



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2014년05월21일
(11) 등록번호 10-1393679
(24) 등록일자 2014년05월02일

(51) 국제특허분류(Int. Cl.)
C12N 15/29 (2006.01)
(21) 출원번호 10-2007-7023530
(22) 출원일자(국제) 2006년01월13일
심사청구일자 2011년01월13일
(85) 번역문제출일자 2007년10월12일
(65) 공개번호 10-2007-0116881
(43) 공개일자 2007년12월11일
(86) 국제출원번호 PCT/US2006/002516
(87) 국제공개번호 WO 2006/101584
국제공개일자 2006년09월28일
(30) 우선권주장
60/662,224 2005년03월15일 미국(US)
(56) 선행기술조사문헌
US6184018 B1
전체 청구항 수 : 총 27 항

(73) 특허권자
비피 코포레이션 노쓰 아메리카 인코포레이티드
미국 77079 텍사스주 휴스턴 웨스트레이크 파크
블러바드 501
(72) 발명자
블럼 데이비드
미국 37205 테네시주 내쉬빌 타가트 애비뉴 118
캠쉬 조슬린
미국 92121 캘리포니아주 샌 디에고 디렉터스 플
레이스 4955
다이카이코 마크
미국 92129 캘리포니아주 샌 디에고 라 트루카 스
트리트 14208
(74) 대리인
김성기, 김진희

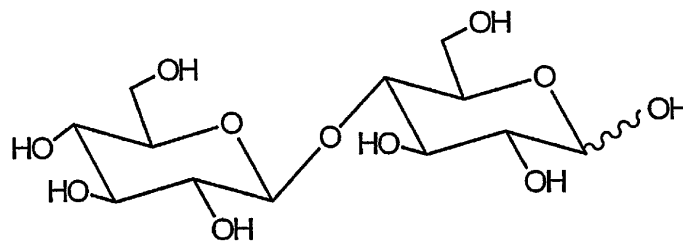
심사관 : 이효진

(54) 발명의 명칭 셀룰라제, 이것을 암호화하는 핵산 및 이들을 제조 및사용하는 방법

(57) 요약

본 발명은 분자 생물학, 세포 생물학 및 생화학 분야에 관한 것이다. 하나의 측면에서, 본 발명은 셀룰라제 활성 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 β -글루코시다제 활성을 보유하는 폴리펩티드, 이러한 폴리펩티드를 암호화하는 폴리뉴클레오티드, 그리고 이러한 폴리뉴클레오티드 및 폴리펩티드를 제조 및 사용하는 방법을 제공한다. 하나의 측면에서, 본 발명은 셀룰라제 활성 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 β -글루코시다제 활성 예를 들어, 열 안정성(thermostable) 및 열 내성(thermotolerant)과 같은 활성을 보유하는 폴리펩티드, 이러한 효소를 암호화하는 폴리뉴클레오티드, 그리고 이러한 폴리뉴클레오티드 및 폴리펩티드를 제조 및 사용하는 방법에 관한 것이다. 본 발명의 폴리펩티드는 약학 분야, 농업 분야 및 산업 분야 등과 같이 다양한 분야에서 사용될 수 있다.

대표도 - 도5



셀로비오스의 구조

특허청구의 범위

청구항 1

서열 번호 1의 핵산 서열로 이루어지고, 셀룰라제 활성을 갖는 하나 이상의 폴리펩티드를 암호화하는 분리된 또는 재조합 핵산.

청구항 2

삭제

청구항 3

삭제

청구항 4

삭제

청구항 5

삭제

청구항 6

삭제

청구항 7

삭제

청구항 8

삭제

청구항 9

삭제

청구항 10

삭제

청구항 11

삭제

청구항 12

삭제

청구항 13

삭제

청구항 14

삭제

청구항 15

삭제

청구항 16

삭제

청구항 17

삭제

청구항 18

삭제

청구항 19

삭제

청구항 20

삭제

청구항 21

삭제

청구항 22

삭제

청구항 23

삭제

청구항 24

삭제

청구항 25

삭제

청구항 26

삭제

청구항 27

삭제

청구항 28

삭제

청구항 29

삭제

청구항 30

삭제

청구항 31

삭제

청구항 32

삭제

청구항 33

삭제

청구항 34

삭제

청구항 35

삭제

청구항 36

삭제

청구항 37

제1항에 기재된 핵산 서열을 포함하는 발현 카세트, 벡터 또는 클로닝 비이클.

청구항 38

삭제

청구항 39

삭제

청구항 40

제1항에 기재된 핵산 서열을 포함하는 형질전환 세포.

청구항 41

제1항에 기재된 서열을 포함하는 트랜스제닉 식물 또는 트랜스제닉 종자.

청구항 42

삭제

청구항 43

삭제

청구항 44

삭제

청구항 45

삭제

청구항 46

삭제

청구항 47

삭제

청구항 48

서열 번호 2의 아미노산 서열로 이루어지고, 셀룰라제, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및 베타-글루코시다제 활성으로 이루어진 군에서 선택되는 하나 이상의 활성을 갖는 분리된 또는 재조합 폴리펩티드.

청구항 49

삭제

청구항 50

삭제

청구항 51

삭제

청구항 52

삭제

청구항 53

삭제

청구항 54

삭제

청구항 55

삭제

청구항 56

삭제

청구항 57

삭제

청구항 58

삭제

청구항 59

삭제

청구항 60

삭제

청구항 61

삭제

청구항 62

삭제

청구항 63

삭제

청구항 64

삭제

청구항 65

삭제

청구항 66

삭제

청구항 67

삭제

청구항 68

삭제

청구항 69

삭제

청구항 70

삭제

청구항 71

삭제

청구항 72

삭제

청구항 73

삭제

청구항 74

삭제

청구항 75

삭제

청구항 76

삭제

청구항 77

삭제

청구항 78

삭제

청구항 79

삭제

청구항 80

삭제

청구항 81

삭제

청구항 82

삭제

청구항 83

삭제

청구항 84

삭제

청구항 85

삭제

청구항 86

삭제

청구항 87

삭제

청구항 88

삭제

청구항 89

삭제

청구항 90

삭제

청구항 91

삭제

청구항 92

삭제

청구항 93

삭제

청구항 94

삭제

청구항 95

삭제

청구항 96

삭제

청구항 97

삭제

청구항 98

삭제

청구항 99

삭제

청구항 100

삭제

청구항 101

삭제

청구항 102

삭제

청구항 103

삭제

청구항 104

삭제

청구항 105

삭제

청구항 106

삭제

청구항 107

삭제

청구항 108

삭제

청구항 109

삭제

청구항 110

삭제

청구항 111

삭제

청구항 112

삭제

청구항 113

삭제

청구항 114

삭제

청구항 115

삭제

청구항 116

삭제

청구항 117

삭제

청구항 118

삭제

청구항 119

삭제

청구항 120

삭제

청구항 121

삭제

청구항 122

삭제

청구항 123

삭제

청구항 124

삭제

청구항 125

삭제

청구항 126

(a) 이중 핵산 서열을 식물 세포 내로 도입함으로써 형질전환된 식물 세포를 생성하는 단계로서, 상기 이중 핵산 서열은 제1항에 기재된 서열로 이루어지는 것인 단계; 및

(b) 형질전환된 세포로부터 트랜스제닉 식물을 생성하는 단계

를 포함하는, 트랜스제닉 식물의 제조 방법.

청구항 127

(a) 프로모터에 작동 가능하게 연결된, 제1항에 기재된 서열로 이루어진 이중 핵산 서열로 식물 세포를 형질전환시키는 단계; 및

(b) 상기 이중 핵산 서열이 식물 세포 내에서 발현되는 조건 하에서 식물을 성장시키는 단계

를 포함하는, 식물 세포 내에서 이중 핵산 서열을 발현시키는 방법.

청구항 128

글루칸, 헤미셀룰로스, 셀룰로스, 리그닌 및 당류로 이루어진 군에서 선택되는 하나 이상을 함유하는 조성물을 가수분해, 분해 또는 파괴하는 방법으로서,

(a) 제48항에 기재된 셀룰라제 활성을 갖는 폴리펩티드, 또는 제1항에 기재된 핵산에 의해 암호화되는 폴리펩티드를 제공하는 단계;

(b) 셀룰로스, 헤미셀룰로스, 리그닌, 당류 및 글루칸으로 이루어진 군에서 선택되는 하나 이상을 함유하는 조성물을 제공하는 단계; 및

(c) 셀룰라제가 글루칸, 헤미셀룰로스, 셀룰로스, 리그닌 및 당류로 이루어진 군에서 선택되는 하나 이상을 함유하는 조성물을 가수분해하거나 분해하거나 파괴하는 조건 하에서, 단계 (a)의 폴리펩티드를 단계 (b)의 조성물과 접촉시키는 단계

를 포함하는 방법.

청구항 129

삭제

청구항 130

삭제

청구항 131

삭제

청구항 132

삭제

청구항 133

삭제

청구항 134

삭제

청구항 135

삭제

청구항 136

삭제

청구항 137

삭제

청구항 138

삭제

청구항 139

삭제

청구항 140

바이오매스, 종이, 목재 또는 목재 제품을 제48항에 기재된 셀룰라제 활성을 갖는 폴리펩티드 또는 제1항에 기재된 핵산에 의해 암호화되는 폴리펩티드와 접촉시키는 단계를 포함하는, 바이오매스, 종이, 목재 또는 목재 제품 내 셀룰로스, 헤미셀룰로스 또는 리그닌의 양을 감소시키는 방법.

청구항 141

삭제

청구항 142

삭제

청구항 143

삭제

청구항 144

셀룰로스, 헤미셀룰로스, 리그닌 및 발효성 당으로 이루어진 군에서 선택되는 하나 이상을 함유하는 조성물을 제48항에 기재된 폴리펩티드 또는 제1항에 기재된 핵산에 의해 암호화되는 폴리펩티드와 접촉시키는 단계를 포함하는, 연료의 제조 방법.

청구항 145

삭제

청구항 146

제48항에 기재된 폴리펩티드 또는 제1항에 기재된 핵산에 의해 암호화되는 폴리펩티드를 포함하는, 셀룰로스가 및 헤미셀룰로스가 중합체를 대사 가능한 탄소 부분으로 탈중합시키기 위한 효소 집합물(enzyme ensemble).

청구항 147

바이오매스를 제48항에 기재된 폴리펩티드 또는 제1항에 기재된 핵산에 의해 암호화되는 폴리펩티드와 접촉시키는 단계를 포함하는, 바이오매스의 효소적 분해 방법.

청구항 148

삭제

청구항 149

삭제

청구항 150

삭제

청구항 151

삭제

청구항 152

삭제

청구항 153

삭제

청구항 154

삭제

청구항 155

삭제

청구항 156

삭제

청구항 157

삭제

청구항 158

삭제

청구항 159

삭제

청구항 160

삭제

청구항 161

제126항에 있어서, 단계 (a)가 식물 세포 원형질체의 전기천공 또는 미세주입에 의해 이중 핵산 서열을 도입하는 것을 추가로 포함하는 것인 제조 방법.

청구항 162

제126항에 있어서, 단계 (a)가 DNA 입자 충격에 의해 또는 아그로박테리움 튜메페이선스(*Agrobacterium tumefaciens*) 숙주를 이용함으로써 이중 핵산 서열을 식물 조직에 직접 도입하는 것을 포함하는 것인 제조 방법.

청구항 163

제128항에 있어서, 상기 조성물이 식물 세포, 박테리아 세포 또는 효모 세포를 포함하는 것인 방법.

청구항 164

제128항에 있어서, 상기 폴리펩티드가 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및 베타-글루코시다제 활성으로 이루어진 군에서 선택되는 하나 이상의 활성을 갖는 것인 방법.

청구항 165

제140항에 있어서, 상기 셀룰라제 활성이 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및 베타-글루코시다제 활성으로 이루어진 군에서 선택되는 하나 이상의 활성을 포함하는 것인 방법.

청구항 166

제144항에 있어서, 셀룰로스, 헤미셀룰로스, 리그닌 및 발효성 당으로 이루어진 군에서 선택되는 하나 이상을 함유하는 조성물이 식물, 식물 생성물 또는 식물 유래물을 포함하는 것인 제조 방법.

청구항 167

제166항에 있어서, 상기 식물 또는 식물 생성물이 사탕수수 식물 또는 식물 생성물, 비트 또는 사탕무, 밀, 옥수수, 대두, 감자, 벼 또는 보리를 포함하는 것인 제조 방법.

청구항 168

제144항에 있어서, 상기 폴리펩티드가 셀룰라제, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및 베타-글루코시다제 활성으로 이루어진 군에서 선택되는 하나 이상의 활성을 갖는 것인 제조 방법.

청구항 169

제144항에 있어서, 상기 연료가 바이오에탄올 또는 가솔린-에탄올 혼합물을 포함하는 것인 제조 방법.

청구항 170

제146항에 있어서, 상기 폴리펩티드가 셀룰라제, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및 베타-글루코시다제 활성으로 이루어진 군에서 선택되는 하나 이상의 활성을 갖는 것인 효소 집합물.

청구항 171

제147항에 있어서, 상기 바이오매스가 농작물로부터 유래하거나, 식품 또는 사료 생산 시의 부산물이거나, 리그노셀룰로오스계, 셀룰로오스계 또는 리그닌 폐기물이거나, 식물 잔해 또는 폐지 또는 폐지 제품인 방법.

청구항 172

제147항에 있어서, 상기 폴리펩티드가 셀룰라제, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및 베타-글루코시다제 활성으로 이루어진 군에서 선택되는 하나 이상의 활성을 갖는 것인 방법.

청구항 173

제171항에 있어서, 상기 식물 잔해가 줄기, 잎, 외피, 껍질, 속대, 목재, 목재 칩, 목재 펄프 및 톱밥을 포함하는 것인 방법.

청구항 174

제171항에 있어서, 상기 폐지가 폐기된 또는 사용이 끝난 복사 용지, 컴퓨터 인쇄 용지, 공책 종이, 메모지, 타자 용지, 신문, 잡지, 판지 및 종이계 포장재를 포함하는 것인 방법.

청구항 175

제147항에 있어서, 바이오매스 물질의 효소적 분해에 의해 바이오에탄올이 생성되는 것인 방법.

명세서

기술 분야

[0001]

정부 지원

[0002]

본 발명은 미국 에너지국과 체결된 계약 즉, DE-FG03-02ER83395 및 DE-FG02-03ER83865 하에서 미국 정부의 지원으로 이루어진 것이다. 정부는 본 발명에 대해 일정 권리를 갖는다.

[0003]

발명의 분야

[0004]

본 발명은 분자 생물학, 세포 생물학 및 생화학 분야에 관한 것이다. 하나의 측면에서, 본 발명은 셀룰라제 활성 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 β -글루코시다제 활성을 보유하는 폴리펩티드, 이러한 폴리펩티드를 암호화하는 폴리뉴클레오티드, 그리고 이러한 폴리뉴클레오티드 및 폴리펩티드를 제조 및 사용하는 방법을 제공한다. 하나의 측면에서, 본 발명은 셀룰라제 활성 예를 들어, 엔도글루카나

제, 셀로바이오하이드롤라제, 만난아제 및/또는 β -글루코시다제 활성 예를 들어, 열 안정성(thermostable) 및 열 내성(thermotolerant)과 같은 활성을 보유하는 폴리펩티드, 이러한 효소를 암호화하는 폴리뉴클레오티드, 그리고 이러한 폴리뉴클레오티드 및 폴리펩티드를 제조 및 사용하는 방법에 관한 것이다. 본 발명의 폴리펩티드는 약학 분야, 농업 분야 및 산업 분야 등과 같이 다양한 분야에서 사용될 수 있다.

배경 기술

- [0005] 셀룰로스는 지구상에서 가장 풍부한 재생 가능 자원이다. 이 셀룰로스는 셀로바이오스(도 5에 나타낸 바와 같은 구조를 가지는 글루코스 이량체)라는 반복 단위를 가지는 β -1,4 글루코스 단위의 선형 사슬로 이루어져 있다. 상기 중합체는 셀룰로스계 중합체를 랜덤하게 가수분해하는 엔도글루카나제(EG), 그리고 셀룰로스로부터 말단 셀로바이오스 잔기를 제거하는 셀로바이오하이드롤라제(CBH)를 포함하는 일군의 효소에 의해 분해된다. 셀로바이오스 및 셀로-올리고당은 β -글루코시다제(BG)에 의해 글루코스로 가수분해된다. 상기 3가지 효소들 모두는 셀룰로스를 글루코스로 완전히 분해하는데에 필요한 효소이다. 상기 3가지 효소들은 각각 기능은 동일하나 구조가 상이한 변이체 관계이다. 뿐만 아니라, 진균 및 박테리아는 상이한 구조적 변이체 이외에도, 동일한 구조적 변이체를 다양한 형태로 생산해 내는 것으로 알려져 있다.
- [0006] 또한, 이 시스템을 종합해보면, 일부 혐기성 박테리아 및 진균은 상기와 같은 효소들을 다수의 효소 복합체 즉, 다수의 효소들이 분자량 2백만 달톤 이상인 효소 골격에 모두 결합되어 있는 복합체의 형태로서 생산하는 것을 알 수 있다. 그렇다면, 단순한 형태의 분자를 생산하는데에 이러한 효소 복합체 시스템이 필요한 이유는 무엇일까? 일부 연구자들은 이와 같이 복잡한 성질은 기질의 저항성으로 인한 것이라고 생각하고 있다. 셀룰로스 사슬은 인접 사슬 간 수소 결합에 의해 결정질의 매트릭스로 패키징되는 미세 섬유를 형성한다. 이러한 구조는 화학 분해 또는 효소 분해에 대해 매우 저항성이다.
- [0007] CBH는 이것의 셀룰로스에 대한 효소 공격 특성으로 인하여, 이와 같은 결정질 셀룰로스의 분해에 중요한 역할을 가지는 효소인 것으로 생각된다. CBH와는 달리, EG는 90° 로 셀룰로스 사슬을 공격하는 개방형 틈(open cleft)을 가진다. CBH는 활성 부위를 포함하는 터널을 통해 직접 사슬을 공격한다. 현재, 학자들은 셀룰로스 사슬이 상기 터널에 들어감과 동시에, 인접한 수소 결합이 파괴되는 것으로 추측하고 있다. 일단 셀로바이오하이드롤라제가 이와 같이 기질 상에서 "발판(foothold)"을 마련하면, 이후 상기 EG는 기질과 결합하고, 보다 용이하게 이 기질을 공격할 수 있게 된다.
- [0008] 공지의 CBH에 있어서 관찰되는 가장 큰 결점은 촉매 활성이 낮다는 점이다. 일부 학자들은 이와 같이 낮은 활성은, 가수분해로 인한 에너지가 역학적 에너지로 바뀌어 수소 결합이 파괴되고, 이에 따라서 이 효소가 기질을 따라서 이동할 수 있게 된다는 사실로부터 기인한 것이라고 주장하고 있다. CBH는 엑소-작용성(exo-acting) 효소로서, 글리코실 가수분해 효소의 90개 군 중 6개 군에서 발견되는 것이다. 이 CBH는 제5군, 제6군, 제7군, 제9군, 제10군 및 제48군을 포함한다. 상기 제5군은 다수의 상이한 형태의 글리코실 가수분해 효소 예를 들어, 셀룰라제, 만난아제 및 자일라제를 포함한다. 이 군에 속하는 대부분의 셀룰라제는 엔도글루카나제이지만, 주로 클로스트리듐 서모셀럼(*Clostridium thermocellum*)으로부터 유래하는, Cel10와 같은 셀로바이오하이드롤라제도 존재한다. 제6군은 엔도글루카나제 또는 셀로바이오하이드롤라제만을 포함하는데, 이 군은 엔도글루카나제보다는 셀로바이오하이드롤라제 일원을 더욱 많이 포함한다. 상기 효소들은 반전 기작을 가지며, 결정학적 분석 결과 이 효소는 8개가 아닌, 7개의 평행 β -사슬을 포함하는 왜곡된 α/β 원통 모양의 구조를 가지는 것으로 밝혀졌다. 제7군 효소는 또한 엔도글루카나제와 셀로바이오하이드롤라제 이 두 가지로 이루어져 있으며(셀로바이오하이드롤라제가 더욱 많음), 유일하게 공지된 일원은 진균으로부터 유래하는 것이다. 상기 효소는 보유 기작을 가지며, 결정 구조는 β -셀리룰 구조를 가진다고 한다. 제9군은 엔도글루카나제, 셀로바이오하이드롤라제 및 β -글루코시다제를 포함하는데, 이 군은 엔도글루카나제를 많이 포함한다. 그러나, 서모비피다 푸스카(*Thermobifida fusca*)는 엔도/엑소-1,4-글루카나제를 생산하는데, 이것의 결정 구조는 $(\alpha/\alpha)_6$ 원통 구조로 폴딩되어 있다고 한다. 이 효소는 엔도- 및 엑소-글루카나제 CBH 둘 다의 특성을 가진다. 제10군은 2개의 일원들 즉, 주를 이루는 셀로바이오하이드롤라제 하나와, 나머지 자일라나제 만을 포함한다. 제10군에 속하는 셀로바이오하이드롤라제 및 자일라나제는 메틸-옴베리페릴 셀로바이오시드에 대해 활성을 가진다. 제48군은 주로 박테리아 및 혐기성 진균 셀로바이오하이드롤라제 및 엔도글루카나제를 포함한다. 이것의 구조는 제9군과 유사하게 $(\alpha/\alpha)_6$ 원통 구조로 폴딩되어 있다.
- [0009] 저렴하고 재생 가능한 차량용 연료의 공급원이 필요한 시대이다. 새로운 연료 공급원은 그것이 연소된 후 무해한 최종 생성물이 생성된다면 금상첨화일 것이다. 에탄올은 석유를 주원료로 하는 연료에 대한 매력적인 대안이

되며, 전분이나 리그노셀룰로스로부터 유래하는 당당류를 발효시킴으로써 얻을 수 있다. 그러나, 에탄올을 생산하는데에는 비용이 많이 들기 때문에, 현재의 경제 사정으로는 이 에탄올 연료를 보편적으로 사용할 수 없다. 비용을 낮추는 것에 초점을 맞춘 연구의 한 분야는 리그노셀룰로스로부터 발효성 당을 생산하는데에 사용할 수 있는 효소의 기술적 효율을 극대화시키는 것이다. 공급 원료를 보다 효율적으로 분해하는 효소의 개발을 통하여, 에탄올 생산 비용을 줄일 수 있을 것으로 생각된다. 보다 효율적인 공정에 의해서도, 미국이 수입 원유에 의존하는 정도와 이러한 의존성과 관련이 있을 수 있는 가격 변동률을 낮출 것이다. 뿐만 아니라, 바이오에탄올과 같은 운송용 청정 연료를 사용하면 지구 온난화에 일부 책임이 있는 것으로 생각되는 CO₂ 방출량도 전체적으로 줄일 수 있다.

[0010] 발명의 개요

[0011] 본 발명은 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제 및/또는 β -글루코시다제(베타-글루코시다제), 그리고 이들을 제조 및 사용하는 방법을 제공한다. 하나의 측면에서, 본 발명의 효소는 촉매 반응 속도가 증가하여, 기질의 가수분해 과정을 개선시킬 수 있다. 이와 같이 촉매 반응 속도의 효율이 증가하면, 산업 분야에서 유용할 수 있는 당을 생산함에 있어서 효율이 증가하게 되는데, 예를 들어, 이렇게 생산된 당은 에탄올을 생산하는 미생물에 의해 사용될 수 있다. 하나의 측면에서, 본 발명은 활성이 높은(예를 들어, 촉매 반응 속도가 증가한) 셀로바이오하이드롤라제, 엔도글루카나제 및 베타-글루코시다제를 제공한다. 본 발명은 효소 비용을 낮춘(예를 들어, 바이오매스(biomass)를 에탄올 공정으로 도입시키는데 드는 비용을 낮춘) 본 발명의 효소를 사용하는 산업 분야에서의 용도(예를 들어, 바이오매스에서 에탄올로 전환시키는 산업 분야에서의 용도)를 제공한다. 그러므로, 본 발명은 임의의 바이오매스로부터 유래하는 바이오에탄올을 포함하는 연료를 비롯한, 바이오에탄-포함 조성물을 및 바이오에탄올을 생산하는 효율적인 방법을 제공한다.

[0012] 하나의 측면에서, 본 발명의 효소는 글루카나제 활성 예를 들어, 엔도글루카나제 활성 예를 들어, 내부 엔도- β -1,4- 및/또는 β -1,3-글루카나제 결합의 가수분해를 촉진하는 활성을 가진다. 하나의 측면에서, 엔도글루카나제 활성(예를 들어, 엔도-1,4-베타-D-글루칸 4-글루카노 하이드롤라제 활성)은, 셀룰로스, 셀룰로스 유도체(예를 들어, 카복시 메틸 셀룰로스 및 하이드록시 에틸 셀룰로스) 리헤닌의 1,4- 및/또는 β -1,3-베타-D-글루코시드 결합, 혼합형 베타-1,3 글루칸 예를 들어, 곡류의 베타-D-글루칸 또는 자일로글루칸과, 셀룰로스 부분을 함유하는 기타 식물 물질에 존재하는 베타-1,4 결합의 가수분해를 포함한다.

[0013] 하나의 측면에서, 본 발명의 효소는 엔도글루카나제(예를 들어, 엔도-베타-1,4-글루카나제, EC 3.2.1.4; 엔도-베타-1,3(1)-글루카나제, EC 3.2.1.6; 엔도-베타-1,3-글루카나제, EC 3.2.1.39) 활성을 보유하며, 셀룰로스 및 글루칸 내에 존재하는 내부 β -1,4- 및/또는 β -1,3-글루코시드 결합을 가수분해하여, 보다 작은 분자량의 글루코스 및 글루코스 올리고머를 생산할 수 있다. 본 발명은 본 발명의 효소를 사용하여 보다 작은 분자량의 글루코스 및 글루코스 올리고머를 생산하는 방법을 제공한다.

[0014] 하나의 측면에서, 본 발명의 효소는 글루칸 예를 들어, 1,4- β - 및/또는 1,3-글리코시드 결합 D-글루코피라노스로부터 형성된 다당류를 생산하는데에 사용된다. 하나의 측면에서, 본 발명의 엔도글루카나제는 식품 산업 예를 들어, 제빵 산업, 과일 및 야채의 가공 과정, 농업용 폐기물을 분해하는 과정, 동물의 사료를 제조하는 경우, 펄프 생산 및 제지, 직물 제조, 그리고 가정용 및 산업용 세정제의 생산에 사용된다. 하나의 측면에서, 본 발명의 효소 예를 들어, 엔도글루카나제는 미생물 예를 들어, 진균 및/또는 박테리아에 의해 생산된다.

[0015] 하나의 측면에서, 본 발명의 효소 예를 들어, 엔도글루카나제는 곡류의 주요 비-전분 다당류인 베타-글루칸(β -글루칸)을 가수분해하는데에 사용된다. 다당류의 글루칸 함량은 주로 다양성과 성장 조건에 따라서 다양할 수 있다. 이러한 다당류의 생리 화학적 특성으로 인하여, 산화 조건 하에서 점성의 용액 또는 겔로 만들어 질 수 있다. 뿐만 아니라, 글루칸은 물과 결합하는 특성이 매우 강하다. 이와 같은 특성들 모두는 몇몇 산업 분야 예를 들어, 양조, 제빵 및 동물의 영양학 분야에서 단점으로 작용하고 있다. 양조 분야에 있어서, 글루칸이 존재하면, 맥아즙 여과 능이 불량해지고 헤이즈(haze)가 형성된다는 문제점이 있다. 제빵 분야(특히, 쿠키 및 크래커 제조)에 있어서는, 글루칸이 존재하면 반죽에 끈기가 생겨, 기계로 다루기 어려워지고 비스킷 크기도 작아질 수 없게 된다. 그러므로, 본 발명의 효소 예를 들어, 엔도글루카나제를 사용하여 β -글루칸 포함 조성물 내 β -글루칸의 양을 줄일 수 있는데, 예를 들어, 본 발명의 효소를 용액 또는 겔의 점성을 줄이는 공정; 조성물 예를 들어, β -글루칸-포함 조성물이 물과 결합하는 능력을 줄이는 공정; 그리고 양조 분야에 있어서는, (예를 들어, 맥아즙 여과능을 증가시키고 헤이즈 형성을 감소시키기 위해서) 반죽 예를 들어, 쿠키, 빵 및 비스킷 등의 제조용 반죽의 끈기를 감소시키는 공정에 사용할 수 있다.

[0016] 뿐만 아니라, 탄수화물(예를 들어, β -글루칸)은 제과 제품을 빠른 속도로 재수화시켜 바삭함을 상실시킬 뿐만

아니라, 유통 기한도 감소시키는 것과 관련되어 있다. 그러므로, 본 발명의 효소 예를 들어, 엔도글루카나제는 바삭함을 유지시키거나, 이를 증가시키거나, 또는 이 바삭함이 없어지는 속도를 늦출 뿐만 아니라, 임의의 탄수화물-함유 식품, 사료 또는 음료 예를 들어, β -글루칸-포함 식품, 사료 또는 음료의 유통 기한을 증가시키는데에 사용된다.

[0017] 본 발명의 효소 예를 들어, 엔도글루카나제는 장 내용물(예를 들어, 동물의 경우에는 반추 동물 또는 인간의 장 내용물) 예를 들어, 곡류 식을 포함하는 내용물의 점성을 감소시키는데에 사용된다. 그러므로, 대안적 측면에서, 본 발명의 효소 예를 들어, 엔도글루카나제는 식품 또는 사료의 소화 능력과, 동물(예를 들어, 인간 또는 가축)의 성장 속도에 긍정적인 영향을 미치는데에 사용되고, 하나의 측면에서는, 사료 전환 효율을 더욱 높이는데에 사용된다. 곡류 식을 이용하는 단위 동물(monogastric animal)의 사료에 있어서, 베타-글루칸은 장 내용물의 점성에 영향을 미치는 요인이므로, 사료의 소화 능력과 동물의 성장 속도에 악영향을 미치게 된다. 반추 동물에 있어서, 베타-글루칸은 섬유 섭취물의 실질적인 성분으로서, 글루칸이 보다 완전히 소화되면, 사료의 전환 효율을 보다 높이는데에 기여할 것이다. 그러므로, 본 발명은 본 발명의 엔도글루카나제를 포함하는 동물 사료 및 식품을 제공하고, 하나의 측면에서, 이 효소는 동물의 소화관 예를 들어, 위 및/또는 소장 내에서 활성을 갖는다.

[0018] 본 발명의 효소 예를 들어, 엔도글루카나제는 셀룰로스 또는 임의의 베타-1,4-결합 글루칸-포함 합성 또는 천연 물질 예를 들어, 임의의 식물 물질에서 발견되는 물질을 분해하는데에 사용된다. 본 발명의 효소 예를 들어, 엔도글루카나제는 예를 들어, 목재 가공, 펄프 및/또는 제지 산업, 직물 제조, 그리고 가정용 및 산업용 세정제, 및/또는 바이오매스 폐기물 처리에 있어서 셀룰로스를 분해하는 산업용 효소로서 사용된다.

[0019] 하나의 측면에서, 본 발명은 본 발명의 효소, 폴리펩티드 또는 폴리뉴클레오티드를 포함하는 조성물(예를 들어, 약학 조성물, 식품, 사료, 약품, 식품 보조제)을 제공하는 것이다. 이러한 조성물은 다양한 형태 예를 들어, 정제, 겔, 환제, 임플란트, 액체, 스프레이, 분말, 식품, 사료 펠릿 또는 캡슐화된 임의의 형태의 것으로서 제형화될 수 있다.

[0020] 본 발명은 적어도 약 10, 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, 2000, 2050, 2100, 2200, 2250, 2300, 2350, 2400, 2450, 2500개 이상의 잔기로 이루어진 부위에 걸쳐서, 본 발명의 대표적인 핵산 예를 들어, 서열 번호 1, 서열 번호 3, 서열 번호 5, 서열 번호 7, 서열 번호 9, 서열 번호 11, 서열 번호 13, 서열 번호 15, 서열 번호 17, 서열 번호 19, 서열 번호 21, 서열 번호 23, 서열 번호 25, 서열 번호 27, 서열 번호 29, 서열 번호 31, 서열 번호 33, 서열 번호 35, 서열 번호 37, 서열 번호 39, 서열 번호 41, 서열 번호 43, 서열 번호 45, 서열 번호 47, 서열 번호 49, 서열 번호 51, 서열 번호 53, 서열 번호 55, 서열 번호 57, 서열 번호 59, 서열 번호 61, 서열 번호 63, 서열 번호 65, 서열 번호 67, 서열 번호 69, 서열 번호 71, 서열 번호 73, 서열 번호 75, 서열 번호 77, 서열 번호 79, 서열 번호 81, 서열 번호 83, 서열 번호 85, 서열 번호 87, 서열 번호 89, 서열 번호 91, 서열 번호 93, 서열 번호 95, 서열 번호 97, 서열 번호 99, 서열 번호 101, 서열 번호 103, 서열 번호 105, 서열 번호 107, 서열 번호 109, 서열 번호 111, 서열 번호 113, 서열 번호 115, 서열 번호 117, 서열 번호 119, 서열 번호 121, 서열 번호 123, 서열 번호 125, 서열 번호 127, 서열 번호 129, 서열 번호 131, 서열 번호 133, 서열 번호 135, 서열 번호 137, 서열 번호 139, 서열 번호 141, 서열 번호 143, 서열 번호 145, 서열 번호 147, 서열 번호 149, 서열 번호 151, 서열 번호 153, 서열 번호 155, 서열 번호 157, 서열 번호 159, 서열 번호 161, 서열 번호 163 및 서열 번호 165와의 서열 동일성이 적어도 약 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 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%, 99% 이상이거나 완벽한(100%) 서열 동일성을 갖는 핵산 서열을 포함하는 분리된 또는 재조합 핵산을 제공하고[이하, 표 1a 및 1b, 2 및 3과, 실시예 1 및 4, 그리고 서열 목록 참조]; 대안적인 측면에서, 이러한 핵산은 셀룰라제 활성 예를 들어, 엔도글루카나제 활성, 셀로바이오히이드롤라제 활성, 만난아제 활성 및/또는 베타-글루코시다제 활성을 보유하는 하나 이상의 폴리펩티드를 암호화하거나, 또는 본 발명의 폴리펩티드에 특이적으로 결합할 수 있는 항체를 생산할 수 있는 폴리펩티드를 암호화하고, 또는 상기 핵산은 셀룰라제 암호화 핵산을 동정 또는 분리하는데에 있어서 프로브로서 사용될 수 있거나, 또는 셀룰라제-발현 핵산의 발현을 억제하는데에 사용될 수 있다[상기 모든 측면들은 "본 발명의 핵산"에 관한 것임]. 하나의 측면에서, 상기 서열의 상동성은 서열 비교 알고리즘을 이용하는 분석법이나 또는 시각적 조사에 의해 결정된다.

[0021] 본 발명의 핵산은 또한 본 발명의 대표적인 효소, 예를 들어 서열 번호 2, 서열 번호 4, 서열 번호 6, 서열 번호 8, 서열 번호 10, 서열 번호 12, 서열 번호 14, 서열 번호 16, 서열 번호 18, 서열 번호 20, 서열 번호 22, 서열 번호 24, 서열 번호 26, 서열 번호 28, 서열 번호 30, 서열 번호 32, 서열 번호 34, 서열 번호 36, 서열 번호 38, 서열 번호 40, 서열 번호 42, 서열 번호 44, 서열 번호 46, 서열 번호 48, 서열 번호 50, 서열 번호 52, 서열 번호 54, 서열 번호 56, 서열 번호 58, 서열 번호 60, 서열 번호 62, 서열 번호 64, 서열 번호 66, 서열 번호 68, 서열 번호 70, 서열 번호 72, 서열 번호 74, 서열 번호 76, 서열 번호 78, 서열 번호 80, 서열 번호 82, 서열 번호 84, 서열 번호 86, 서열 번호 88, 서열 번호 90, 서열 번호 92, 서열 번호 94, 서열 번호 96, 서열 번호 98, 서열 번호 100, 서열 번호 102, 서열 번호 104, 서열 번호 106, 서열 번호 108, 서열 번호 110, 서열 번호 112, 서열 번호 114, 서열 번호 116, 서열 번호 118, 서열 번호 120, 서열 번호 122, 서열 번호 124, 서열 번호 126, 서열 번호 128, 서열 번호 130, 서열 번호 132, 서열 번호 134, 서열 번호 136, 서열 번호 138, 서열 번호 140, 서열 번호 142, 서열 번호 144, 서열 번호 146, 서열 번호 148, 서열 번호 150, 서열 번호 152, 서열 번호 154, 서열 번호 156, 서열 번호 158, 서열 번호 160, 서열 번호 162, 서열 번호 164 및 서열 번호 166에 제시된 서열을 가지는 폴리펩티드를 암호화하는 분리된 또는 재조합 핵산을 포함한다[이하의 표 1a~1c, 2 및 3과, 실시예 1 및 4, 그리고 서열 목록과 이 서열 목록에 제시된 부분 서열, 그리고 이의 변이체 참조]. 하나의 측면에서, 본 발명의 폴리펩티드는 셀룰라제 활성, 예를 들어 엔도글루카나제 활성, 셀로바이오하이드롤라제 활성, 만난아제 활성 및/또는 베타-글루코시다제 활성을 보유한다.

[0022] 하나의 측면에서, 본 발명은 혼합 배양액으로부터 유래하는 핵산이라는 점에서 공통적인 신규성을 가지는 셀룰라제 암호화 핵산 예를 들어, 엔도글루카나제-암호화 핵산, 셀로바이오하이드롤라제-암호화 핵산 및/또는 베타-글루코시다제-암호화 핵산을 제공한다. 본 발명은 적어도 약 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100 또는 1150개 이상의 잔기로 이루어진 부위에 걸쳐서, 본 발명의 폴리뉴클레오티드 예를 들어, 본 발명의 대표적인 핵산 예를 들어, 서열 번호 1, 서열 번호 3, 서열 번호 5, 서열 번호 7, 서열 번호 9, 서열 번호 11, 서열 번호 13, 서열 번호 15, 서열 번호 17, 서열 번호 19, 서열 번호 21, 서열 번호 23, 서열 번호 25, 서열 번호 27, 서열 번호 29, 서열 번호 31, 서열 번호 33, 서열 번호 35, 서열 번호 37, 서열 번호 39, 서열 번호 41, 서열 번호 43, 서열 번호 45, 서열 번호 47, 서열 번호 49, 서열 번호 51, 서열 번호 53, 서열 번호 55, 서열 번호 57, 서열 번호 59, 서열 번호 61, 서열 번호 63, 서열 번호 65, 서열 번호 67, 서열 번호 69, 서열 번호 71, 서열 번호 73, 서열 번호 75, 서열 번호 77, 서열 번호 79, 서열 번호 81, 서열 번호 83, 서열 번호 85, 서열 번호 87, 서열 번호 89, 서열 번호 91, 서열 번호 93, 서열 번호 95, 서열 번호 97, 서열 번호 99, 서열 번호 101, 서열 번호 103, 서열 번호 105, 서열 번호 107, 서열 번호 109, 서열 번호 111, 서열 번호 113, 서열 번호 115, 서열 번호 117, 서열 번호 119, 서열 번호 121, 서열 번호 123, 서열 번호 125, 서열 번호 127, 서열 번호 129, 서열 번호 131, 서열 번호 133, 서열 번호 135, 서열 번호 137, 서열 번호 139, 서열 번호 141, 서열 번호 143, 서열 번호 145, 서열 번호 147, 서열 번호 149, 서열 번호 151, 서열 번호 153, 서열 번호 155, 서열 번호 157, 서열 번호 159, 서열 번호 161, 서열 번호 163 및 서열 번호 165와의 서열 동일성이 적어도 약 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 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%, 99% 이상이거나 완벽한(100%) 서열 동일성을 갖는 서열을 포함하는, 혼합 배양액으로부터 분리된 셀룰로스-분해 효소-암호화 핵산을 제공한다[이하, 표 1a~1c, 2 및 3과, 실시예 1 및 4, 그리고 서열 목록 참조].

[0023] 하나의 측면에서, 본 발명은 공통된 공급원, 예를 들어 환경적 공급원으로부터 유래한다는 점에서 공통된 신규성을 가지는, 셀룰라제 효소 암호화 핵산 예를 들어, 엔도글루카나제 효소 암호화 핵산, 셀로바이오하이드롤라제 효소-암호화 핵산 및/또는 베타-글루코시다제 효소 암호화 핵산, 예를 들어 본 발명의 대표적인 폴리뉴클레오티드 서열[이하의 표 1a~1c, 2 및 3과, 실시예 1 및 4, 그리고 서열 목록 참조], 본 발명의 효소를 비롯한 이 핵산에 의해 암호화되는 폴리펩티드, 예를 들어 본 발명의 대표적인 폴리펩티드, 예를 들어 서열 번호 2, 서열 번호 4, 서열 번호 6, 서열 번호 8, 서열 번호 10, 서열 번호 12, 서열 번호 14, 서열 번호 16, 서열 번호 18, 서열 번호 20, 서열 번호 22, 서열 번호 24, 서열 번호 26, 서열 번호 28, 서열 번호 30, 서열 번호 32, 서열 번호 34, 서열 번호 36, 서열 번호 38, 서열 번호 40, 서열 번호 42, 서열 번호 44, 서열 번호 46, 서열 번호 48, 서열 번호 50, 서열 번호 52, 서열 번호 54, 서열 번호 56, 서열 번호 58, 서열 번호 60, 서열 번호 62, 서열 번호 64, 서열 번호 66, 서열 번호 68, 서열 번호 70, 서열 번호 72, 서열 번호 74, 서열 번호 76, 서열 번호 78, 서열 번호 80, 서열 번호 82, 서열 번호 84, 서열 번호 86, 서열 번호 88, 서열 번호 90, 서열 번호 92, 서열 번호 94, 서열 번호 96, 서열 번호 98, 서열 번호 100, 서열 번호 102, 서열 번호 104, 서열 번호

호 106, 서열 번호 108, 서열 번호 110, 서열 번호 112, 서열 번호 114, 서열 번호 116, 서열 번호 118, 서열 번호 120, 서열 번호 122, 서열 번호 124, 서열 번호 126, 서열 번호 128, 서열 번호 130, 서열 번호 132, 서열 번호 134, 서열 번호 136, 서열 번호 138, 서열 번호 140, 서열 번호 142, 서열 번호 144, 서열 번호 146, 서열 번호 148, 서열 번호 150, 서열 번호 152, 서열 번호 154, 서열 번호 156, 서열 번호 158, 서열 번호 160, 서열 번호 162, 서열 번호 164 또는 서열 번호 166을 제공한다[이하의 표 1a~1c 및 서열 목록 참조]. 하나의 측면에서, 본 발명은 또한 환경적 공급원, 예를 들어 혼합 환경적 공급원으로부터 유래한다는 점에서 공통된 신규성을 가지는, 셀룰라제 효소 암호화 핵산, 예를 들어, 엔도글루카나제 효소 암호화 핵산, 셀로바이오히이드롤라제 효소 암호화 핵산 및/또는 베타-글루코시다제 효소 암호화 핵산을 제공한다.

[0024] 하나의 측면에서, 서열 비교 알고리즘은 BLAST 버전 2.2.2 알고리즘으로서, 필터링 셋팅(filtering setting)이 blastall -p blastp -d"nr pataa" -F F로 설정되고, 다른 모든 선택 사양은 디폴트값으로 설정된 알고리즘이다.

[0025] 본 발명의 다른 측면은 본 발명의 핵산 서열 중 10, 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, 2000, 2050, 2100, 2200, 2250, 2300, 2350, 2400, 2450 또는 2500개 이상의 연속된 염기를 포함하는 분리된 또는 제조합 핵산, 이와 실질적으로 동일한 서열 및 이에 상보적인 서열에 관한 것이다.

[0026] 하나의 측면에서, 본 발명의 분리된 또는 제조합 핵산은 열 안정성인 셀룰라제 활성 예를 들어, 엔도글루카나제 활성, 셀로바이오히이드롤라제 활성, 만난아제 활성 및/또는 베타-글루코시다제 활성을 보유하는 폴리펩티드를 암호화한다. 본 발명의 폴리펩티드는 온도 약 37~약 95℃, 약 55~약 85℃, 약 70~약 95℃, 또는 약 90~약 95℃를 포함하는 조건 하에서 셀룰라제 활성을 보유할 수 있다. 본 발명의 폴리펩티드는 온도 약 1~약 5℃, 약 5~약 15℃, 약 15~약 25℃, 약 25~약 37℃, 약 37~약 95℃, 96℃, 97℃, 98℃ 또는 99℃, 약 55~약 85℃, 약 70~약 75℃, 또는 약 90~약 99℃, 또는 95℃, 96℃, 97℃, 98℃ 또는 99℃ 이상에서 셀룰라제 활성을 보유할 수 있다.

[0027] 다른 측면에서, 본 발명의 분리된 또는 제조합 핵산은 열 내성이며, 셀룰라제 활성 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 활성을 보유하는 폴리펩티드를 암호화한다. 본 발명의 폴리펩티드는 37~약 95℃의 온도 또는 55~약 85℃의 온도 중 임의의 온도에 노출된 후에도 셀룰라제 활성을 유지할 수 있다. 본 발명의 폴리펩티드는 약 1~약 5℃, 약 5~약 15℃, 약 15~약 25℃, 약 25~약 37℃, 약 37~약 95℃, 96℃, 97℃, 98℃ 또는 99℃, 약 55~약 85℃, 약 70~약 75℃, 또는 약 90~약 95℃의 온도에 노출된 후에도 셀룰라제 활성을 유지할 수 있다. 하나의 측면에서, 본 발명의 폴리펩티드는 90~약 99℃, 또는 95℃, 96℃, 97℃, 98℃ 또는 99℃의 온도 및 pH 약 4.5 이상에 노출된 후에도 셀룰라제 활성을 유지할 수 있다.

[0028] 본 발명은 엄중한 조건 하에서 본 발명의 핵산에 혼성화되는 서열을 포함하는 분리된 또는 제조합 핵산을 제공하는데, 여기서, 상기 본 발명의 대표적 핵산의 예로서는 예를 들어, 서열 번호 1, 서열 번호 3, 서열 번호 5, 서열 번호 7, 서열 번호 9, 서열 번호 11, 서열 번호 13, 서열 번호 15, 서열 번호 17, 서열 번호 19, 서열 번호 21, 서열 번호 23, 서열 번호 25, 서열 번호 27, 서열 번호 29, 서열 번호 31, 서열 번호 33, 서열 번호 35, 서열 번호 37, 서열 번호 39, 서열 번호 41, 서열 번호 43, 서열 번호 45, 서열 번호 47, 서열 번호 49, 서열 번호 51, 서열 번호 53, 서열 번호 55, 서열 번호 57, 서열 번호 59, 서열 번호 61, 서열 번호 63, 서열 번호 65, 서열 번호 67, 서열 번호 69, 서열 번호 71, 서열 번호 73, 서열 번호 75, 서열 번호 77, 서열 번호 79, 서열 번호 81, 서열 번호 83, 서열 번호 85, 서열 번호 87, 서열 번호 89, 서열 번호 91, 서열 번호 93, 서열 번호 95, 서열 번호 97, 서열 번호 99, 서열 번호 101, 서열 번호 103, 서열 번호 105, 서열 번호 107, 서열 번호 109, 서열 번호 111, 서열 번호 113, 서열 번호 115, 서열 번호 117, 서열 번호 119, 서열 번호 121, 서열 번호 123, 서열 번호 125, 서열 번호 127, 서열 번호 129, 서열 번호 131, 서열 번호 133, 서열 번호 135, 서열 번호 137, 서열 번호 139, 서열 번호 141, 서열 번호 143, 서열 번호 145, 서열 번호 147, 서열 번호 149, 서열 번호 151, 서열 번호 153, 서열 번호 155, 서열 번호 157, 서열 번호 159, 서열 번호 161, 서열 번호 163 또는 서열 번호 165에 제시된 서열(이하, 표 1a~1c, 2 및 3과, 실시예 1 및 4 참조), 또는 이의 단편이나 부분 서열을 포함한다. 하나의 측면에서, 본 발명의 핵산은 셀룰라제 활성 예를 들어, 엔도글루카나제 활성, 셀로바이오히이드롤라제 활성, 만난아제 활성 및/또는 베타-글루코시다제 활성을 보유하는 폴리펩티드를 암호화한다. 본 발명의 핵산은 길이 약 10, 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 개 이상

인 잔기이거나 또는 전체 길이인 유전자 또는 전사체일 수 있다. 하나의 측면에서, 엄중 조건은 약 65℃의 온도에서 약 15분 동안 0.2×SSC 중에 세척하는 것을 포함하는 세척 단계를 포함한다.

[0029] 본 발명은 셀룰라제 활성 예를 들어, 엔도글루카나제 활성, 셀로바이오히이드롤라제 활성, 만난아제 활성 및/또는 베타-글루코시다제 활성을 보유하는 폴리펩티드를 암호화하는 핵산을 동정 또는 분리하기 위한 핵산 프로브를 제공하는데, 여기서, 상기 프로브는 본 발명의 서열을 포함하는, 약 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950 또는 1000 개 이상의 연속적인 염기 서열들, 또는 이의 단편이나 부분 서열을 포함하며, 상기 프로브는 결합 또는 혼성화에 의해 핵산을 동정한다. 상기 프로브는 본 발명의 서열을 포함하는, 약 10~50개, 약 20~60개, 약 30~70개, 약 40~80개, 또는 약 60~100개 이상의 연속적인 염기 서열들을 포함하는 올리고뉴클레오티드, 또는 이의 단편이나 부분 서열을 포함할 수 있다.

[0030] 본 발명은 셀룰라제 활성 예를 들어, 엔도글루카나제 활성, 셀로바이오히이드롤라제 활성, 만난아제 활성 및/또는 베타-글루코시다제 활성을 보유하는 폴리펩티드를 암호화하는 핵산을 동정 또는 분리하기 위한 핵산 프로브를 제공하는데, 여기서, 상기 프로브는 본 발명의 핵산 잔기를 약 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950 또는 1000개 이상 포함하는 서열을 포함하는 핵산 예를 들어, 본 발명의 대표적인 핵산과의 동일성이 약 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 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% 이상이거나 완벽한(100%) 서열 동일성을 갖는 폴리뉴클레오티드를 포함한다. 하나의 측면에서, 상기 서열 동일성은 서열 비교 알고리즘을 이용하는 분석법 또는 시각적 조사에 의해 측정된다. 대안적인 측면에서, 프로브는 본 발명의 핵산 서열 중 약 10~50개, 약 20~60개, 약 30~70개, 약 40~80개 또는 약 60~100개의 연속된 염기를 포함하는 올리고뉴클레오티드 또는 이의 부분 서열을 포함할 수 있다.

[0031] 본 발명은 셀룰라제 활성 예를 들어, 엔도글루카나제 활성, 셀로바이오히이드롤라제 활성, 만난아제 활성 및/또는 베타-글루코시다제 활성을 보유하는 폴리펩티드를 암호화하는 핵산을 증폭(예를 들어, PCR에 의함)시키기 위한 증폭용 프라이머 쌍을 제공하는데, 상기 프라이머 쌍은 본 발명의 서열을 포함하는 핵산, 또는 이의 단편이나 부분 서열을 증폭시킬 수 있다. 증폭용 프라이머 서열 쌍의 하나 또는 각각의 일원은 약 10~50개 이상의 연속적인 염기 서열, 또는 약 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35 또는 36개 이상의 연속적인 염기 서열을 포함하는 올리고뉴클레오티드를 포함할 수 있다. 본 발명은 증폭용 프라이머 쌍을 제공하는데, 이 프라이머 쌍은 본 발명의 핵산 잔기를, 앞(5' 말단)에서부터 시작하여 약 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35 또는 36개 이상 포함하는 서열을 가지는 제1 일원, 그리고 상기 제1 일원의 상보성 가닥 잔기를, 앞(5' 말단)에서부터 시작하여 약 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35 또는 36개 이상 포함하는 서열을 가지는 제2 일원을 포함한다.

[0032] 본 발명은 본 발명의 증폭용 프라이머 쌍을 이용하는 증폭 반응 예를 들어, 중합 효소 연쇄 반응(PCR)에 의해 생산되는, 셀룰라제 암호화 핵산 예를 들어, 엔도글루카나제-암호화 핵산, 셀로바이오히이드롤라제-암호화 핵산 및/또는 베타-글루코시다제-암호화 핵산을 제공한다. 본 발명은 본 발명의 증폭용 프라이머 쌍을 이용하는 증폭 반응 예를 들어, 중합 효소 연쇄 반응(PCR)에 의해 생산되는, 셀룰라제 암호화 핵산 예를 들어, 엔도글루카나제-암호화 핵산, 셀로바이오히이드롤라제-암호화 핵산 및/또는 베타-글루코시다제-암호화 핵산을 제공한다. 본 발명은 본 발명의 증폭용 프라이머 쌍을 이용하는 증폭 반응 예를 들어, 중합 효소 연쇄 반응(PCR)에 의해 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 제조하는 방법을 제공한다. 하나의 측면에서, 상기 증폭용 프라이머 쌍은 라이브러리 예를 들어, 유전자 라이브러리 예를 들어, 환경 라이브러리(environmental library)로부터 핵산을 증폭시킨다.

[0033] 본 발명은 셀룰라제 활성 예를 들어, 엔도글루카나제 활성, 셀로바이오히이드롤라제 활성, 만난아제 활성 및/또는 베타-글루코시다제 활성을 보유하는 폴리펩티드를 암호화하는 핵산을 증폭시키는 방법을 제공하며, 이 방법은 본 발명의 핵산 서열, 또는 이의 단편이나 부분 서열을 증폭시킬 수 있는 증폭용 프라이머 서열 쌍을 이용하여 주형 핵산을 증폭시키는 단계를 포함한다.

[0034] 본 발명은 본 발명의 핵산이나 이의 부분 서열을 포함하는 발현 카세트를 제공한다. 하나의 측면에서, 상기 발현 카세트는 프로모터에 작동 가능하게 연결된 핵산을 포함할 수 있다. 상기 프로모터는 바이러스 프로모터, 박

테리아 프로모터, 포유동물 프로모터 또는 식물 프로모터일 수 있다. 하나의 측면에서, 상기 식물 프로모터는 감자, 쌀, 옥수수, 밀, 담배 또는 보리의 프로모터일 수 있다. 상기 프로모터는 구성적 프로모터일 수 있다. 상기 구성적 프로모터는 CaMV35S를 포함할 수 있다. 다른 측면에서, 상기 프로모터는 유도성 프로모터일 수 있다. 하나의 측면에서, 상기 프로모터는 조직-특이적 프로모터이거나 또는 환경에 의해 조절되거나 성장함에 따라 조절되는 프로모터일 수 있다. 그러므로, 상기 프로모터는 예를 들어, 종자-특이적, 잎-특이적, 뿌리-특이적, 줄기-특이적 또는 이타-유도성 프로모터일 수 있다. 하나의 측면에서, 상기 발현 카세트는 또한 식물 또는 식물 바이러스 발현 벡터를 포함할 수도 있다.

[0035] 본 발명은 본 발명의 발현 카세트(예를 들어, 벡터) 또는 본 발명의 핵산을 포함하는 클로닝 비이클을 제공한다. 상기 클로닝 비이클은 바이러스 벡터, 플라스미드, 파지, 파지미드, 코스미드, 포스미드, 박테리오파지 또는 인공 염색체일 수 있다. 상기 바이러스 벡터는 아데노바이러스 벡터, 레트로바이러스 벡터 또는 아데노 관련 바이러스 벡터를 포함할 수 있다. 상기 클로닝 비이클은 박테리아의 인공 염색체(BAC), 플라스미드, 박테리오파지 P1-유래 벡터(PAC), 효모 인공 염색체(YAC) 또는 포유동물 인공 염색체(MAC)를 포함할 수 있다.

[0036] 본 발명은 본 발명의 핵산 또는 본 발명의 발현 카세트(예를 들어, 벡터), 또는 본 발명의 클로닝 비이클을 포함하는 형질전환 세포를 제공한다. 하나의 측면에서, 상기 형질전환 세포는 박테리아 세포, 포유 동물 세포, 진균 세포, 효모 세포, 곤충 세포 또는 식물 세포일 수 있다. 하나의 측면에서, 식물 세포는 대두, 유채씨, 유지종자, 토마토, 사탕수수, 곡류, 감자, 밀, 쌀, 옥수수, 담배 또는 보리의 세포일 수 있다.

[0037] 본 발명은 본 발명의 핵산 또는 본 발명의 발현 카세트(예를 들어, 벡터)를 포함하는 트랜스제닉 비인간 동물을 제공한다. 하나의 측면에서, 상기 동물은 마우스, 래트, 돼지, 염소 또는 양이다.

[0038] 본 발명은 본 발명의 핵산 또는 본 발명의 발현 카세트(예를 들어, 벡터)를 포함하는 트랜스제닉 식물을 제공한다. 상기 트랜스제닉 식물은 곡류, 옥수수, 감자, 토마토, 밀, 유지 종자, 유채씨, 대두, 쌀, 보리 또는 담배 식물일 수 있다.

[0039] 본 발명은 본 발명의 핵산 또는 본 발명의 발현 카세트(예를 들어, 벡터)를 포함하는 트랜스제닉 종자를 제공한다. 상기 트랜스제닉 종자는 곡류, 옥수수, 밀알, 유지 종자, 유채씨, 대두씨, 야자씨, 해바라기씨, 참깨씨, 땅콩 또는 담배 식물의 종자일 수 있다.

[0040] 본 발명은 엄중한 조건 하에서 본 발명의 핵산에 혼성화할 수 있거나 또는 이에 상보적인 핵산 서열을 포함하는 안티센스 올리고뉴클레오티드를 제공한다. 본 발명은 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 메세지가 세포 내에서 번역되는 것을 억제하는 방법을 제공하는데, 이 방법은 엄중한 조건 하에서 본 발명의 핵산에 혼성화할 수 있거나 또는 이에 상보적인 핵산 서열을 포함하는 안티센스 올리고뉴클레오티드를 세포에 도입하거나 또는 이를 세포 내에서 발현시키는 단계를 포함한다. 하나의 측면에서, 상기 안티센스 올리고뉴클레오티드의 길이는 약 10~50개, 약 20~60개, 약 30~70개, 약 40~80개, 또는 약 60~100개의 염기 예를 들어, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 또는 100 개 이상의 염기이다. 본 발명은 셀룰라제 효소 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 메세지가 세포 내에서 번역되는 것을 억제하는 방법을 제공하는데, 이 방법은 엄중한 조건 하에서 본 발명의 핵산에 혼성화할 수 있거나 또는 이에 상보적인 핵산 서열을 포함하는 안티센스 올리고뉴클레오티드를 세포에 투여하거나 또는 이를 세포 내에서 발현시키는 단계를 포함한다.

[0041] 본 발명은 본 발명의 서열의 부분 서열을 포함하는, 이중 가닥 억제성 RNA(RNAi 또는 RNA 간섭) 분자(예를 들어, 전사를 억제하는 소 간섭 RNA 또는 siRNA와, 번역을 억제하는 마이크로 RNA 또는 miRNA)를 제공한다. 하나의 측면에서, 상기 siRNA는 길이가 약 21~24개인 잔기이거나 약 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 또는 100개 이상인 이중체 뉴클레오티드이다. 본 발명은 셀룰라제 효소 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소가 세포 내에서 발현되는 것을 억제하는 방법을 제공하며, 이 방법은 이중 가닥 억제성 RNA(siRNA 또는 miRNA)를 세포에 도입하거나 또는 세포 내에서 발현시키는 단계를 포함하고, 상기 RNA는 본 발명의 서열의 부분 서열을 포함한다.

[0042] 본 발명은 약 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325 또는 350개 이상의 잔기로 이루어진 부위, 또는 폴리펩티드 전체 길이에 걸쳐서, 본 발명의 대표적인 폴리펩티드 또는 펩티드에 대한 서열 동일성이 적어도 약 50%, 51%, 52%, 53%,

54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 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% 이상이거나 완벽한(100%) 서열 동일성을 갖는 아미노산 서열을 포함하는 분리된 또는 재조합 폴리펩티드를 제공한다. 하나의 측면에서, 상기 서열 동일성은 서열 비교 알고리즘을 사용하는 분석법 또는 시각적 조사에 의해 측정된다. 본 발명의 대표적인 폴리펩티드 또는 펩티드 서열은 서열 번호 2, 서열 번호 4, 서열 번호 6, 서열 번호 8, 서열 번호 10, 서열 번호 12, 서열 번호 14, 서열 번호 16, 서열 번호 18, 서열 번호 20, 서열 번호 22, 서열 번호 24, 서열 번호 26, 서열 번호 28, 서열 번호 30, 서열 번호 32, 서열 번호 34, 서열 번호 36, 서열 번호 38, 서열 번호 40, 서열 번호 42, 서열 번호 44, 서열 번호 46, 서열 번호 48, 서열 번호 50, 서열 번호 52, 서열 번호 54, 서열 번호 56, 서열 번호 58, 서열 번호 60, 서열 번호 62, 서열 번호 64, 서열 번호 66, 서열 번호 68, 서열 번호 70, 서열 번호 72, 서열 번호 74, 서열 번호 76, 서열 번호 78, 서열 번호 80, 서열 번호 82, 서열 번호 84, 서열 번호 86, 서열 번호 88, 서열 번호 90, 서열 번호 92, 서열 번호 94, 서열 번호 96, 서열 번호 98, 서열 번호 100, 서열 번호 102, 서열 번호 104, 서열 번호 106, 서열 번호 108, 서열 번호 110, 서열 번호 112, 서열 번호 114, 서열 번호 116, 서열 번호 118, 서열 번호 120, 서열 번호 122, 서열 번호 124, 서열 번호 126, 서열 번호 128, 서열 번호 130, 서열 번호 132, 서열 번호 134, 서열 번호 136, 서열 번호 138, 서열 번호 140, 서열 번호 142, 서열 번호 144, 서열 번호 146, 서열 번호 148, 서열 번호 150, 서열 번호 152, 서열 번호 154, 서열 번호 156, 서열 번호 158, 서열 번호 160, 서열 번호 162, 서열 번호 164 및 서열 번호 166의 서열[이하의 표 1a~1c, 2 및 3, 실시예 1 및 4, 그리고 서열 목록 참조]과, 이의 부분 서열 및 변이체를 포함한다. 대표적인 폴리펩티드는 또한 길이 약 10, 15, 20, 25, 30, 35, 40, 45, 50, 75, 80, 85, 90, 95, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550 또는 600개 이상의 잔기로 이루어진 단편, 또는 전체 길이 효소를 포함한다. 본 발명의 폴리펩티드 또는 펩티드는 본 발명의 핵산에 의해 암호화되는 서열을 포함한다. 본 발명의 폴리펩티드 또는 펩티드 서열은 본 발명의 항체에 의해 특이적으로 결합되는 폴리펩티드 또는 펩티드(예를 들어, 에피토프), 또는 본 발명의 항체를 생산할 수 있는 폴리펩티드 또는 펩티드(예를 들어, 면역원)를 포함한다.

[0043] 하나의 측면에서, 본 발명의 폴리펩티드는 하나 이상의 셀룰라제 효소 활성 예를 들어, 엔도글루카나제 활성, 셀로바이오하이드롤라제 활성, 만난아제 활성 및/또는 베타-글루코시다제 효소 활성을 갖는다. 대안적인 측면에서, 본 발명의 폴리뉴클레오티드는 하나 이상의 셀룰라제 효소 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성을 보유하는 폴리펩티드를 암호화한다.

[0044] 하나의 측면에서, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성은 열 안정성이다. 본 발명의 폴리펩티드는 온도 약 1~약 5℃, 약 5~약 15℃, 약 15~약 25℃, 약 25~약 37℃, 약 37~약 95℃, 약 55~약 85℃, 약 70~약 75℃, 또는 약 90~95℃ 이상인 조건 하에서 셀룰라제 활성 예를 들어, 엔도글루카나제 활성, 셀로바이오하이드롤라제 활성, 만난아제 활성 및/또는 베타-글루코시다제 효소 활성을 보유할 수 있다. 다른 측면에서, 셀룰라제 활성 예를 들어, 엔도글루카나제 활성, 셀로바이오하이드롤라제 활성, 만난아제 활성 및/또는 베타-글루코시다제 효소 활성은 열 내성일 수 있다. 본 발명의 폴리펩티드는 온도 37~약 95℃ 또는 55~약 85℃에 노출된 후에 셀룰라제 활성 예를 들어, 엔도글루카나제 활성, 셀로바이오하이드롤라제 활성, 만난아제 활성 및/또는 베타-글루코시다제 효소 활성을 유지할 수 있다. 하나의 측면에서, 본 발명의 폴리펩티드는 90~약 95℃의 온도에서 pH4.5에 노출된 후에도 셀룰라제 활성 예를 들어, 엔도글루카나제 활성, 셀로바이오하이드롤라제 활성, 만난아제 활성 및/또는 베타-글루코시다제 효소 활성을 유지할 수 있다.

[0045] 본 발명의 다른 측면은 본 발명의 폴리펩티드 또는 펩티드 서열 중 적어도 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 125 또는 150개 이상의 연속된 염기를 포함하는 분리된 또는 재조합 폴리펩티드 또는 펩티드, 이와 실질적으로 상동성인 서열, 그리고 이와 상보성인 서열을 제공한다. 이 펩티드는 예를 들어, 면역원성 단편, 모티프(예를 들어, 결합 위치), 신호 서열, 프리프로 서열(prepro sequence) 또는 활성 위치일 수 있다.

[0046] 본 발명은 셀룰라제 활성 예를 들어, 엔도글루카나제 활성, 셀로바이오하이드롤라제 활성, 만난아제 활성 및/또는 베타-글루코시다제 효소 활성을 보유하는 폴리펩티드를 암호화하는 서열 및 신호 서열을 포함하는 분리된 또는 재조합 핵산을 제공하는데, 여기서, 상기 핵산은 본 발명의 서열을 포함한다. 신호 서열은 다른 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 또는 비-셀룰라제 예를 들어, 비-엔도글루카나제, 비-셀로바이오하이드롤라제 및/또는 비-베타-글루코시다제 효소(이종성 효소)로부터 유래될 수 있다. 본 발명은 셀룰라제 활성 예를 들어, 엔도글루카나제 활성, 셀로바이오하이드롤라

제 활성, 만난아제 활성 및/또는 베타-글루코시다제 효소 활성을 보유하는 폴리펩티드를 암호화하는 서열을 포함하는 분리된 또는 재조합 핵산을 제공하며, 여기서, 상기 서열은 신호 서열을 함유하지 않고, 상기 핵산은 본 발명의 서열을 포함한다. 하나의 측면에서, 본 발명은 신호 서열의 전부 또는 일부가 결여된 본 발명의 폴리펩티드를 포함하는 분리된 또는 재조합 폴리펩티드를 제공한다. 하나의 측면에서, 분리된 또는 재조합 폴리펩티드는 이중 신호 서열 예를 들어, 이중 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 신호 서열 또는 비-셀룰라제 예를 들어, 비-엔도글루카나제, 비-셀로바이오하이드롤라제 및/또는 비-베타-글루코시다제 효소 신호 서열을 포함하는, 본 발명의 폴리펩티드를 포함할 수 있다.

[0047] 하나의 측면에서, 본 발명은 본 발명의 신호 서열을 포함하는 제1 도메인과 적어도 제2 도메인을 포함하는 키메라 단백질을 제공한다. 상기 단백질은 융합 단백질일 수 있다. 제2 도메인은 효소를 포함할 수 있다. 상기 효소는 비-효소(non-enzyme)일 수 있다.

[0048] 본 발명은 본 발명의 신호 펩티드(SP), 프리프로 서열 및/또는 촉매 도메인(CD)을 포함하는 적어도 제1 도메인과, 이중 폴리펩티드 또는 펩티드를 포함하는 적어도 제2 도메인을 포함하는 키메라 폴리펩티드를 제공하는데, 여기서, 상기 이중 폴리펩티드 또는 펩티드는 자연적으로는 신호 서열(SP), 프리프로 서열 및/또는 촉매 도메인(CD)과 결합되어 있지 않다. 하나의 측면에서, 상기 이중 폴리펩티드 또는 펩티드는 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소가 아니다. 상기 이중 폴리펩티드 또는 펩티드는 신호 펩티드(SP), 프리프로 서열 및/또는 촉매 도메인(CD)의 아미노 말단, 카복시 말단 또는 이들 양 말단부일 수 있다.

[0049] 본 발명은 키메라 폴리펩티드를 암호화하는 분리된 또는 재조합 핵산을 제공하는데, 여기서, 상기 키메라 폴리펩티드는 본 발명의 신호 펩티드(SP), 프리프로 도메인 및/또는 촉매 도메인(CD)을 포함하는 적어도 제1 도메인과, 이중 폴리펩티드 또는 펩티드를 포함하는 적어도 제2 도메인을 포함하며, 상기 이중 폴리펩티드 또는 펩티드는 자연적으로는 신호 펩티드(SP), 프리프로 도메인 및/또는 촉매 도메인(CD)과 결합되어 있지 않다.

[0050] 본 발명은 본 발명의 폴리펩티드, 예를 들어 서열 번호 2, 서열 번호 4, 서열 번호 6, 서열 번호 8, 서열 번호 10, 서열 번호 12, 서열 번호 14, 서열 번호 16, 서열 번호 18, 서열 번호 20, 서열 번호 22, 서열 번호 24, 서열 번호 26, 서열 번호 28, 서열 번호 30, 서열 번호 32, 서열 번호 34, 서열 번호 36, 서열 번호 38, 서열 번호 40, 서열 번호 42, 서열 번호 44, 서열 번호 46, 서열 번호 48, 서열 번호 50, 서열 번호 52, 서열 번호 54, 서열 번호 56, 서열 번호 58, 서열 번호 60, 서열 번호 62, 서열 번호 64, 서열 번호 66, 서열 번호 68, 서열 번호 70, 서열 번호 72, 서열 번호 74, 서열 번호 76, 서열 번호 78, 서열 번호 80, 서열 번호 82, 서열 번호 84, 서열 번호 86, 서열 번호 88, 서열 번호 90, 서열 번호 92, 서열 번호 94, 서열 번호 96, 서열 번호 98, 서열 번호 100, 서열 번호 102, 서열 번호 104, 서열 번호 106, 서열 번호 108, 서열 번호 110, 서열 번호 112, 서열 번호 114, 서열 번호 116, 서열 번호 118, 서열 번호 120, 서열 번호 122, 서열 번호 124, 서열 번호 126, 서열 번호 128, 서열 번호 130, 서열 번호 132, 서열 번호 134, 서열 번호 136, 서열 번호 138, 서열 번호 140, 서열 번호 142, 서열 번호 144, 서열 번호 146, 서열 번호 148, 서열 번호 150, 서열 번호 152, 서열 번호 154, 서열 번호 156, 서열 번호 158, 서열 번호 160, 서열 번호 162, 서열 번호 164 또는 서열 번호 166의 1~14, 1~15, 1~16, 1~17, 1~18, 1~19, 1~20, 1~21, 1~22, 1~23, 1~24, 1~25, 1~26, 1~27, 1~28, 1~29, 1~30, 1~31, 1~32, 1~33, 1~34, 1~35, 1~36, 1~37, 1~38, 1~40, 1~41, 1~42, 1~43, 1~44, 1~45, 1~46 또는 1~47번 잔기에 제시된 바와 같은 서열로 이루어져 있거나 또는 이를 포함하는 분리된 또는 재조합 신호 서열(예를 들어, 신호 펩티드)를 제공한다[이하의 표 1a~1c, 2 및 3과, 실시예 1 및 4, 그리고 서열 목록 참조]. 하나의 측면에서, 본 발명은 본 발명의 폴리펩티드의 처음 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70개 또는 그 이상의 아미노 말단 잔기를 포함하는 신호 서열을 제공한다.

[0051] 하나의 측면에서, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성은 약 37℃에서 단백질 1 mg 당 약 1~약 1200 유닛, 또는 단백질 1 mg 당 약 100~약 1000 유닛의 비활성(specific activity)을 포함한다. 다른 측면에서, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성은 단백질 1 mg 당 약 100~약 1000 유닛, 또는 단백질 1 mg 당 약 500~약 750 유닛의 비활성을 포함한다. 대안적으로, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성은 37℃에서 단백질 1 mg 당 약 1~약 750 유닛, 또는 단백질 1 mg 당 약 500~약 1200 유닛의 비활성을 포함한다. 하나의 측면에서, 셀룰라제 예를

들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성은 37℃에서 단백질 1 mg 당 약 1~약 500 유닛, 또는 단백질 1 mg 당 약 750~약 1000 유닛의 비활성을 포함한다. 다른 측면에서, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성은 37℃에서 단백질 1 mg 당 약 1~약 250 유닛의 비활성을 포함한다. 대안적으로, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성은 37℃에서 단백질 1 mg 당 약 1~약 100 유닛의 비활성을 포함한다.

[0052] 다른 측면에서, 열 내성은 고온으로 가열한 후에도 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 37℃에서의 비활성의 절반 이상을 유지하는 경우를 포함한다. 대안적으로, 열 내성은 고온으로 가열한 후에도 37℃에서의 비활성이 단백질 1 mg 당 약 1~약 1200 유닛이거나 단백질 1 mg 당 약 500~약 1000 유닛인 경우를 포함할 수 있다. 다른 측면에서, 열 내성은 고온으로 가열한 후에도 37℃에서의 비활성이 단백질 1 mg당 약 1~약 500 유닛인 경우를 포함할 수 있다.

[0053] 본 발명은 본 발명의 분리된 또는 재조합 폴리펩티드를 제공하는데, 여기서, 상기 폴리펩티드는 하나 이상의 글리코실화 부위를 포함한다. 하나의 측면에서, 글리코실화는 N-결합 글리코실화일 수 있다. 하나의 측면에서, 상기 폴리펩티드는 피. 파스토리스(*P. pastoris*) 또는 에스. 폼베(*S. pombe*)에서 발현된 이후에도 글리코실화될 수 있다.

[0054] 하나의 측면에서, 본 발명의 폴리펩티드는 대략 pH 6.5, pH 6, pH 5.5, pH 5, pH 4.5 또는 pH 4 또는 그 이상의 산성도를 포함하는 조건 하에서 셀룰라제, 예를 들어 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 활성을 유지할 수 있다. 다른 측면에서, 본 발명의 폴리펩티드는 대략 pH 7, pH 7.5, pH 8.0, pH 8.5, pH 9, pH 9.5, pH 10, pH 10.5 또는 pH 11 또는 그 이상의 염기성 pH를 포함하는 조건 하에서 셀룰라제, 예를 들어 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 활성을 유지할 수 있다. 하나의 측면에서, 상기 폴리펩티드는 대략 pH 6.5, pH 6, pH 5.5, pH 5, pH 4.5 또는 pH 4 또는 그 이상의 산성도를 포함하는 조건에 노출된 후에도 셀룰라제, 예를 들어 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 활성을 유지할 수 있다. 다른 측면에서, 본 발명의 폴리펩티드는 대략 pH 7, pH 7.5, pH 8.0, pH 8.5, pH 9, pH 9.5, pH 10, pH 10.5 또는 pH 11 또는 그 이상의 염기성 pH를 포함하는 조건에 노출된 후에도 셀룰라제, 예를 들어 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 활성을 유지할 수 있다.

[0055] 하나의 측면에서, 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소는 알칼리성 조건 예를 들어, 장 예를 들어, 소장의 알칼리성 조건 하에서 활성을 가진다. 하나의 측면에서, 본 발명의 폴리펩티드는 위와 같은 산성 pH에 노출된 후에도 활성을 유지할 수 있다.

[0056] 본 발명은 본 발명의 폴리펩티드(펩티드 포함)를 포함하는 단백질 제제를 제공하는데, 여기서, 상기 단백질 제제는 액체, 고체 또는 겔을 포함한다. 본 발명은 본 발명의 폴리펩티드와 제2 단백질 또는 도메인을 포함하는 이중이량체를 제공한다. 상기 이중이량체의 제2 일원은 상이한 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소, 상이한 효소 또는 기타 단백질일 수 있다. 하나의 측면에서, 제2 도메인은 폴리펩티드일 수 있고, 이중이량체는 융합 단백질일 수 있다. 하나의 측면에서, 상기 제2 도메인은 에피토프 또는 태그일 수 있다. 하나의 측면에서, 본 발명은 본 발명의 폴리펩티드를 포함하는 동중이량체를 제공한다.

[0057] 본 발명은 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 활성을 보유하는 고정화된(immobilized) 폴리펩티드(펩티드 포함)를 제공하는데, 여기서, 상기 고정화된 폴리펩티드는 본 발명의 핵산에 의해 암호화되는 폴리펩티드, 또는 본 발명의 폴리펩티드와 제2 도메인을 포함하는 폴리펩티드를 포함한다. 하나의 측면에서, 본 발명의 폴리펩티드는 세포, 금속, 수지, 중합체, 세라믹, 유리, 미세 전극, 흑연 입자, 비드, 겔, 평판, 어레이 또는 모세관 상에 고정화될 수 있다.

[0058] 본 발명은 또한 본 발명의 고정화된 핵산 예를 들어, 본 발명의 프로브를 포함하는 어레이를 제공한다. 본 발명은 또한 본 발명의 항체를 포함하는 어레이를 제공한다.

[0059] 본 발명은 본 발명의 폴리펩티드에 특이적으로 결합하거나, 또는 본 발명의 핵산에 의해 암호화되는 폴리펩티드에 특이적으로 결합하는 분리된 또는 재조합 항체를 제공한다. 이와 같이 본 발명의 항체는 모노클로날 항체 또는 폴리클로날 항체일 수 있다. 본 발명은 본 발명의 항체 예를 들어, 본 발명의 폴리펩티드 또는 본 발명의 핵산에 의해 암호화되는 폴리펩티드에 특이적으로 결합하는 항체를 포함하는 하이브리도마를 제공한다. 본 발명은

이러한 항체를 암호화하는 핵산을 제공한다.

- [0060] 본 발명은 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성을 보유하는 폴리펩티드를 분리 또는 동정하는 방법을 제공하는데, 여기서, 상기 방법은 (a) 본 발명의 항체를 제공하는 단계; (b) 폴리펩티드를 포함하는 시료를 제공하는 단계; 및 (c) 항체가 폴리펩티드에 특이적으로 결합할 수 있는 조건 하에서 상기 단계 (b)의 시료와 상기 단계 (a)의 항체를 접촉시켜, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성을 보유하는 폴리펩티드를 분리 또는 동정하는 단계를 포함한다.
- [0061] 본 발명은 항-셀룰라제 효소 항체 예를 들어, 항-엔도글루카나제 효소 항체, 항-셀로바이오하이드롤라제 효소 항체 및/또는 항-베타-글루코시다제 효소 항체를 생산하는 방법을 제공하는데, 여기서, 이 방법은 비인간 동물에 본 발명의 핵산 또는 본 발명의 폴리펩티드 또는 이의 부분 서열을 체액 면역 반응을 일으키기에 충분한 양만큼 투여하여, 항-셀룰라제 효소 항체 예를 들어, 항-엔도글루카나제 효소 항체, 항-셀로바이오하이드롤라제 효소 항체 및/또는 항-베타-글루코시다제 효소 항체를 생산하는 단계를 포함한다. 본 발명은 항-셀룰라제 면역 반응 예를 들어, 항-엔도글루카나제 면역 반응, 항-셀로바이오하이드롤라제 면역 반응 및/또는 항-베타-글루코시다제 면역 반응(세포성 또는 체액 면역 반응)을 일으키는 방법을 제공하는데, 여기서, 이 방법은 비인간 동물에 본 발명의 핵산 또는 본 발명의 폴리펩티드 또는 이의 부분 서열을 면역 반응(세포성 또는 체액 면역 반응)을 일으키는데에 충분한 양만큼 투여하는 단계를 포함한다.
- [0062] 본 발명은 다음과 같은 단계들을 포함하는 재조합 폴리펩티드를 생산하는 방법을 제공한다: (a) 프로모터에 작동 가능하게 연결된 본 발명의 핵산을 제공하는 단계; 및 (b) 폴리펩티드를 발현시킬 수 있는 조건 하에서 상기 단계 (a)의 핵산을 발현시켜, 재조합 폴리펩티드를 생산하는 단계. 하나의 측면에서, 본 발명의 방법은 단계 (a)의 핵산으로 숙주 세포를 형질전환한 후, 상기 단계 (a)의 핵산을 발현시켜, 형질전환된 세포 내에서 재조합 폴리펩티드를 생산하는 단계를 추가로 포함할 수 있다.
- [0063] 본 발명은 다음과 같은 단계들을 포함하는, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성을 보유하는 폴리펩티드를 동정하는 방법을 제공한다: (a) 본 발명의 폴리펩티드 또는 본 발명의 핵산에 의해 암호화되는 폴리펩티드를 제공하는 단계; (b) 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 기질을 제공하는 단계; 및 (c) 상기 단계 (a)의 폴리펩티드 또는 이의 단편이나 변이체와 상기 단계 (b)의 기질을 접촉시켜, 기질 양의 감소 또는 반응 생성물의 양의 증가를 검출하는 단계[여기서, 기질의 양의 감소 또는 반응 생성물의 양의 증가를 통하여, 상기 폴리펩티드가 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성을 보유하는지 여부를 확인할 수 있음]. 하나의 측면에서, 상기 기질은 셀룰로스-포함 화합물이다.
- [0064] 본 발명은 다음의 단계들을 포함하는 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 기질의 동정 방법을 제공한다: (a) 본 발명의 폴리펩티드 또는 본 발명의 핵산에 의해 암호화되는 폴리펩티드를 제공하는 단계; (b) 시험 기질을 제공하는 단계; 및 (c) 상기 단계 (a)의 폴리펩티드와 상기 단계 (b)의 시험 기질을 접촉시켜, 기질의 양의 감소 또는 반응 생성물의 양의 증가를 검출하는 단계[여기서, 기질의 양의 감소 또는 반응 생성물의 양의 증가를 통하여, 상기 시험 기질이 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 기질과 같은 시험 기질임을 확인할 수 있음].
- [0065] 본 발명은 다음의 단계들을 포함하는, 시험 화합물이 폴리펩티드에 특이적으로 결합하는지 여부를 결정하는 방법을 제공한다: (a) 핵산을 폴리펩티드로 번역시킬 수 있는 조건 하에서 핵산을 포함하는 벡터 또는 핵산을 발현시키거나[여기서, 상기 핵산은 본 발명의 핵산을 포함함], 또는 본 발명의 폴리펩티드를 제공하는 단계; (b) 시험 화합물을 제공하는 단계; (c) 이 시험 화합물과 폴리펩티드를 접촉시키는 단계; 및 (d) 상기 단계 (b)의 시험 화합물이 폴리펩티드에 특이적으로 결합하는지 여부를 결정하는 단계.
- [0066] 본 발명은 다음의 단계들을 포함하는, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성의 조절제를 동정하는 방법을 제공한다: (a) 본 발명의 폴리펩티드 또는 본 발명의 핵산에 의해 암호화되는 폴리펩티드를 제공하는 단계; (b) 시험 화합물을 제공하는 단계; (c) 상기 단계 (a)의 폴리펩티드와 상기 단계 (b)의 시험 화합물을 접촉시켜, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 활성을 측정하는 단계로서, 여기서 시험 화합물의 부재 하에서 측정된 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-

타-글루코시다제 효소의 활성에 비교한, 시험 화합물의 존재 하에서 측정된 상기 효소들의 활성에 있어서의 변화는 상기 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 활성을 조정하는 시험 화합물이 존재하는 증거가 되는 것인 단계. 하나의 측면에서, 상기 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 활성은, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 기질을 제공하여, 이 기질의 양이 감소하는지 아니면 반응 생성물의 양이 증가하는지, 또는 시질의 양이 증가하는지 아니면 반응 생성물의 양이 감소하는지를 확인함으로써 측정될 수 있다. 시험 화합물이 존재하지 않을 경우의 기질 또는 반응 생성물의 양에 비교한, 시험 화합물이 존재할 경우의 기질 양의 감소 또는 반응 생성물 양의 증가 여부는, 이 시험 화합물이 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성에 대한 활성화 인자인지 여부를 확인해준다. 시험 화합물이 존재하지 않을 경우의 기질 또는 반응 생성물의 양에 비교한, 시험 화합물이 존재할 경우의 기질 양의 감소 또는 반응 생성물 양의 증가 여부는, 이 시험 화합물이 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성의 억제제인지 여부를 확인해준다.

[0067] 본 발명은 프로세서 및 데이터 저장 장치를 포함하는 컴퓨터 시스템을 제공하는데, 여기서, 상기 데이터 저장 장치는 본 발명의 폴리펩티드 서열 또는 핵산 서열(예를 들어, 본 발명의 핵산에 의해 암호화되는 폴리펩티드 또는 펩티드)을 저장하고 있다. 하나의 측면에서, 본 발명의 컴퓨터 시스템은 또한 하나 이상의 참조 서열이 저장되어 있는 데이터 저장 장치와 서열 비교 알고리즘을 포함할 수 있다. 다른 측면에서, 본 발명의 서열 비교 알고리즘은 다형성을 나타내는 컴퓨터 프로그램을 포함한다. 하나의 측면에서, 본 발명의 컴퓨터 시스템은 또한 상기 서열에 존재하는 하나 이상의 특징들을 동정하는 식별자(identifier)를 포함할 수도 있다. 본 발명은 본 발명의 핵산 서열 또는 폴리펩티드 서열이 저장되어 있는 컴퓨터 판독 가능 매체를 제공한다. 본 발명은 다음의 단계들을 포함하는, 서열의 특징을 동정하는 방법을 제공한다: (a) 서열 내 하나 이상의 특징들을 동정하는 컴퓨터 프로그램을 이용하여 서열을 판독하는 단계[여기서, 상기 서열은 본 발명의 폴리펩티드 서열 또는 핵산 서열을 포함함]; 및 (b) 컴퓨터 프로그램을 이용하여 서열 내의 하나 이상의 특징들을 동정하는 단계. 본 발명은 다음의 단계들을 포함하는, 제2 서열과 제1 서열의 비교 방법을 제공한다: (a) 서열을 비교하는 컴퓨터 프로그램을 이용하여 제1 서열과 제2 서열을 판독하는 단계[여기서, 상기 제1 서열은 본 발명의 폴리펩티드 서열 또는 핵산 서열을 포함함]; 및 (b) 컴퓨터 프로그램을 이용하여 상기 제1 서열과 제2 서열 간의 차이점을 측정하는 단계. 상기 제1 서열과 제2 서열 간의 차이점을 측정하는 단계는 또한 다형성을 동정하는 단계를 포함할 수도 있다. 하나의 측면에서, 이 방법은 또한 서열 내의 하나 이상의 특징들을 동정하는 식별자를 포함할 수도 있다. 다른 측면에서, 이 방법은 컴퓨터 프로그램을 이용하여 제1 서열을 판독하는 단계 및 이 서열 내의 하나 이상의 특징들을 확인하는 단계를 포함할 수 있다.

[0068] 본 발명은 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 활성을 보유하는 폴리펩티드를 암호화하는 핵산을 환경 시료로부터 분리 또는 회수하는 방법을 제공하는데, 여기서, 이 방법은 (a) 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 활성을 보유하는 폴리펩티드를 암호화하는 핵산을 증폭시키기 위한 증폭용 프라이머 서열 쌍을 제공하는 단계[여기서, 상기 프라이머 쌍은 본 발명의 핵산을 증폭시킬 수 있음]; (b) 환경 시료로부터 핵산을 분리하거나, 또는 이 환경 시료 내에 존재하는 핵산이 증폭용 프라이머 쌍에 혼성화할 수 있도록 이 시료를 처리하는 단계; 및 (c) 상기 단계 (b)의 핵산과 상기 단계 (a)의 증폭용 프라이머 쌍을 결합시키고, 이 환경 시료로부터 핵산을 증폭시켜, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 활성을 보유하는 폴리펩티드를 암호화하는 핵산을 환경 시료로부터 분리 또는 회수하는 단계를 포함한다. 증폭용 프라이머 서열 쌍의 하나 또는 각각의 일원은 본 발명의 증폭용 프라이머 서열 쌍을 포함하는 올리고뉴클레오타이드 예를 들어, 본 발명의 연속된 염기 서열을 약 10~50개 이상 가지는 올리고뉴클레오타이드를 포함할 수 있다.

[0069] 본 발명은 다음의 단계들을 포함하는, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 활성을 보유하는 폴리펩티드를 암호화하는 핵산을 환경 시료로부터 분리 또는 회수하는 방법을 제공한다: (a) 본 발명의 핵산 또는 이의 부분 서열을 포함하는 폴리뉴클레오타이드 프로브를 제공하는 단계; (b) 환경 시료로부터 핵산을 분리하거나, 또는 이 환경 시료 내에 존재하는 핵산이 상기 단계 (a)의 폴리뉴클레오타이드 프로브에 혼성화될 수 있도록 이 환경 시료를 처리하는 단계; (c) 상기 단계 (b)의 분리된 핵산 또는 처리된 환경 시료를 상기 단계 (a)의 폴리뉴클레오타이드 프로브와 결합시키는 단계; 및 (d) 상기 단계 (a)의 폴리뉴클레오타이드 프로브와 특이적으로 혼성화하는 핵산을 분리하여, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 활성을 보유하는 폴리펩티드를

암호화하는 핵산을 환경 시료로부터 분리 또는 회수하는 단계. 상기 환경 시료는 물 시료, 액체 시료, 흙 시료, 공기 시료 또는 생물학적 시료를 포함할 수 있다. 하나의 측면에서, 상기 생물학적 시료는 박테리아 세포, 원생 동물 세포, 곤충 세포, 효모 세포, 식물 세포, 진균 세포 또는 포유동물 세포로부터 유래할 수 있다.

[0070] 본 발명은 다음의 단계들을 포함하는, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 활성을 보유하는 폴리펩티드를 암호화하는 핵산의 변이체를 제조하는 방법을 제공한다: (a) 본 발명의 핵산을 포함하는 주형 핵산을 제공하는 단계; 및 (b) 주형 서열에 존재하는 하나 이상의 뉴클레오티드를 변형, 결실 또는 추가하거나, 또는 이러한 작업들을 조합하여 수행함으로써, 주형 핵산의 변이체를 생산하는 단계. 하나의 측면에서, 이 방법은 또한 변이체 핵산을 발현시켜 변이체 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 폴리펩티드를 생산하는 단계를 추가로 포함할 수 있다. 변형, 부가 또는 결실은 오류 빈발 PCR, 셔플링(shuffling), 올리고뉴클레오티드 지정 돌연변이 유발법, 어셈블리 PCR, 유성(sexual) PCR 돌연변이 유발법, 생체내 돌연변이 유발법, 카세트 돌연변이 유발법, 반복적 앙상블 돌연변이 유발법(recursive ensemble mutagenesis), 지수적 앙상블 돌연변이 유발법(exponential ensemble mutagenesis), 위치 특이적 돌연변이 유발법, 유전자 재조합법, 유전자 위치 포화 돌연변이 유발법(Gene Site Saturation Mutagenesis; GSSM), 합성 결찰 재조합법(SLR), 염색체 포화 돌연변이 유발법(CSM) 또는 이 방법들의 조합법을 포함하는 방법에 의해서 도입될 수 있다. 다른 측면에서, 상기 변형, 부가 또는 결실은 재조합법, 반복적 서열 재조합법, 포스포티오에이트 변형 DNA 돌연변이 유발법, 우라실 함유 주형 돌연변이 유발법, 갭 형성 이중체 돌연변이 유발법, 포인트 미스매치 수복 돌연변이 유발법, 수복 결함 숙주 균주 돌연변이 유발법, 화학적 돌연변이 유발법, 방사능에 의한 돌연변이 유발법, 결실 돌연변이 유발법, 제한 선별 돌연변이 유발법, 제한 정제 돌연변이 유발법, 인공 유전자 합성법, 앙상블 돌연변이 유발법, 키메라 핵산 다량체 생성법 및 이들 방법의 조합법을 포함하는 방법에 의해 도입된다.

[0071] 하나의 측면에서, 본 발명의 방법은 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소가 변형되었거나 또는 상이한 활성을 가지게 될 때까지, 또는 주형인 핵산에 의해 암호화되는 폴리펩티드의 안정성과 상이한 안정성을 가지거나 또는 변질된 안정성을 가지게 될 때까지 반복 수행할 수 있다. 하나의 측면에서, 본 발명의 변이체 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 폴리펩티드는 열 내성으로서, 고온에 노출된 후에도 다소의 활성을 유지한다. 다른 측면에서, 본 발명의 변이체 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 폴리펩티드는, 주형 핵산에 의해 암호화되는 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소에 비해, 글리코실화 효율이 증가한다. 대안적으로, 본 발명의 변이체 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 폴리펩티드는 고온에서도 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 활성을 보유하는데, 여기서, 주형 핵산에 의해 암호화되는 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소는 고온에서는 활성을 가지지 않는다. 하나의 측면에서, 본 발명의 방법은 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 암호화하는 서열의 코돈 사용 패턴이 주형 핵산의 경우와 상이하게 변할 때까지 반복적으로 수행될 수 있다. 다른 측면에서, 본 발명의 방법은 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 유전자의 메시지 발현 수준 또는 안정성이 주형 핵산에 의한 경우보다 높거나 또는 낮게 될 때까지 반복적으로 수행될 수 있다.

[0072] 본 발명은 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성을 보유하는 폴리펩티드를 암호화하는 핵산 내에 존재하는 코돈을 변형시켜 이것의 숙주 세포에서의 발현량을 증가시키는 방법을 제공하는데, 이 방법은 다음의 단계들을 포함한다: (a) 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성을 보유하는 폴리펩티드를 암호화하는, 본 발명의 핵산을 제공하는 단계; 및 (b) 상기 단계 (a)의 핵산 내에 존재하는 비선호 코돈 또는 저선호 코돈을 확인하고, 이를 치환된 코돈이 암호화하는 아미노산과 동일한 아미노산을 암호화하는, 중립 사용 코돈 또는 선호 코돈과 치환시킴으로써, 이 핵산을 변형시켜 숙주 세포에서의 발현량을 증가시키는 단계[여기서, 상기 선호 코돈은 숙주 세포 내에 존재하는 유전자 내의 암호화 서열에서 과다 출현하는(over-represented) 코돈이고, 상기 비선호 코돈 또는 저선호 코돈은 숙주 세포 내에 존재하는 유전자 내의 암호화 서열에서 적게 출현하는(under-represented) 코돈임].

[0073] 본 발명은 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제

효소 활성을 보유하는 폴리펩티드를 암호화하는 핵산 내에 존재하는 코돈을 변형시키는 방법을 제공하는데, 이 방법은 다음의 단계들을 포함한다: (a) 본 발명의 핵산을 제공하는 단계; 및 (b) 상기 단계 (a)의 핵산 내에 존재하는 코돈을 확인하여, 이를 치환된 코돈이 암호화하는 아미노산과 동일한 아미노산을 암호화하는 상이한 코돈으로 치환함으로써, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 암호화하는 핵산 내에 존재하는 코돈을 변형시키는 단계.

[0074] 본 발명은 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성을 보유하는 폴리펩티드를 암호화하는 핵산 내에 존재하는 코돈을 변형시켜, 숙주 세포 내에서 이 변형된 코돈의 발현량을 증가시키는 방법을 제공하는데, 이 방법은 다음의 단계들을 포함한다: (a) 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 폴리펩티드를 암호화하는 본 발명의 핵산을 제공하는 단계; 및 (b) 상기 단계 (a)의 핵산 내에 존재하는 비선호 코돈 또는 저선호 코돈을 확인하여, 이를 치환된 코돈이 암호화하는 아미노산과 동일한 아미노산을 암호화하는 중립 사용 코돈 또는 선호 코돈과 치환함으로써, 이 핵산을 변형시켜 숙주 세포에서의 발현량을 증가시키는 단계[여기서, 상기 선호 코돈은 숙주 세포 내에 존재하는 유전자 내의 암호화 서열에서 과다 출현하는 코돈이고, 상기 비선호 코돈 또는 저선호 코돈은 숙주 세포 내에 존재하는 유전자 내의 암호화 서열에서 적게 출현하는 코돈임].

[0075] 본 발명은 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성을 보유하는 폴리펩티드를 암호화하는 핵산 내에 존재하는 코돈을 변형시켜, 숙주 세포 내에서 이 코돈의 발현량을 감소시키는 방법을 제공하는데, 이 방법은 다음의 단계들을 포함한다: (a) 본 발명의 핵산을 제공하는 단계; 및 (b) 상기 단계 (a)의 핵산 내에 존재하는 하나 이상의 선호 코돈을 확인하여, 이를 치환된 코돈이 암호화하는 아미노산과 동일한 아미노산을 암호화하는 저선호 코돈 또는 비선호 코돈과 치환함으로써, 이 핵산을 변형시켜, 숙주 세포에서의 발현량을 감소시키는 단계[여기서, 상기 선호 코돈은 숙주 세포 내에 존재하는 유전자 내의 암호화 서열에서 과다 출현하는 코돈이고, 비선호 코돈 또는 저선호 코돈은 숙주 세포 내에 존재하는 유전자 내의 암호화 서열에서 적게 출현하는 코돈임]. 하나의 측면에서, 숙주 세포는 박테리아 세포, 진균 세포, 곤충 세포, 효모 세포, 식물 세포 또는 포유동물 세포일 수 있다.

[0076] 본 발명은 다수의 변형된 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성 부위 또는 기질 결합 부위를 암호화하는 핵산의 라이브러리를 제조하는 방법을 제공하는데, 여기서, 이 변형된 활성 부위 또는 기질 결합 부위는 제1 활성 부위 또는 제1 기질 결합 부위를 암호화하는 서열을 포함하는 제1 핵산으로부터 유래하고, 이 방법은 다음의 단계들을 포함한다: (a) 제1 활성 부위 또는 제1 기질 결합 부위를 암호화하는 제1 핵산을 제공하는 단계[여기서, 상기 제1 핵산 서열은 엄중한 조건 하에서 본 발명의 핵산에 혼성화되는 서열을 포함하고, 상기 핵산은 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성 부위, 또는 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 기질 결합 부위를 암호화함]; (b) 자연 발생 아미노산 변이체를 암호화하는 돌연변이 유발성 올리고뉴클레오타이드 세트, 제1 핵산 내 다수의 표적 코돈에 제공하는 단계; 및 (c) 돌연변이 유발성 올리고뉴클레오타이드 세트를 사용하여, 돌연변이가 유발된 각각의 아미노산 코돈에 일정 범위의 아미노산 변이체를 암호화하는, 기질 결합 부위 암호화 변이체 핵산 또는 활성 부위 암호화 변이체 핵산 세트를 생성시킴으로써, 다수의 변형된 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성 부위 또는 기질 결합 부위를 암호화하는 핵산의 라이브러리를 제조하는 단계. 하나의 측면에서, 상기 방법은 상기 단계 (a)의 제1 핵산을, 최적화 유도 진화 시스템, 유전자 위치 포화 돌연변이 유발법(GSSM), 합성 결찰 제조법(SLR), 오류 빈발 PCR, 서플링, 올리고뉴클레오타이드 지정 돌연변이 유발법, 어셈블리 PCR, 유성 PCR 돌연변이 유발법, 생체내 돌연변이 유발법, 카세트 돌연변이 유발법, 반복적 양상 돌연변이 유발법, 지수적 양상 돌연변이 유발법, 위치 특이적 돌연변이 유발법, 유전자 재조합법 및 상기 방법들의 조합법을 포함하는 방법에 의해 돌연변이시키는 단계를 포함한다. 다른 측면에서, 상기 방법은 상기 단계 (a)의 제1 핵산 또는 변이체를, 재조합법, 반복적 서열 재조합법, 포스포티오에이트 변형 DNA 돌연변이 유발법, 우라실 함유 주형 돌연변이 유발법, 겔 형성 이중체 돌연변이 유발법, 포인트 미스매치 수복 돌연변이 유발법, 수복 결함 숙주 균주 돌연변이 유발법, 화학적 돌연변이 유발법, 방사능에 의한 돌연변이 유발법, 결실 돌연변이 유발법, 제한 선별 돌연변이 유발법, 제한 정제 돌연변이 유발법, 인공 유전자 합성법, 양상 돌연변이 유발법, 키메라 핵산 다량체 생성법 및 이들 방법의 조합법을 포함하는 방법에 의해 돌연변이 유발시키는 단계를 포함한다.

[0077] 본 발명은 다음의 단계들을 포함하는 소형 분자 제조 방법을 제공한다: (a) 소형 분자를 합성 또는 변형시킬 수 있는 다수의 생합성 효소를 제공하는 단계[여기서, 상기 효소 중 하나는 본 발명의 핵산에 의해 암호화되는 셀

룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드룰라제, 만난아제 및/또는 베타-글루코시다제 효소를 포함함]; (b) 상기 단계 (a)의 효소 중 하나 이상에 대한 기질을 제공하는 단계; 및 (c) 다수의 생체 촉매 반응을 촉진하는 조건 하에서 상기 단계 (b)의 기질과 효소를 반응시켜, 일련의 생체 촉매 반응에 의해 소형 분자를 생성하는 단계. 본 발명은 다음의 단계들을 포함하는 소형 분자의 변형 방법을 제공한다: (a) 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드룰라제, 만난아제 및/또는 베타-글루코시다제 효소를 제공하는 단계[여기서, 상기 효소는 본 발명의 폴리펩티드, 본 발명의 핵산에 의해 암호화되는 폴리펩티드 또는 이의 부분 서열을 포함함]; (b) 소형 분자를 제공하는 단계; 및 (c) 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드룰라제, 만난아제 및/또는 베타-글루코시다제 효소에 의해 촉매화되는 효소 반응을 촉진하는 조건 하에서 상기 단계 (a)의 효소를 상기 단계 (b)의 소형 분자와 반응시켜, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드룰라제, 만난아제 및/또는 베타-글루코시다제 효소 반응에 의해 소형 분자를 변형시키는 단계. 하나의 측면에서, 이 방법은 상기 단계 (a)의 효소에 대한 소형 분자 기질을 다수 포함하므로, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드룰라제, 만난아제 및/또는 베타-글루코시다제 효소에 의해 촉진되는 하나 이상의 효소 반응에 의해 생산되는 변형된 소형 분자 라이브러리를 생성시키는 단계를 포함할 수 있다. 하나의 측면에서, 이 방법은 다수의 생체 촉매 반응이 효소에 의해 촉진되는 조건 하에서, 다수의 부가 효소들을 사용하여 다수의 효소 반응에 의해 생산되는 변형된 소형 분자 라이브러리를 제조하는 단계를 포함한다. 다른 측면에서, 이 방법은 바람직한 활성을 나타내는 특성의 변형된 소형 분자가 라이브러리 내에 존재하는지 여부를 결정하기 위해 라이브러리를 시험하는 단계를 추가로 포함할 수 있다. 이 라이브러리를 시험하는 단계는, 원하는 활성을 보유하는 변형된 특정 소형 분자의 존재에 대해 변형된 소형 분자의 일부를 시험함으로써 라이브러리 내에 존재하는 다수의 변형된 소형 분자 중 일부를 생산하는데에 사용된 생체 촉매 반응 중 하나를 제외하고는 모두 체계적으로 제외시키고, 원하는 활성을 보유하는 변형된 특정 소형 분자를 생성하는 하나 이상의 특정 생체 촉매 반응을 동정하는 단계를 추가로 포함할 수 있다.

[0078] 본 발명은 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드룰라제, 만난아제 및/또는 베타-글루코시다제 효소의 기능성 단편을 결정하는 방법을 제공하는데, 이 방법은 다음의 단계들을 포함한다: (a) 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드룰라제, 만난아제 및/또는 베타-글루코시다제 효소를 제공하는 단계[여기서, 상기 효소는 본 발명의 폴리펩티드, 본 발명의 핵산에 의해 암호화되는 폴리펩티드 또는 이의 부분 서열을 포함함]; 및 (b) 상기 단계 (a)의 서열로부터 다수의 아미노산 잔기를 결실시키고, 나머지 부분 서열이 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드룰라제, 만난아제 및/또는 베타-글루코시다제 효소 활성을 보유하는지 여부에 대해 시험함으로써, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드룰라제, 만난아제 및/또는 베타-글루코시다제 효소의 기능성 단편을 결정하는 단계. 하나의 측면에서, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드룰라제, 만난아제 및/또는 베타-글루코시다제 효소 활성은, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드룰라제, 만난아제 및/또는 베타-글루코시다제 효소 기질을 제공하고, 기질 양의 감소 또는 반응 생성물 양의 증가를 확인함으로써 측정된다.

[0079] 본 발명은 실시간 대사 흐름 분석(real-time metabolic flux analysis)을 사용하여, 신규 또는 변형 표현형을 가지는 전 세포를 조작하는 방법을 제공하며, 이 방법은 다음의 단계들을 포함한다: (a) 세포의 유전자 조성을 변형시켜 변형된 세포를 제조하는 단계[여기서, 상기 유전자 조성은 본 발명의 핵산을 세포에 첨가함으로써 변형됨]; (b) 변형된 세포를 배양하여 다수의 변형된 세포를 생산하는 단계; (c) 상기 단계 (b)의 세포 배양액을 실시간으로 모니터링하여 세포에 관한 하나 이상의 대사 매개 변수를 측정하는 단계; 및 (d) 상기 단계 (c)의 데이터를 분석하여, 측정된 매개 변수가, 유사한 조건 하에서 측정된 비변형 세포에서의 유사한 측정치와 상이한지 여부를 결정함으로써, 실시간 대사 흐름 분석법을 이용하여 세포 내에서 조작된 표현형이 나타나는지 확인하는 단계. 하나의 측면에서, 상기 세포의 유전자 조성은, 세포 내에서 서열을 결실시키거나 또는 서열을 변형시키거나, 또는 유전자의 발현을 억제(knock-out)하는 단계를 포함하는 방법에 의해 변형될 수 있다. 하나의 측면에서, 이 방법은 새로이 조작된 표현형을 포함하는 세포를 선별하는 단계를 추가로 포함할 수 있다. 다른 측면에서, 이 방법은 선별된 세포를 배양하여, 새로이 조작된 표현형을 나타내는 신규의 세포 균주를 생산하는 단계를 포함할 수 있다.

[0080] 본 발명은 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드룰라제, 만난아제 및/또는 베타-글루코시다제 효소 폴리펩티드의 열 내성 또는 열 안정성을 증가시키는 방법을 제공하는데, 이 방법은 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드룰라제, 만난아제 및/또는 베타-글루코시다제 효소 폴리펩티드를 글리코실화시키고[여기서, 상기 폴리펩티드는 본 발명의 폴리펩티드 또는 본 발명의 핵산 서열에 의해 암호화되는 폴리펩티드 중 30개 이상의 연속 아미노산을 포함함], 이로써 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드룰라제, 만난아제 및/또는 베타-글루코시다제 폴리펩티드의 열 내성 또는 열 안정성을 증가시키는 단계를 포함

한다. 하나의 측면에서, 상기 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 비활성은 약 37~약 95℃의 온도에서 열 안정성이거나 열 내성일 수 있다.

[0081] 본 발명은 재조합 셀룰라제, 예를 들어 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 폴리펩티드를 세포 내에서 과발현시키는 방법을 제공하는데, 이 방법은 본 발명의 핵산을 포함하는 핵산이나 본 발명의 핵산 서열을 포함하는 벡터를 발현시키는 단계를 포함하며, 여기서, 상기 서열의 상동성은 서열 비교 알고리즘을 이용하는 분석법 또는 시각적 조사에 의해 측정되고, 과발현은 고효율 프로모터, 이시스트론성 벡터(dicistronic vector) 또는 벡터의 유전자 증폭법에 의해 이루어진다.

[0082] 본 발명은 다음의 단계들을 포함하는, 트랜스제닉 식물을 제조하는 방법을 제공한다: (a) 이중 핵산 서열을 세포에 도입하여[여기서, 상기 이중 핵산 서열은 본 발명의 핵산 서열을 포함함], 형질전환된 식물 세포를 생산하는 단계; 및 (b) 형질전환된 세포로부터 트랜스제닉 식물을 생산하는 단계. 하나의 측면에서, 상기 단계 (a)는 식물 세포 원형질체를 전기 천공시키거나 미세 주입하여 이중 핵산 서열을 도입하는 단계를 추가로 포함할 수 있다. 다른 측면에서, 상기 단계 (a)는 또한 DNA 입자 충격법(DNA particle bombardment)에 의해 식물의 조직에 이중 핵산을 직접 도입시키는 단계를 추가로 포함할 수 있다. 대안적으로, 상기 단계 (a)는 아그로박테리움 튜메페이스스(*Agrobacterium tumefaciens*) 숙주를 이용하여 식물 세포 DNA에 이중 핵산 서열을 도입하는 단계를 추가로 포함할 수 있다. 하나의 측면에서, 상기 식물 세포는 사탕수수, 비트, 대두, 토마토, 감자, 옥수수, 쌀, 밀, 담배 또는 보리 세포일 수 있다.

[0083] 본 발명은 다음의 단계들을 포함하는, 식물 세포 내에서 이중 핵산 서열을 발현시키는 방법을 제공한다: (a) 프로모터에 작동 가능하게 연결된 이중 핵산 서열로 식물 세포를 형질전환시키는 단계[여기서, 상기 이중 핵산 서열은 본 발명의 핵산을 포함함]; (b) 상기 이중 핵산 서열이 식물 세포 내에서 발현되는 조건 하에서 식물을 생육시키는 단계. 본 발명은 다음의 단계들을 포함하는, 이중 핵산 서열을 식물 세포 내에서 발현시키는 방법을 제공한다: (a) 프로모터에 작동 가능하게 연결된 이중 핵산 서열로 식물 세포를 형질전환시키는 단계[여기서, 상기 이중 핵산 서열은 본 발명의 서열을 포함함]; (b) 상기 이중 핵산 서열이 식물 세포 내에서 발현되는 조건 하에서 식물 세포를 생육시키는 단계.

[0084] 본 발명은 본 발명의 폴리펩티드, 또는 본 발명의 핵산에 의해 암호화되는 폴리펩티드를 포함하는 사료 또는 식품을 제공한다. 하나의 측면에서, 본 발명은 본 발명의 폴리펩티드를 포함하는 식품, 사료, 액체 예를 들어, 음료(예를 들어, 과일 주스 또는 맥주), 빵 또는 가루 반죽이나 빵 제품, 또는 음료 전구 물질(예를 들어, 맥아즙)을 제공한다. 본 발명은 본 발명의 폴리펩티드 예를 들어, 본 발명의 핵산에 의해 암호화되는 폴리펩티드를 포함하는 동물용 식품 또는 영양 보충제를 제공한다.

[0085] 하나의 측면에서, 상기 식품 또는 영양 보충제 중에 포함된 폴리펩티드는 글리코실화될 수 있다. 본 발명은 본 발명의 폴리펩티드 예를 들어, 본 발명의 핵산에 의해 암호화되는 폴리펩티드를 포함하는 식용 효소 전달 매트릭스를 제공한다. 하나의 측면에서, 상기 전달 매트릭스는 펠릿을 포함한다. 하나의 측면에서, 상기 폴리펩티드는 글리코실화될 수 있다. 하나의 측면에서, 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성은 열 내성이다. 다른 측면에서, 상기 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성은 열 안정성이다.

[0086] 본 발명은 본 발명의 폴리펩티드를 포함하는 식품, 사료 또는 영양 보충제를 제공한다. 본 발명은 셀룰라제, 예를 들어 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 동물 식이의 영양 보충제로서 사용하는 방법을 제공하는데, 이 방법은 다음의 단계들을 포함한다: 본 발명의 폴리펩티드를 이루는 아미노산 중 30개 이상의 연속 아미노산을 포함하는 셀룰라제, 예를 들어 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 함유하는 영양 보충제를 제조하는 단계; 및 동물에 이 영양 보충제를 투여하는 단계. 상기 동물은 인간, 반추위 동물 또는 단위 동물일 수 있다. 셀룰라제, 예를 들어 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소는 박테리아, 효모, 식물, 곤충, 진균 및 동물로 이루어진 군으로부터 선택되는 유기체 내에서 셀룰라제, 예를 들어 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 암호화하는 폴리뉴클레오티드를 발현시킴으로써 제조될 수 있다. 상기 유기체는 에스. 폼베(*S.pombe*), 에스. 세레비시아에(*S.cerevisiae*), 피치아 파스토리스(*Pichia pastoris*), 이. 콜라이(*E.coli*), 스트렙토마이세스 종(*Streptomyces sp.*), 바실러스 종(*Bacillus sp.*) 및 락토바실러스 종(*Lactobacillus sp.*)으로 이루어진 군으로부터 선택될 수 있다.

[0087] 본 발명은 열 안정성 재조합 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는

베타-글루코시다제 효소 예를 들어, 본 발명의 폴리펩티드를 포함하는 식용 효소 전달 매트릭스를 제공한다. 본 발명은 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 보조제를 동물에 전달하는 방법을 제공하는데, 이 방법은 다음의 단계들을 포함한다: 과립형 식용 담체와 열 안정성 재조합 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 포함하는 펠릿 형태의 식용 효소 전달 매트릭스를 제조하는 단계[여기서, 상기 펠릿은 이에 함유되어 있는 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 수성 매질에 용이하게 분산시킴]; 및 이 식용 효소 전달 매트릭스를 동물에 투여하는 단계. 상기 재조합 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소는 본 발명의 폴리펩티드를 포함할 수 있다. 상기 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소는 글리코실화되어, 펠릿화 조건에서 열 안정성을 가질 수 있다. 상기 전달 매트릭스는 곡물 배아와 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 포함하는 혼합물을 펠릿화함으로써 생산될 수 있다. 상기 펠릿화 조건은 증기 처리를 포함할 수 있다. 상기 펠릿화 조건은 약 5분 동안 약 80℃ 이상의 온도로 유지시키는 것을 포함할 수 있으며, 이때 상기 효소는, 효소 1 mg당 350~약 900 유닛의 비활성을 갖는다.

[0088] 하나의 측면에서, 본 발명은 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소, 또는 본 발명의 핵산에 의해 암호화되는 폴리펩티드를 포함하는 약학 조성물을 제공한다. 하나의 측면에서, 본 발명의 약학 조성물은 소화 보조제로서 작용한다.

[0089] 임의의 측면에서, 셀룰라제-함유 화합물은 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성을 보유하는 본 발명의 폴리펩티드와 pH 약 3.0~9.0, 10.0 또는 11.0 이상에서 접촉하게 된다. 다른 측면에서, 셀룰로스 함유 화합물은 온도 약 55℃, 60℃, 65℃, 70℃, 75℃, 80℃, 85℃ 또는 90℃ 이상에서 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소와 접촉하게 된다.

[0090] 본 발명의 하나 이상의 측면에 관한 상세한 설명은 이하 첨부된 도면과 상세한 설명에 제시되어 있다. 기타 본 발명의 특징, 목적 및 이점들은 이하 상세한 설명, 도면 그리고 청구항으로부터 분명히 알 수 있을 것이다.

[0091] 본원에 인용된 모든 공보, 특허, 특허 출원, GenBank 서열 및 ATCC 기탁 번호는 총체적인 목적을 위해 참고용으로 인용되어 있음을 분명히 하고자 한다.

발명의 상세한 설명

[0108] 본 발명은 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 활성을 보유하는 폴리펩티드, 이 폴리펩티드를 암호화하는 폴리뉴클레오티드와, 이 폴리뉴클레오티드 및 폴리펩티드를 제조 및 사용하는 방법을 제공한다. 본 발명은 또한 셀룰라제 효소 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소와, 이 효소들을 암호화하는 폴리뉴클레오티드, 그리고 이러한 폴리뉴클레오티드와 폴리펩티드의 용도를 제공한다.

[0109] 하나의 측면에서, 본 발명은 촉매 반응 속도가 증가하였고 기질 가수분해 과정이 개선된, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제를 제공한다. 이와 같이 촉매 반응의 효율이 증가함으로써 인하여, 에탄올을 생산하는데에 미생물에 의해 추후 사용될 당의 생산 효율이 증가하게 된다. 하나의 측면에서, 본 발명의 효소를 생산하는 미생물은 에탄올-생산 미생물과 함께 사용된다. 그러므로, 본 발명은 에탄올을 생산하는 방법과, 이 에탄올을 주성분으로 하는 "청정 연료"(예를 들어, 바이오에탄올을 사용하는 운송용 청정 연료)를 생산하는 방법을 제공한다.

[0110] 하나의 측면에서, 본 발명은 본 발명의 효소, 폴리펩티드 또는 폴리뉴클레오티드를 포함하는 조성물(예를 들어, 효소 제제, 사료, 약품, 식품 보조제)를 제공한다. 이러한 조성물은 다양한 형태 예를 들어, 액체, 겔, 환제, 정제, 스프레이, 분말, 식품, 사료 펠릿 또는 나노캡슐화 된 형태를 비롯하여 캡슐화된 형태의 것으로 제형화될 수 있다.

[0111] 셀룰라제 활성 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 활성을 측정하기 위한 검정법 예를 들어, 폴리펩티드가 셀룰라제 활성 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 활성을 보유하는지 여부를 결정하기 위한 검정법에 관하여는 당 업계에 널리 공지되어 있으며, 이는 본 발명의 범위 내에 포함된다[예를 들어, Baker WL, Panow A, Estimation of cellulase activity using a glucose-oxidase- Cu(II) reducing assay for glucose, J Biochem Biophys

Methods. 1991 Dec, 23(4):265-73; Sharrock KR, Cellulase assay methods: a review, J Biochem Biophys Methods. 1988 Oct, 17(2):81-105; Carder JH, Detection and quantitation of cellulase by Congo red staining of substrates in a cup-plate diffusion assay, Anal Biochem. 1986 Feb 15, 153(1):75-9; Canevascini G., A cellulase assay coupled to cellobiose dehydrogenase, Anal Biochem. 1985 Jun, 147(2):419-27; Huang JS, Tang J, Sensitive assay for cellulase and dextranase. Anal Biochem. 1976 Jun, 73(2):369-77 참조].

[0112] 본 발명에 의해 이용되는 반응 조건의 pH는 본 발명이 제공하는 또 다른 가변성 매개 변수이다. 임의의 측면에서, 반응 pH는 약 3.0~약 9.0이다. 다른 측면에서, pH는 약 4.5이거나 약 7.5이거나 약 9이다. 알칼리성 조건 하에서 수행되는 반응의 조건은 또한 예를 들어, 본 발명의 효소를 사용하는 일부 산업 분야나 약학 분야에서 유리할 수도 있다.

[0113] 본 발명은 다양한 형태 및 제형을 띠고 있는, 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 폴리펩티드를 제공한다. 본 발명의 방법에 있어서, 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 폴리펩티드는 다양한 형태와 제형으로 사용된다. 예를 들어, 정제된 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 폴리펩티드는 바이오에탄올 생산 또는 약학적 용도나 식품 보조제 분야에서 사용되고 있는 효소 제제 중에 사용될 수 있다. 대안적으로, 본 발명의 효소는 바이오에탄올을 생산하는 공정, 청정 연료를 생산하는 공정, 유기성 폐기물(biowaste)을 처리하는 공정, 식품, 액체 또는 사료 등을 가공하는 공정 등에서 직접 사용될 수 있다.

[0114] 대안적으로, 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 폴리펩티드는 당 업계에 공지된 방법을 사용하여 미생물 내에서 발현될 수 있다. 다른 측면에서, 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 폴리펩티드는 본 발명의 방법에 사용되기 이전에 고형 지지체 상에 고정화될 수 있다. 고형 지지체 상에 효소를 고정하는 방법은 당 업계에 널리 공지되어 있다[예를 들어, J. Mol. Cat. B: Enzymatic 6 (1999) 29-39; Chivata와 다수 Biocatalysis: Immobilized cells and enzymes, J Mol. Cat. 37 (1986) 1-24; Sharma와 다수, Immobilized Biomaterials Techniques and Applications, Angew. Chem. Int. Ed. Engl. 21 (1982) 837-54; Laskin (Ed.), Enzymes and Immobilized Cells in Biotechnology].

[0115] **핵산, 프로브 및 억제 분자**

[0116] 본 발명은 분리 및 재조합 핵산[예를 들어, 이하 표 1a~1c, 2 및 3과, 실시예 1 및 4, 그리고 서열 목록 참조]; 본 발명의 대표적인 폴리뉴클레오티드 서열을 포함하는, 폴리펩티드를 암호화하는 핵산[예를 들어, 표 1a~1c 및 서열 목록 참조] 예를 들어, 본 발명의 핵산을 포함하는 발현 카세트 예를 들어, 발현 벡터 및 다양한 클로닝 비이클을 제공한다. 본 발명은 또한 본 발명의 핵산을 이용하여 신규의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 폴리펩티드 서열을 발견, 동정 또는 분리하는 방법을 포함하기도 한다. 본 발명은 또한 본 발명의 핵산을 이용하여 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 암호화 유전자와 전사체의 발현을 억제하는 방법도 포함한다.

[0117] 뿐만 아니라, 본 발명은 예를 들어, 합성 결찰 재조합법, 최적화된 유도 진화 시스템 및/또는 포화 돌연변이 유발법 예를 들어, 유전자 위치 포화 돌연변이 유발법(GSSM)에 의해, 본 발명의 핵산을 변형시키는 방법 예를 들어, 본 발명의 핵산 변이체를 생산하는 방법을 제공한다. "포화 돌연변이 유발법(saturation mutagenesis)", 유전자 위치 포화 돌연변이 유발법(Gene Site Saturation Mutagenesis) 또는 "GSSM"이란 용어는, 이하에 상세히 기술되어 있는 바와 같이, 축퇴성 올리고뉴클레오티드 프라이머를 사용하여 점 돌연변이를 폴리뉴클레오티드에 도입시키는 방법을 포함한다. "최적화 유도 진화 시스템(optimized directed evolution system)" 또는 "최적화 유도 진화"란 용어는, 이하에 상세히 기술되어 있는 바와 같이, 관련된 핵산 서열 예를 들어, 관련된 유전자의 단편들을 재조합하는 방법을 포함한다. "합성 결찰 재조합법(synthetic ligation reassembly)" 또는 "SLR"이란 용어는, 이하에 상세히 기술되어 있는 바와 같이, 확률에 의하지 않는 방식으로 올리고뉴클레오티드 단편을 결찰시키는 방법을 포함한다. "변이체"란 용어는, 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 활성을 여전히 보유하는 하나 이상의 염기쌍, 각각의 코돈, 인트론, 엑손, 또는 아미노산 잔기에서 변형이 일어난 본 발명의 폴리뉴클레오티드 또는 폴리펩티드를 의미하는 것이다. 변이체는 예를 들어, 오류 빈발 PCR, 서플링, 올리고뉴클레오티드 지정 돌연변이 유발법, 어셈블리

PCR, 유성 PCR 돌연변이 유발법, 생체내 돌연변이 유발법, 카세트 돌연변이 유발법, 반복적 양상불 돌연변이 유발법, 지수적 양상불 돌연변이 유발법, 위치 특이적 돌연변이 유발법, 유전자 재조합법, GSSM 및 상기 방법들의 임의의 조합법을 포함하는 다수의 방법에 의해 생산될 수 있다.

[0118] 본 발명의 핵산은 예를 들어, cDNA 라이브러리의 클로닝 및 발현, PCR에 의한 메세지 또는 게놈형 DNA 증폭법 등에 의해 생산, 분리 및/또는 조작될 수 있다. 예를 들어, 본 발명의 대표적인 서열은 처음에는 환경적 공급원으로부터 유래하였다. 그러므로, 하나의 측면에서, 본 발명은, 예를 들어, 환경의 혼합 배양액 또는 박테리아 공급원과 같은 일반적인 공급원으로부터 유래한다는 점에서 공통된 신규성을 가지는 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소-암호화 핵산, 및 이 핵산에 의해 암호화되는 폴리펩티드를 제공한다.

[0119] 본 발명의 방법을 실시함에 있어서, 상동성 유전자는 본원에 기술된 바와 같이, 주형 핵산을 조작함으로써 변형될 수 있다. 본 발명은 당 업계에 공지된 임의의 방법, 프로토콜 또는 장치(과학 논문과 특허 문헌에 자세히 기술되어 있음)와 연계해서 실시될 수 있다.

[0120] 본원에 사용된 "핵산" 또는 "핵산 서열"이란 어구는, 올리고뉴클레오티드, 뉴클레오티드 또는 폴리펩티드, 또는 이것들 중 임의의 것의 단편, 단일 가닥 또는 이중 가닥일 수 있으며 센스 또는 안티센스 (상보성) 가닥일 수 있는 유전자형 또는 합성 기원의 DNA 또는 RNA, 펩티드 핵산(PNA), 또는 천연 또는 합성 기원의 임의의 DNA-유사 또는 RNA-유사 물질을 의미하는 것이다. "핵산" 또는 "핵산 서열"이란 어구는, 올리고뉴클레오티드, 뉴클레오티드 또는 폴리뉴클레오티드, 또는 이것들 중 임의의 것의 단편, 단일 가닥 또는 이중 가닥일 수 있으며 센스 또는 안티센스 가닥일 수 있는 유전자형 또는 합성 기원의 DNA 또는 RNA(예를 들어, mRNA, rRNA, tRNA, iRNA), 펩티드 핵산(PNA), 또는 천연 또는 합성 기원의 임의의 DNA-유사 또는 RNA-유사 물질 예를 들어, iRNA 및 리보핵 단백질(예를 들어, 이중 가닥 iRNA 예를 들어, iRNP)를 포함한다. 상기 용어는 천연 뉴클레오티드의 공지된 유사체를 함유하는 핵산 즉, 올리고뉴클레오티드를 포함한다. 상기 용어는 또한 합성 주쇄를 가지는 핵산-유사 구조를 포함한다[예를 들어, Mata (1997) Toxicol. Appl. Pharmacol. 144:189-197; Strauss- Soukup (1997) Biochemistry 36:8692-8698; Samstag (1996) Antisense Nucleic Acid Drug Dev 6:153-156 참조]. "올리고뉴클레오티드"는 화학적으로 합성될 수 있는 단일 가닥 폴리데옥시리보뉴클레오티드 또는 2개의 상보성 폴리데옥시뉴클레오티드 가닥을 포함한다. 이러한 합성 올리고뉴클레오티드는 5' 말단에 인산기를 가지지 않으므로, 키나제의 존재하에 ATP와 함께 인산염을 첨가하지 않고서도 다른 올리고뉴클레오티드에 결합되지 않을 것이다. 합성 올리고뉴클레오티드는 탈인산화되지 않은 단편에 결합될 수 있다.

[0121] 특정 폴리펩티드 또는 단백질"의 암호화 서열" 또는 특정 폴리펩티드 또는 단백질을 암호화하는 뉴클레오티드"란, 적당한 조절 서열의 제어하에 있을 때, 폴리펩티드 또는 단백질로 전사 및 번역되는 핵산 서열이다. "유전자"란 용어는, 폴리펩티드 사슬을 생산하는데에 관여하는 DNA 분절을 의미하는 것으로서; 이는 암호화 부위(리더 및 트레일러)의 앞 부분과 뒷부분에 위치하는 부위들을 포함하고, 가능한 경우에는, 각각의 암호화 분절들(엑손) 사이에 개입된 서열(인트론)도 포함한다. 프로모터에서 전사를 개시하는 RNA 중합 효소가 암호화 서열을 mRNA로 전사할 때, 이 프로모터 서열은 암호화 서열에 "작동 가능하도록 결합되어 있는" 상태라고 한다. 본원에 사용된 "작동 가능하도록 결합되어 있는"이란, 2개 이상의 핵산(예를 들어, DNA) 분절들간의 기능상 관계를 의미하는 것이다. 이는 곧, 전사된 서열에 대한 전사 조절 서열의 기능상 관계를 의미하는 것일 수도 있다. 예를 들어, 만일 프로모터가 적당한 숙주 세포 또는 기타 발현 시스템 내에서 암호화 서열의 전사를 촉진하거나 조정한다면, 이 프로모터는 암호화 서열 예를 들어, 본 발명의 핵산에 작동 가능하도록 결합되어 있는 것이다. 일반적으로, 전사된 서열에 작동 가능하도록 결합되어 있는 프로모터 전사 조절 서열은 전사된 서열과 물리적으로 인접하여 있는 것으로서, 다시 말하면, 이 서열은 시스-작용(cis-acting) 위치에 있는 것이다. 그러나, 일부 전사 조절 서열 예를 들어, 인핸서는 이것이 강화하고자 하는 전사와 관계된 암호화 서열과 물리적으로 인접하여 존재하거나 또는 이 암호화 서열에 매우 근접하여 위치하고 있을 필요는 없다.

[0122] 본원에 사용된 "발현 카세트"라는 용어는, 구조 유전자를 발현시킬 수 있는 뉴클레오티드 서열(즉, 본 발명의 단백질 암호화 서열 예를 들어, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 암호화 서열)을 수용할 수 있는 숙주 내에서 발현되는 뉴클레오티드 서열을 의미하는 것이다. 발현 카세트는 적어도 폴리펩티드 암호화 서열과 작동 가능하도록 결합되어 있는 프로모터를 포함하고; 경우에 따라, 적어도, 기타 서열 예를 들어, 전사 종결 신호와 작동 가능하도록 결합되어 있는 프로모터를 포함하기도 한다. 발현을 실현하는데에 필요하거나 도움이 되는 부가 요소 예를 들어, 인핸서 및 알파-인자도 사용될 수 있다. 그러므로, 발현 카세트로서는 또한 플라스미드, 발현 벡터, 재조합 바이러스, 그리고 임의의 형태의 재조합 "나출 DNA" 벡터 등을 포함하기도 한다. "벡터"는 세포를 감염, 형질 감염, 일시적으로 또는 영

구적으로 형질 도입시킬 수 있는 핵산을 포함한다. 벡터는 나출 핵산, 또는 단백질이나 지질과 함께 복합체를 형성하는 핵산일 수 있음을 알 수 있을 것이다. 경우에 따라, 벡터는 바이러스 또는 박테리아 핵산 및/또는 단백질, 및/또는 막(예를 들어, 세포막, 바이러스 지질 외피 등)을 포함하기도 한다. 벡터로서는 DNA 단편이 결합하여 복제될 수 있는 레플리콘(예를 들어, RNA 레플리콘 및 박테리오파지)을 포함하나, 이에 한정되는 것은 아니다. 그러므로, 벡터는 RNA, 자율적 자가-복제 환형 또는 선형 DNA 또는 RNA(예를 들어, 플라스미드 및 바이러스 등, 예를 들어, 미국 특허 제5,217,879호 참조)를 포함하며, 또한 발현 및 비-발현 플라스미드 둘 다를 포함하나, 이에 한정되는 것은 아니다. 재조합 미생물 또는 세포 배양액이 "발현 벡터"를 수용하는 것으로서 기술되어 있을 경우, 이는 잉여 염색체성 환형 및 선형 DNA 및 숙주의 염색체에 통합된 DNA 둘 다를 포함하는 것이다. 벡터가 숙주 세포에 의해 유지되고 있는 경우, 벡터는 유사 분열중 자율적 구조물로서 세포에 의해 안정하게 복제될 수 있거나, 또는 숙주의 게놈 내에 통합된다.

[0123] 본원에 사용된 "재조합체"란 용어는, 천연의 환경에서는 인접하여 존재하지 않던 "주쇄" 핵산에 인접하여 위치하는 핵산을 포함한다. 하나의 측면에서, 핵산이 "증폭"되었다 함은, 핵산 주쇄 분자의 무리중에 존재하는 핵산 삽입부의 수가 약 5% 이상임을 의미할 것이다. 본 발명에 의한 주쇄 분자는 핵산 예를 들어, 발현 벡터, 자가-복제 핵산, 바이러스, 통합 핵산 및 기타 벡터, 또는 목적 핵산 삽입부를 유지하거나 조작하는데 사용되는 핵산을 포함한다. 하나의 측면에서, 증폭된 핵산이란, 재조합 주쇄 분자의 무리중에 존재하는 핵산 삽입부의 수가 약 15% 이상인 경우를 말한다. 하나의 측면에서, 증폭된 핵산이란, 재조합 주쇄 분자의 무리중에 존재하는 핵산 삽입부의 수가 약 50% 이상인 경우를 말한다. 하나의 측면에서, 증폭된 핵산이란, 재조합 주쇄 분자의 무리중에 존재하는 핵산 삽입부의 수가 약 90% 이상인 경우를 말한다.

[0124] 본 발명의 하나의 측면은 본 발명의 서열 중 하나를 포함하는 분리된 또는 재조합 핵산, 또는 본 발명의 핵산 중 적어도 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400 또는 500 개 이상의 연속된 염기를 포함하는 단편에 관한 것이다. 분리된 또는 재조합 핵산은 DNA 예를 들어, cDNA, 게놈형 DNA 및 합성 DNA를 포함할 수 있다. 상기 DNA는 이중 가닥형 또는 단일 가닥형일 수 있으며, 만일 단일 가닥일 경우에는 암호화 가닥 또는 비-암호화(안티센스) 가닥일 수 있다. 대안적으로, 분리된 또는 재조합 핵산은 RNA를 포함한다.

[0125] 본 발명의 분리된 또는 재조합 핵산은 본 발명의 폴리펩티드 중 하나, 또는 본 발명의 폴리펩티드 중 하나의 연속적 아미노산을 적어도 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100 또는 150개 이상 포함하는 단편을 제조하는데 사용될 수 있다. 그러므로, 본 발명의 다른 측면은 본 발명의 폴리펩티드 중 하나, 또는 본 발명의 폴리펩티드 중 하나의 연속적 아미노산을 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100 또는 150개 이상 포함하는 단편을 암호화하는 분리된 또는 재조합 핵산에 관한 것이다. 이러한 핵산의 암호화 서열은, 유전자 암호의 중복성 또는 축퇴성으로 인하여, 본 발명의 핵산 중 하나의 암호화 서열 중 하나와 동일할 수 있거나, 또는 본 발명의 폴리펩티드 중 하나의 연속적 아미노산을 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100 또는 150개 이상 가지는 본 발명의 폴리펩티드를 암호화하는 상이한 암호화 서열일 수 있다. 유전자 암호는 당 업자에게 널리 공지되어 있으며, 예를 들어, 문헌[B. Lewin, Genes VI, Oxford University Press, 1997, 214 페이지]에서 살펴볼 수 있다.

[0126] 본 발명의 핵산 암호화 폴리펩티드는 다음의 것들을 포함하나, 이에 한정되는 것은 아니다: 본 발명의 핵산의 암호화 서열 및 부가 암호화 서열 예를 들어, 리더 서열 또는 프로-단백질 서열 및 비-암호화 서열 예를 들어, 인트론 또는 암호화 서열의 5' 및/또는 3' 말단에 존재하는 비-암호화 서열. 그러므로, 본원에 사용된 "폴리펩티드를 암호화하는 폴리뉴클레오티드"라는 용어는, 폴리펩티드에 대한 암호화 서열을 포함하는 폴리뉴클레오티드와, 부가 암호화 서열 및/또는 비-암호화 서열을 포함하는 폴리뉴클레오티드를 포함한다.

[0127] 하나의 측면에서, 본 발명의 핵산 서열은 통상의 기술 예를 들어, 위치 지정 돌연변이 유발법이나 당 업자에게 익숙한 기타 기술을 이용하여 돌연변이 유발되어, 그 결과 본 발명의 폴리뉴클레오티드에 침묵 변이가 도입된다. 본원에 사용된 "침묵 변이"란, 예를 들어, 폴리뉴클레오티드에 의해 암호화되는 아미노산 서열을 변형시키지 않는 변이를 포함한다. 이러한 변이는 숙주 유기체 내에서 자주 발생하는 코돈 또는 코돈 쌍의 도입에 의해 폴리펩티드를 암호화하는 벡터를 함유하는 숙주 세포에 의해 생산되는 폴리펩티드의 수준을 증가시키는데에 바람직할 수 있다.

[0128] 본 발명은 또한 본 발명의 폴리펩티드에 아미노산 치환, 부가, 결실, 융합 및 절단을 일으키는 뉴클레오티드 변이가 발생한 폴리뉴클레오티드에 관한 것이다. 이러한 뉴클레오티드 변이는 기술 예를 들어, 위치 지정 돌연변이 유발법, 랜덤한 화학적 돌연변이 유발법, 엑소뉴클레아제 III 결실법 및 기타 재조합 DNA 기술을 사용하여 도입될 수 있다. 대안적으로, 이러한 뉴클레오티드 변이는, 고 염중도, 중간 정도의 염중도 또는 저 염중도의

조건하에서, 본 발명의 서열(또는 이에 상보성인 서열) 중 하나의 연속된 염기를 적어도 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400 또는 500개 이상 포함하는 프로브에 특이적으로 혼성화하는 핵산을 동정함으로써 분리되는 자연 발생 대립 형질 변이체일 수 있다.

[0129] 일반적인 기술

[0130] 본 발명을 실시하는데에 사용되는 핵산은, 그것이 RNA, siRNA, miRNA, 안티센스 핵산, cDNA, 게놈형 DNA, 벡터, 바이러스 또는 이의 혼성체인 아니든, 다양한 공급원으로부터 분리되어, 유전적으로 조작되고, 증폭 및/또는 재조합적으로 발현/생산될 수 있다. 상기 핵산들로부터 생산된 재조합 폴리펩티드(예를 들어, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소)는 각각 분리 또는 클로닝되고, 목적으로 하는 활성에 대해 시험될 수도 있다. 임의의 재조합 발현 시스템 예를 들어, 박테리아, 포유동물, 효모, 곤충 또는 식물 세포 발현 시스템이 사용될 수 있다.

[0131] 대안적으로, 상기와 같은 핵산은 널리 공지된 화학적 합성 기술에 의해 시험관 내에서 합성될 수 있다[예를 들어, Adams (1983) J. Am. Chem. Soc. 105:661; Belousov (1997) Nucleic Acids Res. 25:3440-3444; Frenkel (1995) Free Radic. Biol. Med. 19:373-380; Blommers (1994) Biochemistry 33:7886-7896; Narang (1979) Meth. Enzymol. 68:90; Brown (1979) Meth. Enzymol. 68:109; Beaucage (1981) Tetra. Lett. 22:1859; 미국 특허 제4,458,066호].

[0132] 핵산을 조작하는 기술 예를 들어, 서브클로닝, 프로브 표지화 (예를 들어, 클레노우 중합 효소를 사용하는 랜덤-프라이머 표지화 방법, Nick 번역 및 증폭), 서열 결정 및 혼성화 등과 같은 기술은 과학 논문 및 특허 문헌에 자세히 기술되어 있다[예를 들어, Sambrook, ed., MOLECULAR CLONING: A LABORATORY MANUAL (2ND ED.), Vols. 1-3, Cold Spring Harbor Laboratory, (1989); CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Ausubel, ed. John Wiley & Sons, Inc., New York (1997); LABORATORY TECHNIQUES IN BIOCHEMISTRY AND MOLECULAR BIOLOGY: HYBRIDIZATION WITH NUCLEIC ACID PROBES, Part I. Theory and Nucleic Acid Preparation, Tijssen, ed. Elsevier, N. Y. (1993) 참조].

[0133] 본 발명의 방법을 실시하는데에 사용되는 핵산의 제조 및 조작 방법으로서 유용한 기타 방법은 게놈형 시료로부터 클로닝하는 방법으로서, 원하는 경우에는 예를 들어, 게놈형 클론 또는 cDNA 클론으로부터 분리 또는 증폭된 삽입체를 스크리닝 및 재 클로닝하는 방법이 있다. 본 발명의 방법에 사용되는 핵산의 공급원은 예를 들어, 포유동물 인공 염색체(MAC)(예를 들어, 미국 특허 제5,721,118호; 동 제6,025,155호); 인간 인공 염색체(예를 들어, Rosenfeld (1997) Nat. Genet. 15:333-335); 효모 인공 염색체(YAC); 박테리아 인공 염색체(BAC); P1 인공 염색체(예를 들어, Woon (1998) Genomics 50:306-316); P1-유래 벡터(PAC)(예를 들어, Kern (1997) Biotechniques 23:120-124); 코스미드, 재조합 바이러스, 파지 또는 플라스미드에 포함되어 있는 게놈형 라이브러리 또는 cDNA 라이브러리를 포함한다.

[0134] 하나의 측면에서, 본 발명의 폴리펩티드를 암호화하는 핵산은 번역된 폴리펩티드 또는 이의 단편의 분비를 유도할 수 있는 리더 서열과 함께 적당한 단계에서 조립된다.

[0135] 본 발명은 융합 단백질과 이를 암호화하는 핵산을 제공한다. 본 발명의 폴리펩티드는 원하는 특성 예를 들어, 강화된 안정성 또는 단순화된 정제 특성을 부여하는 이중 펩티드 또는 폴리펩티드 예를 들어, N-말단 동정 펩티드에 융합될 수 있다. 본 발명의 펩티드 또는 폴리펩티드는 또한 예를 들어, 면역원성이 더욱 큰 펩티드를 생산하거나, 재조합 합성된 펩티드를 더욱 용이하게 분리하거나, 항체 및 항체-발현 B 세포를 더욱 용이하게 동정 및 분리하기 위해서, 이 펩티드 또는 폴리펩티드에 결합되어 있는 하나 이상의 부가 도메인을 보유하는 융합 단백질로서 합성 및 발현될 수도 있다. 검출 및 정제 촉진 도메인으로서 예를 들어, 고정화된 금속상에서 정제될 수 있도록 만들어주는 금속 킬레이트화 펩티드 예를 들어, 폴리히스티딘 구역 및 히스티딘-트립토판 모듈, 고정화된 면역 글로불린상에서 정제될 수 있도록 만들어주는 단백질 A 도메인, 그리고 FLAGS 연장/친화성 정제 시스템(Immunex Corp, Seattle WA)에 이용되는 도메인을 포함한다. 절단 가능한 링커 서열 예를 들어, 인자 Xa 또는 엔테로키나제(Invitrogen, San Diego CA)를 정제용 도메인과 모터프-포함 펩티드 또는 폴리펩티드 사이에 포함시켜 정제를 촉진할 수 있다. 예를 들어, 발현 벡터는 "6개의 히스티딘 잔기→티오리독신→엔테로키나제 절단 위치"에 결합되어 있는 에피토프-암호화 핵산 서열을 포함할 수 있다[예를 들어, Williams (1995) Biochemistry 34:1787-1797; Dobeli (1998) Protein Expr. Purif. 12:404-414 참조]. 히스티딘 잔기는 검출 및 정제를 촉진하는 반면에, 엔테로키나제 절단 위치는 융합 단백질의 나머지 부분으로부터 에피토프를 정제하는 수단을 제공한다. 융합 단백질을 암호화하는 벡터와 관련된 기술 및 융합 단백질의 사용과 관련된 기술은 과학 논문 및 특허 문헌에 상세히 기술되어 있다[예를 들어, Kroll (1993) DNA Cell. Biol., 12:441-53].

- [0136] 전사 및 번역 제어 서열
- [0137] 본 발명은 발현(예를 들어, 전사 또는 번역) 제어 서열(들) 예를 들어, 프로모터 또는 인핸서에 작동 가능하도록 결합되어 있어서, RNA 합성/발현을 유도 또는 조정하는 본 발명의 핵산(예를 들어, DNA) 서열을 제공한다. 이러한 발현 제어 서열은 발현 벡터 내에 존재할 수 있다. 대표적인 박테리아 프로모터로서는 lacI, lacZ, T3, T7, gpt, 람다 P_R, P_L 및 trp를 포함한다. 대표적인 진핵 생물 프로모터로서는 CMV 즉시 초기 프로모터, HSV 티미딘 키나제 프로모터, 초기 및 후기 SV40 프로모터, 레트로바이러스의 LTR 프로모터 및 마우스의 메탈로티오네신 I 프로모터를 포함한다.
- [0138] 본원에 사용된 "프로모터"란 용어는, 세포 예를 들어, 식물 또는 동물 세포 내에서 암호화 서열의 전사를 유도할 수 있는 모든 서열을 포함한다. 그러므로, 본 발명의 구조물에 사용되는 프로모터는 유전자의 전사 타이밍 및/또는 속도를 조절 또는 조정하는 것에 관여하는 시스-작용 전사 제어 요소 및 조절 서열을 포함한다. 예를 들어, 프로모터는, 전사 조절에 관여하는 시스-작용 전사 제어 요소 예를 들어, 인핸서, 프로모터, 전사 종결 인자, 복제 기원, 염색체 통합 서열, 5' 및 3' 비번역 부위 또는 인트론 서열일 수 있다. 이러한 시스-작용 서열은 단백질 또는 기타 생체 분자와 상호 작용하여, 전사를 수행(작동 개시/작동 종료, 조절 또는 조정 등)할 수 있다. "구성적" 프로모터는 대부분의 환경적 조건과 발달 상태 또는 세포 분화시에 연속적으로 발현을 유도하는 프로모터이다. "유도 가능한" 또는 "조절 가능한" 프로모터는, 환경 조건 또는 발달 조건의 영향에 따라서 본 발명의 핵산의 발현을 유도한다. 유도성 프로모터에 의해 전사를 진행할 수 있는 환경 조건의 예로서는, 혐기 조건, 고온, 건조 또는 빛의 존재 여부를 포함한다.
- [0139] "조직-특이적" 프로모터는 예를 들어, 식물이나 동물의 특정 세포 또는 조직 또는 기관 내에서만 활성을 보유하는 전사 제어 요소이다. 조직-특이적 조절은, 소정의 조직에 특이적인 단백질을 암호화하는 유전자가 발현되는지 여부를 확인할 수 있는 임의의 내인성 요인에 의해서 이루어질 수 있다. 이러한 요인은 포유동물 및 식물에 존재하여 특이 조직을 발달시킬 수 있는 것으로 알려져 있다.
- [0140] 박테리아 내에서 폴리펩티드를 발현시키는데에 적당한 프로모터로서는 이. 콜라이의 lac 또는 trp 프로모터, lacI 프로모터, lacZ 프로모터, T3 프로모터, T7 프로모터, gpt 프로모터, 람다 P_R 프로모터, 람다 P_L 프로모터, 해당 효소 예를 들어, 3-포스포글리세레이트 키나제(PGK)를 암호화하는 오페론으로부터 유래하는 프로모터, 그리고 산성 포스파타제 프로모터를 포함한다. 진핵 생물 프로모터로서는 CMV 즉시 초기 프로모터, HSV 티미딘 키나제 프로모터, 열 충격 프로모터, 초기 및 후기 SV40 프로모터, 레트로바이러스의 LTR 프로모터 및 마우스의 메탈로티오네신-I 프로모터를 포함한다. 원핵 생물 또는 진핵 생물 세포 또는 이것들의 바이러스 유전자의 발현을 제어하는 것으로 공지되어 있는 기타 프로모터도 사용될 수 있다. 박테리아 내에서 폴리펩티드 또는 이의 단편을 발현시키는데에 적당한 프로모터로서는 이. 콜라이의 lac 또는 trp 프로모터, lacI 프로모터, lacZ 프로모터, T3 프로모터, T7 프로모터, gpt 프로모터, 람다 P_R 프로모터, 람다 P_L 프로모터, 해당 효소 예를 들어, 3-포스포글리세레이트 키나제(PGK)를 암호화하는 오페론으로부터 유래하는 프로모터, 그리고 산성 포스파타제 프로모터를 포함한다. 진균 프로모터로서는 α-인자 프로모터를 포함한다. 진핵 생물 프로모터로서는 CMV 즉시 초기 프로모터, HSV 티미딘 키나제 프로모터, 열 충격 프로모터, 초기 및 후기 SV40 프로모터, 레트로바이러스의 LTR 프로모터 및 마우스의 메탈로티오네신-I 프로모터를 포함한다. 원핵 생물 또는 진핵 생물 또는 이것들의 바이러스 내에서 유전자 발현을 제어하는 것으로 공지되어 있는 기타 프로모터도 사용될 수 있다.
- [0141] 조직-특이적 식물 프로모터
- [0142] 본 발명은 조직-특이적 방식으로 발현될 수 있는 예를 들어, 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 조직-특이적 방식으로 발현시킬 수 있는 발현 카세트를 제공한다. 본 발명은 또한 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 조직-특이적 방식으로 발현시키는 식물 또는 종자를 제공한다. 상기 조직-특이성은 종자-특이적, 줄기-특이적, 잎-특이적, 뿌리-특이적, 열매-특이적일 수 있다.
- [0143] "식물"이란 용어는 전체 식물, 식물의 일부(예를 들어, 잎, 줄기, 꽃 또는 뿌리 등), 식물의 원형질체, 종자 및 식물의 세포와 이의 자손을 포함한다. 본 발명의 방법에 사용될 수 있는 식물의 군은 일반적으로 형질전환 기술에 순응할 수 있는 고등 식물군 예를 들어, 속씨 식물(외떡잎 식물 및 쌍떡잎 식물), 및 겉씨 식물과 같이 광범위하다. 이 식물군은 다양한 수준의 배수성 예를 들어, 배수체, 이배체, 반수체 및 반접합체 상태의 식물을 포함한다. 본원에 사용된 "트랜스제닉 식물"이란, 이중 핵산 서열 예를 들어, 본 발명의 핵산 및 다양한 재조합 구조물(예를 들어, 발현 카세트)이 삽입된 식물 또는 식물 세포를 포함한다.

- [0144] 하나의 측면에서, 구성적 프로모터 예를 들어, CaMV35S 프로모터는 식물의 일부 또는 종자의 특정 부분 또는 전체 식물에서 발현하는데에 사용될 수 있다. 예를 들어, 과발현의 경우, 식물의 일부 조직 또는 전체 조직 예를 들어, 재생산된 식물 내에서 핵산을 발현시킬 식물 프로모터 단편이 사용될 수 있다. 이러한 프로모터를 본원에서는 "구성적" 프로모터라고 부르며, 이 프로모터는 대부분의 환경 조건과 발달 단계 또는 세포 분화 중에 활성을 갖는다. 구성적 프로모터의 예로서는, 꽃 양배추 모자이크 바이러스(CaMV) 35S 전사 개시 부위, 아그로박테리움 튜메페이션스의 T-DNA로부터 유래하는 1'- 또는 2'-프로모터, 그리고 당 업자에게 공지된 다양한 식물 유전자로부터 유래하는 기타 전사 개시 부위를 포함한다. 이러한 유전자로서는 예를 들어, 아라비도시스(*Arabidopsis*)의 ACT11(Huang (1996) Plant Mol. Biol. 33:125-139); 아라비도시스의 Cat3(GenBank No. U43147, Zhong (1996) Mol. Gen. Genet. 251:196-203); 유채(*Brassica napus*)의 스테아로일-아실 운반 단백질 불포화 효소(desaturase)를 암호화하는 유전자(Genbank No. X74782, Solocombe (1994) Plant Physiol. 104:1167-1176); 옥수수의 Gpc1(GenBank No. X15596; Martinez (1989) J. Mol. Biol. 208:551-565); 옥수수의 Gpc2(GenBank No. U45855, Manjunath (1997) Plant Mol. Biol. 33:97-112); 미국 특허 제4,962,028호; 동 제 5,633,440호에 기술된 식물 프로모터를 포함한다.
- [0145] 본 발명은 예를 들어, 토마모바이러스(tobamovirus) 하위 게놈형 프로모터(subgenomic promoter)(Kumagai (1995) Proc. Natl. Acad. Sci. USA 92:1679-1683; 감염된 쌀 식물의 체관 세포에서만 복제하며, 강력한 체관-특이적 리포터 유전자 발현을 유도하는 프로모터를 가지는 쌀의 통그로(tungro) 간상 바이러스(RTBV); 유관속 세포, 잎의 잎살 세포 및 근단에서만 고도의 활성을 나타내는 카사바 베인 모자이크 바이러스(cassava vein mosaic virus)(CVMV) 프로모터(Verdaguer (1996) Plant Mol. Biol. 31:1129-1139)를 포함할 수 있는 바이러스로부터 유래하는 조직-특이적 또는 구성적 프로모터를 사용한다.
- [0146] 하나의 측면에서, 식물 프로모터는 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소-발현 핵산을 특정 조직, 기관 또는 세포 유형 내에서 발현시키거나(즉, 조직-특이적 프로모터), 또는 더욱 정확한 환경에 따른 제어 또는 발달에 따른 제어하에 있을 수 있거나, 또는 유도성 프로모터의 제어하에 있을 수 있다. 전사시킬 수 있는 환경적 조건의 예로서는, 혐기성 조건, 고온, 빛의 존재 여부, 또는 화학 물질/호르몬이 분무되었는지 여부를 포함한다. 예를 들어, 본 발명은 옥수수의 건조-유도성 프로모터(Busk (1997) 상동); 온도 하강, 건조 및 고농도 염에 따라서 작동이 유도되는 감자의 프로모터(Kirch (1997) Plant Mol. Biol. 33:897 909)를 포함한다.
- [0147] 하나의 측면에서, 조직-특이적 프로모터는 그 조직 내에서의 발달 단계의 주기 중 임의의 시기에만 전사를 촉진한다. 예를 들어, 아라비도시스 LEAFY 유전자 프로모터에 관하여 특성 규명한 문헌[Blazquez (1998) Plant Cell 10:791-800]을 참조하십시오. 또한, 에이.탈리아나(*A.thaliana*) 화아(floral meristem)의 동일성을 결정하는 유전자인 AP1의 프로모터 부위 내에 존재하는 보존적 서열 모티프를 인지하는 전사 인자인 SPL3에 관하여 기술하고 있는 문헌[Cardon (1997) Plant J 12:367-77]; 및 분열 조직 프로모터인 eIF4에 관하여 기술하고 있는 문헌[Mandel (1995) Plant Molecular Biology, Vol. 29, pp 995-1004]을 참조하십시오. 특정 조직의 생존 주기를 통하여 활성을 보유하는 조직 특이적 프로모터를 사용할 수 있다. 하나의 측면에서, 본 발명의 핵산은, 주로 면 섬유 세포에서만 활성을 보유하는 프로모터에 작동 가능하도록 결합되어 있다. 하나의 측면에서, 본 발명의 핵산은 면 섬유 세포 연장 단계 동안 주로 활성을 보유하는 프로모터에 작동 가능하도록 결합되어 있다[Rinehart (1996) 상동]. 핵산은 바람직하게는 면 섬유 세포(상동)에서 발현되는 Fb12A 유전자 프로모터에 작동 가능하도록 결합될 수 있다. 또한, 면 섬유-특이적 프로모터 및 트랜스제닉 면 식물의 구성 방법에 관하여 기술하고 있는 문헌[John (1997) Proc. Natl. Acad. Sci. USA 89:5769-5773; John의 다수, 미국 특허 제 5,608,148호 및 동 제5,602,321호]을 참조하십시오. 뿌리-특이적 프로모터는 또한 본 발명의 핵산을 발현하는데에도 사용될 수 있다. 뿌리-특이적 프로모터의 예로서는 알콜 탈수소효소 유전자로부터 유래하는 프로모터를 포함한다[DeLisle (1990) Int. Rev. Cytol. 123:39-60]. 본 발명의 핵산을 발현시키는데에 사용될 수 있는 기타 프로모터로서는 예를 들어, 난세포-특이적 프로모터, 배아-특이적 프로모터, 배젖-특이적 프로모터, 외피-특이적 프로모터, 종피-특이적 프로모터, 또는 이들의 조합; 잎-특이적 프로모터(예를 들어, 옥수수 내 잎-특이적 프로모터에 관하여 기술하고 있는 Busk (1997) Plant J. 11:1285 1295 참조); 아그로박테리움 리조제네스(*Agrobacterium rhizogenes*)로부터 유래하는 ORF13 프로모터(뿌리에서 고도의 활성을 나타냄)(예를 들어, Hansen (1997) 상동 참조); 옥수수 화분 특이적 프로모터(예를 들어, Guerrero (1990) Mol. Gen. Genet. 224:161 168); 과일 숙성기, 노화기 및 낙엽기 동안에 활성을 보유하는 토마토 프로모터 외에도, 그 정도는 낮지만, 낙화기 동안에 활성을 보유하는 프로모터도 사용될 수 있으며(예를 들어, Blume (1997) Plant J. 12:731 746 참조); 감자 SK2 유전자로부터 유래하는 암술-특이적 프로모터(예를 들어, Ficker (1997) Plant Mol. Biol. 35:425 431 참조); 트

랜스제닉 알팔파의 휴면중인 어린 싹과 꽃의 어린 눈의 정점에 있는 표피 조직에서 활성을 나타내므로 외래 유전자를 왕성하게 성장하고 있는 어린 싹 또는 섬유의 표피층으로 발현시킬 수 있는 유용한 도구가 되는, 완두 유래 Blec4 유전자; 난세포-특이적 BEL1 유전자(예를 들어, Reiser (1995) Cell 83:735-742, GenBank No. U39944); 및/또는, 식물 프로모터 부위는 분열 조직 및/또는 신속하게 분열중인 세포에서 높은 수준으로 전사를 진행시킬 수 있다고 기술하고 있는 미국 특허 제5,589,583호(Klee)에 개시된 프로모터를 포함한다.

[0148] 하나의 측면에서, 식물 호르몬 예를 들어, 옥신에 노출될 때 유도성이 되는 식물 프로모터는 본 발명의 핵산을 발현시키는데에 사용된다. 예를 들어, 본 발명은 대두(*Glycine max L.*)의 옥신-반응 요소 E1 프로모터 단편 (AuxREs)(Liu (1997) Plant Physiol. 115:397-407); 옥신-반응성 아라비도시스 GST6 프로모터(살리실산과 과산화수소에도 반응성임)(Chen (1996) Plant J. 10: 955-966); 담배로부터 유래하는 옥신-유도성 parC 프로모터 (Sakai (1996) 37:906-913); 식물 비오틴 반응 요소(Streit (1997) Mol. Plant Microbe Interact. 10:933-937); 그리고 스트레스 호르몬인 앱사이신산에 반응성인 프로모터(Sheen (1996) Science 274:1900-1902)을 사용할 수 있다.

[0149] 본 발명의 핵산은 또한 식물에 도포될 수 있는 화학 시약 예를 들어, 제초제 또는 항생제에 노출될 때에 유도될 수 있는 식물 프로모터에 작동 가능하도록 결합될 수도 있다. 예를 들어, 벤젠설포아미드 제초제의 안정제에 의해 활성화되는 옥수수 In2-2 프로모터가 사용될 수 있으며[De Veylder (1997) Plant Cell Physiol. 38:568-577]; 상이한 제초제의 안정제를 사용할 경우, 독특한 유전자 발현 패턴 예를 들어, 뿌리, 배수 조직 및 어린 싹의 정점에 존재하는 분열 조직의 발현을 유도하게 된다. 암호화 서열은 예를 들어, 테트라사이클린-유도성 프로모터 예를 들어, 아베나 사티바 엘.(*Avena sativa L.*)(귀리)의 아르기닌 탈탄소 효소 유전자(Masgrau (1997) Plant J. 11:465-473); 또는 살리실산 반응성 요소(Stange (1997) Plant J. 11:1315-1324)의 제어하에 있을 수 있다. 화학적으로 유도되는(예를 들어, 호르몬이나 살생물제로 유도되는) 프로모터 즉, 들에 있는 트랜스제닉 식물에 도포될 수 있는 화학 물질에 반응성인 프로모터를 사용할 경우, 본 발명의 폴리펩티드 발현은 식물 발달 단계 중 특정 단계에서 유도될 수 있다. 그러므로, 본 발명은 또한 본 발명의 폴리펩티드를 암호화하는 유도성 유전자를 함유하며, 곡물의 발달 단계중 임의의 단계에서 유도될 수 있는 트랜스제닉 식물[숙주 범위는 표적 식물 중 예를 들어, 옥수수, 쌀, 보리, 대두, 토마토, 밀, 감자 또는 기타 곡물에 한정됨]을 제공하기도 한다.

[0150] 당업자는 조직-특이적 식물 프로모터가 표적 조직 이외의 조직 내에서 작동 가능하도록 결합되어 있는 서열을 발현시킬 수 있음을 알게 될 것이다. 그러므로, 하나의 측면에서, 조직-특이적 프로모터는 표적 조직 또는 세포 유형 내에서 우선적으로 발현을 유도하되, 기타 조직에서도 어느 정도 발현을 유도할 수 있는 프로모터이다.

[0151] 본 발명의 핵산은 또한 화학 시약에 노출될 때 유도성을 가지는 식물 프로모터에 작동 가능하도록 결합될 수 있다. 이러한 시약들로서는 예를 들어, 트랜스제닉 식물에 사용 예를 들어, 분무될 수 있는 제초제, 합성 옥신 또는 항생제를 포함한다. 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소-생산 핵산의 유도성 발현을 통하여, 최적의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소가 발현되고/발현되거나, 이 효소의 활성을 보유하는 식물을 선별할 수 있도록 이 식물은 더욱 성장하게 될 것이다. 그러므로, 식물 일부의 발달은 제어될 수 있다. 이러한 방식으로, 본 발명은 식물과 식물 일부의 수확을 촉진하는 수단을 제공한다. 예를 들어, 다양한 구체예에서, 벤젠설포아미드 제초제의 안정제에 의해 활성화되는 옥수수 In2-2 프로모터가 사용되며[De Veylder (1997) Plant Cell Physiol. 38:568-577]; 상이한 제초제의 안정제를 사용하면, 독특한 유전자 발현 패턴 예를 들어, 뿌리, 배수 조직 및 어린 싹의 정점에 존재하는 분열 조직의 발현을 유도하게 된다. 본 발명의 암호화 서열은 예를 들어, 아베나 사티바 엘.(귀리)의 아르기닌 탈탄소 효소 유전자(Masgrau (1997) Plant J. 11:465-473); 또는 살리실산 반응성 요소(Stange (1997) Plant J. 11:1315-1324)를 포함하는 트랜스제닉 담배 식물에 대하여 기술된 바와 같이, 테트라사이클린-유도성 프로모터의 제어하에 있을 수 있다.

[0152] 몇몇 측면에서, 적당한 폴리펩티드 발현에는 암호화 부위의 3' 말단에 폴리아데닐화 부위를 필요로 할 수 있다. 폴리아데닐화 부위는 천연 유전자, 다양한 기타 식물(또는 동물 등) 유전자, 또는 아그로박테리아의 T-DNA의 유전자로부터 유래될 수 있다.

[0153] 발현 벡터 및 클로닝 비이클

[0154] 본 발명은 본 발명의 핵산 예를 들어, 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 암호화하는 서열을 포함하는 발현 벡터 및 클로닝 비이클을 제공한다. 본 발명의 발현 벡터 및 클로닝 비이클은 바이러스 입자, 바칼로바이러스, 파지, 플라스미드, 파지미드, 코스미드, 포스미드, 박테리아의 인공 염색체, 바이러스 DNA(예를 들어, 백시니아 바이

러스, 아데노바이러스, 천연두 바이러스, 가성 광견병 바이러스 및 SV40의 유도체), P1-계 인공 염색체, 효소 플라스미드, 효모 인공 염색체, 및 목적으로 하는 특정 숙주(예를 들어, 바실러스(*Bacillus*), 아스퍼질러스(*Aspergillus*) 및 효모)에 특이적인 기타 임의의 벡터를 포함할 수 있다. 본 발명의 벡터로서는 염색체, 비-염색체 및 합성 DNA 서열을 포함할 수 있다. 다수의 적당한 벡터에 관하여는 당 업자에게 공지되어 있으며, 또한 시판되고 있다. 대표적인 벡터로서는 다음의 것들을 포함한다: 박테리아 - pQE™ 벡터(Qiagen), pBLUESCRIPT™ 플라스미드, pNH 벡터(람다-ZAP 벡터)(Stratagene); ptrc99a, pKK223-3, pDR540, pRIT2T(Pharmacia); 진핵 생물 - pXT1, pSG5(Stratagene), pSVK3, pBPV, pMSG, pSVLSV40(Pharmacia). 그러나, 이들 벡터가 숙주 내에서 복제 가능하고 생존 가능한 이상 임의의 기타 플라스미드 또는 기타 벡터가 사용될 수 있다. 소수 또는 다수의 벡터가 본 발명에 사용될 수 있다. "플라스미드"는 시판되거나 제한 없이 공중이 이용 가능할 수 있으며, 또는 공지된 방법에 따라서 입수 가능한 플라스미드로 구성될 수도 있다. 본원에 기술된 것과 균등한 플라스미드는 당 업계에 공지되어 있으며, 이에 관하여는 당업자가 명백히 알고 있을 것이다.

[0155] 발현 벡터는 프로모터, 번역 개시를 위한 리보솜 결합 위치 및 전사 종결 인자를 포함할 수 있다. 벡터는 또한 증폭 발현에 적절한 서열을 포함할 수도 있다. 포유동물의 발현 벡터는 복제 기원, 임의의 필요한 리보솜 결합 위치, 폴리아데닐화 위치, 스플라이싱 공여 위치 및 수용 위치, 전사 종결 서열, 그리고 5' 말단에 축적하는 비 전사 서열을 포함할 수 있다. 몇몇 측면에서, SV40의 스플라이싱을 통하여 유래된 DNA 서열과 폴리아데닐화 위치는 필요한 비 전사 유전자 요소를 제공하는데에 사용될 수 있다.

[0156] 하나의 측면에서, 발현 벡터는 벡터를 함유하는 숙주 세포를 선별할 수 있도록 만드는 하나 이상의 선별 마커 유전자를 포함한다. 이러한 선별 마커로서는 디하이드로폴레이트 환원 효소를 암호화하는 유전자 또는 진핵 생물 세포 배양액에 네오마이신 내성을 부여하는 유전자, 이. 콜라이에 테트라사이클린 또는 암피실린 내성을 부여하는 유전자, 및 에스. 세레비지아에 TRP1 유전자를 포함한다. 프로모터 부위는 클로람페니콜 전이 효소(CAT) 벡터 또는 기타 선별 마커를 포함하는 벡터를 사용하여 임의의 바람직한 유전자로부터 선택될 수 있다.

[0157] 하나의 측면에서, 진핵 생물 세포 내에서 폴리펩티드 또는 이이 단편을 발현하기 위한 벡터는 발현 수준을 증가시키는 인핸서를 포함한다. 인핸서는 DNA의 시스-작용 요소로서, 그 길이가 약 10~약 300 bp일 수 있다. 상기 인핸서는 프로모터 상에서 작용하여 이것의 전사량을 증가시킬 수 있다. 대표적인 인핸서로서는 복제 기원의 뒷 부분에 존재하는 SV40 인핸서(100~270bp), 거대 세포 바이러스의 초기 프로모터 인핸서, 폴리오마 바이러스의 복제 기원 뒷부분에 존재하는 폴리오마 인핸서, 그리고 아데노바이러스 인핸서를 포함한다.

[0158] 핵산 서열은 다양한 방법에 의해 벡터에 삽입될 수 있다. 일반적으로, 이 서열은 삽입부와 벡터를 적당한 제한 엔도뉴클레아제로 절단함으로써 벡터 내 원하는 위치에 결합된다. 대안적으로, 삽입부와 벡터 둘 다에 존재하는 블런트 말단부(blunt end)를 결합할 수도 있다. 다수의 클로닝 기술이 당 업계에 공지되어 있다[예를 들어, Ausubel 및 Sambrook]. 이러한 방법과 기타 방법은 당 업자의 기술 범주에 속하는 것이다.

[0159] 벡터는 플라스미드, 바이러스 입자 또는 파지의 형태일 수 있다. 기타 벡터로서는 염색체, 비-염색체 및 합성 DNA 서열, SV40의 유도체; 박테리아 플라스미드, 파지 DNA, 바콜로바이러스, 효모의 플라스미드, 플라스미드와 파지 DNA의 조합체로부터 유래하는 벡터, 바이러스 DNA 예를 들어, 백시니아 바이러스, 아데노바이러스, 천연두 바이러스 및 가성 광견병 바이러스의 DNA를 포함한다. 원핵 생물 및 진핵 생물 숙주와 함께 사용되는 다양한 클로닝 및 발현 벡터에 관하여는 예를 들어, 문헌[Sambrook]에 기술되어 있다.

[0160] 사용될 수 있는 특정 박테리아 벡터로서는, 널리 공지된 클로닝 벡터인 pBR322(ATCC 37017), pKK223-3(Pharmacia Fine Chemicals, Uppsala, Sweden), GEM1(Promega Biotec, Madison, WI, USA) pQE70, pQE60, pQE-9(Qiagen), pD10, psiX174 pBLUESCRIPT II KS, pNH8A, pNH16a, pNH18A, pNH46A(Stratagene), ptrc99a, pKK223-3, pKK233-3, DR540, pRIT5(Pharmacia), pKK232-8 및 pCM7의 유전자 요소를 포함하는 시판 플라스미드를 포함한다. 특정 진핵 생물 벡터로서는 pSV2CAT, pOG44, pXT1, pSG(Stratagene), pSVK3, pBPV, pMSG 및 pSVL(Pharmacia)를 포함한다. 그러나, 기타 임의의 벡터는 그것이 숙주 세포 내에서 복제 가능하고 생존 가능한 한 사용될 수 있다.

[0161] 본 발명의 핵산은 발현 카세트, 벡터 또는 바이러스 내에서 발현될 수 있으며, 또한 식물 세포와 종자 내에서도 일시적으로나 안정적으로 발현될 수 있다. 하나의 대표적인 일시적 발현 시스템은 수퍼코일 DNA를 함유하는 에피솜 미니-염색체의 발현에 의해 핵 내에서 생산되는 에피솜 발현 시스템 예를 들어, 꽃양배추 모자이크 바이러스(CaMV)의 바이러스 RNA를 사용한다[예를 들어, Covey (1990) Proc. Natl. Acad. Sci. USA 87:1633-1637 참조]. 대안적으로, 암호화 서열 즉, 본 발명의 서열 전부 또는 이 서열에 속하는 단편들은 식물 숙주 세포 게놈 내에 삽입되어, 숙주 염색체 DNA의 내부 부분의 일부가 될 수 있다. 센스 또는 안티센스 전사체는 이러한 방

식으로 발현될 수 있다. 본 발명의 핵산으로부터 유래하는 서열(예를 들어, 프로모터나 암호화 부위)을 포함하는 벡터는 식물 세포나 종자에 선별 가능한 표현형을 부여하는 마커 유전자를 포함할 수 있다. 예를 들어, 상기 마커 유전자는 살생물제(biocide) 내성 예를 들어, 항생제 내성 예를 들어, 카나마이신, G418, 블레오마이신, 하이그로마이신에 대한 내성, 또는 제초제 내성 예를 들어, 클로로설프론 또는 바스타(Basta)에 대한 내성을 암호화할 수 있다.

[0162] 식물 내에서 핵산과 단백질을 발현시킬 수 있는 발현 벡터는 당 업계에 널리 공지되어 있으며, 예를 들어, 아그로박테리움 종, 감자 바이러스 X(예를 들어, Angell (1997) EMBO J. 16:3675-3684), 담배 모자이크 바이러스(예를 들어, Casper (1996) Gene 173:69-73), 토마토 부시 스톨트 바이러스(tomato bushy stunt virus)(예를 들어, Hillman (1989) Virology 169:42-50), 담배 식각바이러스(예를 들어, Dolja (1997) Virology 234:243-252), 콩 골든 모자이크 바이러스(예를 들어, Morinaga (1993) Microbiol Immunol. 37:471-476), 꽃 양배추 모자이크 바이러스(예를 들어, Cecchini (1997) Mol. Plant Microbe Interact. 10:1094-1101), 옥수수 Ac/Ds 전위 인자(예를 들어, Rubin (1997) Mol. Cell. Biol. 17:6294-6302; Kunze (1996) Curr. Top. Microbiol. Immunol. 204:161-194), 그리고 옥수수 억제제-돌연변이 인자(Spm) 전위 인자(예를 들어, Schlappi (1996) Plant Mol. Biol. 32:717-725)로부터 유래하는 벡터; 및 이의 유도체를 포함할 수 있다.

[0163] 하나의 측면에서, 발현 벡터는 2개의 유기체 내에서 유지될 수 있도록 만들어주는 2개의 복제 시스템 예를 들어, 포유동물 또는 곤충 세포에 있어서는 발현용 복제 시스템, 그리고 원핵 생물 숙주에 있어서는 클로닝 및 증폭용 복제 시스템을 가질 수 있다. 뿐만 아니라, 통합용 발현 벡터에 있어서, 발현 벡터는 숙주 세포 계통과 상동성인 하나 이상의 서열을 함유할 수 있다. 이는 발현 구조물에 축적하는 2개의 상동성 서열을 함유할 수 있다. 통합 벡터는 벡터 내에 존재하는 통합용으로 적당한 상동성 서열을 선택함으로써 숙주 세포 내에 존재하는 특정 위치에 유도될 수 있다. 통합 벡터 구조물은 당 업계에 널리 공지되어 있다.

[0164] 본 발명의 발현 벡터는 또한 형질전환된 박테리아 군주의 선별을 가능하게 만들어주는 선별 마커 유전자 예를 들어, 박테리아가 약물 예를 들어, 암피실린, 클로람페니콜, 에리트로마이신, 카나마이신, 네오마이신 및 테트라사이클린에 대해 내성을 가지도록 만들어주는 유전자를 포함할 수도 있다. 선별 마커는 또한 생합성 유전자 예를 들어, 히스티딘, 트립토판 및 루신 생합성 경로에 관여하는 유전자를 포함할 수도 있다.

[0165] 발현 벡터 내에 존재하는 DNA 서열은 RNA 합성을 유도하도록 적당한 발현 제어 서열(들)(프로모터)에 작동 가능하도록 결합되어 있다. 특별히 명명된 박테리아 프로모터로서는 lacI, lacZ, T3, T7, gpt, 람다 P_R, P_L 및 trp 을 포함한다. 진핵 생물의 프로모터로서는 CMV 즉시 초기 프로모터, HSV 티미딘 키나제 프로모터, 초기 및 후기 SV40 프로모터, 레트로바이러스의 LTR 프로모터 및 마우스의 메탈로티오네인-I 프로모터를 포함한다. 적당한 벡터 및 프로모터의 선별법은 당 업자의 기술 수준에서 잘 알려져 있는 것이다. 발현 벡터는 또한 번역 개시를 담당하는 리보솜 결합 위치와 전사 종결 인자를 포함하기도 한다. 상기 벡터는 또한 증폭 발현에 적당한 서열을 포함할 수도 있다. 프로모터 부위는 클로람페니콜 전이효소(CAT) 벡터 또는 선별 마커를 포함하는 기타 벡터를 사용하여 임의의 바람직한 유전자로부터 선별될 수 있다. 뿐만 아니라, 하나의 측면에서, 발현 벡터는 형질전환된 숙주 세포 선별을 위한 표현 특성 예를 들어, 진핵 생물 세포 배양액에 있어서는 디하이드로폴레이트 환원 효소 또는 네오마이신 내성, 또는 이. 콜라이의 경우에는 테트라사이클린이나 암피실린 내성을 제공하는 하나 이상의 선별 마커 유전자를 함유한다.

[0166] 포유동물 발현 벡터는 또한 복제 기원, 임의의 필수 리보솜 결합 위치, 폴리아데닐화 위치, 스플라이싱 공여 및 수용 위치, 전사 종결 서열 및 5' 말단-축적 비 전사 서열을 포함한다. 몇몇 측면에서, SV40 스플라이싱 위치 및 폴리아데닐화 위치로부터 유래된 DNA 서열은 필요한 비 전사 유전 요소를 제공하는데에 사용될 수 있다.

[0167] 진핵 생물 세포 내에서 폴리펩티드 또는 이의 단편을 발현시키기 위한 벡터는 또한 발현 수준을 증가시키는 인핸서를 함유하기도 한다. 인핸서는 DNA의 시스-작용성 요소로서, 일반적으로 그 길이가 약 10~약 300 bp에 달하며, DNA의 전사를 증가시키는 프로모터에서 작용을 한다. 그 예로서는 복제 기원의 뒷부분에 위치하는 SV40 인핸서(100~270bp), 거대 세포 바이러스의 초기 프로모터 인핸서, 복제 기원의 뒷부분에 위치하는 폴리오마 바이러스 인핸서 및 아데노바이러스 인핸서를 포함한다.

[0168] 뿐만 아니라, 발현 벡터는 벡터를 함유하는 숙주 세포를 선별할 수 있도록 만들어주는 하나 이상의 선별 마커 유전자를 함유할 수 있다. 이러한 선별 마커로서는 디하이드로폴레이트 환원 효소를 암호화하는 유전자 또는 진핵 생물 세포 배양액에 네오마이신 내성을 부여해주는 유전자, 이. 콜라이 및 에스. 세레비지아에 TRP1 유전자에 테트라사이클린 또는 암피실린 내성을 부여해주는 유전자를 포함한다.

- [0169] 몇몇 측면에서, 본 발명의 폴리펩티드 중 하나를 암호화하는 핵산, 또는 약 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100 또는 150개 이상의 연속적 아미노산을 포함하는 이의 단편은 적당한 단계에서 번역된 폴리펩티드 또는 이의 단편의 분비를 유도할 수 있는 리더 서열과 함께 조립된다. 하나의 측면에서, 이 핵산은 본 발명의 폴리펩티드 중 하나, 또는 약 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100 또는 150개 이상의 연속적 아미노산을 포함하는 이의 단편이 이중 펩티드 또는 폴리펩티드 예를 들어, 원하는 특성 예를 들어, 안정성의 증가 및 정제의 간편함을 부여하는 N-말단 동정용 펩티드에 융합되어 있는 융합 폴리펩티드를 암호화할 수 있다.
- [0170] 적당한 DNA 서열은 다양한 방법에 의해 벡터 내에 삽입될 수 있다. 일반적으로, DNA 서열은 삽입부와 벡터를 적당한 제한 엔도뉴클레아제로 분해함으로써, 벡터내 원하는 위치에 결합된다. 대안적으로, 상기 삽입부 및 벡터 둘 다에 존재하는 블런트 말단부를 결합할 수도 있다. 다양한 클로닝 기술에 관하여는 문헌[Ausubel 외 다수, Current Protocols in Molecular Biology, John Wiley 503 Sons, Inc. 1997 and Sambrook 외 다수, Molecular Cloning: A Laboratory Manual 2nd Ed., Cold Spring Harbor Laboratory Press (1989)]에 개시되어 있다. 이러한 방법들 및 기타 방법들은 당 업자의 기술 범주 내에 속하는 것이다.
- [0171] 벡터는 예를 들어, 플라스미드, 바이러스 입자 또는 파지의 형태일 수 있다. 기타 벡터로서는 염색체, 비-염색체 및 합성 DNA 서열, SV40의 유도체; 박테리아 플라스미드, 파지 DNA, 마칼로바이러스, 효모 플라스미드, 플라스미드와 파지 DNA의 조합체로부터 유래하는 벡터, 바이러스 DNA 예를 들어, 백시니아 바이러스, 아데노바이러스, 천연두 바이러스 및 가성 광견병 바이러스 DNA를 포함한다. 원핵 생물 및 진핵 생물 숙주를 사용하는 다양한 클로닝 및 발현 벡터에 관하여는 문헌[Sambrook 외 다수, Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor, N.Y., (1989)]에 개시되어 있다.
- [0172] 숙주 세포 및 형질전환된 세포
- [0173] 본 발명은 또한 본 발명의 핵산 서열 예를 들어, 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 암호화하는 서열 또는 본 발명의 벡터를 포함하는 형질전환 세포를 제공한다. 숙주 세포는 당 업자에게 널리 알려진 숙주 세포 중 임의의 것 예를 들어, 원핵 생물 세포, 진핵 생물 세포 예를 들어, 박테리아 세포, 진균 세포, 효모 세포, 포유동물 세포, 곤충 세포 또는 식물 세포일 수 있다. 대표적인 박테리아 세포로서는 스트렙토마이세스, 스태필로코커스(*Staphylococcus*) 또는 바실러스 중 중 임의의 중 또는 대표적인 종인 이. 콜라이, 바실러스 서브틸리스(*Bacillus subtilis*), 바실러스 세레우스(*Bacillus cereus*), 살모넬라 타이피무리움(*Salmonella typhimurium*)을 포함한다. 대표적인 곤충 세포로서는 스포도프테라(*Spodoptera*) 또는 드로소필라(*Drosophila*) 중 중 임의의 중 예를 들어, 드로소필라 S2 및 스포도프테라 Sf9을 포함한다. 대표적인 동물 세포로서는 CHO, COS 또는 바우어즈(Bowes) 흑색종 또는 임의의 마우스나 인간 세포주를 포함한다. 적당한 숙주 세포를 선택함은 당 업자의 능력 범주 내에 있다. 다양한 고등 식물 종을 형질전환하는 기술에 관하여는 당 업계에 널리 공지되어 있으며, 기술 논문 및 과학 논문에서 살펴볼 수 있다. 예를 들어, 문헌[Weising (1988) Ann. Rev. Genet. 22:421-477; 미국 특허 제5,750,870호]을 참조하십시오.
- [0174] 벡터는 다양한 기술 중 임의의 기술 예를 들어, 형질전환 기술, 형질 감염 기술, 형질 도입법, 바이러스 감염, 유전자 총 또는 Ti-매개 유전자 운반 기술을 사용하여 숙주 세포 내에 도입될 수 있다. 구체적인 방법으로서 인산칼슘 형질 감염법, DEAE-텍스트란 매개 형질 감염법, 리포펙션(lipofection) 또는 전기 천공법을 포함한다 [Davis, L., Dibner, M., Battey, I., Basic Methods in Molecular Biology, (1986)].
- [0175] 하나의 측면에서, 본 발명의 핵산 또는 벡터는 스크리닝을 위하여 세포에 도입되며, 그 결과 이 핵산은 핵산의 추후 발현에 적당한 방식으로 세포에 들어가게 된다. 도입 방법은 표적화된 세포 유형에 의해 대부분 결정된다. 대표적인 방법로서는 $CaPO_4$ 침전법, 리포솜 융합법, 리포펙션(예를 들어, LIPOFECTIN™), 전기 천공법 및 바이러스 감염법 등을 포함한다. 후보 핵산은 숙주 세포의 게놈에 안정적으로 통합화될 수 있거나[예를 들어, 레트로바이러스 도입법에 의한], 또는 세포질 내에 일시적으로 또는 안정적으로 존재할 수 있다[즉, 통상의 플라스미드, 표준적인 조절 서열 및 선별 마커 등을 이용함]. 다수의 약학적으로 중요한 스크리닝에는 인간 또는 진형적인 포유동물 세포 표적이 필요하므로, 이러한 표적을 형질 감염시킬 수 있는 레트로바이러스 벡터가 사용될 수 있다.
- [0176] 적당한 경우, 조작된 숙주 세포는 프로모터 활성화, 형질전환체 선별 또는 본 발명의 유전자를 증폭하는데에 적당하게 개질된 통상의 영양 배지 중에서 배양될 수 있다. 적당한 숙주 균주를 형질전환시키고, 이 숙주 균주를 적당한 세포 밀도가 되도록 생육시킨 후, 선별된 프로모터는 적당한 수단(예를 들어, 온도 변화 또는 화학적 유

도)에 의해 유도될 수 있으며, 세포는 이 세포가 원하는 폴리펩티드 또는 이의 단편을 생산할 수 있게 되는 기간 동안 추가로 배양될 수 있다.

[0177] 세포는 원심 분리에 의해 수집될 수 있으며, 또한 물리적 또는 화학적 수단에 의해 파괴될 수 있고, 결과로 생성된 미정제 추출물은 추가의 정제를 위해 보관하게 된다. 단백질 발현에 사용되는 미생물 세포는 임의의 편리한 방법 예를 들어, 동결-해동 순환법, 초음파 처리법, 기계적 파괴법 또는 세포 용해 제제를 사용하는 방법에 의해 파괴될 수 있다. 이러한 방법은 당 업자에게 널리 공지되어 있다. 발현된 폴리펩티드 또는 이의 단편은 예를 들어, 황산암모늄 침전법이나 에탄올 침전법, 산 추출법, 음이온 또는 양이온 교환 크로마토그래피, 포스포셀룰로스 크로마토그래피, 소수성 상호 작용 크로마토그래피, 친화성 크로마토그래피, 하이드록시아파타이트 크로마토그래피 및 렉틴 크로마토그래피와 같은 방법에 의해 재조합 세포 배양액으로부터 회수 및 정제될 수 있다. 단백질 리폴딩 단계는, 필요에 따라서, 폴리펩티드의 원심 분리를 종결하였을 때에 진행될 수 있다. 원한다면, 고 성능 액체 크로마토그래피(HPLC)는 최종 정제 단계에 사용될 수 있다.

[0178] 숙주 세포 내에 존재하는 구조물은 재조합 서열에 의해 암호화되는 유전자 산물을 생산하기 위한 통상의 방식으로 사용될 수 있다. 재조합체 생산 과정에 사용되는 숙주에 따라서, 박터를 함유하는 숙주 세포에 의해 생산된 폴리펩티드는 글리코실화될 수 있거나, 또는 글리코실화되지 않을 수도 있다. 본 발명의 폴리펩티드는 또한 개시 메티오닌 아미노산 잔기를 포함할 수 있거나 또는 포함하지 않을 수도 있다.

[0179] 무 세포 번역 시스템(cell-free translation system)은 또한 본 발명의 폴리펩티드를 생산하는데 사용될 수도 있다. 무 세포 번역 시스템은 폴리펩티드를 암호화하는 핵산 또는 이의 단편에 작동 가능하도록 결합되어 있는 프로모터를 포함하는 DNA 구조물로부터 전사된 mRNA를 사용할 수 있다. 몇몇 측면에서, DNA 구조물은 선형화된 후, 시험관 내 전사 반응을 수행할 수 있다. 이후, 전사된 mRNA는 적당한 무 세포 번역 추출물 예를 들어, 토끼 적혈구 추출물과 함께 항온 처리되어, 목적으로 하는 폴리펩티드 또는 이의 단편을 생산하게 된다.

[0180] 발현 박터는 하나 이상의 선별 마커를 함유하여, 형질전환된 숙주 세포의 선별용인 표현형적 특성 예를 들어, 디하이드로폴레이트 환원 효소 또는 네오마이신 내성(진핵 생물 세포 배양액의 경우), 또는 테트라사이클린 또는 암피실린 내성(이. 콜라이의 경우)을 제공할 수 있다.

[0181] 목적으로 하는 폴리뉴클레오티드 예를 들어, 본 발명의 핵산을 함유하는 숙주 세포는 프로모터를 활성화하거나, 형질전환체를 선별하거나 또는 유전자를 증폭시키기에 적당하게 개질된 통상의 영양 배지 중에서 배양될 수 있다. 배양 조건 예를 들어, 온도 및 pH 등은 발현을 위해 선택된 숙주 세포에 대해 이미 적용되었던 것으로서, 당 업자에게 명백할 것이다. 이후, 특정의 효소 활성을 보유하는 것으로 확인된 클론은 서열 결정되어, 강화된 활성을 보유하는 효소를 암호화하는 폴리뉴클레오티드 서열을 동정할 수 있다.

[0182] 본 발명은 재조합 셀룰라제, 예를 들어 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 세포 내에서 과발현시키는 방법을 제공하는데, 이 방법은 본 발명의 핵산, 예를 들어 약 100개 이상의 잔기로 이루어진 부위에 걸쳐서 본 발명의 대표적인 서열과 적어도 약 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 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% 이상 상동성인 핵산 서열을 포함하는 핵산, 또는 엄중한 조건 하에서 본 발명의 핵산 서열에 혼성화하는 핵산을 포함하는 박터를 발현시키는 단계를 포함하며, 여기서, 상기 서열 동일성은 서열 비교 알고리즘을 이용하는 분석법 또는 시각적 조사에 의해 측정된다. 과발현은 임의의 수단, 예를 들어, 고효성 프로모터 또는 이시스트론성 박터를 사용하거나, 또는 박터의 유전자 증폭법을 사용하여 이루어질 수 있다.

[0183] 본 발명의 핵산은 임의의 시험관 내 또는 생체내 발현 시스템 내에서 발현 또는 과발현될 수 있다. 임의의 세포 배양 시스템은 재조합 단백질, 예를 들어 박테리아, 곤충, 효모, 진균 또는 포유동물 배양액을 발현시키거나 또는 과발현시키는 데 사용될 수 있다. 과발현은 프로모터, 인핸서, 박터[예를 들어, 레플리콘 박터, 이시스트론성 박터(예를 들어, Gurtu (1996) Biochem. Biophys. Res. Commun. 229:295-8) 사용], 배지 및 배양 시스템 등을 적당히 선택함으로써 이루어질 수 있다. 하나의 측면에서, 선별 마커, 예를 들어 글루타민 합성 효소(예를 들어, Sanders (1987) Dev. Biol. Stand. 66:55-63)를 이용한 세포 시스템 내 유전자 증폭법은 본 발명의 폴리펩티드를 과발현시키는 데 사용된다. 숙주 세포는 당 업자에게 널리 알려진 숙주 세포 중 임의의 것, 예를 들어 원핵 생물 세포, 진핵 생물 세포, 포유동물 세포, 곤충 세포 또는 식물 세포일 수 있다. 적당한 숙주를 선택하는 것은 당 업자의 기술 범주 내에 속한다.

[0184] 박터는 다양한 기술 예를 들어, 형질전환 기술, 형질 감염 기술, 형질 도입 기술, 바이러스 감염 기술, 유전자

충 또는 Ti-매개 유전자 운반 기술 중 임의의 기술을 이용하여 숙주 세포내에 도입될 수 있다. 구체적인 방법으로는 인산칼슘 형질 감염법, DEAE-덱스트란 매개 형질 감염법, 리포펙션 또는 전기 천공법을 포함한다[Davis, L., Dibner, M., Battey, I., Basic Methods in Molecular Biology, (1986)].

[0185] 적당한 경우, 조작된 숙주 세포는 프로모터를 활성화하거나, 형질전환체를 선별하거나 또는 본 발명의 유전자를 증폭시키기에 적당하게 개질된 통상의 영양 배지 중에서 배양될 수 있다. 적당한 숙주 균주를 형질전환시키고, 이 숙주 균주를 적당한 세포 밀도가 될 때까지 생육시킨 후, 선별된 프로모터는 적당한 수단(예를 들어, 온도 변화 또는 화학적 유도)에 의해 유도될 수 있으며, 세포는 이 세포가 원하는 폴리펩티드 또는 이의 단편을 생산할 수 있게 되는 기간 동안 추가로 배양될 수 있다.

[0186] 세포는 원심 분리에 의해 수집될 수 있으며, 또한 물리적 또는 화학적 수단에 의해 파괴될 수 있고, 결과로 생성된 미정제 추출물은 추가의 정제를 위해 보관하게 된다. 단백질 발현에 사용되는 미생물 세포는 임의의 편리한 방법 예를 들어, 동결-해동 순환법, 초음파 처리법, 기계적 파괴법 또는 세포 용해 제제를 사용하는 방법에 의해 파괴될 수 있다. 이러한 방법은 당 업자에게 널리 공지되어 있다. 발현된 폴리펩티드 또는 이의 단편은 예를 들어, 황산암모늄 침전법이나 에탄올 침전법, 산 추출법, 음이온 또는 양이온 교환 크로마토그래피, 포스포셀룰로스 크로마토그래피, 소수성 상호 작용 크로마토그래피, 친화성 크로마토그래피, 하이드록실아파타이트 크로마토그래피 및 렉틴 크로마토그래피와 같은 방법에 의해 재조합 세포 배양액으로부터 회수 및 정제될 수 있다. 단백질 리폴딩 단계는, 필요에 따라서, 폴리펩티드의 원심 분리를 종결하였을 때에 진행될 수 있다. 원한다면, 고 성능 액체 크로마토그래피(HPLC)는 최종 정제 단계에 사용될 수 있다.

[0187] 재조합 단백질을 발현시키는데에 다양한 포유동물 세포 배양 시스템이 사용될 수 있다. 포유동물 발현 시스템의 예로서는 원숭이 신장 섬유아세포의 COS-7 세포주(Gluzman, Cell, 23:175, 1981)와, 적합한 벡터로부터 단백질을 발현할 수 있는 기타 세포주 예를 들어, C127, 3T3, CHO, HeLa 및 BHK 세포주를 포함한다.

[0188] 숙주 세포내에 존재하는 구조물은 재조합 서열에 의해 암호화되는 유전자 산물을 생산하기 위해 통상의 방식으로 사용될 수 있다. 재조합체 생산 방법에서 사용되는 숙주에 따라서, 벡터를 함유하는 숙주 세포에 의해 생산되는 폴리펩티드는 글리코실화될 수 있거나 또는 글리코실화될 수 없다. 본 발명의 폴리펩티드는 또한 개시 메티오닌 아미노산 잔기를 포함할 수 있거나 또는 포함하지 않을 수도 있다.

[0189] 대안적으로, 본 발명의 폴리펩티드, 또는 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100 또는 150개 이상의 연속적 아미노산을 포함하는 이의 단편은 예를 들어, 이하에 기술되어 있는 바와 같이, 통상적인 펩티드 합성기에 의해 합성되어 생산될 수 있다. 다른 측면에서, 폴리펩티드 단편 또는 일부는 펩티드 합성에 의해 상응하는 전체 길이 폴리펩티드를 생산하는데에 사용될 수 있으므로; 이 단편은 전체 길이 폴리펩티드를 생산하기 위한 중간체로서 사용될 수 있다.

[0190] 무 세포 번역 시스템은 또한 본 발명의 폴리펩티드 또는 이의 단편을 암호화하는 핵산에 작동 가능하도록 결합되어 있는 프로모터를 포함하는 DNA 구조물로부터 전사된 mRNA를 이용하여, 본 발명의 폴리펩티드 중 하나, 또는 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100 또는 150개 이상의 연속적 아미노산을 포함하는 이의 단편을 생산하는데에 사용될 수도 있다. 몇몇 측면에서, 상기 DNA 구조물은 선형화된 후 시험관 내 전사 반응을 수행할 수 있다. 이후, 전사된 mRNA는 적당한 무 세포 번역 추출물 예를 들어, 토끼 적혈구 추출물과 함께 항온 처리되어, 목적으로 하는 폴리펩티드 또는 이의 단편을 생산하게 된다.

[0191] 핵산의 증폭

[0192] 본 발명을 실시함에 있어서, 본 발명의 핵산과 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 암호화하는 핵산, 또는 본 발명의 변형된 핵산은 증폭 방법 예를 들어, PCR에 의해 재생산될 수 있다. 증폭법은 또한 본 발명의 핵산을 클로닝하거나 또는 증폭하는데에 사용될 수 있다. 그러므로, 본 발명은 본 발명의 핵산을 증폭시키기 위한 증폭용 프라이머 서열의 쌍을 제공한다. 당 업자는 이러한 서열 중 임의의 부분 또는 전체 길이 서열에 대한 증폭용 프라이머 서열 쌍을 디자인할 수 있다.

[0193] 하나의 측면에서, 본 발명은 본 발명의 증폭용 프라이머 쌍 예를 들어, 본 발명의 핵산 중 약 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 또는 25 개 이상의 잔기[앞(5' 말단)에서부터 시작]에 의해 제시되는 프라이머 쌍과, 이의 상보성 가닥 중 약 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 또는 25 개 이상의 잔기[앞(5' 말단)에서부터 시작]에 의해 제시되는 프라이머 쌍을 증폭시킴으로써 증폭되는 핵산을 제공한다. 본 발명은 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성을

보유하는 폴리펩티드를 암호화하는 핵산을 증폭시키기 위한 증폭용 프라이머 서열 쌍을 제공하는데, 여기서, 상기 프라이머 쌍은 본 발명의 서열을 포함하는 핵산, 또는 이의 단편이나 부분 서열을 증폭시킬 수 있다. 증폭용 프라이머 서열 쌍 중 하나 또는 이의 각 일원은 적어도 약 10~50개의 연속된 염기 서열, 또는 적어도 약 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 또는 25개 이상의 연속된 염기 서열을 포함하는 올리고뉴클레오티드를 포함할 수 있다. 본 발명은 증폭용 프라이머 쌍을 제공하는데, 여기서, 상기 프라이머 쌍은 본 발명의 핵산 잔기 중 적어도 약 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 또는 25개 이상의 잔기[앞(5' 말단)에서부터 시작]에 의해 제시되는 서열을 가지는 제1 일원, 그리고 상기 제1 일원의 상보성 가닥 중 적어도 약 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 또는 25개 이상의 잔기[앞(5' 말단)에서부터 시작]에 의해 제시되는 서열을 가지는 제2 일원을 포함한다.

[0194] 본 발명은 본 발명의 증폭용 프라이머 쌍을 이용하는 증폭 반응 예를 들어, 중합 효소 연쇄 반응(PCR)에 의해 생산되는, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제 및/또는 베타-글루코시다제 효소를 제공한다. 본 발명은 본 발명의 증폭용 프라이머 쌍을 이용하는 증폭 반응 예를 들어, PCR에 의해 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제 및/또는 베타-글루코시다제 효소를 생산하는 방법을 제공한다. 하나의 측면에서, 증폭용 프라이머 쌍은 라이브러리 예를 들어, 유전자 라이브러리 예를 들어, 환경 라이브러리로부터 핵산을 증폭시킨다.

[0195] 증폭 반응은 또한 시료 내 핵산의 양(예를 들어, 세포 시료 중의 메세지 양)을 정량하거나, 핵산을 표지화하거나(예를 들어, 이 핵산을 어레이 또는 블롯팅에 사용하기 위함), 핵산을 검출하거나 또는 시료 중 특정 핵산의 양을 정량하는데에 사용될 수도 있다. 본 발명의 하나의 측면에서, 세포 또는 cDNA 라이브러리로부터 유래하는 메세지는 증폭된다.

[0196] 당 업자는 적당한 올리고뉴클레오티드 증폭용 프라이머를 선택하여 디자인할 수 있다. 증폭 방법은 또한 당 업계에 널리 공지되어 있으며, 그 예로서는 중합 효소 연쇄 반응, PCR[예를 들어, PCR PROTOCOLS, A GUIDE TO METHODS AND APPLICATIONS, ed. Innis, Academic Press, N. Y. (1990) and PCR STRATEGIES (1995), ed. Innis, Academic Press, Inc., N.Y.], 리가제 연쇄 반응(LCR)[예를 들어, Wu (1989) Genomics 4:560; Landegren (1988) Science 241:1077; Barringer (1990) Gene 89: 117]; 전사 증폭법[예를 들어, Kwok (1989) Proc. Natl. Acad. Sci. USA 86:1173]; 및 자기-지속적 서열 복제(self-sustained sequence replication)[예를 들어, Guatelli (1990) Proc. Natl. Acad. Sci. USA 87:1874]; Q 베타 레플리카제 증폭법[예를 들어, Smith (1997) J. Clin. Microbiol. 35:1477-1491], 자동화 Q-베타 레플리카제 증폭 검정법[예를 들어, Burg (1996) Mol. Cell. Probes 10:257-271] 및 기타 RNA 중합 효소 매개 기술[예를 들어, NASBA, Cangene, Mississauga, Ontario; 및 Berger (1987) Methods Enzymol. 152:307-316; Sambrook; Ausubel; 미국 특허 제4,683,195호 및 동 제4,683,202호; Sooknanan (1995) Biotechnology 13:563-564]을 포함한다.

[0197] 핵산 및 폴리펩티드의 서열 동일성 측정

[0198] 본 발명은 약 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 또는 1550개 이상의 잔기로 이루어진 부위에 걸쳐서, 본 발명의 대표적인 핵산과 적어도 약 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 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% 이상 상동성이거나 완벽한(100%) 서열 동일성을 갖는 서열을 포함하는 핵산을 제공한다[이하, 표 1a~1c, 2 및 3과, 실시예 1 및 4, 그리고 서열 목록 참조]. 본 발명은 본 발명의 대표적인 폴리펩티드와 적어도 약 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 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% 이상 상동성이거나 완벽한(100%) 서열 동일성을 갖는 서열을 포함하는 폴리펩티드를 제공한다[이하, 표 1a~1c, 2 및 3과, 실시예 1 및 4, 그리고 서열 목록 참조]. 서열 동일성(상동성)의 정도는 임의의 컴퓨터 프로그램 및 관련 매개 변수 예를 들어, 본원에 개시된 매개 변수 예를 들어, BLAST 2.2.2. 또는 FASTA 버전 3.0t78(디폴트 매개 변수 포함)을 사용하여 측정될 수 있다.

[0199] 본 발명의 핵산 서열은 본 발명의 대표적인 서열 중 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400 또는 500개 이상의 연속적 뉴클레오티드 및 이와 실질적으로 동일한 서열을 포함할 수 있다. 본 발명의 핵산 서열과 상동성인 서열 및 이의 단편은, 이들 서열과 약 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 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% 이상 상동성인 서열일 수 있다. 상동성(서열 동일성)은 FASTA 버전 3.0t78(디폴트 매개 변수 포함)을 비롯하여, 본원에 개시된 컴퓨터 프로그램 및 매개 변수 중 임의의 것을 사용하여 측정될 수 있다. 상동성 서열은 또한 본 발명의 핵산 서열 중 우리단이 티미딘으로 치환된 RNA 서열을 포함한다. 상동성 서열은 본원에 개시된 방법들 중 임의의 방법을 사용하여 얻어질 수 있거나, 또는 서열 결정 오류를 수정함으로써 얻어질 수 있다. 본 발명의 핵산 서열은 통상의 단일 문자 방식으로 표시될 수 있거나[Stryer, Lubert. Biochemistry, 3rd Ed., W. H Freeman & Co., New York], 또는 서열 내 뉴클레오타이드의 상동성을 기록하는 기타 임의의 방식으로 표시될 수 있다.

[0200]

다양한 측면에서, 본원에서 기술하고 있는 서열 비교 프로그램은 본 발명의 측면 즉, 핵산 또는 폴리펩티드 서열이 본 발명의 범위 내에 포함되는지 여부를 결정하는데에 사용된다. 그러나, 단백질 및/또는 핵산 서열의 동일성(상동성)은 당 업계에 공지된 임의의 서열 비교 알고리즘이나 프로그램을 이용하여 평가할 수 있다. 이러한 알고리즘 및 프로그램으로는 TBLASTN, BLASTP, FASTA, TFASTA 및 CLUSTALW[예를 들어, Pearson and Lipman, Proc. Natl. Acad. Sci. USA 85(8):2444-2448, 1988; Altschul외 다수, J. Mol. Biol. 215(3):403-410, 1990; Thompson Nucleic Acids Res. 22(2):4673-4680, 1994; Higgins외 다수, Methods Enzymol. 266:383-402, 1996; Altschul외 다수, J. Mol. Biol. 215(3):403-410, 1990; Altschul외 다수, Nature Genetics 3:266-272, 1993]를 포함하나, 이에 한정되는 것은 아니다.

[0201]

하나의 측면에서, 상동성 또는 동일성은 서열 분석용 소프트웨어를 사용하여 측정된다[예를 들어, Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705]. 이러한 소프트웨어는 다양한 형태의 결실, 치환 및 기타 변형에 상동성 등급을 부여함으로써 유사한 서열들을 매칭시킨다. 하나의 측면에서, 2개 이상의 핵산 또는 폴리펩티드 서열을 말할 때 "상동성" 및 "동일성"이란 용어는, 다수의 서열 비교 알고리즘 또는 수동적 정렬 및 시각적 조사에 의해 측정되는 지정된 부위 또는 비교 윈도우에 대하여 비교하고 최대로 상응시키기 위해 정렬하였을 때, 아미노산 잔기가 동일하거나 특정 비율의 아미노산 잔기를 가지는 2개 이상의 서열 또는 부분 서열, 또는 아미노산 잔기가 동일한 뉴클레오타이드를 의미하는 것이다. 하나의 측면에서, 서열 비교에 있어서, 하나의 서열은 시험 서열의 비교 기준인 참조 서열로서의 역할을 갖는다. 서열 비교 알고리즘을 사용할 때, 시험 및 참조 서열은 컴퓨터에 입력되고, 부분 서열의 좌표가 지정되며, 필요에 따라서 서열 알고리즘 프로그램 매개 변수가 지정되기도 한다. 디폴트 프로그램 매개 변수가 사용될 수 있으며, 또는 대안적 매개 변수가 지정될 수도 있다. 이후, 서열 비교 알고리즘은 프로그램 매개 변수에 입각하여, 참조 서열에 대한 시험 서열의 서열 상동성%를 계산한다.

[0202]

본원에 사용된 "비교 윈도우"는, 2개의 서열들이 최적으로 정렬된 이후, 서열이 인접하는 위치와 동일한 수의 위치로 이루어진 참조 서열과 비교될 수 있는 인접 위치 중 임의의 하나로 이루어진 분절에 대한 참조 기준을 포함하는데, 여기서, 상기 인접 위치는 20~600개, 일반적으로 약 50~약 200개, 더욱 일반적으로는 약 100~약 150개로 이루어진 군으로부터 선택된다. 비교를 위하여 서열을 정렬하는 방법에 관하여는 당 업계에 널리 공지되어 있다. 예를 들어, 국소 상동성 알고리즘(Smith & Waterman, Adv. Appl. Math. 2:482, 1981), 상동성 정렬 알고리즘(Needleman & Wunsch, J. Mol. Biol. 48:443, 1970), 유사성 검색 방법(person & Lipman, Proc. Nat'l. Acad. Sci. USA 85:2444, 1988), 알고리즘의 컴퓨터에 의한 실행(GAP, BESTFIT, FASTA 및 TFASTA, Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), 또는 수동식 정렬법 및 시각에 의한 관찰에 의해 비교용 서열의 최적 정렬법이 수행될 수 있다. 상동성 또는 동일성을 측정하기 위한 기타 알고리즘으로는 예를 들어, BLAST 프로그램(Basic Local Alignment Search Tool at the National Center for Biological Information) 이외에도, ALIGN, AMAS(Analysis of Multiply Aligned Sequences), AMPS(Protein Multiple Sequence Alignment), ASSET(Aligned Segment Statistical Evaluation Tool), BANDS, BESTSCOR, BIOSCAN(Biological Sequence Comparative Analysis Node), BLIMPS(BLOCKS IMPROVED SEARCHER), FASTA, 인터벌스 & 포인트(Intervals & Points), BMB, CLUSTAL V, CLUSTAL W, CONSENSUS, LCONSENSUS, WCONSENSUS, TMAALTM-워터맨 알고리즘(Smith-Waterman algorithm), DARWIN, 라스베가스 알고리즘, FNAT(Forced Nucleotide Alignment Tool), 프레임얼라인(Framealign), 프레임서치(Framesearch), DYNAMIC, FILTER, FSAP(Fristensky Sequence Analysis Package), GAP(Global Alignment Program), GENAL, GIBBS, GenQuest, ISSC(Sensitive Sequence Comparison), LALIGN(Local Sequence Alignment), LCP(Local Content Program), MACAW(Multiple Alignment Construction & Analysis Workbench), MAP(Multiple Alignment Program), MBLKP, MBLKN, PIMA(Pattern-Induced Multi-sequence Alignment), SAGA(Sequence Alignment by Genetic Algorithm) 및 WHAT-IF를 포함한다. 이러한 정렬 프로그램은 또한 실질적으로 동일한 서열을 가지는 폴리뉴클레오타이드 서열

을 동정하기 위한 게놈 데이터베이스를 스크리닝하는데에도 사용될 수 있다. 다수의 게놈 데이터베이스를 입수할 수 있는데, 예를 들어, 인간 게놈의 상당 부분은 인간 게놈 서열 결정 프로젝트(Gibbs, 1995)의 일환으로서 얻을 수 있다. 21개 이상의 기타 게놈들 예를 들어, 엠.제니탈륨(*M. genitalium*)(Fraser와 다수, 1995), 엠.제너쉬(*M. jannaschii*)(Bult와 다수, 1996), 에이치.인플루엔자에(*H. influenzae*)(Fleischmann와 다수, 1995), 이.콜라이(Blattner와 다수, 1997) 및 효모(에스. 세레비지아에)(Mewes와 다수, 1997) 및 디.멜라노가스터(*D. melanogaster*)(Adams와 다수, 2000)의 게놈들은 이미 서열 결정되어 있다. 뿐만 아니라, 표준적인 유기체 예를 들어, 마우스, 씨.엘레강스(*C. elegans*) 및 아라비도시스 종의 게놈을 서열 결정하는 방법은 상당히 진보되었다. 약간의 기능상 정보와 함께 게놈 정보를 포함하는 몇몇 데이터베이스는 상이한 조직에 의해 보존되고 있으며, 인터넷을 통하여도 입수할 수 있다.

[0203] 하나의 측면에서, BLAST 및 BLAST 2.0 알고리즘이 사용되는데, 이 알고리즘은 문헌[Altschul와 다수, Nuc. Acids Res. 25:3389-3402, 1977 및 Altschul와 다수, J. Mol. Biol. 215:403-410, 1990]에 각각 개시되어 있다. BLAST 분석 수행용 소프트웨어는 국립 생명 과학 정보 센터(National Center for Biotechnology Information)를 통해 공중이 입수할 수 있다. 이 알고리즘은 의문 서열 내 단어 길이 W인 짧은 단어 수를 확인함으로써 고 스코어링 서열 쌍(high scoring sequence pairs; HSPs)을 우선 동정하는 단계를 포함하며, 이때 이 서열 쌍은 데이터 베이스 서열 내의 동일한 길이를 가지는 단어와 정렬될 때, 몇몇 양의 값을 가지는 최소 스코어 T와 부합하거나 또는 이를 만족시킨다. T는 인접하여 존재하는 단어의 스코어 최소값을 의미한다(Altschul와 다수, 상동). 이와 같은 처음의 인접 단어 히트(word hits)는 히트를 포함하는 더욱 긴 HSP를 찾아내기 위한 검색을 개시하기 위한 시드(seed)로서의 역할을 갖는다. 단어 히트는 누적 정렬 스코어가 증가함에 따라서 각 서열을 따라 양 방향으로 연장된다. 누적 스코어는 뉴클레오티드 서열, 매개 변수 M(매치되는 잔기의 쌍에 대한 보상 스코어로서, 항상 > 0임)을 사용하여 계산된다. 아미노산 서열에 있어서, 스코어링 매트릭스는 누적 스코어를 계산하는데에 사용된다. 이 단어 히트를 각 방향으로 연장하던 것을 다음과 같은 때에 중지시킨다: 즉, 누적 정렬 스코어가 최대도 도달하는 수치로부터 얻어지는 양 X만큼 줄어들 때; 하나 이상의 네가티브-스코어링 잔기 정렬이 축적됨에 따라서 누적 스코어가 0 이하에 도달할 때; 또는 양 서열 중 어느 하나의 서열 말단에 도달할 때. BLAST 알고리즘 매개 변수인 W, T 및 X는 정렬 결과의 정확성과 속도를 결정한다. BLASTN 프로그램(뉴클레오티드 서열에 대한 프로그램)은 디폴트값으로서 단어 길이(W) = 11, 기대치(E) = 10, M = 5, N = -4를 사용하고, 양 가닥을 비교하는 것이다. 아미노산 서열의 경우, BLASTP 프로그램은 디폴트값으로서 단어 길이 = 3, 기대치(E) = 10를 사용하고, BLOSUM62 스코어링 매트릭스[Henikoff & Henikoff, Proc. Natl. Acad. Sci. USA 89:10915, 1989]는 정렬값(B) = 50, 기대치(E) = 10, M = 5, N = -4를 사용하고, 양 가닥을 비교하는 것이다.

[0204] BLAST 알고리즘은 또한 2개 서열 간 유사성을 통계학적으로 분석하는 것이다[예를 들어, Karlin & Altschul, Proc. Natl. Acad. Sci. USA 90:5873, 1993 참조]. BLAST 알고리즘에 의해 유사성을 측정하는 한 가지 수단은 확률의 총합 중 가장 작은 값(P(N))으로서, 이는 2개의 뉴클레오티드 또는 아미노산 서열이 우연히 매치될 확률을 표시하는 방법이다. 예를 들어, 시험 핵산을 참조 서열과 비교하였을 때 최소 확률 총합이 약 0.2 미만일 경우, 더욱 바람직하게는 약 0.01 미만일 경우, 그리고 가장 바람직하게는 약 0.001 미만일 경우, 그 핵산은 참조 서열과 유사한 것으로 간주된다.

[0205] 하나의 측면에서, 단백질 및 핵산 서열의 상동성은 기본 국소 정렬 검색 도구(Basic Local Alignment Search Tool; "BLAST")를 사용하여 평가된다. 특히, 5개의 구체적인 BLAST 프로그램이 다음과 같은 기능을 수행하는데에 사용된다:

[0206] (1) BLASTP 및 BLAST3은 단백질 서열 데이터베이스에 대해 아미노산 의문 서열(query sequence)을 비교하는데 사용됨;

[0207] (2) BLASTN는 뉴클레오티드 서열 데이터베이스에 대해 뉴클레오티드 의문 서열을 비교하는데 사용됨;

[0208] (3) BLASTX는 단백질 서열 데이터베이스에 대해 의문 뉴클레오티드 서열(양쪽 서열 모두 해당함)의 6-프레임 개념 번역 산물을 비교하는데 사용됨;

[0209] (4) TBLASTN는 6개의 해독 틀(양쪽 가닥 모두 해당함) 내 번역된 뉴클레오티드 서열 데이터베이스에 대해 의문 단백질 서열을 비교하는데 사용됨; 및

[0210] (5) TBLASTX는 뉴클레오티드 서열 데이터베이스의 6-프레임 번역에 대해 뉴클레오티드 의문 서열의 6-프레임 번역 결과를 비교하는데 사용됨.

- [0211] BLAST 프로그램은 의문 아미노산 서열 또는 핵산 서열과, 단백질 또는 핵산 서열 데이터베이스로부터 얻은 시험 서열 사이의 유사한 분절(본원에서는 "고-스코어링 분절 쌍"이라 칭함)을 확인함에 따라서 상동성인 서열을 동정한다. 하나의 측면에서, 고-스코어링 분절 쌍은 당 업계에 공지된 다수의 스코어링 매트릭스 수단에 의해 동정된다(즉, 정렬된다). 하나의 측면에서, 사용된 스코어링 매트릭스는 BLOSUM62 매트릭스[Gonnet (1992) Science 256:1443-1445; Henikoff and Henikoff (1993) Proteins 17:49-61]이다. 하나의 측면에서, 소수의 경우이긴 하지만 PAM 또는 PAM250 매트릭스도 사용될 수 있다[예를 들어, Schwartz and Dayhoff, eds., 1978, Matrices for Detecting Distance Relationships: Atlas of Protein Sequence and Structure, Washington: National Biomedical Research Foundation]. BLAST 프로그램은 미국 국립 의학 도서관에서 입수할 수 있다.
- [0212] 상기 알고리즘에서 사용된 매개 변수들은 서열의 길이와 상동성 정도에 따라서 맞추어질 수 있다. 몇몇 측면에서, 상기 매개 변수는 사용자의 지시 없이도 알고리즘에 의해 사용되는 디폴트 매개 변수일 수 있다.
- [0213] 컴퓨터 시스템 및 컴퓨터 프로그램 제품
- [0214] 본 발명은 본 발명의 핵산과 폴리펩티드 서열을 수록 또는 저장하고 있는 컴퓨터, 컴퓨터 시스템, 컴퓨터 판독 가능 매체 및 컴퓨터 프로그램 제품 등을 제공한다. 뿐만 아니라, 예를 들어, 컴퓨터 내에서 서열 동일성을 측정 및 동정하기 위하여(핵산이 본 발명의 범위에 속하는지 여부를 결정하기 위하여), 그리고 구조적 상동성 및 모티프를 측정 및 동정하기 위하여 본 발명의 방법을 실시함에 있어서, 본 발명의 핵산 또는 폴리펩티드 서열은 컴퓨터에 의해 판독 및 접근할 수 있는 임의의 매체 상에 저장, 기록 및 조작될 수 있다.
- [0215] 본원에 사용된 "기록된" 및 "저장된"이란 용어는, 컴퓨터 매체 상에 정보를 저장한 방식을 의미하는 것이다. 당업자는 컴퓨터 판독 가능 매체 상에 정보를 기록하여 본 발명의 핵산 및/또는 폴리펩티드 서열 중 하나 이상을 포함하는 제품을 생산하는 임의의 공지된 방법을 용이하게 채택할 수 있다. 본원에 사용된 "컴퓨터", "컴퓨터 프로그램" 및 "프로세서"란 용어는, 광범위한 의미로 사용되는 것으로서, 이하 상세히 기술된 바와 같은 장치들을 모두 포함하는 의미이다. 특정 폴리펩티드 또는 단백질의 "암호화 서열" 또는 이를 "암호화하는 서열"이란, 적당한 조절 서열의 제어하에 있을 때 폴리펩티드 또는 단백질로 전사 및 번역되는 핵산 서열을 의미한다.
- [0216] 본 발명의 폴리펩티드는 본 발명의 대표적인 서열 및 이와 실질적으로 상동성인 서열, 그리고 이러한 서열들의 임의의 부분 서열(단편)을 포함한다. 하나의 측면에서, 실질적으로 동일하거나, 상동성인 폴리펩티드 서열이란, 본 발명의 대표적인 서열과 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 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% 이상 상동성이거나(동일하거나), 또는 완벽한(100%) 서열 동일성을 갖는 서열을 의미한다.
- [0217] 상동성(서열 동일성)은 본원에 기술된 컴퓨터 프로그램 및 매개 변수 중 임의의 것을 사용하여 측정될 수 있다. 본 발명의 핵산 및 폴리펩티드 서열은 컴퓨터에 의해 판독 및 접근할 수 있는 임의의 매체 상에 저장, 기록 및 조작될 수 있다. 본원에 사용된 "기록된" 및 "저장된"이란 용어는, 컴퓨터 매체 상에 정보를 저장하는 방식을 나타내는 것이다. 당업자는 컴퓨터 판독 가능 매체 상에 정보를 기록하여 본 발명의 핵산 서열 중 하나 이상, 본 발명의 폴리펩티드 서열 중 하나 이상을 포함하는 제품을 생산하는 임의의 공지된 방법을 용이하게 채택할 수 있다. 본 발명의 다른 측면은 본 발명의 핵산 또는 폴리펩티드 서열이 2, 5, 10, 15 또는 20개 이상 기록되어 있는 컴퓨터 판독 가능 매체에 관한 것이다.
- [0218] 본 발명의 다른 측면은 본 발명의 핵산 서열 중 하나 이상이 기록되어 있는 컴퓨터 판독 가능 매체에 관한 것이다. 본 발명의 다른 측면은 본 발명의 폴리펩티드 서열 중 하나 이상이 기록되어 있는 컴퓨터 판독 가능 매체에 관한 것이다. 본 발명의 다른 측면은 상기 제시한 핵산 또는 폴리펩티드 서열 중 2, 5, 10, 15 또는 20개 이상이 기록되어 있는 컴퓨터 판독 가능 매체에 관한 것이다.
- [0219] 컴퓨터 판독 가능 매체는 자기에 의해 읽을 수 있는 매체, 빛에 의해 읽을 수 있는 매체, 전기에 의해 읽을 수 있는 매체 및 자기/광학 매체를 포함한다. 예를 들어, 컴퓨터 판독 가능 매체는 하드 디스크, 플로피 디스크, 자기 테이프, CD-ROM, 디지털 다목적 디스크(DVD), 랜덤 액세스 메모리(RAM) 또는 리드 온리 메모리(Read Only Memory; ROM), 그리고 당 업계에게 공지된 기타 유형의 매체일 수 있다.
- [0220] 본 발명의 측면은 본원에 개시된 서열 정보를 저장하고 조작하는 시스템(예를 들어, 인터넷을 중심으로 하는 시스템) 예를 들어, 컴퓨터 시스템을 포함한다. 컴퓨터 시스템 (100)의 일례에 관하여는 도 1에 블록 선도로서 나타내었다. 본원에 사용된 "컴퓨터 시스템"이란, 본 발명의 핵산 서열의 뉴클레오티드 서열, 또는 본 발명의 폴리펩티드 서열을 분석하는데 사용되는 하드웨어 부품, 소프트웨어 부품 및 데이터 저장 부품을 의미하는 것이

다. 하나의 측면에서, 컴퓨터 시스템(100)은 서열 데이터를 처리, 접근 및 조작하기 위한 프로세서를 포함한다. 이 프로세서(105)는 중앙 처리 장치 예를 들어, 인텔 코퍼레이션의 펜티엄 III, 또는 썬, 모토로라, 컴팩, AMD 또는 인터내셔널 비지니스 머신(International Business Machines)사의 유사한 프로세서 중 널리 공지된 임의의 유형의 것일 수 있다.

[0221] 하나의 측면에서, 컴퓨터 시스템(100)은 프로세서(105)와 데이터 저장용인 하나 이상의 내부 데이터 저장 부품(110), 그리고 데이터 저장 부품에 저장되어 있는 데이터 검색용인 하나 이상의 데이터 검색 장치를 포함하는 일반적인 용도의 시스템이다. 당업자는 현재 시판중인 컴퓨터 시스템 중 임의의 것이 적당하다는 사실을 쉽게 알 수 있다.

[0222] 하나의 특정 측면에서, 컴퓨터 시스템(100)은 메인 메모리(115)(하나의 측면에서는, RAM)에 연결되어 있는 버스(bus)에 결합되어 있는 프로세서(105)와, 데이터가 기록되어 있는 하나 이상의 내부 데이터 저장 장치(110) 예를 들어, 하드 드라이브 및/또는 기타 컴퓨터 판독 가능 매체를 포함한다. 몇몇 측면에서, 상기 컴퓨터 시스템(100)은 또한 내부 데이터 저장 장치(110)에 저장되어 있는 데이터를 판독하기 위한 하나 이상의 데이터 검색 장치(118)를 추가로 포함한다.

[0223] 상기 데이터 검색 장치(118)로서는 예를 들어, 플로피 디스크 드라이브, 컴팩트 디스크 드라이브, 자기 테이프 드라이브, 또는 원격 데이터 저장 시스템과 연결될 수 있는(예를 들어, 인터넷을 통하여) 모뎀이 있다. 몇몇 측면에서, 내부 데이터 저장 장치(110)는 제어 논리 및/또는 데이터가 기록되어 있는, 분리 가능한 컴퓨터 판독 가능 매체 예를 들어, 플로피 디스크, 컴팩트 디스크, 자기 테이프 등이다. 컴퓨터 시스템(100)은 제어 논리 및/또는 데이터 검색 장치에 삽입된 데이터 저장 부품으로부터 유래하는 데이터를 판독하기 위하여, 적당한 소프트웨어에 의해 프로그래밍되거나, 또는 이 소프트웨어를 포함하는 것이 유리할 수 있다.

[0224] 컴퓨터 시스템(100)은 컴퓨터 사용자에게 출력 결과를 보여주는 데 사용되는 디스플레이(120)를 포함한다. 또한, 상기 컴퓨터 시스템(100)은 다른 컴퓨터 시스템(125a~c)과 네트워크로 연결되거나 또는 광역 네트워크로 연결되어 있어서, 이 컴퓨터 시스템(100)에 한꺼번에 접근할 수 있음을 알아야 할 것이다.

[0225] 본 발명의 핵산 서열의 뉴클레오티드 서열, 또는 본 발명의 폴리펩티드 서열에 접근하여 이를 처리하는 소프트웨어(예를 들어, 검색 도구, 비교 도구 및 모델링 도구 등)는 실행중 메인 메모리(115)에 담겨있을 수 있다.

[0226] 몇몇 측면에서, 컴퓨터 시스템(100)은 컴퓨터 판독 가능 매체에 저장되어 있는 본 발명의 핵산 서열 또는 본 발명의 폴리펩티드 서열을, 컴퓨터 판독 가능 매체에 저장되어 있던 참조 뉴클레오티드 또는 폴리펩티드 서열(들)과 비교하기 위한, 서열 비교 알고리즘을 포함할 수도 있다. "서열 비교 알고리즘"이란, 데이터 저장 수단에 저장되어 있던 기타 뉴클레오티드 서열 및/또는 화합물과 뉴클레오티드 서열을 비교하기 위해 컴퓨터 시스템(100) 상에서 실행되는(국소적 또는 광역적으로 실행되는) 하나 이상의 프로그램을 의미하는 것이다. 예를 들어, 서열 비교 알고리즘은, 컴퓨터 판독 가능 매체에 저장되어 있는 본 발명의 핵산 서열의 뉴클레오티드 서열, 또는 본 발명의 폴리펩티드 서열을, 컴퓨터 판독 가능 매체에 저장되어 있던 참조 서열과 비교하여 상동성 또는 구조적 모티프를 확인할 수 있다.

[0227] 도 2는 신규 서열과 데이터베이스에 있던 서열 사이의 상동성 수준을 측정하기 위하여, 신규 뉴클레오티드 또는 단백질 서열과 서열 데이터베이스를 비교하는 공정(200)의 한 측면을 나타내는 공정도이다. 서열 데이터베이스는 컴퓨터 시스템(100) 내에 저장된 개인적 데이터베이스일 수 있거나, 또는 인터넷을 통하여 입수할 수 있는 GENBANK와 같은 공중 데이터베이스일 수 있다.

[0228] 상기 공정(200)은 출발 단계(201)로부터 시작하여, 이후 단계(202) 즉, 비교될 신규 서열이 컴퓨터 시스템(100) 내 메모리에 저장되는 단계로 이동한다. 전술한 바와 같이, 상기 메모리는 임의의 유형의 메모리 예를 들어, RAM 또는 내부 저장 장치일 수 있다.

[0229] 이후, 공정(200)은 단계(204) 즉, 서열 데이터베이스가 분석 및 비교를 위해 공개되는 단계로 진행된다. 이후, 공정(200)은 단계(206) 즉, 데이터베이스에 저장되어 있던 제1 서열을 컴퓨터 상에서 메모리로 읽어내는 단계로 이동한다. 이후, 단계(210)에서는 비교가 실행되는데, 그 결과 상기 제1 서열이 제2 서열과 동일한지 여부를 결정하게 된다. 이 단계는 신규 서열과 데이터베이스의 제1 서열을 정확히 비교하기 위한 목적으로만 행하여지는 것은 아님에 주목해야 할 것이다. 상기 서열들이 동일하지 않더라도 2개의 뉴클레오티드 또는 단백질 서열을 비교할 수 있는, 널리 공지된 방법이 당업자에게 공지되어 있다. 예를 들어, 2개의 시험 서열 사이 상동성 수준을 증가시키기 위해, 하나의 서열에 갭을 도입시킬 수 있다. 갭 또는 다른 특징이 비교중인 서열에 도입되었는지 여부를 제어하는 매개 변수는 일반적으로 컴퓨터 시스템의 사용자에게 의해 입력된다.

- [0230] 일단, 2개 서열을 상기 단계(210)에서 비교하면, 이 2개의 서열이 동일한지 여부를 단계(210)에서 확실히 확인할 수 있게 된다. 물론, "동일하다"는 용어는, 완전히 동일한 서열에 국한되는 것은 아니다. 사용자에게 의해 입력된 상동성 매개 변수 내에 포함되는 서열은 공정(200)에서 "동일한" 것으로 마킹될 것이다.
- [0231] 만일 상기 2개의 서열이 동일하다고 판명되면, 공정(200)은 단계(214) 즉, 데이터베이스로부터 유래하는 서열의 이름을 사용자에게 알려주는 단계로 이동한다. 이 단계는 명명된 서열이 입력된 상동성을 사용자에게 알려준다. 일단, 저장된 서열의 명칭이 사용자에게 알려지면, 공정(200)은 결정 단계(218) 즉, 여전히 그 서열이 데이터베이스 내에 존재하는지 여부를 결정하는 단계로 이동하게 된다. 데이터베이스 내에 서열이 더 이상 존재하지 않으면, 공정(200)은 최종 단계(220)에서 종결된다. 그러나, 만일 여전히 그 서열이 데이터베이스 내에 존재하면, 공정(200)은 다시 단계(214) 즉, 화살표가 데이터베이스의 다음 서열로 이동하여, 신규 서열과 비교될 수 있는 단계로 이동하게 된다. 이러한 방식으로, 신규 서열은 정렬되고, 데이터베이스 내 모든 서열과 비교된다.
- [0232] 만일, 서열들이 상동성을 갖지 않는 것으로 결정 단계(212)에서 측정된다면, 공정(200)은 즉시 결정 단계(218)로 이동하게 되고, 그 결과 기타 임의의 서열이 비교용 데이터베이스에서 얻을 수 있는지 여부를 결정할 수 있다.
- [0233] 그러므로, 본 발명의 하나의 측면은 프로세서, 본 발명의 핵산 서열 또는 본 발명의 폴리펩티드 서열을 담고 있는 데이터 저장 장치, 본 발명의 핵산 서열 또는 본 발명의 폴리펩티드 서열과 비교될 참조 뉴클레오티드 서열 또는 폴리펩티드 서열, 그리고 서열 비교자(sequence comparer)(비교 실행용)를 검색 가능하도록 담고 있는 데이터 저장 장치를 포함하는 컴퓨터 시스템에 관한 것이다. 상기 서열 비교 자는 비교된 서열들간 상동성 수준을 알려줄 수 있거나, 또는 전술한 핵산 코드에 존재하는 구조적 모티프, 본 발명의 핵산 서열 또는 본 발명의 폴리펩티드 서열을 확인시켜줄 수 있거나, 또는 이러한 핵산 코드와 폴리펩티드 코드를 비교하여주는 서열 내 구조적 모티프를 확인시켜줄 수 있다. 몇몇 측면에서, 데이터 저장 장치는 본 발명의 핵산 서열 또는 본 발명의 폴리펩티드 서열 중 2, 5, 10, 15, 20, 25, 30 또는 40개 이상의 서열을 저장할 수 있다.
- [0234] 본 발명의 다른 측면은 본 발명의 핵산 서열, 또는 본 발명의 폴리펩티드 서열 및 참조 서열 사이의 상동성 수준을 측정하는 방법에 관한 것이다. 이 방법은 상동성 수준을 측정하는 컴퓨터 프로그램을 사용함으로써, 핵산 코드 또는 폴리펩티드 코드와, 참조 뉴클레오티드 서열 또는 폴리펩티드 서열을 판독하는 단계; 및 컴퓨터 프로그램을 이용하여 이 핵산 코드 또는 폴리펩티드 코드와, 참조 뉴클레오티드 서열 또는 폴리펩티드 서열 사이의 상동성을 측정하는 단계를 포함한다. 상기 컴퓨터 프로그램은 상동성 수준을 측정하는 다수의 컴퓨터 프로그램 중 임의의 것 예를 들어, 본원에 특정하여 열거한 프로그램(예를 들어, 디폴트 매개 변수 또는 임의의 변형 매개 변수를 포함하는 BLAST2N)일 수 있다. 이 방법은 전술한 컴퓨터 시스템을 이용하여 실시될 수 있다. 이 방법은 또한 컴퓨터 프로그램을 이용하여 전술한 본 발명의 핵산 서열, 또는 본 발명의 폴리펩티드 서열 중 2, 5, 10, 15, 20, 25, 30 또는 40개 이상을 판독한 후, 핵산 코드 또는 폴리펩티드 코드와 참조 뉴클레오티드 서열 또는 폴리펩티드 서열 사이의 상동성을 측정함으로써 실시될 수도 있다.
- [0235] 도 3은 2개 서열이 상동성인지 여부를 결정하기 위하여 컴퓨터 내에서 행하여지는 공정(250)의 한 측면을 나타내는 것이다. 이 공정(250)은 출발 단계(252)에서 시작된 후, 단계(254) 즉, 비교될 제1 서열이 메모리에 저장되는 단계로 이동한다. 이후, 비교될 제2 서열은 단계(256)에서 메모리에 저장된다. 이후, 공정(250)은 단계(260) 즉, 제1 서열의 첫 번째 문자가 판독되는 단계로 이동한 다음, 단계(262) 즉, 제2 서열의 첫 번째 문자가 판독되는 단계로 이동하게 된다. 만일 서열이 뉴클레오티드 서열이면, 그 문자는 일반적으로, A, T, C, G 또는 U 중의 어느 하나일 것임을 이해하여야 할 것이다. 만일 서열이 단백질 서열이면, 단일 문자 아미노산 코드는 본 발명의 한 측면을 이룰 것이며, 또한 제1 서열 및 제2 서열은 용이하게 비교될 수 있을 것이다.
- [0236] 이후, 결정 단계(264)에서는 상기 2개의 문자가 동일한지 여부에 대해 측정된다. 만일 상기 2개의 문자들이 동일하면, 공정(250)은 단계(268) 즉, 제1 서열과 제2 서열의 다음 문자가 판독되는 단계로 이동하게 된다. 이후, 상기 다음 문자가 동일한지 여부를 결정한다. 만일 상기 다음 문자가 동일하다면, 공정(250)은 2개의 문자가 동일하지 않을 때까지 이와 같은 주기를 계속해서 반복 수행하게 된다. 만일 상기 다음 문자가 동일하지 않다고 판명되면, 공정(250)은 결정 단계(274)로 이동하여, 그 결과 양 서열 중 어느 하나에 판독될 더 이상의 문자가 존재하는지 여부를 결정하게 된다.
- [0237] 만일 판독될 문자가 더 이상 존재하지 않는다면, 공정(250)은 단계(276) 즉, 제1 서열 및 제2 서열 간 상동성 수준을 사용자에게 알려주는 단계로 이동하게 된다. 상동성 수준은, 제1 서열 내 서열의 총수와 동일한 수의 서열들 사이에 존재하는 문자의 비율을 계산함으로써 측정된다. 그러므로, 만일 처음 100개의 뉴클레오티드 서열

에 존재하는 모든 문자가 제2 서열에 존재하는 모든 문자와 함께 정렬되면, 상동성 수준은 100%가 될 것이다.

[0238] 대안적으로, 컴퓨터 프로그램은 본 발명에 제시된 핵산 서열의 뉴클레오티드 서열을 하나 이상의 참조 뉴클레오티드 서열과 비교하여, 본 발명의 핵산 코드가 하나 이상의 위치에서 참조 핵산 서열과 상이한지 여부를 결정하는 컴퓨터 프로그램일 수 있다. 경우에 따라, 이러한 프로그램은 참조 폴리뉴클레오티드 서열 또는 본 발명의 핵산 서열 중 어느 한 서열에 비해, 삽입, 결실 또는 치환된 뉴클레오티드의 길이 및 상동성을 기록한다. 하나의 측면에서, 컴퓨터 프로그램은 참조 뉴클레오티드 서열에 비해, 본 발명의 핵산 서열이 단일의 뉴클레오티드 다형성(SNP)을 가지는지 여부를 결정하는 프로그램이다.

[0239] 그러므로, 본 발명의 다른 측면은 본 발명의 핵산 서열이 하나 이상의 뉴클레오티드에서 참조 뉴클레오티드 서열과 상이한지 여부를 결정하는 방법에 관한 것으로서, 이 방법은 핵산 서열 간 차이를 동정하는 컴퓨터 프로그램을 이용하여 참조 뉴클레오티드 서열과 핵산 코드를 판독하는 단계; 및 컴퓨터 프로그램을 이용하여 핵산 코드와 참조 뉴클레오티드 서열 간 차이를 동정하는 단계를 포함한다. 동일한 측면에서, 상기 컴퓨터 프로그램은 단일 뉴클레오티드 다형성을 동정하는 프로그램이다. 이 방법은 전술한 컴퓨터 시스템과 도 3에 도시된 방법에 의해 실시될 수 있다. 이 방법은 또한 본 발명의 핵산 서열 중 2, 5, 10, 15, 20, 25, 30 또는 40개 이상의 핵산 서열과 참조 뉴클레오티드 서열을 컴퓨터 프로그램을 이용하여 판독하는 단계; 및 핵산 코드와 참조 뉴클레오티드 서열 간 차이를 컴퓨터 프로그램을 이용하여 동정하는 단계에 의해 실시될 수도 있다.

[0240] 다른 측면에서, 컴퓨터를 바탕으로 하는 시스템은 또한 본 발명의 핵산 서열 또는 본 발명의 폴리펩티드 서열 내에 존재하는 특징들을 동정하기 위한 식별자를 포함할 수도 있다. "식별자(Identifier)"란, 본 발명의 핵산 서열 또는 본 발명의 폴리펩티드 서열 내에 존재하는 임의의 특징을 동정하는 하나 이상의 프로그램을 의미한다. 하나의 측면에서, 상기 식별자는 본 발명의 핵산 서열 내에 존재하는 개방 해독 틀을 동정하는 프로그램을 포함할 수 있다.

[0241] 도 4는 서열 내에 특징이 존재하는지 여부를 결정하기 위한 식별자 공정(300)의 한 측면을 나타내는 공정도이다. 이 공정(300)은 출발 단계(302)에서 개시되고, 이후 단계(304) 즉, 특징을 가지고 있는지에 대해 체크될 제1 서열이 컴퓨터 시스템(100) 내에 존재하는 메모리(115)에 저장되는 단계로 이동하게 된다. 이후, 상기 공정(300)은 단계(306) 즉, 서열 특징에 관한 데이터베이스가 공개되는 단계로 이동한다. 이러한 데이터베이스는 이 특징의 명칭과 함께 각 특징의 속성에 관한 목록을 포함할 것이다. 예를 들어, 특징의 명칭이 "개시 코돈"이면, 그 속성은 "ATG"가 될 것이다. 다른 예로서는, 특징의 명칭이 "TAATAA 박스"이면, 이 특징의 속성은 "TAATAA"가 될 것이다. 이러한 데이터베이스의 예는 위스콘신 대학교 유전학 컴퓨터 그룹에 의해 개발되었다. 대안적으로, 이 특징들은 폴리펩티드의 구조적 모티프 예를 들어, 알파 나선 구조, 베타 병풍 구조 또는 폴리펩티드의 기능상 모티프 예를 들어, 효소 활성 부위, 나선부-선회부-나선부 모티프 또는 당 업자에게 공지된 기타 모티프 일 수 있다.

[0242] 일단 이 특징들에 관한 데이터베이스가 단계(306)에서 공개되면, 공정(300)은 단계(308) 즉, 제1 특징이 데이터베이스로부터 판독되는 단계로 이동한다. 이후, 제1 특징의 속성과 제1 서열은 단계(310)에서 비교된다. 이후 결정 단계(316)에서는 이 특징의 속성이 제1 서열에서 발견되는지 여부를 결정하게 된다. 만일 속성이 발견되면, 공정(300)은 단계(318) 즉, 발견된 특징의 명칭을 사용자에게 알려주는 단계로 이동하게 된다.

[0243] 이후, 공정(300)은 결정 단계(320) 즉, 이동한 특징이 데이터베이스에 존재하는지 여부를 결정하는 단계로 이동한다. 만일 더 이상의 특징이 존재하지 않으면, 공정(300)은 최종 단계(324)에서 종결될 것이다. 그러나, 만일 아직도 그 특징이 데이터베이스에 존재하면, 공정(300)은 단계(326)에서 다음 서열의 특징을 판독하고, 그 주기는 다시 단계(310) 즉, 다음 특징의 속성이 제1 서열과 비교되는 단계로 되돌아오게 된다. 만일 결정 단계(316)에서 그 특징의 속성이 제1 서열에서 발견되지 않으면, 공정(300)은 결정 단계(320)로 직접 이동되고, 데이터 베이스 내에 여전히 그 특징이 존재하는지 여부를 결정하게 되는 것이다.

[0244] 그러므로, 본 발명의 다른 측면은, 본 발명의 핵산 서열 또는 본 발명의 폴리펩티드 서열 내에 존재하는 특징을 동정하는 방법에 관한 것으로서, 이 방법은 핵산 코드(들) 또는 폴리펩티드 코드(들)에 있던 특징들을 동정하는 컴퓨터 프로그램을 이용하여, 이 핵산 코드(들) 또는 폴리펩티드 코드(들)를 해독하는 단계; 및 이 핵산 코드(들) 내에 존재하는 특징들을 컴퓨터 프로그램을 이용하여 동정하는 단계를 포함한다. 하나의 측면에서, 컴퓨터 프로그램은 개방 해독 틀을 동정하는 컴퓨터 프로그램을 포함한다. 이 방법은 컴퓨터 프로그램을 이용하여, 본 발명의 단일 서열 또는 본 발명의 핵산 서열 중 2, 5, 10, 15, 20, 25, 30 또는 40개 이상의 핵산 서열, 또는 본 발명의 폴리펩티드 서열을 해독한 후, 컴퓨터 프로그램을 이용하여 핵산 코드 또는 폴리펩티드 코드 내 특징들을 동정함으로써 실시될 수 있다.

- [0245] 본 발명의 핵산 서열 또는 본 발명의 폴리펩티드 서열은 다양한 방식으로 다양한 데이터 프로세서 프로그램에 저장되어 조작될 수 있다. 예를 들어, 본 발명의 핵산 서열 또는 본 발명의 폴리펩티드 서열은 워드 프로세서 파일 예를 들어, 마이크로소프트 워드(Microsoft WORD)TM 또는 워드퍼펙트(WORDBERFECT)TM에 텍스트로서 저장될 수 있거나, 또는 당 업자에게 익숙한 다양한 데이터베이스 프로그램에 ASCII 파일 예를 들어, DB2TM, 사이베이스(SYBASE)TM 또는 오라클(ORACLE)TM로서 저장될 수 있다. 뿐만 아니라, 다수의 컴퓨터 프로그램들과 데이터베이스가 본 발명의 핵산 서열 또는 본 발명의 폴리펩티드 서열과 비교될 참조 뉴클레오티드 서열 또는 폴리펩티드 서열의 서열 비교 알고리즘, 식별자 또는 공급원으로서 사용될 수 있다. 다음과 같은 목록은 본 발명을 제한하고자 하는 것이 아니라, 본 발명의 핵산 서열 또는 본 발명의 폴리펩티드 서열에 유용한 프로그램 및 데이터베이스에는 어떠한 것들이 있는지를 알려주고자 하는 것이다.
- [0246] 사용 가능한 프로그램 및 데이터베이스로서는 다음의 것들을 포함하나, 이에 한정되는 것은 아니다: 맥패턴(MACPATTERN)TM(EMBL), 디스커버리베이스(DISCOVERYBASE)TM(Molecular Applications Group), 제네마인(GENEMINE)TM(Molecular Applications Group), 룩(LOOK)TM(Molecular Applications Group), 맥룩(MACLOOK)TM(Molecular Applications Group), BLAST 및 BLAST2(NCBI), BLASTN 및 BLASTX(Altschul와 다수, J. Mol. Biol. 215: 403, 1990), FASTA(Pearson and Lipman, Proc. Natl. Acad. Sci. USA, 85: 2444, 1988), FASTDB(Brutlag와 다수 Comp. App. Biosci. 6:237-245, 1990), 카탈리스트(CATALYST)TM(Molecular Simulations Inc.), 카탈리스트/셰이프(Catalyst/SHAPE)TM(Molecular Simulations Inc.), 세리우스².DB엑세스(Cerius².DBAccess)TM(Molecular Simulations Inc.), 하이포젠(HYPOGEN)TM(Molecular Simulations Inc.), 인사이트 II(INSIGHT II)TM(Molecular Simulations Inc.), 디스커버(DISCOVER)TM(Molecular Simulations Inc.), 참(CHARMM)TM(Molecular Simulations Inc.), 펠릭스(FELIX)TM(Molecular Simulations Inc.), 델피(DELPHI)TM(Molecular Simulations Inc.), 퀀텀(Quantum)TM(Molecular Simulations Inc.), 호몰로지(Homology)(Molecular Simulations Inc.), 모델러(MODELER)TM(Molecular Simulations Inc.), 아이시스(ISIS)TM(Molecular Simulations Inc.), 퀀타/프로테인 디자인(Quanta/Protein Design)(Molecular Simulations Inc.), 웹랩(WebLab)(Molecular Simulations Inc.), 웹랩 다양성탐색기(WebLab Diversity Explorer)(Molecular Simulations Inc.), 진 익스플로러(Gene Explorer)(Molecular Simulations Inc.), SeqFold(Molecular Simulations Inc.), MDL 허용 화학 물질 디렉토리 데이터베이스(MDL Available Chemicals Directory database), MDL 약품 데이터 리포트 데이터베이스(MDL Drug Data Report data base), 종합 약품 화학 데이터베이스(Comprehensive Medicinal Chemistry database), 더웬트 세계 약품 인덱스 데이터베이스(Derwent's World Drug Index database), 바이오바이트마스터파일 데이터베이스(BioByteMasterFile database), Genbank 데이터베이스 및 Genseq 데이터베이스. 본 발명에 개시된 다수의 기타 프로그램 및 데이터베이스는 당 업자에게 익숙한 것일 것이다.
- [0247] 상기 프로그램을 이용하여 검출될 수 있는 모티프로서는 루신 지퍼 암호화 서열, 나선부-선회부-나선부 모티프, 글리코실화 부위, 유비퀴틴화 위치, 알파 나선 구조 및 베타 병풍 구조, 암호화된 단백질의 분비를 유도하는 신호 펩티드 암호화 신호 서열, 전사 조절 관여 서열 예를 들어, 호메오박스, 산성 확장부(acidic stretches), 효소 활성 부위, 기질 결합 부위 및 효소 절단 위치를 포함한다.
- [0248] 핵산의 혼성화
- [0249] 본 발명은 엄중한 조건 하에서 본 발명의 대표적인 서열[예를 들어, 서열 번호 1, 서열 번호 3, 서열 번호 5, 서열 번호 7, 서열 번호 9, 서열 번호 11, 서열 번호 13, 서열 번호 15, 서열 번호 17, 서열 번호 19, 서열 번호 21, 서열 번호 23, 서열 번호 25, 서열 번호 27, 서열 번호 29, 서열 번호 31, 서열 번호 33, 서열 번호 35, 서열 번호 37, 서열 번호 39, 서열 번호 41, 서열 번호 43, 서열 번호 45, 서열 번호 47, 서열 번호 49, 서열 번호 51, 서열 번호 53, 서열 번호 55, 서열 번호 57, 서열 번호 59, 서열 번호 61, 서열 번호 63, 서열 번호 65, 서열 번호 67, 서열 번호 69, 서열 번호 71, 서열 번호 73, 서열 번호 75, 서열 번호 77, 서열 번호 79, 서열 번호 81, 서열 번호 83, 서열 번호 85, 서열 번호 87, 서열 번호 89, 서열 번호 91, 서열 번호 93, 서열 번호 95, 서열 번호 97, 서열 번호 99, 서열 번호 101, 서열 번호 103, 서열 번호 105, 서열 번호 107, 서열 번호 109, 서열 번호 111, 서열 번호 113, 서열 번호 115, 서열 번호 117, 서열 번호 119, 서열 번호 121, 서열 번호 123, 서열 번호 125, 서열 번호 127, 서열 번호 129, 서열 번호 131, 서열 번호 133, 서열 번호 135, 서열 번호 137, 서열 번호 139, 서열 번호 141, 서열 번호 143, 서열 번호 145, 서열 번호 147, 서열 번호 149, 서열 번호 151, 서열 번호 153, 서열 번호 155, 서열 번호 157, 서열 번호 159, 서열 번호 161, 서열 번호 163 또는 서열 번호 165](이하, 표 1a~1c, 2 및 3과, 실시예 1 및 4, 그리고 서열 목록 참조)에 혼성화하는 분리된 또는 재조합 핵산을 제공한다. 엄중한 조건이란, 매우 엄중한 조건, 중간 정도의 엄중한 조건 및

/또는 낮은 염증도의 조건 예를 들어, 본원에 기술된 바와 같이 염증도의 조건이 높은 경우와 낮은 경우일 수 있다. 하나의 측면에서, 이하에 기술된 바와 같이, 핵산이 본 발명의 범위 내에 속하는지 여부를 결정하는 조건은 세척 조건의 염증도가 어느 정도인지에 따라서 결정된다.

[0250] "혼성화"란, 염기쌍 형성 과정을 통하여 핵산 가닥이 상보성 사슬과 결합하는 과정을 의미한다. 혼성화 반응은 감수성이고 선택적이어서, 목적으로 하는 특정 서열은 그것이 시료 중에 낮은 농도로 존재할지라도 동정될 수 있다. 적당히 염증한 조건이란, 예를 들어, 예비 혼성화 및 혼성화 용액 중 염 또는 포름아미드의 농도, 또는 혼성화 온도로써 정의될 수 있으며, 이에 관하여는 당 업계에 널리 공지되어 있다. 대안적인 측면에서, 염증도는 염 농도의 감소, 포름아미드 농도의 증가, 또는 혼성화 온도의 상승에 의해 증가될 수 있다. 대안적 측면에서, 본 발명의 핵산은 본원에 제시된 바와 같이, 다양한 염증도의 조건(예를 들어, 고, 중 및 저) 하에서 혼성화되는 능력에 의해서 정의된다.

[0251] 하나의 측면에서, 고도 염증도의 조건 하에서 이루어지는 혼성화는 약 37~42℃에서 약 50%의 포름아미드로써 이루어진다. 하나의 측면에서, 혼성화 조건은 낮은 염증도의 조건 즉, 약 30~35℃의 온도에서 약 35~25%의 포름아미드로써 이루어지는 조건을 포함한다. 하나의 측면에서, 혼성화 조건은 고도 염증도의 조건 예를 들어, 42℃ 및 50% 포름아미드, 5× SSPE, 0.3% SDS 및 200 n/ml의 전단 및 변성 연어 정자 DNA를 사용하는 경우를 포함한다. 하나의 측면에서, 혼성화 조건은 이와 같은 낮은 염증도의 조건을 포함하되, 다만, 35℃로 감소된 온도 및 35%의 포름아미드 중에서 이루어진다. 구체적인 염증도 수준에 해당하는 온도 범위는 목적 핵산의 퓨린:피리미딘 비율을 계산하고, 이에 따라서 온도를 맞춤으로써 좁혀질 수 있다. 상기 범위 및 조건에 대한 변형예에 관하여는 당 업계에 널리 공지되어 있다.

[0252] 대안적 측면에서, 염증한 조건하에서 혼성화되는 특성을 가지는지에 의해 정의되는 본 발명의 핵산은 본 발명의 핵산 중 약 5개의 잔기 내지 전체 길이인 핵산일 수 있는데; 예를 들어, 상기 핵산은 길이가 5, 10, 15, 20, 25, 30, 35, 40, 50, 55, 60, 65, 70, 75, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950 또는 1000개 이상의 잔기일 수 있다. 전체 길이 핵산보다 길이가 짧은 핵산도 포함된다. 이와 같은 핵산은 예를 들어, 혼성화 프로브, 표지화 프로브, PCR용 올리고뉴클레오타이드 프로브, siRNA 또는 miRNA(단일 가닥 또는 이중 가닥), 항체 결합 펩티드(에피토프)를 암호화하는 안티센스 또는 서열, 모티프 및 활성 부위 등으로서 유용할 수 있다.

[0253] 하나의 측면에서, 본 발명의 핵산은 약 37~42℃ 및 약 50%의 포름아미드를 사용하는 고도로 염증한 조건 하에서 혼성화되는 특성을 가지는지에 의해 정의된다. 하나의 측면에서, 본 발명의 핵산은 약 30~35℃ 및 약 35~25%의 포름아미드를 사용하는 낮은 염증도의 조건 하에서 혼성화되는 특성을 가지는지에 의해 정의된다.

[0254] 대안적으로, 본 발명의 핵산은 약 42℃ 및 약 50%의 포름아미드, 5× SSPE, 0.3% SDS 및 반복적 서열 블로킹 핵산 예를 들어, cot-1 또는 연어 정자 DNA(예를 들어, 200 n/ml의 전단 및 변성 연어 정자 DNA)를 사용하는 고도로 염증한 조건 하에서 혼성화되는 특성을 가지는지에 의해 정의된다. 하나의 측면에서, 본 발명의 핵산은 낮은 온도 즉, 35~42℃에서 35% 또는 40%의 포름아미드를 사용하는 낮은 염증도의 조건 하에서 혼성화되는 특성을 가지는지에 의해 정의된다.

[0255] 핵산 혼성화 반응에 있어서, 특정 수준의 염증도를 이루는 데에 이용되는 조건은 혼성화될 핵산의 특성에 따라서 달라질 것이다. 예를 들어, 핵산의 혼성화 부위의 길이, 상보성 정도, 뉴클레오타이드 서열 조성(예를 들어, GC 대 AT 함량) 및 핵산 유형(예를 들어, RNA 대 DNA)는 혼성화 조건을 선택함에 있어서 고려될 수 있는 사항들이다. 부가적인 고려 사항은 핵산 중 하나가 예를 들어, 필터상에 고정되는지 여부이다.

[0256] 혼성화는 낮은 염증도, 중간 정도의 염증도 또는 높은 염증도의 조건 하에서 수행될 수 있다. 핵산 혼성화의 일례로서, 고정화된 변성 핵산을 함유하는 중합체 막은 우선, 0.9M NaCl, 50mM NaH₂PO₄, pH7.0, 5.0mM Na₂EDTA, 0.5% SDS, 10× 덴하르트 및 0.5mg/ml의 폴리리보아데닐산으로 이루어진 용액 중 45℃의 온도에서 30분 동안 예비적으로 혼성화된다. 대략 2×10^7 cpm(비활성 = $4 \sim 9 \times 10^8$ cpm/ug)의 ³²P 말단-표지화 올리고뉴클레오타이드 프로브는 이후 상기 용액에 첨가된다. 12~16 시간 동안 항온 처리한 후, 상기 막을 실온에서 0.5% SDS를 함유하는 1× SET(150mM NaCl, 20mM Tris 염산염, pH 7.8, 1mM Na₂EDTA) 중 실온에서 30분 동안 세척한 후, 신선한 1× SET 중 온도 -10℃에서 30분 더 세척한다(올리고뉴클레오타이드 프로브 결합). 이후, 상기 막을 방사능 사진 촬영 필름에 노출시켜, 혼성화 신호를 검출한다. 전술한 혼성화 조건 모두는 고도 염증도의 조건인 것으로 간주된다.

[0257] 혼성화 이후, 필터를 세척하여 비특이적으로 결합되어 있던 임의의 검출 가능 프로브를 제거할 수 있다. 이 필터를 세척할 때에 이용된 염증도는 또한 혼성화될 핵산의 특성, 혼성화될 핵산의 길이, 상보성 정도, 뉴클레오

티드 서열 조성(예를 들어, GC 대 AT 함량) 및 핵산의 유형(예를 들어, RNA 대 DNA)에 따라서 달라질 수도 있다. 점차적으로 높아지는 염중도의 조건하의 세척의 예로서는 다음의 경우들을 포함한다: $2\times$ SSC, 0.1% SDS, 실온에서 15분 동안(낮은 염중도); $0.1\times$ SSC, 0.5% SDS, 실온에서 30분~1시간(중간 정도의 염중도); $0.1\times$ SSC, 0.5% SDS, 혼성화 온도~68℃에서 15~30분 동안(높은 염중도); 그리고 0.15M NaCl, 72℃에서 15분 동안(매우 높은 염중도). 가장 낮은 염중도에서의 세척은 실온 및 $0.1\times$ SSC 중에서 수행될 수 있다. 상기 예들은 단지 필터를 세척하는데에 이용될 수 있는 조건의 한 세트에 관하여 예를 든 것이다. 당 업자는 상이한 염중도의 세척에 대하여 다양한 방법이 존재함을 알 것이다. 기타 몇몇 예를 이하에 들어놓았다.

[0258] 하나의 측면에서, 혼성화 조건은 세척 단계 즉, 실온 및 $1\times$ 150mM NaCl, 20mM Tris 염산염, pH7.8, 1 mM Na_2EDTA , 0.5% SDS를 포함하는 용액 중에서 30분 동안 세척한 후, 다시 새 용액 중에서 30분 동안 세척하는 것을 포함한다.

[0259] 프로브에 혼성화되는 핵산은 방사능 사진 촬영 또는 기타 통상의 기술에 의해 동정된다.

[0260] 프로브 서열에 대한 서열 동일성(상동성)의 수준이 줄어든 핵산을 동정하도록 상기 과정은 변형될 수 있다. 예를 들어, 검출 가능한 프로브에 대한 서열 동일성(상동성)이 감소한 핵산을 얻기 위해서는, 덜 염중한 조건을 적용시킬 수 있다. 예를 들어, 혼성화 온도는 Na^+ 농도 약 1M인 혼성화 완충액 중 68℃에서 42℃까지 5℃씩 낮아질 수 있다. 혼성화 이후, 필터는 혼성화 온도에서 $2\times$ SSC 및 0.5% SDS를 사용하여 세척될 수 있다. 이러한 조건들에 있어서 50℃를 초과하는 경우에는 "중간 정도" 조건으로 간주하고, 50℃ 이하인 경우에는 "낮은" 조건으로 간주한다. "중간 정도의" 혼성화 조건의 구체예는 상기 혼성화가 55℃에서 진행될 때이다. "낮은 염중도" 혼성화 조건의 구체예는 상기 혼성화가 45℃에서 진행될 때이다.

[0261] 대안적으로, 혼성화는 42℃의 온도에서 포름아미드를 함유하는 완충액 예를 들어, $6\times$ SSC 중에서 수행될 수 있다. 이러한 경우, 프로브에 대한 상동성 수준이 감소한 클론을 동정하기 위한 혼성화 완충액 중 포름아미드 농도는 50%에서 0%까지 5%씩 감소될 수 있다. 혼성화 이후, 필터는 50℃에서 $6\times$ SSC 및 0.5% SDS로 세척될 수 있다. 이러한 조건에 있어서 포름아미드가 25%를 초과하는 경우 "중간 정도의" 조건으로 간주하고, 포름아미드가 25% 이하인 경우 "낮은" 조건으로 간주한다. "중간 정도의" 혼성화 조건에 관한 구체예는 상기 혼성화가 30%의 포름아미드에서 진행되는 경우이다. "낮은 염중도" 혼성화 조건의 구체예는 상기 혼성화가 10% 포름아미드에서 진행되는 경우이다.

[0262] 그러나, 혼성화 방식이 무엇인지가 중요하지 않을 수 있다 - 즉, 핵산이 본 발명의 범위 내에 포함되는지 여부를 결정하는 조건을 제시할 세척 조건의 염중도의 경우가 바로 그러하다. 본 발명의 범위 내에 포함되는 핵산을 동정하는데에 사용되는 세척 조건으로서의 다음의 경우들을 포함한다: 염 농도 = 약 0.02 몰, pH = 7 및 온도 = 약 50℃ 이상 또는 약 55℃~약 60℃; 또는 염 농도 = 약 0.15 M NaCl, 온도 = 72° C에서 약 15분; 또는 염 농도 = 약 $0.2\times$ SSC, 온도 = 약 50℃ 이상 또는 약 55℃~약 60℃에서 약 15~약 20분; 또는 혼성화 복합체가 염 농도 약 $2\times$ SSC인 용액(0.1% SDS 함유)으로 2회 세척된 후(실온에서 15분 동안), $0.1\times$ SSC인 용액(0.1% SDS 함유)에서 2회 세척(68℃에서 15분 동안)되는 경우; 또는 이와 균등한 조건. SSC 완충액 및 균등한 조건에 관한 문헌[Sambrook, Tijssen 및 Ausubel]을 참조하시오.

[0263] 본 방법은 본 발명의 핵산을 분리 또는 동정하는데에 사용될 수 있다. 예를 들어, 상기 방법들은, 본 발명의 서열 중 하나와, 적어도 약 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400 또는 500개 이상의 연속된 염기를 포함하는 이의 단편, 그리고 이와 상보성인 서열로 이루어진 군으로부터 선택되는 핵산 서열과의 서열 동일성이 적어도 약 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 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% 이상인 서열을 가지는 핵산을 분리 또는 동정하는데에 사용될 수 있다. 서열 동일성(상동성)은 정렬 알고리즘을 이용하여 측정될 수 있다. 예를 들어, 상동성 폴리뉴클레오티드는 본원에 개시된 암호화 서열 중 어느 하나의 자연 발생 대립 형질 변이체인 암호화 서열을 가질 수 있다. 이러한 대립 형질 변이체는 본 발명의 핵산과 비교하였을 때, 하나 이상의 뉴클레오티드가 치환, 결실 또는 부가되어 있을 수 있다. 뿐만 아니라, 상기 방법들은, 본 발명의 폴리펩티드, 또는 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100 또는 150개 이상의 연속 아미노산을 포함하는 이의 단편과의 서열 동일성(상동성)이 약 99%, 95%, 90% 이상, 85% 이상, 80% 이상, 75% 이상, 70% 이상, 65% 이상, 60% 이상, 55% 이상, 또는 50% 이상인 폴리펩티드를 암호화하는 핵산을 분리하는데에 사용될 수 있다[여기서, 상기 서열 동일성은 서열 정렬 알고리즘(예를 들어, 디폴트 매개 변수를 가지는 FASTA 버전 3.0t78 알고리즘)을 사용

하여 측정됨].

[0264] 올리고뉴클레오타이드 프로브 및 이를 사용하는 방법

[0265] 본 발명은 또한 예를 들어, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성을 보유하는 폴리펩티드를 암호화하는 핵산 또는 이의 단편을 동정, 증폭 또는 분리하거나, 또는 이 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 유전자를 동정하는데에 사용될 수 있는 핵산 프로브를 제공한다. 하나의 측면에서, 상기 프로브는 본 발명의 핵산을 이루는 약 10개 이상의 연속된 염기들을 포함한다. 대안적으로, 본 발명의 프로브는 본 발명의 핵산에 제시된 연속된 염기 서열 약 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 120, 130, 150개, 또는 약 10~50개, 약 20~60개, 또는 약 30~70개 이상일 수 있다. 상기 프로브는 결합 및/또는 혼성화에 의해 핵산을 동정한다. 이 프로브는 본 발명의 어레이(이하 논의된 부분 참조) 예를 들어, 모세관형 어레이에 사용될 수 있다. 본 발명의 프로브는 또한 기타 핵산 또는 폴리펩티드를 분리하는데에도 사용될 수 있다.

[0266] 본 발명의 분리된 또는 재조합 핵산, 이와 상보성인 서열 또는 본 발명의 서열 중 하나의 적어도 약 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400 또는 500개 이상의 연속된 염기를 포함하는 이의 단편, 또는 이에 상보성인 서열도, 생물학적 시료 예를 들어, 흙 시료가 본 발명의 핵산 서열을 가지는 유기체, 또는 이 핵산을 얻을 수 있는 유기체를 함유하는지 여부를 결정하기 위한 프로브로서 사용될 수 있다. 이러한 방법에 있어서, 상기 핵산이 분리된 유기체를 보유할 가능성이 있는 생물학적 시료를 얻을 수 있으며, 그 핵산은 이 시료로부터 얻을 수 있다. 상기 핵산은 그 핵산에 존재하는 서열과 상보성인 임의의 서열에 특이적으로 혼성화될 수 있는 조건 하에서 프로브와 접촉된다.

[0267] 필요할 경우, 프로브를 상보성 서열에 특이적으로 혼성화시킬 수 있는 조건은, 이 프로브를 상보성 서열을 함유하는 것으로 공지된 시료로부터 유래하는 상보성 서열 및 상보성 서열을 함유하지 않는 대조군 서열과 접촉시킴으로써 결정될 수 있다. 혼성화 조건 예를 들어, 혼성화 완충액의 염 농도, 혼성화 완충액의 포름아미드 농도, 또는 혼성화 온도는 다양하여, 이 프로브를 상보성 핵산에 특이적으로 혼성화시킬 수 있는 조건을 결정할 수 있다.

[0268] 만일 시료가 핵산이 분리된 유기체를 함유하면, 이 프로브의 특이적 혼성화가 검출된다. 혼성화는 프로브를 검출 가능한 제제 예를 들어, 방사성 동위 원소, 형광 염료 또는 검출 가능한 생성물의 생산을 촉진할 수 있는 효소로 표지화함으로써 검출될 수 있다.

[0269] 시료 중에 상보성 핵산이 존재하는지 여부를 검출하기 위해 표지화된 프로브를 사용하는 다수의 방법들은 당업자에게 널리 공지되어 있다. 이러한 방법으로서는 서던 블롯(Southern Blots), 노던 블롯(Northern Blots), 콜로니 혼성화 방법(colony hybridization procedures) 및 닷 블롯(dot blot)을 포함한다. 이러한 방법들 각각에 대한 프로토콜은 문헌[Ausubel 외 다수 Current Protocols in Molecular Biology, John Wiley 503 Sons, Inc. (1997) 및 Sambrook 외 다수, Molecular Cloning: A Laboratory Manual 2nd Ed., Cold Spring Harbor Laboratory Press (1989)]에 제공되어 있다.

[0270] 대안적으로, 하나 이상의 프로브(핵산 시료중에 존재하는 임의의 상보성 서열에 특이적으로 혼성화될 수 있는 프로브 중 하나 이상)는 증폭 반응에 사용될 수 있으며, 그 결과 이 시료가 본 발명의 핵산 서열을 함유하는 유기체(예를 들어, 핵산이 분리된 유기체)를 함유하는지 여부를 결정할 수 있다. 하나의 측면에서, 프로브는 올리고뉴클레오타이드를 포함한다. 하나의 측면에서, 증폭 반응은 PCR 반응을 포함할 수 있다. PCR 프로토콜에 관하여는 문헌[Ausubel and Sambrook, 상동]에 기술되어 있다. 대안적으로, 증폭 반응으로서 리가제 연쇄 반응, 3SR 또는 가닥 치환 반응을 포함할 수 있다[Barany, F., "The Ligase Chain Reaction in a PCR World", PCR Methods and Applications 1:5-16, 1991; E. Fahy 외 다수, "Self-sustained Sequence Replication (3SR): An Isothermal Transcription-based Amplification System Alternative to PCR", PCR Methods and Applications 1:25-33, 1991; 및 Walker G.T. 외 다수, "Strand Displacement Amplification-an Isothermal in vitro DNA Amplification Technique", Nucleic Acid Research 20:1691-1696, 1992 참조]. 이러한 방법에서, 시료 중에 존재하는 핵산은 프로브와 접촉하고, 증폭 반응이 진행되며, 그 결과 증폭 생성물이 검출된다. 증폭 반응 생성물은 반응 생성물에 대해서 겔 전기 영동을 수행하고, 이 겔을 개입 제제 예를 들어, 브롬화아티딘으로 염색함으로써 검출될 수 있다. 대안적으로, 상기 프로브 중 하나 이상은 방사성 동위 원소로 표지화될 수 있으며, 방사성 증폭 반응 생성물의 존재는 겔 전기 영동 이후 방사선 촬영법에 의해 검출될 수 있다.

- [0271] 본 발명의 서열 말단부에 인접하여 존재하는 서열로부터 유래된 프로브는 또한 염색체 위킹법(chromosome walking procedures)에 사용되어, 본 발명의 서열에 인접하여 위치하고 있는 게놈 서열을 함유하는 클론을 동정할 수 있다. 이러한 방법을 통하여, 숙주 유기체로부터 부가의 단백질을 암호화하는 유전자를 분리할 수 있다.
- [0272] 하나의 측면에서, 본 발명의 분리된 또는 재조합 핵산, 이에 상보성인 서열, 또는 본 발명의 서열 중 하나의 적어도 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400 또는 500 개 이상의 연속된 염기 서열을 포함하는 이의 단편, 또는 이에 상보성인 서열은 프로브로서 사용되어, 관련된 핵산을 동정 및 분리할 수 있다. 몇몇 측면에서, 관련된 핵산은 핵산이 분리된 유기체 이외의 유기체로부터 유래하는 cDNA 또는 게놈형 DNA일 수 있다. 예를 들어, 기타 유기체는 관련된 유기체일 수 있다. 이러한 방법에서, 핵산 시료는 프로브가 관련 서열과 특이적으로 혼성화할 수 있는 조건 하에서 이 프로브와 접촉된다. 이후, 이 프로브와 관련 유기체로부터 유래하는 핵산의 혼성화 여부는 전술한 방법들 중 임의의 방법을 사용하여 확인된다.
- [0273] 검출 가능한 프로브에 혼성화되는 핵산 예를 들어, cDNA 또는 게놈형 DNA를 동정하는데에 사용된 혼성화 조건의 엄중도를 변화시킴으로써, 프로브에 대한 상동성 수준이 상이한 핵산은 동정 및 분리될 수 있다. 엄중도는 프로브의 용융 온도 이하의 여러 온도에서 혼성화를 진행시킴으로써 달라질 수 있다. 용융 온도 T_m 이란, (특정 이온세기 및 pH 하에서) 표적 서열 중 50%가 완전히 상보성인 프로브에 혼성화되는 온도이다. 매우 엄중한 조건은 특정 프로브에 대한 T_m 과 동일하거나, 또는 이보다 약 5°C 낮은 온도로서 선택된다. 프로브의 용융 온도는 다음의 공식을 이용하여 계산될 수 있다:
- [0274] 길이 14~70 뉴클레오티드인 프로브에 있어서, 용융 온도(T_m)는 다음의 식을 이용하여 계산된다: $T_m = 81.5 + 16.6(\log[Na^+]) + 0.41(G+C \text{ 비율}) - (600/N)[\text{식 중, } N \text{은 프로브 길이임}]$.
- [0275] 만일 혼성화가 포름아미드를 함유하는 용액 중에서 실행되면, 용융 온도는 다음의 등식을 이용하여 계산될 수 있다: $T_m = 81.5 + 16.6(\log[Na^+]) + 0.41(G+C \text{ 비율}) - (0.63\% \text{ 포름아미드}) - (600/N)[\text{식 중, } N \text{은 프로브 길이임}]$.
- [0276] 예비 혼성화는 6×SSC, 5×덴하르트 시약, 0.5% SDS, 100 μg의 변성 단편화 연어 정자 DNA 또는 6×SSC, 5×덴하르트 시약, 0.5% SDS, 100 μg의 변성 단편화 연어 정자 DNA, 50% 포름아미드 중에서 수행될 수 있다. SSC와 덴하르트 용액에 대한 공식은 문헌[Sambrook의 다수, 상동]에 개시되어 있다.
- [0277] 하나의 측면에서, 혼성화는 상기 나열한 예비 혼성화 용액에 검출 가능한 프로브를 첨가함으로써 수행된다. 프로브가 이중 가닥 DNA를 포함하는 경우, 이 DNA는 혼성화 용액에 첨가되기 이전에 변성된다. 하나의 측면에서, 상기 필터는 이 프로브가 그것과 상보성인 서열 또는 이의 상동성 서열을 함유하는 cDNA 또는 게놈형 DNA에 혼성화되기에 충분한 시간 동안, 혼성화 용액과 접촉된다. 길이 200 뉴클레오티드 이상인 프로브에 있어서, 혼성화는 T_m 보다 낮은 15~25°C에서 수행될 수 있다. 더욱 짧은 프로브 예를 들어, 올리고뉴클레오티드 프로브에 있어서, 혼성화는 T_m 보다 낮은 5~10°C에서 수행될 수 있다. 하나의 측면에서, 6×SSC에서의 혼성화의 경우, 혼성화는 약 68°C에서 수행된다. 일반적으로, 50% 포름아미드 함유 용액에서의 혼성화의 경우, 혼성화는 약 42°C에서 수행된다.
- [0278] 셀룰라제 효소의 발현 억제
- [0279] 본 발명은 본 발명의 핵산 예를 들어, 셀룰라제 효소-암호화 핵산 예를 들어, 안티센스, siRNA, miRNA, 리보자임을 포함하는 핵산에 상보성인(예를 들어, 안티센스 서열인) 핵산을 제공한다. 안티센스 서열을 포함하는 본 발명의 핵산은 셀룰라제 효소-암호화 유전자의 이동, 스플라이싱 또는 전사를 억제할 수 있다. 이러한 억제는 게놈형 DNA 또는 메신저 RNA를 표적화함으로써 실현될 수 있다. 표적화된 핵산의 전사 또는 기능은 예를 들어, 혼성화 및/또는 절단에 의해 억제될 수 있다. 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 생산 또는 작용을 방지하거나 억제하는 경우 중 어느 하나의 경우에, 본 발명에 의해 제공되는 억제제의 대표적인 군으로서는 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 유전자 또는 메세지 중 어느 하나와 결합할 수 있는 올리고뉴클레오티드를 포함한다. 결합은 서열 특이적 혼성화를 통하여 이루어질 수 있다. 기타 유용한 억제제 군으로서는 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 메세지를 불활성화시키거나 또는 절단할 수 있는 올리고뉴클레오티드를 포함한다. 상기 올리고뉴클레오티드는 이러한 절단을 유도하는 효소 활성 예를 들어, 리보자임 효소 활성을 가질 수 있다. 상기 올리고뉴클레오티드는 상보성 핵산을 절단할 수 있는 조성물 또는 효소에 컨쥬게이트화될 수 있거나 또는 화학적으로 변형될 수 있다. 다수의 상이한 올리고뉴클레오티드 풀은 목적으로 하는 활성을 보유하는 올리고뉴클레오티드인지 여부

에 대해 스크리닝될 수 있다. 그러므로, 본 발명은 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 발현을 핵산 및/또는 단백질 수준으로 억제하기 위한, 다양한 조성물 예를 들어, 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 서열과, 본 발명의 항-셀룰라제 항체 예를 들어, 항-엔도글루카나제, 항-셀로바이오히이드롤라제 및/또는 항-베타-글루코시다제 항체 및 안티센스, siRNA, miRNA 및 리보자임을 포함한다.

[0280] 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 발현의 억제는 다양한 산업 분야에서 이용될 수 있다. 예를 들어, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 발현을 억제하면, 부패 속도를 늦추거나 또는 예방할 수 있다. 하나의 측면에서, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 발현 및/또는 활성을 억제하는 본 발명의 조성물을 사용함에 있어서 예를 들어, 항체, 안티센스 올리고뉴클레오티드, 리보자임, siRNA 및 miRNA는 부패의 속도를 늦추거나 또는 예방하는데에 사용된다. 그러므로, 하나의 측면에서, 본 발명은 식물 또는 식물의 생산물(예를 들어, 곡류, 알곡, 과일, 종자, 뿌리 및 잎 등)에 본 발명의 항체, 안티센스 올리고뉴클레오티드, 리보자임, siRNA 및 miRNA를 가하여 부패를 늦추거나 또는 예방하는 단계를 포함하는 방법과, 본 발명의 항체, 안티센스 올리고뉴클레오티드, 리보자임, siRNA 및 mRNA를 포함하는 부패 지연 또는 예방용 조성물을 제공한다. 이러한 조성물은 또한 식물(예를 들어, 트랜스제닉 식물) 또는 기타 유기체(예를 들어, 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 유전자로 형질전환된 박테리아 또는 기타 미생물)에 의해 발현될 수도 있다.

[0281] 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 발현을 억제하기 위한 본 발명의 조성물(예를 들어, 안티센스, iRNA, 리보자임 및 항체)은 약학 조성물 예를 들어, 항-병원체 제제로서 사용될 수 있거나, 또는 기타 치료법 예를 들어, 살모넬라(*Salmonella*)에 대한 항-미생물 제제로서 사용될 수도 있다.

[0282] *안티센스 올리고뉴클레오티드*

[0283] 본 발명은 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 메세지와 결합할 수 있는 안티센스 올리고뉴클레오티드를 제공하는데, 하나의 측면에서, 이 올리고뉴클레오티드는 mRNA를 표적화함으로써 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성을 억제할 수 있다. 안티센스 올리고뉴클레오티드를 디자인하는 기법에 관하여는 과학 논문 및 특허 문헌에 잘 기술되어 있으며, 당 업자는 본 발명의 신규 시약을 이용하여 이러한 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 올리고뉴클레오티드를 디자인할 수 있다. 예를 들어, 유효한 안티센스 올리고뉴클레오티드에 대해 스크리닝하기 위한 유전자 워킹(gene walking)/RNA 맵핑 프로토콜에 관하여는 당 업계에 널리 공지되어 있으며, 예를 들어, 유력한 안티센스 서열을 선택할 수 있는 용이하고도 신뢰성이 있는 방법을 제공하기 위해 표준적인 분자학적 기술을 바탕으로 하는, RNA 맵핑 검정법에 관하여 기술되어 있는 문헌[Ho (2000) *Methods Enzymol.* 314:168-183]을 참조하십시오. 또한 문헌[Smith(2000) *Eur. J. Pharm. Sci.* 11:191-198]도 참조하십시오.

[0284] 자연 발생 핵산은 안티센스 올리고뉴클레오티드로서 사용된다. 이 안티센스 올리고뉴클레오티드는 임의의 길이를 가질 수 있는데; 예를 들어, 대안적인 측면에서, 상기 안티센스 올리고뉴클레오티드의 길이는 약 5~100, 약 10~80, 약 15~60 및 약 18~40 잔기일 수 있다. 최적의 길이는 통상의 스크리닝법에 의해 결정될 수 있다. 상기 안티센스 올리고뉴클레오티드는 임의의 농도로 제공될 수 있다. 최적 농도는 통상의 스크리닝법에 의해 측정될 수 있다. 이와 같이 잠재적인 문제점을 안고 있는 다양한 합성, 비 자연 발생 뉴클레오티드와 핵산 유사체가 공지되어 있다. 예를 들어, 비-이온성 주쇄를 포함하는 펩티드 핵산(PNA) 예를 들어, N-(아미노에틸)글리신 단위가 사용될 수 있다. 포스포로티오에이트 결합을 가지는 안티센스 올리고뉴클레오티드도 사용될 수 있다[WO 97/03211; WO 96/39154; Mata (1997) *Toxicol Appl Pharmacol* 144:189-197; *Antisense Therapeutics*, ed. Agrawal (Humana Press, Totowa, NJ., 1996) 참조]. 본 발명에 의해 제공되는 합성 DNA 주쇄의 유사체를 가지는 안티센스 올리고뉴클레오티드는 또한, 전술한 바와 같은 포스포로-디티오에이트, 메틸포스포네이트, 포스포라미데이트, 알킬 포스포트리에스테르, 설파메이트, 3'-티오아세탈, 메틸렌(메틸이미노), 3'-N-카바메이트 및 카르바산모폴리노 핵산을 포함할 수도 있다.

[0285] 임의의 표적 예를 들어, 본 발명의 센스 및 안티센스 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 서열에 대한 적당한 결합 친화성과 특이성을 가지는 특정 올리

고뉴클레오티드에 대하여 신속하게 스크리닝할 수 있는 조합적인 화학적 방법을 사용하여 다수의 올리고뉴클레오티드를 합성할 수 있다[예를 들어, Gold (1995) J. of Biol. Chem. 270:13581-13584 참조].

[0286] 억제 리보자임

[0287] 본 발명은 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 메세지와 결합할 수 있는 리보자임을 제공한다. 상기 리보자임은 예를 들어, mRNA를 표적화함으로써, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성을 억제할 수 있다. 리보자임을 디자인하고, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소-특이적 안티센스 표적화용 서열을 선별하는 기법은 과학 논문 및 특허 문헌에 널리 공지되어 있고, 당 업자는 본 발명의 신규 시약을 이용하여 이러한 리보자임을 디자인할 수 있다. 리보자임은 표적 RNA를 절단하는 RNA의 효소 부위에 매우 인접하여 존재하는 리보자임의 표적 RNA 결합 부분을 통하여, 표적 RNA와 결합함으로써 작용을 한다. 그러므로, 리보자임은 상보성 염기 쌍 형성 과정을 통하여 표적 RNA를 인지하고 이에 결합하며, 일단 올바른 위치에 결합하면, 효소적으로 작용하여 상기 표적 RNA를 절단해서 이를 불활성화시킨다. 이러한 방식으로 표적 RNA를 절단하면, 암호화 서열이 절단될 경우 이 표적 RNA가 암호화된 단백질의 합성을 유도하는 능력이 파괴될 것이다. 리보자임이 그것의 RNA 표적에 결합하여 이를 절단한 후에는, 이 리보자임은 그 RNA로부터 유리되어 새로운 표적에 다시 결합하여 이를 절단할 수 있다.

[0288] 몇몇 경우, 리보자임의 효소적 특성은 치료적 처치를 실시하는데에 필요한 리보자임의 유효 농도를 안티센스 올리고뉴클레오티드의 유효 농도보다 낮출 수 있다는 점에서, 다른 기술 예를 들어, 안티센스 기술[핵산 분자를 단순히 핵산 표적에 결합시켜 이것의 전사, 번역 또는 다른 분자와의 결합 능력을 억제시키는 기술]에 비해 본 발명의 기술을 유리하도록 만들 수 있다. 이와 같은 잠재적인 이점은 리보자임이 효소적으로 작용할 수 있음을 시사한다. 그러므로, 단일의 리보자임 분자는 표적 RNA의 다수의 분자들을 절단할 수 있다. 하나의 측면에서, 리보자임은 매우 특이적인 억제제로서, 이의 억제 특이성은 염기 쌍 형성 결합 기작에 따라서 달라질 뿐만 아니라, 이 리보자임 분자와 결합하는 RNA의 발현을 억제하는 기작에 따라서도 달라진다. 그러므로, 억제는 RNA 표적을 절단함으로써 유발되므로, 특이성은 비-표적화 RNA의 절단율에 대한 표적화 RNA의 절단율의 비율로서 정의된다. 이와 같은 절단 기작은 염기 쌍 형성에 관여하는 요소 이외의 요소들에 따라서 달라진다. 그러므로, 리보자임의 작용 특이성은 동일한 RNA 위치에 안티센스 올리고뉴클레오티드가 결합하는 특이성보다 더욱 클 수 있다.

[0289] 본 발명의 리보자임 예를 들어, 효소적 리보자임 RNA 분자는 해머헤드 모티프(hammerhead motif), 헤어핀 모티프, 간염 델타 바이러스 모티프, 제I군 인트론 모티프 및/또는 RNaseP-유사 RNA의 형태로서, RNA 가이드 서열과 함께 생성될 수 있다. 해머헤드 모티프의 예는 문헌[예를 들어, Rossi (1992) Aids Research and Human Retroviruses 8:183; hairpin motifs by Hampel (1989) Biochemistry 28:4929, 및 Hampel (1990) Nuc. Acids Res. 18:299]에 개시되어 있으며; 간염 델타 바이러스 모티프에 관하여는 문헌[Perrotta (1992) Biochemistry 31:16; the RNaseP motif by Guerrier-Takada (1983) Cell 35:849]에 기술되어 있고; 또한 제I군 인트론에 관하여는 문헌[Cech, 미국 특허 제4,987,071호]에 기술되어 있다. 이와 같은 특정 모티프에 관한 열거는 본 발명의 범위를 한정하고자 하는 것은 아니다. 당 업자는 본 발명의 리보자임 예를 들어, 본 발명의 효소적 RNA 분자는 표적 유전자 RNA 부위 중 하나 이상의 부위와 상보성인 특이적 기질 결합 부위를 가질 수 있음을 알 것이다. 본 발명의 리보자임은 RNA 절단 활성을 이 리보자임에 부여하는 기질 결합 부위 내에 또는 이의 주변에 존재하는 뉴클레오티드 서열을 가질 수 있다.

[0290] RNA 간섭(RNA interference; RNAi)

[0291] 하나의 측면에서, 본 발명은 RNA 억제 분자 즉, 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 서열을 포함하는 "RNAi" 분자를 제공한다. 상기 RNAi 분자는 이중 가닥 RNA(dsRNA) 분자 예를 들어, siRNA 및/또는 miRNA를 포함할 수 있다. 상기 RNAi 분자 예를 들어, siRNA 및/또는 miRNA는 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 유전자의 발현을 억제할 수 있다. 하나의 측면에서, 상기 RNAi 분자 예를 들어, siRNA 및/또는 miRNA는 그 길이가 약 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 또는 25개 이상인 이중체 뉴클레오티드이다. 본 발명은 어떠한 특정 작용 기작에 의해서도 제한되지 않지만, 상기 RNAi는 세포에 도입되어, 유사하거나 동일한 서열 예를 들어, 내인성 mRNA의 단일 가닥 RNA(ssRNA)의 분해를 유도할 수 있다. 세포가 이중 가닥 RNA(dsRNA)에 노출될 때, 상동성 유전자로부터 유래하는 mRNA는 RNA 간섭(RNAi)이라고 불리는 과정에 의해 선택적으로 분해된다. RNAi 뒤에 숨겨져 있을 수 있는 기본적인 기작은 특이적 유전자 서열과 매치되는 이중 가닥

RNA(dsRNA)를 짧은 간섭 RNA(특정 유전자 서열과 매치되는 mRNA의 분해를 촉진함)라 불리는 짧은 조각들로 분해하는 것이다. 하나의 측면에서, 본 발명의 RNAi는 유전자-침묵 치료제에 사용될 수 있다[예를 들어, Shuey (2002) Drug Discov. Today 7: 1040-1046 참조]. 하나의 측면에서, 본 발명은 본 발명의 RNAi 분자 예를 들어, siRNA 및/또는 miRNA를 이용하여 RNA를 선택적으로 분해하는 방법을 제공한다. 이 방법은 시험관 내, 생체 외 또는 생체내에서 실시될 수 있다. 하나의 측면에서, 본 발명의 RNAi 분자는 세포, 기관 또는 동물 내에서 기능 상실 돌연변이(loss-of-function mutation)를 유발시키는데에 사용될 수 있다. RNA를 선택적으로 분해하기 위해, 이 RNAi 분자 예를 들어, siRNA 및/또는 miRNA를 제조 및 사용하는 방법에 관하여는 당 업계에 널리 공지되어 있으며, 예를 들어, 미국 특허 제6,506,559호; 동 제6,511,824호; 동 제6,515,109호; 및 동 제6,489,127호에도 개시되어 있다.

[0292] 핵산의 변형 - 본 발명의 변이체 효소의 제조

[0293] 본 발명은 본 발명의 핵산 변이체 예를 들어, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 암호화하는 핵산 변이체를 생산하는 방법을 제공한다. 이러한 방법은 주형 핵산에 의해 암호화되는, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 활성 또는 안정성과 상이하거나 변형된 활성 또는 안정성을 가지는 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 생산하기 위하여, 반복 수행될 수 있거나, 또는 다양하게 병행하여 실시될 수 있다. 이러한 방법은 또한 예를 들어, 유전자/메세지 발현, 메세지 번역 또는 메세지 안정성을 변화시키기 위하여, 반복 수행될 수 있거나, 또는 다양하게 병행하여 실시될 수 있다. 다른 측면에서, 세포의 유전자 조성은 예를 들어, 상동성 유전자를 세포 외에서 변형시킨 후, 이를 세포에 다시 삽입함으로써 변화될 수 있다.

[0294] 예를 들어, 하나의 측면에서, 본 발명은 서열 번호 163의 하나 이상의 뉴클레오티드 염기의 잔기를 변형시킨 서열을 가지는 분리된 또는 재조합 핵산을 제공하는데, 여기서 상기 변형은 다음과 같은 변화들 중 하나 이상을 포함한다: 265~267번 위치 중 임의의 위치에 존재하는 뉴클레오티드가 변경되어 CGT, CGC, CGA, CGG, AGA 또는 AGG로 되는 것; 307~309번 위치 중 임의의 위치에 존재하는 뉴클레오티드가 변경되어 GGT, GGC, GGA 또는 GGG로 되는 것; 328~330번 위치 중 임의의 위치에 존재하는 뉴클레오티드가 변경되어 GGT, GGC, GGA 또는 GGG로 되는 것; 340~342번 위치 중 임의의 위치에 존재하는 뉴클레오티드가 변경되어 TTA, TTG, CTT, CTC, CTA 또는 CTG로 되는 것; 469~471번 위치 중 임의의 위치에 존재하는 뉴클레오티드가 변경되어 TCT, TCC, TCA, TCG, AGT 또는 AGC로 되는 것; 1441~1443번 위치 중 임의의 위치에 존재하는 뉴클레오티드가 변경되어 TTT 또는 TTC로 되는 것; 1648~1650번 위치 중 임의의 위치에 존재하는 뉴클레오티드가 변경되어 AAT 또는 AAC로 되는 것; 또는, 1768~1770번 위치 중 임의의 위치에 존재하는 뉴클레오티드가 변경되어 CGT, CGC, CGA, CGG, AGA 또는 AGG로 되는 것. 다른 측면에서, 본 발명은 서열 번호 164의 하나 이상의 아미노산의 잔기를 변형시킨 서열을 가지는 분리된 또는 재조합 폴리펩티드를 제공하는데, 여기서 상기 변형은 다음과 같은 변화들 중 하나 이상을 포함한다: 89번 아미노산 위치의 메티오닌이 아르기닌으로 변경되는 것; 103번 아미노산 위치의 페닐알라닌이 글리신으로 변경되는 것; 110번 아미노산 위치의 프롤린이 글리신으로 변경되는 것; 114번 아미노산 위치의 티로신이 루신으로 변경되는 것; 157번 아미노산 위치의 알라닌이 세린으로 변경되는 것; 481번 아미노산 위치의 트립토판이 페닐알라닌으로 변경되는 것; 550번 아미노산 위치의 프롤린이 아스파라긴으로 변경되는 것; 또는 590번 아미노산 위치의 글리신이 아르기닌으로 변경되는 것.

[0295] 다른 측면에서, 본 발명은 본 발명의 대표적인 서열(예를 들어, 서열 번호 1, 서열 번호 3, 서열 번호 5, 서열 번호 7, 서열 번호 9 및 서열 번호 11의 서열 등)에 뉴클레오티드 잔기 서열 변형이 일어난 서열을 가지는 분리된 또는 재조합 핵산을 제공하는데, 여기서, 상기 변형은 다음과 같은 변화들 중 하나 이상을 포함한다: 서열 번호 163의 265~267번 위치 중 임의의 위치와 대등한 위치의 뉴클레오티드가 변경되어 CGT, CGC, CGA, CGG, AGA 또는 AGG로 되는 것; 서열 번호 163의 307~309번 위치 중 임의의 위치와 대등한 위치의 뉴클레오티드가 변경되어 GGT, GGC, GGA 또는 GGG로 되는 것; 서열 번호 163의 328~330번 위치 중 임의의 위치와 대등한 위치의 뉴클레오티드가 변경되어 GGT, GGC, GGA 또는 GGG로 되는 것; 서열 번호 163의 340~342번 위치 중 임의의 위치와 대등한 위치의 뉴클레오티드가 변경되어 TTA, TTG, CTT, CTC, CTA 또는 CTG로 되는 것; 서열 번호 163의 469~471번 위치 중 임의의 위치와 대등한 위치의 뉴클레오티드가 변경되어 TCT, TCC, TCA, TCG, AGT 또는 AGC로 되는 것; 서열 번호 163의 1441~1443번 위치 중 임의의 위치와 대등한 위치의 뉴클레오티드가 변경되어 TTT 또는 TTC로 되는 것; 서열 번호 163의 1648~1650번 위치 중 임의의 위치와 대등한 위치의 뉴클레오티드가 변경되어 AAT 또는 AAC로 되는 것; 또는 서열 번호 163의 1768~1770번 위치 중 임의의 위치와 대등한 위치의 뉴클레오티드가 변경되어 CGT, CGC, CGA, CGG, AGA 또는 AGG로 되는 것. 다른 측면에서, 본 발명은 본 발명의 임의

의 핵산에 뉴클레오티드 잔기 서열 변형이 일어난 서열을 가지는 분리된 또는 재조합 핵산을 제공하는데, 여기서, 상기 변형은 다음과 같은 변화들 중 하나 이상을 포함한다: 서열 번호 163의 265~267번 위치 중 임의의 위치와 대등한 위치의 뉴클레오티드가 변경되어 CGT, CGC, CGA, CGG, AGA 또는 AGG로 되는 것; 서열 번호 163의 307~309번 위치 중 임의의 위치와 대등한 위치의 뉴클레오티드가 변경되어 GGT, GGC, GGA 또는 GGG로 되는 것; 서열 번호 163의 328~330번 위치 중 임의의 위치와 대등한 위치의 뉴클레오티드가 변경되어 GGT, GGC, GGA 또는 GGG로 되는 것; 서열 번호 163의 340~342번 위치 중 임의의 위치와 대등한 위치의 뉴클레오티드가 변경되어 TTA, TTG, CTT, CTC, CTA 또는 CTG로 되는 것; 서열 번호 163의 469~471번 위치 중 임의의 위치와 대등한 위치의 뉴클레오티드가 변경되어 TCT, TCC, TCA, TCG, AGT 또는 AGC로 되는 것; 서열 번호 163의 1441~1443번 위치 중 임의의 위치와 대등한 위치의 뉴클레오티드가 변경되어 TTT 또는 TTC로 되는 것; 서열 번호 163의 1648~1650번 위치 중 임의의 위치와 대등한 위치의 뉴클레오티드가 변경되어 AAT 또는 AAC로 되는 것; 또는, 서열 번호 163의 1768~1770번 위치 중 임의의 위치와 대등한 위치의 뉴클레오티드가 변경되어 CGT, CGC, CGA, CGG, AGA 또는 AGG로 되는 것.

[0296] 다른 측면에서, 본 발명은 본 발명의 대표적인 서열(예를 들어, 서열 번호 2, 서열 번호 4, 서열 번호 6, 서열 번호 8 및 서열 번호 10의 서열 등)에 아미노산 잔기 변형이 일어난 서열을 가지는 분리된 또는 재조합 핵산을 제공하는데, 여기서, 상기 변형은 다음과 같은 변화들 중 하나 이상을 포함한다: 서열 번호 164의 89번 아미노산 위치의 메티오닌과 대등한 위치의 아미노산이 아르기닌으로 변경되는 것; 서열 번호 164의 103번 아미노산 위치의 페닐알라닌과 대등한 위치의 아미노산이 글리신으로 변경되는 것; 서열 번호 164의 110번 아미노산 위치의 프롤린과 대등한 위치의 아미노산이 글리신으로 변경되는 것; 서열 번호 164의 114번 아미노산 위치의 티로신과 대등한 위치의 아미노산이 루신으로 변경되는 것; 서열 번호 164의 157번 아미노산 위치의 알라닌과 대등한 위치의 아미노산이 세린으로 변경되는 것; 서열 번호 164의 481번 아미노산 위치의 트립토판과 대등한 위치의 아미노산이 페닐알라닌으로 변경되는 것; 서열 번호 164의 550번 아미노산 위치의 프롤린과 대등한 위치의 아미노산이 아스파라긴으로 변경되는 것; 또는 서열 번호 164의 590번 아미노산 위치의 글리신과 대등한 위치의 아미노산이 아르기닌으로 변경되는 것.

[0297] 다른 측면에서, 본 발명은 본 발명의 임의의 폴리펩티드에 아미노산 잔기 변형이 일어난 서열을 가지는 분리된 또는 재조합 폴리펩티드를 제공하는데, 여기서, 상기 변형은 다음과 같은 변화들 중 하나 이상을 포함한다: 서열 번호 164의 89번 아미노산 위치의 메티오닌과 대등한 위치의 아미노산이 아르기닌으로 변경되는 것; 서열 번호 164의 103번 아미노산 위치의 페닐알라닌과 대등한 위치의 아미노산이 글리신으로 변경되는 것; 서열 번호 164의 110번 아미노산 위치의 프롤린과 대등한 위치의 아미노산이 글리신으로 변경되는 것; 서열 번호 164의 114번 아미노산 위치의 티로신과 대등한 위치의 아미노산이 루신으로 변경되는 것; 서열 번호 164의 157번 아미노산 위치의 알라닌과 대등한 위치의 아미노산이 세린으로 변경되는 것; 서열 번호 164의 481번 아미노산 위치의 트립토판과 대등한 위치의 아미노산이 페닐알라닌으로 변경되는 것; 서열 번호 164의 550번 아미노산 위치의 프롤린과 대등한 위치의 아미노산이 아스파라긴으로 변경되는 것; 또는 서열 번호 164의 590번 아미노산 위치의 글리신과 대등한 위치의 아미노산이 아르기닌으로 변경되는 것.

[0298] 본 발명의 핵산은 임의의 수단에 의해 변형될 수 있다. 예를 들어, 랜덤한 방법 또는 확률에 의한 방법, 또는 확률에 의하지 않는 방법 또는 "유도 진화(directed evolution)" 방법에 관하여는 예를 들어, 미국 특허 제 6,361,974호를 참조하시오. 유전자를 랜덤하게 돌연변이시키는 방법은 당 업계에 널리 공지되어 있다[미국 특허 제 5,830,696호 참조]. 예를 들어, 돌연변이원은 유전자를 랜덤하게 돌연변이시키는 데에 사용될 수 있다. 돌연변이원으로서의 예를 들어, 자외 광선 또는 감마 복사선, 또는 화학적 돌연변이원 예를 들어, 미토마이신, 아질산, 광활성화 소라렌(psoralen)을 포함하며, 이들 물질을 단독으로 투여하거나 또는 조합하여 투여함으로써 재조합에 의해 수복될 수 있는 DNA 파괴를 유도한다. 기타 화학적 돌연변이원으로서의 예를 들어, 중아황산나트륨, 아질산, 하이드록실아민, 히드라진 또는 포름산을 포함한다. 기타 돌연변이원으로서의 뉴클레오티드 전구체의 유사체 예를 들어, 니트로소구아니딘, 5-브로모우라실, 2-아미노퓨린 또는 아크리딘이 있다. 상기 제제는 뉴클레오티드 전구체 대신 PCR 반응에 첨가될 수 있으며, 이로 말미암아 서열을 돌연변이시킬 수 있게 된다. 개입제(intercalating agent) 예를 들어, 프로플라빈, 아크리플라빈 및 퀴나크린 등도 사용될 수 있다.

[0299] 분자 생물학 분야의 임의의 기술 예를 들어, 랜덤 PCR 돌연변이 유발법(예를 들어, Rice (1992) Proc. Natl. Acad. Sci. USA 89:5467-5471 참조); 또는, 조합형 다중 카세트 돌연변이 유발법(combinatorial multiple cassette mutagenesis)(예를 들어, Cramer (1995) Biotechniques 18:194-196 참조)이 사용될 수 있다. 대안적으로, 핵산 예를 들어, 유전자는 랜덤한 단편화 즉, "확률에 의한" 단편화 이후에 재조립될 수 있다[예를 들어, 미국 특허 제 6,291,242호; 동 제 6,287,862호; 동 제 6,287,861호; 동 제 5,955,358호; 동 제 5,830,721호; 동 제

5,824,514호; 동 제5,811,238호; 동 제5,605,793호를 참조하시오]. 대안적 측면에서, 변형, 부가 또는 결실은 오류 빈발 PCR, 서플링, 올리고뉴클레오타이드 지정 돌연변이 유발법, 재어셈블리 PCR, 유성 PCR 돌연변이 유발법, 생체내 돌연변이 유발법, 카세트 돌연변이 유발법, 반복적 양상불 돌연변이 유발법, 지수적 양상불 돌연변이 유발법, 위치 특이적 돌연변이 유발법, 유전자 재조합법, 유전자 위치 포화 돌연변이 유발법(GSSM), 합성 결찰 재조합법(SLR), 재조합법, 반복적 서열 재조합법, 포스포티오에이트 변형 DNA 돌연변이 유발법, 우라실 함유 주형 돌연변이 유발법, 갭 형성 이중체 돌연변이 유발법, 포인트 미스매치 수복 돌연변이 유발법, 수복 결함 숙주 균주 돌연변이 유발법, 화학적 돌연변이 유발법, 방사능에 의한 돌연변이 유발법, 결실 돌연변이 유발법, 제한 선별 돌연변이 유발법, 제한 정제 돌연변이 유발법, 인공 유전자 합성법, 양상불 돌연변이 유발법, 키메라 핵산 다량체 생성법, 염색체 포화 돌연변이 유발법(CSM) 및 이들 방법의 조합법을 포함하는 방법에 의해 도입된다.

[0300] 다음과 같은 문헌들에는 다양한 반복적 재조합법 및/또는 본 발명의 방법에 도입될 수 있는 방법에 관하여 기술되어 있다: Stemmer(1999) "Molecular breeding of viruses for targeting and other clinical properties" *Tumor Targeting* 4:1-4; Ness (1999) *Nature Biotechnology* 17:893-896; Chang (1999) "Evolution of a cytokine using DNA family shuffling" *Nature Biotechnology* 17:793-797; Minshull (1999) "Protein evolution by molecular breeding" *Current Opinion in Chemical Biology* 3:284-290; Christians (1999) "Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling" *Nature Biotechnology* 17:259-264; Cramer (1998) "DNA shuffling of a family of genes from diverse species accelerates directed evolution" *Nature* 391:288-291; Cramer (1997) "Molecular evolution of an arsenate detoxification pathway by DNA shuffling," *Nature Biotechnology* 15:436-438; Zhang (1997) "Directed evolution of an effective fucosidase from a galactosidase by DNA shuffling and screening" *Proc. Natl. Acad. Sci. USA* 94:4504-4509; Patten 외 다수 (1997) "Applications of DNA Shuffling to Pharmaceuticals and Vaccines" *Current Opinion in Biotechnology* 8:724-733; Cramer 외 다수 (1996) "Construction and evolution of antibody-phage libraries by DNA shuffling" *Nature Medicine* 2:100-103; Gates 외 다수 (1996) "Affinity selective isolation of ligands from peptide libraries through display on a lac repressor 'headpiece dimer'" *Journal of Molecular Biology* 255:373-386; Stemmer (1996) "Sexual PCR and Assembly PCR" In: *The Encyclopedia of Molecular Biology*. VCH Publishers, New York, pp.447-457; Cramer and Stemmer (1995) "Combinatorial multiple cassette mutagenesis creates all the permutations of mutant and wildtype cassettes" *BioTechniques* 18:194-195; Stemmer 외 다수 (1995) "Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides" *Gene*, 164:49-53; Stemmer (1995) "The Evolution of Molecular Computation" *Science* 270: 1510; Stemmer (1995) "Searching Sequence Space" *Bio/Technology* 13:549-553; Stemmer (1994) "Rapid evolution of a protein in vitro by DNA shuffling" *Nature* 370:389-391; 및 Stemmer (1994) "DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution." *Proc. Natl. Acad. Sci. USA* 91:10747-10751.

[0301] 다양성을 유발시키는 돌연변이 방법으로서의 예를 들어, 위치 지정 돌연변이 유발법[Ling 외 다수 (1997) "Approaches to DNA mutagenesis: an overview" *Anal Biochem.* 254(2): 157-178; Dale 외 다수 (1996) "Oligonucleotide-directed random mutagenesis using the phosphorothioate method" *Methods Mol. Biol.* 57:369-374; Smith (1985) "In vitro mutagenesis" *Ann. Rev. Genet.* 19:423-462; Botstein & Shortle (1985) "Strategies and applications of in vitro mutagenesis" *Science* 229:1193-1201; Carter (1986) "Site-directed mutagenesis" *Biochem. J.* 237:1-7; 및 Kunkel (1987) "The efficiency of oligonucleotide directed mutagenesis" in *Nucleic Acids & Molecular Biology* (Eckstein, F. and Lilley, D. M. J. eds., Springer Verlag, Berlin)]; 우라실 함유 주형을 사용하는 돌연변이 유발법[Kunkel (1985) "Rapid and efficient site-specific mutagenesis without phenotypic selection" *Proc. Natl. Acad. Sci. USA* 82:488-492; Kunkel 외 다수 (1987) "Rapid and efficient site-specific mutagenesis without phenotypic selection" *Methods in Enzymol.* 154, 367-382; 및 Bass 외 다수 (1988) "Mutant Trp repressors with new DNA-binding specificities" *Science* 242:240-245); 올리고뉴클레오타이드 지정 돌연변이 유발법[*Methods in Enzymol.* 100: 468-500 (1983); *Methods in Enzymol.* 154: 329-350 (1987); Zoller (1982) "Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any DNA fragment" *Nucleic Acids Res.* 10:6487-6500; Zoller & Smith (1983) "Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors" *Methods in Enzymol.* 100:468-500; 및 Zoller (1987) *Oligonucleotide-directed mutagenesis: a simple method using two*

oligonucleotide primers and a single-stranded DNA template" *Methods in Enzymol.* 154:329-350]; 포스포로티오에이트 변형 DNA 돌연변이 유발법[Taylor (1985) "The use of phosphorothioate-modified DNA in restriction enzyme reactions to prepare nicked DNA" *Nucl. Acids Res.* 13: 8749-8764; Taylor (1985) "The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA" *Nucl. Acids Res.* 13: 8765-8787 (1985); Nakamaye (1986) "Inhibition of restriction endonuclease Nci I cleavage by phosphorothioate groups and its application to oligonucleotide-directed mutagenesis" *Nucl. Acids Res.* 14: 9679-9698; Sayers (1988) "Y-T Exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis" *Nucl. Acids Res.* 16:791-802; 및 Sayers와 다수 (1988) "Strand specific cleavage of phosphorothioate-containing DNA by reaction with restriction endonucleases in the presence of ethidium bromide" *Nucl. Acids Res.* 16: 803-814]; 갭 형성 이중체 DNA를 사용하는 돌연변이 유발법[Kramer와 다수 (1984) "The gapped duplex DNA approach to oligonucleotide-directed mutation construction" *Nucl. Acids Res.* 12: 9441-9456; Kramer & Fritz (1987) *Methods in Enzymol.* "Oligonucleotide-directed construction of mutations via gapped duplex DNA" 154:350-367; Kramer (1988) "Improved enzymatic in vitro reactions in the gapped duplex DNA approach to oligonucleotide-directed construction of mutations" *Nucl. Acids Res.* 16: 7207; 및 Fritz (1988) "Oligonucleotide-directed construction of mutations: a gapped duplex DNA procedure without enzymatic reactions in vitro" *Nucl. Acids Res.* 16: 6987-6999]을 포함한다.

[0302] 본 발명을 실시하는데에 사용될 수 있는 부가적인 방법으로는 포인트 미스매치 수복[Kramer (1984) "Point Mismatch Repair" *Cell* 38:879-887], 수복 결합 숙주 균주를 사용하는 돌연변이 유발법[Carter와 다수 (1985) "Improved oligonucleotide site-directed mutagenesis using M13 vectors" *Nucl. Acids Res.* 13: 4431-4443; 및 Carter (1987) "Improved oligonucleotide-directed mutagenesis using M1 3 vectors" *Methods in Enzymol.* 154: 382-403], 결실 돌연변이 유발법[Eghtedarzadeh (1986) "Use of oligonucleotides to generate large deletions" *Nucl. Acids Res.* 14: 5115], 제한 선별법 및 제한 정제법[Wells와 다수 (1986) "Importance of hydrogen-bond formation in stabilizing the transition state of subtilisin" *Phil. Trans. R. Soc. Lond. A* 317: 415-423], 전체 유전자 합성에 의한 돌연변이 유발법[Nambiar와 다수 (1984) "Total synthesis and cloning of a gene coding for the ribonuclease S protein" *Science* 223: 1299-1301; Sakamar and Khorana (1988) "Total synthesis and expression of a gene for the α -subunit of bovine rod outer segment guanine nucleotide-binding protein (transducin)" *Nucl. Acids Res.* 14: 6361-6372; Wells와 다수 (1985) "Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites" *Gene* 34:315-323; 및 Grundstrom와 다수 (1985) "Oligonucleotide-directed mutagenesis by microscale "shot-gun gene synthesis" *Nucl. Acids Res.* 13: 3305-3316], 이중 가닥 파괴 수복법[Mandecki (1986); Arnold (1993) "Protein engineering for unusual environments" *Current Opinion in Biotechnology* 4:450-455. "Oligonucleotide-directed double-strand break repair in plasmids of *Escherichia coli*: a method for site-specific mutagenesis" *Proc. Natl. Acad. Sci. USA*, 83:7177-7181]을 포함한다. 전술한 방법에 관한 부가적인 설명은, 다양한 돌연변이 유발법에서 일어날 수 있는 문제점들을 해결하는 유용한 제어 방법에 관하여 기술되어 있는 문헌[*Methods in Enzymology* Volume 154]에서 살펴볼 수 있다.

[0303] 본 발명의 실시예에 사용될 수 있는 방법은 예를 들어, 문헌[미국 특허 제 5,605,793호, Stemmer(Feb. 25, 1997), "Methods for In Vitro Recombination"; 미국 특허 제5,811,238호, Stemmer와 다수 (1998년 9월 22일) "Methods for Generating Polynucleotides having Desired Characteristics by Iterative Selection and Recombination"; 미국 특허 제5,830,721호, Stemmer와 다수(1998년 11월 3일), "DNA Mutagenesis by Random Fragmentation and Reassembly"; 미국 특허 제5,834,252호, Stemmer와 다수(1998년 11월 10일) "End-Complementary Polymerase Reaction"; 미국 특허 제5,837,458호, Minshull와 다수 (1998년 11월 17일), "Methods and Compositions for Cellular and Metabolic Engineering"; WO 95/22625, Stemmer and Cramer, "Mutagenesis by Random Fragmentation and Reassembly"; WO 96/33207, Stemmer and Lipschutz "End Complementary Polymerase Chain Reaction"; WO 97/20078, Stemmer and Cramer "Methods for Generating Polynucleotides having Desired Characteristics by Iterative Selection and Recombination"; WO 97/35966, Minshull and Stemmer, "Methods and Compositions for Cellular and Metabolic Engineering"; WO 99/41402, Punnonen와 다수 "Targeting of Genetic Vaccine Vectors"; WO 99/41383, Punnonen와 다수 "Antigen Library Immunization"; WO 99/41369, Punnonen와 다수 "Genetic Vaccine Vector Engineering"; WO 99/41368, Punnonen와 다수 "Optimization of Immunomodulatory Properties of Genetic Vaccines"; EP 752008, Stemmer

and Cramer, "DNA Mutagenesis by Random Fragmentation and Reassembly"; EP 0932670, Stemmer "Evolving Cellular DNA Uptake by Recursive Sequence Recombination"; WO 99/23107, Stemmer와 다수, "Modification of Virus Tropism and Host Range by Viral Genome Shuffling"; WO 99/21979, Apt와 다수, "Human Papillomavirus Vectors"; WO 98/31837, del Cardayre와 다수 "Evolution of Whole Cells and Organisms by Recursive Sequence Recombination"; WO 98/27230, Patten and Stemmer, "Methods and Compositions for Polypeptide Engineering"; WO 98/27230, Stemmer와 다수, "Methods for Optimization of Gene Therapy by Recursive Sequence Shuffling and Selection," WO 00/00632, "Methods for Generating Highly Diverse Libraries," WO 00/09679, "Methods for Obtaining in Vitro Recombined Polynucleotide Sequence Banks and Resulting Sequences," WO 98/42832, Arnold와 다수, "Recombination of Polynucleotide Sequences Using Random or Defined Primers," WO 99/29902, Arnold와 다수, "Method for Creating Polynucleotide and Polypeptide Sequences," WO 98/41653, Vind, "An in Vitro Method for Construction of a DNA Library," WO 98/41622, Borchert와 다수, "Method for Constructing a Library Using DNA Shuffling", 및 WO 98/42727 by Pati and Zarling, "Sequence Alterations using Homologous Recombination."에 기술되어 있다.

[0304] 본 발명을 실시하는데 사용될 수 있는 방법(다수의 다양성 유발 방법에 관한 상세한 설명을 제공함)에 관하여는 예를 들어, 문헌[미국 특허 출원 제(USSN) 09/407,800, "SHUFFLING OF CODON ALTERED GENES", Patten와 다수 1999년 9월 28일 출원; "EVOLUTION OF WHOLE CELLS AND ORGANISMS BY RECURSIVE SEQUENCE RECOMBINATION", del Cardayre와 다수, 미국 특허 제6,379,964호; "OLIGONUCLEOTIDE MEDIATED NUCLEIC ACID RECOMBINATION", Cramer와 다수, 미국 특허 제6,319,714호; 동 제6,368,861호; 동 제6,376,246호; 동 제6,423,542호; 동 제426,224호 및 PCT/US00/01203; "USE OF CODON-VARIED OLIGONUCLEOTIDE SYNTHESIS FOR SYNTHETIC SHUFFLING", Welch와 다수, 미국 특허 제6,436,675호; "METHODS FOR MAKING CHARACTER STRINGS, POLYNUCLEOTIDES & POLYPEPTIDES HAVING DESIRED CHARACTERISTICS", Selifonov와 다수, 2000년 1월 18일 출원, (PCT/US00/01202), 및 예를 들어, "METHODS FOR MAKING CHARACTER STRINGS, POLYNUCLEOTIDES & POLYPEPTIDES HAVING DESIRED CHARACTERISTICS", Selifonov와 다수, (2000년 7월 18일 출원)(미국 특허 출원 번호: 09/618,579); "METHODS OF POPULATING DATA STRUCTURES FOR USE IN EVOLUTIONARY SIMULATIONS", Selifonov 및 Stemmer, 2000년 1월 18일 출원(PCT/US00/01138); 및 "SINGLE-STRANDED NUCLEIC ACID TEMPLATE-MEDIATED RECOMBINATION AND NUCLEIC ACID FRAGMENT ISOLATION", Affholter(2000년 9월 16일 출원)(미국 특허 출원 번호: 09/656,549); 및 미국 특허 제6,177,263호; 동 제6,153,410호]에 기술되어 있다.

[0305] "확률에 의하지 않는" 즉, "유도 진화" 방법으로서의 예를 들어, 포화 돌연변이 유발법 예를 들어, 유전자 위치 포화 돌연변이 유발법(GSSM), 합성 결찰 재조립법(SLR) 또는 이들 방법의 조합법을 포함하며, 상기 방법에 의해서 본 발명의 핵산을 변형시켜, 신규의 특성 또는 변형된 특성(예를 들어, 고도로 산성 또는 알칼리성인 조건 하, 및 고온 또는 저온에서의 활성)을 가지는 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 생산한다. 변형된 핵산에 의해 암호화되는 폴리펩티드는, 글루칸 가수분해 활성 또는 기타 활성에 대해 시험하기에 앞서서, 활성에 대해 스크리닝될 수 있다. 임의의 시험 방식 또는 방법은 예를 들어, 모세관 어레이 플랫폼을 사용하여 실시될 수 있다. 예를 들어, 미국 특허 제6,361,974호; 동 제6,280,926호; 및 동 제5,939,250호를 참조하시오.

[0306] 유전자 위치 포화 돌연변이 유발법 또는 GSSM

[0307] 본 발명은 또한 본원 및 미국 특허 제6,171,820호 및 동 제6,579,258호에 기술된 바와 같이, 유전자 위치 포화 돌연변이 유발법 또는 GSSM을 이용하여 효소를 제조하는 방법을 제공한다. 하나의 측면에서, 축퇴성 N,N,G/T 서열을 함유하는 코돈 프라이머는 점 돌연변이를 본 발명의 폴리뉴클레오티드 예를 들어, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 또는 항체에 도입하여, 단일 아미노산 치환부의 전체 영역이 각각의 아미노산 위치 예를 들어, 변형되도록 표적화된 리간드 결합 위치 또는 효소 활성 부위 내 아미노산 잔기에 나타내어지는 일군의 자(progeny) 폴리펩티드를 생산하는데에 사용된다. 이러한 올리고뉴클레오티드는 연속적인 제1의 상동성 서열, 축퇴성 N,N,G/T 서열을 포함할 수 있으며, 경우에 따라 제2의 상동성 서열을 포함할 수 있다. 이러한 올리고뉴클레오티드를 사용하여 생성된 하류 자 번역 생산물은 폴리펩티드 사슬을 따라서 존재하는 각각의 아미노산 위치에 일어날 수 있는 모든 아미노산 변화를 포함하는데, 그 이유는 축퇴성 N,N,G/T 서열이 20개 아미노산 전부에 대한 코돈을 포함하기 때문일 것이다. 하나의 측면에서, 이와 같이 (예를 들어, 하나의 축퇴성 N,N,G/T 카세트로 이루어진) 축퇴성 올리고뉴클레오티드 중 어느 하나는 모(parent) 폴리뉴클레오티드 주형 중에 존재하는 각각의 기원 코돈을 전 범위에 걸쳐 코돈 치환시키는 데에 사용된다. 다른 측면에서, 2개 이상의 축퇴성 카세트를 사용하는데, 이때 상기 카세트 중 어느 하나는 동

일한 올리고뉴클레오타이드 또는 동일하지 않은 올리고뉴클레오타이드 내에서 모 폴리뉴클레오타이드 주형 중 2개 이상의 기원 코돈을 전 범위에 걸쳐 코돈 치환시키는데에 사용된다. 예를 들어, 하나 이상의 N,N,G/T 서열은 하나의 올리고뉴클레오타이드 중에 포함되며, 그 결과 하나 이상의 위치에 아미노산 돌연변이를 도입할 수 있게 된다. 이와 같은 N,N,G/T 서열의 다수성(plurality)은 직접적으로 연결되어 나타날 수 있거나, 또는 하나 이상의 부가 뉴클레오타이드 서열(들)에 의해 분리되어 나타날 수도 있다. 다른 측면에서, 부가 및 결실을 도입하는데에 실용적인(serviceable) 올리고뉴클레오타이드는 단독으로 사용될 수 있거나, 아니면 N,N,G/T 서열을 포함하는 코돈과 함께 사용되어, 아미노산 부가, 결실 및/또는 치환을 경우에 따라 조합하여 도입하거나 순열(permutation) 방식으로 도입할 수 있다.

[0308] 하나의 측면에서, 연속적인 N,N,G/T 삼중 코돈 즉, 축퇴성(N,N,G/T)_n 서열을 함유하는 올리고뉴클레오타이드를 사용하여 2개 이상의 연속적 아미노산 위치를 동시에 돌연변이시킨다. 다른 측면에서, N,N,G/T 서열보다 축퇴성이 떨어지는 축퇴성 카세트를 사용한다. 예를 들어, 몇몇 경우에 있어서(예를 들어, 올리고뉴클레오타이드 내에서) N 하나만으로 이루어진 축퇴성 삼중 코돈 서열을 사용하는 것이 바람직할 수 있는데, 이때 상기 N은 삼중 코돈의 첫 번째, 두 번째 또는 세 번째 위치에 존재할 수 있다. 기타 염기들을 조합하여 포함하거나 순열 방식으로 포함하는 기타 임의의 염기들은 삼중 코돈의 나머지 두 위치에 사용될 수 있다. 대안적으로, 몇몇 경우에 있어서, (예를 들어, 올리고뉴클레오타이드 내에서) 축퇴성 N, N, N 삼사용하는 것이 바람직할 수 있다.

[0309] 하나의 측면에서, 축퇴성 삼중 코돈(예를 들어, N,N,G/T 삼중 코돈)을 사용하면, 전 범위의 생성 가능한 천연 아미노산(전체 20개의 아미노산에 포함)을 폴리펩티드 내에 존재하는 각각의 모든 아미노산 위치로 체계적이고 용이하게 치환시킬 수 있다[대안적 측면에서, 이러한 방법은 아미노산 잔기 또는 코돈 위치마다 일어날 수 있는 거의 모든 치환을 발생시킨다]. 예를 들어, 100개의 아미노산 폴리펩티드에 있어서, 2000개의 별개 중(즉, 위치 당 올 수 있는 20개의 아미노산 × 100개의 아미노산 위치)이 생성될 수 있다. 축퇴성 N,N,G/T 삼중 코돈을 함유하는 올리고뉴클레오타이드 군 또는 올리고뉴클레오타이드를 사용할 경우에는, 32개의 개별 서열이 20개의 천연 아미노산 전부를 암호화할 수 있다. 그러므로, 이와 같은 하나 이상의 올리고뉴클레오타이드를 사용하여 모 폴리뉴클레오타이드 서열에 대해 포화 돌연변이 유발법이 진행되고 있는 반응 용기 중에서는, 20개의 개별 폴리펩티드를 암호화하는 32개의 개별 자 폴리뉴클레오타이드가 생성된다. 이와는 대조적으로, 위치 지정 돌연변이 유발법에서 비-축퇴성 올리고뉴클레오타이드를 사용하면, 반응 용기당 하나의 자 폴리펩티드만이 생성된다. 비축퇴성 올리고뉴클레오타이드는 개시된 축퇴성 프라이머와 함께 사용될 수 있는데; 예를 들어, 비축퇴성 올리고뉴클레오타이드는 작용중인 폴리뉴클레오타이드 내에서 특정의 점 돌연변이를 유발시키는데에 사용될 수 있다. 이로써, 특이적 침묵 점 돌연변이, 상응하는 아미노산에 변화를 일으키는 점 돌연변이, 그리고 종결 코돈을 생성하고 그에 따라서 폴리펩티드 단편의 발현이 결정되는 점 돌연변이를 유발하는 하나의 수단을 제공한다.

[0310] 하나의 측면에서, 각각의 포화 돌연변이 유발 반응 용기는 20개 이상의 자 폴리펩티드(예를 들어, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소) 분자를 암호화하는 폴리뉴클레오타이드를 포함하므로, 20개의 천연 아미노산 모두는 모 폴리뉴클레오타이드 내에서 돌연변이 유발된 코돈의 위치에 해당하는 하나의 특정 아미노산 위치에 존재하게 된다[기타 측면은 전부 20개 미만의 천연 아미노산 조합을 이용함]. 각각의 포화 돌연변이 유발 반응 용기로부터 생성되는 32배 축퇴성인 자 폴리펩티드는 클론 증폭될 수 있으며(예를 들어, 발현 벡터를 사용하여 적당한 숙주 예를 들어, 이. 콜라이 숙주에 클로닝되어 증폭될 수 있으며), 또한 발현 스크리닝될 수도 있다. (알칼리성 또는 산성 조건 하에서 증가하는 글루칸 가수분해 활성을 모 폴리펩티드의 글루칸 가수분해 활성과 비교하였을 때), 바람직한 변화를 적당히 나타내는 각각의 자 폴리펩티드가 스크리닝에 의해 동정될 때, 상기 자 폴리펩티드는 서열 결정되어, 이것에 포함된, 상응하는 바람직한 아미노산 치환을 동정할 수 있다.

[0311] 하나의 측면에서, 본원에 기술된 바와 같이, 포화 돌연변이 유발법을 이용하여 모 폴리펩티드 내 각각의 아미노산 위치를 모두 돌연변이시키면, 하나 이상의 아미노산 위치에 바람직한 아미노산 변화가 일어남을 확인할 수 있다. 하나 이상의 신규 자 분자(상기와 같은 바람직한 아미노산 치환 모두 또는 일부를 조합하여 포함하는 자 분자)가 생성될 수 있다. 예를 들어, 만일 2개의 특이적인 바람직한 아미노산 변화가 폴리펩티드 내 3개의 아미노산 위치에서 각각 동정되면, 순열에 의하면 각 위치(기원 아미노산과 차이가 없는 위치, 그리고 2개의 바람직한 변화가 각각 발생한 위치)와 3개의 위치에서 치환이 일어날 가능성은 3가지가 된다. 그러므로, 전체 가능성은 3×3×3 또는 27가지이며, 이들 중 7가지는 6개의 단일 점 돌연변이(즉, 3개의 위치에서 각각 2개씩)가 발생하고 그 이외의 위치에는 변이가 발생하지 않는 것으로 확인되었다.

[0312] 또 다른 측면에서, 위치-포화 돌연변이 유발법은 스크리닝함과 동시에, 서플링, 키메라 형성, 재조합법 및 기타 돌연변이 유발 방법과 함께 사용될 수 있다. 본 발명은 포화 돌연변이 유발법과 같은 임의의 돌연변이 유발 방

법을 반복적으로 사용한다. 하나의 구체예에서, 임의의 돌연변이 유발 방법(들)은 스크리닝과 함께 반복적으로 실시된다.

[0313] 본 발명은 또한 적당한 코돈 프라이머(축퇴성 N, N, N 서열 함유)를 사용하여 점 돌연변이를 폴리뉴클레오티드에 도입시켜, 전 범위의 단일 아미노산 치환부가 각각의 아미노산 위치에 제시되는 일군의 자 폴리펩티드를 생성시키는 방법을 제공한다[유전자 위치 포화 돌연변이 유발법(GSSM)]. 사용된 올리고머는 제1의 상동성 서열, 축퇴성 N, N, N 서열을 연속적으로 포함하고, 하나의 측면에서는(반드시 그러한 것은 아니지만), 제2의 상동성 서열을 포함하기도 한다. 이와 같은 올리고머를 사용하여 얻어진 하류 자 번역 생산물은 폴리펩티드 사슬을 따라서 존재하는 각각의 아미노산 위치에 일어날 수 있는 아미노산 변이를 모두 포함하는데, 그 이유는 N, N, N 서열의 축퇴성이 20개 아미노산 전부에 대한 코돈을 포함하기 때문일 것이다.

[0314] 하나의 측면에서, 이러한 축퇴성 올리고머(하나의 축퇴성 N, N, N 카세트로 이루어짐) 중 하나는 모 폴리뉴클레오티드 주형 내 기원 코돈 각각을 전 범위 코돈 치환시키는데에 사용된다. 다른 측면에서, 동일한 올리고머 또는 그렇지 않은 경우, 모 폴리뉴클레오티드 주형 내에 존재하는 2개 이상의 기원 코돈을 전 범위 코돈 치환시키기 위해, 2개 이상의 축퇴성 N, N, N 카세트가 사용된다. 그러므로, 하나 이상의 위치에 아미노산 돌연변이를 도입하기 위해서, 하나 이상의 N, N, N 서열은 하나의 올리고머 내에 포함될 수 있는 것이다. 이와 같은 N, N, N 서열의 다수성은 직접적으로 연결되어 나타날 수 있거나, 또는 하나 이상의 부가 뉴클레오티드 서열(들)에 의해 분리되어 나타날 수도 있다. 다른 측면에서, 부가 및 결실을 도입하는데에 실용적인 올리고머는 단독으로 사용될 수 있거나, 아니면 N, N, N 서열을 포함하는 코돈과 함께 사용되어, 아미노산 부가, 결실 및/또는 치환을 경우에 따라 조합하여 도입하거나 순열 방식으로 도입할 수 있다.

[0315] 하나의 측면에서, 연속적 N, N, N 삼중 코돈 즉, 축퇴성 (N, N, N)_n 서열을 함유하는 올리고머를 사용하여 2개 이상의 연속적 아미노산 위치들을 동시에 돌연변이시킬 수 있다. 다른 측면에서, 본 발명은 이 N, N, N 서열보다 축퇴성이 떨어지는 축퇴성 카세트를 사용한다. 예를 들어, 몇몇 경우에 (예를 들어, 올리고머 내에서) N 하나만으로 이루어진 축퇴성 삼중 코돈 서열을 사용하는 것이 바람직할 수 있는데, 이때 상기 N은 삼중 코돈의 첫 번째, 두 번째 또는 세 번째 위치에 존재할 수 있다. 기타 염기들을 조합하여 포함하거나 순열 방식으로 포함하는 기타 임의의 염기들은 삼중 코돈의 나머지 두 위치에 사용될 수 있다. 대안적으로, 몇몇 경우에 있어서, (예를 들어, 올리고머 내에서) 축퇴성 N, N, N 삼중 코돈 서열, N, N, G/T 또는 N, N, G/C 삼중 코돈 서열을 사용하는 것이 바람직할 수 있다.

[0316] 하나의 측면에서, 축퇴성 삼중 코돈(예를 들어, N,N,G/T 삼중 코돈 또는 N, N, G/C 삼중 코돈 서열)을 사용하면 몇 가지 이유에서 유리하다. 하나의 측면에서, 본 발명은, 전 범위에 생성 가능한 천연 아미노산(전체 20개의 아미노산에 포함)을 폴리펩티드 내에 존재하는 각각의 모든 아미노산 위치로 체계적이고 용이하게 치환시킬 수 있다. 그러므로, 100개의 아미노산 폴리펩티드에 있어서, 2000개의 별개 중(즉, 위치당 올 수 있는 20개의 아미노산 × 100개의 아미노산 위치)이 생성될 수 있다. 축퇴성 N,N,G/T 삼중 코돈 또는 N, N, G/C 삼중 코돈 서열을 함유하는 올리고머를 사용할 경우에는, 32개의 개별 서열이 20개의 아미노산을 암호화할 수 있음을 알 수 있다. 그러므로, 이와 같은 하나의 올리고머를 사용하여 모 폴리뉴클레오티드 서열에 대해 포화 돌연변이 유발법이 진행되고 있는 반응 용기 중에서는, 20개의 개별 폴리펩티드를 암호화하는 32개의 개별 자 폴리뉴클레오티드가 생성된다. 이와는 대조적으로, 위치 지정 돌연변이 유발법에서 비-축퇴성 올리고머를 사용하면, 반응 용기당 하나의 자 폴리펩티드 생성물만이 생성된다.

[0317] 본 발명은 또한 본원에 개시된 축퇴성 프라이머와 함께 사용될 수도 있는 비축퇴성 올리고머를 사용한다. 몇몇 경우, 비축퇴성 올리고머는 작용중인 폴리뉴클레오티드 내에서 특정의 점 돌연변이를 유발시키는데에 사용될 수 있음을 알 수 있다. 이로써, 특이적 침묵 점 돌연변이, 아미노산에 변화를 일으키는 점 돌연변이, 그리고 종결 코돈을 생성하고 그에 따라서 폴리펩티드 단편의 발현이 결정되는 점 돌연변이를 유발하는 하나의 수단을 제공한다.

[0318] 그러므로, 본 발명의 하나의 측면에서, 각각의 포화 돌연변이 유발 반응 용기는 20개 이상의 자 폴리펩티드 분자를 암호화하는 폴리뉴클레오티드를 포함하므로, 20개의 아미노산 모두는 모 폴리뉴클레오티드 내에서 돌연변이 유발된 코돈의 위치에 해당하는 하나의 특정 아미노산 위치에 존재하게 된다. 각각의 포화 돌연변이 유발 반응 용기로부터 생성되는 32배 축퇴성인 자 폴리펩티드는 클론 증폭될 수 있으며(예를 들어, 발현 벡터를 사용하여 적당한 숙주 예를 들어, 이. 콜라이 숙주에 클로닝되어 증폭될 수 있으며), 또한 발현 스크리닝될 수도 있다. (모 폴리펩티드의 경우와 비교하였을 때), 바람직한 변화를 적당히 나타내는 각각의 자 폴리펩티드가 스크리닝에 의해 동정될 때, 상기 자 폴리펩티드는 서열 결정되어, 이것에 포함된, 상응하는 바람직한 아미노산

치환을 동정하도록 서열 결정될 수 있다.

- [0319] 하나의 측면에서, 본원에 기술된 바와 같이, 포화 돌연변이 유발법을 이용하여 모 폴리펩티드 내 각각의 아미노산 위치를 모두 돌연변이시키면, 하나 이상의 아미노산 위치에 바람직한 아미노산 변화가 일어남을 확인할 수 있다. 하나 이상의 신규 자 분자(상기와 같은 바람직한 아미노산 치환 모두 또는 일부를 조합하여 포함하는 자 분자)가 생성될 수 있다. 예를 들어, 만일 2개의 특이적인 아미노산의 바람직한 변화가 폴리펩티드 내 3개의 아미노산 위치에서 각각 동정되면, 순열에 의하면 각 위치(기원 아미노산과 차이가 없는 위치, 그리고 2개의 바람직한 변화가 각각 발생한 위치)와 3개의 위치에서 치환이 일어날 가능성은 3가지가 된다. 그러므로, 전체 가능성은 $3 \times 3 \times 3$ 또는 27가지이며, 이들 중 7가지는 6개의 단일 점 돌연변이(즉, 3개의 위치에서 각각 2개씩)가 발생하고 그 이외의 위치에는 변이가 발생하지 않는 것으로 확인되었다.
- [0320] 본 발명은 부가적인 돌연변이 유발 방법과 함께 실시되는데, 예를 들어, 2개 이상의 관련 폴리뉴클레오티드가 적당한 숙주 세포에 도입되어, 하이브리드 폴리뉴클레오티드가 재조합법 및 반복적 재분류법에 의해 생산된다.
- [0321] 유전자 전 서열에 걸쳐서 돌연변이 유발법을 수행하는 이외에, 본 발명은 폴리뉴클레오티드 서열 내에 존재하는 임의의 수의 염기 각각을 치환하는데에 사용될 수 있는 돌연변이 유발법을 제공하며, 하나의 측면에서, 이러한 돌연변이 유발법에서 돌연변이될 염기의 수는 15~100,000의 모든 정수이다. 그러므로, 분자에 전체적으로 존재하는 모든 위치들을 돌연변이시키는 대신에, 모든 염기 또는 일정 수의 염기(하나의 측면에서, 총 15~100,000개의 하위 세트)를 돌연변이시킬 수 있다. 하나의 측면에서, 개별 뉴클레오티드는 폴리뉴클레오티드 서열에 걸쳐 존재하는 각각의 위치 또는 위치 군들을 돌연변이시키는데에 사용된다. 돌연변이될 3개 위치 군은 코돈일 수 있다. 돌연변이는 이중 카세트(돌연변이 유발 카세트라고도 함)를 함유하는 돌연변이 유발 프라이머를 사용하여 도입될 수 있다. 대표적 카세트는 1~500개의 염기를 가질 수 있다. 이러한 이중 카세트 내에 존재하는 각각의 뉴클레오티드 위치는 N, A, C, G, T, A/C, A/G, A/T, C/G, C/T, G/T, C/G/T, A/G/T, A/C/T, A/C/G 또는 E이며, 여기서, E는 A, C, G 또는 T 이외의 임의의 염기이다(E는 디자이너 올리고머라 칭할 수 있음).
- [0322] 하나의 측면에서, 포화 돌연변이 유발법은 돌연변이될 한정된 폴리뉴클레오티드 서열(여기서, 이 돌연변이될 서열은 하나의 측면에 의하면 길이 약 1~100,000개 염기임) 내 돌연변이 유발 카세트(여기서, 각각의 카세트는 하나의 측면에 의하면 길이 약 1~500개 염기임)의 완전한 세트를 돌연변이 유발시키는 과정을 포함한다. 그러므로, 돌연변이군(1~100개의 돌연변이군)은 돌연변이가 유발될 각 카세트에 도입된다. 하나의 카세트에 도입될 돌연변이 군은, 포화 돌연변이 유발 과정 중 제1 라운드 진행시 제2 카세트에 도입될 돌연변이의 제2 군과 상이하거나 또는 이와 동일할 수 있다. 이러한 군들은 특정 코돈의 결실, 부가, 분류 및 특정 뉴클레오티드 카세트의 분류에 의해 구체화된다.
- [0323] 하나의 측면에서, 돌연변이가 유발될 특정 서열은 전체 유전자, 경로, cDNA, 전체 개방 해독 틀(ORF) 및 전체 프로모터, 인핸서, 억제제/트랜스 활성화 인자(transactivator), 복제 기원, 인트론, 작동 인자 또는 임의의 폴리뉴클레오티드 작용군을 포함한다. 일반적으로, 이러한 목적으로 사용될 "특정 서열"은 15 염기-폴리뉴클레오티드 서열과, 길이가 15~15,000 염기인 폴리뉴클레오티드 서열과 같은 임의의 폴리뉴클레오티드일 수 있다[본 발명은 상기와 같은 수들 사이에 존재하는 모든 정수를 구체적으로 지명함]. 코돈 분류시 고려될 수 있는 사항으로서는 축퇴성 돌연변이 유발 카세트에 의해 암호화되는 아미노산 유형들을 포함한다.
- [0324] 하나의 측면에서, 돌연변이 유발 카세트에 도입될 수 있는 돌연변이들을 분류함에 있어서, 본 발명은 구체적으로 각 위치에서 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 및 20 개의 아미노산을 암호화하는 축퇴성 코돈 치환부(축퇴성 올리고머 사용)와, 이에 의해 암호화되는 폴리펩티드 라이브러리를 제공한다.
- [0325] **합성 결찰 제조법(SLR)**
- [0326] 본 발명은, 신규의 특성 또는 변형된 특성을 가지는 본 발명의 폴리펩티드 예를 들어, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 또는 항체를 생산하기 위하여 확률에 의하지 않는 유전자 변형 시스템("합성 결찰 제조법" 또는 간단히 "SLR", "유도 진화법"이라고 칭함)을 제공한다.
- [0327] SLR은 올리고뉴클레오티드 단편들을 확률에 의하지 않고 서로 결찰시키는 방법이다. 이 방법은 핵산 구성 블록(nucleic acid building block)이 랜덤하게 서플링되거나, 연결되거나 또는 키메라화되지 않고, 확률에 의하지 않고 조립된다는 점에서, 확률에 의한 올리고뉴클레오티드 서플링과는 다르다. 예를 들어, 미국 특허 제 6,773,900호; 동 제 6,740,506호; 동 제 6,713,282호; 동 제 6,635,449호; 동 제 6,605,449호; 및 동 제 6,537,776

호를 참조하시오. 하나의 측면에서, SLR은 다음의 단계들을 포함한다: (a) 상동성 유전자를 암호화하는 서열을 포함하는 주형 폴리뉴클레오타이드를 제공하는 단계; (b) 다수의 구성 블록 폴리뉴클레오타이드를 제공하는 단계로서, 여기서 상기 구성 블록 폴리뉴클레오타이드는 소정의 서열에서 주형 폴리뉴클레오타이드와 교차하여 재조립되도록 디자인될 뿐만 아니라, 이 구성 블록 폴리뉴클레오타이드는 상동성 유전자의 변이체인 서열과 이 변이체 서열에 측접하는 주형 폴리뉴클레오타이드에 상동성인 서열을 포함하는 것인 단계; 그리고 (c) 구성 블록 폴리뉴클레오타이드와 주형 폴리뉴클레오타이드를 결합시켜, 이 구성 블록 폴리뉴클레오타이드가 주형 폴리뉴클레오타이드와 교차 재조립되어, 상동성 유전자 서열 변이가 일어난 폴리뉴클레오타이드를 생성하는 단계.

[0328] SLR은 재배열될 폴리뉴클레오타이드 사이의 상동성 수준이 높은지 여부에 의존하지 않는다. 그러므로, 이 방법은 10^{100} 개 이상의 상이한 키메라로 이루어진 자 분자의 라이브러리(또는 세트)를 확률에 의하지 않고 생산하는데에 사용될 수 있다. SLR은 10^{1000} 개 이상의 상이한 자 키메라로 이루어진 라이브러리를 제조하는데에 사용될 수 있다. 그러므로, 본 발명의 측면들은 디자인에 의해 선택된 전체 조립 순서를 단축시켜 최종 키메라 핵산 분자 세트를 생산하는, 확률에 의하지 않는 방법을 포함한다. 이 방법은, 상호 결합이 가능한 말단부를 가지는 특정 핵산 구성 블록 다수 개를 디자인하는 단계와, 이러한 핵산 구성 블록을 조립하여 조립체의 순서를 전체적으로 디자인할 수 있는 단계를 포함한다.

[0329] 만일, 조립될 핵산 구성 블록의 상호 결합 가능한 말단부가 이 구성 블록을 소정의 순서로 커플링시킬 수 있을 경우, 이 말단부를 순서가 정해진 형태의 조립체와 같은 유형에 "실용적인" 것으로서 간주된다. 그러므로, 핵산 구성 블록이 커플링될 수 있는 조립체의 전체적인 순서는 결합 가능한 말단부를 디자인함으로써 정해진다. 하나 이상의 조립 단계가 진행되면, 상기와 같은 핵산 구성 블록이 커플링될 수 있는 조립체의 전체적인 순서는 또한 조립의 연속적 단계(들)에 의해 결정되기도 한다. 하나의 측면에서, 어닐링된 구성 블록 조각들은 효소 예를 들어, 리가제(예를 들어, T4 DNA 리가제)로 처리되어, 구성 조각들을 공유적으로 결합시킬 수 있게 된다.

[0330] 하나의 측면에서, 올리고뉴클레오타이드 구성 블록은 최종 키메라 폴리뉴클레오타이드의 자 세트를 생산하기 위한 기초 단위로서 사용되는 선조 핵산 서열 주형 세트를 분석함으로써 디자인된다. 그러므로, 이와 같은 모 올리고뉴클레오타이드 주형은 돌연변이 예를 들어, 키메라 형성 또는 서플링될 핵산 구성 블록을 디자인하는 것을 도와주는 서열 정보의 공급원으로서 사용된다. 이러한 방법의 하나의 측면에서, 다수의 모 핵산 주형의 서열은 하나 이상의 분계점을 선택하도록 정렬된다. 상기 분계점은 상동성인 구역에 존재할 수 있으며, 또한 하나 이상의 뉴클레오타이드로 이루어져 있다. 하나의 측면에서, 이와 같은 분계점들은 선조 주형들 중 2개 이상에 의해 공유된다. 그러므로, 상기 분계점은 모 폴리뉴클레오타이드를 재배열하기 위하여, 생성될 올리고뉴클레오타이드 구성 블록의 경계를 유추하는데에 사용될 수 있다. 이와 같이 선조 분자 내에서 확인 및 선택된 분계점은 최종 키메라 자 분자의 조립에 있어서 유력한 키메라 형성 지점으로 사용된다. 분계점은 2개 이상의 모 폴리뉴클레오타이드 서열에 의해 공유되는 상동성 구역(하나 이상의 상동성 뉴클레오타이드 염기로 이루어짐)일 수 있다. 대안적으로, 분계점은 최소한 모 폴리뉴클레오타이드 서열의 절반에 의해 공유되는 상동성 구역일 수 있거나, 또는 모 폴리뉴클레오타이드 서열 중 3분의 2 이상에 의해 공유되는 상동성 구역일 수 있다. 하나의 측면에서, 실용적 분계점은 모 폴리뉴클레오타이드 서열 중 4분의 3 이상에 의해 공유되는 상동성 구역이거나 모 폴리뉴클레오타이드 서열의 대부분에 의해 공유될 수 있다. 하나의 측면에서, 분계점은 모 폴리뉴클레오타이드 서열의 전부에 의해 공유되는 상동성 구역이다.

[0331] 하나의 측면에서, 결합 재조립법은 자손 키메라 폴리뉴클레오타이드의 소모적 라이브러리(exhaustive library)를 생산하기 위하여 철저하게 수행된다. 다시 말하면, 핵산 구성 블록을 모든 가능한 순서로 조합한 결과가 최종 키메라 핵산 분자 세트에 제시된다는 것이다. 이와 동시에, 다른 측면에 의하면, 각 조합체에 있어서의 조립 순서(즉, 각 구성 블록을 각 최종 생성 키메라 핵산 서열의 5'→3'로 조립할 때의 순서)는 전술한 바와 같이 디자인(또는 확률에 의하지 않고 생산)된다. 본 발명이 확률에 의하지 않는다는 특징으로 인하여, 원치 않는 부산물이 생성될 가능성은 상당히 줄어든다.

[0332] 다른 측면에서, 결합 재조립 방법은 체계적으로 수행된다. 예를 들어, 이 방법은 예를 들어, 일일이 체계적으로 스크리닝될 수 있는 구획들에 의해 체계적으로 구분된 자 분자의 라이브러리를 제조하기 위해 진행된다. 다시 말해서, 본 발명은 연속적 단계의 조립 반응을 선택적이고 적절하게 이용하고, 특정의 핵산 구성 블록을 선택적이고 적절하게 사용함으로써, 자 생산물의 특정 세트가 몇몇 반응 용기 내에서 각각 생산되도록 이 세트를 디자인할 수 있다. 이로써, 관찰 및 스크리닝 방법이 체계적으로 수행될 수 있다. 그러므로, 이 방법은 다수의 자 분자를, 더욱 작은 규모의 군에서 체계적으로 관찰할 수 있도록 만들어 준다. 특히 선조 분자 간 상동성 수준이 낮을 때, 이 방법을 통하여 매우 융통성 있고, 철저하며, 체계적이기까지 한 방식으로 키메라를 만들 수 있기

때문에, 이러한 방법은 다수의 자 분자로 이루어진 라이브러리(또는 세트)를 생산해 낼 수 있다. 본 발명의 결합 제조방법이 확률에 의하지 않는다는 특성으로 인하여, 하나의 측면에서 생성된 자 분자는 디자인에 따라서 선택되는 전체적인 조립 순서를 가지는 최종 키메라 핵산 분자 라이브러리를 포함한다. 포화 돌연변이 유발법 및 최적화 유도 진화법은 또한 상이한 자 분자 종을 만드는데에 사용될 수도 있다. 본 발명은 분계점 선택에 관한 한 선택권 및 제어 능, 핵산 구성 블록의 크기 및 수, 그리고 커플링 규모 및 디자인을 제공함을 알 수 있다. 뿐만 아니라, 분자 간 상동성은 본 발명을 실시하는데에 그다지 필요하지 않다. 사실, 분계점은 분자 간 상동성이 거의 없거나 아예 없는 구역에서 선택될 수도 있는 것이다. 예를 들어, 코돈의 느슨함(codon wobble) 즉, 코돈의 축퇴성으로 인하여, 뉴클레오타이드 치환은 이 뉴클레오타이드에 상응하는 선조 주형 내에서 원래부터 암호화되던 아미노산을 변형시키지 않고서도 핵산 구성 블록에 도입될 수 있다. 대안적으로, 코돈은 기원 아미노산의 변형을 지시하도록 변형될 수 있다. 본 발명에 의하면, 이러한 치환은 핵산 구성 블록에 도입되어, 분자 간 상동성 분계점의 출현 빈도를 더욱 늘릴 수 있으며, 그 결과 구성 블록 간에 보다 많은 수의 커플링이 형성될 수 있게 되어, 더욱 많은 수의 자 키메라 분자가 생성될 수 있게 되는 것이다.

[0333] 합성 유전자 제조방법

[0334] 하나의 측면에서, 본 발명은 합성 유전자 제조방법이라고 하는 확률에 의하지 않는 방법 즉, 어느 정도는 확률에 의한 서플링과 관련되어 있지만, 핵산 구성 블록이 랜덤하게 서플링되거나 또는 결합되거나 또는 키메라 형성되지 않고, 오히려 확률에 의존하지 않고서 조립되는 방법을 제공한다. 예를 들어, 미국 특허 제6,537,776호를 참조하시오.

[0335] 합성 유전자 제조방법은 서플링될 폴리뉴클레오타이드 간 상동성 수준이 높은지 여부에 의존하지는 않는다. 본 발명은 10^{100} 개 이상의 상이한 키메라로 이루어진 자 분자 라이브러리(또는 세트)를 확률에 의하지 않는 방식으로 생산하는데에 사용될 수 있다. 합성 유전자 제조방법은 10^{1000} 개 이상의 상이한 자 키메라로 이루어진 라이브러리를 제조하는 경우에도 사용될 수 있을 것이다.

[0336] 그러므로, 하나의 측면에서, 본 발명은 확률에 의하지 않고 완성된 키메라 핵산 분자 세트를 생산하는 방법을 제공하는데, 여기서, 상기 키메라 핵산 분자의 전체적인 조립 순서는 디자인에 의해 선택되며, 이 방법은 상호 결합이 가능한 말단부를 가지는 특정 핵산 구성 블록 다수 개를 디자인하는 단계와, 이러한 핵산 구성 블록을 조립하여 전체적인 조립 순서를 디자인할 수 있는 단계를 포함한다.

[0337] 조립될 핵산 구성 블록의 상호 결합 가능한 말단부가 구성 블록을 소정의 순서로 커플링시킬 수 있을 경우, 이 말단부는 순서가 정해진 형태의 조립체에 대하여 "실용적인" 것으로서 간주된다. 그러므로, 하나의 측면에서, 핵산 구성 블록이 커플링될 수 있는 조립체의 전체적인 순서는 결합 가능한 말단부를 디자인함으로써 정해지며, 만일 하나 이상의 조립 단계가 진행되면, 상기와 같은 핵산 구성 블록이 커플링될 수 있는 전체적인 조립 순서는 또한 조립 단계(들)의 연속적인 순서에 의해 결정되기도 한다. 본 발명의 하나의 측면에서, 어닐링된 구성 블록 조각들은 효소 예를 들어, 리가제(예를 들어, T4 DNA 리가제)로 처리되어, 구성 블록 조각들을 공유적으로 결합시킬 수 있게 된다.

[0338] 다른 측면에서, 핵산 구성 블록은 최종 키메라 핵산 분자의 자 세트를 생산하기 위한 기초 단위로서 사용되는 선조 핵산 주형 세트 서열을 분석함으로써 디자인된다. 이와 같은 선조 핵산 주형은 돌연변이 즉, 키메라 형성 또는 서플링될 핵산 구성 블록을 디자인하는 것을 도와주는 서열 정보의 공급원으로서 사용된다.

[0339] 하나의 구체예에서, 본 발명은 관련 유전자 군과, 이로 인하여 암호화되는 관련 생산물 군의 키메라를 형성한다. 특정 구체예에서, 암호화되는 생산물은 효소이다. 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소는 본원에 개시된 방법에 따라서 돌연변이 유발될 수 있다.

[0340] 그러므로, 본 발명의 하나의 측면에 의하면, 복수 개의 선조 핵산 주형의 서열(예를 들어, 본 발명의 폴리뉴클레오타이드 서열)은 하나 이상의 분계점을 선택하기 위하여 정렬되며, 이때 상기 분계점은 상동성이 있는 구역에 존재한다. 상기 분계점은 생성될 핵산 구성 블록의 경계를 유추하는데에 사용될 수 있다. 그러므로, 선조 분자에서 확인 및 선택된 분계점은 자 분자의 조립에 있어서 유력한 키메라 형성 지점으로 사용될 수 있다.

[0341] 하나의 측면에서, 실용적 분계점은 2개 이상의 선조 주형에 의해 공유되는 상동성 구역(하나 이상의 상동성 뉴클레오타이드 염기로 이루어짐)이지만, 이 분계점은 최소한 모 주형의 절반에 의해 공유되는 상동성 구역일 수 있거나, 또는 모 주형 중 3분의 2 이상에 의해 공유되는 상동성 구역일 수 있거나, 또는 모 주형의 4분의 3 이상

에 의해 공유되는 상동성 구역일 수 있으며, 하나의 측면에 의하면, 선조 주형의 대부분에 의해 공유되는 상동성 구역일 수 있다. 하나의 측면에서, 보다 실용적인 분계점은 모 주형의 전부에 의해 공유되는 상동성 구역이다.

[0342] 하나의 측면에서, 유전자 재조합 방법은 소모적 라이브러리를 제조하기 위하여 철저하게 수행된다. 다시 말하면, 핵산 구성 블록을 모든 가능한 순서로 조합한 결과가 최종 키메라 핵산 분자 세트에 제시된다는 것이다. 이와 동시에, 각 조합체에 있어서의 조립 순서(즉, 각 구성 블록을 각 최종 생성 키메라 핵산 서열의 5'→3'로 조립할 때의 순서)는 디자인(또는 확률에 의하지 않고 생산)된다. 본 발명의 방법이 확률에 의하지 않는다는 특징으로 인하여, 원치 않는 부산물이 생성될 가능성은 상당히 줄어든다.

[0343] 다른 측면에서, 본 방법은 예를 들어, 일일이 체계적으로 스크리닝될 수 있는 구획들에 의해 체계적으로 구분된 라이브러리를 제조하기 위해 진행된다. 다시 말해서, 본 발명은 연속적 단계의 조립 반응을 선택적이고 적절하게 이용하고, 특정의 핵산 구성 블록을 선택적이고 적절하게 사용함으로써, 자 생산물의 특정 세트가 몇몇 반응 용기 내에서 각각 생산되도록 실험에 의해 디자인될 수 있다. 이로써, 관찰 및 스크리닝 방법이 체계적으로 수행될 수 있다. 그러므로, 이 방법은 다수의 자 분자를, 더욱 작은 규모의 군에서 체계적으로 관찰할 수 있도록 만들어 준다.

[0344] 특히 선조 분자 간 상동성 수준이 낮을 때, 이 방법을 통하여 매우 융통성 있고, 철저하며, 체계적이기까지 한 방식으로 키메라를 만들 수 있기 때문에, 본 발명의 방법은 다수의 자 분자로 이루어진 라이브러리(또는 세트)를 생산해 낼 수 있다. 본 발명의 유전자 재조합법의 확률에 의하지 않는다는 특징으로 인하여, 하나의 측면에서 생성된 자 분자는 디자인에 따라서 선택되는 전체적인 조립 순서를 가지는 최종 키메라 핵산 분자 라이브러리를 포함한다. 특정 측면에서, 이와 같이 생산된 라이브러리는 $10^3 \sim 10^{1000}$ 개 이상의 상이한 자 분자 종류로 이루어져 있다.

[0345] 하나의 측면에서, 전술한 바와 같이 생산된 완성 키메라 핵산 분자는 폴리펩티드를 암호화하는 폴리뉴클레오티드로 이루어져 있다. 하나의 측면에 의하면, 이러한 폴리뉴클레오티드는 인공적인 유전자일 수 있다. 다른 측면에 의하면, 이러한 폴리뉴클레오티드는 인공적인 유전자 경로일 수 있다. 본 발명은 본 발명에 의해 형성된 하나 이상의 인공적인 유전자를 인공적 유전자 경로 예를 들어, 진핵 생물 유기체(식물 포함) 내에서 작동 가능한 경로에 도입할 수 있다.

[0346] 다른 구체예에서, 구성 블록이 생성되는 단계의 합성에 의한 특징으로 인하여, 추후에 시험관 내 방법(예를 들어, 돌연변이 유발법) 또는 생체내 방법(예를 들어, 숙주 유기체의 유전자 스플라이싱 능력을 이용하는 방법)에서 제거될 수도 있는, 뉴클레오티드(예를 들어, 코돈, 인트론 또는 조절 서열일 수 있는 하나 이상의 뉴클레오티드)를 디자인 및 도입시킬 수 있는 것이다. 다수의 경우, 이와 같은 뉴클레오티드를 도입시키면, 실용적인 분계점을 생성시킬 수 있다는 중요한 이점 이외에도, 기타 다수의 이유로 인하여 바람직할 수도 있다.

[0347] 그러므로, 다른 측면에 의해서, 본 발명은 인트론을 도입시키는데에 사용될 수 있는 핵산 구성 블록을 제공한다. 그러므로, 본 발명은 본 발명의 인공적 유전자에 도입될 수 있는 기능적 인트론을 제공한다. 본 발명은 또한 본 발명의 인공적 유전자 경로를 도입시킬 수 있는 기능적 인트론을 제공한다. 그러므로, 본 발명은 인공적으로 도입된 하나(또는 그 이상)의 인트론(들)을 함유하는 인공적 유전자인 키메라 폴리뉴클레오티드의 생산한다.

[0348] 본 발명은 또한 인공적으로 도입된 하나(또는 그 이상)의 인트론(들)을 함유하는 인공적 유전자 경로인 키메라 폴리뉴클레오티드의 생산 방법을 제공한다. 하나의 측면에서, 인공적으로 도입된 인트론(들)은 하나 이상의 숙주 세포 내에서, 주로 자연 발생 인트론이 유전자 스플라이싱에서 기능적으로 사용되는 방식으로 유전자 스플라이싱을 수행한다. 본 발명은 또한 재조합 및/또는 스플라이싱을 위해 숙주 유기체에 도입될 인공적 인트론-함유 폴리뉴클레오티드를 생산하는 방법을 제공한다.

[0349] 본 발명에 의해 생산되는 인공적 유전자는 또한 다른 핵산과의 재조합에 있어서의 기질로서 사용될 수도 있다. 이와 유사하게, 본 발명을 이용하여 형성된 인공적 유전자 경로는 또한 다른 핵산과의 재조합에 대한 기질로서 사용될 수도 있다. 하나의 측면에서, 재조합법은 인공적, 인트론-함유 유전자와 핵산 사이의 상동성 구역에 의해 촉진되거나 또는 여기에서 발생하며, 여기서, 상기 상동성 구역은 재조합 파트너로서 사용된다. 하나의 측면에서, 재조합 파트너는 또한 본 발명에 의해 생산되는 핵산 예를 들어, 인공적 유전자 또는 인공적 유전자 경로일 수도 있다. 재조합법은 인공적 유전자 내에 존재하는 하나의(또는 그 이상의) 인공적으로 도입된 인트론(들)에 존재하는 상동성 구역에 의해 촉진되거나 또는 여기에서 발생할 수 있다.

- [0350] 하나의 측면에서, 본 발명의 합성 유전자 제조법 방법은 복수 개의 핵산 구성 블록을 이용하는데, 하나의 측면에서, 이들 블록은 각각 2개의 결합 가능한 말단부를 가진다. 각 핵산 구성 블록 상에 존재하는 상이 2개의 결합 가능한 말단부는 2개의 블런트 말단부로 존재할 수 있거나(즉, 오버행 말단부를 가지지 않을 수 있거나), 또는 하나의 측면에서는, 하나의 블런트 말단부(blunt end)와 하나의 오버행 말단부(overhang end)로 존재할 수 있거나, 하나의 측면에서는 2개의 오버행 말단부로 존재할 수 있다. 하나의 측면에서, 이러한 용도로써 유용한 오버행 말단부는 3'-오버행 말단부 또는 5'-오버행 말단부일 수 있다. 그러므로, 핵산 구성 블록은 3'-오버행 말단부나 5'-오버행 말단부, 또는 2개의 3'-오버행 말단부, 또는 2개의 5'-오버행 말단부를 보유할 수 있다. 핵산 구성 블록이 조립되어 완성된 키메라 핵산 분자를 형성하는 과정의 전체적인 순서는 의도에 따른 실험 디자인에 의해 결정되는 것이지, 랜덤하게 결정되는 것은 아니다.
- [0351] 하나의 측면에서, 핵산 구성 블록은 2개의 단일 가닥 핵산(단일 가닥 올리고머라고도 칭함)을 화학적으로 합성하고, 이 핵산들을 접착시켜, 서로 어닐링되어 이중 가닥 핵산 구성 블록을 형성하게 만듦으로써 생산된다. 이중 가닥 핵산 구성 블록은 다양한 크기를 가질 수 있다. 이러한 구성 블록의 크기는 작거나 클 수 있다. 구성 블록의 대표적인 크기는 1 염기쌍(어떠한 오버행 부분도 포함하지 않음)에서부터 100,000 염기쌍(어떠한 오버행 부분도 포함하지 않음)까지로 다양하다. 기타 대표적인 크기는, 최소 한계치 1~10,000 bp(이 사이의 정수)이고, 최고 한계치 2~100,000bp(이 사이의 정수)이다.
- [0352] 본 발명에 실용적인 이중 가닥 핵산 구성 블록을 생산할 수 있는 다수의 방법이 존재하며; 이러한 방법에 관하여는 당 업계에 공지되어 있고, 당 업자에 의해 용이하게 실시될 수 있다. 하나의 측면에 의하면, 이중 가닥 핵산 구성 블록은, 먼저 2개의 단일 가닥 핵산을 생성시킨 후, 이 핵산을 어닐링시켜 이중 가닥 핵산 구성 블록을 형성함으로써 생산된다. 이중 가닥 핵산 구성 블록 중 2개의 가닥은 오버행 부분을 형성하는 부위로부터 떨어져 존재하는 모든 뉴클레오티드에서 상보성일 수 있으므로; 임의의 오버행 부분으로부터 떨어져 존재하는 뉴클레오티드에는 미스매치가 발생하지 않는다. 다른 측면에 의하면, 이중 가닥 핵산 구성 블록을 이루는 2개의 가닥은 오버행 부분을 형성하는 임의의 부위로부터 떨어져 존재하는 모든 뉴클레오티드보다 상동성 정도가 더욱 낮다. 그러므로, 이러한 측면에 의하면, 이중 가닥 핵산 구성 블록은 코돈 축퇴성을 도입하는데에 사용될 수 있다. 하나의 측면에서, 코돈 축퇴성은 본원에 기술된 위치-포화 돌연변이 유발법을 이용하거나, 하나 이상의 N, N, G/T 카세트를 이용하거나, 아니면 하나 이상의 N, N, N 카세트를 이용하여 도입된다.
- [0353] 본 발명의 생체내 재조합법은 특정 폴리뉴클레오티드 또는 서열의 미지의 하이브리드 또는 대립 형질 풀에 대하여 맵핑 방식으로 수행될 수 있다. 그러나, 특정 폴리뉴클레오티드의 실제 DNA 또는 RNA 서열이 무엇인지는 알 필요가 없다. 혼합 유전자군 내에서의 재조합법을 이용하는 접근 방법은 임의의 유용한 단백질 예를 들어, 본 발명의 셀룰라제 또는 이의 변이체를 생산하는데에 유용할 수 있다. 이러한 접근 방법은 변형된 특이성이나 활성을 보유하는 단백질을 생산하는데에 사용될 수 있다. 이 접근 방법은 또한 하이브리드 핵산 서열 예를 들어, 유전자의 프로모터 부위, 인트론, 엑손, 인핸서 서열, 3' 비번역 부위 또는 5' 비번역 부위를 생산하는데에도 유용할 수 있다. 그러므로, 이 접근 방법은 발현율이 증가한 유전자를 생산하는데에 사용될 수 있다. 상기 접근 방법은 또한 반복적 DNA 서열을 연구하는데에도 유용할 수 있다. 마지막으로, 상기 접근 방법은 본 발명의 리보자임이나 앵타머를 제조하는데에 유용할 수 있다.
- [0354] 하나의 측면에서, 본원에 기술된 발명은 매우 복잡한 선행 서열 예를 들어, DNA, RNA 또는 단백질을 재조합을 통하여 배향 분자 진화(directed molecular evolution)시킬 수 있도록 환원적 재분류법, 재조합법 및 선별법을 반복 수행하는 것에 관한 것이다.
- [0355] *최적화된 유도 진화 시스템*
- [0356] 본 발명은 신규 특성 또는 변형된 특성을 가지는, 본 발명의 폴리펩티드 예를 들어, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 또는 항체를 생산하기 위한 유전자 변형 시스템 즉, 확률에 의하지 않는 유전자 변형 시스템, 소위 "최적화된 유도 진화 시스템"을 제공한다. 하나의 측면에서, 최적화된 유도 진화는 재조합 과정을 통하여 핵산을 배향 분자 진화시킬 수 있도록 환원적 재분류법, 재조합법 및 선별법을 반복 수행하는 것에 관한 것이다.
- [0357] 최적화된 유도 진화를 통하여 진화된 키메라 서열의 다수의 군집을 생산할 수 있는데, 이때 생산된 군집에는 소정의 횟수만큼 교차 현상이 일어난 서열이 상당 수 존재한다. 교차 현상은 서열 변이가 하나의 모 변이체로부터 다른 모 변이체로 발생한 키메라 서열 중 한 지점에서 일어난다. 이러한 지점은 보통 2개의 모 변이체로부터 유래하는 올리고뉴클레오티드가 함께 결합되어 단일 서열을 형성하는 분기점에 존재한다. 이 방법은 올리고뉴클레오티드 서열의 정확한 농도를 계산할 수 있게 하여, 최종 키메라 서열 군집에 원하는 횟수만큼 교차 현상이 일

어날 수 있도록 만든다. 이로써 교차 현상이 소정의 횟수만큼 일어난 키메라 변이체를 더욱 잘 선택할 수 있다.

[0358] 뿐만 아니라, 이 방법은 다른 시스템과는 달리, 다량의 생산 가능한 단백질 변이체 공간을 찾아내는 편리한 수단이 되기도 한다. 이미 기술한 바와 같이, 예를 들어, 반응을 통해 10^{13} 개의 키메라 분자가 생산되면, 이와 같이 다수의 키메라 변이체를 특정 활성화에 대하여 시험하는 것이 매우 어려워진다. 더욱이, 자 군집의 상당 부분에서는 특정 활성을 높은 수준으로 나타낼 것 같지 않은 단백질을 생산하는 교차 현상이 다수 일어날 것이다. 이 방법을 사용함으로써, 키메라 분자의 군집에는 교차 현상이 특정 횟수만큼 일어난 변이체가 다량 생산될 수 있다. 그러므로, 반응을 통해 여전히 10^{13} 개의 키메라 분자가 생산될 수 있다고 하더라도, 추가의 분석을 위해 선택된 각 분자에서는 대부분 예를 들어, 3회의 교차 현상만이 일어날 것이다. 결과로 생성된 자 군집은 교차 현상이 소정 횟수만큼만 일어나도록 편향될 수 있기 때문에, 키메라 분자간의 기능상 다양성에 대한 경계는 줄어든다. 이로써, 기원 모 폴리뉴클레오타이드로부터 유래하는 올리고뉴클레오타이드 중 어느 올리고뉴클레오타이드가 특정 성질에 기여할 수 있는 지를 알아낼 때에 사용될 다수의 변수들을 제공할 수 있게 되는 것이다.

[0359] 키메라 자손 폴리뉴클레오타이드 서열을 생산하는 한 가지 방법은 각 모 서열의 단편이나 일부에 상응하는 올리고뉴클레오타이드를 생산하는 것이다. 하나의 측면에서, 각 올리고뉴클레오타이드는 독특한 중첩 부위를 포함하는데, 그 결과, 올리고뉴클레오타이드는 서로 혼합되어 각 올리고뉴클레오타이드 단편이 올바른 순서로 조립된 신규 변이체가 생성된다. 대안적으로, 본 발명의 방법을 실시하는 절차에 관하여는 미국 특허 제6,773,900호; 동 제6,740,506호; 동 제6,713,282호; 동 제6,635,449호; 동 제6,605,449호; 동 제6,537,776호; 및 동 제6,361,974호에서 살펴볼 수 있다.

[0360] 각각의 모 변이체에 대해 생산된 올리고뉴클레오타이드의 수는 최종적으로 생산된 키메라 분자에서 일어난 교차 반응의 총 수와 관련이 있다. 예를 들어, 고온에서 보다 큰 활성을 나타내는 키메라 변이체를 찾기 위해서는 3개의 모 뉴클레오타이드 서열 변이체가 결합 반응을 진행하도록 제공될 수 있을 것이다. 일례로서, 각 모 변이체를 이루는 각 부분에 상응하는 50개의 올리고뉴클레오타이드 서열로 이루어진 세트가 생산될 수 있다. 그러므로, 결합 제조립 과정 중, 각각의 키메라 서열 내에는 50회 이하의 교차 현상이 일어날 수 있는 것이다. 생산된 키메라 폴리뉴클레오타이드 각각이 각각의 모 변이체(순서가 뒤바뀐 변이체)로부터 유래하는 올리고뉴클레오타이드를 함유하게 될 확률은 매우 낮다. 만일, 각각의 올리고뉴클레오타이드 단편이 동 몰량으로 결합 반응에 존재하게 되면, 동일한 모 폴리뉴클레오타이드로부터 유래하는 올리고뉴클레오타이드내 위치 중 일부는 그 다음 위치에서 결합될 것이므로, 교차 현상은 일어나지 않을 것이다. 만일 각각의 모 폴리뉴클레오타이드로부터 유래하는 각각의 올리고뉴클레오타이드 농도가 이러한 예에서의 임의의 결합 단계 동안 일정하게 유지되면, 동일한 모 변이체로부터 유래하는 올리고뉴클레오타이드가 키메라 서열 내에서 결합되어 교차가 일어나지 않게 될 확률은(모 변이체가 3개 일 경우) 3분의 1로 줄어든다.

[0361] 그러므로, 확률 밀도 함수(probability density function; PDF)는 결합 반응을 구성하는 각 단계 중 각 변이체의 농도, 각 변이체에 상응하는 올리고뉴클레오타이드 수, 모 변이체의 세트 수가 주어졌을 때, 결합 반응을 구성하는 각 단계가 진행되는 동안에 발생할 수 있는 교차 현상이 일어날 군집을 예측하도록 측정될 수 있다. 이 PDF를 측정하는 저변에 깔려있는 통계학과 수학에 관하여는 이하 기술되어 있다. 이 방법을 이용함으로써, 확률 밀도 함수를 계산할 수 있으며, 그 결과 특정 결합 반응으로 인한 소정 수의 교차 현상이 일어나는 키메라 자 군집은 증가하게 된다. 뿐만 아니라, 교차 현상이 일어나는 횟수가 미리 결정될 수 있으며, 이후 결합 반응을 구성하는 각 단계가 진행되는 동안 각각의 모 올리고뉴클레오타이드의 출발 양을 계산하도록 프로그래밍된 시스템을 통하여, 교차 현상이 일어나는 서열의 횟수에 중점을 둔 확률 밀도 함수를 구할 수 있게 되는 것이다. 이러한 방법은 재조합을 통하여 폴리펩티드를 암호화하는 핵산의 배향 분자 진화를 가능하도록 하는 환원적 재분류법, 재조합법 및 선별법을 반복적으로 수행하는 것에 원리를 두고 있다. 이 시스템을 통하여, 다수의 진화된 키메라 서열을 생산할 수 있는데, 여기서, 생산된 키메라 군집에는 소정 횟수의 교차 현상이 발생한 서열이 상당량 존재한다. 교차 현상은 서열 변이가 하나의 모 변이체에서 다른 모 변이체로 발생한 키메라 서열의 한 지점에서 일어난다. 이러한 지점은 보통 2개의 모 변이체에서 유래하는 올리고뉴클레오타이드가 함께 결합되어 단일의 서열을 형성하는 분기점에 존재한다. 이 방법은 올리고뉴클레오타이드 서열의 정확한 농도를 계산할 수 있게 하여, 최종 키메라 서열 군집에 원하는 횟수만큼 교차 현상이 일어날 수 있도록 만든다. 이로써 교차 현상이 소정의 횟수만큼 일어난 키메라 변이체를 더욱 잘 선택할 수 있게 된다.

[0362] 뿐만 아니라, 이 방법은 다른 시스템과는 달리, 다량의 생산 가능한 단백질 변이체 공간을 찾아내는 편리한 수단이 되기도 한다. 이 방법을 사용함으로써, 키메라 분자의 군집에는 교차 현상을 특정 횟수만큼 나타내는 변이체가 다량 생산될 수 있다. 그러므로, 반응을 통해 여전히 10^{13} 개의 키메라 분자가 생산될 수 있다고 하더라도,

추가 분석을 위해 선택된 각 분자에는 대부분 예를 들어, 3회의 교차 현상만이 일어날 것이다. 결과로 생성된 자 군집은 교차 현상이 소정 횟수만큼만 일어나도록 편향될 수 있기 때문에, 키메라 분자 간의 기능상 다양성에 대한 경계는 줄어든다. 이로써, 기원 모 폴리뉴클레오타이드로부터 유래하는 올리고뉴클레오타이드 중 어느 올리고뉴클레오타이드가 특정 성질에 기여할 수 있는 지를 알아낼 때에 사용될 다수의 변수들을 제공할 수 있게 되는 것이다.

[0363] 하나의 측면에서, 키메라 자손 폴리뉴클레오타이드 서열을 생산하는 한 가지 방법은 각 모 서열의 단편이나 일부에 상응하는 올리고뉴클레오타이드를 생산하는 것이다. 하나의 측면에서, 각 올리고뉴클레오타이드는 독특한 중첩 부위를 포함하는데, 그 결과, 올리고뉴클레오타이드는 서로 혼합되어 각 올리고뉴클레오타이드 단편이 올바른 순서로 조립된 신규 변이체가 생성된다. 예를 들어, 미국 특허 제6,773,900호; 동 제6,740,506호; 동 제6,713,282호; 동 제6,635,449호; 동 제6,605,449호; 동 제6,537,776호; 및 동 제6,361,974호를 참조하시오.

[0364] 교차 현상의 측정

[0365] 본 발명의 한 측면은, 원하는 교차 확률 밀도 함수(PDF), 제조된 모 유전자의 수, 그리고 제조된 단편의 수를 입력하는 시스템 및 소프트웨어를 포함한다. 이 프로그램의 출력 결과는, 제조된 유전자를 생산하는 레시피와 이 유전자의 측정 교차 PDF를 측정할 수 있는 "단편 PDF"이다. 하나의 측면에서, 본원에 기술된 방법은 MATLAB™(The Mathworks, Natick, Massachusetts) 프로그래밍 언어 및 기술에 의한 계산 방법과 관련된 개발 환경에서 수행된다.

[0366] 반복적 방법

[0367] 본 발명의 임의의 절차는 반복적으로 진행될 수 있는데, 예를 들어, 변형되었거나 신규 셀룰라제 표현형 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 암호화하는 본 발명의 핵산은 동정, 재분리, 재변형 및 (활성에 대하여) 재 시험될 수 있다. 이러한 방법은 원하는 표현형이 얻어질 때까지 반복적으로 진행될 수 있다. 예를 들어, 전체적인 생화학적 동화 경로 또는 이화 경로 예를 들어, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성은 세포 내로 조작되어 도입될 수 있다.

[0368] 이와 유사하게, 만일 특정 올리고뉴클레오타이드가 원하는 특성(예를 들어, 신규 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 표현형)에 전혀 영향을 미치지 못하는 것으로 확인되면, 이는 제거될 서열을 포함하는 더욱 큰 모 올리고뉴클레오타이드를 합성함으로써 가변 요인으로서 제거될 수 있다. 상기 서열을 더욱 큰 서열 내에 통합하면 임의의 교차 현상이 일어나는 것을 방지할 수 있게 되므로, 자 폴리뉴클레오타이드 내에 존재하는 서열에서는 더 이상 어떠한 변이도 일어나지 않을 것이다. 어느 올리고뉴클레오타이드가 원하는 특성과 가장 관련성이 많은지, 그리고 어느 올리고뉴클레오타이드가 관련성이 없는지 여부를 결정하는 반복적 방법을 통하여, 특정 성질 또는 활성을 제공할 수 있는 가능한 단백질 변이체 모 두를 보다 효과적으로 찾을 수 있다.

[0369] 생체내 서플링

[0370] 다양한 측면에서, 분자의 생체내 서플링은, 본 발명의 폴리펩티드 변이체 예를 들어, 본 발명의 항체 또는 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 등을 제공하는 본 발명의 방법에 사용된다. 생체내 서플링은 다량체를 제조하게 될 세포의 원래 성질을 이용하여 수행될 수 있다. 생체내 제조법이 분자의 다양성에 대한 원래의 주요 접근 방법을 제공하였던 반면에, 유전자 재조합법은 1) 상동성의 인지; 2) 재조합 키아스마를 형성시키는 가닥 절단, 가닥 침입 및 대사 단계; 및 마지막으로 3) 키아스마를 개별 재조합 분자로 분배하는 것을 포함하는, 비교적 복잡한 과정을 포함한다. 키아스마 형성에는 상동성 서열을 인지하는 과정이 필요하다.

[0371] 다른 측면에서, 본 발명은 최소한 제1 폴리뉴클레오타이드와 제2 폴리뉴클레오타이드로부터 하이브리드 폴리뉴클레오타이드를 생산하는 방법을 포함한다. 본 발명은, 부분적으로 서열 동일성인 하나 이상의 부위를 공유하는 제2 폴리뉴클레오타이드와 적어도 제1 폴리뉴클레오타이드(예를 들어, 이들 폴리뉴클레오타이드 중 하나 또는 둘 다, 예를 들어, 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소-암호화 서열)를 적당한 숙주 세포에 도입함으로써 하이브리드 폴리뉴클레오타이드를 생산하는데 사용될 수 있다. 부분적으로 서열 동일성을 가지는 부위는 하이브리드 폴리뉴클레오타이드를 생산하는 서열 재조직을 유도하는 방법을 촉진한다. 본원에 사용된 "하이브리드 폴리뉴클레오타이드"란 용어는, 본 발명의 방법에 의해 생산된 것으로서, 2개 이상의 기원 폴리뉴클레오타이드 서열로부터 유래하는 서열을 함유하는 임의의 뉴클레오타

드 서열을 의미하는 것이다. 이러한 하이브리드 폴리뉴클레오티드는 DNA 분자간 서열 통합을 촉진하는 분자간 재조합 현상으로부터 유래할 수 있다. 뿐만 아니라, 이러한 하이브리드 폴리뉴클레오티드는 DNA 분자 내 뉴클레오티드 서열을 변형시키기 위해서 반복 서열을 이용하는 분자 내 환원적 재분류법으로부터 유래할 수도 있다.

[0372] 하나의 측면에서, 생체내 재분류법은 "분자 간(inter-molecular)" 방법(총칭, "재조합법"이라 칭함)에 중점을 두는 방법으로서, 박테리아의 경우에는, 일반적으로 "RecA-의존성" 현상으로서 간주된다. 본 발명은 서열을 재조합 및 재분류하기 위한 숙주 세포의 재조합 방법, 또는 세포가 환원적 방법을 매개하여 결실에 의해 세포 내에서 준-반복 서열(quasi-repeated sequences)의 복잡성을 감소시키는 능력에 의존할 수 있다. 이와 같은 "환원적 재분류" 방법은 "분자 내", RecA-독립적 과정에 의해 진행된다.

[0373] 본 발명의 다른 측면에서, 신규의 폴리뉴클레오티드는 환원적 재분류법에 의해 생산될 수 있다. 이 방법은 연속적 서열(기원 암호화 서열)을 함유하는 구조물을 생산하는 단계, 이 연속적 서열을 적당한 벡터에 삽입하는 단계와 이 벡터를 적당한 숙주 세포에 도입시키는 단계를 포함한다. 각각의 분자 특성을 재분류하는 방법은 상동성인 부위를 가지는 구조물 내 연속적 서열 사이, 또는 준-반복 단위 사이에서 이루어지는 조합적 방법에 의해 일어난다. 재분류 방법은 반복 서열의 복잡성과 그 정도를 재조합 및/또는 감소시켜, 신규의 분자 종을 생산해 낸다. 다양한 처리 방법을 적용하여 재분류 비율을 증가시킬 수 있다. 이러한 처리법으로서는 자외선, DNA 손상 화학 물질 및/또는 "유전자 불안정성" 수준이 증가한 숙주 세포주를 사용하는 방법을 포함한다. 그러므로, 재분류 방법은 준-반복 서열을 상동성 재조합시키거나, 또는 이의 원래의 특성을 이용하여 이 서열 나름대로의 진화를 유도하는 단계를 포함할 수 있다.

[0374] 반복 서열 또는 "준-반복" 서열은 유전자 불안정성의 한 요인이 된다. 하나의 측면에서, "준-반복부"는 서열의 기원 단위 구조에 제한되지 않는 반복부이다. 준-반복 단위는 구조물 중 서열 어레이 즉, 유사한 서열의 연속적 단위로서 제공될 수 있다. 일단 결찰되면, 연속적 서열들 간 접합부는 근본적으로 눈에 띄지 않게 되며, 결과로 형성된 구조물의 준-반복적 특성은 분자적 수준에서 연속적으로 된다. 세포 내에서의 결실 과정은 준-반복 서열 간에서 작용하는 구조물의 복잡성을 감소시킨다. 이러한 준-반복 단위는 서열 간 차이를 나타낼 수 있는, 주형의 실질적으로 무제한적인 레퍼토리를 제공한다. 그러므로, 하나의 측면에서, 상기 준-반복부를 함유하는 구조물은 결실(및 잠재적으로는 삽입) 현상이 준-반복 단위 내 임의의 위치에서 실제로 일어날 수 있도록 분자의 융통성(molecular elasticity)을 충분히 제공한다.

[0375] 준-반복 서열이 모두 동일한 방향으로 결찰될 때(예를 들어, 5'→3' 또는 그 반대), 세포는 각각의 단위들을 구별할 수 없다. 결과적으로, 환원적 과정이 서열 전체에 걸쳐 진행될 수 있다. 이와는 반대로, 예를 들어, 이 단위들이 5'→3' 보다는 5'→5' 방향으로 제공될 때, 이러한 역전 현상은 인접 단위가 끝나는 부분을 추론할 수 있도록 만들어 주며, 그 결과 결실부가 형성됨으로써 개별 단위의 손실이 촉진될 것이다. 그러므로, 본 발명에 있어서는 상기 서열이 동일한 방향으로 존재하는 것이 바람직하다. 준-반복 서열의 랜덤한 배향은 재분류 효율을 상실시킬 것이지만, 이 서열의 일관성 있는 배향은 그 효율을 최고로 만들어줄 것이다. 그러나, 동일한 방향으로 존재하는 연속적 서열이 소수로 존재할 경우에는, 신규 분자를 효율적으로 회수하는데 있어서 충분한 융통성이 발휘될 수 있을 것이다. 구조물은 준-반복 서열을 동일한 방향으로 배열시켜 그 효율을 더욱 높임으로써 제조될 수 있다.

[0376] 서열은 다음의 사항들을 포함하는 다양한 방법 중 임의의 방법을 사용하여 5'→3' 방향으로 배향시켜 조립될 수 있다:

[0377] a) 단일 가닥으로 제조될 때 배향을 결정할 폴리-A 두부 및 폴리-T 미부를 포함하는 프라이머의 사용. 이는 처음 몇몇 프라이머 염기가 RNA로부터 생성되고, 그 결과 RNaseH로 용이하게 제거됨으로써 이루어짐.

[0378] b) 독특한 제한 절단 위치를 포함하는 프라이머의 사용. 다수의 위치, 독특한 서열 배터리 및 반복 합성 단계 및 결찰 단계가 필요함.

[0379] c) 티올화되며, 엑소뉴클레아제를 사용하여 적당히 꼬리 부분을 형성한 분자를 생산할 수 있는 내부의 몇몇 프라이머 염기들.

[0380] 하나의 측면에서, 상기 재분류된 서열의 회수 방법은 반복 지수(RI)가 감소된 클로닝 벡터를 동정하는 것과 관련되어 있다. 이후, 상기 재분류된 암호화 서열은 증폭법에 의해 회수될 수 있다. 생산물은 다시 클로닝되어 발현된다. RI가 감소된 클로닝 벡터를 회수하는 방법은 다음의 사항들에 의해 영향받을 수 있다:

[0381] 1) 구조물의 복잡성이 감소하였을 때에만 안정하게 유지되는 벡터의 사용.

- [0382] 2) 물리적 방법에 의해 단축된 벡터의 물리적 회수 방법. 이 경우, 클로닝 벡터는 표준적인 절차에 의한 아가로스 겔 또는 저 분자량 컷-오프를 사용하는 컬럼 상에서 크기별 분획화에 의하거나 또는 표준적인 플라스미드 분리법에 의해 회수될 것이다.
- [0383] 3) 삽입부의 크기가 작아졌을 때 선택될 수 있는 방해 유전자를 함유하는 벡터의 회수.
- [0384] 4) 발현 벡터 및 적당한 선별법을 이용하는 직접 선별 기술의 사용.
- [0385] 관련된 유기체로부터 유래하는 암호화 서열(예를 들어, 유전자)은 고도의 상동성을 나타내며, 상당히 다양한 단백질 생산물을 암호화할 수 있다. 이러한 유형의 서열은 준-반복부로서 본 발명에 특히 유용하다. 그러나, 이하에 기술된 예들은 거의 동일한 기원 암호화 서열(준-반복부)의 재분류법에 관하여 기술하고 있는데, 이 방법은 거의 동일한 반복부를 사용하는 것에 한정되는 것은 아니다.
- [0386] 이하의 예는 본 발명의 예시적 방법을 기술하는 것이다. 3개의 독특한 종으로부터 유래하는 핵산 서열(준-반복부)의 암호화에 관하여 기술되어 있다. 각각의 서열은 독특한 특성군을 가지는 단백질을 암호화한다. 각각의 서열은 서열 내 독특한 위치에 존재하는 단일 염기쌍 또는 소수의 염기쌍에 의해 차이가 난다. 준-반복부 서열은 개별적으로 또는 총체적으로 증폭되어 랜덤한 조합체에 결합되며, 그 결과, 결합된 분자 군집에는 가능한 모든 순열과 조합이 적용될 수 있다. 준-반복 단위의 수는 조합 조건에 의해 조절될 수 있다. 구조물 내 준-반복 단위의 평균 수를 반복 지수(repetitive index; RI)라 정의한다.
- [0387] 일단 형성되면, 구조물은 공지된 프로토콜에 따라서 아가로스 겔 상에서 크기별로 분획화되고, 클로닝 벡터에 삽입되며, 적당한 숙주 세포내에 형질 감염될 수 있거나 또는 될 수 없다. 이후, 상기 세포들은 증식하게 되며, 이에 따라서 "환원적 재분류법"이 진행된다. 환원적 재분류법의 효율은 원하는 경우 DNA 손상을 도입함으로써 촉진될 수 있다. "분자 내" 기작에 의해 반복 서열 간에 결실이 발생함으로 인하여 RI가 감소되는지 여부, 또는 "분자 간" 기작을 통한 재조합-유사 현상에 의해 RI가 감소되는지 여부는 중요하지 않다. 최종 결과물은 생성 가능한 모든 조합체에 분자를 재분류시킨 결과물이다.
- [0388] 경우에 따라, 본 방법은 소정의 거대 분자 예를 들어, 단백질성 수용체, 올리고당, 비리온 또는 기타 소정의 화합물 또는 구조물을 이용하여 서플링된 폴의 라이브러리 일원을 스크리닝함으로써, 특정 반응물(예를 들어, 효소의 촉매 도메인)에 결합하거나 상호작용하거나, 또는 특정 반응을 촉진하는 능력을 가지는 각각의 서플링된 라이브러리 일원을 동정하는 단계를 추가로 포함한다.
- [0389] 이러한 라이브러리로부터 동정되는 폴리펩티드는 치료용, 진단용, 연구용 그리고 관련 목적(예를 들어, 촉매, 수용액 등의 삼투압을 증가시키기 위한 용질)으로 사용될 수 있으며/있거나, 하나 이상의 서플링 및/또는 선별의 반복 과정을 추가로 진행시킬 수 있다.
- [0390] 다른 측면에서, 재조합 또는 재분류를 수행하기 이전에 또는 이를 수행하는 동안, 본 발명의 방법에 의해 생산된 폴리뉴클레오티드는 돌연변이를 기원 폴리뉴클레오티드에 도입시키는 것을 촉진하는 제제 또는 방법으로 처리될 수 있다. 이러한 돌연변이의 도입을 통하여, 결과로 생성된 하이브리드 폴리뉴클레오티드와 이로부터 암호화되는 폴리펩티드의 다양성을 증가시킬 것이다. 돌연변이를 촉진시키는 제제 또는 방법으로서의 다음의 것들을 포함할 수 있으나, 이에 한정되는 것은 아니다: (+)-CC-1065, 또는 합성 유사체 예를 들어, (+)-CC-1065-(N3-아데닌) (Sun and Hurley, (1992) 참조); DNA 합성을 억제할 수 있는 N-아세틸화 또는 탈아세틸화 4'-플루오로-4-아미노비페닐 부가물(예를 들어, van de Poll와 다수 (1992) 참조); 또는 DNA 합성을 억제할 수 있는 N-아세틸화 또는 탈아세틸화 4-아미노비페닐 부산물(예를 들어, van de Poll와 다수 (1992), pp. 751-758 참조); DNA 복제를 억제할 수 있는 3가 크롬, 3가 크롬염, 다환형 방향족 탄화수소(PAH) DNA 부산물 예를 들어, 7-브로모메틸-벤즈[α]안트라센 ("BMA"), 트리스(2,3-디브로모프로필)포스페이트("트리스-BP"), 1,2-디브로모-3-클로로프로판("DBCP"), 2-브로모아크롤레인(2BA), 벤조[α]피렌-7,8-디하이드로디올-9-10-에폭시드("BPDE"), 백금(II) 할로젠 염, N-하이드록시-2-아미노-3-메틸이미다조[4,5-f]-퀴놀린("N-하이드록시-IQ") 및 N-하이드록시-2-아미노-1-메틸-6-페닐이미다조[4,5-f]-피리딘("N-하이드록시-PhIP"). PCR 증폭 반응을 늦추거나 또는 정지시키는 대표적인 수단으로서 UV 광선 (+)-CC-1065 및 (+)-CC-1065-(N3-아데닌)을 포함한다. 특히, 추가의 공정을 수행하기 이전에 폴리뉴클레오티드를 포함하는 용액을 가열하는 단계를 포함하는 방법에 의해 방출되거나 또는 제거될 수 있는, 폴리뉴클레오티드 또는 폴리뉴클레오티드 풀로부터 유래하는 DNA 부산물을 포함하는 폴리뉴클레오티드 또는 DNA 부산물도 포함된다.
- [0391] 다른 측면에서, 본 발명은 하이브리드 또는 재-분류된 폴리뉴클레오티드를 생산하는 본 발명에 의한 조건 하에서, 야생형 단백질을 암호화하는 이중 가닥 주형 폴리뉴클레오티드를 포함하는 시료를 처리함으로써 생물 활성

을 보유하는 재조합 단백질을 생산하는 방법에 관한 것이다.

[0392] 서열 변이체의 생산

[0393] 본 발명은 또한 본 발명의 핵산 서열(예를 들어, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 핵산 서열)의 서열 변이체를 생산하는 추가의 방법을 제공한다. 본 발명은 또한 본 발명의 핵산과 폴리펩티드를 이용하여 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 분리하는 추가의 방법을 제공하기도 한다. 하나의 측면에서, 본 발명은 전술한 바와 같이, 임의의 방법 예를 들어, 랜덤한 방법 또는 확률에 의하는 방법, 또는 확률에 의하지 않는 방법 즉, "유도 진화" 방법에 의해 변형될 수 있는, 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 암호화 서열(예를 들어, 유전자, cDNA 또는 메세지)의 변이체를 제공한다.

[0394] 분리된 변이체는 자연 발생될 수 있다. 변이체는 또한 시험관 내에서 생산될 수도 있다. 변이체는 유전자 조작 기술 예를 들어, 위치 배향 돌연변이 유발법, 랜덤한 화학적 돌연변이 유발법, 엑소뉴클레아제 III 결실법과 표준적인 클로닝 기술을 이용하여 생산될 수 있다. 대안적으로, 이러한 변이체, 단편, 유사체 또는 유도체들은 화학적 합성법이나 변형 방법을 이용하여 생산될 수 있다. 변이체를 생산하는 기타의 방법도 당 업자에게 공지되어 있다. 이러한 방법으로서의 천연 분리체로부터 얻은 핵산 서열이 변형되어 산업 분야 또는 실험실에서의 사용 가치를 높인 폴리펩티드를 암호화하는 핵산을 생산하는 방법을 포함한다. 이러한 방법에 있어서, 천연 분리체로부터 얻어지는 서열과 상이한 뉴클레오티드를 하나 이상 가지는 다수의 변이체 서열이 생산되고 특성 규명된다. 이러한 뉴클레오티드 간 차이점은 천연 분리체로부터 유래하는 핵산에 의해 암호화되는 폴리펩티드에 비해 아미노산에 차이가 있을 수 있다.

[0395] 예를 들어, 변이체는 오류 유발 PCR을 이용하여 생산될 수 있다. 오류 유발 PCR에 관한 하나의 측면에서, PCR은 DNA 중합 효소의 복사체 생성 신뢰도가 낮은 조건 하에서 수행되며, 그 결과 PCR 생산물의 전체 길이를 따라서 점 돌연변이가 높은 비율로 발생하게 된다. 오류 유발 PCR에 관하여는 문헌[예를 들어, Leung (1989) Technique 1:11-15 및 Caldwell (1992) PCR Methods Applic. 2:28-33]에 개시되어 있다. 간단히 말해서, 이러한 방법에 있어서, 돌연변이될 핵산은 PCR 프라이머, 반응 완충액, $MgCl_2$, $MnCl_2$, Taq 중합 효소, 그리고 적당한 농도의 dNTP와 혼합되어, PCR 생산물의 전체 길이를 따라서 점 돌연변이가 높은 비율로 발생하게 되는 것이다. 예를 들어, 반응은 돌연변이될 핵산 20 fmole, 각 PCR 프라이머 30pmole, 50mM KCl, 10mM Tris HCl(pH 8.3) 및 0.01% 젤라틴, 7mM $MgCl_2$, 0.5mM $MnCl_2$, 5 유닛의 Taq 중합 효소, 0.2mM dGTP, 0.2mM dATP, 1mM dCTP 및 1mM dTTP를 포함하는 반응 완충액을 사용하여 수행될 수 있다. PCR은 94℃에서 1분 동안, 45℃에서 1분 동안, 그리고 72℃에서 1분 동안의 과정을 거치는 주기를 30회 수행함으로써 진행될 수 있다. 그러나, 이와 같은 매개 변수들은 적당히 바뀔 수도 있음을 이해할 것이다. 돌연변이 유발된 핵산은 적당한 벡터에 클로닝되고, 이 돌연변이 유발된 핵산에 의해 암호화되는 폴리펩티드의 활성이 평가된다.

[0396] 하나의 측면에서, 변이체는 올리고뉴클레오티드 배향 돌연변이 유발법을 통하여 임의의 클로닝된 목적 DNA 내에 위치 특이적 돌연변이를 발생시킴으로써 생산된다. 올리고뉴클레오티드 돌연변이 유발법에 관하여는 예를 들어, 문헌[Reidhaar-Olson (1988) Science 241:53-57]에 개시되어 있다. 간단히 말해서, 이러한 방법에 있어서는, 클로닝될 DNA에 도입될 하나 이상의 돌연변이를 보유하고 있는 다수의 이중 가닥 올리고뉴클레오티드가 합성된 후, 돌연변이를 유발시키기 위해 클로닝된 DNA에 삽입된다. 하나의 측면에서, 돌연변이 유발된 DNA를 함유하는 클론은 회수 및 발현되고, 이 DNA에 의해 암호화되는 폴리펩티드의 활성이 평가된다.

[0397] 변이체를 생산하는 또 다른 방법으로서의 어셈블리 PCR이 있다. 어셈블리 PCR은 작은 DNA 단편의 혼합물로부터 PCR 생산물을 조립하는 과정을 포함한다. 다수의 상이한 PCR 반응은 동일한 바이알 내에서 동시에 발생하는데, 이때 하나의 반응에서 생성된 생산물은 다른 반응의 생산물을 프라이밍한다. 어셈블리 PCR에 관하여는 예를 들어, 미국 특허 제5,965,408호에 개시되어 있다.

[0398] 하나의 측면에서, 유성 PCR 돌연변이 유발법은 본 발명의 변이체를 생산하는 대표적인 방법이다. 하나의 측면에서, 유성 PCR 돌연변이 유발법 유도 상동성 재조합법은, 서열 동일성을 바탕으로 하는 DNA 분자의 랜덤한 단편화 이후, PCR 반응에서의 프라이머 연장에 따른 교차 반응에 관한 고정된 결과로서, 상이하되 매우 관련성이 많은 DNA 서열의 DNA 분자간에서 일어난다(시험관 내). 유성 PCR 돌연변이 유발법에 관하여는 예를 들어, 문헌[Stemmer (1994) Proc. Natl. Acad. Sci. USA 91:10747-10751]에 기술되어 있다. 간단히 말해서, 이러한 방법에 있어서, 재조합될 다수의 핵산은 DNase로 분해되어, 평균 크기가 50~200 뉴클레오티드인 단편이 생산된다.

원하는 평균 크기를 가지는 단편을 정제하고, 이를 PCR 혼합물 중에 재현탁시킨다. PCR은 핵산 단편들 간 재조합을 촉진하는 조건 하에서 수행된다. 예를 들어, PCR은 농도 10~30 ng/ μ l의 정제된 단편들을, 각각의 dNTP 0.2mM, 2.2mM MgCl₂, 50mM KCL, 10mM 트리스 HCl, pH 9.0 및 0.1% 트리톤 X-100으로 이루어진 용액 중에 재현탁함으로써 수행될 수 있다. 반응 혼합물의 100:1 희석물 당 Taq 중합 효소 2.5 유닛을 첨가하고, 다음과 같은 방식으로 PCR을 수행한다: 94℃에서 60초, 94℃에서 30초, 50~55℃에서 30초, 72℃에서 30초(30~45회), 그리고 72℃에서 5분. 그러나, 이러한 매개 변수들은 적당히 변형될 수 있음을 이해할 것이다. 몇몇 측면에서, 올리고뉴클레오티드는 PCR 반응에 포함될 수 있다. 기타 측면에서, DNA 중합 효소 I의 Klenow 단편은 PCR 반응의 제 1 세트에 사용될 수 있으며, Taq 중합 효소는 PCR 반응의 후속 세트에서 사용될 수 있다. 재조합 서열은 분리되고, 이 서열들이 암호화하는 폴리펩티드의 활성이 평가된다.

[0399] 하나의 측면에서, 변이체는 생체내 돌연변이 유발법에 의해 생산된다. 몇몇 측면에서, 목적 서열 내 랜덤 돌연변이는 목적 서열을, DNA 수복 경로 중 하나 이상의 과정에 돌연변이가 발생한 박테리아 균주 예를 들어, 이. 콜라이 균주에서 증식시킴으로써 생산된다. 이와 같은 "돌연변이 유발자(mutator)" 균주는 야생형인 부모 균주의 경우보다 랜덤 돌연변이 발생률이 더 높다. 이러한 균주들 중 하나에서 DNA를 증식시킴으로써 DNA 내에 랜덤한 돌연변이를 영구적으로 발생시킬 수 있다. 생체내 돌연변이 유발법에 사용하기에 적당한 돌연변이 유발자 균주에 관하여는 PCT 공보 WO 91/16427(1991년 10월 31일 발행, 발명의 명칭: "Methods for Phenotype Creation from Multiple Gene Populations")에 개시되어 있다.

[0400] 변이체는 또한 카세트 돌연변이 유발법을 이용하여 생산될 수도 있다. 카세트 돌연변이 유발법에 있어서, 이중가닥 DNA 분자 중 작은 부위는 천연 서열과 상이한 합성 올리고뉴클레오티드 "카세트"로 치환된다. 상기 올리고뉴클레오티드는 종종 완전하고/완전하거나 부분적으로 랜덤화된 천연 서열을 함유한다.

[0401] 반복적 양상불 돌연변이 유발법도 변이체를 생산하는데에 사용될 수 있다. 반복적 양상불 돌연변이 유발법은, 돌연변이체의 일원들이 아미노산 서열에 있어서 차이가 나는, 표현형이 관련된 돌연변이체의 다양한 군집을 생산하기 위해 개발된 단백질 조작용(단백질 돌연변이용) 알고리즘이다. 이 방법은 조합 카세트 돌연변이 유발법의 연속적 라운드를 제어하는 피드백 기작을 이용한다. 반복적 양상불 돌연변이 유발법에 관하여는 예를 들어, 문헌[Arkin (1992) Proc. Natl. Acad. Sci. USA 89:7811-7815]에 개시되어 있다.

[0402] 몇몇 측면에서, 변이체는 지수적 양상불 돌연변이 유발법을 이용하여 생산된다. 지수적 양상불 돌연변이 유발법은 독특하고 기능을 가지는 돌연변이체를 다수 보유하는 조합형 라이브러리를 제조하는 방법으로서, 여기서, 잔기의 소 그룹은 동시에 랜덤화되어, 각각의 변형된 위치에 있는 아미노산을 동정하며, 이때 상기 아미노산은 기능성 단백질을 합성한다. 지수적 양상불 돌연변이 유발법에 관하여는 예를 들어, 문헌[Delegrave (1993) Biotechnology Res. 11:1548-1552]에 개시되어 있다. 랜덤 및 위치 지정 돌연변이 유발법에 관하여는 예를 들어, 문헌[Arnold (1993) Current Opinion in Biotechnology 4:450-455]에 개시되어 있다.

[0403] 몇몇 측면에서, 변이체는 별도의 폴리펩티드를 암호화하는 다수의 핵산 중 일부가 융합되어 키메라 폴리펩티드를 암호화하는 키메라 핵산 서열을 생산하는 서플링 방법에 의해 생산된다[미국 특허 제5,965,408호, 1996년 7월 9일 출원, 발명의 명칭: "Method of DNA Reassembly by Interrupting Synthesis" 및 미국 특허 제5,939,250호, 1996년 5월 22일 출원, 발명의 명칭: "Production of Enzymes Having Desired Activities by Mutagenesis"].

[0404] 본 발명의 폴리펩티드의 변이체는, 본 발명의 서열의 폴리펩티드를 이루는 아미노산 잔기들 중 하나 이상이 보존 또는 비보존 아미노산 잔기(하나의 측면에서, 보존된 아미노산 잔기)로 치환되고, 이와 같이 치환된 아미노산 잔기는 유전자 암호에 의해 암호화되는 것일 수 있거나 또는 아닐 수 있는 변이체일 수 있다.

[0405] 하나의 측면에서, 보존적 치환은 폴리펩티드 내 소정의 아미노산이 유사한 특성을 가지는 다른 아미노산으로 치환되는 경우이다. 하나의 측면에서, 본 발명의 보존적 치환은 다음과 같은 치환들을 포함한다: 지방족 아미노산 예를 들어, 알라닌, 발린, 루신 및 이소루신의 다른 지방족 아미노산으로의 치환; 세린의 트레오닌으로의 치환 또는 그 반대; 산성 잔기 예를 들어, 아스파르트산 및 글루탐산의 다른 산성 잔기로의 치환; 아미드기를 보유하는 잔기 예를 들어, 아스파라긴 및 글루타민의 다른 잔기(아미드 기를 보유하는 잔기)로의 치환; 염기성 잔기 예를 들어, 리신 및 아르기닌의 다른 염기성 잔기로의 치환; 그리고 방향족 잔기 예를 들어, 페닐알라닌, 티로신의 다른 방향족 잔기로의 치환.

[0406] 기타 변이체로서는 본 발명의 폴리펩티드를 이루는 아미노산 잔기 중 하나 이상이 치환기를 포함하는 변이체가 있다. 하나의 측면에서, 기타 변이체로서는 폴리펩티드가 다른 화합물 예를 들어, 폴리펩티드의 반감기를 증가

시키는 화합물(예를 들어, 폴리에틸렌 글리콜)과 결합되어 있는 변이체가 있다. 부가의 변이체로서는 부가의 아미노산이 폴리펩티드 예를 들어, 리더 서열, 분비 서열, 프로 단백질 서열, 또는 폴리펩티드의 정제, 증폭 또는 안정화를 촉진하는 서열에 융합되어 있는 변이체가 있다.

[0407] 몇몇 측면에서, 단편, 유도체 및 유사체는 본 발명의 폴리펩티드와 동일한 생물학적 기능 또는 활성을 보유한다. 다른 측면에서, 단편, 유도체 또는 유사체는 프로 단백질을 포함하므로, 이 단편, 유도체 또는 유사체는 이 프로 단백질의 일부를 절단하여 활성인 폴리펩티드를 생산함으로써 활성화될 수 있다.

[0408] 숙주 세포 내에서 단백질 발현 수준을 높이기 위한 코돈의 최적화

[0409] 본 발명은 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소-암호화 핵산을 변형시켜 코돈 사용 방식을 바꾸는(예를 들어, 최적화시키는) 방법을 제공한다. 하나의 측면에서, 본 발명은 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 암호화하는 핵산 내 코돈을 변형시켜 이 코돈의 숙주 세포에서의 발현율을 증감시키는 방법을 제공한다. 본 발명은 또한 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 암호화하는 핵산의 숙주 세포에서의 발현율을 증가시키도록 변형된, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 암호화하는 핵산, 이와 같이 변형된 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소, 그리고 이와 같이 변형된 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 제조하는 방법을 제공한다. 이러한 방법은 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소-암호화 핵산 내 "비선호" 또는 "저선호" 코돈을 확인하는 단계와, 이와 같은 비선호 또는 저선호 코돈 중 하나 이상을 치환될 코돈에 의해 생성되는 아미노산과 동일한 아미노산을 암호화하는 "선호 코돈"으로 치환하는 과정을 포함하며, 이때 핵산 내에 존재하는 하나 이상의 비선호 코돈이나 저선호 코돈은 동일한 아미노산을 암호화하는 선호 코돈으로 치환된다. 선호 코돈은 숙주 세포 내 유전자의 암호화 서열에 과다 출현하는 코돈이고, 비선호 또는 저선호 코돈은 숙주 세포 내 암호화 서열에 적게 출현하는 코돈이다.

[0410] 본 발명의 핵산, 발현 카세트 및 벡터를 발현시키는 숙주 세포로서는 박테리아, 효모, 진균, 식물 세포, 곤충 세포 및 포유동물 세포를 포함한다(상기 참조). 그러므로, 본 발명은 이러한 세포들 모두에서의 코돈 사용 방식, 코돈-변형 핵산 및 이 코돈-변형된 핵산에 의해 생산되는 폴리펩티드를 최적화하는 방법을 제공한다. 대표적인 숙주 세포로서는 그람 음성 박테리아 예를 들어, 에스케리치아 콜라이(*Escherichia coli*); 그람 양성 박테리아 예를 들어, 스트렙토마이세스 종(*Streptomyces sp.*), 락토바실러스 가세리(*Lactobacillus gasseri*), 락토코커스 락티스(*Lactococcus lactis*), 락토코커스 크레모리스(*Lactococcus cremoris*), 바실러스 서브틸리스(*Bacillus subtilis*), 바실러스 세레우스(*Bacillus cereus*)를 포함한다. 대표적인 숙주 세포로서는 또한 진핵 생물 유기체 예를 들어, 다양한 효모 예를 들어, 사카로마이세스 종(*Saccharomyces sp.*) 예를 들어, 사카로마이세스 세레비시아에(*Saccharomyces cerevisiae*), 쉬조사카로마이세스 폼베(*Schizosaccharomyces pombe*), 피치아 파스토리스(*Pichia pastoris*) 및 클루이베로마이세스 락티스(*Kluyveromyces lactis*), 한세놀라 폴리모르파(*Hansenula polymorpha*), 아스퍼질러스 나이저(*Aspergillus niger*), 및 포유동물 세포 및 세포주, 그리고 곤충 세포 및 세포주를 포함한다. 그러므로, 본 발명은 이러한 유기체와 종 내에서 발현에 최적화된 핵산 및 폴리펩티드를 포함한다.

[0411] 예를 들어, 박테리아 세포로부터 분리된 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 암호화하는 핵산 코돈은 핵산이 상기 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소가 유래된 박테리아와 상이한 박테리아 세포, 효모, 진균, 식물 세포, 곤충 세포 및 포유동물 세포 내에서 최적으로 발현되도록 변형된다. 코돈을 최적화하는 방법은 당 업계에 널리 공지되어 있다[예를 들어, 미국 특허 제5,795,737호; Baca (2000) Int. J. Parasitol. 30:113-118; Hale (1998) Protein Expr. Purif. 12:185-188; Narum (2001) Infect. Immun. 69:7250-7253]. 또한, 문헌[Narum (2001) Infect. Immun. 69:7250-7253(마우스 시스템 내 코돈을 최적화하는 것에 관하여 기술함); Outchkourov (2002) Protein Expr. Purif. 24:18-24(효모 내 코돈을 최적화하는 것에 관하여 기술함); Feng (2000) Biochemistry 39:15399-15409(이. 콜라이 내 코돈을 최적화하는 것에 관하여 기술함); Humphreys (2000) Protein Expr. Purif. 20:252-264(이. 콜라이 내에서의 분비에 영향을 미치는 코돈의 사용 방식을 최적화하는 것에 관하여 기술함)]을 참조하시오.

[0412] 트랜스제닉 비인간 동물

- [0413] 본 발명은 본 발명의 핵산, 폴리펩티드(예를 들어, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소), 발현 카세트 또는 벡터, 또는 형질 감염 또는 형질전환된 세포를 포함하는 트랜스제닉 비인간 동물을 제공한다. 본 발명은 또한 이러한 트랜스제닉 비인간 동물을 제조 및 사용하는 방법을 제공한다.
- [0414] 트랜스제닉 비인간 동물 예를 들어, 개, 염소, 토끼, 양, 돼지(예를 들어, 모든 돼지, 식용 돼지 및 관련 동물), 소, 래트 및 마우스는 본 발명의 핵산을 포함할 수 있다. 이러한 동물들은 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성을 연구하기 위한 생체내 모델이나, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성을 생체내에서 변이시키는 제제에 대한 스크리닝 모델로서 사용될 수 있다. 트랜스제닉 비인간 동물 내에서 발현되는 폴리펩티드의 암호화 서열은 구성적으로 디자인될 수 있거나, 아니면 조직-특이적, 발달-특이적 또는 유도성 전사 조절 인자의 제어 하에 있도록 디자인될 수 있다.
- [0415] 트랜스제닉 비인간 동물은 당 업계에 공지된 임의의 방법을 이용하여 디자인 및 생산될 수 있으며; 예를 들어, 형질전환된 세포 및 난자, 그리고 트랜스제닉 마우스, 래트, 토끼, 양, 돼지 및 소를 제조 및 사용하는 것에 관하여 기술하고 있는 미국 특허 제6,211,428호; 동 제6,187,992호; 동 제6,156,952호; 동 제6,118,044호; 동 제6,111,166호; 동 제6,107,541호; 동 제5,959,171호; 동 제5,922,854호; 동 제5,892,070호; 동 제5,880,327호; 동 제5,891,698호; 동 제5,639,940호; 동 제5,573,933호; 동 제5,387,742호; 동 제5,087,571호를 참조하시오. 또한, 예를 들어, 문헌[Pollock (1999) J. Immunol. Methods 231:147-157(트랜스제닉 낙농 동물의 밀크 중에 재조합 단백질을 생성하는 것에 관하여 기술함); Baguisi (1999) Nat. Biotechnol. 17:456-461(트랜스제닉 염소의 생산에 관하여 기술함)]을 참조하시오. 미국 특허 제6,211,428호에는, 뇌에 DNA 서열을 포함하는 핵산 구조물을 발현시키는 트랜스제닉 비인간 포유동물의 제조 및 사용 방법에 관하여 기술되어 있다. 미국 특허 제5,387,742호에는, 클로닝된 재조합 또는 합성 DNA 서열을 수정된 마우스 난자에 주입하고, 이 주입된 난자를 가임인 암컷 마우스에 착상시킨 다음, 이 트랜스제닉 마우스를 생육하는 것에 관하여 기술되어 있다. 미국 특허 제6,187,992호에는 트랜스제닉 마우스의 제조 및 사용에 관하여 기술되어 있다.
- [0416] "넉아웃 동물(knockout animal)"도 본 발명의 방법을 실시하는데에 사용될 수 있다. 예를 들어, 하나의 측면에서, 본 발명의 트랜스제닉 동물 또는 변형 동물로서는, 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소, 또는 이와 같은 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 포함하는 융합 단백질을 발현하는 유전자로 치환된 내인성 유전자를 발현시키지 않도록 조작된 "넉아웃 동물" 예를 들어, "넉아웃 마우스"를 포함한다.
- [0417] 트랜스제닉 식물 및 종자
- [0418] 본 발명은 본 발명의 핵산, 폴리펩티드(예를 들어, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소), 발현 카세트 또는 벡터, 또는 형질 감염되었거나 형질전환된 세포를 포함하는 트랜스제닉 식물 및 종자를 제공한다. 본 발명은 또한 본 발명의 핵산 및/또는 폴리펩티드(예를 들어, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소)를 포함하는 식물의 생산물 예를 들어, 오일, 종자, 잎 및 추출물 등을 제공한다. 트랜스제닉 식물은 쌍떡잎 식물(복엽) 또는 외떡잎 식물(단엽)일 수 있다. 본 발명은 또한 이와 같은 트랜스제닉 식물 및 종자를 생산 및 이용하는 방법을 제공한다. 본 발명의 폴리펩티드를 발현하는 트랜스제닉 식물 또는 식물 세포는 당 업계에 공지된 임의의 방법에 의해 구성될 수 있다. 예를 들어, 미국 특허 제6,309,872호를 참조하시오.
- [0419] 본 발명의 핵산 및 발현 구조물은 임의의 수단을 통하여 식물 세포내에 도입될 수 있다. 예를 들어, 핵산 또는 발현 구조물은 원하는 식물 숙주 계통에 도입될 수 있거나, 또는 핵산 또는 발현 구조물은 에피솜일 수 있다. 원하는 식물 계통에의 도입은, 숙주의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 생산이 내인성 전사 또는 번역 제어 요소에 의해 조절되도록 수행될 수 있다. 본 발명은 또한 예를 들어, 상동성 재조합법에 의한 유전자 서열의 삽입에 의해 내인성 유전자의 발현이 파괴된 "넉아웃 식물"을 제공한다. "넉아웃" 식물을 생산하는 방법에 관하여는 당 업계에 널리 공지되어 있다[Strepp (1998) Proc Natl. Acad. Sci. USA 95:4368-4373; Miao (1995) Plant J 7:359-365]. 이하, 트랜스제닉 식물에 관한 논의를 참조하시오.
- [0420] 본 발명의 핵산은 근본적으로 임의의 식물 예를 들어, 전분-생산 식물 예를 들어, 감자, 토마토, 대두, 비트, 옥수수, 밀, 쌀 및 보리 등에 원하는 특성을 부여하는데에 사용될 수 있다. 본 발명의 핵산은 식물의 대사 경로

를 조작하여 숙주의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 발현을 최적화하거나 변형시키는데에 사용될 수 있다. 이로써, 식물 내에서의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성을 변화시킬 수 있다. 대안적으로, 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소는 천연의 상태에서 이 식물에 의해 생산되지 않는 화합물을 생산할 트랜스제닉 식물을 생산하는데에 사용될 수 있다. 이로써, 신규 생산물 생산시 드는 비용을 낮출 수 있다.

[0421] 하나의 측면에서, 트랜스제닉 식물을 생산하는 첫 번째 단계는 식물 세포 내에서 발현되는 발현 구조물을 제조하는 것이다. 이에 사용되는 기술은 당 업계에 널리 공지되어 있다. 이러한 기술은 프로모터 즉, 리보솜이 mRNA에 효율적으로 결합하는 것을 촉진하는 암호화 서열을 선별 및 클로닝하는 단계와, 적당한 유전자 종결 인자 서열을 선별하는 단계를 포함할 수 있다. 하나의 예시적인 구성 프로모터로서는, 꽃 양배추 모자이크 바이러스로부터 유래하는 CaMV35S(일반적으로 식물 내에서 높은 수준으로 발현됨)이 있다. 기타 프로모터로서는 식물의 내부 또는 외부 환경에 보다 특이적으로 반응하는 것이 있다. 대표적인 광선-유도성 프로모터로서는 주요 엽록소 a/b 결합 단백질 암호화하는 cab 유전자로부터 유래하는 프로모터가 있다.

[0422] 하나의 측면에서, 핵산은 식물 세포 내에서 더욱 다량으로 발현되도록 변형된다. 예를 들어, 본 발명의 서열은 식물에서 살펴볼 수 있는 A-T 뉴클레오티드 쌍에 비해 더 높은 비율의 A-T 뉴클레오티드 쌍을 가질 것이며, 이들 중 일부는 G-C 뉴클레오티드 쌍을 더 선호한다. 그러므로, 암호화 서열 내에 존재하는 A-T 뉴클레오티드는 식물 세포 내에서 유전자 생산물의 생산을 증가시키는 아미노산 서열을 거의 변이시키지 않고서도 G-C 뉴클레오티드와 치환될 수 있다.

[0423] 선별 가능한 마커 유전자는 트랜스 유전자가 성공적으로 통합화된 식물 세포 또는 조직을 동정하기 위해 유전자 구성물에 가하여질 수 있다. 식물 세포에 유전자를 통합하여 이를 발현시키는 과정은 희귀한 현상(표적화 조직 또는 세포 내에서 적은 비율로 발생하는 현상)이기 때문에, 이는 필수적인 절차이다. 선별 가능한 마커 유전자는 보통 식물에 독성인 제제 예를 들어, 항생제나 제초제에 대한 내성을 제공하는 단백질을 암호화한다. 적당한 항생제 또는 제초제를 함유하는 배지에서 생육할 때 선별 가능한 마커 유전자가 통합된 식물 세포만이 생존할 것이다. 기타 삽입된 유전자에 있어서, 마커 유전자는 또한 정확한 기능을 수행하기 위해서 프로모터와 종결 서열을 필요로 하기도 한다.

[0424] 하나의 측면에서, 트랜스제닉 식물 또는 종자를 제조하는데에는 프로모터와 종결 인자 서열을 배치함과 아울러, 본 발명의 서열을 통합하는 과정을 포함하며, 경우에 따라서는 마커 유전자를 표적 발현 구조물(예를 들어, 플라스미드)에 통합하는 과정을 포함하기도 한다. 이는 적당한 방법을 통하여 변형된 유전자를 식물로 운반하는 과정을 포함할 수 있다. 예를 들어, 구조물은 기술 예를 들어, 식물 세포의 원형질체를 전기 천공법 및 미세 주입법을 이용하여 식물 세포의 게놈형 DNA에 직접 도입될 수 있으며, 또는 구조물은 탄도학적 방법 예를 들어, DNA 입자 발사법을 사용하여 식물 조직에 직접 도입될 수 있다. 예를 들어, 밀에 트랜스유전자를 도입하기 위해 입자 발사법을 사용하는 것에 관하여 논의하고 있는 문헌[Christou (1997) Plant Mol. Biol. 35:197-203; Pawlowski (1996) Mol. Biotechnol. 6:17-30; Klein (1987) Nature 327:70-73; Takumi (1997) Genes Genet. Syst. 72:63-69]과, YAC을 식물 세포에 도입하는데에 입자 발사법을 사용하는 것에 관하여 논의하고 있는 문헌[Adam (1997), 상동]을 참조하시오. 예를 들어, 문헌[Rinehart (1997), 상동]에는 트랜스제닉 면 식물을 생산하기 위하여 입자 발사법을 사용하는 것에 관하여 논의되어 있다. 입자를 가속화하는 장치에 관하여는 미국 특허 제5,015,580호에 기술되어 있으며; 또한, 바이오래드(BioRad; Biolistics) PDS-2000 입자 가속 장치가 시판되고 있고; 또한, 문헌[John, 미국 특허 제5,608,148호; 및 Ellis, 미국 특허 제5,681,730호(나자 식물의 입자-매개 형질전환에 관하여 기술함)]을 참조하시오.

[0425] 하나의 측면에서, 원형질체는 고정화되어 핵산 예를 들어, 발현 구조물이 주입될 수 있다. 비록 곡류의 원형질체로부터 식물이 재생되는 것은 쉽지 않지만, 식물 재생은 원형질체 유도 캘러스로부터의 체성 배아 발생법을 이용하는 콩과 식물 재생법을 통해 이루어질 수 있다. 조직화된 조직은 유전자 총 기술을 이용하여 나출 DNA로 형질전환될 수 있는데, 여기서, DNA는, 세포 크기의 100분의 1에 해당하는 발포체로서 DNA를 세포와 기관의 심부까지 운반하는 텅스텐 미립 추진체(microprojectile) 상에 코팅되어 있다. 이후, 형질전환된 조직은 일반적으로 체성 배아 발생법에 의해 재생이 유도된다. 이러한 기술은 옥수수 및 쌀을 포함하는 몇몇 곡류 종 내에서 성공적으로 진행된다.

[0426] 핵산 예를 들어, 발현 구조물은 또한 재조합 바이러스를 사용하여 식물 세포에 도입될 수도 있다. 식물 세포는 바이러스 벡터 예를 들어, 담배 모자이크 바이러스 유래 벡터를 사용하여 형질전환될 수 있다[Rouwendaal (1997)

Plant Mol. Biol. 33:989-999), 및 Porta (1996) "Use of viral replicons for the expression of genes in plants," Mol. Biotechnol. 5:209-221 참조].

[0427] 대안적으로, 핵산 예를 들어, 발현 구조물은 적당한 T-DNA 측접 부위와 결합하여 종래의 아그로박테리움 튜메페이션스 숙주 벡터 내에 도입될 수 있다. 이 아그로박테리움 튜메페이션스 숙주의 병독성 작용은 세포가 박테리아에 의해 감염될 때, 구조물과 인접한 마커를 식물 세포 DNA에 삽입시킬 것이다. 아그로박테리움 튜메페이션스-매개 형질전환 기술 예를 들어, 이원성 벡터를 사용하여 약독화시키는 기술은 과학 논문에 상세히 기술되어 있다. 예를 들어, 문헌[Horsch (1984) Science 233:496-498; Fraley (1983) Proc. Natl. Acad. Sci. USA 80:4803 (1983); Gene Transfer to Plants, Potrykus, ed. (Springer-Verlag, Berlin 1995)]을 참조하십시오. 에이.튜메페이션스 세포의 DNA는 박테리아 염색체와 Ti (종양-유도성) 플라스미드라고 알려진 다른 구조물에 포함된다. 상기 Ti 플라스미드는 감염 과정에서 식물 세포에 운반되는 T-DNA(길이 약 20 kb)라고 불리는 DNA 확장 부위, 감염 과정을 유도하는 일련의 vir(병독성) 유전자를 포함한다. 에이.튜메페이션스는 식물의 상처를 통하여만 감염될 수 있으며: 식물 뿌리나 줄기에 상처가 나면, 임의의 화학 신호를 내보내게 되고, 이에 반응하여, 에이.튜메페이션스의 vir 유전자는 활성화되어, Ti 플라스미드로부터 식물의 염색체까지 T-DNA를 운반하는데에 필요한 일련의 현상들을 유도한다. 이후, 상기 T-DNA는 상처를 통해 식물 세포에 도입된다. 하나의 가설로서는 식물 DNA가 복제 또는 전사될 때까지 T-DNA는 대기하고 있다가, 노출된 식물 DNA에 이 T-DNA가 삽입된다는 설이 있다. 에이.튜메페이션스를 트랜스 유전자 벡터로 사용하기 위해서는, T-DNA의 종양-유도 구획은 제거되어야 하며, 이때 T-DNA 경계 부위와 vir 유전자는 유지되어야 한다. 이후, 상기 트랜스 유전자는 T-DNA 경계 부위 즉, 식물 세포로 운반되어 식물의 염색체 내에 통합되는 부위 사이에 삽입된다.

[0428] 본 발명은 본 발명의 핵산을 사용하여 외떡잎 식물 예를 들어, 중요한 곡류를 형질전환시키는 방법을 제공한다 [Hiei (1997) Plant Mol. Biol. 35:205-218]. 또한 예를 들어, T-DNA의 계놈형 DNA로의 통합에 관하여 논의하고 있는 문헌[Horsch, Science (1984) 233:496; Fraley (1983) Proc. Natl. Acad. Sci. USA 80:4803; Thykjaer (1997) 상동; Park (1996) Plant Mol. Biol. 32: 1135-1148]을 참조하십시오. 또한, 곡류나 기타 외떡잎 식물의 세포 내에서 기능을 가지는 유전자를 포함하는 DNA의 안정한 통합화 방법에 관하여 논의하고 있는 문헌[D'Halluin, 미국 특허 제5,712,135호]을 참조하십시오.

[0429] 하나의 측면에서, 세 번째 단계는 통합된 표적 유전자를 다음 세대에 전달할 수 있는 전체 식물을 선별 및 재생하는 단계를 포함한다. 이러한 재생 기술은 조직 배양 성장 배지 중에서 임의의 식물성 호르몬을 조작하는 원리를 이용할 수 있다. 하나의 측면에서, 이 방법은 원하는 뉴클레오티드 서열과 함께 도입된 살생물제 및/또는 제초제 마커를 사용한다. 배양된 원형질체로부터 식물을 재생시키는 것에 관하여는 문헌[Evans외 다수, Protoplasts Isolation and Culture, Handbook of Plant Cell Culture, pp. 124-176, MacMillan Publishing Company, New York, 1983; 및 Binding, Regeneration of Plants, Plant Protoplasts, pp. 21-73, CRC Press, Boca Raton, 1985]에 기술되어 있다. 재생은 또한 식물 캘러스, 외식편, 기관 또는 이의 일부로써 이루어질 수도 있다. 이러한 재생 기술에 관하여는 문헌[Klee (1987) Ann. Rev. of Plant Phys. 38:467-486]에 개시되어 있다. 트랜스제닉 조직 예를 들어, 미성숙 배아로부터 전체 식물을 얻기 위해서는, 이 조직들은 영양분과 호르몬을 함유하는 연속 배지와 같은 제어 환경 조건 하에서 생육될 수 있다(이 과정을 조직 배양이라고 함). 일단 전체 식물이 생산되어 종자를 생산하게 되면, 자손의 평가가 시작된다.

[0430] 하나의 측면에서, 발현 카세트가 트랜스제닉 식물에 안정하게 통합화된 후, 이 발현 카세트는 유성 교잡에 의해 다른 식물에도 도입될 수 있다. 교잡될 종에 따라서 다수의 표준적 육종 기술 중 임의의 기술이 사용될 수 있다. 본 발명의 핵산의 트랜스제닉 발현으로 인해 표현형에 변화가 올 수 있기 때문에, 본 발명의 재조합 핵산을 포함하는 식물은 제2의 식물과 유성 교잡될 수 있으며, 그 결과 최종 산물을 얻을 수 있다. 그러므로, 본 발명의 종자는 본 발명의 2개의 트랜스제닉 식물 사이의 교잡, 또는 본 발명의 식물과 다른 식물 사이의 교잡으로부터 유래될 수 있다. 양쪽 부모 식물이 본 발명의 폴리펩티드(예를 들어, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소)를 발현시킬 때, 바람직한 효과(예를 들어, 본 발명의 폴리펩티드를 발현시켜, 개화 거동이 바뀐 식물이 생산되는 효과)는 강화될 수 있다. 원하는 효과는 표준적인 증식 방법에 의해 다음 식물 세대에 전달될 수 있다.

[0431] 하나의 측면에서, 본 발명의 핵산 및 폴리펩티드는 임의의 식물 또는 종자에서 발현되거나 또는 여기에 삽입될 수 있다. 본 발명의 트랜스제닉 식물은 쌍떡잎 식물이거나 또는 외떡잎 식물일 수 있다. 본 발명의 외떡잎 트랜스제닉 식물로서는 목초 예를 들어, 방목용 목초(블루 그래스, 포아(Poa)), 마소의 먹이용인 목초 예를 들어, 김의털(festuca), 쥐보리(lolium), 북방형 목초 예를 들어, 아그로티스(Agrostis), 및 곡류 예를 들어, 밀, 귀리, 호밀, 보리, 쌀, 수수 및 옥수수(강냉이)가 있다. 본 발명의 쌍떡잎 트랜스제닉 식물의 예로서는 담배, 콩

과 식물 예를 들어, 루핀(*lupin*), 감자, 사탕 무우, 완두콩, 콩 및 대두와, 십자화과 식물(브라시카세아에(*Brassicaceae*)과 식물) 예를 들어, 꽃양배추, 유채씨, 및 매우 밀접하게 관련된 모델 유기체인 아라비도시스 탈리아나(*Arabidopsis thaliana*)가 있다. 그러므로, 본 발명의 트랜스제닉 식물 및 종자로서는 광범위한 식물 예를 들어, 아나카르디움(*Anacardium*), 아라키스(*Arachis*), 아스파라거스(*Asparagus*), 아트로파(*Atropa*), 아베나(*Avena*), 브라시카(*Brassica*), 사이트러스(*Citrus*), 사이트룰러스(*Citrullus*), 캡시쿰(*Capsicum*), 카타머스(*Carthamus*), 코코스(*Cocos*), 커피아(*Coffea*), 큐커미스(*Cucumis*), 큐커비타(*Cucurbita*), 다우커스(*Daucus*), 엘라에이스(*Elaeis*), 프래가리아(*Fragaria*), 글리신(*Glycine*), 가시피움(*Gossypium*), 헬리안투스(*Helianthus*), 헤테로칼리스(*Heterocallis*), 호르데움(*Hordeum*), 하이오스시아머스(*Hyoscyamus*), 락투카(*Lactuca*), 리눔(*Linum*), 롤륨(*Lolium*), 루피너스(*Lupinus*), 리코페르시콘(*Lycopersicon*), 말루스(*Malus*), 마니호트(*Manihot*), 마조라나(*Majorana*), 메디카고(*Medicago*), 니코티아나(*Nicotiana*), 올레아(*Olea*), 오라이자(*Oryza*), 패니움(*Panimum*), 패니세툼(*Pannisetum*), 페르세아(*Persea*), 파세올루스(*Phaseolus*), 피스타치아(*Pistachia*), 피섬(*Pisum*), 파이러스(*Pyrus*), 프루너스(*Prunus*), 라파누스(*Raphanus*), 리시누스(*Ricinus*), 세케일(*Secale*), 세네시오(*Senecio*), 시네피스(*Sinapis*), 솔라눔(*Solanum*), 소르검(*Sorghum*), 테오브로머스(*Theobromus*), 트리곤넬라(*Trigonella*), 트리티쿰(*Triticum*), 비시아(*Vicia*), 비티스(*Vitis*), 비그나(*Vigna*) 및 지아(*Zea*) 속에 속하는 종을 포함하나, 이에 한정되는 것은 아니다.

[0432] 대안적인 구체예에서, 본 발명의 핵산은 섬유 세포를 함유하는 식물 예를 들어, 목화, 판야 나무(카포크(*Kapok*), 세이바 펜탄듀라(*Ceiba pentandura*)), 사막 버드나무(*desert willow*), 크레어스트 덩굴(*creosote bush*), 윈터 팻(*winterfat*), 발사(*balsa*), 모시, 양마, 삼, 로젤, 주트, 사이잘 섬유 및 아마를 포함하는 식물 내에서 발현된다. 대안적인 구체예에서, 본 발명의 트랜스제닉 식물은 가시피움 속의 일원 예를 들어, 가시피움 종들 예를 들어, 지.아르보레움(*G. arboreum*); 지.허바세움(*G. herbaceum*), 지.바바덴스(*G. barbadense*) 및 지.허수텀(*G. hirsutum*)의 일원들일 수 있다.

[0433] 본 발명은 또한 본 발명의 폴리펩티드(예를 들어, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 또는 항체)를 다량으로 생산하는데 사용되는 트랜스제닉 식물을 제공한다. 예를 들어, 문헌[Palmgren (1997) Trends Genet. 13:348; Chong (1997) Transgenic Res. 6:289-296](아그로박테리움 튜메피케이션스-매개 일 절편 형질전환 방법을 통해, 옥신-유도성인 2방향 만노핀 합성 효소(*mas 1', 2'*) 프로모터를 사용하여 트랜스제닉 감자 식물 내에서 인간의 모유 단백질인 베타-카세인을 생산하는 방법)을 참조하시오. 공지된 방법을 사용하였을 때, 당 업자는 트랜스제닉 식물 내에서 트랜스 유전자 mRNA 또는 단백질의 증감 여부를 확인함으로써 본 발명의 식물에 대해 스크리닝할 수 있다. mRNA 또는 단백질을 확인 및 정량하는 방법에 관하여는 당 업계에 널리 공지되어 있다.

[0434] 폴리펩티드 및 펩티드

[0435] 하나의 측면에서, 본 발명은 본 발명의 대표적인 서열에 대하여 서열 동일성을 가지는(예를 들어, 서열 동일성이 약 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 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% 이상이거나 완벽한(100%) 서열 동일성을 갖는) 분리된 또는 제조한 폴리펩티드, 예를 들어 서열 번호 2, 서열 번호 4, 서열 번호 6, 서열 번호 8, 서열 번호 10, 서열 번호 12, 서열 번호 14, 서열 번호 16, 서열 번호 18, 서열 번호 20, 서열 번호 22, 서열 번호 24, 서열 번호 26, 서열 번호 28, 서열 번호 30, 서열 번호 32, 서열 번호 34, 서열 번호 36, 서열 번호 38, 서열 번호 40, 서열 번호 42, 서열 번호 44, 서열 번호 46, 서열 번호 48, 서열 번호 50, 서열 번호 52, 서열 번호 54, 서열 번호 56, 서열 번호 58, 서열 번호 60, 서열 번호 62, 서열 번호 64, 서열 번호 66, 서열 번호 68, 서열 번호 70, 서열 번호 72, 서열 번호 74, 서열 번호 76, 서열 번호 78, 서열 번호 80, 서열 번호 82, 서열 번호 84, 서열 번호 86, 서열 번호 88, 서열 번호 90, 서열 번호 92, 서열 번호 94, 서열 번호 96, 서열 번호 98, 서열 번호 100, 서열 번호 102, 서열 번호 104, 서열 번호 106, 서열 번호 108, 서열 번호 110, 서열 번호 112, 서열 번호 114, 서열 번호 116, 서열 번호 118, 서열 번호 120, 서열 번호 122, 서열 번호 124, 서열 번호 126, 서열 번호 128, 서열 번호 130, 서열 번호 132, 서열 번호 134, 서열 번호 136, 서열 번호 138, 서열 번호 140, 서열 번호 142, 서열 번호 144, 서열 번호 146, 서열 번호 148, 서열 번호 150, 서열 번호 152, 서열 번호 154, 서열 번호 156, 서열 번호 158, 서열 번호 160, 서열 번호 162, 서열 번호 164 또는 서열 번호 166에 제시된 서열[이하의 표 1a~1c, 2 및 3과, 실시예 1 및 4, 그리고 서열 목록 참조]을 가지는 단백질을 제공한다. 서열 동일성 %는 폴리펩티드 전체 길이에 걸쳐서 해당할 수 있거나, 아니면 이 상동성은 약 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650 또는 700개 이상의 잔기로 이루어

진 부위에 걸쳐서 해당할 수 있다.

[0436] 본 발명의 폴리펩티드는 또한 대표적인 폴리펩티드의 전체 길이보다 더욱 짧을 수도 있다. 대안적 측면에서, 본 발명은 크기가 폴리펩티드 중 약 5개 잔기 및 폴리펩티드 전체 길이 사이인 폴리펩티드(펩티드, 단편) 예를 들어, 효소 예를 들어, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 제공하며; 이의 크기는 예를 들어, 약 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650 또는 700 개 이상의 잔기이며, 예를 들어, 본 발명의 대표적인 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 연속적 잔기들로 이루어져 있다. 본 발명의 펩티드(예를 들어, 본 발명의 대표적인 폴리펩티드의 부분 서열)는 예를 들어, 표지화 프로브, 항원(면역원), 저항원(toleragen), 모티프, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성화 부위(예를 들어, "촉매 도메인"), 신호 서열 및/또는 프리프로 도메인으로서 유용할 수 있다.

[0437] 대안적 측면에서, 셀룰라제 활성 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 활성을 보유하는 본 발명의 폴리펩티드는 예를 들어, 셀룰라제 활성 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 활성과 상관 있는 아미노산 잔기들 즉, 특정 구조적 요소를 공유하는 폴리펩티드 속에 해당하는 일원이다. 이와 같이 공유하는 구조적 요소는 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 변이체를 통상적으로 생산하는데 사용될 수 있다. 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 공유적 구조 요소들은, 본 발명의 폴리펩티드 속 범주에 속하는 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 변이체를 통상적으로 생산함에 있어서 가이드로서 사용될 수 있다.

[0438] 본원에 사용된 "셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제"란 용어는, 셀룰로스(예를 들어, 본 발명의 대표적인 폴리펩티드로서, 이하 표 1a~1c, 2 및 3과, 실시예 1 및 4를 참조하시오)의 전체적 또는 부분적 파괴 및/또는 가수분해, 또는 셀룰로스나 리그노셀룰로스 물질 예를 들어, 리그노셀룰로스를 포함하는 바이오매스 물질의 임의의 변형을 촉진할 수 있는 임의의 폴리펩티드 또는 효소를 포함하는 의미이다.

[0439] 몇몇 측면에서, 본 발명의 폴리펩티드는 대안적 효소 활성 예를 들어, 이하 표 3a~3e에 제시된 활성을 가질 수 있다. 이하에 제시된 바와 같이, 예를 들어, 서열 번호 164에 제시된 서열을 가지는 폴리펩티드, 예를 들어, 서열 번호 163에 의해 암호화되는 폴리펩티드는 알칼리성 엔도글루카나제/셀룰라제 활성을 가질 수 있고; 서열 번호 110에 제시된 서열을 가지는 폴리펩티드, 예를 들어, 서열 번호 109에 의해 암호화되는 폴리펩티드는 자일라나제 활성을 가질 수 있고; 서열 번호 12에 제시된 서열을 가지는 폴리펩티드, 예를 들어, 서열 번호 11에 의해 암호화되는 폴리펩티드는 NAD 결합 산화 환원 효소 활성을 가질 수 있고; 서열 번호 118에 제시된 서열을 가지는 폴리펩티드, 예를 들어, 서열 번호 117에 의해 암호화되는 폴리펩티드는 단쇄(short chain) 탈수소 효소 활성을 가질 수 있고; 서열 번호 14에 제시된 서열을 가지는 폴리펩티드, 예를 들어, 서열 번호 13에 의해 암호화되는 폴리펩티드는 NADH 의존성 탈수소 효소 활성을 가질 수 있고; 서열 번호 138에 제시된 서열을 가지는 폴리펩티드, 예를 들어, 서열 번호 137에 의해 암호화되는 폴리펩티드는 펩티다제 활성을 가질 수 있고; 서열 번호 162에 제시된 서열을 가지는 폴리펩티드, 예를 들어, 서열 번호 161에 의해 암호화되는 폴리펩티드는 셀룰라제 활성 이외에도 알칼리성 엔도글루카나제 활성을 가질 수 있고; 서열 번호 42에 제시된 서열을 가지는 폴리펩티드, 예를 들어, 서열 번호 41에 의해 암호화되는 폴리펩티드는 시스테이닐 tRNA 합성 효소 활성을 가질 수 있고; 서열 번호 32에 제시된 서열을 가지는 폴리펩티드, 예를 들어, 서열 번호 31에 의해 암호화되는 폴리펩티드는 셀로텍스트린 포스포릴라제 활성을 가질 수 있고; 서열 번호 50에 제시된 서열을 가지는 폴리펩티드, 예를 들어, 서열 번호 49에 의해 암호화되는 폴리펩티드는 fdhd/narq 산화 환원 효소 활성을 가질 수 있고; 서열 번호 54에 제시된 서열을 가지는 폴리펩티드, 예를 들어, 서열 번호 53에 의해 암호화되는 폴리펩티드는 라디칼 S-아데노실메티오닌(SAM) 활성을 가질 수 있고; 서열 번호 58에 제시된 서열을 가지는 폴리펩티드, 예를 들어, 서열 번호 57에 의해 암호화되는 폴리펩티드는 서브틸리신 유사 프로테아제 활성을 가질 수 있다.

서열 번호	효소 활성 알칼리성 엔도글루카나제 / 셀룰라제 ORF 001 - 제1군 (β- 글루코시다제)	신호 펩티드 절단 위치	신호 서열	공급원	E C 번호
103,104	ORF 001 - 제1군 (β- 글루코시다제)	1-30	MSCRTLMRRVGVGILLWGGFLRTGVSVTG	미 지	3.2.1.21
101, 102	ORF 003 - 제5군 (셀룰라제)	1-29	MRNHLNVPFYFFIFFFLASIFTVCSSTA	미 지	3.2.1.4
103, 104	제5군 (셀룰라제)	1-20	MLIGGLLVLLGFSSGGRQA	미 지	3.2.1.4
105, 106	제5군 (셀룰라제)			미 지	3.2.1.4
107, 108	제5군 (셀룰라제)	1-32	MEKQICSNVFTMLIGLLVLLGFSSGGRQA	미 지	3.2.1.4
109, 110	제10군 (자인라나제)	1-28	MKTHSFNLRSRITLLTAALLFIGATAGA	미 지	3.2.1.8
11, 12	ORF 003 - NAD 결합 산화환원효소			미 지	1.1.1.18
111, 112	제5군 (셀룰라제)	1-22	MRRLLITLATAVAAILSTTSCS	미 지	3.2.1.4
113, 114	ORF 003 - 제10군	1-27	MKVTRTAVAGIVAAVLTIGTSTASA	미 지	3.2.1.8
115, 116	ORF 004 - 단쇄 탈수소 효소			미 지	1.1.1.100
117, 118	ORF 011 - 단쇄 탈수소 효소	1-19	MPKVMMLVTGGSRGIGAAVA	미 지	1...
119, 120	ORF 002 - 산화환원효소			미 지	1.4.3.16

[0442]

[표 3b]

121,	ORF 004 – 제5군 (셀룰라제)	미 지	3.2.1.4
122,	ORF 006 – 제1군	미 지	3.2.1.21
123,	(β- 글루코시다제)	미 지	3.2.1.21
124,	ORF 009 – 제1군	미 지	1.1.1.100
125,	(β- 글루코시다제)	미 지	1...
126,	ORF 004 – 단쇄	미 지	1.1.1.18
127,	탈수소 효소	미 지	3.2.1.4
128,	ORF 010 – 단쇄	미 지	3.5.1.
129,	탈수소 효소	미 지	1...
130,	ORF 005 – NADH 의존적	미 지	3.2.1.21
131,	탈수소 효소	미 지	3.2.1.21
132,	ORF 007 – 제5군 (셀룰라제)	미 지	3.2.1.21
133,	ORF 006 – 제1군	미 지	3.2.1.4
134,	(β- 글루코시다제)	미 지	3.2.1.21
135,	ORF 001 – 셀룰라제 (글리코실	미 지	3.2.1.4
136,	하이드롤라제 제5군)	미 지	3.5.1.
137,	ORF 001 – 펙티다제 _M37	미 지	1...
138,	ORF 001 – 트레오닌	미 지	3.2.1.21
139,	탈수소 효소	미 지	3.2.1.21
140,	ORF 005 – 제1군	미 지	3.2.1.21
141,	(β- 글루코시다제)	미 지	3.2.1.21
142,	ORF 003 – 제1군	미 지	3.2.1.21
143,	(β- 글루코시다제)	미 지	3.2.1.21
144,	ORF 002 – 제1군	미 지	3.2.1.21
145,	(β- 글루코시다제)	미 지	3.2.1.21
146,	제10군 (자일라나제)	미 지	3.2.1.8
147,	제5군 (셀룰라제)	미 지	3.2.1.4
148,	ORF 007 – 제1군	미 지	3.2.1.21
149,	(β- 글루코시다제)	미 지	3.2.1.4
150,	제5군 (셀룰라제)	미 지	3.2.1.4
151,	ORF 007 – 제1군	미 지	3.2.1.21
152,	(β- 글루코시다제)	미 지	3.2.1.4

[0443]

153, 154, 155, 156, 157, 158, 159, 160	제5군 (셀룰라제) 제5군 (셀룰라제) 제5군 (셀룰라제) 제10군 (자일라나제)	미 지 미 지 미 지 미 지	3.2.1.4 3.2.1.4 3.2.1.4 3.2.1.8
161, 162, 165, 166	알칼리성 엔도글루카나제 / 셀룰라제 자일라나제	미 지	
17, 18	ORF 005 – β- 락타마제	1-30	MSORTLMSRRVGWGLLWGGLFLRTGSVTG
19, 20	ORF 008 – 제10군 (자일라나제)	1-23	MRYVLISCLALASLCAQPLPVST
21, 22	ORF 001 – 제5군 (셀룰라제)	1-20	MPVLFALFLVASSCAAQSLA
23, 24	ORF 003 – 제16군 + CBM ORF 001 –	1-26	MYKRLLSSVLIIMLLLSAWSPIVQA
25, 26	ORF 002 – 제1군 (β- 글루코시다제)		
27, 28	ORF 004 – 제1군 (β- 글루코시다제)		
29, 30	ORF 008 – 제1군 (β- 글루코시다제)		
3, 4	ORF 002 – 셀로덱스트린 인산화 효소		
31, 32	ORF 006 – 제1군 (β- 글루코시다제)		
33, 34	ORF 007 – 제5군 (셀룰라제)	1-23	MNKILFSSLLIFAGICPALQA
35, 36			

[0446] [표 3d]

37, 38	ORF 011 - 제1군 (β-글루코시다제)				미 지	3.2.1.21
39, 40	ORF 004 - 초징 산화환원 효소				미 지	4.1.1.
41, 42	ORF 004 - 시스템이닐 RNA 합성 효소				미 지	6.1.1.16
43, 44	ORF 011 - ORF 006 - 제1군				미 지	3.2.1.21
45, 46	(β-글루코시다제)				미 지	3.2.1.21
47, 48	ORF 002 - 제1군 (β-글루코시다제)				미 지	3.2.1.21
49, 50	ORF 006 - fdh/marq 산화환원 효소				미 지	3.2.1.21
5, 6	ORF 012 - 제6군 (셀룰라제)	1-29		MTRRSIVRSSSNKWLVLGAAL	미 지	3.2.1.91
51, 52	ORF 001 - 제5군 (셀룰라제)	1-20		MSRGILLVMSVLSGAALA	미 지	3.2.1.4
53, 54	ORF 002 - 라디칼 SAM 군				미 지	1...
55, 56	ORF 004 - 제1군 (β-글루코시다제)				미 지	3.2.1.21
57, 58	ORF 001 - 서브틸라신 유사 프로테아제				미 지	3.2.1.4
59, 60	제5군 (셀룰라제)				미 지	3.2.1.4
61, 62	제5군 (셀룰라제) ORF 1	1-52		MWTPARSTLAGSSEIPLMTMINFPNRKDSRMSLWIKL GILCMAGTVMVHG	미 지	3.2.1.4
63, 64	제5군 (셀룰라제) ORF 4	1-24		MKRREFMLGGAGVAAALASTLGVSA	미 지	3.2.1.4
65, 66	제10군 (자일라나제)	1-39		MNTLLPRRRLWSSTAILRTLAAAGALAAAGMWLAPVSAAN A	미 지	3.2.1.8
67, 68	제5군 (셀룰라제) - ORF 2	1-23		MKYIFSIIIMMLIGFIPVYGFG	미 지	3.2.1.4

[0447]

[0448]

[표 3e]

69, 70	제26군 (만나제) - ORF4	1-20	MSFKNHLLSLIVLFFSA	미 지	3.2.1.78
7, 8	ORF 003 - 이소스티레이드 탈수소 효소			미 지	1.1.1.42
71, 72	제5군 (셀룰라제)	1-21	MKLLKLIFLITVIFSDVSA	미 지	3.2.1.4
73, 74	제10군 (자일라나제)			미 지	3.2.1.21
75, 76	제5군 (셀룰라제)	1-21	MLRKLIVSVFGVMLTSA AAA	미 지	3.2.1.4
77, 78	제5군 (셀룰라제)	1-23	MKRKRVFHSLIVFFLMIGSFTSCGSA	미 지	3.2.1.4
79, 80	제5군 (셀룰라제)	1-25	MKYKAFIYLIVLILFYISINIVANA	미 지	3.2.1.4
81, 82	제5군 (셀룰라제)	1-25	MNLLAQYFSGLELILFISIFVSSA	미 지	3.2.1.4
83, 84	ORF 008 - 탈수소 효소			미 지	3.5.4.25
85, 86	ORF 008 - 제1군 (β-글루코시다제)			미 지	3.2.1.21
87, 88	제5군 (셀룰라제)	1-23	MRKSVFTLAVFLSALFAFTSQN	미 지	3.2.1.4
89, 90	제5군 (셀룰라제)	1-29	MKRSVSIFIACLIMTTLTISGVAAPESA	미 지	3.2.1.4
9, 10	ORF 004 - 제10군 (자일라나제)	1-26	MRSVRIVTFALAAALAVPLVTSTATA	미 지	3.2.1.8
91, 92	ORF 001 - 제3군			미 지	3.2.1.52
93, 94	ORF 002 - 알파-람노시다제			미 지	3.2.1.21
95, 96	ORF 001 - 제3군			미 지	3.2.1.31
97, 98	ORF 003 - 베타-글루코시다제			미 지	
99, 100	ORF 012 - 제1군 (β-글루코시다제)			미 지	3.2.1.21

[0449]

[0450]

본원에 사용된 "아미노산" 또는 "아미노산 서열"이란, 올리고펩티드, 펩티드, 폴리펩티드 또는 단백질 서열이나, 이 중 임의의 것의 단편, 일부 또는 서브유닛, 그리고 자연 발생 또는 합성 분자를 의미한다. "아미노산" 또는 "아미노산 서열"로서는 올리고펩티드, 펩티드, 폴리펩티드 또는 단백질 서열, 또는 이 중 임의의 것의 단편, 일부 또는 서브유닛, 그리고 자연 발생 또는 합성 분자를 포함한다. 본원에 사용된 "폴리펩티드"란 용어는, 각각이 펩티드 결합이나 변형된 펩티드 결합에 의해 결합된 아미노산 즉, 펩티드 아이소스티어(peptide isostere)를 의미하고, 이는 20개 유전자-암호화 아미노산 이외의 변형된 아미노산을 함유할 수 있다. 상기 폴리펩티드는 천연에서 진행되는 가공 예를 들어, 번역 후 가공이나, 당 업계에 널리 공지되어 있는 화학적 변형 기술에 의해 변형될 수 있다. 변형은 폴리펩티드 중 임의의 위치 예를 들어, 펩티드 주쇄, 아미노산 측쇄 및 아미노 또는 카복실 말단부에서 일어날 수 있다. 동일한 유형의 변형이 소정의 폴리펩티드 중 몇몇 위치에서 동일한 정도로 또는 다양하게 존재할 수 있음을 알 수 있을 것이다. 뿐만 아니라, 소정의 폴리펩티드에는 여러 가지 유형의 변형이 일어날 수도 있다. 변형의 예로서는 아세틸화, 아실화, ADP-리보실화, 아미드화, 플라빈의 공유 결합, 헴 부분의 공유 결합, 뉴클레오티드 또는 뉴클레오티드 유도체의 공유 결합, 지질 또는 지질 유도체의 공유 결합, 포스파티딜이노시톨의 공유 결합, 가교 고리화 반응, 이황화 결합 형성, 탈메틸화, 공유 가교 결합의 형성, 시스테인의 형성, 피로글루타메이트의 형성, 포르밀화, 감마-카복실화, 글리코실화, GPI 앵커 형성(GPI anchor formation), 하이드록실화, 요드화, 메틸화, 미리스톨화, 산화, 페그화, 글루칸 가수분해 효소 가공, 인산화, 프레닐화, 라세미화, 셀레노일화, 황산화 및 운반-RNA를 매개로 하는, 아미노산의 단백질로의 첨가 반응 예를 들어, 아르기닐화를 포함한다[Creighton, T.E., Proteins - Structure and Molecular Properties 2nd Ed., W.H. Freeman and Company, New York (1993); Posttranslational Covalent Modification of Proteins, B.C. Johnson, Ed., Academic Press, New York, pp. 1-12 (1983)]. 본 발명의 펩티드 및 폴리펩티드는 또한 이하에 더욱 상세히 설명되어 있는 모든 "모의체(mimetic)" 및 "펩티도모의체(peptidomimetic)" 형태의 것을 포함한다.

- [0451] 본원에 사용된 "분리된"이란 용어는, 어느 물질(예를 들어, 본 발명의 단백질 또는 핵산)이 그것의 기원이 되는 환경(예를 들어, 자연 발생 물질의 경우에는 천연 환경)으로부터 분리되는 경우를 의미한다. 예를 들어, 살아있는 동물에 존재하는 자연 발생 폴리뉴클레오타이드 또는 폴리펩티드는 분리된 것이 아니지만, 천연의 시스템 중 공존하는 물질중 일부 또는 전부로부터 분리된 자연 발생 폴리뉴클레오타이드 또는 폴리펩티드는 분리된 것이다. 이러한 폴리뉴클레오타이드는 벡터의 일부가 될 수 있으며/있거나, 이러한 폴리뉴클레오타이드 또는 폴리펩티드는 조성물의 일부일 수 있으며, 또한 이러한 벡터나 조성물은 그것의 기원이 되는 천연 환경의 일부가 아니므로 여전히 분리된 상태인 것이다. 본원에 사용된 "정제된"이란 용어는 완전히 순수할 필요는 없는 경우로서, 이는 상대적인 정의이다. 라이브러리로부터 유래하는 각각의 핵산은 통상적으로 전기 영동을 통하여 균질하게 정제된다. 이러한 클론으로부터 얻어지는 서열은 라이브러리 또는 전체 인간 DNA로부터 직접 얻어질 수는 없다. 본 발명의 정제된 핵산은 유기체 내 게놈형 DNA의 나머지로부터 $10^4 \sim 10^6$ 배 이상으로까지 많이 정제된다. 하나의 측면에서, "정제된"이란 용어에는 게놈형 DNA의 나머지로부터 정제된 핵산, 또는 라이브러리나 기타 환경에 존재하는 기타 서열로부터 1배 이상(예를 들어, 하나의 측면에서는 2배 또는 3배 이상, 또는 4배 또는 5배 이상) 더 많이 정제된 핵산을 포함한다.
- [0452] "제조합" 폴리펩티드 또는 단백질이란, 제조합 DNA 기술에 의해 생산된 폴리펩티드 또는 단백질; 즉, 원하는 폴리펩티드나 단백질을 암호화하는 외인성 DNA 구조물에 의해 형질전환된 세포로부터 생산된 폴리펩티드 또는 단백질을 의미하는 것이다. "합성" 폴리펩티드 또는 단백질은 화학적 합성법에 의해 제조된 것이다. 고상 화학적 펩티드 합성법은 또한 본 발명의 폴리펩티드 또는 단백질을 합성하는데에 사용될 수도 있다. 이러한 방법은 1960년대 초부터 당 업계에 공지되어 왔으며[Merrifield, R. B., J. Am. Chem. Soc., 85:2149-2154, 1963, 및 Stewart, J. M. and Young, J. D., Solid Phase Peptide Synthesis, 2nd Ed., Pierce Chemical Co., Rockford, Ill., pp. 11-12 참조], 최근에는 시판중인 실험실용 펩티드의 디자인과 합성용 키트에 사용되고 있다[Cambridge Research Biochemicals]. 이와 같은 시판중인 실험실용 키트는 일반적으로 문헌[H.M. Geysen외 다수, Proc. Natl. Acad. Sci., USA, 81:3998 (1984)]에 교시된 사항을 활용한 것으로서, 단일 평판에 연결되어 있는 다수의 "막대" 또는 "핀"의 끝 부분에서 펩티드를 합성할 수 있는 것이다.
- [0453] 2개의 핵산 또는 폴리펩티드에 있어서 "실질적으로 동일한"이란 어구는, 공지의 서열 비교 알고리즘 중 어느 하나 또는 시각적 조사를 통하여 측정하였을 경우, 2 이상의 서열을 최대 상응성에 대해 비교 및 정렬할 때 얻어지는 뉴클레오타이드 또는 아미노산 잔기(서열)의 상동성이 예를 들어, 약 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 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% 이상인 경우를 의미하는 것이다. 대안적 측면에서, 실질적 상동성은 약 100개 이상의 잔기로 이루어진 부위에 걸쳐서 존재하고, 가장 일반적으로 서열들은 약 150~200개 이상의 잔기에 걸친 부위가 실질적으로 동일하다. 다른 측면에서, 서열은 암호화 부위 전체에 걸쳐서 실질적으로 동일하다.
- [0454] 뿐만 아니라, "실질적으로 동일한" 아미노산 서열은 하나 이상의 보존적 또는 비-보존적 아미노산 치환, 결실 또는 삽입이 일어났다는 점에서 참조 서열과 상이한 서열이다. 하나의 측면에서, 상기 치환은 분자의 활성 부위가 아닌 위치에서 일어나거나, 분자의 활성 부위인 위치에서 일어나되, 단, 이 폴리펩티드는 그것의 기능상(효소적) 특성을 본질적으로 보유하고 있다. 예를 들어, 보존적 아미노산 치환은 하나의 아미노산을 동일한 군에 속하는 다른 아미노산으로 치환한다[예를 들어, 소수성 아미노산 예를 들어, 이소루신, 발린, 루신 또는 메티오닌을 다른 아미노산에 대해 치환하거나, 또는 하나의 극성 아미노산을 다른 아미노산에 대해 치환하는데, 예를 들어, 아르기닌을 리신으로, 글루탐산을 아스파르트산으로, 또는 글루타민을 아스파라긴으로 치환함]. 하나 이상의 아미노산은 예를 들어, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 폴리펩티드로부터 결실될 수 있으며, 그 결과, 이 폴리펩티드의 생물 활성을 거의 변형시키지 않고 이 폴리펩티드의 구조가 변형된다. 예를 들어, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 생물 활성에 필요하지 않은 아미노- 또는 카복시-말단 아미노산이 제거될 수 있다. 본 발명의 변형된 폴리펩티드 서열은 다수의 방법 예를 들어, 변형된 폴리펩티드 서열과 기질을 접촉시키는 단계 및 변형된 폴리펩티드가 검정법에서 이의 특이 기질의 양을 감소시키는지 여부, 또는 기능성 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 폴리펩티드와 기질의 효소 반응에 따른 생성 생성물의 양을 증가시키는지 여부를 결정하는 단계를 포함하는 방법에 의해, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 생물 활성에 대해 검정될 수 있다.
- [0455] 본원에 사용된 "단편"이란, 2개 이상의 상이한 구조로 존재할 수 있는 자연 발생 단백질의 일부를 의미한다. 단

편은 자연 발생 단백질과 동일하거나 또는 실질적으로 동일한 아미노산 서열을 가질 수 있다. 자연 발생 단백질과 상이한 3차원적 구조를 가지는 단편도 포함된다. 이의 예로서는 절단되면 상당히 높은 활성을 보유하는 성숙한 효소를 생산하도록 변형될 수 있는 "프로-형(pro-form)" 분자 예를 들어, 저 활성 프로단백질이 있다.

[0456] 하나의 측면에서, 본 발명은 단백질 및 펩티드 예를 들어, 본 발명의 셀룰라제의 결정(3차원) 구조를 제공하는데, 이 구조는 당 업계에 널리 공지된 통상의 절차에 따라서 만들어지고 분석되며[예를 들어, MacKenzie (1998) Crystal structure of the family 7 endoglucanase I (Cel7B) from *Humicola insolens* at 2.2 Å resolution and identification of the catalytic nucleophile by trapping of the covalent glycosyl-enzyme intermediate, *Biochem. J.* 335:409-416; Sakon (1997) Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*, *Nat. Struct. Biol.* 4:810-818; Varrot (1999) Crystal structure of the catalytic core domain of the family 6 cellobiohydrolase II, Cel6A, from *Humicola insolens*, at 1.92 Å resolution, *Biochem. J.* 337:297-304]; 또한, 본 발명의 셀룰라제 변이체를 통상적으로 생산함에 있어서의 가이드, 그리고 본 발명의 범위 내에 포함되는 효소 종을 동종함에 있어서의 가이드로서 특징의 구조적 요소를 나타내고 이를 동정할 수 있도록 해준다.

[0457] 본 발명의 폴리펩티드 및 펩티드는 천연 공급원으로부터 분리될 수 있거나, 합성될 수 있거나, 또는 재조합적으로 생산될 수 있는 폴리펩티드이다. 펩티드 및 단백질은 시험관 내 또는 생체내에서 재조합적으로 발현될 수 있다. 본 발명의 펩티드 및 폴리펩티드는 당 업계에 공지된 임의의 방법을 사용하여 제조 및 분리될 수 있다. 본 발명의 폴리펩티드 및 펩티드는 또한 당 업계에 널리 공지된 화학적 방법을 사용하여 전체적 또는 부분적으로 합성될 수도 있다. 예를 들어, 문헌[Caruthers (1980) *Nucleic Acids Res. Symp. Ser.* 215-223; Horn (1980) *Nucleic Acids Res. Symp. Ser.* 225-232; Banga, A.K., *Therapeutic Peptides and Proteins, Formulation, Processing and Delivery Systems* (1995) Technomic Publishing Co., Lancaster, PA]을 참조하시오. 예를 들어, 펩티드 합성법은 다양한 고상 기술[예를 들어, Robert(1995) *Science* 269:202; Merrifield(1997) *Methods Enzymol.* 289:3-13 참조]을 이용하여 수행될 수 있으며, 자동화된 합성법은 제조자에 의해 제공되는 지침에 따라서 예를 들어, ABI 431A 펩티드 합성기(Perkin Elmer)를 사용하여 진행될 수 있다.

[0458] 본 발명의 펩티드 및 폴리펩티드는 또한 글리코실화될 수도 있다. 글리코실화는 번역 후에 화학적 기작 또는 세포내 생합성 기작에 의해서 이루어질 수 있는데, 여기서, 상기 세포내 생합성 기작은 공지의 글리코실화 모티프를 이용하며, 상기 모티프는 서열에 원래 존재하고 있는 것이거나 펩티드로서 첨가될 수 있거나, 또는 핵산 암호화 서열에 첨가될 수 있는 것이다. 글리코실화는 O-결합 또는 N-결합될 수 있다.

[0459] 상기 정의한 바와 같이, 본 발명의 펩티드 및 폴리펩티드는 모든 "모의체" 및 "펩티도모의체" 형태의 것을 포함한다. "모의체" 또는 "펩티도모의체"란 용어는, 본 발명의 폴리펩티드와 실질적으로 동일한 구조적 특징 및/또는 기능상의 특징을 가지는 합성 화학 화합물을 의미하는 것이다. 모의체는 합성 아미노산만으로 이루어질 수 있거나, 아미노산의 비 천연 유사체로 이루어질 수 있거나, 또는 일부는 천연 펩티드 아미노산으로, 일부는 아미노산의 비 천연 유사체로 이루어진 키메라 분자일 수 있다. 보존적 치환이 모의체 구조 및/또는 활성을 실질적으로 바꾸지 않는 한, 상기 모의체는 임의의 양의 천연 아미노산의 보존적 치환을 포함할 수도 있을 것이다. 본 발명의 폴리펩티드의 보존적 변이체나 이 폴리펩티드 속 일원인 본 발명의 폴리펩티드(본 발명의 대표적인 서열과의 서열 동일성이 약 50% 이상인 폴리펩티드)를 사용하여 통상의 실험을 할 경우, 이 모의체가 본 발명의 범위에 속하는지 여부 즉, 이 모의체의 구조 및/또는 기능이 실질적으로 변형되었는지 여부를 확인할 수 있다. 그러므로, 하나의 측면에서, 모의체 조성물이 셀룰라제 활성 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성을 가지면 이 모의체 조성물은 본 발명의 범위에 속하는 것이다.

[0460] 본 발명의 폴리펩티드 모의체 조성물은 비 천연 구조 성분을 경우에 따라 조합하여 포함할 수 있다. 대안적 측면에서, 본 발명의 모의체 조성물은 다음과 같은 3가지 구조 군 중 하나 또는 전부를 포함한다: a) 천연 아미드 결합("펩티드 결합") 이외의 잔기 결합 군; b) 자연 발생 아미노산 잔기를 대신하는 비 천연 잔기; 또는 c) 2차 구조를 모의시키는 잔기 즉, 2차 구조 예를 들어, 베타 선회부, 감마 선회부, 베타 병풍 구조, 알파 나선 구조 등을 유도 또는 안정화시키는 잔기. 예를 들어, 본 발명의 폴리펩티드는 그 잔기 전부 또는 일부가 천연 펩티드 결합 이외의 화학적 수단에 의해 결합될 때 모의체로서 특성 규명될 수 있다. 각각의 펩티도모의체 잔기는 펩티드 결합, 기타 화학 결합 또는 커플링 수단 예를 들어, 글루타르알데히드, N-하이드록시숙신이미드 에스테르, 2-작용성 말레이미드, N,N'-디시클로헥실카르보디이미드(DCC) 또는 N,N'-디이소프로필카르보디이미드(DIC)에 의해 결합될 수 있다. 통상의 아미드 결합("펩티드 결합")에 대안적일 수 있는 결합기로서는 예를 들어, 케토메틸렌

(예를 들어, $-C(=O)-NH-$ 에 대한 $-C(=O)-CH_2-$), 아미노메틸렌(CH_2-NH), 에틸렌, 올레핀($CH=CH$), 에테르(CH_2-O), 티오에테르(CH_2-S), 테트라졸(CN_4-), 티아졸, 레트로아미드, 티오아미드 또는 에스테르(예를 들어, Spatola (1983), Chemistry and Biochemistry of Amino Acids, Peptides and Proteins, Vol. 7, pp 267-357, "Peptide Backbone Modifications," Marcell Dekker, NY)를 포함한다.

[0461] 본 발명의 폴리펩티드는 또한 자연 발생되는 아미노산 잔기 대신에 비 천연 잔기 전부 또는 일부를 함유함으로써 모의체로서 특성 규명될 수 있다. 비 천연 잔기는 과학 및 특허 문헌에 널리 기술되어 있으며; 천연 아미노산 잔기의 모의체로서 유용한 소수의 대표적인 비 천연 조성물과 이에 대한 지침은 이하에 기술되어 있다. 방향족 아미노산의 모의체는 예를 들어, D- 또는 L-나필알라닌; D- 또는 L-페닐글리신; D- 또는 L-2 티에닐알라닌; D- 또는 L-1, -2, 3-, 또는 4- 피레네일알라닌; D- 또는 L-3 티에네일알라닌; D- 또는 L-(2-피리디닐)-알라닌; D- 또는 L-(3-피리디닐)-알라닌; D- 또는 L-(2-피라지닐)-알라닌; D- 또는 L-(4-이소프로필)-페닐글리신; D-(트리플루오로메틸)-페닐글리신; D-(트리플루오로메틸)-페닐알라닌; D-p-플루오로-페닐알라닌; D- 또는 L-p-비페닐 페닐알라닌; D- 또는 L-p-메톡시-비페닐페닐알라닌; D- 또는 L-2-인돌(알킬)알라닌; 및, D- 또는 L-알킬아미닌 [여기서, 알킬은 치환 또는 비 치환 메틸, 에틸, 프로필, 헥실, 부틸, 펜틸, 이소프로필, 이소-부틸, sec-이소틸, 이소-펜틸 또는 비-산성 아미노산일 수 있음]에 의해 치환됨으로써 생산될 수 있다. 비 천연 아미노산의 방향족 고리로서는 예를 들어, 티아졸릴, 티오펜릴, 피라졸릴, 벤지마다졸릴, 나프틸, 푸라닐, 피롤릴 및 피리딜 방향족 고리를 포함한다.

[0462] 산성 아미노산의 모의체는 예를 들어, 음 전하를 유지하는 비-카복실레이트 아미노산; (포스포노)알라닌; 황산화 트레오닌에 의해 생산될 수 있다. 카복실 측기(예를 들어, 아스파르틸 또는 글루타미드)는 또한 카르보디이미드($R'-N-C-N-R'$) 예를 들어, 1-시클로헥실-3(2-모폴리닐-(4-에틸)카르보디이미드 또는 1-에틸-3(4-아조니아-4,4-디메틸펜틸)카르보디이미드와 반응함으로써 선택적으로 변형될 수도 있다. 아스파르틸 또는 글루타미드는 또한 암모늄 이온과의 반응에 의해서 아스파라기닐 및 글루타미닐 잔기로 전환될 수도 있다. 염기성 아미노산 모의체는 예를 들어, (리신 및 아르기닌에 더하여) 아미노산인 오르니틴, 시트룰린 또는 (구아니디노)-아세트산 또는 (구아니디노)알킬-아세트산[여기서, 알킬은 상기 정의한 바와 같음] 과의 치환에 의해 생산될 수 있다. 니트릴 유도체(예를 들어, COOH 대신 CN-부분 함유)는 아스파라긴 또는 글루타민에 대해 치환될 수 있다. 아스파라기닐 및 글루타미닐 잔기는 상응하는 아스파르틸 또는 글루타미닐 잔기로 탈아민화될 수 있다. 하나의 측면에서, 알칼리성 조건 하에서 아르기닌 잔기 모의체는 아르기닐과 예를 들어, 하나 이상의 통상적인 시약 예를 들어, 페닐글리옥살, 2,3-부탄디온, 1,2-시클로-헥산디온 또는 닐히드린을 반응시킴으로써 생성될 수 있다. 티로신 잔기 모의체는 티로실을 예를 들어, 방향족 디아조늄 화합물 또는 테트라니트로메탄과 반응시킴으로써 생성될 수 있다. N-아세틸아미디졸 및 테트라니트로메탄이 사용되면 O-아세틸 티로실 종과 3-니트로 유도체가 각각 생성될 수 있다. 시스테인 잔기 모의체는 시스테인닐 잔기와 예를 들어, 알파-할로아세테이트 예를 들어, 2-클로로아세트산 또는 클로로아세타미드 및 상응하는 아민을 반응시킴으로써 생성될 수 있으며; 그 결과, 카복시메틸 또는 카복시아미도메틸 유도체가 생성된다. 시스테인 잔기 모의체는 또한 시스테인닐 잔기를 예를 들어, 브로모-트리플루오로아세톤, 알파-브로모-베타-(5-이미도조일)프로피온산; 인산클로로아세틸, N-알킬말레이미드, 이황화 3-니트로-2-피리딜; 이황화 메틸 2-피리딜; p-클로로머큐리벤조에이트; 2-클로로머큐리-4-니트로페놀; 또는 클로로-7-니트로벤조-옥사-1,3-디아졸과 반응시킴으로써 생성될 수도 있다. 리신 모의체는 리시닐과 예를 들어, 숙신산 또는 기타 카복시산 무수물을 반응시킴으로써 생성될 수 있으며, 이때 아미노 말단 잔기는 상기과 같이 반응하여 변형될 수 있다. 리신 및 기타 알파-아미노-함유 잔기 모의체는 또한 이미도에스테르 예를 들어, 메틸 피콜리니미데이트, 인산피리독살, 피리독살, 클로로보로하이드라이드, 트리니트로벤젠설포산, O-메틸이소우레아, 2,4-펜탄디온과의 반응, 그리고 글리옥실레이트와의 트랜스아미다제-촉매 반응에 의해 생성될 수도 있다. 메티오닌의 모의체는 예를 들어, 설폭시화메티오닌과의 반응에 의해 생성될 수 있다. 프롤린 모의체로서는 예를 들어, 피페콜산, 티아졸리딘 카복시산, 3- 또는 4-하이드록시 프롤린, 테하이드로프롤린, 3- 또는 4-메틸프롤린 또는 3,3-디메틸프롤린을 포함한다. 히스티딘 잔기 모의체는 히스티딜을 예를 들어, 디에틸프로카보네이트 또는 브롬화 파라-브로모페나실과 반응시킴으로써 생성될 수 있다. 기타 모의체로서는 예를 들어, 프롤린과 리신의 히드록실화에 의해 생성된 것들; 세틸 또는 트레오닐 잔기의 히드록실 기의 인산화에 의해 생성된 것들; 리신, 아르기닌 및 히스티딘의 알파-아미노기의 메틸화에 의해 생성된 것들; N-말단 아민의 아세틸화에 의해 생성된 것들; 주쇄 아미드 잔기의 메틸화 또는 N-메틸 아미노산과의 치환에 의해 생성된 것들; 또는 C-말단 카복실기의 아미드화에 의해 생성된 것들을 포함한다.

[0463] 하나의 측면에서, 본 발명의 폴리펩티드의 잔기 예를 들어, 아미노산은 또한 반대의 키랄성을 가지는 아미노산 (또는 펩티도모의체 잔기)에 의해 치환될 수도 있다. 하나의 측면에서, L-형(이는 또한 화학 실체의 구조에 따

라서 R-형 또는 S-형이라고도 칭함)로 자연 발생되는 임의의 아미노산은 동일하되, 반대의 키랄성을 가지는 (D-아미노산) 즉, R-형 또는 S-형이라고 불리는 화학 구조형 또는 펩티도모의체인 아미노산과 치환될 수 있다.

[0464] 본 발명은 또한 천연의 가공 예를 들어, 번역 후 가공(예를 들어, 인산화, 아실화 등)이나, 화학적 변형 기술에 의해 본 발명의 폴리펩티드를 변형시키는 방법을 제공하며, 그 결과, 변형된 폴리펩티드가 생성된다. 변형은 폴리펩티드 중 임의의 위치 예를 들어, 펩티드 주쇄, 아미노산 측쇄 및 아미노 또는 카복실 말단부에서 일어날 수 있다. 동일한 유형의 변형이 소정의 폴리펩티드 중 몇몇 위치에서 동일한 정도로 또는 다양하게 존재할 수 있음을 알 수 있을 것이다. 뿐만 아니라, 소정의 폴리펩티드에는 여러 가지 유형의 변형이 일어날 수도 있다. 하나의 측면에서, 변형의 예로서는 아세틸화, 아실화, ADP-리보실화, 아마이드화, 플라빈의 공유 결합, 헴 부분의 공유 결합, 뉴클레오타이드 또는 뉴클레오타이드 유도체의 공유 결합, 지질 또는 지질 유도체의 공유 결합, 포스파티딜이노시톨의 공유 결합, 가교 고리화 반응, 이황화 결합 형성, 탈메틸화, 공유 가교 결합의 형성, 시스테인의 형성, 피로글루타메이트의 형성, 포르밀화, 감마-카복실화, 글리코실화, GPI 앵커 형성, 하이드록실화, 요드화, 메틸화, 미리스톨화, 산화, 페그화, 단백질 분해 가공, 인산화, 프레닐화, 라세미화, 셀레노일화, 황산화 및 운반-RNA를 매개로 하는 아미노산의 단백질로의 첨가 반응 예를 들어, 아르기닐화를 포함한다[예를 들어, Creighton, T.E., Proteins - Structure and Molecular Properties 2nd Ed., W.H. Freeman and Company, New York (1993); Posttranslational Covalent Modification of Proteins, B.C. Johnson, Ed., Academic Press, New York, pp. 1-12 (1983)].

[0465] 고상 화학적 펩티드 합성법은 또한 본 발명의 폴리펩티드 또는 단편을 합성하는데에 사용될 수도 있다. 이러한 방법은 1960년대 초부터 당 업계에 공지되어 왔으며[Merrifield, R. B., J. Am. Chem. Soc., 85:2149-2154, 1963, 및 Stewart, J. M. and Young, J. D., Solid Phase Peptide Synthesis, 2nd Ed., Pierce Chemical Co., Rockford, Ill., pp. 11-12)], 최근에는 시판중인 실험실용 펩티드의 디자인과 합성용 키트에 사용되고 있다 [Cambridge Research Biochemicals]. 이와 같은 시판중인 실험실용 키트는 일반적으로 문헌[H.M. Geysen의 다수, Proc. Natl. Acad. Sci., USA, 81:3998 (1984)]에 교시된 사항을 활용한 것으로서, 단일 평판에 연결되어 있는 다수의 "막대" 또는 "핀"의 끝 부분에서 펩티드를 합성할 수 있는 것이다. 이러한 시스템이 사용될 경우, 막대 또는 핀이 존재하는 평판은 거꾸로 뒤집혀 이평판 상의 막대나 핀과 부합하는 웰 또는 용기가 존재하는 제 2의 평판에 삽입되는데, 이때 상기 웰 또는 용기에는 적당한 아미노산을 핀이나 막대의 끝 부분에 부착시키거나 고정하기 위한 용액이 담겨져 있다. 이러한 공정에 따른 단계를 반복함으로써, 즉, 상기 막대 및 핀의 끝 부분을 거꾸로 하여 적당한 용액에 담그는 과정을 반복함으로써, 아미노산은 원하는 펩티드로 구성된다. 뿐만 아니라, 다수의 시판중인 Fmoc 펩티드 합성 시스템이 시판되고 있다. 예를 들어, 폴리펩티드 또는 단편의 조립은 어플라이드 바이오시스템즈 인코포레이션(Applied Biosystems, Inc.)의 모델 431 에이(Model 431 A)TM 자동화 펩티드 합성기를 사용하여 고상 지지체상에서 수행될 수 있다. 이러한 장치는 기타 공지의 기술을 사용하여 커플링될 수 있는 일련의 단편을 합성하거나 또는 펩티드를 직접 합성하여 본 발명의 펩티드를 용이하게 구할 수 있게 해준다.

[0466] 본 발명의 폴리펩티드로서는 활성 형태 또는 비활성 형태인 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 포함한다. 예를 들어, 본 발명의 폴리펩티드는 예를 들어, "활성" 성숙 단백질을 생산하는 프로 단백질-가공 효소 예를 들어, 프로 단백질 전환 효소에 의해 프리프로 서열을 "성숙"시키기 전 또는 가공하기 전의 프로 단백질을 포함한다. 본 발명의 폴리펩티드로서는 기타의 이유 예를 들어, 번역 후 가공 예를 들어, 엔도- 또는 엑소-펩티다제 또는 단백질 분해 효소의 작용, 인산화 공정, 아마이드화, 글리코실화 또는 황산화, 이량체화 등에 의해 "활성화"되기 이전이기 때문에 비활성인 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 포함한다. 본 발명의 폴리펩티드는 효소의 모든 활성 형태의 것 예를 들어, 활성 부분 서열 예를 들어, 효소의 촉매 도메인 또는 활성 부위를 포함한다.

[0467] 본 발명은 고정화된 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소, 항-셀룰라제 예를 들어, 항-엔도글루카나제, 항-셀로바이오히이드롤라제 및/또는 항-베타-글루코시다제 항체 및 이의 단편을 포함한다. 본 발명은 예를 들어, 본 발명의 우성 네가티브 돌연변이체 또는 항-셀룰라제 예를 들어, 항-엔도글루카나제, 항-셀로바이오히이드롤라제 및/또는 항-베타-글루코시다제 항체를 사용하여, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성을 억제하는 방법을 제공한다. 본 발명은 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 포함하는 이중 복합체 예를 들어, 융합 단백질 및 이중이량체 등을 포함한다.

- [0468] 본 발명의 폴리펩티드는 다양한 조건 예를 들어, 극한의 pH 및/또는 온도와 산화제가 존재하는 등의 조건 하에서 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성을 가질 수 있다. 본 발명은 상이한 촉매 효과 예를 들어, 온도, 산화제의 존재 및 세척 조건 교환시에 안정성을 가지는 대안적인 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 제제를 생산하는 방법을 제공한다. 하나의 측면에서, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 변이체는 위치 지정 돌연변이 유발법 및/또는 랜덤한 돌연변이 유발법과 같은 기술을 이용하여 생산될 수 있다. 하나의 측면에서, 유도 진화는 대안적인 특이성과 안정성을 가지는 다양한 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 변이체를 생산하는데에 사용될 수 있다.
- [0469] 본 발명의 단백질은 또한 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 조정제 예를 들어, 이 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성의 활성화제 또는 억제제를 동정하는 연구용 시약으로서 유용하다. 간단히 말해서, 시험 시료(화합물, 발효액 및 추출물 등)는 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 검정법에 첨가되고, 그 결과 이 효소들이 기질 분해를 억제하는 능력이 평가된다. 이러한 방식으로 동정된 억제제는 원치않는 단백 분해를 감소 또는 방지하기 위한 연구 및 산업 분야에 사용될 수 있다. 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 사용할 때, 억제제는이에 결합하여 활성 범위를 증폭시킬 수 있게 된다.
- [0470] 본 발명의 효소는 또한 단백질을 분해하거나 또는 단백질의 서열 결정에 있어서 연구용 시약으로서도 유용하다. 예를 들어, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소는 이 폴리펩티드를 보다 작은 단편들로 분해하는데에 사용될 수 있으며, 그 결과, 예를 들어, 자동화 서열 결정기를 사용하여 서열 결정할 수 있게 된다.
- [0471] 본 발명은 또한 본 발명의 핵산, 폴리펩티드 및 항체를 사용하여 신규의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 발견하는 방법도 제공한다. 하나의 측면에서, 파지미드 라이브러리는 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 발현 여부를 바탕으로 하여 스크리닝된다. 다른 측면에서, 람다 파지 라이브러리는 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 발현 여부를 바탕으로 하여 스크리닝된다. 파지 또는 파지미드 라이브러리의 스크리닝을 통하여, 독성인 클론을 검출할 수 있고; 기질로의 접근 방식을 개선할 수 있으며; 라이브러리를 다량으로 분해함으로 인해 발생할 수 있는 오차의 발생 가능성을 없앴으로써 숙주를 조작할 필요성을 감소시킬 수 있고; 또한 클론의 밀도가 낮을 경우 이 클론을 더 신속하게 성장시킬 수도 있다. 파지 또는 파지미드 라이브러리의 스크리닝은 액상 또는 고상에서 진행될 수 있다. 하나의 측면에서, 본 발명은 액상 스크리닝법을 제공한다. 이로써 검정 조건을 보다 융통성 있게 설정할 수 있으며; 기질도 더욱 융통성 있게 선택할 수 있고; 약한 클론에 대한 선택성을 한층 더 높일 수 있으며; 또한 고상 스크리닝에 대한 자동화를 용이하게 만들 수 있다.
- [0472] 본 발명은 본 발명의 단백질 및 핵산과 수 천개의 생체 촉매 반응을 실행할 수 있는 로봇 자동화 시스템을 사용하는 스크리닝 방법과, 단기간 예를 들어, 하루 동안 진행되는 스크리닝 검정법, 그리고 정확성과 재현 가능성의 수준을 높일 수 있는 스크리닝 방법을 제공한다(이하, 어레이에 관한 논의 참조). 결과적으로, 유도체 화합물 라이브러리는 몇 주만 지나면 얻을 수 있게 된다. 분자 예를 들어, 소형 분자의 변형에 관한 추가의 교시에 관하여는 PCT/US94/09174 및 미국 특허 제6,245,547호를 참조하시오.
- [0473] 하나의 측면에서, 본 발명의 폴리펩티드 또는 단편은 생화학적 증폭 방법 또는 정제 방법을 통하여 얻어진다. 상동성일 수 있는 폴리펩티드 또는 이의 단편은, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 검정법(예를 들어, 이하, 실시예 1, 2 및 3 참조), 겔 전기 영동법 및/또는 미세 서열 결정법에 의해 결정될 수 있다. 본 발명의 각 폴리펩티드 또는 이의 단편 서열은 전술한 프로그램 중 임의의 것을 이용하여, 본 발명의 대표적인 폴리펩티드, 또는 예를 들어, 약 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100 또는 150개 이상의 연속적 아미노산을 포함하는 이의 단편과 비교될 수 있다.
- [0474] 본 발명의 다른 측면은 본 발명의 폴리펩티드의 효소적 기능을 보유하는 본 발명의 단편 또는 변이체를 동정하는 검정법에 관한 것이다. 예를 들어, 상기 폴리펩티드의 단편 또는 변이체는 생화학 반응을 촉진하는데에 사용

될 수 있는데, 이는 곧, 상기 단편 또는 변이체가 본 발명의 폴리펩티드의 효소적 활성을 보유함을 시사하는 것이다. 변이체 단편이 본 발명의 폴리펩티드의 효소적 활성을 보유하는지 여부를 결정하기 위한 대표적인 검정법은 다음의 단계들을 포함한다: 폴리펩티드 단편 또는 변이체가 기능을 할 수 있는 조건 하에서 이 폴리펩티드 단편 또는 변이체를 기질 분자와 접촉시키는 단계, 및 기질 수준의 감소 또는 폴리펩티드와 기질 사이의 반응을 통한 특정 반응 생성물 수준의 증가를 확인하는 단계.

[0475] 본 발명은 효소의 독특한 촉매적 특성을 찾아낸다. 생체 촉매(즉, 정제 또는 미정제 효소, 죽은 세포 또는 살아 있는 세포)를 화학적 형질전환법에서 사용할 경우에는 일반적으로 특정의 출발 화합물과 반응하는 특정의 생체 촉매를 동정하는 과정이 필요하지만, 본 발명은 다수의 출발 화합물 예를 들어, 소형 분자에 존재하는 작용기에 특이적인 반응 조건 및 선택된 생체 촉매를 이용한다. 각각의 생체 촉매는 하나의 작용기 또는 몇몇 관련된 작용기에 특이적인 것으로서, 이러한 작용기를 함유하는 다수의 출발 화합물과 반응할 수 있다.

[0476] 하나의 측면에서, 생체 촉매 반응은 단일의 출발 화합물로부터 유도체 군을 생산한다. 이러한 유도체는 생체 촉매 반응의 다른 라운드에 참가하여, 유도체 화합물의 제2의 군을 생성할 수 있다. 기원 소형 분자 또는 화합물에 일어나는 수천 가지의 변이는 각각의 생체 촉매 유도체화를 진행시킴으로써 일어날 수 있다.

[0477] 효소는 이와 같은 분자의 나머지 과정 즉, 통상의 화학적 방법을 통해서는 이루기 매우 어려운 과정에 영향을 미치지 않고서 출발 화합물의 특정 위치에서 반응한다. 이와 같이 생체 촉매 특이성을 높임으로써, 라이브러리 내에 존재하는 단일의 활성 화합물을 동정하는 수단이 제공된다. 이 라이브러리는 이것을 생산하는데 이용되는 일련의 생체 촉매 반응 즉, "생합성 과정(biosynthetic history)"에 의해 특정 규명된다. 생물 활성에 대해 라이브러리를 스크리닝하고, 생합성 과정을 추적함으로써, 활성 화합물을 생산하는 특이적인 반응의 순서를 확인할 수 있다. 반응 순서는 반복되고, 합성된 화합물의 구조는 측정된다. 이러한 방식의 확인법은 기타 합성법 및 스크리닝 방법과는 달리, 고정화 기술을 필요로 하지 않으며, 화합물은 실질적으로 임의의 유형의 스크리닝 검정법을 사용하여 용액 중에서 자유롭게 합성된 후 시험될 수 있다. 작용기 상에서의 효소 반응 특이성 정도가 높으면 생체 촉매 반응에 의해 생산된 라이브러리를 이루는 특정 효소 반응을 "이끌어" 낼 수 있다는 점에 주목해야 한다.

[0478] 하나의 측면에서, 공정에 따른 단계는 수천 가지의 생체 촉매 반응 및/또는 스크리닝 검정법을 하루 만에 수행할 수 있고, 정확도와 재현 가능성 수준을 높일 수 있는 로봇 자동화법을 이용하여 진행된다. 로봇 자동화법은 또한 어떤 폴리펩티드가 본 발명의 범위 내에 포함되는지 여부를 결정하기 위해 셀룰라제 활성에 대해 스크리닝하는데 사용될 수도 있다. 결과적으로, 하나의 측면에서, 유도체 화합물 라이브러리는 "통상의" 화학적 또는 효소적 스크리닝 방법을 이용하여 생산할 경우에는 수년이 걸릴 수 있는 유도체 화합물을 수주일 이내에 생산할 수 있도록 만들어준다.

[0479] 특정 측면에서, 본 발명은 소형 분자를 변형시키는 방법을 제공하는데, 이 방법은 본원에 기술된 폴리뉴클레오티드에 의해 암호화되는 폴리펩티드 또는 이의 효소적으로 활성인 단편을 소형 분자와 접촉시켜 변형된 소형 분자를 생산하는 단계를 포함한다. 변형된 소형 분자 라이브러리는, 원하는 활성을 나타내는 변형된 소형 분자가 라이브러리 내에 존재하는지 여부를 확인하기 위해 시험된다. 원하는 활성을 보유하는 변형된 소형 분자를 생산하는 특이적 생체 촉매 반응은 라이브러리의 일부를 생산하는데 사용되는 생체 촉매 반응을 각각 체계적으로 생략한 후, 라이브러리의 일부에서 생산되는 소형 분자를, 원하는 활성을 보유하는 변형된 소형 분자가 존재하는지 아니면 존재하지 않는지 여부에 대해 시험함으로써 확인된다. 원하는 활성을 보유하는 변형된 소형 분자를 생산하는 특이적 생체 촉매 반응은 반복 수행되기도 한다. 생체 촉매 반응은 소형 분자의 구조 내에서 살펴볼 수 있는 분명한 구조적 부분과 반응하는 생체 촉매 군과 함께 진행되며, 여기서, 각각의 생체 촉매는 하나의 구조적 부분 또는 관련 구조 부분의 군에 특이적이고; 각각의 생체 촉매는 분명한 구조적 부분을 함유하는 다수의 상이한 소형 분자와 반응한다.

[0480] 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제 및/또는 베타-글루코시다제 효소 신호 서열, 프리프로 도메인 및 촉매 도메인

[0481] 본 발명은 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 신호 서열(예를 들어, 신호 펩티드(SP)), 프리프로 도메인(prepro domain) 및 촉매 도메인(CD)을 제공한다. 상기 본 발명의 SP, 프리프로 도메인 및/또는 CD는 분리된 또는 재조합 펩티드일 수 있거나, 또는 예를 들어, 융합 단백질의 일부 예를 들어, 키메라 단백질 내 이중 도메인 일 수 있다. 본 발명은 이러한 촉매 도메인(CD), 프리프로 도메인 및 신호 서열(SP, 예를 들어, 본 발명의 폴리펩티드의 아미노 말단 잔기를 포함하거나/이로 이루어져 있는 서열을 가지는 펩티드)을 암호화하는 핵산을 제공한다.

- [0482] 본 발명은 본 발명의 폴리펩티드, 예를 들어 본 발명의 대표적인 폴리펩티드의 1~14, 1~15, 1~16, 1~17, 1~18, 1~19, 1~20, 1~21, 1~22, 1~23, 1~24, 1~25, 1~26, 1~27, 1~28, 1~29, 1~30, 1~31, 1~32, 1~33, 1~34, 1~35, 1~36, 1~37, 1~38, 1~40, 1~41, 1~42, 1~43, 1~44, 1~45, 1~46 또는 1~47 번 잔기 또는 그 이상에 제시된 바와 같은 서열로 이루어져 있거나 또는 이 서열을 포함하는 분리된 또는 재조합 신호 서열(예를 들어, 신호 펩티드)을 제공한다[이하, 표 3, 실시예 1 및 4, 그리고 서열 목록 참조]. 예를 들어, 상기 표 3a~3e에는 본 발명의 대표적인 신호(리더) 서열이 제시되어 있는데, 예를 들어, 서열 번호 164에 제시된 서열을 가지는 폴리펩티드 예를 들어, 서열 번호 163에 의해 암호화되는 폴리펩티드는 아미노 말단에 있는 30개의 잔기 즉, MSCRTLMSSRRVWGGLLWGGLFLRTGSVTG를 포함하는(또는 이로써 이루어져 있는) 신호 서열을 가진다. 이와 유사하게 부가의 신호 서열은 표 3a~3e에 제시되어 있다.
- [0483] 하나의 측면에서, 본 발명은 본 발명의 폴리펩티드 중 처음 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69 또는 70개 이상의 아미노 말단 잔기를 포함하는 신호 서열을 제공한다.
- [0484] 본 발명은 신호 서열 및/또는 프리프로 서열을 가지거나 또는 가지지 않는 폴리펩티드를 포함한다. 본 발명은 이중 신호 서열 및/또는 프리프로 서열을 가지는 폴리펩티드를 포함한다. 상기 프리프로 서열(이중 프리프로 도메인으로 사용되는 본 발명의 서열 포함)은 단백질의 아미노 말단 또는 카복시 말단에 위치할 수 있다. 본 발명은 또한 본 발명의 서열을 포함하는 분리된 또는 재조합 신호 서열, 프리프로 서열 및 촉매 도메인(예를 들어, "활성 부위")을 포함한다. 본 발명의 신호 서열을 포함하는 폴리펩티드는 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소일 수 있거나, 또는 다른 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소일 수 있거나, 또는 다른 효소나 다른 폴리펩티드일 수 있다. "프리프로" 도메인 서열 및 신호 서열을 동정하는 방법에 관하여는 당 업계에 널리 공지되어 있다[예를 들어, Van de Ven (1993) Crit. Rev. Oncog. 4(2): 115-136 참조]. 예를 들어, 프리프로 서열을 동정하기 위해서, 단백질은 세포 외 공간으로부터 정제되고, N-말단 단백질 서열은 결정되어 가공되지 않은 형태의 것과 비교된다.
- [0485] 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 신호 서열(SP) 및/또는 프리프로 서열은 분리된 또는 재조합 펩티드일 수 있거나, 아니면 다른 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 또는 비-셀룰라제 예를 들어, 비-엔도글루카나제, 비-셀로바이오하이드롤라제 및/또는 비-베타-글루코시다제 폴리펩티드에 결합된 서열 예를 들어, 융합(키메라) 단백질일 수 있다. 하나의 측면에서, 본 발명은 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 신호 서열을 포함하는 폴리펩티드를 제공한다. 하나의 측면에서, 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 신호 서열인 SP 및/또는 프리프로 서열을 포함하는 폴리펩티드는 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소에 이중인 서열을 포함한다[예를 들어, 본 발명의 SP 및/또는 프리프로 서열, 및 다른 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소로부터 유래하는 서열이나, 비-셀룰라제 예를 들어, 비-엔도글루카나제, 비-셀로바이오하이드롤라제 및/또는 비-베타글루코시다제 단백질로부터 유래하는 서열을 포함하는 융합 단백질의 형태임]. 하나의 측면에서, 본 발명은 이중 SP 및/또는 프리프로 서열 예를 들어, 효소의 신호 서열을 가지는 서열을 포함하는 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 제공한다. 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소는 벡터 예를 들어, pPIC 시리즈 벡터(Invitrogen, Carlsbad, CA) 내에 이중 SP 및/또는 프리프로 서열을 포함할 수 있다.
- [0486] 하나의 측면에서, 본 발명의 SP 및/또는 프리프로 서열은 다음과 같이 신규의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 폴리펩티드를 동정함으로써 확인된다. 단백질이 분류되고 적당한 세포 내 위치로 이동되는 경로를 종종 단백질 표적화 경로(protein targeting pathway)라고 부른다. 이와 같이 표적화 시스템 전부에 있어서 가장 중요한 요소 중 하나는 새로 합성된 폴리펩티드의 아미노 말단에 있는 짧은 아미노산 서열(소위 신호 서열이라고 불리는 서열)이다. 이 신호 서열은 단백질을 세포 내 적당한 위치로 유도하고, 이동 중 제거되거나, 또는 단백질이 그 목적지에 도달할 때에 제거된다. 대부분의 리소솜, 막, 또는 분비형 단백질은 이를 세포질 내 망상 조직 강으로 이동시키는 아미노-말단 신호 서

열을 가진다. 신호 서열은 길이 약 10~65 이상의 아미노산 잔기로서 다양할 수 있다. 신호 서열을 인지하는 다양한 방법에 관하여는 당 업자에게 공지되어 있다. 예를 들어, 하나의 측면에서, 신규의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 신호 펩티드는 시그널 P(SignalP)라고 불리는 방법에 의해 동정된다. 시그널P는 신호 펩티드 및 이의 절단 위치를 모두 인지하는 통합된 신경망을 이용한다[Nielsen (1997) "Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites." Protein Engineering 10:1-6].

[0487] 몇몇 측면에서, 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소는 SP 및/또는 프리프로 서열 또는 "도메인"을 가지지 않는다. 하나의 측면에서, 본 발명은 SP 및/또는 프리프로 도메인의 전부 또는 일부가 결합된 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 제공한다. 하나의 측면에서, 본 발명은 상이한 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 핵산 서열에 작동 가능하도록 결합되어 있는, 하나의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소로부터 유래하는 신호 서열(SP) 및/또는 프리프로 서열을 암호화하는 핵산 서열을 제공하며, 바람직하게는 비-셀룰라제 예를 들어, 비-엔도글루카나제, 비-셀로바이오히이드롤라제 및/또는 비-베타-글루코시다제 단백질로부터 유래하는 신호 서열(SP) 및/또는 프리프로 도메인을 암호화하는 핵산 서열을 제공할 수도 있다.

[0488] 본 발명은 또한 본 발명의 신호 서열(SP), 프리프로 도메인 및/또는 촉매 도메인(CD)과 이종 서열을 포함하는 분리된 또는 재조합 폴리펩티드를 제공한다. 상기 이종 서열은 자연적으로는 (예를 들어, 효소에) SP, 프리프로 도메인 및/또는 CD와 함께 결합되어 있지 않다. 천연의 상태에서 SP, 프리프로 도메인 및/또는 CD와 결합되어 있지 않은 서열은 SP, 프리프로 도메인 및/또는 CD의 아미노 말단, 카복시 말단, 및/또는 SP 및/또는 CD의 양 말단부 상에 존재할 수 있다. 하나의 측면에서, 본 발명은 본 발명의 신호 서열(SP), 프리프로 도메인 및/또는 촉매 도메인(CD)을 포함하는 폴리펩티드를 포함하는(또는 이로써 이루어져 있는) 분리된 또는 재조합 폴리펩티드를 제공하는데, 단, 여기서, 이 폴리펩티드는 천연의 상태에서 결합되어 있는 임의의 서열(예를 들어, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 서열)과 결합되어 있지 않다. 이와 유사하게, 하나의 측면에서, 본 발명은 이러한 폴리펩티드를 암호화하는 분리된 또는 재조합 핵산을 제공한다. 그러므로, 하나의 측면에서, 본 발명의 분리된 또는 재조합 핵산은 본 발명의 신호 서열(SP), 프리프로 도메인 및/또는 촉매 도메인(CD)의 암호화 서열, 그리고 이종 서열(즉, 자연적으로는 본 발명의 신호 서열(SP), 프리프로 도메인 및/또는 촉매 도메인(CD)과 결합되어 있지 않는 서열)을 포함한다. 이종 서열은 SP, 프리프로 도메인 및/또는 CD 암호화 서열의 3' 말단, 5' 말단 및/또는 양 말단에 존재할 수 있다.

[0489] *하이브리드(키메라) 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제 및/또는 베타-글루코시다제 효소 및 펩티드 라이브러리*

[0490] 하나의 측면에서, 본 발명은 본 발명의 서열을 포함하는 하이브리드 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 및 융합 단백질 예를 들어, 펩티드 라이브러리를 제공한다. 본 발명의 펩티드 라이브러리는 표적 예를 들어, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 기질, 수용체 및 효소의 펩티드 조절제(예를 들어, 활성화 인자 또는 억제제)를 분리하는데에 사용될 수 있다. 본 발명의 펩티드 라이브러리는 표적의 형태상 결합 파트너 예를 들어, 리간드 예를 들어, 시토킨 및 호르몬 등을 동정하는데에 사용될 수 있다. 하나의 측면에서, 본 발명은 본 발명의 신호 서열(SP), 프리프로 도메인 및/또는 촉매 도메인(CD)이나, 상기의 것들의 조합체와 이종 서열을 포함하는 키메라 단백질을 제공한다(상기 참조).

[0491] 하나의 측면에서, 본 발명의 융합 단백질(예를 들어, 펩티드 부분)은 (선형 펩티드에 비해) 형태적으로 안정화되어 표적에 대한 결합 친화도가 더욱 높을 수 있다. 본 발명은 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소와 기타 펩티드 예를 들어, 공지의 펩티드 및 랜덤한 펩티드로 이루어진 융합체를 제공한다. 이 융합체를 구성하는 요소들은 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 구조를 거의 망가뜨리지 않는 방식으로 융합될 수 있으며, 펩티드는 대사적으로 또는 구조적으로 또는 형태적으로 안정화된다. 이로써, 세포 내에 융합체가 존재하는지 여부와 이 융합체의 양을 용이하게 모니터링할 수 있는 펩티드 라이브러리를 제조할 수 있다.

[0492] 본 발명의 아미노산 서열 변이체는 일정한 변이 특징 즉, 이 변이체를 자연 발생된 형태와 동떨어지게 만드는

특정 예를 들어, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 서열의 대립 형질 또는 중간 변이에 따른 특징에 의해 특정 규명될 수 있다. 하나의 측면에서, 본 발명의 변이체는 자연 발생 유사체와 동일한 성질의 생물 활성을 나타낸다. 대안적으로, 상기 변이체는 변형된 특성을 가지는지 여부에 대하여 선택될 수 있다. 하나의 측면에서, 아미노산 서열 변이를 도입시키기 위한 위치나 부위가 예정될 경우, 돌연변이 자체는 예정되어 있을 필요가 없다. 예를 들어, 정해진 위치에서의 돌연변이를 최적으로 수행하기 위해서는, 표적 코돈이나 부위에서 랜덤한 돌연변이 유발법이 수행될 수 있는 것이며, 발현된 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 변이체는 원하는 활성을 최적으로 조합한 활성을 보유하는지 여부에 대하여 스크리닝될 수 있다. 본원에 기술된 바와 같이, 공지의 서열을 가지는 DNA 내 예정된 위치에 치환 돌연변이를 일으키는 기술은 널리 공지되어 있으며, 그 예로서는 M13 프라이머 돌연변이 유발법 및 PCR 돌연변이 유발법이 있다. 돌연변이체를 스크리닝하는 방법은 예를 들어, 글루칸 가수분해 검정법을 이용하여 수행될 수 있다. 대안적 측면에서, 아미노산 치환은 단일 잔기에서 일어날 수 있으며; 삽입은 약 1~20개의 아미노산에서 일어날 수 있는데, 다만, 더욱 큰 부분이 삽입될 수도 있다. 결실은 약 1~20, 30, 40, 50, 60 또는 70개 이상의 잔기에서 일어날 수 있다. 최적의 특성을 가지는 최종 유도체를 얻기 위해서는, 치환, 결실, 삽입 또는 이러한 현상을 경우에 따라 조합하여 적용시킬 수 있는 것이다. 일반적으로, 이러한 변화들은 분자가 최소한으로 변형된 소수의 아미노산에서 발생된다. 그러나, 더욱 큰 변화는 임의의 환경에서 관용될 수 있다.

[0493] 본 발명은 폴리펩티드 주쇄 구조, 2차 또는 3차 구조 예를 들어, 알파-나선 구조 또는 베타-병풍 구조가 변형된, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 제공한다. 하나의 측면에서, 이러한 변화 또는 소수성은 변성될 수 있다. 하나의 측면에서, 측쇄 덩어리는 변형된다. 기능 또는 면역원성에 일어난 실질적인 변화는 덜 보존적인(less conservative) 치환부를 선택함으로써 일어나게 된다. 예를 들어, 변형이 일어난 구역에 존재하는 폴리펩티드 주쇄의 구조 예를 들어, 알파-나선 구조 또는 베타-병풍 구조; 활성 위치에 있을 수 있는 분자의 전하 또는 소수성 위치; 또는 측쇄에 더욱 많은 영향을 미치는 치환이 일어날 수 있다. 본 발명은 (a) 소수성 잔기 예를 들어, 세릴 또는 트레오닐이 소수성 잔기 예를 들어, 루실, 이소루실, 페닐알라닌, 발릴 또는 알라닌에 대해(또는 ~에 의해) 치환되는 경우; (b) 시스테인 또는 프롤린이 임의의 기타 잔기에 대해(또는 ~에 의해) 치환되는 경우; (c) 양전하를 띠는 측쇄를 가지는 잔기 예를 들어, 리실, 아르기닌 또는 히스티딘이 음전하를 띠는 잔기 예를 들어, 글루타미나 아스파틸에 대해(또는 ~에 의해) 치환되는 경우; 또는 (d) 벌키한 측쇄를 가지는 잔기 예를 들어, 페닐알라닌이 측쇄를 가지지 않는 아미노산 예를 들어, 글리신에 대해(또는 ~에 의해) 치환될 경우, 본 발명의 폴리펩티드를 치환시킨다. 변이체는 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 특성이 필요에 따라서 변성되도록 선택될 수 있지만, 이 변이체는 동일한 특징의 생물 활성(즉, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성)을 나타낼 수 있다.

[0494] 하나의 측면에서, 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소는 에피토프 또는 정제용 태그, 신호 서열 또는 기타 융합 서열 등을 포함한다. 하나의 측면에서, 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소는 랜덤 펩티드에 융합되어 융합 폴리펩티드를 형성할 수 있다. 본원에 있어서, "융합된" 또는 "작동 가능하도록 결합된"이란, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 구조의 안정성이 파괴되는 것을 최소화하도록, 예를 들어, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성을 보유하도록, 랜덤 펩티드 및 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소가 함께 결합되어 있음을 의미하는 것이다. 상기 융합 폴리펩티드(또는 이 융합 폴리펩티드를 암호화하는 융합 폴리뉴클레오티드)는 추가의 성분 예를 들어, 다수의 루프에 존재하는 다수의 펩티드를 포함할 수도 있다.

[0495] 하나의 측면에서, 펩티드와 이 펩티드를 암호화하는 핵산은 완전히 랜덤한 방식으로, 또는 예를 들어, 일반적으로는 뉴클레오티드/잔기가 출현할 때마다 즉, 이것이 존재하는 위치마다 랜덤화가 편중되어 이루어지는 방식으로 랜덤화된다. "랜덤화된"이란, 각각의 핵산 및 펩티드가 각각 본질적으로 랜덤한 뉴클레오티드 및 아미노산으로 이루어져 있음을 의미하는 것이다. 하나의 측면에서, 펩티드를 합성하는 핵산은 화학적으로 합성될 수 있으므로, 임의의 위치에 임의의 뉴클레오티드를 통합할 수 있다. 그러므로, 핵산이 발현되어 펩티드를 형성할 때, 임의의 아미노산 잔기는 임의의 위치에 통합될 수 있다. 합성 과정은 랜덤화된 핵산을 생산하고, 일정 길이의 핵산에 걸쳐서 생성될 수 있는 핵산들의 조합 모두 또는 대부분을 형성하여, 랜덤화된 핵산으로 이루어진 라이브러리를 생성할 수 있도록 디자인될 수 있다. 이 라이브러리는 랜덤화된 발현 생산물의 구조적으로 매우 다양

한 군을 형성하여, 확률적으로 충분한 범위의 세포 반응에 영향을 미쳐 원하는 반응을 나타내는 하나 이상의 세포를 제공할 수 있다. 그러므로, 본 발명은 충분히 큰 상호 작용 라이브러리(interaction library)를 제공하므로, 라이브러리 일원 중 하나 이상은 몇몇 분자, 단백질 또는 기타 인자에 대한 친화성을 가지도록 만드는 구조를 가지게 될 것이다.

[0496] 하나의 측면에서, 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소는 단일 펩티드, 탄수화물 결합 분자, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 촉매 도메인, 링커 및/또는 다른 촉매 도메인을 포함하는 다중 도메인 효소(multidomain enzyme)이다.

[0497] 본 발명은 생물학적으로 활성인 하이브리드 폴리펩티드(예를 들어, 하이브리드 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소)를 암호화할 수 있는 키메라 폴리펩티드를 생산하는 방법 및 서열을 제공한다. 하나의 측면에서, 기원 폴리뉴클레오티드(예를 들어, 본 발명의 대표적인 핵산)는 생물학적으로 활성인 폴리펩티드를 암호화한다. 하나의 측면에서, 본 발명의 방법은 기원 폴리뉴클레오티드 서열을 통합하여, 결과로 생성된 하이브리드 폴리뉴클레오티드가 생물학적으로 활성인 기원 폴리펩티드(예를 들어, 본 발명의 셀룰라제 또는 항체)로부터 유래하는(그러나, 상이한) 활성을 나타내는 폴리펩티드를 암호화하도록 만드는, 세포 내 과정을 이용함으로써 신규의 하이브리드 폴리펩티드를 생산한다. 예를 들어, 기원 폴리뉴클레오티드는 상이한 미생물로부터 유래하거나 또는 이 미생물에서 관찰되는 특정 효소(예를 들어, 셀룰라제)를 암호화할 수 있다. 하나의 유기체로부터 유래하는 제1 폴리뉴클레오티드 또는 이의 변이체에 의해 암호화되는 효소는 예를 들어, 특정 환경 조건 예를 들어, 고 염도 하에서 효율적으로 작용할 수 있다. 상이한 유기체로부터 유래하는 제2 폴리뉴클레오티드 또는 이의 변이체에 의해 암호화되는 효소는 상이한 환경 조건 예를 들어, 초고온 하에서 효율적으로 작용할 수 있다. 상기 제1 및 제2 기원 폴리뉴클레오티드로부터 유래하는 서열들을 함유하는 하이브리드 폴리뉴클레오티드는 기원 폴리뉴클레오티드들에 의해 암호화되는 두 가지 효소들의 특성을 모두 나타내는 효소를 암호화할 수 있다. 그러므로, 본 발명의 하이브리드 폴리뉴클레오티드에 의해 암호화되는 효소는 상기 제1 및 제2 폴리뉴클레오티드에 의해 암호화되는 각각의 효소가 공유하는 환경 조건 예를 들어, 고 염도 및 초고온 하에서 효율적으로 작용할 수 있다.

[0498] 하나의 측면에서, 본 발명의 방법에 의해 생산되는 하이브리드 폴리펩티드는 기원 효소에서는 살펴볼 수 없는 특화된 효소 활성을 나타낼 수 있다. 예를 들어, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 암호화하는 폴리뉴클레오티드의 재조합 및/또는 환원적 재분류법 수행 후, 결과로 생성되는 하이브리드 폴리뉴클레오티드에 의해 암호화되는 하이브리드 폴리펩티드는, 기원 효소 각각으로부터 유래하는 특화된 비-셀룰라제 예를 들어, 비-엔도글루카나제, 비-셀로바이오하이드롤라제 및/또는 비-베타-글루코시다제 효소 활성 예를 들어, 가수분해 효소, 펩티다제, 포스포릴라제 등과 같은 활성에 대해 스크리닝될 수 있다. 하나의 측면에서, 하이브리드 폴리펩티드는 기원 모 폴리펩티드와 하이브리드 폴리펩티드를 구별할 수 있는 화학적 작용 특성 예를 들어, 하이브리드 폴리펩티드가 작용을 하는 온도, pH 또는 염 농도에서 작용을 하는지 여부를 확인하기 위해 스크리닝된다.

[0499] 하나의 측면에서, 본 발명은 생물학적으로 활성인 하이브리드 폴리펩티드를 생산하는 방법 및 이러한 폴리펩티드가 강화된 활성을 나타내는지 여부를 스크리닝하는 방법에 관한 것으로서, 이 방법은 다음의 단계들을 포함한다:

[0500] 1) 작동 가능하도록 결합되어 있는 적어도 제1 폴리뉴클레오티드와 작동 가능하도록 결합되어 있는 제2 폴리뉴클레오티드를 적당한 숙주 세포에 도입시키는 단계[여기서, 적어도 상기 제1 폴리뉴클레오티드 및 제2 폴리뉴클레오티드는 부분적으로 서열 동일성을 가지는 하나 이상의 부위를 공유함];

[0501] 2) 서열의 재조직을 촉진하여 작동 가능하도록 결합된 하이브리드 폴리뉴클레오티드를 생성하는 조건 하에서 숙주 세포를 생육시키는 단계;

[0502] 3) 하이브리드 폴리뉴클레오티드에 의해 암호화되는 하이브리드 폴리펩티드를 발현시키는 단계;

[0503] 4) 강화된 생물 활성의 확인을 촉진하는 조건 하에서 하이브리드 폴리펩티드를 스크리닝하는 단계; 및

[0504] 5) 하이브리드 폴리펩티드를 암호화하는 폴리뉴클레오티드를 분리하는 단계.

[0505] 셀룰라제 효소의 분리 및 발견

[0506] 본 발명은 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제

효소와, 이 효소를 암호화하는 핵산을 분리 및 발견하는 방법을 제공한다. 폴리뉴클레오티드 또는 효소는 각각의 유기체("분리주"), 한정된 배지 중에서 생육한 유기체 군("증폭 배양액") 또는 배양되지 않은 유기체("환경 시료")로부터 분리될 수 있다. 유기체는 예를 들어, 생체내 바이오패닝(biopanning)에 의해 분리될 수 있다(이하 참조). 환경 시료로부터 신규의 생물 활성을 암호화하는 폴리펩티드를 구하기 위해 배양액-독립적 연구법을 사용하는 방법을 통하여 생물 다양성을 가지는 미개발 자원에 접근할 수 있기 때문에, 이 방법이 가장 바람직하다. 폴리뉴클레오티드 또는 효소는 또한 다수의 유기체 중 임의의 하나 예를 들어, 박테리아로부터 분리될 수도 있다. 전체 세포에 더하여, 폴리뉴클레오티드 또는 효소는 또한 이러한 유기체 예를 들어, 박테리아 배양액으로부터 유래하는 미정제 효소 제제로부터 분리될 수도 있다.

[0507] "환경 라이브러리(environmental library)"란, 환경 시료로부터 생산되는 것으로서, 적당한 원핵 생물 숙주 내에서 증식될 수 있는 클로닝 벡터 내에 존재하는, 자연 발생 유기체 게놈의 집합체를 나타낸다. 클로닝된 DNA는 처음에 환경 시료로부터 직접 추출되기 때문에, 라이브러리는 순수한 배양액 중에서 생육할 수 있는 소량의 원핵 생물의 것으로 한정되지 않는다. 뿐만 아니라, 이러한 시료 중에 존재하는 환경적 DNA를 표준화시킴으로써, 기원 시료 중에 존재하는 모든 종으로부터 유래하는 DNA와 더욱 동일하게 만들 수 있다. 이로써 우세한 종에 비해 몇 배나 더 적게 생성되는 소량의 시료 구성 성분으로부터 원하는 유전자를 찾아내는 효율을 더 높일 수 있다.

[0508] 하나의 측면에서, 하나 이상의 미배양 미생물로부터 생산되는 유전자 라이브러리는 목적으로 하는 활성을 보유하는지 여부에 대해 스크리닝된다. 목적 생물 활성 분자를 암호화하는 유력한 경로는 유전자 발현 라이브러리 형태로 원핵 생물세포 내에 우선 포착된다. 하나의 측면에서, 목적으로 하는 활성을 암호화하는 폴리뉴클레오티드는 이러한 라이브러리로부터 분리되어 숙주 세포에 도입된다. 숙주 세포는 신규하거나 강화된 활성을 보유하는, 잠재적으로 활성인 생물 분자를 생산하는 환원적 재분류법 및/또는 재조합법을 촉진하는 조건 하에서 생육된다.

[0509] 생체내 바이오패닝법은 FACS계 기구 및 비-광학계(예를 들어, 자기계) 기구를 이용하여 수행될 수 있다. 하나의 측면에서, 복합 유전자 라이브러리는 전사된 RNA를 안정화시키는 요소를 함유하는 벡터와 함께 구성된다. 예를 들어, RNA의 전사된 부위에 측정하도록 디자인된 2차 구조 예를 들어, 헤어핀 구조를 이루는 서열을 포함시키면, 벡터의 안정성을 증가시킬 수 있고, 따라서 이 벡터의 세포 내 반감기도 늘릴 수 있게 된다. 바이오패닝 방법에 사용되는 프로브 분자는 프로브가 표적 분자에 결합할 때에만 형광성을 나타내는 리포터 분자로 표지화된 올리고뉴클레오티드로 이루어져 있다. 이와 같은 프로브는 몇몇 형질전환 방법 중 어느 하나를 이용하여 라이브러리로부터 재조합 세포 내에 도입된다. 상기 프로브 분자는 전사된 표적 mRNA에 결합하여, DNA/RNA 이중 이중체 분자를 형성한다. 프로브를 표적에 결합시키면, 스크리닝 방법 중 FACS 기계에 의해 검출 및 분류되는 형광 신호를 나타내게 될 것이다.

[0510] 하나의 측면에서, 서브클로닝은 목적 서열을 추가로 분리하기 위해 수행된다. 서브클로닝에 있어서, DNA의 일부는 (일반적으로 제한 효소에 의해) 증폭 및 분해되어, 원하는 서열을 잘라내게 되며, 원하는 서열은 수용 벡터에 결합되어 증폭된다. 서브클로닝의 각 단계에서, 구조 단백질을 암호화하는 DNA가 제거되지 않았음을 확인하기 위해서, 이 DNA의 일부를 목적으로 하는 활성을 보유하는지 여부에 대해 관찰한다. 삽입부는 서브클로닝의 임의의 단계에서 예를 들어, 겔 전기 영동법을 수행한 후 벡터에 결합시킴으로써 정제될 수 있거나, 또는 수용 벡터를 함유하는 세포 및 수용 벡터를 함유하지 않는 세포가 예를 들어, 수용 벡터를 함유하지 않는 세포를 사멸시킬 항생제를 함유하는 선택 배지 중에 있을 경우에 정제될 수 있다. cDNA 삽입부를 벡터에 서브클로닝시키는 구체적인 방법은 당 업계에 널리 공지되어 있다[Sambrook의 다수, Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory Press (1989)]. 다른 측면에서, 본 발명의 효소는 서브클론이다. 이러한 서브클론은 예를 들어, 길이, 돌연변이, 태그 또는 표지가 존재하는지에 있어서 모 클론과 상이할 수 있다.

[0511] 폴리뉴클레오티드가 발견, 분리 또는 제조될 수 있는 미생물로서는 원핵 생물 미생물 예를 들어, 진정 박테리아와 고 박테리아, 그리고 하등 진핵 미생물 예를 들어, 진균, 일부 조류 및 원생 동물을 포함한다. 폴리뉴클레오티드는 환경 시료로부터 발견, 분리 또는 생산될 수 있는데, 이러한 경우, 핵산은 유기체를 배양하지 않고서도 회수될 수 있거나, 또는 하나 이상의 배양 유기체로부터 회수될 수 있다. 하나의 측면에서, 이러한 미생물은 극한 미생물(extremophiles) 예를 들어, 호극고온성 미생물(hyperthermophiles), 호냉성 미생물(psychrophiles), 내냉성 미생물(psychrotrophs), 호염성 미생물(halophiles), 고압성 미생물(barophiles) 및 호산성 미생물(acidophiles)일 수 있다. 극한 미생물로부터 분리되는 효소를 암호화하는 폴리뉴클레오티드를 사용할 수 있다. 본 발명의 효소는 100℃ 이상의 온도에서 작용할 수 있는데, 그 예로서는, 지상 온천 및 심해 열수구에서 발견

되는 것, 또는 0℃ 이하의 온도에서 작용할 수 있는 것 예를 들어, 북극의 바다에서 발견되는 것, 포화 염 환경에서 작용할 수 있는 것 예를 들어, 사해에서 발견되는 것, pH 약 0에서 작용할 수 있는 것 예를 들어, 석탄 침적물 및 지열 황-풍부 온천에서 발견되는 것, 또는 pH 11 이상에서 작용할 수 있는 것 예를 들어, 하수 슬러지에서 발견되는 것이 있다. 하나의 측면에서, 본 발명의 효소는 광범위한 온도 및 pH에서 활성이 높다.

[0512] 전술한 바와 같이 선별 및 분리된 폴리뉴클레오티드는 적당한 숙주 세포에 도입된다. 적당한 숙주 세포는 재조합 및/또는 환원적 재분류법을 촉진할 수 있는 임의의 세포이다. 하나의 측면에서, 이와 같이 선별된 폴리뉴클레오티드는 적당한 제어 서열을 포함하는 벡터 내에 존재한다. 숙주 세포는 고등 진핵 생물 세포 예를 들어, 포유동물 세포이거나 하등 진핵 생물 세포 예를 들어, 효모 세포일 수 있거나, 또는 하나의 측면에서, 숙주 세포는 원핵 생물 세포 예를 들어, 박테리아 세포일 수 있다. 구조물을 숙주 세포에 도입하는 방법은 인산칼슘 형질 감염법, DEAE-텍스트란 매개 형질 감염법 또는 전기 천공법에 의해 실시될 수 있다.

[0513] 대표적인 숙주로서는 전술한 바와 같이, 박테리아 세포 예를 들어, 이. 콜라이, 스트렙토마이세스, 살모넬라 타이피뮤리움; 진균 세포 예를 들어, 효모; 곤충 세포 예를 들어, 드로소필라 S2 및 스포도프테라 Sf9; 동물 세포 예를 들어, CHO, COS 또는 바우어즈 흑색종; 아테노바이러스; 및 식물 세포를 포함한다. 본원에 교시하는 바에 따르면, 숙주를 선별하는 방법도 당 업자의 기술 범주에 속할 것이다.

[0514] 재조합 단백질을 발현하는데에 다양한 포유동물 세포 배양 시스템이 사용될 수 있으며; 포유 동물 발현 시스템의 예로서는 원숭이 신장 섬유아세포의 COS-7 세포주(Gluzman 1981: "SV40-transformed simian cells support the replication of early SV40 mutants")와, 적합한 벡터를 발현할 수 있는 기타 세포주 예를 들어, C127, 3T3, CHO, HeLa 및 BHK 세포주를 포함한다. 포유동물 발현 벡터는 복제 기원, 적당한 프로모터 그리고 인헨서를 포함할 수 있으며, 또한 임의의 필요한 리보솜 결합 위치, 폴리아데닐화 위치, 스플라이싱 공여 위치 및 수용 위치, 전사 종결 서열 및 5' 측접 비전사 서열을 포함할 수 있다. SV40 스플라이싱 위치 및 폴리아데닐화 위치로부터 유래하는 DNA 서열은 필요한 비전사 유전 요소를 제공하는데에 사용될 수 있다.

[0515] 다른 측면에서, 본 발명의 핵산, 폴리펩티드 및 방법은 생화학적 경로에 사용되거나, 또는 하나 이상의 오페론 또는 유전자 클러스터 또는 이의 일부로부터 생화학적 경로를 암호화하는 신규의 폴리뉴클레오티드를 생산하는데에 사용된다. 예를 들어, 박테리아 및 다수의 진핵 생물은 유전자 생산물이 서로 관련된 과정에 관여하는, 유전자 조절 합작 기구를 가진다. 유전자는 단일 염색체상에 존재하는 "유전자 클러스터"라고 불리는 구조에 모여 있으며, 또한 전체 클러스터의 전사를 개시하는 단일의 조절 서열 예를 들어, 단일의 프로모터의 제어하에 함께 전사된다. 그러므로, 유전자 클러스터는 일반적으로 기능이 동일하거나 또는 관련된 인접 유전자들의 군이다[유전자 클러스터에 의해 암호화되는 생화학적 경로의 예로서는 폴리케타이드(polyketide)가 있음].

[0516] 하나의 측면에서, 유전자 클러스터 DNA는 상이한 유기체로부터 분리되어 벡터 예를 들어, 절찰된 유전자 클러스터로부터 검출 가능한 단백질 또는 단백질-관련 어레이 활성의 발현을 제어 및 조절할 수 있는 발현 조절 서열을 함유하는 벡터에 절찰된다. 외인성 DNA 도입 능력이 매우 큰 벡터는 이러한 유전자 클러스터와 함께 사용하면 적당할 수 있고, 이. 콜라이의 f-인자(즉, 수정 인자)를 포함하는 것으로 본원에 기술되어 있다. 상기 이. 콜라이의 f-인자는 접합시 그 인자를 높은 빈도수로 이동시키는 플라스미드로서, 큰 DNA 절편 예를 들어, 혼합 미생물 시료로부터 유래하는 유전자 클러스터를 안정적으로 증식시키는데에 이상적이다. 본 발명의 하나의 측면은 클로닝 벡터("포스미드(fosmid)") 또는 박테리아 인공 염색체(BAC) 벡터를 사용하는 것이다. 이 벡터는 게놈형 DNA의 큰 분절을 안정적으로 통합시킬 수 있는 이. 콜라이 f-인자로부터 유래한다. 혼합 미배양 환경 시료로부터 유래하는 DNA와 통합할 때, 이러한 통합 과정은 큰 게놈 단편을 안정한 "환경적 DNA 라이브러리"의 형태로 만들 수 있다. 본 발명에 사용되는 다른 유형의 벡터로서는 코스미드 벡터가 있다. 코스미드 벡터는 원래 게놈형 DNA의 큰 분절을 클로닝하고 증식시키도록 디자인된 것이었다. 코스미드 벡터에 클로닝하는 방법에 관하여는 문헌[Sambrook 외 다수, Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory Press (1989)]에 상세히 기술되어 있다. 일단 적당한 벡터에 절찰되면, 상이한 폴리케타이드 합성 효소 유전자 클러스터를 함유하는 2개 이상의 벡터는 적당한 숙주 세포에 도입될 수 있다. 유전자 클러스터에 의해 공유되는 부분적 서열 동일성 부위는 서열을 재조직화하여 하이브리드 유전자 클러스터를 만드는 과정을 촉진할 것이다. 이후, 신규의 하이브리드 유전자 클러스터는 기원 유전자 클러스터에서는 살피볼 수 없는 활성이 강화되었는지 여부에 대해 스크리닝될 수 있다.

[0517] 다양한 효소 활성에 대해 스크리닝하는 방법은 당 업자에게 널리 공지되어 있으며, 본 발명의 상세한 설명을 통하여서도 기술하고 있다(예를 들어, 이하 실시예 1, 2 및 3 참조). 이러한 방법은 본 발명의 폴리펩티드 및 폴리뉴클레오티드를 분리할 때에 사용될 수 있다.

- [0518] 하나의 측면에서, 본 발명은 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 활성을 변형시키기 위해서 전 세포 접근 방법을 이용하여 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 또는 화합물을 발견 및 분리하는 방법을 제공한다(이하 참조). 게놈형 DNA 라이브러리로부터 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제를 암호화하는 추정 클론을 스크리닝할 수 있다.
- [0519] 스크리닝 방법 및 "온-라인" 모니터링 장치
- [0520] 예를 들어, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성에 대해 폴리펩티드를 스크리닝하거나, 유력한 조절제 예를 들어, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성의 활성화 인자 또는 억제제로서의 화합물을 본 발명의 폴리펩티드에 결합하는 항체 또는 본 발명의 핵산과 혼성화되는 핵산에 대해 스크리닝하거나, 본 발명의 폴리펩티드를 발현하는 세포에 대해 스크리닝하기 위한, 본 발명의 방법을 실시함에 있어서, 본 발명의 폴리펩티드 및 핵산과 함께 다양한 장치와 방법들이 사용될 수 있다. 이하에 상세히 기술되어 있는 시료 스크리닝용 어레이 방법에 더하여, 본 발명의 방법을 실시하는데에 대안적인 방법도 사용될 수 있다. 이러한 방법으로서의 예를 들어, 질량 분광 분석법, 크로마토그래피 예를 들어, 고 처리량 HPLC 및 기타 액체 크로마토그래피, 그리고 더욱 작은 규모의 방법 예를 들어, 1536-웰 평판법, 384-웰 평판법을 포함한다. 고 처리량 스크리닝 장치는 본 발명의 방법을 실시할 수 있도록 맞추어 사용할 수 있다[예를 들어, 미국 특허 출원 제 20020001809호 및 동 제20050272044호 참조].
- [0521] 모세관 어레이
- [0522] 본 발명의 핵산 또는 폴리펩티드는 어레이에 고정화되거나 도포될 수 있다. 어레이는 조성물(예를 들어, 소형 분자, 항체 및 핵산 등)의 라이브러리가 본 발명의 핵산 또는 폴리펩티드에 결합하거나 또는 이의 활성을 조정하는 능력을 가지는지 여부를 스크리닝하거나 모니터링하는데에 사용될 수 있다. 모세관 어레이 예를 들어, 기가 매트릭스(GIGAMATRIX)TM(Diversa Corporation, San Diego, CA); 그리고 예를 들어, 미국 특허 출원 제 20020080350호 A1; WO 0231203 A; WO 0244336 A에 개시된 어레이는 시료를 수용하고 스크리닝하는 대안적인 장치를 제공한다. 하나의 측면에서, 모세관 어레이는 인접 모세관 어레이와 연결된 다수의 모세관을 포함하는데, 여기서, 각각의 모세관은 시료를 수용하기 위한 루멘(lumen)을 한정하는 하나 이상의 벽을 포함한다. 상기 루멘은 그 벽이 액체 또는 시료를 수용하기 위한 루멘을 형성하도록 원통형, 사각형, 육각형 또는 임의의 기타 기하학적 형태를 가질 수 있다. 모세관 어레이를 이루는 모세관들은 서로 인접하여 존재함으로써, 평면 구조를 이룬다. 모세관은 융합되어 나란히 결합될 수 있거나(예를 들어, 모세관이 유리로 만들어진 경우), 접착제, 본드 또는 집게로 나란히 결합될 수 있다. 뿐만 아니라, 모세관 어레이는 어레이 내 인접하는 모세관 사이에 배치되어 있는 틈새 간 물질을 포함하여, 다수의 관통형 구멍을 포함하는 고체 평면 장치를 이룰 수 있다.
- [0523] 모세관 어레이는 임의의 수 예를 들어, 100~4,000,000개의 모세관들로 각각 이루어져 있다. 뿐만 아니라, 약 100,000개 이상의 개별 모세관을 가지는 모세관 어레이는 표준 크기로 성형되어, 표준 실험실용 장치에 딱 들어맞는 마이크로타이어(Microtiter)[®] 평판의 형태를 가질 수 있다. 루멘은 모세관 작용 또는 가느다란 바늘을 사용하는 미세 주입법을 이용하여 수동 또는 자동으로 충전된다. 이후, 목적 시료는 각각의 모세관으로부터 분리되어 추가의 분석 또는 특성 규명에 사용된다. 예를 들어, 가느다란 침상의 프로브는 루멘에 물질을 가하거나 또는 빼내기 위해 선택된 모세관과 유체 소통이 가능하도록 배치된다.
- [0524] 단일-팟 스크리닝 검정법(single-pot screening assay)에서, 검정 성분은 모세관 어레이에 삽입되기 이전에 혼합되어 목적 용액으로 만들어진다. 어레이의 적어도 일부가 목적 용액에 침지될 때, 이 루멘은 모세관 작용에 의해 충전된다. 각 모세관에서의 화학 또는 생물 반응 및/또는 활성은 검출 가능한 현상에 대해 모니터링된다. 검출 가능한 현상을 종종 "히트(hit)"라 칭하며, 이는 일반적으로 광학적 검출법에 의해 "비-히트(non-hit)" 생성 모세관과 구별될 수 있다. 그러므로, 모세관 어레이를 통하여 "히트"를 다량으로 동시에 검출할 수 있는 것이다.
- [0525] 다중-팟 스크리닝 검정법(multi-pot screening assay)"에 있어서, 폴리펩티드 또는 핵산 예를 들어, 리간드는 제1 성분 즉, 모세관 어레이를 이루는 모세관의 적어도 부분에 도입되는 성분 첨가될 수 있다. 제1 성분 첨가 후, 기포가 모세관에 도입될 수 있다. 이후, 제2 성분이 모세관에 도입될 수 있는데, 이 경우 상기 제2 성분은 기포에 의해 제1 성분과 분리된다. 이후, 상기 제1 성분 및 제2 성분은 정수압을 모세관 어레이의 양쪽 면에 가하여 기포를 파괴시킴으로써 혼합될 수 있다. 이후, 모세관 어레이는 2개의 성분이 반응하였는지 또는 반응하지

않았는지 여부를 나타내는 현상에 대해 모니터링된다.

[0526] 결합 스크리닝 검정법에 있어서, 목적 시료는 검출 가능한 입자로 표지화된 제1 액체로서 모세관 어레이의 모세관에 도입될 수 있는데, 이때, 모세관의 루멘은 이 루멘에 검출 가능한 입자를 결합시키기 위한 결합 물질로 코팅되어 있다. 그 다음, 제1 액체는 모세관으로부터 제거될 수 있는데, 이 경우, 결합되어 있던 검출 가능 입자는 모세관 내에 남아있게 되고, 제2 액체는 모세관에 도입될 수 있다. 이후, 상기 모세관은 입자가 제2 액체와 반응하였는지 또는 반응하지 않았는지 여부를 나타내는 현상에 대해 모니터링된다.

[0527] 어레이 또는 "바이오칩"

[0528] 본 발명의 핵산 또는 폴리펩티드는 어레이에 고정화되거나 또는 도포될 수 있다. 어레이는 조성물(예를 들어, 소형 분자, 항체 및 핵산 등)이 본 발명의 핵산 또는 폴리펩티드와 결합하여 이의 활성을 조정할 수 있는 능력을 갖는지 여부에 대해 스크리닝하거나 또는 이 조성물 라이브러리를 모니터링하는데에 사용될 수 있다. 예를 들어, 본 발명의 하나의 측면에서, 모니터링될 매개 변수는 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 유전자의 전사체 발현 여부이다. 하나 이상의, 또는 모든 세포의 전사체는 이 세포의 전사체를 포함하는 시료, 또는 세포의 전사체를 대표하는 핵산 또는 이에 상보성인 핵산을 혼성화하거나, 어레이 또는 "바이오칩"상에 고정화된 핵산에 혼성화함으로써 측정될 수 있다. 마이크로칩상에 존재하는 핵산의 "어레이"를 이용함으로써, 세포의 전사체 중 일부 또는 전부는 동시에 정량될 수 있다. 대안적으로, 계통형 핵산을 포함하는 어레이는 또한 본 발명의 방법에 의해 만들어진, 새로이 조작된 균주의 유전자형을 측정하는데에 사용될 수도 있다. "폴리펩티드 어레이"는 또한 다수의 단백질을 동시에 정량하는데에도 사용될 수 있다. 본 발명은 임의의 공지 "어레이"("마이크로어레이" 또는 "핵산 어레이" 또는 "폴리펩티드 어레이" 또는 "항체 어레이" 또는 "바이오칩") 또는 이의 변이체를 사용하여 실시될 수 있다. 어레이는 일반적으로 다수의 "스팟(spot)" 또는 "표적 요소(target element)"인데, 각각의 표적 요소는 하나 이상의 생물 분자 예를 들어, 올리고뉴클레오타이드를 일정량만큼 포함하고, 시료 분자 예를 들어, mRNA 전사체와 특이적으로 결합하는 기질 표면의 한정된 구역 상에 고정화되어 있다.

[0529] 본원에 사용된 "어레이" 또는 "마이크로어레이" 또는 "바이오칩" 또는 "칩"이란 용어는, 다수의 표적 요소를 의미하는데, 각 표적 요소는 하나 이상의 폴리펩티드(항체 포함) 또는 핵산을 일정량만큼 포함하고, 기질 표면의 특정 구역 상에 고정화되어 있다(이하, 보다 상세히 설명함).

[0530] 본 발명의 방법을 실시함에 있어서, 임의의 공지 어레이 및/또는 아레이를 제조 및 사용하는 방법의 전부 또는 일부, 또는 이의 변형은 예를 들어, 문헌[미국 특허 제6,277,628호; 동 제6,277,489호; 동 제6,261,776호; 동 제6,258,606호; 동 제6,054,270호; 동 제6,048,695호; 동 제6,045,996호; 동 제6,022,963호; 동 제6,013,440호; 동 제5,965,452호; 동 제5,959,098호; 동 제5,856,174호; 동 제5,830,645호; 동 제5,770,456호; 동 제5,632,957호; 동 제5,556,752호; 동 제5,143,854호; 동 제5,807,522호; 동 제5,800,992호; 동 제5,744,305호; 동 제5,700,637호; 동 제5,556,752호; 동 제5,434,049; 예를 들어, WO 99/51773; WO 99/09217; WO 97/46313; WO 96/17958; 예를 들어, Johnston (1998) Curr. Biol. 8:R171-R174; Schummer (1997) Biotechniques 23:1087-1092; Kern (1997) Biotechniques 23:120-124; Solinas-Toldo (1997) Genes, Chromosomes & Cancer 20:399-407; Bowtell (1999) Nature Genetics Supp. 21:25-32]에 기술되어 있다. 또한, 미국 특허 출원 제20010018642호; 동 제20010019827호; 동 제20010016322호; 동 제20010014449호; 동 제20010014448호; 동 제20010012537호; 동 제20010008765호를 참조하시오.

[0531] 항체 및 항체계 스크리닝 방법

[0532] 본 발명은 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소에 특이적으로 결합하는 분리된 또는 재조합 항체를 제공한다. 이러한 항체는 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 또는 관련 폴리펩티드를 분리, 동정 또는 정량하는데에 사용될 수 있다. 이러한 항체는 본 발명의 범위 내에 포함되는 기타 폴리펩티드 또는 기타 관련 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 분리하는데에 사용될 수 있다. 항체는 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 활성 부위에 결합하도록 디자인될 수 있다. 그러므로, 본 발명은 본 발명의 항체를 이용하여 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 억제하는 방법을 제공한다[본 발명의 항-셀룰라제 예를 들어, 항-엔도글루카나제, 항-셀로바이오하이드롤라제 및/또는 항-베타-글루코시다제 효소 조성물을 사용하는 것에 관하여 전술한 사항 참조].

- [0533] "항체"란 용어는 면역 글로불린 유전자 또는 면역 글로불린 유전자들 또는 이의 단편으로부터 유래하거나, 이를 모델링하거나 또는 실질적으로 이것에 의해 암호화되는 펩티드 또는 폴리펩티드로서, 항원이나 에피토프에 특이적으로 결합할 수 있는 펩티드 또는 폴리펩티드를 포함한다[예를 들어, Fundamental Immunology, Third Edition, W.E. Paul, ed., Raven Press, N. Y. (1993); Wilson (1994) J. Immunol. Methods 175:267-273; Yarmush (1992) J. Biochem. Biophys. Methods 25:85-97]. 항체라는 용어는 항원에 결합하는 능력을 가지는 항원 결합부 즉, "항원 결합 위치"(예를 들어, 단편, 부분 서열, 상보성 결정 부위(CDR))를 포함하며, 그 예로서는 (i) Fab 단편, 즉, VL, VH, CL 및 CH1 도메인으로 이루어진 1가 단편; (ii) F(ab')₂ 단편, 즉, 2개의 Fab 단편이 힌지부에서 이황화 결합에 의해 결합되어 있는 2가 단편; (iii) VH 및 CH1 도메인으로 이루어진 Fd 단편; (iv) 항체의 단일 팔을 이루는 VL 및 VH 도메인으로 이루어진 Fv 단편; (v) VH 도메인으로 이루어진 dAb 단편(Ward와 다수, (1989) Nature 341 :544-546); 그리고 (vi) 분리된 상보성 결정 부위(CDR)를 포함한다. 단일 사슬 항체도 "항체"라는 용어에 포함된다.
- [0534] 본 발명은 본 발명의 폴리펩티드의 면역원성 단편(예를 들어, 부분 서열)을 포함하는 본 발명의 효소(예를 들어, 펩티드)의 단편을 제공한다. 본 발명은 본 발명의 폴리펩티드 또는 펩티드와 애주반트 또는 담체를 포함하는 조성물을 제공한다.
- [0535] 항체는 면역 침전법, 염색, 면역 친화성 컬럼 등에서 사용될 수 있다. 원한다면, 특정 항원을 암호화하는 핵산 서열은 면역화 이후, 폴리펩티드 또는 핵산을 분리하고, 폴리펩티드를 증폭 또는 클로닝하여 본 발명의 어레이 상에 고정화시킴으로써 생산될 수 있다. 대안적으로, 본 발명의 방법은 변형된 세포에 의해 생산된 항체의 구조를 변형시키는데에 사용될 수 있는데, 이 경우 예를 들어, 항체의 친화성은 증가하거나 또는 감소할 수 있다. 뿐만 아니라, 항체를 생산 또는 변형시키는 능력은 본 발명의 방법에 의해 세포 내에 조작되어 도입된 표현형일 수 있다.
- [0536] 면역화, 항체(폴리클로날 항체 및 모노클로날 항체)의 생산 및 분리 방법에 관하여는 당 업자에게 공지되어 있으며, 과학 및 특허 문헌에 개시되어 있다[예를 들어, Coligan, CURRENT PROTOCOLS IN IMMUNOLOGY, Wiley/Greene, NY (1991); Stites (eds.) BASIC AND CLINICAL IMMUNOLOGY (7th ed.) Lange Medical Publications, Los Altos, CA ("Stites"); Goding, MONOCLONAL ANTIBODIES: PRINCIPLES AND PRACTICE (2d ed.) Academic Press, New York, NY (1986); Kohler (1975) Nature 256:495; Harlow (1988) ANTIBODIES, A LABORATORY MANUAL, Cold Spring Harbor Publications, New York 참조]. 항체는 또한 동물을 이용하는 통상의 생체내 방법 이외에도 예를 들어, 파지 디스플레이 라이브러리를 발현하는 재조합 항체 결합 위치를 이용하여 시험관 내에서 생산될 수도 있다. 예를 들어, 문헌[Hoogenboom (1997) Trends Biotechnol. 15:62-70; Katz (1997) Annu. Rev. Biophys. Biomol. Struct. 26:27-45]을 참조하십시오.
- [0537] 본 발명의 폴리펩티드 또는 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100 또는 150개 이상의 연속적 아미노산을 포함하는 이의 단편은 또한 폴리펩티드 또는 단편에 특이적으로 결합하는 항체를 생산하는데에도 사용될 수 있다. 결과로 생성된 항체는 폴리펩티드를 분리 또는 정제하거나, 또는 이 폴리펩티드가 생물 시료 중에 존재하는지 여부를 결정하기 위한 면역 친화성 크로마토그래피 방법에서 사용될 수 있다. 이러한 방법에 있어서, 단백질 제제 예를 들어, 추출물, 또는 생물 시료는 본 발명의 폴리펩티드 중 하나, 또는 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100 또는 150개 이상의 연속적 아미노산을 포함하는 이의 단편과 특이적으로 결합할 수 있는 항체와 접촉된다.
- [0538] 면역 친화성 방법에 있어서, 항체는 고정 지지체 예를 들어, 비드 또는 기타 컬럼 매트릭스 상에 부착된다. 단백질 제제는 항체가 본 발명의 폴리펩티드 중 하나 또는 이의 단편과 특이적으로 결합하는 조건 하에서 항체와 접촉하게 된다. 세척하여 비-특이적으로 결합된 단백질을 제거한 후, 특이적으로 결합한 폴리펩티드를 용리시킨다.
- [0539] 생물 시료 중 단백질이 항체와 결합하는 능력은 당 업자에게 익숙한 다양한 방법들 중 임의의 방법을 사용하여 측정될 수 있다. 예를 들어, 결합은 검출 가능한 표지 예를 들어, 형광성 제제, 효소 표지 또는 방사성 동위 원소로 항체를 표지화함으로써 측정될 수 있다. 대안적으로, 항체가 시료와 결합하였는지 여부는 검출 가능한 표지가 결합되어 있는 2차 항체를 사용하여 확인될 수 있다. 구체적인 검정법으로서는 ELISA 검정법, 샌드위치 검정법, 방사성 면역 검정법 및 웨스턴 블롯을 포함한다.
- [0540] 본 발명의 폴리펩티드 또는 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100 또는 150개 이상의 연속적 아미노산을 포함하는 이의 단편에 대하여 생산되는 폴리클로날 항체는 폴리펩티드를 동물에 직접 주사하거나 또는 폴리펩티드를 동물 예를 들어, 인간 이외의 동물에 투여함으로써 얻어진다. 이처럼 얻어진 항체는 폴리펩티드 자체

와 결합할 수 있다. 이러한 방식으로, 본 발명의 폴리펩티드의 단편만을 암호화하는 서열조차도 원래 폴리펩티드 전체에 결합할 수 있는 항체를 생산하는데에 사용될 수 있다. 이후, 이러한 항체는 이 폴리펩티드를 발현하는 세포로부터 폴리펩티드를 분리하는데에 사용될 수 있다.

[0541] 모노클로날 항체를 제조함에 있어서, 연속적 세포주 배양액에 의해 생산되는 항체를 제공하는 임의의 기술이 사용될 수 있다. 이러한 기술의 예로서는 하이브리도마 기술(Kohler and Milstein, Nature, 256:495-497, 1975), 트리오마(trioma) 기술, 인간 B-세포 하이브리도마 기술(Kozbor 외 다수, Immunology Today 4:72, 1983) 및 EBV-하이브리도마 기술(Cole 외 다수, 1985, Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96)을 포함한다.

[0542] 단일 사슬 항체를 생산하는 기술(미국 특허 제4,946,778호)을 적용하여 본 발명의 폴리펩티드, 또는 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100 또는 150개 이상의 연속적 아미노산을 포함하는 이의 단편에 대한 단일 사슬 항체를 생산할 수 있다. 대안적으로, 트랜스제닉 마우스를 사용하여 상기 폴리펩티드 또는 이의 단편에 대하여 인간화된 항체를 발현시킬 수 있다.

[0543] 본 발명의 폴리펩티드, 또는 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100 또는 150개 이상의 연속적 아미노산을 포함하는 이의 단편에 대하여 생성된 항체는 기타 유기체 및 시료로부터 유래하는 유사한 폴리펩티드들에 대해 스크리닝하는데에 사용될 수 있다. 이러한 기술에 있어서, 유기체로부터 유래하는 폴리펩티드는 항체와 접촉하게 되고, 이 항체에 특이적으로 결합하는 폴리펩티드가 검출된다. 전술한 방법 중 임의의 방법을 사용하여 항체 결합 여부를 확인할 수 있다. 이러한 스크리닝 방법 중 하나에 관하여는 문헌["Methods for Measuring Cellulase Activities", Methods in Enzymology, Vol. 160, pp. 87-116]에 기술되어 있다.

[0544] 키트

[0545] 본 발명은 본 발명의 조성물 예를 들어, 핵산, 발현 카세트, 벡터, 세포, 트랜스제닉 종자 또는 식물 또는 식물의 일부, 폴리펩티드(예를 들어, 셀룰라제 효소) 및/또는 항체를 포함하는 키트를 제공한다. 이 키트는 또한 본원에 기술된 바와 같이, 본 발명의 조성물을 사용하는 방법과 산업적 용도, 의학적 용도 그리고 섭생시 사용 방법에 관하여 교시하고 있는 지시 사항을 포함할 수 있다.

[0546] 전 세포 조작 및 대사 매개 변수의 측정

[0547] 본 발명의 방법은 신규의 표현형 예를 들어, 신규하거나 또는 변형된 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성을 보유하는 신규 세포 변종을 개발하기 위하여, 세포의 유전자 조성을 변형시킴으로써 세포 즉, 전 세포를 진화시키거나 또는 조작하는 방법을 제공한다. 미국 특허 출원 제20040033975호를 참조하시오.

[0548] 유전자 조성은 본 발명의 핵산 예를 들어, 본 발명의 효소에 대한 암호화 서열이 존재하는 세포를 첨가함으로써 변형될 수 있다. 예를 들어, WO 0229032 및 WO 0196551을 참조하시오.

[0549] 신규 표현형을 검출하기 위해서는, 변형된 세포의 하나 이상의 대사 매개 변수를 세포 내에서 "실시간" 또는 "온-라인" 시간 프레임으로 모니터링한다. 하나의 측면에서, 다수의 세포 예를 들어, 세포 배양액은 "실시간" 또는 "온-라인"으로 모니터링된다. 하나의 측면에서, 다수의 대사 매개 변수는 "실시간" 또는 "온-라인"으로 모니터링된다. 대사 매개 변수는 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 사용하여 모니터링될 수 있다.

[0550] 대사 흐름 분석법(Metabolic flux analysis; MFA)은 공지의 생화학적 원리를 바탕으로 하는 방법이다. 선형 독립 대사 매트릭스는 세포 내 대사 물질에 대한 질량 보존의 법칙과 의사-정상 상태 가설(PSSH)을 바탕으로 하여 구성된다. 본 발명의 방법을 실시함에 있어서, 대사 네트워크는 다음의 사항들을 고려하면서 구축된다:

[0551] * 모든 경로의 기질, 생산물 및 중간 대사 물질의 동일성

[0552] * 경로의 대사 물질을 상호 전환시키는 모든 화학 반응의 동일성, 경로 반응의 화학 양론

[0553] * 반응을 촉진하는 모든 효소의 동일성, 효소 반응 역학

[0554] * 경로를 구성하는 성분들 사이의 조절 상호 작용 예를 들어, 알로스테릭 상호 작용, 효소-효소 상호 작용 등

[0555] * 효소의 세포 내 구획화 또는 효소의 기타 초 분자학적 조직화, 및

[0556] * 대사 물질, 효소 또는 효과기 분자의 임의의 농도 구배 또는 이들의 이동에 대한 확산 장벽의 존재.

- [0557] 일단 소정의 변종에 대해 대사 네트워크가 구축되면, 매트릭스 개념에 의한 수학적 제시가 도입되어, 온-라인 대사 데이터를 얻을 수 있을 경우 세포 내 대사 흐름을 측정할 수 있게 된다. 대사 표현형은 세포 내 전체 대사 네트워크의 변화에 달려있다. 대사 표현형은 환경 조건, 유전자 조절, 발달 단계 및 유전자형 등에 관한 경로 활용 변화에 달려있다. 본 발명의 방법에 관한 하나의 측면에서, 온-라인 MFA 계산 이후, 세포, 이 세포의 표현형 그리고 기타 특성의 동적인 거동은 경로 이용률을 조사함으로써 분석된다. 예를 들어, 만일 효모 발효시 글루코스 공급량이 증가하고 산소 공급량이 감소하면, 호흡 경로 활용도는 감소하고/감소하거나 정지할 것이고, 발효 경로의 활용도는 낮아질 것이다. 세포 배양액의 생리적 상태는 경로 분석 이후에 제어 가능하게 될 것이다. 본 발명의 방법은 바람직한 방향으로 세포의 생리적 단계가 진행되도록 제어하는 유도 물질의 활용 방식과, 기질 공급량 및 온도가 어떻게 변화하는지를 확인하는데 도움이 될 수 있다. 발효 방식이 어떻게 조작되는지를 측정하는 것을 도울 수 있다. 본 발명의 방법을 실시함에 있어서, MFA 결과는 또한 전사체 및 프로테옴(proteome) 데이터와 비교되어, 대사 조작 또는 유전자 서플링 등에 대한 실험 및 프로토콜을 디자인할 수 있다.
- [0558] 본 발명의 방법을 실시함에 있어서, 세포 내에서의 신규하거나 개선된 특성들을 포함하는 임의의 변형된 표현형 또는 신규 표현형이 부여되어 검출될 수 있다. 대사 또는 성장에 관한 임의의 측면이 모니터링될 수 있다.
- [0559] *mRNA 전사체 발현의 모니터링*
- [0560] 본 발명의 하나의 측면에서, 조작된 표현형으로서는 mRNA 전사체(예를 들어, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 메세지)의 발현이 증감되는 경우, 또는 세포 내에서 신규한 전사체(예를 들어, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소)를 생산하는 경우를 포함한다. 이와 같이 발현량의 증감은 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소가 존재하는지 여부에 대해 시험하거나, 또는 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성 검정법에 의해 추적될 수 있다. mRNA 전사체 또는 메세지는 또한 당 업계에 공지된 임의의 방법 예를 들어, 노던 블롯, 정량적 증폭 반응, 어레이와의 혼성화 등에 의해 검출 및 정량될 수 있다. 정량적 증폭 반응으로서는 예를 들어, 정량적 PCR 예를 들어, 정량적 역 전사 중합 효소 연쇄 반응, 또는 RT-PCR; 정량적 실시간 RT-PCR, 또는 "실시간 역학적 RT-PCR"을 포함한다[예를 들어, Kreuzer (2001) Br. J. Haematol. 114:313-318; Xia (2001) Transplantation 72:907-914 참조].
- [0561] 본 발명의 하나의 측면에서, 조작된 표현형은 상동성 유전자의 발현을 억제(knock out)시킴으로써 나타나게 된다. 유전자의 암호화 서열 또는 하나 이상의 전사 조절 요소 예를 들어, 프로모터나 인핸서는 녹 아웃될 수 있다. 그러므로, 전사체의 발현은 완전히 억제되거나 또는 단지 발현량이 감소하기만 할 수 있다.
- [0562] 본 발명의 하나의 측면에서, 조작된 표현형은 상동성 유전자의 발현량이 증가된 경우를 포함한다. 이는 시스-작용 또는 트랜스-작용 네가티브 제어 요소 예를 들어, 전사 조절 요소를 녹 아웃시키거나, 또는 포지티브 제어 요소를 돌연변이시킴으로써 이루어질 수 있다. 세포의 하나 이상의 전사체 또는 모든 전사체는 세포의 전사체, 또는 세포의 전사체를 대표하거나 이에 상응하는 핵산을 포함하는 시료를 혼성화함으로써, 또는 어레이 상에 고정화된 핵산에 혼성화함으로써 확인될 수 있다.
- [0563] *폴리펩티드, 펩티드 및 아미노산 발현의 모니터링*
- [0564] 본 발명의 하나의 측면에서, 조작된 표현형으로서는 폴리펩티드(예를 들어, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소)의 발현이 증감되는 경우, 또는 세포 내에서 신규 폴리펩티드가 생산되는 경우를 포함한다. 이와 같은 발현량의 증감은 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소의 존재량을 측정하거나, 또는 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성 검정법에 의해 추적될 수 있다. 폴리펩티드, 펩티드 및 아미노산은 또한 당 업계에 공지된 임의의 방법 예를 들어, 핵 자기 공명법(NMR), 분광법, 방사선 사진술(단백질 방사능 표지화), 전기 영동, 모세관 전기 영동, 고성능 액체 크로마토그래피(HPLC), 박층 크로마토그래피(TLC), 하이퍼디퓨전 크로마토그래피(hyperdiffusion chromatography), 다양한 면역학적 방법 예를 들어, 면역 침전법, 면역 확산법, 면역-전기 영동법, 방사성 면역 검정법(RIA), 효소-결합 면역 흡착 검정법(ELISA), 면역-형광 검정법, 겔 전기 영동법(예를 들어, SDS-PAGE), 항체를 이용한 염색법, 형광 활성화 세포 분류기(FACS), 열분해 질량 분광법, 푸리에-변형 적외선 분광법(Fourier-Transform Infrared Spectrometry), 라만 분광법, GC-MS 및 LC-전기 방사법 및 캡-LC-직렬-전기 방사 질량 분광법 등에 의해 검출 및 정량될 수 있다. 신규의 생물 활성은 또한 미국 특허 제6,057,103

호에 개시된 방법 또는 이의 변형을 사용하여 스크리닝될 수도 있다. 뿐만 아니라, 이하 더욱 상세히 논의되어 있는 바와 같이, 세포 폴리펩티드 중 하나 이상 또는 전부는 단백질 어레이를 사용하여 측정될 수 있다.

[0565] 산업용, 에너지 합성용, 제약학적 용도 및 기타 용도

[0566] 본 발명의 폴리펩티드(예를 들어, 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 활성을 보유하는 폴리펩티드)는 셀룰로스의 분해를 촉진할 수 있다. 본 발명의 효소는 매우 선택적인 촉매일 수 있다. 본 발명은 본 발명의 효소를 사용하는 산업 분야 예를 들어, 제약학 분야 또는 영양(다이어트) 보조제 산업 분야, 에너지 산업 분야(예를 들어, "청정" 바이오 연료 제조), 식품 및 사료 제조 분야 예를 들어, 식품 및 사료 제품의 제조 분야, 및 식품 및 사료 첨가물의 제조 분야에 사용될 수 있다. 하나의 측면에서, 본 발명은 의료 산업 분야(예를 들어, 약품이나 식품 보조제 또는 보충제를 제조하는 경우)에서 본 발명의 효소를 사용하는 방법을 제공한다. 뿐만 아니라, 본 발명은 바이오에탄올을 비롯한 "청정" 연료 생산에 있어서 본 발명의 효소를 사용하는 방법을 제공한다.

[0567] 본 발명의 효소는 입체 선택성, 위치 선택성 및 화학 선택성을 나타내는 반응을 촉진할 수 있다. 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소는 다양한 용매 중에서 작용하고, 극한의 pH(예를 들어, 고 pH 및 저 pH) 및 극한의 온도(예를 들어, 고온 및 저온), 극한의 염분 수준(예를 들어, 고 염도 및 저 염도)에서 작용하며, 화합물 본래의 생리적인 기질과 구조적으로 관련되어 있지 않은 화합물과의 반응을 촉매하도록 조작될 수 있다.

[0568] 바이오매스 전환 및 청정 바이오 연료의 생산

[0569] 본 발명은 바이오매스(예를 들어, 리그노셀룰로스 물질)를 연료(예를 들어, 바이오에탄올) 및 화학 물질로 전환시키는 효소 및 방법을 제공한다. 그러므로, 본 발명의 조성물 및 방법은 석유계 제품에 대한 효율적이고 지속적인 대체물을 제공한다. 본 발명은 천연 바이오매스 전환과 관련된 화학적 순환 과정에 참여하는 본 발명의 효소를 발현시키는 유기체를 제공한다. 하나의 측면에서, 전환용 효소 및 전환 방법은 셀룰로스계 및 헤미셀룰로스계 중합체를 대사 가능한 탄소 부분으로 효율적으로 분해시키기 위한 효소 집합물에서 사용된다. 전술한 바와 같이, 본 발명은 중요한 신규의 "바이오매스 전환" 및 대체 에너지 산업 공정을 가능하게 하는 가장 효율적인 효소를 발견하고 이를 활용하는 방법을 제공한다.

[0570] 하나의 측면에서, 셀룰라제 활성, 예를 들어 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 활성을 보유하는 본 발명의 폴리펩티드, 예를 들어 단백질은 리그노셀룰로스 바이오매스를 에탄올로 전환시키는 방법에 사용된다. 본 발명은 또한 리그노셀룰로스 바이오매스를 포함하는 조성물로부터 에탄올("바이오에탄올")을 생산하는 방법을 제공한다. 리그노셀룰로스 바이오매스 물질은 농업 작물로부터, 식품 또는 사료 생산시의 부산물로서, 또는 리그노셀룰로스계 폐기물, 예를 들어 식물 잔류물 및 폐지로서 얻을 수 있다. 본 발명의 폴리펩티드로 처리하는 데 적당한 식물 잔류물의 예로서는, 줄기, 잎, 외피, 껍질 및 속대(cob) 등과, 목재, 목재 칩, 목재 펄프 및 톱밥을 포함한다. 본 발명의 폴리펩티드로 처리하는 데 적당한 폐지로서는 버려지는 복사 용지, 컴퓨터 인쇄 용지, 공책 종이, 메모지 및 타자 용지 등과, 신문, 잡지, 판지 및 종이계 포장재를 포함한다.

[0571] 하나의 측면에서, 본 발명의 효소 및 방법은 바이오매스로부터 에탄올을 생산하는 보다 "통상적인" 방법, 예를 들어, 반응기 내에서 건조된 리그노셀룰로스 물질에 촉매(강산의 희석 용액과 금속 염으로 이루어짐)를 처리하여 리그노셀룰로스 물질을 가수분해하는 단계를 포함하는 방법과 함께 사용될 수 있으며; 이로써, 셀룰로스를 가수분해하는데 있어서 활성화 에너지 또는 반응 온도를 낮추어, 수율이 높은 당을 얻을 수 있다[예를 들어, 미국 특허 제6,660,506호 및 동 제6,423,145호 참조].

[0572] 본 발명의 효소를 사용하는 단계를 통합한 기타 대표적인 방법은, 셀룰로스를 글루코스로 거의 분해시키지 않고서, 헤미셀룰로스를 거의 분해하도록 선택된 온도 및 압력에서, 헤미셀룰로스, 셀룰로스 및 리그닌을 함유하는 리그노셀룰로스 물질을 사용하여 수성 매질 중에서 가수분해의 첫 번째 단계를 진행시킴으로써, 이 헤미셀룰로스, 셀룰로스 및 리그닌을 함유하는 리그노셀룰로스 물질을 가수분해시키는 단계를 포함한다. 이 단계를 통하여, 헤미셀룰로스와, 셀룰로스 및 리그닌을 함유하는 고상을 분해하여 얻어진 용해 단당류를 액상이 함유하고 있는 슬러리를 얻을 수 있다. 가수분해의 두 번째 단계는 대부분의 셀룰로스가 분해되는 조건을 포함할 수 있는데, 여기서, 상기 단계를 통하여 셀룰로스의 용해된 분해 산물/가용성 분해 산물을 함유하는 액체 수성 상을 얻을 수 있다. 예를 들어, 미국 특허 제5,536,325호를 참조하시오. 본 발명의 효소는 이러한 대표적인 공정 중의 임의의 단계에 첨가될 수 있다.

- [0573] 본 발명의 효소를 사용하는 단계를 통합한 기타 대표적인 방법은, 약 0.4~2%의 강산을 사용하는 묽은 산 가수분해 방법 중 하나 이상의 단계에 의해 리그노셀룰로스 함유 바이오매스 물질을 가공하는 단계; 및 산으로 가수분해된 바이오매스 물질의 미반응 고체 리그노셀룰로스 성분을 알칼리성 탈 리그닌 처리법으로 처리하여, 생분해 열 가소성 물질 및 이의 유도체에 대한 전구체를 얻는 단계를 포함한다. 예를 들어, 미국 특허 제6,409,841호를 참조하시오. 본 발명의 효소는 이러한 대표적인 방법 중 임의의 단계에 첨가될 수 있다.
- [0574] 본 발명의 효소를 사용하는 단계를 통합한 기타 대표적인 방법은, 예비 가수분해 반응기 내에서 리그노셀룰로스 물질을 예비 가수분해시키는 단계; 산성 액체를 고체 리그노셀룰로스 물질에 첨가하여 혼합물을 만드는 단계; 이 혼합물을 반응 온도로 가열하는 단계; 리그노셀룰로스 물질이 용해된 부분(리그노셀룰로스 물질로부터 유래한 리그닌을 약 20% 이상 함유함)과 셀룰로스를 함유하는 고체 분획으로 분류되기에 충분한 시간 동안 반응 온도를 유지시키는 단계; 반응 온도 또는 이에 가까운 온도에 있을 때 고체 분획으로부터 용해된 부분을 제거하는 단계[여기서, 고체 분획 중 셀룰로스는 효소 분해가 더욱 잘 이루어지게 만들어짐]; 그리고, 용해된 부분을 회수하는 단계를 포함한다. 예를 들어, 미국 특허 제5,705,369호를 참조하시오. 본 발명의 효소는 이러한 대표적인 방법의 임의의 단계에 첨가될 수 있다.
- [0575] 본 발명은 본 발명의 효소나 방법을 사용하여 제조된 연료 등급의 알콜과 배합된 액체 탄화수소를 주성분으로 하는 모터 연료 조성물(예를 들어, 스파크 점화 모터용 연료 조성물)을 제공한다. 하나의 측면에서, 본 발명의 효소를 사용하여 제조된 연료로서는 예를 들어, 석탄 가스 액체-에탄올 배합물 또는 천연 가스 액체-에탄올 배합물을 포함한다. 하나의 측면에서, 보조 용매는 바이오매스-유래 2-메틸테트라하이드로푸란(MTHF)이다. 예를 들어, 미국 특허 제6,712,866호를 참조하시오.
- [0576] 리그노셀룰로스를 효소에 의해 분해하는 본 발명의 방법 예를 들어, 리그노셀룰로스 물질로부터 에탄올을 생산하기 위한 본 발명의 방법은 또한 바이오매스 물질을 초음파 처리하는 단계를 포함할 수 있다. 예를 들어, 미국 특허 제6,333,181호를 참조하시오.
- [0577] 본 발명의 효소를 사용하여 에탄올을 포함하는 바이오 연료를 생산하는 기타 대표적인 방법은, 적어도 헤미셀룰로스 및 셀룰로스를 포함하는 리그노셀룰로스 공급 원료를 함유하는 출발 물질을 전 처리하는 단계를 포함한다. 하나의 측면에서, 상기 출발 물질로서는 감자, 대두(유채씨), 보리, 호밀, 옥수수, 귀리, 밀, 비트 또는 사탕수수 또는 기타 성분 또는 폐기물 또는 식품 또는 사료 생산에 따른 부산물을 포함한다. 출발 물질("공급 원료")은 식물의 섬유 구조를 파괴하여 헤미셀룰로스 및 셀룰로스를 최소한 부분적으로 가수분해하는 조건에서 반응한다. 파괴 조건으로서의 예를 들어, 출발 물질을 약 5초~60분 동안 평균 온도 180~270℃ 및 pH 0.5~2.5에 두거나; 또는 약 5초~120초 동안 온도 220~270℃ 및 pH 0.5~2.5에 두는 경우, 또는 이에 버금가는 경우를 포함할 수 있다. 이로써, 본 발명의 효소 예를 들어, 셀룰라제 효소에 의해 분해될 수 있는 순응성이 증가한 공급 원료를 얻게 된다. 미국 특허 제6,090,595호를 참조하시오.
- [0578] 리그노셀룰로스 물질의 셀룰라제에 의한 가수분해를 위한 조건으로서 대표적인 조건은 온도 약 30~48℃ 및/또는 pH 약 4.0~6.0에서의 반응을 포함한다. 기타 대표적인 조건으로서의 온도는 약 30~60℃ 및 pH 약 4.0~8.0인 경우를 포함한다.
- [0579] 동물 사료 및 식품 또는 사료 첨가물
- [0580] 인간용 식품 보조제 또는식이 보조제, 또는 식품 보충 성분 및 첨가물을 제공하는 이외에도, 본 발명은 예를 들어, 셀룰라제 활성 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 활성을 보유하는 본 발명의 폴리펩티드 예를 들어, 단백질, 및/또는 본 발명의 항체를 사용하여, 동물 사료 및 식품, 식품 첨가물 또는 사료 첨가물을 처리하는 조성물 및 방법도 제공한다. 본 발명은 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 및/또는 본 발명의 항체를 포함하는 동물 사료, 식품 및 첨가물을 제공한다. 이 동물은 임의의 가축이나 동물일 수 있다.
- [0581] 본 발명의 동물 사료 첨가물은 사료 성분들과 용이하게 혼합될 수 있는 과립화된 효소 제품일 수 있다. 대안적으로, 본 발명의 사료 첨가물은 예비-혼합 성분을 형성할 수 있다. 본 발명의 과립화된 효소 제품은 코팅되거나 또는 코팅되지 않을 수 있다. 효소 과립의 입도는 사료와 예비-혼합 성분의 입도에 따라서 조정될 수 있다. 이로써, 효소를 사료에 혼합하는 안전하고도 편리한 수단을 제공한다. 대안적으로, 본 발명의 동물 사료 첨가물은 안정화된 액체형 조성물일 수 있다. 이는 수성 또는 유성 슬러리일 수 있다. 예를 들어, 미국 특허 제6,245,546호를 참조하시오.

- [0582] 동물 사료 또는 식품의 개질에 있어서, 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소는 시험관 내에서(사료 또는 식품의 성분들을 개질시킴으로써) 또는 생체내에서 식품 또는 사료를 가공할 수 있다. 본 발명의 폴리펩티드는 동물 사료 또는 식품 조성물에 첨가될 수 있다.
- [0583] 하나의 측면에서, 본 발명의 효소는 다른 효소 예를 들어, 베타-갈락토시다제, 카탈라제, 락카제, 기타 셀룰라제, 엔도글리코시다제, 엔도-베타-1,4-락카제, 아밀로글루코시다제, 기타 글루코시다제, 글루코스 이소머라제, 글리코실트랜스퍼라제, 리파제, 포스포리파제, 리폭시게나제, 베타-락카제, 엔도-베타-1,3(4)-락카제, 큐티나제, 퍼옥시다제, 아밀라제, 글루코아밀라제, 펙티나제, 리덕타제, 옥시다제, 디카르복실라제, 페놀록시다제, 리그니나제, 폴루라나제, 아라비나나제, 헤미셀룰라제, 만난아제, 자일로락카제, 자일라나제, 펙틴 아세틸 에스테라제, 람노갈락투로난 아세틸 에스테라제, 프로테아제, 펩티다제, 프로티나제, 폴리갈락투로나제, 람노갈락투로나제, 갈락타나제, 펙틴 라이아제, 트랜스글루타미나제, 펙틴 메틸에스테라제, 기타 셀로바이오하이드롤라제 및/또는 트랜스글루타미나제와 함께 첨가된다. 이러한 효소의 분해 산물은 동물에 의해 더욱 분해되기 쉽다. 그러므로, 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소는 사료 또는 식품의 가용 에너지를 내거나, 또는 셀룰로스로 분해함으로써 식품 또는 사료를 분해할 수 있는 능력을 부여하는데에 기여한다.
- [0584] 다른 측면에서, 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소는 트랜스제닉 사료용 곡물(예를 들어, 트랜스제닉 식물 및 종자 등) 예를 들어, 알곡, 곡류, 옥수수, 대두, 유채씨 및 루핀 등에서 직접적으로 효소를 발현시켜 공급될 수 있다. 전술한 바와 같이, 본 발명은 본 발명의 폴리펩티드를 암호화하는 핵산 서열을 포함하는 트랜스제닉 식물, 식물 일부 및 식물 세포를 제공한다. 하나의 측면에서, 상기 핵산은 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소가 회수될 수 있는 양만큼 발현된다. 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소는 임의의 식물 또는 식물의 일부로부터 회수될 수 있다. 대안적으로, 재조합 폴리펩티드를 함유하는 식물 또는 식물의 일부는 식품 또는 사료의 품질을 개선할 목적으로 예를 들어, 영양적 가치 및 기호성(palatability) 등을 개선할 목적으로 사용될 수 있다.
- [0585] 하나의 측면에서, 본 발명의 효소 전달 매트릭스는 다수의 개별 입자, 펠릿 또는 과립의 형태를 가진다. "과립(granule)"이란, 예를 들어, 펠릿화, 압출 또는 유사한 조밀화 기술을 통하여 매트릭스로부터 수분을 제거한, 압착되었거나 또는 조밀한 입자를 의미한다. 이와 같이 입자의 압착 또는 조밀화를 통하여 입자들이 내부에서 서로 뭉치는 것을 촉진할 수도 있다. 예를 들어, 과립은 펠릿 밀(pellet mill)에서 주로 알갱이로 이루어진 기질을 펠릿화함으로써 제조될 수 있다. 이렇게 제조된 펠릿은 동물 사료에 있어서 애췌반트로서 사용하기에 적당한 과립 크기로 분쇄되거나 또는 붕괴된다. 매트릭스 자체가 동물 사료에 사용하도록 승인을 받았으므로, 이 매트릭스는 동물 사료 내에 효소를 전달하기 위한 회석제로서 사용될 수 있다.
- [0586] 하나의 측면에서, 본 발명의 효소 전달용 매트릭스 및 본 발명의 방법에 포함되는 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소는 본원에 기술된 바와 같이, 열 안정성 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소이므로, 고온 및/또는 증기가 펠릿화된 효소 전달 매트릭스를 제조하는데에 적용될 수 있는 제조 과정 중 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소가 불활성화되는 것을 막을 수 있다. 본 발명의 효소 전달 매트릭스를 함유하는 사료를 분해하는 동안, 수성의 침지액은 활성 효소를 방출시킬 것이다. 열 안정성 효소 및 열 안정성인 영양 보충제의 다른 종류도 또한 임의의 유형의 수성 조건 하에서의 방출을 위해 전달 매트릭스 내에 통합될 수 있다.
- [0587] 하나의 측면에서, 다수의 상이한 목적 예를 들어, 동물 사료에 풍미 또는 영양 보충제를 첨가하거나, 동물 사료 보충물과 효소가 위 내 환경에 방출되는 것을 늦추는 등의 목적으로 효소 매트릭스 입자에 코팅이 이루어진다. 하나의 측면에서, 코팅은 기능상의 목적을 위한 경우 예를 들어, 매트릭스 입자로부터 효소가 방출되는 것을 늦추거나, 또는 효소가 방출될 조건을 제어하는 것이 바람직할 경우에는 언제든지 이루어진다. 코팅 물질 조성물은 그것에 민감한 제제(예를 들어, 열, 산 또는 염기, 효소 또는 기타 화학 물질)에 의해 선택적으로 분해될 수 있다. 대안적으로, 상기와 같은 상이한 분해 제제들에 민감한 2개 이상의 코팅이 매트릭스 입자에 연속적으로 가하여질 수 있다.
- [0588] 본 발명은 또한 효소-방출 매트릭스를 제조하기 위한 방법에 관한 것이기도 하다. 본 발명에 따르면, 상기 방법

은 효소-방출 매트릭스로서 사용하기에 적당한 입도를 가지는, 주로 알갱이로 이루어진 기질의 별도 입자들을 다수 제공하는 단계를 포함하는데, 이때, 상기 입자는 본 발명의 아미노산 서열에 의해 암호화되는 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 포함한다. 하나의 측면에서, 상기 방법은 효소-방출 매트릭스 입자를 과립으로 조밀화 또는 압착시키는 단계를 포함하는데, 하나의 측면에 따르면, 이러한 과정은 펠릿화에 의해 이루어진다. 곰팡이 억제제 및 점착제가 사용될 경우, 이들은 임의의 적당한 시간에 첨가될 수 있으며, 하나의 측면에 의하면, 이들은 주로 알갱이로 이루어진 기질을 펠릿화하기 전에, 상기 알갱이로 이루어진 기질과 바람직한 비율로 혼합된다. 하나의 측면에서, 펠릿 체분 사료 내 수분 함량은 최종 생산물 내 수분 함량에 관하여 전술한 바와 같은 범위 내에 있으며, 하나의 측면에 따르면 약 14~15%이다. 하나의 측면에서, 수분은 공급 원료에 효소의 수성 제제의 형태로서 첨가되어, 이와 같은 수분 함량에 피드백(feedback)을 제공한다. 하나의 측면에서, 펠릿 체분시의 온도는 증기에 의해 약 82℃까지 올라간다. 상기 펠릿 체분은 공급 원료를 충분히 작업하여 펠릿으로 만드는 임의의 조건 하에서 실행될 수 있다. 펠릿 공정은 효소-함유 조성물로부터 수분을 제거하는데 있어서 비용면에서 효율적인 공정이다.

[0589] 본 발명의 조성물 및 방법은 프리바이오틱스(prebiotics) 즉, 고 분자량의 당 예를 들어, 프록토-올리고당(FOS); 갈락토-올리고당(GOS), GRAS(일반적으로 안전하다고 승인된; Generally Recognized As Safe) 물질을 투여하여 실시될 수 있다. 이러한 프리바이오틱스는 몇몇 프로바이오틱 락트산 박테리아(LAB)에 의해 대사화될 수 있다. 이 프리바이오틱스는 다수의 장내 미생물에 의해서는 분해되지 않는다.

[0590] *식품의 처리 및 식품 가공*

[0591] 본 발명은 본 발명의 효소들을 함유하는 식품 및 사료와, 이 식품과 사료를 가공할 때 본 발명의 효소를 사용하는 방법을 제공한다. 본 발명의 셀룰라제 효소 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소는 식품 가공 산업에 있어서 여러모로 사용된다. 본 발명은 셀룰로스-포함 조성물 예를 들어, 식물 세포, 박테리아 세포, 효모 세포, 곤충 세포 또는 동물 세포, 또는 임의의 식물 또는 식물의 일부, 또는 임의의 식품 또는 사료, 폐기물 등을 가수분해하는 방법을 제공한다.

[0592] 예를 들어, 본 발명은 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 포함하는 사료 또는 식품 예를 들어, 사료, 액체 예를 들어, 음료(예를 들어, 과일 주스 또는 맥주), 빵 또는 가루 반죽이나 빵 제품, 또는 마실 것(예를 들어, 맥주) 또는 음료의 전구 물질(예를 들어, 맥아즙)을 제공한다.

[0593] 본 발명의 식품 처리 방법은 또한 기타 효소 예를 들어, 트립토판나제 또는 티로신 디카르복실라제, 락카제, 카탈라제, 락카제, 기타 셀룰라제, 엔도글리코시다제, 엔도-베타-1,4-락카제, 아밀로글루코시다제, 기타 글루코시다제, 글루코스 이소머라제, 글리코실트랜스퍼라제, 리파제, 포스포리파제, 리폭시게나제, 베타-락카제, 엔도-베타-1,3(4)-락카제, 큐티나제, 퍼옥시다제, 아밀라제, 글루코아밀라제, 펙티나제, 리덕타제, 옥시다제, 디카르복실라제, 페놀록시다제, 리그니나제, 폴루라나제, 아라비나나제, 헤미셀룰라제, 만난아제, 자일로락카제, 자일라나제, 펙틴 아세틸 에스터라제, 람노갈락투로난 아세틸 에스터라제, 프로테아제, 펩티다제, 프로티나제, 폴리갈락투로나제, 람노갈락투로나제, 갈락타나제, 펙틴 라이아제, 트랜스글루타미나제, 펙틴 메틸에스터라제, 기타 셀로바이오하이드롤라제 및/또는 트랜스글루타미나제를 경우에 따라 조합하여 사용하는 것도 포함할 수 있다.

[0594] 하나의 측면에서, 본 발명은 액체(액화) 및 과립형 전분을 가수분해하는 효소 및 방법을 제공한다. 이러한 전분은 임의의 공급원 예를 들어, 비트, 사탕수수, 감자, 옥수수, 밀, 마일로, 수수, 호밀 또는 벌거(bulgher)로부터 유래될 수 있다. 본 발명은 임의의 식물성 전분 공급원 예를 들어, 액화에 유용한 알곡 전분 공급원(예를 들어, 바이오에탄올 제조용) 예를 들어, 액화에 적당한 전분을 생산하는 것으로 알려져 있는 기타 알곡 또는 식물 공급원에 적용될 수 있다. 본 발명의 방법은 임의의 천연 물질 예를 들어, 쌀, 쌀아 쌀, 옥수수, 보리, 마일로, 밀, 콩과 식물, 감자, 비트, 사탕수수 및 고구마로부터 전분을 액화시키는 과정(예를 들어, 바이오에탄올 제조시)을 포함한다. 이러한 액화 과정은 실질적으로 전분을 가수분해하여 시럽으로 만든다. 액화시 온도 범위는 전분을 액화하는데에 효과적인 것으로 알려진 임의의 액화 온도일 수 있다. 예를 들어, 전분의 온도는 약 80~약 115℃, 약 100~약 110℃ 및 약 105~약 108℃일 수 있다. 본 발명의 효소 및 방법을 사용하여 제조된 바이오에탄올은 알콜 음료를 비롯한 식품 및 사료를 제조하는데에 사용되는 이외에도, 연료로서 사용될 수 있거나, 또는 전술한 바와 같이 연료 중(예를 들어, 자동차 연료)에 사용될 수 있다.

[0595] *폐기물 처리*

[0596] 본 발명은 폐기물 처리에 사용되는 효소를 제공한다. 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이

오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소는 광범위한 폐기물 처리 또는 관련 산업 분야 예를 들어, 바이오매스를 전환시켜 연료를 생산하는 폐기물 처리 분야에 사용될 수 있다. 예를 들어, 하나의 측면에서, 본 발명은 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소를 사용하는 고체 및/또는 액체 폐기물 분해 방법을 제공한다. 이 방법은 실질적으로 미처리된 고체 폐기물의 질량과 부피를 줄이는 과정을 포함할 수 있다. 고체 폐기물은 온도를 조절하면서 효소 용액(예를 들어, 본 발명의 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소 포함) 존재시 효소에 의한 분해 방법으로 처리될 수 있다. 그 결과, 첨가된 미생물로부터 감지할 수 있을 정도의 박테리아 발효가 일어나지 않는 반응이 진행된다. 고체 폐기물은 액화형 폐기물 및 임의의 잔류성 고체 폐기물로 전환된다. 결과로 생성된 액화 폐기물은 상기 임의의 잔류성 고체 폐기물로부터 분리될 수 있다. 예를 들어, 미국 특허 제5,709,796호를 참조하시오.

[0597] 하나의 측면에서, 본 발명의 조성물 및 방법은 예를 들어, 동물로부터 나오는 폐기물 용탕이 예를 들어, 돼지 농장, 기타 동물로부터 나오는 폐기물 처리 시스템, 또는 임의의 산업 분야 또는 식품 가공 분야에서 악취 제거, 악취 예방 또는 악취 감소에 사용된다,

[0598] 바이오매스(예를 들어, 리그노셀룰로스 물질)를 연료(예를 들어, 바이오에탄올)로 전환하는 방법 및 이에 사용되는 효소는, 시에서 나오는 고체 폐기물 예를 들어, 시 자체에서 직접 회수해 온 폐기물이나, 미리 땅에 매립해 두었다가 회수해 온 시의 고체 폐기물, 또는 예를 들어, 상당량의 셀룰로스 물질을 함유하는 하수 슬러지 케이크 형태의 하수 슬러지의 처리/재생 공정에 통합되어 사용될 수 있다. 하수 슬러지 케이크는 일반적으로 재생 가능한 물질(알루미늄, 유리 및 플라스틱 등)을 거의 함유하지 않을 것이므로, 이 하수 슬러지 케이크는 (폐기물의 셀룰로스 성분 중 중금속 함량을 낮추기 위해) 진한 황산으로 직접 처리되어, 에탄올 생산 시스템 내에서 처리될 수 있다. 예를 들어, 미국 특허 제6,267,309호 및 동 제5,975,439호를 참조하시오.

[0599] 본 발명의 효소를 사용하여 폐기물로부터 유기 물질 및 무기 물질을 회수하기 위한 대표적인 방법은 이 고체 유기 물질을 살균한 후, 이것에 열 및 압력을 가하여 연화시키는 과정을 포함한다. 상기와 같은 대표적인 방법은 우선, 폐기물을 교반한 다음 이를 열 및 압력에 노출시켜 이 폐기물을 살균하고 또한 여기에 함유되어 있던 유기 물질을 연화함으로써 수행될 수 있다. 하나의 측면에서, 일정 압력 하에서 가열한 후, 구멍난 챔버로부터 상기 압력을 갑자기 빼주어 이 용기의 구멍을 통해 연화된 유기 물질을 밖으로 배출시켜, 고체 무기 물질과 이 유기 물질을 분리시킬 수 있다. 연화 및 살균된 유기 물질은 이후 본 발명의 효소를 사용하여 발효 챔버 내에서 발효되어 매쉬가 된다. 이 매쉬를 원심 분리, 증류 컬럼 및/또는 헤파 분해기에 의해 추가로 가공하여, 연료 예를 들어, 에탄올 및 메탄올, 그리고 동물 사료 보충물을 회수할 수 있게 된다. 예를 들어, 미국 특허 제 6,251,643호를 참조하시오.

[0600] 본 발명의 효소는 또한 산업 폐기물, 또는 동물의 사육 시설로부터 배출되는 배설물의 악취를 감소시키기 위한 공정, 예를 들어 전처리 공정에서 사용될 수 있다. 예를 들어, 본 발명의 효소는 배설물 수용 시설에서 동물 배설물을 처리하여, 악취가 줄어든 다량의 유기 물질의 효율적 분해를 증대시키는 데 사용될 수 있다. 이러한 공정은 또한 황화물 이용 박테리아 및 유기물 분해 박테리아 및 (본 발명의 효소 이외의) 분해 효소를 접종하는 과정을 포함할 수도 있다. 예를 들어, 미국 특허 제5,958,758호를 참조할 수 있다.

[0601] 본 발명의 효소는 또한 이동 시스템 예를 들어, 수성의 유해한 폐기물을 생물적으로 환경 정화시키기 위한 회분형 반응기에서 사용되기도 한다(예를 들어, 미국 특허 제5,833,857호). 회분형 반응기는 순환식으로 작동하는 큰 용기로서, 이 용기에서는 박테리아(예를 들어, 본 발명의 효소를 발현하는 박테리아)가 이 반응기에 공급되는 영양분에 의해 효율적인 상태로 유지된다. 이러한 시스템은 유출물이 반응기에 전달될 수 있거나, 또는 반응기가 폐수 처리 시스템에 설치된 경우에 사용될 수 있다. 본 발명의 효소는 또한 작은 규모로 사용되는 처리 시스템 또는 일시적으로 멀리 떨어져 있는 시설 예를 들어, 휴대용, 고 용량, 고 효율 및 다방면의 폐수 처리 시스템에 사용될 수도 있다.

[0602] 본 발명의 폐기물 처리 공정은 기타 효소들 예를 들어, 기타 셀룰라제 예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 효소, 카탈라제, 락카제, 기타 셀룰라제, 엔도글리코시다제, 엔도-베타-1,4-락카제, 아밀로글루코시다제, 기타 글루코시다제, 글루코스 이소머라제, 글리코실트랜스퍼라제, 리파제, 포스포리파제, 리폭시게나제, 베타-락카제, 엔도-베타-1,3(4)-락카제, 큐티나제, 퍼옥시다제, 아밀라제, 글루코아밀라제, 펙티나제, 리덕타제, 옥시다제, 디카르복실라제, 페놀록시다제, 리그니나제, 폴루라나제, 피타제, 아라비나나제, 헤미셀룰라제, 만난아제, 자일로락카제, 자일라나제, 펙틴 아세틸 에스터라제, 람노갈락투로난 아세틸 에스터라제, 프로테아제, 펩티다제, 프로티나제, 폴리갈락투로나제, 람노갈락투

로나제, 갈락타나제, 펙틴 라이아제, 트랜스글루타미나제, 펙틴 메틸에스터라제, 기타 셀로바이오히이드롤라제 및/또는 트랜스글루타미나제를 경우에 따라 조합하여 사용하는 것을 포함할 수 있다.

[0603] 세제 조성물

[0604] 본 발명은 본 발명의 폴리펩티드(예를 들어, 셀룰라제, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 활성을 보유하는 효소)를 하나 이상 포함하는 세제 조성물과 이 조성물을 제조하는 방법 그리고 이 조성물을 사용하는 방법을 제공한다. 본 발명은 세제 조성물을 제조 및 사용하는 모든 방법들[예를 들어, 미국 특허 제6,413,928호; 동 제6,399,561호; 동 제6,365,561호; 및 동 제6,380,147호]를 통합한다. 세제 조성물은 1부 및 2부 수성 조성물, 비수성 액체 조성물, 성형 고체(cast solid), 과립 형태, 입자 형태, 압축 정제, 겔 및/또는 페이스트 및 슬러리 형태일 수 있다. 본 발명은 또한 상기와 같은 세제 조성물을 사용하여 음식물의 찌든 때, 음식물 잔류물 및 기타 소량의 식품 조성물을 신속하게 제거할 수 있는 방법을 제공한다. 본 발명의 효소는 전분 다당류를 촉매 가수분해하여 말라붙은 얼룩 제거에 효과적일 수 있다. 본 발명의 효소는 설거지용 세제 및 식물 세탁용 세제로서 사용될 수 있다.

[0605] 실제 활성 효소 함량은 세제 용액이 바람직한 효소 활성을 가진다고 하였을 때, 세제 조성물의 제조 방법에 따라서 달라지는데 이는 중요한 문제가 아니다. 하나의 측면에서, 최종 용액 중에 존재하는 글루코시다제의 양은 세제 조성물 1g당 약 0.001~0.5mg이다. 본 발명의 방법 및 생산물에 사용하기 위해 선택된 특정 효소는 최종 효용 조건 예를 들어, 생산물의 물리적 형태, pH 적용 범위, 온도 적용 범위 및 분해 또는 변형될 오염원의 종류와 같은 조건에 따라서 달라진다. 효소는 효용 조건에 대하여 주어진 임의의 세트에 대해 최적 활성 및 안정성을 제공하도록 선택될 수 있다. 하나의 측면에서, 본 발명의 폴리펩티드는 pH 약 4~약 12, 및 온도 약 20~약 95℃에서 활성을 갖는다. 본 발명의 세제는 양이온계, 반극성 비이온계 또는 쌍극자 이온계 계면 활성제 또는 이의 혼합물을 포함할 수 있다.

[0606] 본 발명의 효소(예를 들어, 셀룰라제, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코시다제 활성을 보유하는 효소)는 분말 세제 및 액체 세제(pH 4.0~12.0이고, 함량은 약 0.01~5 중량%(바람직하게는 0.1~0.5 중량%))로 제형화될 수 있다. 이러한 세제 조성물은 또한 기타 효소 예를 들어, 공지된 프로테아제, 셀룰라제, 리파제 또는 엔도글리코시다제, 그리고 강화제 및 안정화제를 포함할 수도 있다. 본 발명의 효소를 통상의 세정용 조성물에 첨가할 경우, 사용상 어떠한 제한도 따르지 않는다. 다시 말하면, 세제에 적당한 온도 및 pH는 또한 pH가 상기 범위 내이고, 온도가 전술한 효소의 변성 온도 이하인 한, 본 발명의 조성물에도 적당한 것이다. 뿐만 아니라, 본 발명의 폴리펩티드는 세제 없이도 단독으로, 또는 강화제 및 안정화제와 함께 세정용 조성물에 사용될 수 있다.

[0607] 본 발명은 단단한 표면을 세정하는 세제 조성물, 직물을 세정하는 세제 조성물, 설거지용 조성물, 구강 세정용 조성물, 의치 세정용 조성물 및 콘택트 렌즈 세정 용액을 비롯한 세정 조성물을 제공한다.

[0608] 하나의 측면에서, 본 발명은 세척에 충분한 조건 하에서 본 발명의 폴리펩티드와 사물을 접촉시키는 단계를 포함하는, 사물을 세척하는 방법을 제공한다. 본 발명의 폴리펩티드는 세제 첨가물로서 포함될 수 있다. 본 발명의 세제 조성물은 예를 들어, 본 발명의 폴리펩티드를 포함하는, 손 세탁용 또는 기계 세탁용 세제 조성물의 형태로 제형화될 수 있다. 오염된 직물을 전 처리하는데에 적당한 세탁용 첨가제는 본 발명의 폴리펩티드를 포함할 수 있다. 직물 유연제 조성물은 본 발명의 폴리펩티드를 포함할 수 있다. 대안적으로, 본 발명의 폴리펩티드는 가정에서 일반적으로 단단한 표면 세정용으로서 사용되는 세제 조성물로서 제형화될 수 있다. 대안적인 측면에서, 본 발명의 세제 첨가물 및 세제 조성물은 하나 이상의 기타 효소 예를 들어, 프로테아제, 리파제, 큐티나제, 기타 글루코시다제, 탄수화물 분해 효소, 기타 셀룰라제, 펙티나제, 만난아제, 아라비나제, 갈락타나제, 자일라나제, 옥시다제 예를 들어, 탁타제 및/또는 퍼옥시다제를 포함할 수 있다. 본 발명의 효소(들)의 특성은 선택된 세제와 혼화 가능한 것으로 선택되며(즉, pH-최적, 기타 효소 및 비-효소 성분 등과의 혼화성), 이 효소(들)는 유효량만큼 존재한다. 하나의 측면에서, 본 발명의 효소는 직물로부터 악취를 풍기게 하는 물질들을 제거하는데에 사용된다. 본 발명을 실시하는데에 사용될 수 있는 다양한 세제 조성물 및 이를 제조하는 방법에 관하여는 예를 들어, 미국 특허 제6,333,301호; 동 제6,329,333호; 동 제6,326,341호; 동 제6,297,038호; 동 제6,309,871호; 동 제6,204,232호; 동 제6,197,070호; 동 제5,856,164호에 개시되어 있다.

[0609] 본 발명의 세제 및 이와 관련 방법은 또한 기타 효소들 예를 들어, 트립토판나제 또는 티로신 디카르복실라제, 락카제, 카탈라제, 락카제, 기타 셀룰라제, 엔도글리코시다제, 엔도-베타-1,4-락카제, 아밀로글루코시다제, 기타 글루코시다제, 글루코스 이소머라제, 글리코실트랜스퍼라제, 리파제, 포스포리파제, 리폭시게나제, 베타-락카제, 엔도-베타-1,3(4)-락카제, 큐티나제, 퍼옥시다제, 아밀라제, 글루코아밀라제, 펙티나제, 리덕타제, 옥시

다제, 디카르복실라제, 페놀록시다제, 리그니나제, 풀루라나제, 아라비나나제, 헤미셀룰라제, 만난아제, 자일로 락카제, 자일라나제, 펙틴 아세틸 에스터라제, 람노갈락투로난 아세틸 에스터라제, 프로테아제, 펩티다제, 프로 티나제, 폴리갈락투로나제, 람노갈락투로나제, 갈락타나제, 펙틴 라이아제, 트랜스글루타미나제, 펙틴 메틸에스 터라제, 기타 셀로바이오히이드롤라제 및/또는 트랜스글루타미나제를 경우에 따라 조합하여 사용하는 것을 포함 할 수도 있다.

[0610] 직물 및 천의 처리

[0611] 본 발명은 본 발명의 폴리펩티드 예를 들어, 셀룰라제, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/ 또는 베타-글루코시다제 활성을 보유하는 효소 중 하나 이상을 사용하여 직물 및 천을 처리하는 방법을 제공한다. 본 발명의 폴리펩티드는 당 업계에 널리 공지된 임의의 직물-처리 방법에서 사용될 수 있다[예를 들어, 미 국 특허 제6,077,316호 참조]. 예를 들어, 하나의 측면에서, 직물의 촉감과 외견은, 직물과 본 발명의 효소 용 액을 접촉시키는 단계를 포함하는 방법에 의해 개선된다. 하나의 측면에서, 상기 직물은 가압하에서 용액으로 처리된다.

[0612] 하나의 측면에서, 본 발명의 효소는 직물을 직조할 때 또는 그 후, 아니면 디사이징(desizing) 단계 중, 또는 직물 가공 단계 중 하나 이상의 추가의 단계를 수행하는 중에 도입된다. 직물을 직조할 때에, 실은 굉장한 기계 적 장력에 노출된다. 직조 기계에서 직조하기 이전에, 경사는 종종 전분이나 전분 유도체로 사이징(sizing)되어 코팅되며, 이에 따라서 인장 강도가 증가하게 되고 파열이 방지된다. 본 발명의 효소는 이와 같은 사이징 전분 또는 전분 유도체를 제거하기 위해 가하여진다. 천의 직조 과정을 마치면, 이후 직물은 디사이징 단계로 도입될 수 있다. 이후에도 하나 이상의 추가 직물 처리 단계가 수행될 수 있다. 디사이징은 천으로부터 풀을 제거하는 과정이다. 직조 후, 직물을 추가로 처리하기 전에 풀 코팅을 제거해야 하며, 이와 같이 제거해야 균일하고 세척 에 강한 직물을 얻을 수 있다. 본 발명은 본 발명의 효소 작용에 의해 풀을 효소로 가수분해하는 단계를 포함하 는 디사이징 방법을 제공한다.

[0613] 본 발명의 효소(예를 들어, 셀룰라제, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코 시다제 활성을 보유하는 효소)는 세제 첨가물 예를 들어, 수성 조성물의 형태인 세제 첨가물로서, 직물 예를 들 어, 면-함유 직물을 디사이징하는데에 사용될 수 있다. 본 발명은 남색으로 염색된 데님 직물 및 의복에 스톤 워시(stonewash) 느낌을 주는 방법을 제공한다. 천을 제조하기 위하여, 직물을 자른 후 이를 천이나 옷으로 재 봉한 후 피니쉬 처리할 수 있다. 특히, 데님 진을 제조하기 위해서 효소를 사용하는 독특한 피니쉬 처리 방법이 개발되었다. 보통 데님 옷의 피니쉬 처리는, 직물을 부드럽게 만들고 면을 후속의 효소 피니쉬 처리 단계에 더 욱 적합하도록 만들기 위해 아밀로스 분해 효소를 옷에 적용시키는 동안, 효소를 사용하는 디사이징 단계로부터 개시된다. 본 발명은, 본 발명의 효소를 사용하여, 데님 옷을 피니쉬 처리하는 방법(예를 들어, "바이오-스토닝 과정(bio-stoning process)"), 효소를 사용하여 디사이징하는 방법 및 직물을 부드럽게 만드는 방법을 제공한다. 본 발명은 디사이징 및/또는 피니쉬 처리 공정에서 데님 옷을 신속하게 부드럽게 만드는 방법을 제공 한다.

[0614] 본 발명은 또한 본 발명의 효소(예를 들어, 셀룰라제, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/ 또는 베타-글루코시다제 활성을 보유하는 효소)를 포함하는 소독제를 제공한다.

[0615] 본 발명의 직물 또는 천의 처리 방법은 또한 기타 효소들 예를 들어, 트립토판나제 또는 티로신 디카르복실라제, 락카제, 카탈라제, 락카제, 기타 셀룰라제, 엔도글리코시다제, 엔도-베타-1,4-락카제, 아밀로 글루코시다제, 기타 글루코시다제, 글루코스 이소머라제, 글리코실트랜스퍼라제, 리파제, 포스포리파제, 리폭시 게나제, 베타-락카제, 엔도-베타-1,3(4)-락카제, 큐티나제, 피옥시다제, 아밀라제, 글루코아밀라제, 펙티나제, 리덕타제, 옥시다제, 디카르복실라제, 페놀록시다제, 리그니나제, 풀루라나제, 아라비나나제, 헤미셀룰라제, 만 난아제, 자일로락카제, 자일라나제, 펙틴 아세틸 에스터라제, 람노갈락투로난 아세틸 에스터라제, 프로테아제, 펩티다제, 프로티나제, 폴리갈락투로나제, 람노갈락투로나제, 갈락타나제, 펙틴 라이아제, 트랜스글루타미나제, 펙틴 메틸에스터라제, 기타 셀로바이오히이드롤라제 및/또는 트랜스글루타미나제를 경우에 따라 조합하여 사용 하는 것을 포함할 수도 있다.

[0616] 종이 및 펄프 처리

[0617] 본 발명의 효소(예를 들어, 셀룰라제, 엔도글루카나제, 셀로바이오히이드롤라제, 만난아제 및/또는 베타-글루코 시다제 활성을 보유하는 효소)는 종이 또는 펄프 처리 또는 종이의 탈묵(deinking)에 사용될 수 있다. 예를 들 어, 하나의 측면에서, 본 발명은 본 발명의 효소를 사용하는 종이 처리 방법을 제공한다. 하나의 측면에서, 본

발명의 효소는 종이에 있는 전분을 변형시켜 이를 액화 형태의 것으로 전환시키는데에 사용될 수 있다. 다른 측면에서, 화학적 및 효소적 탈묵 공정시 재활용 복사 용지의 종이 성분들을 처리하는 방법에 관한 것이다. 하나의 측면에서, 본 발명의 효소는 다른 효소 예를 들어, 기타 셀룰라제(기타 엔도글루카나제, 셀로바이오하이드롤라제 및/또는 베타-글루코시다제 포함)와 함께 사용될 수 있다. 목재, 종이, 종이 제품 또는 펄프는 다음과 같은 3가지 과정에 의해서 처리될 수 있다: 1) 본 발명의 효소의 존재 하에서 분해하는 과정; 2) 탈묵 화학 물질 및 본 발명의 효소를 사용하여 분해하는 과정, 및/또는 3) 본 발명의 효소에 침지시킨 후 분해하는 과정. 본 발명의 효소에 의해 처리된 재활용 종이는 셀룰라제만으로 처리된 종이에 비해, 토너 입자가 제거되었기 때문에 백색도(brightness)가 더욱 뛰어나다. 본 발명은 어떠한 특정 기작에 의해 한정되지 않지만, 본 발명의 효소의 기능은 펄프 현탁시, 표면 활성화 제제로서의 거동으로 기인할 수 있다.

[0618] 본 발명은 본 발명의 폴리펩티드 중 하나 이상을 사용하여 종이 및 종이 펄프를 처리하는 방법을 제공한다. 본 발명의 폴리펩티드는 당 업계에 널리 공지된 임의의 종이-처리 방법 또는 펄프-처리 방법에 사용될 수 있다[예를 들어, 미국 특허 제6,241,849호; 동 제6,066,233호; 및 동 제5,582,681호]. 예를 들어, 하나의 측면에서, 본 발명은 염료를 함유하는 인쇄 용지를 탈묵 및 탈색시키는 방법을 제공하는데, 이 방법은 인쇄된 종이를 겉쪽하게 만들어 펄프 슬러리를 형성시키는 단계, 및 본 발명의 효소의 존재 하에서(기타 효소도 첨가될 수 있음) 이 펄프 슬러리로부터 잉크를 제거하는 단계를 포함한다. 다른 측면에서, 본 발명은 본 발명의 효소(기타 효소 예를 들어, 펙티나제 효소도 포함할 수 있음)를 포함하는 효소 혼합물을 펄프에 첨가한 후, 펄프를 효소로 처리하는 반응을 유발시키는 조건 하에서 이 펄프를 처리함으로써, 펄프 예를 들어, 2차 섬유로부터 제조되는 펄프의 여수도(freeness)를 강화하는 방법을 제공한다. 효소로 처리된 펄프의 여수도는 백색도를 떨어뜨리지 않고서도 이 2차 섬유 펄프의 초기 여수도에 비해 증가하게 된다.

[0619] 본 발명의 종이, 목재 또는 펄프 처리 또는 재생 방법은 또한 기타 효소들 예를 들어, 트립토판나제 또는 티로신 디카르복실라제, 락카제, 카탈라제, 락카제, 기타 셀룰라제, 엔도글리코시다제, 엔도-베타-1,4-락카제, 아밀로글루코시다제, 기타 글루코시다제, 글루코스 이소머라제, 글리코실트랜스퍼라제, 리파제, 포스포리파제, 리폭시게나제, 베타-락카제, 엔도-베타-1,3(4)-락카제, 큐티나제, 피옥시다제, 아밀라제, 글루코아밀라제, 펙티나제, 리덕타제, 옥시다제, 디카르복실라제, 페놀록시다제, 리그니나제, 폴루라나제, 아라비나나제, 헤미셀룰라제, 만난아제, 자일로락카제, 자일라나제, 펙틴 아세틸 에스테라제, 람노갈락투로난 아세틸 에스테라제, 프로테아제, 펩티다제, 프로티나제, 폴리갈락투로나제, 람노갈락투로나제, 갈락타나제, 펙틴 라이아제, 트랜스글루타미나제, 펙틴 메틸에스테라제, 기타 셀로바이오하이드롤라제 및/또는 트랜스글루타미나제를 경우에 따라 조합하여 사용하는 것을 포함할 수도 있다.

[0620] *해리: 리그노셀룰로스 물질의 처리*

[0621] 본 발명은 또한 섬유의 특성을 개선시키는 데에 유효한 양만큼의 본 발명의 폴리펩티드(예를 들어, 셀룰라제, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 활성을 보유하는 효소)로 섬유를 처리하는, 리그노셀룰로스 섬유의 처리 방법을 제공한다. 본 발명의 효소는 또한 특히, pH7 이상에서 해리 또는 재생 공정이 진행되는 경우와, 본 발명의 효소가 강화된 전분을 분해함으로써 폐기물의 분해를 촉진할 수 있는 경우, 전분 강화 폐 용지 및 판지로부터 리그노셀룰로스 물질 예를 들어, 펄프, 종이 및 판지를 생산 또는 재생하는데에 사용될 수도 있다. 본 발명의 효소는 전분-코팅된 인쇄 용지로부터 제지용 펄프를 생산하는 공정에서도 유용할 수 있다. 이 공정은 예를 들어, WO 95/14807에 기술되어 있는 바와 같이 진행될 수 있다. 대표적인 공정은, 종이를 분해하여 펄프를 만드는 과정, 분해시키기 이전, 분해시키는 동안 또는 분해시킨 후에 전분-분해 효소를 사용하여 이 펄프를 처리하는 과정, 그리고 분해 및 효소 처리 후 펄프로부터 잉크 입자를 분리하는 과정을 포함한다. 미국 특허 제6,309,871호 및 본원에 인용된 기타 미국 특허를 참조하십시오. 그러므로, 본 발명은 재생 용지 펄프를 효소에 의해 탈묵하는 방법을 포함하는데, 이 방법에 있어서 폴리펩티드는 섬유 표면을 효율적으로 탈묵하는데에 유효한 양으로 가하여진다.

[0622] *양조 및 발효*

[0623] 본 발명은 본 발명의 효소 예를 들어, 셀룰라제, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 활성을 보유하는 효소를 포함하는, 맥주의 양조(예를 들어, 발효) 방법을 제공한다. 하나의 대표적인 방법에 있어서, 전분-함유 원료는 분해 및 가공되어 맥아즙을 형성하게 된다. 본 발명의 효소는 발효 과정 중 임의의 시점에 사용된다. 예를 들어, 본 발명의 효소는 보리 맥아즙을 가공하는데에 사용될 수 있다. 맥주 양조시 필요한 주원료는 보리 맥아즙이다. 이는 3가지 단계로 이루어진 공정일 수 있다. 첫 번째, 보리 알곡을 침지시켜 수분 함량을 증가시킬 수 있다(예를 들어, 수분 함량 약 40%까지 증가). 두 번째, 지베렐린의 제

어 하에 효소 합성이 촉진될 때, 15~25℃에서 3~6일 동안 항온 처리하여 이 알곡을 발아시킬 수 있다. 이 시기에 효소 수준은 상당히 높아진다. 하나의 측면에서, 본 발명의 효소는 공정 중 이 시기의(또는 다른 시기의) 단계에 첨가된다. 효소의 작용으로 인하여, 발효 가능한 환원당이 증가하게 된다. 그 결과를 당화력(diastatic power; DP)으로 나타낼 수 있다[이 경우, 당화력은 5일 경과시 12℃에서, 약 80에서 190으로 상승함].

[0624] 본 발명의 효소는 예를 들어, 미국 특허 제5,762,991호; 동 제5,536,650호; 동 제5,405,624호; 동 제5,021,246호; 동 제4,788,066호에 개시된 바와 같은, 임의의 맥주 제조 공정에서 사용될 수 있다.

[0625] 지하에서 생성된 유체의 생산량 증가

[0626] 본 발명은 또한 본 발명의 효소(예를 들어, 셀룰라제, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 활성을 보유하는 효소)를 사용하는 방법을 포함하기도 하는데, 이 방법은 점성이고, 전분을 함유하며, 생산 공정 중 형성된 유해 유체를 제거함으로써 지하에서 생성되는 유체의 생산량을 증가시키는 데; 이러한 유체는 깊이 파인 구덩이 주변을 둘러싸고 있는 땅의 지하에서 형성될 수 있다. 그러므로, 본 발명의 이러한 방법은 깊이 파인 구덩이로부터 흘러들어올 수 있는 유체를 생성한다. 본 발명의 이와 같은 방법은 또한 유체가 기대 유속 이하로 생성되어 유체의 생산량을 감소시키는, 유해 유체로 인한 문제점을 해결하기 위한 것이기도 하다. 하나의 측면에서, 본 발명은 수성 유체와 본 발명의 폴리펩티드를 함께 배합하는 단계; 효소 처리액을 깊이 파인 구덩이의 원하는 위치에 펌핑하는 단계; 효소 처리함으로써 점성이 있고, 전분을 함유하는 유해한 유체로 분해하여, 이 유체가 지하 형성부로부터 구덩이 표면으로 배출될 수 있는 단계를 포함하며; 여기서, 상기 효소 처리액은 전분 함유 유체의 알파 글루코시드 결합을 효과적으로 공격한다.

[0627] 본 발명의 지하 생성 효소 처리 공정은 또한 기타 효소들 예를 들어, 트립토판나제 또는 티로신 디카르복실라제, 락카제, 카탈라제, 락카제, 기타 셀룰라제, 엔도글리코시다제, 엔도-베타-1,4-락카제, 아밀로 글루코시다제, 기타 글루코시다제, 글루코스 이소머라제, 글리코실트랜스퍼라제, 리파제, 포스포리파제, 리폭시 게나제, 베타-락카제, 엔도-베타-1,3(4)-락카제, 큐티나제, 퍼옥시다제, 아밀라제, 글루코아밀라제, 펙티나제, 리덕타제, 옥시다제, 디카르복실라제, 페놀옥시다제, 리그니나제, 풀루라나제, 아라비나나제, 헤미셀룰라제, 만 난아제, 자일로라카제, 자일라나제, 펙틴 아세틸 에스터라제, 람노갈락투로난 아세틸 에스터라제, 프로테아제, 펩티다제, 프로티나제, 폴리갈락투로나제, 람노갈락투로나제, 갈락타나제, 펙틴 라이아제, 트랜스글루타미나제, 펙틴 메틸에스터라제, 기타 셀로바이오하이드롤라제 및/또는 트랜스글루타미나제를 경우에 따라 조합하여 사용하는 것을 포함할 수도 있다.

[0628] 약학 조성물 및 식품 보조제

[0629] 본 발명은 또한 본 발명의 셀룰라제(예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 활성을 보유하는 효소)를 포함하는 약학 조성물 및 식품 보조제(예를 들어, 다이어트 보조제)을 제공한다. 셀룰라제 활성은 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 활성을 포함한다. 하나의 측면에서, 셀룰로스 또는 리그노셀룰로스 성분의 함량이 높은 사료 및 식품의 소화 능력을 개선시키기 위하여, 약학 조성물 및 식품 보조제(예를 들어, 다이어트 보조제)는 경구 섭취용으로 제형화 된다.

[0630] 치주 질환 치료용 화합물은 본 발명의 효소 예를 들어, 미국 특허 제6,776,979호에 개시되어 있는 바와 같은 효소를 포함할 수 있다. 산성 대장 증후군(acidic gut syndrome)을 치료 또는 예방하는 조성물 및 방법은 본 발명의 효소 예를 들어, 미국 특허 제6,468,964호에 개시되어 있는 효소를 포함할 수 있다.

[0631] 다른 측면에서, 상처 드레싱 및 임플란트 등은 항균성 효소(예를 들어, 항생제 작용성 효소) 예를 들어, 본 발명의 효소(예를 들어, 본 발명의 대표적인 서열)를 포함한다. 본 발명의 효소는 또한 알지네이트 드레싱, 항균 막 드레싱, 화상 환자 드레싱, 압박 붕대, 진단 도구, 젤 드레싱, 하이드로-선택성 드레싱(hydro-selective dressing), 하이드로젤(소포) 드레싱, 하이드로콜로이드 드레싱, I.V. 드레싱, 절개 드레이프, 저 점착성 드레싱, 악취 흡수 드레싱, 페이스트 붕대, 수술 후 드레싱, 상처 관리, 피부 보호, 투명 필름 드레싱 및/또는 상처 봉합 시에 사용될 수도 있다. 본 발명의 효소는 압박 레사, 다리의 궤양, 화상, 당뇨병에 의한 족부 궤양, 상처, IV 고정, 수 상처 및 기타 소소한 상처를 치료하기 위한 환부 소독 및 환부 침상 준비에 사용될 수 있다. 본 발명의 효소는 살균 효소 창상 세정용 조성물 예를 들어, 연고에 사용될 수 있다. 다양한 측면에서, 셀룰라제는 정제, 젤, 환약, 임플란트, 액체, 스프레이, 분말, 식품, 사료 펠릿 또는 캡슐화된 제형으로 제형화된다.

[0632] 생물 방어용

[0633] 다른 측면에서, 본 발명의 셀룰라제(예를 들어, 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는

베타-글루코시다제 활성을 보유하는 효소)는 생물 방어(예를 들어, 리그노셀룰로스 물질을 포함하는, 포자 또는 박테리아 박멸)에 사용될 수 있다. 본 발명의 셀룰라제를 생물 방어시에 사용하면, 이 셀룰라제를 이용하여 현재 알려지지 않은 임의의 제제 또는 미래의 생물전 제제에 대처하는 제제를 신속히 개발할 수 있다는 점에서 매우 유리하다. 뿐만 아니라, 본 발명의 셀룰라제는 오염된 환경으로부터 오염원을 제거하는데에 사용될 수 있다. 하나의 측면에서, 본 발명은 셀룰라제 활성을 보유하는 폴리펩티드를 포함하는 생물 방어 제제 또는 생물 해독제를 제공하며, 여기서, 상기 폴리펩티드는 본 발명의 서열(예를 들어, 본 발명의 대표 서열), 또는 본 발명의 핵산에 의해 암호화되는 폴리펩티드(예를 들어, 본 발명의 대표 서열)를 포함하고, 경우에 따라, 상기 폴리펩티드는 엔도글루카나제, 셀로바이오하이드롤라제, 만난아제 및/또는 베타-글루코시다제 활성을 가진다.

[0634] 참고 문헌 목록

1. Sambrook, J. and Russell, D.W. 2001. *Molecular Cloning: A Laboratory Manual*. Third Edition. Cold Spring Harbor Laboratory Press, New York.
2. Benhar, I. *Biotechnological applications of phage and cell display*. *Biotechnology Advances* 19, 1-13. 2001.
3. Coutinho, P. M. and Henrissat, B. Carbohydrate-Active Enzymes server at URL: <http://afmb.cnrs-mrs.fr/~cazy/CAZY/index.html>. 1999.
4. Felix, C. R. and L. G. Ljungdahl. 1993. The cellulosome: the exocellular organelle of *Clostridium*. *Annu. Rev. Microbiol* 47:791-819.:791-819.
5. Gray, K. A., T. H. Richardson, K. Kretz, J. M. Short, F. Bartnek, Knowles R., L. Kan, Swanson P.E., and Robertson D.E. 2001. Rapid evolution of reversible denaturation and elevated melting temperature in a microbial haloalkane dehalogenase. *Advanced Synthesis and Catalysis* 343:607-617.
6. Guttman, A., F. T. Chen, R. A. Evangelista, and N. Cooke. 1996. High-resolution capillary gel electrophoresis of reducing oligosaccharides labeled with 1-aminopyrene-3,6,8-trisulfonate. *Anal. Biochem* 233:234-242.
7. Harjunpaa, V., A. Teleman, A. Koivula, L. Ruohonen, T. T. Teeri, O. Teleman, and T. Drakenberg. 1996. Cello-oligosaccharide hydrolysis by cellobiohydrolase II from *Trichoderma reesei*. Association and rate constants derived from an analysis of progress curves. *Eur. J Biochem* 240:584-591.
8. Himmel, M. E., M. F. Ruth, and C. E. Wyman. 1999. Cellulase for commodity products from cellulosic biomass. *Curr. Opin. Biotechnol* 10:358-364.
9. Kerr, R. A. 1998. GEOLOGY: The Next Oil Crisis Looms Large--and Perhaps Close. *Science* 281:1128.
10. Kerr, R. A. 2000. OIL OUTLOOK: USGS Optimistic on World Oil Prospects. *Science* 289:237.
11. King, R. W., K. D. Lustig, P. T. Stukenberg, T. J. McGarry, and M. W. Kirschner. 1997. Expression cloning in the test tube. *Science* 277:973-974.
12. Kuritz, T. 1999. An easy colorimetric assay for screening and qualitative assessment of deiodination and dehalogenation by bacterial cultures. *Lett. Appl Microbiol* 28:445-447.
13. Lundberg, K. S., P. L. Kretz, G. S. Provost, and J. M. Short. 1993. The use of selection in recovery of transgenic targets for mutation analysis. *Mutat. Res.* 301:99-105.
14. MacKenzie, L. F., G. Sulzenbacher, C. Divne, T. A. Jones, H. F. Woldike, M. Schulein, S. G. Withers, and G. J. Davies. 1998. Crystal structure of the family 7 endoglucanase I (Cel7B) from *Humicola insolens* at 2.2 Å resolution and identification of the catalytic nucleophile by trapping of the covalent glycosyl-enzyme intermediate. *Biochem J* 335:409-416.
15. Richardson, T. H., X. Tan, G. Frey, W. Callen, M. Cabell, D. Lam, J. Macomber, J. M. Short, D. E. Robertson, and C. Miller. 2002. A novel, high performance enzyme for starch liquefaction. Discovery and optimization of a low pH, thermostable alpha-amylase. *J Biol Chem* 277:26501-26507.
16. Sakon, J., D. Irwin, D. B. Wilson, and P. A. Karplus. 1997. Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*. *Nat. Struct. Biol* 4:810-818.
17. Short, J. M., J. M. Fernandez, J. A. Sorge, and W. D. Huse. 1988. Lambda ZAP: a bacteriophage lambda expression vector with in vivo excision properties. *Nucleic Acids Res.* 16:7583-7600.
18. Snustad, D. P., J. P. Hunsperger, B. M. Chereskin, and J. Messing. 1988. Maize glutamine synthetase cDNAs: isolation by direct genetic selection in *Escherichia coli*. *Genetics* 120:1111-1123.
19. Varrot, A., S. Hastrup, M. Schulein, and G. J. Davies. 1999. Crystal structure of the catalytic core domain of the family 6 cellobiohydrolase II, Cel6A, from *Humicola insolens*, at 1.92 Å resolution. *Biochem J* 337:297-304.
20. Yano, T., S. Oue, and H. Kagamiyama. 1998. Directed evolution of an aspartate aminotransferase with new substrate specificities. *Proc. Natl. Acad. Sci U. S. A* 95:5511-5515.
21. Zverlov, V. V., G. A. Velikodvorskaya, and W. H. Schwarz. 2002. A newly described cellulosomal cellobiohydrolase, CelO, from *Clostridium thermocellum*: investigation of the exo-mode of hydrolysis, and binding capacity to crystalline cellulose. *Microbiology* 148:247-255.

하기 실시예는 예시를 위해 제시된 것으로서 청구범위에 청구된 발명을 제한하는 것이 아니다.

실시예

[0639] 실시예 1: 기가매트릭스™ 스크리닝

[0640] 하나의 측면에서, 본 발명의 방법은 다이버사 코퍼레이션(Diversa Corporation)의 소유인 기가매트릭스™ 플랫폼을 사용한다[PCT 특허 공보 WO 01/38583; 미국 특허 출원 제20050046833호; 동 제20020080350호; 미국 특허 제6,918,738호; 의장 등록 제D480,814호]. 예를 들어, 하나의 측면에서, 기가매트릭스™는 폴리펩티드가 셀룰라제 활성을 보유하는지 여부와, 이것이 본 발명의 범위 내에 포함되는지 여부를 확인하는 방법, 또는 셀룰라제 활성을 보유하는 폴리펩티드를 동정 및 분리하는 방법에 사용된다.

[0641] 기가매트릭스™ 플랫폼은 100,000개의 웰을 가지는 미세 평판(통상의 96 웰 평판과 치수가 동일함) 상에 장착되어 있는 초 고 처리량의 스크린을 포함할 수 있다. 본 실시예에서, 기가매트릭스™ 스크리닝은 CBH에 의해 이미 확인된 활성을 바탕으로 하여 2개의 기질을 사용함으로써 수행되었다. 메틸-움베리페틸 셀로바이오시드(MUC) 및 메틸움베리페틸 락토시드(MUL)를 시험하였다. 기질은 세포로 확산되고, 그에 따라서 형광성은 더욱 용이하게 검출될 수 있었기 때문에, 상이한 클론에 대한 파지미드가 스크리닝되었다. 베타-글루코시다제가 결합된 숙주 균주를 사용하여 락토시드 기질에 대한 활성을 감소시켰다. 상기 락토시드 기질은 소수의 히트를 발생시켰는데, 이를 통하여 셀로바이오스 기질보다 더욱 특이하다는 사실을 알 수 있었다. 뿐만 아니라, 락토시드 기질은 소수의 베타-갈락토시다제 히트를 발생시켰다. 이러한 기질들을 스크리닝에 사용할 수 있는지 여부를 시험하기 위하여, 라이브러리가 종전의 스크리닝 프로그램으로부터 엔도글루카나제 히트를 발생시켰다는 사실에 근거하여 14개의 라이브러리를 스크리닝용으로 선택하였다. 스크리닝된 라이브러리 중에서, 11개의 라이브러리로부터는 총 50개의 1차 히트가 발생하였다. 2차적 스크리닝은 아가 평판 상에 클론을 도말한 후, 콜로니를 찍어서 배지와 MUL을 함유하는 384 웰 평판에 접종하는 단계를 포함한다. MUL에 대해 활성인 클론은 비활성 클론의 백그라운드로부터 분화된다. 이후, 각각의 클론들을 밤새도록 생육시키고, 형광도를 측정하여, 활성이 가장 큰 히트를 선택해서 서열 결정하였다.

[0642] 이 히트로부터 유래하는 모든 게놈형 클론 삽입부를 서열 결정하였다. 일반적으로, 히트는 1, 2, 5, 6, 10 및 16번을 포함하는 일부 상이한 글리코실 가수분해 효소 군으로부터 발생하였다. 기타 몇몇 히트는 개방 해독 틀이 임의의 공지된 글리코실 가수분해 효소 군에 상동성이 아닌 경우에 발견되었다. 또한, 히트 중 일부는 GTP 사이클로화이드롤라제 유전자를 암호화하였다.

[0643] [표 1a]

[0644]

기가매트릭스™ 히트에 관한 요약		
효소 번호	개방 해독틀 서열 번호	가장 상관성이 높은 BLAST
1	서열 번호 22(예를 들어, 서열 번호 21에 의해 암호화됨)	ORF001-제5군(셀룰라제)
1a	서열 번호 24(예를 들어, 서열 번호 23에 의해 암호화됨)	ORF003-제16군+CBM
2	서열 번호 26(예를 들어, 서열 번호 25에 의해 암호화됨)	ORF001-제1군(β -글루코시다제)
3	서열 번호 92(예를 들어, 서열 번호 91에 의해 암호화됨)	ORF001-제3군
3a	서열 번호 94(예를 들어, 서열 번호 93에 의해 암호화됨)	ORF002-알파-람노시다제
4	서열 번호 96(예를 들어, 서열 번호 95에 의해 암호화됨)	ORF001-제3군
4a	서열 번호 98(예를 들어, 서열 번호 97에 의해 암호화됨)	ORF003-베타-글루쿠로니다제
5	서열 번호128(예를 들어, 서열 번호 127에 의해 암호화됨)	ORF004-단쇄 탈수소 효소
5a	서열 번호130(예를 들어, 서열 번호 129에 의해 암호화됨)	ORF010-단쇄 탈수소 효소
6	서열 번호116(예를 들어, 서열 번호 115에 의해 암호화됨)	ORF004-단쇄 탈수소 효소
6a	서열 번호118(예를 들어, 서열 번호 117에 의해 암호화됨)	ORF011-단쇄 탈수소 효소

7	서열 번호 40(예를 들어, 서열 번호 39에 의해 암호화됨)	ORF004-추정 산화환원효소
8	서열 번호 42(예를 들어, 서열 번호 41에 의해 암호화됨)	ORF004-시스테인 tRNA 합성효소
8a	서열 번호 44(예를 들어, 서열 번호 43에 의해 암호화됨)	ORF001-가설단백질
9	서열 번호 54(예를 들어, 서열 번호 53에 의해 암호화됨)	ORF002-라디칼 SAM 군
10	서열 번호134(예를 들어, 서열 번호 133에 의해 암호화됨)	ORF006-제1군 (β-글루코시다제)
11	서열 번호 58(예를 들어, 서열 번호 57에 의해 암호화됨)	ORF001-서브틸리신 유사 프로테아제
12	서열 번호 46(예를 들어, 서열 번호 45에 의해 암호화됨)	ORF006-제1군 (β-글루코시다제)

[0645]

[표 1b]

[0646]

13	서열 번호 8(예를 들어, 서열 번호 7에 의해 암호화됨)	ORF003-이소시트레이트 탈수소효소
13a	서열 번호 10(예를 들어, 서열 번호 9에 의해 암호화됨)	ORF004-제10군 (자일라나제)
14	서열 번호 48(예를 들어, 서열 번호 47에 의해 암호화됨)	ORF002-제1군 (β-글루코시다제)
14a	서열 번호 50(예를 들어, 서열 번호 49에 의해 암호화됨)	ORF006-fdhd/narq 산화환원효소
15	서열 번호 4(예를 들어, 서열 번호 3에 의해 암호화됨)	ORF008-제1군 (β-글루코시다제)
15a	서열 번호 6(예를 들어, 서열 번호 5에 의해 암호화됨)	ORF012-제6군 (셀룰라제)
16	서열 번호136(예를 들어, 서열 번호135에 의해 암호화됨)	ORF001-셀룰라제 (글리코실 하이드롤라제 제5군)
17	서열 번호 56(예를 들어, 서열 번호 55에 의해 암호화됨)	ORF004-제1군 (β-글루코시다제)
18	서열 번호126(예를 들어, 서열 번호125에 의해 암호화됨)	ORF009-제1군 (β-글루코시다제)
19	서열 번호120(예를 들어, 서열 번호119에 의해 암호화됨)	ORF002-산화환원효소
19a	서열 번호122(예를 들어, 서열 번호111에 의해 암호화됨)	ORF004-제5군 (셀룰라제)
20	서열 번호124(예를 들어, 서열 번호123에 의해 암호화됨)	ORF006-제1군 (β-글루코시다제)
21	서열 번호132(예를 들어, 서열 번호131에 의해 암호화됨)	ORF007-제5군 (셀룰라제)
22	서열 번호 38(예를 들어, 서열 번호 37에 의해 암호화됨)	ORF011-제1군 (β-글루코시다제)
22a	서열 번호 36(예를 들어, 서열 번호 35에 의해 암호화됨)	ORF007-제5군 (셀룰라제)
23	서열 번호138(예를 들어, 서열 번호137에 의해 암호화됨)	ORF001-펩티다제 M37
24	서열 번호146(예를 들어, 서열 번호145에 의해 암호화됨)	ORF002-제1군 (β-글루코시다제)
25	서열 번호 22(예를 들어, 서열 번호 51에 의해 암호화됨)	ORF001-제5군 (셀룰라제)

[0647] [표 1c]

[0648]

26	서열 번호 20(예를 들어, 서열 번호 19에 의해 암호화됨)	ORF008-제10군 (자일라나제)
26a	서열 번호 18(예를 들어, 서열 번호 17에 의해 암호화됨)	ORF005-β-락타마제
27	서열 번호 16(예를 들어, 서열 번호 15에 의해 암호화됨)	ORF007-제1군 (β-글루코시다제)
27a	서열 번호 14(예를 들어, 서열 번호 13에 의해 암호화됨)	ORF005-NADH 의존적 탈수소 효소
27b	서열 번호 12(예를 들어, 서열 번호 11에 의해 암호화됨)	ORF003-NAD 결합 산화환원효소
28	서열 번호 28(예를 들어, 서열 번호 27에 의해 암호화됨)	ORF002-제1군 (β-글루코시다제)
29	서열 번호114(예를 들어, 서열 번호113에 의해 암호화됨)	ORF003-제10군
30	서열 번호 34(예를 들어, 서열 번호 33에 의해 암호화됨)	ORF006-제1군 (β-글루코시다제)
30a	서열 번호 32(예를 들어, 서열 번호 31에 의해 암호화됨)	ORF002- 셀로텍스트린 포스포릴라제
31	서열 번호 30(예를 들어, 서열 번호 29에 의해 암호화됨)	ORF004-제1군 (β-글루코시다제)
32	서열 번호100(예를 들어, 서열 번호 99에 의해 암호화됨)	ORF012-제1군 (β-글루코시다제)
33	서열 번호 84(예를 들어, 서열 번호 83에 의해 암호화됨)	ORF008-탈수소 효소
34	서열 번호102(예를 들어, 서열 번호101에 의해 암호화됨)	ORF003-제5군 (셀룰라제)
35	서열 번호140(예를 들어, 서열 번호139에 의해 암호화됨)	ORF001-트레오닌 탈수소 효소
36	서열 번호142(예를 들어, 서열 번호141에 의해 암호화됨)	ORF005-제1군 (β-글루코시다제)
37	서열 번호144(예를 들어, 서열 번호143에 의해 암호화됨)	ORF003-제1군 (β-글루코시다제)
38	서열 번호 2(예를 들어, 서열 번호 1에 의해 암호화됨)	ORF001-제1군 (β-글루코시다제)
39	서열 번호 86(예를 들어, 서열 번호 85에 의해 암호화됨)	ORF008-제1군 (β-글루코시다제)
축약어: CBM-탄수화물 결합 모듈		

[0649] 효소의 성질 및 기질 활성의 규명

[0650]

기가매트릭스™ 스크리닝에서 발견된 39 히트(상기 표 1a~1c 참조)를 우선 셀로헥사오스에 대해 스크리닝하여 셀룰로스 올리고머에 대한 작용 패턴을 측정하였다. 계놈형 클론은 다수의 개방 해독 틀을 가지고 있을 가능성이 있는 전체 DNA 삽입부를 가지는 클론으로서 정의한다. 예를 들어, 상기 표 1a~1c에 있어서, 이러한 계놈형 클론은 2개의 개방 해독 틀(효소 22 및 22a)을 함유하며, 여기서, 상기 개방 해독 틀은 서열 번호 37 및 서열 번호 35에 각각 나타난 서열을 가진다. 이러한 계놈형 클론으로서 다른 것은 3개의 개방 해독 틀(효소 27, 27a 및 27b)을 함유한다. 서브클론은 계놈형 클론으로부터 유래하며, 이는 1개의 개방 해독 틀만을 함유할 수 있다. 계놈형 클론을 항생제를 함유하는 TB 배지 중에서 밤새도록 생육시키고, 세포를 용해한 다음, 그 용해물을 원심 분리에 의해 정제하였다. 서브 클론을 OD600이 0.5가 될 때까지 생육시킨 후(이 경우 적당한 유도제로 유도시킴), 이 세포를 용해하기 전에 3시간 더 생육시킨 다음에, 용해물을 정제하였다. 계놈형 클론은 일반적으로 서브 클론보다 활성이 낮을 것이지만, 클론이 다수 존재할 경우에는 활성을 측정하기가 더 용이하다. 박막 크로마토그래피(TLC)를 이용하여 종말점에서의 반응에 대한 기초 연구를 수행하였다(일반적으로 24시간 동안 진행). 또한, 효소를 인산 팽창 셀룰로스(PASC)에 대해 시험하였는데, 이 PASC는 결정질의 셀룰로스로서, 산 처리

에 의해 팽창되면 더욱 비결정질의 것이 된다.

[0651] 환경 라이브러리로부터 클로닝된 다수의 셀룰라제는 PASC에 대해 활성이었지만, 셀룰로스 뿐만 아니라, 셀로트리오스 및/또는 글루코스도 방출하였다. 기가메트릭스™에 의한 관찰 결과 찾아낸 계놈형 클론도 PASC와 셀룰로스 기질 예를 들어, 셀로헥사오스에 대해 시험하였다(Seikagaku, Japan). 박층 크로마토그래피(TLC) 실험 결과, 몇몇 계놈형 클론들은 셀로헥사오스를 가수분해할 수 있었음을 알 수 있었다(도 6 및 7 참조). 이들 클론 중에서, 다수는 최종 생산물로서 글루코스를 생성할 수 있었는데, 이는 이들 클론이 글리코실 가수분해 효소 제1군(베타-글루코시다제를 포함함)과 서열 상동성을 갖는다는 사실과 부합되는 것이다. 일부 효소들은 셀로바이오스 및/또는 더욱 큰 단편을 생성하였지만, 생산 패턴의 정확한 성질은 TLC 실험만으로는 알 수 없었기 때문에, 모세관 전기 영동법(CE)을 실행하였다.

[0652] 실시예 2: 모세관 전기 영동법

[0653] 몇몇 측면에서, 모세관 전기 영동법(CE)은 효소 활성에 대해 스크리닝하기 위한 검정법에서 사용되는데, 예를 들어, 상기 CE는 폴리펩티드가 셀룰라제 활성을 보유하는지 여부와 본 발명의 범위 내에 포함되는지 여부를 확인하거나, 또는 셀룰라제 활성을 보유하는 폴리펩티드를 확인 및 분리하기 위한 방법에 사용된다. 모세관 전기 영동법(CE)은 실행 시간이 적게 들고 검정 감도도 뛰어나다는 이점을 제공한다. 상기 CE법은 1-아미노피렌-3,6,8-트리설포네이트(APTS)를 형광단으로 사용하였으며, 당과 당의 올리고머와 함께 사용하기에 적합하였다 [Guttman (1996) High-resolution capillary gel electrophoresis of reducing oligosaccharides labeled with 1-aminopyrene-3,6,8-trisulfonate. Anal. Biochem 233:234-242]. 셀로헥사오스에 활성인 것으로 나타나는 효소를 인산 팽창 셀룰로스 및 셀로헥사오스에 대해 시험하였다. 유전자를 서브클로닝한 후 발현시키고, 니켈-킬레이트화 컬럼을 사용하여 이를 부분적으로 정제하였다. 효소를 기질과 함께 1 시간 동안 항온 처리한 후, 이로부터 얻어진 생성물을 10cm 또는 48cm의 모세관을 사용하여 분석하였다. 10cm 및 48cm의 모세관을 통하여 2 분 및 9분 경과시 셀로헥사오스를 용리시켰다. 48cm 모세관에서는 생성물이 더욱 잘 분리되었는데, 이는 곧, 여기에 소량의 당이 존재하고 혼합물 중에 오염물이 존재함을 시사하는 것이다. CE법은 TLC 검출법을 이용하였을 때 셀로헥사오스에 대한 활성이 뛰어났다는 기가메트릭스™ 관찰 결과를 바탕으로 하여 효소에 대한 연구를 수행하였다.

[0654] 효소 22/22a(상기 표 1a~1c 참조)는 주로 셀로바이오스를 방출하는 PASC(데이터는 도 8에 그래프의 형태로 요약하여 나타냄)상에서 우수한 성능을 나타내었다. 뿐만 아니라, 효소 22/22a는 아비셀(AVICEL)® 미세 결정질 셀룰로스(MCC)(FMC Corporation, Philadelphia, PA)(데이터는 도 9에 그래프 형태로서 요약하여 나타냄)로부터 셀로바이오스를 방출할 수 있었다. 서열 분석 결과, 효소 22 및 효소 21(글리코실 가수분해 효소 제5군에 속함)는 약 92% 상동성임을 알 수 있었다. 제5군은 주로 엔도글루카나제를 포함하지만, 셀로바이오하이드롤라제도 약간 존재한다. 클로스트리움 서모셀룸(*Clostridium thermocellum*)로부터 유래하는 Cel10는, 비결정질 및 결정질 셀룰로스로부터 셀로바이오스만을 방출할 때의 활성을 바탕으로 하였을 때 셀로바이오하이드롤라제인 것으로 규명되었다[Zverlov (2002) A newly described cellulosomal cellobiohydrolase, Cel10, from *Clostridium thermocellum*: investigation of the exo-mode of hydrolysis, and binding capacity to crystalline cellulose. Microbiology 148:247-255].

[0655] 이러한 3가지 효소 전부는 액시도서머스 셀룰로라이티쿠스(*Acidothermus cellulolyticus*)로부터 유래하는 엔도글루카나제의 경우에 비해 기질 결합 부위에 보다 인접하여 존재하는 삽입부를 가진다. 이 삽입부는 기질 결합 부위를 감싸주어 이 효소를 엔도글루카나제에서 셀로바이오하이드롤라제로 전환시키는 루프를 형성할 수 있었다. 이 효소를 셀로헥사오스에 대해 시험하였을 때, 이 효소는 주로 소량의 셀로트리오스를 포함하는 셀로바이오스를 생산하였다. 이러한 결과들은 셀로바이오하이드롤라제가 셀로헥사오스 기질로부터 셀로바이오스와 셀로트리오스 모두를 생산하는 능력을 가진다는 사실에 의해 설명된다[Harjunpaa (1996) Cello-oligosaccharide hydrolysis by cellobiohydrolase II from *Trichoderma reesei*. Association and rate constants derived from an analysis of progress curves. Eur. J Biochem 240:584-591].

[0656] 실시예 3: 서열을 바탕으로 하는 관찰

[0657] 본 발명은 본 발명의 서열을 이용하여 셀룰라제 예를 들어, 셀로바이오하이드롤라제를 동정 및 분리하는 방법을 제공한다. 하나의 대표적인 방법에 있어서, 셀로바이오하이드롤라제를 함유하는 3개의 글리코실 가수분해 효소군의 보존적 부위에 상동성인 프라이머는 진균 시료로부터 유래하는 폴리뉴클레오티드 라이브러리 또는 DNA를 스크리닝하는데 사용되었다. 프라이머는 48개의 보존 부위 군에 대해 디자인되었으며, 96개의 라이브러리는 스크리닝되어 1개의 히트를 확정하였다. 뿐만 아니라, 프라이머는 제6군과 제7군에 대해 디자인되었다. 진균의

라이브러리는 상기 프라이머를 사용하여 스크리닝되었으며, 그 결과 제6군에 대해서는 1개의 히트를, 그리고 제7군에 대해서는 56개의 히트를 얻었다. 제7군 히트 중 하나는 전체 길이 서열을 추출하기 위한 연구를 위하여 선택되었다. 전체 길이 서열은 성공적으로 얻어졌으며, 페니실리움 잔티넬룸(*Penicillium janthinellum*)의 엑소-셀로바이오하이드롤라제 I과 73% 상동성이었다.

[0658] 실시예 4: 셀로바이오하이드롤라제 활성을 보유하는 효소의 유전자 조작

[0659] 본 실시예는 본 발명의 대표적인 효소를 유전자 조작하는 방법에 관하여 기술하고 있다. 상기 효소는 바이오매스를 연료 및 화학 물질로 전환시키는 경우와, 석유계 제품의 효율적이고 지속 가능한 대체물을 제조하는 경우에 사용될 수 있다. 상기 효소는 천연 바이오매스 전환을 포함하는 화학적 순환 반응에 참여하기 때문에, 유기체(예를 들어, 미생물 예를 들어, 박테리아) 내에서 발현될 수 있다. 하나의 측면에서, 상기 효소는 셀룰로소계 및 헤미셀룰로소계 중합체를, 대사 가능한 탄소 부분으로 효율적으로 분해하기 위한 "효소 집합물(enzyme ensemble)"에 사용된다. 전술한 바와 같이, 본 발명은 이와 같이 중요한 신규의 "바이오매스 전환" 및 대체 에너지 산업을 수행할 수 있도록 하기 위해서, 효소를 가장 효율적으로 발견하고 사용하는 방법을 제공한다.

[0660] 돌연변이 유발에 따른 관찰 결과와 확률에 의하지 않는 유도 진화 방법["다이렉트 에볼루션(DIRECTEVOLUTION)@이라 칭함, 예를 들어, 유전자 위치 포화 돌연변이 유발법(GSSM)(전술한 바와 같이, 미국 특허 제6,171,820호 및 동 제6,579,258호 참조) 및 조정 가능한 유전자 재조립법(Tunable GeneReassembly; TGR)(예를 들어, 미국 특허 제6,537,776호 참조)을 포함하여, 미국 특허 제6,939,689호에 개시됨]과 같은 방법을 사용한다. 그 결과는 셀룰로스를 글루코스로 환원시키는 중요한 효소 성분 즉, 셀로바이오하이드롤라제의 발견과 이의 최적화에 중점을 두고 있다.

[0661] 효소 발견 스크리닝은 다이버사 코퍼레이션의 기가매트릭스™ 고 처리량 발현 스크리닝 플랫폼(전술함)을 사용하여 수행되며, 그 결과, 메틸움베리페릴 셀로바이오시드를 기질로서 이용하는 셀로바이오하이드롤라제이 동정되었다. 총 100개의 복합 환경 라이브러리를 스크리닝한 결과, 이들 라이브러리 중 25개는 주로 글리코실 가수분해 효소 제5군과 제10군으로부터 셀로바이오하이드롤라제 히트를 얻었음을 알 수 있었다. 이러한 히트는 아비셀® 미세 결정질 셀룰로스(MCC)(FMC Corporation, Philadelphia, PA)에 대한 활성에 대하여 특성이 규명되었다. 상기 MCC 성능 분석 결과를 바탕으로 하여, 서열 번호 162의 효소(예를 들어, 서열 번호 161에 의해 암호화됨)를 유전자 위치 포화 돌연변이 유발법(GSSM) 기술을 사용하여 최적화하기 위한 후보 효소로서 선택하였다. 그러나, GSSM 진화를 수행하기 이전에, 신호 서열(1~30번 아미노산)을 서열 번호 162로부터 제거하고, 개시 메티오닌을 부가하였다. 이와 같은 무 신호 서열(이하, "야생형"이라 칭하며, 서열 번호 164로서 나타냄)(예를 들어, 서열 번호 163에 의해 암호화됨)은 GSSM 기술을 사용하여 최적화된 무 서열이었다. 전술한 바와 같이, GSSM 기술은 단백질 내 모든 아미노산을 19개의 기타 아미노산으로 연속 돌연변이시킬 수 있다. 점유계 검정법을 사용하여 돌연변이체를 스크리닝하였으며, 하나의 아미노산이 변이된 잠재 과변이체를 동정하였다. 과변이체를 조합하여 발현하는 신규 라이브러리에 상기 과변이체들을 조합하여 넣었다. 이 라이브러리를 스크리닝한 결과, 시판용 후보 효소 중 몇몇을 동정할 수 있었다.

[0662] 연구 결과 요약

[0663] 기가매트릭스™ 스크리닝

[0664] 기가매트릭스™(GMx) 스크리닝 플랫폼은 통상의 96 웰 평판과 치수가 동일한 100,000 웰 미세 평판을 바탕으로 하는 고 처리량 방법에 사용되는 것이다(상세한 설명은 제II 단계 출원 참조). 상기 스크리닝법은 형광 기질을 사용한다. 상기 GMx 스크리닝은 셀룰라제에 의해 이미 확인된 활성을 바탕으로 하여 2개의 기질을 사용하여 수행되었다. 메틸움베리페릴 셀로바이오시드(MUC)는 스크리닝 기질로서 사용하였다. 뿐만 아니라, 두 개의 기질 모두에 대한 활성을 보유하며 베타-글루코시다제인 것으로 추정되는 클론을 제거하기 위해서, 레조루핀-베타-글루코피라노시드도 스크리닝에 포함되었다.

[0665] 표적 라이브러리의 증폭된 파지 또는 파지미드를 스크리닝하였다. 베타-갈락토시다제 유전자가 결여된 2개의 숙주 균주(CEH6 및 GAL631)은 기질에 대한 내인성 숙주 활성을 감소시키기 위하여 사용하였다. 이러한 라이브러리가 종전의 스크리닝 프로그램으로부터 셀룰라제 히트를 발생시킨다는 사실을 바탕으로 하여 스크리닝하기 위해 100개의 라이브러리를 선택하였다. 스크리닝된 라이브러리 중에서, 69개의 라이브러리로부터 총 355개의 1차 히트가 얻어졌다.

[0666] 2차 스크리닝법은 아가 평판 상에 클론을 도말한 후, 콜로니를 찍어서, 배지 및 메틸움베리페릴 셀로바이오시드(MUC)("브레이크아웃(breakout)")를 함유하는 384 웰 평판에 접종하였다. 도 10은 통상의 기가매트릭스™(GMx)

브레이크아웃을 나타내는 데이터를 그래프로 나타낸 것이다. 이 데이터를 얻기 위해서, MUC에 활성인 클론(즉, 메틸옴베리페릴 셀로바이오시드를 가수분해할 수 있는 클론)을 비활성 클론의 백그라운드로부터 분화하였다. 각각의 클론을 발색도록 생육시킨 다음, 형광도를 측정하고, 가장 활성인 히트를 골라내어 서열 결정하였다. 도 10에서, X축은 시료명용; Y축은 상대적 형광도 단위를 나타낸다. 포지티브 "히트"를 아가 평판 상에 도말한 후, 콜로니를 찍어내어 LB + 항생체 + 50 μ M의 MUC를 함유하는 384 웰 평판에 접종시킨 후, 이를 발색도록 생육시켰다.

[0667] [표 2a]

[0668]

기가매트릭스™(GMx) 히트에 관한 요약		
효소 번호	개방 해독틀 서열 번호	클론 군의 특성
40	서열 번호104(예를 들어, 서열 번호103에 의해 암호화됨)	제5군(셀룰라제)
41	서열 번호108(예를 들어, 서열 번호107에 의해 암호화됨)	제5군(셀룰라제)
42	서열 번호112(예를 들어, 서열 번호111에 의해 암호화됨)	제5군(셀룰라제)
H7	서열 번호 60(예를 들어, 서열 번호 59에 의해 암호화됨)	제5군(셀룰라제)
43	서열 번호 82(예를 들어, 서열 번호 81에 의해 암호화됨)	제5군(셀룰라제)
44	서열 번호 78(예를 들어, 서열 번호 77에 의해 암호화됨)	제5군(셀룰라제)
45	서열 번호 68(예를 들어, 서열 번호 67에 의해 암호화됨)	제5군(셀룰라제)-ORF2

[0669] [표 2b]

[0670]

45a	서열 번호 70(예를 들어, 서열 번호 69에 의해 암호화됨)	제26군(만난아제)-ORF4
46	서열 번호 74(예를 들어, 서열 번호 73에 의해 암호화됨)	제10군(자일라나제)
47	서열 번호110(예를 들어, 서열 번호109에 의해 암호화됨)	제10군(자일라나제)
48	서열 번호106(예를 들어, 서열 번호105에 의해 암호화됨)	제5군(셀룰라제)
49	서열 번호 66(예를 들어, 서열 번호 65에 의해 암호화됨)	제10군(자일라나제)
50	서열 번호 72(예를 들어, 서열 번호 71에 의해 암호화됨)	제5군(셀룰라제)
51	서열 번호 80(예를 들어, 서열 번호 79에 의해 암호화됨)	제5군(셀룰라제)
H8	서열 번호 62(예를 들어, 서열 번호 61에 의해 암호화됨)	제5군(셀룰라제)-ORF1
H8a	서열 번호 64(예를 들어, 서열 번호 63에 의해 암호화됨)	제5군(셀룰라제)-ORF4
52	서열 번호 76(예를 들어, 서열 번호 75에 의해 암호화됨)	제5군(셀룰라제)
53	서열 번호160(예를 들어, 서열 번호159에 의해 암호화됨)	제5군(셀룰라제)
54	서열 번호 88(예를 들어, 서열 번호 87에 의해 암호화됨)	제5군(셀룰라제)
55	서열 번호148(예를 들어, 서열 번호147에 의해 암호화됨)	제10군(자일라나제)
56	서열 번호 90(예를 들어, 서열 번호 89에 의해 암호화됨)	제5군(셀룰라제)
57	서열 번호152(예를 들어, 서열 번호151에 의해 암호화됨)	제5군(셀룰라제)
58	서열 번호150(예를 들어, 서열 번호149에 의해 암호화됨)	제5군(셀룰라제)
59	서열 번호154(예를 들어, 서열 번호153에 의해 암호화됨)	제5군(셀룰라제)
H6	서열 번호158(예를 들어, 서열 번호157에 의해 암호화됨)	제5군(셀룰라제)
60	서열 번호156(예를 들어, 서열 번호155에 의해 암호화됨)	제5군(셀룰라제)

[0671] 히트로부터 유래하는 모든 계놈형 클론 삽입부를 서열 결정하였다. 상기 표 1a~1c에 나타난 바와 같이, 일부 계놈형 클론들은 하나 이상의 개방 해독 틀을 포함하였다. 예를 들어, 하나의 계놈형 클론은 2개의 개방 해독 틀(효소 H8및 H8a)을 포함하며, 이 개방 해독 틀은 각각 서열 번호 67 및 서열 번호 69에 나타난 서열들을 가진다. 스크리닝된 17개의 라이브러리로부터는 총 25개의 글리코실 가수분해 효소 히트가 발생하였다. 일반적으로, 상기 히트는 제5군 및 제10군을 포함하여 몇몇 상이한 글리코실 가수분해 효소 군으로부터 얻어졌다. 표 2a 및 2b(상기)에는 히트와 이의 정체가 나열되어 있다. 몇몇 기타 히트는 상기 개방 해독 틀이 임의의 공지된 글리코

실 가수분해 효소 군과 상동성이지 않은 경우에 발견되었다. 뿐만 아니라, 히트 중 일부는 GTP 가수분해 효소 유전자가 기질 분해와 상관없이 형광성을 나타내는 시스템에 있어서 잘못된 포지티브 신호인 것으로 공지된 GTP 가수분해 효소 유전자를 암호화하였다. 전반적으로, 이러한 스크리닝법은 MUC에 대해 활성이었던 효소를 성공적으로 동정하였다.

[0672] 특성 규명

[0673] 기가메트릭스™ 스크리닝으로부터 확인된 유전자를 서열 결정한 후, 데이터를 분석하였다. 소프트웨어 시스템을 사용하여 개방 해독 틀(ORF)에 명칭을 달았다. 상기 ORF를 적당한 벡터에 서브 클로닝하였으며, 이때 C-말단 His-태그를 암호화하는 DNA도 함께 도입하였다. DNA 구조물을 적당한 이. 콜라이 숙주로 형질전환시킨 다음, 특성 규명을 위한 연구를 수행하기 위하여 발현시켰다. 유전자 생성물을 인산 팽창 셀룰로스(PASC)에 대해 시험하였다. 이 PASC는 결정질의 셀룰로스로서, 산 처리에 의해 팽창되면 더욱 비결정질의 것이 된다. PASC를 아비셀® 미세 결정질 셀룰로스(MCC)로부터 제조하였다. 서브 클론을 생육 후, 발현시킨 다음, 용해하였다. 용해물을 PASC와 함께 항온 처리하고, 이 반응 생성물을 비신코닌산(bicinchoninic acid; BCA) 환원당 검정법을 통하여 분석하였다. 가장 활성인 서브 클론을 보다 큰 규모의 성장 및 정제를 위해 선택하였다. 이러한 서브 클론의 비활성을 PASC에 대해 측정하였다.

[0674] 상기 서브 클론은 또한 모세관 전기 영동법(CE)에 의해 분석하였다. 용해물을 기질과 함께 30 시간 동안 항온 처리하였다. 반응 생성물을 형광단 1-아미노피렌-3,6,8-트리설포네이트(APTS)로 유도체화하였다. 48cm 모세관을 사용하여 생성물을 분석하였다. 셀로바이오스는 6분 경과시에 용리시켰다. 도 11은 모세관 전기 영동(CE) 분석법에 의해 선택된 효소의 PASC에 대한 활성을 나타내는 데이터를 그래프로 도시한 것이다. 시료 H9~H1은 개별 클론이다. 도 11에 있어서, 다수 시료의 반응 생성물 프로파일은 가공 효소를 나타내는 것이었다. 가공 효소의 셀로바이오스/(글루코스 + 셀로트리오스) 비율은 10인 것으로 정의된다. 활성이 가장 큰 2개의 잠재적 가공 효소의 PASC에 대한 비활성은 각각 0.35 및 0.04 U/mg이었다.

[0675] 피치아 내 진균 CBH

[0676] 새로이 발견된 제6군 및 제7군 진균 셀로바이오하이드롤라제 유전자를 피. 파스토리스에 형질전환시키고, 이 형질전환체를 고체 아가 평판 상에 도말하였다. 각각의 구조물에 대해 160개의 콜로니를 선별하였다. 시료를 생육 및 유도시키고, 상청액을 β-갈락토시다제의 존재하에서 PASC와 함께 항온 처리하였다. 글루코스-산화 효소 검정법을 사용하여 반응 생성물을 분석하였다. 글리코실 가수분해 효소 제6군 셀로바이오하이드롤라제는 피. 파스토리스 내에서 성공적으로 이중 발현되었다.

[0677] 엑소-엔도 작용성 셀룰라제

[0678] 야생형 효소 즉, 효소 스크리닝을 통하여 구해 낸 제9군 글리코실 가수분해 효소는 서모모노스포라 푸스카(*Thermomonospora fusca*) E4의 상동체이다. E4는 엔도-활성 및 엑소-활성 모두 가지는 것으로 확인되었다. 야생형 효소로 처음 시험을 한 결과, 이 효소는 PASC 및 아비셀® 미세 결정질 셀룰로스(MCC) 둘 다에 대해 활성인 것으로 확인되었다. 반응 생성물을 HPLC 분석한 결과, 1차 생성물은 글루코스와 셀로바이오스임을 알 수 있었다. 야생형 효소는 글리코실 가수분해 효소 제9군 촉매 도메인, 제3군 셀룰로스 결합 도메인 및 3개의 박테리아 Ig-유사 도메인(세포 부착에 관여하는 것으로 추측됨)을 포함하는 다중 도메인 단백질이다. 야생형 효소의 부가 서브 클론 변이체 3개의 도메인이 활성에 미치는 효과에 대해 시험하였다. 야생형 효소는 다음의 것들과 함께 서브 클로닝되었다: 1) 촉매 도메인 단독(CD); 2) 촉매 및 탄수화물 도메인(CCD); 및 3) 촉매 및 탄수화물 결합 도메인 + 11개의 하류 아미노산(CCD + 11). 전체 길이 단백질 및 3개의 서브 클론 변이체를 아비셀® 미세 결정질 셀룰로스(MCC) 상에서 검정한 후, 반응 생성물을 BCA 환원 당 검정법으로 분석하였으며, 데이터는 도 12에 그래프의 형태로 요약하였다. 도 12에 나타난 데이터는 아비셀® 미세 결정질 셀룰로스(MCC)와 함께 74 시간 동안 37°C 및 pH5에서, 야생형 효소 및 절단 돌연변이체의 BCA에 의해 얻어진 것이다. CBH1은 포지티브 대조군이다. 네가티브 대조군은 삼입부를 포함하지 않는 숙주이다.

[0679] 야생형 효소 즉, 전체 길이 단백질[서열 번호 164, 예를 들어, 서열 번호 163에 의해 암호화됨]이 가장 활성이었다. 전체 길이 단백질을 GSSM 진화용으로서 선택하였다. 촉매 도메인 및 탄수화물 결합 도메인을 진화시켰다.

[0680] GSSM 스크리닝

[0681] GSSM 기술(전술함)을 사용하여 표적 단백질의 촉매 도메인 및 탄수화물 결합 도메인의 아미노산을 기타 19개의 아미노산으로 신속하고 연속적으로 돌연변이시켰다. GSSM의 목적은 불용성 미세 결정질 셀룰로스 상에서 가수분해되는 정도가 증가한 돌연변이체를 동정하는데에 있다. 로봇에 의한 스크리닝 방법은 이러한 GSSM 스크리닝 방

법을 촉진하기 위해 개발된 것이다.

[0682] 돌연변이 구조물로부터 유래하는 DNA를 DH10b 숙주 세포로 형질전환시켰다. 각각의 콜로니들을 찍어내어 (얇은) 96 웰 평판[150uL LB/암피실린 함유]에 접종하였다[자동 콜로니 채집 시스템 사용]. 이 평판을 37℃에서 24 시간 동안 항온 처리하였다(400 rpm). 배양액 15uL을 유도용 평판의 각 웰로 옮겼다. 이 유도용 평판의 각 웰에는 135uL LB/암피실린과 1.1mM IPTG가 담겨져 있었다. 유도용 평판을 37℃ 및 400rpm에서 24 시간 동안 항온 처리하였다. 이 평판을 원심 분리한 후, 상청액을 버렸다.

[0683] 이 시점에서 본 검정법의 자동화 단계를 개시하였다. 세포를 용해시켜 로봇에 의해 재현탁하였다. 150uL의 용해 완충액(125uL 물 + 25uL BPER + 0.2mg/mL 리소자임 + 20 유닛/mL DNaseI)을 각 웰에 첨가하였다. 15uL의 용해 물을 반응 평판의 각 웰로 옮겼다. 반응 평판의 각 웰은 반응 혼합물 185uL를 함유하였다[1% 아비셀® 미세 결정 질 셀룰로스(MCC), 50mM 아세트산나트륨 완충액, pH5.0]. 반응 평판을 37℃ 및 습도 95%에서 30시간 동안 항온 처리하였다. 항온 처리 후, 평판을 원심 분리하여 15 uL의 상청액을 BCA 평판으로 옮겼다. 이 BCA 평판은 웰당 50uL의 시약 A, 50uL의 시약 B, 그리고 80uL의 400 mM 탄산염 완충액(pH10)을 함유하였다. 이 평판을 고무 밀봉재로 씌우고, 이를 80℃에서 30분 동안 항온 처리한 후에, 원심 분리하여 식히고, A₅₆₀에서의 흡광도를 측정하였다.

[0684] 결과

[0685] 80개 이상의 랜덤한 돌연변이 콜로니들을 각각의 아미노산에 대해 스크리닝하였다. 1차 GSSM™ 스크리닝 데이터의 일례를 도 13에 그래프로서 나타내었다. 컬럼 6은 야생형 시료를 함유하였고, 컬럼 12는 숙주/벡터 네가티브 대조군을 함유하였다. 아비셀® 미세 결정 질 셀룰로스(MCC)와 함께 30시간 동안 항온 처리한 후, 야생형 시료로부터 발생한 신호는 약 0.53이었다(표준 편차 = 0.07). 네가티브 대조군의 평균 신호는 0.29였다. 포지티브 대조군의 평균값 + (2×표준 편차) 보다 큰 신호를 발생시키는 시료는 1차 히트를 발생시키는 것으로 간주되었다. 이러한 스크리닝 평판으로부터 2차 확인 스크리닝에 대해 약 10개의 1차 히트가 선택되었다.

[0686] 1차 히트를 2차 검정법을 통해 다시 확인하였다. 이 검정법은 1차 스크리닝법과 동일하였다. 그러나, 시료는 4쌍 준비해 검정법을 수행하였다. 2차 GSSM 스크리닝 데이터의 일례를 도 14에 그래프로서 나타내었다. 웰 E3-H3, A4-D4, A7-D7 중 시료는 평균적으로 그 활성이 야생형의 경우보다 더 높았다. 시료를 4쌍 준비해 검정법을 수행한 결과, 12개의 웰로부터 3개의 히트를 얻었다. 이러한 시료들은 도 13에서의 E4, G2 및 H3 웰에 1차 히트를 보였다(평판 29805-AA89 BCA 평판).

[0687] 2차 스크리닝 결과 77개의 히트가 발생하였다. 이 시료들을 서열 결정하였다. 이 시료 중 35개는 아미노산에 변이가 일어났으며, 22개는 트랜스포손이 삽입되었고, 또한 나머지는 야생형이거나 결실이 일어났다.

[0688] 2차 스크리닝으로부터 얻은 히트를 추가로 분석하였다. GSSM 과변이체를 티.푸스카의 결정 구조상에 맵핑하였다. 아미노산 위치, 아미노산 변이 및 몇 배 개선되었는지를 나타내는 스코어를 바탕으로 하여 시료에 우선 순위를 매겼다. 8개의 과변이체를 GSSM 스크리닝으로부터 선택하고, 전술한 바와 같이, 유전자 재조합 진화 즉, 조정 가능한 유전자 재조합법(TGR)을 위해 선택하였으며, 이에 관하여는 예를 들어, 미국 특허 제 6,537,776호를 참조하시오.

[0689] [표 2c]

위치 배향 돌연 변이 유발 재조합법용으로 선택된 과변이체

관 기	구 아미노산	신규 아미노산
89	M	R
103	F	G

[0690]

[0691] [표 2d]

110	P	G
114	Y	L
157	A	S
481	W	F
550	P	N
590	G	R

[0692]

[0693] 과변이체의 배합

[0694] 유전자 재조합 기술(조정 가능 유전자 재조합법(TGR))을 이용하여, 상기 표 2c~2d에 나타난 과변이체를 배합하여, 최상의 활성을 나타내는 후보 변이체를 동정하였다. 활성 검정법은 그 반응물을 야생형 효소에 비해 과변이체의 활성이 증가하도록 더욱 희석시켰다는 점을 제외하고는, GSSM 스크리닝법과 동일하게 행하여 졌다. 도 15는 혼합된 또는 "배합된" GSSM™ 스크리닝 검정법으로부터 얻은 데이터를 그래프의 형태로 나타내는 것이다.

[0695] 요약하면, 본 발명은 서열 번호 164(예를 들어, 서열 번호 163에 의해 암호화됨)를 바탕으로 하는 다음과 같은 서열을 가지며, 셀룰라제 활성을 나타내는 효소를 제공한다.

간 기	기원 아미노산	기원 아미노산 암호화 코돈	신규 아미노산 (GSSM 전 수행 후)	신규 아미노산 암호화 코돈
89	M	ATG	R	CGT, CGC, CGA, CGG, AGA, AGG
103	F	TTT, TTC	G	GGT, GGC, GGA, GGG
110	P	CCA, CCC, CCG, CCT	G	GGT, GGC, GGA, GGG
114	Y	TAT, TAC	L	TTA, TTG, CTT, CTC, CTA, CTG
157	A	GCT, GCC, GCA, GCG	S	TCT, TCC, TCA, TCG, AGT, AGC
481	W	TGG	F	TTT, TTC
550	P	CCA, CCC, CCG, CCT	N	AAT, AAC
590	G	GGT, GGC, GGA, GGG	R	CGT, CGC, CGA, CGG, AGA, AGG

[0696]

[0697] 본 발명의 다수의 측면에 관하여 기술하였다. 하지만, 본 발명의 사상과 범위를 벗어나지 않고서 다수의 변형이 가하여질 수 있음을 이해하게 될 것이다. 그러므로, 다른 측면은 다음과 같은 청구의 범위 내에 존재하는 것이다.

도면의 간단한 설명

[0092] 이하 도면은 본 발명의 측면들을 예시적으로 나타내고 있는 것으로서, 청구항에 의해 포함되어 있는 본 발명의 범위를 제한하는 의미는 아니다.

[0093] 도 1은 컴퓨터 시스템의 블록 선도이다.

[0094] 도 2는 신규의 뉴클레오티드 또는 단백질 서열과 서열 데이터베이스를 비교하여 신규 서열과 데이터베이스 중 서열 사이의 상동성 수준을 측정하는 공정에 관한 한 측면을 나타내는 공정도이다.

[0095] 도 3은 2개의 서열들이 상동성인지 여부를 결정하기 위하여 컴퓨터 내에서 실행되는 공정의 한 측면을 나타내는 공정도이다.

[0096] 도 4는 서열 내에 특징이 존재하는지 여부를 관찰하기 위한 식별자(identifier)의 한 측면을 나타내는 공정도이다.

[0097] 도 5는 셀로바이오스의 구조를 도시하는 것이다.

[0098] 도 6 및 도 7은 이하 실시예 1에서 상세히 논의되고 있는 바와 같이, 셀로헥사오스로부터 얻은 반응 생성물을 TLC 분석한 결과를 나타내는 것이다.

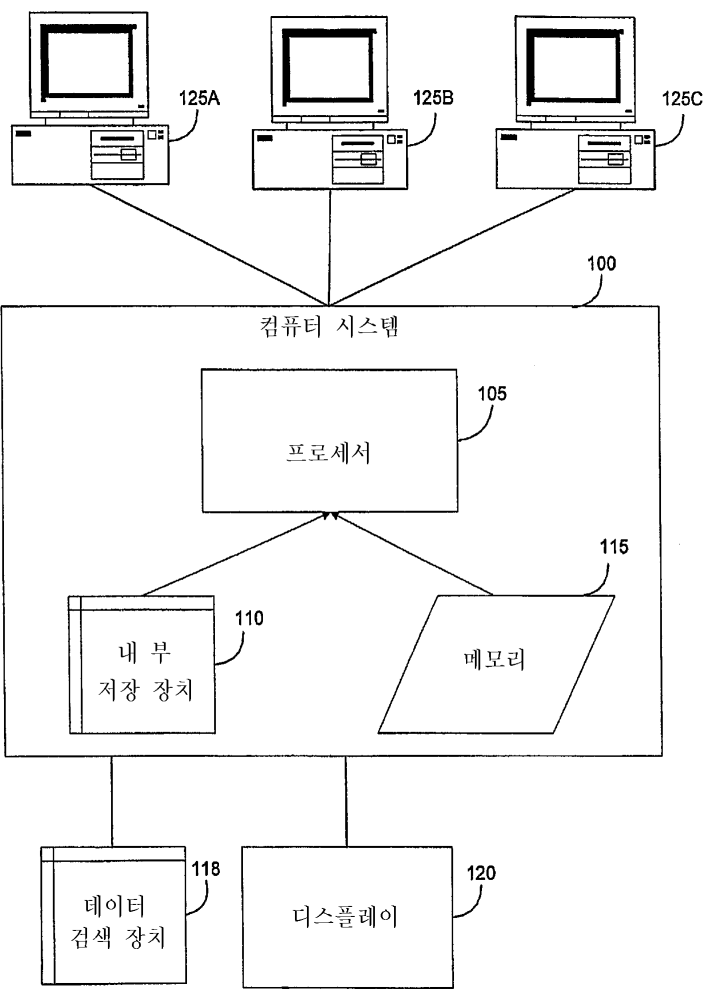
[0099] 도 8은 이하 실시예 2에서 상세히 논의되고 있는 바와 같이, 본 발명의 대표적인 효소인 22/22a(CBH)에 의해

PASC로부터 셀로바이오스가 방출된다는 사실을 보여주는 데이터를 그래프 형태로 나타내는 것이다.

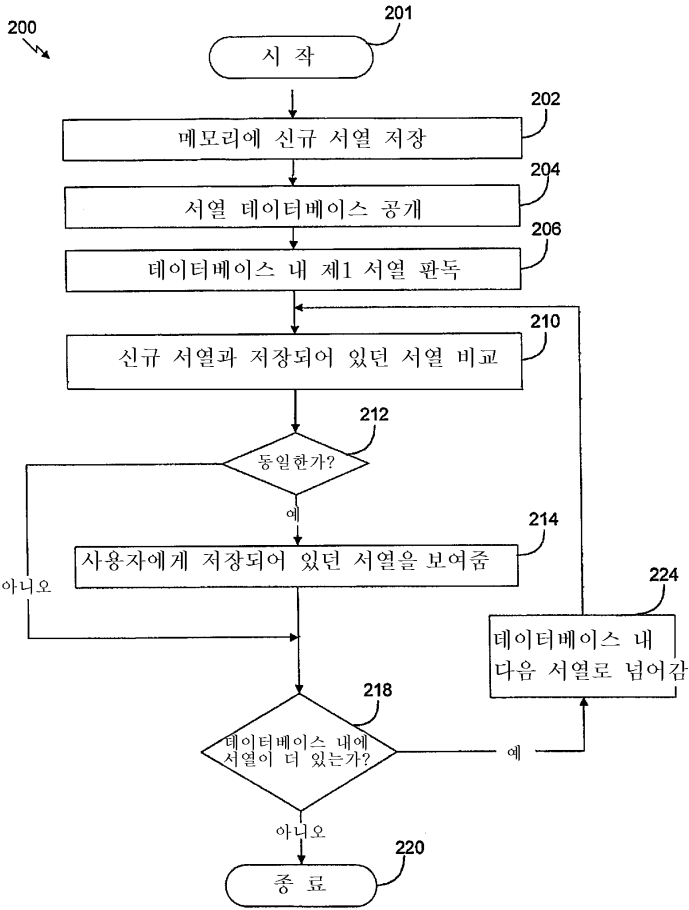
- [0100] 도 9는 이하 실시예 2에서 상세히 논의되고 있는 바와 같은, 본 발명의 대표적인 효소인 22/22a(CBH)에 의해 아비셀(AVICEL)® MCC로부터 셀로바이오스가 방출된다는 사실을 보여주는 데이터를 그래프 형태로 나타내는 것이다.
- [0101] 도 10은 이하 실시예 4에서 상세히 논의되고 있는 바와 같이, 메틸룸베리페틸 셀로바이오시드를 가수분해할 수 있는 효소를 발현하는 활성 클론이 동정될 경우, 통상의 기가매트릭스(GIGAMATRIX)™ 브레이크아웃(breakout)을 보여주는 데이터를 그래프 형태로 나타내는 것이다.
- [0102] 도 11은 이하 실시예 4에서 상세히 논의되고 있는 바와 같이, 모세관 전기 영동(CE) 분석법에 의한 인산-팽창 셀룰로스(PASC)에 대하여 선별된 효소의 활성을 보여주는 데이터를 그래프 형태로 나타내는 것이다.
- [0103] 도 12는 이하 실시예 4에 상세히 논의되고 있는 바와 같이, BCA 환원당 검정법에 의해 반응 생성물이 분석되는 경우, 아비셀(AVICEL)® 미세 결정질 셀룰로스(MCC) 내 서브클론 변이체 및 본 발명의 대표적인 효소를 분석한 결과로부터 얻어진 데이터를 그래프 형태로 나타내는 것이다.
- [0104] 도 13은 이하 실시예 4에 상세히 논의되고 있는 바와 같이, 1차 GSSM 스크리닝 검정법으로부터 얻어진 그래프 형태의 데이터를 나타내는 것이다.
- [0105] 도 14는 이하 실시예 4에 상세히 논의되고 있는 바와 같이, 2차 GSSM 스크리닝 검정법으로부터 얻어진 데이터를 그래프 형태로 나타내는 것이다.
- [0106] 도 15는 이하 실시예 4에 상세히 논의되고 있는 바와 같이, 혼합된 즉, "배합된(blended)" GSSM 스크리닝 검정법으로부터 얻어진 데이터를 그래프 형태로 나타내는 것이다.
- [0107] 상기와 같은 도면에 나타낸 참조 부호들은 동일한 요소들을 나타낸다.

도면

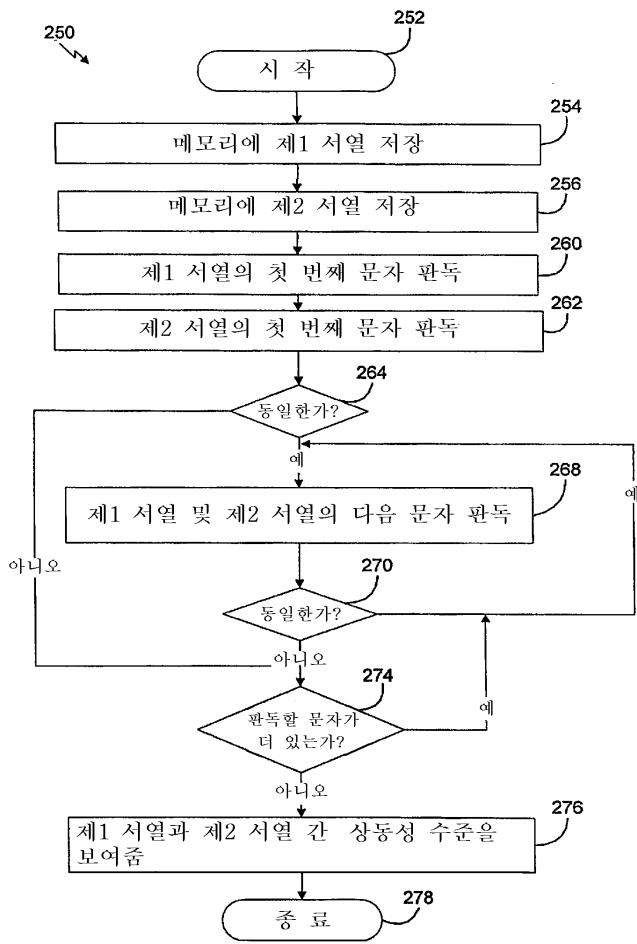
도면1



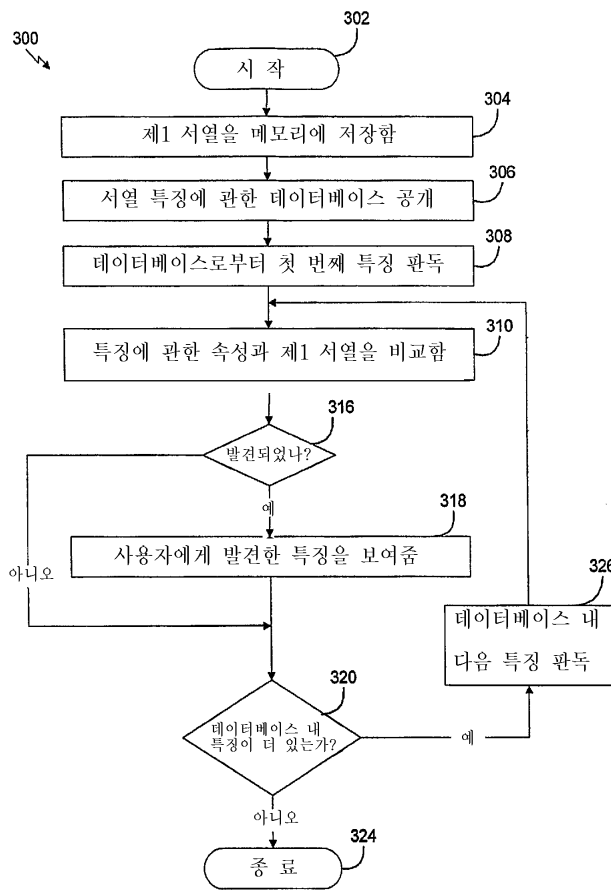
도면2



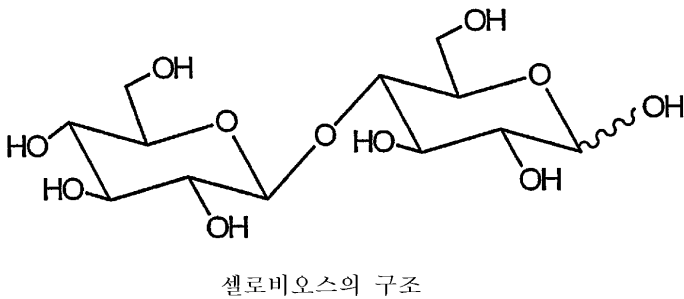
도면3



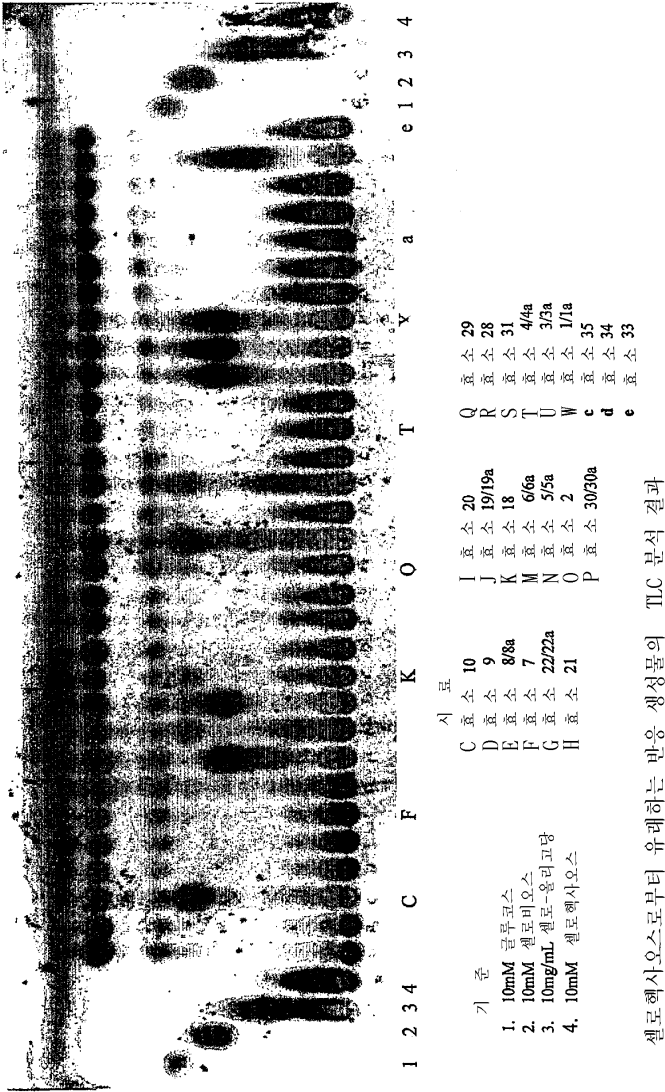
도면4



도면5



도면6



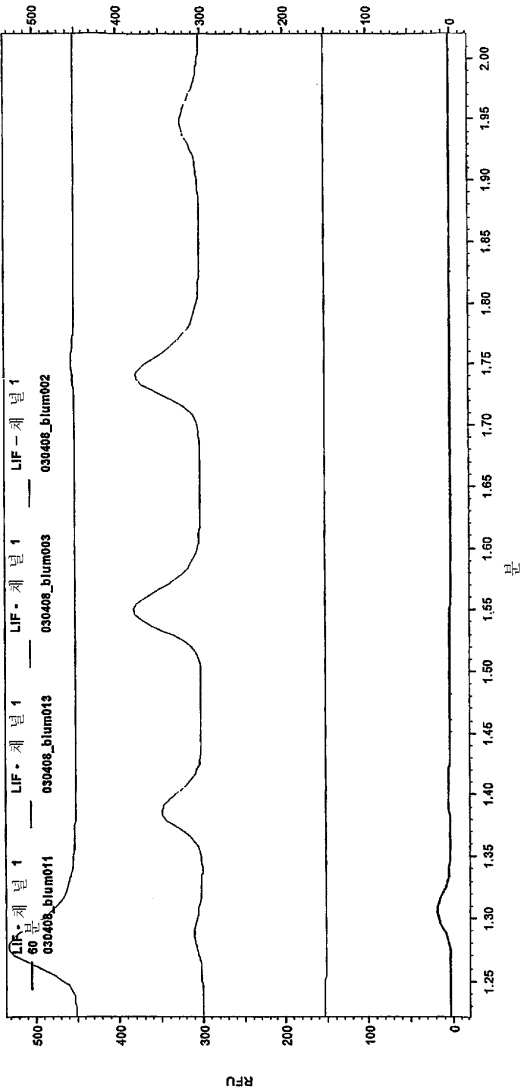
도면7



기 준		시 료 -BD #	
1.	10mM 클루코스	A	호 소 32
2.	10mM 셀로비오스	B	호 소 17
3.	10mg/mL 셀로-올리고당	C	호 소 16
4.	10mM 셀로헥사오스	D	호 소 15/15a
		E	호 소 26/26a
		F	호 소 25
		G	호 소 27/27a/27b
		H	호 소 27/27a/27b
		I	호 소 14/14a
		J	호 소 13/13a
		K	호 소 12
		L	호 소 11
		O	호 소 11
		P	네가티브 대조군

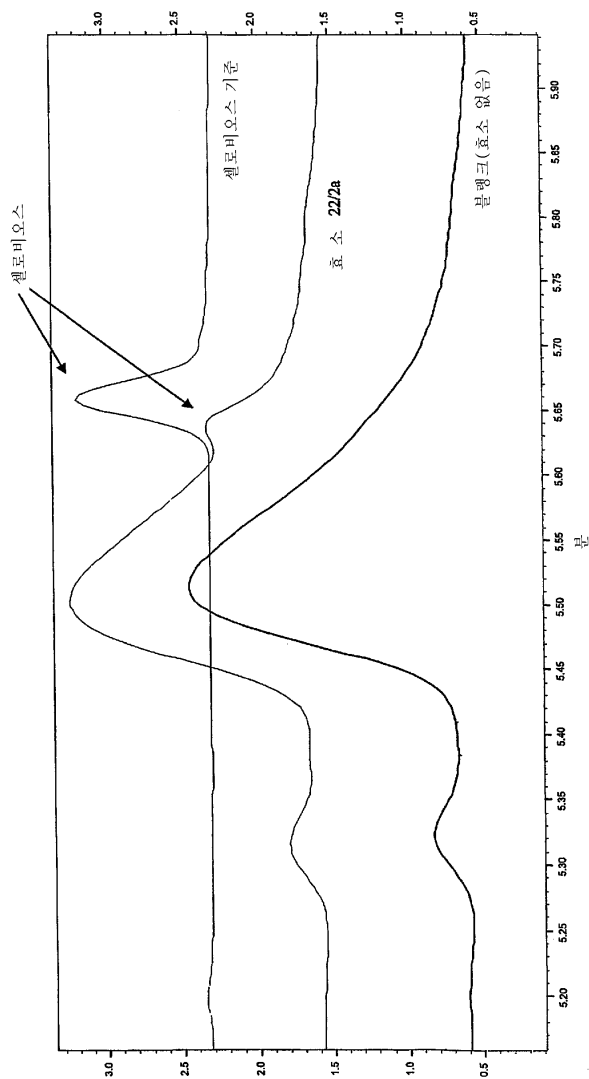
셀로헥사오스로부터 유래하는 반응 생성물의 TLC 분석 결과

도면8



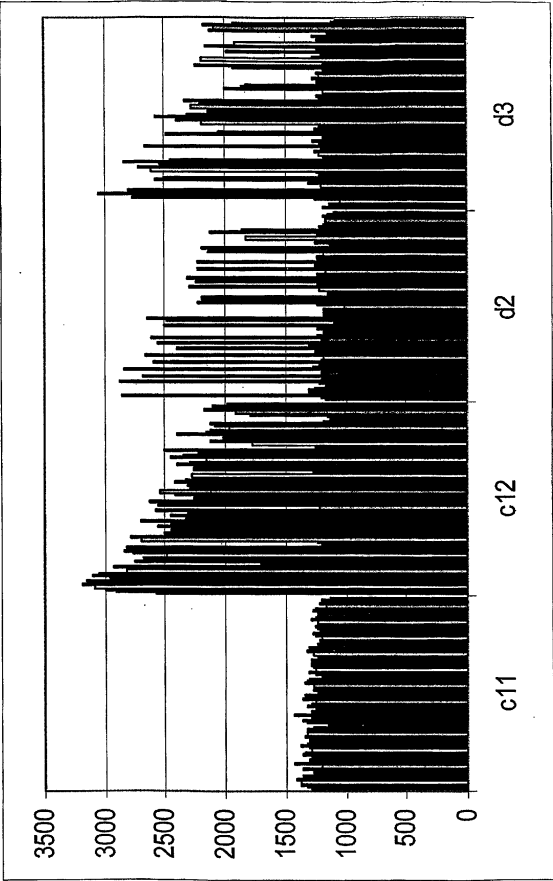
효소 22/22a에 의해 PASC로부터 셀로비오스가 방출되었을 때의 결과

도면9

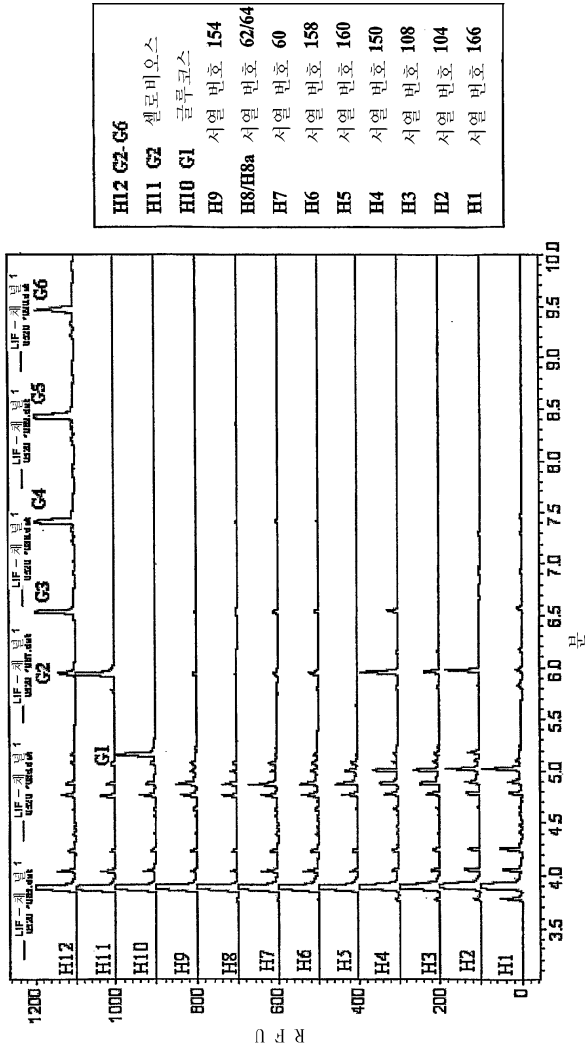


CBH에 의하여 아비셀 MCC로부터 셀로비오스가 방출되었을 때의 결과

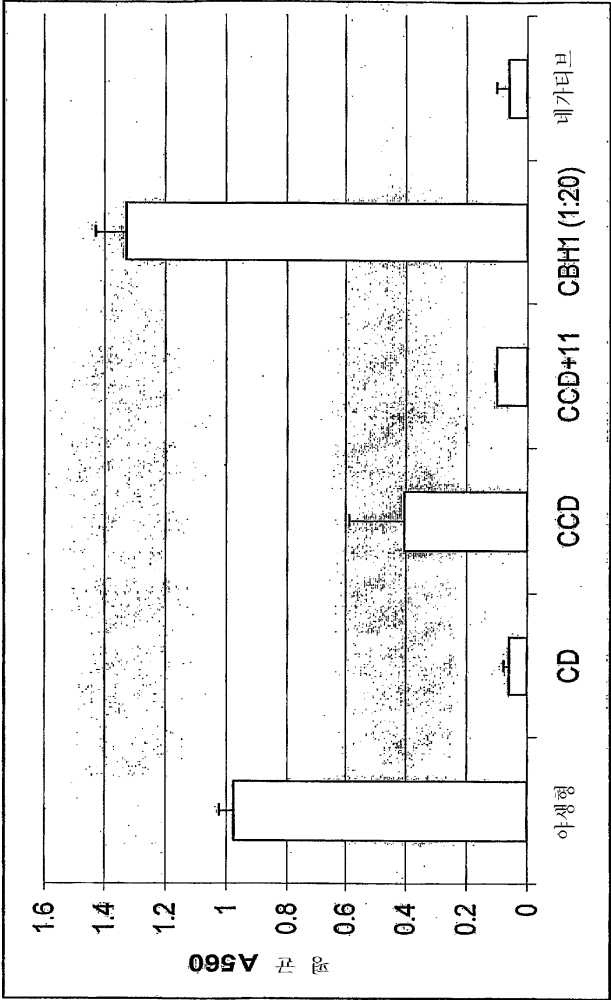
도면10



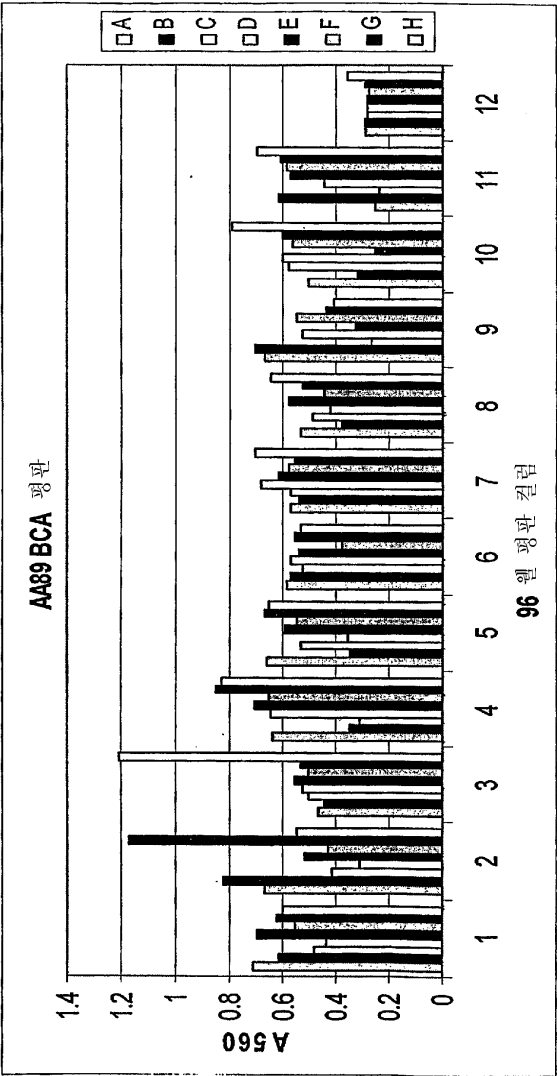
도면11



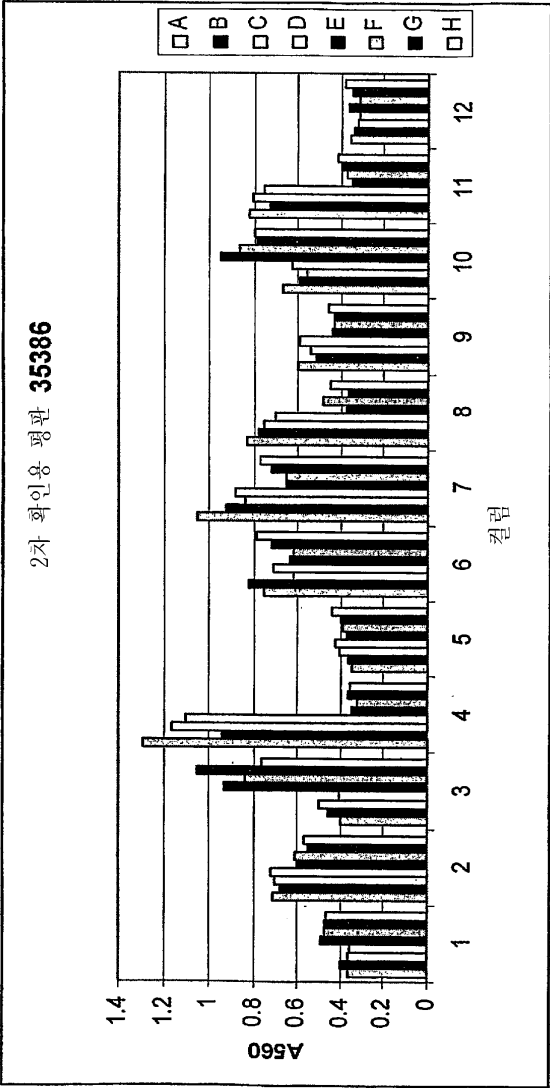
도면12



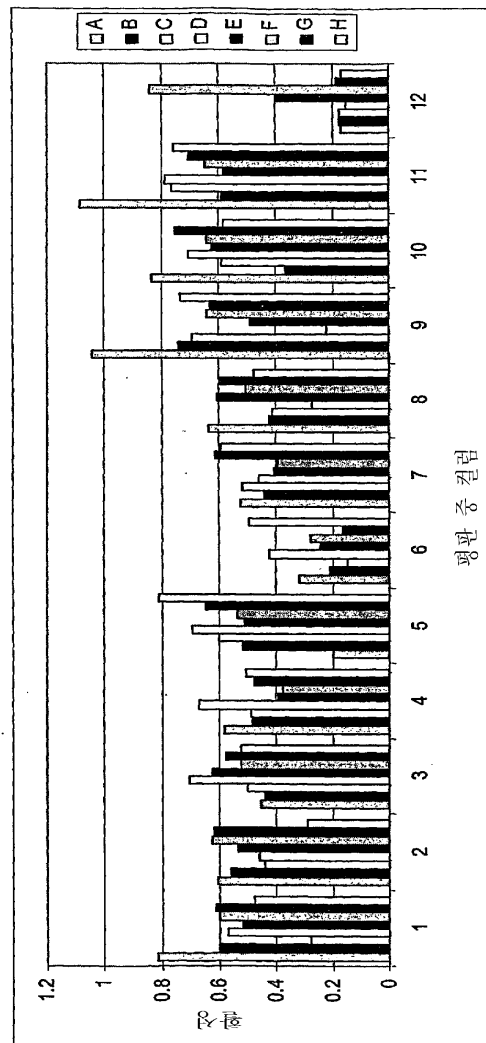
도면13



도면14



도면15



서열 목록

SEQUENCE LISTING

<110> DIVERSA CORPORATION

BLUM, David

GEMSCH, Joslin

DYCAICO, Mark

<120> CELLULASES, NUCLEIC ACIDS ENCODING THEM
AND METHODS FOR MAKING AND USING THEM

<130> 564462014240

<140> Not Yet Assigned

<141> Filed Herewith

<150> US 60/662,224

<151> 2005-03-15

<160> 166

<170> PatentIn version 3.1

<210> 1

<211> 1323

<212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 1

atgtcaacct ataaatttcc gcacaacttt ttttggggag ccgcaaccgc gtcttatcag	60
atcgaaggcg catggaacga ggatggcaaa ggcgaaatcca tttgggatcg cttcagccat	120
acgcccgga aggtcaccaa tgccgatacc ggtgacatcg cctgtgacca ctatcacctg	180
tgggaggaag atatgcacct tatgcgcaa cttgggttga aggcgtaccg cttttccact	240
tcatggcccc gtgtgatccc ggcgggccgc agacgggtga atgtcaaagg gctggatttc	300
tacgatcgcc tggatgatgg tctgtgcgcc gcgaacatcg aaccgttcct caccctgtat	360
cactgggacc tgccgcaggc tcttcaagac gaaggcggct gggataatcg caacaccgcc	420
catgcctttg ccgattatgc cgattgatg gtgaaacgac ttggcgaccg tatccgctat	480
tggacgacgt tcaacgaacc cagcgttgtg gcgttcaatg gtcattactc aggcctgcac	540
gccccgggca ttcaagatgc ccgtgttacc cgccaggtgg tgcacattt gctggtggcg	600
catgggttgg ctgtgcaggc gatccgcggc gcaaaactcca aagtggatgt gggcatcgtg	660

cttaatttat ggcccgccga acccgattcg gactccccg aagatgccgc cgccgccgaa 720

gccgcctgga accggcacga gacctgttc cttgaccca tctttaaggc gcattatccc 780

gtatctgccc ttgatgcgat tggggaggat atccccgca tccacgacgg cgatctggcg 840

ttgatctctc aggaattgga tttgtcggc atcaactatt actcccgcca tgtgtcagt 900

gccacaaaag aaataggcag gcttcccgaa tcggaatata ctgaaatggg ctgggaagta 960

tgcgccccg cactccgccc cctgtgtgtc aagatccata acgattaccg tttgccgcc 1020

atctatatca ccgaaaacgg atcggcattc aaggacgaag ttaacgcaga cggaaaggtt 1080

catgaccgcg ggcggttga ttacctgaaa caacacctga ttcaactttg ctttgccatg 1140

caggacggcg tggatgtgcg cggctacatg gcttgggtccc tgctggataa tttcgagtgg 1200

ggtcacggct tttccaagcg ctttggcttg gtccatgtgg attacgagag ccagaagcgg 1260

attattaaag actcgggtga atggtatgca agtgtgatac ggaagaacga ggttgttgaa 1320

taa 1323

<210> 2
 <211> 440
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> DOMAIN
 <222> (2)...(438)
 <223> Glycosyl hydrolase family 1

<220>
 <221> SITE
 <222> (10)...(24)
 <223> Glycosyl hydrolases family 1 N-terminal signature. Prosite id = PS00653

<220>

<221> SITE

<222> (351)...(354)

<223> N-glycosylation site. Prosite id = PS00001

<400> 2

Met Ser Thr Tyr Lys Phe Pro His Asn Phe Phe Trp Gly Ala Ala Thr
1 5 10 15

Ala Ser Tyr Gln Ile Glu Gly Ala Trp Asn Glu Asp Gly Lys Gly Glu
20 25 30

Ser Ile Trp Asp Arg Phe Ser His Thr Pro Gly Lys Val Thr Asn Ala
35 40 45

Asp Thr Gly Asp Ile Ala Cys Asp His Tyr His Arg Trp Glu Glu Asp
50 55 60

Ile Ala Leu Met Arg Gln Leu Gly Leu Lys Ala Tyr Arg Phe Ser Thr
65 70 75 80

Ser Trp Pro Arg Val Ile Pro Ala Gly Arg Arg Arg Val Asn Val Lys
85 90 95

Gly Leu Asp Phe Tyr Asp Arg Leu Val Asp Gly Leu Cys Ala Ala Asn
100 105 110

Ile Glu Pro Phe Leu Thr Leu Tyr His Trp Asp Leu Pro Gln Ala Leu
115 120 125

Gln Asp Glu Gly Gly Trp Asp Asn Arg Asn Thr Ala His Ala Phe Ala
130 135 140

Asp Tyr Ala Ala Leu Met Val Lys Arg Leu Gly Asp Arg Ile Arg Tyr
145 150 155 160

Trp Thr Thr Phe Asn Glu Pro Ser Val Val Ala Phe Asn Gly His Tyr

165

170

175

Ser Gly Ser His Ala Pro Gly Ile Gln Asp Ala Arg Val Thr Arg Gln
180 185 190

Val Val His His Leu Leu Val Ala His Gly Leu Ala Val Gln Ala Ile
195 200 205

Arg Gly Ala Asn Ser Lys Val Asp Val Gly Ile Val Leu Asn Leu Trp
210 215 220

Pro Ala Glu Pro Asp Ser Asp Ser Pro Glu Asp Ala Ala Ala Glu
225 230 235 240

Ala Ala Trp Asn Arg His Glu Thr Leu Phe Leu Asp Pro Ile Phe Lys
245 250 255

Ala His Tyr Pro Val Ser Ala Leu Asp Ala Ile Gly Glu Asp Met Pro
260 265 270

Arg Ile His Asp Gly Asp Leu Ala Leu Ile Ser Gln Glu Leu Asp Phe
275 280 285

Val Gly Ile Asn Tyr Tyr Ser Arg His Val Val Ser Ala Thr Lys Glu
290 295 300

Ile Gly Arg Leu Pro Glu Ser Glu Tyr Thr Glu Met Gly Trp Glu Val
305 310 315 320

Cys Ala Pro Ala Leu Arg Arg Leu Leu Val Lys Ile His Asn Asp Tyr
325 330 335

Arg Leu Pro Pro Ile Tyr Ile Thr Glu Asn Gly Ser Ala Phe Lys Asp
340 345 350

Glu Val Asn Ala Asp Gly Lys Val His Asp Pro Arg Arg Leu Asp Tyr

355 360 365

Leu Lys Gln His Leu Ile Gln Leu Cys Leu Ala Met Gln Asp Gly Val
370 375 380

Asp Val Arg Gly Tyr Met Ala Trp Ser Leu Leu Asp Asn Phe Glu Trp
385 390 395 400

Gly His Gly Phe Ser Lys Arg Phe Gly Leu Val His Val Asp Tyr Glu
405 410 415

Ser Gln Lys Arg Ile Ile Lys Asp Ser Gly Glu Trp Tyr Ala Ser Val
420 425 430

Ile Arg Lys Asn Glu Val Val Glu
435 440

<210> 3
<211> 1389
<212> DNA
<213> Unknown

<220>
<223> Obtained from environmental sample

<400> 3
atgagcgtc cgagtccgc ccgccccgtg tcctttctc cccgcttcgt gtggggagcc 60

gcggccgc at cctatcaaat cgaggcgcc gtccgggagg acggcaagg cccttcggtg 120

tgggacatgt tcigcgagaa gccgggagcc gtcttcgagg ggcacgacgg ggcggtggct 180

tgcgatcact accaccgtta ccggaagac gtggccctga tgcggcagat tgggctccag 240

gcttaccgcc tgagcgtgtg ctggcccagg gtgctgccc aggggaccgg gcagcccaac 300

gagaaggggc tcgacttcta ctcccggctc gtcgacgct tgctcgagg ggggatcacg 360

ccttgggtca cctttttca ctgggactac ccactagccc tataccaccg gggaggctgg 420

ctcaatcggg atagctcaga ctggttcggc gactacgagg gtctgattgc ggagcgcctc 480

tccgatcggg tgagccactt cttcaccag aacgagcccc aggtgtacat cggcttcggg 540

cacctcgagg ggaaacacgc gccgggcgat acccttcccc tgtcgcagat gctgctggcc 600

ggtcaccaca gcctgctgc ccatggaaag gccgtgcagg cgctgcgcgc ccacggcaag 660

cagcagctgc gggttggata cgctccggtg gggatgccgc tgcacccgt cagcagctcc 720

gccgaagacg tggcggctgc acgcaccgcc actttccgcg tccgagagaa gaattcctgg 780

aacaacgctt ggtggatgga cccggtgtac ctggtgagt accccgcca agggctcgag 840

ttctacgggc gagacgtccc cgcgatccgg tccggagaca tggaactcat ccggcaaccc 900

ttggactttt tcggcgtcaa catctaccag agcaccccg tcgcgcgcgc gggggcgccc 960

cagggttcg aggtcgtccg gcatccgacg ggccaccca tcaccgctt caactggcgc 1020

gttacccac aggccttgta ttgggggcgc cggttcttct acgagcgcta tggcaagccc 1080

atcgtcatta cggaaaacgg gctttcctgc cgagacgtga tcgcccttga cggcaaggtg 1140

cacgatccgt cccgcatcga cttcaccacg cgctacctgc gcgagctcca ccgcgccatc 1200

gccgaaggca acgaggtgga gggctacttc cactggtcca tcatggaaa cttcgaatgg 1260

gctgccggat accgagaacg cttcgggctc gttcacgtgg attacgagac cctggtgagg 1320

acaccaagg actctgcggc gtggtaccgc caggtcatcc agagcaacgg ggccgtgctg 1380

ttcgattga 1389

<210> 4

<211> 462

<212> PRT

<213> Unknown

<220>

<223> Obtained from environmental sample

<220>

<221> DOMAIN

<222> (8)...(458)

<223> Glycosyl hydrolase family 1

<220>

<221> SITE

<222> (16)...(30)

<223> Glycosyl hydrolases family 1 N-terminal signature. Prosite id = PS00653

<220>

<221> SITE

<222> (366)...(374)

<223> Glycosyl hydrolases family 1 active site. Prosite id = PS00572

<400> 4

Met	Ser	Ala	Pro	Ser	Pro	Ala	Arg	Pro	Val	Ser	Phe	Pro	Pro	Arg	Phe
1				5					10					15	

Val	Trp	Gly	Ala	Ala	Ala	Ala	Ser	Tyr	Gln	Ile	Glu	Gly	Ala	Val	Arg
			20					25					30		

Glu	Asp	Gly	Lys	Gly	Pro	Ser	Val	Trp	Asp	Met	Phe	Cys	Glu	Lys	Pro
		35					40					45			

Gly	Ala	Val	Phe	Glu	Gly	His	Asp	Gly	Ala	Val	Ala	Cys	Asp	His	Tyr
	50					55					60				

His	Arg	Tyr	Arg	Glu	Asp	Val	Ala	Leu	Met	Arg	Gln	Ile	Gly	Leu	Gln
65						70				75				80	

Ala	Tyr	Arg	Leu	Ser	Val	Cys	Trp	Pro	Arg	Val	Leu	Pro	Glu	Gly	Thr
				85					90					95	

Gly	Gln	Pro	Asn	Glu	Lys	Gly	Leu	Asp	Phe	Tyr	Ser	Arg	Leu	Val	Asp
			100					105					110		

Ala Leu Leu Glu Ala Gly Ile Thr Pro Trp Val Thr Leu Phe His Trp
115 120 125

Asp Tyr Pro Leu Ala Leu Tyr His Arg Gly Gly Trp Leu Asn Arg Asp
130 135 140

Ser Ser Asp Trp Phe Gly Glu Tyr Ala Gly Leu Ile Ala Glu Arg Leu
145 150 155 160

Ser Asp Arg Val Ser His Phe Phe Thr Gln Asn Glu Pro Gln Val Tyr
165 170 175

Ile Gly Phe Gly His Leu Glu Gly Lys His Ala Pro Gly Asp Thr Leu
180 185 190

Pro Leu Ser Gln Met Leu Leu Ala Gly His His Ser Leu Leu Ala His
195 200 205

Gly Lys Ala Val Gln Ala Leu Arg Ala His Gly Lys Gln Gln Leu Arg
210 215 220

Val Gly Tyr Ala Pro Val Gly Met Pro Leu His Pro Val Ser Glu Ser
225 230 235 240

Ala Glu Asp Val Ala Ala Ala Arg Thr Ala Thr Phe Arg Val Arg Glu
245 250 255

Lys Asn Ser Trp Asn Asn Ala Trp Trp Met Asp Pro Val Tyr Leu Gly
260 265 270

Glu Tyr Pro Ala Gln Gly Leu Glu Phe Tyr Gly Arg Asp Val Pro Ala
275 280 285

Ile Arg Ser Gly Asp Met Glu Leu Ile Arg Gln Pro Leu Asp Phe Phe
290 295 300

Gly Val Asn Ile Tyr Gln Ser Thr Pro Val Arg Ala Ala Gly Ala Pro
305 310 315 320

Gln Gly Phe Glu Val Val Arg His Pro Thr Gly His Pro Ile Thr Ala
325 330 335

Phe Asn Trp Pro Val Thr Pro Gln Ala Leu Tyr Trp Gly Pro Arg Phe
340 345 350

Phe Tyr Glu Arg Tyr Gly Lys Pro Ile Val Ile Thr Glu Asn Gly Leu
355 360 365

Ser Cys Arg Asp Val Ile Ala Leu Asp Gly Lys Val His Asp Pro Ser
370 375 380

Arg Ile Asp Phe Thr Thr Arg Tyr Leu Arg Glu Leu His Arg Ala Ile
385 390 395 400

Ala Glu Gly Asn Glu Val Glu Gly Tyr Phe His Trp Ser Ile Met Asp
405 410 415

Asn Phe Glu Trp Ala Ala Gly Tyr Arg Glu Arg Phe Gly Leu Val His
420 425 430

Val Asp Tyr Glu Thr Leu Val Arg Thr Pro Lys Asp Ser Ala Ala Trp
435 440 445

Tyr Arg Gln Val Ile Gln Ser Asn Gly Ala Val Leu Phe Asp
450 455 460

<210> 5

<211> 1098

<212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 5

atgactcgga ggtctatcgt gcgtttcttct tccaacaagt ggcttgtcct tgccggtgcg	60
gcgctgctcg cctgcaccgc cctcgggtgc aagaaaaag gcgagagcgg tgacgtcgcc	120
tcggccccgg ggcaggccca ggcgggcggc aagcagccgt ttcccacga tgcgccgac	180
accgaaccgc ccgtccgcc ccctcgtagc ggcaatcctc tggtagggcg caagctcttc	240
gtcgacccgg aatctttggc catgttgtag gcgaacaagc tgcggcgac cgacccggag	300
aaggcggcga ttttgtagc catgccccag cagccccagg ctttgtgat gggcgagtgg	360
aacacgaaca tcttcgcgc ggtcgagcat ttcgtggctc gcgccaagc ggagggcgcc	420
gtgccgtca tgcgccta caacatccc caccgcgact gcgggcagta ctctcagggt	480
gggctttcct ccaaggagc ttaccagcgc tggattcgga acgtcgccgc ggggattggc	540
agcgtatcag cggctgctgt gctcgagccc gacgcgctc gccacttcca ggagtgtttg	600
accgaggagc agagcgccga gcgcatgttc ctgctcagc acgccgtcaa ggtgctgcgc	660
caaatccga agacggccgt gtacctgat gccgggcac gcgctgggt gccggtggag	720
gagatggcgc agcgctcaa gctcggggc atcgagcac ccatggctt ttcgtcaac	780
acctgaact acgtgggcac cgaggagaac gccgcttac gccacaagc cgtcgaggcc	840
ctgggtggga acgtgcgctt cgtcatcgac acgagccga atggggcggg cccctacgag	900
gaggccaaga acgccgagga gagctggtgc aaccgcccc gtcgcaagat cggcaagccg	960
ccgaccaccg agacggggga tccctcacc gacggattcc tttggctgaa gcgccgggc	1020
gagtcggacg gtcagtcaa cggcgggccc aaggccggtg tgttctggt ggagcaggt	1080

ctccagcagg cccagtaa

1098

<210> 6
<211> 365
<212> PRT
<213> Unknown

<220>
<223> Obtained from environmental sample

<220>
<221> SIGNAL
<222> (1)...(29)

<220>
<221> DOMAIN
<222> (81)...(358)
<223> Glycosyl hydrolases family 6

<220>
<221> SITE
<222> (187)...(196)
<223> Glycosyl hydrolases family 6 signature 2. Prosite id = PS00656

<220>
<221> SITE
<222> (263)...(266)
<223> N-glycosylation site. Prosite id = PS00001

<400> 6
Met Thr Arg Arg Ser Ile Val Arg Ser Ser Ser Asn Lys Trp Leu Val
1 5 10 15

Leu Ala Gly Ala Ala Leu Leu Ala Cys Thr Ala Leu Gly Cys Lys Lys
20 25 30

Lys Gly Glu Ser Gly Asp Val Ala Ser Ala Pro Gly Gln Ala Gln Ala
35 40 45

Gly Gly Lys Gln Pro Phe Pro Asp Asp Ala Pro Ile Thr Glu Pro Pro
50 55 60

Ala Pro Pro Pro Arg Ser Gly Asn Pro Leu Val Gly Ala Lys Leu Phe
65 70 75 80

Val Asp Pro Glu Ser Leu Ala Met Leu Gln Ala Asn Lys Leu Arg Arg
85 90 95

Thr Asp Pro Glu Lys Ala Ala Ile Leu Asp Arg Ile Ala Gln Gln Pro
100 105 110

Gln Ala Leu Trp Met Gly Glu Trp Asn Thr Asn Ile Phe Arg Ala Val
115 120 125

Glu His Phe Val Ala Arg Ala Lys Ala Glu Gly Ala Val Pro Val Met
130 135 140

Ile Ala Tyr Asn Ile Pro His Arg Asp Cys Gly Gln Tyr Ser Gln Gly
145 150 155 160

Gly Leu Ser Ser Lys Glu Ala Tyr Gln Arg Trp Ile Arg Asn Val Ala
165 170 175

Ala Gly Ile Gly Ser Asp Ala Ala Val Val Val Leu Glu Pro Asp Ala
180 185 190

Leu Gly His Phe Gln Glu Cys Leu Thr Glu Glu Gln Ser Ala Glu Arg
195 200 205

Met Phe Leu Leu Ser Asp Ala Val Lys Val Leu Arg Gln Asn Pro Lys
210 215 220

Thr Ala Val Tyr Leu Asp Ala Gly His Ala Arg Trp Val Pro Val Glu
225 230 235 240

Glu Met Ala Glu Arg Leu Lys Leu Ala Gly Ile Glu His Ala His Gly
245 250 255

Phe Ser Leu Asn Thr Ser Asn Tyr Val Gly Thr Glu Glu Asn Ala Ala

260 265 270

Tyr Gly His Lys Leu Val Glu Ala Leu Gly Gly Asn Val Arg Phe Val
275 280 285

Ile Asp Thr Ser Arg Asn Gly Ala Gly Pro Tyr Glu Glu Ala Lys Asn
290 295 300

Ala Glu Glu Ser Trp Cys Asn Pro Pro Gly Arg Lys Ile Gly Lys Pro
305 310 315 320

Pro Thr Thr Glu Thr Gly Asp Pro Leu Ile Asp Gly Phe Leu Trp Leu
325 330 335

Lys Arg Pro Gly Glu Ser Asp Gly Gln Cys Asn Gly Gly Pro Lys Ala
340 345 350

Gly Val Phe Trp Leu Glu Gln Ala Leu Gln Gln Ala Gln
355 360 365

<210> 7

<211> 2649

<212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 7

atgcaaggaa agaaaattga ttccattaac tcaaggttgt tagttcctga ttatccaatc 60

gttccttca ttgaggaga tggtagcggc cctgatattc ggctgtcttc agtcagggtg 120

ctggatgttg ctgttgacag ggcatattcc ggcaagcgaa aacttctctg gaaagaggtg 180

ctggctggcg aaaaggcatt tacaataacc gggtcctggc ttccggagga aactcttaga 240

gcatttcgtg aatatcatgt tggaattaaa ggccactca ctacgccagt tgggtggggga 300

attcgttctc tcaatgtagc cctcaggcaa gagcttgact tgtatgttg cctgaggcca 360

gtcaaatggt ttaagggtgt accaagtcct ctaaaagatc cttccaaagt ggatatgcat 420

atitccgcg aaaacactga agatatitct gcaggtattg aatttatgca tggatgaaccg 480

gaggccctga aagttaagaa atttcttacc gaagaaatgg gaatcaagaa gtttcggttt 540

cccatacat cctccattgg tatcaagcct atctcactcg aaggaaacaga gcgtcttgta 600

agagcttcca ttcaatatgc acttgacagg aagttgcctt ccgtaacatt ggttcataaa 660

ggcaatatca tgaattcac cgagggggca ttcaaaaaat ggggttatga acttgccgaa 720

agagaatttg gcgacagggt ttttcatgg tcaatgtatg accgtatcgc cgatgaacat 780

ggaacggaag aagctggcaa agtgcaatcc gaagcgattg caaaaggtaa actcctgata 840

aaggatgtga ttgtgatgc tttctgcag caatactac tcaggcctgc cgagtacagc 900

gttatcgcaa ccatgaacct gaatggcgat tatatcagcg atgcactggc agctatggtg 960

gggggtatag gaattgctcc cggagccaat attaaccatc aaactggcca tgcagctttt 1020

gaagcaacac acggcacggc tcccaaatat gccaaccttg atcaggtaaa ccctggctca 1080

glaatactaa gtggcgcgct gatgctcgaa tacatgggct ggaacgaagc cgctcagctc 1140

attaccaatg gattggaggc taccattcaa cagaaactgg taacctatga tttccatcgc 1200

ttaatggaag gtgctacaaa gttgaagact tcagaatttg gcgatgctgt gatccggccg 1260

gcacgttccg cctgggcgga cacggctgcc gatgccctct ccgggcggcg gcgtcgtgcg 1320

cggaacggcg ggcttgttc cccgccgcg gcctgtcgcc gggggcgggt acgggactca 1380

gcgcttgccg gcctcctca ggggtgactg cagggcgaag aaggccggt tgcggacgaa 1440

cttctccgtc atgaccgtgg cgctgccctc accctcgaag aagaccggca cccacagta 1500

cttgtcggtg aagccccaga tggatgaagga gttgcagtcg ttcacggcca ggcaggccga 1560

cagtgcctgc tggtagtagt cggcctgctg ccgcagctgc tccttgggtg gcttgccgct 1620

cgccgggagg tccatgcgga cgtcgatctc ggtgatggcg gtcctcagac cgaggtcggc 1680

gaaccctgc aggttctgct gcaggtcgcc cgggaagccg tagcgggtgc tcaggtggcc 1740

ctgggcgccc aatccgtgga gcggcacgcc ctgctccagc atctcctggg cgagctcgta 1800

gtaggcgtcg ctcttggcgt tgatgccctc gacgtttag tcgttgagga acagcttggc 1860

ctcggggtcg gcctcgtggg cccagcggaa ggcgtccgag acgatctccg ggccgagctc 1920

acgtatccag atgttctcgt cggtcgcgag ctggcctgg tcgttgaaga tctcgttggc 1980

cacgtccac tgcctgagct tgccggcgta gcggccgacg accgtgtcga tgtggtcctt 2040

gaggatggcg cgcagttcct ccttgggtgaa gtcgccctcc tccagccatt cggggttctg 2100

gtgtgccac aggagggtgt gcccgccgac ggcctggcgg ttcgctggg cgaactcgac 2160

gatggcgtcg gcctcctcga agcgttactg gtcgcgtcgc ggggtgatga actccactt 2220

catctggttc tcggcggaga ccgagttgaa ctgctggccc aggatcttcc ggtacttctt 2280

gtcgaaggtg aagggttccg ggtagtctctg ttcgaggtgg tggccgccc cggccgccc 2340

ggagcctatg aagaaccctt cgggggcccgc ccagcgcagg cggtcgaact tggcgttga 2400

gtggggcgcg gcctcgtggt cggcggacgg ctggccgtg gccgtcgacg tcaccagcgg 2460

gacggccagc gcggcgccga gagcaaaggt gacgatgcgg acgatctca tcagaggtcc 2520

ctcattcgat cgcggctccg aaagttttcg gaggattacc ggaatgttc agggacctta 2580

aggcgcccgg agccgggtcg tcaacggttt ggcccgccc ggtcgaagct tctcccgacc 2640

aggcgttga

2649

<210> 8

<211> 882

<212> PRT

<213> Unknown

<220>

<223> Obtained from environmental sample

<220>

<221> DOMAIN

<222> (20)...(417)

<223> Isocitrate/isopropylmalate dehydrogenase

<220>

<221> SITE

<222> (310)...(329)

<223> Isocitrate and isopropylmalate dehydrogenases signature. Prosite id = PS00470

<220>

<221> SITE

<222> (868)...(871)

<223> N-glycosylation site. Prosite id = PS00001

<400> 8

Met	Gln	Gly	Lys	Lys	Ile	Asp	Phe	Ile	Asn	Ser	Arg	Leu	Leu	Val	Pro
1				5					10					15	

Asp	Tyr	Pro	Ile	Val	Pro	Phe	Ile	Glu	Gly	Asp	Gly	Thr	Gly	Pro	Asp
			20					25					30		

Ile	Trp	Arg	Ala	Ser	Val	Arg	Val	Leu	Asp	Val	Ala	Val	Asp	Arg	Ala
			35					40					45		

Tyr	Ser	Gly	Lys	Arg	Lys	Leu	Leu	Trp	Lys	Glu	Val	Leu	Ala	Gly	Glu
	50					55					60				

Lys	Ala	Phe	Thr	Asn	Thr	Gly	Ser	Trp	Leu	Pro	Glu	Glu	Thr	Leu	Arg
65						70				75					80

Ala Phe Arg Glu Tyr His Val Gly Ile Lys Gly Pro Leu Thr Thr Pro
85 90 95

Val Gly Gly Gly Ile Arg Ser Leu Asn Val Ala Leu Arg Gln Glu Leu
100 105 110

Asp Leu Tyr Val Cys Leu Arg Pro Val Lys Trp Phe Lys Gly Val Pro
115 120 125

Ser Pro Leu Lys Asp Pro Ser Lys Val Asp Met His Ile Phe Arg Glu
130 135 140

Asn Thr Glu Asp Ile Tyr Ala Gly Ile Glu Phe Met His Gly Glu Pro
145 150 155 160

Glu Ala Leu Lys Val Lys Lys Phe Leu Thr Glu Glu Met Gly Ile Lys
165 170 175

Lys Phe Arg Phe Pro Asp Thr Ser Ser Ile Gly Ile Lys Pro Ile Ser
180 185 190

Leu Glu Gly Thr Glu Arg Leu Val Arg Ala Ser Ile Gln Tyr Ala Leu
195 200 205

Asp Arg Lys Leu Pro Ser Val Thr Leu Val His Lys Gly Asn Ile Met
210 215 220

Lys Phe Thr Glu Gly Ala Phe Lys Lys Trp Gly Tyr Glu Leu Ala Glu
225 230 235 240

Arg Glu Phe Gly Asp Arg Val Phe Thr Trp Ser Met Tyr Asp Arg Ile
245 250 255

Ala Asp Glu His Gly Thr Glu Glu Ala Gly Lys Val Gln Ser Glu Ala
260 265 270

Ile Ala Lys Gly Lys Leu Leu Ile Lys Asp Val Ile Ala Asp Ala Phe

275

280

285

Leu Gln Gln Ile Leu Leu Arg Pro Ala Glu Tyr Ser Val Ile Ala Thr
290 295 300

Met Asn Leu Asn Gly Asp Tyr Ile Ser Asp Ala Leu Ala Ala Met Val
305 310 315 320

Gly Gly Ile Gly Ile Ala Pro Gly Ala Asn Ile Asn His Gln Thr Gly
325 330 335

His Ala Val Phe Glu Ala Thr His Gly Thr Ala Pro Lys Tyr Ala Asn
340 345 350

Leu Asp Gln Val Asn Pro Gly Ser Val Ile Leu Ser Gly Ala Leu Met
355 360 365

Leu Glu Tyr Met Gly Trp Asn Glu Ala Ala Gln Leu Ile Thr Asn Gly
370 375 380

Leu Glu Ala Thr Ile Gln Gln Lys Leu Val Thr Tyr Asp Phe His Arg
385 390 395 400

Leu Met Glu Gly Ala Thr Lys Leu Lys Thr Ser Glu Phe Gly Asp Ala
405 410 415

Val Ile Arg Pro Ala Arg Ser Ala Trp Ala Asp Thr Ala Ala Asp Ala
420 425 430

Leu Ser Gly Arg Arg Arg Arg Ala Arg Asn Gly Gly Leu Val Ala Pro
435 440 445

Pro Ala Ala Cys Arg Arg Gly Arg Val Arg Asp Ser Ala Leu Ala Arg
450 455 460

Leu Leu Gln Gly Gly Leu Gln Gly Glu Glu Gly Arg Leu Ala Asp Glu

465 470 475 480
 Leu Leu Arg His Asp Arg Gly Ala Ala Leu Thr Leu Glu Glu Asp Arg
 485 490 495
 His Pro Arg Val Leu Val Gly Glu Ala Pro Asp Gly Glu Gly Val Ala
 500 505 510
 Val Val His Gly Gln Ala Gly Arg Gln Cys Leu Leu Val Val Val Gly
 515 520 525
 Leu Leu Pro Gln Leu Leu Leu Gly Gly Leu Ala Ala Arg Arg Glu Val
 530 535 540
 His Ala Asp Val Asp Leu Gly Asp Gly Gly Leu Gln Thr Glu Val Gly
 545 550 555 560
 Glu Pro Leu Gln Val Leu Leu Gln Val Ala Arg Glu Ala Val Ala Gly
 565 570 575
 Ala Gln Val Ala Leu Gly Ala Glu Ser Val Glu Arg His Ala Leu Leu
 580 585 590
 Gln His Leu Leu Gly Glu Leu Val Val Gly Val Ala Leu Gly Val Asp
 595 600 605
 Ala Leu Asp Val Val Val Val Glu Glu Gln Leu Gly Leu Gly Val Gly
 610 615 620
 Leu Val Gly Pro Ala Glu Gly Val Arg Asp Asp Leu Arg Ala Glu Leu
 625 630 635 640
 Thr Tyr Pro Asp Val Leu Val Gly Ala Gln Leu Gly Leu Val Val Glu
 645 650 655
 Asp Leu Val Gly His Val Pro Leu Leu Asp Leu Ala Gly Val Ala Ala
 660 665 670

Asp Asp Arg Val Asp Val Val Leu Glu Asp Gly Ala Gln Phe Leu Leu
675 680 685

Gly Glu Val Ala Leu Leu Gln Pro Phe Gly Val Leu Ala Val Pro Gln
690 695 700

Glu Gly Val Pro Ala His Gly Leu Ala Val Pro Leu Gly Glu Leu Asp
705 710 715 720

Asp Gly Val Gly Leu Leu Glu Ala Val Leu Val Ala Leu Gly Val Asp
725 730 735

Glu Leu Pro Leu His Leu Val Leu Gly Gly Asp Arg Val Glu Leu Leu
740 745 750

Ala Gln Asp Leu Pro Val Leu Leu Val Glu Gly Glu Gly Val Arg Val
755 760 765

Val Leu Phe Glu Val Val Ala Ala Ala Gly Arg Arg Gly Ala Tyr Glu
770 775 780

Glu Pro Phe Gly Gly Gly Pro Ala Gln Ala Val Glu Leu Gly Val Gly
785 790 795 800

Val Gly Arg Gly Leu Val Val Gly Gly Arg Leu Gly Arg Gly Arg Arg
805 810 815

Arg His Gln Arg Asp Gly Gln Arg Gly Gly Glu Ser Lys Gly Asp Asp
820 825 830

Ala Asp Gly Ser His Gln Arg Ser Leu Ile Arg Ser Arg Leu Arg Lys
835 840 845

Phe Ser Glu Asp Tyr Arg Asn Val Ser Gly Thr Leu Arg Arg Pro Glu
850 855 860

Pro Gly Arg Gln Arg Phe Gly Pro Ala Arg Ser Lys Leu Leu Pro Thr
865 870 875 880

Arg Arg

<210> 9
<211> 1134
<212> DNA
<213> Unknown

<220>
<223> Obtained from environmental sample

<400> 9
atgagatccg tccgcatcgt cacctttgct ctccgcccg cgctggccgt cccgctggtg 60

acgtcgacgg ccacggccaa gccgtccgcc gaccacgagg ccgcgcccc ctccaacgcc 120

aagttcgacc gcctgcgtg ggccgcccc gaagggttct tcataggctc cgcggcggcc 180

ggcggcggcc accacctga acaggactac ccggaccctc tcaccttga caagaagtac 240

cggaagatcc tgggccagca gttcaactcg gtctccgcc agaaccagat gaagtgggag 300

ttcatccacc ccgagcgga ccagtaccgc ttcgaggagg ccgacgcat cgtcgagttc 360

gcccagcgga accgccaggc cgtgcgcggg cacacctcc tgtggcacag ccagaacccc 420

gaatggctgg aggaggcgga cttaccaag gaggaactgc gcgccatcct caaggaccac 480

atcgacacgg tcgtcggccg ctacgccggc aagatccagc agtgggacgt ggccaacgag 540

atcttcaacg accaggccga gctgcgcacc gacgagaaca tctggatacg tgagctcggc 600

ccggagatcg tcgcggacgc cttccgctgg gccacgagg ccgaccccga ggccaagctg 660

ttcttcaacg actacaacgt cgagggcac aacgccaaga gcgacgccta ctacagctc 720

gcccaggaga tgctggagca gggcgtgccg ctccacggat tcggcgccca gggccacctg 780
 agcaccgcgt acggcttccc gggcgacctg cagcagaacc tgcagcggtt cgccgacctc 840
 ggtcttgaga ccgccatcac cgagatcgac gtccgcatgg acctcccggc gagcggcaag 900
 cccaccaagg agcagctgcg gcagcaggcc gactactacc agcaggcact gtcggcctgc 960
 ctggccgtga acgactgcaa ctcttcacc atctggggct tcaccgaaa gtactcgtgg 1020
 gtgccggtct tcttcgaggg tgagggcagc gccacggtca tgacggagaa gttcgtccgc 1080
 aagccggcct tcttcgcct gcagtcacc ctgaaggagg cgcgcaagcg ctga 1134

<210> 10
 <211> 377
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> SIGNAL
 <222> (1)...(26)

<220>
 <221> DOMAIN
 <222> (44)...(371)
 <223> Glycosyl hydrolase family 10

<400> 10
 Met Arg Ser Val Arg Ile Val Thr Phe Ala Leu Ala Ala Ala Leu Ala
 1 5 10 15

Val Pro Leu Val Thr Ser Thr Ala Thr Ala Lys Pro Ser Ala Asp His
 20 25 30

Glu Ala Ala Pro His Ser Asn Ala Lys Phe Asp Arg Leu Arg Trp Ala

35

40

45

Ala Pro Glu Gly Phe Phe Ile Gly Ser Ala Ala Ala Gly Gly Gly His
50 55 60

His Leu Glu Gln Asp Tyr Pro Asp Pro Phe Thr Phe Asp Lys Lys Tyr
65 70 75 80

Arg Lys Ile Leu Gly Gln Gln Phe Asn Ser Val Ser Ala Glu Asn Gln
85 90 95

Met Lys Trp Glu Phe Ile His Pro Glu Arg Asp Gln Tyr Arg Phe Glu
100 105 110

Glu Ala Asp Ala Ile Val Glu Phe Ala Gln Arg Asn Arg Gln Ala Val
115 120 125

Arg Gly His Thr Leu Leu Trp His Ser Gln Asn Pro Glu Trp Leu Glu
130 135 140

Glu Gly Asp Phe Thr Lys Glu Glu Leu Arg Ala Ile Leu Lys Asp His
145 150 155 160

Ile Asp Thr Val Val Gly Arg Tyr Ala Gly Lys Ile Gln Gln Trp Asp
165 170 175

Val Ala Asn Glu Ile Phe Asn Asp Gln Ala Glu Leu Arg Thr Asp Glu
180 185 190

Asn Ile Trp Ile Arg Glu Leu Gly Pro Glu Ile Val Ala Asp Ala Phe
195 200 205

Arg Trp Ala His Glu Ala Asp Pro Glu Ala Lys Leu Phe Leu Asn Asp
210 215 220

Tyr Asn Val Glu Gly Ile Asn Ala Lys Ser Asp Ala Tyr Tyr Glu Leu

225 230 235 240

Ala Gln Glu Met Leu Glu Gln Gly Val Pro Leu His Gly Phe Gly Ala
245 250 255

Gln Gly His Leu Ser Thr Arg Tyr Gly Phe Pro Gly Asp Leu Gln Gln
260 265 270

Asn Leu Gln Arg Phe Ala Asp Leu Gly Leu Glu Thr Ala Ile Thr Glu
275 280 285

Ile Asp Val Arg Met Asp Leu Pro Ala Ser Gly Lys Pro Thr Lys Glu
290 295 300

Gln Leu Arg Gln Gln Ala Asp Tyr Tyr Gln Gln Ala Leu Ser Ala Cys
305 310 315 320

Leu Ala Val Asn Asp Cys Asn Ser Phe Thr Ile Trp Gly Phe Thr Asp
325 330 335

Lys Tyr Ser Trp Val Pro Val Phe Phe Glu Gly Glu Gly Ser Ala Thr
340 345 350

Val Met Thr Glu Lys Phe Val Arg Lys Pro Ala Phe Phe Ala Leu Gln
355 360 365

Ser Thr Leu Lys Glu Ala Arg Lys Arg
370 375

<210>	11
<211>	1080
<212>	DNA
<213>	Unknown

<220>
<223> Obtained from environmental sample

<400> 11

atgccctgga gctcatcaac gggacctgca cctatgacga gtaaccgcc cctcaaacgc 60

ccccctgcgt tccgtctggt cggcacgggc atcggctcac tgcacgccgc cggaatttcc 120

cggatgcctc agcttgccac gctgggggcc atctgtgggc ttgataacca cgccgtgaat 180

gccctagcca cagctacgg ggtagaaaaa accacatctc gctatgagga ttactgaac 240

gacccggcc ttgatgtcat cgatctgtgc gttcctcac atgaacacat gcccatggcc 300

attgccgcgc cccggggccg aaaacatctc ctcacgaaa aacctttggc cgcacacctg 360

gaagaggccg atgcaatcct cgaggccgtg aaaagcgcgc gtgtaacgt gatgatggga 420

cacaaccagc gttactacgc ccatcacgcc agggctaaag cattggctga cgccggggtc 480

atcgaaaaac cctacatgat cgtagcttcg gttcatgtgc acgggcagat tgatggtttt 540

cgccgctttc ttaagcacgc cgggggtggc acgttgatcg attcgggagt gcaccgcttc 600

gacctcattc gctggatcat ggggtgaagtc gagaccgtct tcgctcaaac gggtcgcttc 660

ctccagatgc aaatggaagg agaagactgc gcggtgtca cctccgctt ccgcagcgga 720

gccatcgga gcttctcatg cagctggagc gccaaaggcc ctgttccaga agaaacattg 780

caaattttcg gccctatgg ttcgatttat accgaagacc acaccgcac cttacgcctt 840

tacaccgaaa gaccacccc cgaactggaa gacgtaaggc agtttgtctt cccggtcgat 900

caggctgagt ccatccgccg catgattgaa gcgcacttca ccagcctgca acaggggtta 960

ccccctcga tcaccggtat ggacggacgc gttcccttg agctcagcat ggcctcctat 1020

cgctcggctc aaaccggcca gcctgttcat ctcccttc agagaggaaa ccagaaatga 1080

<210> 12

<211> 359

<212> PRT

<213> Unknown

<220>

<223> Obtained from environmental sample

<220>

<221> DOMAIN

<222> (22)...(141)

<223> Oxidoreductase family, NAD-binding Rossmann fold

<220>

<221> DOMAIN

<222> (153)...(260)

<223> Oxidoreductase family, C-terminal alpha/beta domain

<400> 12

Met Pro Trp Ser Ser Ser Thr Gly Pro Ala Pro Met Thr Ser Asn Pro
1 5 10 15

Pro Leu Lys Arg Pro Leu Arg Ile Gly Leu Val Gly Thr Gly Ile Gly
20 25 30

Ser Leu His Ala Ala Gly Ile Ser Arg Met Pro Gln Leu Ala Thr Leu
35 40 45

Gly Ala Ile Cys Gly Leu Asp Thr His Ala Val Asn Ala Leu Ala Thr
50 55 60

Arg Tyr Gly Val Glu Lys Thr Thr Ser Arg Tyr Glu Asp Leu Leu Asn
65 70 75 80

Asp Pro Gly Leu Asp Val Ile Asp Leu Cys Val Pro His Asp Glu His
85 90 95

Met Pro Met Ala Ile Ala Ala Ala Arg Ala Gly Lys His Leu Leu Ile
100 105 110

Glu Lys Pro Leu Ala Arg Thr Leu Glu Glu Ala Asp Ala Ile Leu Glu
115 120 125

Ala Val Lys Ser Ala Gly Val Thr Leu Met Met Gly His Asn Gln Arg
130 135 140

Tyr Tyr Ala His His Ala Arg Ala Lys Ala Leu Val Asp Ala Gly Val
145 150 155 160

Ile Gly Lys Pro Tyr Met Ile Val Ala Ser Val His Val His Gly Gln
165 170 175

Ile Asp Gly Phe Arg Arg Phe Leu Lys His Ala Gly Gly Gly Thr Leu
180 185 190

Ile Asp Ser Gly Val His Arg Phe Asp Leu Ile Arg Trp Ile Met Gly
195 200 205

Glu Val Glu Thr Val Phe Ala Gln Thr Gly Arg Phe Leu Gln Met Gln
210 215 220

Met Glu Gly Glu Asp Cys Ala Val Val Thr Leu Arg Phe Arg Ser Gly
225 230 235 240

Ala Ile Gly Ser Phe Ser Cys Ser Trp Ser Ala Lys Gly Pro Val Pro
245 250 255

Glu Glu Thr Leu Gln Ile Phe Gly Pro Tyr Gly Ser Ile Tyr Thr Glu
260 265 270

Asp His Thr Arg Thr Leu Arg Leu Tyr Thr Glu Arg Pro Thr Pro Glu
275 280 285

Leu Glu Asp Val Arg Gln Phe Val Phe Pro Val Asp Gln Ala Glu Ser
290 295 300

Ile Arg Arg Met Ile Glu Ala His Phe Thr Ser Leu Gln Gln Gly Leu
305 310 315 320

Pro Pro Pro Ile Thr Gly Met Asp Gly Arg Ala Ser Leu Glu Leu Ser
325 330 335

Met Ala Ser Tyr Arg Ser Ala Gln Thr Gly Gln Pro Val His Leu Pro
340 345 350

Leu Gln Arg Gly Asn Gln Lys
355

<210> 13

<211> 1038

<212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 13

atgagcccg tgccggtgc tgcacgcgc gccgggcaaa ttgccagcg cgggcattta 60

cccgggttc tggaagctgg cgccgaaatt accgttctgt gcgataattc ctttctcag 120

cttgaagaaa ttggggccaa atttcacgtt caccgggtct accgcgactg gcacgccatg 180

ctggatgccg gcggattcga agccgtcacc atttgtaccc cgccttctct ccatgccgag 240

atggccatcg aatgtgcccg cagagggttg catgtactgg tagaaaaacc catggctgta 300

aatctccaac aatgcgatca aatgatcgcc gcgtctgaac aggccggaac catcttaatg 360

gtctgcata accagcgctt tatggaggca catcgtctgg ccaaagaaat cttgatgcc 420

ggcctcctcg gcaggctcta cctggcgcac ggggtctttg gccacggcgg cccggagggt 480

tggagcccaa ccagcaatg gtacttcga cctgaccgcg ccggcgctgg cgtgatcgct 540

gacctgggtg atcataaact tgacctgatc cgctggctca ccgggcaaga aattaccgcg 600

gtgggagcac tgggcgccac ctttgaaaag caaacctcgc ttgaagactc tgctgtgatg 660

ctggttcacc tticggaggg tactctcgcc accatccagg taagctgggt gttcaggcct 720

gactgggaaa acagcctggt ccttcgagga gaacgggggg tgctcgccat cccactgat 780

gcctcgcaac ccctgcgggt ctcttacata tcttcttcgg gtcaggatcat tgaaagtacg 840

catcggttgcg actccggcga tacctccggc tggttcggag cgatccgggc atttctcacc 900

gcgatcgaaa aaagcgtccc cgctcccat t gacggaaaag aagggcgtgc tgtcatggcg 960

gcagttctgg cggccacacg ctccattcaa aaacatacga tcatttctat aaccgaggta 1020

gaaaccatcc atgactga 1038

<210> 14
 <211> 345
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> DOMAIN
 <222> (4)...(123)
 <223> Oxidoreductase family, NAD-binding Rossmann fold

<220>
 <221> DOMAIN
 <222> (135)...(248)
 <223> Oxidoreductase family, C-terminal alpha/beta domain

<400> 14
 Met Ser Pro Val Arg Val Ala Val Ile Gly Ala Gly Gln Ile Ala Gln
 1 5 10 15

Arg Gly His Leu Pro Gly Leu Leu Glu Ala Gly Ala Glu Ile Thr Val
 20 25 30

Leu Cys Asp Asn Ser Leu Pro Gln Leu Glu Glu Ile Gly Ala Lys Phe
35 40 45

His Val His Arg Val Tyr Arg Asp Trp His Ala Met Leu Asp Ala Gly
50 55 60

Gly Phe Glu Ala Val Thr Ile Cys Thr Pro Pro Phe Leu His Ala Glu
65 70 75 80

Met Ala Ile Glu Cys Ala Arg Arg Gly Leu His Val Leu Val Glu Lys
85 90 95

Pro Met Ala Val Asn Leu Gln Gln Cys Asp Gln Met Ile Ala Ala Ser
100 105 110

Glu Gln Ala Gly Thr Ile Leu Met Val Ser His Asn Gln Arg Phe Met
115 120 125

Glu Ala His Arg Leu Ala Lys Glu Ile Leu Asp Ala Gly Leu Leu Gly
130 135 140

Arg Leu Tyr Leu Ala His Gly Val Phe Gly His Gly Gly Pro Glu Val
145 150 155 160

Trp Ser Pro Thr Gln Gln Trp Tyr Phe Arg Pro Asp Arg Ala Gly Ala
165 170 175

Gly Val Ile Ala Asp Leu Gly Tyr His Lys Leu Asp Leu Ile Arg Trp
180 185 190

Leu Thr Gly Gln Glu Ile Thr Ala Val Gly Ala Leu Gly Ala Thr Phe
195 200 205

Glu Lys Gln Thr Ser Leu Glu Asp Ser Ala Val Met Leu Val His Leu
210 215 220

Ser Glu Gly Thr Leu Ala Thr Ile Gln Val Ser Trp Val Phe Arg Pro
225 230 235 240

Asp Trp Glu Asn Ser Leu Val Leu Arg Gly Glu Arg Gly Val Leu Ala
245 250 255

Ile Pro Thr Asp Ala Ser Gln Pro Leu Arg Val Ser Tyr Ile Ser Ser
260 265 270

Ser Gly Gln Val Ile Glu Ser Thr His Arg Cys Asp Ser Gly Asp Thr
275 280 285

Ser Gly Trp Phe Gly Ala Ile Arg Ala Phe Leu Thr Ala Ile Glu Lys
290 295 300

Ser Ala Pro Ala Pro Ile Asp Gly Lys Glu Gly Arg Ala Val Met Ala
305 310 315 320

Ala Val Leu Ala Ala Thr Arg Ser Ile Gln Lys His Thr Ile Ile Ser
325 330 335

Ile Thr Glu Val Glu Thr Ile His Asp
340 345

<210> 15

<211> 1347

<212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 15

atgactgacc atcgttttcc aaaaggattc atctggggaa ccgctacggc gtctttccag 60

attgaaggcg ccacccgcga agatggccgg ggcgaatcca tctgggaccg cttctgcgcc 120

acgccgggga aaattgtcac gggcgaaacc ggcgatcctg cctgcgactc ctatcatcgt 180

tacctgaag acatgccct gatgaagct atgtcgtca atggttacg cttttcaatc	240
gcctggcctc gcgtcattcc tgacggagac ggtaaagtct gtcaggccgg gctcgactac	300
tacgatcgtg tggtagatgc tctcctggcg gagaatatcc aacctttat caccctgtac	360
cactgggacc tgccccaggc attacaggat cggggtggct ggggcaaccg tgccacggtt	420
gaggcgttca ctgctacgt agatattgtg gtttctcgcc tgggtgaccg cgtaaagta	480
tggatgacac acaacgaacc ctggtgtgta tccattttga gccatgagct tggatgaacat	540
gccccgggt tgaaggaccg aaaactggcc ctccaggagg cgcaccatgt cctcgtttct	600
cacggcctgg ccgtgccat catccgccag cgttgtaaag aggcgcaggt tggcatcgtg	660
ttgaatttt cactgctta cccggccacc gatagcctgg ccgaccagat ggccaccgt	720
cagcaccacg cccggtttaa cctctggttc ctgatccca tcgccggcg cggtaccg	780
caggatgcct ggaagggtta cggagccgat gttcccga tgaggcctga tgacatgcag	840
atcatcgccg ccccatcga ctctctgggc gtcaatttct acagtcgggc ggtctgccac	900
gatccggccg ggggcgaagg ttcccgggtg ctcaatgtgc gcagtaaac cgaggccacc	960
gatcgagact gggagattta cctcaggcg ctctacgatt tactcatctg gatccacaat	1020
ggataccagt tcagagatat ttacattacc gagaatggcg cctcatacaa cgatgtggtc	1080
tccccgatg ggaagtga cgatcctaaa cgtctggact atctgaaac ccatctggcc	1140
atggctctgc ggcccatcga agcgggcgtt cactgcgtg gttatttctg ctggagcttg	1200
atggacaact tcgaatgggc catgggcacc agcagccgat tcgggttggc ctacaccgac	1260
ttcactacc agaagcgtat tctcaaagac agtgggtctt ggtttggcga agtggcacgg	1320
gcaaacgcct taatcgacct tccctga	1347

<210> 16
 <211> 448
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> DOMAIN
 <222> (2)...(444)
 <223> Glycosyl hydrolase family 1

<220>
 <221> SITE
 <222> (10)...(24)
 <223> Glycosyl hydrolases family 1 N-terminal signature. Prosite id = PS00653

<220>
 <221> SITE
 <222> (352)...(360)
 <223> Glycosyl hydrolases family 1 active site. Prosite id = PS00572

<400> 16
 Met Thr Asp His Arg Phe Pro Lys Gly Phe Ile Trp Gly Thr Ala Thr
 1 5 10 15

Ala Ser Phe Gln Ile Glu Gly Ala Thr Arg Glu Asp Gly Arg Gly Glu
 20 25 30

Ser Ile Trp Asp Arg Phe Cys Ala Thr Pro Gly Lys Ile Val Thr Gly
 35 40 45

Glu Thr Gly Asp Pro Ala Cys Asp Ser Tyr His Arg Tyr Pro Glu Asp
 50 55 60

Ile Ala Leu Met Lys Ala Met Ser Leu Asn Gly Tyr Arg Phe Ser Ile
 65 70 75 80

Ala Trp Pro Arg Val Ile Pro Asp Gly Asp Gly Lys Val Cys Gln Ala

85

90

95

Gly Leu Asp Tyr Tyr Asp Arg Val Val Asp Ala Leu Leu Ala Glu Asn
100 105 110

Ile Gln Pro Phe Ile Thr Leu Tyr His Trp Asp Leu Pro Gln Ala Leu
115 120 125

Gln Asp Arg Gly Gly Trp Gly Asn Arg Ala Thr Val Glu Ala Phe Thr
130 135 140

Arg Tyr Val Asp Ile Val Val Ser Arg Leu Gly Asp Arg Val Lys Tyr
145 150 155 160

Trp Met Thr His Asn Glu Pro Trp Cys Val Ser Ile Leu Ser His Glu
165 170 175

Leu Gly Glu His Ala Pro Gly Leu Lys Asp Arg Lys Leu Ala Leu Gln
180 185 190

Val Ala His His Val Leu Val Ser His Gly Leu Ala Val Pro Ile Ile
195 200 205

Arg Gln Arg Cys Lys Glu Ala Gln Val Gly Ile Val Leu Asn Phe Ser
210 215 220

Pro Ala Tyr Pro Ala Thr Asp Ser Leu Ala Asp Gln Met Ala Thr Arg
225 230 235 240

Gln His His Ala Arg Phe Asn Leu Trp Phe Leu Asp Pro Ile Ala Gly
245 250 255

Arg Gly Tyr Pro Gln Asp Ala Trp Glu Gly Tyr Gly Ala Asp Val Pro
260 265 270

Ala Met Arg Pro Asp Asp Met Gln Ile Ile Ala Ala Pro Ile Asp Phe

275 280 285

Leu Gly Val Asn Phe Tyr Ser Arg Ala Val Cys His Asp Pro Ala Gly
290 295 300

Gly Glu Gly Ser Arg Val Leu Asn Val Arg Ser Lys Thr Glu Ala Thr
305 310 315 320

Asp Arg Asp Trp Glu Ile Tyr Pro Gln Ala Leu Tyr Asp Leu Leu Ile
325 330 335

Trp Ile His Asn Gly Tyr Gln Phe Arg Asp Ile Tyr Ile Thr Glu Asn
340 345 350

Gly Ala Ser Tyr Asn Asp Val Val Ser Pro Asp Gly Lys Val His Asp
355 360 365

Pro Lys Arg Leu Asp Tyr Leu Lys Arg His Leu Ala Met Ala Leu Arg
370 375 380

Ala Ile Glu Ala Gly Val Pro Leu Arg Gly Tyr Phe Cys Trp Ser Leu
385 390 395 400

Met Asp Asn Phe Glu Trp Ala Met Gly Thr Ser Ser Arg Phe Gly Leu
405 410 415

Ala Tyr Thr Asp Phe Thr Thr Gln Lys Arg Ile Leu Lys Asp Ser Gly
420 425 430

Leu Trp Phe Gly Glu Val Ala Arg Ala Asn Ala Leu Ile Asp Leu Pro
435 440 445

<210> 17

<211> 1215

<212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 17

atgcggtacg tgctgatttc ctgccttgcg ctggcttccc tgtgcgcga gcctcttct	60
gtttcacgc ctgaaaaaga gggcttctcg gcggagcgcc tcgggcggat gcaccggtat	120
ttcgagaacc tgacgaaaac cggagagcgg cctggcgcca tcacgtgat cgtgcgcaac	180
gggcgcacgc tggactggcg cacgttcggg ctgcgcgacg tcgagaacaa tctgccgatg	240
gagaaggaca cgatcgtcca catctactcg atgacgaagc cggtagacgtc cgtggccgtg	300
atgatgctgg tggaggaggg caggctggcg ctggacgacc gggtaggaca gttcattccc	360
gagttcaagg ggaatgaagg gtacaagggc ggcacggtgg agcggccgga gctggaggac	420
gcggcgcggc cgatcacggt gaagcatctg ctgacgcaca cgagcgggct gagctacggc	480
tggggcaacg acaacgtctc cgcatgtac cgcaaggccg acccgctcgg cgcgccgagc	540
ctgaaagagt ttatcgacag gctggtgaaa ctgccgtgg cattccaccc gggcgagcgt	600
tacgagtatt cgatgtcgat cgacgtgctg ggctacctgg tggaggctgt ctccggcgag	660
ccgttcgac agttcgtgga gaagcggatc acggggccgc tgaagatgaa cgacacgcat	720
ttcagactgc cggaggcgaa gcgggcgcgg ctggcgaaga tctactcgcg gcgcgagggg	780
aagctgacgg cgcagcgagg cctgcagacg ggaggcgttc cgtacggcgg catggggctg	840
tactcgacga tcggcgacta tgcgcggttc gcgcagatgc tgttgaacgg cggccatctc	900
gacggagtgc gccctgtggg gcggaagacg gtggatctga tgatgatgaa ccatctgggc	960
ggactgtcga agccgacgat cggcggcgat gattcagcgg gattcggact gggcggagcg	1020
gtgcggatcg atccggcgaa atcgggccgt ccgggcacgg aaggactctt cggctgggac	1080

ggggcggctt cgacgtatit cgggtggac cgaaagaga agctggcgat gctgctgttc 1140

ctgcaatgga tgccgtttga tcaggggacg ctgaacctgt acgagacgct ggtctaccaa 1200

gctctggtgg actga 1215

<210> 18
 <211> 404
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> DOMAIN
 <222> (37)...(400)
 <223> Beta-lactamase

<220>
 <221> SITE
 <222> (43)...(46)
 <223> N-glycosylation site. Prosite id = PS00001

<220>
 <221> SITE
 <222> (167)...(170)
 <223> N-glycosylation site. Prosite id = PS00001

<220>
 <221> SITE
 <222> (240)...(243)
 <223> N-glycosylation site. Prosite id = PS00001

<400> 18
 Met Arg Tyr Val Leu Ile Ser Cys Leu Ala Leu Ala Ser Leu Cys Ala
 1 5 10 15

Gln Pro Leu Pro Val Ser Thr Pro Glu Lys Glu Gly Phe Ser Ala Glu
 20 25 30

Arg Leu Gly Arg Met His Arg Tyr Phe Glu Asn Leu Thr Lys Thr Gly
 35 40 45

Glu Arg Pro Gly Ala Ile Thr Leu Ile Val Arg Asn Gly Arg Ile Val
50 55 60

Asp Trp Arg Thr Phe Gly Leu Arg Asp Val Glu Asn Asn Leu Pro Met
65 70 75 80

Glu Lys Asp Thr Ile Val His Ile Tyr Ser Met Thr Lys Pro Val Thr
85 90 95

Ser Val Ala Val Met Met Leu Val Glu Glu Gly Arg Leu Ala Leu Asp
100 105 110

Asp Arg Val Asp Lys Phe Ile Pro Glu Phe Lys Gly Met Lys Val Tyr
115 120 125

Lys Gly Gly Thr Val Glu Arg Pro Glu Leu Glu Asp Ala Ala Arg Pro
130 135 140

Ile Thr Val Lys His Leu Leu Thr His Thr Ser Gly Leu Ser Tyr Gly
145 150 155 160

Trp Gly Asn Asp Asn Val Ser Ala Met Tyr Arg Lys Ala Asp Pro Leu
165 170 175

Gly Ala Pro Ser Leu Lys Glu Phe Ile Asp Arg Leu Val Lys Leu Pro
180 185 190

Leu Ala Phe His Pro Gly Glu Arg Tyr Glu Tyr Ser Met Ser Ile Asp
195 200 205

Val Leu Gly Tyr Leu Val Glu Ala Val Ser Gly Glu Pro Phe Asp Gln
210 215 220

Phe Val Glu Lys Arg Ile Thr Gly Pro Leu Lys Met Asn Asp Thr His

225 230 235 240

 Phe Arg Leu Pro Glu Ala Lys Arg Ala Arg Leu Ala Lys Ile Tyr Ser
 245 250 255

 Arg Arg Glu Gly Lys Leu Thr Ala Gln Arg Gly Leu Gln Thr Gly Gly
 260 265 270

 Val Pro Tyr Gly Gly Met Gly Leu Tyr Ser Thr Ile Gly Asp Tyr Ala
 275 280 285

 Arg Phe Ala Gln Met Leu Leu Asn Gly Gly His Leu Asp Gly Val Arg
 290 295 300

 Leu Leu Gly Arg Lys Thr Val Asp Leu Met Met Met Asn His Leu Gly
 305 310 315 320

 Gly Leu Ser Lys Pro Thr Ile Gly Gly Asp Asp Ser Ala Gly Phe Gly
 325 330 335

 Leu Gly Gly Ala Val Arg Ile Asp Pro Ala Lys Ser Gly Arg Pro Gly
 340 345 350

 Thr Glu Gly Leu Phe Gly Trp Asp Gly Ala Ala Ser Thr Tyr Phe Arg
 355 360 365

 Val Asp Arg Lys Glu Lys Leu Ala Met Leu Leu Phe Leu Gln Trp Met
 370 375 380

 Pro Phe Asp Gln Gly Thr Leu Asn Leu Tyr Glu Thr Leu Val Tyr Gln
 385 390 395 400

 Ala Leu Val Asp

<210> 19
 <211> 1794
 <212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 19

atgccggttt tgitcgccct gtttcttggt gcctcgtcct gcgcggcgca gtcgctggcc	60
gggccggttt ccctgcttgg cggagatgcg ggcgcgcgct tccgtatac cgggccatcg	120
gcggcgcgcg cgagcggtc ggccgaatgg gtggcggtgg agaacaatgcc gttcacgcac	180
gcctggcgcg tgcgcacgaa tccgtgccg gagagcgcg gcaacgaatg ggacctgcgc	240
atccgcgccc gcggagcgcg ggctgtttcg gcaggggaca agatcctggc cgagtcttgg	300
atgcgtgcg tggagcccga aaacggcgac tgcattctgc gcctgaacgt ggagcgcgac	360
gggtcgccgt ggaccaaacc catcagcaac ccctaccggt tgggccggga gtggcgcgcg	420
ttccgctgc tgttcgagat gcgggagagc tacccgccc gcggtacat gatcgatttc	480
tggatgggcc agcaggtgca gacggcgga gtggcgggga tttccctgct gaattacggt	540
ccgcaggcca cggccgagca gcttggcctg gaccggtttt atgagggcg gcggcgcgac	600
gccgctggc ggagggcg cgagcagcg atcgaggaga tccggaagc gggcatgac	660
atcgtggcg tgacccgga cggcgagccg atcgaggcg ctgaaatccg ggcaagctg	720
aagcgcgacg cgttcgggtg gggcacggct gtggcgcat cacggcttct ggggacggga	780
acggacagcg agcgtaccg caacttcac gcgagaact tcaacatggc ggtgctcgag	840
aacgacctga aatggggccc gttcgaagag aaccgcaacc gcgcgatgaa cgcgtgcgc	900
tggctgcatg agaacgggat cacgtggatc gcgggcaca atctcgtctg gccgggctgg	960
cggatgatgc cgaacgacgt gcgaacctg gcgaacaac ccgagcgct gcggcagcg	1020

attctggacc gcatcggga cacggccacg gccacgcgcg ggctgggtgt gcactgggac 1080

gtcgtcaacg agccggtggc cgagcgcgac gtgctgaaca ttctgggcga cgaggtgatg 1140

gcggaactgt tccgcgccgc gaaggagtgc gatcccagg cgaggatgtt catcaatgag 1200

tacgacattc tggcggcgaa cggggccaat ctgcggaagc agaacgcgta ttaccgatg 1260

atcgagatgc tgitgaagct cgagcgcgcg gtggagggca tcggcttcca gggccacttc 1320

gacacggcca cgccgccgga gcggtatgtg gagatcatga accggtacgc ccggctcggg 1380

ctgccgatcg ccatcaccga gtacatttc gccacggcgg acgaggagct gcaggcgag 1440

ttcacgcgcg acctgatgat tctcgccttc agccatccgg cggtttcgga ctctctgatg 1500

tggggcttct gggaaggag ccaactggaag ccgctgggcg ccatgatccg gcgcgactgg 1560

agcgagaagc cgatgtaccg cgtctggcgc gagctgatct tcgagcgtg gcagacggat 1620

gaaacaggcg tgacgccgga gcacggtgcc atctacgtgc ggggcttcaa gggcgactac 1680

gagatcacgg tgaaggcggg cgggcaggaa gtccgggtgc cgtacacgt gaaagaagac 1740

ggccagggtc tgtgggtgac ggtgggcggg gcttctgaag agcgcgtgca gtaa 1794

<210> 20
 <211> 597
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> SIGNAL
 <222> (1)...(20)

<220>
 <221> DOMAIN

<222> (235)...(533)

<223> Glycosyl hydrolase family 10

<220>

<221> SITE

<222> (467)...(477)

<223> Glycosyl hydrolases family 10 active site. Prosite id = PS00591

<400> 20

Met	Pro	Val	Leu	Phe	Ala	Leu	Phe	Leu	Val	Ala	Ser	Ser	Cys	Ala	Ala
1			5					10						15	

Gln	Ser	Leu	Ala	Gly	Pro	Val	Ser	Leu	Leu	Gly	Gly	Asp	Ala	Gly	Ala
		20						25					30		

Ala	Phe	Arg	Tyr	Thr	Gly	Pro	Ser	Ala	Gly	Ala	Ala	Ser	Gly	Ser	Ala
		35					40					45			

Glu	Trp	Val	Ala	Val	Glu	Asn	Met	Pro	Phe	Thr	His	Ala	Trp	Arg	Leu
	50					55					60				

Arg	Thr	Asn	Pro	Leu	Pro	Glu	Ser	Gly	Gly	Asn	Glu	Trp	Asp	Leu	Arg
65					70					75				80	

Ile	Arg	Ala	Arg	Gly	Ala	Ala	Ala	Val	Ser	Ala	Gly	Asp	Lys	Ile	Leu
			85						90					95	

Ala	Glu	Phe	Trp	Met	Arg	Cys	Val	Glu	Pro	Glu	Asn	Gly	Asp	Cys	Ile
			100						105					110	

Leu	Arg	Leu	Asn	Val	Glu	Arg	Asp	Gly	Ser	Pro	Trp	Thr	Lys	Ser	Ile
		115							120				125		

Ser	Asn	Pro	Tyr	Pro	Val	Gly	Arg	Glu	Trp	Arg	Arg	Phe	Arg	Val	Leu
	130						135					140			

Phe	Glu	Met	Arg	Glu	Ser	Tyr	Ala	Ala	Gly	Gly	Tyr	Met	Ile	Asp	Phe

145	150	155	160
Trp Met Gly Gln Gln Val Gln Thr Ala Glu Val Gly Gly Ile Ser Leu			
	165	170	175
Leu Asn Tyr Gly Pro Gln Ala Thr Ala Glu Gln Leu Gly Leu Asp Arg			
	180	185	190
Phe Tyr Glu Gly Ala Ala Ala Asp Ala Ala Trp Arg Gln Ala Ala Glu			
	195	200	205
Gln Arg Ile Glu Glu Ile Arg Lys Ala Gly Met Ile Ile Val Ala Val			
	210	215	220
Thr Pro Asp Gly Glu Pro Ile Glu Gly Ala Glu Ile Arg Ala Lys Leu			
	225	230	235
Lys Arg His Ala Phe Gly Trp Gly Thr Ala Val Ala Ala Ser Arg Leu			
	245	250	255
Leu Gly Thr Gly Thr Asp Ser Glu Arg Tyr Arg Asn Phe Ile Arg Glu			
	260	265	270
Asn Phe Asn Met Ala Val Leu Glu Asn Asp Leu Lys Trp Gly Pro Phe			
	275	280	285
Glu Glu Asn Arg Asn Arg Ala Met Asn Ala Leu Arg Trp Leu His Glu			
	290	295	300
Asn Gly Ile Thr Trp Ile Arg Gly His Asn Leu Val Trp Pro Gly Trp			
	305	310	315
Arg Trp Met Pro Asn Asp Val Arg Asn Leu Ala Asn Asn Pro Glu Ala			
	325	330	335
Leu Arg Gln Arg Ile Leu Asp Arg Ile Arg Asp Thr Ala Thr Ala Thr			
	340	345	350

Arg Gly Leu Val Val His Trp Asp Val Val Asn Glu Pro Val Ala Glu
355 360 365

Arg Asp Val Leu Asn Ile Leu Gly Asp Glu Val Met Ala Asp Trp Phe
370 375 380

Arg Ala Ala Lys Glu Cys Asp Pro Glu Ala Arg Met Phe Ile Asn Glu
385 390 395 400

Tyr Asp Ile Leu Ala Ala Asn Gly Ala Asn Leu Arg Lys Gln Asn Ala
405 410 415

Tyr Tyr Arg Met Ile Glu Met Leu Leu Lys Leu Glu Ala Pro Val Glu
420 425 430

Gly Ile Gly Phe Gln Gly His Phe Asp Thr Ala Thr Pro Pro Glu Arg
435 440 445

Met Leu Glu Ile Met Asn Arg Tyr Ala Arg Leu Gly Leu Pro Ile Ala
450 455 460

Ile Thr Glu Tyr Asp Phe Ala Thr Ala Asp Glu Glu Leu Gln Ala Gln
465 470 475 480

Phe Thr Arg Asp Leu Met Ile Leu Ala Phe Ser His Pro Ala Val Ser
485 490 495

Asp Phe Leu Met Trp Gly Phe Trp Glu Gly Ser His Trp Lys Pro Leu
500 505 510

Gly Ala Met Ile Arg Arg Asp Trp Ser Glu Lys Pro Met Tyr Arg Val
515 520 525

Trp Arg Glu Leu Ile Phe Glu Arg Trp Gln Thr Asp Glu Thr Gly Val
530 535 540

Thr Pro Glu His Gly Ala Ile Tyr Val Arg Gly Phe Lys Gly Asp Tyr
545 550 555 560

Glu Ile Thr Val Lys Ala Gly Gly Gln Glu Val Arg Val Pro Tyr Thr
565 570 575

Leu Lys Glu Asp Gly Gln Val Leu Trp Val Thr Val Gly Gly Ala Ser
580 585 590

Glu Glu Arg Val Gln
595

<210> 21

<211> 1032

<212> DNA

<213> Clostridium thermocellum

<400> 21

atggtgagtt ttaaagcagg tataaattta ggcggatgga ttcacaata tcaagttttc 60

agcaaagagc atttcgatac attcattacg gagaaggaca ttgaaactat tgcagaagca 120

gggtttgacc atgtcagact gccttttgat tatccaatta tcgagtctga tgacaatgtg 180

ggagaatata aagaagatgg gctttcttat attgaccggg gccttgagtg gtgtaaaaaa 240

tacaatttgg ggcttgtgtt ggatatgcat cacgtcccg ggtaccgctt tcaagatttt 300

aagacaagca ccttgtttga agatccgaac cagcaaaaga gatttgttga catatggaga 360

tttttagcca agcgttacat aaatgaacgg gaacatattg cctttgaact gttaaatgaa 420

gttgttgagc ctgacagtac ccgctggaac aagttgatgc ttgagtgtgt aaaagcaatc 480

agggaattg attccaccag gtggctttac attgggggca ataactataa cagtcctgat 540

gagcttaaaa acctgcaga tattgatgat gattacatag ttacaattt ccatttttac 600

aatccttttt tctttacgca tcagaaagcc cactggtcgg aaagtgccat ggcgtacaac 660

aggactgtaa aaatccggg acaatatgag ggaattgaag agtttgtgaa aaataatcct 720

aagtacagtt ttatgatgga attgaataac ctgaagctga ataaagagct ttgcgcaaa 780

gatttaaaac cagcaattga gttcagggaa aagaaaaaat gcaaaactata ttgcggggag 840

tttggcgtaa ttgccattgc tgacctggag tccaggataa aatggcatga agattatata 900

agtcttctag aggagtatga tatcggcggc gcggtgtgga actacaaaaa aatggatttt 960

gaaatttata atgaggatag aaaacctgtc tcgcaagaat tggtaaatat actggcgaga 1020

agaaaaactt ga 1032

<210> 22
 <211> 343
 <212> PRT
 <213> Clostridium thermocellum

<220>
 <221> DOMAIN
 <222> (1)...(323)
 <223> Cellulase (glycosyl hydrolase family 5)

<220>
 <221> SITE
 <222> (15)...(32)
 <223> Cytosolic fatty-acid binding proteins signature. Prosite id = PS00214

<220>
 <221> SITE
 <222> (135)...(144)
 <223> Glycosyl hydrolases family 5 signature. Prosite id = PS00659

<220>
 <221> SITE
 <222> (223)...(226)
 <223> N-glycosylation site. Prosite id = PS00001

<400> 22

Met Val Ser Phe Lys Ala Gly Ile Asn Leu Gly Gly Trp Ile Ser Gln
1 5 10 15

Tyr Gln Val Phe Ser Lys Glu His Phe Asp Thr Phe Ile Thr Glu Lys
20 25 30

Asp Ile Glu Thr Ile Ala Glu Ala Gly Phe Asp His Val Arg Leu Pro
35 40 45

Phe Asp Tyr Pro Ile Ile Glu Ser Asp Asp Asn Val Gly Glu Tyr Lys
50 55 60

Glu Asp Gly Leu Ser Tyr Ile Asp Arg Cys Leu Glu Trp Cys Lys Lys
65 70 75 80

Tyr Asn Leu Gly Leu Val Leu Asp Met His His Ala Pro Gly Tyr Arg
85 90 95

Phe Gln Asp Phe Lys Thr Ser Thr Leu Phe Glu Asp Pro Asn Gln Gln
100 105 110

Lys Arg Phe Val Asp Ile Trp Arg Phe Leu Ala Lys Arg Tyr Ile Asn
115 120 125

Glu Arg Glu His Ile Ala Phe Glu Leu Leu Asn Glu Val Val Glu Pro
130 135 140

Asp Ser Thr Arg Trp Asn Lys Leu Met Leu Glu Cys Val Lys Ala Ile
145 150 155 160

Arg Glu Ile Asp Ser Thr Arg Trp Leu Tyr Ile Gly Gly Asn Asn Tyr
165 170 175

Asn Ser Pro Asp Glu Leu Lys Asn Leu Ala Asp Ile Asp Asp Asp Tyr
180 185 190

Ile Val Tyr Asn Phe His Phe Tyr Asn Pro Phe Phe Phe Thr His Gln
195 200 205

Lys Ala His Trp Ser Glu Ser Ala Met Ala Tyr Asn Arg Thr Val Lys
210 215 220

Tyr Pro Gly Gln Tyr Glu Gly Ile Glu Glu Phe Val Lys Asn Asn Pro
225 230 235 240

Lys Tyr Ser Phe Met Met Glu Leu Asn Asn Leu Lys Leu Asn Lys Glu
245 250 255

Leu Leu Arg Lys Asp Leu Lys Pro Ala Ile Glu Phe Arg Glu Lys Lys
260 265 270

Lys Cys Lys Leu Tyr Cys Gly Glu Phe Gly Val Ile Ala Ile Ala Asp
275 280 285

Leu Glu Ser Arg Ile Lys Trp His Glu Asp Tyr Ile Ser Leu Leu Glu
290 295 300

Glu Tyr Asp Ile Gly Gly Ala Val Trp Asn Tyr Lys Lys Met Asp Phe
305 310 315 320

Glu Ile Tyr Asn Glu Asp Arg Lys Pro Val Ser Gln Glu Leu Val Asn
325 330 335

Ile Leu Ala Arg Arg Lys Thr
340

<210> 23

<211> 3966

<212> DNA

<213> Clostridium thermocellum

<400> 23

atgtataaaa gattattgtc gtcagtactg ataattatgc tggtattatc agcctggctcg 60

ccaatatccg tacaagcttc tgatggaatc aatgacatta gaggtcattg ggctgaagaa	120
gacttgaaca aatggatgga aaaaggtatt ttggtgggct accaggatgg gacgataagg	180
cccgataata atatcacaag agccgaattt gtcacattaa ttaacaaggt tttcgggctt	240
tatgaattaa gccgggagca attcgcagat gttgaagact caaaatggta ttcccgtaga	300
atattaaaag ccagggctgc gggatatatt gcaggttatg gaagcaatgt tttcaaacct	360
gacaattata ttacaagaca agaagccgtt gttataatcg cgaaagtttt tgaacttcaa	420
agcggcagca attatacaag caagtttaaa gatggaagtc tggtaaagga atacgcaaaa	480
gattccgtta gcgcgttggg tgaaaaaggc tacatagcag gttatgaaga tggcactttc	540
aggccggaca actacattac ccgtgcagaa acaataaaaa ttctgaataa aattattcct	600
tccttgata acgagaaaagg agattataaa aatgaagaag tagccggaaa cgctctgatt	660
aacaccgaag gaggttattt aaaagatacc gtaataaacg gggatttgta tcttgctcag	720
ggaattcaga acggcgatgt tacccttgac ggtgtgaatg taaaaggaac ggttttcgta	780
aatgggtggag gaagcgacag catacatitt ataaatcga aaataaacag ggttgtgtgc	840
aataaaacag gaggtagaat tgtaacttcc ggcaatacct cggttgaaag tgttgtcgtt	900
aatccgggtg caaaacttga agaaaaagaa ttgacgggcg acggctttta aaacgttaca	960
gtcgattctc aactttcagc cggcaatgaa ataatatgtg tcggggattt tgaacaggtc	1020
gatgttctgg cggatgatgc cttgctggaa accaaagagg caaaaatgaa actgagaata	1080
ttcgccaaa ggattaaagt aaatggaaag gcaatagaaa aatcatcaaa gaactatatt	1140
gtaaacgggg aacttatatc aactgaggaa gaacccggtc cttccgacgc acccggtgcg	1200
gaagacgac aaaattcagg tagtccgggc tcatcgacta atcctgcacc aaccaagaat	1260

ccgaatgaag agtggcgctc ggtttggagc gatgagttta acggttctga aataaatatg 1320

gctaattgga gctatgacga cccgaccaac ggaagatgga acggggaagt acaatcctac 1380

acacaaaaca atgcctatat caaagacggc gcgttggtta ttgaagcaag aaaagaagac 1440

attacggaac caagcggiga gacttatcat tatacatcgt caaagctgat taccaaaggc 1500

aaaaagtcac ggaagtacgg aaaatttgaa ataagggcaa aaatgccaca gggacaaggt 1560

atatggcctg caatctggat gatgccgga gacgaaccct tctacggaac atggccaaag 1620

tgcggcgaaa tagatattat ggagcttttg ggccacgagc ctgataaaat ttatggaacg 1680

atccattttg gagagcctca taaagaatcc cagggaacgt ataccttgcc ggaaggccag 1740

acttttgctg atgatttcca cgtttattcg attgaatggg aaccgggaga aatacgtgg 1800

tatatagacg gcaagctgta tcatgtcgt aatgactggg actcgaggga cccgtacctt 1860

gccgatgact acacttatcc cgcacctttt gaccagaatt tcttcttgat tctcaatata 1920

tccgttgggt gcggctggcc gggatatcct gacgaaacga cagttttccc gcagcaaatg 1980

gttgtggact atgtgagagt atatcaaaaa gataaatatc ctcacaggga aaaaccggca 2040

aaggaagaag tgaagccaag agagcctctt gaggacggca attatatcta taacggcggg 2100

tttgatgtgg atgattctgc agcagttggt gtggacgggtg ttcctatac gtcttactgg 2160

acattcttaa cagcatccgg tggagctgcg acagtcaatg tagaggaagg tgttatgcac 2220

gtacagatag aaaacggagg gacaaccgac tacggcgtac aattgcttca agctccgatt 2280

catcttgaaa aaggcgcaaa atataagca tcttttgaca tgaaagctga aaatccaagg 2340

caggtaaaac tgaaaatagg cggagacggc gacaggggat ggaaagatta tgcggctatt 2400

ccaccgttta cggttctcaac agagatgacc aactatgagt ttgagtttac tatgaaagat 2460

gataccgatg ttaaggcagc gtttgagttt aatatgggtt tggacgataa tgatgtctgg 2520

attgacaatg ttaaactgat taaaacagaa gatgcgccgg ttatagatcc ttccgaaata 2580

gcaagacctc cgcttctttc cggcaactat atatacaacg gtacctttga ccaaggtccg 2640

aacagaatgg gattctggaa ttttgttgat gatagcactg caaaggctac atactatatt 2700

ggaagcgatg ttaatgagcg caggtttgaa acaagaatag aaaaaggcgg aacatcgagg 2760

ggagccataa gattggttca gccgggaatt aacattgaaa acggcaaaac atacaaggtt 2820

agcttcgaag ccagtgcggc aaatacaaga actattgagg tggaaattgc aagcaatctt 2880

cacaacagca gcatttttgc gacaactttt gaaataagca aagagagcaa gatatacgaa 2940

tttgagttaa caatggacaa agattcggac aagaacggag aacttaggtt caatctgggc 3000

ggaagcaacg tgaacgtcta tattgataat gtcgttatga aaagagtaag taccgatgaa 3060

gttgaaggaa acctgatttt aaacggcgta tttaacggcc tggcaggctg gggatatgga 3120

gcgtatgaac ctggatcggc agattttgaa agtcatgagg aacaatttag ggcaattatt 3180

agctctgtcg gtaatgaagg ttggaatgta cagttgtatc aggataatgt tccgctggaa 3240

caagggcaaa cctacgaagt ttcttttgat gcaaaatcaa cgattgacag aaagataatt 3300

gttcagctgc aaaggaacgg tacttcggat aataattggg actcctatct ctatcaagaa 3360

gttgaactta ctaatgaact taaaacattc aaatatgaat ttacaatgag taaacctaca 3420

gattcggcgt caagatttaa ttttgctttg ggtaatactg aaaacaaaac ttatgtctct 3480

catgaaataa taattgacaa tgtttagta agaaaagttg cgactccttc tgcgctgata 3540

ttgaacggaa cctttgacga tggaatggat catggctgc tatactgggg agacggtgaa 3600

ggcaattgcg atgtaactga cggagagctt gaaattaaca ttaccaaggt aggtaccgcg 3660

gattacatgc cgcagattaa acaggaaaac atagcgttgc aagagggtgt gacgtatact 3720

ttgtctctta aagcagagagc gcttgaggca agaagtatta aagtgacat attggattct 3780

tcttataact ggtatggcgg aactattttc gatttaacaa cggaagatgc cgtatacacg 3840

tttacattta cccaagcaa gtcgataaat aacggtgtct taactataaa ttaggtacc 3900

atagaaggta agacatccgc cgcaactact gtctatcttg atgatatttt gctggaacaa 3960

cagtaa 3966

<210> 24
 <211> 1321
 <212> PRT
 <213> Clostridium thermocellum

<220>
 <221> SIGNAL
 <222> (1)...(26)

<220>
 <221> DOMAIN
 <222> (30)...(71)
 <223> S-layer homology domain

<220>
 <221> DOMAIN
 <222> (88)...(130)
 <223> S-layer homology domain

<220>
 <221> DOMAIN
 <222> (149)...(192)
 <223> S-layer homology domain

<220>
 <221> DOMAIN
 <222> (445)...(666)
 <223> Glycosyl hydrolases family 16

<220>
 <221> DOMAIN
 <222> (693)...(849)
 <223> Carbohydrate binding domain

<220>
 <221> DOMAIN
 <222> (868)...(1016)
 <223> Carbohydrate binding domain

<220>
 <221> DOMAIN
 <222> (1023)...(1173)
 <223> Carbohydrate binding domain

<220>
 <221> DOMAIN
 <222> (1177)...(1321)
 <223> Carbohydrate binding domain

<220>
 <221> SITE
 <222> (146)...(149)
 <223> N-glycosylation site. Prosite id = PS00001

<220>
 <221> SITE
 <222> (285)...(288)
 <223> N-glycosylation site. Prosite id = PS00001

<220>
 <221> SITE
 <222> (296)...(299)
 <223> N-glycosylation site. Prosite id = PS00001

<220>
 <221> SITE
 <222> (322)...(325)
 <223> N-glycosylation site. Prosite id = PS00001

<220>
 <221> SITE
 <222> (440)...(443)
 <223> N-glycosylation site. Prosite id = PS00001

<220>
 <221> SITE
 <222> (448)...(451)
 <223> N-glycosylation site. Prosite id = PS00001

<220>
 <221> SITE
 <222> (648)...(651)
 <223> N-glycosylation site. Prosite id = PS00001

<220>
 <221> SITE
 <222> (886)...(889)
 <223> N-glycosylation site. Prosite id = PS00001

<220>
 <221> SITE
 <222> (976)...(979)
 <223> N-glycosylation site. Prosite id = PS00001

<220>
 <221> SITE
 <222> (1123)...(1126)
 <223> N-glycosylation site. Prosite id = PS00001

<220>
 <221> SITE
 <222> (1172)...(1175)
 <223> N-glycosylation site. Prosite id = PS00001

<220>
 <221> SITE
 <222> (1200)...(1203)
 <223> N-glycosylation site. Prosite id = PS00001

<220>
 <221> SITE
 <222> (1231)...(1234)
 <223> N-glycosylation site. Prosite id = PS00001

<400> 24
 Met Tyr Lys Arg Leu Leu Ser Ser Val Leu Ile Ile Met Leu Leu Leu
 1 5 10 15

Ser Ala Trp Ser Pro Ile Ser Val Gln Ala Ser Asp Gly Ile Asn Asp

20

25

30

Ile Arg Gly His Trp Ala Glu Glu Asp Leu Asn Lys Trp Met Glu Lys
35 40 45

Gly Ile Leu Val Gly Tyr Gln Asp Gly Thr Ile Arg Pro Asp Asn Asn
50 55 60

Ile Thr Arg Ala Glu Phe Val Thr Leu Ile Asn Lys Val Phe Gly Leu
65 70 75 80

Tyr Glu Leu Ser Arg Glu Gln Phe Ala Asp Val Glu Asp Ser Lys Trp
85 90 95

Tyr Ser Arg Glu Ile Leu Lys Ala Arg Ala Ala Gly Tyr Ile Ala Gly
100 105 110

Tyr Gly Ser Asn Val Phe Lys Pro Asp Asn Tyr Ile Thr Arg Gln Glu
115 120 125

Ala Val Val Ile Ile Ala Lys Val Phe Glu Leu Gln Ser Gly Ser Asn
130 135 140

Tyr Thr Ser Lys Phe Lys Asp Gly Ser Leu Val Lys Glu Tyr Ala Lys
145 150 155 160

Asp Ser Val Ser Ala Leu Val Glu Lys Gly Tyr Ile Ala Gly Tyr Glu
165 170 175

Asp Gly Thr Phe Arg Pro Asp Asn Tyr Ile Thr Arg Ala Glu Thr Ile
180 185 190

Lys Ile Leu Asn Lys Ile Ile Pro Ser Leu Tyr Asn Glu Lys Gly Asp
195 200 205

Tyr Lys Asn Glu Glu Val Ala Gly Asn Ala Leu Ile Asn Thr Glu Gly

210

215

220

Val Ile Leu Lys Asp Thr Val Ile Asn Gly Asp Leu Tyr Leu Ala Gln
225 230 235 240

Gly Ile Gln Asn Gly Asp Val Thr Leu Asp Gly Val Asn Val Lys Gly
245 250 255

Thr Val Phe Val Asn Gly Gly Gly Ser Asp Ser Ile His Phe Ile Asn
260 265 270

Thr Lys Ile Asn Arg Val Val Val Asn Lys Thr Gly Val Arg Ile Val
275 280 285

Thr Ser Gly Asn Thr Ser Val Glu Ser Val Val Val Lys Ser Gly Ala
290 295 300

Lys Leu Glu Glu Lys Glu Leu Thr Gly Asp Gly Phe Lys Asn Val Thr
305 310 315 320

Val Asp Ser Gln Leu Ser Ala Gly Asn Glu Ile Ile Phe Val Gly Asp
325 330 335

Phe Glu Gln Val Asp Val Leu Ala Asp Asp Ala Leu Leu Glu Thr Lys
340 345 350

Glu Ala Lys Met Lys Leu Arg Ile Phe Gly Gln Arg Ile Lys Val Asn
355 360 365

Gly Lys Ala Ile Glu Lys Ser Ser Lys Asn Tyr Ile Val Asn Gly Glu
370 375 380

Leu Ile Ser Thr Glu Glu Glu Pro Gly Pro Ser Asp Ala Pro Gly Ala
385 390 395 400

Glu Asp Asp Gln Asn Ser Gly Ser Pro Gly Ser Ser Thr Asn Pro Ala
405 410 415

Pro Thr Lys Asn Pro Asn Glu Glu Trp Arg Leu Val Trp Ser Asp Glu
420 425 430

Phe Asn Gly Ser Glu Ile Asn Met Ala Asn Trp Ser Tyr Asp Asp Pro
435 440 445

Thr Asn Gly Arg Trp Asn Gly Glu Val Gln Ser Tyr Thr Gln Asn Asn
450 455 460

Ala Tyr Ile Lys Asp Gly Ala Leu Val Ile Glu Ala Arg Lys Glu Asp
465 470 475 480

Ile Thr Glu Pro Ser Gly Glu Thr Tyr His Tyr Thr Ser Ser Lys Leu
485 490 495

Ile Thr Lys Gly Lys Lys Ser Trp Lys Tyr Gly Lys Phe Glu Ile Arg
500 505 510

Ala Lys Met Pro Gln Gly Gln Gly Ile Trp Pro Ala Ile Trp Met Met
515 520 525

Pro Glu Asp Glu Pro Phe Tyr Gly Thr Trp Pro Lys Cys Gly Glu Ile
530 535 540

Asp Ile Met Glu Leu Leu Gly His Glu Pro Asp Lys Ile Tyr Gly Thr
545 550 555 560

Ile His Phe Gly Glu Pro His Lys Glu Ser Gln Gly Thr Tyr Thr Leu
565 570 575

Pro Glu Gly Gln Thr Phe Ala Asp Asp Phe His Val Tyr Ser Ile Glu
580 585 590

Trp Glu Pro Gly Glu Ile Arg Trp Tyr Ile Asp Gly Lys Leu Tyr His
595 600 605

Val Ala Asn Asp Trp Tyr Ser Arg Asp Pro Tyr Leu Ala Asp Asp Tyr
610 615 620

Thr Tyr Pro Ala Pro Phe Asp Gln Asn Phe Phe Leu Ile Leu Asn Ile
625 630 635 640

Ser Val Gly Gly Gly Trp Pro Gly Tyr Pro Asp Glu Thr Thr Val Phe
645 650 655

Pro Gln Gln Met Val Val Asp Tyr Val Arg Val Tyr Gln Lys Asp Lys
660 665 670

Tyr Pro His Arg Glu Lys Pro Ala Lys Glu Glu Val Lys Pro Arg Glu
675 680 685

Pro Leu Glu Asp Gly Asn Tyr Ile Tyr Asn Gly Gly Phe Asp Val Asp
690 695 700

Asp Ser Ala Ala Val Gly Val Asp Gly Val Pro Tyr Thr Ser Tyr Trp
705 710 715 720

Thr Phe Leu Thr Ala Ser Gly Gly Ala Ala Thr Val Asn Val Glu Glu
725 730 735

Gly Val Met His Val Gln Ile Glu Asn Gly Gly Thr Thr Asp Tyr Gly
740 745 750

Val Gln Leu Leu Gln Ala Pro Ile His Leu Glu Lys Gly Ala Lys Tyr
755 760 765

Lys Ala Ser Phe Asp Met Lys Ala Glu Asn Pro Arg Gln Val Lys Leu
770 775 780

Lys Ile Gly Gly Asp Gly Asp Arg Gly Trp Lys Asp Tyr Ala Ala Ile
785 790 795 800

Pro Pro Phe Thr Val Ser Thr Glu Met Thr Asn Tyr Glu Phe Glu Phe
805 810 815

Thr Met Lys Asp Asp Thr Asp Val Lys Ala Arg Phe Glu Phe Asn Met
820 825 830

Gly Leu Asp Asp Asn Asp Val Trp Ile Asp Asn Val Lys Leu Ile Lys
835 840 845

Thr Glu Asp Ala Pro Val Ile Asp Pro Ser Glu Ile Ala Arg Pro Pro
850 855 860

Leu Leu Ser Gly Asn Tyr Ile Tyr Asn Gly Thr Phe Asp Gln Gly Pro
865 870 875 880

Asn Arg Met Gly Phe Trp Asn Phe Val Val Asp Ser Thr Ala Lys Ala
885 890 895

Thr Tyr Tyr Ile Gly Ser Asp Val Asn Glu Arg Arg Phe Glu Thr Arg
900 905 910

Ile Glu Lys Gly Gly Thr Ser Arg Gly Ala Ile Arg Leu Val Gln Pro
915 920 925

Gly Ile Asn Ile Glu Asn Gly Lys Thr Tyr Lys Val Ser Phe Glu Ala
930 935 940

Ser Ala Ala Asn Thr Arg Thr Ile Glu Val Glu Ile Ala Ser Asn Leu
945 950 955 960

His Asn Ser Ser Ile Phe Ala Thr Thr Phe Glu Ile Ser Lys Glu Ser
965 970 975

Lys Ile Tyr Glu Phe Glu Phe Thr Met Asp Lys Asp Ser Asp Lys Asn
980 985 990

Gly Glu Leu Arg Phe Asn Leu Gly Gly Ser Asn Val Asn Val Tyr Ile
995 1000 1005

Asp Asn Val Val Met Lys Arg Val Ser Thr Asp Glu Val Glu Gly
1010 1015 1020

Asn Leu Ile Leu Asn Gly Val Phe Asn Gly Leu Ala Gly Trp Gly
1025 1030 1035

Tyr Gly Ala Tyr Glu Pro Gly Ser Ala Asp Phe Glu Ser His Glu
1040 1045 1050

Glu Gln Phe Arg Ala Ile Ile Ser Ser Val Gly Asn Glu Gly Trp
1055 1060 1065

Asn Val Gln Leu Tyr Gln Asp Asn Val Pro Leu Glu Gln Gly Gln
1070 1075 1080

Thr Tyr Glu Val Ser Phe Asp Ala Lys Ser Thr Ile Asp Arg Lys
1085 1090 1095

Ile Ile Val Gln Leu Gln Arg Asn Gly Thr Ser Asp Asn Asn Trp
1100 1105 1110

Asp Ser Tyr Phe Tyr Gln Glu Val Glu Leu Thr Asn Glu Leu Lys
1115 1120 1125

Thr Phe Lys Tyr Glu Phe Thr Met Ser Lys Pro Thr Asp Ser Ala
1130 1135 1140

Ser Arg Phe Asn Phe Ala Leu Gly Asn Thr Glu Asn Lys Thr Tyr
1145 1150 1155

Ala Pro His Glu Ile Ile Ile Asp Asn Val Val Val Arg Lys Val

1160 1165 1170

Ala Thr Pro Ser Ala Leu Ile Leu Asn Gly Thr Phe Asp Asp Gly
1175 1180 1185

Met Asp His Trp Leu Leu Tyr Trp Gly Asp Gly Glu Gly Asn Cys
1190 1195 1200

Asp Val Thr Asp Gly Glu Leu Glu Ile Asn Ile Thr Lys Val Gly
1205 1210 1215

Thr Ala Asp Tyr Met Pro Gln Ile Lys Gln Glu Asn Ile Ala Leu
1220 1225 1230

Gln Glu Gly Val Thr Tyr Thr Leu Ser Leu Lys Ala Arg Ala Leu
1235 1240 1245

Glu Ala Arg Ser Ile Lys Val Asp Ile Leu Asp Ser Ser Tyr Asn
1250 1255 1260

Trp Tyr Gly Gly Thr Ile Phe Asp Leu Thr Thr Glu Asp Ala Val
1265 1270 1275

Tyr Thr Phe Thr Phe Thr Gln Ser Lys Ser Ile Asn Asn Gly Val
1280 1285 1290

Leu Thr Ile Asn Leu Gly Thr Ile Glu Gly Lys Thr Ser Ala Ala
1295 1300 1305

Thr Thr Val Tyr Leu Asp Asp Ile Leu Leu Glu Gln Gln
1310 1315 1320

<210> 25

<211> 1347

<212> DNA

<213> Clostridium thermocellum

<400> 25

atgtcaaaga taactttccc aaaagatttc atatgggggtt ctgcaacagc agcatatcag	60
attgaagggtg catacaacga agacggcaaa ggtgaatcta tatgggaccg tttttccac	120
acgccaggaa atatagcaga cggacatacc ggcgatgttg catgcgacca ctatcatcgt	180
tatgaagaag atatcaaat aatgaaagaa atcgggtatta aatcatacag gttttccatc	240
tcatggccca gaatctttcc tgaaggaaca ggtaaattaa atcaaaaggg actggatfff	300
tacaaaaggc tcacaaatct gtttctggaa aacggaatta tgcctgcaat cactctttat	360
cactgggacc ttccccaaaa gtttcaggat aaaggcggat ggaaaaaccg ggacaccacc	420
gattatfita cagaatactc tgaagtaata tttaaaaatc tggagatat cgttccaata	480
tggtttactc acaatgaacc cgggtgtgtt tctttgcttg gccactffff aggaattcat	540
gccccggga taaaagacct ccgcacttca ttggaagtct cgcacaatct tctttgtcc	600
cacggcaagg ccgtgaaact gtttagagaa atgaatatg acgcccacat tggaaatagct	660
ctcaatttat ctiaccatta tcccgcattc gaaaaagctg aggatattga agcagcggaa	720
ttgtcatttt ctctggcggg aagggtgtat ctggatcctg tgctaaaagg ccggtatcct	780
gaaaacgcat tgaacttta taaaaagaag ggtattgagc tttctttccc tgaagatgac	840
ctgaaactta tcagtcagcc aatagacttc atagcattca acaattattc ttcggaattt	900
ataaaatatg atccgtccag tgagtcaggf tttcacctg caaactccat attagaaaag	960
ttcgaaaaaa cagatatggg ctggatcata taccctgaag gcttgatga tctgcttatg	1020
ctccttgaca gggattatgg aaagccaaac attgttatca gcgaaaacgg agccgccttc	1080

aaagatgaaa taggtagcaa cggaagata gaagacacaa agagaatcca atatcttaaa 1140

gattatctga cccaggctca cagggaatt caggacgggt taaacttaaa agcatactac 1200

ttgtggtcgc ttttgacaa ctttgaatgg gcttacgggt acaacaagag attcggaatc 1260

gttcacgtaa attttgatac gttggaaaga aaaataaagg atagcggcta ctgtacaaa 1320

gaagtaatca aaaacaacgg tttttaa 1347

<210> 26
 <211> 448
 <212> PRT
 <213> Clostridium thermocellum

<220>
 <221> DOMAIN
 <222> (2)...(448)
 <223> Glycosyl hydrolase family 1

<220>
 <221> SITE
 <222> (10)...(24)
 <223> Glycosyl hydrolases family 1 N-terminal signature. Prosite id = PS00653

<220>
 <221> SITE
 <222> (225)...(228)
 <223> N-glycosylation site. Prosite id = PS00001

<220>
 <221> SITE
 <222> (299)...(302)
 <223> N-glycosylation site. Prosite id = PS00001

<220>
 <221> SITE
 <222> (356)...(364)
 <223> Glycosyl hydrolases family 1 active site. Prosite id = PS00572

<400> 26
 Met Ser Lys Ile Thr Phe Pro Lys Asp Phe Ile Trp Gly Ser Ala Thr
 1 5 10 15

Ala Ala Tyr Gln Ile Glu Gly Ala Tyr Asn Glu Asp Gly Lys Gly Glu
20 25 30

Ser Ile Trp Asp Arg Phe Ser His Thr Pro Gly Asn Ile Ala Asp Gly
35 40 45

His Thr Gly Asp Val Ala Cys Asp His Tyr His Arg Tyr Glu Glu Asp
50 55 60

Ile Lys Ile Met Lys Glu Ile Gly Ile Lys Ser Tyr Arg Phe Ser Ile
65 70 75 80

Ser Trp Pro Arg Ile Phe Pro Glu Gly Thr Gly Lys Leu Asn Gln Lys
85 90 95

Gly Leu Asp Phe Tyr Lys Arg Leu Thr Asn Leu Leu Leu Glu Asn Gly
100 105 110

Ile Met Pro Ala Ile Thr Leu Tyr His Trp Asp Leu Pro Gln Lys Leu
115 120 125

Gln Asp Lys Gly Gly Trp Lys Asn Arg Asp Thr Thr Asp Tyr Phe Thr
130 135 140

Glu Tyr Ser Glu Val Ile Phe Lys Asn Leu Gly Asp Ile Val Pro Ile
145 150 155 160

Trp Phe Thr His Asn Glu Pro Gly Val Val Ser Leu Leu Gly His Phe
165 170 175

Leu Gly Ile His Ala Pro Gly Ile Lys Asp Leu Arg Thr Ser Leu Glu
180 185 190

Val Ser His Asn Leu Leu Leu Ser His Gly Lys Ala Val Lys Leu Phe
195 200 205

Arg Glu Met Asn Ile Asp Ala Gln Ile Gly Ile Ala Leu Asn Leu Ser
210 215 220

Tyr His Tyr Pro Ala Ser Glu Lys Ala Glu Asp Ile Glu Ala Ala Glu
225 230 235 240

Leu Ser Phe Ser Leu Ala Gly Arg Trp Tyr Leu Asp Pro Val Leu Lys
245 250 255

Gly Arg Tyr Pro Glu Asn Ala Leu Lys Leu Tyr Lys Lys Lys Gly Ile
260 265 270

Glu Leu Ser Phe Pro Glu Asp Asp Leu Lys Leu Ile Ser Gln Pro Ile
275 280 285

Asp Phe Ile Ala Phe Asn Asn Tyr Ser Ser Glu Phe Ile Lys Tyr Asp
290 295 300

Pro Ser Ser Glu Ser Gly Phe Ser Pro Ala Asn Ser Ile Leu Glu Lys
305 310 315 320

Phe Glu Lys Thr Asp Met Gly Trp Ile Ile Tyr Pro Glu Gly Leu Tyr
325 330 335

Asp Leu Leu Met Leu Leu Asp Arg Asp Tyr Gly Lys Pro Asn Ile Val
340 345 350

Ile Ser Glu Asn Gly Ala Ala Phe Lys Asp Glu Ile Gly Ser Asn Gly
355 360 365

Lys Ile Glu Asp Thr Lys Arg Ile Gln Tyr Leu Lys Asp Tyr Leu Thr
370 375 380

Gln Ala His Arg Ala Ile Gln Asp Gly Val Asn Leu Lys Ala Tyr Tyr
385 390 395 400

Leu Trp Ser Leu Leu Asp Asn Phe Glu Trp Ala Tyr Gly Tyr Asn Lys

405

410

415

Arg Phe Gly Ile Val His Val Asn Phe Asp Thr Leu Glu Arg Lys Ile
420 425 430

Lys Asp Ser Gly Tyr Trp Tyr Lys Glu Val Ile Lys Asn Asn Gly Phe
435 440 445

<210> 27

<211> 1362

<212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 27

atggcaaaca agataacctt tcctgaaaat tttctgtggg gcgcggcaac ggcttcgtac 60

cagatcgaag gcgcctggaa caaacatggt aaaggcgaat ccacctggga tcgcttttca 120

cacacgcccg gtaagatcag gaacaacgat acgggcgatg tagcaaatga ccattatcgc 180

ctctgaaaaa aagacattgg cttgatgaag aagatcgggt tgaaggctta tcgattttcc 240

atttcgtggc cgcgtattct tcctgctgga agaggcaagg tcaatcaaag agggctggat 300

ttttacaaca agatcgtaga tgagctgctg aaagcagata tcatccatt tgttactctc 360

aatcactggg acctgcccca aaaactggaa gatgagggcg gctggccggc cgtttctact 420

gccgatgctt ttattgaata cacagatgtg atcacccgct cccttggcga ccgcgcaaag 480

aattggatca ctcaaatga acctgccgtc gttgcctgga tgggatactc cactggccaa 540

cacgcacccg gactgaagga ctatgggctt ggtgcccgcg ccgcgcatca cctgttgctc 600

tcacatggac aggtgtgacc ggtcattcgc agcaatagcg cgggggcaga agtgggaatt 660

acgctcgata ttagctggcg gatcgtgcc tcaaacagcc gcgccgaccg ggagctggtc 720

cgtgaggatg atgggagggtg gtccgctgg ttgccgacc cgctttacgg gcgcggatat 780

ccctccgata aggtgtctga ttactaag ttgggagcac tgcccaacgg acttgatatt 840

gtgcaggcag gcgacatgga cagcatcgcg acaccgactg attttatggg gctaaactac 900

tactcccgaa atgtctaccg cgcggacggt gcagataatg atccgcaaac tgttttccca 960

caaccgaaga tgcccgaaca ctggaccgag atgggctggg aaatttaccg ggatgggctg 1020

accaacattc tgggacgct ctatttcaac tatcagccgc gaaactata cgtcacagaa 1080

aacggcgcca gttactccac gcctcctgat gataagggga atgtcgcgga tgaactccgc 1140

atccattatc tggagacaca ttttgagct gcctatcggg ccattcaaatt gggcgtgcct 1200

ctggcaggat acttcgtctg gtccctcatg gacaactttg agtggctcatg gggctatatg 1260

caacgctttg gactcatctg ggtggattat gagaccctaa aacgcacttt aaaggatagc 1320

gcaaaatggt ataagcgct gatcaagaag aatgggctct aa 1362

<210> 28
 <211> 453
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> DOMAIN
 <222> (3)...(453)
 <223> Glycosyl hydrolase family 1

<220>
 <221> SITE
 <222> (11)...(25)
 <223> Glycosyl hydrolases family 1 N-terminal signature. Prosite id = PS00653

<220>
 <221> SITE
 <222> (49)...(52)
 <223> N-glycosylation site. Prosite id = PS00001

<220>
 <221> SITE
 <222> (361)...(369)
 <223> Glycosyl hydrolases family 1 active site. Prosite id = PS00572

<400> 28
 Met Ala Asn Lys Ile Thr Phe Pro Glu Asn Phe Leu Trp Gly Ala Ala
 1 5 10 15

Thr Ala Ser Tyr Gln Ile Glu Gly Ala Trp Asn Lys His Gly Lys Gly
 20 25 30

Glu Ser Thr Trp Asp Arg Phe Ser His Thr Pro Gly Lys Ile Arg Asn
 35 40 45

Asn Asp Thr Gly Asp Val Ala Asn Asp His Tyr Arg Leu Trp Lys Lys
 50 55 60

Asp Ile Gly Leu Met Lys Lys Ile Gly Leu Lys Ala Tyr Arg Phe Ser
 65 70 75 80

Ile Ser Trp Pro Arg Ile Leu Pro Ala Gly Arg Gly Lys Val Asn Gln
 85 90 95

Arg Gly Leu Asp Phe Tyr Asn Lys Ile Val Asp Glu Leu Leu Lys Ala
 100 105 110

Asp Ile Ile Pro Phe Val Thr Leu Asn His Trp Asp Leu Pro Gln Lys
 115 120 125

Leu Glu Asp Glu Gly Gly Trp Pro Ala Arg Ser Thr Ala Asp Ala Phe
 130 135 140

Ile Glu Tyr Thr Asp Val Ile Thr Arg Ser Leu Gly Asp Arg Ala Lys
145 150 155 160

Asn Trp Ile Thr His Asn Glu Pro Ala Val Val Ala Trp Met Gly Tyr
165 170 175

Ser Thr Gly Gln His Ala Pro Gly Leu Lys Asp Tyr Gly Leu Gly Ala
180 185 190

Arg Ala Ala His His Leu Leu Leu Ser His Gly Gln Ala Val Pro Val
195 200 205

Ile Arg Ser Asn Ser Ala Gly Ala Glu Val Gly Ile Thr Leu Asp Ile
210 215 220

Ser Trp Arg Ile Ala Ala Ser Asn Ser Arg Ala Asp Arg Glu Leu Val
225 230 235 240

Arg Glu Asp Asp Gly Arg Trp Phe Arg Trp Phe Ala Asp Pro Leu Tyr
245 250 255

Gly Arg Gly Tyr Pro Ser Asp Lys Val Ser Asp Phe Thr Lys Leu Gly
260 265 270

Ala Leu Pro Asn Gly Leu Asp Phe Val Gln Ala Gly Asp Met Asp Thr
275 280 285

Ile Ala Thr Pro Thr Asp Phe Met Gly Leu Asn Tyr Tyr Ser Arg Asn
290 295 300

Val Tyr Arg Ala Asp Gly Ala Asp Asn Asp Pro Gln Thr Val Phe Pro
305 310 315 320

Gln Pro Lys Met Pro Glu His Trp Thr Glu Met Gly Trp Glu Ile Tyr
325 330 335

Pro Asp Gly Leu Thr Asn Ile Leu Gly Arg Val Tyr Phe Asn Tyr Gln

340 345 350

Pro Arg Lys Leu Tyr Val Thr Glu Asn Gly Ala Ser Tyr Ser Thr Pro
355 360 365

Pro Asp Asp Lys Gly Asn Val Ala Asp Glu Leu Arg Ile His Tyr Leu
370 375 380

Arg Thr His Phe Ala Ala Ala Tyr Arg Ala Ile Gln Met Gly Val Pro
385 390 395 400

Leu Ala Gly Tyr Phe Val Trp Ser Leu Met Asp Asn Phe Glu Trp Ser
405 410 415

Trp Gly Tyr Met Gln Arg Phe Gly Leu Ile Trp Val Asp Tyr Glu Thr
420 425 430

Gln Lys Arg Thr Leu Lys Asp Ser Ala Lys Trp Tyr Lys Arg Val Ile
435 440 445

Lys Lys Asn Gly Leu
450

<210> 29
<211> 1362
<212> DNA
<213> Unknown

<220>
<223> Obtained from environmental sample

<400> 29
atggcgaaca aaattacctt tccgaaaat ttcttttggg gcgcggcaac agcctcctac 60

cagatcgaag gtgcgtggga caaacatggc aagggtgaat ccatctggga tcgcttttcg 120

cataccctg gcaagatcag aaataatgat acgggcgatg ttgccaatga tcattatcgt 180

ctctggaaaa aagacattgg ctgatgaag aagatcggct tgaaggcata tcgtttttcc	240
atttcgtggc cgcgtgttct tcccgccgga cgcggcaaag tcaatcagaa gggactggat	300
ttctataaca ggctggtaga tgctctgttg aaagaagata tcatccatt tgtgactctc	360
aatcactggg acctgcccc aaagctggag gaggaaggcg gttggccggt tcgctccacc	420
gcagatgcct ttgtggaata cacagacgtg gtcacacgtt ccctcggcga ccgcgtaaag	480
aattggatca cgcataatga gcctgccgtc gttgcctgga tgggatattc cacaggtcaa	540
cacgcacccg gtttgaagga ctatgggctt ggtgtgcgcg ccgcgcatca tctgctgctc	600
tcccacgggc aggcgggtgcc agtcatccgc agtaacagcg ccgatgcaga agtgggcatt	660
acgttgata ttagctggcg gattcctgcc tccaatagcc gagcagaccg ggaattggtc	720
cgtaaagatg acggactatg gttccgctgg ttcgccgac cgctttatgg gcgcggatac	780
ccctcgata aagtcaccga ttttaciaag atcggcgcgc tgccaatgg tctggacttt	840
atgcaagccg gtgatatgga tgcgatgcc acgccaaccg atttcatggg gctgaactat	900
tatttccgaa atgtctaccg cgcgaatggc gaagacaatg atccgcaggt cgttttccca	960
caaccaaaga tgccgaaca ctggacggag atgggctggg aaatctatcc ggatggactg	1020
acgaacatcc tgggacgct tttttcaat taccagccac ataaactgta taccacagag	1080
aacggcgcga gctactccac cccgccgat gaaaagggga atgtcgccga tgagctccgc	1140
actcattatt tacggacaca cttcgaggct gcctaccggg cgattcagat gggcgtgcct	1200
ctggcaggat actttgtctg gtccctcatg gacaactttg agtggctctg gggatatatg	1260
cagcgtttg ggctcatctg ggtggactac gagacacaga aacgcaccct gaaggatagc	1320
gccaagtggg acaaacgtgt gatcaggaag aatgggtttt ag	1362

<210> 30
 <211> 453
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> DOMAIN
 <222> (3)...(453)
 <223> Glycosyl hydrolase family 1

<220>
 <221> SITE
 <222> (11)...(25)
 <223> Glycosyl hydrolases family 1 N-terminal signature. Prosite id = PS00653

<220>
 <221> SITE
 <222> (49)...(52)
 <223> N-glycosylation site. Prosite id = PS00001

<220>
 <221> SITE
 <222> (361)...(369)
 <223> Glycosyl hydrolases family 1 active site. Prosite id = PS00572

<400> 30
 Met Ala Asn Lys Ile Thr Phe Pro Glu Asn Phe Leu Trp Gly Ala Ala
 1 5 10 15

Thr Ala Ser Tyr Gln Ile Glu Gly Ala Trp Asp Lys His Gly Lys Gly
 20 25 30

Glu Ser Ile Trp Asp Arg Phe Ser His Thr Pro Gly Lys Ile Arg Asn
 35 40 45

Asn Asp Thr Gly Asp Val Ala Asn Asp His Tyr Arg Leu Trp Lys Lys
 50 55 60

Asp Ile Gly Leu Met Lys Lys Ile Gly Leu Lys Ala Tyr Arg Phe Ser
65 70 75 80

Ile Ser Trp Pro Arg Val Leu Pro Ala Gly Arg Gly Lys Val Asn Gln
85 90 95

Lys Gly Leu Asp Phe Tyr Asn Arg Leu Val Asp Ala Leu Leu Lys Glu
100 105 110

Asp Ile Ile Pro Phe Val Thr Leu Asn His Trp Asp Leu Pro Gln Lys
115 120 125

Leu Glu Glu Glu Gly Gly Trp Pro Val Arg Ser Thr Ala Asp Ala Phe
130 135 140

Val Glu Tyr Thr Asp Val Val Thr Arg Ser Leu Gly Asp Arg Val Lys
145 150 155 160

Asn Trp Ile Thr His Asn Glu Pro Ala Val Val Ala Trp Met Gly Tyr
165 170 175

Ser Thr Gly Gln His Ala Pro Gly Leu Lys Asp Tyr Gly Leu Gly Val
180 185 190

Arg Ala Ala His His Leu Leu Leu Ser His Gly Gln Ala Val Pro Val
195 200 205

Ile Arg Ser Asn Ser Ala Asp Ala Glu Val Gly Ile Thr Leu Asp Ile
210 215 220

Ser Trp Arg Ile Pro Ala Ser Asn Ser Arg Ala Asp Arg Glu Leu Val
225 230 235 240

Arg Lys Asp Asp Gly Leu Trp Phe Arg Trp Phe Ala Asp Pro Leu Tyr
245 250 255

Gly Arg Gly Tyr Pro Ser Asp Lys Val Thr Asp Phe Thr Lys Ile Gly

260 265 270

Ala Leu Pro Asn Gly Leu Asp Phe Met Gln Ala Gly Asp Met Asp Ala
275 280 285

Ile Ala Thr Pro Thr Asp Phe Met Gly Leu Asn Tyr Tyr Phe Arg Asn
290 295 300

Val Tyr Arg Ala Asn Gly Glu Asp Asn Asp Pro Gln Val Val Phe Pro
305 310 315 320

Gln Pro Lys Met Pro Glu His Trp Thr Glu Met Gly Trp Glu Ile Tyr
325 330 335

Pro Asp Gly Leu Thr Asn Ile Leu Gly Arg Val Tyr Phe Asn Tyr Gln
340 345 350

Pro His Lys Leu Tyr Ile Thr Glu Asn Gly Ala Ser Tyr Ser Thr Pro
355 360 365

Pro Asp Glu Lys Gly Asn Val Ala Asp Glu Leu Arg Thr His Tyr Leu
370 375 380

Arg Thr His Phe Ala Ala Ala Tyr Arg Ala Ile Gln Met Gly Val Pro
385 390 395 400

Leu Ala Gly Tyr Phe Val Trp Ser Leu Met Asp Asn Phe Glu Trp Ser
405 410 415

Trp Gly Tyr Met Gln Arg Phe Gly Leu Ile Trp Val Asp Tyr Glu Thr
420 425 430

Gln Lys Arg Thr Leu Lys Asp Ser Ala Lys Trp Tyr Lys Arg Val Ile
435 440 445

Arg Lys Asn Gly Phe

450

<210> 31
<211> 1167
<212> DNA
<213> Unknown

<220>
<223> Obtained from environmental sample

<400> 31
atggaagacc gccgcacta ttacagcgac gaccatctct ggggtgtact gtgcgtgacc 60

gcctacatca aggaaactgg ggactttgca ttcttgacg agaaagtta cttttacgag 120

aaggaccggg tcgagggcgt gtctgtgctg gatcacgtta aacgggcctt gacctttacc 180

cgcaacaaca tcgggaaaca tggctgcct ctctcggct ttgcggattg gaacgacacg 240

atcaatctgg cgaagggcgc cgagtctctt ttcacgtcgc atctatatgg acgcgcgctg 300

ctggagttta ttgatctgct cacatatctt ggcaagaacg atgaagccga tgaatggcag 360

cgagcccacg ttgagatgca gtcccgctc gaaaaacatg cctgggatgg cgaatggtat 420

ttcatgtact ttgaccacga cggcagcccg gttgggtctc acacgaatca gtatggaaag 480

atccatctca acggacagag ctgggctgtg ctttcgggct ttgcctctcc gcagcgcgcc 540

cgccaggcca tggactcggg ttacaagcat ctcaacacaa agcacggcat caagctctcc 600

acgccgggct acaatggcta tgacccaac tacggcggcg tgaccaccta cccaccggga 660

gcaaaggaaa acggcggcat cttctgcac ccgaatccct gggccatgat cgcagagacc 720

atgctcgggg atggcgatcg cgcctacgag tattactcgc agatcaaccc ggccggcaag 780

aacgatgaca tcgacctgta cgaggtcgag ccatatgttt acgctcaaaa catcctgggc 840

gatgagcatc cgagttcgg gctgggacgc aactcgtggc tctcgggtac ggcattcctgg 900

tgctatcagg ctgccacaca gtggatcctc ggaatccgcg ccgactatga agggctgcgc 960

atcgaccctg gcattccgtc caagtgggat gggttcaagg caacgcgcct gtatcgcggc 1020

gtgaagtaca acattacggt caccaaccgc aagcacatct gcaaaggcgt ggaaaaagtt 1080

ctggtcaacg gcaaaccggt tgaggggaat gtggtccggg cagacgtggg ttgctgcgaa 1140

gtgaacgtgg aagttacctt aggataa 1167

<210> 32
 <211> 388
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> SITE
 <222> (79)...(82)
 <223> N-glycosylation site. Prosite id = PS00001

<220>
 <221> SITE
 <222> (349)...(352)
 <223> N-glycosylation site. Prosite id = PS00001

<400> 32
 Met Glu Asp Arg Pro His Tyr Tyr Ser Asp Asp His Leu Trp Gly Val
 1 5 10 15

Leu Cys Val Thr Ala Tyr Ile Lys Glu Thr Gly Asp Phe Ala Phe Leu
 20 25 30

Asp Glu Lys Val His Phe Tyr Glu Lys Asp Pro Val Glu Gly Val Ser
 35 40 45

Val Leu Asp His Val Lys Arg Ala Leu Thr Phe Thr Arg Asn Asn Ile

50

55

60

Gly Lys His Gly Leu Pro Leu Leu Gly Phe Ala Asp Trp Asn Asp Thr
65 70 75 80

Ile Asn Leu Ala Lys Gly Ala Glu Ser Leu Phe Thr Ser His Leu Tyr
85 90 95

Gly Arg Ala Leu Leu Glu Phe Ile Asp Leu Leu Thr Tyr Leu Gly Lys
100 105 110

Asn Asp Glu Ala Asp Glu Trp Gln Arg Ala His Val Glu Met Gln Ser
115 120 125

Arg Val Glu Lys His Ala Trp Asp Gly Glu Trp Tyr Phe Met Tyr Phe
130 135 140

Asp His Asp Gly Ser Pro Val Gly Ser His Thr Asn Gln Tyr Gly Lys
145 150 155 160

Ile His Leu Asn Gly Gln Ser Trp Ala Val Leu Ser Gly Phe Ala Ser
165 170 175

Pro Gln Arg Ala Arg Gln Ala Met Asp Ser Val Tyr Lys His Leu Asn
180 185 190

Thr Lys His Gly Ile Lys Leu Ser Thr Pro Gly Tyr Asn Gly Tyr Asp
195 200 205

Pro Asn Tyr Gly Gly Val Thr Thr Tyr Pro Pro Gly Ala Lys Glu Asn
210 215 220

Gly Gly Ile Phe Leu His Pro Asn Pro Trp Ala Met Ile Ala Glu Thr
225 230 235 240

Met Leu Gly Asp Gly Asp Arg Ala Tyr Glu Tyr Tyr Ser Gln Ile Asn

245

250

255

Pro Ala Gly Lys Asn Asp Asp Ile Asp Leu Tyr Glu Val Glu Pro Tyr
260 265 270

Val Tyr Ala Gln Asn Ile Leu Gly Asp Glu His Pro Gln Phe Gly Leu
275 280 285

Gly Arg Asn Ser Trp Leu Ser Gly Thr Ala Ser Trp Cys Tyr Gln Ala
290 295 300

Ala Thr Gln Trp Ile Leu Gly Ile Arg Ala Asp Tyr Glu Gly Leu Arg
305 310 315 320

Ile Asp Pro Cys Ile Pro Ser Lys Trp Asp Gly Phe Lys Ala Thr Arg
325 330 335

Leu Tyr Arg Gly Val Lys Tyr Asn Ile Thr Val Thr Asn Pro Lys His
340 345 350

Ile Cys Lys Gly Val Glu Lys Val Leu Val Asn Gly Lys Pro Val Glu
355 360 365

Gly Asn Val Val Arg Ala Asp Val Gly Leu Arg Glu Val Asn Val Glu
370 375 380

Val Thr Leu Gly
385

<210> 33

<211> 1362

<212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 33

atggcaata aaattctctt ccccgagaac tttctctggg gcacggcgac cgcattctac	60
cagatcgagg gggcttgga taaacatggt aaggcgagt cgacctgga ccgttttacg	120
catacacctg gaaagatcaa aaacaatgat acgggcgatg tagcagatga ccattatcga	180
ttatgaaaa aagatatcgg cttagatgaag aagctcggct tgaaggctta tcgtttttcg	240
acttcctggc cgcgggtgct gccggccggg cgcggtaaga gcaatcaaaa aggactcgat	300
ttctacagca agctggttga tgagtgtcta aaagcaaata tcattccatt cgtgacattg	360
aatcactggg acatcccaca aaagtggag gacgagggtg gctgggccgt gcgtcaacg	420
gctgaggcat ttgtggaata tgccgatctc atgtcgcga cgcttgaga ccgcgtcaag	480
aactggatca cgcacaacga accggccgtc gtcgcctgga tgggatacgg gatgggcac	540
cacgcgccgg gcttaacgga tttctcgatt gcggtgccgg tctcgcac	600
tcgcacggat ggcccggtcc tgtgattcgc ggtaacagcc cgcatgccga ggtgggcatt	660
accctcaaca ttcaatgggg cgaagcagca tccaacagcc ggcccgacct aaacgccctg	720
cgctgaacg atggacagtg gttccgctgg ttgccgac cggtttatgg ccgcgctat	780
ccttcgcagc tgggtgctga tttcgagaaa atgggcgcgc tgccgaacgg catgaatttc	840
gtgcaacctg gcgatatgga tgtcatgcc acgccaaccg atttcctcgg gctcaattat	900
tattccgcc atgtgcatcg cgtcaacaca ccggataacg atcaacaggt tgtgtttgcc	960
aaacagcagg gtcccagaaa ctggaccgag atgggctggg agatccatcc tgatggattg	1020
gccggaattt tatccagagc gtatttcaat taccagccgc gcaaagtata tgtgactgaa	1080
aacggtgcc gctattccac cgcgcccgat gagaatggtt ttgtcaacga cattcaccgc	1140

gtcaattatc tacggacgca cttcgggct gcccatcgcg ccctgcaggc gggcgtgcca 1200

ttggcaggat acttcgtctg gtcaatgctc gataacttcg aatggagtca cgggtacagc 1260

cagcgctttg gcatcgttta tgtggactat caaaccaga agcgttactt gaaagacagc 1320

gccaagtggg acaaagggtg catcaaaaag aatgggttct aa 1362

<210> 34
 <211> 453
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> DOMAIN
 <222> (3)...(453)
 <223> Glycosyl hydrolase family 1

<220>
 <221> SITE
 <222> (11)...(25)
 <223> Glycosyl hydrolases family 1 N-terminal signature. Prosite id = PS00653

<220>
 <221> SITE
 <222> (49)...(52)
 <223> N-glycosylation site. Prosite id = PS00001

<220>
 <221> SITE
 <222> (332)...(335)
 <223> N-glycosylation site. Prosite id = PS00001

<220>
 <221> SITE
 <222> (361)...(369)
 <223> Glycosyl hydrolases family 1 active site. Prosite id = PS00572

<400> 34
 Met Ala Asn Lys Ile Leu Phe Pro Glu Asn Phe Leu Trp Gly Thr Ala
 1 5 10 15

Thr Ala Ser Tyr Gln Ile Glu Gly Ala Trp Asp Lys His Gly Lys Gly
20 25 30

Glu Ser Thr Trp Asp Arg Phe Thr His Thr Pro Gly Lys Ile Lys Asn
35 40 45

Asn Asp Thr Gly Asp Val Ala Asp Asp His Tyr Arg Leu Trp Lys Lys
50 55 60

Asp Ile Gly Leu Met Lys Lys Leu Gly Leu Lys Ala Tyr Arg Phe Ser
65 70 75 80

Thr Ser Trp Pro Arg Val Leu Pro Ala Gly Arg Gly Lys Ser Asn Gln
85 90 95

Lys Gly Leu Asp Phe Tyr Ser Lys Leu Val Asp Glu Leu Leu Lys Ala
100 105 110

Asn Ile Ile Pro Phe Val Thr Leu Asn His Trp Asp Ile Pro Gln Lys
115 120 125

Leu Glu Asp Glu Gly Gly Trp Ala Val Arg Ser Thr Ala Glu Ala Phe
130 135 140

Val Glu Tyr Ala Asp Leu Met Ser Arg Thr Leu Gly Asp Arg Val Lys
145 150 155 160

Asn Trp Ile Thr His Asn Glu Pro Ala Val Val Ala Trp Met Gly Tyr
165 170 175

Gly Met Gly Ile His Ala Pro Gly Leu Thr Asp Phe Ser Ile Ala Val
180 185 190

Pro Val Ser His His Leu Leu Leu Ser His Gly Trp Ala Val Pro Val
195 200 205

Ile Arg Gly Asn Ser Pro Asp Ala Glu Val Gly Ile Thr Leu Asn Ile
210 215 220

Gln Trp Gly Glu Ala Ala Ser Asn Ser Arg Ala Asp Leu Asn Ala Leu
225 230 235 240

Arg Leu Asn Asp Gly Gln Trp Phe Arg Trp Phe Ala Asp Pro Val Tyr
245 250 255

Gly Arg Gly Tyr Pro Ser Asp Val Val Ala Asp Phe Glu Lys Met Gly
260 265 270

Ala Leu Pro Asn Gly Met Asn Phe Val Gln Pro Gly Asp Met Asp Val
275 280 285

Ile Ala Thr Pro Thr Asp Phe Leu Gly Leu Asn Tyr Tyr Ser Arg His
290 295 300

Val His Arg Val Asn Thr Pro Asp Asn Asp Gln Gln Val Val Phe Ala
305 310 315 320

Lys Gln Gln Gly Pro Glu Asn Trp Thr Glu Met Gly Trp Glu Ile His
325 330 335

Pro Asp Gly Leu Ala Gly Ile Leu Ser Arg Ala Tyr Phe Asn Tyr Gln
340 345 350

Pro Arg Lys Val Tyr Val Thr Glu Asn Gly Ala Ser Tyr Ser Thr Ala
355 360 365

Pro Asp Glu Asn Gly Ile Val Asn Asp Ile His Arg Val Asn Tyr Leu
370 375 380

Arg Thr His Phe Ala Ala Ala His Arg Ala Leu Gln Ala Gly Val Pro
385 390 395 400

Leu Ala Gly Tyr Phe Val Trp Ser Met Leu Asp Asn Phe Glu Trp Ser
405 410 415

His Gly Tyr Ser Gln Arg Phe Gly Ile Val Tyr Val Asp Tyr Gln Thr
420 425 430

Gln Lys Arg Tyr Leu Lys Asp Ser Ala Lys Trp Tyr Lys Gly Val Ile
435 440 445

Lys Lys Asn Gly Phe
450

<210> 35
<211> 1116
<212> DNA
<213> Unknown

<220>
<223> Obtained from environmental sample

<400> 35
atgaataaaa tctcaaaact cttcagcagc ctgctgcttt ttgcaggcat ctgtcccgcg 60

cttcaggcag agccagtaga aacctacttt ccctgtccc gcgggatcaa catgagccac 120

tggtctcttc aagtgaatga aaacattccc gaccgttcca cctatgtgac ggagcgggat 180

ttgcaatttc tgcgggcagc cggtttcgac catgtgctgc tgccaatcga tgaggtcgaa 240

ctctgggatg aagagggcaa tcagatcgag gaggcctggc aatacatgca taactttctc 300

cgttggagcc gaaagaacga tctccgggtc attctcgacc tgcacacggt attgtccac 360

cattcaacg cggtaaatat gggagaggtc aatacactct tcaatgatcc cagggaacag 420

gaaaagttcc tcaacctatg ggaacaaatc atggatgccg tgggtcacca tccgaatgag 480

tttctcgct atgaaatgct caatgaggcg gtcgcggaag atgatgaaga ctggaatctg 540

ctctcaacc gcgccattgt ccgcatccgg gaccgtgagc cttatcgggt gctgattgcg 600

gggtcgaact ggiggcagca tgccgaccgg gtccccaacc tgaggctccc gaaaggagac 660

ccaatatca tcatacagttt tcatttttat tccccitttc tcttaccca ctaccgcagt 720

agctggactg cgatgcaggc gtaccagggc ttcgtccaat accctggcaa aaccatacct 780

tccatacatc tcgaaggcat gaactaccg gagtccttcg ttcataatgtg ggaagcgac 840

aatcggtact atgacatcca ttccatgtat gccgaaatgg tcccggcggg gcgttttgcc 900

gaaaagttag gacttcggct ctattgcgga gaattcgggg ccatgaagac cgttgatcgc 960

gcccagatgc tgcagtggta tcgggatgtt gtcactgtat ttaataaatt gggtattccc 1020

tatactgcct gggattatca gggaaccttc ggaatccgag atgagctgac cggtagagccc 1080

gatcatgaaa tgatcgatat tctctcggg cgtga 1116

<210> 36
<211> 371
<212> PRT
<213> Unknown

<220>
<223> Obtained from environmental sample

<220>
<221> SIGNAL
<222> (1)...(23)

<220>
<221> DOMAIN
<222> (39)...(350)
<223> Cellulase (glycosyl hydrolase family 5)

<220>
<221> SITE
<222> (37)...(40)

<223> N-glycosylation site. Prosite id = PS00001

<400> 36

Met Asn Lys Ile Leu Lys Leu Phe Ser Ser Leu Leu Leu Phe Ala Gly
1 5 10 15

Ile Cys Pro Ala Leu Gln Ala Glu Pro Val Glu Thr Tyr Phe Pro Leu
20 25 30

Ser Arg Gly Ile Asn Met Ser His Trp Leu Ser Gln Val Asn Glu Asn
35 40 45

Ile Pro Asp Arg Ser Thr Tyr Val Thr Glu Arg Asp Leu Gln Phe Leu
50 55 60

Arg Ala Ala Gly Phe Asp His Val Arg Leu Pro Ile Asp Glu Val Glu
65 70 75 80

Leu Trp Asp Glu Glu Gly Asn Gln Ile Glu Glu Ala Trp Gln Tyr Met
85 90 95

His Asn Phe Leu Arg Trp Ser Arg Lys Asn Asp Leu Arg Val Ile Leu
100 105 110

Asp Leu His Thr Val Leu Ser His His Phe Asn Ala Val Asn Met Gly
115 120 125

Glu Val Asn Thr Leu Phe Asn Asp Pro Arg Glu Gln Glu Lys Phe Leu
130 135 140

Asn Leu Trp Glu Gln Ile Met Asp Ala Val Gly His His Pro Asn Glu
145 150 155 160

Phe Leu Ala Tyr Glu Met Leu Asn Glu Ala Val Ala Glu Asp Asp Glu
165 170 175

Asp Trp Asn Leu Leu Leu Asn Arg Ala Ile Val Arg Ile Arg Asp Arg
180 185 190

Glu Pro Tyr Arg Val Leu Ile Ala Gly Ser Asn Trp Trp Gln His Ala
195 200 205

Asp Arg Val Pro Asn Leu Arg Leu Pro Lys Gly Asp Pro Asn Ile Ile
210 215 220

Ile Ser Phe His Phe Tyr Ser Pro Phe Leu Phe Thr His Tyr Arg Ser
225 230 235 240

Ser Trp Thr Ala Met Gln Ala Tyr Gln Gly Phe Val Gln Tyr Pro Gly
245 250 255

Lys Thr Ile Pro Ser Ile His Leu Glu Gly Met Asn Tyr Pro Glu Ser
260 265 270

Phe Val His Met Trp Glu Ala His Asn Arg Tyr Tyr Asp Ile His Ser
275 280 285

Met Tyr Ala Glu Met Val Pro Ala Val Arg Phe Ala Glu Lys Leu Gly
290 295 300

Leu Arg Leu Tyr Cys Gly Glu Phe Gly Ala Met Lys Thr Val Asp Arg
305 310 315 320

Ala Gln Met Leu Gln Trp Tyr Arg Asp Val Val Thr Val Phe Asn Lys
325 330 335

Leu Gly Ile Pro Tyr Thr Ala Trp Asp Tyr Gln Gly Thr Phe Gly Ile
340 345 350

Arg Asp Glu Leu Thr Gly Glu Pro Asp His Glu Met Ile Asp Ile Leu
355 360 365

Leu Gly Arg

370

<210> 37
 <211> 1383
 <212> DNA
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<400> 37
 atgagcaaac tcccaaat cctctttgga gccggcacct caagtatca gatcgaaggt 60

 gcctggaata tagatggcaa aggtccctcc atttgggatt tccacactcg ccatcccggc 120

 gcggtttatc ggatgcacaa cggggatatg gcctgcgac attatcatcg gtatcgaacg 180

 gatatcgagc tgatgcagaa gatcggccta gaggcttacc gcttttccat aaactggccc 240

 cgggtttctgc cggaaggagc cgggtccgcc aatgaagcag gtctggactt ttacgaccgg 300

 ctggtggacg cactgttga agcggaatt cagccttga tcacccttta tcaactggaa 360

 ctccctggg ctctccact gcgcgggggt tggtcaatc gggacatgcc cgaccacatt 420

 gagaactacg ccgcttgggt cgccaggtgc ctcggtgacc gggtagaaaa ctggattact 480

 ttgaatgagc ctcaagtttt catcgggctt ggctatgcca gcgggggtca tgccccggc 540

 tataagttgt ccttgcggga gtgcctggc ggttcccacc atgccgtgct tccccacc 600

 cgggcagtca aggcgatccg ggccaactgc gaaggcagcg tccagatcgg ctacgccccg 660

 gtgggtgttg tctgccgacc ggaaacggag tcggcagcag acattgagc tgccccccag 720

 gccacctacc atatcaacac tccagcacc cacactcccg acaatctgat cggtgcctc 780

 tggaacagca cttggtggat agatccaatg gttctgggga agtatccga acacgggctg 840

aaagcctttg aaagctatct gccggacaac attcaggccg aactggatgc cgtattcgaa 900

ccgacggact ttgtcggttc caacatctac cacggccgca cggtcgggc caagcaggat 960

ggtggttttg agtttatcga cttccgccc ggagccccc gcaccacat gggctgggac 1020

atcaccccg acatcctcta ctggggagga aagtatcttt acgaacgcta tggcaagccg 1080

atgtttatca cggaaaacgg cattgccgtc ccggaactgg tgaatgatga aggccaggtc 1140

gaggataccg tccgtgagca ataatgaag ctgcacctgc gtgggctgca gcgggcccgc 1200

gatgaaggca tcccctatgc cggatacttc cactggctcc tgctcgacaa cttcagtggt 1260

gaacaaggct actcccagcg ctttggcatg gtctacgtcg actaccagac ccaggaacgt 1320

atcctcaaac gttcgggcca gcatttcgct gccatcgcc gggaatcac cggaaccgcc 1380

taa 1383

<210> 38
 <211> 460
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> DOMAIN
 <222> (1)...(458)
 <223> Glycosyl hydrolase family 1

<220>
 <221> SITE
 <222> (7)...(21)
 <223> Glycosyl hydrolases family 1 N-terminal signature. Prosite id = PS00653

<220>
 <221> SITE
 <222> (266)...(269)
 <223> N-glycosylation site. Prosite id = PS00001

<220>

<221> SITE

<222> (366)...(374)

<223> Glycosyl hydrolases family 1 active site. Prosite id = PS00572

<400> 38

Met Ser Lys Leu Pro Lys Phe Leu Phe Gly Ala Gly Thr Ser Ser Tyr
1 5 10 15

Gln Ile Glu Gly Ala Trp Asn Ile Asp Gly Lys Gly Pro Ser Ile Trp
20 25 30

Asp Phe His Thr Arg His Pro Gly Ala Val Tyr Arg Met His Asn Gly
35 40 45

Asp Met Ala Cys Asp His Tyr His Arg Tyr Arg Thr Asp Ile Glu Leu
50 55 60

Met Gln Lys Ile Gly Leu Glu Ala Tyr Arg Phe Ser Ile Asn Trp Pro
65 70 75 80

Arg Val Leu Pro Glu Gly Thr Gly Ala Ala Asn Glu Ala Gly Leu Asp
85 90 95

Phe Tyr Asp Arg Leu Val Asp Ala Leu Leu Glu Ala Gly Ile Gln Pro
100 105 110

Trp Ile Thr Leu Tyr His Trp Glu Leu Pro Trp Ala Leu His Leu Arg
115 120 125

Gly Gly Trp Leu Asn Arg Asp Met Pro Asp His Ile Glu Asn Tyr Ala
130 135 140

Ala Leu Val Ala Arg Cys Leu Gly Asp Arg Val Lys Asn Trp Ile Thr
145 150 155 160

Leu Asn Glu Pro Gln Val Phe Ile Gly Leu Gly Tyr Ala Ser Gly Val
165 170 175

His Ala Pro Gly Tyr Lys Leu Ser Leu Arg Glu Cys Leu Val Gly Ser
180 185 190

His His Ala Val Leu Ser His His Arg Ala Val Lys Ala Ile Arg Ala
195 200 205

Asn Cys Glu Gly Ser Val Gln Ile Gly Ser Ala Pro Val Gly Val Val
210 215 220

Cys Arg Pro Glu Thr Glu Ser Ala Ala Asp Ile Glu Ala Ala Arg Gln
225 230 235 240

Ala Thr Tyr His Ile Asn Thr Pro Ser Thr His Thr Pro Asp Asn Leu
245 250 255

Ile Gly Cys Leu Trp Asn Ser Thr Trp Trp Ile Asp Pro Met Val Leu
260 265 270

Gly Lys Tyr Pro Glu His Gly Leu Lys Ala Phe Glu Ser Tyr Leu Pro
275 280 285

Asp Asn Ile Gln Ala Glu Leu Asp Ala Val Phe Glu Pro Thr Asp Phe
290 295 300

Val Gly Ser Asn Ile Tyr His Gly Arg Thr Val Arg Ala Lys Gln Asp
305 310 315 320

Gly Gly Phe Glu Phe Ile Asp Leu Pro Pro Gly Ser Pro Arg Thr Thr
325 330 335

Met Gly Trp Asp Ile Thr Pro Asp Ile Leu Tyr Trp Gly Gly Lys Tyr
340 345 350

Leu Tyr Glu Arg Tyr Gly Lys Pro Met Phe Ile Thr Glu Asn Gly Ile

355

360

365

Ala Val Pro Glu Leu Val Asn Asp Glu Gly Gln Val Glu Asp Thr Val
370 375 380

Arg Glu Gln Tyr Met Lys Leu His Leu Arg Gly Leu Gln Arg Ala Arg
385 390 395 400

Asp Glu Gly Ile Pro Tyr Ala Gly Tyr Phe His Trp Ser Leu Leu Asp
405 410 415

Asn Phe Glu Trp Glu Gln Gly Tyr Ser Gln Arg Phe Gly Met Val Tyr
420 425 430

Val Asp Tyr Gln Thr Gln Glu Arg Ile Leu Lys Arg Ser Gly Gln His
435 440 445

Phe Ala Ala Ile Val Arg Glu Ile Thr Gly Thr Ala
450 455 460

<210> 39

<211> 1521

<212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 39

gtgctcgccc ataaccgctc gcaccgtgaa gaactcctca atcgccggcc ggttgaattc 60

atcagcgccc tggaggcccg gggcgagctc cagcgcatca ccgccgaggt ggaccctac 120

ctcgagatca ccgagatctg cgatcgacc ctgcgcgccg gcggcccggc gctgctgttc 180

gagaacgtca aggggcacga catgcctctg ctgggcaacc tcttcggcac gccgaagcgg 240

gttgccctcg gcatgggcca ggactccgtg gccgccctgc gcgaagtggg cgagctgctc 300

gccttctca aggagccgga gcctcccaag ggctttcgcg acgcctggga caagctgccg 360

atcttcaagc aggtgatgag catggggccg aagaaggtcc gctcggcgcc ggtgcaggaa 420

aaggtgtacg agggcgacga ggtcgacctc gaccgcctgc cgatccagca ctgctggccc 480

ggcgacgccg cgccccigtg cacctggccg ctggtgatca cccgcgggcc ccacaagaag 540

cgccagaacc tcggcatcta ccgccagcag aagctgtcga agaaccggct gatcatgcgc 600

tggctctccc accgcggcgg ggcgctggac ttcttgaggt tccagaagcc ccaccccgcc 660

gagcccttcc cggtggcggg ggcgctgggc gccgaccggc cgaccatcct cggcgcggtg 720

accccggtgc cggattcgtc ctccgagtac gccttcgccg ggctgtgcgc cggctcgcgc 780

accgagctgg tcaagtgcgg ccacgccgac ctggacgtgc cggcctcgcc ggagatcatc 840

ctggaggggt tcatctaccc ggatgacatg gccccgagg gccctacgg cgaccatacc 900

ggctactaca acgaggtgga taccttcccg gtcttcacgg tgacgcgtat gaccatgcgc 960

cgcatgccca tctatcactc cacctacacc ggccggccgc ccgacgagcc ggcatcctt 1020

gggctggcgc tcaacgaggt gtctgtgccg atcctgcgcc gccagttccc ggagatcgtc 1080

gacttctacc tgccgccgga gggctgctcc taccgcatgg cggtggtgac catgaagaag 1140

cagtaccggg gccacgcaa gcgggtgatg atgggcgtgt ggagcttcct gcgccagttc 1200

atgtacacca agttcgtggt ggtgctcgac gacgacgtca gcgccggga ctgggaggac 1260

gtgatctggg ccatcaccac ccgcatggac ccggcccggg acaccgtggt ggtggagaac 1320

accccatcg actacctgga cttcgcctcg ccggtctccg gcctcggttc caagatgggc 1380

ctggatgccca ccagcaagtg gcccgcgag accgaccgcg agtggggggt gcccatcgtc 1440

atggacgagg ccgtcaagcc ccgctcagc gagcgtgga acgagctggg catcgagtc 1500

ccccgacaaca cgacccccctg a

1521

<210> 40

<211> 506

<212> PRT

<213> Unknown

<220>

<223> Obtained from environmental sample

<220>

<221> DOMAIN

<222> (21)...(445)

<223> 3-octaprenyl-4-hydroxybenzoate carboxy-lyase

<220>

<221> SITE

<222> (5)...(8)

<223> N-glycosylation site. Prosite id = PS00001

<400> 40

Met	Leu	Ala	His	Asn	Arg	Ser	His	Arg	Glu	Glu	Leu	Leu	Asn	Arg	Arg
1				5					10					15	

Pro	Val	Glu	Phe	Ile	Ser	Ala	Leu	Glu	Ala	Arg	Gly	Glu	Leu	Gln	Arg
		20					25					30			

Ile	Thr	Ala	Glu	Val	Asp	Pro	Tyr	Leu	Glu	Ile	Thr	Glu	Ile	Cys	Asp
	35					40					45				

Arg	Thr	Leu	Arg	Ala	Gly	Gly	Pro	Ala	Leu	Leu	Phe	Glu	Asn	Val	Lys
	50				55					60					

Gly	His	Asp	Met	Pro	Leu	Leu	Gly	Asn	Leu	Phe	Gly	Thr	Pro	Lys	Arg
65				70				75						80	

Val	Ala	Leu	Gly	Met	Gly	Gln	Asp	Ser	Val	Ala	Ala	Leu	Arg	Glu	Val
			85					90					95		

Gly Glu Leu Leu Ala Phe Leu Lys Glu Pro Glu Pro Pro Lys Gly Phe
100 105 110

Arg Asp Ala Trp Asp Lys Leu Pro Ile Phe Lys Gln Val Met Ser Met
115 120 125

Gly Pro Lys Lys Val Arg Ser Ala Pro Val Gln Glu Lys Val Tyr Glu
130 135 140

Gly Asp Glu Val Asp Leu Asp Arg Leu Pro Ile Gln His Cys Trp Pro
145 150 155 160

Gly Asp Ala Ala Pro Leu Val Thr Trp Pro Leu Val Ile Thr Arg Gly
165 170 175

Pro His Lys Lys Arg Gln Asn Leu Gly Ile Tyr Arg Gln Gln Lys Leu
180 185 190

Ser Lys Asn Arg Leu Ile Met Arg Trp Leu Ser His Arg Gly Gly Ala
195 200 205

Leu Asp Phe Leu Glu Phe Gln Lys Ala His Pro Gly Glu Pro Phe Pro
210 215 220

Val Ala Val Ala Leu Gly Ala Asp Pro Ala Thr Ile Leu Gly Ala Val
225 230 235 240

Thr Pro Val Pro Asp Ser Leu Ser Glu Tyr Ala Phe Ala Gly Leu Leu
245 250 255

Arg Gly Ser Arg Thr Glu Leu Val Lys Cys Gly His Ala Asp Leu Asp
260 265 270

Val Pro Ala Ser Ala Glu Ile Ile Leu Glu Gly Phe Ile Tyr Pro Asp
275 280 285

Asp Met Ala Pro Glu Gly Pro Tyr Gly Asp His Thr Gly Tyr Tyr Asn
290 295 300

Glu Val Asp Thr Phe Pro Val Phe Thr Val Thr Arg Met Thr Met Arg
305 310 315 320

Arg Asp Ala Ile Tyr His Ser Thr Tyr Thr Gly Arg Pro Pro Asp Glu
325 330 335

Pro Ala Ile Leu Gly Leu Ala Leu Asn Glu Val Phe Val Pro Ile Leu
340 345 350

Arg Arg Gln Phe Pro Glu Ile Val Asp Phe Tyr Leu Pro Pro Glu Gly
355 360 365

Cys Ser Tyr Arg Met Ala Val Val Thr Met Lys Lys Gln Tyr Pro Gly
370 375 380

His Ala Lys Arg Val Met Met Gly Val Trp Ser Phe Leu Arg Gln Phe
385 390 395 400

Met Tyr Thr Lys Phe Val Val Val Leu Asp Asp Asp Val Ser Ala Arg
405 410 415

Asp Trp Glu Asp Val Ile Trp Ala Ile Thr Thr Arg Met Asp Pro Ala
420 425 430

Arg Asp Thr Val Val Val Glu Asn Thr Pro Ile Asp Tyr Leu Asp Phe
435 440 445

Ala Ser Pro Val Ser Gly Leu Gly Ser Lys Met Gly Leu Asp Ala Thr
450 455 460

Ser Lys Trp Pro Gly Glu Thr Asp Arg Glu Trp Gly Val Pro Ile Val
465 470 475 480

Met Asp Glu Ala Val Lys Ala Arg Val Ser Glu Arg Trp Asn Glu Leu
485 490 495

Gly Ile Glu Leu Pro Asp Asn Thr Thr Pro
500 505

<210> 41
<211> 1410
<212> DNA
<213> Unknown

<220>
<223> Obtained from environmental sample

<400> 41
atgaagacgc cttcgatcta cgataccatg acgcggtcgg tgcagccgtt gacacccgcc 60

gacggcgaca ccttcgctt ttattgctgc ggccccaccg tctacgggcc ggcgcatgtc 120

ggcaatttcc gcaccttcat cattcaggac gtgctgacgc gcgttatcga agggtcgggc 180

ctcaaaacga gacacgtacg caacatcacc gatgtggacg acaaaacat ccgccaatcg 240

caagcggaag gaaaatctct gaaaatcttc acagggtact ggctggaacg gtccacgcc 300

gattgcgacg cgtgaatct gctgcgccg cagtcgagc ccggcgccgt tgaccatata 360

ccggcgcaaa tccgatgat cgaacaactg atcgaaaaag gccacgccta cgtggcggac 420

gacaactcgg tctattatcg cggtgcttcg ttcgaagcgt acggccggtt gtcacgcctg 480

caagaacgac acatcaccac cggctgcgcc gaacacgcgc ataccgacga tgaatacgag 540

cgcgaaatcg ccgccgactt cgccttgttg aaagcgata aatccgagga cggcccgaac 600

gcgtggccga gcccgctggg cgacggacga cccggctggc acatcgagt cagcgccatg 660

tccgtcgagt atctgggcga gacattcgat ctgcacggcg gcggcgtgga cctgaccttc 720

```

ccccaccacg aaaacgaaat cgcgcaaagc gaagccgcca cgggcaagcc cttcgcgcgt    780

atctggttcc attccgcgca tctcatggtc gaaggccaca agatgtccaa ggcctcggc    840

aacctgttta cgctcgacga tatccgcgcg cgcggattcg acgcatgac cctgcgctat    900

gtcctgcttt cgggcaatta ccgccaaccc ctcaatttca cgtgggactc ccttaacgcc    960

gcgcaaagcg ccttacgccg cctgcgtcag ctcaaccacg atctccagca ggcggcgggc   1020

aagacggtcg cgcccgtga tacttcgtgg gggccgttcg aaccggtgta cgacgcgctt   1080

gccgacaacc tgaacacgcc cgacgccctc ggccgcttat tctccgcct gcacagcatc   1140

gagcgcgcgc ttaacggcaa ggaaaggacg gccgaagagg ccgcctcgc ccgtgcgcag   1200

ttcctgcggg tcatggacct ttctggtttc agcctggacg cgccgccgac cgccgaagcg   1260

cccgaagaag tgctgctgct ggcgcgacgag cgatgggacg ctaaacaagc gcgcgatttc   1320

gtccgcgccg acgccttgcg caaacaggtc accgacctcg gctggacat ccgcgacgcc   1380

aaagacggct acgaactcgt ccaagagtaa                                     1410

```

<210> 42
 <211> 469
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> DOMAIN
 <222> (16)...(323)
 <223> tRNA synthetases class I (C) catalytic domain

<220>
 <221> SITE
 <222> (69)...(72)
 <223> N-glycosylation site. Prosite id = PS00001

<220>

<221> SITE

<222> (316)...(319)

<223> N-glycosylation site. Prosite id = PS00001

<400> 42

Met Lys Thr Pro Ser Ile Tyr Asp Thr Met Thr Arg Ser Val Gln Pro
1 5 10 15

Leu Thr Pro Ala Asp Gly Asp Thr Phe Arg Phe Tyr Cys Cys Gly Pro
20 25 30

Thr Val Tyr Gly Pro Ala His Val Gly Asn Phe Arg Thr Phe Ile Ile
35 40 45

Gln Asp Val Leu Arg Arg Val Ile Glu Gly Ser Gly Leu Lys Thr Arg
50 55 60

His Val Arg Asn Ile Thr Asp Val Asp Asp Lys Thr Ile Arg Gln Ser
65 70 75 80

Gln Ala Glu Gly Lys Ser Leu Lys Ile Phe Thr Gly Tyr Trp Leu Glu
85 90 95

Arg Phe His Ala Asp Cys Asp Ala Leu Asn Leu Leu Arg Pro His Val
100 105 110

Glu Pro Gly Ala Val Asp His Ile Pro Ala Gln Ile Arg Met Ile Glu
115 120 125

Gln Leu Ile Glu Lys Gly His Ala Tyr Val Ala Asp Asp Asn Ser Val
130 135 140

Tyr Tyr Arg Val Ala Ser Phe Glu Ala Tyr Gly Arg Leu Ser Arg Leu
145 150 155 160

Gln Glu Arg His Ile Thr Thr Gly Cys Ala Glu His Ala His Thr Asp

165

170

175

Asp Glu Tyr Glu Arg Glu Ser Ala Ala Asp Phe Ala Leu Trp Lys Ala
180 185 190

His Lys Ser Glu Asp Gly Pro Asn Ala Trp Pro Ser Pro Trp Gly Asp
195 200 205

Gly Arg Pro Gly Trp His Ile Glu Cys Ser Ala Met Ser Val Glu Tyr
210 215 220

Leu Gly Glu Thr Phe Asp Leu His Gly Gly Gly Val Asp Leu Thr Phe
225 230 235 240

Pro His His Glu Asn Glu Ile Ala Gln Ser Glu Ala Ala Thr Gly Lys
245 250 255

Pro Phe Ala Arg Ile Trp Phe His Ser Ala His Leu Met Val Glu Gly
260 265 270

His Lys Met Ser Lys Ser Leu Gly Asn Leu Phe Thr Leu Asp Asp Ile
275 280 285

Arg Ala Arg Gly Phe Asp Ala Met Thr Leu Arg Tyr Val Leu Leu Ser
290 295 300

Gly Asn Tyr Arg Gln Pro Leu Asn Phe Thr Trp Asp Ser Leu Asn Ala
305 310 315 320

Ala Gln Ser Ala Leu Arg Arg Leu Arg Gln Leu Asn His Asp Leu Gln
325 330 335

Gln Ala Ala Gly Lys Thr Val Ala Pro Ala Asp Thr Ser Trp Gly Pro
340 345 350

Phe Glu Pro Val Tyr Asp Ala Leu Ala Asp Asn Leu Asn Thr Pro Asp

355 Ala Leu Gly Arg Leu Phe Ser Ala Leu His Ser Ile Glu Arg Ala Leu
370 375 380

Asn Gly Lys Glu Arg Thr Ala Glu Glu Ala Ala Leu Ala Arg Ala Gln
385 390 395 400

Phe Leu Arg Val Met Asp Leu Phe Gly Phe Ser Leu Asp Ala Pro Pro
405 410 415

Thr Ala Glu Ala Pro Glu Glu Val Arg Ala Leu Ala Gln Gln Arg Trp
420 425 430

Asp Ala Lys Gln Ala Arg Asp Phe Val Arg Ala Asp Ala Leu Arg Lys
435 440 445

Gln Val Thr Asp Leu Gly Trp Thr Ile Arg Asp Ala Lys Asp Gly Tyr
450 455 460

Glu Leu Val Gln Glu
465

<210> 43
<211> 984
<212> DNA
<213> Unknown

<220>
<223> Obtained from environmental sample

<400> 43
atgacgactg aaaccaaadc caaactgtac ttgcataaag tgaacggcca gaaaggactg 60

gacctgcgcc agacctatca gcgcgacttc accgtgaccg aggcgtatcg cgatacgtg 120

ccggatatgc agaacgcttc cgaggcgttg cagggggcca atgtcgccat ccagaaagtc 180

ggcgtatcca atttcaagct gccactcaag taccgcaccc acacgggcga accgaccacg 240

ctggaaacca gcgtaacggg cagcgtatcc ctgaagccgg gcctgaaggg catcaacatg 300

tcccgcgtca tgcggacctt ctacgacttc caggacgacg tgttcacgct cgacacgctg 360

gcccgtatac tggaagcgta caaacgggat gtcgacagca acgacgcaca tcttcggctg 420

agttttctct acccgctgct tcaaaaaagt ctgcgcagcg aattattcgg ctggcaatat 480

taccaggtcg cattcgaggg acggatcgat gccgaaaac gagtccgcac gttcattcat 540

tttgacttcg tgtattcctc cgctgtccc tgttcggctg aactggccga acacgcgcgg 600

gaagtgcgcg gcctatacag catccccac tcgcaacgca gcaaggcgcg cgtcttcgtg 660

gaagttcagc ccggcgccga actcaccatc gaagacgtgc acatgcactg cctgaacgcg 720

ctccaaacgg aaacgcaagt gatggtcaaa cgcgaaagac agcaggcggtt cgctgaaatg 780

aacggcgccg ccatcaaatt cgctgaagac gccgcccgtc tgatctatga gcagttcgac 840

caggatccgc gcatcaagga ttctgaaatc gcctgcgcgc atctggaatc cttgcactcg 900

cacgacgccg tatcggtcat cgccaaaggc gtgcccggcg gcttccgcgc cgacttctcg 960

gacttcaaga gtctgatctg ctaa 984

<210> 44
 <211> 327
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> DOMAIN
 <222> (39)...(308)
 <223> Uncharacterized ACR, COG1469

<220>
 <221> SITE
 <222> (45)...(48)
 <223> N-glycosylation site. Prosite id = PS00001

<220>
 <221> SITE
 <222> (100)...(103)
 <223> N-glycosylation site. Prosite id = PS00001

<400> 44
 Met Thr Thr Glu Thr Lys Ser Lys Leu Tyr Leu His Lys Val Asn Gly
 1 5 10 15

Gln Lys Gly Leu Asp Leu Arg Gln Thr Tyr Gln Arg Asp Phe Thr Val
 20 25 30

Thr Glu Ala Tyr Arg Asp Thr Leu Pro Asp Met Gln Asn Ala Ser Glu
 35 40 45

Ala Leu Gln Gly Ala Asn Val Ala Ile Gln Lys Val Gly Val Ser Asn
 50 55 60

Phe Lys Leu Pro Leu Lys Tyr Arg Thr His Thr Gly Glu Pro Thr Thr
 65 70 75 80

Leu Glu Thr Ser Val Thr Gly Ser Val Ser Leu Lys Pro Gly Leu Lys
 85 90 95

Gly Ile Asn Met Ser Arg Val Met Arg Thr Phe Tyr Asp Phe Gln Asp
 100 105 110

Asp Val Phe Thr Leu Asp Thr Leu Ala Arg Ile Leu Glu Ala Tyr Lys
 115 120 125

Arg Asp Val Asp Ser Asn Asp Ala His Leu Arg Leu Ser Phe Ser Tyr
 130 135 140

Pro Leu Leu Gln Lys Ser Leu Arg Ser Glu Leu Phe Gly Trp Gln Tyr
145 150 155 160

Tyr Gln Val Ala Phe Glu Gly Arg Ile Asp Ala Glu Asn Arg Val Arg
165 170 175

Thr Phe Ile His Phe Asp Phe Val Tyr Ser Ser Ala Cys Pro Cys Ser
180 185 190

Ala Glu Leu Ala Glu His Ala Arg Glu Val Arg Gly Leu Tyr Ser Ile
195 200 205

Pro His Ser Gln Arg Ser Lys Ala Arg Val Phe Val Glu Val Gln Pro
210 215 220

Gly Ala Glu Leu Thr Ile Glu Asp Val His Met His Cys Leu Asn Ala
225 230 235 240

Leu Gln Thr Glu Thr Gln Val Met Val Lys Arg Glu Asp Glu Gln Ala
245 250 255

Phe Ala Glu Met Asn Gly Ala Ala Ile Lys Phe Val Glu Asp Ala Ala
260 265 270

Arg Leu Ile Tyr Glu Gln Phe Asp Gln Asp Pro Arg Ile Lys Asp Phe
275 280 285

Glu Ile Ala Cys Ala His Leu Glu Ser Leu His Ser His Asp Ala Val
290 295 300

Ser Val Ile Ala Lys Gly Val Pro Gly Gly Phe Arg Ala Asp Phe Ser
305 310 315 320

Asp Phe Lys Ser Leu Ile Cys
325

<210> 45

<211> 1377

<212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 45

atgacacaac tggcttttcc atctaacttc atctggggaa cagctacttc cgcttaccaa	60
atcgaaggcg cctggaacgc agacggcaag ggcgaaatcta tttgggatcg cttttcccat	120
acgcagggga agatcattga cggcagcaac ggcatgttgg cctgcgatca ctaccaccgc	180
tggcgcgagg acgtggccct catgagagac ttgggtatgc aggcatatcg cttctccatc	240
tcctggccac gcatcctgcc caccggtcat ggacagatca atcaggctgg gctggacttt	300
tacaatgcc tggtaggacgg gttgctggaa gctggcatca agccctttgc caccctctac	360
cactgggacc tgccgctggc gctacaggct gacggcggct ggccggagcg ctccacggcc	420
aaggcctttg tcgaatacgc cgacgtgttc agccgcgcgc tggcgatcg ggtgaagagc	480
tggatcacc ataacgaacc gtggtgcatc agcatgtga gccatcaaat tggggagcat	540
gcgccggct ggccggactg gcaggctgcg ttggcgccg cgcaccacgt cctcctttcg	600
catggttggg ccgtgccgga actgcgtcgc aacagccgcg atgcagaaat cggcatcacg	660
ttgaacttta ccccgccgga gccagcttcg aacagcgcag ccgatttcaa ggcctatcgc	720
cagttcgatg gctacttcaa ccgtggttc ctggaccgc tctatggccg ccactatccg	780
gcagatatgg tgcacgatta catcgcgcaa ggctacctgc catcacaggg ttgactttc	840
gtggaagctg gtgacctgga cgcgatcgcg acgcgcaccg atttcctggg tgtgaactat	900
tacacgcgcg aagtgttccg tagccaggaa atcccagaga gtgagaacgc gccgcgcaca	960

gtcttgcgcg cgccacagga agagtggaca gagatgggct gggaagtgtg tcctgagggc 1020

ctctacaggt tgctcaatcg gttgcacttt gaataccagc cgcgcaagct ctacgtgacc 1080

gagagcggtt gcagctactc cgatggaccc ggccccaacg gtcggatacc ggaccaacgc 1140

cgtatcaact acctgcgcga tcacttcgca gcggcgcatc aggcgataca atgcggcgtc 1200

ccgctggcgc gctacttcgt ctggtcgttc atggacaact tcgagtgggc caaagggtac 1260

acccaacgtt ttggtatcgt atgggtggat tatcaatcgc aacgacggat accgaaagac 1320

agcgcctact ggtatcgcca tgctcgtgcc gccaacgcgg tgcaagttcc tgattag 1377

<210> 46
 <211> 458
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> DOMAIN
 <222> (2)...(454)
 <223> Glycosyl hydrolase family 1

<220>
 <221> SITE
 <222> (10)...(24)
 <223> Glycosyl hydrolases family 1 N-terminal signature. Prosite id = PS00653

<400> 46
 Met Thr Gln Leu Ala Phe Pro Ser Asn Phe Ile Trp Gly Thr Ala Thr
 1 5 10 15

Ser Ala Tyr Gln Ile Glu Gly Ala Trp Asn Ala Asp Gly Lys Gly Glu
 20 25 30

Ser Ile Trp Asp Arg Phe Ser His Thr Gln Gly Lys Ile Ile Asp Gly
 35 40 45

Ser Asn Gly Asp Val Ala Cys Asp His Tyr His Arg Trp Arg Glu Asp
50 55 60

Val Ala Leu Met Arg Asp Leu Gly Met Gln Ala Tyr Arg Phe Ser Ile
65 70 75 80

Ser Trp Pro Arg Ile Leu Pro Thr Gly His Gly Gln Ile Asn Gln Ala
85 90 95

Gly Leu Asp Phe Tyr Asn Arg Leu Val Asp Gly Leu Leu Glu Ala Gly
100 105 110

Ile Lys Pro Phe Ala Thr Leu Tyr His Trp Asp Leu Pro Leu Ala Leu
115 120 125

Gln Ala Asp Gly Gly Trp Pro Glu Arg Ser Thr Ala Lys Ala Phe Val
130 135 140

Glu Tyr Ala Asp Val Val Ser Arg Ala Leu Gly Asp Arg Val Lys Ser
145 150 155 160

Trp Ile Thr His Asn Glu Pro Trp Cys Ile Ser Met Leu Ser His Gln
165 170 175

Ile Gly Glu His Ala Pro Gly Trp Arg Asp Trp Gln Ala Ala Leu Ala
180 185 190

Ala Ala His His Val Leu Leu Ser His Gly Trp Ala Val Pro Glu Leu
195 200 205

Arg Arg Asn Ser Arg Asp Ala Glu Ile Gly Ile Thr Leu Asn Phe Thr
210 215 220

Pro Ala Glu Pro Ala Ser Asn Ser Ala Ala Asp Phe Lys Ala Tyr Arg

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

Ser Gln Arg Arg Ile Pro Lys Asp Ser Ala Tyr Trp Tyr Arg Asp Val
435 440 445

Val Ala Ala Asn Ala Val Gln Val Pro Asp
450 455

<210> 47
<211> 1353
<212> DNA
<213> Unknown

<220>
<223> Obtained from environmental sample

<400> 47
atgaaaaaat acctttttcc tgaaaatttt ttatgggggtg ctgccacagc ttcgtatcaa 60

atcgaagggtt ctccctctgc tgatggcaaa ggtgaatcga tatgggaccg tttttctcac 120

acaccgggga acatttggaa cgctgaaacc ggggatatcg cctgcgatca ttaccggcgt 180

tacgtggatg atgtaaagct gatttcacaa atcgggctta acgcgtaccg tttttcaatt 240

tcctggccca gggatattcc agaggggaga ggaaaagcaa atgaaaaagg actcgatttt 300

taccgcaggt tgattgaaca gctgcagcaa catcgaatca aaacggcagt gacactttac 360

cactgggatc ttccacaagt tctgcaggat cgcggcgggt gggcaaaccg tgatacggcg 420

aagtattttt ctgagtatgc cacctttctc ttgaaaaac tcgatctccc cgttgacatg 480

tggattactc ttaacgaacc atgggttata gctattctgg ggcatgcttt tggatccac 540

gctccaggga tgagtgaact cagcacagcc ctccaggtct cgcataacct gcttctgggg 600

cacgggttgg cggttaaagc atttcgggag tctaagaggg gtgatgaacc ggtaggtatt 660

```

acccttaacc ttgccccggt tgaaccgctg accgaaaagc cgcgcgatct aaaggcagct    720

ttactttctg acggttttat gaaccgctgg taccttgatc cctgtttcaa aggtggttac    780

cctgaagata tgatggatat ctattcccgg aactttgaac tgcccaaat tgaaaagggg    840

gatgctcagg ttattgccga accgatcgac ttcttgggca taaataacta taccagggtt    900

ctcgtggaag ccagcgggtga tgaaaatgcc ttatgggca accctgtcaa cccccagggc    960

tctgaatata ctgaaatggg ttgggaggtt tatccgcagg gtctctacga cctgctgacc   1020

agggttcacc gggattacgg gccaatgccg ctatatataa ctgaaaacgg ggcagccttt   1080

cccgatgaac ttgacagcaa tgggcagata gatgatccaa ggcggataaa ttacctggaa   1140

acttatcttc atcagtgctg gaaggcagtt caggacgggtg tgcctctaaa aggctatttt   1200

gtctggacce tgatggataa cttcgagtgg gctttcgggtt tcagcaagcg atttgggctc   1260

atatacgtag attaccagga tcagaaacgt tacttgaaaa acagcgccta ctggtatagc   1320

aaggttattg ggcgaaacgg cctcgagcta taa                                1353

```

<210> 48
 <211> 450
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> DOMAIN
 <222> (4)...(448)
 <223> Glycosyl hydrolase family 1

<220>
 <221> SITE
 <222> (10)...(24)
 <223> Glycosyl hydrolases family 1 N-terminal signature. Prosite id = PS00653

<220>
 <221> SITE
 <222> (300)...(303)
 <223> N-glycosylation site. Prosite id = PS00001

<220>
 <221> SITE
 <222> (356)...(364)
 <223> Glycosyl hydrolases family 1 active site. Prosite id = PS00572

<400> 48
 Met Lys Lys Tyr Leu Phe Pro Glu Asn Phe Leu Trp Gly Ala Ala Thr
 1 5 10 15

Ala Ser Tyr Gln Ile Glu Gly Ser Pro Ser Ala Asp Gly Lys Gly Glu
 20 25 30

Ser Ile Trp Asp Arg Phe Ser His Thr Pro Gly Asn Ile Trp Asn Ala
 35 40 45

Glu Thr Gly Asp Ile Ala Cys Asp His Tyr Arg Arg Tyr Val Asp Asp
 50 55 60

Val Lys Leu Ile Ser Gln Ile Gly Leu Asn Ala Tyr Arg Phe Ser Ile
 65 70 75 80

Ser Trp Pro Arg Val Phe Pro Glu Gly Arg Gly Lys Ala Asn Glu Lys
 85 90 95

Gly Leu Asp Phe Tyr Arg Arg Leu Ile Glu Gln Leu Gln Gln His Arg
 100 105 110

Ile Lys Thr Ala Val Thr Leu Tyr His Trp Asp Leu Pro Gln Val Leu
 115 120 125

Gln Asp Arg Gly Gly Trp Ala Asn Arg Asp Thr Ala Lys Tyr Phe Ser
 130 135 140

Glu Tyr Ala Thr Phe Leu Phe Glu Lys Leu Asp Leu Pro Val Asp Met
145 150 155 160

Trp Ile Thr Leu Asn Glu Pro Trp Val Ile Ala Ile Leu Gly His Ala
165 170 175

Phe Gly Ile His Ala Pro Gly Met Ser Asp Phe Ser Thr Ala Leu Gln
180 185 190

Val Ser His Asn Leu Leu Leu Gly His Gly Leu Ala Val Lys Ala Phe
195 200 205

Arg Glu Ser Lys Arg Gly Asp Glu Pro Val Gly Ile Thr Leu Asn Leu
210 215 220

Ala Pro Val Glu Pro Leu Thr Glu Lys Pro Ala Asp Leu Lys Ala Ala
225 230 235 240

Leu Leu Ser Asp Gly Phe Met Asn Arg Trp Tyr Leu Asp Pro Leu Phe
245 250 255

Lys Gly Gly Tyr Pro Glu Asp Met Met Asp Ile Tyr Ser Arg Asn Phe
260 265 270

Glu Leu Pro Lys Ile Glu Lys Gly Asp Ala Gln Val Ile Ala Glu Pro
275 280 285

Ile Asp Phe Leu Gly Ile Asn Asn Tyr Thr Arg Val Leu Val Glu Ala
290 295 300

Ser Gly Asp Glu Asn Ala Phe Met Gly Asn Pro Val Asn Pro Gln Gly
305 310 315 320

Ser Glu Tyr Thr Glu Met Gly Trp Glu Val Tyr Pro Gln Gly Leu Tyr
325 330 335

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

340 345 350

Ile Thr Glu Asn Gly Ala Ala Phe Pro Asp Glu Leu Asp Ser Asn Gly
355 360 365

Gln Ile Asp Asp Pro Arg Arg Ile Asn Tyr Leu Glu Thr Tyr Leu His
370 375 380

Gln Cys Trp Lys Ala Val Gln Asp Gly Val Pro Leu Lys Gly Tyr Phe
385 390 395 400

Val Trp Thr Leu Met Asp Asn Phe Glu Trp Ala Phe Gly Phe Ser Lys
405 410 415

Arg Phe Gly Leu Ile Tyr Val Asp Tyr Gln Asp Gln Lys Arg Tyr Leu
420 425 430

Lys Asn Ser Ala Tyr Trp Tyr Ser Lys Val Ile Gly Arg Asn Gly Leu
435 440 445

Glu Leu
450

<210> 49
<211> 591
<212> DNA
<213> Unknown

<220>
<223> Obtained from environmental sample

<400> 49
atggactttg agcgggcagt tgacaggaat atcattagat tacgctcttc gttaaaggaa 60

gaaatgaagg atctagttgc agttgaagct ccggttaacaa tatittttaa tggcagcgag 120

ctggtaaccc tgctctgcac cccggagaaa attgatcggt tggccctcgg tttccttcat 180

tcagaagggc tgcttaactc acttgatgat cttagtatga tcaggaccag ggagagcgaa 240

ggcctgggtt aaattgaact taaagaggcc tcgccggcac ttgataaatt atacgggaag 300

aggacaatta cticcgggtt cggttaaggga acaatttttt ttaatgttct cgattctctg 360

cgcagtaaac cactcgacgg aaagcttgtg attacaaccg aagagattca taaattaatg 420

gatgacctgc aggggcgggc ggaactgttc aaggctaccg ggggtgttca cagcgctgcg 480

cttgccgaca gaaaggaaat actctttttc agtgaagata tcggccgcca taatgctatc 540

gataaaattg tgggagagtg tttgctggag ggggtatctc ctgaagataa g 591

<210> 50
<211> 197
<212> PRT
<213> Unknown

<220>
<223> Obtained from environmental sample

<220>
<221> DOMAIN
<222> (17)...(195)
<223> FdhD/NarQ family

<220>
<221> SITE
<222> (37)...(40)
<223> N-glycosylation site. Prosite id = PS00001

<400> 50
Met Asp Phe Glu Arg Ala Val Asp Arg Asn Ile Ile Arg Leu Arg Ser
1 5 10 15

Ser Leu Lys Glu Glu Met Lys Asp Leu Val Ala Val Glu Ala Pro Val
20 25 30

Thr Ile Phe Leu Asn Gly Ser Glu Leu Val Thr Leu Leu Cys Thr Pro
35 40 45

Glu Lys Ile Asp Arg Leu Ala Leu Gly Phe Leu His Ser Glu Gly Leu
50 55 60

Leu Asn Ser Leu Asp Asp Leu Ser Met Ile Arg Thr Arg Glu Ser Glu
65 70 75 80

Gly Leu Val Glu Ile Glu Leu Lys Glu Ala Ser Pro Ala Leu Asp Lys
85 90 95

Leu Tyr Gly Lys Arg Thr Ile Thr Ser Gly Cys Gly Lys Gly Thr Ile
100 105 110

Phe Phe Asn Val Leu Asp Ser Leu Arg Ser Lys Pro Leu Asp Gly Lys
115 120 125

Leu Val Ile Thr Thr Glu Glu Ile His Lys Leu Met Asp Asp Leu Gln
130 135 140

Gly Arg Ala Glu Leu Phe Lys Ala Thr Gly Gly Val His Ser Ala Ala
145 150 155 160

Leu Ala Asp Arg Lys Glu Ile Leu Phe Phe Ser Glu Asp Ile Gly Arg
165 170 175

His Asn Ala Ile Asp Lys Ile Val Gly Glu Cys Leu Leu Glu Gly Val
180 185 190

Ser Pro Glu Asp Lys
195

<210> 51

<211> 1014

<212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 51

atgtccagg gcatcctgat cctcgtcatg ctgtctgttc tgagcggcgc ggcgctggcc	60
caaccggcgc ggctgccgcc gcgttcgccg gtgcagcgtc gcatcaacct gggcaatatg	120
ctggaagcgc cggaggaggc ctggtggggg ctgcgcgtcg agcgcgacta cctgacgacg	180
atcgccgggg ccgggttcga tgcggtgcgc atcccataa gctggtcaac ccatgctgcc	240
agcgagccgc cctacacatc cgatccggtc ttcttcgccc gcgttgatga agtcgtcggc	300
tgggcgctgg cggacgggct gaaggccatc atcaacgtgc accactacga ggagatgatg	360
agcgatccgg cggggcatct ccccgggctg cgcgcgctgt gggcgagat cgcggagcac	420
tacgccgact acccgccgcg gctgatgttc gagctgctca acgaaccgtt cgaggcgctg	480
acgccgctgc ggtggaacga gtacgccgc gatctgatcg cgctgatccg ccagaccaac	540
ccggggcgca ccctgatcgt cggcgggggc tgggtgaaca gtgtggaagg gctgatgcag	600
ctccgcctgc cggatgatcc cgatctgtcg gcgacgttcc attactacca cccgttcgag	660
ttcacgcac agggggcgga gtggtcaccg gaagtgactg acctgagcgg gatcgcttgg	720
gggacgggcg aggaacggct cgatctggag tccaatatcc gtattgcgcg ggcctgggcg	780
gtgtacaacc ggcgcccgtc gctgttgggc gaattcggcg tctatggccg ggtggccgat	840
ctcgattcgc gcctgcgctg gacgacggcg gtgcgcgccg aggccgaggc gcaggcatc	900
ggctggtgct actgggaatt cggcggggc ttgggcattt acgacccgga aagccggacg	960
ttcaaccgcg tgtaccgcgc gctgatcccg caggccgggc cggcgcgccc ctga	1014

<210> 52

<211> 337

<212> PRT

<213> Unknown

<220>

<223> Obtained from environmental sample

<220>

<221> SIGNAL

<222> (1)...(20)

<220>

<221> DOMAIN

<222> (38)...(314)

<223> Cellulase (glycosyl hydrolase family 5)

<220>

<221> SITE

<222> (150)...(159)

<223> Glycosyl hydrolases family 5 signature. Prosite id = PS00659

<400> 52

Met	Ser	Arg	Gly	Ile	Leu	Ile	Leu	Val	Met	Leu	Ser	Val	Leu	Ser	Gly
1				5					10					15	

Ala	Ala	Leu	Ala	Gln	Pro	Ala	Gly	Leu	Pro	Pro	Arg	Ser	Pro	Val	Gln
			20					25					30		

Arg	Cys	Ile	Asn	Leu	Gly	Asn	Met	Leu	Glu	Ala	Pro	Glu	Glu	Gly	Trp
	35					40					45				

Trp	Gly	Leu	Arg	Val	Glu	Arg	Asp	Tyr	Leu	Thr	Thr	Ile	Ala	Gly	Ala
	50					55					60				

Gly	Phe	Asp	Ala	Val	Arg	Ile	Pro	Ile	Ser	Trp	Ser	Thr	His	Ala	Ala
65					70				75					80	

Ser	Glu	Pro	Pro	Tyr	Thr	Ile	Asp	Pro	Ala	Phe	Phe	Ala	Arg	Val	Asp
				85					90					95	

Glu	Val	Val	Gly	Trp	Ala	Leu	Ala	Asp	Gly	Leu	Lys	Ala	Ile	Ile	Asn

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

290 295 300

Trp Glu Phe Ala Ala Gly Phe Gly Ile Tyr Asp Pro Glu Ser Arg Thr
305 310 315 320

Phe Asn Pro Leu Tyr Arg Ala Leu Ile Pro Gln Ala Gly Pro Ala Arg
325 330 335

Pro

<210> 53
<211> 1377
<212> DNA
<213> Unknown

<220>
<223> Obtained from environmental sample

<400> 53
atgtggatgg ttcaagcgac atctttaatt caaaaataca atgtgcctgg cccacgctac 60

accagttatc caacggttcc ttattgggaa agtgagaatt tttcactaaa gcagtggcaa 120

caaacgctca aaaaatcctt tgatgagtcg aatcaaagtg aaggcatcag tctgtatatc 180

catttgccat ttgcgaaag tttatgcacc tttctgtggtt gccataaacg tgtgactaaa 240

aagcatgaga tggaaaagcc ttatatccaa gcggtattaa aagaatggga tttatatatgc 300

caacttttgg tggataaacc tgtcattaaa gaaattcatt tgggtggggg aactccgaca 360

ttttttagtc ctgaacattt aacgcagctg attaagggga tattggctaa agccgaagtt 420

gcagatgagc atgagtttag ttttgaagga catccaaca atacgacacg tgaacatttg 480

caagcgctct atgatgttgg atttcgacgt gtcagttatg gcgtgcagga ctataacgaa 540

actgtgcaaa aagccattca cgcattcag ccctatgaaa atgttaaaaa tgtcaccgag 600

tgggcgcgtg agattggcta tacctctatt tcgcatgatt tggcttttgg cctgccgttt 660
 caaagtttag acgatgtctt aaatacgatt gatcaaacca ataccttaat gccggatcgt 720
 ttggctttgt atagctatgc ccatgtgccca tggattaaag gcaatggta acgcggtttt 780
 aaagatgctg atgtcccgaa agacgagatt aaacgtcaat gttatgagga aggcaaaaaa 840
 aaattattag aacatggcta tcatgaaatt ggtatggatc attttgctct agaacaagac 900
 agtatgtatc agtcttttaa agcagggagc ttgcatcgta atttcatggg ttataccgca 960
 tcgaaaacgc aagtgatgat tgggcttggg atttcatcaa ttagtgacag ttggtacagc 1020
 tttagcgaaa acgtgaaaac attagatgaa tattatacct tgctagaaaa aaatcagatt 1080
 cccgtgttta aagggcattgt cttgaatcag gaagatttga tcatccgtaa acatatttta 1140
 aatttgatgt gtggcttcca aacctcatgg gcaaatcccg atatgcaatt tcctgaaatt 1200
 cagtctgttt tggcacaatt agcagaaatg cagcaagatg gtttgattca aattgaagac 1260
 gcatcggta cagttttaga agcgggcaag ccttttgttc gaaatatttg tatggccttt 1320
 gatttaagac tcaagcgcaa caagcctgag aatcggattt tttcatgac gatttaa 1377

<210> 54
 <211> 458
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> DOMAIN
 <222> (59)...(233)
 <223> Radical SAM superfamily

<220>

<221> DOMAIN
 <222> (316)...(431)
 <223> HemN C-terminal region

<220>
 <221> SITE
 <222> (33)...(36)
 <223> N-glycosylation site. Prosite id = PS00001

<220>
 <221> SITE
 <222> (51)...(54)
 <223> N-glycosylation site. Prosite id = PS00001

<220>
 <221> SITE
 <222> (155)...(158)
 <223> N-glycosylation site. Prosite id = PS00001

<220>
 <221> SITE
 <222> (181)...(184)
 <223> N-glycosylation site. Prosite id = PS00001

<220>
 <221> SITE
 <222> (200)...(203)
 <223> N-glycosylation site. Prosite id = PS00001

<400> 54
 Met Trp Met Val Gln Ala Thr Ser Leu Ile Gln Lys Tyr Asn Val Pro
 1 5 10 15

Gly Pro Arg Tyr Thr Ser Tyr Pro Thr Val Pro Tyr Trp Glu Ser Glu
 20 25 30

Asn Phe Ser Leu Lys Gln Trp Gln Gln Thr Leu Lys Lys Ser Phe Asp
 35 40 45

Glu Ser Asn Gln Ser Glu Gly Ile Ser Leu Tyr Ile His Leu Pro Phe
 50 55 60

Cys Glu Ser Leu Cys Thr Phe Cys Gly Cys His Lys Arg Val Thr Lys

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

260 265 270

Gln Cys Tyr Glu Glu Gly Lys Lys Lys Leu Leu Glu His Gly Tyr His
275 280 285

Glu Ile Gly Met Asp His Phe Ala Leu Glu Gln Asp Ser Met Tyr Gln
290 295 300

Ser Phe Lys Ala Gly Ser Leu His Arg Asn Phe Met Gly Tyr Thr Ala
305 310 315 320

Ser Lys Thr Gln Val Met Ile Gly Leu Gly Ile Ser Ser Ile Ser Asp
325 330 335

Ser Trp Tyr Ser Phe Ala Gln Asn Val Lys Thr Leu Asp Glu Tyr Tyr
340 345 350

Thr Leu Leu Glu Lys Asn Gln Ile Pro Val Phe Lys Gly His Val Leu
355 360 365

Asn Gln Glu Asp Leu Ile Ile Arg Lys His Ile Leu Asn Leu Met Cys
370 375 380

Gly Phe Gln Thr Ser Trp Ala Asn Pro Asp Met Gln Phe Pro Glu Ile
385 390 395 400

Gln Ser Val Leu Ala Gln Leu Ala Glu Met Gln Gln Asp Gly Leu Ile
405 410 415

Gln Ile Glu Asp Ala Ser Val Thr Val Leu Glu Ala Gly Lys Pro Phe
420 425 430

Val Arg Asn Ile Cys Met Ala Phe Asp Leu Arg Leu Lys Arg Asn Lys
435 440 445

Pro Glu Asn Arg Ile Phe Ser Met Thr Ile

450

455

<210> 55

<211> 1389

<212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 55

atgagcgctt cgagtcctc ccgccccctg tccttcccag agcagttcgt ctggggtgct 60

gccgcggcct cctaccaagt cgaggcgcc gtccacgagg acggaagg cccctcgtc 120

tgggacatgt tctgcgagaa gcccgagcg gtcttcagg ggcacgacgg ggcggtggct 180

tgcgaccact atcacgcta ccgagaggac gtggcgttga tgcgacaggt ggcctgcac 240

gcctaccgcc tgagcgtgtg ctggccccga gtgtccccgg agggcgctcg gcagcccaac 300

gagaagggcc tcgacttcta ctgcgggtt gtggacgcgc tgctcgagc agggattacg 360

ccctgggtaa cgctttttca ttgggactac cccttggtc tctatcacg ggggggctgg 420

ctcaaccggg atagcgcgga ttggtttgcc gactacgagg gcctaatacg cgtcgcctc 480

tccgaccggg tgcagcattt cttcactcag aacgagcccc aggtctatat cggcttcgga 540

cacctcgagg gtaagcatgc tccaggagac accttgcca tgtcccaggt gctgcttgcg 600

gggcatcata gcctactggc gcacggcaag gccgtgcagg cgctccgcgc ccaggcgaag 660

cagcagctgc gcgtcggcta cgctccgtc ggcatgcccc tccatccctt cacggactcg 720

gccgaggacg tggccgctgc gcggaaggcg accttttggg ttcgggagaa gaactcctgg 780

aacaacgcct ggtggatgga cccggtgttc ttgggtgagt acccggtca ggcctcgc 840

ttcttcggcc gggacgtgcc gcaggtgcgc gagggagaca tgcagtcacat cgcgcagccc 900

ttggacttct ttggggtaaa catctaccag agcaccctccg tgcgcgcgtc tagcgccgaa 960

agcggcttcg aggtcgtccc ccatccaacg ggctatccta tcactgcctt caactggccg 1020

atcacgcccc aggccttcta ctgggggtccg cgctttctct acgagcgcta ccagaagccg 1080

atcgtcatca cggagaacgg actgtcctgt cgggacgtcg tcgctgtgga cgggaaggtt 1140

cacgatccgg ctgcacatga ttccaccacc cgctatctgc gcgagctcca ccgagccgtc 1200

gcggacggcg tcgcggtcga gggctacttc cactggtcca tcatggacaa ctgcgaatgg 1260

gtgcccggct accgcgagcg gttcgggctc attcacgtcg actacgagac cctggcgcg 1320

acgccaagg cgtccgtcgc gtggtatcgc aaggtaatcg agagcaacgg agcgaccctt 1380

ttcggatga 1389

<210> 56
 <211> 462
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> DOMAIN
 <222> (8)...(458)
 <223> Glycosyl hydrolase family 1

<220>
 <221> SITE
 <222> (16)...(30)
 <223> Glycosyl hydrolases family 1 N-terminal signature. Prosite id = PS00653

<220>
 <221> SITE
 <222> (366)...(374)
 <223> Glycosyl hydrolases family 1 active site. Prosite id = PS00572

<400> 56

Met Ser Ala Ser Ser Pro Ser Arg Pro Leu Ser Phe Pro Glu Gln Phe
1 5 10 15

Val Trp Gly Ala Ala Ala Ala Ser Tyr Gln Val Glu Gly Ala Val His
20 25 30

Glu Asp Gly Lys Gly Pro Ser Val Trp Asp Met Phe Cys Glu Lys Pro
35 40 45

Gly Ala Val Phe Gln Gly His Asp Gly Ala Val Ala Cys Asp His Tyr
50 55 60

His Arg Tyr Arg Glu Asp Val Ala Leu Met Arg Gln Val Gly Leu His
65 70 75 80

Ala Tyr Arg Leu Ser Val Cys Trp Pro Arg Val Leu Pro Glu Gly Val
85 90 95

Gly Gln Pro Asn Glu Lys Gly Leu Asp Phe Tyr Ser Arg Leu Val Asp
100 105 110

Ala Leu Leu Glu Ala Gly Ile Thr Pro Trp Val Thr Leu Phe His Trp
115 120 125

Asp Tyr Pro Leu Ala Leu Tyr His Arg Gly Gly Trp Leu Asn Arg Asp
130 135 140

Ser Ala Asp Trp Phe Ala Glu Tyr Ala Gly Leu Ile Ala Asp Arg Leu
145 150 155 160

Ser Asp Arg Val Gln His Phe Phe Thr Gln Asn Glu Pro Gln Val Tyr
165 170 175

Ile Gly Phe Gly His Leu Glu Gly Lys His Ala Pro Gly Asp Thr Leu

180	185	190
Pro Met Ser Gln Val Leu Leu Ala Gly His His Ser Leu Leu Ala His		
195	200	205
Gly Lys Ala Val Gln Ala Leu Arg Ala Gln Ala Lys Gln Gln Leu Arg		
210	215	220
Val Gly Tyr Ala Pro Val Gly Met Pro Leu His Pro Phe Thr Asp Ser		
225	230	235 240
Ala Glu Asp Val Ala Ala Ala Arg Lys Ala Thr Phe Trp Val Arg Glu		
245	250	255
Lys Asn Ser Trp Asn Asn Ala Trp Trp Met Asp Pro Val Phe Leu Gly		
260	265	270
Glu Tyr Pro Ala Gln Gly Leu Ala Phe Phe Gly Arg Asp Val Pro Gln		
275	280	285
Val Arg Glu Gly Asp Met Gln Leu Ile Ala Gln Pro Leu Asp Phe Phe		
290	295	300
Gly Val Asn Ile Tyr Gln Ser Thr Pro Val Arg Ala Ser Ser Ala Glu		
305	310	315 320
Ser Gly Phe Glu Val Val Pro His Pro Thr Gly Tyr Pro Ile Thr Ala		
325	330	335
Phe Asn Trp Pro Ile Thr Pro Gln Ala Leu Tyr Trp Gly Pro Arg Phe		
340	345	350
Phe Tyr Glu Arg Tyr Gln Lys Pro Ile Val Ile Thr Glu Asn Gly Leu		
355	360	365
Ser Cys Arg Asp Val Val Ala Val Asp Gly Lys Val His Asp Pro Ala		

370

375

380

Arg Ile Asp Phe Thr Thr Arg Tyr Leu Arg Glu Leu His Arg Ala Val
385 390 395 400

Ala Asp Gly Val Ala Val Glu Gly Tyr Phe His Trp Ser Ile Met Asp
405 410 415

Asn Phe Glu Trp Ala Ala Gly Tyr Arg Glu Arg Phe Gly Leu Ile His
420 425 430

Val Asp Tyr Glu Thr Leu Ala Arg Thr Pro Lys Ala Ser Ala Ala Trp
435 440 445

Tyr Arg Lys Val Ile Glu Ser Asn Gly Ala Thr Leu Phe Gly
450 455 460

<210> 57

<211> 414

<212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 57

atgattgctt catctatgtt ctatggaacg gttcgtggaa tacaagagct aactcaaac 60

gttattgcat tggataccgc aatggtttcg cttaccagag ttgctgacgg aagtgatatt 120

gagtttgata gagttattga acgctcgatt gaaaacgtaa ccgaactatc aggtaagcta 180

actgattaca tggatttagt aacggagttt gctagaactg gtaaaacaat agatgaatct 240

tttaatttag ctaatacaac acaaatgtta atgaatattt ctgaattaac agcagatgaa 300

tcagtaaata gttaactgc cgcaatgatt gcttttaata ttaacgcaga tgatagtatt 360

agaattgctg ataagttgaa tgaggttaac aatcagcc tcctttgtg gtaa 414

<210> 58
<211> 137
<212> PRT
<213> Unknown

<220>
<223> Obtained from environmental sample

<220>
<221> SITE
<222> (52)...(55)
<223> N-glycosylation site. Prosite id = PS00001

<220>
<221> SITE
<222> (86)...(89)
<223> N-glycosylation site. Prosite id = PS00001

<220>
<221> SITE
<222> (93)...(96)
<223> N-glycosylation site. Prosite id = PS00001

<220>
<221> SITE
<222> (133)...(136)
<223> N-glycosylation site. Prosite id = PS00001

<400> 58
Met Ile Ala Ser Ser Met Phe Tyr Gly Thr Val Arg Gly Ile Gln Glu
1 5 10 15

Leu Thr Gln Asn Val Ile Ala Leu Asp Thr Ala Met Val Ser Leu Thr
20 25 30

Arg Val Ala Asp Gly Ser Asp Phe Glu Phe Asp Arg Val Ile Glu Arg
35 40 45

Ser Ile Glu Asn Val Thr Glu Leu Ser Gly Lys Leu Thr Asp Tyr Met
50 55 60

Asp Leu Val Thr Glu Phe Ala Arg Thr Gly Lys Thr Ile Asp Glu Ser
65 70 75 80

Phe Asn Leu Ala Asn Thr Thr Gln Met Leu Met Asn Ile Ser Glu Leu
85 90 95

Thr Ala Asp Glu Ser Val Asn Ser Leu Thr Ala Ala Met Ile Ala Phe
100 105 110

Asn Ile Asn Ala Asp Asp Ser Ile Arg Ile Ala Asp Lys Leu Asn Glu
115 120 125

Val Asn Asn Ile Ser Leu Leu Leu Trp
130 135

<210> 59
<211> 1044
<212> DNA
<213> Unknown

<220>
<223> Obtained from environmental sample

<400> 59
atgagaattt ttgaaggatt tcagcgaggt gtaaacttg gcggctggat ctcccagttc 60

gacaagtacg accatgagca tttccgcagc ttattacgg aaaatgacat cgccgccatt 120

gcagctcttg gttttgacca tgtccgcgtg ccggtggatt ataactgtct ggaggatgag 180

gagggcaacc gcatcgacag cggatttgc tacctgagaa gctgctacga gtggtgccgc 240

aaacacgacc tgaacatgct ggtgatctt cagagtgtc acggctactc cttcgatccg 300

ctgaaaaaag atatggaccg caaacgcttc ttctatgccg aagctctgca ggagcgtttt 360

ctgaagctct gggagcagat ctgtgaaacc tttaaagacg atcctgtgca cgtggcattc 420

gagccgctga atgagatcgt tttaggagag gtcgcagacg cctggaacgt gatgatccgc 480

aaatatatca agaccgtccg cgccatctgc ccggagcact atctggctct tggagcgtg 540

cactacagcc acgttaccac catccctctt cttgaggcac cggcagatga caagatcgtc 600

ttcaacttcc acigtctacga gccgctggtc ttacccacc agggcgcata ctggctggag 660

gatatgattc cggatttccg catgacctat cctgccacca tggaagagtt ctacgaagca 720

acaaagaaga tcctgcaaaa catgagtcg gatggattta aggatttcga tcaggagatg 780

ggtcgggct tctttgagaa gatcttcaca ccggccctga aacgtgccga gcaggacaat 840

gtagccctct actgcggcga gtacggcgtc attgatctgg cagataacca tgccaagatc 900

cgctggctca aagacatcca caccacctic tccaaatag gcatcggaag tgccctctgg 960

aactacaagg gcaaggattt cggctatgta gatgatcgct tcgccgagtg cagagaagca 1020

tttatcgagt gcctgaaggc ctga 1044

<210> 60
 <211> 347
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> DOMAIN
 <222> (14)...(330)
 <223> Cellulase (glycosyl hydrolase family 5)

<400> 60
 Met Arg Ile Phe Glu Gly Phe Gln Arg Gly Val Asn Leu Gly Gly Trp
 1 5 10 15

Ile Ser Gln Phe Asp Lys Tyr Asp His Glu His Phe Arg Ser Phe Ile

20 25 30

Thr Glu Asn Asp Ile Ala Ala Ile Ala Ala Leu Gly Phe Asp His Val
35 40 45

Arg Val Pro Val Asp Tyr Asn Val Leu Glu Asp Glu Glu Gly Asn Arg
50 55 60

Ile Asp Ser Gly Phe Val Tyr Leu Arg Ser Cys Tyr Glu Trp Cys Arg
65 70 75 80

Lys His Asp Leu Asn Met Leu Val Asp Leu His Glu Cys Tyr Gly Tyr
85 90 95

Ser Phe Asp Pro Leu Lys Lys Asp Met Asp Arg Lys Arg Phe Phe Tyr
100 105 110

Ala Glu Ala Leu Gln Glu Arg Phe Leu Lys Leu Trp Glu Gln Ile Cys
115 120 125

Glu Thr Phe Lys Asp Asp Pro Val His Val Ala Phe Glu Pro Leu Asn
130 135 140

Glu Ile Val Leu Gly Glu Val Ala Asp Ala Trp Asn Val Met Ile Arg
145 150 155 160

Lys Tyr Ile Lys Thr Val Arg Ala Ile Cys Pro Glu His Tyr Leu Val
165 170 175

Leu Gly Ser Val His Tyr Ser His Val Thr Thr Ile Pro Leu Leu Glu
180 185 190

Ala Pro Ala Asp Asp Lys Ile Val Phe Asn Phe His Cys Tyr Glu Pro
195 200 205

Leu Val Phe Thr His Gln Gly Ala Tyr Trp Leu Glu Asp Met Ile Pro

210

215

220

Asp Phe Arg Met Thr Tyr Pro Ala Thr Met Glu Glu Phe Tyr Glu Ala
225 230 235 240

Thr Lys Lys Ile Leu Pro Asn Met Ser Pro Asp Gly Phe Lys Asp Phe
245 250 255

Asp Gln Glu Met Gly Pro Gly Phe Phe Glu Lys Ile Phe Thr Pro Ala
260 265 270

Leu Lys Arg Ala Glu Gln Asp Asn Val Ala Leu Tyr Cys Gly Glu Tyr
275 280 285

Gly Val Ile Asp Leu Ala Asp Asn His Ala Lys Ile Arg Trp Leu Lys
290 295 300

Asp Ile His Thr Thr Phe Ser Lys Tyr Gly Ile Gly Ser Ala Leu Trp
305 310 315 320

Asn Tyr Lys Gly Lys Asp Phe Gly Tyr Val Asp Asp Arg Phe Ala Glu
325 330 335

Cys Arg Glu Ala Phe Ile Glu Cys Leu Lys Ala
340 345

<210> 61

<211> 1230

<212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 61

ttggtatgga caccagctcg atcaacgctt gctggatctt ctgaaatccc actaatgaca 60

atgaatatat tccccaatag aaaagactca cgaatgtccc tctggatcaa gcttggcata 120

ctttgtatga tggctggaac ggtgatggtt cacggagcgc agactggtca aggagaagca	180
acaatgaatc aagcaaatgg cttcaaggta agcaacggga ccaatatcag ccattggttg	240
tcccagtgtt ttgaaacaat gccaccccg cgcgatttt tctccgaact ggatgttacc	300
ttcatccgct cgctggggat ggatcatttc cgtcttcagg tggacgagaa ggaactttgg	360
acggaggatc ttgagaagat tcccgaagcg tgggattacc tcaggaatgc tctaagctgg	420
gctagaaagc atgagcttcg tgtgattgtg gatcttcacg tcgtgcggtc ccatcacttt	480
aatgcggcaa atgaaggggg aaccaacact ctgtgggatg atccggaggc gcaggaaagt	540
ttctcaacc ttggaggca gctttcgca gagctcgct acaccgatgt ggactgggtg	600
gcctatgaga tcatgaatga ggccgtcgcg gatgatccgg aggactggaa tcgtctcatc	660
gccaaagccc actccttgat ccgcgagcgt gagccaagc gcacactcgt catcggatec	720
aaccggtggc aaattccgtc aacgttcccg gatctgaaga ttccggacgg agatccgaac	780
atcctcctga gtttccattt ctacgcgct ctgcttttca cccactatcg ggcaacctgg	840
gttgctttt acgattatga tgggcccgtt tcctatcctg gcaggatcgt tgatgatgca	900
gctcttgaga aaaatgatta tactcctgca ttcaaagaca agattcgtgc gttgaatggt	960
gtgtatgaca tcgacgtctc cgaaaaagaa atgcagccgg ctatcgaata cgcaaacag	1020
aaagggttac cactgtattg cggagagtgg ggatgttttc atgctgtgga aagaaaacaa	1080
cgcttgcaat ggtacaaaga tatatccact attttgaac gcaatgggat cgcccatgcc	1140
acatgggatt acaaggcgca gttcggcatt gtggacactt ggacactagg tgttgattgg	1200
aatttggtag gagcaatcct gtcagagtag	1230

<210> 62
 <211> 409
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> DOMAIN
 <222> (62)...(390)
 <223> Cellulase (glycosyl hydrolase family 5)

<220>
 <221> SITE
 <222> (73)...(76)
 <223> N-glycosylation site. Prosite id = PS00001

<400> 62
 Met Val Trp Thr Pro Ala Arg Ser Thr Leu Ala Gly Ser Ser Glu Ile
 1 5 10 15

Pro Leu Met Thr Met Asn Ile Phe Pro Asn Arg Lys Asp Ser Arg Met
 20 25 30

Ser Leu Trp Ile Lys Leu Gly Ile Leu Cys Met Met Ala Gly Thr Val
 35 40 45

Met Val His Gly Ala Gln Thr Gly Gln Gly Glu Ala Thr Met Asn Gln
 50 55 60

Ala Asn Gly Phe Lys Val Ser Asn Gly Thr Asn Ile Ser His Trp Leu
 65 70 75 80

Ser Gln Cys Phe Glu Thr Met Pro Pro Arg Arg Gly Phe Phe Ser Glu
 85 90 95

Leu Asp Val Ile Phe Ile Arg Ser Leu Gly Met Asp His Phe Arg Leu
 100 105 110

Pro Val Asp Glu Lys Glu Leu Trp Thr Glu Asp Leu Glu Lys Ile Pro
115 120 125

Glu Ala Trp Asp Tyr Leu Arg Asn Ala Leu Ser Trp Ala Arg Lys His
130 135 140

Glu Leu Arg Val Ile Val Asp Leu His Val Val Arg Ser His His Phe
145 150 155 160

Asn Ala Ala Asn Glu Gly Gly Thr Asn Thr Leu Trp Asp Asp Pro Glu
165 170 175

Ala Gln Glu Ser Phe Leu Asn Leu Trp Arg Gln Leu Ser Ala Glu Leu
180 185 190

Ala Tyr Thr Asp Val Asp Trp Val Ala Tyr Glu Ile Met Asn Glu Ala
195 200 205

Val Ala Asp Asp Pro Glu Asp Trp Asn Arg Leu Ile Ala Lys Ala His
210 215 220

Ser Leu Ile Arg Glu Arg Glu Pro Arg Arg Thr Leu Val Ile Gly Ser
225 230 235 240

Asn Arg Trp Gln Ile Pro Ser Thr Phe Pro Asp Leu Lys Ile Pro Asp
245 250 255

Gly Asp Pro Asn Ile Leu Leu Ser Phe His Phe Tyr Ala Pro Leu Leu
260 265 270

Phe Thr His Tyr Arg Ala Thr Trp Val Ala Phe Tyr Asp Tyr Asp Gly
275 280 285

Pro Val Ser Tyr Pro Gly Arg Ile Val Asp Asp Ala Ala Leu Glu Lys
290 295 300

Asn Asp Tyr Thr Pro Ala Phe Lys Asp Lys Ile Arg Ala Leu Asn Gly
305 310 315 320

Val Tyr Asp Ile Asp Ala Leu Glu Lys Glu Met Gln Pro Ala Ile Glu
325 330 335

Tyr Ala Lys Gln Lys Gly Leu Pro Leu Tyr Cys Gly Glu Trp Gly Cys
340 345 350

Phe His Ala Val Glu Arg Lys Gln Arg Leu Gln Trp Tyr Lys Asp Ile
355 360 365

Ser Thr Ile Leu Lys Arg Asn Gly Ile Ala His Ala Thr Trp Asp Tyr
370 375 380

Lys Gly Glu Phe Gly Ile Val Asp Thr Trp Thr Leu Gly Val Asp Trp
385 390 395 400

Asn Leu Val Gly Ala Ile Leu Ser Glu
405

<210> 63

<211> 1152

<212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 63

atgaaacgga gggaattcat gttggggggt gcgggtgttg ctgcgttggc atcgactctt 60

ggagtctccg ccggttcac ttccgggcag ggagtgaacg agaatgtgag ggtataccgg 120

aatgcgattc cccgttgag ggggttcaac ctcatgccct tttctcggc aatgagcacc 180

aaccggaat acaatggtct gacggtgccg gaggatgacc taaactggat ccgcgactgg 240

ggttttgact atgtccggct tccgattgat tactggattc tggttgattc cgattggcga 300

gatgcaaagc gcatgcgggt agaggatgtt cgcaaggccg accagaaggg atattcacgg 360

ctggacgctg tgattgaagc ctgtatcgcg aagggtttgc acctcaacct gaatatgcat 420

cgggtgtccc ggiattgcat caatggctgg gaactggagc cctataacct cttcaaggat 480

gagcaggcgg aggatgattt tgtctacat tgggagttgc tcgcgagacg ctataaggga 540

atcgatcctt cgctgctgag tttcaatctg ctgaatgagg ctccaatcc tggagacaag 600

atgtcgtcgg aggattatcg tcgggtgatg cttcgatccg ctgctgttat tcgggggata 660

agccccgacc gcatgattat tgtggacggg ctggaaatcg gtaaatacgt tgttccaggg 720

ctgatgcatg agccatttgc ccaagctgtt catgcctacg agccccacga gttgagccat 780

tataatgcgc cttggacgtc ggtgtttatg ggtattcctg agccatcctg gccgacagtt 840

cgtttggatg gttctctgtt cgaccgcaag cgactggagt tgtatttcgc gccgtggggg 900

gagtttgtcc gccagggggg aggggtccac tgtggggaga ccggttgcta cattcatacg 960

ccccatcggg tgtttctgtc ctggttcgaa gatgttttgg atatcctgac cggatacgac 1020

atagggtggg ctctatggaa tttccgggga gatttcggaa tacttgattc caaacgcaag 1080

gatgtgcaat atgtcgattg gtatggacac cagctcgatc aacgcttgct ggatcttctg 1140

aaatcccact aa 1152

<210> 64
 <211> 383
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>

<221> SIGNAL

<222> (1)...(24)

<220>

<221> DOMAIN

<222> (48)...(357)

<223> Cellulase (glycosyl hydrolase family 5)

<400> 64

Met Lys Arg Arg Glu Phe Met Leu Gly Gly Ala Gly Val Ala Ala Leu
1 5 10 15

Ala Ser Thr Leu Gly Val Ser Ala Gly Ser Thr Ser Gly Gln Gly Val
20 25 30

Asn Glu Asn Val Arg Val Tyr Arg Asn Ala Ile Pro Arg Trp Arg Gly
35 40 45

Phe Asn Leu Met Pro Phe Phe Ser Ala Met Ser Thr Asn Pro Glu Tyr
50 55 60

Asn Gly Leu Thr Val Pro Glu Asp Asp Leu Asn Trp Ile Arg Asp Trp
65 70 75 80

Gly Phe Asp Tyr Val Arg Leu Pro Ile Asp Tyr Trp Ile Leu Val Asp
85 90 95

Ser Asp Trp Arg Asp Ala Lys Arg Met Arg Val Glu Asp Val Arg Lys
100 105 110

Ala Asp Gln Lys Gly Tyr Ser Arg Leu Asp Ala Val Ile Glu Ala Cys
115 120 125

Ile Ala Lys Gly Leu His Leu Asn Leu Asn Met His Arg Cys Pro Gly
130 135 140

Tyr Cys Ile Asn Gly Trp Glu Leu Glu Pro Tyr Asn Leu Phe Lys Asp

145	150	155	160
Glu Gln Ala Glu Asp Asp Phe Val Tyr His Trp Glu Leu Leu Ala Arg	165	170	175
Arg Tyr Lys Gly Ile Asp Pro Ser Leu Leu Ser Phe Asn Leu Leu Asn	180	185	190
Glu Ala Pro Asn Pro Gly Asp Lys Met Ser Ser Glu Asp Tyr Arg Arg	195	200	205
Val Met Leu Arg Ser Ala Ala Val Ile Arg Gly Ile Ser Pro Asp Arg	210	215	220
Met Ile Ile Val Asp Gly Leu Glu Ile Gly Lys Ser Val Val Pro Gly	225	230	235
Leu Met His Glu Pro Phe Ala Gln Ala Val His Ala Tyr Glu Pro His	245	250	255
Glu Leu Ser His Tyr Asn Ala Pro Trp Thr Ser Val Phe Met Gly Ile	260	265	270
Pro Glu Pro Ser Trp Pro Thr Val Arg Leu Asp Gly Ser Leu Phe Asp	275	280	285
Arg Lys Arg Leu Glu Leu Tyr Phe Ala Pro Trp Gly Glu Leu Val Arg	290	295	300
Gln Gly Val Gly Val His Cys Gly Glu Thr Gly Cys Tyr Ile His Thr	305	310	315
Pro His Arg Val Phe Leu Ser Trp Phe Glu Asp Val Leu Asp Ile Leu	325	330	335
Thr Gly Tyr Asp Ile Gly Trp Ala Leu Trp Asn Phe Arg Gly Asp Phe	340	345	350

Gly Ile Leu Asp Ser Lys Arg Lys Asp Val Gln Tyr Val Asp Trp Tyr
355 360 365

Gly His Gln Leu Asp Gln Arg Leu Leu Asp Leu Leu Lys Ser His
370 375 380

<210> 65
<211> 1131
<212> DNA
<213> Unknown

<220>
<223> Obtained from environmental sample

<400> 65
atgaacacac tctaccacg gcggcgactg tggctctcca cggcgatcct gcgcacgctg 60

gcggccgggg cgctggcggc cggatatgtc ctggcacccg tcagtgccgc caacgcggcc 120

accaccctcg gtgcctcggc ggccgagaag ggccggtact tcggtgcggc cgtcgggacg 180

tacaagtcca acgacagcac ctacatgtcg gtgctgaacc gcgagttcaa cagcctggtc 240

gccgagaacg agatgaagtg ggacgcgacc gagccccagc gcggcgtgtt caactacagc 300

gccggggacc gcatcgtcaa ccaccccga tcccaggga tgaaggtacg cggacacgcc 360

ctgttgtggc acgcccagca gccacgctgg acggagggcc tgtccggcgg cgacctgcgc 420

aacgccgcga tcaaccatgt caccaggtg gccagccact tccgggggca gatctactcc 480

tgggacgtgg tgaacgaggc ttctgccgac ggtggcagcg gtgcccggcg ggactcgaac 540

ctccagcgca ccggcaacga ctggatcgag gcggcggtcc gtgccgcccg ggcagccgat 600

cccaacgcca agctctgcta caacgactac aacaccgacg ggatcaacgc gaagtcacc 660

ggcgtctaca acatggtgcg tgacttcaag tcccgtaggg tgccgatcga ctgcgtgggc 720

ttccagtcac acctggggcac caccctcccc ggtagtacc aggccaacct tcagcgcttc 780

gccgacctgg gcgtcgacgt ggagatcacc gagctggaca tcaccaggg cggaaccag 840

gccaacatgt acggcgccgt caccgcgcc tgctggcga tctcgcgtg caccggcatc 900

accgtgtggg gggtacggga ctgcgactcc tggcgtggtg gggacaacgc cctgctgttc 960

gactgcgcg gcaacaagaa gcccgctac acggcgtcc tcgacgcct caacagcggc 1020

tcgaacccga accccaaccc caccggcaac cggtgcgca acgaggcctc cggtcgatgc 1080

ctggacgtca acggcgcaag ctccgccaac gggtcacaaa tgatccaaag a 1131

<210> 66
 <211> 377
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> SIGNAL
 <222> (1)...(39)

<220>
 <221> DOMAIN
 <222> (42)...(337)
 <223> Glycosyl hydrolase family 10

<220>
 <221> SITE
 <222> (99)...(102)
 <223> N-glycosylation site. Prosite id = PS00001

<220>
 <221> SITE
 <222> (268)...(278)
 <223> Glycosyl hydrolases family 10 active site. Prosite id = PS00591

<220>

<221> SITE

<222> (375)...(378)

<223> N-glycosylation site. Prosite id = PS00001

<400> 66

Met Asn Thr Leu Leu Pro Arg Arg Arg Leu Trp Ser Ser Thr Ala Ile
1 5 10 15

Leu Arg Thr Leu Ala Ala Gly Ala Leu Ala Ala Gly Met Val Leu Ala
20 25 30

Pro Val Ser Ala Ala Asn Ala Ala Thr Thr Leu Gly Ala Ser Ala Ala
35 40 45

Glu Lys Gly Arg Tyr Phe Gly Ala Ala Val Gly Thr Tyr Lys Phe Asn
50 55 60

Asp Ser Thr Tyr Met Ser Val Leu Asn Arg Glu Phe Asn Ser Leu Val
65 70 75 80

Ala Glu Asn Glu Met Lys Trp Asp Ala Thr Glu Pro Gln Arg Gly Val
85 90 95

Phe Asn Tyr Ser Ala Gly Asp Arg Ile Val Asn His Ala Arg Ser Gln
100 105 110

Gly Met Lys Val Arg Gly His Ala Leu Leu Trp His Ala Gln Gln Pro
115 120 125

Arg Trp Thr Glu Gly Leu Ser Gly Gly Asp Leu Arg Asn Ala Ala Ile
130 135 140

Asn His Val Thr Gln Val Ala Ser His Phe Arg Gly Gln Ile Tyr Ser
145 150 155 160

Trp Asp Val Val Asn Glu Ala Phe Ala Asp Gly Gly Ser Gly Ala Arg

165

170

175

Arg Asp Ser Asn Leu Gln Arg Thr Gly Asn Asp Trp Ile Glu Ala Ala
180 185 190

Phe Arg Ala Ala Arg Ala Ala Asp Pro Asn Ala Lys Leu Cys Tyr Asn
195 200 205

Asp Tyr Asn Thr Asp Gly Ile Asn Ala Lys Ser Thr Gly Val Tyr Asn
210 215 220

Met Val Arg Asp Phe Lys Ser Arg Gly Val Pro Ile Asp Cys Val Gly
225 230 235 240

Phe Gln Ser His Leu Gly Thr Thr Leu Pro Gly Asp Tyr Gln Ala Asn
245 250 255

Leu Gln Arg Phe Ala Asp Leu Gly Val Asp Val Glu Ile Thr Glu Leu
260 265 270

Asp Ile Thr Gln Gly Gly Asn Gln Ala Asn Met Tyr Gly Ala Val Thr
275 280 285

Arg Ala Cys Leu Ala Ile Ser Arg Cys Thr Gly Ile Thr Val Trp Gly
290 295 300

Val Arg Asp Cys Asp Ser Trp Arg Gly Gly Asp Asn Ala Leu Leu Phe
305 310 315 320

Asp Cys Ala Gly Asn Lys Lys Pro Ala Tyr Thr Ala Val Leu Asp Ala
325 330 335

Leu Asn Ser Gly Ser Asn Pro Asn Pro Asn Pro Thr Gly Asn Arg Leu
340 345 350

Arg Asn Glu Ala Ser Gly Arg Cys Leu Asp Val Asn Gly Ala Ser Ser

355

360

365

Ala Asn Gly Ser Gln Met Ile Gln Arg
370 375

<210> 67

<211> 1023

<212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 67

atgaaatata tatatttcgta tataataatg atgatttttaa tcggttttat accggtctat 60

ggattcggcg attcacctga ccaaacatac tctctccct tcctcagcgt agaaggaaat 120

tcattcgtcg atgaaaacgg tgaggagggtt attttgctgg gtgtatcgtt tcccgatccc 180

aatcgattgg atgatgctac tcaatggaac aaacgggtatt tccaggcagc aaaagattgg 240

aactgtaatg tcgtcagaat accggttcat ccgcaaagat ggctgggaaag gggaaaagaa 300

aattatctga aacttttaga taagggtatc gaggggccg gtgaactcgg tatgtacgtg 360

atcattgact ggcacactat cggcaatccg attaccgaag tgttcttcgg cgagctctat 420

aatacgaccc agaccgaaac gtcccggttc tggagaacaa tagcggagcg atatgcaggt 480

aatcccggtt ttgcatttta tgaattgttt aatgaaccga ccgattataa cggtcggctc 540

gggaggatga ccgggatca atataaagaa ttcacgaag agatcattta tataatttat 600

gcacacgacg aaaccgtgat accgcttgta ggctggttcg attggggata tgatctcagg 660

aatgttagag ataatccgat aaatgccccg ggtatcgcgt atgttactca cccgtatccg 720

caaaagcggg accaaccgtg ggaagaaaaa tgggaaaggg atttcggttt cgtagccgac 780

acctaccctg tgtttgctac cgagttcgga tttatgagtg aggatggttt ggggtgcacat 840

attcccggtta tcggtgatga aacatacggg gaagcgatca tcagttactt caatgagaaa 900

ggatatatcgt ggacggcctg ggtgttcgat ccgctctggt cgccgcagct tattaagac 960

tggtatttta ccccgaccgc gcagggacag ttttttaaag agaagctaata ggagtgaat 1020

taa 1023

<210> 68
 <211> 340
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> SIGNAL
 <222> (1)...(23)

<220>
 <221> DOMAIN
 <222> (40)...(317)
 <223> Cellulase (glycosyl hydrolase family 5)

<220>
 <221> SITE
 <222> (143)...(146)
 <223> N-glycosylation site. Prosit id = PS00001

<400> 68
 Met Lys Tyr Ile Phe Ser Tyr Ile Ile Met Met Ile Leu Ile Gly Phe
 1 5 10 15

Ile Pro Val Tyr Gly Phe Gly Asp Ser Pro Asp Gln Thr Tyr Ser Leu
 20 25 30

Pro Phe Leu Ser Val Glu Gly Asn Ser Phe Val Asp Glu Asn Gly Glu

35

40

45

Glu Val Ile Leu Arg Gly Val Ser Phe Pro Asp Pro Asn Arg Leu Asp
50 55 60

Asp Ala Thr Gln Trp Asn Lys Arg Tyr Phe Gln Ala Ala Lys Asp Trp
65 70 75 80

Asn Cys Asn Val Val Arg Ile Pro Val His Pro Gln Arg Trp Arg Glu
85 90 95

Arg Gly Lys Glu Asn Tyr Leu Lys Leu Leu Asp Lys Gly Ile Glu Trp
100 105 110

Ala Gly Glu Leu Gly Met Tyr Val Ile Ile Asp Trp His Thr Ile Gly
115 120 125

Asn Pro Ile Thr Glu Val Phe Phe Gly Glu Leu Tyr Asn Thr Thr Gln
130 135 140

Thr Glu Thr Phe Arg Phe Trp Arg Thr Ile Ala Glu Arg Tyr Ala Gly
145 150 155 160

Asn Pro Val Val Ala Phe Tyr Glu Leu Phe Asn Glu Pro Thr Asp Tyr
165 170 175

Asn Gly Arg Leu Gly Arg Met Thr Trp Asp Gln Tyr Lys Glu Phe Ile
180 185 190

Glu Glu Ile Ile Tyr Ile Ile Tyr Ala His Asp Glu Thr Val Ile Pro
195 200 205

Leu Val Gly Gly Phe Asp Trp Gly Tyr Asp Leu Arg Asn Val Arg Asp
210 215 220

Asn Pro Ile Asn Ala Pro Gly Ile Ala Tyr Val Thr His Pro Tyr Pro

225 230 235 240

Gln Lys Arg Asp Gln Pro Trp Glu Glu Lys Trp Glu Arg Asp Phe Gly
245 250 255

Phe Val Ala Asp Thr Tyr Pro Val Phe Ala Thr Glu Phe Gly Phe Met
260 265 270

Ser Glu Asp Gly Leu Gly Ala His Ile Pro Val Ile Gly Asp Glu Thr
275 280 285

Tyr Gly Glu Ala Ile Ile Ser Tyr Phe Asn Glu Lys Gly Ile Ser Trp
290 295 300

Thr Ala Trp Val Phe Asp Pro Leu Trp Ser Pro Gln Leu Ile Lys Asp
305 310 315 320

Trp Tyr Phe Thr Pro Thr Arg Gln Gly Gln Phe Phe Lys Glu Lys Leu
325 330 335

Met Glu Leu Asn
340

<210> 69

<211> 1182

<212> DNA

<213> Unknown

 $\langle 220 \rangle$

<223> Obtained from environmental sample

<400> 69

atgagtttta aaaaccacat acttttgtcg ctcctcatag tattgctttt cttttcagcg 60

tgcgatatcg aagaaccgat cgccggagat tatcatacac ttgtggatca aaacgctata 120

tcgcacaccc ggcattatt caccaacctc gaacgtatcc gtcacgatca tatectcttc 180

ggatcatcagg atgcgcttgc atacggtgtt cactggcgca acgatgagcc gggtcgatcg 240

gatgtattcg aagtaaccgg ttctgtatcct gcggtgtatg gctgggagat tggcgatatt 300

gaacttgggt caccggaaaa tctggataac gtaaacttcg atcaaatgca gggctggatt 360

cgcaagggt acgaacgagg cggtataatt acgattagct ggcatatgaa caatccggca 420

tcgggtgggt attcgtggga tgtgaatgga ggtcataaag cggtaactaa gatacttccc 480

ggcggagaac ttcacgatac gtttaaagaa tggttgata cgtttgcaa attcggaag 540

agccagattg cttttccga aacaaataat gaacacctta tcccgtcat attccggccg 600

tatcatgaaa acaccggaag ctggttctgg tggggcgccg accactgtac acctgaagaa 660

tataaaaagt tatggcgatt taccgtcgaa tacctgcgcg atgtaaaagg tgttcacaat 720

ctctctggg cgtattcacc tgccggcaat gctgcggatt cagaggaagc atattttgct 780

cggtatcccg gcgacgacta tggtgatatt attggattcg acgattacgg cagtgtgcgg 840

aaaccgtatc aaatcgaacg ttttactaac cggattcgaa cgattgtaa cttcgccgaa 900

gcacgaaata aaatcccggc aataacggaa accggctatg aaactatccc cgatccgcaa 960

tgggtggacgg gtacattgct tagtgcactt gatcacgatt tgacaaccg gagaatagca 1020

tacgtacttg tgtggcgaaa ttcaacaat gctaccgacc ggcagaatca ttattacgt 1080

ccgtatcccg gacatccaag tgctgacgat ttatcgcgt tcaggaatca cccgttgata 1140

gttttcgaag atgatctgcc gggatatgtat aactaccgt aa 1182

<210> 70

<211> 393

<212> PRT

<213> Unknown

<220>

<223> Obtained from environmental sample

<220>
 <221> SIGNAL
 <222> (1)...(20)

<220>
 <221> DOMAIN
 <222> (35)...(351)
 <223> Glycosyl hydrolase family 26

<220>
 <221> SITE
 <222> (355)...(358)
 <223> N-glycosylation site. Prosite id = PS00001

<400> 70
 Met Ser Phe Lys Asn His Ile Leu Leu Ser Leu Leu Ile Val Leu Leu
 1 5 10 15

Phe Phe Ser Ala Cys Asp Ile Glu Glu Pro Ile Ala Gly Asp Tyr His
 20 25 30

Thr Leu Val Asp Gln Asn Ala Ile Ser His Thr Arg Ala Leu Phe Thr
 35 40 45

Asn Leu Glu Arg Ile Arg His Asp His Ile Leu Phe Gly His Gln Asp
 50 55 60

Ala Leu Ala Tyr Gly Val His Trp Arg Asn Asp Glu Pro Gly Arg Ser
 65 70 75 80

Asp Val Phe Glu Val Thr Gly Ser Tyr Pro Ala Val Tyr Gly Trp Glu
 85 90 95

Ile Gly Asp Ile Glu Leu Gly Ala Pro Glu Asn Leu Asp Asn Val Asn
 100 105 110

Phe Asp Gln Met Gln Gly Trp Ile Arg Glu Gly Tyr Glu Arg Gly Gly

115	120	125
Ile Ile Thr Ile Ser Trp His Met Asn Asn Pro Ala Ser Gly Gly Asp		
130	135	140
Ser Trp Asp Val Asn Gly Gly His Lys Ala Val Thr Lys Ile Leu Pro		
145	150	155 160
Gly Gly Glu Leu His Asp Thr Phe Lys Glu Trp Leu Asp Thr Phe Ala		
165	170	175
Lys Phe Ala Lys Ser Gln Ile Ala Phe Pro Glu Thr Asn Asn Glu His		
180	185	190
Leu Ile Pro Val Ile Phe Arg Pro Tyr His Glu Asn Thr Gly Ser Trp		
195	200	205
Phe Trp Trp Gly Ala Asp His Cys Thr Pro Glu Glu Tyr Lys Lys Leu		
210	215	220
Trp Arg Phe Thr Val Glu Tyr Leu Arg Asp Val Lys Gly Val His Asn		
225	230	235 240
Leu Leu Trp Ala Tyr Ser Pro Ala Gly Asn Ala Ala Asp Ser Glu Glu		
245	250	255
Ala Tyr Phe Ala Arg Tyr Pro Gly Asp Asp Tyr Val Asp Ile Ile Gly		
260	265	270
Phe Asp Asp Tyr Gly Ser Val Arg Lys Pro Tyr Gln Ile Glu Arg Phe		
275	280	285
Thr Asn Arg Ile Arg Thr Ile Val Asn Phe Ala Glu Ala Arg Asn Lys		
290	295	300
Ile Pro Ala Ile Thr Glu Thr Gly Tyr Glu Thr Ile Pro Asp Pro Gln		

305 310 315 320

Trp Trp Thr Gly Thr Leu Leu Ser Ala Leu Asp His Asp Leu Thr Thr
325 330 335

Arg Arg Ile Ala Tyr Val Leu Val Trp Arg Asn Ser Asn Asn Ala Thr
340 345 350

Asp Arg Gln Asn His Tyr Tyr Ala Pro Tyr Pro Gly His Pro Ser Ala
355 360 365

Asp Asp Phe Ile Ala Phe Arg Asn His Pro Leu Ile Val Phe Glu Asp
370 375 380

Asp Leu Pro Gly Met Tyr Thr Leu Pro
385 390

<210>	71
<211>	1089
<212>	DNA
<213>	Unknown

<220>
<223> Obtained from environmental sample

<400> 71	
atgaaacttt taaaactttt aatcttttc cttattacgg taatttttc tgatgtttcg	60
gctcaaaactt ttcaaataca aaaaggcaag aacattagcc attggctgtc ccaaagtaaa	120
agaaggggag aagagcgaaa agagtctttt actaagaatg acgtagaatt tattgcaggc	180
atcggtttcg atcatattcg tattcctatt gacgaggagc aaatgtggga tgaaaaaggc	240
aacaaagagc ctgaagcgtt tcagttgctg cacaacgcga tagaatggag caggcaatcg	300
aacttaaaag ttattgtgga cctgcatatt ttgaggtcgc attatttcaa cgcggaagaa	360

aaaccgcttt ttacggacct taaagctcag gaacgttttt accaatgttg ggcggatctg 420

tctggatgaat tgaaaaaata tccgaataga ctgggtggctt atgaattaat gaacgaacct 480

gtagccgatg atccggaaga ctggaataga attgtaagag aatcagtaaa aaggctaagg 540

gtgcttgagc ccaatagggt tattgtaatc gggcttaacc gatggcagca ttatgacact 600

ctgaaggatt tatacgtgcc ggaaaacgac aaaaacatca ttttaagctt tcatttttat 660

aaccctatgt tgcttacgca ttacagggcc agctgggtaa atttcggcga ttaccagggt 720

cccgtaact acccgggaca gttggtagac tcaaagcatt tgtcgggact gagcgaagat 780

ttaagaaaga aagtcgagca aaacaatggc gtttataata aggctcggat tgagaaaatg 840

atagccgaag ccgttgctgt agcaaaaaag cacaacctcc ctttgtattg tggatgaatg 900

ggcgcctacg aaaaagcgcc aaggagccc aggtacaat ggtacagaga catggtggat 960

gtgtgaaca aaaacaatat tgcctggact acctgggact ataaaggagg cttcggcata 1020

gttgacgcca aaggaaacaa agacgaacag ttgatcaatg tattaacagg aaaagagaaa 1080

aaaatgtaa 1089

<210> 72
 <211> 362
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> SIGNAL
 <222> (1)...(21)

<220>
 <221> DOMAIN
 <222> (22)...(340)

<223> Cellulase (glycosyl hydrolase family 5)

<220>

<221> SITE

<222> (31)...(34)

<223> N-glycosylation site. Prosite id = PS00001

<220>

<221> SITE

<222> (154)...(163)

<223> Glycosyl hydrolases family 5 signature. Prosite id = PS00659

<400> 72

Met	Lys	Leu	Leu	Lys	Leu	Leu	Ile	Phe	Leu	Leu	Ile	Thr	Val	Ile	Phe
1				5					10					15	

Ser	Asp	Val	Ser	Ala	Gln	Thr	Phe	Gln	Ile	Gln	Lys	Gly	Lys	Asn	Ile
		20						25						30	

Ser	His	Trp	Leu	Ser	Gln	Ser	Lys	Arg	Arg	Gly	Glu	Glu	Arg	Lys	Glu
		35					40					45			

Phe	Phe	Thr	Lys	Asn	Asp	Val	Glu	Phe	Ile	Ala	Gly	Ile	Gly	Phe	Asp
		50				55						60			

His	Ile	Arg	Ile	Pro	Ile	Asp	Glu	Glu	Gln	Met	Trp	Asp	Glu	Lys	Gly
65					70					75					80

Asn	Lys	Glu	Pro	Glu	Ala	Phe	Gln	Leu	Leu	His	Asn	Ala	Ile	Glu	Trp
				85						90				95	

Ser	Arg	Gln	Ser	Asn	Leu	Lys	Val	Ile	Val	Asp	Leu	His	Ile	Leu	Arg
			100							105				110	

Ser	His	Tyr	Phe	Asn	Ala	Glu	Glu	Lys	Pro	Leu	Phe	Thr	Asp	Pro	Lys
		115						120					125		

Ala	Gln	Glu	Arg	Phe	Tyr	Gln	Cys	Trp	Ala	Asp	Leu	Ser	Gly	Glu	Leu

130

135

140

Lys Lys Tyr Pro Asn Thr Leu Val Ala Tyr Glu Leu Met Asn Glu Pro
145 150 155 160

Val Ala Asp Asp Pro Glu Asp Trp Asn Arg Ile Val Arg Glu Ser Val
165 170 175

Lys Arg Leu Arg Val Leu Glu Pro Asn Arg Val Ile Val Ile Gly Ser
180 185 190

Asn Arg Trp Gln His Tyr Asp Thr Leu Lys Asp Leu Tyr Val Pro Glu
195 200 205

Asn Asp Lys Asn Ile Ile Leu Ser Phe His Phe Tyr Asn Pro Met Leu
210 215 220

Leu Thr His Tyr Arg Ala Ser Trp Val Asn Phe Gly Asp Tyr Gln Gly
225 230 235 240

Pro Val Asn Tyr Pro Gly Gln Leu Val Asp Ser Lys His Leu Ser Gly
245 250 255

Leu Ser Glu Asp Leu Arg Lys Lys Val Glu Gln Asn Asn Gly Val Tyr
260 265 270

Asn Lys Ala Arg Ile Glu Lys Met Ile Ala Glu Ala Val Ala Val Ala
275 280 285

Lys Lys His Asn Leu Pro Leu Tyr Cys Gly Glu Trp Gly Ala Tyr Glu
290 295 300

Lys Ala Pro Arg Glu Pro Arg Leu Gln Trp Tyr Arg Asp Met Val Asp
305 310 315 320

Val Leu Asn Lys Asn Asn Ile Ala Trp Thr Thr Trp Asp Tyr Lys Gly
325 330 335

Gly Phe Gly Ile Val Asp Ala Lys Gly Asn Lys Asp Glu Gln Leu Ile
340 345 350

Asn Val Leu Thr Gly Lys Glu Lys Lys Met
355 360

<210> 73
<211> 1146
<212> DNA
<213> Unknown

<220>
<223> Obtained from environmental sample

<400> 73
gtggatatta ccggacatcc cgaccacatc gccttcgcgc gggaagttgc cgagcaaagc 60

atggtcttgc tgcaaaaccg tgcgaacctc gcccccttt cggtatctga ctattccacc 120

attgccgtga tcggcccgaa tgccaatgac actttgctgg gttcttacag cggcgttccg 180

aaaacctact acacggtact cgacgggata cggtcctatg tcggtgaccg ggcgaatgtg 240

gtttacgtc aggggccgaa gataacaaa cccggccatc gggaggacaa tgaagtattt 300

ccaccggatc ctgaaaacga ccggagacga ctggccgaag cgatagctgt cgccgagaac 360

gccgacctga tcatcctcgc gatcggcggc aatgaactga cgggacgaga ggcatgggcg 420

gcgcatcatc ccggtgatcg accggatctg tcgttgctcg gtttcagga ggatcttgtt 480

gacgcagttg gagcgatggg ggttccatct gtcgcattgg ttttcggtgc acggccgctg 540

gacctggca atgtcgcca aaaaattgat gtggtcttcc aaaactgta cctgggccag 600

gaaaccggca atgccgtcgc caatgtgctg ttggcgagg tgtcacctc cgccaaactc 660

```

cccatcagct tcccgcggac tgccgggcac attcctgcct actacaatta caaaccatcg      720

gtctgcacggg tctacctttt tgacgatgtc actccgcgtt accatttcgg gtacggcctc      780

agctatacga cgtttgaata cggggaaccg cagctatcgg atacactact gtctggcgat      840

ggtgaaataa ccctctacgt tgaagttacc aacaccggag agcgaggcgg ttcggaagtc      900

gtgcaactgt acatcaacca cgaatacaga tccgtcacc ggccggtaaa ggagctcaag      960

ggattcgaaa aggtgtatct cgagccgaat gaaactgccg gtgtatcggt caccatcact     1020

tcagatcagt tgaggttctg gaatatcgac atggagtta cgcgtgaatc cggtaaagtg     1080

aacctgatgg tcggctcatc cagccgtgac gaagacctgc agacgacggc aatttttctt     1140

gaataa                                           1146

```

<210> 74
 <211> 381
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> DOMAIN
 <222> (21)...(264)
 <223> Glycosyl hydrolase family 3 C terminal domain

<220>
 <221> SITE
 <222> (49)...(52)
 <223> N-glycosylation site. Prosite id = PS00001

<220>
 <221> SITE
 <222> (335)...(338)
 <223> N-glycosylation site. Prosite id = PS00001

<400> 74
 Met Asp Ile Thr Gly His Pro Asp His Ile Ala Phe Ala Arg Glu Val

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

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

<210> 75
<211> 1014

<212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 75

atgctgcgca agttgatcgt ctcggtcttc ggcttcgtca tgctgactag tgcggcagcg	60
gcgagactc ctccgcctt agcggaatcc ggcctgtc tccggcgcg aatgaacgtt	120
ctgggctacg acccaatctg gcacgacccg aagaaaggtc ggctcgaaga gcggcacttc	180
gccgagattc gcaagggcgg cttcgacttc gtccgggtga acctccacgg gttcaacat	240
atgaagccg cggacaaact cagtccggag ttcttgagcc gcgtggactg gatcgtgaag	300
cagccagtgc cgcggggcct gtcggtcatc ctagacgagc atgaatatga ggaatgctcg	360
gacgagctgc caatgtgccg gcggcggttg gcggcattct ggacgcaggt cgcgccgcgc	420
tacaagggcg cgccgatac ggttctgttc gagcttctca atgagccgca cgacaagttg	480
gatgccgaca cctggaacgc cttgtttccc gacatcctgg ccatcgtgcg gcagtcgaac	540
ccgaagcgcc gcgtggtgat cgcccgact cagtggaaca acttcagcca gctggacacg	600
ctcaagctgc cggcagacga ccggaacatc gtcgtcacct tccattatta cgatccgttc	660
ccgtttacc accagggcgc gccgtgggtt ccggacatgc tcaaggtgaa aggcacgag	720
tggaagcccc agcagaggcg gaagatcgcc gaggacttcg gcaaggtcgc ggaatggtcg	780
cagaaaaccg gccgcgaaat cttgtctggc gagttcgggg cctacgatgt gagcggtacg	840
ccaaccgcca tgcgttcagc ttatacgga gcggtggcgc gcgaggcgga acgccacggc	900
ttcgcttggg cctactggca gttcgacagc aatttcctgg cttgggacat gaagacaaac	960

ggctgggtcg agccgatcca caaggcactc atccccgagg cgaagcagcc ttag 1014

<210> 76
<211> 337
<212> PRT
<213> Unknown

<220>
<223> Obtained from environmental sample

<220>
<221> DOMAIN
<222> (37)...(316)
<223> Cellulase (glycosyl hydrolase family 5)

<220>
<221> SITE
<222> (150)...(159)
<223> Glycosyl hydrolases family 5 signature. Prosite id = PS00659

<400> 76
Met Leu Arg Lys Leu Ile Val Ser Val Phe Gly Phe Val Met Leu Thr
1 5 10 15

Ser Ala Ala Ala Ala Gln Thr Pro Pro Ala Leu Ala Glu Ser Ala Pro
20 25 30

Ala Leu Arg Arg Gly Met Asn Val Leu Gly Tyr Asp Pro Ile Trp His
35 40 45

Asp Pro Lys Lys Gly Arg Phe Glu Glu Arg His Phe Ala Glu Ile Arg
50 55 60

Lys Gly Gly Phe Asp Phe Val Arg Val Asn Leu His Gly Phe Lys His
65 70 75 80

Met Asn Ala Ala Asp Lys Leu Ser Pro Glu Phe Leu Ser Arg Val Asp
85 90 95

Trp Ile Val Lys His Ala Ser Ala Ala Gly Leu Ser Val Ile Leu Asp

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

290 295 300

Tyr Trp Gln Phe Asp Ser Asn Phe Leu Ala Trp Asp Met Lys Thr Asn
305 310 315 320

Gly Trp Val Glu Pro Ile His Lys Ala Leu Ile Pro Glu Ala Lys Gln
325 330 335

Pro

<210> 77
<211> 1125
<212> DNA
<213> Unknown

<220>
<223> Obtained from environmental sample

<400> 77
atgaaaagga aacgggtttt tattcattct ctaatcgtat tttttttaat gattggttct 60

tttacttctt gtggatcagt cgccgatgat gccgaagaag ggtttgatat ttttagagga 120

accaatatcg ctcatagggt atcacaaagt aatgcaaggg gcgaagagcg aaaaaatttc 180

tttaccgaaa atgatataaa atttattgct gatgctgggt ttgatcatat tcgtttgcc 240

attgacgagg ttcatctctg ggatgagaat atgaaccggc accaagatgc atttgatctt 300

atgcatgact gtattaagtg gtcagagaaa catggtctta gggttgtagt ggatttgc 360

attattcggt cacattatct tgttgagat gataatacac tatgggatga aagacatgaa 420

caggaagagt ttgttgatat ttggatggag ttatcatctg aactatctca atattcaa 480

tcattagtag cttatgagtt aatgaatgaa cctgtagccc cttctcatga tgattgga 540

agtttggttg cggaactat agaggcaatt cgtaaagttg aacctgagag atatattgta 600

gttggatcaa atatgtggca aggtattgat acatttgagt atttgaagt tcccgaat 660

gatgatagaa taattcttag ttttcatttt tatgatccct ttattttgac tcattatact 720

gcatcttggg ggiatttaag agattactca gggcctgtta actatccggg atatcttgtt 780

acaaatgacc agctgttga tatgtcaaac gaaatgcaa agttaattag ggagtttcag 840

acaaattttg atatttatac cattgaagaa ctgatatcta ttccatatag tattgcaaag 900

gaaaaagggt tgaattata ttgtggagag ttggtgcaa ttgatcaggc tccaagagat 960

gcgagattgg catggtacag agatgttgtt caggtttttg agcgatatgg tatagctcat 1020

gccaaactgga attacaaaga ttatggtacg ttgggataa agaactatag cgaggagata 1080

gatcaggaac tgtttgaaat cttaattgga aaaaacata aatag 1125

<210> 78
<211> 374
<212> PRT
<213> Unknown

<220>
<223> Obtained from environmental sample

<220>
<221> SIGNAL
<222> (1)...(28)

<220>
<221> DOMAIN
<222> (25)...(353)
<223> Cellulase (glycosyl hydrolase family 5)

<220>
<221> SITE
<222> (165)...(174)
<223> Glycosyl hydrolases family 5 signature. Prosite id = PS00659

<220>

<221> SITE
 <222> (360)...(363)
 <223> N-glycosylation site. Prosite id = PS00001

<400> 78

Met Lys Arg Lys Arg Val Phe Ile His Ser Leu Ile Val Phe Phe Leu
 1 5 10 15

Met Ile Gly Ser Phe Thr Ser Cys Gly Ser Val Ala Asp Asp Ala Glu
 20 25 30

Glu Gly Phe Asp Ile Phe Arg Gly Thr Asn Ile Ala His Trp Leu Ser
 35 40 45

Gln Ser Asn Ala Arg Gly Glu Glu Arg Lys Asn Phe Phe Thr Glu Asn
 50 55 60

Asp Ile Lys Phe Ile Ala Asp Ala Gly Phe Asp His Ile Arg Leu Pro
 65 70 75 80

Ile Asp Glu Val His Phe Trp Asp Glu Asn Met Asn Arg His Gln Asp
 85 90 95

Ala Phe Asp Leu Met His Asp Cys Ile Lys Trp Ser Glu Lys His Gly
 100 105 110

Leu Arg Val Val Val Asp Leu His Ile Ile Arg Ser His Tyr Phe Val
 115 120 125

Gly Asp Asp Asn Thr Leu Trp Asp Glu Arg His Glu Gln Glu Lys Phe
 130 135 140

Val Asp Ile Trp Met Glu Leu Ser Ser Glu Leu Ser Gln Tyr Ser Asn
 145 150 155 160

Ser Leu Val Ala Tyr Glu Leu Met Asn Glu Pro Val Ala Pro Ser His
 165 170 175

Asp Asp Trp Asn Ser Leu Val Ala Glu Thr Ile Glu Ala Ile Arg Lys
180 185 190

Val Glu Pro Glu Arg Tyr Ile Val Val Gly Ser Asn Met Trp Gln Gly
195 200 205

Ile Asp Thr Phe Glu Tyr Leu Glu Val Pro Glu Asn Asp Asp Arg Ile
210 215 220

Ile Leu Ser Phe His Phe Tyr Asp Pro Phe Ile Leu Thr His Tyr Thr
225 230 235 240

Ala Ser Trp Gly Tyr Leu Arg Asp Tyr Ser Gly Pro Val Asn Tyr Pro
245 250 255

Gly Tyr Leu Val Thr Asn Asp Gln Leu Leu Asp Met Ser Asn Glu Met
260 265 270

Gln Lys Leu Ile Arg Glu Phe Gln Thr Asn Phe Asp Ile Tyr Thr Ile
275 280 285

Glu Glu Leu Ile Ser Ile Pro Tyr Ser Ile Ala Lys Glu Lys Gly Leu
290 295 300

Lys Leu Tyr Cys Gly Glu Phe Gly Ala Ile Asp Gln Ala Pro Arg Asp
305 310 315 320

Ala Arg Leu Ala Trp Tyr Arg Asp Val Val Gln Val Phe Glu Arg Tyr
325 330 335

Gly Ile Ala His Ala Asn Trp Asn Tyr Lys Asp Tyr Gly Thr Phe Gly
340 345 350

Ile Lys Asn Tyr Ser Glu Glu Ile Asp Gln Glu Leu Phe Glu Ile Leu
355 360 365

Ile Gly Thr Lys His Lys
370

<210> 79
<211> 1017
<212> DNA
<213> Unknown

<220>
<223> Obtained from environmental sample

<400> 79
atgaaatata aagctatitt tatatacctt attgttttga ttctatttta ctcaattaat 60

atttatgcta atgcagaaaa caacccctc cccttcctca gtgtcgaagg aaacaggttc 120

gtcgaatgaag atggaaatag ggtaatcctg cgaggtgtat cgttccccga tcccgaaccg 180

ctggctgagg caactcaatg gaacaagcga tacttccagg cggcaaaaga ctggaactgt 240

aatgtcgtcc ggattcctgt ccatccacag aaatggcggg aaagaggcga ggaaaattat 300

ctgaaacttt tagataaggg aattcaatgg gcgggtgaac tcgggatgta tgtaatcatc 360

gactggcata ccacggtaa tccgataacc gaagtatttt tccgcgaact atacaatacg 420

tcacgtgcgg agaccttcca gttctggaga acaatcgtg agcgtatgc cggtaaccgc 480

gttgttgctt tctatgaact gttcaatgaa ccgaccgact acaacggccg tctcggaaga 540

atgaactggg atcagtataa agagtttacc gaggagataa ttacatcat ctattctcac 600

gacgatacag ttatccctct cgttgccggt ttcgactggg cgtatgaact ccgcatata 660

aaagataaac ctatagattt tcccggcatc gcttatgtga ctaccccta tcccagaaa 720

cgcgatccgc catgggaaga aaaatgggaa gaggatttcg ggtttgccgc cgatatgtat 780

ccggtgtttg caaccgagtt cggtttcatg ggggaggatg aattaggtgc acacataccc 840

gtcatcggcg atgaaacata cggcgaagcc attatcgatt acttttataa aaaggggata 900

tcgtggactg catgggtatt cgatccgctt tggtcgccgc agcttattag agactggtat 960

tttaccctgt cccgacaggg gcagtttttt aaagagaagt tgatggagtt gaattag 1017

<210> 80
<211> 338
<212> PRT
<213> Unknown

<220>
<223> Obtained from environmental sample

<220>
<221> SIGNAL
<222> (1)...(25)

<220>
<221> DOMAIN
<222> (38)...(315)
<223> Cellulase (glycosyl hydrolase family 5)

<220>
<221> SITE
<222> (141)...(144)
<223> N-glycosylation site. Prosite id = PS00001

<400> 80
Met Lys Tyr Lys Ala Ile Phe Ile Tyr Leu Ile Val Leu Ile Leu Phe
1 5 10 15

Tyr Ser Ile Asn Ile Tyr Ala Asn Ala Glu Asn Asn Pro Leu Pro Phe
20 25 30

Leu Ser Val Glu Gly Asn Arg Phe Val Asp Glu Asp Gly Asn Thr Val
35 40 45

Ile Leu Arg Gly Val Ser Phe Pro Asp Pro Asp Arg Leu Ala Glu Ala

50

55

60

Thr Gln Trp Asn Lys Arg Tyr Phe Gln Ala Ala Lys Asp Trp Asn Cys
65 70 75 80

Asn Val Val Arg Ile Pro Val His Pro Gln Lys Trp Arg Glu Arg Gly
85 90 95

Glu Glu Asn Tyr Leu Lys Leu Leu Asp Lys Gly Ile Gln Trp Ala Gly
100 105 110

Glu Leu Gly Met Tyr Val Ile Ile Asp Trp His Thr Ile Gly Asn Pro
115 120 125

Ile Thr Glu Val Phe Phe Arg Glu Leu Tyr Asn Thr Ser Arg Ala Glu
130 135 140

Thr Phe Gln Phe Trp Arg Thr Ile Ala Glu Arg Tyr Ala Gly Asn Pro
145 150 155 160

Val Val Ala Phe Tyr Glu Leu Phe Asn Glu Pro Thr Asp Tyr Asn Gly
165 170 175

Arg Leu Gly Arg Met Asn Trp Asp Gln Tyr Lys Glu Phe Ile Glu Glu
180 185 190

Ile Ile His Ile Ile Tyr Ser His Asp Asp Thr Val Ile Pro Leu Val
195 200 205

Ala Gly Phe Asp Trp Ala Tyr Glu Leu Arg His Ile Lys Asp Lys Pro
210 215 220

Ile Asp Phe Pro Gly Ile Ala Tyr Val Thr His Pro Tyr Pro Gln Lys
225 230 235 240

Arg Asp Pro Pro Trp Glu Glu Lys Trp Glu Glu Asp Phe Gly Phe Ala

245

250

255

Ala Asp Met Tyr Pro Val Phe Ala Thr Glu Phe Gly Phe Met Gly Glu
260 265 270

Asp Glu Leu Gly Ala His Ile Pro Val Ile Gly Asp Glu Thr Tyr Gly
275 280 285

Glu Ala Ile Ile Asp Tyr Phe Tyr Lys Lys Gly Ile Ser Trp Thr Ala
290 295 300

Trp Val Phe Asp Pro Leu Trp Ser Pro Gln Leu Ile Arg Asp Trp Tyr
305 310 315 320

Phe Thr Pro Ser Arg Gln Gly Gln Phe Phe Lys Glu Lys Leu Met Glu
325 330 335

Leu Asn

<210> 81

<211> 1119

<212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 81

atgaatttac ttgctcaata cttttccgga ctatttctga tttttttgat ctcaattttt 60

ttcgtaggtt ctgcagcgaa tcatcattat gaaaaaata cagtcaacga attttctgat 120

gatgtaaatc aaacaacatt agtccttcaa cccgggatat ccgaagccca gaataactcaa 180

aacctgccgc ggatttcggt tgaaggaaac caatttgtgg atgaatcggg aaacacagtc 240

acatttcagg gtgtcagttg tgccgatccg cacaggctta ataatgccgg ccaatggaaa 300

cgggaactgt ttgaagaaat cgcaactgg ggagcaaacg tcgttcgtct gcccatacac 360
 ccgctctggt ggcggaacg gggagaggag caatacctcg aatggattga tgaagccgtg 420
 gagtgggcca aagagctgga gatgtacctc atcatcgact ggcacagtat cgggaacctg 480
 cggacagaac tcttttcag ggatatctac aacaccaccc gccgtgaaac ttatgaattc 540
 tggagctga tttcgatcg ctatgctgat gaaaccacaa ttgcctttta cgaaatcttt 600
 aatgaacca cacggcagca gggcaggctg ggaaccatga cctggaagca atggaaggaa 660
 attctaaccg acattatcac aatcatttat gccacaatc ctgatgcgat tccgctggta 720
 gcaggtttta actgggcgta tgaccttact ccggtccgcc actcaccct cgattttgaa 780
 ggtattgcct atgttaccca cccatatccg caaaaaagaa gcaggccctg ggttccaaaa 840
 tgggaagaag atttcggttt tgtgctgac aaatatectg tatttgccac tgaattcggc 900
 tatatgaggg agtatgagcg gggcgctcat gtgcccgtaa tcggggacga agaatatggg 960
 gaaatcctca tcaattatit ccgcgaaaaa gggatttcgt ggacagcctg ggtattcgat 1020
 ccaagctggt cgccacagct cattcaggat tgggattata caccacacg ctcaggtgag 1080
 ttttcagaa atgcatgag aacgaaaaac aatgaataa 1119

<210> 82
 <211> 372
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> SIGNAL
 <222> (1)...(25)

<220>
 <221> DOMAIN
 <222> (70)...(347)
 <223> Cellulase (glycosyl hydrolase family 5)

<220>
 <221> SITE
 <222> (43)...(46)
 <223> N-glycosylation site. Prosite id = PS00001

<220>
 <221> SITE
 <222> (173)...(176)
 <223> N-glycosylation site. Prosite id = PS00001

<400> 82
 Met Asn Leu Leu Ala Gln Tyr Phe Ser Gly Leu Phe Leu Ile Phe Leu
 1 5 10 15

Ile Ser Ile Phe Phe Val Ser Ser Ala Ala Asn His His Tyr Glu Lys
 20 25 30

Asn Thr Val Asn Glu Phe Ser Asp Asp Val Asn Gln Thr Thr Leu Val
 35 40 45

Leu Gln Pro Gly Ile Ser Glu Ala Gln Asn Thr Gln Asn Leu Pro Arg
 50 55 60

Ile Ser Val Glu Gly Asn Gln Phe Val Asp Glu Ser Gly Asn Thr Val
 65 70 75 80

Thr Phe Gln Gly Val Ser Val Ala Asp Pro His Arg Leu Asn Asn Ala
 85 90 95

Gly Gln Trp Lys Arg Glu Leu Phe Glu Glu Ile Ala Asn Trp Gly Ala
 100 105 110

Asn Val Val Arg Leu Pro Ile His Pro Leu Trp Trp Arg Glu Arg Gly

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

305 310 315 320

Glu Ile Leu Ile Asn Tyr Phe Arg Glu Lys Gly Ile Ser Trp Thr Ala
325 330 335

Trp Val Phe Asp Pro Ser Trp Ser Pro Gln Leu Ile Gln Asp Trp Asp
340 345 350

Tyr Thr Pro Thr Arg Ser Gly Glu Phe Phe Arg Asn Ala Met Arg Thr
355 360 365

Lys Asn Asn Glu
370

<210>	83
<211>	1089
<212>	DNA
<213>	Unknown

<220>
<223> Obtained from environmental sample

<400> 83	
atgagccttg gctgactgc aatcgagttg atcaatcgcg cccgcgccga tctgcgactg	60
ggcgtgccga tcgttctgcg cgagggcgac gtgcaggcgc tgggtctggc ggtcgagcca	120
gtaaccgagg cgcggctggg tgggctgcbc gggctggggc cagggctggt gcttgcaatc	180
acgcagcgcc gcgccacgac actgaaggcg cgcgcctatg atgaggatct tgcgcgagtg	240
gtggtgcccg agggggtagg ctgcgactgg ctgcgggcgg tggcggaccc ctccgacgat	300
ctgcgctttc cgatgaaggg cccgctgatg accgctcgcg agggcacggc cgcgctgcat	360
cgcgctgcac ttcaactggt gaaatccgcg cagcttcttc cggccgcact tgttcagccg	420
cttgcggatc ccgaggcgct gcccgtcacg gggctgacag tgctcgatat cgccgatgtc	480

agccgtgaat tggcgcgcga gacagtgttg tatccagtgg tgcacgcgcg ctgtccgatg 540

ctggcggcgc aagcgggccc cggtgcatac ttccgacccc gcgacggcgg cgttgagcat 600

tacgccatcg agatcggccg gcccgaaccgt gccgcggccc tgctcacgcg gctgcattcg 660

gcctgtttca caggcgatgt gctgggctcg ctcaaatgcg attgcggccc gcaactgcag 720

gcagcactcg cgcagatggg cgaggaaggc gcgggggtgc tgctctatct caatcaggag 780

ggtcgcggca tcgggcttgc caacaagatg cgcgcctatt cgctgcagga tcagggtttt 840

gacacggtcg aggccaatca ccgtctgggg ttcgaggatg acgagcggga ttccgcac 900

ggggccgcgc ttctgcggcg gatgggggtt tctcgggcgc ggctgctgac caacaaccct 960

cggaaggtga acatgctgaa tgcgcatcgg gtcgaagtgg tggaacgggt gccgcttcgg 1020

gtgggcgaga cggtcgagaa ccgcgcctat cttgccacca aggccgcaa atccgggcat 1080

ctgtttgtga 1089

<210> 84
 <211> 362
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> DOMAIN
 <222> (168)...(339)
 <223> GTP cyclohydrolase II

<400> 84
 Met Ser Leu Gly Leu Thr Ala Ile Glu Leu Ile Asn Arg Ala Arg Ala
 1 5 10 15

Asp Leu Arg Leu Gly Val Pro Ile Val Leu Arg Glu Gly Asp Val Gln
 20 25 30

Ala Leu Val Leu Ala Val Glu Pro Val Thr Glu Ala Arg Leu Gly Gly
35 40 45

Leu Arg Gly Leu Gly Pro Gly Leu Val Leu Ala Ile Thr Gln Arg Arg
50 55 60

Ala Thr Thr Leu Lys Ala Arg Ala Tyr Asp Glu Asp Leu Ala Arg Val
65 70 75 80

Val Val Pro Glu Gly Val Gly Cys Asp Trp Leu Arg Ala Val Ala Asp
85 90 95

Pro Ser Asp Asp Leu Arg Phe Pro Met Lys Gly Pro Leu Met Thr Ala
100 105 110

Arg Glu Gly Thr Ala Ala Leu His Arg Ala Ala Leu Gln Leu Val Lys
115 120 125

Ser Ala Gln Leu Leu Pro Ala Ala Leu Val Gln Pro Leu Ala Asp Pro
130 135 140

Glu Ala Leu Pro Val Thr Gly Leu Thr Val Leu Asp Ile Ala Asp Val
145 150 155 160

Ser Arg Glu Leu Ala Arg Glu Thr Val Leu Tyr Pro Val Val His Ala
165 170 175

Arg Leu Pro Met Leu Ala Ala Gln Ala Gly Arg Val His Ile Phe Arg
180 185 190

Pro Arg Asp Gly Gly Val Glu His Tyr Ala Ile Glu Ile Gly Gln Pro
195 200 205

Asp Arg Ala Ala Pro Val Leu Thr Arg Leu His Ser Ala Cys Phe Thr

210

215

220

Gly Asp Val Leu Gly Ser Leu Lys Cys Asp Cys Gly Pro Gln Leu Gln
225 230 235 240

Ala Ala Leu Ala Gln Met Gly Glu Glu Gly Ala Gly Val Leu Leu Tyr
245 250 255

Leu Asn Gln Glu Gly Arg Gly Ile Gly Leu Ala Asn Lys Met Arg Ala
260 265 270

Tyr Ser Leu Gln Asp Gln Gly Phe Asp Thr Val Glu Ala Asn His Arg
275 280 285

Leu Gly Phe Glu Asp Asp Glu Arg Asp Phe Arg Ile Gly Ala Ala Leu
290 295 300

Leu Arg Arg Met Gly Phe Ser Arg Ala Arg Leu Leu Thr Asn Asn Pro
305 310 315 320

Arg Lys Val Asn Met Leu Asn Ala His Arg Val Glu Val Val Glu Arg
325 330 335

Val Pro Leu Arg Val Gly Glu Thr Val Glu Asn Arg Ala Tyr Leu Ala
340 345 350

Thr Lys Ala Ala Lys Ser Gly His Leu Leu
355 360

<210> 85

<211> 1284

<212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 85

gtgaacaccg cgcatcgcat cgaattccct cggcaattta tcttcgggtc cgccactgct	60
gctcaccaag tggagggcaa caacgttcac aatgattggt gggcccacga gcatgccacc	120
gacacgaatg ccgtggagcc gtccggcctc gcctgcgacc actttcggcg ctttgccgac	180
gacttcgcc tcttacgcca actcggacag ccagcgcacc gcctgtcgtt ggaatggagc	240
cgcatcgaa cggcaccggt tgaaatcgat cgttcggcat tgtcccacta ccgccgagtc	300
ctgggtactt tgcgagacct cggaatcgag ccatgggtca ccatccacca cttcacttgc	360
cctcgtggt tcttggaaca gggagggttt acacgcatgg attcagcgcg ctctctcgtt	420
cgccataccg aacgcgtggc gagggagtgc tccgacctag tcacaaactg gtgcaccata	480
aatgagccaa acgtcgtggc agaactcgtt tatcgcttcg gatactttcc gccgcggttg	540
caggacgatg agctggcagc ggaagtgtc accaacttct ttcgcttaca cgctgaaatg	600
gcagaagttt tgcgcgtca cgcgcagaga tcggcgcaaa tcggtatcac ctttgcatg	660
caagcacacg agccgttcg catcgaaagc gaagcggacc gcgcactggc ggcgcggcgc	720
gacgccgaga ccaacggcgt catgtcaac gccttgcgaa ccggtgtatt cgcctaccg	780
ggacgggagc cgttggaat ccctggactg aaaacgtcat cgaccttcgt ggggtccag	840
tactattcgc ggttcgcta cgacccgag tcgcaaggtc cagcaatgcc cgacttcgag	900
cgcacctca gccaaatggg atgggagggt tatcctgagg ggttcggccc cttgctcgag	960
cgcgcagcag aaactggact cgaagtgatc gtcacagaga acgggatggc gcacgacgat	1020
gaccgtgtgc gcgtgcgttt tatcgccgac cacttgctgg tcgttcaccg cttcttgaa	1080
cgcggtgtgc gcatcggagg gtacttttac tggtcgacca tggacaactt cgaatggaac	1140
ttcgggtacg gaccgaagtt cggcctgatc gaagtggacc gttctaccct ggaacgcagg	1200

ccgcgcgcaa gcgcgtatit ctcccgtagac atgatccagc agcgagtgcg cgacgacgac 1260

ctgggtcgagc actggactcg ctga 1284

<210> 86
<211> 427
<212> PRT
<213> Unknown

<220>
<223> Obtained from environmental sample

<220>
<221> DOMAIN
<222> (5)...(417)
<223> Glycosyl hydrolase family 1

<220>
<221> SITE
<222> (334)...(342)
<223> Glycosyl hydrolases family 1 active site. Prosite id = PS00572

<400> 86
Met Asn Thr Ala His Arg Ile Glu Phe Pro Arg Gln Phe Ile Phe Gly
1 5 10 15

Ser Ala Thr Ala Ala His Gln Val Glu Gly Asn Asn Val His Asn Asp
20 25 30

Trp Trp Ala His Glu His Ala Thr Asp Thr Asn Ala Val Glu Pro Ser
35 40 45

Gly Leu Ala Cys Asp His Phe Arg Arg Phe Ala Asp Asp Phe Arg Leu
50 55 60

Leu Arg Gln Leu Gly Gln Pro Ala His Arg Leu Ser Leu Glu Trp Ser
65 70 75 80

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

85

90

95

Tyr Arg Arg Val Leu Gly Thr Leu Arg Asp Leu Gly Ile Glu Pro Trp
100 105 110

Val Thr Ile His His Phe Thr Cys Pro Arg Trp Phe Val Glu Gln Gly
115 120 125

Gly Phe Thr Arg Met Asp Ser Ala Arg Ser Leu Val Arg His Thr Glu
130 135 140

Arg Val Ala Arg Glu Phe Ser Asp Leu Val Thr Asn Trp Cys Thr Ile
145 150 155 160

Asn Glu Pro Asn Val Val Ala Glu Leu Gly Tyr Arg Phe Gly Tyr Phe
165 170 175

Pro Pro Arg Leu Gln Asp Asp Glu Leu Ala Ala Glu Val Leu Thr Asn
180 185 190

Phe Phe Arg Leu His Ala Glu Met Ala Glu Val Leu Arg Ala His Ala
195 200 205

Gln Arg Ser Ala Gln Ile Gly Ile Thr Leu Ala Met Gln Ala His Glu
210 215 220

Pro Leu Arg Ile Glu Ser Glu Ala Asp Arg Ala Leu Ala Ala Arg Arg
225 230 235 240

Asp Ala Glu Thr Asn Gly Val Met Leu Asn Ala Leu Arg Thr Gly Val
245 250 255

Phe Ala Tyr Pro Gly Arg Glu Pro Val Glu Ile Pro Gly Leu Lys Thr
260 265 270

Ser Ser Thr Phe Val Gly Val Gln Tyr Tyr Ser Arg Val Arg Tyr Asp

275 280 285

Ala Glu Ser Gln Gly Pro Ala Met Pro Asp Phe Glu Arg Thr Leu Ser

290 295 300

Gln Met Gly Trp Glu Val Tyr Pro Glu Gly Phe Gly Pro Leu Leu Glu
305 310 315 320

Arg Ala Ala Glu Thr Gly Leu Glu Val Ile Val Thr Glu Asn Gly Met
325 330 335

Ala His Asp Asp Asp Arg Val Arg Val Arg Phe Ile Ala Asp His Leu
340 345 350

Arg Val Val His Arg Leu Leu Glu Arg Gly Val Arg Ile Gly Gly Tyr
355 360 365

Phe Tyr Trp Ser Thr Met Asp Asn Phe Glu Trp Asn Phe Gly Tyr Gly
370 375 380

Pro Lys Phe Gly Leu Ile Glu Val Asp Arg Ser Thr Leu Glu Arg Arg
385 390 395 400

Pro Arg Arg Ser Ala Tyr Phe Phe Arg Asp Met Ile Gln Gln Arg Val
405 410 415

Leu Asp Asp Asp Leu Val Glu His Trp Thr Arg
420 425

<210>	87
<211>	1167
<212>	DNA
<213>	Unknown

<220>
<223> Obtained from environmental sample

<400> 87

atgagaaaga gtgtgttcac cctcgccgtg tttttgtcgg cactgtttgc attcacgtct	60
tgtcagaaca agagccagaa cgaggctcaa gaccaggcag gacaagtcaa taacttccgc	120
atcaagcgcg gcacgaacat cagccactgg ctgtcgagc cggagcagcg cggtaggct	180
cgcagactgc atatccagga ggacgacttc gcccgctctgg aagagctggg cttcgacttc	240
gtgcgcatcc ccatcgacga ggtgcagtgc tgggacgagc agggcaacaa gctgcccag	300
gcgtgggagc tgctgaacaa cgccctcgac tggagcaaga agcacaacct gcgtgccatc	360
gtcgacctgc acatcatccg tgcgcactat ttcaatgccg tgaatgagc agaccaggcc	420
gccaatacc tcttcacctc tgaggaggca caggaaggac tcctaacct gtggcgccag	480
ctctccgagt tctgaagga ccgcagcaac gactgggtgg cctacgagtt catgaacgag	540
ccggtagccc ctgagcacga gatgtggaac cagctggtag ccaaggtaca caaggccctg	600
cgcgaactgg aaccccgagc tacactcgtc gtcggctcga acatgtggca gggacacgag	660
acgatgaagt atctgaaagt gcccaggggc gataagaaca tcctcctctc gttccactac	720
tacaaccga tgctgctgac gcactacggt gcctggtggt cgccgctgtg tgctgcctac	780
aagggtgaag tgaactatcc cgggtgtgtc gtgtcgaagg aagactacga tgccgtcct	840
gtgcatca aggatcagct gaagcccttt accgaggaag tatggaacat cgacaagatc	900
cgtgagcagt tcaaggatgc catcgaggcc gccaaagaa atgacctgca actgttctgc	960
ggcgagtggg gtgtctatga gcccgtaggac cgtgagctgg cctacaaatg gtatcgtgac	1020
gtgtgacgg tggtcgacga gttcaacatc gcctggacga cctggtgcta cgatgctgac	1080
ttcggtttct gggatcagca gcgcactgc taaaagact atccgctggt ggagctcctg	1140
atgtcaggaa agaaactggg agaataag	1167

<210> 88
 <211> 388
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> SIGNAL
 <222> (1)...(23)

<220>
 <221> DOMAIN
 <222> (48)...(365)
 <223> Cellulase (glycosyl hydrolase family 5)

<220>
 <221> SITE
 <222> (23)...(26)
 <223> N-glycosylation site. Prosit id = PS00001

<220>
 <221> SITE
 <222> (46)...(49)
 <223> N-glycosylation site. Prosit id = PS00001

<400> 88
 Met Arg Lys Ser Val Phe Thr Leu Ala Val Phe Leu Ser Ala Leu Phe
 1 5 10 15

Ala Phe Thr Ser Cys Gln Asn Lys Ser Gln Asn Glu Ala Gln Asp Gln
 20 25 30

Ala Gly Gln Val Asn Asn Phe Arg Ile Lys Arg Gly Thr Asn Ile Ser
 35 40 45

His Trp Leu Ser Gln Ser Glu Gln Arg Gly Glu Ala Arg Arg Leu His
 50 55 60

Ile Gln Glu Asp Asp Phe Ala Arg Leu Glu Glu Leu Gly Phe Asp Phe

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

260 265 270

Lys Glu Asp Tyr Asp Ala Ala Pro Ala Ala Ile Lys Asp Gln Leu Lys
275 280 285

Pro Phe Thr Glu Glu Val Trp Asn Ile Asp Lys Ile Arg Glu Gln Phe
290 295 300

Lys Asp Ala Ile Glu Ala Ala Lys Lys Tyr Asp Leu Gln Leu Phe Cys
305 310 315 320

Gly Glu Trp Gly Val Tyr Glu Pro Val Asp Arg Glu Leu Ala Tyr Lys
325 330 335

Trp Tyr Arg Asp Val Leu Thr Val Phe Asp Glu Phe Asn Ile Ala Trp
340 345 350

Thr Thr Trp Cys Tyr Asp Ala Asp Phe Gly Phe Trp Asp Gln Gln Arg
355 360 365

His Cys Tyr Lys Asp Tyr Pro Leu Val Glu Leu Leu Met Ser Gly Lys
370 375 380

Lys Leu Gly Glu
385

<210> 89
<211> 1500
<212> DNA
<213> Unknown

<220>
<223> Obtained from environmental sample

<400> 89
atgaaacgtt cagtctctat ctttatcgca tgtttattaa tgacagtatt aacaattagc 60

gggtgtcgcg caccagaagc atctgcagca ggggcgaaaa cgctgtagc ccttaatggc 120

cagcttagca ttaaaggtag tcagctagtc aatcaaacg gaaaaccggt gcagctgaag 180

gggatcagct cacacggttt gcagtgggtt gccgattatg tcaataaaga cactttaaaa 240

tggctaagag acgattgggg aattaccgtc ttccgggcgg caatgtacac ggctgacggc 300

ggttatatcg agaatccgtc tgtgaaaaat aaagtcaaag aagctgttga agcggcaaaa 360

gagctcggga tatatgtcat cattgactgg catatttta atgacggcaa tccaaatcaa 420

aataaagaga aggcgaagga attctttaag gaaatgtcaa gcctttacgg aagctcacca 480

aacgttatat atgaaattgc taatgaaccg aacggtgatg taaattggaa gcgcgatatc 540

aaaccgtatg cggaagaagt gatttctgtt atccgtaaaa atgacccgga taacatcatt 600

attaccggaa caggcacttg gagccaggat gtcaacgatg ctgcggatga tcagcttaag 660

gatgcaaacg tcatgtacgc gcttcatttt tatgccggtt cacacggcca gtttttaagg 720

gataaagcgg actatgcgct cagcaaagga gctccgattt ttgtaacgga atgggggacg 780

agtgacgctt ccggaatgg aggggtatac cttgaccagt cgagggaatg gctgaattat 840

ctcgacagca agaaaatcag ctgggtaaac tggaaccttt ctgataagca ggaatcatcc 900

tcagctttaa agccgggggc atctaaaaca ggcggtggc cgttatcaga tttatccgt 960

tcaggacat ttgtaagaga aaacattcgc ggctcccaa attcgagtga agacagatct 1020

gagacaccaa agcaagagaa acccgcacag gaaaacagca tctctgtgca atacagaaca 1080

ggggatggaa gtgtgaacag caaccaaattc cgtctcaga tcaatgtgaa aaacaacagc 1140

aagaccaccg ttaacttaaa aaatgtaact gtccgtact ggtataacac gaaaaacaaa 1200

ggccaaaact tcgactgtga ttacgcgaag atcggtatgca gcaatgtgac gcacaagttt 1260

gtgacattac ataaacctgt aaaagggtgca gatgcctatc tgggaacttgg gtttagaaac 1320

gggacgctgt caccgggagc aagcaccgga gaaattcaaa ttcgtcttca caatgaggac 1380

tggagcaatt attcacaagc cggggattat tcttttttcc agtcgaatac gtttaaagat 1440

acaaaaaaaa tcacattata taataacgga aaactgattt gggaacaga acccaaatag 1500

<210> 90

<211> 499

<212> PRT

<213> Unknown

<220>

<223> Obtained from environmental sample

<220>

<221> SIGNAL

<222> (1)...(29)

<220>

<221> DOMAIN

<222> (47)...(301)

<223> Cellulase (glycosyl hydrolase family 5)

<220>

<221> DOMAIN

<222> (356)...(437)

<223> Cellulose binding domain

<220>

<221> SITE

<222> (164)...(173)

<223> Glycosyl hydrolases family 5 signature. Prosite id = PS00659

<220>

<221> SITE

<222> (296)...(299)

<223> N-glycosylation site. Prosite id = PS00001

<220>

<221> SITE

<222> (339)...(342)

<223> N-glycosylation site. Prosite id = PS00001

<220>

<221> SITE

<222> (383)...(386)

<223> N-glycosylation site. Prosite id = PS00001

<220>

<221> SITE

<222> (393)...(396)

<223> N-glycosylation site. Prosite id = PS00001

<220>

<221> SITE

<222> (421)...(424)

<223> N-glycosylation site. Prosite id = PS00001

<220>

<221> SITE

<222> (446)...(449)

<223> N-glycosylation site. Prosite id = PS00001

<220>

<221> SITE

<222> (470)...(473)

<223> N-glycosylation site. Prosite id = PS00001

<400> 90

Met Lys Arg Ser Val Ser Ile Phe Ile Ala Cys Leu Leu Met Thr Val
1 5 10 15

Leu Thr Ile Ser Gly Val Ala Ala Pro Glu Ala Ser Ala Ala Gly Ala
20 25 30

Lys Thr Pro Val Ala Leu Asn Gly Gln Leu Ser Ile Lys Gly Thr Gln
35 40 45

Leu Val Asn Gln Asn Gly Lys Pro Val Gln Leu Lys Gly Ile Ser Ser
50 55 60

His Gly Leu Gln Trp Phe Gly Asp Tyr Val Asn Lys Asp Thr Leu Lys
65 70 75 80

Trp Leu Arg Asp Asp Trp Gly Ile Thr Val Phe Arg Ala Ala Met Tyr
85 90 95

Thr Ala Asp Gly Gly Tyr Ile Glu Asn Pro Ser Val Lys Asn Lys Val
100 105 110

Lys Glu Ala Val Glu Ala Ala Lys Glu Leu Gly Ile Tyr Val Ile Ile
115 120 125

Asp Trp His Ile Leu Asn Asp Gly Asn Pro Asn Gln Asn Lys Glu Lys
130 135 140

Ala Lys Glu Phe Phe Lys Glu Met Ser Ser Leu Tyr Gly Ser Ser Pro
145 150 155 160

Asn Val Ile Tyr Glu Ile Ala Asn Glu Pro Asn Gly Asp Val Asn Trp
165 170 175

Lys Arg Asp Ile Lys Pro Tyr Ala Glu Glu Val Ile Ser Val Ile Arg
180 185 190

Lys Asn Asp Pro Asp Asn Ile Ile Ile Thr Gly Thr Gly Thr Trp Ser
195 200 205

Gln Asp Val Asn Asp Ala Ala Asp Asp Gln Leu Lys Asp Ala Asn Val
210 215 220

Met Tyr Ala Leu His Phe Tyr Ala Gly Thr His Gly Gln Phe Leu Arg
225 230 235 240

Asp Lys Ala Asp Tyr Ala Leu Ser Lys Gly Ala Pro Ile Phe Val Thr
245 250 255

Glu Trp Gly Thr Ser Asp Ala Ser Gly Asn Gly Gly Val Tyr Leu Asp
260 265 270

Gln Ser Arg Glu Trp Leu Asn Tyr Leu Asp Ser Lys Lys Ile Ser Trp
275 280 285

Val Asn Trp Asn Leu Ser Asp Lys Gln Glu Ser Ser Ser Ala Leu Lys
290 295 300

Pro Gly Ala Ser Lys Thr Gly Gly Trp Pro Leu Ser Asp Leu Ser Ala
305 310 315 320

Ser Gly Thr Phe Val Arg Glu Asn Ile Arg Gly Ser Gln Asn Ser Ser
325 330 335

Glu Asp Arg Ser Glu Thr Pro Lys Gln Glu Lys Pro Ala Gln Glu Asn
340 345 350

Ser Ile Ser Val Gln Tyr Arg Thr Gly Asp Gly Ser Val Asn Ser Asn
355 360 365

Gln Ile Arg Pro Gln Ile Asn Val Lys Asn Asn Ser Lys Thr Thr Val
370 375 380

Asn Leu Lys Asn Val Thr Val Arg Tyr Trp Tyr Asn Thr Lys Asn Lys
385 390 395 400

Gly Gln Asn Phe Asp Cys Asp Tyr Ala Lys Ile Gly Cys Ser Asn Val
405 410 415

Thr His Lys Phe Val Thr Leu His Lys Pro Val Lys Gly Ala Asp Ala
420 425 430

Tyr Leu Glu Leu Gly Phe Arg Asn Gly Thr Leu Ser Pro Gly Ala Ser
435 440 445

Thr Gly Glu Ile Gln Ile Arg Leu His Asn Glu Asp Trp Ser Asn Tyr
450 455 460

Ser Gln Ala Gly Asp Tyr Ser Phe Phe Gln Ser Asn Thr Phe Lys Asp
465 470 475 480

Thr Lys Lys Ile Thr Leu Tyr Asn Asn Gly Lys Leu Ile Trp Gly Thr
485 490 495

Glu Pro Lys

<210> 91

<211> 1725

<212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 91

atgctgaaat taagtataa cctaactttc ttgaaaagca aaccattttt tcttaatgaa 60

aaagaaatga agtgggtgga gaaaacactt caatccatgt ccttacatga aaaagtaggg 120

caattatfff gtccattgg cgggttcagat aataaacaag aattagaagc ctttattaag 180

gaatatcatc ctggcggcat catgtaccgt cctaatacag gagcaaaaat acaggaaaca 240

catcggttgt tacaagagct atccccggtt cctttattaa tttctgctaa cttagaggcc 300

ggtggtaatg ggattgctac ggatgggtact tacttcggaa agcaaatgca ggtggctgca 360

acagataatg aagaaatggc ctataaatta ggattagtgt ctggccgtga aggccgtgtg 420

gccggttgta actgggcttt tgcaccaatt gttgatattg atatgaacta tcgaaacca 480

attacaaacg taagaacgta tgggtctgac ccaattagag ttgcccgaat gtctaaagct 540

tttatgaagg gaattcatga aagcggactc gcagcagctg ttaagcattt cccaggggat 600

ggagtggatg atagagatca gcatctttta tcatctgtaa acaccttata taccgaagaa 660

tgggatcaaa cctttgggat ggtttatcaa gaaatgatag acagtggggc aaaatcgatt 720

atggcgggcc atatcatgct ccctgaatat tcaagagaac tattgccggg tattgaagac 780

gaacaaatca tgccccccac actagcacca gagttactta atggtttatt aagggaagaa 840

ttaggtttta atggtttaat cgtgactgat gcatcccta tgtaggggtt cactacttcg 900

gaaagaagag aaattgctgt tcctaaggcg attgcttcgg gctgtgatat gtttctcttc 960

aaccgtaaca taaaagaaga ttatgagttc atgctgaatg gaattgaaac tggaattcta 1020

accttgaaa gagtagatga agctgttact agagtacttg ctcttaaagc atctctaggt 1080

ctgaatgtac aaaaggaatt gggaatatta gtacctgaag aagcggaatt gtcggtatta 1140

caatctgaag aacatttgga ttgggcaaga gaatgtgcag accaatcggt tacattagta 1200

aaggatacac aaaaactgct gcctattagt gctgatcagt ataaacgggt tcgactttat 1260

gtattgggtg atcaagaagg agggctaaag gaaggcggct ccgtcactca accgtttatc 1320

gattctctta aaaatgctgg ctttgaagia gatttatata atgacaagca agttaatttc 1380

caagaactgt ttatgagtgt aaacgagttt aaaaagaact atgatctgat catttatgtc 1440

gccaaccttg aaaccgctag taaccaaagc acagtcagaa ttaattggca gcagccgcta 1500

aatgccaacg ctccatgggt tggttaaagat ataccgacat tatttatttc ggttgctaac 1560

ccataccatc tacaggacgt accaatgggt aagacctata taaatgctta ttcactaat 1620

gaatatgtgg tagaagcaat ttagataaaa atcttaggaa aatcagagtt taaaggaag 1680

aatcccgctg atccgttttg tgggaaatgg gataccagac tttaa 1725

<210> 92

<211> 574

<212> PRT
<213> Unknown

<220>
<223> Obtained from environmental sample

<220>
<221> DOMAIN
<222> (87)...(320)
<223> Glycosyl hydrolase family 3 N terminal domain

<220>
<221> SITE
<222> (7)...(10)
<223> N-glycosylation site. Prosite id = PS00001

<220>
<221> SITE
<222> (495)...(498)
<223> N-glycosylation site. Prosite id = PS00001

<400> 92
Met Leu Lys Leu Ser Asp Asn Leu Thr Phe Leu Lys Ser Lys Pro Phe
1 5 10 15

Phe Leu Asn Glu Lys Glu Met Lys Trp Val Glu Lys Thr Leu Gln Ser
20 25 30

Met Ser Leu His Glu Lys Val Gly Gln Leu Phe Cys Pro Ile Gly Gly
35 40 45

Ser Asp Asn Lys Gln Glu Leu Glu Ala Phe Ile Lys Glu Tyr His Pro
50 55 60

Gly Gly Ile Met Tyr Arg Pro Asn Thr Gly Ala Lys Ile Gln Glu Thr
65 70 75 80

His Arg Leu Leu Gln Glu Leu Ser Pro Val Pro Leu Leu Ile Ser Ala
85 90 95

Asn Leu Glu Ala Gly Gly Asn Gly Ile Ala Thr Asp Gly Thr Tyr Phe
100 105 110

Gly Lys Gln Met Gln Val Ala Ala Thr Asp Asn Glu Glu Met Ala Tyr
115 120 125

Lys Leu Gly Leu Val Ala Gly Arg Glu Gly Arg Val Ala Gly Cys Asn
130 135 140

Trp Ala Phe Ala Pro Ile Val Asp Ile Asp Met Asn Tyr Arg Asn Pro
145 150 155 160

Ile Thr Asn Val Arg Thr Tyr Gly Ser Asp Pro Ile Arg Val Ala Gln
165 170 175

Met Ser Lys Ala Phe Met Lys Gly Ile His Glu Ser Gly Leu Ala Ala
180 185 190

Ala Val Lys His Phe Pro Gly Asp Gly Val Asp Asp Arg Asp Gln His
195 200 205

Leu Leu Ser Ser Val Asn Thr Leu Ser Thr Glu Glu Trp Asp Gln Thr
210 215 220

Phe Gly Met Val Tyr Gln Glu Met Ile Asp Ser Gly Ala Lys Ser Ile
225 230 235 240

Met Ala Gly His Ile Met Leu Pro Glu Tyr Ser Arg Glu Leu Leu Pro
245 250 255

Gly Ile Glu Asp Glu Gln Ile Met Pro Ala Thr Leu Ala Pro Glu Leu
260 265 270

Leu Asn Gly Leu Leu Arg Glu Lys Leu Gly Phe Asn Gly Leu Ile Val
275 280 285

Thr Asp Ala Ser Pro Met Leu Gly Phe Thr Thr Ser Glu Arg Arg Glu

290

295

300

Ile Ala Val Pro Lys Ala Ile Ala Ser Gly Cys Asp Met Phe Leu Phe
305 310 315 320

Asn Arg Asn Ile Lys Glu Asp Tyr Glu Phe Met Leu Asn Gly Ile Glu
325 330 335

Thr Gly Ile Leu Thr Leu Glu Arg Val Asp Glu Ala Val Thr Arg Val
340 345 350

Leu Ala Leu Lys Ala Ser Leu Gly Leu Asn Val Gln Lys Glu Leu Gly
355 360 365

Ile Leu Val Pro Glu Glu Ala Glu Leu Ser Val Leu Gln Ser Glu Glu
370 375 380

His Leu Asp Trp Ala Arg Glu Cys Ala Asp Gln Ser Val Thr Leu Val
385 390 395 400

Lys Asp Thr Gln Lys Leu Leu Pro Ile Ser Ala Asp Gln Tyr Lys Arg
405 410 415

Val Arg Leu Tyr Val Leu Gly Asp Gln Glu Gly Gly Leu Lys Glu Gly
420 425 430

Gly Ser Val Thr Gln Pro Phe Ile Asp Ser Leu Lys Asn Ala Gly Phe
435 440 445

Glu Val Asp Leu Tyr Asn Asp Lys Gln Val Asn Phe Gln Glu Leu Phe
450 455 460

Met Ser Val Asn Glu Phe Lys Lys Asn Tyr Asp Leu Ile Ile Tyr Val
465 470 475 480

Ala Asn Leu Glu Thr Ala Ser Asn Gln Thr Thr Val Arg Ile Asn Trp
485 490 495

Gln Gln Pro Leu Asn Ala Asn Ala Pro Trp Phe Val Lys Asp Ile Pro
500 505 510

Thr Leu Phe Ile Ser Val Ala Asn Pro Tyr His Leu Gln Asp Val Pro
515 520 525

Met Val Lys Thr Tyr Ile Asn Ala Tyr Ser Ser Asn Glu Tyr Val Val
530 535 540

Glu Ala Ile Val Asp Lys Ile Leu Gly Lys Ser Glu Phe Lys Gly Lys
545 550 555 560

Asn Pro Val Asp Pro Phe Cys Gly Lys Trp Asp Thr Arg Leu
565 570

<210> 93

<211> 546

<212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 93

atgagaataa aaaattttaa aacgaaccgt atcacaacc cgctgggatt tgatatagga 60

aaaccacgta tatcttttgt cacttatgac actacggcta aaaagcaaac agcagcgcaa 120

atacaggttg cgctagatca agagtttacg aacctaacat ttgacagtgg gaaaagcacg 180

gagatagata gtctagcata cgaactgcc a tttcaattag agtcttacac tcgctactac 240

tggcgtgtga ccgtttgggc ggataatggg gatgtggcca caagtgaaat tgcttggttt 300

gaaacagcca aactaggcga ttcttgggag gccaaagtga ttacccccga ttttgataag 360

gaaatccatc ccgtactatc aagggaattt gatttgtcaa aagaagtcgt tctgcccgt 420

gcctatgttt gcggtttggg attatatgaa atggagatta atggtctaaa ggctggggat 480

gaatatctga cccctaattt caacgcctat gataaatggc tgcagtacca gacctatgat 540

attaca 546

<210> 94

<211> 182

<212> PRT

<213> Unknown

<220>

<223> Obtained from environmental sample

<220>

<221> SITE

<222> (51)...(54)

<223> N-glycosylation site. Prosite id = PS00001

<400> 94

Met	Arg	Ile	Lys	Asn	Leu	Lys	Thr	Asn	Arg	Ile	Thr	Asn	Pro	Leu	Gly
1			5					10					15		

Phe	Asp	Ile	Gly	Lys	Pro	Arg	Ile	Ser	Phe	Val	Thr	Tyr	Asp	Thr	Thr
	20						25						30		

Ala	Lys	Lys	Gln	Thr	Ala	Ala	Gln	Ile	Gln	Val	Ala	Leu	Asp	Gln	Glu
	35						40					45			

Phe	Thr	Asn	Leu	Thr	Phe	Asp	Ser	Gly	Lys	Ser	Thr	Glu	Ile	Asp	Ser
	50					55					60				

Leu	Ala	Tyr	Glu	Leu	Pro	Phe	Gln	Leu	Glu	Ser	Tyr	Thr	Arg	Tyr	Tyr
65					70				75					80	

Trp	Arg	Val	Thr	Val	Trp	Ala	Asp	Asn	Gly	Asp	Val	Ala	Thr	Ser	Glu
				85					90					95	

Ile Ala Trp Phe Glu Thr Ala Lys Leu Gly Asp Ser Trp Glu Ala Lys
100 105 110

Trp Ile Thr Pro Asp Phe Asp Lys Glu Ile His Pro Val Leu Ser Arg
115 120 125

Glu Phe Asp Leu Ser Lys Glu Val Val Ser Ala Arg Ala Tyr Val Cys
130 135 140

Gly Leu Gly Leu Tyr Glu Met Glu Ile Asn Gly Leu Lys Ala Gly Asp
145 150 155 160

Glu Tyr Leu Thr Pro Asn Phe Asn Ala Tyr Asp Lys Trp Leu Gln Tyr
165 170 175

Gln Thr Tyr Asp Ile Thr
180

<210> 95

<211> 2298

<212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 95

atgatcaatc aagatataaa acaattaatc tcacaaatga ccttgaaga aaaagctggt 60

ctttgttctg gattagattt ttggaattta aaaggtatcg aaagactggg aataccctcg 120

ataatggtaa ccgatgggcc gcatggactc cgtaaacaaa aaatgggagc agatcattta 180

gggctgtttg acagtattcc tgcgacatgt ttcccatctg cagccggttt agctagtact 240

tggaataaag agttaatata tgaagttggg gttgcattag gaaaggaatg ccaggcagag 300

gatgtggcaa tacttcttgg ccctggagca aacattaagc gtcaccctt ttgtggcaga 360

aactttgaat atttttcgga agatccattc ctttcatcag aaatggctgc gtcccatatc 420

aagggtgttc aaagtgaggg ggttgggaca tcacttaagc acttcgctgc aaataatcaa 480

gaacaccgaa gaatgtcgac agatgctatt gtggatgaaa ggacgttgcg agaaatatat 540

ttggccagct ttgaaaacgc tgtaagaaa ggcagccat ggactgtgat gtgcgcctac 600

aacaaggta atggagactt tgcacagaa aataaaacat tgttaactga catcctgcga 660

gatgagtggg gctttgaagg aattgttgtt tctgactggg gggcggttaa tgaacctgtt 720

gacggattaa atgccgggtt agacctggaa atgccttcaa gtagtgggat tggtgaaaag 780

aaaatcatca atgctgtaag aaatggtcag cttttagaag ataaactaga tcaggcagtt 840

gaaagaattc tacgtattat cttaatggca gtagaaaaca agaaagaaac cgctgactat 900

gataaagaac aacatcataa gcttcaaga aaagcagcaa gtgaaagtat ggttttatta 960

aagaatgaag ataatatcct gccgttaaag aaagaaggaa ccatttcgat tattggttca 1020

tttgccaaaa aaccaaggta tcaaggcggg ggaagctcac acattaaccc gacaaaagctt 1080

gaaaatatct atgaagaaat agagaaaaca gcgggccaaa atgtgaacgt tttatacgcg 1140

gaaggatatc atcttgaaaa ggatttaatc gatgatcaat taattgaaga ggcaaaaaaa 1200

acggcagcaa aatccgatgt aaccgtattg ttgtaggtc ttcctgaccg atatgaatct 1260

gaaggatatg atagagagca cctgaatata ccggagaatc accgtctttt agtcgaagcg 1320

gttgcggaag tacaaaagaa tatagttgtt gtactaagta atggggcacc gcttgttatg 1380

ccatggcttg ataaggtgaa ggggctgctg gaaagttacc tgggaggtca ggcaactagga 1440

ggtgcgattg cagacatcct attcggagaa gttaatccaa gtggaaagct tgccgaaact 1500

tttcccgtaa aattaggta caatccttct tatctcaact ttccaggaga gagggataaa 1560

gttgagtata aagaaggcat ctttgttggc tatcggtatt acgatacaaa acagattgag 1620

ccgctgtttc catttgata tggtttaagc tatacaaact ttgaatataa aaaccttgta 1680

attgataaaa aagaaataaa agatacagaa attgtcacag ttaccgtgaa tgtgaaaaat 1740

acaggaaaag tgcctgggaa agaaatcatc cagtttatatg taaaagatat aaaaagcagt 1800

glagttcgtc ctgaaaaaga gttaaaaggc ttggaaagg tttccttaca gcctggggaa 1860

gacaaaacta tttcctttaa attgataaa cgcgcatttg catattaca cacggaattg 1920

aaggattggc atgtagaatc aggagaattt gaaattttgg tggggaaatc gtccagagaa 1980

attgaactaa cagaaaaaat tatggttcac tctacttccc cagttttctt ggaggttcac 2040

cgaaattcca cggtcggaga tcttttaact gatccaattc tagtgaaaa agctaattgct 2100

ctaattagag agctaacaaa aggaagtcca ttatttgatg ctgggtcaga tcacggagag 2160

gggtgcagaaa tgatggaagc gatgttaaaa tacatgcctt tgcgtgctct tatgaatttt 2220

agtgggtggag acattaccga agagaaacta actgaattta ttaaggaact taattcaact 2280

aattttgtaa gcctttaa 2298

<210> 96
 <211> 765
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> DOMAIN
 <222> (30)...(252)
 <223> Glycosyl hydrolase family 3 N terminal domain

<220>

<221> DOMAIN
 <222> (317)...(531)
 <223> Glycosyl hydrolase family 3 C terminal domain

<220>
 <221> SITE
 <222> (214)...(217)
 <223> N-glycosylation site. Prosite id = PS00001

<220>
 <221> SITE
 <222> (221)...(238)
 <223> Glycosyl hydrolases family 3 active site. Prosite id = PS00775

<220>
 <221> SITE
 <222> (692)...(695)
 <223> N-glycosylation site. Prosite id = PS00001

<220>
 <221> SITE
 <222> (750)...(753)
 <223> N-glycosylation site. Prosite id = PS00001

<220>
 <221> SITE
 <222> (769)...(772)
 <223> N-glycosylation site. Prosite id = PS00001

<400> 96
 Met Ile Asn Gln Asp Ile Lys Gln Leu Ile Ser Gln Met Thr Leu Glu
 1 5 10 15

Glu Lys Ala Gly Leu Cys Ser Gly Leu Asp Phe Trp Asn Leu Lys Gly
 20 25 30

Ile Glu Arg Leu Gly Ile Pro Ser Ile Met Val Thr Asp Gly Pro His
 35 40 45

Gly Leu Arg Lys Gln Lys Met Gly Ala Asp His Leu Gly Leu Phe Asp
 50 55 60

Ser Ile Pro Ala Thr Cys Phe Pro Ser Ala Ala Gly Leu Ala Ser Thr

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

260 265 270

Glu Asp Lys Leu Asp Gln Ala Val Glu Arg Ile Leu Arg Ile Ile Leu
275 280 285

Met Ala Val Glu Asn Lys Lys Glu Thr Ala Asp Tyr Asp Lys Glu Gln
290 295 300

His His Lys Leu Ala Arg Lys Ala Ala Ser Glu Ser Met Val Leu Leu
305 310 315 320

Lys Asn Glu Asp Asn Ile Leu Pro Leu Lys Lys Glu Gly Thr Ile Ser
325 330 335

Ile Ile Gly Ser Phe Ala Lys Lys Pro Arg Tyr Gln Gly Gly Gly Ser
340 345 350

Ser His Ile Asn Pro Thr Lys Leu Glu Asn Ile Tyr Glu Glu Ile Glu
355 360 365

Lys Thr Ala Gly Gln Asn Val Asn Val Leu Tyr Ala Glu Gly Tyr His
370 375 380

Leu Glu Lys Asp Leu Ile Asp Asp Gln Leu Ile Glu Glu Ala Lys Lys
385 390 395 400

Thr Ala Ala Lys Ser Asp Val Thr Val Leu Phe Val Gly Leu Pro Asp
405 410 415

Arg Tyr Glu Ser Glu Gly Tyr Asp Arg Glu His Leu Asn Ile Pro Glu
420 425 430

Asn His Arg Leu Leu Val Glu Ala Val Ala Glu Val Gln Lys Asn Ile
435 440 445

Val Val Val Leu Ser Asn Gly Ala Pro Leu Val Met Pro Trp Leu Asp

450

455

460

Lys Val Lys Gly Leu Leu Glu Ser Tyr Leu Gly Gly Gln Ala Leu Gly
465 470 475 480

Gly Ala Ile Ala Asp Ile Leu Phe Gly Glu Val Asn Pro Ser Gly Lys
485 490 495

Leu Ala Glu Thr Phe Pro Val Lys Leu Gly Asp Asn Pro Ser Tyr Leu
500 505 510

Asn Phe Pro Gly Glu Arg Asp Lys Val Glu Tyr Lys Glu Gly Ile Phe
515 520 525

Val Gly Tyr Arg Tyr Tyr Asp Thr Lys Gln Ile Glu Pro Leu Phe Pro
530 535 540

Phe Gly Tyr Gly Leu Ser Tyr Thr Asn Phe Glu Tyr Lys Asn Leu Val
545 550 555 560

Ile Asp Lys Lys Glu Ile Lys Asp Thr Glu Ile Val Thr Val Thr Val
565 570 575

Asn Val Lys Asn Thr Gly Lys Val Pro Gly Lys Glu Ile Ile Gln Leu
580 585 590

Tyr Val Lys Asp Ile Lys Ser Ser Val Val Arg Pro Glu Lys Glu Leu
595 600 605

Lys Gly Phe Gly Lys Val Ser Leu Gln Pro Gly Glu Asp Lys Thr Ile
610 615 620

Ser Phe Lys Leu Asp Lys Arg Ala Phe Ala Tyr Tyr Asn Thr Glu Leu
625 630 635 640

Lys Asp Trp Tyr Val Glu Ser Gly Glu Phe Glu Ile Leu Val Gly Lys
645 650 655

Ser Ser Arg Glu Ile Glu Leu Thr Glu Lys Ile Met Val His Ser Thr
660 665 670

Ser Pro Val Phe Leu Glu Val His Arg Asn Ser Thr Val Gly Asp Leu
675 680 685

Leu Thr Asp Pro Ile Leu Gly Glu Lys Ala Asn Ala Leu Ile Arg Glu
690 695 700

Leu Thr Lys Gly Ser Pro Leu Phe Asp Ala Gly Ser Asp His Gly Glu
705 710 715 720

Gly Ala Glu Met Met Glu Ala Met Leu Lys Tyr Met Pro Leu Arg Ala
725 730 735

Leu Met Asn Phe Ser Gly Gly Asp Ile Thr Glu Glu Lys Leu Thr Glu
740 745 750

Phe Ile Lys Glu Leu Asn Ser Thr Asn Phe Val Ser Leu
755 760 765

<210> 97
<211> 615
<212> DNA
<213> Unknown

<220>
<223> Obtained from environmental sample

<400> 97
atgttataacc caattataac tgaaactcgc agtatcatcg atttaaattgg tatctggaaa 60

tttaaattag ataatgggtga aggactgcag gaaaaatggt atgaaaacgg attaacagac 120

acgatcagta tggtgtgacc atcttccttt aatgatattg gagtaaattgc cagtatacgc 180

gatcatgttg gctgggtatg gtatgagcgg gaattttctg tccccgcat ctttcaatct 240

gagcgtgtgg ttttgcgatt cggttccgca acacatctag ctaaggtttt cgtaaagtgt 300

gaacttggtt ttgaacataa gggcgggttt ttaccgtttg aagcagaaat aaataagttt 360

ttacaaaaag ggaaaaatcg aataacggtt gctgtcaaca atattcttga ttactcaact 420

ttaccgtttg gcacagtaat agaaaaggat attcctggag ttggcaaagt aatacgcaat 480

cagccaaatt ttgacttctt caactacgct ggcttgcacc gtccagtga aatataact 540

acaccgacta cttatgtgaa ggatgtaacc attgtaacgg aaatagatgg acaggttcac 600

tattcaattg attaa 615

<210> 98
 <211> 204
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> DOMAIN
 <222> (9)...(182)
 <223> Glycosyl hydrolases family 2, sugar binding domain

<220>
 <221> SITE
 <222> (56)...(59)
 <223> N-glycosylation site. Prosite id = PS00001

<400> 98
 Met Leu Tyr Pro Ile Ile Thr Glu Thr Arg Ser Ile Ile Asp Leu Asn
 1 5 10 15

Gly Ile Trp Lys Phe Lys Leu Asp Asn Gly Glu Gly Leu Gln Glu Lys
 20 25 30

Trp Tyr Glu Asn Gly Leu Thr Asp Thr Ile Ser Met Ala Val Pro Ser

35

40

45

Ser Phe Asn Asp Ile Gly Val Asn Ala Ser Ile Arg Asp His Val Gly
50 55 60

Trp Val Trp Tyr Glu Arg Glu Phe Ser Val Pro Ala Ile Leu Gln Ser
65 70 75 80

Glu Arg Val Val Leu Arg Phe Gly Ser Ala Thr His Leu Ala Lys Val
85 90 95

Phe Val Asn Gly Glu Leu Val Val Glu His Lys Gly Gly Phe Leu Pro
100 105 110

Phe Glu Ala Glu Ile Asn Lys Phe Leu Gln Lys Gly Lys Asn Arg Ile
115 120 125

Thr Val Ala Val Asn Asn Ile Leu Asp Tyr Ser Thr Leu Pro Val Gly
130 135 140

Thr Val Ile Glu Lys Asp Ile Pro Gly Val Gly Lys Val Ile Arg Asn
145 150 155 160

Gln Pro Asn Phe Asp Phe Phe Asn Tyr Ala Gly Leu His Arg Pro Val
165 170 175

Lys Ile Tyr Thr Thr Pro Thr Thr Tyr Val Lys Asp Val Thr Ile Val
180 185 190

Thr Glu Ile Asp Gly Gln Val His Tyr Ser Ile Asp
195 200

<210> 99

<211> 1404

<212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 99

atgaatcatt ccctttcatt tccgcatcc ttgtatggg gcgcggcaac cgcaagctac	60
caactggaag gatcaacca aggcgtggac ggctgcgccc agtccgtctg ggatatgcac	120
tgccgaagat ccggcgcgat caaggacggc tcgaacggat tcgtcgctg cgatcactac	180
catcgctatc gcgaggatgt ggcgctcatg aacgagcttg gcttgaatgc ctatcgattc	240
tcaatcatgt ggccccgcgt catgcccga ggcaccggcg cggatgaacga gaaggcatg	300
gattttctacg atcggttggg tgatgaactg ctgcccgcg gcatcacacc ttgggttact	360
ttgttccact gggactttcc cctagccttg ttccaacgcg gtggtggct gaatgcggat	420
tccccgaat ggtttgagga ttacactcgg gaagtggta aacgcttgtc ggatcgtgtg	480
catcactggc taacgtcaa cgaaccggcg tgcttcattg agtttggcca ccgtaccggc	540
atgatgcac ccggcttgca actggcggac aaggaaacct gccgggtctg gcaccatgcc	600
atgctggccc acggtcgcgc cgttcgcgt attcgccagg aatccgtgca tccatcacc	660
caggctggct acgcgccgt cttccgact accatcccgg acactgaaga tctgccgac	720
atcgaagcgg ccggacctc gatgtttgct catcaggccg gcaacctgtt cgatacgcg	780
tggaacctcg acccctgctt tcggggcgcg tatccggaga tcatgatgca gtattggggc	840
gatgccgcgc cgcgcatcca ggacggcgac atggagtga tccgtcagga actcgatatt	900
ctcggcctga atatitacca gtccgagcgc attcgggccc gtgcggatgg cgcacccgag	960
gtggtgccat accctgcgga ttatccgcgc aaccagctcg gttggcccat cacgccggag	1020
gccctgcgct gggcgacctt ctttctcttt gaggagtacg ggaaacctt gatcatcaca	1080

gaaaacggaa tcaccctcga cgacaagccc aatgcagacg gcgaggtgaa tgatgtccag 1140

cggatcgctt ttctgaatga ctatcttagc ggtctccagc gcagcgtgga cgacggcatc 1200

cctgtactgg gctatttcca ctggtcgctg tgcgacaact ttgagtgggc agaaggctat 1260

gtccctcgct tcggcctgat ccatgtggac tatgccagtc aacgcagaac catcaaggcc 1320

tcaggacggt ttaccgcga catcattcgg ggccagacag ccacgccctg catcgcccaa 1380

tccagtcagc cgaaacaac ctaa 1404

<210> 100
 <211> 467
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> DOMAIN
 <222> (3)...(454)
 <223> Glycosyl hydrolase family 1

<220>
 <221> SITE
 <222> (2)...(5)
 <223> N-glycosylation site. Prosite id = PS00001

<220>
 <221> SITE
 <222> (11)...(25)
 <223> Glycosyl hydrolases family 1 N-terminal signature. Prosite id = PS00653

<400> 100
 Met Asn His Ser Leu Ser Phe Pro Pro Ser Phe Val Trp Gly Ala Ala
 1 5 10 15

Thr Ala Ser Tyr Gln Leu Glu Gly Ser Thr Gln Gly Val Asp Gly Cys
 20 25 30

Ala Glu Ser Val Trp Asp Met His Cys Arg Arg Ser Gly Ala Ile Lys
35 40 45

Asp Gly Ser Asn Gly Phe Val Ala Cys Asp His Tyr His Arg Tyr Arg
50 55 60

Glu Asp Val Ala Leu Met Asn Glu Leu Gly Leu Asn Ala Tyr Arg Phe
65 70 75 80

Ser Ile Met Trp Pro Arg Val Met Pro Glu Gly Thr Gly Ala Val Asn
85 90 95

Glu Lys Gly Met Asp Phe Tyr Asp Arg Leu Val Asp Glu Leu Leu Ala
100 105 110

Ala Gly Ile Thr Pro Trp Val Thr Leu Phe His Trp Asp Phe Pro Leu
115 120 125

Ala Leu Phe Gln Arg Gly Gly Trp Leu Asn Ala Asp Ser Pro Gln Trp
130 135 140

Phe Glu Asp Tyr Thr Arg Glu Val Val Lys Arg Leu Ser Asp Arg Val
145 150 155 160

His His Trp Leu Thr Leu Asn Glu Pro Ala Cys Phe Ile Glu Phe Gly
165 170 175

His Arg Thr Gly Met His Ala Pro Gly Leu Gln Leu Ala Asp Lys Glu
180 185 190

Ala Cys Arg Val Trp His His Ala Met Leu Ala His Gly Arg Ala Val
195 200 205

Arg Ala Ile Arg Gln Glu Ser Val His Pro Ser Pro Gln Val Gly Tyr

210

215

220

Ala Pro Val Phe Arg Thr Thr Ile Pro Asp Thr Glu Asp Pro Ala Asp
225 230 235 240

Ile Glu Ala Ala Arg Thr Ser Met Phe Ala His Gln Ala Gly Asn Leu
245 250 255

Phe Asp Thr Arg Trp Asn Leu Asp Pro Cys Phe Arg Gly Ala Tyr Pro
260 265 270

Glu Ile Met Met Gln Tyr Trp Gly Asp Ala Ala Pro Arg Ile Gln Asp
275 280 285

Gly Asp Met Glu Leu Ile Arg Gln Glu Leu Asp Phe Leu Gly Leu Asn
290 295 300

Ile Tyr Gln Ser Glu Arg Ile Arg Ala Gly Ala Asp Gly Ala Pro Glu
305 310 315 320

Val Val Pro Tyr Pro Ala Asp Tyr Pro Arg Asn Gln Leu Gly Trp Pro
325 330 335

Ile Thr Pro Glu Ala Leu Arg Trp Ala Thr Leu Phe Leu Phe Glu Glu
340 345 350

Tyr Gly Lys Pro Leu Ile Ile Thr Glu Asn Gly Ile Thr Leu Asp Asp
355 360 365

Lys Pro Asn Ala Asp Gly Glu Val Asn Asp Val Gln Arg Ile Ala Phe
370 375 380

Leu Asn Asp Tyr Leu Ser Gly Leu Gln Arg Ser Val Asp Asp Gly Ile
385 390 395 400

Pro Val Leu Gly Tyr Phe His Trp Ser Leu Cys Asp Asn Phe Glu Trp
405 410 415

Ala Glu Gly Tyr Val Pro Arg Phe Gly Leu Ile His Val Asp Tyr Ala
420 425 430

Ser Gln Arg Arg Thr Ile Lys Ala Ser Gly Arg Phe Tyr Arg Asp Ile
435 440 445

Ile Arg Gly Gln Thr Ala Thr Pro Cys Ile Ala Gln Ser Ser Gln Pro
450 455 460

Glu Thr Thr
465

<210> 101
<211> 1101
<212> DNA
<213> Unknown

<220>
<223> Obtained from environmental sample

<400> 101
atgagaaatc atctgaatgt acccttttac tttatcttct ttttttaaat agcgtcaata 60

tttacagtct gttcatcatc aactgcttct gataacaatg agcatccacc gccagtggaa 120

gtcgcggatc aggacgcttt tcgtgatgct tttgaagtga atgaattact tggacgcggt 180

attaatctgg gtaatgcctt tgaagcgccc aatgaaggcg aatggggaat ggtaatccag 240

gaagagtttc ttgatctgat acttcagca ggttttgagt ctgtacgaat tccgattcgc 300

tggaatgccc atgccagtga aagtcacctt ttcaccattc aacgatcgtt tttgatcgg 360

gttgatgaag tcatccaatg gtcgtggat cgtggccttt ctgtaatgat caatattcat 420

cactacaatg aactgatgca aaaccgcag cagcaccggc agcggttttt gcgactctgg 480

aaccagattg ctacacacta taaagattat ccggataatc tggtttttga aatccttaat 540

gaacctcatg ataatctgac tccttctatc tggaatagtt atttgaggga tgctattggc 600

atgattcgcc agacaaaccc acgcagggtt atcgctatcg gaacagcaaa ctgggggtggt 660

ttcggagcat tatcacaact tgaaatcccc tcaaacgac gccagatcat tgcaactgtt 720

cattattatg aacccttcag gttcacccat cagggggctg aatgggcagg accggaaca 780

aacgattggc tggggacacg atgggatgga tcggatgagg aaaaatttga tattgaaagt 840

ggttttgatg ccgtacagtc ctgggcagtg acaaataacc ggcctgttca tctcggagaa 900

ttcggtgctt acagtactgc cgataatgaa tcacgcgaac gctggacaac ctttgttcgg 960

gaatccgctg agcaacgcaa tticagctgg gcatactggg aatttgcagc cggttttggg 1020

atctatgacc gtaatcagtg gcaatggagg gattatctgt tgagggttt gataccggat 1080

agcccgggtcc tgttgagta a 1101

<210> 102
<211> 366
<212> PRT
<213> Unknown

<220>
<223> Obtained from environmental sample

<220>
<221> SIGNAL
<222> (1)...(29)

<220>
<221> DOMAIN
<222> (64)...(342)
<223> Cellulase (glycosyl hydrolase family 5)

<220>
<221> SITE
<222> (176)...(185)

<223> Glycosyl hydrolases family 5 signature. Prosite id = PS00659

<220>

<221> SITE

<222> (313)...(316)

<223> N-glycosylation site. Prosite id = PS00001

<220>

<221> SITE

<222> (332)...(335)

<223> N-glycosylation site. Prosite id = PS00001

<400> 102

Met	Arg	Asn	His	Leu	Asn	Val	Pro	Phe	Tyr	Phe	Ile	Phe	Phe	Phe	Leu
1				5					10					15	

Ile	Ala	Ser	Ile	Phe	Thr	Val	Cys	Ser	Ser	Ser	Thr	Ala	Ser	Asp	Asn
			20					25						30	

Asn	Glu	His	Pro	Pro	Pro	Val	Glu	Val	Ala	Asp	Gln	Asp	Ala	Phe	Arg
		35					40					45			

Asp	Ala	Phe	Glu	Val	Asn	Glu	Leu	Leu	Gly	Arg	Gly	Ile	Asn	Leu	Gly
	50						55					60			

Asn	Ala	Leu	Glu	Ala	Pro	Asn	Glu	Gly	Glu	Trp	Gly	Met	Val	Ile	Gln
65					70					75					80

Glu	Glu	Phe	Leu	Asp	Leu	Ile	Leu	Ala	Ala	Gly	Phe	Glu	Ser	Val	Arg
				85					90					95	

Ile	Pro	Ile	Arg	Trp	Asn	Ala	His	Ala	Ser	Glu	Ser	His	Pro	Phe	Thr
			100					105					110		

Ile	Gln	Arg	Ser	Phe	Phe	Asp	Arg	Val	Asp	Glu	Val	Ile	Gln	Trp	Ser
		115					120					125			

Leu	Asp	Arg	Gly	Leu	Ser	Val	Met	Ile	Asn	Ile	His	His	Tyr	Asn	Glu

130

135

140

Leu Met Gln Asn Pro Gln Gln His Arg Gln Arg Phe Leu Arg Leu Trp
145 150 155 160

Asn Gln Ile Ala Thr His Tyr Lys Asp Tyr Pro Asp Asn Leu Val Phe
165 170 175

Glu Ile Leu Asn Glu Pro His Asp Asn Leu Thr Pro Ser Ile Trp Asn
180 185 190

Ser Tyr Leu Arg Asp Ala Ile Gly Met Ile Arg Gln Thr Asn Pro Arg
195 200 205

Arg Val Ile Ala Ile Gly Thr Ala Asn Trp Gly Gly Phe Gly Ala Leu
210 215 220

Ser Gln Leu Glu Ile Pro Ser Asn Asp Arg Gln Ile Ile Ala Thr Val
225 230 235 240

His Tyr Tyr Glu Pro Phe Arg Phe Thr His Gln Gly Ala Glu Trp Ala
245 250 255

Gly Pro Glu Thr Asn Asp Trp Leu Gly Thr Arg Trp Asp Gly Ser Asp
260 265 270

Glu Glu Lys Phe Asp Ile Glu Ser Gly Phe Asp Ala Val Gln Ser Trp
275 280 285

Ala Val Thr Asn Asn Arg Pro Val His Leu Gly Glu Phe Gly Ala Tyr
290 295 300

Ser Thr Ala Asp Asn Glu Ser Arg Glu Arg Trp Thr Thr Phe Val Arg
305 310 315 320

Glu Ser Ala Glu Gln Arg Asn Phe Ser Trp Ala Tyr Trp Glu Phe Ala
325 330 335

Ala Gly Phe Gly Ile Tyr Asp Arg Asn Gln Trp Gln Trp Arg Asp Tyr
340 345 350

Leu Leu Arg Ala Leu Ile Pro Asp Ser Pro Val Leu Leu Glu
355 360 365

<210> 103
<211> 1101
<212> DNA
<213> Unknown

<220>
<223> Obtained from environmental sample

<400> 103
atgctgataa ttggaggcct tcttgtttta ctgggatttt cttcttgctg gcggcaggca 60

gaacctgctg ctgactcttt caggggggttt catgactttg acatcaggcg tgggggtgaac 120

atcagccact ggttgtcgca gagtggaagg cgtggtgctg atcgggagcg gttctttacc 180

agggcggatg tggaggccat cgccggcttc ggttatgac acattcgttt gccattgat 240

gaggagcaga tgtgggatga gtcgggcaac aaggaaccac gtgcctttga attgctgcat 300

gaagccattg gctgggcttt ggacaatgag ctgagggtca ttgtcgacct gcacatcatc 360

aggctgcact attttaatgc gcctgagaac ccgctttgga ccgatcgtgc tgaacagtgtg 420

aaatttggtg agatgtggcg acagttgtct gatgagctgc agggctatcc gctcgatagg 480

gtggcctatg aattgatgaa tgaggccgtg gctgatgac cgacgattg gaaccggctt 540

gtggctgaga cgatggaggc gctacggatg ctggaaccgg agcgcaagat tgtcattggc 600

tccaaccgct ggcagtctgt gcatacatct cctgacctgg tgatcccga taatgaccg 660

catatcatat tgagttttca ctctacgaa ccatttctgc tgacgcacca caaggcctcc 720

tggacacaca tccgtgatta caccgtccg gtgaactatc cgggtttgac ttagaccgcg 780

accacactgg aggggttgc tgaagaactg gtgaccgga ttggccatca caatggggtg 840

tataaaaag aaacgatga ggagatgatc atgatccac tgcaatatgc caagaccgg 900

gggtccccc ttattgtgg agagtgggga tgtttccga ccatgcccc ggagatgcgc 960

ctgcaatgt acgccgatgt gcgtgcgatc ctggaaaagc atgagattgc ctgggcaaac 1020

tgggattaca aggttggtt cgggtgtgtt gaccgcaacg gcgaaccca ccatgattta 1080

ttggaagtgc tcttaaaata a 1101

<210> 104
 <211> 366
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> SIGNAL
 <222> (1)...(20)

<220>
 <221> DOMAIN
 <222> (42)...(349)
 <223> Cellulase (glycosyl hydrolase family 5)

<220>
 <221> SITE
 <222> (40)...(43)
 <223> N-glycosylation site. Prosite id = PS00001

<400> 104
 Met Leu Ile Ile Gly Gly Leu Leu Val Leu Leu Gly Phe Ser Ser Cys
 1 5 10 15

Gly Arg Gln Ala Glu Pro Ala Ala Asp Ser Phe Arg Gly Phe His Asp
20 25 30

Phe Asp Ile Arg Arg Gly Val Asn Ile Ser His Trp Leu Ser Gln Ser
35 40 45

Gly Arg Arg Gly Ala Asp Arg Glu Ala Phe Phe Thr Arg Ala Asp Val
50 55 60

Glu Ala Ile Ala Gly Phe Gly Tyr Asp His Ile Arg Leu Pro Ile Asp
65 70 75 80

Glu Glu Gln Met Trp Asp Glu Ser Gly Asn Lys Glu Pro Arg Ala Phe
85 90 95

Glu Leu Leu His Glu Ala Ile Gly Trp Ala Leu Asp Asn Glu Leu Arg
100 105 110

Val Ile Val Asp Leu His Ile Ile Arg Ser His Tyr Phe Asn Ala Pro
115 120 125

Glu Asn Pro Leu Trp Thr Asp Arg Ala Glu Gln Leu Lys Phe Val Glu
130 135 140

Met Trp Arg Gln Leu Ser Asp Glu Leu Gln Gly Tyr Pro Leu Asp Arg
145 150 155 160

Val Ala Tyr Glu Leu Met Asn Glu Ala Val Ala Asp Asp Pro Asp Asp
165 170 175

Trp Asn Arg Leu Val Ala Glu Thr Met Glu Ala Leu Arg Met Leu Glu
180 185 190

Pro Glu Arg Lys Ile Val Ile Gly Ser Asn Arg Trp Gln Ser Val His
195 200 205

Thr Phe Pro Asp Leu Val Ile Pro Asp Asn Asp Pro His Ile Ile Leu
210 215 220

Ser Phe His Phe Tyr Glu Pro Phe Leu Leu Thr His His Lys Ala Ser
225 230 235 240

Trp Thr His Ile Arg Asp Tyr Thr Gly Pro Val Asn Tyr Pro Gly Leu
245 250 255

Thr Val Asp Pro Thr His Leu Glu Gly Leu Ser Glu Glu Leu Val Thr
260 265 270

Arg Ile Gly His His Asn Gly Val Tyr Thr Lys Glu Thr Met Glu Glu
275 280 285

Met Ile Met Ile Pro Leu Gln Tyr Ala Lys Asp Arg Gly Leu Pro Leu
290 295 300

Tyr Cys Gly Glu Trp Gly Cys Phe Pro Thr Met Pro Gln Glu Met Arg
305 310 315 320

Leu Gln Trp Tyr Ala Asp Val Arg Ala Ile Leu Glu Lys His Glu Ile
325 330 335

Ala Trp Ala Asn Trp Asp Tyr Lys Gly Gly Phe Gly Val Val Asp Arg
340 345 350

Asn Gly Glu Pro His His Asp Leu Leu Glu Val Leu Leu Lys
355 360 365

<210> 105

<211> 1047

<212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 105
atgcaacact tcatacaagg cgtcaacctg ggaggctggc tctcccaata ccagaaatac 60

gaccatgagc acttcgcac ctatcacacc cggcgcgata tcgaacaaat cgcatacctgg 120

ggcttcgacc acatccgctt gccggtcgat tatccggttc tcgaatcgga cgacgcgccc 180

ggatatctatc atgaagatgg ctttcctat cttgactctt gcctggaatg gtgccaggcc 240

gctgggctgg cagtcgtctt cgacctgcat catgccccg gctacagttt cacgaacacg 300

ctcaagcctg aaacctgca cctgaacgta ctcttcgagc aggaaatcgc ccaaaatcga 360

tttatcgccc tcigggaac cattgttcgg cgtaccagg ccgccggctt gcctatcatc 420

tttgaactac tgaatgaaat ggtgctgcca gacagcggcc cctggaacgc cctggcccac 480

aaaaccgtcg ccgccctcg acagatttcg cccgattgca aaatcatgat tggcggcaat 540

aactacaacg ccgcatccga actcaaaaac ataacctgc acaacgaccc caacatccta 600

tacaccttc atttctacga accggccctg ttcaccacc agaaagcccc ctgggtgcag 660

attgctgtcg aatacaacca ggaactcgaa taccctggct cgtacacaa cctggccgcc 720

tttctccgc gcaatccca ctatcaagaa tcctatggat ggcaggtaa ccgccgtatc 780

gaccgcgacc tctgtctga attcaccaa ccgccctgg actttgtcca gcagaccggg 840

cgcgacctgt actcgggtga attcggcgtc attgaatag tcgagcctgc cagccgcaa 900

aactggcacg ccgacctgt ggacatcctg cgccagcaga agattggccg cgccgtctgg 960

acttataaac aaatggattt tggcctgggt gacgcggacg gcaagggtgt cgacccaaa 1020

cttctcgaaa tctgtgtca atcctga 1047

<210> 106

<211> 348

<212> PRT

<213> Unknown

<220>

<223> Obtained from environmental sample

<220>

<221> DOMAIN

<222> (2)...(330)

<223> Cellulase (glycosyl hydrolase family 5)

<220>

<221> SITE

<222> (192)...(195)

<223> N-glycosylation site. Prosite id = PS00001

<400> 106

Met	Gln	His	Phe	Ile	Asn	Gly	Val	Asn	Leu	Gly	Gly	Trp	Leu	Ser	Gln
1				5					10					15	

Tyr	Gln	Lys	Tyr	Asp	His	Glu	His	Phe	Arg	Thr	Phe	Ile	Thr	Arg	Arg
			20					25					30		

Asp	Ile	Glu	Gln	Ile	Ala	Ser	Trp	Gly	Phe	Asp	His	Ile	Arg	Leu	Pro
		35					40					45			

Val	Asp	Tyr	Pro	Val	Leu	Glu	Ser	Asp	Asp	Ala	Pro	Gly	Ile	Tyr	His
		50				55						60			

Glu	Asp	Gly	Phe	Ala	Tyr	Leu	Asp	Ser	Cys	Leu	Glu	Trp	Cys	Gln	Ala
65					70					75					80

Ala	Gly	Leu	Ala	Val	Val	Phe	Asp	Leu	His	His	Ala	Pro	Gly	Tyr	Ser
				85					90					95	

Phe	Thr	Asn	Thr	Leu	Lys	Pro	Glu	Thr	Leu	His	Leu	Asn	Val	Leu	Phe
				100					105				110		

Glu	Gln	Glu	Ile	Ala	Gln	Asn	Arg	Phe	Ile	Ala	Leu	Trp	Glu	Thr	Ile

115	120	125
Val Arg Arg Tyr Gln Ala Ala Gly Leu Pro Ile Ile Phe Glu Leu Leu		
130	135	140
Asn Glu Met Val Leu Pro Asp Ser Gly Pro Trp Asn Ala Leu Ala His		
145	150	155 160
Lys Thr Val Ala Ala Leu Arg Gln Ile Ser Pro Asp Cys Lys Ile Met		
165	170	175
Ile Gly Gly Asn Asn Tyr Asn Ala Ala Ser Glu Leu Lys Asn Ile Thr		
180	185	190
Leu His Asn Asp Pro Asn Ile Leu Tyr Thr Phe His Phe Tyr Glu Pro		
195	200	205
Ala Leu Phe Thr His Gln Lys Ala Pro Trp Val Gln Ile Ala Val Glu		
210	215	220
Tyr Asn Gln Glu Leu Glu Tyr Pro Gly Ser Tyr Thr Asn Leu Ala Ala		
225	230	235 240
Phe Leu Arg Arg Asn Pro His Tyr Gln Glu Ser Tyr Gly Trp Gln Val		
245	250	255
Asn Arg Arg Ile Asp Arg Asp Leu Leu Leu Glu Phe Thr Gln Pro Ala		
260	265	270
Leu Asp Phe Val Gln Gln Thr Gly Arg Asp Leu Tyr Cys Gly Glu Phe		
275	280	285
Gly Val Ile Glu Tyr Val Glu Pro Ala Ser Arg Gln Asn Trp His Ala		
290	295	300
Asp Leu Leu Asp Ile Leu Arg Gln Gln Lys Ile Gly Arg Ala Val Trp		

305 310 315 320

Thr Tyr Lys Gln Met Asp Phe Gly Leu Val Asp Ala Asp Gly Lys Val
325 330 335

Val Asp Pro Lys Leu Leu Glu Ile Leu Cys Gln Ser
340 345

<210>	107
<211>	1137
<212>	DNA
<213>	Unknown

<220>
<223> Obtained from environmental sample

<400> 107	
atggaaaagc aaatctgttc aaatgttttc agtacgatgc tgataattgg aggccttctt	60
gttttactgg gattttcttc ttgcgggcgg caggcagaac ctgctgctga ctctttcagg	120
gggtttcacg actttgacat caggcgcggg gtgaacatca gccattggtt gtcgcagagt	180
ggaaggcgtg gtgctgatcg ggaggcggtc ttaccaggg cggatgtgga ggccatcgcc	240
ggcttcggtt atgatcacat tcgtttgcc atcgatgaag agcagatgtg ggatgagtcg	300
ggcaacaagg agccacgtgc ctttgaattg ctgcatgagg ccattggctg ggctttggac	360
aatgagctca gggtcattgt tgacctgcac atcatcaggt cgcactattt taatgcgcct	420
gagaaccgcg tttggaccga tcgtgctgaa cagttgaaat ttgttgagat gtggcgacag	480
ttgtctgatg agctgcaggg ctatccgctc gatagggtgg cctatgaatt gatgaatgag	540
gccgtggctg atgatccgga cgattggaac cggcttgtgg ctgagacgat ggaggcgcta	600
cggatgctgg aaccggagcg caagattgtc attggctcca accgctggca gtctgtgcat	660

acatttcctg acctggtgat cccggataat gaccgcata tcataattgag ttttcacttc 720

tacgaacat ttctgctgac gcaccacaag gcctcctgga cacacatccg tgattacacc 780

ggtccggtga aciatccggg ttgactgia gaccgcaccc acctggaggg gtgtctgaa 840

gaactggtga cccggattgg ccatcacaat ggggtgtata caaaagaaac gatggaggag 900

atgatcatga tccactgca atatgcaaaa gaacgggggc tcccctgta ttgcggggag 960

tgggatgtt tccgacat gccccaggag atgcgcctgc aatggtacgc cgatgtgcgt 1020

gcgatcctgg aaaagcatga gattgcctgg gaaactggg attacaaggg tggtttcggt 1080

gtggttgacc gcaacggcga accccacat gatttattgg aagtcttact aaaataa 1137

<210> 108
 <211> 378
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> SIGNAL
 <222> (1)...(32)

<220>
 <221> DOMAIN
 <222> (54)...(361)
 <223> Cellulase (glycosyl hydrolase family 5)

<220>
 <221> SITE
 <222> (52)...(55)
 <223> N-glycosylation site. Prosite id = PS00001

<400> 108
 Met Glu Lys Gln Ile Cys Ser Asn Val Phe Ser Thr Met Leu Ile Ile
 1 5 10 15

Gly Gly Leu Leu Val Leu Leu Gly Phe Ser Ser Cys Gly Arg Gln Ala
20 25 30

Glu Pro Ala Ala Asp Ser Phe Arg Gly Phe His Asp Phe Asp Ile Arg
35 40 45

Arg Gly Val Asn Ile Ser His Trp Leu Ser Gln Ser Gly Arg Arg Gly
50 55 60

Ala Asp Arg Glu Ala Phe Phe Thr Arg Ala Asp Val Glu Ala Ile Ala
65 70 75 80

Gly Phe Gly Tyr Asp His Ile Arg Leu Pro Ile Asp Glu Glu Gln Met
85 90 95

Trp Asp Glu Ser Gly Asn Lys Glu Pro Arg Ala Phe Glu Leu Leu His
100 105 110

Glu Ala Ile Gly Trp Ala Leu Asp Asn Glu Leu Arg Val Ile Val Asp
115 120 125

Leu His Ile Ile Arg Ser His Tyr Phe Asn Ala Pro Glu Asn Pro Leu
130 135 140

Trp Thr Asp Arg Ala Glu Gln Leu Lys Phe Val Glu Met Trp Arg Gln
145 150 155 160

Leu Ser Asp Glu Leu Gln Gly Tyr Pro Leu Asp Arg Val Ala Tyr Glu
165 170 175

Leu Met Asn Glu Ala Val Ala Asp Asp Pro Asp Asp Trp Asn Arg Leu
180 185 190

Val Ala Glu Thr Met Glu Ala Leu Arg Met Leu Glu Pro Glu Arg Lys
195 200 205

Ile Val Ile Gly Ser Asn Arg Trp Gln Ser Val His Thr Phe Pro Asp
210 215 220

Leu Val Ile Pro Asp Asn Asp Pro His Ile Ile Leu Ser Phe His Phe
225 230 235 240

Tyr Glu Pro Phe Leu Leu Thr His His Lys Ala Ser Trp Thr His Ile
245 250 255

Arg Asp Tyr Thr Gly Pro Val Asn Tyr Pro Gly Leu Thr Val Asp Pro
260 265 270

Thr His Leu Glu Gly Leu Ser Glu Glu Leu Val Thr Arg Ile Gly His
275 280 285

His Asn Gly Val Tyr Thr Lys Glu Thr Met Glu Glu Met Ile Met Ile
290 295 300

Pro Leu Gln Tyr Ala Lys Glu Arg Gly Leu Pro Leu Tyr Cys Gly Glu
305 310 315 320

Trp Gly Cys Phe Pro Thr Met Pro Gln Glu Met Arg Leu Gln Trp Tyr
325 330 335

Ala Asp Val Arg Ala Ile Leu Glu Lys His Glu Ile Ala Trp Ala Asn
340 345 350

Trp Asp Tyr Lys Gly Gly Phe Gly Val Val Asp Arg Asn Gly Glu Pro
355 360 365

His His Asp Leu Leu Glu Val Leu Leu Lys
370 375

<210> 109

<211> 1248

<212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 109

atgaagacac atagcttcaa ctcagatca cggatcacct tgttgaccgc ggcactgctt	60
ttcatcgggg caacggccgg ggccgccacg acacctatca ccctcaaaga cgcctacaaa	120
gaccatttcc ttatgggtgt agccatcaac cgcctgattg caatgggcga tacgaatgtc	180
cgggccgaca acgccagcgg gaccccgaa cagctcaagg gggacattgc cctggtcaag	240
gcgcagtcca acctgatcgt caatgagaac gatctgaaac cgatttcat tcacccgagg	300
ccaggaccgg acgggtacga cttcgcccca gcggatgcct tcgtgaagtt cggcatggac	360
aacaatatgt atatcgtggg ccacaccctc ctctggcaca gccagggtgc caactggttc	420
ttccaagggt ctgctccggc gactccggaa acgccacctg ctgccacgga cgcggcggtc	480
gcaccccgcg gcggacgagg aggtcgcggc gggattaccg gcccctggc gaccccgag	540
gagttgatcg aacgcatcgg cgagcacatt cacaccgtcg tcggccgcta taagggaag	600
atcaaggctt gggacgtcgt caacgaagcc ctgcgcgacg gcggcaccga gaccctgcga	660
agcacgtact ggacccaaat catcgggccc gaattcatcg ccatggcctt tcgattcgcc	720
cacgaagccg atccgatgc gatccttcgt tacaacgatt atggcctgga gaacctgcc	780
aagcgtgaga aactcaagaa gctgatcgcg tcgctccagg agcagaacgt tccggttcat	840
gccatcgga cgaacacca tatcagcgtc tccacgacgt tcgaaagaat ggatgagacc	900
ttgagggacc tggcatccat cgggttccc gtccacatca ccgaactgga tgtcaacgcc	960
gccgcggggg gccagagggg caccaatcgc gacattgccg gcactgccga gcgtacggcg	1020
ggcggcgtgg tcagtgaagc cgacaagcgg ctggccgacg cctacgcgaa tctcttcgc	1080

gcgatcatga agcacaagga ctcggtgaag atggtcacgt tctggggcgt caatgacgcg 1140

gtttcgtggc tcgcacgcgg caccgccgtg ctgttcgacg gcaacaatca gcccaagccg 1200

gctttcgatg cggtcattcg cgtcgccacg gaggcggcac agaactga 1248

<210> 110

<211> 415

<212> PRT

<213> Unknown

<220>

<223> Obtained from environmental sample

<220>

<221> SIGNAL

<222> (1)...(28)

<220>

<221> DOMAIN

<222> (34)...(409)

<223> Glycosyl hydrolase family 10

<220>

<221> SITE

<222> (312)...(322)

<223> Glycosyl hydrolases family 10 active site. Prosite id = PS00591

<400> 110

Met Lys Thr His Ser Phe Asn Leu Arg Ser Arg Ile Thr Leu Leu Thr
1 5 10 15

Ala Ala Leu Leu Phe Ile Gly Ala Thr Ala Gly Ala Ala Thr Thr Pro
20 25 30

Ile Thr Leu Lys Asp Ala Tyr Lys Asp His Phe Leu Met Gly Val Ala
35 40 45

Ile Asn Arg Leu Ile Ala Met Gly Asp Thr Asn Val Arg Ala Asp Asn
50 55 60

Ala Ser Arg Thr Pro Glu Gln Leu Lys Gly Asp Ile Ala Leu Val Lys
65 70 75 80

Ala Gln Phe Asn Leu Ile Val Asn Glu Asn Asp Leu Lys Pro Ile Leu
85 90 95

Ile His Pro Arg Pro Gly Pro Asp Gly Tyr Asp Phe Ala Pro Ala Asp
100 105 110

Ala Phe Val Lys Phe Gly Met Asp Asn Asn Met Tyr Ile Val Gly His
115 120 125

Thr Leu Leu Trp His Ser Gln Val Pro Asn Trp Phe Phe Gln Gly Ser
130 135 140

Ala Pro Ala Thr Pro Glu Thr Pro Pro Ala Ala Thr Asp Ala Ala Val
145 150 155 160

Ala Pro Arg Gly Gly Arg Gly Gly Arg Gly Gly Ile Thr Gly Pro Leu
165 170 175

Ala Thr Arg Glu Glu Leu Ile Glu Arg Met Arg Glu His Ile His Thr
180 185 190

Val Val Gly Arg Tyr Lys Gly Lys Ile Lys Val Trp Asp Val Val Asn
195 200 205

Glu Ala Leu Ala Asp Gly Gly Thr Glu Thr Leu Arg Ser Thr Tyr Trp
210 215 220

Thr Gln Ile Ile Gly Pro Glu Phe Ile Ala Met Ala Phe Arg Phe Ala
225 230 235 240

His Glu Ala Asp Pro Asp Ala Ile Leu Arg Tyr Asn Asp Tyr Gly Leu
245 250 255

Glu Asn Pro Ala Lys Arg Glu Lys Leu Lys Lys Leu Ile Ala Ser Leu
260 265 270

Gln Glu Gln Asn Val Pro Val His Ala Ile Gly Thr Gln Thr His Ile
275 280 285

Ser Val Ser Thr Thr Phe Glu Arg Met Asp Glu Thr Leu Arg Asp Leu
290 295 300

Ala Ser Ile Gly Leu Pro Val His Ile Thr Glu Leu Asp Val Asn Ala
305 310 315 320

Ala Ala Gly Gly Gln Arg Gly Thr Asn Ala Asp Ile Ala Gly Thr Ala
325 330 335

Glu Arg Thr Ala Gly Gly Val Val Ser Glu Ala Asp Lys Arg Leu Ala
340 345 350

Asp Ala Tyr Ala Asn Leu Phe Arg Ala Ile Met Lys His Lys Asp Ser
355 360 365

Val Lys Met Val Thr Phe Trp Gly Val Asn Asp Ala Val Ser Trp Leu
370 375 380

Ala Arg Gly Thr Pro Leu Leu Phe Asp Gly Asn Asn Gln Pro Lys Pro
385 390 395 400

Ala Phe Asp Ala Val Ile Arg Val Ala Thr Glu Ala Ala Gln Asn
405 410 415

<210> 111

<211> 1131

<212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 111

atgcgaagac tgatcacat catccttgcg acggctgtcg caatcttata gaccacatca	60
tgctccaaga ccgtgaacg agagggcttc ttgatcaagc gaggaacaa ctcagccat	120
tggctctccc agagcaagga aaggggagag gctcgcaggc tccatatcca ggaggatgac	180
tttgctcgcc tcgacagcct cggtttcgac catgtgcgca tccctgtcga cgaggaacaa	240
ctctgggacg aggatggcaa caagtcaca gaagcatggg aactgctcga ttctgcctc	300
gacatggcgc gcaagtacaa cctgcgcgct atcgtggacc ttacatcat ccgcgccat	360
tacttcaacg ccgtcaacga aggcgcgtcg aatactctct tcaccagcga ggaggcgag	420
cagggcctga tcaaccttig gtaccagctt tccgacttcc tcaaggaccg cagcgtcgac	480
tgggttgctt acgagttcat gaacgagcca gtcgctcctg agcatgagca atggaacgcc	540
ctcgtcgcaa aggtgcacaa ggcgcttcgt gagcgtgaac cggagcgtag cctcgtgatc	600
ggttctaacc tgtggcaggg tcaccagacc ttcaagtacc tccgcgtacc tgagaatgac	660
ccgaacatca tcttgagctt ccactactac aacccttcga tcttcacca caacatggct	720
ccgtggactc cggtgggcaa atataccggt tccatcaatt atccgggcgt catcgtctct	780
gctgaggatt acgtgcgca gagccctgag gtgcaggccg aggtgaagca gtatacggag	840
atggtctgga accgcgacac gatctacagc cagatgaagg atcgatcga ggtggctgcc	900
agctatggac tgcagctctt ctgcggcgaa tggggcgtgt atgaacctgt cgaccgtgag	960
cttgcatagc catggaccaa ggataatgctg tcggtgttcg acgagttcga catcgcatgg	1020
acgacctggt gttacgatgc cgacttcggc ttctgggacc aggcgaaaca tgatttcaag	1080

gacaagcctc ttgtcgatct cctgatgggt tccaagggtc ttgaacaata g 1131

<210> 112
<211> 376
<212> PRT
<213> Unknown

<220>
<223> Obtained from environmental sample

<220>
<221> SIGNAL
<222> (1)...(22)

<220>
<221> DOMAIN
<222> (39)...(353)
<223> Cellulase (glycosyl hydrolase family 5)

<220>
<221> SITE
<222> (37)...(40)
<223> N-glycosylation site. Prosite id = PS00001

<400> 112
Met Arg Arg Leu Ile Thr Ile Ile Leu Ala Thr Ala Val Ala Ile Leu
1 5 10 15

Ser Thr Thr Ser Cys Ser Lys Thr Ala Glu Arg Glu Gly Phe Leu Ile
20 25 30

Lys Arg Gly Thr Asn Leu Ser His Trp Leu Ser Gln Ser Lys Glu Arg
35 40 45

Gly Glu Ala Arg Arg Leu His Ile Gln Glu Asp Asp Phe Ala Arg Leu
50 55 60

Asp Ser Leu Gly Phe Asp His Val Arg Ile Pro Val Asp Glu Glu Gln
65 70 75 80

Leu Trp Asp Glu Asp Gly Asn Lys Leu Thr Glu Ala Trp Glu Leu Leu

85

90

95

Asp Phe Ala Leu Asp Met Ala Arg Lys Tyr Asn Leu Arg Ala Ile Val
100 105 110

Asp Leu His Ile Ile Arg Ala His Tyr Phe Asn Ala Val Asn Glu Gly
115 120 125

Ala Ser Asn Thr Leu Phe Thr Ser Glu Glu Ala Gln Gln Gly Leu Ile
130 135 140

Asn Leu Trp Tyr Gln Leu Ser Asp Phe Leu Lys Asp Arg Ser Val Asp
145 150 155 160

Trp Val Ala Tyr Glu Phe Met Asn Glu Pro Val Ala Pro Glu His Glu
165 170 175

Gln Trp Asn Ala Leu Val Ala Lys Val His Lys Ala Leu Arg Glu Arg
180 185 190

Glu Pro Glu Arg Thr Leu Val Ile Gly Ser Asn Leu Trp Gln Gly His
195 200 205

Gln Thr Phe Lys Tyr Leu Arg Val Pro Glu Asn Asp Pro Asn Ile Ile
210 215 220

Leu Ser Phe His Tyr Tyr Asn Pro Ser Ile Leu Thr His Asn Met Ala
225 230 235 240

Pro Trp Thr Pro Val Gly Lys Tyr Thr Gly Ser Ile Asn Tyr Pro Gly
245 250 255

Val Ile Val Ser Ala Glu Asp Tyr Ala Ala Gln Ser Pro Glu Val Gln
260 265 270

Ala Glu Val Lys Gln Tyr Thr Glu Met Val Trp Asn Arg Asp Thr Ile

275	280	285
Tyr Ser Gln Met Lys Asp Ala Ile Glu Val Ala Ala Ser Tyr Gly Leu 290 295 300		
Gln Leu Phe Cys Gly Glu Trp Gly Val Tyr Glu Pro Val Asp Arg Glu 305 310 315 320		
Leu Ala Tyr Ala Trp Thr Lys Asp Met Leu Ser Val Phe Asp Glu Phe 325 330 335		
Asp Ile Ala Trp Thr Thr Trp Cys Tyr Asp Ala Asp Phe Gly Phe Trp 340 345 350		
Asp Gln Ala Lys His Asp Phe Lys Asp Lys Pro Leu Val Asp Leu Leu 355 360 365		
Met Gly Ser Lys Gly Leu Glu Gln 370 375		
<210> 113 <211> 1095 <212> DNA <213> Unknown		
<220> <223> Obtained from environmental sample		
<400> 113 atgaaggtga cccgaacagc tgtcgcgggc attgtcgccg cagcggtcct catcacgac 60 ggcacgtcga ccgcgtcggc tgaggatgaa ccaaccagcg agaacacgtc gacggatcag 120 ccgttgcgcg tcttggcagc caaagccggg atcgcggttcg gcacggccgt cgacatgaac 180 gcgtacaaca acgacgcgac ctaccgtgag ctcgctggcc aggagttctc gagcgtcacg 240 gccgagaacg tcatgaagtg gcagctcttc gagccgcagc gaggggtcta caactggggt 300		

ccggccgac agctcgtgcg cgtagccaac gagaacggcc agaaggtgcg cgggcacacg 360

ctcatctggc acaaccagct gccacactgg cttaccagcg gagtcgcctc cggtagatc 420

acaccggacg agctccggca gctcctgagg aaccacatct tcacggtgat gcgccacttc 480

aaggcgagaga tccaccagtg ggatgtcgcc aacgaggtca tcgacgacag cggcaacctg 540

cgcaacacga tctggctgca gaacctgggt ccgagctaca tcgcggaacg gttccggtgg 600

gctcgcaagg ccgacccgga cgccgccctc tatctgaacg actacaacgt cgagggcccg 660

aacgccaagg ccgatgcgta ctacgccctg gtcaagcagc tcctcgccga cgacgtgccg 720

gtggacggct tcggaatata ggggcacctc ggtgtgcagt tcggcttctg gcccgcgagt 780

gcggtggccg acaacatggg gcgcttcgag gcactcggcc tgcagacggc ggtcaccgag 840

gcggatgtcc ggatgatcat gccgcccgc gaggacaagc tggccgcaca ggcacgtggc 900

tacagcacgt tggccagggt ctgcctgatg gccaaagcgtt gcaggtcgtt caccgtctgg 960

ggcttcaccg acaagtactc ctgggttccg ggcaccttcc ccggccaggg cgcggcgaac 1020

ctcctggccg aggacttcca gccaagccg gcttactacg ccgtccagga tgacctcgcg 1080

cgcgccggac ggtag 1095

<210> 114
 <211> 364
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> SIGNAL
 <222> (1)...(27)

<220>

<221> DOMAIN

<222> (41)...(359)

<223> Glycosyl hydrolase family 10

<220>

<221> SITE

<222> (35)...(38)

<223> N-glycosylation site. Prosite id = PS00001

<400> 114

Met Lys Val Thr Arg Thr Ala Val Ala Gly Ile Val Ala Ala Ala Val
1 5 10 15

Leu Ile Thr Ile Gly Thr Ser Thr Ala Ser Ala Glu Asp Glu Pro Thr
20 25 30

Ser Glu Asn Thr Ser Thr Asp Gln Pro Leu Arg Val Leu Ala Ala Lys
35 40 45

Ala Gly Ile Ala Phe Gly Thr Ala Val Asp Met Asn Ala Tyr Asn Asn
50 55 60

Asp Ala Thr Tyr Arg Glu Leu Val Gly Gln Glu Phe Ser Ser Val Thr
65 70 75 80

Ala Glu Asn Val Met Lys Trp Gln Leu Leu Glu Pro Gln Arg Gly Val
85 90 95

Tyr Asn Trp Gly Pro Ala Asp Gln Leu Val Arg Val Ala Asn Glu Asn
100 105 110

Gly Gln Lys Val Arg Gly His Thr Leu Ile Trp His Asn Gln Leu Pro
115 120 125

Thr Trp Leu Thr Ser Gly Val Ala Ser Gly Glu Ile Thr Pro Asp Glu
130 135 140

Leu Arg Gln Leu Leu Arg Asn His Ile Phe Thr Val Met Arg His Phe
145 150 155 160

Lys Gly Glu Ile His Gln Trp Asp Val Ala Asn Glu Val Ile Asp Asp
165 170 175

Ser Gly Asn Leu Arg Asn Thr Ile Trp Leu Gln Asn Leu Gly Pro Ser
180 185 190

Tyr Ile Ala Asp Ala Phe Arg Trp Ala Arg Lys Ala Asp Pro Asp Ala
195 200 205

Ala Leu Tyr Leu Asn Asp Tyr Asn Val Glu Gly Pro Asn Ala Lys Ala
210 215 220

Asp Ala Tyr Tyr Ala Leu Val Lys Gln Leu Leu Ala Asp Asp Val Pro
225 230 235 240

Val Asp Gly Phe Gly Ile Gln Gly His Leu Gly Val Gln Phe Gly Phe
245 250 255

Trp Pro Ala Ser Ala Val Ala Asp Asn Met Gly Arg Phe Glu Ala Leu
260 265 270

Gly Leu Gln Thr Ala Val Thr Glu Ala Asp Val Arg Met Ile Met Pro
275 280 285

Pro Asp Glu Asp Lys Leu Ala Ala Gln Ala Arg Gly Tyr Ser Thr Leu
290 295 300

Val Gln Gly Cys Leu Met Ala Lys Arg Cys Arg Ser Phe Thr Val Trp
305 310 315 320

Gly Phe Thr Asp Lys Tyr Ser Trp Val Pro Gly Thr Phe Pro Gly Gln
325 330 335

Gly Ala Ala Asn Leu Leu Ala Glu Asp Phe Gln Pro Lys Pro Ala Tyr
340 345 350

Tyr Ala Val Gln Asp Asp Leu Ala Arg Ala Gly Arg
355 360

<210> 115
<211> 774
<212> DNA
<213> Unknown

<220>
<223> Obtained from environmental sample

<400> 115
atggacttgc agctaggcgg aaagcgcgtg ctgatcacgg gtgctgctcaa aggcacgcgc 60

ctggcctgcg ccgtcgcctt tgcgcgcgag ggtgccgacc cgattctggt ggcgcgcgat 120

gatgcggcgt tgcacacgc cactccgcc atccgcgaac aaagcggccg cgcggcacat 180

gccatcacgc tggacctggc cctgcctggc gcggcggaag agctggccaa ggaaaccggc 240

cccacgcaca tactggtcaa caacgcgggc gcggtgcccg gcggcgcgct ggaccagggt 300

caagacgaac gctggcgcgc gggctgggaa ttgaaagtgc acggctacat cagcctggcg 360

cgctgtact accgcacat gcgcgaagcg gcgcggggcg tcacgcgcaa catcatcggc 420

atggcgggcg cggcgccccg cgccgactac atctgcggcg cggcggccaa tgcctcactg 480

attgccttta cccgcgcgct gggcggcgaa gcgccccgcc acggcgtgcg cgtctttggc 540

gtcaaccctt cgcgcacgcg gaccgaccgc gtgctgacct tggcccgca acgcgcgcag 600

gcgcgctggg gcgacgaaac ccgttggcag gaaacgtgt cgacactgcc cttcaaccgg 660

ctgatggaac ccgccgaagt ggccgacatg attgtgttcg gcgcctcgcc gcgcgcgggt 720

tacctgagcg gcacggcat cgacctggac ggcggcgaac agtacgcgaa atag

774

<210> 116
<211> 257
<212> PRT
<213> Unknown

<220>
<223> Obtained from environmental sample

<220>
<221> DOMAIN
<222> (8)...(172)
<223> short chain dehydrogenase

<220>
<221> SITE
<222> (159)...(162)
<223> N-glycosylation site. Prosite id = PS00001

<400> 116
Met Asp Leu Gln Leu Gly Gly Lys Arg Val Leu Ile Thr Gly Ala Ser
1 5 10 15

Lys Gly Ile Gly Leu Ala Cys Ala Val Ala Phe Ala Arg Glu Gly Ala
20 25 30

Asp Pro Ile Leu Val Ala Arg Asp Asp Ala Ala Leu His His Ala Thr
35 40 45

Ser Ala Ile Arg Glu Gln Ser Gly Arg Ala Ala His Ala Ile Thr Leu
50 55 60

Asp Leu Ala Leu Pro Gly Ala Ala Glu Lys Leu Ala Lys Glu Thr Gly
65 70 75 80

Pro Ile Asp Ile Leu Val Asn Asn Ala Gly Ala Val Pro Gly Gly Ala
85 90 95

Leu Asp Gln Val Gln Asp Glu Arg Trp Arg Ala Gly Trp Glu Leu Lys
100 105 110

Val His Gly Tyr Ile Ser Leu Ala Arg Cys Tyr Tyr Pro His Met Arg
115 120 125

Glu Ala Gly Ala Gly Val Ile Ala Asn Ile Ile Gly Met Ala Gly Ala
130 135 140

Ala Pro Arg Ala Asp Tyr Ile Cys Gly Ala Ala Ala Asn Ala Ser Leu
145 150 155 160

Ile Ala Phe Thr Arg Ala Leu Gly Gly Glu Ala Pro Arg His Gly Val
165 170 175

Arg Val Phe Gly Val Asn Pro Ser Arg Thr Arg Thr Asp Arg Val Leu
180 185 190

Thr Leu Ala Arg Gln Arg Ala Gln Ala Arg Trp Gly Asp Glu Thr Arg
195 200 205

Trp Gln Glu Thr Leu Ser Asp Leu Pro Phe Asn Arg Leu Met Glu Pro
210 215 220

Ala Glu Val Ala Asp Met Ile Val Phe Gly Ala Ser Pro Arg Ala Gly
225 230 235 240

Tyr Leu Ser Gly Thr Val Ile Asp Leu Asp Gly Gly Glu Gln Tyr Ala
245 250 255

Lys

<210> 117

<211> 747

<212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 117

atgcccaaag tcatgctcgt taccggcggc agccgtggca tcggcgccgc cgtcgccaag	60
ctggccgcgc gccgcggcta cgcggtcggc atcaactacc gcacccattc cgacgcgcgc	120
gacgccgtcg tggccgagat ccagcaggcg ggccggcaccg cgctggccat ccaggccgac	180
gtgtcgcaag aagatgacgt gctgcacatg ttccgcacgc tggacgagcg cctgggccgc	240
atcgacgcgc tggtaataa cgccggcatc ctggaaacgc agatgcgcct ggaccagatg	300
gaagcggacc gcctgtcgcg cgtgtgttcc accaacgtca tcggcgcttt cctgtgtgcg	360
cgcaagcgg tgcgcaggat gtcgacgcgc catggcggcg tggcgggcgc catcgtcaac	420
gtgtcttcgg cgccggcgcg cctgggctcg cccaatgaat acgtggatta cgcggcctcc	480
aagggcgcg tggacacgat gaccatcggc ctgtccaaag aggtagcgcc cgaaggtatc	540
cgcgtgaatg gcgtagcccc cggcaccatc tacaccgaca tgcacgaag cggcggcgag	600
ccgggccggg tggatcgctt gaaaagcgtg atcccgtgc ggccggcgcg ctcggtggaa	660
gaagtggcgg gcgccgtcat gtggctgttt tccgaagaag ccggtatatac cagcggtcgc	720
ttcatcgacg tgiccgcgcg tagttga	747

<210> 118

<211> 248

<212> PRT

<213> Unknown

<220>

<223> Obtained from environmental sample

<220>
 <221> DOMAIN
 <222> (3)...(176)
 <223> short chain dehydrogenase

<220>
 <221> SITE
 <222> (142)...(145)
 <223> N-glycosylation site. Prosite id = PS00001

<220>
 <221> SITE
 <222> (146)...(174)
 <223> Short-chain dehydrogenases/reductases family signature. Prosite id = PS00061

<400> 118
 Met Pro Lys Val Met Leu Val Thr Gly Gly Ser Arg Gly Ile Gly Ala
 1 5 10 15

Ala Val Ala Lys Leu Ala Ala Arg Arg Gly Tyr Ala Val Gly Ile Asn
 20 25 30

Tyr Arg Thr His Ser Asp Ala Ala Asp Ala Val Val Ala Glu Ile Gln
 35 40 45

Gln Ala Gly Gly Thr Ala Leu Ala Ile Gln Ala Asp Val Ser Gln Glu
 50 55 60

Asp Asp Val Leu His Met Phe Arg Thr Leu Asp Glu Arg Leu Gly Arg
 65 70 75 80

Ile Asp Ala Leu Val Asn Asn Ala Gly Ile Leu Glu Thr Gln Met Arg
 85 90 95

Leu Asp Gln Met Glu Ala Asp Arg Leu Leu Arg Val Leu Ser Thr Asn
 100 105 110

Val Ile Gly Ala Phe Leu Cys Ala Arg Glu Ala Val Arg Arg Met Ser
 115 120 125

Thr Arg His Gly Gly Val Gly Gly Ala Ile Val Asn Val Ser Ser Ala
130 135 140

Ala Ala Arg Leu Gly Ser Pro Asn Glu Tyr Val Asp Tyr Ala Ala Ser
145 150 155 160

Lys Gly Ala Leu Asp Thr Met Thr Ile Gly Leu Ser Lys Glu Val Ala
165 170 175

Pro Glu Gly Ile Arg Val Asn Gly Val Arg Pro Gly Thr Ile Tyr Thr
180 185 190

Asp Met His Ala Ser Gly Gly Glu Pro Gly Arg Val Asp Arg Leu Lys
195 200 205

Ser Val Ile Pro Leu Arg Arg Gly Gly Ser Val Glu Glu Val Ala Gly
210 215 220

Ala Val Met Trp Leu Phe Ser Glu Glu Ala Gly Tyr Thr Ser Gly Ser
225 230 235 240

Phe Ile Asp Val Ser Gly Gly Ser
245

<210> 119

<211> 1611

<212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 119

atgcaaaagc ggtatgacgt cattgtcgtg ggcagcggga tcgccggcct cagttttgcg 60

ctaaaagtgc ccaaggcggg gcatcgcgta gggattttga ccaaaaaaga ccgtgctgaa 120

agcaacacca attatgccca aggcggcatc gcggcagtca cttcgagac agatgatttc 180

gagctgcatg tgcaggacac attgaccgcg ggagatggac tctgcgacga ggcagtcgtc 240

cgcacgatta tcggcgaggc tcccggcga atccaggagc tgatcgattt gggggtggcc 300

ttctcacatt tggaagatgg acgggtttcc ctccatcgcg aaggggttca ctgaaaagg 360

cgcattcttc acgttcagga tgtcaccggc aaagcgattg aagaagcct cctccatgcc 420

atcgaacagt cggcgtgat cgacctgaat gacacgtct ttgccatcga ctactgact 480

gaacgcaagc tggcgtggc gggctttgag gtggaagggtg ctaaaaaccg ggtggtcgga 540

ctctatgcgc tcgatgaagc cactcaggag gttcacgtat ttgaggctcc agtcgtcatg 600

ctggcaacgg gaggcgtcgg gcaggtctac ctctacagca ccaacccaag gatcgcgacc 660

ggtgatggat tggccatggc ttaccgggct ggccgcaaa tccgcaacct cgagtgtatc 720

caatttcate ctacagcgct atacaccacc accaatgacc gctttctgat cagcgaagcc 780

gtccggggtg aaggggccat cctccgcaat caggaggag aggccttcat ggctcgtac 840

gatgaccgca aggacctgc ccccgggat attgtggcca gagcaattga cagtgaatg 900

aagcagtccg gtcacatcca tgtctggctc gacatcactc atcgggatga aaccgatctg 960

cgggagcgtt tcccaacat ttctgaggcc tgcctgaagg tcggagtcaa catggcgcaa 1020

tcctccatcc cgttgggttc ggcgatgcac tacctctgcg gaggcgtagc caccgacctc 1080

aatgcggcca ccgacatcac tggactgttt gcctgtgggg aagttgcctg cacgggattg 1140

catggtgcca accgtctgc cagcaacagc ctgctggagg cagtggatcat ggcgcaccgg 1200

gcctccgtcg cagtggatgc atacctcaac agcaaaccct accgtatgc acaattgccg 1260

gaatggacgg atggcaacgt gcaggacagc gacgagcgtg tcgtgatcag ccacaactgg 1320

gatgaactca aacgcacgat gtgggactac gtgggcatcg tccgcaccac caageggctt 1380

cagcgcgcgc aacgacgcat tcgtcacctc cagcaggaaa tcgaagagta ttactggaat 1440

ttcaaggttg agtcctccct tctggagtta cggaatctgg ttgtgggtggc ggatctgggtt 1500

atccactgtg ccctccaacg ccatgagagc cgtggcctgc attgcaccgc ggattatccc 1560

ggcaagttgc ccaccccgat caataccgcc gttcgcagaa gaaccgtta a 1611

<210> 120
 <211> 536
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> DOMAIN
 <222> (6)...(260)
 <223> FAD dependent oxidoreductase

<220>
 <221> DOMAIN
 <222> (6)...(380)
 <223> Pyridine nucleotide-disulphide oxidoreductase

<220>
 <221> DOMAIN
 <222> (6)...(391)
 <223> FAD binding domain

<220>
 <221> DOMAIN
 <222> (440)...(534)
 <223> Fumarate reductase/succinate dehydrogenase flavoprotein C-terminal domain

<400> 120
 Met Gln Lys Arg Tyr Asp Val Ile Val Val Gly Ser Gly Ile Ala Gly
 1 5 10 15

Leu Ser Phe Ala Leu Lys Val Ala Lys Ala Gly His Arg Val Gly Ile
20 25 30

Leu Thr Lys Lys Asp Arg Ala Glu Ser Asn Thr Asn Tyr Ala Gln Gly
35 40 45

Gly Ile Ala Ala Val Thr Ser Gln Thr Asp Asp Phe Glu Leu His Val
50 55 60

Gln Asp Thr Leu Thr Ala Gly Asp Gly Leu Cys Asp Glu Ala Val Val
65 70 75 80

Arg Thr Ile Ile Gly Glu Ala Pro Ala Arg Ile Gln Glu Leu Ile Asp
85 90 95

Leu Gly Val Ala Phe Ser His Leu Glu Asp Gly Arg Val Ser Leu His
100 105 110

Arg Glu Gly Gly His Ser Lys Arg Arg Ile Leu His Val Gln Asp Val
115 120 125

Thr Gly Lys Ala Ile Glu Glu Ala Leu Leu His Ala Ile Glu Gln Ser
130 135 140

Pro Leu Ile Asp Leu Asn Glu His Val Phe Ala Ile Asp Leu Leu Thr
145 150 155 160

Glu Arg Lys Leu Ala Leu Ala Gly Phe Glu Val Glu Gly Ala Lys Asn
165 170 175

Arg Val Val Gly Leu Tyr Ala Leu Asp Glu Ala Thr Gln Glu Val His
180 185 190

Val Phe Glu Ala Pro Val Val Met Leu Ala Thr Gly Gly Val Gly Gln
195 200 205

Val Tyr Leu Tyr Ser Thr Asn Pro Arg Ile Ala Thr Gly Asp Gly Leu
210 215 220

Ala Met Ala Tyr Arg Ala Gly Ala Glu Ile Arg Asn Leu Glu Cys Ile
225 230 235 240

Gln Phe His Pro Thr Ala Leu Tyr Thr Thr Thr Asn Asp Arg Phe Leu
245 250 255

Ile Ser Glu Ala Val Arg Gly Glu Gly Ala Ile Leu Arg Asn Gln Glu
260 265 270

Gly Glu Ala Phe Met Ala Arg Tyr Asp Asp Arg Lys Asp Leu Ala Pro
275 280 285

Arg Asp Ile Val Ala Arg Ala Ile Asp Ser Glu Met Lys Gln Ser Gly
290 295 300

Ser Ser His Val Trp Leu Asp Ile Thr His Arg Asp Glu Thr Asp Leu
305 310 315 320

Arg Glu Arg Phe Pro Asn Ile Phe Glu Ala Cys Leu Lys Val Gly Val
325 330 335

Asn Met Ala Gln Ser Ser Ile Pro Val Val Pro Ala Met His Tyr Leu
340 345 350

Cys Gly Gly Val Ala Thr Asp Leu Asn Ala Ala Thr Asp Ile Thr Gly
355 360 365

Leu Phe Ala Cys Gly Glu Val Ala Cys Thr Gly Leu His Gly Ala Asn
370 375 380

Arg Leu Ala Ser Asn Ser Leu Leu Glu Ala Val Val Met Ala His Arg
385 390 395 400

Ala Ser Val Ala Val Asp Ala Tyr Leu Asn Ser Lys Pro His Arg Tyr

405

410

415

Ala Gln Leu Pro Glu Trp Thr Asp Gly Asn Val Gln Asp Ser Asp Glu
420 425 430

Arg Val Val Ile Ser His Asn Trp Asp Glu Leu Lys Arg Thr Met Trp
435 440 445

Asp Tyr Val Gly Ile Val Arg Thr Thr Lys Arg Leu Gln Arg Ala Gln
450 455 460

Arg Arg Ile Arg His Leu Gln Gln Glu Ile Glu Glu Tyr Tyr Trp Asn
465 470 475 480

Phe Lys Val Glu Ser Ser Leu Leu Glu Leu Arg Asn Leu Val Val Val
485 490 495

Ala Asp Leu Val Ile His Cys Ala Leu Gln Arg His Glu Ser Arg Gly
500 505 510

Leu His Cys Thr Arg Asp Tyr Pro Gly Lys Leu Pro Thr Pro Ile Asn
515 520 525

Thr Ala Val Arg Arg Arg Thr Gly
530 535

<210> 121

<211> 990

<212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 121

atgccttttg atgccattgg agaaagcttc cgtgccagcc agcaactccc getgatcaag 60

gtcgacggca accgtttcgt gattcgggag accggtgagc cgatcgtctt ccggggcgctc 120

tccgcctccg acccggtcgc gctactggaa cgcggtcaat ggggtcgccg ttactttgaa 180

gagatggcca agtggaatgc caacgttgig cgcattcctg ttcacccggc agactggcgt 240

aatctcggcg aagacatcta tctcgcccta ctcgaccagg cgattgaatg gtcggctgaa 300

ctcgcatgc acgtcatcat cgactggcac actatcggca atattctgac cggatattat 360

caccgcgaca tttatgaaac caccgtgat gagacttacc gtttttgta caccatcgcc 420

attcgttatc agggtaaccc gacagtggcc ttttatgaac tctacaatga gcccaccaac 480

cgaggcggtc gcatgggccc ctttccttgg gaagaatatg ccagttcat cgaagggtg 540

atttccatgc tctacgcat cgacgacacc gtattccac tggtcgtgg ctctgactgg 600

ggatatgatt tgagctatgt tgcggaacgc ccgatccgtt ttccaggagt cgcctatgtc 660

accacacctt accgcagaa gcgccccgag ccttggaac cgatctggca ggaggaatgg 720

ggttttgtcg ccgacaccta tcccatgac gccactgagt ttggcttcat gagtgaggac 780

ggtcccgag cccacaaccc ggttatcggg gatgaacact atggcgaatc ggtcatccgc 840

tttttcgagg aacgcggcat ttctggacg gcctgggtgt ttgatcctct ctggtcaccc 900

cagcttttcg aagactggga aacctatacc ccacccggc aaggccgatt ctttaaacag 960

aaaatgatgg aactgaatcc cccgcgttga 990

<210> 122
 <211> 329
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>

<221> DOMAIN

<222> (25)...(302)

<223> Cellulase (glycosyl hydrolase family 5)

<400> 122

Met Pro Phe Asp Ala Ile Gly Glu Ser Phe Arg Ala Ser Gln Gln Leu
1 5 10 15

Pro Leu Ile Lys Val Asp Gly Asn Arg Phe Val Ile Ala Glu Thr Gly
20 25 30

Glu Pro Ile Val Phe Arg Gly Val Ser Ala Ser Asp Pro Ala Ala Leu
35 40 45

Leu Glu Arg Gly Gln Trp Gly Arg Arg Tyr Phe Glu Glu Met Ala Lys
50 55 60

Trp Asn Ala Asn Val Val Arg Ile Pro Val His Pro Ala Asp Trp Arg
65 70 75 80

Asn Leu Gly Glu Asp Ile Tyr Leu Ala Leu Leu Asp Gln Ala Ile Glu
85 90 95

Trp Ser Ala Glu Leu Gly Met His Val Ile Ile Asp Trp His Thr Ile
100 105 110

Gly Asn Ile Leu Thr Gly Ile Tyr His Arg Asp Ile Tyr Glu Thr Thr
115 120 125

Arg Asp Glu Thr Tyr Arg Phe Trp Tyr Thr Ile Ala Ile Arg Tyr Gln
130 135 140

Gly Asn Pro Thr Val Ala Phe Tyr Glu Leu Tyr Asn Glu Pro Thr Asn
145 150 155 160

Arg Gly Gly Arg Met Gly Pro Leu Pro Trp Glu Glu Tyr Ala Gln Phe

165

170

175

Ile Glu Gly Leu Ile Ser Met Leu Tyr Ala Ile Asp Asp Thr Val Ile
180 185 190

Pro Leu Val Ala Gly Phe Asp Trp Gly Tyr Asp Leu Ser Tyr Val Ala
195 200 205

Glu Arg Pro Ile Arg Phe Pro Gly Val Ala Tyr Val Thr His Pro Tyr
210 215 220

Pro Gln Lys Arg Pro Glu Pro Trp Glu Pro Ile Trp Gln Glu Glu Trp
225 230 235 240

Gly Phe Val Ala Asp Thr Tyr Pro Met Ile Ala Thr Glu Phe Gly Phe
245 250 255

Met Ser Glu Asp Gly Pro Gly Ala His Asn Pro Val Ile Gly Asp Glu
260 265 270

His Tyr Gly Glu Ser Val Ile Arg Phe Phe Glu Glu Arg Gly Ile Ser
275 280 285

Trp Thr Ala Trp Val Phe Asp Pro Leu Trp Ser Pro Gln Leu Phe Glu
290 295 300

Asp Trp Glu Thr Tyr Thr Pro Thr Arg Gln Gly Arg Phe Phe Lys Gln
305 310 315 320

Lys Met Met Glu Leu Asn Pro Pro Arg
325

<210> 123

<211> 1398

<212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 123

atgccgatga gcacagaaac gacttttcct tctgatttca cctggggcgc agcaacagcc	60
gcctaccaga tcgaaggggg cgatcgcgct ggcgggcgcg gccgttccgt gtgggacatg	120
ttttgcgaga aacgaggagc tatttgggag ggccatacgg ggcagcgagc gagtctgcat	180
cttcagcgct ggctgagga cgtaatgttg atgcaacagc tcggactgcg gggctatcgt	240
tttagcgta gctggccgcg cgtcttcccg acaggagtcg gcaaagtcaa ccgtgaaggg	300
ttggcctttt acgatcagct cgtagacgcc ttgctcgagg ccggcatcac cccctttata	360
acgtatattc attgggactt cccgctcgat ttgtaccacc gaggcggctg gttgaatcgc	420
gacagcgccg actggtttgc ctctacgcc gagtgcctcg gcaaggcact gggcgacagg	480
gtcaagcact gggtagacct caacagcccg caggttttca taggcctcgg tcattacgaa	540
gggcgtcatg ccccggggtt gaagctctcc atcgcggaat tgctgcgctg cgggcaccac	600
gccttgctcg cgacgggaa ggccgtgcaa gccctgcgcg cttccgtcga cggccctgc	660
aagattggat ttgctccgtt ggggattccc aagcttccgg cgagtgagag ctgagaggat	720
atcgccgcgg cagaaaaggc ccagttcgcg gcgggagcgc cgccgtattg gacgctgagc	780
tggtgggcgg atccggtgtt tcaggggaca tatcccgtg atgcctgcca ggctctcgga	840
gcggacgcgc cgcaggtggc cgatcacgac atgagcatca tcagcgagcc gactgatttc	900
ctgggcctca acctttatca aggggtggtg gtgcgtgccg atcacacggg tcaaccagaa	960
acggtgccgt ttccgccggg attccccgtg actgcgtca actgggccgt aaccccagag	1020
gcgctgtatt ggggcccgcg ctttgccttc gaacgtaca aaaagccgat tcacatcacg	1080

gaaaacgggc tatcctgtcg tgactggccg tcgctcgacg ggcacgtcca cgacgccgac 1140

cgcatcgact tcattggccc gcacttgcgc gcagcgcatc gagccattcg cgatgggata 1200

ccgatcgaag gctacttcca ctggctgtcg atcgacaact tcgagtgggc agaaggctac 1260

aaggaacgct tcgggctcat ttacgtcgac tatacgagcg gcgagcgcat tccgaaggac 1320

tcgtaccact ggtaccagaa ggtcattgcc tccgaggggc gggcagcgct cggcgcgccc 1380

agtgtgtctc gcccataa 1398

<210> 124
 <211> 465
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> DOMAIN
 <222> (5)...(454)
 <223> Glycosyl hydrolase family 1

<220>
 <221> SITE
 <222> (13)...(27)
 <223> Glycosyl hydrolases family 1 N-terminal signature. Prosite id = PS00653

<400> 124
 Met Pro Met Ser Thr Glu Thr Thr Phe Pro Ser Asp Phe Thr Trp Gly
 1 5 10 15

Ala Ala Thr Ala Ala Tyr Gln Ile Glu Gly Gly Asp Arg Ala Gly Gly
 20 25 30

Arg Gly Arg Ser Val Trp Asp Met Phe Cys Glu Lys Arg Gly Ala Ile
 35 40 45

Trp Glu Gly His Thr Gly Gln Arg Ala Ser Leu His Leu Gln Arg Trp

50

55

60

Arg Glu Asp Val Met Leu Met Gln Gln Leu Gly Leu Arg Gly Tyr Arg
65 70 75 80

Phe Ser Val Ser Trp Pro Arg Val Phe Pro Thr Gly Val Gly Lys Val
85 90 95

Asn Arg Glu Gly Leu Ala Phe Tyr Asp Gln Leu Val Asp Ala Leu Leu
100 105 110

Glu Ala Gly Ile Thr Pro Phe Ile Thr Leu Phe His Trp Asp Phe Pro
115 120 125

Leu Asp Leu Tyr His Arg Gly Gly Trp Leu Asn Arg Asp Ser Ala Asp
130 135 140

Trp Phe Ala Ser Tyr Ala Glu Cys Leu Gly Lys Ala Leu Gly Asp Arg
145 150 155 160

Val Lys His Trp Val Thr Leu Asn Glu Pro Gln Val Phe Ile Gly Leu
165 170 175

Gly His Tyr Glu Gly Arg His Ala Pro Gly Leu Lys Leu Ser Ile Ala
180 185 190

Glu Met Leu Arg Cys Gly His His Ala Leu Leu Ala His Gly Lys Ala
195 200 205

Val Gln Ala Leu Arg Ala Ser Val Asp Gly Pro Cys Lys Ile Gly Phe
210 215 220

Ala Pro Val Gly Ile Pro Lys Leu Pro Ala Ser Glu Ser Ser Glu Asp
225 230 235 240

Ile Ala Ala Ala Arg Lys Ala Gln Phe Ala Ala Gly Ala Pro Pro Tyr

245

250

255

Trp Thr Leu Ser Trp Trp Ala Asp Pro Val Phe Gln Gly Thr Tyr Pro
260 265 270

Ala Asp Ala Cys Gln Ala Leu Gly Ala Asp Ala Pro Gln Val Ala Asp
275 280 285

His Asp Met Ser Ile Ile Ser Glu Pro Thr Asp Phe Leu Gly Leu Asn
290 295 300

Leu Tyr Gln Gly Val Val Val Arg Ala Asp His Thr Gly Gln Pro Glu
305 310 315 320

Thr Val Pro Phe Pro Pro Gly Phe Pro Val Thr Ala Leu Asn Trp Ala
325 330 335

Val Thr Pro Glu Ala Leu Tyr Trp Gly Pro Arg Phe Ala Phe Glu Arg
340 345 350

Tyr Lys Lys Pro Ile His Ile Thr Glu Asn Gly Leu Ser Cys Arg Asp
355 360 365

Trp Pro Ser Leu Asp Gly His Val His Asp Ala Asp Arg Ile Asp Phe
370 375 380

Met Ala Arg His Leu Arg Ala Ala His Arg Ala Ile Arg Asp Gly Ile
385 390 395 400

Pro Ile Glu Gly Tyr Phe His Trp Ser Ala Ile Asp Asn Phe Glu Trp
405 410 415

Ala Glu Gly Tyr Lys Glu Arg Phe Gly Leu Ile Tyr Val Asp Tyr Thr
420 425 430

Ser Gly Glu Arg Ile Pro Lys Asp Ser Tyr His Trp Tyr Gln Lys Val

435

440

445

Ile Ala Ser Glu Gly Arg Ala Ala Leu Gly Ala Pro Ser Ala Ala Arg
450 455 460

Pro
465

<210> 125

<211> 1350

<212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 125

atgtcagatg ccgccccgac tgatccgaaa tccgcaatgc ccagacgctc ggacttcccc 60

gagggttttg tcttcggcgc ggccaccgcg gcctatcaga tcgagggcca tgccttcggc 120

ggcgcggggc cctgccattg ggacagcttc gccgcaaccg ggcgtaacgt ggtcggcaat 180

gaggatggcg cgcgcgcctg cgagcattac acccgctggc cgcaggatct ggacctgac 240

cgcgaggccg ggctcgacgc ctaccgcttc tcgacctcct gggcgcgggt gatgcccgat 300

ggcgtgacct tgaacccga ggggctggat ttctacgacc gcctcgtcga tggcatgctc 360

gagcgcgggc taaagcccta tctcaccttc taccattggg aattgccctc ggcgcttgcc 420

gacaggggcg gctggaccaa tcgacacacg gccgagcgct ttgccgattt cgcagcggtg 480

gtgatggagc ggttgggcag ccgcgtcgcc cgcacggcca ccatcaacga gccatggtgc 540

gtgagctggc tctcgattt cgaaggccat cagcgcccg gcctgcgcga catccgtgcc 600

accgcacgcg ccatgcatca tgtgcaactg gcgcacggcc tcgcgctcgg gaagctgcgc 660

gcgcaggggc atggcaatct cggcatcgtg ctgaatttct cggaaatcat tcccgcggg 720

cgagagcacg cgaaggcggc tgatctcggc gacgcaatct cgaaccgctg gttcatcgag 780

tcagtgcgc gtggcaccta tcccgatgtg gtctcggagg gtcctgggcaa gcacatgcc 840

gagggctggc aggatgacat gaaaaccatc gcggcccgcc tcgactggct ggggtgtgaac 900

tactacacc gcggcatcgt cgcgcatgac ccggacgcgt cctggccctc gacccgagcg 960

gaggaggggc cctgcccga gacgcagatg ggctgggaga tctacccga gggcttgcgc 1020

aacctgctgg tgcgcatggc gcgcgactat gtggcgacc ttcccatggt cgtgaccgaa 1080

aacgggatgg cctgggccga cgaggtcgcg gatggcgccg tcagagatac gatccgcacc 1140

gaatatgtcg cagcccatct caacgcgacc cgcgaggcgc tggccggcgg ggcgaatatc 1200

gaaggtttct tctattggtc gctgctcgac aattacgaat gggccttcgg ctatgccaag 1260

cgcttcggcc tcgtccatgt cgatttcgac acgatggcac gcacgccgaa agcctcctac 1320

cacgcgctga gggccgcgct gcagggttga 1350

<210> 126
 <211> 449
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> DOMAIN
 <222> (15)...(443)
 <223> Glycosyl hydrolase family 1

<220>
 <221> SITE
 <222> (235)...(238)
 <223> N-glycosylation site. Prosite id = PS00001

<220>
 <221> SITE
 <222> (361)...(369)
 <223> Glycosyl hydrolases family 1 active site. Prosite id = PS00572

<220>
 <221> SITE
 <222> (393)...(396)
 <223> N-glycosylation site. Prosite id = PS00001

<400> 126
 Met Ser Asp Ala Ala Pro Thr Asp Pro Lys Ser Ala Met Pro Arg Arg
 1 5 10 15

Ser Asp Phe Pro Glu Gly Phe Val Phe Gly Ala Ala Thr Ala Ala Tyr
 20 25 30

Gln Ile Glu Gly His Ala Phe Gly Gly Ala Gly Pro Cys His Trp Asp
 35 40 45

Ser Phe Ala Ala Thr Gly Arg Asn Val Val Gly Asn Glu Asp Gly Ala
 50 55 60

Arg Ala Cys Glu His Tyr Thr Arg Trp Pro Gln Asp Leu Asp Leu Ile
 65 70 75 80

Arg Glu Ala Gly Leu Asp Ala Tyr Arg Phe Ser Thr Ser Trp Ala Arg
 85 90 95

Val Met Pro Asp Gly Val Thr Leu Asn Pro Glu Gly Leu Asp Phe Tyr
 100 105 110

Asp Arg Leu Val Asp Gly Met Leu Glu Arg Gly Leu Lys Pro Tyr Leu
 115 120 125

Thr Leu Tyr His Trp Glu Leu Pro Ser Ala Leu Ala Asp Arg Gly Gly
 130 135 140

Trp Thr Asn Arg Asp Thr Ala Glu Arg Phe Ala Asp Phe Ala Ala Val
 145 150 155 160

Val Met Glu Arg Leu Gly Ser Arg Val Ala Arg Thr Ala Thr Ile Asn
 165 170 175

Glu Pro Trp Cys Val Ser Trp Leu Ser His Phe Glu Gly His His Ala
 180 185 190

Pro Gly Leu Arg Asp Ile Arg Ala Thr Ala Arg Ala Met His His Val
 195 200 205

Gln Leu Ala His Gly Leu Ala Leu Gly Lys Leu Arg Ala Gln Gly His
 210 215 220

Gly Asn Leu Gly Ile Val Leu Asn Phe Ser Glu Ile Ile Pro Ala Gly
 225 230 235 240

Arg Glu His Ala Lys Ala Ala Asp Leu Gly Asp Ala Ile Ser Asn Arg
 245 250 255

Trp Phe Ile Glu Ser Val Ala Arg Gly Thr Tyr Pro Asp Val Val Leu
 260 265 270

Glu Gly Leu Gly Lys His Met Pro Glu Gly Trp Gln Asp Asp Met Lys
 275 280 285

Thr Ile Ala Ala Pro Leu Asp Trp Leu Gly Val Asn Tyr Tyr Thr Arg
 290 295 300

Gly Ile Val Ala His Asp Pro Asp Ala Ser Trp Pro Ser Thr Arg Ala
 305 310 315 320

Glu Glu Gly Pro Leu Pro Lys Thr Gln Met Gly Trp Glu Ile Tyr Pro
 325 330 335

Glu Gly Leu Arg Asn Leu Leu Val Arg Met Ala Arg Asp Tyr Val Gly

340 345 350

Asp Leu Pro Met Val Val Thr Glu Asn Gly Met Ala Trp Ala Asp Glu
355 360 365

Val Ala Asp Gly Ala Val Arg Asp Thr Ile Arg Thr Glu Tyr Val Ala
370 375 380

Ala His Leu Asn Ala Thr Arg Glu Ala Leu Ala Gly Gly Ala Asn Ile
385 390 395 400

Glu Gly Phe Phe Tyr Trp Ser Leu Leu Asp Asn Tyr Glu Trp Ala Phe
405 410 415

Gly Tyr Ala Lys Arg Phe Gly Leu Val His Val Asp Phe Asp Thr Met
420 425 430

Ala Arg Thr Pro Lys Ala Ser Tyr His Ala Leu Arg Ala Ala Leu Gln
435 440 445

Gly

<210> 127

<211> 774

<212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 127

atggacttgc agctaggcgg aaagcgctg ctgatcacgg gtgcgtccaa aggcacgcgc 60

ctggcctgcg ccgtcgcctt tgcgcgcgag ggtgccgacc cgattctggt ggcgcgcgat 120

gatgcggcgt tgcacacgc cagtcgcc atccgcgaac aaagcgccg cgcggcacat 180

gccatcacgc tggacctggc cctgcctggc gcggcggaag agctggccaa ggaaaccggc 240

cccatcgaca tactgggtcaa caacgcgggc gcggtgcccgc gcggcgcgct ggaccaggtg 300

caagacgaac gctggcgcgc gggctgggaa ttgaaagtgc acggctacat cagcctggcg 360

cgctgctact acccgacat gcgcgaagcg gcgcggggcg tcatcgccaa catcatcggc 420

atggcgggcg cggcgccccg cgccgactac atctgcggcg cggcggccaa tgcctcactg 480

attgccttta ccgcgcgct ggggtggcgaa gcgccccgcc acggcgtgcg cgtctttggc 540

gtcaaccctt cgcgcacgcg gaccgaccgc gtgctgacct tggcccggca acgcgcgcag 600

gcgcgctggg gcgacgaac gcgttggcag gaaacgctgt cggacctgcc cttcaaccgg 660

ctgatggaac ccgccgaagt ggccgacatg attgtgttcg gcgcctcgcc acgcgcgggt 720

tacctgagcg gcacggtcat cgacctggac ggcgcggaac agtacgcgaa atag 774

<210> 128
<211> 257
<212> PRT
<213> Unknown

<220>
<223> Obtained from environmental sample

<220>
<221> DOMAIN
<222> (8)...(172)
<223> short chain dehydrogenase

<220>
<221> SITE
<222> (159)...(162)
<223> N-glycosylation site. Prosite id = PS00001

<400> 128
Met Asp Leu Gln Leu Gly Gly Lys Arg Val Leu Ile Thr Gly Ala Ser
1 5 10 15

Lys Gly Ile Gly Leu Ala Cys Ala Val Ala Phe Ala Arg Glu Gly Ala
20 25 30

Asp Pro Ile Leu Val Ala Arg Asp Asp Ala Ala Leu His His Ala Thr
35 40 45

Ser Ala Ile Arg Glu Gln Ser Gly Arg Ala Ala His Ala Ile Thr Leu
50 55 60

Asp Leu Ala Leu Pro Gly Ala Ala Glu Lys Leu Ala Lys Glu Thr Gly
65 70 75 80

Pro Ile Asp Ile Leu Val Asn Asn Ala Gly Ala Val Pro Gly Gly Ala
85 90 95

Leu Asp Gln Val Gln Asp Glu Arg Trp Arg Ala Gly Trp Glu Leu Lys
100 105 110

Val His Gly Tyr Ile Ser Leu Ala Arg Cys Tyr Tyr Pro His Met Arg
115 120 125

Glu Ala Gly Ala Gly Val Ile Ala Asn Ile Ile Gly Met Ala Gly Ala
130 135 140

Ala Pro Arg Ala Asp Tyr Ile Cys Gly Ala Ala Ala Asn Ala Ser Leu
145 150 155 160

Ile Ala Phe Thr Arg Ala Leu Gly Gly Glu Ala Pro Arg His Gly Val
165 170 175

Arg Val Phe Gly Val Asn Pro Ser Arg Thr Arg Thr Asp Arg Val Leu
180 185 190

Thr Leu Ala Arg Gln Arg Ala Gln Ala Arg Trp Gly Asp Glu Thr Arg
195 200 205

Trp Gln Glu Thr Leu Ser Asp Leu Pro Phe Asn Arg Leu Met Glu Pro
210 215 220

Ala Glu Val Ala Asp Met Ile Val Phe Gly Ala Ser Pro Arg Ala Gly
225 230 235 240

Tyr Leu Ser Gly Thr Val Ile Asp Leu Asp Gly Gly Glu Gln Tyr Ala
245 250 255

Lys

<210> 129
<211> 747
<212> DNA
<213> Unknown

<220>
<223> Obtained from environmental sample

<400> 129
atgccccaaag tcatgctcgt taccggcggc agccgtggca tcggcgccgc cgtcgccaag 60

ctggcgccgc gccgcggcta cgcggtcggc atcaactacc gcacccattc cgacgcgcgc 120

gacgccgtcg tggccgaaat ccagcaggcg ggccggcaccg cgctggccat ccaggccgac 180

gtgtcgcagg aagacgatgt gctgcacatg ttccgcacgc tggacgagcg cctgggccgc 240

atcgacgcgc tggtaataa gccggcatc ctggaaacgc agatgcgcct ggaccagatg 300

gaagccgacc gcctgctcgc cgtgtgttcc accaacgtca tcggcgcttt cctatgtgcg 360

cgcaagccg tgcgcaggat gtcgacgcgc catggcggcg tgggcggcgc catcgtcaac 420

gtgtcttcgg cggcggcgcg cctgggctcg ccaatgaat acgtggatta cgcggcctcc 480

aaggcgcgcg tggacacgat gaccatcggc ctgtcgaaag aggtggcgcc cgaaggtatc 540
 cgcgatgaatg gcgtgcgccc cggcaccatc tacaccgaca tgcacgcaag cggcggcgag 600
 ccgggcccggg tggatcgctt gaaaagcgtg atcccgtgc ggcgcgcgcg ctcggtggaa 660
 gaagtggcgg gcgccgtcat gtggctgttt tccgaagaag ccggctatac cagcggttcg 720
 ttcacgcacg tgtccggcgg tagttga 747

<210> 130
 <211> 248
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> DOMAIN
 <222> (3)...(176)
 <223> short chain dehydrogenase

<220>
 <221> SITE
 <222> (142)...(145)
 <223> N-glycosylation site. Prosite id = PS00001

<220>
 <221> SITE
 <222> (146)...(174)
 <223> Short-chain dehydrogenases/reductases family signature. Prosite id = PS00061

<400> 130
 Met Pro Lys Val Met Leu Val Thr Gly Gly Ser Arg Gly Ile Gly Ala
 1 5 10 15

Ala Val Ala Lys Leu Ala Ala Arg Arg Gly Tyr Ala Val Gly Ile Asn
 20 25 30

Tyr Arg Thr His Ser Asp Ala Ala Asp Ala Val Val Ala Glu Ile Gln
 35 40 45

Gln Ala Gly Gly Thr Ala Leu Ala Ile Gln Ala Asp Val Ser Gln Glu
50 55 60

Asp Asp Val Leu His Met Phe Arg Thr Leu Asp Glu Arg Leu Gly Arg
65 70 75 80

Ile Asp Ala Leu Val Asn Asn Ala Gly Ile Leu Glu Thr Gln Met Arg
85 90 95

Leu Asp Gln Met Glu Ala Asp Arg Leu Leu Arg Val Leu Ser Thr Asn
100 105 110

Val Ile Gly Ala Phe Leu Cys Ala Arg Glu Ala Val Arg Arg Met Ser
115 120 125

Thr Arg His Gly Gly Val Gly Gly Ala Ile Val Asn Val Ser Ser Ala
130 135 140

Ala Ala Arg Leu Gly Ser Pro Asn Glu Tyr Val Asp Tyr Ala Ala Ser
145 150 155 160

Lys Gly Ala Leu Asp Thr Met Thr Ile Gly Leu Ser Lys Glu Val Ala
165 170 175

Pro Glu Gly Ile Arg Val Asn Gly Val Arg Pro Gly Thr Ile Tyr Thr
180 185 190

Asp Met His Ala Ser Gly Gly Glu Pro Gly Arg Val Asp Arg Leu Lys
195 200 205

Ser Val Ile Pro Leu Arg Arg Gly Gly Ser Val Glu Glu Val Ala Gly
210 215 220

Ala Val Met Trp Leu Phe Ser Glu Glu Ala Gly Tyr Thr Ser Gly Ser

225	230	235	240
Phe Ile Asp Val Ser Gly Gly Ser			
245			
<210> 131			
<211> 1041			
<212> DNA			
<213> Unknown			
<220>			
<223> Obtained from environmental sample			
<400> 131			
gtggaaacct attttccct gcaccgcggg atcaacatga gccactggct ttcgcaagt	60		
aatgaaaaca ttcccgaccg ttccacctat gtgacggagc gggacctgca atttttgcgg	120		
gcagcgggct tcgaccatgt gcgtctgccg atcgatgaga tcgaactctg ggatgaggag	180		
ggccatcaga tcgaggagcg ctggcaatac atgcacaact ttatgcgctg gagccgaaag	240		
aatgacctcc gggttattct cgacctgcac acggtattgt cccaccactt caacgcgatc	300		
aacatgggag aggtcaacac cctctttaat gatcccaagg aacaggaaaa attcctcaat	360		
ctctgggagc aaatcatgga tgccgtaggg caccaccca acgagtttct cgcttatgaa	420		
atgctcaatg aggcggtcgc ggaagatgat gaagactgga acctgctcct caaccgtgcg	480		
attgaacgca tccgggaacg tgagccgcat cgcgttctga ttgccggggc caactggtgg	540		
cagcatgccg cccgcgttcc caacctgagg ctccccctg gtgatccaa catcatcatc	600		
agttttcact ttactcacc ctttctcttc acgcactatc gcagcagctg gactgccatg	660		
cgggcatacc agggtttctg ccaatacccc ggcatcatta ttcccgccat ccatctcgaa	720		
ggaatgaact atccggagtc ctttgtccaa atgtgggaag agcacaatca gtattacgac	780		

atccattcaa tgtatgccga aatggtcccg gcggtgcgtt ttgccgaaaa gctgggcctt 840
 cggtctctatt gcggcgaatt tggagccatg aagaccgttg atcgtgccca aatgctgcag 900
 tggatcggg atgtggtcag agtctttgaa atgttggaca ttccctacac tgcctgggat 960
 tatcaggga cctttggaat ccgcgatgag ctgaccgtg agcctgatca tgaactgac 1020
 gacattctcc tcggccgcta a 1041

<210> 132
 <211> 346
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> DOMAIN
 <222> (14)...(325)
 <223> Cellulase (glycosyl hydrolase family 5)

<220>
 <221> SITE
 <222> (12)...(15)
 <223> N-glycosylation site. Prosite id = PS00001

<400> 132
 Met Glu Thr Tyr Phe Pro Leu His Arg Gly Ile Asn Met Ser His Trp
 1 5 10 15

Leu Ser Gln Val Asn Glu Asn Ile Pro Asp Arg Ser Thr Tyr Val Thr
 20 25 30

Glu Arg Asp Leu Gln Phe Leu Arg Ala Ala Gly Phe Asp His Val Arg
 35 40 45

Leu Pro Ile Asp Glu Ile Glu Leu Trp Asp Glu Glu Gly His Gln Ile
 50 55 60

Glu Glu Ala Trp Gln Tyr Met His Asn Phe Met Arg Trp Ser Arg Lys
65 70 75 80

Asn Asp Leu Arg Val Ile Leu Asp Leu His Thr Val Leu Ser His His
85 90 95

Phe Asn Ala Ile Asn Met Gly Glu Val Asn Thr Leu Phe Asn Asp Pro
100 105 110

Lys Glu Gln Glu Lys Phe Leu Asn Leu Trp Glu Gln Ile Met Asp Ala
115 120 125

Val Gly His His Pro Asn Glu Phe Leu Ala Tyr Glu Met Leu Asn Glu
130 135 140

Ala Val Ala Glu Asp Asp Glu Asp Trp Asn Leu Leu Leu Asn Arg Ala
145 150 155 160

Ile Glu Arg Ile Arg Glu Arg Glu Pro His Arg Val Leu Ile Ala Gly
165 170 175

Ala Asn Trp Trp Gln His Ala Ala Arg Val Pro Asn Leu Arg Leu Pro
180 185 190

Pro Gly Asp Pro Asn Ile Ile Ile Ser Phe His Phe Tyr Ser Pro Phe
195 200 205

Leu Phe Thr His Tyr Arg Ser Ser Trp Thr Ala Met Arg Ala Tyr Gln
210 215 220

Gly Phe Val Gln Tyr Pro Gly Ile Thr Ile Pro Ala Ile His Leu Glu
225 230 235 240

Gly Met Asn Tyr Pro Glu Ser Phe Val Gln Met Trp Glu Glu His Asn
245 250 255

Gln Tyr Tyr Asp Ile His Ser Met Tyr Ala Glu Met Val Pro Ala Val
260 265 270

Arg Phe Ala Glu Lys Leu Gly Leu Arg Leu Tyr Cys Gly Glu Phe Gly
275 280 285

Ala Met Lys Thr Val Asp Arg Ala Gln Met Leu Gln Trp Tyr Arg Asp
290 295 300

Val Val Arg Val Phe Glu Met Leu Asp Ile Pro Tyr Thr Ala Trp Asp
305 310 315 320

Tyr Gln Gly Thr Phe Gly Ile Arg Asp Glu Leu Thr Gly Glu Pro Asp
325 330 335

His Glu Leu Ile Asp Ile Leu Leu Gly Arg
340 345

<210> 133
<211> 1377
<212> DNA
<213> Unknown

<220>
<223> Obtained from environmental sample

<400> 133
atgacacaac tggcttttcc atctaacttc atctggggaa cagctacttc cgcttaccaa 60

atcgaaggcg cctggaacgc agacggcaag ggccaatcta tttgggatcg cttttccat 120

acgcagggga agatcattga cggcagcaac ggcgatgtgg cctgcgatca ctaccaccgc 180

tggcgcgagg acgtggccct catgagagac ttgggtatgc aggcatatcg ctttccatc 240

tcctggccac gcatcctgcc caccggtcat ggacagatca atcaggctgg gctggacttt 300

tacaatcgcc tggtaggacg gttgctggaa gctggcatca agccctttgc caccctctac 360

cactgggacc tgccgtggc gctacaggct gacggcggct ggccggagcg ctccacggcc 420

aaggcctttg tcgaatacgc cgacgtggc agccgcgcgc tggcgatcg ggtgaagagc 480

tggatcacc ataacgaacc gtggtgcatc agcatgctga gccatcaa tggggagcat 540

gcgcccggct ggccggactg gcaggtgctg ttggcgccg cgcaccacgt cctcctttcg 600

catggttgg ccgtgccga actgcgtgc aacagccgc atgcagaaat cggcatcacg 660

ttgaacttta ccccgccga gccagcttcg aacagcgcag ccgatttcaa ggcctatcgc 720

cagttgatg gctacttcaa ccgtggttc ctggaccgc tctatggccg ccactatccg 780

gcagatatgg tgcacgatta catcgcgaa ggtacctgc catcacaggg ttgactttc 840

gtggaagctg gtgacctga cgcgatcgc acgcgcaccg atttctggg tgtgaactat 900

tacacgcgc aagtggtcg tagccaggaa atcccagaga gtgagaacgc gccgcgcaca 960

gtcttcgcgc gccacagga agagtgga gagatgggct gggaagtga tcctgagggc 1020

ctctacaggt tgctcaatcg gttgcacttt gaataccagc cgcgcaagct ctacgtgacc 1080

gagagcgggt gcagctactc cgatggacc ggccccaacg gtcggatacc ggaccaacgc 1140

cgatatcaact acctgcgca tcacttcgca gcggcgcatc aggcgataca atgcggcgctc 1200

ccgtggccg gctacttcgt ctggtcgttc atggacaact tcgagtgggc caaagggtac 1260

acccaacgtt ttggtatcgt atgggtggat tatcaatgc aacgacggat accgaaagac 1320

agcgctact ggtatcgca tgcgtcgcc gccaacgcgg tgcaagttcc tgattag 1377

<210> 134

<211> 458

<212> PRT

<213> Unknown

<220>

<223> Obtained from environmental sample

<220>

<221> DOMAIN

<222> (2)...(454)

<223> Glycosyl hydrolase family 1

<220>

<221> SITE

<222> (10)...(24)

<223> Glycosyl hydrolases family 1 N-terminal signature. Prosite id = PS00653

<400> 134

Met	Thr	Gln	Leu	Ala	Phe	Pro	Ser	Asn	Phe	Ile	Trp	Gly	Thr	Ala	Thr
1			5					10					15		

Ser	Ala	Tyr	Gln	Ile	Glu	Gly	Ala	Trp	Asn	Ala	Asp	Gly	Lys	Gly	Glu
			20					25					30		

Ser	Ile	Trp	Asp	Arg	Phe	Ser	His	Thr	Gln	Gly	Lys	Ile	Ile	Asp	Gly
			35				40					45			

Ser	Asn	Gly	Asp	Val	Ala	Cys	Asp	His	Tyr	His	Arg	Trp	Arg	Glu	Asp
		50				55					60				

Val	Ala	Leu	Met	Arg	Asp	Leu	Gly	Met	Gln	Ala	Tyr	Arg	Phe	Ser	Ile
65					70				75						80

Ser	Trp	Pro	Arg	Ile	Leu	Pro	Thr	Gly	His	Gly	Gln	Ile	Asn	Gln	Ala
				85					90					95	

Gly	Leu	Asp	Phe	Tyr	Asn	Arg	Leu	Val	Asp	Gly	Leu	Leu	Glu	Ala	Gly
			100					105					110		

Ile	Lys	Pro	Phe	Ala	Thr	Leu	Tyr	His	Trp	Asp	Leu	Pro	Leu	Ala	Leu
				115				120				125			

Gln Ala Asp Gly Gly Trp Pro Glu Arg Ser Thr Ala Lys Ala Phe Val
130 135 140

Glu Tyr Ala Asp Val Val Ser Arg Ala Leu Gly Asp Arg Val Lys Ser
145 150 155 160

Trp Ile Thr His Asn Glu Pro Trp Cys Ile Ser Met Leu Ser His Gln
165 170 175

Ile Gly Glu His Ala Pro Gly Trp Arg Asp Trp Gln Ala Ala Leu Ala
180 185 190

Ala Ala His His Val Leu Leu Ser His Gly Trp Ala Val Pro Glu Leu
195 200 205

Arg Arg Asn Ser Arg Asp Ala Glu Ile Gly Ile Thr Leu Asn Phe Thr
210 215 220

Pro Ala Glu Pro Ala Ser Asn Ser Ala Ala Asp Phe Lys Ala Tyr Arg
225 230 235 240

Gln Phe Asp Gly Tyr Phe Asn Arg Trp Phe Leu Asp Pro Leu Tyr Gly
245 250 255

Arg His Tyr Pro Ala Asp Met Val His Asp Tyr Ile Ala Gln Gly Tyr
260 265 270

Leu Pro Ser Gln Gly Leu Thr Phe Val Glu Ala Gly Asp Leu Asp Ala
275 280 285

Ile Ala Thr Arg Thr Asp Phe Leu Gly Val Asn Tyr Tyr Thr Arg Glu
290 295 300

Val Val Arg Ser Gln Glu Ile Pro Glu Ser Glu Asn Ala Pro Arg Thr
305 310 315 320

Val Leu Arg Ala Pro Gln Glu Glu Trp Thr Glu Met Gly Trp Glu Val
325 330 335

Tyr Pro Glu Gly Leu Tyr Arg Leu Leu Asn Arg Leu His Phe Glu Tyr
340 345 350

Gln Pro Arg Lys Leu Tyr Val Thr Glu Ser Gly Cys Ser Tyr Ser Asp
355 360 365

Gly Pro Gly Pro Asn Gly Arg Ile Pro Asp Gln Arg Arg Ile Asn Tyr
370 375 380

Leu Arg Asp His Phe Ala Ala Ala His Gln Ala Ile Gln Cys Gly Val
385 390 395 400

Pro Leu Ala Gly Tyr Phe Val Trp Ser Phe Met Asp Asn Phe Glu Trp
405 410 415

Ala Lys Gly Tyr Thr Gln Arg Phe Gly Ile Val Trp Val Asp Tyr Gln
420 425 430

Ser Gln Arg Arg Ile Pro Lys Asp Ser Ala Tyr Trp Tyr Arg Asp Val
435 440 445

Val Ala Ala Asn Ala Val Gln Val Pro Asp
450 455

<210> 135

<211> 987

<212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 135

atggttgagc ctgccgatca gaggcatctt tcagatgctt ttcaggtaaa tcgcactctt 60

ggaaaaggca tcaatcttgg taacacactg gaggtccaa atgaaggcga gtggggattg 120

acaattcgcg aggagtattt tgatgaagtg aaacaagccg gatttgaatc cgtgcgtatt 180

ccgatacgat ggaatgctca tgctctggaa ggttttccat atacgataga tgaatctttt 240

tttgaccggg ttgatgaagt tattggctgg gcttttgatc gtgatcttgc agtcattgatt 300

aacattcatc actacaacga attgatggag cagccacagg atcaccggga tcgctttttg 360

aaactttggg agcaaattgc tgcgcactat aaagagtacc cggaagaact ggtattcgag 420

attttaaacg aacccacga taatctgacc ccggctatct ggaatagctt ttggctgat 480

gctctcggta ttatcgcca aaccaatcca ggaagggtta ttgcagtcgg aacagctgaa 540

tggggcgggt tcgggagttt gcaggatctt gagctgcctg ataatgaccg ccagataatc 600

accaccgttc attactataa cccatttcat ttcacgcac agggggcaga ttgggttga 660

gatgaagcgg atcagtggct tggaaccgaa tgggatggag cagatcatga aaaagctgaa 720

gttgacagcg attttgactc tgtggaacag tgggcccga atcatgaccg gccaatacac 780

gtgggagagt tcggagcttt cagcggcga gatgatttgt cacgtgaaca gtggacggca 840

tacgtacgtg agtcttcgga gaaccggcag tttagctggg cgtattggga gtttgggtca 900

gggttcggtg cctatgatcc cggttccgga gaatggcgtg aatatttact ccgggcgtta 960

atccccgaca gtccggtgat tgattaa 987

<210> 136

<211> 328

<212> PRT

<213> Unknown

<220>

<223> Obtained from environmental sample

<220>

<221> DOMAIN

<222> (27)...(306)

<223> Cellulase (glycosyl hydrolase family 5)

<220>

<221> SITE

<222> (17)...(20)

<223> N-glycosylation site. Prosite id = PS00001

<220>

<221> SITE

<222> (139)...(148)

<223> Glycosyl hydrolases family 5 signature. Prosite id = PS00659

<400> 136

Met	Val	Glu	Pro	Ala	Asp	Gln	Ser	His	Phe	Ser	Asp	Ala	Phe	Gln	Val
1			5						10					15	

Asn	Arg	Thr	Leu	Gly	Lys	Gly	Ile	Asn	Leu	Gly	Asn	Thr	Leu	Glu	Ala
		20					25						30		

Pro	Asn	Glu	Gly	Glu	Trp	Gly	Leu	Thr	Ile	Arg	Glu	Glu	Tyr	Phe	Asp
		35					40						45		

Glu	Val	Lys	Gln	Ala	Gly	Phe	Glu	Ser	Val	Arg	Ile	Pro	Ile	Arg	Trp
	50					55					60				

Asn	Ala	His	Ala	Leu	Glu	Gly	Phe	Pro	Tyr	Thr	Ile	Asp	Glu	Ser	Phe
65					70					75					80

Phe	Asp	Arg	Val	Asp	Glu	Val	Ile	Gly	Trp	Ala	Phe	Asp	Arg	Asp	Leu
			85						90					95	

Ala	Val	Met	Ile	Asn	Ile	His	His	Tyr	Asn	Glu	Leu	Met	Glu	Gln	Pro
		100						105					110		

Gln Asp His Arg Asp Arg Phe Leu Lys Leu Trp Glu Gln Ile Ala Ala
115 120 125

His Tyr Lys Glu Tyr Pro Glu Glu Leu Val Phe Glu Ile Leu Asn Glu
130 135 140

Pro His Asp Asn Leu Thr Pro Ala Ile Trp Asn Ser Phe Leu Ala Asp
145 150 155 160

Ala Leu Gly Ile Ile Arg Gln Thr Asn Pro Gly Arg Val Ile Ala Val
165 170 175

Gly Thr Ala Glu Trp Gly Gly Phe Gly Ser Leu Gln Asp Leu Glu Leu
180 185 190

Pro Asp Asn Asp Arg Gln Ile Ile Thr Thr Val His Tyr Tyr Asn Pro
195 200 205

Phe His Phe Thr His Gln Gly Ala Asp Trp Val Gly Asp Glu Ala Asp
210 215 220

Gln Trp Leu Gly Thr Glu Trp Asp Gly Ala Asp His Glu Lys Ala Glu
225 230 235 240

Val Asp Ser Asp Phe Asp Ser Val Glu Gln Trp Ala Arg Asn His Asp
245 250 255

Arg Pro Ile His Val Gly Glu Phe Gly Ala Phe Ser Ala Ala Asp Asp
260 265 270

Leu Ser Arg Glu Gln Trp Thr Ala Tyr Val Arg Glu Ser Ser Glu Asn
275 280 285

Arg Gln Phe Ser Trp Ala Tyr Trp Glu Phe Gly Ser Gly Phe Gly Ala
290 295 300

Tyr Asp Pro Gly Ser Gly Glu Trp Arg Glu Tyr Leu Leu Arg Ala Leu

305	310	315	320
Ile Pro Asp Ser Pro Val Ile Asp			
	325		
<210> 137			
<211> 702			
<212> DNA			
<213> Unknown			
<220>			
<223> Obtained from environmental sample			
<400> 137			
atgagccacc gatcgaggaa attcaacggc cagccactga tgggtgccga agacggccac			60
ttcgtgctcg gattcgggag cgacgacgag gccaccacc gactgcgcgt tcagctaccg			120
gatgagcgag tctgggagaa gaatctgcgt ccggaatcgc gcgagttcga tattcagcgg			180
atcgacggct tgccgcaaga ccaggtcacc ccacccact ccgtgctggc gagaatccga			240
gaggacgctt cgtgtgcgag ccgtgcccgc gaacgacgag atccgaggac cgactggacc			300
gatggctgga tctggccggc cgaggccgc atttccggcg tgtacggcag ccagcgcatc			360
ctcaacggtg agcctcgaa cccgactgg gggctggata tcgccgcgcc aaccggcagc			420
ccggtcgtgg cgctgccgg cggcacgctc agcctgactc atccggacat gtatttttcc			480
ggcggcacc tgtaatcga ccacggtcac ggcctggtgt ctgcgttcct ccacctgagt			540
gaaatcctgg tcgaggaagg gcagcgggtc gagcaggggg atctgatcgc acgattggc			600
gccaccggtc gtgccaccgg gccgcacctg gactggcgga tcaatctcgg cgatgtacgc			660
gtggaccac agctgctgct gccgccgatg gacgcgcagt ga			702
<210> 138			
<211> 233			
<212> PRT			

<213> Unknown

<220>

<223> Obtained from environmental sample

<220>

<221> DOMAIN

<222> (127)...(223)

<223> Peptidase family M23

<400> 138

Met Ser His Arg Ser Gln Glu Phe Asn Gly Gln Pro Leu Met Val Ser
1 5 10 15

Glu Asp Gly His Phe Val Leu Gly Phe Gly Arg Asp Asp Glu Ala Thr
20 25 30

His Arg Leu Arg Val Gln Leu Pro Asp Glu Arg Val Trp Glu Lys Asn
35 40 45

Leu Arg Pro Glu Ser Arg Glu Phe Asp Ile Gln Arg Ile Asp Gly Leu
50 55 60

Pro Gln Asp Gln Val Thr Pro Pro His Ser Val Leu Ala Arg Ile Arg
65 70 75 80

Glu Asp Ala Ser Leu Ser Arg Arg Ala Arg Glu Arg Arg Asp Pro Arg
85 90 95

Thr Asp Trp Thr Asp Gly Trp Ile Trp Pro Ala Glu Gly Arg Ile Ser
100 105 110

Gly Val Tyr Gly Ser Gln Arg Ile Leu Asn Gly Glu Pro Arg Asn Pro
115 120 125

His Trp Gly Leu Asp Ile Ala Ala Pro Thr Gly Ser Pro Val Val Ala
130 135 140

Pro Ala Gly Gly Ile Val Ser Leu Thr His Pro Asp Met Tyr Phe Ser
145 150 155 160

Gly Gly Thr Leu Leu Ile Asp His Gly His Gly Leu Val Ser Ala Phe
165 170 175

Leu His Leu Ser Glu Ile Leu Val Glu Glu Gly Gln Arg Val Glu Gln
180 185 190

Gly Asp Leu Ile Ala Arg Ile Gly Ala Thr Gly Arg Ala Thr Gly Pro
195 200 205

His Leu Asp Trp Arg Ile Asn Leu Gly Asp Val Arg Val Asp Pro Gln
210 215 220

Leu Leu Leu Pro Pro Met Asp Ala Gln
225 230

<210> 139
<211> 351
<212> DNA
<213> Unknown

<220>
<223> Obtained from environmental sample

<400> 139
atggaaaaaa ttctcgttat cggatgcgcg ggccagatag gctcagagct tacgctcgaa 60

cttcgtaaga ttatgggtga tgacaatgtg gtggctactg acattaagcc ggccagcaag 120

gaaattaccg agggcggccc ctttgaaatt cttgatgtgc tcgacacca ccggcttttt 180

ggcactgtaa gccgcaacaa gatcaccag atttatcacc ttgcagccat ctttcgggc 240

aatgccgaga aaaaaccact tgcaagctgg cacattaaca tggagagttt gctcaacgtg 300

cttgaactgg cccgtgaact gaagcttcat aaaattttct ggccaagctc a 351

<210> 140
<211> 117
<212> PRT
<213> Unknown

<220>
<223> Obtained from environmental sample

<400> 140
Met Glu Lys Ile Leu Val Ile Gly Cys Ala Gly Gln Ile Gly Ser Glu
1 5 10 15

Leu Thr Leu Glu Leu Arg Lys Ile Tyr Gly Asp Asp Asn Val Val Ala
20 25 30

Thr Asp Ile Lys Pro Ala Ser Lys Glu Ile Thr Glu Gly Gly Pro Phe
35 40 45

Glu Ile Leu Asp Val Leu Asp Thr His Arg Leu Phe Gly Thr Val Ser
50 55 60

Arg Asn Lys Ile Thr Gln Ile Tyr His Leu Ala Ala Ile Leu Ser Gly
65 70 75 80

Asn Ala Glu Lys Lys Pro Leu Ala Ser Trp His Ile Asn Met Glu Ser
85 90 95

Leu Leu Asn Val Leu Glu Leu Ala Arg Glu Leu Lys Leu His Lys Ile
100 105 110

Phe Trp Pro Ser Ser
115

<210> 141
<211> 1350
<212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 141

atgctgtcct atacgagtcg gttcccaaag aactttgtct ggggtgtggc gacggcggcg	60
ccgcagatcg agggcgctgc gcgagaagac ggaaaggcg aatcgatatg ggatcgcttt	120
tgccgctgc ccgaaagggt ccacaatggc gatactctcg atgttgcgtg cgaccactac	180
caccggttcc gggaggattt cgcgctcatg cgagacttgg gcgtgcgcca ctaccggctt	240
tcgcttgctt gggcccgcat attcccgac ggcgacggcg cattgaacca gcgcggagtg	300
gatttctacc accggctctt tgaggccatg atcgagcacg ggattacgcc ttgggtgacg	360
ctctttcact gggatttgcc gcaggcgctc gaggaccgcg gcggtgtgtg tgagcgtctc	420
accgtcgatg cattcgggcg ctacgtgac accgttgtga aggcgttttg cgatcgctg	480
aagaattgga tcacctgaa cgaaatccgc tgettccagt tgctcgctta cgatctctgc	540
atcaaggccc cgggcccga ggtctcgcg gcgcagctca accagacctc tcatcacgcg	600
ctgatctgcc atgggcatgg cgctcgggcg gtccgcgaac acggcgggcg aggcgtcgc	660
gtcgggctta ccgacaacag cgacgtatgc gtgcccgtca ccgagaccgc gcccacatc	720
attgcggcca gatcctggta tgcgtcgca aatattcatc tgctcgatcc gatctatcg	780
ggcgagtatg cgccgaata cctcgaacgc tgcggtgcgg acgcgcccc ggtggccgag	840
gacgatttcg cgctgatttc aatgccgacg gattttctcg ggctgaatgt atataggcg	900
acctttgtgc gtgccgacgc ggagggcagg ccggaggaga ttaaactgcc gcggaattac	960
ccgcgcgagg atagcgctg gttgaatatt gtgcccagt cgatgtactg ggccacacgg	1020

ctggcgcggg aaacctacgg cgtgagatca atctacatca ccgaaaacgg ctgcggctac 1080

gacgacgagc ccgtcgacgg cggcgagggtg ctcgacctgc atcgacgcga tttctgcgc 1140

aaccaccttc ggggaattgca tcgcgccata ggcgacggcg tgcccgttga cgggtatatt 1200

ctctggtcct tcatggacaa ctacgagtgg gaggacgggt atgcgcggcg gttcggcatc 1260

gttcacgtcg acttcgaaag ccagaaacgg actccaaaac tctcggcgcg ctattacgcg 1320

caggtaatga aagaaaacgg gatcctgtga 1350

<210> 142
 <211> 449
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> DOMAIN
 <222> (4)...(448)
 <223> Glycosyl hydrolase family 1

<400> 142
 Met Leu Ser Tyr Thr Ser Pro Phe Pro Lys Asn Phe Val Trp Gly Val
 1 5 10 15

Ala Thr Ala Ala Pro Gln Ile Glu Gly Ala Ala Arg Glu Asp Gly Lys
 20 25 30

Gly Glu Ser Ile Trp Asp Arg Phe Cys Arg Val Pro Gly Lys Val His
 35 40 45

Asn Gly Asp Thr Leu Asp Val Ala Cys Asp His Tyr His Arg Phe Arg
 50 55 60

Glu Asp Phe Ala Leu Met Arg Asp Leu Gly Val Arg His Tyr Arg Leu

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

260 265 270

Ala Asp Ala Pro Gln Val Ala Glu Asp Asp Phe Ala Leu Ile Ser Met
275 280 285

Pro Thr Asp Phe Leu Gly Leu Asn Val Tyr Thr Ala Thr Phe Val Arg
290 295 300

Ala Asp Ala Glu Gly Arg Pro Glu Glu Ile Lys Leu Pro Arg Asn Tyr
305 310 315 320

Pro Arg Ala Asp Ser Ala Trp Leu Asn Ile Val Pro Gln Ser Met Tyr
325 330 335

Trp Ala Thr Arg Leu Ala Arg Glu Thr Tyr Gly Val Arg Ser Ile Tyr
340 345 350

Ile Thr Glu Asn Gly Cys Gly Tyr Asp Asp Glu Pro Val Asp Gly Gly
355 360 365

Glu Val Leu Asp Leu His Arg Arg Asp Phe Leu Arg Asn His Leu Arg
370 375 380

Glu Leu His Arg Ala Ile Gly Asp Gly Val Pro Val Asp Gly Tyr Phe
385 390 395 400

Leu Trp Ser Phe Met Asp Asn Tyr Glu Trp Glu Asp Gly Tyr Ala Arg
405 410 415

Arg Phe Gly Ile Val His Val Asp Phe Glu Ser Gln Lys Arg Thr Pro
420 425 430

Lys Leu Ser Ala Arg Tyr Tyr Ala Gln Val Met Lys Glu Asn Arg Ile
435 440 445

Leu

<210> 143
 <211> 1188
 <212> DNA
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<400> 143
 atgaccatca ctttccccga cgggttcttg tgggggacgg cgacggccgc ccaccaggtg 60

 gagggcggca actggaacac cgactggttg gcctacgagc acgccccggg caccgcctgc 120

 gcggagccgt ccggcgatgc gtgcgaccac tggcaccgct acccgaggga catcgccctc 180

 ctgcgccgcg tcgggttcag tgcctaccgc ttctcggttg aatgggctcg catcgagccc 240

 gaggaagggc atttctcccg cgccaccctc gaccactacc ggcgcatgat cgcttgcctg 300

 cgcgaccacg ggttggtccc ggtggtgacc ttccaccact tcaccacccc ccgtggggcc 360

 gggcccgggg gctgctggtc cgaccgggtc accgccgagc gcttcgcccg ttactgcgag 420

 cgcacgtggc ccgccctcgg cgacgagatc gcgatggcct gcacgatcaa cgagccgaac 480

 atcgtggcca cctcgggta cttctcggc gagttcccgc cggccgtcgc cgaccccgac 540

 cgctaccggc aggcgaacga cacgtgatc cgcgcccatc gcctcgcta cgaggcgctg 600

 aaggccgggc ccggcgagtt cccgtcggc ctaccctgt cgatggccga gttcgtcgcc 660

 gagcccggcg gcgaggccca cctcgcccag gtccggcaca cgatggagga catcttctg 720

 gaggcgccc gggcgacga cttcatcggg gtgcagacct acagccgat gcgttcggt 780

 cccgactcgc cgatcccgt cgggccggcc gagggcgtcg aggtcgtcca gatggggtac 840

gagtactggc cgtgggcgct cgagcgacg atccggcgcg ccgccgaggt caccggcacg 900

gcggtccacg tcaccgagaa cggcatcggt accgccgacg acacgcagcg ggtcgcttac 960

gtcaccgagg ccctccgggg gctgcggcgc tgcctcgacg acggcatcga cgtccgcagc 1020

tacttctact ggacgtgct cgacaacttc gattggacgc gcggctacgt gccgacgttc 1080

gggtcgtcg ccgtcgaccg caccaccag cgcgggtcgg tgaagccgag cgcggtgtgg 1140

ctcggcgagg tcgcccgcac gaaccgcctc gagctcccg accgtga 1188

<210> 144
 <211> 395
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> DOMAIN
 <222> (1)...(390)
 <223> Glycosyl hydrolase family 1

<220>
 <221> SITE
 <222> (9)...(23)
 <223> Glycosyl hydrolases family 1 N-terminal signature. Prosite id = PS00653

<220>
 <221> SITE
 <222> (188)...(191)
 <223> N-glycosylation site. Prosite id = PS00001

<400> 144
 Met Thr Ile Thr Phe Pro Asp Gly Phe Trp Trp Gly Thr Ala Thr Ala
 1 5 10 15

Ala His Gln Val Glu Gly Gly Asn Trp Asn Thr Asp Trp Trp Ala Tyr
 20 25 30

Glu His Ala Pro Gly Thr Arg Cys Ala Glu Pro Ser Gly Asp Ala Cys
35 40 45

Asp His Trp His Arg Tyr Pro Glu Asp Ile Ala Leu Leu Ala Ala Leu
50 55 60

Gly Phe Ser Ala Tyr Arg Phe Ser Val Glu Trp Ala Arg Ile Glu Pro
65 70 75 80

Glu Glu Gly His Phe Ser Arg Ala Thr Leu Asp His Tyr Arg Arg Met
85 90 95

Ile Ala Cys Cys Arg Asp His Gly Leu Ala Pro Val Val Thr Phe His
100 105 110

His Phe Thr Thr Pro Arg Trp Ala Ala Ala Gly Gly Cys Trp Ser Asp
115 120 125

Pro Val Thr Ala Glu Arg Phe Ala Arg Tyr Cys Glu Arg Thr Val Ala
130 135 140

Ala Leu Gly Asp Glu Ile Ala Met Ala Cys Thr Ile Asn Glu Pro Asn
145 150 155 160

Ile Val Ala Thr Leu Gly Tyr Phe Leu Gly Glu Phe Pro Pro Ala Val
165 170 175

Ala Asp Pro Asp Arg Tyr Arg Gln Ala Asn Asp Thr Leu Ile Arg Ala
180 185 190

His Arg Leu Ala Tyr Glu Ala Leu Lys Ala Gly Pro Gly Glu Phe Pro
195 200 205

Val Gly Leu Thr Leu Ser Met Ala Glu Phe Val Ala Glu Pro Gly Gly
210 215 220

Glu Ala His Leu Ala Gln Val Arg His Thr Met Glu Asp Ile Phe Leu
225 230 235 240

Glu Ala Ala Arg Gly Asp Asp Phe Ile Gly Val Gln Thr Tyr Ser Arg
245 250 255

Met Arg Phe Gly Pro Asp Ser Pro Ile Pro Leu Gly Pro Ala Glu Gly
260 265 270

Val Glu Val Val Gln Met Gly Tyr Glu Tyr Trp Pro Trp Ala Leu Glu
275 280 285

Ala Thr Ile Arg Arg Ala Ala Glu Val Thr Gly Thr Ala Val His Val
290 295 300

Thr Glu Asn Gly Ile Gly Thr Ala Asp Asp Thr Gln Arg Val Ala Tyr
305 310 315 320

Val Thr Glu Ala Leu Arg Gly Leu Arg Arg Cys Leu Asp Asp Gly Ile
325 330 335

Asp Val Arg Ser Tyr Phe Tyr Trp Thr Leu Leu Asp Asn Phe Glu Trp
340 345 350

Thr Arg Gly Tyr Val Pro Thr Phe Gly Leu Val Ala Val Asp Arg Thr
355 360 365

Thr Gln Arg Arg Ser Val Lys Pro Ser Ala Val Trp Leu Gly Glu Val
370 375 380

Ala Arg Thr Asn Arg Leu Glu Leu Pro Asp Arg
385 390 395

<210> 145

<211> 1386

<212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 145

atgtcgtttc cgagaaattt cctgtgggga tcagccacct cctcctacca aatcgaaggc	60
gcctggcaag aagacggcaa aggcccaaat atctgggacg tgttttcaca caccgccggg	120
aaagtcgcca atggcgacac cggatgatat gccatcgacc actaccaccg ataccgagac	180
gacgttgccc tgatggctga gcttggactt caggcatacc gtttctcggt ctctggggc	240
agaataatgc cggaaggagc agggcccatc gagcaacggg gtctggactt ctacgaccgc	300
ctcattgatg cactgctgga gaaaaacatc caacccatgg ccacctcta cactgggat	360
ttaccagccg cactgcaaga cagagggggg tgactaacc gcgacagcg gtcttggttt	420
gtgactact cagccgttgt tcacgacgt ttttctgacc gggtgggaat gtgggcaacg	480
ttgaacgagc cgtgggtgtc tgcatTTTTG ggccacggaa ctggcatcca cgcacctggc	540
atcacaagcc cccacgcggc gttcgccgcg gggcatcacc tgcttctggg gcatggcaag	600
gccatccaag cgatgcgcgc tcaatcgtct agcacccaac tgggaattgt ttgaacctc	660
gccccgtgt atctgaagg tgacaccct gctgaccacc cggctcacac ctccgtggca	720
ctacacgatg ccattttgaa tgggttgttg acagagccgc ttctgcgctc cagatacccc	780
gacctgttc ttcaactagg cgacatggtg acaaaaaaca tccacgacgg tgacctgcc	840
atcatggccg agccgattga ctggatgggc atcaactact accaggacat tagatttgtg	900
gccactgatg ttgccccac ggctaaccg atggccctc cgggtaacga cctgccgggc	960
accgtcgggg tggagcctgc gccagcaatc ggaacatca ccagctttgg ctggtccacc	1020
acccccgacg gactgcgagt actgttggtg ggcctggatg aggaatacga caacctccg	1080

ccgatattca ttaccgaaaa cgggtgtgct tacgattacc ccgtcgagga cgggtgtcgtc 1140

aacgacaccc ttcgtgtcac atacatgcga gaacacctca ccgcgtttgtc gcaggccatt 1200

gaggcgggtg tgaatgtccg gggctatatg cactgggtctc tgttcgacaa cttcgagtgg 1260

gccgaagggt atcgccaacg ctttggcatg gtgcacgtcg actttgagac cttggagcgg 1320

actcccaaag cctcagctca ctactattca cgtgtcatca caaataacgc cctctctgac 1380

gactga 1386

<210> 146
 <211> 461
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> DOMAIN
 <222> (1)...(458)
 <223> Glycosyl hydrolase family 1

<220>
 <221> SITE
 <222> (7)...(21)
 <223> Glycosyl hydrolases family 1 N-terminal signature. Prosite id = PS00653

<220>
 <221> SITE
 <222> (337)...(340)
 <223> N-glycosylation site. Prosite id = PS00001

<220>
 <221> SITE
 <222> (386)...(389)
 <223> N-glycosylation site. Prosite id = PS00001

<400> 146
 Met Ser Phe Pro Arg Asn Phe Leu Trp Gly Ser Ala Thr Ser Ser Tyr
 1 5 10 15

Gln Ile Glu Gly Ala Trp Gln Glu Asp Gly Lys Gly Pro Asn Ile Trp
20 25 30

Asp Val Phe Ser His Thr Pro Gly Lys Val Ala Asn Gly Asp Thr Gly
35 40 45

Asp Ile Ala Ile Asp His Tyr His Arg Tyr Arg Asp Asp Val Ala Leu
50 55 60

Met Ala Glu Leu Gly Leu Gln Ala Tyr Arg Phe Ser Phe Ser Trp Ala
65 70 75 80

Arg Ile Met Pro Glu Gly Ala Gly Pro Ile Glu Gln Arg Gly Leu Asp
85 90 95

Phe Tyr Asp Arg Leu Ile Asp Ala Leu Leu Glu Lys Asn Ile Gln Pro
100 105 110

Met Ala Thr Leu Tyr His Trp Asp Leu Pro Ala Ala Leu Gln Asp Arg
115 120 125

Gly Gly Trp Thr Asn Arg Asp Ser Ala Ser Trp Phe Ala Asp Tyr Ser
130 135 140

Ala Val Val His Asp Ala Phe Ser Asp Arg Val Gly Met Trp Ala Thr
145 150 155 160

Leu Asn Glu Pro Trp Val Ser Ala Phe Leu Gly His Gly Thr Gly Ile
165 170 175

His Ala Pro Gly Ile Thr Ser Pro His Ala Ala Phe Ala Ala Gly His
180 185 190

His Leu Leu Leu Gly His Gly Lys Ala Ile Gln Ala Met Arg Ala Gln
195 200 205

Ser Ser Ser Thr Gln Leu Gly Ile Val Leu Asn Leu Ala Pro Val Tyr
210 215 220

Leu Glu Gly Asp Thr Pro Ala Asp His Pro Ala His Thr Ser Val Ala
225 230 235 240

Leu His Asp Ala Ile Leu Asn Gly Leu Trp Thr Glu Pro Leu Leu Arg
245 250 255

Ser Arg Tyr Pro Asp Leu Leu Leu Gln Leu Gly Asp Met Val Thr Lys
260 265 270

Asn Ile His Asp Gly Asp Leu Ala Ile Met Ala Glu Pro Ile Asp Trp
275 280 285

Met Gly Ile Asn Tyr Tyr Gln Asp Ile Arg Phe Val Ala Thr Asp Val
290 295 300

Ala Pro Thr Ala Asn Pro Met Ala Pro Pro Gly Asn Asp Leu Pro Gly
305 310 315 320

Thr Val Gly Val Glu Pro Ala Pro Ala Ile Gly Asn Ile Thr Ser Phe
325 330 335

Gly Trp Ser Thr Thr Pro Asp Gly Leu Arg Val Leu Leu Val Gly Leu
340 345 350

Asp Glu Glu Tyr Asp Asn Leu Pro Pro Ile Phe Ile Thr Glu Asn Gly
355 360 365

Cys Ala Tyr Asp Tyr Pro Val Glu Asp Gly Val Val Asn Asp Thr Leu
370 375 380

Arg Val Thr Tyr Met Arg Glu His Leu Thr Ala Leu Ser Gln Ala Ile
385 390 395 400

Glu Ala Gly Val Asn Val Arg Gly Tyr Met His Trp Ser Leu Phe Asp
405 410 415

Asn Phe Glu Trp Ala Glu Gly Tyr Arg Gln Arg Phe Gly Met Val His
420 425 430

Val Asp Phe Glu Thr Leu Glu Arg Thr Pro Lys Ala Ser Ala His Tyr
435 440 445

Tyr Ser Arg Val Ile Thr Asn Asn Ala Leu Ser Asp Asp
450 455 460

<210> 147
<211> 1242
<212> DNA
<213> Unknown

<220>
<223> Obtained from environmental sample

<400> 147
atgctaaaag ttttacgtaa acctattatt tctggattag ctttagctct attattgccg 60

gcaggggcag ctggtgccga aactaatatt tcaaagaagc caaatataag tggattaacc 120

gcgccgaat tagaccaaag atataaagat tctttcacca ttggtgctgc ggttgagccg 180

tatcaattat tagatgcaaa agattcacia atgctaaagc ggcatTTTTaa tagtatcgta 240

gcagagaatg tcatgaagcc tagtagttta cagccagtag aaggacaatt caactgggag 300

ccggctgata aacttggtca gtttgccaag gaaaatggaa tggacatgcg aggtcatacg 360

cttgtctggc atagccaggt accggattgg ttctttgaag atgcggcagg aaatccaatg 420

gtgttttggg aaaatggcag gcaagtgggt gccgatccat caaagcttca ggaaaacaaa 480

gagctcttac ttagccgatt acaaaatcat attcaggcag tcgtaacgcg ttataaagat 540

gatataaaat cttagggatgt tgtcaatgaa gtaatcgatg aatggggcgg acattctgaa 600

gggctgcgtc aaictccatg gticctcatc accggaacgg actatattaa agttgctttt 660

gaaactgcaa gagaatatgc agctccagac gctaagctgt atatcaatga ttacaataca 720

gaagtagaac caaaaaggac gcacctttat aacttagtaa aaagtttaaa agaagaacag 780

aacgttccga ttgatggigt tgggcacag tctcacattc aaattggctg gccttcagaa 840

aaagaaattg aagatactat taatatgttt gcagatcttg gtttagataa ccaaatacacc 900

gagcttgatg ttagtatgta tggctggccg gtaaggtcgt atccaactta tgatgcgac 960

ccagaactta aattcatgga tcaagcagct cgttatgac gtttatttaa gttatatgag 1020

aaattaggag ataaaatcag taatgtgaca ttctggggta ttgcggataa ccatacatgg 1080

ctgaatgacc gcgcagatgt ttactatgat gaaaatggaa atgttgtatt agatagagaa 1140

acaccaagag tagaaagagg agcaggaaaa gatgcgccat ttgtatttga tcctgaatac 1200

aatgtaaac cagcttattg ggcaattatc gatcacaaat aa 1242

<210> 148
 <211> 413
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> SIGNAL
 <222> (1)...(26)

<220>
 <221> DOMAIN
 <222> (43)...(413)

<223> Glycosyl hydrolase family 10

<220>

<221> SITE

<222> (29)...(32)

<223> N-glycosylation site. Prosite id = PS00001

<220>

<221> SITE

<222> (35)...(38)

<223> N-glycosylation site. Prosite id = PS00001

<220>

<221> SITE

<222> (298)...(308)

<223> Glycosyl hydrolases family 10 active site. Prosite id = PS00591

<220>

<221> SITE

<222> (353)...(356)

<223> N-glycosylation site. Prosite id = PS00001

<220>

<221> SITE

<222> (362)...(365)

<223> N-glycosylation site. Prosite id = PS00001

<400> 148

Met Leu Lys Val Leu Arg Lys Pro Ile Ile Ser Gly Leu Ala Leu Ala
1 5 10 15

Leu Leu Leu Pro Ala Gly Ala Ala Gly Ala Glu Thr Asn Ile Ser Lys
20 25 30

Lys Pro Asn Ile Ser Gly Leu Thr Ala Pro Gln Leu Asp Gln Arg Tyr
35 40 45

Lys Asp Ser Phe Thr Ile Gly Ala Ala Val Glu Pro Tyr Gln Leu Leu
50 55 60

Asp Ala Lys Asp Ser Gln Met Leu Lys Arg His Phe Asn Ser Ile Val
65 70 75 80

Ala Glu Asn Val Met Lys Pro Ser Ser Leu Gln Pro Val Glu Gly Gln
85 90 95

Phe Asn Trp Glu Pro Ala Asp Lys Leu Val Gln Phe Ala Lys Glu Asn
100 105 110

Gly Met Asp Met Arg Gly His Thr Leu Val Trp His Ser Gln Val Pro
115 120 125

Asp Trp Phe Phe Glu Asp Ala Ala Gly Asn Pro Met Val Val Trp Glu
130 135 140

Asn Gly Arg Gln Val Val Ala Asp Pro Ser Lys Leu Gln Glu Asn Lys
145 150 155 160

Glu Leu Leu Leu Ser Arg Leu Gln Asn His Ile Gln Ala Val Val Thr
165 170 175

Arg Tyr Lys Asp Asp Ile Lys Ser Trp Asp Val Val Asn Glu Val Ile
180 185 190

Asp Glu Trp Gly Gly His Ser Glu Gly Leu Arg Gln Ser Pro Trp Phe
195 200 205

Leu Ile Thr Gly Thr Asp Tyr Ile Lys Val Ala Phe Glu Thr Ala Arg
210 215 220

Glu Tyr Ala Ala Pro Asp Ala Lys Leu Tyr Ile Asn Asp Tyr Asn Thr
225 230 235 240

Glu Val Glu Pro Lys Arg Thr His Leu Tyr Asn Leu Val Lys Ser Leu
245 250 255

Lys Glu Glu Gln Asn Val Pro Ile Asp Gly Val Gly His Gln Ser His
260 265 270

Ile Gln Ile Gly Trp Pro Ser Glu Lys Glu Ile Glu Asp Thr Ile Asn
275 280 285

Met Phe Ala Asp Leu Gly Leu Asp Asn Gln Ile Thr Glu Leu Asp Val
290 295 300

Ser Met Tyr Gly Trp Pro Val Arg Ser Tyr Pro Thr Tyr Asp Ala Ile
305 310 315 320

Pro Glu Leu Lys Phe Met Asp Gln Ala Ala Arg Tyr Asp Arg Leu Phe
325 330 335

Lys Leu Tyr Glu Lys Leu Gly Asp Lys Ile Ser Asn Val Thr Phe Trp
340 345 350

Gly Ile Ala Asp Asn His Thr Trp Leu Asn Asp Arg Ala Asp Val Tyr
355 360 365

Tyr Asp Glu Asn Gly Asn Val Val Leu Asp Arg Glu Thr Pro Arg Val
370 375 380

Glu Arg Gly Ala Gly Lys Asp Ala Pro Phe Val Phe Asp Pro Glu Tyr
385 390 395 400

Asn Val Lys Pro Ala Tyr Trp Ala Ile Ile Asp His Lys
405 410

<210> 149

<211> 1068

<212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 149

atgacccgaa tgcgcgggat aaacatgggc ggctggctca gccaaattga cgccatacag 60

gaaaaagacc ctgatacatt tcccgaaca gacaaacata tggaaacttt tatccagcag 120

aaggattttg ccaatgtcag gagatggggt ttcgatcatg tgcgaattcc aattgacgcg 180

tatctgttct ttaccgaaaa aggagagccg attgaaaaca ggcttgccaa tcttgaccgc 240

gccgtagagt atgcgctgcc cgccggcctc aacatgatat tggacctcca cgagtgtccg 300

gggcacgatt tttcgggaagc agtaaaaagc cctgtccaaa aacttttctc gggagatgac 360

acctggataa ggaaaactga aaaaatatgg gcttgccttg ccgagcgtaa tctcaaaag 420

ggccacgtcc tttttgagac gctcaatgag cctgtcgtc ccaccgcgga gatttgaac 480

aatgttaagg acaggctctg ccgcgaaata cggtccacg cccctggtc gactataatc 540

accggtcca acatgtggaa ctcagcggca accttcgaca gcctcacgcc ctttgacgac 600

gacaacatga tctacagcgt acatttttac gagccgtgc ttttcacga ccagaacgca 660

ttgtggatcg acaatccgga aatcaggatc gcaaggccgt atccgggcga ttacggtccc 720

ggctttgtcc ccaaagacgg ttgacgctg tcggacggcg tctggaacag ggatcgtctc 780

gccggcgcat tagcgcccggt gaacgcgttc aggaaaaagt acaatgcgaa gattatctgt 840

aacgagttcg gcgtttacgc gcccgtagac cttcaatcgc agctgcgctg gtatgaagat 900

ctgctctcaa tctcaatga gacggggatc ggtttcacgt actggaacta taaaaatctc 960

gacttcggga taatttccat aggggagaag ctgcacgaag cccttcgca gtacgacaat 1020

agcgatcgaa taataaatc ggttcttgaa gtgttaaaaa agtattag 1068

<210> 150

<211> 355

<212> PRT

<213> Unknown

<220>

<223> Obtained from environmental sample

<220>

<221> DOMAIN

<222> (24)...(325)

<223> Cellulase (glycosyl hydrolase family 5)

<220>

<221> SITE

<222> (145)...(154)

<223> Glycosyl hydrolases family 5 signature. Prosite id = PS00659

<220>

<221> SITE

<222> (310)...(313)

<223> N-glycosylation site. Prosite id = PS00001

<220>

<221> SITE

<222> (350)...(353)

<223> N-glycosylation site. Prosite id = PS00001

<400> 150

Met	Thr	Arg	Met	Arg	Gly	Ile	Asn	Met	Gly	Gly	Trp	Leu	Ser	Gln	Ile
1				5					10					15	

Asp	Ala	Ile	Gln	Glu	Lys	Asp	Pro	Asp	Thr	Phe	Pro	Gly	Thr	Asp	Lys
			20					25					30		

His	Met	Glu	Thr	Phe	Ile	Gln	Gln	Lys	Asp	Phe	Ala	Asn	Val	Arg	Arg
		35						40				45			

Trp	Gly	Phe	Asp	His	Val	Arg	Ile	Pro	Ile	Asp	Ala	Tyr	Leu	Phe	Phe
	50					55				60					

Thr	Glu	Lys	Gly	Glu	Pro	Ile	Glu	Asn	Arg	Leu	Ala	Asn	Leu	Asp	Arg
65					70				75					80	

Ala	Val	Glu	Tyr	Ala	Leu	Pro	Ala	Gly	Leu	Asn	Met	Ile	Leu	Asp	Leu

85

90

95

His Glu Cys Pro Gly His Asp Phe Ser Glu Ala Val Lys Ser Pro Val
100 105 110

Gln Lys Leu Phe Ser Gly Asp Asp Thr Trp Ile Arg Lys Thr Glu Lys
115 120 125

Ile Trp Ala Cys Leu Ala Glu Arg Tyr Ser Gln Lys Gly His Val Leu
130 135 140

Phe Glu Thr Leu Asn Glu Pro Val Ala Pro Thr Ala Glu Ile Trp Asn
145 150 155 160

Asn Val Lys Asp Arg Leu Cys Arg Glu Ile Arg Leu His Ala Pro Trp
165 170 175

Ser Thr Ile Ile Thr Gly Ser Asn Met Trp Asn Ser Ala Ala Thr Phe
180 185 190

Asp Ser Leu Thr Pro Phe Asp Asp Asp Asn Met Ile Tyr Ser Val His
195 200 205

Phe Tyr Glu Pro Leu Leu Phe Thr His Gln Asn Ala Leu Trp Ile Asp
210 215 220

Asn Pro Glu Ile Arg Ile Ala Arg Pro Tyr Pro Gly Asp Tyr Gly Pro
225 230 235 240

Gly Phe Val Pro Lys Asp Gly Leu Thr Leu Ser Asp Gly Val Trp Asn
245 250 255

Arg Asp Arg Leu Ala Gly Ala Leu Ala Pro Val Asn Ala Phe Arg Lys
260 265 270

Lys Tyr Asn Ala Lys Ile Ile Cys Asn Glu Phe Gly Val Tyr Ala Pro

275	280	285
Val Asp Leu Gln Ser Gln Leu Arg Trp Tyr Glu Asp Leu Leu Ser Ile		
290	295	300
Leu Asn Glu Thr Gly Ile Gly Phe Thr Tyr Trp Asn Tyr Lys Asn Leu		
305	310	315 320
Asp Phe Gly Ile Ile Ser Ile Gly Glu Lys Leu His Glu Ala Leu Pro		
325	330	335
Gln Tyr Asp Asn Ser Asp Arg Ile Asn Lys Ser Val Leu Glu Val Leu		
340	345	350
Lys Lys Tyr		
355		
<210> 151		
<211> 1068		
<212> DNA		
<213> Unknown		
<220>		
<223> Obtained from environmental sample		
<400> 151		
atgaccagaa tgcgcggaat aaacatgggc ggctggctca gccagattga cgccatacag	60	
gaaaaagacc ccgataaatt tcccggaata gacaaacaca tggaacatt tatcggttcc	120	
aatgatTTTT ccaatgtcag gaaatggggt ttcgatcatg tgcgaatccc gattgacgcg	180	
tacTTTTTT ttaccgatca ggaagccccg attgaaaaca ggcttgcca tattgacaac	240	
gccgtaaaat acgcgcggag caacggcctc aaggtgatat tggacctcca cgagtgtccg	300	
gggcatgatt tttcggacgc ggcaaaaggc cctgtccaga aacttttctc cggagatgac	360	
acttatataa aaaagaccga aaaaatatgg gcatgtctgg ccgagcgtta ttcgaaaaac	420	

gacaacgtcc tctatgagac tctcaacgag cctgtcgccc ccacgcctga gatttggaac 480

actgttaagg acaggctctg ccgggaaata cgctgcacg cccctgggc gacgataatc 540

accggttcca atatgtggaa ttggccgagc accttgaca gcctgacgcc ctttgacgac 600

gacaacgtga tctacagcgt gcatttttac gagccgtgc tttttacga ccagaacgcg 660

ccctggatca acaattctga aatcaggatc acaaggccgt atccgggcga ttacggcccc 720

ggctttgtcc gcaatacgg cttaactctg tcagccggcg tctggaacag ggacaggctg 780

gcgaaggaat tcgcgccgt gaacgcgttc aggaaaaaat acaaggcgca ggttatatgc 840

gacgaattcg gcgtttacgc gcctgtcgag attgaatcg agcttcgatg gtatgaggat 900

ttgctctcga tctcagga gatgggtata ggttttctg actggaacta taaaaacctg 960

gactttggga taatttccat aggggagaag ctgcacgaaa gccttctgca gtacggcaac 1020

ggcgacagga taaatcatat ggttcttgac ttgctaaaga agtactaa 1068

<210> 152
 <211> 355
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> DOMAIN
 <222> (19)...(325)
 <223> Cellulase (glycosyl hydrolase family 5)

<220>
 <221> SITE
 <222> (145)...(154)
 <223> Glycosyl hydrolases family 5 signature. Prosite id = PS00659

<220>

<221> SITE

<222> (227)...(230)

<223> N-glycosylation site. Prosite id = PS00001

<400> 152

Met	Thr	Arg	Met	Arg	Gly	Ile	Asn	Met	Gly	Gly	Trp	Leu	Ser	Gln	Ile
1				5					10					15	

Asp	Ala	Ile	Gln	Glu	Lys	Asp	Pro	Asp	Lys	Phe	Pro	Gly	Ile	Asp	Lys
			20					25					30		

His	Met	Glu	Thr	Phe	Ile	Gly	Ser	Asn	Asp	Phe	Ser	Asn	Val	Arg	Lys
		35					40					45			

Trp	Gly	Phe	Asp	His	Val	Arg	Ile	Pro	Ile	Asp	Ala	Tyr	Leu	Phe	Phe
	50					55					60				

Thr	Asp	Gln	Glu	Ala	Pro	Ile	Glu	Asn	Arg	Leu	Val	His	Ile	Asp	Asn
65					70					75					80

Ala	Val	Lys	Tyr	Ala	Arg	Ser	Asn	Gly	Leu	Lys	Val	Ile	Leu	Asp	Leu
				85					90					95	

His	Glu	Cys	Pro	Gly	His	Asp	Phe	Ser	Asp	Ala	Ala	Lys	Gly	Pro	Val
			100						105				110		

Gln	Lys	Leu	Phe	Ser	Gly	Asp	Asp	Thr	Tyr	Ile	Lys	Lys	Thr	Glu	Lys
		115					120					125			

Ile	Trp	Ala	Cys	Leu	Ala	Glu	Arg	Tyr	Ser	Lys	Asn	Asp	Asn	Val	Leu
	130						135					140			

Tyr	Glu	Thr	Leu	Asn	Glu	Pro	Val	Ala	Pro	Thr	Pro	Glu	Ile	Trp	Asn
145					150					155					160

Thr	Val	Lys	Asp	Arg	Leu	Cys	Arg	Glu	Ile	Arg	Leu	His	Ala	Pro	Trp
				165					170					175	

Ala Thr Ile Ile Thr Gly Ser Asn Met Trp Asn Trp Pro Ser Thr Phe
180 185 190

Asp Ser Leu Thr Pro Phe Asp Asp Asp Asn Val Ile Tyr Ser Val His
195 200 205

Phe Tyr Glu Pro Leu Leu Phe Thr His Gln Asn Ala Pro Trp Ile Asn
210 215 220

Asn Ser Glu Ile Arg Ile Thr Arg Pro Tyr Pro Gly Asp Tyr Gly Pro
225 230 235 240

Gly Phe Val Arg Lys Tyr Gly Leu Thr Leu Ser Ala Gly Val Trp Asn
245 250 255

Arg Asp Arg Leu Ala Lys Glu Phe Ala Pro Val Asn Ala Phe Arg Lys
260 265 270

Lys Tyr Lys Ala Gln Val Ile Cys Asp Glu Phe Gly Val Tyr Ala Pro
275 280 285

Val Glu Ile Glu Ser Gln Leu Arg Trp Tyr Glu Asp Leu Leu Ser Ile
290 295 300

Leu Arg Glu Met Gly Ile Gly Phe Ser Tyr Trp Asn Tyr Lys Asn Leu
305 310 315 320

Asp Phe Gly Ile Ile Ser Ile Gly Glu Lys Leu His Glu Ser Leu Leu
325 330 335

Gln Tyr Gly Asn Gly Asp Arg Ile Asn His Met Val Leu Asp Leu Leu
340 345 350

Lys Lys Tyr
355

<210> 153
 <211> 1068
 <212> DNA
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<400> 153
 atgcaaagaa tgcgaggctt aaatatggc ggctggctca gccagattga cgccatacag 60

 gaaaaggacc ctgagggtt tcccgaata gacaaacaca tggaaacatt cattgtttcc 120

 ggagattttt acaatatcag gaaatgggtt ttcgacctg tgcggcttcc cattgactcg 180

 tacctgttct ttacgaaga cgatgcccc attgagaaca ggtttgccca tcttgaccgc 240

 gccgtacaat tcgcaagag caacagcctc aagctgatat tggacctcca cgagtgtccg 300

 ggacacgatt tttccgaagc cgcgaaagga cccgtccaga aacttttttc gggagatgac 360

 gtttacataa aaaaaaccga gaaaatctgg gcctgcctcg ccgagcgta ttcgaaaaac 420

 gacctgtac tctttgagac tctcaacgaa cctgtcgtc ccactgccga aatttggaac 480

 aaggtaagg acaggtctg cagagtaatc cgcattccag cgccttggtc gaccataatc 540

 accggtccca atatgtgaa ctgcgctcc gccttcgacg gtcttacgcc ctttgacgat 600

 ggcaacgtga tctacagct gcatttttac gagccgtgc tttttacga tcagaacgcg 660

 ccgtggatcg acaatccga gatcaggacg gcaaggccct atccggcgga ttacggcccc 720

 gcccttgctc gcaatacgg tatggcgag tcggccggca tctggaacaa gaaacggctt 780

 gcaaaagaat ttgagccgt ggacgcgttc aggaaaaat acaaggcgcg cgttatctgt 840

 aacgagtttg gcgtgtacgc ccccgccgat ctggaatcg agcttcgctg gtatgaggat 900

ctgctctcaa tctcaacgg gatgcagata ggttactcgt actggaacta caaaaatctg 960

gatttcggaa taatttccat aggggagaaa ctgcacgaaa gactttcgca gtatgacaac 1020

gacgagcgga taaaccaccc ggtgctgaat gtgctgaaga aatattaa 1068

<210> 154

<211> 355

<212> PRT

<213> Unknown

<220>

<223> Obtained from environmental sample

<220>

<221> DOMAIN

<222> (19)...(325)

<223> Cellulase (glycosyl hydrolase family 5)

<220>

<221> SITE

<222> (145)...(154)

<223> Glycosyl hydrolases family 5 signature. Prosite id = PS00659

<400> 154

Met Gln Arg Met Arg Gly Leu Asn Ile Gly Gly Trp Leu Ser Gln Ile
1 5 10 15

Asp Ala Ile Gln Glu Lys Asp Pro Glu Gly Phe Pro Gly Ile Asp Lys
20 25 30

His Met Glu Thr Phe Ile Val Ser Gly Asp Phe Tyr Asn Ile Arg Lys
35 40 45

Trp Gly Phe Asp His Val Arg Leu Pro Ile Asp Ser Tyr Leu Phe Phe
50 55 60

Thr Glu Asp Asp Ala Pro Ile Glu Asn Arg Phe Ala His Leu Asp Arg
65 70 75 80

Ala Val Gln Phe Ala Lys Ser Asn Ser Leu Lys Leu Ile Leu Asp Leu
85 90 95

His Glu Cys Pro Gly His Asp Phe Ser Glu Ala Ala Lys Gly Pro Val
100 105 110

Gln Lys Leu Phe Ser Gly Asp Asp Val Tyr Ile Lys Lys Thr Glu Lys
115 120 125

Ile Trp Ala Cys Leu Ala Glu Arg Tyr Ser Lys Asn Asp His Val Leu
130 135 140

Phe Glu Thr Leu Asn Glu Pro Val Ala Pro Thr Ala Glu Ile Trp Asn
145 150 155 160

Lys Val Lys Asp Arg Leu Cys Arg Val Ile Arg Ile His Ala Pro Trp
165 170 175

Ser Thr Ile Ile Thr Gly Ser Asn Met Trp Asn Ser Pro Ser Ala Phe
180 185 190

Asp Gly Leu Thr Pro Phe Asp Asp Gly Asn Val Ile Tyr Ser Val His
195 200 205

Phe Tyr Glu Pro Leu Leu Phe Thr His Gln Asn Ala Pro Trp Ile Asp
210 215 220

Asn Pro Glu Ile Arg Thr Ala Arg Pro Tyr Pro Gly Asp Tyr Gly Pro
225 230 235 240

Gly Leu Val Arg Lys Tyr Gly Met Ala Gln Ser Ala Gly Ile Trp Asn
245 250 255

Lys Lys Arg Leu Ala Lys Glu Phe Glu Pro Val Asp Ala Phe Arg Lys
260 265 270

Lys Tyr Lys Ala Arg Val Ile Cys Asn Glu Phe Gly Val Tyr Ala Pro

275	280	285
Ala Asp Leu Glu Ser Gln Leu Arg Trp Tyr Glu Asp Leu Leu Ser Ile 290	295	300
Leu Asn Gly Met Gln Ile Gly Tyr Ser Tyr Trp Asn Tyr Lys Asn Leu 305	310	315 320
Asp Phe Gly Ile Ile Ser Ile Gly Glu Lys Leu His Glu Arg Leu Ser 325	330	335
Gln Tyr Asp Asn Asp Glu Arg Ile Asn His Pro Val Leu Asn Val Leu 340	345	350
Lys Lys Tyr 355		
<210> 155		
<211> 954		
<212> DNA		
<213> Unknown		
<220>		
<223> Obtained from environmental sample		
<400> 155		
atgttaaagg attccggttt ttataagggc atcaatctcg gcggctggct gtcccagtgc	60	
gactacagcg aggagcgctt gaacagcttc atcaccgaaa aggactttga ggtgatcgcc	120	
tcctgggggtt ttgaccacgt ccgcctcccg gtggactata atgcatcca ggatgcggaa	180	
ggccgcatga tggagaaagg ccttgacgc atcgacgccg cgcttcggtt ttgtgagaag	240	
accgggcttc acatggttct cgacctgcat aagacaccgg gcttttcctt cgaccgcag	300	
gagcaggaga tgggattctt ccggtcggcg cccgaccagc agctcttcta cacgatctgg	360	
gagagccttg ctgcccggta tgcagacaaa tcggagatac tcattgttga tcttctgaac	420	

gagatcacgg agccggcgta tctggaggac tggaaccgga tttccgcgga atgcatccgc 480

cgcattccggc gtacgatgcc ggacgtccga attctggtcg gaagctatca ccacaatgcc 540

gtcagcgcgg taaaggacct gcctgcgccg gcagacgata aggtttttta cagctttcac 600

tgttacgacc ctcacaccta taccaccag ggcgcttact ggatgccgga tgactttgac 660

atcgatgcaa gagtttcctt ccgcgacacc ggcgttacct ccgtcttctt cgaaaagctg 720

tttgctcgcg ccgttgaaaa ggcgaggcg gaaggacgg aactgtactg cggagaatac 780

ggcgtcatcg acattgttcc gccggaggat gccgttctct ggttccggac cattcatgag 840

gtctttgaag cattcgggat tgcaagaagc gtctggagct ataaggaaat ggatttcggt 900

ctcgcgcacc cccgcatgga tgcggtccgg gcagagctgc tgacctgtct ctga 954

<210> 156
 <211> 317
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> DOMAIN
 <222> (14)...(302)
 <223> Cellulase (glycosyl hydrolase family 5)

<400> 156
 Met Leu Lys Asp Ser Gly Phe Tyr Lys Gly Ile Asn Leu Gly Gly Trp
 1 5 10 15

Leu Ser Gln Cys Asp Tyr Ser Glu Glu Arg Leu Asn Ser Phe Ile Thr
 20 25 30

Glu Lys Asp Phe Glu Val Ile Ala Ser Trp Gly Phe Asp His Val Arg

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

225 230 235 240

Phe Ala Ser Ala Val Glu Lys Ala Gln Ala Glu Gly Thr Glu Leu Tyr
245 250 255

Cys Gly Glu Tyr Gly Val Ile Asp Ile Val Pro Pro Glu Asp Ala Val
260 265 270

Leu Trp Phe Arg Thr Ile His Glu Val Phe Glu Ala Phe Gly Ile Ala
275 280 285

Arg Ser Val Trp Ser Tyr Lys Glu Met Asp Phe Gly Leu Ala Asp Pro
290 295 300

Arg Met Asp Ala Val Arg Ala Glu Leu Leu Thr Cys Leu
305 310 315

<210> 157

<211> 954

<212> DNA

<213> Unknown

 $\langle 220 \rangle$

<223> Obtained from environmental sample

<400> 157

atgttaaagg attccggttt ttataagggc atcaatctcg gcggtctggct gtcccagtgc 60

gactacagcg aggagcgcct gaacagcttc atcaccgaaa aagactttga ggtgatcgcc 120

tcctgggggtt ttgaccacgt cegtctgccg gtggactata atgtcatcca ggatgcggaa 180

ggccgcatga tggaggaagg cctcgcacgc atcgacgccg cgcttcggtt ttgtgaaaag 240

accgggcttc acatggttct cgacctgcat aagacaccgg gcttttcttt cgaccgcgag 300

gagcaggaga tgggattctt cccgtcggcg cccgaccagc agcgcttcta cacgatctgg 360

gagagccttg ctgcccgta tgcagacaaa tcggagatgc tcatgttcga tcttctgaac 420

gagatcacgg agccggcgta tctgaaggac tggaaccgga tttccgcgga atgcatccgc 480

cgcacccggc gtacgatgcc ggacgtccgg attctggtcg gaagctatca ccacaatgcc 540

gtcagcgcg taaaggacct gcctgcgccg gcggacgacc gggtttttta cagctttcac 600

tgttacgacc ctcacaccta taccaccag ggcgcttact ggatgccgga tgactttgac 660

atcgatgcaa gagtttcctt ccgcgacatc ggcgtcaccc ccgccttctt cgaagagctg 720

tttgcacatg ccgttgaaaa ggcgaagggtg gaagggacgg aactgtactg cggagaatac 780

ggcgatcatg acattgttcc gccggaggat gccgttctct ggttccggac cattcatgag 840

gtctttgaga aatacgggat tgcaagaagc gtctggagct ataaggaaat ggatttcggt 900

ctctccgacc ccgcatgga cgcggtccgg gcagagctgc tgacctgtct ctga 954

<210> 158
 <211> 317
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> DOMAIN
 <222> (14)...(302)
 <223> Cellulase (glycosyl hydrolase family 5)

<400> 158
 Met Leu Lys Asp Ser Gly Phe Tyr Lys Gly Ile Asn Leu Gly Gly Trp
 1 5 10 15

Leu Ser Gln Cys Asp Tyr Ser Glu Glu Arg Leu Asn Ser Phe Ile Thr
 20 25 30

Glu Lys Asp Phe Glu Val Ile Ala Ser Trp Gly Phe Asp His Val Arg

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

225 230 235 240

Phe Ala Ser Ala Val Glu Lys Ala Lys Val Glu Gly Thr Glu Leu Tyr
245 250 255

Cys Gly Glu Tyr Gly Val Ile Asp Ile Val Pro Pro Glu Asp Ala Val
260 265 270

Leu Trp Phe Arg Thr Ile His Glu Val Phe Glu Lys Tyr Gly Ile Ala
275 280 285

Arg Ser Val Trp Ser Tyr Lys Glu Met Asp Phe Gly Leu Ser Asp Pro
290 295 300

Arg Met Asp Ala Val Arg Ala Glu Leu Leu Thr Cys Leu
305 310 315

<210>	159
<211>	1023
<212>	DNA
<213>	Unknown

<220>
<223> Obtained from environmental sample

<400> 159
atgaatccaa cattcagttc cgtaccggca ttaaaggagc tgttttgcggc ggacttcaac 60

atcggggcg ggtgaatcc gacgacgatc cggacgcagg aggcgttgct ggcttatcat 120

tttaacagcc tgactgcgga gaacgagatg aagttcgtca gcgtgcatcc ggaggagcag 180

acctatacct tcgaggcggc ggaccggctg gtcgaattcg cccgagagca cggcatggcc 240

atgcggggac acacgtcgt atggcataac cagacgtccg attggctgtt ccaggatcgc 300

caaggcggga gggtaa gcaaggagtgctg ctcggaaggc tccgggagca tattcatacc 360

atagtaggcc ggtacaagaa cgagatctac gcctgggacg tcgtcaacga ggtcatcgcg	420
gacgaagggg aggcgctgct gcgcacttcc aaatggacgg aaatcgcggg acctgaattt	480
atcgctaaag cgttcgagta tgcacatgag gcggatccac aggcgctgtt gttttataac	540
gactacaacg aatcgaatcc tctgaaacgc gataaaattt acacactcgt tcattcgctg	600
ctggagcaag gggtgccgat ccatggcatc ggattacaag cgactggaa cctgtacgat	660
ccatcggttg atgagattaa ggcagcgatt gagaagtatg cttcgctggg ttgcagctg	720
cagctgacgg agctggatct ctcatgttc cgcttcgatg accggcgaac cgatttgacc	780
gcgccagagc cggggatgct ggagcaacag gccgagcgtt atgaagccgt gttccggctg	840
ttgctggagt atcgtgacgt catcagcggc gttaccttct ggggagcggc ggatgattat	900
acctggctgg acaattttcc ggtgcgcggc cggaagaact ggccgtttct gttcgaigcc	960
cagcaccagc cgaaggcagc ttatcacctg gtggcggcat tggctgcgga gcaacgagca	1020
taa	1023

<210> 160
 <211> 340
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> DOMAIN
 <222> (10)...(335)
 <223> Glycosyl hydrolase family 10

<220>
 <221> SITE
 <222> (91)...(94)
 <223> N-glycosylation site. Prosit id = PS00001

<220>

<221> SITE

<222> (185)...(188)

<223> N-glycosylation site. Prosite id = PS00001

<400> 160

Met Asn Pro Thr Phe Ser Ser Val Pro Ala Leu Lys Glu Leu Phe Ala
1 5 10 15

Ala Asp Phe Asn Ile Gly Ala Ala Val Asn Pro Thr Thr Ile Arg Thr
20 25 30

Gln Glu Ala Leu Leu Ala Tyr His Phe Asn Ser Leu Thr Ala Glu Asn
35 40 45

Glu Met Lys Phe Val Ser Val His Pro Glu Glu Gln Thr Tyr Thr Phe
50 55 60

Glu Ala Ala Asp Arg Leu Val Glu Phe Ala Arg Glu His Gly Met Ala
65 70 75 80

Met Arg Gly His Thr Leu Val Trp His Asn Gln Thr Ser Asp Trp Leu
85 90 95

Phe Gln Asp Arg Gln Gly Gly Arg Val Ser Lys Glu Val Leu Leu Gly
100 105 110

Arg Leu Arg Glu His Ile His Thr Ile Val Gly Arg Tyr Lys Asn Glu
115 120 125

Ile Tyr Ala Trp Asp Val Val Asn Glu Val Ile Ala Asp Glu Gly Glu
130 135 140

Ala Leu Leu Arg Thr Ser Lys Trp Thr Glu Ile Ala Gly Pro Glu Phe
145 150 155 160

Ile Ala Lys Ala Phe Glu Tyr Ala His Glu Ala Asp Pro Gln Ala Leu

165

170

175

Leu Phe Tyr Asn Asp Tyr Asn Glu Ser Asn Pro Leu Lys Arg Asp Lys
180 185 190

Ile Tyr Thr Leu Val His Ser Leu Leu Glu Gln Gly Val Pro Ile His
195 200 205

Gly Ile Gly Leu Gln Ala His Trp Asn Leu Tyr Asp Pro Ser Leu Asp
210 215 220

Glu Ile Lys Ala Ala Ile Glu Lys Tyr Ala Ser Leu Gly Leu Gln Leu
225 230 235 240

Gln Leu Thr Glu Leu Asp Leu Ser Met Phe Arg Phe Asp Asp Arg Arg
245 250 255

Thr Asp Leu Thr Ala Pro Glu Pro Gly Met Leu Glu Gln Gln Ala Glu
260 265 270

Arg Tyr Glu Ala Val Phe Arg Leu Leu Leu Glu Tyr Arg Asp Val Ile
275 280 285

Ser Gly Val Thr Phe Trp Gly Ala Ala Asp Asp Tyr Thr Trp Leu Asp
290 295 300

Asn Phe Pro Val Arg Gly Arg Lys Asn Trp Pro Phe Leu Phe Asp Ala
305 310 315 320

Gln His Gln Pro Lys Ala Ala Tyr His Arg Val Ala Ala Leu Ala Ala
325 330 335

Glu Gln Arg Ala
340

<210> 161
<211> 2820
<212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 161

atgtcctgcc gcacctgat gagtaggcgt gtaggatggg gacttttatt gtggggaggt	60
ttattcctca gaaccggttc ggttacagga caaacttaca attatgccga agtcctgcag	120
aaatctatgt ttttctacga atgtcaggag tctaaaattg ccccgggcaa tcgggtgaca	180
tggcgagcta atgcagccat gaacgatggg agcgatgttg gaaaagacct gacaggagga	240
tggtttgatg caggtgacca tgtgaaattt aattttccca tggcgtttac cgctacggcg	300
ctggcgtagg gagctattga ctttgctcag ggatacatta gttccgggca aatgcaatac	360
ctgaaacgta atctgcgcta cgtcaatgac tatttcatta aatgtcacac agcccccaac	420
gaattgtatg gtcaggtggg taatggaggc cttgacatg ccttttgggg accacccgaa	480
gtcatgcgca tggctaggcc tgcctataaa attgatgcgt caaaaccgg atcagatctg	540
gctgccgaaa cagctgctgc aatggctgcc gccagcattg ttttcaaac cgacgacct	600
acctatagcg ctactttgct gaatcatgca aaacagctgt tttcttttgc cgaaacctat	660
aaaggaaaat attccgacgc tattaccgat gctgcaggat attataactc ctggagcggc	720
tataacgatg aactgggatg gggagctata tggctttacc gggctaccgg cgatgcaacc	780
tatctatcta aggcagaatc ctattacgac aatctgggta atcagggtca ggaaccggtt	840
aaagcctaca aatggacatc tgcattggat gacaaatcct atggctgtta tgcctactg	900
gccaaattga caggtaagga aaaatacaaa attgacgccg aacgttttct cgactattgg	960
accgatgggtt ataatgggtc ccggattact tataccccgg gaggactcgc tttcctcgat	1020

atatggggat cgttgcgcta tgctatgaat actgcctttg ttgctgccta ctatgccgat 1080
 gcagccactt cagctgctaa aaccacaaaa tatctcaact ttgctaaaca acaactgcat 1140
 tatgctcttg gatccaatcc gagcaacaga agctatgtct gtggctttgg caataatcct 1200
 cccgttaatc ctcacatag aggtgcacac ggagcatggt ctaataatgt tcaaggacct 1260
 cctaccgaaa cagacatat cctctacggc gcattagtgg gtggaccagg cagtaatgac 1320
 tcctatactg acgaccgatc caattacacc aataacgaag tagcatgtga ctacaatgct 1380
 cttttctccg gactgcttgc aaagttcgtc attgattatg gaggcacacc gttagccaac 1440
 ttccctgttc gtgaaacccc aaaagatgaa tatttcgttg aagcaaaagc aaacgctaca 1500
 ggaaccaatt tctccgaatg gacggtatgg gtatacaacc aactgcatg gccagcccggt 1560
 gaaggttctg aatataaatt cagattatac gtaaataatt cggaaggact ggctgcaggc 1620
 tatactgcct caaattatgt tgtgcaaacc aataatgccg gtgtggtaaa ctttacccaa 1680
 cttttagctg ctgatgcagc taacggcatc tattataccg aagtaacctt taaacctggt 1740
 accgaaattt atcctggcgg gcaacagtat gacaagaagg aagctcagat gcgtattagt 1800
 ttgccaatg ctccggcttc tgcatgggat ccgactaacg acccgtcatg ggcgggaatc 1860
 acctctacct tgaacaaaat gccgggtata cccatgtatg tagatggtgt aaaggtattt 1920
 ggtaatgagc ctgtcccagg tcagacagtt cccgtcaccc gagtaaccgt atcgctacc 1980
 acctgagtc tgactgtagg acagaccagt aactcaccg ctaccgtatc gccggctaatt 2040
 gctaccaaca aaaacgtcac ctggagcagc agcaatacca gcgtagccac ggtaagctca 2100
 acaggcggtg tcacagccgt agcagccggt tcggccacca tcaccgtaac cacagtcgat 2160

ggcgctaaaa cagccacctg cgccgtaacg gtaacaggca gcaccaacgt tcccgtcacc 2220

ggagtaaccg tatcgcccac cacgctgagt ctgaccgtag ggcagaccgc taccctcacc 2280

gtaccgtat cgccggctaa tgctaccaac aagaacgtta cctggagcag cagcaatacc 2340

agcgtagcca cggtaaagttc aacaggcgta gttactgccg tagcggccgg ttcggccacc 2400

atcacgtaa ccaccgtcga tggagctaaa accgtacct gcaccgtaac ggtaacgggc 2460

agcactaccg taccctcac cggcgtaact gtatcgcta ccacctgag tctgaccgtt 2520

ggacaaaccg ctaccctgac cgctaccgta tcgccagctg atgctacaa caagaacgtc 2580

acctggagca gcagcaatac cagcgtagcc acggttaagct caacaggcgt agtcactgcc 2640

gtagcggccg gttcagctac catcacctg accacagtcg atggggctaa aactgctacc 2700

tgtgccgtga ccgtaacgc cggaggttcc accacccct gcagtaatcc ggtaagcaaa 2760

accctacctc tggtagagga tggtgccggc gaattcaggt tgagtaatag tttaaattaa 2820

<210> 162
 <211> 939
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> SIGNAL
 <222> (1)...(30)

<220>
 <221> DOMAIN
 <222> (34)...(469)
 <223> Glycosyl hydrolase family 9

<220>
 <221> DOMAIN
 <222> (491)...(576)

<223> Cellulose binding domain

<220>

<221> DOMAIN

<222> (738)...(816)

<223> Bacterial Ig-like domain (group 2)

<220>

<221> DOMAIN

<222> (825)...(903)

<223> Bacterial Ig-like domain (group 2)

<220>

<221> DOMAIN

<222> (651)...(729)

<223> Bacterial Ig-like domain (group 2)

<400> 162

Met	Ser	Cys	Arg	Thr	Leu	Met	Ser	Arg	Arg	Val	Gly	Trp	Gly	Leu	Leu
1				5					10					15	

Leu	Trp	Gly	Gly	Leu	Phe	Leu	Arg	Thr	Gly	Ser	Val	Thr	Gly	Gln	Thr
		20					25						30		

Tyr	Asn	Tyr	Ala	Glu	Val	Leu	Gln	Lys	Ser	Met	Phe	Phe	Tyr	Glu	Cys
	35					40					45				

Gln	Glu	Ser	Lys	Ile	Ala	Pro	Gly	Asn	Arg	Val	Thr	Trp	Arg	Ala	Asn
	50					55					60				

Ala	Ala	Met	Asn	Asp	Gly	Ser	Asp	Val	Gly	Lys	Asp	Leu	Thr	Gly	Gly
65				70					75					80	

Trp	Phe	Asp	Ala	Gly	Asp	His	Val	Lys	Phe	Asn	Phe	Pro	Met	Ala	Phe
			85						90					95	

Thr	Ala	Thr	Ala	Leu	Ala	Trp	Gly	Ala	Ile	Asp	Phe	Ala	Gln	Gly	Tyr
		100					105						110		

Ile Ser Ser Gly Gln Met Gln Tyr Leu Lys Arg Asn Leu Arg Tyr Val
115 120 125

Asn Asp Tyr Phe Ile Lys Cys His Thr Ala Pro Asn Glu Leu Tyr Gly
130 135 140

Gln Val Gly Asn Gly Gly Leu Asp His Ala Phe Trp Gly Pro Pro Glu
145 150 155 160

Val Met Arg Met Ala Arg Pro Ala Tyr Lys Ile Asp Ala Ser Lys Pro
165 170 175

Gly Ser Asp Leu Ala Ala Glu Thr Ala Ala Ala Met Ala Ala Ala Ser
180 185 190

Ile Val Phe Lys Ser Asp Asp Pro Thr Tyr Ser Ala Thr Leu Leu Asn
195 200 205

His Ala Lys Gln Leu Phe Ser Phe Ala Glu Thr Tyr Lys Gly Lys Tyr
210 215 220

Ser Asp Ala Ile Thr Asp Ala Ala Gly Tyr Tyr Asn Ser Trp Ser Gly
225 230 235 240

Tyr Asn Asp Glu Leu Val Trp Gly Ala Ile Trp Leu Tyr Arg Ala Thr
245 250 255

Gly Asp Ala Thr Tyr Leu Ser Lys Ala Glu Ser Tyr Tyr Asp Asn Leu
260 265 270

Gly Asn Gln Gly Gln Glu Pro Val Lys Ala Tyr Lys Trp Thr Ile Ala
275 280 285

Trp Asp Asp Lys Ser Tyr Gly Cys Tyr Ala Leu Leu Ala Lys Leu Thr
290 295 300

Gly Lys Glu Lys Tyr Lys Ile Asp Ala Glu Arg Phe Leu Asp Tyr Trp
305 310 315 320

Thr Asp Gly Tyr Asn Gly Ser Arg Ile Thr Tyr Thr Pro Gly Gly Leu
325 330 335

Ala Phe Leu Asp Ile Trp Gly Ser Leu Arg Tyr Ala Met Asn Thr Ala
340 345 350

Phe Val Ala Ala Tyr Tyr Ala Asp Ala Ala Thr Ser Ala Ala Lys Thr
355 360 365

Thr Lys Tyr Leu Asn Phe Ala Lys Gln Gln Leu His Tyr Ala Leu Gly
370 375 380

Ser Asn Pro Ser Asn Arg Ser Tyr Val Cys Gly Phe Gly Asn Asn Pro
385 390 395 400

Pro Val Asn Pro His His Arg Gly Ala His Gly Ala Trp Ser Asn Asn
405 410 415

Val Gln Gly Pro Pro Thr Glu Thr Arg His Ile Leu Tyr Gly Ala Leu
420 425 430

Val Gly Gly Pro Gly Ser Asn Asp Ser Tyr Thr Asp Asp Arg Ser Asn
435 440 445

Tyr Thr Asn Asn Glu Val Ala Cys Asp Tyr Asn Ala Leu Phe Ser Gly
450 455 460

Leu Leu Ala Lys Phe Val Ile Asp Tyr Gly Gly Thr Pro Leu Ala Asn
465 470 475 480

Phe Pro Val Arg Glu Thr Pro Lys Asp Glu Tyr Phe Val Glu Ala Lys
485 490 495

Ala Asn Ala Thr Gly Thr Asn Phe Ser Glu Trp Thr Val Trp Val Tyr
500 505 510

Asn His Thr Ala Trp Pro Ala Arg Glu Gly Ser Glu Tyr Lys Phe Arg
515 520 525

Leu Tyr Val Asn Ile Ser Glu Gly Leu Ala Ala Gly Tyr Thr Ala Ser
530 535 540

Asn Tyr Val Val Gln Thr Asn Asn Ala Gly Val Val Asn Phe Thr Gln
545 550 555 560

Leu Leu Ala Ala Asp Ala Ala Asn Gly Ile Tyr Tyr Thr Glu Val Thr
565 570 575

Phe Lys Pro Gly Thr Glu Ile Tyr Pro Gly Gly Gln Gln Tyr Asp Lys
580 585 590

Lys Glu Ala Gln Met Arg Ile Ser Leu Pro Asn Ala Pro Ala Ser Ala
595 600 605

Trp Asp Pro Thr Asn Asp Pro Ser Trp Ala Gly Ile Thr Ser Thr Leu
610 615 620

Lys Gln Met Pro Gly Ile Pro Met Tyr Val Asp Gly Val Lys Val Phe
625 630 635 640

Gly Asn Glu Pro Val Pro Gly Gln Thr Val Pro Val Thr Gly Val Thr
645 650 655

Val Ser Pro Thr Thr Leu Ser Leu Thr Val Gly Gln Thr Ser Thr Leu
660 665 670

Thr Ala Thr Val Ser Pro Ala Asn Ala Thr Asn Lys Asn Val Thr Trp
675 680 685

Ser Ser Ser Asn Thr Ser Val Ala Thr Val Ser Ser Thr Gly Val Val
690 695 700

Thr Ala Val Ala Ala Gly Ser Ala Thr Ile Thr Val Thr Thr Val Asp
705 710 715 720

Gly Ala Lys Thr Ala Thr Cys Ala Val Thr Val Thr Gly Ser Thr Asn
725 730 735

Val Pro Val Thr Gly Val Thr Val Ser Pro Thr Thr Leu Ser Leu Thr
740 745 750

Val Gly Gln Thr Ala Thr Leu Thr Ala Thr Val Ser Pro Ala Asn Ala
755 760 765

Thr Asn Lys Asn Val Thr Trp Ser Ser Ser Asn Thr Ser Val Ala Thr
770 775 780

Val Ser Ser Thr Gly Val Val Thr Ala Val Ala Ala Gly Ser Ala Thr
785 790 795 800

Ile Thr Val Thr Thr Val Asp Gly Ala Lys Thr Ala Thr Cys Thr Val
805 810 815

Thr Val Thr Gly Ser Thr Thr Val Pro Val Thr Gly Val Thr Val Ser
820 825 830

Pro Thr Thr Leu Ser Leu Thr Val Gly Gln Thr Ala Thr Leu Thr Ala
835 840 845

Thr Val Ser Pro Ala Asp Ala Thr Asn Lys Asn Val Thr Trp Ser Ser
850 855 860

Ser Asn Thr Ser Val Ala Thr Val Ser Ser Thr Gly Val Val Thr Ala
865 870 875 880

Val Ala Ala Gly Ser Ala Thr Ile Thr Val Thr Thr Val Asp Gly Ala
885 890 895

Lys Thr Ala Thr Cys Ala Val Thr Val Thr Ala Gly Gly Ser Thr Thr
900 905 910

Pro Cys Ser Asn Pro Val Ser Lys Thr Leu Pro Leu Val Gln Asp Gly
915 920 925

Ala Gly Glu Phe Arg Leu Ser Asn Ser Phe Asn
930 935

<210> 163

<211> 2733

<212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 163

atgcaaactt acaattatgc cgaagtcctg cagaaatcta tgtttttcta cgaatgtcag 60

gagtctaaaa ttccccggg caatcgggtg acatggcgag ctaatgcagc catgaacgat	120
gggagcgatg ttggaaaaga cctgacagga ggatggtttg atgcaggatga ccatgtgaaa	180
tttaattttc ccatggcggtt taccgctacg gcgctggcgt ggggagctat tgactttgct	240
cagggataca ttagttccgg gcaaatgcaa tacctgaaac gtaatctgcg ctacgtcaat	300
gactatttca ttaaatgtca cacagcccc aacgaattgt atggtcaggt gggtaatgga	360
ggccttgacc atgccttttg gggaccacc gaagtcatgc gcatggctag gcctgcctat	420
aaaattgatg cgtcaaaacc cggatcagat ctggctgccg aaacagctgc tgcaatggct	480
gccgccagca ttgttttcaa atccgacgat cctacctata gcgctacttt gctgaatcat	540
gcaaaacagc tgttttcttt tgccgaaacc tataaaggaa aatattccga cgctattacc	600
gatgctgcag gatattataa ctctgggagc ggctataacg atgaactggt atggggagct	660
atatggcttt accgggctac cggcgatgca acctatctat ctaaggcaga atcctattac	720
gacaatctgg gtaatcaggg tcaggaaccc gttaaagcct acaaatggac cattgcatgg	780
gatgacaaat cciatggctg ttatgcccia ctggccaaat tgacaggtaa ggaaaaatac	840
aaaattgacg ccgaacgttt tctcgactat tggaccgatg gttataatgg ttcccggatt	900
acttatacc cgaggaggact cgctttctc gatatatggg gatcgttgcg ctatgctatg	960
aatactgcct ttgttgctgc ctactatgcc gatgcagcca cttcagctgc taaaaccaca	1020
aaatatctca actttgctaa acaacaactg cattatgctc ttggatcaa tccgagcaac	1080
agaagctatg tctgtggctt tggcaataat cctcccgtta atcctacca tagagggtgca	1140
cacggagcat ggtctaataa tgttcaagga cctctaccg aaacacgaca tatectctac	1200
ggcgcattag tgggtggacc aggcagtaat gactcctata ctgacgaccg atccaattac	1260

accaataacg aagtagcatg tgactacaat gctcttttct cggactgct tgcaaagttc	1320
gtcattgatt atggaggcac accgttagcc aacttccctg ttcgtgaaac cccaaaagat	1380
gaatatttcg ttgaagcaaa agcaaacgct acaggaacca atttctccga atggacggta	1440
tgggtataca accacactgc atggccagcc cgtgaaggtt ctgaatataa attcagatta	1500
tacgtaaata ttctggaagg actggctgca ggctatactg cctcaaatta tgttgtgcaa	1560
accaataatg ccggtgtggt aaactttacc caacttttag ctgctgatgc agctaacggc	1620
atctattata ccgaagtaac ctttaaacct ggtaccgaaa tttatcctgg cgggcaacag	1680
tatgacaaga aggaagctca gatgcgtatt agtttgccca atgctccggc ttctgcatgg	1740
gacccgacta acgaccgctc atgggcggga atcacctcta ccttgaaaca aatgccgggt	1800
atacccatgt atgtagatgg tgtaaaggta tttggtaatg agcctgtccc aggtcagaca	1860
gttcccgta ccggagtaac cgtatcgct accaccctga gtctgactgt aggacagacc	1920
agtacactca ccgtaccgt atcgccggct aatgtacca aaaaaacgt cacctggagc	1980
agcagcaata ccagcgtagc cacggtaacg tcaacaggcg ttgtcacagc cgtagcagcc	2040
ggttcggcca ccatcacctg aaccacagtc gatggcgcta aaacagccac ctgcgccgta	2100
acggtaacag gcagcaccaa cggtcccgct accggagtaa ccgtatcgcc caccacgtg	2160
agtctgaccg tagggcagac cgctaccctc accgtaccg tatcgccggc taatgctacc	2220
aacaagaacg ttacctggag cagcagcaat accagcgtag ccacggttag ttcaacaggc	2280
gtagtactg ccgtagcggc cggttcggcc accatcacct taaccacgt cgatggagct	2340
aaaaccgcta cctgcaccgt aacggtaacg ggcagcacta ccgtaccgt caccggcgta	2400

actgtatcgc ctaccacct gagtctgacc gttggacaaa ccgtaccct gaccgtacc 2460

gtatcgccag ctgatgctac caacaagaac gtcacctgga gcagcagcaa taccagcgta 2520

gccacggtaa gtcacacagg cgtagtcact gccgtagcgg ccggttcagc taccatcacc 2580

gtgaccacag tcgatggggc taaaactgct acctgtgccg tgaccgtaac cgccggaggt 2640

tccaccacc cctgcagtaa tccggtaagc aaaaccctac ctctggtaca ggatggtgcc 2700

ggcgaattca ggttgagtaa tagttttaat taa 2733

<210> 164
<211> 910
<212> PRT
<213> Unknown

<220>
<223> Obtained from environmental sample

<220>
<221> DOMAIN
<222> (5)...(440)
<223> Glycosyl hydrolase family 9

<220>
<221> DOMAIN
<222> (462)...(547)
<223> Cellulose binding domain

<220>
<221> DOMAIN
<222> (709)...(787)
<223> Bacterial Ig-like domain (group 2)

<220>
<221> DOMAIN
<222> (796)...(874)
<223> Bacterial Ig-like domain (group 2)

<220>
<221> DOMAIN
<222> (622)...(700)

<223> Bacterial Ig-like domain (group 2)

<400> 164

Met Gln Thr Tyr Asn Tyr Ala Glu Val Leu Gln Lys Ser Met Phe Phe
1 5 10 15

Tyr Glu Cys Gln Glu Ser Lys Ile Ala Pro Gly Asn Arg Val Thr Trp
20 25 30

Arg Ala Asn Ala Ala Met Asn Asp Gly Ser Asp Val Gly Lys Asp Leu
35 40 45

Thr Gly Gly Trp Phe Asp Ala Gly Asp His Val Lys Phe Asn Phe Pro
50 55 60

Met Ala Phe Thr Ala Thr Ala Leu Ala Trp Gly Ala Ile Asp Phe Ala
65 70 75 80

Gln Gly Tyr Ile Ser Ser Gly Gln Met Gln Tyr Leu Lys Arg Asn Leu
85 90 95

Arg Tyr Val Asn Asp Tyr Phe Ile Lys Cys His Thr Ala Pro Asn Glu
100 105 110

Leu Tyr Gly Gln Val Gly Asn Gly Gly Leu Asp His Ala Phe Trp Gly
115 120 125

Pro Pro Glu Val Met Arg Met Ala Arg Pro Ala Tyr Lys Ile Asp Ala
130 135 140

Ser Lys Pro Gly Ser Asp Leu Ala Ala Glu Thr Ala Ala Ala Met Ala
145 150 155 160

Ala Ala Ser Ile Val Phe Lys Ser Asp Asp Pro Thr Tyr Ser Ala Thr
165 170 175

Leu Leu Asn His Ala Lys Gln Leu Phe Ser Phe Ala Glu Thr Tyr Lys
180 185 190

Gly Lys Tyr Ser Asp Ala Ile Thr Asp Ala Ala Gly Tyr Tyr Asn Ser
195 200 205

Trp Ser Gly Tyr Asn Asp Glu Leu Val Trp Gly Ala Ile Trp Leu Tyr
210 215 220

Arg Ala Thr Gly Asp Ala Thr Tyr Leu Ser Lys Ala Glu Ser Tyr Tyr
225 230 235 240

Asp Asn Leu Gly Asn Gln Gly Gln Glu Pro Val Lys Ala Tyr Lys Trp
245 250 255

Thr Ile Ala Trp Asp Asp Lys Ser Tyr Gly Cys Tyr Ala Leu Leu Ala
260 265 270

Lys Leu Thr Gly Lys Glu Lys Tyr Lys Ile Asp Ala Glu Arg Phe Leu
275 280 285

Asp Tyr Trp Thr Asp Gly Tyr Asn Gly Ser Arg Ile Thr Tyr Thr Pro
290 295 300

Gly Gly Leu Ala Phe Leu Asp Ile Trp Gly Ser Leu Arg Tyr Ala Met
305 310 315 320

Asn Thr Ala Phe Val Ala Ala Tyr Tyr Ala Asp Ala Ala Thr Ser Ala
325 330 335

Ala Lys Thr Thr Lys Tyr Leu Asn Phe Ala Lys Gln Gln Leu His Tyr
340 345 350

Ala Leu Gly Ser Asn Pro Ser Asn Arg Ser Tyr Val Cys Gly Phe Gly
355 360 365

Asn Asn Pro Pro Val Asn Pro His His Arg Gly Ala His Gly Ala Trp
370 375 380

Ser Asn Asn Val Gln Gly Pro Pro Thr Glu Thr Arg His Ile Leu Tyr
385 390 395 400

Gly Ala Leu Val Gly Gly Pro Gly Ser Asn Asp Ser Tyr Thr Asp Asp
405 410 415

Arg Ser Asn Tyr Thr Asn Asn Glu Val Ala Cys Asp Tyr Asn Ala Leu

420	425	430
Phe Ser Gly Leu Leu Ala Lys Phe Val Ile Asp Tyr Gly Gly Thr Pro		
435	440	445
Leu Ala Asn Phe Pro Val Arg Glu Thr Pro Lys Asp Glu Tyr Phe Val		
450	455	460
Glu Ala Lys Ala Asn Ala Thr Gly Thr Asn Phe Ser Glu Trp Thr Val		
465	470	475
Trp Val Tyr Asn His Thr Ala Trp Pro Ala Arg Glu Gly Ser Glu Tyr		
485	490	495
Lys Phe Arg Leu Tyr Val Asn Ile Ser Glu Gly Leu Ala Ala Gly Tyr		
500	505	510
Thr Ala Ser Asn Tyr Val Val Gln Thr Asn Asn Ala Gly Val Val Asn		
515	520	525
Phe Thr Gln Leu Leu Ala Ala Asp Ala Ala Asn Gly Ile Tyr Tyr Thr		
530	535	540
Glu Val Thr Phe Lys Pro Gly Thr Glu Ile Tyr Pro Gly Gly Gln Gln		
545	550	555
Tyr Asp Lys Lys Glu Ala Gln Met Arg Ile Ser Leu Pro Asn Ala Pro		
565	570	575
Ala Ser Ala Trp Asp Pro Thr Asn Asp Pro Ser Trp Ala Gly Ile Thr		
580	585	590
Ser Thr Leu Lys Gln Met Pro Gly Ile Pro Met Tyr Val Asp Gly Val		
595	600	605
Lys Val Phe Gly Asn Glu Pro Val Pro Gly Gln Thr Val Pro Val Thr		
610	615	620
Gly Val Thr Val Ser Pro Thr Thr Leu Ser Leu Thr Val Gly Gln Thr		
625	630	635
		640

Ser Thr Leu Thr Ala Thr Val Ser Pro Ala Asn Ala Thr Asn Lys Asn
645 650 655

Val Thr Trp Ser Ser Ser Asn Thr Ser Val Ala Thr Val Ser Ser Thr
660 665 670

Gly Val Val Thr Ala Val Ala Ala Gly Ser Ala Thr Ile Thr Val Thr
675 680 685

Thr Val Asp Gly Ala Lys Thr Ala Thr Cys Ala Val Thr Val Thr Gly
690 695 700

Ser Thr Asn Val Pro Val Thr Gly Val Thr Val Ser Pro Thr Thr Leu
705 710 715 720

Ser Leu Thr Val Gly Gln Thr Ala Thr Leu Thr Ala Thr Val Ser Pro
725 730 735

Ala Asn Ala Thr Asn Lys Asn Val Thr Trp Ser Ser Ser Asn Thr Ser
740 745 750

Val Ala Thr Val Ser Ser Thr Gly Val Val Thr Ala Val Ala Ala Gly
755 760 765

Ser Ala Thr Ile Thr Val Thr Thr Val Asp Gly Ala Lys Thr Ala Thr
770 775 780

Cys Thr Val Thr Val Thr Gly Ser Thr Thr Val Pro Val Thr Gly Val
785 790 795 800

Thr Val Ser Pro Thr Thr Leu Ser Leu Thr Val Gly Gln Thr Ala Thr
805 810 815

Leu Thr Ala Thr Val Ser Pro Ala Asp Ala Thr Asn Lys Asn Val Thr
820 825 830

Trp Ser Ser Ser Asn Thr Ser Val Ala Thr Val Ser Ser Thr Gly Val
835 840 845

Val Thr Ala Val Ala Ala Gly Ser Ala Thr Ile Thr Val Thr Thr Val
850 855 860

Asp Gly Ala Lys Thr Ala Thr Cys Ala Val Thr Val Thr Ala Gly Gly
865 870 875 880

Ser Thr Thr Pro Cys Ser Asn Pro Val Ser Lys Thr Leu Pro Leu Val
885 890 895

Gln Asp Gly Ala Gly Glu Phe Arg Leu Ser Asn Ser Phe Asn
900 905 910

<210> 165

<211> 1347

<212> DNA

<213> Unknown

<220>

<223> Obtained from environmental sample

<400> 165

atgacaatta acaacaaaac tacagcgagt cctagtattc ccagcaccca caattccctc 60

ccgtcgcttc gcacactgtt taccaccagc ctgctcacgc tggccctgac cgcctgcggt 120

ggtttcttcca gcagcgacaa ggacccttca agtccagct ccagtgaatc atcaagtcc 180

agcgaatcct cgagctcagc ttccagcgaa tctctgagca gtgagtcag cagtagctct 240

tccgcgggcc atttctccat cgagccggac ttccagctct acagcctggc caacttcccg 300

gtgggcgtgg cggtctccgc cgccaacgag aacgacagca tcttcaacag tccggatgcc 360

gccgaacgtc aggccgttat tattgagcac ttctctcagc tcaccgccgg caacatcatg 420

aaaatgagct acctgcagcc gagtcaaggc aacttcacct tcgatgacgc cgacgagttg 480

gttaacttcg cccaagccaa tggcatgacc gtacacggcc actccacat ctggcacgcg 540

gactaccaag taccgaactt catgagaaac tttgaagggt accaggagga atgggcagaa 600

attctgaccg atcacgtcac taccatcatc gagcacttcc ccgacgatgt ggtcatcagc 660

tgggacgtgg tgaacgaggc tgtcgatcaa ggcacggcga acggctggcg ccattcggtg 720

ttctacaatg cattcgacgc cccggaagaa ggcgacattc ccgaatacat caaagtcgct 780

ttccgcgcgc cgcgcgaggc tgacgccaac gtagacctct actacaacga ctacgacaat 840

accgccaatg cccagcgctt ggccaaaaca ctgcaaattg ccgaggtact ggacgccgaa 900

ggcaccattg acggcgctcg tttccagatg cagcctaca tggattaccg gaccccgacc 960

cattttgaaa acgccttccg gcaagtcgtc gacctggggc tcaaagtga agttaccgag 1020

ctggacgtat ccgtagtcaa cccctacggc ggcggaagcac ctccacaacc ggaatacgac 1080

aaagaactgg ccggcgcgca aaaactgcgc ttctgccaaa tcgccgaagt ttacatgaac 1140

actgtacccg aggagttacg cgggtggcttc accgtctggg gcctgaccga tgatgaaagt 1200

tggtgatgc aacagttcag aaacgccacc ggcgccgact acgacgacgt ctggccgtta 1260

ctgttcaatg ccgacaaatc cgccaaaccg gcaactgcaag gcgtggccga cgcctttacc 1320

ggacaaacct gcacctccga gttctaa 1347

<210> 166
 <211> 448
 <212> PRT
 <213> Unknown

<220>
 <223> Obtained from environmental sample

<220>
 <221> SIGNAL
 <222> (1)...(45)

<400> 166
 Met Thr Ile Asn Asn Lys Thr Thr Ala Ser Pro Ser Ile Pro Ser Thr
 1 5 10 15

His Asn Ser Leu Pro Ser Leu Arg Thr Leu Phe Thr Thr Ser Leu Leu
20 25 30

Thr Leu Ala Leu Thr Ala Cys Gly Gly Ser Ser Ser Ser Asp Lys Asp
35 40 45

Pro Ser Ser Ser Ser Ser Ser Glu Ser Ser Ser Ser Ser Glu Ser Ser
50 55 60

Ser Ser Ala Ser Ser Glu Ser Ser Ser Ser Glu Ser Ser Ser Ser Ser
65 70 75 80

Ser Ala Gly His Phe Ser Ile Glu Pro Asp Phe Gln Leu Tyr Ser Leu
85 90 95

Ala Asn Phe Pro Val Gly Val Ala Val Ser Ala Ala Asn Glu Asn Asp
100 105 110

Ser Ile Phe Asn Ser Pro Asp Ala Ala Glu Arg Gln Ala Val Ile Ile
115 120 125

Glu His Phe Ser Gln Leu Thr Ala Gly Asn Ile Met Lys Met Ser Tyr
130 135 140

Leu Gln Pro Ser Gln Gly Asn Phe Thr Phe Asp Asp Ala Asp Glu Leu
145 150 155 160

Val Asn Phe Ala Gln Ala Asn Gly Met Thr Val His Gly His Ser Thr
165 170 175

Ile Trp His Ala Asp Tyr Gln Val Pro Asn Phe Met Arg Asn Phe Glu
180 185 190

Gly Asp Gln Glu Glu Trp Ala Glu Ile Leu Thr Asp His Val Thr Thr
195 200 205

Ile Ile Glu His Phe Pro Asp Asp Val Val Ile Ser Trp Asp Val Val
210 215 220

Asn Glu Ala Val Asp Gln Gly Thr Ala Asn Gly Trp Arg His Ser Val
225 230 235 240

Phe Tyr Asn Ala Phe Asp Ala Pro Glu Glu Gly Asp Ile Pro Glu Tyr
245 250 255

Ile Lys Val Ala Phe Arg Ala Ala Arg Glu Ala Asp Ala Asn Val Asp
260 265 270

Leu Tyr Tyr Asn Asp Tyr Asp Asn Thr Ala Asn Ala Gln Arg Leu Ala
275 280 285

Lys Thr Leu Gln Ile Ala Glu Val Leu Asp Ala Glu Gly Thr Ile Asp
290 295 300

Gly Val Gly Phe Gln Met His Ala Tyr Met Asp Tyr Pro Ser Leu Thr
305 310 315 320

His Phe Glu Asn Ala Phe Arg Gln Val Val Asp Leu Gly Leu Lys Val
325 330 335

Lys Val Thr Glu Leu Asp Val Ser Val Val Asn Pro Tyr Gly Gly Glu
340 345 350

Ala Pro Pro Gln Pro Glu Tyr Asp Lys Glu Leu Ala Gly Ala Gln Lys
355 360 365

Leu Arg Phe Cys Gln Ile Ala Glu Val Tyr Met Asn Thr Val Pro Glu
370 375 380

Glu Leu Arg Gly Gly Phe Thr Val Trp Gly Leu Thr Asp Asp Glu Ser
385 390 395 400

Trp Leu Met Gln Gln Phe Arg Asn Ala Thr Gly Ala Asp Tyr Asp Asp
405 410 415

Val Trp Pro Leu Leu Phe Asn Ala Asp Lys Ser Ala Lys Pro Ala Leu
420 425 430

Gln Gly Val Ala Asp Ala Phe Thr Gly Gln Thr Cys Thr Ser Glu Phe
435 440 445

??

??

??

??

1

2