(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2020년06월02일
(11) 등록번호 10-2117225
(24) 등록일자 2020년05월26일

(51) 국제특허분류(Int. Cl.)
C12P 7/64 (2006.01) *A23L 33/115* (2016.01)
C07C 69/52 (2006.01) *C10L 1/02* (2006.01)
C12N 15/52 (2006.01) *C12N 15/79* (2006.01)
C12N 9/02 (2006.01) *C12N 9/10* (2006.01)
C12N 9/16 (2006.01)
(52) CPC특허분류
C12P 7/6463 (2013.01)
A23L 33/115 (2016.08)
(21) 출원번호 10-2019-7008914(분할)
(22) 출원일자(국제) 2012년02월02일
심사청구일자 2019년04월26일
(85) 번역문제출일자 2019년03월27일
(65) 공개번호 10-2019-0036571
(43) 공개일자 2019년04월04일
(62) 원출원 특허 10-2013-7023181
원출원일자(국제) 2012년02월02일
심사청구일자 2017년02월01일
(86) 국제출원번호 PCT/US2012/023696
(87) 국제공개번호 WO 2012/106560
국제공개일자 2012년08월09일
(30) 우선권주장
61/438,969 2011년02월02일 미국(US)
(뒷면에 계속)
(56) 선행기술조사문헌
WO20100063032 A2*
(뒷면에 계속)
전체 청구항 수 : 총 22 항

(73) 특허권자
테라비아 홀딩스 인코포레이티드
미합중국, 캘리포니아 94080, 사우쓰 샌프란시스코, 게이트웨이 블러바드 225
(72) 발명자
프랭클린, 스코트
미국 94080 캘리포니아 사우스 샌프란시스코 게이트웨이 블러바드 225
소만치, 아라빈드
미국 94080 캘리포니아 사우스 샌프란시스코 게이트웨이 블러바드 225
(뒷면에 계속)
(74) 대리인
특허법인 남앤남

심사관 : 황상필

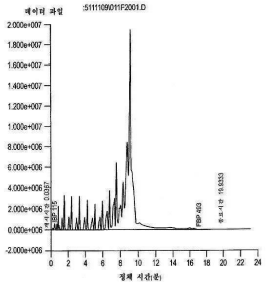
(54) 발명의 명칭 **재조합 유지성 미생물로부터 생산된 맞춤 오일**

(57) 요 약

오일, 연료, 함유 화학 제품 및 기타 다른 화합물을 재조합 미생물, 예를 들어 오일 함유 미생물 내에서 생산하기 위한 방법 및 조성물, 그리고 이와 같은 미생물을 저비용으로 배양하는 방법이 제공된다. 예를 들어 리파제, 수크로스 운반체, 수크로스 인버타제, 프럭토키나제, 다당류 분해 효소, 케토 아실-ACP 합성 효소, 지방 아실-

(뒷면에 계속)

대 표 도



ACP 티오에스테라제, 지방 아실-CoA/알데하이드 환원 효소, 지방 아실-CoA 환원 효소, 지방 알데하이드 환원 효소, 지방산 하이드록실라제, 불포화 효소, 지방 알데하이드 탈카보틸화 효소 및/또는 아실기 운반 단백질을 암호화하는 외인성 유전자를 함유하는 미세 조류 세포는, 교통 수단의 연료, 예를 들어 재생 디젤, 바이오디젤 및 재생 제트 연료, 그리고 함유 화학 제품, 예를 들어 기능성 유체, 계면 활성제, 비누 및 윤활제를 제조하는데 유용하다.

(52) CPC특허분류
C07C 69/52 (2013.01)
C10L 1/02 (2013.01)
C12N 15/52 (2013.01)
C12N 15/79 (2013.01)
C12N 9/0071 (2013.01)
C12N 9/1029 (2013.01)
C12N 9/16 (2013.01)
C10L 2200/0476 (2013.01)
C10L 2290/26 (2013.01)

(72) 발명자
위, 자니스
미국 94080 캘리포니아 사우스 샌프란시스코 게이트웨이 블러바드 225
루덴코, 죠지
미국 94080 캘리포니아 사우스 샌프란시스코 게이트웨이 블러바드 225
모즐리, 제프리
미국 94080 캘리포니아 사우스 샌프란시스코 게이트웨이 블러바드 225
라키트스키, 왈트
미국 94080 캘리포니아 사우스 샌프란시스코 게이트웨이 블러바드 225
자오, 신후아
미국 94080 캘리포니아 사우스 샌프란시스코 게이트웨이 블러바드 225
바트, 리야즈
미국 94080 캘리포니아 사우스 샌프란시스코 게이트웨이 블러바드 225

(56) 선행기술조사문헌
WO2002008403 A2*
WO20000061740 A1*
WO2009076559 A1
JP2010528627 A
WO2010063032 A2*
*는 심사관에 의하여 인용된 문헌

(30) 우선권주장
61/476,691 2011년04월18일 미국(US)
61/484,458 2011년05월10일 미국(US)
61/548,616 2011년10월18일 미국(US)

명 세 서

청구범위

청구항 1

트리아실글리세라이드를 포함하는 오일, 또는 상기 오일로부터 생산되는 생성물을 제조하는 방법으로서, 이 방법은,

프로토테카 또는 클로렐라 속의 재조합 유지성 미세 조류의 세포를 배양하는 단계로서, 상기 세포는 활성 β-케토아실-ACP 합성 효소 II를 암호화하는 제1 외인성 유전자에 의해 암호화되는 효소를 발현하도록 작동가능한 하나 이상의 외인성 유전자를 포함하고, 임의로 상기 세포는 지방 아실-ACP 티오에스테라제 또는 불포화 효소를 암호화하는 제2 외인성 유전자를 추가로 포함하는 단계, 및

상기 세포로부터 오일을 회수하는 단계를 포함하고,

임의로 식품, 연료 또는 화학 생성물을 생성하기 위하여 상기 오일을 추가로 가공하는 단계를 포함하며,

상기 오일은 상기 하나 이상의 재조합 유전자로 인하여 변형된 지방산 프로필을 가지는 것인 방법.

청구항 2

제1항에 있어서, 외인성 불포화 효소는 스테아로일-ACP 불포화 효소(stearoyl-ACP desaturase (SAD)) 또는 지방 아실 불포화 효소(fatty acyl desaturase (FAD))이고, 또는 외인성 지방 아실-ACP 티오에스테라제는 FATA인 방법.

청구항 3

제1항 또는 제2항에 있어서, 상기 하나 이상의 외인성 유전자는 세포의 핵 게놈의 동종성 재조합 부위 내로 통합되고, 상기 통합은 상기 동종성 재조합 부위에 의해 암호화되는 내인성 유전자 생성물의 발현을 감소시키거나 제거하거나, 또는 상기 내인성 유전자의 발현이 RNA 유도 기작에 의해 감소되는 것인 방법.

청구항 4

제3항에 있어서, 상기 동종성 재조합 부위는 내인성 지방 아실-ACP 티오에스테라제 또는 내인성 불포화 효소를 암호화하고, 임의로 상기 내인성 불포화 효소는 스테아로일-ACP 불포화 효소(SAD) 또는 지방 아실 불포화 효소(FAD)인 방법.

청구항 5

제1항 또는 제2항에 있어서, 미세 조류는 프로토테카 속에서 선택되는 것인 방법.

청구항 6

제5항에 있어서, 미세 조류는 프로토테카 윅커헤미, 프로토테카 스태그노라, 프로토테카 포르토리센시스, 프로토테카 모리포르미스 또는 프로토테카 조프피인 방법.

청구항 7

제6항에 있어서, 미세 조류는 프로토테카 모리포르미스인 방법.

청구항 8

재조합 유지성 미세 조류의 세포로서, 상기 세포는 활성 β-케토아실-ACP 합성 효소 II를 암호화하는 제1 외인성 유전자에 의해 암호화되는 효소를 발현하도록 작동가능한 하나 이상의 외인성 유전자를 포함하고, 임의로 상기 세포는 지방 아실-ACP 티오에스테라제 또는 불포화 효소를 암호화하는 제2 외인성 유전자를 추가로 포함하며, 상기 세포는 하나 이상의 재조합 유전자의 발현으로 인하여 비-재조합 미세조류 세포의 지방산 프로필과 비교할 때 변형된 지방산 프로필을 가지는 것인, 세포.

청구항 9

제8항에 있어서, 외인성 불포화 효소는 스테아로일-ACP 불포화 효소(SAD) 또는 지방 아실 불포화 효소(FAD)이고, 또는 외인성 지방 아실-ACP 티오에스테라제는 FATA인 세포.

청구항 10

제8항 또는 제9항에 있어서, 상기 하나 이상의 외인성 유전자는 세포의 핵 게놈의 동종성 재조합 부위 내로 통합되고, 상기 통합은 상기 동종성 재조합 부위에 의해 암호화되는 내인성 유전자 생성물의 발현을 감소시키거나 제거하거나, 또는 상기 내인성 유전자의 발현이 RNA 유도 기작에 의해 감소되는 것인 세포.

청구항 11

제10항에 있어서, 상기 동종성 재조합 부위는 내인성 지방 아실-ACP 티오에스테라제 또는 내인성 불포화 효소를 암호화하고, 임의로 상기 내인성 불포화 효소는 스테아로일-ACP 불포화 효소(SAD) 또는 지방 아실 불포화 효소(FAD)인 세포.

청구항 12

제8항 또는 제9항에 있어서, 미세 조류는 프로토테카 속에서 선택되는 것인 세포.

청구항 13

제12항에 있어서, 미세 조류는 프로토테카 윅커해미, 프로토테카 스태그노라, 프로토테카 포르토리센시스, 프로토테카 모리포르미스 또는 프로토테카 조프피인 세포.

청구항 14

제13항에 있어서, 미세 조류는 프로토테카 모리포르미스인 세포.

청구항 15

하기의 단계들을 포함하는 제조 방법에 의해 생산되는, 트리아실글리세라이드를 포함하는 오일:

재조합 유지성 미세 조류의 세포를 배양하는 단계로서, 상기 세포는 활성 β-케토아실-ACP 합성 효소 II를 암호화하는 제1 외인성 유전자에 의해 암호화되는 효소를 발현하도록 작동가능한 하나 이상의 외인성 유전자를 포함하고, 임의로 상기 세포는 지방 아실-ACP 티오에스테라제 또는 불포화 효소를 암호화하는 제2 외인성 유전자를 추가로 포함하는 단계, 및

상기 세포로부터 오일을 회수하는 단계를 포함하고,

임의로 식품, 연료 또는 화학 생성물을 생성하기 위하여 상기 오일을 추가로 가공하는 단계를 포함하며,

상기 오일은 하나 이상의 재조합 유전자로 인하여 변형된 지방산 프로필을 가지는 것인 오일.

청구항 16

제15항에 있어서, 외인성 불포화 효소는 스테아로일-ACP 불포화 효소(SAD) 또는 지방 아실 불포화 효소(FAD)이고, 또는 외인성 지방 아실-ACP 티오에스테라제는 FATA인 오일.

청구항 17

제15항 또는 제16항에 있어서, 상기 하나 이상의 외인성 유전자는 세포의 핵 게놈의 동종성 재조합 부위 내로 통합되고, 상기 통합은 상기 동종성 재조합 부위에 의해 암호화되는 내인성 유전자 생성물의 발현을 감소시키거나 제거하거나, 또는 상기 내인성 유전자의 발현이 RNA 유도 기작에 의해 감소되는 것인 오일.

청구항 18

제17항에 있어서, 상기 동종성 재조합 부위는 내인성 지방 아실-ACP 티오에스테라제 또는 내인성 불포화 효소를 암호화하고, 임의로 상기 내인성 불포화 효소는 스테아로일-ACP 불포화 효소(SAD) 또는 지방 아실 불포화 효소(FAD)인 오일.

청구항 19

제15항 또는 제16항에 있어서, 미세 조류는 프로토테카 속에서 선택되는 것인 오일.

청구항 20

제19항에 있어서, 미세 조류는 프로토테카 윅커해미, 프로토테카 스태그노라, 프로토테카 포르토리센시스, 프로토테카 모리포르미스 또는 프로토테카 조프피인 오일.

청구항 21

제20항에 있어서, 미세 조류는 프로토테카 모리포르미스인 오일.

청구항 22

식품, 연료 또는 화학 생성물을 생성하기 위하여 제15항의 오일을 추가로 가공함으로써 제조된 생성물.

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

삭제

청구항 38

삭제

청구항 39

삭제

청구항 40

삭제

청구항 41

삭제

청구항 42

삭제

청구항 43

삭제

청구항 44

삭제

청구항 45

삭제

청구항 46

삭제

청구항 47

삭제

청구항 48

삭제

청구항 49

삭제

청구항 50

삭제

청구항 51

삭제

청구항 52

삭제

청구항 53

삭제

청구항 54

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

발명의 설명

기 술 분 야

[0001] 관련 출원의 상호 참조

[0002] 본 출원은 35 U.S.C. 119(e)에 따라서 2011년 2월 2일자로 출원된 미국 가출원 제61/438,969호, 2011년 4월 18일자로 출원된 미국 가출원 제61/476,691호, 2011년 5월 10일자로 출원된 미국 가출원 제61/484,458호, 및 2011년 10월 18일자로 출원된 미국 가출원 제61/548,616호의 이익을 주장한다. 이들 각각의 출원은 모든 목적을 위해 본원에 전체가 참조로 포함되어 있다.

[0003] 서열목록에 대한 참조

[0004] 본 출원은 상세한 설명의 마지막에 나타낸 서열 목록을 포함한다.

[0005] 기술분야

[0006] 본 발명은 미생물로부터 만들어진 오일, 연료 및 함유 화학 제품의 제조에 관한 것이다. 특히, 본원은 오일 함유 미세 조류; 지질, 지방산 에스테르, 지방산, 알데하이드, 알코올 및 알칸을 포함하는 유용한 화합물의 제조를 위해 상기 미세 조류를 배양하는 방법; 및 상기 미세 조류를 유전적으로 변경시켜 제조 효율을 개선시키고 상기 미세 조류에 의해 생성된 오일의 유형과 조성을 변경시키는 방법 및 시약에 관한 것이다.

배 경 기 술

[0007] 화석 연료는 썩은 식물 및 동물로부터 형성되어 수억년 이상 동안 지각에서 열 및 압력에 노출됨에 의해 원유, 석탄, 천연 가스 또는 중유로 전환된, 매장된 유기 물질의 연소성 지질학적 퇴적물에 대한 일반적인 용어이다. 화석 연료는 한정된 재생가능하지 않은 자원이다. 세계 경제의 에너지에 대한 수요 증가는 또한 탄화수소 비용에 대한 가중된 압력을 부여하였다. 에너지외에도 플라스틱 및 화학 물질 제조업체를 포함하는 많은 산업은 이들의 제조 공정을 위한 공급원료로서 탄화수소의 유용성에 크게 의존한다. 현재 공급 자원에 대한 비용 효율이 높은 대용물이 에너지 및 이들의 원료 비용에 대한 상승된 압력을 완화시키는데 일조할 수 있었다.

[0008] PCT 공개 제2008/151149호는 오일 제조용의 미세 조류를 배양하기 위한 방법 및 재료를 기술하고 있고 특히 미세 조류 클로렐라 프로토테코이데스(*Chlorella protothecoides*)에 의해 생성된 오일로부터 디젤 연료의 제조를 예시하고 있다. 연료, 화학 물질, 식품 및 기타 다른 용도를 위한 개선된 오일 제조 방법, 특히 단쇄의 길이 및 더 높은 포화도를 갖고 안료를 포함하지 않는 오일을 보다 높은 수율 및 효율로 제조하는 방법이 여전히 요구되고 있다. 본 발명은 상기 필요성을 충족시킨다.

[0009] 폴리우레탄은 카바메이트(우레탄) 결합을 포함하는 화합물이다. 통상적으로, 폴리우레탄은 유기 단위들로 이루어진 중합체이다. 이 중합체는 이소시아네이트부를 포함하는 제1 유기 단위($C(O)N-R_1-NC(O)$) 및 하이드록실기를 포함하는 제2 유기 단위($HO-R_2-OH$)의 반응에 의해 제조된다. 폴리우레탄은 화학식 $-[C(O)NH-R_1-NHC(O)-O-R_2-O]_m-$(식 중, 아래 첨자 m은 중합체 중에 함유된 단량체의 수를 나타내는 수임)으로서 표시될 수 있다. R_1 및 R_2는 동일하거나 상이할 수 있지만, 통상적으로는 상이하다. 폴리우레탄은 다수의 상이한 분야, 예를 들어 가요성 재료 및 경직성 재료에 사용된다. 폴리우레탄은 신발, 자동차, 비행기, 부싱(bushing), 개스킷, 접착제, 양탄자, 스판덱스 섬유 및 전자 제품의 하우징 등에 사용된다.

발명의 내용

해결하려는 과제

[0010] 트리아실글리세라이드를 포함하는 천연 오일 또는 상기 천연 오일로부터 생성된 생성물의 제조 방법을 제공한다.

과제의 해결 수단

[0011] 본 발명의 예시적인 구체예는 변경된 글리세로지질 프로필을 생산하는 유지성 세포와, 이 세포로부터 생산되는 생성물을 제공한다. 유지성 세포의 예로서는 제II형 지질 생합성 경로를 가지는 미생물 세포를 포함한다. 구체예는 또한 이 세포를 사용하여 얻을 수 있는 천연 오일을 특징으로 한다. 구체예는 단백질, 예를 들어 지방 아

실-ACP 티오에스테라제를 암호화하는 외인성 유전자를 발현하는 재조합 세포를 포함한다. 본 발명은 또한 상기 세포로부터 생성되는 지질 및 오일계 생성물, 예를 들어 연료, 예를 들어 바이오디젤, 재생 디젤 및 제트 연료, 식품용 오일 및 화학 물질을 제공한다.

[0012] 제1 양태에서, 본 발명은 C8:0이 3% 이상인 지질 프로필을 가지는 미세 조류 세포를 제공한다. 몇몇 경우에 있어서, 지질 프로필은 C8:0이 12% 이상이다. 몇몇 구체예에서, 세포는 재조합 세포이다. 몇몇 경우에 있어서, 재조합 세포는 사슬 길이가 C8인 지방 아실-ACP 기질에 대해 가수 분해 활성을 가지는 아실-ACP 티오에스테라제 단백질을 암호화하는 외인성 유전자를 포함한다. 몇몇 구체예에서, 외인성 유전자는 쿠페아 팔루스트리스(*Cuphea palustris*) 아실-ACP 티오에스테라제를 암호화한다. 몇몇 경우에 있어서, 세포는 프로토테카(*Prototheca*) 세포이다. 몇몇 경우에 있어서, 세포는 표 1에 제시된 미세 조류들로부터 선택되는 미세 조류 속 또는 종에 속하는 것이다.

[0013] 제2 양태에서, 본 발명은 C10:0이 4% 이상인 지질 프로필을 가지는 미세 조류 세포를 제공한다. 몇몇 경우에 있어서, 지질 프로필은 C10:0이 18% 이상이다. 몇몇 경우에 있어서, 지질 프로필은 C10:0이 20% 이상이다. 몇몇 경우에 있어서, 미세 조류 세포의 지질 내용물은 C12:0을 추가로 포함한다. 몇몇 경우에 있어서, C10:0 대 C12:0의 비율은 3:1 이상이다. 몇몇 경우에 있어서, 미세 조류 세포의 지질 내용물은 C14:0을 추가로 포함한다. 몇몇 경우에 있어서, C10:0 대 C14:0의 비율은 10:1 이상이다. 몇몇 구체예에서, 세포는 재조합 세포이다. 몇몇 경우에 있어서, 재조합 세포는 사슬 길이가 C10인 지방 아실-ACP 기질에 대해 가수 분해 활성을 가지는 지방 아실-ACP 티오에스테라제 단백질을 암호화하는 외인성 유전자를 포함한다. 몇몇 구체예에서, 외인성 유전자는 쿠페아 후커리아나(*Cuphea hookeriana*) 및 얼머스 아메리카나(*Ulmus americana*)로 이루어진 군으로부터 선택되는 종으로부터 유래하는 지방 아실-ACP 티오에스테라제 단백질을 암호화한다. 몇몇 구체예에서, 지방 아실-ACP 티오에스테라제 단백질은 쿠페아 후커리아나 및 얼머스 아메리카나로 이루어진 군으로부터 선택되는 종으로부터 유래한다. 몇몇 경우에 있어서, 지방 아실-ACP 티오에스테라제 유전자는 사슬 길이가 C10인 지방 아실-ACP 기질에 대해 가수 분해 활성을 가지는 쿠페아 후커리아나 및 얼머스 아메리카나로부터 유래하는 지방 아실-ACP 티오에스테라제 유전자로 이루어진 군으로부터 선택된다.

[0014] 몇몇 경우에 있어서, 세포는 프로토테카 세포이다. 몇몇 구체예에서, 세포는 표 1에 제시된 미세 조류들로부터 선택되는 미세 조류 속 또는 종에 속하는 것이다.

[0015] 제3 양태에서, 본 발명은 C12:0이 13% 이상인 지질 프로필을 가지는 미세 조류 세포를 제공한다. 몇몇 경우에 있어서, 세포는 재조합 세포이다. 몇몇 구체예에서, 재조합 세포는 사슬 길이가 C12인 지방 아실-ACP 기질에 대해 가수 분해 활성을 가지는 지방 아실-ACP 티오에스테라제 단백질을 암호화하는 외인성 유전자를 포함한다. 몇몇 구체예에서, 지방 아실-ACP 티오에스테라제 단백질은 움벨루라리아 캘리포니카(*Umbellularia californica*) 및 신나모뭄 캠포라(*Cinnamomum camphora*)로 이루어진 군으로부터 선택되는 종으로부터 유래한다. 몇몇 경우에 있어서, 지방 아실-ACP 티오에스테라제 유전자는 사슬 길이가 C12인 지방 아실-ACP 기질에 대해 가수 분해 활성을 가지는, 움벨루라리아 캘리포니카 및 신나모뭄 캠포라로부터 유래하는 지방 아실-ACP 티오에스테라제 유전자로 이루어진 군으로부터 선택된다. 몇몇 구체예에서, 세포는 프로토테카 세포이다.

[0016] 제4 양태에서, 본 발명은 C14:0이 10% 이상인 지질 프로필을 가지는 미세 조류 세포를 제공한다. 몇몇 경우에 있어서, 지질 프로필은 C14:0이 35% 이상이다. 몇몇 경우에 있어서, 미세 조류 세포의 지질 내용물은 C12:0을 추가로 포함한다. 몇몇 경우에 있어서, C14:0 대 C12:0의 비율은 3:1 이상이다. 몇몇 경우에 있어서, 세포는 재조합 세포이다. 몇몇 구체예에서, 재조합 세포는 사슬 길이가 C14인 지방 아실-ACP 기질에 대해 가수 분해 활성을 가지는 지방 아실-ACP 티오에스테라제 단백질을 암호화하는 외인성 유전자를 포함한다. 몇몇 구체예에서, 지방 아실-ACP 티오에스테라제 단백질은 신나모뭄 캠포라 및 얼머스 아메리카나로 이루어진 군으로부터 선택되는 종으로부터 유래한다. 몇몇 경우에 있어서, 지방 아실-ACP 티오에스테라제 유전자는 사슬 길이가 C14인 지방 아실-ACP 기질에 대해 가수 분해 활성을 가지는 신나모뭄 캠포라 및 얼머스 아메리카나로부터 유래하는 지방 아실-ACP 티오에스테라제 유전자로 이루어진 군으로부터 선택된다. 몇몇 경우에 있어서, 세포는 프로토테카 세포이다. 몇몇 구체예에서, 세포는 표 1에 제시된 미세 조류들로부터 선택되는 미세 조류 속 또는 종에 속하는 것이다.

[0017] 제5 양태에서, 본 발명은 C16:0이 15% 이상인 지질 프로필을 가지는 미세 조류 세포를 제공한다. 몇몇 경우에 있어서, 지질 프로필은 C16:0이 39% 이상이다. 몇몇 경우에 있어서, 지질 프로필은 C16:0이 67% 이상이다. 몇몇 경우에 있어서, 세포는 재조합 세포이다. 몇몇 구체예에서, 재조합 세포는 사슬 길이가 C16인 지방 아실-ACP 기질에 대해 가수 분해 활성을 가지는 지방 아실-ACP 티오에스테라제 단백질을 암호화하는 외인성 유전자를 포함

한다. 몇몇 구체예에서, 지방 아실-ACP 티오에스테라제 단백질은 쿠페아 후커리아나 및 얼머스 아메리카나로부터 선택되는 종으로부터 유래한다. 몇몇 구체예에서, 재조합 세포는 사슬 길이가 C16인 지방 아실-ACP 기질에 대해 가수 분해 활성을 가지는, 쿠페아 후커리아나 및 얼머스 아메리카나로 이루어진 군으로부터 선택되는 종으로부터 유래하는 지방 아실-ACP 티오에스테라제 단백질을 암호화하는 외인성 유전자를 포함한다. 몇몇 경우에 있어서, 세포는 프로토테카 세포이다. 몇몇 구체예에서, 미세 조류 세포는 내인성 불포화 효소 유전자를 추가로 포함하는데, 여기서 상기 내인성 불포화 효소 유전자는 돌연 변이되지 않은 불포화 효소보다 활성이 작거나 불활성인 불포화 효소를 암호화하도록 돌연 변이되었거나, 또는 상기 내인성 불포화 효소는 미세 조류 세포 게놈으로부터 결실되었다.

[0018] 제6 양태에서, 본 발명은 지질 프로필이 포화 지방산 60% 이상인 미세 조류 세포를 제공한다. 몇몇 경우에 있어서, 미세 조류 세포는 지질 프로필이 포화 지방산 85% 이상이다. 몇몇 경우에 있어서, 세포는 재조합 세포이다. 몇몇 구체예에서, 재조합 세포는 사슬 길이가 C10 내지 C16인 지방 아실-ACP 기질에 대해 가수 분해 활성을 가지는 지방 아실-ACP 티오에스테라제 단백질을 암호화하는 외인성 유전자를 포함한다. 몇몇 경우에 있어서, 세포는 프로토테카 세포이다.

[0019] 제7 양태에서, 본 발명은 C18:0이 19% 이상인 지질 프로필을 가지는 미세 조류 세포를 제공한다. 몇몇 경우에 있어서, 지질 프로필은 C18:0 27% 이상이다. 몇몇 경우에 있어서, 세포는 재조합 세포이다, 몇몇 구체예에서, 재조합 세포는 사슬 길이가 C18인 지방 아실-ACP 기질에 대해 가수 분해 활성을 가지는 지방 아실-ACP 티오에스테라제 단백질을 암호화하는 외인성 유전자를 포함한다. 몇몇 구체예에서, 지방 아실-ACP 티오에스테라제 단백질은 브라시카 나푸스(*Brassica napus*)로부터 선택되는 종으로부터 유래한다. 몇몇 구체예에서, 재조합 세포는 사슬 길이가 C16인 지방 아실-ACP 기질에 대해 가수 분해 활성을 가지는 브라시카 나푸스로부터 유래하는 지방 아실-ACP 티오에스테라제 단백질을 암호화하는 외인성 유전자를 포함한다.

[0020] 제8 양태에서, 본 발명은 쿠페아 후커리아나, 움벨루라리아 캘리포니카, 신나모뭄 캠포라, 쿠페아 팔루스트리스, 쿠페아 란세올라타(*Cuphea lanceolata*), 아이리스 게르마니카(*Iris germanica*), 미리스티카 프래그란스(*Myristica fragrans*), 가르시니아 망고스타나(*Garcinia mangostana*), 엘라에이스 기니엔시스(*Elaeis guiniensis*), 브라시카 나푸스, 리시누스 커뮤니스(*Ricinus communis*) 및 얼머스 아메리카나로 이루어진 군으로부터 유래하는 지방 아실-ACP 티오에스테라제 단백질을 암호화하는 외인성 유전자를 포함하는 미세 조류 세포를 제공한다.

[0021] 제9 양태에서, 본 발명은 발현 구조물을 포함하는 미세 조류 세포를 제공하는데, 여기서 상기 발현 구조물은, 내인성 유전자가 돌연 변이되지 않은 유전자 생성물보다 활성이 작거나 불활성인 유전자 생성물을 암호화하도록 돌연 변이된 것, 내인성 유전자가 미세 조류 세포 게놈으로부터 결실된 것 및 RNA 유도 기작을 통한 것으로 이루어진 방법으로부터 선택된 내인성 유전자의 발현을 하향 조절한다. 몇몇 경우에 있어서, 상기 방법은 RNA 유도 기작, 예를 들어 RNAi, 안티센스 및/또는 dsRNA이다. 몇몇 경우에 있어서, 내인성 유전자는 불포화 효소 유전자이다. 몇몇 구체예에서, 불포화 효소 유전자는 델타 12 지방산 불포화 효소 유전자이다. 몇몇 경우에 있어서, 세포는 프로토테카 세포이다.

[0022] 제10 양태에서, 본 발명은 상기 양태들 중 임의의 것에 기술된 미세 조류 세포를 제공하는데, 여기서 상기 미세 조류 세포는 탄소원으로서 스타키오스, 라피노스 또는 멜리비오스를 사용하여 배양된다.

[0023] 제11 양태에서, 본 발명은 지질 프로필이 C18:2 2% 이하인 미세 조류 세포를 제공한다. 몇몇 경우에 있어서, 본 발명은 지질 프로필이 C18:2 7% 이하인 미세 조류 세포를 제공한다.

[0024] 제12 양태에서, 본 발명은 지질을 제조하는 방법을 제공한다. 하나의 구체예에서, 본 발명의 방법은 (a) 세포가 건조 중량으로 지질이 20% 이상일 때까지 상기 논의한 바와 같은 미세 조류 세포를 배양하는 단계; 및 (b) 수용성 바이오매스 성분들로부터 지질을 분리하는 단계를 포함한다.

[0025] 제13 양태에서, 본 발명은 지질을 제조하는 다른 방법을 제공한다. 하나의 구체예에서, 본 발명의 방법은 (a) 2개의 상이한 아실-ACP 티오에스테라제를 암호화하는 2개의 상이한 외인성 유전자를 함유하는 미세 조류 세포를 배양하는 단계, 및 (b) 수용성 바이오매스 성분들로부터 지질을 분리하는 단계를 포함한다. 몇몇 경우에 있어서, 외인성 유전자들 중 하나 이상은 표 4에 제시된 티오에스테라제로 이루어진 군으로부터 선택되는 지방 아실-ACP 티오에스테라제를 암호화한다.

[0026] 제14 양태에서, 본 발명은 오일계 생성물을 제조하는 방법을 제공한다. 하나의 구체예에서, 본 발명의 방법은 (a) 세포가 건조 중량으로 지질이 10% 이상일 때까지 상기 논의한 바와 같은 미세 조류 세포를 배양하는 단계;

(b) 미세 조류 세포로부터 지질을 분리하는 단계; (c) 지질을 대상으로, 비누화; 복분해; 산 가수 분해; 알칼리 가수 분해; 효소 가수 분해; 촉매 가수 분해; 고온 응축수 가수 분해; 촉매 가수 분해 반응(이 경우, 지질은 글리세롤과 지방산으로 분할됨); 지방 질소 화합물을 생성하는 아민화 반응; 1 염기산 및 2 염기산을 생성하는 오존 분해 반응; 효소 분할 및 가압 분할로 이루어진 군으로부터 선택되는 트리글리세라이드 분할 반응; 가수 분해 반응 이후의 축합 반응; 수소 가공 반응; 수소 가공 반응 및 탈산소화 반응 및/또는 수소 가공 반응 이전 또는 이와 동시에 진행되는 축합 반응; 가스 제거 반응; 수소화 분해 반응, 수소화, 연속적 수소화-수소화 분해 반응, 연속적 수소화 분해-수소화 반응 및 수소화-수소화 분해 복합 반응으로 이루어진 군으로부터 선택되는 탈산소화 반응; 탈산소화 반응 후 축합 반응; 에스테르화 반응; 에스테르 교환 반응; 에스테르 결합 전이 반응; 하이드록실화 반응; 및 하이드록실화 반응 후 축합 반응으로 이루어진 군으로부터 선택되는 하나 이상의 화학 반응을 진행시키는 단계; 그리고 (d) 임의로 기타 다른 성분들로부터 반응 생성물을 분리하여, 오일계 생성물을 제조하는 단계를 포함한다.

[0027] 몇몇 경우에 있어서, 오일계 생성물은 비누 또는 연료 생성물로부터 선택된다. 몇몇 구체예에서, 오일계 생성물은 바이오디젤, 재생 디젤 및 제트 연료로 이루어진 군으로부터 선택되는 연료 생성물이다. 몇몇 경우에 있어서, 연료 생성물은, (i) 총 카로티노이드 0.025mcg/g 내지 0.3mcg/g, 바람직하게는 0.05mcg/g 내지 0.244mcg/g; (ii) 라이코펜 0.005mcg/g 미만, 바람직하게는 0.003mcg/g 미만; (iii) 베타 카로틴 0.005mcg/g 미만, 바람직하게는 0.003mcg/g 미만; (iv) 엽록소 A 0.025mcg/g 내지 0.3mcg/g, 바람직하게는 0.045mcg/g 내지 0.268mcg/g; (v) 감마 토코페롤 35mcg/g 내지 175mcg/g, 바람직하게는 38.3mcg/g 내지 164mcg/g; (vi) 브라시카스테롤, 캄페스테롤, 스티그나스테롤 또는 베타-시토스테롤 0.25% 미만; (vii) 총 토코트리에놀 225mcg/g 내지 350mcg/g, 바람직하게는 249.6mcg/g 내지 325.3mcg/g; (viii) 루테인 0.0025mcg/g 내지 0.05mcg/g, 바람직하게는 0.003mcg/g 내지 0.039mcg/g; 또는 (ix) 토코페롤 50mcg/g 내지 300mcg/g, 바람직하게는 60.8mcg/g 내지 261.7mcg/g의 속성들 중 하나 이상의 속성을 가지는 바이오디젤이다. 몇몇 경우에 있어서, 연료 생성물은 T10 내지 T90이 20℃, 40℃ 또는 60℃ 이상인 재생 디젤이다. 몇몇 경우에 있어서, 연료 생성물은 HRJ-5 및/또는 ASTM 사양 D1655와 부합하는 제트 연료이다.

[0028] 제15 양태에서, 본 발명은 (a) 지질 프로필이 C8:0 1% 내지 5% 이상, 바람직하게는 3% 이상, C10:0 2.5% 이상, 바람직하게는 4% 이상, C12:0 10% 이상, 바람직하게는 13% 이상, C14:0 10% 이상 및/또는 포화 지방산 60% 이상이고, (b) (i) 총 카로티노이드 0.025mcg/g 내지 0.3mcg/g, 바람직하게는 0.05mcg/g 내지 0.244mcg/g; (ii) 라이코펜 0.005mcg/g 미만, 바람직하게는 0.003mcg/g 미만; (iii) 베타 카로틴 0.005mcg/g 미만, 바람직하게는 0.003mcg/g 미만; (iv) 엽록소 A 0.025mcg/g 내지 0.3mcg/g, 바람직하게는 0.045mcg/g 내지 0.268mcg/g; (v) 감마 토코페롤 35mcg/g 내지 175mcg/g, 바람직하게는 38.3mcg/g 내지 164mcg/g; (vi) 브라시카스테롤, 캄페스테롤, 스티그나스테롤 또는 베타-시토스테롤 0.25% 미만; (vii) 총 토코트리에놀 225mcg/g 내지 350mcg/g, 바람직하게는 249.6mcg/g 내지 325.3mcg/g; (viii) 루테인 0.0025 mcg/g 내지 0.05mcg/g, 바람직하게는 0.003mcg/g 내지 0.039mcg/g; 또는 (ix) 토코페롤 50mcg/g 내지 300mcg/g, 바람직하게는 60.8mcg/g 내지 261.7mcg/g의 속성들 중 하나 이상을 포함하는 트리글리세라이드 오일을 제공한다.

[0029] 제16 양태에서, 본 발명은 C8:C10 지방산 비율이 5:1 이상인 미세 조류로부터 분리된 트리글리세라이드 오일을 제공한다. 몇몇 구체예에서, 트리글리세라이드 오일은 미세 조류 세포(예를 들어, 프로토테카 속에 속하는 미세 조류 세포)로부터 분리되는데, 여기서 상기 미세 조류 세포는 외인성 유전자를 포함한다. 관련 양태에서, 본 발명은 포화 지방산을 60% 이상 포함하는, 미세 조류로부터 분리되는 트리글리세라이드 오일을 제공한다.

[0030] 다른 관련 양태에서, 본 발명은 C16:14 지방산 비율이 약 2:1인 지질 프로필을 가지는 미세 조류로부터 분리된 트리글리세라이드 오일을 제공한다. 다른 관련 양태에서, 본 발명은 미세 조류 세포에 의해 생산되는 트리글리세라이드 오일을 제공하는데, 여기서 상기 미세 조류 세포는 외인성 유전자를 포함한다. 몇몇 경우에 있어서, 미세 조류는 프로토테카 속에 속하는 것이다.

[0031] 또 다른 관련 양태에서, 본 발명은 C12:14 지방산 비율이 약 3:1인 지질 프로필을 가지는 미세 조류로부터 분리된 트리글리세라이드 오일을 제공한다. 다른 관련 양태에서, 본 발명은 미세 조류 세포에 의해 생산되는 트리글리세라이드 오일을 제공하는데, 여기서 상기 미세 조류 세포는 외인성 유전자를 포함한다. 몇몇 경우에 있어서, 미세 조류는 프로토테카 속에 속하는 것이다.

[0032] 제17 양태에서, 본 발명은 (a) C12 미만 지방산이 1% 미만; C14:0 지방산이 1% 내지 10%, 바람직하게는 2% 내지 7%; C16:0 지방산이 20% 내지 35%, 바람직하게는 23% 내지 30%; C18:0 지방산이 5% 내지 20%, 바람직하게는 7% 내지 15%; C18:1 지방산이 35% 내지 60%, 바람직하게는 40% 내지 55%; 그리고 C18:2 지방산이 1% 내지 20%, 바

람직하게는 2% 내지 15%인 지질 프로필; 및 (b)(i) 총 카로티노이드 0.025mcg/g 내지 0.3mcg/g, 바람직하게는 0.05mcg/g 내지 0.244mcg/g; (ii) 라이코펜 0.005mcg/g 미만, 바람직하게는 0.003mcg/g 미만; (iii) 베타 카로틴 0.005mcg/g 미만, 바람직하게는 0.003mcg/g 미만; (iv) 엽록소 A 0.025mcg/g 내지 0.3mcg/g, 바람직하게는 0.045mcg/g 내지 0.268mcg/g; (v) 감마 토코페롤 35mcg/g 내지 175mcg/g, 바람직하게는 38.3mcg/g 내지 164mcg/g; (vi) 브라시카스테롤, 캄페스테롤, 스티그나스테롤 또는 베타-시토스테롤 0.25% 미만; (vii) 총 토코트리에놀 225mcg/g 내지 350mcg/g, 바람직하게는 249.6mcg/g 내지 325.3mcg/g; (viii) 루테인 0.0025mcg/g 내지 0.05mcg/g, 바람직하게는 0.003mcg/g 내지 0.039mcg/g; 또는 (ix) 토코페롤 50mcg/g 내지 300mcg/g, 바람직하게는 60.8mcg/g 내지 261.7mcg/g의 속성들 중 하나 이상을 가지는 트리글리세라이드 오일을 제공한다.

[0033] 몇몇 경우에 있어서, 트리글리세라이드 오일은 하나 이상의 외인성 유전자를 포함하는 미생물로부터 분리된다. 몇몇 구체예에서, 하나 이상의 외인성 유전자는 지방 아실-ACP 티오에스테라제를 암호화한다. 몇몇 경우에 있어서, 지방 아실-ACP 티오에스테라제는 사슬 길이가 C14인 지방 아실-ACP 기질에 대해 가수 분해 활성을 가진다. 몇몇 구체예에서, 미생물은 발현 구조물을 추가로 포함하는데, 여기서 상기 발현 구조물은, 내인성 유전자가 돌연 변이되지 않은 유전자 생성물보다 활성이 작거나 불활성인 유전자 생성물을 암호화하도록 돌연 변이된 것, 내인성 유전자가 미세 조류 세포 게놈으로부터 결실된 것 및 RNA 유도 기작을 통한 것으로 이루어진 방법으로부터 선택된 내인성 유전자의 발현을 하향 조절한다. 몇몇 구체예에서, 내인성 유전자는 불포화 효소를 암호화한다. 몇몇 경우에 있어서, 불포화 효소는 스테아로일-아실기 운반 단백질 불포화 효소(SAD) 또는 지방산 불포화 효소(FAD)이다.

[0034] 제18 양태에서, 본 발명은 C12 미만 지방산이 1% 미만; C14:0 지방산이 2% 내지 7%; C16:0 지방산이 23% 내지 30%; C18:0 지방산이 7% 내지 15%; C18:1 지방산이 40% 내지 55%; 그리고 C18:2 지방산이 2% 내지 15%인 지질 프로필을 포함하는 트리글리세라이드 오일을 제조하는 방법을 제공하는데, 여기서 상기 트리글리세라이드 오일은 하나 이상의 외인성 유전자를 포함하는 미생물로부터 분리된다. 몇몇 경우에 있어서, 트리글리세라이드 오일은 C14:0이 3% 내지 5%; C16:0이 25% 내지 27%; C18:0이 10% 내지 15%; 그리고 C18:1이 40% 내지 45%인 지질 프로필을 포함한다. 몇몇 구체예에서, 하나 이상의 외인성 유전자는 지방 아실-ACP 티오에스테라제를 암호화한다. 몇몇 경우에 있어서, 지방 아실-ACP 티오에스테라제는 사슬 길이가 C14인 지방 아실-ACP 기질에 대해 가수 분해 활성을 가진다.

[0035] 제19 양태에서, 본 발명은 리시놀레산을 생산하는 미생물 세포를 제공한다. 몇몇 경우에 있어서, 미생물 세포는 미세 조류 세포이다. 몇몇 경우에 있어서, 미생물 세포와 미세 조류 세포는 지방산 하이드록실라제를 암호화하는 외인성 유전자를 포함한다. 몇몇 구체예에서, 지방산 하이드록실라제는 올리에이트 12-하이드록실라제이다. 몇몇 경우에 있어서, 지방산 하이드록실라제는 리시누스 커뮤니스로부터 유래한다. 몇몇 경우에 있어서, 미생물 세포는 프로토테카 속, 예를 들어 프로토테카 모리포르미스(*Prototheca moriformis*)에 속하는 것이다.

[0036] 제20 양태에서, 본 발명은 알파-갈락토시다제를 암호화하는 외인성 유전자를 포함하는 미세 조류 세포를 제공한다. 몇몇 경우에 있어서, 미세 조류 세포는 알파-갈락토시다제를 암호화하는 외인성 유전자를 포함하는데, 여기서 상기 알파-갈락토시다제는 분비된다. 몇몇 구체예에서, 알파-갈락토시다제를 암호화하는 외인성 유전자는 사카로마이세스(*Saccharomyces*), 아스퍼질러스(*Aspergillus*) 및 시아몹시스(*Cyamopsis*)로 이루어진 군으로부터 선택되는 속으로부터 유래한다.

[0037] 제21 양태에서, 본 발명은, (a) 종속 영양 조건에서 고정 탄소원의 존재 하에 미세 조류 세포를 배양하는 단계로서, 여기서 상기 미세 조류 세포는 알파-갈락토시다제를 암호화하는 외인성 유전자를 포함하고, 상기 고정 탄소원은 라피노스, 스타키오스 및 멜리비오스로 이루어진 군으로부터 선택되는 단계; (b) 지질이 아닌 성분들로부터 지질을 분리하여, 지질 조성물을 제조하는 단계를 포함하는 지질 조성물을 제조하는 방법을 제공한다. 몇몇 경우에 있어서, 미세 조류 세포는 프로토테카 속에 속하는 것이다.

[0038] 제22 양태에서, 본 발명은 THIC 효소를 암호화하는 외인성 유전자를 포함하는 미세 조류 세포를 제공한다. 몇몇 경우에 있어서, THIC 효소는 코코믹사(*Coccomyxa*), 아라비돕시스(*Arabidopsis*) 및 시네코시스티스(*Synechocystis*)로 이루어진 군으로부터 선택되는 속으로부터 유래한다.

[0039] 다른 양태에서, 본 발명은 THIC 효소를 암호화하는 외인성 유전자를 발현하는 단계를 포함하는, 티아민의 부재 하에 미세 조류 세포를 배양하는 방법을 제공한다. 몇몇 경우에 있어서, THIC 효소는 코코믹사, 아라비돕시스 및 시네코시스티스로 이루어진 군으로부터 선택되는 속으로부터 유래한다.

[0040] 다른 양태에서, 본 발명은 본원에 기술된 바와 같이 미세 조류 세포를 제공하는데, 여기서 상기 미세 조류 세포

는 수크로스 인버타제, 알파-갈락토시다제 및 THIC 효소로 이루어진 군으로부터 선택되는 다른 외인성 유전자를 추가로 포함한다.

[0041] 본 발명의 다른 양태는 미생물 세포로부터 분리된 하이드록실화 오일을 제공한다. 하나의 양태에서, 하이드록실화 오일은 미생물 세포로부터 분리되며, 여기서 상기 미생물 세포는 미세 조류 세포(예를 들어, 프로토테카 속에 속하는 미세 조류 세포)이다. 추가의 양태에서, 하이드록실화 오일은 프로토테카 모리포르미스 세포로부터 분리된다. 하나의 구체예에서, 하이드록실화 오일은 하이드록실화 트리글리세라이드이다. 하이드록실화 트리글리세라이드는 피마자유와 화학적으로 유사하거나 동일할 수 있다.

[0042] 본 발명의 추가의 양태는 하이드록실화 지방산이다. 하이드록실화 지방산의 하나의 구체예는 리시놀레산이다.

[0043] 또 다른 양태에서, 미생물의 하이드록실화 오일 또는 하이드록실화 지방산은 추가로 하이드록실화된다. 리시놀레산이 하이드록실화될 때, 2개의 하이드록실기를 함유하는 지방산이 제공된다.

[0044] 본 발명의 또 다른 양태는 하이드록실화 오일 및/또는 하이드록실화 지방산과, 이소시아네이트부를 함유하는 화합물을 반응시켜 폴리우레탄을 형성함으로써 제조되는 조성물을 제공한다.

[0045] 본 발명의 다른 양태는 C18:2가 20% 이상인 지질 프로필을 가지는 미세 조류 세포를 제공한다. 몇몇 경우에 있어서, 미세 조류 세포는 C18:2가 30% 이상인 지질 프로필을 가진다. 몇몇 경우에 있어서, 미세 조류 세포는 C18:2가 40% 이상인 지질 프로필을 가진다. 몇몇 경우에 있어서, 미세 조류 세포는 C18:2가 50% 이상인 지질 프로필을 가진다.

[0046] 본 발명의 다른 양태는, (a) 배양 배지 중에서 미세 조류 세포를 배양하고, 당 농도를 모니터하는 단계; (b) 배양 배지의 당 농도가 약 1g/l 미만에 이를 때, 배양 배지에 제1 당 용액을 약 2시간 내지 약 24시간 동안 연속 공급율 약 2g/h/l 내지 약 10g/h/l 로 첨가하는 단계; (c) 배양 배지에 제2 당 용액을 첨가하여, 배양 배지의 당 농도를 약 15g/l 내지 약 20g/l로 유지하는 단계; 및 (d) 미세 조류 바이오매스로부터 지질을 분리하는 단계를 포함하는 지질 제조 방법을 제공한다. 몇몇 경우에 있어서, 제1 당 용액은 배양 배지에 공급율 약 4g 수크로스/h/l 내지 약 6g 수크로스/h/l로 첨가된다. 몇몇 경우에 있어서, 제1 당 용액은 배양 배지에 공급율 약 5.25g 수크로스/h/l로 첨가된다. 몇몇 경우에 있어서, 당은 수크로스 또는 글루코스이다.

[0047] 본 발명의 다른 양태는 C18:3이 10% 이상인 지질 프로필을 가지는 미세 조류 세포를 제공한다. 몇몇 경우에 있어서, 미세 조류 세포는 C18:3이 20% 이상인 지질 프로필을 가진다. 몇몇 경우에 있어서, 미세 조류 세포는 C18:3이 30% 이상인 지질 프로필을 가진다. 몇몇 경우에 있어서, 미세 조류 세포는 C18:3이 40% 이상인 지질 프로필을 가진다. 몇몇 경우에 있어서, 미세 조류 세포는 C18:3이 50% 이상인 지질 프로필을 가진다.

[0048] 본 발명의 다른 양태는 리놀레산 또는 리놀렌산을 포함하는 트리글리세라이드를 생산하는 미생물을 제공하는데, 여기서 상기 미생물은 β-케토아실-ACP 합성 효소 II(KAS II) 효소를 암호화하는 재조합 핵산을 포함한다. 몇몇 경우에 있어서, 미생물은 스테아로일 ACP 불포화 효소(SAD) 효소를 암호화하는 재조합 핵산을 추가로 포함한다. 몇몇 경우에 있어서, 미생물은 올리에이트 특이적 티오에스테라제 효소를 암호화하는 재조합 핵산을 추가로 포함한다. 몇몇 경우에 있어서, 미생물은 지방산 불포화 효소(FAD) 효소를 암호화하는 재조합 핵산을 추가로 포함한다. 몇몇 경우에 있어서, 미생물은 글리세로지질 불포화 효소를 암호화하는 재조합 핵산을 추가로 포함한다.

[0049] 상기 논의한 조작 미생물에 있어서, KAS II 효소는 서열 번호 175, 서열 번호 176, 서열 번호 177, 서열 번호 178 및 서열 번호 179로 이루어진 군으로부터 선택되는 아미노산 서열을 포함할 수 있다. 몇몇 경우에 있어서, SAD 효소는 서열 번호 172, 서열 번호 196, 서열 번호 197, 서열 번호 198, 서열 번호 199, 서열 번호 200 및 서열 번호 201로 이루어진 군으로부터 선택되는 아미노산 서열을 포함할 수 있다. 몇몇 경우에 있어서, 올리에이트 특이적 티오에스테라제 효소는 서열 번호 195, 서열 번호 202, 서열 번호 203, 서열 번호 204, 서열 번호 205 및 서열 번호 206으로 이루어진 군으로부터 선택되는 아미노산 서열을 포함할 수 있다. 몇몇 경우에 있어서, FAD 효소는 서열 번호 181, 서열 번호 182, 서열 번호 183, 서열 번호 184 및 서열 번호 185로 이루어진 군으로부터 선택되는 아미노산 서열을 포함할 수 있다. 몇몇 경우에 있어서, FAD 효소는 서열 번호 207, 서열 번호 208, 서열 번호 209, 서열 번호 210, 서열 번호 211 및 서열 번호 212로 이루어진 군으로부터 선택되는 아미노산 서열을 포함하는 Δ12 FAD 효소이다. 몇몇 경우에 있어서, FAD 효소는 서열 번호 213, 서열 번호 214, 서열 번호 215, 서열 번호 216, 서열 번호 217, 서열 번호 218, 서열 번호 219, 서열 번호 220 및 서열 번호 221로 이루어진 군으로부터 선택되는 아미노산 서열을 포함하는 Δ15 FAD 효소이다. 몇몇 경우에 있어서, 글리세로지질 불포화 효소는 ω-6 지방산 불포화 효소, ω-3 지방산 불포화 효소 또는 ω-6 올리에이트 불포화 효소다.

[0050] 상기 논의한 조작 미생물들 중 임의의 하나는 수크로스 이용 경로 효소를 암호화하는 재조합 핵산을 추가로 포함할 수 있다. 몇몇 경우에 있어서, 수크로스 이용 경로 효소는 수크로스 인버타제이다.

[0051] 상기 논의한 조작 미생물들 중 임의의 하나에 있어서, 미생물은 기타 다른 지방산에 비하여 리놀레산 또는 리놀렌산의 비율이 증가하도록 추가로 조작될 수 있다.

[0052] 상기 논의한 조작 미생물들 중 임의의 하나에 있어서, 미생물은 C18 기질에 특이적이거나 이 기질을 선호하는 티오에스테라제를 과발현하도록 추가로 조작될 수 있다.

[0053] 상기 논의한 조작 미생물들 중 임의의 하나에 있어서, 미생물은 C8 내지 C16 기질에 특이적이거나 이 기질을 선호하는 티오에스테라제의 발현이 감소하도록 추가로 조작될 수 있다.

[0054] 상기 논의한 조작 미생물들 중 임의의 하나에 있어서, 미생물은 미세 조류 세포일 수 있다. 몇몇 경우에 있어서, 미세 조류 세포는 프로토테카 속에 속하는 것이다. 몇몇 경우에 있어서, 미세 조류는 프로토테카 모리포르미스 세포이다.

[0055] 본 발명의 다른 양태는 상기 논의한 조작 미생물들 중 임의의 하나에 의해 생산된 오일을 제공한다.

[0056] 본 발명의 다른 양태는 재조합 핵산들 중 하나 이상을 미생물에 도입하여 리놀레산 또는 리놀렌산을 포함하는 트리글리세라이드를 생산하게 함으로써, 상기 논의한 조작 미생물들을 제조하는 방법을 제공한다.

[0057] 본 발명의 다른 양태는 트리아실글리세라이드를 포함하는 천연 오일 또는 천연 오일로부터 생성되는 생성물을 제조하는 방법을 제공하는데, 여기서 이 방법은 (i)(a) β-케토아실-ACP 합성 효소 I, β-케토아실-ACP 합성 효소 II, 스테아로일 ACP 불포화 효소, 지방산 불포화 효소 또는 아실-ACP 티오에스테라제를 암호화하는 하나 이상의 유전자에 의해 암호화되는 효소의 발현을 감소시키거나 이 효소를 발현하지 않거나(여기서, 세포는 임의로 β-케토아실-ACP 합성 효소 I, β-케토아실-ACP 합성 효소 II, 스테아로일 ACP 불포화 효소, 지방산 불포화 효소 또는 아실-ACP 티오에스테라제를 암호화하는 유전자의 복사체 2개의 발현을 감소시키거나 이 복사체 2개를 발현하지 않도록 작동 가능한 재조합 핵산을 포함함); 또는 (b) 활성 β-케토아실-ACP 합성 효소 I, β-케토아실-ACP 합성 효소 II, 스테아로일 ACP 불포화 효소, 지방산 불포화 효소 또는 아실-ACP 티오에스테라제를 암호화하는 외인성 유전자 생성물을 발현하거나; 또는 (c) β-케토아실-ACP 합성 효소 I 또는 β-케토아실-ACP 합성 효소 II를 암호화하는 하나 이상의 유전자에 의해 암호화되는 효소의 발현을 감소시키거나 또는 이 효소를 발현하지 않고, 활성 스테아로일 ACP 불포화 효소, 지방산 불포화 효소 또는 아실-ACP 티오에스테라제를 암호화하는 외인성 유전자의 생성물을 발현하거나; 또는 (d) 스테아로일 ACP 불포화 효소 또는 지방산 불포화 효소를 암호화하는 하나 이상의 유전자에 의해 암호화되는 효소의 발현을 감소시키거나 또는 이 효소를 발현하지 않고, 활성 β-케토아실-ACP 합성 효소 I, β-케토아실-ACP 합성 효소 II 또는 아실-ACP 티오에스테라제를 암호화하는 외인성 유전자의 생성물을 발현하도록 작동 가능한 재조합 핵산을 포함하는 세포인, 재조합 미생물의 세포를 배양하는 단계; 그리고 (ii) 세포로부터 천연 오일을 회수하는 단계를 포함하고, 임의로는 식품, 연료 또는 화학 생성물을 생성하기 위하여 천연 오일을 가공하는 단계로서, 천연 오일은 재조합 핵산으로 인하여 변경된 지방산 프로필을 가지는 단계를 추가로 포함한다.

[0058] 몇몇 경우에 있어서, 미생물은 제II형 지방산 생합성 경로를 통해 지방산을 합성한다. 몇몇 경우에 있어서, 미생물은 미세 조류이다. 몇몇 경우에 있어서, 미세 조류는 절대 종속 영양 생물이다. 몇몇 경우에 있어서, 미세 조류는 프로토테카 또는 클로렐라(*Chlorella*) 종이다. 몇몇 경우에 있어서, 미세 조류는 프로토테카 윅커해미(*Prototheca wickerhamii*), 프로토테카 스태그노라(*Prototheca stagnora*), 프로토테카 포르토리센시스(*Prototheca portoricensis*), 프로토테카 모리포르미스 또는 프로토테카 조프피(*Prototheca zopfii*)이다. 몇몇 경우에 있어서, 미세 조류는 클로렐라 케슬러리(*Chlorella kessleri*), 클로렐라 루테오비리디스(*Chlorella luteoviridis*), 클로렐라 프로토테코이데스(*Chlorella protothecoides*) 또는 클로렐라 불가리스(*Chlorella vulgaris*)이다. 몇몇 경우에 있어서, 세포는 활성 수크로스 인버타제를 발현하는 재조합 세포이다. 몇몇 경우에 있어서, 배양은 종속 영양 배양이다. 몇몇 경우에 있어서, 지방산 불포화 효소는 ω-6 지방산 불포화 효소, ω-3 지방산 불포화 효소 또는 ω-6 올리에이트 불포화 효소 또는 델타 12 지방산 불포화 효소 중 하나 이상이다. 몇몇 경우에 있어서, 세포는 건조 세포 중량으로 50% 이상, 60% 이상, 70% 이상 또는 50% 내지 90%의 트리글리세라이드를 포함하도록 배양된다. 몇몇 경우에 있어서, 오일은 발색 분자 500ppm, 50ppm 또는 5ppm 미만을 포함한다. 몇몇 경우에 있어서, 재조합 핵산은 안정적으로 통합된다. 몇몇 경우에 있어서, 재조합 핵산은 미생물의 염색체 내에 안정적으로 통합된다. 몇몇 경우에 있어서, 세포는 하나 이상의 선택가능한 마커를 추가로 포함한다.

[0059] 몇몇 경우에 있어서, 세포는 효소를 암호화하는 유전자 전사체를 표적화하는 안티센스, RNAi 또는 dsRNA의 발현을 통해 효소의 발현을 감소시키거나 이 효소를 발현하지 않도록 작동 가능한 재조합 핵산을 포함한다. 몇몇 경우에 있어서, β-케토아실-ACP 합성 효소 I, β-케토아실-ACP 합성 효소 II, 스테아로일 ACP 불포화 효소 또는 지방산 불포화 효소를 암호화하는 하나 이상의 유전자에 의해 암호화되는 효소의 발현을 감소시키거나 이 효소를 발현하지 않는 것은, 활성 β-케토아실-ACP 합성 효소 I, β-케토아실-ACP 합성 효소 II, 스테아로일 ACP 불포화 효소, 지방산 불포화 효소 또는 아실-ACP 티오에스테라제를 암호화하는 하나 이상의 유전자로 하나 이상의 유전자를 분할(interrupt) 또는 치환하는 것을 원인으로 한다. 몇몇 경우에 있어서, 재조합 세포는 올리에이트 12-하이드록실라제를 암호화하는 외인성 유전자를 추가로 포함하여, 리시놀레산을 합성한다. 몇몇 경우에 있어서, 재조합 세포는 KASII 유전자에 의해 암호화되는 β-케토아실-ACP 합성 효소 II의 발현을 감소시키거나 이 효소를 발현하지 않도록 작동 가능하고, 또한 아실-ACP 티오에스테라제를 암호화하는 외인성 유전자의 생성물을 발현하도록 작동 가능한 핵산을 포함한다. 몇몇 경우에 있어서, 세포는 40%, 50%, 60%, 70% 또는 80% 이상의 C16 지방산을 가지는 것을 특징으로 하는 지방산 프로필을 나타내는 오일을 생산한다. 몇몇 경우에 있어서, 세포는 C16:0이 50% 내지 75% 이상인 것을 특징으로 하는 지방산 프로필을 가지는 오일을 생산한다. 몇몇 경우에 있어서, 세포는 C18:1이 20% 내지 40% 이상인 것을 추가의 특징으로 하는 지방산 프로필을 가지는 오일을 생산한다. 몇몇 경우에 있어서, 아실-ACP 티오에스테라제를 암호화하는 외인성 유전자는 C8 내지 C16 지방 아실 사슬의 가수 분해 활성이 세포의 천연 아실-ACP 티오에스테라제보다 큰 활성 아실-ACP 티오에스테라제를 생산한다. 몇몇 경우에 있어서, 아실-ACP 티오에스테라제를 암호화하는 외인성 유전자는 KASII 유전자를 분할한다. 몇몇 경우에 있어서, 재조합 세포는 β-케토아실-ACP 합성 효소 I을 암호화하는 하나 이상의 유전자에 의해 암호화되는 효소의 발현을 감소시키거나 이 효소를 발현하지 않도록 작동 가능한 핵산을 포함한다.

[0060] 몇몇 경우에 있어서, 생산된 오일은 재조합 핵산으로 인하여 평균 지방산 사슬 길이가 더 짧은 것을 특징으로 하는 지방산 프로필을 가진다. 몇몇 경우에 있어서, 재조합 세포는 하나 이상의 FAD 유전자에 의해 암호화되는 지방산 불포화 효소의 발현을 감소시키거나 이 효소를 발현하지 않도록 작동 가능하며, 활성 스테아로일 ACP 불포화 효소를 암호화하는 스테아로일-ACP 불포화 효소 외인성 유전자의 생성물을 발현하도록 작동 가능한 핵산을 포함한다. 몇몇 경우에 있어서, 핵산은 지방산 불포화 효소 유전자의 복사체 다수 개에 의해 암호화되는 지방산 불포화 효소의 발현을 감소시키거나 이 효소를 발현하지 않도록 작동 가능하다. 몇몇 경우에 있어서, 스테아로일-ACP 불포화 효소 외인성 유전자는 지방산 불포화 효소 유전자의 암호화 영역 내 위치에 재조합된다. 몇몇 경우에 있어서, 생산된 오일은 올레산이 증가한 지방산 프로필을 가진다. 몇몇 경우에 있어서, 올레산은 지방산의 50%, 60%, 70%, 80% 또는 90% 이상을 포함한다. 몇몇 경우에 있어서, 재조합 세포는 활성 β-케토아실-ACP 합성 효소 II를 암호화하는 β-케토아실-ACP 합성 효소 II 외인성 유전자의 생성물을 발현하도록 작동 가능한 핵산을 포함한다. 몇몇 경우에 있어서, 생산된 오일은 재조합 핵산으로 인하여 C18:1 지방산이 증가하였고, C16 지방산은 감소한 지방산 프로필을 특징으로 한다. 몇몇 경우에 있어서, 재조합 세포는 RNA 간섭에 의해 스테아로일-ACP 불포화 효소를 암호화하는 하나 이상의 유전자에 의해 암호화되는 효소의 발현을 감소시키거나 이 효소를 발현하지 않도록 작동 가능한 핵산을 포함한다. 몇몇 경우에 있어서, 생산된 오일은 C18:0 지방산이 증가한 것을 특징으로 하는 지방산 프로필을 가진다. 몇몇 경우에 있어서, 생산된 오일은 C18:0이 50%, 60%, 70%, 80% 또는 90% 이상인 지방산 프로필을 특징으로 한다. 몇몇 경우에 있어서, 생산된 오일은 C18:0이 50% 내지 70% 이상인 지방산 프로필을 특징으로 한다. 몇몇 경우에 있어서, 생산된 오일은 C18:1이 20% 내지 40% 이상인 지방산 프로필을 추가의 특징으로 한다. 몇몇 경우에 있어서, 세포는 β-케토아실-ACP 합성 효소 I, β-케토아실-ACP 합성 효소 II, 스테아로일 ACP 불포화 효소 또는 지방산 불포화 효소를 암호화하는 유전자의 복사체 2개의 발현을 감소시키거나 이 복사체 2개를 발현하지 않도록 작동 가능한 재조합 핵산을 포함한다. 몇몇 경우에 있어서, 핵산은 활성 ω-3 지방산 불포화 효소 및/또는 ω-6 올리에이트 불포화 효소를 암호화하는 지방산 불포화 효소 외인성 유전자의 생성물을 발현하도록 작동 가능하다. 몇몇 경우에 있어서, 생산된 오일은 리놀레산, 리놀렌산 또는 이것들 둘 다의 수준이 상승한 것을 특징으로 하는 지방산 프로필을 가진다. 몇몇 경우에 있어서, 오일의 지방산 프로필은 리놀레산, 리놀렌산 또는 이것들 둘 다가 10%, 20%, 30%, 40% 또는 50% 이상인 것을 특징으로 한다. 몇몇 경우에 있어서, 오일의 추가 가공으로서는 정제, 표백, 탈취, 복분해, 에스테르 결합 전이, 수소화, 가수 분해, 수소화, 탈산소화, 수소 크래킹, 이성체화, 하이드록실화, 에스테르 교환, 아미드화, 설폰화 및 가황 처리 중 하나 이상을 포함한다. 몇몇 경우에 있어서, 오일은 식품용 오일, 지방산, 지방 알코올, 윤활제, 비누, 지방산 에스테르, 지방산 에톡실레이트, 지방 아민, 알킬 클로라이드, 지방 알코올 에톡실레이트, 지방 알코올 설페이트, 지방산 알카놀아미드, 설폰화된 오일, 가황 오일, 디젤 연료, 제트 연료, 가솔린, 연료 블렌드스톡(fuel blendstock), 연료 첨가제, 윤활제 첨가제 또는 코팅제를 형성하도록 가공된다.

[0061] 본 발명의 다른 양태는 상기 논의한 방법에 의해 얻을 수 있는 천연 오일을 제공한다.

[0062] 본 발명의 다른 양태는 상기 논의한 천연 오일로부터 제조된 생성물을 제공한다. 몇몇 경우에 있어서, 생성물은 식품용 오일, 지방산, 지방 알코올, 윤활제, 비누, 지방산 에스테르, 지방산 에톡실레이트, 지방 아민, 알킬 클로라이드, 지방 알코올 에톡실레이트, 지방 알코올 설페이트, 지방산 알카놀아미드, 설폰화 오일, 가황 오일, 디젤 연료, 제트 연료, 가솔린, 연료 블렌드스톡, 연료 첨가제, 화학 첨가제 또는 코팅제를 포함한다.

[0063] 본 발명의 다른 양태는 (a) β-케토아실-ACP 합성 효소 I, β-케토아실-ACP 합성 효소 II, 스테아로일 ACP 불포화 효소, 지방산 불포화 효소 또는 아실-ACP 티오에스테라제를 암호화하는 하나 이상의 유전자에 의해 암호화되는 효소의 발현을 감소시키거나 이 효소를 발현하지 않거나(여기서, 세포는 임의로 β-케토아실-ACP 합성 효소 I, β-케토아실-ACP 합성 효소 II, 스테아로일 ACP 불포화 효소, 지방산 불포화 효소 또는 아실-ACP 티오에스테라제를 암호화하는 유전자의 복사체 2개의 발현을 감소시키거나 이 복사체 2개를 발현하지 않도록 작동 가능한 재조합 핵산을 포함함); 또는 (b) 활성 β-케토아실-ACP 합성 효소 I, β-케토아실-ACP 합성 효소 II, 스테아로일 ACP 불포화 효소, 지방산 불포화 효소 또는 아실-ACP 티오에스테라제를 암호화하는 외인성 유전자 생성물을 발현하거나; 또는 (c) β-케토아실-ACP 합성 효소 I, β-케토아실-ACP 합성 효소 II를 암호화하는 하나 이상의 유전자에 의해 암호화되는 효소의 발현을 감소시키거나 또는 이것을 발현하지 않고, 활성 스테아로일 ACP 불포화 효소, 지방산 불포화 효소 또는 아실-ACP 티오에스테라제를 암호화하는 외인성 유전자의 생성물을 발현하거나; 또는 (d) 스테아로일 ACP 불포화 효소 또는 지방산 불포화 효소를 암호화하는 하나 이상의 유전자에 의해 암호화되는 효소의 발현을 감소시키거나 이 효소를 발현하지 않고, 활성 β-케토아실-ACP 합성 효소 I, β-케토아실-ACP 합성 효소 II 또는 아실-ACP 티오에스테라제를 암호화하는 외인성 유전자의 생성물을 발현하도록 작동 가능한 재조합 핵산을 포함하는 재조합 세포를 제공한다.

[0064] 몇몇 경우에 있어서, 미생물은 제II형 지방산 생합성 경로를 통해 지방산을 합성한다. 몇몇 경우에 있어서, 미생물은 미세 조류이다. 몇몇 경우에 있어서, 미세 조류는 절대 종속 영양 생물이다. 몇몇 경우에 있어서, 미세 조류는 프로토테카 또는 클로렐라 종이다. 몇몇 경우에 있어서, 미세 조류는 프로토테카 윅커해미, 프로토테카 스태그노라, 프로토테카 포르토리센시스, 프로토테카 모리포르미스 또는 프로토테카 조프피이다. 몇몇 경우에 있어서, 미세 조류는 클로렐라 케슬러리, 클로렐라 루테오비리디스, 클로렐라 프로토테코이데스 또는 클로렐라 불가리스이다. 몇몇 경우에 있어서, 세포는 활성 수크로스 인버타제를 발현하는 재조합 세포이다. 몇몇 경우에 있어서, 세포는 종속 영양 생장을 할 수 있다. 몇몇 경우에 있어서, 지방산 불포화 효소는 ω-6 지방산 불포화 효소, ω-3 지방산 불포화 효소 또는 ω-6 올리에이트 불포화 효소 또는 델타 12 지방산 불포화 효소 중 하나 이상이다. 몇몇 경우에 있어서, 세포는 건조 세포 중량을 기준으로 50% 이상, 60% 이상, 70% 이상 또는 50% 내지 90%의 트리글리세라이드를 포함하도록 배양될 수 있다. 몇몇 경우에 있어서, 재조합 핵산은 안정적으로 통합된다. 몇몇 경우에 있어서, 재조합 핵산은 미생물의 염색체 내에 안정적으로 통합된다. 몇몇 경우에 있어서, 세포는 하나 이상의 선택가능한 마커를 추가로 포함한다. 몇몇 경우에 있어서, 세포는 효소를 암호화하는 유전자 전사체를 표적화하는 안티센스, RNAi 또는 dsRNA의 발현을 통해 효소의 발현을 감소시키거나 이 효소를 발현하지 않도록 작동 가능한 재조합 핵산을 포함한다. 몇몇 경우에 있어서, β-케토아실-ACP 합성 효소 I, β-케토아실-ACP 합성 효소 II, 스테아로일 ACP 불포화 효소 또는 지방산 불포화 효소를 암호화하는 하나 이상의 유전자에 의해 암호화되는 효소의 발현을 감소시키거나 이 효소를 발현하지 않는 것은, 활성 β-케토아실-ACP 합성 효소 I, β-케토아실-ACP 합성 효소 II, 스테아로일 ACP 불포화 효소, 지방산 불포화 효소 또는 아실-ACP 티오에스테라제를 암호화하는 하나 이상의 유전자로 하나 이상의 유전자를 분할 또는 치환하는 것을 원인으로 한다.

[0065] 몇몇 경우에 있어서, 재조합 세포는 올리에이트 12-하이드록실라제를 암호화하는 외인성 유전자를 추가로 포함하여, 리시놀레산을 합성한다. 몇몇 경우에 있어서, 재조합 세포는 KASII 유전자에 의해 암호화되는 β-케토아실-ACP 합성 효소 II의 발현을 감소시키거나 이 효소를 발현하지 않도록 작동 가능하고, 아실-ACP 티오에스테라제를 암호화하는 외인성 유전자의 생성물을 발현하도록 작동 가능한 핵산을 포함한다. 몇몇 경우에 있어서, 세포는 C16 지방산이 40%, 50%, 60%, 70% 또는 80% 이상인 것을 특징으로 하는 지방산 프로필을 가지는 오일을 생산한다. 몇몇 경우에 있어서, 세포는 C16:0이 50% 내지 75% 이상인 것을 특징으로 하는 지방산 프로필을 가지는 오일을 생산한다. 몇몇 경우에 있어서, 세포는 C18:1이 20% 내지 40% 이상인 것을 추가의 특징으로 하는 지방산 프로필을 가지는 오일을 생산한다. 몇몇 경우에 있어서, 아실-ACP 티오에스테라제를 암호화하는 외인성 유전자는 C8 내지 C16 지방 아실 사슬의 가수 분해 활성이 세포의 천연 아실-ACP 티오에스테라제보다 큰 활성 아실-ACP 티오에스테라제를 생산한다. 몇몇 경우에 있어서, 아실-ACP 티오에스테라제를 암호화하는 외인성 유전자는 KASII 유전자를 분할한다. 몇몇 경우에 있어서, 재조합 세포는 β-케토아실-ACP 합성 효소 I을 암호화하는 하나 이상의 유전자에 의해 암호화되는 효소의 발현을 감소시키거나 이 효소를 발현하지 않도록 작동 가능한 핵산을 포함한다. 몇몇 경우에 있어서, 생산된 오일은 재조합 핵산으로 인하여 평균 지방산 사슬 길이가 더 짧은 것을

특징으로 하는 지방산 프로필을 가진다. 몇몇 경우에 있어서, 재조합 세포는 하나 이상의 FAD 유전자에 의해 암호화되는 지방산 불포화 효소의 발현을 감소시키거나 이 효소를 발현하지 않도록 작동 가능하며, 활성 스테아로일 ACP 불포화 효소를 암호화하는 스테아로일-ACP 불포화 효소 외인성 유전자의 생성물을 발현하도록 작동 가능한 핵산을 포함한다. 몇몇 경우에 있어서, 핵산은 지방산 불포화 효소 유전자의 복사체 다수 개에 의해 암호화되는 지방산 불포화 효소의 발현을 감소시키거나 이 효소를 발현하지 않도록 작동 가능하다. 몇몇 경우에 있어서, 스테아로일-ACP 불포화 효소 외인성 유전자는 지방산 불포화 효소 유전자의 암호화 영역 내 위치에 재조합된다.

[0066] 몇몇 경우에 있어서, 생산된 오일은 올레산이 증가한 지방산 프로필을 가진다. 몇몇 경우에 있어서, 올레산은 지방산의 50%, 60%, 70%, 80% 또는 90% 이상을 포함한다. 몇몇 경우에 있어서, 재조합 세포는 활성 β-케토아실-ACP 합성 효소 II를 암호화하는 β-케토아실-ACP 합성 효소 II 외인성 유전자의 생성물을 발현하도록 작동 가능한 핵산을 포함한다. 몇몇 경우에 있어서, 생산된 오일은 재조합 핵산으로 인하여 C18:1 지방산이 증가하였고, C16 지방산은 감소한 지방산 프로필을 특징으로 한다. 몇몇 경우에 있어서, 재조합 세포는 RNA 간섭에 의해 스테아로일 ACP 불포화 효소를 암호화하는 하나 이상의 유전자에 의해 암호화되는 효소의 발현을 감소시키거나 이 효소를 발현하지 않도록 작동 가능한 핵산을 포함한다. 몇몇 경우에 있어서, 생산된 오일은 C18:0 지방산이 증가한 것을 특징으로 하는 지방산 프로필을 가진다. 몇몇 경우에 있어서, 생산된 오일은 C18:0이 50%, 60%, 70%, 80% 또는 90% 이상인 지방산 프로필을 특징으로 한다. 몇몇 경우에 있어서, 생산된 오일은 C18:0이 50% 내지 75% 이상인 지방산 프로필을 특징으로 한다. 몇몇 경우에 있어서, 생산된 오일은 C18:1이 20% 내지 40% 이상인 지방산 프로필을 추가의 특징으로 한다. 몇몇 경우에 있어서, 세포는 β-케토아실-ACP 합성 효소 I, β-케토아실-ACP 합성 효소 II, 스테아로일 ACP 불포화 효소 또는 지방산 불포화 효소를 암호화하는 유전자의 복사체 2개의 발현을 감소시키거나 이 복사체 2개를 발현하지 않도록 작동 가능한 재조합 핵산을 포함한다. 몇몇 경우에 있어서, 핵산은 활성 ω-3 지방산 불포화 효소 및/또는 ω-6 올리에이트 불포화 효소를 암호화하는 지방산 불포화 효소 외인성 유전자의 생성물을 발현하도록 작동 가능하다. 몇몇 경우에 있어서, 생산된 오일은 리놀레산, 리놀렌산 또는 이것들 둘 다의 수준이 상승한 것을 특징으로 하는 지방산 프로필을 가진다. 몇몇 경우에 있어서, 오일의 지방산 프로필은 리놀레산, 리놀렌산 또는 이것들 둘 다가 10%, 20%, 30%, 40% 또는 50% 이상인 것을 특징으로 한다.

[0067] 본 발명의 다른 양태는 전술한 세포로부터 생산된 오일 함유 생성물 또는 천연 오일을 제공한다.

[0068] 본 발명의 다른 양태는 리시놀레산을 포함하는 트리아실글리세라이드를 포함하는 천연 오일, 또는 천연 오일로부터 생산되는 생성물을 제조하는 방법을 제공하는데, 이 방법은 재조합 미생물의 세포를 배양하는 단계를 포함하며, 이 세포는 활성 올리에이트 12-하이드록실라제를 암호화하는 외인성 유전자의 생성물을 발현하여 리시놀레산을 합성하도록 작동 가능한 재조합 핵산을 포함한다.

[0069] 몇몇 경우에 있어서, 미생물은 제II형 지방산 생합성 경로를 가진다. 몇몇 경우에 있어서, 미생물은 미세 조류이다. 몇몇 경우에 있어서, 미세 조류는 절대 종속 영양 생물이다. 몇몇 경우에 있어서, 미세 조류는 프로토테카 종이다. 몇몇 경우에 있어서, 미세 조류는 프로토테카 윅커해미, 프로토테카 스태그노라, 프로토테카 포르토리센시스, 프로토테카 모리포르미스 또는 프로토테카 조프피이다. 몇몇 경우에 있어서, 미세 조류는 클로렐라 케슬러리, 클로렐라 루테오비리디스, 클로렐라 프로토테코이데스 또는 클로렐라 불가리스이다. 몇몇 경우에 있어서, 세포는 활성 수크로스 인버타제를 발현하는 재조합 세포이다. 몇몇 경우에 있어서, 배양은 종속 영양 배양이다. 몇몇 경우에 있어서, 활성 올리에이트 12-하이드록실라제를 암호화하는 외인성 유전자의 생성물을 발현하도록 작동 가능한 재조합 핵산을 가지지 않는 세포는 올레산을 40%, 50%, 60%, 70%, 80% 또는 90% 이상 생산한다. 몇몇 경우에 있어서, 세포는 올레산 생산을 증강하여 올리에이트 12-하이드록실라제에 대한 기질 수준을 상승시키도록 작동 가능한 재조합 핵산을 추가로 포함한다. 몇몇 경우에 있어서, 세포는 (a) 활성 스테아로일 ACP 불포화 효소를 암호화하는 외인성 유전자의 생성물을 발현하고, 지방산 불포화 효소를 암호화하는 하나 이상의 유전자에 의해 암호화되는 효소의 발현을 감소시키거나 이 효소를 발현하지 않거나; 또는 (b) 활성 β-케토아실-ACP 합성 효소 I을 암호화하는 외인성 유전자의 생성물을 발현하고, 활성 아실-ACP 티오에스테라제를 암호화하는 외인성 유전자의 생성물을 발현하도록 작동 가능한 재조합 핵산을 포함한다.

[0070] 본 발명의 다른 양태는 상기 논의한 방법들 중 임의의 것에 의하여 생산된 생성물을 제공한다.

[0071] 본 발명의 다른 양태는 활성 올리에이트 12-하이드록실라제를 암호화하는 외인성 유전자의 생성물을 발현하여 리시놀레산을 합성하도록 작동 가능한 재조합 핵산을 포함하는 미생물 세포를 제공한다.

[0072] 몇몇 경우에 있어서, 미생물은 제II형 지방산 생합성 경로를 가진다. 몇몇 경우에 있어서, 미생물은 미세 조류

이다. 몇몇 경우에 있어서, 미세 조류는 절대 종속 영양 생물이다. 몇몇 경우에 있어서, 미세 조류는 프로토테카 종이다. 몇몇 경우에 있어서, 미세 조류는 프로토테카 윅커해미, 프로토테카 스태그노라, 프로토테카 포르토리센시스, 프로토테카 모리포르미스 또는 프로토테카 조프피이다. 몇몇 경우에 있어서, 미세 조류는 클로렐라 케슬러리, 클로렐라 루테오비리디스, 클로렐라 프로토테코이데스 또는 클로렐라 불가리스이다. 몇몇 경우에 있어서, 세포는 활성 수크로스 인버타제를 발현하는 재조합 세포이다. 몇몇 경우에 있어서, 세포는 종속 영양 생장을 할 수 있다. 몇몇 경우에 있어서, 활성 올리에이트 12-하이드록실라제를 암호화하는 외인성 유전자의 생성물을 발현하도록 작동 가능한 재조합 핵산을 가지지 않는 세포는 올레산을 40%, 50%, 60%, 70%, 80% 또는 90% 이상 생산한다. 몇몇 경우에 있어서, 세포는 올레산 생산을 증강하여 올리에이트 12-하이드록실라제에 대한 기질의 수준을 상승시키도록 작동 가능한 재조합 핵산을 추가로 포함한다. 몇몇 경우에 있어서, 세포는 (a) 활성 스테아로일 ACP 불포화 효소를 암호화하는 외인성 유전자의 생성물을 발현하고, 지방산 불포화 효소를 암호화하는 하나 이상의 유전자에 의해 암호화되는 효소의 발현을 감소시키거나 이 효소를 발현하지 않거나; 또는 (b) 활성 β-케토아실-ACP 합성 효소 I을 암호화하는 외인성 유전자의 생성물을 발현하고, 활성 아실-ACP 티오에스테라제를 암호화하는 외인성 유전자의 생성물을 발현하도록 작동 가능한 재조합 핵산을 포함한다.

[0073] 본 발명의 다른 양태는 상기 논의한 오일을 포함하는 식품을 제공한다.

[0074] 본 발명의 이와 같은 양태와 다른 양태, 그리고 구체예는 첨부된 도면, 바로 다음에 기술된 도면의 간단한 설명, 이하 본 발명의 상세한 설명에 기술되어 있으며, 이하 실시예에 예시되어 있다. 상기 및 본원의 명세서 전반에 걸쳐 논의되어 있는 특징들 중 임의의 것 또는 모두는 본 발명의 다양한 구체예에서 조합될 수 있다.

도면의 간단한 설명

[0075] 도 1은 프로토테카 트리글리세라이드 오일로부터 생산된 재생 디젤의 크로마토그램을 나타낸다.

도 2는 리시놀레산 표준 및 야생형 대조군과 비교하여, 대표적인 양성 유전자 이식 클론의 GC 정체 시간을 나타낸다.

도 3은 실시예 18에 기술한 바와 같이 표적화 유전자가 파괴된 프로토테카 모리포르미스 FADc 대립 유전자 및 표적화 유전자가 파괴되지 않은 프로토테카 모리포르미스 FADc 대립 유전자의 PstI 제한 맵을 나타낸다.

도 4는 실시예 18에 기술된 서던 블럿의 결과들을 나타낸다.

발명을 실시하기 위한 구체적인 내용

[0076] 본 발명의 예시적인 구체예는 변경된 글리세로지질 프로필을 생산하는 유지성 세포와, 이 세포로부터 생산되는 생성물을 특징으로 한다. 유지성 세포의 예로서는 제II형 지질 생합성 경로를 가지는 미생물 세포를 포함한다. 구체예는 단백질, 예를 들어 지방 아실-ACP 티오에스테라제, 지방산 불포화 효소, 케토-아실 합성 효소를 암호화하는 하나 이상의 외인성 유전자를 발현하고, 임의로는 유사한 활성을 가지는 단백질을 암호화하는 하나 이상의 내인성 유전자가 녹다운된 재조합 세포를 포함한다. 결과적으로, 몇몇 구체예는 이전에는 결코 얻을 수 없었던 천연 오일을 특징으로 한다. 본 발명은 또한 상기 세포로부터 지질 및 오일계 생성물, 예를 들어 연료, 예를 들어 바이오디젤, 재생 디젤 및 제트 연료, 식품용 오일 및 화학 물질을 제조하는 방법을 제공한다.

[0077] 본 발명의 구체예에 의하여 생산된 오일은 여러 분야들 중 교통 수단 연료, 함유 화학 제품 및/또는 식품과 화장품 산업에서 사용될 수 있다. 예를 들어 지질의 에스테르 결합 전이 반응으로 바이오디젤로서 유용한 장쇄 지방산 에스테르가 생성될 수 있다. 기타 다른 효소 및 화학 공정은 지방산, 알데하이드, 알코올, 알칸 및 알켄을 생성하도록 조작될 수 있다. 몇몇 분야에 있어서, 재생 디젤, 제트 연료 또는 기타 다른 탄화수소 화합물이 생산된다. 본 발명은 또한 지질 생산성과 수율을 증가시키고/증가시키거나, 본원에 기술된 조성물을 더욱 비용 효율적으로 생산하기 위하여 미세 조류를 배양하는 방법도 제공한다.

[0078] 본 발명의 하나의 구체예는 트리글리세라이드를 포함하는 천연 오일을 생산하거나, 또는 천연 오일로부터 생산되는 생성물을 생산하는 방법을 제공한다. 천연 오일은 식물성 오일이 아니거나 종자로부터 유래하는 오일이 아닐 수 있다. 본 발명의 방법은 맞춤 오일, 즉 세포 내 재조합 핵산이 존재함으로 인하여 변경된 지방산 프로필을 가지는 오일을 생산하도록 재조합 미생물의 세포를 배양하는 단계를 포함한다. 이후, 천연 오일은 추가로 가공되어, 식품, 연료 또는 화학 생성물로 생산될 수 있다. 세포 내 재조합 핵산은 (a) β-케토아실-ACP 합성 효소 I, β-케토아실-ACP 합성 효소 II, 스테아로일 ACP 불포화 효소, 지방산 불포화 효소 또는 아실-ACP 티오에스테라제를 암호화하는 하나 이상의 유전자에 의해 암호화되는 효소의 발현을 감소시키거나 이 효소를 발현하지

않거나(여기서, 세포는 임의로 β-케토아실-ACP 합성 효소 I, β-케토아실-ACP 합성 효소 II, 스테아로일 ACP 불포화 효소, 지방산 불포화 효소 또는 아실-ACP 티오에스테라제를 암호화하는 유전자의 복사체 2개(즉 이배체 유기체의 대립 유전자 2개)의 발현을 감소시키거나 이 복사체 2개를 발현하지 않도록 작동 가능한 재조합 핵산을 포함함); 또는 (b) 활성 β-케토아실-ACP 합성 효소 I, β-케토아실-ACP 합성 효소 II, 스테아로일 ACP 불포화 효소, 지방산 불포화 효소 또는 아실-ACP 티오에스테라제를 암호화하는 외인성 유전자의 생성물을 발현하거나; 또는 (c) β-케토아실-ACP 합성 효소 I 또는 β-케토아실-ACP 합성 효소 II를 암호화하는 하나 이상의 유전자에 의해 암호화되는 효소의 발현을 감소시키거나 이 효소를 발현하지 않고, 활성 스테아로일 ACP 불포화 효소, 지방산 불포화 효소 또는 아실-ACP 티오에스테라제를 암호화하는 외인성 유전자의 생성물을 발현하거나; 또는 (d) 스테아로일 ACP 불호화 효소 또는 지방산 불포화 효소를 암호화하는 하나 이상의 유전자에 의해 암호화되는 효소의 발현을 감소시키거나 이 효소를 발현하지 않고, 활성 β-케토아실-ACP 합성 효소 I, β-케토아실-ACP 합성 효소 II 또는 아실-ACP 티오에스테라제를 암호화하는 외인성 유전자의 생성물을 발현하도록 작동 가능하다.

[0079] 하나 이상의 지방산 불포화 효소를 암호화하는 재조합 핵산이 세포 내에 존재할 경우, 핵산은 ω-6 지방산 불포화 효소, ω-3 지방산 불포화 효소 또는 ω-6 올리에이트 불포화 효소 또는 델타 12 지방산 불포화 효소 중 하나 이상을 암호화할 수 있다.

[0080] 세포가 효소의 발현을 감소시키거나 효소를 발현하지 않도록 작동 가능한 재조합 핵산을 포함할 경우, 이와 같이 효소의 발현을 감소시키거나 효소를 발현하지 않도록 하는 것은, 효소를 암호화하는 유전자 전사체를 표적화하는 안티센스, RNAi 또는 dsRNA의 발현을 통해 이루어질 수 있거나, 또는 기타 다른 적당한 수단, 예를 들어 유도적 돌연 변이, 완전 결실 또는 부분 결실에 의해 이루어질 수 있다. 따라서, β-케토아실-ACP 합성 효소 I, β-케토아실-ACP 합성 효소 II, 스테아로일 ACP 불포화 효소 또는 지방산 불포화 효소를 암호화하는 하나 이상의 유전자에 의해 암호화되는 효소의 발현을 감소시키거나 이 효소의 발현을 지지(alimentation)하는 것은, 활성 β-케토아실-ACP 합성 효소 I, β-케토아실-ACP 합성 효소 II, 스테아로일 ACP 불포화 효소, 지방산 불포화 효소 또는 아실-ACP 티오에스테라제를 암호화하는 하나 이상의 유전자로 하나 이상의 유전자를 분할 또는 치환하는 것을 원인으로 한다.

[0081] 바람직하게, 재조합 핵산은 세포, 예를 들어 세포 염색체 또는 에피좀 내에 안정적으로 통합된다. 핵산이 안정적으로 통합된 세포의 선택은, 본원에 기술된 바와 같이 선택가능한 마커, 예를 들어 수크로스 인버타제, 항생제 저항성 유전자 또는 티아민 영양 요구성 보완을 이용하여 도움받을 수 있다.

[0082] 바람직하게, 미생물은 제II형 지방산 생합성 경로를 통해 지방산을 합성하는 것일 수 있다. 예를 들어 미생물은 미세 조류일 수 있으나, 또는 보통 제I형 지방산 생합성 경로를 가지는 것으로서, 유전자 조작 기술을 이용하여 제II형 유전 기구가 도입된 미생물(예를 들어, 오일 생산 효모)일 수도 있다. 미생물은 종속 영양 생물일 수 있으며, 특정 구체예에서 절대 종속 영양 생물일 수 있다. 미세 조류가 사용되는 경우, 이 미세 조류는 프로토테카 또는 클로렐라 종일 수 있다. 예시적 종으로서는 프로토테카 윅커해미, 프로토테카 스태그노라, 프로토테카 포르토리센시스, 프로토테카 모리포르미스, 프로토테카 조프피, 클로렐라 케슬러리, 클로렐라 루테오비리디스, 클로렐라 프로토테코이데스 및 클로렐라 불가리스를 포함한다. 수크로스 공급원료, 예를 들어 사탕수수 즙과 기타 다른 본원에 기술된 것들을 사용할 수 있도록 하기 위해서, 재조합 세포는 수크로스 인버타제 유전자를 포함하여 활성 수크로스 인버타제를 발현하는 재조합 핵산을 포함할 수 있다. 수크로스 인버타제는 미생물에 의해 배지에 분비될 수 있다.

[0083] 배양은, 예를 들어 고정 탄소원, 예를 들어 글루코스나 수크로스를 사용하는 생물 반응기 내에서 수행되는 종속 영양 배양일 수 있다. 배양은, 세포 중 트리글리세라이드가 건조 세포 중량으로 50% 이상, 60% 이상, 70% 이상, 또는 50% 내지 90%가 될 때까지 지속될 수 있다. 이는 다음에 기술된 바와 같이 질소를 제한하면서 배양하는 것을 수반할 수 있다.

[0084] 세포에 의해 생산된 오일은 세포로부터 추출될 수 있다. 하나의 구체예에서, 오일은 발색 분자 500ppm, 50ppm 또는 5ppm 미만을 포함한다. 임의로 오일은, 예를 들어 LC-MS에 의해 자체의 지방산 프로필에 대해 분석될 수 있다. 오일은 또한 실시예 19의 표 60 내지 표 63의 오일의 특성들 중 하나 이상을 가질 수도 있다.

[0085] 특정 구체예에서, 재조합 세포는 KASII 유전자에 의해 암호화되는 β-케토아실-ACP 합성 효소 II의 발현을 감소시키거나 이 효소를 발현하지 않고, 아실-ACP 티오에스테라제를 암호화하는 외인성 유전자의 생성물을 발현하도록 작동 가능한 핵산을 포함한다. 결과적으로, 세포는 C16 지방산이 40%, 50%, 60% 또는 70% 이상인 것(예를 들어, 팔미트산)을 특징으로 하는 지방산 프로필을 가지는 오일을 생산할 수 있다. 따라서, 오일은 길이가 더 단

쇄 방향으로 옮겨진 지방산 분포를 가질 수 있다. 지방산 분포의 변화는 평균 지방산 길이의 감소 또는 기타 다른 분포와 관련된 통계학적 특징에 의해 특징지어질 수 있다. 예를 들어 평균 지방산 길이를 산정하기 위해서, 트리글리세라이드를 구성하는 각각의 검출 가능 지방산의 백분율은 지방산 내 탄소의 수로 곱하여 지고, 이 곱의 합은 100으로 나누어진다. 아실-ACP 티오에스테라제를 암호화하는 외인성 유전자는 C8 내지 C16 지방 아실 사슬의 가수 분해 활성이 세포의 천연 아실-ACP-티오에스테라제보다 큰 활성 아실-ACP 티오에스테라제를 생산할 수 있다. 아실-ACP 티오에스테라제를 암호화하는 외인성 유전자는 KASII 유전자를 분할할 수 있다. 이러한 방식으로, 아실-ACP 티오에스테라제의 삽입은 또한 β-케토아실-ACP 합성 효소 II의 발현을 하나의 단계로 억제할 수도 있다. 실시예 15 및 16을 참조한다.

[0086] 다른 특정 구체예에서, 재조합 세포는 β-케토아실-ACP 합성 효소 I을 암호화하는 하나 이상의 유전자에 의해 암호화되는 효소의 발현을 감소시키거나 이 효소를 발현하지 않도록 작동 가능한 핵산을 포함한다. 결과적으로, 생산된 오일은 재조합 핵산이 결여된 비교 대상 세포보다 짧은 지방산 사슬 길이의 분포를 특징으로 하는 지방산 프로필을 가진다. 이와 같은 특징은 평균 지방산 사슬 길이가 감소한 것으로서 표현될 수 있다. 재조합 세포는 하나 이상의 FAD 유전자에 의해 암호화되는 지방산 불포화 효소의 발현을 감소시키거나 이 효소를 발현하지 않으며, 활성 스테아로일 ACP 불포화 효소를 암호화하는 스테아로일-ACP 불포화 효소 외인성 유전자의 생성물을 발현하도록 작동 가능한 핵산을 포함할 수 있다. 임의로, 핵산은 지방산 불포화 효소 유전자의 복사체(예를 들어, 대립 유전자) 다수 개에 의해 암호화되는 지방산 불포화 효소의 발현을 감소시키거나 이 효소를 발현하지 않도록 작동 가능하다. 특정 구체예에서, 스테아로일-ACP 불포화 효소 외인성 유전자는 지방산 불포화 효소 유전자의 암호화 영역 내 위치에 재조합된다. 결과적으로, 생산된 오일은 핵산이 결여된 비교 대상 세포에 의해 생산된 오일에 비하여 올레산 수준이 상승할 수 있다. 올레산은 지방산 프로필에 있어서 지방산의 50%, 60%, 70%, 80% 또는 90% 이상을 포함한다. 실시예 10을 참조한다.

[0087] 다른 특정 구체예에서, 재조합 세포는 활성 β-케토아실-ACP 합성 효소 II를 암호화하는 β-케토아실-ACP 합성 효소 II 외인성 유전자의 생성물을 발현하도록 작동 가능한 핵산을 포함한다. 결과적으로, 생산된 오일은 재조합 핵산으로 인하여 C18:1 지방산은 상승하고, C16 지방산은 감소하는 지방산 프로필을 특징으로 할 수 있다. KASII 유전자의 과발현으로 비 형질 전환 세포 내 C18 지방산의 백분율이 약 68%에서 약 84%까지 증가된 실시예 13을 참조한다. 관련 구체예에서, 증가율은 70% 이상, 75% 내지 85% 또는 70% 내지 90%이다.

[0088] 다른 특정 구체예에서, 재조합 세포는 RNA 간섭에 의해 스테아로일 ACP 불포화 효소를 암호화하는 하나 이상의 유전자에 의해 암호화되는 효소의 발현을 감소시키거나 이 효소를 발현하지 않도록 작동 가능한 핵산을 포함한다. 결과적으로, 생산된 오일은 C18:0 지방산이 증가한 것을 특징으로 하는 지방산 프로필을 가질 수 있다. C18:0 지방산은 프로필에 있어서 지방산의 50%, 60%, 70%, 80% 또는 90% 이상일 수 있다. 실시예 12를 참조한다.

[0089] 다른 특정 구체예에서, 세포는 β-케토아실-ACP 합성 효소 I, β-케토아실-ACP 합성 효소 II, 스테아로일 ACP 불포화 효소 또는 지방산 불포화 효소를 암호화하는 유전자의 복사체 2개(예를 들어, 2개의 대립 유전자)의 발현을 감소시키거나 이 유전자의 복사체를 발현하지 않도록 작동 가능한 재조합 핵산을 포함한다. 내인성 KASI 대립 유전자가 프로토테카 내에서 녹아웃된 실시예 14를 참조한다. 결과적으로, 증가율은 내인성 KASI의 파괴로 인하여 총 C14 지방산 백분율이 약 35%에서 400%까지이고, C16 지방산 백분율이 약 30%에서 50%까지인 것으로 관찰되었다.

[0090] 다른 특정 구체예에서, 세포는 활성 ω-3 지방산 불포화 효소 및/또는 ω-6 올리에이트 불포화 효소를 암호화하는 지방산 불포화 효소 외인성 유전자의 생성물을 발현하도록 작동 가능한 재조합 핵산을 포함한다. 결과적으로, 리놀레산, 리놀렌산 또는 이것들 둘 다의 수준의 상승은 상기 세포에 의해 생성되고, 세포 지질의 지방산 프로필을 통하여 확인될 수 있다. 예를 들어 세포는 리놀레산, 리놀렌산 또는 이것들 둘 다를 10%, 20%, 30%, 40% 또는 50% 이상 가질 수 있다. 예를 들어 재조합 Δ15 불포화 효소는 실시예 11에서와 같이 발현될 수 있다. 결과적으로, C18:3 지방산(즉 리놀렌산)은 약 2배 내지 17배, 또는 그 이상 증가할 수 있다.

[0091] 다른 구체예에서, 재조합 미생물의 세포가 배양된다. 세포는 활성 올리에이트 12-하이드록실라제를 암호화하는 외인성 유전자의 생성물을 발현하여 리시놀레산을 합성하도록 작동하는 재조합 핵산을 포함한다. 이 유전자는 상기 언급한 구체예들 중 임의의 것에 존재할 수 있다. 실시예 7을 참조한다. 12-하이드록실라제에 바람직한 기질은 올레산이다, 따라서, 바람직한 구체예에서, 올레산 생산을 증가시키도록 작동하는 재조합 핵산을 세포 내에 포함시킴으로써 리시놀레산을 고수율로 얻을 수 있다. 제한하는 것은 아니지만, 세포는 활성 스테아로일 ACP 불포화 효소를 암호화하는 외인성 유전자의 생성물을 발현하고, 지방산 불포화 효소를 암호화하는 하나 이상의

유전자에 의해 암호화되는 효소의 발현을 감소시키거나 이 효소를 발현하지 않거나; 또는 활성 β-케토아실-ACP 합성 효소 I을 암호화하는 외인성 유전자의 생성물을 발현하고, 활성 아실-ACP 티오에스테라제를 암호화하는 외인성 유전자의 생성물을 발현하도록 작동 가능한 재조합 핵산을 포함한다.

[0092] 본 발명의 구체예들 중 임의의 것에 따르면, 오일은 정제, 표백, 탈취, 복분해, 에스테르 결합 전이, 수소화, 가수 분해, 탈산소화, 수소 크래킹, 이성체화, 하이드록실화, 에스테르 교환, 아미드화, 설폰화 및 가황 처리 중 하나 이상에 의해 추출 및 추가 가공될 수 있다. 오일은, 예를 들어 식품용 오일, 지방산, 지방 알코올, 윤활제, 비누, 지방산 에스테르, 지방산 에톡실레이트, 지방 아민, 알킬 클로라이드, 지방 알코올 에톡실레이트, 지방 알코올 설페이트, 지방산 알카놀아미드, 설폰화된 오일, 가황 오일, 디젤, 제트, 가솔린, 블렌드스톡, 첨가제, 윤활제 또는 페인트를 형성하도록 가공될 수 있다.

[0093] 본원에 언급된 구체예들 중 임의의 것은 식품 또는 식품용 오일로서 유용할 수 있다. 전 유기체가 식품에 혼입될 수 있다. 유기체는 가공되지 않은 것, 일부가 용해된 것, 대부분이 용해된 것 또는 전체적으로 용해된 것일 수 있다. 식품 중 유지성 유기체를 제조 및 사용하는 방법은 WO2011/150411호, WO2010/12093호, WO2011/130578호 및 WO2011/130576호에 교시되어 있다. 대안적으로, 유기체로부터 추출된 오일 및 임의로 정제된 오일은 식품용 오일, 예를 들어 조리 식품, 예를 들어 스프레드, 소스, 당과 제품 및 냉동 당과 제품 중 식품용 오일 성분으로서 사용될 수 있다. 특정 구체예에서, 유지성 세포 또는 식품용 오일은 C18:0 50% 내지 70% 및 C18:1(예를 들어, 올리에이트) 20% 내지 40%를 포함한다. 다른 특정 구체예에서, 유지성 세포 또는 식품용 오일은 C16:0 50% 내지 70% 및 C18:1(예를 들어, 올리에이트) 20% 내지 40%를 포함한다.

[0094] 본 발명의 상세한 설명은 독자의 편리를 위해 여러 섹션으로 나누어진다. 섹션 I은 본원에 사용되는 용어의 정의를 제공한다. 섹션 II는 본 발명의 방법에 유용한 배양 조건의 설명을 제공한다. 섹션 III은 유전자 조작 방법 및 재료의 설명을 제공한다. 섹션 IV는 수크로스를 이용가능할 수 있게 하는 유전자 조작의 설명을 제공한다. 섹션 V는 지질 생합성을 변형시키기 위한 유전자 조작의 설명을 제공한다. 섹션 VI은 연료 및 화학물질을 제조하는 방법을 기술한다. 섹션 VII은 발명의 실시예 및 구체예를 개시한다. 본 발명의 상세한 설명에 이어서 발명의 다양한 양태 및 구체예를 예시하는 실시예가 제공된다.

[0095] **I. 정의**

[0096] 달리 정의되지 않는다면, 본원에 사용된 모든 기술적 및 과학적 용어는 본 발명이 속하는 분야의 당업자가 통상적으로 이해하는 의미를 갖는다. 본원에 사용된 바와 같이, 하기의 용어는 달리 특정되지 않는다면 이들 본래의 의미를 갖는다.

[0097] "미세 조류에서 활성(active in microalgae)"은 미세 조류에서 기능성인 핵산을 말한다. 예를 들어, 유전자 이식 미세 조류에 항생제 내성을 부여하는 항생제 내성 유전자를 구동하기 위해 사용된 프로모터는 미세 조류에서 활성이다.

[0098] "면적 백분율"은 FAME GC/FID 검출 방법을 사용하여 관찰된 피크의 면적을 말하며, 상기 방법에서 샘플 내 모든 지방산은 검출 전에 지방산 메틸 에스테르(FAME)로 전환된다. 예를 들어, 분리 피크는 C14:1과 같은 임의의 기타 다른 지방산과 비교하여 불포화성(C14:0)을 갖지 않는 14개 탄소 원자의 지방산에 대해서 관찰된다. 각각의 부류의 FAME에 대한 피크 면적은 혼합물 내 이의 백분율 조성과 정비례하고 샘플 내 존재하는 모든 피크의 합계를 기준으로 계산된다(즉, [특정 피크하의 면적/모든 측정된 피크의 총 면적] X 100). 본 발명의 오일 및 세포의 지질 프로필을 말하는 경우, "4% 이상의 C8 내지 C14"는 세포 내 또는 추출된 글리세로지질 조성물 내의 총 지방산의 4% 이상이 탄소 원자 8개, 10개, 12개 또는 14개를 포함하는 사슬 길이를 가짐을 의미한다.

[0099] "무균성"은 기타 다른 생존 유기체에 의해 오염되지 않은 유기체의 배양물이다.

[0100] "바이오디젤"은 디젤 엔진에서 연료로서 사용하기에 적당한 생물학적으로 생산된 지방산 알킬 에스테르이다.

[0101] "바이오매스"는 세포의 생장 및/또는 번식에 의해 생산된 물질이다. 바이오매스는 세포에 의해 분비되는 화합물을 포함하지만 이에 제한되지 않는, 세포외 물질뿐만 아니라 세포 및/또는 세포 내 내용물을 함유할 수 있다.

[0102] "생물 반응기"는 임의로 현탁된 세포가 배양된 인클로저(enclosure) 또는 부분적 인클로저이다.

[0103] "셀룰로스 물질"은 셀룰로스 및 임의로는 헤미셀룰로스를 포함하는 생물학적 물질이다. 이는 당, 예를 들어 글루코스 및 자일로스로 분해될 수 있으므로, 임의로는 부가의 화합물, 예를 들어 이당류, 올리고당, 리그닌, 푸르푸랄 및 기타 다른 화합물을 포함할 수 있다. 셀룰로스 물질의 공급원에 관한 비제한적 예로서는 사탕수수 바가스(sugar cane bagasses), 사탕무 펄프, 옥수수 대, 목재 칩, 톱밥 및 스위치그래스(switchgrass)를 포함

한다.

[0104] “동시배양”, 및 이의 변형어구, 예를 들어 “공동 배양하다” 및 “공동 발효시키다”는 동일한 생물 반응기에 2가지 이상의 세포 유형을 배양함을 말한다. 2가지 이상의 세포 유형은 둘 다 미세 조류와 같은 미생물일 수 있거나 상이한 세포 유형과 함께 배양되는 미세 조류 세포일 수 있다.

[0105] 본원에 사용된 “발색된 분자” 또는 “발색 불순물”은 추출된 오일에 색을 부여하는 임의의 화합물을 말한다. “발색된 분자” 또는 “발색 불순물”은, 예를 들어 엽록소 a, 엽록소 b, 라이코펜, 토코페롤, 캄페스테롤, 토코트리에놀 및 카로티노이드, 예를 들어 베타 카로틴, 루테인, 제아잔틴, 아스타잔틴을 포함한다. 이와 같은 분자들은 바람직하게 미생물 바이오매스 또는 추출된 오일 중에 500ppm 이하, 250ppm 이하, 100ppm 이하, 75ppm 이하 또는 25ppm 이하의 농도로 존재한다. 다른 구체예에서, 미생물 바이오매스 또는 추출된 오일에 존재하는 엽록소의 양은 500mg/kg 미만, 100mg/kg 미만, 10mg/kg 미만, 1mg/kg 미만, 0.5mg/kg 미만, 0.1mg/kg 미만, 0.05mg/kg 미만 또는 0.01mg/kg 미만이다.

[0106] “배양된(cultivated)”, 및 이의 변형어, 예를 들어 “배양되는(cultured)” 및 “발효된”은 선택되고/선택되거나 제어된 조건의 사용에 의한 하나 이상의 세포의 생장의 의도적인 조성(세포 크기, 세포 함량 및/또는 세포 활성의 증가) 및/또는 번식(propagation)(세포 수의 증가)을 말한다. 생장 및 번식 둘 다의 조합은 “증식(proliferation)”이라고 일컬어진다. 선택되고/선택되거나 제어된 조건의 예는 생물 반응기에서 규정된 배지(pH, 이온 강도 및 탄소원과 같은 공지된 특성을 가짐), 특정 온도, 산소분압, 이산화탄소 수준 및 생장의 사용을 포함한다. “배양된”은 자연에서 또는 다르게는 사람 간섭이 없는 미생물의 생장 또는 번식을 말하지 않고, 예를 들어 궁극적으로 화석화되어 지질학적 미정제 오일을 생성하는 유기체의 자연적인 생장은 배양이 아니다.

[0107] “불포화 효소”는 지질 합성 경로에서 이중 결합(불포화)을 트리글리세라이드 분자의 지방산 사슬에 도입하는 역할을 담당하는 효소를 말한다. 그 예로서는 스테아로일-아실기 운반 단백질 불포화 효소(SAD) 및 지방산 불포화 효소(FAD)(지방 아실 불포화 효소라고도 알려짐)를 포함하나, 이에 제한되는 것은 아니다.

[0108] “발현 벡터” 또는 “발현 구조물” 또는 “플라스미드” 또는 “재조합 DNA 구조물”은 숙주 세포 내에 핵산을 도입하기 위한 비이클이다. 핵산은 특정 핵산의 전사 및/또는 번역을 허용하는 일련의 특정 핵산 요소를 사용하여 사람의 개입을 통하여, 예를 들어 재조합 수단 또는 직접 화학 합성법에 의해 생성된 것일 수 있다. 발현 벡터는 플라스미드, 바이러스 또는 핵산 단편이나 기타 다른 적당한 비이클의 일부일 수 있다. 통상적으로, 발현 벡터는 프로모터에 작동 가능하도록 결합되어 전사될 핵산을 포함한다.

[0109] “외인성 유전자”는 (예를 들어, 형질 전환/형질 감염에 의해) 세포 내에 도입된 RNA 및/또는 단백질의 발현을 암호화하는 핵산으로서, 또한 “이식 유전자”라고도 칭하여진다. 외인성 유전자를 포함하는 세포는 부가의 외인성 유전자(들)가 도입될 수 있는 재조합 세포라고도 칭하여질 수 있다. 상기 외인성 유전자는 형질 전환될 세포에 대해 상이한 종(및 따라서 이종성) 또는 동일한 종(따라서 동종성)으로부터 유래한 것일 수 있다. 따라서, 외인성 유전자는 유전자의 내인성 복사체에 비해 세포의 게놈 내에서 상이한 위치를 차지하거나 상이한 제어 하에 있는 동종성 유전자를 포함할 수 있다. 외인성 유전자는 세포 내에서 하나 초과의 복사체에 존재할 수 있다. 외인성 유전자는 게놈(핵 또는 색소체) 내로의 삽입물로서 또는 에피좀 분자로서 세포 내에 유지될 수 있다.

[0110] “외인성으로 제공된”은 세포 배양의 배양 배지에 공급된 분자를 말한다.

[0111] 본 발명의 내용에 따라서, “지방산”은 글리세로지질 중 유리 지방산, 지방산 염 또는 지방 아실부를 의미할 것이다.

[0112] “고정 탄소원”은 본원에서 배양된 미생물에 의해 사용될 수 있는 배양 배지 내 고체 또는 액체 형태로 주위 온도 및 압력에서 존재하는 탄소를 함유하는 분자(들), 통상적으로 유기 분자이다. 따라서, 이산화탄소는 고정된 탄소원이 아니다.

[0113] 배양 조건과 관련된 “종속 영양성”은 실질적으로 빛이 존재하지 않는 조건 하에 고정된 탄소원을 이용하거나 대사하면서 배양하는 것이다.

[0114] “파쇄액”은 물리적으로 파괴된 바이오매스다.

[0115] “수소:탄소 비율”은 원자 대 원자를 기준으로 분자 내 수소 원자 대 탄소 원자의 비율이다. 상기 비율은 탄화수소 분자 내 탄소 원자 및 수소 원자의 수를 말하는데 사용될 수 있다. 예를 들어, 최고 비율의 탄화수소는 메탄 CH4(4:1)이다.

[0116] “유도성 프로모터”는 특정 자극에 응답하여 작동적으로 연결된 유전자의 전사를 매개하는 프로모터이다. 이와 같은 프로모터의 예는 pH 또는 질소 수준이 변하는 조건 하에서 유도되는 프로모터 서열일 수 있다.

[0117] “작동 가능한 연결”은 2개의 핵산 서열, 예를 들어 제어 서열(통상적으로 프로모터)과 연결된 서열(통상적으로 단백질을 암호화하는 서열, 또는 암호화 서열이라고 불림) 간의 기능적 연결이다. 프로모터는 이것이 유전자의 전사를 매개할 수 있다면 외인성 유전자와 작동가능하게 연결되어 있다.

[0118] “지질 변형 효소”는 지질의 공유 구조를 변경하거나, 아니면 세포 내 지방산 프로필을 변경시킬 수 있는 효소를 말한다. 지질 변형 효소의 예로서는 리파제, 지방 아실-ACP 티오에스테라제, 지방 아실-CoA/알데하이드 환원 효소, 지방 아실-CoA 환원 효소, 지방 알데하이드 환원 효소, 불포화 효소, 예를 들어 스테아로일 아실기 운반 단백질 불포화 효소(SAD) 및 지방 아실 불포화 효소(FAD), 및 지방 알데하이드 탈카보닐화 효소를 포함한다.

[0119] “지질 경로 효소”는 지질 대사, 즉 지질 합성, 변형 또는 분해에 역할을 하는 임의의 효소 및 운반 단백질뿐만 아니라 지질을 화학적으로 변형시키는 임의의 단백질이다.

[0120] “지질 프로필” 또는 “글리세로지질 프로필”은 사슬 길이 및/또는 포화 패턴의 관점에서, 세포 내 지방산의 분포 또는 세포로부터 유래하는 오일의 지방산의 분포를 말한다. 이와 같은 내용 중, 포화 패턴은 포화된 산 대 불포화된 산의 척도, 또는 세포의 다양한 지방산 내 이중 결합의 위치 분포를 더욱 자세히 분석한 결과를 포함할 수 있다.

[0121] “용해(lysis)”는 유기체의 일체성을 손상시키는 흔히, 기계적, 화학 물질, 바이러스 또는 삼투압 기작에 의해 적어도 일부의 세포 내 내용물을 방출하기에 충분한 생물학적 유기체의 세포질 막 및 임의로 세포 벽을 파괴하는 것이다.

[0122] “용해(lysing)”은 용해 공정이다.

[0123] “미세 조류”는 엽록체 또는 색소체를 함유하고 임의로 광합성을 수행할 수 있는 미생물 유기체, 또는 광합성을 수행할 수 있는 원핵 미생물 유기체이다. 미세 조류는 고정 탄소원으로부터 유일하게 생존할 수 있는 종속 영양 생물뿐만 아니라 에너지로서 고정 탄소원을 대사시킬 수 없는 절대 광합성 독립 영양 생물을 포함한다. 미세 조류는 2개의 독특한 세포 유형의 단순한 다중세포 광합성 미생물인, 예를 들어 볼복스(Volvox)와 같은 미생물뿐만 아니라 세포 분열 후 단시간에 자매 세포로부터 분리되는 단일 세포 유기체, 예를 들어 클라미도모나스(Chlamydomonas)를 포함한다. 미세 조류는 클로렐라, 두날리엘라(Dunaliella), 및 프로토테카와 같은 세포를 포함한다. 미세 조류는 또한 아그메넬룸(Agmenellum), 아나바에나(Anabaena), 및 피로보트리스(Pyrobotrys)와 같은 세포-세포 접착을 나타내는 기타 다른 미생물 광합성 유기체를 포함한다. 미세 조류는 또한 특정 디노플라겔레이트 조류 종 및 프로토테카 속의 종과 같은 광합성을 수행하는 능력을 상실한 절대 종속 영양 미생물을 포함한다.

[0124] 지방산과 관련된 내용 중 본원에 사용된 “중간쇄”는 C10 내지 C16 지방산을 말한다. 지방산과 관련된 내용 중 “단쇄”는 C6 내지 C10 지방산을 말하는 한편, “장쇄”은 C17 이상인 지방산을 말한다. 이와 같은 범위는 달리 지정되지 않는 한 정확히 제한되는 것으로 의도되지 않는다.

[0125] “천연 오일”은 트리글리세라이드의 지방산 프로필이 실질적으로 변경되도록, 다른 천연 또는 합성 오일 또는 분획과 오일이 배합되지 않는 경우, 유기체로부터 얻어진 대부분의 트리글리세라이드 오일을 의미할 것이다. 본원에 있어서, “분별”은 오일 자체의 지방산 프로필을 유기체에 의해 생성되지만, 달성된 지방산 프로필과 비교하였을 때 변경하는 방식으로, 오일로부터 물질을 제거하는 것을 의미한다. 오일을 대상으로 하여 자체의 트리글리세라이드 프로필을 실질적으로 변화시키지 않는 최소의 가공, 예를 들어 정제, 표백 및/또는 고무 성질 제거가 행해진 경우, 천연 오일은 유기체로부터 얻어지는 오일을 포함한다. 천연 오일은 또한 “에스테르 교환 반응이 진행되지 않은 천연 오일”일 수 있는데, 이는 천연 오일이 자체의 아실 결합에 지방산이 재분포되어 글리세롤로 가공되는 공정이 진행되지 않게 되고, 본질적으로 유기체로부터 회수되었을 때와 동일한 배열로 유지되는 경우를 의미한다.

[0126] 2개의 단백질 또는 유전자와 관련하여 “자연적으로 공동 발현된”은 단백질 또는 이들의 유전자가 이들이 유래된 조직 또는 유기체에서 자연적으로 공동 발현됨을 의미하며, 이는 예를 들어 2개의 단백질을 암호화하는 유전자가 공통된 조절 서열의 제어 하에 있기 때문이거나 이들이 동일한 자극에 응답하여 발현되기 때문이다.

[0127] “삼투압 쇼크”는 삼투압의 급작스런 감소 후 용액 내 세포의 파열이다. 삼투압 쇼크는 때때로 상기 세포의 세포 성분을 용액 내로 방출하도록 유도된다.

[0128] “다당류 분해 효소”는 임의의 다당류의 가수 분해 또는 당화를 촉매할 수 있는 임의의 효소이다. 예를 들어, 셀룰라제는 셀룰로스의 가수 분해를 촉매한다.

[0129] “다당류” 또는 “글리칸”은 글리코시드 결합에 의해 서로 함께 연결된 단당류로 구성된 탄수화물이다. 셀룰로스는 특정 식물 세포 벽을 구성하는 다당류다. 셀룰로스는 보다 큰 이당류 및 올리고당뿐만 아니라 자일로스 및 글루코스와 같은 단당류를 생성하도록 효소에 의해 해중합될 수 있다.

[0130] “프로모터”는 핵산의 전사를 지시하는 핵산 제어 서열이다. 본원에 사용된 바와 같이, 프로모터는 전사 개시 부위 근처에 필수 핵산 서열, 예를 들어 중합 효소 II형 프로모터의 경우에, TATA 요소를 포함한다. 프로모터는 또한 임의로 전사 개시 부위로부터 수천 염기쌍 정도로 멀리 위치할 수 있는 원거리 인핸서 또는 리프레서 요소를 포함한다.

[0131] “재조합체”는 외인성 핵산의 도입 또는 고유 핵산의 변경으로 인해 변형된 세포, 핵산, 단백질 또는 벡터이다. 따라서, 예를 들어 재조합 세포는 천연(비재조합체) 형태의 세포 내에서 발견되지 않는 유전자를 발현할 수 있거나 비재조합 세포에 의해 발현되는 유전자와는 상이하게 천연 유전자를 발현할 수 있다. 재조합 세포는 유전자 생성물 또는 세포 내 활성 유전자 생성물의 수준을 감소시키는 억제 요소, 예를 들어 돌연 변이, 녹아웃, 안티센스, 간섭 RNA(RNAi) 또는 dsRNA를 암호화하는 재조합 핵산(이에 제한되는 것은 아님)을 포함할 수 있다. “재조합 핵산”은 예를 들어, 중합 효소, 리가아제, 엑소뉴클레아제 및 엔도뉴클레아제를 사용하는 핵산의 조작에 의해 일반적으로 시험관내 형성된 본래의 핵산이거나, 다르게는 통상적으로 자연적으로 발견되지 않는 형태로 존재한다. 재조합 핵산은, 예를 들어 2개 이상의 핵산을 작동가능하게 연결시키도록 하여 생성될 수 있다. 따라서, 통상적으로 자연적으로 연결되어 있지 않은 DNA 분자를 연결함에 의해 시험관내 형성된 분리된 핵산 또는 발현 벡터는 둘 다 본 발명의 목적을 위한 재조합체인 것으로 고려된다. 일단 재조합 핵산이 제조되고 숙주 세포 또는 유기체로 도입되면, 이것은 숙주 세포의 생체내 세포 기구를 사용하여 복제할 수 있지만, 일단 재조합적으로 생성되면 당해 핵산은 후속적으로 세포 내에서 복제된다 하더라도 여전히 본 발명의 목적을 위한 재조합체인 것으로 고려된다. 유사하게, “재조합 단백질”은, 즉 재조합 핵산의 발현을 통해 재조합 기술을 사용하여 제조된 단백질이다.

[0132] “재생 디젤”은 지질의 수소화 및 탈산소화를 통해 천연 오일로부터 생성된 알칸(예를 들어, C10:0, C12:0, C14:0, C16:0 및 C 18:0)의 혼합물이다.

[0133] “당화”는 바이오매스, 통상적으로 셀룰로스 또는 리그노셀룰로스 바이오매스를 글루코스 및 자일로스와 같은 단량체성 당으로 전환시키는 공정이다. “당화된” 또는 “해중합된” 셀룰로스 물질 또는 바이오매스는 당화를 통해 단량체 당으로 전환되는 셀룰로스 물질 또는 바이오매스를 말한다.

[0134] “푸르푸랄 종”은 동일한 염기성 구조 특징을 보유한 2-푸란카복스알데하이드 또는 유도체이다.

[0135] 본원의 구체예에 따르면(반드시 선행 기술에 관해 논의된 사항에 의하지는 않음) 재조합 균주를 형성하기 위한 균주의 형질 전환과 관련하였을 때, “안정적인” 또는 “안정적으로 통합된”은 균주의 세포가 10세대 이상 동안 재조합 핵산을 보유하고 있는 경우를 의미할 것이다. 예를 들어 재조합 균주가 선택압의 존재 하에서 배양을 가능하게 하는 선택가능한 마커를 가지는 경우, 재조합 핵산은 선택압의 부재시 10세대 배양된 후에도 유지된다.

[0136] “수크로스 이용 유전자”는 발현되는 경우 에너지원으로서 수크로스를 이용하는 세포의 능력을 도와주는 유전자이다. 수크로스 이용 유전자에 의해 암호화된 단백질은 본원에서 “수크로스 이용 효소”로 지칭되고 수크로스 수송체, 수크로스 인버타제, 및 글루코키나제 및 프럭토키나제와 같은 헥소키나제를 포함한다.

[0137] **II. 배양**

[0138] 본 발명은 일반적으로 미생물, 특히 제II형 지방산 생합성 경로를 가지는 유지성 미생물, 예를 들어 트리글리세라이드를 생산하는 미세 조류의 배양에 관한 것이다. 하나의 구체예에서, 미생물은 절대 종속 영양 생물이다. 미생물은 예를 들어 본원에 개시된 유전자 조작 방법을 바탕으로 생산된 재조합 미생물일 수 있다. 독자들의 편의를 위하여, 본 섹션은 하위 섹션들로 세분화된다. 하위 섹션 1은 미생물의 종과 균주를 기술한다. 하위 섹션 2는 배양에 유용한 생물 반응기를 기술한다. 하위 섹션 3은 배양을 위한 배지를 기술한다. 하위 섹션 4는 본 발명의 예시적인 배양 방법에 따른 오일 생성을 기술한다.

[0139] **1. 미생물 종 및 균주**

[0140] 비록 이하에 제시된 예시적인 구체예들이 다수의 미생물에 적용될 수 있지만, 프로토테카는 지질의 생산에 사용

되는 바람직한 미생물이다. 중요한 점은, 예를 들어 프로토테카를 사용하는, 본원에 기술된 유전자 조작 방법은 기타 다른 미생물(예를 들어, 클로렐라 소로키니아나(Chlorella sorokiniana), 클로렐라 불가리스, 클로렐라 엘립소이데아(Chlorella ellipsoidea), 클로렐라 케슬러리, 두날리엘라 터티올렉타(Dunaliella tertiolecta), 볼복스 카르테리(Volvox carteri), 해마토코커스 플루비알리스(Haematococcus pluvialis), 클로스테리움 페라세로섬-스트리고섬-리토랄레 복합체(Closterium peracerosum-strigosum-littorale complex), 두날리엘라 비리디스(Dunaliella viridis), 두날리엘라 살리나(Dunaliella salina), 고니움 펙토랄레(Gonium pectorale), 파에오닥틸룸 트리코르누툼(Phaeodactylum tricornutum), 카에토세로스(Chaetoceros), 실린드로테카 푸시포르미스(Cylindrotheca fusiformis), 암피디늄 종(Amphidinium sp.), 심비오디늄 마이크로아드리악티큠(Symbiodinium microadriacticum), 나노클로롭시스(Nannochloropsis), 시클로텔라 크립티카(Cyclotella cryptica), 나비큘라 사프로필라(Navicula saprophila) 또는 탈라시오시라 슈도나나(Thalassiosira pseudonana))에 적용 가능하다는 것이다.

[0141] 절대 종속 영양 미세 조류, 예를 들어 프로토테카로부터 얻어진 지질 또는 오일은 일반적으로 색소 함량이 낮을 수 있으며(예를 들어, 엽록소 및 임의의 카로티노이드의 수준은 검출이 안 될 정도로 낮을 수 있으며, 예를 들어, 발색 분자, 발색 불순물 또는 엽록소와 카로티노이드 농도의 합이 500ppm, 50ppm 또는 5ppm 미만일 수 있으며), 임의의 경우에는 기타 다른 미세 조류로부터 유래하는 지질보다 색소를 훨씬 적게 함유한다. 더욱이, 본 발명에 의해 제공된 재조합 프로토테카 세포는 기타 다른 미생물로부터 지질을 생산하는 것과 비교하여 보다 큰 수율 및 효율로 저렴하게 지질을 생산하는데 사용될 수 있다. 본 발명의 방법에 사용하기 위한 예시적인 프로토테카 균주가 포함된다. 추가로, 상기 미세 조류는 종속 영양적으로 생장하고 프로토테카 윅커해미, 프로토테카 스태그노라(UTEX 327을 포함함), 프로토테카 포르토리센시스, 프로토테카 모리포르미스(UTEX 균주 1441, 1435를 포함함), 및 프로토테카 조프피로서 유전자 조작될 수 있다. 프로토테카 속의 종은 절대 종속 영양 생물이다.

[0142] 오일, 연료 및 함유 화학 제품을 생산하는데 적당한 지질 또는 탄화수소를 생산하는 것 이외에, 본 발명의 구체예에 사용될 미생물을 선택하는 것에 영향을 미치는 고려 사항으로서는, (1) 고 지질 함량(세포 중량 백분율); (2) 생장의 용이성; (3) 유전자 조작의 용이성; 및 (4) 바이오매스 가공의 용이성을 포함한다. 특정 구체예에서, 야생형 미생물 또는 유전자 조작된 미생물은 지질이 40% 이상, 45% 이상, 50% 이상, 55% 이상, 60% 이상, 65% 이상 또는 70% 이상 또는 그 이상인 세포를 생산한다. 기타 특정 구체예에서, 야생형 미생물 또는 유전자 조작된 미생물은 40% 내지 80% 또는 50% 내지 90%의 트리글리세라이드를 포함하는 세포를 생산한다. 바람직한 유기체는 종속 영양 조건 하에서(빛이 부재하고 당이 존재하는 조건 하에서) 생장한다.

[0143] 본 발명을 실시하는데 사용될 수 있는 조류의 예로서는 이하 표 1에 나열된 조류를 포함하나, 이에 제한되는 것은 아니다.

표 1

[0144] 조류의 예

아크난테스 오리엔탈리스(Achnanthes orientalis), 아그메넬룸, 암피프로라 히알린(Amphiprora hyaline), 암포라 코페이포르미스(Amphora coffeiformis), 암포라 코페이포르미스 리니아(Amphora coffeiformis linea), 암포라 코페이포르미스 펑타타(Amphora coffeiformis punctata), 암포라 코페이포르미스 타일로리(Amphora coffeiformis taylori), 암포라 코페이포르미스 테뉴이스(Amphora coffeiformis tenuis), 암포라 델리카티시마(Amphora delicatissima), 암포라 델리카티시마 카피타타(Amphora delicatissima capitata), 암포라 종(Amphora sp.), 아나바에나, 안키스트로데스무스(Ankistrodesmus), 안키스트로데스무스 팔카투스(Ankistrodesmus falcatus), 보에켈로비아 후글란디(Boekelovia hooglandii), 보로디넬라 종(Borodinella sp.), 보트리오코커스 브라우니(Botryococcus braunii), 보트리오코커스 수데티쿠스(Botryococcus sudeticus), 카르테리아(Carteria), 카에토세로스 그라실리스(Chaetoceros gracilis), 카에토세로스 무엘러리(Chaetoceros muelleri), 카에토세로스 무엘러리 서브살섬(Chaetoceros muelleri subsalsum), 카에토세로스 종(Chaetoceros sp.), 클로렐라 아니트라타(Chlorella anitrata), 클로렐라 안타르크티카(Chlorella Antarctica), 클로렐라 아우레오비리디스(Chlorella aureoviridis), 클로렐라 칸디다(Chlorella candida), 클로렐라 캡슐라테(Chlorella capsulate), 클로렐라 데시카테(Chlorella desiccate), 클로렐라 엘립소이데아(Chlorella ellipsoidea), 클로렐라 에머소니(Chlorella emersonii), 클로렐라 푸스카(Chlorella fusca), 클로렐라 푸스카 변종 바큐올라타(Chlorella fusca var. vacuolata), 클로렐라 글루코트로파(Chlorella glucotropha), 클로렐라 인

퓨시오늄(Chlorella infusionum), 클로렐라 인퓨시오늄 변종 악토필라(Chlorella infusionum var. actophila), 클로렐라 인퓨시오늄 변종 옥세노필라(Chlorella infusionum var. auxenophila), 클로렐라 케슬러리, 클로렐라 로보포라(Chlorella lobophora)(균주 SAG 37.88), 클로렐라 루테오비리디스, 클로렐라 루테오비리디스 변종 아우레오비리디스(Chlorella luteoviridis var. aureoviridis), 클로렐라 루테오비리디스 변종 루테센스(Chlorella luteoviridis var. lutescens), 클로렐라 미니아타(Chlorella miniata), 클로렐라 미뉴티시마(Chlorella minutissima), 클로렐라 뮤타빌리스(Chlorella mutabilis), 클로렐라 녹터나(Chlorella nocturna), 클로렐라 파르바(Chlorella parva), 클로렐라 포토필리아(Chlorella photophila), 클로렐라 프링스헤이미(Chlorella pringsheimii), 클로렐라 프로토테코이데스(Chlorella protothecoides)(UTEX 균주 1806, 411, 264, 256, 255, 250, 249, 31, 29, 25, 및CCAP 균주 211/17 및 211/8d 중 임의의 것을 포함함), 클로렐라 프로토테코이데스 변종 아시디콜라(Chlorella protothecoides var. acidicola), 클로렐라 레귤라리스(Chlorella regularis), 클로렐라 레귤라리스 변종 미니마(Chlorella regularis var. minima), 클로렐라 레귤라리스 변종 움브리카타(Chlorella regularis var. umbricata), 클로렐라 레이시글리(Chlorella reisiglii), 클로렐라 사카로필라(Chlorella saccharophila), 클로렐라 사카로필라 변종 엘립소이데아(Chlorella saccharophila var. ellipsoidea), 클로렐라 살리나(Chlorella salina), 클로렐라 심플렉스(Chlorella simplex), 클로렐라 소로키니아나, 클로렐라 종(Chlorella sp.), 클로렐라 스파에리카(Chlorella sphaerica), 클로렐라 스티그마토포라(Chlorella stigmatophora), 클로렐라 바니엘리(Chlorella vanniellii), 클로렐라 불가리스, 클로렐라 불가리스, 클로렐라 불가리스 에프. 테르티아(Chlorella vulgaris f. tertia), 클로렐라 불가리스 변종 오토트로피카(Chlorella vulgaris var. autotrophica), 클로렐라 불가리스 변종 비리디스(Chlorella vulgaris var. viridis), 클로렐라 불가리스 변종 불가리스(Chlorella vulgaris var. vulgaris), 클로렐라 불가리스 변종 불가리스 에프. 테르티아(Chlorella vulgaris var. vulgaris f. tertia), 클로렐라 불가리스 변종 불가리스 에프. 비리디스(Chlorella vulgaris var. vulgaris f. viridis), 클로렐라 잔텔라(Chlorella xanthella), 클로렐라 조핀지엔시스(Chlorella zofingiensis), 클로렐라 트레복시오이데스(Chlorella trebouxioides), 클로렐라 불가리스, 클로로코쿰 인퓨시오늄(Chlorococcum infusionum), 클로로코쿰 종(Chlorococcum sp.), 클로로고늄(Chlorogonium), 크로오모나스 종(Chroomonas sp.), 크리소스파에라 종(Chrysosphaera sp.), 크리코스파에라 종(Cricosphaera sp.), 크립토모나스 종(Cryptomonas sp.), 시클로텔라 크립티카(Cyclotella cryptica), 시클로텔라 메네기니아나(Cyclotella meneghiniana), 시클로텔라 종(Cyclotella sp.), 두날리엘라 종(Dunaliella sp.), 두날리엘라 바르다윌(Dunaliella bardawil), 두날리엘라 비오큘라타(Dunaliella bioculata), 두날리엘라 그라뉼라테(Dunaliella granulate), 두날리엘라 마리타임(Dunaliella maritime), 두날리엘라 미뉴타(Dunaliella minuta), 두날리엘라 파르바(Dunaliella parva), 두날리엘라 페이르세이(Dunaliella peircei), 두날리엘라 프리몰렉타(Dunaliella primolecta), 두날리엘라 살리나(Dunaliella salina), 두날리엘라 테리콜라(Dunaliella terricola), 두날리엘라 터티올렉타(Dunaliella tertiolecta), 두날리엘라 비리디스(Dunaliella viridis), 두날리엘라 터티올렉타, 에레모스파에라 비리디스(Eremosphaera viridis), 에레모스파에라 종(Eremosphaera sp.), 엘립소이돈 종(Ellipsoidon sp.), 유글레나(Euglena), 프란세이아 종(Franceia sp.), 프라길라리아 크로토넨시스(Fragilaria crotonensis), 프라길라리아 종(Fragilaria sp.), 글레오캡사 종(Gleocapsa sp.), 글로에오탐니온 종(Gloeothamnion sp.), 하이메노모나스 종(Hymenomonas sp.), 이소크리시스 에이에프에프. 갈바나(Isochrysis aff. galbana), 이소크리시스 갈바나(Isochrysis galbana), 레포신클리스(Lepocinclis), 마이크락티늄(Micractinium), 마이크락티늄(UTEX LB 2614), 모노라피듐 미뉴텀(Monoraphidium minutum), 모노라피듐 종(Monoraphidium sp.), 나노클로리스 종(Nannochloris sp.), 나노클로롭시스 살리나(Nannochloropsis salina), 나노클로롭시스 종(Nannochloropsis sp.), 나비큘라 아셉타타(Navicula acceptata), 나비큘라 비스칸테라에(Navicula biskanterae), 나비큘라 슈도테넬로이데스(Navicula pseudotenelloides), 나비큘라 펠리큘로사(Navicula pelliculosa), 나비큘라 사프로필라(Navicula saprophila), 나비큘라 종(Navicula sp.), 네프로클로리스 종(Nephrochloris sp.), 네프로셀미스 종(Nephroselmis sp.), 니츠시아 커뮤니스(Nitschia communis), 니츠시아 알렉산드리나(Nitzschia alexandrina), 니츠시아 커뮤니스(Nitzschia communis), 니츠시아 디시파타(Nitzschia dissipata), 니츠시아 프러스튤럼(Nitzschia frustulum), 니츠시아 한츠시아나(Nitzschia hantzschiana), 니츠시아 인컨스피큐아(Nitzschia inconspicua), 니츠시아 인터미디아(Nitzschia intermedia), 니츠시아 마이크로세팔라(Nitzschia microcephala), 니츠시아 푸실라(Nitzschia pusilla), 니츠시아 푸실라 엘립티카(Nitzschia pusilla elliptica), 니츠시아 푸실라 모노엔시스(Nitzschia pusilla monoensis), 니츠시아 쿼드랭귤라(Nitzschia quadrangular), 니츠시아 종(Nitzschia sp.), 오크로모나스 종(Ochromonas sp.), 오오시스티스 파르바(Oocystis parva), 오오시스티스 푸실라(Oocystis pusilla), 오오시스티스 종(Oocystis sp.), 오실라토리아 림네티카(Oscillatoria limnetica), 오실라토리아 종(Oscillatoria sp.), 오실라토리아 수브레비스(Oscillatoria subbrevis), 파셰리아 악시도필라(Pascheria acidophila), 파블로바 종(Pavlova sp.), 파구스(Phagus), 포르미디움(Phormidium), 플라티모나스 종(Platymonas sp.), 플루로크리시스 카르테라에(Pleurochrysis carterae), 플루로크리시스 덴타테(Pleurochrysis dentate), 플루로크리시스 종(Pleurochrysis sp.), 프로토테카 윅커해미, 프로토테카 스태그노라, 프로토테카 포르토리센시스, 프로토테카 모리포르미스, 프로토테카 조프피, 피라미모나스 종(Pyramimonas sp.), 피로보트리스, 사르시노이드 크리소피트(Sarcinoid chrysophyte), 세네데스무스 아르마터스(Scenedesmus armatus), 스피로자이라(Spirogyra), 스피룰리나 플라텐시스(Spirulina platensis), 스티코코커스 종(Stichococcus sp.), 시네코코커스 종(Synechococcus sp.), 테트라에드론(Tetraedron), 테트라셀미스 종(Tetraselmis sp.), 테트라셀미스 수에시카(Tetraselmis suecica), 탈라시오시라 웨이스플로기(Thalassiosira weissflogii) 및 비리디엘라 프리데리시아나(Viridiella fridericiana)

[0146] 2. 생물 반응기

[0147] 미생물은 유전자 조작을 수행하고 탄화수소(예를 들어, 지질, 지방산, 알데하이드, 알코올 및 알칸)를 생성할 2가지 목적을 위해 배양한다. 전자 유형의 배양은 소규모 및 처음에는 적어도 출발 미생물이 생장할 수 있는 조건 하에서 수행한다. 탄화수소를 생산하기 위한 배양은 생물 반응기에서 통상적으로 대규모(예를 들어, 10,000L, 40,000L, 100,000L 또는 보다 대형의 생물 반응기)로 수행한다. 미세 조류, 예를 들어 프로토테카 종은 통상적으로 생물 반응기내 액체 배지 중에서 본 발명의 방법으로 배양한다. 통상적으로, 생물 반응기는 빛이 입사되지 않도록 한다.

[0148] 생물 반응기 또는 발효기를 사용하여 다양한 단계의 세포의 생리학적 사이클을 통해 미세 조류 세포를 배양한다. 생물 반응기는 종속 영양 생장 및 번식 방법에 사용하기 위한 많은 이점을 제공한다. 바이오매스를 생산하기 위해 미세 조류를 바람직하게 액체 중에서, 예를 들어 현탁 배양으로 대량으로 발효시킨다. 강철 발효기와 같은 생물 반응기는 매우 큰 배양 용적(40,000리터 이상의 용량을 갖는 생물 반응기가 본 발명의 다양한 구체예에서 사용됨)을 수용할 수 있다. 생물 반응기는 또한 통상적으로 온도, pH, 산소분압 및 이산화탄소 수준과 같은 배양 조건의 제어를 허용한다. 예를 들어, 생물 반응기는 통상적으로, 예를 들어 관에 부착된 포트를 사용하여 산소 또는 질소와 같은 가스 성분이 액체 배양물을 통해 기포 주입될 수 있도록 하는 형태를 취할 수 있다. 배양 배지의 pH, 미량 요소의 종류 및 농도 및 기타 다른 배지 성분과 같은 기타 다른 배양 파라미터가 또한 생물 반응기를 사용하여 보다 용이하게 조작될 수 있다.

[0149] 회전 블레이드 및 임펠러, 진동(rocking) 기구, 교반 막대, 가압 가스 주입용 수단과 같은 장치가 장착된 생물 반응기를 사용하여 배양물이 혼합되도록 할 수 있다. 혼합은 연속적이거나 간헐적일 수 있다. 예를 들어, 몇몇 구체예에서, 상기 세포의 수에 있어서 원하는 증가가 이루어질 때까지 가스 주입 및 배지 주입의 난류 체계는 미세 조류의 번식 동안에 유지되지 않는다.

[0150] 생물 반응기 포트를 사용하여 가스, 고체, 반고체 및 액체를, 미세 조류를 함유하는 생물 반응기 챔버에 도입하거나 추출할 수 있다. 많은 생물 반응기는 하나 초과의 포트(예를 들어, 배지 주입을 위한 하나 및 샘플링을 위한 또 다른 하나)를 갖는 한편, 단지 하나의 물질이 포트를 출입할 필요는 없다. 예를 들어, 포트를 사용하여 배양 배지가 생물 반응기로 유입될 수 있도록 하고 이후에 샘플링, 가스 주입, 가스 배출 또는 기타 다른 목적을 위해 사용될 수 있다. 바람직하게, 샘플링 포트는 배양물의 무균 특성을 변경, 손상시키지 않고 반복적으로 사용될 수 있다. 샘플링 포트는 샘플의 흐름이 중지되고 개시되거나 연속 샘플링의 수단을 제공하도록 하는 밸브 또는 기타 다른 장치로 구성될 수 있다. 생물 반응기는 통상적으로 배양물의 접종을 가능하게 하는 하나 이상의 포트를 갖고 상기 포트는 또한 배지 또는 가스 주입과 같은 기타 다른 목적을 위해 사용될 수 있다.

[0151] 생물 반응기 포트는 미세 조류의 배양물의 가스 함량이 조작되도록 할 수 있다. 설명하자면, 생물 반응기 용적의 일부는 액체보다는 가스일 수 있고, 생물 반응기의 가스 주입구는 가스가 생물 반응기로 펌핑될 수 있도록 한다. 생물 반응기 내로 이롭게 펌핑될 수 있는 가스는 공기, 산소, 공기/CO2 혼합물, 비활성 가스(예를 들어, 아르곤) 및 기타 다른 가스를 포함한다. 생물 반응기는 통상적으로 사용자가 생물 반응기로의 가스 주입 속도를 조절할 수 있도록 구성될 수 있다. 상기에 언급된 바와 같이, 생물 반응기로의 가스 주입 증가를 사용하여 배양물의 혼합을 증가시킬 수 있다.

[0152] 증가된 가스 유입은 또한 배양물의 탁도에 영향을 미친다. 난기류는 수성 배양 배지의 수위 하부에 가스 주입 포트를 위치시켜 생물 반응기로 주입되는 가스가 배양물 표면으로 기포 주입되도록 함으로써 성취할 수 있다. 하나 이상의 가스 배출 포트는 가스가 배출되도록 하여 생물 반응기 내 압력 증강을 차단한다. 바람직하게, 가스 배출 포트는 생물 반응기로 오염 미생물이 들어가는 것을 차단하는 “원-웨이” 밸브와 연결된다.

[0154] 3. 배지

[0155] 미생물 배양 배지는 통상적으로 고정 질소원, 고정 탄소원, 미량 요소, 임의로 pH 유지를 위한 완충제 및 포스페이트(통상적으로 포스페이트 염으로서 제공됨)와 같은 성분들을 함유한다. 기타 다른 성분은 특히 해양 미세 조류를 위해 염화나트륨과 같은 염을 포함할 수 있다. 질소원은, 예를 들어 제한 없이 분자 질소, 질산염(nitrate), 질산 염(nitrate salt), 암모니아(순수 형태 또는 염 형태, 예를 들어 (NH4)2SO4 및 NH4OH), 단백질, 대두 밀, 옥수수 침지액, 및 효모 추출물을 포함하는 유기 및 무기 질소원을 함유한다. 미량 요소의 예로서는 아연, 붕소, 코발트, 구리, 망간, 및 몰리브덴을 포함하고, 예를 들어 이의 각각의 형태는 $ZnCl_2$, H_3BO_3, $CoCl_2 \cdot 6H_2O$, $CuCl_2 \cdot 2H_2O$, $MnCl_2 \cdot 4H_2O$ 및 $(NH_4)_6Mo_7O_{24} \cdot 4H_2O$이다.

[0156] 본 발명의 방법에 따라 유용한 미생물은 전세계적으로 다양한 장소 및 환경에서 발견된다. 기타 다른 종 및 이들의 궁극적인 진화적 분기점으로부터의 분리 결과로서, 최적의 생장 및 지질 및/또는 탄화수소 성분들의 최적

의 생산을 위한 특정 생장 배지는 예측하기 어려울 수 있다. 몇몇 경우에, 특정 미생물 균주는 몇몇 억제 성분의 존재 또는 특정 미생물 균주에 의해 요구되는 일부 필수 영양 요구성 물질의 부재 때문에 특정 생장 배지 상에서 생장하지 않을 수 있다.

[0157] 고체 및 액체 생장 배지는 일반적으로 다양한 공급원으로부터 이용 가능하며, 광범위하게 다양한 미생물 균주에 적합한 특정 배지를 제조하기 위한 지침은, 예를 들어 조류의 배양 수집물을 위하여 택사스 대학교 오스틴 캠퍼스(University of Texas at Austin; 1 University Station A6700, Austin, Texas, 78712-0183, (UTEX))에서 운영하는 온라인 사이트(http://www.utex.org/)에서 찾을 수 있다. 예를 들어, 다양한 담수 및 염수 배지는 본원에 참조로서 포함된 문헌[PCT Pub. No. 2008/151149]에 기재된 것들을 포함한다.

[0158] 특정 예에서, 프로테오스 배지(Proteose Medium)는 무균 배양에 적합하고, 1L 용적의 배지(pH 약 6.8)는 1리터의 Bristol 배지에 1g의 프로테오스 펩톤을 첨가하여 제조할 수 있다. Bristol 배지는 수용액 중에 2.94mM NaNO3, 0.17mM CaCl2·2H2O, 0.3mM MgSO4·7H2O, 0.43mM, 1.29mM KH2PO4, 및 1.43mM NaCl을 포함한다. 1.5% 한천 배지에 대해, 15g의 한천을 1L의 용액에 첨가할 수 있다. 상기 용액을 덮고 오토클레이브 처리하며, 이어서 사용 전에 냉장 온도에 저장한다. 또 다른 예는 프로토테카 분리 배지(PIM)이며, 이는 10g/L의 수소칼륨 프탈레이트(KHP), 0.9g/L의 수산화나트륨, 0.1g/L의 황산마그네슘, 0.2g/L의 인산수소칼륨, 0.3g/L의 염화암모늄, 10g/L의 글루코스, 0.001g/L의 티아민 염산염, 20g/L 한천, 0.25g/L의 5-플루오로사이토신을 pH 범위 5.0 내지 5.2로 포함한다(문헌[Pore, 1973, App. Microbiology, 26: 648-649] 참조). 본 발명의 방법에 사용하기에 적합한 기타 다른 배지는 상기의 URL에 문의하거나 미생물 배양물(예를 들어, SAG, CCAP, 또는 CCALA)을 유지하는 기타 다른 기관에 문의하여 용이하게 확인할 수 있다. SAG는 괴팅겐 대학교(University of _; _, Germany)에서의 조류의 배양 수집물을 말하고, CCAP는 스코틀랜드 해양과학협회(Scottish Association for Marine Science; Scotland, United Kingdom)에 의해 관리되는 조류 및 원생동물의 배양 수집물을 말하며, CCALA는 식물연구원(Institute of Botany; _, Czech Republic)의 조류 연구소의 배양 수집물을 말한다. 추가로, 미국 특허 제5,900,370호는 프로토테카 종의 종속 영양 발효에 적합한 배지 제형 및 조건을 기술한다.

[0159] 오일 생산을 위해, 고정 탄소원의 선택이 중요한데, 그 이유는 오일 생산을 경제적으로 하기 위해서는 고정 탄소원의 비용이 충분히 낮아야 하기 때문이다. 따라서, 적합한 탄소원은, 예를 들어 아세테이트, 플로리도사이드, 프럭토스, 갈락토스, 글루쿠론산, 글루코스, 글리세롤, 락토스, 만노스, N-아세틸글루코사민, 람노스, 수크로스, 및/또는 자일로스를 포함할 수 있고, 상기 화합물을 함유하는 공급원료의 선택이 본 발명의 구체예의 방법의 중요한 양태이다. 본 발명의 방법에 따라 유용한 적합한 공급원료는, 예를 들어 흑액, 옥수수 전분, 해중합된 셀룰로스 물질, 유청, 당밀, 감자, 수수, 수크로스, 사탕무, 사탕수수, 쌀 및 밀을 포함할 수 있다. 탄소원은 또한 수크로스와 해중합된 사탕무 펄프의 혼합물과 같은 혼합물로서 제공될 수 있다. 하나 이상의 탄소원(들)은 하나 이상의 외인성으로 제공된 고정 탄소원(들)의 약 50 μM 이상, 약 100 μM 이상, 약 500 μM 이상, 약 5mM 이상, 약 50mM 이상, 및 약 500mM 이상의 농도로 제공될 수 있다. 발효를 위한 공급원료로서 고도로 농축된 탄소원이 바람직하다. 예를 들어 몇몇 구체예에서, 배양 단계 이전 공급원료 중 글루코스의 수준은 300g/L 이상, 400g/L 이상, 500g/L 이상 또는 600g/L 이상, 또는 그 이상의 글루코스 수준으로 유가식 배양(fed batch cultivation)에 첨가되는데, 여기서 고도로 농축된 고정 탄소원은 세포가 생장하고 지질을 축적함에 따라서 시간 경과에 따라 세포에 공급된다. 다른 구체예에서, 배양 이전 수크로스는 500g/L 이상, 600g/L 이상, 700g/L 이상, 800g/L 이상 또는 그 이상의 수크로스 수준으로 유가 배양에 첨가되는데, 여기서 고도로 농축된 고정 탄소원은 세포가 생장하고 지질을 축적함에 따라서 시간 경과에 따라 세포에 공급된다. 고도로 농축된 고정 탄소원, 예를 들어 수크로스의 비제한적인 예로서는 진한 감자 당즙, 사탕수수 즙, 사탕무 즙 및 당밀을 포함한다. 본 발명의 목적을 위해 특히 관심이 있는 탄소원은 셀룰로스(해중합된 형태), 글리세롤, 수크로스 및 수수를 포함하고, 이들 각각은 하기에서 보다 상세히 논의된다.

[0160] 본 발명에 따라, 미생물은 공급원료로서 해중합된 셀룰로스 바이오매스를 사용하여 배양할 수 있다. 셀룰로스 바이오매스(예를 들어, 여물, 예를 들어 옥수수 대)는 저렴하고 용이하게 입수할 수 있지만, 상기 공급원료는 효모 생장에 억제 작용을 하는 것으로 밝혀졌고, 효모는 셀룰로스 물질로부터 생성된 5탄당(예를 들어, 헤미셀룰로스 유래의 자일로스)을 사용할 수 없다. 대조적으로, 적어도 일부 미세 조류는 가공된 셀룰로스 물질 상에서 생장할 수 있다. 셀룰로스 물질은 일반적으로 약 40% 내지 60%의 셀룰로스, 약 20% 내지 40%의 헤미셀룰로스 및 10% 내지 30%의 리그닌을 포함한다.

[0161] 셀룰로스 물질은 농작물, 즉, 식물 일부, 1차 식품 또는 섬유 생성물과 함께 야생에서 제거되지 않는 주로 줄기 및 잎뿐만 아니라 초본성 및 목재 에너지 작물로부터의 잔류물을 포함한다. 이의 예로서는 농업 폐기물, 예를 들어 사탕수수 바가스, 쌀겨, 옥수수 섬유(줄기, 잎, 껍질 및 속대를 포함함), 밀짚, 쌀짚, 사탕무 펄프, 감귤

펄프, 감귤 과피; 산림 폐기물, 예를 들어 경목재 및 연목재 간벌, 및 수목 작업으로부터의 경목재 및 연목재 잔류물; 목재 폐기물, 예를 들어 톱 가루 폐기물(목재 칩, 톱밥) 및 펄프 가루 폐기물; 도시 폐기물, 예를 들어 시의 고형 폐기물의 종이 분획물, 도시 목재 폐기물 및 도시 녹색 폐기물, 예를 들어 시의 깎아낸 풀; 및 목재 건축 폐기물을 포함한다. 추가의 셀룰로스는 스위치그래스, 잡종 포플러 목재 및 억제, 섬유 케인(cane) 및 섬유 수수와 같은 전용 셀룰로스 작물을 포한한다. 상기 물질로부터 생성된 5탄당은 자일로스를 포함한다.

[0162] 셀룰로스 물질은 상기 물질내 함유된 당(들)을 미생물이 이용할 수 있는 효율이 증가되도록 처리될 수 있다. 본 발명의 구체예는 산 폭쇄(acid explosion) 후 셀룰로스 물질을 처리하여 물질이 미생물(예를 들어, 미세 조류 및 유지성 효모)의 종속 영양성 배양에 사용하기에 적합하도록 하는 방법을 제공한다. 상기 논의된 바와 같이, 리그노셀룰로스 바이오매스는 셀룰로스, 베타 1,4 연결된 글루코스(6탄당)의 결정성 중합체, 헤미셀룰로스, 자일로스(5탄당)로 주로 구성된 보다 느슨하게 회합된 중합체 및 보다 적은 정도의 만노스, 갈락토스, 아라비노스, 리그닌, 시나필 알코올 및 이의 유도체로 구성된 복합 방향족 중합체, 및 알파 1,4-연결된 폴리갈락투론산의 선형 사슬인 펙틴을 포함하는, 다양한 분획물로 구성된다. 셀룰로스 및 헤미셀룰로스의 중합체 구조 때문에 이들 내 당(예를 들어, 단량체 글루코스 및 자일로스)은 많은 미생물에 의해 효율적으로 사용(대사)될 수 있는 형태로 존재 하지 않는다. 상기 미생물을 위해 중합체를 구성하는 단량체 당을 생성하도록 셀룰로스 바이오매스를 추가로 가공하면 셀룰로스 물질이 공급원료(탄소원)로서 효율적으로 사용되는 것을 보장하는 데 매우 도움이 될 수 있다.

[0163] 셀룰로스 또는 셀룰로스 바이오매스는, 바이오매스가 승온 및 승압에서 희석 황산(또는 기타 다른 산)으로 처리되는 "폭쇄(explosion)"로 호칭되는 공정을 거친다. 상기 공정은, 바이오매스가 셀룰로스 및 헤미셀룰로스 분획물의 글루코스 및 자일로스 단량체로의 효소 가수 분해를 효율적으로 거칠 수 있도록 바이오매스에 영향을 미친다. 생성된 단량체 당은 셀룰로스 당으로 호칭된다. 셀룰로스 당은 이어서 미생물에 의해 이용되어 다양한 대사물(예를 들어, 지질)을 생성할 수 있다. 산 폭쇄 단계는 헤미셀룰로스 분획물이 구성 단당류로 부분 가수 분해되도록 한다. 이들 당은 추가의 처리로 바이오매스로부터 완전히 유리될 수 있다. 몇몇 구체예에서, 추가의 처리는 열수 처리이고, 이는 폭쇄된 물질을 고온수로 세척하여 염과 같은 오염물을 제거하는 것을 포함한다. 이러한 단계는 상기 공정에 사용되는 보다 희석된 당 농도로 인해 셀룰로스 에탄올 발효에 대해서는 필요하지 않다. 다른 구체예에서, 추가의 처리는 추가의 산 처리이다. 또 다른 구체예에서, 추가의 처리는 폭쇄된 물질의 효소 가수 분해이다. 이들 처리는 또한 임의의 조합으로 사용될 수 있다. 처리 유형은 유리되는 당의 유형(예를 들어, 5탄당 대 6탄당), 및 이들이 당해 공정에서 유리되는 단계에 영향을 미칠 수 있다. 결과로서, 이들이 주로 5탄당 또는 6탄당이든 상관없이 상이한 당 스트림이 형성될 수 있다. 따라서 이들 농축된 5탄당 또는 6탄당 스트림은 상이한 탄소 이용 능력을 갖는 특이적 미생물에 지시될 수 있다.

[0164] 본 발명의 방법에서는 통상적으로 에탄올 발효에서 달성되는 것 보다 높은 세포 밀도로 발효가 이루어질 수 있다. 종속 영양성 셀룰로스 오일 생산을 위한 배양물의 보다 높은 밀도 때문에, 고정 탄소원(예를 들어, 셀룰로스 유래 당 스트림(들))은 바람직하게 농축된 형태로 존재한다. 해중합된 셀룰로스 물질의 글루코스 수준은 바람직하게 배양 단계 전에 300g/리터 이상, 400g/리터 이상, 500g/리터 이상 또는 600g/리터 이상이며, 상기 배양 단계는 임의로 유가식 배양이고, 여기서 상기 물질은 세포가 생장하고 지질을 축적함에 따라서 시간 경과에 따라 세포에 공급된다. 따라서, 리그노셀룰로스 오일의 생산 동안에 매우 높은 세포 밀도를 생성하고 유지하기 위해서, 탄소 공급원료(들)가 고농축된 형태로 종속 영양 배양물로 전달될 수 있다. 그러나, 유지성 미생물에 대한 기질이 아니고 유지성 미생물에 의해 대사되지 않는 공급 스트림에서 임의의 성분은 생물 반응기에 축적할 것이며, 이는 성분이 독성을 가지거나 원하는 최종 생성물의 생산을 억제하는 경우 문제점을 유발할 수 있다. 폭쇄 공정 및 후속적인 중화 공정 둘 다에서 셀룰로스 물질의 생성으로부터 유래된 푸르푸랄 및 하이드록시메틸 푸르푸랄 및 염과 같은 리그닌 및 리그닌 유래 부산물, 탄수화물 유래 부산물, 및 심지어 비대사된 펜토스/헥소스 당은 에탄올 발효에서 문제점을 제시할 수 있고, 이들 효과는 초기 공급원료에서 이들의 농도가 높은 공정에서 상당히 증폭된다. 본 발명에 기술된 리그노셀룰로스 오일의 대규모 생산에 사용될 수 있는 6탄당에 대한 300g/L 범위(또는 이보다 높은)의 당 농도를 성취하기 위해, 이들 독성 물질의 농도는 셀룰로스 바이오매스의 에탄올 발효물에 통상적으로 존재하는 농도보다 20배 높을 수 있다.

[0165] 셀룰로스 물질의 폭쇄 공정 처리는 상당한 양의 황산, 열 및 압력을 사용함으로써 탄수화물의 부산물, 즉 푸르푸랄 및 하이드록시메틸 푸르푸랄을 유리시킨다. 푸르푸랄 및 하이드록시메틸 푸르푸랄은 자일로스의 푸르푸랄 및 물로의 탈수를 통한 헤미셀룰로스의 가수 분해 동안에 생성된다. 본 발명의 몇몇 구체예에서, 이들 부산물(예를 들어, 푸르푸랄 및 하이드록시메틸 푸르푸랄)은 생물 반응기로 도입하기 전에 당화된 리그노셀룰로스 물질로부터 제거된다. 본 발명의 특정 구체예에서, 탄수화물의 부산물의 제거 공정은 폭쇄된 셀룰로스 물질의 열

수 처리이다. 추가로, 본 발명은 푸르푸랄 또는 하이드록시메틸 푸르푸랄과 같은 화합물에 내성일 수 있는 균주가 리그노셀룰로스 오일 생산에 사용되는 방법을 제공한다. 다른 구체예에서, 본 발명은 또한 발효 배지에서 푸르푸랄에 내성일 수 있는 것뿐만 아니라 실제로 리그노셀룰로스 오일의 생산 동안에 당해 부산물을 대사시킬 수 있는 미생물 및 방법을 제공한다.

[0166] 상기 폭쇄 공정은 또한 상당한 수준의 염을 생성한다. 예를 들어, 폭쇄를 위한 통상적인 조건은, 폭쇄된 셀룰로스 바이오매스가 10:1의 물:고체(건조 중량)의 비율로 재현탁되는 경우 5mS/cm 초과의 전도성을 초래할 수 있다. 본 발명의 임의의 구체예에서, 희석된 폭쇄된 바이오매스는 효소 당화를 거치고, 생성된 상청액은 생물반응기에 사용하기 위해 25배까지 농축된다. 농축된 당 스트림(들)에서 염 수준(전도성에 의해 측정됨)은 허용가능하지 않게 높을 수 있다(1.5M까지의 Na+ 당량). 추가의 염은 또한 후속적인 효소 당화 공정을 위해 폭쇄된 물질의 중화시 생성된다. 본 발명의 구체예는 이들 염을 제거하기 위한 방법을 제공하여 생성된 농축된 셀룰로스 당 스트림(들)이 리그노셀룰로스 오일을 생산하기 위한 종속 영양 공정에 사용될 수 있도록 한다. 몇몇 구체예에서, 상기 염을 제거하는 방법은, 예를 들어 이에 제한되지 않지만 DOWEX Marathon MR3과 같은 수지를 사용한 탈이온화이다. 임의의 구체예에서 수지 단계를 사용하는 탈이온화는 당화 전에, 바이오매스의 당 농축 또는 pH 조정 및 열수 처리 전, 또는 이들의 임의의 조합 전에 일어나며, 다른 구체예에서 상기 단계는 이들 공정 중 하나 이상 후에 수행한다. 다른 구체예에서, 폭쇄 공정 자체는 허용가능하지 않게 높은 수준으로 염이 생성되는 것을 피하도록 변화시킨다. 예를 들어, 셀룰로스 바이오매스의 황산(또는 기타 다른 산) 폭쇄에 대한 대안법은 기계적 펄핑으로 셀룰로스 바이오매스가 효소 가수 분해(당화)에 민감해지도록 하는 것이다. 또 다른 구체예에서, 고수준의 염에 내성인 미생물의 본래의 균주 또는 고수준의 염에 내성을 갖는 유전자 조작된 균주가 사용된다.

[0167] 유지성 미생물을 사용하는 종속 영양 리그노셀룰로스 오일 생산에 사용하기 위한 폭쇄된 셀룰로스 바이오매스를 제조하는 방법에 대한 바람직한 구체예는 하기에 기술한다. 제1 단계는 재현탁된 폭쇄된 셀룰로스 바이오매스의 pH를 5.0 내지 5.3의 범위로 조정하고 이어서 셀룰로스 바이오매스를 3회 세척하는 것을 포함한다. 이러한 세척 단계는 탈염 및 이온 교환 수지, 역삼투압, 열수 처리(전술한 바와 같음) 또는 단지 반복적인 탈이온수 중의 재현탁 및 원심분리의 사용을 포함하는 다양한 수단에 의해 성취될 수 있다. 이러한 세척 단계는 전도성이 100 μS/cm 내지 300 μS/cm인 셀룰로스 스트림을 초래하고 상당한 양의 푸르푸랄 및 하이드록시메틸 푸르푸랄을 제거시킨다. 이러한 세척 단계로부터 경사분리(decant)는 헤미셀룰로스 분획물로부터 유리된 5탄당을 농축시키는 단계를 배제할 수 있다. 제2 단계는 세척된 셀룰로스 바이오매스의 효소 당화를 포함한다. 바람직한 구체예에서, 아셀레라제(Genencor)를 사용한다. 제3 단계는 상기 당화된 바이오매스의 원심분리 또는 경사분리 및 세정을 통한 당의 회수를 포함한다. 생성한 바이오매스(고체)는 연료로서 사용될 수 있거나 폐기물로 보내질 수 있는 에너지가 밀접한 리그닌 풍부 성분이다. 원심분리/경사분리 및 세정 공정에서 회수된 당 스트림을 수집한다. 제4 단계는 투과물의 회수와 함께 오염 고체를 제거하기 위한 마이크로여과를 포함한다. 제5 단계는 진공 증발기를 사용하여 성취될 수 있는 농축 단계를 포함한다. 이러한 단계는 P'2000(Sigma/Fluka)과 같은 소포제의 첨가를 임의로 포함할 수 있으며, 이는 생성된 당 공급원료의 단백질 함량으로 인해 때때로 필요하다.

[0168] 본 발명의 방법의 다른 구체예에서, 탄소원은 글리세롤이며, 이는 바이오디젤 에스테르 결합 전이로부터 산성화 및 비산성화된 글리세롤 부산물을 포함한다. 하나의 구체예에서, 탄소원은 글리세롤 및 하나 이상의 기타 다른 탄소원을 포함한다. 몇몇 경우에, 글리세롤 및 하나 이상의 기타 다른 고정 탄소원 모두는 발효 초기에 미생물에 제공된다. 몇몇 경우에, 글리세롤 및 하나 이상의 기타 다른 고정 탄소원이 소정의 비율로 동시에 미생물에 제공된다. 몇몇 경우에, 글리세롤 및 하나 이상의 기타 다른 고정 탄소원은 발효 과정 동안에 소정의 속도로 미생물에 공급된다.

[0169] 몇몇 미세 조류는 글루코스의 존재 하에서보다 글리세롤의 존재 하에서 보다 신속하게 세포 분열을 진행한다(PCT 공개 제2008/151149호 참조). 이들 경우에, 처음에 세포에 글리세롤을 공급하여 세포 밀도를 신속하게 증가시키고, 이어서 글루코스를 공급하여 지질을 축적시키는 2단계 생장 공정은 지질이 생산되는 효율을 개선시킬 수 있다. 에스테르 결합 전이 공정의 글리세롤 부산물의 사용은 생산 공정으로 도입되는 경우 상당한 경제적인 이점을 제공할 수 있다. 글리세롤 및 글루코스의 혼합물과 같은 기타 다른 공급 방법이 또한 제공된다. 상기 혼합물의 공급은 또한 동일한 경제적 이득을 취득한다. 추가로, 본 발명은 미세 조류에 글리세롤과의 다양한 조합으로 수크로스와 같은 대안적인 당을 공급하는 방법을 제공한다.

[0170] 본 발명의 방법의 다른 구체예에서, 탄소원은 전화당이다. 전화당은 수크로스에 비하여 결정화되려는 성향이 약하므로, 저장 및 유가식 발효에 이점을 제공할 수 있으며, 미세 조류를 포함하는 미생물의 종속 영양 배양의 경우, 농축된 탄소원을 필요로 한다. 하나의 구체예에서, 탄소원은 전화당, 바람직하게는 농축된 형태의 전화당이

며, 바람직하게는 배양 단계, 임의로는 유가식 배양 단계 이전에 800g/리터 이상, 900g/리터 이상, 1000g/리터 이상 또는 1100g/리터 이상으로 존재한다. 전화당, 바람직하게 농축된 형태의 전화당은 세포가 생장하고 지질을 축적함에 따라서 시간 경과에 따라 세포에 공급된다.

[0171] 본 발명의 방법의 다른 구체예에서, 탄소원은 수크로스이며, 이는 사탕수수 가공 유래의 진한 감자 당즙과 같이 수크로스를 함유하는 복합 공급원료를 포함한다. 종속 영양 오일 생산용 배양물의 밀도가 더 높기 때문에, 고정된 탄소원(예를 들어, 수크로스, 글루코스 등)은 바람직하게 농축된 형태이며, 바람직하게는 배양 단계, 임의로는 세포가 생장하고 지질을 축적함에 따라서 시간 경과에 따라 물질이 세포에 공급되는 유가식 배양 단계 이전에 500g/리터 이상, 600g/리터 이상, 700g/리터 이상 또는 800g/리터 이상으로 존재한다. 몇몇 경우에 있어서, 탄소원은 진한 감자 당즙의 형태, 바람직하게는 농축된 형태의 수크로스이며, 바람직하게는 배양 단계, 임의로는 유가식 배양 단계 이전에 60% 이상의 고체 또는 약 770g/리터 당, 70% 이상의 고체 또는 약 925g/리터 당, 또는 80% 이상의 고체 또는 약 1125g/리터 당으로 존재한다. 농축된 진한 감자 당즙은 세포가 생장하고 지질을 축적함에 따라서 시간 경과에 따라 세포에 공급된다.

[0172] 하나의 구체예에서, 배양 배지는 하나 이상의 수크로스 이용 효소를 추가로 포함한다. 몇몇 경우에 있어서, 수크로스 이용 효소는 수크로스 인버타제이다. 수크로스 인버타제 효소는 미생물 군집에 의해 발현되는 외인성 수크로스 인버타제 유전자에 의해 암호화되는 분비성 수크로스 인버타제 효소일 수 있다. 분비성 수크로스 인버타제는 미생물에 의해 배양 배지 중에 분비될 수 있으며, 그 결과 배지 중 수크로스는 미생물에 의해 사용될 글루코스와 프럭토스로 전환된다. 이하에 기술된 바와 같이, 수크로스 인버타제는 재조합될 수 있으며, 이에 의해 순수하거나 복합적인 수크로스 공급 원료를 생장 또는 오일 생산용 고정 탄소원으로서 사용할 수 있는 능력을 미생물에 부여할 수 있다. 몇몇 경우에 하기 섹션 IV에 보다 상세하게 기술된 바와 같이, 미세 조류는 수크로스 수송체, 수크로스 인버타제, 헥소키나제, 글루코키나제 또는 프럭토키나제와 같은 수크로스 이용 효소를 발현하도록 유전자 조작하였다.

[0173] 수크로스를 함유하는 복합 공급원료는 사탕수수 가공 유래의 폐기 당밀을 포함하며, 사탕수수 가공의 낮은 가치의 폐기물의 사용은 탄화수소 및 기타 다른 오일의 생성에 상당한 비용 절감을 제공할 수 있다. 본 발명의 방법에 유용한 수크로스를 함유하는 다른 복합 공급원료는 수수 시럽 및 순수한 수수를 포함하는 수수이다. 수수 시럽은 달콤한 수수 줄기의 즙으로부터 생산된다. 이의 당 프로필은 주로 글루코스(덱스트로스), 프럭토스 및 수크로스로 이루어진다.

[0175] **4. 오일 생산**

[0176] 본 발명의 방법에 따른 오일 생산을 위해, 예를 들어 빛이 배양물에 입사할 수 없도록 하는 매우 대용량(40,000리터 이상)의 발효기를 사용하는 경우에서와 같이 암실에서 세포를 배양하는 것이 바람직하다. 종속 영양 종은, 고정 탄소원을 함유하는 배지에서 빛의 부재 하에 오일 생산을 위해 생장되고 번식되며, 상기 생장은 종속 영양 생장으로서 공지되어 있다.

[0177] 하나의 예로서, 지질 생산 미세 조류 세포의 접종물을 배지에 도입하고, 세포가 번식을 시작하기 전에 지체(lag) 기간(지체 단계)이 있다. 지체 기간 후, 번식률은 일정하게 증가하고 대수기(log) 또는 지수기로 진입한다. 지수기에 이어서 차례로 질소와 같은 영양물의 감소, 및 독성 물질의 증가 및 쿼럼센싱(quorum sensing) 기작으로 번식이 느려진다. 상기와 같이 느려진 후, 번식은 멈추고 세포는 정지기 또는 일정한 생장 상태로 진입하며, 이는 세포에 제공된 특정 환경에 좌우된다. 지질 풍부 바이오매스를 수득하기 위해, 상기 배양물은 통상적으로 질소 또는 다른 주요 영양물(탄소 이외의 다른 영양물)이 고갈되게 함으로써 조기에 종결되어 세포가 과량으로 존재하는 탄소원을 지질로, 특히 트리글리세라이드로 전환하도록 하는 지수기의 말기 후에 잘 수거한다. 배양 조건 파라미터는 총 오일 생산, 생산된 지질 종의 조합 및/또는 특정 오일의 생산을 최적화하도록 조작될 수 있다.

[0178] 본원에 개시된 세포에 의한 지질 생산은 영양물이 공급되거나 세포 분열의 부재 하에 계속적인 지질 생산을 가능하도록 하기에 여전히 이용가능한 정지기를 포함하는, 대수기 또는 이후 기간 동안에 일어날 수 있다.

[0179] 바람직하게, 본원에 기술되고/기술되거나 당업계에 공지된 조건을 사용하여 생장한 미생물은 건조 세포 중량으로 약 20% 내지 30%, 30% 내지 40%, 40% 내지 50%, 50% 내지 60%, 60% 내지 70%, 또는 80% 내지 90% 이상의 트리글리세라이드를 포함한다. 가공 조건은 특정 용도에 적합한 지질의 수율을 증가시키고/증가시키거나 생산 비용을 절감시키도록 조정될 수 있다. 예를 들어, 임의의 구체예에서, 고정 탄소 에너지, 예를 들어 글루코스를 과량으로 제공하면서, 미세 조류는 하나 이상의 영양분, 예를 들어 질소, 인 또는 황의 제한 농도의 존재 하에

서 배양한다. 질소 제한은 질소가 과량으로 제공된 배양물에서의 미생물 지질 수율 이상으로 미생물 지질 수율(건조 세포 중량의 그램 당 생산된 지질의 양을 측정)을 증가시키는 경향이 있다. 특정 구체예에서, 지질 수율의 증가는 약 10%, 50%, 100%, 200%, 또는 500% 이상이다. 상기 미생물은 전체 배양 기간 중 일부 동안 또는 전체 기간 동안 제한된 양의 영양물의 존재 하에 배양할 수 있다. 특정 구체예에서, 영양물 농도는 총 배양 기간 동안 2회 이상 제한 농도와 비제한 농도 사이에서 순환한다. 세포의 지질 함량은 과량의 탄소를 제공하지만 질소를 제한하거나 제공하지 않으면서 증가된 기간 동안 배양을 계속함으로써 증가시킬 수 있다.

[0180] 다른 구체예에서, 지질 수율은 지질 경로 효소(예를 들어, 지방산 합성 효소의 조효소 또는 보결 분자단)를 위한 하나 이상의 조인자(들)의 존재 하에 지질 생산 미생물(예를 들어, 미세 조류)을 배양함으로써 증가된다. 일반적으로, 조인자(들)의 농도는 조인자(들)의 부재 하의 미생물 지질 수율 이상으로 미생물 지질(예를 들어, 지방산) 수율을 증가시키기에 충분하다. 특정 구체예에서, 상기 조인자(들)는 배양물에 조인자(들)를 암호화하는 외인성 유전자를 함유하는 미생물(예를 들어, 미세 조류)을 포함시킴으로써 배양물에 제공된다. 대안적으로, 조인자(들)는 조인자의 합성에 관여하는 단백질을 암호화하는 외인성 유전자를 함유하는 미생물(예를 들어, 미세 조류)를 포함시킴으로써 배양물에 제공될 수 있다. 임의의 구체예에서, 적합한 조인자는 예를 들어, 비오틴 또는 판토테네이트와 같은 지질 경로 효소에 의해 요구되는 임의의 비타민을 포함한다. 본 발명에 사용하기에 적합한 조인자를 암호화하거나 상기 조인자의 합성에 관여하는 유전자는 널리 공지되어 있고 전술한 것들과 같은 기술 및 구조물을 사용하여 미생물(예를 들어, 미세 조류)에 도입될 수 있다.

[0181] 본원에 기술된 생물 반응기, 배양 조건 및 종속 영양 생장 및 번식 방법의 특정 예를 임의의 적합한 방식으로 조합하여 미생물 생장 및 지질 및/또는 단백질 생산의 효율을 개선시킬 수 있다.

[0182] 건조 중량으로 높은 백분율의 오일/지질이 축적된 미세 조류 바이오매스는 당업계에 공지된 상이한 배양 방법을 사용하여 생성되었다(PCT 공개 제2008/151149호 참조). 본원에 기술되고 본 발명에 따라 유용한 배양 방법에 의해 생성된 미세 조류 바이오매스는 건조 중량으로 10% 이상의 미세 조류 오일을 포함한다. 몇몇 구체예에서, 미생물 바이오매스는 건조 중량을 기준으로 25%, 50%, 60%, 70% 이상, 또는 80% 이상의 미세 조류 오일을 포함한다. 몇몇 구체예에서, 미세 조류 바이오매스는 건조 중량으로 10% 내지 90%의 미세 조류 오일, 25% 내지 75%의 미세 조류 오일, 40% 내지 75%의 미세 조류 오일, 75% 내지 85% 또는 50% 내지 70%의 미세 조류 오일을 함유한다.

[0183] 본원에 기술된 바이오매스의 미세 조류 오일, 또는 본 발명의 방법 및 조성물에 사용하기 위한 바이오매스로부터 추출된 오일은 하나 이상의 별개의 지방산 에스테르 측쇄를 갖는 글리세로지질을 포함할 수 있다. 글리세로지질은 다양한 길이 및 다양한 포화도를 가질 수 있는 1개, 2개 또는 3개의 지방산 분자에 에스테르화된 글리세롤 분자로 구성된다. 지방산 분자(및 미세 조류 오일)의 길이 및 포화 특성은 하기 섹션 IV에서 보다 상세하게 기술된 바와 같은 배양 조건을 통해 또는 지질 경로 조작을 통해 본 발명의 구체예의 미세 조류 오일 중 지방산 분자의 특성 또는 비율을 변형시키기 위해 조작될 수 있다. 특성과 비율의 특정 변형은 미생물 트리글리세라이드의 지방산 분포 변화, 예를 들어 지방산의 사슬 길이 프로필, 포화 프로필 및 하이드록실화의 변화를 포함한다. 이와 같이 생산된 오일은 천연 오일을 포함할 수 있다. 대안적으로, 미생물 오일의 특이적 배합물은 2가지 이상의 미세 조류 종으로부터 바이오매스 또는 조류 오일을 함께 혼합함으로써, 또는 대두, 평지씨, 카놀라, 야자, 팜핵, 코코넛, 옥수수, 폐기 식물, 오구나무(Chinese tallow), 올리브, 해바라기, 면화씨, 닭 지방, 우지, 돈지(porcine tallow), 미세 조류, 거대 조류, 미생물, 쿠페아(Cuphea), 아마, 땅콩, 초이스 백색 그리스(choice white grease), 라드, 양구슬냉이(Camelina sativa), 겨자씨, 캐슈 너트, 귀리, 루핀, 양마, 금잔화, 대마, 커피, 아마인(아마), 헤이즐넛, 등대풀속, 호박씨, 고수, 동백나무, 참깨, 홍화, 벼, 유동, 코코아, 코프라, 양귀비, 아주까리씨, 피칸, 호호바, 마카다미아, 브라질 너트, 아보카도, 석유 또는 상기 오일 중 임의의 증류 분획물과 같은 기타 다른 공급원 유래의 오일과 본 발명의 조류 오일을 배합함으로써 조류의 단일 종 내에서 제조할 수 있다.

[0184] 오일 조성물, 즉 글리세로지질의 지방산 성분의 특성 및 비율은 또한 미생물의 2가지 이상의 별개 종으로부터 유래한 바이오매스 또는 오일을 조합함으로써 조작할 수 있다. 몇몇 구체예에서, 미세 조류의 2가지 이상의 별개 종은 상이한 글리세로지질 프로필을 갖는다. 미세 조류의 별개 종은 바람직하게 종속 영양 조건 하에 본원에 기술된 바와 같이 함께 배양하거나 별도로 배양하여 각각의 오일을 생성할 수 있다. 미세 조류의 상이한 종은 세포의 글리세로지질 내 상이한 백분율의 별개의 지방산 성분을 함유할 수 있다.

[0185] 일반적으로, 프로토테카 균주는 사슬 길이가 C8 내지 C14인 지방산을 매우 적게 갖거나 또는 전혀 갖지 않는다. 예를 들어, 프로토테카 모리포르미스(UTEX 1435), 프로토테카 크루가니(Prototheca krugani)(UTEX 329), 프로

토테카 스태그노라(UTEX 1442) 및 프로토테카 조프피(UTEX 1438)는 C8 지방산을 함유하지 않고(또는 검출가능하지 않은 양으로 함유함) C10 지방산을 0% 내지 0.01%, C12 지방산을 0.03% 내지 2.1%, C14 지방산을 1.0% 내지 1.7%를 함유한다.

[0186] 몇몇 경우에 있어서, 사슬 길이가 C8 또는 C8 내지 10인 지방 아실-ACP 기질에 대한 활성을 가지는 지방 아실-ACP 티오에스테라제를 암호화하는 이식 유전자 함유 미생물 균주는 사슬 길이가 C8인 지방산을 1.5% 이상, 3.0% 이상, 10% 이상, 12% 이상, 또는 그 이상 가진다. 다른 경우에 있어서, 사슬 길이가 C10인 지방 아실-ACP 기질에 대한 활성을 가지는 지방 아실 ACP 티오에스테라제를 암호화하는 이식 유전자 함유 미생물 균주는 사슬 길이가 C10인 지방산을 5.0% 이상, 10.0% 이상, 24% 이상, 29% 이상, 또는 그 이상 가진다. 다른 경우에 있어서, 사슬 길이가 C12인 지방 아실-ACP 기질에 대한 활성을 가지는 지방 아실-ACP 티오에스테라제를 암호화하는 이식 유전자 함유 미생물 균주는 사슬 길이가 C12인 지방산을 5% 이상, 15% 이상, 34% 이상, 50% 이상, 또는 그 이상 가진다. 다른 경우에 있어서, 사슬 길이가 C14인 지방 아실-ACP 기질에 대한 활성을 가지는 지방 아실-ACP 티오에스테라제를 암호화하는 이식 유전자 함유 미생물 균주는 사슬 길이가 C14인 지방산을 2.0% 이상, 7% 이상, 10% 이상, 15% 이상, 30% 이상, 또는 43% 이상, 또는 그 이상 가진다. 다른 경우에 있어서, 사슬 길이가 C16인 지방 아실-ACP 기질에 대한 활성을 가지는 지방 아실-ACP 티오에스테라제를 암호화하는 이식 유전자 함유 미생물 균주는 사슬 길이가 C16인 지방산을 30% 이상, 40% 이상, 66% 이상, 또는 그 이상 가진다. 또 다른 경우에 있어서, 사슬 길이가 C18인 지방 아실-ACP 기질, 특히 C18:0에 대한 활성을 가지는 지방 아실-ACP 티오에스테라제를 암호화하는 이식 유전자 함유 미생물 균주는 C18:0 지방산 수준이 5% 이상, 10% 이상, 26% 이상, 40% 이상, 또는 그 이상이다. 이와 같은 예들 중 임의의 것에 있어서, 미생물은 미세 조류, 예를 들어 프로토테카일 수 있다.

[0187] 비제한적인 예에 있어서, 사슬 길이가 C8인 지방 아실-ACP 기질에 대하여 활성을 가지는 지방 아실-ACP 티오에스테라제를 암호화하는 이식 유전자 함유 미생물 균주는 사슬 길이가 C8인 지방산을 1% 내지 20%, 바람직하게는 1.8% 내지 13% 가진다. 다른 비제한적인 예에서, 사슬 길이가 C10인 지방 아실-ACP 기질에 대한 활성을 가지는 지방 아실-ACP 티오에스테라제를 암호화하는 이식 유전자 함유 미생물 균주는 사슬 길이가 C10인 지방산을 1% 내지 40%, 바람직하게는 1.91% 내지 30% 가진다. 다른 비제한적인 예에서, 사슬 길이가 C12인 지방 아실-ACP 기질에 대한 활성을 가지는 지방 아실-ACP 티오에스테라제를 암호화하는 이식 유전자 함유 미생물 균주는 사슬 길이가 C12인 지방산을 10% 내지 60%, 바람직하게는 13.55% 내지 55% 가진다. 기타 비제한적인 예에서, 사슬 길이가 C14인 지방 아실-ACP 기질에 대한 활성을 가지는 지방 아실-ACP 티오에스테라제를 암호화하는 이식 유전자 함유 미생물 균주는 사슬 길이가 C14인 지방산을 1% 내지 50%, 바람직하게는 2.59% 내지 43.27% 가진다. 다른 비제한적인 예에서, 탄소 사슬 길이가 다양한 지방 아실-ACP 기질에 대한 특이성이 넓은 지방 아실-ACP 티오에스테라제를 암호화하는 이식 유전자 함유 미생물 균주는 사슬 길이가 C16인 지방산을 70% 이하 가진다. 다른 경우에서, 사슬 길이가 C16인 지방 아실-ACP 기질에 대한 활성을 가지는 지방 아실-ACP 티오에스테라제를 암호화하는 이식 유전자 함유 미생물 균주는 사슬 길이가 C16인 지방산을 75% 이하, 바람직하게는 67.42% 이하 가진다. 몇몇 경우에 있어서, 사슬 길이가 C8 내지 C14인 지방 아실-ACP 기질에 대한 활성을 가지는 지방 아실-ACP 티오에스테라제를 암호화하는 이식 유전자 함유 미생물 균주는 (C8 내지 C14) 지방산을 1% 내지 790%, 또는 약 2% 내지 80% 가진다. 몇몇 경우에 있어서, 사슬 길이가 C12 내지 C14인 지방 아실-ACP 기질에 대한 활성을 가지는 지방 아실-ACP 티오에스테라제를 암호화하는 이식 유전자 함유 미생물 균주는 C12 내지 C14 지방산을 50% 이상 또는 60% 이상 가진다. 몇몇 경우에 있어서, 특정 사슬 길이의 원하는 지방산이 증가하게 되므로, 일정하면서 높은 선택압 하에 이식 유전자 함유 미생물 균주가 외인성 유전자를 보유하도록 유지시키는 것은 유리하다. 외인성 유전자 보유의 고수준은 또한 본원에 개시된 방법 및 상동성 재조합 벡터를 사용하여 세포의 핵내 염색체에 외인성 유전자를 삽입함으로써 성취될 수 있다. 핵내 염색체에 통합된 외인성 유전자를 함유하는 재조합 세포는 본 발명의 목적이다. 이와 같은 예들 중 임의의 것에 있어서, 미생물은 미세 조류, 예를 들어 프로토테카일 수 있다.

[0188] 임의로, 미생물 오일은 또한 미세 조류에 의해 생산되거나, 배양 배지로부터 유래하는 미세 조류 오일에 통합되는 기타 다른 성분들을 포함할 수도 있다. 이러한 기타 다른 구성 성분들은 미세 조류를 배양하는데 사용되는 배양 조건, 미세 조류의 종, 바이오매스로부터 미세 조류 오일을 회수할 때 사용되는 추출 방법, 그리고 미세 조류 오일 조성에 영향을 미칠 수 있는 기타 다른 인자들에 따라서 다양한 양으로 존재할 수 있다. 이와 같은 성분들의 비제한적인 예로서는 카로티노이드(오일 중 0.025mcg/g 내지 0.3mcg/g, 바람직하게는 0.05mcg/g 내지 0.244mcg/g으로 존재), 엽록소 A(오일 중 0.025mcg/g 내지 0.3mcg/g, 바람직하게는 0.045mcg/g 내지 0.268mcg/g으로 존재); 총 엽록소(오일 중 0.03mcg/g 미만, 바람직하게는 0.025mcg/g 미만으로 존재); 감마 토코페롤(오일 중 35mcg/g 내지 175mcg/g, 바람직하게는 38.3mcg/g 내지 164mcg/g으로 존재); 총 토코페롤(오일

중 50mcg/g 내지 300mcg/g, 바람직하게는 60.8mcg/g 내지 261.7mcg/g으로 존재); 브라시카스테롤, 캄페스테롤, 스티그마스테롤 또는 베타시토스테롤(0.5% 미만, 바람직하게는 0.25% 미만으로 존재); 총 토코트리에놀(오일 중 300mcg/g 미만으로 존재); 그리고 총 토코트리에놀(오일 중 225mcg/g 내지 350mcg/g, 바람직하게는 249.6mcg/g 내지 325.3mcg/g으로 존재)을 포함한다.

[0189] 기타 다른 성분으로서는 인지질, 토코페롤, 토코트리에놀, 카로티노이드(예를 들어, 알파-카로틴, 베타-카로틴 및 라이코펜 등), 잔토필(예를 들어, 루테인, 제아잔틴, 알파-크립토잔틴 및 베타-크립토잔틴) 및 다양한 유기 또는 무기 화합물을 포함할 수 있으나, 이에 제한되는 것은 아니다. 몇몇 경우에 있어서, 프로토테카 종으로부터 추출된 오일은 오일 1g당 0.001마이크로그램 내지 0.05마이크로그램의 루테인, 바람직하게는 0.003마이크로그램 내지 0.039마이크로그램의 루테인, 오일 1g당 0.005마이크로그램 미만의 라이코펜, 바람직하게는 0.003마이크로그램 미만의 라이코펜; 오일 1g당 0.005마이크로그램 미만의 베타 카로틴, 바람직하게는 0.003마이크로그램 미만의 베타 카로틴을 포함한다.

[0191] **III. 유전자 조작 방법 및 재료**

[0192] 본 발명은 프로토테카 세포를 유전자 변형시키기 위한 방법 및 재료 및 본 발명의 방법에 유용한 재조합 숙주 세포를 제공하고, 당해 숙주 세포는 재조합 프로토테카 모리포르미스, 프로토테카 조프피, 프로토테카 크루가니 및 프로토테카 스태그노라 숙주 세포를 포함하지만, 이에 제한되지 않는다. 이들 방법 및 재료의 기재는 독자의 편의를 위해 하위 섹션으로 나눈다. 하위 섹션 1에서, 형질 전환 방법을 기술한다. 하위 섹션 2에서 상동성 재조합을 사용한 유전자 조작 방법을 기술한다. 하위 섹션 3에서 발현 벡터 및 성분을 기술한다.

[0193] **1. 조작 방법 - 형질 전환**

[0194] 세포는, 예를 들어 유전자 총(biolistics), 전기천공(문헌[Maruyama et al. (2004), Biotechnology Techniques 8:821-826] 참조), 유리 비드 형질 전환 및 탄화규소 휘스커 형질 전환을 포함하는 임의의 적합한 기술에 의해 형질 전환될 수 있다. 사용될 수 있는 다른 방법은 원형질체를 형성하고 CaCl2 및 폴리에틸렌 글리콜(PEG)을 사용하여 재조합 DNA를 미세 조류 세포에 도입하는 것을 수반한다(문헌[Kim et al. (2002), Mar. Biotechnol. 4:63-73] 참조, 상기 문헌은 클로렐라 엘립소이데아의 형질 전환을 위한 상기 방법의 사용을 보고함). 미세 조류의 동시 형질전환을 사용하여 2가지의 별개 벡터 분자를 세포에 동시에 도입할 수 있다(예를 들어, 문헌[Protist 2004 Dec;155(4):381-93] 참조).

[0195] 또한 유전자 총 방법(예를 들어, 문헌[Sanford, Trends In Biotech. (1988) 6:299 302], 미국 특허 제4,945,050호 참조); 전기천공[문헌[Fromm et al., Proc. Nat'l. Acad. Sci. (USA) (1985) 82:5824 5828]; 레이져 빔, 미세주입 또는 DNA를 미세 조류로 도입할 수 있는 임의의 기타 다른 방법을 프로토테카 세포의 형질 전환에 사용할 수 있다.

[0196] **2. 조작 방법 - 상동성 재조합**

[0197] 상동성 재조합은 상동성 영역을 정렬하고 교환하는 상보적 DNA 서열의 능력이다. 표적화된 게놈 서열(“주형”)과 상동성인 서열을 함유하는 이식 유전자 DNA(“공여체”)는 유기체에 도입된 다음, 상응하는 게놈 상동성 서열 부위에서 게놈으로 재조합을 진행한다.

[0198] 숙주 유기체에서 상동성 재조합을 수행하기 위한 능력은 분자 유전학적 수준에서 수행될 수 있는 것에 대한 많은 실제 연관성을 갖고 맞춤 오일을 생성할 수 있는 유지성 미생물의 생성에 유용하다. 이의 매우 자연적인 상동성 재조합이란 정확한 유전자 표적화 사건이고, 따라서 동일한 표적화 서열로 생성된 대부분의 이식 유전자 계열은 표현형의 관점에서 근본적으로 동일하고 극소수의 형질 전환 사건의 스크리닝을 필요로 한다. 상동성 재조합은 또한 숙주 염색체로의 유전자 삽입 사건을 표적화하여, 잠재적으로 심지어 유전자 선별 부재 하에서도 우수한 유전학적 안정성을 가져온다. 상이한 염색체 유전자좌는 심지어 이종성 프로모터/UTR로부터 유전자 발현에 영향을 줄 가능성이 있기 때문에, 상동성 재조합은 잘 모르는 게놈 환경 내에서 유전자좌를 탐색하고 유전자 발현에 대한 이들 환경의 영향을 평가하는 방법일 수 있다.

[0199] 상동성 재조합을 사용하는 특히 유용한 유전자 조작 접근법은 고도로 특이적인 방식으로 이종성 유전자 발현을 구동시키는 프로모터/UTR과 같은 특이적 숙주 조절 요소를 선택하는 것이다. 예를 들어, 선택 마커를 암호화하는 이종 유전자로 불포화 효소 유전자/유전자 군을 제거 또는 녹아웃하는 것은 숙주 세포 내에서 생산된 포화 지방산의 전체 백분율을 증가시킬 것으로 예상될 수 있다. 실시예 6은 상동성 재조합 표적화 구조물과, (프로토테카 모리포르미스 내에서 발생되는) 이와 같은 불포화 효소 유전자 제거 또는 녹아웃의 작업 실시예를 기술한다. 내인성 유전자의 발현을 감소시키기 위한 다른 접근법으로서는 유전자 발현을 RNA 유도 하향 조절 또는 침

묵을 이용하는 접근법, 예를 들어 RNAi 또는 안티센스 접근법, 그리고 dsRNA 접근법(이에 제한되는 것은 아님)이 있다. 안티센스, RNAi, dsRNA 접근법은 당업계에 널리 공지되어 있으며, mRNA로서 발현될 때 헤어핀 RNA를 형성시킬 발현 구조물 또는 안티센스 배향으로 전사될 표적 유전자의 일부를 함유하는 발현 구조물의 도입을 포함한다. 3가지 접근법 모두는 표적 유전자의 발현의 감소를 가져올 것이다. 실시예 6은 또한 발현 구조물과, RNAi 및 안티센스 접근법에 의해 내인성 프로토테카 모리포르미스 델타 12 불포화 효소 유전자(FADc)를 하향 조절하는 작업 실시예를 기술한다.

[0200] 상동성 재조합은 정확한 유전자 표적화 사건이기 때문에, 충분한 측접(flanking) 영역이 확인되었다면, 관심의 유전자 또는 영역 내 임의의 뉴클레오티드(들)를 정확히 변형시키는데 사용될 수 있다. 따라서, 상동성 재조합은 RNA 및/또는 단백질의 유전자 발현에 영향을 주는 조절 서열을 변형시키기 위한 수단으로서 사용될 수 있다. 이것은 또한 기질 특이성, 친화성 및 Km과 같은 효소 활성을 변형시킴에 따라서 숙주 세포의 대사에서 원하는 변화에 영향을 주고자 단백질 암호화 영역을 변형시키는데 사용될 수 있다. 상동성 재조합은 숙주 게놈을 조작하여 유전자 표적화, 유전자 전환, 유전자 결실, 유전자 복제, 유전자 역위를 가져오고, 프로모터, 인핸서 및 3'UTR과 같은 유전자 발현 조절 요소를 교환하기 위한 강력한 수단을 제공한다.

[0201] 상동성 재조합은 내인성 숙주 세포 게놈 내에 관심의 유전자 또는 영역을 "표적화"하기 위한 내인성 서열 조각을 함유하는 표적화 구조물을 사용하여 성취될 수 있다. 상기 표적화 서열은 관심의 유전자 또는 영역의 5'에 위치하거나 관심의 유전자/영역의 3'에 위치하거나 심지어 관심의 유전자/영역을 측접할 수 있다. 상기 표적화 구조물은 추가의 벡터 골격과 함께 고차코일화된 플라스미드 DNA로서, 벡터 골격을 갖지 않는 PCR 생성물로서 또는 선형화된 분자로서 숙주 세포로 형질 전환될 수 있다. 몇몇 경우에, 이것은 먼저 제한 효소에 이식 유전자 DNA(공여체 DNA)내의 상동성 서열을 노출시키는 것이 유리할 수 있다. 상기 단계는 재조합 효율을 증가시킬 수 있고 원하지 않는 사건의 발생을 감소시킬 수 있다. 재조합 효율을 증가시키는 기타 다른 방법은 PCR을 사용하여 표적화된 게놈 서열에 상동성인 선형 말단을 함유하는 형질 전환 이식 유전자 DNA를 생성시키는 것을 포함한다.

[0202] 비제한적 예시를 위하여, 상동성 재조합에 유용한 공여 DNA 서열 영역은 프로토테카 모리포르미스 내 DNA의 KE858 영역을 포함한다. KE858은 단백질의 운반 RNA(tRNA) 군과 상동성을 공유하는 단백질의 암호화 영역의 일부를 포함하는, 1.3kb 크기의 게놈 단편이다. 서던 블럿은, KE858 서열이 프로토테카 모리포르미스(UTEX 1435) 게놈 내 단일 복사체로 존재한다는 것을 보여주었다. 이 영역과, 이 영역을 상동성 재조합 표적화에 이용하는 실시예는 PCT 출원 PCT/US2009/066142호에 기술되어 있다. 유용한 공여체 DNA의 다른 영역은 본원에서 "6S"라고 칭하여지는 게놈 서열(서열 번호 82 및 서열 번호 84의 공여체 서열)이다. 6S 서열은 6S rRNA 서열이 아니라는 것을 주목한다. 이 서열을 프로토테카 모리포르미스 내 상동성 재조합에 사용하는 것은 이하 실시예에 기술되어 있다.

[0203] **3. 벡터 및 벡터 성분**

[0204] 본 발명에 따라 미생물을 형질 전환시키기 위한 벡터는 본원에 개시된 바를 고려하여 당업자에게 익숙한 공지된 기술에 의해 제조할 수 있다. 벡터는 통상적으로 하나 이상의 유전자를 함유하며, 여기서 각각의 유전자는 원하는 생성물(유전자 생성물)의 발현을 암호화하고 유전자 발현을 조절하거나 유전자 생성물을 재조합 세포 내 특정 위치에 표적화하는 하나 이상의 제어 서열에 작동적으로 연결되어 있다. 독자를 돕기 위해, 본 하위 섹션은 여러 하위 섹션으로 나누어진다. 하위 섹션 A는 통상적으로 벡터 상에 함유된 제어 서열뿐만 아니라 본 발명에 의해 제공된 신규 제어 서열을 기술한다. 하위 섹션 B는 벡터 내에 통상적으로 함유된 유전자뿐만 아니라 신규 코돈의 최적화 방법 및 본 발명에 의해 제공된 것들을 사용하여 제조된 유전자를 기술한다.

[0205] **A. 제어 서열**

[0206] 제어 서열은 암호화 서열의 발현을 조절하거나 유전자 생성물을 세포 내부 또는 외부의 특정 위치로 지시하는 핵산이다. 발현을 조절하는 제어 서열은, 예를 들어 암호화 서열의 전사를 조절하는 프로모터 및 암호화 서열의 전사를 종결시키는 종결자를 포함한다. 다른 제어 서열은 폴리아데닐화 신호를 암호화하는 암호화 서열의 말단에 위치한 3' 비번역 서열이다. 유전자 생성물을 특정 위치로 지시하는 제어 서열은 단백질이 세포 내부 또는 외부의 특정 위치에 부착되는 위치로 지시하는 신호 펩티드를 암호화하는 것들을 포함한다.

[0207] 따라서, 미세 조류 내 외인성 유전자의 발현을 위한 예시적인 벡터 설계는 미세 조류내에서 활성인 프로모터에 작동 가능하도록 결합된 원하는 유전자 생성물(예를 들어, 선택가능한 마커, 지질 경로 변형 효소 또는 수크로스 이용 효소)에 대한 암호화 서열을 함유한다. 대안적으로, 벡터가 관심의 암호화 서열과 작동가능하게 연결된

프로모터를 함유하지 않는 경우, 상기 암호화 서열은 이것이 벡터 통합 지점에서 내인성 프로모터와 작동가능하게 연결되도록 세포 내로 형질 전환될 수 있다. 형질 전환의 프로모터 부재 방법은 미세 조류에서 작동하는 것으로 입증되었다(예를 들어, 문헌[Plant Journal 14:4, (1998), pp.441-447] 참조).

[0208] 많은 프로모터, 예를 들어 형질 전환된 조류에 내인성이 아닌 프로모터(즉, 다른 조류 유래의 프로모터, 고등 식물 유래의 프로모터 및 식물 바이러스 또는 조류 바이러스 유래의 프로모터)뿐만 아니라 형질 전환된 조류에 내인성인 프로모터는 미세 조류 내에서 활성이다. 미세 조류 내에서 활성인 예시적인 외인성 및/또는 내인성 프로모터(또한 미세 조류에서 기능하는 항생제 내성 유전자)는 PCT 공개 제2008/151149호 및 본원에 인용된 참조 문헌에 기술되어 있다.

[0209] 외인성 유전자를 발현시키는데 사용되는 프로모터는 상기 유전자에 자연적으로 연결되는 프로모터일 수 있거나 이종성 프로모터일 수 있다. 몇몇 프로모터는 하나 초과의 미세 조류 종에서 활성이다. 기타 다른 프로모터는 종 특이적이다. 예시적인 프로모터로서는 하기 실시예에 사용된 클라미도모나스 레인하르드티(Chlamydomonas reinhardtii) 유래의 β-튜불린과 같은 프로모터, 및 다수의 미세 조류 종에서 활성인 것으로 밝혀진 콜리플라워 모자이크 바이러스(CMV) 및 클로렐라 바이러스와 같은 바이러스 프로모터를 포함한다(예를 들어, 문헌[Plant Cell Rep. 2005 Mar;23(10-11):727-35]; [J Microbiol. 2005 Aug;43(4):361-5]; [Mar Biotechnol (NY). 2002 Jan;4(1):63-73] 참조). 프로토테카에서 외인성 유전자의 발현에 사용하기에 적합한 다른 프로모터는 클로렐라 소로키니아나 글루타메이트 탈수소화 효소 프로모터/5 'UTR이다. 임의로, 프로모터를 함유하는 상기 서열중 10개, 20개, 30개, 40개, 50개, 또는 60개 이상의 뉴클레오티드가 사용된다. 프로토테카에서 외인성 유전자의 발현에 유용한 예시적인 프로모터는 본 출원의 서열 목록에 열거되어 있으며, 예를 들어 클로렐라 HUP1 유전자의 프로모터(서열번호 1) 및 클로렐라 엘립소이데아 질산염 환원 효소 프로모터(서열번호 2)가 있다. 클로렐라 바이러스 프로모터는 또한 프로토테카에서 미국 특허 제6,395,965호의 서열번호 1 내지 서열번호 7과 같은 유전자를 발현시키는데 사용될 수 있다. 프로토테카에서 활성인 추가의 프로모터는 예를 들어 문헌[Biochem Biophys Res Commun. 1994 Oct 14;204(1): 187-94]; [Plant Mol Biol. 1994 Oct;26(1):85-93]; [Virology. 2004 Aug 15;326(1): 150-9]; 및 [Virology. 2004 Jan 5;318(1):214-23]에서 찾을 수 있다. 기타 다른 유용한 프로모터는 이하 실시예에 상세히 기술되어 있다.

[0210] 프로모터는 일반적으로 지속성(constitutive)이거나 유도성인 것으로서 특징지어질 수 있다. 지속성 프로모터는 일반적으로 동일한 수준에서 항상(또는 세포 수명 사이클의 특정 시점에서) 발현을 구동하는데 활성이거나 발현을 구동하도록 작용한다. 역으로, 유도성 프로모터는 활성이거나(또는 불활성화되거나) 단지 자극에 응답하여 상당히 상향 조절되거나 하향 조절된다. 상기 프로모터 유형 둘 다 본 발명의 방법에 적용된다. 본 발명에서 유용한 유도성 프로모터는 외인성으로 제공된 소분자(예를 들어, 서열번호 1에서와 같이 글루코스), 온도(열 또는 냉), 배양 배지 내 질소 부재 등과 같은 자극에 응답하여 작동 가능하도록 결합된 유전자의 전사를 매개하는 것들을 포함한다. 적합한 프로모터는 근본적으로 사일런트 유전자의 전사를 활성화시킬 수 있거나 바람직하게 낮은 수준으로 전사된 작동 가능하도록 결합된 유전자의 전사를 상당히 상향 조절할 수 있다. 이하 실시예에는 프로토테카 세포 내에서 유용한 부가적인 유도성 프로모터를 기술한다.

[0211] 종결 영역 제어 서열의 내포는 임의적이며, 상기 종결 영역은 비교적 상호교환가능하므로 사용되는 경우 상기 선택은 주로 편의를 위한 것이다. 상기 종결 영역은 전사 개시 영역(프로모터)에 고유한 것일 수 있거나, 관심의 DNA 서열에 고유한 것일 수 있거나, 다른 공급원으로부터 수득할 수 있다. 예를 들어, 문헌[Chen 및 Orozco, Nucleic Acids Res. (1988) 16:8411]를 참조한다.

[0212] 본 발명은 또한 제어 서열과 재조합 유전자, 그리고 관심의 유전자 생성물을 특정 세포 구획, 예를 들어 엽록체, 색소체, 미토콘드리아 또는 소포체로 보내는 것을 가능하게 하는 유전자를 함유하는 벡터를 제공한다. 뿐만 아니라, 본 발명의 구체예는 제어 서열과 재조합 유전자, 그리고 세포 외부로 단백질의 분비를 가능하게 하는 유전자를 함유하는 벡터를 포함한다.

[0213] 프로토테카 핵 게놈에서 발현되는 단백질은 색소체 표적화 신호를 사용하여 색소체로 표적화될 수 있다. 클로렐라에 내인성인 색소체 표적화 서열은, 예를 들어 색소체로 표적화되는 단백질을 암호화하는 클로렐라 핵 게놈 내 유전자들은 공지되어 있고, 이는 예를 들어 GenBank 수탁 번호 제AY646197호 및 제AF499684호를 참조하며, 하나의 구체예에서 이러한 제어 서열은 단백질의 발현을 프로토테카 색소체로 표적화하는 본 발명의 벡터 내에 사용된다.

[0214] 하기 실시예는 숙주 세포 내 올바른 구획으로 이종성 단백질을 표적화하는 조류 색소체 표적화 서열의 용도를 기술한다. cDNA 라이브러리는 프로토테카 모리포르미스 및 클로렐라 프로토테코디에스(Chlorella

protothecodies) 세포를 사용하여 제조되었으며 PCT 출원 PCT/US2009/066142호에 기술되어 있다.

[0215] 본 발명의 다른 구체예에서, 프로토테카 내에서 폴리펩티드의 발현은 소포체로 표적화된다. 발현 벡터 내 적절한 보유 또는 분류 신호의 내포는 단백질이 소포체(ER)에 보유되고 골지체로 하류로 가지 않도록 하는 것을 보장한다. 예를 들어, IMPACTVECTOR 1.3 벡터(Wageningen UR- Plant Research International 제조)는 널리 공지된 KDEL 보유 또는 분류 신호를 포함한다. 상기 벡터와 함께, ER 보유는 이것이 발현 수준을 5배 이상으로 개선시킴이 보고되었다는 점에서 실제 이점을 갖는다. 이에 대한 주요 이유는 ER이 세포질 내 존재하는 것보다 발현된 단백질의 번역 후 분해에 관여하는 보다 낮은 농도이고/농도이거나 상이한 프로테아제를 함유하는 것으로 나타난다. 녹색 미세 조류에서 기능하는 ER 보유 신호는 공지되어 있다. 예를 들어, 문헌[Proc Natl Acad Sci U S A. 2005 Apr 26;102(17):6225-30]을 참조한다.

[0216] 본 발명의 다른 구체예에서, 폴리펩티드는 배양 배지 내로 세포 외부로의 분비를 위해 표적화된다. 프로토테카에서 본 발명의 방법에 따라 사용될 수 있는 클로렐라 중에서 활성인 분비 신호에 대한 예는 문헌[Hawkins et al., Current Microbiology Vol. 38 (1999), pp. 335-341]을 참조한다.

[0217] **B. 유전자 및 코돈 최적화**

[0218] 통상적으로, 유전자는 프로모터, 암호화 서열 및 종결 제어 서열을 포함한다. 재조합 DNA 기술에 의해 조립되는 경우, 유전자는 발현 카세트로 호칭될 수 있고, 재조합 유전자를 숙주 세포로 도입하는데 사용되는 벡터로의 편리한 삽입을 위한 제한 부위에 측접할 수 있다. 상기 발현 카세트는 상동성 재조합에 의해 발현 카세트의 게놈으로의 안정적인 통합을 촉진시키는, 게놈 또는 기타 다른 핵산 표적 유래의 DNA 서열에 측접할 수 있다. 대안적으로, 상기 벡터 및 이의 발현 카세트는 통합되지 않은 상태(예를 들어, 에피좀)로 유지될 수 있고, 이 경우에 상기 벡터는 통상적으로, 이종성 벡터 DNA의 복제를 제공할 수 있는 복제 기점을 포함한다.

[0219] 벡터상에 존재하는 통상의 유전자는 단백질을 암호화하는 유전자이고, 이의 발현은 상기 단백질을 함유하는 재조합 세포가 상기 단백질을 발현하지 않는 세포로부터 구분될 수 있도록 한다. 상기 유전자 및 이의 상응하는 유전자 생성물은 선택가능한 마커 또는 선택 마커로서 호칭된다. 임의의 광범위하게 다양한 선택가능한 마커는 프로토테카를 형질 전환시키는데 유용한 이식 유전자 구조물에 사용될 수 있다. 적합한 선택가능한 마커의 예로서는 G418 내성 유전자, 질산염 환원 효소 유전자(문헌[Dawson et al. (1997), Current Microbiology 35:356-362] 참조), 하이그로마이신 포스포트랜스퍼라제 유전자(HPT; 문헌[Kim et al. (2002), Mar. Biotechnol. 4:63-73] 참조), 네오마이신 포스포트랜스퍼라제 유전자, 및 플레오마이신에 대해 내성을 부여하는 ble 유전자(문헌[Huang et al. (2007), Appl. Microbiol. Biotechnol. 72:197-205] 참조)를 포함한다. 항생제에 대한 미세 조류의 민감성을 측정하는 방법은 널리 공지되어 있다. 예를 들어 문헌[Mol Gen Genet. 1996 Oct 16;252(5):572-9]은 본원에 기술된 바와 같은 수크로스 인버타제 및 또한 본원에 기술된 바와 같은 티아민 영양요구성 보완을 기술한다.

[0220] 항생제를 바탕으로 하지 않는 기타 다른 선택가능한 마커도 또한 일반적으로 미세 조류(예를 들어, 프로토테카종)를 형질 전환하는데 유용한 이식 유전자 구조물 내에 사용될 수 있다. 이전에는 미세 조류에 의해 이용될 수 없었던 임의의 탄소원을 이용하는 능력을 부여하는 유전자도 또한 선택가능한 마커로서 사용될 수 있다. 예를 들어, 프로토테카 모리포르미스 균주는 통상적으로 적어도 수크로스 상에서 잘 생장하지 못한다. 수크로스 인버타제 유전자를 함유하는 구조물의 사용으로 양성 형질 전환체가 탄소 기질인 수크로스 상에서 생장하는 능력을 부여할 수 있다. 기타 다른 선택가능한 마커와 함께 선택가능한 마커로서 수크로스를 활용하는 것에 관한 부연설명은 이하 섹션 IV에 논의되어 있다.

[0221] 본 발명의 목적을 위해, 본 발명의 재조합 숙주 세포를 제조하는데 사용되는 발현 벡터는 상기 유전자들 중 하나가 선택가능한 마커인 경우, 2개 이상, 흔히 3개의 유전자를 포함할 것이다. 예를 들어, 본 발명의 유전자 조작된 프로토테카는 선택가능한 마커뿐만 아니라 하나 이상의 외인성 유전자, 예를 들어 수크로스 인버타제 유전자 또는 아실 ACP-티오에스테라제 유전자를 포함하는 본 발명의 벡터로 형질 전환될 수 있다. 하나 또는 2개의 유전자는 지질 수율 및 지방산 에스테르로의 전환을 증진시키기 위해 상기 유전자들의 발현의 상대 타이밍이 제어될 수 있게 하는 유도성 프로모터를 사용하여 발현될 수 있다. 2개 이상의 외인성 유전자의 발현은 동일한 유도성 프로모터의 제어 하에 또는 상이한 유도성(또는 지속성) 프로모터의 제어 하에 있을 수 있다. 후자 상황에서, 제1 외인성 유전자의 발현은 제1 기간 동안 유도될 수 있고(제1 기간 동안에 제2 외인성 유전자의 발현은 유도되거나 유도되지 않을 수 있음) 제2 외인성 유전자의 발현은 제2 기간 동안 유도될 수 있다(제2 기간 동안에 제1 외인성 유전자의 발현은 유도되거나 유도되지 않을 수 있음).

[0222] 다른 구체예에서, (임의의 선택가능한 마커에 추가로) 2개 이상의 외인성 유전자는 지방 아실-ACP 티오에스테라제 및 지방 아실-CoA/알데하이드 환원 효소이고, 상기 조합 작용은 알코올 생성물을 생성시킨다. 추가로 제한없이 알데하이드를 생성하기 위한 지방 아실-ACP 티오에스테라제 및 지방 아실-CoA 환원 효소를 포함하는 외인성 유전자의 기타 다른 조합이 제공된다. 하나의 구체예에서, 벡터는 알칸을 생성하기 위한 지방 아실-ACP 티오에스테라제, 지방 아실-CoA 환원 효소 및 지방 알데하이드 탈카보닐화 효소의 조합을 제공한다. 이들 각각의 구체예에서, 외인성 유전자 중 하나 이상은 유도성 프로모터를 사용하여 발현될 수 있다.

[0223] 2개 이상의 외인성 유전자를 발현하는 본 발명의 기타 다른 예시적인 벡터는 수크로스 수송체 및 수크로스 인버타제 효소 둘 다를 암호화하는 것들 및 선택가능한 마커 및 분비된 수크로스 인버타제 둘 다를 암호화하는 것들을 포함한다. 벡터중 어느 하나의 유형으로 형질 전환된 재조합 프로토테카는 탄소원으로서 사탕수수(및 사탕수수 유래 당)를 사용하는 조작된 능력으로 인하여 보다 낮은 제조 비용으로 지질을 생산한다. 전술한 2개의 외인성 유전자의 삽입은, 유도적 돌연변이 유발 및/또는 무작위 돌연변이 유발을 통해 다당류 생합성의 파괴와 조합될 수 있으며, 이는 지질 생산으로 훨씬 더 많은 탄소가 유입되게 한다. 개별적으로 및 조합적으로, 영양성 전환, 지질 생산을 변화시키기 위한 조작 및 외인성 효소를 사용한 처리는 미생물에 의해 생산되는 지질의 조성을 변형시킨다. 상기 변형은 미생물 내에서 생산되는 지질의 양, 기타 다른 지질과 비교하여 생산되는 하나 이상의 탄화수소 종의 양, 및/또는 생산되는 지질 종의 유형의 변화일 수 있다. 예를 들어, 미세 조류는 보다 높은 양 및/또는 %의 TAG(트리아실글리세라이드)를 생산하도록 조작될 수 있다.

[0224] 재조합 단백질의 최적의 발현을 위해, 형질 전환될 숙주 세포에 의해 우선적으로 사용되는 코돈을 가지는 mRNA를 생산하는 암호화 서열을 사용하는 것이 이롭다. 따라서, 이식 유전자의 적절한 발현은, 이식 유전자의 코돈 사용 빈도가 이식 유전자가 발현되는 유기체의 특이적 코돈 편향에 부합하는 것을 필요로 할 수 있다. 상기 효과의 기저를 이루는 정확한 기작은 많지만, 상기 필요성이 충족되는 경우 이식 유전자 전령 RNA(mRNA)의 보다 효율적인 번역과 커플링되는, 세포 내에서 합성되는 단백질과 이용가능하게 아미노아실화된 tRNA 풀의 적당한 균형을 포함한다. 이식 유전자 내 코돈 사용 빈도가 최적화되지 않은 경우, 이용가능한 tRNA 풀은 이식 유전자 mRNA의 가능한 불안정성 및 리보솜 멈춤(stalling) 종결을 초래하는 이종성 mRNA의 효율적인 번역을 가능하게 하기에 충분하지 않다.

[0225] 본 발명은 프로토테카 내에서 재조합 단백질의 성공적인 발현을 위해 유용한 코돈 최적화된 핵산을 제공한다. 프로토테카 종에서 코돈 사용 빈도는 프로토테카 모리포르미스로부터 분리된 cDNA 서열을 연구함으로써 분석되었다. 상기 분석은 24,000개 코돈에 걸친 탐색을 나타내며, 이의 결과는 하기의 표 2에 나타내었다.

표 2

프로토테카 균주에서 바람직한 코돈 사용 빈도

Ala	GCG	345 (0.36)	Asn	AAT	8 (0.04)
	GCA	66 (0.07)		AAC	201 (0.96)
	GCT	101 (0.11)			
	GCC	442 (0.46)	Pro	CCG	161 (0.29)
				CCA	49 (0.09)
Cys	TGT	12 (0.10)		CCT	71 (0.13)
	TGC	105 (0.90)		CCC	267 (0.49)
Asp	GAT	43 (0.12)	Gln	CAG	226 (0.82)
	GAC	316 (0.88)		CAA	48 (0.18)
Glu	GAG	377 (0.96)	Arg	AGG	33 (0.06)
	GAA	14 (0.04)		AGA	14 (0.02)
				CGG	102 (0.18)
Phe	TTT	89 (0.29)		CGA	49 (0.08)
	TTC	216 (0.71)		CGT	51 (0.09)
				CGC	331 (0.57)
Gly	GGG	92 (0.12)			
	GGA	56 (0.07)	Ser	AGT	16 (0.03)
	GGT	76 (0.10)		AGC	123 (0.22)
	GGC	559 (0.71)		TCG	152 (0.28)
				TCA	31 (0.06)
His	CAT	42 (0.21)		TCT	55 (0.10)
	CAC	154 (0.79)		TCC	173 (0.31)
Ile	ATA	4 (0.01)	Thr	ACG	184 (0.38)
	ATT	30 (0.08)		ACA	24 (0.05)
	ATC	338 (0.91)		ACT	21 (0.05)
				ACC	249 (0.52)
Lys	AAG	284 (0.98)			
	AAA	7 (0.02)	Val	GTG	308 (0.50)
				GTA	9 (0.01)
Leu	TTG	26 (0.04)		GTT	35 (0.06)
	TTA	3 (0.00)		GTC	262 (0.43)
	CTG	447 (0.61)			
	CTA	20 (0.03)	Trp	TGG	107 (1.00)
	CTT	45 (0.06)			
	CTC	190 (0.26)	Tyr	TAT	10 (0.05)
				TAC	180 (0.95)
Met	ATG	191 (1.00)			
			Stop	TGA/TAG/TAA	

[0226]

[0227] 다른 구체예에서, 재조합 벡터 내 유전자는 프로토테카 균주 이외의 다른 미세 조류 균주를 참조하여 코돈 최적화되었다. 예를 들어, 미세 조류 내 발현용 유전자를 재암호화하는 방법은 미국 특허 제7,135,290호에 기재되어 있다. 코돈 최적화를 위한 추가의 정보는, 예를 들어 GenBank의 코돈 사용 빈도 데이타베이스에서 입수할 수 있다.

[0228] 코돈 최적화 유전자를 가지는 구체예와 관련하여, 최적화된 유전자는 바람직하게 10% 이상, 더욱 바람직하게는 20%, 40%, 60%, 80%, 100% 또는 200% 이상으로 최적화되는 유전자의 유전자 생성물의 발현을 증가시키도록 최적화된다.

[0229] 본 발명의 방법 및 재료는 임의의 외인성 유전자의 프로토테카로의 도입을 허용하는 한편, 수크로스 이용 및 지질 경로 변형에 관한 유전자는 하기의 섹션에서 논의되는 바와 같이 특히 관심이 있는 것이다.

[0231] **IV. 선택가능한 마커**

[0232] 1. <u>수크로스 이용</u>

[0233] 하나의 구체예에서, 본 발명의 재조합 세포는 하나 이상의 외인성 수크로스 이용 유전자를 추가로 함유한다. 다양한 구체예에서, 하나 이상의 유전자는 프럭토키나제, 글루코키나제, 헥소키나제, 수크로스 인버타제 및 수크로스 수송체로 이루어진 군으로부터 선택되는 하나 이상의 단백질을 암호화한다. 예를 들어, 수크로스 수송체 및 수크로스 인버타제의 발현은 프로토테카가 수크로스를 배양 배지로부터 세포에 수송하여 수크로스를 가수 분해시켜 글루코스 및 프럭토스를 생성하도록 한다. 임의로, 내인성 헥소키나제 활성이 프럭토스의 최대 인산화에 불충분한 경우에도 또한 프럭토키나제는 발현될 수 있다. 적합한 수크로스 수송체의 예로서는 Genbank 수탁 번호 제CAD91334호, 제CAB92307호, 및 제CAA53390호가 있다. 적합한 프럭토키나제의 예로서는 Genbank 수탁 번호 제P26984호, 제P26420호 및 제CAA43322호가 있다.

[0234] 하나의 구체예에서, 본 발명은 수크로스 인버타제를 분비하는 숙주 세포를 제공한다. 수크로스 인버타제의 분비는 수크로스를 세포 내로 수송할 수 있는 수송체의 발현에 대한 필요성을 제거한다. 이것은 분비된 인버타제가 수크로스 분자의 글루코스 분자 및 프럭토스 분자로의 전환을 촉매하기 때문이며, 이들 분자 둘 다는 본 발명에 의해 제공된 미생물에 의해 수송되고 이용될 수 있다. 예를 들어, 분비 신호(예를 들어, 서열번호 4(효모 유래), 서열번호 5(고등 식물 유래), 서열번호 6(진핵 공통 분비 신호) 및 서열번호 7(고등 식물 및 진핵 공통 서열 유래의 신호 서열의 조합체))와 함께 수크로스 인버타제(예를 들어, 서열번호 3)의 발현은 세포 외부에서 인버타제 활성을 생성시킨다. 본원에 개시된 유전자 조작 방법에 의해 가능해진 이러한 단백질의 발현은 이미 에너지원으로서 세포외 글루코스를 이용할 수 있는 세포가 세포외 에너지원으로 수크로스를 이용할 수 있게 한다.

[0235] 수크로스 함유 배지에서 인버타제를 발현하는 프로토테카 종은 오일 생산을 위해 바람직한 미세 조류 종이다. 이러한 완전히 활성인 단백질의 발현 및 세포외 표적화는, 생성된 숙주 세포가 수크로스 상에서 생장할 수 있게 하는 반면에, 이 숙주 세포의 형질 전환되지 않은 대응 세포는 수크로스 상에서 생장할 수 없다. 그러므로, 본 발명의 구체예는 코돈 최적화 인버타제 유전자, 예를 들어 효모 인버타제 유전자(이에 제한되는 것은 아님)가 미세 조류 세포의 게놈에 통합되어 이 인버타제 유전자가 발현되는(이 유전자의 발현 여부는 인버타제 활성과 수크로스 가수 분해에 의해 평가됨) 재조합 미세 조류(예를 들어, 프로토테카) 세포를 제공한다. 인버타제 유전자는 재조합 세포 내 선택가능한 마커로서 유용하므로, 세포는 수크로스 상에서 생장할 수 있는 반면, 이 세포의 형질 전환되지 않은 대응 세포는 수크로스 상에서 생장할 수 없고, 조류 분자 유전학에 대하여 강력한 선택가능한 마커로서 인버타제를 사용하여 재조합 숙주 세포를 선택하는 방법에서도 유용하다.

[0236] 프로토테카에서 수크로스 인버타제의 성공적인 발현은 또한, 이것이 이종성(재조합) 단백질이 조류 세포에서 발현되고 완전한 활성 및 기능적 형태로서 세포 외부로 및 배양 배지 내로 성공적으로 이동할 수 있음을 입증한다는 점에서 본 발명의 또 다른 양태를 설명한다. 따라서, 본 발명의 구체예는 미세 조류 내 광범위하고 다양한 집합체의 이종성 단백질을 발현하고 이들을 숙주 세포 외부로 분비시키기 위한 방법 및 시약을 제공한다. 상기 단백질은, 예를 들어 산업용 효소, 예를 들어 리파제, 프로테아제, 셀룰라제, 펙티나제, 아밀라제, 에스테라제, 산화 환원 효소, 트랜스퍼라제, 락타제, 이소머라제 및 인버타제, 및 치료학적 단백질, 예를 들어 생장 인자, 사이토킨, 2개의 경쇄 및 2개의 중쇄를 포함하는 전장 항체, Fab, scFv(단일쇄 가변 단편), 카멜리드형 항체, 항체 단편, 항체-단편 융합체, 항체-수용체 융합체, 인슐린, 인터페론 및 인슐린형 생장 인자를 포함한다.

[0237] 프로토테카 내 수크로스 인버타제의 성공적인 발현은 또한, 이것이 프로토테카 내 단백질의 분비를 지시하기 위한 조류 내 진균류 수송 펩티드의 사용을 위한 방법 및 시약을 제공하고 펩티드가 프로토테카 세포에서 수송 펩티드로서 기능할 수 있는지의 여부 및 펩티드가 기능하는 능력을 결정하기 위한 방법 및 시약을 제공한다는 점에서 본 발명의 다른 양태를 설명한다. 본 발명의 방법 및 시약은 단백질을 세포 외로 성공적으로 수송할 수 있는 기타 다른 수송 펩티드를 확인하기 위한 도구 및 플랫폼으로서 사용될 수 있고, 이때 효모 인버타제가 이들 방법에서 우수한 용도를 갖는다. 본 실시예에서 입증된 바와 같이, 내인성 효모 인버타제 수송 펩티드의 제거, 숙주 조류에 내인성이거나 기타 다른 공급원(진핵, 원핵 및 바이러스)으로부터 유래하는 기타 다른 수송 펩티드에 의한 이의 대체로 관심의 임의의 펩티드가 세포로부터 단백질의 방출을 유도하는 수송 펩티드로서 기능할 수 있는지의 여부를 확인할 수 있다.

[0238] 적합한 수크로스 인버타제의 예로서는 Genbank 수탁 번호 제CAB95010호, 제NP_012104호 및 제CAA06839호에 의해 확인된 것들을 포함한다. 적합한 인버타제의 비제한적인 예는 하기 표 3에 열거되어 있다. 각각의 열거된 인버타제에 대한 아미노산 서열은 하기의 서열 목록에 포함된다. 몇몇 경우에, 본 발명의 방법 및 벡터에 사용하기에 적합한 외인성 수크로스 이용 유전자는 표 3로부터 선택되는 수크로스 인버타제와 적어도 40%, 50%, 60%, 75%, 또는 90% 이상의 아미노산 동일성을 갖는 수크로스 인버타제를 암호화한다.

표 3

수크로스 인버타제

기재	유기체	GenBank 수탁 번호	서열번호
인버타제	*키코리움 인티부스*	Y11124	서열번호 20
인버타제	*쉬조사카로마이세스 폼베*	AB011433	서열번호 21
베타-프럭토푸라노시다제 (인버타제)	*피키아 아노말라*	X80640	서열번호 22
인버타제	*데바리오마이세스 옥시덴탈리스*	X17604	서열번호 23
인버타제	*오리재 사티바*	AF019113	서열번호 24
인버타제	*알리움 세파*	AJ006067	서열번호 25
인버타제	*베타 불가리스 아종 불가리스*	AJ278531	서열번호 26
베타-프럭토푸라노시다제 (인버타제)	*비피도박테리움 브레베 UCC2003*	AAT28190	서열번호 27
인버타제	*사카로마이세스 세레비지아에*	NP_012104	서열번호 8 (뉴클레오티드) 서열번호 28 (아미노산)
인버타제 A	*지모모나스 모빌리스*	AAO38865	서열번호 29

[0239]

[0240] 프로토테카에 의한 배양 배지로의 인버타제의 분비는 세포가 이들이 순수 시약 등급의 글루코스 상에 생장하는 바와 같이 사탕수수 가공으로부터 비롯된 폐기 당밀 상에서 생장할 수 있게 하고, 사탕수수 가공의 이러한 낮은 가치의 폐기물의 사용은 지질 및 기타 다른 오일의 제조에서 상당한 비용 절감을 제공할 수 있다. 따라서, 본 발명은 (i) 수크로스 및 (ii) 수크로스 인버타제 효소를 포함하는 배양 배지 및 프로토테카 미생물 집단을 함유하는 미생물 배양물을 제공한다. 다양한 구체예에서 배양물 내 수크로스는 수수, 사탕무, 사탕수수, 당밀 또는 해중합된 셀룰로스 물질(이것은 임의로 리그닌을 함유할 수 있음)로부터 유래한다. 다른 양태에서, 본 발명의 방법 및 시약은 재조합 미세 조류 또는 기타 다른 미생물에 의해 사용될 수 있는 공급원료의 수 및 유형을 상당히 증가시킨다. 본원에 예시된 미생물은 이들이 수크로스를 이용할 수 있도록 변형되는 한편, 본 발명의 방법 및 시약이 적용되어 셀룰로스와 같은 공급원료가 셀룰라제, 펙티나제, 이소머라제 등을 분비하는 능력을 갖는 본 발명의 조작된 숙주 미생물에 의해 이용될 수 있도록 하고 효소 반응의 분해 생성물은 더 이상 단순히 내성이 아니고 숙주에 의해 탄소원으로 이용된다. 이에 관한 예는 이하, 및 올리고당 내 α-갈락토실 결합, 예를 들어 라피노스 및 스타키오스(이들 올리고당 2개는 농업 폐기물 스트림에서 발견됨) 내에 함유된 α-갈락토실 결합을 가수 분해하는 능력을 부여하는, 분비성 α-갈락토시다제를 발현하도록 조작된 미생물에 관한 실시예에 기술되어 있다.

[0241] **2. 알파-갈락토시다제 발현**

[0242] 전술한 바와 같이, 수크로스 인버타제의 발현으로 프로토테카 세포가 수크로스를 (이당류 수크로스 내 프럭토스 및 글루코스 분자 사이 α-결합을 가수 분해하는 효소를 통해) 탄소원으로서 더욱 효과적으로 이용하는 능력을 부여하는 한편, 올리고당 내 기타 다른 유형의 α-결합을 가수 분해하는 기타 다른 효소의 발현으로 프로토테카 세포가 기타 다른 탄소원을 이용하는 능력을 부여할 수 있다. 이러한 효소의 발현(그리고 결과로 얻어진, 프로토테카 및 기타 다른 미세 조류 세포는 일반적으로 이용할 수 없는 탄소원을 이용하는 능력)은, 이러한 탄소원 상에서 생장할 수 있는 양성 클론을 선택할 수 있게 함으로써, 이러한 유전자 이식 프로토테카 세포에 대한 선

택가능한 마커로서 사용될 수 있다.

[0243] 하나의 구체예에서, 본 발명의 재조합 프로토테카 세포는 다당류 분해 효소를 암호화하는 하나 이상의 외인성 유전자를 추가로 함유한다. 다양한 구체예에서, 다당류 분해 효소를 암호화하는 하나 이상의 유전자는 분비된 α-갈락토시다제를 암호화하는 유전자이다. 프로토테카 세포 내에서 외인성 분비 α-갈락토시다제의 발현으로, 이와 같이 형질 전환된 균주가 D-갈락토실 결합, 예를 들어 갈락토스 및 글루코스 단당류 단위 사이의 α-결합을 함유하는 당(탄소원) 상에서 생장할 수 있는 능력을 부여한다. 외인성 분비 α-갈락토시다제를 발현하는 프로토테카 균주는 이당류, 예를 들어 멜리비오스(α-D-갈락토스-글루코스로 구성된 이당류)를 이용할 수 있을 것이다.

[0244] 당, 예를 들어 라피노스(α-결합된 갈락토스-글루코스-프럭토스로 이루어진 삼당류) 및 스타키오스(2개의 α-결합 D-갈락토스 단위와, 이어서 α-결합 글루코스 및 프럭토스를 포함하는 사당류)는 농업 폐기물 스트림, 예를 들어 사탕무 펄프(라피노스) 및 대두 밀(스타키오스) 내에 상당한 비율로 존재한다. 이와 같은 농업 잔류물은, 이 잔류물을 이용할 수 있는 미생물(예를 들어, 프로토테카)에 의해 오일로 전환하는데 사용되기 전 상당한 탄소원을 대표한다.

[0245] 프로토테카 균주는 올리고당, 예를 들어 라피노스 및 스타키오스를 임의의 상당한 양으로 이용할 수 없거나 전혀 이용할 수 없다. 라피노스와 스타키오스의 경우, (전술한 바와 같이) 수크로스 인버타제를 발현하는 유전자 이식 균주는 수크로스의 α-갈락토실 유도체 내 프럭토스와 글루코스 간 α-결합을 가수 분해하는 능력을 가지지만, 상기 수크로스 인버타제는 이와 같은 당 내 잔여 α-결합은 절단하지 않을 것이며, 생성된 이당류는 이용될 수 없으므로, 올리고당의 나머지는 이용되지 않은 상태로 남게 된다. 다른 구체예에서, 본 발명의 재조합 프로토테카 세포는 수크로스 인버타제를 암호화하는 외인성 유전자와 α-갈락토시다제를 암호화하는 외인성 유전자 둘 다를 포함한다. 따라서, 수크로스 인버타제와 α-갈락토시다제 둘 다를 발현하는 균주는 올리고당, 예를 들어 라피노스와 스타키오스를 완전히 가수 분해하여 구성 단량체들을 소모할 수 있을 것이다. 뿐만 아니라, α-갈락토시다제 암호화 유전자도 형질 전환에 있어서 선택가능한 마커로서 사용될 수 있다. 외인성 α-갈락토시다제 유전자를 함유하는 클론은 멜리비오스 상에서 생장하는 능력을 가질 것이다. 프로토테카 균주에 사용하기 적당한 α-갈락토시다제 유전자의 예로서는 사카로마이세스 칼베르젠시스로부터 유래하는 MEL1 유전자와 아스퍼질러스 나이저로부터 유래하는 AglC 유전자를 포함한다. 흥미로운 점은, 유전자가 프로토테카 균주 내 선호 코돈 사용 빈도에 따라서 최적화될지라도, 모든 α-갈락토시다제 유전자가 프로토테카 종 내에서 작용하는 것으로 관찰되지는 않는다는 점이다. 이하 실시예는 에스. 칼베르젠시스로부터 유래하는 코돈 최적화 MEL1 유전자와 에이. 나이저로부터 유래하는 AglC 유전자로 형질 전환되되, 고등 식물인 시아몹시스 테트라고노로바(Cyamopsis tetragonoloba)(구아 콩)로부터 유래하는 α-갈락토시다제 암호화 유전자로는 형질 전환되지 않을 때, 유전자 이식 프로토테카 세포가 멜리비오스 상에서 생장하는 능력을 입증한다.

[0246] **3. 티아민 영양 요구성 보완**

[0247] 프로토테카 균주, 예를 들어 프로토테카 모리포르미스는 티아민 영양 요구성인 것으로 알려져 있는데(예를 들어, 문헌[Ciferri, O. (1956) Nature, v.178, pp. 1475-1476] 참조), 이는 상기 균주가 생장용 영양 배지 중에 티아민을 필요로 함을 의미한다. 티아민 영양 요구성은 돌연 변이, 또는 티아민 생합성 경로에 관여하는 효소들의 발현 부족의 결과일 수 있다. 이후, 티아민 생합성 경로에서 결여된 효소(들)를 발현하는 보완 유전자 이식 균주는 티아민이 첨가되지 않는 경우에도 생장할 수 있으므로, 영양 배지의 비용을 줄일 수 있을 뿐만 아니라, 생성된 미세 조류 바이오매스를 동물 먹이로서 사용함에 있어서 더욱 바람직하게 만들 수 있다. 티아민 생합성 경로 효소로 보완하는 것은 또한 유전자 이식된 유전자가 티아민을 함유하지 않는 평판/배지 상에서 생장할 수 있는 능력을 부여하므로, 선택가능한 마커로서 사용될 수도 있다.

[0248] 하나의 구체예에서, 본 발명의 재조합 프로토테카 세포는 티아민 생합성 경로 효소를 암호화하는 하나 이상의 외인성 유전자를 추가로 함유한다. 다른 구체예에서, 본 발명의 재조합 프로토테카 세포는 조류, 식물 또는 시아노박테리아 공급원으로부터 유래하는 하이드록시메틸피리미딘 포스페이트 합성 효소를 암호화하는 외인성 유전자를 포함한다. 또 다른 구체예에서, 하이드록시메틸피리미딘 포스페이트 합성 효소는 THIC 유전자에 의해 암호화된다. 또 다른 구체예에서, THIC 유전자는 코코믹사(Coccomyxa) C-169 THIC, 아라비돕시스 탈리아나(Arabidopsis thaliana) THIC, 또는 시네코시스티스 종(Synechocystis sp.) PCC 6803 thiC이다. 이하 실시예는 티아민 원 영양성이 복구된 프로토테카 모리포르미스 UTEX 1435의 조작에 대해 상세히 기술한다.

[0250] **V. 지질 경로 조작**

[0251] 수크로스 함유 공급원료와 같은 공급원료를 사용하는 미세 조류 또는 기타 다른 미생물의 능력을 변형시키는 것에 추가로, 본 발명은 또한 생산된 지질의 특성 및/또는 비율을 변화시키도록 변형된 재조합 미세 조류 또는 기타 다른 미생물을 제공한다. 상기 경로는 추가로 또는 대안적으로 지방산 경로에서 지질 및 중간체의 효소 가공을 통해 생성된 다양한 지질 분자의 특성 및/또는 비율을 변화시키도록 변형될 수 있다. 다양한 구체예에서, 본 발명의 재조합 프로토테카 세포는 이들의 형질 전환되지 않은 대응 세포에 비해 단위 용적당 및/또는 단위 시간당 최적화된 지질 수율, 탄소 사슬 길이(예를 들어, 재생 디젤 생산 또는 지질 공급원료를 필요로 하는 산업적 화학 제품), (임의로 0까지) 감소 또는 증가된 수 및/또는 위치의 이중 결합, 지방산의 하이드록실화 및 지질의 특정 종 또는 독특한 지질 집단의 수소:탄소 비율의 증가를 가진다.

[0252] 특정 구체예에서, 지방산 합성으로의 대사에서 분기점을 제어하는 하나 이상의 주요 효소는 지질 생산을 개선시키기 위해 상향 조절되거나 하향 조절되었다. 상향조절은, 예를 들어 전사를 증가시키는 강한 프로모터 및/또는 인핸서 요소를 사용하여, 관심의 효소를 암호화하는 유전자가 발현되는 발현 구조물로 세포를 형질 전환시킴으로써 성취될 수 있다. 상기 구조물은 선택가능한 마커를 포함하여 상기 형질 전환체는 선택될 수 있고 이에 의해 유전자 유지, 및 가능하게는 구조물의 증폭과 암호화된 효소의 발현 수준의 증가를 가져올 수 있다. 본 발명의 방법에 따른 상향 조절을 위해 적합한 효소의 예는 피루베이트를 아세틸-CoA로 전환시키는 역할을 하는 피루베이트 탈수소화 효소 (예를 들어, 미세 조류 유래의 일부는 Genbank 수탁 번호 제NP_415392호; 제AAA53047호; 제Q1XDM1호; 및 제CAF05587호를 포함함)를 포함한다. 피루베이트 탈수소화 효소의 상향조절은 아세틸-CoA의 생산을 증가시킬 수 있고, 이에 의해 지방산 합성을 증가시킬 수 있다. 아세틸-CoA 카복실라제는 지방산 합성에서 초기 단계를 촉매한다. 따라서, 상기 효소는 지방산의 생산을 증가시키기 위해 상향조절될 수 있다(예를 들어, 미세 조류 유래의 일부는 Genbank 수탁 번호 제BAA94752호; 제AAA75528호; 제AAA81471호; 제YP_537052호; 제YP_536879호; 제NP_045833호; 및 제BAA57908호를 포함함). 지방산 생산은 또한 지방산 합성 동안에 성장하는 아실 사슬을 운반하는 아실기 운반 단백질(ACP)의 상향 조절에 의해 증가될 수 있다(예를 들어, 미세 조류 유래의 일부는 Genbank 수탁 번호 제A0T0F8호; 제P51280호; 제NP_849041호; 제YP_874433호를 포함함). 글리세롤-3-포스페이트 아실트랜스퍼라제는 지방산 합성의 속도 제한 단계를 촉매한다. 상기 효소의 상향 조절은 지방산 생산을 증가시킬 수 있다(예를 들어, 미세 조류 유래의 일부는 Genbank 수탁 번호 제AAA74319호; 제AAA33122호; 제AAA37647호; 제P44857호; 및 제ABO94442호를 포함함).

[0253] 유전자의 상향 조절 및/또는 하향 조절은 지방산 생합성 경로의 유전자의 발현을 제어하는 범용 조절제에 적용될 수 있다. 따라서, 지방산 합성의 하나 이상의 범용 조절제는 적절히 각각 다수의 지방산 합성 유전자의 발현을 억제하거나 증진시키고 궁극적으로 지질 생산을 증가시키도록 상향 또는 하향 조절될 수 있다. 이의 예로서는 스테롤 조절 요소 결합 단백질(SREBP), 예를 들어 SREBP-1a 및 SREBP-1c(예를 들어, Genbank 수탁 번호 제NP_035610호 및 제Q9WTN3호를 참조)를 포함한다.

[0254] 본 발명은 또한 지질 변형 효소, 예를 들어 지방 아실-ACP 티오에스테라제(표 4 참조), 지방 아실-CoA/알데하이드 환원 효소(표 6 참조), 지방 아실-CoA 환원 효소(표 7 참조), 지방 알데하이드 탈카보닐화 효소(표 8 참조), 지방 알데하이드 환원 효소, 불포화 효소(예를 들어, 스테아로일-ACP 불포화 효소 및 지방 아실 불포화 효소 및 스쿠알렌 합성 효소(GenBank 수탁 번호 제AF205791호를 참조))를 암호화하는 하나 이상의 외인성 유전자를 함유하도록 변형된 재조합 프로토테카 세포를 제공한다. 지방 아실-ACP 티오에스테라제는 통상적으로 지질을 직접 화학 변형시키지는 않지만, 본 발명의 구체예에 따른 조작으로 세포의 지방산 프로필을, 특히 사슬 길이 및 이중 결합 분포 측면에서 변화시킬 수 있다. 몇몇 구체예에서, 지방 아실-ACP 티오에스테라제와 자연적으로 공동 발현되는 아실기 운반 단백질을 암호화하는 유전자는 임의로 기타 다른 지질 변형 효소를 암호화하는 하나 이상의 유전자와 함께, 프로토테카 또는 기타 다른 미세 조류 세포 또는 미생물 세포로 형질 전환된다. 다른 구체예에서, ACP와 지방 아실-ACP 티오에스테라제는 이것들이 특정 조직이나 유기체 내에서 자연적으로 공동 발현되든지 발현되지 않든지 상관없이, 본 발명의 방법 및 미생물에서 상기 ACP와 지방 아실-ACP 티오에스테라제가 함께 사용될 때 이점을 부여하는 서로에 대한 친화성을 가질 수 있다. 따라서, 본 발명의 구체예는 상기 효소들의 자연적으로 공동 발현되는 쌍들뿐만 아니라, ACP로부터 길이 특이적 탄소 사슬의 절단을 촉진하기 위해 서로 상호작용하기 위한 친화성을 공유하는 쌍들을 고려한다.

[0255] 또 다른 구체예에서, 불포화 효소를 암호화하는 외인성 유전자는 지질 포화와 관련하여 변형을 제공하기 위해 기타 다른 지질 변형 효소를 암호화하는 하나 이상의 유전자와 함께 미세 조류 또는 기타 다른 미생물 세포로 형질 전환된다. 다른 구체예에서, 내인성 불포화 효소 유전자는 (예를 들어, 부가의 유전자 유래 복사체들을 도입함으로써) 미세 조류 세포 또는 기타 다른 미생물 세포 내에서 과발현된다. 스테아로일-ACP 불포화 효소(예를 들어, GenBank 수탁 번호 제AAF15308호; 제ABM45911호; 및 제 AAY86086호 참조)는, 예를 들어 스테아로일-

ACP의 올레오일-ACP로의 전환을 촉매한다. 상기 유전자의 상향 조절은 세포에 의해 생산된 단일불포화 지방산의 비율을 증가시킬 수 있는 반면, 하향 조절은 단일 불포화 지방의 비율을 감소시킬 수 있다. 예시적 목적으로, 스테아로일-ACP 불포화 효소(SAD)는 C18:0 전구체로부터 C18:1 지방산을 합성하는데 관여한다. 불포화 효소의 다른 군으로서는 지방 아실 불포화 효소(FAD), 예를 들어 델타 12 지방산 불포화 효소(Δ12 FAD)가 있다. 이러한 불포화 효소는 또한 지질 포화도에 대한 변형을 제공한다. 예시적 목적으로, 델타 12 지방산 불포화 효소는 C18:1 전구체로부터 C18:2 지방산을 합성하는데 관여한다. 유사하게, ω-6 지방산 불포화 효소, ω-3 지방산 불포화 효소, 또는 ω-6-올리에이트 불포화 효소와 같은 하나 이상의 글리세로지질 불포화 효소의 발현은 포화 지방산에 대한 불포화 지방산의 비율을 변화시키기 위해 제어될 수 있다. 몇몇 구체예에서, 불포화 효소는 불포화 효소가 특정 탄소 길이 기질 또는 특정 범위 내 탄소 길이를 갖는 기질 내에서 위치 특이적 변형을 만들 수 있도록 원하는 탄소 사슬 길이와 관련하여 선택될 수 있다. 다른 구체예에서, 만일 원하는 지방산 프로필이 단일불포화 지방(예를 들어, C16:1 및/또는 C18:1)에 있어서 증가하면, SAD의 과발현 또는 이종 SAD의 발현은 (예를 들어, 돌연 변이, RNAi, 안티센스 또는 내인성 불포화 효소 유전자의 녹아웃 등을 통하여) 지방 아실 불포화 효소(FAD) 또는 다른 불포화 효소 유전자의 침묵 또는 불활성화와 결합될 수 있다.

[0256] 다른 구체예에서, 미세 조류 세포 또는 기타 다른 미생물 세포는 돌연 변이된 내인성 불포화 효소 유전자를 가지도록 변형되었는데, 이 경우 상기 돌연 변이는 상기 유전자 또는 불포화 효소가 불활성이 되도록 만든다. 몇몇 경우에 있어서, 돌연 변이된 내인성 불포화 효소 유전자는 지방산 불포화 효소(FAD)이다. 다른 경우에 있어서, 돌연 변이된 내인성 불포화 효소 유전자는 스테아로일 아실기 운반 단백질 불포화 효소(SAD)이다. 이하 실시예 6은 프로토테카 내에서 스테아로일-ACP 불포화 효소 및 델타 12 지방산 불포화 효소의 표적화된 제거 또는 녹아웃을 기술한다. 실시예 6은 또한 내인성 불포화 효소 유전자의 발현을 감소시키는 RNAi 또는 안티센스 구조물의 사용을 기술한다.

[0257] 몇몇 경우에 있어서, 원하는 지질 프로필을 생성하는 유전자 이식된 세포를 획득하기 위하여 유전자 조작 기술들 중 하나 이상을 함께 사용하는 것이 유리할 수 있다. 하나의 구체예에서, 미세 조류 세포 또는 기타 다른 미생물 세포는 돌연 변이된 내인성 불포화 효소 유전자와 하나 이상의 외인성 유전자를 포함한다. 비제한적인 실시예에서, 내인성 불포화 효소 유전자가 돌연 변이된 미세 조류 세포 또는 기타 다른 미생물 세포는 또한 외인성 지방 아실-ACP 티오에스테라제 유전자 및/또는 수크로스 인버타제 유전자를 발현할 수도 있다. 이하 실시예 6은 내인성 SAD가 표적화 제거 또는 녹아웃되었고, 신나모뮴 캠포라 C14 선호 티오에스테라제 및 수크로스 인버타제를 발현하는 유전자 이식 프로토테카 세포를 기술한다. 이와 같은 경우, 유전자 이식 프로토테카 세포는 우지에서 살펴볼 수 있는 지질 프로필과 매우 유사한 지질 프로필을 생성한다. 우지는 통상적으로 녹인 소 또는 양의 지방으로부터 유래하는 것으로서, 실온에서 고체이며, 식품, 화장품 및 화학 산업 분야에서 다양한 용도로 사용된다. 우지의 지방산 프로필은, C14:0 4%; C16:0 26%; C16:1 3%; C18:0 14%; C18:1 41%; C18:2 3%; 및 C18:3 1%이다. 이하 실시예 6에 나타낸 바와 같이, 내인성 SAD가 표적화 제거 또는 녹아웃되었고, 씨. 캠포라 C14 선호 티오에스테라제를 발현하는 유전자 이식 프로토테카 세포의 클론은 지질 프로필이, C12 및 이 보다 짧은 탄소 사슬 길이를 가지는 지방산 1% 미만; C14:0 2.74% 내지 6.13%; C16:0 23.07% 내지 25.69%; C18:0 7.02% 내지 11.08%; C18:1 42.03% 내지 51.21%; 및 C18:2 9.37% 내지 13.45%(면적 %로 표시함)이다. 몇몇 경우에 있어서, 유전자 이식 프로토테카 세포는 지질 프로필이, C14:0 3% 내지 5%; C16:0 25% 내지 27%; C18:0 10% 내지 15%; 및 C18:1 40% 내지 45%이다.

[0258] 특정 구체예에서, 본 발명의 미생물은 아실-ACP 티오에스테라제, 아실-CoA/알데하이드 환원 효소, 지방 아실-CoA 환원 효소, 지방 알데하이드 환원 효소, 지방 알데하이드 탈카보닐화 효소 또는 자연적으로 공동 발현되는 아실기 운반 단백질로부터 선택되는 하나 이상의 외인성 유전자를 발현하도록 유전자 조작된다. 적합한 발현 방법은 기타 다른 방법 중에서 유도성 발현 및 구분된 발현을 포함하는 리파제 유전자의 발현과 관련하여 상기 기술되어 있다. 지방 아실-ACP 티오에스테라제는 지질 합성동안에 아실기 운반 단백질(ACP)로부터 지방산을 절단한다. 추가의 효소 가공을 통해, 절단된 지방산은 이어서 조효소와 조합되어 아실-CoA 분자를 생성한다. 이러한 아실-CoA는 알코올을 생성하기 위한 지방 아실-CoA/알데하이드 환원 효소에 대한 것뿐만 아니라 알데하이드를 생성하기 위한 지방 아실-CoA 환원 효소의 효소 활성에 대한 기질이다. 상기 확인된 지방 아실-CoA 환원 효소의 작용에 의해 생산된 알데하이드는 알코올을 생성하기 위한 지방 알데하이드 환원 효소 또는 알칸 또는 알켄을 생성하기 위한 지방 알데하이드 탈카보닐화 효소에 의한 추가의 효소 활성에 대한 기질이다.

[0259] 몇몇 구체예에서, 본원에 기술된 방법에 의해 생성되는 지방산, 글리세로지질, 또는 상응하는 1급 알코올, 알데하이드, 알칸 또는 알켄은 8개, 10개, 12개, 또는 14개의 탄소 원자를 함유한다. 산업적 적용을 위한 디젤, 바이오디젤, 재생 디젤 또는 제트 연료, 또는 상응하는 1급 알코올, 알데하이드, 알칸 및 알켄을 생산하기 위한

바람직한 지방산은 8개 내지 14개의 탄소 원자를 함유한다. 특정 구체예에서, 기타 다른 상응하는 탄화수소 분자뿐만 아니라 상기 지방산은 포화되거나(어떠한 탄소-탄소 이중 결합 또는 삼중 결합을 갖지 않음); 단일 불포화되거나(단일 이중 결합); 다중 불포화되거나(2개 이상의 이중 결합); 선형(환형이 아님)이거나 분지형이다. 연료 생산을 위해 보다 큰 포화도가 바람직하다.

[0260] 상기 직접 기술된 효소는 특정 수의 탄소 원자를 함유하는 기질의 가수 분해에 대해 우선적인 특이성을 갖는다. 예를 들어, 지방 아실-ACP 티오에스테라제는 ACP로부터 12개의 탄소 원자를 갖는 지방산을 절단하는 것을 선호할 수 있다. 몇몇 구체예에서, ACP 및 길이 특이적 티오에스테라제는 이들이 조합체로서 특히 유용하게 하는 서로에 대해 친화성을 가질 수 있다(예를 들어, 외인성 ACP 및 티오에스테라제 유전자는 이들이 유래하는 특정 조직 또는 유기체에서 자연적으로 공동 발현될 수 있다). 따라서, 다양한 구체예에서, 본 발명의 재조합 프로토테카 세포는 기질 내 함유된 탄소 원자의 수와 관련하여 효소 활성(예를 들어, ACP로부터 지방산의 절단, 아실-CoA의 알데하이드 또는 알코올로의 환원, 또는 알데하이드의 알칸으로의 전환)을 촉매하기 위한 특이성을 갖는 단백질을 암호화하는 외인성 유전자를 함유할 수 있다. 다양한 구체예에서 효소 특이성은 8개 내지 34개의 탄소 원자, 바람직하게는 8개 내지 18개의 탄소 원자, 보다 바람직하게는 8개 내지 14개의 탄소 원자를 갖는 기질에 대한 것일 수 있다. 바람직한 특이성은 더 많은 탄소 원자, 즉 18개 탄소 원자보다는 더 적은, 즉 12개 탄소 원자를 갖는 기질에 대한 것이다.

[0261] 본 발명의 미생물 및 방법과 함께 사용하기에 적합한 기타 다른 지방 아실-ACP 티오에스테라제는 표 4에 열거된 것들을 제한없이 포함한다.

표 4

[0262] 지방 아실-ACP 티오에스테라제 및 GenBank 수탁 번호

움벨루라리아 캘리포니카 지방 아실-ACP 티오에스테라제(GenBank #AAC49001)
신나모뭄 캄포라 지방 아실-ACP 티오에스테라제(GenBank #Q39473)
움벨루라리아 캘리포니카 지방 아실-ACP 티오에스테라제(GenBank #Q41635)
미리스티카 프래그란스 지방 아실-ACP 티오에스테라제(GenBank #AAB71729)
미리스티카 프래그란스 지방 아실-ACP 티오에스테라제(GenBank #AAB71730)
엘라에이스 기닌시스 지방 아실-ACP 티오에스테라제(GenBank #ABD83939)
엘라에이스 기닌시스 지방 아실-ACP 티오에스테라제(GenBank #AAD42220)
엘라에이스 기닌시스 지방 아실-ACP 티오에스테라제(GenBank#AAD42220.2)
파퓰러스 토멘토사 지방 아실-ACP 티오에스테라제(GenBank #ABC47311)
아라비돕시스 탈리아나 지방 아실-ACP 티오에스테라제(GenBank #NP_172327)
아라비돕시스 탈리아나 지방 아실-ACP 티오에스테라제(GenBank #CAA85387)
아라비돕시스 탈리아나 지방 아실-ACP 티오에스테라제(GenBank #CAA85388)
고시피움 히르서텀 지방 아실-ACP 티오에스테라제(GenBank #Q9SQI3)
쿠페아 란세올라타 지방 아실-ACP 티오에스테라제(GenBank #CAA54060)
쿠페아 후커리아나 지방 아실-ACP 티오에스테라제(GenBank #AAC72882)
쿠페아 칼로필라 아종 메소스테몬 지방 아실-ACP 티오에스테라제(GenBank #ABB71581)
쿠페아 란세올라타 지방 아실-ACP 티오에스테라제(GenBank #CAC19933)
엘라에이스 기닌시스 지방 아실-ACP 티오에스테라제(GenBank #AAL15645)
쿠페아 후커리아나 지방 아실-ACP 티오에스테라제(GenBank #Q39513)
쿠페아 후커리아나 지방 아실-ACP 티오에스테라제(GenBank#Q39513.1)
고시피움 히르서텀 지방 아실-ACP 티오에스테라제(GenBank #AAD01982)
비티스 비니페라 지방 아실-ACP 티오에스테라제(GenBank #CAN81819)
가르시니아 망고스타나 지방 아실-ACP 티오에스테라제(GenBank #AAB51525)
가르시니아 망고스타나 지방 아실-ACP 티오에스테라제(GenBank#AAB51525.1)
브라시카 준세아 지방 아실-ACP 티오에스테라제(GenBank #ABI18986)
마듀카 론지폴리아 지방 아실-ACP 티오에스테라제(GenBank #AAX51637)
브라시카 나푸스 지방 아실-ACP 티오에스테라제(GenBank #ABH11710)
브라시카 나푸스 지방 아실-ACP 티오에스테라제(GenBank#CAA52070.1)
오리자 사티바(인디카 컬티바군) 지방 아실-ACP 티오에스테라제(GenBank #EAY86877)
오리자 사티바(자포니카 컬티바군) 지방 아실-ACP 티오에스테라제(GenBank #NP_001068400)
오리자 사티바(인디카 컬티바군) 지방 아실-ACP 티오에스테라제(GenBank #EAY99617)
쿠페아 후커리아나 지방 아실-ACP 티오에스테라제(GenBank #AAC49269)
얼머스 아메리카나 지방 아실-ACP 티오에스테라제(GenBank #AAB71731)
쿠페아 란세올라타 지방 아실-ACP 티오에스테라제(GenBank #CAB60830)
쿠페아 팔루스트리스 지방 아실-ACP 티오에스테라제(GenBank #AAC49180)
아이리스 게르마니카 지방 아실-ACP 티오에스테라제(GenBank #AAG43858)
아이리스 게르마니카 지방 아실-ACP 티오에스테라제(GenBank #AAG43858.1)
쿠페아 팔루스트리스 지방 아실-ACP 티오에스테라제(GenBank #AAC49179)
미리스티카 프래그란스 지방 아실-ACP 티오에스테라제(GenBank# AAB71729)
미리스티카 프래그란스 지방 아실-ACP 티오에스테라제(GenBank# AAB717291.1)
쿠페아 후커리아나 지방 아실-ACP 티오에스테라제(GenBank #U39834)
움벨루라리아 캘리포니카 지방 아실-ACP 티오에스테라제(GenBank # M94159)
신나모뭄 캄포라 지방 아실-ACP 티오에스테라제(GenBank #U31813)
리시누스 커뮤니스 지방 아실-ACP 티오에스테라제(GenBank#ABS30422.1)

[0263] 이하 실시예는 프로토테카 종 내에서 쿠페아 후커리아나, 움벨루라리아 캘리포니카, 신나모뭄 캄포라, 쿠페아 팔루스트리스, 쿠페아 란세올라타, 아이리스 게르마니카, 미리스티카 프래그란스, 가르시니아 망고스타나, 엘라에이스 기니엔시스, 브라시카 나푸스, 리시누스 커뮤니스 및 얼머스 아메리카나로부터 유래하는 이종 지방 아실-ACP 티오에스테라제의 성공적인 표적화 및 발현을 기술한다. 추가로, 지방산 프로필의 변경은 이러한 이종 지방 아실-ACP 티오에스테라제를 발현하는 숙주 세포 내에서 확인되었다. 실시예들에서 나타낸 바와 같이, 프로토테카 내에서 이러한 이종 티오에스테라제의 발현은 몇몇 종자 곡물 오일을 배합하더라도 시판중인 종자 곡물로

부터 현재는 얻을 수 없는, 진정 독특한 지방산 프로필을 가지는 오일/지질을 생산할 수 있는 유전자 이식 미세조류를 생성한다. 표 5는 통상의 시판 종자 오일의 지방산 프로필을 나타낸다. 이하 시판 중인 종자 오일 데이터는 모두 문헌[US Pharmacopeias Food and Chemicals Codes, 7th Ed. 2010-2011]에서 수집되었다. 우지 데이터는 문헌[National Research Council: Fat Content and Composition of Animal Products (1976)]으로부터 수집한 것이다.

표 5

시판중인 종자 오일의 지질 프로필

	C8:0	C10:0	C12:0	C14:0	C16:0	C18:0	C18:1	C18:0-diOH	C18:1-OH	C18:2	C18:3 α
알.커뮤니스 (피마자 오일)	0	0	0	0	0.9-1.6	1.0-1.8	3.7-6.7	0.4-1.3	83.6-89.0	0	0.2-0.6
씨.뉴시페라 (코코넛 오일)	5.0-9.0	4.0-8.0	44-52	15-21	8.0-11.0	1.0-4.0	5.0-8.0	0	0	0-2.5	0
제트.메이스 (옥수수 오일)	0	0	0	< 1.0	8.0-19.0	0.5-4.0	19-50	0	0	38-65	< 2.0
지.바르바덴스 (목화씨 오일)	0	0	< 0.1	0.5-2.0	17-29	1.0-4.0	13-44	0	0	40-63	0.1-2.1
비.라파, 비.나푸스, 비.준세아 (카놀라)	0	0	< 0.1	< 0.2	< 6.0	< 2.5	> 50	0	0	< 40	< 14
오.유로피아 (올리브)	0	0	0	< 0.1	6.5-20.0	0.5-5.0	56-85	0	0	3.5-20.0	< 1.2
에이.하이포가에아(땅콩)	0	0	< 0.1	< 0.2	7.0-16.0	1.3-6.5	35-72	0	0	13.0-43	< 0.6
이.기넌시스 (팜핵 오일)	3.0-5.0	2.5-6.0	40-52	14.0-18.0	7.0-10.0	1.0-3.0	11.0-19.0	0	0	0.5-4.0	0
이.기넌시스 (팜 오일)	0	0	0	0.5-5.9	32.0-47.0	2.0-8.0	34-44	0	0	7.2-12.0	0
씨.팅크토러스 (홍화)	0	0	< 0.1	< 0.1	2.0-10.0	1.0-10.0	7.0-16.0	0	0	72-81	< 1.5
에이치.아너스 (해바라기)	0	0	< 0.1	< 0.5	3.0-10.0	1.0-10.0	14-65	0	0	20-75	< 0.5
지.맥스 (대두)	0	0	< 0.1	< 0.5	7.0-12.0	2.0-5.5	19-30	0	0	48-65	5.0-10.0
엘. 우시타티시멈 (솔린-플랙스)	0	0	< 0.1	< 0.5	2.0-9.0	2.0-5.0	8.0-60	0	0	40-80	< 5.0
비.파킬 (시어넛)	0	0	0	0	3.8-4.1	41.2-56.8	34.0-46.9	0	0	3.7-6.5	0
우지				4	26	14	41			3	1

[0264]

[0266] 예를 들어, 이러한 통상의 종자 오일 중 어느 것도 C8 또는 C10 지방산을 고량으로 함유하지 않으며, 코코넛 오일과 팜핵 오일은 가장 큰 공급원이지만, 이것들은 둘 다 (C8:C10 지방산) 비율이 약 1:1이다. 실시예에 나타낸 바와 같이, 쿠페아 팔루스트리스 C:8 선호 티오에스테라제로 형질 전환된 프로토테카는 12% 초과인 C8 지방산 수준을 달성할 수 있었을 뿐만 아니라, C8:C10 지방산의 비율은 약 5:1이었다. 지방산 수준의 변화는, 다양한 상업적 용도로서 맞춤 지방산 프로필을 가지는 오일을 생산하는데 유용하다. 뿐만 아니라, 상이한 지방산 사슬 길이간 비율의 변화는 추가의 고비용 화학 공정(예를 들어, 에스테르화, 증류, 분별 및 재에스테르화)을 수행하지 않은 오일에서는 상업상 얻어질 수 없는 것이다. 다른 예로서, 팜 오일은 C16:0 지방산을 가장 높이(32% 내지 47%)함유하는 오일이지만, 팜 오일은 C14:0 지방산의 함량이 매우 낮다. 유.아메리카나 티오에스테라제를 함유하는 프로토테카는 C16:0 지방산 함유율이 약 33% 내지 38%, C14:0 지방산 함유율이 약 10% 내지 16%인 것으로 획득하였다(C16:0 대 C14:0 비율은 약 2:1). 16:0 지방산을 높은 비율로 포함하는 종자 오일은 일반적으로 14:0 지방산은 그다지 많이 함유하지 않으므로, 기존의 오일을 상업상 수준으로 배합하였을 때 이러한 지방산 프로필은 상업상 비실용적이었다.

[0267] 이하 실시예는 또한 하나의 클론에서 2개 이상의 지방 아실-ACP 티오에스테라제의 성공적인 표적화 및 발현을 기술한다. 지방산 프로필 변화는 이러한은 클론에서 확인되었고, 2개의 티오에스테라제가 하나의 클론에서 공동 발현되었는지 여부에 의존적이었으며, 지방산 프로필은 상이한 방식으로 영향을 받았다. 예를 들어, 상기 표 5에 의하면, 코코넛 오일과 팜핵 오일 둘 다는 C12:C14 비율이 대략 3:1이다. 이하 실시예에 기술된 바와 같이, 2개의 이종 티오에스테라제 유전자를 함유하는 프로토테카 형질 전환체는 C12:C14 지방산 수준의 비율은 대략 5:1로 생성할 수 있었다. 이러한 종류의 C12:C14 지방산 비율은 (즉, 종자 오일을 배합할 때) 상업상 비실용적

이었다.

[0268] 유전자 이식 미세 조류에 의해 생산된 오일의 다른 신규 양태는 지방산의 포화도이다. 팜 오일은 현재 포화 오일의 가장 큰 공급원으로서 총 포화 오일 대 불포화 오일의 비율은 52% 내지 48%이다. 이하 실시예에 나타낸 바와 같이, 유.아메리카나 및 씨.캠포라로부터 유래하는 이종 티오에스테라제를 가지는 프로토테카는 그것이 생산하는 오일 중 포화 오일의 총 수준을 60% 초과로 획득하였다. 또한 이하 실시예에 나타낸 바와 같이, 유.아메리카나로부터 유래하는 이종 티오에스테라제를 가지는 프로토테카는 그것이 생산하는 오일 중 총 포화 오일 수준을 86% 초과로 획득하였다.

[0269] 본 발명의 미생물 및 방법과 함께 사용하기에 적합한 지방 아실-CoA/알데하이드 환원 효소는 제한없이 표 6에 열거된 것들을 포함한다.

표 6

GenBank 수탁 번호로 열거된 지방 아실-CoA/알데하이드 리덕타제
AAC45217, YP_047869, BAB85476, YP_001086217, YP_580344, YP_001280274, YP_264583, YP_436109, YP_959769, ZP_01736962, ZP_01900335, ZP_01892096, ZP_01103974, ZP_01915077, YP_924106, YP_130411, ZP_01222731, YP_550815, YP_983712, YP_001019688, YP_524762, YP_856798, ZP_01115500, YP_001141848, NP_336047, NP_216059, YP_882409, YP_706156, YP_001136150, YP_952365, ZP_01221833, YP_130076, NP_567936, AAR88762, ABK28586, NP_197634, CAD30694, NP_001063962, BAD46254, NP_001030809, EAZ10132, EAZ43639, EAZ07989, NP_001062488, CAB88537, NP_001052541, CAH66597, CAE02214, CAH66590, CAB88538, EAZ39844, AAZ06658, CAA68190, CAA52019, 및 BAC84377

[0270]

[0271] 본 발명의 미생물 및 방법과 함께 사용하기에 적합한 지방 아실-CoA 환원 효소는 제한 없이 표 7에 기재된 것들을 포함한다.

표 7

GenBank 수탁 번호로 열거된 지방 아실-CoA 환원 효소
NP_187805, ABO14927, NP_001049083, CAN83375, NP_191229, EAZ42242, EAZ06453, CAD30696, BAD31814, NP_190040, AAD38039, CAD30692, CAN81280, NP_197642, NP_190041, AAL15288, 및 NP_190042

[0272]

[0273] 본 발명의 미생물 및 방법과 함께 사용하기에 적합한 지방 알데하이드 탈카보닐화 효소는 제한 없이 표 8에 열거된 것들을 포함한다.

표 8

GenBank 수탁 번호로 열거된 지방 알데하이드 탈카보닐화 효소
NP_850932, ABN07985, CAN60676, AAC23640, CAA65199, AAC24373, CAE03390, ABD28319, NP_181306, EAZ31322, CAN63491, EAY94825, EAY86731, CAL55686, XP_001420263, EAZ23849, NP_200588, NP_001063227, CAN83072, AAR90847, 및 AAR97643

[0274]

[0275] 자연적으로 공동 발현된 지방 아실-ACP 티오에스테라제 및 아실기 운반 단백질의 조합은 본 발명의 미생물 및 방법과 사용하기에 적합하다.

[0276] 탄화수소 또는 지질 변형 효소의 추가의 예는 미국 특허 제6,610,527호; 제6,451,576호; 제6,429,014호; 제6,342,380호; 제6,265,639호; 제6,194,185호; 제6,114,160호; 제6,083,731호; 제6,043,072호; 제5,994,114호; 제5,891,697호; 제5,871,988호; 제6,265,639호 중 어느 하나에 포함되거나 참조되거나, 함유되거나 참조된 핵산 서열에 의해 암호화된 아미노산 서열을 포함하고, 추가로 GenBank 수탁 번호 제AAO18435호; 제ZP_00513891호; 제Q38710호; 제AAK60613호; 제AAK60610호; 제AAK60611호; 제NP_113747호; 제CAB75874호; 제AAK60612호; 제AAF20201호; 제BAA11024호; 제AF205791호; 및 제CAA03710호에 기술되어 있다.

[0277] 지질 생합성 경로의 기타 다른 효소들 또한 본 발명의 미생물 및 방법과 함께 사용하기 적당하다. 예를 들어, 케토 아실-ACP 합성 효소(Kas)는 지질 생합성 경로에 있어서 상기 나열된 효소들 중 일부와 함께 작용한다. Kas 효소에는 상이한 군들이 있다: Kas I은 지속적으로 성장하는(ever-growing) 아실 ACP 사슬과 말로닐-ACP 간 연속 축합 단계에 참여한다. Kas II는 통상적으로 C16:0-ACP로부터 C18:0-ACP(예를 들어, 말로닐-ACP)에 이르게 하는 최종 축합 단계에 참여한다. 그러므로, C16 내지 C18:0 지방산(그리고 이의 불포화 유도체)을 주로 합성하는 고등 식물 및 일부 미세 조류 종/균주에 있어서, Kas II 효소는 FatA 유전자(아실-ACP 티오에스테라제)의 생성물과 상호 작용한다.

[0278] 아실-ACP는 지방산 생합성 동안 성장중인 지방산 사슬을 ACP로부터 유리시키며, 대부분의 식물 종에 있어서 이는 C16:0에서 C18:0 단계로의 연장을 종결시키는 역할을 하는 FatA 유전자 군의 일원에 의해 수행된다. 길이가 더욱 단쇄의 지방산을 합성하는 종(예를 들어, 쿠페아, 엘라에이스, 미리스티카 또는 움벨루라리아)에서, FatB 유전자에 의해 암호화되는 아실-ACP 티오에스테라제의 상이한 군은 상기 종결 단계를 수행한다. Kas II 효소와 아실-ACP 티오에스테라제 간 상호 작용은 지방산 사슬 연장을 올바르게 종결하는데 중요하다. 결과적으로, 길이가 더욱 단쇄의 지질을 생합성할 수 있는 진화 FatB 유전자를 가지는 고등 식물 종(그리고 미세 조류 종)에 있어서, Kas IV 유전자라고 칭하는 Kas 유전자의 부가 군의 상응하는 공동 진화가 있었다. Kas IV 유전자는 길이가 4개 내지 14개 탄소인 특정한 크기 범위의 지방산의 사슬 길이 연장에 관여한다.

[0279] 본 발명의 미생물 및 방법과 사용하기에 적합한 기타 다른 효소는 표 4, 표 6 내지 표 8에 열거된 단백질 중 하나와 70% 이상의 아미노산 동일성을 갖고 상응하는 원하는 효소 활성(예를 들어, 아실기 운반 단백질로부터 지방산의 절단, 아실-CoA의 알데하이드 또는 알코올로의 환원, 또는 알데하이드의 알칸으로의 전환)을 나타내는 것들을 포함한다. 추가의 구체예에서, 효소 활성은 전술된 서열 중 하나와 약 75% 이상, 약 80% 이상, 약 85% 이상, 약 90% 이상, 약 95% 이상, 또는 약 99% 이상의 동일성을 갖는 서열 중에 존재하고 이 모두는 완전히 제시된 것처럼 본원에 참조로 포함된다.

[0280] 발현되는 외인성 유전자의 원하는 조합을 선택함으로써, 이후에 수성 바이오매스로부터 추출될 수 있는, 미생물에 의해 생성되는 생성물을 조작할 수 있다. 예를 들어, 상기 미생물은 (i) 지방 아실-ACP 티오에스테라제를 암호화하는 외인성 유전자; 및, 임의로, (ii) 자연적으로 공동 발현된 아실기 운반 단백질 또는 다르게는 지방 아실-ACP 티오에스테라제(또는 역으로)에 대해 친화성을 갖는 아실기 운반 단백질; 및, 임의로, (iii) 지방 아실-CoA/알데하이드 환원 효소 또는 지방 아실-CoA 환원 효소를 암호화하는 외인성 유전자; 및, 임의로, (iv) 지방 알데하이드 환원 효소 또는 지방 알데하이드 탈카보닐화 효소를 암호화하는 외인성 유전자 중 하나 이상을 함유할 수 있다. 미생물은 또한 외인성 스테아로일 ACP 불포화 효소, 지방산 불포화 효소, β-케토아실-ACP 합성 효소 I(예를 들어, KASI 유전자에 의해 암호화됨), β-케토아실-ACP 합성 효소 II(예를 들어, KASII 유전자에 의해 암호화됨), 또는 올리에이트-12 하이드록실라제 중 하나 이상을 함유할 수도 있다. 본원에 기재된 배양 조건 하의 미생물은 ACP에 연결된 지방산을 합성하고, 지방 아실-ACP 티오에스테라제는 ACP로부터 지방산의 절단을 촉매하여 추가의 효소 가공을 통해 지방 아실-CoA 분자를 생산한다. 존재하는 경우, 지방 아실-CoA/알데하이드 환원 효소는 아실-CoA의 알코올로의 환원을 촉매한다. 유사하게, 존재하는 경우 지방 아실-CoA 환원 효소는 아실-CoA의 알데하이드로의 환원을 촉매한다. 지방 아실-CoA 환원 효소를 암호화하는 외인성 유전자가 존재 하고 알데하이드 생성물을 생산하기 위해 발현되는 상기 구체예에서, 제3 외인성 유전자에 의해 암호화된 지방 알데하이드 환원 효소는 알데하이드의 알코올로의 환원을 촉매한다. 유사하게, 지방 알데하이드 탈카보닐화 효소는 존재하는 경우 알데하이드의 알칸 또는 알켄으로의 전환을 촉매한다.

[0281] 다른 구체예에서, 미생물은, (i) 지방 아실-ACP 티오에스테라제를 암호화하는 외인성 유전자; (ii) 임의로, 지방산 아실-ACP 티오에스테라제에 대한 친화성을 가지는 아실기 운반 단백질 또는 자연적으로 공동 발현된 아실기 운반 단백질; (iii) 돌연 변이된 내인성 불포화 효소 유전자(여기서, 상기 돌연 변이는 불포화 효소 유전자 또는 불포화 효소 단백질을 불활성이 되도록 만듦(예를 들어, 상기 불포화 효소 유전자 또는 단백질을 불포화 효소 녹아웃 요소 또는 불포화 효소 억제 요소, 예를 들어 표적화된 RNAi, 안티센스 또는 dsRNA 구조물로 만듦)); (iv) 내인성 스테아로일 아실기 운반 단백질 불포화 효소의 과발현 또는 이종 SAD의 발현; 및 (v) 상기

(i) 내지 (iv)의 임의의 조합을 포함할 수 있다.

[0282] 상기 효소, 예를 들어 지방 아실 ACP 티오에스테라제를 암호화하는 유전자는 클로렐라 프로토테코이데스와 같이 상당한 지질 생산을 나타내는 것으로 이미 공지된 세포로부터 수득될 수 있다. 지질 생산에 역할을 하는 것으로 이미 공지된 유전자, 예를 들어 이중 결합을 포화시키는 효소를 암호화하는 유전자는 개별적으로 수용 세포로 형질 전환될 수 있다. 그러나, 본 발명을 실시하기 위해 어느 유전자가 필요한지에 관해 우선적으로 추측할 필요는 없다. 미세 조류 내 지질 생산을 변형(개선)시킬 수 있는 유전자를 확인하는 방법은 PCT 공개 제 2008/151149호에 기술되어 있다.

[0283] 따라서, 본 발명은 동일 종의 야생형 세포에 비해 변화된 수준으로 지질 경로 효소를 발현하도록 유전자 조작된 프로토테카 세포를 제공한다. 몇몇 경우에, 상기 세포는 2개 세포가 동일한 조건 하에서 생장하는 경우 야생형 세포에 비해 더 많은 지질을 생산한다. 몇몇 경우에, 상기 세포는 야생형 세포보다 더 높은 수준 또는 더 낮은 수준으로 지질 경로 효소를 발현하도록 유전자 조작되고/조작되거나 선택되었다. 몇몇 경우에, 지질 경로 효소는 피루베이트 탈수소화 효소, 아세틸-CoA 카복실라제, 아실기 운반 단백질 및 글리세롤-3 포스페이트 아실트랜스퍼라제로 이루어진 군으로부터 선택된다. 몇몇 경우에, 상기 세포는 야생형 세포보다 더 낮은 수준으로 지질 경로 효소를 발현하도록 유전자 조작되고/조작되거나 선택되었다. 세포가 더 낮은 수준으로 지질 경로 효소를 발현하는 하나 이상의 구체예에서, 지질 경로 효소는 시트레이트 합성 효소를 포함한다.

[0284] 몇몇 구체예에서, 상기 세포는 야생형 세포와 비교하여 변화된 수준으로 지방산 합성의 범용 조절제를 발현하도록 유전자 조작되고/조작되거나 선택됨으로써 다수의 지방산 합성 유전자의 발현 수준은 야생형 세포와 비교하여 변화된다. 몇몇 경우에, 지질 경로 효소는 지방산을 변형시키는 효소를 포함한다. 몇몇 경우에, 지질 경로 효소는 스테아로일-ACP 불포화 효소 및 글리세로지질 불포화 효소로부터 선택된다. 몇몇 경우에 있어서, 세포는 지질 경로 효소를 더 낮은 수준으로 발현하거나 특이적 지질 경로 효소를 전혀 발현하지 않도록 유전자 조작되고/조작되거나 선택되었다(즉, 지질 경로 효소가 녹아웃되어 외인성 유전자로 치환되거나, RNAi 또는 안티센스 방법을 이용하여 이 효소의 발현이 감소되었다). 다른 구체예에서, 지질 경로 효소는 불포화 효소 유전자, 예를 들어 스테아로일-ACP 불포화 효소 또는 지방산 불포화 효소(FAD) 유전자(이에 제한되는 것은 아님)의 이종 발현물이다. 실시예 6은 프로토테카 모리포르미스 유전적 배경에서 올레아 유로파에아로부터 이종 스테아로일-ACP의 발현을 기술한다.

[0285] 다른 구체예에서, 본 발명은 하나 이상의 외인성 유전자를 함유하는 오일 생산 미생물에 관한 것으로서, 여기서 상기 외인성 유전자는 지방 아실-ACP 티오에스테라제, 지방 아실-CoA 환원 효소, 지방 알데하이드 환원 효소, 지방 아실-CoA/알데하이드 환원 효소, 지방 알데하이드 탈카보닐화 효소, 불포화 효소 및 아실기 운반 단백질로 이루어진 군으로부터 선택되는 단백질(들)을 암호화한다. 다른 구체예에서, 내인성 불포화 효소 유전자는 상기 외인성 유전자들 중 하나 이상을 함유하는 미생물 내에서 과발현된다. 하나의 구체예에서, 외인성 유전자는 자극에 응답하여 유도성이거나 억제성인 프로모터에 작동가능하게 연결되어 있다. 몇몇 경우에, 상기 자극은 외부적으로 제공되는 작은 분자, 열, 냉 및 배양 배지에서 제한되거나 부재인 질소로 이루어진 군으로부터 선택된다. 몇몇 경우에, 외인성 유전자는 세포 구획에서 발현된다. 몇몇 구체예에서, 세포 구획은 엽록체, 색소체 및 미토콘드리아로 이루어진 군으로부터 선택된다. 몇몇 구체예에서 미생물은 프로토테카 모리포르미스, 프로토테카 크루가니, 프로토테카 스태그노라 또는 프로토테카 조프피이다.

[0286] 하나의 구체예에서, 외인성 유전자는 지방산 아실-ACP 티오에스테라제를 암호화한다. 몇몇 경우에, 외인성 유전자에 의해 암호화된 티오에스테라제는 아실기 운반 단백질(ACP)로부터 8개 내지 18개의 탄소 지방산의 절단을 촉매한다. 몇몇 경우에, 외인성 유전자에 의해 암호화된 티오에스테라제는 ACP로부터 10개 내지 14개 탄소 지방산의 절단을 촉매한다. 하나의 구체예에서, 외인성 유전자에 의해 암호화된 티오에스테라제는 ACP로부터 12개 탄소 지방산의 절단을 촉매한다.

[0287] 하나의 구체예에서, 외인성 유전자는 지방 아실-CoA/알데하이드 환원 효소를 암호화한다. 몇몇 경우에, 외인성 유전자에 의해 암호화된 환원 효소는 8개 내지 18개의 탄소 지방 아실-CoA의 상응하는 1급 알코올로의 환원을 촉매한다. 몇몇 경우에, 외인성 유전자에 의해 암호화된 환원 효소는 10개 내지 14개의 탄소 지방 아실-CoA의 상응하는 1급 알코올로의 환원을 촉매한다. 하나의 구체예에서, 외인성 유전자에 의해 암호화된 환원 효소는 12개 탄소의 지방 아실-CoA의 도데칸올로의 환원을 촉매한다.

[0288] 본 발명은 또한 2개의 외인성 유전자를 함유하는 재조합 프로토테카 또는 기타 다른 세포를 제공하고, 여기서 제1 외인성 유전자는 지방 아실-ACP 티오에스테라제를 암호화하고 제2 외인성 유전자는 지방 아실-CoA 환원 효소, 지방 아실-CoA/알데하이드 환원 효소, 및 아실기 운반 단백질로 이루어진 군으로부터 선택되는 단백질을 암

호화한다. 몇몇 경우에, 2개의 외인성 유전자는 각각 자극에 응답하여 유도성인 프로모터와 작동가능하게 연결되어 있다. 몇몇 경우에, 각각의 프로모터는 배양 배지 내 제한된 질소 또는 질소 부재와 같은 동일한 자극에 응답하여 유도성이다. 배양 배지에서 질소의 제한 또는 완전한 부재는 프로토테카 종과 같은 몇몇 미생물에서 오일 생산을 자극하고, 높은 수준으로 오일 생산을 유도하기 위한 촉발제로서 사용될 수 있다. 본원에 개시된 유전자 조작 방법과 조합하여 사용되는 경우, 건조 세포 중량의 %로서 지질은 30% 이상, 40% 이상, 50% 이상, 60% 이상, 70% 이상 및 75% 이상, 80% 이상, 85% 이상 또는 75% 내지 90%와 같은 높은 수준으로 상승될 수 있고, 본원에 개시된 방법은 이러한 수준의 지질을 갖는 세포에 대해 제공하고, 여기서 지질은 C8 내지 C14 4% 이상, C8 0.3% 이상, C10 2% 이상, C12 2% 이상, 및 C14 2% 이상이다. 몇몇 구체예에서, 상기 세포는 건조 세포 중량을 기준으로 지질이 25% 초과이고, C8 내지 C14 10% 이상, C8 내지 C14 20% 이상, C8 내지 C14 30% 이상, C8 내지 C14 10% 내지 30% 및 C8 내지 C14 20% 내지 30%인 지질을 함유한다.

[0289] 본원에 개시된 신규 오일은 mid-사슬 지방산(예를 들어, 팜오일, 팜핵 오일 및 코코넛 오일)의 함량이 높은 기타 다른 자연적으로 존재하는 오일과 구분된다. 예를 들어, 카로티노이드와 같은 오염물 수준은 본 발명의 오일에서 보다 팜오일 및 팜핵 오일에서 훨씬 더 높다. 특히 팜오일 및 팜핵 오일은 본 발명의 오일에서 훨씬 더 높은 양으로 알파 및 베타 카로틴 및 라이코펜을 함유한다. 추가로, 20개 초과의 상이한 카로티노이드가 팜오일 및 팜핵 오일에서 발견된 반면 실시예는 본 발명의 오일이 매우 소량의 카로티노이드 종을 매우 낮은 수준으로 함유함을 입증한다. 추가로, 토코트리에놀과 같은 비타민 E 화합물의 수준은 본 발명의 오일에서 보다 팜오일, 팜핵 및 코코넛오일에서 훨씬 더 높다.

[0290] 하나의 구체예에서, 제1 외인성 유전자에 의해 암호화된 티오에스테라제는 ACP로부터 8개 내지 18개 탄소 지방산의 절단을 촉매한다. 몇몇 구체예에서, 제2 외인성 유전자는 8개 내지 18개 탄소 지방 아실-CoA의 상응하는 1급 알코올로의 환원을 촉매하는 지방 아실-CoA/알데하이드 환원 효소를 암호화한다. 몇몇 경우에, 제1 외인성 유전자에 의해 암호화된 티오에스테라제는 ACP로부터 10개 내지 14개 탄소 지방산의 절단을 촉매하고, 제2 외인성 유전자에 의해 암호화된 환원 효소는 10개 내지 14개 탄소 지방 아실-CoA의 상응하는 1급 알코올로의 환원을 촉매하며, 여기서 티오에스테라제 및 환원 효소는 동일한 탄소 사슬 길이에 작용한다. 하나의 구체예에서, 제1 외인성 유전자에 의해 암호화된 티오에스테라제는 ACP로부터 12개 탄소 지방산의 절단을 촉매하고, 제2 외인성 유전자에 의해 암호화된 환원 효소는 12개 탄소 지방 아실-CoA의 도데칸올로의 환원을 촉매한다. 몇몇 구체예에서, 제2 외인성 유전자는 8개 내지 18개 탄소 지방 아실-CoA의 상응하는 알데하이드로의 환원을 촉매하는 지방 아실-CoA 환원 효소를 암호화한다. 몇몇 구체예에서, 제2 외인성 유전자는 자연적으로 지방 아실-ACP 티오에스테라제와 함께 공동 발현되는 아실기 운반 단백질을 암호화한다.

[0291] 몇몇 구체예에서, 제2 외인성 유전자는 지방 아실-CoA 환원 효소를 암호화하고, 미생물은 지방 알데하이드 탈카보닐화 효소를 암호화하는 제3 외인성 유전자를 추가로 함유한다. 몇몇 경우에, 제1 외인성 유전자에 의해 암호화된 티오에스테라제는 ACP로부터 8개 내지 18개 탄소 지방산의 절단을 촉매하고, 제2 외인성 유전자에 의해 암호화된 환원 효소는 8개 내지 18개 탄소 지방 아실-CoA의 상응하는 지방 알데하이드로의 환원을 촉매하며, 제3 외인성 유전자에 의해 암호화된 탈카보닐화 효소는 8개 내지 18개 탄소 지방 알데하이드의 상응하는 알칸으로의 전환을 촉매하고, 여기서 티오에스테라제, 환원 효소 및 탈카보닐화 효소는 동일한 탄소 사슬 길이에 작용한다.

[0292] 몇몇 구체예에서, 제2 외인성 유전자는 아실기 운반 단백질을 암호화하고, 미생물은 지방 아실-CoA 환원 효소 및 지방 아실-CoA/알데하이드 환원 효소로 이루어진 군으로부터 선택되는 단백질을 암호화하는 제3 외인성 유전자를 추가로 함유한다. 몇몇 경우에, 제3 외인성 유전자는 지방 아실-CoA 환원 효소를 암호화하고, 미생물은 지방 알데하이드 탈카보닐화 효소를 암호화하는 제4 외인성 유전자를 추가로 함유한다.

[0293] 본 발명은 또한 배양 배지에서 재조합 프로토테카 세포의 집단을 배양하는 것을 포함하는 알코올을 생산하는 방법을 제공하며, 여기서 상기 세포는 (i) 지방 아실-ACP 티오에스테라제를 암호화하는 제1 외인성 유전자, 및 (ii) 지방 아실-CoA/알데하이드 환원 효소를 암호화하는 제2 외인성 유전자를 함유하고, 상기 세포는 아실기 운반 단백질(ACP)에 연결된 지방산을 합성하며, 지방 아실-ACP 티오에스테라제는 ACP로부터 지방산의 절단을 촉매하여 추가의 가공을 통해 지방 아실-CoA를 생산하고, 지방 아실-CoA/알데하이드 환원 효소는 아실-CoA의 알코올로의 환원을 촉매한다.

[0294] 본 발명은 또한 프로토테카 세포에서 지질 분자를 생산하는 방법을 제공한다. 하나의 구체예에서, 상기 방법은 배양 배지에서 프로토테카 세포 집단을 배양하는 것을 포함하며, 상기 세포는 (i) 지방 아실-ACP 티오에스테라제를 암호화하는 제1 외인성 유전자, 및 (ii) 지방 아실-CoA 환원 효소를 암호화하는 제2 외인성 유전자를 함유하고, 상기 미생물은 아실기 운반 단백질(ACP)에 연결된 지방산을 합성하며, 지방 아실-ACP 티오에스테라제는

ACP로부터 지방산의 절단을 촉매하여 추가의 가공을 통해 지방 아실-CoA를 생산하고, 지방 아실-CoA 환원 효소는 아실-CoA의 알데하이드로의 환원을 촉매한다.

[0295] 본 발명은 또한 프로토테카 세포에서 특정된 탄소 사슬 길이를 갖는 지방산 분자를 생산하는 방법을 제공한다. 하나의 구체예에서, 상기 방법은 배양 배지에서 지질 생산 프로토테카 세포 집단을 배양하는 것을 포함하며, 여기서 상기 미생물은 8개, 10개, 12개 또는 14개 탄소 원자와 같은 특정 탄소 사슬 길이에 특이적이거나 우선적인 활성을 갖는 지방 아실-ACP 티오에스테라제를 암호화하는 외인성 유전자를 함유하고, 상기 미생물은 아실기 운반 단백질(ACP)에 연결된 지방산을 합성하며, 상기 티오에스테라제는 지방산이 특정 탄소 사슬 길이로 합성되는 경우 ACP로부터 지방산의 절단을 촉매한다.

[0296] 전술한 다양한 구체예에서, 프로토테카 세포는 지질 경로 효소를 암호화하는 하나 이상의 외인성 유전자 또는 유전자 생성물의 발현을 억제하는 억제 요소, 예를 들어 RNA 간섭 요소를 함유할 수 있다. 몇몇 경우에, 지질 경로 효소는 스테아로일-ACP 불포화 효소, 글리세로지질 불포화 효소, 피루베이트 탈수소화 효소, 아세틸-CoA 카복실라제, 아실기 운반 단백질 및 글리세롤-3 포스페이트 아실트랜스퍼라제로 이루어진 군으로부터 선택된다. 다른 경우에, 프로토테카 세포는 지방 아실-ACP 티오에스테라제, 지방 아실-CoA/알데하이드 환원 효소, 지방 아실-CoA 환원 효소, 지방 알데하이드 환원 효소, 지방 알데하이드 탈카보닐화 효소 및/또는 아실기 운반 단백질로 이루어진 군으로부터 선택되는 지질 변형 효소를 함유한다.

[0297] 본 발명은 또한 하이드록실화된 지방산을 생성하는 하이드록실라제를 암호화하는 이종 유전자를 함유하는 미생물 세포를 제공한다. 미생물 세포는 제II형 지방산 합성 경로를 포함할 수 있다. 예를 들어, 미생물 세포는 미세 조류 세포일 수 있다. 몇몇 구체예에서, 미세 조류 세포는 상기 표 1에 나열된 미세 조류 세포들로부터 선택된다. 다른 구체예에서, 미세 조류 세포는 프로토테카 속에 속하는 것이다. 또 다른 구체예에서, 미세 조류 세포는 프로토테카 모리포르미스이다. 하이드록실라제는 하이드록실기(-OH)를 기질에 부가하는 효소이다. 지방산 하이드록실라제는 일부 고등 식물에서 발견되는 자연적으로 생성되는 효소이다. 고등 식물에서 발견되는 자연적으로 생성되는 하이드록실라제의 비제한적 예로서는 리시놀레산을 생산하는데 관여하는 리시누스 커뮤니스로부터 유래하는 올리에이트 12-하이드록실라제가 있다. 실시예 7은 프로토테카 세포 내에서 하이드록실라제의 이종 발현의 예, 특히 프로토테카 모리포르미스 세포 내에서 리시누스 커뮤니스 올리에이트 12-하이드록실라제의 발현을 기술한다.

[0299] **VI. 연료 및 화학 물질 생산**

[0300] 본 발명의 방법에 따른 연료의 생산을 위해, 본 발명의 세포에 의해 생산된 지질을 임의의 편리한 수단에 의해 수거하거나, 다르게는 수집한다. 지질은 전체 세포 추출에 의해 분리될 수 있다. 상기 세포를 우선 파쇄시키고 이어서 세포 내 및 세포막/세포벽 결합 지질 및 세포외 탄화수소를, 예를 들어 전술된 바와 같은 원심분리를 사용하여 세포 덩어리로부터 분리할 수 있다. 미생물내에서 생성된 세포 내 지질은 몇몇 구체예에서 미생물의 세포를 용해시킨 후 추출된다. 일단 추출된 후, 지질을 추가로 정제하여 오일, 연료 또는 함유 화학 제품을 생성한다.

[0301] 배양을 완료한 후, 미생물을 발효 브로스로부터 분리할 수 있다. 임의로, 상기 분리는 원심분리로 수행하여 농축된 페이스트를 생성한다. 원심분리는 미생물로부터 상당양의 세포 내 물을 제거하지 못하며, 건조 단계가 아니다. 그 다음, 상기 바이오매스는 임의로 세척액(예를 들어, 탈이온수)으로 세척하여 발효 브로스 및 잔해를 제거할 수 있다. 임의로, 세척된 미생물 바이오매스는 또한 세포 파쇄 전에 건조(오븐 건조, 동결 건조 등)될 수 있다. 대안적으로, 세포는 발효가 완료된 경우 발효 브로스의 일부 또는 전부로부터 분리 없이 용해될 수 있다. 예를 들어, 상기 세포는 세포가 용해된 경우 세포외 액체에 대한 세포의 비율이 1:1(v:v) 미만일 수 있다.

[0302] 지질을 함유하는 미생물은 용해시켜 용해물을 생성할 수 있다. 본원에 상세히 기술된 바와 같이, 미생물 용해시키는 단계(또한 세포 용해로서 언급됨)는 열 유도된 용해, 염기의 첨가, 산의 첨가, 프로테아제와 같은 효소 및 아밀라제와 같은 다당류 분해 효소의 사용, 초음파의 사용, 기계적 용해, 삼투압 충격의 사용, 용균성 바이러스에 의한 감염 및/또는 하나 이상의 용균성 유전자의 발현을 포함하는 임의의 간편한 수단에 의해 성취될 수 있다. 용해를 수행하여 미생물에 의해 생성된 세포 내 분자를 방출시킨다. 미생물을 용해시키기 위한 이들 방법 각각은 단일 방법으로 또는 동시에 또는 연속적으로 조합하여 사용할 수 있다. 세포 파쇄 정도는 현미경 분석에 의해 관찰할 수 있다. 본원에 기술된 방법중 하나 이상을 사용하여, 통상적으로 70% 초과의 세포 파괴가 관찰된다. 바람직하게, 세포 파괴는 80% 초과, 보다 바람직하게는 90% 초과 및 가장 바람직하게는 약 100%이다.

[0303] 특정 구체예에서, 미생물은, 예를 들어 추출 또는 추가의 가공을 위해 세포 지질 및/또는 탄화수소의 노출을 증

가시키기 위해 생장 후 용해시킨다. 리파제 발현(예를 들어, 유도성 프로모터를 통함) 및 세포 용해의 타이밍은 지질 및/또는 탄화수소의 수율을 최적화하기 위해 조정될 수 있다. 다수의 용해 기술은 하기에 기술되어 있다. 이들 기술은 개별적으로 또는 조합하여 사용할 수 있다.

[0304] 본 발명의 하나의 구체예에서, 미생물을 용해시키는 단계는 미생물을 함유하는 세포 현탁액의 가열을 포함한다. 이러한 구체예에서, 미생물을 함유하는 발효 브로스(또는 발효 브로스로부터 분리된 미생물 현탁액)는 미생물, 즉 미생물의 세포벽 및 막이 분해하거나 파괴될 때까지 가열한다. 통상적으로, 적용되는 온도는 50℃ 이상이다. 보다 효율적인 세포 용해를 위해, 보다 고온, 예를 들어 30℃ 이상, 60℃ 이상, 70℃ 이상, 80℃ 이상, 90℃ 이상, 100℃ 이상, 110℃ 이상, 120℃ 이상, 130℃ 이상 또는 그 이상의 온도가 사용된다. 열 처리에 의한 세포 용해는 미생물을 비등시킴으로써 수행할 수 있다. 대안적으로, 열 처리(비등 없이)는 오토클레이브에서 수행할 수 있다. 열 처리된 용해물은 추가의 처리를 위해 냉각시킬 수 있다. 세포 파쇄는 또한 증기 처리, 즉 가압된 증기의 부가를 통해 수행할 수 있다. 세포 파쇄를 위한 미세 조류의 증기 처리는, 예를 들어 미국 특허 제6,750,048호에 기술되어 있다. 몇몇 구체예에서, 증기 처리는 발효기로 증기를 살포하고 브로스를 약 90분 미만, 바람직하게는 약 60분 미만, 보다 바람직하게는 약 30분 미만 동안 원하는 온도에서 유지시킴으로써 성취할 수 있다.

[0305] 본 발명의 다른 구체예에서, 미생물을 용해시키는 단계는 염기를, 미생물을 함유하는 세포 현탁액에 첨가하는 것을 포함한다. 상기 염기는 사용되는 미생물의 단백질성 화합물의 적어도 일부를 가수 분해하기에 충분히 강해야만 한다. 단백질을 가용화시키는데 유용한 염기는 화학 기술 분야에 공지되어 있다. 본 발명의 방법에 유용한 예시적인 염기로서는 리튬, 나트륨, 칼륨, 칼슘의 수산화물, 탄산염 및 중탄산염, 및 이들의 혼합물을 포함하지만 이에 제한되지 않는다. 바람직한 염기는 KOH이다. 세포 파쇄를 위한 미세 조류의 염기 처리는, 예를 들어 미국 특허 제6,750,048호에 기술되어 있다.

[0306] 본 발명의 다른 구체예에서, 미생물을 용해시키는 단계는 산을, 미생물을 함유하는 세포 현탁액에 첨가하는 것을 포함한다. 산 용해는 10mN 내지 500mN 또는 바람직하게 40nM 내지 160nM의 농도로 산을 사용하여 수행할 수 있다. 산 용해는 바람직하게 실온 초과(예를 들어, 40℃ 내지 160℃, 및 바람직하게 50℃ 내지 130℃의 온도)에서 수행한다. 적당한 온도(예를 들어, 실온 내지 100℃ 및 특히 실온 내지 65℃)에 대해서, 산 처리는 초음파 처리 또는 기타 다른 세포 파쇄 방법과 유용하게 조합할 수 있다.

[0307] 본 발명의 다른 구체예에서, 미생물을 용해시키는 단계는 효소를 사용함으로써 미생물을 용해시키는 것을 포함한다. 미생물을 용해시키기 위해 바람직한 효소로서는 프로테아제 및 다당류 분해 효소, 예를 들어 헤미셀룰라제(예를 들어, 아스퍼질러스 나이저; Sigma Aldrich, 미국 미주리주 세인트루이스 소재; #H2125), 펙티나제(예를 들어, 리조푸스 종 유래의 펙티나제; Sigma Aldrich, 미국 미주리주 세인트루이스 소재; #P2401), 만나웨이(Mannaway) 4.0 L(Novozymes), 셀룰라제(예를 들어, 트리코더마 비리데 유래의 셀룰라제; Sigma Aldrich, 미국 미주리주 세인트루이스 소재; #C9422), 및 드리셀라제(예를 들어, 바시디오마이세테스 종 유래의 드리셀라제; Sigma Aldrich, 미국 미주리주 세인트루이스 소재; #D9515)가 있다.

[0308] 본 발명의 다른 구체예에서, 용해는, 예를 들어 임의로 클로렐라 또는 클로렐라 바이러스 유래의 다당류 분해 효소와 같은 셀룰라제, 또는 스트렙토마이세스 그리세우스(Streptomyces griseus) 프로테아제, 키모트립신, 프로테이나제 K, 문헌[Degradation of Polylactide by Commercial Proteases, Oda Yet al., Journal of Polymers and the Environment, Volume 8, Number 1, January 2000, pp. 29-32(4)]에 열거된 프로테아제, 알칼라제 2.4 FG(Novozymes), 및 플라보우르짐 100 L(Novozymes)과 같은 프로테아제와 같은 효소를 사용하여 성취한다. 또한 이전의 프로테아제 및 다당류 분해 효소의 임의의 조합을 비롯한 프로테아제 및 다당류 분해 효소의 임의의 조합을 사용할 수 있다.

[0309] 다른 구체예에서, 용해는 익스펠러 프레스를 사용하여 수행할 수 있다. 이러한 공정에서, 바이오매스는 고압에서 스크류형 장치를 통해 강제 주입하고 세포를 용해시키며, 세포 내 지질이 방출되게 하고 세포 내에서 단백질 및 섬유(및 기타 다른 성분)로부터 분리되도록 한다.

[0310] 본 발명의 다른 구체예에서, 미생물을 용해시키는 단계는 초음파, 즉 초음파 처리를 사용하여 수행한다. 따라서, 세포는 또한 고주파음으로 용해시킬 수 있다. 상기 음은 전자적으로 생성될 수 있고 금속 팁을 통해 적절히 농축된 세포 현탁액으로 운반된다. 상기 초음파 처리(또는 초음파 처리(ultrasonication))는 세포 현탁액 중 공극의 형성을 기준으로 세포 완전성을 붕괴시킨다.

[0311] 본 발명의 다른 구체예에서, 미생물을 용해시키는 단계는 기계적 용해에 의해 수행한다. 세포는 기계적으로 용

해시키고, 임의로 균질화하여 탄화수소(예를 들어, 지질) 수집을 촉진시킬 수 있다. 예를 들어, 압력 분쇄기를 사용하여 세포 함유 슬러리를 제한된 오리피스 밸브를 통과하도록 펌핑시킬 수 있다. 고압(1500바 이하)을 적용함에 이어서 출구 노즐을 통해 즉시 확장시킨다. 세포 파쇄는 밸브상에 충돌, 오리피스 내 액체 고전단, 및 방출시 급작스런 압력 강하로 세포 파열의 유도의 3가지 상이한 기작에 의해 성취한다. 상기 방법은 세포 내 분자를 방출시킨다. 대안적으로, 볼 밀을 사용할 수 있다. 볼 밀에서, 세포는 비드와 같은 작은 연마 입자와 함께 현탁 상태로 교반시킨다. 전단력 때문에 세포는 분해되고, 비드 사이에서 분쇄되며 비드와 충돌한다. 상기 비드는 세포를 파쇄하여 세포 내용물을 방출시킨다. 세포는, 예를 들어 블렌딩(예를 들어, 고속 또는 와링 블렌더(Waring blender)), 프렌치 프레스 또는 심지어 약한 세포 벽의 경우에 세포 파쇄를 위한 원심분리의 사용에 의하는 것과 같은 전단력에 의해 파쇄될 수 있다.

[0312] 본 발명의 다른 구체예에서, 미생물을 용해시키는 단계는 삼투압 충격을 적용함으로써 수행한다.

[0313] 본 발명의 다른 구체예에서, 미생물을 용해시키는 단계는 용균성 바이러스로 미생물을 감염시키는 것을 포함한다. 본 발명에 사용하기 위해 적합한 미생물을 용해시키기 위한 광범위하게 다양한 바이러스가 공지되어 있고, 특정 미생물에 대한 특정 용균성 바이러스의 선택 및 용도는 당업계의 기술 수준 내에 있다. 예를 들어, 파라메시움 부르사리아 클로렐라 바이러스(paramecium bursaria chlorella virus; PBCV-1)는 특정 단일 세포 진핵 클로렐라형 녹색 조류 내에서 복제하고 이를 용해시키는 대형의 20면체 플라크 형성 이중가닥의 DNA 바이러스군의 프로토타입이다(피코드나비리대(Phycodnaviridae) 계열, 클로로바이러스(Chlorovirus) 속). 따라서, 임의의 민감성 미세 조류는 적합한 클로렐라 바이러스로 배양물을 감염시킴으로써 용해시킬 수 있다. 클로렐라 바이러스로 클로렐라 종을 감염시키는 방법은 공지되어 있다. 예를 들어, 문헌[Adv. Virus Res. 2006;66:293-336]; [Virology, 1999 Apr 25;257(1): 15-23]; [Virology, 2004 Jan 5;318(1):214-23]; [Nucleic Acids Symp. Ser. 2000;(44): 161-2]; [J. Virol. 2006 Mar;80(5):2437-44]; 및 [Annu. Rev. Microbiol. 1999;53:447-94]을 참조한다.

[0314] 본 발명의 다른 구체예에서, 미생물을 용해시키는 단계는 자가용해를 포함한다. 이러한 구체예에서, 본 발명에 따른 미생물은 미생물을 용해시킬 용균성 단백질을 생성하도록 유전자 조작된다. 이러한 용균성 유전자는 유도성 프로모터를 사용하여 발현시켜 세포가 처음에 발효기 내에서 바람직한 밀도로 생장한 후 프로모터를 유도하여 용균성 유전자를 발현시킴으로써 세포를 용해시킬 수 있게 할 수 있다. 하나의 구체예에서, 상기 용균성 유전자는 다당류 분해 효소를 암호화한다. 다른 특정 구체예에서, 상기 용균성 유전자는 용균성 바이러스 유래의 유전자이다. 따라서, 예를 들어 클로렐라 바이러스 유래의 용균성 유전자는 조류 세포에서 발현될 수 있고, 문헌[Virology 260, 308-315 (1999)]; [FEMS Microbiology Letters 180 (1999) 45-53]; [Virology 263, 376-387 (1999)]; 및 [Virology 230, 361-368 (1997)]을 참조한다. 용균성 유전자의 발현은 바람직하게 유도성 프로모터, 예를 들어 소분자, 광, 열의 존재와 같은 자극 및 기타 다른 자극에 의해 유도되는 미세 조류에서 활성인 프로모터를 사용하여 수행한다.

[0315] 상기 방법에 의해 생성된 세포 용해물로부터 지질을 분리하기 위한 다양한 방법이 이용가능하다. 예를 들어, 지방 알데하이드, 지방 알코올 및 탄화수소(예를 들어, 알칸)와 같은 지질 및 지질 유도체는 헥산과 같은 소수성 용매로 추출할 수 있다(문헌[Frenz et al. 1989, Enzyme Microb. Technol., 11:717] 참조). 지질 및 지질 유도체는 또한 액화(예를 들어, 문헌[Sawayama et al. 1999, Biomass and Bioenergy 17:33-39] 및 [Inoue et al. 1993, Biomass Bioenergy 6(4):269-274] 참조); 오일 액화(예를 들어, 문헌[Minowa et al. 1995, Fuel 74(12): 1735-1738] 참조); 및 초임계 CO2 추출(예를 들어, 문헌[Mendes et al. 2003, Inorganica Chimica Acta 356:328-334] 참조)을 사용하여 추출할 수 있다. Miao 및 Wu는 클로렐라 프로토테오코이데스의 배양으로부터 미세 조류 지질 회수의 프로토콜을 기술하며, 여기서 상기 세포는 원심분리에 의해 수거되고 증류수로 세척되며 동결 건조에 의해 건조된다. 생성된 세포 분말을 모르타르에서 분쇄하고 이어서 n-헥산으로 추출하였다(문헌[Miao 및 Wu, Biosource Technology (2006) 97:841-846]).

[0316] 따라서, 본 발명의 미생물에 의해 생성된 지질, 지질 유도체 및 탄화수소는 유기 용매로 추출하여 회수할 수 있다. 몇몇 경우에, 바람직한 유기 용매는 헥산이다. 통상적으로, 유기 용매는 용해 성분의 사전 분리 없이 용해물에 직접 첨가한다. 하나의 구체예에서, 전술한 방법 중 하나 이상의 방법에 의해 생성된 용해물은 지질 및/또는 탄화수소 성분이 유기 용매와 용액을 형성하도록 하기에 충분한 시간 동안 유기 용매와 접촉시킨다. 몇몇 경우에, 상기 용액을 이어서 추가로 정제하여 특정의 원하는 지질 또는 탄화수소 성분을 회수할 수 있다. 헥산 추출 방법은 당업계에 널리 공지되어 있다.

[0317] 본원에 기술된 세포에 의해 생성된 지방 알데하이드, 지방 알코올 및 탄화수소(예를 들어, 알칸)와 같은 지질

및 지질 유도체는 전술된 바와 같은 리파제를 포함하여 하나 이상의 효소를 사용함으로써 변형시킬 수 있다. 탄화수소가 세포의 세포외 환경에 존재하는 경우, 효소가 탄화수소를 변형시키거나 탄화수소 전구체로부터 이의 합성을 완성하도록 하는 조건 하에서 하나 이상의 효소를 상기 환경에 첨가할 수 있다. 대안적으로, 상기 탄화수소는 효소와 같은 하나 이상의 촉매의 첨가 전에 세포 물질로부터 부분적으로 또는 완전하게 분리될 수 있다. 상기 촉매는 외인성으로 첨가되고, 이들의 활성은 세포 외부에서 또는 시험관내에서 발생한다.

[0318] 따라서, 본원에 기술된 바와 같이 생체 내에서 세포에 의해 생성되거나 시험관 내에서 효소로 변형된 지질 및 탄화수소는 임의로 통상적인 수단에 의해 추가로 가공될 수 있다. 상기 가공은 탄화수소 분자의 크기를 감소시키고 따라서 이의 수소:탄소 비율을 증가시키기 위한 "크래킹(cracking)"을 포함할 수 있다. 촉매 및 열 크래킹 방법은 통상적으로 탄화수소 및 트리글리세라이드 오일 가공에 사용한다. 촉매 방법은 고체 산 촉매와 같은 촉매의 사용을 포함한다. 촉매는 실리카-알루미나 또는 제올라이트일 수 있고, 이로써 탄소-탄소 결합을 이종용해 또는 비대칭 분해시켜 카보양이온(carbocation) 및 수소화물 음이온을 생성할 수 있다. 이들 반응성 중간체를 이어서 다른 탄화수소와 함께 재배열하거나 수소화 전달을 진행시킨다. 따라서 상기 반응은 중간체를 재생성시켜 자가 증폭 사슬 기작을 유도할 수 있다. 또한 탄화수소는 여기서의 탄소-탄소 이중 결합 또는 삼중 결합의 수를 임의로 0까지 감소하도록 가공할 수 있다. 또한 탄화수소를 가공하여 여기서의 고리 또는 환형 구조를 제거할 수 있다. 탄화수소를 또한 가공하여 수소:탄소 비율을 증가시킬 수 있다. 이것은 수소의 첨가("수소화") 및/또는 탄화수소의 보다 작은 탄화수소로의 "크래킹"을 포함할 수 있다.

[0319] 열 방법은 탄화수소 크기를 감소시키기 위한 상승된 온도 및 압력의 사용을 포함한다. 약 800℃의 승온 및 약 700kPa의 압력을 사용할 수 있다. 이들 조건은 때때로 수소 풍부 탄화수소 분자(양성자 플럭스와는 구분됨)를 언급하는데 사용되는 용어 "가벼운" 탄화수소 분자를 생성하는 한편, 또한 응축에 의해 상대적으로 수소가 고갈된 보다 무거운 탄화수소 분자를 생성한다. 상기 방법은 균형 분해 또는 대칭적 분해를 제공하고, 전술된 바와 같이 임의로 효소로 포화될 수 있는 알켄을 생성한다.

[0320] 촉매 및 열 방법은 탄화수소 가공 및 정유를 위한 식물에서의 표준 방법이다. 따라서, 본원에 기술된 바와 같은 세포에 의해 생성된 탄화수소가 수집될 수 있고, 통상적인 수단에 의해 가공되거나 정제될 수 있다. 미세 조류 생산 탄화수소의 수소크래킹에 대한 보고에 대해 문헌[Hillen et al. (Biotechnology and Bioengineering, Vol. XXIV: 193-205 (1982))]을 참조한다. 대안적인 구체예에서, 상기 분획물은 다른 촉매, 예를 들어 유기 화합물, 열 및/또는 무기 화합물로 처리한다. 지질의 바이오디젤로의 가공을 위해, 에스테르 결합 전이 공정이 본 섹션의 하기에 기술된 바와 같이 사용된다.

[0321] 본 발명의 방법을 통해 생성되는 탄화수소는 다양한 산업 응용 분야에서 유용하다. 예를 들어, 거의 모든 유형의 세제 및 세정 제조물에서 사용되는 음이온 계면활성제인 선형 알킬벤젠 설포네이트(LAS)의 생산은 일반적으로 10개 내지 14개의 탄소 원자 사슬을 포함하는 탄화수소를 사용한다. 예를 들어, 미국 특허 제6,946,430호; 제5,506,201호; 제6,692,730호; 제6,268,517호; 제6,020,509호; 제6,140,302호; 제5,080,848호; 및 제5,567,359호를 참조한다. LAS와 같은 계면활성제는 퍼스널 케어 조성물 및 세제(예를 들어, 미국 특허 제5,942,479호; 제6,086,903호; 제5,833,999호; 제6,468,955호; 및 제6,407,044호에 기술된 것들)의 제조에 사용할 수 있다.

[0322] 화석 연료로부터 유래된 출발 물질을 대체할 수 있는 재생가능한 생물학적 출발 물질이 이용가능하고 이의 사용이 바람직할 수 있기 때문에, 바이오디젤, 재생 디젤 및 제트 연료와 같은 연료에서의 생물학적 유래의 탄화수소 성분의 사용에 대한 관심이 증대되고 있다. 생물학적 물질로부터 탄화수소 성분을 생산하기 위한 방법이 시급히 요구되고 있다. 본 발명은 바이오디젤, 재생 디젤 및 제트 연료를 제조하기 위해 생물학적 재료로서 본원에 기술된 방법에 의해 생성된 지질을 사용하여, 바이오디젤, 재생 디젤 및 제트 연료를 생산하는 방법을 제공함으로써 상기 필요성을 충족시킨다.

[0323] 통상적인 디젤 연료는 파라핀 탄화수소가 풍부한 석유 증류물이다. 이들은 370°F 내지 780°F 정도로 넓은 비등 범위를 가지며, 이는 디젤 엔진 차량과 같은 압축 점화 엔진에서 연소용으로서 적합하다. 미국 재료 시험 협회(American Society of Testing and Materials; ASTM)는 세탄수, 혼탁점, 인화점, 점도, 아닐린 점, 황 함량, 물 함량, 재 함량, 동판 부식 및 잔류 탄소와 같은 허용 가능한 범위의 기타 다른 연료 특성과 함께 비등 범위에 따라 디젤 등급을 확립한다. 기술적으로, 적절한 ASTM 사양을 충족하는 바이오매스 또는 기타 다른 것으로부터 유래된 임의의 탄화수소 증류 재료는 디젤 연료(ASTM D975), 제트 연료(ASTM D1655), 또는 이것이 지방산 메틸 에스테르인 경우 바이오디젤(ASTM D6751)로서 정의될 수 있다.

[0324] 추출 후, 본원에 기술된 미생물 바이오매스로부터 회수된 지질 및/또는 탄화수소 성분은 디젤 차량 및 제트 엔

진에 사용하기 위한 연료를 제조하기 위해 화학 처리될 수 있다.

[0325] 바이오디젤은 금색 내지 암갈색으로 생산 공급원료에 따라 색상이 다양한 액체이다. 이것은 실제로 수불혼화성이고, 고비등점 및 저증기압을 갖는다. 바이오디젤은 디젤-엔진 차량에 사용하기 위한 디젤 등가의 가공된 연료를 말한다. 바이오디젤은 생분해성이고 비독성이다. 통상적인 디젤 연료에 대한 바이오디젤의 추가의 이득은 보다 낮은 엔진 마모이다. 통상적으로, 바이오디젤은 C14 내지 C18 알킬 에스테르를 포함한다. 다양한 가공은 본원에 기술된 바와 같이 생성되고 분리된 바이오매스 또는 지질을 디젤 연료로 전환시킨다. 바이오디젤을 생산하기 위한 바람직한 방법은 본원에 기술된 바와 같이 지질의 에스테르 결합 전이에 의한 것이다. 바이오디젤로서 사용하기 위한 바람직한 알킬 에스테르는 메틸 에스테르 또는 에틸 에스테르이다.

[0326] 본원에 기술된 방법에 의해 생성된 바이오디젤은 대부분의 현재 디젤-엔진 차량에서 임의의 농도로 단독으로 사용되거나 통상적인 디젤 연료와 배합하여 사용될 수 있다. 통상적인 디젤 연료(석유 디젤)와 배합되는 경우, 바이오디젤은 약 0.1% 내지 약 99.9%로 존재할 수 있다. 많은 국가에서는 임의의 연료 혼합물에서 바이오디젤의 양을 지칭하기 위해 "B" 인자로서 공지된 시스템을 사용한다. 예를 들어, 20%의 바이오디젤을 함유하는 연료는 B20으로 표시된다. 순수한 바이오디젤은 B100으로 언급된다.

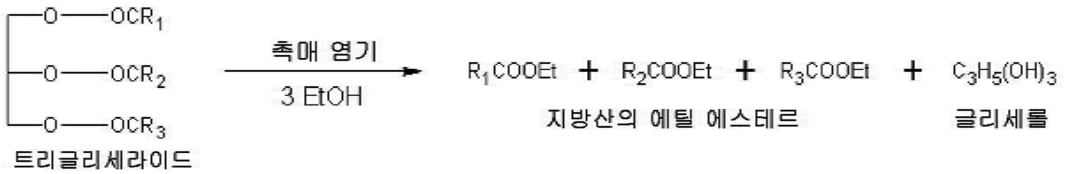
[0327] 바이오디젤은 또한 가정용 및 상업용 보일러에서 가열 연료로서 사용될 수 있다. 기존의 오일 보일러는 고무 부품을 함유할 수 있고 바이오디젤 상에서의 운행을 위해서는 전환을 필요로 할 수 있다. 전환 과정은 일반적으로 비교적 단순하고, 바이오디젤이 강한 용매이기 때문에 고무 부품의 합성 부품으로의 교환을 포함한다. 이의 강한 용해력 때문에, 바이오디젤의 연소는 보일러의 효율을 증가시킬 것이다. 바이오디젤은 순수한 초저황 디젤(ULSD) 연료의 윤활성을 증가시키기 위해 디젤 제형 중에 첨가제로서 사용될 수 있고, 이는 이것이 실제로 황을 내포하지 않기 때문에 유리하다. 바이오디젤은 석유디젤 보다 양호한 용매이고, 이전에 석유디젤 상에서 운행된 차량의 연료관 중에 잔류 침적물을 분해시키는데 사용될 수 있다.

[0328] 바이오디젤은 오일 풍부 바이오매스 중에 함유된 트리글리세라이드의 에스테르 결합 전이에 의해 생성될 수 있다. 따라서, 본 발명의 다른 양태에서, 바이오디젤을 생산하기 위한 방법이 제공된다. 바람직한 구체예에서, 바이오디젤을 생산하는 방법은 (a) 본원에 개시된 방법을 사용하여 지질 함유 미생물을 배양하는 단계, (b) 지질 함유 미생물을 용해시켜 용해물을 생성시키는 단계, (c) 용해된 미생물로부터 지질을 분리하는 단계, 및 (d) 지질 조성물을 에스테르 결합 전이하여 바이오디젤을 생성시키는 단계를 포함한다. 미생물의 생장 방법, 미생물을 용해시켜 용해물을 생성하는 방법, 유기 용매를 포함하는 배지에서 용해물을 처리하여 이종성 혼합물을 형성시키는 방법 및 처리된 용해물을 지질 조성물로 분리하는 방법은 상기 기술되었고 또한 바이오디젤을 생산하는 방법에 사용될 수 있다.

[0329] 바이오디젤의 지질 프로필은 일반적으로 공급원료 오일의 지질 프로필과 고도로 유사하다. 본 발명의 방법 및 조성물에 의해 제공된 기타 다른 오일은 에스테르 결합 전이시켜 바이오디젤을 생산할 수 있고, 이때 지질 프로필은 (a) C8 내지 C14 4% 이상; (b) C8 0.3% 이상; (c) C10 2% 이상; (d) C12 2% 이상; 및 (e) C8 내지 C14 30% 이상을 포함한다.

[0330] 지질 조성물은 에스테르 결합 전이되어 바이오디젤로서 유용한 장쇄의 지방산 에스테르를 생산할 수 있다. 바람직한 에스테르 결합 전이 반응은 하기에 개괄되어 있으며, 염기 촉매된 에스테르 결합 전이 및 재조합 리파제를 사용하는 에스테르 결합 전이를 포함한다. 염기 촉매된 에스테르 결합 전이 공정에서, 트리아실글리세라이드는 알칼리성인 촉매, 통상적으로 수산화칼륨의 존재 하에 메탄올 또는 에탄올과 같은 알코올과 반응시킨다. 상기 반응은 부산물로서 메틸 에스테르 또는 에틸 에스테르 및 글리세린(글리세롤)을 형성한다.

[0331] 동물 및 식물 오일은 통상적으로 3가 알코올인 글리세롤과 유리된 지방산의 에스테르인 트리글리세라이드로 구성된다. 에스테르 결합 전이에서, 트리글리세라이드(TAG) 중의 글리세롤은 메탄올 또는 에탄올과 같은 단쇄 알코올로 대체된다. 통상적인 반응식은 다음과 같다.



트리글리세라이드

[0332]

[0333] 상기 반응에서, 알코올은 염기로 탈양성자화되어 보다 강한 친핵성이 되게 한다. 통상적으로, 에탄올 또는 메탄올은 매우 과량(50배 이하)으로 사용된다. 정상적으로 상기 반응은 상당히 느리게 진행하거나 전혀 진행하지 않

을 것이다. 산 또는 염기뿐만 아니라 열을 사용하여 반응의 보다 신속한 진행을 도와줄 수 있다. 산 또는 염기는 에스테르 결합 전이 반응에 의해 소비되지 않으며, 따라서 이들은 반응물이 아니고 촉매이다. 거의 모든 바이오디젤은 이것이 단지 저온 및 저압을 필요로 하고 98% 초과의 전환 수율을 나타내므로(단, 출발 오일은 수분 및 유리된 지방산이 낮음) 염기 촉매된 기술을 사용하여 생산되었다.

[0334] 에스테르 결합 전이는 또한 상기 개시된 바와 같이 염기 대신 리파제와 같은 효소를 사용하여 수행하였다. 리파제로 촉매된 에스테르 결합 전이는, 예를 들어 실온 내지 80℃의 온도 및 1:1 초과, 바람직하게는 약 3:1인 TAG 대 저급 알코올의 몰비에서 수행할 수 있다. 에스테르 결합 전이에 사용하기에 적합한 리파제는 표 9에 기재된 것들을 포함하지만 이에 제한되지 않는다. 에스테르 결합 전이에 유용한 리파제의 기타 다른 예는, 예를 들어 미국 특허 제4,798,793호; 제4,940,845호; 제5,156,963호; 제5,342,768호; 제5,776,741호 및 WO 89/01032호에서 찾을 수 있다. 상기 리파제로서는 리조푸스(Rhizopus), 아스퍼질러스(Aspergillus), 칸디다(Candida), 무코르(Mucor), 슈도모나스(Pseudomonas), 리조무코르(Rhizomucor), 칸디다 및 휴미콜라(Humicola) 미생물에 의해 생성된 리파제 및 췌장 리파제를 포함하지만 이에 제한되지 않는다.

표 9

[0335] 에스테르 결합 전이에 사용하기에 적합한 리파제

아스퍼질러스 나이저 리파제 ABG73614, 칸디다 안타크티카(Candida antarctica) 리파제 B (novozym-435) CAA83122, 칸디다 실린드라세아(Candida cylindracea) 리파제 AAR24090, 칸디다 리폴리티카(Candida lipolytica) 리파제(리파제 L; Amano Pharmaceutical Co., Ltd.), 칸디다 루고사(Candida rugosa) 리파제(예를 들어, 리파제-OF; Meito Sangyo Co., Ltd.), 무코르 미에헤이(Mucor miehei) 리파제(Lipozyme IM 20), 슈도모나스 플루오레센스(Pseudomonas fluorescens) 리파제 AAA25882, 리조푸스 자포니카스(Rhizopus japonicas) 리파제(리파제 A-10FG) Q7M4U7_1, 리조무코르 미에헤이(Rhizomucor miehei) 리파제 B34959, 리조푸스 오리재(Rhizopus oryzae) 리파제(리파제 F) AAF32408, 세라티아 마르세센스(Serratia marcescens) 리파제(SM Enzyme) ABI13521, 써모마이세스 라누기노사(Thermomyces lanuginosa) 리파제 CAB58509, 리파제 P(Nagase ChemteX Corporation), 및 리파제 QLM(Meito Sangyo Co., Ltd., Nagoya, Japan)

[0336] 바이오디젤용으로 적합한 지방산 에스테르의 제조를 위해 리파제를 사용하는 것에 대한 문제점은 리파제 가격이 강염기 공정에 의해 사용되는 수산화나트륨(NaOH)의 가격보다 훨씬 고가라는 것이다. 상기 문제점은 재순환될 수 있는 고정화된 리파제를 사용함으로써 해결되었다. 그러나, 리파제 기반 공정이 생산 단가의 측면에서 강염기 공정과 경쟁을 하도록 하기 위해서는 최소의 주기 수 동안 재순환된 후 고정화된 리파제의 활성은 유지되어야만 한다. 고정화된 리파제는 에스테르 결합 전이에 통상적으로 사용되는 저급 알코올에 의해 독성이 된다. 미국 특허 제6,398,707호(Wu et al., 2002년 6월 4일자로 등록됨)는 고정화된 리파제의 활성을 증진시키는 방법 및 감소된 활성을 갖는 고정화된 리파제를 재생시키는 방법을 기술한다. 몇몇 적합한 방법은 일정 기간 동안, 바람직하게는 0.5시간 내지 48시간, 보다 바람직하게는 0.5시간 내지 1.5시간 동안 3개 이상의 탄소 원자수를 갖는 알코올 중에 고정화된 리파제를 침지시키는 것을 포함한다. 몇몇 적합한 방법은 또한 불활성화된 고정화된 리파제를 3개 이상의 탄소 원자수를 갖는 알코올로 세척함에 이어서 식물성 오일 중에 0.5시간 내지 48시간 동안 불활성화된 고정화된 리파제를 침지시키는 것을 포함한다.

[0337] 특정 구체예에서, 재조합 리파제는 리파제가 작용하는 지질을 생산하는 동일한 미생물 내에서 발현된다. 적합한 재조합 리파제는 상기 표 9에 열거되고/열거되거나 표 9에 열거된 GenBank 수탁 번호를 갖는 것들, 또는 상기 표 9에 열거된 리파제 중 하나와 70% 이상의 아미노산 동일성을 갖고 리파제 활성을 나타내는 폴리펩티드를 포함한다. 추가의 구체예에서, 효소 활성은 전술된 서열 중 하나와 약 75% 이상, 약 80% 이상, 약 85% 이상, 약 90% 이상, 약 95% 이상, 또는 약 99% 이상 동일성을 갖는 서열 중에 존재하고 상기 서열 모두는 완전히 제시된 것처럼 본원에 참조로 포함된다. 리파제 및 선택가능한 마커를 암호화하는 DNA는 바람직하게 코돈 최적화된 cDNA이다. 미세 조류 중에서 발현하기 위해 유전자를 재암호화하는 방법은 미국 특허 제7,135,290호에 기술되어 있다.

[0338] 바이오디젤에 대한 통상의 국제 표준은 EN 14214이다. ASTM D6751은 미국 및 캐나다에서 참조되는 가장 일반적인 바이오디젤 표준이다. 독일은 DIN EN 14214를 사용하고, UK는 BS EN 14214을 준수할 것을 요구한다. 생성물이 이들 표준에 부합하는지의 여부를 결정하기 위한 기본 산업적 시험은 통상적으로 가스 크로마토그래피, HPLC 등을 포함한다. 품질 표준을 충족하는 바이오디젤은 매우 비독성이고, 독성 등급(LD50)은 50mL/kg을 초과한다.

[0339] ASTM 표준을 충족하는 바이오디젤이 비독성이어야만 하지만, 결정화하고/결정화하거나 침전하고 용액으로부터

침강물로서 나오는 경향이 있는 오염물이 있을 수 있다. 침강물 형성은 특히 바이오디젤이 보다 저온에서 사용되는 경우 문제가 된다. 침강 또는 침전은 연료 흐름을 감소시키고, 연료관을 막히게 하며, 필터를 막히게 하는 것 등과 같은 문제를 유발할 수 있다. 고품질의 생성물을 제조하기 위해 바이오디젤에서 상기 오염물 및 침강물을 제거하는 것을 다루는 공정들은 당업계에 널리 공지되어 있다. 상기 공정들의 예로서는 인지질 및 유리 지방산과 같은 오염물을 제거하기 위한 오일의 전처리(예를 들어, 고무 성질 제거(degumming), 가성 정제 및 실리카 흡착 여과) 및 냉여과를 포함하지만 이에 제한되지 않는다. 냉여과는 생성 후 바이오디젤에 존재하는 임의의 미립자 및 침강물을 제거하기 위해 특별히 개발된 공정이다. 상기 공정은 바이오디젤을 냉각시키고 연료가 보다 저온에서 사용되는 경우 형성될 수 있는 임의의 침강물 또는 침전물을 여과 제거한다. 상기 공정은 당업계에 널리 공지되어 있고 미국 특허 출원 공개 제2007-0175091호에 기술되어 있다. 적합한 방법은 바이오디젤을 약 38 ℃ 미만의 온도로 냉각시켜 불순물 및 오염물이 바이오디젤 액체 중에서 미립자로서 침전하도록 하는 것을 포함할 수 있다. 이어서 규조토 또는 기타 다른 여과 재료를 냉각된 바이오디젤에 첨가하여 슬러리를 형성시킬 수 있고, 이것은 이어서 압력 리프(pressure leaf) 또는 기타 다른 유형의 필터를 통해 여과하여 미립자를 제거할 수 있다. 이어서 여과된 바이오디젤은 연마(polish) 필터에 통과시켜 임의의 잔류 침강물 및 규조토를 제거하여 최종 바이오디젤 생성물을 생성할 수 있다.

[0340] 실시예 9는 프로토테카 모리포르미스로부터 유래한 트리글리세라이드 오일을 사용하는 바이오디젤의 제조법을 기술한다. 실시예 9에서 생산된 바이오디젤의 ASTM D6751 A1 방법에 의한 콜드 소킹 여과능(Cold Soak Filterability)은 300ml의 용적에 대해 120초였다. 상기 시험은 16시간 동안 40°F로 냉각시키고 실온으로 가온되도록 하며, 스테인리스강 지지체를 가진 0.7마이크론 유리 섬유 필터를 사용하여 진공 여과하는, 300ml의 B100의 여과를 포함한다. 본 발명의 오일은 에스테르 결합 전이되어 120초 미만, 100초 미만 및 90초 미만의 콜드 소킹 시간을 가진 바이오디젤을 생성할 수 있다.

[0341] 후속 공정은 또한 바이오디젤이 특히 냉온에서 사용될 경우 이용될 수 있다. 상기 공정은 윈터리제이션(winterization) 및 분별을 포함한다. 2개 공정은 혼탁점(바이오디젤이 결정화하기 시작하는 온도)을 저하시킴으로써 연료의 냉각 흐름 및 동결 방지 수행능을 개선시키도록 설계된다. 바이오디젤의 동결 방지 처리에는 다수의 접근법이 있다. 한가지 접근법은 바이오디젤을 석유 디젤과 배합하는 것이다. 다른 접근법은 바이오디젤의 혼탁점을 저하시킬 수 있는 첨가제를 사용하는 것이다. 다른 접근법은 첨가제 중에 혼합하고 포화물이 결정화하도록 하며, 이어서 결정을 여과 제거함으로써 무분별하게 포화된 메틸 에스테르를 제거하는 것이다. 분별은 선택적으로 메틸 에스테르를 개별 성분 또는 분획으로 분리하여 특정 메틸 에스테르가 제거 또는 포함될 수 있게 한다. 분별 방법은 우레아 분별, 용매 분별 및 열 증류를 포함한다.

[0342] 본 발명의 방법에 의해 제공되는 다른 가치있는 연료는 C10:0, C12:0, C14:0, C16:0 및 C18:0과 같은 알칸을 포함하여 바이오디젤과는 구분될 수 있는 재생 디젤이다. 고품질의 재생 디젤은 ASTM D975 표준에 부합한다. 본 발명의 방법에 의해 생성된 지질은 공급원료로서 작용하여 재생 디젤을 생성할 수 있다. 따라서, 본 발명의 또 다른 양태에서, 재생 디젤을 생성하는 방법이 제공된다. 재생 디젤은 열수 가공(수소 처리); 수소화처리; 및 간접적인 액화의 3가지 이상의 공정에 의해 생성될 수 있다. 이들 공정으로 비에스테르 증류물을 생성한다. 이들 공정 동안에, 본원에 기술된 바와 같이 생성되고 분리된 트리아실글리세라이드는 알칸으로 전환된다.

[0343] 하나의 구체예에서, 재생 디젤을 생성하는 방법은 (a) 본원에 개시된 방법을 사용하여 지질 함유 미생물을 배양하고 (b) 미생물을 용해시켜 용해물을 생성하며, (c) 용해된 미생물로부터 지질을 분리하고 (d) 지질을 탈산소화하고 수소화처리하여 알칸을 생성함으로써 재생 디젤이 생성되도록 하는 것을 포함한다. 재생 디젤을 제조하기에 적합한 지질은 헥산과 같은 유기 용매를 사용하거나 미국 특허 제5,928,696호에 기술된 것들과 같은 기타 다른 방법을 통해 미생물 바이오매스로부터 추출을 통해 수득할 수 있다. 몇몇 적합한 방법으로서는 기계적 프레싱 및 원심분리를 포함할 수 있다.

[0344] 몇몇 방법에서, 미생물 지질은 먼저 수소화처리와 연계하여 크래킹하여 각각 탄소 사슬 길이를 감소시키고 이중 결합을 포화시킨다. 상기 물질을 또한 이어서 수소화처리와 연계하여 이성체화한다. 이어서 나프타 분획을 증류를 통해 제거하고 이어서 추가 증류하여 디젤 연료에 요구되는 성분을 증발시키고 증류시킴으로써 ASTM D975 표준에 충족하도록 하고 D975 표준을 충족시키기 위해 원하는 것보다 무거운 성분을 잔류시킨다. 트리글리세라이드 오일을 포함하는 오일을 화학적으로 변형시키는 수소 처리, 수소 크래킹, 탈산소화 및 이성체화 방법은 당업계에 널리 공지되어 있다. 예를 들어, 유럽 특허 출원 EP1741768(A1); EP1741767(A1); EP1682466(A1); EP1640437(A1); EP1681337(A1); EP1795576(A1); 및 미국 특허 제7,238,277호; 제6,630,066호; 제6,596,155호; 제6,977,322호; 제7,041,866호; 제6,217,746호; 제5,885,440호; 제6,881,873호를 참조한다.

[0345] 재생 디젤을 생성하기 위한 방법의 하나의 구체예에서, 알칸을 생성하기 위한 지질의 처리는 지질 조성물의 수소 처리에 의해 수행한다. 열수 가공에서, 통상적으로 바이오매스는 승온 및 승압에서 물 중에서 반응시켜 오일 및 잔류 고체를 형성시킨다. 전환 온도는 통상적으로 300°F 내지 660°F이고, 이때 압력은 주로 액체로서 물을 유지하기에 충분한 100 표준 대기압 내지 170 표준 대기압(atm)이다. 반응 시간은 대략 15분 내지 30분 정도이다. 반응을 종결한 후 유기물을 물로부터 분리한다. 이로써 디젤에 적합한 증류물이 생성된다.

[0346] 재생 디젤을 제조하는 몇몇 방법에서, 트리글리세라이드를 처리하는 제1 단계는 이중 결합을 포화시키기 위한 수소 가공에 이어서, 수소 및 촉매의 존재 하에 승온에서 탈산소화시키는 것이다. 몇몇 방법에서, 수소화 및 탈산소화는 동일한 반응에서 발생한다. 기타 다른 방법에서, 탈산소화는 수소화 전에 일어난다. 이어서 이성체화는 임의로 또한 수소 및 촉매의 존재 하에서 수행한다. 나프타 성분은 바람직하게 증류를 통해 제거한다. 예를 들어, 미국 특허 제5,475,160호 (트리글리세라이드의 수소화); 제5,091,116호(탈산소화, 수소화 및 가스 제거); 제6,391,815호(수소화); 및 제5,888,947호(이성체화)를 참조한다.

[0347] 트리글리세라이드의 수소화를 위한 하나의 적합한 방법은 구리, 아연, 마그네슘 및 란탄염의 수용액 및 알칼리 금속 또는 바람직하게는 탄산암모늄의 다른 용액을 제조하는 것을 포함한다. 상기 2개의 용액을 약 20℃ 내지 약 85℃의 온도로 가열하고 침전 용기 내 pH가 5.5 내지 7.5로 유지되어 촉매를 형성하도록 하는 비율로 침전 용기 내로 함께 계량하여 부가할 수 있다. 추가의 물은 초기에 침전 용기 내에 사용할 수 있거나 염 용액 및 침전 용액과 동시에 첨가할 수 있다. 이어서 생성된 침전물을 약 300℃에서 완전히 세척하고 건조시키며 하소시키고 약 100℃ 내지 약 400℃의 온도 범위에서 수소 중에서 활성화시킬 수 있다. 이어서 하나 이상의 트리글리세라이드는 반응기 내 전술된 촉매의 존재 하에서 수소와 접촉시키고 반응시킬 수 있다. 상기 반응기는 살수층(trickle bed) 반응기, 고정 상 가스-고체 반응기, 충진된 기포 탑 반응기, 연속 교반 탱크 반응기, 슬러리 상 반응기 또는 당업계에 공지된 임의의 기타 다른 적합한 반응기 유형일 수 있다. 상기 방법은 배치식으로 또는 연속식으로 수행할 수 있다. 반응 온도는 통상적으로 약 170℃ 내지 약 250℃의 범위인 한편, 반응 압력은 통상적으로 약 300psig 내지 약 2000psig 범위이다. 더욱이, 본 발명의 방법에서 수소 대 트리글리세라이드의 몰 비는 통상적으로 약 20:1 내지 약 700:1 범위이다. 상기 방법은 통상적으로 시간당 중량 공간 속도(WHSV)에서 수행하고 이의 범위는 약 0.1hr-1 내지 약 5hr-1이다. 당업자는 반응을 위해 요구되는 시간이 사용되는 온도, 수소 대 트리글리세라이드의 몰비, 및 수소 분압에 따라 다양할 것임을 인지할 것이다. 상기 수소화 공정에 의해 생성되는 생성물은 지방 알코올, 글리세롤, 미량의 파라핀 및 미반응 트리글리세라이드를 포함한다. 이들 생성물은 통상적으로, 예를 들어 증류, 추출, 여과, 결정화 등과 같은 통상적인 수단에 의해 분리한다.

[0348] 석유 정유업자는 수소화처리를 사용하여 공급물을 수소로 처리함으로써 불순물을 제거한다. 수소화처리 전환 온도는 통상적으로 300°F 내지 700°F이다. 압력은 통상적으로 40atm 내지 100atm이다. 반응 시간은 통상적으로 대략 10분 내지 60분 정도이다. 고체 촉매를 사용하여 특정 반응 속도를 증가시키고 특정 생성물에 대한 선택성을 개선시키며 수소 소비를 최적화한다.

[0349] 오일의 탈산소화를 위해 적합한 방법은 오일을 약 350°F 내지 약 550°F 범위의 온도로 가열하고 연속적으로 가열된 오일을 약 5분 이상 동안 적어도 약 대기압 내지 그 이상의 범위 압력 하에서 질소와 접촉시키는 것을 포함한다.

[0350] 이성체화를 위해 적합한 방법은 알칼리 이성체화 및 당업계에 공지된 기타 다른 오일 이성체화를 사용하는 것을 포함한다.

[0351] 수소 처리 및 수소 가공은 궁극적으로 트리글리세라이드 공급물의 분자량 감소를 유도한다. 상기 트리글리세라이드 분자는 수소 가공 조건 하에서 4개의 탄화수소 분자, 즉 프로판 분자 및 통상적으로 C8 내지 C18 범위인 3개의 보다 무거운 탄화수소 분자로 감소한다.

[0352] 따라서, 하나의 구체예에서, 본 발명의 지질 조성물 상에서 수행되는 하나 이상의 화학 반응(들)의 생성물은 ASTM D975 재생 디젤을 포함하는 알칸 혼합물이다. 미생물에 의한 탄화수소의 생성은 문헌[Metzger et al. Appl Microbiol Biotechnol (2005) 66: 486-496] 및 [A Look Back at the U.S. Department of Energy's Aquatic Species Program: Biodiesel from Algae, NREL/TP-580-24190, John Sheehan, Terri Dunahay, John Benemann 및 Paul Roessler (1998)]에 의해 검토되었다.

[0353] 디젤 연료의 증류 특성은 T10-T90(각각 10% 및 90%의 용적이 증류되는 온도)의 관점에서 기술한다. 재생 디젤은 프로토테카 모리포르미스 트리글리세라이드 오일로부터 생성되었고 실시예 9에 기술된다. 실시예 9에 생성된 물질의 T10-T90은 57.9℃였다. 본원에 개시된 오일의 수소 처리, 이성체화 및 기타 다른 공유 변형 방법, 및 본원

에 개시된 증류 및 분별(예를 들어, 냉여과) 방법을 사용하여, 본원에 개시된 방법에 따라 생성된 트리글리세라이드 오일을 사용하여 기타 다른 T10-T90 범위가 예를 들어 20℃, 25℃, 30℃, 35℃, 40℃, 45℃, 50℃, 60℃ 및 65℃인 재생 디젤 조성물을 생성할 수 있다.

[0354] 실시예 9에서 생성된 물질의 T10은 242.1℃이다. 본원에 개시된 오일의 수소 처리, 이성체화 및 기타 다른 공유 변형 방법, 및 본원에 개시된 증류 및 분별(예를 들어, 냉여과) 방법을 사용하여 기타 다른 T10 값, 예를 들어 T10이 180 내지 295, 190 내지 270, 210 내지 250, 225 내지 245, 및 290 이상의 T10인 재생 디젤 조성물을 생성할 수 있다.

[0355] 실시예 9에서 생성된 물질의 T90은 300℃이다. 본원에 개시된 오일의 수소 처리, 이성체화 및 기타 다른 공유 변형 방법, 및 본원에 개시된 증류 및 분별(예를 들어, 냉여과) 방법을 사용하여 기타 다른 T90 값, 예를 들어 T90이 280 내지 380, 290 내지 360, 300 내지 350, 310 내지 340, 및 290 이상인 재생 디젤 조성물을 생성할 수 있다.

[0356] 실시예 9에서 생성된 물질의 FBP는 300℃이다. 본원에 개시된 오일의 수소 처리, 이성체화 및 기타 다른 공유 변형 방법, 및 본원에 개시된 증류 및 분별(예를 들어, 냉여과) 방법을 사용하여 기타 다른 FBP 값, 예를 들어 FBP가 290 내지 400, 300 내지 385, 310 내지 370, 315 내지 360, 및 300 이상인 재생 디젤 조성물을 생성할 수 있다.

[0357] 본 발명의 방법 및 조성물에 의해 제공된 기타 다른 오일은 수소처리, 이성체화 및 기타 다른 공유 변형의 조합을 거칠 수 있고, (a) C8 내지 C14 4% 이상; (b) C8 0.3% 이상; (c) C10 2% 이상; (d) C12 2% 이상; 및 (e) C8 내지 C14 30% 이상을 포함하는 지질 프로필을 갖는 오일을 포함한다.

[0358] 통상적인 초저황 디젤은 2단계 공정에 의해 임의의 형태의 바이오매스로부터 생성될 수 있다. 먼저, 바이오매스를 수소 및 일산화탄소가 풍부한 가스 혼합물인 합성가스로 전환시킨다. 이어서, 합성가스를 촉매적으로 액체로 전환시킨다. 통상적으로, 액체의 생성은 피셔-트로프슈(Fischer-Tropsch)(FT) 합성을 사용하여 성취한다. 상기 기술은 석탄, 천연 가스 및 중유에 적용한다. 따라서, 재생 디젤을 생성하기 위한 방법의 또 다른 바람직한 구체예에서, 알칸을 생성하기 위한 액체 조성물의 처리는 액체 조성물의 간접적인 액화에 의해 수행한다.

[0359] 본 발명은 또한 제트 연료를 제조하는 방법을 제공한다. 제트 연료는 담황색 색상으로 투명하다. 가장 일반적인 연료는 국제 표준 세트의 사양으로 제조되는 항공기 A-1로서 분류되는 무연/파라핀 오일계 연료이다. 제트 연료는 다수의 상이한, 가능하게는 천 개 이상의 많은 상이한 탄화수소의 혼합물이다. 이들의 크기 범위(분자량 또는 탄소수)는 생성물에 대한 요건, 예를 들어 동결점 또는 발연점에 의해 제한된다. 등유형(Kerosene-type) 항공기 연료(제트 A 및 제트 A-1을 포함함)는 약 8개 내지 16개 탄소수의 탄소수 분포를 갖는다. 와이드-컷 또는 나프타형 항공기 연료(제트 B를 포함함)는 통상적으로 약 5개 내지 15개 탄소의 탄소수 분포를 갖는다.

[0360] 2개의 항공기(제트 A 및 제트 B)는 다수의 첨가제를 함유할 수 있다. 유용한 첨가제로서는 항산화제, 대전방지제, 부식 억제제, 및 연료 시스템 아이싱 억제제(FSII)를 포함하지만 이에 제한되지 않는다. 항산화제는 검화(gumming)를 방지하고 일반적으로 알킬화된 페놀, 예를 들어 AO-30, AO-31, 또는 AO-37을 기초로 한다. 대전방지제는 정전기를 소멸시켜 스파킹을 차단한다. 활성 성분으로서 디노닐나프틸설폰산(DINNSA)을 갖는 Stadis 450이 하나의 예이다. 부식 억제제, 예를 들어 DCI-4A는 대중 및 군사 연료용으로 사용되고, DCI-6A는 군사 연료용으로 사용된다. FSII 제제는, 예를 들어 Di-EGME를 포함한다.

[0361] 본 발명의 하나의 구체예에서, 제트 연료는 기존의 제트 연료와 조류 연료를 배합함으로써 생성된다. 본 발명의 방법에 의해 생성된 지질은 공급원료로서 작용하여 제트 연료를 생성할 수 있다. 따라서, 본 발명의 다른 양태에서, 제트 연료를 제조하는 방법이 제공된다. 이와 함께 본 발명의 방법에 의해 생성된 지질로부터 제트 연료를 제조하는 2가지 방법으로서 유동 촉매 크래킹(FCC); 및 수소첨가탈산소화(HDO)가 제공된다.

[0362] 유동 촉매 크래킹(FCC)은 중질 원유 분획물 유래의 올레핀, 특히 프로필렌을 제조하기 위해 사용되는 하나의 방법이다. 본 발명의 방법에 의해 생성된 지질은 올레핀으로 전환시킬 수 있다. 상기 방법은 FCC 영역을 통해 생성된 지질을 유동시키고 제트 연료로서 유용한 올레핀으로 구성된 생성물 스트림을 수집함을 포함한다. 생성된 지질은 크래킹 조건에서 크래킹 촉매와 접촉시켜 제트 연료로서 유용한 올레핀 및 탄화수소를 포함하는 생성물 스트림을 제공한다.

[0363] 하나의 구체예에서, 제트 연료를 제조하는 방법은 (a) 본원에 개시된 방법을 사용하여 지질 함유 미생물을 배양하는 단계, (b) 지질 함유 미생물을 용해시켜 용해물을 생성하는 단계, (c) 용해물로부터 지질을 분리하는 단계, 및 (d) 지질 조성물을 처리함으로써 제트 연료를 생성하는 단계를 포함한다. 제트 연료를 제조하는 방법

의 하나의 구체예에서, 상기 지질 조성물은 유체 촉매 크래킹 영역을 통류할 수 있고 이는 하나의 구체예에서 지질 조성물을 크래킹 조건에서 크래킹 촉매와 접촉시켜 C2 내지 C5 올레핀을 포함하는 생성물 스트림을 제공함을 포함할 수 있다.

[0364] 상기 방법의 특정 구체예에서, 지질 조성물에 존재할 수 있는 임의의 오염물을 제거하는 것이 바람직할 수 있다. 따라서, 유체 촉매 크래킹 영역을 통해 지질 조성물을 유동시키기 전에, 지질 조성물을 전처리한다. 전처리는 지질 조성물을 이온 교환 수지와 접촉시키는 것을 포함할 수 있다. 이온 교환 수지는 산성 이온 교환 수지, 예를 들어 Amberlyst™-15이고, 지질 조성물이 통류(상향 통류 또는 하향 통류)하는 반응기 내 상으로서 사용될 수 있다. 기타 다른 전처리는 지질 조성물의 산, 예를 들어 황산, 아세트산, 질산 또는 염산과의 접촉에 의한 약산 세척을 포함할 수 있다. 접촉은 일반적으로 주위 온도 및 대기압에서 희석 산 용액으로 수행된다.

[0365] 임의로 전처리된 지질 조성물은 탄화수소 성분이 올레핀으로 크래킹되는 FCC 영역으로 유동한다. 촉매 크래킹은 반응 영역중의 지질 조성물을 미분된 미립자 물질로 구성된 촉매와 접촉시킴으로써 성취된다. 상기 반응은 수소 크래킹이 아닌 촉매 크래킹이고, 부가되는 수소의 부재 또는 수소의 소비 하에서 수행된다. 크래킹 반응이 진행함으로써 상당한 양의 코크가 촉매상에 침적된다. 상기 촉매는 재생 영역에서 촉매 유래의 코크를 연소시킴으로써 고온에서 재생된다. 본원에서, “코크 촉매”로서 언급되는 코크 함유 촉매는 반응 영역으로부터 재생될 재생 영역으로 연속으로 수송되고, 재생 영역 유래의 필수적 코크 부재의 재생된 촉매에 의해 대체된다. 다양한 가스 스트림에 의한 촉매 입자의 유동화는 반응 영역과 재생 영역간에 촉매의 수송을 허용한다. 촉매의 유동화 스트림에서 본원에 기술된 지질 조성물의 탄화수소와 같은 탄화수소를 크래킹하고, 반응 영역 및 재생 영역간에 촉매를 수송하고 재생기 내 코크를 연소시키는 방법은 FCC 방법 기술 분야의 당업자에게 널리 공지되어 있다. C2 내지 C5 올레핀을 제조하기 위해 지질 조성물을 크래킹하는데 유용한 예시적인 FCC 적용 및 촉매는 이의 전체가 본원에 참조로 포함된 미국 특허 제6,538,169호 및 제7,288,685호에 기술되어 있다.

[0366] 적합한 FCC 촉매는 일반적으로 동일한 매트릭스에 존재할 수 있거나 존재하지 않을 수 있는 2개 이상의 성분을 포함한다. 몇몇 구체예에서, 2개의 성분 둘 다는 전체 반응 용기를 거쳐 순환할 수 있다. 제1 성분은 일반적으로 유동화된 촉매 크래킹의 분야에 사용되는 임의의 널리 공지된 촉매, 예를 들어 활성 무정형 점토형 촉매 및/또는 고활성의 결정 분자체를 포함한다. 분자체 촉매는 원하는 생성물에 대한 이의 매우 개선된 선택성 때문에 무정형 촉매보다 바람직할 수 있다. 몇몇 바람직한 구체예에서, 제올라이트는 FCC 공정에서 분자체로서 사용할 수 있다. 바람직하게, 제1 촉매 성분은 실리카 또는 알루미나 및 카올린과 같은 불활성 충전제를 포함하는, 대형 공극 제올라이트(예를 들어 Y형 제올라이트), 활성 알루미나 물질, 결합제 물질을 포함한다.

[0367] 하나의 구체예에서, 본 발명의 지질 조성물의 크래킹은 FCC 영역의 라이저(riser) 섹션 또는 대안적으로 리프트 섹션에서 발생한다. 상기 지질 조성물은 노즐에 의해 라이저로 도입되어 지질 조성물의 신속한 증발을 유도한다. 촉매와 접촉시키기 전에, 상기 지질 조성물은 통상적으로 약 149℃ 내지 약 316℃(300°F 내지 600°F)의 온도를 가질 것이다. 상기 촉매는 블렌딩 용기로부터 라이저로 유동하여 여기서 촉매는 약 2초 이하의 시간 동안 지질 조성물과 접촉한다.

[0368] 배합된 촉매 및 반응된 지질 조성물 증기는 이어서 출구를 통해 라이저의 최상부로부터 배출되고, 올레핀, 및 상당한 양의 코크로 덮여지고 일반적으로 “코크된 촉매”로 언급되는 촉매 입자의 수집물을 포함하는 크래킹된 생성물 증기 스트림으로 분리된다. 원하는 생성물이 바람직하지 못한 기타 다른 생성물로 추가로 전환되는 것을 촉진시킬 수 있는, 지질 조성물과 촉매의 접촉시간을 최소화하기 위해, 스월 암 배열 장치(swirl arm arrangement)와 같은 임의의 배열된 분리기를 사용하여 생성물 스트림으로부터 신속하게 코크된 촉매를 제거할 수 있다. 분리기, 예를 들어 스월 암 분리기는 챔버의 하한부에 위치한 스트립핑 영역(stripping zone)과 함께 챔버의 챔버의 상한부에 위치한다. 스월 암 배열 장치에 의해 분리된 촉매는 스트립핑 영역으로 떨어진다. 가벼운 올레핀 및 몇몇 촉매를 포함하는 크래킹된 탄화수소를 포함하는 크래킹된 생성물 증기 스트림은 사이클론과 소통하는 도관을 통해 챔버로부터 배출된다. 상기 사이클론은 생성물 증기 스트림으로부터 잔류 촉매 입자를 제거하여 입자 농도를 매우 낮은 수준으로 감소시킨다. 이어서 상기 생성물 증기는 분리 용기의 최상부로부터 배출된다. 사이클론에 의해 분리된 촉매는 분리 용기에 이어서 스트립팅 영역으로 복귀한다. 스트립핑 영역은 증기와의 향류 접촉에 의해 촉매 표면으로터 흡착된 탄화수소를 제거한다.

[0369] 낮은 탄화수소 분압은 가벼운 올레핀이 우선적으로 생성되도록 작용한다. 따라서, 라이저 압력은 약 172kPa 내지 241kPa(25psia 내지 35psia)로 설정되며, 이 때 탄화수소 분압은 약 35kPa 내지 172kPa(5psia 내지 25psia)이고 바람직한 탄화수소 분압은 약 69kPa 내지 138kPa(10psia 내지 20psia)이다. 탄화수소에 대한 상대적으로 낮은 분압은 희석제가 지질 조성물의 10중량% 내지 55중량% 및 바람직하게는 지질 조성물의 약 15중량% 정도인

희석제로서 증기를 사용함으로써 성취된다. 기타 다른 희석제, 예를 들어 건조 가스를 사용하여 등가의 탄화수소 분압에 도달할 수 있다.

[0370] 라이저 출구에서 크래킹된 스트림의 온도는 약 510℃ 내지 621℃ (950°F 내지 1150°F)이다. 그러나, 566℃(1050°F) 초과의 라이저 출구 온도는 보다 많은 건조 가스 및 보다 많은 올레핀을 생성시킨다. 반면에, 566℃(1050°F) 미만의 라이저 출구 온도는 에틸렌 및 프로필렌을 적게 생성시킨다. 따라서, 약 566℃ 내지 약 630℃의 바람직한 온도, 약 138kPa 내지 약 240kPa(20psia 내지 35psia)의 바람직한 압력에서 FCC 공정을 수행하는 것이 바람직하다. 상기 공정의 다른 조건은 약 5 내지 약 20 및 바람직하게는 약 10 내지 약 15로 다양할 수 있는 지질 조성물에 대한 촉매 비율이다.

[0371] 제트 연료를 생성하는 방법의 하나의 구체예에서, 상기 지질 조성물은 FCC 반응기의 리프트 섹션으로 도입된다. 리프트 섹션에서 온도는 매우 고온이고 약 700℃(1292°F) 내지 약 760℃(1400°F)의 범위이며, 지질 조성물에 대한 촉매의 비율은 약 100 내지 약 150일 것이다. 이것은 지질 조성물의 리프트 섹션으로의 도입이 상당양의 프로필렌 및 에틸렌을 생성시킬 것으로 예상된다.

[0372] 본원에 기술된 바와 같이 생성된 지질 조성물 또는 지질을 사용하여 제트 연료를 생성시키는 방법의 또 다른 구체예에서, 지질 조성물 또는 지질의 구조는 수소탈산소화(HDO)로서 언급되는 공정에 의해 파괴된다. HDO는 수소에 의한 산소의 제거를 의미하는데, 즉 산소는 물질의 구조를 파괴시키면서 제거된다. 올레핀 이중 결합을 수소화시키고 임의의 황 및 질소 화합물을 제거한다. 황 제거는 소위 수소탈황화(HDS)로 불린다. 원료(지질 조성물 또는 지질)의 전처리 및 순도는 촉매의 작용 수명에 기여한다.

[0373] 일반적으로, HDO/HDS 단계에서, 수소는 공급 원료(지질 조성물 또는 지질)와 혼합하고 이어서 상기 혼합물을 단일상 또는 2개상 공급 원료로서 병류로서 촉매 상에 통과시킨다. HDO/HDS 단계 후, 생성물 분획을 분리하고 별도의 이성체화 반응기에 통과시킨다. 생물학적 출발 물질용 이성체화 반응기는 병류 반응기로서 문헌(FI 100248)에 기술되어 있다.

[0374] 탄화수소 공급물, 예를 들어 본원에서의 지질 조성물 또는 지질을 수소화시킴에 의한 연료의 제조 방법은 또한 지질 조성물 또는 지질을 병류로서 제1 수소화 영역을 통해 수소 가스와 함께 통과시킴으로써 수행될 수 있고, 이후 탄화수소 유출물은 수소 가스를 탄화수소 유출물에 대하여 향류로서 제2 수소화 영역에 통과시킴으로써 제2 수소화 영역에서 추가로 수소화시킨다. 지질 조성물을 크래킹하여 C2 내지 C5 올레핀을 제조하는데 유용한 예시적인 HDO 적용 및 촉매는 전체가 참조로 포함되어 있는 미국 특허 제7,232,935호에 기술되어 있다.

[0375] 통상적으로, 수소탈산소화 단계에서, 본원에서의 지질 조성물 또는 지질과 같은 생물학적 성분의 구조는 산소, 질소, 인 및 황 화합물로 분해되고 가스로서의 가벼운 탄화수소는 제거되며, 올레핀 결합은 수소화된다. 상기 방법의 제2 단계에서, 즉 소위 이성체화 단계에서, 이성체화는 낮은 온도에서 탄화수소 사슬을 분지화하고 파라핀의 수행능을 개선시키기 위하여 수행된다.

[0376] 제1 단계, 즉 크래킹 공정의 HDO 단계에서, 수소 가스, 및 수소화될 본원의 지질 조성물 또는 지질은 HDO 촉매 상 시스템에 병류로서 또는 향류로서 통과시키고, 상기 촉매 상 시스템은 하나 이상의 촉매 상(들), 바람직하게 1개 내지 3개의 촉매 상을 포함한다. HDO 단계는 통상적으로 병류 방식으로 작동한다. 2개 이상의 촉매 상을 포함하는 HDO 촉매 상 시스템의 경우에, 상기 상중 하나 이상은 향류 원리를 사용하여 작동할 수 있다. HDO 단계에서, 압력은 20바 내지 150바, 바람직하게는 50바 내지 100바로 다양하고, 온도는 200℃ 내지 500℃, 바람직하게는 300℃ 내지 400℃의 범위로 다양하다. HDO 단계에서, 원소주기율표의 VII 족 및/또는 VIB족 금속을 함유하는 공지된 수소화 촉매가 사용될 수 있다. 바람직하게, 수소화 촉매는 지지된 Pd, Pt, Ni, NiMo 또는 CoMo 촉매이고, 상기 지지체는 알루미나 및/또는 실리카이다. 통상적으로, NiMo/Al2O3 및 CoMo/Al2O3 촉매가 사용된다.

[0377] HDO 단계 전에, 본원에서 지질 조성물 또는 지질은 임의로 보다 온화한 조건 하에서 예비수소화로 처리하여 이중 결합의 부반응을 회피할 수 있다. 상기 예비수소화는 예비수소화 촉매의 존재 하에 50℃ 내지 400℃의 온도 및 1바 내지 200바의 수소 압력, 바람직하게는 150℃ 내지 250℃의 온도 및 10바 내지 100바의 수소 압력에서 수행한다. 상기 촉매는 원소주기율표의 VIII족 및/또는 VIB족 유래의 금속을 함유할 수 있다. 바람직하게, 예비수소화 촉매는 지지된 Pd, Pt, Ni, NiMo 또는 CoMo 촉매이고, 상기 지지체는 알루미나 및/또는 실리카이다.

[0378] 수소를 함유하는 HDO 단계 유래의 가스 스트림을 냉각시킨 후, 일산화탄소, 이산화탄소, 질소, 인 및 황 화합물, 가벼운 가스 탄화수소 및 기타 다른 불순물을 이로부터 제거한다. 압축시킨 후 정제된 수소 또는 재순환된 수소를 제1 촉매 상으로 및/또는 촉매 상 사이로 복귀시켜 인출된 가스 스트림을 구성하도록 한다. 물은 응축 액체로부터 제거한다. 상기 액체는 제1 촉매 상 또는 촉매 상 사이로 통과시킨다.

[0379] HDO 단계 후, 생성물은 이성체화 단계를 거친다. 탄화수소를 이성체화 촉매와 접촉시키기 전에 불순물을 가능한 한 완전히 제거하는 것이 공정을 위해 필요하다. 상기 이성체화 단계는 임의의 스트리핑 단계를 포함하고, 이때 HDO 단계 유래의 반응 생성물은 수증기 또는 가벼운 탄화수소, 질소 또는 수소와 같은 적합한 가스를 스트리핑시킴으로써 정제될 수 있다. 임의의 스트리핑 단계는 이성체화 촉매의 상부 유니트에서 향류 방식(여기서, 가스 및 액체는 서로 접촉함)으로 수행하거나, 향류 원리를 사용하는 별도의 스트리핑 유니트에서 실제 이성체화 반응기 전에 수행한다.

[0380] 스트리핑 단계 후, 수소 가스 및 본원에서의 수소화된 지질 조성물 또는 지질, 및 임의로 n-파라핀 혼합물을 하나 또는 여러 가지의 촉매 상(들)을 포함하는 반응성 이성체화 유니트에 통과시킨다. 이성체화 단계의 촉매 상은 병류 또는 향류 방식으로 작동할 수 있다.

[0381] 향류 원리가 이성체화 단계에 적용되는 것이 상기 공정을 위해 중요하다. 이성화체 단계에서, 이것은 임의의 스트리핑 단계 또는 이성체화 반응 단계 중 하나 또는 둘 다를 향류 방식으로 진행함에 의해 수행한다. 이성체화 단계에서, 압력은 20바 내지 150바, 바람직하게는 20바 내지 100바의 범위로 다양하고, 온도는 200℃ 내지 500℃, 바람직하게는 300℃ 내지 400℃이다. 이성체화 단계에서, 당업계에 공지된 이성체화 촉매가 사용될 수 있다. 적합한 이성체화 촉매는 분자체 및/또는 VII족 금속 및/또는 담체를 함유한다. 바람직하게, 이성체화 촉매는 SAPO-11 또는 SAPO41 또는 ZSM-22 또는 ZSM-23 또는 페리어라이트(ferrierite) 및 Pt, Pd 또는 Ni 및 A1203 또는 SiO2를 함유한다. 통상적인 이성체화 촉매는, 예를 들어 Pt/SAPO-11/A1203, Pt/ZSM-22/A1203, Pt/ZSM-23/A1203 및 Pt/SAPO-11/SiO2이다. 이성체화 단계 및 HDO 단계는 동일한 압력 용기 또는 별도의 압력 용기에서 수행할 수 있다. 임의의 예비수소화는 별도의 압력 용기 또는 HDO 및 이성체화 단계와 동일한 압력 용기에서 수행할 수 있다.

[0382] 따라서, 하나의 구체예에서, 하나 이상의 화학 반응 생성물은 HRJ-5를 포함하는 알칸 혼합물이다. 다른 구체예에서, 하나 이상의 화학적 반응 생성물은 ASTM D1655 제트 연료를 포함하는 알칸 혼합물이다. 몇몇 구체예에서, ASTM 1655 제트 연료의 사양을 준수하는 조성물은 황 함량이 10ppm 미만이다. 다른 구체예에서, ASTM 1655 제트 연료의 사양을 준수하는 조성물은 205℃ 미만의 증류 곡선의 T10값을 갖는다. 또 다른 구체예에서, ASTM 1655 제트 연료의 사양을 준수하는 조성물은 300℃ 미만의 최종 비점(FBP)을 갖는다. 다른 구체예에서, ASTM 1655 제트 연료의 사양을 준수하는 조성물은 38℃ 이상의 인화점을 갖는다. 다른 구체예에서, ASTM 1655 제트 연료의 사양을 준수하는 조성물은 밀도가 775K/M3 내지 840K/M3이다. 다른 구체예에서, ASTM 1655 제트 연료의 사양을 준수하는 조성물은 동결점이 -47℃ 미만이다. 다른 구체예에서, ASTM 제트 연료의 사양을 준수하는 조성물은 순 연소열(Heat of Combustion)이 42.8MJ/K 이상이다. 다른 구체예에서, ASTM 1655 제트 연료의 사양을 준수하는 조성물은 수소 함량이 13.4질량% 이상이다. 다른 구체예에서, ASTM 1655 제트 연료의 사양을 준수하는 조성물은 Hg의 3mm 미만인, 260℃에서의 정량적 중량 JFTOT에 의하여 시험된 바와 같이, 열 안정성을 갖는다. 다른 구체예에서, ASTM 1655 제트 연료의 사양을 준수하는 조성물은 7mg/dl 미만인 실제 고무를 갖는다.

[0383] 따라서, 본 발명은 미세 조류 지질의 화학적 변형으로 다양한 산업적 및 기타 다른 적용 분야에 유용한 생성물을 생성하는 다양한 방법을 개시한다. 본원에 개시된 방법에 의해 생성된 오일을 변형시키는 방법의 예로서는 오일 가수 분해, 오일의 수소화처리 및 오일의 에스테르화를 포함하지만 이에 제한되지 않는다. 미세 조류 지질의 기타 다른 화학 변형으로서는 에폭시화, 산화, 가수 분해, 황산화, 설폰화, 에톡시화, 프로폭시화, 아미드화 및 비누화를 포함하나, 이에 제한되는 것은 아니다. 미세 조류 오일의 변형은 기본 함유 화학 제품을 생성하고 이는 추가로 원하는 기능을 위해 선택된 유도체 함유 화학 제품으로 변형시킬 수 있다. 연료 생성 공정과 관련하여 전술된 것과 유사한 방식으로, 이들 화학적 변형은 또한 본원에 기술된 미생물 배양물로부터 생성된 오일 상에서 수행할 수 있다. 염기성 함유 화학 제품의 예로서는 비누, 지방산, 지방 에스테르, 지방 알코올, 지방 질소 화합물, 지방산 메틸 에스테르 및 글리세롤을 포함하나, 이에 제한되는 것은 아니다. 유도성 함유 화학 제품의 예로서는 지방 니트릴, 에스테르, 다이머산, 쿼트(quat), 계면 활성제, 지방 알카놀아미드, 지방 알코올 설페이트, 수지, 유화제, 지방 알코올, 올레핀, 시추 이수, 폴리올, 폴리우레탄, 폴리아크릴레이트, 고무, 양초, 화장품, 금속 비누, 비누, 알파-설폰화 메틸 에스테르, 지방 알코올 설페이트, 지방 알코올 에톡실레이트, 지방 알코올 에테르 설페이트, 이미다졸린, 계면 활성제, 세제, 에스테르, 쿼트, 오존 분해 생성물, 지방 아민, 지방 알카놀아미드, 에톡시설페이트, 모노글리세라이드, 디글리세라이드, 트리글리세라이드(예를 들어, 중간쇄 트리글리세라이드), 윤활제, 유압유, 그리스, 유전성 유체, 이형제, 금속 가공 유체, 열 전달 유체, 기타 다른 기능성 유체, 산업용 화학 물질(예를 들어, 클리너, 직물 가공 보조제, 가소제, 안정화제, 첨가제), 표면 코팅제, 페인트 및 래커, 전기 와이어 절연제 및 고급 알칸을 포함하나, 이에 제한되는 것은 아니다.

[0384] 본 발명의 방법에 의해 생성된 글리세로지질 유래의 지방산 성분의 가수 분해는 기타 다른 유용한 화학 물질을 생성하도록 유도체화될 수 있는 유리 지방산을 생성시킨다. 가수 분해는 산 또는 염기일 수 있는 촉매 및 물의 존재 하에 일어난다. 유리된 지방산을 유도체화하여 미국 특허 제5,304,664호(고도로 황화된 지방산); 제7,262,158호(세정 조성물); 제7,115,173호(직물 연화 조성물); 제6,342,208호(피부 처리용 에멀젼); 제7,264,886호(방수 조성물); 제6,924,333호 (페인트 첨가제); 제6,596,768호(지질 풍부 반추동물 공급원료); 및 제6,380,410호(세제 및 세정제용 계면활성제)에 보고된 바와 같은 다양한 생성물을 수득할 수 있다.

[0385] 본 발명의 하나의 구체예에서 가수 분해와 관련하여, 트리글리세라이드 오일은 임의로 먼저 물 또는 수산화나트륨과 같은 액체 매질 중에서 가수 분해하여 글리세롤 및 비누를 수득한다. 다양한 적합한 트리글리세라이드 가수 분해 방법이 있으며, 여기에는 비누화, 산 가수 분해, 알칼리성 가수 분해, 효소 가수 분해(본원에서 분리(splitting)로서 언급됨), 및 고온 압축수를 사용한 가수 분해가 포함되지만 이에 제한되지 않는다. 당업자는 트리글리세라이드 오일이 함유 화학 제품을 생성하기 위해 가수 분해될 필요가 없고, 오히려 상기 오일이 기타 다른 공지된 방법에 의해 원하는 함유 화학 제품으로 직접 전환될 수 있음을 인식할 것이다. 예를 들어, 트리글리세라이드 오일은 에스테르화를 통해 메틸 에스테르 지방산으로 직접 전환될 수 있다.

[0386] 몇몇 구체예에서, 본원에 개시된 방법에 의해 생성된 오일의 촉매 가수 분해는 오일을 글리세롤 및 지방산으로 분리시킴으로써 일어난다. 상기 논의된 바와 같이, 지방산은 이어서 여러 가지 기타 다른 변형을 통해 수득된 유도체 함유 화학 제품으로 추가로 가공될 수 있다. 예를 들어, 하나의 구체예에서, 지방산은 아민화 반응을 거쳐서 지방 질소 화합물을 생성할 수 있다. 다른 구체예에서, 지방산은 일염기성 산 및 이염기성 산을 생성하도록 오존용해를 거칠 수 있다.

[0387] 다른 구체예에서, 가수 분해는 본원에서 생성된 오일 분리를 통해 일어나서 함유 화학 제품을 생성할 수 있다. 본 발명의 몇몇 바람직한 구체예에서, 트리글리세라이드 오일은 기타 다른 공정을 수행하기 전에 분리될 수 있다. 당업자는 많은 적합한 트리글리세라이드 분리 방법, 예를 들어 효소 분리 및 압력 분리(이에 제한되지 않음)가 있음을 인식할 것이다.

[0388] 일반적으로, 오일의 효소 분리 방법은 물/오일 혼합물에 대해 작용하는 생촉매로서 효소인 리파제를 사용한다. 효소 분리는 이어서 오일 또는 지방을 각각 글리세롤 및 유리 지방산으로 분리한다. 이어서 상기 글리세롤은 물상으로 이동하는 반면, 상기 유기상에는 유리 지방산이 풍부해진다.

[0389] 효소 분리 반응은 일반적으로 유기 및 수성 상간의 상경계에서 일어나며, 여기서 효소는 단지 상경계에 존재한다. 이어서 상경계와 만나는 트리글리세라이드는 분리 반응에 기여하거나 이에 관여한다. 반응이 진행됨에 따라, 유리 지방산과는 대조적으로 글리세라이드로서 여전히 화학적으로 결합되어 있는 지방산의 작업 밀도 또는 농도는 상경계에서 감소하여 반응은 느려진다. 특정 구체예에서, 효소 분리는 실온에서 일어날 수 있다. 당업자는 오일을 원하는 지방산으로 분리시키는데 적합한 조건을 알 것이다.

[0390] 예를 들어, 반응 속도는 계면 경계 표면을 증가시킴으로써 가속화될 수 있다. 일단 반응이 완료되면, 유리 지방산을 이어서 효소로부터 유리된 유기상으로부터 분리하고, 글리세라이드로서 여전히 화학적으로 결합되어 있는 지방산을 함유하는 잔류물은 역 공급되거나 재순환되고 새로운 오일 또는 지방과 혼합되어 분리를 거친다. 이어서 상기 방식으로, 재순환된 글리세라이드는 추가의 효소 분리를 거친다. 몇몇 구체예에서, 유리 지방산은 상기 방식으로 부분적으로 분리된 오일 또는 지방으로부터 추출된다. 상기 방식으로, 화학적으로 결합된 지방산(트리글리세라이드)이 분리 공정으로 복귀하거나 역 공급되는 경우, 효소 소비는 급격히 감소할 수 있다.

[0391] 분리 정도는 측정된 산 값을, 소정의 오일 또는 지방에 대해 계산될 수 있는 이론적으로 가능한 산 값으로 나눈 비율로서 결정한다. 바람직하게, 산 값은 표준 일반 방법에 따른 적정을 사용하여 측정한다. 대안적으로, 수성 글리세롤 상의 밀도는 분리 정도에 대한 척도로서 취할 수 있다.

[0392] 하나의 구체예에서, 본원에 기술된 바와 같은 분리 방법은 또한 생성된 오일의 알칼리 정제 공정 유래의 소위 비누-스톡에 함유된 모노글리세라이드, 디글리세라이드 및 트리글리세라이드에 적합하다. 상기 방식으로, 비누스톡은 사전 비누화 없이 중성 오일의 지방산으로 정량적으로 전환될 수 있다. 상기 목적을 위해, 비누에 화학적으로 결합된 지방산을 바람직하게는 분리 전에 산의 부가를 통해 방출시킨다. 특정 구체예에서, 분리 공정을 위해 물 및 효소에 추가로 완충액이 사용된다.

[0393] 하나의 구체예에서, 본 발명의 방법에 따라 생성된 오일은 또한 가수 분해 방법으로서 비누화를 거칠 수 있다. 동물성 및 식물성 오일은 통상적으로 3가 알코올인 글리세롤과 지방산의 에스테르인 트리아실글리세롤(TAG)로 이루어진다. 알칼리성 가수 분해 반응에서, TAG내 글리세롤이 제거되어 나트륨 또는 칼륨과 같은 알칼리 금속

양이온과 연합할 수 있는 3개의 카복실산 음이온을 남겨 두어 지방산 염을 생성한다. 상기 도식에서, 카복실산 성분들은 글리세롤부로부터 절단되고 하이드록실기로 대체된다. 상기 반응에 사용되는 염기(예를 들어, KOH)의 양은 원하는 비누화 정도에 의해 결정된다. 예를 들어, 상기 목적이 본래 TAG 조성물 중에 존재하는 오일의 일부를 포함하는 비누 생성물을 제조하기 위한 것인 경우, 모든 TAG를 지방산 염으로 전환시키기에 불충한 양의 염기가 반응 혼합물로 도입된다. 통상적으로, 상기 반응은 수용액 중에서 수행하여 서서히 진행시키지만 가열에 의해 촉진될 수 있다. 지방산 염의 침전은 수용성 알칼리 금속 할라이드(예를 들어, NaCl 또는 KCl)와 같은 염을 반응 혼합물에 부가하여 촉진될 수 있다. 바람직하게, 상기 염기는 알칼리 금속 수산화물(예를 들어, NaOH 또는 KOH)이다. 대안적으로, 예를 들어 트리에탄올아민 및 아미노메틸프로판올을 포함하는 알칸올아민과 같은 기타 다른 염기를 상기 반응 도식에 사용할 수 있다. 몇몇 경우에, 이들 대안이 투명한 비누 생성물을 제조하는데 바람직할 수 있다. 하나의 구체예에서, 비누화를 거친 지질 조성물은 본원에 기술된 바와 같이 제조된 우지 모의체(즉 우지와 유사한 지질 조성물), 또는 다른 트리글리세라이드 오일과 우지 모의체의 배합물이다.

[0394] 몇몇 방법에서, 화학적 변형의 제1 단계는 수소 및 촉매의 존재 하에서 승온에서 이중 결합을 포화시키기 위한 수소화처리에 이어서 탈산소화일 수 있다. 다른 방법에서, 수소화 및 탈산소화는 동일한 반응 중에서 일어날 수 있다. 또 다른 방법에서, 탈산소화는 수소화전에 일어난다. 이어서 이성체화는 또한 수소 및 촉매의 존재 하에서 임의로 수행할 수 있다. 최종적으로, 가스 및 나프타 성분은 원한다면 제거할 수 있다. 예를 들어, 미국 특허 5,475,160호(트리글리세라이드의 수소화); 제5,091,116호(탈산소화, 수소화 및 가스 제거); 제6,391,815호(수소화); 및 제5,888,947호(이성체화)를 참조한다.

[0395] 본 발명의 몇몇 구체예에서, 트리글리세라이드 오일은 부분적으로 또는 완전히 탈산소화된다. 탈산소화 반응은 지방산, 지방 알코올, 폴리올, 케톤 및 알데하이드(이에 제한되지 않음)를 포함하는 원하는 생성물을 형성시킨다. 일반적으로, 임의의 특정 이론에 구애되지 않고, 탈산소화 반응은 지방산 또는 지방산 에스테르로부터 산소를 적어도 부분적으로 제거하여 추가의 가공에 의해 원하는 화학 물질로 용이하게 전환될 수 있는 지방 알코올과 같은 반응 생성물을 생성시키는 반응으로서, 제한 없이 수소용해, 수소화, 연속 수소화-수소용해, 연속 수소용해-수소화 및 조합된 수소화-수소용해 반응을 포함하는 다양하고 상이한 반응 경로의 조합을 포함한다. 예를 들어, 하나의 구체예에서, 지방 알코올은 FCC 반응을 통해 올레핀으로 또는 응축 반응을 통해 고급 알칸으로 전환될 수 있다.

[0396] 하나의 상기 화학적 변형은 글리세로지질 또는 유리 지방산의 지방산 성분에서 이중 결합으로 수소를 첨가하는 수소화이다. 수소화 과정으로 액체 오일을 특정 응용에 보다 적합할 수 있는 반고체 또는 고체 지방으로 전환하게 할 수 있다.

[0397] 본원에 기술된 방법에 의해 생성된 오일의 수소화는 미국 특허 제7,288,278호(식품 첨가제 또는 의약); 제5,346,724호(윤활 생성물); 제5,475,160호(지방 알코올); 제5,091,116호(식용 오일); 제6,808,737호(마가린 및 스프레드용 구조적 지방); 제5,298,637호(감소된 칼로리 지방 대용물); 제6,391,815호(수소화 촉매 및 황 흡착제); 제5,233,099호 및 제5,233,100호(지방 알코올); 제4,584,139호(수소화 촉매); 제6,057,375호(소포제); 및 제7,118,773호(식용 에멀젼 스프레드)에 보고된 바와 같이 본원에 제공된 방법 및/또는 재료 중 하나 이상과 연계하여 수행할 수 있다.

[0398] 당업자는 탄수화물을 수소화시키기 위해 다양한 방법이 사용될 수 있음을 인지할 것이다. 하나의 적합한 방법은 수소화 반응기에서 수소화된 생성물을 형성하기 위해 충분한 조건 하에서 적합한 가스 및 촉매와 혼합된 수소, 또는 수소와 탄수화물을 접촉시키는 것을 포함한다. 수소화 촉매는 일반적으로 단독으로, 또는 W, Mo, Au, Ag, Cr, Zn, Mn, Sn, B, P, Bi, 및 이의 합금 또는 조합과 같은 프로모터와 함께 Cu, Re, Ni, Fe, Co, Ru, Pd, Rh, Pt, Os, Ir, 및 이의 합금 또는 조합을 포함할 수 있다. 기타 다른 효과적인 수소화 촉매 물질은 지지된 니켈 또는 레늄으로 변형된 루테늄을 포함한다. 하나의 구체예에서, 수소화 촉매는 또한 원하는 촉매의 기능성에 따라, 지지체 중 임의의 하나를 포함한다. 수소화 촉매는 당업자에게 공지된 방법에 의해 제조할 수 있다.

[0399] 몇몇 구체예에서, 수소화 촉매는 지지된 VIII족 금속 촉매 및 금속 스폰지 물질(예를 들어, 스폰지 니켈 촉매)을 포함한다. 라니 니켈은 본 발명에 사용하기에 적합한 활성화된 스폰지 니켈 촉매의 예를 제공한다. 다른 구체예에서, 본 발명의 수소화 반응은 니켈-레늄 촉매 또는 텅스텐 변형 니켈 촉매를 포함하는 촉매를 사용하여 수행한다. 본 발명의 수소화 반응을 위해 적합한 촉매의 예로서는 탄소 지지된 니켈-레늄 촉매가 있다.

[0400] 하나의 구체예에서, 적합한 라니 니켈 촉매는, 예를 들어 약 25중량%의 수산화나트륨을 함유하는 알칼리 수용액으로 대략 균등 중량의 니켈 및 알루미늄의 합금을 처리함으로써 제조할 수 있다. 상기 알루미늄은 선택적으로 알칼리 수용액에 의해 용해되어 알루미늄이 최소량인 대부분 니켈을 포함하는 스폰지형 물질을 생성한다. 초기

합금은 약 1중량% 내지 2중량%가 형성된 스폰지 니켈 촉매 중에 잔류하도록 하는 양으로 프로모터 금속(즉, 몰리브덴 또는 크롬)을 포함한다. 다른 구체예에서, 수소화 촉매는 물 중에서 루테늄(III) 니트로실질산염, 루테늄(III) 클로라이드의 용액을 사용하여 적합한 지지체 물질을 침지시키기 위해 제조한다. 이어서 상기 용액을 건조시켜 수분 함량이 약 1중량% 미만인 고체를 형성한다. 상기 고체는 이어서 4시간 동안 회전볼 로내 300℃(하소되지 않음) 또는 400℃(하소됨)에서 수소 스트림 중에 대기압에서 환원될 수 있다. 촉매를 냉각시키고 질소로 촉매가 불활성이도록 한 후, 질소 중에 5용적%의 산소를 2시간 동안 촉매 상에 통과시킨다.

[0401] 특정 구체예에서, 본원에 기술된 촉매는 촉매 지지체를 포함한다. 상기 촉매 지지체는 촉매를 안정화시키고 지지한다. 사용되는 촉매 지지체의 유형은 선택된 촉매 및 반응 조건에 좌우된다. 본 발명에 적합한 지지체로서는 탄소, 실리카, 실리카-알루미나, 지르코니아, 티타니아, 세리아, 바나디아, 질화물, 질화붕소, 헤테로다중산, 하이드록시애퍼타이트, 산화아연, 크로미아, 제올라이트, 탄소 나노튜브, 탄소 풀러렌, 및 이들의 임의의 조합을 포함하지만 이에 제한되지 않는다.

[0402] 본 발명에 사용되는 촉매는 당업자에게 공지된 통상적인 방법을 사용하여 제조할 수 있다. 적합한 방법은 초기 습윤화, 증발 침지, 화학적 증착, 세척-피복, 마그네트론 스퍼터링 기술 등을 포함할 수 있지만 이에 제한되지 않는다.

[0403] 수소화 반응을 수행하기 위한 조건은 출발 물질 및 원하는 생성물의 유형에 따라 다양할 것이다. 본 개시 내용의 이득과 함께 당업자는 적절한 반응 조건을 인지할 것이다. 일반적으로, 수소화 반응은 80℃ 내지 250℃, 바람직하게는 90℃ 내지 200℃, 가장 바람직하게는 100℃ 내지 150℃의 온도에서 수행한다. 몇몇 구체예에서, 수소화 반응은 500KPa 내지 14000Kpa의 압력에서 수행한다.

[0404] 본 발명의 수소용해 반응에 사용되는 수소는 외부 소수, 재순환된 수소, 동일계 생성 수소 및 이의 임의의 조합을 포함할 수 있다. 본원에 사용된 바와 같이, 용어 “외부 수소”는 바이오매스 반응 자체 기원이 아니며, 오히려 다른 공급원으로부터 시스템에 첨가되는 수소를 언급한다.

[0405] 본 발명의 몇몇 구체예에서, 출발 탄수화물을, 원하는 고급 탄화수소로 보다 용이하게 전환될 것인 소분자로 전환시키는 것이 바람직하다. 상기 전환을 위해 적합한 하나의 방법은 수소용해 반응을 통한 것이다. 탄수화물의 수소용해를 수행하기 위한 다양한 방법이 공지되어 있다. 하나의 적합한 방법은 소분자 또는 폴리올을 포함하는 반응 생성물을 형성하기에 충분한 조건 하에서 수소용해 반응기 내에서 적합한 가스 및 수소용해 촉매와 혼합된 수소, 또는 수소와 탄수화물을 접촉시키는 것을 포함한다. 본원에서, 용어 “소분자 또는 폴리올”은 출발 탄수화물보다 더 적은 양의 탄소 원자 또는 산소 원자를 포함할 수 있는, 소분자량을 갖는 임의의 분자를 포함한다. 하나의 구체예에서, 반응 생성물은 폴리올 및 알코올을 포함하는 보다 작은 분자를 포함한다. 당업자는 수소용해 반응을 수행하기 위한 적절한 방법을 선택할 수 있을 것이다.

[0406] 몇몇 구체예에서, 5개 및/또는 6개 탄소 당 또는 당 알코올은 수소용해 촉매를 사용하여 프로필렌 글리콜, 에틸렌 글리콜 및 글리세롤로 전환시킬 수 있다. 수소용해 촉매는 단독으로 또는 Au, Ag, Cr, Zn, Mn, Sn, Bi, B, O, 및 이의 합금 또는 임의의 조합과 같은 프로모터와 함께 Cr, Mo, W, Re, Mn, Cu, Cd, Fe, Co, Ni, Pt, Pd, Rh, Ru, Ir, Os, 및 이의 합금 또는 조합을 포함할 수 있다. 수소용해 촉매는 또한 전이 금속(예를 들어, 크롬, 몰리브덴, 텅스텐, 레늄, 망간, 구리, 카드뮴) 또는 VIII족 금속(예를 들어, 철, 코발트, 니켈, 백금, 팔라듐, 로듐, 루테늄, 이리듐 및 오스뮴)을 함유하는 탄소성 피로중합체 촉매를 포함할 수 있다. 특정 구체예에서, 수소용해 촉매는 알칼리성 토금속 산화물과 조합되거나 촉매 활성 지지체에 부착된 임의의 상기 금속을 포함할 수 있다. 특정 구체예에서, 수소용해 반응에 기술된 촉매는 수소화 반응을 위해 전술된 바와 같은 촉매 지지체를 포함할 수 있다.

[0407] 수소용해 반응을 수행하기 위한 조건은 출발 물질 및 원하는 생성물의 유형에 따라 다양할 것이다. 당업자는 본 개시 내용의 이득과 함께 반응을 수행하기 위해 사용하는 적절한 조건을 인지할 것이다. 일반적으로, 이들 수소용해 반응은 110℃ 내지 300℃, 바람직하게는 170℃ 내지 220℃, 가장 바람직하게는 200℃ 내지 225℃의 온도에서 수행한다. 몇몇 구체예에서, 수소용해 반응은 염기성 조건 하, 바람직하게는 pH 8 내지 13, 훨씬 더 바람직하게는 pH 10 내지 12에서 수행한다. 몇몇 구체예에서, 수소용해 반응은 60KPa 내지 16500Kpa의 범위, 바람직하게는 1700KPa 내지 14000KPa의 범위, 훨씬 더 바람직하게는 4800KPa 내지 11000KPa의 압력에서 수행한다.

[0408] 본 발명의 수소용해 반응에 사용되는 수소는 외부 수소, 재순환된 수소, 동일계 생성된 수소 및 이의 임의의 조합을 포함할 수 있다.

[0409] 몇몇 구체예에서, 상기 논의된 반응 생성물은 응축 반응기에서 응축 반응을 통해 고급 탄화수소로 전환시킬 수

있다. 상기 구체예에서, 반응 생성물의 응축은 고급 탄화수소를 형성할 수 있는 촉매의 존재 하에서 일어난다. 이론에 의해 제한되고자 하는 것은 아니지만, 고급 탄화수소의 생산은 탄소-탄소 또는 탄소-산소 결합의 형성을 포함하는 단계적 부가 반응을 통해 진행하는 것으로 여겨진다. 생성된 반응 생성물은 하기에서 보다 상세하게 기술된 바와 같이 상기 부를 함유하는 임의의 수의 화합물을 포함한다.

[0410] 특정 구체예에서, 적합한 응축 촉매로서는 산 촉매, 염기 촉매 또는 산/염기 촉매를 포함한다. 본원에 사용된 바와 같이, 용어 “산/염기 촉매”는 산 및 염기 기능 둘 다를 갖는 촉매를 나타낸다. 몇몇 구체예에서, 상기 응축 촉매는 제한없이, 제올라이트, 카바이드, 질화물, 지르코니아, 알루미나, 실리카, 알루미노실리케이트, 포스페이트, 산화티탄, 산화아연, 산화바나듐, 산화란탄, 산화이트륨, 산화스캔듐, 산화마그네슘, 산화세륨, 산화바륨, 산화칼슘, 수산화물, 헤테로다중산, 무기산, 산 변형된 수지, 염기 변형된 수지, 및 이들의 임의의 조합을 포함할 수 있다. 몇몇 구체예에서, 응축 촉매는 또한 변형제를 포함할 수 있다. 적합한 변형제는 La, Y, Sc, P, B, Bi, Li, Na, K, Rb, Cs, Mg, Ca, Sr, Ba, 및 이들의 임의의 조합을 포함한다. 몇몇 구체예에서, 상기 응축 촉매는 또한 금속을 포함할 수 있다. 적합한 금속은 Cu, Ag, Au, Pt, Ni, Fe, Co, Ru, Zn, Cd, Ga, In, Rh, Pd, Ir, Re, Mn, Cr, Mo, W, Sn, Os, 이들의 합금 및 임의의 조합을 포함한다.

[0411] 특정 구체예에서, 응축 반응에 기술된 촉매는 수소화 반응을 위해 전술된 바와 같은 촉매 지지체를 포함할 수 있다. 특정 구체예에서, 응축 촉매는 자가 지지형이다. 본원에 사용된 바와 같이, 상기 용어 “자가 지지형”은 촉매가 지지체로서 작용하기 위해 다른 물질을 필요로 하지 않음을 의미한다. 다른 구체예에서, 상기 응축 촉매는 촉매를 현탁시키기에 적합한 별도의 지지체와 연계하여 사용된다. 하나의 구체예에서, 응축 촉매 지지체는 실리카이다.

[0412] 응축 반응이 일어나는 조건은 출발 물질 및 원하는 생성물의 유형에 따라 다양할 것이다. 당업자는 본 개시 내용의 이득과 함께 반응을 수행하는데 사용하는 적절한 조건을 인지할 것이다. 몇몇 구체예에서, 응축 반응은 제안된 반응에 대한 열역학을 유리하게 하는 온도에서 수행한다. 응축 반응을 위한 온도는 특정 출발 폴리올 또는 알코올에 따라 다양할 것이다. 몇몇 구체예에서, 응축 반응을 위한 온도는 80℃ 내지 500℃, 바람직하게는 125℃ 내지 450℃, 가장 바람직하게는 125℃ 내지 250℃의 범위이다. 몇몇 구체예에서, 응축 반응은 0Kpa 내지 9000KPa 범위, 바람직하게는 0KPa 내지 7000KPa 범위, 훨씬 더 바람직하게는 0KPa 내지 5000KPa 범위의 압력에서 수행한다.

[0413] 본 발명에 의해 형성되는 고급 알칸은 4개 내지 30개의 탄소 원자를 갖는 분지쇄 또는 직쇄 알칸, 4개 내지 30개의 탄소 원자를 갖는 분지쇄 또는 직쇄 알켄, 5개 내지 30개의 탄소 원자를 갖는 사이클로알칸, 5개 내지 30개의 탄소 원자를 갖는 사이클로알켄, 아릴, 융합된 아릴, 알코올 및 케톤을 포함하지만 이에 제한되지 않는다. 적합한 알칸은 부탄, 펜탄, 펜텐, 2-메틸부탄, 헥산, 헥센, 2-메틸펜탄, 3-메틸펜탄, 2,2-디메틸부탄, 2,3-디메틸부탄, 헵탄, 헵텐, 옥탄, 옥텐, 2,2,4-트리메틸펜탄, 2,3-디메틸헥산, 2,3,4-트리메틸펜탄, 2,3-디메틸펜탄, 노난, 노넨, 데칸, 데센, 운데칸, 운데센, 도데칸, 도데센, 트리데칸, 트리데센, 테트라데칸, 테트라데센, 펜타데칸, 펜타데센, 노닐데칸, 노닐데센, 에이코산, 에이코센, 운에이코산, 운에이코센, 도에이코산, 도에이코센, 트리에이코산, 트리에이코센, 테트라에이코산, 테트라에이코센 및 이들의 이성체를 포함하지만 이에 제한되지 않는다. 이들 생성물중 일부는 연료로서 사용하기에 적합할 수 있다.

[0414] 몇몇 구체예에서, 사이클로알칸 및 사이클로알켄은 비치환된다. 다른 구체예에서, 사이클로알칸 및 사이클로알켄은 1치환된다. 또 다른 구체예에서, 사이클로알칸 및 사이클로알켄은 다치환된다. 치환된 사이클로알칸 및 사이클로알켄을 포함하는 구체예에서, 치환된 기는 제한없이, 1개 내지 12개의 탄소 원자를 갖는 분지쇄 또는 직쇄 알킬, 1개 내지 12개의 탄소 원자를 갖는 분지쇄 또는 직쇄 알킬렌, 페닐 및 이의 임의의 조합을 포함한다. 적합한 사이클로알칸 및 사이클로알켄은 사이클로펜탄, 사이클로펜텐, 사이클로헥산, 사이클로헥센, 메틸-사이클로펜탄, 메틸-사이클로펜텐, 에틸-사이클로펜탄, 에틸-사이클로펜텐, 에틸-사이클로헥산, 에틸-사이클로헥센, 이들의 이성체 및 임의의 조합을 포함하지만 이에 제한되지 않는다.

[0415] 몇몇 구체예에서, 형성된 아릴은 비치환된다. 다른 구체예에서, 형성된 아릴은 1치환된다. 치환된 아릴을 포함하는 구체예에서, 치환된 기는 제한없이 1개 내지 12개의 탄소 원자를 갖는 분지쇄 또는 직쇄 알킬, 1개 내지 12개의 탄소 원자를 갖는 분지쇄 또는 직쇄의 알킬렌, 페닐, 및 이들의 임의의 조합을 포함한다. 본 발명에 적합한 아릴로서는 벤젠, 톨루엔, 크실렌, 에틸 벤젠, 파라크실렌, 메타크실렌 및 이들의 임의의 조합을 포함한다.

[0416] 본 발명에서 생성된 알코올은 4개 내지 30개의 탄소 원자를 갖는다. 몇몇 구체예에서, 알코올은 환형이다. 다른 구체예에서, 알코올은 분지형다. 다른 구체예에서, 알코올은 직쇄이다. 본 발명에 적합한 알코올로서는 부탄올,

펜탄올, 헥산올, 헵탄올, 옥탄올, 노난올, 데칸올, 운데칸올, 도데칸올, 트리데칸올, 테트라데칸올, 펜타데칸올, 헥사데칸올, 헵틸데칸올, 옥틸데칸올, 노닐데칸올, 에이코산올, 운에이코산올, 도에이코산올, 트리에이코산올, 테트라에이코산올 및 이들의 이성체를 포함하지만 이에 제한되지 않는다.

[0417] 본 발명에서 생성된 케톤은 4개 내지 30개의 탄소 원자를 갖는다. 하나의 구체예에서, 케톤은 환형이다. 다른 구체예에서, 케톤은 분지형이다. 다른 구체예에서, 케톤은 직쇄이다. 본 발명에 적합한 케톤은 부탄온, 펜탄온, 헥산온, 헵탄온, 옥탄온, 노난온, 데칸온, 운데칸온, 도데칸온, 트리데칸온, 테트라데칸온, 펜타데칸온, 헥사데칸온, 헵틸데칸온, 옥틸데칸온, 노닐데칸온, 에이코산온, 운에이코산온, 도에이코산온, 트리에이코산온, 테트라에이코산온, 및 이들의 이성체를 포함하지만 이에 제한되지 않는다.

[0418] 다른 이러한 화학적 변형은 에스테르 교환이다. 자연적으로 생성된 글리세로지질은 지방산 성분의 균일한 분포를 갖지 않는다. 오일과 관련하여, 에스테르 교환은 상이한 글리세로지질의 2개의 에스테르간의 아실 라디칼의 교환을 언급한다. 에스테르 교환 공정은 글리세로지질의 혼합물의 지방산 성분이 재배열되어 분포 패턴을 변형시킬 수 있는 기작을 제공한다. 에스테르 교환은 널리 공지된 화학적 공정이고, 일반적으로 알칼리 금속 또는 알칼리 금속 알킬레이트(예를 들어, 나트륨 메톡사이드)와 같은 촉매의 존재 하에서 일정 시간(예를 들어, 30분) 동안 오일의 혼합물을 가열(약 200℃로)하는 것을 포함한다. 이러한 공정을 사용하여 오일 혼합물의 지방산 성분의 분포 패턴을 무작위화할 수 있거나, 원하는 분포 패턴을 생성하도록 지시될 수 있다. 지질의 화학 변형에 대한 상기 방법은, 지질로서 건조 세포 중량%가 20% 이상인 미생물 바이오매스와 같은 본원에 제공된 물질에 대해 수행할 수 있다.

[0419] 지방산의 특정 분포 패턴이 요구되는 지시된 에스테르 교환은 일어날 수 있는 몇몇 TAG의 융점 미만의 온도에서 오일 혼합물을 유지시킴으로써 수행할 수 있다. 이것은 이들 TAG를 선택적으로 결정화하며, 이는 이들이 결정화함에 따라 반응 혼합물로부터 이들을 효과적으로 제거한다. 상기 방법은, 예를 들어 오일 중 대부분의 지방산이 침전할 때까지 계속할 수 있다. 지시된 에스테르 교환 방법은 예를 들어 장쇄 지방산의 단쇄 대응물로의 치환을 통해 보다 낮은 칼로리 함량을 갖는 생성물을 제조하기 위해 사용될 수 있다. 지시된 에스테르 교환은 또한 원치않는 트랜스 이성체를 생성시킬 수 있는 수소화에 의지하는 것 없이 식품 부가물 또는 제품(예를 들어, 마가린)에서 요구되는 원하는 용융 특성 및 구조적 특징을 제공할 수 있는 지방 혼합물을 갖는 생성물을 제조하기 위해 사용될 수 있다.

[0420] 본원에 기술된 방법에 의해 생성된 오일의 에스테르 교환은 미국 특허 제6,080,853호(비분해성 지방 대용물); 제4,288,378호(땅콩 버터 안정화제); 제5,391,383호(식용 스프레이 오일); 제6,022,577호(식품 생성물에 대한 식용 지방); 제5,434,278호(식품 생성물에 대한 식용 지방);제 5,268,192호(저칼로리 너트 생성물); 제5,258,197호(감소된 칼로리의 식용 조성물); 제4,335,156호(식용 지방 생성물); 제7,288,278호(식품 부가물 또는 의약); 제7,115,760호(분별 방법); 제6,808,737호(구조적 지방); 제5,888,947호(엔진 윤활제); 제5,686,131호(식용 오일 혼합물); 및 제4,603,188호(경화성 우레탄 조성물)에 보고된 바와 같이 방법 및/또는 물질 중 하나 이상과 함께 또는 생성물을 생성하기 위해 수행될 수 있다.

[0421] 본 발명에 따른 하나의 구체예에서, 상기된 바와 같은 오일의 에스테르 결합 전이에 이어서 미국 특허 제6,465,642호에 보고된 바와 같이 에스테르 결합 전이된 생성물을 폴리올과 반응시켜 폴리올 지방산 폴리에스테르를 생성시킨다. 상기 에스테르화 및 분리 공정은, 저급 알킬 에스테르를 비누의 존재 하에 폴리올과 반응시키는 단계; 생성물 혼합물로부터 잔여 비누를 제거하는 단계; 생성물 혼합물을 물 세척 및 건조시켜 불순물을 제거하는 단계; 정제를 위해 생성물 혼합물을 표백시키는 단계; 생성물 혼합물중에 폴리올 지방산 폴리에스테르로부터 미반응된 저급 알킬 에스테르의 적어도 일부를 분리하는 단계; 및 분리된 미반응된 저급 알킬 에스테르를 재순환하는 단계를 포함할 수 있다.

[0422] 에스테르 결합 전이는 또한 미국 특허 제6,278,006호에 보고된 바와 같은 단쇄 지방산 에스테르를 갖는 미생물 바이오매스 상에서 수행할 수 있다. 일반적으로, 에스테르 결합 전이는 단쇄 지방산 에스테르를 적합한 촉매의 존재 하에서 오일에 부가하고 혼합물을 가열시킴으로써 수행할 수 있다. 몇몇 구체예에서, 상기 오일은 반응 혼합물의 약 5중량% 내지 약 90중량%를 포함한다. 몇몇 구체예에서, 단쇄 지방산 에스테르는 반응 혼합물의 약 10중량% 내지 약 50중량%일 수 있다. 촉매의 비제한적인 예로서는 염기 촉매; 나트륨 메톡사이드; 산 촉매, 예를 들어 무기산(예를 들어, 황산 및 산성화된 점토), 유기산(예를 들어, 메탄 설폰산, 벤젠설폰산 및 톨루엔설폰산), 및 산성 수지(예를 들어, Amberlyst 15)를 포함한다. 나트륨 및 마그네슘과 같은 금속 및 금속 수소화물은 또한 유용한 촉매이다.

[0423] 다른 당해 화학 변형은 하이드록실화이며, 이는 물을 이중 결합에 부가하여 하이드록실부의 포화 및 혼입을 가

져온다. 상기 하이드록실화 공정은 글리세로지질의 하나 이상의 지방산 성분을 하이드록시 지방산으로 전환시키기 위한 기작을 제공한다. 하이드록실화는, 예를 들어 미국 특허 제5,576,027호에 보고된 방법을 통해 수행할 수 있다. 피마자유를 포함하는 하이드록실화된 지방산 및 이의 유도체는 식품 첨가제, 계면활성제, 안료 습윤제, 소포제, 방수 첨가제, 가소화제, 화장품 유화제 및/또는 탈취제를 포함하는 여러 가지의 산업 응용 분야, 및 전자기기, 약제, 페인트, 잉크, 접착제 및 윤활제에서의 성분으로서 유용하다. 글리세라이드의 하이드록실화를 수행할 수 있는 방법의 한가지 예는 다음과 같다: 지방은 바람직하게 헵탄과 조합하여 약 30℃ 내지 50℃로 가열하고 30분 이상 동안 온도에 유지시킬 수 있으며; 이어서 아세트산을 혼합물에 첨가하고 황산 수용액, 그 다음 과산화수소 수용액을 소량의 증분으로 1시간에 걸쳐서 혼합물에 첨가하며; 수성 과산화수소 후에 온도를 이어서 약 60℃ 이상으로 증가시키고 6시간 이상 동안 교반시킬 수 있으며; 교반 후 상기 혼합물이 침강할수 있게 하고 반응에 의해 형성된 저급 수성 층을 제거하는 한편, 반응에 의해 형성된 상부 헵탄층을 약 60℃의 온도를 갖는 고온수로 세척할 수 있고; 세척된 헵탄 층을 이어서 수산화칼륨 수용액을 사용하여 pH 약 5 내지 7로 중화시키고 이어서 진공 증류에 의해 제거할 수 있으며; 반응 생성물을 이어서 100℃에서 진공 건조시킬 수 있고 건조된 생성물을 진공 조건 하에 증기 탈취시키고 규조토를 사용하여 약 50℃ 내지 60℃에서 여과할 수 있다.

[0424] 본원에 기술된 방법에 의해 생성된 미생물 오일의 하이드록실화는 미국 특허 제6,590,113호(오일계 피복제 및 잉크); 제4,049,724호(하이드록실화 공정); 제6,113,971호(올리브유 버터); 제4,992,189호(윤활제 및 루브 첨가제); 제5,576,027호(하이드록실화된 밀크); 및 제6,869,597호(화장품)에 보고된 바와 같이 방법 및/또는 물질 중 하나 이상과 함께 또는 생성물을 생성하기 위해 수행될 수 있다. 리시놀레산의 하이드록실화는 폴리올을 제공한다.

[0425] 하이드록실화된 글리세로지질은 에스톨리드로 전환시킬 수 있다. 에스톨리드는 글리세로지질로 이루어지고, 여기서 하이드록실화된 지방산 성분은 다른 지방산 분자에 에스테르화시켰다. 하이드록실화된 글리세로지질의 에스톨리드로의 전환은 글리세로지질 및 지방산의 혼합물을 가온시키고 상기 혼합물을 문헌[Isbell et al., JAOCS 71(2): 169-174 (1994)]에 기술된 바와 같이 무기산과 접촉시킴으로써 수행할 수 있다. 에스톨리드는 미국 특허 제7,196,124호(탄성 물질 및 바닥덮개); 제5,458,795호(고온 응용을 위한 증점 오일); 제5,451,332호(산업적 응용을 위한 유체); 제5,427,704호(연료 첨가제); 및 제5,380,894호(윤활제, 그리스, 가소제 및 인쇄 잉크)(이에 제한되는 것은 아님)에 보고된 것들을 포함하는, 다양한 응용 분야에서 유용하다.

[0426] 이와 같은 화학 변형의 다른 예로서는 올레핀 복분해가 있다. 올레핀 복분해에서, 촉매는 알켄(올레핀) 내에 있는 알킬리덴 탄소를 분리하며, 이와 같이 분리된 탄소들 각각과 상이한 알킬리덴 탄소들 간 쌍을 이룸으로써 새로운 알켄을 형성한다. 올레핀 복분해 반응은, 에틸렌을 사용한 분해(ethenolysis), 자가 복분해에 의한(알켄 결합을 통한) 지방산의 가교, 및 교차 복분해에 의하여 유도체화된 알켄과 함께 신규 작용기를 지방산에 통합시킴으로써, 알켄에 존재하는 불포화 지방산 알킬 사슬을 절두하는 것과 같은 공정 기작을 제공한다.

[0427] 기타 다른 반응, 예를 들어 에스테르 결합 전이 및 수소화와 관련하여, 올레핀 복분해는 불포화 글리세로지질을 다양한 최종 생성물로 전환할 수 있다. 이러한 생성물은 왁스용 글리세로지질 올리고머; 윤활제용 단쇄 글리세로지질; 화학 물질 및 중합체용 동종이작용성 및 이종이작용성 알킬 사슬; 생물 연료용 단쇄 에스테르; 및 제트 연료용 단쇄 탄화수소를 포함한다. 올레핀 복분해는, 예를 들어 미국 특허 제7,119,216호, 미국 특허 공개 제2010/0160506호 및 미국 특허 공개 제2010/0145086호에 보고된 방법 및 촉매를 사용하여 트리아실글리세롤과 지방산 유도체 상에서 수행될 수 있다.

[0428] 바이오오일(bio-oil)의 올레핀 복분해는 일반적으로 기타 다른 알켄의 존재 하(교차 복분해) 또는 부재 하(자가 복분해)에, 불활성 조건 하에서 Ru 촉매 용액을 약 10ppm 내지 250ppm의 로딩율로 불포화 지방산 에스테르에 첨가하는 것을 포함한다. 반응은 통상적으로 수 시간에서 수 일에 이르기까지 진행될 수 있으며, 궁극적으로는 알켄 생성물을 분배한다. 올레핀 복분해가 지방산 유도체 상에서 어떻게 수행될 수 있는지에 관한 일례는 다음과 같다: 톨루엔 중 제1 세대 그럽스 촉매(Grubbs Catalyst)(디클로로[2(1-메틸에톡시-α-O)페닐]메틸렌-α-C)(트리사이클로헥실-포스핀) 용액(촉매 로딩율 = 222ppm)은 탈기 및 무수 메틸 올리에이트가 담긴 용기에 첨가될 수 있다. 그 다음, 상기 용기는 약 60psig의 에틸렌 가스로 가압된 후, 약 30℃ 이하의 온도에서 3시간 동안 유지될 수 있는데, 이로써 대략 수율 50%의 메틸 9-데세노에이트가 생성될 수 있다.

[0429] 본원에 기술된 방법에 의해 생산된 오일의 올레핀 복분해는 특허 출원 PCT/US07/081427호(α-올레핀 지방산) 및 미국 특허 출원 제12/281,938호(바셀린), 제12/281,931호(페인트볼 건 캡슐), 제12/653,742호(가소제 및 윤활제), 제12/422,096호(이작용성 유기 화합물) 및 제11/795,052호(양초)에 보고된 바와 같은 방법 및/또는 물질

중 하나 이상과 함께 또는 생성물을 생성하기 위해 수행될 수 있다.

[0430] 미생물 오일상에서 수행될 수 있는 기타 다른 화학적 반응은 미국 특허 제6,051,539호에 보고된 바와 같이 트리아실글리세롤과 사이클로프로판화제를 반응시켜 유동성 및/또는 산화 안정성을 증진시키는는 것; 미국 특허 제6,770,104호에 보고된 바와 같이 트리아실글리세롤로부터 왁스를 제조하는 것; 및 문헌[“The effect of fatty acid composition on the acrylation kinetics of epoxidized triacylglycerols”, Journal of the American Oil Chemists' Society, 79:1, 59-63, (2001)] 및 [Free Radical Biology and Medicine, 37:1, 104-114 (2004)]에 보고된 바와 같이 트리아실글리세롤을 에폭시화시키는 것을 포함한다.

[0431] 전술된 바와 같이 연료 및 화학적 생성물에 대한 오일 함유 미생물 바이오매스의 생성은 탈지된 바이오매스 밀(meal)을 생성시킨다. 탈지된 밀은 조류 오일을 제조하는데 있어서의 부산물이고, 농장 동물, 예를 들어 반추동물, 가금류, 돼지 및 양식물고기에 대한 동물 사료로서 유용하다. 생성된 밀은 오일 함량이 감소하였지만 여전히 고품질의 단백질, 탄수화물, 섬유, 재, 잔류 오일 및 동물 사료에 적당한 기타 다른 영양물을 함유한다. 세포는 주로 오일 분리 공정에 의해 용해되기 때문에, 탈지된 밀은 상기 동물에 의해 쉽게 소화될 수 있다. 탈지된 밀은 임의로 동물 사료에서 곡물과 같은 기타 다른 성분과 배합될 수 있다. 탈지된 밀은 분말성 점도를 갖기 때문에, 이것은 압출기 또는 확장기 또는 시판되는 다른 유형의 기계를 사용하여 펠렛으로 압축될 수 있다.

[0432] 피마자 오일은 아주까리로부터 분리되는, 자연적으로 생성되는 오일이다. 피마자 오일의 가수 분해로 리시놀레산이 생성된다. 아주까리는 다량의 리신을 함유하므로, 아주까리로부터의 피마자 오일의 생산은 어렵다. 리신은 화학 무기 금지 협약(Chemical Weapons Convention) 중 스케쥴 1 화합물로서 나열된 매우 위험한 독소이다. 그러므로, 아주까리로부터 피마자유를 생산할 때에는 세심한 주의를 기울여야 한다. 미세 조류 세포로부터 분리된 하이드록실화 오일은 본 발명의 하나의 구체예에 의해 제공된다. 이러한 방식으로, 리시놀레산이 생산될 수 있다. 하나의 구체예에서, 하이드록실화된 오일은 하이드록실화된 트리글리세라이드이다. 본 발명의 하이드록실화된 트리글리세라이드는 피마자 오일과 화학적으로 유사할 수 있다. 실시예 7에 나타낸 바와 같이, 본 발명은 하이드록실화된 미생물 오일을 제공한다. 실시예 7의 오일의 GC/MS 분석 결과는, 본 발명자들이 리시놀레산(12-하이드록시-9-cis-옥타데센산)을 제조하였음을 나타낸다.

[0433] 본 발명의 하나의 구체예에 따른 지방산은 하이드록실화된 지방산이다. 하이드록실화된 지방산의 하나의 구체예는 리시놀레산이다.

[0434] 미생물의 하이드록실화 오일 또는 하이드록실화 지방산은 추가로 하이드록실화될 수 있다. 리시놀레산이 추가로 하이드록실화될 때, 2개의 하이드록실기를 함유하는 지방산, 폴리올이 제공된다.

[0435] 본 발명은 폴리올(예를 들어, 하이드록실화된 오일 및/또는 하이드록실화된 지방산)과, 이소시아네이트부를 함유하는 화합물을 반응시킴으로써 제조된 조성물을 제공한다. 피마자 오일과 이소시아네이트를 사용하여 폴리우레탄이 생산되었다. 폴리우레탄은 오늘날 사용되는 제품에 두루 존재한다. 폴리우레탄은 자동차, 장난감, 운동기구, 소비자용 전자 제품, 신발, 매트리스, 쿠션, 접착제, 건축 자재 등에서 발견된다. 현재, 피마자유로 제조된 폴리우레탄이 BASF, Itoh Oil 등으로부터 시판되고 있다. 하이드록실화된 대두 오일로 제조된 폴리우레탄은 Cargil, Dow, Bayer 등으로부터 시판되고 있다.

[0436] 하나의 구체예에서, 미생물 세포에 의해 생산된 리시놀레산은, 당업계에 공지된 방법에 따라서, 함유 화학 제품, 예를 들어 리시놀레 에스테르, 리시놀레 아미드, 폴리우레탄, 폴리우레탄 폼 또는 폴리우레탄 부품으로 추가로 가공될 수 있다. 예를 들어 미국 특허 제6194475호, 제4266617호, 제6403664호 및 제4058492호, 그리고 미국 특허 출원 제20100227151호를 참조한다.

[0437] 상기에서 상세하게 기술된 본 발명은 하기의 실시예에 예시되지만, 이는 청구된 발명을 설명하기 위해 제공될 뿐 이에 제한되지 않는다.

[0438] **VII. 실시예**

[0439] **실시예 1: 프로토테카를 배양하는 방법**

[0440] 프로토테카 균주는 건조 세포 중량을 기준으로 높은 백분율의 오일을 성취하기 위해 배양하였다. 동결보존된 세포를 실온에서 해동시키고 500㎕의 세포를 4.5ml의 배지(4.2g/L K2HPO4, 3.1g/L NaH2PO4, 0.24g/L MgSO4·7H2O, 0.25g/L 시트르산 1수화물, 0.025g/L CaCl2 2H2O, 2g/L 효모 추출물) + 2% 글루코스에 첨가하고 6웰 평판에서 교반(200rpm)하면서 7일 동안 28℃에서 생장시켰다. 건조 세포 중량은 1ml의 배양물을 예비 칭량된 에펜도르프 튜브에서 5분 동안 14,000rpm에서 원심분리하여 측정하였다. 배양 상청액을 버리고 수득한 세포 펠렛을

1ml의 탈이온수로 세척하였다. 배양물을 다시 원심분리하고, 상청액을 버리며, 세포 펠렛을 동결될 때까지 -80℃에 두었다. 이어서 샘플을 24시간 동안 동결건조시키고 건조 세포 중량을 계산하였다. 배양물 중 총 지질을 측정하기 위해, 3ml의 배양물을 제거하고 제조업자의 프로토콜에 따라 Ankom 시스템(Ankom Inc., Macedon, NY)을 사용하여 분석하였다. 샘플은 제조업자의 프로토콜에 따라 Amkom XT10 추출기로 용매 추출을 하였다. 총 지질은 산 가수 분해된 건조 샘플 및 용매 추출된 건조 샘플간의 질량 차이로서 측정하였다. 오일 건조 세포 중량의 백분율 측정은 표 10에 나타내어져 있다.

표 10

[0441] 건조 세포 중량을 기준으로 한 오일 백분율

종	균주	% 오일
프로토테카 스태그노라	UTEX 327	13.14
프로토테카 모리포르미스	UTEX 1441	18.02
프로토테카 모리포르미스	UTEX 1435	27.17

[0442] 프로토테카 속의 다수의 균주들로부터 얻은 미세 조류 샘플의 유전자형을 분석하였다. 게놈 DNA를 다음과 같이 조류 바이오매스로부터 분리하였다. 세포(대략 200mg)를 14,000 x g에서 5분 동안 액체 배양물로부터 원심분리하였다. 세포를 이어서 멸균 증류수에 재현탁시키며, 14,000 x g에서 5분 동안 원심분리하고 상청액을 버렸다. 직경이 약 2mm인 단일 유리 비드를 바이오매스에 첨가하고 튜브를 15분 이상 동안 -80℃에 두었다. 샘플을 제거하고 150μl의 분쇄 완충액(1% 사르코실(Sarkosyl), 0.25M 수크로스, 50mM NaCl, 20mM EDTA, 100mM Tris-HCl, pH 8.0, RNase A 0.5ug/ul)을 첨가하였다. 펠렛을 간단히 와류시켜 재현탁시키고, 이어서 40㎕의 5M NaCl을 첨가하였다. 샘플을 간단히 와류시키고, 이어서 66μl의 5% CTAB(세틸 트리메틸암모늄 브로마이드)를 첨가하며, 최종적으로 간단히 와류시켰다. 다음에 샘플을 10분 동안 65℃에서 항온처리하고, 이후 이들을 10분 동안 14,000 x g에서 원심분리하였다. 상기 상청액을 새로운 튜브로 옮기고, 300μl의 페놀:클로로포름:이소아밀 알코올 12:12:1로 1회 추출함에 이어서 14,000 x g에서 5분 동안 원심분리하였다. 생성된 수성 상을 0.7용적의 이소프로판올(약 190㎕)을 함유하는 새로운 튜브로 옮기고, 뒤집어서 혼합하며, 30분 동안 실온에서 또는 밤새 4℃에서 항온처리하였다. DNA를 10분 동안 14,000 x g에서 원심분리를 통해 회수하였다. 이어서 생성된 펠렛을 70% 에탄올로 2회 세척함에 이어서 100% 에탄올로 최종 세척하였다. 펠렛을 실온에서 20분 내지 30분 동안 공기 건조시킴에 이어서 50μl의 10mM TrisCl, 1mM EDTA(pH 8.0) 중에 재현탁시켰다.

[0443] 상기된 바와 같이 제조된 5μl의 총 조류 DNA를 10mM Tris(pH 8.0) 중에서 1:50으로 희석시켰다. 최종 용적 20㎕의 PCR 반응을 다음과 같이 준비하였다. 10μl의 2 x iProof HF 마스터 믹스(BIO-RAD)를 0.4μl의 프라이머 SZ02613 (10mM 스톡 농도에서 5'-TGTTGAAGAATGAGCCGGCGAC-3'(서열번호 9))에 첨가하였다. 상기 프라이머 서열은 GenBank 수탁 번호 제L43357호에서 567번 내지 588번이고, 고등 식물 및 조류 색소체 게놈에서 고도로 보존되어 있다. 이어서 0.4μl의 프라이머 SZ02615(10mM 스톡 농도에서 5'-CAGTGAGCTATTACGCACTC-3'(서열번호 10))를 첨가하였다. 상기 프라이머 서열은 GenBank 수탁 번호 제L43357호에서 1112번 내지 1093번 위치에 상보적이고, 고등 식물 및 조류 색소체 게놈에서 고도로 보존되어 있다. 다음에, 5μl의 희석된 총 DNA 및 3.2μl의 dH2O를 첨가하였다. PCR 반응을, 35사이클 동안 98℃, 45초; 98℃, 8초; 53℃, 12초; 72℃, 20초에 이어서 1분 동안 72℃에서 수행하며, 25℃에 유지시켜 수행한다. PCR 생성물의 정제를 위해, 20μl의 10mM Tris(pH 8.0)를 각각의 반응에 첨가하고, 이어서 40μl의 페놀:클로로포름:이소아밀 알코올 12:12:1로 추출하며, 와류시키고 5분 동안 14,000 x g에서 원심분리하였다. PCR 반응을 S-400 칼럼(GE Healthcare)에 적용하고 3,000 x g에서 2분 동안 원심분리하였다. 정제된 PCR 생성물을 이어서 PCR8/GW/TOPO로 TOPO 클로닝하고, 양성 클론을 LB/Spec 평판을 위해 선택하였다. 정제된 플라스미드 DNA를 M13 정방향 및 역방향 프라이머를 사용하여 양 방향으로 서열분석하였다. 전체적으로, 12개의 프로토테카 균주는 서열분석된 이들의 23S rRNA DNA를 갖도록 선택하고, 상기 서열을 서열 목록 중에 열거하였다. 균주 및 서열 번호에 대한 요약은 하기에 포함된다. 상기 서열은 UTEX 1435(서열번호 15) 서열로부터 총 분기점에 대해 분석하였다. 2개의 쌍(UTEX 329/UTEX 1533 및 UTEX 329/UTEX 1440)이 가장 분기된 것으로서 나타났다. 양 경우에, 쌍형성 정렬은 75.0% 쌍형성 서열 동일성을 나타내었다. UTEX 1435에 대한 서열 동일성 백분율은 또한 하기에 포함된다.

종	균주	% nt 동일성	서열번호
프로토테카 크루에가니	UTEX 329	75.2	서열번호 11
프로토테카 윅커해미	UTEX 1440	99	서열번호 12
프로토테카 스태그노라	UTEX 1442	75.7	서열번호 13
프로토테카 모리포르미스	UTEX288	75.4	서열번호 14
프로토테카 모리포르미스	UTEX 1439; 1441; 1435; 1437	100	서열번호 15
프로토테카 윅커해미	UTEX 1533	99.8	서열번호 16
프로토테카 모리포르미스	UTEX 1434	75.9	서열번호 17
프로토테카 조프피	UTEX 1438	75.7	서열번호 18
프로토테카 모리포르미스	UTEX 1436	88.9	서열번호 19

[0444]

[0445] 상기 열거된 균주의 서브세트 유래의 지질 샘플은 HPLC를 사용하여 지질 프로필에 대해 분석하였다. 결과는 하기 표 11에 나타내어져 있다. 대안적으로, 지질 프로필은 실시예 11에 개략적으로 기술된 방법을 이용하여 측정될 수 있다.

표 11

프로토테카 종 내 지질 사슬의 다양성

균주	C14:0	C16:0	C16:1	C18:0	C18:1	C18:2	C18:3	C20:0	C20:1
UTEX 327	0	12.01	0	0	50.33	17.14	0	0	0
UTEX 1441	1.41	29.44	0 70	3.05	57.72	12.37	0.97	0.33	0
UTEX 1435	1 09	25.77	0	2.75	54.01	11.90	2.44	0	0

[0446]

[0447] (용매 추출을 통해 또는 익스펠러 프레스를 사용하여) 프로토테카 모리포르미스 UTEX 1435로부터 추출된 오일을 대상으로, 카로티노이드, 엽록소, 토코페롤, 기타 다른 스테롤 및 토코트리에놀에 대해 분석하였다. 그 결과는 하기 표 12에 요약되어 있다.

표 12

[0448] 프로토테카 모리포르미스(UTEX 1435)로부터 추출된 오일 중 카로티노이드, 엽록소, 토코페롤/스테롤 및 토코트리에놀 분석 결과

	압착된 오일 (mcg/ml)	용매 추출된 오일(mcg/ml)
cis-루테인	0.041	0.042
trans-루테인	0.140	0.112
trans-제아잔틴	0.045	0.039
cis-제아잔틴	0.007	0.013
t-알파-크립토잔틴	0.007	0.010
t-베타-크립토잔틴	0.009	0.010
t-알파-카로틴	0.003	0.001
c-알파-카로틴	검출되지 않음	검출되지 않음
t-베타-카로틴	0.010	0.009
9-cis-베타-카로틴	0.004	0.002
라이코펜	검출되지 않음	검출되지 않음
총 카로티노이드	**0.267**	**0.238**
엽록소	<0.01mg/kg	<0.01mg/kg
토코페롤 및 스테롤		
	압착된 오일 (mg/100g)	용매추출된 오일(mg/100g)
감마 토코페롤	0.49	0.49
캄페스테롤	6.09	6.05
스티그마스테롤	47.6	47.8
베타-시토스테롤	11.6	11.5
기타 다른 스테롤	445	446

토코트리에놀		
	프레스된 오일 (mg/g)	용매추출된 오일(mg/g)
알파 토코트리에놀	0.26	0.26
베타 토코트리에놀	<0.01	<0.01
감마 토코트리에놀	0.10	0.10
델타 토코트리에놀	<0.01	<0.01
총 토코트리에놀	0.36	0.36

[0449] 표준 식물성 오일 가공 방법을 이용하여, 프로토테카 모리포르미스(4개의 개별 분획)로부터 추출된 오일을 정제한 후 표백하였다. 간단히 말해서, 프로토테카 모리포르미스로부터 추출된 미정제 오일을 수평 디캔터 내에서 투명하게 만들었는데, 이때 오일로부터 고체가 분리되었다. 그 다음, 투명한 오일을 시트르산과 물이 담긴 탱크로 옮긴 후, 이를 약 24시간 동안 방치하여 침강시켰다. 24시간 경과 후, 탱크 내 혼합물은 2개의 별도 층들을 형성하였다. 아래 층은 물과 고무질로 이루어졌는데, 이후 이것들을 경사분리(decantation)에 의해 제거한 다음, 고무 성질이 제거된 오일을 표백 탱크로 옮겨 담았다. 그 다음, 오일을 일정량의 추가 시트르산과 함께 가열하였다. 이후, 표백 점토를 상기 표백 탱크 내에 첨가하였으며, 이 혼합물을 진공 하에서 추가로 가열하여 잔류하던 물을 전부 증발시켰다. 그 다음, 혼합물을 잎상 필터를 통해 펌핑하여 표백 점토를 제거하였다. 이후, 여과된 오일을 최종 5㎛ 연마 필터(polishing filter)에 통과시킨 다음, 이를 수집하여 사용시까지 보관하여 두었다. 그 다음, 정제 및 표백(RB) 오일을 카로티노이드, 엽록소, 스테롤, 토코트리에놀 및 토코페롤에 대해 분석하였다. 이와 같은 분석 결과는 하기 표 13에 요약되어 있다. “nd”는 검출되지 않았음을 의미하고, 검출 감도는 하기에 나열된 바와 같았다.

[0450] **검출 감도**

[0451] 카로티노이드(mcg/g) nd = <0.003mcg/g

[0452] 엽록소(mcg/g) nd = <0.03mcg/g

[0453] 스테롤(%) nd = 0.25%

[0454] 토코페롤(mcg/g); nd = 3mcg/g

표 13

정제 및 표백된 프로토테카 모리포르미스 오일 중 카로티노이드, 엽록소, 스테롤, 토코트리에놀 및 토코페롤 분석 결과

	분획 A	분획 B	분획 C	분획 D
카로티노이드 (mcg/g)				
루테인	0.025	0.003	nd	0.039
제아잔틴	nd	nd	nd	nd
cis-루테인/제아잔틴	nd	nd	nd	nd
trans-알파-크립토잔틴	nd	nd	nd	nd
trans-베타-크립토잔틴	nd	nd	nd	nd
trans-알파-카로틴	nd	nd	nd	nd
cis-알파-카로틴	nd	nd	nd	nd
trans-베타-카로틴	nd	nd	nd	nd
cis-베타-카로틴	nd	nd	nd	nd
라이코펜	nd	nd	nd	nd
미확인	0.219	0.066	0.050	0.026
총 카로티노이드	**0.244**	**0.069**	**0.050**	**0.065**
엽록소 (mcg/g)				
엽록소 A	0.268	0.136	0.045	0.166
엽록소 B	nd	nd	nd	nd
총 엽록소	**0.268**	**0.136**	**0.045**	**0.166**
스테롤 (%)				
브라시카스테롤	nd	nd	nd	nd
캄페스테롤	nd	nd	nd	nd
스티그마스테롤	nd	nd	nd	nd
베타-시토스테롤	nd	nd	nd	nd
총 스테롤	nd	nd	nd	nd
토코페롤 (mcg/g)				
알파-토코페롤	23.9	22.8	12.5	8.2
베타-토코페롤	3.72	nd	nd	nd
감마-토코페롤	164	85.3	43.1	38.3
델타-토코페롤	70.1	31.1	18.1	14.3
총 토코페롤	**262**	**139.2**	**73.7**	**60.8**
토코트리에놀 (mcg/g)				
알파-토코트리에놀	190	225	253	239
베타-토코트리에놀	nd	nd	nd	nd
감마-토코트리에놀	47.3	60.4	54.8	60.9
델타-토코트리에놀	12.3	16.1	17.5	15.2
총 토코트리에놀	**250**	**302**	**325**	**315**

[0455]

[0456] 프로토테카 모리포르미스 오일의 동일한 4개 분획을 또한 미량 원소에 대해 분석하였으며, 그 결과는 하기 표 14에 요약되어 있다.

표 14

정제 및 표백된 프로토테카 모리포르미스 오일의 원소분석

	분획 A	분획 B	분획 C	분획 D
원소분석 (ppm)				
칼슘	0.08	0.07	< 0.04	0.07
인	< 0.2	0.38	< 0.2	0.33
나트륨	< 0.5	0.55	< 0.5	< 0.5
칼륨	1.02	1.68	< 0.5	0.94
마그네슘	< 0.04	< 0.04	< 0.04	0.07
망간	< 0.05	< 0.05	< 0.05	< 0.05
철	< 0.02	< 0.02	< 0.02	< 0.02
아연	< 0.02	< 0.02	< 0.02	< 0.02
구리	< 0.05	< 0.05	< 0.05	< 0.05
황	2.55	4.45	2.36	4.55
납	< 0.2	< 0.2	< 0.2	< 0.2
규소	0.37	0.41	0.26	0.26
니켈	< 0.2	< 0.2	< 0.2	< 0.2
유기 염화물	< 1.0	< 1.0	< 1.0	2.2
무기 염화물	< 1.0	< 1.0	< 1.0	< 1.0
질소	4.4	7.8	4.2	6.9
리듐	< 0.02	< 0.02	< 0.02	< 0.02
붕소	0.07	0.36	0.09	0.38
알루미늄	--	< 0.2	< 0.2	< 0.2
바나듐	< 0.05	< 0.05	<0.05	< 0.05
로비본드 색 (°L)				
레드	5.0	4.3	3.2	5.0
옐로우	70.0	70.0	50.0	70.0
모노&디글리세라이드(HPLC에 의해 분석)(%)				
디글리세라이드	1.68	2.23	1.25	1.61
모노글리세라이드	0.03	0.04	0.02	0.03
유리 지방산(FFA)	1.02	1.72	0.86	0.83
비누	0	0	0	
산화 및 중합된 트리글리세라이드				
산화트리글리세라이드(%)	3.41	2.41	4.11	1.00
중합트리글리세라이드(%)	1.19	0.45	0.66	0.31
과산화물가 (meg/kg)	0.75	0.80	0.60	1.20
p-아니시딘가(단위 무)	5.03	9.03	5.44	20.1
물 및 기타 불순물 (%)				
칼 피셔 수분	0.8	0.12	0.07	0.18
총극성 화합물	5.02	6.28	4.54	5.23
비누화될 수 없는 물질	0.92	1.07	0.72	1.04
불용성 불순물	<0.01	<0.01	0.01	< 0.01
총 오일 (%)				
중성 오일	98.8	98.2	99.0	98.9

[0457]

[0459] **실시예 2: 프로토테카의 일반적 유전자 총 형질 전환법**

[0460] 씨셸 골드 마이크로캐리어 550(Seashell Gold Microcarriers 550) 나노미터를 제조자 프로토콜에 따라서 제조하였다. 플라스미드(20㎍)를 결합 완충액 50㎕ 및 S550d 골드 캐리어 60㎕(30㎎)와 혼합하고 나서, 이 혼합물을 1분 동안 얼음 속에서 항온 처리하였다. 여기에 침전 완충액(100㎕)을 첨가하고, 이 혼합물을 얼음 속에서 1분 더 항온 처리하였다. 이를 와류시킨 다음, 에펜도르프 5415C 마이크로 원심 분리로 10초 동안 10,000rpm에서 회전시켜 DNA 코팅된 입자를 펠릿화하였다. 골드 펠릿을 냉각 100% 에탄올 500㎕로 1회 세정한 후, 마이크로 원심 분리로 간단히 회전시켜 펠릿화한 다음, 얼음 냉각된 에탄올 50㎕로 재현탁하였다. 간단히 초음파 처리한 후(1초 내지 2초), DNA 코팅된 입자들 10㎕를 즉시 캐리어 막으로 이동시켰다.

[0461] 프로토테카 균주를 2% 글루코스 함유 프로테오스 배지(2g/L 효모 추출물, 2.94mM NaNO3, 0.17mM CaCl2·2H2O, 0.3mM MgSO4·7H2O, 0.4mM K2HPO4, 1.28mM KH2PO4, 0.43mM NaCl) 중에서, 세포 밀도가 2 x 106세포/㎖가 될 때까지 선회 진탕기에서 생육하였다. 세포를 회수한 후, 멸균 증류수로 1회 세정하고 나서, 배지 50㎕ 중에 재현탁하였다. 1 x 107 세포를 비선택성 프로테오스 배지 평판 중앙 3분의 1에 해당하는 부위에 도말하였다. 세포를 PDS-1000/He 유전자 총 입자 전달 시스템(Bio-Rad)으로 발사하였다. 파열 판(rupture disk)(1350psi)을 사용하였으며, 평판을 스크리닝/매크로캐리어 조립체 6㎝ 아래에 놓았다. 25℃에서 12시간 내지 24시간 동안 세포를 회수하였다. 회수시, 고무 주걱으로 세포를 평판으로부터 긁어내고 나서, 이를 배지 100㎕와 혼합한 다음, 적절하게 선택된 항생제를 함유하는 평판상에 도말하였다. 25℃에서 7일 내지 10일 동안 배양한 다음, 형질 전환된

세포를 나타내는 콜로니들을 평판 상에서 가시적으로 확인하였다. 콜로니들을 골라낸 다음 (항생제 또는 탄소원 함유) 선택 한천 평판 상에 점 찍었다(제2 선택 라운드용).

[0462] **실시예 3: 프로토테카 내 다양한 티오에스테라제의 발현**

[0463] 프로토테카 종 내 이종 티오에스테라제 유전자를 발현하는 방법과 이에 따른 효과는, 본원에 참조로 포함되어 있는 PCT 출원 PCT/US2009/066142호에 이미 기술되어 있다. 고등 식물 종으로부터 유래하는 기타 다른 티오에스테라제 유전자/유전자 생성물의 효과를 더욱 자세히 관찰하였다. 이러한 티오에스테라제는 다음과 같은 고등 식물로부터 유래하는 티오에스테라제를 포함한다:

[0464]

		GenBank		
[0465]	종		수탁 번호	특이성 서열 번호
[0466]	신나모뭄 캠포라	Q39473	C14	서열 번호 30 ~ 31
[0467]	움벨루라리아 캘리포니카	Q41635	C10-C12	서열 번호 34 ~ 35
[0468]	쿠페아 후커리아나	AAC49269	C8-C10	서열 번호 32 ~ 33
[0469]	쿠페아 팔루스트리스	AAC49179	C8	서열 번호 36 ~ 37
[0470]	쿠페아 란세올라타	CAB60830	C10	서열 번호 38 ~ 39
[0471]	아이리스 게르마니카	AAG43858.1	C14	서열 번호 40 ~ 41
[0472]	미리스티카 프래그란스	AAB717291.1	C14	서열 번호 42 ~ 43
[0473]	쿠페아 팔루스트리스	AAC49180	C14	서열 번호 44 ~ 45
[0474]	얼머스 아메리카나	AAB71731	광범위	서열 번호 46 ~ 47
[0475]	미리스티카 프래그란스	AAB71729	광범위	서열 번호 145 ~ 146
[0476]	가르시니아 망고스타나	AAB51525.1	C16	서열 번호 147 ~ 148
[0477]	쿠페아 후커리아나	Q39513.1	C16	서열 번호 149 ~ 150
[0478]	엘라에이스 기니엔시스	AAD42220.2	C16	서열 번호 151 ~ 152
[0479]	브라시카 나푸스	CAA52070.1	C18	서열 번호 153 ~ 154
[0480]	리시누스 커뮤니스	ABS30422.1	C18:1	서열 번호 155 ~ 156

[0481] 모든 경우에서, 상기 티오에스테라제 구조물 각각은 유전자 총 입자 발사법을 사용하여 프로토테카 모리포르미스(UTEX 1435)로 형질 전환되었다. 기타 다른 형질 전환법, 예를 들어 PCT 출원 PCT/US2009/066142호에 개시된 상동성 재조합법은 또한 관심의 유전자의 이종 발현에 적당할 것이다. 실시예 2에 기술된 방법을 사용하여 프로토테카 모리포르미스(UTEX 1435)를 상기 티오에스테라제 구조물 각각으로 형질 전환하였다. 구조물 각각은 NeoR 유전자를 함유하였으며, 양성 클론 선택은 100㎍/㎖ G418을 사용하여 수행하였다. 모든 암호화 영역은 프로토테카 모리포르미스 UTEX 1435(표 2 참조) 핵내 유전자에 고유한 코돈 편향 사용(codon bias)을 반영하도록 최적화된 코돈이었다. 사용된 구조물에 대한 아미노산 서열과 cDNA 서열은 둘 다 서열 동일성 목록에 나열되어 있다. 달리 특정하지 않는 한, 고등 식물 티오에스테라제 각각에 대한 운반 펩티드(transit peptide)는 프로토테카 모리포르미스 델타 12 지방산 불포화 효소(서열 번호 48) 또는 클로렐라 프로토테코이데스 스테아로일 ACP 불포화 효소(서열 번호 49)로부터 유래하는 조류 코돈 최적화 운반 펩티드로 대체되었다. 모든 티오에스테라제 구조물은 클라미도마나스 레인하르드티 베타-튜불린 프로모터/5'UTR에 의해 작동되었다. 선택된 양성 클론의 생장 및 지질 생산을 야생형(형질 전환되지 않은) 프로토테카 모리포르미스(UTEX 1435)와 비교하였다. 야생형 및 선택된 양성 클론을 2% 글루코스 G418 평판 상에서 생육하였다. 선택된 양성 클론에 대한 각각의 구조물의 지질 프로필 분석 결과는 하기 표 15에 요약되어 있다(면적%로 표시).

표 15

다양한 이종 티오에스테라제를 발현하는 프로토테카 모리포르미스의 지질 프로필

지방산	UTEX 1435 wt	티오에스테라제								
		유.칼리포니카	씨.캠포라	아이.갤르마니카	엠.프레그란스	씨.팔루스트리스 C8:0	씨.후커리아나	씨.란세올라타	씨.팔루스트리스 C14:0	유.아메리카나
C8:0	0	0	0	0		3.1	1.8	0	0	.09
C10:0	0.02	.07	.02	.01	.09	.56	6.85	1.91	.01	2.85
C12:0	0.05	14	1.82	.09	.05	.25	.2	.29	.06	.74
C14:0	1.65	3	17.3	2.59	5.31	1.45	1.8	1.83	2.87	10.45
C16:0	28.0	21.4	24.3	26.52	31.08	22.84	23.9	25.55	27.23	33.3
C18:0	2.9	2.9	2.7	3.11	2.71	3.24	2.8	3.26	3.62	3.47
C18:1	53.8	45.2	41.3	49.96	39.77	56.62	49.8	55.43	51.04	38.71
C18:2	10.95	10	9.7	11.86	14.17	8.24	9.7	8.17	10.81	7.38
C18:3 α	0.8	.86	.8	.40	.64	.61	.9	.58	.97	.52
총포화물 (면적%)	32.62	44.97	46.14	32.32	39.24	31.44	37.35	32.84	33.79	50.9

[0482]

[0483] 상기 결과들은 발현된 티오에스테라제 전부가 지방산 프로필에 어느 정도까지 영향을 주었음을 나타낸다. "총 포화물" 행을 살펴보면, 포화도는 티오에스테라제, 예를 들어 유.캘리포니카, 씨.캠포라 및 가장 눈에 띄게는 유.아메리카나로부터 유래하는 티오에스테라제들 중 몇 개의 발현에 의해 깊은 영향을 받았다. 이러한 총 포화물의 백분율 변화는, 고등 식물로부터 유래하는 티오에스테라제의 이종 발현이 지질 사슬 길이에 확실하게 영향을 줄 수 있으며; 또한 미세 조류에 의해 생성되는 지질 프로필의 기타 다른 속성, 즉 지방산의 포화도에도 영향을 줄 수 있다는 점에서 예상하지 못했던 것이다.

[0484] 씨.팔루스트리스 C8 티오에스테라제, 씨.후커리아나 티오에스테라제, 유.캘리포니카 및 씨.캠포라 티오에스테라제로 형질 전환된 선택 클론을, G418의 양에 변화를 주고(25㎎/L 내지 50㎎/L), 온도에도 변화를 주면서(22℃ 내지 25℃) 추가로 생육하였으며, 이러한 클론을 대상으로 지질 프로필을 측정하였다. 표 16은 각각의 티오에스테라제를 함유하는 대표적인 클론의 지질 프로필(면적%)을 요약한다. 유.아메리카나 티오에스테라제를 함유하는 제2 구조물을 구성하여, 전술한 유전자 총 발사법 이용하여 이를 프로토테카 모리포르미스(UTEX 1435)로 형질 전환되었다. 제2 구조물을 상동성 재조합법을 통해 세포 내에 도입하였다. 프로토테카 종 내에서의 상동성 재조합법은 PCT 출원 PCT/US2009/66142호에 이미 기술되어 있다. 사용된 상동성 DNA는 프로토테카 모리포르미스 UTEX 1435로부터 유래하는 6S 게놈 DNA 서열로부터 유래한 것이었다(서열 번호 92 및 서열 번호 84에 제시된 공여체 서열). 선택 요인은 씨.레인하르드티 베타 튜불린 프로모터에 의해 작동되는, 에스.세레비지아에로부터 유래하는 코돈 최적화 suc2 유전자를 이용하여 수크로스 상에서 생장하는 능력이었다. 천연의 유.아메리카나 운반 펩티드는 클로렐라 프로토테코이데스(UTEX 250) 스테아로일 ACP 불포화 효소 운반 펩티드에 의해 대체되었다. 이 구조물의 cDNA는 서열 목록에 서열 번호 50으로서 나열되어 있다. 양성 클론의 선택은 2% 수크로스 평판 상에서 수행되었으며, 지질 프로필 측정을 위한 생성된 배양물도 또한 2% 수크로스 함유 배지 상에서 생육하였다. 상동적으로 재조합된 이종 유.아메리카나 티오에스테라제를 함유하는 프로토테카 모리포르미스 균주에 대한 대표적인 지질 프로필은 표 16에 요약되어 있다.

표 16

이종 티오에스테라제 유전자를 함유하는 프로토테카 모리포르미스 균주의 지질 프로필

	씨. 팔루스트리스 C8	씨. 후커리아나	씨. 캠포라	유. 아메리카나 2
C8:0	12.28	2.37	0	0
C10:0	2.17	12.09	0.02	4.69
C12:0	0.34	0.33	3.81	1.02
C14:0	1.59	2.08	32.73	16.21
C16:0	15.91	20.07	24.03	38.39
C18:0	1.59	1.57	1.21	2.83
C18:1	50.64	41.80	18.64	27.22
C18:2	13.02	16.37	16.57	7.65
C18:3 α	1.52	1.75	1.66	0.74
총포화물	33.88	38.51	61.80	63.14

[0485]

[0486] 전술한 클론을 사용하였을 때, 나타낸 이종 티오에스테라제 유전자를 함유하는 형질 전환체 모두는 지방산 프로필에 어느 정도까지 영향을 미쳤으며, 포화된 지방산의 총 백분율도 야생형(형질 전환되지 않은) 프로토테카 모리포르미스와 비교하였을 때 변하였다. 상동성 재조합법에 의해 도입된 유.아메리카나 티오에스테라제를 함유하는 프로토테카 모리포르미스는 총 포화물이 가장 많이 증가하였다.

[0487] 뿐만 아니라, 외인성 씨.후커리아나, 씨.캠포라, 유.캘리포니카 또는 유.아메리카나 티오에스테라제를 함유하는 유전자 이식 클론은 신규 지질 프로필에 대해 평가하였다. 씨.후커리아나 티오에스테라제 함유 클론은 22℃에서 2% 글루코스, 25㎎/㎖ G418를 함유하는 배지 중에서 생육되었을 때, C8:0 5.10%; C10:0 18.28%; C12:0 0.41%; C14:0 1.76%; C16:0 16.31%; C18:0 1.40%; C18:1 40.49%; 및 C18:2 13.16%의 지질 프로필을 나타내었다. 씨.캠포라 티오에스테라제 함유 클론(또한 외인성 수크로스 인버타제도 함유함)은 25℃에서 2% 수크로스를 함유하는 배지 중에서 생육되었을 때, C10:0 0.04%; C12:0 6.01%; C14:0 35.98%; C16:0 19.42%; C18:0 1.48%; C18:1 25.44%; 및 C18:2 9.34%의 지질 프로필을 나타내었다. 유.캘리포니카 티오에스테라제 함유 클론은 22℃에서 2% 글루코스와 25㎎/㎖ 내지 100㎎/㎖ G418을 함유하는 배지 중에서 생육되었을 때, C8:0 0%; C10:0 0.11%; C12:0 34.01%; C14:0 5.75%; C16:0 14.02%; C18:0 1.10%; C18:1 28.93%; 및 C18:2 13.01%의 지질 프로필을 나타내었다. 유.아메리카나 티오에스테라제 함유 클론은 28℃에서 2% 글루코스를 함유하는 배지 중에서 생육되었을 때, C10:0 1.54%; C12:0 0.43%; C14:0 7.56%; C16:0 39.45%; C18:0 2.49%; C18:1 38.49%; 및 C18:2 7.88%의 지질 프로필을 나타내었다.

[0488] 고등 식물로부터 유래하는 부가의 티오에스테라제도 또한 프로토테카 모리포르미스 UTEX 1435 유전적 배경에 도입하였으며, 코돈 최적화 cDNA 서열과 아미노산 서열은 전술한 서열 목록에 나열되어 있다. 이러한 부가의 티오에스테라제는 미리스티카 프래그란스 유래의 광범위한 특이성을 가지는 티오에스테라제(C14:0 내지 C18:0), 가르시니아 망고스타나 유래의 C16:0 선호 티오에스테라제, 쿠페아 후커리아나 유래의 C16:0 선호 티오에스테라제, 엘라에이스 기니엔시스 유래의 C16:0 선호 티오에스테라제, 브라시카 나푸스 유래의 C18:0 선호 티오에스테라제 및 리시누스 커뮤니스 유래의 C18:1 선호 티오에스테라제를 포함한다. 상기 이식 유전자/형질 전환체 각각으로부터 유래하는 생성 유전자 이식 클론과 발현 구조물에 대한 상세한 설명은 이하에 기술되어 있다.

[0489] 미리스티카 프래그란스로부터 유래하며 광범위한 특이성을 가지는 티오에스테라제(C14:0 내지 C18:0)를, 전술한 방법을 사용하여 프로토테카 모리포르미스 UTEX 1435 유전적 배경에 도입하였다. 2개의 상이한 발현 구조물들을 테스트하였으며, 이 구조물들은 각각 상이한 색소체 표적화 서열을 함유한다. 구조물 2개에 있어서, 에스.세레비지아에 수크로스 인버타제 유전자인 suc2는, 유일한 탄소원으로서 수크로스를 함유하는 평판 상에서 생장하는 능력을 양성 형질 전환체에 부여하는 선택가능한 마커로서 사용되었다. 두 개의 발현 구조물, 즉 pSZ1318 및

pSZ1317은 6S 게놈 영역에 대한 5'(서열 번호 82) 및 3'(서열 번호 84) 상동성 재조합 표적화 서열(구조물에 측접)을 함유하였는데, 상기 재조합 표적화 서열은, 씨.레인하르드티 β-튜불린 프로모터/5'UTR 및 클로렐라 불가리스 질산염 환원 효소 3'UTR의 제어 하에 있는 에스.세레비지아에 suc2 수크로스 인버타제 암호화 영역과 핵내 게놈으로의 통합 과정에 관여하였다. 이러한 에스.세레비지아에 suc2 발현 카세트는 서열 번호 159로서 나열되어 있다. pSZ1318은, 프로토테카 모리포르미스 Amt03 프로모터(서열 번호 89) 및 씨.불가리스 질산염 환원 효소 3'UTR의 제어 하에 있는, 프로토테카 모리포르미스 델타 12 FAD 유래 운반 펩티드(서열 번호 48)에 의해 천연 운반 펩티드가 대체된 엠.프래그란스 티오에스테라제 암호화 영역을 함유하였다. 프로토테카 모리포르미스 델타 12 FAD 유래 운반 펩티드를 함유하는, 코돈 최적화 엠.프래그란스 티오에스테라제는 서열 번호 145로서 나열되어 있다. pSZ1317은, 프로토테카 모리포르미스 Amt03 프로모터(서열 번호 89) 및 씨.불가리스 질산염 환원 효소 3'UTR의 제어 하에 있는, 클로렐라 프로토테코이데스 스테아로일 ACP 불포화 효소 유래 운반 펩티드(서열 번호 49)에 의해 천연 운반 펩티드가 대체된 엠.프래그란스 암호화 영역을 함유하였다. 씨.프로토테코이데스 스테아로일 ACP 불포화 효소 유래 운반 펩티드를 함유하는, 코돈 최적화 엠.프래그란스 티오에스테라제는 서열 번호 158로서 나열되어 있다. 두개의 발현 구조물, 즉 pSZ1318 및 pSZ1317은 프로토테카 세포로 형질 전환되었으며, 선택은 유일한 탄소원으로서 수크로스를 함유하였던 평판 상에서 수행되었다. 각각의 형질 전환 단계로부터 양성 클론을 선택하여, 이 클론은 지질 생산을 위한 질소 제한 조건 하에 유일한 탄소원으로서 수크로스를 함유하는 배지 중에서 생육하였다. 선택된 양성 클론의 하위 세트의 지질 프로필을 전술한 직접 에스테르 결합 전이 방법을 사용하여 측정하였으며, 그 결과는 표 17에 요약되어 있다.

표 17

미리스티카 프래그란스 광범위 특이성 티오에스테라제 유전자 이식 프로토테카 세포의 지질 프로필

[0490]

균주	C10:0	C12:0	C14:0	C16:0	C18:0	C18:1	C18:2
야생형	0.01	0.03	1.17	25.86	2.84	58.33	9.16
pSZ1318 클론A	0.03	0.23	16.09	37.72	6.11	27.39	9.98
pSZ1318 클론B	0.03	0.22	15.74	37.17	6.23	28.16	9.94
pSZ1318 클론C	0.03	0.22	14.97	36.05	5.87	30.48	9.86
pSZ1317 클론A	0.02	0.21	15.23	36.62	5.11	31.83	8.76
pSZ1317 클론B	0.03	0.27	18.06	38.88	5.64	26.11	8.90
pSZ1317 클론C	0.02	0.24	16.19	37.02	5.61	29.52	9.19

[0491] 미리스티카 프래그란스 티오에스테라제 이식 유전자를 함유하는 양성 클론은 지질 프로필에 변화가 나타났다. 그러나, 상기 요약된 결과들은 예상외의 결과를 나타내었으며, 고등 식물에서 미리스티카 프래그란스 티오에스테라제는 C16:0 지방 아실-ACP에 대해 상당한 활성을 나타낸 반면에(Voelker et al.,1997), 프로토테카 세포에서 미리스티카 프래그란스 티오에스테라제는 영향력의 크기가 차이가 났었으며(C14:0 > C18:0 > C16:0), 단지 C16:0만에 대해서는 영향이 미치는 범위가 더욱 넓었다.

[0492] 가르시니아 망고스타나 유래 C16:0 선호 티오에스테라제를 프로토테카 모리포르미스 UTEX 1435 유전적 배경에 도입하였으며, 이 경우 코돈 최적화 cDNA 서열과 아미노산 서열은 전술한 바와 같은 서열 목록 내에 나열되어 있다. 발현 구조물은 6S 게놈 영역에 대한 5'(서열 번호 82) 및 3'(서열 번호 84) 상동성 재조합 표적화 서열(구조물에 측접)을 포함하였는데, 이 재조합 표적화 서열은, 씨.레인하르드티 β-튜불린 프로모터/5'UTR 및 클로렐라 불가리스 질산염 환원 효소 3'UTR의 제어 하에 있는 에스.세레비지아에 suc2 수크로스 인버타제 암호화 영역과 핵내 게놈으로의 통합 과정에 관여하였다. 이러한 에스.세레비지아에 suc2 발현 카세트는 서열 번호 159로서 나열되어 있으며, 선택 마커로서 작용하였다. 지.망고스타나 암호화 영역은 프로토테카 모리포르미스 Amt03 프로모터/5'UTR(서열 번호 89) 및 씨.불가리스 질산염 환원 효소 3'UTR의 제어 하에 두었다. 상기 지.망고스타나 천연 운반 펩티드는 또한 씨.프로토테코이데스 스테아로일 불포화 효소 유래 운반 펩티드(서열 번호 49)로 대체되었으며, 운반 펩티드로 대체된 상기 티오에스테라제의 cDNA 서열은 서열 번호 147로서 나열되어 있다. 전체 가르시니아 망고스타나 발현 카세트는 pSZ1452로 명명하였으며, 이 발현 카세트는 프로토테카 모리포르미스 유전적 배경으로 형질 전환되었다. 양성 클론을, 유일한 탄소원으로서 수크로스를 함유하는 평판 상에서 스크리닝하였다. 양성 클론의 하위 세트를 선택하여, 지질 생산 조건 하에서 생육한 다음, 전술한 바와 같은 직접 에스테르 결합 전이 방법을 사용하여 지질 프로필을 측정하였다. 선택된 클론의 지질 프로필은 하기 표 18에 요약되어 있다.

표 18

가르시니아 망고스타나 C16:0 선호 티오에스테라제 유전자 이식 프로토테카 세포의 지질 프로필

균주	C10:0	C12:0	C14:0	C16:0	C18:0	C18:1	C18:2
야생형	0.01	0.03	1.17	25.86	2.84	58.33	9.16
pSZ1452 클론A	0.02	0.07	5.52	62.77	4.36	18.99	6.29
pSZ1452 클론B	0.02	0.08	5.69	61.66	4.76	19.28	6.54
pSZ1452 클론C	0.01	0.05	3.44	57.97	4.21	24.76	7.38

[0493]

[0494] 상기 결과는, 지.망고스타나 티오에스테라제 이식 유전자를 함유하는 형질 전환체가 C16:0 지방산 수준에 상당한 영향을 미쳤으며, 이보다는 그 정도가 약하지만, C14:0 지방산과 C18:0 지방산 수준에도 영향을 미쳤고, 아울러 야생형에 비해서 C18:1 지방산 수준은 극도로 감소하였음을 나타낸다.

[0495] 쿠페아 후커리아나 유래 C16:0 선호 티오에스테라제를 프로토테카 모리포르미스 UTEX1435 유전적 배경에 도입하였다. 2개의 발현 구조물을 만들었는데, 이것들 중 하나는 천연 쿠페아 후커리아나 C16 선호 티오에스테라제 운반 펩티드 서열을 포함하였고(pSZ1417), 다른 하나는 천연 운반 펩티드 서열이 씨.프로토테코이데스 스테아로일-ACP 불포화 효소 유래 운반 펩티드(서열 번호 49)로 대체된 것이었다(pSZ1462). 천연 운반 펩티드를 함유하는 씨.후커리아나 티오에스테라제의 암호화 서열은 서열 번호 149로서 나열되어 있으며, 운반 펩티드가 대체된 씨.후커리아나 티오에스테라제의 암호화 서열은 서열 번호 160으로서 나열되어 있다. 상기 발현 구조물은 둘 다 6S 게놈 영역에 대한 5'(서열 번호 82) 및 3'(서열 번호 84) 상동성 재조합 표적화 서열(구조물에 측접)을 포함하였는데, 상기 재조합 표적화 서열은, 씨.레인하르드티 β-튜불린 프로모터/5'UTR 및 클로렐라 불가리스 질산염 환원 효소 3'UTR의 제어 하에 있는 에스.세레비지아에 suc2 수크로스 인버타제 암호화 영역과 핵내 게놈으로의 통합 과정에 관여하였다. 이러한 에스.세레비지아에 suc2 발현 카세트는 서열 번호 159로서 나열되어 있으며, 선택 마커로서 작용하였다. 상기 2개의 구조물에서, 씨.후커리아나 암호화 영역은 프로토테카 모리포르미스 Amt03 프로모터/5'UTR(서열 번호 89) 및 씨.불가리스 질산염 환원 효소 3'UTR의 제어 하에 두었다. 상기 구조물은 둘 다 프로토테카 모리포르미스 유전적 배경으로 형질 전환되었으며, 양성 클론은 유일한 탄소원으로서 수크로스를 함유하는 평판 상에서 스크리닝하였다. 양성 클론의 하위 세트를 선택하여 지질 생산 조건 하에서 생육하였으며, 지질 프로필은 전술한 직접 에스테르 결합 전이 방법을 사용하여 측정하였다. 선택된 클론의 지질 프로필은 하기 표 19에 요약되어 있다.

표 19

쿠페아 후커리아나 C16:0 선호 티오에스테라제 유전자 이식 프로토테카 세포의 지질 프로필

균주	C10:0	C12:0	C14:0	C16:0	C18:0	C18:1	C18:2
야생형	0.01	0.03	1.17	25.86	2.84	58.33	9.16
pSZ1417 클론A	0.02	0.06	4.21	55.29	2.59	26.87	9.02
pSZ1417 클론B	0.02	0.06	4.12	54.57	2.31	26.43	10.45
pSZ1417 클론C	0.01	0.05	3.59	53.18	2.60	29.02	9.48
pSZ1462 클론A	0.02	0.11	10.62	67.42	2.18	12.95	5.13
pSZ1462 클론B	0.03	0.11	8.88	66.83	2.30	15.32	5.16
pSZ1462 클론C	0.03	0.11	9.28	66.65	2.27	15.19	5.14
pSZ1462 클론D	0.02	0.09	8.30	66.36	2.27	16.52	5.01

[0496]

[0497] 상기 결과는, 쿠페아 후커리아나 C16:0 선호 티오에스테라제 구조물들 중 어느 하나를 함유하는 형질 전환체가 C16:0 지방산 수준에 상당한 영향을 미쳤으며, 이보다는 그 정도가 약하지만, C14:0 지방산 수준에도 영향을 미쳤고, 아울러 야생형에 비해서 C18:1 지방산 수준은 극도로 감소하였음을 나타낸다. 2개의 구조물 내 운반 펩티드의 차이는 pSZ1417 형질 전환체와 비교하였을 때, pSZ1462 형질 전환체 내 C16:0 지방산 수준이 증가하였음을 설명해줄 수 있다.

[0498] 엘라에이스 기니엔시스(아프리카 기름 야자)로부터 유래하는 2개의 C16:0 선호 티오에스테라제(GenBank 수탁 번호 제AAD422220.2호(서열 번호 152) 및 제ABD83939호(서열 번호 162)인 아미노산 서열과 상응하며, 각각 이.기니엔시스 팔미토일-ACP 티오에스테라제 및 이.기니엔시스 팔미토일-ACP 티오에스테라제 PATE라 칭함)를, 프로토테카 모리포르미스 UTEX 1435 유전적 배경에 도입하였다. 코돈 최적화 cDNA 서열과 아미노산 서열은 상기 지정한 바와 같은 서열 목록에 나열되어 있다. 상기 2개의 티오에스테라제는 아미노산 동일성을 상당한 수준으로 공유하지만(94% 초과), 이 테오에스테라제 각각의 아프리카 기름 야자수 내 역할은 여전히 규명되지 않았다. 이.기니엔시스 팔미토일-ACP 티오에스테라제를 암호화하는 구조물은 pSZ1437로, 그리고 이.기니엔시스 팔미토일-ACP 티오에스테라제 PATE를 암호화하는 구조물은 pSZ1436으로 명명하였다. 상기 발현 구조물은 둘 다 6S 게놈 영역에 대한 5'(서열 번호 82) 및 3'(서열 번호 84) 상동성 재조합 표적화 서열(구조물에 측접)을 함유하였는데, 상기 재조합 표적화 서열은, 씨.레인하르드티 β-튜불린 프로모터/5'UTR 및 클로렐라 불가리스 질산염 환원 효소 3'UTR의 제어 하에 있는 에스.세레비지아에 suc2 수크로스 인버타제 암호화 영역과 핵내 게놈으로의 통합 과정에 관여하였다. 이러한 에스.세레비지아에 suc2 발현 카세트는 서열 번호 159로서 나열되어 있으며, 선택 마커로서 작용하였다. 상기 2개의 구조물에서, 이.기니엔시스 티오에스테라제 암호화 영역은 프로토테카 모리포르미스 Amt03 프로모터/5'UTR(서열 번호 89) 및 씨.불가리스 질산염 환원 효소 3'UTR의 제어 하에 두었다. 상기 구조물은 둘 다 프로토테카 모리포르미스 유전적 배경으로 형질 전환되었으며, 양성 클론은 유일한 탄소원으로서 수크로스를 함유하는 평판 상에서 스크리닝하였다. 양성 클론의 하위 세트를 선택한 후 지질 생산 조건 하에서 생육하였으며, 지질 프로필은 전술한 직접 에스테르 결합 전이 방법을 사용하여 측정하였다. 선택된 클론의 지질 프로필은 하기 표 20에 요약되어 있다.

표 20

엘라에이스 기니엔시스 C16:0 선호 티오에스테라제 유전자 이식 프로토테카 세포의 지질 프로필

균주	C10:0	C12:0	C14:0	C16:0	C18:0	C18:1	C18:2
야생형	0.01	0.03	1.17	25.86	2.84	58.33	9.16
pSZ1437 클론 A	0.02	0.07	3.82	55.86	3.12	24.47	10.48
pSZ1437 클론 B	0.02	0.05	3.01	53.23	3.47	28.70	9.26
pSZ1437 클론 C	0.02	0.06	3.33	53.20	3.30	26.64	11.19
pSZ1437 클론 D	0.02	0.05	3.08	52.88	3.60	27.94	10.16
pSA1437 클론 E	0.02	0.05	3.01	52.84	3.48	28.46	9.87
pSZ1436 클론 A	0.01	0.04	1.48	29.54	3.33	52.26	10.58
pSZ1436 클론 B	0.01	0.04	1.48	29.43	3.33	52.11	10.76
pSZ1436 클론 C	0.01	0.04	1.50	29.25	3.38	52.07	10.89
pSZ1436 클론 D	0.01	0.04	1.51	29.18	3.41	51.80	11.17
pSZ1436 클론 E	0.01	0.04	1.54	29.14	3.56	51.43	11.42

[0499]

[0500] pSZ1437에 의해 암호화된 이.기니엔시스 C16:0 선호 티오에스테라제는 C16:0 지방산 수준에 상당한 영향을 미쳤으며, 이보다는 그 정도가 약하지만, C14:0 지방산 수준에도 영향을 미쳤고, 야생형에 비해서 C18:1 지방산 수준은 극도로 감소하였음을 나타낸다. 놀라운 점은, pSZ1436에 의해 암호화된 이.기니엔시스 C16:0 선호 티오에스테라제 PATE는, 이 pSZ1436에 의해 암호화된 티오에스테라제에 대한 아미노산 동일성이 상당한 수준이었음에도 불구하고, C16:0 또는 C14:0 지방산 수준에 대해 비교적 작은 활성을 나타냈다는 점이다.

[0501] 브라시카 나푸스 유래 C18:0 선호 티오에스테라제를 프로토테카 모리포르미스 UTEX 1435 유전적 배경에 도입하였으며, 코돈 최적화된 cDNA 서열 및 아미노산 서열은 상기 지정된 바와 같은 서열 목록 내에 나열되어 있다. 발현 구조물은 6S 게놈 영역에 대한 5'(서열 번호 82) 및 3'(서열 번호 84) 상동성 재조합 표적화 서열(구조물에 측접)을 함유하였는데, 이 재조합 표적화 서열은, 씨.레인하르드티 β-튜불린 프로모터/5'UTR 및 클로렐라 불가리스 질산염 환원 효소 3'UTR의 제어 하에 있는 에스.세레비지아에 suc2 수크로스 인버타제 암호화 영역과 핵내 게놈으로의 통합 과정에 관여하였다. 이러한 에스.세레비지아에 suc2 발현 카세트는 서열 번호 159로서 나열되어 있으며, 선택 마커로서 작용하였다. 비.나푸스 암호화 영역은 프로토테카 모리포르미스 Amt03 프로모터/5'UTR(서열 번호 89) 및 씨.불가리스 질산염 환원 효소 3'UTR의 제어 하에 두었다. 전체 브라시카 나푸스 발현 카세트를 pSZ1358로 명명하였으며, 이는 프로토테카 모리포르미스 유전적 배경으로 형질 전환되었다. 양성 클론

은 유일한 탄소원으로서 수크로스를 함유하는 평판 상에서 스크리닝하였다. 양성 클론의 하위 세트를 선택한 후 지질 생산 조건 하에서 생육하였으며, 지질 프로필을 전술한 직접 에스테르 결합 전이 방법을 사용하여 측정하였다. 선택된 클론의 지질 프로필은 하기 표 21에 요약되어 있다.

표 21

브라시카 나푸스 C18:0 선호 티오에스테라제 유전자 이식 프로토테카 세포의 지질 프로필

[0502]

균주	C10:0	C12:0	C14:0	C16:0	C18:0	C18:1	C18:2
야생형	0.00	0.04	1.18	25.44	3.42	57.97	6.98
pSZ1358 클론A	0.07	0.31	1.51	33.27	27.26	27.37	7.50
pSZ1358 클론B	0.07	0.33	1.60	34.73	26.71	26.52	7.32

[0503] 상기 결과는, 브라시카 나푸스 C18:0 선호 티오에스테라제 이식 유전자를 함유하는 형질 전환체가 C18:0 지방산 수준에 상당한 영향을 미쳤으며, 이보다는 그 정도가 약하지만, C16:0 지방산 수준에도 영향을 미쳤고, 아울러 야생형에 비해서 C18:1 지방산 수준은 극도로 감소하였음을 나타낸다.

[0504] 리시누스 커뮤니스 유래 지방 아실-ACP 티오에스테라제를 프로토테카 모리포르미스 UTEX 1435 유전적 배경에 도입하였으며, 코돈 최적화 cDNA 서열 및 아미노산 서열은 상기 지정된 서열 목록에 나열되어 있다. 발현 구조물은 6S 게놈 영역에 대한 5'(서열 번호 82) 및 3'(서열 번호 84) 상동성 재조합 표적화 서열(구조물에 측접)을 포함하였는데, 상기 재조합 표적화 서열은, 씨.레인하르드티 β-튜불린 프로모터/5'UTR 및 클로렐라 불가리스 질산염 환원 효소 3'UTR의 제어 하에 있는 에스.세레비지아에 suc2 수크로스 인버타제 암호화 영역과 핵내 게놈으로의 통합 과정에 관여하였다. 이러한 에스.세레비지아에 suc2 발현 카세트는 서열 번호 159로서 나열되어 있으며, 선택 마커로서 작용하였다. 알.커뮤니스 암호화 영역은 프로토테카 모리포르미스 Amt03 프로모터/5'UTR (서열 번호 89) 및 씨.불가리스 질산염 환원 효소 3'UTR의 제어 하에 두었다. 리시누스 커뮤니스 천연 운반 펩티드는 또한 씨.프로토테코이데스 스테아로일 불포화 효소 유래 운반 펩티드(서열 번호 49)로 대체되었으며, 운반 펩티드가 대체된 티오에스테라제의 cDNA 서열은 서열 번호 155로서 나열되어 있다. 전체 리시누스 커뮤니스 발현 카세트는 pSZ1375로 명명되었으며, 이는 프로토테카 모리포르미스 유전적 배경으로 형질 전환되었다. 양성 클론은 유일한 탄소원으로서 수크로스를 함유하는 평판 상에서 스크리닝하였다. 양성 클론의 하위 세트를 선택한 후 지질 생산 조건 하에서 생육하였으며, 지질 프로필은 전술한 직접 에스테르 결합 전이 방법을 사용하여 측정하였다. 선택된 클론의 지질 프로필은 하기 표 22에 요약되어 있다.

표 22

리시누스 커뮤니스 ACP-티오에스테라제 유전자 이식 프로토테카 세포의 지질 프로필

[0505]

균주	C10:0	C12:0	C14:0	C16:0	C18:0	C18:1	C18:2
야생형	0.01	0.03	0.98	24.65	3.68	62.48	6.26
pSZ1375 클론A	0.01	0.03	0.91	18.34	2.55	67.93	8.35
pSZ1375 클론B	0.01	0.03	0.97	18.51	2.47	67.83	8.25
pSZ1375 클론C	0.01	0.03	0.93	18.65	2.84	67.58	7.90
pSZ1375 클론D	0.01	0.03	0.92	18.90	2.30	67.48	8.37

[0506] 상기 결과는, 리시누스 커뮤니스 티오에스테라제 이식 유전자를 함유하는 형질 전환체가 C16:0 지방산 수준에 상당한 영향을 미쳤으며, 이보다는 그 정도가 약하지만, C18:0 지방산 수준에도 영향을 미쳤음을 나타낸다. 또한, 야생형의 수준에 비해서 C18:1 지방산 수준도 동반하여 증가하였다.

[0508] **실시예 4: 다수의 외인성 이종 티오에스테라제 유전자를 사용하는 프로토테카의 형질 전환**

[0509] 전술한 방법을 사용하여 미세 조류 균주인 프로토테카 모리포르미스(UTEX1435)를 형질 전환하여 하나의 클론에서 다수의 티오에스테라제가 발현되도록 만들었다. 하나의 클론에서 다수의 티오에스테라제가 발현됨에 따라서, 미세 조류는, 임의의 단일 티오에스테라제만이 단독으로 발현될 때와 지방산 프로필이 완전히 상이한 오일을 생

산하게 되었다(이전 실시예에서 입증됨). 프로토테카 모리포르미스(UTEX 1435)는 처음에, 상동성 재조합법을 사용하여 신나모뮴 캠포라 티오에스테라제(C14 선호 티오에스테라제) 유전자와 함께 수크로스 인버타제 유전자, 그리고 에스.세레비지아에 유래 suc2 유전자(선택은 수크로스 상에서 생장하는 능력을 바탕으로 함)로 형질 전환되었다. 이와 같은 상동성 재조합 구조물에 사용된 DNA는 상기 섹션 III에 기술된 바와 같이 프로토테카 모리포르미스 게놈 DNA의 KE858 영역으로부터 유래한 것이다. 이 유전자 구조물의 관련 부분은 서열 목록 중 서열 번호 51로서 나열하였다. 수크로스 함유 평판 상에서 양성 클론을 스크리닝하였다. 그 다음, 양성 클론을 3개의 카세트들 중 하나로 다시 형질 전환하였는데, 이 경우, 카세트들은 각각 항생제 G418에 대한 내성과, (1) 쿠페아 후커리아나 유래 티오에스테라제 유전자(C8-10 선호), 서열 번호 52; (2) 움벨루라리아 캘리포니카 유래 티오에스테라제 유전자(C12 선호), 서열 번호 53; 또는 얼머스 아메리카나 유래 티오에스테라제(광범위; C10-C16 선호), 서열 번호 54의 부가의 티오에스테라제를 암호화하였다. 각각의 구조물의 관련 부분의 서열도 서열 목록에 포함된다. 티오에스테라제 유전자 둘 다를 발현하는 클론을 50㎍/㎖ G418과 수크로스를 함유하는 배지 상에서 스크리닝하였다. 양성 클론을 선택하고 나서, 이의 생장과 지질 프로필을 평가하였다. 표 23은 대표적인 양성 클론의 지질 프로필을 요약한 것이다(면적%로 표시).

표 23

다수의 티오에스테라제로 형질 전환된 프로토테카 모리포르미스의 지질 프로필

[0510]

지방산	UTEX 1435	UTEX 1435 + 씨.캠포라 TE	UTEX 1435 + 씨.캠포라 TE 유전적 배경		
			+ 씨.후커리아나 TE	+ 유.캘리포니카 TE	+ 유.아메리카나 TE
C8:0	0	0	0.19	0	0.06
C10:0	0.02	0.02	2.16	0.07	1.87
C12:0	0.05	0.66	0.53	13.55	1.61
C14:0	1.65	10.52	7.64	8.0	14.58
C16:0	28.0	22.56	22.31	19.98	29.53
C18:0	2.9	6.67	3.23	2.24	2.93
C18:1	53.8	47.78	48.54	42.55	37.3
C18:2	10.95	12.3	11.76	10.13	8.9
C18:3 α	0.8	0.93	0.91	0.91	0.76
총 포화물 (면적%)	32.62	40.43	36.06	43.84	50.58

[0511] 추가로, 씨.캠포라 및 유.캘리포니카 티오에스테라제를 함유하는 이중 티오에스테라제 클론을 22℃에서 50mg/L G418과 2% 수크로스를 함유하는 배지 중에서 생육하였다. 이와 같은 생육 조건 하에 있는 상기 균주로부터 얻어진 지방산 프로필은, C8:0(0); C10:0(0.10); C12:0(31.03); C14:0(7.47); C16:0(15.20); C18:0(0.90); C18:1(30.60); C18:2(12.44); 및 C18:3α(1.38)(총 포화물 = 54.7)이었다.

[0512] 2개의 상동성 재조합 구조물(하나는 6S 영역을 표적화하며, 다른 하나는 KE858 영역을 표적화함)과 씨.캠포라 티오에스테라제를 함유하는 이중 티오에스테라제 클론을 제조하였다. 양성 클론으로서 대표적인 것의 지방산 프로필은, C8:0 0%; C10:0 0.06%; C12:0 5.91%; C14:0 43.27%; C16:0 19.63%; C18:0 0.87%; C18:1 13.96%; 및 C18:2 13.78%(총 포화물 = 69.74%)이었다. 이 클론의 C12 내지 C14 수준은 49% 이상이었는데, 이는 야생형 세포 내 C12 내지 C14 수준의 37배 이상에 해당하는 수준이었다.

[0513] 상기 데이터는 다수의 티오에스테라제가 미세 조류 내에서 성공적으로 공동 발현될 수 있음을 보여준다. 다수의 티오에스테라제가 공동 발현되면, 지방산 프로필에 변화가 생겼는데, 즉, 지방산 프로필은 야생형 균주의 지방산 프로필과 상당히 상이할 뿐만 아니라, 개별적 티오에스테라제 중 임의의 하나를 발현시킴으로써 얻어지는 지방산 프로필과도 상이하였다. 중첩 사슬 길이가 특이적인 티오에스테라제를 다수 개 발현하면, 이와 같은 특이적 지방산을 누적 증가시킬 수 있었다.

[0514] 프로토테카 모리포르미스 내에서 이종 티오에스테라제를 발현(어느 하나만을 단독으로 발현하거나 함께 발현)시키면, 숙주 균주의 지방산/지질 프로필을 변화시켰을 뿐만 아니라, 현재 다양한 종자 작물로부터 얻을 수 있는 오일과 비교하였을 때(표 5) 상기 프로필은 현재 이용되고 있는 기타 다른 시스템 내에서 발견되지 않는, 진정

으로 유일한 오일의 프로필이었다. 유전자 이식된 균주는 형질 전환되지 않은 야생형 균주와 확연한 차이를 보였을 뿐만 아니라, 이들 유전자 이식 균주는 표 5에 보여준 시판 오일 중 임의의 것의 프로필과는 확연히 상이한 프로필을 나타냈다. 예를 들어, 코코넛 오일과 팜핵 오일의 C8 내지 C10 지방산 수준은 5.5% 내지 17%였다. 씨.팔루스트리스 C8 선호 티오에스테라제 또는 씨.후커리아나 C10 선호 티오에스테라제를 발현하는 유전자 이식 균주는 각각 C8 내지 C10 지방산을 3.66% 내지 8.65%의 범위 내에서 축적하였다. 이와 같은 C8 내지 C10 지방산 수준은 코코넛 오일과 팜핵 오일의 C8 내지 C10 지방산 수준과 비슷하였으나, 유전자 이식 조류 균주는 상당히 더 높은 수준의 C12:0 지방산이 결여되었으며, 이 균주는 또한 C16:0(유전자 이식체 중 23% 대 코코넛 또는 팜핵 오일 중 11% 내지 16%) 및/또는 C18:1(유전자 이식체 중 50% 내지 57% 대 코코넛 또는 팜핵 오일 중 8% 내지 19%) 수준이 매우 높았다.

[0515] **쿠페아 라이티(Cuphea wrightii) 티오에스테라제를 발현시켜 균주 UTEX1435 내 라우레이트 및 미리스테이트 풍부 오일 생산**

[0516] 쿠페아 라이티 종자는 C10:0 지방산 25% 이상 및 C12:0 지방산 65% 이상을 함유하는 오일을 축적하는 것으로 보였다. 2개의 FatB 티오에스테라제, CwFatB1(GenBank 수탁 번호 제U56103호) 및 CwFatB2(GenBank 수탁 번호 제U56104호)는 쿠페아 라이티(문헌[Leonard et al,, Plant Mol. Biol. 34(4):669-79 (1997)]에 기술됨)로부터 클로닝하여, 아라비돕시스 탈리아나 내에서 발현하였다(문헌[Leonard et al,, Plant J. 13(5):621-8 (1998)]에 기술됨). CwFatB1 및 CwFatB2를 발현하는 에이.탈리아나 유전자 이식 계열의 지방산 프로필에 있어서는, C12:0 지방산 종이 16% 내지 25% 증가되었음이 확인되었다(Leonard et al, 1998, 상동). 본 발명의 발명자들은, 균주 UTEX1435가 쿠페아 라이티 티오에스테라제, CwFatB1 및 CwFatB2을 발현시킴으로써 라우레이트 및 미리스테이트 풍부 오일을 생산하는 능력을 입증하였다. 본원에 기술된 실시예에서는, CwFatB1 및 CwFatB2를 발현하는 유전자 이식 균주를 전술된 형질 전환법을 사용하여 생산하였다.

[0517] CwFatB1 및 CwFatB2의 아미노산 서열을 이하에 보였다(예측 엽록체 표적화 서열에 밑줄을 그어 나타냄). 이와 같은 1차 아미노산 서열을 형질 전환 구조물에 삽입될 해당 유전자를 합성하는데 사용하였다. 2개의 유전자의 뉴클레오티드 서열을 균주 UTEX 1435(전술한 바와 같이 선호 코돈 사용 빈도 이용) 내에서 발현용으로서 최적화하였다.

[0518] CwFatB1(U56103):

MVAAAASSAFFSVPTPGTSPKPGKFGNWPSSLSVPFKPDNGGFVKANASAHPKANGS
AVNLKSGSLETPPRSFINQLPDLSMLLSKITTVFGAAEKQWKRPGMLVEPFGVDRIFQ
DGVFFRQSFSIRSYEIGVDRTASIETLMNIFQETSLNHCKSIGLLNDGFGRTPEMCKRD
LIWVVTKIQVEVNRYPTWGDTIEVNTWVSESGKNGMGRDWLISDCRTGEILIRATSV
WAMMNQNTRRLSKFPYEVRQEIAPHFVDSAPVIEDDRKLHKLDVKTGDSIRDGLTPR
WNDLDVNQHVNNVKYIGWILKSVPIEVFETQELCGVTLEYRRECGRDSVLESVTTM
DPAKEGDRCVYQHLLRLEDGADITIGRTEWRPKNAGANGAISSGKTSNGNSVS
(서열 번호 186)

[0519]

[0520] CwFatB2(U56104):

MVVAAAASSAFFPVPAPRPTPKPGKFGNWPSSLSQPFKPKSNPNGRFQVKANVSPHP
KANGSAVSLKSGSLNTLEDPPSSPPPRTFLNQLPDWSRLRTAITTVFVAAEKQFTRLD
RKSKRPDMLVDWFGSETIVQDGLVFRERFSIRSYEIGADRTASIETLMNHLQDTSLNH
CKSVGLLNDGFGRTPEMCTRDLIWVLTKMQIVVNRYPTWGDTVEINSWFSQSGKIG
MGREWLISDCNTGEILVRATSAWAMMNQKTRRFSKLPCEVRQEIAPHFVDAPPVIED
NDRKLHKFDVKTGDSICKGLTPGWNDFDVNQHVSNVKYIGWILESMPTEVLETQEL
CSLTLEYRRECGRESVVESVTSMNPSKVGDRSQYQHLLRLEDGADIMKGRTEWRPK
NAGTNRAIST (서열 번호 187)

[0521]

[0522] **씨.라이티 티오에스테라제를 사용하는 UTEX1435의 형질 전환**: 본 실시예에서는, UTEX1435 균주를 씨.라이티 FatB1 및 FatB2 티오에스테라제를 발현하는 카세트가 도입된 수용체 균주로서 사용하였다. 형질 전환 구조물은 티오에스테라제와 함께, 수크로스 상에서 선택될 수 있도록 만드는 카세트(사카로마이세스 세레비지아에 suc2 유전자)를 함유하였다. 세포를, 전술한 바와 같이 유전자 총을 이용하여 형질 전환하였다. 세포를 2% 수크로스 함유 배지 상에서 직접 형질 전환하였다. 형질 전환 구조물을 제조하여, 티오에스테라제 발현이 씨.레인하르드티 B-튜불린 프로모터 또는 내인성 UTEX 1435 Amt3 프로모터에 의해 진행되도록 만들었다.

[0523] 티오에스테라제 카세트의 부가 버전(천연 고등 식물 운반 펩티드가 조류 운반 펩티드에 의해 치환된 것)을 제조하였다. 이 구조물에 사용된 운반 펩티드는 다음과 같이 명명하였다: TP1은 UTEX250으로부터 유래하는 스테아로

일 ACP 불포화 효소에 대한 운반 펩티드를 암호화하는 것이고, TP2는 UTEX1435로부터 유래하는 스테아로일 ACP 불포화 효소에 대한 운반 펩티드를 암호화하는 것이며, TP3은 UTEX1435로부터 유래된 델타 12 지방산 불포화 효소의 운반 펩티드를 암호화하는 것이고, 또한 TP4는 UTEX1435로부터 유래된 이소펜테닐 디포스페이트 합성 효소의 운반 펩티드를 암호화하는 것이다. 본 실시예에 사용된 구조물은 이하 표 24에 나열하였다.

표 24

TE 구조물

구조물	설 명
Const.1	6S-CrbTub_suc2_nr::CrbTub_CwFatB1_nr-6S
Const.2	6S-CrbTub_suc2_nr::CrbTub_CwFatB2_nr-6S
Const.3	6S-CrbTub_suc2_nr::Amt3_CwFatB1_nr-6S
Const.4	6S-CrbTub_suc2_nr::Amt3_CwFatB2_nr-6S
Const.5	6S-CrbTub_suc2_nr::CrbTub_TP1-CwFatB1_nr-6S
Const.6	6S-CrbTub_suc2_nr::CrbTub_TP2-CwFatB1_nr-6S
Const.7	6S-CrbTub_suc2_nr::CrbTub_TP3-CwFatB1_nr-6S
Const.8	6S-CrbTub_suc2_nr::CrbTub_TP4-CwFatB1_nr-6S
Const.9	6S-CrbTub_suc2_nr::CrbTub_TP1-CwFatB2_nr-6S
Const.10	6S-CrbTub_suc2_nr::CrbTub_TP2-CwFatB2_nr-6S
Const.11	6S-CrbTub_suc2_nr::CrbTub_TP3-CwFatB2_nr-6S
Const.12	6S-CrbTub_suc2_nr::CrbTub_TP4-CwFatB2_nr-6S

[0524]

[0525] **suc2 및 씨.라이티 FatB1 티오에스테라제를 발현하는 형질 전환 DNA (Const.1)**: 형질 전환 구조물의 서열, 즉 6S-CrbTub_suc2_nr::CrbTub_CwFatB1_nr-6S(Const.1으로 명명)을 이하에 제시하였다. 관련 제한 위치는 굵은 글씨체의 소문자에 밑줄을 그어 표시하였으며, SapI, KpnI, AscI, MfeI, BamHI, EcoRI, SpeI, AscI, XhoI, SacI 및 SapI를 5'→3' 방향으로 각각 표시하였다. SapI 위치는 형질 전환 DNA의 5' 및 3' 말단을 한정하는 것이었다. 구조물의 5' 및 3' 측접부에 존재하는 밑줄 친 서열은, 상동성 재조합을 통해 형질 전환 DNA를 표적 통합할 수 있는, UTEX1435 유래 게놈 DNA를 나타내는 것이다(6S 영역). 5'→3' 방향으로 진행되는 씨.레인하르드티 B-튜불린 프로모터(에스.세레비지아에 suc2 유전자(수크로스 가수 분해 활성을 암호화하므로, 균주가 수크로스 상에서 생육될 수 있도록 허용)를 발현시킴)는 소문자로 박스 안에 표시하였다. suc2 에 대한 개시 인자 ATG 및 종결 인자 TGA는 이탤릭체의 굵은 대문자로 표시하였으며, 암호화 영역은 소문자 이탤릭체로 표시하였다. 클로렐라 불가리스 질산염 환원 효소 3'UTR은 굵은 글씨체의 소문자로 표시하였으며, 이 뒤에는 스페이서 영역이 존재한다. 씨.라이티 TE(CwFatB1)를 발현시키는 씨.레인하르드티 B-튜불린 프로모터는 글씨에 박스를 쳐서 표시하였다. 티오에스테라제(CwFatB1)의 개시 인자 ATG 또는 종결 인자 TGA는 굵은 이탤릭체 대문자로 표시하였으며, 암호화 영역 나머지 부분은 소문자 이탤릭체로 표시하였다. 티오에스테라제의 예측 색소체 표적화 서열은 서열 내 개시 인자 ATG 및 AscI 위치 사이에 존재한다. 씨.불가리스 질산염 환원 효소 3'UTR은 굵은 글시로 표시하였다.

[0526]

구조물 1

gctcttcgccgccgccactcctgctcgagcgcgcccgcgcgtgcgccgccagcgccttggccttttcgccgcgctcgtgcgcgtcg
ctgatgtccatcaccaggtccatgaggtctgccttgcgccggctgagccactgcttcgtccgggcggccaagaggagcatgaggga
ggactcctggtccagggtcctgacgtggtcgcggctctgggagcgggccagcatcatctggctctgccgcaccgaggccgcctcca
actggtcctccagcagccgcagtcgccgccgaccctggcagaggaagacaggtgaggggggtatgaattgtacagaacaaccacg
agccttgtctaggcagaatccctaccagtcatggctttacctggatgacggcctgcgaacagctgtccagcgaccctcgctgccgccg
cttctcccgcacgcttctttccagcaccgtgatggcgcgagccagcgccgcacgctggcgctgcgcttcgccgatctgaggacagtc
ggggaactctgatcagtctaaaccccctgcgcgttagtgttgccatcctttgcagaccggtgagagccgacttgttgtgcgccacccc



[0527]

*gcgccatcagcagcggcaagaccagcaacggcaacagcgtgagc**TGA**ttaattaa**actcgaggcagcagcagctcggatagt**
atcgacacactctggacgctggtcgtgtgatggactgttgccgccacacttgctgccttgacctgtgaatatccctgccgctttta
tcaaacagcctcagtgtgtttgatcttgtgtgtacgcgcttttgcgagttgctagctgcttgtgctatttgcgaatacccaccccag
catcccctttccctcgtttcatatcgcttgcatcccaaccgcaacttatctacgctgtcctgctatccctcagcgctgctcctgctcct
gctcactgccccctcgcacagccttggtttgggctccgcctgtattctcctggtactgcaacctgtaaaccagcactgcaatgctg
atgcacgggaagtagtgggatgggaacacaaatggaaagcttgagctcttgttttccagaaggagttgctccttgagcctttcattc
tcagcctcgataacctccaaagccgctctaattgtggaggggtttcgaatttaaaagcttggaatgttggttcgtgcgtctggaacaagc
ccagacttgtgctcactgggaaaaggaccatcagctccaaaaaacttgccgctcaaaccgcgtacctctgcttccgcgcaatctgccct
gttgaaatcgccaccacattcatattgtgacgcttgagcagtctgtaattgcctcagaatgtggaatcatctgccccctgtgcgagcccat
gccaggcatgtcgcgggcgaggacacccgccactcgtacagcagaccattatgctacctcacaatagttcataacagtgaccatatttc
tcgaagctccccaacgagcacctccatgctctgagtggccacccccgcgccctggtgcttgcggagggcaggtcaaccggcatggg
gctaccgaaatccccgaccggatcccaccacccccgcgatggaagaatctctccccgggatgtgggcccaccaccagcacaacct
gctggcccaggcgagcgtcaaaccatacctacacaaatatccttggcatcggccctgaattccttctgccgctctgctaccccggtgcttct
gtccgaagcagggttgctagggatcgctccgagtccgcaaaccctttgtcgcgtggcggggcttgttcgagcttgaagagc*

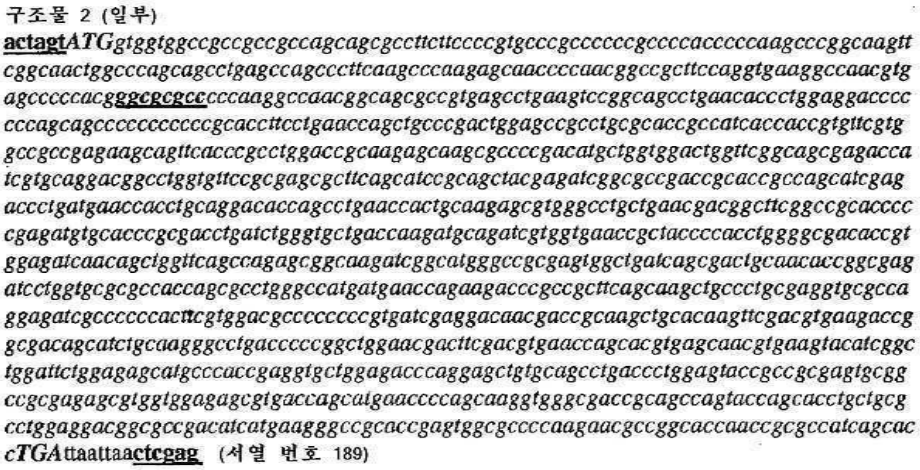
(서열 번호 188)

[0528]

[0529] suc2 및 씨.라이티 FatB2 티오에스테라제를 발현하는 형질 전환 DNA(Const.2): Const.1 유래 CwFatB1 유전자를

코돈 최적화 CwFatB2 유전자로 치환함으로써, SpeI 및 AscI 제한 위치(굵은 글씨체의 소문자에 밑줄 쳐 표시)를 이용하는 형질 전환 구조물, 6S-CrbTub_suc2_nr::CrbTub_CwFatB2_nr-6S(Const.2로 명명)을 제조하였다. 티오에스테라제(CwFatB2)의 개시 인자 ATG 및 종결 인자 TGA는 굵은 이탤릭체의 대문자로 표시하였으며, 암호화 영역의 나머지 부분은 소문자 이탤릭체로 표시하였다. 티오에스테라제의 예측 색소체 표적화 서열은 서열 내 개시 인자 ATG와 AscI 위치 사이에 존재한다.

구조물 2 (일부)



(서열 번호 189)

[0530]

[0531] **amt3 프로모터에 의해 작동되는, suc2 및 씨.라이티 FatB1 및 FatB2 티오에스테라제를 발현하는 형질 전환 DNA(Const.3 & 4)**: Const. 1 및 Const. 2 유래 티오에스테라제를 작동시키는 CrbTub 프로모터를, UTEX1435 유래 Amt3 프로모터(EcoRI 및 SpeI 제한 단편(이하 굵은 글씨체의 소문자에 밑줄 쳐 표시))로 치환하여, 형질 전환 구조물 6S-CrbTub_suc2_nr::Amt3_CwFatB1_nr-6S(Const.3로 명명), 그리고 6S-CrbTub_suc2_nr::Amt3_CwFatB2_nr-6S(Const.4로 명명)를 제조하였다. Amt3 프로모터 영역은 소문자에 박스를 쳐서 표시하였다.

구조물 3 및 4 (일부)



(서열 번호 190)

[0532]

[0533] **조류 운반 펩티드의 제어 하에 씨.라이티 FatB1 티오에스테라제를 발현하는 형질 전환 DNA**: Const.1 유래 CwFatB1의 천연 운반 펩티드를 이하에 SpeI 및 AscI 제한 단편(굵은 글씨체의 소문자에 밑줄 쳐서 표시)으로서 보인 상응 조류 운반 펩티드로 치환하여, 형질 전환 구조물 6S-CrbTub_suc2_nr::CrbTub_TP1-CwFatB1_nr-6S(Const.5로 명명); 구조물 6S-CrbTub_suc2_nr::CrbTub_TP2-CwFatB1_nr-6S(Const.6로 명명); 구조물 6S-CrbTub_suc2_nr::CrbTub_TP3-CwFatB1_nr-6S(Const.7로 명명) 및 구조물 6S-CrbTub_suc2_nr::CrbTub_TP4-CwFatB1_nr-6S(Const.8로 명명)을 제조하였다. 생성된 조류 운반 펩티드 서열은 이하 서열 내 개시 인자 ATG와 AscI 위치 사이에 존재한다.

구조물 5 - 12 (일부)
TP1 (UTEX250 스테아로일 ACP 불포화 효소 운반 펩티드 서열)
actagtATGgccaccgcatccacttttctcggcgttcaatgcccgctgcggcgacctgcgtcgctcggcgggctccgggccccggcgcccagcgaggcccctccccgtgcgcgggcgcgcc (서열 번호 191)
TP2 (UTEX1435 스테아로일 ACP 불포화 효소 운반 펩티드 서열)

[0534]

actagtATGgcttccgcggcattcaccatgtcggcgtgccccgcgatgactggcagggcccctggggcacgtcgctccggacggccagtcgccacccgcctgagggggcgcgccc (서열 번호 192)
TP3 (UTEX1435 델타 12 지방산 불포화 효소 운반 펩티드 서열)
actagtATGgctatcaagacgaacaggcagcctgtggagaagcctccgttcacgatcgggacgctgcgcaaggccatccccgcgcactgttcgagcgctcggcgcttcgtgggcgcgcc (서열 번호 193)
TP4 (UTEX1435 이소펜테닐 디포스페이트 합성 효소 운반 펩티드 서열)
actagtATGacgttcggggtcgcccctcccggccatgggccgcggtgtctcccttccccggcccagggtcgcggtgcgcgcccagtcggcgagtcaggttttggagagcgggcgcgcc (서열 번호 194)

[0535]

[0536] **조류 운반 펩티드의 제어 하에 씨.라이티 FatB2 티오에스테라제를 발현하는 형질 전환 DNA**: Const.2 유래 CwFatB2의 천연 운반 펩티드를, 상기 SpeI 및 AscI 제한 단편(굵은 글씨체의 소문자에 밑줄 쳐 표시)으로 보인 상응하는 조류 운반 펩티드로 치환하여, 형질 전환 구조물 6S-CrbTub_suc2_nr::CrbTub_TP1-CwFatB2_nr-6S(Const.9로 명명); 구조물 6S-CrbTub_suc2_nr::CrbTub_TP2-CwFatB2_nr-6S (Const.10로 명명); 구조물 6S-CrbTub_suc2_nr::CrbTub_TP3-CwFatB2_nr-6S(Const.11로 명명) 및 구조물 6S-CrbTub_suc2_nr::CrbTub_TP4-CwFatB2_nr-6S(Const.12로 명명)을 제조하였다. 조류 운반 펩티드 서열은 상기 서열 내 개시 인자 ATG와 AscI 위치 사이에 존재한다.

[0537] **쿠페아 라이티 티오에스테라제를 발현하는 균주의 지방산 프로필**: 전술한 구조물로 형질 전환된 균주를, 전술한 바와 같은 오일 생산 가능 조건 하에서 생육하였다. 야생형 UTEX 1435를 글루코스를 바탕으로 생육하였는데, 이 때, UTEX 1435의 형질 전환을 통해 생산된 모든 유전자 이식 계열을 수크로스 상에서 생육하였다. 테스트된 각각의 구조물에 있어서, 4개의 형질 전환체를 지방산 프로필에 미치는 영향력에 대해 분석하였다. 유전자 이식 균주의 지방산 프로필을 이하 표 25 내지 표 28에 나타내었다.

[0538] CwFatB1 및 CwFatB2를 발현하는 에이.탈리아나의 유전자 이식 계열(표 25)은, C16:0 지방산 축적에 상당히 영향을 받았으며, 아울러 C14:0 및 C12:0 지방산의 축적에도 영향을 받았음이 확인되었다(Leonard et al, 1998, 상동).

[0539] 표 26에서 볼 수 있는 바와 같이, 유전자 이식 UTEX 1435 계열(천연 고등 식물 운반 펩티드와 함께 CwFatB1(Const.1 & Const.3)을 발현)은, C14:0 지방산 축적에 1차적으로 영향을 받았으며, 이보다 그 정도는 덜하지만 C12:0 지방산 축적에도 영향을 받았다. 유전자 이식 UTEX1435 계열(천연 고등 식물 운반 펩티드와 함께 CwFatB2(Const.2 & Const.4)를 발현)은, C12:0 및 C14:0 지방산 축적에 상당히 영향을 받았으며, 이 경우, C12:0에 대한 영향력은 C14:0에 대한 영향력보다 컸음이 확인되었다. 2개의 프로모터, 즉 CrbTub(Const. 1 & Const.2) 및 Amt3(Const.3 & Const.4)를 비교함으로써, CwFatB1 및 CwFatB2를 발현하는 유전자 이식 계열은, Amt3 프로모터에 의해 작동될 때 C10:0, C12:0, C14:0 및 C16:0 지방산에 대해 상당히 큰 영향을 받았음이 입증되었다.

[0540] CwFatB1 티오에스테라제의 발현이 4개의 상이한 조류 엽록체 표적화 서열(Const. 5, 6, 7 및 8)에 의해 작동되는 유전자 이식 계열을 분석한 결과, 조류 운반 펩티드들 중 임의의 것은, 천연 고등 식물 운반 펩티드의 경우보다 티오에스테라제를 색소체에 더욱 효율적으로 표적화하는 것을 확인할 수 있었다(구조물 5 내지 8 내 C12:0 및 C14:0 수준(표 27)과, 구조물 1 내 C12:0 및 C14:0 수준(표 26)을 비교함). 이와 같은 유전자 이식 계열을 추가로 분석한 결과, 4개의 조류 운반 펩티드 TP2(UTEX1435 스테아로일 ACP 불포화 효소 엽록체 표적화 서열) 및 TP3(UTEX 1435 델타 12 지방산 불포화 효소 엽록체 표적화 서열)은, C12:0 및 C14:0 축적에 더욱 큰 영향을 미쳤음이 확인되었다. 이와 같은 2개의 운반 펩티드(TP1 & TP2)는 UTEX 1435 내 CwFatB1을 색소체에 더욱 효율적으로 표적화하는 것으로 보였다.

[0541] CwFatB2 티오에스테라제의 발현이 4개의 상이한 조류 엽록체 표적화 서열(Const. 9, 10, 11 및 12)에 의해 작동되는 유전자 이식 계열을 분석한 결과, 조류 운반 펩티드들 중 오로지 하나만이 천연 고등 식물 운반 펩티드(즉 TP-1)에 우수한 성능을 부여하였음이 입증되었다(C12:0 및 C14:0 지방산에 더욱 큰 영향을 미치는 것을 통해 알 수 있음(Const.9 및 Const.2 비교)).

[0542] 씨.라이티 티오에스테라제가 UTEX 1435 내에서 발현될 때, 이 효소의 영향력은, 아라비돕시스 내에서 발현될 때

와 상당히 상이하였다. CwFatB1 및 CwFatB2를 발현하는 에이.탈리아나의 유전자 이식 계열은, C16:0 지방산의 축적과 아울러 C14:0 및 C12:0 지방산의 축적에도 상당히 영향을 받았음을 나타내었다(표 25). 그러나, UTEX 1435 내에서 발현된 CwFatB1 및 CwFatB2의 C16:0 지방산에 대한 영향력 수준은 동일하지 않다. CwFatB1을 발현하는 모든 UTEX 1435 유전자 이식 계열과, CwFatB2를 발현하는 일부 UTEX 1435 유전자 이식 계열(Const. 10, 11, 12) 내 C12:C14 비율은, 이러한 티오에스테라제를 발현하는 에이.탈리아나 유전자 이식 계열 내 C12:C14 비율과 유사하였다(표 29). 그러나, UTEX 1435 유전자 이식 계열의 C14:C16 비율은, 에이.탈리아나 유전자 이식 계열의 C14:C16 비율과 비교하였을 때 상당히 작았다(표 29). Const.2, 4, 9를 사용하여 제조된 UTEX 1435 유전자 이식 계열 내 C12:C14 및 C14:C16 비율은, 동일한 티오에스테라제를 발현하는 에이.탈리아나 유전자 이식 계열 내 C12:C14 및 C14:C16 비율과는 상당히 상이하였다(표 29). 그러므로, UTEX 1435 내에서 CwFatB1 및 CwFatB2이 발현될 때의 오일 프로필은, 동일한 티오에스테라제를 발현하는 아라비돕시스의 유전자 이식 계열 내 오일 프로필과는 상당히 상이하였다. 또한, 이러한 UTEX 1435 유전자 이식 계열 내 오일 프로필은 야생형 UTEX 1435 내 오일 프로필과 확연히 상이하였다.

[0543] 마지막으로, 본 실시예에 기술된 유전자 이식 계열에 의해 생산된 변형 오일들은 또한, (전술한 바와 같은) 움벨루라리아 캘리포니카 유래 C12:0 특이 티오에스테라제를 발현하는 유전자 이식 UTEX 1435 계열 내 생성되는 라우레이트 풍부 오일들과 상당히 상이하였다.

[0544] 상기 사항들을 종합하여 보았을 때, 이와 같은 데이터들은, (1) UTEX 1435 내에서 쿠페아 라이티 티오에스테라제가 발현되면, 지방산 프로필에 상당한 영향을 미쳐 독특한 오일이 생산됨; (2) 균주 UTEX 1435 내에서 CwFatB1 티오에스테라제가 발현되면, 미리스테이트가 풍부한 오일이 생산됨; (3) UTEX 1435 내에서 CwFatB2 티오에스테라제가 발현되면, 라우레이트와 미리스테이트 둘 다가 풍부한 오일이 생산됨; (4) 조류 내에서 CwFatB1 및 FatB2가 발현되면, 생산된 중간 길이 사슬의 지방산의 절대 수준과 이 지방산의 상호간 상대 비율의 관점에서, 표준 고등 식물 시스템 내에서 나타나는 프로필과 분명히 구분되는 프로필을 나타냄과 같은 특성들을 나타내었다.

표 25

CwFatB1(CwFatB1-Ath) 및 CwFatB2(CwFatB2-Ath) 티오에스테라제를 발현하는 에이.탈리아나 유전자 이식 계열 및 에이.탈리아나 야생형(Ath) 및 UTEX1435 내 지방산 프로필(면적%)

[0545]

샘플 ID	C16:0	C12:0	C14:0	C16:0	C18:0	C18:1	C18:2	C18:3	C20:0	C20:1
UTEX1435	0.01	0.03	0.93	23.83	3.27	61.85	8.08	0.53	0.31	0.08
Ath	0.00	0.00	0.00	8.40	3.80	13.00	29.20	20.10	2.40	19.30
Cw FatB1-Ath	0.00	7.10	24.40	22.80	3.30	4.50	14.10	12.90	3.00	6.00
Cw FatB2-Ath	4.40	16.40	15.30	18.10	3.90	4.90	13.90	13.60	2.80	5.70

표 26

UTEX 1435 유전자 이식 계열 내 지방산 프로필(면적%로 표시)

구조물	샘플 ID	C10:0	C12:0	C14:0	C16:0	C18:0	C18:1	C18:2	C18:3	C20:0	C20:1
Const.1	1A	0.01	0.43	3.17	19.84	1.66	60.34	12.38	0.45	0.23	0.02
	1B	0.01	0.52	3.45	19.81	2.03	60.65	11.42	0.42	0.26	0.02
	1C	0.01	0.59	3.62	20.53	2.24	59.64	11.29	0.42	0.26	0.02
	1D	0.01	0.67	3.92	21.97	1.96	58.62	10.82	0.44	0.24	0.02
Const.2	2A	0.63	7.47	5.64	18.74	2.36	52.11	10.98	0.49	0.28	0.02
	2B	0.82	7.77	5.83	19.64	2.52	51.98	9.24	0.50	0.24	0.02
	2C	0.82	9.57	6.31	18.64	1.66	50.89	10.42	0.43	0.21	0.01
	2D	0.90	10.04	7.11	17.99	2.34	49.03	10.63	0.47	0.28	0.02
Const.3	3A	0.04	2.85	12.09	28.04	2.89	39.02	12.35	1.05	0.25	0.05
	3B	0.03	2.90	13.39	28.01	2.02	41.47	10.01	0.71	0.21	0.05
	3C	0.04	3.30	14.10	27.91	2.09	40.50	9.92	0.71	0.21	0.04
	3D	0.04	3.71	15.10	27.86	2.01	39.56	9.62	0.68	0.21	0.06
Const.4	4A	1.43	11.78	8.87	19.67	2.04	43.80	10.28	0.77	0.21	0.06
	4B	1.39	12.26	9.29	18.68	1.48	44.20	12.18	0.86	0.19	0.08
	4C	1.73	13.42	9.55	18.93	2.05	41.92	10.30	0.81	0.21	0.04
	4D	1.82	14.18	10.07	18.56	1.93	41.22	10.21	0.78	0.19	0.05

Const.1; Const.2; Const.3 및 Const.4. 발현

[0546]

표 27

Const.5; Const.6; Const.7 및 Const.8을 발현하는 UTEX 1435 유전자 이식 계열 내 지방산 프로필 (면적%로 표시)

구조물	샘플 ID	C10:0	C12:0	C14:0	C16:0	C18:0	C18:1	C18:2	C18:3	C20:0	C20:1
Const.5	5A	0.02	0.62	4.33	23.30	2.00	57.43	10.00	0.59	0.26	0.09
	5B	0.02	0.78	5.11	25.06	2.45	55.65	8.76	0.58	0.26	0.07
	5C	0.02	1.20	7.41	24.48	1.87	52.65	10.18	0.60	0.24	0.08
	5D	0.02	1.33	7.56	24.55	1.87	52.59	9.90	0.54	0.26	0.09
Const.6	6A	0.02	0.56	4.01	24.23	2.50	57.04	9.32	0.57	0.28	0.08
	6B	0.02	0.69	4.97	23.41	1.92	55.59	10.98	0.61	0.25	0.09
	6C	0.02	1.14	7.07	25.05	2.11	53.23	9.23	0.60	0.23	0.07
	6D	0.05	5.10	19.88	21.43	1.29	40.40	9.89	0.63	0.20	0.08
Const.7	7A	0.02	1.39	8.36	25.39	2.00	51.37	9.37	0.55	0.22	0.07
	7B	0.02	1.42	7.59	24.77	2.12	53.17	8.87	0.47	0.23	0.08
	7C	0.02	1.49	7.82	24.87	2.10	52.45	9.14	0.53	0.24	0.08
	7D	0.03	2.15	11.01	25.64	1.85	47.35	9.92	0.51	0.23	0.08
Const.8	8A	0.02	0.81	5.28	23.38	2.03	56.10	10.14	0.54	0.26	0.09
	8B	0.02	0.88	5.77	23.58	1.91	54.91	10.58	0.57	0.24	0.09
	8C	0.02	1.27	7.57	24.28	1.93	52.95	9.88	0.54	0.24	0.08
	8D	0.02	1.43	5.02	21.52	2.63	58.32	9.14	0.52	0.28	0.09

[0547]

표 28

Const.9; Const.10; Const.11 및 Const.12를 발현하는 UTEX 1435 유전자 이식 계열 내 지방산 프로필 (면적%로 표시)

구조물	샘플 ID	C10:0	C12:0	C14:0	C16:0	C18:0	C18:1	C18:2	C18:3	C20:0	C20:1
Const.9	9A	0.90	8.96	6.59	19.24	2.20	51.17	8.91	0.56	0.24	0.08
	9B	1.00	9.07	6.31	19.55	2.21	51.43	8.47	0.53	0.25	0.07
	9C	1.06	12.08	8.79	17.57	1.70	46.54	10.29	0.58	0.21	0.07
	9D	1.27	13.05	8.67	17.70	1.76	46.28	9.38	0.57	0.22	0.07
Const.10	10A	0.52	5.55	5.00	20.60	2.03	53.35	10.66	0.71	0.24	0.07
	10B	0.53	5.76	5.16	20.63	1.92	52.80	10.90	0.67	0.23	0.06
	10C	0.47	5.86	5.20	19.54	1.89	53.34	11.41	0.62	0.26	0.06
	10D	0.87	8.59	6.85	19.65	1.98	49.65	10.21	0.69	0.23	0.08
Const.11	11A	0.21	2.48	2.85	20.80	1.99	57.69	11.53	0.70	0.27	0.11
	11B	0.22	2.80	3.01	21.30	1.89	57.28	11.07	0.66	0.27	0.11
	11C	0.29	3.22	3.38	21.33	2.09	56.92	10.42	0.71	0.25	0.05
	11D	0.28	4.01	4.01	18.79	1.69	56.08	12.73	0.64	0.26	0.09
Const.12	12A	0.65	6.01	5.43	21.50	2.10	52.70	9.28	0.75	0.24	0.10
	12B	0.52	6.56	5.62	18.71	1.76	52.59	11.84	0.66	0.26	0.11
	12C	0.78	8.60	6.88	19.01	1.89	50.40	10.45	0.69	0.23	0.08
	12D	0.72	8.75	7.07	17.74	1.54	50.57	11.29	0.69	0.22	0.10

[0548]

표 29

CwFatB1(CwFatB1-Ath) 및 CwFatB2(CwFatB2-Ath)를 발현하는 에이.탈리아나 유전자 이식 계열, 및 CwFatB1(Const.1,3,5,6,7,8) 및 CwFatB2(Const.2,4,9,10,11,12)를 발현하는 UTEX 1435 유전자이식계열 내 C12:C14 및 C14:C16의 비율

샘플 ID	평균 12:14 비율	평균 14:16 비율
CwFatB1-Ath	0.291	1.070
Const.1	0.155	0.172
Const.3	0.233	0.489
Const.5	0.250	0.250
Const.6	0.174	0.397
Const.7	0.185	0.345
Const.8	0.190	0.254
CwFatB2-Ath	1.072	0.845
Const.2	1.396	0.332
Const.4	1.365	0.512
Const.9	1.419	0.414
Const.10	1.152	0.277
Const.11	0.938	0.163
Const.12	1.191	0.328

[0549]

[0550] **실시예 5: 질소 의존 내인성 프로토테카 프로모터의 동정**

[0551] **A. 내인성 질소 의존 프로모터의 동정 및 특성 규명**

[0552] cDNA 라이브러리는 표준 기술을 사용하여 프로토테카 모리포르미스(UTEX 1435)로부터 생성하였다. 상기 프로토테카 모리포르미스를 질소 풍부 조건 하에서 48시간 동안 생육시켰다. 이어서 5%의 접종물(v/v)을 낮은 질소로 이동시키고 상기 세포를 7일 동안 24시간 마다 수거하였다. 배양 약 24시간 후, 배지 내 질소 공급은 완전히 고갈되었다. 수집된 샘플은 드라이아이스 및 이소프로판올을 사용하여 즉시 동결시켰다. 후속적으로 총 RNA는 동결된 세포 펠렛 샘플로부터 분리하고 각각의 샘플로부터 유래하는 부분은 RT-PCR 연구를 위해 보존하였다. 상기 샘플로부터 수거된 총 RNA의 나머지는 폴리A 선택을 거쳤다. 이어서 각각의 조건으로부터 동몰량의 폴리A 선택된 RNA를 모으고, 이를 사용하여 벡터 pcDNA 3.0(Invitrogen) 중에서 cDNA 라이브러리를 생성하였다. 대략적으로 1200개 클론을 생성된 모아진 cDNA 라이브러리로부터 무작위로 집어내고, 양쪽 가닥을 서열분석하였다. 대략적으로 68개의 상이한 cDNA는 이들 1200개의 서열로부터 선택하고, 이를 사용하여 실시간 RT-PCR 연구에 사용하

기 위한 cDNA 특이적 프라이머를 설계하였다.

[0553] 보존된 세포 펠렛 샘플로부터 분리된 RNA를 상기 생성된 cDNA 특이적 프라이머 세트를 사용한 실시간 RT-PCR 연구에서 기질로서 사용하였다. 상기 보존된 RNA를 cDNA로 전환시키고, 68개 유전자 특이적인 프라이머 세트 각각에 대한 RT-PCR의 기질로서 사용하였다. 역치 사이클 또는 CT 수를 사용하여 시간 전반에 걸쳐 수집된 각각의 RNA 샘플 내에서 68개 cDNA 각각에 대한 상대적인 전사체 풍부함을 나타내었다. 질소 풍부와 질소 고갈된 조건 간의 상당한 증가(3배 초과)를 나타내는 cDNA는 잠재적인 유전자로서 표시하고 이의 발현은 질소 고갈에 의해 상향조절된다. 명세서에서 논의된 바와 같이, 질소 고갈/제한은 유지성 미생물에서 지질 합성의 공지된 유도인자이다.

[0554] 질소 고갈/제한 동안에 발현이 상향조절된 cDNA 유래의 추정 프로모터/5'UTR 서열을 확인하기 위해, 총 DNA를 질소 풍부 조건 하에서 생장한 프로토테카 모리포르미스(UTEX 1435)로부터 분리하고, 이어서 454 서열분석 기술(Roche)을 사용하여 서열분석을 거쳤다. 상기 RT-PCR 결과에 의해 상향 조절된 바와 같은 표시된 cDNA는 454 게놈 서열분석 판독으로부터 비롯된 조립된 콘틱에 대한 BLAST를 사용하여 비교하였다. cDNA의 5' 말단은 특이적 콘틱으로 맵핑하고, 가능한 경우 500bp 초과의 5' 측접 DNA를 사용하여 추정적으로 프로모터/UTR을 확인하였다. 이어서 프로모터/5'UTR의 존재는 게놈 DNA의 PCR 증폭을 사용하여 확인하고 클로닝하였다. 개별 cDNA 5' 말단을 사용하여 3' 프라이머를 설계하고 454 콘틱 조립체의 5' 말단을 사용하여 5' 유전자 특이적 프라이머를 설계하였다.

[0555] 제1 스크리닝으로서, 추정 프로모터 중 하나인 Aat2(암모늄 수송체, 서열번호 63)로부터 분리된 5'UTR/프로모터를 클로렐라 프로토테코이데스 스테아로일 ACP 불포화 효소 운반 펩티드와 함께 신나모뮴 캠포라 C14 티오에스테라제 구조물로 클로닝하여, 씨. 소로키니아나 글루타메이트 탈수소화 효소 프로모터를 대체한다. 상기 구조물은 서열번호 81로서 열거되어 있다. 추정 프로모터를 시험하기 위해, 티오에스테라제 구조물을 프로토테카 모리포르미스 세포로 형질 전환시켜 전술된 방법을 사용하여 낮은/무 질소 조건 하에서 C14/C12 지방산 증가에 대해 스크리닝함으로써 실제 프로모터 활성을 확인하였다. cDNA/게놈 스크리닝으로부터 분리된 추정 질소 조정된 프로모터의 유사한 테스트는 동일한 방법을 사용하여 수행할 수 있다.

[0556] cDNA/게놈 스크리닝으로부터 분리된 기타 다른 추정 질소 조절된 프로모터/5' UTR은 다음과 같다.

[0557]	프로모터/5'UTR	서열 번호	증가 배수
[0558]	FatB/A 프로모터/5'UTR	서열 번호 55	n/a
[0559]	NRAMP 금속 운반체 프로모터/5'UTR	서열 번호 56	9.65
[0560]	플랩 플라젤라 연관 단백질 프로모터/5'UTR	서열 번호 57	4.92
[0561]	SulfRed 아황산염 환원 효소 프로모터/5'UTR	서열 번호 58	10.91
[0562]	SugT 당 운반체프로모터/5'UTR	서열 번호 59	17.35
[0563]	Amt03-암모늄 운반체 03 프로모터/5'UTR	서열 번호 60	10.1
[0564]	Amt02-암모늄 운반체 02 프로모터/5'UTR	서열 번호 61	10.76
[0565]	Aat01-아미노산 운반체01 프로모터/5'UTR	서열 번호 62	6.21
[0566]	Aat02-아미노산 운반체02 프로모터/5'UTR	서열 번호 63	6.5
[0567]	Aat03-아미노산 운반체03 프로모터/5'UTR	서열 번호 64	7.87
[0568]	Aat04-아미노산 운반체04 프로모터/5'UTR	서열 번호 65	10.95
[0569]	Aat05-아미노산 운반체05 프로모터/5'UTR	서열 번호 66	6.71

[0570] 증가 배수란, 질소 함량이 낮은 배지 중에서 24시간 동안 배양한 후 cDNA 존재량의 증가 배수를 말한다.

[0571] 이와 같은 추정 프로모터/5'UTR의 잠재적 조절에 대하여 더욱 자세히 관찰하기 위해서, 서열들 중 8개를 선택하여 (1) FatB/A; (2) SulfRed 아황산염 환원 효소; (3) SugT 당 운반체; (4) Amt02-암모늄 운반체 02; (5) Aat01-아미노산 운반체 01; (6) Aat03-아미노산 운반체 03; (7) Aat04-아미노산 운반체 04; 및 (8) Aat05-아미노산 운반체 05에 대해 추가로 테스트하였다. 다양한 시점, 즉 T0(종자); 종자 접종 후 20시간; 32시간; 48시간; 62시간; 및 114시간의 시점에서 프로토테카 모리포르미스 세포로부터 분리된 RNA를 대상으로 일루미나

(Illumina) 서열 결정 판독 결과를 이용하는 고 해상도 전사체 분석법을 수행하였다. T0(종자)의 배지에는 질소가 풍부하였으며, 20시간 경과시로부터 그 이후의 배지에는 질소가 약간 함유되었거나 질소를 아예 함유하지 않았다. 그 다음, 각 시점에서 분리된 RNA로부터 조립 생산된 전사체 콘틱(transcript contig)을, 미리 동정한 8개의 전사체 각각과 독립적으로 블라스팅(blasting)하였다. 그 결과는 하기 표 30에 요약되어 있다.

표 30

8개의 추정 프로모터/5' UTR의 전사체 발현 프로필

[0572]

cDNA		T5	T20	T32	T48	T62	T114
aa trans_01	절대	98	96	321	745	927	1300
	상대	1	0.98	3.28	7.61	9.47	13.28
aa trans_03	절대	7	21	51	137	102	109
	상대	1	2.95	7.2	19.42	14.47	15.45
aa trans_04	절대	1	6	25	90	131	160
	상대	1	5.16	21.29	74.97	109.35	133.31
aa trans_05	절대	109	88	123	210	214	273
	상대	1	0.81	1.13	1.93	1.97	2.51
아몬 trans_02	절대	683	173	402	991	1413	1397
	상대	1	0.25	0.59	1.45	2.07	2.04
fatA/B-1_cDNA	절대	13	36	654	617	544	749
	상대	1	2.8	51.57	48.65	42.9	59.1
sug trans_01	절대	25	25	106	261	266	251
	상대	1	1	4.22	10.4	10.63	10
아황산염환원효소_01	절대	634	238	138	145	163	155
	상대	1	0.38	0.22	0.22	0.26	0.24

[0573] 상기 요약된 결과들로부터, 전사체들 중 몇몇은 시간이 경과함에 따라서 축적량이 증가하였음을 알 수 있었으나, 흥미롭게도 아황산염 환원 효소 mRNA는 시간이 경과함에 따라서 mRNA 축적량이 눈에 띄게 감소하였다.

[0574] 이와 같은 8개의 추정 프로모터/5'UTR 영역들을 자체의 천연 운반 펩티드가 빠지고 클로렐라 프로토테코이데스(UTEX 250) 스테아로일 ACP 불포화 효소 유래 운반 펩티드로 치환된 씨.캠포라 티오에스테라제 암호화 영역의 상류에 클로닝하였다. 각각의 추정 프로모터/5'UTR 영역 구조물을, 6S 게놈 서열 유래 DNA를 사용하는 상동성 재조합을 통하여 프로토테카 모리포르미스 UTEX 1435에 도입하였다. 또한, 이 구조물 내에는 수크로스 함유 배지/평판 상 양성 클론 선택용 에스.세레비지아에 유래 suc2 수크로스 인버타제 유전자가 함유되어 있었다. Aat01에 대한 구조물의 관련 부위들에 대한 cDNA 서열은 서열 목록 중 서열 번호 67로서 나열하였다. 기타 다른 구조물에 있어서, 동일한 골격이 사용되었는데, 단지 추정 프로모터/5'UTR 서열만이 가변적이었다. 씨.레인하르드티 베타 튜불린 프로모터가 씨.캠포라 티오에스테라제 유전자 발현을 작동시키는데 사용된 부가의 제어 유전자 이식 균주를 생산하였다. 이 프로모터는 관심 유전자의 항상적 발현을 작동하는 것으로 보였으므로, 테스트된 다양한 추정 N-조절 프로모터/5'UTR에 의해 작동될 때 동일한 티오에스테라제 메시지 발현을 측정함에 있어서 유용한 제어 수단을 제공하였다.

[0575] 일단 유전자 이식 클론이 생산되면, 3개의 독립적 실험을 수행하였다. 처음 2개의 실험에서는 배양액 상청액 중 정상 상태 티오에스테라제 mRNA 수준(RT-PCR로 측정), 지방산 프로필 및 암모니아 수준을 측정함으로써 8개의 추정 프로모터 모두의 잠재적 질소 조절 가능성을 평가하였다. 처음에 클론을 28℃에서 교반하면서(200rpm) 질소 풍부 종자 배지(1g/L 질산 암모늄 - 15mM 질소(암모니아), 4g/L 효모 추출물) 중에서 24시간 내지 48시간 동안 생육하였으며, 이때, 질소 저 함유 배지(0.2g/L 황산 암모늄 - 3mM 질소(암모니아), 0.2g/L 효모 추출물) 50㎖를 접종하는데에는 200D 유닛(A750)을 적용하였다. 총 6일 동안 24시간 마다 세포의 샘플을 취하되, 저 질소 조건에 대해 스위칭(switching)하기 직전에 샘플을 수집하였다. 이후, 각각의 샘플로부터 취한 세포 일부를 총 RNA 추출(트리졸(Trizol) 시약 사용)에 사용하였다(제조자가 제안하는 방법에 따름). 암모니아 분석 결과, 저 질소 배지 중에서 24시간 경과하였을 때 상청액 중 암모니아 수준이 검출 한계(약 100μM) 이하로 강하되었음을 알 수 있었다.

[0576] 실시간 RT-PCR에 있어서, 모든 RNA 수준을 각 시점에서 프로토테카 모리포르미스(UTEX 1435) 내에서 발현된 내부 제어 RNA의 수준에 대해 정규화하였다. 상기 내부 제어 RNA(cd189라 칭함)는 N-아세틸 오르니틴 아미노기 전이 효소를 암호화하는 ARG9 유전자의 생성물이다. 이 실험에 있어서 실시간 RT-PCR에 사용된 프라이머 세트는 다음과 같았다:

다음과 같은 것들과 특이적인 유전자	프라이머 서열 5' -3'	서열 번호
씨.캠포라 TE 정방향	TACCCCGCCTGGGGCGACAC	서열 번호 68
씨.캠포라 TE 역방향	CTTGCTCAGGCGGCGGGTGC	서열 번호 69
cd189 정방향	CCGGATCTCGGCCAGGGCTA	서열 번호 70
cd189 역방향	TCGATGTCGTGCACCGTCGC	서열 번호 71

[0577]

[0578] 각 시점에서 얻은 형질 전환체 각각의 지질 프로필도 구하여, 이를 RT-PCR 결과와 비교하였다. 암모니아 수준, RT-PCR 결과 및 C12 내지 C14 지방산 수준 변화를 바탕으로 하였을 때, 아미노산 운반체 01(Aat-01), 아미노산 운반체 04(Aat-04) 및 암모늄 운반체 02(Amt-02) 서열은 기능상 질소 조절 가능 프로모터/5'UTR을 함유하였다는 결론을 내릴 수 있었다.

[0579] RT-PCR 결과를 통하여 보았을 때, Aat-01은 정상 상태 씨.캠포라 티오에스테라제 mRNA 수준을, 대조군(씨.레인하르드티 베타 튜불린 프로모터)의 경우보다 4배 이하로 높게 만드는 기능이 있음이 입증되었다. mRNA 수준은 또한 질소 제한과, 눈에 띄는 C12 내지 C14 지방산 수준 증가와 상관되었다. 이와 같은 결과들은, Aat-01 프로모터 연관 5'UTR이 지질 생합성시 단백질 합성을, 대조군 씨.레인하르드티 프로모터의 경우보다 효율적으로 작동시킬 수 있다는 것을 입증하였다. Aat-01 프로모터와 같이, Aat-04 프로모터는, 씨.레인하르드티 제어 프로모터의 경우보다 mRNA를 5배 이하 정도로 많이 축적시킬 수 있었다. 그러나, Aat-04 프로모터 구조물은 C12 내지 C14 지방산 수준에 중간 정도로 영향을 미칠 수 있는 능력을 가지게 되었다. 이러한 데이터를 통하여, Aat-04 프로모터는 질소 고갈에 의해 명확히 조절 가능하지만, 프로모터 연관 UTR은 번역 인핸서로서의 기능이 떨어지는 경향이 있음이 입증되었다. 마지막으로, Amt-02 프로모터는, 대조군 프로모터의 경우보다 mRNA를 3배 이하로 많이 축적시킬 수 있다는 점에서 Aat-01 프로모터와 유사하였다. mRNA 수준은 또한 질소 제한, 눈에 띄는 C12 내지 C14 지방산 수준 증가와 상관되었다. 이러한 사실들을 종합하여 보면, 이와 같은 프로모터들 3개는 모두 질소에 의해 조절된다는 것이 입증되었다.

[0580] **B. 암모늄 운반체 3(amt03) 프로모터 및 다양한 티오에스테라제 발현에 관한 추가 특성 규명**

[0581] 전술한 바와 같이, 부분 cDNA(암모늄 운반체 02 및 03(amt02 및 amt03)이라 칭함)를 동정하였다. 상기 2개의 부분 cDNA와 함께, 제3의 부분 cDNA(암모늄 운반체 01(amt01)이라 칭함)도 동정하였다. 부분 cDNA 및 추정 번역 아미노산 서열의 배열을 비교하였다. 그 결과들은, amt01은 상기 3개의 서열들과의 관련성이 더욱 낮았으며, amt02 및 amt03은 단지 하나의 아미노산만이 상이하다는 것을 보여주었다.

[0582] 전술한 바와 같이, 처음에, 부분 cDNA 서열을 로쉬 454 게놈 DNA 조립체 및 일루미나 전사체 조립체에 대해 블라스팅함으로써 컴퓨터로 프로모터/5'UTR을 구성하였다. amt01, amt02 및 amt03을 암호화하는 cDNA와 상동성을 보이는 전사체 접합체를 동정하였으나, 전사체 접합체들은 상기 3개의 mRNA들에 의해 공유되는 서열을 함유하였으므로, 이 전사체 접합체 3개의 mRNA들간에는 차이가 없었다. 로쉬 454 게놈 DNA 조립체는 amt02 및 amt03 cDNA 서열에 대해 힛트(hit)를 부여하였으며, N-말단 단백질 서열을 함유하였다. PCR을 수행하여 5' 측접 영역을 클로닝하였다. 클론 amt02 및 amt03 프로모터/UTR을 입증하는데 사용된 PCR 프라이머는 다음과 같았다:

[0583] Amt03 정방향: 5'-GGAGGAATTCGGCCGACAGGACGCGCGTCA-3'(서열 번호 85)

[0584] Amt03 역방향: 5'-GGAGACTAGTGGCTGCGACCGGCCTGTG-3'(서열 번호 86)

[0585] Amt02 정방향: 5'-GGAGGAATTCTCACCAGCGGACAAAGCACCG-3'(서열 번호 87)

[0586] Amt02 역방향: 5'-GGAGACTAGTGGCTGCGACCGGCCTCTGG-3'(서열 번호 88)

[0587] 상기 2개의 경우에 있어서, 5' 및 3' 프라이머는, 상기와 같은 프로모터/5'UTR 영역의 기능성을 입증하기 위해 발현 벡터에 예상 클로닝하는데 유용한 제한 위치들을 함유하였다.

[0588] 이와 같이 컴퓨터와 PCR계 방법을 통하여 합체되어 클로닝된 DNA와, amt02(서열 번호 61) 및 amt03(서열 번호

60)을 암호화하는 원 cDNA 간에 쌍 정렬(pair wise alignment)을 수행하였다. 이 정렬의 결과를 통하여, 원 cDNA와, 클로닝된 게놈 서열들 간 상당한 차이가 확인되었는데, 이는 곧, 암모늄 운반체들은 다양한 유전자 군을 대표할 것이라는 것을 암시하는 것이다. 뿐만 아니라, amt03에 대한 병행법을 바탕으로 제조된 프로모터/5'UTR 클론은, 원 amt03 서열과 상이하였던 반면에, amt02 서열과는 동일하였다. amt03 프로모터/UTR 서열(서열 번호 89)의 특성을 규명하는 추가의 실험이 이하에 기술된 바와 같이 수행되었다.

[0589] 추정 프로모터/UTR 서열을 클로닝하여, 4개의 상이한 티오에스테라제를 발현시킴으로써, 상기 동정된 amt03 프로모터//UTR 서열(서열 번호 89)을 테스트하였다. 발현 카세트는 게놈의 6S 위치에 상동성인 상류 및 하류 재조합 서열(각각 서열 번호 82 및 84)을 함유하였다. 카세트는 또한 수크로스 함유 배지에서 양성 클론들을 선택할 수 있도록 하는 에스.세레비지아에 SUC2 수크로스 인버타제 cDNA를 함유하였다. 수크로스 인버타제 발현은 씨.레인하르드티 베타 튜불린 프로모터에 의해 작동되었으며, 이는 또한 씨.불가리스 질산염 환원 효소 3'UTR이 관여하였다. 그 다음, amt03 프로모터/UTR 서열은, 수크로스 인버타제 카세트 하류와, 다음과 같은 티오에스테라제 유전자 4개중 하나로부터 유래하는 인프레임 티오에스테라제 cDNA 서열 사이에 클로닝되었으며, 상기 amt03 프로모터/UTR 서열은 또한 씨.불가리스 질산염 환원 효소 3'UTR을 함유하였다: (1) 씨.캠포라 유래 C14 티오에스테라제; (2) 유.캘리포니카 유래 C12 티오에스테라제; (3) 유.아메리카나 유래 C10 내지 C16 티오에스테라제; 또는 (4) 씨.후커리아나 유래 C10 티오에스테라제. C14 씨.캠포라 티오에스테라제, C12 유.캘리포니카 티오에스테라제, 그리고 C10 내지 C16 유.아메리카나는 모두, 클로렐라 프로토테코이데스 스테아로일 ACP 불포화 효소로부터 유래하는 운반 펩티드를 함유하였다. C10 씨.후커리아나 티오에스테라제는 프로토테카 모리포르미스 델타 12 지방산 불포화 효소(FAD) 유래 운반 펩티드를 함유하였다. 모든 경우에서, 서열들은 프로토테카 모리포르미스 내 발현에 최적화된 코돈이었다. 상기 티오에스테라제 구조물들의 서열을 이하 서열 목록에 기술하였다:

[0590] amt03 프로모터/UTR::씨.캠포라 티오에스테라제 구조물 서열 번호 90

[0591] 씨.캠포라 티오에스테라제 구조물 서열 번호 91

[0592] 유.캘리포니카 티오에스테라제 구조물 서열 번호 92

[0593] 유.아메리카나 티오에스테라제 구조물 서열 번호 93

[0594] 씨.후커리아나 티오에스테라제 구조물 서열 번호 94

[0595] 상기 실시예 2에 기술된 바와 같이, 야생형 프로토테카 모리포르미스 세포로의 유전자 총 형질 전환법을 통하여 유전자 이식 계열을 생산하였으며, 이를 수크로스 함유 평판/배지에서 선택하였다. 이후, 양성 계열을, 자체의 지방산 프로필이 변한 정도에 대해 스크리닝하였다. 그 다음, 4개의 계열들(이 계열들 중 하나는 전술한 4개의 구조물 각각으로 형질 전환하여 생산)을 대상으로 추가 분석을 수행하였다. 계열 76은 씨.캠포라 C14 티오에스테라제를 발현하였으며, 계열 37은 유.캘리포니카 C12 티오에스테라제를 발현하였고, 계열 60은 유.아메리카나 C10 내지 C16 티오에스테라제를 발현하였으며, 또한 계열 56은 씨.후커리아나 C10 티오에스테라제를 발현하였다. 각각의 계열을 유일한 탄소원으로서 수크로스를 함유하는 배지 중에서 48시간 동안 생육하였으며, 14시간, 24시간, 36시간 및 48시간 경과시 세포 샘플(종자 배양액)을 취하여, 지방산 메틸 에스테르로의 직접 에스테르 결합 전이 반응을 통해 지방산 프로필을 측정하고 나서, GC-FID(전술한 바와 같음)로 분석하였으며, 또한 총 RNA를 분리하였다. 48시간 경과 시점이 거의 끝나갈 즈음, 이 세포들을 질소를 함유하지 않거나 낮은 수준으로 함유하는 배양액(이때, 이 배양액은 유일한 탄소원으로서 수크로스를 함유하였으며, 또한 pH는 5.0(시트레이트 완충, 최종 농도 = 0.05M) 또는 7.0(HEPES 완충, 최종 농도 = 0.1M)으로 유지됨)에 접종하였다. (지질을 생산함에 있어서) 12시간, 24시간, 72시간 및 108시간 경과시 배양액 샘플을 취하여 지방산 프로파일링을 수행하였으며, 또한 총 RNA를 분리하였다. 상기 배양액의 암모니아 검정법을 통하여, 질소를 낮은 수준으로 함유하는 배지 중에서 24시간 배양하였을 때의 암모니아 수준은 검출 한계(약 100μM) 이하로 떨어졌음을 알 수 있었다.

[0596] 상기와 같이 샘플을 수집한 각각의 시점에 취한 총 RNA를 대상으로 하여, 티오에스테라제의 mRNA 수준에 대한 실시간 RT-PCR 검정법을 수행하였으며, 이때 구하여진 모든 mRNA 수준은 내부 대조군 RNA(cd189)의 수준에 대해 정규화하였다. 실시간 PCR에 사용된 프라이머 세트를 이하 표 31에 나타내었다.

표 31

실시간 PCR용 프라이머 세트

이하에 특이적인 유전자	프라이머 서열 5'-3'	서열 번호
씨.캠포라 TE정방향	TACCCCGCCTGGGGCGACAC	서열 번호 68
씨.캠포라 TE역방향	CTTGCTCAGGCGGCGGGTGC	서열 번호 69
유.캘리포니카TE 정방향	CTGGGCGACGGCTTCGGCAC	서열 번호 95
유.캘리포니카TE 역방향	AAGTCGCGGCGCATGCCGTT	서열 번호 96
유.아메리카나TE 정방향	CCCAGCTGCTCACCTGCACC	서열 번호 97
유.아메리카나TE 역방향	CACCCAAGGCCAACGGCAGCGCCGTG	서열 번호 98
씨.후커리아나 TE정방향	TACCCCGCCTGGGGCGACAC	서열 번호 99
씨.후커리아나 TE역방향	AGCTTGGACAGGCGGCGGGT	서열 번호 100
cd189 역방향	TCGATGTCGTGCACCGTCGC	서열 번호 71
cd189 정방향	CCGGATCTCGGCCAGGGCTA	서열 번호 70

[0597]

[0598] 종자 배양 단계의 각 시점에서의 지방산 프로필 결과들을 통하여, 티오에스테라제의 영향력이 매우 미약함을 알 수 있었다. 지질 생산 단계 개시시, 지방산 프로필은 상당히 영향을 받았는데, 이 경우 이 지방산 프로필은 pH 5.0인 배양액의 경우와 비교하였을 때 pH 7.0로 유지되는 배양액에서는 훨씬 극적으로 증가하였다. pH 7.0 및 5.0 표적 지방산 축적량 간 차이는 테스트된 각각의 티오에스테라제에 따라서 다양하였으나, 전체 효과는 동일하였으며, pH 5.0에서 생육된 세포들은 축적된 표적 지방산 수준이 상당히 낮았으나, 대조군 야생형 세포보다는 상기 축적된 표적 지방산 수준은 더욱 높았다.

[0599] 티오에스테라제 각각에 대한 정상 상태 mRNA 수준에 대한 배양액 pH의 영향력은 명확하였던 것으로 보아, 이와 같이 매우 밀접하게 상관되어 있는 샘플들로부터 분리된 RNA의 분석 결과는, 지방산 프로필 데이터에 의존적일 것이다. 지방산 축적 데이터와 mRNA 데이터를 종합하였을 때, amt03 프로모터/UTR에 의해 작동되는 티오에스테라제 유전자 발현에 있어서 pH 조절은, 전사 수준 또는 mRNA 안정성중 하나, 또는 이것들 둘 다에 의해 조율되었음을 알 수 있었다. 뿐만 아니라, 유.캘리포니카 mRNA의 정상 상태 수준은, 씨.후커리아나 mRNA의 정상 상태 수준과 비교하였을 때, 4로그만큼 낮았다. 이와 같은 관찰 결과는, 개별 mRNA 서열이 발현을 제어하는 역할을 할 수 있다는 가설과 일치하였다. 이러한 데이터는, 운반체의 amt03 군에 의해 프로토테카 모리포르미스 내에서의 암모늄 흡수가 pH와 직접적으로 상관되어 있음을 암시하는 것이다.

[0600] 프로토테카 모리포르미스 세포를, 유.아메리카나 C10 내지 C16 티오에스테라제를 발현하는 amt03 프로모터/UTR 구조물로 형질 전환하여 생산된 계열들 12개를 대상으로 부가의 지방산 프로필 분석을 수행하였다. 전술한 계열 60은 다음과 같은 분석법의 일부였다. 이하 표 32는, 야생형 대조군과 함께 분석되었던 12개의 계열들 중 3개의 지질 프로필을 보여주는 것이다.

표 32

amt03 프로모터/UTR에 의해 촉발되는 유.아메리카나 TE 함유 형질 전환체의 지방산 프로필

면적%	C8:0	C10:0	C12:0	C14:0	C16:0	C18:0	C18:1	C18:2	총포화물
야생형	0.00	0.01	0.04	1.27	27.20	3.85	58.70	7.18	32.36
계열 40	2.38	20.61	3.41	28.41	29.92	1.91	8.57	3.74	86.64
계열 44	1.50	20.16	4.44	31.88	26.66	1.88	6.95	5.42	86.50
계열 60	0.98	14.56	3.15	27.49	31.76	2.14	12.23	6.36	80.06

[0601]

[0602] 상기 표에 나타낸 바와 같이, 총 포화물의 수준은 야생형과 비교하였을 때, 계열 40의 경우 야생형의 총 포화물 수준에 비하여 2.6배 이상 극적으로 증가하였다(분석된 12개의 계열들의 총 포화물 수준은 약 63% 내지 86%였음). 뿐만 아니라, 유.아메리카나 티오에스테라제가 상기와 같은 수준으로 발현될 때, 이 효소는 불포화물(특히 C18:1 및 C18:2) 수준을 극적으로 감소시켰는데(계열 40 및 44 참조), 이 경우, 계열 44에서 C18:1 수준

은 야생형의 경우와 비교하였을 때 8배 이상 감소하였다. 뿐만 아니라, 유.아메리카나 티오에스테라제(amt03 프로모터에 의해 발현이 촉발됨)는 중간쇄 길이의 지방산 수준을 상당히 증가시켰다. 계열 44는, C10:0 내지 C14:0 수준이 56% 이상(야생형 균주에서 볼 수 있는 수준보다 약 42배 높음)이었으며, C8:0 내지 C14:0 수준은 57% 이상이었다. 유.아메리카나 티오에스테라제를 발현하는 Amt03 프로모터 구조물로 형질 전환된 부가의 균주들의 대표적인 지질 프로필은, C8:0 0.23%; C10:0 9.64%; C12:0 2.62%; C14:0 31.52%; C16:0 37.63%; C18:0 5.34%; C18:1 7.05%; 및 C18:2 5.03%(총 포화물 백분율 = 86.98%)이었다.

[0603] 프로토테카 모리포르미스 세포를 구조물 amt03 프로모터/UTR(서열 번호 89)(씨.후커리아나 C10 티오에스테라제(서열 번호 94)를 발현시킴)로 형질 전환하여 부가 지질 프로필을 구하였다. 이 구조물을 발현하는 양성 클론을 선택하여 pH 7.0 조건 하에서 생육하였다. 양성 클론으로부터 얻은 대표적인 지질 프로필은, C8:0 9.87%; C10:0 23.97%; C12:0 0.46%; C14:0 1.24%; C16:0 10.24%; C18:0 2.45%; C18:1 42.81%; 및 C18:2 7.32%이었다. 이 클론의 C8 내지 C10 백분율은 33.84%였다.

[0604] 이와 같은 데이터들을 종합하였을 때, 이 데이터들은, amt03 프로모터/UTR과 기타 다른 유사 프로모터는 조밀하게 조절되는 프로모터(잠재적으로 독성인 화합물을 발현하고, 유전자 발현을 엄격하게 수행해야 할 때 특히 유용할 수 있음)로서 사용될 수 있음을 암시하고 있다. 프로토테카 모리포르미스가 넓은 pH 범위(pH 5.0 이상 내지7.0)에서 생장할 수 있는 능력은, 이 유기체를 조절 요소, 예를 들어 amt03 프로모터/UTR과 함께 사용되기에 특히 유용하게 만든다. 뿐만 아니라, 상기 지질 프로필 데이터는, amt03 프로모터/UTR이 유전자를 발현하는 인상적인 능력을 입증해주는 것이다.

[0606] **실시예 6: 미세 조류 프로토테카 모리포르미스 내 포화 지방산 수준의 변경**

[0607] **A. 유전자 녹아웃 접근법에 의한, 스테아로일 ACP 불포화 효소 및 델타 12 지방산 불포화 효소 발현의 감소**

[0608] 생물 정보학을 바탕으로 하는 접근법(cDNA 대상)을 사용하는 유전체학적 스크리닝의 일환으로서, 프로토테카 모리포르미스(UTEX 1435) 유래 게놈 DNA의 로쉬 454 서열 결정 및 일루미나 전사체, 지방산 불포화에 관여하는 유전자의 특정 군 2개를 동정하였다: 스테아로일 ACP 불포화 효소(SAD) 및 델타 12 지방산 불포화 효소(Δ12 FAD). 스테아로일 ACP 불포화 효소는 지질을 합성하는 경로, 예를 들어 C18:0 지방산으로부터 C18:1 지방산을 합성하는 경로에 관여하는 일원으로서, 이중 결합을 지방산 사슬에 도입하는 기능을 갖는다. 델타 12 지방산 불포화 효소는 또한 지질을 합성하는 경로, 예를 들어 C18:1 지방산으로부터 C18:2 지방산을 합성하는 경로에 관여하는 일원으로서, 이미 불포화된 지방산에 이중 결합을 도입하는 기능을 갖는다. 생물 정보학적 시도가 행해질 때 동정된 지방산 불포화 효소 유전자 군 2개를 바탕으로 하는 프로브를 사용하는 서던 블럿 분석은, 불포화 효소 유전자 각 군은 다수의 군 일원들로 이루어져 있을 것임을 말해주었다. 뿐만 아니라, 스테아로일 ACP 불포화 효소를 암호화하는 유전자들은 2개의 별도 군들에 속하였다. 이와 같은 결과를 토대로 하여, 3개의 유전자 중단 구조물을, 불포화 효소 각 군에 고도로 보존된 암호화 영역을 표적화함으로써, 잠재적으로 다수의 유전자 군 일원들을 중단하도록 설계하였다.

[0609] (1) 델타 12 지방산 불포화 효소(d12FAD) 군 일원의 암호화 서열의 고도로 보존된 부분들과 (2) SAD의 개별 군 2개 각각을 표적화하는 2개의 구조물(상기 구조물은 각각 상기 각 군으로부터 유래하는 암호화 서열의 보존된 영역을 함유함)을 사용하여 3개의 상동성 재조합 표적화 구조물을 설계하였다. 이와 같은 기법은, 통상적인 유전자 치환 기법보다 선택가능한 마커 유전자(에스.세레비지아에로부터 유래하는 suc2 수크로스 인버타제 카세트로서, 수크로스를 가수 분해하는 능력을 부여함)를 고도로 보존된 암호화 영역(다수의 군에 속하는 일원들을 표적화함)에 더욱 효율적으로 삽입할 것이며, 이 경우, 상동성 재조합은 표적화된 유전자에 측접하는 영역들을 표적화할 것이다.

[0610] 전술한 방법들을 사용하여 유전자 총 형질 전환법에 의해 모든 구조물들을 세포에 도입하였는데 즉, 구조물을 선형화한 다음 세포에 발사하였다. 수크로스 함유 평판/배지에서 형질 전환체를 선택하였으며, 지질 프로필 변화는 전술한 방법을 사용하여 분석하였다. 3개의 표적화 구조물 각각으로부터 유래하는 관련 서열들을 이하에 나열하였다:

[0611] 설 명 서열 번호

[0612] 표적 구조물 유래 d12FAD의 암호화 영역 유래 5' 서열 서열 번호 72

[0613] 표적 구조물 유래 d12FAD의 암호화 영역 유래 3' 서열 서열 번호 73

[0614] d12FAD 표적화 구조물cDNA 서열 서열 번호 74

[0615] SAD2A의 암호화 영역 유래 5' 서열 서열 번호 75

[0616] SAD2A의 암호화 영역 유래 3' 서열 서열 번호 76

[0617] SAD2A 표적화 구조물cDNA 서열 서열 번호 77

[0618] SAD2B의 암호화 영역 유래 5' 서열 서열 번호 78

[0619] SAD2B의 암호화 영역 유래 3' 서열 서열 번호 79

[0620] SAD2B 표적화 구조물 cDNA 서열 서열 번호 80

[0621] 구조물 각각으로 형질 전환하여 얻은, 대표적인 양성 클론들을 골라낸 다음, 이 클론들에 대한 지질 프로필을 측정하여(면적%로 표시) 하기 표 33에 요약하였다.

표 33

불포화 효소 녹아웃시 지질 프로필

지방산	d12FAD KO	SAD2A KO	SAD2B KO	wt UTEX 1435
C8:0	0	0	0	0
C10:0	0.01	0.01	0.01	0.01
C12:0	0.03	0.03	0.03	0.03
C14:0	1.08	0.985	0.795	1.46
C16:0	24.42	25.335	23.66	29.87
C18:0	6.85	12.89	19.555	3.345
C18:1	58.35	47.865	43.115	54.09
C18:2	7.33	10.27	9.83	9.1
C18:3 알파	0.83	0.86	1	0.89
C20:0	0.48	0.86	1.175	0.325

[0622]

[0623] 구조물 각각은 지방산의 원하는 군에 측정 가능한 정도의 영향을 미쳤으며, 3가지 경우 모두에 있어서, 특히 2개의 SAD가 녹아웃되었을 때 C18:0 수준은 눈에 띄게 증가하였다. SAD 녹아웃으로 인하여 생산된 클론들 다수개를 추가로 비교한 결과, SDA2B 녹아웃 계열들은 C18:1 지방산 수준이, SAD2A 녹아웃 계열들에서 관찰되는 C18:1 지방산 수준보다 상당히 많이 감소되었음이 확인되었다.

[0624] 부가의 Δ12 지방산 불포화 효소(FAD) 녹아웃은, 전술한 방법을 사용하였을 때 프로토테카 모리포르미스 배경에서 나타났다. Δ12FAD의 잠재적인 상동성을 확인하기 위해서 추정 FAD를 암호화하는 게놈 영역을 증폭시키기 위한, 다음과 같은 프라이머들을 사용하였다.

[0625] 프라이머 1 5'-TCACTTCATGCCGGCGGTCC-3' 서열 번호 101

[0626] 프라이머 2 5'- GCGCTCCTGCTTGGCTCGAA-3' 서열 번호 102

[0627] 프로토테카 모리포르미스 게놈 DNA를, 상기 프라이머들을 사용하여 게놈 증폭 생성한 서열들은 매우 유사하였는데, 이는 프로토테카 내에 Δ12FAD의 유전자 또는 대립 유전자 다수 개가 존재함을 암시하는 것이었다.

[0628] 이와 같은 결과를 바탕으로 하여, 1개 이상의 Δ12FAD 유전자를 불활성화하고자 하는 2개의 유전자 중단 구조물을 설계하였다. 이때 사용된 기법은, 통상의 유전자 치환 기법이 아니고, 수크로스 인버타제(에스.세레비지아에 유래 suc2) 카세트(선택가능한 마커로서 수크로스를 가수 분해하는 능력 부여)를 고도로 보존된 암호화 영역에 삽입하는 기법이었다. 제1 구조물(pSZ1124로 명명)은 에스.세레비지아에 suc2 유전자와 클로렐라 불가리스 질산염 환원 효소 3'UTR(에스.세레비지아에 suc2 카세트)을 발현하는, 씨.레인하르드티 β-튜불린 프로모터 측접 5' 및 3'게놈표적화 서열을 함유하였다. 제2 구조물(pSZ1125로 명명)은 에스.세레비지아에 suc2 유전자와 클로렐라 불가리스 질산염 환원 효소 3'UTR을 발현하는, 씨.레인하르드티 β-튜불린 프로모터 측접 5' 및 2' 게놈 표적화 서열들을 함유하였다. 이 구조물의 관련 서열을 서열 목록에 나열하였다.

[0629] pSZ1124 (FAD2B) 5' 게놈 표적화 서열 서열 번호 103

[0630] pSZ1124 (FAD2B) 3' 게놈 표적화 서열 서열 번호 104

[0631] 에스.세레비지아에 suc2 카세트 서열 번호 105

[0632] pSZ1125 (FAD2C) 5' 게놈 표적화 서열 서열 번호 106

[0633] pSZ1125 (FAD2C) 3' 게놈 표적화 서열 서열 번호 107

[0634] pSZ1124 및 pSZ1125를 각각 프로토테카 모리포르미스 배경에 도입하고, 수크로스를 가수 분해하는 능력을 바탕으로 양성 클론을 선택하였다. 표 34는 pSZ1124 및 pSZ1125 표적화 벡터를 이용한 유전자 이식 계열 2개의 지질 프로필(면적%, 전술한 방법을 사용하여 구함)을 요약한 것이다.

표 34

[0635] Δ12 FAD 녹아웃시 지질 프로필

	C10:0	C12:0	C14:0	C16:0	C16:1	C18:0	C18:1	C18:2	C18:3α
부모	0.01	0.03	1.15	26.13	1.32	4.39	57.20	8.13	0.61
FAD2B	0.02	0.03	0.80	12.84	1.92	0.86	74.74	7.08	0.33
FAD2C	0.02	0.04	1.42	25.85	1.65	2.44	66.11	1.39	0.22

[0636] FAD2B(pSZ1124) 구조물을 함유하는 유전자 이식체는 지질 프로필(C18:2 수준)에 있어서 매우 이색적이면서 예상외의 결과를 보였는데, 즉, 상기 C18:2 수준은 감소하되, 단지 약 1 면적%까지만 감소하는 것으로 예상되었다. 그러나, C18:1 지방산 수준은, 거의 배타적으로 C16:0 수준을 상당히 낮춰가면서 상당히 증가하였다. FAD2C(pSZ1125) 구조물을 함유하는 유전자 이식체는 또한 지질 프로필도 변화하였는데; C18:2의 수준은, C18:1 수준이 증가하였던 것과는 달리, 상당히 감소하였다.

[0637] *소 우지 모의체*

[0638] 전술한 바와 같은 SAD2B 녹아웃 실험으로부터 생산된 하나의 양성 클론으로서, C14-선호 지방 아실-ACP 티오에스테라제 유전자를 추가로 도입하기 위한 배경으로 사용될 클론을 선택하였다. 씨.캠포라 C14 선호 티오에스테라제를 도입하는데 사용되는 구조물은, 6S 게놈 영역에 대한 표적화 서열(상동성 재조합을 통해 형질 전환 DNA를 표적 통합할 수 있음)을 함유하였으며, 발현 구조물은, neoR 유전자를 발현하는 씨.레인하르드티 β-튜불린 프로모터와 클로렐라 불가리스 질산염 환원 효소 3'UTR, 그리고 이것들 다음에는 코돈 최적화 씨.캠포라 티오에스테라제를 발현하는 제2 씨.레인하르드티 β-튜불린 프로모터와 클로렐라 프로토테코이데스 스테아로일 ACP 불포화 효소 운반 펩티드(제2 클로렐라 불가리스 질산염 환원 효소 3'UTR 함유)를 함유하였다. 5'6S 게놈 공여 서열은 서열 번호 82에 나열하였고; 3'6S 게놈 공여 서열은 서열 번호 84에 나열하였으며; 씨.캠포라 티오에스테라제와 관련된 발현 구조물은 서열 번호 83에 나열하였다.

[0639] 전술한 바와 같은 유전자 총 방법을 사용하여 형질 전환을 수행하였으며, 이때, 세포를 2% 수크로스 함유 평판 상에서 24시간 동안 회수하였다. 이 시간이 경과한 후, 세포를 재현탁하고 나서, 2% 수크로스와 50㎍/㎖ G418(선택용)을 함유하는 평판 상에 재도말하였다. 생산된 양성 클론들 중 9개의 클론들을 지질 생산 및 지질 프로필 분석용으로 선택하였다. 9개의 유전자 이식 클론(SAD2B KO를 함유하며, 씨.캠포라 C14 선호 티오에스테라제를 발현함)을 전술한 바와 같이 배양한 다음, 지질 프로필에 대해 분석하였다. 그 결과들은 하기 표 35에 요약되어 있다. 우지에 대한 지질 프로필도 이하 표 35에 제시하였다(National Research Council 1976: Fat Content and Composition of Animal Product).

표 35

티오에스테라제 형질 전환된 클론의 지질 프로필

	C10:0	C12:0	C14:0	C16:0	C16:1	C18:0	C18:1	C18:2	C18:3	C20
SAD2BKO 씨.캠포라 TE 클론 1	0.01	0.33	6.13	24.24	0.19	11.08	42.03	13.45	0.98	0.73
SAD2BKO 씨.캠포라 TE 클론 2	0.01	0.16	3.42	23.80	0.40	9.40	50.62	10.2	0.62	0.70
SAD2BKO 씨.캠포라 TE 클론 3	0.01	0.20	4.21	25.69	0.40	7.79	50.51	9.37	0.66	0.63
SAD2BKO 씨.캠포라 TE 클론 4	0.01	0.21	4.29	23.57	0.31	9.44	50.07	10.07	0.70	0.70
SAD2BKO 씨.캠포라 TE 클론 5	0.01	0.18	3.87	24.42	0.32	9.24	49.75	10.17	0.71	0.71
SAD2BKO 씨.캠포라 TE 클론 6	0.01	0.28	5.34	23.78	0.33	9.12	49.12	10.00	0.68	0.70
SAD2BKO 씨.캠포라 TE 클론 7	0.01	0.15	3.09	23.07	0.32	10.08	51.21	10.00	0.66	0.74
SAD2BKO 씨.캠포라 TE 클론 8	0.01	0.29	5.33	24.62	0.37	7.02	49.67	10.74	0.69	0.70
SAD2BKO 씨.캠포라 TE 클론 9	0.01	0.12	2.74	25.13	0.30	10.17	50.18	9.42	0.71	0.71
wt UTEX 1435	0.01	0.02	0.96	23.06	0.79	3.14	61.82	9.06	0.46	0.27
SAD2BKO	0.01	0.03	0.80	23.66	0.13	19.56	43.12	9.83	1.00	1.18
우지	0.00	0.00	4.00	26.00	3.00	14.00	41.00	3.00	1.00	0.00

[0640]

[0641] 상기 표 35에서 살펴볼 수 있는 바와 같이, 유전자 이식 계열의 지질 프로필은 우지의 지질 프로필과 매우 유사하였다. 종합하여 보면, 상기 데이터는 특이적 유전자 이식 배경을, (이와 같은 경우에는 SAD2B 녹아웃) C14 선호 티오에스테라제(씨.캠포라 유래)와 통합하여, 우지의 지질 프로필과 유사한 지질 프로필을 나타내는 오일을 생산하는 유전자 이식 조류 균주를 생산하는 것의 이점을 입증하였다.

[0642] **B. 프로토테카 세포 내 델타 12 불포화 효소 유전자(FADc)의 하향 조절에 대한 RNAi 접근법**

[0643] RNAi에 의해 FADc 유전자(델타 12 불포화 효소 유전자) 발현을 하향 조절하는 벡터를 프로토테카 모리포르미스 UTEX 1435 유전적 배경에 도입하였다. 사카로마이세스 세레비지아에 suc2 수크로스 인버타제 유전자를 선택가능한 마커(유일한 탄소원으로서 수크로스를 사용하여 생육할 수 있는 능력을 양성 클론에 부여)로서 사용하였다. 제1 유형의 구조물은, 자체의 제1 인트론에 대해 cis 배열로 결합된 FADc 암호화 영역의 제1 엑손의 일부와, 이것의 다음에 존재하는 제1 엑손의 역방향 반복 단위를 이용하였다. 이와 같은 유형의 구조물은 통상적으로 mRNA로서 발현될 때 이론상 헤어핀 RNA로 형성되었다. 상기 제1 유형의 구조물 2개를 제조하였는데, 이것들 중 하나는 프로토테카 모리포르미스 Amt03 프로모터(서열 번호 89)(pSZ1468로 명명)에 의해 작동되었으며, 제2 구조물은 클라미도모나스 레인하르드티 β-튜불린 프로모터(서열 번호 114)(pSZ1469로 명명)에 의해 작동되었다. 제2 유형의 구조물은, 프로토테카 모리포르미스 Amt03 프로모터(서열 번호 89)(pSZ1470으로 명명)에 의해 작동되거나, 또는 클라미도모나스 레인하르드티 β-튜불린 프로모터(서열 번호 114)(pSZ1471로 명명)에 의해 작동되는, 크기가 큰 FADc 엑손 2(안티센스 배향)를 이용하였다. 4개의 구조물 모두는, 6S 게놈 영역에 대해 상동성인 에스.세레비지아에 suc2 수크로스 인버타제 카세트(서열 번호 159) 및 5'(서열 번호 82) 및 3'(서열 번호 84) 재조합 표적화 서열(구조물에 측접)(핵내 게놈으로의 통합에 사용)을 가졌다. 각 구조물의 관련 부분들과 함께 각 RNAi 구조물의 FADc 부분들의 서열을 다음과 같이 서열 목록으로 나열하였다.

[0644] 설 명 서열 번호

[0645] pSZ1468 FADc RNAi 헤어핀 카세트 서열 번호 163

[0646] pSZ1468 구조물의 관련 부분 서열 번호 164

[0647] pSZ1469 FADc RNAi 헤어핀 카세트 서열 번호 165

[0648] pSZ1469 구조물의 관련 부분 서열 번호 166

[0649] pSZ1470 FADc 엑손 2 RNAi 카세트 서열 번호 167

[0650] pSZ1470 구조물의 관련 부분 서열 번호 168

[0651] pSZ1471 FADc 엑손 2 RNAi 카세트 서열 번호 169

[0652] pSZ1471 구조물의 관련 부분 서열 번호 170

[0653] 4개의 구조물 각각을 프로토테카 모리포르미스 배경에 도입하여 이를 형질 전환하고, 유일한 탄소원으로서 수크로스를 함유하는 평판을 사용하여 양성 클론들을 스크리닝하였다. 각각의 형질 전환체로부터 양성 클론들을 골라낸 다음, 이것들의 하위 세트를 선택하여, pSZ1468, pSZ1469, pSZ1470 및 pSZ1471에 함유된 안티센스 카세트와 헤어핀 구조의 지방산 프로필에 대한 영향력을 분석하였다. 각각의 형질 전환체로부터 선택된 클론들을 지질 생산 조건 하에서 생육하였으며, 이때, 전술한 바와 같이 직접 에스테르 결합 전이 방법을 사용하여 지질 프로필을 분석하였다. 형질 전환체 각각의 대표적인 지질 프로필은 하기 표 36에 요약되어 있다. 야생형 1 및 2 세포들은, 음성 대조군으로서 형질 전환체들 각각과 별도로 생육되는 비형질 전환 프로토테카 모리포르미스 세포였다.

표 36

델타 12 불포화효소 유전자(FADc)의 발현을 하향 조절하기 위한 RNAi 구조물을 함유하는 프로토테카 모리포르미스 세포의 지질 프로필

[0654]

균주	C10:0	C12:0	C14:0	C16:0	C18:0	C18:1	C18:2
야생형 1	0.01	0.03	1.20	27.08	4.01	57.58	7.81
pSZ1468 클론 A	0.01	0.04	1.33	25.95	3.68	65.60	1.25
pSZ1468 클론 B	0.01	0.03	1.18	23.43	2.84	65.32	4.91
pSZ1468 클론 C	0.01	0.04	1.34	23.18	4.27	63.65	5.17
pSZ1468 클론 D	0.01	0.03	1.24	23.00	3.85	61.92	7.62
pSZ1470 클론 A	0.01	0.03	1.23	24.79	4.33	58.43	8.92
pSZ1470 클론 B	0.01	0.03	1.26	24.91	4.14	57.59	9.64
pSZ1470 클론 C	0.01	0.03	1.21	23.35	4.75	58.52	9.70
야생형 2	0.01	0.03	0.98	24.65	3.68	62.48	6.26
pSZ1469 클론 A	0.01	0.03	1.05	21.74	2.71	71.33	1.22
pSZ1469 클론 B	0.01	0.03	1.01	22.60	2.98	70.19	1.27
pSZ1469 클론 C	0.01	0.03	1.03	19.82	2.38	72.95	1.82
pSZ1469 클론 D	0.01	0.03	1.03	20.54	2.66	70.96	2.71
pSZ1471 클론 A	0.01	0.03	1.03	18.42	2.63	66.94	8.55
pSZ1471 클론 B	0.01	0.03	0.94	18.61	2.58	67.13	8.66
pSZ1471 클론 C	0.01	0.03	1.00	18.31	2.46	67.41	8.71
pSZ1471 클론 D	0.01	0.03	0.93	18.82	2.54	66.84	8.77

[0655] 상기 요약된 결과들을 통하여, 헤어핀 구조물인 pSZ1468 및 pSZ1469는 특이적 예상 표현형(즉, 야생형 1 및 야생형 2 각각과 비교하였을 때, C18:2 지방산 수준은 감소하였으며, C18:1 지방산 수준은 증가함)을 보였음을 알 수 있었다. 안티센스 구조물인 pSZ1470 및 pSZ1471은 C18:2 지방산 수준이 감소하지 않았으나, 그 대신, 야생형 1 및 야생형 2와 비교하였을 때 C18:2 지방산 수준은 약간 증가하였고, C16:0 지방산 수준은 약간 감소하였다.

[0656] **C. 외인성 스테아로일-ACP 불포화 효소의 발현**

[0657] 올레아 유로파에아 스테아로일-ACP 불포화 효소(GenBank 수탁 번호 제AAB67840.1호)를 프로토테카 모리포르미스 UTEX1435 유전적 배경에 도입하였다. 발현 구조물은 6S 게놈 영역에 대한 5'(서열 번호 82) 및 3'(서열 번호 84) 상동성 재조합 표적화 서열(구조물에 측접)을 함유하였는데, 이 재조합 표적화 사슬은 씨.레인하르드티 β-튜불린 프로모터/5'UTR 및 클로렐라 불가리스 질산염 환원 효소 3'UTR의 제어 하에 있는 에스.세레비지아에 suc2 수크로스 인버타제 암호화 영역과 핵내 게놈으로의 통합에 사용되었다. 이와 같은 에스.세레비지아에 suc2 발현 카세트를 서열 번호 159로서 나열하였으며, 이를 선택 마커로 사용하였다. 올레아 유로파에아 스테아로일-ACP 불포화 효소 암호화 영역은 프로토테카 모리포르미스 Amt03 프로모터/5'UTR(서열 번호 89) 및 씨.불가리스

질산염 환원 효소 3'UTR의 제어 하에 두었으며, 천연 운반 펩티드는 클로렐라 프로토테코이데스 스테아로일-ACP 불포화 효소 운반 펩티드(서열 번호 49)로 치환되었다. 코돈 최적화된 cDNA 서열과 아미노산 서열(운반 펩티드가 치환됨)을 서열 목록 중 서열 번호 171 및 서열 번호 172로 각각 나열하였다. 전체 오.유로파에아 SAD 발현 카세트를 pSZ1377이라 명명하고, 이를 프로토테카 모리포르미스 유전적 배경에 도입하여 이 균주를 형질 전환하였다. 양성 클론을, 유일한 탄소원으로서 수크로스를 함유하는 평판 상에서 스크리닝하였다. 양성 클론들의 하위 세트를 선택하여, 지질 생산 조건 하에서 생육하였으며, 이때, 지질 프로필은 전술한 바와 같은 직접 에스테르 결합 전이 방법을 통하여 분석하였다. 선택된 클론들의 지질 프로필은 하기 표 37에 요약되어 있다.

표 37

올레아 유로파에아 스테아로일-ACP 불포화 효소 유전자 이식 프로토테카 모리포르미스 세포의 지질 프로필

균주	C14:0	C16:0	C18:0	C18:1	C18:2
야생형	0.88	22.82	3.78	64.43	6.54
pSZ1377 클론 A	0.94	18.60	1.50	69.45	7.67
pSZ1377 클론 B	0.93	18.98	1.35	69.12	7.67
pSZ1377 클론 C	0.93	19.01	2.31	68.56	7.43

[0658]

[0659] 상기 요약된 결과들은, 이종 불포화 효소(이 경우에는 올레아 유로파에아 유래 스테아로일-ACP 불포화 효소)를 도입하면 C18:1 지방산 수준은 증가할 수 있으며, 아울러 C18:0 및 C16:0 지방산 수준은 감소할 수 있음을 입증하는 것이다.

[0660] **실시예 7: 하이드록실화된 지방산을 생산하기 위한 프로토테카 조작**

[0661] 리시누스 커뮤니스 올리에이트 12-하이드록실라제(Rc12hydro)(GenBank 수탁 번호 제AAC49010.1호)를 프로토테카 모리포르미스 UTEX1435 유전적 배경에 도입하였다. 발현 구조물은 6S 게놈 영역에 대한 5'(서열 번호 82) 및 3'(서열 번호 84) 상동성 재조합 표적화 서열(구조물에 측접)을 함유하였는데, 이 재조합 표적화 서열은 씨.레인하르드티 β-튜불린 프로모터/5'UTR 및 클로렐라 불가리스 질산염 환원 효소 3'UTR의 제어 하에 있는 에스.세레비지아에 suc2 수크로스 인버타제 암호화 영역과 핵내 게놈으로의 통합에 사용되었다. 이와 같은 에스.세레비지아에 suc2 발현 카세트를 서열 번호 159에 나열하였으며, 이를 선택 마커로 사용하였다. 리시누스 커뮤니스 올리에이트 12-하이드록실라제 암호화 영역은 프로토테카 모리포르미스 Amt03 프로모터/5'UTR(서열 번호 89) 및 씨.불가리스 질산염 환원 효소 3'UTR의 제어 하에 두었다. 코돈 최적화 cDNA 서열 및 아미노산 서열을 각각 서열 목록 중 서열 번호 173 및 서열 번호 174에 나열하였다. 전체 Rc12hydro 발현 카세트는 pSZ1419로 명명하였으며, 이를 프로토테카 모리포르미스 유전적 배경에 도입하여 이 균주를 형질 전환하였다. 양성 클론을 유일한 탄소원으로서 수크로스를 함유하는 평판 상에서 스크리닝하였다. 이와 같은 양성 클론 하위 세트를 선택하고 나서, 지질 생산 조건 하에 생육하여, 알.커뮤니스 올리에이트 12-하이드록실라제 생성물, 즉 리시놀레산에 대해 스크리닝하였다(GC/MS법 사용).

[0662] 양성 Rc12hydro 유전자 이식체 내에서 생산되는 임의의 리시놀레산을 검출하는데 사용되는 GC/MS법은 다음과 같았다. 양성 클론 및 야생형 대조군 샘플을 건조한 다음, 이를 메탄올-톨루엔 중 4.5 H2SO4 2.2㎖(10:1(v/v)) 중에 현탁하였다. 이후, 이 혼합물을 70℃ 내지 75℃에서 3.5시간 동안 가열하였는데, 이때 중간 중간 초음파 처리 및 와류 처리를 해 주었다. 실온으로 냉각한 다음, 여기에 6% K2CO3(aq) 2㎖ 및 헵탄 2㎖를 첨가한 다음, 이 혼합물을 격렬하게 와류시킨 후, 층들을 원심 분리(900rpm, 2분)에 의해 급속하게 분리하였다. 상층을 분리하고 나서 농축하여 건조하였다(질소 기류 이용). 생성된 오일에 무수 피리딘 500㎕와 BSTFA/1% TMCS(서모 사이언티픽(Thermo Scientific)) 500㎕를 첨가하였다. 생성된 용액을 70℃ 내지 75℃에서 1.5시간 동안 가열한 후, 농축하여 건조하였다(질소 기류 이용). 그 다음, 잔류물을 헵탄 2㎖ 중에 재현탁하고 나서, 다시 농축하였다. 샘플을 취하여 헵탄 2㎖중에 재현탁한 다음, 서모 트레이스 GC 울트라/DSQII 시스템(Thermo Trace GC Ultra/DSQII system) 상에서 GC/MS에 의해 분석하였다(EI 모드, 리시놀레산 메틸의 TMS 베이스 피크(m/z 187)의 선택적 이온 모니터링법 사용). 지방산 메틸 에스테르를 레스텍 Rxi-5Sil MS 컬럼(Restek Rxi-5Sil MS column)(0.25㎜ ID, 길이 30m 및 필름 두께 0.5㎛) 상에서 분리하였다(캐리어 가스로서 헬륨 사용, 유속 = 1㎖/분). 컬럼의 초기 온도는 4분 동안 130℃로 유지시켰으며, 이후 램프를 사용하여 4℃/분의 속도로 가열하여 최종 온도를 240℃로 만

들었다. 샘플 중 리시놀레산의 존재 여부는, 전술한 바와 같이, 정체 시간과 전 스캔 질량 스펙트럼을, 처리된 리시놀레산의 진정한 샘플의 것과 비교함으로써 확인하였다. 도 2는 대표적인 양성 유전자 이식 클론의 GC 정체 시간을, 리시놀레산 표준과 야생형 대조군의 GC 정체 시간과 비교한 결과를 보여주는 것이다. 양성 유전자 이식 클론은 RT:33.22/33.23에서 유도성 피크(리시놀레산 표준에 있어서 유사한 피크와 상응함)를 나타냈는데, 이는 곧, 유전자 이식 클론과 양성 대조군 두가지에 유도성 리시놀레산이 존재한다는 것을 암시한다. 이 피크는 야생형 대조군 샘플에서는 전혀 나타나지 않았다.

[0663] **실시예 8: 대안적 선택가능한 마커를 사용한 프로토테카 조작**

[0664] **A. 프로토테카 모리포르미스 내 분비 가능한 α-갈락토시다제의 발현**

[0665] 프로토테카 종 내에서 이종 수크로스 인버타제 유전자를 발현하는 방법과 발현에 따른 효과에 대해서는 이미, 본원에 참조로 포함된 PCT 특허 출원 PCT/US2009/66142에 기술되어 있다. 본 실시예에서는 기타 다른 이종 다당류 분해 효소의 발현을 관찰하였다. MEL1 유전자(사카로마이세스 칼베르겐시스 유래; NCBI 수탁 번호 제P04824호의 서열에 상응하는 아미노산 서열), AglC 유전자(아스퍼질러스 나이저 유래; NCBI 수탁 번호 제Q9UUZ4호의 서열에 상응하는 아미노산 서열), 그리고 고등 식물인 시아몹시스 테트라고노로바(구아 콩) 유래 α-갈락토시다제(NCBI 수탁 번호 제P14749호의 서열에 상응하는 아미노산 서열)와 같은 α-갈락토시다제 암호화 외인성 유전자들 중 하나를 가지는 프로토테카 모리포르미스 UTEX 1435가 멜리비오스(α-D-gal-glu)를 바탕으로 생육할 수 있는 능력을 테스트하였다. 상기 수탁 번호들을 가지는 서열들과 이에 상응하는 아미노산 서열은 본원에 참조로 포함된 것이다. 모든 경우에 있어서, 프로토테카 모리포르미스 내 선호 코돈 사용 빈도에 따라서 유전자를 최적화하였다. 발현 카세트의 관련 부분들을 서열 목록 번호들과 함께 이하에 나열하였다. 모든 발현 카세트들은 안정적인 게놈 통합을 위한 5' 및 3'Clp 상동성 재조합 표적화 서열, 즉 클라미도모나스 레인하르드티 TUB2 프로모터/5'UTR 및 클로렐라 불가리스 질산염 환원 효소 3'UTR을 사용하였다.

[0666] 에스.칼베르겐시스 MEL1 아미노산 서열 서열 번호108

[0667] 에스.칼베르겐시스 MEL1 아미노산 서열 신호 펩티드 서열 번호 109

[0668] 에스.칼베르겐시스 MEL1 형질 전환 카세트 서열 번호 110

[0669] 에스.칼베르겐시스 MEL1 서열(최적화된 코돈) 서열 번호 111

[0670] 5' Clp 상동성 재조합 표적화 서열 서열 번호 112

[0671] 3' Clp 상동성 재조합 표적화 서열 서열 번호 113

[0672] 클라미도모나스 레인하르드티 TUB2 프로모터/5'UTR 서열 번호 114

[0673] 클로렐라 불가리스 질산염 환원 효소 3'UTR 서열 번호 115

[0674] 에이.나이저 AlgC 아미노산 서열 서열 번호 116

[0675] 에이.나이저 AlgC 아미노산 서열 신호 펩티드 서열 번호 117

[0676] 에이.나이저 AlgC 서열(최적화된 코돈) 서열 번호 118

[0677] 에이.나이저 AlgC 형질 전환 카세트 서열 번호 119

[0678] 씨.테트라고노볼라 α-갈락토시다제 아미노산 서열 서열 번호 120

[0679] 씨.테트라고노볼라 α-갈락토시다제 서열(최적화된 코돈)서열 번호 121

[0680] 씨.테트라고노볼라 α-갈락토시다제 형질 전환 카세트 서열 번호 122

[0681] 프로토테카 모리포르미스 세포를, 에스.칼베르겐시스 MEL1, 에이.나이저 AlgC 또는 씨.테트라고노볼라 α-갈락토시다제 유전자를 함유하는 발현 카세트 3개 각각으로 형질 전환하였다(상기 실시예 2에 기술된 바와 같은, 유전자 총 형질 전환법 이용). 유일한 탄소원으로서 2% 멜리비오스를 함유하는 평판을 사용하여 양성 클론을 스크리닝하였다. 씨.테트라고노볼라 발현 카세트 형질 전환체를 함유하는 콜로니들은 평판 상에 출현하지 않았다. 양성 클론을, 에스.칼베르겐시스 MEL1 형질 전환체와 에이.나이저 AlgC 형질 전환체를 함유하는 평판으로부터 골라냈다. 형질 전환 DNA의 통합 여부는 씨.불가리스 3'UTR 및 3'Clp 상동성 재조합 표적화 서열의 일부를 표적화하는 프라이머를 사용하는 PCR을 통해 확인하였다.

[0682] 5'프라이머 씨.불가리스 3'UTR:하류 Clp 서열(서열 번호 123)

[0683] ACTGCAATGCTGATGCACGGGA

[0684] 3'프라이머 씨.불가리스 3'UTR:하류 Clp 서열(서열 번호 124)

[0685] TCCAGGTCCTTTTCGCACT

[0686] 음성 대조군으로서, 형질 전환되지 않은 프로토테카 모리포르미스 세포 유래 게놈 DNA도 상기 프라이머 세트를 사용하여 증폭하였다. 야생형 세포로부터 유래하는 게놈 DNA로부터는 그 어떤 생성물도 증폭되지 않았다.

[0687] 에스.칼베르겐시스 MEL1 형질 전환체와 에이.나이저 AlgC 형질 전환체 각각으로부터 유래하는 몇 개의 양성 클론들(PCR로 확인)을 대상으로, 지질 배지 중 유일한 탄소원으로서 멜리비오스 상에서 생육할 수 있는 능력에 대해 테스트하였다. 이와 같이 선택된 클론들을 상기 실시예 1에 기술된 기본 배지(유일한 탄소원으로서 멜리비오스 포함)와 조건 하에서 3일 동안 생육하였다. α-갈락토시다제 암호화 유전자들 중 어느 하나를 함유하는 클론들은 모두 이 시기 동안 튼튼하게 생육되었던 반면에, 사카로마이세스 세레비지아에 SUC2 수크로스 인버타제를 발현하는 프로토테카 모리포르미스와 형질 전환되지 않은 야생형 균주는 둘 다 멜리비오스 배지 상에서는 잘 자라지 못했다. 이러한 결과들은, α-갈락토시다제 암호화 유전자가 형질 전환에 있어서 선택가능한 마커로서 사용될 수 있음을 암시하는 것이다. 또한, 이 데이터들은 에스.칼베르겐시스 MEL1 (서열 번호 109) 또는 에이.나이저 AlgC (서열 번호 117) 내에 존재하는 천연 신호 펩티드가 단백질을 프로토테카 모리포르미스 세포 내 주변 세포질에 표적화하는데 유용함을 암시하는 것이다.

[0688] **B. 프로토테카 내 THIC 유전자 상보체 티아민 영양 요구성**

[0689] 외인성 THIC 유전자의 발현을 이용하여 프로토테카 모리포르미스 세포 내 티아민 원 영양성을 관찰하였다. 식물과 조류 내 티아민 생합성은 통상적으로 색소체 내에서 수행되므로, 단백질 생산에 관여하는 대부분의 핵내 암호화 단백질은 색소체에 효율적으로 표적화되어야 할 것이다. 프로토테카 모리포르미스 세포의 DNA 서열 결정 및 전사체 서열 결정 결과, 티아민 생합성 효소를 암호화하는 유전자 모두(THIC 제외)는 게놈 내에 존재한다는 것이 규명되었다. 생화학적 수준으로 티아민 영양 요구성에 관여하는 병터를 나누기 위해서, (1) 2μM 티아민 하이드로클로라이드의 존재; (2) 티아민 부재; (3) 티아민은 존재하지 않되, 2μM 하이드록시에틸 티아졸(THZ)은 존재; (4) 티아민은 존재하지 않되, 2μM 2-메틸-4-아미노-5-(아미노메틸)피리미딘(PYR)은 존재; 그리고 (5) 티아민은 존재하지 않되, 2μM THZ 및 2μM PYR은 존재와 같은 5개의 상이한 조건 하에서의 프로토테카 모리포르미스 세포의 생장을 관찰하였다. 상기와 같은 5개의 상이한 조건 하에서 생육 실험을 수행한 결과들은, 프로토테카 모리포르미스 세포가 새로운 방식의 합성을 수행할 수 있지만, 만일 PYR 전구체가 제공되면 오로지 티아민 피로포스페이트(TPP)만을 생산할 수 있음을 암시한다. 이와 같은 결과는 프로토테카 모리포르미스의 티아민 영양 요구성은, (THIC 효소에 의해 촉매화되는 전환 과정인) 아미노이미다졸 리보뉴클레오티드로부터 하이드록시메틸피리미딘 포스페이트(HMP-P)로의 합성 과정을 진행시킬 수 없기 때문에 발생하는 것이라는 가설과 일치한다.

[0690] 상기 실시예 2에 기술된 유전자 총 형질 전환법을 통하여, 선택 마커로서 코코믹사 C-169 THIC(JGI 단백질 ID 30481에 상응하는 아미노산 서열로서 본원에는 참조로 포함됨) 및 에스.세레비지아에 SUC2 수크로스 인버타제를 발현하는, 프로토테카 모리포르미스 세포를 형질 전환하였다. 상기 발현 구조물은 코코믹사 C-169 THIC 유래 천연 운반 펩티드 서열, 게놈 DNA의 6S 영역에 상동성인 상류 및 하류 재조합 표적화 서열, 씨.레인하르드티 TUB2 프로모터/5'UTR 영역(서열 번호 104) 및 클로렐라 불가리스 질산염 환원 효소 3'UTR(서열 번호 115)을 함유하였다. 상기 에스.세레비지아에 SUC2 발현은 또한 씨.레인하르드티 TUB2 프로모터/5'UTR 영역(서열 번호 114)에 의해 작동되었으며, 클로렐라 불가리스 질산염 환원 효소 3'UTR(서열 번호 115)을 함유하였다. 유전자는 프로토테카 모리포르미스 내 선호 코돈 사용 빈도에 따라서 최적화하였다. 관련 발현 카세트 서열을 이하 서열 목록에 상세히 나열하였다:

[0691] 코코믹사 C-169 THIC 아미노산 서열 서열 번호 125

[0692] 코코믹사 C-169 THIC 아미노산 서열 천연 운반 펩티드 서열 번호 126

[0693] 코코믹사 C-169 THIC 형질 전환 카세트 서열 번호 127

[0694] 코코믹사 C-169 THIC 서열(최적화된 코돈) 서열 번호 128

[0695] 에스.세레비지아에 SUC2 서열(최적화된 코돈) 서열 번호 129

[0696] 5'6S 상동성 재조합 표적화 서열 서열 번호 82

[0697] 3'6S 상동성 재조합 표적화 서열 서열 번호 84

[0698] 양성 클론은, 티아민을 함유하지 않고 유일한 탄소원으로서 수크로스를 함유하는 평판 상에서 선택되었다. 양성 클론은, 코코믹사 C-169 THIC 유전자 내에서 결합하는 5'프라이머와, 6S 위치 내 형질 전환 DNA의 하류에 어닐링되는 3'프라이머를 사용하는 PCR로써 확인되었다. PCR로 확인된 양성 클론은 또한, 서던 블럿 검정법을 사용하여 확인되기도 하였다.

[0699] 야생형 프로토테카 모리포르미스 세포의 티아민 영양 요구성을 관찰하기 위하여는 우선, 내부 티아민 보존 세포를 고갈시켜야 했다. 티아민을 함유하지 않는 배지 중에서의 생장을 테스트하기 위해서 우선, 세포를 2μM 티아민을 함유하는 배지 중 고정 상에서 생육한 다음, 이 세포를 티아민 불포함 배지 중 750㎚에서의 광학 밀도(OD750)가 약 0.05가 되도록 희석하였다. 그 다음, 희석된 세포들을 티아민 불포함 배지 중 고정 상에서 다시 생육하였다(약 2일 내지 3일). 이와 같은 티아민 결핍 세포를, 티아민을 함유하지 않는 배지 중 생장 연구를 위해 배양액에 접종하였다. 야생형 세포를, 탄소원으로서 글루코스를 함유하는(또한 티아민을 함유하거나 함유하지 않는) 배지 중에서 생육한 다음, 천연 운반 펩티드 코코믹사 C-169 THIC 구조물을 포함하는 양성 클론을 유일한 탄소원으로서 수크로스를 함유하는 배지 중에서 생육하였다. 생장은 750㎚에서의 흡광도를 모니터링함으로써 측정하였다. 생장 실험의 결과를 통하여, 티아민 불포함 배지 중 야생형 세포의 생장과 비교하였을 때 티아민 불포함 배지 중 이식 유전자를 발현하는 균주의 생장이 훨씬 컸음을 알 수 있었다. 그러나, 형질 전환체의 세포 생장 속도와 세포 밀도는 티아민 함유 배지 중 야생형 세포의 생장 속도와 세포 밀도에 못미쳤다. 또한, 티아민 불포함 배지 중 형질 전환체 클론의 생장량과, 통합된 코코믹사 효소의 복사체 수 간에는 강한 상관 관계가 있었다(즉, 티아민 불포함 배지 중 이식 유전자의 복사체 수는 더욱 많았으며, 이 배지 중에서 세포는 더욱 잘 생장하였다).

[0700] 코코믹사 THIC, 아라비돕시스 탈리아나 THIC 유전자, 그리고 시네코시스티스 종 PCC 6803 thiC 유전자를 함유하는 발현 구조물을 사용하여 부가의 형질 전환체를 제조하였다. 코코믹사 및 에이.탈리아나 THIC 유전자의 경우, 천연 운반 펩티드 서열은 클로렐라 프로토테코이데스 스테아로일-ACP 불포화 효소(SAD) 유전자 유래 운반 펩티드 서열과 치환되었다. 시네코시스티스 종은 시아노박테리아이며, thiC 단백질은 천연 운반 펩티드 서열을 함유하지 않는다. 시네코시스티스 종 thiC 구조물에 있어서, 클로렐라 프로토테코이데스 SAD 유전자 유래 운반 펩티드 서열을 시네코시스티스 종 thiC의 N-말단에 융합하였다. 모든 경우에 있어서, 서열들은 프로토테카 모리포르미스 내 발현에 대해 최적화된 코돈이었다. 상기 구조물들 3개는 모두 게놈의 6S 영역에 상동성인 상류 및 하류 재조합 표적화 서열(서열 번호 82 및 84), 클로렐라 프로테코이데스 액틴 프로모터/5'UTR, 그리고 클로렐라 프로토테코이데스 EF1A 유전자 3'UTR을 함유하였다. 3개의 구조물은 모두 씨.레인하르드티 TUB2 프로모터/5'UTR(서열 번호 114)에 의해 작동되는 neoR 유전자를 함유하였으며, 또한 G418에 의해 선택되도록 하는 씨.불가리스 3'UTR(서열 번호 115)을 함유하였다. 에이.탈리아나 THIC의 아미노산 서열은 NCBI 수탁 번호 제NP_180524호에 해당하고, 시네코시스티스 종 thiC의 아미노산 서열은 NCBI 수탁 번호 제NP_442586호에 해당하며, 이들 서열은 둘 다 본원에 참조로 포함된 것이다. 관련된 발현 카세트 서열은 이하 서열 목록에 상세히 나열하였다:

[0701] 씨.프로토테코이데스 운반 펩티드 함유 코코믹사 THIC 발현 구조물

[0702] 서열 번호 130

[0703] 씨.프로토테코이데스 운반 펩티드 함유 코코믹사 THIC 서열 번호 131

[0704] 씨.프로토테코이데스 액틴 프로모터 /5'UTR 서열 번호 132

[0705] 씨.프로토테코이데스 EF1A 3'UTR 서열 번호 133

[0706] 에이.탈리아나 THIC 발현 구조물 서열 번호 134

[0707] 씨.프로토테코이데스 운반 펩티드 함유 에이.탈리아나 THIC 서열 번호 135

[0708] 천연 운반 펩티드 함유 에이.탈리아나 THIC 아미노산 서열 서열 번호136

[0709] 시네코시스티스 종 thiC 발현 구조물 서열 번호 137

[0710] 씨.프로토테코이데스 운반 펩티드 함유 시네코시스티스 종 thiC

[0711] 서열 번호 138

[0712] 시네코시스티스 종 thiC 아미노산 서열 서열 번호 139

[0713] neoR 유전자 서열 번호 140

[0714] 양성 클론을, G418을 함유하는 평판 상에서 스크리닝하였으며, 각각의 형질 전환으로부터 유래하는 몇몇 클론들을 골라내어 PCR로 확인하였다. 코코믹사 C-169(씨.프로토테코이데스 운반 펩티드 함유), 에이.탈리아나 및 시네코시스티스 종 PCC 6803 THIC 유전자를 함유하는 형질 전환 DNA 구조물을 각각 게놈의 6S 위치에 통합한 결과는, 다음과 같은 프라이머들을 사용하는 PCR 분석법을 통하여 확인되었다.

[0715] 5'THIC 코코믹사 확인 프라이머 서열(서열 번호 141)

[0716] ACGTCGCGACCCATGCTTCC

[0717] 3'THIC 확인 프라이머 서열(서열 번호 142)

[0718] GGGTGATCGCCTACAAGA

[0719] 5'THIC 에이.탈리아나 확인 프라이머 서열(서열 번호 143)

[0720] GCGTCATCGCCTACAAGA

[0721] 5'thiC 시네코시스티스 종 확인 프라이머 서열(서열 번호 144)

[0722] CGATGCTGTGCTACGTGA

[0723] G418 함유 배지 중 상이한 구조물 각각의 형질 전환체로부터 선택 확인된 양성 클론을 사용하여, (전술한 바와 같이) 티아민 결핍 세포를 대상으로 하는 생장 실험을 수행하였다. 모든 형질 전환체는 티아민 불포함 배지 중에서 (다양한 견고성 정도로) 생육할 수 있었다. 티아민 불포함 배지 중 형질 전환체의 생장과 티아민 함유 배지 중 야생형 세포의 생장을 비교한 결과, 티아민 불포함 배지 중 생장을 지속할 수 있는 능력에 있어서, (1) 에이.탈리아나 형질 전환체; (2) 코코믹사 C-169(씨.프로토테코이데스 운반 펩티드 함유) 형질 전환체; 및 (3) 시네코시스티스 종 형질 전환체와 같은 순위가 매겨졌다. 이와 같은 결과들은, 에이.탈리아나 THIC 복사체 하나는 프로토테카 모리포르미스 세포 내 티아민 영양 요구성을 보완할 수 있었으며, 코코믹사 C-169(천연 운반 펩티드 서열 또는 씨.프로토테코이데스 유래 운반 펩티드 서열 중 어느 하나를 포함) 복사체 다수 개와 시네코시스티스 종 THIC 는 티아민 부재 하에 신속하게 생장할 수 있어야 함을 암시하는 것이다. 상이한 공급원으로부터 유래하는 상이한 THIC로 인한 결과가 다양하였으므로, 임의의 특정 THIC 유전자가 프로토테카 종 내에 존재하는 병터를 완전히 보완할 수 있는 능력은 예측 불가능하였다.

[0724] 3개의 THIC 아미노산 서열의 정렬을 수행하였다. 시네코시스티스 종 유래 thiC와, 코코믹사 및 에이.탈리아나 유래 THIC 간 서열 보존 수준은 유의적이었던 반면에(아미노산 수준에서 41% 동일), 시아노박테리아 단백질에는 조류 및 식물 단백질에 잘 보존되어 있는, N-말단 존재 도메인이 존재하지 않았다. 도메인이 존재하지 않음에도 불구하고(그리고 이로 인하여 구조상에도 차이가 있을 수 있음에도 불구하고), 시네코시스티스 종 thiC 를 발현하는 구조물은 프로토테카 모리포르미스 세포 내 티아민 원 영양성을 최소한 부분적으로나마 복구할 수 있었다.

[0725] **실시예 9: 연료 생산**

[0726] **A. 익스펠러 프레스 및 프레스 보조 장치를 사용한 미세 조류로부터 오일의 추출**

[0727] DCW를 기준으로 38%의 오일을 함유하는 미세 조류 바이오매스는 드럼 건조기를 사용하여 건조시켜 수분 함량이 5% 내지 5.5%가 되게 하였다. 상기 바이오매스는 프렌치 L250 프레스에 공급하였다. 30.4kg(67lbs.)의 바이오매스를 프레스를 통해 공급하고 오일은 회수하지 않았다. 프레스 보조기로서 다양한 백분율의 스위치그래스와 조합된 동일한 건조된 미세 조류 바이오매스를 프레스를 통해 공급하였다. 건조된 미생물 바이오매스와 20% w/w의 스위치그래스의 조합으로 최상의 총 백분율의 오일을 회수하였다. 이어서 프레스된 케이크는 헥산 추출을 거치고 20% 스위치그래스 조건에 대한 최종 수율은 총 이용가능한 오일(중량으로 계산함)의 61.6%이었다. 상기 50% 이상의 오일 건조 세포 중량을 갖는 바이오매스는 오일을 유리시키기 위해 스위치그래스와 같은 프레싱 보조기의 사용을 필요로 하지 않는다. 익스펠러 프레스를 사용하여 미세 조류로부터 오일을 추출하는 기타 다른 방법은 본원에 참조로 포함되어 있는 PCT 특허 출원 PCT/US2010/31108호에 기술되어 있다.

[0728] **B. 프로토테카 오일로부터 바이오디젤의 생성**

[0729] 전술된 방법에 따라 생성된 프로토테카 모리포르미스 UTEX 1435 유래의 고무 성질이 제거된 오일을 에스테르 결

합 전이를 거쳐서 지방산 메틸 에스테르를 생성하였다. 결과는 하기 표 38에 나타내어져 있다.

[0730] 오일의 지질 프로필은 다음과 같다:

[0731]
C10:0 0.02
C12:0 0.06
C14:0 1.81
C14.1 0.07
C16:0 24.53
C16:1 1.22
C18:0 2.34
C18:1 59.21
C18:2 8.91
C18:3 0.28
C20:0 0.23
C20:1 0.10
C20:1 0.08
C21:0 0.02
C22:0 0.06
C24:0 0.10

표 38

프로토테카 모리포르미스 트리글리세라이드 오일 유래의 바이오디젤 프로필

방법	시험		결과	단위
ASTM D6751 A1	바이오디젤 배합 연료의 콜드 소킹 여과능	여과 시간	120	초
		여과된 용적	300	ml
ASTM D93	펜스키-마르텐스 폐쇄 컵 인화점(Pensky-Martens Closed Cup Flash Point)	사용된 과정	A	
		보정된 인화점	165.0	°C
ASTM D2709	중간 증류 연료중 물 및 침강물 (원심분리 방법)	침강 및 물	0.000	용적%
EN 14538	ICP OES에 의한 Ca 및 Mg 함량의 측정	(Ca 및 Mg)의 합	<1	mg/kg
EN 14538	ICP OES에 의한 Ca 및 Mg 함량의 측정	(Na 및 K)의 합	<1	mg/kg
ASTM D445	역학적 /동력학적 점도	140°F/40℃에서의 역학적 점도	4.873	mm^2/s
ASTM D874	윤활 오일 첨가제 유래의 황화된 재	황화된 재	< 0.005	중량%
ASTM D5453	자외선 형광에 의한 가벼운 탄화수소, 스파크 점화 엔진 연료, 디젤 엔진 연료 및 엔진오일 중 총 황의 측정	황, mg /kg	1.7	mg/kg
ASTM D130	부식 – 구리 스트립	바이오디젤 -Cu 부식 50℃ (122°F)/3 시간	1a	
ASTM D2500	혼탁점	혼탁점	6	°C
ASTM D4530	잔류 마이크로 탄소	평균 마이크로 방법 잔류 탄소	< 0.10	중량%
ASTM D664	전위차 적정에 의한 석유 생성물의 산가	사용된 과정	A	
		산가	0.20	mg KOH/g
ASTM	가스 크로마토그래피에 의한	유리 글리세린	< 0.005	중량%

[0732]

방법	시험		결과	단위
D6584	B-100 바이오디젤 메틸 에스테르 중 유리된 것 및 총 글리세린의 측정	총 글리세린	0.123	중량%
ASTM D4951	ICP-AES에 의한 윤활 오일 중 첨가 요소	인	0.000200	중량%
ASTM D1160	감압하에 석유 생성물의 증류	IBP	248	°C
		5%회수시의 AET	336	°C
		10%회수시의 AET	338	°C
		20%회수시의 AET	339	°C
		30%회수시의 AET	340	°C
		40%회수시의 AET	342	°C
		50%회수시의 AET	344	°C
		60%회수시의 AET	345	°C
		70%회수시의 AET	347	°C
		80%회수시의 AET	349	°C
		90%회수시의 AET	351	°C
		95%회수시의 AET	353	°C
		FBP	362	°C
		% 회수됨	98.5	%
		% 손실	1.5	%
		% 잔사	0.0	%
		냉각 트랩 용적	0.0	ml
		IBP	248	°C
EN 14112	산화 안정성 측정 (가속화된 산화 시험)	산화 안정성	> 12	시간
		작업 온도 (통상적으로 110℃)	110	°C
ASTM D4052	디지털 밀도 측정기에 의한 액체의 밀도	60°F에서의 API 중량	29.5	°API
ASTM D 6890	점화 지연(ID) 및 유래된 세탄가(DCN)의 측정	유래된 세탄가 (DCN)	> 61.0	

[0733]

[0734] 바이오디젤의 지질 프로필은 원료공급 오일의 지질 프로필과 매우 유사하였다. 본 발명의 방법 및 조성물에 의해 제공된 기타 다른 오일은 에스테르 결합 전이를 거쳐서 (a) C8 내지 C14 4% 이상; (b) C8 0.3% 이상; (c) C10 2% 이상; (d) C12 2% 이상; 및 (e) C8 내지 C14 30% 이상을 포함하는 지질 프로필을 갖는 바이오디젤을 생성할 수 있다.

[0735] 생성된 바이오디젤의 ASTM D6751 A1 방법에 의한 콜드 소킹 여과능은 300ml의 용적에 대해 120초였다. 이러한 시험은 16시간 동안 40°F로 냉각되고 실온으로 가온되도록 하며 스테인리스강 지지체와 함께 0.7마이크론 유리 섬유 필터를 사용하여 진공 여과하는 300ml의 B100의 여과를 포함한다. 본 발명의 오일은 에스테르 결합 전이하여 바이오디젤을 생성시킬 수 있으며, 이 때 콜드 소킹 시간은 120초 미만, 100초 미만 및 90초 미만이다.

[0736] **C. 재생 디젤의 생산**

[0737] 전술된 방법에 따라 제조되고 본 실시예에서 바이오디젤을 제조하는데 사용되는 오일과 동일한 지질 프로필을 갖는, 프로토테카 모리포르미스 UTEX 1435 유래의 고무 성질이 제거된 오일은 에스테르 결합 전이를 거쳐서 재생 디젤을 생성하였다.

[0738] 상기 오일을 먼저 수소화처리하여 산소 및 글리세롤 골격을 제거하여 n-파라핀을 생성하였다. 상기 n-파라핀은 이어서 크래킹 및 이성체화를 거쳤다. 상기 물질의 크로마토그램은 도 1에 나타내어져 있다. 상기 물질은 이어서 냉 여과를 거쳐서 C18 물질의 약 5%를 제거하였다. 냉 여과 후, 총 용적의 물질을 인화점으로 컷팅하고, 인

화점, ASTM D-86 증기 분포, 혼탁점 및 점도에 대해 평가하였다. 인화점은 63℃이었고, 점도는 2.86cSt(센티스토크)이었으며; 혼탁점은 4℃이었다. ASTM D86 증류가는 표 39에 나타내어져 있다.

표 39

ASTM D86 증류가

판독 결과(°C) :

부피	온도
IBP	173
5	217.4
10	242.1
15	255.8
20	265.6
30	277.3
40	283.5
50	286.6
60	289.4
70	290.9
80	294.3
90	300
95	307.7
FBP	331.5

[0739]

[0740] 생성된 물질의 T10-T90은 57.9℃이었다. 본원에 개시된 오일의 수소화처리, 이성체화 및 기타 다른 공유적 변형 방법, 및 본원에 개시된 증류 및 분별(예를 들어, 냉여과) 방법을 사용하여, 본원에 개시된 방법에 따라 생성된 트리글리세라이드 오일을 사용하는 20℃, 25℃, 30℃, 35℃, 40℃, 45℃, 50℃, 60℃ 및 65℃와 같은 기타 다른 T10-T90 범위를 갖는 재생 디젤 조성물 생성할 수 있다.

[0741] 생성된 물질의 T10은 242.1℃이었다. 본원에 개시된 오일의 수소화처리, 이성체화 및 기타 다른 공유적 변형 방법, 및 본원에 개시된 증류 및 분별(예를 들어, 냉여과) 방법을 사용하여, 180 내지 295, 190 내지 270, 210 내지 250, 225 내지 245, 및 290 이상의 T10과 같은 기타 다른 T10 값을 갖는 재생 디젤 조성물을 생성할 수 있다.

[0742] 생성된 물질의 T90은 300℃이었다. 본원에 개시된 오일의 수소화처리, 이성체화 및 기타 다른 공유적 변형 방법, 및 본원에 개시된 증류 및 분별(예를 들어, 냉여과) 방법을 사용하여, 280 내지 380, 290 내지 360, 300 내지 350, 310 내지 340, 및 290 이상의 T90과 같은 기타 다른 T90 값을 갖는 재생 디젤 조성물을 생성할 수 있다.

[0743] 생성된 물질의 FBP는 300℃이었다. 본원에 기술된 오일의 수소화처리, 이성체화 및 기타 다른 공유적 변형 방법, 및 본원에 기술된 증류 및 분별(예를 들어, 냉여과) 방법을 사용하여, 290 내지 400, 300 내지 385, 310 내지 370, 315 내지 360, 및 300 이상의 FBP와 같은 기타 다른 FBP 값을 갖는 재생 디젤 조성물을 생성할 수 있다.

[0744] (a) C8 내지 C14 4% 이상; (b) C8 0.3% 이상; (c) C10 2% 이상; (d) C12 2% 이상; 및 (e) C8 내지 C14 30% 이상을 포함하는 지질 프로필을 갖는 오일을 포함하는, 본 발명의 방법 및 조성물에 의해 제공된 기타 다른 오일은 조합된 수소화처리, 이성체화 및 기타 다른 공유적 변형을 거칠 수 있다.

[0745] 본 발명은 이의 특정 구체예와 관련하여 기술되었지만, 추가의 변형이 가능함이 이해될 것이다. 본원은 일반적으로 본 발명의 원리에 따르고 본 발명이 속하는 기술 분야에서 공지되거나 통상적인 실시 범위 내에 있고 이전에 제시된 필수적 특징에 적용될 수 있는 바와 같이, 본 개시 내용을 벗어나는 것을 포함하는 본 발명의 임의의 변형, 용도 또는 적용을 포함하는 것으로 의도된다.

[0746] **실시예 10: C18:2 및 C18:3 글리세로지질을 생산하기 위한 미생물 조작**

[0747] 조류 및 식물 내에서의 지질 합성은, 색소체 피루베이트 탈수소화 효소 복합체를 통해 글루코스 또는 기타 다른

탄소원을 아세틸 CoA로 전환하는 것으로 시작되었다. 그 다음, 중탄산염을 기질로서 사용하는 아세틸 CoA 카복실라제(ACCase)는 3-C 화합물, 말로닐 CoA를 생성하였다. 그 다음, b-케토아실-ACP(아실기 운반 단백질) 합성효소 III(KAS III)은, 말로닐 CoA와 아세틸 CoA 간 제1 축합 반응을 촉진하여 4-C 화합물을 생성하였다. 연속적 2-C의 부가(C16:0를 이용함)는 KAS I에 의해 촉진되었다. 마지막으로, C18:0으로의 2-C 연장 과정은 KAS II에 의해 촉진되었다. 티오에스테라제(TE)는 아실-ACP의 연장을 종결시켰다. 가용성 효소, 스테아로일 ACP 불포화효소(SAD)는 C18:0-ACP 기질에 대하여 활성을 가졌으며, 또한 D9 위치에 이중 결합을 형성하여 올리에이트-ACP를 생성하였다. 생성된 C18:1 지방산은 올리에이트 또는 광범위한 특이성을 가지는 TE 중 어느 하나의 작용으로 ACP로부터 유리되었다.

[0748] 색소체 내 ACP로부터 일단 유리된 모든 지방산은 ER, 즉 지질(TAG) 생합성이 일어나는 곳으로 운반되었다. 포괄적으로 말하면, 고등 식물 내 ER에서 일어나는 지질 생합성 경로는 2개가 존재하는데, 이들 2개의 경로는 어느 정도 기질을 공유하는 것이 분명하다. 지방 아실 CoA 독립 경로는, 매우 선택적인 기질 특이성을 나타낼 수 있는 아실리소포스파티딜콜린 아실 트랜스퍼라제를 사용하는 포스파티딜 콜린(P-콜린) 부들 사이에 지방 아실 기들을 운반하며, 궁극적으로는 상기 지방 아실기들을 디아실글리세롤(DAG)로 운반하였다. 효소 디아실글리세롤 아실트랜스퍼라제(DGAT)는 최종 운반 과정, 즉 아실 CoA 기질 유래 지방 아실 기를, DAG를 거쳐 마지막으로 트리아실 글리세롤로 운반하는 과정을 수행하였다. 다른 한편, 지방 아실 CoA 종속 경로는, 기질로서 지방 아실 CoA를 사용하여 지방 아실기를, 각각 글리세롤 포스페이트 아실트랜스퍼라제(GPAT), 리소포스파티딘산 아실트랜스퍼라제(LPAAT) 및 DGAT의 작용을 통하여, 글리세롤-3-포스페이트, 리소포스파티딘산(LPA) 및 DAG로 운반하였다.

[0749] C18:2 및 C18:3을 포함하는 TAG를 합성하도록 미생물을 조작하는데 유용한 효소로서는 b-케토아실-ACP 합성 효소 II(KAS II), 스테아로일 ACP 불포화 효소(SAD), 티오에스테라제, 예를 들어 올리에이트 특이적 티오에스테라제, 지방산 불포화 효소(FAD), 그리고 글리세로지질 불포화 효소, 예를 들어 ω-6 지방산 불포화 효소, ω-3 지방산 불포화 효소, 또는 ω-6 올리에이트 불포화 효소를 포함한다. 상이한 효소들은 미생물 내에서 단독으로, 또는 C18:2 또는 C18:3 지방산을 증가시키거나 TAG를 생산하면서 과발현될 수 있었다. KAS II 효소 활성의 발현이 증가함에 따라서, 팔미테이트(C16:0)로부터 스테아레이트(C18:0) 쪽으로, 그리고 그 이상의 프로필을 가지는 지방산으로 탄소가 축적되었다. 몇몇 후보 KAS II 효소의 아미노산 서열을 이하 표 40에 제시하였다. 개시된 KAS II 서열은 올레 지방산, 리놀레 지방산 또는 리놀렌 지방산의 수준을 특이적으로 상승시키는 고등 식물 종으로부터 유래하였다. 당업자는 KAS II에 대한 기타 다른 유전자, 예를 들어 자트로파 컬카스 유전자(GenBank 수탁 번호 제ABJ90469.2호), 글리신 맥스 유전자(GenBank 수탁 번호 제AAW88762.1호), 엘라에이스 올레이페라 유전자(GenBank 수탁 번호 제ACQ41833.1호), 아라비돕시스 탈리아나 유전자(GenBank 수탁 번호 제AAL91174호), 비티스 비니페라 유전자(GenBank 수탁 번호 제CBI27767호) 및 고시피움 히르서텀 유전자(GenBank 수탁 번호 제ADK23940.1호)(이에 제한되는 것은 아님)를 동정할 수 있을 것이다.

표 40

[0750]

예시적 KAS II 효소	
KAS II 효소	서열 번호
리시누스 커뮤니스	서열 번호 175
헬리안투스 아너스	서열 번호 176
브라시카 나푸스	서열 번호 177
글리신 맥스	서열 번호 178
피.모리포르미스	서열 번호 179

[0751] 리놀레 지방산과 리놀렌 지방산의 수준을 높이기 위해, 증가된 수준의 스테아레이트를 올레산(C18:1)으로 전환시키는 것은, 하나 이상의 지질 경로 효소를 미생물 내에서 과발현시킴으로써 이루어질 수 있었다. 불포화물의 수준을 상승시킬 수 있는 활성을 가지는 부가 효소 2가지로서는 스테아로일 ACP 불포화 효소(SAD) 및 올리에이트 특이적 티오에스테라제가 있다. 이와 같은 효소들 중 하나 또는 둘 다의 작용을 통하여, 증가된 수준의 스테아레이트(C18:0)를 올레산으로 전환시키는 것은, 우선 리놀레 지방산과 리놀렌 지방산을 형성시킴으로써 이루어질 수 있었다.

[0752] 올레산 수준을 상승시키기 위해 과발현하는데 유용한 예시적 SAD 효소의 아미노산 서열을 표 41에 제시하였다. 뿐만 아니라, 피.모리포르미스로부터 유래하는 내인성 SAD(서열 번호 180)는 또한 C18:1 수준을 증가시키는데

효율적이었다. 본원에 개시된, 올레 지방산, 리놀레 지방산 또는 리놀렌 지방산의 수준을 특이적으로 상승시키는 SAD 서열은, 고등 식물 종으로부터 유래하였다. 당업자는 SAD에 대한 기타 다른 유전자들을 동정할 수 있을 것이다.

표 41

[0753] 예시적 SAD 효소

SAD 효소	서열 번호	GenBank ID 번호
리시누스 커뮤니스	서열 번호 196	ACG59946.1
헬리안투스 아너스	서열 번호 197	AAB65145.1
브라시카 준세아	서열 번호 198	AAD40245.1
글리신 맥스	서열 번호 199	ACJ39209.1
올레아 유로파에아	서열 번호 200	AAB67840.1
베르니시아 포르디	서열 번호 201	ADC32803.1

[0754] SAD 효소는 C18:0-ACP 기질에 대해 활성을 가졌으며, D9 위치에서 탄소-탄소 이중 결합을 형성하여, 올리에이트-ACP를 생성하였다. 생성된 C18:1 지방산은, 올리에이트 또는 광범위한 특이성을 가지는 TE 중 어느 하나의 작용을 통하여 ACP로부터 유리되었다. 본 발명의 발명자들은, 본원의 실시예에서 올레아 유로파에아 스테아로일-ACP 불포화 효소(수탁 번호 제AAB67840.1호; 서열 번호 172) 또는 카르타머스 팅크토리우스 ACP 티오에스테라제(수탁 번호 제AAA33019.1호; 서열 번호 195)를 과발현시켰을 경우, C18:1 지방산의 축적량이 증가되었다는 것을 보여주었다. 올레 지방산을 증가시키는데 유용한, 예시적 올리에이트 티오에스테라제의 아미노산 서열을 표 42에 제시하였다. 당업자는 올리에이트 티오에스테라제에 대한 다른 유전자를 동정할 수 있을 것이다.

표 42

[0755] 올레 지방산 생산을 증가시키기 위한 예시적 티오에스테라제

티오에스테라제 효소	서열 번호	GenBank ID 번호
헬리안투스 아너스	서열 번호 202	AAL79361.1
브라시카 라파	서열 번호 203	AAC49002.1
자트로파 컬카스	서열 번호 204	ABX82799.3
제아 메이스	서열 번호 205	ACG40089.1
제아 메이스	서열 번호 206	ACG42559.1

[0756] 지방산 불포화 효소는 미생물 내에 리놀레 지방산과 리놀렌 지방산의 축적을 증가시키는데에 유용한 부가 효소이다. 특히, 2개의 효소 즉, FAD 2 및 FAD 3의 활성은 미생물 내 리놀레 지방산과 리놀렌 지방산의 축적을 증가시켰다. 예시적 FAD 2 및 FAD 3 효소의 아미노산 서열을 이하 표 43에 제시하였다.

표 43

[0757] 예시적 FAD2 및 FAD3 효소

FAD 2 및 FAD 3 효소	서열 번호
리누스 유시타티시멈 12 불포화 효소	서열 번호 181
리누스 유시타티시멈 15 불포화 효소	서열 번호 182
리누스 유시타티시멈 15 불포화 효소	서열 번호 183
칼타머스 팅크토리우스 12 불포화 효소	서열 번호 184
헬리안투스 아너스 12 불포화 효소	서열 번호 185

[0758] 표 43에 나열된 효소들 이외에, 예시적 Δ12 FAD 효소의 아미노산 서열을 표 44에 나열하였다. 미생물 내 과발현에 적당한 기타 다른 Δ12 FAD 효소는 표 45에 나열하였다. 예시적 Δ15 FAD 및 불포화 지방산과 TAG의 수준을 증가시키는데 유용한 기타 다른 효소를 표 46에 나열하였다. 부가의 글리세로지질 불포화 효소는 표 47에 나열하였다.

표 44

[0759] 리놀레 지방산 생산을 증가시키기 위한 예시적 Δ12 FAD 효소

Δ12 FAD 효소	서열 번호	GenBank ID 번호
칼타머스 팅크토리우스	서열 번호 207	ADM48790.1
고시피움 히르서텀	서열 번호 208	CAA71199.1
글리신 맥스	서열 번호 209	BAD89862.1
제아 메이스	서열 번호 210	ABF50053.1
프로토테카 모리포르미스 대립유전자 1	서열 번호 211	
프로토테카 모리포르미스 대립유전자 2	서열 번호 212	

표 45

[0760] 리놀레산 또는 리놀렌산을 증가시키기 위한, 미생물 내 과발현에 적당한 부가 Δ12 FAD 효소

Δ12 FAD 효소	GenBank 수탁 번호
베르노니아 갈라멘시스	AAF04094.1
베르노니아 갈라멘시스	AAF04093.1
라이티아 팅크토리아	ADK47520.1
올레아 유로파에아	AAW63041.1
베르니시아 포르디	AAN87573
아라비돕시스 탈리아나	AAA32782.1
카멜리나 사티바	ADU18247.1
카멜리나 사티바	ADU18248.1
카멜리나 사티바	ADU18249.1
칼타머스 팅크토리우스	ADK94440.1
글리신 맥스	BAD89862.1
글리신 맥스	DQ532371.1
고시피움 히르서텀	AAL37484.1
리넘 유시타티시멈	ACF49507.1
리넘 유시타티시멈	ACF49508.1
오에노테라 비에티스	ACB47482
사카로마이세스 세레비지아에	NP_011460.1
제아 메이스	ACG37433.1
브라시카 라파	CAD30827.1

표 46

[0761] 리놀렌 지방산 생산을 증가시키기 위한 예시적 Δ15 FAD 효소

Δ15 FAD 효소	서열 번호	GenBank ID 번호
브라시카 라파	서열 번호 213	AAA32994.1
카멜리나 사티바	서열 번호 214	
카멜리나 사티바	서열 번호 215	
글리신 맥스	서열 번호 216	ACF19424.1
베르니시아 포르디	서열 번호 217	AAF12821.1
리시누스 커뮤니스	서열 번호 218	EEF36775.1
리넘 유시타티시멈	서열 번호 219	ADV92272.1
프로토테카 모리포르미스 대립유전자 1	서열 번호 220	
프로토테카 모리포르미스 대립유전자 2	서열 번호 221	

표 47

[0762] 불포화 TAG의 수준 또는 리놀렌산을 증가시키기 위한 미생물 내 과발현에 적당한 글리세로지질 불포화 효소

글리세로지질 불포화 효소	GenBank 수탁 번호
글리신 맥스	ACD69577.1
글리신 맥스 컬티바 볼라니아	ACS15381.1
글리신 소자	P48621.1
아라비돕시스 탈리아나	NP_180559.1
리넘 그랜디플로럼	BAG70949.1
제아 메이스	BAA22440.1
올레아 유로파에아	ABG88130.2
자트로파 컬카스	ABX82798.1
베르니시아 포르디	CAB45155.1
베르니시아 포르디	AAD13527.1

[0763] 본원에 개시된 아미노산 서열들은 본원에 개시된 방법을 사용하여 미생물 내에서 발현되었다. 암호화 서열은 미생물 내 발현에 최적화될 수 있었다. 예를 들어 피.모리포르미스 내 발현에 있어서 본원의 표 2에 개시된 선호 코돈 사용 빈도가 적용되었다.

[0764] **실시예 11: 리놀렌 불포화 지방산 및 글리세로지질의 생산을 증가시키기 위한 미생물 조작**

[0765] 실시예 10에 기술된 바와 같이, Δ15 불포화 효소는 C18:2(리놀레) 지방산 또는 지방 아실 분자들의 15번 위치에 이중 결합을 형성하여, C18:3(리놀렌) 지방산 또는 지방 아실 분자가 생산되는 것을 촉진한다. 리놀렌 불포화 지방산이 풍부한 오일을 생산하는 임의의 고등 식물 종, 예를 들어 브라시카 나푸스(Bn), 카멜리나 사티바(Cs), 그리고 리넘 유시타티시멈은, 미생물 내에서 발현되어 지방산 프로필에 영향을 미칠 수 있는 Δ15 불포화 효소를 암호화하는 유전자 원이다. 본 실시예는, 형질 전환된 미생물의 지방산 프로필에 있어서 리놀렌산이 증가되도록 미생물을 조작하기 위해 Δ15 불포화 효소를 암호화하는 폴리뉴클레오티드를 사용하는 것에 관하여 기술하고 있다.

[0766] 프로토테카 모리포르미스 UTEX 1435, 균주 A의, (고급 오일을 생산하도록) 전통적으로 돌연 변이된 파생 균주를 실시예 2에 상세히 기술된 유전자 총 형질 전환법에 따라서 표 49에 나열된 플라스미드 구조물 각각으로 개별 형질 전환하였다. 각각의 구조물은 6S 게놈 영역에 대한 5'(서열 번호 82) 및 3'(서열 번호 84) 상동성 재조합 표적화 서열(구조물에 측접함)을 함유하였으며, 여기서, 상기 재조합 표적화 서열은, 씨.레인하르드티 β-튜불린 프로모터/5'UTR 및 클로렐라 불가리스 질산염 환원 효소 3'UTR의 제어 하에 있는 에스.세레비지아에 suc2 수크로스 인버타제 암호화 영역과 핵내 게놈으로의 통합에 사용되었다. 이와 같은 에스.세레비지아에 suc2 발현 카세트는 서열 번호 159로서 나열되었으며, 수크로스 인버타제 유전자는 선택 마커로서 사용되었다. 단백질 암호화 영역은 모두 표 2에 따라서 프로토테카 모리포르미스 UTEX 1435 핵내 유전자에 고유한 코돈 편향 사용을 반영하도록 최적화된 코돈이었다. 브라시카 나푸스(Bn FAD3, GenBank 수탁 번호 제AAA32994호), 카멜리나 사티바 FAD-7 및 리넘 유시타티시멈(Lu FAD3A 및 Lu FAD3B, GenBank 수탁 번호 제ABA02172호 및 제ABA02173호) 유래 불포화 효소 유전자의 암호화 영역은 각각 프로토테카 모리포르미스 Amt03 프로모터/5'UTR(서열 번호 89) 및 씨.불가리스 질산염 환원 효소 3'UTR의 제어 하에 두었다. 플래그(FLAG)® 에피토프 서열은 재조합 불포화 효소 유전자 서열의 N-말단 세포질 루프 내에서 암호화되었다.

표 49

프로토테카 모리포르미스(UTEX 1435) 균주 A를 형질 전환하는데 사용된 플라스미드 구조물

플라스미드 구조물	관련 서열 요소	서열 번호
pSZ2124	6S::Crβtub:ScSuc2:Cvnr::PmAmt03:3xFlag-BnFad3:Cvnr::6S	서열 번호 222
pSZ2125	6S::Crβtub:ScSuc2:Cvnr::PmAmt03:3xFlag-CsFad7-1:Cvnr::6S	서열 번호 223
pSZ2126	6S::Crβtub:ScSuc2:Cvnr::PmAmt03:3xFlag-LuFad3A:Cvnr::6S	서열 번호 224
pSZ2127	6S::Crβtub:ScSuc2:Cvnr::PmAmt03:3xFlag-LuFad3B:Cvnr::6S	서열 번호 225

[0767]

[0768] 구조물 pSZ2124, pSZ2125, pSZ2126 및 pSZ2127 각각은 균주 A에 개별적으로 도입되어 이 균주를 형질 전환하였다. 1차 형질 전환체를 유일한 탄소원으로서 수크로스를 함유하는 한천 평판 상에서 선택하였다. 각각의 형질 전환체를 클론으로 정제하였으며, 이 클론을 (실시예 1에 개시된 조건과 유사한) 지질 생산에 적당한 조건 하에서 생육하였다(pH 7.0). 각각의 형질 전환체로부터 유래하는 무수 바이오매스로 지질 샘플을 제조하였다. 각각의 형질 전환체로부터 유래하는 무수 바이오매스 20㎎ 내지 40㎎을 MeOH 중 5% H2SO4 2㎖ 중에 재현탁하고 나서, 여기에 적당한 내부 표준(C19:0)을 적당량 함유하는 톨루엔 200ul을 첨가하였다. 혼합물을 간단히 초음파 처리하여 바이오매스를 분산한 다음, 이를 3.5시간 동안 70℃ 내지 75℃에서 가열하였다. 여기에 헵탄 2㎖를 첨가하여 지방산 메틸 에스테르를 추출한 후, 다시 6% K2CO3(aq) 2㎖를 첨가하여 산을 중화하였다. 상기 혼합물을 격렬하게 교반하고, 상층 일부를, 표준 FAME GC/FID법(지방산 메틸 에스테르 가스 크로마토그래피 불꽃 이온화 검출법)을 사용하는 가스 크로마토그래피 분석을 위해서, Na2SO4(무수)가 담긴 바이알에 옮겨 담았다. 표준 지방산 메틸 에스테르 가스 크로마토그래피 불꽃 이온화(FAME GC/FID) 검출법을 사용하여 지방산 프로필을 분석하였다. pSZ2124, pSZ2125, pSZ2126 및 pSZ2127으로 균주 A를 형질 전환하여 생성된 대표적인 클론 세트로부터 얻어진 지방산 프로필(총 지방산의 면적%로 표시)을 표 50에 나타내었다. 비교용으로서, 형질 전환되지 않은 균주 A 대조군 세포로부터 얻은 지질의 지방산 프로필을 표 50에 추가로 제시하였다.

표 50

고등 식물의 외인성 불포화 효소를 발현하도록 조작된 프로토테카 모리포르미스 세포의 불포화 C18:1, C18:2 및 C18:3 지방산 프로필

		총 지방산%		
균주	샘플	C18:1	C18:2	C18:3
형질전환되지 않은 균주 A	1	55.68	7.99	0.63
	2	55.31	8.16	0.7
pSZ2124 브라시카 나푸스 FAD3	형질전환체 1	60.42	1.62	7.69
	형질전환체 2	59.7	0.6	8.49
	형질전환체 3	60.56	1.19	8.15
	형질전환체 4	59.85	0.8	7.9
pSZ2125 카멜리나 사티바 FAD7	형질전환체 1	57.11	9.45	1.4
	형질전환체 2	57.5	8.56	1.39
	형질전환체 3	56.27	8.78	1.39
	형질전환체 4	52.57	9.39	1.7
pSZ2126 리넘 유시타티시멈 FAD3A	형질전환체 1	58.97	0.84	9.67
	형질전환체 2	57.93	1.36	11.92
	형질전환체 3	59.37	0.58	10.33
	형질전환체 4	59.05	0.49	10.24
pSZ2127 리넘 유시타티시멈 FAD3B	형질전환체 1	59.6	1.67	8.67
	형질전환체 2	59.73	1.02	8.6
	형질전환체 3	60.04	1.6	8.74
	형질전환체 4	58.57	0.89	9.05

[0769]

[0770] 형질 전환되지 않은 프로토테카 모리포르미스(UTEX 1435) 균주 A는 C18:3 지방산 1% 미만을 비롯한 지방산 프로필을 나타냈다. 이와는 반대로, 고등 식물 지방산 불포화 효소를 발현하는 균주 A의 지방산 프로필의 경우에는 C18:3 지방산 조성이 약 2배 내지 17배 증가하였음이 보였다. 리넘 유시타티시멈의 FAD3A 또는 FAD3B, 또는 브라시카 나푸스의 FAD3 유전자 생성물을 발현하는 조작 균주는 최대의 C18:3 증가 정도를 보였다(표 50). 18:3 대 총 C18 불포화물의 비율은, 형질 전환되지 않은 균주 내에서는 약 1%였으며, 형질 전환된 균주 내에서는 약 2% 내지 17%였다. 18:2 대 총 C18:0의 비율은, 형질 전환되지 않은 균주 내에서는 약 12% 내지 13%였으며, 형질 전환된 균주 내에서는 약 1% 내지 15%였다(FAD3A 형질 전환체 내 최저 수준). 이와 같은 데이터는, Δ15 불포화 효소 지방산 불포화 효소의 외부 발현이, 조작된 미생물의 지방산 프로필을 바꾸도록(특히 미생물 세포 내 18:3 지방산 농도가 증가되도록) 허용하는 폴리뉴클레오티드의 유용성과 효율을 입증하였다.

[0771] **실시예 12: 헤어핀 RNA 접근법을 통해 스테아르산 및 스테아레이트의 생산을 증가시키기 위한 미생물 조작**

[0772] 스테아로일 ACP 불포화 효소(SAD)는 지질 합성 경로의 일부이다. 이 효소는 지방 아실 사슬에 이중 결합을 도입하는 역할을 한다. 예를 들어 SAD 효소는 C18:0 지방산(스테아르산)으로부터 C18:1 지방산이 합성되는 과정을 촉진한다. 실시예 6에 나타낸 바와 같이, 표적화 유전자 중단을 통해 프로토테카 모리포르미스의 SAD2 대립유전자를 중단시키면, 조작된 미생물의 지방산 프로필에 있어서 C18:0 지방산 수준이 측정 가능할 정도로 증가되었다. 본 실시예는, 형질 전환된 미생물의 지방산 프로필에 있어서 포화 C18:0 지방산이 증가하도록 미생물을 조작하기 위해, SAD2의 발현을 하향 조절하는 헤어핀 RNA를 암호화하는 폴리뉴클레오티드를 사용하는 것에 관하여 기술하고 있다.

[0773] 표 51에 나열된 4개의 구조물 pSZ2139 내지 pSZ2142를, 프로토테카 모리포르미스 SAD2 유전자 생성물의 발현이 감소되도록 설계하였다. 각각의 구조물은 프로토테카 모리포르미스 SAD2 mRNA 전사체에 대해 표적화된 헤어핀 RNA를 암호화하는 상이한 핵산 서열(이 경우, 상기 헤어핀 구조의 줄기부 길이는 180 내지 240 염기쌍임), 6S 게놈 영역에 대한 5'(서열 번호 82) 및 3'(서열 번호 84) 상동성 재조합 표적화 서열(구조물에 측접함)을 함유하였는데, 이 재조합 표적화 서열은, 씨.레인하르드티 β-튜불린 프로모터/5'UTR 및 클로렐라 불가리스 질산염 환원 효소 3'UTR의 제어 하에 있는 에스.세레비지아에 suc2 수크로스 인버타제 암호화 영역과 핵내 게놈으로의

통합에 사용되었다. 이와 같은 에스.세레비지아에 suc2 발현 카세트는 서열 번호 159로서 나열하였으며, 이는 선택 마커로서 사용되었다. 각 구조물의 SAD2 RNA 헤어핀을 암호화하는 폴리뉴클레오티드 서열은 씨.레인하르드티 β-튜불린 프로모터/5'UTR 및 씨. 불가리스 질산염 환원 효소 3'UTR의 제어 하에 두었다.

표 51

[0774]

프로토테카 모리포르미스(UTEX 1435) 균주 A를 형질 전환하는데 사용되는 플라스미드 구조물		
플라스미드 구조물	관련 서열 요소	서열 번호
SZ2139 헤어핀 A	6S::Cr β tub:ScSuc2:Cvnr:Cr β tub:PmSAD2-hpA:Cvnr::6S	서열 번호 226
SZ2140 헤어핀 B	6S::Cr β tub:ScSuc2:Cvnr:Cr β tub:PmSAD2-hpB:Cvnr::6S	서열 번호 227
SZ2141 헤어핀 C	6S::Cr β tub:ScSuc2:Cvnr:Cr β tub:PmSAD2-hpC:Cvnr::6S	서열 번호 228
SZ2142 헤어핀 D	6S::Cr β tub:ScSuc2:Cvnr:Cr β tub:PmSAD2-hpD:Cvnr::6S	서열 번호 229

[0775] (고급 오일을 생산하도록) 전통적 방식으로 돌연 변이된 프로토테카 모리포르미스 UTEX 1435 파생 균주, 균주 A에, 실시예 2에 상세히 기술된 유전자 총 형질 전환법에 따라서 표 51에 나열된 플라스미드 구조물들을 각각 도입하여 상기 균주를 형질 전환하였다. 유일한 탄소원으로서 수크로스를 함유하는 한천 평판 상에서 1차 형질 전환체를 선택한 다음, 클론으로서 정제한 후, 표준 지질 생산 조건 하에서 생육하였다. 실시예 11에 기술된 바와 같은 직접 에스테르 결합 전이법을 이용하여 지방산 프로필을 측정하였다. 균주 A를 pSZ2139, pSZ2140, pSZ2141 및 pSZ2142로 형질 전환하여 생성된 대표적인 클론 세트로부터 얻어진 지방산 프로필(총 지방산의 면적%로 표시)을 이하 표 52에 제시하였다. 비교를 위해서, 형질 전환되지 않은 균주 A 대조군 세포로부터 얻은 지질의 지방산 프로필을 표 52에 추가로 제시하였다.

표 52

[0776]

스테아로일 ACP 불포화 효소 유전자/유전자 생성물을 표적화하는 헤어핀 RNA 구조물을 발현하도록 조작된 프로토테카 모리포르미스 세포의 C18:0, C18:1 및 C18:2 지방산 프로필

		총 지방산%			
균주/플라스미드 구조물	형질전환체	C18:0	C18:1	C18:2	C18 포화:C18 불포화 비율%
균주 A	비형질전환	2.77	60.74	7.27	4
균주 A/pSZ2139 헤어핀A	형질전환체 1	6.39	51.69	9.06	11
	형질전환체 2	5.49	52.89	9.25	9
	형질전환체 3	3.39	56.12	8.85	5
	형질전환체 4	3.24	54.55	8.62	5
균주 A/pSZ2140 헤어핀B	형질전환체 1	22.14	36.14	8.13	50
	형질전환체 2	17.19	41.17	8.31	35
	형질전환체 3	9.45	49.81	8.79	16
	형질전환체 4	5.61	53.8	9.02	9
균주 A/pSZ2141 헤어핀C	형질전환체 1	20.7	40.96	6.45	44
	형질전환체 2	16.33	45.57	7.31	31
	형질전환체 3	13.43	44.79	9.04	25
	형질전환체 4	12.7	46.25	9.98	23
	형질전환체 5	8.47	50.65	9.12	14
균주 A/pSZ2142 헤어핀D	형질전환체 1	26.99	30.93	8.31	69
	형질전환체 2	10.96	47.27	9.9	19
	형질전환체 3	8.64	50.77	11.7	14
	형질전환체 4	7.67	49.76	9.39	13

[0777] 상기 표 52에 제시된 데이터는, 숙주 유기체의 C18:0 및 C18:1 지방산 프로필에 SAD2 헤어핀 RNA 구조물 발현이

미치는 분명한 영향력을 보여주는 것이다. SAD2 헤어핀 RNA 구조물을 포함하는 균주 A 형질 전환체의 지방산 프로필은, 불포화 C18:1 지방산의 백분율은 감소하면서 포화 C18:0 지방산의 백분율은 증가하였음을 입증하는 것이다. 형질 전환되지 않은 균주의 지방산 프로필 중 C18:0는 약 3%였다. 형질 전환된 균주의 지방산 프로필은 C18:0 3% 이상 내지 거의 27%까지였다. C18:0 대 총 C18 불포화물의 비율은, 형질 전환되지 않은 균주 내에서는 약 4%였으며, 형질 전환된 균주 내에서는 약 5% 내지 69%였다. 이와 같은 데이터는, 프로토테카 모리포르미스 내 폴리뉴클레오티드 SAD RNA 헤어핀 구조물이 성공적으로 발현 및 사용된 결과, 조작된 숙주 미생물 내 포화 지방산 백분율이 바뀌었음(특히 미생물 세포 내 C18:0 지방산 농도는 증가되고 C18:1 지방산 농도는 감소하였음)을 입증하였다.

[0778] **실시예 13: β-케토아실-ACP 합성 효소 II 유전자의 과발현을 통한 미세 조류의 지방산 프로필 변경**

[0779] β-케토아실-ACP 합성 효소 II(KASII)는 지방산 생합성시 C16:0-ACP로부터 C18:0-ACP로의 2-탄소 연장을 촉진하였다. 플라스미드 구조물을 제조하여, KASII 유전자가 발현될 경우 숙주 세포의 지방산 프로필이 영향받을 수 있는지 여부를 평가하였다. 프로토테카 모리포르미스 UTEX 1435 또는 고등 식물(글리신 맥스: GenBank 수탁 번호 제AAW88763호; 헬리안터스 아너스; GenBank 수탁 번호 제ABI18155호; 및 리시누스 커뮤니스: GenBank 수탁 번호 제AAA33872호)로부터 KASII 유전자 서열의 공급원을 선택하였다.

[0780] (고급 오일을 생산하도록) 전통적 방식으로 돌연 변이된 프로토테카 모리포르미스 UTEX 1435 파생 균주, 균주 A에, 실시예 2의 방법에 따라서 표 53에 나열된 플라스미드 구조물들 중 하나를 각각 도입하여 이 균주를 형질 전환하였다. 각각의 구조물은 6S 게놈 영역에 대한 5'(서열 번호 82) 및 3'(서열 번호 84) 상동성 재조합 표적화 서열(구조물에 측접함)을 함유하였는데, 이 재조합 표적화 서열은, 씨.레인하르드티 β-튜불린 프로모터/5'UTR 및 클로렐라 불가리스 질산염 환원 효소 3'UTR의 제어 하에 있는 에스.세레비지아에 suc2 수크로스 인버타제 암호화 영역과 핵내 게놈으로의 통합에 사용되었다. 이와 같은 에스.세레비지아에 suc2 발현 카세트는 서열 번호 29로서 나열하였으며, 이는 선택 마커로서 사용되었다. 각각의 구조물에 있어서, KASII 암호화 영역은 프로토테카 모리포르미스 Amt03 프로모터/5'UTR(서열 번호 37) 및 씨.불가리스 질산염 환원 효소 3'UTR의 제어 하에 두었다. 각각의 KASII 효소의 천연 운반 펩티드를 클로렐라 프로토테코이데스 스테아로일-ACP 불포화 효소 운반 펩티드(서열 번호 54)로 치환하였다. 모든 단백질 암호화 영역은 표 2에 따라서 프로토테카 모리포르미스 UTEX 1435 핵내 유전자에 고유한 코돈 편향 사용을 반영하도록 최적화된 코돈이었다.

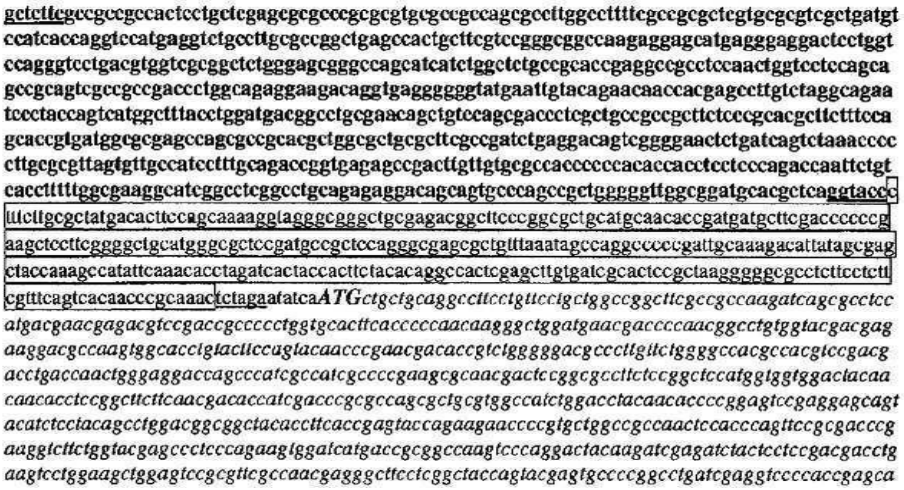
표 53

[0781]

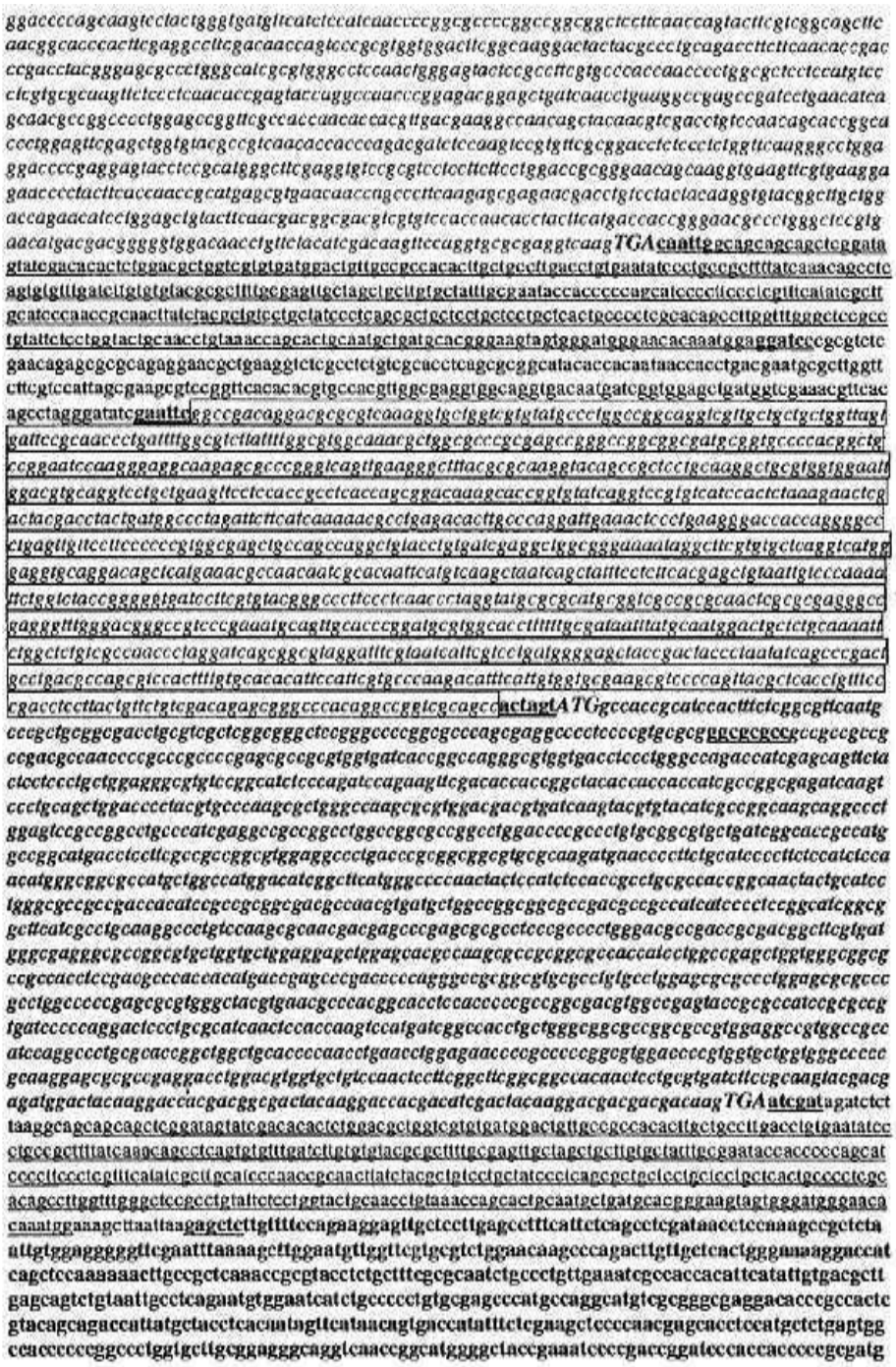
프로토테카 모리포르미스(UTEX 1435) 균주 A 를 형질 전환하는데 사용되는 플라스미드 구조물			
플라스미드 구조물	KASII 효소의 공급원	서열 요소	서열 번호
SZ1747	글리신 맥스 (*Glm*)	6S::β-tub:suc2:nr::Amt03:S106SAD:GlmKASII:nr::6S	서열 번호 230
SZ1750	헬리안터스 아너스 (*Ha*)	6S::β-tub:suc2:nr::Amt03:S106SAD:HaKASII:nr::6S	서열 번호 231
SZ1754	*리시누스 커뮤니스* (*Rc*)	6S::β-tub:suc2:nr::Amt03:S106SAD:RcKASII:nr::6S	서열 번호 232
SZ2041	프로토테카 모리포르미스 (*Pm*)	6S::β-tub:suc2:nr::Amt03:S106SAD:PmKASII:nr::6S	서열 번호 233

[0782] 구조물 6S::β-Tub:suc2:nr::Amt03:S106SAD :PmKASII:nr::6S 내 관련 제한 위치들은 굵은 글씨체의 소문자에 밑줄을 쳐서 표시하였으며, 각각 5'-3' BspQ 1, Kpn I, Xba I, Mfe I, BamH I, EcoR I, Spe I, Asc I, Cla I, Sac I, BspQ I에 해당하였다. BspQI 위치들은 형질 전환 DNA의 5' 및 3' 말단을 한정하였다. 굵은 글씨체의 소문자 서열은 상동성 재조합을 통하여 6S 위치에서의 표적화 통합을 허용하는, 균주 A 유래 게놈 DNA를 나타낸다. 5'→3' 방향으로 진행되는, 효모 수크로스 인버타제 유전자(균주 A가 수크로스를 대사할 수 있는 능력을 부여함)의 발현을 작동하는 씨.레인하르드티 β-튜불린 프로모터는 박스를 쳐서 표시하였다. 인버타제에 대한 개시 인자 ATG 및 종결 인자 TGA는 굵은 이탤릭체의 대문자로 표시한 반면에, 암호화 영역은 이탤릭체의

소문자로 표시하였다. 클로렐라 불가리스 질산염 환원 효소 3'UTR은 소문자에 밑줄을 쳐서 표시하였으며, 그 다음에 존재하는 피.모리포르미스의 내인성 amt03 프로모터는 이탤릭체에 박스를 쳐서 표시하였다. PmKASII의 개시 인자 ATG 및 종결 인자 TGA 코돈은 굵은 이탤릭체의 대문자로 표시하였으며, 유전자의 나머지 부분은 굵은 이탤릭체로 표시하였다. 클로렐라 프로토테코이데스 S106 스테아로일-ACP 불포화 효소 운반 펩티드는 개시 인자 ATG와 Asc I 위치 사이에 위치하였다. 씨.불가리스 질산염 환원 효소 3'UTR은 또한 소문자에 밑줄을 쳐서 표시하였으며, 그 다음에 존재하는 균주 A 6S 게놈 영역은 굵은 소문자로 표시하였다. 구조물 6S::β-tub:suc2:nr::Amt03:S106SAD:PmKASII:nr::6S의 관련 뉴클레오티드 서열은 서열 번호 234로서 서열 목록에 제공하였다. 클로렐라 프로토테코이데스 S106 스테아로일-ACP 불포화 효소 운반 펩티드를 포함하는 PmKASII의 코돈 최적화된 서열은 서열 목록 중 서열 번호 235로서 제공하였다. 서열 번호 236은 서열 번호 235의 단백질 번역물을 제공하였다.



[0783]



[0784]

ggaagaatctctccccgggatgtgggcccaccaccagcacaacctgctggcccaggcgagcgtcaaacccatacccacacaaatatccttgg
catcggccctgaattccttctgccgctctgctacccggtgcttctgtccgaagcagggggttgctagggatcgctccgagtccgcaaacccttg
tcgcgtggcggggcttgttcgagcttgaagagc

[0785]

[0786] 각각의 플라스미드 구조물을 균주 A에 도입하여 이 균주를 별도로 형질 전환하고 나서, 양성 클론을 유일한 탄소원으로서 수크로스를 포함하는 한천 평판 상에서 선택하였다. 상기 실시예들에서와 같이, 1차 형질 전환체는 클론으로서 정제하였으며, 또한 pH 7의 표준 지질 생산 조건 하에 생육하였고, 지질 샘플은 각각의 형질 전환체의 무수 바이오매스로 제조하였다. 지방산 프로필을 실시예 11에 기술된 직접 에스테르 결합 전이법을 사용하여 측정하였다. 균주 A를 형질 전환하여 생성된 대표적인 클론 세트에서 얻은 지방산 프로필(총 지방산의 면적%로 표시)과, 몇몇 양성 형질 전환체의 지방산 프로필(면적%로 표시)을 형질 전환되지 않은 균주 A 대조군의 지방산 프로필과 비교한 결과를, 이하 표 54에 제시된 각각의 플라스미드 구조물에 대해 요약하였다.

표 54

KAS II 유전자를 과발현하도록 조작된 프로토테카 모리포르미스 세포의 지방산 프로필

플라스미드 구조물	KASII 공급원	형질전환체	%C14:0	%C16:0	%C18:0	%C18:1	%C18:2
무	과발현 안됨	1	1.36	28.69	2.92	56.36	8.16
		2	1.35	28.13	3.57	55.63	8.79
		3	1.22	25.74	2.82	60.6	7.31
		4	1.22	25.74	2.82	60.6	7.31
pSZ1747	글리신 맥스	1	2.23	25.34	2.69	57.35	9.53
		2	2.18	25.46	2.74	57.35	9.46
		3	2.18	25.33	2.89	57.34	9.5
		4	2.2	25.69	2.66	57.28	9.43
		5	2.17	25.38	3.03	56.99	9.72
pSZ1750	에이치. 아너스	1	2.43	26.82	2.72	55.17	9.87
		2	2.44	27.14	2.62	54.89	9.81
		3	2.61	26.9	2.67	54.43	10.25
		4	1.96	30.32	2.87	53.87	8.26
		5	2.55	27.64	2.98	53.82	10.07
pSZ1754	리시누스 커뮤니스	1	1.84	24.41	2.89	59.26	9.08
		2	1.3	25.04	2.81	58.75	9.65
		3	1.27	25.98	2.76	58.33	9.22
		4	1.95	25.34	2.77	58.15	9.22
		5	1.3	26.53	2.75	57.87	9.09
pSZ2041	피. 모리포르미스	1	1.63	11.93	3.62	70.95	9.64
		2	1.85	11.63	3.34	69.88	10.93
		3	1.84	12.01	3.81	69.56	10.45
		4	1.63	14.22	3.72	68.86	9.6
		5	1.67	15.04	3.05	68.63	9.24

[0787]

[0788] pSZ2041을 사용하여, 표 2에 제시된 코돈 사용 빈도를 이용하여 추가로 코돈 최적화된 프로토테카 모리포르미스(UTEX 1435) KASII 유전자를 과발현하였을 때, C16:0 사슬 길이는 명확히 감소됨과 아울러 C18:1 지방산의 길이는 증가하는 것이 관찰되었다. β-튜불린 프로모터에 의해 작동되는 프로토테카 모리포르미스(UTEX 1435) KASII 유전자를 발현하는 구조물로 형질 전환하였을 때에도 유사한 지방산 프로필 변화가 관찰되었다.

[0789] 이와 같은 결과들은, 코돈 최적화된 프로토테카 지질 생합성 유전자가 외부 과발현될 때 유전자 조작된 미세 조류의 지방산 프로필이 변경될 수 있음을 보여주는 것이다. 특히 형질 전환되지 않은 세포 내에서의 KASII 유전자의 과발현은 C18 지방산의 백분율을 약 68%에서 약 84%로 증가시킬 수 있었다.

[0790] **실시예 14: 조작된 프로토테카 내 중간 길이 사슬 지방산 수준의, KASI 대립 유전자의 표적화 녹아웃을 통한 변화**

[0791] β-케토아실-ACP 합성 효소 I(KASI)은, 지방산 생합성시 C4:0, C6:0, C8:0, C10:0, C12:0 및 C14:0 지방 아실-ACP 분자의 2-탄소 연장을 촉진하였다. 본 실시예에서는, 녹아웃 플라스미드 구조물인 pSZ2014를 제조하여, KASI 유전자 위치가 표적 중단되었을 때 숙주 세포의 지방산 프로필에 미치는 영향을 평가하였다.

[0792] (고급 오일을 생산하도록) 전통적 방식으로 돌연 변이된 프로토테카 모리포르미스 UTEX 1435 파생 균주, 균주 A에, 실시예 2에 기술된 유전자 총 형질 전환법에 따라서 pSZ2014 구조물을 도입하여 이 균주를 형질 전환하였다. pSZ2014는 씨.레인하르드티 β-튜불린 프로모터 및 클로렐라 불가리스 질산염 환원 효소 3'UTR의 제어 하에 있는 에스.세레비지아에 *suc2* 인버타제 발현 카세트를 포함하였는데, 상기 발현 카세트는, 프로토테카 모리포르미스 게놈의 KASI 위치에 통합되는데 사용되는 구조물을 표적화하는 KASI 유전자 특이적 상동성 영역의 양 말단 중 어느 한쪽에 측접하였다. 이와 같은 에스.세레비지아에 *suc2* 발현 카세트는 서열 번호 159로서 나열하였으며, 또한 선택 마커로 사용되었다. 핵내 게놈 통합용 KASI 위치에 대한 표적화 영역과 관련된 서열을 이하에 보였으며, 또한 서열 번호 238과 서열 번호 239에 나열하였다. pSZ2014내 관련 제한 위치들(굵은 글씨체

의 소문자에 밑줄을 쳐서 표시함)로서는 각각 5'→3' *BspQ* 1, *Kpn* I, AscI, *Xho* I, *Sac* I, *BspQ* I이 있으며, 이것들을 이하 서열에 보였다. *BspQI* 위치는 형질 전환 DNA의 5' 말단 및 3' 말단을 한정하였다. 굵은 글씨체의 소문자로 표시한 서열은, 상동성 재조합을 통해 KASI 위치에서의 표적화 통합을 허용하는, 균주 A 유래 게놈 DNA를 나타내는 것이다. 5'→3' 방향으로 진행될 때, 씨.레인하르드티 b-튜불린 프로모터(코돈 최적화 효모 수크로스 인버타제 유전자를 발현시킴)(균주 A에 수크로스를 대사할 수 있는 능력을 부여함)는 글자에 박스를 쳐서 표시하였다. *suc2* 인버타제에 대한 개시 인자 ATG 코돈과 종결 인자 TGA 코돈은 굵은 이탤릭체 대문자로 표시한 반면에, 암호화 영역은 이탤릭체 소문자로 표시하였다. 클로렐라 불가리스 질산염 환원 효소 3'UTR은 소문자에 밑줄을 쳐서 표시하였다. pSZ2014의 형질 전환 서열을 이하에 보였으며, 서열 번호 237로서 나열하였다.

[0793]

gctcttcgctcaccgcgtgaattgctgtcccaaacgtaagcatcatcgtggctcggtcacgcgatcctggatccgggggatcctagaccgctggtggagagc
gctgccgtcggattggtggcaagtaagattgcgcaggttggcgaagggagagaccaaaaccggaggctggaagcgggcacaacatcgtattattgcgt
atagtagagcagtggcagtcgcatttcgaggtccgcaacggatctcgcaagctcgctacgctcacagtcggagcgaaaggggaccactgccccctgccaga
atggtcgcgaccctctcccctcgccggccccgcctgcaacacgcagtgcgtatccggcaagcgggctgtcgccttcaaccgccccatgttggcgtccggg
ctcgatcaggtgcgctgaggggggtttggtgtgcccgcgcctctgggcccgtgtcggccgtgcggacgtggggccctgggcagtggatcagcagggtttg
cgtgcaaatgcctataccggcgattgaatagcgatgaacgggatacggttgcgctcactccatgcccatgcgaccccgtttctgtccgccagccgtggtcg
ccgggctgcgaagcgggaccccacccagcgcattgtgatcaccggaatgggcgtgggtaccctttcttgcgctatgacacttccagcaaaaggtagggc
gggctgcgagacggcttcccggcgctgcatgcaacaccgatgatgcttcgaccccccgaagctccttcggggctgcatgggcgctccgatgccgctccagggc
gagcgctgtttaaatagccaggccccccgattgcaaagacattatagcgagctaccaaagccatattcaaacacctagatcactaccacttctacacaggccactcga
gcttgtgatcgcactccgctaagggggcgcctcttcctcttcgtttcagtcacaaccgcaaacggcgcgccATGctgctgcaggccttcctgttccttgctggcc
ggcttcgccgccaagatcagcgcctccatgacgaacgagacgtccgaccgccccctggtgcacttcaccccccaacaagggctggatgaacgaccccaacg
gcctgtggtacgacgagaaggacgccaagtggcacctgtacttccagtacaaccgaacgacaccgtctgggggacgccctgttctggggccacgccacg
tccgacgacctgaccaactgggaggaccagcccatcgccatcgccccgaagcgcaacgactccggcgccttctccggctccatggtggtggactacaaca
acacctccggcttcttcaacgacaccatcgaccgcgccagcgctgcgtggccatctggacctacaacaccccggagtccgaggagcagtacatctcctaca
gcctggacggcggctacaccttcaccgagtaccagaagtaccccgtgctggccgccaactccaccgagttccgcgaccccgaagggtcttctggtacgagccc
tcccagaagtggatcatgaccgcggccaagtcccaggactacaagatcgagatctactcctccgacgacctgaagtcctggaagctggagtccgcgttcgc
caacgagggcttcctcggctaccagtacgagtgccccggcctgatcgaggtccccaccgagcaggacccccagcaagtcctactgggtgatgttcatctccat
caaccccggcgccccggccgcggctccttcaaccagtacttcgtcggcagcttcaacggcaccccacttcgaggccttcgacaaccaggtcccgcgtggtgga
cttcggcaaggactactacgccctgcagaccttcttcaacaccgaccccgacctacgggagcgcccttgggcatcgcgtgggcctccaactgggagtacttccgc
cttcgtgcccaccaaccctggcgctcctccatgtccctcgtgcgcaagttctcccgcaacaccgagtaccaggcccaccccggagacggagctgatcaaccctg
aaggccgagccgatcctgaacatcagcaacgccggccccctggagccggttcgccaccaacaccacgttgacgaaggccaacagctacaacgtcgacctg
tccaacagcaccggcaccctggagttcgagctggtgtacgccgtcaacaccaccccagacgatctccaagtccgtgttcgcggacctctccctctggttcaagg
gcctggaggaccccgaggagtacctccgcatgggcttcgaggtgtccgcgtcctccttcttcctggaccgcgggaacagcaaggtgaagttcgtgaaggag
aaccctacttcaccaaccgcatgagcgtgaacaaccagcccttcaagagcgagaacgacctgtcctactacaaggtgtacggcttgctggaccagaacat
cctggagctgtacttcaacgacggcgacgtcgtgtccaccaacacctacttcatgaccaccgggaacgccctgggctccgtgaacatgacgacggggtgg
acaacctgttctacatcgacaagttccagggtgcgcgaggtcaagTGAcaattggcagcagcagctcggatagtatcgacacactctggacgctggtcgtgtg
atggactgttgccgccacacttgctgccttgacctgtgaatatccctgccgcttttatcaaacagcctcagtgtgtttgatcttgtgtgtacgcgcttttgcgagttgctag
ctgcttgtgctatttgcgaataccacccccagcatccccttccctcgtttcatatcgcttgcatcccaaccgcaacttatctacgctgtcctgctatcccagcgctgctc
ctgctcctgctcactgcccctcgcacagccttgctttggctccgcctgtattctcctggtactgcaacctgtaaaccagcactgcaatgctgatgcacgggaagtag
tggatgggaacacaaatggaggatcgtagagctccaacctgcatccgcctggcgctcgaggacgccggcgtctcgcccgacgaggtcaactacgtcaacg
cgcacgccacctccaccctggtgggcgacaaggccgaggtgcgcgcggtcaagtcggtctttggcgacatgaagggcatcaagatgaacgccaccaag
tccatgatcgggcactgcctgggcgccgccggcggcatggaggccgtcgccacgctcatggccatccgcaccggctgggtgcacccccaccatcaaccac
gacaacccccatcgccgaggtcgacggcctggacgtcgtcgccaacgccaaggcccagcacaaaatcaacgtcgccatctccaactccttcggcttcggc
gggcacaactccgtcgtcgccttcgcgccttccgcgagtaggcggagcgagcgcgcttggctgaggagggaggcggggtgcgagccctttggctgcgc

[0794]

gcgatactctcccgcacgagcagactccacgcgcctgaatctacttgtcaacgagcaaccgtgtgttttgtccgtggccattcttattatttctccgactgtg
gccgtactctgtttggctgtgcaagcaccgaagagc

[0795] 플라스미드 구조물 pSZ2014를 균주 A에 도입하여 이 균주를 형질 전환하고 나서, 유일한 탄소원으로서 수크로스를 함유하는 평판 상에서 선택하였다. 1차 형질 전환체를 클론으로서 정제한 다음, 이를 표준적인 지질 생산 조건 하에서 생육하였다. 각각의 형질 전환체로부터 얻은 무수 바이오매스로부터 지질 샘플을 제조하였다. 실시예 11에 기술된 직접 에스테르 결합 전이법을 사용하여 지방산 프로필을 측정하였다. 몇몇 양성 형질 전환체의 지방산 프로필(총 지방산의 면적%로 표시)을, 형질 전환되지 않은 균주 A 대조군의 지방산 프로필과 비교하였으며, 그 결과는 하기 표 55에 요약되어 있다.

표 55

내인성 KASI 대립 유전자를 중단하도록 조작된, 선택가능한 마커 포함 프로토테카 모리포르미스 세포의 지방산 프로필

	형질전환체	% C14:0	% C16:0	% C18:0	% C18:1	% C18:2
균주 A	비형질전환 대조군	1.22	25.61	2.82	60.76	7.44
균주 A pSZ2014	형질전환체 1	1.65	32.55	2.17	53.99	7.43
	형질전환체 2	2.25	30.04	2.57	55.86	7.12
	형질전환체 3	3.51	31.22	1.90	53.85	7.00
	형질전환체 4	4.09	31.51	2.57	53.14	6.21
	형질전환체 5	4.68	34.47	1.94	49.75	6.49
	형질전환체 6	5.68	37.98	1.83	44.76	6.75
	형질전환체 7	5.82	37.82	1.93	44.84	6.44

[0796]

[0797] 상기 표 55에 나타낸 바와 같이, KASI 대립 유전자의 표적화 중단에 의하여 형질 전환된 미생물의 지방산 프로필이 영향을 받았음을 알 수 있다. pSZ2014 형질 전환 벡터를 포함하는 균주 A의 지방산 프로필은, C18:1 지방산이 감소됨과 아울러, C14:0 및 C16:0 지방산의 조성이 증가하였음을 보여주었다. 모든 형질 전환체에서, C18:0 지방산은 감소하였다. 몇몇 형질 전환체에 있어서, KASI 대립 유전자가 중단되었을 경우에는, 형질 전환되지 않은 균주 A 유기체의 지방산 프로필에 비하여 C18:2 지방산의 백분율이 감소한 지방산 프로필을 추가로 얻을 수 있었다.

[0798] 그러므로, 본 발명의 발명자들은 내인성 KASI를 중단함으로써 총 C14 지방산 백분율을 약 35%에서 400%로 증가시켰으며, C16 지방산의 백분율은 약 30%에서 약 50%로 증가시켰다.

[0799] 상기 데이터는 숙주 미생물의 지방산 프로필을 변경함에 있어서 내인성 KASI 대립 유전자의 표적화 유전자 중단의 유용성을 입증하는 것이다.

[0800] **실시예 15: 프로토테카 내 지방산 프로필을 변경시키기 위한 유전자 변형 접근법의 통합**

[0801] 본 실시예에서는, KASII 대립 유전자를 녹아웃시키고 아울러 중간쇄인 지방산의 가수 분해를 선호하는 외인성 티오에스테라제를 과발현하도록 만드는 유전자 변형법을 통합 수행하여, 미생물 내 숙주 유기체의 지방산 프로필이 미생물 내에서 변경되는 것을 입증하고 있다.

[0802] 우선, (고급 오일을 생산하기 위하여) 전통적 방식에 따라서 돌연 변이된 프로토테카 모리포르미스 균주(UTEX 1435) 파생 균주, 즉 균주 C를, 실시예 2에 상세히 기술된 유전자 총 형질 전환법에 따라서 플라스미드 구조물 pSZ1283으로 형질 전환하였다. PCT 특허 출원 PCT/US2011/038463 및 PCT/US2011/038463에 이미 개시된 pSZ1283(서열 번호 256)은, 쿠페아 라이티 FATB2(CwTE2) 티오에스테라제의 암호화 서열, 6S 게놈 영역에 대한 5'(서열 번호 82) 및 3'(서열 번호 84) 상동성 재조합 표적화 서열(구조물에 측접함)을 함유하였으며, 이 재조합 표적화 서열은, 씨.레인하르드티 β-튜불린 프로모터/5'UTR 및 클로렐라 불가리스 질산염 환원 효소 3'UTR의 제어 하에 있는 에스.세레비지아에 *suc2* 수크로스 인버타제 암호화 영역과 핵내 게놈으로의 통합에 사용되었다. 이와 같은 에스.세레비지아에 *suc2* 발현 카세트는 서열 번호 159로서 나열하였으며, 이는 선택 마커로서 사용되었다. *Cw*TE2 암호화 서열은 프로토테카 모리포르미스 Amt03 프로모터/5'UTR(서열 번호 89) 및 씨.불가리스 질산염 환원 효소 3'UTR(서열 번호 32)의 제어 하에 두었다. *CwTE2* 및 *suc2*의 단백질 암호화 영역은 표 2에 따른 프로토테카 모리포르미스 UTEX 1435 핵내 유전자에 고유한 코돈 편향 사용을 반영하도록 최적화된 코돈이었다.

[0803] pSZ1283을 균주 C에 도입하여 이 균주를 형질 전환하고 나서, 양성 클론을, 유일한 탄소원으로서 수크로스를 포함하는 한천 평판 상에서 선택하였다. 그 다음, 1차 형질 전환체를 클론으로서 정제한 다음, 단일 형질 전환체 즉, 균주 B를 추가 유전자 변형을 위해서 선택하였다. 이와 같이 유전자 조작된 균주를 플라스미드 구조물 pSZ2110(서열 번호 240)(KASII 대립 유전자 1 위치를 중단함)으로 형질 전환하였다. pSZ2110(KASII'5::CrbTub:NeoR:nr::KASII-3'으로 표시)는, G418에 대한 내성을 부여하고, 씨.레인하르드티 β-튜불린 프로모터와 클로렐라 불가리스 질산염 환원 효소 3'UTR(여기서, 상기 발현 카세트는 프로토테카 모리포르미스 게놈의 KASII 위치에 통합할 때 사용되는 구조물을 표적화하는, KASII 유전자 특이적 상동성 영역 양쪽 말단 중 어느 한쪽에 측접)의 제어 하에 있는 네오마이신 내성(NeoR) 발현 카세트를 포함하였다. pSZ2110 구조

물 내 관련 제한 위치로서는, 5'→3' 방향으로, *BspQ* 1, KpnI, XbaI, *Mfe*I, *BamH*I, *EcoR*I, *Spe*I, *XhoI*, *Sac* I 및 *BspQ*I이 있으며, 이것들은 굵은 글씨체의 소문자에 밑줄을 쳐서 표시하였다. BspQI 위치는 형질 전환 DNA의 5' 및 3' 말단을 한정한다. 형질 전환 구조물의 5' 및 3' 말단에 존재하는 굵은 글씨체의 소문자로 표시된 서열은, 상동성 재조합을 통해 KASII 대립 유전자 1 위치로의 통합을 표적화하는 UTEX 1435 유래 게놈 DNA를 나타낸다. 씨.레인하르드티 β-튜불린은 소문자에 박스를 쳐서 표시하였다. *NeoR* 에 대한 개시 인자 ATG와 종결 인자 TGA는 이탤릭체의 대문자로 표시하였으며, 암호화 영역은 이탤릭체의 소문자로 표시하였다.

[0804] 3'UTR은 소문자에 밑줄을 쳐서 표시하였다.

gctcttcgctggcctgttgccggacgatccgtgtcgtcgagactgcattttgttttgggtgtggggctggggtactggatggcttg
agggcatgactttttctgatggagaagattgcaatgagatcatttgggtcgtctatttgtttgctgtgcaagagggtttactggtat
ctggcaccagcttttggcccgtgcccgtttgatggacgcgtgacaggcaggcgtcctggaaagcacagacaccgtacgtacga
ccttgacctccccccttctccacacggcaggtgcgaggctgcccacggcgtcgaggcgggcggtgcgccgggcatggtcccg
catcgcgcgcgcggcggccgcggccgacgcaaacccgcccgccctgagcgccgcgtggtcatcacgggccagggcgtggt
gaccagcctgggccagacgatcgagcagttttacagcagcctgctggagggcgtgagcggcatctcgcagatacagaagttc
gacaccacgggctacacgacgacgatcgcgggcgagatcaagtcgctgcagctggacccgtacgtgcccaagcgctgggcg
aagcgcgtggacgacgtgataaagtacgtctacatcgcgggcaagcaggcgctggagagcgccggcctgccgatcgaggcg
gcgggcctggcgggcgcgggcctggacccggcgctgtgcggcgtgctcatcggcaccgccatggcgggcatgacgtctttcg
cggcggcgtggaggcgctgacgcgcggtaccctttcttgcgctatgacacttccagcaaaaggtagggcgggctgcgagacgg
cttcccggcgctgcatgcaacaccgatgatgcttcgaccccccgaagctccttcggggctgcatgggcgctccgatgccgctccagg
gcgagcgctgtttaaatagccaggccccgattgcaaagacattatagcgagctaccaaagccatattcaaacacctagatcactacc
acttctacacaggccactcgagcttgtgatcgcactccgctaaggggcgcctcttcctcttcgtttcagtcacaacccgcaaac**tctag**
aatatca*ATGatcgagcaggacggccatccacgccggctccccgccgcctgggtggagcgcctgttcggctacgactgggccc*
agcagaccatcggctgctccgacgccgccgttccgcctgtccgcccagggccgccccgtgctgttcgtgaagaccgacctgtc
cggcgccctgaacgagctgcaggacgaggccgcccgcctgtcctggctggccaccaccgggcgtgccctgcgccgccgtgctgg
acgtggtgaccgaggccggccgcgactggctgctgtgggcgaggtgcccggccaggacctgctgtccttccccaccctggccccg
ccgagaaggtgtccatcatggccgacgccatgcgccgcctgcacacccttggaccccgccacctgccccttcgaccaccaggcca
agcaccgcatcgagcgcgcccgcacccgcatggaggccggcctggtggaccaggacgacctggacgaggagcaccagggc
ctggccccgccgagctgttcgcccgcctgaaggcccgcatgcccgacggcgaggacctggtggtgacccacggcgacgcctg
cctgcccaacatcatggtggagaacggccgcttctccggcttcatcgactgcggccgcctgggcgtggccgaccgctaccaggac
atcgccctggccacccgcgacatcgccgaggagctgggcggcgagtgggccgaccgcttcctggtgctgtacggcatcgccgcc
*cccgactcccagcgcatcgccttctaccgcctgctggacgagttcttc*TGA**caattg**gcagcagcagctcggatagtatcgacaca
ctctggacgctggtcgtgtgatggactgttgccgccacacttgctgccttgacctgtgaatatccctgccgcttttatcaaacagcctcagt
gtgtttgatcttgtgtgtacgcgcttttgcgagttgctagctgcttgtgctatttcgcgaataccacccccagcatccccttccctcgtttcatat
cgcttgcatcccaaccgcaacttatctacgctgtcctgctatcccctcagcgctgctcctgctcctgctcactgcccctcgcacagccttgg
tttgggctccgcctgtattctcctggtactgcaacctgtaaaccagcactgcaatgctgatgcacgggaagtagtgggatgggaacaca
aatggaggatccactagttctagagcggccgccaccgcggtg**gagctc**ggcggcgtgcgcaagatgaacccctttgcatcccct
tctccatctccaacatgggcggcgcgatgctggcgatggcatcggcttcatgggccccaactactccatctccacggcctgcg
cgacgggcaactactgcatcctgggcgcggcggaccacatccggcgcggcgacgcaaacgtgatgctggccggcggcgcgg
acgcggccatcatcccctcgggcatcggcggcttcatcgcgtgcaaggcgctgagcaagcgcaacgacgagcccgagcgcg
cgagccggccctgggacgccgaccgcgacggcttcgtcatgggcgagggcgccggcgtgctggtgctggaggagctggagc
acgccaagcgccgcggcgcgaccatttttggctgaattagttggcggcgggccacctcggacgcgcaccacatgaccgagcc
cgacccgcagggccgcggcgtgcgcctctgcctcgagcgcgcgctcgagcgcgcgcgcctcgcgcccgagcgcgtcggctac
gtcaacgcgcacggcaccagcacgcccgcgggcgacgtggccgagtaccgcgccatccgcgccgtcatcccgcaggactca
ctacgcatcaactccacaaagtccatgatcgggcacctgctcggcggcgccggcgcggtcgaggccgtggccgccatccagg
ccctgcgcaccggctggctccaccccaacttgaacctcgagaaccccgcgcctggcgtcgaccccgtcgtgctcgtgggctctt
cc

[0805]

[0806] 균주 B를 pSZ2110으로 형질 전환하고 나서, 양성 클론을, G418을 함유하는 선택 한천 평판 상에서 선택하였다. 이후, 1차 형질 전환체를 클론으로서 정제한 다음, 표준 지질 생산 조건(pH 5.0 및 pH 7.0) 하에서 탄소원으로서 수크로스를 기반으로 생육하였다. 실시예 11에 기술된 바와 같이 각각의 형질 전환체로부터 얻어진 무수 바이오매스로부터 지질 샘플을 제조하였다. 5개의 양성 형질 전환체(T1-T5)의 지방산 프로필(총 지방산의 면적 %로 표시), 유일한 탄소원으로서 수크로스를 기반으로 생육된 균주 B(U1)의 프로필, 그리고 유일한 탄소원으로서 글루코스를 기반으로 생육된 비형질 전환 UTEX 1435(U1)의 프로필을 이하 표 56에 제시하였다.

표 56

내인성 KASII 유전자 생성물을 소모시키고, 쿠페아 라이티 티오에스테라제를 발현하도록 중복 조작된 프로토테카 모리포르미스(UTEX 1435)의 지방산 프로필

		균주						
지방산	pH	UTEX 1435	균주 B	균주 B pSZ2011				
		U1	U1	T1	T2	T3	T4	T5
% C10:0	pH 5.0	0.01	*	0.02	0.03	0.03	0.03	0.07
	pH 7.0	0.01	5.35	4.94	4.79	4.70	4.10	4.12
% C12:0	pH 5.0	0.04	*	0.09	0.39	0.41	0.41	0.86
	pH 7.0	0.04	27.06	25.31	25.03	25.02	23.54	23.47
% C14:0	pH 5.0	1.30	*	0.92	1.13	1.14	1.15	1.50
	pH 7.0	1.56	15.20	14.49	14.22	14.50	15.84	15.85
% C16:0	pH 5.0	25.89	*	36.07	35.05	35.35	35.23	35.05
	pH 7.0	29.80	13.89	14.96	14.88	15.11	22.62	22.94
% C18:0	pH 5.0	2.84	*	1.76	1.79	1.84	1.82	1.80
	pH 7.0	3.00	1.44	1.57	1.71	1.49	1.37	1.33
% C18:1	pH 5.0	60.34	*	49.82	50.21	49.97	50.01	49.13
	pH 7.0	54.96	28.57	29.84	30.45	30.27	24.18	23.96
% C18:2	pH 5.0	7.40	*	8.39	8.59	8.47	8.52	8.75
	pH 7.0	8.15	6.85	7.19	7.16	7.15	6.51	6.65

* 테스트하지 않음

[0807]

[0808] 표 56에 보인 바와 같이, 프로토테카 모리포르미스(UTEX 1435) 균주 B 내에서 *Cw*TE2가 발현되었을 때의 영향으로 인하여, 형질 전환된 미생물의 지방산 프로필에는 눈에 띄는 변화가 일어났다. *Cw*TE2를 발현하고, Amt03 프로모터에 의해 CwTE2가 발현되는 것이 촉진되도록 pH 7.0에서 배양된 균주 B의 지방산 프로필에서는, 형질 전환되지 않은 UTEX 1435의 지방산 프로필에 비하여, C10:0, C12:0 및 C14:0 지방산의 조성이 증가함과 아울러, C16:0 및 C18:1 지방산의 조성은 감소되었음이 확인되었다. 이후, C16:0 지방산에서 C18:0 지방산으로의 2-탄소 연장을 촉진하는 효소를 암호화하는 KASII 대립 유전자를 중단하도록 균주 B를 변형시킨 결과, 새로 조작된 균주가 pH 5.0에서 생육되었을 때 이 균주의 지질 프로필에 있어서 C16:0 지방산은 증가하였으며, 아울러 C18:1 지방산은 감소하였다. 배양 배지의 pH는 Amt03 프로모터의 활성에 최적이 아니므로, pH 5.0에서 형질 전환체가 증식하였음은 곧, 티오에스테라제가 지방산 프로필을 바구는데에 기여하는 것과는 별도로, KASII 대립 유전자 녹아웃이 일어났을 때의 영향력을 증명하는 것이다. pH 7.0에서 지질을 생산하여 *Cw*TE2를 발현시키면, pSZ2011 형질 전환체는 UTEX 1435 균주의 지방산 프로필에 비하여, C10:0, C12:0 및 C14:0 지방산의 조성은 증가하고, 아울러 C16:0 및 C18:1 지방산 조성은 감소하는 지방산 프로필을 보였다. 몇몇 pSZ2011 형질 전환체가 pH 7.0에서 배양되었을 때 이 형질 전환체는, pH 7.0에서 배양된 부모 균주 B의 지방산 프로필에 비하여, C16:0 지방산은 증가하고, C18:1 지방산 조성은 더욱 감소한 지방산 프로필을 나타내었다.

[0809] 이와 같은 데이터는, 숙주 유기체의 지방산 프로필에 영향을 미치도록 만드는 중복 유전자 변형의 유용성을 입증하는 것이다. 뿐만 아니라, 본 실시예는 숙주 미생물의 지방산 프로필을 바꾸는데 있어서 내인성 KASII 대립 유전자의 유전자 중단을 표적화하는 재조합 폴리뉴클레오티드의 용도를 입증하는 것이다.

[0810] **실시예 16: 프로토테카의 팔미트산 조성을 변경시키기 위한 유전자 변형 접근법의 통합**

[0811] 본 실시예는, KASII 대립 유전자를 녹아웃시키고, 이와 아울러 C14 및 C16 지방산의 가수 분해에 선호적인 특이성을 나타내는 외인성 티오에스테라제를 과발현하는 유전자 변형법들을 통합하면, 숙주 유기체의 지방산 프로필이 미생물 내에서 바뀐다는 사실을 입증하였다.

[0812] 실시예 2에 상세히 기술된 바와 같은 유전자 총 형질 전환법에 따라서 플라스미드 구조물 pSZ2004를 사용하여,

(고급 오일을 생산하도록) 전통적 방식으로 돌연 변이된 프로토테카 모리포르미스(UTEX 1435) 균주 A를 형질 전환하였다. pSZ2004(KASII_5'_btub-SUC2-nr_2X_Amt03-Ch16TE2-nr_KASII_3'으로 표시)는, 쿠페아 후커리아나 지방 아실-ACP 티오에스테라제의 암호화 서열(Ch16TE2, GenBank #Q39513), 5' 및 3' 상동성 재조합 표적화 서열(구조물에 측접)(핵내 게놈의 KASII 위치에 표적화 통합하는데 사용됨), 그리고 씨.레인하르드티 β-튜불린 프로모터/5'UTR 및 클로렐라 불가리스 질산염 환원 효소 3'UTR의 제어 하에 있는 에스.세레비지아에 *suc2* 수크로스 인버타제 암호화 영역을 포함하였다. Ch16TE2는 C14 및 C16 지방산에 대해 선호적인 특이성을 보이는 티오에스테라제다. 이러한 에스.세레비지아에 *suc2* 발현 카세트를 서열 번호 159에 나열하였으며, 이를 또한 선택 마커로서 사용하였다. Ch16TE 암호화 서열은, 직렬로 반복되는 프로토테카 모리포르미스 Amt03 프로모터/5'UTR(서열 번호 89)와, 씨.불가리스 질산염 환원 효소 3'UTR의 제어 하에 두었다. Ch16TE의 단백질 암호화 영역 및 *suc2*는, 표 2에 따라서 프로토테카 모리포르미스 UTEX 1435 핵내 유전자에 고유한 코돈 편향 사용을 반영하도록 최적화된 코돈이었다. pSZ2004는 서열 목록에 서열 번호 249로서 제시하였다.

[0813] pSZ2004로 형질 전환하고 나서, 1차 형질 전환체를, 유일한 탄소원으로서 수크로스를 함유하는 평판 상에서 선택하였다. 각각의 형질 전환체를 클론으로서 정제하여, 표준 지질 생산 조건(pH 7.0, 실시예 1에 개시된 조건과 유사한 조건) 하에서 생육하였다. 실시예 11에 기술된 바와 같은 표준 지방산 메틸 에스테르 가스 크로마토그래피 불꽃 이온화(FAME GC/FID) 검출법을 사용하여 지방산 프로필을 분석하였다. 형질 전환 벡터 pSZ2004로 형질 전환하여 생산된 대표적인 클론으로부터 얻어진 지방산 프로필(총 지방산의 면적%로 표시)을 표 57에 보였다. 지질 생산 조건(유일한 탄소원으로서 글루코스 포함) 하에서 생육된, 형질 전환되지 않은 균주로부터 얻은 지질의 지방산 프로필을 표 57에 추가로 제시하였다.

표 57

내인성 KASII 유전자 생성물을 소모하고, 쿠페아 후커리아나 티오에스테라제를 발현하도록 중복 조작된 프로토테카 모리포르미스(UTEX 1435)의 지방산 프로필

[0814]

지방산	UTEX 1435	pSZ2004
% C10:0	0.01	0.00
% C12:0	0.04	0.09
% C14:0	1.27	6.42
% C16:0	27.20	69.97
% C18:0	3.85	1.84
% C18:1	58.70	13.69
% C18:2	7.18	7.15

[0815] 상기 표 57에 보인 바와 같이, KASII 대립 유전자를, 선택가능한 마커 및 C14/C16 선호 티오에스테라제를 발현하기 위한 발현 카세트로 표적 중단시키면, 형질 전환된 미생물의 지방산 프로필에 영향을 미쳤다. pSZ2004 형질 전환 벡터를 포함하는 균주는, C14:0 및 C16:0 지방산의 조성이 증가함과 아울러, C18:0 및 C18:1 지방산은 감소하는 지방산 프로필을 보였다. 형질 전환되지 않은 프로토테카 모리포르미스(UTEX 1435) 균주는, C16 지방산 약 27%와 C18:1 지방산 약 58%를 포함하는 지방산 프로필을 나타냈다. 이와는 대조적으로, 쿠페아 후커리아나 지방 아실-ACP 티오에스테라제를 발현시킬 수 있는 카세트와 선택가능한 마커에 의해 KASII 위치가 중단된 균주의 지방산 프로필은, C16 지방산 약 70%와 지방산 약 14%를 포함하였다. C16:0의 수준은 2.5배 이상 증가하였다. 이 데이터는, 외인성 유전자를 과발현하고, 내인성 유전자를 소모시키는 유전자 변형법들은 숙주 유기체 내 지방산 프로필을 변경하도록 통합될 수 있음을 보여주는 것이다.

[0816] 비교를 위하여, 수크로스 인버타제 유전자를 발현시킬 수 있는 카세트에 의해 KASII 위치가 중단된 균주의 지방산 프로필은 C16 지방산이 약 35%이고 C18:1 지방산이 약 50%였다.

[0817] 이와 같은 데이터는, 특히 티오에스테라제 효소를 외부적으로 발현시켜, 미생물 세포 내 C14 및 C16 지방산의 농도를 증가시키고, C18:0 및 C18:1 지방산을 감소시키는 것과 같이, 조작된 미생물의 지방산 프로필을 바꾸는 것을 허용하는 폴리뉴클레오티드로 KASII 대립 유전자를 표적화 중단함으로써, 지방산 프로필을 상기한 바와 같이 바꾸는데 있어서 상기 폴리뉴클레오티드의 유용성과 효율을 입증한다.

[0818] **실시예 17: 리놀레 불포화 지방산 및 글리세로지질을 생산하기 위한 미생물 조작**

[0819] 임의의 Δ12 지방산 불포화 효소는 C18:1 지방산 또는 지방 아실 분자에 이중 결합을 형성하고, 이로써 C18:2 (리놀레) 지방산 또는 지방 아실 분자를 생성하는 것을 촉진할 수 있었다. 리놀레 불포화 지방산이 풍부한 오일을 생산하는 임의의 식물 종, 예를 들어 고시피움 히르서텀, 칼타머스 팅크토리우스, 글리신 맥스, 헬리안터스 아너스 및 제아 메이스는, 미생물 내에서 발현되어 지방산 프로필에 영향을 미칠수 있는 Δ12 불포화 효소 암호화 유전자 공급원이다. 본 실시예는 형질 전환된 미생물의 지방산 프로필에 있어서 리놀레산이 증가한 미생물을 조작함에 있어서, Δ12 불포화 효소를 암호화하는 폴리뉴클레오티드의 용도를 기술하였다.

[0820] (고급 오일을 생산하도록) 전통적 방식으로 돌연 변이된 프로토테카 모리포르미스 UTEX 1435 파생 균주, 균주 A에, 실시예 2에 상세히 기술된 유전자 총 형질 전환법에 따라서 표 58에 나열된 플라스미드 구조물들 중 하나를 도입하여 이 균주를 형질 전환하였다. 각각의 구조물은 6S 게놈 영역에 대한 5'(서열 번호 82) 및 3'(서열 번호 84) 상동성 재조합 표적화 서열(구조물에 측접함)을 함유하였는데, 이 재조합 표적화 서열은, 씨.레인하르드티 β-튜불린 프로모터/5'UTR 및 클로렐라 불가리스 질산염 환원 효소 3'UTR의 제어 하에 있는 에스.세레비지아에 *suc2* 수크로스 인버타제 암호화 영역과 핵내 게놈으로의 통합에 사용되었다. 이와 같은 에스.세레비지아에 *suc2* 발현 카세트는 서열 번호 159로서 나열하였으며, 이는 선택 마커로서 사용되었다. 모든 단백질 암호화 영역은, 표 2에 따라서 프로토테카 모리포르미스 UTEX 1435에 고유한 코돈 편향 사용을 반영하도록 최적화된 코돈이었다. 고시피움 히르서텀 유래 불포화 효소 유전자(*Gh*, GenBank 수탁 번호 제CAA71199호), 칼타머스 팅크토리우스 유래 불포화 효소 유전자(*Ct* GenBank 수탁 번호 제ADM48789호), 글리신 맥스 유래 불포화 효소 유전자(*Gm*, GenBank 수탁 번호 제BAD89862호), 헬리안투스 아너스 유래 불포화 효소 유전자(*Ha*, GenBank 수탁 번호 제AAL68983호) 및 제아 메이스 유래 불포화 효소 유전자(*Zm*, GenBank 수탁 번호 제ABF50053호)의 암호화 영역들은 각각 프로토테카 모리포르미스 Amt03 프로모터/5'UTR(서열 번호 89) 및 씨.불가리스 질산염 환원 효소 3'UTR의 제어 하에 두었다.

표 58

프로토테카 모리포르미스(UTEX 1435) 균주 A를 형질 전환하는데 사용된 플라스미드 구조물

[0821]

플라스미드 구조물	관련 서열 요소	서열 번호
pSZ2150	6S::Crβtub:ScSuc2:Cvnr::PmAmt03:CtFad2-2:Cvnr::6S	서열 번호 241
pSZ2151	6S::Crβtub:ScSuc2:Cvnr::PmAmt03:GlmFad2-2:Cvnr::6S	서열 번호 242
pSZ2152	6S::Crβtub:ScSuc2:Cvnr::PmAmt03:HaFad2:Cvnr::6S	서열 번호 243
pSZ2153	6S::Crβtub:ScSuc2:Cvnr::PmAmt03:ZmFad2:Cvnr::6S	서열 번호 244
pSZ2172	6S::Crβtub:ScSuc2:Cvnr::PmAmt03:GhFad2:Cvnr::6S	서열 번호 245

[0822] 표 58에 나열된 구조물 각각을 개별적으로 균주 A에 도입하여 이 균주를 형질 전환하였다. 유일한 탄소원으로서 수크로스를 함유하는 평판 상에서 1차 형질 전환체를 선택하였다. 각각의 형질 전환체를 클론으로서 정제하였으며, 이를 표준 지질 생산 조건(pH 7.0, 실시예 1에 기술된 조건과 유사한 조건) 하에서 생육하였다. 실시예 11에 기술된 바와 같은 표준 지방산 메틸 에스테르 가스 크로마토그래피 불꽃 이온화(FAME GC/FID) 검출법을 사용하여 지방산 프로필을 분석하였다. 표 58의 상응 균주 A 형질 전환을 통해 생산된 대표적인 클론 세트로부터 얻어진 지방산 프로필을 표 59에 보였다. 비교를 위하여, 형질 전환되지 않은 균주 A 대조군 세포로부터 얻은 지질의 지방산 프로필을 표 59에 추가로 제시하였다.

표 59

외인성 FAD불포화효소를 발현하도록 조작된 프로토테카 모리포르미스 세포의 C18:1, C18:2 및 C18:3 지방산 프로필

균주	샘플	총 지방산%			총 C18 불포화 지방산 (총 지방산%)	C18 다중 불포화 지방산/총C18 불포화 지방산
		C18:1	C18:2	C18:3		
균주A	비형질전환	55.37	8.18	0.7	64.25	13.82
균주A pSZ2150 *Ct* FAD2-2	형질전환체 1	58.29	11.96	0.75	71	17.90
	형질전환체 2	59.00	10.29	0.84	70.13	15.87
	형질전환체 3	52.41	10.36	1.16	63.93	18.02
	형질전환체 4	56.77	10.17	0.8	67.74	16.19
	형질전환체 5	57.19	10.15	0.79	68.13	16.06
균주A pSZ2151 *Glm* FAD2-2	형질전환체 1	58.09	10.32	0.86	69.27	16.14
	형질전환체 2	59.6	11.87	0.86	72.33	17.60
	형질전환체 3	58.93	11.54	0.83	71.3	17.35
	형질전환체 4	58.4	12.29	0.9	71.59	18.42
	형질전환체 5	58.27	10.8	0.83	69.9	16.64
	형질전환체 6	58.85	10.48	0.82	70.15	16.11
균주A pSZ2152 *Ha* FAD2	형질전환체 1	59.30	10.02	0.82	70.14	15.45
	형질전환체 2	58.45	9.87	0.81	69.13	15.45
	형질전환체 3	59.38	9.89	0.79	70.06	15.24
	형질전환체 4	59.54	9.79	0.81	70.14	15.11
	형질전환체 5	59.07	9.92	0.82	69.81	15.38
	형질전환체 6	59.57	10.02	0.55	70.14	15.07
균주A pSZ2153 *Zm* FAD2	형질전환체 1	64.30	11.18	0.89	76.37	15.80
	형질전환체 2	58.54	10.49	0.88	69.91	16.26
	형질전환체 3	58.80	9.95	0.81	69.56	15.47
	형질전환체 4	56.18	10.81	1.16	68.15	17.56
	형질전환체 5	58.82	10.02	0.83	69.67	15.57
	형질전환체 6	58.72	10.06	0.86	69.64	15.68
균주A pSZ2172 *Gh*-FLAG-FAD2	형질전환체 1	55.71	10.85	0.84	67.4	17.34
	형질전환체 2	56.12	10.29	0.75	67.16	16.44
	형질전환체 3	54.14	12	0.96	67.1	19.31
	형질전환체 4	55.72	11.68	0.75	68.15	18.24

[0823]

[0824] 형질 전환되지 않은 프로토테카 모리포르미스(UTEX 1435) 균주는 C18:2 지방산 8.5% 미만을 포함하는 지방산 프로필을 나타냈다. 표 59에 보인 바와 같이, 고등 식물 지방산 불포화 효소를 발현하는 균주 A의 지질 프로필에서는 C18:2 지방산이 증가하였다. 총 C18 불포화 지방산은 약 64%에서 약 67% 내지 72%로 증가하였다. 이와 유사하게, 총 C18 다중 불포화 지방산(C18:2 및 C18:3) 대 합하여진 총 C18 불포화 지방산(C18:1, C18:2 및 C18:3)의 비율은 14% 미만에서 19% 이상으로 증가하였다. 이와 같은 데이터는, 조작된 미생물의 지방산 프로필을 바꾸는데 있어서 Δ12 불포화 효소 지방산 불포화 효소의 외부 발현을 허용하는 폴리뉴클레오티드의 유용성과 효율을 입증하는 것이다.

[0825] **실시예 18: 지방산 불포화 효소의 다수 대립 유전자 중단을 통한 조작 미생물의 지방산 수준의 변경**

[0826] 본 실시예는, 형질 전환된 미생물의 지방산 프로필이 변경되도록 미생물을 조작함에 있어서, 선택가능한 마커와, 외인성 SAD 효소를 암호화하는 서열을 포함하는 형질 전환 카세트로 프로토테카 모리포르미스의 FADc 위치를 중단할 때의 형질 전환 벡터의 용도를 기술하고 있다.

[0827] (고급 오일을 생산하기 위하여) 전통적 방식에 따라서 돌연 변이된 프로토테카 모리포르미스(UTEX 1435) 파생

균주, 즉 균주 C를, 실시예 2에 상세히 기술된 유전자 총 형질 전환법에 따라서 형질 전환 구조물 pSZ1499(서열 번호 246)으로 형질 전환하였다. pSZ1499는 프로토테카 모리포르미스 UTEX 1435 내 발현에 대해 코돈 최적화된 올레아 유로파에아 스테아로일-ACP 불포화 효소 유전자의 뉴클레오티드 서열을 포함하였다. 상기 pSZ1499 발현 구조물은 씨.레인하르드티 β -튜불린 프로모터/5'UTR 및 클로렐라 불가리스 질산염 환원 효소 3'UTR의 제어 하에 있는 에스.세레비지아에 *suc2* 수크로스 인버타제 암호화 영역과 핵내 게놈에의 통합에 사용되는, FADc 게놈 영역에 대한 5'(서열 번호 247) 및 3'(서열 번호 248) 상동성 재조합 표적화 서열(구조물에 측접함)을 함유하였다. 이와 같은 에스.세레비지아에 *suc2* 발현 카세트는 서열 번호 159로서 나열하였으며, 이는 선택 마커로서 사용되었다. 올레아 유로파에아 스테아로일-ACP 불포화 효소 암호화 영역은 프로토테카 모리포르미스 Amt03 프로모터/5'UTR(서열 번호 89) 및 씨.불가리스 질산염 환원 효소 3'UTR의 제어 하에 두었으며, 천연 운반 펩티드는 클로렐라 프로토테코이데스 스테아로일-ACP 불포화 효소 운반 펩티드(서열 번호 49)로 치환하였다. 전체 오.유로파에아 SAD 발현 카세트는 pSZ1499라고 명명하였으며, FADc5'_btub-Suc2-nr_amt03-S106SAD-OeSAD-nr-FADc3'로서 표시될 수 있다.

[0828] 유일한 탄소원으로서 수크로스를 함유하는 평판 상에서 1차 형질 전환체를 선택하였다. 각각의 형질 전환체를 클론으로서 정제하여, 이를 표준 지질 생산 조건(pH 7.0, 실시예 1에 기술된 조건과 유사한 조건) 하에서 생육하였다. 실시예 11에 기술된 바와 같은 표준 지방산 메틸 에스테르 가스 크로마토그래피 불꽃 이온화(FAME GC/FID) 검출법을 사용하여 지방산 프로필을 분석하였다. 형질 전환 벡터로 형질 전환하여 생산된 대표적인 클론 세트로부터 얻어진 지방산 프로필을 표 60에 보였다. 유일한 탄소원으로서 글루코스를 포함하는 지질 생산 조건(pH 5.0) 하에서 생육된, 형질 전환되지 않은 균주 C로부터 얻은 지질의 지방산 프로필을 표 60에 추가로 제시하였다.

표 60

내인성 FAD_C대립유전자를 녹아웃하고, 오.유로파에아 스테아로일-ACP 불포화효소를 발현하도록 중복 조작된 프로토테카 모리포르미스(UTEX 1435)의 지방산 프로필

균주	형질전환체	% C16:0	% C18:0	% C18:1	% C18:2
균주 C	비형질전환	28.50	3.72	57.70	7.04
균주 C	비형질전환	28.57	3.69	57.61	7.07
균주 C pSZ1499	형질전환체 1	20.37	1.13	74.38	0.01
	형질전환체 2	19.98	1.16	74.60	0.00
	형질전환체 3	20.10	1.16	74.70	0.00
	형질전환체 4	21.13	1.21	73.86	0.00
	형질전환체 5	19.95	1.11	74.58	0.00
	형질전환체 6	20.20	1.14	74.61	0.00
	형질전환체 7	20.72	1.15	74.15	0.00
	형질전환체 8	20.06	1.11	74.44	0.00
	형질전환체 9	19.86	1.18	74.88	0.00

[0829]

[0830] 표 60에 나타낸 바와 같이, 균주 C를 pSZ1499로 형질 전환하였을 경우, 형질 전환된 미생물의 지방산 프로필은 영향을 받았다. 형질 전환되지 않은 프로토테카 모리포르미스(UTEX 1435) 균주 C의 지방산 프로필은, C18:1 지방산 60% 미만과, C18:2 지방산 7% 이상이었다. 이와는 대조적으로, pSZ1499 로 형질 전환된 균주 C는 지방산 프로필에 있어서 C18:1 지방산의 조성은 증가하였고, 이와 아울러 C18:0 및 C18:2 지방산의 조성은 감소하였다. pSZ1499로 형질 전환된 균주 C의 지방산 프로필에 있어서 C18:2 지방산은 확인되지 않았다. pSZ1499 형질 전환체 내 C18:2 지방산이 검출 가능할 정도로 확인되지 않음은 곧, 다수의 FADc 게놈 위치로의 통합에 사용되는 상동성 재조합 표적화 서열을 보유하는 pSZ1499로 형질 전환되었을 때에는 FAD 활성을 나타내지 않았다는 것을 암시하는 것이었다.

[0831] 다수의 FADc 대립유전자들이 pSZ1499 형질 전환 벡터에 의해 중단되었음을 입증하기 위하여 서던 블럿 분석법을 수행하였다. 표준 분자 생물학적 방법을 사용하여 게놈 DNA를 균주 C와 pSZ1499 형질 전환체로부터 추출하였다. 각각의 샘플로부터 얻은 DNA를 제한 효소 PstI로 절단한 후 0.8% 아가로스 겔 상에서 이동시켰다. 이 겔로부터

얻은 DNA를 나일론+ 막(애머샴(Amersham)) 상에 옮긴 다음, 이 DNA를 FADc 2' 영역과 상응하는 P32-표지화 폴리뉴클레오티드 프로브와 혼성화하였다. 도 3은, pSZ1499 형질 전환 카세트, 프로토테카 모리포르미스(UTEX 1435)의 서열 결정된 FADc 대립 유전자 2개, 그리고 크기를 예측할 수 있는 pSZ1499 형질 전환 벡터에 의해 중단된 대립 유전자들의 맵을 보여주는 것이다. FADc 대립 유전자 1은 PstI 제한 위치를 포함하였던 반면에, FADc 대립 유전자 2는 그렇지 않았다. SAD 카세트를 통합하면, PstI 제한 위치는 중단된 FADc 대립 유전자에 도입될 것이고, 그 결과, 어떤 대립 유전자가 중단되었는지 여부에 상관 없이 서던 블럿으로 약 6kb의 단편이 확인되었다. 도 4는 서던 블럿 분석 결과를 보여주는 것이다. 약 6kb에서의 혼성화 밴드는 형질 전환체 둘 다에서 확인되었다. 중단되지 않은 대립 유전자를 나타낼, 크기가 더 작은 혼성화 밴드는 확인되지 않았다. 이와 같은 결과들은 FADc 대립 유전자 둘 다는 pSZ1499에 의해 중단되었음을 보여주는 것이다.

[0832] FADc 지방산 불포화 효소의 대립 유전자 두 개를 SAD 발현 카세트로 소모시키면, 지방산 프로필에 있어서 C18:1이 약 74%였다. 종합해보면, 이러한 데이터는 조작된 미생물의 지방산 프로필을 바꾸는데 있어서, FAD 대립 유전자의 녹아웃과 스테아로일-ACP 불포화 효소의 외부 발현을 허용하는, 폴리뉴클레오티드의 유용성과 효율을 입증해주는 것이다.

[0833] **실시예 19: 조작된 미생물로부터 생산된 가공 오일의 특징들**

[0834] 프로토테카 모리포르미스(UTEX 1435)를 형질 전환 벡터 pSZ1500(서열 번호 250)으로 형질 전환하는 방법과 그 효율은 이미 PCT 특허 출원 PCT/US2011/038463 및 PCT/US2011/038463에 기술되어 있다.

[0835] (고급 오일을 생산하기 위하여) 전통적 방식에 따라서 돌연 변이된 프로토테카 모리포르미스(UTEX 1435) 파생 균주, 즉 균주 C를, 실시예 2에 상세히 기술된 유전자 총 형질 전환법에 따라서 형질 전환 구조물 pSZ1500으로 형질 전환하였다. 유일한 탄소원으로서 수크로스를 함유하는 한천 평판 상에서 1차 형질 전환체를 선택하고, 이를 클론으로 정제한 다음, 단일 조작된 계열인 균주 D를 분석용으로 선택하였다. 균주 D는 본원에 기술된 바와 같이 생육하였다. 그 다음, 표준적 방법을 사용하여, 제조된 바이오매스로부터의 오일 헥산 추출법을 수행하였는데, 이 경우, 생성된 트리글리세라이드 오일은 잔여 헥산이 존재하지 않는 것으로 확인되었다. 익스펠러 프레스를 사용하여 미세 조류로부터 오일을 추출하는 기타 다른 방법은 본원에 참조로 포함되어 있는 PCT 특허 출원 PCT/US2010/31108호에 기술되어 있다.

[0836] 이후, 널리 공지된 식물성 오일 가공법을 사용하여 균주 D의 바이오매스로부터 추출된 오일을 정제, 표백 및 탈취하였다. 이와 같은 방법들을 통하여 오일 샘플(RBD469)을 제조하였는데, 이 샘플을 대상으로, 미국 유지 화학회, 미국 재료 시험 협회 및 국제 표준화 기구에 규정된 방법에 따라서 분석 테스트 프로토콜 다수 개를 수행하였다. 이러한 분석 결과들은 하기 표 60에 요약되어 있다.

[0837] [표 60]

오일 샘플 RBD469에 대한 분석 결과

방법 번호	테스트 설명	결과	단위
AOCS Ca 3a-46	불용성 불순물	<0.01	%
AOCS Ca 5a-40	유리 지방산(올레산)	0.02	%
AOCS Ca 5a-40	산가	0.04	mg KOH/g
AOCS CA 9f-57	중성 오일	98.9	%
D97	혼탁점	-15	deg C
D97	유동점	-18	deg C
	칼 피셔 수분	0.01	%
AOCS Cc 13d-55 (수정)	염록소	<0.01	ppm
	요오드가	78.3	g I_2/100g
AOCS Cd 8b-90	과산화물가	0.31	meq/kg
ISO 6885	p-아니시딘가	0.65	
AOCS Cc 18-80	낙하 용융점(메틀러)	6.2	deg C
AOCS Cd 11d-96	트리글리세라이드	98.6	%
AOCS Cd 11d-96	모노글리세라이드	<0.01	%
AOCS Cd 11d-96	디글리세라이드	0.68	%
AOCS Cd 20-91	총 극성 화합물	2.62	%
IUPAC, 2.507 및 2.508	산화 및 중합된 트리글리세라이드	17.62	%
AOCS Cc 9b-55	인화점	244	℃
AOCS Cc 9a-48	발연점	232	℃
AOCS Cd 12b-92	산화 안정성 지수 랜시맷 (110℃)	31.6	시간
AOCS Ca 6a-40	비누화되지 않은 물질	2.28	%

[0838] *

[0839] 프로토테카 모리포르미스 균주 D RBD469 오일의 동일한 분획을 미량 원소 함량, 고체 지방 함량 및 로비본드 색상에 대해 분석하였다(AOCS 방법에 따름). 이와 같은 분석 결과들을 하기의 표 61, 표 62 및 표 63에 제시하였다.

표 61

RBD469 오일의 ICP 원소 분석

방법 번호	테스트 설명	결과(ppm)
AOCS Ca 20-99 AOCS Ca 17-01 (수정)	인	1.09
	칼슘	0.1
	마그네슘	0.04
	철	<0.02
	황	28.8
	구리	<0.05
	칼륨	<0.50
	나트륨	<0.50
	규소	0.51
	붕소	0.06
	알루미늄	<0.20
	납	<0.20
	리튬	<0.02
	니켈	<0.20
	바나듐	<0.05
	아연	<0.02
	비소	<0.20
	수은	<0.20
	카드뮴	<0.03
	크롬	<0.02
	망간	<0.05
	은	<0.05
	티타늄	<0.05
	셀레늄	<0.50
UOP779	유기 염화물	<1
UOP779	무기 염화물	7.24
AOCS Ba 4e-93	질소	6.7

[0840]

표 62

RBD469 오일의 고체 지방 함량

방법 번호	고체 지방 함량	결과
AOCS Cd 12b-93	고체 지방 함량, 10℃	0.13%
AOCS Cd 12b-93	고체 지방 함량, 15℃	0.13%
AOCS Cd 12b-93	고체 지방 함량, 20℃	0.28%
AOCS Cd 12b-93	고체 지방 함량, 25℃	0.14%
AOCS Cd 12b-93	고체 지방 함량, 30℃	0.08%
AOCS Cd 12b-93	고체 지방 함량, 35℃	0.25%

[0841]

표 63

RBD469 오일의 로비본드 색상

방법 번호	색상	결과	단위
AOCS Cc 13j-97	적색	2	유닛
AOCS Cc 13j-97	황색	27	유닛

[0842]

[0843] RBD469 오일을 대상으로 에스테르 결합 전이법을 수행하여 지방산 메틸 에스테르(FAME)를 생산하였다. 얻어진 RBD469의 지방산 메틸 에스테르 프로필을 표 64에 나타내었다.

표 64

RBD469 오일의 지방산 메틸 에스테르 프로필

지방산	면적%
C10	0.01
C12:0	0.04
C14:0	0.64
C15:0	0.08
C16:0	8.17
C16:1 이소	0.39
C16:1	0.77
C17:0	0.08
C18:0	1.93
C18:1	85.88
C18:1	0.05
C18:2	0.05
C20:0	0.3
C20:1	0.06
C20:1	0.44
C22:0	0.11
C23:0	0.03
C24:0	0.1
확인된 총 FAME	99.13

[0844]

[0845] **실시예 20: 지방산 프로필이 변경된 조작 미세 조류**

[0846] 전술한 바와 같이, 프로토테카 종 내 특이적 내인성 지질 경로 효소를 녹아웃 또는 녹다운하도록 이종 유전자를 통합시키면, 조작된 미생물의 지방산 프로필이 바뀔수 있었다. 본 실시예에서는 숙주 세포의 지질 프로필이, 내인성 지방 아실-ACP 티오에스테라제 유전자인 FATA1을 녹아웃 또는 녹다운함으로써 영향을 받을 수 있는지 여부에 대해 평가하기 위해 플라스미드 구조물을 제조하였다.

[0847] **A. 내인성 프로토테카 모리포르미스 티오에스테라제 유전자를 녹아웃함으로써 지질 프로필 변경**

[0848] (고급 오일을 생산하도록) 전통적 방식으로 돌연 변이된 프로토테카 모리포르미스 UTEX 1435 파생 균주, 균주 A에, 실시예 2의 방법을 사용하여 표 65의 플라스미드 구조물들 중 하나를 도입하여 이 균주를 형질 전환하였다. 각각의 구조물은, 씨.레인하르드티 β-튜불린 프로모터/5'UTR 및 클로렐라 불가리스 질산염 환원 효소 3'UTR의 제어 하에 있는 에스.세레비지아에 *suc2* 수크로스 인버타제 암호화 영역과 내인성 FATA1 유전자를 중단하는, 핵 내 게놈으로의 통합에 사용되는 영역을 함유하였다. 이와 같은 에스.세레비지아에 *suc2* 발현 카세트는 서열 번호 159로서 나열하였으며, 이는 선택 마커로서 사용되었다. 모든 단백질 암호화 영역은, 표 2에 따라서 프로토

테카 모리포르미스 UTEX 1435 핵내 유전자에 고유한 코돈 편향 사용을 반영하도록 최적화된 코돈이었다. 핵내 게놈 통합에 사용되는 FATA1 유전자에 대한 표적화 영역과 관련된 서열들을 이하에 나타내었다.

[0849]

설 명	서열 번호
[0850] FATA1 위치로의 통합에 사용되는 5' 서열	서열 번호 251
[0851] FATA1 위치로의 통합에 사용되는 3' 서열	서열 번호 252

표 65

[0852] 프로토테카 모리포르미스(UTEX 1435) 균주 A를 형질 전환하는데 사용된 플라스미드 구조물

플라스미드 구조물	서열 요소
1	FATA1-CrbTub_yInv_nr-FATA1
2	FATA1-CrbTub_yInv_nr::amt03_CwTE2_nr-FATA1

[0853] 구조물 FATA1-CrbTub_yInv_nr-FATA1(서열 번호 253) 내 관련 제한 위치들은 굵은 글씨체의 소문자에 밑줄을 쳐서 표시하였는데, 이 제한 위치들로서는 각각 5'→3' 방향의 *BspQ* I, *Kpn* I, *Asc* I, *Mfe* I, *Sac* I, *BspQ* I이 있다. *BspQI* 위치는 형질 전환 DNA의 5' 및 3' 말단을 한정한다. 굵은 글씨체의 소문자로 표시한 서열은, 상동성 재조합을 통하여 FATA1 위치에서 표적화 통합이 일어나는 것을 허용하는 균주 A 유래 게놈 DNA를 나타낸다. 5'→3' 방향으로 진행하였을 때, (균주 A에 수크로스를 대사하는 능력을 부여하는) 효모 수크로스 인버타제 유전자의 발현을 작동하는 씨.레인하르드티 β-튜불린 프로모터는 글자에 박스를 쳐서 표시하였다. 인버타제에 대한 개시 인자 ATG와 종결 인자 TGA는 굵은 이탤릭체의 대문자로 표시하였으며, 암호화 영역은 이탤릭체 소문자로 표시하였다. 클로렐라 불가리스 질산염 환원 효소 3'UTR은 소문자에 밑줄을 쳐서 표시하였고, 그 다음에 존재하는 굵은 글씨체의 소문자는 균주 A의 FATA1 게놈 영역을 표시하는 것이었다.

[0854]

gctcttcggagtcactgtgccactgagttcgactggtagctgaatggagtcgctgctccactaaacgaattgtcagcaccgcca
gccggccgaggacccgagtcatagcgagggtagtagcgcgccatggcaccgaccagcctgcttgccagtactggcgtctcttc
cgcttctctgtggtcctctgcgcgctccagcgcgtgcgcttttccggtggatcatgcggtccgtggcgcaccgcagcggccgctg
cccatgcagcgccgctgcttccgaacagtggcggtcagggccgcacccgcggtagccgtccgtccggaacccgcccaagagt
tttgggagcagcttgagccctgcaagatggcggaggacaagcgcatcttcctggaggagcaccggtgcgtggaggtccgggg
ctgaccggccgtcgcattcaacgtaatcaatcgcatgatgatcagaggacacgaagtcttggtggcggtggccagaaacact
gtccattgcaagggcatagggatgcgttccttcacctctcatttctcatttctgaatccctccctgctcactctttctcctcctccttc
ccgttcacgcagcattcggggtaccctttcttgcgctatgacacttccagcaaaaggtagggcgggctgcgagacggcttcccggc
gctgcatgcaacaccgatgatgcttcgaccccccgaagctccttcggggctgcatgggcgctccgatgccgctccagggcgagcgc

tgtttaaatagccaggccccccgattgcaaagacattatagcgagctaccaaagccatattcaaacacctagatcactaccacttctacac
aggccactcgagcttgtgatcgcactccgctaaggggcgcctcttcctcttcgtttcagtcacaacccgcaaacgggcgccc*ATG*
ctgctgcaggccttcctgttcctgctggccggcttcgccgccaagatcagcgcctccatgacgaacgagacgtccgaccgccccct
ggtgcacttcaccccccaacaagggctggatgaacgaccccaacggcctgtggtacgacgagaaggacgccaagtggcacctgt
acttccagtacaaccccgaacgacaccgtctgggggacgcccttgttctggggccacgccacgtccgacgacctgaccaactggg
aggaccagccccatcgccatcgccccgaagcgcaacgactccggcgccttctccggctccatggtggtggactacaacaacacct
ccggcttcttcaacgacaccatcgaccccgccgccagcgctgcgtggccatctggacctacaacaccccggagtccgaggagcagt
acatctcctacagcctggacggcggctacaccttcaccgagtaccagaagaaccccgtgctggccgccaactccacccagttcc
gcgacccgaaggtcttctggtacgagccctcccagaagtggatcatgaccgcggccaagtcccaggactacaagatcgagatct
actcctccgacgacctgaagtcctggaagctggagtccgcgttcgccaacgagggcttcctcggctaccagtacgagtgccccgg
cctgatcgaggtccccaccgagcaggaccccagcaagtcctactgggtgatgttcatctccatcaacccgggcgccccggccgg
cggctccttcaaccagtacttcgtcggcagcttcaacggcaccccacttcgaggccttcgacaaccagtccgcgtggtggacttcg
gcaaggactactacgccctgcagaccttcttcaacaccgaccccgacctacgggagcgccctgggcatcgcgtgggcctccaact
gggagtactccgccttcgtgcccaccaaccccttggcgctcctccatgtccctcgtgcgcaagttctccctcaacaccgagtaccag
gccaaccccggagacggagctgatcaacctgaaggccgagccgatcctgaacatcagcaacgccggccccctggagccggttcg
ccaccaacaccacgttgacgaaggccaacagctacaacgtcgacctgtccaacagcaccggcaccctggagttcgagctggtg
tacgccgtcaacaccaccccagacgatctccaagtccgtgttcgcggacctctccctctggttcaagggcctggaggaccccgagg
agtacctccgcatgggcttcgaggtgtccgcgtcctcctttcctggaccgcgggaacagcaaggtgaagttcgtgaaggagaa
ccccctacttcaccaaccgcatgagcgtgaacaaccagccctttcaagagcgagaacgacctgtcctactacaaggtgtacggcttg
ctggaccagaacatcctggagctgtacttcaacgacggcgacgtcgtgtccaccaacacctacttcatgaccaccgggaacgcc
*ctgggctccgtgaacatgacgacgggggtggacaacctgttctacatcgacaagttccaggtgcgcgaggtcaag**TGA**caatt*
ggcagcagctcggatagtatcgacacactctggacgctggtcgtgtgatggactgttgccgccacacttgctgccttgacctgtga
atatccctgccgcttttatcaaacagcctcagtgtgttgatcttgtgtgtacgcgcttttgcgagttgctagctgcttgtgctatttgcgaata
ccacccccagcatccccttccctcgtttcatatcgcttgcatcccaaccgcaacttatctacgctgtcctgctatccctcagcgctgctcct
gctcctgctcactgcccctcgcacagccttggtttgggctccgcctgtattctcctggtactgcaacctgtaaaccagcactgcaatgctg
atgcacgggaagtagtgggatgggaacacaaatggaggatcgta**gagctc**actagtatcgatttcgaa**gacagggtggttggctgg**
atggggaaacgctggtcgcgggattcgatcctgctgcttatatcctccctggaagcacacccacgactctgaagaagaaaacg
tgcacacacacaacccaacccggccgaatatttgcttccttatcccgggtccaagagagactgcgatgccccctcaatcagcat
cctcctccctgccgcttcaatcttccctgcttgcctgcgcccgcggtgcgccgtctgcccgcccagtcagtcactcctgcacaggc
ccettgtgcgcagtgctcctgtaccctttaccgctccttccattctgcgaggcccccctattgaatgtattcgttgcctgtgtggcca
agcgggctgctgggcgcgccgccgtcgggcagtgctcggcgactttggcggaagccgattgttcttctgtaagccacgcgcttg
ctgctttgggaagagaagggggggggtactgaatggatgaggaggagaaggagggtattggtattatctgagttgggtgaa
gagc

[0855]

[0856] 쿠페아 라이티 ACP-티오에스테라제 2(CwTE2) 유전자(수탁 번호 제U56104호)를, 균주 A의 FATA1 위치에 도입하기 위하여, 프로토테카 모리포르미스 Amt03 프로모터/5'UTR(서열 번호 89) 및 씨.불가리스 질산염 환원 효소 3'UTR의 제어 하에 있는 *CwTE2* 유전자의 단백질 암호화 영역을 발현하는 구조물을 제조하였다. 균주 A 내에서 발현된 구조물은 FATA1-CrbTub_yInv_nr::amt03_CwTE2_nr-FATA1(서열 번호 254)이라 표시하였다.

[0857] 구조물 FATA1-CrbTub_yInv_nr::amt03_CwTE2_nr-FATA1 내 관련 제한 위치들은 굵은 글씨체의 소문자에 밑줄을 쳐서 표시하였으며, 이 제한 위치들로서는 각각 5'→3' 방향의 *BspQ* I, *Kpn* I, *Asc* I, *Mfe* I, *BamH* I, *EcoR* I, *Spe* I, *Asc* I, *Pac* I, *Sac* I, *BspQ* I이 있다. *BspQI* 위치들은 형질 전환 DNA의 5' 및 3' 말단을 한정하였다. 굵은 글씨체의 소문자로 표시한 서열은 상동성 재조합을 통하여 FATA1 위치에서의 표적화 통합을 허용하는 균주 A 유래 게놈 DNA를 나타낸다. 5'→3' 방향으로 진행하였을 때, (균주 A에 수크로스를 대사하는 능력을 부여하는) 효모 수크로스 인버타제 유전자의 발현을 작동하는 씨.레인하르드티 β-튜불린 프로모터는 문자에 박스를 쳐서 표시하였다. 인버타제에 대한 개시 인자 ATG와 종결 인자 TGA는 굵은 이탤릭체의 대문자로 표시하였으며, 암호화 영역은 이탤릭체 소문자로 표시하였다. 클로렐라 불가리스 질산염 환원 효소 3'UTR은 소문자에 밑줄을 쳐서 표시하였고, 그 다음에 존재하는 프로토테카 모리포르미스의 내인성 Amt03 프로모터는 이탤릭체 문자에 박스를 쳐서 표시하였다. 씨.라이티 ACP-티오에스테라제의 개시 인자 ATG 및 종결 인자 TGA 코돈은 굵은 이탤릭체의 대문자로 표시하였으며, ACP-티오에스테라제 암호화 영역의 나머지 부분은 굵은 이탤릭체로 표시하였

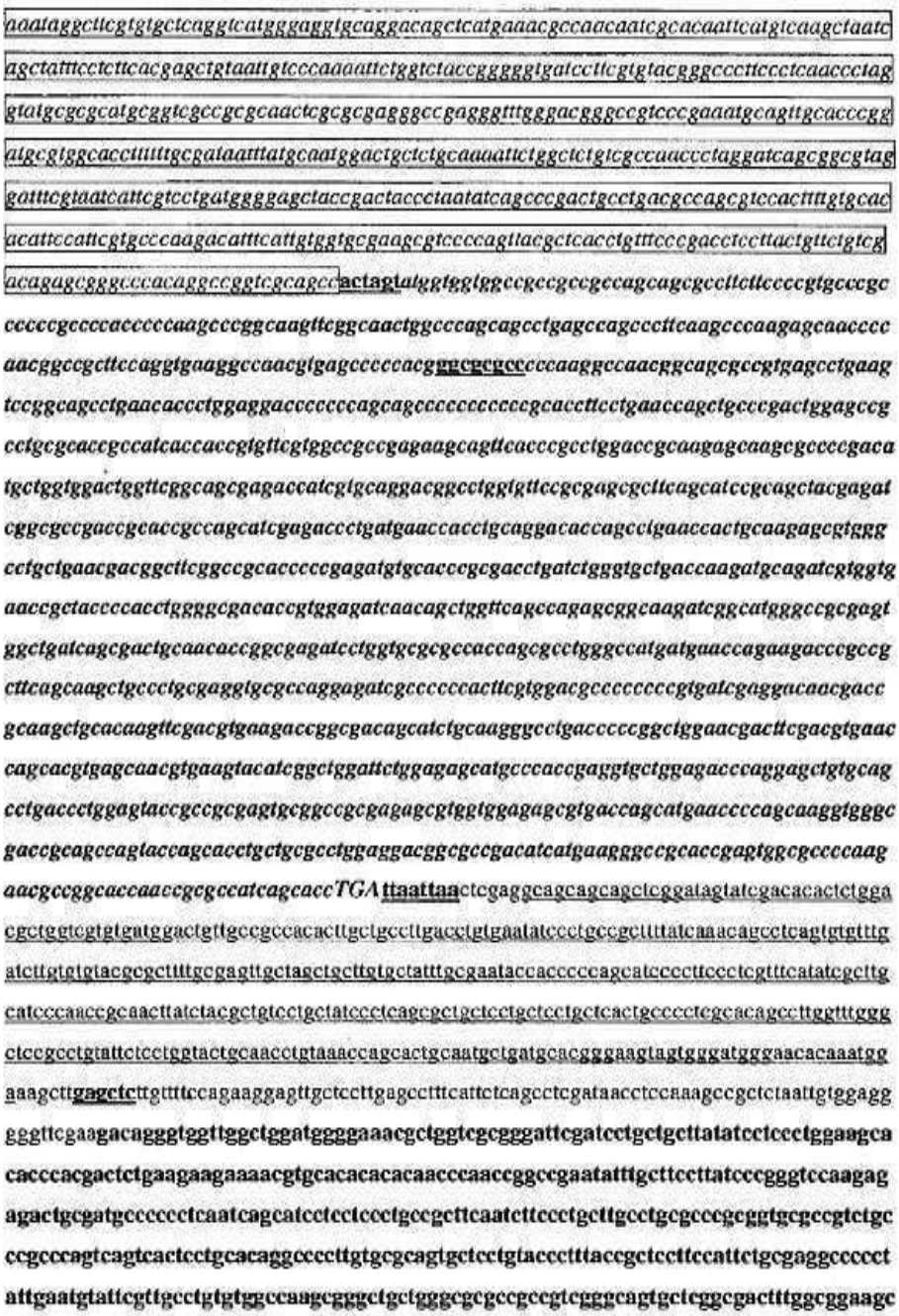
다. 씨.불가리스 질산염 환원 효소 3'UTR은 또한 소문자에 밑줄을 쳐서 표시하였으며, 그 다음에 존재하는 균주 A FATA1 게놈 영역은 굵은 글씨체의 소문자로 표시하였다.

gctcttcggagtcactgtgccactgagttcgactggtagctgaatggagtcgctgctccactaaacgaattgtcagcaccgcca
gccggccgaggacccgagtcatagcgagggtagtagcgcgccatggcaccgaccagcctgcttgccagtactggcgtctcttc
cgcttctctgtggtcctctgcgcgctccagcgcgtgcgcttttccggtggatcatgcggtccgtggcgcaccgcagcggccgctg
cccatgcagcgccgctgcttccgaacagtggcggtcagggccgcacccgcggtagccgtccgtccggaacccgcccaagagt
tttgggagcagcttgagccctgcaagatggcggaggacaagcgcatcttcctggaggagcaccggtgcgtggaggtccgggg
ctgaccggccgtcgcattcaacgtaatcaatcgcatgatgatcagaggacacgaagtcttggtggcggtggccagaaacact
gtccattgcaagggcatagggatgcgttccttcacctctcatttctcatttctgaatccctccctgctcactcttttctcctcctcttc
ccgttcacgcagcattcggggtacc<u>ctttcttgcgctatgacacttccagcaaaaggtagggcgggctgcgagacggcttcccggc</u>
<u>gctgcatgcaacaccgatgatgcttcgaccccccgaagctccttcggggctgcatgggcgctccgatgccgctccagggcgagcgc</u>
<u>tgtttaaatagccaggcccccgattgcaaagacattatagcgagctaccaaagccatattcaaacacctagatcactaccacttctacac</u>

[0858]

<u>aggccactcgagcttgtgatcgcactccgctaaggggggcgcctcttcctcttcgtttcagtcacaaccccgcaaacggcgccgccc</u>*ATG*
ctgctgcaggccttcctgttcctgctggccggcttcgccgccaagatcagcgcctccatgacgaacgagacgtccgaccgccccct
ggtgcacttcaccccccaacaagggctggatgaacgaccccaacggcctgtggtacgacgagaaggacgccaagtggcacctgt
acttccagtacaaccccgaacgacaccgtctgggggacgcccttgttctggggccacgccacgtccgacgacctgaccaactggg
aggaccagcccatcgccatcgccccgaagcgcaacgactccggcgccttctccggctccatggtggtggactacaacaacacct
ccggcttcttcaacgacaccatcgacccgcgccagcgctgcgtggccatctggacctacaacaccccggagtccgaggagcagt
acatctcctacagcctggacggcggctacaccttcaccgagtaccagaagaaccccgtgctggccgccaactccaccccagttcc
gcgacccgaaggtcttctggtacgagcccctccccagaagtggatcatgaccgcggccaagtccccaggactacaagatcgagatct
actcctccgacgacctgaagtcctggaagctggagtccgcgttcgccaacgagggcttcctcggctaccagtacgagtgccccgg
cctgatcgaggtccccaccgagcaggaccccagcaagtcctactgggtgatgttcatctccatcaacccccggcgccccggccgg
cggctccttcaaccagtacttcgtcggcagcttcaacggcaccccacttcgaggccttcgacaaccagtcccgcgtggtggacttcg
gcaaggactactacgccctgcagaccttcttcaacaccgacccgacctacgggagcgccctgggcatcgcgtgggcctccaact
gggagtacttccgcctcgtgcccaccaaccccttggcgctcctccatgtcccttcgtgcgcaagttctccctcaacaccgagtaccag
gccaacccggagacggagctgatcaacctgaaggccgagccgatcctgaacatcagcaacgccggccccctggagccggttcg
ccaccaacaccacgttgacgaaggccaacagctacaacgtcgacctgtccaacagcaccggcacccctggagttcgagctggtg
tacgccgtcaacaccacccagacgatctccaagtccgtgttcgcggacctctcccctctggttcaagggcctggaggaccccgagg
agtacctccgcatgggcttcgaggtgtccgcgtcctccttcttcctggaccgcgggaacagcaaggtgaagttcgtgaaggagaa
ccctacttcaccaaccgcatgagcgtgaacaaccagcccttcaagagcgagaacgacctgtcctactacaaggtgtacggcttg
ctggaccagaacatcctggagctgtacttcaacgacggcgacgtcgtgtccaccaacacctacttcatgaccaccgggaacgcc
*ctgggctccgtgaacatgacgacggggtggacaacctgttctacatcgacaagttccaggtgcgcgaggtcaag***TGA**caatt**
ggcagcagcagctcggatagtatcgacacactctggacgctggtcgtgtgatggactgttgccgccacacttgctgccttgacctgtga
atatccctgccgcttttatcaaacagcctcagtgtgtttgatcttgtgtgtacgcgcttttgcgagttgctagctgcttgtgctatttgcgaata
ccacccccagcatccccttccctcgtttcatatcgcttgcatcccaaccgcaacttatctacgctgtcctgctatcccagcgctgctcct
gctcctgctcactgcccctcgcacagccttggtttgggctccgcctgtattctcctggtactgcaacctgtaaaccagcactgcaatgctg
atgcacgggaagtagtgggatgggaacacaaatggaggatccgcgtctcgaacagagcgcgcagaggaacgctgaaggtctcg
cctctgtcgcacctcagcgcggcatacaccacaataaccacctgacgaatgcgcttggttcttcgtccattagcgaagcgtccggttca
cacacgtgccacgttggcgaggtggcaggtgacaatgatcggtggagctgatggtcgaaacgttcacagcctagggatatc**gaattc**
ggccgacaggacgcgcgtcaaaggtgctggtcgtgtatgccctggccggcaggtcgttgctgctgctggttagtgattccgcaaacc
ctgattttggcgtcttatttggcgtggcaaacgctggcgcccgcgagccgggccggcggcgatgcggtgccccacggctgccgg
aatccaagggaggcaagagcgcccgggtcagttgaagggctttacgcgcaaggtacagccgctcctgcaaggctgcgtggtgg
aattggacgtgcaggtcctgctgaagttcctccaccgcctcaccagcggacaaagcaccggtgtatcaggtccgtgtcatccactc
taaagagctcgactacgacctactgatggccctagattcttcatcaaaaacgcctgagacacttgcccaggattgaaactccctga
agggaccaccaggggccctgagttgttccttccccccgtggcgagctgccagccaggctgtacctgtgatcgaggctggcgggga

[0859]



[0860]

cgattgttcttctgtaagccacgcgcttgctgctttgggaagagaaggggggggtactgaatggatgaggaggagaaggag

[0861]

gggtattggtattatctgagttgggtgaagagc

[0862] 플라스미드 구조물 1 또는 2를 개별적으로 균주 A에 도입하여 이 균주를 형질 전환하고 나서, 유일한 탄소원으로서 수크로스를 포함하는 한천 평판 상에서 양성 클론들을 선택하였다. 상기 실시예들에서와 같이, 1차 형질 전환체를 클론으로서 정제한 다음, 표준적 지질 생산 조건(pH 7) 하에서 생육하고 나서, 각각의 형질 전환체 유래 무수 바이오매스로 지질 샘플들을 제조하였다. 실시예 11에 기술된 바와 같은 직접 에스테르 결합 전이법을 사용하여 지방산 프로필을 측정하였다. 구조물 1로 형질 전환하여 생성된 대표적인 클론 세트들로부터 얻어진 지방산 프로필(총 지방산의 면적%로 표시)을, 형질 전환되지 않은 균주 A 대조군의 지방산 프로필과 비교한 결과를 표 66에 제시하였다. 구조물 2로 형질 전환하여 생성된 대표적인 클론 세트들로부터 얻어진 지방산 프로필(총 지방산의 면적%로 표시)을, 형질 전환되지 않은 균주 A 대조군의 지방산 프로필과 비교한 결과를 표 67에 제시하였다.

표 66

내인성 FATA1 대립 유전자를 중단하는 선택가능한 마커를 포함하는 프로토테카 모리포르미스 세포의 지방산 프로필

형질 전환	% C14:0	% C16:0	% C18:0	% C18:1	% C18:2
야생형	1.23	25.68	2.83	60.54	7.52
형질전환체 1	0.86	16.95	1.75	68.44	9.78
형질전환체 2	0.85	17.33	1.71	68.57	9.31
형질전환체 3	0.82	17.40	1.78	68.55	9.22
형질전환체 4	0.84	17.43	1.78	68.25	9.53
형질전환체 5	0.75	17.64	2.02	69.02	8.61

[0863]

[0864] 표 66에 제시된 결과들은, 숙주의 내인성 FATA1 대립 유전자가 소모될 경우, 조작된 미세 조류의 지방산 프로필이 바뀐다는 것을 보여주는 것이다. 선택가능한 마커를 내인성 FATA1 대립 유전자에 표적화하였을 때, 형질 전환된 미생물의 지방산 프로필에 상기 표적화가 미치는 영향력으로 인하여 C16:0 지방산은 명확히 감소하였으며, 이와 아울러 C18:1 지방산은 증가하였다.

표 67

내인성 FATA1 대립 유전자를 중단하는 외인성 티오에스테라제 및 선택가능한 마커를 함유하는 프로토테카 모리포르미스 세포의 지방산 프로필

	형질전환체	탄소 공급원	% C10:0	% C12:0	% C14:0	% C16:0	% C18:0	% C18:1	% C18:2
균주 A	야생형	글루코스	0.01	0.04	1.38	28.83	3.00	56.05	8.21
	야생형	글루코스	0.01	0.04	1.50	29.38	3.00	55.29	8.23
	야생형	글루코스/프럭토스	0.01	0.05	1.48	28.58	3.20	57.14	7.27
	야생형	글루코스/프럭토스	0.01	0.04	1.54	29.05	3.23	56.47	7.32
>2 복사체	1	글루코스/프럭토스	4.29	19.98	9.17	20.68	3.47	34.38	6.37
	2	글루코스/프럭토스	3.11	16.17	9.91	15.97	1.57	45.72	5.81
	3	수크로스	4.84	24.22	11.56	19.48	2.67	29.56	6.02
	4	수크로스	3.24	16.67	10.39	16.34	1.43	44.41	6.00
1-2 복사체	1	글루코스/프럭토스	0.18	1.64	1.85	14.43	2.12	70.30	7.63
	2	글루코스/프럭토스	0.18	1.56	1.74	13.56	2.25	71.04	7.72
	3	수크로스	0.19	1.69	1.89	13.79	3.15	69.97	7.68
	4	수크로스	0.15	1.26	1.49	13.44	2.73	71.46	7.77

[0865]

[0866] 숙주 FATA1 대립 유전자에 선택가능한 마커만을 표적화하는 것과 아울러, 외인성 티오에스테라제와 함께 선택가능한 마커를 통합하였더니, 조작된 미세 조류의 지방산 프로필이 바뀌었다. 상기 표 67에 보인 바와 같이, FATA1 대립 유전자를 중단하는 외인성 티오에스테라제 유전자를 표적화한 결과, C16:0 지방산 생산이 명확히 감소하였다. FATA1 위치에서 *CwTE2* 티오에스테라제를 발현하면 또한, 분석된 형질 전환체 내에서 나타나는 외인성 티오에스테라제 활성의 수준에 의존적으로 중간쇄 지방산과 C18:1 지방산 생산에 영향을 미쳤다. 표적 통합 위치에서 증폭된 이식 유전자의 복사체 수와 티오에스테르 수준은 서로 조화가 매우 잘 되었다(이는 지방산 프로필에 대한 영향력이나 재조합 단백질 축적 여부(웨스턴 블럿팅에 의해 평가됨)에 의해 규명됨).

[0867] *Cw*TE2 유전자를 증폭시킨 유전자 이식 계열에서는 C10:0 내지 C14:0 지방산이 눈에 띄게 증가하였고, 이와 아울러 C18:1 지방산은 감소하였다. 이와는 대조적으로, *Cw*TE2를 약간 증폭시켰거나 아예 증폭시키지 않은 형질 전환체들은 외인성 티오에스테라제를 더욱 낮은 수준으로 발현하였으며, 그 결과, 중간쇄 지방산은 약간 증가하였고, C18:1 지방산의 증가에는 더욱 큰 영향을 미쳤다.

[0868] 종합하여 보면, 상기 데이터는 숙주의 내인성 FATA1 대립 유전자가 표적 중단되면 조작된 미세 조류의 지질 프

로필이 바뀐다는 것을 보여준다. 이 데이터는, FATA 대립 유전자를 표적 중단시켜, 조작된 미생물 세포의 지방산 프로필을 바꿀수 있는(특히 C16 지방산 농도를 감소시키고, C18:1 지방산 농도를 증가시킬 수 있는) 폴리뉴클레오티드의 유용성과 효율을 입증하였다. 상기 데이터는 또한, FATA 대립 유전자의 표적 중단을 허용함과 동시에, 조작된 미생물 세포의 지방산 프로필을 바꾸는(특히 C16 지방산의 농도를 감소시키는) 외인성 티오에스테라제를 발현하는, 폴리뉴클레오티드의 유용성과 효율도 입증하였다.

[0869] **B. 내인성 프로토테카 모리포르미스 티오에스테라제 유전자의 녹다운에 의한 지질 프로필의 변화**

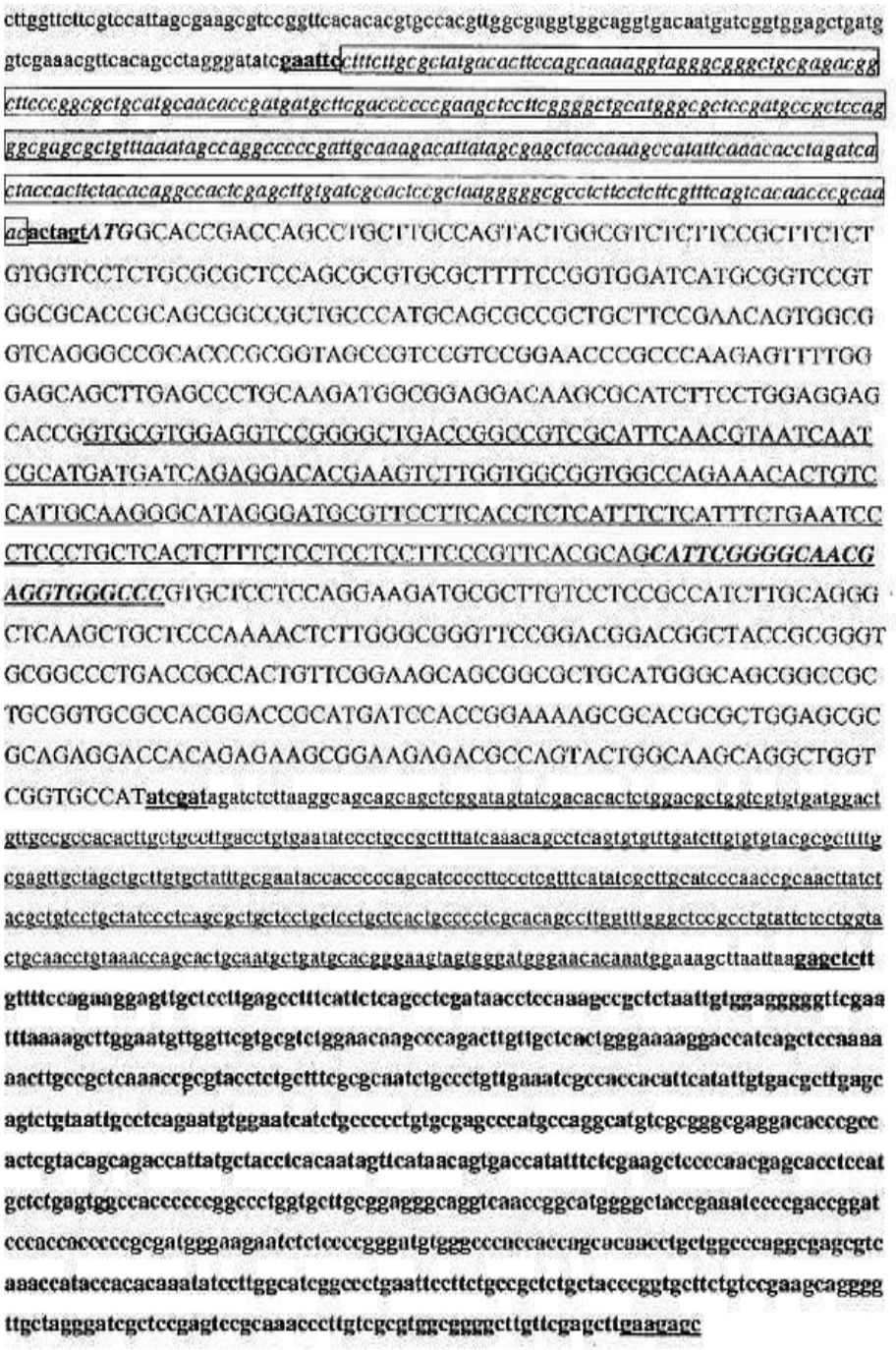
[0870] RNAi에 의해 프로토테카 모리포르미스 FATA1 유전자 발현을 하향 조절하는 구조물을, 프로토테카 모리포르미스 UTEX 1435 균주 A 유전적 배경에 도입하였다. 사카로마이세스 세레비지아에 *suc2* 수크로스 인버타제 유전자를, 유일한 탄소원으로서 수크로스를 바탕으로 생육할 수 있는 능력을 부여하는 선택가능한 마커로서 이용하였다. 구조물은 FatA1 암호화 영역의 제1 엑손 → 내인성 인트론 → 제1 엑손의 역방향 반복 단위를 이용하였다. 6S 게놈 영역에 대한 5' 및 3' 상동성 재조합 표적화 서열(구조물에 측접)(각각 서열 번호 82 및 84로 나열됨)을 헤어핀 구조물의 통합을 위해 핵내 게놈에 포함시켰다. 이 구조물을 6S::β-Tub:suc2:nr:: β-tub:hairpinFatA:nr::6S로 명명하였다.

[0871] 6S::β-Tub:suc2:nr:: β-tub:hairpin FatA:nr::6S 내 관련 제한 위치들은 굵은 글씨체의 소문자에 밑줄을 쳐서 표시하였으며, 이 제한 위치로서는 각각 5'→3' 방향의 *BspQ* I, *Kpn* I, *Mfe* I, *BamH* I, *EcoR* I, *Spe* I, *Xho* I, *Sac* I, *BspQ* I이 있다. *BspQI* 위치는 형질 전환 DNA의 5' 말단 및 3' 말단을 한정한다. 굵은 글씨체의 소문자로 표시한 서열은 상동성 재조합을 통하여 6S 위치에서의 표적 통합을 허용하는 균주 A 유래 게놈 DNA를 나타낸다. 5'→3' 방향으로 진행되는, 효모 수크로스 인버타제 유전자(균주 A가 수크로스를 대사할 수 있는 능력 부여)의 발현을 작동하는 씨.레인하르드티 β-튜불린 프로모터는 글자에 박스를 쳐서 표시하였다. 인버타제에 대한 개시 인자 ATG 및 종결 인자 TGA는 굵은 이탤릭체의 대문자로 표시하였으며, 암호화 영역은 이탤릭체의 소문자로 표시하였다. 클로렐라 불가리스 질산염 환원 효소 3'UTR은 소문자에 밑줄을 쳐서 표시하였으며, 그 다음에 존재하는 제2 씨.레인하르드티 β-튜불린 프로모터(헤어핀 FatA1의 발현을 작동함)는 이탤릭체에 박스를 쳐서 표시하였다. FatA1의 개시 인자 ATG 코돈은 굵은 이탤릭체의 대문자로 표시하였으며, FatA1 암호화 영역의 제1 엑손의 나머지 부분은 대문자로 표시하였다. FatA 유전자의 인트론은 대문자에 밑줄을 쳐서 표시하였으며, 굵은 이탤릭체 대문자에 밑줄을 쳐서 표시한 링커 영역은 FatA1 인트론/역행 제1 엑손 접합부에 삽입하여, 이 벡터 내에서의 RNA 스플라이싱을 보조하도록 만들었다. FatA1의 역행 제1 엑손은 대문자로 표시하였다. 씨.불가리스 질산염 환원 효소 3'UTR도 역시 소문자에 밑줄을 쳐서 표시하였으며, 그 다음에 존재하는 균주 A의 6S 게놈 영역은 굵은 글씨체의 소문자로 표시하였다. 이 RNAi 구조물의 FATA 부분들의 서열은 서열 번호 255로서 나열하였다.

[0872] **gctcttcgccgccgccactcctgctcgagcgcgcccgcgcgtgcgccgccagcgccttggccttttcgccgcgctcgtgcgcgtc**
gctgatgtccatcaccaggtccatgaggtctgccttgcgccggctgagccactgcttcgtccgggcggccaagaggagcatga
gggaggactcctggtccagggtcctgacgtggtcgcggctctgggagcgggccagcatcatctggctctgccgcaccgaggc
cgcctccaactggtcctccagcagccgcagtcgccgccgaccctggcagaggaagacaggtgaggggggtatgaattgtaca

gaacaaccacgagccttgtctaggcagaatccctaccagtcatggctttacctggatgacggcctgcgaacagctgtccagcg
accctcgctgccgccgcttctcccgcacgcttctttccagcaccgtgatggcgcgagccagcgccgcacgctggcgctgcgctt
cgccgatctgaggacagtcggggaactctgatcagtctaaaccccctttgcgcgttagtgttgccatcctttgcagaccggtgag
agccgacttgttgtgcgccacccccccacaccacctcctcccagaccaattctgtcacctttttggcgaaggcatcggcctcggcc
tgcagagaggacagcagtgcccagccgctgggggttggcggatgcacgctcaggtaccgcttcttgcgctatgacacttccagca
aaaggtagggcgggctgcgagacggcttcccggcgctgcatgcaacaccgatgatgcttcgaccccccgaagctccttcggggctg
catgggcgctccgatgccgctccagggcgagcgctgtttaaatagccaggcccccgattgcaaagacattatagcgagctaccaaag
ccatattcaaacacctagatcactaccacttctacacaggccactcgagcttgtgatcgcactccgctaaggggcgcctcttcctcttc
gtttcagtcacaacccgcaaactctagaatatcaATGctgctgcaggccttcctgttcctgctggccggcttcgccgccaagatcag
cgcctccatgacgaacgagacgtccgaccgccccctggtgcacttcaccccctaacaagggctggatgaacgaccccaacggcc
tgtggtacgacgagaaggacgccaagtggcacctgtacttccagtacaacccgaacgacaccgtctgggggacgccctttgttctg
ggccacgccacgtccgacgacctgaccaactgggaggaccagcccatcgccatcgccccgaagcgcaacgactccggcgc
cttctccggctccatggtggtggactacaacaacacctccggcttcttcaacgacaccatcgaccccgcgccagcgctgcgtggcca
tctggacctacaacaccccggagtccgaggagcagtacatctcctacagcctggacggcggctacaccttcaccgagtaccaga
agaaccccgtgctggccgccaactccaccgagttccgcgaccccgaaggtcttctggtacgagccctcccagaagtggatcatgac
cgcggccaagtcccaggactacaagatcgagatctacctccgacgacctgaagtcctggaagctggagtccgcgttcgccaa
cgagggcttcctcggctaccagtacgagtgccccggcctgatcgaggtccccaccgagcaggaccccagcaagtcctactgggt
gatgttcatctccatcaaccccggcgccccggccggcggctccttcaaccagtacttcgtcggcagcttcaacggcaccccacttcg
aggccttcgacaaccagtccgcgtggtggacttcggcaaggactactacgcctgcagaccttcttcaacaccgaccccgaccta
cgggagcgccctgggcatcgcgtgggcctccaactgggagtactccgccttcgtgcccaccaacccctggcgctcctccatgtccc
tcgtgcgcaagttctcctcaacaccgagtaccaggccaaccccggagacggagctgatcaacctgaaggccgagccgatcctg
aacatcagcaacgccggccccctggagccggttcgccaccaacaccacgttgacgaaggccaacagctacaacgtcgacctgtc
caacagcaccggcaccctggagttcgagctggtgtacgccgtcaacaccacccagacgatctccaagtccgtgttcgcggacctc
tccctctggttcaagggcctggaggacccgaggaglacctccgcatgggcttcgaggtgtccgcgtcctccttcttcctggaccgc
gggaacagcaaggtgaagttcgtgaaggagaaccccctacttcaccaaccgcatgagcgtgaacaaccagccccttcaagagcg
agaacgacctgtcctactacaaggtgtacggcttgctggaccagaacatcctggagctgtacttcaacgacggcgacgtcgtgtcc
accaacacctacttcatgaccaccgggaacgccctgggctccgtgaacatgacgacgggggtggacaacctgttctacatcgac
aagttccaggtgcgcgaggtcaagTGAcaattgcgcagcagcagctcggatagtatcgacacactctggacgctggtcgtgtgat
ggactgttgccgccacacttgctgccttgacctgtgaatatccctgccgcttttatcaaacagcctcagtgtgtttgatcttgtgtgtacgcg
cttttgcgagttgctagctgcttgtgctatttgcgaataccacccccagcatccccttccctcgtttcatatcgcttgcatcccaaccgcaac
ttatctacgctgtcctgctatccctcagcgctgctcctgctcctgctcactgcccctcgcacagccttggtttgggctccgcctgtattctcc
tggtactgcaacctgtaaaccagcactgcaatgctgatgcacgggaagtagtgggatgggaacacaaatggaggatccgcgtctcg
aacagagcgcgcagaggaacgctgaaggtctcgcctctgtcgcacctcagcgcggcatacaccacaataaccacctgacgaatgcg

[0873]



[0874]

[0875] 6S::β-Tub:suc2:nr:: β-tub:hairpin FatA:nr::6S를 발현하였을 경우, 헤어핀 RNA가 형성되어 표적 FatA 유전자 생성물이 침묵되었다. 구조물 6S::β-Tub:suc2:nr:: β-tub:hairpin FatA:nr::6S를 균주 A에 도입하여 이 균주를 형질 전환하고 나서, 유일한 탄소원으로서 수크로스를 포함하는 한천 평판 상에서 양성 클론들을 선택하였다. 1차 형질 전환체를 클론으로서 정제한 다음, 표준적 지질 생산 조건(pH 5.0) 하에서 생육하고 나서, 각각의 형질 전환체 유래 무수 바이오매스로 지질 샘플들을 제조하였다. 실시예 11에 기술된 바와 같은 직접 에스테르 결합 전이법을 사용하여 지방산 프로필을 측정하였다. 형질 전환하여 생성된 대표적인 클론 세트들로부터 얻어진 지방산 프로필(총 지방산의 면적%로 표시)을, 형질 전환되지 않은 균주 A 대조군의 지방산 프로필과 비교한 결과를 표 68에 제시하였다.

표 68

FATA의 발현을 하향 조절하는 RNA 헤어핀 구조물을 함유하는 프로토테카 모리포르미스 세포의 지방산 프로필

형질전환체	% C10:0	% C12:0	% C14:0	% C16:0	% C16:1	% C18:0	% C18:1	% C18:2
비형질전환	0.01	0.03	1.23	25.68	0.96	2.83	60.54	7.52
형질전환체 1	0.01	0.03	0.71	15.10	1.05	1.67	72.08	8.27
형질전환체 2	0.01	0.03	0.81	15.66	1.16	1.56	70.03	9.61
형질전환체 3	0.01	0.03	1.09	22.67	1.05	2.12	63.18	8.66
형질전환체 4	0.01	0.04	1.14	23.31	1.01	2.23	62.83	8.26

[0876]

[0877] 표 68에 제시된 데이터는, FATA 헤어핀 RNA 구조물의 발현이 숙주 유기체의 C16 및 C18:1 지방산 프로필에 미치는 명확한 영향력을 보여주는 것이다. FATA 헤어핀 RNA 구조물을 포함하는 균주 A 형질 전환체의 지방산 프로필은, C18:1 지방산 백분율이 증가하였음과 아울러, C16 지방산은 감소하였음이 확인되었다. 이러한 데이터는, 조작된 숙주 미생물의 지방산 프로필을 바꾸는데 있어서(특히 미생물 세포 내 C18:1 지방산 농도를 증가시키고, C16 지방산의 농도를 감소시키는데 있어서) 프로토테카 모리포르미스 내 폴리뉴클레오티드 FATA RNA 헤어핀 구조물의 성공적인 발현과 용도를 입증하는 것이다.

[0878] **실시예 21: 클로렐라 소로키니아나의 조작**

[0879] 본 발명에 의하여, 본원에 기술된 바와 같이 Dawson(도슨) et al.에 의해 교시된 방법과 벡터를 변형함으로써 클로렐라 소로키니아나 내에서 재조합 유전자를 발현시킬 수 있었다. 간단히 말해서, 문헌[Dawson *et al.*, *Current Microbiology* Vol. 35 (1997) pp. 356-362]에는 플라스미드 DNA로 클로렐라 소로키니아나를 핵내에서 안정적으로 형질 전환하는 것에 대해 개시되어 있다. 도슨은 미세 입자 투사법(microprojectile bombardment)과 같은 형질 전환법을 사용하여, 전 클로렐라 불가리스 질산염 환원 효소 유전자(NR, GenBank 수탁 번호 제U39931호)를 암호화하는 플라스미드 pSV72-NRg를 돌연 변이 클로렐라 소로키니아나(NR 돌연 변이체)에 도입하였다. 상기 NR 돌연 변이체는 질소원으로서 질산염을 사용하지 않으면 생육할 수 없었다. 질산염 환원 효소는 질산염이 아질산염으로 전환되는 것을 촉진한다. 형질 전환하기 전, 클로렐라 소로키니아나 NR 돌연 변이체는 유일한 질소원으로서 질산염(NO_3^-)을 포함하는 배양 배지상에서 마이크로콜로니 단계 이상으로는 생육하지 못하였다. NR 돌연 변이체 클로렐라 소로키니아나 내 클로렐라 불가리스 NR 유전자 생성물의 발현은 질산염 대사 결핍을 복구하는 선택가능한 마커로서 사용되었다. pSV72-NRg 플라스미드로 형질 전환하여, 클로렐라 불가리스 NR 유전자 생성물을 안정적으로 발현하는 NR 돌연 변이 클로렐라 소로키니아나(유일한 탄소원으로서 질산염을 포함하는 한천 평판 상에서 마이크로콜로니 단계 이상으로 생육할 수 있었음)를 제조하였다. 안정적인 형질 전환체의 DNA 평가는 서던 블럿 분석법에 의해 수행되었으며, 안정적인 형질 전환체의 RNA 평가는 RNase 보호법(RNase protection)에 의해 수행되었다. 형질 전환된 클로렐라 소로키니아나(NR 돌연 변이체)의 선택과 유지는, 0.2g/L $MgSO_4$, 0.67g/L KH_2PO_4, 3.5g/L K_2HPO_4, 1.0g/L $Na_3C_6H_5O_7 \cdot H_2O$ 및 16.0g/L 한천, 적당한 질소원(예를 들어, NO_3^-), 미세 영양소 및 탄소원을 포함하는 한천 평판(pH 7.4) 상에서 수행되었다. 도슨은 또한 클로렐라 소로키니아나 및 클로렐라 소로키니아나 NR 돌연 변이체의 액체 배양 배지 중 번식에 대해서도 개시하였다. 도슨은 클로렐라 불가리스 질산염 환원 효소 유전자의 3' UTR/종결 인자와 프로모터, 그리고 플라스미드 pSV72-NRg는 클로렐라 소로키니아나 NR 돌연 변이체 중 이종 유전자를 발현하는데 적당하였다고 개시하였다. 도슨은 또한 클로렐라 불가리스 질산염 환원 효소 유전자 생성물의 발현은 클로렐라 소로키니아나 NR 돌연 변이체 내에서 선택가능한 마커로서 사용하기 적당하였다고 개시하였다.

[0880] 본 발명의 하나의 구체예에서는, 선택가능한 마커로서 사용될 클로렐라 불가리스 질산염 환원 효소(*Cv*NR) 유전자 생성물을 암호화하는 뉴클레오티드 서열을 포함하는 벡터 pSV72-NRg를 구성하고, 지질 생합성 경로 발현 카세트 서열을 추가로 포함하도록 변형하여, 형질 전환 벡터를 제조하였다. 지질 생합성 경로 발현 카세트는 표 70으로부터 선택된 하나 이상의 지질 생합성 경로 단백질을 암호화하였는데, 여기서, 각각의 단백질 암호화 서

열은 표 69에 따른 클로렐라 소로키니아나의 핵내 유전자에 고유한 코돈 편향 사용을 반영하도록 클로렐라 소로키니아나 내 발현에 대하여 최적화된 코돈이었다. 표 70의 지질 생합성 경로 단백질 각각에 있어서, 코돈 최적화 유전자 서열은 각각 단백질 암호화 서열 상류의 *Cv*NR 프로모터에 작동 가능하도록 결합될 수 있었고, 단백질 암호화 서열의 하류 또는 3' 영역에서 *Cv*NR 3'UTR/종결 인자에 작동 가능하도록 결합될 수 있었다. 형질 전환 구조물은 추가로는 형질 전환 벡터의 표적화 게놈 통합에 사용되는 클로렐라 소로키니아나 게놈에 대한 상동성 영역을 포함할 수 있었다. 상동성 영역은 내인성 지질 생합성 경로 유전자의 하나 이상의 게놈 위치를 중단하는 것으로 선택할 수 있었다. 클로렐라 소로키니아나는, 널리 공지된 형질 전환 기술, 예를 들어 미세 입자 투사법 또는 기타 다른 공지된 방법을 통해 형질 전환 벡터를 사용하여 안정적으로 형질 전환되었다. *Cv*NR 유전자 생성물의 활성은, 예를 들어 클로렐라 소로키니아나 NR 돌연 변이 균주의 질소 동화 결핍을 복구하고, 형질 전환 벡터를 안정적으로 발현하는 클로렐라 소로키니아나 NR 돌연 변이체를 선택하는 선택가능한 마커로서 사용될 수 있었다. 클로렐라 소로키니아나 지질 생산에 적당한 생육 배지는 0.5g/L KH_2PO_4, 0.5g/L K_2HPO_4, 0.25g/L $MgSO_4$-7H2O와, 보충 미량 영양소, 그리고 적당한 질소원과 탄소원을 포함하였으나, 이에 제한되는 것은 아니다(Patterson, *Lipids* Vol.5:7 (1970), pp.597-600). 클로렐라 소로키니아나 지질의 지방산 프로필은 본원에 기술된 분석 방법과 표준적 지질 추출법을 통하여 평가될 수 있었다.

[0881] **실시예 22 내지 실시예 44: 도입 및 표**

[0882] 이하 실시예 22 내지 실시예 44는 본 발명에 따라서 다양한 미생물을 조작하는 것에 관해 기술하고 있다. 미생물의 지방산 프로필을 변경하기 위하여, 미생물을 유전자 변형할 수 있었는데, 이때, 내인성 또는 외인성 지질 생합성 경로 효소는 발현 또는 과발현되었거나, 발현이 약화되었다. 미생물 지질 자체의 지방산 프로필을 지방산 불포화 정도에 대해서 바꾸고, 지방산 사슬 길이를 감소 또는 증가시키기 위하여 미생물을 유전자 조작하는 단계는, 형질 전환 벡터(예를 들어, 플라스미드)의 설계 및 구성, 하나 이상의 벡터를 사용하는 미생물의 형질 전환, 형질 전환된 미생물(형질 전환체)의 선택, 형질 전환된 미생물의 생육, 그리고 조작된 미생물에 의해 생산된 지질의 지방산 프로필 분석을 포함한다.

[0883] 숙주 유기체의 지방산 프로필을 바꾸는 이식 유전자는 다수의 진핵 미생물 내에서 발현될 수 있었다. 진핵 미생물, 예를 들어 클라미도모나스 레인하르드티, 클로렐라 엘립소이데아, 클로렐라 사카로필라, 클로렐라 불가리스, 클로렐라 케슬러리, 클로렐라 소로키니아나, 해마토코커스 플루비알리스, 고니움 펙토랄레, 볼복스 칼테리, 두날리엘라 터티올렉타, 두날리엘라 비리디스, 두날리엘라 살리나, 클로스테리움 페라세로섬-스트리고섬-리토랄레 복합체, 나노클로롭시스 종, 탈라시오시라 슈도나나, 파에오닥틸룸 트리코르누툼, 나비큘라 사프로필라, 실린드로테카 후시포르미스, 시클로텔라 크립티카, 심비오디늄 마이크로아드리악티큠, 암피디늄 종, 카에토세로스 종, 모르티에렐라 알피나, 그리고 야로위아 리폴리티카 내에서 이식 유전자를 발현하는 예는 과학 문헌에서 살펴볼 수 있다. 이와 같은 발현 기술들은 본 발명에 교시된 기술들 즉, 지방산 프로필에 변화가 있는 조작 미생물을 생산하기 위한 기술들과 통합될 수 있었다.

[0884] 숙주 유기체의 지방산 프로필에 변화를 일으키는 이식 유전자는 또한 다수의 원핵 미생물 내에서 발현될 수도 있었다. 유지성 미생물, 예를 들어 로도코커스 오파쿠스 내에서 이식 유전자를 발현하는 것에 관한 예도 문헌에서 살펴볼 수 있다. 이와 같은 발현 기술들은 본 발명에 교시되어 있는 기술들 즉, 지방산 프로필에 변화가 있는 조작 미생물을 생산하기 위한 기술들과 통합될 수 있었다.

표 69

코돈 선호 목록

아미노산	코돈	클로렐라 소로키니아나	클로렐라 불가리스	클로렐라 엘립소이데아	클로렐라 제슬러리	두날리엘라 터티올렉타	볼복스 카르테리	헤미도코커스 플루비알리스
Ala	GCG	0.20	0.25	0.15	0.14	0.09	0.25	0.21
Ala	GCA	0.05	0.24	0.32	0.10	0.17	0.13	0.27
Ala	GCT	0.12	0.16	0.26	0.18	0.31	0.26	0.17
Ala	GCC	0.63	0.35	0.27	0.58	0.43	0.36	0.35
Arg	AGG	0.03	0.09	0.10	0.09	0.26	0.08	0.14
Arg	AGA	0.04	0.05	0.14	0.01	0.09	0.03	0.05
Arg	CGG	0.06	0.19	0.09	0.06	0.06	0.17	0.15
Arg	CGA	0.00	0.10	0.08	0.00	0.08	0.08	0.10
Arg	CGT	0.06	0.09	0.37	0.14	0.12	0.22	0.13
Arg	CGC	0.81	0.48	0.22	0.71	0.40	0.43	0.42
Asn	AAT	0.04	0.16	0.43	0.06	0.27	0.23	0.21
Asn	AAC	0.96	0.84	0.57	0.94	0.73	0.77	0.79
Asp	GAT	0.13	0.25	0.47	0.12	0.40	0.35	0.27
Asp	GAC	0.87	0.75	0.53	0.88	0.60	0.65	0.73
Cys	TGT	0.06	0.13	0.43	0.09	0.20	0.17	0.27
Cys	TGC	0.94	0.87	0.57	0.91	0.80	0.83	0.64
End	TGA	0.00	0.72	0.14	0.14	0.36	0.24	0.70
End	TAG	0.33	0.11	0.29	0.00	0.00	0.18	0.22
End	TAA	0.67	0.17	4.00	0.86	0.64	0.59	0.09
Gln	CAG	0.42	0.40	0.15	0.40	0.27	0.29	0.33
Gln	CAA	0.04	0.04	0.21	0.40	0.27	0.07	0.10
Glu	GAG	0.53	0.50	0.33	0.40	0.27	0.53	0.49
Glu	GAA	0.02	0.06	0.31	0.40	0.27	0.11	0.07
Gly	GGG	0.04	0.16	0.19	0.08	0.10	0.12	0.22
Gly	GGA	0.02	0.11	0.13	0.07	0.13	0.12	0.11
Gly	GGT	0.03	0.12	0.39	0.24	0.25	0.23	0.15
Gly	GGC	0.91	0.61	0.29	0.96	0.51	0.53	0.52
His	CAT	0.14	0.16	0.30	0.08	0.25	0.35	0.27

[0885]

His	CAC	0.86	0.84	0.70	0.93	0.75	0.65	0.73
Ile	ATA	0.00	0.04	0.07	0.01	0.04	0.08	0.09
Ile	ATT	0.15	0.30	0.63	0.29	0.31	0.35	0.29
Ile	ATC	0.85	0.65	0.65	0.69	0.65	0.57	0.62
Leu	TTG	0.03	0.07	0.03	0.05	0.14	0.14	0.16
Leu	TTA	0.00	0.01	0.32	0.00	0.02	0.03	0.02
Leu	CTG	0.72	0.61	0.34	0.61	0.60	0.45	0.53
Leu	CTA	0.01	0.03	0.03	0.04	0.04	0.07	0.07
Leu	CTT	0.04	0.08	0.16	0.06	0.06	0.14	0.09
Leu	CTC	0.20	0.20	0.12	0.24	0.14	0.17	0.13
Lys	AAG	0.98	0.94	0.54	0.98	0.90	0.90	0.84
Lys	AAA	0.02	0.06	0.46	0.02	0.10	0.10	0.16
Met	ATG	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Phe	TTT	0.28	0.32	0.42	0.31	0.24	0.27	0.35
Phe	TTC	0.72	0.68	0.58	0.69	0.76	0.73	0.65
Pro	CCG	0.18	0.31	0.09	0.07	0.04	0.34	0.15
Pro	CCA	0.06	0.17	0.36	0.07	0.04	0.20	0.24
Pro	CCT	0.10	0.14	0.25	0.17	0.04	0.19	0.29
Pro	CCC	0.66	0.38	0.29	0.69	0.04	0.27	0.32
Ser	AGT	0.03	0.04	0.14	0.02	0.08	0.08	0.07
Ser	AGC	0.27	0.38	0.18	0.18	0.31	0.27	0.31
Ser	TCG	0.12	0.14	0.08	0.10	0.02	0.19	0.10
Ser	TCA	0.03	0.08	0.14	0.08	0.09	0.09	0.14
Ser	TCT	0.09	0.11	0.26	0.18	0.19	0.14	0.13
Ser	TCC	0.47	0.24	0.20	0.44	0.30	0.24	0.24
Thr	ACG	0.11	0.20	0.13	0.05	0.12	0.27	0.19
Thr	ACA	0.01	0.20	0.32	0.07	0.20	0.12	0.23
Thr	ACT	0.12	0.13	0.29	0.12	0.24	0.20	0.18
Thr	ACC	0.76	0.47	0.26	0.76	0.44	0.41	0.40
Trp	TGG	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Tyr	TAT	0.07	0.15	0.43	0.27	0.28	0.24	0.19
Tyr	TAC	0.93	0.85	0.57	0.73	0.72	0.76	0.81
Val	GTG	0.71	0.54	0.37	0.60	0.54	0.46	0.67
Val	GTA	0.00	0.05	0.25	0.03	0.09	0.07	0.09
Val	GTT	0.11	0.14	0.24	0.09	0.14	0.17	0.09
Val	GTC	0.18	0.27	0.14	0.28	0.23	0.30	0.21

[0886]

아미노산	코돈	클로스테리움 셀라세토설-스트리고설-리토랄레 복합체	두날리엘라 비리디스	두날리엘라 살리나	고니움 펙토랄레	파에오닥틸룸 트리코르누툼	카에토세로스 컴프레섬
Ala	GCG	0.48	0.13	0.15	0.43	0.15	0.08
Ala	GCA	0.10	0.27	0.20	0.09	0.10	0.37
Ala	GCT	0.15	0.25	0.27	0.08	0.23	0.36
Ala	GCC	0.26	0.35	0.39	0.41	0.52	0.18

[0887]

Arg	AGG	0.04	0.25	0.22	0.11	0.02	0.14
Arg	AGA	0.00	0.06	0.05	0.00	0.04	0.29
Arg	CGG	0.18	0.08	0.12	0.40	0.10	0.00
Arg	CGA	0.00	0.06	0.06	0.05	0.12	0.19
Arg	CGT	0.13	0.15	0.13	0.08	0.41	0.18
Arg	CGC	0.64	0.39	0.43	0.35	0.31	0.00
Asn	AAT	0.04	0.17	0.23	0.07	0.30	0.58
Asn	AAC	0.96	0.83	0.77	0.93	0.65	0.42
Asp	GAT	0.30	0.38	0.40	0.11	0.41	0.53
Asp	GAC	0.70	0.62	0.60	0.89	0.59	0.47
Cys	TGT	0.06	0.24	0.17	0.20	0.39	0.44
Cys	TGC	0.94	0.76	0.83	0.90	0.61	0.56
End	TGA	0.75	0.31	0.37	0.50	0.06	0.50
End	TAG	0.00	0.15	0.14	0.00	0.13	0.00
End	TAA	0.25	0.54	0.49	0.50	0.81	0.50
Gln	CAG	0.53	0.36	0.32	0.31	0.23	0.16
Gln	CAA	0.09	0.12	0.08	0.07	0.14	0.19
Glu	GAG	0.31	0.44	0.51	0.56	0.21	0.28
Glu	GAA	0.06	0.00	0.09	0.07	0.42	0.37
Gly	GGG	0.31	0.14	0.10	0.18	0.08	0.12
Gly	GGA	0.06	0.11	0.12	0.09	0.34	0.33
Gly	GGT	0.09	0.22	0.22	0.07	0.30	0.39
Gly	GGC	0.53	0.54	0.56	0.65	0.28	0.16
His	CAT	0.33	0.25	0.25	0.43	0.28	0.84
His	CAC	0.67	0.75	0.75	0.57	0.72	0.16
Ile	ATA	0.03	0.03	0.03	0.07	0.03	0.12
Ile	ATT	0.23	0.25	0.31	0.33	0.51	0.65
Ile	ATC	0.74	0.72	0.66	0.59	0.46	0.23
Leu	TTG	0.04	0.11	0.12	0.04	0.26	0.11
Leu	TTA	0.00	0.01	0.01	0.00	0.02	0.14
Leu	CTG	0.31	0.60	0.61	0.84	0.15	0.05
Leu	CTA	0.01	0.05	0.04	0.01	0.05	0.08
Leu	CTT	0.04	0.07	0.08	0.05	0.18	0.51
Leu	CTC	0.60	0.16	0.14	0.26	0.34	0.11
Lys	AAG	0.86	0.87	0.89	0.93	0.75	0.52
Lys	AAA	0.14	0.13	0.11	0.07	0.25	0.48
Met	ATG	1.00	1.00	1.00	1.00	1.00	1.00
Phe	TTT	0.09	0.25	0.29	0.10	0.44	0.65
Phe	TTC	0.91	0.75	0.71	0.90	0.56	0.35
Pro	CCG	0.28	0.10	0.08	0.53	0.29	0.05
Pro	CCA	0.15	0.10	0.17	0.09	0.12	0.45
Pro	CCT	0.12	0.10	0.30	0.04	0.20	0.33
Pro	CCC	0.44	0.10	0.45	0.34	0.40	0.17
Ser	AGT	0.04	0.09	0.06	0.02	0.12	0.14
Ser	AGC	0.05	0.31	0.32	0.20	0.12	0.07

[0888]

Ser	TCG	0.22	0.04	0.06	0.42	0.19	0.08
Ser	TCA	0.16	0.08	0.10	0.09	0.06	0.31
Ser	TCT	0.05	0.17	0.15	0.07	0.15	0.23
Ser	TCC	0.47	0.31	0.30	0.20	0.35	0.18
Thr	ACG	0.30	0.16	0.13	0.42	0.23	0.10
Thr	ACA	0.06	0.21	0.18	0.03	0.13	0.38
Thr	ACT	0.22	0.18	0.23	0.08	0.19	0.27
Thr	ACC	0.42	0.46	0.46	0.47	0.45	0.25
Trp	TGG	1.00	1.00	1.00	1.00	1.00	1.00
Tyr	TAT	0.07	0.16	0.21	0.12	0.18	0.67
Tyr	TAC	0.93	0.84	0.79	0.88	0.82	0.33
Val	GTG	0.50	0.64	0.62	0.57	0.22	0.30
Val	GTA	0.02	0.03	0.05	0.04	0.09	0.27
Val	GTT	0.06	0.11	0.11	0.04	0.22	0.10
Val	GTC	0.42	0.22	0.23	0.35	0.47	0.33

[0889]

아미노산	코돈	실린드로덱카 후시포르미스	암피디늄 칼데라에	심비오디늄 마이크로 아드리악티쿰	나노 클로롭시스 종	시클로텔라 크립티카	나비큘라 펠리큘로사	탈라시오시라 슈도나나	씨. 레인하르드티
Ala	GCG	0.07	0.17	0.22	0.24	0.11	0.00	0.11	0.35
Ala	GCA	0.14	0.33	0.26	0.10	0.16	0.13	0.25	0.08
Ala	GCT	0.35	0.29	0.20	0.17	0.45	0.44	0.33	0.13
Ala	GCC	0.43	0.20	0.32	0.48	0.27	0.44	0.30	0.43
Arg	AGG	0.09	0.15	0.27	0.00	0.09	0.05	0.18	0.05
Arg	AGA	0.14	0.03	0.27	0.00	0.05	0.10	0.17	0.01
Arg	CGG	0.06	0.08	0.09	0.00	0.04	0.05	0.06	0.20
Arg	CGA	0.16	0.18	0.09	0.29	0.08	0.35	0.11	0.04
Arg	CGT	0.34	0.18	0.09	0.14	0.47	0.20	0.34	0.09
Arg	CGC	0.22	0.40	0.18	0.57	0.28	0.25	0.15	0.62
Asn	AAT	0.42	0.37	0.21	0.00	0.25	0.47	0.43	0.09
Asn	AAC	0.58	0.63	0.79	1.00	0.75	0.53	0.57	0.91
Asp	GAT	0.54	0.54	0.50	0.20	0.52	0.20	0.56	0.14
Asp	GAC	0.46	0.46	0.50	0.80	0.48	0.80	0.44	0.86
Cys	TGT	0.44	0.75	0.50	0.00	0.29	0.10	0.54	0.10
Cys	TGC	0.56	0.25	0.50	1.00	0.71	0.90	0.46	0.90
End	TGA	0.13	0.50	1.00	0.00	0.10	0.00	0.31	0.27
End	TAG	0.10	0.00	0.00	0.00	0.00	0.00	0.38	0.22
End	TAA	0.77	0.50	0.00	1.00	0.90	1.00	0.31	0.52
Gln	CAG	0.12	0.33	0.28	0.41	0.19	0.21	0.16	0.38
Gln	CAA	0.25	0.15	0.17	0.00	0.17	0.28	0.19	0.04
Glu	GAG	0.23	0.41	0.50	0.59	0.38	0.17	0.40	0.55
Glu	GAA	0.39	0.10	0.06	0.00	0.26	0.34	0.26	0.03
Gly	GGG	0.06	0.19	0.32	0.10	0.10	0.03	0.12	0.11
Gly	GGA	0.47	0.10	0.12	0.05	0.45	0.28	0.51	0.06

[0890]

Gly	GGT	0.35	0.34	0.16	0.25	0.22	0.13	0.23	0.11
Gly	GGC	0.12	0.37	0.40	0.60	0.24	0.56	0.14	0.72
His	CAT	0.39	0.12	0.40	0.00	0.42	1.00	0.50	0.11
His	CAC	0.61	0.88	0.60	1.00	0.58	0.00	0.50	0.89
Ile	ATA	0.06	0.05	0.00	0.00	0.04	0.00	0.08	0.03
Ile	ATT	0.42	0.53	0.38	0.14	0.53	0.73	0.38	0.22
Ile	ATC	0.52	0.42	0.63	0.86	0.42	0.27	0.54	0.75
Leu	TTG	0.26	0.35	0.39	0.22	0.20	0.16	0.29	0.04
Leu	TTA	0.09	0.01	0.00	0.00	0.03	0.00	0.05	0.01
Leu	CTG	0.09	0.22	0.39	0.09	0.06	0.12	0.08	0.73
Leu	CTA	0.05	0.00	0.04	0.00	0.03	0.04	0.06	0.03
Leu	CTT	0.37	0.31	0.13	0.04	0.39	0.36	0.20	0.05
Leu	CTC	0.13	0.12	0.04	0.65	0.29	0.32	0.32	0.15
Lys	AAG	0.60	0.93	0.85	1.00	0.70	0.83	0.76	0.95
Lys	AAA	0.40	0.07	0.15	0.00	0.30	0.17	0.24	0.05
Met	ATG	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Phe	TTT	0.37	0.21	0.25	0.20	0.31	0.78	0.38	0.16
Phe	TTC	0.63	0.79	0.75	0.80	0.69	0.22	0.62	0.84
Pro	CCG	0.11	0.14	0.18	0.08	0.10	0.21	0.16	0.33
Pro	CCA	0.33	0.42	0.09	0.08	0.16	0.29	0.31	0.08
Pro	CCT	0.32	0.22	0.41	0.25	0.35	0.21	0.31	0.13
Pro	CCC	0.24	0.22	0.32	0.58	0.39	0.29	0.23	0.47
Ser	AGT	0.12	0.13	0.09	0.00	0.09	0.13	0.18	0.04
Ser	AGC	0.09	0.24	0.14	0.13	0.08	0.28	0.11	0.35
Ser	TCG	0.13	0.03	0.05	0.00	0.15	0.25	0.17	0.25
Ser	TCA	0.12	0.25	0.05	0.00	0.12	0.08	0.12	0.05
Ser	TCT	0.30	0.16	0.23	0.13	0.39	0.25	0.23	0.07
Ser	TCC	0.24	0.19	0.45	0.75	0.18	0.03	0.19	0.25
Thr	ACG	0.09	0.14	0.10	0.28	0.10	0.18	0.21	0.30
Thr	ACA	0.15	0.28	0.10	0.00	0.15	0.09	0.19	0.08
Thr	ACT	0.39	0.12	0.10	0.17	0.33	0.41	0.28	0.10
Thr	ACC	0.37	0.47	0.70	0.56	0.43	0.32	0.32	0.52
Trp	TGG	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Tyr	TAT	0.38	0.32	0.20	0.00	0.38	0.20	0.39	0.10
Tyr	TAC	0.62	0.68	0.80	1.00	0.62	0.80	0.61	0.90
Val	GTG	0.11	0.65	0.67	0.31	0.16	0.18	0.29	0.67
Val	GTA	0.06	0.05	0.00	0.00	0.09	0.09	0.16	0.03
Val	GTT	0.38	0.08	0.11	0.15	0.42	0.09	0.28	0.07
Val	GTC	0.46	0.21	0.22	0.54	0.33	0.64	0.27	0.22

[0891]

아미노산	코돈	야로위아 리폴리티카	모르티에렐라 알피나	로도코커스 오파커스
Ala	GCG	0.08	0.14	0.35

[0892]

Ala	GCA	0.11	0.12	0.14
Ala	GCT	0.35	0.29	0.09
Ala	GCC	0.46	0.45	0.43
Arg	AGG	0.05	0.05	0.05
Arg	AGA	0.13	0.06	0.02
Arg	CGG	0.12	0.06	0.26
Arg	CGA	0.52	0.09	0.12
Arg	CGT	0.11	0.32	0.11
Arg	CGC	0.07	0.42	0.44
Asn	AAT	0.17	0.15	0.21
Asn	AAC	0.83	0.85	0.79
Asp	GAT	0.35	0.42	0.24
Asp	GAC	0.65	0.58	0.76
Cys	TGT	0.46	0.13	0.26
Cys	TGC	0.54	0.87	0.74
End	TGA	0.16	0.05	0.72
End	TAG	0.38	0.25	0.17
End	TAA	0.46	0.70	0.11
Gln	CAG	0.33	0.36	0.28
Gln	CAA	0.08	0.06	0.06
Glu	GAG	0.44	0.49	0.45
Glu	GAA	0.14	0.09	0.22
Gly	GGG	0.05	0.03	0.18
Gly	GGA	0.28	0.29	0.15
Gly	GGT	0.32	0.32	0.20
Gly	GGC	0.34	0.36	0.48
His	CAT	0.34	0.27	0.20
His	CAC	0.66	0.73	0.80
Ile	ATA	0.03	0.01	0.05
Ile	ATT	0.44	0.33	0.14
Ile	ATC	0.53	0.66	0.81
Leu	TTG	0.09	0.27	0.09
Leu	TTA	0.02	0.00	0.01
Leu	CTG	0.37	0.26	0.41
Leu	CTA	0.05	0.02	0.03
Leu	CTT	0.18	0.12	0.06
Leu	CTC	0.29	0.32	0.40
Lys	AAG	0.84	0.91	0.80
Lys	AAA	0.16	0.09	0.20
Met	ATG	1.00	1.00	1.00
Phe	TTT	0.38	0.39	0.09
Phe	TTC	0.62	0.61	0.91
Pro	CCG	0.10	0.07	0.52
Pro	CCA	0.10	0.08	0.09
Pro	CCT	0.32	0.36	0.07

[0893]

Pro	CCC	0.47	0.49	0.32
Ser	AGT	0.07	0.05	0.08
Ser	AGC	0.11	0.14	0.23
Ser	TCG	0.16	0.32	0.33
Ser	TCA	0.08	0.08	0.07
Ser	TCT	0.28	0.12	0.05
Ser	TCC	0.30	0.29	0.24
Thr	ACG	0.11	0.17	0.28
Thr	ACA	0.14	0.10	0.11
Thr	ACT	0.26	0.23	0.07
Thr	ACC	0.49	0.49	0.53
Trp	TGG	1.00	1.00	1.00
Tyr	TAT	0.18	0.20	0.18
Tyr	TAC	0.82	0.80	0.82
Val	GTG	0.33	0.22	0.37
Val	GTA	0.05	0.02	0.05
Val	GTT	0.26	0.27	0.10
Val	GTC	0.36	0.49	0.49

[0894]

표 70

[0895] 지질 생합성 경로 단백질들

3-케토아실 ACP 합성 효소 쿠페아 후커리아나 3-케토아실-ACP 합성 효소(GenBank 수탁 번호제AAC68861.1호), 쿠페아 라이티 베타-케토아실-ACP 합성 효소 II(GenBank 수탁 번호 제AAB37271.1호), 쿠페아 란세올라타 베타-케토아실-ACP 합성 효소 IV(GenBank 수탁 번호 제CAC59946.1호), 쿠페아 라이티 베타-케토아실-ACP 합성 효소 II(GenBank 수탁 번호 제AAB37270.1호), 리시누스 커뮤니스 케토아실-ACP 합성 효소(GenBank 수탁 번호 제XP_002516228호), 고시피움 히르서텀 케토아실-ACP 합성 효소(GenBank 수탁 번호 제ADK23940.1호), 글리신 맥스 색소체 3-케토-아실-ACP 합성 효소 II-A(GenBank 수탁 번호 제AAW88763.1호), 엘라에이스 기니엔시스 베타-케토아실-ACP 합성 효소 II(GenBank 수탁 번호 제AAF26738.2호), 헬리안터스 아너스 색소체 3-케토-아실-ACP 합성 효소 I(GenBank 수탁 번호 제ABM53471.1호), 글리신 맥스 3-케토-아실-ACP 합성 효소 I(GenBank 수탁 번호 제NP_001238610.1호), 헬리안터스 아너스 색소체 3-케토-아실-ACP 합성 효소 II(GenBank 수탁 번호 제ABI18155.1호), 브라시카 나푸스 베타-케토아실-ACP 합성 효소 2(GenBank 수탁 번호 제AAF61739.1호), 페릴라 프루테센스 베타-케토아실-ACP 합성 효소 II(GenBank 수탁 번호 제AAC04692.1호), 헬리안터스 아너스 베타-케토아실-ACP 합성 효소 II(GenBank 수탁 번호 제ABI18155호), 리시누스 커뮤니스 베타-케토아실-ACP 합성 효소 II(GenBank 수탁 번호 제AAA33872호), 해마토코커스 플루비알리스 베타-케토아실 아실기 운반 단백질 합성 효소(GenBank 수탁 번호 제HM560033.1호), 자트로파 컬카스 베타 케토아실-ACP 합성 효소 I(GenBank 수탁 번호 제ABJ90468.1호), 파퓰러스 트리코카르파 베타-케토아실-ACP 합성 효소 I(GenBank 수탁 번호 제XP_002303661.1호), 코리안드럼 사티범 베타-케토아실-ACP 합성 효소 I(GenBank 수탁 번호 제AAK58535.1호), 아라비돕시스 탈리아나 3-옥소아실-[아실기-운반 단백질] 합성 효소 I(GenBank 수탁 번호 제NP_001190479.1호), 비티스 비니페라 3-옥소아실-[아실기-운반 단백질] 합성 효소 I(GenBank 수탁 번호 제XP_002272874.2호) 지방 아실-ACP 티오에스테라제 움벨루라리아 캘리포니카 지방 아실-ACP 티오에스테라제(GenBank 수탁 번호 제AAC49001호), 신나모뭄 캠포라 지방 아실-ACP 티오에스 프래그란스 테라제(GenBank 수탁 번호 제Q39473호), 움벨루라리아 캘리포니카 지방 아실-ACP 티오에스테라제(GenBank 수탁 번호 제Q41635호), 미리스티카 프래그란스지방 아실-ACP 티오에스테라제(GenBank 수탁 번호 제AAB71729호), 미리스티카 프래그란스 지방 아실-ACP 티오에스테라제(GenBank 수탁 번호 제AAB71730호), 엘라에이스 기니엔시스 지방 아실-ACP 티오에스테라제(GenBank 수탁 번호 제ABD83939호), 엘라에이스 기니엔시스 지방 아실-ACP 티오에스테라제(GenBank 수탁 번호 제AAD42220호), 파퓰러스 토멘토사 지방 아실-ACP 티오에스테라제(GenBank 수탁 번호 제ABC47311호), 아라비돕시스 탈리아나 지방 아실-ACP 티오에스테라제(GenBank 수탁 번호 제NP_172327호), 아라비돕시스 탈리아나 지방 아실-ACP 티오에스테라제(GenBank 수탁 번호 제CAA85387

호), 아라비돕시스 탈리아나 지방 아실-ACP 티오에스테라제(GenBank 수탁 번호 제CAA85388호), 고시피움 히르서텀 지방 아실-ACP 티오에스테라제(GenBank 수탁 번호 제Q9SQI3호), 쿠페아 란세올라타 지방 아실-ACP 티오에스테라제(GenBank 수탁 번호 제CAA54060호), 쿠페아 후커리아나 지방 아실-ACP 티오에스테라제(GenBank 수탁 번호 제AAC72882호), 쿠페아 칼로필라 아종 메소스테몬 지방 아실-ACP 티오에스테라제(GenBank 수탁 번호 제ABB71581호), 쿠페아 란세올라타 지방 아실-ACP 티오에스테라제(GenBank 수탁 번호 제CAC19933호), 엘라에이스 기니엔시스 지방 아실-ACP 티오에스테라제(GenBank 수탁 번호 제AAL15645호), 쿠페아 후커리아나 지방 아실-ACP 티오에스테라제(GenBank 수탁 번호 제Q39513호), 고시피움 히르서텀 지방 아실-ACP 티오에스테라제(GenBank 수탁 번호 제AAD01982호), 비티스 비니페라 지방 아실-ACP 티오에스테라제(GenBank 수탁 번호 제CAN81819호), 가르시니아 망고스타나 지방 아실-ACP 티오에스테라제(GenBank 수탁 번호 제AAB51525호), 브라시카 준세아 지방 아실-ACP 티오에스테라제(GenBank 수탁 번호 제ABI18986호), 마듀카 론지폴리아 지방 아실-ACP 티오에스테라제(GenBank 수탁 번호 제AAX51637호), 브라시카 나푸스 지방 아실-ACP 티오에스테라제(GenBank 수탁 번호 제ABH11710호), 브라시카 나푸스 지방 아실-ACP 티오에스테라제(GenBank 수탁 번호 제CAA52070.1호), 오리자 사티바(인디카 컬티바 군) 지방 아실-ACP 티오에스테라제(GenBank 수탁 번호 제EAY86877호), 오리자 사티바(자포니카 컬티바 군) 지방 아실-ACP 티오에스테라제(GenBank 수탁 번호 제NP_001068400호), 오리자 사티바(인디카 컬티바 군) 지방 아실-ACP 티오에스테라제(GenBank 수탁 번호 제EAY99617호), 쿠페아 후커리아나 지방 아실-ACP 티오에스테라제(GenBank 수탁 번호

제AAC49269호), 얼머스 아메리카나 지방 아실-ACP 티오에스테라제(GenBank 수탁 번호 제AAB71731호), 쿠페아 란세올라타 지방 아실-ACP 티오에스테라제(GenBank 수탁 번호 제CAB60830호), 쿠페아 팔루스트리스 지방 아실-ACP 티오에스테라제(GenBank 수탁 번호 제AAC49180호), 아이리스 게르마니카 지방 아실-ACP 티오에스테라제(GenBank 수탁 번호 제AAG43858호, 아이리스 게르마니카 지방 아실-ACP 티오에스테라제(GenBank 수탁 번호 제AAG43858.1호), 쿠페아 팔루스트리스 지방 아실-ACP 티오에스테라제(GenBank 수탁 번호 제AAC49179호), 미리스티카 프래그란스 지방 아실-ACP 티오에스테라제(GenBank 수탁 번호 제AAB71729호), 미리스티카 프래그란스 지방 아실-ACP 티오에스테라제(GenBank 수탁 번호 제AAB717291.1호), 쿠페아 후커리아나 지방 아실-ACP 티오에스테라제 GenBank 수탁 번호 제U39834호), 움벨루라리아 캘리포니카 지방 아실-ACP 티오에스테라제(GenBank 수탁 번호 제M94159호), 신나모뭄 캠포라 지방 아실-ACP 티오에스테라제(GenBank 수탁 번호 제U31813호), 리시누스 커뮤니스 지방 아실-ACP 티오에스테라제(GenBank 수탁 번호 제ABS30422.1호), 헬리안터스 아너스 아실-ACP 티오에스테라제(GenBank 수탁 번호 제AAL79361.1호), 자트로파 컬카스 아실-ACP 티오에스테라제(GenBank 수탁 번호 제ABX82799.3호), 제아 메이스 올레오일-아실기 운반 단백질 티오에스테라제(GenBank 수탁 번호 제ACG40089.1호), 해마토코커스 플루비알리스 지방 아실-ACP 티오에스테라제(GenBank 수탁 번호 제HM560034.1호)

불포화 효소

리넘 유시타티시멈 지방산 불포화 효소 3C(GenBank 수탁 번호 제ADV92272.1호), 리시누스 커뮤니스 오메가-3 지방산 불포화 효소, 소포체, 추정(GenBank 수탁 번호 제EEF36775.1호), 베르니시아 포르디 오메가-3 지방산 불포화 효소(GenBank 수탁 번호 제AAF12821호), 글리신 맥스 엽록소오메가 3 지방산 불포화 효소 아형 2(GenBank 수탁 번호 제ACF19424.1호), 프로토테카 모리포르미스 FAD-D 오메가 3 불포화 효소(서열 번호 221), 프로토테카 모리포르미스 리놀리에이트 불포화 효소(서열 번호 220), 칼타머스 팅크토리우스 델타 12 불포화 효소(GenBank 수탁 번호 제ADM48790.1호), 고시피움 히르서텀 오메가-6 불포화 효소(GenBank 수탁 번호 제CAA71199.1호), 글리신 맥스 마이크로좀 불포화 효소(GenBank 수탁 번호 제BAD89862.1호), 제아 메이스 지방산 불포화 효소(GenBank 수탁 번호 제ABF50053.1호), 브라시카 나파 리놀레산 불포화 효소(GenBank 수탁 번호 제AAA32994.1호), 카멜리나 사티바 오메가-3 불포화 효소(서열 번호 214), 프로토테카 모리포르미스 델타 12 불포화 효소 대립 유전자 2(서열 번호 212), 카멜리나 사티바 오메가-3 FAD7-1(서열 번호 215), 헬리안터스 아너스 스테아로일-ACP 불포화 효소(GenBank 수탁 번호 제AAB65145.1호), 리시누스 커뮤니스 스테아로일-ACP 불포화 효소(GenBank 수탁 번호 제AACG59946.1호), 브라시카 준세아 색소체 델타-9-스테아로일-ACP 불포화 효소(GenBank 수탁 번호 제AAD40245.1호), 글리신 맥스 스테아로일-ACP 불포화 효소(GenBank 수탁 번호 제ACJ39209.1호), 올레아 유로파에아 스테아로일-ACP 불포화 효소(GenBank 수탁 번호 제AAB67840.1호), 베르니시아 포르디 스테아로일-아실기-운반 단백질 불포화 효소(GenBank 수탁 번호 제ADC32803.1호), 데스큐라이니아 소피아 델타-12 지방산 불포화 효소(GenBank 수탁 번호 제ABS86964.2호), 유포비아 라가스카에 델타 12-올레산 불포화 효소(GenBank 수탁 번호 제AAS57577.1호), 클로렐라 불가리스 델타 12 지방산 불포화 효소(GenBank 수탁 번호 제ACF98528호), 클로렐라 불가리스 오메가-3 지방산 불포화 효소(GenBank 수탁 번호 제BAB78717호), 해마토코커스 플루비알리스 오메가-3 지방산 불포화 효소(GenBank 수탁 번호 제HM560035.1호), 해마토코커스 플루비알리스 스테아로일-ACP-불포화 효소 GenBank 수탁 번호 제EF586860.1호, 해마토코커스 플루비알리스 스테아로일-ACP-불포화 효소 GenBank 수탁 번호 제EF523479.1호

올리에이트 12-하이드록실라제 효소

리시누스 커뮤니스 올리에이트 12-하이드록실라제(GenBank 수탁 번호 제AAC49010.1호),

피사리아 린데이메리 올리에이트 12-하이드록실라제(GenBank 수탁 번호 제ABQ01458.1호),

피사리아 린데이메리 돌연 변이 이작용성 올리에이트 12-하이드록실라제:불포화 효소(GenBank 수탁 번호 제ACF17571.1호), 피사리아 린데이메리 이작용성 올리에이트 12-하이드록실라제:불포화 효소(GenBank 수탁 번호 제ACQ42234.1호), 피사리아 린데이메리 이작용성 올리에이트 12-하이드록실라제:불포화 효소(GenBank 수탁 번호 제AAC32755.1호), 아라비돕시스 라이라타 아종 라이라타(GenBank 수탁 번호 제XP_002884883.1호)

[0896] **실시예 22: 클로렐라 불가리스의 조작**

[0897] 본 발명에 따른 재조합 유전자의 발현을, 본원에 기술된 바와 같이 Chow(초우) 및 Tung(퉁) et al.에 의해 교시된 방법과 벡터를 변형하여 클로렐라 불가리스 내에서 수행할 수 있었다. 간단히 말해서, 문헌[Chow 및 Tung *et al., Plant Cell Reports*, Volume 18 (1999), pp. 778-780]에는 클로렐라 불가리스를 플라스미드 DNA로 핵내에서 안정적으로 형질 전환하는 것에 대해 개시되어 있다. 초우 및 퉁은, 전기 천공법과 같은 형질 전환법을 사용하여 플라스미드 pIG121-Hm(GenBank 수탁 번호 제AB489142호)을 클로렐라 불가리스에 도입하였다. pIG121-Hm의 뉴클레오티드 서열은, 베타-글루큐로니다제(GUS) 단백질 암호화 서열 상류의 CaMV 35S 프로모터에 작동 가능하도록 결합되었으며, 또한 GUS 단백질 암호화 서열 하류의 노팔린 합성 효소(*nos*) 유전자의 3'UTR/종결 인자에도 작동 가능하도록 결합된, GUS 리포터 유전자 생성물을 암호화하는 서열을 포함하였다. 플라스미드 pIG121-Hm 서

열은 하이그로마이신 B 항생제 내성 카세트를 추가로 포함하였다. 이 하이그로마이신 B 항생제 내성 카세트는 하이그로마이신 포스포트랜스퍼라제(*hpt*, GenBank 수탁 번호 제BAH24259호) 유전자 생성물을 암호화하는 서열에 작동 가능하도록 결합된 CaMV 35S 프로모터를 포함하였다. 형질 전환하기 전, 클로렐라 불가리스는 50ug/㎖ 하이그로마이신 B를 포함하는 배양 배지 중에서 번식될 수 없었다. pIG121-Hm 플라스미드로 형질 전환을 수행한 결과, 50ug/㎖ 하이그로마이신 B를 포함하는 배양 배지중에서 번식한 클로렐라 불가리스 형질 전환체를 제조하였다. 클로렐라 불가리스 내 *hpt* 유전자 생성물의 발현은, 50ug/㎖의 하이그로마이신 B의 존재 하에 형질 전환된 클로렐라 불가리스의 번식을 가능하게 만들었는데, 이로써 클로렐라 불가리스 내에서 사용될 선택가능한 마커로서 하이그로마이신 B 내성 카세트의 유용성이 확립되었다. GUS 리포터 유전자의 검출 가능한 활성은, CaMV 35S 프로모터 및 nos 3'UTR이 클로렐라 불가리스 내 이종 유전자 발현을 가능하게 만드는데 적당함을 암시하였다. 안정적인 형질 전환체의 게놈 DNA는 서던 블럿 분석법으로 평가하였다. 형질 전환된 클로렐라 불가리스의 선택과 유지는 YA 배지(한천 및 4g/L 효모 추출물)를 포함하는 한천 평판 상에서 수행하였다. 액체 배양 배지 중 클로렐라 불가리스의 번식은 초우 및 통에 의해 기술된 바와 같이 수행되었다. YA 배지 이외의 배지 중에서 클로렐라 불가리스의 번식에 대해서는 이미 개시된 바 있다(예를 들어, 문헌[Chader et al., *Revue des Energies Renouvelabes*, Volume 14 (2011), pp. 21-26] 및 [Illman et al., *Enzyme and Microbial Technology*, Vol. 27 (2000), pp. 631-635] 참조). 초우 및 통은, 플라스미드 pIG121-Hm, CaMV 35S 프로모터 및 아그로박테리움 튜미파시엔스 노팔린 합성 효소 유전자 3'UTR/종결 인자가 클로렐라 불가리스 내 이종 유전자 발현을 가능하게 만드는데 적당하다고 개시하였다. 뿐만 아니라, 초우 및 통은, 하이그로마이신 B 내성 카세트는 클로렐라 불가리스 내 선택가능한 마커로서 사용되기 적당하였다고 개시하였다. 부가의 플라스미드, 프로모터, 3'UTR/종결 인자, 그리고 클로렐라 불가리스 내 이종 유전자 발현을 가능하게 하는데 적당한 선택가능한 마커에 대해서는 문헌[Chader *et al.*, *Revue des Energies Renouvelabes*, Volume 14 (2011), pp. 21-26]에 기술되어 있다.

[0898] 본 발명의 하나의 구체예에서, 선택가능한 마커로 사용되는 하이그로마이신 B 유전자 생성물을 암호화하는 뉴클레오티드 서열을 포함하는 pIG121-Hm을 구성하고 지질 생합성 경로 발현 카세트 서열을 추가로 포함하도록 변형하여, 형질 전환 벡터를 제조하였다. 지질 생합성 경로 발현 카세트는 표 70으로부터 선택되는 하나 이상의 지질 생합성 경로 단백질을 암호화하였는데, 이 경우, 각각의 단백질 암호화 서열은, 표 69에 의한 클로렐라 불가리스의 핵내 유전자에 고유한 코돈 편향 사용을 반영하도록 클로렐라 불가리스 내에서의 발현에 대해 코돈 최적화되었다. 표 70의 지질 생합성 경로 단백질 각각에 있어서, 코돈 회적화된 유전자 서열은 각각 단백질 암호화 서열 상류의 CaMV 35S 프로모터에 작동 가능하도록 결합될 수 있었으며, 또한 단백질 암호화 서열의 하류 또는 3' 영역에 있는 아그로박테리움 튜미파시엔스 노팔린 합성 효소 유전자 3'UTR/종결 인자에 작동 가능하도록 결합될 수 있었다. 형질 전환 구조물은 추가로는 클로렐라 불가리스 게놈에 상동성인 영역(형질 전환 벡터의 표적화 게놈 통합용)을 포함할 수 있었다. 상동성 영역은 내인성 지질 생합성 경로 유전자의 하나 이상의 게놈 위치를 중단하는 것으로 선택할 수 있었다. 널리 공지된 형질 전환 기술, 예를 들어 전기 천공법 또는 기타 다른 공지된 방법을 통하여 클로렐라 불가리스는 형질 전환 벡터로 안정적으로 형질 전환될 수 있었다. 하이그로마이신 B 내성 유전자 생성물의 활성은, 예를 들어 하이그로마이신을 포함하는 한천 배지 상에서 형질 전환 벡터로 형질 전환된 클로렐라 불가리스를 선택하기 위한 마커로서 사용될 수 있었다. 클로렐라 불가리스 지질 생산에 적당한 생장 배지는 미량 금속이 보충되어 있으며, 임의로는 1.5g/L $NaNO_3$도 함유하는 BG11 배지(0.04g/L KH_2PO_4, 0.075g/L $CaCl_2$, 0.036g/L 시트르산, 0.006g/L 구연산 철 암모늄, 1mg/L EDTA 및 0.02g/L Na_2CO_3)를 포함하나, 이에 제한되는 것은 아니다. 지질 생산용으로서 클로렐라 불가리스를 배양하는데 적당한 부가 배지는, 예를 들어 와타나베 배지(1.5g/L KNO_3, 1.25g/L KH_2PO_4, 1.25 g l^{-1} $MgSO_4 \cdot 7H_2O$, 20 mg l^{-1} $FeSO_4 \cdot 7H_2O$와 미량 영양소, 그리고 저 질소 배지(203mg/L $(NH_4)_2HPO_4$, 2.236g/L KCl, 2.465g/L $MgSO_4$, 1.361g/L KH_2PO_4 및 10mg/L $FeSO_4$ 포함))를 포함한다(Illman et al., *Enzyme and Microbial Technology*, Vol. 27 (2000), pp. 631-635). 클로렐라 불가리스 지질의 지방산 프로필은 본원에 기술된 바와 같은 표준적인 지질 추출 및 분석 방법을 통하여 평가할 수 있었다.

[0899] **실시예 23: 클로렐라 엘립소이데아의 조작**

[0900] 본 발명에 따라서, 재조합 유전자의 발현은 본원에 기술된 바와 같이 Chen(첸) et al.에 의하여 교시된 방법과 벡터를 변형하여 클로렐라 엘립소이데아 내에서 이루어질 수 있었다. 간단히 말해서, 문헌[Chen *et al.*, *Current Genetics*, Vol. 39:5 (2001), pp. 365-370]에는 클로렐라 엘립소이데아를 플라스미드 DNA로 안정적으로 형질 전환하는 것에 대해 개시되어 있다. 첸은 전기 천공법과 같은 형질 전환법을 사용하여 플라스미드

pBinUWNP-1을 클로렐라 엘립소이데아에 도입하였다. pBinUWNP-1의 뉴클레오티드 서열은 호중구 펩티드-1(NP-1) 단백질 암호화 영역 상류의 제아 메이스 유비퀴틴(ubi1) 유전자 프로모터에 작동 가능하도록 결합되어 있으며, 또한 NP-1 단백질 암호화 영역 하류의 노팔린 합성 효소(*nos*) 유전자의 3'UTR/종결 인자에 작동 가능하도록 결합된, NP-1 토끼 유전자 생성물 암호화 서열을 포함하였다. 플라스미드 pBinUWNP-1의 서열은 또한 G418 항생제 내성 카세트를 추가로 포함하였다. 이 G418 항생제 내성 카세트는 아미노글리코시드 3'-포스포트랜스퍼라제(*aph 3'*) 유전자 생성물을 암호화하는 서열을 포함하

[0901] 였다. *aph 3'* 유전자 생성물은 항생제 G418에 대한 내성을 부여하였다. 형질 전환하기 전, 클로렐라 엘립소이데아는 30ug/㎖ G418을 포함하는 배양 배지 중에서 번식할 수 없었다. pBinUWNP-1 플라스미드로 형질 전환하여, 30ug/㎖ G418을 포함하는 선택 배양 배지 중에서 번식한 클로렐라 엘립소이데아 형질 전환체를 제조하였다. 클로렐라 엘립소이데아 내에서 *aph 3'* 유전자 생성물을 발현하면, 30ug/㎖ G418 존재 하에 형질 전환된 클로렐라 엘립소이데아가 번식할 수 있었으므로, 클로렐라 엘립소이데아 내에서 사용될 G418 항생제 내성 카세트의 선택 가능한 마커로서의 유용성이 확립되었다. NP-1 유전자 생성물의 검출 가능한 활성은, ubi1 프로모터와 *nos* 3'UTR이 클로렐라 엘립소이데아 내에서의 이종 유전자 발현을 가능하게 만드는데 적당하다는 것을 나타낸다. 안정적인 형질 전환체의 게놈 DNA는 서던 블럿 분석법에 의해 평가하였다. 형질 전환된 클로렐라 엘립소이데아의 선택 및 유지는 납(Knop) 배지(0.2g/L K_2HPO_4, 0.2g/L $MgSO_4 \cdot 7H_2O$, 0.12g/L KCl 및 10mg/L FeCl3(pH 6.0 내지 8.0)를 포함하고, 0.1% 효모 추출물과 0.2% 글루코스가 보충됨)로서, 15ug/㎖ G418(액체 배양액용) 또는 30ug/㎖ G418(1.8% 한천을 포함하는 고체 배양액용)을 포함하는 배지에서 수행되었다. 납 배지 이외의 배지 중에서 클로렐라 엘립소이데아를 번식시키는 것에 대해서는 개시된 바 있다(문헌[Cho *et al.*, Fisheries Science, Vol. 73:5 (2007), pp. 1050-1056], [Jarvis 및 Brown, *Current Genetics*, Vol. 19 (1991), pp.317-321] 및 [Kim *et al.*, *Marine Biotechnology*, Vol. 4 (2002), pp.63-73] 참조). 부가의 플라스미드, 프로모터, 3'UTR/종결 인자 및 클로렐라 엘립소이데아 내에서 이종 유전자를 발현하는데 적당한 선택가능한 마커에 대해서도 개시된 바 있다(문헌[Jarvis 및 Brown 및 Kim *et al.*, *Marine Biotechnology*, Vol. 4 (2002), pp.63-73] 참조). 첸은, 플라스미드 pBinUWNP-1, ubi1 프로모터, 그리고 아그로박테리움 튜미파시엔스 노팔린 합성 효소 유전자 3'UTR/종결 인자가 클로렐라 엘립소이데아 내에서 외인성 유전자 발현을 가능하게 하는데 적당하다고 개시하였다. 뿐만 아니라, 첸은, pBinUWNP-1 상에서 암호화된 G418 내성 카세트가 클로렐라 엘립소이데아 내에서 선택가능한 마커로서 사용하기 적당하였다고 개시하였다.

[0902] 본 발명의 하나의 구체예에서, *aph 3'* 유전자 생성물을 암호화하고, G418에 대한 내성을 부여하며, 선택가능한 마커로서 사용되는 뉴클레오티드 서열을 포함하는 벡터 pBinUWNP-1을 구성하고, 지질 생합성 경로 발현 카세트 서열을 추가로 포함하도록 변형하여, 형질 전환 벡터를 생산하였다. 지질 생합성 경로 발현 카세트는 표 70으로부터 선택된 하나 이상의 지질 생합성 경로 단백질을 암호화하였는데, 여기서, 각각의 단백질 암호화 서열은 표 69에 의한 클로렐라 엘립소이데아의 핵내 유전자에 고유한 코돈 편향 사용을 반영하도록 클로렐라 엘립소이데아 내에서의 발현에 대해 코돈 최적화되었다. 표 70의 지질 생합성 경로 단백질 각각에 있어서, 코돈 회적화된 유전자 서열은 각각 단백질 암호화 서열 상류의 제아 메이스 ubi1 프로모터에 작동 가능하도록 결합될 수 있으며, 또한 단백질 암호화 서열의 하류 또는 3' 영역에 있는 아그로박테리움 튜미파시엔스 노팔린 합성 효소 유전자 3'UTR/종결 인자에 작동 가능하도록 결합될 수 있었다. 형질 전환 구조물은 추가로 클로렐라 엘립소이데아 게놈에 상동성인 영역(형질 전환 벡터의 표적화 게놈 통합용)을 포함하였다. 상동성 영역은 내인성 지질 생합성 경로 유전자의 하나 이상의 게놈 위치를 중단하는 것으로 선택할 수 있었다. 널리 공지된 형질 전환 기술, 예를 들어 전기 천공법 또는 기타 다른 공지된 방법을 통하여 클로렐라 엘립소이데아는 형질 전환 벡터로 안정적으로 형질 전환되었다. *aph*3' 유전자 생성물의 활성은, 예를 들어 G418을 포함하는 납 한천 배지 상에서 형질 전환 벡터로 형질 전환된 클로렐라 엘립소이데아를 선택하기 위한 마커로서 사용될 수 있었다. 클로렐라 엘립소이데아 지질 생산에 적당한 생장 배지는 납 배지와, 자비스(Jarvis) 및 브라운(Brown) 및 Kim et al.에 의해 개시된 배양 배지들을 포함하나, 이에 제한되는 것은 아니다. 클로렐라 엘립소이데아 지질의 지방산 프로필은 표준적인 지질 추출법과 본원에 기술된 분석 방법을 통하여 평가될 수 있었다.

[0903] **실시예 24: 클로렐라 케슬러리의 조작**

[0904] 본 발명에 따라서, 재조합 유전자의 발현은 본원에 기술된 El-Sheekh(엘-쉬크) et al.에 의해 교시된 방법과 벡터를 변형하여 클로렐라 케슬러리 내에서 이루어졌다. 간단히 말해서, 문헌[El-Sheekh et al., Biologia Plantarium, Vol. 42:2 (1999), pp. 209-216]에는, 플라스미드 DNA로 클로렐라 케슬러리를 안정적으로 형질 전환하는 것에 대해 개시되어 있다. 엘-쉬크는 미세 입자 투사법과 같은 형질 전환법을 사용하여 플라스미드 pBI121(GenBank 수탁 번호 제AF485783호)를 클로렐라 케슬러리에 도입하였다. 플라스미드 pBI121은 카나마이신/

네오마이신 항생제 내성 카세트를 포함하였다. 이 카나마이신/네오마이신 항생제 내성 카세트는, 아그로박테리움 튜미파시엔스 노팔린 합성 효소(nos) 유전자 프로모터, 네오마이신 포스포트랜스퍼라제 II(nptII) 유전자 생성물을 암호화하는 서열(GenBank 수탁 번호 제AAL92039호)(카나마이신 및 G418에 대한 내성 부여용), 그리고 아그로박테리움 튜미파시엔스 노팔린 합성 효소(nos) 유전자의 3'UTR/종결 인자를 포함하였다. pBI121은, CaMV 35S 프로모터에 작동 가능하도록 결합되고, 또한 nos 유전자의 3'UTR/종결 인자에 작동 가능하도록 결합된 베타-글루쿠로니다제(GUS) 리포터 유전자 생성물을 암호화하는 서열을 추가로 포함하였다. 형질 전환하기 전, 클로렐라 케슬러리는 15ug/L 카나마이신을 포함하는 배양 배지 중에서 번식할 수 없었다. 상기 균주를 pBI121 플라스미드로 형질 전환하여, 15㎎/L 카나마이신을 포함하는 선택 배양 배지중에서 번식한 클로렐라 케슬러리 형질 전환체를 제조하였다. 클로렐라 캐슬러리 내 nptII 유전자 생성물의 발현은, 이 균주가 15㎎/L 카나마이신의 존재 하에서 번식할 수 있도록 만들 수 있으므로, 카나마이신/네오마이신 항생제 내성 카세트의 클로렐라 케슬러리 내에서 사용될 선택가능한 마커로서의 유용성이 확립되었다. GUS 유전자 생성물의 검출 가능한 활성은, CaMV 35S 프로모터와 nos 3'UTR이 클로렐라 케슬러리 내에서 이종 유전자 발현을 가능하도록 만드는데 적당하다는 것을 암시하였다. 안정적인 형질 전환체의 게놈 DNA는 서던 블럿 분석법으로 평가하였다. 엘-쉬크에 의해 개시된 바와 같이, 형질 전환된 클로렐라 케슬러리의 선택과 유지는, YEG 배지(1% 효모 추출물, 1% 글루코스) 및 15㎎/L 카나마이신을 포함하는 반고체 한천 평판 상에서 수행되었다. 엘-쉬크는 또한 YEG 액체 배양 배지 중 클로렐라 케슬러리의 번식에 대해 개시하였다. 지질 생산을 위해 클로렐라 케슬러리를 배양하는데 적당한 부가의 배지는 문헌[Sato et al., BBA Molecular and Cell Biology of Lipids, Vol. 1633 (2003), pp. 27-34]에 개시되어 있다. 엘-쉬크는, 플라스미드 pBI121, CaMV 프로모터, 그리고 노팔린 합성 효소 유전자 3'UTR/종결 인자가 클로렐라 케슬러리 내에서 이종 유전자 발현을 가능하게 만드는데 적당하다고 개시하였다. 뿐만 아니라, 엘-쉬크는, pBI121 상에 암호화된 카나마이신/네오마이신 내성 카세트가 클로렐라 케슬러리 내에서 선택가능한 마커로서 사용하기 적당하였다고 개시하였다.

[0905] 본 발명의 하나의 구체예에서, 선택가능한 마커로서 사용할 카나마이신/네오마이신 내성 유전자 생성물을 암호화하는 뉴클레오티드 서열을 포함하는 벡터 pBI121을 구성하고, 지질 생합성 경로 발현 카세트 서열을 추가로 포함하도록 변형하여, 형질 전환 벡터를 제조하였다. 지질 생합성 경로 발현 카세트는 표 70으로부터 선택된 하나 이상의 지질 생합성 경로 단백질을 암호화하였는데, 여기서, 각각의 단백질 암호화 서열은, 표 69에 의하여 클로렐라 케슬러리의 핵내 유전자에 고유한 코돈 편향 사용을 반영하도록 클로렐라 케슬러리 내에서의 발현에 대해 코돈 최적화되었다. 표 70의 지질 생합성 경로 단백질 각각에 있어서, 코돈 최적화 유전자 서열은 각각, 단백질 암호화 상류의 CaMV 35S 프로모터에 작동 가능하도록 결합될 수 있으며, 또한 단백질 암호화 서열의 하류 또는 3' 영역에서 아그로박테리움 튜미파시엔스 노팔린 합성 효소 유전자 3'UTR/종결 인자에 작동 가능하도록 결합될 수 있었다. 형질 전환 구조물은 추가로는 형질 전환 벡터로의 표적화 게놈 통합용인, 클로렐라 케슬러리 게놈에 상동성인 영역을 포함할 수 있었다. 상동성 영역은 내인성 지질 생합성 경로 유전자의 하나 이상의 게놈 위치를 중단하는 것으로 선택할 수 있었다. 형질 전환 벡터로 클로렐라 케슬러리를 안정적으로 형질 전환하는 것은, 널리 알려진 형질 전환 기술, 예를 들어 미세 입자 투사법 또는 기타 다른 공지된 방법들을 통하여 이루어졌다. nptII 유전자 생성물의 활성은, 카나마이신이나 네오마이신을 포함하는 YEG 한천 배지(이에 제한되는 것은 아님) 상에서 형질 전환 벡터로 형질 전환된 클로렐라 케슬러리를 선택하기 위한 마커로서 사용될 수 있었다. 클로렐라 케슬러리 지질 생산에 적당한 생장 배지는 YEG 배지와, Sato et al.에 의해 개시된 배양 배지를 포함하나, 이에 제한되는 것은 아니다. 클로렐라 케슬러리 지질의 지방산 프로필은 본원에 기술된 분석 방법과 표준적인 지질 추출법을 통하여 평가될 수 있었다.

[0906] **실시예 25: 두날리엘라 터티올렉타의 조작**

[0907] 본 발명에 따라서, 두날리엘라 터티올렉타 내 재조합 유전자의 발현은 본원에 기술된 바와 같이, Walker(워커) et al.에 의해 교시된 방법과 벡터를 변형함으로써 이루어질 수 있었다. 간단히 말해서, 문헌[Walker et al., Journal of Applied Phycology, Vol. 17 (2005), pp. 363-368]에는 두날리엘라 터티올렉타를 플라스미드 DNA로 안정적으로 핵내 형질 전환하는 것에 대해 개시되어 있다. 워커는 전기 천공법과 같은 형질 전환법을 사용하여 플라스미드 pDbleFLAG1.2를 두날리엘라 터티올렉타에 도입하였다. pDbleFLAG1.2는, 스트렙토알로테이쿠스 힌두스타너스(Streptoalloteichus hindustanus) 블레오마이신 결합 단백질(ble)을 암호화하는 서열을 포함하는 블레오마이신 항생제 내성 카세트(항생제 플레오마이신에 대한 내성 부여)를 암호화하는 서열로서, 두날리엘라 터티올렉타 리불로스-1,5-비스포스페이트 카복실라제/옥시게나제 소형 서브유닛 유전자(rbcS1, GenBank 수탁 번호 제AY530155호)의 3'UTR과 프로모터에 작동 가능하도록 결합된 서열을 포함하였다. 형질 전환하기 전, 두날리엘라 터티올렉타는 1㎎/L 플레오마이신을 포함하는 배양 배지 내에서 번식할 수 없었다. 이 균주를 pDbleFLAG1.2 플라스미드를 사용하여 형질 전환하여, 1㎎/L 플레오마이신을 포함하는 선택 배양 배지 내에서 번식한 두날리엘

라 터티올렉타 형질 전환체를 제조하였다. ble 유전자 생성물의 두날리엘라 터티올렉타 내 발현은, 1mg/L 플레오마이신의 존재 하에서 번식을 가능하게 만들었으며, 이로써 두날리엘라 터티올렉타 내에서 사용될 선택가능한 마커로서 블레오마이신 항생제 내성 카세트의 유용성이 확립되었다. 안정적인 형질 전환체의 게놈 DNA는 서던 블럿 분석법으로 평가하였다. 워커에 의해 개시된 바와 같이, 형질 전환된 두날리엘라 터티올렉타의 선택 및 유지는, 4.5g/L NaCl 및 1mg/L 페오마이신을 추가로 포함하는 두날리엘라 배지(DM, 문헌[Provasoli et al., Archiv fur Mikrobiologie, Vol. 25 (1957), pp. 392-428]에 개시된 바와 같음) 중에서 이루어졌다. 지질을 생산하기 위하여 두날리엘라 터티올렉타를 배양하는데 적당한 부가의 배지는 문헌[Takagi et al., Journal of Bioscience and Bioengineering, Vol. 101:3 (2006), pp. 223-226] 및 [Massart 및 Hanston, Proceedings Venice 2010, Third International Symposium on Energy from Biomass and Waste]에 기술되어 있다. 워커는, 플라스미드 pDbleFLAG1.2, 두날리엘라 터티올렉타 리불로스-1,5-비스포스페이트 카복실라제/옥시게나제 소형 서브유닛 유전자의 3'UTR 및 프로모터가, 두날리엘라 터티올렉타 내 이종 발현을 가능하게 하는데 적당하다고 개시하였다. 뿐만 아니라, 워커는 pDbleFLAG1.2 상에 암호화된 블레오마이신 내성 카세트가 두날리엘라 터티올렉타 내 선택가능한 마커로서 사용하기 적당하였다고 개시하였다.

[0908] 본 발명의 하나의 구체예에서는, 선택가능한 마커로서 사용되는 ble 유전자 생성물을 암호화하는 뉴클레오티드 서열을 포함하는 벡터 pDbleFLAG1.2를 구성하고, 지질 생합성 경로 발현 카세트 서열을 추가로 포함하도록 변형하여, 형질 전환 벡터를 제조하였다. 지질 생합성 경로 발현 카세트는 표 70으로부터 선택된 하나 이상의 지질 생합성 경로 단백질을 암호화하였는데, 여기서, 각각의 단백질 암호화 서열은, 표 69에 의하여 두날리엘라 터티올렉타의 핵내 유전자에 고유한 코돈 편향 사용을 반영하도록 두날리엘라 터티올렉타 내에서의 발현에 대해 코돈 최적화되었다. 표 70의 지질 생합성 경로 단백질 각각에 있어서, 코돈 최적화 유전자 서열은 각각, 단백질 암호화 서열 상류의 rbcS1 프로모터에 작동 가능하도록 결합될 수 있으며, 또한 단백질 암호화 서열의 하류 또는 3' 영역에 있는 rbcS1 3'UTR/종결 인자에 작동 가능하도록 결합될 수 있었다. 형질 전환 구조물은 형질 전환 벡터의 표적화 게놈 통합에 사용될 두날리엘라 터티올렉타 게놈에 대한 상동성 영역을 추가로 포함할 수도 있었다. 상동성 영역은 내인성 지질 생합성 경로 유전자의 하나 이상의 게놈 위치를 중단하는 것으로 선택할 수 있었다. 두날리엘라 터티올렉타를 형질 전환 벡터로 안정적으로 형질 전환하는 것은, 널리 공지된 형질 전환 기술, 예를 들어 전기 천공법 또는 기타 다른 공지된 방법들을 사용하여 이루어질 수 있었다. ble 유전자 생성물의 활성은, 예를 들어 페오마이신을 포함하는 DM 배지에서 형질 전환 벡터로 형질 전환된 두날리엘라 터티올렉타를 선택하기 위한 마커로서 사용될 수 있었다. 두날리엘라 터티올렉타 지질 생산에 적당한 생장 배지로서는 DM 배지와, Takagi et al. 및 Massart 및 Hanston에 의해 개시된 배양 배지를 포함하나, 이에 제한되는 것은 아니다. 두날리엘라 터티올렉타 지질의 지방산 프로필은 표준적인 지질 추출법과 본원에 기술된 분석 방법들을 통하여 평가될 수 있었다.

[0909] **실시예 26: 볼복스 카르테리의 조작**

[0910] 본 발명에 따라서 볼복스 카르테리 내 재조합 유전자의 발현은 본원에 기술된 바와 같이, Hallman(홀맨) 및 Rappel(라펠) et al.에 의해 교시된 방법과 벡터를 변형함으로써 이루어질 수 있었다. 간단히 말해서, 문헌[Hallman 및 Rappel et al., The Plant Journal, Volume 17 (1999), pp. 99-109]에는 볼복스 카르테리를 플라스미드 DNA로 안정적으로 핵내 형질 전환하는 것을 기술한다. 홀맨과 라펠은 미세 입자 투사법과 같은 형질 전환법을 사용하여 pzeoE 플라스미드를 볼복스 카르테리에 도입하였다. pzeoE 플라스미드는, 스트렙토알로테이쿠스 힌두스타너스 블레오마이신 결합 단백질(ble)을 암호화하는 서열을 포함하는 블레오마이신 항생제 내성 카세트(항생제 제오신에 대한 내성 부여)를 암호화하는 서열로서, 볼복스 카르테리 베타-튜불린 유전자(GenBank 수탁 번호 제L24547호)의 3'UTR과 프로모터에 작동 가능하도록 결합된 서열을 포함하였다. 형질 전환하기 전, 볼복스 카르테리는 1.5ug/mL 제오신을 포함하는 배양 배지 내에서 번식할 수 없었다. 이 균주를 pzeoE 플라스미드를 사용하여 형질 전환하여, 20ug/mL 이상의 제오신을 포함하는 선택 배양 배지 내에서 번식된 볼복스 카르테리 형질 전환체를 제조하였다. ble 유전자 생성물의 볼복스 카르테리 내 발현은, 20ug/mL 제오신의 존재 하에서 번식을 가능하게 만들었으며, 이로써 볼복스 카르테리 내에서 사용될 선택가능한 마커로서 블레오마이신 항생제 내성 카세트의 유용성이 확립되었다. 안정적인 형질 전환체의 게놈 DNA는 서던 블럿 분석법으로 평가하였다. 홀맨과 라펠에 의해 개시된 바와 같이, 형질 전환된 볼복스 카르테리의 선택 및 유지는, 1mg/L 페오마이신을 포함하는 볼복스 배지(VM, 문헌[Provasoli 및 Pintner, The Ecology of Algae, Special Publication No. 2 (1959)], [Tyron, C.A. 및 Hartman, R.T., eds., Pittsburgh: Univeristy of Pittsburgh, pp. 88-96]에 개시된 바와 같음) 중에서 이루어졌다. 지질을 생산하기 위하여 볼복스 카르테리를 배양하는데 적당한 배지는 문헌[Starr in Starr R,C,. Dev Biol Suppl., Vol. 4 (1970), pp.59-100]에 기술되어 있다. 홀맨과 라펠은 플라스미드 pzeoE, 볼복스 카르테리 베타-튜불린 유전자의 3'UTR 및 프로모터가, 볼복스 카르테리 내 이종 발현을 가

능하게 하는데 적당하다고 개시하였다. 뿐만 아니라, 홀맨과 라펠은 pzeoE 상에 암호화된 블레오마이신 내성 카세트가 볼복스 카르테리 내 선택가능한 마커로서 사용하기 적당하였다고 개시하였다. 부가의 플라스미드, 프로모터, 3'UTR/종결 인자, 그리고 볼복스 카르테리 내 이종 유전자 발현을 가능하게 하고, 볼복스 카르테리 내 선택 마커로서 사용되기 적당한 선택가능한 마커에 대해서도 개시된 바 있다(예를 들어, 문헌[Hallamann 및 Sumper, Proceedings of the National Academy of Sciences, Vol. 91 (1994), pp 11562-11566] 및 [Hallman 및 Wodniok, Plant Cell Reports, Volume 25 (2006), pp. 582-581]을 참조한다).

[0911] 본 발명의 하나의 구체예에서는, 선택가능한 마커로서 사용되는 ble 유전자 생성물을 암호화하는 뉴클레오티드 서열을 포함하는 벡터 pzeoE를 구성하고, 지질 생합성 경로 발현 카세트 서열을 추가로 포함하도록 변형하여, 형질 전환 벡터를 제조하였다. 지질 생합성 경로 발현 카세트는 표 70으로부터 선택된 하나 이상의 지질 생합성 경로 단백질을 암호화하였는데, 여기서, 각각의 단백질 암호화 서열은, 표 69에 의하여 볼복스 카르테리의 핵내 유전자에 고유한 코돈 편향 사용을 반영하도록 볼복스 카르테리 내에서의 발현에 대해 코돈 최적화되었다. 표 70의 지질 생합성 경로 단백질 각각에 있어서, 코돈 최적화 유전자 서열은 각각, 단백질 암호화 서열 상류의 볼복스 카르테리 베타-튜불린 프로모터에 작동 가능하도록 결합될 수 있으며, 또한 단백질 암호화 서열의 하류 또는 3' 영역에 있는 볼복스 카르테리 베타-튜불린 3'UTR/종결 인자에 작동 가능하도록 결합될 수 있었다. 형질 전환 구조물은 형질 전환 벡터의 표적화 게놈 통합에 사용될 볼복스 카르테리 게놈에 대한 상동성 영역을 추가로 포함할 수도 있었다. 상동성 영역은 내인성 지질 생합성 경로 유전자의 하나 이상의 게놈 위치를 중단하는 것으로 선택할 수 있었다. 당업자는 볼복스 카르테리 게놈 내 서열에 존재하는 이와 같은 상동성 영역을 동정할 수 있다(문헌[Prochnik et al., Science, Vol. 329:5988 (2010), pp223-226] 참조). 볼복스 카르테리를 형질 전환 벡터로 안정적으로 형질 전환하는 것은, 널리 공지된 형질 전환 기술, 예를 들어 미세 입자 투사법 또는 기타 다른 공지된 방법들을 사용하여 이루어졌다. ble 유전자 생성물의 활성은, 예를 들어 제오신을 포함하는 VM 배지 상에서 형질 전환 벡터로 형질 전환된 볼복스 카르테리를 선택하기 위한 마커로서 사용될 수 있었다. 볼복스 카르테리 지질 생산에 적당한 생장 배지로서는 VM 배지와, 스타(Starr)에 의해 개시된 배양 배지를 포함하나, 이에 제한되는 것은 아니다. 볼복스 카르테리 지질의 지방산 프로필은 표준적인 지질 추출법과 본원에 기술된 분석 방법들을 통하여 평가될 수 있었다.

[0912] **실시예 27: 해마토코커스 플루비알리스의 조작**

[0913] 본 발명에 따라서 해마토코커스 플루비알리스 내 재조합 유전자의 발현은 본원에 기술된 바와 같이, Steinbrenner(슈타인브레너) 및 Sandmann(샌드만) et al.에 의해 교시된 방법과 벡터를 변형함으로써 이루어질 수 있었다. 간단히 말해서, 문헌[Steinbrenner 및 Sandmann et al., Applied and Environmental Microbiology, Vol. 72:12 (2006), pp.7477-7484]에는 해마토코커스 플루비알리스를 플라스미드 DNA로 안정적으로 핵내 형질 전환하는 것에 대해 개시되어 있다. 슈타인브레너는 미세 입자 투사법과 같은 형질 전환법을 사용하여 pPlat-pds-L504R을 해마토코커스 플루비알리스에 도입하였다. pPlat-pds-L504R 플라스미드는, 해마토코커스 플루비알리스 피토엔 불포화 효소 유전자(Pds, GenBank 수탁 번호 제AY781170호)의 3'UTR, 단백질 암호화 서열 및 프로모터를 포함하는 노르플루라존 내성 카세트를 포함하였으며, 여기서, Pds의 단백질 암호화 서열은 504번 위치가 변형되어(즉 루신이 아르기닌으로 바뀌어), 제초제 노르플루라존에 대한 내성을 부여하는 유전자 생성물(Pds-L504)을 암호화하였다. pPlat-pds-L504R로 형질 전환하기 전, 해마토코커스 플루비알리스는 5uM 노르플루라존을 포함하는 배지 상에서 번식할 수 없었다. pPlat-pds-L504R 플라스미드로 형질 전환하여, 5uM 노르플루라존을 포함하는 선택 배양 배지중에서 번식한 헤마토코커스 플루비알리스 형질 전환체를 제조하였다. 해마토코커스 플루비알리스 내 Pds-L504R 유전자 생성물의 발현은, 5uM 노르플루라존의 존재 하에서 번식이 가능하도록 만들었는데, 이로써 해마토코커스 플루비알리스 내에서 사용되는 선택가능한 마커로서 노르플루라존 제초제 내성 카세트의 유용성이 확립되었다. 안정적인 형질 전환체의 게놈 DNA는 서던 블럿 분석법으로 평가하였다. 슈타인브레너에 의해 개시된 바와 같이, 형질 전환된 해마토코커스 플루비알리스의 선택 및 유지는, 2.42g/L 트리스-아세테이트 및 5mM 노르플루라존이 보충된 OHA 배지(OHM(0.41g/L KNO3, 0.03g/L Na2HPO4, 0.246g/L MgSO4·7H2O, 0.11g/L CaCl2·2H2O, 2.62mg/L Fe(III)citrate x H2O, 0.011mg/L CoCl2·6H2O, 0.012mg/L CuSO4·5H2O, 0.075mg/L Cr2O3, 0.98mg/L MnCl2·4H2O, 0.12mg/L Na2MoO4 x 2H2O, 0.005mg/L SeO2 및 25mg/L 바이오틴, 17.5 mg/L 티아민, 그리고 15mg/L 비타민 B12))를 포함하는 한천 평판 상에서 이루어졌다. 액체 배양액 중 해마토코커스 플루비알리스의 번식은 슈타인브레너 및 샌드만에 의해 기저 배지(문헌[Kobayashi et al., Applied and Environmental Microbiology, Vol. 59 (1993), pp.867-873]에 기술되어 있는 기저 배지)를 사용하여 이루어졌다. 슈타인브레너와 샌드만은, pPlat-pds-L504R 플라스미드, 해마토코커스 플루비알리스 파이토엔 불포화 효소 유전자의 3'UTR과 프로모터는 해마토코커스 플루비알리스 내 이종 발현을 가능하게 하는데 적당하였다고 개시하였다. 뿐만 아니라, 슈타인브레너와 샌드만은, pPlat-pds-L504R상에서 암호화되는 노르플루라존 내성 카세트는

해마토코커스 플루비알리스 내 선택가능한 마커로서 사용하기 적당하였다고 개시하였다. 부가의 플라스미드, 프로모터, 3'UTR/종결 인자, 그리고 해마토코커스 플루비알리스 내에서 이종 유전자를 발현하는데 적당한 선택가능한 마커에 대해서는 문헌[Kathiresan et al., Journal of Phycology, Vol. 45 (2009), pp 642-649]에 개시되어 있다.

[0914] 본 발명의 하나의 구체예에서, 선택가능한 마커로서 사용되는 Pds-L504R 유전자 생성물을 암호화하는 뉴클레오티드 서열을 포함하는 벡터 pPlat-pds-L504R을 구성하고, 이 벡터가 지질 생합성 경로 발현 카세트 서열을 추가로 포함하도록 변형하여, 형질 전환 벡터를 제조하였다. 지질 생합성 경로 발현 카세트는 표 70으로부터 선택되는 하나 이상의 지질 생합성 경로 단백질을 암호화하는데, 이 경우, 각각의 단백질 암호화 서열은, 표 69에 의하여 해마토코커스 플루비알리스의 핵내 유전자에 고유한 코돈 편향 사용을 반영하도록 해마토코커스 플루비알리스 내에서의 발현에 대해 코돈 최적화되었다. 표 70의 지질 생합성 경로 단백질 각각에 있어서, 코돈 최적화 유전자 서열은 각각, 단백질 암호화 서열 상류의 해마토코커스 플루비알리스 pds 유전자 프로모터에 작동 가능하도록 결합될 수 있으며, 또한 단백질 암호화 서열의 하류 또는 3' 영역에 있는 해마토코커스 플루비알리스 pds 유전자 3'UTR/종결 인자에 작동 가능하도록 결합될 수 있었다. 형질 전환 구조물은 형질 전환 벡터의 표적화 게놈 통합에 사용될 해마토코커스 플루비알리스 게놈에 대한 상동성 영역을 추가로 포함할 수도 있었다. 상동성 영역은 내인성 지질 생합성 경로 유전자의 하나 이상의 게놈 위치를 중단하는 것으로 선택할 수 있었다. 해마토코커스 플루비알리스를 형질 전환 벡터로 안정적으로 형질 전환하는 것은, 널리 공지된 형질 전환 기술, 예를 들어 미세 입자 투사법 또는 기타 다른 공지된 방법들을 사용하여 이루어졌다. Pds-L504R 유전자 생성물의 활성은, 예를 들어 노르플루라존을 포함하는 OHA 배지에서 형질 전환 벡터로 형질 전환된 해마토코커스 플루비알리스에 대하여 선택하기 위한 마커로서 사용될 수 있었다. 해마토코커스 플루비알리스 지질 생산에 적당한 생장 배지는 기저 배지 및 문헌[Kobayashi et al., Kathiresan et al, and Gong and Chen, Journal of Applied Phycology, Vol. 9:5 (1997), pp. 437-444]에 기술된 배양 배지를 포함하나, 이에 제한되는 것은 아니다. 해마토코커스 플루비알리스 지질의 지방산 프로필은 표준적인 지질 추출법과 본원에 기술된 분석 방법을 통하여 평가될 수 있었다.

[0915] **실시예 28: 클로스테리움 페라세로섬-스트리고섬-리토랄레 복합체의 조작**

[0916] 본 발명에 따라서 클로스테리움 페라세로섬-스트리고섬-리토랄레 복합체 내 재조합 유전자 발현은 본원에 기술된 바와 같은 Abe(아베) et al.에 의해 교시된 방법과 벡터를 변형하여 이루어질 수 있었다. 간단히 말하면, 문헌[Abe et al., Plant Cell Physiology, Vol. 52:9 (2011), pp. 1676-1685]에는, 클로스테리움 페라세로섬-스트리고섬-리토랄레 복합체를 플라스미드 DNA로 안정적으로 핵내 형질 전환하는 것에 관하여 개시되어 있다. 아베는 미세 입자 투사법과 같은 형질 전환법을 사용하여 플라스미드 pSA106을 클로스테리움 페라세로섬-스트리고섬-리토랄레 복합체에 도입하였다. 플라스미드 pSA106은, 스트렙토알로테이쿠스 힌두스타너스 블레오마이신 결합 단백질 유전자(ble, GenBank 수탁 번호 제CAA37050호)를 암호화하는 서열을 포함하는 블레오마이신 내성 카세트로서, 클로스테리움 페라세로섬-스트리고섬-리토랄레 복합체 엽록소 a/b 결합 단백질 유전자(CAB, GenBank 수탁 번호 제AB363403호)의 3'UTR과 프로모터에 작동 가능하도록 결합된 카세트를 포함하였다. Psa106으로 형질 전환하기 전, 클로스테리움 페라세로섬-스트리고섬-리토랄레 복합체는 3ug/mL 플레오마이신을 포함하는 배지에서 번식할 수 없었다. pSA106을 사용하여 형질 전환하여, 3ug/mL 플레오마이신을 포함하는 선택 배양 배지 내에서 번식한 클로스테리움 페라세로섬-스트리고섬-리토랄레 복합체 형질 전환체를 제조하였다. ble 유전자 생성물의 클로스테리움 페라세로섬-스트리고섬-리토랄레 복합체 내 발현은, 3ug/mL 플레오마이신의 존재 하에서 번식을 가능하게 만들었는데, 이로써 클로스테리움 페라세로섬-스트리고섬-리토랄레 복합체 내에서 사용될 선택가능한 마커로서 블레오마이신 항생제 내성 카세트의 유용성이 확립되었다. 안정적인 형질 전환체의 게놈 DNA는 서던 블럿 분석법으로 평가하였다. 아베에 의해 개시된 바에 의하면, 형질 전환된 클로스테리움 페라세로섬-스트리고섬-리토랄레 복합체의 선택 및 유지는 우선, C 배지(0.1g/L KNO3, 0.015g/L Ca(NO3)2·4H2O, 0.05g/L 글리세로포스페이트 -Na2, 0.04g/L MgSO4·7H2O, 0.5g/L 트리스(하이드록실메틸) 아미노메탄, 미량 무기물, 바이오틴, 비타민 B1 및 B12)를 포함하는 탑 한천에서 수행되었으며, 이후에는 이 복합체를 분리하여, C 배지를 포함하고 플레오마이신이 보충된 한천 평판으로 옮겼다고 개시하였다. 아베에 의해 개시된 바에 의하면, 액체 배양액 중 클로스테리움 페라세로섬-스트리고섬-리토랄레 복합체의 번식은 C 배지 중에서 수행되었다. 클로스테리움 페라세로섬-스트리고섬-리토랄레 복합체의 번식에 적당한 부가의 액체 배양 배지에 대해서는 문헌[Sekimoto et al., DNA Research, 10:4 (2003), pp. 147-153]에 기술되어 있다. 아베는, 클로스테리움 페라세로섬-스트리고섬-리토랄레 복합체 CAB 유전자의 3'UTR, 프로모터 및 pSA106 플라스미드는 클로스테리움 페라세로섬-스트리고섬-리토랄레 복합체 내에서 이종 유전자 발현을 가능하게 하는데 적당하였다고 개시하였다. 뿐만 아니라, 아베는 pSA106 상에 암호화된 블레오마이신 내성 카세트가 클로스테리움 페라세로섬-스트리고섬-리토랄레 복합체 내

에서 선택가능한 마커로서 사용되기 적당하였다고 개시하였다. 부가의 플라스미드, 프로모터, 3'UTR/종결 인자, 그리고 클로스테리움 페라세로섬-스트리고섬-리토랄레 복합체 내에서 이종 유전자 발현을 가능하게 만드는데 적당한 선택가능한 마커에 대해서는 문헌[Abe et al., Plant Cell Physiology, Vol. 49 (2008), pp. 625-632]에 개시된 바 있다.

[0917] 본 발명의 하나의 구체예에서, 선택가능한 마커로서 사용되는 ble 유전자 생성물을 암호화하는 뉴클레오티드 서열을 포함하는 벡터 pSA106을 구성하고, 지질 생합성 경로 발현 카세트 서열을 추가로 포함하도록 변형하여, 형질 전환 벡터를 제조하였다. 지질 생합성 경로 발현 카세트는 표 70으로부터 선택되는 하나 이상의 지질 생합성 경로 단백질을 암호화하는데, 이 경우, 각각의 단백질 암호화 서열은 표 69에 의하여 클로스테리움 페라세로섬-스트리고섬-리토랄레 복합체의 핵내 유전자에 고유한 코돈 편향 사용을 반영하도록 이 클로스테리움 페라세로섬-스트리고섬-리토랄레 복합체 내 발현에 대해 코돈 최적화되었다. 표 70의 지질 생합성 경로 단백질 각각에 있어서, 코돈 최적화 유전자 서열은 각각 단백질 암호화 서열 상류의 클로스테리움 페라세로섬-스트리고섬-리토랄레 복합체 CAB 유전자 프로모터에 작동 가능하도록 결합될 수 있으며, 단백질 암호화 서열의 하류 또는 3' 영역에 있는 클로스테리움 페라세로섬-스트리고섬-리토랄레 복합체 CAB 유전자 3'UTR/종결 인자에 작동 가능하도록 결합될 수 있었다. 형질 전환 구조물은 형질 전환 벡터의 표적화 게놈 통합에 사용되는 클로스테리움 페라세로섬-스트리고섬-리토랄레 복합체 게놈에 상동성인 영역들을 추가로 포함할 수 있었다. 내인성 지질 생합성 경로 유전자의 하나 이상의 게놈 위치를 중단하도록 상동성 영역들을 선택할 수 있었다. 클로스테리움 페라세로섬-스트리고섬-리토랄레 복합체는, 널리 공지된 형질 전환 기술, 예를 들어 미세 입자 투사법 또는 기타 다른 공지된 방법을 통하여 형질 전환 벡터로 안정적으로 형질 전환되었다. Ble 유전자 생성물의 활성은, 플레오마이신을 포함하는 C 배지(이에 제한되는 것은 아님) 상에서 형질 전환 벡터로 형질 전환된 클로스테리움 페라세로섬-스트리고섬-리토랄레 복합체를 선택하는 마커로서 사용될 수 있었다. 클로스테리움 페라세로섬-스트리고섬-리토랄레 복합체 지질 생산에 적당한 생장 배지는 C 배지 및 Abe et al. 및 Sekimoto et al.에 의해 개시된 바와 같은 배양 배지를 포함하나, 이에 제한되는 것은 아니다. 클로스테리움 페라세로섬-스트리고섬-리토랄레 복합체 지질의 지방산 프로필은 본원에 기술된 분석 방법과 표준적인 지질 추출법을 통하여 평가될 수 있었다.

[0918] **실시예 29: 두날리엘라 비리디스의 조작**

[0919] 본 발명에 의한 두날리엘라 비리디스 내에서의 재조합 유전자의 발현은 본원에 기술된 바와 같이 Sun(선) et al.에 의해 교시된 방법과 벡터를 변형함으로써 이루어질 수 있었다. 간단히 말해서, 문헌[Sun et al., Gene, Vol. 377 (2006), pp.140-149]에는 두날리엘라 비리디스를 플라스미드 DNA로 안정적으로 형질 전환하는 것에 관하여 개시되어 있다. 선은 전기 천공법과 같은 형질 전환법을 사용하여 전 두날리엘라 비리디스 질산염 환원 효소 유전자를 암호화하는 플라스미드 pDVNR을 돌연 변이 두날리엘라 비리디스(두날리엘라 비리디스 NR-돌연 변이체)에 도입하였다. 상기 NR-돌연 변이체는 질소원으로서 질산염을 사용하지 않았을 때에는 생장할 수 없었다. 질산염 환원 효소는 질산염이 아질산염으로 전환되는 것을 촉진한다. 형질 전환하기 전, 두날리엘라 비리디스 NR 돌연 변이체는 유일한 질소원으로서 질산염(NO3-)을 포함하는 배양 배지 중에서 번식할 수 없었다. NR 돌연 변이 두날리엘라 비리디스 내 두날리엘라 비리디스 NR 유전자 생성물의 발현은 질산염 대사 결핍을 복구하는 선택가능한 마커로서 사용되었다. pDVNR 플라스미드로 형질 전환하여, 유일한 탄소원으로서 질산염을 포함하는 한천 평판 상에서 생장할 수 있는 두날리엘라 비리디스 NR 유전자 생성물을 안정적으로 발현하는 NR-돌연 변이 두날리엘라 비리디스를 제조하였다. 안정적인 형질 전환체의 DNA는 서던 블럿 분석법에 의해 평가하였다. 형질 전환된 두날리엘라 비리디스(NR 돌연 변이체)의 선택과 유지는 5mM KNO3를 포함하는 한천 평판 상에서 수행되었다. 선은 또한 액체 배양 배지 내에서 두날리엘라 비리디스 NR 돌연 변이체와 두날리엘라 비리디스를 번식시키는 것에 관하여 개시하였다. 두날리엘라 비리디스의 번식에 적당한 부가의 배지에 대해서는 문헌[Gordillo et al., Journal of Applied Phycology, Vol. 10:2 (1998), pp. 135-144] 및 [Moulton 및 Burford, Hydrobiologia, Vols. 204-205:1 (1990), pp. 401-408]에 개시되어 있다. 선은, 두날리엘라 비리디스 질산염 환원 효소 유전자의 3' UTR/종결 인자, 프로모터 및 플라스미드 pDVNR은 두날리엘라 비리디스 NR-돌연 변이체 내에서 이종 발현을 가능하게 하는데 적당하였다고 개시하였다. 선은 또한 두날리엘라 비리디스 질산염 환원 효소 유전자 생성물의 발현은 두날리엘라 비리디스 NR 돌연 변이체 내 선택가능한 마커로서 사용하기 적당하였다고 개시하였다.

[0920] 본 발명의 하나의 구체예에서, 선택가능한 마커로서 사용되는 두날리엘라 비리디스 질산염 환원 효소(DvNR) 유전자 생성물을 암호화하는 뉴클레오티드 서열을 포함하는 벡터 pDVNR을 구성하고, 지질 생합성 경로 발현 카세트 서열을 추가로 포함하도록 변형하여, 형질 전환 벡터를 제조하였다. 지질 생합성 경로 발현 카세트는 표 70으로부터 선택되는 하나 이상의 지질 생합성 경로 단백질을 암호화하였는데, 이 경우, 각각의 단백질 암호화 서

열은, 표 69에 따라서, 두날리엘라 비리디스의 핵내 유전자에 고유한 코돈 편향 사용을 반영하도록 두날리엘라 비리디스 내 발현에 대해 코돈 최적화되었다. 표 70의 지질 생합성 경로 단백질 각각에 있어서, 코돈 최적화 유전자 서열은 각각 단백질 암호화 서열 상류의 DvNR 프로모터에 작동 가능하도록 결합될 수 있었으며, 또한 단백질 암호화 서열의 하류 또는 3' 영역에 있는 DvNR 3'UTR/종결 인자에 작동 가능하도록 결합될 수 있었다. 형질 전환 구조물은 형질 전환 벡터의 표적화 게놈 통합용인 두날리엘라 비리디스 게놈에 대해 상동성인 영역을 추가로 포함할 수 있었다. 상동성 영역은 내인성 지질 생합성 경로 유전자의 하나 이상의 게놈 위치를 중단하는 것으로 선택할 수 있었다. 두날리엘라 비리디스 NR 돌연 변이체를 형질 전환 벡터로 안정적으로 형질 전환하는 것은, 전기 천공법과 같은 널리 공지된 형질 전환 기술 또는 기타 다른 공지된 방법을 통해 이루어졌다. DvNR 유전자 생성물의 활성은, 예를 들어 두날리엘라 비리디스 NR 돌연 변이 균주의 질소 동화 결핍을 복구하고, 형질 전환 벡터를 안정적으로 발현하는 두날리엘라 비리디스 NR-돌연 변이체를 선택하는 선택가능한 마커로서 사용될 수 있었다. 두날리엘라 비리디스 지질 생산에 적당한 생장 배지는 Sun et al., Moulton 및 Burford, 그리고 Gordillo et al.에 의해 개시된 것들을 포함하나, 이에 제한되는 것은 아니다. 두날리엘라 비리디스 지질의 지방산 프로필은 표준적인 지질 추출법과 본원에 기술된 분석 방법들을 통하여 평가될 수 있었다.

[0921] **실시예 30: 두날리엘라 살리나의 조작**

[0922] 본 발명에 의하여 두날리엘라 살리나 내에서 재조합 유전자를 발현하는 것은 본원에 기술된 바와 같이 Geng(겡) et al.에 의해 교시된 방법 및 벡터를 변형하여 이루어질 수 있었다. 간단히 말해서, 문헌[Geng et al., Journal of Applied Phycology, Vol. 15 (2003), pp. 451-456]에는 두날리엘라 살리나를 플라스미드 DNA로 안정적으로 형질 전환하는 것에 대해 개시되어 있다. 겡은 전기 천공법과 같은 형질 전환법을 사용하여 pUWHBsAg-CAT 플라스미드를 두날리엘라 살리나에 도입하였다. pUWHBsAg-CAT은, HBsAG 단백질 암호화 영역 상류의 제아 메이스 ubi1 프로모터에 작동 가능하도록 결합되고, HBsAG 단백질 암호화 영역 하류의 아그로박테리움 튜미파시엔스 노팔린 합성 효소 유전자(nos)의 3'UTR/종결 인자에 작동 가능하도록 결합된 B형 간염 표면 항원을 암호화하는 서열을 포함하는, B형 간염 표면 항원(HBsAG) 발현 카세트를 포함하였다. pUWHBsAg-CAT는 또한, 유인원 바이러스 40 프로모터 및 인핸서와 작동 가능하도록 결합된, 클로람페니콜 아세틸트랜스퍼라제(CAT) 유전자 생성물(항생제 클로람페니콜에 대한 내성 부여)을 암호화하는 서열을 포함하는 클로람페니콜 내성 카세트를 추가로 포함하였다. pUWHBsAg-CAT으로 형질 전환하기 전, 두날리엘라 살리나는 60㎎/L 클로람페니콜을 포함하는 배지 상에서 번식할 수 없었다. pUWHBsAg-CAT 플라스미드로 형질 전환하여, 60㎎/L 클로람페니콜을 포함하는 선택 배양 배지 중에서 번식한 두날리엘라 살리나의 형질 전환체를 제조하였다. 두날리엘라 살리나 내 CAT 유전자 생성물의 발현을 통하여, 60㎎/L 클로람페니콜의 존재 하에서 상기 균주는 번식할 수 있었는데, 이로써 두날리엘라 살리나 내에서 클로람페니콜 내성 카세트의 선택가능한 마커로서의 유용성이 확립되었다. HBsAg 유전자 생성물의 검출 가능한 활성은, ubi1 프로모터와 nos 3'UTR/종결 인자가 두날리엘라 살리나 내 유전자 발현을 가능하게 하는데 적당하다는 것을 암시하였다. 안정적인 형질 전환체의 게놈 DNA는 서던 블럿 분석법에 의해 평가하였다. 겡은, 형질 전환된 두날리엘라 살리나의 선택 및 유지는 존슨 배지(Johnson뭩 medium)(J1, 문헌[Borowitzka 및 Borowitzka (eds), Micro-algal Biotechnology. Cambridge University Press, Cambridge, pp. 460-461]에 개시됨)와 60㎎/L 클로람페니콜을 포함하는 한천 평판 상에서 수행되었다고 개시하였다. 두날리엘라 살리나의 액체 중 번식은 겡에 의하여 J1 배지(60㎎/L 클로람페니콜 함유) 중에서 수행되었다. J1 배지 이외의 배지 중에서의 두날리엘라 살리나의 번식에 대해서는 문헌[Feng et al., Mol. Bio. Reports, Vol. 36 (2009), pp.1433-1439] 및 [Borowitzka et al., Hydrobiologia, Vols. 116-117:1 (1984), pp. 115-121]에 개시되어 있다. 두날리엘라 살리나 내에서 이종 유전자 발현을 가능하게 하는데 적당한 부가의 플라스미드, 프로모터 3'UTR/종결 인자, 그리고 선택가능한 마커는 Feng et al.에 의해 개시된 바 있다. 겡은 플라스미드 pUWHBsAg-CAT, ubi1 프로모터 및 아그로박테리움 튜미파시엔스 노팔린 합성 효소 유전자 3'UTR/종결 인자는 두날리엘라 살리나 내에서 외인성 유전자 발현을 가능하게 하는데 적당하다고 개시하였다. 뿐만 아니라, 겡은 pUWHBsAg-CAT 상에 암호화되는 CAT 내성 카세트는 두날리엘라 살리나 내 선택가능한 마커로서 사용하기 적당하였다고 개시하였다.

[0923] 본 발명의 하나의 구체예에서, 선택가능한 마커로서 사용되는 CAT 유전자 생성물을 암호화하는 뉴클레오티드 서열을 포함하는 벡터 pUWHBsAg-CAT을 구성되고, 이것이 지질 생합성 경로 발현 카세트 서열을 추가로 포함하도록 변형하여, 형질 전환 벡터를 제조하였다. 지질 생합성 경로 발현 카세트는 표 70으로부터 선택되는 하나 이상의 지질 생합성 경로 단백질을 암호화하였는데, 여기서, 각각의 단백질 암호화 서열은 표 69에 따라서 두날리엘라 살리나의 핵내 유전자에 고유한 코돈 편향 사용을 반영하도록 두날리엘라 살리나 내 발현에 대해 코돈 최적화되었다. 표 70의 지질 생합성 경로 단백질 각각에 있어서, 코돈 최적화 유전자 서열은 각각, 단백질 암호화 서열 상류의 ubi1 프로모터에 작동 가능하도록 결합될 수 있으며, 단백질 암호화 서열의 하류 또는 3' 영역에서 아그로박테리움 튜미파시엔스 노팔린 합성 효소 유전자 3'UTR/종결 인자에 작동 가능하도록 결합될 수 있었다. 형질

전환 구조물은 형질 전환 벡터의 표적화 게놈 통합에 사용되는, 두날리엘라 살리나 게놈에 상동성인 영역을 추가로 포함할 수 있었다. 상동성 영역은 내인성 지질 생합성 경로 유전자의 하나 이상의 게놈 위치를 중단하는 것으로 선택할 수 있었다. 널리 공지된 형질 전환 기술, 예를 들어 전기 천공법 또는 기타 다른 공지된 방법들을 사용하여 형질 전환 벡터로 두날리엘라 살리나를 안정적으로 형질 전환하였다. CAT 유전자 생성물의 활성은, 예를 들어 클로람페니콜을 포함하는 J1 배지 중에서 형질 전환 벡터로 형질 전환된 두날리엘라 살리나를 선택하는 선택가능한 마커로서 사용될 수 있었다. 두날리엘라 살리나 지질 생산에 적당한 생장 배지는 J1 배지와 Feng et al.와 Borowitzka et al.에 의해 개시된 배양 배지들을 포함하나, 이에 제한되는 것은 아니다. 두날리엘라 살리나 지질의 지방산 프로필은 표준적인 지질 추출법과 본원에 기술된 분석 방법을 통하여 평가될 수 있었다.

[0924] **실시예 31: 고니움 펙토랄의 조작**

[0925] 본 발명에 따라서 고니움 펙토랄 내에서 재조합 유전자를 발현하는 것은, 본원에 기술된 바와 같이 Lerche(레르헤) 및 Hollman et al.에 의해 교시된 바와 같은 방법과 벡터를 변형하여 수행될 수 있다. 간단히 말해서, 문헌[Lerche 및 Hallman et al., BMC Biotechnology, Volume 9:64, 2009]에는 고니움 펙토랄레를 플라스미드 DNA로 안정적으로 핵내 형질 전환하는 것에 대해 개시되어 있다. 레르헤는 미세 입자 투사법과 같은 형질 전환법을 사용하여 플라스미드 pPmr3을 고니움 펙토랄레에 도입하였다. 플라스미드 pPmr3는 스트렙토마이세스 리모서스의 아미노글리코시드 3'-포스포트랜스퍼라제(aphVIII) 유전자 생성물을 암호화하는 서열(GenBank 수탁 번호 제AAB03856호)(항생제 파로모마이신에 대한 내성 부여)을 포함하며, aphVIII 단백질 암호화 영역 상류의 볼복스 카르테리 hsp70A-rbcS3 하이브리드 프로모터에 작동 가능하도록 결합되고, aphVIII 단백질 암호화 영역 하류의 볼복스 카르테리 rbcS3 유전자의 3'UTR/종결 인자에 작동 가능하도록 결합된 파로모마이신 내성 카세트를 포함하였다. pPmr3으로 형질 전환하기 전, 고니움 펙토랄레는 0.06ug/㎖ 파로모마이신을 포함하는 배지에서 번식할 수 없었다. pPmr3으로 형질 전환하여, 0.75ug/㎖ 이상의 파로모마이신을 포함하는 선택 배양 배지 중에서 번식될 수 있었던 고니움 펙토랄레의 형질 전환체를 제조하였다. 고니움 펙토랄레 내 aphVIII 유전자 생성물이 발현됨으로 인하여 0.75ug/㎖ 이상의 파로모마이신의 존재 하에서 번식될 수 있었는데, 이로써 파로모마이신 항생제 내성 카세트의 고니움 펙토랄레 내에서 사용되는 선택가능한 마커로서의 유용성이 확립되었다. 안정적인 형질 전환체의 게놈 DNA는 서던 블럿 분석법으로 평가하였다. 레르헤 및 홀맨은, 형질 전환된 고니움 펙토랄레의 선택 및 유지는, 1.0ug/㎖ 파로모마이신을 함유하는 액체 자보르스키 배지(Jaworski뭩 medium)(20mg/L Ca(NO3)2·4H2O, 12.4mg/L KH2PO4, 50mg/L MgSO4·7H2O, 15.9mg/L NaHCO3, 2.25mg/L EDTA-FeNa, 2.25mg/L EDTA Na2, 2.48g/L H3BO3, 1.39g/L MnCl2.4H2O, 1mg/L (NH4)6MO7O24.4H2O, 0.04mg/L 비타민 B12, 0.04mg/L 티아민-HCl, 0.04mg/L 바이오틴, 80mg/L NaNO3, 36mg/L Na4HPO4.12H2O) 중에서 수행되었다고 개시하였다. 고니움 펙토랄레 내에서 이종 유전자 발현을 가능하게 만드는데 적당한 부가의 플라스미드, 프로모터, 3'UTR/종결 인자 및 선택가능한 마커는 레르헤 및 홀맨에 의해 추가로 기술되었다. 레르헤 및 홀맨은, 플라스미드 pPmr3, 볼복스 카르테리 hsp70A-rbcS3 하이브리드 프로모터, 그리고 볼복스 카르테리 rbcS3 유전자의 3'UTR/종결 인자는 고니움 펙토랄레 내 외인성 유전자 발현을 가능하게 만드는데 적당하다고 개시하였다. 뿐만 아니라 레르헤 및 홀맨은, 파로모마이신 내성 카세트 암호화 pPmr3은 고니움 펙토랄레 내 선택가능한 마커로서 사용하기 적당하였다고 개시하였다.

[0926] 본 발명의 하나의 구체예에서, 선택가능한 마커로서 사용되는 aphVIII 유전자 생성물을 암호화하는 뉴클레오티드 서열을 포함하는 벡터 pPmr3을 구성하고, 지질 생합성 경로 발현 카세트 서열을 추가로 포함하도록 변형하여, 형질 전환 벡터를 제조하였다. 지질 생합성 경로 발현 카세트는 표 70으로부터 선택되는 하나 이상의 지질 생합성 경로 단백질을 암호화하였는데, 여기서, 각각의 단백질 암호화 서열은 표 69에 의하여 고니움 펙토랄레의 핵내 유전자에 고유한 고돈 편향 사용을 반영하도록 고니움 펙토랄레 내 발현에 대해서 코돈 최적화되었다. 표 70의 지질 생합성 경로 단백질 각각에 있어서, 코돈 최적화 유전자 서열은 각각 단백질 암호화 서열 상류의 볼복스 카르테리 hsp70A-rbcS3 하이브리드 프로모터에 작동 가능하도록 결합될 수 있었으며, 또한 단백질 암호화 서열의 하류 또는 3' 영역에서 볼복스 카르테리 rbcS3 유전자 3'UTR/종결 인자에 작동 가능하도록 결합될 수 있었다. 형질 전환 구조물은 형질 전환 벡터의 표적화 게놈 통합에 사용되는 고니움 펙토랄레 게놈에 상동성인 영역을 추가로 포함할 수 있었다. 상동성 영역은 내인성 지질 생합성 경로 유전자의 하나 이상의 게놈 위치를 중단하는 것으로 선택할 수 있었다. 널리 공지된 형질 전환 기술, 예를 들어 미세 입자 투사법 또는 기타 다른 공지된 방법을 통하여 형질 전환 벡터로 고니움 펙토랄레를 안정적으로 형질 전환할 수 있었다. aphVIII 유전자 생성물의 활성은, 예를 들어 파로모마이신을 포함하는 자보르스키 배지 중에서 형질 전환 벡터로 형질 전환된 고니움 펙토랄레를 선택하는 선택가능한 마커로서 사용될 수 있었다. 고니움 펙토랄레 지질 생산에 적당한 생장 배지는 자보르스키 배지와 문헌[Stein, American Journal of Botany, Vol. 45:9 (1958), pp. 664-672]에 개시된 배지를 포함하였다. 고니움 펙토랄레 지질의 지방산 프로필은 표준적인 지질 추출법과 본원

에 기술된 분석 방법들을 통하여 평가될 수 있었다.

[0927] **실시예 32: 파에오닥틸룸 트리코르누툼의 조작**

[0928] 본 발명에 따라서, 재조합 유전자의 발현은 본원에 개시된 바와 같이 Apt(앱트) et al.에 의해 교시된 방법 및 벡터를 변형함으로써 파에오닥틸룸 트리코르누툼 내에서 이루어질 수 있었다. 간단히 말해서, 문헌[Apt et al., Molecular and General Genetics, Vol. 252 (1996), pp. 572-579]에는 파에오닥틸룸 트리코르누툼을 벡터 DNA로 안정적으로 핵내 형질 전환하는 것에 대해 개시되어 있다. 앱트는 미세 입자 투사법과 같은 형질 전환 기술을 사용하여 파에오닥틸룸 트리코르누툼에 플라스미드 pfcpA를 도입하였다. 플라스미드 pfcpA는 항생제 플레오마이신과 제오신에 대한 내성을 나타내는, 스트렙토알로테이쿠스 힌두스타너스 블레오마이신 결합 단백질(ble)을 암호화하는 서열을 포함하고, ble 단백질 암호화 영역 상류의 파에오닥틸룸 트리코르누툼 후코잔틴 엽록소 a 결합 단백질 유전자(fcpA)의 프로모터에 작동 가능하도록 결합될 수 있었으며, 또한 ble 단백질 암호화 영역의 하류 또는 3' 영역에서 파에오닥틸룸 트리코르누툼 fcpA 유전자의 3'UTR/종결 인자에 작동 가능하도록 결합될 수 있는 블레오마이신 내성 카세트를 포함하였다. pfcpA로 형질 전환하기 전에, 파에오닥틸룸 트리코르누툼은 50ug/㎖ 제오신을 포함하는 배지에서 번식할 수 없었다. pfcpA로 형질 전환하여, 50ug/㎖ 제오신을 포함하는 선택 배양 배지에서 번식할 수 있는 파에오닥틸룸 트리코르누툼의 형질 전환체를 제조하였다. 파에오닥틸룸 트리코르누툼 내 ble 유전자 생성물이 발현됨으로써, 50ug/㎖ 제오신의 존재 하에서 번식할 수 있었으며, 이로써 파에오닥틸룸 트리코르누툼 내에서 사용되는 선택가능한 마커로서 블레오마이신 항생제 내성 카세트의 유용성이 확립되었다. 안정적인 형질 전환체의 게놈 DNA는 서던 블럿 분석법에 의해 평가되었다. 앱트는, 형질 전환된 파에오닥틸룸 트리코르누툼의 선택 및 유지는 LDM 배지(문헌 [Starr and Zeikus, Journal of Phycology, Vol. 29, Supplement, (1993)]에 개시됨)와 50mg/L 제오신을 함유하는 한천 평판상에서 수행되었다고 개시하였다. 앱트는, 50mg/L 제오신을 함유하는 LDM 배지 중 파에오닥틸룸 트리코르누툼 형질 전환체의 액체 중 번식에 대해 개시하였다. LDM 배지 이외의 배지 중에서 파에오닥틸룸 트리코르누툼을 번식시키는 것에 대해서는 문헌 [Zaslavskaia et al., Science, Vol. 292 (2001), pp. 2073-2075], 및 [Radokovits et al., Metabolic Engineering, Vol. 13 (2011), pp. 89-95]에 기술되어 있다. 파에오닥틸룸 트리코르누툼 내에서의 이종 유전자 발현을 가능하게 만드는데 적당한 부가 플라스미드, 프로모터, 3'UTR/종결 인자 및 선택가능한 마커에 대해서는 Apt et al., Zaslavskaia et al. 및 Radokovits et al.에 의해 개시된 바 있다. 앱트는, 플라스미드 pfcpA와, 파에오닥틸룸 트리코르누툼 fcpA 프로모터, 그리고 3'UTR/종결 인자는 파에오닥틸룸 트리코르누툼 내에서 외인성 유전자 발현을 가능하게 하는데 적당하다고 개시하였다. 뿐만 아니라, 앱트는, pfcpA상에서 암호화되는 블레오마이신 내성 카세트는 파에오닥틸룸 트리코르누툼 내 선택가능한 마커로서 사용하기 적당하였다고 개시하였다.

[0929] 본 발명의 하나의 구체예에서, 선택가능한 마커로서 사용되는 ble 유전자 생성물을 암호화하는 뉴클레오티드 서열을 포함하는 벡터 pfcpA를 구성하고, 지질 생합성 경로 발현 카세트 서열을 추가로 포함하도록 변형하여, 형질 전환 벡터를 생산하였다. 지질 생합성 경로 발현 카세트는 표 70으로부터 선택되는 하나 이상의 지질 생합성 경로 단백질을 암호화하였는데, 여기서, 단백질 암호화 서열은 각각 표 69에 따라서 파에오닥틸룸 트리코르누툼의 핵내 유전자에 고유한 코돈 편향 사용을 반영하도록, 파에오닥틸룸 트리코르누툼 내 발현에 대해 코돈 최적화되었다. 표 70의 지질 생합성 경로 단백질 각각에 있어서, 코돈 최적화된 유전자 서열은 각각 단백질 암호화 서열 상류의 파에오닥틸룸 트리코르누툼 fcpA 유전자 프로모터에 작동 가능하도록 결합될 수 있었으며, 또한 단백질 암호화 서열의 하류 또는 3' 영역에서 파에오닥틸룸 트리코르누툼 fcpA 유전자 3'UTR/종결 인자에 작동 가능하도록 결합될 수 있었다. 형질 전환 구조물은 형질 전환 벡터의 표적화 게놈 통합에 사용되는 파에오닥틸룸 트리코르누툼 게놈에 상동성인 영역을 추가로 포함할 수 있었다. 상동성 영역은 내인성 지질 생합성 경로 유전자의 하나 이상의 게놈 위치를 중단하는 것으로 선택할 수 있었다. 당업자는 파에오닥틸룸 트리코르누툼 게놈의 서열 내에 존재하는 이와 같은 상동성 영역들을 동정할 수 있다(문헌[Bowler et al., Nature, Vol. 456 (2008), pp. 239-244]에 개시된 바 참조). 파에오닥틸룸 트리코르누툼을 형질 전환 벡터로 안정적으로 형질 전환하는 것은 널리 공지된 형질 전환 기술, 예를 들어 미세 입자 투사법 또는 기타 다른 공지된 방법을 통하여 이루어질 수 있었다. ble 유전자 생성물의 활성은, 예를 들어 파로모마이신을 포함하는 LDM 배지 중에서 형질 전환 벡터로 형질 전환된 파에오닥틸룸 트리코르누툼을 선택하는 마커로서 사용될 수 있었다. 파에오닥틸룸 트리코르누툼 지질 생산에 적당한 생장 배지는 Radokovits et al.에 의해 개시된 f/2 배지를 포함하나, 이에 제한되는 것은 아니다. 파에오닥틸룸 트리코르누툼 지질의 지방산 프로필은 표준적인 지질 추출법과 본원에 개시된 분석 방법을 통하여 평가될 수 있었다.

[0930] **실시예 33: 카에토세로스 종의 조작**

[0931] 본 발명에 따라서, 카에토세로스 종 내 재조합 유전자의 발현은 본원에 기술된 Yamaguchi(야마구치) et al.에 의해 교시된 방법과 벡터를 변형하여 이루어질 수 있었다. 간단히 말해서, 문헌[Yamaguchi et al., Phycological Research, Vol. 59:2 (2011), pp.113-119]에는 카에토세로스 종을 플라스미드 DNA로 핵내 안정적으로 형질 전환하는 것에 대해 개시되어 있다. 야마구치는 미세 입자 투사법과 같은 형질 전환법을 사용하여 플라스미드 pTpfcp/nat를 카에토세로스 종에 도입하였다. pTpfcp/nat는, nat 단백질 암호화 영역 상류의 탈라시오시라 슈도나나 후코잔틴 엽록소 a/c 결합 단백질 유전자(fcp) 프로모터에 작동 가능하도록 결합되고, 또한 3' 영역(nat 단백질 암호화 서열의 하류)에서 탈라시오시라 슈도나나 fcp 유전자 3'UTR/종결 인자에 작동 가능하도록 결합된, 누르세오트리신 아세틸트랜스퍼라제(nat) 유전자 생성물을 암호화하는 서열(GenBank 수탁 번호 제 AAC60439호)을 포함하는 누르세오트리신 내성 카세트를 포함하였다. nat 유전자 생성물은 항생제 누르세오트리신에 대한 내성을 부여한다. pTpfcp/nat으로 형질 전환하기 전, 카에토세로스 종은 500ug/㎖ 누르세오트리신을 포함하는 배지 상에서 번식될 수 없었다. pTpfcp/nat으로 형질 전환하여, 500ug/㎖ 누르세오트리신을 포함하는 선택 배양 배지에서 번식한 카에토세로스 종의 형질 전환체를 제조하였다. 카에토세로스 종 내에서 nat 유전자 생성물이 발현되면, 500ug/㎖ 누르세오트리신의 존재 하에 번식될 수 있었는데, 이로써 카에토세로스 종 내에서 사용되는 선택가능한 마커로서의 누르세오트리신 항생제 내성 카세트의 유용성이 확립되었다. 안정적인 형질 전환체의 게놈 DNA는 서던 블럿 분석법으로 평가하였다. 야마구치는, 형질 전환된 카에토세로스 종의 선택과 유지는 f/2 배지(문헌[Guilard, R.R., Culture of Phytoplankton for feeding marine invertebrates, In Culture of Marine Invertebrate Animals, Smith and Chanley (eds) 1975, Plenum Press, New York, pp. 26-60]에 개시됨)와 500ug/㎖ 누르세오트리신을 포함하는 한천 평판 상에서 수행되었다고 개시하였다. 야마구치에 의해 수행된 카에토세로스 종 형질 전환체의 액체 중 번식은 500ug/㎖ 누르세오트리신을 함유하는 f/2 배지 중에서 수행되었다. 부가의 배양 배지 중 카에토세로스 종의 번식에 대해서는, 예를 들어 문헌[Napolitano et al., Journal of the World Aquaculture Society, Vol. 21:2 (1990), pp. 122-130], 및 [Volkman et al., Journal of Experimental Marine Biology and Ecology, Vol. 128:3 (1989), pp. 219-240]에 개시되어 있다. 카에토세로스 종 내 이종 유전자 발현을 가능하게 하는데 적당한 부가의 플라스미드, 프로모터, 3'UTR/종결 인자, 그리고 선택가능한 마커는 Yamaguchi et al.에 의한 동일한 문헌에 개시되었다. 야마구치는 플라스미드 pTpfcp/nat, 탈라시오시라 슈도나나 fcp 프로모터, 그리고 3'UTR/종결 인자가, 카에토세로스 종 내 외인성 유전자 발현을 가능하게 하는데 적당하다고 개시하였다. 뿐만 아니라, 야마구치는, pTpfcp/nat 상에서 암호화되는 누르세오트리신 내성 카세트가 카에토세로스 종 내 선택가능한 마커로서 사용되기 적당하였다고 개시하였다.

[0932] 본 발명의 하나의 구체예에서, 선택가능한 마커로서 사용되는 nat 유전자 생성물을 암호화하는 뉴클레오티드 서열을 포함하는 벡터 pTpfcp/nat를 구성하고, 지질 생합성 경로 발현 카세트 서열을 추가로 포함하도록 변형하여, 형질 전환 벡터를 생산하였다. 지질 생합성 경로 발현 카세트는 표 70으로부터 선택되는 하나 이상의 지질 생합성 경로 단백질을 암호화하였는데, 이 경우, 각각의 단백질 암호화 서열은, 표 69에 따라서 카에토세로스 컴프레섬의 핵내 유전자에 고유한 코돈 편향 사용을 반영하도록, 밀접하게 관련된 카에토세로스 컴프레섬 내 발현에 대해 코돈 최적화되었다. 표 70의 지질 생합성 경로 단백질 각각에 있어서, 코돈 최적화 유전자 서열은 각각 단백질 암호화 서열 상류의 탈라시오시라 슈도나나 fcp 유전자 프로모터에 작동 가능하도록 결합될 수 있었으며, 단백질 암호화 서열의 하류 또는 3' 영역에서 탈라시오시라 슈도나나 fcp 유전자 3'UTR/종결 인자와 작동 가능하도록 결합될 수 있었다. 형질 전환 구조물은 형질 전환 벡터의 표적화 게놈 통합에 사용되는, 카에토세로스 종 게놈에 대한 상동성 영역을 추가로 포함할 수 있었다. 상동성 영역은 내인성 지질 생합성 경로 유전자의 하나 이상의 게놈 위치를 중단하는 것으로 선택할 수 있었다. 카에토세로스 종을 형질 전환 벡터로 안정적으로 형질 전환하는 것은 널리 공지된 형질 전환법, 예를 들어 미세 입자 투사법 또는 기타 다른 공지된 방법을 통하여 이루어졌다. nat 유전자 생성물의 활성은, 예를 들어 누르세오트리신을 포함하는 f/2 한천 배지 중에서 형질 전환 벡터로 형질 전환된 카에토세로스 종을 선택하는 선택가능한 마커로서 사용될 수 있었다. 카에토세로스 종 지질 생산에 적당한 생산 배지는 f/2 배지와 Napolitano et al. 및 Volkman et al.에 의해 개시된 배양 배지를 포함하나, 이에 제한되는 것은 아니다. 카에토세로스 종 지질의 지방산 프로필은 표준적인 지질 추출법과 본원에 기술된 분석 방법을 통하여 평가될 수 있었다.

[0933] **실시예 34: 실린드로테카 후지포르미스의 조작**

[0934] 본 발명에 따라서 재조합 유전자를 실린드로테카 후지포르미스 내에서 발현하는 것은, 본원에 기술된 Poulsen(폴센) 및 Kroger(크로거) et al.에 의해 교시된 방법과 벡터를 변형함으로써 이루어질 수 있었다. 간단히 말해서, 문헌[Poulsen 및 Kroger et al., FEBS Journal, Vol. 272 (2005), pp.3413-3423]에는 실린드로테카 후지포르미스를 플라스미드 DNA로 형질 전환하는 것에 대해 개시되어 있다. 폴센과 크로거는 미세 입자 투사법과 같은

형질 전환법을 사용하여 pCF-ble 플라스미드를 실린드로테카 후지포르미스에 도입하였다. 플라스미드 pCF-ble은, 항생제 제오신 및 플레오마이신에 대한 내성을 부여하는, 스트렙토알로테이쿠스 힌두스타너스 블레오마이신 결합 단백질(ble)을 암호화하는 서열을 포함하고, ble 단백질 암호화 영역 상류의 실린드로테카 후지포르미스 후코잔틴 엽록소 a/c 결합 단백질 유전자(fcpA, GenBank 수탁 번호 제AY125580호) 프로모터에 작동 가능하도록 결합되었고, 3' 영역(ble 단백질 암호화 영역의 하류)에서 실린드로테카 후지포르미스 fcpA 유전자 3'UTR/종결 인자에 작동 가능하도록 결합된, 블레오마이신 내성 카세트를 포함하였다. pCF-ble로 형질 전환하기 전, 실린드로테카 후지포르미스는 1mg/㎖ 제오신을 포함하는 배지 상에서 번식할 수 없었다. pCF-ble로 형질 전환하여, 1mg/㎖ 제오신을 포함하는 선택 배양 배지 중에서 번식한 실린드로테카 후지포르미스 형질 전환체를 제조하였다. 실린드로테카 후지포르미스 내 ble 유전자 생성물의 발현으로 인하여 1mg/㎖ 제오신의 존재 하에서 번식될 수 있었는데, 이로써 실린드로테카 후지포르미스 내에서 사용되는 선택가능한 마커로서의 블레오마이신 항생제 내성 카세트의 유용성이 확립되었다. 폴센 및 크로거는, 형질 전환된 실린드로테카 후지포르미스의 선택 및 유지는 1mg/㎖ 제오신과 인공 해양수 배지를 포함하는 한천 평판 상에서 수행되었다고 개시하였다. 폴센 및 크로거는, 1mg/㎖ 제오신을 포함하는 인공 해양수 배지 중 실린드로테카 후지포르미스 형질 전환체의 액체 중 번식에 대해 개시하였다. 부가의 배양 배지 중 실린드로테카 후지포르미스의 번식에 대해서는, 예를 들어 문헌 [Liang et al., Journal of Applied Phycology, Vol. 17:1 (2005), pp. 61-65], 및 [Orcutt 및 Patterson, Lipids, Vol. 9:12 (1974), pp. 1000-1003]에 기술되어 있다. 카에토세로스 종 내 이종 유전자 발현을 가능하게 만드는 부가 플라스미드, 프로모터 및 3'UTR/종결 인자는 폴센 및 크로거에 의한 동일 문헌에 개시되었다. 폴센 및 크로거는, 플라스미드 pCF-ble, 실린드로테카 후지포르미스 fcp 프로모터 및 3'UTR/종결 인자는 실린드로테카 후지포르미스 내 외인성 유전자 발현을 가능하게 하는데 적당하다고 개시하였다. 뿐만 아니라, 폴센 및 크로거는, pCF-ble 상에 암호화된 블레오마이신 내성 카세트는 실린드로테카 후지포르미스 내 선택가능한 마커로서 사용하기 적당하였다고 개시하였다.

[0935] 본 발명의 하나의 구체예에서, 선택가능한 마커로서 사용되는 ble 유전자 생성물 암호화 뉴클레오티드 서열을 포함하는 벡터 pCF-ble를 구성하고, 지질 생합성 경로 발현 카세트 서열을 추가로 포함하도록 변형하여, 형질 전환 벡터를 생산하였다. 지질 생합성 경로 발현 카세트는 표 70으로부터 선택된 하나 이상의 지질 생합성 경로 단백질을 암호화하였는데, 여기서, 각각의 단백질 암호화 서열은, 표 69에 따른 실린드로테카 후지포르미스의 핵내 유전자에 고유한 코돈 편향 사용을 반영하도록 실린드로테카 후지포르미스 내 발현에 대해 코돈 최적화되었다. 표 70의 지질 생합성 경로 단백질 각각에 있어서, 코돈 최적화된 유전자 서열은 각각, 단백질 암호화 서열 상류의 실린드로테카 후지포르미스 fcp 유전자 프로모터에 작동 가능하도록 결합될 수 있었고, 단백질 암호화 서열의 하류 또는 3' 영역에서 실린드로테카 후지포르미스 fcp 유전자 3'UTR/종결 인자에 작동 가능하도록 결합될 수 있었다. 형질 전환 구조물은 형질 전환 벡터의 표적화 게놈 통합에 사용되는 실린드로테카 후지포르미스 게놈에 상동성인 영역을 추가로 포함할 수 있었다. 상동성 영역은 내인성 지질 생합성 경로 유전자의 하나 이상의 게놈 위치를 중단하는 것으로 선택할 수 있었다. 형질 전환 벡터를 사용하는 실린드로테카 후지포르미스의 안정적인 형질 전환은 널리 공지된 형질 전환 기술, 예를 들어 미세 입자 투사법 또는 기타 다른 공지된 방법을 통하여 이루어졌다. Ble 유전자 생성물의 활성은, 예를 들어 제오신을 포함하는 인공 해양수 한천 배지 중 형질 전환 벡터로 형질 전환된 실린드로테카 후지포르미스를 선택하는 선택가능한 마커로서 사용될 수 있었다. 실린드로테카 후지포르미스 지질 생산에 적당한 생장 배지는 Liang et al. 및 Orcutt 및 Patterson에 의해 개시된 배지와 인공 해양수를 포함하나, 이에 제한되는 것은 아니다. 실린드로테카 후지포르미스 지질의 지방산 프로필은 표준적인 지질 추출법과 본원에 기술된 분석 방법을 통하여 평가될 수 있었다.

[0936] **실시예 35: 암피디니움 종의 조작**

[0937] 본 발명에 따라서 암피디니움 종 내 재조합 유전자의 발현은 본원에 기술된 ten Lohuis(텐 로휘스) 및 Miller et al.에 의해 교시된 방법과 벡터를 변형하여 이루어질 수 있었다. 간단히 말해서, 문헌[ten Lohuis 및 Miller et al., The Plant Journal, Vol. 13:3 (1998), pp. 427-435]에는, 플라스미드 DNA로 암피디니움 종을 안정적으로 형질 전환하는 것에 대해 개시되어 있다. 텐 로휘스는 탄화 규소 위스커의 존재 하에서 교반과 같은 형질 전환 기술을 사용하여 플라스미드 pMT NPT/GUS를 암피디니움 종에 도입하였다. pMT NPT/GUS는, nptII 단백질 암호화 영역 상류의 5' 또는 아그로박테리움 튜미파시엔스 노팔린 합성 효소(nos) 유전자 프로모터에 작동 가능하도록 결합되었고, 3' 영역(nptII 단백질 암호화 영역의 하류)에서 nos 유전자의 3'UTR/종결 인자에 작동 가능하도록 결합된, 네오마이신 포스포트랜스퍼라제 II(nptII) 유전자 생성물을 암호화하는 서열(GenBank 수탁 번호 제AAL92039호)을 포함하는, 네오마이신 내성 카세트를 포함하였다. nptII 유전자 생성물은 항생제 G418에 대한 내성을 부여하였다. pMT NPT/GUS 플라스미드는 CaMV 35S 프로모터에 작동 가능하도록 결합되었으며, 또한 CaMV 35S 3'UTR/종결 인자에도 작동 가능하도록 결합된 베타-글루쿠로니다제(GUS) 리포터 유전자 생성물을 암호화하

는 서열을 추가로 포함하였다. pMT NPT/GUS로 형질 전환하기 전, 암피디니움 종은 3㎎/㎖ G418을 포함하는 배지에서 번식할 수 없었다. pMT NPT/GUS로 형질 전환하여, 3㎎/㎖ G418을 포함하는 선택 배양 배지 중에서 번식한 암피디니움 종의 형질 전환체를 제조하였다. nptII 유전자 생성물의 암피디니움 종 내 발현을 통하여, 3㎎/㎖ G418의 존재 하에 번식이 가능하게 되었는데, 이로써 암피디니움 종 내에서 사용되는 선택가능한 마커로서의 네오마이신 항생제 내성 카세트의 유용성이 확립되었다. GUS 리포터 유전자의 검출 가능한 활성은, CaMV 35S 프로모터 및 3'UTR은 암피디니움 종 내 유전자 발현을 가능하게 하는데 적당하다는 것을 암시하는 것이다. 안정적인 형질 전환체의 게놈 DNA는 서던 블럿 분석법에 의해 평가하였다. 텐 로휘스와 밀러는, F/2 풍부 용액(시그마(Sigma)에 의해 공급) 및 3㎎/㎖ G418이 보충된 해양수 포함 배지 중 암피디니움 종 형질 전환체의 액체 중 번식과, F/2 풍부 용액 및 3㎎/㎖ G418이 보충된 해양수 포함 한천 배지에서의 암피디니움 종 형질 전환체의 선택 및 유지에 대해 개시하였다. 부가의 배양 배지 중 암피디니움 종의 번식은, 예를 들어 문헌[Mansour et al., Journal of Applied Phycology, Vol. 17:4 (2005) pp. 287-v300]에 개시되어 있다. 암피디니움 종 내에서 이종 유전자 발현을 가능하게 하는 부가의 프로모터, 3'UTR/종결 인자 및 선택가능한 마커를 포함하는 부가의 플라스미드는 텐 로휘스 및 밀러에 의한 동일 문헌에 개시되어 있다. 텐 로휘스 및 밀러는, 플라스미드 pMT NPT/GUS, 프로모터 그리고 nos 및 CaMV 35S 유전자의 3'UTR/종결 인자가 암피디니움 종 내 외인성 유전자 발현을 가능하게 하는데 적당하다고 개시하였다. 뿐만 아니라, 텐 로휘스 및 밀러는 pMT NPT/GUS 상에서 암호화된 네오마이신 내성 카세트가 암피디니움 종 내 선택가능한 마커로서 사용하기 적당하였다고 개시하였다.

[0938] 본 발명의 하나의 구체예에서, 선택가능한 마커로 사용되는 nptII 유전자 생성물 암호화 뉴클레오티드 서열을 포함하는 벡터 pMT NPT/GUS를 구성하고, 지질 생합성 경로 발현 카세트 서열을 추가로 포함하도록 변형하여, 형질 전환 벡터를 생산하였다. 지질 생합성 경로 발현 카세트는 표 70으로부터 선택되는 하나 이상의 지질 생합성 경로 단백질을 암호화하였는데, 이 경우, 각각의 단백질 암호화 서열은, 표 69에 따라서 상기 균주와 밀접한 관련이 있는 종인 암피디니움 카르테라에의 핵내 유전자에 고유한 코돈 편향 사용을 반영하도록, 암피디니움 종 내 발현에 대해서 코돈 최적화되었다. 표 70의 지질 생합성 경로 단백질 각각에 있어서, 코돈 최적화 유전자 서열은 각각, 단백질 암호화 서열 상류의 아그로박테리움 튜미파시엔스 노팔린 합성 효소(nos) 유전자 프로모터에 작동 가능하도록 결합될 수 있었으며, 단백질 암호화 서열의 하류 또는 3' 영역에서 nos 3'UTR/종결 인자에 작동 가능하도록 결합될 수 있었다. 형질 전환 구조물은 형질 전환 벡터의 표적화 게놈 통합에 사용되는 암피디니움 종 게놈에 상동성인 영역들을 추가로 포함할 수 있었다. 상동성 영역은 내인성 지질 생합성 경로 유전자들의 하나 이상의 게놈 위치를 중단하는 것으로 선택할 수 있었다. 형질 전환 벡터를 사용하는 암피디니움 종의 안정적인 형질 전환은 널리 공지된 형질 전환 기술, 예를 들어 규소 섬유 매개 미세 주입법 또는 기타 다른 공지된 방법을 통하여 이루어질 수 있었다. nptII 유전자 생성물의 활성은, 예를 들어 G418을 포함하는 해양수 한천 배지 중 형질 전환 벡터로 형질 전환된 암피디니움 종을 선택하는 선택가능한 마커로서 사용될 수 있었다. 암피디니움 종 지질 생산에 적당한 생장 배지는 Mansour et al. 및 텐 로휘스 및 밀러에 의해 개시된 배지와 인공 해양수를 포함하였으나, 이에 제한되는 것은 아니었다. 암피디니움 종 지질의 지방산 프로필은 표준적인 지질 추출법과 본원에 기술된 분석 방법들을 통하여 평가될 수 있었다.

[0939] **실시예 36: 심비오디늄 마이크로아드리악티큠의 조작**

[0940] 본 발명에 따라서 심비오디늄 마이크로아드리악티큠 내 재조합 유전자의 발현은 본원에 기술된 바와 같이 ten Lohuis 및 Miller et al.에 의해 교시되는 방법과 벡터를 변형함으로써 이루어질 수 있었다. 간단히 말해서, 문헌[ten Lohuis 및 Miller et al., The Plant Journal, Vol. 13:3 (1998), pp. 427-435]에는, 플라스미드 DNA로 심비오디늄 마이크로아드리악티큠을 안정적으로 형질 전환하는 것에 대해 개시되어 있다. 텐 로휘스는, 규소 섬유 매개 미세 주입법과 같은 형질 전환 기술을 사용하여 플라스미드 pMT NPT/GUS를 심비오디늄 마이크로아드리악티큠에 도입하였다. pMT NPT/GUS는, nptII 단백질 암호화 영역의 5'또는 상류의 아그로박테리움 튜미파시엔스 노팔린 합성 효소(nos) 유전자 프로모터에 작동 가능하도록 결합되었고, 3' 영역 nptII 단백질 암호화 영역의 하류)에서 nos 유전자의 3'UTR/종결 인자에 작동 가능하도록 결합된, 네오마이신 포스포트랜스퍼라제 II(nptII) 유전자 생성물을 암호화하는 서열(GenBank 수탁 번호 제AAL92039호)을 포함하는 네오마이신 내성 카세트를 포함하였다. nptII 유전자 생성물은 항생제 G418에 대한 내성을 부여하였다. pMT NPT/GUS 플라스미드는 CaMV 35S 프로모터에 작동 가능하도록 결합되었고, 추가로 CaMV 35S 3'UTR/종결 인자에도 작동 가능하도록 결합된, 베타-글루쿠로니다제(GUS) 리포터 유전자 생성물을 암호화하는 서열을 추가로 포함하였다. pMT NPT/GUS로 형질 전환하기 전, 심비오디늄 마이크로아드리악티큠은 3㎎/㎖ G418을 포함하는 배지에서 번식할 수 없었다. pMT NPT/GUS로 형질 전환하여, 3㎎/㎖ G418을 포함하는 선택 배양 배지 내 번식한 심비오디늄 마이크로아드리악티큠 형질 전환체를 제조하였다. nptII 유전자 생성물의 심비오디늄 마이크로아드리악티큠 내 발현으로 인하여, 3㎎/㎖ G418의 존재 하에 번식될 수 있었으며, 이로써, 심비오디늄 마이크로아드리악티큠 내 사용되는 선택가능

한 마커로서의 네오마이신 항생제 내성 카세트의 유용성이 확립되었다. GUS 리포터 유전자의 검출 가능한 활성은, CaMV 35S 프로모터 및 3'UTR은 심비오디늄 마이크로아드리악티큠 내 유전자 발현을 가능하게 하는데 적당하다는 것을 암시하였다. 안정적인 형질 전환체의 게놈 DNA는 서던 블럿 분석법에 의해 평가되었다. 텐 로휘스 및 밀러는, F/2 풍부 용액(공급자 시그마에 의해 제공) 및 3㎎/㎖ G418이 보충된, 해양수 포함 배지 중 심비오디늄 마이크로아드리악티큠 형질 전환체의 액체 중 번식과, F/2 풍부 용액 및 3㎎/㎖ G418이 보충된, 해양수 포함 한천 배지에서의 심비오디늄 마이크로아드리악티큠 형질 전환체의 선택과 유지에 대해 개시하였다. 부가의 배양 배지 중 심비오디늄 마이크로아드리악티큠의 번식에 대해서는, 예를 들어 문헌[Iglesias-Prieto et al., Proceedings of the National Academy of Sciences, Vol. 89:21 (1992) pp. 10302-10305]에 개시되어 있다. 심비오디늄 마이크로아드리악티큠 내 이종 유전자 발현을 가능하게 하는 부가의 프로모터, 3'UTR/종결 인자 및 선택가능한 마커를 포함하는 부가의 플라스미드에 대해서는 텐 로휘스 및 밀러에 의한 동일한 문헌에 개시되어 있다. 텐 로휘스와 밀러는, 플라스미드 pMT NPT/GUS와, nos의 3'UTR/종결 인자 및 프로모터, 그리고 CaMV 35S 유전자는 심비오디늄 마이크로아드리악티큠 내 외인성 유전자 발현을 가능하게 하는데에 적당하다고 개시하였다. 뿐만 아니라, 텐 로휘스 및 밀러는, pMT NPT/GUS상에 암호화된 네오마이신 내성 카세트가 심비오디늄 마이크로아드리악티큠 내 선택가능한 마커로서 사용하기에 적당하였다고 개시하였다.

[0941] 본 발명의 하나의 구체예에서, 선택가능한 마커로 사용되는 nptII 유전자 생성물 암호화 뉴클레오티드 서열을 포함하는 벡터 pMT NPT/GUS를 구성하고, 지질 생합성 경로 발현 카세트 서열을 추가로 포함하도록 변형하여, 형질 전환 벡터를 생산하였다. 지질 생합성 경로 발현 카세트는 표 70으로부터 선택되는 하나 이상의 지질 생합성 경로 단백질을 암호화하였는데, 이 경우, 각각의 단백질 암호화 서열은, 표 69에 따라서 심비오디늄 마이크로아드리악티큠의 핵내 유전자에 고유한 코돈 편향 사용을 반영하도록, 심비오디늄 마이크로아드리악티큠 내 발현에 대해서 코돈 최적화되었다. 표 70의 지질 생합성 경로 단백질 각각에 있어서, 코돈 최적화 유전자 서열은 각각, 단백질 암호화 서열 상류의 아그로박테리움 튜미파시엔스 노팔린 합성 효소(nos) 유전자 프로모터에 작동 가능하도록 결합될 수 있었으며, 단백질 암호화 서열의 하류 또는 3' 영역에서 nos 3'UTR/종결 인자에 작동 가능하도록 결합될 수 있었다. 형질 전환 구조물은 형질 전환 벡터의 표적화 게놈 통합에 사용되는 심비오디늄 마이크로아드리악티큠 게놈에 상동성인 영역을 추가로 포함할 수 있었다. 상동성 영역은 내인성 지질 생합성 경로 유전자들의 하나 이상의 게놈 위치를 중단하는 것으로 선택할 수 있었다. 형질 전환 벡터를 사용하는 심비오디늄 마이크로아드리악티큠의 안정적인 형질 전환은 널리 공지된 형질 전환 기술, 예를 들어 규소 섬유 매개 미세 주입법 또는 기타 다른 공지된 방법을 통하여 이루어졌다. nptII 유전자 생성물의 활성은, 예를 들어 G418을 포함하는 해양수 한천 배지 중 형질 전환 벡터로 형질 전환된 심비오디늄 마이크로아드리악티큠을 선택하는 선택가능한 마커로서 사용될 수 있었다. 심비오디늄 마이크로아드리악티큠 지질 생산에 적당한 생장 배지는 Iglesias-Prieto et al. 및 텐 로휘스 및 밀러에 의해 개시된 배지와 인공 해양수를 포함하나, 이에 제한되는 것은 아니다. 심비오디늄 마이크로아드리악티큠 지질의 지방산 프로필은 표준적인 지질 추출법과 본원에 기술된 분석 방법들을 통하여 평가될 수 있었다.

[0942] **실시예 37: 나노클로롭시스 종의 조작**

[0943] 본 발명에 따라서 재조합 유전자의 나노클로롭시스 종 W2J3B 내 발현은 본원에 기술된 바와 같이 Kilian(킬리안) et al.에 의해 교시된 방법 및 벡터를 변형하여 이루어질 수 있었다. 간단히 말해서, 문헌[Kilian et al., Proceedings of the National Academy of Sciences, Vol. 108:52 (2011) pp.21265-21269]에는, 형질 전환 구조물을 사용하여 나노클로롭시스를 핵내 안정적으로 형질 전환하는 것에 대해 개시되어 있다. 킬리안은 전기 천공법과 같은 형질 전환법을 사용하여 형질 전환 구조물 C2를 나노클로롭시스 종 W2J3B에 도입하였다. 상기 C2 형질 전환 구조물은, 스트렙토알로테이쿠스 힌두스타너스 블레오마이신 결합 단백질(ble)에 대한 암호화 서열(항생제 플레오마이신과 제오신에 대한 내성 부여)을 포함하고, ble 단백질 암호화 영역 상류의 나노클로롭시스 종 W2J3B 비올라잔틴/엽록소 a 결합 단백질 유전자 VCP2의 프로모터에 작동 가능하도록 결합되어 있으며, ble 단백질 암호화 영역 하류의 나노클로롭시스 종 W2J3B 비올라잔틴/엽록소 a 결합 유전자 VCP1의 3'UTR/종결 인자에 작동 가능하도록 결합된, 블레오마이신 내성 카세트를 포함하였다. C2로 형질 전환하기 전, 나노클로롭시스 종 W2J3B는 2ug/㎖ 제오신을 포함하는 배지에서 번식할 수 없었다. C2로 형질 전환하여, 2ug/㎖ 제오신을 포함하는 선택 배양 배지 중 번식된 나노클로롭시스 종 W2J3B의 형질 전환체를 제조하였다. 나노클로롭시스 종 W2J3B 내 ble 유전자 생성물의 발현을 통하여, 2ug/㎖ 제오신의 존재 하에서 번식이 가능하게 되었는데, 이로써 나노클로롭시스에서 사용되는 선택가능한 마커로서의 블레오마이신 항생제 내성 카세트의 유용성이 확립되었다. 안정적인 형질 전환체의 게놈 DNA는 PCR에 의해 평가되었다. 킬리안은, 미량 금속, 비타민 및 포스페이트 용액을 5배 많은 수준으로 포함하고 2ug/㎖ 제오신을 추가로 함유하는 F/2 배지(문헌[Guilard and Ryther, Canadian Journal of Microbiology, Vol. 8 (1962), pp. 229-239]에 개시됨) 내에서 나노클로롭시스

종 W2J3B 형질 전환체의 액체 중 번식에 대해 개시하였다. 킬리안은 또한, 2㎎/㎖ 제오신과 인공 해양수를 포함하는 한천 F/2 배지에서의 나노클로롭시스 종 W2J3B 형질 전환체의 선택과 유지에 대해서도 개시하였다. 부가 배양 배지 중 나노클로롭시스의 번식에 대해서는, 예를 들어 문헌[Chiu et al., Bioresour Technol., Vol. 100:2 (2009), pp. 833-838] 및 [Pal et al., Applied Microbiology and Biotechnology, Vol. 90:4 (2011), pp. 1429-1441]에 기술되었다. 나노클로롭시스 종 W2J3B 내에서 이종 유전자 발현을 가능하게 하는 부가 프로모터 및 3'UTR/종결 인자와, 형질 전환체의 선택에 사용되는 선택가능한 마커를 포함하는 부가의 형질 전환 구조물은 킬리안에 의한 동일 문헌에 기술되어 있다. 킬리안은, 형질 전환 구조물 C2, 나노클로롭시스 종 W2J3B 비올라잔틴/엽록소 a 결합 단백질 유전자 VCP2의 프로모터, 그리고 나노클로롭시스 종 W2J3B 비올라잔틴/엽록소 a 결합 단백질 유전자 VCP1의 3'UTR/종결 인자는 나노클로롭시스 종 W2J3B 내 외인성 유전자 발현을 가능하게 하는데 적당하다고 개시하였다. 뿐만 아니라, 킬리안은, C2 상에서 암호화되는 블레오마이신 내성 카세트가 나노클로롭시스 종 W2J3B 내 선택가능한 마커로 사용하기 적당하였다고 개시하였다.

[0944] 본 발명의 하나의 구체예에서, 선택가능한 마커로 사용되는 ble 유전자 생성물을 암호화하는 뉴클레오티드 서열을 포함하는 형질 전환 구조물 C2를 구성하고, 지질 생합성 경로 발현 카세트 서열을 추가로 포함하도록 변형하여, 형질 전환 벡터를 생산하였다. 지질 생합성 경로 발현 카세트는 표 70으로부터 선택되는 하나 이상의 지질 생합성 경로 단백질을 암호화하였는데, 각각의 단백질 암호화 서열은, 표 69에 따라서 나노클로롭시스 종의 핵내 유전자에 고유한 코돈 편향 사용을 반영하도록 나노클로롭시스 종 W2J3B 내에서의 발현에 대해 코돈 최적화되었다. 표 70의 지질 생합성 경로 단백질 각각에 있어서, 코돈 최적화 유전자 서열은 각각, 단백질 암호화 서열 상류의 나노클로롭시스 종 W2J3B VCP2 유전자 프로모터에 작동 가능하도록 결합될 수 있었으며, 단백질 암호화 서열의 하류 또는 3' 영역에서 나노클로롭시스 종 W2J3B VCP1 유전자 3'UTR/종결 인자에 작동 가능하도록 결합될 수 있었다. 형질 전환 구조물은, 형질 전환 벡터의 표적화 게놈 통합에 사용되는 나노클로롭시스 종 W2J3B 게놈에 상동성인 영역을 추가로 포함할 수 있었다. 상동성 영역은 내인성 지질 생합성 경로 유전자의 하나 이상의 게놈 위치를 중단하는 것으로 선택할 수 있었다. 형질 전환 벡터로 나노클로롭시스 종 W2J3B를 안정적으로 형질 전환하는 것은, 널리 공지된 형질 전환 기술, 예를 들어 전기 천공법 또는 기타 다른 공지된 방법을 통하여 이루어졌다. ble 유전자 생성물의 활성은, 예를 들어 제오신을 포함하는 F/2 배지 중에서 형질 전환 벡터로 형질 전환된 나노클로롭시스 종 W2J3B를 선택하는 선택가능한 마커로서 사용될 수 있었다. 나노클로롭시스 종 W2J3B 지질 생산에 적당한 생장 배지는 F/2 배지와 Chiu et al.과 Pal et al.에 의해 개시된 배지를 포함하였으나, 이에 제한되는 것은 아니다. 나노클로롭시스 종 W2J3B 지질의 지방산 프로필은 표준적 지질 추출법과 본원에 기술된 분석 방법을 통하여 평가될 수 있었다.

[0945] **실시예 38: 시클로텔라 크립티카의 조작**

[0946] 본 발명에 따라서 재조합 유전자의 발현은 본원에 개시된 바와 같이 Dunahay(듀나헤이) et al.에 의해 교시된 방법 및 벡터를 변형함으로써 시클로텔라 크립티카 내에서 이루어질 수 있었다. 간단히 말해서, 문헌[Dunahay et al., Journal of Phycology, Vol. 31 (1995), pp. 1004-1012]에는 시클로텔라 크립티카를 플라스미드 DNA로 안정적으로 형질 전환하는 것이 개시되어 있다. 듀나헤이는 미세 입자 투사법과 같은 형질 전환법을 사용하여 플라스미드 pACCNPT5.1를 시클로텔라 크립티카에 도입하였다. 플라스미드 pACCNPT5.1은, nptII 암호화 영역 상류의 시클로텔라 크립티카 아세틸-CoA 카복실라제(ACCase) 유전자(GenBank 수탁 번호 제L20784호)의 프로모터에 작동 가능하도록 결합되어 있으며, 또한 3' 영역(nptII 암호화 영역의 하류)에서 시클로텔라 크립티카 ACCase 유전자의 3'UTR/종결 인자에 작동 가능하도록 결합되어 있는, 네오마이신 포스포트랜스퍼라제 II(nptII) 유전자 생성물의 암호화 서열을 포함하는, 네오마이신 내성 카세트를 포함하였다. nptII 유전자 생성물은 항생제 G418에 대한 내성을 부여하였다. pACCNPT5.1로 형질 전환하기 전, 시클로텔라 크립티카는 100ug/㎖ G418을 포함하는 50% 인공 해양수 배지에서 번식할 수 없었다. pACCNPT5.1로 형질 전환하여, 100ug/㎖ G418을 포함하는 50% 인공 해양수 선택 배지에서 번식된 시클로텔라 크립티카 형질 전환체를 제조하였다. 시클로텔라 크립티카 내 nptII 유전자 생성물의 발현으로 인하여 이 균주는 100ug/㎖ G418의 존재 하에서 번식할 수 있었는데, 이로써 시클로텔라 크립티카에서 사용되는 선택가능한 마커로서의 네오마이신 항생제 내성 카세트의 유용성이 확립되었다. 안정적인 형질 전환체의 게놈 DNA는 서던 블럿 분석법에 의해 평가하였다. 듀나헤이는, 1.07㎜ 규산화 나트륨과 100ug/㎖ G418이 보충된 인공 해양수 배지(ASW, 문헌[Brown, L., Phycologia, Vol. 21 (1982), pp. 408-410]에 개시됨) 내 시클로텔라 크립티카의 액체 중 번식에 대해 개시하였다. 듀나헤이는 또한, 100ug/㎖ G418을 함유하는 ASW 배지를 포함하는 한천 평판 상에서의 시클로텔라 크립티카 형질 전환체의 선택 및 유지에 대해서도 개시하였다. 부가의 배양 배지 중에서의 시클로텔라 크립티카의 번식에 대해서는, 예를 들어 문헌[Sriharan et al., Applied Biochemistry and Biotechnology, Vol. 28-29:1 (1991), pp. 317-326] 및 [Pahl et al., Journal of Bioscience and Bioengineering, Vol. 109:3 (2010), pp. 235-239]에 기술되어 있다. 듀나헤이는, 플라스미드

pACCNPT5.1과 시클로텔라 크립티카 아세틸-CoA 카복실라제(ACCase) 유전자의 프로모터는 시클로텔라 크립티카 내에서 외인성 유전자의 발현을 가능하게 하는데 적당하다고 개시하였다. 뿐만 아니라, 듀나헤이는, pACCNPT5.1 상에서 암호화된 네오마이신 내성 카세트가 시클로텔라 크립티카 내 선택가능한 마커로서 사용하기 적당하였다고 개시하였다.

[0947] 본 발명의 하나의 구체예에서, nptII 유전자 생성물을 암호화하는 뉴클레오티드 서열(선택가능한 마커로서 사용)을 포함하는 벡터 pACCNPT5.1을 구성하고, 지질 생합성 경로 발현 카세트 서열을 추가로 포함하도록 변형하여, 형질 전환 벡터를 생산하였다. 지질 생합성 경로 발현 카세트는 표 70으로부터 선택되는 하나 이상의 지질 생합성 경로 단백질을 암호화하였는데, 이때, 각각의 단백질 암호화 서열은, 표 69에 의한 시클로텔라 크립티카의 핵내 유전자에 고유한 코돈 편향 사용을 반영하도록, 시클로텔라 크립티카 내에서의 발현에 대해 코돈 최적화되었다. 표 70의 지질 생합성 경로 단백질 각각에 있어서, 코돈 최적화된 유전자 서열은 각각 단백질 암호화 서열 상류의 시클로텔라 크립티카 ACCase 프로모터에 작동 가능하도록 결합될 수 있었으며, 또한 단백질 암호화 서열의 하류 또는 3' 영역에서 시클로텔라 크립티카 ACCase 3'UTR/종결 인자에 작동 가능하도록 결합될 수 있었다. 형질 전환 구조물은 형질 전환 벡터의 표적화 게놈 통합에 사용되는 시클로텔라 크립티카 게놈에 상동성인 영역을 추가로 포함할 수도 있었다. 상동성 영역은 내인성 지질 생합성 경로 유전자의 하나 이상의 게놈 위치를 중단하는 것으로 선택할 수 있었다. 시클로텔라 크립티카를 형질 전환 벡터로 안정적으로 형질 전환하는 것은, 널리 공지된 형질 전환 기술, 예를 들어 미세 입자 투사법 또는 기타 다른 공지된 방법을 통하여 이루어졌다. nptII 유전자 생성물의 활성은, 예를 들어 G418을 포함하는 한천 ASW 배지 중에서 형질 전환 벡터로 형질 전환된 시클로텔라 크립티카를 선택하는 마커로서 사용될 수 있었다. 시클로텔라 크립티카 지질 생산에 적당한 생장 배지는 ASW 배지와, Sriharan et al.(1991)와 Pahl et al.에 의해 개시된 배지를 포함하나 이에 제한되는 것은 아니다. 시클로텔라 크립티카 지질의 지방산 프로필은 표준적인 지질 추출법과 본원에 개시된 분석 방법을 통하여 평가될 수 있었다.

[0948] **실시예 39: 나비큘라 사프로필라의 조작**

[0949] 본 발명에 따라서 나비큘라 사프로필라 내에서 재조합 유전자를 발현하는 것은, 본원에 기술된 Dunahay et al.에 의해 교시된 방법과 벡터를 변형함으로써 이루어질 수 있었다. 간단히 말해서, 문헌[Dunahay et al., Journal of Phycology, Vol. 31 (1995), pp. 1004-1012]에는, 플라스미드 DNA를 사용하여 나비큘라 사프로필라를 안정적으로 형질 전환하는 것이 개시되어 있다. 더나헤이는 미세 입자 투사법과 같은 형질 전환법을 사용하여 플라스미드 pACCNPT5.1를 나비큘라 사프로필라에 도입하였다. 플라스미드 pACCNPT5.1는, nptII 암호화 영역 상류의 시클로텔라 크립티카 아세틸-CoA 카복실라제(ACCase) 유전자(GenBank 수탁 번호 제L20784호)의 프로모터에 작동 가능하게 연결되었으며, 3' 영역(nptII 암호화 영역의 하류)에서 시클로텔라 크립티카 ACCase 유전자의 3'UTR/종결 인자에 작동 가능하도록 결합된, 네오마이신 포스포트랜스퍼라제 II(nptII) 유전자 생성물의 암호화 서열을 포함하는 네오마이신 내성 카세트를 포함하였다. nptII 유전자 생성물은 항생제 G418에 대한 내성을 부여하였다. pACCNPT5.1로 형질 전환하기 전, 나비큘라 사프로필라는 100ug/㎖ G418을 포함하는 인공 해양수 배지에서는 번식할 수 없었다. pACCNPT5.1로 형질 전환한 후에는, 100ug/㎖ G418을 포함하는 인공 해양수 선택 배지 중에서 번식한 나비큘라 사프로필라 형질 전환체를 제조하였다. 나비큘라 사프로필라 내 nptII 유전자 생성물의 발현으로 인하여, G418의 존재 하에서 상기 균주는 번식할 수 있었는데, 이로써 네오마이신 항생제 내성 카세트의 나비큘라 사프로필라 내에서 사용되는 선택가능한 마커로서의 유용성이 확립되었다. 안정적인 형질 전환체의 게놈 DNA는 서던 블럿 분석법에 의해 평가하였다. 더나헤이는, 1.07mM 규산화나트륨과 100ug/㎖ G418이 보충된 인공 해양수 배지(ASW, 문헌[Brown, L., Phycologia, Vol. 21 (1982), pp. 408-410]에 기술됨) 중 나비큘라 사프로필라의 액체 중 번식에 대해 개시하였다. 더나헤이는 또한, 100ug/㎖ G418와 함께 ASW 배지를 포함하는 한천 평판에서의 나비큘라 사프로필라 형질 전환체의 선택 및 유지에 대해 개시하였다. 부가의 배양 배지 중에서의 나비큘라 사프로필라의 번식에 대해서는, 예를 들어 문헌[Tadros 및 Johansen, Journal of Phycology, Vol. 24:4 (1988), pp. 445-452] 및 [Sriharan et al., Applied Biochemistry and Biotechnology, Vol. 20-21:1 (1989), pp. 281-291]에 기술된 바 있다. 더나헤이는, 플라스미드 pACCNPT5.1과 시클로텔라 크립티카 아세틸-CoA 카복실라제(ACCase) 유전자의 프로모터는 나비큘라 사프로필라 내에서 외인성 유전자를 발현시킬 수 있는데에 적당하다고 개시하였다. 뿐만 아니라, 더나헤이는, pACCNPT5.1상에서 암호화된 네오마이신 내성 카세트는 나비큘라 사프로필라 내 선택가능한 마커로서 사용하기 적당하다고 개시하였다.

[0950] 본 발명의 하나의 구체예에서, 선택가능한 마커로서 사용되는, nptII 유전자 생성물 암호화 뉴클레오티드 서열을 포함하는 벡터 pACCNPT5.1를 구성하고, 지질 생합성 경로 발현 카세트 서열을 추가로 포함하도록 변형하여, 형질 전환 벡터를 생산하였다. 지질 생합성 경로 발현 카세트는 표 70으로부터 선택되는 하나 이상의 지질 생합

성 경로 단백질을 암호화하였는데, 이 경우, 각각의 단백질 암호화 서열은, 표 69에 따라서, 나비큘라 사프로필라와 밀접하게 관련된 나비큘라 펠리큘로사의 핵내 유전자에 고유한 코돈 편향 사용을 반영하도록 나비큘라 사프로필라 내 발현에 대해 코돈 최적화되었다. 표 70의 지질 생합성 경로 단백질 각각에 있어서, 코돈 최적화 유전자 서열은 각각, 단백질 암호화 서열 상류의 시클로텔라 크립티카 ACCase 유전자 프로모터에 작동 가능하도록 결합될 수 있었으며, 또한 단백질 암호화 서열의 하류 또는 3' 영역에서 시클로텔라 크립티카 ACCase 유전자 3'UTR/종결 인자에 작동 가능하도록 결합될 수 있었다. 형질 전환 구조물은 또한 형질 전환 벡터의 표적화 게놈 통합에 사용되는 나비큘라 사프로필라 게놈에 상동성인 영역을 추가로 포함할 수 있었다. 상동성 영역은 내인성 지질 생합성 경로 유전자의 하나 이상의 게놈 위치를 중단하는 것으로 선택될 수 있었다. 나비큘라 사프로필라를 형질 전환 벡터로 안정적으로 형질 전환하는 것은, 널리 공지된 형질 전환 기술, 예를 들어 미세 입자 투사법 또는 기타 다른 공지된 방법을 통하여 이루어졌다. nptII 유전자 생성물의 활성은, 예를 들어 G418을 포함하는 한천 ASW 배지에서 형질 전환 벡터로 형질 전환된 나비큘라 사프로필라를 선택하는 선택가능한 마커로서 사용될 수 있었다. 나비큘라 사프로필라 지질 생산에 적당한 생장 배지로서는 ASW 배지 및 Sriharan et al.(1989) 및 태드로스(Tadros) 및 요한슨(Johansen)에 의해 개시된 배지를 포함하나, 이에 제한되는 것은 아니다. 나비큘라 사프로필라 지질의 지방산 프로필은 표준적 지질 추출법과 본원에 기술된 분석 방법을 통하여 평가될 수 있었다.

[0951] **실시예 40: 탈라시오시라 슈도나나의 조작**

[0952] 본 발명에 따라서, 재조합 유전자의 탈라시오시라 슈도나나 내 발현은 본원에 기술된 바와 같이 Poulsen(풀센) et al.에 의해 교시된 방법 및 벡터를 변형하여 이루어질 수 있었다. 간단히 말해서, 문헌[Poulsen et al., Journal of Phycology, Vol. 42 (2006), pp. 1059-1065]에는, 플라스미드 DNA를 사용하여 탈라시오시라 슈도나나를 안정적으로 형질 전환하는 것에 대해 개시되어 있다. 풀센은 미세 입자 투사법과 같은 형질 전환법을 사용하여 플라스미드 pTpfcp/nat를 탈라시오시라 슈도나나에 도입하였다. pTpfcp/nat는, nat 단백질 암호화 영역 상류의 탈라시오시라 슈도나나 후코잔틴 엽록소 a/c 결합 단백질 유전자(fcp) 프로모터에 작동 가능하도록 결합되고, 또한 3' 영역(nat 단백질 암호화 서열의 하류)에서 탈라시오시라 슈도나나 fcp 유전자 3'UTR/종결 인자에 작동 가능하도록 결합된, 누르세오트리신 아세틸트랜스퍼라제(nat) 유전자 생성물을 암호화하는 서열(GenBank 수탁 번호 제AAC60439호)을 포함하는, 누르세오트리신 내성 카세트를 포함하였다. nat 유전자 생성물은 항생제 누르세오트리신에 대한 내성을 부여하였다. pTpfcp/nat로 형질 전환하기 전, 탈라시오시라 슈도나나는 10ug/㎖ 누르세오트리신을 포함하는 배지에서 번식할 수 없었다. pTpfcp/nat로 형질 전환하여, 100ug/㎖ 누르세오트리신을 포함하는 선택 배양 배지에서 번식한 탈라시오시라 슈도나나 형질 전환체를 제조하였다. 탈라시오시라 슈도나나 내에서 nat 유전자 생성물이 발현됨으로 인하여, 100ug/㎖ 누르세오트리신의 존재하에 상기 균주는 번식할 수 있었는데, 이로써 탈라시오시라 슈도나나에서 사용되는 선택가능한 마커로서의 누르세오트리신 항생제 내성 카세트의 유용성이 확립되었다. 안정적인 형질 전환체의 게놈 DNA는 서던 블럿 분석법에 의해 평가하였다. 풀센은, 형질 전환된 탈라시오시라 슈도나나의 선택 및 유지는 100ug/㎖ 누르세오트리신 함유 개질 ESAW 배지(문헌[Harrison et al., Journal of Phycology, Vol. 16 (1980), pp. 28-35]에 기술됨)를 포함하는 액체 배양액 중에서 수행되었다고 개시하였다. 부가의 배양 배지 중 탈라시오시라 슈도나나의 번식에 대해서는, 예를 들어 문헌[Volkman et al., Journal of Experimental Marine Biology and Ecology, Vol. 128:3 (1989), pp. 219-240]에 기술된 바 있다. 탈라시오시라 슈도나나 내에서 사용하기 적당한 부가의 선택가능한 마커를 포함하는 부가의 플라스미드에 대해서는 풀센에 의한 동일한 문헌에 기술되어 있다. 풀센은, 플라스미드 pTpfcp/nat, 그리고 탈라시오시라 슈도나나 fcp 프로모터 및 3'UTR/종결 인자는 탈라시오시라 슈도나나 내에서 외인성 유전자의 발현을 가능하게 하는데 적당하였다고 개시하였다. 뿐만 아니라, 풀센은, pTpfcp/nat 상에서 암호화된 누르세오트리신 내성 카세트는 탈라시오시라 슈도나나 내에서 선택가능한 마커로서 사용하기 적당하였다고 개시하였다.

[0953] 본 발명의 하나의 구체예에서, 선택가능한 마커로서 사용되는, nat 유전자 생성물 암호화 뉴클레오티드 서열을 포함하는 벡터 pTpfcp/nat를 구성하고, 지질 생합성 경로 발현 카세트 서열을 추가로 포함하도록 변형하여, 형질 전환 벡터를 생산하였다. 지질 생합성 경로 발현 카세트는 표 70으로부터 선택되는 하나 이상의 지질 생합성 경로 단백질을 암호화하였는데, 이 경우, 각각의 단백질 암호화 서열은, 표 69에 따라서 탈라시오시라 슈도나나의 핵내 유전자에 고유한 코돈 편향 사용을 반영하도록, 탈라시오시라 슈도나나 내 발현에 대해 코돈 최적화되었다. 표 70의 지질 생합성 경로 단백질 각각에 있어서, 코돈 최적화 유전자 서열은 각각, 단백질 암호화 서상류의 탈라시오시라 슈도나나 fcp 유전자 프로모터에 작동 가능하도록 결합될 수 있었으며, 또한 단백질 암호화 서열의 하류 또는 3' 영역에서 탈라시오시라 슈도나나 fcp 유전자 3'UTR/종결 인자에 작동 가능하도록 결합될 수 있었다. 형질 전환 구조물은 형질 전환 벡터의 표적화 게놈 통합에 사용되는 탈라시오시라 슈도나나 게놈에 대해 상동성인 영역을 추가로 포함할 수도 있었다. 상동성 영역은 내인성 지질 생합성 경로 유전자의 하나 이상

의 게놈 위치를 중단하는 것으로 선택할 수 있었다. 당업자는, 탈라시오시라 슈도나나 게놈의 서열 내에서 이와 같은 상동성 영역을 동정할 수 있었다(문헌[Armbrust et al., Science, Vol. 306: 5693 (2004): pp. 79-86] 참조). 탈라시오시라 슈도나나를 형질 전환 벡터로 안정적으로 형질 전환하는 것은, 널리 공지된 형질 전환 기술, 예를 들어 미세 입자 투사법 또는 기타 다른 공지된 방법을 통하여 이루어졌다. nat 유전자 생성물의 활성은, 예를 들어 누르세오트리신을 포함하는 ESAW 한천 배지 내에서 형질 전환 벡터로 형질 전환된 탈라시오시라 슈도나나를 선택하는 마커로서 사용될 수 있었다. 탈라시오시라 슈도나나 지질 생산에 적당한 생장 배지로서는 ESAW 배지와 Volkman et al. 및 Harrison et al.에 의해 개시된 배양 배지를 포함하나, 이에 제한되는 것은 아니다. 탈라시오시라 슈도나나 지질의 지방산 프로필은 표준적 지질 추출법과 본원에 기술된 분석 방법을 통하여 평가될 수 있었다.

[0954] **실시예 41: 클라미도모나스 레인하르드티의 조작**

[0955] 본 발명에 따라서, 클라미도모나스 레인하르드티 내에서 재조합 유전자의 발현은 본원에 기술된 바와 같이 Cerutti(세루티) et al.에 의해 교시된 방법과 벡터를 변형하여 이루어질 수 있었다. 간단히 말해서, 문헌[Cerutti et al., Genetics, Vol. 145:1 (1997), pp. 97-110]에는, 형질 전환 벡터로 클라미도모나스 레인하르드티를 핵내 안정적으로 형질 전환하는 것에 대해 개시되어 있다. 세루티는 미세 입자 투사법과 같은 형질 전환법을 사용하여 형질 전환 구조물 P[1030]을 클라미도모나스 레인하르드티에 도입하였다. 구조물 P[1030]는, 아미노글루코시드 3'-아데닐트랜스퍼라제(aadA) 단백질 암호화 영역 상류의 클라미도모나스 레인하르드티 리불로스-1,5-비스포스페이트 카복실라제/옥시게나제 소형 서브유닛 유전자(RbcS2, GenBank 수탁 번호 제X04472호) 프로모터에 작동 가능하도록 결합되고, 3' 영역(aadA 단백질 암호화 서열의 하류)에서 클라미도모나스 레인하르드티 RbcS2 유전자 3'UTR/종결 인자에 작동 가능하도록 결합된, aadA 유전자 생성물 암호화 서열을 포함하는, 스펙티노마이신 내성 카세트를 포함하였다. aadA 유전자 생성물은, 항생제 스펙티노마이신에 대한 내성을 부여하였다. P[1030]로 형질 전환하기 전, 클라미도모나스 레인하르드티는 90ug/㎖ 스펙티노마이신을 포함하는 배지에서는 번식할 수 없었다. P[1030]로 형질 전환하여, 90ug/㎖ 스펙티노마이신을 포함하는 선택 배양 배지에서 번식한 클라미도모나스 레인하르드티 형질 전환체가 제조되었는데, 이로써 클라미도모나스 레인하르드티 내에서 사용되는 선택가능한 마커로서의 스펙티노마이신 항생제 내성 카세트의 유용성이 확립되었다. 안정적인 형질 전환체의 게놈 DNA는 서던 블럿 분석법에 의해 평가되었다. 세루티는, 형질 전환된 클라미도모나스 레인하르드티의 선택 및 유지는 90ug/㎖ 스펙티노마이신을 함유하고 트리스-아세테이트-포스페이트 배지(TAP, 문헌[Harris, The Chlamydomonas Sourcebook, Academic Press, San Diego, 1989]에 개시됨)를 포함하는 한천 평판 상에서 수행되었다고 개시하였다. 세루티는 또한, 클라미도모나스 레인하르드티의 TAP 액체 배양액(90ug/㎖ 스펙티노마이신 함유) 중 번식에 대해서도 개시하였다. 대안적 배양 배지 중 클라미도모나스 레인하르드티의 번식에 대해서는, 예를 들어 문헌[Dent et al., African Journal of Microbiology Research, Vol. 5:3 (2011), pp. 260-270] 및 [Yantao et al., Biotechnology and Bioengineering, Vol. 107:2 (2010), pp. 258-268]에 개시되어 있다. 클라미도모나스 레인하르드티에 사용하기 적당한 부가의 선택가능한 마커와 다수의 조절 서열(예를 들어, 클라미도모나스 레인하르드티 내 이종 유전자 발현을 촉진하는데 적당한 3'UTR 및 프로모터)을 포함하는 부가의 구조물은 당업계에 공지되어 있으며, 문헌[Radakovits et al., Eurkaryotic Cell, Vol. 9:4 (2010), pp. 486-501]에 기술되어 있다. 세루티는, 형질 전환 벡터 P[1030]과 클라미도모나스 레인하르드티 프로모터 및 3'UTR/종결 인자는 클라미도모나스 레인하르드티 내 외인성 유전자 발현을 가능하게 하는데 적당하다고 개시하였다. 뿐만 아니라, 세루티는, P[1030] 상에 암호화되는 스펙티노마이신 내성 카세트가 클라미도모나스 레인하르드티 내에서 선택가능한 마커로서 사용되기 적당하였다고 개시하였다.

[0956] 본 발명의 하나의 구체예에서, 선택가능한 마커로서 사용되는 aadA 유전자 생성물 암호화 뉴클레오티드 서열을 포함하는 벡터 P[1030]를 구성하고, 지질 생합성 경로 발현 카세트 서열을 추가로 포함하도록 변형하여, 형질 전환 벡터를 생산하였다. 지질 생합성 경로 발현 카세트는 표 70으로부터 선택되는 하나 이상의 지질 생합성 경로 단백질을 암호화하였는데, 이 경우, 각각의 단백질 암호화 서열은, 표 69에 따라서 클라미도모나스 레인하르드티의 핵내 유전자에 고유한 코돈 편향 사용을 반영하도록, 클라미도모나스 레인하르드티 내 발현에 대해 코돈 최적화되었다. 표 70의 지질 생합성 경로 단백질 각각에 있어서, 코돈 최적화 유전자 서열은 각각, 단백질 암호화 서열 상류의 클라미도모나스 레인하르드티 RbcS2 프로모터에 작동 가능하도록 결합될 수 있었으며, 단백질 암호화 서열의 하류 또는 3' 영역에서 클라미도모나스 레인하르드티 RbcS2 3'UTR/종결 인자에 작동 가능하도록 결합될 수 있었다. 형질 전환 구조물은 형질 전환 벡터의 표적화 게놈 통합에 사용되는, 클라미도모나스 레인하르드티 게놈에 상동성인 영역을 추가로 포함할 수 있었다. 상동성 영역은 내인성 지질 생합성 경로 유전자의 하나 이상의 게놈 위치를 중단하는 것으로 선택할 수 있었다. 당업자는, 클라미도모나스 레인하르드티 게놈의 서열 내에서 이와 같은 상동성 영역을 동정할 수 있다(문헌[Merchant et al., Science, Vol. 318:5848 (2007),

pp. 245-250] 참조). 클라미도모나스 레인하르드티를 형질 전환 벡터로 안정적으로 형질 전환하는 것은, 널리 공지된 형질 전환 기술, 예를 들어 미세 입자 투사법 또는 기타 다른 공지된 방법들을 통하여 이루어졌다. aadA 유전자 생성물의 활성은, 예를 들어 스펙티노마이신을 포함하는 TAP 한천 배지에서 형질 전환 벡터로 형질 전환된 클라미도모나스 레인하르드티를 선택하는 마커로서 사용될 수 있었다. 클라미도모나스 레인하르드티 지질 생산에 적당한 생장 배지는 ESAW 배지와, Yantao et al. 및 Dent et al.에 의해 개시된 배양 배지를 포함하나, 이에 제한되는 것은 아니다. 클라미도모나스 레인하르드티 지질의 지방산 프로필은 표준적 지질 추출법과 본원에 기술된 분석 방법을 통하여 평가될 수 있었다.

[0957] **실시예 42: 야로위아 리폴리티카의 조작**

[0958] 본 발명에 따라서 재조합 유전자의 발현은 본원에 기술된 바와 같이 Fickers(픽커스) et al.에 의해 교시된 방법과 벡터를 변형하여 야로위아 리폴리티카 내에서 이루어질 수 있었다. 간단히 말해서, 문헌[Fickers et al., Journal of Microbiological Methods, Vol. 55 (2003), pp. 727-737]에는, 플라스미드 DNA로 야로위아 리폴리티카를 안정적으로 핵내 형질 전환하는 것에 대해 개시되어 있다. 픽커스는 아세트산 리튬 형질 전환법을 사용하여 플라스미드 JMP123을 야로위아 리폴리티카에 도입하였다. 플라스미드 JMP123은, hph 단백질 암호화 영역 상류의 야로위아 리폴리티카 LIP2 유전자 프로모터(GenBank 수탁 번호 제AJ012632호)에 작동 가능하도록 결합되고, 또한 hph 단백질 암호화 영역 하류의 야로위아 리폴리티카 LIP2 유전자 3'UTR/종결 인자에 작동 가능하도록 결합된, 하이그로마이신 B 포스포트랜스퍼라제 유전자 생성물(hph)을 암호화하는 서열을 포함하는 하이그로마이신 B 내성 카세트를 포함하였다. JMP123으로 형질 전환하기 전, 야로위아 리폴리티카는 100ug/㎖ 하이그로마이신을 포함하는 배지에서 번식할 수 없었다. JMP123으로 형질 전환하여, 100ug/㎖ 하이그로마이신을 포함하는 배지에서 번식할 수 있었던, 형질 전환된 야로위아 리폴리티카가 생산되었는데, 이로써, 야로위아 리폴리티카 내에서 사용되는 선택가능한 마커로서의 하이그로마이신 B 항생제 내성 카세트의 유용성이 확립되었다. 야로위아 리폴리티카 LIP2 유전자의 3'UTR/종결 인자와 프로모터의 JMP123상에 제공된 뉴클레오티드 서열은, LIP2 위치로 hph 암호화 서열이 상동성 재조합되는데 사용되는 공여 서열로서 사용되었다. 안정적인 형질 전환체의 게놈 DNA는 서던 블럿에 의해 평가되었다. 픽커스는, 형질 전환된 야로위아 리폴리티카의 선택 및 유지는 100ug/㎖ 하이그로마이신을 함유하는 표준 YPD 배지(효모 추출 펩톤 덱스트로스)를 포함하는 한천 평판에서 수행되었다. 형질 전환된 야로위아 리폴리티카의 액체 배양은, 하이그로마이신을 함유하는 YPD 배지 중에서 수행되었다. 야로위아 리폴리티카를 배양하는데 사용되는 기타 다른 배지와 기술에 대해서는 개시되었으며, 야로위아 리폴리티카 내에서 사용되는 선택가능한 마커, 3'UTR, 프로모터 및 기타 다른 다수의 플라스미드도, 예를 들어 문헌[Pignede et al., Applied and Environmental Biology, Vol. 66:8 (2000), pp. 3283-3289], [Chuang et al., New Biotechnology, Vol. 27:4 (2010), pp. 277-282], 및 [Barth 및 Gaillardin, (1996), In: K,W. (Ed.), Nonconventional Yeasts in Biotecnology. Sprinter-Verlag, Berlin-Heidelber, pp. 313-388]에 개시되었다. 픽커스는, 형질 전환 벡터 JMP123 및 야로위아 리폴리티카 LIP2 유전자 프로모터, 그리고 3'UTR/종결 인자는 야로위아 리폴리티카 내에서 이종 유전자 발현을 가능하게 하는데 적당하다고 개시하였다. 뿐만 아니라, 픽커스는, JMP123상에서 암호화되는 하이그로마이신 내성 카세트가 야로위아 리폴리티카 내 선택가능한 마커로서 사용되기 적당하였다고 개시하였다.

[0959] 본 발명의 하나의 구체예에서, 선택가능한 마커로서 사용되는 hph 유전자 생성물 암호화 뉴클레오티드 서열을 포함하는 벡터 JMP123을 구성하고, 지질 생합성 경로 발현 카세트 서열을 추가로 포함하도록 변형하여, 형질 전환 벡터를 생산하였다. 지질 생합성 경로 발현 카세트는 표 70으로부터 선택되는 하나 이상의 지질 생합성 경로 단백질을 암호화하였는데, 이 경우, 각각의 단백질 암호화 서열은, 표 69에 따라서 야로위아 리폴리티카의 핵내 유전자에 고유한 코돈 편향 사용을 반영하도록, 야로위아 리폴리티카 내 발현에 대해 코돈 최적화되었다. 표 70의 지질 생합성 경로 단백질 각각에 있어서, 코돈 최적화 유전자 서열은 각각, 단백질 암호화 서열 상류의 야로위아 리폴리티카 LIP2 유전자 프로모터에 작동 가능하도록 결합될 수 있었으며, 단백질 암호화 서열의 하류 또는 3' 영역에서 야로위아 리폴리티카 LIP2 유전자 3'UTR/종결 인자에 작동 가능하도록 결합될 수 있었다. 형질 전환 구조물은 형질 전환 벡터의 표적화 게놈 통합에 사용되는, 야로위아 리폴리티카 게놈에 상동성인 영역을 추가로 포함할 수 있었다. 상동성 영역은 내인성 지질 생합성 경로 유전자의 하나 이상의 게놈 위치를 중단하는 것으로 선택할 수 있었다. 당업자는, 야로위아 리폴리티카 게놈의 서열 내에서 이와 같은 상동성 영역을 동정할 수 있다(문헌[Dujun et al., Nature, Vol. 430 (2004), pp. 35-44] 참조). 야로위아 리폴리티카를 형질 전환 벡터로 안정적으로 형질 전환하는 것은, 널리 공지된 형질 전환 기술, 예를 들어 아세트산 리튬 형질 전환법 또는 기타 다른 공지된 방법들을 통하여 이루어졌다. hph 유전자 생성물의 활성은, 예를 들어 하이그로마이신을 포함하는 YPD 배지에서 형질 전환 벡터로 형질 전환된 야로위아 리폴리티카를 선택하는 마커로서 사용될 수 있었다. 야로위아 리폴리티카 지질 생산에 적당한 생장 배지는 YPD 배지와, Chuang et al.에 의해 개시된 배양 배

지를 포함하나, 이에 제한되는 것은 아니다. 야로위아 리폴리티카 지질의 지방산 프로필은 표준적 지질 추출법과 본원에 기술된 분석 방법을 통하여 평가될 수 있었다.

[0960] **실시예 43: 모르티에렐라 알파인의 조작**

[0961] 본 발명에 따라서, 재조합 유전자의 모르티에렐라 알파인 내 발현은 본원에 기술된 바와 같이 Mackenzie(맥켄지) et al.에 의해 교시된 방법과 벡터를 변형하여 이루어질 수 있었다. 간단히 말해서, 문헌[Mackenzie et al., Applied and Environmental Microbiology, Vol. 66 (2000), pp. 4655-4661]에는, 플라스미드 DNA로 모르티에렐라 알피나를 안정적으로 핵내에서 형질 전환하는 것에 대해 개시되어 있다. 맥켄지는 원형질체 형질 전환법을 사용하여 플라스미드 pD4를 모르티에렐라 알피나에 도입하였다. 플라스미드 pD4는, hpt 단백질 암호화 영역 상류의 모르티에렐라 알피나 히스톤 H4.1 유전자 프로모터(GenBank 수탁 번호 제AJ249812호)에 작동 가능하도록 결합되고, hpt 단백질 암호화 영역 하류의 아스퍼질러스 니둘란스 N-(5'-포스포리보실)안트라닐레이트 이소머라제(trpC) 유전자 3'UTR/종결 인자에 작동 가능하도록 결합된, 하이그로마이신 B 포스포트랜스퍼라제 유전자 생성물(hpt) 암호화 서열을 포함하는 하이그로마이신 B 내성 카세트를 포함하였다. pD4로 형질 전환하기 전, 모르티에렐라 알피나는 300ug/㎖ 하이그로마이신을 포함하는 배지에서 번식할 수 없었다. pD4로 형질 전환하였을 때, 300ug/㎖ 하이그로마이신을 포함하는 배지에서 번식된, 형질 전환된 모르티에렐라 알피나가 생산되었는데, 이로써, 모르티에렐라 알피나 내에서 사용되는 선택가능한 마커로서의 하이그로마이신 B 항생제 내성 카세트의 유용성이 확립되었다. 안정적인 형질 전환체의 게놈 DNA는 서던 블럿에 의해 평가되었다. 맥켄지는, 형질 전환된 모르티에렐라 알피나의 선택 및 유지는 하이그로마이신을 포함하는 PDA(감자 덱스트로스 한천) 배지에서 이루어졌다고 개시하였다. 형질 전환된 모르티에렐라 알피나의 액체 배양(맥켄지에 의함)은 PDA 배지 중 또는 S2GYE 배지(5% 글루코스, 0.5% 효모 추출물, 0.18% NH4SO4, 0.02% MgSO4-7H2O, 0.0001% FeCl3- 6H2O, 0.1% 미량 원소, 10mM K2HPO4-NaH2PO4 포함)(하이그로마이신 함유) 중에서 수행되었다. 모르티에렐라 알피나를 배양하는데 사용되는 기타 다른 배지와 기술, 그리고 모르티에렐라 알피나에서 사용되는 선택가능한 마커, 3'UTR, 프로모터 및 기타 다른 플라스미드에 대해서는, 예를 들어 문헌[Ando et al., Applied and Environmental Biology, Vol. 75:17 (2009) pp. 5529-35] 및 [Lu et al., Applied Biochemistry and Biotechnology, Vol. 164:7 (2001), pp. 979-90]에 개시되어 있다. 맥켄지는, 형질 전환 벡터 pD4와 모르티에렐라 알피나 하스톤 H4.1 프로모터, 그리고 에이.니둘란스 trpC 유전자 3'UTR/종결 인자는 모르티에렐라 알피나 내에서의 이종 유전자 발현을 가능하게 하는데 적당하다고 개시하였다. 뿐만 아니라, 맥켄지는, pD4상에서 암호화된 하이그로마이신 내성 카세트는 모르티에렐라 알피나 내에서 선택가능한 마커로서 사용하기 적당하였다고 개시하였다.

[0962] 본 발명의 하나의 구체예에서, 선택가능한 마커로서 사용되는 hpt 유전자 생성물 암호화 뉴클레오티드 서열을 포함하는 벡터 pD4를 구성하고, 지질 생합성 경로 발현 카세트 서열을 추가로 포함하도록 변형히여, 형질 전환 벡터를 생산하였다. 지질 생합성 경로 발현 카세트는 표 70으로부터 선택되는 하나 이상의 지질 생합성 경로 단백질을 암호화하였는데, 이 경우, 각각의 단백질 암호화 서열은, 표 69에 따라서 모르티에렐라 알피나의 핵내 유전자에 고유한 코돈 편향 사용을 반영하도록, 모르티에렐라 알피나 내 발현에 대해 코돈 최적화되었다. 표 70의 지질 생합성 경로 단백질 각각에 있어서, 코돈 최적화 유전자 서열은 각각, 단백질 암호화 서열 상류의 모르티에렐라 알피나 히스톤 H4.1 유전자 프로모터에 작동 가능하도록 결합될 수 있었으며, 단백질 암호화 영역의 하류 또는 3' 영역에서 에이.니둘란스 trpC 3'UTR/종결 인자에 작동 가능하도록 결합될 수 있었다. 형질 전환 구조물은 형질 전환 벡터의 표적화 게놈 통합에 사용되는, 모르티에렐라 알피나 게놈에 상동성인 영역을 추가로 포함할 수 있었다. 상동성 영역은 내인성 지질 생합성 경로 유전자의 하나 이상의 게놈 위치를 중단하는 것으로 선택될 수 있었다. 당업자는, 모르티에렐라 알파인 게놈의 서열 내에서 이와 같은 상동성 영역을 동정할 수 있다(문헌[Wang et al., PLOS One, Vol. 6:12 (2011)] 참조). 모르티에렐라 알피나를 형질 전환 벡터로 안정적으로 형질 전환하는 것은, 널리 공지된 형질 전환 기술, 예를 들어 원형질체 형질 전환법 또는 기타 다른 공지된 방법들을 통하여 이루어졌다. hpt 유전자 생성물의 활성은, 예를 들어 하이그로마이신을 포함하는 PDA 배지에서 형질 전환 벡터로 형질 전환된 모르티에렐라 알피나를 선택하는 마커로서 사용될 수 있었다. 모르티에렐라 알피나 지질 생산에 적당한 생장 배지는 S2GYE 배지와, Lu et al. 및 Ando et al.에 의해 개시된 배양 배지를 포함하나, 이에 제한되는 것은 아니다. 모르티에렐라 알피나 지질의 지방산 프로필은 표준적 지질 추출법과 본원에 기술된 분석 방법을 통하여 평가될 수 있었다.

[0963] **실시예 44: 로도코커스 오파쿠스 PD630의 조작**

[0964] 본 발명에 따라서 재조합 유전자의 로도코커스 오파쿠스 PD630 내 발현은 본원에 기술된 바와 같이 Kalscheuer(칼슈어) et al.에 의해 교시된 방법과 벡터를 변형하여 이루어질 수 있었다. 간단히 말하면, 문헌[Kalscheuer

et al., Applied and Environmental Microbiology, Vol. 52 (1999), pp. 508-515]에는, 로도코커스 오파쿠스를 플라스미드 DNA로 안정적으로 형질 전환하는 것에 대해 개시되어 있다. 칼슈어는 전기 천공법과 같은 형질 전환 법을 사용하여 플라스미드 pNC9501을 로도코커스 오파쿠스 PD630에 도입하였다. 플라스미드 pNC9501은, 유전자의 프로모터와 3' 종결 인자 서열을 포함하는, 스트렙토마이세스 아주레우스 23S rRNA A1067 메틸트랜스퍼라제 유전자의 전 뉴클레오티드 서열을 포함하는 티오스트렙톤 내성(thior) 카세트를 포함하였다. pNC9501으로 형질 전환하기 전, 로도코커스 오파쿠스는 1mg/㎖ 티오스트렙톤을 포함하는 배지에서 번식할 수 없었다. 로도코커스 오파쿠스 PD630을 pNC9501로 형질 전환하여, 1mg/㎖ 티오스트렙톤을 포함하는 배양 배지에서 번식된 형질 전환체가 생산되었는데, 이로써, 로도코커스 오파쿠스 PD630 내 선택가능한 마커로서의 티오스트렙톤 내성 카세트의 용도가 확립되었다. 칼슈어에 의해 개시된 제2 플라스미드 pAK68은, 내성 thior 카세트와, 랄스토니아 유트로파 베타-케토티올라제 유전자(phaB) 서열, 아세토아세틸-CoA 환원 효소 유전자(phaA) 서열 및 폴리 3-하이드록시알칸산 합성 효소 유전자(phaC) 서열(폴리하이드록시알카노에이트 생합성용; lacZ 프로모터에 의해 작동됨)을 포함하였다. 폴리하이드록시알카노에이트 생합성 경로가 결핍된 로도코커스 오파쿠스 PD630 균주를 pAK68로 형질 전환하여, 형질 전환된 로도코커스 오파쿠스 PD630(형질 전환되지 않은 균주보다 많은 양의 폴리하이드록시알카노에이트를 생산)을 생산하였다. 도입된 랄스토니아 유트로파 phaB, phaA 및 phaC 효소의 검출 가능한 활성은, pAK68 플라스미드상에서 암호화되는 조절 요소가 로도코커스 오파쿠스 PD630 내 이종 유전자 발현에 적당하였음을 암시하였다. 칼슈어는, 형질 전환된 로도코커스 오파쿠스 PD630의 선택 및 유지는, 표준적 루리아 브로스(Luria Broth)(LB) 배지, 영양 브로스(NB), 또는 티오스트렙톤을 포함하는 무기염 배지(MSM) 상에서 이루어졌다고 개시하였다. 로도코커스 오파쿠스 PD630를 배양하는데 사용되는 기타 다른 배지와 기술에 대해서는, 예를 들어 문헌[Kurosawa et al., Journal of Biotechnology, Vol. 147:3-4 (2010), pp. 212-218] 및 [Alverez et al., Applied Microbial and Biotechnology, Vol. 54:2 (2000), pp.218-223]에 개시되어 있다. 칼슈어는, 형질 전환 벡터 pNC9501 및 pAK68, 스트렙토마이세스 아주레우스 23S rRNA A1067 메틸트랜스퍼라제 유전자의 프로모터 및 lacZ 유전자가 로도코커스 오파쿠스 PD630 내 이종 유전자 발현을 가능하게 하는데 적당하다고 개시하였다. 뿐만 아니라, 칼슈어는, pNC9501 및 pAK68상에서 암호화되는 thior 카세트가 로도코커스 오파쿠스 PD630 내 선택가능한 마커로서 사용하기 적당하였다고 개시하였다.

[0965] 본 발명의 하나의 구체예에서, 선택가능한 마커로서 사용되는 thior 유전자 생성물 암호화 뉴클레오티드 서열을 포함하는 벡터 pNC9501을 구성하고, 지질 생합성 경로 발현 카세트 서열을 추가로 포함하도록 변형하여, 형질 전환 벡터를 생산하였다. 지질 생합성 경로 발현 카세트는 표 70으로부터 선택되는 하나 이상의 지질 생합성 경로 단백질을 암호화하였는데, 이 경우, 각각의 단백질 암호화 서열은, 표 69에 따라서 로도코커스 오파쿠스의 핵내 유전자에 고유한 코돈 편향 사용을 반영하도록, 로도코커스 오파쿠스 PD630 내 발현에 대해 코돈 최적화되었다. 표 70의 지질 생합성 경로 단백질 각각에 있어서, 코돈 최적화 유전자 서열은 각각, 단백질 암호화 서열 상류의 lacZ 유전자 프로모터에 작동 가능하도록 결합될 수 있었다. 형질 전환 구조물은 형질 전환 벡터의 표적화 게놈 통합에 사용되는, 로도코커스 오파쿠스 PD630 게놈에 상동성인 영역을 추가로 포함할 수 있었다. 상동성 영역은 내인성 지질 생합성 경로 유전자의 하나 이상의 게놈 위치를 중단하는 것으로 선택할 수 있었다. 당업자는, 로도코커스 오파쿠스 PD630 게놈의 서열 내에서 이와 같은 상동성 영역을 동정할 수 있다(문헌[Holder et al., PLOS Genetics, Vol. 7:9 (2011)] 참조). 로도코커스 오파쿠스 PD630을 형질 전환 벡터로 형질 전환하는 것은, 널리 공지된 형질 전환 기술, 예를 들어 전기 천공법 또는 기타 다른 공지된 방법들을 통하여 이루어졌다. 스트렙토마이세스 아주레우스 23S rRNA A1067 메틸트랜스퍼라제 유전자 생성물의 활성은, 예를 들어 티오스트렙톤을 포함하는 LB 배지에서 형질 전환 벡터로 형질 전환된 로도코커스 오파쿠스 PD630을 선택하는 마커로서 사용될 수 있었다. 로도코커스 오파쿠스 PD630 지질 생산에 적당한 생장 배지는 Kurosawa et al. 및 Albarez et al.에 의해 개시된 배양 배지를 포함하나, 이에 제한되는 것은 아니다. 로도코커스 오파쿠스 PD630 지질의 지방산 프로필은 표준적 지질 추출법과 본원에 기술된 분석 방법을 통하여 평가될 수 있었다.

[0966] 본원에 인용된 모든 참고 문헌, 예를 들어 특허, 특허 출원 및 공보, 예를 들어 GenBank 수탁 번호는, 사전에 구체적으로 인용되었든 아니든, 본원에 그 자체로서 참조로 포함되어 있는 것이다. 본원에 인용된 공보는, 본 발명과 연관되어 사용될 수 있는 시약, 방법 및 개념을 기술 및 개시하고자 하는 목적으로 인용되었다. 본원에 개시된 것들 중 그 어떤 것도, 본원에 인용된 참고 문헌들이 본원에 기술된 본 발명과 관련하여 선행 기술에 해당함을 인정하는 것으로서 해석되지 않는다. 특히 다음과 같은 특허 출원은 모든 목적을 위하여 그 자체로서 본원에 참조로 포함되었다: PCT 특허 출원 PCT/US2008/065563(2008년 6월 2일 출원; 발명의 명칭 “Production of Oil in Microorganisms”), PCT 특허 출원 PCT/US2010/31108(2010년 4월 14일 출원; 발명의 명칭 “Methods of Microbial Oil Extraction and Separation”), 그리고 PCT 특허 출원 PCT/US2009/066142(2009년 11월 30일 출원; 발명의 명칭 “Production of Tailored Oils in Heterotrophic Microorganisms”).

[0967] **서열 목록**

[0968] 서열 번호 1

[0969] 클로렐라로부터의 HUP 프로모터 (GenBank 수탁 번호 X55349의 서브시퀀스)

GATCAGACGGGCCTGACCTGCGAGATAATCAAGTGCTCGTAGGCAACCAACTCA
GCAGCTGCTTGGTGTTGGGTCTGCAGGATAGTGTTGCAGGGCCCCAAGGACAGC
AGGGGAACTTACACCTTGTCCCCGACCCAGTTTTATGGAGTGCATTGCCTCAAGA
GCCTAGCCGGAGCGCTAGGCTACATACTTGCCGCACCGGTATGAGGGGATATAG
TACTCGCACTGCGCTGTCTAGTGAGATGGGCAGTGCTGCCCATAAACAACTGGCT
GCTCAGCCATTTGTTGGCGGACCATTCTGGGGGGGCCAGCAATGCCTGACTTTCG
GGTAGGGTGAAAACTGAACAAAGACTACCAAAACAGAATTTCTTCCTCCTTGGA
GGTAAGCGCAGGCCGGCCCGCCTGCGCCCACATGGCGCTCCGAACACCTCCATA
GCTGTAAGGGCGCAAACATGGCCGGACTGTTGTCAGCACTCTTTCATGGCCATAC
AAGGTCATGTCGAGATTAGTGCTGAGTAAGACACTATCACCCCATGTTCGATTGA
AGCCGTGACTTCATGCCAACCTGCCCCTGGGCGTAGCAGACGTATGCCATCATGA
CCACTAGCCGACATGCGCTGTCTTTTGCCACCAAAACAACTGGTACACCGCTCGA
AGTCGTGCCGCACACCTCCGGGAGTGAGTCCGGCGACTCCTCCCCGGCGGGCCG
CGGCCCTACCTGGGTAGGGTCGCCATACGCCCACGACCAAACGACGCAGGAGGG
GATTGGGGTAGGGAATCCCAACCAGCCTAACCAAGACGGCACCTATAATAATAG
GTGGGGGGGACTAACAGCCCTATATCGCAAGCTTTGGGTGCCTATCTTGAGAAGCA
CGAGTTGGAGTGGCTGTGTACGGTCGACCCTAAGGTGGGTGTGCCGCAGCCTGA
AACAAAGCGTCTAGCAGCTGCTTCTATAATGTGTCAGCCGTTGTGTTTCAGTTAT
ATTGTATGCTATTGTTTGTTCGTGCTAGGGTGGCGCAGGCCCACCTACTGTGGCG
GGCCATTGGTTGGTGCTTGAATTGCCTCACCATCTAAGGTCTGAACGCTCACTCA
AACGCCTTTGTACAACTGCAGAACTTTCCTTGGCGCTGCAACTACAGTGTGCAAA
CCAGCACATAGCACTCCCTTACATCACCCAGCAGTACAACA

[0970]

[0971] 서열 번호 2

[0972] AY307383으로부터의 클로렐라 엘립소이데아 질산염 환원 효소 프로모터

CGCTGCGCACCAGGGCCGCCAGCTCGCTGATGTCGCTCCAAATGCGGTCCCCCGA
TTTTTTGTTCTTCATCTTCTCCACCTTGGTGGCCTTCTTGGCCAGGGCCTTCAGCTG
CATGCGCACAGACCGTTGAGCTCCTGATCAGCATCCTCAGGAGGCCCTTTGACAA
GCAAGCCCCTGTGCAAGCCCATTCACGGGGTACCAGTGGTGCTGAGGTAGATGG
GTTTGAAAAGGATTGCTCGGTCGATTGCTGCTCATGGAATTGGCATGTGCATGCA
TGTTCACAATATGCCACCAGGCTTTGGAGCAAGAGAGCATGAATGCCTTCAGGC
AGGTTGAAAGTTCCTGGGGGTGAAGAGGCAGGGCCGAGGATTGGAGGAGGAAA
GCATCAAGTCGTCGCTCATGCTCATGTTTTCAGTCAGAGTTTGCCAAGCTCACAG
GAGCAGAGACAAGACTGGCTGCTCAGGTGTTGCATCGTGTGTGTGGTGGGGGGG
GGGGGGTTAATACGGTACGAAATGCACTTGGAATTCCCACCTCATGCCAGCGGA
CCCACATGCTTGAATTCGAGGCCTGTGGGGTGAGAAATGCTCACTCTGCCCTCGT
TGCTGAGGTACTTCAGGCCGCTGAGCTCAAAGTCGATGCCCTGCTCGTCTATCAG
GGCCTGCACCTCTGGGCTGACCGGCTCAGCCTCCTTCGCGGGCATGGAGTAGGCG
CCGGCAGCGTTCATGTCCGGGCCCAGGGCAGCGGTGGTGCCATAAATGTCGGTG
ATGGTGGGGAGGGGGGCCGTCGCCACACCATTGCCGTTGCTGGCTGACGCATGC
ACATGTGGCCTGGCTGGCACCGGCAGCACTGGTCTCCAGCCAGCCAGCAAGTGG
CTGTTCAGGAAAGCGGCCATGTTGTTGGTCCCTGCGCATGTAATTCCCCAGATCA
AAGGAGGGAACAGCTTGGATTTGATGTAGTGCCCAACCGGACTGAATGTGCGAT

[0973]

GGCAGGTCCCTTTGAGTCTCCCGAATTACTAGCAGGGCACTGTGACCTAACGCAG
CATGCCAACCGCAAAAAAATGATTGACAGAAAATGAAGCGGTGTGTCAATATTT
GCTGTATTTATTCGTTTTAATCAGCAACCAAGTTCGAAACGCAACTATCGTGGTG
ATCAAGTGAACCTCATCAGACTTACCTCGTTCGGCAAGGAAACGGAGGCACCAA
ATTCCAATTTGATATTATCGCTTGCCAAGCTAGAGCTGATCTTTGGGAAACCAAC
TGCCAGACAGTGGACTGTGATGGAGTGCCCCGAGTGGTGGAGCCTCTTCGATTCG
GTTAGTCATTACTAACGTGAACCCTCAGTGAAGGGACCATCAGACCAGAAAGAC
CAGATCTCCTCCTCGACACCGAGAGAGTGTTGCGGCAGTAGGACGACAAG

[0974]

[0975] 서열 번호 3

[0976] 효모 수크로스 인버타제

MTNETSDRPLVHFTPNKGWMNDPNGLWYDEKDAKWHLYFQYNPNDTVWGTPLF
WGHATSDDLTNWEDQPIAIAPKRNDSGAFSGSMVVDYNNTSGFFNDTIDPRQRCVAI
WTYNTPESEEQYISYSLDGGYTFTEYQKNPVLAANSTQFRDPKVFWYEPSQKWIMT
AAKSQDYKIEIYSSDDLKSWKLESAFANEGFLGYQYECPGLIEVPTEQDPSKSYWVM
FISINPGAPAGGSFNQYFVGSFNGTHFEAFDNQSRVVDFGKDYYALQTFFNTDPTYGS
ALGIAWASNWEYSAFVPTNPWRSSMSLVRKFSLNTEYQANPETELINLKAEPILNISN
AGPWSRFATNTTLTKANSYNVDLSNSTGTLEFELVYAVNTTQTISKSVFADLSLWFK
GLEDPEEYLRMGFEVSASSFFLDRGNSKVKFVKENPYFTNRMSVNNQPFKSENDLSY
YKVYGLLDQNILELYFNDGDVVSTNTYFMTTGNALGSVNMTTGVDNLFYIDKFQVR
EVK

[0977]

[0978] 서열 번호 4

[0979] 효모 분비 신호

[0980] MLLQAFLFLLAGFAAKISAS

[0981] 서열 번호 5

[0982] 고등 식물 분비 신호

[0983] MANKSLLLLLLLGSLASG

[0984] 서열 번호 6

[0985] 콘센서스 진핵 분비 신호

[0986] MARLPLAALG

[0987] 서열 번호 7

[0988] 조합 고등 식물/진핵 분비 신호

[0989] MANKLLLLLLLLLPLAASG

[0990] 서열 번호 8

[0991] 에스.세레비지아에 수크로스 인버타제 NP_012104

GAATTCCCCAACATGGTGGAGCACGACACTCTCGTCTACTCCAAGAATATCAAAG
ATACAGTCTCAGAAGACCAAAGGGCTATTGAGACTTTTCAACAAAGGGTAATAT
CGGGAAACCTCCTCGGATTCCATTGCCCAGCTATCTGTCACTTCATCAAAAGGAC
AGTAGAAAAGGAAGGTGGCACCTACAAATGCCATCATTGCGATAAAGGAAAGGC
TATCGTTCAAGATGCCTCTGCCGACAGTGGTCCCAAAGATGGACCCCCACCCACG
AGGAGCATCGTGGAAAAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGAT

[0992]

TGATGTGAACATGGTGGAGCACGACACTCTCGTCTACTCCAAGAATATCAAAGAT
ACAGTCTCAGAAGACCAAAGGGCTATTGAGACTTTTCAACAAAGGGTAATATCG
GGAAACCTCCTCGGATTCCATTGCCCAGCTATCTGTCACTTCATCAAAAGGACAG
TAGAAAAGGAAGGTGGCACCTACAAATGCCATCATTGCGATAAAGGAAAGGCTA
TCGTTCAAGATGCCTCTGCCGACAGTGGTCCCAAAGATGGACCCCCACCCACGAG
GAGCATCGTGGAAAAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGATTG
ATGTGATATCTCCACTGACGTAAGGGATGACGCACAATCCCACTATCCTTCGCAA
GACCCTTCCTCTATATAAGGAAGTTCATTTCATTTGGAGAGGACACGCTGAAATC
ACCAGTCTCTCTCTACAAATCTATCTCTGGCGCGCCATATCAATGCTTCTTCAGGC
CTTTCTTTTTCTTCTTGCTGGTTTTGCTGCCAAGATCAGCGCCTCTATGACGAACG
AAACCTCGGATAGACCACTTGTGCACTTTACACCAAACAAGGGCTGGATGAATG
ACCCCAATGGACTGTGGTACGACGAAAAAGATGCCAAGTGGCATCTGTACTTTC
AATACAACCCGAACGATACTGTCTGGGGGACGCCATTGTTTTGGGGCCACGCCAC
GTCCGACGACCTGACCAATTGGGAGGACCAACCAATAGCTATCGCTCCGAAGAG
GAACGACTCCGGAGCATTCTCGGGTTCCATGGTGGTTGACTACAACAATACTTCC
GGCTTTTTCAACGATACCATTGACCCGAGACAACGCTGCGTGGCCATATGGACTT
ACAACACACCGGAGTCCGAGGAGCAGTACATCTCGTATAGCCTGGACGGTGGAT
ACACTTTTACAGAGTATCAGAAGAACCCTGTGCTTGCTGCAAATTCGACTCAGTT
CCGAGATCCGAAGGTCTTTTGGTACGAGCCCTCGCAGAAGTGGATCATGACAGC
GGCAAAGTCACAGGACTACAAGATCGAAATTTACTCGTCTGACGACCTTAAATCC
TGGAAGCTCGAATCCGCGTTCGCAAACGAGGGCTTTCTCGGCTACCAATACGAAT
GCCCAGGCCTGATAGAGGTCCCAACAGAGCAAGATCCCAGCAAGTCCTACTGGG
TGATGTTTATTTCCATTAATCCAGGAGCACCGGCAGGAGGTTCTTTTAATCAGTA
CTTCGTCGGAAGCTTTAACGGAACTCATTTCGAGGCATTTGATAACCAATCAAGA
GTAGTTGATTTTGGAAAGGACTACTATGCCCTGCAGACTTTCTTCAATACTGACC
CGACCTATGGGAGCGCTCTTGGCATTGCGTGGGCTTCTAACTGGGAGTATTCCGC
ATTCGTTCCTACAAACCCTTGGAGGTCCTCCATGTCGCTCGTGAGGAAATTCTCTC
TCAACACTGAGTACCAGGCCAACCCGGAAACCGAACTCATAAACCTGAAAGCCG
AACCGATCCTGAACATTAGCAACGCTGGCCCCTGGAGCCGGTTTGCAACCAACA
CCACGTTGACGAAAGCCAACAGCTACAACGTCGATCTTTCGAATAGCACCGGTA
CACTTGAATTTGAACTGGTGTATGCCGTCAATACCACCCAAACGATCTCGAAGTC
GGTGTTCGCGGACCTCTCCCTCTGGTTTAAAGGCCTGGAAGACCCCGAGGAGTAC
CTCAGAATGGGTTTCGAGGTTTCTGCGTCCTCCTTCTTCCTTGATCGCGGGAACAG
CAAAGTAAAATTTGTTAAGGAGAACCCATATTTTACCAACAGGATGAGCGTTAA
CAACCAACCATTCAAGAGCGAAAACGACCTGTCGTACTACAAAGTGTATGGTTT
GCTTGATCAAAATATCCTGGAACTCTACTTCAACGATGGTGATGTCGTGTCCACC
AACACATACTTCATGACAACCGGGAACGCACTGGGCTCCGTGAACATGACGACG
GGTGTGGATAACCTGTTCTACATCGACAAATTCCAGGTGAGGGAAGTCAAGTGA
GATCTGTCGATCGACAAGCTCGAGTTTCTCCATAATAATGTGTGAGTAGTTCCCA
GATAAGGGAATTAGGGTTCCTATAGGGTTTCGCTCATGTGTTGAGCATATAAGAA
ACCCTTAGTATGTATTTGTATTTGTAAAATACTTCTATCAATAAAATTTCTAATTC
CTAAAACCAAAATCCAGTACTAAAATCCAGATCCCCCGAATTAA

[0993]

[0994] 서열 번호 9

TGTTGAAGAATGAGCCGGCGAC

[0995]

[0996] 서열 번호 10

CAGTGAGCTATTACGCACTC

[0997]

[0998] 서열 번호 11

[0999] UTEX 329 프로토테카 크로에가니

TGTTGAAGAATGAGCCGGCGAGTTAAAAAGAGTGGCATGGTTAAAGAAAATACT
CTGGAGCCATAGCGAAAGCAAGTTTAGTAAGCTTAGGTCATTCTTTTTAGACCCG
AAACCGAGTGATCTACCCATGATCAGGGTGAAGTGTTAGTAAAATAACATGGAG
GCCCGAACCGACTAATGTTGAAAAATTAGCGGATGAATTGTGGGTAGGGGCGAA
AAACCAATCGAACTCGGAGTTAGCTGGTTCTCCCCGAAATGCGTTTAGGCGCAGC
AGTAGCAGTACAAATAGAGGGGTAAAGCACTGTTTCTTTTGTGGGCTTCGAAAGT
TGTACCTCAAAGTGGCAAACTCTGAATACTCTATTTAGATATCTACTAGTGAGAC
CTTGGGGGATAAGCTCCTTGGTCAAAAGGGAAACAGCCCAGATCACCAGTTAAG
GCCCCAAAATGAAAATGATAGTGACTAAGGATGTGGGTATGTCAAAACCTCCAG
CAGGTTAGCTTAGAAGCAGCAATCCTTTCAAGAGTGCGTAATAGCTCACTG

[1000]

[1001] 서열 번호 12

[1002] UTEX 1440 프로토테카 윅커해미

TGTTGAAGAATGAGCCGGCGACTTAAAATAAATGGCAGGCTAAGAGATTTAATA
ACTCGAAACCTAAGCGAAAGCAAGTCTTAATAGGGCGTCAATTTAACAAAACTT
TAAATAAATTATAAAGTCATTTATTTTAGACCCGAACCTGAGTGATCTAACCATG
GTCAGGATGAAACTTGGGTGACACCAAGTGGAAGTCCGAACCGACCGATGTTGA
AAAATCGGCGGATGAACTGTGGTTAGTGGTGAAATACCAGTCGAACTCAGAGCT
AGCTGGTTCTCCCCGAAATGCGTTGAGGCGCAGCAATATATCTCGTCTATCTAGG
GGTAAAGCACTGTTTCGGTGCGGGCTATGAAAATGGTACCAAATCGTGGCAAAC
TCTGAATACTAGAAATGACGATATATTAGTGAGACTATGGGGGATAAGCTCCAT
AGTCGAGAGGGAAACAGCCCAGACCACCAGTTAAGGCCCCAAAATGATAATGAA
GTGGTAAAGGAGGTGAAAATGCAAATACAACCAGGAGGTTGGCTTAGAAGCAGC
CATCCTTTAAAGAGTGCGTAATAGCTCACTG

[1003]

[1004] 서열 번호 13

[1005] UTEX 1442 프로토테카 스태그노라

TGTTGAAGAATGAGCCGGCGAGTTAAAAAAAATGGCATGGTTAAAGATATTTCT
CTGAAGCCATAGCGAAAGCAAGTTTTACAAGCTATAGTCATTTTTTTTAGACCCG
AAACCGAGTGATCTACCCATGATCAGGGTGAAGTGTTGGTCAAATAACATGGAG
GCCCGAACCGACTAATGGTGAAAAATTAGCGGATGAATTGTGGGTAGGGGCGAA
AAACCAATCGAACTCGGAGTTAGCTGGTTCTCCCCGAAATGCGTTTAGGCGCAGC
AGTAGCAACACAAATAGAGGGGTAAAGCACTGTTTCTTTTGTGGGCTTCGAAAGT
TGTACCTCAAAGTGGCAAACTCTGAATACTCTATTTAGATATCTACTAGTGAGAC
CTTGGGGGATAAGCTCCTTGGTCAAAAGGGAAACAGCCCAGATCACCAGTTAAG
GCCCCAAAATGAAAATGATAGTGACTAAGGACGTGAGTATGTCAAAACCTCCAG
CAGGTTAGCTTAGAAGCAGCAATCCTTTCAAGAGTGCGTAATAGCTCACTG

[1006]

[1007] 서열 번호 14

[1008] UTEX 288 프로토테카 모리포르미스

TGTTGAAGAATGAGCCGGCGAGTTAAAAAGAGTGGCATGGTTAAAGATAATTCT
CTGGAGCCATAGCGAAAGCAAGTTTAACAAGCTAAAGTCACCCTTTTTAGACCCG
AAACCGAGTGATCTACCCATGATCAGGGTGAAGTGTTGGTAAAATAACATGGAG

[1009]

GCCCGAACCGACTAATGGTGAAAAATTAGCGGATGAATTGTGGGTAGGGGCGAA
AAACCAATCGAACTCGGAGTTAGCTGGTTCTCCCCGAAATGCGTTTAGGCGCAGC
AGTAGCAACACAAATAGAGGGGTAAAGCACTGTTTCTTTTGTGGGCTTCGAAAGT
TGTACCTCAAAGTGGCAAACTCTGAATACTCTATTTAGATATCTACTAGTGAGAC
CTTGGGGGATAAGCTCCTTGGTCAAAAGGGAAACAGCCCAGATCACCAGTTAAG
GCCCCAAAATGAAAATGATAGTGACTAAGGATGTGGGTATGTTAAAACCTCCAG
CAGGTTAGCTTAGAAGCAGCAATCCTTTCAAGAGTGCGTAATAGCTCACTG

[1010]

[1011] 서열 번호 15

[1012] UTEX 1439, UTEX 1441, UTEX 1435, UTEX 1437 프로토테카 모리포르미스

TGTTGAAGAATGAGCCGGCGACTTAAAATAAATGGCAGGCTAAGAGAATTAATA
ACTCGAAACCTAAGCGAAAGCAAGTCTTAATAGGGCGCTAATTTAACAAAACAT
TAAATAAAATCTAAAGTCATTTATTTTAGACCCGAACCTGAGTGATCTAACCATG
GTCAGGATGAAACTTGGGTGACACCAAGTGGAAGTCCGAACCGACCGATGTTGA
AAAATCGGCGGATGAACTGTGGTTAGTGGTGAAATACCAGTCGAACTCAGAGCT
AGCTGGTTCTCCCCGAAATGCGTTGAGGCGCAGCAATATATCTCGTCTATCTAGG
GGTAAAGCACTGTTTCGGTGCGGGCTATGAAAATGGTACCAAATCGTGGCAAAC
TCTGAATACTAGAAATGACGATATATTAGTGAGACTATGGGGGATAAGCTCCAT
AGTCGAGAGGGAAACAGCCCAGACCACCAGTTAAGGCCCCAAAATGATAATGAA
GTGGTAAAGGAGGTGAAAATGCAAATACAACCAGGAGGTTGGCTTAGAAGCAGC
CATCCTTTAAAGAGTGCGTAATAGCTCACTG

[1013]

[1014] 서열 번호 16

[1015] UTEX 1533 프로토테카 윅커해미

TGTTGAAGAATGAGCCGTCGACTTAAAATAAATGGCAGGCTAAGAGAATTAATA
ACTCGAAACCTAAGCGAAAGCAAGTCTTAATAGGGCGCTAATTTAACAAAACAT
TAAATAAAATCTAAAGTCATTTATTTTAGACCCGAACCTGAGTGATCTAACCATG
GTCAGGATGAAACTTGGGTGACACCAAGTGGAAGTCCGAACCGACCGATGTTGA
AAAATCGGCGGATGAACTGTGGTTAGTGGTGAAATACCAGTCGAACTCAGAGCT
AGCTGGTTCTCCCCGAAATGCGTTGAGGCGCAGCAATATATCTCGTCTATCTAGG
GGTAAAGCACTGTTTCGGTGCGGGCTATGAAAATGGTACCAAATCGTGGCAAAC
TCTGAATACTAGAAATGACGATATATTAGTGAGACTATGGGGGATAAGCTCCAT
AGTCGAGAGGGAAACAGCCCAGACCACCAGTTAAGGCCCCAAAATGATAATGAA
GTGGTAAAGGAGGTGAAAATGCAAATACAACCAGGAGGTTGGCTTAGAAGCAGC
CATCCTTTAAAGAGTGCGTAATAGCTCACTG

[1016]

[1017] 서열 번호 17

[1018] UTEX 1434 프로토테카 모리포르미스

TGTTGAAGAATGAGCCGGCGAGTTAAAAAGAGTGGCGTGGTTAAAGAAAATTCT
CTGGAACCATAGCGAAAGCAAGTTTAACAAGCTTAAGTCACTTTTTTTAGACCCG
AAACCGAGTGATCTACCCATGATCAGGGTGAAGTGTTGGTAAAATAACATGGAG
GCCCGAACCGACTAATGGTGAAAAATTAGCGGATGAATTGTGGGTAGGGGCGAA
AAACCAATCGAACTCGGAGTTAGCTGGTTCTCCCCGAAATGCGTTTAGGCGCAGC
AGTAGCAACACAAATAGAGGGGTAAAGCACTGTTTCTTTTGTGGGCTCCGAAAG
TTGTACCTCAAAGTGGCAAACTCTGAATACTCTATTTAGATATCTACTAGTGAGA
CCTTGGGGGGATAAGCTCCTTGGTCGAAAGGGAAACAGCCCAGATCACCAGTTAA

[1019]

GGCCCCAAAATGAAAATGATAGTGACTAAGGATGTGAGTATGTCAAAACCTCCA
GCAGGTTAGCTTAGAAGCAGCAATCCTTTCAAGAGTGCGTAATAGCTCACTG

[1020]

[1021] 서열 번호 18

[1022] UTEX 1438 프로토테카 조프피이

TGTTGAAGAATGAGCCGGCGAGTTAAAAAGAGTGGCATGGTTAAAGAAAATTCT
CTGGAGCCATAGCGAAAGCAAGTTTAACAAGCTTAAGTCACTTTTTTTAGACCCG
AAACCGAGTGATCTACCCATGATCAGGGTGAAGTGTTGGTAAAATAACATGGAG
GCCCGAACCGACTAATGGTGAAAAATTAGCGGATGAATTGTGGGTAGGGGCGAA
AAACCAATCGAACTCGGAGTTAGCTGGTTCTCCCCGAAATGCGTTTAGGCGCAGC
AGTAGCAACACAAATAGAGGGGTAAAGCACTGTTTCTTTCGTGGGCTTCGAAAG
TTGTACCTCAAAGTGGCAAACTCTGAATACTCTATTTAGATATCTACTAGTGAGA
CCTTGGGGGGATAAGCTCCTTGGTCAAAAGGGAAACAGCCCAGATCACCAGTTAA
GGCCCCAAAATGAAAATGATAGTGACTAAGGATGTGAGTATGTCAAAACCTCCA
GCAGGTTAGCTTAGAAGCAGCAATCCTTTCAAGAGTGCGTAATAGCTCACTG

[1023]

[1024] 서열 번호 19

[1025] UTEX 1436 프로토테카 모리포르미스

TGTTGAAGAATGAGCCGGCGACTTAGAAAAGGTGGCATGGTTAAGGAAATATTC
CGAAGCCGTAGCAAAAGCGAGTCTGAATAGGGCGATAAAATATATTAATATTTA
GAATCTAGTCATTTTTTCTAGACCCGAACCCGGGTGATCTAACCATGACCAGGAT
GAAGCTTGGGTGATACCAAGTGAAGGTCCGAACCGACCGATGTTGAAAAATCGG
CGGATGAGTTGTGGTTAGCGGTGAAATACCAGTCGAACCCGGAGCTAGCTGGTT
CTCCCCGAAATGCGTTGAGGCGCAGCAGTACATCTAGTCTATCTAGGGGTAAAGC
ACTGTTTCGGTGCGGGCTGTGAGAACGGTACCAAATCGTGGCAAACTCTGAATAC
TAGAAATGACGATGTAGTAGTGAGACTGTGGGGGATAAGCTCCATTGTCAAGAG
GGAAACAGCCCAGACCACCAGCTAAGGCCCCAAAATGGTAATGTAGTGACAAAG
GAGGTGAAAATGCAAATACAACCAGGAGGTTGGCTTAGAAGCAGCCATCCTTTA
AAGAGTGCGTAATAGCTCACTG

[1026]

[1027] 서열 번호 20

[1028] 치코리움 인티부스 인버타제: Genbank 수탁 번호 Y11124

MSNSSNASESLFPATSEQPYRTAFHFQPPQNWMNDPNGPMCYNGVYHLFYQYNPFG
PLWNLRMYWAHSVSHDLINWIHLDLAFAPTEPFDINGCLSGSATVLPGNKPIMLYTG
IDTENRQVQNLAVPKDLSDPYLREWVKHTGNPIISLPEEIQPDDFRDPTTTWLEEDGT
WRLLVGSQKDKTGIAFLYHSGDFVNWTKSDSPLHKVSGTGMWECVDFFPVWVDST
NGVDTSIINPSNRVKHVLKLGIQDHGKDCYLIGKYSADKENYVPEDELTLSTLRLDY
GMYYASKSFFDPVKNRRIMTAWVNESDSEADVIARGWSGVQSFPRSLWLDKNQKQ
LLQWPIEEIEMLHQNEVSFHNKKLDGGSSLEVLGITASQADVKISFKLANLEEAEELD
PSWVDPQLICSENDASKKGKFGPFGLLALASSDLREQTAIFFRVFRKNGRYVVLMCS
DQSRSSMKNGIEKRTYGAFVDIDPQQDEISLRTLIDHSIVESFGGRGKTCITTRVYPTL
AIGEQARLFAFNHGTESVEISELSAWSMKKAQMKVEEP

[1029]

[1030] 서열 번호 21

[1031] 스키조사카로마이세스 폼베 인버타제: Genbank 수탁 번호 AB011433

MFLKYILASGICLVSLLSSTNAAPRHLYVKRYPVIYNASNITEVSNSTTVPPPPFVNTT
APNGTCLGNYNEYLPSGYYNATDRPKIHFTPSSGFMNDPNGLVYTGGVYHMFFQYS
PKTLTAGEVHWGHTVSKDLIHWENYPIAIYPDEHENGVLSLPFSGSAVVDVHNSSGL
FSNDTIPEERIVLIYTDHWTGVAERQAIAYTTDGGYTFKKYSGNPVLDINSLQFRDPK
VIWDFDANRWVMIVAMSQNYGIAFYSSYDLIHWTELSVFSTSGYLGLQYECPGMAR
VPVEGTDEYKWVLFISINPGAPLGGSVVQYFVGDWNGTNFVPDDGQTRFVDLGKDF
YASALYHSSSANADVIGVGWASNWQYTNQAPTQVFRSAMTVARKFTLRDVPQNPM
TNLTSLIQTPLNVSLLRDETLFTAPVINSSSSLSGSPITLPSNTAFEFNVTLSINYTEGCT
TGYCLGRIIIDSDDPYRLQSISVDVDFAASTLVINRAKAQMGWFNSLFTPSFANDIYIY
GNVTLYGIVDNGLLELYVNNGEKTYTNDFFFLQGATPGQISFAAFQGVSFNNVTVTP
LKTIWNC

[1032]

[1033] 서열 번호 22

[1034] 피카 아노말라 베타-프럭토푸라노시다제 (인버타제): Genank 수탁 번호 X80640

MIQLSPLLLLPLFSVFNSIADASTEYLRPQIHLTPDQGWMNDPNGMFYDRKDKLWHV
YFQHNPDKKSIWATPVTWGHSTSKDLLTWDYHGNALEPENDDEGIFSGSVVVDRNN
TSGFFNDSTDPEQRIVAIYTNNAQLQTQEIAYSLDKGYSFIKYDQNPVINVNSSQQRD
PKVLWHDESNQWIMVVAKTQEFKVQIYGSPDLKKWDLKSNFTSNGYLGFQYECPG
LFKLPIENPLNDTVTSKWVLLLAINPGSPLGGSINEYFIGDFDGTTFHPDDGATRFMDI
GKDFYAFQSFDNTEPEDGALGLAWASNWQYANTVPTENWRSSMSLVRNYTLKYVD
VNPENYGLTLIQKPVYDTKETRLNETLKTLETINEYEVNDLKLDKSSFVATDFNTERN
ATGVFEFDLKFTQTDLKMGYSNMTTQFGLYIHSQTVKGSQETLQLVFDTLSTTWYID
RTTQHSFQRNSPVFTERISTYVEKIDTTDQGNVYTLYGVVDRNILELYFNDGSIAMTN
TFFFREGKIPTSFEVVCDSEKSFITIDELSVRELARK

[1035]

[1036] 서열 번호 23

[1037] 데바리오마이세스 옥시덴탈리스 인버타제: Genank 수탁 번호 X17604

MVQVLSVLVIPLLTLFFGYVASSSIDLSVDTSEYNRPLIHFTPEKGWMNDPNGLFYDK
TAKLWHLYFQYNPNATAWGQPLYWGHATSNDLVHWDEHEIAIGPEHDNEGIFSGSI
VVDHNNTSGFFNSSIDPNQRIVAIYTNNIPDLQTQDIAFSLDGGYTFTKYENNPVIDVS
SNQFRDPKVFWHERFKSMDHGCSEIARVKIQIFGSANLKNWVLNSNFSSGYYGNQY
GMSRLIEVPIENSDKSKWVMFLAINPGSPLGGSINQYFVGDFDGFQFVPDDSQTRFVD
IGKDFYAFQTFSEVEHGVLGLAWASNWQYADQVPTNPWRSSTSLARNYTLRYVIQM
LKLTANIDKSVLPDSINVVDKLKKKNVKLTNKKPIKTNFKGSTGLFDFNITFKVLNLN
VSPGKTHFDILINSQELNSSVDSIKIGFDSSQSLFYIDRHIPNVEFPRKQFFTDKLAAYL
EPLDYDQDLRVFSLYGIVDKNIIELYFNDGTVAMTNTFFMGEGKYPHDIQIVTDTEEP
LFELESVIIRELNK

[1038]

[1039] 서열 번호 24

[1040] 오리자 사티바 인버타제: Genank 수탁 번호 AF019113

MATSRLTPAYDLKNAAAAVYTPLPEQPHSAEVEIRDRKPFKIISAIILSSLLLLALILVA
VNYQAPPSHSSGDNSQPAAVMPPSRGVSQGVSEKAFRGASGAGNGVSFAWSNLMLS
WQRTSYHFQPVKNWMNDPNGPLYYKGWYHLFYQYNPDSAVWGNITWGHAVSTD
LINWLHLPFAMVPDQWYDVNGVWTGSATILPDGRIVMLYTGDTDDYVQDQNLAFP
ANLSDPLLVDWVKYPNNPVIYPPPGIGVKDFRDPTTAGTAGMQNGQRLVTIGSKVG
KTGISLVYETTNFTTFKLLYGVLHAVPGTGMWECVDLYPVSTTGENGLDTSVNGLG
VKHVLKTSLDDDKHDYYALGTYDPVKNKWTPDNPDLDVGIGLRLDYGKYYAARTF
[1041]

YDQNKQRRILWGWIGETDLEAVDLMKGWASLQAIPRTIVFDKKTGTNVLQRPEEEV
ESWSSGDPITQRRIFEPGSVVPIHVSGATQLDITASFEVDETLLETTSESHDAGYDCSN
SGGAGTRGSLGPFGLLVVADEKLSELTPVYLYVAKGGDGKAKAHLCAYQTRSSMAS
GVEKEVYGSAVPVLDGENYSARILIDHSIVESFAQAGRTCVRSRDYPTKDIYGAARCF
FFNNATEASVRASLKAWQMKSFIRPYPFIPDQKS
[1042]

[1043] 서열 번호 25

[1044] 알리움 케파 인버타제: Genank 수탁 번호 AJ006067

MSSDDLESPPSSYLPIPPSDEFHDQPPPLRSWLRLLSIPLALMFLLFLATFLSNLESPPSD
SGLVSDPVTFDVNPAVVRRGKDAGVSDKTSGVDSGFVLDPVAVDANSVVVHRGKD
AGVSDKTSGVDSGLLKDSPLGPYPWTNQMLSWQRTGFHFQPVKNWMNDPNGPLYY
KGWYHFFYQYNPEGAVWGNIAWGHAVSRDLVHWTHLPLAMVPDQWYDINGVWT
GSATILPDGQIVMLYTGATNESVQVQNLAVPADQSDTLLLRWKKSEANPILVPPPGIG
DKDFRDPTTAWYEPSDDTWRIVIGSKDSSHSGIAIVYSTKDFINYKLIPGILHAVERVG
MWECVDFYPVATADSSHANHGLDPSARPSPAVKHVLKASMDDDRHDYYAIGTYDP
AQNTWVPDDASVDVGIGLRYDWGKFYASKTFYDHAKKRRILWSWIGETDSETADIA
KGWASLQGVPRTVLLDVKTGSNLITWPVVEIESLRTRPRDFSGITVDAGSTFKLDVG
GAAQLDIEAEFKISSEELEAVKEADVSYNCSSSGGAAERGVLGPFGLLVLANQDLTE
QTATYFYVSRGMDGGLNTHFCQDEKRSSKASDIVKRIVGHSVPVLDGESFALRILVD
HSIVESFAQGGRASATSRVYPTEAIYNNARVFVFNNATGAKVTAQSLKVWHMSTAI
NEIYDPATSVM
[1045]

[1046] 서열 번호 26

[1047] 베타 불가리스 아종 불가리스 인버타제: Genank 수탁 번호 AJ278531

LFYQYNPNGVIWGPPVWGHSTSKDLVNWVPQPLTMEPEMAANINGSWSGSATILPG
NKPAILFTGLDPKYEQVQVLAYPKDTSDPNLKEWFLAPQNPVMFPTPQNQINATSFR
DPTTAWRLPDGVWRLLIGSKRGQRGLSLLFRSRDFVHWVQAKHPLYSDKLSGMWE
CPDFFPVYANGDQMGVDTSIIGSHVKHVLKNSLDITKHDIYTIGDYNIKKDAYTPDIG
YMNDSSLRYDYGKYYASKTFFDDAKKERILLGWANESSSVEDDIKKGWSGIHTIPRK
IWLDKLGKQLIQWPIANIEKLRQKPVNIYRKVLKGGSQIEVSGITAAQADVEISFKIKD
LKNVEKFDASWTSPQLLCSKKGASVKGGLGPFGLLTLASXGLEEYTAVFFRIFKAYD
NKFVVLMCSDQSRSSLNPTNDKTTYGTFVDVNPIREGLSLRVLIDHSVVESFGAKGK
NVITARVYPTLAINEKAHLYVFNRGTSNVEITGLTAWSMKKANIA
[1048]

[1049] 서열 번호 27

[1050] 비피도박테리움 브레베 UCC2003 베타-프럭토푸라노시다제 (인버타제): Genank 수탁 번호 AAT28190

MTDFTPETPVLTPIRDHAAELAKAEAGVAEMAAKRNNRWYPKYHIASNGGWINDPN
GLCFYKGRWHVFYQLHPYGTQWGPMHWGHVSSTDMLNWKREPIMFAPSLEQEKD
GVFSGSAVIDDNGDLRFYYTGHRWANGHDNTGGDWQVQMTALPDNDELTSATKQ
GMIIDCPTDKVDHHYRDPKVWKTGDTWYMTFGVSSEDKRGQMWLFSSKDMVRWE
YERVLFQHPDPDVFMLECPDFFPIKDKDGNEKWVIGFSAMGSKPSGFMNRNVNNAG
YMIGTWEPGGEFKPETEFRLWDCGHNYYAPQSFNVDGRQIVYGWMSPFVQPIPMED
DGWCGQLTLPREITLDDDGDVVTAPVAEMEGLREDTLDHGSITLDMDGEQVIADDA
EAVEIEMTIDLAASTADRAGLKIHATEDGAYTYVAYDDQIGRVVVDRQAMANGDH
GYRAAPLTDAELASGKLDLRVFVDRGSVEVYVNGGHQVLSSYSYASEGPRAIKLVA
EFGNLKVESLKLHHMKSIGLE
[1051]

[1052] 서열 번호 28

[1053] 사카로마이세스 세레비지아에 인버타제: Genank 수탁 번호 NP_012104

MLLQAFLFLLAGFAAKISASMTNETSDRPLVHFTPNKGWMNDPNGLWYDEKDAKW
HLYFQYNPNDTVWGTPLFWGHATSDDLTNWEDQPIAIAPKRNDSGAFSGSMVVDY
NNTSGFFNDTIDPRQRCVAIWTYNTPESEEQYISYSLDGGYTFTEYQKNPVLAANSTQ
FRDPKVFWYEPSQKWIMTAAKSQDYKIEIYSSDDLKSWKLESAFANEGFLGYQYECP
GLIEVPTEQDPSKSYWVMFISINPGAPAGGSFNQYFVGSFNGTHFEAFDNQSRVVDFG
KDYYALQTFFNTDPTYGSALGIAWASNWEYSAFVPTNPWRSSMSLVRKFSLNTEYQ
ANPETELINLKAEPILNISNAGPWSRFATNTTLTKANSYNVDLSNSTGTLEFELVYAV
NTTQTISKSVFADLSLWFKGLEDPEEYLRMGFEVSASSFFLDRGNSKVKFVKENPYFT
NRMSVNNQPFKSENDLSYYKVYGLLDQNILELYFNDGDVVSTNTYFMTTGNALGSV
NMTTGVDNLFYIDKFQVREVK

[1054]

[1055] 서열 번호 29

[1056] 지모모나스 모빌리스 인버타제 A: Genank 수탁 번호 AY171597

MESPSYKNLIKAEDAQKKAGKRLLSSEWYPGFHVTPLTGWMNDPNGLIFFKGEYHL
FYQYYPFAPVWGPMHWGHAKSRDLVHWETLPVALAPGDLFDRDGCFSGCAVDNN
GVLTLIYTGHIVLSNDSPDAIREVQCMATSIDGIHFQKEGIVLEKAPMPQVAHFRDPR
VWKENDHWFMVVGYRTDDEKHQGIGHVALYRSENLKDWIFVKTLLGDNSQLPLGK
RAFMWECPDFFSLGNRSVLMFSPQGLKASGYKNRNLFQNGYILGKWQAPQFTPETS
FQELDYGHDFYAAQRFEAKDGRQILIAWFDMWENQKPSQRDGWAGCMTLPRKLDL
IDNKIVMTPVREMEILRQSEKIESVVTLSDAEHPFTMDSPLQEIELIFDLEKSSAYQAG
LALRCNGKGQETLLYIDRSQNRIILDRNRSGQNVKGIRSCPLPNTSKVRLHIFLDRSSIE
IFVGDDQTQGLYSISSRIFPDKDSLKGRLFAIEGYAVFDSFKRWTLQDANLAAFSSDA
C

[1057]

[1058] 서열 번호 30

[1059] 프로토테카 모리포르미스 델타 12 지방산 불포화 효소 운반 펩티드를 지니 는 신나모뮴 캠포라 FATB1 (Genbank Q39473) 아미노산 서열

MAIKTNRQPVEKPPFTIGTLRKAIPAHCFERSALRGRAPDWSMLFAVITTIFSAAEKQ
WTNLEWKPKPNPPQLLDDHFGPHGLVFRRTFAIRSYEVGPDRSTSIVAVMNHLQEAA
LNHAKSVGILGDGFGTTLEMSKRDLIWVVKRTHVAVERYPAWGDTVEVECWVGAS
GNNGRRHDFLVRDCKTGEILTRCTSLSVMMNTRTRRLSKIPEEVRGEIGPAFIDNVAV
KDEEIKKPQKLNDSTADYIQGGLTPRWNDLDINQHVNNIKYVDWILETVPDSIFESHH
ISSFTIEYRRECTMDSVLQSLTTVSGGSSEAGLVCEHLLQLEGGSEVLRAKTEWRPKL
TDSFRGISVIPAESSV

[1060]

[1061] 서열 번호 31

[1062] 프로토테카 모리포르미스 델타 12 지방산 불포화 효소 운반 펩티드를 지니는 신나모뮴 캠포라 FATB1 cDNA의 관련 코돈 최적화 발현 구조물

GGTACCCGCCTGCAACGCAAGGGCAGCCACAGCCGCTCCCACCCGCCGCTGAAC
CGACACGTGCTTGGGCGCCTGCCGCCTGCCTGCCGCATGCTTGTGCTGGTGAGGC
TGGGCAGTGCTGCCATGCTGATTGAGGCTTGGTTCATCGGGTGGAAGCTTATGTG
TGTGCTGGGCTTGCATGCCGGGCAATGCGCATGGTGGCAAGAGGGCGGCAGCAC
TTGCTGGAGCTGCCGCGGTGCCTCCAGGTGGTTCAATCGCGGCAGCCAGAGGGA

[1063]

TTTCAGATGATCGCGCGTACAGGTTGAGCAGCAGTGTCAGCAAAGGTAGCAGTTT
GCCAGAATGATCGGTTCAGCTGTTAATCAATGCCAGCAAGAGAAGGGGTCAAGT
GCAAACACGGGCATGCCACAGCACGGGCACCGGGGAGTGGAATGGCACCACCA
AGTGTGTGCGAGCCAGCATCGCCGCCTGGCTGTTTCAGCTACAACGGCAGGAGTC
ATCCAACGTAACCATGAGCTGATCAACACTGCAATCATCGGGCGGGCGTGATGC
AAGCATGCCTGGCGAAGACACATGGTGTGCGGATGCTGCCGGCTGCTGCCTGCT
GCGCACGCCGTTGAGTTGGCAGCAGGCTCAGCCATGCACTGGATGGCAGCTGGG
CTGCCACTGCAATGTGGTGGATAGGATGCAAGTGGAGCGAATACCAAACCCTCT
GGCTGCTTGCTGGGTTGCATGGCATCGCACCATCAGCAGGAGCGCATGCGAAGG
GACTGGCCCCATGCACGCCATGCCAAACCGGAGCGCACCGAGTGTCCACACTGT
CACCAGGCCCGCAAGCTTTGCAGAACCATGCTCATGGACGCATGTAGCGCTGAC
GTCCCTTGACGGCGCTCCTCTCGGGTGTGGGAAACGCAATGCAGCACAGGCAGC
AGAGGCGGCGGCAGCAGAGCGGCGGCAGCAGCGGCGGGGGCCACCCTTCTTGCG
GGGTCGCGCCCCAGCCAGCGGTGATGCGCTGATCCCAAACGAGTTCACATTCATT
TGCATGCCTGGAGAAGCGAGGCTGGGGCCTTTGGGCTGGTGCAGCCCGCAATGG
AATGCGGGACCGCCAGGCTAGCAGCAAAGGCGCCTCCCCTACTCCGCATCGATG
TTCCATAGTGCATTGGACTGCATTTGGGTGGGGCGGCCGGCTGTTTCTTTCGTGTT
GCAAAACGCGCCAGCTCAGCAACCTGTCCCGTGGGTCCCCCGTGCCGATGAAAT
CGTGTGCACGCCGATCAGCTGATTGCCCGGCTCGCGAAGTAGGCGCCCTCCTTTC
TGCTCGCCCTCTCTCCGTCCCGCCTCTAGAATATCAATGATCGAGCAGGACGGCC
TCCACGCCGGCTCCCCCGCCGCCTGGGTGGAGCGCCTGTTCGGCTACGACTGGGC
CCAGCAGACCATCGGCTGCTCCGACGCCGCCGTGTTCCGCCTGTCCGCCCAGGGC
CGCCCCGTGCTGTTCGTGAAGACCGACCTGTCCGGCGCCCTGAACGAGCTGCAGG
ACGAGGCCGCCCGCCTGTCCTGGCTGGCCACCACCGGCGTGCCCTGCGCCGCCGT
GCTGGACGTGGTGACCGAGGCCGGCCGCGACTGGCTGCTGCTGGGCGAGGTGCC
CGGCCAGGACCTGCTGTCCTCCCACCTGGCCCCCGCCGAGAAGGTGTCCATCATG
GCCGACGCCATGCGCCGCCTGCACACCCTGGACCCCGCCACCTGCCCCTTCGACC
ACCAGGCCAAGCACCGCATCGAGCGCGCCCGCACCCGCATGGAGGCCGGCCTGG
TGGACCAGGACGACCTGGACGAGGAGCACCAGGGCCTGGCCCCCGCCGAGCTGT
TCGCCCGCCTGAAGGCCCGCATGCCCGACGGCGAGGACCTGGTGGTGACCCACG
GCGACGCCTGCCTGCCCAACATCATGGTGGAGAACGGCCGCTTCTCCGGCTTCAT
CGACTGCGGCCGCCTGGGCGTGGCCGACCGCTACCAGGACATCGCCCTGGCCAC
CCGCGACATCGCCGAGGAGCTGGGCGGCGAGTGGGCCGACCGCTTCCTGGTGCT
GTACGGCATCGCCGCCCCCGACTCCCAGCGCATCGCCTTCTACCGCCTGCTGGAC
GAGTTCTTCTGACAATTGGCAGCAGCAGCTCGGATAGTATCGACACACTCTGGAC
GCTGGTCGTGTGATGGACTGTTGCCGCCACACTTGCTGCCTTGACCTGTGAATAT
CCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGTGTGTACGCGCTTT
TGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCCAGCATCCCCTTCC
CTCGTTTCATATCGCTTGCATCCCAACCGCAACTTATCTACGCTGTCCTGCTATCC
CTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCTCGCACAGCCTTGGTTTGGGCTC
CGCCTGTATTCTCCTGGTACTGCAACCTGTAAACCAGCACTGCAATGCTGATGCA
CGGGAAGTAGTGGGATGGGAACACAAATGGAGGATCCCGCGTCTCGAACAGAGC
GCGCAGAGGAACGCTGAAGGTCTCGCCTCTGTCGCACCTCAGCGCGGCATACAC
CACAATAACCACCTGACGAATGCGCTTGGTTCTTCGTCCATTAGCGAAGCGTCCG
GTTCACACACGTGCCACGTTGGCGAGGTGGCAGGTGACAATGATCGGTGGAGCT
GATGGTCGAAACGTTCACAGCCTAGGGATATCGAATTCCTTTCTTGCGCTATGAC
ACTTCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTTCCCGGCGCTGCATGCA
ACACCGATGATGCTTCGACCCCCCGAAGCTCCTTCGGGGCTGCATGGGCGCTCCG
ATGCCGCTCCAGGGCGAGCGCTGTTTAAATAGCCAGGCCCCCGATTGCAAAGAC
ATTATAGCGAGCTACCAAAGCCATATTCAAACACCTAGATCACTACCACTTCTAC

[1064]

ACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGGGCGCCTCTTCCTCTTC
GTTTCAGTCACAACCCGCAAACACTAGTATGGCTATCAAGACGAACAGGCAGCC
TGTGGAGAAGCCTCCGTTCACGATCGGGACGCTGCGCAAGGCCATCCCCGCGCA
CTGTTTCGAGCGCTCGGCGCTTCGTGGGCGCGCCCCCGACTGGTCCATGCTGTTC
GCCGTGATCACCACCATCTTCTCCGCCGCCGAGAAGCAGTGGACCAACCTGGAGT
GGAAGCCCAAGCCCAACCCCCCCCCAGCTGCTGGACGACCACTTCGGCCCCCACG
GCCTGGTGTTCCGCCGCACCTTCGCCATCCGCAGCTACGAGGTGGGCCCCGACCG
CTCCACCAGCATCGTGGCCGTGATGAACCACCTGCAGGAGGCCGCCCTGAACCA
CGCCAAGTCCGTGGGCATCCTGGGCGACGGCTTCGGCACCACCCTGGAGATGTCC
AAGCGCGACCTGATCTGGGTGGTGAAGCGCACCCACGTGGCCGTGGAGCGCTAC
CCCGCCTGGGGCGACACCGTGGAGGTGGAGTGCTGGGTGGGCGCCTCCGGCAAC
AACGGCCGCCGCCACGACTTCCTGGTGCGCGACTGCAAGACCGGCGAGATCCTG
ACCCGCTGCACCTCCCTGAGCGTGATGATGAACACCCGCACCCGCCGCCTGAGCA
AGATCCCCGAGGAGGTGCGCGGCGAGATCGGCCCCGCCTTCATCGACAACGTGG
CCGTGAAGGACGAGGAGATCAAGAAGCCCCAGAAGCTGAACGACTCCACCGCCG
ACTACATCCAGGGCGGCCTGACCCCCCGCTGGAACGACCTGGACATCAACCAGC
ACGTGAACAACATCAAGTACGTGGACTGGATCCTGGAGACCGTGCCCGACAGCA
TCTTCGAGAGCCACCACATCTCCTCCTTCACCATCGAGTACCGCCGCGAGTGCAC
CATGGACAGCGTGCTGCAGTCCCTGACCACCGTGAGCGGCGGCTCCTCCGAGGC
CGGCCTGGTGTGCGAGCACCTGCTGCAGCTGGAGGGCGGCAGCGAGGTGCTGCG
CGCCAAGACCGAGTGGCGCCCCAAGCTGACCGACTCCTTCCGCGGCATCAGCGT
GATCCCCGCCGAGTCCAGCGTGATGGACTACAAGGACCACGACGGCGACTACAA
GGACCACGACATCGACTACAAGGACGACGACGACAAGTGATGACTCGAGGCAGC
AGCAGCTCGGATAGTATCGACACACTCTGGACGCTGGTCGTGTGATGGACTGTTG
CCGCCACACTTGCTGCCTTGACCTGTGAATATCCCTGCCGCTTTTATCAAACAGCC
TCAGTGTGTTTGATCTTGTGTGTACGCGCTTTTGCGAGTTGCTAGCTGCTTGTGCT
ATTTGCGAATACCACCCCCAGCATCCCCTTCCCTCGTTTCATATCGCTTGCATCCC
AACCGCAACTTATCTACGCTGTCCTGCTATCCCTCAGCGCTGCTCCTGCTCCTGCT
CACTGCCCCTCGCACAGCCTTGGTTTGGGCTCCGCCTGTATTCTCCTGGTACTGCA
ACCTGTAAACCAGCACTGCAATGCTGATGCACGGGAAGTAGTGGGATGGGAACA
CAAATGGAAAGCTT

[1065]

[1066] 서열 번호 32

[1067] 프로토테카 모리포르미스 델타 12 지방산 불포화 효소 운반 펩티드를 지니는 쿠페아 후커리아나 FATB2 (Genbank AAC49269) 아미노산 서열

MAIKTNRQPVEKPPFTIGTLRKAIPAHCFERSALRGRAQLPDWSRLLTAITTVFVKSK
RPDMHDRKSKRPDMLVDSFGLESTVQDGLVFRQSFSIRSYEIGTDRTASIETLMNHLQ
ETSLNHCKSTGILLDGFGRTLEMCKRDLIWVVIKMQIKVNRYPAWGDTVEINTRFSR
LGKIGMGRDWLISDCNTGEILVRATSAYAMMNQKTRRLSKLPYEVHQEIVPLFVDSP
VIEDSDLKVHKFKVKTGDSIQKGLTPGWNDLDVNQHVSNVKYIGWILESMPTEVLET
QELCSLALEYRRECGRDSVLESVTAMDPSKVGVRSQYQHLLRLEDGTAIVNGATEW
RPKNAGANGAISTGKTSNGNSVS

[1068]

[1069] 서열 번호 33

[1070] 프로토테카 모리포르미스 델타 12 지방산 불포화 효소 운반 펩티드를 지니는 쿠페아 후커리아나 FATB2 cDNA의 관련 코돈 최적화 발현 구조물

GGTACCCGCCTGCAACGCAAGGGCAGCCACAGCCGCTCCCACCCGCCGCTGAAC
CGACACGTGCTTGGGCGCCTGCCGCCTGCCTGCCGCATGCTTGTGCTGGTGAGGC

[1071]

TGGGCAGTGCTGCCATGCTGATTGAGGCTTGGTTCATCGGGTGGAAGCTTATGTG
TGTGCTGGGCTTGCATGCCGGGCAATGCGCATGGTGGCAAGAGGGCGGCAGCAC
TTGCTGGAGCTGCCGCGGTGCCTCCAGGTGGTTCAATCGCGGCAGCCAGAGGGA
TTTCAGATGATCGCGCGTACAGGTTGAGCAGCAGTGTCAGCAAAGGTAGCAGTTT
GCCAGAATGATCGGTTCAGCTGTTAATCAATGCCAGCAAGAGAAGGGGTCAAGT
GCAAACACGGGCATGCCACAGCACGGGCACCGGGGAGTGGAATGGCACCACCA
AGTGTGTGCGAGCCAGCATCGCCGCCTGGCTGTTTCAGCTACAACGGCAGGAGTC
ATCCAACGTAACCATGAGCTGATCAACACTGCAATCATCGGGCGGGCGTGATGC
AAGCATGCCTGGCGAAGACACATGGTGTGCGGATGCTGCCGGCTGCTGCCTGCT
GCGCACGCCGTTGAGTTGGCAGCAGGCTCAGCCATGCACTGGATGGCAGCTGGG
CTGCCACTGCAATGTGGTGGATAGGATGCAAGTGGAGCGAATACCAAACCCTCT
GGCTGCTTGCTGGGTTGCATGGCATCGCACCATCAGCAGGAGCGCATGCGAAGG
GACTGGCCCCATGCACGCCATGCCAAACCGGAGCGCACCGAGTGTCCACACTGT
CACCAGGCCCGCAAGCTTTGCAGAACCATGCTCATGGACGCATGTAGCGCTGAC
GTCCCTTGACGGCGCTCCTCTCGGGTGTGGGAAACGCAATGCAGCACAGGCAGC
AGAGGCGGCGGCAGCAGAGCGGCGGCAGCAGCGGCGGGGGCCACCCTTCTTGCG
GGGTCGCGCCCCAGCCAGCGGTGATGCGCTGATCCCAAACGAGTTCACATTCATT
TGCATGCCTGGAGAAGCGAGGCTGGGGCCTTTGGGCTGGTGCAGCCCGCAATGG
AATGCGGGACCGCCAGGCTAGCAGCAAAGGCGCCTCCCCTACTCCGCATCGATG
TTCCATAGTGCATTGGACTGCATTTGGGTGGGGCGGCCGGCTGTTTCTTTCGTGTT
GCAAAACGCGCCAGCTCAGCAACCTGTCCCGTGGGTCCCCCGTGCCGATGAAAT
CGTGTGCACGCCGATCAGCTGATTGCCCGGCTCGCGAAGTAGGCGCCCTCCTTTC
TGCTCGCCCTCTCTCCGTCCCGCCTCTAGAATATCAATGATCGAGCAGGACGGCC
TCCACGCCGGCTCCCCCGCCGCCTGGGTGGAGCGCCTGTTCGGCTACGACTGGGC
CCAGCAGACCATCGGCTGCTCCGACGCCGCCGTGTTCCGCCTGTCCGCCCAGGGC
CGCCCCGTGCTGTTCGTGAAGACCGACCTGTCCGGCGCCCTGAACGAGCTGCAGG
ACGAGGCCGCCCGCCTGTCCTGGCTGGCCACCACCGGCGTGCCCTGCGCCGCCGT
GCTGGACGTGGTGACCGAGGCCGGCCGCGACTGGCTGCTGCTGGGCGAGGTGCC
CGGCCAGGACCTGCTGTCCTCCCACCTGGCCCCCGCCGAGAAGGTGTCCATCATG
GCCGACGCCATGCGCCGCCTGCACACCCTGGACCCCGCCACCTGCCCCTTCGACC
ACCAGGCCAAGCACCGCATCGAGCGCGCCCGCACCCGCATGGAGGCCGGCCTGG
TGGACCAGGACGACCTGGACGAGGAGCACCAGGGCCTGGCCCCCGCCGAGCTGT
TCGCCCGCCTGAAGGCCCGCATGCCCGACGGCGAGGACCTGGTGGTGACCCACG
GCGACGCCTGCCTGCCCAACATCATGGTGGAGAACGGCCGCTTCTCCGGCTTCAT
CGACTGCGGCCGCCTGGGCGTGGCCGACCGCTACCAGGACATCGCCCTGGCCAC
CCGCGACATCGCCGAGGAGCTGGGCGGCGAGTGGGCCGACCGCTTCCTGGTGCT
GTACGGCATCGCCGCCCCCGACTCCCAGCGCATCGCCTTCTACCGCCTGCTGGAC
GAGTTCTTCTGACAATTGGCAGCAGCAGCTCGGATAGTATCGACACACTCTGGAC
GCTGGTCGTGTGATGGACTGTTGCCGCCACACTTGCTGCCTTGACCTGTGAATAT
CCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGTGTGTACGCGCTTT
TGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCCAGCATCCCCTTCC
CTCGTTTCATATCGCTTGCATCCCAACCGCAACTTATCTACGCTGTCCTGCTATCC
CTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCTCGCACAGCCTTGGTTTGGGCTC
CGCCTGTATTCTCCTGGTACTGCAACCTGTAAACCAGCACTGCAATGCTGATGCA
CGGGAAGTAGTGGGATGGGAACACAAATGGAGGATCCCGCGTCTCGAACAGAGC
GCGCAGAGGAACGCTGAAGGTCTCGCCTCTGTCGCACCTCAGCGCGGCATACAC
CACAATAACCACCTGACGAATGCGCTTGGTTCTTCGTCCATTAGCGAAGCGTCCG
GTTCACACACGTGCCACGTTGGCGAGGTGGCAGGTGACAATGATCGGTGGAGCT
GATGGTCGAAACGTTCACAGCCTAGGGATATCGAATTCCTTTCTTGCGCTATGAC
ACTTCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTTCCCGGCGCTGCATGCA

[1072]

ACACCGATGATGCTTCGACCCCCCGAAGCTCCTTCGGGGCTGCATGGGCGCTCCG
ATGCCGCTCCAGGGCGAGCGCTGTTTAAATAGCCAGGCCCCCGATTGCAAAGAC
ATTATAGCGAGCTACCAAAGCCATATTCAAACACCTAGATCACTACCACTTCTAC
ACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGGGCGCCTCTTCCTCTTC
GTTTCAGTCACAACCCGCAAACACTAGTATGGCTATCAAGACGAACAGGCAGCC
TGTGGAGAAGCCTCCGTTCACGATCGGGACGCTGCGCAAGGCCATCCCCGCGCA
CTGTTTCGAGCGCTCGGCGCTTCGTGGGCGCGCCCAGCTGCCCGACTGGAGCCGC
CTGCTGACCGCCATCACCACCGTGTTCGTGAAGTCCAAGCGCCCCGACATGCACG
ACCGCAAGTCCAAGCGCCCCGACATGCTGGTGGACAGCTTCGGCCTGGAGTCCA
CCGTGCAGGACGGCCTGGTGTTCCGCCAGTCCTTCTCCATCCGCTCCTACGAGAT
CGGCACCGACCGCACCGCCAGCATCGAGACCCTGATGAACCACCTGCAGGAGAC
CTCCCTGAACCACTGCAAGAGCACCGGCATCCTGCTGGACGGCTTCGGCCGCACC
CTGGAGATGTGCAAGCGCGACCTGATCTGGGTGGTGATCAAGATGCAGATCAAG
GTGAACCGCTACCCCGCCTGGGGCGACACCGTGGAGATCAACACCCGCTTCAGC
CGCCTGGGCAAGATCGGCATGGGCCGCGACTGGCTGATCTCCGACTGCAACACC
GGCGAGATCCTGGTGCGCGCCACCAGCGCCTACGCCATGATGAACCAGAAGACC
CGCCGCCTGTCCAAGCTGCCCTACGAGGTGCACCAGGAGATCGTGCCCCTGTTCG
TGGACAGCCCCGTGATCGAGGACTCCGACCTGAAGGTGCACAAGTTCAAGGTGA
AGACCGGCGACAGCATCCAGAAGGGCCTGACCCCCGGCTGGAACGACCTGGACG
TGAACCAGCACGTGTCCAACGTGAAGTACATCGGCTGGATCCTGGAGAGCATGC
CCACCGAGGTGCTGGAGACCCAGGAGCTGTGCTCCCTGGCCCTGGAGTACCGCC
GCGAGTGCGGCCGCGACTCCGTGCTGGAGAGCGTGACCGCCATGGACCCCAGCA
AGGTGGGCGTGCGCTCCCAGTACCAGCACCTGCTGCGCCTGGAGGACGGCACCG
CCATCGTGAACGGCGCCACCGAGTGGCGCCCCAAGAACGCCGGCGCCAACGGCG
CCATCTCCACCGGCAAGACCAGCAACGGCAACTCCGTGTCCATGGACTACAAGG
ACCACGACGGCGACTACAAGGACCACGACATCGACTACAAGGACGACGACGAC
AAGTGACTCGAGGCAGCAGCAGCTCGGATAGTATCGACACACTCTGGACGCTGG
TCGTGTGATGGACTGTTGCCGCCACACTTGCTGCCTTGACCTGTGAATATCCCTGC
CGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGTGTGTACGCGCTTTTGCGAG
TTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCCAGCATCCCCTTCCCTCGTT
TCATATCGCTTGCATCCCAACCGCAACTTATCTACGCTGTCCTGCTATCCCTCAGC
GCTGCTCCTGCTCCTGCTCACTGCCCCTCGCACAGCCTTGGTTTGGGCTCCGCCTG
TATTCTCCTGGTACTGCAACCTGTAAACCAGCACTGCAATGCTGATGCACGGGAA
GTAGTGGGATGGGAACACAAATGGAAAGCTT

[1073]

[1074] 서열 번호 34

[1075] 프로토테카 모리포르미스 델타 12 지방산 불포화 효소 운반 펩티드를 지니는 움벨루라리아 캘리포니카 FATB1 (Genbank Q41635) 아미노산 서열

MAIKTNRQPVEKPPFTIGTLRKAIPAHCFERSALRGRAPDWSMLFAVITTIFSAAEKQ
WTNLEWKPKPKLPQLLDDHFGLHGLVFRRTFAIRSYEVGPDRSTSILAVMNHMQEA
TLNHAKSVGILGDGFGTTLEMSKRDLMWVVRRTHVAVERYPTWGDTVEVECWIGA
SGNNGMRRDFLVRDCKTGEILTRCTSLSVLMNTRTRRLSTIPDEVRGEIGPAFIDNVA
VKDDEIKKLQKLNDSTADYIQGGLTPRWNDLDVNQHVNNLKYVAWVFETVPDSIFE
SHHISSFTLEYRRECTRDSVLRSLTTVSGGSSEAGLVCDHLLQLEGGSEVLRARTEWR
PKLTDSFRGISVIPAEPRV

[1076]

[1077] 서열 번호 35

[1078] 프로토테카 모리포르미스 델타 12 지방산 불포화 효소 운반 펩티드를 지니는 움벨루라리아 캘리포니카 FATB1 cDNA의 관련 코돈 최적화 발현 구조물

GGTACCCGCCTGCAACGCAAGGGCAGCCACAGCCGCTCCCACCCGCCGCTGAAC
CGACACGTGCTTGGGCGCCTGCCGCCTGCCTGCCGCATGCTTGTGCTGGTGAGGC
TGGGCAGTGCTGCCATGCTGATTGAGGCTTGGTTCATCGGGTGGAAGCTTATGTG
TGTGCTGGGCTTGCATGCCGGGCAATGCGCATGGTGGCAAGAGGGCGGCAGCAC
TTGCTGGAGCTGCCGCGGTGCCTCCAGGTGGTTCAATCGCGGCAGCCAGAGGGA
TTTCAGATGATCGCGCGTACAGGTTGAGCAGCAGTGTCAGCAAAGGTAGCAGTTT
GCCAGAATGATCGGTTCAGCTGTTAATCAATGCCAGCAAGAGAAGGGGTCAAGT
GCAAACACGGGCATGCCACAGCACGGGCACCGGGGAGTGGAATGGCACCACCA
AGTGTGTGCGAGCCAGCATCGCCGCCTGGCTGTTTCAGCTACAACGGCAGGAGTC
ATCCAACGTAACCATGAGCTGATCAACACTGCAATCATCGGGCGGGCGTGATGC
AAGCATGCCTGGCGAAGACACATGGTGTGCGGATGCTGCCGGCTGCTGCCTGCT
GCGCACGCCGTTGAGTTGGCAGCAGGCTCAGCCATGCACTGGATGGCAGCTGGG
CTGCCACTGCAATGTGGTGGATAGGATGCAAGTGGAGCGAATACCAAACCCTCT
GGCTGCTTGCTGGGTTGCATGGCATCGCACCATCAGCAGGAGCGCATGCGAAGG
GACTGGCCCCATGCACGCCATGCCAAACCGGAGCGCACCGAGTGTCCACACTGT
CACCAGGCCCGCAAGCTTTGCAGAACCATGCTCATGGACGCATGTAGCGCTGAC
GTCCCTTGACGGCGCTCCTCTCGGGTGTGGGAAACGCAATGCAGCACAGGCAGC
AGAGGCGGCGGCAGCAGAGCGGCGGCAGCAGCGGCGGGGGCCACCCTTCTTGCG
GGGTCGCGCCCCAGCCAGCGGTGATGCGCTGATCCCAAACGAGTTCACATTCATT
TGCATGCCTGGAGAAGCGAGGCTGGGGCCTTTGGGCTGGTGCAGCCCGCAATGG
AATGCGGGACCGCCAGGCTAGCAGCAAAGGCGCCTCCCCTACTCCGCATCGATG
TTCCATAGTGCATTGGACTGCATTTGGGTGGGGCGGCCGGCTGTTTCTTTCGTGTT
GCAAAACGCGCCAGCTCAGCAACCTGTCCCGTGGGTCCCCCGTGCCGATGAAAT
CGTGTGCACGCCGATCAGCTGATTGCCCGGCTCGCGAAGTAGGCGCCCTCCTTTC
TGCTCGCCCTCTCTCCGTCCCGCCTCTAGAATATCAATGATCGAGCAGGACGGCC
TCCACGCCGGCTCCCCCGCCGCCTGGGTGGAGCGCCTGTTCGGCTACGACTGGGC
CCAGCAGACCATCGGCTGCTCCGACGCCGCCGTGTTCCGCCTGTCCGCCCAGGGC
CGCCCCGTGCTGTTCGTGAAGACCGACCTGTCCGGCGCCCTGAACGAGCTGCAGG
ACGAGGCCGCCCGCCTGTCCTGGCTGGCCACCACCGGCGTGCCCTGCGCCGCCGT
GCTGGACGTGGTGACCGAGGCCGGCCGCGACTGGCTGCTGCTGGGCGAGGTGCC
CGGCCAGGACCTGCTGTCCTCCCACCTGGCCCCCGCCGAGAAGGTGTCCATCATG
GCCGACGCCATGCGCCGCCTGCACACCCTGGACCCCGCCACCTGCCCCTTCGACC
ACCAGGCCAAGCACCGCATCGAGCGCGCCCGCACCCGCATGGAGGCCGGCCTGG
TGGACCAGGACGACCTGGACGAGGAGCACCAGGGCCTGGCCCCCGCCGAGCTGT
TCGCCCGCCTGAAGGCCCGCATGCCCGACGGCGAGGACCTGGTGGTGACCCACG
GCGACGCCTGCCTGCCCAACATCATGGTGGAGAACGGCCGCTTCTCCGGCTTCAT
CGACTGCGGCCGCCTGGGCGTGGCCGACCGCTACCAGGACATCGCCCTGGCCAC
CCGCGACATCGCCGAGGAGCTGGGCGGCGAGTGGGCCGACCGCTTCCTGGTGCT
GTACGGCATCGCCGCCCCCGACTCCCAGCGCATCGCCTTCTACCGCCTGCTGGAC
GAGTTCTTCTGACAATTGGCAGCAGCAGCTCGGATAGTATCGACACACTCTGGAC
GCTGGTCGTGTGATGGACTGTTGCCGCCACACTTGCTGCCTTGACCTGTGAATAT
CCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGTGTGTACGCGCTTT
TGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCCAGCATCCCCTTCC
CTCGTTTCATATCGCTTGCATCCCAACCGCAACTTATCTACGCTGTCCTGCTATCC
CTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCTCGCACAGCCTTGGTTTGGGCTC
CGCCTGTATTCTCCTGGTACTGCAACCTGTAAACCAGCACTGCAATGCTGATGCA
CGGGAAGTAGTGGGATGGGAACACAAATGGAGGATCCCGCGTCTCGAACAGAGC
GCGCAGAGGAACGCTGAAGGTCTCGCCTCTGTCGCACCTCAGCGCGGCATACAC
CACAATAACCACCTGACGAATGCGCTTGGTTCTTCGTCCATTAGCGAAGCGTCCG

[1079]

GTTCACACACGTGCCACGTTGGCGAGGTGGCAGGTGACAATGATCGGTGGAGCT
GATGGTCGAAACGTTCACAGCCTAGGGATATCGAATTCCTTTCTTGCGCTATGAC
ACTTCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTTCCCGGCGCTGCATGCA
ACACCGATGATGCTTCGACCCCCCGAAGCTCCTTCGGGGCTGCATGGGCGCTCCG
ATGCCGCTCCAGGGCGAGCGCTGTTTAAATAGCCAGGCCCCCGATTGCAAAGAC
ATTATAGCGAGCTACCAAAGCCATATTCAAACACCTAGATCACTACCACTTCTAC
ACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGGGCGCCTCTTCCTCTTC
GTTTCAGTCACAACCCGCAAACACTAGTATGGCTATCAAGACGAACAGGCAGCC
TGTGGAGAAGCCTCCGTTCACGATCGGGACGCTGCGCAAGGCCATCCCCGCGCA
CTGTTTCGAGCGCTCGGCGCTTCGTGGGCGCGCCCCCGACTGGTCCATGCTGTTC
GCCGTGATCACCACCATCTTCAGCGCCGCCGAGAAGCAGTGGACCAACCTGGAG
TGGAAGCCCAAGCCCAAGCTGCCCCAGCTGCTGGACGACCACTTCGGCCTGCAC
GGCCTGGTGTTCCGCCGCACCTTCGCCATCCGCTCCTACGAGGTGGGCCCCGACC
GCAGCACCTCCATCCTGGCCGTGATGAACCACATGCAGGAGGCCACCCTGAACC
ACGCCAAGAGCGTGGGCATCCTGGGCGACGGCTTCGGCACCACCCTGGAGATGT
CCAAGCGCGACCTGATGTGGGTGGTGCGCCGCACCCACGTGGCCGTGGAGCGCT
ACCCCACCTGGGGCGACACCGTGGAGGTGGAGTGCTGGATCGGCGCCAGCGGCA
ACAACGGCATGCGCCGCGACTTCCTGGTGCGCGACTGCAAGACCGGCGAGATCC
TGACCCGCTGCACCTCCCTGAGCGTGCTGATGAACACCCGCACCCGCCGCCTGAG
CACCATCCCCGACGAGGTGCGCGGCGAGATCGGCCCCGCCTTCATCGACAACGT
GGCCGTGAAGGACGACGAGATCAAGAAGCTGCAGAAGCTGAACGACTCCACCGC
CGACTACATCCAGGGCGGCCTGACCCCCCGCTGGAACGACCTGGACGTGAACCA
GCACGTGAACAACCTGAAGTACGTGGCCTGGGTGTTCGAGACCGTGCCCGACAG
CATCTTCGAGTCCCACCACATCAGCTCCTTCACCCTGGAGTACCGCCGCGAGTGC
ACCCGCGACTCCGTGCTGCGCAGCCTGACCACCGTGAGCGGCGGCAGCTCCGAG
GCCGGCCTGGTGTGCGACCACCTGCTGCAGCTGGAGGGCGGCAGCGAGGTGCTG
CGCGCCCGCACCGAGTGGCGCCCCAAGCTGACCGACTCCTTCCGCGGCATCAGC
GTGATCCCCGCCGAGCCCCGCGTGATGGACTACAAGGACCACGACGGCGACTAC
AAGGACCACGACATCGACTACAAGGACGACGACGACAAGTGATGACTCGAGGC
AGCAGCAGCTCGGATAGTATCGACACACTCTGGACGCTGGTCGTGTGATGGACT
GTTGCCGCCACACTTGCTGCCTTGACCTGTGAATATCCCTGCCGCTTTTATCAAAC
AGCCTCAGTGTGTTTGATCTTGTGTGTACGCGCTTTTGCGAGTTGCTAGCTGCTTG
TGCTATTTGCGAATACCACCCCCAGCATCCCCTTCCCTCGTTTCATATCGCTTGCA
TCCCAACCGCAACTTATCTACGCTGTCCTGCTATCCCTCAGCGCTGCTCCTGCTCC
TGCTCACTGCCCCTCGCACAGCCTTGGTTTGGGCTCCGCCTGTATTCTCCTGGTAC
TGCAACCTGTAAACCAGCACTGCAATGCTGATGCACGGGAAGTAGTGGGATGGG
AACACAAATGGAAAGCTT

[1080]

[1081] 서열 번호 36

[1082] 프로토테카 모리포르미스 델타 12 지방산 불포화 효소 운반 펩티드를 지니는 쿠페아 팔루스트리스 C8 선호 티오에스테라제 (Genbank AAC49179) 아미노산 서열

MAIKTNRQPVEKPPFTIGTLRKAIPAHCFERSALRGRAPANGSAVTLKSGSLNTQEDT
LSSSPPPRAFFNQLPDWSMLLTAITTVFVAPEKRWTMFDRKSKRPNMLMDSFGLERV
VQDGLVFRQSFSIRSYEICADRTASIETVMNHVQETSLNQCKSIGLLDDGFGRSPEMC
KRDLIWVVTRMKIMVNRYPTWGDTIEVSTWLSQSGKIGMGRDWLISDCNTGEILVR
ATSVYAMMNQKTRRFSKLPHEVRQEFAPHFLDSPPAIEDNDGKLQKFDVKTGDSIRK
GLTPGWYDLDVNQHVSNVKYIGWILESMPTEVLETQELCSLTLEYRRECGRDSVLES
VTSMDPSKVGDRFQYRHLLRLEDGADIMKGRTEWRPKNAGTNGAISTGKT

[1083]

[1084] 서열 번호 37

[1085] 프로토테카 모리포르미스 델타 12 지방산 불포화 효소 운반 펩티드를 지니는 쿠페아 팔루스트리스 C8 선호 티오에스테라제 cDNA의 관련 코돈 최적화 발현 구조물

GGTACCCGCCTGCAACGCAAGGGCAGCCACAGCCGCTCCCACCCGCCGCTGAAC
CGACACGTGCTTGGGCGCCTGCCGCCTGCCTGCCGCATGCTTGTGCTGGTGAGGC
TGGGCAGTGCTGCCATGCTGATTGAGGCTTGGTTCATCGGGTGGAAGCTTATGTG
TGTGCTGGGCTTGCATGCCGGGCAATGCGCATGGTGGCAAGAGGGCGGCAGCAC
TTGCTGGAGCTGCCGCGGTGCCTCCAGGTGGTTCAATCGCGGCAGCCAGAGGGA
TTTCAGATGATCGCGCGTACAGGTTGAGCAGCAGTGTCAGCAAAGGTAGCAGTTT
GCCAGAATGATCGGTTCAGCTGTTAATCAATGCCAGCAAGAGAAGGGGTCAAGT
GCAAACACGGGCATGCCACAGCACGGGCACCGGGGAGTGGAATGGCACCACCA
AGTGTGTGCGAGCCAGCATCGCCGCCTGGCTGTTTCAGCTACAACGGCAGGAGTC
ATCCAACGTAACCATGAGCTGATCAACACTGCAATCATCGGGCGGGCGTGATGC
AAGCATGCCTGGCGAAGACACATGGTGTGCGGATGCTGCCGGCTGCTGCCTGCT
GCGCACGCCGTTGAGTTGGCAGCAGGCTCAGCCATGCACTGGATGGCAGCTGGG
CTGCCACTGCAATGTGGTGGATAGGATGCAAGTGGAGCGAATACCAAACCCTCT
GGCTGCTTGCTGGGTTGCATGGCATCGCACCATCAGCAGGAGCGCATGCGAAGG
GACTGGCCCCATGCACGCCATGCCAAACCGGAGCGCACCGAGTGTCCACACTGT
CACCAGGCCCGCAAGCTTTGCAGAACCATGCTCATGGACGCATGTAGCGCTGAC
GTCCCTTGACGGCGCTCCTCTCGGGTGTGGGAAACGCAATGCAGCACAGGCAGC
AGAGGCGGCGGCAGCAGAGCGGCGGCAGCAGCGGCGGGGGCCACCCTTCTTGCG
GGGTCGCGCCCCAGCCAGCGGTGATGCGCTGATCCCAAACGAGTTCACATTCATT
TGCATGCCTGGAGAAGCGAGGCTGGGGCCTTTGGGCTGGTGCAGCCCGCAATGG
AATGCGGGACCGCCAGGCTAGCAGCAAAGGCGCCTCCCCTACTCCGCATCGATG
TTCCATAGTGCATTGGACTGCATTTGGGTGGGGCGGCCGGCTGTTTCTTTCGTGTT
GCAAAACGCGCCAGCTCAGCAACCTGTCCCGTGGGTCCCCCGTGCCGATGAAAT
CGTGTGCACGCCGATCAGCTGATTGCCCGGCTCGCGAAGTAGGCGCCCTCCTTTC
TGCTCGCCCTCTCTCCGTCCCGCCTCTAGAATATCAATGATCGAGCAGGACGGCC
TCCACGCCGGCTCCCCCGCCGCCTGGGTGGAGCGCCTGTTCGGCTACGACTGGGC
CCAGCAGACCATCGGCTGCTCCGACGCCGCCGTGTTCCGCCTGTCCGCCCAGGGC
CGCCCCGTGCTGTTCGTGAAGACCGACCTGTCCGGCGCCCTGAACGAGCTGCAGG
ACGAGGCCGCCCGCCTGTCCTGGCTGGCCACCACCGGCGTGCCCTGCGCCGCCGT
GCTGGACGTGGTGACCGAGGCCGGCCGCGACTGGCTGCTGCTGGGCGAGGTGCC
CGGCCAGGACCTGCTGTCCTCCCACCTGGCCCCCGCCGAGAAGGTGTCCATCATG
GCCGACGCCATGCGCCGCCTGCACACCCTGGACCCCGCCACCTGCCCCTTCGACC
ACCAGGCCAAGCACCGCATCGAGCGCGCCCGCACCCGCATGGAGGCCGGCCTGG
TGGACCAGGACGACCTGGACGAGGAGCACCAGGGCCTGGCCCCCGCCGAGCTGT
TCGCCCGCCTGAAGGCCCGCATGCCCGACGGCGAGGACCTGGTGGTGACCCACG
GCGACGCCTGCCTGCCCAACATCATGGTGGAGAACGGCCGCTTCTCCGGCTTCAT
CGACTGCGGCCGCCTGGGCGTGGCCGACCGCTACCAGGACATCGCCCTGGCCAC
CCGCGACATCGCCGAGGAGCTGGGCGGCGAGTGGGCCGACCGCTTCCTGGTGCT
GTACGGCATCGCCGCCCCCGACTCCCAGCGCATCGCCTTCTACCGCCTGCTGGAC
GAGTTCTTCTGACAATTGGCAGCAGCAGCTCGGATAGTATCGACACACTCTGGAC
GCTGGTCGTGTGATGGACTGTTGCCGCCACACTTGCTGCCTTGACCTGTGAATAT
CCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGTGTGTACGCGCTTT
TGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCCAGCATCCCCTTCC
CTCGTTTCATATCGCTTGCATCCCAACCGCAACTTATCTACGCTGTCCTGCTATCC
CTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCTCGCACAGCCTTGGTTTGGGCTC
CGCCTGTATTCTCCTGGTACTGCAACCTGTAAACCAGCACTGCAATGCTGATGCA

[1086]

CGGGAAGTAGTGGGATGGGAACACAAATGGAGGATCCCGCGTCTCGAACAGAGC
GCGCAGAGGAACGCTGAAGGTCTCGCCTCTGTCGCACCTCAGCGCGGCATACAC
CACAATAACCACCTGACGAATGCGCTTGGTTCTTCGTCCATTAGCGAAGCGTCCG
GTTCACACACGTGCCACGTTGGCGAGGTGGCAGGTGACAATGATCGGTGGAGCT
GATGGTCGAAACGTTCACAGCCTAGGGATATCGAATTCCTTTCTTGCGCTATGAC
ACTTCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTTCCCGGCGCTGCATGCA
ACACCGATGATGCTTCGACCCCCCGAAGCTCCTTCGGGGCTGCATGGGCGCTCCG
ATGCCGCTCCAGGGCGAGCGCTGTTTAAATAGCCAGGCCCCCGATTGCAAAGAC
ATTATAGCGAGCTACCAAAGCCATATTCAAACACCTAGATCACTACCACTTCTAC
ACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGGGCGCCTCTTCCTCTTC
GTTTCAGTCACAACCCGCAAACACTAGTATGGCTATCAAGACGAACAGGCAGCC
TGTGGAGAAGCCTCCGTTCACGATCGGGACGCTGCGCAAGGCCATCCCCGCGCA
CTGTTTCGAGCGCTCGGCGCTTCGTGGGCGCGCCCCCGCGAACGGCAGCGCGGTG
ACCCTGAAGTCGGGCTCCCTGAACACCCAGGAGGACACGCTGAGCTCGTCCCCC
CCCCCCCGCGCGTTCTTCAACCAGCTGCCCGACTGGAGCATGCTGCTGACCGCGA
TCACCACGGTCTTCGTGGCGCCCGAGAAGCGCTGGACCATGTTCGACCGCAAGTC
GAAGCGCCCCAACATGCTGATGGACTCCTTCGGCCTGGAGCGCGTGGTCCAGGA
CGGCCTGGTGTTCCGCCAGAGCTTCTCGATCCGCTCCTACGAGATCTGCGCGGAC
CGCACCGCGAGCATCGAGACGGTGATGAACCACGTCCAGGAGACCTCGCTGAAC
CAGTGCAAGTCCATCGGCCTGCTGGACGACGGCTTCGGCCGCAGCCCCGAGATG
TGCAAGCGCGACCTGATCTGGGTGGTCACCCGCATGAAGATCATGGTGAACCGC
TACCCCACGTGGGGCGACACCATCGAGGTCTCGACGTGGCTGTCCCAGAGCGGC
AAGATCGGCATGGGCCGCGACTGGCTGATCTCGGACTGCAACACCGGCGAGATC
CTGGTGCGCGCGACGTCCGTCTACGCGATGATGAACCAGAAGACCCGCCGCTTC
AGCAAGCTGCCCCACGAGGTGCGCCAGGAGTTCGCGCCCCACTTCCTGGACTCGC
CCCCCGCGATCGAGGACAACGACGGCAAGCTGCAGAAGTTCGACGTCAAGACGG
GCGACTCCATCCGCAAGGGCCTGACCCCCGGCTGGTACGACCTGGACGTGAACC
AGCACGTGAGCAACGTCAAGTACATCGGCTGGATCCTGGAGTCGATGCCCACCG
AGGTCCTGGAGACGCAGGAGCTGTGCTCCCTGACCCTGGAGTACCGCCGCGAGT
GCGGCCGCGACTCGGTGCTGGAGAGCGTCACCAGCATGGACCCCTCGAAGGTGG
GCGACCGCTTCCAGTACCGCCACCTGCTGCGCCTGGAGGACGGCGCGGACATCA
TGAAGGGCCGCACCGAGTGGCGCCCCAAGAACGCGGGCACGAACGGCGCGATCT
CCACCGGCAAGACGTGACTCGAGGCAGCAGCAGCTCGGATAGTATCGACACACT
CTGGACGCTGGTCGTGTGATGGACTGTTGCCGCCACACTTGCTGCCTTGACCTGT
GAATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGTGTGTACG
CGCTTTTGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCCAGCATCC
CCTTCCCTCGTTTCATATCGCTTGCATCCCAACCGCAACTTATCTACGCTGTCCTG
CTATCCCTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCTCGCACAGCCTTGGTTT
GGGCTCCGCCTGTATTCTCCTGGTACTGCAACCTGTAAACCAGCACTGCAATGCT
GATGCACGGGAAGTAGTGGGATGGGAACACAAATGGAAAGCTTGAGCTC

[1087]

[1088] 서열 번호 38

[1089] 프로토테카 모리포르미스 델타 12 지방산 불포화 효소 운반 펩티드를 지니는 쿠페아 란세올라타 C10 선호 티오에스테라제 (Genbank CAB60830) 아미노산 서열

MAIKTNRQPVEKPPFTIGTLRKAIPAHCFERSALRGRAPANGSAVNLKSGSLNTQEDT
SSSPPPRAFLNQLPDWSMLLTAITTVFVAAEKQWTMLDRKSKRPDMLVDSVGLKSIV
RDGLVSRQSFLIRSYEIGADRTASIETLMNHLQETSINHCKSLGLLNDGFGRTPGMCK
NDLIWVLTKMQIMVNRYPTWGDTVEINTWFSQSGKIGMASDWLISDCNTGEILIRAT
SVWAMMNQKTRRFSRLPYEVRQELTPHFVDSPHVIEDNDQKLHKFDVKTGDSIRKG

[1090]

LTPRWNDLDVNQHVSNVKYIGWILESMPIEVLETQELCSLTVEYRRECGMDSVLESV
TAVDPSENGGRSQYKHLLRLEDGTDIVKSRTEWRPKNAGTNGAISTSTAKTSNGNSA
SDDDDKLG

[1091]

[1092] 서열 번호 39

[1093] 프로토테카 모리포르미스 델타 12 지방산 불포화 효소 운반 펩티드를 지니는 쿠페아 란세올라타 C10 선호 티오에스테라제의 관련 코돈 최적화 코딩 영역

ACTAGTATGGCTATCAAGACGAACAGGCAGCCTGTGGAGAAGCCTCCGTTCACG
ATCGGGACGCTGCGCAAGGCCATCCCCGCGCACTGTTTCGAGCGCTCGGCGCTTC
GTGGGCGCGCCCCCGCGAACGGCAGCGCGGTGAACCTGAAGTCGGGCTCCCTGA
ACACCCAGGAGGACACGAGCTCGTCCCCCCCCCCCCGCGCGTTCCTGAACCAGCT
GCCCGACTGGAGCATGCTGCTGACCGCGATCACCACCGTCTTCGTGGCGGCGGA
GAAGCAGTGGACGATGCTGGACCGCAAGTCGAAGCGCCCCGACATGCTGGTGGA
CTCCGTCGGCCTGAAGAGCATCGTGCGCGACGGCCTGGTCTCGCGCCAGTCCTTC
CTGATCCGCAGCTACGAGATCGGCGCGGACCGCACCGCGTCGATCGAGACCCTG
ATGAACCACCTGCAGGAGACGTCCATCAACCACTGCAAGAGCCTGGGCCTGCTG
AACGACGGCTTCGGCCGCACCCCCGGCATGTGCAAGAACGACCTGATCTGGGTG
CTGACCAAGATGCAGATCATGGTCAACCGCTACCCCACGTGGGGCGACACCGTC
GAGATCAACACGTGGTTCTCGCAGTCCGGCAAGATCGGCATGGCGAGCGACTGG
CTGATCTCGGACTGCAACACCGGCGAGATCCTGATCCGCGCGACCTCCGTGTGGG
CGATGATGAACCAGAAGACGCGCCGCTTCAGCCGCCTGCCCTACGAGGTCCGCC
AGGAGCTGACCCCCCACTTCGTGGACTCGCCCCACGTCATCGAGGACAACGACC
AGAAGCTGCACAAGTTCGACGTGAAGACCGGCGACTCCATCCGCAAGGGCCTGA
CGCCCCGCTGGAACGACCTGGACGTCAACCAGCACGTGTCGAACGTGAAGTACA
TCGGCTGGATCCTGGAGTCCATGCCCATCGAGGTCCTGGAGACCCAGGAGCTGTG
CTCGCTGACCGTGGAGTACCGCCGCGAGTGCGGCATGGACTCCGTGCTGGAGTC
GGTCACGGCGGTGGACCCCAGCGAGAACGGCGGCCGCAGCCAGTACAAGCACCT
GCTGCGCCTGGAGGACGGCACCGACATCGTCAAGTCGCGCACCGAGTGGCGCCC
CAAGAACGCGGGCACGAACGGCGCGATCTCCACCAGCACCGCGAAGACGTCGAA
CGGCAACTCCGCGAGCGATGACGATGACAAGCTGGGATGACTCGAG

[1094]

[1095] 서열 번호 40

[1096] 클로렐라 프로토테코이데스 스테아로일 ACP 불포화 효소 엽록체 운반 펩티드를 지니는 아이리스 게르마니카 C14 선호 티오에스테라제 (Genbank AAG43858.1) 아미노산 서열

MATASTFSAFNARCGDLRRSAGSGPRRPARPLPVRGRAAQAATRVNGSKVGLKTDT
NKLEDAPFIPSSAPRTFYNQLPDWSVLLAAITTIFLAAEKQWTLIDWKRGGPDMLSDA
FGLPKIIENGLLYRQKFSIRSYEIGADQTASIETLMNHLQETALNHVKCAGLLGNGFG
STPEMSKMNLIWVVTKMQVLVEHYPSWGDVIEVDTWAAASGKNGMRRDWHVRD
WQTGQTIMRASSNWVMMNQNTRRLSKFPEEVRAEIEPYFMERAPVIDDDNRKLPKL
DDDTADHVRNGLTPRWSDLDVNQHVKNVKYIGWILESAPISILESHELASMTLEYRR
ECGRDSVLQSLTSVSNNCTDGSEELPIECQHLLRNEGGSEIVKGRTEWRPKKCGPFGA
GRP

[1097]

[1098] 서열 번호 41

[1099] 클로렐라 프로토테코이데스 스테아로일 ACP 불포화 효소 운반 펩티드를 지니는 아이리스 게르마니카 C14 선호 티오에스테라제의 관련 코돈 최적화 코딩 영역

ACTAGTATGGCCACCGCATCCACTTTCTCGGCGTTCAATGCCCGCTGCGGCGACC
TGCGTCGCTCGGCGGGCTCCGGGCCCCGGCGCCCAGCGAGGCCCCTCCCCGTGCG
CGGGCGCGCCGCCCAGGCGGCCACCCGCGTGAACGGCAGCAAGGTGGGCCTGAA
GACCGACACCAACAAGCTGGAGGACGCGCCCTTCATCCCCTCGTCCGCCCCCCGC
ACCTTCTACAACCAGCTGCCCGACTGGAGCGTCCTGCTGGCGGCCATCACCACCA
TCTTCCTGGCGGCCGAGAAGCAGTGGACCCTGATCGACTGGAAGCGCGGCGGCC
CCGACATGCTGTCGGACGCGTTCGGCCTGCCCAAGATCATCGAGAACGGCCTGCT
GTACCGCCAGAAGTTCTCCATCCGCAGCTACGAGATCGGCGCCGACCAGACCGC
CTCGATCGAGACCCTGATGAACCACCTGCAGGAGACCGCGCTGAACCACGTCAA
GTGCGCCGGCCTGCTGGGCAACGGCTTCGGCTCCACCCCCGAGATGAGCAAGAT
GAACCTGATCTGGGTGGTCACCAAGATGCAGGTGCTGGTCGAGCACTACCCCTCG
TGGGGCGACGTGATCGAGGTGGACACCTGGGCGGCCGCGTCCGGCAAGAACGGC
ATGCGCCGCGACTGGCACGTCCGCGACTGGCAGACCGGCCAGACCATCATGCGC
GCCAGCTCGAACTGGGTGATGATGAACCAGAACACCCGCCGCCTGTCCAAGTTC
CCCGAGGAGGTCCGCGCCGAGATCGAGCCCTACTTCATGGAGCGCGCCCCCGTG
ATCGACGACGACAACCGCAAGCTGCCCAAGCTGGACGACGACACCGCGGACCAC
GTGCGCAACGGCCTGACCCCCCGCTGGAGCGACCTGGACGTGAACCAGCACGTC
AAGAACGTGAAGTACATCGGCTGGATCCTGGAGTCGGCCCCCATCTCCATCCTGG
AGAGCCACGAGCTGGCCTCGATGACCCTGGAGTACCGCCGCGAGTGCGGCCGCG
ACTCCGTCCTGCAGAGCCTGACCTCGGTGTCCAACAACTGCACCGACGGCAGCG
AGGAGCTGCCCATCGAGTGCCAGCACCTGCTGCGCAACGAGGGCGGCTCGGAGA
TCGTCAAGGGCCGCACCGAGTGGCGCCCCAAGAAGTGCGGCCCCTTCGGCGCCG
GCCGCCCCTGACTCGAG

[1100]

[1101] 서열 번호 42

[1102] 프로토테카 모리포르미스 델타 12 지방산 불포화 효소 엽록체 운반 펩티드를 지니는 미리스티카 프래그란스 지

방산 티오에스테라제 (Genbank AAB717291.1) 아미노산 서열

MAIKTNRQPVEKPPFTIGTLRKAIPAHCFERSALRGRAANAHTVPKINGNKAGLLTPM
ESTKDEDIVAAPTVAPKRTFINQLPDWSMLLAAITTIFLAAEKQWTNLDWKPRRPDM
LVDFDPFSLGRFVQDGLIFRQNFSIRSYEIGADRTASIETLMNHLQETALNHVRCIGLL
DDGFGSTPEMTRRDLIWVVTRMQVLVDRYPSWGDVIEVDSWVTPSGKNGMKREWF
LRDCKTGEILTRATSVWVMMNKRTRRLSKIPEEVRVEIEPYFVEHGVLDEDSRKLPK
LNDNTANYIRRGLAPRWSDLDVNQHVNNVKYIGWILESVPSSLLESHELYGMTLEY
RKECGKDGLLQSLTAVASDYGGGSLEAGVECDHLLRLEDGSEIMRGKTEWRPKRAA
NTTYFGSVDDIPPANNA

[1103]

[1104] 서열 번호 43

[1105] 프로토테카 모리포르미스 델타 12 지방산 불포화 효소 엽록체 운반 펩티드를 지니는 미리스티카 프래그란스 지방산 티오에스테라제의 관련 코돈 최적화 코딩 영역

ACTAGTATGGCTATCAAGACGAACAGGCAGCCTGTGGAGAAGCCTCCGTTCACG
ATCGGGACGCTGCGCAAGGCCATCCCCGCGCACTGTTTCGAGCGCTCGGCGCTTC
GTGGGCGCGCCGCCAACGCCCACACCGTGCCCAAGATCAACGGCAACAAGGCCG
GCCTGCTGACCCCCATGGAGAGCACCAAGGACGAGGACATCGTCGCGGCCCCCA
CCGTGGCGCCCAAGCGCACCTTCATCAACCAGCTGCCCGACTGGTCGATGCTGCT
GGCCGCGATCACCACCATCTTCCTGGCGGCCGAGAAGCAGTGGACCAACCTGGA
CTGGAAGCCCCGCCGCCCCGACATGCTGGTCGACTTCGACCCCTTCTCCCTGGGC
CGCTTCGTGCAGGACGGCCTGATCTTCCGCCAGAACTTCAGCATCCGCTCGTACG
AGATCGGCGCGGACCGCACCGCCTCCATCGAGACCCTGATGAACCACCTGCAGG

[1106]

AGACCGCGCTGAACCACGTCCGCTGCATCGGCCTGCTGGACGACGGCTTCGGCA
GCACCCCCGAGATGACCCGCCGCGACCTGATCTGGGTGGTCACCCGCATGCAGG
TCCTGGTGGACCGCTACCCCTCGTGGGGCGACGTGATCGAGGTCGACTCCTGGGT
GACCCCCAGCGGCAAGAACGGCATGAAGCGCGAGTGGTTCCTGCGCGACTGCAA
GACCGGCGAGATCCTGACCCGCGCCACCTCGGTCTGGGTGATGATGAACAAGCG
CACCCGCCGCCTGTCCAAGATCCCCGAGGAGGTCCGCGTGGAGATCGAGCCCTA
CTTCGTCGAGCACGGCGTGCTGGACGAGGACTCGCGCAAGCTGCCCAAGCTGAA
CGACAACACCGCCAACTACATCCGCCGCGGCCTGGCGCCCCGCTGGTCCGACCTG
GACGTCAACCAGCACGTGAACAACGTCAAGTACATCGGCTGGATCCTGGAGAGC
GTGCCCAGCAGCCTGCTGGAGTCGCACGAGCTGTACGGCATGACCCTGGAGTAC
CGCAAGGAGTGCGGCAAGGACGGCCTGCTGCAGTCCCTGACCGCCGTCGCCAGC
GACTACGGCGGCGGCTCGCTGGAGGCCGGCGTGGAGTGCGACCACCTGCTGCGC
CTGGAGGACGGCTCCGAGATCATGCGCGGCAAGACCGAGTGGCGCCCCAAGCGC
GCCGCGAACACCACCTACTTCGGCAGCGTCGACGACATCCCCCCCGCCAACAAC
GCGTGACTCGAG

[1107]

[1108] 서열 번호 44

[1109] 클로렐라 프로토테코이데스 스테아로일 ACP 불포화 효소 운반 펩티드를 지니는 쿠페아 팔루스트리스 C14 선호 티오에스테라제 (Genbank AAC49180) 아미노산 서열

MATASTFSAFNARCGDLRRSAGSGPRRPARPLPVRGRASMLLSAVTTVFGVAEKQW
PMLDRKSKRPDMLVEPLGVDRIVYDGVSFRQSFSIRSYEIGADRTASIETLMNMFQET
SLNHCKIIGLLNDGFGRTPEMCKRDLIWVVTKMQIEVNRYPTWGDTIEVNTWVSASG
KHGMGRDWLISDCHTGEILIRATSVWAMMNQKTRRLSKIPYEVRQEIEPQFVDSAPV
IVDDRKFHKLDLKTGDSICNGLTPRWTDLDVNQHVNNVKYIGWILQSVPTEVFETQE
LCGLTLEYRRECGRDSVLESVTAMDPSKEGDRSLYQHLLRLEDGADIVKGRTEWRP
KNAGAKGAILTGKTSNGNSIS

[1110]

[1111] 서열 번호 45

[1112] 클로렐라 프로토테코이데스 스테아로일 ACP 불포화 효소 운반 펩티드를 지니는 쿠페아 팔루스트리스 C14 선호 티오에스테라제의 관련 코돈 최적화 코딩 영역

ACTAGTATGGCCACCGCATCCACTTTCTCGGCGTTCAATGCCCGCTGCGGCGACC
TGCGTCGCTCGGCGGGCTCCGGGCCCCGGCGCCCAGCGAGGCCCCTCCCCGTGCG
CGGGCGCGCCAGCATGCTGCTGTCGGCGGTGACCACGGTCTTCGGCGTGGCCGA
GAAGCAGTGGCCCATGCTGGACCGCAAGTCCAAGCGCCCCGACATGCTGGTCGA
GCCCCTGGGCGTGGACCGCATCGTCTACGACGGCGTGAGCTTCCGCCAGTCGTTC
TCCATCCGCAGCTACGAGATCGGCGCCGACCGCACCGCCTCGATCGAGACGCTG
ATGAACATGTTCCAGGAGACCTCCCTGAACCACTGCAAGATCATCGGCCTGCTGA
ACGACGGCTTCGGCCGCACGCCCGAGATGTGCAAGCGCGACCTGATCTGGGTCG
TGACCAAGATGCAGATCGAGGTGAACCGCTACCCCACGTGGGGCGACACCATCG
AGGTCAACACGTGGGTGAGCGCCTCGGGCAAGCACGGCATGGGCCGCGACTGGC
TGATCTCCGACTGCCACACCGGCGAGATCCTGATCCGCGCGACGAGCGTCTGGGC
GATGATGAACCAGAAGACCCGCCGCCTGTCGAAGATCCCCTACGAGGTGCGCCA
GGAGATCGAGCCCCAGTTCGTCGACTCCGCCCCCGTGATCGTGGACGACCGCAA
GTTCCACAAGCTGGACCTGAAGACGGGCGACAGCATCTGCAACGGCCTGACCCC
CCGCTGGACGGACCTGGACGTGAACCAGCACGTCAACAACGTGAAGTACATCGG
CTGGATCCTGCAGTCGGTCCCCACCGAGGTGTTCGAGACGCAGGAGCTGTGCGG
CCTGACCCTGGAGTACCGCCGCGAGTGCGGCCGCGACTCCGTGCTGGAGAGCGT
CACGGCCATGGACCCCTCGAAGGAGGGCGACCGCTCCCTGTACCAGCACCTGCT

[1113]

GCGCCTGGAGGACGGCGCGGACATCGTGAAGGGCCGCACCGAGTGGCGCCCCAA
GAACGCCGGCGCCAAGGGCGCCATCCTGACGGGCAAGACCAGCAACGGCAACTC
GATCTCCTGACTCGAG

[1114]

[1115] 서열 번호 46

[1116] 클로렐라 프로토테코이데스 스테아로일 ACP 불포화 효소 운반 펩티드를 지니는 얼머스 아메리카나 광범위 특이성 티오에스테라제 (Genbank AAB71731) 아미노산 서열

MATASTFSAFNARCGDLRRSAGSGPRRPARPLPVRGRAQLPDWSMLLAAITTLFLAA
EKQWMMLDWKPKRPDMLVDPFGLGRFVQDGLVFRNNFSIRSYEIGADRTASIETLM
NHLQETALNHVKSVGLLEDGLGSTREMSLRNLIWVVTKMQVAVDRYPTWGDEVQV
SSWATAIGKNGMRREWIVTDFRTGETLLRATSVWVMMNKLTRRISKIPEEVWHEIGP
SFIDAPPLPTVEDDGRKLTRFDESSADFIRKGLTPRWSDLDINQHVNNVKYIGWLLES
APPEIHESHEIASLTLEYRRECGRDSVLNSATKVSDSSQLGKSAVECNHLVRLQNGGE
IVKGRTVWRPKRPLYNDGAVVDVPAKTS

[1117]

[1118] 서열 번호 47

[1119] 클로렐라 프로토테코이데스 스테아로일 ACP 불포화 효소 운반 펩티드를 지니는 얼머스 아메리카나 광범위 특이성 티오에스테라제의 관련 코돈 최적화 코딩 영역

ACTAGTATGGCCACCGCATCCACTTTCTCGGCGTTCAATGCCCGCTGCGGCGACC
TGCGTCGCTCGGCGGGCTCCGGGCCCCGGCGCCCAGCGAGGCCCCTCCCCGTGCG
CGGGCGCGCCCAGCTGCCCGACTGGAGCATGCTGCTGGCCGCGATCACCACCCT
GTTCCTGGCGGCCGAGAAGCAGTGGATGATGCTGGACTGGAAGCCCAAGCGCCC
CGACATGCTGGTGGACCCCTTCGGCCTGGGCCGCTTCGTGCAGGACGGCCTGGTG
TTCCGCAACAACTTCAGCATCCGCAGCTACGAGATCGGCGCGGACCGCACCGCC
AGCATCGAGACCCTGATGAACCACCTGCAGGAGACCGCCCTGAACCACGTGAAG
AGCGTGGGCCTGCTGGAGGACGGCCTGGGCAGCACCCGCGAGATGAGCCTGCGC
AACCTGATCTGGGTGGTGACCAAGATGCAGGTGGCGGTGGACCGCTACCCCACC
TGGGGCGACGAGGTGCAGGTGAGCAGCTGGGCGACCGCCATCGGCAAGAACGG
CATGCGCCGCGAGTGGATCGTGACCGACTTCCGCACCGGCGAGACCCTGCTGCG
CGCCACCAGCGTGTGGGTGATGATGAACAAGCTGACCCGCCGCATCAGCAAGAT
CCCCGAGGAGGTGTGGCACGAGATCGGCCCCAGCTTCATCGACGCGCCCCCCCT
GCCCACCGTGGAGGACGACGGCCGCAAGCTGACCCGCTTCGACGAGAGCAGCGC
CGACTTCATCCGCAAGGGCCTGACCCCCCGCTGGAGCGACCTGGACATCAACCA
GCACGTGAACAACGTGAAGTACATCGGCTGGCTGCTGGAGAGCGCGCCCCCCGA
GATCCACGAGAGCCACGAGATCGCCAGCCTGACCCTGGAGTACCGCCGCGAGTG
CGGCCGCGACAGCGTGCTGAACAGCGCCACCAAGGTGAGCGACAGCAGCCAGCT
GGGCAAGAGCGCCGTGGAGTGCAACCACCTGGTGCGCCTGCAGAACGGCGGCGA
GATCGTGAAGGGCCGCACCGTGTGGCGCCCCAAGCGCCCCCTGTACAACGACGG
CGCCGTGGTGGACGTGCCCGCCAAGACCAGCTGACTCGAG

[1120]

[1121] 서열 번호 48

[1122] 코돈 최적화 프로토테카 모리포르미스 (UTEX 1435) 델타 12 지방산 불포화 효소 운반 펩티드 cDNA 서열

ACTAGTATGGCTATCAAGACGAACAGGCAGCCTGTGGAGAAGCCTCCGTTCACG
ATCGGGACGCTGCGCAAGGCCATCCCCCGCGCACTGTTTCGAGCGCTCGGCGCTTC
GTGGGCGCGCC

[1123]

[1124] 서열 번호 49

[1125] 코돈 최적화 클로렐라 프로토테코이데스 (UTEX 250) 스테아로일 ACP 불포화 효소 운반 펩티드 cDNA 서열

ACTAGTATGGCCACCGCATCCACTTTCTCGGCGTTCAATGCCCGCTGCGGCGACC
TGCGTCGCTCGGCGGGCTCCGGGCCCCGGCGCCCAGCGAGGCCCCTCCCCGTGCG
CGGGCGCGCC

[1126]

[1127] 서열 번호 50

[1128] 얼머스 아메리카나 광범위 특이성 티오에스테라제의 코돈 최적화 코딩 영역의 적절한 상동 재조합 발현 구조물

GCTCTTCGGCCGCCGCCACTCCTGCTCGAGCGCGCCCGACTCGCGCTCCGCCTGC
GCCCGCGCGTGCGCCGCCAGCGCCTTGGCCTTTTCGCCGCGCTCGTGCGCGTCGC
TGATGTCCATCACCAGGTCCATGAGGTCTGCCTTGCGCCGGCTGAGCCACTGCTT
CGTCCGGGCGGCCAAGAGGAGCATGAGGGAGGACTCCTGGTCCAGGGTCCTGAC
GTGGTCGCGGCTCTGGGAGCGGGCCAGCATCATCTGGCTCTGCCGCACCGAGGC
CGCCTCCAACTGGTCCTCCAGCAGCCGCAGTCGCCGCCGACCCTGGCAGAGGAA
GACAGGTGAGGGGGGTATGAATTGTACAGAACAACCACGAGCCTTGTCTAGGCA
GAATCCCTACCAGTCATGGCTTTACCTGGATGACGGCCTGCGAACAGCTGTCCAG
CGACCCTCGCTGCCGCCGCTTCTCCCGCACGCTTCTTTCCAGCACCGTGATGGCG
CGAGCCAGCGCCGCACGCTGGCGCTGCGCTTCGCCGATCTGAGGACAGTCGGGG
AACTCTGATCAGTCTAAACCCCCTTGCGCGTTAGTGTTGCCATCCTTTGCAGACC
GGTGAGAGCCGACTTGTTGTGCGCCACCCCCCACACCACCTCCTCCCAGACCAAT
TCTGTCACCTTTTTGGCGAAGGCATCGGCCTCGGCCTGCAGAGAGGACAGCAGTG
CCCAGCCGCTGGGGGTTGGCGGATGCACGCTCAGGTACCCTTTCTTGCGCTATGA
CACTTCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTTCCCGGCGCTGCATGC
AACACCGATGATGCTTCGACCCCCCGAAGCTCCTTCGGGGCTGCATGGGCGCTCC
GATGCCGCTCCAGGGCGAGCGCTGTTTAAATAGCCAGGCCCCCGATTGCAAAGA
CATTATAGCGAGCTACCAAAGCCATATTCAAACACCTAGATCACTACCACTTCTA
CACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGGGCGCCTCTTCCTCTT
CGTTTCAGTCACAACCCGCAAACGGCGCGCCATGCTGCTGCAGGCCTTCCTGTTC
CTGCTGGCCGGCTTCGCCGCCAAGATCAGCGCCTCCATGACGAACGAGACGTCC
GACCGCCCCCTGGTGCACTTCACCCCCAACAAGGGCTGGATGAACGACCCCAAC
GGCCTGTGGTACGACGAGAAGGACGCCAAGTGGCACCTGTACTTCCAGTACAAC
CCGAACGACACCGTCTGGGGGACGCCCTTGTTCTGGGGCCACGCCACGTCCGAC
GACCTGACCAACTGGGAGGACCAGCCCATCGCCATCGCCCCGAAGCGCAACGAC
TCCGGCGCCTTCTCCGGCTCCATGGTGGTGGACTACAACAACACCTCCGGCTTCT
TCAACGACACCATCGACCCGCGCCAGCGCTGCGTGGCCATCTGGACCTACAACA
CCCCGGAGTCCGAGGAGCAGTACATCTCCTACAGCCTGGACGGCGGCTACACCTT
CACCGAGTACCAGAAGAACCCCGTGCTGGCCGCCAACTCCACCCAGTTCCGCGA
CCCGAAGGTCTTCTGGTACGAGCCCTCCCAGAAGTGGATCATGACCGCGGCCAA
GTCCCAGGACTACAAGATCGAGATCTACTCCTCCGACGACCTGAAGTCCTGGAA
GCTGGAGTCCGCGTTCGCCAACGAGGGCTTCCTCGGCTACCAGTACGAGTGCCCC
GGCCTGATCGAGGTCCCCACCGAGCAGGACCCCAGCAAGTCCTACTGGGTGATG
TTCATCTCCATCAACCCCGGCGCCCCGGCCGGCGGCTCCTTCAACCAGTACTTCG
TCGGCAGCTTCAACGGCACCCACTTCGAGGCCTTCGACAACCAGTCCCGCGTGGT
GGACTTCGGCAAGGACTACTACGCCCTGCAGACCTTCTTCAACACCGACCCGACC
TACGGGAGCGCCCTGGGCATCGCGTGGGCCTCCAACTGGGAGTACTCCGCCTTCG
TGCCCACCAACCCCTGGCGCTCCTCCATGTCCCTCGTGCGCAAGTTCTCCCTCAAC

[1129]

ACCGAGTACCAGGCCAACCCGGAGACGGAGCTGATCAACCTGAAGGCCGAGCCG
ATCCTGAACATCAGCAACGCCGGCCCCTGGAGCCGGTTCGCCACCAACACCACG
TTGACGAAGGCCAACAGCTACAACGTCGACCTGTCCAACAGCACCGGCACCCTG
GAGTTCGAGCTGGTGTACGCCGTCAACACCACCCAGACGATCTCCAAGTCCGTGT
TCGCGGACCTCTCCCTCTGGTTCAAGGGCCTGGAGGACCCCGAGGAGTACCTCCG
CATGGGCTTCGAGGTGTCCGCGTCCTCCTTCTTCCTGGACCGCGGGAACAGCAAG
GTGAAGTTCGTGAAGGAGAACCCCTACTTCACCAACCGCATGAGCGTGAACAAC
CAGCCCTTCAAGAGCGAGAACGACCTGTCCTACTACAAGGTGTACGGCTTGCTGG
ACCAGAACATCCTGGAGCTGTACTTCAACGACGGCGACGTCGTGTCCACCAACA
CCTACTTCATGACCACCGGGAACGCCCTGGGCTCCGTGAACATGACGACGGGGG
TGGACAACCTGTTCTACATCGACAAGTTCCAGGTGCGCGAGGTCAAGTGACAATT
GGCAGCAGCAGCTCGGATAGTATCGACACACTCTGGACGCTGGTCGTGTGATGG
ACTGTTGCCGCCACACTTGCTGCCTTGACCTGTGAATATCCCTGCCGCTTTTATCA
AACAGCCTCAGTGTGTTTGATCTTGTGTGTACGCGCTTTTGCGAGTTGCTAGCTGC
TTGTGCTATTTGCGAATACCACCCCCAGCATCCCCTTCCCTCGTTTCATATCGCTT
GCATCCCAACCGCAACTTATCTACGCTGTCCTGCTATCCCTCAGCGCTGCTCCTGC
TCCTGCTCACTGCCCCTCGCACAGCCTTGGTTTGGGCTCCGCCTGTATTCTCCTGG
TACTGCAACCTGTAAACCAGCACTGCAATGCTGATGCACGGGAAGTAGTGGGAT
GGGAACACAAATGGAGGATCCCGCGTCTCGAACAGAGCGCGCAGAGGAACGCT
GAAGGTCTCGCCTCTGTCGCACCTCAGCGCGGCATACACCACAATAACCACCTGA
CGAATGCGCTTGGTTCTTCGTCCATTAGCGAAGCGTCCGGTTCACACACGTGCCA
CGTTGGCGAGGTGGCAGGTGACAATGATCGGTGGAGCTGATGGTCGAAACGTTC
ACAGCCTAGGGATATCGAATTCCTTTCTTGCGCTATGACACTTCCAGCAAAAGGT
AGGGCGGGCTGCGAGACGGCTTCCCGGCGCTGCATGCAACACCGATGATGCTTC
GACCCCCCGAAGCTCCTTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCAGGGCG
AGCGCTGTTTAAATAGCCAGGCCCCCGATTGCAAAGACATTATAGCGAGCTACC
AAAGCCATATTCAAACACCTAGATCACTACCACTTCTACACAGGCCACTCGAGCT
TGTGATCGCACTCCGCTAAGGGGGCGCCTCTTCCTCTTCGTTTCAGTCACAACCC
GCAAACACTAGTATGGCCACCGCATCCACTTTCTCGGCGTTCAATGCCCGCTGCG
GCGACCTGCGTCGCTCGGCGGGCTCCGGGCCCCGGCGCCCAGCGAGGCCCCTCC
CCGTGCGCGGGCGCGCCCAGCTGCCCGACTGGAGCATGCTGCTGGCCGCGATCA
CCACCCTGTTCCTGGCGGCCGAGAAGCAGTGGATGATGCTGGACTGGAAGCCCA
AGCGCCCCGACATGCTGGTGGACCCCTTCGGCCTGGGCCGCTTCGTGCAGGACGG
CCTGGTGTTCCGCAACAACTTCAGCATCCGCAGCTACGAGATCGGCGCGGACCGC
ACCGCCAGCATCGAGACCCTGATGAACCACCTGCAGGAGACCGCCCTGAACCAC
GTGAAGAGCGTGGGCCTGCTGGAGGACGGCCTGGGCAGCACCCGCGAGATGAGC
CTGCGCAACCTGATCTGGGTGGTGACCAAGATGCAGGTGGCGGTGGACCGCTAC
CCCACCTGGGGCGACGAGGTGCAGGTGAGCAGCTGGGCGACCGCCATCGGCAAG
AACGGCATGCGCCGCGAGTGGATCGTGACCGACTTCCGCACCGGCGAGACCCTG
CTGCGCGCCACCAGCGTGTGGGTGATGATGAACAAGCTGACCCGCCGCATCAGC
AAGATCCCCGAGGAGGTGTGGCACGAGATCGGCCCCAGCTTCATCGACGCGCCC
CCCCTGCCCACCGTGGAGGACGACGGCCGCAAGCTGACCCGCTTCGACGAGAGC
AGCGCCGACTTCATCCGCAAGGGCCTGACCCCCCGCTGGAGCGACCTGGACATC
AACCAGCACGTGAACAACGTGAAGTACATCGGCTGGCTGCTGGAGAGCGCGCCC
CCCGAGATCCACGAGAGCCACGAGATCGCCAGCCTGACCCTGGAGTACCGCCGC
GAGTGCGGCCGCGACAGCGTGCTGAACAGCGCCACCAAGGTGAGCGACAGCAGC
CAGCTGGGCAAGAGCGCCGTGGAGTGCAACCACCTGGTGCGCCTGCAGAACGGC
GGCGAGATCGTGAAGGGCCGCACCGTGTGGCGCCCCAAGCGCCCCCTGTACAAC
GACGGCGCCGTGGTGGACGTGCCCGCCAAGACCAGCGATGACGATGACAAGCTG
GGATGACTCGAGTTAATTAACTCGAGGCAGCAGCAGCTCGGATAGTATCGACAC

[1130]

ACTCTGGACGCTGGTCGTGTGATGGACTGTTGCCGCCACACTTGCTGCCTTGACC
TGTGAATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGTGTGT
ACGCGCTTTTGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCCAGC
ATCCCCTTCCCTCGTTTCATATCGCTTGCATCCCAACCGCAACTTATCTACGCTGT
CCTGCTATCCCTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCTCGCACAGCCTTG
GTTTGGGCTCCGCCTGTATTCTCCTGGTACTGCAACCTGTAAACCAGCACTGCAA
TGCTGATGCACGGGAAGTAGTGGGATGGGAACACAAATGGAAAGCTGTAGAGCT
CCTTGTTTTCCAGAAGGAGTTGCTCCTTGAGCCTTTCATTCTCAGCCTCGATAACC
TCCAAAGCCGCTCTAATTGTGGAGGGGGTTCGAATTTAAAAGCTTGGAATGTTGG
TTCGTGCGTCTGGAACAAGCCCAGACTTGTTGCTCACTGGGAAAAGGACCATCAG
CTCCAAAAAACTTGCCGCTCAAACCGCGTACCTCTGCTTTCGCGCAATCTGCCCT
GTTGAAATCGCCACCACATTCATATTGTGACGCTTGAGCAGTCTGTAATTGCCTC
AGAATGTGGAATCATCTGCCCCCTGTGCGAGCCCATGCCAGGCATGTCGCGGGC
GAGGACACCCGCCACTCGTACAGCAGACCATTATGCTACCTCACAATAGTTCATA
ACAGTGACCATATTTCTCGAAGCTCCCCAACGAGCACCTCCATGCTCTGAGTGGC
CACCCCCCGGCCCTGGTGCTTGCGGAGGGCAGGTCAACCGGCATGGGGCTACCG
AAATCCCCGACCGGATCCCACCACCCCCGCGATGGGAAGAATCTCTCCCCGGGA
TGTGGGCCCACCACCAGCACAACCTGCTGGCCCAGGCGAGCGTCAAACCATACC
ACACAAATATCCTTGGCATCGGCCCTGAATTCCTTCTGCCGCTCTGCTACCCGGT
GCTTCTGTCCGAAGCAGGGGTTGCTAGGGATCGCTCCGAGTCCGCAAACCCTTGT
CGCGTGGCGGGGCTTGTTCGAGCTTGTTCGAGCTTGAAGAGCCTCTAGAGTCGAC
CTGCAGGCATGCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAAT
TGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAA
GCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGC
CCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACG
CGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCC

[1131]

[1132] 서열 번호 51

[1133] 신나모뮴 캠포라 C14 선호 티오에스테라제의 코돈 최적화 코딩 영역의 적절한 상동 재조합 발현 구조물

GAATTCGCCCTCCCGTGATCACACAGGTGCCTTGCGAGCGTGATCACACTATTTT
GGGGGTCCTACAGTACTGAAATGGTGAGAAGTCGTACTGAAATCAAGGATGAAC
AATGAAAATGGTGCTGTGGTGGCTTCTCAAAGGTCAAGAATCAGTCGCTCGCGTC
AGGAAATCGCGGCGTCAACCAGCGTGGGCGCGGTCAGTGGCCCCGCACTGGTCA
CCATAGCCTCTCCTGCCACAGTAGCGATCCCCTGGGCGTTCACTCTCAGCAGCGG
CTGTACTGCCTCCCAGATTTTCTTCTTCTGGACCTGCGGGCGTGAGAGGATGAGC
AGGGTGGGCCAAGGGCTCAATCCTGAACGGCCCTCATTCGGTTTCCAATCCCACA
ACACATACCCACAGCAGGTCAGACCACGCATTCCACCATGCGCACCAATAACGT
GTCCTTACCTGATTGGGTGTGGCAGGCTCCGTGGACAGGAGTGCCTCGTCCCCCG
CCCAGACCCGCTCCCCCGTCACGGCGGCGTCCGGGACCCGCAGCGGCTCCACCG
CGGTGTGATCCGCGTTGGCGGCGCAGAGCAGCATCCCAGCCGATTTGACCCCGC
GCATGCTCCGAGGCTTGAGGTTGGCCAGCACCACCACCCGCCGGCCGACAAGGT
CCTCCAGGGTCACGTGCCGGACCAGGCCACTCACGATGGTGCGAGGGCCCCCCT
CCTCGCCGAGGTCGATCTGCTCGACGTACAGACTGCGACATGCGTGGCGAGTGGT
CATCAGAAGGAAGCAGGTGTGCAGAAGGGGCACGTGGTTGGTATTGAGAGTAGC
CAAAGCTTTGTGCCAATCAGAAAGTCAACGCAGCTGCCTGCCTGGCTCGCGTACA
ATTCCTTTCTTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGGGCTGCGAGACG
GCTTCCCGGCGCTGCATGCAACACCGATGATGCTTCGACCCCCCGAAGCTCCTTC
GGGGCTGCATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGTTTAAATAGCC

[1134]

AGGCCCCCGATTGCAAAGACATTATAGCGAGCTACCAAAGCATATTCAAACACC
TAGATCACTACCACTTCTACACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAA
GGGGGCGCCTCTTCCCTTCGTTTCAGTCACAACCCGCAAACGGCGCGCCATGCTG
CTGCAGGCCTTCCTGTTCCTGCTGGCCGGCTTCGCCGCCAAGATCAGCGCCTCCA
TGACGAACGAGACGTCCGACCGCCCCCTGGTGCACTTCACCCCCAACAAGGGCT
GGATGAACGACCCCAACGGCCTGTGGTACGACGAGAAGGACGCCAAGTGGCACC
TGTACTTCCAGTACAACCCGAACGACACCGTCTGGGGGGACGCCCTTGTTCTGGGG
CCACGCCACGTCCGACGACCTGACCAACTGGGAGGACCAGCCCATCGCCATCGC
CCCGAAGCGCAACGACTCCGGCGCCTTCTCCGGCTCCATGGTGGTGGACTACAAC
AACACCTCCGGCTTCTTCAACGACACCATCGACCCGCGCCAGCGCTGCGTGGCCA
TCTGGACCTACAACACCCCGGAGTCCGAGGAGCAGTACATCTCCTACAGCCTGG
ACGGCGGCTACACCTTCACCGAGTACCAGAAGAACCCCGTGCTGGCCGCCAACT
CCACCCAGTTCCGCGACCCGAAGGTCTTCTGGTACGAGCCCTCCCAGAAGTGGAT
CATGACCGCGGCCAAGTCCCAGGACTACAAGATCGAGATCTACTCCTCCGACGA
CCTGAAGTCCTGGAAGCTGGAGTCCGCGTTCGCCAACGAGGGCTTCCTCGGCTAC
CAGTACGAGTGCCCCGGCCTGATCGAGGTCCCCACCGAGCAGGACCCCAGCAAG
TCCTACTGGGTGATGTTCATCTCCATCAACCCCGGCGCCCCGGCCGGCGGCTCCT
TCAACCAGTACTTCGTCGGCAGCTTCAACGGCACCCACTTCGAGGCCTTCGACAA
CCAGTCCCGCGTGGTGGACTTCGGCAAGGACTACTACGCCCTGCAGACCTTCTTC
AACACCGACCCGACCTACGGGAGCGCCCTGGGCATCGCGTGGGCCTCCAACTGG
GAGTACTCCGCCTTCGTGCCCACCAACCCCTGGCGCTCCTCCATGTCCCTCGTGC
GCAAGTTCTCCCTCAACACCGAGTACCAGGCCAACCCGGAGACGGAGCTGATCA
ACCTGAAGGCCGAGCCGATCCTGAACATCAGCAACGCCGGCCCCTGGAGCCGGT
TCGCCACCAACACCACGTTGACGAAGGCCAACAGCTACAACGTCGACCTGTCCA
ACAGCACCGGCACCCTGGAGTTCGAGCTGGTGTACGCCGTCAACACCACCCAGA
CGATCTCCAAGTCCGTGTTCGCGGACCTCTCCCTCTGGTTCAAGGGCCTGGAGGA
CCCCGAGGAGTACCTCCGCATGGGCTTCGAGGTGTCCGCGTCCTCCTTCTTCCTG
GACCGCGGGAACAGCAAGGTGAAGTTCGTGAAGGAGAACCCCTACTTCACCAAC
CGCATGAGCGTGAACAACCAGCCCTTCAAGAGCGAGAACGACCTGTCCTACTAC
AAGGTGTACGGCTTGCTGGACCAGAACATCCTGGAGCTGTACTTCAACGACGGC
GACGTCGTGTCCACCAACACCTACTTCATGACCACCGGGAACGCCCTGGGCTCCG
TGAACATGACGACGGGGGTGGACAACCTGTTCTACATCGACAAGTTCCAGGTGC
GCGAGGTCAAGTGATTAATTAACTCGAGGCAGCAGCAGCTCGGATAGTATCGAC
ACACTCTGGACGCTGGTCGTGTGATGGACTGTTGCCGCCACACTTGCTGCCTTGA
CCTGTGAATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGTGT
GTACGCGCTTTTGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCCA
GCATCCCCTCCCTCGTTTCATATCGCTTGCATCCCAACCGCAACTTATCTACGCTG
TCCTGCTATCCCTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCTCGCACAGCCTT
GGTTTGGGCTCCGCCTGTATTCTCCTGGTACTGCAACCTGTAAACCAGCACTGCA
ATGCTGATGCACGGGAAGTAGTGGGATGGGAACACAAATGGAAAGCTTGAGCTC
CTTTCTTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTT
CCCGGCGCTGCATGCAACACCGATGATGCTTCGACCCCCCGAAGCTCCTTCGGGG
CTGCATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGTTTAAATAGCCAGGCC
CCCGATTGCAAAGACATTATAGCGAGCTACCAAAGCCATATTCAAACACCTAGA
TCACTACCACTTCTACACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGG
GCGCCTCTTCCTCTTCGTTTCAGTCACAACCCGCAAACACTAGTATGACGTTCGG
GGTCGCCCTCCCGGCCATGGGCCGCGGTGTCTCCCTTCCCCGGCCCAGGGTCGCG
GTGCGCGCCCAGTCGGCGAGTCAGGTTTTGGAGAGCGGGCGCGCCCCCGACTGG
TCCATGCTGTTCGCCGTGATCACCACCATCTTCTCCGCCGCCGAGAAGCAGTGGA
CCAACCTGGAGTGGAAGCCCAAGCCCAACCCCCCCCCAGCTGCTGGACGACCACT

[1135]

TCGGCCCCCACGGCCTGGTGTTCCGCCGCACCTTCGCCATCCGCAGCTACGAGGT
GGGCCCCGACCGCTCCACCAGCATCGTGGCCGTGATGAACCACCTGCAGGAGGC
CGCCCTGAACCACGCCAAGTCCGTGGGCATCCTGGGCGACGGCTTCGGCACCAC
CCTGGAGATGTCCAAGCGCGACCTGATCTGGGTGGTGAAGCGCACCCACGTGGC
CGTGGAGCGCTACCCCGCCTGGGGCGACACCGTGGAGGTGGAGTGCTGGGTGGG
CGCCTCCGGCAACAACGGCCGCCGCCACGACTTCCTGGTGCGCGACTGCAAGAC
CGGCGAGATCCTGACCCGCTGCACCTCCCTGAGCGTGATGATGAACACCCGCACC
CGCCGCCTGAGCAAGATCCCCGAGGAGGTGCGCGGCGAGATCGGCCCCGCCTTC
ATCGACAACGTGGCCGTGAAGGACGAGGAGATCAAGAAGCCCCAGAAGCTGAA
CGACTCCACCGCCGACTACATCCAGGGCGGCCTGACCCCCCGCTGGAACGACCT
GGACATCAACCAGCACGTGAACAACATCAAGTACGTGGACTGGATCCTGGAGAC
CGTGCCCGACAGCATCTTCGAGAGCCACCACATCTCCTCCTTCACCATCGAGTAC
CGCCGCGAGTGCACCATGGACAGCGTGCTGCAGTCCCTGACCACCGTGAGCGGC
GGCTCCTCCGAGGCCGGCCTGGTGTGCGAGCACCTGCTGCAGCTGGAGGGCGGC
AGCGAGGTGCTGCGCGCCAAGACCGAGTGGCGCCCCAAGCTGACCGACTCCTTC
CGCGGCATCAGCGTGATCCCCGCCGAGTCCAGCGTGATGGACTACAAGGACCAC
GACGGCGACTACAAGGACCACGACATCGACTACAAGGACGACGACGACAAGTG
ACTCGAGGCAGCAGCAGCTCGGATAGTATCGACACACTCTGGACGCTGGTCGTG
TGATGGACTGTTGCCGCCACACTTGCTGCCTTGACCTGTGAATATCCCTGCCGCTT
TTATCAAACAGCCTCAGTGTGTTTGATCTTGTGTGTACGCGCTTTTGCGAGTTGCT
AGCTCTTGTGCTATTTGCGAATACCACCCCCAGCATCCCCTTCCCTCGTTTCATAT
CGCTTGCATCCCAACCGCAACTTATCTACGCTGTCCTGCTATCCCTCAGCGCTGCT
CCTGCTCCTGCTCACTGCCCCTCGCACAGCCTTGGTTTGGGCTCCGCCTGTATTCT
CCTGGTACTGCAACCTGTAAACCAGCACTGCAATGCTGATGCACGGGAAGTAGT
GGGATGGGAACACAAATGGAAAGCTGGTACCCGTACCCATCAGCATCCGGGTGA
ATCTTGGCCTCCAAGATATGGCCAATCCTCACATCCAGCTTGGCAAAATCGACTA
GACTGTCTGCAAGTGGGAATGTGGAGCACAAGGTTGCTTGTAGCGATCGACAGA
CTGGTGGGGTACATTGACAGGTGGGCAGCGCCGCATCCATCGTGCCTGACGCGA
GCGCCGCCGGTTGCTCGCCCGTGCCTGCCGTCAAAGAGCGGCAGAGAAATCGGG
AACCGAAAACGTCACATTGCCTGATGTTGTTACATGCTGGACTAGACTTTCTTGG
CGTGGGTCTGCTCCTCGCCAGGTGCGCGACGCCTCGGGGCTGGGTGCGAGGGAG
CCGTGCGGCCACGCATTTGACAAGACCCAAAGCTCGCATCTCAGACGGTCAACC
GTTCGTATTATACATTCAACATATGGTACATACGCAAAAAGCATGCCAACGATGA
CATAGGCGAATTC

[1136]

[1137] 서열 번호 52

[1138] 클로렐라 프로토테코이데스 스테아로일 ACP 불포화 효소 운반 펩티드를 지니는 쿠페아 후커리아나 C10 선호 티오에스테라제의 코돈 최적화 코딩 영역에 대한 관련 발현 구조물

GGTACCCGCCTGCAACGCAAGGGCAGCCACAGCCGCTCCCACCCGCCGCTGAAC
CGACACGTGCTTGGGCGCCTGCCGCCTGCCTGCCGCATGCTTGTGCTGGTGAGGC
TGGGCAGTGCTGCCATGCTGATTGAGGCTTGGTTCATCGGGTGGAAGCTTATGTG
TGTGCTGGGCTTGCATGCCGGGCAATGCGCATGGTGGCAAGAGGGCGGCAGCAC
TTGCTGGAGCTGCCGCGGTGCCTCCAGGTGGTTCAATCGCGGCAGCCAGAGGGA
TTTCAGATGATCGCGCGTACAGGTTGAGCAGCAGTGTCAGCAAAGGTAGCAGTTT
GCCAGAATGATCGGTTCAGCTGTTAATCAATGCCAGCAAGAGAAGGGGTCAAGT
GCAAACACGGGCATGCCACAGCACGGGCACCGGGGAGTGGAATGGCACCACCA
AGTGTGTGCGAGCCAGCATCGCCGCCTGGCTGTTTCAGCTACAACGGCAGGAGTC
ATCCAACGTAACCATGAGCTGATCAACACTGCAATCATCGGGCGGGCGTGATGC
AAGCATGCCTGGCGAAGACACATGGTGTGCGGATGCTGCCGGCTGCTGCCTGCT

[1139]

GCGCACGCCGTTGAGTTGGCAGCAGGCTCAGCCATGCACTGGATGGCAGCTGGG
CTGCCACTGCAATGTGGTGGATAGGATGCAAGTGGAGCGAATACCAAACCCTCT
GGCTGCTTGCTGGGTTGCATGGCATCGCACCATCAGCAGGAGCGCATGCGAAGG
GACTGGCCCCATGCACGCCATGCCAAACCGGAGCGCACCGAGTGTCCACACTGT
CACCAGGCCCGCAAGCTTTGCAGAACCATGCTCATGGACGCATGTAGCGCTGAC
GTCCCTTGACGGCGCTCCTCTCGGGTGTGGGAAACGCAATGCAGCACAGGCAGC
AGAGGCGGCGGCAGCAGAGCGGCGGCAGCAGCGGCGGGGGCCACCCTTCTTGCG
GGGTCGCGCCCCAGCCAGCGGTGATGCGCTGATCCCAAACGAGTTCACATTCATT
TGCATGCCTGGAGAAGCGAGGCTGGGGCCTTTGGGCTGGTGCAGCCCGCAATGG
AATGCGGGACCGCCAGGCTAGCAGCAAAGGCGCCTCCCCTACTCCGCATCGATG
TTCCATAGTGCATTGGACTGCATTTGGGTGGGGCGGCCGGCTGTTTCTTTCGTGTT
GCAAAACGCGCCAGCTCAGCAACCTGTCCCGTGGGTCCCCCGTGCCGATGAAAT
CGTGTGCACGCCGATCAGCTGATTGCCCGGCTCGCGAAGTAGGCGCCCTCCTTTC
TGCTCGCCCTCTCTCCGTCCCGCCTCTAGAATATCAATGATCGAGCAGGACGGCC
TCCACGCCGGCTCCCCCGCCGCCTGGGTGGAGCGCCTGTTCGGCTACGACTGGGC
CCAGCAGACCATCGGCTGCTCCGACGCCGCCGTGTTCCGCCTGTCCGCCCAGGGC
CGCCCCGTGCTGTTCGTGAAGACCGACCTGTCCGGCGCCCTGAACGAGCTGCAGG
ACGAGGCCGCCCGCCTGTCCTGGCTGGCCACCACCGGCGTGCCCTGCGCCGCCGT
GCTGGACGTGGTGACCGAGGCCGGCCGCGACTGGCTGCTGCTGGGCGAGGTGCC
CGGCCAGGACCTGCTGTCCTCCCACCTGGCCCCCGCCGAGAAGGTGTCCATCATG
GCCGACGCCATGCGCCGCCTGCACACCCTGGACCCCGCCACCTGCCCCTTCGACC
ACCAGGCCAAGCACCGCATCGAGCGCGCCCGCACCCGCATGGAGGCCGGCCTGG
TGGACCAGGACGACCTGGACGAGGAGCACCAGGGCCTGGCCCCCGCCGAGCTGT
TCGCCCGCCTGAAGGCCCGCATGCCCGACGGCGAGGACCTGGTGGTGACCCACG
GCGACGCCTGCCTGCCCAACATCATGGTGGAGAACGGCCGCTTCTCCGGCTTCAT
CGACTGCGGCCGCCTGGGCGTGGCCGACCGCTACCAGGACATCGCCCTGGCCAC
CCGCGACATCGCCGAGGAGCTGGGCGGCGAGTGGGCCGACCGCTTCCTGGTGCT
GTACGGCATCGCCGCCCCCGACTCCCAGCGCATCGCCTTCTACCGCCTGCTGGAC
GAGTTCTTCTGACAATTGGCAGCAGCAGCTCGGATAGTATCGACACACTCTGGAC
GCTGGTCGTGTGATGGACTGTTGCCGCCACACTTGCTGCCTTGACCTGTGAATAT
CCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGTGTGTACGCGCTTT
TGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCCAGCATCCCCTTCC
CTCGTTTCATATCGCTTGCATCCCAACCGCAACTTATCTACGCTGTCCTGCTATCC
CTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCTCGCACAGCCTTGGTTTGGGCTC
CGCCTGTATTCTCCTGGTACTGCAACCTGTAAACCAGCACTGCAATGCTGATGCA
CGGGAAGTAGTGGGATGGGAACACAAATGGAGGATCCCGCGTCTCGAACAGAGC
GCGCAGAGGAACGCTGAAGGTCTCGCCTCTGTCGCACCTCAGCGCGGCATACAC
CACAATAACCACCTGACGAATGCGCTTGGTTCTTCGTCCATTAGCGAAGCGTCCG
GTTCACACACGTGCCACGTTGGCGAGGTGGCAGGTGACAATGATCGGTGGAGCT
GATGGTCGAAACGTTCACAGCCTAGGGATATCGAATTCCTTTCTTGCGCTATGAC
ACTTCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTTCCCGGCGCTGCATGCA
ACACCGATGATGCTTCGACCCCCCGAAGCTCCTTCGGGGCTGCATGGGCGCTCCG
ATGCCGCTCCAGGGCGAGCGCTGTTTAAATAGCCAGGCCCCCGATTGCAAAGAC
ATTATAGCGAGCTACCAAAGCCATATTCAAACACCTAGATCACTACCACTTCTAC
ACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGGGCGCCTCTTCCTCTTC
GTTTCAGTCACAACCCGCAAACACTAGTATGGCCACCGCATCCACTTTCTCGGCG
TTCAATGCCCGCTGCGGCGACCTGCGTCGCTCGGCGGGCTCCGGGCCCCGGCGCC
CAGCGAGGCCCCTCCCCGTGCGCGGGCGCGCCCAGCTGCCCGACTGGAGCCGCC
TGCTGACCGCCATCACCACCGTGTTCGTGAAGTCCAAGCGCCCCGACATGCACGA
CCGCAAGTCCAAGCGCCCCGACATGCTGGTGGACAGCTTCGGCCTGGAGTCCAC

[1140]

CGTGCAGGACGGCCTGGTGTTCCGCCAGTCCTTCTCCATCCGCTCCTACGAGATC
GGCACCGACCGCACCGCCAGCATCGAGACCCTGATGAACCACCTGCAGGAGACC
TCCCTGAACCACTGCAAGAGCACCGGCATCCTGCTGGACGGCTTCGGCCGCACCC
TGGAGATGTGCAAGCGCGACCTGATCTGGGTGGTGATCAAGATGCAGATCAAGG
TGAACCGCTACCCCGCCTGGGGCGACACCGTGGAGATCAACACCCGCTTCAGCC
GCCTGGGCAAGATCGGCATGGGCCGCGACTGGCTGATCTCCGACTGCAACACCG
GCGAGATCCTGGTGCGCGCCACCAGCGCCTACGCCATGATGAACCAGAAGACCC
GCCGCCTGTCCAAGCTGCCCTACGAGGTGCACCAGGAGATCGTGCCCCTGTTCGT
GGACAGCCCCGTGATCGAGGACTCCGACCTGAAGGTGCACAAGTTCAAGGTGAA
GACCGGCGACAGCATCCAGAAGGGCCTGACCCCCGGCTGGAACGACCTGGACGT
GAACCAGCACGTGTCCAACGTGAAGTACATCGGCTGGATCCTGGAGAGCATGCC
CACCGAGGTGCTGGAGACCCAGGAGCTGTGCTCCCTGGCCCTGGAGTACCGCCG
CGAGTGCGGCCGCGACTCCGTGCTGGAGAGCGTGACCGCCATGGACCCCAGCAA
GGTGGGCGTGCGCTCCCAGTACCAGCACCTGCTGCGCCTGGAGGACGGCACCGC
CATCGTGAACGGCGCCACCGAGTGGCGCCCCAAGAACGCCGGCGCCAACGGCGC
CATCTCCACCGGCAAGACCAGCAACGGCAACTCCGTGTCCATGGACTACAAGGA
CCACGACGGCGACTACAAGGACCACGACATCGACTACAAGGACGACGACGACA
AGTGACTCGAGGCAGCAGCAGCTCGGATAGTATCGACACACTCTGGACGCTGGT
CGTGTGATGGACTGTTGCCGCCACACTTGCTGCCTTGACCTGTGAATATCCCTGC
CGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGTGTGTACGCGCTTTTGCGAG
TTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCCAGCATCCCCTTCCCTCGTT
TCATATCGCTTGCATCCCAACCGCAACTTATCTACGCTGTCCTGCTATCCCTCAGC
GCTGCTCCTGCTCCTGCTCACTGCCCCTCGCACAGCCTTGGTTTGGGCTCCGCCTG
TATTCTCCTGGTACTGCAACCTGTAAACCAGCACTGCAATGCTGATGCACGGGAA
GTAGTGGGATGGGAACACAAATGGAAAGCTTGAGCTC

[1141]

[1142] 서열 번호 53

[1143] 클로렐라 프로토테코이데스 스테아로일 ACP 불포화 효소 운반 펩티드를 지니는 움벨루라리아 캘리포니카 C12 선호 티오에스테라제의 코돈 최적화 코딩 영역에 대한 관련 발현 구조물

GGTACCCGCCTGCAACGCAAGGGCAGCCACAGCCGCTCCCACCCGCCGCTGAAC
CGACACGTGCTTGGGCGCCTGCCGCCTGCCTGCCGCATGCTTGTGCTGGTGAGGC
TGGGCAGTGCTGCCATGCTGATTGAGGCTTGGTTCATCGGGTGGAAGCTTATGTG
TGTGCTGGGCTTGCATGCCGGGCAATGCGCATGGTGGCAAGAGGGCGGCAGCAC
TTGCTGGAGCTGCCGCGGTGCCTCCAGGTGGTTCAATCGCGGCAGCCAGAGGGA
TTTCAGATGATCGCGCGTACAGGTTGAGCAGCAGTGTCAGCAAAGGTAGCAGTTT
GCCAGAATGATCGGTTCAGCTGTTAATCAATGCCAGCAAGAGAAGGGGTCAAGT
GCAAACACGGGCATGCCACAGCACGGGCACCGGGGAGTGGAATGGCACCACCA
AGTGTGTGCGAGCCAGCATCGCCGCCTGGCTGTTTCAGCTACAACGGCAGGAGTC
ATCCAACGTAACCATGAGCTGATCAACACTGCAATCATCGGGCGGGCGTGATGC
AAGCATGCCTGGCGAAGACACATGGTGTGCGGATGCTGCCGGCTGCTGCCTGCT
GCGCACGCCGTTGAGTTGGCAGCAGGCTCAGCCATGCACTGGATGGCAGCTGGG
CTGCCACTGCAATGTGGTGGATAGGATGCAAGTGGAGCGAATACCAAACCCTCT
GGCTGCTTGCTGGGTTGCATGGCATCGCACCATCAGCAGGAGCGCATGCGAAGG
GACTGGCCCCATGCACGCCATGCCAAACCGGAGCGCACCGAGTGTCCACACTGT
CACCAGGCCCGCAAGCTTTGCAGAACCATGCTCATGGACGCATGTAGCGCTGAC
GTCCCTTGACGGCGCTCCTCTCGGGTGTGGGAAACGCAATGCAGCACAGGCAGC
AGAGGCGGCGGCAGCAGAGCGGCGGCAGCAGCGGCGGGGGCCACCCTTCTTGCG
GGGTCGCGCCCCAGCCAGCGGTGATGCGCTGATCCCAAACGAGTTCACATTCATT

[1144]

TGCATGCCTGGAGAAGCGAGGCTGGGGCCTTTGGGCTGGTGCAGCCCGCAATGG
AATGCGGGACCGCCAGGCTAGCAGCAAAGGCGCCTCCCCTACTCCGCATCGATG
TTCCATAGTGCATTGGACTGCATTTGGGTGGGGCGGCCGGCTGTTTCTTTCGTGTT
GCAAAACGCGCCAGCTCAGCAACCTGTCCCGTGGGTCCCCCGTGCCGATGAAAT
CGTGTGCACGCCGATCAGCTGATTGCCCGGCTCGCGAAGTAGGCGCCCTCCTTTC
TGCTCGCCCTCTCTCCGTCCCGCCTCTAGAATATCAATGATCGAGCAGGACGGCC
TCCACGCCGGCTCCCCCGCCGCCTGGGTGGAGCGCCTGTTCGGCTACGACTGGGC
CCAGCAGACCATCGGCTGCTCCGACGCCGCCGTGTTCCGCCTGTCCGCCCAGGGC
CGCCCCGTGCTGTTCGTGAAGACCGACCTGTCCGGCGCCCTGAACGAGCTGCAGG
ACGAGGCCGCCCGCCTGTCCTGGCTGGCCACCACCGGCGTGCCCTGCGCCGCCGT
GCTGGACGTGGTGACCGAGGCCGGCCGCGACTGGCTGCTGCTGGGCGAGGTGCC
CGGCCAGGACCTGCTGTCCTCCCACCTGGCCCCCGCCGAGAAGGTGTCCATCATG
GCCGACGCCATGCGCCGCCTGCACACCCTGGACCCCGCCACCTGCCCCTTCGACC
ACCAGGCCAAGCACCGCATCGAGCGCGCCCGCACCCGCATGGAGGCCGGCCTGG
TGGACCAGGACGACCTGGACGAGGAGCACCAGGGCCTGGCCCCCGCCGAGCTGT
TCGCCCGCCTGAAGGCCCGCATGCCCGACGGCGAGGACCTGGTGGTGACCCACG
GCGACGCCTGCCTGCCCAACATCATGGTGGAGAACGGCCGCTTCTCCGGCTTCAT
CGACTGCGGCCGCCTGGGCGTGGCCGACCGCTACCAGGACATCGCCCTGGCCAC
CCGCGACATCGCCGAGGAGCTGGGCGGCGAGTGGGCCGACCGCTTCCTGGTGCT
GTACGGCATCGCCGCCCCCGACTCCCAGCGCATCGCCTTCTACCGCCTGCTGGAC
GAGTTCTTCTGACAATTGGCAGCAGCAGCTCGGATAGTATCGACACACTCTGGAC
GCTGGTCGTGTGATGGACTGTTGCCGCCACACTTGCTGCCTTGACCTGTGAATAT
CCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGTGTGTACGCGCTTT
TGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCCAGCATCCCCTTCC
CTCGTTTCATATCGCTTGCATCCCAACCGCAACTTATCTACGCTGTCCTGCTATCC
CTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCTCGCACAGCCTTGGTTTGGGCTC
CGCCTGTATTCTCCTGGTACTGCAACCTGTAAACCAGCACTGCAATGCTGATGCA
CGGGAAGTAGTGGGATGGGAACACAAATGGAGGATCCCGCGTCTCGAACAGAGC
GCGCAGAGGAACGCTGAAGGTCTCGCCTCTGTCGCACCTCAGCGCGGCATACAC
CACAATAACCACCTGACGAATGCGCTTGGTTCTTCGTCCATTAGCGAAGCGTCCG
GTTCACACACGTGCCACGTTGGCGAGGTGGCAGGTGACAATGATCGGTGGAGCT
GATGGTCGAAACGTTCACAGCCTAGGGATATCGAATTCCTTTCTTGCGCTATGAC
ACTTCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTTCCCGGCGCTGCATGCA
ACACCGATGATGCTTCGACCCCCCGAAGCTCCTTCGGGGCTGCATGGGCGCTCCG
ATGCCGCTCCAGGGCGAGCGCTGTTTAAATAGCCAGGCCCCCGATTGCAAAGAC
ATTATAGCGAGCTACCAAAGCCATATTCAAACACCTAGATCACTACCACTTCTAC
ACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGGGCGCCTCTTCCTCTTC
GTTTCAGTCACAACCCGCAAACACTAGTATGGCCACCGCATCCACTTTCTCGGCG
TTCAATGCCCGCTGCGGCGACCTGCGTCGCTCGGCGGGCTCCGGGCCCCGGCGCC
CAGCGAGGCCCCTCCCCGTGCGCGGGCGCGCCCCCGACTGGTCCATGCTGTTCGC
CGTGATCACCACCATCTTCAGCGCCGCCGAGAAGCAGTGGACCAACCTGGAGTG
GAAGCCCAAGCCCAAGCTGCCCCAGCTGCTGGACGACCACTTCGGCCTGCACGG
CCTGGTGTTCCGCCGCACCTTCGCCATCCGCTCCTACGAGGTGGGCCCCGACCGC
AGCACCTCCATCCTGGCCGTGATGAACCACATGCAGGAGGCCACCCTGAACCAC
GCCAAGAGCGTGGGCATCCTGGGCGACGGCTTCGGCACCACCCTGGAGATGTCC
AAGCGCGACCTGATGTGGGTGGTGCGCCGCACCCACGTGGCCGTGGAGCGCTAC
CCCACCTGGGGCGACACCGTGGAGGTGGAGTGCTGGATCGGCGCCAGCGGCAAC
AACGGCATGCGCCGCGACTTCCTGGTGCGCGACTGCAAGACCGGCGAGATCCTG
ACCCGCTGCACCTCCCTGAGCGTGCTGATGAACACCCGCACCCGCCGCCTGAGCA
CCATCCCCGACGAGGTGCGCGGCGAGATCGGCCCCGCCTTCATCGACAACGTGG

[1145]

CCGTGAAGGACGACGAGATCAAGAAGCTGCAGAAGCTGAACGACTCCACCGCCG
ACTACATCCAGGGCGGCCTGACCCCCCGCTGGAACGACCTGGACGTGAACCAGC
ACGTGAACAACCTGAAGTACGTGGCCTGGGTGTTCGAGACCGTGCCCGACAGCA
TCTTCGAGTCCCACCACATCAGCTCCTTCACCCTGGAGTACCGCCGCGAGTGCAC
CCGCGACTCCGTGCTGCGCAGCCTGACCACCGTGAGCGGCGGCAGCTCCGAGGC
CGGCCTGGTGTGCGACCACCTGCTGCAGCTGGAGGGCGGCAGCGAGGTGCTGCG
CGCCCGCACCGAGTGGCGCCCCAAGCTGACCGACTCCTTCCGCGGCATCAGCGTG
ATCCCCGCCGAGCCCCGCGTGATGGACTACAAGGACCACGACGGCGACTACAAG
GACCACGACATCGACTACAAGGACGACGACGACAAGTGACTCGAGGCAGCAGC
AGCTCGGATAGTATCGACACACTCTGGACGCTGGTCGTGTGATGGACTGTTGCCG
CCACACTTGCTGCCTTGACCTGTGAATATCCCTGCCGCTTTTATCAAACAGCCTCA
GTGTGTTTGATCTTGTGTGTACGCGCTTTTGCGAGTTGCTAGCTGCTTGTGCTATT
TGCGAATACCACCCCCAGCATCCCCTTCCCTCGTTTCATATCGCTTGCATCCCAAC
CGCAACTTATCTACGCTGTCCTGCTATCCCTCAGCGCTGCTCCTGCTCCTGCTCAC
TGCCCCTCGCACAGCCTTGGTTTGGGCTCCGCCTGTATTCTCCTGGTACTGCAACC
TGTAAACCAGCACTGCAATGCTGATGCACGGGAAGTAGTGGGATGGGAACACAA
ATGGAAAGCTTGAGCTC

[1146]

[1147] 서열 번호 54

[1148] 클로렐라 프로토테코이데스 스테아로일 ACP 불포화 효소 운반 펩티드를 지니는 얼머스 아메리카나 광범위 특이

성 티오에스테라제의 코돈 최적화 코딩 영역에 대한 관련 발현 구조물

GGTACCCGCCTGCAACGCAAGGGCAGCCACAGCCGCTCCCACCCGCCGCTGAAC
CGACACGTGCTTGGGCGCCTGCCGCCTGCCTGCCGCATGCTTGTGCTGGTGAGGC
TGGGCAGTGCTGCCATGCTGATTGAGGCTTGGTTCATCGGGTGGAAGCTTATGTG
TGTGCTGGGCTTGCATGCCGGGCAATGCGCATGGTGGCAAGAGGGCGGCAGCAC
TTGCTGGAGCTGCCGCGGTGCCTCCAGGTGGTTCAATCGCGGCAGCCAGAGGGA
TTTCAGATGATCGCGCGTACAGGTTGAGCAGCAGTGTCAGCAAAGGTAGCAGTTT
GCCAGAATGATCGGTTCAGCTGTTAATCAATGCCAGCAAGAGAAGGGGTCAAGT
GCAAACACGGGCATGCCACAGCACGGGCACCGGGGAGTGGAATGGCACCACCA
AGTGTGTGCGAGCCAGCATCGCCGCCTGGCTGTTTCAGCTACAACGGCAGGAGTC
ATCCAACGTAACCATGAGCTGATCAACACTGCAATCATCGGGCGGGCGTGATGC
AAGCATGCCTGGCGAAGACACATGGTGTGCGGATGCTGCCGGCTGCTGCCTGCT
GCGCACGCCGTTGAGTTGGCAGCAGGCTCAGCCATGCACTGGATGGCAGCTGGG
CTGCCACTGCAATGTGGTGGATAGGATGCAAGTGGAGCGAATACCAAACCCTCT
GGCTGCTTGCTGGGTTGCATGGCATCGCACCATCAGCAGGAGCGCATGCGAAGG
GACTGGCCCCATGCACGCCATGCCAAACCGGAGCGCACCGAGTGTCCACACTGT
CACCAGGCCCGCAAGCTTTGCAGAACCATGCTCATGGACGCATGTAGCGCTGAC
GTCCCTTGACGGCGCTCCTCTCGGGTGTGGGAAACGCAATGCAGCACAGGCAGC
AGAGGCGGCGGCAGCAGAGCGGCGGCAGCAGCGGCGGGGGCCACCCTTCTTGCG
GGGTCGCGCCCCAGCCAGCGGTGATGCGCTGATCCCAAACGAGTTCACATTCATT
TGCATGCCTGGAGAAGCGAGGCTGGGGCCTTTGGGCTGGTGCAGCCCGCAATGG
AATGCGGGACCGCCAGGCTAGCAGCAAAGGCGCCTCCCCTACTCCGCATCGATG
TTCCATAGTGCATTGGACTGCATTTGGGTGGGGCGGCCGGCTGTTTCTTTCGTGTT
GCAAAACGCGCCAGCTCAGCAACCTGTCCCGTGGGTCCCCCGTGCCGATGAAAT
CGTGTGCACGCCGATCAGCTGATTGCCCGGCTCGCGAAGTAGGCGCCCTCCTTTC
TGCTCGCCCTCTCTCCGTCCCGCCTCTAGAATATCAATGATCGAGCAGGACGGCC
TCCACGCCGGCTCCCCCGCCGCCTGGGTGGAGCGCCTGTTCGGCTACGACTGGGC
CCAGCAGACCATCGGCTGCTCCGACGCCGCCGTGTTCCGCCTGTCCGCCCAGGGC

[1149]

CGCCCCGTGCTGTTCGTGAAGACCGACCTGTCCGGCGCCCTGAACGAGCTGCAGG
ACGAGGCCGCCCGCCTGTCCTGGCTGGCCACCACCGGCGTGCCCTGCGCCGCCGT
GCTGGACGTGGTGACCGAGGCCGGCCGCGACTGGCTGCTGCTGGGCGAGGTGCC
CGGCCAGGACCTGCTGTCCTCCCACCTGGCCCCCGCCGAGAAGGTGTCCATCATG
GCCGACGCCATGCGCCGCCTGCACACCCTGGACCCCGCCACCTGCCCCTTCGACC
ACCAGGCCAAGCACCGCATCGAGCGCGCCCGCACCCGCATGGAGGCCGGCCTGG
TGGACCAGGACGACCTGGACGAGGAGCACCAGGGCCTGGCCCCCGCCGAGCTGT
TCGCCCGCCTGAAGGCCCGCATGCCCGACGGCGAGGACCTGGTGGTGACCCACG
GCGACGCCTGCCTGCCCAACATCATGGTGGAGAACGGCCGCTTCTCCGGCTTCAT
CGACTGCGGCCGCCTGGGCGTGGCCGACCGCTACCAGGACATCGCCCTGGCCAC
CCGCGACATCGCCGAGGAGCTGGGCGGCGAGTGGGCCGACCGCTTCCTGGTGCT
GTACGGCATCGCCGCCCCCGACTCCCAGCGCATCGCCTTCTACCGCCTGCTGGAC
GAGTTCTTCTGACAATTGGCAGCAGCAGCTCGGATAGTATCGACACACTCTGGAC
GCTGGTCGTGTGATGGACTGTTGCCGCCACACTTGCTGCCTTGACCTGTGAATAT
CCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGTGTGTACGCGCTTT
TGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCCAGCATCCCCTTCC
CTCGTTTCATATCGCTTGCATCCCAACCGCAACTTATCTACGCTGTCCTGCTATCC
CTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCTCGCACAGCCTTGGTTTGGGCTC
CGCCTGTATTCTCCTGGTACTGCAACCTGTAAACCAGCACTGCAATGCTGATGCA
CGGGAAGTAGTGGGATGGGAACACAAATGGAGGATCCCGCGTCTCGAACAGAGC
GCGCAGAGGAACGCTGAAGGTCTCGCCTCTGTCGCACCTCAGCGCGGCATACAC
CACAATAACCACCTGACGAATGCGCTTGGTTCTTCGTCCATTAGCGAAGCGTCCG
GTTCACACACGTGCCACGTTGGCGAGGTGGCAGGTGACAATGATCGGTGGAGCT
GATGGTCGAAACGTTCACAGCCTAGGGATATCGAATTCCTTTCTTGCGCTATGAC
ACTTCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTTCCCGGCGCTGCATGCA
ACACCGATGATGCTTCGACCCCCCGAAGCTCCTTCGGGGCTGCATGGGCGCTCCG
ATGCCGCTCCAGGGCGAGCGCTGTTTAAATAGCCAGGCCCCCGATTGCAAAGAC
ATTATAGCGAGCTACCAAAGCCATATTCAAACACCTAGATCACTACCACTTCTAC
ACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGGGCGCCTCTTCCTCTTC
GTTTCAGTCACAACCCGCAAACACTAGTATGGCCACCGCATCCACTTTCTCGGCG
TTCAATGCCCGCTGCGGCGACCTGCGTCGCTCGGCGGGCTCCGGGCCCCGGCGCC
CAGCGAGGCCCCTCCCCGTGCGCGGGCGCGCCCAGCTGCCCGACTGGAGCATGC
TGCTGGCCGCGATCACCACCCTGTTCCTGGCGGCCGAGAAGCAGTGGATGATGCT
GGACTGGAAGCCCAAGCGCCCCGACATGCTGGTGGACCCCTTCGGCCTGGGCCG
CTTCGTGCAGGACGGCCTGGTGTTCCGCAACAACTTCAGCATCCGCAGCTACGAG
ATCGGCGCGGACCGCACCGCCAGCATCGAGACCCTGATGAACCACCTGCAGGAG
ACCGCCCTGAACCACGTGAAGAGCGTGGGCCTGCTGGAGGACGGCCTGGGCAGC
ACCCGCGAGATGAGCCTGCGCAACCTGATCTGGGTGGTGACCAAGATGCAGGTG
GCGGTGGACCGCTACCCCACCTGGGGCGACGAGGTGCAGGTGAGCAGCTGGGCG
ACCGCCATCGGCAAGAACGGCATGCGCCGCGAGTGGATCGTGACCGACTTCCGC
ACCGGCGAGACCCTGCTGCGCGCCACCAGCGTGTGGGTGATGATGAACAAGCTG
ACCCGCCGCATCAGCAAGATCCCCGAGGAGGTGTGGCACGAGATCGGCCCCAGC
TTCATCGACGCGCCCCCCCTGCCCACCGTGGAGGACGACGGCCGCAAGCTGACC
CGCTTCGACGAGAGCAGCGCCGACTTCATCCGCAAGGGCCTGACCCCCCGCTGG
AGCGACCTGGACATCAACCAGCACGTGAACAACGTGAAGTACATCGGCTGGCTG
CTGGAGAGCGCGCCCCCCGAGATCCACGAGAGCCACGAGATCGCCAGCCTGACC
CTGGAGTACCGCCGCGAGTGCGGCCGCGACAGCGTGCTGAACAGCGCCACCAAG
GTGAGCGACAGCAGCCAGCTGGGCAAGAGCGCCGTGGAGTGCAACCACCTGGTG
CGCCTGCAGAACGGCGGCGAGATCGTGAAGGGCCGCACCGTGTGGCGCCCCAAG
CGCCCCCTGTACAACGACGGCGCCGTGGTGGACGTGCCCGCCAAGACCAGCGAT

[1150]

GACGATGACAAGCTGGGATGACTCGAGGCAGCAGCAGCTCGGATAGTATCGACA
CACTCTGGACGCTGGTCGTGTGATGGACTGTTGCCGCCACACTTGCTGCCTTGAC
CTGTGAATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGTGTG
TACGCGCTTTTGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCCAGC
ATCCCCTTCCCTCGTTTCATATCGCTTGCATCCCAACCGCAACTTATCTACGCTGT
CCTGCTATCCCTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCTCGCACAGCCTTG
GTTTGGGCTCCGCCTGTATTCTCCTGGTACTGCAACCTGTAAACCAGCACTGCAA
TGCTGATGCACGGGAAGTAGTGGGATGGGAACACAAATGGAAAGCTTGAGCTC

[1151]

[1152] 서열 번호 55

[1153] 프로토테카 모리포르미스 FatB/A 프로모터/5'UTR

CCTGTCGATCGAAGAGAAGGAGACATGTGTACATTATTGGTGTGAGGGCGCTGA
ATCGGCCATTTTTTAAAATGATCACGCTCATGCCAATAGACGCGGCACATAACGA
CGTTCAAACCCCCGCCAAAGCCGCGGACAACCCCATCCCTCCACACCCCCCACAC
AAAGAACCCGCCACCGCTTACCTTGCCCACGAGGTAGGCCTTTCGTTGCGCAAAA
CCGGCCTCGGTGATGAATGCATGCCCGTTCCTGACGAGCGCTGCCCGGGCCAACA
CGCTCTTTTGCTGCGTCTCCTCAGGCTTGGGGGCCTCCTTGGGCTTGGGTGCCGCC
ATGATCTGCGCGCATCAGAGAAACGTTGCTGGTAAAAAGGAGCGCCCGGCTGCG
CAATATATATATAGGCATGCCAACACAGCCCAACCTCACTCGGGAGCCCGTCCCA
CCACCCCCAAGTCGCGTGCCTTGACGGCATACTGCTGCAGAAGCTTCATGAGAAT
GATGCCGAACAAGAGGGGCACGAGGACCCAATCCCGGACATCCTTGTCGATAAT
GATCTCGTGAGTCCCCATCGTCCGCCCGACGCTCCGGGGAGCCCGCCGATGCTCA
AGACGAGAGGGCCCTCGACCAGGAGGGGCTGGCCCGGGCGGGCACTGGCGTCG
AAGGTGCGCCCGTCGTTCGCCTGCAGTCCTATGCCACAAAACAAGTCTTCTGACG
GGGTGCGTTTGCTCCCGTGCGGGCAGGCAACAGAGGTATTCACCCTGGTCATGGG
GAGATCGGCGATCGAGCTGGGATAAGAGATACTTCTGGCAAGCAATGACAACTT
GTCAGGACCGGACCGTGCCATATATTTCTCACCTAGCGCCGCAAAACCTAACAAT
TTGGGAGTCACTGTGCCACTGAGTTCGACTGGTAGCTGAATGGAGTCGCTGCTCC
ACTAAACGAATTGTCAGCACCGCCAGCCGGCCGAGGACCCGAGTCATAGCGAGG
GTAGTAGCGCGCC

[1154]

[1155] 서열 번호 56

[1156] 프로토테카 모리포르미스 NRAMP 금속 운반체 프로모터/5'UTR

ACTAATTGCAATCGTGCAGTAATCATCGATATGGTCACAAGTAGATCCCCTACTG
ACACCCTCTCGTACATGTAGGCAATGTCATCGGCGCCGTCCTGCTGACCGATGCC
GACGTAGCAGAGCAGACCCGGGCCGATCTGGGATACGAGCCGGCCCTCCACCTG
CGCTCGAGGTGGAATCAAGTAAATAACCAATACACTTTTCGACACCACACAGAG
TTGCACGGACGGTGGCGTACCTCTACGCTCGCGCTCTTCACGCGCTGGACGACCG
CACGCATGAGCCCGGGTGGCTTGGTCTGGGCTGCAAAAATGCACAACAAACAAG
TATCAGACGCTCATGGATGCACACGCGCTCCCAAGCACGCTCAGACTAAATATTA
CAGTAGCTCGTATCTGATAAGATATCGAGACATACCGCTCAACTCACCCGCAAAC
TGCGCCCCGCCAGGTGATGCGCACAGGGCCCCACCATGCGATCCATCGCATCGCT
CCTCGAGGGCGCTATCACGTGGCCGGAGAGCGTTCACAGCGTACGCCACTGTATC
TGGGCGGTATGCGGTCCGTCAACATGGAGACAGATACCCGCACCACCACCTTGC
AAGCTCTTCCATATTGGAAGTAGAAAATTGTAATTGTATCATCGCACGAGGGGCC
AACTTGCCGTCGGCGAGCTGGGCGACGAACACCACCTGGACGTTGTCGAGACTC
GCTCGTGCCGTGCGCCGGGCCGCTGGGTATCCAGACCGTCGCC

[1157]

[1158] 서열 번호 57

[1159] 프로토테카 모리포르미스 FLAP 편모-관련 단백질 프로모터/5'UTR

CAACGACAACCAGCAGGCAACTCGGTCAGCGACCCAACACGCGAGTCAAATTGT
TGCGTGTTCTTGCCTTGTCTATTTACTGTGATAGCAAGACTGTCGGTCAGTCAATA
CCGCGGTGCGCACGTCGGGGTGCCAAGCCTAGCAGAGCACGGGACGGCTGGTGC
TGTGCGCCAGCTCAGCTCGCTTCGCGACCAATTGTAGGACCGGCAAAGTCACCAA
AACATGCCAGCGGTGCGATTCAATTGGTCATGAGCTCTACAAAATTGTTTTGTGC
GTCGCGCAGGTATCCAACGGCGCGGCAGAGAAAGTTTGACAGCTCTCGATTTCAT
CTCGGAAAAATGGGGAGAATTTATGACACACAAGTGCGCAGGCGGCCCAGGCGG
CCAGCATATTCTGGCGTGACCTGGGCCGCCCACAAAATGCTTGGATGCACTCTAA
AATAATTATATTTGCCATGAACAAGGGAAGAGTTACCGCACCCAGCCCTAGACTT
GGGCGCCCGAGCAAGGTTACGTCAAGCCACCTTCGCCCATCGCCCAACTCCGTAT
TCCCCGACAGCCGCACGTGGCCCTCGCCGGAATGAACCCTGAATCGGCATCACG
CCACGCGTTCGCCAATCGTTCCGCTCTCTGGCTTCATCGGCCTGCGCCTTCACGTC
GTGGTCACGACAGTGCATTCATACTTCCATTTGCACCTCGGCACACACTTTTACG
CATCGCCTACCCTTGCTGCGGCAGTCTAGGGTCACTTTGCAGCCATGGGACAGTG
CTACACCACCGTCGGTGCGCAAAGCTATTTCAAGTGAACCGTGGGCGGAAAAAA
GGAATGTACACTGTCTCAACCGACTCCTACAATTGTTTACCATGCAGATCAGAGC
TCGACGGCCATCATCGAGCAGGTGTGGGGCCTTGGTGGCGCGGCGCGGGGCCCC
AGGGCGTCGCAGGCATTGATGGCACTCTGAGACTTTCGCACGCGCATGAGGGAC
CCCATCAAGAGAAGAGTGTGTCTTTATGTCCCCATTCATGATGATGTATCTTGTG
ATTGTCGCAGTTTGGCAAGTTTAACCGGATCGCCGCTCCAGGTGTGGCGTGGCGG
ATTTTTCTAGGGGTGCTTGAGCAGTCG

[1160]

[1161] 서열 번호 58

[1162] 프로토테카 모리포르미스 SulfRed Sulfite 환원 효소 프로모터/5'UTR

GGCCCAGGGCCCTGCGGATGGCCCACACCAGATCTAGCCTCTCTTATGCCATGCC
CGCCTCGCTGCCCGTCGTATCCCCCCGCCGATCCGCGCGTAGGGGACCGCGGCCT
GACCCACGCCACGAAAGAGCTTTGCTCCTCAATTTCTCGCCAACAGAACCGTATC
AAACGCTCAACGCCTATCCCGAACAATCCGTATTCACACCAAATCGAGTATACCG
GACTGGTTTGCCTAGTCTTGAAGGAAATGATCCCGTCCATGCTCGGAAGGGGGA
GCGGGCGGAGGATCCTACTCATCTCTGAAATGGGATTGGTCCGAAGATGGGTTG
GGCAAGCACGTGCCAAACCCCAGCGAGTTGCTGACGAGCAGGCTCATCCAATCC
CCCGGCGAATCCTCCCTCACGCCCCGCATGCATACAAGTCCCTCCCACACGCCCC
CTCCCATCCATTTTCGCCTGGTCCGAACGCGAGCGGCGTCGAGGCGGACCACTTG
CTCCGCAGCGCCGTCTGGGTCTCCACCCCACAGCGGCTTTGCTGCCAGAGGCACC
CCCCTTGCCCCACCTCCTCTTGCAGCC

[1163]

[1164] 서열 번호 59

[1165] 프로토테카 모리포르미스 SugT 당 운반체 프로모터/5'UTR

CCAGGCAGGCGGTAGGGTTGCCGATTGCTTGAGCGAATTGGAAGATATAATTTTT
TGTGGTGTCCCTGGACGCTGTTTGTGGCGCTCCTTTTTGGAGAAGATTGCGTGGG
GGAGCTTTCCATGTACCACGCTTCCTTCTGAAAGGATTCTGGCCGAGTCCTGATG
AGCCCAAAGAAAACACCTGCCTTTCAGTGCTGGCACTCTGAAAACGTCAACAGA
TGATTATACATGTCACAAAAGGCAGCCGATTAGGAACGGGAGCTCTGGCCGTTC
GTTTGGCTGCCTGGGCTGATTGAAGTGATCCACCCTGTTCGAATGAAGGCGGTCG
AGTCGAATTATCGACCGGAGCTGTCGGGAAGGCGTCCGGGGGCAGAGTGAGGTGC

[1166]

TGCGGCCTGGTTGTCGTTCAAAAAGACCCCGGTAGCCCAACAATCACGAACGAA
AGGAATATAATTGCTTGCATACTATACATTCAGTTTCTATGTGGCGGGTAGACAA
GTCTCATGGGCTTCTAAAGGCTGTCCCTTGAAGGCTACTTATAAAAACTTGCTGC
GCCATGGCACGGATCGCGCTTGCGCAGGCTGCAACCCTGCGCGCAAGGTCAAAT
ACACAGCAAAAGATACTAACAGAATTTCTAAAAACATTTAAATATTTGTTTCGAC
CAGCCAATTGTGGTCGTAGGCACGCAAAAGACTTTGTTTTGCGCCCACCGAGCAT
CCACGCTGGCAGTCAAGCCAGTCCGATGTGCATTGCGTGGCAGCATCGAGGAGC
ATCAAAAACCTCGTGCACGCTTTTCTGTCAATCATCATCAACCACTCCACCATGT
ATACCCGATGCATCGCGGTGCGCAGCGCGCCACGCGTCCCAGACCCGCCCAAAA
ACCCAGCAGCGGCGAAAGCAAATCTTCACTTGCCCGAAACCCCGAGCAGCGGCA
TTCACACGTGGGCGAAAACCCCACTTGCCCTAACAGGCGTATGTCTGCTGTCACG
ATGCCTGACAACGGTATTATAGATATACACTGATTAATGTTTGAGTGTGTGCGAG
TCGCGAATCAGGAATGAATTGCTAGTAGGCACTCCGACCGGGCGGGGGCCGAGG
GACCA

[1167]

[1168] 서열 번호 60

[1169] 프로토테카 모리포르미스 Amt03-암모늄 운반체 프로모터/5'UTR

GGCCGACAGGACGCGCGTCAAAGGTGCTGGGCGTGTATGCCCTGGTCGGCAGGT
CGTTGCTGTTGCTGCGCTCGTGGTTCCGCAACCCTGATTTTGGCGTCTTATTCTGG
CGTGGCAAGCGCTGACGCCCGCGAGCCGGGCCGGCGGCGATGCGGTGTCTCACG
GCTGCCGAGCTCCAAGGGAGGCAAGAGCGCCCGGATCAGCTGAAGGGCTTTACA
CGCAAGGTACAGCCGCTCCTGCAAGGCTGCGTGGTGGACTTGAACCTGTAGGTCC
TCTGCTGAAGTTCCTCCACTACCTCACCAGGCCCAGCAGACCAAAGCACAGGCTT
TTCAGGTCCGTGTCATCCACTCTAAAACACTCGACTACGACCTACTGATGGCCCT
AGATTCTTCATCAACAATGCCTGAGACACTTGCTCAGAATTGAAACTCCCTGAAG
GGACCACCAGAGGCCCTGAGTTGTTCCTTCCCCCCGTGGCGAGCTGCCAGCCAGG
CTGTACCTGTGATCGAGGCTGGCGGGAAAATAGGCTTCGTGTGCTCAGGTCATGG
GAGGTGCAGGACAGCTCATGAAACGCCAACAATCGCACAATTCATGTCAAGCTA
ATCAGCTATTTCCTCTTCACGAGCTGTAATTGTCCCAAAATTCTGGTCTACCGGGG
GTGATCCTTCGTGTACGGGCCCTTCCCTCAACCCTAGGTATGCGCGCATGCGGTC
GCCGCGCAACTCGCGCGAGGGCCGAGGGTTTGGGACGGGCCGTCCCGAAATGCA
GTTGCACCCGGATGCGCGGCGCCTTTCTTGCGATAATTTATGCAATGGACTGCTC
TGCAAATTTCTGGGTCTGTCGCCAACCCTAGGATCAGCGGCGTAGGATTTCGTAA
TCATTCGTCCTGATGGGGAGCTACCGACTACCCTAATATCAGCCCGGCTGCCTGA
CGCCAGCGTCCACTTTTGCGTACACATTCCATTCGTGCCCAAGACATTTCATTGTG
GTGCGAAGCGTCCCCAGTTACGCTCACCTGTTTCCCGACCTCCTTACTGTTCTGTC
GACAGAGCGGGCCCACAGGCCGGTCGCAGCC

[1170]

[1171] 서열 번호 61

[1172] 프로토테카 모리포르미스 Amt02-암모늄 운반체 프로모터/5'UTR

TCACCAGCGGACAAAGCACCGGTGTATCAGGTCCGTGTCATCCACTCTAAAGAG
CTCGACTACGACCTACTGATGGCCCTAGATTCTTCATCAAAAACGCCTGAGACAC
TTGCCCAGGATTGAAACTCCCTGAAGGGACCACCAGGGGCCCTGAGTTGTTCCTT
CCCCCCGTGGCGAGCTGCCAGCCAGGCTGTACCTGTGATCGGGGCTGGCGGGAA
AACAGGCTTCGTGTGCTCAGGTTATGGGAGGTGCAGGACAGCTCATTAAACGCC
AACAATCGCACAATTCATGGCAAGCTAATCAGTTATTTCCCATTAACGAGCTATA
ATTGTCCCAAAATTCTGGTCTACCGGGGGTGATCCTTCGTGTACGGGCCCTTCCCT
CAACCCTAGGTATGCGCACATGCGGTCGCCGCGCAACGCGCGCGAGGGCCGAGG
[1173]

GTTTGGGACGGGCCGTCCCGAAATGCAGTTGCACCCGGATGCGTGGCACCTTTTT
TGCGATAATTTATGCAATGGACTGCTCTGCAAAATTCTGGCTCTGTCGCCAACCC
TAGGATCAGCGGTGTAGGATTTCGTAATCATTCGTCCTGATGGGGAGCTACCGAC
TGCCCTAGTATCAGCCCGACTGCCTGACGCCAGCGTCCACTTTTGTGCACACATT
CCATTCGTGCCCAAGACATTTCATTGTGGTGCGAAGCGTCCCCAGTTACGCTCAC
CTGATCCCCAACCTCCTTATTGTTCTGTCGACAGAGTGGGCCCAGAGGCCGGTCG
CAGCC
[1174]

[1175] 서열 번호 62

[1176] 프로토테카 모리포르미스 Aat01-아미노산 운반체 프로모터/5'UTR

CGAAGGGGTCTGCATCGATTCGCGCGGTCTGGAGGCCAGCGTGACTGCTCGCGA
AAATGCTCTGCCGTGTCGGGCTCTGGCTGGGGCGGCCAGAGATCTCACCGTGCCA
CACGCAACTGCCGCACTCTGTGCCCGCCACCTGGCGCGCACATGCGACCTCTTCC
CCGTCATACCCTCTCCTCATGTGATCTTTCCACACGAGTGACGCAGGTGCGCGGA
GTGGAGGGAATCAGGACGTTTTCAAGGTACCTGCTCGAGCCGTACCAACAGCTG
CCGCCCGGCAAGGAAGAGATCGAGGCAGAGATTGCCCGGCTGGAGGCCCGGATA
ACGGAGCTCAAGAGCAAGCTGTCCGAGTGAGACCGCCCAGGTGCACGTGTCGAC
TCGCTATGACATGTACTCGACACAACATGAGGAATTCATCGAATTTGTAGGAAGC
GGGCATTGGTACGGGAGTGGGAAAGCGAAAAAACCTCCCTCCGGCAGTGCCATC
TGCCGGAGTCGAACGTTGATAGGGTTCTCGTGACAGGGTGTGACCTCTCAGCCTT
GCATCAATTAAACGCTATAGACATTATCAGTAACCGTGAATCCCGCATTGGATGC
CACCCGCGCGACCATTGGGGACCTGCATTACAGATCTAGGTGAGATGACAGCGA
GGCAACTTCGGCCCGCGGCCCAGCTTGCGGCGCACCAATATTGGTCACGGGAAG
CCACACACCGACCATAAATGAATACTTGTAAGCTATGTCAACCGATCAATGGCGT
CGAAAGTGTGCCACGAGGATCCATCTGGCGGGGCGGCGTGGCGCACAAGCGCAG
TCGCAATTTCTCGGACCCATCTGACCTAGGCCCAGCGCCGCGGGAGAAATCCCCG
GCGGGTCCTCCACGCAGTAACCCTAATGAGTATCGAGCGCCGACCATTTACACCA
TCGCCCCCGAAATCCTTCCGACATTATTATTATCTTTTAGATCTTGGAACAGACTC
TGCCAACC
[1177]

[1178] 서열 번호 63

[1179] 프로토테카 모리포르미스 Aat02-아미노산 운반체 프로모터/5'UTR

AGAGAGCGGAGGTGGGGTTGTGAGGTGGGGTTGCTGACCAGGAGCTCGCGTCGC
CGAGCGCGACTCGCACACGGTCCAGTTACCCCCCCCTCCGCCCAAACGCAAGCCT
CCCATCTTGATGCCTTTCCGGCCACCTATACTATTTCTTAGTTCGCTGTAACATCC
AGACCGTCCTGAATAATAACAATGCCCTGTGTCAAGTGCATTCCTAAAAAAATTC
TGTCCCAACCAACAATCCCACCTGAAATACCACCAGCCCTGCCCAGTACACTCTT
CCAATACCATCTCCCTACCTCCACGCGCAAGCGACCCCCATGCGCGACCAGGCTC
GAAAGTGATTTATGACTTGAGACGAGCGAGTGGCGGCGCGGTCGACTGCCTTTTC
ATCACGTGCCGTACGTCGGCGACCGCTAGGGCTTTGCACGGCAACGCACGGCTTC
GCCAACCCGACCAGCCAGGACCTCGACTACTCTACCGCGAATTCGCCTCAAGAA
GTCGCCAAATGTGCCATACACCATTCCTTACAGCACTGTTCAAACTTGATGCCAA
TTTTGACATTCGGGTTGCTCGTTGGCTGCGCCCACATCGGCCGTGAGTGCAGCAG
GCGGGATCGGACACGGAGGACGCGGCGTCACGCCCCGAACGCAGCCCGTAACTC
TACATCAACACGACGTGTTGCGTAATCCCGCCCGGCTGCGCATCGTGCCAACCCA
TTCGCGATGGATGGTCGGAAAATGGTGTGCCAACTGCCCTGAGGGAGGCTCTCG
CGAAACGGGCACGTCCCTGAAACCGAAACTGTGGCCTTGTCGTCGGCCACGCAA
GCACGTGGACCCTAAACACCAAGAAAATCAGTAAACAAGGTTGACATCCTCTAC
[1180]

GGGCGAATTGTTTGCCCAACCCTTCATCGCACACTGCCATTATAATGCATCTAGC
TCGGCGACAAGTTTAGAAAAGGCAGGCTGCATTGTTCCATTTCGCCGTGGCGGCG
TGGGTGCCCATTTTACGAGGTTTGGGCTCCCGGGCAGCGACCGAGCCAGGTCGA
GTCCCTCTCGCCCGTCGACAATGTTGCGAACCCCACAAGCGGCTAACAACAACTT
GATGGTACCTGTACACTGCCAATTCCTTCTTCCCCGGCCGAGGTTTACACGTGAT
GGCCATGGCTTCGCATTCAGGCCGACTTCCCATTCCGACTTTCCAGAGGGTCCGC
GGACGCTGGGGGGTTGGCTGCCTGAGGCCCACCCTTTGTTCCCCGCGTCCCGACAA
ACACAATTGCGTTACATAAGGGGGAGCCGCCCCCGTTCAGAGTGCAGAAATCTTT
CACTATATTTTCCAGTCGTCAGCGAAATCAAGT

[1181]

[1182] 서열 번호 64

[1183] 프로토테카 모리포르미스 Aat03-아미노산 운반체 프로모터/5'UTR

GATGGTGGGGTGTCTGCCTTGGGCTGGGTGATGGAGGCTGGTGGTGCGCGGGTTT
CCTGATGCATTCTATCTACGCAGTGTCATGGTGTCCATTCCACACACCAGTACAC
CCTTACACTAAGGATCCATCCCTCCTTCCCTCTTCAGGACTACATGGACCCCACG
AGCTACCGACCGGGCTTTCTCAAAAACGTCAAGGTCATGTTTGACATGCGGGACG
TGGTGGACGACGTGCAAGGTGCGTCCGGAGTGCGCGCAAATGAGCAAGTCGGGC
AATGTGTCGGGGTGGGCACCGGGGCTGGAGATCCGCGATCCCCGAGAAAACGCC
GTACCACCCCCCGCGCTATTCCCTCGATTGCGCGCAGATGTGGTGACCGACACGG
GGGACAACCTGGCGGACATGGGGCGCCGGACCTGGAAGCACGCCAAGTCGCACA
CGGGGAGGCTCGTGCAGTCCCCCCCATCGTACCTCAAGGGTCTCTTTGGTCGCGA
TCCAAAGTACGCTGGTGGCATGGCATGCCCGAAATGAACATCATGTGTGATCTCC
GATTGCCAATGGCCACCTCCACGGACCACCTTGCAGGCGGAAGCGCAATCCAGG
GCCCGAGCCTGACGAGGACGGAGACTCCTCGTCCAGCGCGGGGTCCCCGACCCG
ACGCAGCAGCCGACCCCTGCTAACCCGGCAACGATCGGACCAGCAACCTTGCTG
TAGTTCCGATCCGTGATGACGGGCATTGCCGCCGCTCGATCCGCTTTGATGACTG
TCTATTATTTGCGCGGAGCCCCCTCGGAACCCTACCCCGCTCTTGCAAGCCCCTTG
CATCGGAGATCCTCGTGCGCCCGCCATGACCCCACTGGATTGCCCAACATCCTTC
TTTATCGTGTAAAATGTGATTCCTCGGCTGCAATCGACTGGCCTTCGCTTCTGGCC
CCAAGAGGGCTCGAACGTGCGGCAGCGAGGGCGCTGACACACCCAAGCCCTAGG
GCTTTCAACGTCGGCTGCCAGGCCGGATAGGGGGATCGCCTCCTTTCCACCACCC
ACCTACGAGGGATTCGAGTCGGCTTCCAGCTCAGCTATTCGGCCGCGCCCCCGGC
CCTGCAGACGTCCTCCAGTTTCCGAACAGGTCGCTCTCAGAACACCTGCCGCGGC
TGCGATACGGCAGGCTCTCAAAGCGTCGAC

[1184]

[1185] 서열 번호 65

[1186] 프로토테카 모리포르미스 Aat04-아미노산 운반체 프로모터/5'UTR

CGCGTGGAGCGGTGCGTGCGGATGCCGCGCGCCTGCCAAGGCCTTTTGTATGCCT
GGCCTGGGAAGTTTCCTGACTGAAGCATCTTCAAGATGCTCTCTCACGACCAGCG
ACACCAACACCGTCACTTTTTGCCCCTCCTGCCGCAGGTGCCACTTTCTACTTTGA
CGTCTTCTCCAGGCGGTACATTGCGGGACTGAGCGCCAATTCGGCCAAGAACAG
CGCTGTCGACTTGAGGAGGCAGGGGTCCGTCGACTCTGCCGAGTGACACGCCTTC
GACCCGACTGTACTACGGCCTGCTGAAGAGTGGGTCTCGCCGGCCGGCGTGACC
GGCCCTGTGCCCACAATCGACCATCTATTCGCTCCTTGTCATCTGGCGCCGTCAAT
TGCCCGCGACTTGACGGCAACTGGCTCGATCGAGTCGTATTGAAAAAGCACGTTT
TGTCCTACAGGGCCGCGGTCCGTTACCAACGTGGTTCTCGTTAGGTTTTCGTCGG
GCGGTGGTGCGCGAACTGTCCGATGCCATCCCGGCAAACCCCAGCAAGGTCGCC
AGTCTGGTTCTGACGCAATAGAGTGCGTTTTGGGCCAGTCTAAAAATTCGTCTGG

[1187]

CATGACGTGGCTCCACATCGTACCCGGAGCCTGCCTTGGTAATGTGAGGCACCGG
TGCCAACTCCATTATGGCAGGCATCGAGCGCGCAGGTGAGTACATGACCTTCCGT
GAATTGGGAAGGCGAGCTTGTGTAACGCCTGCGATCGTGCCAGTGAGGCATCGT
AAACTCAAAATATTTTGTAGAAAGTGTCTGATGCCTGGTGAGGCTGCGTAGGGCA
AGGGCAAGCCCTTGGCAGATGGGTAATGGGTCCGGACCTCACAACAGCAACCCC
GCGTCCCCCTTAGGGCCCCTGAGGCTCGATGGCAGGGCCAGCGAGCCCGCGGCC
AAAGGGCGCCATCCCACGGTCGCCCAACGACTCCACGGGTCCTATACCTCATCTT
GAATGGCACTAAAAACTATAGAATATCGGGCACTGGTGGGCGTCTGGGGTACAG
CTGGCCGAGCGCAGTGGCAAACCCTAGGTCCCGCCTCAAGGGCGATTCCCGGGT
CAATGACACGCAAGCAAGATCACATGGCGCGGTCCGCCTCGCGGCTCCACACCC
AGGCCCTAGTTTCGCAACCCATAAATATCGCCCCGATACCATCATAAGCCAGCAA
ATAATTTTTTATCAGAGTTCCAAACCTCCTCAGCTGTGGGAAACCAGCCCACTCT
GAACG

[1188]

[1189] 서열 번호 66

[1190] 프로토테카 모리포르미스 Aat05-아미노산 운반체 프로모터/5'UTR

CCGAGCAGTTCATGGCCAAGTACAAGGACTAGAGACCGGAGGTCGGTAGGCTGA
ATGGAGCTGGCGTCGTCGTGCGCGACGTGCACGCGATGCGATACTACGACCCCA
CAAACGCATGCCTCCCATCTTGATGCCTTTCCGGCCATTTATACTATTTCTCATTT
CGCTGTAACATCTTGAATAATAGAATTGCCCTGTGTCAAGTGGATTCCAAGAAAT
ATTCTGTCCCAACAAAACAACCCAACCTGAAAACAACCTCAAATACCACCAGCC
CTGCCCACCTGCCCAGTACACTTTTCCAATACCATCTCCCTACCTTCACGCGCAAG
CGGCACCCATGCGCGACCAGGCTCGAAAGGATTTCACGACTCAGGACGAGCGAG
TGGCGGCGCGACCGCCTGCCTGTTCGTCACGTGCCGTACGTCGGCGACCGCTAGA
GCTTTGCCTGGCAACCCCCGGCTTCGTCAACCCGGCCAGCCAGGATCTCGACCAC
TCTACCGCGAAATCGCCTCAAGAAGTCGCCAAAAGTGCCGTACACCATGCTTCGC
AGCGCTGTTCAAACTTGATGCCAATCTTGACAATCAGGTTGCTCGTTGGCTGCGT
CCACATCGGCCGTGATTGCAGCAGGCGGGGATCGGACACGGAGGACGCGGCGTC
ACGCCGCGAACGCAGCCCGTAACTCTACATCAACGCGATATGTTGCGTAATCCCG
CCCGGCTGCGCATTGTGACAACCCATTCGCGATGGATGGTCGGAAAATGGTGTGC
CAACTGCCCTGAGGGACTCTCTCGCGAAACGGGCACGTCCCTGTATCCGAAACTG
TGGCATGGCCTTGTCGACCACGCAAGCACGTGGACCCTAACACCACGAAAATAA
GTAAAAAAGGTTGACATCCTCTACGAGCGAATTGTTTGCTCGACCCTTCATCGCA
CACTGTCATTATAATGCATCTAGCTCGGCGACAAGTTTAAAAAAGGCAGGCTGCA
TTATTCCATTTTGCCGTGGCGGCATGGGTGCCCATTTTATGAGGTTTGGGCTCTTG
GGCAGCGACCGAGCCAGGTTGAGTCCCTCTCGCCCGTCGACAACGTTCCAAAGC
CCATAAGTGGCTAATAAACAACTTGATGGTACCTGTACACTGCCAGTTCCTTCTT
CCCCGGCCGAGGTTTACACGTGATGGCCATGGCTTCGCGTTTCAGGCTGACTTCC
CATTCCGACTTTCCAGAGGGTCCGCGGACGCCGGGGGTTGGCTGCGTGAGGCCC
ACCCCTTGTTCCCCGCGTCCCGACAAACACAATTGCGTTACATAAGGGGGAAGCC
GCCCCCCGTTCAGAGTGCAAACATCTTTCATTATATTTTTCAGTCGTCAGCGAAAT
CAAGTATGTCGCTGACAGGCATGAAGGCC

[1191]

[1192] 서열 번호 67

[1193] 프로토테카 모리포르미스 Aat01의 추정 프로모터/5'UTR을 시험하기 위한 리포터 구조물의 관련 부분

GCTCTTCGGCCGCCGCCACTCCTGCTCGAGCGCGCCCGACTCGCGCTCCGCCTGC
GCCCGCGCGTGCGCCGCCAGCGCCTTGGCCTTTTCGCCGCGCTCGTGCGCGTCGC

[1194]

TGATGTCCATCACCAGGTCCATGAGGTCTGCCTTGCGCCGGCTGAGCCACTGCTT
CGTCCGGGCGGCCAAGAGGAGCATGAGGGAGGACTCCTGGTCCAGGGTCCTGAC
GTGGTCGCGGCTCTGGGAGCGGGCCAGCATCATCTGGCTCTGCCGCACCGAGGC
CGCCTCCAACTGGTCCTCCAGCAGCCGCAGTCGCCGCCGACCCTGGCAGAGGAA
GACAGGTGAGGGGGGTATGAATTGTACAGAACAACCACGAGCCTTGTCTAGGCA
GAATCCCTACCAGTCATGGCTTTACCTGGATGACGGCCTGCGAACAGCTGTCCAG
CGACCCTCGCTGCCGCCGCTTCTCCCGCACGCTTCTTTCCAGCACCGTGATGGCG
CGAGCCAGCGCCGCACGCTGGCGCTGCGCTTCGCCGATCTGAGGACAGTCGGGG
AACTCTGATCAGTCTAAACCCCCTTGCGCGTTAGTGTTGCCATCCTTTGCAGACC
GGTGAGAGCCGACTTGTTGTGCGCCACCCCCCACACCACCTCCTCCCAGACCAAT
TCTGTCACCTTTTTGGCGAAGGCATCGGCCTCGGCCTGCAGAGAGGACAGCAGTG
CCCAGCCGCTGGGGGGTTGGCGGATGCACGCTCAGGTACCCTTTCTTGCGCTATGA
CACTTCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTTCCCGGCGCTGCATGC
AACACCGATGATGCTTCGACCCCCCGAAGCTCCTTCGGGGCTGCATGGGCGCTCC
GATGCCGCTCCAGGGCGAGCGCTGTTTAAATAGCCAGGCCCCCGATTGCAAAGA
CATTATAGCGAGCTACCAAAGCCATATTCAAACACCTAGATCACTACCACTTCTA
CACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGGGCGCCTCTTCCTCTT
CGTTTCAGTCACAACCCGCAAACGGCGCGCCATGCTGCTGCAGGCCTTCCTGTTC
CTGCTGGCCGGCTTCGCCGCCAAGATCAGCGCCTCCATGACGAACGAGACGTCC
GACCGCCCCCTGGTGCACTTCACCCCCAACAAGGGCTGGATGAACGACCCCAAC
GGCCTGTGGTACGACGAGAAGGACGCCAAGTGGCACCTGTACTTCCAGTACAAC
CCGAACGACACCGTCTGGGGGACGCCCTTGTTCTGGGGCCACGCCACGTCCGAC
GACCTGACCAACTGGGAGGACCAGCCCATCGCCATCGCCCCGAAGCGCAACGAC
TCCGGCGCCTTCTCCGGCTCCATGGTGGTGGACTACAACAACACCTCCGGCTTCT
TCAACGACACCATCGACCCGCGCCAGCGCTGCGTGGCCATCTGGACCTACAACA
CCCCGGAGTCCGAGGAGCAGTACATCTCCTACAGCCTGGACGGCGGCTACACCTT
CACCGAGTACCAGAAGAACCCCGTGCTGGCCGCCAACTCCACCCAGTTCCGCGA
CCCGAAGGTCTTCTGGTACGAGCCCTCCCAGAAGTGGATCATGACCGCGGCCAA
GTCCCAGGACTACAAGATCGAGATCTACTCCTCCGACGACCTGAAGTCCTGGAA
GCTGGAGTCCGCGTTCGCCAACGAGGGCTTCCTCGGCTACCAGTACGAGTGCCCC
GGCCTGATCGAGGTCCCCACCGAGCAGGACCCCAGCAAGTCCTACTGGGTGATG
TTCATCTCCATCAACCCCGGCGCCCCGGCCGGCGGCTCCTTCAACCAGTACTTCG
TCGGCAGCTTCAACGGCACCCACTTCGAGGCCTTCGACAACCAGTCCCGCGTGGT
GGACTTCGGCAAGGACTACTACGCCCTGCAGACCTTCTTCAACACCGACCCGACC
TACGGGAGCGCCCTGGGCATCGCGTGGGCCTCCAACTGGGAGTACTCCGCCTTCG
TGCCCACCAACCCCTGGCGCTCCTCCATGTCCCTCGTGCGCAAGTTCTCCCTCAAC
ACCGAGTACCAGGCCAACCCGGAGACGGAGCTGATCAACCTGAAGGCCGAGCCG
ATCCTGAACATCAGCAACGCCGGCCCCTGGAGCCGGTTCGCCACCAACACCACG
TTGACGAAGGCCAACAGCTACAACGTCGACCTGTCCAACAGCACCGGCACCCTG
GAGTTCGAGCTGGTGTACGCCGTCAACACCACCCAGACGATCTCCAAGTCCGTGT
TCGCGGACCTCTCCCTCTGGTTCAAGGGCCTGGAGGACCCCGAGGAGTACCTCCG
CATGGGCTTCGAGGTGTCCGCGTCCTCCTTCTTCCTGGACCGCGGGAACAGCAAG
GTGAAGTTCGTGAAGGAGAACCCCTACTTCACCAACCGCATGAGCGTGAACAAC
CAGCCCTTCAAGAGCGAGAACGACCTGTCCTACTACAAGGTGTACGGCTTGCTGG
ACCAGAACATCCTGGAGCTGTACTTCAACGACGGCGACGTCGTGTCCACCAACA
CCTACTTCATGACCACCGGGAACGCCCTGGGCTCCGTGAACATGACGACGGGGG
TGGACAACCTGTTCTACATCGACAAGTTCCAGGTGCGCGAGGTCAAGTGACAATT
GGCAGCAGCAGCTCGGATAGTATCGACACACTCTGGACGCTGGTCGTGTGATGG
ACTGTTGCCGCCACACTTGCTGCCTTGACCTGTGAATATCCCTGCCGCTTTTATCA
AACAGCCTCAGTGTGTTTGATCTTGTGTGTACGCGCTTTTGCGAGTTGCTAGCTGC

[1195]

TTGTGCTATTTGCGAATACCACCCCCAGCATCCCCTTCCCTCGTTTCATATCGCTT
GCATCCCAACCGCAACTTATCTACGCTGTCCTGCTATCCCTCAGCGCTGCTCCTGC
TCCTGCTCACTGCCCCTCGCACAGCCTTGGTTTGGGCTCCGCCTGTATTCTCCTGG
TACTGCAACCTGTAAACCAGCACTGCAATGCTGATGCACGGGAAGTAGTGGGAT
GGGAACACAAATGGAGGATCCCGCGTCTCGAACAGAGCGCGCAGAGGAACGCT
GAAGGTCTCGCCTCTGTCGCACCTCAGCGCGGCATACACCACAATAACCACCTGA
CGAATGCGCTTGGTTCTTCGTCCATTAGCGAAGCGTCCGGTTCACACACGTGCCA
CGTTGGCGAGGTGGCAGGTGACAATGATCGGTGGAGCTGATGGTCGAAACGTTC
ACAGCCTAGGGATATCGAATTCATCGAATTTGTAGGAAGCGGGCATTGGTACGG
GAGTGGGAAAGCGAAAAAACCTCCCTCCGGCAGTGCCATCTGCCGGAGTCGAAC
GTTGATAGGGTTCTCGTGACAGGGTGTGACCTCTCAGCCTTGCATCAATTAAACG
CTATAGACATTATCAGTAACCGTGAATCCCGCATTGGATGCCACCCGCGCGACCA
TTGGGGACCTGCATTACAGATCTAGGTGAGATGACAGCGAGGCAACTTCGGCCC
GCGGCCCAGCTTGCGGCGCACCAATATTGGTCACGGGAAGCCACACACCGACCA
TAAATGAATACTTGTAAGCTATGTCAACCGATCAATGGCGTCGAAAGTGTGCCAC
GAGGATCCATCTGGCGGGGCGGCGTGGCGCACAAGCGCAGTCGCAATTTCTCGG
ACCCATCTGACCTAGGCCCAGCGCCGCGGGAGAAATCCCCGGCGGGTCCTCCAC
GCAGTAACCCTAATGAGTATCGAGCGCCGACCATTTACACCATCGCCCCCCGAAA
TCCTTCCGACATTATTATTATCTTTTAGATCTTGGAACAGACTCTGCCAACCACTA
GTATGGCCACCGCATCCACTTTCTCGGCGTTCAATGCCCGCTGCGGCGACCTGCG
TCGCTCGGCGGGCTCCGGGCCCCGGCGCCCAGCGAGGCCCCTCCCCGTGCGCGG
GCGCGCCCCCGACTGGTCCATGCTGTTCGCCGTGATCACCACCATCTTCTCCGCC
GCCGAGAAGCAGTGGACCAACCTGGAGTGGAAGCCCAAGCCCAACCCCCCCCAG
CTGCTGGACGACCACTTCGGCCCCCACGGCCTGGTGTTCCGCCGCACCTTCGCCA
TCCGCAGCTACGAGGTGGGCCCCGACCGCTCCACCAGCATCGTGGCCGTGATGA
ACCACCTGCAGGAGGCCGCCCTGAACCACGCCAAGTCCGTGGGCATCCTGGGCG
ACGGCTTCGGCACCACCCTGGAGATGTCCAAGCGCGACCTGATCTGGGTGGTGA
AGCGCACCCACGTGGCCGTGGAGCGCTACCCCGCCTGGGGCGACACCGTGGAGG
TGGAGTGCTGGGTGGGCGCCTCCGGCAACAACGGCCGCCGCCACGACTTCCTGG
TGCGCGACTGCAAGACCGGCGAGATCCTGACCCGCTGCACCTCCCTGAGCGTGAT
GATGAACACCCGCACCCGCCGCCTGAGCAAGATCCCCGAGGAGGTGCGCGGCGA
GATCGGCCCCGCCTTCATCGACAACGTGGCCGTGAAGGACGAGGAGATCAAGAA
GCCCCAGAAGCTGAACGACTCCACCGCCGACTACATCCAGGGCGGCCTGACCCC
CCGCTGGAACGACCTGGACATCAACCAGCACGTGAACAACATCAAGTACGTGGA
CTGGATCCTGGAGACCGTGCCCGACAGCATCTTCGAGAGCCACCACATCTCCTCC
TTCACCATCGAGTACCGCCGCGAGTGCACCATGGACAGCGTGCTGCAGTCCCTGA
CCACCGTGAGCGGCGGCTCCTCCGAGGCCGGCCTGGTGTGCGAGCACCTGCTGC
AGCTGGAGGGCGGCAGCGAGGTGCTGCGCGCCAAGACCGAGTGGCGCCCCAAGC
TGACCGACTCCTTCCGCGGCATCAGCGTGATCCCCGCCGAGTCCAGCGTGATGGA
CTACAAGGACCACGACGGCGACTACAAGGACCACGACATCGACTACAAGGACGA
CGACGACAAGTGACTCGAGTTAATTAACTCGAGGCAGCAGCAGCTCGGATAGTA
TCGACACACTCTGGACGCTGGTCGTGTGATGGACTGTTGCCGCCACACTTGCTGC
CTTGACCTGTGAATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCT
TGTGTGTACGCGCTTTTGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACC
CCCAGCATCCCCTTCCCTCGTTTCATATCGCTTGCATCCCAACCGCAACTTATCTA
CGCTGTCCTGCTATCCCTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCTCGCACA
GCCTTGGTTTGGGCTCCGCCTGTATTCTCCTGGTACTGCAACCTGTAAACCAGCA
CTGCAATGCTGATGCACGGGAAGTAGTGGGATGGGAACACAAATGGAAAGCTGT
AGAGCTCCTTGTTTTCCAGAAGGAGTTGCTCCTTGAGCCTTTCATTCTCAGCCTCG
ATAACCTCCAAAGCCGCTCTAATTGTGGAGGGGGTTCGAATTTAAAAGCTTGGAA

[1196]

TGTTGGTTCGTGCGTCTGGAACAAGCCCAGACTTGTTGCTCACTGGGAAAAGGAC
CATCAGCTCCAAAAAACTTGCCGCTCAAACCGCGTACCTCTGCTTTCGCGCAATC
TGCCCTGTTGAAATCGCCACCACATTCATATTGTGACGCTTGAGCAGTCTGTAAT
TGCCTCAGAATGTGGAATCATCTGCCCCCTGTGCGAGCCCATGCCAGGCATGTCG
CGGGCGAGGACACCCGCCACTCGTACAGCAGACCATTATGCTACCTCACAATAG
TTCATAACAGTGACCATATTTCTCGAAGCTCCCCAACGAGCACCTCCATGCTCTG
AGTGGCCACCCCCCGGCCCTGGTGCTTGCGGAGGGCAGGTCAACCGGCATGGGG
CTACCGAAATCCCCGACCGGATCCCACCACCCCCGCGATGGGAAGAATCTCTCCC
CGGGATGTGGGCCCACCACCAGCACAACCTGCTGGCCCAGGCGAGCGTCAAACC
ATACCACACAAATATCCTTGGCATCGGCCCTGAATTCCTTCTGCCGCTCTGCTACC
CGGTGCTTCTGTCCGAAGCAGGGGTTGCTAGGGATCGCTCCGAGTCCGCAAACCC
TTGTCGCGTGGCGGGGCTTGTTCGAGCTTGTTCGAGCTTGAAGAGCCTCTAGAGT
CGACCTGCAGGCATGCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTG
AAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGT
AAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCAC
TGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCA
ACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCC

[1197]

[1198] 서열 번호 68

[1199] 씨. 캠포라 티오에스테라제 정방향 프라이머

[1200] TACCCCGCCTGGGGCGACAC

[1201] 서열 번호 69

[1202] 씨. 캠포라 티오에스테라제 역방향 프라이머

[1203] CTTGCTCAGGCGGCGGGTGC

[1204] 서열 번호 70

[1205] cd189 정방향 프라이머

[1206] CCGGATCTCGGCCAGGGCTA

[1207] 서열 번호 71

[1208] cd189 역방향 프라이머

[1209] TCGATGTCGTGCACCGTCGC

[1210] 서열 번호 72

[1211] 프로토테카 모리포르미스 델타 12 FAD 녹아웃 상동 재조합 표적화 구조물의 5' 공여체 DNA 서열

[1212] GCTCTTCGGGTTTGCTCACCCGCGAGGTCGACGCCCAGCATGGCTATCAAGACGA
ACAGGCAGCCTGTGGAGAAGCCTCCGTTCACGATCGGGACGCTGCGCAAGGCCA
TCCCCGCGCACTGTTTCGAGCGCTCGGCGCTTCGTAGCAGCATGTACCTGGCCTT
TGACATCGCGGTCATGTCCCTGCTCTACGTCGCGTCGACGTACATCGACCCTGCG
CCGGTGCCTACGTGGGTCAAGTATGGCGTCATGTGGCCGCTCTACTGGTTCTTCC
AGGTGTGTGTGAGGGTTGTGGTTGCCCGTATCGAGGTCCTGGTGGCGCGCATGGG
GGAGAAGGCGCCTGTCCCGCTGACCCCCCCCGGCTACCCTCCCGGCACCTTCCAGG
GCGCCTTCGGCACGGGTGTCTGGGTGTGCGCGCACGAGTGCGGCCACCAGGCCTT

[1213] TTCCTCCAGCCAGGCCATCAACGACGGCGTGGGCCTGGTGTTCCACAGCCTGCTG
CTGGTGCCCTACTACTCCTGGAAGCACTCGCACCGCCGCCACCACTCCAACACGG
GGTGCCTGGACAAGGACGAGGTGTTTGTGCCGCCGCACCGCGCAGTGGCGCACG
AGGGCCTGGAGTGGGAGGAGTGGCTGCCCATCCGCATGGGCAAGGTGCTGGTCA
CCCTGACCCTGGGCTGGCCGCTGTACCTCATGTTCAACGTCGCCTCGCGGCCGTA
CCCGCGCTTCGCCAACCACTTTGACCCGTGGTCGCCCATCTTCAGCAAGCGCGAG
GTACCCTTTCTTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGGGCTGCGAGAC
GGCTTCCCGGCGCTGCATGCAACACCGATGATGCTTCGACCCCCCGAAGCTCCTT
CGGGGCTGCATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGTTTAAATAGC
CAGGCCCCCGATTGCAAAGACATTATAGCGAGCTACCAAAGCCATATTCAAACA
CCTAGATCACTACCACTTCTACACAGGCCACTCGAGCTTGTGATCGCACTCCGCT
AAGGGGGCGCCTCTTCCTCTTCGTTTCAGTCACAACCCGCAAACGGCGCGCC

[1214] 서열 번호 73

[1215] 프로토테카 모리포르미스 델타 12 FAD 녹아웃 상동 재조합 표적화 구조물의 3' 공여체 DNA 서열

CAATTGGCAGCAGCAGCTCGGATAGTATCGACACACTCTGGACGCTGGTCGTGTG
ATGGACTGTTGCCGCCACACTTGCTGCCTTGACCTGTGAATATCCCTGCCGCTTTT
ATCAAACAGCCTCAGTGTGTTTGATCTTGTGTGTACGCGCTTTTGCGAGTTGCTAG
CTGCTTGTGCTATTTGCGAATACCACCCCCAGCATCCCCTTCCCTCGTTTCATATC
GCTTGCATCCCAACCGCAACTTATCTACGCTGTCCTGCTATCCCTCAGCGCTGCTC
CTGCTCCTGCTCACTGCCCCTCGCACAGCCTTGGTTTGGGCTCCGCCTGTATTCTC
CTGGTACTGCAACCTGTAAACCAGCACTGCAATGCTGATGCACGGGAAGTAGTG
GGATGGGAACACAAATGGAGCATCGAGGTGGTCATCTCCGACCTCGCGTTGGTG
GCGGTGCTCAGCGGGCTCAGCGTGCTGGGCCGCACCATGGGCTGGGCCTGGCTG
GTCAAGACCTACGTGGTGCCCTACATGATCGTGAACATGTGGCTGGTGCTCATCA
CGCTGCTCCAGCACACGCACCCGGCCCTGCCGCACTACTTCGAGAAGGACTGGG
ACTGGCTACGCGGCGCCATGGCCACCGTCGACCGCTCCATGGGCCCGCCCTTCAT
GGACAGCATCCTGCACCACATCTCCGACACCCACGTGCTGCACCACCTCTTCAGC
ACCATCCCGCACTACCACGCCGAGGAGGCCTCCGCCGCCATCCGGCCCATCCTGG
GCAAGTACTACCAATCCGACAGCCGCTGGGTCGGCCGCGCCCTGTGGGAGGACT
GGCGCGACTGCCGCTACGTCGTCCCCGACGCGCCCGAGGACGACTCCGCGCTCTG
GTTCCACAAGTGAGCGCGCCTGCGCGAGGACGCAGAACAACGCTGCCGCCGTGT
CTTTTGCACGCGCGACTCCGGCGCTTCGCTGGTGGCACCCCCATAAAGAAACCCT
CAATTCTGTTTGTGGAAGACACGGTGTACCCCCACCCACCCACCTGCACCTCTAT
TATTGGTATTATTGACGCGGGAGTGGGCGTTGTACCCTACAACGTAGCTTCTCTA
GTTTTCAGCTGGCTCCCACCATTGTAAAGAGCCTCTAGAGTCGACCTGCAGGCAT
GCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCT
CACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGC
CTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAG
TCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGA
GGCGGTTTGCGTATTGGGCGCTCTTCC

[1216]

[1217] 서열 번호 74

[1218] 프로토테카 모리포르미스 델타 12 FAD 녹아웃 상동 재조합 표적화 구조물

GCTCTTCGGGTTTGCTCACCCGCGAGGTCGACGCCCAGCATGGCTATCAAGACGA
ACAGGCAGCCTGTGGAGAAGCCTCCGTTCACGATCGGGACGCTGCGCAAGGCCA

[1219]

TCCCCGCGCACTGTTTCGAGCGCTCGGCGCTTCGTAGCAGCATGTACCTGGCCTT
TGACATCGCGGTCATGTCCCTGCTCTACGTCGCGTCGACGTACATCGACCCTGCG
CCGGTGCCTACGTGGGTCAAGTATGGCGTCATGTGGCCGCTCTACTGGTTCTTCC
AGGTGTGTGTGAGGGTTGTGGTTGCCCGTATCGAGGTCCTGGTGGCGCGCATGGG
GGAGAAGGCGCCTGTCCCGCTGACCCCCCCGGCTACCCTCCCGGCACCTTCCAGG
GCGCCTTCGGCACGGGTGTCTGGGTGTGCGCGCACGAGTGCGGCCACCAGGCCTT
TTCCTCCAGCCAGGCCATCAACGACGGCGTGGGCCTGGTGTTCCACAGCCTGCTG
CTGGTGCCCTACTACTCCTGGAAGCACTCGCACCGCCGCCACCACTCCAACACGG
GGTGCCTGGACAAGGACGAGGTGTTTGTGCCGCCGCACCGCGCAGTGGCGCACG
AGGGCCTGGAGTGGGAGGAGTGGCTGCCCATCCGCATGGGCAAGGTGCTGGTCA
CCCTGACCCTGGGCTGGCCGCTGTACCTCATGTTCAACGTCGCCTCGCGGCCGTA
CCCGCGCTTCGCCAACCACTTTGACCCGTGGTCGCCCATCTTCAGCAAGCGCGAG
GTACCCTTTCTTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGGGCTGCGAGAC
GGCTTCCCGGCGCTGCATGCAACACCGATGATGCTTCGACCCCCCGAAGCTCCTT
CGGGGCTGCATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGTTTAAATAGC
CAGGCCCCCGATTGCAAAGACATTATAGCGAGCTACCAAAGCCATATTCAAACA
CCTAGATCACTACCACTTCTACACAGGCCACTCGAGCTTGTGATCGCACTCCGCT
AAGGGGGCGCCTCTTCCTCTTCGTTTCAGTCACAACCCGCAAACGGCGCGCCATG
CTGCTGCAGGCCTTCCTGTTCCTGCTGGCCGGCTTCGCCGCCAAGATCAGCGCCT
CCATGACGAACGAGACGTCCGACCGCCCCCTGGTGCACTTCACCCCCAACAAGG
GCTGGATGAACGACCCCAACGGCCTGTGGTACGACGAGAAGGACGCCAAGTGGC
ACCTGTACTTCCAGTACAACCCGAACGACACCGTCTGGGGGACGCCCTTGTTCTG
GGGCCACGCCACGTCCGACGACCTGACCAACTGGGAGGACCAGCCCATCGCCAT
CGCCCCGAAGCGCAACGACTCCGGCGCCTTCTCCGGCTCCATGGTGGTGGACTAC
AACAACACCTCCGGCTTCTTCAACGACACCATCGACCCGCGCCAGCGCTGCGTGG
CCATCTGGACCTACAACACCCCGGAGTCCGAGGAGCAGTACATCTCCTACAGCCT
GGACGGCGGCTACACCTTCACCGAGTACCAGAAGAACCCCGTGCTGGCCGCCAA
CTCCACCCAGTTCCGCGACCCGAAGGTCTTCTGGTACGAGCCCTCCCAGAAGTGG
ATCATGACCGCGGCCAAGTCCCAGGACTACAAGATCGAGATCTACTCCTCCGAC
GACCTGAAGTCCTGGAAGCTGGAGTCCGCGTTCGCCAACGAGGGCTTCCTCGGCT
ACCAGTACGAGTGCCCCGGCCTGATCGAGGTCCCCACCGAGCAGGACCCCAGCA
AGTCCTACTGGGTGATGTTCATCTCCATCAACCCCGGCGCCCCGGCCGGCGGCTC
CTTCAACCAGTACTTCGTCGGCAGCTTCAACGGCACCCACTTCGAGGCCTTCGAC
AACCAGTCCCGCGTGGTGGACTTCGGCAAGGACTACTACGCCCTGCAGACCTTCT
TCAACACCGACCCGACCTACGGGAGCGCCCTGGGCATCGCGTGGGCCTCCAACT
GGGAGTACTCCGCCTTCGTGCCCACCAACCCCTGGCGCTCCTCCATGTCCCTCGT
GCGCAAGTTCTCCCTCAACACCGAGTACCAGGCCAACCCGGAGACGGAGCTGAT
CAACCTGAAGGCCGAGCCGATCCTGAACATCAGCAACGCCGGCCCCTGGAGCCG
GTTCGCCACCAACACCACGTTGACGAAGGCCAACAGCTACAACGTCGACCTGTC
CAACAGCACCGGCACCCTGGAGTTCGAGCTGGTGTACGCCGTCAACACCACCCA
GACGATCTCCAAGTCCGTGTTCGCGGACCTCTCCCTCTGGTTCAAGGGCCTGGAG
GACCCCGAGGAGTACCTCCGCATGGGCTTCGAGGTGTCCGCGTCCTCCTTCTTCC
TGGACCGCGGGAACAGCAAGGTGAAGTTCGTGAAGGAGAACCCCTACTTCACCA
ACCGCATGAGCGTGAACAACCAGCCCTTCAAGAGCGAGAACGACCTGTCCTACT
ACAAGGTGTACGGCTTGCTGGACCAGAACATCCTGGAGCTGTACTTCAACGACG
GCGACGTCGTGTCCACCAACACCTACTTCATGACCACCGGGAACGCCCTGGGCTC
CGTGAACATGACGACGGGGGTGGACAACCTGTTCTACATCGACAAGTTCCAGGT
GCGCGAGGTCAAGTGACAATTGGCAGCAGCAGCTCGGATAGTATCGACACACTC
TGGACGCTGGTCGTGTGATGGACTGTTGCCGCCACACTTGCTGCCTTGACCTGTG
AATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGTGTGTACGC

[1220]

GCTTTTGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCCAGCATCCC
CTTCCCTCGTTTCATATCGCTTGCATCCCAACCGCAACTTATCTACGCTGTCCTGC
TATCCCTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCTCGCACAGCCTTGGTTTG
GGCTCCGCCTGTATTCTCCTGGTACTGCAACCTGTAAACCAGCACTGCAATGCTG
ATGCACGGGAAGTAGTGGGATGGGAACACAAATGGAGCATCGAGGTGGTCATCT
CCGACCTCGCGTTGGTGGCGGTGCTCAGCGGGCTCAGCGTGCTGGGCCGCACCAT
GGGCTGGGCCTGGCTGGTCAAGACCTACGTGGTGCCCTACATGATCGTGAACATG
TGGCTGGTGCTCATCACGCTGCTCCAGCACACGCACCCGGCCCTGCCGCACTACT
TCGAGAAGGACTGGGACTGGCTACGCGGCGCCATGGCCACCGTCGACCGCTCCA
TGGGCCCGCCCTTCATGGACAGCATCCTGCACCACATCTCCGACACCCACGTGCT
GCACCACCTCTTCAGCACCATCCCGCACTACCACGCCGAGGAGGCCTCCGCCGCC
ATCCGGCCCATCCTGGGCAAGTACTACCAATCCGACAGCCGCTGGGTCGGCCGC
GCCCTGTGGGAGGACTGGCGCGACTGCCGCTACGTCGTCCCCGACGCGCCCGAG
GACGACTCCGCGCTCTGGTTCCACAAGTGAGCGCGCCTGCGCGAGGACGCAGAA
CAACGCTGCCGCCGTGTCTTTTGCACGCGCGACTCCGGCGCTTCGCTGGTGGCAC
CCCCATAAAGAAACCCTCAATTCTGTTTGTGGAAGACACGGTGTACCCCCACCCA
CCCACCTGCACCTCTATTATTGGTATTATTGACGCGGGAGTGGGCGTTGTACCCT
ACAACGTAGCTTCTCTAGTTTTCAGCTGGCTCCCACCATTGTAAAGAGCCTCTAG
AGTCGACCTGCAGGCATGCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGT
GTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAA
GTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGC
TCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCG
GCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCC

[1221]

[1222] 서열 번호 75

[1223] 프로토테카 모리포르미스 SAD2A 녹아웃 상동 재조합 표적화 구조물의 5' 공여체 DNA 서열

GCTCTTCCGCCTGGAGCTGGTGCAGAGCATGGGTCAGTTTGCGGAGGAGAGGGT
GCTCCCCGTGCTGCACCCCGTGGACAAGCTGTGGCAGCCGCAGGACTTCCTGCCC
GACCCCGAGTCGCCCGACTTCGAGGACCAGGTGGCGGAGCTGCGCGCGCGCGCC
AAGGACCTGCCCGACGAGTACTTTGTGGTGCTGGTGGGCGACATGATCACGGAG
GAGGCGCTGCCGACCTACATGGCCATGCTCAACACCTTGGACGGTGTGCGCGAC
GACACGGGCGCGGCTGACCACCCGTGGGCGCGCTGGACGCGGCAGTGGGTGGCC
GAGGAGAACCGGCACGGCGACCTGCTGAACAAGTACTGTTGGCTGACGGGGCGC
GTCAACATGCGGGCCGTGGAGGTGACCATCAACAACCTGATCAAGAGCGGCATG
AACCCGCAGACGGACAACAACCCTTACTTGGGCTTCGTCTACACCTCCTTCCAGG
AGCGCGCCACCAAGTAGGTACC

[1224]

[1225] 서열 번호 76

[1226] 프로토테카 모리포르미스 SAD2A 녹아웃 상동 재조합 표적화 구조물의 3' 공여체 DNA 서열

CAATTGGCAGCAGCAGCTCGGATAGTATCGACACACTCTGGACGCTGGTCGTGTG
ATGGACTGTTGCCGCCACACTTGCTGCCTTGACCTGTGAATATCCCTGCCGCTTTT
ATCAAACAGCCTCAGTGTGTTTGATCTTGTGTGTACGCGCTTTTGCGAGTTGCTAG
CTGCTTGTGCTATTTGCGAATACCACCCCCAGCATCCCCTTCCCTCGTTTCATATC
GCTTGCATCCCAACCGCAACTTATCTACGCTGTCCTGCTATCCCTCAGCGCTGCTC
CTGCTCCTGCTCACTGCCCCTCGCACAGCCTTGGTTTGGGCTCCGCCTGTATTCTC
CTGGTACTGCAACCTGTAAACCAGCACTGCAATGCTGATGCACGGGAAGTAGTG

[1227]

GGATGGGAACACAAATGGAAGGATCGTAGAGCTCCAGCCACGGCAACACCGCGC
GCCTGGCGGCCGAGCACGGCGACAAGGGCCTGAGCAAGATCTGCGGGCTGATCG
CCAGCGACGAGGGCCGGCACGAGATCGCCTACACGCGCATCGTGGACGAGTTCT
TCCGCCTCGACCCCGAGGGCGCCGTCGCCGCCTACGCCAACATGATGCGCAAGC
AGATCACCATGCCCGCGCACCTCATGGACGACATGGGCCACGGCGAGGCCAACC
CGGGCCGCAACCTCTTCGCCGACTTCTCCGCCGTCGCCGAGAAGATCGACGTCTA
CGACGCCGAGGACTACTGCCGCATCCTGGAGCACCTCAACGCGCGCTGGAAGGT
GGACGAGCGCCAGGTCAGCGGCCAGGCCGCCGCGGACCAGGAGTACGTTCTGGG
CCTGCCCCAGCGCTTCCGGAAACTCGCCGAGAAGACCGCCGCCAAGCGCAAGCG
CGTCGCGCGCAGGCCCGTCGCCTTCTCCTGGAGAGAAGAGCCTCTAGAGTCGACC
TGCAGGCATGCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATT
GTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAG
CCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCC
CGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGC
GCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCC

[1228]

[1229] 서열 번호 77

[1230] 프로토테카 모리포르미스 SAD2A 녹아웃 상동 재조합 표적화 구조물

GCTCTTCCGCCTGGAGCTGGTGCAGAGCATGGGTCAGTTTGCGGAGGAGAGGGT
GCTCCCCGTGCTGCACCCCGTGGACAAGCTGTGGCAGCCGCAGGACTTCCTGCCC
GACCCCGAGTCGCCCGACTTCGAGGACCAGGTGGCGGAGCTGCGCGCGCGCGCC
AAGGACCTGCCCGACGAGTACTTTGTGGTGCTGGTGGGCGACATGATCACGGAG
GAGGCGCTGCCGACCTACATGGCCATGCTCAACACCTTGGACGGTGTGCGCGAC
GACACGGGCGCGGCTGACCACCCGTGGGCGCGCTGGACGCGGCAGTGGGTGGCC
GAGGAGAACCGGCACGGCGACCTGCTGAACAAGTACTGTTGGCTGACGGGGCGC
GTCAACATGCGGGCCGTGGAGGTGACCATCAACAACCTGATCAAGAGCGGCATG
AACCCGCAGACGGACAACAACCCTTACTTGGGCTTCGTCTACACCTCCTTCCAGG
AGCGCGCCACCAAGTAGGTACCCTTTCTTGCGCTATGACACTTCCAGCAAAAGGT
AGGGCGGGCTGCGAGACGGCTTCCCGGCGCTGCATGCAACACCGATGATGCTTC
GACCCCCCGAAGCTCCTTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCAGGGCG
AGCGCTGTTTAAATAGCCAGGCCCCCGATTGCAAAGACATTATAGCGAGCTACC
AAAGCCATATTCAAACACCTAGATCACTACCACTTCTACACAGGCCACTCGAGCT
TGTGATCGCACTCCGCTAAGGGGGCGCCTCTTCCTCTTCGTTTCAGTCACAACCC
GCAAACGGCGCGCCATGCTGCTGCAGGCCTTCCTGTTCCTGCTGGCCGGCTTCGC
CGCCAAGATCAGCGCCTCCATGACGAACGAGACGTCCGACCGCCCCCTGGTGCA
CTTCACCCCCAACAAGGGCTGGATGAACGACCCCAACGGCCTGTGGTACGACGA
GAAGGACGCCAAGTGGCACCTGTACTTCCAGTACAACCCGAACGACACCGTCTG
GGGGACGCCCTTGTTCTGGGGCCACGCCACGTCCGACGACCTGACCAACTGGGA
GGACCAGCCCATCGCCATCGCCCCGAAGCGCAACGACTCCGGCGCCTTCTCCGGC
TCCATGGTGGTGGACTACAACAACACCTCCGGCTTCTTCAACGACACCATCGACC
CGCGCCAGCGCTGCGTGGCCATCTGGACCTACAACACCCCGGAGTCCGAGGAGC
AGTACATCTCCTACAGCCTGGACGGCGGCTACACCTTCACCGAGTACCAGAAGA
ACCCCGTGCTGGCCGCCAACTCCACCCAGTTCCGCGACCCGAAGGTCTTCTGGTA
CGAGCCCTCCCAGAAGTGGATCATGACCGCGGCCAAGTCCCAGGACTACAAGAT
CGAGATCTACTCCTCCGACGACCTGAAGTCCTGGAAGCTGGAGTCCGCGTTCGCC
AACGAGGGCTTCCTCGGCTACCAGTACGAGTGCCCCGGCCTGATCGAGGTCCCCA
CCGAGCAGGACCCCAGCAAGTCCTACTGGGTGATGTTCATCTCCATCAACCCCGG
CGCCCCGGCCGGCGGCTCCTTCAACCAGTACTTCGTCGGCAGCTTCAACGGCACC
CACTTCGAGGCCTTCGACAACCAGTCCCGCGTGGTGGACTTCGGCAAGGACTACT

[1231]

ACGCCCTGCAGACCTTCTTCAACACCGACCCGACCTACGGGAGCGCCCTGGGCAT
CGCGTGGGCCTCCAACTGGGAGTACTCCGCCTTCGTGCCCACCAACCCCTGGCGC
TCCTCCATGTCCCTCGTGCGCAAGTTCTCCCTCAACACCGAGTACCAGGCCAACC
CGGAGACGGAGCTGATCAACCTGAAGGCCGAGCCGATCCTGAACATCAGCAACG
CCGGCCCCTGGAGCCGGTTCGCCACCAACACCACGTTGACGAAGGCCAACAGCT
ACAACGTCGACCTGTCCAACAGCACCGGCACCCTGGAGTTCGAGCTGGTGTACG
CCGTCAACACCACCCAGACGATCTCCAAGTCCGTGTTCGCGGACCTCTCCCTCTG
GTTCAAGGGCCTGGAGGACCCCGAGGAGTACCTCCGCATGGGCTTCGAGGTGTC
CGCGTCCTCCTTCTTCCTGGACCGCGGGAACAGCAAGGTGAAGTTCGTGAAGGA
GAACCCCTACTTCACCAACCGCATGAGCGTGAACAACCAGCCCTTCAAGAGCGA
GAACGACCTGTCCTACTACAAGGTGTACGGCTTGCTGGACCAGAACATCCTGGA
GCTGTACTTCAACGACGGCGACGTCGTGTCCACCAACACCTACTTCATGACCACC
GGGAACGCCCTGGGCTCCGTGAACATGACGACGGGGGTGGACAACCTGTTCTAC
ATCGACAAGTTCCAGGTGCGCGAGGTCAAGTGACAATTGGCAGCAGCAGCTCGG
ATAGTATCGACACACTCTGGACGCTGGTCGTGTGATGGACTGTTGCCGCCACACT
TGCTGCCTTGACCTGTGAATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTT
TGATCTTGTGTGTACGCGCTTTTGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAAT
ACCACCCCCAGCATCCCCTTCCCTCGTTTCATATCGCTTGCATCCCAACCGCAACT
TATCTACGCTGTCCTGCTATCCCTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCT
CGCACAGCCTTGGTTTGGGCTCCGCCTGTATTCTCCTGGTACTGCAACCTGTAAA
CCAGCACTGCAATGCTGATGCACGGGAAGTAGTGGGATGGGAACACAAATGGAA
GGATCGTAGAGCTCCAGCCACGGCAACACCGCGCGCCTGGCGGCCGAGCACGGC
GACAAGGGCCTGAGCAAGATCTGCGGGCTGATCGCCAGCGACGAGGGCCGGCAC
GAGATCGCCTACACGCGCATCGTGGACGAGTTCTTCCGCCTCGACCCCGAGGGCG
CCGTCGCCGCCTACGCCAACATGATGCGCAAGCAGATCACCATGCCCGCGCACCT
CATGGACGACATGGGCCACGGCGAGGCCAACCCGGGCCGCAACCTCTTCGCCGA
CTTCTCCGCCGTCGCCGAGAAGATCGACGTCTACGACGCCGAGGACTACTGCCGC
ATCCTGGAGCACCTCAACGCGCGCTGGAAGGTGGACGAGCGCCAGGTCAGCGGC
CAGGCCGCCGCGGACCAGGAGTACGTTCTGGGCCTGCCCCAGCGCTTCCGGAAA
CTCGCCGAGAAGACCGCCGCCAAGCGCAAGCGCGTCGCGCGCAGGCCCGTCGCC
TTCTCCTGGAGAGAAGAGCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGGCGT
AATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACAC
AACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGC
TAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGT
CGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTA
TTGGGCGCTCTTCC

[1232]

[1233] 서열 번호 78

[1234] 프로토테카 모리포르미스 SAD2B 녹아웃 상동 재조합 표적화 구조물의 5' 공여체 DNA 서열

GCTCTTCCCGCCTGGAGCTGGTGCAGAGCATGGGGCAGTTTGCGGAGGAGAGGG
TGCTCCCCGTGCTGCACCCCGTGGACAAGCTGTGGCAGCCGCAGGACTTCCTGCC
CGACCCCGAGTCGCCCGACTTCGAGGACCAGGTGGCGGAGCTGCGCGCGCGCGC
CAAGGACCTGCCCGACGAGTACTTTGTGGTGCTGGTGGGCGACATGATCACGGA
GGAGGCGCTGCCGACCTACATGGCCATGCTCAACACCTTGGACGGTGTGCGCGA
CGACACGGGCGCGGCTGACCACCCGTGGGCGCGCTGGACGCGGCAGTGGGTGGC
CGAGGAGAACCGGCACGGCGACCTGCTGAACAAGTACTGTTGGCTGACGGGGCG
CGTCAACATGCGGGCCGTGGAGGTGACCATCAACAACCTGATCAAGAGCGGCAT

[1235]

GAACCCGCAGACGGACAACAACCCTTACTTGGGCTTCGTCTACACCTCCTTCCAG
GAGCGCGCCACCAAGTAGGTACC

[1236]

[1237] 서열 번호 79

[1238] 프로토테카 모리포르미스 SAD2B 녹아웃 상동 재조합 표적화 구조물의 3' 공여체 DNA 서열

CAGCCACGGCAACACCGCGCGCCTTGCGGCCGAGCACGGCGACAAGAACCTGAG
CAAGATCTGCGGGCTGATCGCCAGCGACGAGGGCCGGCACGAGATCGCCTACAC
GCGCATCGTGGACGAGTTCTTCCGCCTCGACCCCGAGGGCGCCGTCGCCGCCTAC
GCCAACATGATGCGCAAGCAGATCACCATGCCCGCGCACCTCATGGACGACATG
GGCCACGGCGAGGCCAACCCGGGCCGCAACCTCTTCGCCGACTTCTCCGCGGTCG
CCGAGAAGATCGACGTCTACGACGCCGAGGACTACTGCCGCATCCTGGAGCACC
TCAACGCGCGCTGGAAGGTGGACGAGCGCCAGGTCAGCGGCCAGGCCGCCGCGG
ACCAGGAGTACGTCCTGGGCCTGCCCCAGCGCTTCCGGAAACTCGCCGAGAAGA
CCGCCGCCAAGCGCAAGCGCGTCGCGCGCAGGCCCGTCGCCTTCTCCTGGAGAA
GAGCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGGCGTAATCATGGTCATAGC
TGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGG
AAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAAT
TGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCAT
TAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCC

[1239]

[1240] 서열 번호 80

[1241] 프로토테카 모리포르미스 SAD2B 녹아웃 상동 재조합 표적화 구조물

GCTCTTCCCGCCTGGAGCTGGTGCAGAGCATGGGGCAGTTTGCGGAGGAGAGGG
TGCTCCCCGTGCTGCACCCCGTGGACAAGCTGTGGCAGCCGCAGGACTTCCTGCC
CGACCCCGAGTCGCCCGACTTCGAGGACCAGGTGGCGGAGCTGCGCGCGCGCGC
CAAGGACCTGCCCGACGAGTACTTTGTGGTGCTGGTGGGCGACATGATCACGGA
GGAGGCGCTGCCGACCTACATGGCCATGCTCAACACCTTGGACGGTGTGCGCGA
CGACACGGGCGCGGCTGACCACCCGTGGGCGCGCTGGACGCGGCAGTGGGTGGC
CGAGGAGAACCGGCACGGCGACCTGCTGAACAAGTACTGTTGGCTGACGGGGCG
CGTCAACATGCGGGCCGTGGAGGTGACCATCAACAACCTGATCAAGAGCGGCAT
GAACCCGCAGACGGACAACAACCCTTACTTGGGCTTCGTCTACACCTCCTTCCAG
GAGCGCGCCACCAAGTAGGTACCCTTTCTTGCGCTATGACACTTCCAGCAAAAGG
TAGGGCGGGCTGCGAGACGGCTTCCCGGCGCTGCATGCAACACCGATGATGCTT
CGACCCCCCGAAGCTCCTTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCAGGGC
GAGCGCTGTTTAAATAGCCAGGCCCCCGATTGCAAAGACATTATAGCGAGCTAC
CAAAGCCATATTCAAACACCTAGATCACTACCACTTCTACACAGGCCACTCGAGC
TTGTGATCGCACTCCGCTAAGGGGGCGCCTCTTCCTCTTCGTTTCAGTCACAACCC
GCAAACGGCGCGCCATGCTGCTGCAGGCCTTCCTGTTCCTGCTGGCCGGCTTCGC
CGCCAAGATCAGCGCCTCCATGACGAACGAGACGTCCGACCGCCCCCTGGTGCA
CTTCACCCCCAACAAGGGCTGGATGAACGACCCCAACGGCCTGTGGTACGACGA
GAAGGACGCCAAGTGGCACCTGTACTTCCAGTACAACCCGAACGACACCGTCTG
GGGGACGCCCTTGTTCTGGGGCCACGCCACGTCCGACGACCTGACCAACTGGGA
GGACCAGCCCATCGCCATCGCCCCGAAGCGCAACGACTCCGGCGCCTTCTCCGGC
TCCATGGTGGTGGACTACAACAACACCTCCGGCTTCTTCAACGACACCATCGACC
CGCGCCAGCGCTGCGTGGCCATCTGGACCTACAACACCCCGGAGTCCGAGGAGC
AGTACATCTCCTACAGCCTGGACGGCGGCTACACCTTCACCGAGTACCAGAAGA
ACCCCGTGCTGGCCGCCAACTCCACCCAGTTCCGCGACCCGAAGGTCTTCTGGTA

[1242]

CGAGCCCTCCCAGAAGTGGATCATGACCGCGGCCAAGTCCCAGGACTACAAGAT
CGAGATCTACTCCTCCGACGACCTGAAGTCCTGGAAGCTGGAGTCCGCGTTCGCC
AACGAGGGCTTCCTCGGCTACCAGTACGAGTGCCCCGGCCTGATCGAGGTCCCCA
CCGAGCAGGACCCCAGCAAGTCCTACTGGGTGATGTTCATCTCCATCAACCCCGG
CGCCCCGGCCGGCGGCTCCTTCAACCAGTACTTCGTCGGCAGCTTCAACGGCACC
CACTTCGAGGCCTTCGACAACCAGTCCCGCGTGGTGGACTTCGGCAAGGACTACT
ACGCCCTGCAGACCTTCTTCAACACCGACCCGACCTACGGGAGCGCCCTGGGCAT
CGCGTGGGCCTCCAACTGGGAGTACTCCGCCTTCGTGCCCACCAACCCCTGGCGC
TCCTCCATGTCCCTCGTGCGCAAGTTCTCCCTCAACACCGAGTACCAGGCCAACC
CGGAGACGGAGCTGATCAACCTGAAGGCCGAGCCGATCCTGAACATCAGCAACG
CCGGCCCCTGGAGCCGGTTCGCCACCAACACCACGTTGACGAAGGCCAACAGCT
ACAACGTCGACCTGTCCAACAGCACCGGCACCCTGGAGTTCGAGCTGGTGTACG
CCGTCAACACCACCCAGACGATCTCCAAGTCCGTGTTCGCGGACCTCTCCCTCTG
GTTCAAGGGCCTGGAGGACCCCGAGGAGTACCTCCGCATGGGCTTCGAGGTGTC
CGCGTCCTCCTTCTTCCTGGACCGCGGGAACAGCAAGGTGAAGTTCGTGAAGGA
GAACCCCTACTTCACCAACCGCATGAGCGTGAACAACCAGCCCTTCAAGAGCGA
GAACGACCTGTCCTACTACAAGGTGTACGGCTTGCTGGACCAGAACATCCTGGA
GCTGTACTTCAACGACGGCGACGTCGTGTCCACCAACACCTACTTCATGACCACC
GGGAACGCCCTGGGCTCCGTGAACATGACGACGGGGGTGGACAACCTGTTCTAC
ATCGACAAGTTCCAGGTGCGCGAGGTCAAGTGACAATTGGCAGCAGCAGCTCGG
ATAGTATCGACACACTCTGGACGCTGGTCGTGTGATGGACTGTTGCCGCCACACT
TGCTGCCTTGACCTGTGAATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTT
TGATCTTGTGTGTACGCGCTTTTGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAAT
ACCACCCCCAGCATCCCCTTCCCTCGTTTCATATCGCTTGCATCCCAACCGCAACT
TATCTACGCTGTCCTGCTATCCCTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCT
CGCACAGCCTTGGTTTGGGCTCCGCCTGTATTCTCCTGGTACTGCAACCTGTAAA
CCAGCACTGCAATGCTGATGCACGGGAAGTAGTGGGATGGGAACACAAATGGAC
AGCCACGGCAACACCGCGCGCCTTGCGGCCGAGCACGGCGACAAGAACCTGAGC
AAGATCTGCGGGCTGATCGCCAGCGACGAGGGCCGGCACGAGATCGCCTACACG
CGCATCGTGGACGAGTTCTTCCGCCTCGACCCCGAGGGCGCCGTCGCCGCCTACG
CCAACATGATGCGCAAGCAGATCACCATGCCCGCGCACCTCATGGACGACATGG
GCCACGGCGAGGCCAACCCGGGCCGCAACCTCTTCGCCGACTTCTCCGCGGTCGC
CGAGAAGATCGACGTCTACGACGCCGAGGACTACTGCCGCATCCTGGAGCACCT
CAACGCGCGCTGGAAGGTGGACGAGCGCCAGGTCAGCGGCCAGGCCGCCGCGG
ACCAGGAGTACGTCCTGGGCCTGCCCCAGCGCTTCCGGAAACTCGCCGAGAAGA
CCGCCGCCAAGCGCAAGCGCGTCGCGCGCAGGCCCGTCGCCTTCTCCTGGAGAA
GAGCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGGCGTAATCATGGTCATAGC
TGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGG
AAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAAT
TGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCAT
TAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCC

[1243]

[1244] 서열 번호 81

[1245] 피. 모리포르미스 Aat02 프로모터/5'UTR::씨. 프로토테코이데스 스테아로일 ACP 불포화 효소 PTS::신나모뮴 캠포라 C14 티오에스테라제::씨. 불가리스 질산염 환원 효소 3'UTR

AGAGAGCGGAGGTGGGGTTGTGAGGTGGGGTTGCTGACCAGGAGCTCGCGTCGC
CGAGCGCGACTCGCACACGGTCCAGTTACCCCCCCCTCCGCCCAAACGCAAGCCT
CCCATCTTGATGCCTTTCCGGCCACCTATACTATTTCTTAGTTCGCTGTAACATCC
AGACCGTCCTGAATAATAACAATGCCCTGTGTCAAGTGCATTCCTAAAAAAATTC

[1246]

TGTCCCAACCAACAATCCCACCTGAAATACCACCAGCCCTGCCCAGTACACTCTT
CCAATACCATCTCCCTACCTCCACGCGCAAGCGACCCCCATGCGCGACCAGGCTC
GAAAGTGATTTATGACTTGAGACGAGCGAGTGGCGGCGCGGTCGACTGCCTTTTC
ATCACGTGCCGTACGTCGGCGACCGCTAGGGCTTTGCACGGCAACGCACGGCTTC
GCCAACCCGACCAGCCAGGACCTCGACTACTCTACCGCGAATTCGCCTCAAGAA
GTCGCCAAATGTGCCATACACCATTCCTTACAGCACTGTTCAAACTTGATGCCAA
TTTTGACATTCGGGTTGCTCGTTGGCTGCGCCCACATCGGCCGTGAGTGCAGCAG
GCGGGATCGGACACGGAGGACGCGGCGTCACGCCCCGAACGCAGCCCGTAACTC
TACATCAACACGACGTGTTGCGTAATCCCGCCCGGCTGCGCATCGTGCCAACCCA
TTCGCGATGGATGGTCGGAAAATGGTGTGCCAACTGCCCTGAGGGAGGCTCTCG
CGAAACGGGCACGTCCCTGAAACCGAAACTGTGGCCTTGTCGTCGGCCACGCAA
GCACGTGGACCCTAAACACCAAGAAAATCAGTAAACAAGGTTGACATCCTCTAC
GGGCGAATTGTTTGCCCAACCCTTCATCGCACACTGCCATTATAATGCATCTAGC
TCGGCGACAAGTTTAGAAAAGGCAGGCTGCATTGTTCCATTTCGCCGTGGCGGCG
TGGGTGCCCATTTTACGAGGTTTGGGCTCCCGGGCAGCGACCGAGCCAGGTCGA
GTCCCTCTCGCCCGTCGACAATGTTGCGAACCCCACAAGCGGCTAACAACAACTT
GATGGTACCTGTACACTGCCAATTCCTTCTTCCCCGGCCGAGGTTTACACGTGAT
GGCCATGGCTTCGCATTCAGGCCGACTTCCCATTCCGACTTTCCAGAGGGTCCGC
GGACGCTGGGGGTTGGCTGCCTGAGGCCCACCCTTTGTTCCCCGCGTCCCGACAA
ACACAATTGCGTTACATAAGGGGGAGCCGCCCCCGTTCAGAGTGCAGAAATCTTT
CACTATATTTTCCAGTCGTCAGCGAAATCAAGTACTAGTATGGCCACCGCATCCA
CTTTCTCGGCGTTCAATGCCCGCTGCGGCGACCTGCGTCGCTCGGCGGGCTCCGG
GCCCCGGCGCCCAGCGAGGCCCCTCCCCGTGCGCGGGCGCGCCCCCGACTGGTC
CATGCTGTTCGCCGTGATCACCACCATCTTCTCCGCCGCCGAGAAGCAGTGGACC
AACCTGGAGTGGAAGCCCAAGCCCAACCCCCCCCCAGCTGCTGGACGACCACTTC
GGCCCCCACGGCCTGGTGTTCCGCCGCACCTTCGCCATCCGCAGCTACGAGGTGG
GCCCCGACCGCTCCACCAGCATCGTGGCCGTGATGAACCACCTGCAGGAGGCCG
CCCTGAACCACGCCAAGTCCGTGGGCATCCTGGGCGACGGCTTCGGCACCACCCT
GGAGATGTCCAAGCGCGACCTGATCTGGGTGGTGAAGCGCACCCACGTGGCCGT
GGAGCGCTACCCCGCCTGGGGCGACACCGTGGAGGTGGAGTGCTGGGTGGGCGC
CTCCGGCAACAACGGCCGCCGCCACGACTTCCTGGTGCGCGACTGCAAGACCGG
CGAGATCCTGACCCGCTGCACCTCCCTGAGCGTGATGATGAACACCCGCACCCGC
CGCCTGAGCAAGATCCCCGAGGAGGTGCGCGGCGAGATCGGCCCCGCCTTCATC
GACAACGTGGCCGTGAAGGACGAGGAGATCAAGAAGCCCCAGAAGCTGAACGA
CTCCACCGCCGACTACATCCAGGGCGGCCTGACCCCCCGCTGGAACGACCTGGA
CATCAACCAGCACGTGAACAACATCAAGTACGTGGACTGGATCCTGGAGACCGT
GCCCGACAGCATCTTCGAGAGCCACCACATCTCCTCCTTCACCATCGAGTACCGC
CGCGAGTGCACCATGGACAGCGTGCTGCAGTCCCTGACCACCGTGAGCGGCGGC
TCCTCCGAGGCCGGCCTGGTGTGCGAGCACCTGCTGCAGCTGGAGGGCGGCAGC
GAGGTGCTGCGCGCCAAGACCGAGTGGCGCCCCAAGCTGACCGACTCCTTCCGC
GGCATCAGCGTGATCCCCGCCGAGTCCAGCGTGATGGACTACAAGGACCACGAC
GGCGACTACAAGGACCACGACATCGACTACAAGGACGACGACGACAAGTGACTC
GAGTTAATTAACTCGAGGCAGCAGCAGCTCGGATAGTATCGACACACTCTGGAC
GCTGGTCGTGTGATGGACTGTTGCCGCCACACTTGCTGCCTTGACCTGTGAATAT
CCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGTGTGTACGCGCTTT
TGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCCAGCATCCCCTTCC
CTCGTTTCATATCGCTTGCATCCCAACCGCAACTTATCTACGCTGTCCTGCTATCC
CTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCTCGCACAGCCTTGGTTTGGGCTC
CGCCTGTATTCTCCTGGTACTGCAACCTGTAAACCAGCACTGCAATGCTGATGCA
CGGGAAGTAGTGGGATGGGAACACAAATGGAAAGCTT

[1247]

[1248] 서열 번호 82

[1249] 5' 6S 영역 게놈 공여체 서열

GCTCTTCGCCGCCGCCACTCCTGCTCGAGCGCGCCCGCGCGTGCGCCGCCAGCGC
CTTGGCCTTTTCGCCGCGCTCGTGCGCGTCGCTGATGTCCATCACCAGGTCCATG
AGGTCTGCCTTGCGCCGGCTGAGCCACTGCTTCGTCCGGGCGGCCAAGAGGAGC
ATGAGGGAGGACTCCTGGTCCAGGGTCCTGACGTGGTCGCGGCTCTGGGAGCGG
GCCAGCATCATCTGGCTCTGCCGCACCGAGGCCGCCTCCAACTGGTCCTCCAGCA
GCCGCAGTCGCCGCCGACCCTGGCAGAGGAAGACAGGTGAGGGGGGTATGAATT
GTACAGAACAACCACGAGCCTTGTCTAGGCAGAATCCCTACCAGTCATGGCTTTA
CCTGGATGACGGCCTGCGAACAGCTGTCCAGCGACCCTCGCTGCCGCCGCTTCTC
CCGCACGCTTCTTTCCAGCACCGTGATGGCGCGAGCCAGCGCCGCACGCTGGCGC
TGCGCTTCGCCGATCTGAGGACAGTCGGGGAACTCTGATCAGTCTAAACCCCCTT
GCGCGTTAGTGTTGCCATCCTTTGCAGACCGGTGAGAGCCGACTTGTTGTGCGCC
ACCCCCCACACCACCTCCTCCCAGACCAATTCTGTCACCTTTTTGGCGAAGGCAT
CGGCCTCGGCCTGCAGAGAGGACAGCAGTGCCCAGCCGCTGGGGGTTGGCGGAT
GCACGCTCAGGTACC

[1250]

[1251] 서열 번호 83

[1252] 신나모뮴 캠포라 티오에스테라제에 대한 관련 발현 구조물

[1253] (βtub::neo::nitred::βtub::씨. 캠포라 TE::nitred)

CTTTCTTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTT
CCCGGCGCTGCATGCAACACCGATGATGCTTCGACCCCCCGAAGCTCCTTCGGGG
CTGCATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGTTTAAATAGCCAGGCC
CCCGATTGCAAAGACATTATAGCGAGCTACCAAAGCCATATTCAAACACCTAGA
TCACTACCACTTCTACACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGG
GCGCCTCTTCCTCTTCGTTTCAGTCACAACCCGCAAACTCTAGAATATCAATGATC
GAGCAGGACGGCCTCCACGCCGGCTCCCCCGCCGCCTGGGTGGAGCGCCTGTTC
GGCTACGACTGGGCCCAGCAGACCATCGGCTGCTCCGACGCCGCCGTGTTCCGCC
TGTCCGCCCAGGGCCGCCCCGTGCTGTTCGTGAAGACCGACCTGTCCGGCGCCCT
GAACGAGCTGCAGGACGAGGCCGCCCGCCTGTCCTGGCTGGCCACCACCGGCGT
GCCCTGCGCCGCCGTGCTGGACGTGGTGACCGAGGCCGGCCGCGACTGGCTGCT
GCTGGGCGAGGTGCCCGGCCAGGACCTGCTGTCCTCCCACCTGGCCCCCGCCGAG
AAGGTGTCCATCATGGCCGACGCCATGCGCCGCCTGCACACCCTGGACCCCGCCA
CCTGCCCCTTCGACCACCAGGCCAAGCACCGCATCGAGCGCGCCCGCACCCGCAT
GGAGGCCGGCCTGGTGGACCAGGACGACCTGGACGAGGAGCACCAGGGCCTGG
CCCCCGCCGAGCTGTTCGCCCGCCTGAAGGCCCGCATGCCCGACGGCGAGGACC
TGGTGGTGACCCACGGCGACGCCTGCCTGCCCAACATCATGGTGGAGAACGGCC
GCTTCTCCGGCTTCATCGACTGCGGCCGCCTGGGCGTGGCCGACCGCTACCAGGA
CATCGCCCTGGCCACCCGCGACATCGCCGAGGAGCTGGGCGGCGAGTGGGCCGA
CCGCTTCCTGGTGCTGTACGGCATCGCCGCCCCCGACTCCCAGCGCATCGCCTTC
TACCGCCTGCTGGACGAGTTCTTCTGACAATTGGCAGCAGCAGCTCGGATAGTAT
CGACACACTCTGGACGCTGGTCGTGTGATGGACTGTTGCCGCCACACTTGCTGCC
TTGACCTGTGAATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTT
GTGTGTACGCGCTTTTGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACCC
CCAGCATCCCCTTCCCTCGTTTCATATCGCTTGCATCCCAACCGCAACTTATCTAC
GCTGTCCTGCTATCCCTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCTCGCACAG
CCTTGGTTTGGGCTCCGCCTGTATTCTCCTGGTACTGCAACCTGTAAACCAGCACT

[1254]

GCAATGCTGATGCACGGGAAGTAGTGGGATGGGAACACAAATGGAGGATCCCGC
GTCTCGAACAGAGCGCGCAGAGGAACGCTGAAGGTCTCGCCTCTGTCGCACCTC
AGCGCGGCATACACCACAATAACCACCTGACGAATGCGCTTGGTTCTTCGTCCAT
TAGCGAAGCGTCCGGTTCACACACGTGCCACGTTGGCGAGGTGGCAGGTGACAA
TGATCGGTGGAGCTGATGGTCGAAACGTTCACAGCCTAGGGATATCGAATTCCTT
TCTTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTTCCC
GGCGCTGCATGCAACACCGATGATGCTTCGACCCCCCGAAGCTCCTTCGGGGCTG
CATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGTTTAAATAGCCAGGCCCCC
GATTGCAAAGACATTATAGCGAGCTACCAAAGCCATATTCAAACACCTAGATCA
CTACCACTTCTACACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGGGCG
CCTCTTCCTCTTCGTTTCAGTCACAACCCGCAAACACTAGTATGGCCACCGCATCC
ACTTTCTCGGCGTTCAATGCCCGCTGCGGCGACCTGCGTCGCTCGGCGGGCTCCG
GGCCCCGGCGCCCAGCGAGGCCCCTCCCCGTGCGCGGGCGCGCCCCCGACTGGT
CCATGCTGTTCGCCGTGATCACCACCATCTTCTCCGCCGCCGAGAAGCAGTGGAC
CAACCTGGAGTGGAAGCCCAAGCCCAACCCCCCCCAGCTGCTGGACGACCACTT
CGGCCCCCACGGCCTGGTGTTCCGCCGCACCTTCGCCATCCGCAGCTACGAGGTG
GGCCCCGACCGCTCCACCAGCATCGTGGCCGTGATGAACCACCTGCAGGAGGCC
GCCCTGAACCACGCCAAGTCCGTGGGCATCCTGGGCGACGGCTTCGGCACCACC
CTGGAGATGTCCAAGCGCGACCTGATCTGGGTGGTGAAGCGCACCCACGTGGCC
GTGGAGCGCTACCCCGCCTGGGGCGACACCGTGGAGGTGGAGTGCTGGGTGGGC
GCCTCCGGCAACAACGGCCGCCGCCACGACTTCCTGGTGCGCGACTGCAAGACC
GGCGAGATCCTGACCCGCTGCACCTCCCTGAGCGTGATGATGAACACCCGCACCC
GCCGCCTGAGCAAGATCCCCGAGGAGGTGCGCGGCGAGATCGGCCCCGCCTTCA
TCGACAACGTGGCCGTGAAGGACGAGGAGATCAAGAAGCCCCAGAAGCTGAAC
GACTCCACCGCCGACTACATCCAGGGCGGCCTGACCCCCCGCTGGAACGACCTG
GACATCAACCAGCACGTGAACAACATCAAGTACGTGGACTGGATCCTGGAGACC
GTGCCCGACAGCATCTTCGAGAGCCACCACATCTCCTCCTTCACCATCGAGTACC
GCCGCGAGTGCACCATGGACAGCGTGCTGCAGTCCCTGACCACCGTGAGCGGCG
GCTCCTCCGAGGCCGGCCTGGTGTGCGAGCACCTGCTGCAGCTGGAGGGCGGCA
GCGAGGTGCTGCGCGCCAAGACCGAGTGGCGCCCCAAGCTGACCGACTCCTTCC
GCGGCATCAGCGTGATCCCCGCCGAGTCCAGCGTGATGGACTACAAGGACCACG
ACGGCGACTACAAGGACCACGACATCGACTACAAGGACGACGACGACAAGTGA
CTCGAGGCAGCAGCAGCTCGGATAGTATCGACACACTCTGGACGCTGGTCGTGT
GATGGACTGTTGCCGCCACACTTGCTGCCTTGACCTGTGAATATCCCTGCCGCTTT
TATCAAACAGCCTCAGTGTGTTTGATCTTGTGTGTACGCGCTTTTGCGAGTTGCTA
GCTGCTTGTGCTATTTGCGAATACCACCCCCAGCATCCCCTTCCCTCGTTTCATAT
CGCTTGCATCCCAACCGCAACTTATCTACGCTGTCCTGCTATCCCTCAGCGCTGCT
CCTGCTCCTGCTCACTGCCCCTCGCACAGCCTTGGTTTGGGCTCCGCCTGTATTCT
CCTGGTACTGCAACCTGTAAACCAGCACTGCAATGCTGATGCACGGGAAGTAGT
GGGATGGGAACACAAATGGAAAGCTT

[1255]

[1256] 서열 번호 84

[1257] 3' 6S 영역 게놈 공여체 서열

[1258]
GAGCTCCTTGTTTTCCAGAAGGAGTTGCTCCTTGAGCCTTTCATTCTCAGCCTCGA
TAACCTCCAAAGCCGCTCTAATTGTGGAGGGGGTTCGAATTTAAAAGCTTGGAAT
GTTGGTTCGTGCGTCTGGAACAAGCCCAGACTTGTTGCTCACTGGGAAAAGGACC
ATCAGCTCCAAAAAACTTGCCGCTCAAACCGCGTACCTCTGCTTTCGCGCAATCT
GCCCTGTTGAAATCGCCACCACATTCATATTGTGACGCTTGAGCAGTCTGTAATT
GCCTCAGAATGTGGAATCATCTGCCCCCTGTGCGAGCCCATGCCAGGCATGTCGC

[1259]
GGGCGAGGACACCCGCCACTCGTACAGCAGACCATTATGCTACCTCACAATAGTT
CATAACAGTGACCATATTTCTCGAAGCTCCCCAACGAGCACCTCCATGCTCTGAG
TGGCCACCCCCCGGCCCTGGTGCTTGCGGAGGGCAGGTCAACCGGCATGGGGCT
ACCGAAATCCCCGACCGGATCCCACCACCCCCGCGATGGGAAGAATCTCTCCCC
GGGATGTGGGCCCACCACCAGCACAACCTGCTGGCCCAGGCGAGCGTCAAACCA
TACCACACAAATATCCTTGGCATCGGCCCTGAATTCCTTCTGCCGCTCTGCTACCC
GGTGCTTCTGTCCGAAGCAGGGGTTGCTAGGGATCGCTCCGAGTCCGCAAACCCT
TGTCGCGTGGCGGGGCTTGTTCGAGCTTGAAGAGC

[1260] 서열 번호 85

[1261] Amt03 정방향 프라이머

[1262] 5'-GGAGGAATTCGGCCGACAGGACGCGCGTCA-3'

[1263] 서열 번호 86

[1264] Amt03 역방향 프라이머

[1265] 5'-GGAGACTAGTGGCTGCGACCGGCCTGTG-3'

[1266] 서열 번호 87

[1267] Amt02 정방향 프라이머

[1268] 5'-GGAGGAATTCTCACCAGCGGACAAAGCACCG-3'

[1269] 서열 번호 88

[1270] Amt02 역방향 프라이머

[1271] 5'-GGAGACTAGTGGCTGCGACCGGCCTCTGG-3'

[1272] 서열 번호 89

[1273] Amt03 프로모터/UTR 서열

[1274]
GGCCGACAGGACGCGCGTCAAAGGTGCTGGTCGTGTATGCCCTGGCCGGCAGGT
CGTTGCTGCTGCTGGTTAGTGATTCCGCAACCCTGATTTTGGCGTCTTATTTTGGC
GTGGCAAACGCTGGCGCCCGCGAGCCGGGCCGGCGGCGATGCGGTGCCCCACGG
CTGCCGGAATCCAAGGGAGGCAAGAGCGCCCGGGTCAGTTGAAGGGCTTTACGC
GCAAGGTACAGCCGCTCCTGCAAGGCTGCGTGGTGGAATTGGACGTGCAGGTCC
TGCTGAAGTTCCTCCACCGCCTCACCAGCGGACAAAGCACCGGTGTATCAGGTCC
GTGTCATCCACTCTAAAGAGCTCGACTACGACCTACTGATGGCCCTAGATTCTTC
ATCAAAAACGCCTGAGACACTTGCCCAGGATTGAAACTCCCTGAAGGGACCACC
AGGGGCCCTGAGTTGTTCCTTCCCCCCGTGGCGAGCTGCCAGCCAGGCTGTACCT
GTGATCGAGGCTGGCGGGAAAATAGGCTTCGTGTGCTCAGGTCATGGGAGGTGC
AGGACAGCTCATGAAACGCCAACAATCGCACAATTCATGTCAAGCTAATCAGCT
ATTTCCTCTTCACGAGCTGTAATTGTCCCAAAATTCTGGTCTACCGGGGGTGATCC
TTCGTGTACGGGCCCTTCCCTCAACCCTAGGTATGCGCGCATGCGGTCGCCGCGC
AACTCGCGCGAGGGCCGAGGGTTTGGGACGGGCCGTCCCGAAATGCAGTTGCAC
CCGGATGCGTGGCACCTTTTTTGCGATAATTTATGCAATGGACTGCTCTGCAAAA
TTCTGGCTCTGTCGCCAACCCTAGGATCAGCGGCGTAGGATTTCGTAATCATTCG
TCCTGATGGGGAGCTACCGACTACCCTAATATCAGCCCGACTGCCTGACGCCAGC
GTCCACTTTTGTGCACACATTCCATTCGTGCCCAAGACATTTCATTGTGGTGCGAA

GCGTCCCCAGTTACGCTCACCTGTTTCCCGACCTCCTTACTGTTCTGTCGACAGAG
CGGGCCCACAGGCCGGTCGCAGCC

[1275]

[1276] 서열 번호 90

[1277] amt 프로모터/UTR의 관련 서열::씨. 캠포라 티오에스테라제 발현 구조물

GCTCTTCGGCCGCCGCCACTCCTGCTCGAGCGCGCCCGACTCGCGCTCCGCCTGC
GCCCGCGCGTGCGCCGCCAGCGCCTTGGCCTTTTCGCCGCGCTCGTGCGCGTCGC
TGATGTCCATCACCAGGTCCATGAGGTCTGCCTTGCGCCGGCTGAGCCACTGCTT
CGTCCGGGCGGCCAAGAGGAGCATGAGGGAGGACTCCTGGTCCAGGGTCCTGAC
GTGGTCGCGGCTCTGGGAGCGGGCCAGCATCATCTGGCTCTGCCGCACCGAGGC
CGCCTCCAACTGGTCCTCCAGCAGCCGCAGTCGCCGCCGACCCTGGCAGAGGAA
GACAGGTGAGGGGGGTATGAATTGTACAGAACAACCACGAGCCTTGTCTAGGCA
GAATCCCTACCAGTCATGGCTTTACCTGGATGACGGCCTGCGAACAGCTGTCCAG
CGACCCTCGCTGCCGCCGCTTCTCCCGCACGCTTCTTTCCAGCACCGTGATGGCG
CGAGCCAGCGCCGCACGCTGGCGCTGCGCTTCGCCGATCTGAGGACAGTCGGGG
AACTCTGATCAGTCTAAACCCCCTTGCGCGTTAGTGTTGCCATCCTTTGCAGACC
GGTGAGAGCCGACTTGTTGTGCGCCACCCCCCACACCACCTCCTCCCAGACCAAT
TCTGTCACCTTTTTGGCGAAGGCATCGGCCTCGGCCTGCAGAGAGGACAGCAGTG
CCCAGCCGCTGGGGGTTGGCGGATGCACGCTCAGGTACCCTTTCTTGCGCTATGA
CACTTCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTTCCCGGCGCTGCATGC
AACACCGATGATGCTTCGACCCCCCGAAGCTCCTTCGGGGCTGCATGGGCGCTCC
GATGCCGCTCCAGGGCGAGCGCTGTTTAAATAGCCAGGCCCCCGATTGCAAAGA
CATTATAGCGAGCTACCAAAGCCATATTCAAACACCTAGATCACTACCACTTCTA
CACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGGGCGCCTCTTCCTCTT
CGTTTCAGTCACAACCCGCAAACGGCGCGCCATGCTGCTGCAGGCCTTCCTGTTC
CTGCTGGCCGGCTTCGCCGCCAAGATCAGCGCCTCCATGACGAACGAGACGTCC
GACCGCCCCCTGGTGCACTTCACCCCCAACAAGGGCTGGATGAACGACCCCAAC
GGCCTGTGGTACGACGAGAAGGACGCCAAGTGGCACCTGTACTTCCAGTACAAC
CCGAACGACACCGTCTGGGGGACGCCCTTGTTCTGGGGCCACGCCACGTCCGAC
GACCTGACCAACTGGGAGGACCAGCCCATCGCCATCGCCCCGAAGCGCAACGAC
TCCGGCGCCTTCTCCGGCTCCATGGTGGTGGACTACAACAACACCTCCGGCTTCT
TCAACGACACCATCGACCCGCGCCAGCGCTGCGTGGCCATCTGGACCTACAACA
CCCCGGAGTCCGAGGAGCAGTACATCTCCTACAGCCTGGACGGCGGCTACACCTT
CACCGAGTACCAGAAGAACCCCGTGCTGGCCGCCAACTCCACCCAGTTCCGCGA
CCCGAAGGTCTTCTGGTACGAGCCCTCCCAGAAGTGGATCATGACCGCGGCCAA
GTCCCAGGACTACAAGATCGAGATCTACTCCTCCGACGACCTGAAGTCCTGGAA
GCTGGAGTCCGCGTTCGCCAACGAGGGCTTCCTCGGCTACCAGTACGAGTGCCCC
GGCCTGATCGAGGTCCCCACCGAGCAGGACCCCAGCAAGTCCTACTGGGTGATG
TTCATCTCCATCAACCCCGGCGCCCCGGCCGGCGGCTCCTTCAACCAGTACTTCG
TCGGCAGCTTCAACGGCACCCACTTCGAGGCCTTCGACAACCAGTCCCGCGTGGT
GGACTTCGGCAAGGACTACTACGCCCTGCAGACCTTCTTCAACACCGACCCGACC
TACGGGAGCGCCCTGGGCATCGCGTGGGCCTCCAACTGGGAGTACTCCGCCTTCG
TGCCCACCAACCCCTGGCGCTCCTCCATGTCCCTCGTGCGCAAGTTCTCCCTCAAC
ACCGAGTACCAGGCCAACCCGGAGACGGAGCTGATCAACCTGAAGGCCGAGCCG
ATCCTGAACATCAGCAACGCCGGCCCCTGGAGCCGGTTCGCCACCAACACCACG
TTGACGAAGGCCAACAGCTACAACGTCGACCTGTCCAACAGCACCGGCACCCTG
GAGTTCGAGCTGGTGTACGCCGTCAACACCACCCAGACGATCTCCAAGTCCGTGT
TCGCGGACCTCTCCCTCTGGTTCAAGGGCCTGGAGGACCCCGAGGAGTACCTCCG
CATGGGCTTCGAGGTGTCCGCGTCCTCCTTCTTCCTGGACCGCGGGAACAGCAAG

[1278]

GTGAAGTTCGTGAAGGAGAACCCCTACTTCACCAACCGCATGAGCGTGAACAAC
CAGCCCTTCAAGAGCGAGAACGACCTGTCCTACTACAAGGTGTACGGCTTGCTGG
ACCAGAACATCCTGGAGCTGTACTTCAACGACGGCGACGTCGTGTCCACCAACA
CCTACTTCATGACCACCGGGAACGCCCTGGGCTCCGTGAACATGACGACGGGGG
TGGACAACCTGTTCTACATCGACAAGTTCCAGGTGCGCGAGGTCAAGTGACAATT
GGCAGCAGCAGCTCGGATAGTATCGACACACTCTGGACGCTGGTCGTGTGATGG
ACTGTTGCCGCCACACTTGCTGCCTTGACCTGTGAATATCCCTGCCGCTTTTATCA
AACAGCCTCAGTGTGTTTGATCTTGTGTGTACGCGCTTTTGCGAGTTGCTAGCTGC
TTGTGCTATTTGCGAATACCACCCCCAGCATCCCCTTCCCTCGTTTCATATCGCTT
GCATCCCAACCGCAACTTATCTACGCTGTCCTGCTATCCCTCAGCGCTGCTCCTGC
TCCTGCTCACTGCCCCTCGCACAGCCTTGGTTTGGGCTCCGCCTGTATTCTCCTGG
TACTGCAACCTGTAAACCAGCACTGCAATGCTGATGCACGGGAAGTAGTGGGAT
GGGAACACAAATGGAGGATCCCGCGTCTCGAACAGAGCGCGCAGAGGAACGCT
GAAGGTCTCGCCTCTGTCGCACCTCAGCGCGGCATACACCACAATAACCACCTGA
CGAATGCGCTTGGTTCTTCGTCCATTAGCGAAGCGTCCGGTTCACACACGTGCCA
CGTTGGCGAGGTGGCAGGTGACAATGATCGGTGGAGCTGATGGTCGAAACGTTC
ACAGCCTAGGGATATCGAATTCGGCCGACAGGACGCGCGTCAAAGGTGCTGGTC
GTGTATGCCCTGGCCGGCAGGTCGTTGCTGCTGCTGGTTAGTGATTCCGCAACCC
TGATTTTGGCGTCTTATTTTGGCGTGGCAAACGCTGGCGCCCGCGAGCCGGGCCG
GCGGCGATGCGGTGCCCCACGGCTGCCGGAATCCAAGGGAGGCAAGAGCGCCCG
GGTCAGTTGAAGGGCTTTACGCGCAAGGTACAGCCGCTCCTGCAAGGCTGCGTG
GTGGAATTGGACGTGCAGGTCCTGCTGAAGTTCCTCCACCGCCTCACCAGCGGAC
AAAGCACCGGTGTATCAGGTCCGTGTCATCCACTCTAAAGAGCTCGACTACGACC
TACTGATGGCCCTAGATTCTTCATCAAAAACGCCTGAGACACTTGCCCAGGATTG
AAACTCCCTGAAGGGACCACCAGGGGCCCTGAGTTGTTCCTTCCCCCCGTGGCGA
GCTGCCAGCCAGGCTGTACCTGTGATCGAGGCTGGCGGGAAAATAGGCTTCGTG
TGCTCAGGTCATGGGAGGTGCAGGACAGCTCATGAAACGCCAACAATCGCACAA
TTCATGTCAAGCTAATCAGCTATTTCCTCTTCACGAGCTGTAATTGTCCCAAAATT
CTGGTCTACCGGGGGTGATCCTTCGTGTACGGGCCCTTCCCTCAACCCTAGGTAT
GCGCGCATGCGGTCGCCGCGCAACTCGCGCGAGGGCCGAGGGTTTGGGACGGGC
CGTCCCGAAATGCAGTTGCACCCGGATGCGTGGCACCTTTTTTGCGATAATTTAT
GCAATGGACTGCTCTGCAAAATTCTGGCTCTGTCGCCAACCCTAGGATCAGCGGC
GTAGGATTTCGTAATCATTCGTCCTGATGGGGAGCTACCGACTACCCTAATATCA
GCCCGACTGCCTGACGCCAGCGTCCACTTTTGTGCACACATTCCATTCGTGCCCA
AGACATTTCATTGTGGTGCGAAGCGTCCCCAGTTACGCTCACCTGTTTCCCGACC
TCCTTACTGTTCTGTCGACAGAGCGGGCCCACAGGCCGGTCGCAGCCACTAGTAT
GGCCACCGCATCCACTTTCTCGGCGTTCAATGCCCGCTGCGGCGACCTGCGTCGC
TCGGCGGGCTCCGGGCCCCGGCGCCCAGCGAGGCCCCTCCCCGTGCGCGGGCGC
GCCCCCGACTGGTCCATGCTGTTCGCCGTGATCACCACCATCTTCTCCGCCGCCG
AGAAGCAGTGGACCAACCTGGAGTGGAAGCCCAAGCCCAACCCCCCCCAGCTGC
TGGACGACCACTTCGGCCCCCACGGCCTGGTGTTCCGCCGCACCTTCGCCATCCG
CAGCTACGAGGTGGGCCCCGACCGCTCCACCAGCATCGTGGCCGTGATGAACCA
CCTGCAGGAGGCCGCCCTGAACCACGCCAAGTCCGTGGGCATCCTGGGCGACGG
CTTCGGCACCACCCTGGAGATGTCCAAGCGCGACCTGATCTGGGTGGTGAAGCG
CACCCACGTGGCCGTGGAGCGCTACCCCGCCTGGGGCGACACCGTGGAGGTGGA
GTGCTGGGTGGGCGCCTCCGGCAACAACGGCCGCCGCCACGACTTCCTGGTGCG
CGACTGCAAGACCGGCGAGATCCTGACCCGCTGCACCTCCCTGAGCGTGATGAT
GAACACCCGCACCCGCCGCCTGAGCAAGATCCCCGAGGAGGTGCGCGGCGAGAT
CGGCCCCGCCTTCATCGACAACGTGGCCGTGAAGGACGAGGAGATCAAGAAGCC
CCAGAAGCTGAACGACTCCACCGCCGACTACATCCAGGGCGGCCTGACCCCCCG

[1279]

CTGGAACGACCTGGACATCAACCAGCACGTGAACAACATCAAGTACGTGGACTG
GATCCTGGAGACCGTGCCCGACAGCATCTTCGAGAGCCACCACATCTCCTCCTTC
ACCATCGAGTACCGCCGCGAGTGCACCATGGACAGCGTGCTGCAGTCCCTGACC
ACCGTGAGCGGCGGCTCCTCCGAGGCCGGCCTGGTGTGCGAGCACCTGCTGCAG
CTGGAGGGCGGCAGCGAGGTGCTGCGCGCCAAGACCGAGTGGCGCCCCAAGCTG
ACCGACTCCTTCCGCGGCATCAGCGTGATCCCCGCCGAGTCCAGCGTGATGGACT
ACAAGGACCACGACGGCGACTACAAGGACCACGACATCGACTACAAGGACGAC
GACGACAAGTGACTCGAGTTAATTAACTCGAGGCAGCAGCAGCTCGGATAGTAT
CGACACACTCTGGACGCTGGTCGTGTGATGGACTGTTGCCGCCACACTTGCTGCC
TTGACCTGTGAATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTT
GTGTGTACGCGCTTTTGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACCC
CCAGCATCCCCTTCCCTCGTTTCATATCGCTTGCATCCCAACCGCAACTTATCTAC
GCTGTCCTGCTATCCCTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCTCGCACAG
CCTTGGTTTGGGCTCCGCCTGTATTCTCCTGGTACTGCAACCTGTAAACCAGCACT
GCAATGCTGATGCACGGGAAGTAGTGGGATGGGAACACAAATGGAAAGCTGTAG
AGCTCCTTGTTTTCCAGAAGGAGTTGCTCCTTGAGCCTTTCATTCTCAGCCTCGAT
AACCTCCAAAGCCGCTCTAATTGTGGAGGGGGTTCGAATTTAAAAGCTTGGAATG
TTGGTTCGTGCGTCTGGAACAAGCCCAGACTTGTTGCTCACTGGGAAAAGGACCA
TCAGCTCCAAAAAACTTGCCGCTCAAACCGCGTACCTCTGCTTTCGCGCAATCTG
CCCTGTTGAAATCGCCACCACATTCATATTGTGACGCTTGAGCAGTCTGTAATTG
CCTCAGAATGTGGAATCATCTGCCCCCTGTGCGAGCCCATGCCAGGCATGTCGCG
GGCGAGGACACCCGCCACTCGTACAGCAGACCATTATGCTACCTCACAATAGTTC
ATAACAGTGACCATATTTCTCGAAGCTCCCCAACGAGCACCTCCATGCTCTGAGT
GGCCACCCCCCGGCCCTGGTGCTTGCGGAGGGCAGGTCAACCGGCATGGGGCTA
CCGAAATCCCCGACCGGATCCCACCACCCCCGCGATGGGAAGAATCTCTCCCCG
GGATGTGGGCCCACCACCAGCACAACCTGCTGGCCCAGGCGAGCGTCAAACCAT
ACCACACAAATATCCTTGGCATCGGCCCTGAATTCCTTCTGCCGCTCTGCTACCC
GGTGCTTCTGTCCGAAGCAGGGGTTGCTAGGGATCGCTCCGAGTCCGCAAACCCT
TGTCGCGTGGCGGGGCTTGTTCGAGCTTGTTCGAGCTTGAAGAGCCTCTAGAGTC
GACCTGCAGGCATGCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGA
AATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTA
AAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACT
GCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAA
CGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCC

[1280]

[1281] 서열 번호 91

[1282] 코돈-최적화 씨. 캠포라 티오에스테라제 서열

ATGGCCACCGCATCCACTTTCTCGGCGTTCAATGCCCGCTGCGGCGACCTGCGTC
GCTCGGCGGGCTCCGGGCCCCGGCGCCCAGCGAGGCCCCTCCCCGTGCGCGGGC
GCGCCCCCGACTGGTCCATGCTGTTCGCCGTGATCACCACCATCTTCTCCGCCGC
CGAGAAGCAGTGGACCAACCTGGAGTGGAAGCCCAAGCCCAACCCCCCCCAGCT
GCTGGACGACCACTTCGGCCCCCACGGCCTGGTGTTCCGCCGCACCTTCGCCATC
CGCAGCTACGAGGTGGGCCCCGACCGCTCCACCAGCATCGTGGCCGTGATGAAC
CACCTGCAGGAGGCCGCCCTGAACCACGCCAAGTCCGTGGGCATCCTGGGCGAC
GGCTTCGGCACCACCCTGGAGATGTCCAAGCGCGACCTGATCTGGGTGGTGAAG
CGCACCCACGTGGCCGTGGAGCGCTACCCCGCCTGGGGCGACACCGTGGAGGTG
GAGTGCTGGGTGGGCGCCTCCGGCAACAACGGCCGCCGCCACGACTTCCTGGTG
CGCGACTGCAAGACCGGCGAGATCCTGACCCGCTGCACCTCCCTGAGCGTGATG
ATGAACACCCGCACCCGCCGCCTGAGCAAGATCCCCGAGGAGGTGCGCGGCGAG

[1283]

ATCGGCCCCGCCTTCATCGACAACGTGGCCGTGAAGGACGAGGAGATCAAGAAG
CCCCAGAAGCTGAACGACTCCACCGCCGACTACATCCAGGGCGGCCTGACCCCC
CGCTGGAACGACCTGGACATCAACCAGCACGTGAACAACATCAAGTACGTGGAC
TGGATCCTGGAGACCGTGCCCGACAGCATCTTCGAGAGCCACCACATCTCCTCCT
TCACCATCGAGTACCGCCGCGAGTGCACCATGGACAGCGTGCTGCAGTCCCTGAC
CACCGTGAGCGGCGGCTCCTCCGAGGCCGGCCTGGTGTGCGAGCACCTGCTGCA
GCTGGAGGGCGGCAGCGAGGTGCTGCGCGCCAAGACCGAGTGGCGCCCCAAGCT
GACCGACTCCTTCCGCGGCATCAGCGTGATCCCCGCCGAGTCCAGCGTGATGGAC
TACAAGGACCACGACGGCGACTACAAGGACCACGACATCGACTACAAGGACGAC
GACGACAAGTGA

[1284]

[1285] 서열 번호 92

[1286] 코돈-최적화 유. 캘리포니카 티오에스테라제 서열

GGCGCGCCCCCGACTGGTCCATGCTGTTCGCCGTGATCACCACCATCTTCAGCGC
CGCCGAGAAGCAGTGGACCAACCTGGAGTGGAAGCCCAAGCCCAAGCTGCCCCA
GCTGCTGGACGACCACTTCGGCCTGCACGGCCTGGTGTTCCGCCGCACCTTCGCC
ATCCGCTCCTACGAGGTGGGCCCCGACCGCAGCACCTCCATCCTGGCCGTGATGA
ACCACATGCAGGAGGCCACCCTGAACCACGCCAAGAGCGTGGGCATCCTGGGCG
ACGGCTTCGGCACCACCCTGGAGATGTCCAAGCGCGACCTGATGTGGGTGGTGC
GCCGCACCCACGTGGCCGTGGAGCGCTACCCCACCTGGGGCGACACCGTGGAGG
TGGAGTGCTGGATCGGCGCCAGCGGCAACAACGGCATGCGCCGCGACTTCCTGG
TGCGCGACTGCAAGACCGGCGAGATCCTGACCCGCTGCACCTCCCTGAGCGTGCT
GATGAACACCCGCACCCGCCGCCTGAGCACCATCCCCGACGAGGTGCGCGGCGA
GATCGGCCCCGCCTTCATCGACAACGTGGCCGTGAAGGACGACGAGATCAAGAA
GCTGCAGAAGCTGAACGACTCCACCGCCGACTACATCCAGGGCGGCCTGACCCC
CCGCTGGAACGACCTGGACGTGAACCAGCACGTGAACAACCTGAAGTACGTGGC
CTGGGTGTTCGAGACCGTGCCCGACAGCATCTTCGAGTCCCACCACATCAGCTCC
TTCACCCTGGAGTACCGCCGCGAGTGCACCCGCGACTCCGTGCTGCGCAGCCTGA
CCACCGTGAGCGGCGGCAGCTCCGAGGCCGGCCTGGTGTGCGACCACCTGCTGC
AGCTGGAGGGCGGCAGCGAGGTGCTGCGCGCCCGCACCGAGTGGCGCCCCAAGC
TGACCGACTCCTTCCGCGGCATCAGCGTGATCCCCGCCGAGCCCCGCGTGATGGA
CTACAAGGACCACGACGGCGACTACAAGGACCACGACATCGACTACAAGGACGA
CGACGACAAGTGACTCGAG

[1287]

[1288] 서열 번호 93

[1289] 코돈-최적화 유. 아메리카나 티오에스테라제 서열

GGCGCGCCCAGCTGCCCGACTGGAGCATGCTGCTGGCCGCGATCACCACCCTGTT
CCTGGCGGCCGAGAAGCAGTGGATGATGCTGGACTGGAAGCCCAAGCGCCCCGA
CATGCTGGTGGACCCCTTCGGCCTGGGCCGCTTCGTGCAGGACGGCCTGGTGTTC
CGCAACAACTTCAGCATCCGCAGCTACGAGATCGGCGCGGACCGCACCGCCAGC
ATCGAGACCCTGATGAACCACCTGCAGGAGACCGCCCTGAACCACGTGAAGAGC
GTGGGCCTGCTGGAGGACGGCCTGGGCAGCACCCGCGAGATGAGCCTGCGCAAC
CTGATCTGGGTGGTGACCAAGATGCAGGTGGCGGTGGACCGCTACCCCACCTGG
GGCGACGAGGTGCAGGTGAGCAGCTGGGCGACCGCCATCGGCAAGAACGGCAT
GCGCCGCGAGTGGATCGTGACCGACTTCCGCACCGGCGAGACCCTGCTGCGCGC
CACCAGCGTGTGGGTGATGATGAACAAGCTGACCCGCCGCATCAGCAAGATCCC
CGAGGAGGTGTGGCACGAGATCGGCCCCAGCTTCATCGACGCGCCCCCCCTGCC
CACCGTGGAGGACGACGGCCGCAAGCTGACCCGCTTCGACGAGAGCAGCGCCGA

[1290]

CTTCATCCGCAAGGGCCTGACCCCCCGCTGGAGCGACCTGGACATCAACCAGCA
CGTGAACAACGTGAAGTACATCGGCTGGCTGCTGGAGAGCGCGCCCCCCGAGAT
CCACGAGAGCCACGAGATCGCCAGCCTGACCCTGGAGTACCGCCGCGAGTGCGG
CCGCGACAGCGTGCTGAACAGCGCCACCAAGGTGAGCGACAGCAGCCAGCTGGG
CAAGAGCGCCGTGGAGTGCAACCACCTGGTGCGCCTGCAGAACGGCGGCGAGAT
CGTGAAGGGCCGCACCGTGTGGCGCCCCAAGCGCCCCCTGTACAACGACGGCGC
CGTGGTGGACGTGCCCCGCCAAGACCAGCGATGACGATGACAAGCTGGGATGACT
CGAG

[1291]

[1292] 서열 번호 94

[1293] 코돈-최적화 씨. 후커리아나 티오에스테라제 서열

ACTAGTATGGCTATCAAGACGAACAGGCAGCCTGTGGAGAAGCCTCCGTTCACG
ATCGGGACGCTGCGCAAGGCCATCCCCGCGCACTGTTTCGAGCGCTCGGCGCTTC
GTGGGCGCGCCCAGCTGCCCGACTGGAGCCGCCTGCTGACCGCCATCACCACCGT
GTTCGTGAAGTCCAAGCGCCCCGACATGCACGACCGCAAGTCCAAGCGCCCCGA
CATGCTGGTGGACAGCTTCGGCCTGGAGTCCACCGTGCAGGACGGCCTGGTGTTC
CGCCAGTCCTTCTCCATCCGCTCCTACGAGATCGGCACCGACCGCACCGCCAGCA
TCGAGACCCTGATGAACCACCTGCAGGAGACCTCCCTGAACCACTGCAAGAGCA
CCGGCATCCTGCTGGACGGCTTCGGCCGCACCCTGGAGATGTGCAAGCGCGACCT
GATCTGGGTGGTGATTAAGATGCAGATCAAGGTGAACCGCTACCCCGCCTGGGG
CGACACCGTGGAGATCAACACCCGCTTCAGCCGCCTGGGCAAGATCGGCATGGG
CCGCGACTGGCTGATCTCCGACTGCAACACCGGCGAGATCCTGGTGCGCGCCACC
AGCGCCTACGCCATGATGAACCAGAAGACCCGCCGCCTGTCCAAGCTGCCCTAC
GAGGTGCACCAGGAGATCGTGCCCCTGTTCGTGGACAGCCCCGTGATCGAGGAC
TCCGACCTGAAGGTGCACAAGTTCAAGGTGAAGACCGGCGACAGCATCCAGAAG
GGCCTGACCCCCGGCTGGAACGACCTGGACGTGAACCAGCACGTGTCCAACGTG
AAGTACATCGGCTGGATCCTGGAGAGCATGCCCACCGAGGTGCTGGAGACCCAG
GAGCTGTGCTCCCTGGCCCTGGAGTACCGCCGCGAGTGCGGCCGCGACTCCGTGC
TGGAGAGCGTGACCGCCATGGACCCCAGCAAGGTGGGCGTGCGCTCCCAGTACC
AGCACCTGCTGCGCCTGGAGGACGGCACCGCCATCGTGAACGGCGCCACCGAGT
GGCGCCCCAAGAACGCCGGCGCCAACGGCGCCATCTCCACCGGCAAGACCAGCA
ACGGCAACTCCGTGTCCATGGACTACAAGGACCACGACGGCGACTACAAGGACC
ACGACATCGACTACAAGGACGACGACGACAAGTGACTCGAG

[1294]

[1295] 서열 번호 95

[1296] 유. 캘리포니카 티오에스테라제 정방향 프라이머

[1297] CTGGGCGACGGCTTCGGCAC

[1298] 서열 번호 96

[1299] 유. 캘리포니카 티오에스테라제 역방향 프라이머

[1300] AAGTCGCGGCGCATGCCGTT

[1301] 서열 번호 97

[1302] 유. 아메리카나 티오에스테라제 정방향 프라이머

[1303] CCCAGCTGCTCACCTGCACC

[1304] 서열 번호 98

[1305] 유. 아메리카나 티오에스테라제 역방향 프라이머

[1306] CACCCAAGGCCAACGGCAGCGCCGTG

[1307] 서열 번호 99

[1308] 씨. 후커리아나 티오에스테라제 정방향 프라이머

[1309] TACCCCGCCTGGGGCGACAC

[1310] 서열 번호 100

[1311] 씨. 후커리아나 티오에스테라제 역방향 프라이머

[1312] AGCTTGGACAGGCGGCGGGT

[1313] 서열 번호 101

[1314] FAD 프라이머 1

[1315] 5'-TCACTTCATGCCGGCGGTCC-3'

[1316] 서열 번호 102

[1317] FAD 프라이머 2

[1318] 5'- GCGCTCCTGCTTGGCTCGAA-3'

[1319] 서열 번호 103

[1320] pSZ1124 (FAD2B) 5' 게놈 표적화 서열

[1321]
GCTCTTCGAGACGTGGTCTGAATCCTCCAGGCGGGTTTCCCCGAGAAAGAAAGG
GTGCCGATTTCAAAGCAGAGCCATGTGCCGGGCCCTGTGGCCTGTGTTGGCGCCT
ATGTAGTCACCCCCCCTCACCCAATTGTCGCCAGTTTGCGCAATCCATAAACTCA
AAACTGCAGCTTCTGAGCTGCGCTGTTCAAGAACACCTCTGGGGTTTGCTCACCC
GCGAGGTCGACGCCCAGCATGGCTATCAAGACGAACAGGCAGCCTGTGGAGAAG
CCTCCGTTCACGATCGGGACGCTGCGCAAGGCCATCCCCGCGCACTGTTTCGAGC
GCTCGGCGCTTCGTAGCAGCATGTACCTGGCCTTTGACATCGCGGTCATGTCCCT
GCTCTACGTCGCGTCGACGTACATCGACCCTGCGCCGGTGCCTACGTGGGTCAAG
TATGGCGTCATGTGGCCGCTCTACTGGTTCTTCCAGGTGTGTGTGAGGGTTGTGG
TTGCCCGTATCGAGGTCCTGGTGGCGCGCATGGGGGAGAAGGCGCCTGTCCCGCT
GACCCCCCCGGCTACCCTCCCGGCACCTTCCAGGGCGCCTTCGGCACGGGTGTCT
GGGTGTGCGCGCACGAGTGCGGCCACCAGGCCTTTTCCTCCAGCCAGGCCATCAA
CGACGGCGTGGGCCTGGTGTTCCACAGCCTGCTGCTGGTGCCCTACTACTCCTGG
AAGCACTCGCACCGGGTACC

[1322] 서열 번호 104

[1323] pSZ1124 (FAD2B) 3' 게놈 표적화 서열

[1324]
CCGCCACCACTCCAACACGGGGTGCCTGGACAAGGACGAGGTGTTTGTGCCGCC
GCACCGCGCAGTGGCGCACGAGGGCCTGGAGTGGGAGGAGTGGCTGCCCATCCG
CATGGGCAAGGTGCTGGTCACCCTGACCCTGGGCTGGCCGCTGTACCTCATGTTC

[1325]
AACGTCGCCTCGCGGCCGTACCCGCGCTTCGCCAACCACTTTGACCCGTGGTCGC
CCATCTTCAGCAAGCGCGAGCGCATCGAGGTGGTCATCTCCGACCTGGCGCTGGT
GGCGGTGCTCAGCGGGCTCAGCGTGCTGGGCCGCACCATGGGCTGGGCCTGGCT
GGTCAAGACCTACGTGGTGCCCTACCTGATCGTGAACATGTGGCTCGTGCTCATC
ACGCTGCTCCAGCACACGCACCCGGCGCTGCCGCACTACTTCGAGAAGGACTGG
GACTGGCTGCGCGGCGCCATGGCCACCGTGGACCGCTCCATGGGCCCGCCCTTCA
TGGACAACATCCTGCACCACATCTCCGACACCCACGTGCTGCACCACCTCTTCAG
CACCATCCCGCACTACCACGCCGAGGAGGCCTCCGCCGCCATCAGGCCCATCCTG
GGCAAGTACTACCAGTCCGACAGCCGCTGGGTCGGCCGCGCCCTGTGGGAGGAC
TGGCGCGACTGCCGCTACGTCGTCCCGGACGCGCCCGAGGACGACTCCGCGCTCT
GGTTCCACAAGTGAGTGAGTGAGAAGAGC

[1326] 서열 번호 105

[1327] 씨. 세레비지아에 suc2 카세트

CTTTCTTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTT
CCCGGCGCTGCATGCAACACCGATGATGCTTCGACCCCCCGAAGCTCCTTCGGGG
CTGCATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGTTTAAATAGCCAGGCC
CCCGATTGCAAAGACATTATAGCGAGCTACCAAAGCCATATTCAAACACCTAGA
TCACTACCACTTCTACACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGG
GCGCCTCTTCCTCTTCGTTTCAGTCACAACCCGCAAACGGCGCGCCATGCTGCTG
CAGGCCTTCCTGTTCCTGCTGGCCGGCTTCGCCGCCAAGATCAGCGCCTCCATGA
CGAACGAGACGTCCGACCGCCCCCTGGTGCACTTCACCCCCAACAAGGGCTGGA
TGAACGACCCCAACGGCCTGTGGTACGACGAGAAGGACGCCAAGTGGCACCTGT
ACTTCCAGTACAACCCGAACGACACCGTCTGGGGGACGCCCTTGTTCTGGGGCCA
CGCCACGTCCGACGACCTGACCAACTGGGAGGACCAGCCCATCGCCATCGCCCC
GAAGCGCAACGACTCCGGCGCCTTCTCCGGCTCCATGGTGGTGGACTACAACAA
CACCTCCGGCTTCTTCAACGACACCATCGACCCGCGCCAGCGCTGCGTGGCCATC
TGGACCTACAACACCCCGGAGTCCGAGGAGCAGTACATCTCCTACAGCCTGGAC
GGCGGCTACACCTTCACCGAGTACCAGAAGAACCCCGTGCTGGCCGCCAACTCC
ACCCAGTTCCGCGACCCGAAGGTCTTCTGGTACGAGCCCTCCCAGAAGTGGATCA
TGACCGCGGCCAAGTCCCAGGACTACAAGATCGAGATCTACTCCTCCGACGACCT
GAAGTCCTGGAAGCTGGAGTCCGCGTTCGCCAACGAGGGCTTCCTCGGCTACCA
GTACGAGTGCCCCGGCCTGATCGAGGTCCCCACCGAGCAGGACCCCAGCAAGTC
CTACTGGGTGATGTTCATCTCCATCAACCCCGGCGCCCCGGCCGGCGGCTCCTTC
AACCAGTACTTCGTCGGCAGCTTCAACGGCACCCACTTCGAGGCCTTCGACAACC
AGTCCCGCGTGGTGGACTTCGGCAAGGACTACTACGCCCTGCAGACCTTCTTCAA
CACCGACCCGACCTACGGGAGCGCCCTGGGCATCGCGTGGGCCTCCAACTGGGA
GTACTCCGCCTTCGTGCCCACCAACCCCTGGCGCTCCTCCATGTCCCTCGTGCGCA
AGTTCTCCCTCAACACCGAGTACCAGGCCAACCCGGAGACGGAGCTGATCAACC
TGAAGGCCGAGCCGATCCTGAACATCAGCAACGCCGGCCCCTGGAGCCGGTTCG
CCACCAACACCACGTTGACGAAGGCCAACAGCTACAACGTCGACCTGTCCAACA
GCACCGGCACCCTGGAGTTCGAGCTGGTGTACGCCGTCAACACCACCCAGACGA
TCTCCAAGTCCGTGTTCGCGGACCTCTCCCTCTGGTTCAAGGGCCTGGAGGACCC
CGAGGAGTACCTCCGCATGGGCTTCGAGGTGTCCGCGTCCTCCTTCTTCCTGGAC
CGCGGGAACAGCAAGGTGAAGTTCGTGAAGGAGAACCCCTACTTCACCAACCGC
ATGAGCGTGAACAACCAGCCCTTCAAGAGCGAGAACGACCTGTCCTACTACAAG
GTGTACGGCTTGCTGGACCAGAACATCCTGGAGCTGTACTTCAACGACGGCGAC
GTCGTGTCCACCAACACCTACTTCATGACCACCGGGAACGCCCTGGGCTCCGTGA
ACATGACGACGGGGGTGGACAACCTGTTCTACATCGACAAGTTCCAGGTGCGCG

[1328]

AGGTCAAGTGACAATTGGCAGCAGCAGCTCGGATAGTATCGACACACTCTGGAC
GCTGGTCGTGTGATGGACTGTTGCCGCCACACTTGCTGCCTTGACCTGTGAATAT
CCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGTGTGTACGCGCTTT
TGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCCAGCATCCCCTTCC
CTCGTTTCATATCGCTTGCATCCCAACCGCAACTTATCTACGCTGTCCTGCTATCC
CTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCTCGCACAGCCTTGGTTTGGGCTC
CGCCTGTATTCTCCTGGTACTGCAACCTGTAAACCAGCACTGCAATGCTGATGCA
CGGGAAGTAGTGGGATGGGAACACAAATGGA

[1329]

[1330] 서열 번호 106

[1331] pSZ1125 (FAD2C) 5' 게놈 표적화 서열

GCTCTTCGAGGGGCTGGTCTGAATCCTTCAGGCGGGTGTTACCCGAGAAAGAAA
GGGTGCCGATTTCAAAGCAGACCCATGTGCCGGGCCCTGTGGCCTGTGTTGGCGC
CTATGTAGTCACCCCCCCCTCACCCAATTGTCGCCAGTTTGCGCACTCCATAAACTC
AAAACAGCAGCTTCTGAGCTGCGCTGTTCAAGAACACCTCTGGGGTTTGCTCACC
CGCGAGGTCGACGCCCAGCATGGCTATCAAGACGAACAGGCAGCCTGTGGAGAA
GCCTCCGTTCACGATCGGGACGCTGCGCAAGGCCATCCCCGCGCACTGTTTCGAG
CGCTCGGCGCTTCGTAGCAGCATGTACCTGGCCTTTGACATCGCGGTCATGTCCC
TGCTCTACGTCGCGTCGACGTACATCGACCCTGCACCGGTGCCTACGTGGGTCAA
GTACGGCATCATGTGGCCGCTCTACTGGTTCTTCCAGGTGTGTTTGAGGGTTTTGG
TTGCCCGTATTGAGGTCCTGGTGGCGCGCATGGAGGAGAAGGCGCCTGTCCCGCT
GACCCCCCCCGGCTACCCTCCCGGCACCTTCCAGGGCGCCTTCGGCACGGGTGTCT
GGGTGTGCGCGCACGAGTGCGGCCACCAGGCCTTTTCCTCCAGCCAGGCCATCAA
CGACGGCGTGGGCCTGGTGTTCCACAGCCTGCTGCTGGTGCCCTACTACTCCTGG
AAGCACTCGCACCGGGTACC

[1332]

[1333] 서열 번호 107

[1334] pSZ1125 (FAD2C) 3' 게놈 표적화 서열

CCGCCACCACTCCAACACGGGGTGCCTGGACAAGGACGAGGTGTTTGTGCCGCC
GCACCGCGCAGTGGCGCACGAGGGCCTGGAGTGGGAGGAGTGGCTGCCCATCCG
CATGGGCAAGGTGCTGGTCACCCTGACCCTGGGCTGGCCGCTGTACCTCATGTTC
AACGTCGCCTCGCGGCCGTACCCGCGCTTCGCCAACCACTTTGACCCGTGGTCGC
CCATCTTCAGCAAGCGCGAGCGCATCGAGGTGGTCATCTCCGACCTGGCGCTGGT
GGCGGTGCTCAGCGGGCTCAGCGTGCTGGGCCGCACCATGGGCTGGGCCTGGCT
GGTCAAGACCTACGTGGTGCCCTACCTGATCGTGAACATGTGGCTCGTGCTCATC
ACGCTGCTCCAGCACACGCACCCGGCGCTGCCGCACTACTTCGAGAAGGACTGG
GACTGGCTGCGCGGCGCCATGGCCACCGTGGACCGCTCCATGGGCCCGCCCTTCA
TGGACAACATCCTGCACCACATCTCCGACACCCACGTGCTGCACCACCTCTTCAG
CACCATCCCGCACTACCACGCCGAGGAGGCCTCCGCCGCCATCAGGCCCATCCTG
GGCAAGTACTACCAGTCCGACAGCCGCTGGGTCGGCCGCGCCCTGTGGGAGGAC
TGGCGCGACTGCCGCTACGTCGTCCCGGACGCGCCCGAGGACGACTCCGCGCTCT
GGTTCCACAAGTGAGTGAGTGAGAAGAGC

[1335]

[1336] 서열 번호 108

[1337] 에스. 칼베르겐시스 MEL1 아미노산 서열

MFAFYFLTACISLKGVFGVSPSYNGLGLTPQMGWDNWNTFACDVSEQLLLDTADRI
[1338] SDLGLKDMGYKYIILDDCWSSGRDSDGFLVADEQKFPNGMGHVADHLHNNSFLFG

MYSSAGEYTCAGYPGSLGREEEDAQFFANNRVDYLKYDNCYNKGQFGTPEISYHRY
KAMSDALNKTGRPIFYSLCNWGQDLTFYWGSGIANSWRMSGDVTAEFTRPDSRCPC
DGDEYDCKYAGFHCSIMNILNKAAPMGQNAGVGGWNDLDNLEVGVGNLTDDEEK
AHFSMWAMVKSPLIIGANVNNLKASSYSIYSQASVIAINQDSNGIPATRVWRYYVSD
TDEYGQGEIQMWSGPLDNGDQVVALLNGGSVSRPMNTTLEEIFFDSNLGSKKLTST
WDIYDLWANRVDNSTASAILGRNKTATGILYNATEQSYKDGLSKNDTRLFGQKIGSL
[1339] SPNAILNTTVPAHGIAFYRLRPSS

[1340] 서열 번호 109

[1341] 에스. 칼베르겐시스 MEL1 천연 신호 펩티드 아미노산 서열

[1342] MFAFYFLTACISLKGVFG

[1343] 서열 번호 110

[1344] 에스. 칼베르겐시스 MEL1 발현 카세트의 관련 서열

GCGGCCGCGTGGACGAGGGCTACAACCCCGCCTACGGCGCGCGCCCGCTGCGCC
GCGCCATCATGCGCCTGCTGGAGGACGCGCTGGCCGAGCGCATGCTCGCCGGCG
ACGTCAAGGAGGGCGACTCGGTCATCATGGACGTCGATGGCGATGGCGCCATCA
GCGTCCTCAACGGAGACCGCACCCACACCACCACCATCGACTCCTCCCCGGCGG
GCATCTCGTAGACGCGAGAGGGAGGGAGGGAGGATTTTCAGAGGGAGATGAGA
CGAGGACTGGGCCGGGGGCCTTGGGGCGCTGCTGGAGCGTGGTGAGAGCGCGGC
GGACGTGCCTTTTCTTCTTCCGTGCGCGCGCTCTTGGCCATTGATCCCCGATTCGC
GCCCGCATCCCCCCACTGCCCCCATCATCTTGCCTGTTGTCGTGGCACTGACATA
AACCCCCTGCGCTGCGCTGCTCCGCTACTATTGATATAGGTCTCACGCGCCAATC
TTTTTTGCTCCGGGTAACCGTCTGGACGCCAGAATTCCTTTCTTGCGCTATGACAC
TTCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTTCCCGGCGCTGCATGCAAC
ACCGATGATGCTTCGACCCCCCGAAGCTCCTTCGGGGCTGCATGGGCGCTCCGAT
GCCGCTCCAGGGCGAGCGCTGTTTAAATAGCCAGGCCCCCGATTGCAAAGACAT
TATAGCGAGCTACCAAAGCCATATTCAAACACCTAGATCACTACCACTTCTACAC
AGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGGGCGCCTCTTCCTCTTCGT
TTCAGTCACAACCCGCAAACACTAGTATGTTCGCGTTCTACTTCCTGACGGCCTG
CATCTCCCTGAAGGGCGTGTTCGGCGTCTCCCCCTCCTACAACGGCCTGGGCCTG
ACGCCCCAGATGGGCTGGGACAACTGGAACACGTTCGCCTGCGACGTCTCCGAG
CAGCTGCTGCTGGACACGGCCGACCGCATCTCCGACCTGGGCCTGAAGGACATG
GGCTACAAGTACATCATCCTGGACGACTGCTGGTCCTCCGGCCGCGACTCCGACG
GCTTCCTGGTCGCCGACGAGCAGAAGTTCCCCAACGGCATGGGCCACGTCGCCG
ACCACCTGCACAACAACTCCTTCCTGTTCGGCATGTACTCCTCCGCGGGCGAGTA
CACGTGCGCCGGCTACCCCGGCTCCCTGGGCCGCGAGGAGGAGGACGCCCAGTT
CTTCGCGAACAACCGCGTGGACTACCTGAAGTACGACAACTGCTACAACAAGGG
CCAGTTCGGCACGCCCGAGATCTCCTACCACCGCTACAAGGCCATGTCCGACGCC
CTGAACAAGACGGGCCGCCCCATCTTCTACTCCCTGTGCAACTGGGGCCAGGACC
TGACCTTCTACTGGGGCTCCGGCATCGCGAACTCCTGGCGCATGTCCGGCGACGT
CACGGCGGAGTTCACGCGCCCCGACTCCCGCTGCCCCTGCGACGGCGACGAGTA
CGACTGCAAGTACGCCGGCTTCCACTGCTCCATCATGAACATCCTGAACAAGGCC
GCCCCCATGGGCCAGAACGCGGGCGTCGGCGGCTGGAACGACCTGGACAACCTG
GAGGTCGGCGTCGGCAACCTGACGGACGACGAGGAGAAGGCGCACTTCTCCATG
TGGGCCATGGTGAAGTCCCCCCTGATCATCGGCGCGAACGTGAACAACCTGAAG
GCCTCCTCCTACTCCATCTACTCCCAGGCGTCCGTCATCGCCATCAACCAGGACT
CCAACGGCATCCCCGCCACGCGCGTCTGGCGCTACTACGTGTCCGACACGGACG

[1345]

AGTACGGCCAGGGCGAGATCCAGATGTGGTCCGGCCCCCTGGACAACGGCGACC
AGGTCGTGGCGCTGCTGAACGGCGGCTCCGTGTCCCGCCCCATGAACACGACCCT
GGAGGAGATCTTCTTCGACTCCAACCTGGGCTCCAAGAAGCTGACCTCCACCTGG
GACATCTACGACCTGTGGGCGAACCGCGTCGACAACTCCACGGCGTCCGCCATCC
TGGGCCGCAACAAGACCGCCACCGGCATCCTGTACAACGCCACCGAGCAGTCCT
ACAAGGACGGCCTGTCCAAGAACGACACCCGCCTGTTCGGCCAGAAGATCGGCT
CCCTGTCCCCCAACGCGATCCTGAACACGACCGTCCCCGCCCACGGCATCGCGTT
CTACCGCCTGCGCCCCTCCTCCTGACAATTGAAGCAGCAGCAGCTCGGATAGTAT
CGACACACTCTGGACGCTGGTCGTGTGATGGACTGTTGCCGCCACACTTGCTGCC
TTGACCTGTGAATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTT
GTGTGTACGCGCTTTTGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACCC
CCAGCATCCCCTTCCCTCGTTTCATATCGCTTGCATCCCAACCGCAACTTATCTAC
GCTGTCCTGCTATCCCTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCTCGCACAG
CCTTGGTTTGGGCTCCGCCTGTATTCTCCTGGTACTGCAACCTGTAAACCAGCACT
GCAATGCTGATGCACGGGAAGTAGTGGGATGGGAACACAAATGGAAAGCTTGAG
CTCAGAATAGTATCGGGTGATGCGAAGTCAGAACCAGGCAGGGCCTGTCGCCTG
AGGTGGCAACGATGGGAAGCAATCAATCTGGGTACAGTCGTCCGCACGATCCCG
TGATCTCCCCCACCGACACCTATCCCCGCCCATCCCGGCCCACCCTTTCAGTCCCC
TCAGCATGCATTGTGCACCGCGACAAAGCATGTCTGCTCGTGCACTGGTTCAGGC
CACGGCGCACCGAGTCCTCGCCCTTCGCAGAGTGATCACCCTCCCCGGAACCAGC
CACGCTCGCTGCTGCGGGCCGATCAGCCGCGCGCACTCCCTGCAACTAGGGACA
ACTCAGGCAACCACGCGCCTCACAAGCATGGCCGCCGTGGCATCCAACCCGCTC
GTGACGGTGGGTGCGCAAGTGCCAGGGGCCTCGTCGTCACGGCGTGCATCCTCG
AGGGATGCGATCCGGCAACTATATGTCGTTTATCTCCCCACCAATCACAGGATGA
GCCCCTGTCTAGA

[1346]

[1347] 서열 번호 111

[1348] 코돈-최적화 에스. 칼베르겐시스 MEL1 서열

ATGTTCGCGTTCTACTTCCTGACGGCCTGCATCTCCCTGAAGGGCGTGTTCGGCGT
CTCCCCCTCCTACAACGGCCTGGGCCTGACGCCCCAGATGGGCTGGGACAACTGG
AACACGTTCGCCTGCGACGTCTCCGAGCAGCTGCTGCTGGACACGGCCGACCGC
ATCTCCGACCTGGGCCTGAAGGACATGGGCTACAAGTACATCATCCTGGACGACT
GCTGGTCCTCCGGCCGCGACTCCGACGGCTTCCTGGTCGCCGACGAGCAGAAGTT
CCCCAACGGCATGGGCCACGTCGCCGACCACCTGCACAACAACTCCTTCCTGTTC
GGCATGTACTCCTCCGCGGGCGAGTACACGTGCGCCGGCTACCCCGGCTCCCTGG
GCCGCGAGGAGGAGGACGCCCAGTTCTTCGCGAACAACCGCGTGGACTACCTGA
AGTACGACAACTGCTACAACAAGGGCCAGTTCGGCACGCCCGAGATCTCCTACC
ACCGCTACAAGGCCATGTCCGACGCCCTGAACAAGACGGGCCGCCCCATCTTCTA
CTCCCTGTGCAACTGGGGCCAGGACCTGACCTTCTACTGGGGCTCCGGCATCGCG
AACTCCTGGCGCATGTCCGGCGACGTCACGGCGGAGTTCACGCGCCCCGACTCCC
GCTGCCCCTGCGACGGCGACGAGTACGACTGCAAGTACGCCGGCTTCCACTGCTC
CATCATGAACATCCTGAACAAGGCCGCCCCCATGGGCCAGAACGCGGGCGTCGG
CGGCTGGAACGACCTGGACAACCTGGAGGTCGGCGTCGGCAACCTGACGGACGA
CGAGGAGAAGGCGCACTTCTCCATGTGGGCCATGGTGAAGTCCCCCCTGATCATC
GGCGCGAACGTGAACAACCTGAAGGCCTCCTCCTACTCCATCTACTCCCAGGCGT
CCGTCATCGCCATCAACCAGGACTCCAACGGCATCCCCGCCACGCGCGTCTGGCG
CTACTACGTGTCCGACACGGACGAGTACGGCCAGGGCGAGATCCAGATGTGGTC
CGGCCCCCTGGACAACGGCGACCAGGTCGTGGCGCTGCTGAACGGCGGCTCCGT
GTCCCGCCCCATGAACACGACCCTGGAGGAGATCTTCTTCGACTCCAACCTGGGC
[1349]

TCCAAGAAGCTGACCTCCACCTGGGACATCTACGACCTGTGGGCGAACCGCGTC
GACAACTCCACGGCGTCCGCCATCCTGGGCCGCAACAAGACCGCCACCGGCATC
CTGTACAACGCCACCGAGCAGTCCTACAAGGACGGCCTGTCCAAGAACGACACC
CGCCTGTTCGGCCAGAAGATCGGCTCCCTGTCCCCCAACGCGATCCTGAACACGA
CCGTCCCCGCCCACGGCATCGCGTTCTACCGCCTGCGCCCCTCCTCCTGA
[1350]

[1351] 서열 번호 112

[1352] 5' Clp 상동 재조합 표적화 서열

GCGGCCGCGTGGACGAGGGCTACAACCCCGCCTACGGCGCGCGCCCGCTGCGCC
GCGCCATCATGCGCCTGCTGGAGGACGCGCTGGCCGAGCGCATGCTCGCCGGCG
ACGTCAAGGAGGGCGACTCGGTCATCATGGACGTCGATGGCGATGGCGCCATCA
GCGTCCTCAACGGAGACCGCACCCACACCACCACCATCGACTCCTCCCCGGCGG
GCATCTCGTAGACGCGAGAGGGAGGGAGGGAGGATTTTCAGAGGGAGATGAGA
CGAGGACTGGGCCGGGGGCCTTGGGGCGCTGCTGGAGCGTGGTGAGAGCGCGGC
GGACGTGCCTTTTCTTCTTCCGTGCGCGCGCTCTTGGCCATTGATCCCCGATTCGC
GCCCGCATCCCCCCACTGCCCCCATCATCTTGCCTGTTGTCGTGGCACTGACATA
AACCCCCTGCGCTGCGCTGCTCCGCTACTATTGATATAGGTCTCACGCGCCAATC
TTTTTTGCTCCGGGTAACCGTCTGGACGCCAGAATTC
[1353]

[1354] 서열 번호 113

[1355] 3' Clp 상동 재조합 표적화 서열

GAGCTCAGAATAGTATCGGGTGATGCGAAGTCAGAACCAGGCAGGGCCTGTCGC
CTGAGGTGGCAACGATGGGAAGCAATCAATCTGGGTACAGTCGTCCGCACGATC
CCGTGATCTCCCCCACCGACACCTATCCCCGCCCATCCCGGCCCACCCTTTCAGTC
CCCTCAGCATGCATTGTGCACCGCGACAAAGCATGTCTGCTCGTGCACTGGTTCA
GGCCACGGCGCACCGAGTCCTCGCCCTTCGCAGAGTGATCACCCTCCCCGGAACC
AGCCACGCTCGCTGCTGCGGGCCGATCAGCCGCGCGCACTCCCTGCAACTAGGG
ACAACTCAGGCAACCACGCGCCTCACAAGCATGGCCGCCGTGGCATCCAACCCG
CTCGTGACGGTGGGTGCGCAAGTGCCAGGGGCCTCGTCGTCACGGCGTGCATCCT
CGAGGGATGCGATCCGGCAACTATATGTCGTTTATCTCCCCACCAATCACAGGAT
GAGCCCCTGTCTAGA
[1356]

[1357] 서열 번호 114

[1358] 클라미도모나스 레인하르드티 TUB2 프로모터/5'UTR

CTTTCTTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTT
CCCGGCGCTGCATGCAACACCGATGATGCTTCGACCCCCCGAAGCTCCTTCGGGG
CTGCATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGTTTAAATAGCCAGGCC
CCCGATTGCAAAGACATTATAGCGAGCTACCAAAGCCATATTCAAACACCTAGA
TCACTACCACTTCTACACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGG
GCGCCTCTTCCTCTTCGTTTCAGTCACAACCCGCAAAC
[1359]

[1360] 서열 번호 115

[1361] 클로렐라 불가리스 질산염 환원 효소 3'UTR

[1362]
GCAGCAGCAGCTCGGATAGTATCGACACACTCTGGACGCTGGTCGTGTGATGGA
CTGTTGCCGCCACACTTGCTGCCTTGACCTGTGAATATCCCTGCCGCTTTTATCAA

[1363]
ACAGCCTCAGTGTGTTTGATCTTGTGTGTACGCGCTTTTGCGAGTTGCTAGCTGCT
TGTGCTATTTGCGAATACCACCCCCAGCATCCCCTTCCCTCGTTTCATATCGCTTG
CATCCCAACCGCAACTTATCTACGCTGTCCTGCTATCCCTCAGCGCTGCTCCTGCT
CCTGCTCACTGCCCCTCGCACAGCCTTGGTTTGGGCTCCGCCTGTATTCTCCTGGT
ACTGCAACCTGTAAACCAGCACTGCAATGCTGATGCACGGGAAGTAGTGGGATG
GGAACACAAATGGAAAGCTT

[1364] 서열 번호 116

[1365] 에이. 나이저 *A1gC* 아미노산 서열

[1366]
MIGSSHAVVALGLFTLYGHSAAAPAIGASNSQTIVTNGTSFALNGDNVSYRFHVNSS
TGDLISDHFGGVVSGTIPSPVEPAVNGWVGMPGRIRREFPDQGRGDFRIPAVRIRESA
GYTVSDLQYVSHEVIEGKYALPGLPATFGDAQDATTLVVHLYDNYSSVAADLSYSIF
PKYDAIVRSVNVTNQGPGNITIEALASISIDFPYEDLDMVSLRGDWAREANVQRSKV
QYGVQGFGSSTGYSSHLHNPFLAIVDPATTESQGEAWGFNLVYTGSFSAQVEKGSQG
FTRALLGFNPDQLSWNLGPGETLTSPECVAVYSDKGLGSVSRKFHRLYRNHLMKSK
FATSDRPVLLNSWEGVYFDYNQSSIETLAEESAALGVHLFVMDDGWFGDKYPRVSD
NAGLGDWMPNPARFPDGLTPVVQDITNLTVNGTESTKLRFGIWVEPEMVNPNSTLY
HEHPEWALHAGPYPRTERRNQLVLNLALPAVQDFIIDFMTNLLQDTGISYVKWDNN
RGIHETPSPSTDHQYMLGLYRVFDTLTTRFPDVLWEGCASGGGRFDAGMLQYVPQI
WTSDNTDAIDRITIQFGTSLAYPPSAMGAHLSAVPNAQTGRTVPFTFRAHVAMMGGS
FGLELDPATVEGDEIVPELLALAEKVNPIILNGDLYRLRLPQDSQWPAALFVSQDGAQ
AVLFYFQVQPNVNHAVPWVRLQGLDPKADYTVDGDQTYSGATLMNLGLQYSFDTE
YGSKVVFLERQ

[1367] 서열 번호 117

[1368] 에이. 나이저 *A1gC* 천연 신호 펩티드 아미노산 서열

[1369]
MIGSSHAVVALGLFTLYGHSAAAPAIGA

[1370] 서열 번호 118

[1371] 코돈-최적화 에이. 나이저 *A1gC* 코딩 서열

[1372]
ATGATCGGCTCCTCCCACGCGGTCGTCGCCCTGGGCCTGTTCACCCTGTACGGCC
ACTCCGCCGCCGCGCCCGCCATCGGCGCCTCCAACTCCCAGACCATCGTCACGAA
CGGCACCTCCTTCGCCCTGAACGGCGACAACGTGTCCTACCGCTTCCACGTGAAC
TCCTCCACGGGCGACCTGATCTCCGACCACTTCGGCGGCGTGGTGTCCGGCACCA
TCCCCTCCCCCGTGGAGCCCGCGGTCAACGGCTGGGTGGGCATGCCCGGCCGCAT
CCGCCGCGAGTTCCCCGACCAGGGCCGCGGCGACTTCCGCATCCCCGCGGTGCGC
ATCCGCGAGTCCGCCGGCTACACCGTCTCCGACCTGCAGTACGTGTCCCACGAGG
TGATCGAGGGCAAGTACGCGCTGCCCGGCCTGCCCGCCACGTTCGGCGACGCCC
AGGACGCCACCACCCTGGTGGTGCACCTGTACGACAACTACTCCTCCGTCGCGGC
CGACCTGTCCTACTCCATCTTCCCCAAGTACGACGCGATCGTCCGCTCCGTGAAC
GTGACCAACCAGGGCCCCGGCAACATCACCATCGAGGCGCTGGCCTCCATCTCC
ATCGACTTCCCCTACGAGGACCTGGACATGGTGTCCCTGCGCGGCGACTGGGCCC
GCGAGGCGAACGTGCAGCGCTCCAAGGTCCAGTACGGCGTGCAGGGCTTCGGCT
CCTCCACCGGCTACTCCTCCCACCTGCACAACCCCTTCCTGGCGATCGTCGACCC
CGCGACCACCGAGTCCCAGGGCGAGGCCTGGGGCTTCAACCTGGTCTACACCGG
CTCCTTCTCCGCCCAGGTCGAGAAGGGCTCCCAGGGCTTCACGCGCGCCCTGCTG
GGCTTCAACCCCGACCAGCTGTCCTGGAACCTGGGCCCCGGCGAGACGCTGACG

TCCCCCGAGTGCGTCGCCGTCTACTCCGACAAGGGCCTGGGCTCCGTCTCCCGCA
AGTTCCACCGCCTGTACCGCAACCACCTGATGAAGTCCAAGTTCGCCACGTCCGA
CCGCCCCGTGCTGCTGAACTCCTGGGAGGGCGTCTACTTCGACTACAACCAGTCC
TCCATCGAGACGCTGGCGGAGGAGTCCGCCGCCCTGGGCGTGCACCTGTTCGTCA
TGGACGACGGCTGGTTCGGCGACAAGTACCCCCGCGTGTCCGACAACGCCGGCC
TGGGCGACTGGATGCCCAACCCCGCCCGCTTCCCCGACGGCCTGACGCCCGTCGT
GCAGGACATCACCAACCTGACCGTCAACGGCACCGAGTCCACCAAGCTGCGCTT
CGGCATCTGGGTGGAGCCCGAGATGGTGAACCCCAACTCCACCCTGTACCACGA
GCACCCCGAGTGGGCGCTGCACGCGGGCCCCTACCCCCGCACCGAGCGCCGCAA
CCAGCTGGTCCTGAACCTGGCCCTGCCCGCGGTCCAGGACTTCATCATCGACTTC
ATGACCAACCTGCTGCAGGACACCGGCATCTCCTACGTCAAGTGGGACAACAAC
CGCGGCATCCACGAGACGCCCTCCCCCTCCACGGACCACCAGTACATGCTGGGCC
TGTACCGCGTGTTCGACACGCTGACCACGCGCTTCCCCGACGTCCTGTGGGAGGG
CTGCGCGTCCGGCGGCGGCCGGTTCGACGCCGGCATGCTGCAGTACGTCCCCCAG
ATCTGGACGTCCGACAACACGGACGCGATCGACCGCATCACCATCCAGTTCGGC
ACCTCCCTGGCCTACCCCCCCTCCGCCATGGGCGCCCACCTGTCCGCGGTGCCCA
ACGCCCAGACCGGCCGCACGGTGCCCTTCACCTTCCGCGCCCACGTCGCGATGAT
GGGCGGCTCCTTCGGCCTGGAGCTGGACCCCGCCACCGTGGAGGGCGACGAGAT
CGTGCCCGAGCTGCTGGCGCTGGCCGAGAAGGTGAACCCCATCATCCTGAACGG
CGACCTGTACCGCCTGCGCCTGCCCCAGGACTCCCAGTGGCCCGCGGCCCTGTTC
GTGTCCCAGGACGGCGCCCAGGCCGTCCTGTTCTACTTCCAGGTGCAGCCCAACG
TCAACCACGCCGTCCCCTGGGTCCGCCTGCAGGGCCTGGACCCCAAGGCCGACTA
CACGGTGGACGGCGACCAGACGTACTCCGGCGCGACCCTGATGAACCTGGGCCT
GCAGTACTCCTTCGACACCGAGTACGGCTCCAAGGTGGTGTTCCTGGAGCGCCAG
TAA

[1373]

[1374] 서열 번호 119

[1375] 에이. 나이저 *A1gC* 발현 카세트의 관련 서열

GCGGCCGCGTGGACGAGGGCTACAACCCCGCCTACGGCGCGCGCCCGCTGCGCC
GCGCCATCATGCGCCTGCTGGAGGACGCGCTGGCCGAGCGCATGCTCGCCGGCG
ACGTCAAGGAGGGCGACTCGGTCATCATGGACGTCGATGGCGATGGCGCCATCA
GCGTCCTCAACGGAGACCGCACCCACACCACCACCATCGACTCCTCCCCGGCGG
GCATCTCGTAGACGCGAGAGGGAGGGAGGGAGGATTTTCAGAGGGAGATGAGA
CGAGGACTGGGCCGGGGGCCTTGGGGCGCTGCTGGAGCGTGGTGAGAGCGCGGC
GGACGTGCCTTTTCTTCTTCCGTGCGCGCGCTCTTGGCCATTGATCCCCGATTCGC
GCCCGCATCCCCCCACTGCCCCCATCATCTTGCCTGTTGTCGTGGCACTGACATA
AACCCCCTGCGCTGCGCTGCTCCGCTACTATTGATATAGGTCTCACGCGCCAATC
TTTTTTGCTCCGGGTAACCGTCTGGACGCCAGAATTCCTTTCTTGCGCTATGACAC
TTCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTTCCCGGCGCTGCATGCAAC
ACCGATGATGCTTCGACCCCCCGAAGCTCCTTCGGGGCTGCATGGGCGCTCCGAT
GCCGCTCCAGGGCGAGCGCTGTTTAAATAGCCAGGCCCCCGATTGCAAAGACAT
TATAGCGAGCTACCAAAGCCATATTCAAACACCTAGATCACTACCACTTCTACAC
AGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGGGCGCCTCTTCCTCTTCGT
TTCAGTCACAACCCGCAAACACTAGTATGATCGGCTCCTCCCACGCGGTCGTCGC
CCTGGGCCTGTTCACCCTGTACGGCCACTCCGCCGCCGCGCCCGCCATCGGCGCC
TCCAACTCCCAGACCATCGTCACGAACGGCACCTCCTTCGCCCTGAACGGCGACA
ACGTGTCCTACCGCTTCCACGTGAACTCCTCCACGGGCGACCTGATCTCCGACCA
CTTCGGCGGCGTGGTGTCCGGCACCATCCCCTCCCCCGTGGAGCCCGCGGTCAAC
GGCTGGGTGGGCATGCCCGGCCGCATCCGCCGCGAGTTCCCCGACCAGGGCCGC

[1376]

GGCGACTTCCGCATCCCCGCGGTGCGCATCCGCGAGTCCGCCGGCTACACCGTCT
CCGACCTGCAGTACGTGTCCCACGAGGTGATCGAGGGCAAGTACGCGCTGCCCG
GCCTGCCCGCCACGTTCGGCGACGCCCAGGACGCCACCACCCTGGTGGTGCACCT
GTACGACAACTACTCCTCCGTCGCGGCCGACCTGTCCTACTCCATCTTCCCCAAG
TACGACGCGATCGTCCGCTCCGTGAACGTGACCAACCAGGGCCCCGGCAACATC
ACCATCGAGGCGCTGGCCTCCATCTCCATCGACTTCCCCTACGAGGACCTGGACA
TGGTGTCCCTGCGCGGCGACTGGGCCCGCGAGGCGAACGTGCAGCGCTCCAAGG
TCCAGTACGGCGTGCAGGGCTTCGGCTCCTCCACCGGCTACTCCTCCCACCTGCA
CAACCCCTTCCTGGCGATCGTCGACCCCGCGACCACCGAGTCCCAGGGCGAGGC
CTGGGGCTTCAACCTGGTCTACACCGGCTCCTTCTCCGCCCAGGTCGAGAAGGGC
TCCCAGGGCTTCACGCGCGCCCTGCTGGGCTTCAACCCCGACCAGCTGTCCTGGA
ACCTGGGCCCCGGCGAGACGCTGACGTCCCCCGAGTGCGTCGCCGTCTACTCCGA
CAAGGGCCTGGGCTCCGTCTCCCGCAAGTTCCACCGCCTGTACCGCAACCACCTG
ATGAAGTCCAAGTTCGCCACGTCCGACCGCCCCGTGCTGCTGAACTCCTGGGAGG
GCGTCTACTTCGACTACAACCAGTCCTCCATCGAGACGCTGGCGGAGGAGTCCGC
CGCCCTGGGCGTGCACCTGTTCGTCATGGACGACGGCTGGTTCGGCGACAAGTAC
CCCCGCGTGTCCGACAACGCCGGCCTGGGCGACTGGATGCCCAACCCCGCCCGCT
TCCCCGACGGCCTGACGCCCGTCGTGCAGGACATCACCAACCTGACCGTCAACG
GCACCGAGTCCACCAAGCTGCGCTTCGGCATCTGGGTGGAGCCCGAGATGGTGA
ACCCCAACTCCACCCTGTACCACGAGCACCCCGAGTGGGCGCTGCACGCGGGCC
CCTACCCCCGCACCGAGCGCCGCAACCAGCTGGTCCTGAACCTGGCCCTGCCCGC
GGTCCAGGACTTCATCATCGACTTCATGACCAACCTGCTGCAGGACACCGGCATC
TCCTACGTCAAGTGGGACAACAACCGCGGCATCCACGAGACGCCCTCCCCCTCCA
CGGACCACCAGTACATGCTGGGCCTGTACCGCGTGTTCGACACGCTGACCACGCG
CTTCCCCGACGTCCTGTGGGAGGGCTGCGCGTCCGGCGGCGGCCGGTTCGACGCC
GGCATGCTGCAGTACGTCCCCCAGATCTGGACGTCCGACAACACGGACGCGATC
GACCGCATCACCATCCAGTTCGGCACCTCCCTGGCCTACCCCCCCTCCGCCATGG
GCGCCCACCTGTCCGCGGTGCCCAACGCCCAGACCGGCCGCACGGTGCCCTTCAC
CTTCCGCGCCCACGTCGCGATGATGGGCGGCTCCTTCGGCCTGGAGCTGGACCCC
GCCACCGTGGAGGGCGACGAGATCGTGCCCGAGCTGCTGGCGCTGGCCGAGAAG
GTGAACCCCATCATCCTGAACGGCGACCTGTACCGCCTGCGCCTGCCCCAGGACT
CCCAGTGGCCCGCGGCCCTGTTCGTGTCCCAGGACGGCGCCCAGGCCGTCCTGTT
CTACTTCCAGGTGCAGCCCAACGTCAACCACGCCGTCCCCTGGGTCCGCCTGCAG
GGCCTGGACCCCAAGGCCGACTACACGGTGGACGGCGACCAGACGTACTCCGGC
GCGACCCTGATGAACCTGGGCCTGCAGTACTCCTTCGACACCGAGTACGGCTCCA
AGGTGGTGTTCCTGGAGCGCCAGTAACAATTGAAGCAGCAGCAGCTCGGATAGT
ATCGACACACTCTGGACGCTGGTCGTGTGATGGACTGTTGCCGCCACACTTGCTG
CCTTGACCTGTGAATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATC
TTGTGTGTACGCGCTTTTGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCAC
CCCCAGCATCCCCTTCCCTCGTTTCATATCGCTTGCATCCCAACCGCAACTTATCT
ACGCTGTCCTGCTATCCCTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCTCGCAC
AGCCTTGGTTTGGGCTCCGCCTGTATTCTCCTGGTACTGCAACCTGTAAACCAGC
ACTGCAATGCTGATGCACGGGAAGTAGTGGGATGGGAACACAAATGGAAAGCTT
GAGCTCAGAATAGTATCGGGTGATGCGAAGTCAGAACCAGGCAGGGCCTGTCGC
CTGAGGTGGCAACGATGGGAAGCAATCAATCTGGGTACAGTCGTCCGCACGATC
CCGTGATCTCCCCCACCGACACCTATCCCCGCCCATCCCGGCCCACCCTTTCAGTC
CCCTCAGCATGCATTGTGCACCGCGACAAAGCATGTCTGCTCGTGCACTGGTTCA
GGCCACGGCGCACCGAGTCCTCGCCCTTCGCAGAGTGATCACCCTCCCCGGAACC
AGCCACGCTCGCTGCTGCGGGCCGATCAGCCGCGCGCACTCCCTGCAACTAGGG
ACAACTCAGGCAACCACGCGCCTCACAAGCATGGCCGCCGTGGCATCCAACCCG

[1377]

CTCGTGACGGTGGGTGCGCAAGTGCCAGGGGCCTCGTCGTCACGGCGTGCATCCT
CGAGGGATGCGATCCGGCAACTATATGTCGTTTATCTCCCCACCAATCACAGGAT
GAGCCCCTGTCTAGA

[1378]

[1379] 서열 번호 120

[1380] 씨. 테트라고노볼라 α-갈락토시다제 아미노산 서열

MATHYSIIGGMIIVVLLMIIGSEGGRLLEKKNRTSAEAEHYNVRRYLAENGLGQTPPM
GWNSWNHFGCDINENVVRETADAMVSTGLAALGYQYINLDDCWAELNRDSEGNM
VPNAAAFPSGIKALADYVHSKGLKLGVYSDAGNQTCSKRMPGSLGHEEQDAKTFAS
WGVDYLKYDNCENLGISVKERYPPMGKALLSSGRPIFFSMCEWGWEDPQIWAKSIG
NSWRTTGDIEDNWNSMTSIADSNDKWASYAGPGGWNDPDMLEVGNGGMTTEEYR
SHFSIWALAKAPLLVGCDIRAMDDTTHELISNAEVIAVNQDKLGVQGKKVKSTNDLE
VWAGPLSDNKVAVILWNRSSSRATVTASWSDIGLQQGTTVDARDLWEHSTQSLVSG
EISAEIDSHACKMYVLTPRS

[1381]

[1382] 서열 번호 121

[1383] 코돈 최적화 씨. 테트라고노볼라 α-갈락토시다제 서열

ATGGCCACCCACTACTCCATCATCGGCGGCATGATCATCGTCGTCCTGCTGATGA
TCATCGGCTCCGAGGGCGGCCGGCTGCTGGAGAAGAAGAACCGCACCTCCGCCG
AGGCGGAGCACTACAACGTGCGCCGCTACCTGGCCGAGAACGGCCTGGGCCAGA
CCCCCCCCATGGGCTGGAACTCCTGGAACCACTTCGGCTGCGACATCAACGAGA
ACGTCGTCCGCGAGACGGCGGACGCCATGGTGTCCACCGGCCTGGCCGCCCTGG
GCTACCAGTACATCAACCTGGACGACTGCTGGGCGGAGCTGAACCGCGACTCCG
AGGGCAACATGGTGCCCAACGCCGCCGCGTTCCCCTCCGGCATCAAGGCGCTGG
CGGACTACGTCCACTCCAAGGGCCTGAAGCTGGGCGTCTACTCCGACGCGGGCA
ACCAGACCTGCTCCAAGCGCATGCCCGGCTCCCTGGGCCACGAGGAGCAGGACG
CGAAGACGTTCGCCTCCTGGGGCGTCGACTACCTGAAGTACGACAACTGCGAGA
ACCTGGGCATCTCCGTGAAGGAGCGCTACCCCCCCATGGGCAAGGCGCTGCTGTC
CTCCGGCCGCCCCATCTTCTTCTCCATGTGCGAGTGGGGCTGGGAGGACCCCCAG
ATCTGGGCCAAGTCCATCGGCAACTCCTGGCGCACCACCGGCGACATCGAGGAC
AACTGGAACTCCATGACCTCCATCGCCGACTCCAACGACAAGTGGGCCTCCTACG
CCGGCCCCGGCGGCTGGAACGACCCCGACATGCTGGAGGTCGGCAACGGCGGCA
TGACGACGGAGGAGTACCGCTCCCACTTCTCCATCTGGGCCCTGGCCAAGGCGCC
CCTGCTGGTGGGCTGCGACATCCGCGCGATGGACGACACGACGCACGAGCTGAT
CTCCAACGCCGAGGTGATCGCGGTGAACCAGGACAAGCTGGGCGTGCAGGGCAA
GAAGGTCAAGTCCACGAACGACCTGGAGGTCTGGGCCGGCCCCCTGTCCGACAA
CAAGGTGGCGGTGATCCTGTGGAACCGCTCCTCCTCCCGCGCCACCGTCACCGCG
TCCTGGTCCGACATCGGCCTGCAGCAGGGCACCACCGTCGACGCGCGCGACCTGT
GGGAGCACTCCACGCAGTCCCTGGTGTCCGGCGAGATCTCCGCCGAGATCGACTC
CCACGCCTGCAAGATGTACGTCCTGACGCCCCGCTCCTAA

[1384]

[1385] 서열 번호 122

[1386] 씨. 테트라고노볼라 α-갈락토시다제 발현 카세트의 관련 서열

GCGGCCGCGTGGACGAGGGCTACAACCCCGCCTACGGCGCGCGCCCGCTGCGCC
GCGCCATCATGCGCCTGCTGGAGGACGCGCTGGCCGAGCGCATGCTCGCCGGCG
ACGTCAAGGAGGGCGACTCGGTCATCATGGACGTCGATGGCGATGGCGCCATCA
GCGTCCTCAACGGAGACCGCACCCACACCACCACCATCGACTCCTCCCCGGCGG

[1387]

GCATCTCGTAGACGCGAGAGGGAGGGAGGGAGGATTTTCAGAGGGAGATGAGA
CGAGGACTGGGCCGGGGGCCTTGGGGCGCTGCTGGAGCGTGGTGAGAGCGCGGC
GGACGTGCCTTTTCTTCTTCCGTGCGCGCGCTCTTGGCCATTGATCCCCGATTCGC
GCCCGCATCCCCCCACTGCCCCCATCATCTTGCCTGTTGTCGTGGCACTGACATA
AACCCCCTGCGCTGCGCTGCTCCGCTACTATTGATATAGGTCTCACGCGCCAATC
TTTTTTGCTCCGGGTAACCGTCTGGACGCCAGAATTCCTTTCTTGCGCTATGACAC
TTCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTTCCCGGCGCTGCATGCAAC
ACCGATGATGCTTCGACCCCCCGAAGCTCCTTCGGGGCTGCATGGGCGCTCCGAT
GCCGCTCCAGGGCGAGCGCTGTTTAAATAGCCAGGCCCCCGATTGCAAAGACAT
TATAGCGAGCTACCAAAGCCATATTCAAACACCTAGATCACTACCACTTCTACAC
AGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGGGCGCCTCTTCCTCTTCGT
TTCAGTCACAACCCGCAAACACTAGTATGGCCACCCACTACTCCATCATCGGCGG
CATGATCATCGTCGTCCTGCTGATGATCATCGGCTCCGAGGGCGGCCGGCTGCTG
GAGAAGAAGAACCGCACCTCCGCCGAGGCGGAGCACTACAACGTGCGCCGCTAC
CTGGCCGAGAACGGCCTGGGCCAGACCCCCCCCCATGGGCTGGAACTCCTGGAAC
CACTTCGGCTGCGACATCAACGAGAACGTCGTCCGCGAGACGGCGGACGCCATG
GTGTCCACCGGCCTGGCCGCCCTGGGCTACCAGTACATCAACCTGGACGACTGCT
GGGCGGAGCTGAACCGCGACTCCGAGGGCAACATGGTGCCCAACGCCGCCGCGT
TCCCCTCCGGCATCAAGGCGCTGGCGGACTACGTCCACTCCAAGGGCCTGAAGCT
GGGCGTCTACTCCGACGCGGGCAACCAGACCTGCTCCAAGCGCATGCCCGGCTC
CCTGGGCCACGAGGAGCAGGACGCGAAGACGTTCGCCTCCTGGGGCGTCGACTA
CCTGAAGTACGACAACTGCGAGAACCTGGGCATCTCCGTGAAGGAGCGCTACCC
CCCCATGGGCAAGGCGCTGCTGTCCTCCGGCCGCCCCATCTTCTTCTCCATGTGC
GAGTGGGGCTGGGAGGACCCCCAGATCTGGGCCAAGTCCATCGGCAACTCCTGG
CGCACCACCGGCGACATCGAGGACAACTGGAACTCCATGACCTCCATCGCCGAC
TCCAACGACAAGTGGGCCTCCTACGCCGGCCCCGGCGGCTGGAACGACCCCGAC
ATGCTGGAGGTCGGCAACGGCGGCATGACGACGGAGGAGTACCGCTCCCACTTC
TCCATCTGGGCCCTGGCCAAGGCGCCCCTGCTGGTGGGCTGCGACATCCGCGCGA
TGGACGACACGACGCACGAGCTGATCTCCAACGCCGAGGTGATCGCGGTGAACCC
AGGACAAGCTGGGCGTGCAGGGCAAGAAGGTCAAGTCCACGAACGACCTGGAG
GTCTGGGCCGGCCCCCTGTCCGACAACAAGGTGGCGGTGATCCTGTGGAACCGCT
CCTCCTCCCGCGCCACCGTCACCGCGTCCTGGTCCGACATCGGCCTGCAGCAGGG
CACCACCGTCGACGCGCGCGACCTGTGGGAGCACTCCACGCAGTCCCTGGTGTCC
GGCGAGATCTCCGCCGAGATCGACTCCCACGCCTGCAAGATGTACGTCCTGACGC
CCCGCTCCTAACAATTGAAGCAGCAGCAGCTCGGATAGTATCGACACACTCTGG
ACGCTGGTCGTGTGATGGACTGTTGCCGCCACACTTGCTGCCTTGACCTGTGAAT
ATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGTGTGTACGCGCT
TTTGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCCAGCATCCCCTT
CCCTCGTTTCATATCGCTTGCATCCCAACCGCAACTTATCTACGCTGTCCTGCTAT
CCCTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCTCGCACAGCCTTGGTTTGGGC
TCCGCCTGTATTCTCCTGGTACTGCAACCTGTAAACCAGCACTGCAATGCTGATG
CACGGGAAGTAGTGGGATGGGAACACAAATGGAAAGCTTGAGCTCAGAATAGTA
TCGGGTGATGCGAAGTCAGAACCAGGCAGGGCCTGTCGCCTGAGGTGGCAACGA
TGGGAAGCAATCAATCTGGGTACAGTCGTCCGCACGATCCCGTGATCTCCCCCAC
CGACACCTATCCCCGCCCATCCCGGCCCACCCTTTCAGTCCCCTCAGCATGCATT
GTGCACCGCGACAAAGCATGTCTGCTCGTGCACTGGTTCAGGCCACGGCGCACC
GAGTCCTCGCCCTTCGCAGAGTGATCACCCTCCCCGGAACCAGCCACGCTCGCTG
CTGCGGGCCGATCAGCCGCGCGCACTCCCTGCAACTAGGGACAACTCAGGCAAC
CACGCGCCTCACAAGCATGGCCGCCGTGGCATCCAACCCGCTCGTGACGGTGGG
TGCGCAAGTGCCAGGGGCCTCGTCGTCACGGCGTGCATCCTCGAGGGATGCGAT

[1388]

CCGGCAACTATATGTCGTTTATCTCCCCACCAATCACAGGATGAGCCCCTGTCTA
GA

[1389]

[1390] 서열 번호 123

[1391] 5' 프라이머 씨. 불가리스 3'UTR:하류 Clp 서열

[1392] 5’-ACTGCAATGCTGATGCACGGGA-3’

[1393] 서열 번호 124

[1394] 3' 프라이머 씨. 불가리스 3'UTR:하류 Clp 서열

[1395] 5’-TCCAGGTCCTTTTCGCACT-3’

[1396] 서열 번호 125

[1397] 코코믹사 C-169 THIC 아미노산 서열

MTTNLAKLPLGGLSSRSSIAGAPLRVNSHSRERHLGAKTTAIAAPERLDYLDNAEEA
RLQQTDAFAELKALSSRQSVNRPQKGELSFRQSPTFQDCFPGSEKCYREVEHDGKTL
KVPFRRVHLQDDNGHFDLYDTSGPQGVNPREGLPKIRSSWVEPREARGDKVQTQQY
YAKQGIITEEMAFCAARERMDPEFIRSEVARGRAIIPANKRHLELEPTVVGRNFLVKV
NANIGNSAVSSSIEEEVEKLQWSTIWGADTAMDLSTGNNIHETREWVMRNSPVPVGT
VPIYQCLEKAGGIVENITWELFRETLIEQAEQGVDYFTIHAGVLLRYIPLTANRVTGIV
SRGGSIHAKLCLLDHTENFAYMHWDEILDICAQYDITLSIGDGLRPGCIADANDAAQF
AELKTQGELTRRAWAKDVQVMNEGPGHVPLHKIPENMEKQLDWCSEAPFYTLGPL
ATDIAPAYDHITSAIGAATIGALGTALLCYVTPKEHLGLPDRDDVKAGVIAYKIAAHA
ADLAKGHPYAQEWDNALSKARFEFRWYDQFHLSLDPVTARLFHDATLPQEPAKTA
HFCSMCGPKFCSMQITQDLREYAKNHQMEEDEAIQTGMAEMSEQFKSIGAEVYLDE
AVLEHA

[1398]

[1399] 서열 번호 126

[1400] 코코믹사 C-169 THIC 천연 운반 펩티드 아미노산 서열

[1401] MTTNLAKLPLGGLSSRSSIAGAPLRVNSHSRERHLGAKTT

[1402] 서열 번호 127

[1403] 코코믹사 C-169 THIC 발현 카세트의 관련 서열

GCTCTTCGCCGCCGCCACTCCTGCTCGAGCGCGCCCGCGCGTGCGCCGCCAGCGC
CTTGGCCTTTTCGCCGCGCTCGTGCGCGTCGCTGATGTCCATCACCAGGTCCATG
AGGTCTGCCTTGCGCCGGCTGAGCCACTGCTTCGTCCGGGCGGCCAAGAGGAGC
ATGAGGGAGGACTCCTGGTCCAGGGTCCTGACGTGGTCGCGGCTCTGGGAGCGG
GCCAGCATCATCTGGCTCTGCCGCACCGAGGCCGCCTCCAACTGGTCCTCCAGCA
GCCGCAGTCGCCGCCGACCCTGGCAGAGGAAGACAGGTGAGGGGGGTATGAATT
GTACAGAACAACCACGAGCCTTGTCTAGGCAGAATCCCTACCAGTCATGGCTTTA
CCTGGATGACGGCCTGCGAACAGCTGTCCAGCGACCCTCGCTGCCGCCGCTTCTC
CCGCACGCTTCTTTCCAGCACCGTGATGGCGCGAGCCAGCGCCGCACGCTGGCGC
TGCGCTTCGCCGATCTGAGGACAGTCGGGGAACTCTGATCAGTCTAAACCCCCTT
GCGCGTTAGTGTTGCCATCCTTTGCAGACCGGTGAGAGCCGACTTGTTGTGCGCC
ACCCCCCACACCACCTCCTCCCAGACCAATTCTGTCACCTTTTTGGCGAAGGCAT
CGGCCTCGGCCTGCAGAGAGGACAGCAGTGCCCAGCCGCTGGGGGTTGGCGGAT

[1404]

GCACGCTCAGGTACCCTTTCTTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGG
GCTGCGAGACGGCTTCCCGGCGCTGCATGCAACACCGATGATGCTTCGACCCCCC
GAAGCTCCTTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGT
TTAAATAGCCAGGCCCCCGATTGCAAAGACATTATAGCGAGCTACCAAAGCCAT
ATTCAAACACCTAGATCACTACCACTTCTACACAGGCCACTCGAGCTTGTGATCG
CACTCCGCTAAGGGGGCGCCTCTTCCTCTTCGTTTCAGTCACAACCCGCAAACGG
CGCGCCATGCTGCTGCAGGCCTTCCTGTTCCTGCTGGCCGGCTTCGCCGCCAAGA
TCAGCGCCTCCATGACGAACGAGACGTCCGACCGCCCCCTGGTGCACTTCACCCC
CAACAAGGGCTGGATGAACGACCCCAACGGCCTGTGGTACGACGAGAAGGACGC
CAAGTGGCACCTGTACTTCCAGTACAACCCGAACGACACCGTCTGGGGGACGCC
CTTGTTCTGGGGCCACGCCACGTCCGACGACCTGACCAACTGGGAGGACCAGCC
CATCGCCATCGCCCCGAAGCGCAACGACTCCGGCGCCTTCTCCGGCTCCATGGTG
GTGGACTACAACAACACCTCCGGCTTCTTCAACGACACCATCGACCCGCGCCAGC
GCTGCGTGGCCATCTGGACCTACAACACCCCGGAGTCCGAGGAGCAGTACATCT
CCTACAGCCTGGACGGCGGCTACACCTTCACCGAGTACCAGAAGAACCCCGTGC
TGGCCGCCAACTCCACCCAGTTCCGCGACCCGAAGGTCTTCTGGTACGAGCCCTC
CCAGAAGTGGATCATGACCGCGGCCAAGTCCCAGGACTACAAGATCGAGATCTA
CTCCTCCGACGACCTGAAGTCCTGGAAGCTGGAGTCCGCGTTCGCCAACGAGGG
CTTCCTCGGCTACCAGTACGAGTGCCCCGGCCTGATCGAGGTCCCCACCGAGCAG
GACCCCAGCAAGTCCTACTGGGTGATGTTCATCTCCATCAACCCCGGCGCCCCGG
CCGGCGGCTCCTTCAACCAGTACTTCGTCGGCAGCTTCAACGGCACCCACTTCGA
GGCCTTCGACAACCAGTCCCGCGTGGTGGACTTCGGCAAGGACTACTACGCCCTG
CAGACCTTCTTCAACACCGACCCGACCTACGGGAGCGCCCTGGGCATCGCGTGG
GCCTCCAACTGGGAGTACTCCGCCTTCGTGCCCACCAACCCCTGGCGCTCCTCCA
TGTCCCTCGTGCGCAAGTTCTCCCTCAACACCGAGTACCAGGCCAACCCGGAGAC
GGAGCTGATCAACCTGAAGGCCGAGCCGATCCTGAACATCAGCAACGCCGGCCC
CTGGAGCCGGTTCGCCACCAACACCACGTTGACGAAGGCCAACAGCTACAACGT
CGACCTGTCCAACAGCACCGGCACCCTGGAGTTCGAGCTGGTGTACGCCGTCAAC
ACCACCCAGACGATCTCCAAGTCCGTGTTCGCGGACCTCTCCCTCTGGTTCAAGG
GCCTGGAGGACCCCGAGGAGTACCTCCGCATGGGCTTCGAGGTGTCCGCGTCCTC
CTTCTTCCTGGACCGCGGGAACAGCAAGGTGAAGTTCGTGAAGGAGAACCCCTA
CTTCACCAACCGCATGAGCGTGAACAACCAGCCCTTCAAGAGCGAGAACGACCT
GTCCTACTACAAGGTGTACGGCTTGCTGGACCAGAACATCCTGGAGCTGTACTTC
AACGACGGCGACGTCGTGTCCACCAACACCTACTTCATGACCACCGGGAACGCC
CTGGGCTCCGTGAACATGACGACGGGGGTGGACAACCTGTTCTACATCGACAAG
TTCCAGGTGCGCGAGGTCAAGTGACAATTGGCAGCAGCAGCTCGGATAGTATCG
ACACACTCTGGACGCTGGTCGTGTGATGGACTGTTGCCGCCACACTTGCTGCCTT
GACCTGTGAATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGT
GTGTACGCGCTTTTGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCC
AGCATCCCCTTCCCTCGTTTCATATCGCTTGCATCCCAACCGCAACTTATCTACGC
TGTCCTGCTATCCCTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCTCGCACAGCC
TTGGTTTGGGCTCCGCCTGTATTCTCCTGGTACTGCAACCTGTAAACCAGCACTGC
AATGCTGATGCACGGGAAGTAGTGGGATGGGAACACAAATGGAGGATCCCGCGT
CTCGAACAGAGCGCGCAGAGGAACGCTGAAGGTCTCGCCTCTGTCGCACCTCAG
CGCGGCATACACCACAATAACCACCTGACGAATGCGCTTGGTTCTTCGTCCATTA
GCGAAGCGTCCGGTTCACACACGTGCCACGTTGGCGAGGTGGCAGGTGACAATG
ATCGGTGGAGCTGATGGTCGAAACGTTCACAGCCTAGGGATATCGAATTCCTTTC
TTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTTCCCG
GCGCTGCATGCAACACCGATGATGCTTCGACCCCCCGAAGCTCCTTCGGGGCTGC
ATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGTTTAAATAGCCAGGCCCCC

[1405]

GATTGCAAAGACATTATAGCGAGCTACCAAAGCCATATTCAAACACCTAGATCA
CTACCACTTCTACACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGGGCG
CCTCTTCCTCTTCGTTTCAGTCACAACCCGCAAACACTAGTATGACCACCAACCT
GGCTAAGCTGCCTCTCGGCGGCCTCTCTTCTCGGTCGTCCATCGCCGGGGCTCCTC
TGCGCGTGAACAGCCACAGCCGCGAGCGCCACCTGGGCGCCAAGACCACCGCGA
TCGCCGCTCCCGAGCGGCTGGACTACCTCGACAACGCGGAGGAGGCGCGGCTGC
AGCAGACGGACGCCTTCGCCGAGCTGAAGGCGCTGTCGAGCCGCCAGTCCGTGA
ACCGGCCCCAAAAGGGCGAGCTGTCGTTCCGGCAGTCGCCGACCTTTCAGGACT
GCTTCCCTGGCTCGGAGAAGTGCTACCGCGAGGTGGAGCACGATGGCAAGACGC
TCAAGGTGCCTTTTCGCCGGGTGCACCTGCAGGACGACAACGGCCATTTCGACCT
GTACGACACCTCTGGCCCCCAGGGCGTGAACCCCCGCGAGGGTCTGCCCAAGAT
CCGCTCCTCCTGGGTGGAGCCGCGCGAGGCGCGGGGCGACAAGGTGCAAACGCA
GCAGTACTACGCCAAGCAAGGCATCATCACCGAGGAGATGGCCTTCTGCGCTGC
GCGCGAGCGCATGGACCCCGAGTTCATCCGGTCTGAGGTGGCCCGGGGTCGGGC
CATCATCCCTGCGAACAAGCGCCATCTGGAGCTGGAGCCGACCGTCGTGGGCCG
CAACTTTCTGGTCAAGGTGAACGCCAACATCGGCAACTCTGCGGTCAGCTCGAGC
ATCGAGGAGGAGGTGGAGAAGCTGCAGTGGAGCACCATCTGGGGTGCCGACACG
GCCATGGACCTGAGCACCGGCAACAACATCCACGAGACGCGCGAGTGGGTGATG
CGCAACAGCCCGGTCCCTGTGGGTACGGTCCCTATCTACCAGTGCCTGGAGAAGG
CGGGCGGCATCGTGGAGAACATCACCTGGGAGCTGTTCCGGGAGACGCTGATCG
AGCAGGCCGAGCAGGGCGTCGACTACTTTACCATCCACGCGGGGGTCCTGCTCC
GCTACATCCCCCTGACGGCGAACCGCGTCACCGGCATCGTCAGCCGCGGCGGCTC
TATCCATGCCAAGCTCTGCCTGCTGGACCACACCGAGAACTTCGCCTACATGCAC
TGGGACGAGATCCTGGACATCTGCGCCCAGTACGACATCACCCTGTCGATCGGCG
ACGGCCTGCGCCCCGGCTGCATCGCGGACGCCAACGACGCTGCCCAGTTTGCTGA
GCTGAAGACCCAGGGCGAGCTGACCCGCCGCGCCTGGGCGAAGGACGTGCAGGT
GATGAACGAGGGTCCCGGCCATGTGCCGCTGCACAAGATCCCCGAGAACATGGA
GAAGCAGCTGGACTGGTGCAGCGAGGCGCCCTTCTACACGCTCGGCCCCCTGGC
GACGGACATCGCGCCTGCGTACGACCACATCACCTCCGCGATCGGCGCCGCCAC
GATCGGCGCGCTGGGGACCGCGCTGCTGTGTTACGTGACCCCGAAGGAGCACCT
GGGCCTCCCCGATCGCGACGACGTGAAGGCCGGGGTGATCGCCTACAAGATCGC
GGCGCACGCGGCTGACCTGGCGAAGGGTCACCCCTACGCTCAGGAGTGGGATAA
CGCCCTCTCTAAGGCGCGCTTCGAGTTCCGGTGGTACGACCAGTTCCACCTGTCG
CTGGACCCCGTCACCGCCCGCCTGTTCCATGACGCGACCCTGCCGCAGGAGCCCG
CCAAGACCGCCCACTTCTGCTCCATGTGCGGCCCCAAGTTCTGCTCCATGCAGAT
CACCCAGGACCTCCGCGAGTACGCCAAGAACCATCAAATGGAGGAGGACGAGGC
GATCCAAACCGGCATGGCCGAGATGTCCGAGCAGTTCAAGAGCATCGGCGCGGA
GGTCTACCTCGACGAGGCGGTGCTGGAGCACGCGTAAGCAGCAGCAGCTCGGAT
AGTATCGACACACTCTGGACGCTGGTCGTGTGATGGACTGTTGCCGCCACACTTG
CTGCCTTGACCTGTGAATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTG
ATCTTGTGTGTACGCGCTTTTGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATAC
CACCCCCAGCATCCCCTTCCCTCGTTTCATATCGCTTGCATCCCAACCGCAACTTA
TCTACGCTGTCCTGCTATCCCTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCTCG
CACAGCCTTGGTTTGGGCTCCGCCTGTATTCTCCTGGTACTGCAACCTGTAAACC
AGCACTGCAATGCTGATGCACGGGAAGTAGTGGGATGGGAACACAAATGGAAA
GCTTGAGCT**C**TTGTTTTCCAGAAGGAGTTGCTCCTTGAGCCTTTCATTCTCAGCCT
CGATAACCTCCAAAGCCGCTCTAATTGTGGAGGGGGTTCGAATTTAAAAGCTTGG
AATGTTGGTTCGTGCGTCTGGAACAAGCCCAGACTTGTTGCTCACTGGGAAAAGG
ACCATCAGCTCCAAAAAACTTGCCGCTCAAACCGCGTACCTCTGCTTTCGCGCAA
TCTGCCCTGTTGAAATCGCCACCACATTCATATTGTGACGCTTGAGCAGTCTGTA

[1406]

ATTGCCTCAGAATGTGGAATCATCTGCCCCCTGTGCGAGCCCATGCCAGGCATGT
CGCGGGCGAGGACACCCGCCACTCGTACAGCAGACCATTATGCTACCTCACAAT
AGTTCATAACAGTGACCATATTTCTCGAAGCTCCCCAACGAGCACCTCCATGCTC
TGAGTGGCCACCCCCCGGCCCTGGTGCTTGCGGAGGGCAGGTCAACCGGCATGG
GGCTACCGAAATCCCCGACCGGATCCCACCACCCCCGCGATGGGAAGAATCTCT
CCCCGGGATGTGGGCCCACCACCAGCACAACCTGCTGGCCCAGGCGAGCGTCAA
ACCATACCACACAAATATCCTTGGCATCGGCCCTGAATTCCTTCTGCCGCTCTGCT
ACCCGGTGCTTCTGTCCGAAGCAGGGGTTGCTAGGGATCGCTCCGAGTCCGCAAA
CCCTTGTCGCGTGGCGGGGCTTGTTCGAGCTGAAGAGC

[1407]

[1408] 서열 번호 128

[1409] 코돈-최적화 코코믹사 C-169 THIC 서열

ATGACCACCAACCTGGCTAAGCTGCCTCTCGGCGGCCTCTCTTCTCGGTCGTCCA
TCGCCGGGGCTCCTCTGCGCGTGAACAGCCACAGCCGCGAGCGCCACCTGGGCG
CCAAGACCACCGCGATCGCCGCTCCCGAGCGGCTGGACTACCTCGACAACGCGG
AGGAGGCGCGGCTGCAGCAGACGGACGCCTTCGCCGAGCTGAAGGCGCTGTCGA
GCCGCCAGTCCGTGAACCGGCCCCAAAAGGGCGAGCTGTCGTTCCGGCAGTCGC
CGACCTTTCAGGACTGCTTCCCTGGCTCGGAGAAGTGCTACCGCGAGGTGGAGCA
CGATGGCAAGACGCTCAAGGTGCCTTTTCGCCGGGTGCACCTGCAGGACGACAA
CGGCCATTTCGACCTGTACGACACCTCTGGCCCCCAGGGCGTGAACCCCCGCGAG
GGTCTGCCCAAGATCCGCTCCTCCTGGGTGGAGCCGCGCGAGGCGCGGGGCGAC
AAGGTGCAAACGCAGCAGTACTACGCCAAGCAAGGCATCATCACCGAGGAGATG
GCCTTCTGCGCTGCGCGCGAGCGCATGGACCCCGAGTTCATCCGGTCTGAGGTGG
CCCGGGGGTCGGGCCATCATCCCTGCGAACAAGCGCCATCTGGAGCTGGAGCCGA
CCGTCGTGGGCCGCAACTTTCTGGTCAAGGTGAACGCCAACATCGGCAACTCTGC
GGTCAGCTCGAGCATCGAGGAGGAGGTGGAGAAGCTGCAGTGGAGCACCATCTG
GGGTGCCGACACGGCCATGGACCTGAGCACCGGCAACAACATCCACGAGACGCG
CGAGTGGGTGATGCGCAACAGCCCGGTCCCTGTGGGTACGGTCCCTATCTACCAG
TGCCTGGAGAAGGCGGGCGGCATCGTGGAGAACATCACCTGGGAGCTGTTCCGG
GAGACGCTGATCGAGCAGGCCGAGCAGGGCGTCGACTACTTTACCATCCACGCG
GGGGTCCTGCTCCGCTACATCCCCCTGACGGCGAACCGCGTCACCGGCATCGTCA
GCCGCGGCGGCTCTATCCATGCCAAGCTCTGCCTGCTGGACCACACCGAGAACTT
CGCCTACATGCACTGGGACGAGATCCTGGACATCTGCGCCCAGTACGACATCACC
CTGTCGATCGGCGACGGCCTGCGCCCCGGCTGCATCGCGGACGCCAACGACGCT
GCCCAGTTTGCTGAGCTGAAGACCCAGGGCGAGCTGACCCGCCGCGCCTGGGCG
AAGGACGTGCAGGTGATGAACGAGGGTCCCGGCCATGTGCCGCTGCACAAGATC
CCCGAGAACATGGAGAAGCAGCTGGACTGGTGCAGCGAGGCGCCCTTCTACACG
CTCGGCCCCCTGGCGACGGACATCGCGCCTGCGTACGACCACATCACCTCCGCGA
TCGGCGCCGCCACGATCGGCGCGCTGGGGACCGCGCTGCTGTGTTACGTGACCCC
GAAGGAGCACCTGGGCCTCCCCGATCGCGACGACGTGAAGGCCGGGGTGATCGC
CTACAAGATCGCGGCGCACGCGGCTGACCTGGCGAAGGGTCACCCCTACGCTCA
GGAGTGGGATAACGCCCTCTCTAAGGCGCGCTTCGAGTTCCGGTGGTACGACCA
GTTCCACCTGTCGCTGGACCCCGTCACCGCCCGCCTGTTCCATGACGCGACCCTG
CCGCAGGAGCCCGCCAAGACCGCCCACTTCTGCTCCATGTGCGGCCCCAAGTTCT
GCTCCATGCAGATCACCCAGGACCTCCGCGAGTACGCCAAGAACCATCAAATGG
AGGAGGACGAGGCGATCCAAACCGGCATGGCCGAGATGTCCGAGCAGTTCAAGA
GCATCGGCGCGGAGGTCTACCTCGACGAGGCGGTGCTGGAGCACGCGTAA

[1410]

[1411] 서열 번호 129

[1412] 코돈-최적화 에스. 세레비지아에 *suc2* 서열

GGCGCGCCATGCTGCTGCAGGCCTTCCTGTTCCTGCTGGCCGGCTTCGCCGCCAA
GATCAGCGCCTCCATGACGAACGAGACGTCCGACCGCCCCCTGGTGCACTTCACC
CCCAACAAGGGCTGGATGAACGACCCCAACGGCCTGTGGTACGACGAGAAGGAC
GCCAAGTGGCACCTGTACTTCCAGTACAACCCGAACGACACCGTCTGGGGGACG
CCCTTGTTCTGGGGCCACGCCACGTCCGACGACCTGACCAACTGGGAGGACCAG
CCCATCGCCATCGCCCCGAAGCGCAACGACTCCGGCGCCTTCTCCGGCTCCATGG
TGGTGGACTACAACAACACCTCCGGCTTCTTCAACGACACCATCGACCCGCGCCA
GCGCTGCGTGGCCATCTGGACCTACAACACCCCGGAGTCCGAGGAGCAGTACAT
CTCCTACAGCCTGGACGGCGGCTACACCTTCACCGAGTACCAGAAGAACCCCGT
GCTGGCCGCCAACTCCACCCAGTTCCGCGACCCGAAGGTCTTCTGGTACGAGCCC
TCCCAGAAGTGGATCATGACCGCGGCCAAGTCCCAGGACTACAAGATCGAGATC
TACTCCTCCGACGACCTGAAGTCCTGGAAGCTGGAGTCCGCGTTCGCCAACGAGG
GCTTCCTCGGCTACCAGTACGAGTGCCCCGGCCTGATCGAGGTCCCCACCGAGCA
GGACCCCAGCAAGTCCTACTGGGTGATGTTCATCTCCATCAACCCCGGCGCCCCG
GCCGGCGGCTCCTTCAACCAGTACTTCGTCGGCAGCTTCAACGGCACCCACTTCG
AGGCCTTCGACAACCAGTCCCGCGTGGTGGACTTCGGCAAGGACTACTACGCCCT
GCAGACCTTCTTCAACACCGACCCGACCTACGGGAGCGCCCTGGGCATCGCGTG
GGCCTCCAACTGGGAGTACTCCGCCTTCGTGCCCACCAACCCCTGGCGCTCCTCC
ATGTCCCTCGTGCGCAAGTTCTCCCTCAACACCGAGTACCAGGCCAACCCGGAGA
CGGAGCTGATCAACCTGAAGGCCGAGCCGATCCTGAACATCAGCAACGCCGGCC
CCTGGAGCCGGTTCGCCACCAACACCACGTTGACGAAGGCCAACAGCTACAACG
TCGACCTGTCCAACAGCACCGGCACCCTGGAGTTCGAGCTGGTGTACGCCGTCAA
CACCACCCAGACGATCTCCAAGTCCGTGTTCGCGGACCTCTCCCTCTGGTTCAAG
GGCCTGGAGGACCCCGAGGAGTACCTCCGCATGGGCTTCGAGGTGTCCGCGTCCT
CCTTCTTCCTGGACCGCGGGAACAGCAAGGTGAAGTTCGTGAAGGAGAACCCCT
ACTTCACCAACCGCATGAGCGTGAACAACCAGCCCTTCAAGAGCGAGAACGACC
TGTCCTACTACAAGGTGTACGGCTTGCTGGACCAGAACATCCTGGAGCTGTACTT
CAACGACGGCGACGTCGTGTCCACCAACACCTACTTCATGACCACCGGGAACGC
CCTGGGCTCCGTGAACATGACGACGGGGGTGGACAACCTGTTCTACATCGACAA
GTTCCAGGTGCGCGAGGTCAAGTGACAATTG

[1413]

[1414] 서열 번호 130

[1415] 씨. 프로토테코이데스 운반 펩티드를 지니는 코코믹사 C-169 THIC 발현 구조물의 관련 서열

GCGGCCGCGTGGACGAGGGCTACAACCCCGCCTACGGCGCGCGCCCGCTGCGCC
GCGCCATCATGCGCCTGCTGGAGGACGCGCTGGCCGAGCGCATGCTCGCCGGCG
ACGTCAAGGAGGGCGACTCGGTCATCATGGACGTCGATGGCGATGGCGCCATCA
GCGTCCTCAACGGAGACCGCACCCACACCACCACCATCGACTCCTCCCCGGCGG
GCATCTCGTAGACGCGAGAGGGAGGGAGGGAGGATTTTCAGAGGGAGATGAGA
CGAGGACTGGGCCGGGGGCCTTGGGGCGCTGCTGGAGCGTGGTGAGAGCGCGGC
GGACGTGCCTTTTCTTCTTCCGTGCGCGCGCTCTTGGCCATTGATCCCCGATTCGC
GCCCGCATCCCCCCACTGCCCCCATCATCTTGCCTGTTGTCGTGGCACTGACATA
AACCCCCTGCGCTGCGCTGCTCCGCTACTATTGATATAGGTCTCACGCGCCAATC
TTTTTTGCTCCGGGTAACCGTCTGGACGCCAGAATTCCTTTCTTGCGCTATGACAC
TTCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTTCCCGGCGCTGCATGCAAC
ACCGATGATGCTTCGACCCCCCGAAGCTCCTTCGGGGCTGCATGGGCGCTCCGAT
GCCGCTCCAGGGCGAGCGCTGTTTAAATAGCCAGGCCCCCGATTGCAAAGACAT

[1416]

TATAGCGAGCTACCAAAGCCATATTCAAACACCTAGATCACTACCACTTCTACAC
AGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGGGCGCCTCTTCCTCTTCGT
TTCAGTCACAACCCGCAAACACTAGTATGGCCACCGCATCCACTTTCTCGGCGTT
CAATGCCCGCTGCGGCGACCTGCGTCGCTCGGCGGGCTCCGGGCCCCGGCGCCC
AGCGAGGCCCCTCCCCGTGCGCGGGCGCGCCGCGATCGCCGCTCCCGAGCGGCT
GGACTACCTCGACAACGCGGAGGAGGCGCGGCTGCAGCAGACGGACGCCTTCGC
CGAGCTGAAGGCGCTGTCGAGCCGCCAGTCCGTGAACCGGCCCCAAAAGGGCGA
GCTGTCGTTCCGGCAGTCGCCGACCTTTCAGGACTGCTTCCCTGGCTCGGAGAAG
TGCTACCGCGAGGTGGAGCACGATGGCAAGACGCTCAAGGTGCCTTTTCGCCGG
GTGCACCTGCAGGACGACAACGGCCATTTCGACCTGTACGACACCTCTGGCCCCC
AGGGCGTGAACCCCCGCGAGGGTCTGCCCAAGATCCGCTCCTCCTGGGTGGAGC
CGCGCGAGGCGCGGGGCGACAAGGTGCAAACGCAGCAGTACTACGCCAAGCAA
GGCATCATCACCGAGGAGATGGCCTTCTGCGCTGCGCGCGAGCGCATGGACCCC
GAGTTCATCCGGTCTGAGGTGGCCCGGGGTCGGGCCATCATCCCTGCGAACAAG
CGCCATCTGGAGCTGGAGCCGACCGTCGTGGGCCGCAACTTTCTGGTCAAGGTGA
ACGCCAACATCGGCAACTCTGCGGTCAGCTCGAGCATCGAGGAGGAGGTGGAGA
AGCTGCAGTGGAGCACCATCTGGGGTGCCGACACGGCCATGGACCTGAGCACCG
GCAACAACATCCACGAGACGCGCGAGTGGGTGATGCGCAACAGCCCGGTCCCTG
TGGGTACGGTCCCTATCTACCAGTGCCTGGAGAAGGCGGGCGGCATCGTGGAGA
ACATCACCTGGGAGCTGTTCCGGGAGACGCTGATCGAGCAGGCCGAGCAGGGCG
TCGACTACTTTACCATCCACGCGGGGGTCCTGCTCCGCTACATCCCCCTGACGGC
GAACCGCGTCACCGGCATCGTCAGCCGCGGCGGCTCTATCCATGCCAAGCTCTGC
CTGCTGGACCACACCGAGAACTTCGCCTACATGCACTGGGACGAGATCCTGGAC
ATCTGCGCCCAGTACGACATCACCCTGTCGATCGGCGACGGCCTGCGCCCCGGCT
GCATCGCGGACGCCAACGACGCTGCCCAGTTTGCTGAGCTGAAGACCCAGGGCG
AGCTGACCCGCCGCGCCTGGGCGAAGGACGTGCAGGTGATGAACGAGGGTCCCG
GCCATGTGCCGCTGCACAAGATCCCCGAGAACATGGAGAAGCAGCTGGACTGGT
GCAGCGAGGCGCCCTTCTACACGCTCGGCCCCCTGGCGACGGACATCGCGCCTGC
GTACGACCACATCACCTCCGCGATCGGCGCCGCCACGATCGGCGCGCTGGGGAC
CGCGCTGCTGTGTTACGTGACCCCGAAGGAGCACCTGGGCCTCCCCGATCGCGAC
GACGTGAAGGCCGGGGTGATCGCCTACAAGATCGCGGCGCACGCGGCTGACCTG
GCGAAGGGTCACCCCTACGCTCAGGAGTGGGATAACGCCCTCTCTAAGGCGCGC
TTCGAGTTCCGGTGGTACGACCAGTTCCACCTGTCGCTGGACCCCGTCACCGCCC
GCCTGTTCCATGACGCGACCCTGCCGCAGGAGCCCGCCAAGACCGCCCACTTCTG
CTCCATGTGCGGCCCCAAGTTCTGCTCCATGCAGATCACCCAGGACCTCCGCGAG
TACGCCAAGAACCATCAAATGGAGGAGGACGAGGCGATCCAAACCGGCATGGCC
GAGATGTCCGAGCAGTTCAAGAGCATCGGCGCGGAGGTCTACCTCGACGAGGCG
GTGCTGGAGCACGCGTAACAATTGAAGCAGCAGCAGCTCGGATAGTATCGACAC
ACTCTGGACGCTGGTCGTGTGATGGACTGTTGCCGCCACACTTGCTGCCTTGACC
TGTGAATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGTGTGT
ACGCGCTTTTGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCCAGC
ATCCCCTTCCCTCGTTTCATATCGCTTGCATCCCAACCGCAACTTATCTACGCTGT
CCTGCTATCCCTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCTCGCACAGCCTTG
GTTTGGGCTCCGCCTGTATTCTCCTGGTACTGCAACCTGTAAACCAGCACTGCAA
TGCTGATGCACGGGAAGTAGTGGGATGGGAACACAAATGGAAAGCTTGAGCTCA
GAATAGTATCGGGTGATGCGAAGTCAGAACCAGGCAGGGCCTGTCGCCTGAGGT
GGCAACGATGGGAAGCAATCAATCTGGGTACAGTCGTCCGCACGATCCCGTGAT
CTCCCCCACCGACACCTATCCCCGCCCATCCCGGCCCACCCTTTCAGTCCCCTCAG
CATGCATTGTGCACCGCGACAAAGCATGTCTGCTCGTGCACTGGTTCAGGCCACG
GCGCACCGAGTCCTCGCCCTTCGCAGAGTGATCACCCTCCCCGGAACCAGCCACG

[1417]

CTCGCTGCTGCGGGCCGATCAGCCGCGCGCACTCCCTGCAACTAGGGACAACTCA
GGCAACCACGCGCCTCACAAGCATGGCCGCCGTGGCATCCAACCCGCTCGTGAC
GGTGGGTGCGCAAGTGCCAGGGGCCTCGTCGTCACGGCGTGCATCCTCGAGGGA
TGCGATCCGGCAACTATATGTCGTTTATCTCCCCACCAATCACAGGATGAGCCCC
TGTCTAGA

[1418]

[1419] 서열 번호 131

[1420] 씨. 프로토테코이데스 운반 펩티드를 지니는 코돈-최적화 코코믹사 C-169 THIC 서열

ACTAGTATGGCCACCGCATCCACTTTCTCGGCGTTCAATGCCCGCTGCGGCGACC
TGCGTCGCTCGGCGGGCTCCGGGCCCCGGCGCCCAGCGAGGCCCCTCCCCGTGCG
CGGGCGCGCCGCGATCGCCGCTCCCGAGCGGCTGGACTACCTCGACAACGCGGA
GGAGGCGCGGCTGCAGCAGACGGACGCCTTCGCCGAGCTGAAGGCGCTGTCGAG
CCGCCAGTCCGTGAACCGGCCCCAAAAGGGCGAGCTGTCGTTCCGGCAGTCGCC
GACCTTTCAGGACTGCTTCCCTGGCTCGGAGAAGTGCTACCGCGAGGTGGAGCAC
GATGGCAAGACGCTCAAGGTGCCTTTTCGCCGGGTGCACCTGCAGGACGACAAC
GGCCATTTCGACCTGTACGACACCTCTGGCCCCCAGGGCGTGAACCCCCGCGAGG
GTCTGCCCAAGATCCGCTCCTCCTGGGTGGAGCCGCGCGAGGCGCGGGGCGACA
AGGTGCAAACGCAGCAGTACTACGCCAAGCAAGGCATCATCACCGAGGAGATGG
CCTTCTGCGCTGCGCGCGAGCGCATGGACCCCGAGTTCATCCGGTCTGAGGTGGC
CCGGGGTCGGGCCATCATCCCTGCGAACAAGCGCCATCTGGAGCTGGAGCCGAC
CGTCGTGGGCCGCAACTTTCTGGTCAAGGTGAACGCCAACATCGGCAACTCTGCG
GTCAGCTCGAGCATCGAGGAGGAGGTGGAGAAGCTGCAGTGGAGCACCATCTGG
GGTGCCGACACGGCCATGGACCTGAGCACCGGCAACAACATCCACGAGACGCGC
GAGTGGGTGATGCGCAACAGCCCGGTCCCTGTGGGTACGGTCCCTATCTACCAGT
GCCTGGAGAAGGCGGGCGGCATCGTGGAGAACATCACCTGGGAGCTGTTCCGGG
AGACGCTGATCGAGCAGGCCGAGCAGGGCGTCGACTACTTTACCATCCACGCGG
GGGTCCTGCTCCGCTACATCCCCCTGACGGCGAACCGCGTCACCGGCATCGTCAG
CCGCGGCGGCTCTATCCATGCCAAGCTCTGCCTGCTGGACCACACCGAGAACTTC
GCCTACATGCACTGGGACGAGATCCTGGACATCTGCGCCCAGTACGACATCACCC
TGTCGATCGGCGACGGCCTGCGCCCCGGCTGCATCGCGGACGCCAACGACGCTG
CCCAGTTTGCTGAGCTGAAGACCCAGGGCGAGCTGACCCGCCGCGCCTGGGCGA
AGGACGTGCAGGTGATGAACGAGGGTCCCGGCCATGTGCCGCTGCACAAGATCC
CCGAGAACATGGAGAAGCAGCTGGACTGGTGCAGCGAGGCGCCCTTCTACACGC
TCGGCCCCCTGGCGACGGACATCGCGCCTGCGTACGACCACATCACCTCCGCGAT
CGGCGCCGCCACGATCGGCGCGCTGGGGACCGCGCTGCTGTTACGTGACCCC
GAAGGAGCACCTGGGCCTCCCCGATCGCGACGACGTGAAGGCCGGGGTGATCGC
CTACAAGATCGCGGCGCACGCGGCTGACCTGGCGAAGGGTCACCCCTACGCTCA
GGAGTGGGATAACGCCCTCTCTAAGGCGCGCTTCGAGTTCCGGTGGTACGACCA
GTTCCACCTGTCGCTGGACCCCGTCACCGCCCGCCTGTTCCATGACGCGACCCTG
CCGCAGGAGCCCGCCAAGACCGCCCACTTCTGCTCCATGTGCGGCCCCAAGTTCT
GCTCCATGCAGATCACCCAGGACCTCCGCGAGTACGCCAAGAACCATCAAATGG
AGGAGGACGAGGCGATCCAAACCGGCATGGCCGAGATGTCCGAGCAGTTCAAGA
GCATCGGCGCGGAGGTCTACCTCGACGAGGCGGTGCTGGAGCACGCGTAACAAT
TG

[1421]

[1422] 서열 번호 132

[1423] 클로렐라 프로토테코이데스 액틴 프로모터/5'UTR

ACTAGAGAGTTTAGGTCCAGCGTCCGTGGGGGGGGACGGGCTGGGAGCTTGGGC
CGGGAAGGGCAAGACGATGCAGTCCCTCTGGGGAGTCACAGCCGACTGTGTGTG
TTGCACTGTGCGGCCCGCAGCACTCACACGCAAAATGCCTGGCCGACAGGCAGG
CCCTGTCCAGTGCAACATCCACGGTCCCTCTCATCAGGCTCACCTTGCTCATTGAC
ATAACGGAATGCGTACCGCTCTTTCAGATCTGTCCATCCAGAGAGGGGAGCAGG
CTCCCCACCGACGCTGTCAAACTTGCTTCCTGCCCAACCGAAAACATTATTGTTT
GAGGGGGGGGGGGGGGGGGCAGATTGCATGGCGGGATATCTCGTGAGGAACAT
CACTGGGACACTGTGGAACACAGTGAGTGCAGTATGCAGAGCATGTATGCTAGG
GGTCAGCGCAGGAAGGGGGCCTTTCCCAGTCTCCCATGCCACTGCACCGTATCCA
CGACTCACCAGGACCAGCTTCTTGATCGGCTTCCGCTCCCGTGGACACCAGTGTG
TAGCCTCTGGACTCCAGGTATGCGTGCACCGCAAAGGCCAGCCGATCGTGCCGAT
TCCTGGGGTGGAGGATATGAGTCAGCCAACTTGGGGCTCAGAGTGCACACTGGG
GCACGATACGAAACAACATCTACACCGTGTCCTCCATGCTGACACACCACAGCTT
CGCTCCACCTGAATGTGGGCGCATGGGCCCGAATCACAGCCAATGTCGCTGCTGC
CATAATGTGATCCAGACCCTCTCCGCCCAGATGCCGAGCGGATCGTGGGCGCTGA
ATAGATTCCTGTTTCGATCACTGTTTGGGTCCTTTCCTTTTCGTCTCGGATGCGCG
TCTCGAAACAGGCTGCGTCGGGCTTTCGGATCCCTTTTGCTCCCTCCGTCACCATC
CTGCGCGCGGGCAAGTTGCTTGACCCTGGGCTGGTACCAGGGTTGGAGGGTATTA
CCGCGTCAGGCCATTCCCAGCCCGGATTCAATTCAAAGTCTGGGCCACCACCCTC
CGCCGCTCTGTCTGATCACTCCACATTCGTGCATACACTACGTTCAAGTCCTGATC
CAGGCGTGTCTCGGGACAAGGTGTGCTTGAGTTTGAATCTCAAGGACCCACTCCA
GCACAGCTGCTGGTTGACCCCGCCCTCGCAA

[1424]

[1425] 서열 번호 133

[1426] 클로렐라 프로토테코이데스 EF1A 3'UTR

[1427]
ACGGAGCGTCGTGCGGGAGGGAGTGTGCCGAGCGGGGAGTCCCGGTCTGTGCGA
GGCCCGGCAGCTGACGCTGGCGAGCCGTACGCCCCGAGGGTCCCCCTCCCCTGC
ACCCTCTTCCCCTTCCCTCTGACGGCCGCGCCTGTTCTTGCATGTTCAGCGAC

[1428] 서열 번호 134

[1429] 에이. 탈리아나 THIC 발현 카세트의 관련 서열

[1430]
GCTCTTCGCCGCCGCCACTCCTGCTCGAGCGCGCCCGCGCGTGCGCCGCCAGCGC
CTTGGCCTTTTCGCCGCGCTCGTGCGCGTCGCTGATGTCCATCACCAGGTCCATG
AGGTCTGCCTTGCGCCGGCTGAGCCACTGCTTCGTCCGGGCGGCCAAGAGGAGC
ATGAGGGAGGACTCCTGGTCCAGGGTCCTGACGTGGTCGCGGCTCTGGGAGCGG
GCCAGCATCATCTGGCTCTGCCGCACCGAGGCCGCCTCCAACTGGTCCTCCAGCA
GCCGCAGTCGCCGCCGACCCTGGCAGAGGAAGACAGGTGAGGGGGGTATGAATT
GTACAGAACAACCACGAGCCTTGTCTAGGCAGAATCCCTACCAGTCATGGCTTTA
CCTGGATGACGGCCTGCGAACAGCTGTCCAGCGACCCTCGCTGCCGCCGCTTCTC
CCGCACGCTTCTTTCCAGCACCGTGATGGCGCGAGCCAGCGCCGCACGCTGGCGC
TGCGCTTCGCCGATCTGAGGACAGTCGGGGAACTCTGATCAGTCTAAACCCCCTT
GCGCGTTAGTGTTGCCATCCTTTGCAGACCGGTGAGAGCCGACTTGTTGTGCGCC
ACCCCCCACACCACCTCCTCCCAGACCAATTCTGTCACCTTTTTGGCGAAGGCAT
CGGCCTCGGCCTGCAGAGAGGACAGCAGTGCCCAGCCGCTGGGGGTTGGCGGAT
GCACGCTCAGGTACCCTTTCTTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGG
GCTGCGAGACGGCTTCCCGGCGCTGCATGCAACACCGATGATGCTTCGACCCCCC
GAAGCTCCTTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGT
TTAAATAGCCAGGCCCCCGATTGCAAAGACATTATAGCGAGCTACCAAAGCCAT

ATTCAAACACCTAGATCACTACCACTTCTACACAGGCCACTCGAGCTTGTGATCG
CACTCCGCTAAGGGGGCGCCTCTTCCTCTTCGTTTCAGTCACAACCCGCAAACTC
TAGAATATCAATGATCGAGCAGGACGGCCTCCACGCCGGCTCCCCCGCCGCCTG
GGTGGAGCGCCTGTTCGGCTACGACTGGGCCCAGCAGACCATCGGCTGCTCCGA
CGCCGCCGTGTTCCGCCTGTCCGCCCAGGGCCGCCCCGTGCTGTTCGTGAAGACC
GACCTGTCCGGCGCCCTGAACGAGCTGCAGGACGAGGCCGCCCGCCTGTCCTGG
CTGGCCACCACCGGCGTGCCCTGCGCCGCCGTGCTGGACGTGGTGACCGAGGCC
GGCCGCGACTGGCTGCTGCTGGGCGAGGTGCCCGGCCAGGACCTGCTGTCCTCCC
ACCTGGCCCCCGCCGAGAAGGTGTCCATCATGGCCGACGCCATGCGCCGCCTGC
ACACCCTGGACCCCGCCACCTGCCCCTTCGACCACCAGGCCAAGCACCGCATCGA
GCGCGCCCGCACCCGCATGGAGGCCGGCCTGGTGGACCAGGACGACCTGGACGA
GGAGCACCAGGGCCTGGCCCCCGCCGAGCTGTTCGCCCGCCTGAAGGCCCGCAT
GCCCGACGGCGAGGACCTGGTGGTGACCCACGGCGACGCCTGCCTGCCCAACAT
CATGGTGGAGAACGGCCGCTTCTCCGGCTTCATCGACTGCGGCCGCCTGGGCGTG
GCCGACCGCTACCAGGACATCGCCCTGGCCACCCGCGACATCGCCGAGGAGCTG
GGCGGCGAGTGGGCCGACCGCTTCCTGGTGCTGTACGGCATCGCCGCCCCCGACT
CCCAGCGCATCGCCTTCTACCGCCTGCTGGACGAGTTCTTCTGACAATTGGCAGC
AGCAGCTCGGATAGTATCGACACACTCTGGACGCTGGTCGTGTGATGGACTGTTG
CCGCCACACTTGCTGCCTTGACCTGTGAATATCCCTGCCGCTTTTATCAAACAGCC
TCAGTGTGTTTGATCTTGTGTGTACGCGCTTTTGCGAGTTGCTAGCTGCTTGTGCT
ATTTGCGAATACCACCCCCAGCATCCCCTTCCCTCGTTTCATATCGCTTGCATCCC
AACCGCAACTTATCTACGCTGTCCTGCTATCCCTCAGCGCTGCTCCTGCTCCTGCT
CACTGCCCCTCGCACAGCCTTGGTTTGGGCTCCGCCTGTATTCTCCTGGTACTGCA
ACCTGTAAACCAGCACTGCAATGCTGATGCACGGGAAGTAGTGGGATGGGAACA
CAAATGGAGGATCCACTAGAGAGTTTAGGTCCAGCGTCCGTGGGGGGGGACGGG
CTGGGAGCTTGGGCCGGGAAGGGCAAGACGATGCAGTCCCTCTGGGGAGTCACA
GCCGACTGTGTGTGTTGCACTGTGCGGCCCGCAGCACTCACACGCAAAATGCCTG
GCCGACAGGCAGGCCCTGTCCAGTGCAACATCCACGGTCCCTCTCATCAGGCTCA
CCTTGCTCATTGACATAACGGAATGCGTACCGCTCTTTCAGATCTGTCCATCCAG
AGAGGGGAGCAGGCTCCCCACCGACGCTGTCAAACTTGCTTCCTGCCCAACCGA
AAACATTATTGTTTGAGGGGGGGGGGGGGGGGGCAGATTGCATGGCGGGATATC
TCGTGAGGAACATCACTGGGACACTGTGGAACACAGTGAGTGCAGTATGCAGAG
CATGTATGCTAGGGGTCAGCGCAGGAAGGGGGCCTTTCCCAGTCTCCCATGCCAC
TGCACCGTATCCACGACTCACCAGGACCAGCTTCTTGATCGGCTTCCGCTCCCGT
GGACACCAGTGTGTAGCCTCTGGACTCCAGGTATGCGTGCACCGCAAAGGCCAG
CCGATCGTGCCGATTCCTGGGGTGGAGGATATGAGTCAGCCAACTTGGGGCTCA
GAGTGCACACTGGGGCACGATACGAAACAACATCTACACCGTGTCCTCCATGCT
GACACACCACAGCTTCGCTCCACCTGAATGTGGGCGCATGGGCCCGAATCACAG
CCAATGTCGCTGCTGCCATAATGTGATCCAGACCCTCTCCGCCCAGATGCCGAGC
GGATCGTGGGCGCTGAATAGATTCCTGTTTCGATCACTGTTTGGGTCCTTTCCTTT
TCGTCTCGGATGCGCGTCTCGAAACAGGCTGCGTCGGGCTTTCGGATCCCTTTTG
CTCCCTCCGTCACCATCCTGCGCGCGGGCAAGTTGCTTGACCCTGGGCTGGTACC
AGGGTTGGAGGGTATTACCGCGTCAGGCCATTCCCAGCCCGGATTCAATTCAAAG
TCTGGGCCACCACCCTCCGCCGCTCTGTCTGATCACTCCACATTCGTGCATACACT
ACGTTCAAGTCCTGATCCAGGCGTGTCTCGGGACAAGGTGTGCTTGAGTTTGAAT
CTCAAGGACCCACTCCAGCACAGCTGCTGGTTGACCCCGCCCTCGCAAACTAGTA
TGGCCACCGCATCCACTTTCTCGGCGTTCAATGCCCGCTGCGGCGACCTGCGTCG
CTCGGCGGGCTCCGGGCCCCGGCGCCCAGCGAGGCCCCTCCCCGTGCGCGGGCG
CGCCGTCCAGGCCGCGGCCACCCGCTTCAAGAAGGAGACGACGACCACCCGCGC
CACGCTGACGTTCGACCCCCCCACGACCAACTCCGAGCGCGCCAAGCAGCGCAA

[1431]

GCACACCATCGACCCCTCCTCCCCCGACTTCCAGCCCATCCCCTCCTTCGAGGAG
TGCTTCCCCAAGTCCACGAAGGAGCACAAGGAGGTGGTGCACGAGGAGTCCGGC
CACGTCCTGAAGGTGCCCTTCCGCCGCGTGCACCTGTCCGGCGGCGAGCCCGCCT
TCGACAACTACGACACGTCCGGCCCCCAGAACGTCAACGCCCACATCGGCCTGG
CGAAGCTGCGCAAGGAGTGGATCGACCGCCGCGAGAAGCTGGGCACGCCCCGCT
ACACGCAGATGTACTACGCGAAGCAGGGCATCATCACGGAGGAGATGCTGTACT
GCGCGACGCGCGAGAAGCTGGACCCCGAGTTCGTCCGCTCCGAGGTCGCGCGGG
GCCGCGCCATCATCCCCTCCAACAAGAAGCACCTGGAGCTGGAGCCCATGATCG
TGGGCCGCAAGTTCCTGGTGAAGGTGAACGCGAACATCGGCAACTCCGCCGTGG
CCTCCTCCATCGAGGAGGAGGTCTACAAGGTGCAGTGGGCCACCATGTGGGGCG
CCGACACCATCATGGACCTGTCCACGGGCCGCCACATCCACGAGACGCGCGAGT
GGATCCTGCGCAACTCCGCGGTCCCCGTGGGCACCGTCCCCATCTACCAGGCGCT
GGAGAAGGTGGACGGCATCGCGGAGAACCTGAACTGGGAGGTGTTCCGCGAGAC
GCTGATCGAGCAGGCCGAGCAGGGCGTGGACTACTTCACGATCCACGCGGGCGT
GCTGCTGCGCTACATCCCCCTGACCGCCAAGCGCCTGACGGGCATCGTGTCCCGC
GGCGGCTCCATCCACGCGAAGTGGTGCCTGGCCTACCACAAGGAGAACTTCGCC
TACGAGCACTGGGACGACATCCTGGACATCTGCAACCAGTACGACGTCGCCCTGT
CCATCGGCGACGGCCTGCGCCCCGGCTCCATCTACGACGCCAACGACACGGCCC
AGTTCGCCGAGCTGCTGACCCAGGGCGAGCTGACGCGCCGCGCGTGGGAGAAGG
ACGTGCAGGTGATGAACGAGGGCCCCGGCCACGTGCCCATGCACAAGATCCCCG
AGAACATGCAGAAGCAGCTGGAGTGGTGCAACGAGGCGCCCTTCTACACCCTGG
GCCCCCTGACGACCGACATCGCGCCCGGCTACGACCACATCACCTCCGCCATCGG
CGCGGCCAACATCGGCGCCCTGGGCACCGCCCTGCTGTGCTACGTGACGCCCAA
GGAGCACCTGGGCCTGCCCAACCGCGACGACGTGAAGGCGGGCGTCATCGCCTA
CAAGATCGCCGCCCACGCGGCCGACCTGGCCAAGCAGCACCCCCACGCCCAGGC
GTGGGACGACGCGCTGTCCAAGGCGCGCTTCGAGTTCCGCTGGATGGACCAGTTC
GCGCTGTCCCTGGACCCCATGACGGCGATGTCCTTCCACGACGAGACGCTGCCCG
CGGACGGCGCGAAGGTCGCCCACTTCTGCTCCATGTGCGGCCCCAAGTTCTGCTC
CATGAAGATCACGGAGGACATCCGCAAGTACGCCGAGGAGAACGGCTACGGCTC
CGCCGAGGAGGCCATCCGCCAGGGCATGGACGCCATGTCCGAGGAGTTCAACAT
CGCCAAGAAGACGATCTCCGGCGAGCAGCACGGCGAGGTCGGCGGCGAGATCTA
CCTGCCCGAGTCCTACGTCAAGGCCGCGCAGAAGTGATACGTACTCGAGACGGA
GCGTCGTGCGGGAGGGAGTGTGCCGAGCGGGGAGTCCCGGTCTGTGCGAGGCCC
GGCAGCTGACGCTGGCGAGCCGTACGCCCCGAGGGTCCCCCTCCCCTGCACCCTC
TTCCCCTTCCCTCTGACGGCCGCGCCTGTTCTTGCATGTTCAGCGACGAGCTCTTG
TTTTCCAGAAGGAGTTGCTCCTTGAGCCTTTCATTCTCAGCCTCGATAACCTCCAA
AGCCGCTCTAATTGTGGAGGGGGTTCGAATTTAAAAGCTTGGAATGTTGGTTCGT
GCGTCTGGAACAAGCCCAGACTTGTTGCTCACTGGGAAAAGGACCATCAGCTCC
AAAAAACTTGCCGCTCAAACCGCGTACCTCTGCTTTCGCGCAATCTGCCCTGTTG
AAATCGCCACCACATTCATATTGTGACGCTTGAGCAGTCTGTAATTGCCTCAGAA
TGTGGAATCATCTGCCCCCTGTGCGAGCCCATGCCAGGCATGTCGCGGGCGAGG
ACACCCGCCACTCGTACAGCAGACCATTATGCTACCTCACAATAGTTCATAACAG
TGACCATATTTCTCGAAGCTCCCCAACGAGCACCTCCATGCTCTGAGTGGCCACC
CCCCGGCCCTGGTGCTTGCGGAGGGCAGGTCAACCGGCATGGGGCTACCGAAAT
CCCCGACCGGATCCCACCACCCCCGCGATGGGAAGAATCTCTCCCCGGGATGTG
GGCCCACCACCAGCACAACCTGCTGGCCCAGGCGAGCGTCAAACCATACCACAC
AAATATCCTTGGCATCGGCCCTGAATTCCTTCTGCCGCTCTGCTACCCGGTGCTTC
TGTCCGAAGCAGGGGTTGCTAGGGATCGCTCCGAGTCCGCAAACCCTTGTCGCGT
GGCGGGGCTTGTTCGAGCTTGAAGAGC

[1432]

[1433] 서열 번호 135

[1434] 씨. 프로토테코이데스 운반 펩티드를 지니는 코돈-최적화 에이.탈리아나 THIC

ATGGCCACCGCATCCACTTTCTCGGCGTTCAATGCCCGCTGCGGCGACCTGCGTC
GCTCGGCGGGCTCCGGGCCCCGGCGCCCAGCGAGGCCCCTCCCCGTGCGCGGGC
GCGCCGTCCAGGCCGCGGCCACCCGCTTCAAGAAGGAGACGACGACCACCCGCG
CCACGCTGACGTTCGACCCCCCCCACGACCAACTCCGAGCGCGCCAAGCAGCGCA
AGCACACCATCGACCCCTCCTCCCCCGACTTCCAGCCCATCCCCTCCTTCGAGGA
GTGCTTCCCCAAGTCCACGAAGGAGCACAAGGAGGTGGTGCACGAGGAGTCCGG
CCACGTCCTGAAGGTGCCCTTCCGCCGCGTGCACCTGTCCGGCGGCGAGCCCGCC
TTCGACAACTACGACACGTCCGGCCCCCAGAACGTCAACGCCCACATCGGCCTG
GCGAAGCTGCGCAAGGAGTGGATCGACCGCCGCGAGAAGCTGGGCACGCCCCGC
TACACGCAGATGTACTACGCGAAGCAGGGCATCATCACGGAGGAGATGCTGTAC
TGCGCGACGCGCGAGAAGCTGGACCCCGAGTTCGTCCGCTCCGAGGTCGCGCGG
GGCCGCGCCATCATCCCCTCCAACAAGAAGCACCTGGAGCTGGAGCCCATGATC
GTGGGCCGCAAGTTCCTGGTGAAGGTGAACGCGAACATCGGCAACTCCGCCGTG
GCCTCCTCCATCGAGGAGGAGGTCTACAAGGTGCAGTGGGCCACCATGTGGGGC
GCCGACACCATCATGGACCTGTCCACGGGCCGCCACATCCACGAGACGCGCGAG
TGGATCCTGCGCAACTCCGCGGTCCCCGTGGGCACCGTCCCCATCTACCAGGCGC
TGGAGAAGGTGGACGGCATCGCGGAGAACCTGAACTGGGAGGTGTTCCGCGAGA
CGCTGATCGAGCAGGCCGAGCAGGGCGTGGACTACTTCACGATCCACGCGGGCG
TGCTGCTGCGCTACATCCCCCTGACCGCCAAGCGCCTGACGGGCATCGTGTCCCG
CGGCGGCTCCATCCACGCGAAGTGGTGCCTGGCCTACCACAAGGAGAACTTCGC
CTACGAGCACTGGGACGACATCCTGGACATCTGCAACCAGTACGACGTCGCCCT
GTCCATCGGCGACGGCCTGCGCCCCGGCTCCATCTACGACGCCAACGACACGGC
CCAGTTCGCCGAGCTGCTGACCCAGGGCGAGCTGACGCGCCGCGCGTGGGAGAA
GGACGTGCAGGTGATGAACGAGGGCCCCGGCCACGTGCCCATGCACAAGATCCC
CGAGAACATGCAGAAGCAGCTGGAGTGGTGCAACGAGGCGCCCTTCTACACCCT
GGGCCCCCTGACGACCGACATCGCGCCCGGCTACGACCACATCACCTCCGCCATC
GGCGCGGCCAACATCGGCGCCCTGGGCACCGCCCTGCTGTGCTACGTGACGCCC
AAGGAGCACCTGGGCCTGCCCAACCGCGACGACGTGAAGGCGGGCGTCATCGCC
TACAAGATCGCCGCCCACGCGGCCGACCTGGCCAAGCAGCACCCCCACGCCCAG
GCGTGGGACGACGCGCTGTCCAAGGCGCGCTTCGAGTTCCGCTGGATGGACCAG
TTCGCGCTGTCCCTGGACCCCATGACGGCGATGTCCTTCCACGACGAGACGCTGC
CCGCGGACGGCGCGAAGGTCGCCCACTTCTGCTCCATGTGCGGCCCCAAGTTCTG
CTCCATGAAGATCACGGAGGACATCCGCAAGTACGCCGAGGAGAACGGCTACGG
CTCCGCCGAGGAGGCCATCCGCCAGGGCATGGACGCCATGTCCGAGGAGTTCAA
CATCGCCAAGAAGACGATCTCCGGCGAGCAGCACGGCGAGGTCGGCGGCGAGAT
CTACCTGCCCGAGTCCTACGTCAAGGCCGCGCAGAAGTGA

[1435]

[1436] 서열 번호 136

[1437] 에이.탈리아나 아미노산 서열 (천연 운반 펩티드 서열을 지님)

MAASVHCTLMSVVCNNKNHSARPKLPNSSLLPGFDVVVQAAATRFKKETTTTRATL
TFDPPTTNSERAKQRKHTIDPSSPDFQPIPSFEECFPKSTKEHKEVVHEESGHVLKVPF
RRVHLSGGEPAFDNYDTSGPQNVNAHIGLAKLRKEWIDRREKLGTPRYTQMYYAKQ
GIITEEMLYCATREKLDPEFVRSEVARGRAIIPSNKKHLELEPMIVGRKFLVKVNANIG
NSAVASSIEEEVYKVQWATMWGADTIMDLSTGRHIHETREWILRNSAVPVGTVPIYQ
ALEKVDGIAENLNWEVFRETLIEQAEQGVDYFTIHAGVLLRYIPLTAKRLTGIVSRGG
SIHAKWCLAYHKENFAYEHWDDILDICNQYDVALSIGDGLRPGSIYDANDTAQFAEL

[1438]

LTQGELTRRAWEKDVQVMNEGPGHVPMHKIPENMQKQLEWCNEAPFYTLGPLTTD
IAPGYDHITSAIGAANIGALGTALLCYVTPKEHLGLPNRDDVKAGVIAYKIAAHAAD
LAKQHPHAQAWDDALSKARFEFRWMDQFALSLDPMTAMSFHDETLPADGAKVAH
FCSMCGPKFCSMKITEDIRKYAEENGYGSAEEAIRQGMDAMSEEFNIAKKTISGEQH
GEVGGEIYLPESYVKAAQK

[1439]

[1440] 서열 번호 137

[1441] 시네코시스티스 종 *thiC* 발현 카세트의 관련 서열

GCTCTTCGCCGCCGCCACTCCTGCTCGAGCGCGCCCGCGCGTGCGCCGCCAGCGC
CTTGGCCTTTTCGCCGCGCTCGTGCGCGTCGCTGATGTCCATCACCAGGTCCATG
AGGTCTGCCTTGCGCCGGCTGAGCCACTGCTTCGTCCGGGCGGCCAAGAGGAGC
ATGAGGGAGGACTCCTGGTCCAGGGTCCTGACGTGGTCGCGGCTCTGGGAGCGG
GCCAGCATCATCTGGCTCTGCCGCACCGAGGCCGCCTCCAACTGGTCCTCCAGCA
GCCGCAGTCGCCGCCGACCCTGGCAGAGGAAGACAGGTGAGGGGGGTATGAATT
GTACAGAACAACCACGAGCCTTGTCTAGGCAGAATCCCTACCAGTCATGGCTTTA
CCTGGATGACGGCCTGCGAACAGCTGTCCAGCGACCCTCGCTGCCGCCGCTTCTC
CCGCACGCTTCTTTCCAGCACCGTGATGGCGCGAGCCAGCGCCGCACGCTGGCGC
TGCGCTTCGCCGATCTGAGGACAGTCGGGGAACTCTGATCAGTCTAAACCCCCTT
GCGCGTTAGTGTTGCCATCCTTTGCAGACCGGTGAGAGCCGACTTGTTGTGCGCC
ACCCCCCACACCACCTCCTCCCAGACCAATTCTGTCACCTTTTTGGCGAAGGCAT
CGGCCTCGGCCTGCAGAGAGGACAGCAGTGCCCAGCCGCTGGGGGTTGGCGGAT
GCACGCTCAGGTACCCTTTCTTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGG
GCTGCGAGACGGCTTCCCGGCGCTGCATGCAACACCGATGATGCTTCGACCCCCC
GAAGCTCCTTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGT
TTAAATAGCCAGGCCCCCGATTGCAAAGACATTATAGCGAGCTACCAAAGCCAT
ATTCAAACACCTAGATCACTACCACTTCTACACAGGCCACTCGAGCTTGTGATCG
CACTCCGCTAAGGGGGCGCCTCTTCCTCTTCGTTTCAGTCACAACCCGCAAACTC
TAGAATATCAATGATCGAGCAGGACGGCCTCCACGCCGGCTCCCCCGCCGCCTG
GGTGGAGCGCCTGTTCGGCTACGACTGGGCCCAGCAGACCATCGGCTGCTCCGA
CGCCGCCGTGTTCCGCCTGTCCGCCCAGGGCCGCCCCGTGCTGTTCGTGAAGACC
GACCTGTCCGGCGCCCTGAACGAGCTGCAGGACGAGGCCGCCCGCCTGTCCTGG
CTGGCCACCACCGGCGTGCCCTGCGCCGCCGTGCTGGACGTGGTGACCGAGGCC
GGCCGCGACTGGCTGCTGCTGGGCGAGGTGCCCGGCCAGGACCTGCTGTCCTCCC
ACCTGGCCCCCGCCGAGAAGGTGTCCATCATGGCCGACGCCATGCGCCGCCTGC
ACACCCTGGACCCCGCCACCTGCCCCTTCGACCACCAGGCCAAGCACCGCATCGA
GCGCGCCCGCACCCGCATGGAGGCCGGCCTGGTGGACCAGGACGACCTGGACGA
GGAGCACCAGGGCCTGGCCCCCGCCGAGCTGTTCGCCCGCCTGAAGGCCCGCAT
GCCCGACGGCGAGGACCTGGTGGTGACCCACGGCGACGCCTGCCTGCCCAACAT
CATGGTGGAGAACGGCCGCTTCTCCGGCTTCATCGACTGCGGCCGCCTGGGCGTG
GCCGACCGCTACCAGGACATCGCCCTGGCCACCCGCGACATCGCCGAGGAGCTG
GGCGGCGAGTGGGCCGACCGCTTCCTGGTGCTGTACGGCATCGCCGCCCCCGACT
CCCAGCGCATCGCCTTCTACCGCCTGCTGGACGAGTTCTTCTGACAATTGGCAGC
AGCAGCTCGGATAGTATCGACACACTCTGGACGCTGGTCGTGTGATGGACTGTTG
CCGCCACACTTGCTGCCTTGACCTGTGAATATCCCTGCCGCTTTTATCAAACAGCC
TCAGTGTGTTTGATCTTGTGTGTACGCGCTTTTGCGAGTTGCTAGCTGCTTGTGCT
ATTTGCGAATACCACCCCCAGCATCCCCTTCCCTCGTTTCATATCGCTTGCATCCC
AACCGCAACTTATCTACGCTGTCCTGCTATCCCTCAGCGCTGCTCCTGCTCCTGCT
CACTGCCCCTCGCACAGCCTTGGTTTGGGCTCCGCCTGTATTCTCCTGGTACTGCA
ACCTGTAAACCAGCACTGCAATGCTGATGCACGGGAAGTAGTGGGATGGGAACA

[1442]

CAAATGGAGGATCCACTAGAGAGTTTAGGTCCAGCGTCCGTGGGGGGGGACGGG
CTGGGAGCTTGGGCCGGGAAGGGCAAGACGATGCAGTCCCTCTGGGGAGTCACA
GCCGACTGTGTGTGTTGCACTGTGCGGCCCGCAGCACTCACACGCAAAATGCCTG
GCCGACAGGCAGGCCCTGTCCAGTGCAACATCCACGGTCCCTCTCATCAGGCTCA
CCTTGCTCATTGACATAACGGAATGCGTACCGCTCTTTCAGATCTGTCCATCCAG
AGAGGGGAGCAGGCTCCCCACCGACGCTGTCAAACTTGCTTCCTGCCCAACCGA
AAACATTATTGTTTGAGGGGGGGGGGGGGGGGGCAGATTGCATGGCGGGATATC
TCGTGAGGAACATCACTGGGACACTGTGGAACACAGTGAGTGCAGTATGCAGAG
CATGTATGCTAGGGGTCAGCGCAGGAAGGGGGCCTTTCCCAGTCTCCCATGCCAC
TGCACCGTATCCACGACTCACCAGGACCAGCTTCTTGATCGGCTTCCGCTCCCGT
GGACACCAGTGTGTAGCCTCTGGACTCCAGGTATGCGTGCACCGCAAAGGCCAG
CCGATCGTGCCGATTCCTGGGGTGGAGGATATGAGTCAGCCAACTTGGGGCTCA
GAGTGCACACTGGGGCACGATACGAAACAACATCTACACCGTGTCCTCCATGCT
GACACACCACAGCTTCGCTCCACCTGAATGTGGGCGCATGGGCCCGAATCACAG
CCAATGTCGCTGCTGCCATAATGTGATCCAGACCCTCTCCGCCCAGATGCCGAGC
GGATCGTGGGCGCTGAATAGATTCCTGTTTCGATCACTGTTTGGGTCCTTTCCTTT
TCGTCTCGGATGCGCGTCTCGAAACAGGCTGCGTCGGGCTTTCGGATCCCTTTTG
CTCCCTCCGTCACCATCCTGCGCGCGGGCAAGTTGCTTGACCCTGGGCTGGTACC
AGGGTTGGAGGGTATTACCGCGTCAGGCCATTCCCAGCCCGGATTCAATTCAAAG
TCTGGGCCACCACCCTCCGCCGCTCTGTCTGATCACTCCACATTCGTGCATACACT
ACGTTCAAGTCCTGATCCAGGCGTGTCTCGGGACAAGGTGTGCTTGAGTTTGAAT
CTCAAGGACCCACTCCAGCACAGCTGCTGGTTGACCCCGCCCTCGCAAACTAGTA
TGGCCACCGCATCCACTTTCTCGGCGTTCAATGCCCGCTGCGGCGACCTGCGTCG
CTCGGCGGGCTCCGGGCCCCGGCGCCCAGCGAGGCCCCTCCCCGTGCGCGGGCG
CGCCATGCGCACCGCGTGGGTCGCCAAGCGCCAGGGCCAGACGAACGTCTCCCA
GATGCACTACGCGCGCAAGGGCGTGATCACCGAGGAGATGGACTACGTGGCGAA
GCGCGAGAACCTGCCCGTGGAGCTGATCAAGGACGAGGTGGCGCGGGGCCGCAT
GATCATCCCCGCCAACATCAACCACACGAACCTGGAGCCCATGGCCATCGGCAT
CGCCTCCAAGTGCAAGGTCAACGCGAACATCGGCGCGTCCCCCAACTCCTCCAAC
ATCGACGAGGAGGTGGAGAAGCTGCTGCTGTCCGTGAAGTACGGCGCGGACACG
GTGATGGACCTGTCCACCGGCGGCGGCGACCTGGACGTGATCCGCACCGCCATC
ATCAACGCCTCCCCCGTGCCCATCGGCACGGTGCCCATCTACCAGGCCCTGGAGT
CCGTGCACGGCTCCATCGAGAACCTGACGCCCGACGACTTCCTGCACATCATCGA
GAAGCACGCCCAGCAGGGCGTGGACTACATGACGATCCACGCGGGCCTGCTGAT
CGAGTACCTGCCCCTGGTGAAGTCCCGCATCACGGGCATCGTGTCCCGCGGCGGC
GGCATCATCGCCAAGTGGATGCTGCACCACCACAAGCAGAACCCCCTGTACACG
CACTTCGACGAGATCATCGAGATCTTCAAGAAGTACGACGTGTCCTTCTCCCTGG
GCGACTCCCTGCGCCCCGGCTGCACCCACGACGCCTCCGACGACGCCCAGCTGTC
CGAGCTGAAGACGCTGGGCCAGCTGACGCGCCGCGCCTGGGAGCACGACGTCCA
GGTCATGGTGGAGGGCCCCGGCCACGTCCCCATCGACCAGATCGAGTTCAACGT
CAAGAAGCAGATGGAGGAGTGCTCCGAGGCCCCCTTCTACGTGCTGGGCCCCCT
GGTCACCGACATCGCGCCCGGCTACGACCACATCACCTCCGCGATCGGCGCCGCC
ATCGCGGGCTGGCACGGCACCGCGATGCTGTGCTACGTGACGCCCAAGGAGCAC
CTGGGCCTGCCCAACGCGGAGGACGTGCGCAACGGCCTGATCGCCTACAAGATC
GCCGCGCACGCCGCGGACATCGCCCGCCACCGCCCCGGCGCGCGCGACCGCGAC
GACGAGCTGTCCAAGGCGCGCTACAACTTCGACTGGAACCGCCAGTTCGAGCTG
TCCCTGGACCCCGAGCGCGCGAAGGAGTACCACGACGAGACGCTGCCCGCGGAC
ATCTACAAGACCGCGGAGTTCTGCTCCATGTGCGGCCCCAAGTTCTGCCCCATGC
AGACCAAGGTGGACGCCGAGATGCTGGAGGAGCTGGAGGTCTTCCTGGCCAAGG
ACAAGGAGATGGTGTCCCAGCGCTGATACGTACTCGAGACGGAGCGTCGTGCGG

[1443]

GAGGGAGTGTGCCGAGCGGGGAGTCCCGGTCTGTGCGAGGCCCGGCAGCTGACG
CTGGCGAGCCGTACGCCCCGAGGGTCCCCCTCCCCTGCACCCTCTTCCCCTTCCCT
CTGACGGCCGCGCCTGTTCTTGCATGTTCAGCGACGAGCTCTTGTTTTCCAGAAG
GAGTTGCTCCTTGAGCCTTTCATTCTCAGCCTCGATAACCTCCAAAGCCGCTCTAA
TTGTGGAGGGGGTTCGAATTTAAAAGCTTGGAATGTTGGTTCGTGCGTCTGGAAC
AAGCCCAGACTTGTTGCTCACTGGGAAAAGGACCATCAGCTCCAAAAAACTTGC
CGCTCAAACCGCGTACCTCTGCTTTCGCGCAATCTGCCCTGTTGAAATCGCCACC
ACATTCATATTGTGACGCTTGAGCAGTCTGTAATTGCCTCAGAATGTGGAATCAT
CTGCCCCCTGTGCGAGCCCATGCCAGGCATGTCGCGGGCGAGGACACCCGCCAC
TCGTACAGCAGACCATTATGCTACCTCACAATAGTTCATAACAGTGACCATATTT
CTCGAAGCTCCCCAACGAGCACCTCCATGCTCTGAGTGGCCACCCCCCGGCCCTG
GTGCTTGCGGAGGGCAGGTCAACCGGCATGGGGCTACCGAAATCCCCGACCGGA
TCCCACCACCCCCGCGATGGGAAGAATCTCTCCCCGGGATGTGGGCCCACCACCA
GCACAACCTGCTGGCCCAGGCGAGCGTCAAACCATACCACACAAATATCCTTGG
CATCGGCCCTGAATTCCTTCTGCCGCTCTGCTACCCGGTGCTTCTGTCCGAAGCAG
GGGTTGCTAGGGATCGCTCCGAGTCCGCAAACCCTTGTCGCGTGGCGGGGCTTGT
TCGAGCTTGAAGAGCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGGCGTAATC
ATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAAC
ATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAA
CTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGT
GCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTG
GGCGCTCTTC

[1444]

[1445] 서열 번호 138

[1446] 씨. 프로토테코이데스 운반 펩티드를 지니는 코돈-최적화 시네코시스티스 종 *thiC* 서열

[1447]
ATGGCCACCGCATCCACTTTCTCGGCGTTCAATGCCCGCTGCGGCGACCTGCGTC
GCTCGGCGGGCTCCGGGCCCCGGCGCCCAGCGAGGCCCCTCCCCGTGCGCGGGC
GCGCCATGCGCACCGCGTGGGTCGCCAAGCGCCAGGGCCAGACGAACGTCTCCC
AGATGCACTACGCGCGCAAGGGCGTGATCACCGAGGAGATGGACTACGTGGCGA
AGCGCGAGAACCTGCCCGTGGAGCTGATCAAGGACGAGGTGGCGCGGGGCCGCA
TGATCATCCCCGCCAACATCAACCACACGAACCTGGAGCCCATGGCCATCGGCAT
CGCCTCCAAGTGCAAGGTCAACGCGAACATCGGCGCGTCCCCCAACTCCTCCAAC
ATCGACGAGGAGGTGGAGAAGCTGCTGCTGTCCGTGAAGTACGGCGCGGACACG
GTGATGGACCTGTCCACCGGCGGCGGCGACCTGGACGTGATCCGCACCGCCATC
ATCAACGCCTCCCCCGTGCCCATCGGCACGGTGCCCATCTACCAGGCCCTGGAGT
CCGTGCACGGCTCCATCGAGAACCTGACGCCCGACGACTTCCTGCACATCATCGA
GAAGCACGCCCAGCAGGGCGTGGACTACATGACGATCCACGCGGGCCTGCTGAT
CGAGTACCTGCCCCTGGTGAAGTCCCGCATCACGGGCATCGTGTCCCGCGGCGGC
GGCATCATCGCCAAGTGGATGCTGCACCACCACAAGCAGAACCCCCTGTACACG
CACTTCGACGAGATCATCGAGATCTTCAAGAAGTACGACGTGTCCTTCTCCCTGG
GCGACTCCCTGCGCCCCGGCTGCACCCACGACGCCTCCGACGACGCCCAGCTGTC
CGAGCTGAAGACGCTGGGCCAGCTGACGCGCCGCGCCTGGGAGCACGACGTCCA
GGTCATGGTGGAGGGCCCCGGCCACGTCCCCATCGACCAGATCGAGTTCAACGT
CAAGAAGCAGATGGAGGAGTGCTCCGAGGCCCCCTTCTACGTGCTGGGCCCCCT
GGTCACCGACATCGCGCCCGGCTACGACCACATCACCTCCGCGATCGGCGCCGCC
ATCGCGGGCTGGCACGGCACCGCGATGCTGTGCTACGTGACGCCCAAGGAGCAC
CTGGGCCTGCCCAACGCGGAGGACGTGCGCAACGGCCTGATCGCCTACAAGATC
GCCGCGCACGCCGCGGACATCGCCCGCCACCGCCCCGGCGCGCGCGACCGCGAC
GACGAGCTGTCCAAGGCGCGCTACAACTTCGACTGGAACCGCCAGTTCGAGCTG

[1448]
TCCCTGGACCCCGAGCGCGCGAAGGAGTACCACGACGAGACGCTGCCCGCGGAC
ATCTACAAGACCGCGGAGTTCTGCTCCATGTGCGGCCCCAAGTTCTGCCCCATGC
AGACCAAGGTGGACGCCGAGATGCTGGAGGAGCTGGAGGTCTTCCTGGCCAAGG
ACAAGGAGATGGTGTCCCAGCGCTGA

[1449] 서열 번호 139

[1450] 시네코시스티스 종 *thiC* 아미노산 서열 (천연 운반 펩티드 서열을 지님)

[1451]
MRTAWVAKRQGQTNVSQMHYARKGVITEEMDYVAKRENLPVELIKDEVARGRMII
PANINHTNLEPMAIGIASKCKVNANIGASPNSSNIDEEVEKLLLSVKYGADTVMDLST
GGGDLDVIRTAIINASPVPIGTVPIYQALESVHGSIENLTPDDFLHIIEKHAQQGVDYM
TIHAGLLIEYLPLVKSRITGIVSRGGGIIAKWMLHHHKQNPLYTHFDEIIEIFKKYDVSF
SLGDSLRPGCTHDASDDAQLSELKTLGQLTRRAWEHDVQVMVEGPGHVPIDQIEFN
VKKQMEECSEAPFYVLGPLVTDIAPGYDHITSAIGAAIAGWHGTAMLCYVTPKEHLG
LPNAEDVRNGLIAYKIAAHAADIARHRPGARDRDDELSKARYNFDWNRQFELSLDP
ERAKEYHDETLPADIYKTAEFCSMCGPKFCPMQTKVDAEMLEELEVFLAKDKEMVS
QR

[1452] 서열 번호 140

[1453] 코돈-최적화 *neoR* 유전자

[1454]
CTTTCTTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTT
CCCGGCGCTGCATGCAACACCGATGATGCTTCGACCCCCCGAAGCTCCTTCGGGG
CTGCATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGTTTAAATAGCCAGGCC
CCCGATTGCAAAGACATTATAGCGAGCTACCAAAGCCATATTCAAACACCTAGA
TCACTACCACTTCTACACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGG
GCGCCTCTTCCTCTTCGTTTCAGTCACAACCCGCAAAC

[1455] 서열 번호 141

[1456] 5'THIC 코코믹사 C-169 확인 프라이머

[1457]
5’-GGGTGATCGCCTACAAGA-3’

[1458] 서열 번호 142

[1459] 3'THIC 확인 프라이머

[1460] 5'-ACGTCGCGACCCATGCTTCC-3'

[1461] 서열 번호 143

[1462] 5'THIC 에이. 탈리아나 확인 프라이머

[1463] 5'- GCGTCATCGCCTACAAGA-3'

[1464] 서열 번호 144

[1465] 5'thiC 시네코시스티스 종 확인 프라이머

[1466] 5'-CGATGCTGTGCTACGTGA-3'

[1467] 서열 번호 145

[1468] pSZ1318로부터의 미리스티카 프래그란스 광범위-기반 C14:0-C18:0 티오에스테라제의 코돈-최적화 코딩 영역

[1469]
ACTAGTATGGCTATCAAGACGAACAGGCAGCCTGTGGAGAAGCCTCCGTTCACG
ATCGGGACGCTGCGCAAGGCCATCCCCGCGCACTGTTTCGAGCGCTCGGCGCTTC
GTGGGCGCGCCGCCAACGCCCACACCGTGCCCAAGATCAACGGCAACAAGGCCG
GCCTGCTGACCCCCATGGAGAGCACCAAGGACGAGGACATCGTCGCGGCCCCCA
CCGTGGCGCCCAAGCGCACCTTCATCAACCAGCTGCCCGACTGGTCGATGCTGCT
GGCCGCGATCACCACCATCTTCCTGGCGGCCGAGAAGCAGTGGACCAACCTGGA
CTGGAAGCCCCGCCGCCCCGACATGCTGGTCGACTTCGACCCCTTCTCCCTGGGC
CGCTTCGTGCAGGACGGCCTGATCTTCCGCCAGAACTTCAGCATCCGCTCGTACG
AGATCGGCGCGGACCGCACCGCCTCCATCGAGACCCTGATGAACCACCTGCAGG
AGACCGCGCTGAACCACGTCCGCTGCATCGGCCTGCTGGACGACGGCTTCGGCA
GCACCCCCGAGATGACCCGCCGCGACCTGATCTGGGTGGTCACCCGCATGCAGG
TCCTGGTGGACCGCTACCCCTCGTGGGGCGACGTGATCGAGGTCGACTCCTGGGT
GACCCCCAGCGGCAAGAACGGCATGAAGCGCGAGTGGTTCCTGCGCGACTGCAA
GACCGGCGAGATCCTGACCCGCGCCACCTCGGTCTGGGTGATGATGAACAAGCG
CACCCGCCGCCTGTCCAAGATCCCCGAGGAGGTCCGCGTGGAGATCGAGCCCTA
CTTCGTCGAGCACGGCGTGCTGGACGAGGACTCGCGCAAGCTGCCCAAGCTGAA
CGACAACACCGCCAACTACATCCGCCGCGGCCTGGCGCCCCGCTGGTCCGACCTG
GACGTCAACCAGCACGTGAACAACGTCAAGTACATCGGCTGGATCCTGGAGAGC
GTGCCCAGCAGCCTGCTGGAGTCGCACGAGCTGTACGGCATGACCCTGGAGTAC
CGCAAGGAGTGCGGCAAGGACGGCCTGCTGCAGTCCCTGACCGCCGTCGCCAGC
GACTACGGCGGCGGCTCGCTGGAGGCCGGCGTGGAGTGCGACCACCTGCTGCGC
CTGGAGGACGGCTCCGAGATCATGCGCGGCAAGACCGAGTGGCGCCCCAAGCGC
GCCGCGAACACCACCTACTTCGGCAGCGTCGACGACATCCCCCCCGCCAACAAC
GCGTGACTCGAG

[1470] 서열 번호 146

[1471] 미리스티카 프래그란스 광범위-기반 C14:0-C18:0 티오에스테라제의 아미노산 서열 (수탁 번호 AAB71729)

[1472]
MVATSAASAFFPVASPSPVKPSMMLGGGGGSDNLDARGIKSRPASSGGLQVKANAH
TVPKINGNKAGLLTPMESTKDEDIVAAPTVAPKRTFINQLPDWSMLLAAITTIFLAAE
KQWTNLDWKPRRPDMLVDFDPFSLGRFVQDGLIFRQNFSIRSYEIGADRTASIETLM
NHLQETALNHVRCIGLLDDGFGSTPEMTRRDLIWVVTRMQVLVDRYPSWGDVIEVD
SWVTPSGKNGMKREWFLRDCKTGEILTRATSVWVMMNKRTRRLSKIPEEVRVEIEP
YFVEHGVLDEDSRKLPKLNDNTANYIRRGLAPRWSDLDVNQHVNNVKYIGWILESV
PSSLLESHELYGMTLEYRKECGKDGLLQSLTAVASDYGGGSLEAGVECDHLLRLED
GSEIMRGKTEWRPKRAANTTYFGSVDDIPPANNA

[1473] 서열 번호 147

[1474] pSZ1452로부터의 가르시니아 망고스타나 C16:0-선호 티오에스테라제의 코돈-최적화 코딩 영역

[1475]
ACTAGTATGGCCACCGCATCCACTTTCTCGGCGTTCAATGCCCGCTGCGGCGACC
TGCGTCGCTCGGCGGGCTCCGGGCCCCGGCGCCCAGCGAGGCCCCTCCCCGTGCG
CGGGCGCGCCGCCAACGCCCAGGCCCCCACCAAGATCAACGGCTCCACCGACGA
CGCCCAGCTGCCCGCCCCCCGCACCTTCATCAACCAGCTGCCCGACTGGTCCATG
CTGCTGGCCGCCATCACCACCGTGTTCCTGGCCGCCGAGAAGCAGTGGATGATGC

TGGACTGGAAGCCCCGCCGCCCCGACATGCTGATCGACACCTTCGGCCTGGGCCG
CATCGTGCAGGACGGCCTGGTGTTCCGCCAGAACTTCTCCATCCGCTCCTACGAG
ATCGGCGCCGACCGCACCGCCTCCATCGAGACCGTGATGAACCACCTGCAGGAG
ACCGCCCTGAACCACGTGAAGACCGCCGGCCTGCTGGGCGACGGCTTCGGCTCC
ACCCCCGAGATGTCCAAGCGCAACCTGATCTGGGTGGTGACCAAGATGCAGGTG
GAGGTGGACCGCTACCCCACCTGGGGCGACGTGGTGCAGGTGGACACCTGGGTG
TCCGCCTCCGGCAAGAACGGCATGCGCCGCGACTGGCTGCTGCGCGACGGCAAC
ACCGGCGAGACCCTGACCCGCGCCTCCTCCGTGTGGGTGATGATGAACAAGCTG
ACCCGCCGCCTGTCCAAGATCCCCGAGGAGGTGCGCGAGGAGATCGGCTCCTAC
TTCGTGAACTCCGACCCCGTGGTGGAGGAGGACGGCCGCAAGGTGACCAAGCTG
GACGACAACACCGCCGACTTCGTGCGCAAGGGCCTGACCCCCAAGTGGAACGAC
CTGGACATCAACCAGCACGTGAACAACGTGAAGTACATCGGCTGGATCCTGGAG
TCCGCCCCCCAGCCCATCCTGGAGACCCGCGAGCTGTCCGCCGTGACCCTGGAGT
ACCGCCGCGAGTGCGGCCGCGACTCCGTGCTGCGCTCCCTGACCGCCGTGTCCGG
CGGCGGCGTGGGCGACCTGGGCCACGCCGGCAACGTGGAGTGCCAGCACGTGCT
GCGCCTGGAGGACGGCGCCGAGATCGTGCGCGGCCGCACCGAGTGGCGCCCCAA
GTACATCAACAACTTCTCCATCATGGGCCAGATCCCCACCGACGCCTCCATGGAC
TACAAGGACCACGACGGCGACTACAAGGACCACGACATCGACTACAAGGACGAC
GACGACAAGTGAATCGAT

[1476]

[1477] 서열 번호 148

[1478] 가르시니아 망고스타나 C16:0-선호 티오에스테라제의 아미노산 서열 (수탁 번호 AAB51525.1)

MVATAATSSFFPLTSPSGDAKSGNPGKGSVSFGSMKSKSAASSRGLQVKANAQAPTK
INGSTDDAQLPAPRTFINQLPDWSMLLAAITTVFLAAEKQWMMLDWKPRRPDMLID
TFGLGRIVQDGLVFRQNFSIRSYEIGADRTASIETVMNHLQETALNHVKTAGLLGDGF
GSTPEMSKRNLIWVVTKMQVEVDRYPTWGDVVQVDTWVSASGKNGMRRDWLLR
DGNTGETLTRASSVWVMMNKLTRRLSKIPEEVREEIGSYFVNSDPVVEEDGRKVTKL
DDNTADFVRKGLTPKWNDLDINQHVNNVKYIGWILESAPQPILETRELSAVTLEYRR
ECGRDSVLRSLTAVSGGGVGDLGHAGNVECQHVLRLEDGAEIVRGRTEWRPKYINN
FSIMGQIPTDAS

[1479]

[1480] 서열 번호 149

[1481] pSZ1417로부터의 쿠페아 후커리아나 C16:0-선호 티오에스테라제의 코돈-최적화 코딩 영역

ACTAGTATGGTGGCCACCGCCGCCTCCTCCGCCTTCTTCCCCCTGCCCTCCGCCGA
CACCTCCTCCCGCCCCGGCAAGCTGGGCAACAAGCCCTCCTCCCTGTCCCCCCTG
AAGCCCAAGTCCACCCCCAACGGCGGCCTGCAGGTGAAGGCCAACGCCTCCGCC
CCCCCCAAGATCAACGGCTCCCCCGTGGGCCTGAAGTCCGGCGGCCTGAAGGGG
CGCGCCACCCAGGAGGACGCCCACTCCGCCCCCCCCCCCCGCACCTTCATCAACC
AGCTGCCCGACTGGTCCATGCTGCTGGCCGCCATCACCACCGTGTTCCTGGCCGC
CGAGAAGCAGTGGATGATGCTGGACTGGAAGCCCAAGCGCCCCGACATGCTGGT
GGACCCCTTCGGCCTGGGCTCCATCGTGCAGGACGGCCTGGTGTTCCGCCAGAAC
TTCTCCATCCGCTCCTACGAGATCGGCGCCGACCGCACCGCCTCCATCGAGACCG
TGATGAACCACCTGCAGGAGACCGCCCTGAACCACGTGAAGATCGCCGGCCTGT
CCAACGACGGCTTCGGCCGCACCCCCGAGATGTACAAGCGCGACCTGATCTGGG
TGGTGGCCAAGATGCAGGTGATGGTGAACCGCTACCCCACCTGGGGCGACACCG
TGGAGGTGAACACCTGGGTGGCCAAGTCCGGCAAGAACGGCATGCGCCGCGACT

[1482]

GGCTGATCTCCGACTGCAACACCGGCGAGATCCTGACCCGCGCCTCCTCCGTGTG
GGTGATGATGAACCAGAAGACCCGCCGCCTGTCCAAGATCCCCGACGAGGTGCG
CAACGAGATCGAGCCCCACTTCGTGGACTCCCCCCCCGTGATCGAGGACGACGA
CCGCAAGCTGCCCAAGCTGGACGAGAAGACCGCCGACTCCATCCGCAAGGGCCT
GACCCCCCGCTGGAACGACCTGGACGTGAACCAGCACGTGAACAACGTGAAGTA
CATCGGCTGGATCCTGGAGTCCACCCCCCCCGAGGTGCTGGAGACCCAGGAGCT
GTGCTCCCTGACCCTGGAGTACCGCCGCGAGTGCGGCCGCGAGTCCGTGCTGGA
GTCCCTGACCGCCATGGACCCCTCCGGCGGCGGCTACGGCTCCCAGTTCCAGCAC
CTGCTGCGCCTGGAGGACGGCGGCGAGATCGTGAAGGGCCGCACCGAGTGGCGC
CCCAAGAACGGCGTGATCAACGGCGTGGTGCCCACCGGCGAGTCCTCCCCCGGC
GACTACTCCATGGACTACAAGGACCACGACGGCGACTACAAGGACCACGACATC
GACTACAAGGACGACGACGACAAGTGAATCGAT

[1483]

[1484] 서열 번호 150

[1485] 쿠페아 후커리아나 C16:0-선호 티오에스테라제의 아미노산 서열 (수탁 번호 Q39513.1)

MVATAASSAFFPLPSADTSSRPGKLGNKPSSLSPLKPKSTPNGGLQVKANASAPPKIN
GSPVGLKSGGLKTQEDAHSAPPPRTFINQLPDWSMLLAAITTVFLAAEKQWMMLDW
KPKRPDMLVDPFGLGSIVQDGLVFRQNFSIRSYEIGADRTASIETVMNHLQETALNHV
KIAGLSNDGFGRTPEMYKRDLIWVVAKMQVMVNRYPTWGDTVEVNTWVAKSGKN
GMRRDWLISDCNTGEILTRASSVWVMMNQKTRRLSKIPDEVRNEIEPHFVDSPPVIED
DDRKLPKLDEKTADSIRKGLTPRWNDLDVNQHVNNVKYIGWILESTPPEVLETQELC
SLTLEYRRECGRESVLESLTAMDPSGGGYGSQFQHLLRLEDGGEIVKGRTEWRPKNG
VINGVVPTGESSPGDYS

[1486]

[1487] 서열 번호 151

[1488] pSZ1437로부터의 엘라에이스 기니엔시스 C16:0-선호 티오에스테라제의 코돈-최적화 코딩 영역

ACTAGTATGGTGGCCTCCATCGTGGCCTGGGCCTTCTTCCCCACCCCCTCCTTCTC
CCCCACCGGGCGCGCCGCCTCCGCCAAGGCCTCCAAGACCATCGGCGAGGGCTC
CGAGAACCTGAACGTGCGCGGCATCATCGCCAAGCCCACCTCCTCCTCCGCCGCC
AAGCAGGGCAAGGTGATGGCCCAGGCCGTGCCCAAGATCAACGGCGCCAAGGTG
GGCCTGAAGGCCGAGTCCCAGAAGGCCGAGGAGGACGCCGCCCCCTCCTCCGCC
CCCCGCACCTTCTACAACCAGCTGCCCGACTGGTCCGTGCTGCTGGCCGCCGTGA
CCACCATCTTCCTGGCCGCCGAGAAGCAGTGGACCCTGCTGGACTGGAAGCCCC
GCCGCCCCGACATGCTGACCGGCGCCTTCTCCCTGGGCAAGATCGTGCAGGACG
GCCTGGTGTTCCGCCAGAACTTCTCCATCCGCTCCTACGAGATCGGCGCCGACCG
CACCGCCTCCATCGAGACCCTGATGAACCACCTGCAGGAGACCGCCCTGAACCA
CGTGCGCAACGCCGGCCTGCTGGGCGACGGCTTCGGCGCCACCCCCGAGATGTC
CAAGCGCAACCTGATCTGGGTGGTGACCAAGATGCAGGTGCTGATCGAGCACTA
CCCCTCCTGGGGCGACGTGGTGGAGGTGGACACCTGGGTGGGCGCCTCCGGCAA
GAACGGCATGCGCCGCGACTGGCACGTGCGCGACTACCGCACCGGCCAGACCAT
CCTGCGCGCCACCTCCATCTGGGTGATGATGGACAAGCACACCCGCAAGCTGTCC
AAGATGCCCGAGGAGGTGCGCGCCGAGATCGGCCCCTACTTCATGGAGCACGCC
GCCATCGTGGACGAGGACTCCCGCAAGCTGCCCAAGCTGGACGACGACACCGCC
GACTACATCAAGTGGGGCCTGACCCCCCGCTGGTCCGACCTGGACGTGAACCAG
CACGTGAACAACGTGAAGTACATCGGCTGGATCCTGGAGTCCGCCCCCATCTCCA

[1489]

TCCTGGAGAACCACGAGCTGGCCTCCATGACCCTGGAGTACCGCCGCGAGTGCG
GCCGCGACTCCGTGCTGCAGTCCCTGACCGCCGTGGCCAACGACTGCACCGGCG
GCCTGCCCGAGGCCTCCATCGAGTGCCAGCACCTGCTGCAGCTGGAGTGCGGCG
CCGAGATCGTGCGCGGCCGCACCCAGTGGCGCCCCCGCCGCGCCTCCGGCCCCA
CCTCCGCCGGCTCCGCCATGGACTACAAGGACCACGACGGCGACTACAAGGACC
ACGACATCGACTACAAGGACGACGACGACAAGTGAATCGAT

[1490]

[1491] 서열 번호 152

[1492] 엘라에이스 기니엔시스 C16:0-선호 티오에스테라제의 아미노산 서열 (수탁 번호 AAD42220.2)

MVASIVAWAFFPTPSFSPTASAKASKTIGEGSENLNVRGIIAKPTSSSAAKQGKVMAQ
AVPKINGAKVGLKAESQKAEEDAAPSSAPRTFYNQLPDWSVLLAAVTTIFLAAEKQ
WTLLDWKPRRPDMLTGAFSLGKIVQDGLVFRQNFSIRSYEIGADRTASIETLMNHLQ
ETALNHVRNAGLLGDGFGATPEMSKRNLIWVVTKMQVLIEHYPSWGDVVEVDTWV
GASGKNGMRRDWHVRDYRTGQTILRATSIWVMMDKHTRKLSKMPEEVRAEIGPYF
MEHAAIVDEDSRKLPKLDDDTADYIKWGLTPRWSDLDVNQHVNNVKYIGWILESAP
ISILENHELASMTLEYRRECGRDSVLQSLTAVANDCTGGLPEASIECQHLLQLECGAEI
VRGRTQWRPRRASGPTSAGSA

[1493]

[1494] 서열 번호 153

[1495] pSZ1358로부터의 브라시카 나푸스 C18:0-선호 티오에스테라제의 코돈-최적화 코딩 영역

ACTAGTATGCTGAAGCTGTCCTGCAACGTGACCAACAACCTGCACACCTTCTCCT
TCTTCTCCGACTCCTCCCTGTTCATCCCCGTGAACCGCCGCACCATCGCCGTGTCC
TCCGGGCGCGCCTCCCAGCTGCGCAAGCCCGCCCTGGACCCCCTGCGCGCCGTGA
TCTCCGCCGACCAGGGCTCCATCTCCCCCGTGAACTCCTGCACCCCCGCCGACCG
CCTGCGCGCCGGCCGCCTGATGGAGGACGGCTACTCCTACAAGGAGAAGTTCAT
CGTGCGCTCCTACGAGGTGGGCATCAACAAGACCGCCACCGTGGAGACCATCGC
CAACCTGCTGCAGGAGGTGGCCTGCAACCACGTGCAGAAGTGCGGCTTCTCCAC
CGACGGCTTCGCCACCACCCTGACCATGCGCAAGCTGCACCTGATCTGGGTGACC
GCCCGCATGCACATCGAGATCTACAAGTACCCCGCCTGGTCCGACGTGGTGGAG
ATCGAGACCTGGTGCCAGTCCGAGGGCCGCATCGGCACCCGCCGCGACTGGATC
CTGCGCGACTCCGCCACCAACGAGGTGATCGGCCGCGCCACCTCCAAGTGGGTG
ATGATGAACCAGGACACCCGCCGCCTGCAGCGCGTGACCGACGAGGTGCGCGAC
GAGTACCTGGTGTTCTGCCCCCGCGAGCCCCGCCTGGCCTTCCCCGAGGAGAACA
ACTCCTCCCTGAAGAAGATCCCCAAGCTGGAGGACCCCGCCCAGTACTCCATGCT
GGAGCTGAAGCCCCGCCGCGCCGACCTGGACATGAACCAGCACGTGAACAACGT
GACCTACATCGGCTGGGTGCTGGAGTCCATCCCCCAGGAGATCATCGACACCCAC
GAGCTGCAGGTGATCACCCTGGACTACCGCCGCGAGTGCCAGCAGGACGACATC
GTGGACTCCCTGACCACCTCCGAGATCCCCGACGACCCCATCTCCAAGTTCACCG
GCACCAACGGCTCCGCCATGTCCTCCATCCAGGGCCACAACGAGTCCCAGTTCCT
GCACATGCTGCGCCTGTCCGAGAACGGCCAGGAGATCAACCGCGGCCGCACCCA
GTGGCGCAAGAAGTCCTCCCGCATGGACTACAAGGACCACGACGGCGACTACAA
GGACCACGACATCGACTACAAGGACGACGACGACAAGTGAATCGAT

[1496]

[1497] 서열 번호 154

[1498] 브라시카 나푸스 C18:0-선호 티오에스테라제의 아미노산 서열 (수탁 번호 CAA52070.1)

MLKLSCNVTNNLHTFSFFSDSSLFIPVNRRTIAVSSSQLRKPALDPLRAVISADQGSISP
VNSCTPADRLRAGRLMEDGYSYKEKFIVRSYEVGINKTATVETIANLLQEVACNHVQ
KCGFSTDGFATTLTMRKLHLIWVTARMHIEIYKYPAWSDVVEIETWCQSEGRIGTRR
DWILRDSATNEVIGRATSKWVMMNQDTRRLQRVTDEVRDEYLVFCPREPRLAFPEE
NNSSLKKIPKLEDPAQYSMLELKPRRADLDMNQHVNNVTYIGWVLESIPQEIIDTHEL
QVITLDYRRECQQDDIVDSLTTSEIPDDPISKFTGTNGSAMSSIQGHNESQFLHMLRLS
ENGQEINRGRTQWRKKSSR

[1499]

[1500] 서열 번호 155

[1501] pSZ1375로부터의 클로렐라 프로토테코이데스 UTEX 250 스테아로일-ACP 불포화 효소 운반 펩티드를 지니는 리시누스 커뮤니스 ACP-티오에스테라제의 코돈-최적화 코딩 서열

ACTAGTATGGCCACCGCATCCACTTTCTCGGCGTTCAATGCCCGCTGCGGCGACC
TGCGTCGCTCGGCGGGCTCCGGGCCCCGGCGCCCAGCGAGGCCCCTCCCCGTGCG
CGGGCGCGCCCGCTCCTACGAGGTGGGCATCAACAAGACCGCCACCGTGGAGAC
CATCGCCAACCTGCTGCAGGAGGTGGGCTGCAACCACGCCCAGTCCGTGGGCTTC
TCCACCGACGGCTTCGCCACCACCACCTCCATGCGCAAGATGCACCTGATCTGGG
TGACCGCCCGCATGCACATCGAGATCTACAAGTACCCCGCCTGGTCCGACGTGGT
GGAGGTGGAGACCTGGTGCCAGTCCGAGGGCCGCATCGGCACCCGCCGCGACTG
GATCCTGACCGACTACGCCACCGGCCAGATCATCGGCCGCGCCACCTCCAAGTG
GGTGATGATGAACCAGGACACCCGCCGCCTGCAGAAGGTGACCGACGACGTGCG
CGAGGAGTACCTGGTGTTCTGCCCCCGCGAGCTGCGCCTGGCCTTCCCCGAGGAG
AACAACCGCTCCTCCAAGAAGATCTCCAAGCTGGAGGACCCCGCCCAGTACTCC
AAGCTGGGCCTGGTGCCCCGCCGCGCCGACCTGGACATGAACCAGCACGTGAAC
AACGTGACCTACATCGGCTGGGTGCTGGAGTCCATCCCCCAGGAGATCATCGAC
ACCCACGAGCTGCAGACCATCACCCTGGACTACCGCCGCGAGTGCCAGCACGAC
GACATCGTGGACTCCCTGACCTCCGTGGAGCCCTCCGAGAACCTGGAGGCCGTGT
CCGAGCTGCGCGGCACCAACGGCTCCGCCACCACCACCGCCGGCGACGAGGACT
GCCGCAACTTCCTGCACCTGCTGCGCCTGTCCGGCGACGGCCTGGAGATCAACCG
CGGCCGCACCGAGTGGCGCAAGAAGTCCGCCCGCATGGACTACAAGGACCACGA
CGGCGACTACAAGGACCACGACATCGACTACAAGGACGACGACGACAAGTGAAT
CGAT

[1502]

[1503] 서열 번호 156

[1504] 리시누스 커뮤니스 ACP-티오에스테라제의 아미노산 서열 (수탁 번호 ABS30422.1)

MLKVPCCNATDPIQSLSSQCRFLTHFNNRPYFTRRPSIPTFFSSKNSSASLQAVVSDISS
VESAACDSLANRLRLGKLTEDGFSYKEKFIVRSYEVGINKTATVETIANLLQEVGCN
HAQSVGFSTDGFATTTSMRKMHLIWVTARMHIEIYKYPAWSDVVEVETWCQSEGRI
GTRRDWILTDYATGQIIGRATSKWVMMNQDTRRLQKVTDDVREEYLVFCPRELRLA
FPEENNRSSKKISKLEDPAQYSKLGLVPRRADLDMNQHVNNVTYIGWVLESIPQEIID
THELQTITLDYRRECQHDDIVDSLTSVEPSENLEAVSELRGTNGSATTTAGDEDCRNF
LHLLRLSGDGLEINRGRTEWRKKSAR

[1505]

[1506] 서열 번호 157

[1507] pSZ1317로부터의 클로렐라 프로토테코이데스 UTEX 250 스테아로일-ACP 운반 펩티드를 지니는 미리스티카 프래그란스 광범위-기반 티오에스테라제의 코돈-최적화 코딩 영역

ATGGCCACCGCATCCACTTTCTCGGCGTTCAATGCCCGCTGCGGCGACCTGCGTC
GCTCGGCGGGCTCCGGGCCCCGGCGCCCAGCGAGGCCCCTCCCCGTGCGCGGGC
GCGCCGCCAACGCCCACACCGTGCCCAAGATCAACGGCAACAAGGCCGGCCTGC
TGACCCCCATGGAGAGCACCAAGGACGAGGACATCGTCGCGGCCCCCACCGTGG
CGCCCAAGCGCACCTTCATCAACCAGCTGCCCGACTGGTCGATGCTGCTGGCCGC
GATCACCACCATCTTCCTGGCGGCCGAGAAGCAGTGGACCAACCTGGACTGGAA
GCCCCGCCGCCCCGACATGCTGGTCGACTTCGACCCCTTCTCCCTGGGCCGCTTC
GTGCAGGACGGCCTGATCTTCCGCCAGAACTTCAGCATCCGCTCGTACGAGATCG
GCGCGGACCGCACCGCCTCCATCGAGACCCTGATGAACCACCTGCAGGAGACCG
CGCTGAACCACGTCCGCTGCATCGGCCTGCTGGACGACGGCTTCGGCAGCACCCC
CGAGATGACCCGCCGCGACCTGATCTGGGTGGTCACCCGCATGCAGGTCCTGGTG
GACCGCTACCCCTCGTGGGGCGACGTGATCGAGGTCGACTCCTGGGTGACCCCCA
GCGGCAAGAACGGCATGAAGCGCGAGTGGTTCCTGCGCGACTGCAAGACCGGCG
AGATCCTGACCCGCGCCACCTCGGTCTGGGTGATGATGAACAAGCGCACCCGCC
GCCTGTCCAAGATCCCCGAGGAGGTCCGCGTGGAGATCGAGCCCTACTTCGTCGA
GCACGGCGTGCTGGACGAGGACTCGCGCAAGCTGCCCAAGCTGAACGACAACAC
CGCCAACTACATCCGCCGCGGCCTGGCGCCCCGCTGGTCCGACCTGGACGTCAAC
CAGCACGTGAACAACGTCAAGTACATCGGCTGGATCCTGGAGAGCGTGCCCAGC
AGCCTGCTGGAGTCGCACGAGCTGTACGGCATGACCCTGGAGTACCGCAAGGAG
TGCGGCAAGGACGGCCTGCTGCAGTCCCTGACCGCCGTCGCCAGCGACTACGGC
GGCGGCTCGCTGGAGGCCGGCGTGGAGTGCGACCACCTGCTGCGCCTGGAGGAC
GGCTCCGAGATCATGCGCGGCAAGACCGAGTGGCGCCCCAAGCGCGCCGCGAAC
ACCACCTACTTCGGCAGCGTCGACGACATCCCCCCCGCCAACAACGCGTGA

[1508]

[1509] 서열 번호 158

[1510] 클로렐라 프로토테코이데스 UTEX 250 스테아로일-ACP 운반 펩티드 서열을 지니는 미리스티카 프래그란스 광범위-기반 티오에스테라제의 아미노산 서열

MATASTFSAFNARCGDLRRSAGSGPRRPARPLPVRGRAANAHTVPKINGNKAGLLTP
MESTKDEDIVAAPTVAPKRTFINQLPDWSMLLAAITTIFLAAEKQWTNLDWKPRRPD
MLVDFDPFSLGRFVQDGLIFRQNFSIRSYEIGADRTASIETLMNHLQETALNHVRCIGL
LDDGFGSTPEMTRRDLIWVVTRMQVLVDRYPSWGDVIEVDSWVTPSGKNGMKREW
FLRDCKTGEILTRATSVWVMMNKRTRRLSKIPEEVRVEIEPYFVEHGVLDEDSRKLP
KLNDNTANYIRRGLAPRWSDLDVNQHVNNVKYIGWILESVPSSLLESHELYGMTLE
YRKECGKDGLLQSLTAVASDYGGGSLEAGVECDHLLRLEDGSEIMRGKTEWRPKRA
ANTTYFGSVDDIPPANNA

[1511]

[1512] 서열 번호 159

[1513] 씨. 레인하르드티 β-튜불린 프로모터/5'UTR 및 씨. 불가리스 질산염 환원 효소 3'UTR을 지니는 에스. 세레비지아에 *suc2* 유전자의 코돈-최적화 발현 카세트의 뉴클레오티드 서열

CTTTCTTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTT
CCCGGCGCTGCATGCAACACCGATGATGCTTCGACCCCCCGAAGCTCCTTCGGGG
CTGCATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGTTTAAATAGCCAGGCC
CCCGATTGCAAAGACATTATAGCGAGCTACCAAAGCCATATTCAAACACCTAGA
TCACTACCACTTCTACACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGG
GCGCCTCTTCCTCTTCGTTTCAGTCACAACCCGCAAACGGCGCGCCATGCTGCTG
CAGGCCTTCCTGTTCCTGCTGGCCGGCTTCGCCGCCAAGATCAGCGCCTCCATGA
CGAACGAGACGTCCGACCGCCCCCTGGTGCACTTCACCCCCAACAAGGGCTGGA
TGAACGACCCCAACGGCCTGTGGTACGACGAGAAGGACGCCAAGTGGCACCTGT

[1514]

ACTTCCAGTACAACCCGAACGACACCGTCTGGGGGACGCCCTTGTTCTGGGGCCA
CGCCACGTCCGACGACCTGACCAACTGGGAGGACCAGCCCATCGCCATCGCCCC
GAAGCGCAACGACTCCGGCGCCTTCTCCGGCTCCATGGTGGTGGACTACAACAA
CACCTCCGGCTTCTTCAACGACACCATCGACCCGCGCCAGCGCTGCGTGGCCATC
TGGACCTACAACACCCCGGAGTCCGAGGAGCAGTACATCTCCTACAGCCTGGAC
GGCGGCTACACCTTCACCGAGTACCAGAAGAACCCCGTGCTGGCCGCCAACTCC
ACCCAGTTCCGCGACCCGAAGGTCTTCTGGTACGAGCCCTCCCAGAAGTGGATCA
TGACCGCGGCCAAGTCCCAGGACTACAAGATCGAGATCTACTCCTCCGACGACCT
GAAGTCCTGGAAGCTGGAGTCCGCGTTCGCCAACGAGGGCTTCCTCGGCTACCA
GTACGAGTGCCCCGGCCTGATCGAGGTCCCCACCGAGCAGGACCCCAGCAAGTC
CTACTGGGTGATGTTCATCTCCATCAACCCCGGCGCCCCGGCCGGCGGCTCCTTC
AACCAGTACTTCGTCGGCAGCTTCAACGGCACCCACTTCGAGGCCTTCGACAACC
AGTCCCGCGTGGTGGACTTCGGCAAGGACTACTACGCCCTGCAGACCTTCTTCAA
CACCGACCCGACCTACGGGAGCGCCCTGGGCATCGCGTGGGCCTCCAACTGGGA
GTACTCCGCCTTCGTGCCCACCAACCCCTGGCGCTCCTCCATGTCCCTCGTGCGCA
AGTTCTCCCTCAACACCGAGTACCAGGCCAACCCGGAGACGGAGCTGATCAACC
TGAAGGCCGAGCCGATCCTGAACATCAGCAACGCCGGCCCCTGGAGCCGGTTCG
CCACCAACACCACGTTGACGAAGGCCAACAGCTACAACGTCGACCTGTCCAACA
GCACCGGCACCCTGGAGTTCGAGCTGGTGTACGCCGTCAACACCACCCAGACGA
TCTCCAAGTCCGTGTTCGCGGACCTCTCCCTCTGGTTCAAGGGCCTGGAGGACCC
CGAGGAGTACCTCCGCATGGGCTTCGAGGTGTCCGCGTCCTCCTTCTTCCTGGAC
CGCGGGAACAGCAAGGTGAAGTTCGTGAAGGAGAACCCCTACTTCACCAACCGC
ATGAGCGTGAACAACCAGCCCTTCAAGAGCGAGAACGACCTGTCCTACTACAAG
GTGTACGGCTTGCTGGACCAGAACATCCTGGAGCTGTACTTCAACGACGGCGAC
GTCGTGTCCACCAACACCTACTTCATGACCACCGGGAACGCCCTGGGCTCCGTGA
ACATGACGACGGGGGTGGACAACCTGTTCTACATCGACAAGTTCCAGGTGCGCG
AGGTCAAGTGACAATTGGCAGCAGCAGCTCGGATAGTATCGACACACTCTGGAC
GCTGGTCGTGTGATGGACTGTTGCCGCCACACTTGCTGCCTTGACCTGTGAATAT
CCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGTGTGTACGCGCTTT
TGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCCAGCATCCCCTTCC
CTCGTTTCATATCGCTTGCATCCCAACCGCAACTTATCTACGCTGTCCTGCTATCC
CTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCTCGCACAGCCTTGGTTTGGGCTC
CGCCTGTATTCTCCTGGTACTGCAACCTGTAAACCAGCACTGCAATGCTGATGCA
CGGGAAGTAGTGGGATGGGAACACAAATGGAGGATCC

[1515]

[1516] 서열 번호 160

[1517] pSZ1462로부터의 클로렐라 프로토테코이데스 UTEX 250 스테아로일-ACP 불포화 효소 운반 펩티드를 지니는 쿠페아 후커리아나 C16:선호 티오에스테라제의 코돈-최적화 코딩 영역

ACTAGTATGGCCACCGCATCCACTTTCTCGGCGTTCAATGCCCGCTGCGGCGACC
TGCGTCGCTCGGCGGGCTCCGGGCCCCGGCGCCCAGCGAGGCCCCTCCCCGTGCG
CGGGCGCGCCACCCAGGAGGACGCCCACTCCGCCCCCCCCCCCCGCACCTTCATC
AACCAGCTGCCCGACTGGTCCATGCTGCTGGCCGCCATCACCACCGTGTTCCTGG
CCGCCGAGAAGCAGTGGATGATGCTGGACTGGAAGCCCAAGCGCCCCGACATGC
TGGTGGACCCCTTCGGCCTGGGCTCCATCGTGCAGGACGGCCTGGTGTTCCGCCA
GAACTTCTCCATCCGCTCCTACGAGATCGGCGCCGACCGCACCGCCTCCATCGAG
ACCGTGATGAACCACCTGCAGGAGACCGCCCTGAACCACGTGAAGATCGCCGGC
CTGTCCAACGACGGCTTCGGCCGCACCCCCGAGATGTACAAGCGCGACCTGATCT
GGGTGGTGGCCAAGATGCAGGTGATGGTGAACCGCTACCCCACCTGGGGCGACA
CCGTGGAGGTGAACACCTGGGTGGCCAAGTCCGGCAAGAACGGCATGCGCCGCG

[1518]

ACTGGCTGATCTCCGACTGCAACACCGGCGAGATCCTGACCCGCGCCTCCTCCGT
GTGGGTGATGATGAACCAGAAGACCCGCCGCCTGTCCAAGATCCCCGACGAGGT
GCGCAACGAGATCGAGCCCCACTTCGTGGACTCCCCCCCCCGTGATCGAGGACGA
CGACCGCAAGCTGCCCAAGCTGGACGAGAAGACCGCCGACTCCATCCGCAAGGG
CCTGACCCCCCGCTGGAACGACCTGGACGTGAACCAGCACGTGAACAACGTGAA
GTACATCGGCTGGATCCTGGAGTCCACCCCCCCCGAGGTGCTGGAGACCCAGGA
GCTGTGCTCCCTGACCCTGGAGTACCGCCGCGAGTGCGGCCGCGAGTCCGTGCTG
GAGTCCCTGACCGCCATGGACCCCTCCGGCGGCGGCTACGGCTCCCAGTTCCAGC
ACCTGCTGCGCCTGGAGGACGGCGGCGAGATCGTGAAGGGCCGCACCGAGTGGC
GCCCCAAGAACGGCGTGATCAACGGCGTGGTGCCCACCGGCGAGTCCTCCCCCG
GCGACTACTCCATGGACTACAAGGACCACGACGGCGACTACAAGGACCACGACA
TCGACTACAAGGACGACGACGACAAGTGAATCGAT

[1519]

[1520] 서열 번호 161

[1521] pSZ1436으로부터의 엘라에이스 기니엔시스 C16:0-선호 티오에스테라제 PATE의 코돈-최적화 코딩 영역

ACTAGTATGGTGGCCTCCATCGCCGCCTCCGCCTTCTTCCCCACCCCCTCCTTCTC
CCCCACCGGGCGCGCCGCCTCCGCCAAGGCCTCCAAGACCATCGGCGAGGGCTC
CGAGAACCTGGACGCCCGCGGCATCATCGCCAAGCCCACCTCCTCCTCCGCCGCC
ATGCAGGGCAAGGTGATGGCCCAGGCCGTGTCCAAGATCAACGGCGCCAAGGTG
GGCCTGAAGGCCGAGTCCCAGAAGGCCGAGGAGGACGCCGCCCCCTCCTCCGCC
CCCCGCACCTTCTACAACCAGCTGCCCGACTGGTCCGTGCTGCTGGCCGCCGTGA
CCACCATCTTCCTGGCCGCCGAGAAGCAGTGGACCCTGCTGGACTGGAAGCCCC
GCCGCCCCGACATGCTGACCGACGCCTTCTCCCTGGGCAAGATCGTGCAGGACG
GCCTGGTGTTCAAGCAGAACTTCTCCATCCGCTCCTACGAGATCGGCGCCGACCG
CACCGCCTCCATCGAGACCCTGATGAACCACCTGCAGGAGACCGCCCTGAACCA
CGTGCGCTCCGCCGGCCTGATGGGCGACGGCTTCGGCGCCACCCCCGAGATGTCC
AAGCGCAACCTGATCTGGGTGGTGACCAAGGTGCGCGTGCTGATCGAGCACTAC
CCCTCCTGGGGCGACGTGGTGGAGGTGGACACCTGGGTGGGCCCCGCCGGCAAG
AACGGCATGCGCCGCGACTGGCACGTGCGCGACCACCGCACCGGCCAGACCATC
CTGCGCGCCACCCGCGTGTGGGTGATGATGAACAAGAACACCCGCAAGCTGTCC
AAGGTGCCCGAGGAGGTGCGCGCCGAGATCGGCCCCTACTTCGTGGAGCGCGCC
GCCATCGTGGACGAGGACTCCCGCAAGCTGCCCAAGCTGGACGAGGACACCACC
GACTACGTGAAGAAGGGCCTGACCCCCCGCTGGTCCGACCTGGACGTGAACCAG
CACGTGAACAACGTGAAGTACATCGGCTGGATCCTGGAGTCCGCCCCCATCTCCA
TCCTGGAGAACCACGAGCTGGCCTCCATGACCCTGGAGTACCGCCGCGAGTGCG
GCCGCGACTCCGTGCTGCAGTCCCTGACCGCCGTGGCCAACGACTGCACCGGCG
GCCTGCCCGAGGCCTCCATCGAGTGCCAGCACCTGCTGCAGCTGGAGTGCGGCG
CCGAGATCGTGCGCGGCCGCACCCAGTGGCGCCCCCGCCGCGCCTCCGGCCCCA
CCTCCGCCGGCTCCGCCATGGACTACAAGGACCACGACGGCGACTACAAGGACC
ACGACATCGACTACAAGGACGACGACGACAAGTGAATCGAT

[1522]

[1523] 서열 번호 162

[1524] 엘라에이스 기니엔시스 C16:0-선호 티오에스테라제 PATE의 아미노산 서열 (GenBank 수탁 번호 ABD83939)

MVASIAASAFFPTPSFSPTASAKASKTIGEGSENLDARGIIAKPTSSSAAMQGKVMAQ
AVSKINGAKVGLKAESQKAEEDAAPSSAPRTFYNQLPDWSVLLAAVTTIFLAAEKQ
WTLLDWKPRRPDMLTDAFSLGKIVQDGLVFKQNFSIRSYEIGADRTASIETLMNHLQ

[1525]

ETALNHVRSAGLMGDGFGATPEMSKRNLIWVVTKVRVLIEHYPSWGDVVEVDTWV
GPAGKNGMRRDWHVRDHRTGQTILRATRVWVMMNKNTRKLSKVPEEVRAEIGPYF
VERAAIVDEDSRKLPKLDEDTTDYVKKGLTPRWSDLDVNQHVNNVKYIGWILESAPI
SILENHELASMTLEYRRECGRDSVLQSLTAVANDCTGGLPEASIECQHLLQLECGAEI
VRGRTQWRPRRASG PTSAGSA

[1526]

[1527] 서열 번호 163

[1528] pSZ1468로부터의 RNAi 발현 카세트의 FADc 부분

ACTAGTATGGCTATCAAGACGAACAGGCAGCCTGTGGAGAAGCCTCCGTTCACG
ATCGGGACGCTGCGCAAGGCCATCCCCGCGCACTGTTTCGAGCGCTCGGCGCTTC
GTAGCAGCATGTACCTGGCCTTTGACATCGCGGTCATGTCCCTGCTCTACGTCGC
GTCGACGTACATCGACCCTGCACCGGTGCCTACGTGGGTCAAGTACGGCATCATG
TGGCCGCTCTACTGGTTCTTCCAGGTGTGTTTGAGGGTTTTGGTTGCCCGTATTGA
GGTCCTGGTGGCGCGCATGGAGGAGAAGGCGCCTGTCCCGCTGACCCCCCCGGC
TACCCTCCCGGCACCTTCCAGGGCGCGTACGGGAAGAACCAGTAGAGCGGCCAC
ATGATGCCGTACTTGACCCACGTAGGCACCGGTGCAGGGTCGATGTACGTCGAC
GCGACGTAGAGCAGGGACATGACCGCGATGTCAAAGGCCAGGTACATGCTGCTA
CGAAGCGCCGAGCGCTCGAAACAGTGCGCGGGGATGGCCTTGCGCAGCGTCCCG
ATCGTGAACGGAGGCTTCTCCACAGGCTGCCTGTTCGTCTTGATAGCCAT

[1529]

[1530] 서열 번호 164

[1531] pSZ1468로부터의 FADc RNAi 발현 카세트의 관련 부분

GCTCTTCGCCGCCGCCACTCCTGCTCGAGCGCGCCCGCGCGTGCGCCGCCAGCGC
CTTGGCCTTTTCGCCGCGCTCGTGCGCGTCGCTGATGTCCATCACCAGGTCCATG
AGGTCTGCCTTGCGCCGGCTGAGCCACTGCTTCGTCCGGGCGGCCAAGAGGAGC
ATGAGGGAGGACTCCTGGTCCAGGGTCCTGACGTGGTCGCGGCTCTGGGAGCGG
GCCAGCATCATCTGGCTCTGCCGCACCGAGGCCGCCTCCAACTGGTCCTCCAGCA
GCCGCAGTCGCCGCCGACCCTGGCAGAGGAAGACAGGTGAGGGGGGTATGAATT
GTACAGAACAACCACGAGCCTTGTCTAGGCAGAATCCCTACCAGTCATGGCTTTA
CCTGGATGACGGCCTGCGAACAGCTGTCCAGCGACCCTCGCTGCCGCCGCTTCTC
CCGCACGCTTCTTTCCAGCACCGTGATGGCGCGAGCCAGCGCCGCACGCTGGCGC
TGCGCTTCGCCGATCTGAGGACAGTCGGGGAACTCTGATCAGTCTAAACCCCCTT
GCGCGTTAGTGTTGCCATCCTTTGCAGACCGGTGAGAGCCGACTTGTTGTGCGCC
ACCCCCCACACCACCTCCTCCCAGACCAATTCTGTCACCTTTTTGGCGAAGGCAT
CGGCCTCGGCCTGCAGAGAGGACAGCAGTGCCCAGCCGCTGGGGGTTGGCGGAT
GCACGCTCAGGTACCCTTTCTTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGG
GCTGCGAGACGGCTTCCCGGCGCTGCATGCAACACCGATGATGCTTCGACCCCCC
GAAGCTCCTTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGT
TTAAATAGCCAGGCCCCCGATTGCAAAGACATTATAGCGAGCTACCAAAGCCAT
ATTCAAACACCTAGATCACTACCACTTCTACACAGGCCACTCGAGCTTGTGATCG
CACTCCGCTAAGGGGGCGCCTCTTCCTCTTCGTTTCAGTCACAACCCGCAAACGG
CGCGCCATGCTGCTGCAGGCCTTCCTGTTCCTGCTGGCCGGCTTCGCCGCCAAGA
TCAGCGCCTCCATGACGAACGAGACGTCCGACCGCCCCCTGGTGCACTTCACCCC
CAACAAGGGCTGGATGAACGACCCCAACGGCCTGTGGTACGACGAGAAGGACGC
CAAGTGGCACCTGTACTTCCAGTACAACCCGAACGACACCGTCTGGGGGACGCC
CTTGTTCTGGGGCCACGCCACGTCCGACGACCTGACCAACTGGGAGGACCAGCC
CATCGCCATCGCCCCGAAGCGCAACGACTCCGGCGCCTTCTCCGGCTCCATGGTG
GTGGACTACAACAACACCTCCGGCTTCTTCAACGACACCATCGACCCGCGCCAGC

[1532]

GCTGCGTGGCCATCTGGACCTACAACACCCCGGAGTCCGAGGAGCAGTACATCT
CCTACAGCCTGGACGGCGGCTACACCTTCACCGAGTACCAGAAGAACCCCGTGC
TGGCCGCCAACTCCACCCAGTTCCGCGACCCGAAGGTCTTCTGGTACGAGCCCTC
CCAGAAGTGGATCATGACCGCGGCCAAGTCCCAGGACTACAAGATCGAGATCTA
CTCCTCCGACGACCTGAAGTCCTGGAAGCTGGAGTCCGCGTTCGCCAACGAGGG
CTTCCTCGGCTACCAGTACGAGTGCCCCGGCCTGATCGAGGTCCCCACCGAGCAG
GACCCCAGCAAGTCCTACTGGGTGATGTTCATCTCCATCAACCCCGGCGCCCCGG
CCGGCGGCTCCTTCAACCAGTACTTCGTCGGCAGCTTCAACGGCACCCACTTCGA
GGCCTTCGACAACCAGTCCCGCGTGGTGGACTTCGGCAAGGACTACTACGCCCTG
CAGACCTTCTTCAACACCGACCCGACCTACGGGAGCGCCCTGGGCATCGCGTGG
GCCTCCAACTGGGAGTACTCCGCCTTCGTGCCCACCAACCCCTGGCGCTCCTCCA
TGTCCCTCGTGCGCAAGTTCTCCCTCAACACCGAGTACCAGGCCAACCCGGAGAC
GGAGCTGATCAACCTGAAGGCCGAGCCGATCCTGAACATCAGCAACGCCGGCCC
CTGGAGCCGGTTCGCCACCAACACCACGTTGACGAAGGCCAACAGCTACAACGT
CGACCTGTCCAACAGCACCGGCACCCTGGAGTTCGAGCTGGTGTACGCCGTCAAC
ACCACCCAGACGATCTCCAAGTCCGTGTTCGCGGACCTCTCCCTCTGGTTCAAGG
GCCTGGAGGACCCCGAGGAGTACCTCCGCATGGGCTTCGAGGTGTCCGCGTCCTC
CTTCTTCCTGGACCGCGGGAACAGCAAGGTGAAGTTCGTGAAGGAGAACCCCTA
CTTCACCAACCGCATGAGCGTGAACAACCAGCCCTTCAAGAGCGAGAACGACCT
GTCCTACTACAAGGTGTACGGCTTGCTGGACCAGAACATCCTGGAGCTGTACTTC
AACGACGGCGACGTCGTGTCCACCAACACCTACTTCATGACCACCGGGAACGCC
CTGGGCTCCGTGAACATGACGACGGGGGTGGACAACCTGTTCTACATCGACAAG
TTCCAGGTGCGCGAGGTCAAGTGACAATTGGCAGCAGCAGCTCGGATAGTATCG
ACACACTCTGGACGCTGGTCGTGTGATGGACTGTTGCCGCCACACTTGCTGCCTT
GACCTGTGAATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGT
GTGTACGCGCTTTTGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCC
AGCATCCCCTTCCCTCGTTTCATATCGCTTGCATCCCAACCGCAACTTATCTACGC
TGTCCTGCTATCCCTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCTCGCACAGCC
TTGGTTTGGGCTCCGCCTGTATTCTCCTGGTACTGCAACCTGTAAACCAGCACTGC
AATGCTGATGCACGGGAAGTAGTGGGATGGGAACACAAATGGAGGATCCCGCGT
CTCGAACAGAGCGCGCAGAGGAACGCTGAAGGTCTCGCCTCTGTCGCACCTCAG
CGCGGCATACACCACAATAACCACCTGACGAATGCGCTTGGTTCTTCGTCCATTA
GCGAAGCGTCCGGTTCACACACGTGCCACGTTGGCGAGGTGGCAGGTGACAATG
ATCGGTGGAGCTGATGGTCGAAACGTTCACAGCCTAGGGATATCGAATTCGGCC
GACAGGACGCGCGTCAAAGGTGCTGGTCGTGTATGCCCTGGCCGGCAGGTCGTT
GCTGCTGCTGGTTAGTGATTCCGCAACCCTGATTTTGGCGTCTTATTTTGGCGTGG
CAAACGCTGGCGCCCGCGAGCCGGGCCGGCGGCGATGCGGTGCCCCACGGCTGC
CGGAATCCAAGGGAGGCAAGAGCGCCCGGGTCAGTTGAAGGGCTTTACGCGCAA
GGTACAGCCGCTCCTGCAAGGCTGCGTGGTGGAATTGGACGTGCAGGTCCTGCTG
AAGTTCCTCCACCGCCTCACCAGCGGACAAAGCACCGGTGTATCAGGTCCGTGTC
ATCCACTCTAAAGAGCTCGACTACGACCTACTGATGGCCCTAGATTCTTCATCAA
AAACGCCTGAGACACTTGCCCAGGATTGAAACTCCCTGAAGGGACCACCAGGGG
CCCTGAGTTGTTCCTTCCCCCCGTGGCGAGCTGCCAGCCAGGCTGTACCTGTGAT
CGAGGCTGGCGGGAAAATAGGCTTCGTGTGCTCAGGTCATGGGAGGTGCAGGAC
AGCTCATGAAACGCCAACAATCGCACAATTCATGTCAAGCTAATCAGCTATTTCC
TCTTCACGAGCTGTAATTGTCCCAAAATTCTGGTCTACCGGGGGTGATCCTTCGT
GTACGGGCCCTTCCCTCAACCCTAGGTATGCGCGCATGCGGTCGCCGCGCAACTC
GCGCGAGGGCCGAGGGTTTGGGACGGGCCGTCCCGAAATGCAGTTGCACCCGGA
TGCGTGGCACCTTTTTTGCGATAATTTATGCAATGGACTGCTCTGCAAAATTCTGG
CTCTGTCGCCAACCCTAGGATCAGCGGCGTAGGATTTCGTAATCATTCGTCCTGA

[1533]

TGGGGAGCTACCGACTACCCTAATATCAGCCCGACTGCCTGACGCCAGCGTCCAC
TTTTGTGCACACATTCCATTCGTGCCCAAGACATTTCATTGTGGTGCGAAGCGTCC
CCAGTTACGCTCACCTGTTTCCCGACCTCCTTACTGTTCTGTCGACAGAGCGGGCC
CACAGGCCGGTCGCAGCCACTAGTATGGCTATCAAGACGAACAGGCAGCCTGTG
GAGAAGCCTCCGTTCACGATCGGGACGCTGCGCAAGGCCATCCCCGCGCACTGTT
TCGAGCGCTCGGCGCTTCGTAGCAGCATGTACCTGGCCTTTGACATCGCGGTCAT
GTCCCTGCTCTACGTCGCGTCGACGTACATCGACCCTGCACCGGTGCCTACGTGG
GTCAAGTACGGCATCATGTGGCCGCTCTACTGGTTCTTCCAGGTGTGTTTGAGGG
TTTTGGTTGCCCGTATTGAGGTCCTGGTGGCGCGCATGGAGGAGAAGGCGCCTGT
CCCGCTGACCCCCCCGGCTACCCTCCCGGCACCTTCCAGGGCGCGTACGGGAAGA
ACCAGTAGAGCGGCCACATGATGCCGTACTTGACCCACGTAGGCACCGGTGCAG
GGTCGATGTACGTCGACGCGACGTAGAGCAGGGACATGACCGCGATGTCAAAGG
CCAGGTACATGCTGCTACGAAGCGCCGAGCGCTCGAAACAGTGCGCGGGGATGG
CCTTGCGCAGCGTCCCGATCGTGAACGGAGGCTTCTCCACAGGCTGCCTGTTCGT
CTTGATAGCCATCTCGAGGCAGCAGCAGCTCGGATAGTATCGACACACTCTGGAC
GCTGGTCGTGTGATGGACTGTTGCCGCCACACTTGCTGCCTTGACCTGTGAATAT
CCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGTGTGTACGCGCTTT
TGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCCAGCATCCCCTTCC
CTCGTTTCATATCGCTTGCATCCCAACCGCAACTTATCTACGCTGTCCTGCTATCC
CTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCTCGCACAGCCTTGGTTTGGGCTC
CGCCTGTATTCTCCTGGTACTGCAACCTGTAAACCAGCACTGCAATGCTGATGCA
CGGGAAGTAGTGGGATGGGAACACAAATGGAAAGCTGTAGAGCTCTTGTTTTCC
AGAAGGAGTTGCTCCTTGAGCCTTTCATTCTCAGCCTCGATAACCTCCAAAGCCG
CTCTAATTGTGGAGGGGGTTCGAATTTAAAAGCTTGGAATGTTGGTTCGTGCGTC
TGGAACAAGCCCAGACTTGTTGCTCACTGGGAAAAGGACCATCAGCTCCAAAAA
ACTTGCCGCTCAAACCGCGTACCTCTGCTTTCGCGCAATCTGCCCTGTTGAAATC
GCCACCACATTCATATTGTGACGCTTGAGCAGTCTGTAATTGCCTCAGAATGTGG
AATCATCTGCCCCCTGTGCGAGCCCATGCCAGGCATGTCGCGGGCGAGGACACC
CGCCACTCGTACAGCAGACCATTATGCTACCTCACAATAGTTCATAACAGTGACC
ATATTTCTCGAAGCTCCCCAACGAGCACCTCCATGCTCTGAGTGGCCACCCCCCG
GCCCTGGTGCTTGCGGAGGGCAGGTCAACCGGCATGGGGCTACCGAAATCCCCG
ACCGGATCCCACCACCCCCGCGATGGGAAGAATCTCTCCCCGGGATGTGGGCCC
ACCACCAGCACAACCTGCTGGCCCAGGCGAGCGTCAAACCATACCACACAAATA
TCCTTGGCATCGGCCCTGAATTCCTTCTGCCGCTCTGCTACCCGGTGCTTCTGTCC
GAAGCAGGGGTTGCTAGGGATCGCTCCGAGTCCGCAAACCCTTGTCGCGTGGCG
GGGCTTGTTCGAGCTTGAAGAGC

[1534]

[1535] 서열 번호 165

[1536] pSZ1469로부터의 RNAi 발현 카세트의 FADc 부분

ACTAGTATGGCTATCAAGACGAACAGGCAGCCTGTGGAGAAGCCTCCGTTCACG
ATCGGGACGCTGCGCAAGGCCATCCCCGCGCACTGTTTCGAGCGCTCGGCGCTTC
GTAGCAGCATGTACCTGGCCTTTGACATCGCGGTCATGTCCCTGCTCTACGTCGC
GTCGACGTACATCGACCCTGCACCGGTGCCTACGTGGGTCAAGTACGGCATCATG
TGGCCGCTCTACTGGTTCTTCCAGGTGTGTTTGAGGGTTTTGGTTGCCCGTATTGA
GGTCCTGGTGGCGCGCATGGAGGAGAAGGCGCCTGTCCCGCTGACCCCCCCGGC
TACCCTCCCGGCACCTTCCAGGGCGCGTACGGGAAGAACCAGTAGAGCGGCCAC
ATGATGCCGTACTTGACCCACGTAGGCACCGGTGCAGGGTCGATGTACGTCGAC
GCGACGTAGAGCAGGGACATGACCGCGATGTCAAAGGCCAGGTACATGCTGCTA

[1537]

CGAAGCGCCGAGCGCTCGAAACAGTGCGCGGGGATGGCCTTGCGCAGCGTCCCG
ATCGTGAACGGAGGCTTCTCCACAGGCTGCCTGTTCGTCTTGATAGCCAT

[1538]

[1539] 서열 번호 166

[1540] pSZ1469로부터의 FADc RNAi 발현 카세트의 관련 부분

GCTCTTCGCCGCCGCCACTCCTGCTCGAGCGCGCCCGCGCGTGCGCCGCCAGCGC
CTTGGCCTTTTCGCCGCGCTCGTGCGCGTCGCTGATGTCCATCACCAGGTCCATG
AGGTCTGCCTTGCGCCGGCTGAGCCACTGCTTCGTCCGGGCGGCCAAGAGGAGC
ATGAGGGAGGACTCCTGGTCCAGGGTCCTGACGTGGTCGCGGCTCTGGGAGCGG
GCCAGCATCATCTGGCTCTGCCGCACCGAGGCCGCCTCCAACTGGTCCTCCAGCA
GCCGCAGTCGCCGCCGACCCTGGCAGAGGAAGACAGGTGAGGGGGGTATGAATT
GTACAGAACAACCACGAGCCTTGTCTAGGCAGAATCCCTACCAGTCATGGCTTTA
CCTGGATGACGGCCTGCGAACAGCTGTCCAGCGACCCTCGCTGCCGCCGCTTCTC
CCGCACGCTTCTTTCCAGCACCGTGATGGCGCGAGCCAGCGCCGCACGCTGGCGC
TGCGCTTCGCCGATCTGAGGACAGTCGGGGAACTCTGATCAGTCTAAACCCCCTT
GCGCGTTAGTGTTGCCATCCTTTGCAGACCGGTGAGAGCCGACTTGTTGTGCGCC
ACCCCCCACACCACCTCCTCCCAGACCAATTCTGTCACCTTTTTGGCGAAGGCAT
CGGCCTCGGCCTGCAGAGAGGACAGCAGTGCCCAGCCGCTGGGGGTTGGCGGAT
GCACGCTCAGGTACCCTTTCTTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGG
GCTGCGAGACGGCTTCCCGGCGCTGCATGCAACACCGATGATGCTTCGACCCCCC
GAAGCTCCTTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGT
TTAAATAGCCAGGCCCCCGATTGCAAAGACATTATAGCGAGCTACCAAAGCCAT
ATTCAAACACCTAGATCACTACCACTTCTACACAGGCCACTCGAGCTTGTGATCG
CACTCCGCTAAGGGGGCGCCTCTTCCTCTTCGTTTCAGTCACAACCCGCAAACGG
CGCGCCATGCTGCTGCAGGCCTTCCTGTTCCTGCTGGCCGGCTTCGCCGCCAAGA
TCAGCGCCTCCATGACGAACGAGACGTCCGACCGCCCCCTGGTGCACTTCACCCC
CAACAAGGGCTGGATGAACGACCCCAACGGCCTGTGGTACGACGAGAAGGACGC
CAAGTGGCACCTGTACTTCCAGTACAACCCGAACGACACCGTCTGGGGGACGCC
CTTGTTCTGGGGCCACGCCACGTCCGACGACCTGACCAACTGGGAGGACCAGCC
CATCGCCATCGCCCCGAAGCGCAACGACTCCGGCGCCTTCTCCGGCTCCATGGTG
GTGGACTACAACAACACCTCCGGCTTCTTCAACGACACCATCGACCCGCGCCAGC
GCTGCGTGGCCATCTGGACCTACAACACCCCGGAGTCCGAGGAGCAGTACATCT
CCTACAGCCTGGACGGCGGCTACACCTTCACCGAGTACCAGAAGAACCCCGTGC
TGGCCGCCAACTCCACCCAGTTCCGCGACCCGAAGGTCTTCTGGTACGAGCCCTC
CCAGAAGTGGATCATGACCGCGGCCAAGTCCCAGGACTACAAGATCGAGATCTA
CTCCTCCGACGACCTGAAGTCCTGGAAGCTGGAGTCCGCGTTCGCCAACGAGGG
CTTCCTCGGCTACCAGTACGAGTGCCCCGGCCTGATCGAGGTCCCCACCGAGCAG
GACCCCAGCAAGTCCTACTGGGTGATGTTCATCTCCATCAACCCCGGCGCCCCGG
CCGGCGGCTCCTTCAACCAGTACTTCGTCGGCAGCTTCAACGGCACCCACTTCGA
GGCCTTCGACAACCAGTCCCGCGTGGTGGACTTCGGCAAGGACTACTACGCCCTG
CAGACCTTCTTCAACACCGACCCGACCTACGGGAGCGCCCTGGGCATCGCGTGG
GCCTCCAACTGGGAGTACTCCGCCTTCGTGCCCACCAACCCCTGGCGCTCCTCCA
TGTCCCTCGTGCGCAAGTTCTCCCTCAACACCGAGTACCAGGCCAACCCGGAGAC
GGAGCTGATCAACCTGAAGGCCGAGCCGATCCTGAACATCAGCAACGCCGGCCC
CTGGAGCCGGTTCGCCACCAACACCACGTTGACGAAGGCCAACAGCTACAACGT
CGACCTGTCCAACAGCACCGGCACCCTGGAGTTCGAGCTGGTGTACGCCGTCAAC
ACCACCCAGACGATCTCCAAGTCCGTGTTCGCGGACCTCTCCCTCTGGTTCAAGG
GCCTGGAGGACCCCGAGGAGTACCTCCGCATGGGCTTCGAGGTGTCCGCGTCCTC
CTTCTTCCTGGACCGCGGGAACAGCAAGGTGAAGTTCGTGAAGGAGAACCCCTA

[1541]

CTTCACCAACCGCATGAGCGTGAACAACCAGCCCTTCAAGAGCGAGAACGACCT
GTCCTACTACAAGGTGTACGGCTTGCTGGACCAGAACATCCTGGAGCTGTACTTC
AACGACGGCGACGTCGTGTCCACCAACACCTACTTCATGACCACCGGGAACGCC
CTGGGCTCCGTGAACATGACGACGGGGGTGGACAACCTGTTCTACATCGACAAG
TTCCAGGTGCGCGAGGTCAAGTGACAATTGGCAGCAGCAGCTCGGATAGTATCG
ACACACTCTGGACGCTGGTCGTGTGATGGACTGTTGCCGCCACACTTGCTGCCTT
GACCTGTGAATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGT
GTGTACGCGCTTTTGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCC
AGCATCCCCTTCCCTCGTTTCATATCGCTTGCATCCCAACCGCAACTTATCTACGC
TGTCCTGCTATCCCTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCTCGCACAGCC
TTGGTTTGGGCTCCGCCTGTATTCTCCTGGTACTGCAACCTGTAAACCAGCACTGC
AATGCTGATGCACGGGAAGTAGTGGGATGGGAACACAAATGGAGGATCCCGCGT
CTCGAACAGAGCGCGCAGAGGAACGCTGAAGGTCTCGCCTCTGTCGCACCTCAG
CGCGGCATACACCACAATAACCACCTGACGAATGCGCTTGGTTCTTCGTCCATTA
GCGAAGCGTCCGGTTCACACACGTGCCACGTTGGCGAGGTGGCAGGTGACAATG
ATCGGTGGAGCTGATGGTCGAAACGTTCACAGCCTAGGGATATCGAATTCCTTTC
TTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTTCCCG
GCGCTGCATGCAACACCGATGATGCTTCGACCCCCCGAAGCTCCTTCGGGGCTGC
ATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGTTTAAATAGCCAGGCCCCC
GATTGCAAAGACATTATAGCGAGCTACCAAAGCCATATTCAAACACCTAGATCA
CTACCACTTCTACACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGGGCG
CCTCTTCCTCTTCGTTTCAGTCACAACCCGCAAACACTAGTATGGCTATCAAGAC
GAACAGGCAGCCTGTGGAGAAGCCTCCGTTCACGATCGGGACGCTGCGCAAGGC
CATCCCCGCGCACTGTTTCGAGCGCTCGGCGCTTCGTAGCAGCATGTACCTGGCC
TTTGACATCGCGGTCATGTCCCTGCTCTACGTCGCGTCGACGTACATCGACCCTG
CACCGGTGCCTACGTGGGTCAAGTACGGCATCATGTGGCCGCTCTACTGGTTCTT
CCAGGTGTGTTTGAGGGTTTTGGTTGCCCGTATTGAGGTCCTGGTGGCGCGCATG
GAGGAGAAGGCGCCTGTCCCGCTGACCCCCCCGGCTACCCTCCCGGCACCTTCCA
GGGCGCGTACGGGAAGAACCAGTAGAGCGGCCACATGATGCCGTACTTGACCCA
CGTAGGCACCGGTGCAGGGTCGATGTACGTCGACGCGACGTAGAGCAGGGACAT
GACCGCGATGTCAAAGGCCAGGTACATGCTGCTACGAAGCGCCGAGCGCTCGAA
ACAGTGCGCGGGGATGGCCTTGCGCAGCGTCCCGATCGTGAACGGAGGCTTCTC
CACAGGCTGCCTGTTCGTCTTGATAGCCATCTCGAGGCAGCAGCAGCTCGGATAG
TATCGACACACTCTGGACGCTGGTCGTGTGATGGACTGTTGCCGCCACACTTGCT
GCCTTGACCTGTGAATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGAT
CTTGTGTGTACGCGCTTTTGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCA
CCCCCAGCATCCCCTTCCCTCGTTTCATATCGCTTGCATCCCAACCGCAACTTATC
TACGCTGTCCTGCTATCCCTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCTCGCA
CAGCCTTGGTTTGGGCTCCGCCTGTATTCTCCTGGTACTGCAACCTGTAAACCAG
CACTGCAATGCTGATGCACGGGAAGTAGTGGGATGGGAACACAAATGGAAAGCT
GTAGAGCTCTTGTTTTCCAGAAGGAGTTGCTCCTTGAGCCTTTCATTCTCAGCCTC
GATAACCTCCAAAGCCGCTCTAATTGTGGAGGGGGTTCGAATTTAAAAGCTTGGA
ATGTTGGTTCGTGCGTCTGGAACAAGCCCAGACTTGTTGCTCACTGGGAAAAGGA
CCATCAGCTCCAAAAAACTTGCCGCTCAAACCGCGTACCTCTGCTTTCGCGCAAT
CTGCCCTGTTGAAATCGCCACCACATTCATATTGTGACGCTTGAGCAGTCTGTAA
TTGCCTCAGAATGTGGAATCATCTGCCCCCTGTGCGAGCCCATGCCAGGCATGTC
GCGGGCGAGGACACCCGCCACTCGTACAGCAGACCATTATGCTACCTCACAATA
GTTCATAACAGTGACCATATTTCTCGAAGCTCCCCAACGAGCACCTCCATGCTCT
GAGTGGCCACCCCCCGGCCCTGGTGCTTGCGGAGGGCAGGTCAACCGGCATGGG
GCTACCGAAATCCCCGACCGGATCCCACCACCCCCGCGATGGGAAGAATCTCTCC

[1542]

CCGGGATGTGGGCCCACCACCAGCACAACCTGCTGGCCCAGGCGAGCGTCAAAC
CATACCACACAAATATCCTTGGCATCGGCCCTGAATTCCTTCTGCCGCTCTGCTAC
CCGGTGCTTCTGTCCGAAGCAGGGGTTGCTAGGGATCGCTCCGAGTCCGCAAACC
CTTGTCGCGTGGCGGGGCTTGTTCGAGCTTGAAGAGC

[1543]

[1544] 서열 번호 167

[1545] pSZ1470으로부터의 RNAi 발현 카세트의 FADc 부분

ACTAGTTCACTTGTGGAACCAGAGCGCGGAGTCGTCCTCGGGCGCGTCCGGGAC
GACGTAGCGGCAGTCGCGCCAGTCCTCCCACAGGGCGCGGCCGACCCAGCGGCT
GTCGGACTGGTAGTACTTGCCCAGGATGGGCCTGATGGCGGCGGAGGCCTCCTC
GGCGTGGTAGTGCGGGATGGTGCTGAAGAGGTGGTGCAGCACGTGGGTGTCGGA
GATGTGGTGCAGGATGTTGTCCATGAAGGGCGGGCCCATGGAGCGGTCCACGGT
GGCCATGGCGCCGCGCAGCCAGTCCCAGTCCTTCTCGAAGTAGTGCGGCAGCGC
CGGGTGCGTGTGCTGGAGCAGCGTGATGAGCACGAGCCACATGTTCACGATCAG
GTAGGGCACCACGTAGGTCTTGACCAGCCAGGCCCAGCCCATGGTGCGGCCCAG
CACGCTGAGCCCGCTGAGCACCGCCACCAGCGCCAGGTCGGAGATGACCACCTC
GATGCGCTCGCGCTTGCTGAAGATGGGCGACCACGGGTCAAAGTGGTTGGCGAA
GCGCGGGTACGGCCGCGAGGCGACGTTGAACATGAGGTACAGCGGCCAGCCCAG
GGTCAGGGTGACCAGCACCTTGCCCATGCGGATGGGCAGCCACTCCTCCCACTCC
AGGCCCTCGTGCGCCACTGCGCGGTGCGGCGGCACAAACACCTCGTCCTTGTCCA
GGCACCCCGTGTTGGAGTGGTGGCGGCGGTGCGAGTGCTTCCAGGAGTAGTAGG
GCACCAGCAGCAGGCTGTGGAACACCAGGCCCACGCCGTCGTTGATGGCCTGGC
TGGAGGAAAAGGCCTGGTGGCCGCACTCGTGCGCGCACACCCAGACACCCGTGC
CGAAGGCGCCCTGGAAGGTGCCGGGAGGGTAGCCGGGGGGGGTCAGCGGGACAG
GCGCCTTCTCCTCCATGCGCGCCACCAGGACCTCAATACGGGCAACCAAAACCCT
CAAACACACCTGGAAGAACCAGTAGAGCGGCCACATGATGCCGTACTTGACCCA
CGTAGGCACCGGTGCAGGGTCGATGTACGTCGACGCGACGTAGAGCAGGGACAT
GACCGCGATGTCAAAGGCCAGGTACATGCTGCTACGAAGCGCCGAGCGCTCGAA
ACAGTGCGCGGGGATGGCCTTGCGCAGCGTCCCGATCGTGAACGGAGGCTTCTC
CACAGGCTGCCTGTTCGTCTTGATAGCCAT

[1546]

[1547] 서열 번호 168

[1548] pSZ1470으로부터의 FADc RNAi 발현 카세트의 관련 부분

GCTCTTCGCCGCCGCCACTCCTGCTCGAGCGCGCCCGCGCGTGCGCCGCCAGCGC
CTTGGCCTTTTCGCCGCGCTCGTGCGCGTCGCTGATGTCCATCACCAGGTCCATG
AGGTCTGCCTTGCGCCGGCTGAGCCACTGCTTCGTCCGGGCGGCCAAGAGGAGC
ATGAGGGAGGACTCCTGGTCCAGGGTCCTGACGTGGTCGCGGCTCTGGGAGCGG
GCCAGCATCATCTGGCTCTGCCGCACCGAGGCCGCCTCCAACTGGTCCTCCAGCA
GCCGCAGTCGCCGCCGACCCTGGCAGAGGAAGACAGGTGAGGGGGGTATGAATT
GTACAGAACAACCACGAGCCTTGTCTAGGCAGAATCCCTACCAGTCATGGCTTTA
CCTGGATGACGGCCTGCGAACAGCTGTCCAGCGACCCTCGCTGCCGCCGCTTCTC
CCGCACGCTTCTTTCCAGCACCGTGATGGCGCGAGCCAGCGCCGCACGCTGGCGC
TGCGCTTCGCCGATCTGAGGACAGTCGGGGAACTCTGATCAGTCTAAACCCCCTT
GCGCGTTAGTGTTGCCATCCTTTGCAGACCGGTGAGAGCCGACTTGTTGTGCGCC
ACCCCCCACACCACCTCCTCCCAGACCAATTCTGTCACCTTTTTGGCGAAGGCAT
CGGCCTCGGCCTGCAGAGAGGACAGCAGTGCCCAGCCGCTGGGGGTTGGCGGAT
GCACGCTCAGGTACCCTTTCTTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGG

[1549]

GCTGCGAGACGGCTTCCCGGCGCTGCATGCAACACCGATGATGCTTCGACCCCCC
GAAGCTCCTTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGT
TTAAATAGCCAGGCCCCCGATTGCAAAGACATTATAGCGAGCTACCAAAGCCAT
ATTCAAACACCTAGATCACTACCACTTCTACACAGGCCACTCGAGCTTGTGATCG
CACTCCGCTAAGGGGGCGCCTCTTCCTCTTCGTTTCAGTCACAACCCGCAAACGG
CGCGCCATGCTGCTGCAGGCCTTCCTGTTCCTGCTGGCCGGCTTCGCCGCCAAGA
TCAGCGCCTCCATGACGAACGAGACGTCCGACCGCCCCCTGGTGCACTTCACCCC
CAACAAGGGCTGGATGAACGACCCCAACGGCCTGTGGTACGACGAGAAGGACGC
CAAGTGGCACCTGTACTTCCAGTACAACCCGAACGACACCGTCTGGGGGACGCC
CTTGTTCTGGGGCCACGCCACGTCCGACGACCTGACCAACTGGGAGGACCAGCC
CATCGCCATCGCCCCGAAGCGCAACGACTCCGGCGCCTTCTCCGGCTCCATGGTG
GTGGACTACAACAACACCTCCGGCTTCTTCAACGACACCATCGACCCGCGCCAGC
GCTGCGTGGCCATCTGGACCTACAACACCCCGGAGTCCGAGGAGCAGTACATCT
CCTACAGCCTGGACGGCGGCTACACCTTCACCGAGTACCAGAAGAACCCCGTGC
TGGCCGCCAACTCCACCCAGTTCCGCGACCCGAAGGTCTTCTGGTACGAGCCCTC
CCAGAAGTGGATCATGACCGCGGCCAAGTCCCAGGACTACAAGATCGAGATCTA
CTCCTCCGACGACCTGAAGTCCTGGAAGCTGGAGTCCGCGTTCGCCAACGAGGG
CTTCCTCGGCTACCAGTACGAGTGCCCCGGCCTGATCGAGGTCCCCACCGAGCAG
GACCCCAGCAAGTCCTACTGGGTGATGTTCATCTCCATCAACCCCGGCGCCCCGG
CCGGCGGCTCCTTCAACCAGTACTTCGTCGGCAGCTTCAACGGCACCCACTTCGA
GGCCTTCGACAACCAGTCCCGCGTGGTGGACTTCGGCAAGGACTACTACGCCCTG
CAGACCTTCTTCAACACCGACCCGACCTACGGGAGCGCCCTGGGCATCGCGTGG
GCCTCCAACTGGGAGTACTCCGCCTTCGTGCCCACCAACCCCTGGCGCTCCTCCA
TGTCCCTCGTGCGCAAGTTCTCCCTCAACACCGAGTACCAGGCCAACCCGGAGAC
GGAGCTGATCAACCTGAAGGCCGAGCCGATCCTGAACATCAGCAACGCCGGCCC
CTGGAGCCGGTTCGCCACCAACACCACGTTGACGAAGGCCAACAGCTACAACGT
CGACCTGTCCAACAGCACCGGCACCCTGGAGTTCGAGCTGGTGTACGCCGTCAAC
ACCACCCAGACGATCTCCAAGTCCGTGTTCGCGGACCTCTCCCTCTGGTTCAAGG
GCCTGGAGGACCCCGAGGAGTACCTCCGCATGGGCTTCGAGGTGTCCGCGTCCTC
CTTCTTCCTGGACCGCGGGAACAGCAAGGTGAAGTTCGTGAAGGAGAACCCCTA
CTTCACCAACCGCATGAGCGTGAACAACCAGCCCTTCAAGAGCGAGAACGACCT
GTCCTACTACAAGGTGTACGGCTTGCTGGACCAGAACATCCTGGAGCTGTACTTC
AACGACGGCGACGTCGTGTCCACCAACACCTACTTCATGACCACCGGGAACGCC
CTGGGCTCCGTGAACATGACGACGGGGGTGGACAACCTGTTCTACATCGACAAG
TTCCAGGTGCGCGAGGTCAAGTGACAATTGGCAGCAGCAGCTCGGATAGTATCG
ACACACTCTGGACGCTGGTCGTGTGATGGACTGTTGCCGCCACACTTGCTGCCTT
GACCTGTGAATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGT
GTGTACGCGCTTTTGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCC
AGCATCCCCTTCCCTCGTTTCATATCGCTTGCATCCCAACCGCAACTTATCTACGC
TGTCCTGCTATCCCTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCTCGCACAGCC
TTGGTTTGGGCTCCGCCTGTATTCTCCTGGTACTGCAACCTGTAAACCAGCACTGC
AATGCTGATGCACGGGAAGTAGTGGGATGGGAACACAAATGGAGGATCCCGCGT
CTCGAACAGAGCGCGCAGAGGAACGCTGAAGGTCTCGCCTCTGTCGCACCTCAG
CGCGGCATACACCACAATAACCACCTGACGAATGCGCTTGGTTCTTCGTCCATTA
GCGAAGCGTCCGGTTCACACACGTGCCACGTTGGCGAGGTGGCAGGTGACAATG
ATCGGTGGAGCTGATGGTCGAAACGTTCACAGCCTAGGGATATCGAATTCGGCC
GACAGGACGCGCGTCAAAGGTGCTGGTCGTGTATGCCCTGGCCGGCAGGTCGTT
GCTGCTGCTGGTTAGTGATTCCGCAACCCTGATTTTGGCGTCTTATTTTGGCGTGG
CAAACGCTGGCGCCCGCGAGCCGGGCCGGCGGCGATGCGGTGCCCCACGGCTGC
CGGAATCCAAGGGAGGCAAGAGCGCCCGGGTCAGTTGAAGGGCTTTACGCGCAA

[1550]

GGTACAGCCGCTCCTGCAAGGCTGCGTGGTGGAATTGGACGTGCAGGTCCTGCTG
AAGTTCCTCCACCGCCTCACCAGCGGACAAAGCACCGGTGTATCAGGTCCGTGTC
ATCCACTCTAAAGAGCTCGACTACGACCTACTGATGGCCCTAGATTCTTCATCAA
AAACGCCTGAGACACTTGCCCAGGATTGAAACTCCCTGAAGGGACCACCAGGGG
CCCTGAGTTGTTCCTTCCCCCCGTGGCGAGCTGCCAGCCAGGCTGTACCTGTGAT
CGAGGCTGGCGGGAAAATAGGCTTCGTGTGCTCAGGTCATGGGAGGTGCAGGAC
AGCTCATGAAACGCCAACAATCGCACAATTCATGTCAAGCTAATCAGCTATTTCC
TCTTCACGAGCTGTAATTGTCCCAAAATTCTGGTCTACCGGGGGTGATCCTTCGT
GTACGGGCCCTTCCCTCAACCCTAGGTATGCGCGCATGCGGTCGCCGCGCAACTC
GCGCGAGGGCCGAGGGTTTGGGACGGGCCGTCCCGAAATGCAGTTGCACCCGGA
TGCGTGGCACCTTTTTTGCGATAATTTATGCAATGGACTGCTCTGCAAAATTCTGG
CTCTGTCGCCAACCCTAGGATCAGCGGCGTAGGATTTCGTAATCATTCGTCCTGA
TGGGGAGCTACCGACTACCCTAATATCAGCCCGACTGCCTGACGCCAGCGTCCAC
TTTTGTGCACACATTCCATTCGTGCCCAAGACATTTCATTGTGGTGCGAAGCGTCC
CCAGTTACGCTCACCTGTTTCCCGACCTCCTTACTGTTCTGTCGACAGAGCGGGCC
CACAGGCCGGTCGCAGCCACTAGTTCACTTGTGGAACCAGAGCGCGGAGTCGTC
CTCGGGCGCGTCCGGGACGACGTAGCGGCAGTCGCGCCAGTCCTCCCACAGGGC
GCGGCCGACCCAGCGGCTGTCGGACTGGTAGTACTTGCCCAGGATGGGCCTGAT
GGCGGCGGAGGCCTCCTCGGCGTGGTAGTGCGGGATGGTGCTGAAGAGGTGGTG
CAGCACGTGGGTGTCGGAGATGTGGTGCAGGATGTTGTCCATGAAGGGCGGGCC
CATGGAGCGGTCCACGGTGGCCATGGCGCCGCGCAGCCAGTCCCAGTCCTTCTCG
AAGTAGTGCGGCAGCGCCGGGTGCGTGTGCTGGAGCAGCGTGATGAGCACGAGC
CACATGTTCACGATCAGGTAGGGCACCACGTAGGTCTTGACCAGCCAGGCCCAG
CCCATGGTGCGGCCCAGCACGCTGAGCCCGCTGAGCACCGCCACCAGCGCCAGG
TCGGAGATGACCACCTCGATGCGCTCGCGCTTGCTGAAGATGGGCGACCACGGG
TCAAAGTGGTTGGCGAAGCGCGGGTACGGCCGCGAGGCGACGTTGAACATGAGG
TACAGCGGCCAGCCCAGGGTCAGGGTGACCAGCACCTTGCCCATGCGGATGGGC
AGCCACTCCTCCCACTCCAGGCCCTCGTGCGCCACTGCGCGGTGCGGCGGCACAA
ACACCTCGTCCTTGTCCAGGCACCCCGTGTTGGAGTGGTGGCGGCGGTGCGAGTG
CTTCCAGGAGTAGTAGGGCACCAGCAGCAGGCTGTGGAACACCAGGCCCACGCC
GTCGTTGATGGCCTGGCTGGAGGAAAAGGCCTGGTGGCCGCACTCGTGCGCGCA
CACCCAGACACCCGTGCCGAAGGCGCCCTGGAAGGTGCCGGGAGGGTAGCCGGG
GGGGTCAGCGGGACAGGCGCCTTCTCCTCCATGCGCGCCACCAGGACCTCAATA
CGGGCAACCAAAACCCTCAAACACACCTGGAAGAACCAGTAGAGCGGCCACATG
ATGCCGTACTTGACCCACGTAGGCACCGGTGCAGGGTCGATGTACGTCGACGCG
ACGTAGAGCAGGGACATGACCGCGATGTCAAAGGCCAGGTACATGCTGCTACGA
AGCGCCGAGCGCTCGAAACAGTGCGCGGGGATGGCCTTGCGCAGCGTCCCGATC
GTGAACGGAGGCTTCTCCACAGGCTGCCTGTTCGTCTTGATAGCCATCTCGAGGC
AGCAGCAGCTCGGATAGTATCGACACACTCTGGACGCTGGTCGTGTGATGGACT
GTTGCCGCCACACTTGCTGCCTTGACCTGTGAATATCCCTGCCGCTTTTATCAAAC
AGCCTCAGTGTGTTTGATCTTGTGTGTACGCGCTTTTGCGAGTTGCTAGCTGCTTG
TGCTATTTGCGAATACCACCCCCAGCATCCCCTTCCCTCGTTTCATATCGCTTGCA
TCCCAACCGCAACTTATCTACGCTGTCCTGCTATCCCTCAGCGCTGCTCCTGCTCC
TGCTCACTGCCCCTCGCACAGCCTTGGTTTGGGCTCCGCCTGTATTCTCCTGGTAC
TGCAACCTGTAAACCAGCACTGCAATGCTGATGCACGGGAAGTAGTGGGATGGG
AACACAAATGGAAAGCTGTAGAGCTCTTGTTTTCCAGAAGGAGTTGCTCCTTGAG
CCTTTCATTCTCAGCCTCGATAACCTCCAAAGCCGCTCTAATTGTGGAGGGGGTT
CGAATTTAAAAGCTTGGAATGTTGGTTCGTGCGTCTGGAACAAGCCCAGACTTGT
TGCTCACTGGGAAAAGGACCATCAGCTCCAAAAAACTTGCCGCTCAAACCGCGT
ACCTCTGCTTTCGCGCAATCTGCCCTGTTGAAATCGCCACCACATTCATATTGTGA

[1551]

CGCTTGAGCAGTCTGTAATTGCCTCAGAATGTGGAATCATCTGCCCCCTGTGCGA
GCCCATGCCAGGCATGTCGCGGGCGAGGACACCCGCCACTCGTACAGCAGACCA
TTATGCTACCTCACAATAGTTCATAACAGTGACCATATTTCTCGAAGCTCCCCAA
CGAGCACCTCCATGCTCTGAGTGGCCACCCCCCGGCCCTGGTGCTTGCGGAGGGC
AGGTCAACCGGCATGGGGCTACCGAAATCCCCGACCGGATCCCACCACCCCCGC
GATGGGAAGAATCTCTCCCCGGGATGTGGGCCCACCACCAGCACAACCTGCTGG
CCCAGGCGAGCGTCAAACCATACCACACAAATATCCTTGGCATCGGCCCTGAATT
CCTTCTGCCGCTCTGCTACCCGGTGCTTCTGTCCGAAGCAGGGGTTGCTAGGGAT
CGCTCCGAGTCCGCAAACCCTTGTCGCGTGGCGGGGCTTGTTCGAGCTTGAAGAG
C

[1552]

[1553] 서열 번호 169

[1554] pSZ1471로부터의 RNAi 발현 카세트의 FADc 부분

ACTAGTTCACTTGTGGAACCAGAGCGCGGAGTCGTCCTCGGGCGCGTCCGGGAC
GACGTAGCGGCAGTCGCGCCAGTCCTCCCACAGGGCGCGGCCGACCCAGCGGCT
GTCGGACTGGTAGTACTTGCCCAGGATGGGCCTGATGGCGGCGGAGGCCTCCTC
GGCGTGGTAGTGCGGGATGGTGCTGAAGAGGTGGTGCAGCACGTGGGTGTCGGA
GATGTGGTGCAGGATGTTGTCCATGAAGGGCGGGCCCATGGAGCGGTCCACGGT
GGCCATGGCGCCGCGCAGCCAGTCCCAGTCCTTCTCGAAGTAGTGCGGCAGCGC
CGGGTGCGTGTGCTGGAGCAGCGTGATGAGCACGAGCCACATGTTCACGATCAG
GTAGGGCACCACGTAGGTCTTGACCAGCCAGGCCCAGCCCATGGTGCGGCCCAG
CACGCTGAGCCCGCTGAGCACCGCCACCAGCGCCAGGTCGGAGATGACCACCTC
GATGCGCTCGCGCTTGCTGAAGATGGGCGACCACGGGTCAAAGTGGTTGGCGAA
GCGCGGGTACGGCCGCGAGGCGACGTTGAACATGAGGTACAGCGGCCAGCCCAG
GGTCAGGGTGACCAGCACCTTGCCCATGCGGATGGGCAGCCACTCCTCCCACTCC
AGGCCCTCGTGCGCCACTGCGCGGTGCGGCGGCACAAACACCTCGTCCTTGTCCA
GGCACCCCGTGTTGGAGTGGTGGCGGCGGTGCGAGTGCTTCCAGGAGTAGTAGG
GCACCAGCAGCAGGCTGTGGAACACCAGGCCCACGCCGTCGTTGATGGCCTGGC
TGGAGGAAAAGGCCTGGTGGCCGCACTCGTGCGCGCACACCCAGACACCCGTGC
CGAAGGCGCCCTGGAAGGTGCCGGGAGGGTAGCCGGGGGGGGTCAGCGGGACAG
GCGCCTTCTCCTCCATGCGCGCCACCAGGACCTCAATACGGGCAACCAAAACCCT
CAAACACACCTGGAAGAACCAGTAGAGCGGCCACATGATGCCGTACTTGACCCA
CGTAGGCACCGGTGCAGGGTCGATGTACGTCGACGCGACGTAGAGCAGGGACAT
GACCGCGATGTCAAAGGCCAGGTACATGCTGCTACGAAGCGCCGAGCGCTCGAA
ACAGTGCGCGGGGATGGCCTTGCGCAGCGTCCCGATCGTGAACGGAGGCTTCTC
CACAGGCTGCCTGTTCGTCTTGATAGCCAT

[1555]

[1556] 서열 번호 170

[1557] pSZ1471로부터의 FADc RNAi 발현 카세트의 관련 부분

GCTCTTCGCCGCCGCCACTCCTGCTCGAGCGCGCCCGCGCGTGCGCCGCCAGCGC
CTTGGCCTTTTCGCCGCGCTCGTGCGCGTCGCTGATGTCCATCACCAGGTCCATG
AGGTCTGCCTTGCGCCGGCTGAGCCACTGCTTCGTCCGGGCGGCCAAGAGGAGC
ATGAGGGAGGACTCCTGGTCCAGGGTCCTGACGTGGTCGCGGCTCTGGGAGCGG
GCCAGCATCATCTGGCTCTGCCGCACCGAGGCCGCCTCCAACTGGTCCTCCAGCA
GCCGCAGTCGCCGCCGACCCTGGCAGAGGAAGACAGGTGAGGGGGGTATGAATT
GTACAGAACAACCACGAGCCTTGTCTAGGCAGAATCCCTACCAGTCATGGCTTTA
CCTGGATGACGGCCTGCGAACAGCTGTCCAGCGACCCTCGCTGCCGCCGCTTCTC
CCGCACGCTTCTTTCCAGCACCGTGATGGCGCGAGCCAGCGCCGCACGCTGGCGC

[1558]

TGCGCTTCGCCGATCTGAGGACAGTCGGGGAACTCTGATCAGTCTAAACCCCCTT
GCGCGTTAGTGTTGCCATCCTTTGCAGACCGGTGAGAGCCGACTTGTTGTGCGCC
ACCCCCCACACCACCTCCTCCCAGACCAATTCTGTCACCTTTTTGGCGAAGGCAT
CGGCCTCGGCCTGCAGAGAGGACAGCAGTGCCCAGCCGCTGGGGGTTGGCGGAT
GCACGCTCAGGTACCCTTTCTTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGG
GCTGCGAGACGGCTTCCCGGCGCTGCATGCAACACCGATGATGCTTCGACCCCCC
GAAGCTCCTTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGT
TTAAATAGCCAGGCCCCCGATTGCAAAGACATTATAGCGAGCTACCAAAGCCAT
ATTCAAACACCTAGATCACTACCACTTCTACACAGGCCACTCGAGCTTGTGATCG
CACTCCGCTAAGGGGGCGCCTCTTCCTCTTCGTTTCAGTCACAACCCGCAAACGG
CGCGCCATGCTGCTGCAGGCCTTCCTGTTCCTGCTGGCCGGCTTCGCCGCCAAGA
TCAGCGCCTCCATGACGAACGAGACGTCCGACCGCCCCCTGGTGCACTTCACCCC
CAACAAGGGCTGGATGAACGACCCCAACGGCCTGTGGTACGACGAGAAGGACGC
CAAGTGGCACCTGTACTTCCAGTACAACCCGAACGACACCGTCTGGGGGACGCC
CTTGTTCTGGGGCCACGCCACGTCCGACGACCTGACCAACTGGGAGGACCAGCC
CATCGCCATCGCCCCGAAGCGCAACGACTCCGGCGCCTTCTCCGGCTCCATGGTG
GTGGACTACAACAACACCTCCGGCTTCTTCAACGACACCATCGACCCGCGCCAGC
GCTGCGTGGCCATCTGGACCTACAACACCCCGGAGTCCGAGGAGCAGTACATCT
CCTACAGCCTGGACGGCGGCTACACCTTCACCGAGTACCAGAAGAACCCCGTGC
TGGCCGCCAACTCCACCCAGTTCCGCGACCCGAAGGTCTTCTGGTACGAGCCCTC
CCAGAAGTGGATCATGACCGCGGCCAAGTCCCAGGACTACAAGATCGAGATCTA
CTCCTCCGACGACCTGAAGTCCTGGAAGCTGGAGTCCGCGTTCGCCAACGAGGG
CTTCCTCGGCTACCAGTACGAGTGCCCCGGCCTGATCGAGGTCCCCACCGAGCAG
GACCCCAGCAAGTCCTACTGGGTGATGTTCATCTCCATCAACCCCGGCGCCCCGG
CCGGCGGCTCCTTCAACCAGTACTTCGTCGGCAGCTTCAACGGCACCCACTTCGA
GGCCTTCGACAACCAGTCCCGCGTGGTGGACTTCGGCAAGGACTACTACGCCCTG
CAGACCTTCTTCAACACCGACCCGACCTACGGGAGCGCCCTGGGCATCGCGTGG
GCCTCCAACTGGGAGTACTCCGCCTTCGTGCCCACCAACCCCTGGCGCTCCTCCA
TGTCCCTCGTGCGCAAGTTCTCCCTCAACACCGAGTACCAGGCCAACCCGGAGAC
GGAGCTGATCAACCTGAAGGCCGAGCCGATCCTGAACATCAGCAACGCCGGCCC
CTGGAGCCGGTTCGCCACCAACACCACGTTGACGAAGGCCAACAGCTACAACGT
CGACCTGTCCAACAGCACCGGCACCCTGGAGTTCGAGCTGGTGTACGCCGTCAAC
ACCACCCAGACGATCTCCAAGTCCGTGTTCGCGGACCTCTCCCTCTGGTTCAAGG
GCCTGGAGGACCCCGAGGAGTACCTCCGCATGGGCTTCGAGGTGTCCGCGTCCTC
CTTCTTCCTGGACCGCGGGAACAGCAAGGTGAAGTTCGTGAAGGAGAACCCCTA
CTTCACCAACCGCATGAGCGTGAACAACCAGCCCTTCAAGAGCGAGAACGACCT
GTCCTACTACAAGGTGTACGGCTTGCTGGACCAGAACATCCTGGAGCTGTACTTC
AACGACGGCGACGTCGTGTCCACCAACACCTACTTCATGACCACCGGGAACGCC
CTGGGCTCCGTGAACATGACGACGGGGGTGGACAACCTGTTCTACATCGACAAG
TTCCAGGTGCGCGAGGTCAAGTGACAATTGGCAGCAGCAGCTCGGATAGTATCG
ACACACTCTGGACGCTGGTCGTGTGATGGACTGTTGCCGCCACACTTGCTGCCTT
GACCTGTGAATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGT
GTGTACGCGCTTTTGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCC
AGCATCCCCTTCCCTCGTTTCATATCGCTTGCATCCCAACCGCAACTTATCTACGC
TGTCCTGCTATCCCTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCTCGCACAGCC
TTGGTTTGGGCTCCGCCTGTATTCTCCTGGTACTGCAACCTGTAAACCAGCACTGC
AATGCTGATGCACGGGAAGTAGTGGGATGGGAACACAAATGGAGGATCCCGCGT
CTCGAACAGAGCGCGCAGAGGAACGCTGAAGGTCTCGCCTCTGTCGCACCTCAG
CGCGGCATACACCACAATAACCACCTGACGAATGCGCTTGGTTCTTCGTCCATTA
GCGAAGCGTCCGGTTCACACACGTGCCACGTTGGCGAGGTGGCAGGTGACAATG

[1559]

ATCGGTGGAGCTGATGGTCGAAACGTTCACAGCCTAGGGATATCGAATTCCTTTC
TTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTTCCCG
GCGCTGCATGCAACACCGATGATGCTTCGACCCCCCGAAGCTCCTTCGGGGCTGC
ATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGTTTAAATAGCCAGGCCCCC
GATTGCAAAGACATTATAGCGAGCTACCAAAGCCATATTCAAACACCTAGATCA
CTACCACTTCTACACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGGGCG
CCTCTTCCTCTTCGTTTCAGTCACAACCCGCAAACACTAGTTCACTTGTGGAACCA
GAGCGCGGAGTCGTCCTCGGGCGCGTCCGGGACGACGTAGCGGCAGTCGCGCCA
GTCCTCCCACAGGGCGCGGCCGACCCAGCGGCTGTCGGACTGGTAGTACTTGCCC
AGGATGGGCCTGATGGCGGCGGAGGCCTCCTCGGCGTGGTAGTGCGGGATGGTG
CTGAAGAGGTGGTGCAGCACGTGGGTGTCGGAGATGTGGTGCAGGATGTTGTCC
ATGAAGGGCGGGCCCATGGAGCGGTCCACGGTGGCCATGGCGCCGCGCAGCCAG
TCCCAGTCCTTCTCGAAGTAGTGCGGCAGCGCCGGGTGCGTGTGCTGGAGCAGCG
TGATGAGCACGAGCCACATGTTCACGATCAGGTAGGGCACCACGTAGGTCTTGA
CCAGCCAGGCCCAGCCCATGGTGCGGCCCAGCACGCTGAGCCCGCTGAGCACCG
CCACCAGCGCCAGGTCGGAGATGACCACCTCGATGCGCTCGCGCTTGCTGAAGA
TGGGCGACCACGGGTCAAAGTGGTTGGCGAAGCGCGGGTACGGCCGCGAGGCGA
CGTTGAACATGAGGTACAGCGGCCAGCCCAGGGTCAGGGTGACCAGCACCTTGC
CCATGCGGATGGGCAGCCACTCCTCCCACTCCAGGCCCTCGTGCGCCACTGCGCG
GTGCGGCGGCACAAACACCTCGTCCTTGTCCAGGCACCCCGTGTTGGAGTGGTGG
CGGCGGTGCGAGTGCTTCCAGGAGTAGTAGGGCACCAGCAGCAGGCTGTGGAAC
ACCAGGCCCACGCCGTCGTTGATGGCCTGGCTGGAGGAAAAGGCCTGGTGGCCG
CACTCGTGCGCGCACACCCAGACACCCGTGCCGAAGGCGCCCTGGAAGGTGCCG
GGAGGGTAGCCGGGGGGGTCAGCGGGACAGGCGCCTTCTCCTCCATGCGCGCCA
CCAGGACCTCAATACGGGCAACCAAAACCCTCAAACACACCTGGAAGAACCAGT
AGAGCGGCCACATGATGCCGTACTTGACCCACGTAGGCACCGGTGCAGGGTCGA
TGTACGTCGACGCGACGTAGAGCAGGGACATGACCGCGATGTCAAAGGCCAGGT
ACATGCTGCTACGAAGCGCCGAGCGCTCGAAACAGTGCGCGGGGATGGCCTTGC
GCAGCGTCCCGATCGTGAACGGAGGCTTCTCCACAGGCTGCCTGTTCGTCTTGAT
AGCCATCTCGAGGCAGCAGCAGCTCGGATAGTATCGACACACTCTGGACGCTGG
TCGTGTGATGGACTGTTGCCGCCACACTTGCTGCCTTGACCTGTGAATATCCCTGC
CGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGTGTGTACGCGCTTTTGCGAG
TTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCCAGCATCCCCTTCCCTCGTT
TCATATCGCTTGCATCCCAACCGCAACTTATCTACGCTGTCCTGCTATCCCTCAGC
GCTGCTCCTGCTCCTGCTCACTGCCCCTCGCACAGCCTTGGTTTGGGCTCCGCCTG
TATTCTCCTGGTACTGCAACCTGTAAACCAGCACTGCAATGCTGATGCACGGGAA
GTAGTGGGATGGGAACACAAATGGAAAGCTGTAGAGCTCTTGTTTTCCAGAAGG
AGTTGCTCCTTGAGCCTTTCATTCTCAGCCTCGATAACCTCCAAAGCCGCTCTAAT
TGTGGAGGGGGTTCGAATTTAAAAGCTTGGAATGTTGGTTCGTGCGTCTGGAACA
AGCCCAGACTTGTTGCTCACTGGGAAAAGGACCATCAGCTCCAAAAAACTTGCC
GCTCAAACCGCGTACCTCTGCTTTCGCGCAATCTGCCCTGTTGAAATCGCCACCA
CATTCATATTGTGACGCTTGAGCAGTCTGTAATTGCCTCAGAATGTGGAATCATC
TGCCCCCTGTGCGAGCCCATGCCAGGCATGTCGCGGGCGAGGACACCCGCCACT
CGTACAGCAGACCATTATGCTACCTCACAATAGTTCATAACAGTGACCATATTTC
TCGAAGCTCCCCAACGAGCACCTCCATGCTCTGAGTGGCCACCCCCCGGCCCTGG
TGCTTGCGGAGGGCAGGTCAACCGGCATGGGGCTACCGAAATCCCCGACCGGAT
CCCACCACCCCCGCGATGGGAAGAATCTCTCCCCGGGATGTGGGCCCACCACCA
GCACAACCTGCTGGCCCAGGCGAGCGTCAAACCATACCACACAAATATCCTTGG
CATCGGCCCTGAATTCCTTCTGCCGCTCTGCTACCCGGTGCTTCTGTCCGAAGCAG

[1560]

GGGTTGCTAGGGATCGCTCCGAGTCCGCAAACCCTTGTCGCGTGGCGGGGCTTGT
TCGAGCTTGAAGAGC

[1561]

[1562] 서열 번호 171

[1563] pSZ1377로부터의 클로렐라 프로토테코이데스 UTEX 250 스테아로일-ACP 운반 펩티드를 지니는 올레아 유로파에아 스테아로일-ACP 불포화 효소의 코돈-최적화 코딩 영역

ACTAGTATGGCCACCGCATCCACTTTCTCGGCGTTCAATGCCCGCTGCGGCGACC
TGCGTCGCTCGGCGGGCTCCGGGCCCCGGCGCCCAGCGAGGCCCCTCCCCGTGCG
CGGGCGCGCCGAGGTGCACGTGCAGGTGACCCACTCCCTGGCCCCCGAGAAGCG
CGAGATCTTCAACTCCCTGAACAACTGGGCCCAGGAGAACATCCTGGTGCTGCTG
AAGGACGTGGACAAGTGCTGGCAGCCCTCCGACTTCCTGCCCGACTCCGCCTCCG
AGGGCTTCGACGAGCAGGTGATGGAGCTGCGCAAGCGCTGCAAGGAGATCCCCG
ACGACTACTTCATCGTGCTGGTGGGCGACATGATCACCGAGGAGGCCCTGCCCAC
CTACCAGACCATGCTGAACACCCTGGACGGCGTGCGCGACGAGACCGGCGCCTC
CCTGACCCCCTGGGCCATCTGGACCCGCGCCTGGACCGCCGAGGAGAACCGCCA
CGGCGACCTGCTGAACAAGTACCTGTACCTGTCCGGCCGCGTGGACATGAAGCA
GATCGAGAAGACCATCCAGTACCTGATCGGCTCCGGCATGGACCCCCGCACCGA
GAACAACCCCTACCTGGGCTTCATCTACACCTCCTTCCAGGAGCGCGCCACCTTC
ATCTCCCACGGCAACACCGCCCGCCTGGCCAAGGAGCACGGCGACCTGAAGCTG
GCCCAGATCTGCGGCATCATCGCCGCCGACGAGAAGCGCCACGAGACCGCCTAC
ACCAAGATCGTGGAGAAGCTGTTCGAGATCGACCCCGACGGCACCGTGCTGGCC
CTGGCCGACATGATGCGCAAGAAGGTGTCCATGCCCGCCCACCTGATGTACGAC
GGCCAGGACGACAACCTGTTCGAGAACTTCTCCTCCGTGGCCCAGCGCCTGGGCG
TGTACACCGCCAAGGACTACGCCGACATCCTGGAGTTCCTGGTGGGCCGCTGGG
ACATCGAGAAGCTGACCGGCCTGTCCGGCGAGGGCCGCAAGGCCCAGGACTACG
TGTGCACCCTGCCCCCCCGCATCCGCCGCCTGGAGGAGCGCGCCCAGTCCCGCGT
GAAGAAGGCCTCCGCCACCCCCTTCTCCTGGATCTTCGGCCGCGAGATCAACCTG
ATGGACTACAAGGACCACGACGGCGACTACAAGGACCACGACATCGACTACAAG
GACGACGACGACAAGTGAATCGAT

[1564]

[1565] 서열 번호 172

[1566] 올레아 유로파에아 스테아로일-ACP 불포화 효소의 아미노산 서열 (수탁 번호 AAB67840.1)

MALKLCFPPHKMPSFPDARIRSHRVFMASTIHSPSMEVGKVKKPFTPPREVHVQVTH
SLAPEKREIFNSLNNWAQENILVLLKDVDKCWQPSDFLPDSASEGFDEQVMELRKRC
KEIPDDYFIVLVGDMITEEALPTYQTMLNTLDGVRDETGASLTPWAIWTRAWTAEEN
RHGDLLNKYLYLSGRVDMKQIEKTIQYLIGSGMDPRTENNPYLGFIYTSFQERATFIS
HGNTARLAKEHGDLKLAQICGIIAADEKRHETAYTKIVEKLFEIDPDGTVLALADMM
RKKVSMPAHLMYDGQDDNLFENFSSVAQRLGVYTAKDYADILEFLVGRWDIEKLT
GLSGEGRKAQDYVCTLPPRIRRLEERAQSRVKKASATPFSWIFGREINL

[1567]

[1568] 서열 번호 173

[1569] pSZ1419로부터의 리시누스 커뮤니스 올리에이트 12-하이드록실라제의 코돈-최적화 코딩 서열

ACTAGTATGGGCGGCGGCGGCCGCATGTCCACCGTGATCACCTCCAACAACTCCG
AGAAGAAGGGCGGCTCCTCCCACCTGAAGCGCGCCCCCCACACCAAGCCCCCCT
TCACCCTGGGCGACCTGAAGCGCGCCATCCCCCCCCCACTGCTTCGAGCGCTCCTT

[1570]

CGTGCGCGGGCGCGCCTCCTTCTCCTACGTGGCCTACGACGTGTGCCTGTCCTTCC
TGTTCTACTCCATCGCCACCAACTTCTTCCCCTACATCTCCTCCCCCCTGTCCTAC
GTGGCCTGGCTGGTGTACTGGCTGTTCCAGGGCTGCATCCTGACCGGCCTGTGGG
TGATCGGCCACGAGTGCGGCCACCACGCCTTCTCCGAGTACCAGCTGGCCGACG
ACATCGTGGGCCTGATCGTGCACTCCGCCCTGCTGGTGCCCTACTTCTCCTGGAA
GTACTCCCACCGCCGCCACCACTCCAACATCGGCTCCCTGGAGCGCGACGAGGTG
TTCGTGCCCAAGTCCAAGTCCAAGATCTCCTGGTACTCCAAGTACTCCAACAACC
CCCCCGGCCGCGTGCTGACCCTGGCCGCCACCCTGCTGCTGGGCTGGCCCCTGTA
CCTGGCCTTCAACGTGTCCGGCCGCCCCTACGACCGCTTCGCCTGCCACTACGAC
CCCTACGGCCCCATCTTCTCCGAGCGCGAGCGCCTGCAGATCTACATCGCCGACC
TGGGCATCTTCGCCACCACCTTCGTGCTGTACCAGGCCACCATGGCCAAGGGCCT
GGCCTGGGTGATGCGCATCTACGGCGTGCCCCTGCTGATCGTGAACTGCTTCCTG
GTGATGATCACCTACCTGCAGCACACCCACCCCGCCATCCCCCGCTACGGCTCCT
CCGAGTGGGACTGGCTGCGCGGCGCCATGGTGACCGTGGACCGCGACTACGGCG
TGCTGAACAAGGTGTTCCACAACATCGCCGACACCCACGTGGCCCACCACCTGTT
CGCCACCGTGCCCCACTACCACGCCATGGAGGCCACCAAGGCCATCAAGCCCAT
CATGGGCGAGTACTACCGCTACGACGGCACCCCCTTCTACAAGGCCCTGTGGCGC
GAGGCCAAGGAGTGCCTGTTCGTGGAGCCCGACGAGGGCGCCCCCACCCAGGGC
GTGTTCTGGTACCGCAACAAGTACATGGACTACAAGGACCACGACGGCGACTAC
AAGGACCACGACATCGACTACAAGGACGACGACGACAAGTGAATCGAT

[1571]

[1572] 서열 번호 174

[1573] 리시누스 커뮤니스 올리에이트 12-하이드록실라제의 아미노산 서열 (수탁 번호 AAC49010.1)

MGGGGRMSTVITSNNSEKKGGSSHLKRAPHTKPPFTLGDLKRAIPPHCFERSFVRSFS
YVAYDVCLSFLFYSIATNFFPYISSPLSYVAWLVYWLFQGCILTGLWVIGHECGHHAF
SEYQLADDIVGLIVHSALLVPYFSWKYSHRRHHSNIGSLERDEVFVPKSKSKISWYSK
YSNNPPGRVLTLAATLLLGWPLYLAFNVSGRPYDRFACHYDPYGPIFSERERLQIYIA
DLGIFATTFVLYQATMAKGLAWVMRIYGVPLLIVNCFLVMITYLQHTHPAIPRYGSS
EWDWLRGAMVTVDRDYGVLNKVFHNIADTHVAHHLFATVPHYHAMEATKAIKPI
MGEYYRYDGTPFYKALWREAKECLFVEPDEGAPTQGVFWYRNKY

[1574]

[1575] 서열 번호 175

[1576] 리시누스 커뮤니스

MSVTCSKENR HAFFSSSTPG TTSSHSRTRR RPKYNSISTP ASQSFFNSLS SSGSSFQQLM
SSCLAFEPCS HYYSSNGLFP NTPLLPKRHP RLHHRLPRSG EAMAVAVQPE KEVATNKKPL
MKQRRVVVTG MGVVSPLGHD IDVYYNNLLD GSSGISQIDS FDCAQFPTRI AGEIKSFSTD
GWVAPKLSKR MDKFMLYMLT AGKKALADGG ITEDMMDELD KARCGVLIGS AMGGMKVFND
AIEALRISYR KMNPFCVPFA TTNMGSAMLA MDLGWMGPNY SISTACATSN FCILNAANHI
IRGEADIMLC GGSDAAIIPI GLGGFVACRA LSQRNDDPTK ASRPWDMNRD GFVMGEGAGV
LLLEELEHAK KRGANIYAEF LGGSFTCDAY HMTEPRPDGV GVILCIEKAL ARSGVSKEEV
NYINAHATST PAGDLKEYEA LMRCFSQNPD LRVNSTKSMI GHLLGAAGAV EAIATIQAIR
TGWVHPNINL ENPEEGVDTK VLVGPKKERL DIKVALSNSF GFGGHNSSII FAPYK

[1577]

[1578] 서열 번호 176

[1579] 헬리안터스 아너스

MAASMVCTWI VAACMSVSSS SSSNSLFSSK RKRRLDRFSL SSRCRRGGRV AMAIAIQPSS
IEMEEETTLT KRKQPPTKQR RVVVTGMGVE TPIGNNPDQF YNNLLQGVSG ITQIEAFDCS

[1580]

SYPTRIAGEI KNFSTDGWVA PKLSKRMDRF MLYMLTAGKK ALADAGISPS DSDEIDKSRC
GVLIGSAMGG MKVFNDAIEA LRVSYRKMNP FCVPFATTNM GSAMLAMDLG WMGPNYSIST
ACATSNFCIL NAANHIIRGE ADMMLCGGSD AVIIPIGLGG FVACRALSER NTDPAKASRP
WDSGRDGFVM GEGAGVLLLE ELEHAKKRGA KIYAEFLGGS FTCDAYHMTE PHPEGAGVIL
CIEKALSQAG VRREDVNYIN AHATSTPAGD LKEYHALLHC FGNNQELRVN STKSMIGHLL
GAAGAVEAVA TVQAIRTGWI HPNINLENPD QGVDTKVLVG SKKERLNVKV GLSNSFGFGG
HNSSILFAPF Q

[1581]

[1582] 서열 번호 177

[1583] 브라시카 나푸스

MAVAMDMEKE AKVDNKPPTE QRRVVVTGMG VETSLGHDPD TFYENLLQGN SGISQIENFD
CSAFPTRIAG EIKSFSTEGW VAPKLSKRMD KFMLYLLTAG KKALVDGGVT EEVMAEFDKA
KCGVLIGSAM GGMKVFQDAI EAMKISYKKM NPFCVPFATT NMGSAMLALD LGWMGPNYSI
STACATSNFC ILNSANHIIK GEADVMLCGG SDSVIIPIGL GGFVACRALS QRNNDPTKAS
RPWDSNRDGF VMGEGAGVLL LEELEHAKKR GATIYAEFLG GSFTCDAYHM TEPRPDGAGV
ILCIERALAD AGISKEQINY INAHATSTPA GDLKEYQALA HCFGQNPEIK VNSTKSMIGH
LLGAAGAVEA VATVQAIRTG WVHPNINLES PDNGVDTNLL VGPEKERLDI KAALSNSFGF
GGHNSSIIFA PYK

[1584]

[1585] 서열 번호 178

[1586] GenBank 수탁 번호 ABR01158.1, ABI18155.1, AAF61739.1 및 AAW88763.1을 각각 지니는 글리신 맥스

MASTTTSSLC TWLVAACMSV TCHADRTKTP HAMFRSSKKS RYSQFNVCRS THSGKTMAVA
LQPTQEITTI KKPPTKQRRV VVTGLGVVTP LGHEPDIFYN NLLDGASGIS EIETFDCAEY
PTRIAGEIKS FSTDGWVAPK LSKRMDKFML YMLTAGKKAL VDGGITDDVM DELNKEKCGV
LIGSAMGGMK VFNDAIEALR ISYKKMNPFC VPFATTNMGS AMLAMDLGWM GPNYSISTAC
ATSNFCILNA ANHIIRGEAD VMLCGGSDAA IIPIGLGGFV ACRALSQRNT DPTKASRPWD
INRDGFVMGE GAGVLLLEEL EHAKERGATI YAEFLGGSFT CDAYHVTEPR PDGAGVILCI
EKALAQSGVS KEDVNYINAH ATSTPAGDLK EYQALMHCFG QNPELRVNST KSMIGHLLGA
AGGVEAVATI QAIRTGWVHP NINLENPDNG VDAKVLVGSK KERLDVKAAL SNSFGFGGHN
SSIIFAPY

[1587]

[1588] 서열 번호 179

[1589] 피. 모리포르미스

MQTAHQRPPTEGHCFGARLPTASRRAVRRAWSRIARAAAAADANPARPERRVVITGQGVVTS
LGQTIEQFYSSLLEGVSGISQIQKFDTTGYTTTIAGEIKSLQLDPYVPKRWAKRVDDVIKYV
YIAGKQALESAGLPIEAAGLAGAGLDPALCGVLIGTAMAGMTSFAAGVEALTRGGVRKMNPF
CIPFSISNMGGAMLAMDIGFMGPNYSISTACATGNYCILGAADHIRRGDANVMLAGGADAAI
IPSGIGGFIACKALSKRNDEPERASRPWDADRDGFVMGEGAGVLVLEELEHAKRRGATILAE
LVGGAATSDAHHMTEPDPQGRGVRLCLERALERARLAPERVGYVNAHGTSTPAGDVAEYRAI
RAVIPQDSLRINSTKSMIGHLLGGAGAVEAVAAIQALRTGWLHPNLNLENPAPGVDPVVLVG
PRKERAEDLDVVLSNSFGFGGHNSCVIFRKYDE

[1590]

[1591] 서열 번호 180

[1592] 피. 모리포르미스 스테아로일 ACP 불포화 효소

MASAVTFACAPPRGAVAAPGRRAASRPLVVRAVASEAPLGVPPSVQRPSPVVYSKLDKQHRL
TPERLELVQSMGQFAEERVLPVLHPVDKLWQPQDFLPDPESPDFEDQVAELRARAKDLPDEY
FVVLVGDMITEEALPTYMAMLNTLDGVRDDTGAADHPWARWTRQWVAEENRHGDLLNKYCWL
TGRVNMRAVEVTINNLIKSGMNPQTDNNPYLGFVYTSFQERATKYSHGNTARLAAEHGDKNL
SKICGLIASDEGRHEIAYTRIVDEFFRLDPEGAVAAYANMMRKQITMPAHLMDDMGHGEANP
GRNLFADFSAVAEKIDVYDAEDYCRILEHLNARWKVDERQVSGQAAADQEYVLGLPQRFRKL
AEKTAAKRKRVARRPVAFSWISGREIMV

[1593]

[1594] 서열 번호 181

[1595] 리누스 유시타티시멈 12 불포화 효소

MGAGGRMPVP PSSKPMKRSP YSKPPFTLGE LKKAIPPHCF KRSIPRSFAY
VAHDLTIAAI FYYIATTYFH LLPSPLNYLA WPVYWACQGC ILTGVWVLAH
ECGHHAFSDY QWLDDMDGFV LHSSLLVPYF SWKHSHRRHH SNTGSLDRDE
VFVPKQKAEI GWYSKYLNNP PGRVITLAVT LTLGWPLYLA FNVSGRPYDR
FACHFDPHGP IYNDRERMEI YLSDAGIFTV CYILYRLVLT KGLVWVVSIY
GVPLLIVNGF LVLITFLQHT HPSLPHYKSS EWDWMRGALS TVDRDYGLLN
TVFHNITDTH VAHHLFSTMP HYHAMEATKA IKPVLGEYYQ FDGTPFVKAM
WREAKECIYV EPDEGDPSQG VFWYNNKL

[1596]

[1597] 서열 번호 182

[1598] 리누스 유시타티시멈 15 불포화 효소

MSPPNSMSPT TNGNGVAMNG AKKQLDFDPS AAPPFKIADI RAAIPPHCWV KNPWRSLSYV
LRDLLVILSF AVAAAKLDSW TFWPLYWVAQ GTMFWAVFVL GHDCGHGSFS DIWLLNNVMG
HILHSSILVP YHGWRISHKT HHQNHGNVEK DESWVPLPEK VYKSLDTSTK FMRFTIPLPM
FAYPIYLWTR SPGKKGSHFN PYSDLFAPNE RAAVLISTLC WTAMALLLCY SSFIYGFAPV
LKIYGVPYLI FVAWLDMVTY LHHHGYEQKL PWYRGKEWSY LRGGLTTVDR DYGVINNIHH
DIGTHVIHHL FPQMPHYHLV EATQAAKHVL GKYYREPKKS GPFPFHLFGY LVRSLGEDHY
VSDTGDVVFY QSDPHIPKFR TSSATTKSKS S

[1599]

[1600] 서열 번호 183

[1602] *리누스 유시타티시멈 15 불포화 효소

MSPPNSMSPA TNGSTNGVAI NGAKKLLDFD PSAAPPFKIA DIRAAIPPHC
WVKNPWRSLS YVLRDLLVIL SFAVAATKLD SWTVWPLYWI AQGTMFWAVF
VLGHDCGHGS FSDSWLLNNV MGHILHSSIL VPYHGWRISH KTHHQNHGNV
EKDESWVPLP EKVYKSLDTG TKFMRFTIPL PMFAYPIYLW RRSPGKKGSH
FNPYSDLFAP NERTSVMIST LCWTAMALLL CYSSFIYGFL PVFKIYGVPY
LIFVAWLDMV TYLHHHGYEQ KLPWYRGKEW SYLRGGLTTV DRDYGVINNI
HHDIGTHVIH HLFPQMPHYH LVEATQAAKH VLGKYYREPK KSGPFPFHLF
GYLVRSLGED HYVSDTGDVV FYQSDPHIPK FPTSATTKSK SS

[1603]

[1604] 서열 번호 184

[1605] 칼타머스 팅크토리우스 12 불포화 효소

MGAGGRMSNP SEGEKKTELE GIQRVPYQKP PFTVGDVKKA IPPHCFNRSV
IRSFSYVVYD LTIASILYYI ATTFIPLLPH PLAYVAWPIY WAVQGCVMTG
VWVIAHECGH HAFSDYQWLD DTVGLILHSV LLVPYFSWKY SHRRHHSNTG
SIEHDEVFVP KLKSGVRSTA KYLNNPPGRI LTLLVTLTLG WPLYLMFNVS
GRYYDRFACH FDPNSPIYSN RERAQIFISD AGIFAVLYGL YRLAAVKGLV
WVLTVYAGPL LVVNGFLVLI TFLQHTHPSL PHYDSTEWDW LRGALATIDR
DYGILNKVFH NITDTHVTHH LFSTMPHYHA MEATKAIIPI LGDYYQFDGT
SVFKAMYRET KECIYVDKDE EVKDGVYWYR NKI

[1606]

[1607] 서열 번호 185

[1608] GenBank 수탁 번호 ABB05230.1, ABA02173.1, ABA02172.1, ADM48789.1 및 AAL68983.1을 각각 지니는 헬리안투스 아너스 12 불포화 효소

[1609]
MGAGGRMSSP NGKEKDGPKP LERALHEKPP FTVGDIKKVI PPHCFKRSVI RSFSYVVYDL
TIASIFYYLA NNYIPLLPNS LAYVAWPVYW IFQGCVLTGV WVIAHECGHH AFSDYQWLDD

[1610]
TVGLILHSAL LVPYFSWKYS HRRHHSNTGS IEHDEVFVPK LKSSVRSTAK YLNNPPGRIL
TLLVTLTMGW PLYLMFNVSG RYYDRFACHF DPNSPIYSNR ERAQIFISDA GILTVFYILF
RLASTKGLVW VLTMYGGPLL VVNGFLVLIT FLQHTHPSLP HYDSTEWDWL RGALATVDRD
YGILNKVFHN ITDTHVTHHL FSTMPHYHAM EATKAIKPIL GDYYQFDGTS IFKAMYRETK
ECIYVDKDED VKDGVYWYRN KI

[1611] 서열 번호 186

[1612] 쿠페아 라이티 FatB1 티오에스테라제: GenBank 수탁 번호 U56103

[1613]
MVAAAASSAFFSVPTPGTSPKPGKFGNWPSSLSVPFKPDNGGFVKANASAHPKANGS
AVNLKSGSLETPPRSFINQLPDLSMLLSKITTVFGAAEKQWKRPGMLVEPFGVDRIFQ
DGVFFRQSFSIRSYEIGVDRTASIETLMNIFQETSLNHCKSIGLLNDGFGRTPEMCKRD
LIWVVTKIQVEVNRYPTWGDTIEVNTWVSESGKNGMGRDWLISDCRTGEILIRATSV
WAMMNQNTRRLSKFPYEVRQEIAPHFVDSAPVIEDDRKLHKLDVKTGDSIRDGLTPR
WNDLDVNQHVNNVKYIGWILKSVPIEVFETQELCGVTLEYRRECGRDSVLESVTTM
DPAKEGDRCVYQHLLRLEDGADITIGRTEWRPKNAGANGAISSGKTSNGNSVS

[1614] 서열 번호 187

[1615] 쿠페아 라이티 FatB2 티오에스테라제: GenBank 수탁 번호 U56104

[1616]
MVVAAAASSAFFPVPAPRPTPKPGKFGNWPSSLSQPFKPKSNPNGRFQVKANVSPHP
KANGSAVSLKSGSLNTLEDPPSSPPPRTFLNQLPDWSRLRTAITTVFVAAEKQFTRLD
RKSKRPDMLVDWFGSETIVQDGLVFRERFSIRSYEIGADRTASIETLMNHLQDTSLNH
CKSVGLLNDGFGRTPEMCTRDLIWVLTKMQIVVNRYPTWGDTVEINSWFSQSGKIG
MGREWLISDCNTGEILVRATSAWAMMNQKTRRFSKLPCEVRQEIAPHFVDAPPVIED
NDRKLHKFDVKTGDSICKGLTPGWNDFDVNQHVSNVKYIGWILESMPTEVLETQEL
CSLTLEYRRECGRESVVESVTSMNPSKVGDRSQYQHLLRLEDGADIMKGRTEWRPK
NAGTNRAIST

[1617] 서열 번호 188

[1618] 형질 전환 구조물, 6S-CrbTub_suc2_nr::CrbTub_CwFatB1_nr-6S, Const.1으로 명명됨

[1619]
gctcttcgccgccgccactcctgctcgagcgcgcccgcgcgtgcgccgccagcgccttggccttttcgccgcgctcgtgcgcgtcg
ctgatgtccatcaccaggtccatgaggtctgccttgcgccggctgagccactgcttcgtccgggcggccaagaggagcatgaggga
ggactcctggtccagggtcctgacgtggtcgcggctctgggagcgggccagcatcatctggctctgccgcaccgaggccgcctcca
actggtcctccagcagccgcagtcgccgccgaccctggcagaggaagacaggtgaggggggtatgaattgtacagaacaaccacg
agccttgtctaggcagaatccctaccagtcatggctttacctggatgacggcctgcgaacagctgtccagcgaccctcgctgccgccg
cttctcccgcacgcttctttccagcaccgtgatggcgcgagccagcgccgcacgctggcgctgcgcttcgccgatctgaggacagtc
ggggaactctgatcagtctaaaccccccttgcgcgttagtgttgccatcctttgcagaccggtgagagccgacttgttgtgcgccacccc
ccacaccacctcctcccagaccaattctgtcacctttttggcgaaggcatcggcctcggcctgcagagaggacagcagtgcccagcc
gctgggggttggcggatgcacgctca**ggtacc**ctttcttgcgctatgacacttccagcaaaaggtagggcgggctgcgagacggctt
cccggcgctgcatgcaacaccgatgatgcttcgaccccccgaagctccttcggggctgcatgggcgctccgatgccgctccagggc
gagcgctgtttaaatagccaggccccgattgcaaagacattatagcgagctaccaaagcatattcaaacacctagatcactaccactt
ctacacaggccactcgagcttgtgatcgcactccgctaaggggggcgcctcttcccttcgtttcagtcacaacccgcaaac**ggcgcgcc**
atatca***ATG***ctgctgcaggccttcctgttcctgctggccggcttcgccgccaagatcagcgcctccatgacgaacgagacgtccg
accgccccctggtgcacttcacccccaacaagggctggatgaacgaccccaacggcctgtggtacgacgagaaggacgccaa
gtggcacctgtacttccagtacaacccgaacgacaccgtctgggggacgcccttgttctggggccacgccacgtccgacgacctg
accaactgggaggaccagcccatcgccatcgccccgaagcgcaacgactccggcgccttctccggctccatggtggtggactac
aacaacacctccggcttcttcaacgacaccatcgacccgcgccagcgctgcgtggccatctggacctacaacaccccggagtcc
gaggagcagtacatctcctacagcctggacggcggctacaccttcaccgagtaccagaagaaccccgtgctggccgccaactcc

acccagttccgcgacccgaaggtcttctggtacgagccctcccagaagtggatcatgaccgcggccaagtcccaggactacaag
atcgagatctactcctccgacgacctgaagtcctggaagctggagtccgcgttcgccaacgagggcttcctcggctaccagtacg
agtgccccggcctgatcgaggtccccaccgagcaggaccccagcaagtcctactgggtgatgttcatctccatcaaccccggcgc
cccggccggcggctccttcaaccagtacttcgtcggcagcttcaacggcacccacttcgaggccttcgacaaccagtcccgcgtg
gtggacttcggcaaggactactacgccctgcagaccttcttcaacaccgacccgacctacgggagcgcccgggcatcgcgtgg
gcctccaactgggagtactccgccttcgtgcccaccaacccctggcgctcctccatgtccctcgtgcgcaagttctccctcaacacc
gagtaccaggccaacccggagacggagctgatcaacctgaaggccgagccgatcctgaacatcagcaacgccggcccctgg
agccggttcgccaccaacaccacgttgacgaaggccaacagctacaacgtcgacctgtccaacagcaccggcaccctggagtt
cgagctggtgtacgccgtcaacaccacccagacgatctccaagtccgtgttcgcggacctctccctctggttcaagggcctggagg
accccgaggagtacctccgcatgggcttcgaggtgtccgcgtcctccttcttcctggaccgcgggaacagcaaggtgaagttcgtg
aaggagaaccccctacttcaccaaccgcatgagcgtgaacaaccagcccttcaagagcgagaacgacctgtcctactacaaggt
gtacggcttgctggaccagaacatcctggagctgtacttcaacgacggcgacgtcgtgtccaccaacacctacttcatgaccaccg
ggaacgccctgggctccgtgaacatgacgacgggggtggacaacctgttctacatcgacaagttccaggtgcgcgaggtcaag
***TGA*caattggcagcagcagctcggatagtatcgacacactctggacgctggtcgtgtgatggactgttgccgccacacttgct**
gccttgacctgtgaatatccctgccgcttttatcaaacagcctcagtgtgtttgatcttgtgtgtacgcgcttttgcgagttgctagc
tgcttgtgctatttgcgaataccacccccagcatccccttccctcgtttcatatcgcttgcatcccaaccgcaacttatctacgctg
tcctgctatccctcagcgctgctcctgctcctgctcactgcccctcgcacagccttggtttgggctccgcctgtattctcctggtact
gcaacctgtaaaccagcactgcaatgctgatgcacgggaagtagtgggatgggaacacaaatggaggatcccgcgtctcgaa
cagagcgcgcagaggaacgctgaaggtctcgcctctgtcgcacctcagcgcggcatacaccacaataaccacctgacgaatgcgct
tggttcttcgtccattagcgaagcgtccggttcacacacgtgccacgttggcgaggtggcaggtgacaatgatcggtggagctgatggt
cgaaacgttcacag**cctagggatatcgaattc**ctttcttgcgctatgacacttccagcaaaaggtagggcgggctgcgagacggctt
cccggcgctgcatgcaacaccgatgatgcttcgaccccccgaagctccttcggggctgcatgggcgctccgatgccgctccagggc
gagcgctgtttaaatagccaggcccccgattgcaaagacattatagcgagctaccaaagccatattcaaacacctagatcactaccact
tctacacaggccactcgagcttgtgatcgcactccgctaaggggcgcctcttcctcttcgtttcagtcacaacccgcaaac**actagt*A***
***TG*gtggccgccgccgccagcagcgccttcttcagcgtgcccaccccggcaccagccccaagcccggcaagttcggcaactg**
*gcccagcagcctgagcgtgccttcaagcccgacaacggcggcttccacgtgaaggccaacgccagcgcccac***ggcgcgcc***
cccaaggccaacggcagcgccgtgaacctgaagtccggcagcctggagaccccccccgcagcttcatcaaccagctgcccg
acctgagcatgctgctgagcaagatcaccaccgtgttcggcgccgccgagaagcagtggaagcgccccggcatgctggtggag
ccctlcggcgtggaccgcatcttccaggacggcgtgttcttccgccagagcttcagcatccgcagctacgagatcggcgtggaccg
caccgccagcatcgagaccctgatgaacatcttccaggagaccagcctgaaccactgcaagagcatcggcctgctgaacgacg
gcttcggccgcacccccgagatgtgcaagcgcgacctgatctgggtggtgaccaagatccaggtggaggtgaaccgctacccca
cctggggcgacaccatcgaggtgaacacctgggtgagcgagagcggcaagaacggcatgggccgcgactggctgatcagcg
actgccgcaccggcgagatcctgatccgcgccaccagcgtgtgggccatgatgaaccagaacacccgccgcctgagcaagttc
ccctacgaggtgcgccaggagatcgccccccacttcgtggacagcgccccgtgatcgaggacgaccgcaagctgcacaagct
ggacgtgaagaccggcgacagcatccgcgacggcctgaccccccgctggaacgacctggacgtgaaccagcacgtgaacaa
cgtgaagtacatcggctggattctgaagtccgtgcccatcgaggtgttcgagacccaggagctgtgcggcgtgaccctggagtac
cgccgcgagtgcggccgcgacagcgtgctggagagcgtgaccaccatggaccccgccaaggagggcgaccgctgcgtgtacc
agcacctgctgcgcctggaggacggcgccgacatcaccatcggccgcaccgagtggcgccccaagaacgccggcgccaacg
*gcgccatcagcagcggcaagaccagcaacggcaacagcgtgagc**TGA**ttaattaa**ctcgaggcagcagcagctcggatagt***
atcgacacactctggacgctggtcgtgtgatggactgttgccgccacacttgctgccttgacctgtgaatatccctgccgctttta
tcaaacagcctcagtgtgtttgatcttgtgtgtacgcgcttttgcgagttgctagctgcttgtgctatttgcgaataccacccccag
catccccttccctcgtttcatatcgcttgcatcccaaccgcaacttatctacgctgtcctgctatccctcagcgctgctcctgctcct
gctcactgcccctcgcacagccttggtttgggctccgcctgtattctcctggtactgcaacctgtaaaccagcactgcaatgctg
atgcacgggaagtagtgggatgggaacacaaatggaaagcttgagctcttgttttccagaaggagttgctccttgagccttttcattc
tcagcctcgataacctccaaagccgctctaattgtggagggggttcgaatttaaaagcttggaatgttggttcgtgcgtctggaacaagc
ccagacttgttgctcactgggaaaaggaccatcagctccaaaaaacttgccgctcaaaccgcgtacctctgctttcgcgcaatctgccct
gttgaaatcgccaccacattcatattgtgacgcttgagcagtctgtaattgcctcagaatgtggaatcatctgccccctgtgcgagcccat
gccaggcatgtcgcgggcgaggacacccgccactcgtacagcagaccattatgctacctcacaatagttcataacagtgaccatatttc
tcgaagctccccaacgagcacctccatgctctgagtggccaccccccggccctggtgcttgcggagggcaggtcaaccggcatggg
gctaccgaaatccccgaccggatcccaccaccccgcgatgggaagaatctctccccgggatgtgggcccaccaccagcacaacct

[1620]

gctggcccaggcgagcgtcaaaccataccacacaaatatccttggcatcggccctgaattccttctgccgctctgctacccggtgcttct
gtccgaagcaggggttgctagggatcgctccgagtccgcaaacccttgtcgcgtggcggggcttgttcgagctt**gaagagc**

[1621]

[1622] 서열 번호 189

[1623] 형질 전환 구조물, 6S-CrbTub_suc2_nr::CrbTub_CwFatB2_nr-6S, Const.2로 명명됨

actagt*ATG*gtggtggccgccgccgccagcagcgccttcttccccgtgcccgcccccgccccacccccaagcccggcaagtt
cggcaactggcccagcagcctgagccagcccttcaagcccaagagcaaccccaacggccgcttccaggtgaaggccaacgtg
agcccccac**ggcgcgcc**cccaaggccaacggcagcgccgtgagcctgaagtccggcagcctgaacaccctggaggacccc
cccagcagccccccccccgcaccttcctgaaccagctgcccgactggagccgcctgcgcaccgccatcaccaccgtgttcgtg
gccgccgagaagcagttcacccgcctggaccgcaagagcaagcgccccgacatgctggtggactggttcggcagcgagacca
tcgtgcaggacggcctggtgttccgcgagcgcttcagcatccgcagctacgagatcggcgccgaccgcaccgccagcatcgag
accctgatgaaccacctgcaggacaccagcctgaaccactgcaagagcgtgggcctgctgaacgacggcttcggccgcacccc
cgagatgtgcacccgcgacctgatctgggtgctgaccaagatgcagatcgtggtgaaccgctacccccacctggggcgacaccgt
ggagatcaacagctggttcagccagagcggcaagatcggcatgggccgcgagtggctgatcagcgactgcaacaccggcgag
atcctggtgcgccgccaccagcgcctgggccatgatgaaccagaagacccgccgcttcagcaagctgccctgcgaggtgcgcca
ggagatcgccccccacttcgtggacgcccccccgtgatcgaggacaacgaccgcaagctgcacaagttcgacgtgaagaccg
gcgacagcatctgcaagggcctgaccccggctggaacgacttcgacgtgaaccagcacgtgagcaacgtgaagtacatcggc
tggattctggagagcatgcccaccgaggtgctggagacccaggagctgtgcagcctgaccctggagtaccgccgcgagtgcgg
ccgcgagagcgtggtggagagcgtgaccagcatgaacccccagcaaggtgggcgaccgcagccagtaccagcacctgctgcg
cctggaggacggcgccgacatcatgaagggccgcaccgagtggcgccccaagaacgccggcaccaaccgcgccatcagcac
c***TGA***ttaattaa**ctcgag**

[1624]

[1625] 서열 번호 190

[1626] 형질 전환 구조물 6S-CrbTub_suc2_nr::Amt3_CwFatB1_nr-6S (Const.3으로 명명됨) 및 6S-CrbTub_suc2_nr::Amt3_CwFatB2_nr-6S (Const.4로 명명됨)

gaattcggcgctatgacacttccagcaaaaggtagggcgggctgcgagacggcttcccggcgctgcatgcaacaccgatgatgctt
cgaccccccgaagctccttcggggctgcatgggcgctccgatgccgctccagggcgagcgctgtttaaatagccaggccccgatt
gcaaagacattatagcgagctaccaaagccatattcaaacacctagatcactaccacttctacacaggccactcgagcttgtgatcgca
ctccgctaaggggggcgcctcttcctcttcgtttcagtcggccgacaggacgcgcgtcaaaggtgctggtcgtgtatgccctggccggc
aggtcgttgctgctgctggttagtgattccgcaaccctgattttggcgtcttattttggcgtggcaaacgctggcgcccgcgagccgggc
cggcggcgatgcggtgccccacggctgccggaatccaagggaggcaagagcgcccgggtcagttgaagggctttacgcgcaagg
tacagccgctcctgcaaggctgcgtggtggaattggacgtgcaggtcctgctgaagttcctccaccgcctcaccagcggacaaagca
ccggtgtatcaggtccgtgtcatccactctaaagagctcgactacgacctactgatggccctagattcttcatcaaaaacgcctgagaca
cttgcccaggattgaaactccctgaagggaccaccaggggccctgagttgttccttcccccgtggcgagctgccagccaggctgta
cctgtgatcgaggctggcgggaaaataggcttcgtgtgctcaggtcatgggaggtgcaggacagctcatgaaacgccaacaatcgc
acaattcatgtcaagctaatcagctatttcctcttcacgagctgtaattgtcccaaaattctggtctaccgggggtgatccttcgtgtacgg
gcccttccctcaaccctaggtatgcgcgcatgcggtcgccgcgcaactcgcgcgagggccgagggtttgggacgggccgtcccga
aatgcagttgcacccggatgcgtggcacctttttgcgataatttatgcaatggactgctctgcaaaattctggctctgtcgccaaccctag
gatcagcggcgtaggatttcgtaatcattcgtcctgatggggagctaccgactaccctaatatcagcccgactgcctgacgccagcgtc
cacttttgtgcacacattccattcgtgcccaagacatttcattgtggtgcgaagcgtccccagttacgctcacctgtttcccgacctccttac
tgttctgtcgacagagcgggcccacaggccggtcgcagcc**actagt**

[1627]

[1628] 서열 번호 191

[1629] TP1 (UTEX 250 스테아로일 ACP 불포화 효소 운반 펩티드 서열)

actagt*ATG*gccaccgcatccacttctcggcgttcaatgcccgctgcggcgacctgcgtcgctcggcgggctccgggccccgg
cgcccagcgaggcccctccccgtgcgc**gggcgcgcc**

[1630]

[1631] 서열 번호 192

[1632] TP2 (UTEX 1435 스테아로일 ACP 불포화 효소 운반 펩티드 서열)

actagt*ATG*gcttccgcggcattcaccatgtcggcgtgccccgcgatgactggcagggcccctggggcacgtcgctccggacg
gccagtcgccacccgcctgaggg**ggcgcgcc**

[1633]

[1634] 서열 번호 193

[1635] TP3 (UTEX 1435 델타 12 지방산 불포화 효소 운반 펩티드 서열)

actagt*ATG*gctatcaagacgaacaggcagcctgtggagaagcctccgttcacgatcgggacgctgcgcaaggccatccccg
cgcactgtttcgagcgctcggcgcttcgt**ggcgcgcc**

[1636]

[1637] 서열 번호 194

[1638] TP4 (UTEX 1435 이소펜테닐 디포스페이트 합성 효소 운반 펩티드 서열)

[1639] **actagt*ATG***acgttcggggtcgcccctcccggccatgggccgcggtgtctcccttccccggcccagggtcgcggtgcgcgccca gtcggcgagtcaggttttggagagcg**ggcgcgcgcc**

[1640] 서열 번호 195

[1641] 카르타머스 팅크토리우스 올레오일-아실 담체 단백질 티오에스테라제; GenBank 수탁 번호 AAA33019.1

[1642]
MLSKGAPAAP AVAAMYNASA KDTTFALTHS RSIGSVSIRR RYNVFLCNSS SSSRKVSPLL
AVATGEQPSG VASLREADKE KSLGNRLRLG SLTEDGLSYK EKFVIRCYEV GINKTATIET
IANLLQEVGG NHAQGVGFST DGFATTTTMR KLHLIWVTAR MHIEIYRYPA WSDVIEIETW
VQGEGKVGTR RDWILKDYAN GEVIGRATSK WVMMNEDTRR LQKVSDDVRE EYLVFCPRTL
RLAFPEENNN SMKKIPKLED PAEYSRLGLV PRRSDLDMNK HVNNVTYIGW ALESIPPEII
DTHELQAITL DYRRECQRDD IVDSLTSREP LGNAAGVKFK EINGSVSPKK DEQDLSRFMH
LLRSAGSGLE INRCRTEWRK KPAKR

[1643] 서열 번호 196

[1644] 리시누스 커뮤니스 스테아로일-아실 담체 단백질 불포화 효소, GenBank 수탁 번호 AACG59946.1

[1645]
MALKLNRLTFPPTHKLPSQSPFQRSPNRVFMAATVGSTSTKEVENLKKPFSPPREVHV
QVTHSMPPQKMEIFKSLEGWAEDNILVHLKPVEKCWQPQDFLPDPESEGFYEQVKEL
RERARELPDDYFVVLVGDMITEEALPTYQTMLNTLDGVRDETGASLTSWAIWTRAW
TAEENRHGDLLNKYLYLSGRVDMKHIEKTIQYLIGSGMDPKTENNPYLGFIYTSFQE
RATFISHGNTARLAKEHGDMKLAQICGIIAADEKRHETAYTKIAEKLFEIDPDGTVMA
LADMMRKKISMPAHLMFDGQDDNLFDHYSSVAQRIGVYTAKDYADILEFLVGRWK
VENLIGLSGEGRRAQDFVCGLPGRIRRLEERAQGRAKQAATVPFSWIFGRELKL

[1646] 서열 번호 197

[1647] 헬리안투스 아너스 스테아로일-ACP 불포화 효소, GenBank 수탁 번호 AAB65145.1

[1648]
MAIRINTATFQSDLYRSFAFPQPKPLRSPKFAMASTIGSATTKVESTKKPFTPPREVHQ
QVLHSMPPQKIEIFKSMEGWAEDNILVHLKPVEKCWQAQDFLPDPASDGFMEQVEE
LRARAKEIPDDYFVVLVGDMITEEALPTYQTMLNTLDGVRDETGASPTSWAIWTRA

[1649]
WTAEENRHGDLLHQYLYLSGRVDMRQIQKTIQYLIGSGMDPRTENSPYLGFIYTSFQ
ERATFISHGNTARHAKEHGDVKLAQMCGIIAADEKRHETAYTKIVEKLFEIDPDGTV
LAFADMMRKKISMPAHLMYDGRDDNLFENFSAVAQRLGVYTAKDYADILEFLVGR
WKVADLTGLSGEGRKAQDYVCGLAPRIRRLEERNSARAKESVNVPFSWIFDREVKL

[1650] 서열 번호 198

[1651] 브라시카 준세아 색소체 델타-9-스테아로일-아실 담체 단백질 불포화 효소, GenBank 수탁 번호 AAD40245.1

[1652]
MALKLNPLASQPYNFPSSARPPISTFRSPKFLCLASSSSPALSSSTKEVESLKKPFTPPK
EVHVQVLHSMPPQKIEIFKSMEDWAEQNLLTHLKDVEKSWQPQDFLPDPASDGFED
QVKELRERARELPDDYFVVLVGDMITEEALPTYQTMLNTLDGVRDETGASPTPWAV
WTRAWTAEENRHGDLLNKYLYLSGRVDMRQIEKTIQYLIGSGMDPRTENNPYLGFI
YTSFQERATFVSHGNTARQAKEHGDLKLAQICGTIAADEKRHETAYTKIVEKLLEIDP
DGTVVAFADMMRKKISMPAHLMYDGRDDNLFDNFSSVAQRLGVYTAKDYADILEF
LAGRWKIESLSGLSGEGNKAQEYLCGLTPRIRRLDERAQARAKKGPKIPFSWIHDRE
VQL

[1653] 서열 번호 199

[1654] 글리신 맥스 스테아로일-ACP 불포화 효소, GenBank 수탁 번호 ACJ39209.1

[1655]
MALRLNPIPTQTFSLPQMASLRSPRFRMASTLRSGSKEVENIKKPFTPPREVHVQVTH
SMPPQKIEIFKSLEDWAEQNILTHLKPVEKCWQPQDFLPDPSSDGFEEQVKELRERAK
ELPDDYFVVLVGDMITEEALPTYQTMLNTLDGVRDETGASLTSWAIWTRAWTAEEN
RHGDLLNKYLYLSGRVDMKQIEKTIQYLIGSGMDPRTENSPYLGFIYTSFQERATFIS
HGNTARLAKEHGDIKLAQICGMIASDEKRHETAYTKIVEKLFEVDPDGTVMAFADM
MRKKIAMPAHLMYDGRDDNLFDSYSSVAQRIGVYTAKDYADILEFLVGRWKVEQL
TGLSGEGRKAQEYICGLPPRIRRLEERAQARVKESSTLKFSWIHDREVLL

[1656] 서열 번호 200

[1657] 올레아 유로파에아 스테아로일-ACP 불포화 효소, GenBank 수탁 번호 AAB67840.1

MALKLCFPPHKMPSFPDARIRSHRVFMASTIHSPSMEVGKVKKPFTPPREVHVQVTH
SLAPEKREIFNSLNNWAQENILVLLKDVDKCWQPSDFLPDSASEGFDEQVMELRKRC
KEIPDDYFIVLVGDMITEEALPTYQTMLNTLDGVRDETGASLTPWAIWTRAWTAEEN
RHGDLLNKYLYLSGRVDMKQIEKTIQYLIGSGMDPRTENNPYLGFIYTSFQERATFIS
HGNTARLAKEHGDLKLAQICGIIAADEKRHETAYTKIVEKLFEIDPDGTVLALADMM
RKKVSMPAHLMYDGQDDNLFENFSSVAQRLGVYTAKDYADILEFLVGRWDIEKLT
GLSGEGRKAQDYVCTLPPRIRRLEERAQSRVKKASATPFSWIFGREINL

[1658]

[1659] 서열 번호 201

[1660] 베르니시아 포르디 스테아로일-아실-담체 단백질 불포화 효소, GenBank 수탁 번호 ADC32803.1

MALKLNPFISQSHKFPTFALPPMANLRSPKFYMASTLREIEHLKKPFMPPREVHVQVT
HSMPPQKIEIFKSLEDWAEQNILVHLKPVEKCWQPQDFLPDPASDDFDEQVKELRER
AKEIPDDYFVVLVGDMITEEALPTYQTMLNTLDGVRDETGASLTSWAIWTRAWTAE
ENRHGDLLNKYLYLSGRVDMRQIEKTIQYLIGSGMDPRTENSPYLGFIYTSFQERATF
ISHGNTARHAKEHGDIKLAQICGTIASDEKRHETAYTKIVEKLFEIDPNGTVLAFADM
MRKKISMPAHLMYDGCNDNLFDHFSAVAQRLGVYTAKDYADILEFLVGRWKVDKL
MGLSAEGQKAQDYVCRLPPRIRRLEERAQGRAKEATIVPFSWIFEREVQL

[1661]

[1662] 서열 번호 202

[1663] 헬리안투스 아너스 아실-ACP 티오에스테라제 FATA1, GenBank 수탁 번호 AAL79361.1

MLSRGVPTATATAYNDVIKETARSVTNSRSIDSVSIRRRNNGFVSNSLCRRVAPVMA
VKVDEQRTGVAVDVTEKRLADRLRMGSLTEDGLSYKERFIIRCYEVGINKTATVETI
ANLLQEVGGNHAQSVGFSTDGFATTTTMRKLNLIWVTSRMHIEIYRYPAWSDVVEIE
TWCQGEGRIGTRRDWIIKDHANGEVIGRATSKWVMMNSETRRLQKVNDDIRDEYLI
FCPKTLRLAFPEENNNSLKKIAKLEDPAECSTLGLVPRRADLDMNKHVNNVTYIGWV
LESIPQEVIDTHELQTITLDYRRECQHDDVVDSLTSSESPAVNGCASGENLSKFLHLLR
SSGEGLELNRGRTEWRKKPAKK

[1664]

[1665] 서열 번호 203

[1666] 브라시카 라파 FATA1, GenBank 수탁 번호 AAC49002.1

MLKLSCNATDKLQTLFSHSHQPDPAHRRTVSSVSCSHLRKPVLDPLRAIVSADQGSVI
RAEQGLGSLADQLRLGSLTEDGLSYKEKFIVRSYEVGSNKTATVETVANLLQEVGCN
HAQSVGFSTDGFATTPTMRKLHLIWVTARMHIEIYKYPAWGDVVEIETWCQSEGRIG
TRRDWILKDVATGEVTGRATSKWVMMNQDTRRLQKVSDDVRDEYLVFCPKELRLA
FPEENNRSLKKIPKLEDPAQYSMIGLKPRRADLDMNQHVNNVTYIGWVLESIPQEIVD
THELQVITLDYRRECQQDDVVDSLTTTTSEIGGTNGSASSGTQGQNDSQFLHLLRLSG
DGQEINRGTTLWRKKPSNL

[1667]

[1668] 서열 번호 204

[1669] 자트로파 컬카스 아실-ACP 티오에스테라제, GenBank 수탁 번호 ABX82799.3

MMLKVPCTAPDQIQSLAPQCRFLNQHRSSSHFTRRRANAVVYSSPVAKISTQPVVSD
GVTKVESGCESLADRLRLGHLTEDGFSYKEKFIVRSYEVGINKTATVETIANFLQEVG
CNHAQSVGFSTDGFATTPTMRKLHLIWVTARMHIEIYKYPAWSDVVEIETWCQSEG
RIGTRRDWILKDYATGQVIGRATSKWVMMNQDTRRLQKVTDDVRDEYLVFCPREL
RLAFPEENNRSLKKISKLEDTTQYSKLGLVPRRADLDMNQHVNNVTYIGWVLESMP
QEIIDTHELQTITLDYRRECQHDDVVDSLTSAESLEGAGLDLHATNGSATAIAGEQDS
RNFLHLLRFSSDGLEINRGRTEWRKKPSR

[1670]

[1671] 서열 번호 205

[1672] 제아 메이스 올레오일-아실 담체 단백질 티오에스테라제, GenBank 수탁 번호 ACG40089.1

MRKLGLIWVTNRMHIEIYKYPAWGDVVEIETWCQEDGKIGTRRDWILKDLCTGEVT
GRATSKWVMMNQNTRRLQRVSDDVRDEVFIHCPKTPRLAFPEENNGSLKKIPNLSDP
AQYSRLGLVPRRADLDMNQHVNNVTYIGWVLESIPQDIIDTHELQTITLDYRRECQH
DDIVDSLTYVEEGEERSMNGSASSVPHTEQRRQFLHCLRFAANGDEINRGRTVWRKL
AR

[1673]

[1674] 서열 번호 206

[1675] 제아 메이스 올레오일-아실 담체 단백질 티오에스테라제, GenBank 수탁 번호 ACG42559.1

MLRCHTPPQCARAPLRHHGRWESPPAAAPAVVVRCARGAPQVSGIEAASPGHAAVT
AALAKAEGGDARPSLAERLRLGNLLEDGLSYKESFIVRCYEVGINKTATVETIANLLQ
EVGCNHAQSVGFSTDGFATTTTMRKLGLIWVTNRMHIEIYKYPAWGDFVEIETWCQ
EDGRIGTRRDWILKDLANGEVIGRATSKWVMMNQNTRRLQRVSDDVRDEVFMHCP
KAPRLAFPEENNGSLKKIPNLSDPAEYSRLGLVPRRADLDMNQHVNNVTYIGWVLES

[1676]

IPQDIIDTHELQTITLDYRRECQQDDIVDSLTCIEEGEEKSMNGSASAAAPHKEERQQF
[1677] LHCLRFAANGHEINRGRTVWRKLAR

[1678] 서열 번호 207

[1679] 칼타머스 팅크토리우스 델타 12 지방산 CtFAD2-3, GenBank 수탁 번호 ADM48790.1

MGAGGRMNVAVSGSEKPHAFKRVPVSKPPFELSDLKKAVPPHCFKRSLVRSFAALF
RDIIIVTALYYLAATIIPVLPKPLTYVAWPLYWFFQGAYLMGLWVIGHECGHHGFSEY
QWLDDTVGFIVHSFILTPYFGFKYSHRTHHANTNSIEYDEVWIPKRKSDKLYSEVLNN
PLGSFIVFVFKIVLGFPLYFVFNLYGRKYEKGITSHFYPYSPIFNDSERFQIFLTDLGVF
GTLYGVYRLALIKGTEWVINFYGMPILFMSGSFILLTYLHHTHPSIPHYDSTEWDWLR
GALATVDRNFGFLNHAFHDATRTHAVHHLFPTIPHYHTFEARQAVMPILGDYYKYD
DTPILQAIWRETKDCTFIEPEEVNGEKKGIYWFYK

[1680]

[1681] 서열 번호 208

[1682] 고시피움 히르서텀 오메가-6 불포화 효소, GenBank 수탁 번호 CAA71199.1

MGAGGRMSVPTSPKKPEFNSLKRVPYSKPPFTLSEIKKAIPPHCFQRSVLRSFSYLLYD
FILASLFYHVATNYFPNLPQALSNVAWPLYWAMQGCILTGVWVIAHECGHHAFSDY
QWLDDTVGLILHSSLLVPYFSWKYSHRRHHSNTGSLERDEVFVPKKKSGLRWWAK
HFNNPPGRFLSITIQLTLGWPLYLAFNVAGRPYDRFACHYDPYGPIFSDRERLQIYISD
AGVLAVAYALYRLVLAKGVGWVISVYGVPLLVVNAFLVMITYLQHTHPSLPHYDSS
EWDWMRGALSTVDRDYGILNKVFHNITDTHVAHHLFSTMPHYHAMVATKAIKPILG
EYYQFDGMPVYKAIWREAKECLYVEPDEGDKDKGVFWFRNKL

[1683]

[1684] 서열 번호 209

[1685] 글리신 맥스 마이크로좀 불포화 효소 GmFAD2-2a, GenBank 수탁 번호 BAD89862.1

MGAGGRTDVPPANRKSEVDPLKRVPFEKPPFSLSQIKKVIPPHCFQRSVFRSFSYVVY
DLTIAFCLYYVATHYFHLLPSPLSFLAWPIYWAVQGCILTGVWVIAHECGHHAFSDY
QLLDDIVGLVLHSGLLVPYFSWKYSHRRHHSNTGSLERDEVFVPKQKSCIKWYSKYL
NNPPGRVLTLAVTLTLGWPLYLALNVSGRPYDRFACHYDPYGPIYSDRERLQIYISDA
GVLAVCYGLFRLAMAKGLAWVVCVYGVPLLVVNGFLVLITFLQHTHPALPHYTSSE
WDWLRGALATVDRDYGILDKVFHNITDTHVAHELFSTMPHYHAMEATKAIKPILGE
YYRFDETPFVKAMWREARECIYVEPDQSTESKGVFWYNNKL

[1686]

[1687] 서열 번호 210

[1688] 제아 메이스 지방산 불포화 효소 FAD2, GenBank 수탁 번호 ABF50053.1

MGAGGRMTEKEREKQEQLARATGGAAMQRSPVEKPPFTLGQIKKAIPPHCFERSVL
KSFSYVVHDLVIAAALLYFALAIIPALPSPLGYAAWPLYWIAQGCVCTGVWVIAHEC
GHHAFSDYSLLDDVVGLVLHSSLMVPYFSWKYSHRRHHSNTGSLERDEVFVPKKKE
ALPWYTPYVYNNPVGRLAHIVVQLTLGWPLYLATNASGRPYPRFACHFDPYGPIYN
DRERAQIFVSDAGVAAVAFGLYKLAAAFGVWWVVRVYAVPLLIVNAWLVLITYLQ
HTHPSLPHYDSSEWDWLRGALATMDRDYGILNRVFHNITDTHVAHHLFSTMPHYHA
MEATKAIRPILGDYYHFDPTPVAKATWREAGECIYVEPEDRKGVFWYNKKF

[1689]

[1690] 서열 번호 211

[1691] 프로토테카 모리포르미스 d12 불포화 효소 대립유전자 1 CHP - check with Riyaz

MAIKTNRQPVEKPPFTIGTLRKAIPAHCFERSALRSSMYLAFDIAVMSLLYVASTYIDP
APVPTWVKYGVMWPLYWFFQGAFGTGVWVVCAHECGHQAFSSSQAINDGVGLVFH
SLLLVPYYSWKHSHRRHHSNTGCLDKDEVFVPPHRAVAHEGLEWEEWLPIRMGKV
LVTLTLGWPLYLMFNVASRPYPRFANHFDPWSPIFSKRERIEVVISDLALVAVLSGLS
VLGRTMGWAWLVKTYVVPYLIVNMWLVLITLLQHTHPALPHYFEKDWDWLRGAM
ATVDRSMGPPFMDNILHHISDTHVLHHLFSTIPHYHAEEASAAIRPILGKYYQSDSRW
VGRALWEDWRDCRYVVPDAPEDDSALWFHK

[1692]

[1693] 서열 번호 212

[1694] 프로토테카 모리포르미스 델타 12 불포화 효소 대립유전자 2

MAIKTNRQPVEKPPFTIGTLRKAIPAHCFERSALRSSMYLAFDIAVMSLLYVASTYIDP
APVPTWVKYGIMWPLYWFFQGAFGTGVWVVCAHECGHQAFSSSQAINDGVGLVFHS
LLLVPYYSWKHSHRRHHSNTGCLDKDEVFVPPHRAVAHEGLEWEEWLPIRMGKVL
VTLTLGWPLYLMFNVASRPYPRFANHFDPWSPIFSKRERIEVVISDLALVAVLSGLSV
LGRTMGWAWLVKTYVVPYMIVNMWLVLITLLQHTHPALPHYFEKDWDWLRGAM
ATVDRSMGPPFMDSILHHISDTHVLHHLFSTIPHYHAEEASAAIRPILGKYYQSDSRW
VGRALWEDWRDCRYVVPDAPEDDSALWFHK

[1695]

[1696] 서열 번호 213

[1697] 브라시카 나파 리놀레산 불포화 효소 FAD3, GenBank 수탁 번호 AAA32994.1

MVVAMDQRSNVNGDSGARKEEGFDPSAQPPFKIGDIRAAIPKHCWVKSPLRSMSYV
TRDIFAVAALAMAAVYFDSWFLWPLYWVAQGTLFWAIFVLGHDCGHGSFSDIPLLN
SVVGHILHSFILVPYHGWRISHRTHHQNHGHVENDESWVPLPEKLYKNLPHSTRMLR
YTVPLPMLAYPIYLWYRSPGKEGSHFNPYSSLFAPSERKLIATSTTCWSIMLATLVYL
SFLVDPVTVLKVYGVPYIIFVMWLDAVTYLHHHGHDEKLPWYRGKEWSYLRGGLT
TIDRDYGIFNNIHHDIGTHVIHHLFPQIPHYHLVDATRAAKHVLGRYYREPKTSGAIPI
HLVESLVASIKKDHYVSDTGDIVFYETDPDLYVYASDKSKIN

[1698]

[1699] 서열 번호 214

[1700] 카멜리나 사티바 오메가-3 FAD7-2

MANLVLSECGIRPLPRIYTTPRSNFVSNNNKPIFKFRPFTSYKTSSSPLA
CSRDGFGKNWSLNVSVPLTTTTPIVDESPLKEEEEEKQRFDPGAPPPFNL
ADIRAAIPKHCWVKNPWKSMSYVLRDVAIVFALAAGASYLNNWIVWPLYW
LAQGTMFWALFVLGHDCGHGSFSNNPRLNNVVGHLLHSSILVPYHGWRIS
HRTHHQNHGHVENDESWHPMSEKIYQSLDKPTRFFRFTLPLVMLAYPFYL
WARSPGKKGSHYHPESDLFLPKEKTDVLTSTACWTAMAALLICLNFVVGP
VQMLKLYGIPYWINVMWLDFVTYLHHHGHEDKLPWYRGKEWSYLRGGLTT
LDRDYGVINNIHHDIGTHVIHHLFPQIPHYHLVEATEAVKPVLGKYYREP
DKSGPLPLHLLGILAKSIKEDHYVSDEGDVVYYKADPNMYGEIKVGAD

[1701]

[1702] 서열 번호 215

[1703] 카멜리나 사티바 오메가-3 FAD7-1

MANLVLSECGIRPLPRIYTTPRSNFVSNNNKPIFKFRPLTSYKTSSPLFC
SRDGFGRNWSLNVSVPLATTTPIVDESPLEEEEEEEKQRFDPGAPPPFNL
ADIRAAIPKHCWVKNPWKSMSYVLRDVAIVFALAAGAAYLNNWIVWPLYW
LAQGTMFWALFVLGHDCGHGSFSNNPRLNNVVGHLLHSSILVPYHGWRIS

[1704]

HRTHHQNHGHVENDESWHPMSEKIYQSLDKPTRFFRFTLPLVMLAYPFYL
WARSPGKKGSHYHPESDLFLPKEKTDVLTSTACWTAMAALLICLNFVVGP
VQMLKLYGIPYWINVMWLDFVTYLHHHGHEDKLPWYRGKEWSYLRGGLTT
LDRDYGVINNIHHDIGTHVIHHLFPQIPHYHLVEATEAVKPVLGKYYREP
DKSGPLPLHLLGILAKSIKEDHYVSDEGDVVYYKADPNMYGEIKVGAD

[1705]

[1706] 서열 번호 216

[1707] 글리신 맥스 엽록체 오메가 3 지방산 불포화 효소 아이소형 2, GenBank 수탁 번호 ACF19424.1

MATWVLSECSLKPLAPVIPRPRTGAVLSSTSKVGFLETNKVLEGSKFQPL
RCNLRERNWGLKVSVPLRIASIEEEEQKSVDVINGSNGVEHEKLPEFDPG
APPPFNLADIRAAIPKHCWVKDPLKSMSYVVRDVIAVFGLASAAAYLNNW
LVWPLYWAAQGTMFWALFVLGHDCGHGSFSNNPKLNSVVGHLLHSSILVP
YHGWRISHRTHHQHHGHVENDESWHPLPEKLFRSLDTVTRMLRFTAPFPL
LAYPVYLWGRSPGKTGSHFDPSSDLFVPNERKDVITSTACWAAMLGLLVG
LGFVMGPIQLLKLYGVPYVIFVMWLDLVTYLHHHGHEDKLPWYRGKEWSY
LRGGLTTLDRDYGLINNIHHDIGTHVIHHLFPQIPHYHLVEATEAAKPVF
GKYYREPKKSSPLPFHLIGELIKSFKTDHFVSDKGDVVYYQTDSEINGSS
KSE

[1708]

[1709] 서열 번호 217

[1710] 베르니시아 포르디 오메가-3 지방산 불포화 효소, GenBank 수탁 번호 AAF12821.1

MAVWALSDCGIRPLPKIYSKPRLAFTSNNPQPTKPPIPRPELRNGSSFKL
SSAGLRDKWVVNLSAPLRGGLVEEDEDNFEGNRVISIDESGGEDFDPDAP
PPDTLADIRAAIPKHRWTKNPEISLSYVVRNVAVVSGLAAVAAYFNDWAF
WSPYWICQGTMLSALFVLGHDCGHGSFSNNPNLDSVVGHLLHYSILVPYH
GWRISYRTHYRHHQHAENDESWHPLSEKIYKGLNNVTRTLRLSLPFSLLA
YPSYLWSRSPGKKGFHFHPSSDLFVPSEKKDVITSTVCWTAMLALLVGSS
FVMGLVQLLKPYGIPYLIFVMWLDLVTYLHHHGHENKIPWYRGEEWSYLR
GRLTTVGRDYGWFNKIHHDIGIHVIHHLFPQIPHYHLIEATEAAKPVLGK
YYREPKKSGPLPFHLLGSLIRSMKEDHYVSDAGDVVYYQKEPKLSGARCL
S

[1711]

[1712] 서열 번호 218

[1713] 리시누스 커뮤니스 오메가-3 지방산 불포화 효소, 소포체, 추정, GenBank 수탁 번호 EEF36775.1

MSAVMLARATNVAANNNVTSRTGSHERPHTPAAAAHFNLSLCLLSPFKGS
PLQLEGILLMKKADSVNAFDALDGDADFDPSDPPPFKIAEIRAAIPKHCW
IKNPWRSLSYVLRDAFVVLALAAAALFLDSWCFWPLYWAAQGTMFWAIFV
LGHDCGHGSFSDSTLLNNVVGHLLHSSILVPYHGWRISHRTHHQNHGNVD
KDESWVPLPEKIYNSLDVSTRIMRFTVPLPLLAYPIYLWTRGPGKDGSHF
SPYSDLFAPNERRAVLTSTVCWILMILLLLYSCFVFGSVQVLKLYGVPYI
IFIMWLDFVTYLHHHGHEQKLPWYRGKEWSYLRGGLTTVDRDYGWINNIH
HDIGTHVIHHLFPQIPHYHLVEATKAAKPVLGKYYREPRKSGPFPLHLFE
NLVKSISQDHYVSDIGDVVYYQTDPELYKKLAKTKDVLEDLESCDALGIL
KCKNCSAKSGSSYSALQFLELKGRAAGAEKGREWKILVFEILTLHHRIIA
QMSFMG

[1714]

[1715] 서열 번호 219

[1716] 리넘 유시타티시멈 지방산 불포화 효소 3C, GenBank 수탁 번호 ADV92272.1

MSPPNSMGMEAAHPTGNGNGVAVMNGASANKPDFDPSEAPPFKIADIRAA
IPPHCWVKNPWRSLSYVLRDAVVILAFAAAALKLDLWAVWPLYWIAQGTM
FWAVFVLGHDCGHGSFSDSWWLNNVVGHILHSAILVPYHGWRISHKTHHQ
NHGNVEKDESWVPLPEKVYKTLDTSTKFMRFTIPLPMFAYPIYLWTRSPG
KKGSHFNPYSDLFAPQERKSVLISTISWISMVLILLYASFLFGFLTVFKV
YTVPYLIFVAWLDMVTYLHHHGHEEKLPWYRGQEWSYLRGGLTTVDRDYG
IINNIHHDIGTHVIHHLFPQIPHYHLVEATKAAKSVLGKYYREPKKSGPF
PFHLFDNLVRSLGEDHYVSDAGDVVFYQSDPEIFKFSKSKSA

[1717]

[1718] 서열 번호 220

[1719] 프로토테카 모리포르미스 FAD-A 리놀리에이트 불포화 효소

MFWALFVVGHDCGHQSFSTSKRLNDAVGLFVHSIIGVPYHGWRISHRTHHNNHGHV
ENDESWYPPTESGLKAMTDMGRQGRFHFPSMLFVYPFYLFWRSPGKTGSHFSPATD
LFALWEAPLIRTSNACQLAWLGALAAGTWALGVLPMLNLYLAPYVISVAWLDLVT
YLHHHGPSDPREEMPWYRGREWSYMRGGLTTIDRDYGLFNKVHHDIGTHVVHHLF
PQIPHYNLCRATKAAKKVLGPYYREPERCPLGLLPVHLLAPLLRSLGQDHFVDDAGS
VLFYRRAEGINPWIQKLLPWLGGARRGADAQRDAAQ

[1720]

[1721] 서열 번호 221

[1722] 프로토테카 모리포르미스 FAD-D 오메가 3 불포화 효소

MSIQFALRAAYIKGTCQRLSGRGAALGLSRDWTPGWTLPRCWPASAAATAPPRARH
QERAIHLTSGRRRHSALASDADERALPSNAPGLVMASQANYFRVRLLPEQEEGELES
WSPNVRHTTLLCKPRAMLSKLQMRVMVGDRVIVTAIDPVNMTVHAPPFDPLPATRF
LVAGEAADMDITVVLNKADLVPEEESAALAQEVASWGPVVLTSTLTGRGLQELERQ
LGSTTAVLAGPSGAGKSSIINALARAARERPSDASVSNVPEEQVVGEDGRALANPPPF
TLADIRNAIPKDCFRKSAAKSLAYLGDLSITGMAVLAYKINSPWLWPLYWFAQGTM
FWALFVVGHDCGHQSFSTSKRLNDALAWLGALAAGTWTWALGVLPMLNLYLAPY
VWLLVTYLHHHGPSDPREEMPWYRGREWSYMRGGLTTIDRDYGLFNKVHHDIGTH
VVHH

[1723]

[1724] 서열 번호 222

[1725] Bn FAD3-3xFLAG에 대한 pSZ2124 형질 전환 서열의 부분, 피. 모리포르미스에서의 발현을 위해 코돈-최적화됨

gctcttcgccgccgccactcctgctcgagcgcgcccgcgcgtgcgccgccagcgccttggccttttcgccgcgctcgtgcgcgtcgc
tgatgtccatcaccaggtccatgaggtctgccttgcgccggctgagccactgcttcgtccgggcggccaagaggagcatgaggga
ggactcctggtccagggtcctgacgtggtcgcggctctgggagcgggccagcatcatctggctctgccgcaccgaggccgcctcca
actggtcctccagcagccgcagtcgccgccgaccctggcagaggaagacaggtgaggggggtatgaattgtacagaacaaccca
cgagccttgtctaggcagaatccctaccagtcatggctttacctggatgacggcctgcgaacagctgtccagcgaccctcgctgcc
gccgcttctcccgcacgcttctttccagcaccgtgatggcgcgagccagcgccgcacgctggcgctgcgcttcgccgatctgagga
cagtcggggaactctgatcagtctaaaccccccttgcgcgttagtgttgccatcctttgcagaccggtgagagccgacttgttgtgcg
ccacccccccacaccacctcctcccagaccaattctgtcacctttttggcgaaggcatcggcctcggcctgcagagaggacagcagt
gcccagccgctgggggttggcggatgcacgctcaggtacc<u>ctttcttgcgctatgacacttccagcaaaaggtagggcgggctgcgag</u>
<u>acggcttcccggcgctgcatgcaacaccgatgatgcttcgaccccccgaagctccttcggggctgcatgggcgctccgatgccgctccagg</u>
<u>gcgagcgctgtttaaatagccaggcccccgattgcaaagacattatagcgagctaccaaagccatattcaaacacctagatcactaccactt</u>

[1726]

ctacacaggccactcgagcttgtgatcgcactccgctaaggggcgcctcttcctcttcgtttcagtcacaacccgcaaactctagaatatca
*ATG*ctgctgcaggccttcctgttcctgctggccggcttcgccgccaagatcagcgcctccatgacgaacgagacgtccgaccgcccct
ggtgcacttcaccccaacaagggctggatgaacgaccccaacggcctgtggtacgacgagaaggacgccaagtggcacctgtactt
ccagtacaacccgaacgacaccgtctgggggacgcccttgttctggggccacgccacgtccgacgacctgaccaactgggaggacca
gcccatcgccatcgccccgaagcgcaacgactccggcgccttctccggctccatggtggtggactacaacaacacctccggcttcttcaa
cgacaccatcgacccgcgccagcgctgcgtggccatctggacctacaacacccccggagtccgaggagcagtacatctcctacagcctg
gacggcggctacaccttcaccgagtaccagaagaaccccgtgctggccgccaactccacccagttccgcgacccgaaggtcttctggt
acgagcccctcccagaagtggatcatgaccgcggccaagtcccaggactacaagatcgagatctactcctccgacgacctgaagtcctg
gaagctggagtccgcgttcgccaacgagggcttcctcggctaccagtacgagtgccccggcctgatcgaggtccccaccgagcaggac
cccagcaagtcctactgggtgatgttcatctccatcaaccccggcgccccggccggcggctccttcaaccagtacttcgtcggcagcttca
acggcacccacttcgaggccttcgacaaccagtcccgcgtggtggacttcggcaaggactactacgccctgcagaccttcttcaacacc
gacccgacctacgggagcgccctgggcatcgcgtgggcctccaactgggagtactccgccttcgtgcccaccaaccccctggcgctcctc
catgtccctcgtgcgcaagttctccctcaacaccgagtaccaggccaacccggagacggagctgatcaacctgaaggccgagccgatc
ctgaacatcagcaacgccggcccctggagccggttcgccaccaacaccacgttgacgaaggccaacagctacaacgtcgacctgtcc
aacagcaccggcaccctggagttcgagctggtgtacgccgtcaacaccacccagacgatctccaagtccgtgttcgcggacctctccct
ctggttcaaggcctggaggaccccgaggagtacctccgcatgggcttcgaggtgtccgcgtcctccttcttcctggaccgcgggaaca
gcaaggtgaagttcgtgaaggagaaccctacttcaccaaccgcatgagcgtgaacaaccagcccttcaagagcgagaacgacctgt
cctactacaaggtgtacggcttgctggaccagaacatcctggagctgtacttcaacgacggcgacgtcgtgtccaccaacacctacttc
atgaccaccgggaacgccctgggctccgtgaacatgacgacgggggtggacaacctgttctacatcgacaagttccaggtgcgcgag
gtcaag*TGA*caattggcagcagcagctcggatagtatcgacacactctggacgctggtcgtgtgatggactgttgccgccacacttgctg
ccttgacctgtgaatatccctgccgcttttatcaaacagcctcagtgtgtttgatcttgtgtgtacgcgcttttgcgagttgctagctgcttgtgct
atttgcgaataccaccccccagcatccccttccctcgtttcatatcgcttgcatcccaaccgcaacttatctacgctgtcctgctatccctcagcg
ctgctcctgctcctgctcactgcccctcgcacagccttggtttgggctccgcctgtattctcctggtactgcaacctgtaaaccagcactgcaat
gctgatgcacgggaagtagtgggatgggaacacaaatggaggatccgcgtctcgaacagagcgcgcagaggaacgctgaaggtctcg
cctctgtcgcacctcagcgcggcatacaccacaataaccacctgacgaatgcgcttggttcttcgtccattagcgaagcgtccggttcacaca
cgtgccacgttggcgaggtggcaggtgacaatgatcggtggagctgatggtcgaaacgttcacagcctagggatatcgaattcggccgac
aggacgcgcgtcaaaggtgctggtcgtgtatgccctggccggcaggtcgttgctgctgctggttagtgattccgcaaccctgattttggc
gtcttattttggcgtggcaaacgctggcgcccgcgaggcgggccggcggcgatgcggtgccccacggctgccggaatccaagggagg
caagagcgcccgggtcagttgaagggctttacgcgcaaggtacagccgctcctgcaaggctgcgtggtggaattggacgtgcaggtc
ctgctgaagttcctccaccgcctcaccagcggacaaagcaccggtgtatcaggtccgtgtcatccactctaaagaactcgactacgacc
tactgatggccctagattcttcatcaaaaacgcctgagacacttgcccaggattgaaactccctgaagggaccaccaggggccctgag
ttgttccttccccccgtggcgagctgccagccaggctgtacctgtgatcgaggctggcgggaaaataggcttcgtgtgctcaggtcatgg
gaggtgcaggacagctcatgaaacgccaacaatcgcacaattcatgtcaagctaatcagctatttcctcttcacgagctgtaattgtcc
caaaattctggtctaccgggggtgatccttcgtgtacgggcccttccctcaaccctaggtatgcgcgcatgcggtcgccgcgcaactcgc
gcgagggccgagggtttgggacgggccgtcccgaaatgcagttgcacccggatgcgtggcaccttttttgcgataatttatgcaatgg
actgctctgcaaaattctggctctgtcgccaaccctaggatcagcggcgtaggatttcgtaatcattcgtcctgatggggagctaccgac
taccctaatatcagcccgactgcctgacgccagcgtccacttttgtgcacacattccattcgtgcccaagacatttcattgtggtgcgaag
cgtccccagttacgctcacctgtttcccgacctccttactgttctgtcgacagagcgggcccacaggccggtcgcagccactagt*ATGg*
tggtggccatggaccagcgctccaacgtgaacggcgactccggcgcccgcaaggaggagggcggcggcgccatggactac
aaggaccacgacggcgactacaaggaccacgacatcgactacaaggacgacgacgacaagttcgacccctccgcccagcc
ccccttcaagatcggcgacatccgcgccgccatccccaagcactgctgggtgaagtccccctgcgctccatgtcctacgtgacc
cgcgacatcttcgccgtggccgccctggccatggccgccgtgtacttcgactcctggttcctgtggccctgtactgggtggccca
gggcacccctgttctgggccatcttcgtgctgggccacgactgcggccacggctccttctccgacatcccccgctgaactccgtgg
tgggccacatcctgcactccttcatcctggtgccctaccacggctggcgcatctcccaccgcacccaccaccagaaccacggcca
cgtggagaacgacgagtcctgggtgcccctgcccgagaagctgtacaagaacctgccccactccacccgcatgctgcgctac
accgtgcccctgcccatgctggcctaccccatctacctgtggtaccgctccccggcaaggagggctcccacttcaaccctactc
ctccctgttcgccccctccgagcgcaagctgatcgccacctccaccacctgctggtccatcatgctggccaccctggtgtacctgtc
cttcctggtggaccccgtgaccgtgctgaaggtgtacggcgtgccctacatcatcttcgtgatgtggctggacgccgtgacctac
ctgcaccaccacggccacgacgagaagctgccctggtaccgcggcaaggagtggtcctacctgcgcggcggcctgaccacca
tcgaccgcgactacggcatcttcaacaacatccaccacgacatcggcacccacgtgatccaccacctgttcccccagatccccc
actaccacctggtggacgccacccgcgccgccaagcacgtgctgggccgctactaccgcgagcccaagacctccggcgccatc
cccatccacctggtggagtccctggtggcctccatcaagaaggaccactacgtgtccgacaccggcgacatcgtgttctacga
***gaccgacccgacctgtacgtgtacgcctccgacaagtccaagatcaacTAG*atcgatcttaaggcagcagcagctcggatagt

[1727]

atcgacacactctggacgctggtcgtgtgatggactgttgccgccacacttgctgccttgacctgtgaatatccctgccgcttttatcaaacag
cctcagtgtgtttgatcttgtgtgtacgcgcttttgcgagttgctagctgcttgtgctatttgcgaataccaccccccagcatccccttccctcgttt
catatcgcttgcatcccaaccgcaacttatctacgctgtcctgctatccctcagcgctgctcctgctcctgctcactgcccctcgcacagccttg
gtttgggctccgcctgtattctcctggtactgcaacctgtaaaccagcactgcaatgctgatgcacgggaagtagtgggatgggaacacaaa
tggaaagcttaattaa**gagctcttgttttccagaaggagttgctccttgagcctttcattctcagcctcgataacctccaaagccgctct**
aattgtggaggggttcgaatttaaaagcttggaatgttggttcgtgcgtctggaacaagcccagacttgttgctcactggagaaa
ggaccatcagctccaaaaaacttgccgctcaaaccgcgtacctctgctttcgcgcaatctgccctgttgaaatcgccaccacattca
tattgtgacgcttgagcagtctgtaattgcctcagaatgtggaatcatctgccccctgtgcgagcccatgccaggcatgtcgcgggc
gaggacacccgccactcgtacagcagaccattatgctacctcacaatagttcataacagtgaccatatttctcgaagctccccaac
gagcacctccatgctctgagtggccacccccggccctggtgcttgcggagggcaggtcaaccggcatggggctaccgaaatccc
cgaccggatcccaccacccccgcgatgggaagaatctctccccgggatgtgggcccacccaccagcacaacctgctggcccaggc
gagcgtcaaaccatacccacacaaatatccttggcatcggccctgaattccttctgccgctctgctacccggtgcttctgtccgaagc
aggggttgctagggatcgctccgagtccgcaaacccttgtcgcgtggcggggcttgttcgagcttgaagagc

[1728]

[1729] 서열 번호 223

[1730] pSZ2125의 부분, 피. 모리포르미스에서의 발현을 위해 코돈-최적화된 카멜리나 사티바 FAD-7-3XFLAG의 코딩 서열

actagt*ATGgccaacctggtgctgtccgagtgcggcatccgccccctgccccgcatctacaccacccccgctccaacttcgtgtccaac*
aacaacaagcccatcttcaagttccgccccctgacctcctacaagacctcctccccccctgttctgctcccgcgacggcttcggccgcaact
ggtccctgaacgtgtccgtgccccctggccaccaccaccccccatcgtggacgagtccccccctggaggaggaggaggaggaggaggaagaag
*cagcgcg****ggcgcgcgcc****atggactacaaggaccacgacggcgactacaaggaccacgacatcgactacaaggacgacgacgacaag*
ttcgaccccggcgcccccccccccttcaacctggccgacatccgcgccgccatccccaagcactgctgggtgaagaacccctggaagtc
catgtcctacgtgctgcgcgacgtggccatcgtgttcgccctggccgccggcgccgcctacctgaacaactggatcgtgtggcccctgta
ctggctggcccagggcaccatgttctgggccctgttcgtgctgggccacgactgcggccacggctccttctccaacaacccccgcctgaa
caacgtggtgggccacctgctgcactcctccatcctggtgccctaccacggctggcgcatctcccaccgcacccaccaccagaaccacg
gccacgtggagaacgacgagtcctggcaccccatgtccgagaagatctaccagtccctggacaagcccacccgcttcttccgcttcacc
ctgccccctggtgatgctggcctacccctttctacctgtgggcccgctccccggcaagaagggctcccactaccaccccgagtccgacctgt
tcctgcccaaggagaagaccgacgtgctgacctccaccgcctgctggaccgccatggccgccctgctgatctgcctgaacttcgtggtg
ggccccgtgcagatgctgaagctgtacggcatcccctactggatcaacgtgatgtggctggacttcgtgacctacctgcaccaccacgg
ccacgaggacaagctgccctggtaccgcggcaaggagtggtcctacctgcgcggcggcctgaccaccctggaccgcgactacggcgt
gatcaacaacatccaccacgacatcggcacccacgtgatccaccacctgttcccccagatcccccactaccacctggtggaggccaccg
aggccgtgaagcccgtgctgggcaagtactaccgcgagcccgacaagtccggccccctgccctgcacctgctgggcatcctggccaa
gtccatcaaggagggaccactacgtgtccgacgagggcgacgtggtgtactacaaggccgaccccaacatgtacggcgagatcaagg
*tgggcgccgac****TAG******atcgat***

[1731]

[1732] 서열 번호 224

[1733] pSZ2126의 부분, 피. 모리포르미스에서의 발현을 위해 코돈-최적화된 리넘 유시타티시멈 FAD3A-3XFLAG의 코딩 서열

actagt*ATGtccccccccaactccatgtcccccgccaccaacggctccaccaacggcgtggccatcaacggcgccaagaagctgctg*
*gacg****ggcgcgcgcc****atggactacaaggaccacgacggcgactacaaggaccacgacatcgactacaaggacgacgacgacaag****ttc***
gacccctccgccgccccccccttcaagatcgccgacatccgcgccgccatcccccccactgctgggtgaagaaccccctggcgctccctg
tcctacgtgctgcgcgacctgctggtgatcctgtccttcgccgtggccgccaccaagctggactcctggaccgtgtggcccctgtactgga
tcgcccagggcaccatgttctgggccgtgttcgtgctgggccacgactgcggccacggctccttctccgactcctggctgctgaacaacg
tgatgggccacatcctgcactcctccatcctggtgccctaccacggctggcgcatctcccacaagacccaccaccagaaccacggcaac
gtggagaaggacgagtcctgggtgcccctgcccgagaaggtgtacaagtccctggacaccggcaccaagttcatgcgcttcaccatcc
ccctgcccatgttcgcctaccccatctacctgtggcgccgctccccggcaagaagggctcccacttcaaccctactccgacctgttcgcc
cccaacgagcgcacctccgtgatgatctccaccctgtgctggaccgccatggccctgctgctgtgctactcctccttcatctacggcttcct
gcccgtgttcaagatctacggcgtgccctacctgatcttcgtggcctggctggacatggtgacctacctgcaccaccacggctacgagca
gaagctgccctggtaccgcggcaaggagtggtcctacctgcgcggcggcctgaccaccgtggaccgcgactacggcgtgatcaacaa
catccaccacgacatcggcacccacgtgatccaccacctgttcccccagatgcccctactaccacctggtggaggccacccaggccgcca
agcacgtgctgggcaagtactaccgcgagcccaagaagtccggccccttcccttccacctgttcggctacctggtgcgctccctgggcg

[1734]

aggaccactacgtgtccgacaccggcgacgtggtgttctaccagtccgacccccacatccccaagttccccacctccgccaccaccaagt
*ccaagtcctccTAG****atcgat***

[1735]

[1736] 서열 번호 225

[1737] pSZ2127의 부분, 피. 모리포르미스에서의 발현을 위해 코돈-최적화된 리넘 유시타티시멈 FAD3B-3XFLAG의 코딩 서열

actagt*ATGtccccccccaactccatgtcccccaccaccaacggcaacggcgtggccatgaacggcgccaagaagcagctggacg****gg***
gcgcgcc*atggactacaaggaccacgacggcgactacaaggaccacgacatcgactacaaggacgacgacgacaag****ttcgacccc***
tccgccgccccccccttcaagatcgccgacatccgcgccgccatcccccccactgctgggtgaagaaccctggcgctccctgtcctacg
tgctgcgcgacctgctggtgatcctgtccttcgccgtggccgccgccaagctggactcctggaccttctggcccctgtactgggtggccca
gggcaccatgttctgggccgtgttcgtgctgggccacgactgcggccacggctccttctccgacatctggctgctgaacaacgtgatggg
ccacatcctgcactcctccatcctggtgccctaccacggctggcgcatctcccacaagacccaccaccagaaccacggcaacgtggaga
aggacgagtcctgggtgcccctgcccgagaaggtgtacaagtccctggacacctccaccaagttcatgcgcttcaccatccccctgccc
atgttcgcctaccccatctacctgtggacccgctccccggcaagaagggctcccacttcaaccctactccgacctgttcgccccaacg
agcgcgccgccgtgctgatctccaccctgtgctggaccgccatggccctgctgctgtgctactcctccttcatctacggcttcgccccgtg
ctgaagatctacggcgtgccctacctgatcttcgtggcctggctggacatggtgacctacctgcaccaccacggctacgagcagaagct
gccctggtaccgcggcaaggagtggtcctacctgcgcggcggcctgaccaccgtggaccgcgactacggcgtgatcaacaacatcca
ccacgacatcggcacccacgtgatccaccacctgttcccccagatgccccactaccacctggtggaggccacccaggccgccaagcac
gtgctgggcaagtactaccgcgagcccaagaagtccggccccttcccttccacctgttcggctacctggtgcgctccctgggcgaggac
cactacgtgtccgacaccggcgacgtggtgttctaccagtccgacccccacatccccaagttccgcacctcctccgccaccaccaagtcc
*aagtcctccTAG****atcgat***

[1738]

[1739] 서열 번호 226

[1740] pSZ2139의 부분, 프로토테카 모리포르미스 SAD2-헤어핀 A

gctcttc**gccgccgccactcctgctcgagcgcgcccgcgcgtgcgccgccagcgccttggccttttcgccgcgctcgtgcgcgtc**
gctgatgtccatcaccaggtccatgaggtctgccttgcgccggctgagccactgcttcgtccgggcggccaagaggagcatga
gggaggactcctggtccagggtcctgacgtggtcgcggctctgggagcgggccagcatcatctggctctgccgcaccgaggc
cgcctccaactggtcctccagcagccgcagtcgccgccgaccctggcagaggaagacaggtgaggggtgtatgaattgtaca
gaacaaccacgagccttgtctaggcagaatccctaccagtcatggctttacctggatgacggcctgcgaacagctgtccagcg
accctcgctgccgccgcttctcccgcacgcttctttccagcaccgtgatggcgcgagccagcgccgcacgctggcgctgcgctt
cgccgatctgaggacagtcggggaactctgatcagtctaaaccccctttgcgcgttagtgttgccatcctttgcagaccggtgag
agccgacttgttgtgcgccaccccccacaccacctcctcccagaccaattctgtcacctttttggcgaaggcatcggcctcggcc
tgcagagaggacagcagtgcccagccgctgggggttggcggatgcacgctcaggtaccctttcttgcgctatgacacttccagca
aaaggtagggcgggctgcgagacggcttcccggcgctgcatgcaacaccgatgatgcttcgacccccgaagctccttcgggggctg
catgggcgctccgatgccgctccagggcgagcgctgtttaaatagccaggcccccgattgcaaagacattatagcgagctaccaaag
ccatattcaaacacctagatcactaccacttctacacaggccactcgagcttgtgatcgcactccgctaaggggggcgcctcttcctcttc
gtttcagtcacaacccgcaaacggcgcgcc*ATGctgctgcaggccttcctgttcctgctggccggcttcgccgccaagatcagcg*
cctccatgacgaacgagacgtccgaccgccccctggtgcacttcacccccaacaagggctggatgaacgaccccaacggcctg
tggtacgacgagaaggacgccaagtggcacctgtacttccagtacaacccgaacgacaccgtctgggggacgcccttgttctgg
ggccacgccacgtccgacgacctgaccaactgggaggaccagcccatcgccatcgccccgaagcgcaacgactccggcgcct
tctccggctccatggtggtggactacaacaacaccctccggcttcttcaacgacaccatcgacccgcgccagcgctgcgtggccatc
tggacctacaacacccccggagtccgaggagcagtacatctcctacagcctggacggcggctacaccttcaccgagtaccagaa
gaaccccgtgctggccgccaactccacccagttccgcgacccgaaggtcttctggtacgagccctcccagaagtggatcatgacc
gcggccaagtcccaggactacaagatcgagatctactcctccgacgacctgaagtcctggaagctggagtccgcgttcgccaac
gagggcttcctcggctaccagtacgagtgccccggcctgatcgaggtcccccaccgagcaggaccccagcaagtcctactgggtg
atgttcatctccatcaaccccggcgccccggccggcggctccttcaaccagtacttcgtcggcagcttcaacggcacccacttcga
ggccttcgacaaccagtcccgcgtggtggacttcggcaaggactactacgccctgcagaccttcttcaacaccgacccgacctac
gggagcgcccgggcatcgcgtgggcctccaactgggagtactccgccttcgtgcccaccaaccccctggcgctcctccatgtccct
cgtgcgcaagttctccctcaacaccgagtaccaggccaacccggagacggagctgatcaacctgaaggccgagccgatcctga

[1741]

acatcagcaacgccggcccctggagccggttcgccaccaacaccacgttgacgaaggccaacagctacaacgtcgacctgtcc
aacagcaccggcaccctggagttcgagctggtgtacgccgtcaacaccacccagacgatctccaagtccgtgttcgcggacctct
ccctctggttcaagggcctggaggaccccgaggagtacctccgcatgggcttcgaggtgtccgcgtcctccttcttcctggaccgcg
ggaacagcaaggtgaagttcgtgaaggagaaccccctacttcaccaaccgcatgagcgtgaacaaccagcccttcaagagcga
gaacgacctgtcctactacaaggtgtacggcttgctggaccagaacatcctggagctgtacttcaacgacggcgacgtcgtgtcca
ccaacacctacttcatgaccaccgggaacgccctgggctccgtgaacatgacgacgggggtggacaacctgttctacatcgaca
*agttccaggtgcgcgaggtcaagTGA*caattggcagcagcagctcggatagtatcgacacactctggacgctggtcgtgtgatgg
actgttgccgccacacttgctgccttgacctgtgaatatccctgccgcttttatcaaacagcctcagtgtgtttgatcttgtgtgtacgcgctt
ttgcgagttgctagctgcttgtgctatttgcgaataccaccccccagcatcccccttccctcgtttcatatcgcttgcatcccaaccgcaacttа
tctacgctgtcctgctatccctcagcgctgctcctgctcctgctcactgcccctcgcacagccttggtttgggctccgcctgtattctcctg
gtactgcaacctgtaaaccagcactgcaatgctgatgccacgggaagtagtgggatgggaacacaaatggaggatcccgcgtctcgaa
cagagcgcgcagaggaacgctgaaggtctcgcctctgtgcgcacctcagcgcggcatacaccacaataaccacctgacgaatgcgct
tggttcttcgtccattagcgaagcgtccggttcacacacgtgccacgttggcgaggtggcaggtgacaatgatcggtggagctgatggt
cgaaacgttcacagcctagggatatcgaattcctttcttgcgctatgacacttccagcaaaaggtagggcgggctgcgagacggcttcc
cggcgctgcatgcaacaccgatgatgcttcgacccccgaagctccttcggggctgcatgggcgctccgatgccgctccagggcga
gcgctgtttaaatagccaggcccccgattgcaaagacattatagcgagctaccaaagccatattcaaacacctagatcactaccacttct
acacaggccactcgagcttgtgatcgcactccgctaaggggggcgcctcttcctcttcgtttcagtcacaacccgcaaacactagt*GG*
GTCGCCGCGCTGCCTCGCGTCCCCTGGTGGTGCACGCCGTCGCCAGCGAGGCCCC
GCTGGGCGTGCCGCCCTCGGTGCAGCGCCCCTCCCCCGTGGTCTACTCCAAGCTG
GACAAGCAACACCGCCTGACGCCCGAGCGCCTGGAGCTGGTGCAGAGCATGGGGC
AGTTTGCGGAGGAGAG*GTGTGTTTGAGGGTTTTGGTTGCCCGTATCGAGGTCCTGGT*
GGCGCGCATGGGGGAGAAGGCGCCTGTCCCGCTGACCCCCCCGGCTACCCTCCCGG
*CACCTTCCAGGGCGCGTACG*ggatcc***CTCTCCTCCACAAACTGCCCCATGCTCTGCACC***
AGCTCCAGGCGCTCGGGCGTCTGCTTGTCCAGCTTGGAGTAGACCA
CGGGGGAGGGGCGCTGCACCGAGGGCGGAACGCCCAGCGGGGCCTCGCTGGCGA
CCGCGCGCACCACCAGGGGACGCGAGGCAGCGCGGCGACCCaagcttgcagcagcagctcgg
atagtatcgacacactctggacgctggtcgtgtgatggactgttgccgccacacttgctgccttgacctgtgaatatccctgccgcttttat
caaacagcctcagtgtgtttgatcttgtgtgtacgcgcttttgcgagttgctagctgcttgtgctatttgcgaataccaccccagcatccc
cttccctcgtttcatatcgcttgcatcccaaccgcaacttatctacgctgtcctgctatccctcagcgctgctcctgctcctgctcactgccc
ctcgcacagccttggtttgggctccgcctgtattctcctggtactgcaacctgtaaaccagcactgcaatgctgatgcacgggaagtagt
gggatgggaacacaaatggaaagctggagctc**ttgttttccagaaggagttgctccttgagcctttcattctcagcctcgataacct**
ccaaagccgctctaattgtggagggggttcgaatttaaaagcttggaatgttggttcgtgcgtctggaacaagcccagacttgtt
gctcactgggaaaaggaccatcagctccaaaaaacttgccgctcaaaccgcgtacctctgctttcgcgcaatctgccctgttga
aatcgccaccacattcatattgtgacgcttgagcagtctgtaattgcctcagaatgtggaatcatctgcccccctgtgcgagccca
tgccaggcatgtcgcgggcgaggacacccgccactcgtacagcagaccattatgctacctcacaatagttcataacagtgacc
atatttctcgaagctccccaacgagcacctccatgctctgagtggccacccccccggccctggtgcttgcggagggcaggtcaac
cggcatggggctaccgaaatccccgaccggatcccaccacccccgcgatgggaagaatctctccccgggatgtgggcccacc
accagcacaacctgctggcccaggcgagcgtcaaaccataccacacaaatatccttggcatcggccctgaattccttctgccg
ctctgctacccggtgcttctgtccgaagcagggttgctagggatcgctccgagtccgcaaacccttgtcgcgtggcggggctt
gttcgagcttgaagagc

[1742]

[1743] 서열 번호 227

[1744] pSZ2140의 부분, 프로토테카 모리포르미스 SAD2-헤어핀 B

actagt*TGCCCGTGCTGCACCCCGTGGACAAGCTGTGGCAGCCGCAGGACTTCCTGCCCGACC*
CCGAGTCGCCCGACTTCGAGGACCAGGTGGCGGAGCTGCGCGCGCGCGCCAAGGACCTGC
CCGACGAGTACTTTGTGGTGCTGGTGGGCGACATGATCACGGAGGAGGCGCTGCCGACCT
ACATGGCCATGCTCAACACCTTGGACGGTGTGCGCGACGACACGGGCGCGGCCGACCACC
CGTGTGTGTTTGAGGGTTTTGGTTGCCCGTATTGAGGTCCTGGTGGCGCGCATGGAGGAGAAG
*GCGCCTGTCCCGCTGACCCCCCCGGCTACCCTCCCGGCACCTTCCAGGGCGCGTACG*ggatcc*A*

[1745]

CGGGTGGTCGGCCGCGCCCGTGTCGTCGCGCACACCGTCCAAGGTGTTGAGCATGGCCAT
GTAGGTCGGCAGCGCCTCCTCCGTGATCATGTCGCCCACCAGCACCACAAAGTACTCGTCG
GGCAGGTCCTTGGCGCGCGCGCGCAGCTCCGCCACCTGGTCCTCGAAGTCGGGCGACTCG
*GGGTCGGGCAGGAAGTCCTGCGGCTGCCACAGCTTGTCCACGGGGTGCAGCACGGGCA*aag
ctt

[1746]

[1747] 서열 번호 228

[1748] pSZ2141의 부분, 프로토테카 모리포르미스 SAD2-헤어핀 C

actagt*GCGCTGGACGCGGCAGTGGGTGGCCGAGGAGAACCGGCACGGCGACCTGCTGAACA*
AGTACTGTTGGCTGACGGGGCGCGTCAACATGCGGGCCGTGGAGGTGACCATCAACAACC
TGATCAAGAGCGGCATGAACCCGCAGACGGACAACAACCCTTACTTGGGCTTCGTCTACAC
CTCCTTCCAGGAGCGCGCGACCAAGTACAGCCACGGCAACACCGCGCGCCTTGCGGCCGA
GCAGTGTGTTTGAGGGTTTTGGTTGCCCGTATCGAGGTCCTGGTGGCGCGCATGGGGGAGAA
*GGCGCCTGTCCCGCTGACCCCCCCGGCTACCCTCCCGGCACCTTCCAGGGCGCGTACG*ggatcc
TGCTCGGCCGCAAGGCGCGCGGTGTTGCCGTGGCTGTACTTGGTCGCGCGCTCCTGGAAG
GAGGTGTAGACGAAGCCCAAGTAAGGGTTGTTGTCCGTCTGCGGGTTCATGCCGCTCTTGA
TCAGGTTGTTGATGGTCACCTCCACGGCCCGCATGTTGACGCGCCCCGTCAGCCAACAGTA
*CTTGTTCAGCAGGTCGCCGTGCCGGTTCTCCTCGGCCACCCACTGCCGCGTCCAGCGC*aagct
t

[1749]

[1750] 서열 번호 229

[1751] pSZ2142의 부분, 프로토테카 모리포르미스 SAD2-헤어핀 D

actagt*GCTGATCGCCAGCGACGAGGGCCGGCACGAGATCGCCTACACGCGCATCGTGGACG*
AGTTCTTCCGCCTCGACCCCGAGGGCGCCGTCGCCGCCTACGCCAACATGATGCGCAAGC
AGATCACCATGCCCGCGCACCTCATGGACGACATGGGCCACGGCGAGGCCAACCCGGGCC
GCAACCTCTTCGCCGACTTCTCCGCGGTCGCCGAGAAGATCGACGTCTACGACGCCGAGG
ACTGTGTGTTTGAGGGTTTTGGTTGCCCGTATCGAGGTCCTGGTGGCGCGCATGGGGGAGAA
*GGCGCCTGTCCCGCTGACCCCCCCGGCTACCCTCCCGGCACCTTCCAGGGCGCGTACG*ggatcc
AGTCCTCGGCGTCGTAGACGTCGATCTTCTCGGCGACCGCGGAGAAGTCGGCGAAGAGGT
TGCGGCCCGGGTTGGCCTCGCCGTGGCCCATGTCGTCCATGAGGTGCGCGGGCATGGTGA
TCTGCTTGCGCATCATGTTGGCGTAGGCGGCGACGGCGCCCTCGGGGTCGAGGCGGAAGA
*ACTCGTCCACGATGCGCGTGTAGGCGATCTCGTGCCGGCCCTCGTCGCTGGCGATCAGC*aa
gctt

[1752]

[1753] 서열 번호 230

[1754] 코돈-최적화 글리신 맥스 KASII을 포함하는 pSZ1747의 부분

ATGGCCACCGCATCCACTTTCTCGGCGTTCAATGCCCGCTGCGGCGACCTGCG
TCGCTCGGCGGGCTCCGGGCCCCGGCGCCCAGCGAGGCCCCTCCCCGTGCG
CGGGCGCGCCTCCACCCACTCCGGCAAGACCATGGCCGTGGCCCTGCAGCCC
ACCCAGGAGATCACCACCATCAAGAAGCCCCCCACCAAGCAGCGCCGCGTGG
TGGTGACCGGCCTGGGCGTGGTGACCCCCCTGGGCCACGAGCCCGACATCTT
CTACAACAACCTGCTGGACGGCGCCTCCGGCATCTCCGAGATCGAGACCTTCG
ACTGCGCCGAGTACCCCACCCGCATCGCCGGCGAGATCAAGTCCTTCTCCACC
GACGGCTGGGTGGCCCCCAAGCTGTCCAAGCGCATGGACAAGTTCATGCTGTA
CATGCTGACCGCCGGCAAGAAGGCCCTGGTGGACGGCGGCATCACCGACGAC
GTGATGGACGAGCTGAACAAGGAGAAGTGCGGCGTGCTGATCGGCTCCGCCA
TGGGCGGCATGAAGGTGTTCAACGACGCCATCGAGGCCCTGCGCATCTCCTAC
AAGAAGATGAACCCCTTCTGCGTGCCCTTCGCCACCACCAACATGGGCTCCGC
CATGCTGGCCATGGACCTGGGCTGGATGGGCCCCAACTACTCCATCTCCACCG

[1755]

CCTGCGCCACCTCCAACTTCTGCATCCTGAACGCCGCCAACCACATCATCCGC
GGCGAGGCCGACGTGATGCTGTGCGGCGGCTCCGACGCCGCCATCATCCCCA
TCGGCCTGGGCGGCTTCGTGGCCTGCCGCGCCCTGTCCCAGCGCAACACCGA
CCCCACCAAGGCCTCCCGCCCCTGGGACATCAACCGCGACGGCTTCGTGATG
GGCGAGGGCGCCGGCGTGCTGCTGCTGGAGGAGCTGGAGCACGCCAAGGAG
CGCGGCGCCACCATCTACGCCGAGTTCCTGGGCGGCTCCTTCACCTGCGACG
CCTACCACGTGACCGAGCCCCGCCCCGACGGCGCCGGCGTGATCCTGTGCAT
CGAGAAGGCCCTGGCCCAGTCCGGCGTGTCCAAGGAGGACGTGAACTACATC
AACGCCCACGCCACCTCCACCCCCGCCGGCGACCTGAAGGAGTACCAGGCCC
TGATGCACTGCTTCGGCCAGAACCCCGAGCTGCGCGTGAACTCCACCAAGTCC
ATGATCGGCCACCTGCTGGGCGCCGCCGGCGGCGTGGAGGCCGTGGCCACC
ATCCAGGCCATCCGCACCGGCTGGGTGCACCCCAACATCAACCTGGAGAACCC
CGACAACGGCGTGGACGCCAAGGTGCTGGTGGGCTCCAAGAAGGAGCGCCTG
GACGTGAAGGCCGCCCTGTCCAACTCCTTCGGCTTCGGCGGCCACAACTCCTC
CATCATCTTCGCCCCCTACATGGACTACAAGGACCACGACGGCGACTACAAGG
ACCACGACATCGACTACAAGGACGACGACGACAAGTGA

[1756]

[1757] 서열 번호 231

[1758] 코돈-최적화 헬리안터스 아너스 KASII을 포함하는 pSZ1750의 관련 부분

ATGGCCACCGCATCCACTTTCTCGGCGTTCAATGCCCGCTGCGGCGACCTGCG
TCGCTCGGCGGGCTCCGGGCCCCGGCGCCCAGCGAGGCCCCTCCCCGTGCG
CGGGCGCGCCTGCCGCCGCGGCGGCCGCGTGGCCATGGCCATCGCCATCCA
GCCCTCCTCCATCGAGATGGAGGAGGAGACCACCCTGACCAAGCGCAAGCAG
CCCCCCACCAAGCAGCGCCGCGTGGTGGTGACCGGCATGGGCGTGGAGACCC
CCATCGGCAACAACCCCGACCAGTTCTACAACAACCTGCTGCAGGGCGTGTCC
GGCATCACCCAGATCGAGGCCTTCGACTGCTCCTCCTACCCCACCCGCATCGC
CGGCGAGATCAAGAACTTCTCCACCGACGGCTGGGTGGCCCCCAAGCTGTCCA
AGCGCATGGACCGCTTCATGCTGTACATGCTGACCGCCGGCAAGAAGGCCCTG
GCCGACGCCGGCATCTCCCCCTCCGACTCCGACGAGATCGACAAGTCCCGCT
GCGGCGTGCTGATCGGCTCCGCCATGGGCGGCATGAAGGTGTTCAACGACGC
CATCGAGGCCCTGCGCGTGTCCTACCGCAAGATGAACCCCTTCTGCGTGCCCT
TCGCCACCACCAACATGGGCTCCGCCATGCTGGCCATGGACCTGGGCTGGAT
GGGCCCCAACTACTCCATCTCCACCGCCTGCGCCACCTCCAACTTCTGCATCC
TGAACGCCGCCAACCACATCATCCGCGGCGAGGCCGACATGATGCTGTGCGG
CGGCTCCGACGCCGTGATCATCCCCATCGGCCTGGGCGGCTTCGTGGCCTGC
CGCGCCCTGTCCGAGCGCAACACCGACCCCGCCAAGGCCTCCCGCCCCTGGG
ACTCCGGCCGCGACGGCTTCGTGATGGGCGAGGGCGCCGGCGTGCTGCTGCT
GGAGGAGCTGGAGCACGCCAAGAAGCGCGGCGCCAAGATCTACGCCGAGTTC
CTGGGCGGCTCCTTCACCTGCGACGCCTACCACATGACCGAGCCCCACCCCG
AGGGCGCCGGCGTGATCCTGTGCATCGAGAAGGCCCTGTCCCAGGCCGGCGT
GCGCCGCGAGGACGTGAACTACATCAACGCCCACGCCACCTCCACCCCCGCC
GGCGACCTGAAGGAGTACCACGCCCTGCTGCACTGCTTCGGCAACAACCAGGA
GCTGCGCGTGAACTCCACCAAGTCCATGATCGGCCACCTGCTGGGCGCCGCC
GGCGCCGTGGAGGCCGTGGCCACCGTGCAGGCCATCCGCACCGGCTGGATC
CACCCCAACATCAACCTGGAGAACCCCGACCAGGGCGTGGACACCAAGGTGCT
GGTGGGCTCCAAGAAGGAGCGCCTGAACGTGAAGGTGGGCCTGTCCAACTCC
TTCGGCTTCGGCGGCCACAACTCCTCCATCCTGTTCGCCCCCTTCCAGATGGA
CTACAAGGACCACGACGGCGACTACAAGGACCACGACATCGACTACAAGGACG
ACGACGACAAGTGA

[1759]

[1760] 서열 번호 232

[1761] 코돈-최적화 리시누스 커뮤니스 KASII을 포함하는 pSZ1754의 부분

ATGGCCACCGCATCCACTTTCTCGGCGTTCAATGCCCGCTGCGGCGACCTGCG
TCGCTCGGCGGGCTCCGGGCCCCGGCGCCCAGCGAGGCCCCTCCCCGTGCG
CGGGCGCGCCTCCCACTACTACTCCTCCAACGGCCTGTTCCCCAACACCCCCC
TGCTGCCCAAGCGCCACCCCCGCCTGCACCACCGCCTGCCCCGCTCCGGCGA
GGCCATGGCCGTGGCCGTGCAGCCCGAGAAGGAGGTGGCCACCAACAAGAAG
CCCCTGATGAAGCAGCGCCGCGTGGTGGTGACCGGCATGGGCGTGGTGTCCC
CCCTGGGCCACGACATCGACGTGTACTACAACAACCTGCTGGACGGCTCCTCC
GGCATCTCCCAGATCGACTCCTTCGACTGCGCCCAGTTCCCCACCCGCATCGC
CGGCGAGATCAAGTCCTTCTCCACCGACGGCTGGGTGGCCCCCAAGCTGTCCA
AGCGCATGGACAAGTTCATGCTGTACATGCTGACCGCCGGCAAGAAGGCCCTG
GCCGACGGCGGCATCACCGAGGACATGATGGACGAGCTGGACAAGGCCCGCT
GCGGCGTGCTGATCGGCTCCGCCATGGGCGGCATGAAGGTGTTCAACGACGC
CATCGAGGCCCTGCGCATCTCCTACCGCAAGATGAACCCCTTCTGCGTGCCCT
TCGCCACCACCAACATGGGCTCCGCCATGCTGGCCATGGACCTGGGCTGGAT
GGGCCCCAACTACTCCATCTCCACCGCCTGCGCCACCTCCAACTTCTGCATCC
TGAACGCCGCCAACCACATCATCCGCGGCGAGGCCGACATCATGCTGTGCGG
CGGCTCCGACGCCGCCATCATCCCCATCGGCCTGGGCGGCTTCGTGGCCTGC
CGCGCCCTGTCCCAGCGCAACGACGACCCCACCAAGGCCTCCCGCCCCTGGG
ACATGAACCGCGACGGCTTCGTGATGGGCGAGGGCGCCGGCGTGCTGCTGCT
GGAGGAGCTGGAGCACGCCAAGAAGCGCGGCGCCAACATCTACGCCGAGTTC
CTGGGCGGCTCCTTCACCTGCGACGCCTACCACATGACCGAGCCCCGCCCCG
ACGGCGTGGGCGTGATCCTGTGCATCGAGAAGGCCCTGGCCCGCTCCGGCGT
GTCCAAGGAGGAGGTGAACTACATCAACGCCCACGCCACCTCCACCCCCGCC
GGCGACCTGAAGGAGTACGAGGCCCTGATGCGCTGCTTCTCCCAGAACCCCG
ACCTGCGCGTGAACTCCACCAAGTCCATGATCGGCCACCTGCTGGGCGCCGC
CGGCGCCGTGGAGGCCATCGCCACCATCCAGGCCATCCGCACCGGCTGGGTG
CACCCCAACATCAACCTGGAGAACCCCGAGGAGGGCGTGGACACCAAGGTGC
TGGTGGGCCCCAAGAAGGAGCGCCTGGACATCAAGGTGGCCCTGTCCAACTC
CTTCGGCTTCGGCGGCCACAACTCCTCCATCATCTTCGCCCCCTACAAGATGG
ACTACAAGGACCACGACGGCGACTACAAGGACCACGACATCGACTACAAGGAC
GACGACGACAAGTGA

[1762]

[1763] 서열 번호 233

[1764] 코돈-최적화 프로토테카 모리포르미스 KASII을 포함하는 pSZ2041 형질 전환 벡터의 관련 부분

GCTCTTCGCCGCCGCCACTCCTGCTCGAGCGCGCCCGCGCGTGCGCCGCCAG
CGCCTTGGCCTTTTCGCCGCGCTCGTGCGCGTCGCTGATGTCCATCACCAGGT
CCATGAGGTCTGCCTTGCGCCGGCTGAGCCACTGCTTCGTCCGGGCGGCCAA
GAGGAGCATGAGGGAGGACTCCTGGTCCAGGGTCCTGACGTGGTCGCGGCTC
TGGGAGCGGGCCAGCATCATCTGGCTCTGCCGCACCGAGGCCGCCTCCAACT
GGTCCTCCAGCAGCCGCAGTCGCCGCCGACCCTGGCAGAGGAAGACAGGTGA
GGGGGGTATGAATTGTACAGAACAACCACGAGCCTTGTCTAGGCAGAATCCCT
ACCAGTCATGGCTTTACCTGGATGACGGCCTGCGAACAGCTGTCCAGCGACCC
TCGCTGCCGCCGCTTCTCCCGCACGCTTCTTTCCAGCACCGTGATGGCGCGAG
CCAGCGCCGCACGCTGGCGCTGCGCTTCGCCGATCTGAGGACAGTCGGGGAA

[1765]

CTCTGATCAGTCTAAACCCCCTTGCGCGTTAGTGTTGCCATCCTTTGCAGACCG
GTGAGAGCCGACTTGTTGTGCGCCACCCCCCACACCACCTCCTCCCAGACCAA
TTCTGTCACCTTTTTGGCGAAGGCATCGGCCTCGGCCTGCAGAGAGGACAGCA
GTGCCCAGCCGCTGGGGGTTGGCGGATGCACGCTCAGGTACCCTTTCTTGCG
CTATGACACTTCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTTCCCGGCG
CTGCATGCAACACCGATGATGCTTCGACCCCCCGAAGCTCCTTCGGGGCTGCA
TGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGTTTAAATAGCCAGGCCCC
CGATTGCAAAGACATTATAGCGAGCTACCAAAGCCATATTCAAACACCTAGATC
ACTACCACTTCTACACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGG
GCGCCTCTTCCTCTTCGTTTCAGTCACAACCCGCAAACTCTAGAATATCAATGCT
GCTGCAGGCCTTCCTGTTCCTGCTGGCCGGCTTCGCCGCCAAGATCAGCGCCT
CCATGACGAACGAGACGTCCGACCGCCCCCTGGTGCACTTCACCCCCAACAAG
GGCTGGATGAACGACCCCAACGGCCTGTGGTACGACGAGAAGGACGCCAAGT
GGCACCTGTACTTCCAGTACAACCCGAACGACACCGTCTGGGGGACGCCCTTG
TTCTGGGGCCACGCCACGTCCGACGACCTGACCAACTGGGAGGACCAGCCCA
TCGCCATCGCCCCGAAGCGCAACGACTCCGGCGCCTTCTCCGGCTCCATGGT
GGTGGACTACAACAACACCTCCGGCTTCTTCAACGACACCATCGACCCGCGCC
AGCGCTGCGTGGCCATCTGGACCTACAACACCCCGGAGTCCGAGGAGCAGTA
CATCTCCTACAGCCTGGACGGCGGCTACACCTTCACCGAGTACCAGAAGAACC
CCGTGCTGGCCGCCAACTCCACCCAGTTCCGCGACCCGAAGGTCTTCTGGTAC
GAGCCCTCCCAGAAGTGGATCATGACCGCGGCCAAGTCCCAGGACTACAAGAT
CGAGATCTACTCCTCCGACGACCTGAAGTCCTGGAAGCTGGAGTCCGCGTTCG
CCAACGAGGGCTTCCTCGGCTACCAGTACGAGTGCCCCGGCCTGATCGAGGT
CCCCACCGAGCAGGACCCCAGCAAGTCCTACTGGGTGATGTTCATCTCCATCA
ACCCCGGCGCCCCGGCCGGCGGCTCCTTCAACCAGTACTTCGTCGGCAGCTT
CAACGGCACCCACTTCGAGGCCTTCGACAACCAGTCCCGCGTGGTGGACTTCG
GCAAGGACTACTACGCCCTGCAGACCTTCTTCAACACCGACCCGACCTACGGG
AGCGCCCTGGGCATCGCGTGGGCCTCCAACTGGGAGTACTCCGCCTTCGTGC
CCACCAACCCCTGGCGCTCCTCCATGTCCCTCGTGCGCAAGTTCTCCCTCAAC
ACCGAGTACCAGGCCAACCCGGAGACGGAGCTGATCAACCTGAAGGCCGAGC
CGATCCTGAACATCAGCAACGCCGGCCCCTGGAGCCGGTTCGCCACCAACACC
ACGTTGACGAAGGCCAACAGCTACAACGTCGACCTGTCCAACAGCACCGGCAC
CCTGGAGTTCGAGCTGGTGTACGCCGTCAACACCACCCAGACGATCTCCAAGT
CCGTGTTCGCGGACCTCTCCCTCTGGTTCAAGGGCCTGGAGGACCCCGAGGA
GTACCTCCGCATGGGCTTCGAGGTGTCCGCGTCCTCCTTCTTCCTGGACCGCG
GGAACAGCAAGGTGAAGTTCGTGAAGGAGAACCCCTACTTCACCAACCGCATG
AGCGTGAACAACCAGCCCTTCAAGAGCGAGAACGACCTGTCCTACTACAAGGT
GTACGGCTTGCTGGACCAGAACATCCTGGAGCTGTACTTCAACGACGGCGACG
TCGTGTCCACCAACACCTACTTCATGACCACCGGGAACGCCCTGGGCTCCGTG
AACATGACGACGGGGGTGGACAACCTGTTCTACATCGACAAGTTCCAGGTGCG
CGAGGTCAAGTGACAATTGGCAGCAGCAGCTCGGATAGTATCGACACACTCTG
GACGCTGGTCGTGTGATGGACTGTTGCCGCCACACTTGCTGCCTTGACCTGTG
AATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGTGTGTACG
CGCTTTTGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCCAGCAT
CCCCTTCCCTCGTTTCATATCGCTTGCATCCCAACCGCAACTTATCTACGCTGT
CCTGCTATCCCTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCTCGCACAGCC
TTGGTTTGGGCTCCGCCTGTATTCTCCTGGTACTGCAACCTGTAAACCAGCACT
GCAATGCTGATGCACGGGAAGTAGTGGGATGGGAACACAAATGGAGGATCCCG
CGTCTCGAACAGAGCGCGCAGAGGAACGCTGAAGGTCTCGCCTCTGTCGCAC
CTCAGCGCGGCATACACCACAATAACCACCTGACGAATGCGCTTGGTTCTTCGT

[1766]

CCATTAGCGAAGCGTCCGGTTCACACACGTGCCACGTTGGCGAGGTGGCAGGT
GACAATGATCGGTGGAGCTGATGGTCGAAACGTTCACAGCCTAGGGATATCGA
ATTCGGCCGACAGGACGCGCGTCAAAGGTGCTGGTCGTGTATGCCCTGGCCG
GCAGGTCGTTGCTGCTGCTGGTTAGTGATTCCGCAACCCTGATTTTGGCGTCTT
ATTTTGGCGTGGCAAACGCTGGCGCCCGCGAGCCGGGCCGGCGGCGATGCG
GTGCCCCACGGCTGCCGGAATCCAAGGGAGGCAAGAGCGCCCGGGTCAGTTG
AAGGGCTTTACGCGCAAGGTACAGCCGCTCCTGCAAGGCTGCGTGGTGGAATT
GGACGTGCAGGTCCTGCTGAAGTTCCTCCACCGCCTCACCAGCGGACAAAGCA
CCGGTGTATCAGGTCCGTGTCATCCACTCTAAAGAACTCGACTACGACCTACTG
ATGGCCCTAGATTCTTCATCAAAAACGCCTGAGACACTTGCCCAGGATTGAAAC
TCCCTGAAGGGACCACCAGGGGCCCTGAGTTGTTCCTTCCCCCCGTGGCGAG
CTGCCAGCCAGGCTGTACCTGTGATCGAGGCTGGCGGGAAAATAGGCTTCGTG
TGCTCAGGTCATGGGAGGTGCAGGACAGCTCATGAAACGCCAACAATCGCACA
ATTCATGTCAAGCTAATCAGCTATTTCCTCTTCACGAGCTGTAATTGTCCCAAAA
TTCTGGTCTACCGGGGGTGATCCTTCGTGTACGGGCCCTTCCCTCAACCCTAG
GTATGCGCGCATGCGGTCGCCGCGCAACTCGCGCGAGGGCCGAGGGTTTGGG
ACGGGCCGTCCCGAAATGCAGTTGCACCCGGATGCGTGGCACCTTTTTTGCGA
TAATTTATGCAATGGACTGCTCTGCAAAATTCTGGCTCTGTCGCCAACCCTAGG
ATCAGCGGCGTAGGATTTCGTAATCATTCGTCCTGATGGGGAGCTACCGACTAC
CCTAATATCAGCCCGACTGCCTGACGCCAGCGTCCACTTTTGTGCACACATTCC
ATTCGTGCCCAAGACATTTCATTGTGGTGCGAAGCGTCCCCAGTTACGCTCACC
TGTTTCCCGACCTCCTTACTGTTCTGTCGACAGAGCGGGCCCACAGGCCGGTC
GCAGCCACTAGTATGGCCACCGCATCCACTTTCTCGGCGTTCAATGCCCGCTG
CGGCGACCTGCGTCGCTCGGCGGGCTCCGGGCCCCGGCGCCCAGCGAGGCC
CCTCCCCGTGCGCGGGCGCGCCGCCGCCGCCGCCGACGCCAACCCCGCCCG
CCCCGAGCGCCGCGTGGTGATCACCGGCCAGGGCGTGGTGACCTCCCTGGGC
CAGACCATCGAGCAGTTCTACTCCTCCCTGCTGGAGGGCGTGTCCGGCATCTC
CCAGATCCAGAAGTTCGACACCACCGGCTACACCACCACCATCGCCGGCGAGA
TCAAGTCCCTGCAGCTGGACCCCTACGTGCCCAAGCGCTGGGCCAAGCGCGT
GGACGACGTGATCAAGTACGTGTACATCGCCGGCAAGCAGGCCCTGGAGTCC
GCCGGCCTGCCCATCGAGGCCGCCGGCCTGGCCGGCGCCGGCCTGGACCCC
GCCCTGTGCGGCGTGCTGATCGGCACCGCCATGGCCGGCATGACCTCCTTCG
CCGCCGGCGTGGAGGCCCTGACCCGCGGCGGCGTGCGCAAGATGAACCCCTT
CTGCATCCCCTTCTCCATCTCCAACATGGGCGGCGCCATGCTGGCCATGGACA
TCGGCTTCATGGGCCCCAACTACTCCATCTCCACCGCCTGCGCCACCGGCAAC
TACTGCATCCTGGGCGCCGCCGACCACATCCGCCGCGGCGACGCCAACGTGA
TGCTGGCCGGCGGCGCCGACGCCGCCATCATCCCCTCCGGCATCGGCGGCTT
CATCGCCTGCAAGGCCCTGTCCAAGCGCAACGACGAGCCCGAGCGCGCCTCC
CGCCCCTGGGACGCCGACCGCGACGGCTTCGTGATGGGCGAGGGCGCCGGC
GTGCTGGTGCTGGAGGAGCTGGAGCACGCCAAGCGCCGCGGCGCCACCATCC
TGGCCGAGCTGGTGGGCGGCGCCGCCACCTCCGACGCCCACCACATGACCGA
GCCCGACCCCCAGGGCCGCGGCGTGCGCCTGTGCCTGGAGCGCGCCCTGGA
GCGCGCCCGCCTGGCCCCCGAGCGCGTGGGCTACGTGAACGCCCACGGCAC
CTCCACCCCCGCCGGCGACGTGGCCGAGTACCGCGCCATCCGCGCCGTGATC
CCCCAGGACTCCCTGCGCATCAACTCCACCAAGTCCATGATCGGCCACCTGCT
GGGCGGCGCCGGCGCCGTGGAGGCCGTGGCCGCCATCCAGGCCCTGCGCAC
CGGCTGGCTGCACCCCAACCTGAACCTGGAGAACCCCGCCCCCGGCGTGGAC
CCCGTGGTGCTGGTGGGCCCCCGCAAGGAGCGCGCCGAGGACCTGGACGTG
GTGCTGTCCAACTCCTTCGGCTTCGGCGGCCACAACTCCTGCGTGATCTTCCG
CAAGTACGACGAGATGGACTACAAGGACCACGACGGCGACTACAAGGACCACG

[1767]

ACATCGACTACAAGGACGACGACGACAAGTGAATCGATAGATCTCTTAAGGCAG
CAGCAGCTCGGATAGTATCGACACACTCTGGACGCTGGTCGTGTGATGGACTG
TTGCCGCCACACTTGCTGCCTTGACCTGTGAATATCCCTGCCGCTTTTATCAAA
CAGCCTCAGTGTGTTTGATCTTGTGTGTACGCGCTTTTGCGAGTTGCTAGCTGC
TTGTGCTATTTGCGAATACCACCCCCAGCATCCCCTTCCCTCGTTTCATATCGCT
TGCATCCCAACCGCAACTTATCTACGCTGTCCTGCTATCCCTCAGCGCTGCTCC
TGCTCCTGCTCACTGCCCCTCGCACAGCCTTGGTTTGGGCTCCGCCTGTATTCT
CCTGGTACTGCAACCTGTAAACCAGCACTGCAATGCTGATGCACGGGAAGTAG
TGGGATGGGAACACAAATGGAAAGCTTAATTAAGAGCTCTTGTTTTCCAGAAGG
AGTTGCTCCTTGAGCCTTTCATTCTCAGCCTCGATAACCTCCAAAGCCGCTCTA
ATTGTGGAGGGGGTTCGAATTTAAAAGCTTGGAATGTTGGTTCGTGCGTCTGGA
ACAAGCCCAGACTTGTTGCTCACTGGGAAAAGGACCATCAGCTCCAAAAAACTT
GCCGCTCAAACCGCGTACCTCTGCTTTCGCGCAATCTGCCCTGTTGAAATCGC
CACCACATTCATATTGTGACGCTTGAGCAGTCTGTAATTGCCTCAGAATGTGGA
ATCATCTGCCCCCTGTGCGAGCCCATGCCAGGCATGTCGCGGGCGAGGACAC
CCGCCACTCGTACAGCAGACCATTATGCTACCTCACAATAGTTCATAACAGTGA
CCATATTTCTCGAAGCTCCCCAACGAGCACCTCCATGCTCTGAGTGGCCACCC
CCCGGCCCTGGTGCTTGCGGAGGGCAGGTCAACCGGCATGGGGCTACCGAAA
TCCCCGACCGGATCCCACCACCCCCCGCGATGGGAAGAATCTCTCCCCGGGAT
GTGGGCCCACCACCAGCACAACCTGCTGGCCCAGGCGAGCGTCAAACCATAC
CACACAAATATCCTTGGCATCGGCCCTGAATTCCTTCTGCCGCTCTGCTACCCG
GTGCTTCTGTCCGAAGCAGGGGTTGCTAGGGATCGCTCCGAGTCCGCAAACCC
TTGTCGCGTGGCGGGGCTTGTTCGAGCTTGAAGAGC

[1768]

[1769] 서열 번호 234

gctcttcgccgccgccactcctgctcgagcgcgcccgcgcgtgcgccgccagcgccttggccttttcgccgcgctcgtgcgcgtcgctgatgt
ccatcaccaggtccatgaggtctgccttgcgccggctgagccactgcttcgtccgggcggccaagaggagcatgagggaggactcctggt
ccagggtcctgacgtggtcgcggctctgggagcgggccagcatcatctggctctgccgcaccgaggccgcctccaactggtcctccagca
gccgcagtcgccgccgaccctggcagaggaagacaggtgaggggggtatgaattgtacagaacaaccacgagccttgtctaggcagaa
tccctaccagtcatggctttacctggatgacggcctgcgaacagctgtccagcgaccctcgctgccgccgcttctcccgcacgcttctttcca
gcaccgtgatggcgcgagccagcgccgcacgctggcgctgcgcttcgccgatctgaggacagtcggggaactctgatcagtctaaacccc
cttgcgcgttagtgttgccatcctttgcagaccggtgagagccgacttgttgtgcgccacccccccacaccacctcctcccagaccaattctgt
cacctttttggcgaaggcatcggcctcggcctgcagagaggacagcagtgcccagccgctgggggttggcggatgcacgctcaggtaccc
tttcttgcgctatgacacttccagcaaaaggtagggcgggctgcgagacggcttcccggcgctgcatgcaacaccgatgatgcttcgaccccccg
aagctccttcggggctgcatgggcgctccgatgccgctccagggcgagcgctgtttaaatagccaggcccccgattgcaaagacattatagcgag
ctaccaaagccatattcaaacacctagatcactaccacttctacacaggccactcgagcttgtgatcgcactccgctaagggggcgcctcttcctctt
cgtttcagtcacaacccgcaaactctagaatatcaATGctgctgcaggccttcctgttcctgctggccggcttcgccgccaagatcagcgcctcc
atgacgaacgagacgtccgaccgccccctggtgcacttcaccccccaacaagggctggatgaacgaccccaacggcctgtggtacgacgag
aaggacgccaagtggcacctgtacttccagtacaacccgaacgacaccgtctgggggacgcccttgttctggggccacgccacgtccgacg
acctgaccaactgggaggaccagcccatcgccatcgccccgaagcgcaacgactccggcgccttctccggctccatggtggtggactacaa
caacacctccggcttcttcaacgacaccatcgacccgcgccagcgctgcgtggccatctggacctacaacacccccggagtccgaggagcagt
acatctcctacagcctggacggcggctacacctttcaccgagtaccagaagaacccgtgctggccgccaactcccacccagttccgcgacccg
aaggtcttctggtacgagccctcccagaagtggatcatgaccgcggccaagtcccaggactacaagatcgagatctactcctccgacgacctg
aagtcctggaagctggagtccgcgttcgccaacgagggcttcctcggctaccagtacgagtgccccggcctgatcgaggtccccaccgagca
ggaccccagcaagtcctactgggtgatgttcatctccatcaaccccggcgccccggccggcggctccttcaaccagtacttcgtcggcagcttc
aacggcacccacttcgaggccttcgacaaccagtcccgcgtggtggacttcggcaaggactactacgccctgcagaccttcttcaacaccgac
ccgacctacgggagcgccctgggcatcgcgtgggcctccaactgggagtactccgccttcgtgcccaccaaccctggcgctcctccatgtcc
ctcgtgcgcaagttctccctcaacaccgagtaccaggccaacccggagacggagctgatcaacctgaaggccgagccgatcctgaacatca
gcaacgccggccctggagccggttcgccaccaacaccacgttgacgaaggccaacagctacaacgtcgacctgtccaacagcaccggca
ccctggagttcgagctggtgtacgccgtcaacaccacccagacgatctccaagtccgtgttcgcggacctctccctctggttcaaggggcctgga
ggacccgaggagtacctccgcatgggcttcgaggtgtccgctcctccttcttcctggaccgcgggaacagcaaggtgaagttcgtgaagga

[1770]

gaaccctacttcaccaaccgcatgagcgtgaacaaccagccccttcaagagcgagaacgacctgtcctactacaaggtgtacggcttgctgg
accagaacatcctggagctgtacttcaacgacggcgacgtcgtgtccaccaacacctacttcatgaccaccgggaacgccctgggctccgtg
aacatgacgacgggggtggacaacctgttctacatcgacaagttccaggtgcgcgaggtcaagTGAcaattgcgacgacgctcggata
gtatcgacacactctggacgctggtcgtgtgatggactgttgccgccacacttgctgccttgacctgtgaatatccctgccgctttttatcaaacagcctc
agtgtgtttgatcttgtgtgtacgcgcttttgcgagttgctagctgcttgtgctatttgcgaataccacccccagcatccccttccctcgtttcatatcgctt
gcatcccaaccgcaacttatctacgctgtcctgctatcctcagcgctgctcctgctcactgcccctcgcacagccttggtttgggctccgcc
tgtattctcctggtactgcaacctgtaaaccagcactgcaatgctgatgcacgggaagtagtgggatgggaacacaaatggaggatccgcgctctc
gaacagagcgcgcagaggaacgctgaaggtctcgcctctgtcgcacctcagcgcggcatacaccacaataaccacctgacgaatgcgcttggtt
cttcgtccattagcgaagcgtccggttcacacacgtgccacgttggcgaggtggcaggtgacaatgatcggtggagctgatggtcgaaacgttcac
agcctagggatatcgaattcggccgacaggacgcgcgtcaaaggtgctggtcgtgtatgccctggccggcaggtcgttgctgctgctggttagt
gattccgcaacccttgattttggcgtcttattttggcgtggcaaacgctggcgcccgcgagccgggccggcggcgatgcggtgccccacggctg
ccggaatccaagggaggcaagagcgcccgggtcagttgaagggctttacgcgcaaggtacagccgctcctgcaaggctgcgtggtggaatt
ggacgtgcaggtcctgctgaagttcctccaccgcctcaccagcggacaaagcaccggtgtatcaggtccgtgtcatccactctaaagaacteg
actacgacctactgatggccctagattcttcatcaaaaacgcctgagacacttgcccaggattgaaactccctgaagggaccaccaggggcc
ctgagttgttccttccccccgtggcgagctgccagccaggctgtacctgtgatcgaggctggcgggaaaataggcttcgtgtgctcaggtcatgg
gaggtgcaggacagctcatgaaacgccaacaatcgcacaatttcatgtcaagctaatcagctatttcctcttcacgagctgtaattgtcccaaaa
ttctggtctaccgggggtgatccttcgtgtacgggcccttccctcaaccctaggtatgcgcgcatgcggtcgccgcgcaactcgcgcgagggcc
gagggtttgggacgggccgtcccgaaatgcagttgcacccggatgcgtggcaccttttttgcgataatttatgcaatggactgctctgcaaaatt
ctggctctgtcgccaaccctaggatcagcggcgtaggatttcgtaatcattcgtcctgatggggagctaccgactaccctaatatcagcccgact
gcctgacgccagcgtccacttttgtgcacacattccatttcgtgcccaagacatttcattgtggtgcgaagcgtccccagttacgctcacctgtttcc
cgacctcctttactgttctgtcgacagagcgggcccacaggccggtcgcagccdactagtATGgccaccgcatccacttttctcggcgttcaatg
cccgctgcggcgacctgcgtcgctcggcgggctccgggccccggcgcccagcgaggccctcccgtgcgcggggcgccgccgccgccg
ccgacgccaaccccgcccgccccgagcgccgcgtggtgatcaccggccagggcgtggtgacctccctgggccagacccatcgagcagttcta
ctcctccctgctggagggcgtgtccggcatctcccagatccagaagttcgacaccaccggctacaccaccaccatcgccggcgagatcaagt
ccctgcagctggaccccctacgtgcccaagcgctgggccaagcgcgtggacgacgtgatcaagtacgtgtacatcgccggcaagcaggccct
ggagtccgccggcctgcccatcgaggccgccggcctggccggcgccggcctggaccccgccctgtgcggcgtgctgatcggcaccgccatg
gccggcatgacctccttcgccgccggcgtggaggccctgaccccgcggcggcgtgcgcaagatgaaccccttctgcatccccttctccatctcca
acatgggcggcgccatgctggccatggacatcggcttcatgggccccaactactccatctccaccgcctgcgccaccggcaactactgcatcc
tgggcgccgccgaccacatccgccgcggcgacgccaacgtgatgctggccggcggcgccgacgccgccatcatccccctccggcatcggcg
gcttcatcgcctgcaaggccctgtccaagcgcaacgacgagcccgagcgcgcctcccgccctgggacgccgaccgcgacggcttcgtgat
gggcgagggcgccggcgtgctggctggaggagctggagcacgccaagcgccgcggcgccaccatcctggccgagctggtgggcggcg
ccgccacctccgacgcccaccacatgaccgagccgaccccccagggccgcggcgtgcgcctgtgcctggagcgcgccctggagcgccgcc
gcctggccccgagcgcgtgggctacgtgaacgcccacggcacctccaccccgccggcgacgtggccgagtaccgcgccatccgcgccg
tgatcccccaggactccctgcgcatcaactccaccaagtccatgatcggccacctgctgggcggcgccggcgccgtggaggccgtggccgcc
atccaggccctgcgcaccggctggctgcaccccaacctgaacctggagaaccccgccccccggcgtggaccccgtggtgctggtgggccccc
gcaaggagcgcgccgaggacctggacgtggtgctgtccaactccttcggcttcggcggccacaactcctgcgtgatcttccgcaagtacgacg
agatggactacaaggaccacgacggcgactacaaggaccacgacatcgactacaaggacgacgacgacaagTGAatcgatagatctct
taaggcagcagcagctcggatagtatcgacacactctggacgctggtcgtgtgatggactgttgccgccacacttgctgccttgacctgtgaatatcc
ctgccgcttttatcaaacagcctcagtgtgtttgatcttgtgtgtacgcgcttttgcgagttgctagctgcttgtgctatttgcgaataccacccccagcat
ccccttccctcgtttcatatcgcttgcatcccaaccgcaacttatctacgctgtcctgctatcctcagcgctgctcctgctcctgctcactgcccctcgc
acagccttggtttgggctccgcctgtattctcctggtactgcaacctgtaaaccagcactgcaatgctgatgcacgggaagtagtgggatgggaaca
caaatggaaagcttaattaagagctcttgttttccagaaggagttgctcctgagcctttcattctcagcctcgataacctccaaagccgcctcta
attgtggaggggggttcgaatttaaaagcttggaatgttggttcgtgcgtctggaacaagcccagacttgttgctcactgggaaaaggaccat
cagctccaaaaaacttgccgctcaaaccgcgtacctctgctttcgcgcaatctgccctgttgaaatcgccaccacattcatattgtgacgctt
gagcagtctgtaattgcctcagaatgtggaatcatctgccccctgtgcgagcccatgccaggcatgtcgcgggcgaggacacccgccactc
gtacagcagaccattatgctacctcacaatagttcataacagtgaccatatttctcgaagctccccaacgagcacctccatgctctgagtgg
ccacccccgcgccctggtgcttgcggagggcaggtcaaccggcatggggctaccgaaatcccgaccggatcccaccaccccgcgatg
ggaagaatctctcccgggatgtgggcccaccaccagcacaacctgctggcccaggcgagcgtcaaaccataccacacaaatatccttgg
catcggccctgaattccttctgccgctctgctacccggtgcttctgtccgaagcaggggttgctagggatcgctccgagtccgcaaaccccttg
tcgcgtggcggggcttgttcgagcttgaagagc

[1771]

[1772] 서열 번호 235

ATGgccaccgcatccactttctcggcgttcaatgcccgctgcggcgacctgcgtcgctcggcgggctccgggcccc
ggcgcccagcgaggcccctccccgtgcgcggqcgccgccgccgccgacgccaaccccgcccgccccg
agcgccgcgtggtgatcaccggccagggcgtggtgacctccctgggccagaccatcgagcagttctactcctccctg
ctggagggcgtgtccggcatctcccagatccagaagttcgacaccaccggctacaccaccaccatcgccggcgag
atcaagtccctgcagctggacccctacgtgcccaagcgctgggccaagcgcgtggacgacgtgatcaagtacgtgt
acatcgccggcaagcaggccctggagtccgccggcctgcccatcgaggccgccggcctggccggcgccggcct
ggaccccgccctgtgcggcgtgctgatcggcaccgccatggccggcatgacctccttcgccgccggcgtggaggc
cctgacccgcggcggcgtgcgcaagatgaaccccttctgcatcccccttctccatctccaacatgggcggcgccatgc
tggccatggacatcggcttcatgggccccaactactccatctccaccgcctgcgccaccggcaactactgcatcctgg
gcgccgccgaccacatccgccgcggcgacgccaacgtgatgctggccggcggcgccgacgccgccatcatccc
ctccggcatcggcggcttcatcgcctgcaaggccctgtccaagcgcaacgacgagcccgagcgcgcctcccgccc
ctgggacgccgaccgcgacggcttcgtgatgggcgagggcgccggcgtgctggtgctggaggagctggagcacg
ccaagcgccgcggcgccaccatcctggccgagctggtgggcggcgccgccacctccgacgcccaccacatgacc
gagcccgacccccagggccgcggcgtgcgcctgtgcctggagcgcgccctggagcgcgcccgcctggcccccg
agcgcgtgggctacgtgaacgcccacggcacctccacccccgccggcgacgtggccgagtaccgcgccatccgc
gccgtgatcccccaggactccctgcgcatcaactccaccaagtccatgatcggccacctgctgggcggcgccggcg
ccgtggaggccgtggccgccatccaggccctgcgcaccggctggctgcaccccaacctgaacctggagaacccc
gcccccggcgtggaccccgtggtgctggtgggcccccgcaaggagcgcgccgaggacctggacgtggtgctgtc
caactccttcggcttcggcggccacaactcctgcgtgatcttccgcaagtacgacgagatggactacaaggaccacg
acggcgactacaaggaccacgacatcgactacaaggacgacgacgacaag***TGA***

[1773]

[1774] 서열 번호 236

MATASTFSAF NARCGDLRRS AGSGPRRPAR PLPVRGRAAA AADANPARPE
RRVVITGQGVVTSLGQTIEQ FYSSLLEGVS GISQIQKFDT TGYTTTIAGE
IKSLQLDPYV PKRWAKRVDDVIKYVYIAGK QALESAGLPI EAAGLAGAGL
DPALCGVLIG TAMAGMTSFA AGVEALTRGGVRKMNPFCIP FSISNMGGAM
LAMDIGFMGP NYSISTACAT GNYCILGAAD HIRRGDANVMLAGGADAAII
PSGIGGFIACKALSKRNDEPERASRPWDADRDGFVMGEGAGVLVLEELEHAKRRG
ATILA ELVGGAATSD AHHMTEPDPQ GRGVRLCLER ALERARLAPE
RVGYVNAHGTSTPAGDVAEY RAIRAVIPQD SLRINSTKSM IGHLLGGAGA
VEAVAAIQAL RTGWLHPNLNLENPAPGVDP VVLVGPRKER AEDLDVVLSN
SFGFGGHNSC VIFRKYDEMD YKDHDGDYKDHDIDYKDDDD K

[1775]

[1776] 서열 번호 237

[1777] 형질 전환 구조물 pSZ2014의 관련 서열

gctcttcgctcaccgcgtgaattgctgtcccaaacgtaagcatcatcgtggctcggtcacgcgatcctggatccggggatccta
gaccgctggtggagagcgctgccgtcggattggtggcaagtaagattgcgcaggttggcgaagggagagaccaaaaccggaggctggaagcggg
cacaacatcgtattattgcgtatagtagagcagtggcagtcgcatttcgaggtccgcaacggatctcgcaagctcgctacgctcacagtaggagaaagg
ggaccactgcccctgccagaatggtcgcgaccctctccctcgccggccccgcctgcaacacgcagtgcgtatccggcaagcgggctgtcgccttcaaccg
cccccatgttggcgtccgggctcgatcaggtgcgctgaggggggtttggtgtgcccgcgcctctgggcccgtgtcggccgtgcggacgtggggccctggg
cagtggatcagcagggtttgcgtgcaaatgcctataccggcgattgaatagcgatgaacgggatacggttgcgctcactccatgcccatgcgaccccgttt

[1778]

ctgtccgccagccgtggtcgcccgggctgcgaagcgggaccccacccagcgcattgtgatcaccggaatgggcgtggggtaccctttcttgcgctatgaca
cttccagcaaaaggtagggcgggctgcgagacggcttcccggcgctgcatgcaacaccgatgatgcttcgaccccccgaagctccttcggggctgcatgggcg
ctccgatgccgctccagggcgagcgctgtttaaatagccaggccccgattgcaaagacattatagcgagctaccaaagccatattcaaacacctagatcactacc
acttctacacaggccactcgagcttgtgatcgcactccgctaaggggcgcctcttcctcttcgtttcagtcacaacccgcaaac**ggcgcgccATG**ctgctgcag
gccttcctgttcctgctggccggcttcgccgccaagatcagcgcctccatgacgaacgagacgtccgaccgccccctggtgcacttcaccccccaacaagggct
ggatgaacgaccccaacggcctgtggtacgacgagaaggacgccaagtggcacctgtacttccagtacaacccgaacgacaccgtctgggggacgccctt
gttctggggccacgccacgtccgacgacctgaccaactgggaggaccagcccatcgccatcgccccgaagcgcaacgactccggcgccttctccggctcc
atggtggtggactacaacaacacctccggcttcttcaacgacaccatcgacccgcgccagcgctgcgtggccatctggacctacaacacccccggagtccga
ggagcagtacatctcctacagcctggacggcggctacacctttcaccgagtaccagaagaacccgtgctggccgccaactccacccagttccgcgacccga
aggtcttctggtacgagccctcccagaagtggatcatgaccgcggccaagtcccaggactacaagatcgagatctactcctccgacgacctgaagtcctgga
agctggagtccgcgttcgccaacgagggcttcctcggctaccagtacgagtgccccggcctgatcgaggtccccaccgagcaggaccccagcaagtcctac
tgggtgatgttcatctccatcaaccccggcgccccggccggcggcgctcctttcaaccagtacttcgtcggcagcttcaacggcacccacttcgaggccttcgacaa
ccagtcccgcgtggtggacttcggcaaggactactacgccctgcagacctttcttcaacaccgacccgacctacgggagcgccctgggcatcgcgtgggcctc
caactgggagtactccgccttcgtgcccaccaaccctggcgctcctccatgtccctcgtgcgcaagttctccctcaacaccgagtaccaggccaacccggag
acggagctgatcaacctgaaggccgagccgatcctgaacatcagcaacgccggccccttggagccggttcgccaccaacaccacgttgacgaaggccaac
agctacaacgtcgacctgtccaacagcaccggcaccctggagttcgagctggtgtacgccgtcaacaccacccagacgatctccaagtccgtgttcgcgga
cctctccctctggttcaagggcctggaggacccgaggagtacctccgcatgggcttcgaggtgtccgcgtcctcctttcttcctggaccgcgggaacagcaag
gtgaagttcgtgaaggagaaccctacttcaccaaccgcatgagcgtgaacaaccagccctttcaagagcgagaacgacctgtcctactacaaggtgtacgg
cttgctggaccagaacatcctggagctgtacttcaacgacggcgacgtcgtgtccaccaacacctacttcatgaccaccgggaacgccctgggctccgtgaa
catgacgacgggggtggacaacctgttctacatcgacaagttccaggtgcgcgaggtcaag**TGA**caattgcagcagcagctcggatagtatcgacacac
tctggacgctggtcgtgtgatggactgttgccgccacacttgctgccttgacctgtgaatatccctgccgcttttatcaaacagcctcagtgtgtttgatcttgtgtgtac
gcgcttttgcgagttgctagctgcttgtgctatttgcgaatacccaccccagcatcccccttccctcgtttcatatcgcttgcatcccaacccgcaacttatctacgctgtcct
gctatccctcagcgctgctcctgctcctgctcactgcccctcgcacagccttggtttgggctccgcctgtattctcctggtactgcaacctgtaaaccagcactgcaat
gctgatgcacgggaagtagtgggatgggaacacaaatggaggatcgta**gagctc**cacctgcatccgcctggcgctcgaggacgccggcgtctcgcccgacg
aggtcaactacgtcaacgcgcacgccacctccacccctggtgggcgacaaggccgaggtgcgcgcggtcaagtcggtctttggcgacatgaagggcatc
aagatgaacgccaccaagtccatgatcgggcactgcctgggcgccgccggcggcatggaggccgtcgccacgctcatggccatccgcaccggctgggt
gcaccccaccatcaaccacgacaaccccatcgccgaggtcgacggcctggacgtcgtcgccaacgccaaggcccagcacaaaatcaacgtcgccatct
ccaactccttcggcttcggcgggcacaactccgtcgtcgccttcgcccttccgcgagtaggcggagcgagcgcgcttggctgaggagggaggcgggg
tgcgagccctttggctgcgcgcgatactctccccgcacgagcagactccacgcgcctgaatctacttgtcaacgagcaaccgtgtgttttgtccgtggccat
tcttattatttctccgactgtggccgtactctgtttggctgtgcaagcaccgaagagc

[1779]

[1780] 서열 번호 238

gctcttcgctcaccgcgtgaattgctgtcccaaacgtaagcatcatcgtggctcggtcacgcgatcctggatccggggatccta
gaccgctggtggagagcgctgccgtcggattggtggcaagtaagattgcgcaggttggcgaagggagagaccaaaaccggaggctggaagcggg
cacaacatcgtattattgcgtatagtagagcagtggcagtcgcatttcgaggtccgcaacggatctcgcaagctcgctacgctcacagtaggagaaagg
ggaccactgcccctgccagaatggtcgcgacgctctccctcgccggcccgcctgcaacacgcagtgcgtatccggcaagcgggctgtcgccttcaaccg
ccccatgttggcgtccgggctcgatcaggtgcgctgaggggggtttggtgtgcccgcgcctctgggcccgtgtcggccgtgcggacgtggggccctggg
cagtggatcagcagggtttgcgtgcaaatgcctataccggcgattgaatagcgatgaacgggatacggttgcgctcactccatgcccatgcgaccccgttt
ctgtccgccagccgtggtcgcccgggctgcgaagcgggaccccacccagcgcattgtgatcaccggaatgggcgtggggtacc

[1781]

[1782] 서열 번호 239

gagctccacctgcatccgcctggcgctcgaggacgccggcgtctcgcccgacgaggtcaactacgtcaacgcgcacgccacctccaccctggtgggcga
caaggccgaggtgcgcgcggtcaagtcggtctttggcgacatgaagggcatcaagatgaacgccaccaagtccatgatcgggcactgcctgggcgccg
ccggcggcatggaggccgtcgccacgctcatggccatccgcaccggctgggtgcaccccaccatcaaccacgacaaccccatcgccgaggtcgacggc
ctggacgtcgtcgccaacgccaaggcccagcacaaaatcaacgtcgccatctccaactccttcggcttcggcgggcacaactccgtcgtcgccttcgcgc
ccttccgcgagtaggcggagcgagcgcgcttggctgaggagggaggcggggtgcgagcccttggctgcgcgcgatactctccccgcacgagcagact
ccacgcgcctgaatctacttgtcaacgagcaaccgtgtgttttgtccgtggccattcttattatttctccgactgtggccgtactctgtttggctgtgcaagcac
cgaagag

[1783]

[1784] 서열 번호 240

[1785] pSZ2110, KASII'5::CrbTub:NeoR:nr::KASII-3'의 관련 서열

gctcttcgctggcctgttgccggacgatccgtgtcgtcgagactgcattttgttttgggtgtggggctggggtactggatggcttg
agggcatgacttttctgatggagaagattgcaatgagatcatttgggtcgtctatttgtttgctgtgcaagagggtttactggtat
**ctggcaccagcttttggcccgtgcccgtttgatggacgcgtgacaggcaggcgtcctggaaagcacagacaccgtacgtacga

[1786]

ccttgacctcccccccttctccacacggcaggtgcgaggctgcccacggcgtcgaggcgggcggtgcgccgggcatggtcccg catcgcgcgcgcggcggccgcggccgacgcaaaccccgcccgccctgagcgccgcgtggtcatcacgggccagggcgtggt gaccagcctgggccagacgatcgagcagttttacagcagcctgctggagggcgtgagcggcatctcgcagatacagaagttc gacaccacgggctacacgacgacgatcgcgggcgagatcaagtcgctgcagctggacccgtacgtgcccaagcgctgggcg aagcgcgtggacgacgtgataaagtacgtctacatcgcgggcaagcaggcgctggagagcgccggcctgccgatcgaggcg gcggggctggcgggcgcggggctggacccggcgctgtgcggcgtgctcatcggcaccgccatggcgggcatgacgtctttcg cggcgggcgtggaggcgctgacgcgcggtaccctttcttgcgctatgacacttccagcaaaaggtagggcgggctgcgagacgg cttcccggcgctgcatgcaacaccgatgatgcttcgacccccccgaagctccttcggggctgcatgggcgctccgatgccgctccagg gcgagcgctgtttaaatagccaggcccccgattgcaaagacattatagcgagctaccaaagccatattcaaacacctagatcactacc acttctacacaggccactcgagcttgtgatcgcactccgctaaggggggcgcctcttcctcttcgtttcagtcacaacccgcaaac**tctag a**atatca*ATGatcgagcaggacggcctccacgccggctccccgccgcctgggtggagcgcctgttcggctacgactgggccc agcagaccatcggctgctccgacgccgccgtgttccgcctgtccgcccagggccgccccgtgctgttcgtgaagaccgacctgtc cggcgccctgaacgagctgcaggacgaggccgcccgcctgtcctggctggccaccaccggcgtgccctgcgccgccgtgctgg acgtggtgaccgaggccggccgcgactggctgctgtgggcgaggtgcccggccaggacctgctgtcctcccacctggccccg ccgagaaggtgtccatcatggccgacgccatgcgccgcctgcacaccctggaccccgccacctgccccttcgaccaccaggcca agcaccgcatcgagcgcgcccgcacccgcatggaggccggcctggtggaccaggacgacctggacgaggagcaccagggc ctggcccccgccgagctgttcgcccgcctgaaggcccgcatgcccgacggcgaggacctggtggtgacccacggcgacgcctg cctgcccaacatcatggtggagaacggccgcttctccggcttcatcgactgcggccgcctgggcgtggccgaccgctaccaggac atcgccctggccacccgcgacatcgccgaggagctgggcggcgagtgggccgaccgcttcctggtgctgtacggcatcgccgcc cccgactcccagcgcatcgccttctaccgcctgctggacgagttcttc*TGA***caattg**gcagcagcagctcggatagtatcgacaca ctctggacgctggtcgtgtgatggactgttgccgccacacttgctgccttgacctgtgaatatccctgccgcttttatcaaacagcctcagt gtgtttgatcttgtgtgtacgcgcttttgcgagttgctagctgcttgtgctatttgcgaataccacccccagcatccccttccctcgtttcatat cgcttgcatcccaaccgcaacttatctacgctgtcctgctatccctcagcgctgctcctgctcctgctcactgcccctcgcacagccttgg tttgggctccgcctgtattctcctggtactgcaacctgtaaaccagcactgcaatgctgatgcacgggaagtagtgggatgggaacaca aatggaggatccactagttctagagcggccgccaccgcggtg**gagctc**ggcggcgtgcgcaagatgaaccccttttgcatcccct tctccatctccaacatgggcggcgcgatgctggcgatggacatcggcttcatgggccccaactactccatctccacggcctgcg cgacgggcaactactgcatcctgggcgcggcggaccacatccggcgcggcgacgcaaacgtgatgctggccggcggcgcgg acgcggccatcatcccctcgggcatcggcggcttcatcgcgtgcaaggcgctgagcaagcgcaacgacgagcccgagcgcg cgagccggccctgggacgccgaccgcgacggcttcgtcatgggcgagggcgccggcgtgctggtgctggaggagctggagc acgccaagcgccgccggcgcgaccattttggctgaattagttggcggcgcggccacctcggacgcgcaccacatgaccgagcc cgacccgcagggccgcggcgtgcgcctctgcctcgagcgcgcgctcgagcgcgcgcgcctcgcgcccgagcgcgtcggctac gtcaacgcgcacggcaccagcacgcccgcgggcgacgtggccgagtaccgcgccatccgcgccgtcatcccgcaggactca ctacgcatcaactccacaaagtccatgatcgggcacctgctcggcggcgccggcgcggtcgaggccgtggccgccatccagg ccctgcgcaccggctggctccaccccaacttgaacctcgagaaccccgcgcctggcgtcgaccccgtcgtgctcgtgggctctt cc**

[1787]

[1788] 서열 번호 241

[1789] CtFad2-2의 코돈-최적화 뉴클레오티드 서열을 포함하는 pSZ2150의 부분

actagtATGggcgccggcggccgcatgtccaacccctccgagggcgagaagaagaccgagctggagggcatccagcgcgt gccctaccagaagccccccttcaccgtgggcgacgggcgccgtgaagaaggccatccccccccactgcttcaaccgctccgt gatccgctccttctcctacgtggtgtacgacctgaccatcgcctccatcctgtactacatcgccaccaccttcatccccctgctgcccc acccccctggcctacgtggcctggcccatctactgggccgtgcagggctgcgtgatgaccggcgtgtgggtgatcgcccacgagtg cggccaccacgccttctccgactaccagtggctggacgacaccgtgggcctgatcctgcactccgtgctgctggtgccctacttctc ctggaagtactcccaccgccgccaccactccaacaccggctccatcgagcacgacgaggtgttcgtgcccaagctgaagtccgg cgtgcgctccaccgccaagtacctgaacaacccccccgggcgcatcctgaccctgctggtgaccctgaccctgggctggccctg tacctgatgttcaacgtgtccggccgctactacgaccgcttcgcctgccacttcgaccccaactcccccatctactccaaccgcgag cgcgcccagatcttcatctccgacgccggcatcttcgccgtgctgtacggcctgtaccgcctggccgccgtgaagggcctggtgtg ggtgctgaccgtgtacgccggccccctgctggtggtgaacggcttcctggtgctgatcaccttcctgcagcacacccaccccctccct gccccactacgactccaccgagtgggactggctgcgcggcgccctggccaccatcgaccgcgactacggcatcctgaacaagg

[1790]

tgttccacaacatcaccgacacccacgtgacccaccacctgttctccaccatgccccactaccacgccatggaggccaccaagg ccatcatccccatcctgggcgactactaccagttcgacggcacctccgtgttcaaggccatgtaccgcgagaccaaggagtgcat ctacgtggacaaggacgaggaggtgaaggacggcgtgtactggtaccgcaacaagatcTAGatcgat

[1791]

[1792] 서열 번호 242

[1793] GlmFad2-2의 코돈-최적화 뉴클레오티드 서열을 포함하는 pSZ2151의 부분

actagtATGggcgccggcggccgcaccgacgtgccccccgccaaccgcaagtccgaggtggacccccctgaagcgcgtgccc
ttcgagaagccccccttctccctgtcccaggggcgcgccatcaagaaggtgatccccccccactgcttccagcgctccgtgttccgc
tccttctcctacgtggtgtacgacctgaccatcgccttctgcctgtactacgtggccacccactacttccacctgctgccctcccccctg
tccttcctggcctggcccatctactgggccgtgcagggctgcatcctgaccggcgtgtgggtgatcgcccacgagtgcggccacca
cgccttctccgactaccagctgctggacgacatcgtgggcctggtgctgcactccggcctgctggtgccctacttctcctggaagtac
tcccaccgccgccaccactccaacaccggctccctggagcgcgacgaggtgttcgtgcccaagcagaagtcctgcatcaagtgg
tactccaagtacctgaacaacccccccggccgcgtgctgaccctggccgtgaccctgaccctgggctggccctgtacctggccct
gaacgtgtccggccgccctacgaccgcttcgcctgccactacgacccctacggccccatctactccgaccgcgagcgcctgcag
atctacatctccgacgccggcgtgctggccgtgtgctacggcctgttccgcctggccatggccaagggcctggcctgggtggtgtg
cgtgtacggcgtgcccctgctggtggtgaacggcttcctggtgctgatcaccttcctgcagcacacccaccccgccctgccccacta
cacctcctccgagtgggactggctgcgcggcgccctggccaccgtggaccgcgactacggcatcctggacaaggtgttccacaa
catcaccgacacccacgtggcccacgagctgttctccaccatgccccactaccacgccatggaggccaccaaggccatcaagc
ccatcctgggcgagtactaccgcttcgacgagaccccttcgtgaaggccatgtggcgcgaggcccgcgagtgcatctacgtgga
gcccgaccagtccaccgagtccaagggcgtgttctggtacaacaacaagctgTAGatcgat

[1794]

[1795] 서열 번호 243

[1796] HaFad2의 코돈-최적화 뉴클레오티드 서열을 포함하는 pSZ2152의 부분

actagtATGggcgccggcggccgcatgtcctcccccaacggcaaggagaaggacggccccaagccccctggagcgcgccctg
cacgagaagccccccttcaccgtgggcgacgggcgcgccatcaagaaggtgatccccccccactgcttcaagcgctccgtgatc
cgctccttctcctacgtggtgtacgacctgaccatcgcctccatcttctactacctggccaacaactacatcccccctgctgcccaactc
cctggcctacgtggcctggcccgtgtactggatcttccagggctgcgtgctgaccggcgtgtgggtgatcgcccacgagtgcggcc
accacgccttctccgactaccagtggctggacgacaccgtgggcctgatcctgcactccgccctgctggtgccctacttctcctgga
agtactcccaccgccgccaccactccaacaccggctccatcgagcacgacgaggtgttcgtgcccaagctgaagtcctccgtgc
gctccaccgccaagtacctgaacaacccccccggccgcatcctgaccctgctggtgaccctgaccatgggctggccctgtacct
gatgttcaacgtgtccggccgctactacgaccgcttcgcctgccacttcgacccccaactcccccatctactccaaccgcgagcgcg
cccagatcttcatctccgacgccggcatcctgaccgtgttctacatcctgttccgcctggcctccaccaagggcctggtgtgggtgct
gaccatgtacggcggccccctgctggtggtgaacggcttcctggtgctgatcaccttcctgcagcacacccacccctccctgcccc
actacgactccaccgagtgggactggctgcgcggcgccctggccaccgtggaccgcgactacggcatcctgaacaaggtgttcc
acaacatcaccgacacccacgtgacccaccacctgttctccaccatgccccactaccacgccatggaggccaccaaggccatca
agcccatcctgggcgactactaccagttcgacggcacctccatcttcaaggccatgtaccgcgagaccaaggagtgcatctacgt
ggacaaggacgaggacgtgaaggacggcgtgtactggtaccgcaacaagatcTAGatcgat

[1797]

[1798] 서열 번호 244

[1800] *ZmFad2의 코돈-최적화 뉴클레오티드 서열을 포함하는 pSZ2153의 부분

actagtATGggcgccggcggccgcatgaccgagaaggagcgcgagaagcaggagcagctggcccgcgccaccggcggcg
ccgccatgcagcgctccccgtggagaagccccccttcaccctgggccaggggcgcgccatcaagaaggccatccccccccac
tgcttcgagcgctccgtgctgaagtccttctcctacgtggtgcacgacctggtgatcgccgccgccctgctgtacttcgccctggccat
catccccgccctgccctcccccctgggctacgccgcctggcccctgtactggatcgcccagggctgcgtgtgcaccggcgtgtggg
tgatcgcccacgagtgcggccaccacgccttctccgactactccctgctggacgacgtggtgggcctggtgctgcactcctccctga
tggtgccctacttctcctggaagtactcccaccgccgccaccactccaacaccggctccctggagcgcgacgaggtgttcgtgccc
aagaagaaggaggccctgccctggtacacccctacgtgtacaacaacccccgtgggccgcctggcccacatcgtggtgcagct

[1801]

gaccctgggctggcccctgtacctggccaccaacgcctccggccgcccctaccccgcttcgcctgccacttcgaccctacggcc
ccatctacaacgaccgcgagcgcgcccagatcttcgtgtccgacgccggcgtggccgccgtggccttcggcctgtacaagctggc
cgccgccttcggcgtgtggtgggtggtgcgcgtgtacgccgtgccccctgctgatcgtgaacgcctggctggtgctgatcacctacct
gcagcacacccaccctccctgccccactacgactcctccgagtgggactggctgcgcggcgccctggccaccatggaccgcga
ctacggcatcctgaaccgcgtgttccacaacatcaccgacacccacgtggcccaccacctgttctccaccatgccccactaccac
gccatggaggccaccaaggccatccgccccatcctgggcgactactaccacttcgacccccacccccgtggccaaggccacctg
gcgcgaggccggcgagtgcatctacgtggagcccgaggaccgcaagggcgtgttctggtacaacaagaagttcTAGatcgat

[1802]

[1803] 서열 번호 245

[1804] 플라스미드 pSZ2172에 함유된 형질 전환 DNA의 뉴클레오티드 서열

gctcttcgccgccgccactcctgctcgagcgcgcccgcgcgtgcgccgccagcgccttggccttttcgccgcgctcgtgcgcgtcgc
tgatgtccatcaccaggtccatgaggtctgccttgcgccggctgagccactgcttcgtccgggcggccaagaggagcatgagggag
gactcctggtccagggtcctgacgtggtcgcggctctgggagcgggccagcatcatctggctctgccgcaccgaggccgcctccaa
ctggtcctccagcagccgcagtcgccgccgaccctggcagaggaagacaggtgaggggggtatgaattgtacagaacaaccacga
gccttgtctaggcagaatccctaccagtcatggctttacctggatgacggcctgcgaacagctgtccagcgaccctcgctgccgccgc
ttctcccgcacgcttctttccagcaccgtgatggcgcgagccagcgccgcacgctggcgctgcgcttcgccgatctgaggacagtcg
gggaactctgatcagtctaaaccccctgcgcgttagtgttgccatcctttgcagaccggtgagagccgacttgttgtgcgccaccccc
acaccacctcctcccagaccaattctgtcacctttttggcgaaggcatcggcctcggcctgcagagaggacagcagtgcccagccgct
gggggttggcggatgcacgctcaggtaccctttcttgcgctatgacacttccagcaaaaggtagggcgggctgcgagacggcttccc
ggcgctgcatgcaacaccgatgatgcttcgaccccccgaagctccttcggggctgcatgggcgctccgatgccgctccagggcgag
cgctgtttaaatagccaggccccccgattgcaaagacattatagcgagctaccaaagccatattcaaacacctagatcactaccacttcta
cacaggccactcgagcttgtgatcgcactccgctaaggggcgcctcttcctcttcgtttcagtcacaacccgcaaactctagaatatca
ATGctgctgcaggccttcctgttcctgctggccggcttcgccgccaagatcagcgcctccatgacgaacgagacgtccgaccgc
ccctggtgcacttcacccccaacaagggctggatgaacgaccccaacggcctgtggtacgacgagaaggacgccaagtggc
acctgtacttccagtacaacccgaacgacaccgtctgggggacgcccttgttctggggccacgccacgtccgacgacctgaccaa
ctgggaggaccagcccatcgccatcgccccgaagcgcaacgactccggcgccttctccggctccatggtggtggactacaacaa
cacctccggcttcttcaacgacaccatcgacccgcgccagcgctgcgtggccatctggacctacaacaccccggagtccgagga
gcagtacatctcctacagcctggacggcggctacaccttcaccgagtaccagaagaacccctgtctggccgccaactccaccca
gttccgcgacccgaaggtcttctggtacgagcccctcccagaagtggatcatgaccgcggccaagtcccaggactacaagatcga
gatctactcctccgacgacctgaagtcctggaagctggagtccgcgttcgccaacgagggcttcctcggctaccagtacgagtgcc
ccggcctgatcgaggtccccaccgagcaggaccccagcaagtcctactgggtgatgttcatctccatcaaccccggcgccccgg
ccggcggctccttcaaccagtacttcgtcggcagcttcaacggcacccacttcgaggccttcgacaaccagtcccgcgtggtggac
ttcggcaaggactactacgccctgcagaccttcttcaacaccgacccgacctacgggagcgccctgggcatcgcgtgggcctcca
actgggagtactccgccttcgtgcccaccaaccctggcgctcctccatgtccctcgtgcgcaagttctccctcaacaccgagtacc
aggccaacccggagacggagctgatcaacctgaaggccgagccgatcctgaacatcagcaacgccggccccctggagccggtt
cgccaccaacaccacgttgacgaaggccaacagctacaacgtcgacctgtccaacagcaccggcaccctggagttcgagctgg
tgtacgccgtcaacaccacccagacgatctccaagtccgtgttcgcggacctctccctctggttcaagggcctggaggaccccga
ggagtacctccgcatgggcttcgaggtgtccgcgtcctccttcttcctggaccgcgggaacagcaaggtgaagttcgtgaaggag
aacccctacttcaccaaccgcatgagcgtgaacaaccagccccttcaagagcgagaacgacctgtcctactacaaggtgtacggc
ttgctggaccagaacatcctggagctgtacttcaacgacggcgacgtcgtgtccaccaacacctacttcatgaccaccgggaacg
*ccctgggctccgtgaacatgacgacgggggtggacaacctgttctacatcgacaagttccaggtgcgcgaggtcaagTGA*caa
ttggcagcagcagctcggatagtatcgacacactctggacgctggtcgtgtgatggactgttgccgccacacttgctgccttgacctgt
gaatatccctgccgcttttatcaaacagcctcagtgtgtttgatcttgtgtgtacgcgcttttgcgagttgctagctgcttgtgctatttgcga
ataccacccccagcatcccccttccctcgtttcatatcgcttgcatcccaaccgcaacttatctacgctgtcctgctatccctcagcgctgct
cctgctcctgctcactgcccctcgcacagccttggtttgggctccgcctgtattctcctggtactgcaacctgtaaaccagcactgcaatg
ctgatgcacgggaagtagtgggatgggaacacaaatggaggatcccgcgtctcgaacagagcgcgcagaggaacgctgaaggtct
cgcctctgtcgcacctcagcgcggcatacaccacaataaccacctgacgaatgcgcttggttcttcgtccattagcgaagcgtccggtt
cacacacgtgccacgttggcgaggtggcaggtgacaatgatcggtggagctgatggtcgaaacgttcacagcctagggatatcgaat
tc*ggccgacaggacgcgcgtcaaaggtgctggtcgtgtatgccctggccggcaggtcgttgctgctgctggttagtgattccgcaat*

[1805]

ccctgattttggcgtcttattttggcgtggcaaacgctggcgcccgcgagccgggccggcggcgatgcggtgccccacggctgcc
ggaatccaagggaggcaagagcgcccgggtcagttgaagggctttacgcgcaaggtacagccgctcctgcaaggctgcgtggt
ggaattggacgtgcaggtcctgctgaagttcctccaccgcctcaccagcggacaaagcaccggtgtatcaggtccgtgtcatcca
ctctaaagaactcgactacgacctactgatggcctagattcttcatcaaaaacgcctgagacacttgcccaggattgaaactccc
tgaagggaccaccaggggccctgagttgttccttccccccgtggcgagctgccagccaggctgtacctgtgatcgaggctggcgg
gaaaataggcttcgtgtgctcaggtcatgggaggtgcaggacagctcatgaaacgccaacaatcgcacaattcatgtcaagcta
atcagctatttcctcttcacgagctgtaattgtcccaaaattctggtctaccgggggtgatccttcgtgtacgggcccttccctcaaccc
taggtatgcgcgcatgcggtcgccgcgcaactcgcgcgagggccgagggtttgggacgggccgtcccgaaatgcagttgcacc
cggatgcgtggcacctttttgcgataatttatgcaatggactgctctgcaaaattctggctctgtcgccaaccctaggatcagcggc
gtaggatttcgtaatcattcgtcctgatggggagctaccgactaccctaatatcagcccgactgcctgacgccagcgtccacttttgt
gcacacattccattcgtgcccaagacatttcattgtggtgcgaagcgtccccagttacgctcacctgtttcccgacctccttactgttct
*gtcgacagagcgggcccacaggccggtcgcagcc*actagt*ATG*ggcgccggcggccgcatgtccgtgcccacctccccaa
gaagcccgagttcaactccctgaagcgcgtgccctactccaagccccccttcacctgtccgaggggcgcgccatggactacaa
*ggaccacgacggcgactacaaggaccacgacatcgactacaaggacgacgacgacaag*atcaagaaggccatccccccccc
actgcttccagcgctccgtgctgcgctccttctcctacctgctgtacgacttcatcctggcctccctgttctaccacgtggccaccaact
acttccccaacctgccccaggccctgtccaacgtggcctggccccctgtactgggccatgcagggctgcatcctgaccggcgtgtgg
gtgatcgcccacgagtgcggccaccacgccttctccgactaccagtggctggacgacaccgtgggcctgatcctgcactcctccct
gctggtgccctacttctcctggaagtactcccaccgccgccaccactccaacaccggctccctggagcgcgacgaggtgttcgtgc
ccaagaagaagtccggcctgcgctggtgggccaagcacttcaacaaccccccccggccgcttcctgtccatcaccatccagctga
ccctgggctggcccctgtacctggccttcaacgtggccggccgcccctacgaccgcttcgcctgccactacgaccctacggcccc
atcttctccgaccgcgagcgcctgcagatctacatctccgacgccggcgtgctggccgtggcctacgccctgtaccgcctggtgctg
gccaagggcgtgggctgggtgatctccgtgtacggcgtgccccctgctggtggtgaacgccttcctggtgatgatcacctacctgca
gcacacccacccctccctgccccactacgactcctccgagtgggactggatgcgcggcgccctgtccaccgtggaccgcgacta
cggcatcctgaacaaggtgttccacaacatcaccgacacccacgtggcccaccacctgttctccaccatgccccactaccacgcc
atggtggccaccaaggccatcaagcccatcctgggcgagtactaccagttcgacggcatgcccgtgtacaaggccatctggcgc
*gaggccaaggagtgcctgtacgtggagcccgacgagggcgacaaggacaagggcgtgttctggttccgcaacaagctg*TAG
atcgatcttaaggcagcagcagctcggatagtatcgacacactctggacgctggtcgtgtgatggactgttgccgccacacttgctgcc
ttgacctgtgaatatccctgccgcttttatcaaacagcctcagtgtgtttgatcttgtgtgtacgcgcttttgcgagttgctagctgcttgtgct
atttgcgaataccacccccagcatcccctttcctcgtttcatatcgcttgcatcccaaccgcaacttatctacgctgtcctgctatccctcag
cgctgctcctgctcctgctcactgcccctcgcacagccttggtttgggctccgcctgtattctcctggtactgcaacctgtaaaccagcac
tgcaatgctgatgcacgggaagtagtgggatgggaacacaaatggaaagcttaattaagagctcttgttttccagaaggagttgctcctt
gagcctttcattctcagcctcgataacctccaaagccgctctaattgtggaggggggttcgaatttaaaagcttggaatgttggttcgtgcgt
ctggaacaagcccagacttgttgctcactgggaaaaggaccatcagctccaaaaaacttgccgctcaaaccgcgtacctctgctttcgc
gcaatctgccctgttgaaatcgccaccacattcatattgtgacgcttgagcagtctgtaattgcctcagaatgtggaatcatctgccccctg
tgcgagcccatgccaggcatgtcgcgggcgaggacacccgccactcgtacagcagaccattatgctacctcacaatagttcataaca
gtgaccatatttctcgaagctccccaacgagcacctccatgctctgagtggccacccccggccctggtgcttgcggagggcaggtca
accggcatggggctaccgaaatccccgaccggatcccaccacccccgcgatgggaagaatctctccccgggatgtgggcccacca
ccagcacaacctgctggcccaggcgagcgtcaaaccatacccacacaaatatccttggcatcggccctgaattccttctgccgctctgct
acccggtgcttctgtccgaagcaggggttgctagggatcgctccgagtccgcaaaccccttgtcgcgtggcggggcttgttcgagcttg
aagagc

[1806]

[1807] 서열 번호 246

[1808] pSZ1499

```
CACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTG
AAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGT
TACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTT
TTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACG
GGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGAT
CTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTT
GGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCC
```

[1809]

ATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGC
TGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAA
GGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAA
GCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGT
GTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGAT
CCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCC
GCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATG
CTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCT
CTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGA
AAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCAC
TCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAA
GGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTT
CAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAA
AAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTA
TTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGAT
GACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGG
GAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGG
CATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGA
AAATACCGCATCAGGCGCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGC
CTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAG
GGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTGATGCATGCTCTTCGGGCTGGTCTG
AATCCTTCAGGCGGGTGTTACCCGAGAAAGAAAGGGTGCCGATTTCAAAGCAGACCCATGTGCCGGGC
CCTGTGGCCTGTGTTGGCGCCTATGTAGTCACCCCCCCTCACCCAATTGTCGCCAGTTTGCGCACTCC
ATAAACTCAAAACAGCAGCTTCTGAGCTGCGCTGTTCAAGAACACCTCTGGGGTTTGCTCACCCGCGA
GGTCGACGCCCAGCATGGCTATCAAGACGAACAGGCAGCCTGTGGAGAAGCCTCCGTTCACGATCGGG
ACGCTGCGCAAGGCCATCCCCGCGCACTGTTTCGAGCGCTCGGCGCTTCGTAGCAGCATGTACCTGGC
CTTTGACATCGCGGTCATGTCCCTGCTCTACGTCGCGTCGACGTACATCGACCCTGCACCGGTGCCTA
CGTGGGTCAAGTACGGCATCATGTGGCCGCTCTACTGGTTCTTCCAGGTGTGTTTGAGGGTTTTGGTT
GCCCGTATTGAGGTCCTGGTGGCGCGCATGGAGGAGAAGGCGCCTGTCCCGCTGACCCCCCCGGCTAC
CCTCCCGGCACCTTCCAGGGCGCCTTCGGCACGGGTGTCTGGGTGTGCGCGCACGAGTGCGGCCACCA
GGCCTTTTCCTCCAGCCAGGCCATCAACGACGGCGTGGGCCTGGTGTTCCACAGCCTGCTGCTGGTGC
CCTACTACTCCTGGAAGCACTCGCACCGGGTACCCTTTCTTGCGCTATGACACTTCCAGCAAAAGGTA
GGGCGGGCTGCGAGACGGCTTCCCGGCGCTGCATGCAACACCGATGATGCTTCGACCCCCCGAAGCTC
CTTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGTTTAAATAGCCAGGCCCCCGA
TTGCAAAGACATTATAGCGAGCTACCAAAGCCATATTCAAACACCTAGATCACTACCACTTCTACACA
GGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGGGCGCCTCTTCCTCTTCGTTTCAGTCACAACCC
GCAAACTCTAGAATATCAATGCTGCTGCAGGCCTTCCTGTTCCTGCTGGCCGGCTTCGCCGCCAAGAT
CAGCGCCTCCATGACGAACGAGACGTCCGACCGCCCCCTGGTGCACTTCACCCCCAACAAGGGCTGGA
TGAACGACCCCAACGGCCTGTGGTACGACGAGAAGGACGCCAAGTGGCACCTGTACTTCCAGTACAAC
CCGAACGACACCGTCTGGGGGACGCCCTTGTTCTGGGGCCACGCCACGTCCGACGACCTGACCAACTG
GGAGGACCAGCCCATCGCCATCGCCCCGAAGCGCAACGACTCCGGCGCCTTCTCCGGCTCCATGGTGG
TGGACTACAACAACACCTCCGGCTTCTTCAACGACACCATCGACCCGCGCCAGCGCTGCGTGGCCATC
TGGACCTACAACACCCCGGAGTCCGAGGAGCAGTACATCTCCTACAGCCTGGACGGCGGCTACACCTT
CACCGAGTACCAGAAGAACCCCGTGCTGGCCGCCAACTCCACCCAGTTCCGCGACCCGAAGGTCTTCT
GGTACGAGCCCTCCCAGAAGTGGATCATGACCGCGGCCAAGTCCCAGGACTACAAGATCGAGATCTAC
TCCTCCGACGACCTGAAGTCCTGGAAGCTGGAGTCCGCGTTCGCCAACGAGGGCTTCCTCGGCTACCA
GTACGAGTGCCCCGGCCTGATCGAGGTCCCCACCGAGCAGGACCCCAGCAAGTCCTACTGGGTGATGT
TCATCTCCATCAACCCCGGCGCCCCGGCCGGCGGCTCCTTCAACCAGTACTTCGTCGGCAGCTTCAAC
GGCACCCACTTCGAGGCCTTCGACAACCAGTCCCGCGTGGTGGACTTCGGCAAGGACTACTACGCCCT
GCAGACCTTCTTCAACACCGACCCGACCTACGGGAGCGCCCTGGGCATCGCGTGGGCCTCCAACTGGG
AGTACTCCGCCTTCGTGCCCACCAACCCCTGGCGCTCCTCCATGTCCCTCGTGCGCAAGTTCTCCCTC
AACACCGAGTACCAGGCCAACCCGGAGACGGAGCTGATCAACCTGAAGGCCGAGCCGATCCTGAACAT
CAGCAACGCCGGCCCCTGGAGCCGGTTCGCCACCAACACCACGTTGACGAAGGCCAACAGCTACAACG
TCGACCTGTCCAACAGCACCGGCACCCTGGAGTTCGAGCTGGTGTACGCCGTCAACACCACCCAGACG
ATCTCCAAGTCCGTGTTCGCGGACCTCTCCCTCTGGTTCAAGGGCCTGGAGGACCCCGAGGAGTACCT
CCGCATGGGCTTCGAGGTGTCCGCGTCCTCCTTCTTCCTGGACCGCGGGAACAGCAAGGTGAAGTTCG
TGAAGGAGAACCCCTACTTCACCAACCGCATGAGCGTGAACAACCAGCCCTTCAAGAGCGAGAACGAC
CTGTCCTACTACAAGGTGTACGGCTTGCTGGACCAGAACATCCTGGAGCTGTACTTCAACGACGGCGA
CGTCGTGTCCACCAACACCTACTTCATGACCACCGGGAACGCCCTGGGCTCCGTGAACATGACGACGG
GGGTGGACAACCTGTTCTACATCGACAAGTTCCAGGTGCGCGAGGTCAAGTGACAATTGGCAGCAGCA
GCTCGGATAGTATCGACACACTCTGGACGCTGGTCGTGTGATGGACTGTTGCCGCCACACTTGCTGCC
TTGACCTGTGAATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGTGTGTACGCGC
TTTTGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCCAGCATCCCCTTCCCTCGTTTCA

[1810]

TATCGCTTGCATCCCAACCGCAACTTATCTACGCTGTCCTGCTATCCCTCAGCGCTGCTCCTGCTCCT
GCTCACTGCCCCTCGCACAGCCTTGGTTTGGGCTCCGCCTGTATTCTCCTGGTACTGCAACCTGTAAA
CCAGCACTGCAATGCTGATGCACGGGAAGTAGTGGGATGGGAACACAAATGGAGGATCCCGCGTCTCG
AACAGAGCGCGCAGAGGAACGCTGAAGGTCTCGCCTCTGTCGCACCTCAGCGCGGCATACACCACAAT
AACCACCTGACGAATGCGCTTGGTTCTTCGTCCATTAGCGAAGCGTCCGGTTCACACACGTGCCACGT
TGGCGAGGTGGCAGGTGACAATGATCGGTGGAGCTGATGGTCGAAACGTTCACAGCCTAGGGATATCG
AATTCGGCCGACAGGACGCGCGTCAAAGGTGCTGGTCGTGTATGCCCTGGCCGGCAGGTCGTTGCTGC
TGCTGGTTAGTGATTCCGCAACCCTGATTTTGGCGTCTTATTTTGGCGTGGCAAACGCTGGCGCCCGC
GAGCCGGGCCGGCGGCGATGCGGTGCCCCACGGCTGCCGGAATCCAAGGGAGGCAAGAGCGCCCGGGT
CAGTTGAAGGGCTTTACGCGCAAGGTACAGCCGCTCCTGCAAGGCTGCGTGGTGGAATTGGACGTGCA
GGTCCTGCTGAAGTTCCTCCACCGCCTCACCAGCGGACAAAGCACCGGTGTATCAGGTCCGTGTCATC
CACTCTAAAGAACTCGACTACGACCTACTGATGGCCCTAGATTCTTCATCAAAAACGCCTGAGACACT
TGCCCAGGATTGAAACTCCCTGAAGGGACCACCAGGGGCCCTGAGTTGTTCCTTCCCCCCGTGGCGAG
CTGCCAGCCAGGCTGTACCTGTGATCGAGGCTGGCGGGAAAATAGGCTTCGTGTGCTCAGGTCATGGG
AGGTGCAGGACAGCTCATGAAACGCCAACAATCGCACAATTCATGTCAAGCTAATCAGCTATTTCCTC
TTCACGAGCTGTAATTGTCCCAAAATTCTGGTCTACCGGGGGTGATCCTTCGTGTACGGGCCCTTCCC
TCAACCCTAGGTATGCGCGCATGCGGTCGCCGCGCAACTCGCGCGAGGGCCGAGGGTTTGGGACGGGC
CGTCCCGAAATGCAGTTGCACCCGGATGCGTGGCACCTTTTTTGCGATAATTTATGCAATGGACTGCT
CTGCAAAATTCTGGCTCTGTCGCCAACCCTAGGATCAGCGGCGTAGGATTTCGTAATCATTCGTCCTG
ATGGGGAGCTACCGACTACCCTAATATCAGCCCGACTGCCTGACGCCAGCGTCCACTTTTGTGCACAC
ATTCCATTCGTGCCCAAGACATTTCATTGTGGTGCGAAGCGTCCCCAGTTACGCTCACCTGTTTCCCG
ACCTCCTTACTGTTCTGTCGACAGAGCGGGCCCACAGGCCGGTCGCAGCCACTAGTATGGCCACCGCA
TCCACTTTCTCGGCGTTCAATGCCCGCTGCGGCGACCTGCGTCGCTCGGCGGGCTCCGGGCCCCGGCG
CCCAGCGAGGCCCCTCCCCGTGCGCGGGCGCGCCGAGGTGCACGTGCAGGTGACCCACTCCCTGGCCC
CCGAGAAGCGCGAGATCTTCAACTCCCTGAACAACTGGGCCCAGGAGAACATCCTGGTGCTGCTGAAG
GACGTGGACAAGTGCTGGCAGCCCTCCGACTTCCTGCCCGACTCCGCCTCCGAGGGCTTCGACGAGCA
GGTGATGGAGCTGCGCAAGGCGCTGCAAGGAGATCCCCGACGACTACTTCATCGTGCTGGTGGGCGACA
TGATCACCGAGGAGGCCCTGCCCACCTACCAGACCATGCTGAACACCCTGGACGGCGTGCGCGACGAG
ACCGGCGCCTCCCTGACCCCCTGGGCCATCTGGACCCGCGCCTGGACCGCCGAGGAGAACCGCCACGG
CGACCTGCTGAACAAGTACCTGTACCTGTCCGGCCGCGTGGACATGAAGCAGATCGAGAAGACCATCC
AGTACCTGATCGGCTCCGGCATGGACCCCCGCACCGAGAACAACCCCTACCTGGGCTTCATCTACACC
TCCTTCCAGGAGCGCGCCACCTTCATCTCCCACGGCAACACCGCCCGCCTGGCCAAGGAGCACGGCGA
CCTGAAGCTGGCCCAGATCTGCGGCATCATCGCCGCCGACGAGAAGCGCCACGAGACCGCCTACACCA
AGATCGTGGAGAAGCTGTTCGAGATCGACCCCGACGGCACCGTGCTGGCCCTGGCCGACATGATGCGC
AAGAAGGTGTCCATGCCCGCCCACCTGATGTACGACGGCCAGGACGACAACCTGTTCGAGAACTTCTC
CTCCGTGGCCCAGCGCCTGGGCGTGTACACCGCCAAGGACTACGCCGACATCCTGGAGTTCCTGGTGG
GCCGCTGGGACATCGAGAAGCTGACCGGCCTGTCCGGCGAGGGCCGCAAGGCCCAGGACTACGTGTGC
ACCCTGCCCCCCCCGCATCCGCCGCCTGGAGGAGCCGCCCAGTCCCGCGTGAAGAAGGCCTCCGCCAC
CCCCTTCTCCTGGATCTTCGGCCGCGAGATCAACCTGATGGACTACAAGGACCACGACGGCGACTACA
AGGACCACGACATCGACTACAAGGACGACGACGACAAGTGAATCGATAGATCTCTTAAGGCAGCAGCA
GCTCGGATAGTATCGACACACTCTGGACGCTGGTCGTGTGATGGACTGTTGCCGCCACACTTGCTGCC
TTGACCTGTGAATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGTGTGTACGCGC
TTTTGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCCAGCATCCCCTTCCCTCGTTTCA
TATCGCTTGCATCCCAACCGCAACTTATCTACGCTGTCCTGCTATCCCTCAGCGCTGCTCCTGCTCCT
GCTCACTGCCCCTCGCACAGCCTTGGTTTGGGCTCCGCCTGTATTCTCCTGGTACTGCAACCTGTAAA
CCAGCACTGCAATGCTGATGCACGGGAAGTAGTGGGATGGGAACACAAATGGAAAGCTTAATTAAGAG
CTCCCGCCACCACTCCAACACGGGGTGCCTGGACAAGGACGAGGTGTTTGTGCCGCCGCACCGCGCAG
TGGCGCACGAGGGCCTGGAGTGGGAGGAGTGGCTCCCATCCGCATGGGCAAGGTGCTGGTCACCCTG
ACCCTGGGCTGGCCGCTGTACCTCATGTTCAACGTCGCCTCGCGGCCGTACCCGCGCTTCGCCAACCA
CTTTGACCCGTGGTCGCCCATCTTCAGCAAGCGCGAGCGCATCGAGGTGGTCATCTCCGACCTGGCGC
TGGTGGCGGTGCTCAGCGGGCTCAGCGTGCTGGGCCGCACCATGGGCTGGGCCTGGCTGGTCAAGACC
TACGTGGTGCCCTACCTGATCGTGAACATGTGGCTCGTGCTCATCACGCTGCTCCAGCACACGCACCC
GGCGCTGCCGCACTACTTCGAGAAGGACTGGGACTGGCTGCGCGGCGCCATGGCCACCGTGGACCGCT
CCATGGGCCCGCCCTTCATGGACAACATCCTGCACCACATCTCCGACACCCACGTGCTGCACCACCTC
TTCAGCACCATCCCGCACTACCACGCCGAGGAGGCCTCCGCCGCCATCAGGCCCATCCTGGGCAAGTA
CTACCAGTCCGACAGCCGCTGGGTCGGCCGCGCCCTGTGGGAGGACTGGCGCGACTGCCGCTACGTCG
TCCCGGACGCGCCCGAGGACGACTCCGCGCTCTGGTTCCACAAGTGAGTGAGTGAGAAGAGCCTCTAG
AGTCGACCTGCAGGCATGCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTAT
CCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGT
GAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGC
TGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCG
CTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAAT
ACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCA

[1811]

GGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAA
AATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGG
AAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTT
CGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCC
AAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCT
TGAGTCCAACCCGGTAAGACACGACTTATCGC

[1812]

[1813] 서열 번호 247

GGGCTGGTCTGAATCCTTCAGGCGGGTGTTACCCGAGAAAGAAAGGGTGCCGATTTCAAAGCAGACCC
ATGTGCCGGGCCCTGTGGCCTGTGTTGGCGCCTATGTAGTCACCCCCCCTCACCCAATTGTCGCCAGT
TTGCGCACTCCATAAACTCAAAACAGCAGCTTCTGAGCTGCGCTGTTCAAGAACACCTCTGGGGTTTG
CTCACCCGCGAGGTCGACGCCCAGCATGGCTATCAAGACGAACAGGCAGCCTGTGGAGAAGCCTCCGT
TCACGATCGGGACGCTGCGCAAGGCCATCCCCGCGCACTGTTTCGAGCGCTCGGCGCTTCGTAGCAGC
ATGTACCTGGCCTTTGACATCGCGGTCATGTCCCTGCTCTACGTCGCGTCGACGTACATCGACCCTGC
ACCGGTGCCTACGTGGGTCAAGTACGGCATCATGTGGCCGCTCTACTGGTTCTTCCAGGTGTGTTTGA
GGGTTTTGGTTGCCCGTATTGAGGTCCTGGTGGCGCGCATGGAGGAGAAGGCGCCTGTCCCGCTGACC
CCCCCGGCTACCCTCCCGGCACCTTCCAGGGCGCCTTCGGCACGGGTGTCTGGGTGTGCGCGCACGAG
TGCGGCCACCAGGCCTTTTCCTCCAGCCAGGCCATCAACGACGGCGTGGGCCTGGTGTTCCACAGCCT
GCTGCTGGTGCCCTACTACTCCTGGAAGCACTCGCACCG

[1814]

[1815] 서열 번호 248

CCGCCACCACTCCAACACGGGGTGCCTGGACAAGGACGAGGTGTTTGTGCCGCC
GCACCGCGCAGTGGCGCACGAGGGCCTGGAGTGGGAGGAGTGGCTGCCCATCCG
CATGGGCAAGGTGCTGGTCACCCTGACCCTGGGCTGGCCGCTGTACCTCATGTTC
AACGTCGCCTCGCGGCCGTACCCGCGCTTCGCCAACCACTTTGACCCGTGGTCGC
CCATCTTCAGCAAGCGCGAGCGCATCGAGGTGGTCATCTCCGACCTGGCGCTGGT
GGCGGTGCTCAGCGGGCTCAGCGTGCTGGGCCGCACCATGGGCTGGGCCTGGCT
GGTCAAGACCTACGTGGTGCCCTACCTGATCGTGAACATGTGGCTCGTGCTCATC
ACGCTGCTCCAGCACACGCACCCGGCGCTGCCGCACTACTTCGAGAAGGACTGG
GACTGGCTGCGCGGCGCCATGGCCACCGTGGACCGCTCCATGGGCCCGCCCTTCA
TGGACAACATCCTGCACCACATCTCCGACACCCACGTGCTGCACCACCTCTTCAG
CACCATCCCGCACTACCACGCCGAGGAGGCCTCCGCCGCCATCAGGCCCATCCTG
GGCAAGTACTACCAGTCCGACAGCCGCTGGGTCGGCCGCGCCCTGTGGGAGGAC
TGGCGCGACTGCCGCTACGTCGTCCCGGACGCGCCCGAGGACGACTCCGCGCTCT
GGTTCCACAAGTGAGTGAGTGA

[1816]

[1817] 서열 번호 249

[1818] pSZ2004

TAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATCAATACCATATTT
TTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCT
GGTATCGGTCTGCGATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATA
AGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGTTTATGCAT
TTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAAC
CGTTATTCATTCGTGATTGCGCCTGAGCGAGGCGAAATACGCGATCGCTGTTAAAAGGACAATTACAA
ACAGGAATCGAGTGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGG
ATATTCTTCTAATACCTGGAACGCTGTTTTTCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAG
GAGTACGGATAAAATGCTTGATGGTCGGAAGTGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATC
TCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTT
CCCATACAAGCGATAGATTGTCGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATA
AATCAGCATCCATGTTGGAATTTAATCGCGGCCTCGACGTTTCCCGTTGAATATGGCTCATATTCTTC
CTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTAT
TTAGAAAAATAAACAAATAGGGGTCAGTGTTACAACCAATTAACCAATTCTGAACATTATCGCGAGCC
CATTTATACCTGAATATGGCTCATAACACCCCTTGTTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACC
TGACCCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGACTCCCCATGCGA
GAGTAGGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGCCCGGG

[1819]



[1820]

CAGCCACTAGTATGGTGGCCACCGCCGCCTCCTCCGCCTTCTTCCCCCTGCCCTCCGCCGACACCTCC
TCCCGCCCCGGCAAGCTGGGCAACAAGCCCTCCTCCCTGTCCCCCCTGAAGCCCAAGTCCACCCCCAA
CGGCGGCCTGCAGGTGAAGGCCAACGCCTCCGCCCCCCCCAAGATCAACGGCTCCCCCGTGGGCCTGA
AGTCCGGCGGCCTGAAGGGGCGCGCCACCCAGGAGGACGCCCACTCCGCCCCCCCCCCCCCGCACCTTC
ATCAACCAGCTGCCCGACTGGTCCATGCTGCTGGCCGCCATCACCACCGTGTTCCTGGCCGCCGAGAA
GCAGTGGATGATGCTGGACTGGAAGCCCAAGCGCCCCGACATGCTGGTGGACCCCTTCGGCCTGGGCT
CCATCGTGCAGGACGGCCTGGTGTTCCGCCAGAACTTCTCCATCCGCTCCTACGAGATCGGCGCCGAC
CGCACCGCCTCCATCGAGACCGTGATGAACCACCTGCAGGAGACCGCCCTGAACCACGTGAAGATCGC
CGGCCTGTCCAACGACGGCTTCGGCCGCACCCCCGAGATGTACAAGCGCGACCTGATCTGGGTGGTGG
CCAAGATGCAGGTGATGGTGAACCGCTACCCCACCTGGGGCGACACCGTGGAGGTGAACACCTGGGTG
GCCAAGTCCGGCAAGAACGGCATGCGCCGCGACTGGCTGATCTCCGACTGCAACACCGGCGAGATCCT
GACCCGCGCCTCCTCCGTGTGGGTGATGATGAACCAGAAGACCCGCCGCCTGTCCAAGATCCCCGACG
AGGTGCGCAACGAGATCGAGCCCCACTTCGTGGACTCCCCCCCCCGTGATCGAGGACGACGACCGCAAG
CTGCCCAAGCTGGACGAGAAGACCGCCGACTCCATCCGCAAGGGCCTGACCCCCCGCTGGAACGACCT
GGACGTGAACCAGCACGTGAACAACGTGAAGTACATCGGCTGGATCCTGGAGTCCACCCCCCCCGAGG
TGCTGGAGACCCAGGAGCTGTGCTCCCTGACCCTGGAGTACCGCCGCGAGTGCGGCCGCGAGTCCGTG
CTGGAGTCCCTGACCGCCATGGACCCCTCCGGCGGCGGCTACGGCTCCCAGTTCCAGCACCTGCTGCG
CCTGGAGGACGGCGGCGAGATCGTGAAGGGCCGCACCGAGTGGCGCCCCAAGAACGGCGTGATCAACG
GCGTGGTGCCCACCGGCGAGTCCTCCCCCGGCGACTACTCCATGGACTACAAGGACCACGACGGCGAC
TACAAGGACCACGACATCGACTACAAGGACGACGACGACAAGTGAATCGATAGATCTCTTAAGGCAGC
AGCAGCTCGGATAGTATCGACACACTCTGGACGCTGGTCGTGTGATGGACTGTTGCCGCCACACTTGC
TGCCTTGACCTGTGAATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGTGTGTAC
GCGCTTTTGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCCAGCATCCCCTTCCCTCGT
TTCATATCGCTTGCATCCCAACCGCAACTTATCTACGCTGTCCTGCTATCCCTCAGCGCTGCTCCTGC
TCCTGCTCACTGCCCCTCGCACAGCCTTGGTTTGGGCTCCGCCTGTATTCTCCTGGTACTGCAACCTG
TAAACCAGCACTGCAATGCTGATGCACGGGAAGTAGTGGGATGGGAACACAAATGGAAAGCTGTAGAA
TTCGGCCGACAGGACGCGCGTCAAAGGTGCTGGTCGTGTATGCCCTGGCCGGCAGGTCGTTGCTGCTG
CTGGTTAGTGATTCCGCAACCCTGATTTTGGCGTCTTTATTTTGGCGTGGCAAACGCTGGCGCCCGCGA
GCCGGGCCGGCGGCGATGCGGTGCCCCACGGCTGCCGGAATCCAAGGGAGGCAAGAGCGCCCGGGTCA
GTTGAAGGGCTTTACGCGCAAGGTACAGCCGCTCCTGCAAGGCTGCGTGGTGGAATTGGACGTGCAGG
TCCTGCTGAAGTTCCTCCACCGCCTCACCAGCGGACAAAGCACCGGTGTATCAGGTCCGTGTCATCCA
CTCTAAAGAACTCGACTACGACCTACTGATGGCCCTAGATTCTTCATCAAAAACGCCTGAGACACTTG
CCCAGGATTGAAACTCCCTGAAGGGACCACCAGGGGCCCTGAGTTGTTCCTTCCCCCCGTGGCGAGCT
GCCAGCCAGGCTGTACCTGTGATCGAGGCTGGCGGGAAAATAGGCTTCGTGTGCTCAGGTCATGGGAG
GTGCAGGACAGCTCATGAAACGCCAACAATCGCACAATTCATGTCAAGCTAATCAGCTATTTCCTCTT
CACGAGCTGTAATTGTCCCAAAATTCTGGTCTACCGGGGGTGATCCTTCGTGTACGGGCCCTTCCCTC
AACCCTAGGTATGCGCGCATGCGGTCGCCGCGCAACTCGCGCGAGGGCCGAGGGTTTGGGACGGGCCG
TCCCGAAATGCAGTTGCACCCGGATGCGTGGCACCTTTTTTGCGATAATTTATGCAATGGACTGCTCT
GCAAAATTCTGGCTCTGTCGCCAACCCTAGGATCAGCGGCGTAGGATTTCGTAATCATTCGTCCTGAT
GGGGAGCTACCGACTACCCTAATATCAGCCCGACTGCCTGACGCCAGCGTCCACTTTTGTGCACACAT
TCCATTCGTGCCCAAGACATTTCATTGTGGTGCGAAGCGTCCCCAGTTACGCTCACCTGTTTCCCGAC
CTCCTTACTGTTCTGTCGACAGAGCGGGCCCACAGGCCGGTCGCAGCCACTAGTATGGTGGCCACCGC
CGCCTCCTCCGCCTTCTTCCCCCTGCCCTCCGCCGCACACCTCCTCCCGCCCCGGCAAGCTGGGCAACA
AGCCCTCCTCCCTGTCCCCCCTGAAGCCCAAGTCCACCCCCAACGGCGGCCTGCAGGTGAAGGCCAAC
GCCTCCGCCCCCCCCCAAGATCAACGGCTCCCCCGTGGGCCTGAAGTCCGGCGGCCTGAAGGGGCGCGC
CACCCAGGAGGACGCCCACTCCGCCCCCCCCCCCCGCACCTTCATCAACCAGCTGCCCGACTGGTCCA
TGCTGCTGGCCGCCATCACCACCGTGTTCCTGGCCGCCGAGAAGCAGTGGATGATGCTGGACTGGAAG
CCCAAGCGCCCCGACATGCTGGTGGACCCCTTCGGCCTGGGCTCCATCGTGCAGGACGGCCTGGTGTT
CCGCCAGAACTTCTCCATCCGCTCCTACGAGATCGGCGCCGACCGCACCGCCTCCATCGAGACCGTGA
TGAACCACCTGCAGGAGACCGCCCTGAACCACGTGAAGATCGCCGGCCTGTCCAACGACGGCTTCGGC
CGCACCCCCGAGATGTACAAGCGCGACCTGATCTGGTGGTGGCCAAGATGCAGGTGATGGTGAACCG
CTACCCCACCTGGGGCGACACCGTGGAGGTGAACACCTGGGTGGCCAAGTCCGGCAAGAACGGCATGC
GCCGCGACTGGCTGATCTCCGACTGCAACACCGGCGAGATCCTGACCCGCGCCTCCTCCGTGTGGGTG
ATGATGAACCAGAAGACCCGCCGCCTGTCCAAGATCCCCGACGAGGTGCGCAACGAGATCGAGCCCCA
CTTCGTGGACTCCCCCCCCGTGATCGAGGACGACGACCGCAAGCTGCCCAAGCTGGACGAGAAGACCG
CCGACTCCATCCGCAAGGGCCTGACCCCCCGCTGGAACGACCTGGACGTGAACCAGCACGTGAACAAC
GTGAAGTACATCGGCTGGATCCTGGAGTCCACCCCCCCCGAGGTGCTGGAGACCCAGGAGCTGTGCTC
CCTGACCCTGGAGTACCGCCGCGAGTGCGGCCGCGAGTCCGTGCTGGAGTCCCTGACCGCCATGGACC
CCTCCGGCGGCGGCTACGGCTCCCAGTTCCAGCACCTGCTGCGCCTGGAGGACGGCGGCGAGATCGTG
AAGGGCCGCACCGAGTGGCGCCCCAAGAACGGCGTGATCAACGGCGTGGTGCCCACCGGCGAGTCCTC
CCCCGGCGACTACTCCATGGACTACAAGGACCACGACGGCGACTACAAGGACCACGACATCGACTACA
AGGACGACGACGACAAGTGAATCGATAGATCTCTTAAGGCAGCAGCAGCTCGGATAGTATCGACACAC
TCTGGACGCTGGTCGTGTGATGGACTGTTGCCGCCACACTTGCTGCCTTGACCTGTGAATATCCCTGC

[1821]

CGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGTGTGTACGCGCTTTTGCGAGTTGCTAGCTGCT
TGTGCTATTTGCGAATACCACCCCCAGCATCCCCTTCCCTCGTTTCATATCGCTTGCATCCCAACCGC
AACTTATCTACGCTGTCCTGCTATCCCTCAGCGCTGCTCCTGCTCCTGCTCACTGCCCCTCGCACAGC
CTTGGTTTGGGCTCCGCCTGTATTCTCCTGGTACTGCAACCTGTAAACCAGCACTGCAATGCTGATGC
ACGGGAAGTAGTGGGATGGGAACACAAATGGAAAGCTGTAGAGCTCGGCGGCGTGCGCAAGATGAACC
CCTTTTGCATCCCCTTCTCCATCTCCAACATGGGCGGCGCGATGCTGGCGATGGACATCGGCTTCATG
GGCCCCAACTACTCCATCTCCACGGCCTGCGCGACGGGCAACTACTGCATCCTGGGCGCGGCGGACCA
CATCCGGCGCGGCGACGCAAACGTGATGCTGGCCGGCGGCGCGGACGCGGCCATCATCCCCTCGGGCA
TCGGCGGCTTCATCGCGTGCAAGGCGCTGAGCAAGCGCAACGACGAGCCCGAGCGCGCGAGCCGGCCC
TGGGACGCCGACCGCGACGGCTTCGTCATGGGCGAGGGCGCCGGCGTGCTGGTGCTGGAGGAGCTGGA
GCACGCCAAGCGCCGCGGCGCGACCATTTTGGCTGAATTAGTTGGCGGCGCGGCCACCTCGGACGCGC
ACCACATGACCGAGCCCGACCCGCAGGGCCGCGGCGTGCGCCTCTGCCTCGAGCGCGCGCTCGAGCGC
GCGCGCCTCGCGCCCGAGCGCGTCGGCTACGTCAACGCGCACGGCACCAGCACGCCCGCGGGCGACGT
GGCCGAGTACCGCGCCATCCGCGCCGTCATCCCGCAGGACTCACTACGCATCAACTCCACAAAGTCCA
TGATCGGGCACCTGCTCGGCGGCGCCGGCGCGGTCGAGGCCGTGGCCGCCATCCAGGCCCTGCGCACC
GGCTGGCTCCACCCCAACTTGAACCTCGAGAACCCCGCGCCTGGCGTCGACCCCGTCGTGCTCGTGGG
GAAGAGCCTCTAGATCTTAAAATGAAGTGAAGTTCCTATACTTTCTAGAGAATAGGAACTTCTATAGT
GAGTCGAATAAGGGCGACACAAAATTTATTCTAAATGCATAATAAATACTGATAACATCTTATAGTTT
GTATTATATTTTGTATTATCGTTGACATGTATAATTTTGATATCAAAAACTGATTTTCCCTTTATTAT
TTTCGAGATTTATTTTCTTAATTCTCTTTAACAAACTAGAAATATTGTATATACAAAAAATCATAAAT
AATAGATGAATAGTTTAATTATAGGTGTTCATCAATCGAAAAAGCAACGTATCTTATTTAAAGTGCGT
TGCTTTTTTCTCATTTATAAGGTTAAATAATTCTCATATATCAAGCAAAGTGACAGGCGCCCTTAAAT
ATTCTGACAAATGCTCTTTCCCTAAACTCCCCCCATAAAAAAACCCGCCGAAGCGGGTTTTTACGTTA
TTTGCGGATTAACGATTACTCGTTATCAGAACCGCCCAGGGGGCCCGAGCTTAAGACTGGCCGTCGTT
TTACAACACAGAAAGAGTTTGTAGAAACGCAAAAAGGCCATCCGTCAGGGGCCTTCTGCTTAGTTTGA
TGCCTGGCAGTTCCCTACTCTCGCCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGG
CTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGC
AGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGT
TTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAAC
CCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGAC
CCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCAC
GCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTT
CAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATC
GCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCT
TGAAGTGGTGGGCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCA
GTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTT
TTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTA
CGGGGTCTGACGCTCAGTGGAACGACGCGCGCGTAACTCACGTTAAGGGATTTTGGTCATGAGCTTGC
GCCGTCCCGTCAAGTCAGCGTAATGCTCTGCTTT

[1822]

[1823] 서열 번호 250

[1824] pSZ1500

CACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTG
AAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGT
TACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTT
TTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACG
GGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGAT
CTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTT
GGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCC
ATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGC
TGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAA
GGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAA
GCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGT
GTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGAT
CCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCC
GCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATG
CTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCT
CTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGA
AAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCAC
TCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAA
GGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTT
CAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAA

[1825]



[1826]

CTGCCAGCCAGGCTGTACCTGTGATCGAGGCTGGCGGGAAAATAGGCTTCGTGTGCTCAGGTCATGGG
AGGTGCAGGACAGCTCATGAAACGCCAACAATCGCACAATTCATGTCAAGCTAATCAGCTATTTCCTC
TTCACGAGCTGTAATTGTCCCAAAATTCTGGTCTACCGGGGGTGATCCTTCGTGTACGGGCCCTTCCC
TCAACCCTAGGTATGCGCGCATGCGGTCGCCGCGCAACTCGCGCGAGGGCCGAGGGTTTGGGACGGGC
CGTCCCGAAATGCAGTTGCACCCGGATGCGTGGCACCTTTTTTGCGATAATTTATGCAATGGACTGCT
CTGCAAAATTCTGGCTCTGTCGCCAACCCTAGGATCAGCGGCGTAGGATTTCGTAATCATTCGTCCTG
ATGGGGAGCTACCGACTACCCTAATATCAGCCCGACTGCCTGACGCCAGCGTCCACTTTTGTGCACAC
ATTCCATTCGTGCCCAAGACATTTCATTGTGGTGCGAAGCGTCCCCAGTTACGCTCACCTGTTTCCCG
ACCTCCTTACTGTTCTGTCGACAGAGCGGGCCCACAGGCCGGTCGCAGCCACTAGTATGGCCACCGCA
TCCACTTTCTCGGCGTTCAATGCCCGCTGCGGCGACCTGCGTCGCTCGGCGGGCTCCGGGCCCCGGCG
CCCAGCGAGGCCCCTCCCCGTGCGCGGGCGCGCCGCCACCGGCGAGCAGCCCTCCGGCGTGGCCTCCC
TGCGCGAGGCCGACAAGGAGAAGTCCCTGGGCAACGCCTGCGCCTGGGCTCCCTGACCGAGGACGGC
CTGTCCTACAAGGAGAAGTTCGTGATCCGCTGCTACGAGGTGGGCATCAACAAGACCGCCACCATCGA
GACCATCGCCAACCTGCTGCAGGAGGTGGGCGGCAACCACGCCCAGGGCGTGGGCTTCTCCACCGACG
GCTTCGCCACCACCACCACCATGCGCAAGCTGCACCTGATCTGGGTGACCGCCCGCATGCACATCGAG
ATCTACCGCTACCCCGCCTGGTCCGACGTGATCGAGATCGAGACCTGGGTGCAGGGCGAGGGCAAGGT
GGGCACCCGCCGCGACTGGATCCTGAAGGACTACGCCAACGGCGAGGTGATCGGCCGCGCCACCTCCA
AGTGGGTGATGATGAACGAGGACACCCGCCGCCTGCAGAAGGTGTCCGACGACGTGCGCGAGGAGTAC
CTGGTGTTCTGCCCCCGCACCCTGCGCCTGGCCTTCCCCGAGGAGAACAACAACTCCATGAAGAAGAT
CCCCAAGCTGGAGGACCCCGCCGAGTACTCCCGCCTGGGCCTGGTGCCCCGCCGCTCCGACCTGGACA
TGAACAAGCACGTGAACAACGTGACCTACATCGGCTGGGCCCTGGAGTCCATCCCCCCCGAGATCATC
GACACCCACGAGCTGCAGGCCATCACCCTGGACTACCGCCGCGAGTGCCAGCGCGACGACATCGTGGA
CTCCCTGACCTCCCGCGAGCCCCTGGGCAACGCCGCCGGCGTGAAGTTCAAGGAGATCAACGGCTCCG
TGTCCCCCAAGAAGGACGAGCAGGACCTGTCCCGCTTCATGCACCTGCTGCGCTCCGCCGGCTCCGGC
CTGGAGATCAACCGCTGCCGCACCGAGTGGCGCAAGAAGCCCGCCAAGCGCATGGACTACAAGGACCA
CGACGGCGACTACAAGGACCACGACATCGACTACAAGGACGACGACGACAAGTGAATCGATAGATCTC
TTAAGGCAGCAGCAGCTCGGATAGTATCGACACACTCTGGACGCTGGTCGTGTGATGGACTGTTGCCG
CCACACTTGCTGCCTTGACCTGTGAATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATC
TTGTGTGTACGCGCTTTTGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCCAGCATCCC
CTTCCCTCGTTTCATATCGCTTGCATCCCAACCGCAACTTATCTACGCTGTCCTGCTATCCCTCAGCG
CTGCTCCTGCTCCTGCTCACTGCCCCTCGCACAGCCTTGGTTTGGGCTCCGCCTGTATTCTCCTGGTA
CTGCAACCTGTAAACCAGCACTGCAATGCTGATGCACGGGAAGTAGTGGGATGGGAACACAAATGGAA
AGCTTAATTAAGAGCTCCCGCCACCACTCCAACACGGGGTGCCTGGACAAGGACGAGGTGTTTGTGCC
GCCGCACCGCGCAGTGGCGCACGAGGGCCTGGAGTGGGAGGAGTGGCTGCCCATCCGCATGGGCAAGG
TGCTGGTCACCCTGACCCTGGGCTGGCCGCTGTACCTCATGTTCAACGTCGCCTCGCGGCCGTACCCG
CGCTTCGCCAACCACTTTGACCCGTGGTCGCCCATCTTCAGCAAGCGCGAGCGCATCGAGGTGGTCAT
CTCCGACCTGGCGCTGGTGGCGGTGCTCAGCGGGCTCAGCGTGCTGGGCCGCACCATGGGCTGGGCCT
GGCTGGTCAAGACCTACGTGGTGCCCTACCTGATCGTGAACATGTGGCTCGTGCTCATCACGCTGCTC
CAGCACACGCACCCGGCGCTGCCGCACTACTTCGAGAAGGACTGGGACTGGCTGCGCGGCGCCATGGC
CACCGTGGACCGCTCCATGGGCCCGCCCTTCATGGACAACATCCTGCACCACATCTCCGACACCCACG
TGCTGCACCACCTCTTCAGCACCATCCCGCACTACCACGCCGAGGAGGCCTCCGCCGCCATCAGGCCC
ATCCTGGGCAAGTACTACCAGTCCGACAGCCGCTGGGTCGGCCGCGCCCTGTGGGAGGACTGGCGCGA
CTGCCGCTACGTCGTCCCGGACGCGCCCGAGGACGACTCCGCGCTCTGGTTCCACAAGTGAGTGAGTG
AGAAGAGCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTG
TGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGG
GGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAA
CCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCT
CTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCAC
TCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGG
CCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTG
ACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAG
GCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTC
CGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGT
AGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCC
GGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGC

[1827]

[1828] 서열 번호 251

[1829] 프로토테카 모리포르미스 FATA1 녹아웃 상동 재조합 표적화 구조물의 5' 공여체 DNA 서열

GCTCTTCGGAGTCACTGTGCCACTGAGTTCGACTGGTAGCTGAATGGAGTCGCTG
CTCCACTAAACGAATTGTCAGCACCGCCAGCCGGCCGAGGACCCGAGTCATAGC
GAGGGTAGTAGCGCGCCATGGCACCGACCAGCCTGCTTGCCAGTACTGGCGTCTC
TTCCGCTTCTCTGTGGTCCTCTGCGCGCTCCAGCGCGTGCGCTTTTCCGGTGGATC
ATGCGGTCCGTGGCGCACCGCAGCGGCCGCTGCCCATGCAGCGCCGCTGCTTCCG
AACAGTGGCGGTCAGGGCCGCACCCGCGGTAGCCGTCCGTCCGGAACCCGCCCA
AGAGTTTTGGGAGCAGCTTGAGCCCTGCAAGATGGCGGAGGACAAGCGCATCTT
CCTGGAGGAGCACCGGTGCGTGGAGGTCCGGGGCTGACCGGCCGTCGCATTCAA
CGTAATCAATCGCATGATGATCAGAGGACACGAAGTCTTGGTGGCGGTGGCCAG
AAACACTGTCCATTGCAAGGGCATAGGGATGCGTTCCTTCACCTCTCATTTCTCA
TTTCTGAATCCCTCCCTGCTCACTCTTTCTCCTCCTCCTTCCCGTTCACGCAGCATT
CGGGGTACC

[1830]

[1831] 서열 번호 252

[1832] 프로토테카 모리포르미스 FATA1 녹아웃 상동 재조합 표적화 구조물의 3' 공여체 DNA 서열

GACAGGGTGGTTGGCTGGATGGGGAAACGCTGGTCGCGGGATTCGATCCTGCTG
CTTATATCCTCCCTGGAAGCACACCCACGACTCTGAAGAAGAAAACGTGCACAC
ACACAACCCAACCGGCCGAATATTTGCTTCCTTATCCCGGGTCCAAGAGAGACTG
CGATGCCCCCCTCAATCAGCATCCTCCTCCCTGCCGCTTCAATCTTCCCTGCTTGC
CTGCGCCCGCGGTGCGCCGTCTGCCCGCCCAGTCAGTCACTCCTGCACAGGCCCC
TTGTGCGCAGTGCTCCTGTACCCTTTACCGCTCCTTCCATTCTGCGAGGCCCCCTA
TTGAATGTATTCGTTGCCTGTGTGGCCAAGCGGGCTGCTGGGCGCGCCGCCGTCG
GGCAGTGCTCGGCGACTTTGGCGGAAGCCGATTGTTCTTCTGTAAGCCACGCGCT
TGCTGCTTTGGGAAGAGAAGGGGGGGGGTACTGAATGGATGAGGAGGAGAAGG
AGGGGTATTGGTATTATCTGAGTTGGGTGAAGAGC

[1833]

[1834] 서열 번호 253

gctcttcggagtcactgtgccactgagttcgactggtagctgaatggagtcgctgctccactaaacgaattgtcagcaccgcca
gccggccgaggacccgagtcatagcgagggtagtagcgcgccatggcaccgaccagcctgcttgccagtactggcgtctcttc
cgcttctctgtggtcctctgcgcgctccagcgcgtgcgcttttccggtggatcatgcggtccgtggcgcaccgcagcggccgctg
cccatgcagcgccgctgcttccgaacagtggcggtcagggccgcacccgcggtagccgtccgtccggaacccgcccaagagt
ttgggagcagcttgagccctgcaagatggcggaggacaagcgcatcttcctggaggagcaccggtgcgtggaggtccgggg
ctgaccggccgtcgcattcaacgtaatcaatcgcatgatgatcagaggacacgaagtcttggtggcggtggccagaaacact
gtccattgcaagggcatagggatgcgttccttcacctctcatttctcatttctgaatcctccctgctcactctttctcctcctcttc
ccgttcacgcagcattcggggtaccctttcttgcgctatgacacttccagcaaaaggtagggcgggctgcgagacggcttcccggc
gctgcatgcaacaccgatgatgcttcgaccccccgaagctccttcggggctgcatgggcgctccgatgccgctccagggcgagcgc
tgtttaaatagccaggcccccgattgcaaagacattatagcgagctaccaaagccatattcaaacacctagatcactaccacttctacac
aggccactcgagcttgtgatcgcactccgctaaggggcgcctcttcctcttcgtttcagtcacaacccgcaaac**ggcgcgcc*ATG***
ctgctgcaggccttcctgttcctgctggccggcttcgccgccaagatcagcgcctccatgacgaacgagacgtccgaccgccccct
ggtgcacttcaccccctaacaagggctggatgaacgaccccaacggcctgtggtacgacgagaaggacgccaagtggcacctgt
acttccagtacaacccgaacgacaccgtctgggggacgcccttgttctggggccacgccacgtccgacgacctgaccaactggg
aggaccagcccatcgccatcgcccgaagcgcaacgactccggcgccttctccggctccatggtggtggactacaacaacaacct
ccggcttcttcaacgacaccatcgacccgcgccagcgctgcgtggccatctggacctacaacacccccggagtccgaggagcagt
acatctcctacagcctggacggcggctacaccttcaccgagtaccagaagaaccccgtgctggccgccaactccaccccagttcc
gcgacccgaaggtcttctggtacgagccctcccagaagtggatcatgaccgcggccaagtcccaggactacaagatcgagatct
actcctccgacgacctgaagtcctggaagctggagtccgcgttcgccaacgagggcttcctcggctaccagtacgagtgccccgg
cctgatcgaggtccccaccgagcaggaccccagcaagtcctactgggtgatgttcatctccatcaaccccggcgccccggccgg

[1835]

cggctccttcaaccagtacttcgtcggcagcttcaacggcacccacttcgaggccttcgacaaccagtcccgcgtggtggacttcg
gcaaggactactacgccctgcagaccttcttcaacaccgacccgacctacgggagcgccctgggcatcgcgtgggcctccaact
gggagtactccgccttcgtgcccaccaaccccttggcgctcctccatgtccctcgtgcgcaagttctccctcaacaccgagtaccag
gccaaccccggagacggagctgatcaacctgaaggccgagccgatcctgaacatcagcaacgccggcccctggagccggttcg
ccaccaacaccacgttgacgaaggccaacagctacaacgtcgacctgtccaacagcaccggcaccctggagttcgagctggtg
tacgccgtcaacaccacccagacgatctccaagtccgtgttcgcggacctctccctctggttcaagggcctggaggaccccgagg
agtacctccgcatgggcttcgaggtgtccgcgtcctccttcttcctggaccgcgggaacagcaaggtgaagttcgtgaaggagaa
cccctacttcaccaaccgcatgagcgtgaacaaccagccccttcaagagcgagaacgacctgtcctactacaaggtgtacggcttg
ctggaccagaacatcctggagctgtacttcaacgacggcgacgtcgtgtccaccaacacctacttcatgaccaccgggaacgcc
*ctgggctccgtgaacatgacgacgggggtggacaacctgttctacatcgacaagttccaggtgcgcgaggtcaag**TGA***caatt**
ggcagcagcagctcggatagtatcgacacactctggacgctggtcgtgtgatggactgttgccgccacacttgctgccttgacctgtga
atatccctgccgcttttatcaaacagcctcagtgtgtttgatcttgtgtgtacgcgcttttgcgagttgctagctgcttgtgctatttgcgaata
ccacccccagcatccccttccctcgtttcatatcgcttgcatcccaaccgcaacttatctacgctgtcctgctatccctcagcgctgctcct
gctcctgctcactgcccctcgcacagccttggtttgggctccgcctgtattctcctggtactgcaacctgtaaaccagcactgcaatgctg
atgcacgggaagtagtgggatgggaacacaaatggaggatcgta**gagctc**actagtatcgatttcgaa**gacagggtggttggctgg**
atggggaaacgctggtcgcgggattcgatcctgctgcttatatcctccctggaagcacacccacgactctgaagaagaaaacg
tgcacacacaacaacccaaccggccgaatatttgcttccttatcccgggtccaagagagactgcgatgccccccctcaatcagcat
cctcctccctgccgcttcaatcttccctgcttgcctgcgcccgcggtgcgccgtctgcccgcccagtcagtcactcctgcacaggc
cccttgtgcgcagtgctcctgtaccctttaccgctccttccattctgcgaggcccccctattgaatgtattcgttgcctgtgtggcca
agcgggctgctgggcgcgccgccgtcgggcagtgctcggcgactttggcggaagccgattgttcttctgtaagccacgcgcttg
ctgctttgggaagagaagggggggggtactgaatggatgaggaggagaaggagggtattggtattatctgagttgggtgaa
gagc

[1836]

[1837] 서열 번호 254

gctcttcggagtcactgtgccactgagttcgactggtagctgaatggagtcgctgctccactaaacgaattgtcagcaccgcca
gccggccgaggacccgagtcatagcgagggtagtagcgcgccatggcaccgaccagcctgcttgccagtactggcgtctcttc
cgcttctctgtggtcctctgcgcgctccagcgcgtgcgcttttccggtggatcatgcggtccgtggcgcaccgcagcggccgctg
cccatgcagcgccgctgcttccgaacagtggcggtcagggccgcacccgcggtagccgtccgtccggaacccgcccaagagt
ttgggagcagcttgagccctgcaagatggcggaggacaagcgcatcttcctggaggagcaccggtgcgtggaggtccgggg
ctgaccggccgtcgcattcaacgtaatcaatcgcatgatgatcagaggacacgaagtcttggtggcggtggccagaaacact
gtccattgcaagggcatagggatgcgttccttcacctctcatttctcatttctgaatccctccctgctcactctttctcctcctcttc
ccgttcacgcagcattcggggtaccctttcttgcgctatgacacttccagcaaaaggtagggcgggctgcgagacggcttcccggc
gctgcatgcaacaccgatgatgcttcgaccccccgaagctccttcggggctgcatgggcgctccgatgccgctccagggcgagcgc
tgtttaaatagccaggcccccgattgcaaagacattatagcgagctaccaaagccatattcaaacacctagatcactaccacttctacac
aggccactcgagcttgtgatcgcactccgctaaggggcgcctcttcctcttcgtttcagtcacaacccgcaaac**ggcgcgcc*ATG***
ctgctgcaggccttcctgttcctgctggccggcttcgccgccaagatcagcgcctccatgacgaacgagacgtccgaccgccccct
ggtgcacttcacccccaacaagggctggatgaacgaccccaacggcctgtggtacgacgagaaggacgccaagtggcacctgt
acttccagtacaacccgaacgacaccgtctgggggacgcccttgttctggggccacgccacgtccgacgacctgaccaactggg
aggaccagcccatcgccatcgccccgaagcgcaacgactccggcgccttctccggctccatggtggtggactacaacaacacct
ccggcttcttcaacgacaccatcgacccgcgccagcgctgcgtggccatctggacctacaacacccccggagtccgaggagcagt
acatctcctacagcctggacggcggctacaccttcaccgagtaccagaagaaccccgtgctggccgccaactccacccagttcc
gcgacccgaaggtcttctggtacgagccctcccagaagtggatcatgaccgcggccaagtcccaggactacaagatcgagatct
actcctccgacgacctgaagtcctggaagctggagtccgcgttcgccaacgagggcttcctcggctaccagtacgagtgccccgg
cctgatcgaggtccccaccgagcaggaccccagcaagtcctactgggtgatgttcatctccatcaaccccggcgccccggccgg
cggctccttcaaccagtacttcgtcggcagcttcaacggcacccacttcgaggccttcgacaaccagtcccgcgtggtggacttcg
gcaaggactactacgccctgcagaccttcttcaacaccgacccgacctacgggagcgccctgggcatcgcgtgggcctccaact
gggagtactccgccttcgtgcccaccaaccctggcgctcctccatgtccctcgtgcgcaagttctccctcaacaccgagtaccag
gccaacccggagacggagctgatcaacctgaaggccgagccgatcctgaacatcagcaacgccggccccctggagccggttcg
ccaccaacaccacgttgacgaaggccaacagctacaacgtcgacctgtccaacagcaccggcaccctggagttcgagctggtg

[1838]



[1839]

attgaatgtattcgttgcctgtgtggccaagcgggctgctgggcgcgccgccgtcgggcagtgctcggcgactttggcggaagc
cgattgttcttctgtaagccacgcgcttgctgctttgggaagagaaggggggggtactgaatggatgaggaggagaaggag
ggtattggtattatctgagttgggtgaagagc

[1840]

[1841] 서열 번호 255

[1842] RNAi 발현 카세트의 FATA 부분

ATGGCACCGACCAGCCTGCTTGCCAGTACTGGCGTCTCTTCCGCTTCTCTGTGGTC
CTCTGCGCGCTCCAGCGCGTGCGCTTTTCCGGTGGATCATGCGGTCCGTGGCGCA
CCGCAGCGGCCGCTGCCCATGCAGCGCCGCTGCTTCCGAACAGTGGCGGTCAGG
GCCGCACCCGCGGTAGCCGTCCGTCCGGAACCCGCCCAAGAGTTTTGGGAGCAG
CTTGAGCCCTGCAAGATGGCGGAGGACAAGCGCATCTTCCTGGAGGAGCACCGG
TGCGTGGAGGTCCGGGGCTGACCGGCCGTCGCATTCAACGTAATCAATCGCATG
ATGATCAGAGGACACGAAGTCTTGGTGGCGGTGGCCAGAAACACTGTCCATTGC
AAGGGCATAGGGATGCGTTCCTTCACCTCTCATTTCTCATTTCTGAATCCCTCCCT
GCTCACTCTTTCTCCTCCTCCTTCCCGTTCACGCAGCATTCGGGGCAACGAGGTGG
GCCCGTGCTCCTCCAGGAAGATGCGCTTGTCCTCCGCCATCTTGCAGGGCTCAAG
CTGCTCCCAAAACTCTTGGGCGGGTTCCGGACGGACGGCTACCGCGGGTGCGGC
CCTGACCGCCACTGTTCGGAAGCAGCGGCGCTGCATGGGCAGCGGCCGCTGCGG
TGCGCCACGGACCGCATGATCCACCGGAAAAGCGCACGCGCTGGAGCGCGCAGA
GGACCACAGAGAAGCGGAAGAGACGCCAGTACTGGCAAGCAGGCTGGTCGGTG
CCAT

[1843]

[1844] 서열 번호 256

[1845] pSZ1283

CACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTC
TTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAG
CCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGG
TGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGAT
CTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATT
ATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTAT
ATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTG
TCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTA
CCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCA
ATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCA
GTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGT
TGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCC
CAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCT
CCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAAT
TCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCT
GAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCA
CATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATC
TTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTA
CTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAG
GGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGT
TATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGC
ACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAA
AATAGGCGTATCACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACAC
ATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCA
GGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTG
TACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATC
AGGCGCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTA

[1846]

[1847]

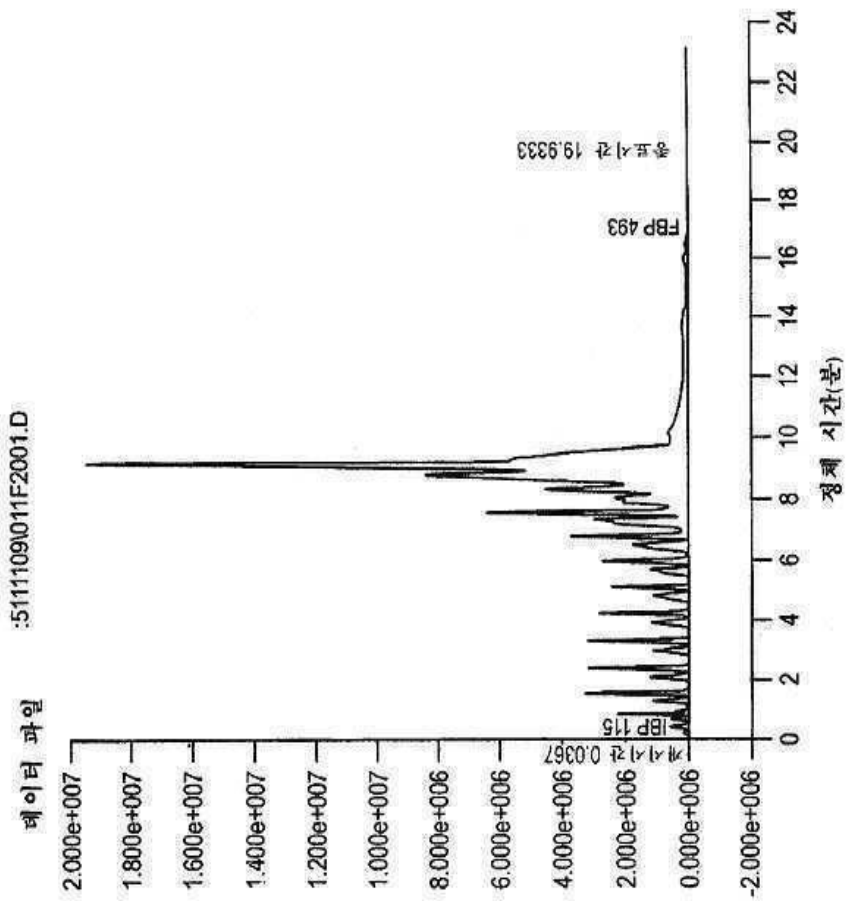

TTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTC
CCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTGATGCATGCTCTTCGCCGCCGCCACTCCTG
CTCGAGCGCGCCCGCGCGTGCGCCGCCAGCGCCTTGGCCTTTTCGCCGCGCTCGTGCGCGTCGCTG
ATGTCCATCACCAGGTCCATGAGGTCTGCCTTGCGCCGGCTGAGCCACTGCTTCGTCCGGGCGGCC
AAGAGGAGCATGAGGGAGGACTCCTGGTCCAGGGTCCTGACGTGGTCGCGGCTCTGGGAGCGGGC
CAGCATCATCTGGCTCTGCCGCACCGAGGCCGCCTCCAACTGGTCCTCCAGCAGCCGCAGTCGCCG
CCGACCCTGGCAGAGGAAGACAGGTGAGGGGGGTATGAATTGTACAGAACAACCACGAGCCTTGT
CTAGGCAGAATCCCTACCAGTCATGGCTTTACCTGGATGACGGCCTGCGAACAGCTGTCCAGCGAC
CCTCGCTGCCGCCGCTTCTCCCGCACGCTTCTTTCCAGCACCGTGATGGCGCGAGCCAGCGCCGCA
CGCTGGCGCTGCGCTTCGCCGATCTGAGGACAGTCGGGGAACTCTGATCAGTCTAAACCCCCTTGC
GCGTTAGTGTTGCCATCCTTTGCAGACCGGTGAGAGCCGACTTGTTGTGCGCCACCCCCCACACCA
CCTCCTCCCAGACCAATTCTGTCACCTTTTTGGCGAAGGCATCGGCCTCGGCCTGCAGAGAGGACA
GCAGTGCCCAGCCGCTGGGGGTTGGCGGATGCACGCTCAGGTACCCTTTCTTGCGCTATGACACTT
CCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTTCCCGGCGCTGCATGCAACACCGATGATGCTT
CGACCCCCCGAAGCTCCTTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGTTT
AAATAGCCAGGCCCCCGATTGCAAAGACATTATAGCGAGCTACCAAAGCCATATTCAAACACCTA
GATCACTACCACTTCTACACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGGGCGCCTCTT
CCTCTTCGTTTCAGTCACAACCCGCAAACGGCGCGCCATGCTGCTGCAGGCCTTCCTGTTCCTGCTG
GCCGGCTTCGCCGCCAAGATCAGCGCCTCCATGACGAACGAGACGTCCGACCGCCCCCTGGTGCA
CTTCACCCCCAACAAGGGCTGGATGAACGACCCCAACGGCCTGTGGTACGACGAGAAGGACGCCA
AGTGGCACCTGTACTTCCAGTACAACCCGAACGACACCGTCTGGGGGACGCCCTTGTTCTGGGGCC
ACGCCACGTCCGACGACCTGACCAACTGGGAGGACCAGCCCATCGCCATCGCCCCGAAGCGCAAC
GACTCCGGCGCCTTCTCCGGCTCCATGGTGGTGGACTACAACAACACCTCCGGCTTCTTCAACGAC
ACCATCGACCCGCGCCAGCGCTGCGTGGCCATCTGGACCTACAACACCCCGGAGTCCGAGGAGCA
GTACATCTCCTACAGCCTGGACGGCGGCTACACTTCACCGAGTACCAGAAGAACCCCGTGCTGGC
CGCCAACTCCACCCAGTTCCGCGACCCGAAGGTCTTCTGGTACGAGCCCTCCCAGAAGTGGATCAT
GACCGCGGCCAAGTCCCAGGACTACAAGATCGAGATCTACTCCTCCGACGACCTGAAGTCCTGGA
AGCTGGAGTCCGCGTTCGCCAACGAGGGCTTCCTCGGCTACCAGTACGAGTGCCCCGGCCTGATCG
AGGTCCCCACCGAGCAGGACCCCAGCAAGTCCTACTGGGTGATGTTCATCTCCATCAACCCCGGCG
CCCCGGCCGGCGGCTCCTTCAACCAGTACTTCGTCGGCAGCTTCAACGGCACCCACTTCGAGGCCT
TCGACAACCAGTCCCGCGTGGTGGACTTCGGCAAGGACTACTACGCCCTGCAGACCTTCTTCAACA
CCGACCCGACCTACGGGAGCGCCCTGGGCATCGCGTGGGCCTCCAACTGGGAGTACTCCGCCTTCG
TGCCCACCAACCCCTGGCGCTCCTCCATGTCCCTCGTGCGCAAGTTCTCCCTCAACACCGAGTACC
AGGCCAACCCGGAGACGGAGCTGATCAACCTGAAGGCCGAGCCGATCCTGAACATCAGCAACGCC
GGCCCCTGGAGCCGGTTCGCCACCAACACCACCTTGACGAAGGCCAACAGCTACAACGTCGACCT
GTCCAACAGCACCGGCACCCTGGAGTTCGAGCTGGTGTACGCCGTCAACACCACCCAGACGATCT
CCAAGTCCGTGTTCGCGGACCTCTCCCTCTGGTTCAAGGGCCTGGAGGACCCCGAGGAGTACCTCC
GCATGGGCTTCGAGGTGTCCGCGTCCTCCTTCTTCCTGGACCGCGGGAACAGCAAGGTGAAGTTCG
TGAAGGAGAACCCCTACTTCACCAACCGCATGAGCGTGAACAACCAGCCCTTCAAGAGCGAGAAC
GACCTGTCCTACTACAAGGTGTACGGCTTGCTGGACCAGAACATCCTGGAGCTGTACTTCAACGAC
GGCGACGTCGTGTCCACCAACACCTACTTCATGACCACCGGGAACGCCCTGGGCTCCGTGAACATG
ACGACGGGGGTGGACAACCTGTTCTACATCGACAAGTTCCAGGTGCGCGAGGTCAAGTGACAATT
GGCAGCAGCAGCTCGGATAGTATCGACACACTCTGGACGCTGGTCGTGTGATGGACTGTTGCCGCC
ACACTTGCTGCCTTGACCTGTGAATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCT
TGTGTGTACGCGCTTTTGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCCAGCATCCC
CTTCCCTCGTTTCATATCGCTTGCATCCCAACCGCAACTTATCTACGCTGTCCTGCTATCCCTCAGC
GCTGCTCCTGCTCCTGCTCACTGCCCCTCGCACAGCCTTGGTTTGGGCTCCGCCTGTATTCTCCTGG
TACTGCAACCTGTAAACCAGCACTGCAATGCTGATGCACGGGAAGTAGTGGGATGGGAACACAAA
TGGAGGATCCCGCGTCTCGAACAGAGCGCGCAGAGGAACGCTGAAGGTCTCGCCTCTGTCGCACC
TCAGCGCGGCATACACCACAATAACCACCTGACGAATGCGCTTGGTTCTTCGTCCATTAGCGAAGC
GTCCGGTTCACACACGTGCCACGTTGGCGAGGTGGCAGGTGACAATGATCGGTGGAGCTGATGGT
CGAAACGTTCACAGCCTAGGGATATCGAATTCCGCCGACAGGACGCGCGTCAAAGGTGCTGGTCG
TGTATGCCCTGGCCGGCAGGTCGTTGCTGCTGCTGGTTAGTGATTCCGCAACCCTGATTTTGGCGTC
TTATTTTGGCGTGGCAAACGCTGGCGCCCGCGAGCCGGGCCGGCGGCGATGCGGTGCCCCACGGC
TGCCGGAATCCAAGGGAGGCAAGAGCGCCCGGGTCAGTTGAAGGGCTTTACGCGCAAGGTACAGC
CGCTCCTGCAAGGCTGCGTGGTGGAATTGGACGTGCAGGTCCTGCTGAAGTTCCTCCACCGCCTCA
CCAGCGGACAAAGCACCGGTGTATCAGGTCCGTGTCATCCACTCTAAAGAGCTCGACTACGACCT
ACTGATGGCCCTAGATTCTTCATCAAAAACGCCTGAGACACTTGCCCAGGATTGAAACTCCCTGAA
GGGACCACCAGGGGCCCTGAGTTGTTCCTTCCCCCCGTGGCGAGCTGCCAGCCAGGCTGTACCTGT
GATCGAGGCTGGCGGGAAAATAGGCTTCGTGTGCTCAGGTCATGGGAGGTGCAGGACAGCTCATG

AAACGCCAACAATCGCACAATTCATGTCAAGCTAATCAGCTATTTCCTCTTCACGAGCTGTAATTG
TCCCAAAATTCTGGTCTACCGGGGGGTGATCCTTCGTGTACGGGCCCTTCCCTCAACCCTAGGTATG
CGCGCATGCGGTCGCCGCGCAACTCGCGCGAGGGCCGAGGGTTTGGGACGGGCCGTCCCGAAATG
CAGTTGCACCCGGATGCGTGGCACCTTTTTTGCGATAATTTATGCAATGGACTGCTCTGCAAAATT
CTGGCTCTGTCGCCAACCCTAGGATCAGCGGCGTAGGATTTCGTAATCATTCGTCCTGATGGGGAG
CTACCGACTACCCTAATATCAGCCCGACTGCCTGACGCCAGCGTCCACTTTTGTGCACACATTCCA
TTCGTGCCCAAGACATTTCATTGTGGTGCGAAGCGTCCCCAGTTACGCTCACCTGTTTCCCGACCTC
CTTACTGTTCTGTCGACAGAGCGGGCCCACAGGCCGGTCGCAGCCACTAGTATGGTGGTGGCCGCC
GCCGCCAGCAGCGCCTTCTTCCCCGTGCCCGCCCCCCGCCCACCCCCAAGCCCGGCAAGTTCGGC
AACTGGCCCAGCAGCCTGAGCCAGCCCTTCAACCCCAAGAGCAACCCCAACGGCCGCTTCCAGGT
GAAGGCCAACGTGAGCCCCCACGGGCGCGCCCCCAAGGCCAACGGCAGCGCCGTGAGCCTGAAGT
CCGGCAGCCTGAACACCCTGGAGGACCCCCCCCAGCAGCCCCCCCCCCCCGCACCTTCCTGAACCAGC
TGCCCGACTGGAGCCGCCTGCGCACCGCCATCACCACCGTGTTCGTGGCCGCCGAGAAGCAGTTCA
CCCGCCTGGACCGCAAGAGCAAGCGCCCCGACATGCTGGTGGACTGGTTCGGCAGCGAGACCATC
GTGCAGGACGGCCTGGTGTTCCGCGAGCGCTTCAGCATCCGCAGCTACGAGATCGGCGCCGACCG
CACCGCCAGCATCGAGACCCTGATGAACCACCTGCAGGACACCAGCCTGAACCACTGCAAGAGCG
TGGGCCTGCTGAACGACGGCTTCGGCCGCACCCCCGAGATGTGCACCCGCGACCTGATCTGGGTGC
TGACCAAGATGCAGATCGTGGTGAACCGCTACCCCACCTGGGGCGACACCGTGGAGATCAACAGC
TGGTTCAGCCAGAGCGGCAAGATCGGCATGGGCCGCGAGTGGCTGATCAGCGACTGCAACACCGG
CGAGATCCTGGTGCGCGCCACCAGCGCCTGGGCCATGATGAACCAGAAGACCCGCCGCTTCAGCA
AGCTGCCCTGCGAGGTGCGCCAGGAGATCGCCCCCCACTTCGTGGACGCCCCCCCCGTGATCGAG
GACAACGACCGCAAGCTGCACAAGTTCGACGTGAAGACCGGCGACAGCATCTGCAAGGGCCTGAC
CCCCGGCTGGAACGACTTCGACGTGAACCAGCACGTGAGCAACGTGAAGTACATCGGCTGGATTC
TGGAGAGCATGCCCACCGAGGTGCTGGAGACCCAGGAGCTGTGCAGCCTGACCCTGGAGTACCGC
CGCGAGTGCGGCCGCGAGAGCGTGGTGGAGAGCGTGACCAGCATGAACCCCAGCAAGGTGGGCG
ACCGCAGCCAGTACCAGCACCTGCTGCGCCTGGAGGACGGCGCCGACATCATGAAGGGCCGCACC
GAGTGGCGCCCCAAGAACGCCGGCACCAACCGCGCCATCAGCACCTGATTAATTAACTCGAGGCA
GCAGCAGCTCGGATAGTATCGACACACTCTGGACGCTGGTCGTGTGATGGACTGTTGCCGCCACAC
TTGCTGCCTTGACCTGTGAATATCCCTGCCGCTTTTATCAAACAGCCTCAGTGTGTTTGATCTTGTG
TGTACGCGCTTTTGCGAGTTGCTAGCTGCTTGTGCTATTTGCGAATACCACCCCCAGCATCCCCTTC
CCTCGTTTCATATCGCTTGCATCCCAACCGCAACTTATCTACGCTGTCCTGCTATCCCTCAGCGCTG
CTCCTGCTCCTGCTCACTGCCCCTCGCACAGCCTTGGTTTGGGCTCCGCCTGTATTCTCCTGGTACT
GCAACCTGTAAACCAGCACTGCAATGCTGATGCACGGGAAGTAGTGGGATGGGAACACAAATGGA
AAGCTTGAGCTCTTGTTTTCCAGAAGGAGTTGCTCCTTGAGCCTTTCATTCTCAGCCTCGATAACCT
CCAAAGCCGCTCTAATTGTGGAGGGGGTTCGAATTTAAAAGCTTGGAATGTTGGTTCGTGCGTCTG
GAACAAGCCCAGACTTGTTGCTCACTGGGAAAAGGACCATCAGCTCCAAAAAACTTGCCGCTCAA
ACCGCGTACCTCTGCTTTCGCGCAATCTGCCCTGTTGAAATCGCCACCACATTCATATTGTGACGCT
TGAGCAGTCTGTAATTGCCTCAGAATGTGGAATCATCTGCCCCCTGTGCGAGCCCATGCCAGGCAT
GTCGCGGGCGAGGACACCCGCCACTCGTACAGCAGACCATTATGCTACCTCACAATAGTTCATAAC
AGTGACCATATTTCTCGAAGCTCCCCAACGAGCACCTCCATGCTCTGAGTGGCCACCCCCCGGCCC
TGGTGCTTGCGGAGGGCAGGTCAACCGGCATGGGGCTACCGAAATCCCCGACCGGATCCCACCAC
CCCCGCGATGGGAAGAATCTCTCCCCCGGGATGTGGGCCCACCACCAGCACAACCTGCTGGCCCAG
GCGAGCGTCAAACCATACCACACAAATATCCTTGGCATCGGCCCTGAATTCCTTCTGCCGCTCTGC
TACCCGGTGCTTCTGTCCGAAGCAGGGGTTGCTAGGGATCGCTCCGAGTCCGCAAACCCTTGTCGC
GTGGCGGGGCTTGTTCGAGCTTGAAGAGCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGGCGT
AATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAG
CCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTG
CGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGC
GCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCG
GTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATC
AGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAA
GGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTC
AAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCC
TCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAG
CGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTG
GGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAG
TCCAACCCGGTAAGACACGACTTATCGC

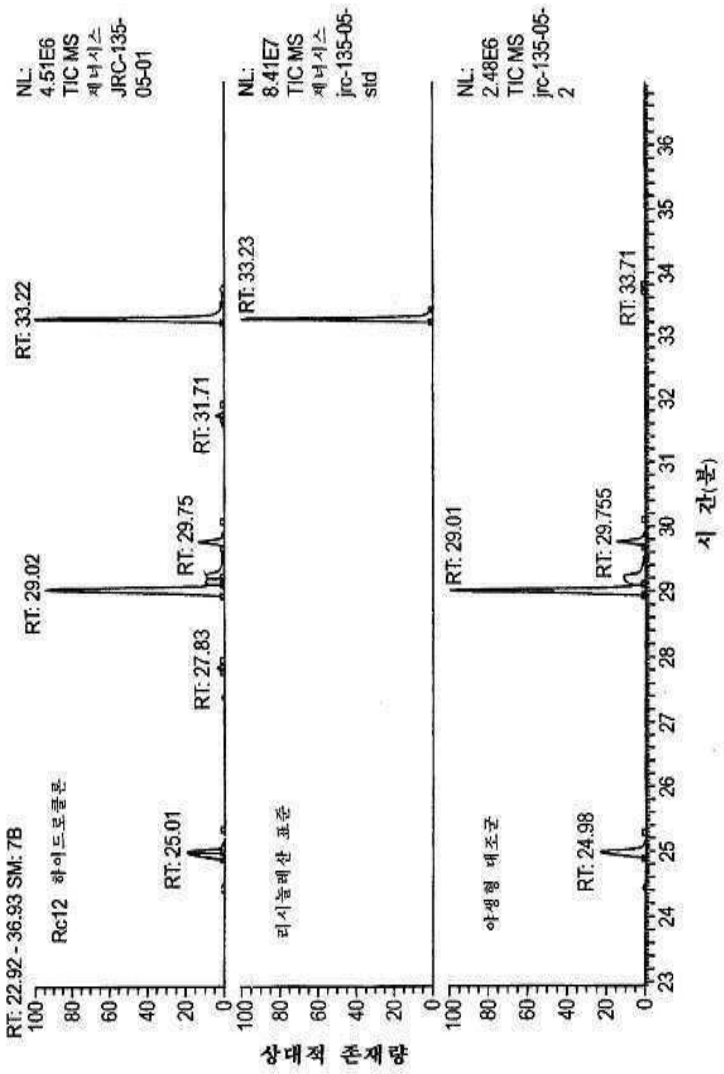
[1848]

도면

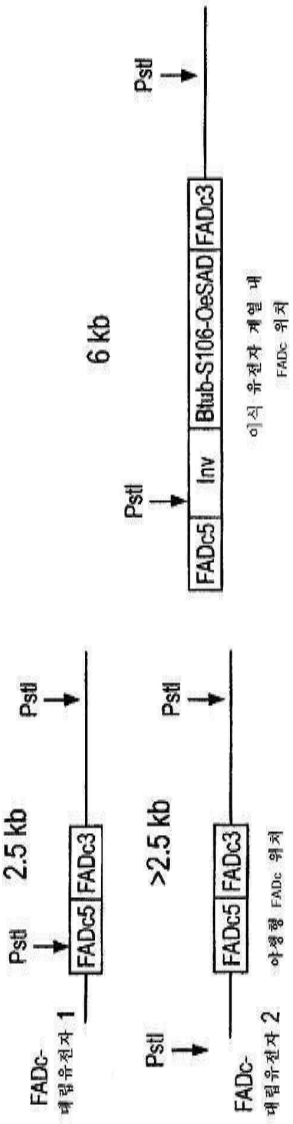
도면1


데이터 파일 :5111109011F2001.D
2.000e+007
1.800e+007
1.600e+007
1.400e+007
1.200e+007
1.000e+007
8.000e+006
6.000e+006
4.000e+006
2.000e+006
0.000e+006
-2.000e+006
0
2
4
6
8
10
12
14
16
18
20
22
24
정체 시간(분)
IBP 115
FBP 493

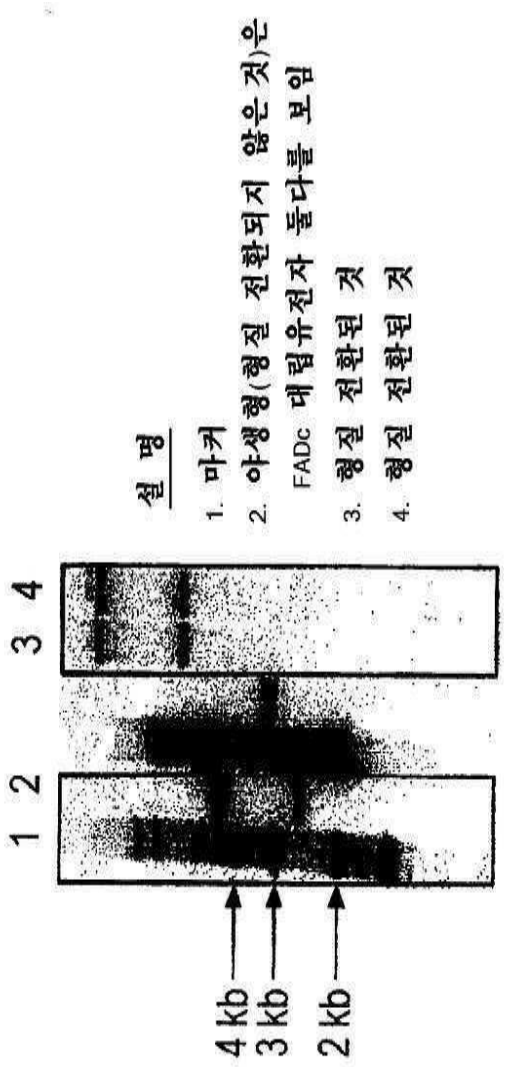
도면2


RT: 22.92 - 36.93 SM: 7B
Rc12 하이드로클론
RT: 25.01
RT: 27.83
RT: 29.02
RT: 29.75
RT: 31.71
RT: 33.22
NL: 4.51E6 TIC MS 제너시스 JRC-135-05-01
리시놀레산 표준
RT: 33.23
NL: 8.41E7 TIC MS 제너시스 jrc-135-05-std
야생형 대조군
RT: 24.98
RT: 29.01
RT: 29.755
RT: 33.71
NL: 2.48E6 TIC MS jrc-135-05-2
상대적 존재량
시 간(분)

도면3


PstI
6 kb
FADc5
Inv
Btub-S106-OeSAD
FADc3
PstI
이식 유전자 재열 내
FADc 위치
PstI
2.5 kb
FADc5
FADc3
PstI
FADc-
대립유전자 1
PstI
>2.5 kb
FADc5
FADc3
PstI
FADc-
대립유전자 2
야생형 FADc 위치

도면4



서 열 목 록

SEQUENCE LISTING

<110> SOLAZYME, INC.

<120> Tailored Oils Produced from Recombinant Oleaginous Microorganisms

<130> 057518-415312 (068PCT)

<140> PCT/US12/023696

<141> 2012-02-02

<150> US 61/438,969

<151> 2011-02-02

<150> US 61/476,691

<151> 2011-04-18

<150> US 61/484,458

<151> 2011-05-10

<150> US 61/548,616

<151> 2011-10-18

<160> 256

<170> PatentIn version 3.5

<210> 1

<211> 1187

<212> DNA

<213> Unknown

<220><223> Chlorella species

<400> 1

gatcagacgg gcctgacctg cgagataatc aagtgctcgt aggcaaccaa ctcagcagct 60

gcttggtgtt gggtctgcag gatagtgttg cagggcccca aggacagcag gggaacttac 120

accttgtccc cgacccagtt ttatggagtg cattgcctca agagcctagc cggagcgcta 180

ggctacatac ttgccgcacc ggtatgaggg gatatagtac tcgcactgcg ctgtctagtg 240

agatgggcag tgctgcccat aaacaactgg ctgctcagcc atttgttggc ggaccattct 300

gggggggcca gcaatgcctg actttcgggt agggtgaaaa ctgaacaaag actaccaaaa 360

cagaatttct tcctccttgg aggtaagcgc aggccggccc gcctgcgccc acatggcgct 420

ccgaacacct ccatagctgt aagggcgcaa acatggccgg actgttgtca gcactctttc 480

atggccatac aaggtcatgt cgagattagt gctgagtaag acactatcac cccatgttcg 540

attgaagccg tgacttcatg ccaacctgcc cctgggcgta gcagacgtat gccatcatga 600

ccactagccg acatgcgctg tcttttgcca ccaaaacaac tggtacaccg ctcgaagtcg 660

tgccgcacac ctccgggagt gagtccggcg actcctcccc ggcgggccgc ggccctacct 720

gggtagggtc gccatacgcc cacgaccaaa cgacgcagga ggggattggg gtagggaatc 780

ccaaccagcc taaccaagac ggcacctata ataataggtg gggggactaa cagccctata 840

tcgcaagctt tgggtgccta tcttgagaag cacgagttgg agtggctgtg tacggtcgac 900

cctaaggtgg gtgtgccgca gcctgaaaca aagcgtctag cagctgcttc tataatgtgt 960

cagccgttgt gtttcagtta tattgtatgc tattgtttgt tcgtgctagg gtggcgcagg 1020

cccacctact gtggcgggcc attggttggt gcttgaattg cctcaccatc taaggtctga 1080

acgctcactc aaacgccttt gtacaactgc agaactttcc ttggcgctgc aactacagtg 1140

tgcaaaccag cacatagcac tcccttacat cacccagcag tacaaca 1187

<210> 2

<211> 1414

<212> DNA

<213> Chlorella ellipsoidea

<400> 2

cgctgcgcac cagggccgcc agctcgctga tgtcgctcca aatgcggtcc cccgattttt 60

tgttcttcat cttctccacc ttggtggcct tcttggccag ggccttcagc tgcatgcgca 120

cagaccgttg agctcctgat cagcatcctc aggaggccct ttgacaagca agcccctgtg 180

caagcccatt cacggggtac cagtggtgct gaggtagatg ggtttgaaaa ggattgctcg 240

gtcgattgct gctcatggaa ttggcatgtg catgcatgtt cacaatatgc caccaggctt 300

tggagcaaga gagcatgaat gccttcaggc aggttgaaag ttcctgggggg tgaagaggca 360

gggccgagga ttggaggagg aaagcatcaa gtcgtcgctc atgctcatgt tttcagtcag 420

agtttgccaa gctcacagga gcagagacaa gactggctgc tcaggtgttg catcgtgtgt 480

gtggtggggg gggggggggtt aatacggtac gaaatgcact tggaattccc acctcatgcc 540

agcggaccca catgcttgaa ttcgaggcct gtggggtgag aaatgctcac tctgccctcg 600

ttgctgaggt acttcaggcc gctgagctca aagtcgatgc cctgctcgtc tatcagggcc 660

tgcacctctg ggctgaccgg ctcagcctcc ttcgcgggca tggagtaggc gccggcagcg 720

ttcatgtccg ggcccagggc agcggtggtg ccataaatgt cggtgatggt ggggaggggg 780

gccgtcgcca caccattgcc gttgctggct gacgcatgca catgtggcct ggctggcacc 840

ggcagcactg gtctccagcc agccagcaag tggctgttca ggaaagcggc catgttgttg 900

gtccctgcgc atgtaattcc ccagatcaaa ggagggaaca gcttggattt gatgtagtgc 960

ccaaccggac tgaatgtgcg atggcaggtc cctttgagtc tcccgaatta ctagcagggc 1020

actgtgacct aacgcagcat gccaaccgca aaaaaatgat tgacagaaaa tgaagcggtg 1080

tgtcaatatt tgctgtattt attcgtttta atcagcaacc aagttcgaaa cgcaactatc 1140

gtggtgatca agtgaacctc atcagactta cctcgttcgg caaggaaacg gaggcaccaa 1200

attccaattt gatattatcg cttgccaagc tagagctgat ctttgggaaa ccaactgcca 1260

gacagtggac tgtgatggag tgccccgagt ggtggagcct cttcgattcg gttagtcatt 1320

actaacgtga accctcagtg aagggaccat cagaccagaa agaccagatc tcctcctcga 1380

caccgagaga gtgttgcggc agtaggacga caag 1414

<210> 3

<211> 512

<212> PRT

<213> Unknown

<220><223> Yeast sequence

<400> 3

Met Thr Asn Glu Thr Ser Asp Arg Pro Leu Val His Phe Thr Pro Asn
1 5 10 15

Lys Gly Trp Met Asn Asp Pro Asn Gly Leu Trp Tyr Asp Glu Lys Asp
20 25 30

Ala Lys Trp His Leu Tyr Phe Gln Tyr Asn Pro Asn Asp Thr Val Trp
35 40 45

Gly Thr Pro Leu Phe Trp Gly His Ala Thr Ser Asp Asp Leu Thr Asn
50 55 60

Trp Glu Asp Gln Pro Ile Ala Ile Ala Pro Lys Arg Asn Asp Ser Gly
65 70 75 80

Ala Phe Ser Gly Ser Met Val Val Asp Tyr Asn Asn Thr Ser Gly Phe
85 90 95

Phe Asn Asp Thr Ile Asp Pro Arg Gln Arg Cys Val Ala Ile Trp Thr
100 105 110

Tyr Asn Thr Pro Glu Ser Glu Glu Gln Tyr Ile Ser Tyr Ser Leu Asp
115 120 125

Gly Gly Tyr Thr Phe Thr Glu Tyr Gln Lys Asn Pro Val Leu Ala Ala
130 135 140

Asn Ser Thr Gln Phe Arg Asp Pro Lys Val Phe Trp Tyr Glu Pro Ser
145 150 155 160

Gln Lys Trp Ile Met Thr Ala Ala Lys Ser Gln Asp Tyr Lys Ile Glu
165 170 175

Ile Tyr Ser Ser Asp Asp Leu Lys Ser Trp Lys Leu Glu Ser Ala Phe
180 185 190

Ala Asn Glu Gly Phe Leu Gly Tyr Gln Tyr Glu Cys Pro Gly Leu Ile
195 200 205

Glu Val Pro Thr Glu Gln Asp Pro Ser Lys Ser Tyr Trp Val Met Phe
210 215 220

Ile Ser Ile Asn Pro Gly Ala Pro Ala Gly Gly Ser Phe Asn Gln Tyr
225 230 235 240
Phe Val Gly Ser Phe Asn Gly Thr His Phe Glu Ala Phe Asp Asn Gln
245 250 255
Ser Arg Val Val Asp Phe Gly Lys Asp Tyr Tyr Ala Leu Gln Thr Phe
260 265 270
Phe Asn Thr Asp Pro Thr Tyr Gly Ser Ala Leu Gly Ile Ala Trp Ala

275 280 285
Ser Asn Trp Glu Tyr Ser Ala Phe Val Pro Thr Asn Pro Trp Arg Ser
290 295 300
Ser Met Ser Leu Val Arg Lys Phe Ser Leu Asn Thr Glu Tyr Gln Ala
305 310 315 320
Asn Pro Glu Thr Glu Leu Ile Asn Leu Lys Ala Glu Pro Ile Leu Asn
325 330 335
Ile Ser Asn Ala Gly Pro Trp Ser Arg Phe Ala Thr Asn Thr Thr Leu

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

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

465 470 475 480

Phe Met Thr Thr Gly Asn Ala Leu Gly Ser Val Asn Met Thr Thr Gly

485 490 495

Val Asp Asn Leu Phe Tyr Ile Asp Lys Phe Gln Val Arg Glu Val Lys

500 505 510

<210> 4

<211> 20

<212> PRT

<213> Unknown

<220><223> Yeast sequence

<400> 4

Met Leu Leu Gln Ala Phe Leu Phe Leu Leu Ala Gly Phe Ala Ala Lys

1 5 10 15

Ile Ser Ala Ser

20

<210> 5

<211> 18

<212> PRT

<213> Unknown

<220><223> higher plants secretion signal

<400> 5

Met Ala Asn Lys Ser Leu Leu Leu Leu Leu Leu Leu Gly Ser Leu Ala

1 5 10 15

Ser Gly

<210> 6

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> synthetic concensus sequence

<400> 6

Met Ala Arg Leu Pro Leu Ala Ala Leu Gly

1 5 10

<210> 7

<211> 20

<212> PRT

<213> Unknown

<220><223> synthetic peptide

<400> 7

Met Ala Asn Lys Leu Leu Leu Leu Leu Leu Leu Leu Leu Leu Pro Leu

1 5 10 15

Ala Ala Ser Gly

20

<210> 8

<211> 2615

<212> DNA

<213> Saccharomyces cerevisiae

<400> 8

gaattcccca acatggtgga gcacgacact ctcgtctact ccaagaatat caaagataca 60

gtctcagaag accaaagggc tattgagact tttcaacaaa gggtaatatc gggaaacctc 120

ctcggattcc attgcccagc tatctgtcac ttcatcaaaa ggacagtaga aaaggaaggt 180

ggcacctaca aatgccatca ttgcgataaa ggaaaggcta tcgttcaaga tgcctctgcc 240

gacagtggtc ccaaagatgg acccccaccc acgaggagca tcgtggaaaa agaagacgtt 300

ccaaccacgt cttcaaagca agtggattga tgtgaacatg gtggagcacg acactctcgt 360

ctactccaag aatatcaaag atacagtctc agaagaccaa agggctattg agacttttca 420

acaaagggta atatcgggaa acctcctcgg attccattgc ccagctatct gtcacttcat 480

caaaaggaca gtagaaaagg aaggtggcac ctacaaatgc catcattgcg ataaaggaaa 540

ggctatcgtt caagatgcct ctgccgacag tggtcccaaa gatggacccc cacccacgag 600

gagcatcgtg gaaaaagaag acgttccaac cacgtcttca aagcaagtgg attgatgtga 660

tatctccact gacgtaaggg atgacgcaca atcccactat ccttcgcaag acccttcctc 720

tatataagga agttcatttc atttggagag gacacgctga aatcaccagt ctctctctac 780

aaatctatct ctggcgcgcc atatcaatgc ttcttcaggc ctttcttttt cttcttgctg 840

gttttgctgc caagatcagc gcctctatga cgaacgaaac ctcggataga ccacttgtgc 900

actttacacc aaacaagggc tggatgaatg accccaatgg actgtggtac gacgaaaaag 960

atgccaagtg gcatctgtac tttcaataca acccgaacga tactgtctgg gggacgccat 1020

tgttttgggg ccacgccacg tccgacgacc tgaccaattg ggaggaccaa ccaatagcta 1080
tcgctccgaa gaggaacgac tccggagcat tctcgggttc catggtggtt gactacaaca 1140
atacttccgg ctttttcaac gataccattg acccgagaca acgctgcgtg gccatatgga 1200
cttacaacac accggagtcc gaggagcagt acatctcgta tagcctggac ggtggataca 1260
cttttacaga gtatcagaag aaccctgtgc ttgctgcaaa ttcgactcag ttccgagatc 1320
cgaaggtctt ttggtacgag ccctcgcaga agtggatcat gacagcggca aagtcacagg 1380
actacaagat cgaaatttac tcgtctgacg accttaaatc ctggaagctc gaatccgcgt 1440

tcgcaaacga gggctttctc ggctaccaat acgaatgccc aggcctgata gaggtcccaa 1500
cagagcaaga tcccagcaag tcctactggg tgatgtttat ttccattaat ccaggagcac 1560
cggcaggagg ttcttttaat cagtacttcg tcggaagctt taacggaact catttcgagg 1620
catttgataa ccaatcaaga gtagttgatt ttggaaagga ctactatgcc ctgcagactt 1680
tcttcaatac tgacccgacc tatgggagcg ctcttggcat tgcgtgggct tctaactggg 1740
agtattccgc attcgttcct acaaaccctt ggaggtcctc catgtcgctc gtgaggaaat 1800
tctctctcaa cactgagtac caggccaacc cggaaaccga actcataaac ctgaaagccg 1860

aaccgatcct gaacattagc aacgctggcc cctggagccg gtttgcaacc aacaccacgt 1920
tgacgaaagc caacagctac aacgtcgatc tttcgaatag caccggtaca cttgaatttg 1980
aactggtgta tgccgtcaat accacccaaa cgatctcgaa gtcggtgttc gcggacctct 2040
ccctctggtt taaaggcctg gaagaccccg aggagtacct cagaatgggt ttcgaggttt 2100
ctgcgtcctc cttcttcctt gatcgcggga acagcaaagt aaaatttgtt aaggagaacc 2160
catattttac caacaggatg agcgttaaca accaaccatt caagagcgaa aacgacctgt 2220
cgtactacaa agtgtatggt ttgcttgatc aaaatatcct ggaactctac ttcaacgatg 2280

gtgatgtcgt gtccaccaac acatacttca tgacaaccgg gaacgcactg ggctccgtga 2340
acatgacgac gggtgtggat aacctgttct acatcgacaa attccaggtg agggaagtca 2400
agtgagatct gtcgatcgac aagctcgagt ttctccataa taatgtgtga gtagttccca 2460
gataagggaa ttagggttcc tatagggttt cgctcatgtg ttgagcatat aagaaaccct 2520
tagtatgtat ttgtatttgt aaaatacttc tatcaataaa atttctaatt cctaaaacca 2580
aaatccagta ctaaaatcca gatcccccga attaa 2615

<210> 9

<211> 22

<212> DNA

<213> Prototheca moriformis

<400> 9

tgttgaagaa tgagccggcg ac 22

<210> 10

<211> 20

<212> DNA

<213> Prototheca moriformis

<400> 10

cagtgagcta ttacgcactc 20

<210> 11

<211> 541

<212> DNA

<213> Prototheca kruegani

<400> 11

tgttgaagaa tgagccggcg agttaaaaag agtggcatgg ttaaagaaaa tactctggag 60

ccatagcgaa agcaagttta gtaagcttag gtcattcttt ttagacccga aaccgagtga 120

tctacccatg atcagggtga agtgttagta aaataacatg gaggcccgaa ccgactaatg 180

ttgaaaaatt agcggatgaa ttgtgggtag gggcgaaaaa ccaatcgaac tcggagttag 240

ctggttctcc ccgaaatgcg tttaggcgca gcagtagcag tacaaataga ggggtaaagc 300

actgtttctt ttgtgggctt cgaaagttgt acctcaaagt ggcaaactct gaatactcta 360

tttagatatc tactagtgag accttggggg ataagctcct tggtcaaaag ggaaacagcc 420

cagatcacca gttaaggccc caaaatgaaa atgatagtga ctaaggatgt gggtatgtca 480

aaacctccag caggttagct tagaagcagc aatcctttca agagtgcgta atagctcact 540

g 541

<210> 12

<211> 573

<212> DNA

<213> Prototheca wickerhamii

<400> 12

tgttgaagaa tgagccggcg acttaaaata aatggcaggc taagagattt aataactcga 60

aacctaagcg aaagcaagtc ttaatagggc gtcaatttaa caaaacttta aataaattat 120

aaagtcattt attttagacc cgaacctgag tgatctaacc atggtcagga tgaaacttgg 180

gtgacaccaa gtggaagtcc gaaccgaccg atgttgaaaa atcggcggat gaactgtggt 240

tagtggtgaa ataccagtcg aactcagagc tagctggttc tccccgaaat gcgttgaggc 300

gcagcaatat atctcgtcta tctaggggta aagcactgtt tcggtgcggg ctatgaaaat 360

ggtaccaaat cgtggcaaac tctgaatact agaaatgacg atatattagt gagactatgg 420

gggataagct ccatagtcga gagggaaaca gcccagacca ccagttaagg ccccaaaatg 480

ataatgaagt ggtaaaggag gtgaaaatgc aaatacaacc aggaggttgg cttagaagca 540

gccatccttt aaagagtgcg taatagctca ctg 573

<210> 13

<211> 541

<212> DNA

<213> Prototheca stagnora

<400> 13

tgttgaagaa tgagccggcg agttaaaaaa aatggcatgg ttaaagatat ttctctgaag 60

ccatagcgaa agcaagtttt acaagctata gtcatttttt ttagacccga aaccgagtga 120

tctacccatg atcagggtga agtgttggtc aaataacatg gaggcccgaa ccgactaatg 180

gtgaaaaatt agcggatgaa ttgtgggtag gggcgaaaaa ccaatcgaac tcggagttag 240

ctggttctcc ccgaaatgcg tttaggcgca gcagtagcaa cacaaataga ggggtaaagc 300

actgtttctt ttgtgggctt cgaaagttgt acctcaaagt ggcaaactct gaatactcta 360

tttagatatc tactagtgag accttggggg ataagctcct tggtcaaaag ggaaacagcc 420

cagatcacca gttaaggccc caaaatgaaa atgatagtga ctaaggacgt gagtatgtca 480

aaacctccag caggttagct tagaagcagc aatcctttca agagtgcgta atagctcact 540

g 541

<210> 14

<211> 541

<212> DNA

<213> Prototheca moriformis

<400> 14

tgttgaagaa tgagccggcg agttaaaaag agtggcatgg ttaaagataa ttctctggag 60

ccatagcgaa agcaagttta acaagctaaa gtcacccttt ttagacccga aaccgagtga 120

tctacccatg atcagggtga agtgttggta aaataacatg gaggcccgaa ccgactaatg 180

gtgaaaaatt agcggatgaa ttgtgggtag gggcgaaaaa ccaatcgaac tcggagttag 240

ctggttctcc ccgaaatgcg tttaggcgca gcagtagcaa cacaaataga ggggtaaagc 300

actgtttctt ttgtgggctt cgaaagttgt acctcaaagt ggcaaactct gaatactcta 360
tttagatatc tactagtgag accttggggg ataagctcct tggtcaaaag ggaaacagcc 420
cagatcacca gttaaggccc caaaatgaaa atgatagtga ctaaggatgt gggtatgtta 480
aaacctccag caggttagct tagaagcagc aatcctttca agagtgcgta atagctcact 540
g 541

<210> 15
<211> 573
<212> DNA
<213> Prototheca moriformis
<400> 15

tgttgaagaa tgagccggcg acttaaaata aatggcaggc taagagaatt aataactcga 60

aacctaagcg aaagcaagtc ttaatagggc gctaatttaa caaaacatta aataaaatct 120
aaagtcattt attttagacc cgaacctgag tgatctaacc atggtcagga tgaaacttgg 180
gtgacaccaa gtggaagtcc gaaccgaccg atgttgaaaa atcggcggat gaactgtggt 240
tagtggtgaa ataccagtcg aactcagagc tagctggttc tccccgaaat gcgttgaggc 300
gcagcaatat atctcgtcta tctaggggta aagcactgtt tcggtgcggg ctatgaaaat 360
ggtaccaaat cgtggcaaac tctgaatact agaaatgacg atatattagt gagactatgg 420
gggataagct ccatagtcga gagggaaaca gcccagacca ccagttaagg ccccaaaatg 480

ataatgaagt ggtaaaggag gtgaaaatgc aaatacaacc aggaggttgg cttagaagca 540
gccatccttt aaagagtgcg taatagctca ctg 573

<210> 16
<211> 573
<212> DNA
<213> Prototheca wickerhamii
<400> 16

tgttgaagaa tgagccgtcg acttaaaata aatggcaggc taagagaatt aataactcga 60
aacctaagcg aaagcaagtc ttaatagggc gctaatttaa caaaacatta aataaaatct 120
aaagtcattt attttagacc cgaacctgag tgatctaacc atggtcagga tgaaacttgg 180
gtgacaccaa gtggaagtcc gaaccgaccg atgttgaaaa atcggcggat gaactgtggt 240

tagtggtgaa ataccagtcg aactcagagc tagctggttc tccccgaaat gcgttgaggc 300

gcagcaatat atctcgtcta tctaggggta aagcactgtt tcggtgcggg ctatgaaaat 360

ggtaccaaat cgtggcaaac tctgaatact agaaatgacg atatattagt gagactatgg 420

gggataagct ccatagtcga gagggaaaca gcccagacca ccagttaagg ccccaaaatg 480

ataatgaagt ggtaaaggag gtgaaaatgc aaatacaacc aggaggttgg cttagaagca 540

gccatccttt aaagagtgcg taatagctca ctg 573

<210> 17

<211> 541

<212> DNA

<213> Prototheca moriformis

<400> 17

tgttgaagaa tgagccggcg agttaaaaag agtggcgtgg ttaaagaaaa ttctctggaa 60

ccatagcgaa agcaagttta acaagcttaa gtcacttttt ttagacccga aaccgagtga 120

tctacccatg atcagggtga agtgttggta aaataacatg gaggcccgaa ccgactaatg 180

gtgaaaaatt agcggatgaa ttgtgggtag gggcgaaaaa ccaatcgaac tcggagttag 240

ctggttctcc ccgaaatgcg tttaggcgca gcagtagcaa cacaaataga ggggtaaagc 300

actgtttctt ttgtgggctc cgaaagttgt acctcaaagt ggcaaactct gaatactcta 360

tttagatatc tactagtgag accttggggg ataagctcct tggtcgaaag ggaaacagcc 420

cagatcacca gttaaggccc caaaatgaaa atgatagtga ctaaggatgt gagtatgtca 480

aaacctccag caggttagct tagaagcagc aatcctttca agagtgcgta atagctcact 540

g 541

<210> 18

<211> 541

<212> DNA

<213> Prototheca zopfii

<400> 18

tgttgaagaa tgagccggcg agttaaaaag agtggcatgg ttaaagaaaa ttctctggag 60

ccatagcgaa agcaagttta acaagcttaa gtcacttttt ttagacccga aaccgagtga 120

tctacccatg atcagggtga agtgttggta aaataacatg gaggcccgaa ccgactaatg 180

gtgaaaaatt agcggatgaa ttgtgggtag gggcgaaaaa ccaatcgaac tcggagttag 240

ctggttctcc ccgaaatgcg tttaggcgca gcagtagcaa cacaaataga ggggtaaagc 300

actgtttctt tcgtgggctt cgaaagttgt acctcaaagt ggcaaactct gaatactcta 360

tttagatatc tactagtgag accttggggg ataagctcct tggtcaaaag ggaaacagcc 420

cagatcacca gttaaggccc caaaatgaaa atgatagtga ctaaggatgt gagtatgtca 480

aaacctccag caggttagct tagaagcagc aatcctttca agagtgcgta atagctcact 540

g 541

<210> 19

<211> 565

<212> DNA

<213> Prototheca moriformis

<400> 19

tgttgaagaa tgagccggcg acttagaaaa ggtggcatgg ttaaggaaat attccgaagc 60

cgtagcaaaa gcgagtctga atagggcgat aaaatatatt aatatttaga atctagtcat 120

tttttctaga cccgaacccg ggtgatctaa ccatgaccag gatgaagctt gggtgatacc 180

aagtgaaggt ccgaaccgac cgatgttgaa aaatcggcgg atgagttgtg gttagcggtg 240

aaataccagt cgaacccgga gctagctggt tctccccgaa atgcgttgag gcgcagcagt 300

acatctagtc tatctagggg taaagcactg tttcggtgcg ggctgtgaga acggtaccaa 360

atcgtggcaa actctgaata ctagaaatga cgatgtagta gtgagactgt gggggataag 420

ctccattgtc aagagggaaa cagcccagac caccagctaa ggccccaaaa tggtaatgta 480

gtgacaaagg aggtgaaaat gcaaatacaa ccaggaggtt ggcttagaag cagccatcct 540

ttaaagagtg cgtaatagct cactg 565

<210> 20

<211> 550

<212> PRT

<213> Chicorium intybus

<400> 20

Met Ser Asn Ser Ser Asn Ala Ser Glu Ser Leu Phe Pro Ala Thr Ser

1 5 10 15

Glu Gln Pro Tyr Arg Thr Ala Phe His Phe Gln Pro Pro Gln Asn Trp

20 25 30

Met Asn Asp Pro Asn Gly Pro Met Cys Tyr Asn Gly Val Tyr His Leu

35 40 45

Phe Tyr Gln Tyr Asn Pro Phe Gly Pro Leu Trp Asn Leu Arg Met Tyr

50 55 60

Trp Ala His Ser Val Ser His Asp Leu Ile Asn Trp Ile His Leu Asp

65 70 75 80

Leu Ala Phe Ala Pro Thr Glu Pro Phe Asp Ile Asn Gly Cys Leu Ser
85 90 95

Gly Ser Ala Thr Val Leu Pro Gly Asn Lys Pro Ile Met Leu Tyr Thr
100 105 110

Gly Ile Asp Thr Glu Asn Arg Gln Val Gln Asn Leu Ala Val Pro Lys
115 120 125

Asp Leu Ser Asp Pro Tyr Leu Arg Glu Trp Val Lys His Thr Gly Asn
130 135 140

Pro Ile Ile Ser Leu Pro Glu Glu Ile Gln Pro Asp Asp Phe Arg Asp
145 150 155 160

Pro Thr Thr Thr Trp Leu Glu Glu Asp Gly Thr Trp Arg Leu Leu Val
165 170 175

Gly Ser Gln Lys Asp Lys Thr Gly Ile Ala Phe Leu Tyr His Ser Gly
180 185 190

Asp Phe Val Asn Trp Thr Lys Ser Asp Ser Pro Leu His Lys Val Ser
195 200 205

Gly Thr Gly Met Trp Glu Cys Val Asp Phe Phe Pro Val Trp Val Asp
210 215 220

Ser Thr Asn Gly Val Asp Thr Ser Ile Ile Asn Pro Ser Asn Arg Val
225 230 235 240

Lys His Val Leu Lys Leu Gly Ile Gln Asp His Gly Lys Asp Cys Tyr
245 250 255

Leu Ile Gly Lys Tyr Ser Ala Asp Lys Glu Asn Tyr Val Pro Glu Asp
260 265 270

Glu Leu Thr Leu Ser Thr Leu Arg Leu Asp Tyr Gly Met Tyr Tyr Ala
275 280 285

Ser Lys Ser Phe Phe Asp Pro Val Lys Asn Arg Arg Ile Met Thr Ala
290 295 300

Trp Val Asn Glu Ser Asp Ser Glu Ala Asp Val Ile Ala Arg Gly Trp
305 310 315 320

Ser Gly Val Gln Ser Phe Pro Arg Ser Leu Trp Leu Asp Lys Asn Gln
325 330 335

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

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

Glu Lys Arg Thr Tyr Gly Ala Phe Val Asp Ile Asp Pro Gln Gln Asp
465 470 475 480
Glu Ile Ser Leu Arg Thr Leu Ile Asp His Ser Ile Val Glu Ser Phe
485 490 495
Gly Gly Arg Gly Lys Thr Cys Ile Thr Thr Arg Val Tyr Pro Thr Leu
500 505 510
Ala Ile Gly Glu Gln Ala Arg Leu Phe Ala Phe Asn His Gly Thr Glu
515 520 525

Ser Val Glu Ile Ser Glu Leu Ser Ala Trp Ser Met Lys Lys Ala Gln
530 535 540
Met Lys Val Glu Glu Pro
545 550
<210> 21

<211> 581

<212> PRT

<213> Schizosaccharomyces pombe

<400> 21

Met Phe Leu Lys Tyr Ile Leu Ala Ser Gly Ile Cys Leu Val Ser Leu
1 5 10 15

Leu Ser Ser Thr Asn Ala Ala Pro Arg His Leu Tyr Val Lys Arg Tyr
20 25 30

Pro Val Ile Tyr Asn Ala Ser Asn Ile Thr Glu Val Ser Asn Ser Thr
35 40 45

Thr Val Pro Pro Pro Pro Phe Val Asn Thr Thr Ala Pro Asn Gly Thr
50 55 60

Cys Leu Gly Asn Tyr Asn Glu Tyr Leu Pro Ser Gly Tyr Tyr Asn Ala
65 70 75 80

Thr Asp Arg Pro Lys Ile His Phe Thr Pro Ser Ser Gly Phe Met Asn
85 90 95

Asp Pro Asn Gly Leu Val Tyr Thr Gly Gly Val Tyr His Met Phe Phe
100 105 110

Gln Tyr Ser Pro Lys Thr Leu Thr Ala Gly Glu Val His Trp Gly His
115 120 125

Thr Val Ser Lys Asp Leu Ile His Trp Glu Asn Tyr Pro Ile Ala Ile
130 135 140

Tyr Pro Asp Glu His Glu Asn Gly Val Leu Ser Leu Pro Phe Ser Gly
145 150 155 160

Ser Ala Val Val Asp Val His Asn Ser Ser Gly Leu Phe Ser Asn Asp
165 170 175

Thr Ile Pro Glu Glu Arg Ile Val Leu Ile Tyr Thr Asp His Trp Thr
180 185 190

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

Thr Phe Lys Lys Tyr Ser Gly Asn Pro Val Leu Asp Ile Asn Ser Leu

210 215 220

Gln Phe Arg Asp Pro Lys Val Ile Trp Asp Phe Asp Ala Asn Arg Trp
225 230 235 240
Val Met Ile Val Ala Met Ser Gln Asn Tyr Gly Ile Ala Phe Tyr Ser
245 250 255
Ser Tyr Asp Leu Ile His Trp Thr Glu Leu Ser Val Phe Ser Thr Ser
260 265 270
Gly Tyr Leu Gly Leu Gln Tyr Glu Cys Pro Gly Met Ala Arg Val Pro
275 280 285

Val Glu Gly Thr Asp Glu Tyr Lys Trp Val Leu Phe Ile Ser Ile Asn
290 295 300
Pro Gly Ala Pro Leu Gly Gly Ser Val Val Gln Tyr Phe Val Gly Asp
305 310 315 320
Trp Asn Gly Thr Asn Phe Val Pro Asp Asp Gly Gln Thr Arg Phe Val
325 330 335
Asp Leu Gly Lys Asp Phe Tyr Ala Ser Ala Leu Tyr His Ser Ser Ser
340 345 350

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

Thr Leu Phe Thr Ala Pro Val Ile Asn Ser Ser Ser Ser Leu Ser Gly
420 425 430
Ser Pro Ile Thr Leu Pro Ser Asn Thr Ala Phe Glu Phe Asn Val Thr
435 440 445
Leu Ser Ile Asn Tyr Thr Glu Gly Cys Thr Thr Gly Tyr Cys Leu Gly
450 455 460

Arg Ile Ile Ile Asp Ser Asp Asp Pro Tyr Arg Leu Gln Ser Ile Ser
465 470 475 480

Val Asp Val Asp Phe Ala Ala Ser Thr Leu Val Ile Asn Arg Ala Lys
485 490 495
Ala Gln Met Gly Trp Phe Asn Ser Leu Phe Thr Pro Ser Phe Ala Asn
500 505 510
Asp Ile Tyr Ile Tyr Gly Asn Val Thr Leu Tyr Gly Ile Val Asp Asn
515 520 525
Gly Leu Leu Glu Leu Tyr Val Asn Asn Gly Glu Lys Thr Tyr Thr Asn
530 535 540

Asp Phe Phe Phe Leu Gln Gly Ala Thr Pro Gly Gln Ile Ser Phe Ala
545 550 555 560
Ala Phe Gln Gly Val Ser Phe Asn Asn Val Thr Val Thr Pro Leu Lys
565 570 575
Thr Ile Trp Asn Cys
580

<210> 22
<211> 550
<212> PRT
<213> Pichia anomala
<400> 22
Met Ile Gln Leu Ser Pro Leu Leu Leu Leu Pro Leu Phe Ser Val Phe
1 5 10 15

Asn Ser Ile Ala Asp Ala Ser Thr Glu Tyr Leu Arg Pro Gln Ile His
20 25 30
Leu Thr Pro Asp Gln Gly Trp Met Asn Asp Pro Asn Gly Met Phe Tyr
35 40 45
Asp Arg Lys Asp Lys Leu Trp His Val Tyr Phe Gln His Asn Pro Asp
50 55 60
Lys Lys Ser Ile Trp Ala Thr Pro Val Thr Trp Gly His Ser Thr Ser
65 70 75 80

```
Lys Asp Leu Leu Thr Trp Asp Tyr His Gly Asn Ala Leu Glu Pro Glu
                85                  90                  95
Asn Asp Asp Glu Gly Ile Phe Ser Gly Ser Val Val Val Asp Arg Asn
            100                 105                 110
Asn Thr Ser Gly Phe Phe Asn Asp Ser Thr Asp Pro Glu Gln Arg Ile
        115                 120                 125
Val Ala Ile Tyr Thr Asn Asn Ala Gln Leu Gln Thr Gln Glu Ile Ala
    130                 135                 140

Tyr Ser Leu Asp Lys Gly Tyr Ser Phe Ile Lys Tyr Asp Gln Asn Pro
145                 150                 155                 160
Val Ile Asn Val Asn Ser Ser Gln Gln Arg Asp Pro Lys Val Leu Trp
                165                 170                 175
His Asp Glu Ser Asn Gln Trp Ile Met Val Val Ala Lys Thr Gln Glu
            180                 185                 190
Phe Lys Val Gln Ile Tyr Gly Ser Pro Asp Leu Lys Lys Trp Asp Leu
        195                 200                 205

Lys Ser Asn Phe Thr Ser Asn Gly Tyr Leu Gly Phe Gln Tyr Glu Cys
    210                 215                 220
Pro Gly Leu Phe Lys Leu Pro Ile Glu Asn Pro Leu Asn Asp Thr Val
225                 230                 235                 240
Thr Ser Lys Trp Val Leu Leu Leu Ala Ile Asn Pro Gly Ser Pro Leu
                245                 250                 255
Gly Gly Ser Ile Asn Glu Tyr Phe Ile Gly Asp Phe Asp Gly Thr Thr
            260                 265                 270

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

325 330 335

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

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

Asn Ser Pro Val Phe Thr Glu Arg Ile Ser Thr Tyr Val Glu Lys Ile
465 470 475 480
Asp Thr Thr Asp Gln Gly Asn Val Tyr Thr Leu Tyr Gly Val Val Asp
485 490 495
Arg Asn Ile Leu Glu Leu Tyr Phe Asn Asp Gly Ser Ile Ala Met Thr
500 505 510
Asn Thr Phe Phe Phe Arg Glu Gly Lys Ile Pro Thr Ser Phe Glu Val
515 520 525

Val Cys Asp Ser Glu Lys Ser Phe Ile Thr Ile Asp Glu Leu Ser Val
530 535 540
Arg Glu Leu Ala Arg Lys
545 550

<210> 23

<211> 533

<212> PRT

<213> Debaryomyces occidentalis

<400> 23

Met Val Gln Val Leu Ser Val Leu Val Ile Pro Leu Leu Thr Leu Phe
1 5 10 15

Phe Gly Tyr Val Ala Ser Ser Ser Ile Asp Leu Ser Val Asp Thr Ser
20 25 30

Glu Tyr Asn Arg Pro Leu Ile His Phe Thr Pro Glu Lys Gly Trp Met
35 40 45

Asn Asp Pro Asn Gly Leu Phe Tyr Asp Lys Thr Ala Lys Leu Trp His
50 55 60

Leu Tyr Phe Gln Tyr Asn Pro Asn Ala Thr Ala Trp Gly Gln Pro Leu
65 70 75 80

Tyr Trp Gly His Ala Thr Ser Asn Asp Leu Val His Trp Asp Glu His
85 90 95

Glu Ile Ala Ile Gly Pro Glu His Asp Asn Glu Gly Ile Phe Ser Gly
100 105 110

Ser Ile Val Val Asp His Asn Asn Thr Ser Gly Phe Phe Asn Ser Ser
115 120 125

Ile Asp Pro Asn Gln Arg Ile Val Ala Ile Tyr Thr Asn Asn Ile Pro
130 135 140

Asp Leu Gln Thr Gln Asp Ile Ala Phe Ser Leu Asp Gly Gly Tyr Thr
145 150 155 160

Phe Thr Lys Tyr Glu Asn Asn Pro Val Ile Asp Val Ser Ser Asn Gln
165 170 175

Phe Arg Asp Pro Lys Val Phe Trp His Glu Arg Phe Lys Ser Met Asp
180 185 190

His Gly Cys Ser Glu Ile Ala Arg Val Lys Ile Gln Ile Phe Gly Ser
195 200 205

Ala Asn Leu Lys Asn Trp Val Leu Asn Ser Asn Phe Ser Ser Gly Tyr
210 215 220

Tyr Gly Asn Gln Tyr Gly Met Ser Arg Leu Ile Glu Val Pro Ile Glu
225 230 235 240
Asn Ser Asp Lys Ser Lys Trp Val Met Phe Leu Ala Ile Asn Pro Gly
245 250 255
Ser Pro Leu Gly Gly Ser Ile Asn Gln Tyr Phe Val Gly Asp Phe Asp
260 265 270
Gly Phe Gln Phe Val Pro Asp Asp Ser Gln Thr Arg Phe Val Asp Ile
275 280 285

Gly Lys Asp Phe Tyr Ala Phe Gln Thr Phe Ser Glu Val Glu His Gly
290 295 300
Val Leu Gly Leu Ala Trp Ala Ser Asn Trp Gln Tyr Ala Asp Gln Val
305 310 315 320
Pro Thr Asn Pro Trp Arg Ser Ser Thr Ser Leu Ala Arg Asn Tyr Thr
325 330 335
Leu Arg Tyr Val Ile Gln Met Leu Lys Leu Thr Ala Asn Ile Asp Lys
340 345 350

Ser Val Leu Pro Asp Ser Ile Asn Val Val Asp Lys Leu Lys Lys Lys
355 360 365
Asn Val Lys Leu Thr Asn Lys Lys Pro Ile Lys Thr Asn Phe Lys Gly
370 375 380
Ser Thr Gly Leu Phe Asp Phe Asn Ile Thr Phe Lys Val Leu Asn Leu
385 390 395 400
Asn Val Ser Pro Gly Lys Thr His Phe Asp Ile Leu Ile Asn Ser Gln
405 410 415

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

465 470 475 480

Lys Asn Ile Ile Glu Leu Tyr Phe Asn Asp Gly Thr Val Ala Met Thr
485 490 495
Asn Thr Phe Phe Met Gly Glu Gly Lys Tyr Pro His Asp Ile Gln Ile
500 505 510
Val Thr Asp Thr Glu Glu Pro Leu Phe Glu Leu Glu Ser Val Ile Ile
515 520 525
Arg Glu Leu Asn Lys
530

<210> 24
<211> 654
<212> PRT
<213> Oryza sativa
<400> 24

Met Ala Thr Ser Arg Leu Thr Pro Ala Tyr Asp Leu Lys Asn Ala Ala
1 5 10 15
Ala Ala Val Tyr Thr Pro Leu Pro Glu Gln Pro His Ser Ala Glu Val
20 25 30
Glu Ile Arg Asp Arg Lys Pro Phe Lys Ile Ile Ser Ala Ile Ile Leu
35 40 45
Ser Ser Leu Leu Leu Leu Ala Leu Ile Leu Val Ala Val Asn Tyr Gln
50 55 60

Ala Pro Pro Ser His Ser Ser Gly Asp Asn Ser Gln Pro Ala Ala Val
65 70 75 80
Met Pro Pro Ser Arg Gly Val Ser Gln Gly Val Ser Glu Lys Ala Phe
85 90 95
Arg Gly Ala Ser Gly Ala Gly Asn Gly Val Ser Phe Ala Trp Ser Asn
100 105 110
Leu Met Leu Ser Trp Gln Arg Thr Ser Tyr His Phe Gln Pro Val Lys
115 120 125

Asn Trp Met Asn Asp Pro Asn Gly Pro Leu Tyr Tyr Lys Gly Trp Tyr

130 135 140

His Leu Phe Tyr Gln Tyr Asn Pro Asp Ser Ala Val Trp Gly Asn Ile
145 150 155 160

Thr Trp Gly His Ala Val Ser Thr Asp Leu Ile Asn Trp Leu His Leu
165 170 175

Pro Phe Ala Met Val Pro Asp Gln Trp Tyr Asp Val Asn Gly Val Trp
180 185 190

Thr Gly Ser Ala Thr Ile Leu Pro Asp Gly Arg Ile Val Met Leu Tyr
195 200 205

Thr Gly Asp Thr Asp Asp Tyr Val Gln Asp Gln Asn Leu Ala Phe Pro
210 215 220

Ala Asn Leu Ser Asp Pro Leu Leu Val Asp Trp Val Lys Tyr Pro Asn
225 230 235 240

Asn Pro Val Ile Tyr Pro Pro Pro Gly Ile Gly Val Lys Asp Phe Arg
245 250 255

Asp Pro Thr Thr Ala Gly Thr Ala Gly Met Gln Asn Gly Gln Arg Leu
260 265 270

Val Thr Ile Gly Ser Lys Val Gly Lys Thr Gly Ile Ser Leu Val Tyr
275 280 285

Glu Thr Thr Asn Phe Thr Thr Phe Lys Leu Leu Tyr Gly Val Leu His
290 295 300

Ala Val Pro Gly Thr Gly Met Trp Glu Cys Val Asp Leu Tyr Pro Val
305 310 315 320

Ser Thr Thr Gly Glu Asn Gly Leu Asp Thr Ser Val Asn Gly Leu Gly
325 330 335

Val Lys His Val Leu Lys Thr Ser Leu Asp Asp Asp Lys His Asp Tyr
340 345 350

Tyr Ala Leu Gly Thr Tyr Asp Pro Val Lys Asn Lys Trp Thr Pro Asp
355 360 365

Asn Pro Asp Leu Asp Val Gly Ile Gly Leu Arg Leu Asp Tyr Gly Lys
370 375 380

Tyr Tyr Ala Ala Arg Thr Phe Tyr Asp Gln Asn Lys Gln Arg Arg Ile
385 390 395 400
Leu Trp Gly Trp Ile Gly Glu Thr Asp Leu Glu Ala Val Asp Leu Met
405 410 415
Lys Gly Trp Ala Ser Leu Gln Ala Ile Pro Arg Thr Ile Val Phe Asp
420 425 430
Lys Lys Thr Gly Thr Asn Val Leu Gln Arg Pro Glu Glu Glu Val Glu
435 440 445

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

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

Asp Gly Glu Asn Tyr Ser Ala Arg Ile Leu Ile Asp His Ser Ile Val
580 585 590
Glu Ser Phe Ala Gln Ala Gly Arg Thr Cys Val Arg Ser Arg Asp Tyr
595 600 605
Pro Thr Lys Asp Ile Tyr Gly Ala Ala Arg Cys Phe Phe Phe Asn Asn
610 615 620
Ala Thr Glu Ala Ser Val Arg Ala Ser Leu Lys Ala Trp Gln Met Lys

625 630 635 640

Ser Phe Ile Arg Pro Tyr Pro Phe Ile Pro Asp Gln Lys Ser
645 650

<210> 25

<211> 690

<212> PRT

<213> Allium cepa

<400> 25

Met Ser Ser Asp Asp Leu Glu Ser Pro Pro Ser Ser Tyr Leu Pro Ile
1 5 10 15

Pro Pro Ser Asp Glu Phe His Asp Gln Pro Pro Pro Leu Arg Ser Trp
20 25 30

Leu Arg Leu Leu Ser Ile Pro Leu Ala Leu Met Phe Leu Leu Phe Leu

35 40 45

Ala Thr Phe Leu Ser Asn Leu Glu Ser Pro Pro Ser Asp Ser Gly Leu
50 55 60

Val Ser Asp Pro Val Thr Phe Asp Val Asn Pro Ala Val Val Arg Arg
65 70 75 80

Gly Lys Asp Ala Gly Val Ser Asp Lys Thr Ser Gly Val Asp Ser Gly
85 90 95

Phe Val Leu Asp Pro Val Ala Val Asp Ala Asn Ser Val Val Val His

100 105 110

Arg Gly Lys Asp Ala Gly Val Ser Asp Lys Thr Ser Gly Val Asp Ser
115 120 125

Gly Leu Leu Lys Asp Ser Pro Leu Gly Pro Tyr Pro Trp Thr Asn Gln
130 135 140

Met Leu Ser Trp Gln Arg Thr Gly Phe His Phe Gln Pro Val Lys Asn
145 150 155 160

Trp Met Asn Asp Pro Asn Gly Pro Leu Tyr Tyr Lys Gly Trp Tyr His

165 170 175

Phe Phe Tyr Gln Tyr Asn Pro Glu Gly Ala Val Trp Gly Asn Ile Ala

180 185 190

Trp Gly His Ala Val Ser Arg Asp Leu Val His Trp Thr His Leu Pro

195 200 205

Leu Ala Met Val Pro Asp Gln Trp Tyr Asp Ile Asn Gly Val Trp Thr

210 215 220

Gly Ser Ala Thr Ile Leu Pro Asp Gly Gln Ile Val Met Leu Tyr Thr

225 230 235 240

Gly Ala Thr Asn Glu Ser Val Gln Val Gln Asn Leu Ala Val Pro Ala

245 250 255

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

260 265 270

Pro Ile Leu Val Pro Pro Pro Gly Ile Gly Asp Lys Asp Phe Arg Asp

275 280 285

Pro Thr Thr Ala Trp Tyr Glu Pro Ser Asp Asp Thr Trp Arg Ile Val

290 295 300

Ile Gly Ser Lys Asp Ser Ser His Ser Gly Ile Ala Ile Val Tyr Ser

305 310 315 320

Thr Lys Asp Phe Ile Asn Tyr Lys Leu Ile Pro Gly Ile Leu His Ala

325 330 335

Val Glu Arg Val Gly Met Trp Glu Cys Val Asp Phe Tyr Pro Val Ala

340 345 350

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

355 360 365

Pro Ser Pro Ala Val Lys His Val Leu Lys Ala Ser Met Asp Asp Asp

370 375 380

Arg His Asp Tyr Tyr Ala Ile Gly Thr Tyr Asp Pro Ala Gln Asn Thr

385 390 395 400

Trp Val Pro Asp Asp Ala Ser Val Asp Val Gly Ile Gly Leu Arg Tyr

405 410 415

Asp Trp Gly Lys Phe Tyr Ala Ser Lys Thr Phe Tyr Asp His Ala Lys

420 425 430

Lys Arg Arg Ile Leu Trp Ser Trp Ile Gly Glu Thr Asp Ser Glu Thr
435 440 445
Ala Asp Ile Ala Lys Gly Trp Ala Ser Leu Gln Gly Val Pro Arg Thr
450 455 460
Val Leu Leu Asp Val Lys Thr Gly Ser Asn Leu Ile Thr Trp Pro Val
465 470 475 480
Val Glu Ile Glu Ser Leu Arg Thr Arg Pro Arg Asp Phe Ser Gly Ile

485 490 495
Thr Val Asp Ala Gly Ser Thr Phe Lys Leu Asp Val Gly Gly Ala Ala
500 505 510
Gln Leu Asp Ile Glu Ala Glu Phe Lys Ile Ser Ser Glu Glu Leu Glu
515 520 525
Ala Val Lys Glu Ala Asp Val Ser Tyr Asn Cys Ser Ser Ser Gly Gly
530 535 540
Ala Ala Glu Arg Gly Val Leu Gly Pro Phe Gly Leu Leu Val Leu Ala

545 550 555 560
Asn Gln Asp Leu Thr Glu Gln Thr Ala Thr Tyr Phe Tyr Val Ser Arg
565 570 575
Gly Met Asp Gly Gly Leu Asn Thr His Phe Cys Gln Asp Glu Lys Arg
580 585 590
Ser Ser Lys Ala Ser Asp Ile Val Lys Arg Ile Val Gly His Ser Val
595 600 605
Pro Val Leu Asp Gly Glu Ser Phe Ala Leu Arg Ile Leu Val Asp His

610 615 620
Ser Ile Val Glu Ser Phe Ala Gln Gly Gly Arg Ala Ser Ala Thr Ser
625 630 635 640
Arg Val Tyr Pro Thr Glu Ala Ile Tyr Asn Asn Ala Arg Val Phe Val
645 650 655
Phe Asn Asn Ala Thr Gly Ala Lys Val Thr Ala Gln Ser Leu Lys Val
660 665 670
Trp His Met Ser Thr Ala Ile Asn Glu Ile Tyr Asp Pro Ala Thr Ser

675 680 685

Val Met

690

<210> 26

<211> 501

<212> PRT

<213> Beta vulgaris

<220><221> misc_feature

<222> (382)..(382)

<223> Xaa can be any naturally occurring amino acid

<400> 26

Leu Phe Tyr Gln Tyr Asn Pro Asn Gly Val Ile Trp Gly Pro Pro Val
1 5 10 15

Trp Gly His Ser Thr Ser Lys Asp Leu Val Asn Trp Val Pro Gln Pro
20 25 30

Leu Thr Met Glu Pro Glu Met Ala Ala Asn Ile Asn Gly Ser Trp Ser
35 40 45

Gly Ser Ala Thr Ile Leu Pro Gly Asn Lys Pro Ala Ile Leu Phe Thr
50 55 60

Gly Leu Asp Pro Lys Tyr Glu Gln Val Gln Val Leu Ala Tyr Pro Lys
65 70 75 80

Asp Thr Ser Asp Pro Asn Leu Lys Glu Trp Phe Leu Ala Pro Gln Asn
85 90 95

Pro Val Met Phe Pro Thr Pro Gln Asn Gln Ile Asn Ala Thr Ser Phe
100 105 110

Arg Asp Pro Thr Thr Ala Trp Arg Leu Pro Asp Gly Val Trp Arg Leu
115 120 125

Leu Ile Gly Ser Lys Arg Gly Gln Arg Gly Leu Ser Leu Leu Phe Arg
130 135 140

Ser Arg Asp Phe Val His Trp Val Gln Ala Lys His Pro Leu Tyr Ser
145 150 155 160

Asp Lys Leu Ser Gly Met Trp Glu Cys Pro Asp Phe Phe Pro Val Tyr

165 170 175

Ala Asn Gly Asp Gln Met Gly Val Asp Thr Ser Ile Ile Gly Ser His

180 185 190

Val Lys His Val Leu Lys Asn Ser Leu Asp Ile Thr Lys His Asp Ile

195 200 205

Tyr Thr Ile Gly Asp Tyr Asn Ile Lys Lys Asp Ala Tyr Thr Pro Asp

210 215 220

Ile Gly Tyr Met Asn Asp Ser Ser Leu Arg Tyr Asp Tyr Gly Lys Tyr

225 230 235 240

Tyr Ala Ser Lys Thr Phe Phe Asp Asp Ala Lys Lys Glu Arg Ile Leu

245 250 255

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

260 265 270

Gly Trp Ser Gly Ile His Thr Ile Pro Arg Lys Ile Trp Leu Asp Lys

275 280 285

Leu Gly Lys Gln Leu Ile Gln Trp Pro Ile Ala Asn Ile Glu Lys Leu

290 295 300

Arg Gln Lys Pro Val Asn Ile Tyr Arg Lys Val Leu Lys Gly Gly Ser

305 310 315 320

Gln Ile Glu Val Ser Gly Ile Thr Ala Ala Gln Ala Asp Val Glu Ile

325 330 335

Ser Phe Lys Ile Lys Asp Leu Lys Asn Val Glu Lys Phe Asp Ala Ser

340 345 350

Trp Thr Ser Pro Gln Leu Leu Cys Ser Lys Lys Gly Ala Ser Val Lys

355 360 365

Gly Gly Leu Gly Pro Phe Gly Leu Leu Thr Leu Ala Ser Xaa Gly Leu

370 375 380

Glu Glu Tyr Thr Ala Val Phe Phe Arg Ile Phe Lys Ala Tyr Asp Asn

385 390 395 400

Lys Phe Val Val Leu Met Cys Ser Asp Gln Ser Arg Ser Ser Leu Asn

405 410 415

Pro Thr Asn Asp Lys Thr Thr Tyr Gly Thr Phe Val Asp Val Asn Pro
420 425 430

Ile Arg Glu Gly Leu Ser Leu Arg Val Leu Ile Asp His Ser Val Val
435 440 445

Glu Ser Phe Gly Ala Lys Gly Lys Asn Val Ile Thr Ala Arg Val Tyr
450 455 460

Pro Thr Leu Ala Ile Asn Glu Lys Ala His Leu Tyr Val Phe Asn Arg
465 470 475 480

Gly Thr Ser Asn Val Glu Ile Thr Gly Leu Thr Ala Trp Ser Met Lys
485 490 495

Lys Ala Asn Ile Ala
500

<210> 27
<211> 518
<212> PRT
<213> Bifidobacterium breve
<400> 27

Met Thr Asp Phe Thr Pro Glu Thr Pro Val Leu Thr Pro Ile Arg Asp
1 5 10 15

His Ala Ala Glu Leu Ala Lys Ala Glu Ala Gly Val Ala Glu Met Ala
20 25 30

Ala Lys Arg Asn Asn Arg Trp Tyr Pro Lys Tyr His Ile Ala Ser Asn
35 40 45

Gly Gly Trp Ile Asn Asp Pro Asn Gly Leu Cys Phe Tyr Lys Gly Arg
50 55 60

Trp His Val Phe Tyr Gln Leu His Pro Tyr Gly Thr Gln Trp Gly Pro
65 70 75 80

Met His Trp Gly His Val Ser Ser Thr Asp Met Leu Asn Trp Lys Arg
85 90 95

Glu Pro Ile Met Phe Ala Pro Ser Leu Glu Gln Glu Lys Asp Gly Val
100 105 110

Phe Ser Gly Ser Ala Val Ile Asp Asp Asn Gly Asp Leu Arg Phe Tyr

115 120 125

Tyr Thr Gly His Arg Trp Ala Asn Gly His Asp Asn Thr Gly Gly Asp

130 135 140

Trp Gln Val Gln Met Thr Ala Leu Pro Asp Asn Asp Glu Leu Thr Ser

145 150 155 160

Ala Thr Lys Gln Gly Met Ile Ile Asp Cys Pro Thr Asp Lys Val Asp

165 170 175

His His Tyr Arg Asp Pro Lys Val Trp Lys Thr Gly Asp Thr Trp Tyr

180 185 190

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

195 200 205

Phe Ser Ser Lys Asp Met Val Arg Trp Glu Tyr Glu Arg Val Leu Phe

210 215 220

Gln His Pro Asp Pro Asp Val Phe Met Leu Glu Cys Pro Asp Phe Phe

225 230 235 240

Pro Ile Lys Asp Lys Asp Gly Asn Glu Lys Trp Val Ile Gly Phe Ser

245 250 255

Ala Met Gly Ser Lys Pro Ser Gly Phe Met Asn Arg Asn Val Asn Asn

260 265 270

Ala Gly Tyr Met Ile Gly Thr Trp Glu Pro Gly Gly Glu Phe Lys Pro

275 280 285

Glu Thr Glu Phe Arg Leu Trp Asp Cys Gly His Asn Tyr Tyr Ala Pro

290 295 300

Gln Ser Phe Asn Val Asp Gly Arg Gln Ile Val Tyr Gly Trp Met Ser

305 310 315 320

Pro Phe Val Gln Pro Ile Pro Met Glu Asp Asp Gly Trp Cys Gly Gln

325 330 335

Leu Thr Leu Pro Arg Glu Ile Thr Leu Asp Asp Asp Gly Asp Val Val

340 345 350

Thr Ala Pro Val Ala Glu Met Glu Gly Leu Arg Glu Asp Thr Leu Asp

355 360 365

His Gly Ser Ile Thr Leu Asp Met Asp Gly Glu Gln Val Ile Ala Asp
370 375 380
Asp Ala Glu Ala Val Glu Ile Glu Met Thr Ile Asp Leu Ala Ala Ser
385 390 395 400
Thr Ala Asp Arg Ala Gly Leu Lys Ile His Ala Thr Glu Asp Gly Ala
405 410 415

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

Ser Ser Tyr Ser Tyr Ala Ser Glu Gly Pro Arg Ala Ile Lys Leu Val
485 490 495
Ala Glu Phe Gly Asn Leu Lys Val Glu Ser Leu Lys Leu His His Met
500 505 510
Lys Ser Ile Gly Leu Glu
515

<210> 28
<211> 532
<212> PRT
<213> Saccharomyces cerevisae
<400> 28
Met Leu Leu Gln Ala Phe Leu Phe Leu Leu Ala Gly Phe Ala Ala Lys
1 5 10 15

Ile Ser Ala Ser Met Thr Asn Glu Thr Ser Asp Arg Pro Leu Val His
20 25 30
Phe Thr Pro Asn Lys Gly Trp Met Asn Asp Pro Asn Gly Leu Trp Tyr
35 40 45
Asp Glu Lys Asp Ala Lys Trp His Leu Tyr Phe Gln Tyr Asn Pro Asn

```
    50                  55                  60
Asp Thr Val Trp Gly Thr Pro Leu Phe Trp Gly His Ala Thr Ser Asp
65                  70                  75                  80

Asp Leu Thr Asn Trp Glu Asp Gln Pro Ile Ala Ile Ala Pro Lys Arg
                85                  90                  95
Asn Asp Ser Gly Ala Phe Ser Gly Ser Met Val Val Asp Tyr Asn Asn
            100                 105                 110
Thr Ser Gly Phe Phe Asn Asp Thr Ile Asp Pro Arg Gln Arg Cys Val
        115                 120                 125
Ala Ile Trp Thr Tyr Asn Thr Pro Glu Ser Glu Glu Gln Tyr Ile Ser
    130                 135                 140

Tyr Ser Leu Asp Gly Gly Tyr Thr Phe Thr Glu Tyr Gln Lys Asn Pro
145                 150                 155                 160
Val Leu Ala Ala Asn Ser Thr Gln Phe Arg Asp Pro Lys Val Phe Trp
                165                 170                 175
Tyr Glu Pro Ser Gln Lys Trp Ile Met Thr Ala Ala Lys Ser Gln Asp
            180                 185                 190
Tyr Lys Ile Glu Ile Tyr Ser Ser Asp Asp Leu Lys Ser Trp Lys Leu
        195                 200                 205

Glu Ser Ala Phe Ala Asn Glu Gly Phe Leu Gly Tyr Gln Tyr Glu Cys
    210                 215                 220
Pro Gly Leu Ile Glu Val Pro Thr Glu Gln Asp Pro Ser Lys Ser Tyr
225                 230                 235                 240
Trp Val Met Phe Ile Ser Ile Asn Pro Gly Ala Pro Ala Gly Gly Ser
                245                 250                 255
Phe Asn Gln Tyr Phe Val Gly Ser Phe Asn Gly Thr His Phe Glu Ala
            260                 265                 270

Phe Asp Asn Gln Ser Arg Val Val Asp Phe Gly Lys Asp Tyr Tyr Ala
        275                 280                 285
Leu Gln Thr Phe Phe Asn Thr Asp Pro Thr Tyr Gly Ser Ala Leu Gly
    290                 295                 300
```

Ile Ala Trp Ala Ser Asn Trp Glu Tyr Ser Ala Phe Val Pro Thr Asn
305 310 315 320
Pro Trp Arg Ser Ser Met Ser Leu Val Arg Lys Phe Ser Leu Asn Thr
325 330 335

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

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

Phe Lys Ser Glu Asn Asp Leu Ser Tyr Tyr Lys Val Tyr Gly Leu Leu
465 470 475 480
Asp Gln Asn Ile Leu Glu Leu Tyr Phe Asn Asp Gly Asp Val Val Ser
485 490 495
Thr Asn Thr Tyr Phe Met Thr Thr Gly Asn Ala Leu Gly Ser Val Asn
500 505 510
Met Thr Thr Gly Val Asp Asn Leu Phe Tyr Ile Asp Lys Phe Gln Val
515 520 525

Arg Glu Val Lys
530

<210> 29

<211> 512

<212> PRT

<213> Zymomonas mobilis

<400> 29

Met Glu Ser Pro Ser Tyr Lys Asn Leu Ile Lys Ala Glu Asp Ala Gln
1 5 10 15

Lys Lys Ala Gly Lys Arg Leu Leu Ser Ser Glu Trp Tyr Pro Gly Phe
20 25 30

His Val Thr Pro Leu Thr Gly Trp Met Asn Asp Pro Asn Gly Leu Ile
35 40 45

Phe Phe Lys Gly Glu Tyr His Leu Phe Tyr Gln Tyr Tyr Pro Phe Ala
50 55 60

Pro Val Trp Gly Pro Met His Trp Gly His Ala Lys Ser Arg Asp Leu
65 70 75 80

Val His Trp Glu Thr Leu Pro Val Ala Leu Ala Pro Gly Asp Leu Phe
85 90 95

Asp Arg Asp Gly Cys Phe Ser Gly Cys Ala Val Asp Asn Asn Gly Val
100 105 110

Leu Thr Leu Ile Tyr Thr Gly His Ile Val Leu Ser Asn Asp Ser Pro
115 120 125

Asp Ala Ile Arg Glu Val Gln Cys Met Ala Thr Ser Ile Asp Gly Ile
130 135 140

His Phe Gln Lys Glu Gly Ile Val Leu Glu Lys Ala Pro Met Pro Gln
145 150 155 160

Val Ala His Phe Arg Asp Pro Arg Val Trp Lys Glu Asn Asp His Trp
165 170 175

Phe Met Val Val Gly Tyr Arg Thr Asp Asp Glu Lys His Gln Gly Ile
180 185 190

Gly His Val Ala Leu Tyr Arg Ser Glu Asn Leu Lys Asp Trp Ile Phe
195 200 205

Val Lys Thr Leu Leu Gly Asp Asn Ser Gln Leu Pro Leu Gly Lys Arg
210 215 220

Ala Phe Met Trp Glu Cys Pro Asp Phe Phe Ser Leu Gly Asn Arg Ser
225 230 235 240

Val Leu Met Phe Ser Pro Gln Gly Leu Lys Ala Ser Gly Tyr Lys Asn
245 250 255

Arg Asn Leu Phe Gln Asn Gly Tyr Ile Leu Gly Lys Trp Gln Ala Pro
260 265 270

Gln Phe Thr Pro Glu Thr Ser Phe Gln Glu Leu Asp Tyr Gly His Asp
275 280 285

Phe Tyr Ala Ala Gln Arg Phe Glu Ala Lys Asp Gly Arg Gln Ile Leu
290 295 300

Ile Ala Trp Phe Asp Met Trp Glu Asn Gln Lys Pro Ser Gln Arg Asp
305 310 315 320

Gly Trp Ala Gly Cys Met Thr Leu Pro Arg Lys Leu Asp Leu Ile Asp
325 330 335

Asn Lys Ile Val Met Thr Pro Val Arg Glu Met Glu Ile Leu Arg Gln
340 345 350

Ser Glu Lys Ile Glu Ser Val Val Thr Leu Ser Asp Ala Glu His Pro
355 360 365

Phe Thr Met Asp Ser Pro Leu Gln Glu Ile Glu Leu Ile Phe Asp Leu
370 375 380

Glu Lys Ser Ser Ala Tyr Gln Ala Gly Leu Ala Leu Arg Cys Asn Gly
385 390 395 400

Lys Gly Gln Glu Thr Leu Leu Tyr Ile Asp Arg Ser Gln Asn Arg Ile
405 410 415

Ile Leu Asp Arg Asn Arg Ser Gly Gln Asn Val Lys Gly Ile Arg Ser
420 425 430

Cys Pro Leu Pro Asn Thr Ser Lys Val Arg Leu His Ile Phe Leu Asp
435 440 445

Arg Ser Ser Ile Glu Ile Phe Val Gly Asp Asp Gln Thr Gln Gly Leu
450 455 460

Tyr Ser Ile Ser Ser Arg Ile Phe Pro Asp Lys Asp Ser Leu Lys Gly

465 470 475 480

Arg Leu Phe Ala Ile Glu Gly Tyr Ala Val Phe Asp Ser Phe Lys Arg

485 490 495

Trp Thr Leu Gln Asp Ala Asn Leu Ala Ala Phe Ser Ser Asp Ala Cys

500 505 510

<210> 30

<211> 360

<212> PRT

<213> Artificial Sequence

<220><223> synthetic construct

<400> 30

Met Ala Ile Lys Thr Asn Arg Gln Pro Val Glu Lys Pro Pro Phe Thr

1 5 10 15

Ile Gly Thr Leu Arg Lys Ala Ile Pro Ala His Cys Phe Glu Arg Ser

20 25 30

Ala Leu Arg Gly Arg Ala Pro Asp Trp Ser Met Leu Phe Ala Val Ile

35 40 45

Thr Thr Ile Phe Ser Ala Ala Glu Lys Gln Trp Thr Asn Leu Glu Trp

50 55 60

Lys Pro Lys Pro Asn Pro Pro Gln Leu Leu Asp Asp His Phe Gly Pro

65 70 75 80

His Gly Leu Val Phe Arg Arg Thr Phe Ala Ile Arg Ser Tyr Glu Val

85 90 95

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

100 105 110

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

115 120 125

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

130 135 140

Lys Arg Thr His Val Ala Val Glu Arg Tyr Pro Ala Trp Gly Asp Thr

145 150 155 160

Val Glu Val Glu Cys Trp Val Gly Ala Ser Gly Asn Asn Gly Arg Arg

165 170 175

His Asp Phe Leu Val Arg Asp Cys Lys Thr Gly Glu Ile Leu Thr Arg

180 185 190

Cys Thr Ser Leu Ser Val Met Met Asn Thr Arg Thr Arg Arg Leu Ser

195 200 205

Lys Ile Pro Glu Glu Val Arg Gly Glu Ile Gly Pro Ala Phe Ile Asp

210 215 220

Asn Val Ala Val Lys Asp Glu Glu Ile Lys Lys Pro Gln Lys Leu Asn

225 230 235 240

Asp Ser Thr Ala Asp Tyr Ile Gln Gly Gly Leu Thr Pro Arg Trp Asn

245 250 255

Asp Leu Asp Ile Asn Gln His Val Asn Asn Ile Lys Tyr Val Asp Trp

260 265 270

Ile Leu Glu Thr Val Pro Asp Ser Ile Phe Glu Ser His His Ile Ser

275 280 285

Ser Phe Thr Ile Glu Tyr Arg Arg Glu Cys Thr Met Asp Ser Val Leu

290 295 300

Gln Ser Leu Thr Thr Val Ser Gly Gly Ser Ser Glu Ala Gly Leu Val

305 310 315 320

Cys Glu His Leu Leu Gln Leu Glu Gly Gly Ser Glu Val Leu Arg Ala

325 330 335

Lys Thr Glu Trp Arg Pro Lys Leu Thr Asp Ser Phe Arg Gly Ile Ser

340 345 350

Val Ile Pro Ala Glu Ser Ser Val

355 360

<210> 31

<211> 4653

<212> DNA

<213> Artificial Sequence

<220><223> synthetic construct

<400> 31

ggtacccgcc tgcaacgcaa gggcagccac agccgctccc acccgccgct gaaccgacac 60

gtgcttgggc gcctgccgcc tgcctgccgc atgcttgtgc tggtgaggct gggcagtgct 120

gccatgctga ttgaggcttg gttcatcggg tggaagctta tgtgtgtgct gggcttgcat 180

gccgggcaat gcgcatggtg gcaagagggc ggcagcactt gctggagctg ccgcggtgcc 240

tccaggtggt tcaatcgcgg cagccagagg gatttcagat gatcgcgcgt acaggttgag 300

cagcagtgtc agcaaaggta gcagtttgcc agaatgatcg gttcagctgt taatcaatgc 360

cagcaagaga aggggtcaag tgcaaacacg ggcatgccac agcacgggca ccggggagtg 420

gaatggcacc accaagtgtg tgcgagccag catcgccgcc tggctgtttc agctacaacg 480

gcaggagtca tccaacgtaa ccatgagctg atcaacactg caatcatcgg gcgggcgtga 540

tgcaagcatg cctggcgaag acacatggtg tgcggatgct gccggctgct gcctgctgcg 600

cacgccgttg agttggcagc aggctcagcc atgcactgga tggcagctgg gctgccactg 660

caatgtggtg gataggatgc aagtggagcg aataccaaac cctctggctg cttgctgggt 720

tgcatggcat cgcaccatca gcaggagcgc atgcgaaggg actggcccca tgcacgccat 780

gccaaaccgg agcgcaccga gtgtccacac tgtcaccagg cccgcaagct ttgcagaacc 840

atgctcatgg acgcatgtag cgctgacgtc ccttgacggc gctcctctcg ggtgtgggaa 900

acgcaatgca gcacaggcag cagaggcggc ggcagcagag cggcggcagc agcggcgggg 960

gccacccttc ttgcggggtc gcgccccagc cagcggtgat gcgctgatcc caaacgagtt 1020

cacattcatt tgcatgcctg gagaagcgag gctggggcct ttgggctggt gcagcccgca 1080

atggaatgcg ggaccgccag gctagcagca aaggcgcctc ccctactccg catcgatgtt 1140

ccatagtgca ttggactgca tttgggtggg gcggccggct gtttctttcg tgttgcaaaa 1200

cgcgccagct cagcaacctg tcccgtgggt cccccgtgcc gatgaaatcg tgtgcacgcc 1260

gatcagctga ttgcccggct cgcgaagtag gcgccctcct ttctgctcgc cctctctccg 1320

tcccgcctct agaatatcaa tgatcgagca ggacggcctc cacgccggct cccccgccgc 1380

ctgggtggag cgcctgttcg gctacgactg ggcccagcag accatcggct gctccgacgc 1440

cgccgtgttc cgcctgtccg cccagggccg ccccgtgctg ttcgtgaaga ccgacctgtc 1500

cggcgccctg aacgagctgc aggacgaggc cgcccgcctg tcctggctgg ccaccaccgg 1560

cgtgccctgc gccgccgtgc tggacgtggt gaccgaggcc ggccgcgact ggctgctgct 1620

gggcgaggtg cccggccagg acctgctgtc ctcccacctg gcccccgccg agaaggtgtc 1680

catcatggcc gacgccatgc gccgcctgca caccctggac cccgccacct gccccttcga 1740

ccaccaggcc aagcaccgca tcgagcgcgc ccgcacccgc atggaggccg gcctggtgga 1800

ccaggacgac ctggacgagg agcaccaggg cctggccccc gccgagctgt tcgcccgcct 1860

gaaggcccgc atgcccgacg gcgaggacct ggtggtgacc cacggcgacg cctgcctgcc 1920

caacatcatg gtggagaacg gccgcttctc cggcttcatc gactgcggcc gcctgggcgt 1980

ggccgaccgc taccaggaca tcgccctggc cacccgcgac atcgccgagg agctgggcgg 2040

cgagtgggcc gaccgcttcc tggtgctgta cggcatcgcc gcccccgact cccagcgcat 2100

cgccttctac cgcctgctgg acgagttctt ctgacaattg gcagcagcag ctcggatagt 2160

atcgacacac tctggacgct ggtcgtgtga tggactgttg ccgccacact tgctgccttg 2220

acctgtgaat atccctgccg cttttatcaa acagcctcag tgtgtttgat cttgtgtgta 2280

cgcgcttttg cgagttgcta gctgcttgtg ctatttgcga ataccacccc cagcatcccc 2340

ttccctcgtt tcatatcgct tgcatcccaa ccgcaactta tctacgctgt cctgctatcc 2400

ctcagcgctg ctcctgctcc tgctcactgc ccctcgcaca gccttggttt gggctccgcc 2460

tgtattctcc tggtactgca acctgtaaac cagcactgca atgctgatgc acgggaagta 2520

gtgggatggg aacacaaatg gaggatcccg cgtctcgaac agagcgcgca gaggaacgct 2580

gaaggtctcg cctctgtcgc acctcagcgc ggcatacacc acaataacca cctgacgaat 2640

gcgcttggtt cttcgtccat tagcgaagcg tccggttcac acacgtgcca cgttggcgag 2700

gtggcaggtg acaatgatcg gtggagctga tggtcgaaac gttcacagcc tagggatatc 2760

gaattccttt cttgcgctat gacacttcca gcaaaaggta gggcgggctg cgagacggct 2820

tcccggcgct gcatgcaaca ccgatgatgc ttcgaccccc cgaagctcct tcggggctgc 2880

atgggcgctc cgatgccgct ccagggcgag cgctgtttaa atagccaggc ccccgattgc 2940

aaagacatta tagcgagcta ccaaagccat attcaaacac ctagatcact accacttcta 3000

cacaggccac tcgagcttgt gatcgcactc cgctaagggg gcgcctcttc ctcttcgttt 3060

cagtcacaac ccgcaaacac tagtatggct atcaagacga acaggcagcc tgtggagaag 3120

cctccgttca cgatcgggac gctgcgcaag gccatccccg cgcactgttt cgagcgctcg 3180

gcgcttcgtg ggcgcgcccc cgactggtcc atgctgttcg ccgtgatcac caccatcttc 3240

tccgccgccg agaagcagtg gaccaacctg gagtggaagc ccaagcccaa ccccccccag 3300

ctgctggacg accacttcgg cccccacggc ctggtgttcc gccgcacctt cgccatccgc 3360

agctacgagg tgggccccga ccgctccacc agcatcgtgg ccgtgatgaa ccacctgcag 3420

gaggccgccc tgaaccacgc caagtccgtg ggcatcctgg gcgacggctt cggcaccacc 3480

ctggagatgt ccaagcgcga cctgatctgg gtggtgaagc gcacccacgt ggccgtggag 3540

cgctaccccg cctggggcga caccgtggag gtggagtgct gggtgggcgc ctccggcaac 3600

aacggccgcc gccacgactt cctggtgcgc gactgcaaga ccggcgagat cctgacccgc 3660

tgcacctccc tgagcgtgat gatgaacacc cgcacccgcc gcctgagcaa gatccccgag 3720

gaggtgcgcg gcgagatcgg ccccgccttc atcgacaacg tggccgtgaa ggacgaggag 3780
atcaagaagc cccagaagct gaacgactcc accgccgact acatccaggg cggcctgacc 3840
ccccgctgga acgacctgga catcaaccag cacgtgaaca acatcaagta cgtggactgg 3900
atcctggaga ccgtgcccga cagcatcttc gagagccacc acatctcctc cttcaccatc 3960
gagtaccgcc gcgagtgcac catggacagc gtgctgcagt ccctgaccac cgtgagcggc 4020
ggctcctccg aggccggcct ggtgtgcgag cacctgctgc agctggaggg cggcagcgag 4080
gtgctgcgcg ccaagaccga gtggcgcccc aagctgaccg actccttccg cggcatcagc 4140

gtgatccccg ccgagtccag cgtgatggac tacaaggacc acgacggcga ctacaaggac 4200
cacgacatcg actacaagga cgacgacgac aagtgatgac tcgaggcagc agcagctcgg 4260
atagtatcga cacactctgg acgctggtcg tgtgatggac tgttgccgcc acacttgctg 4320
ccttgacctg tgaatatccc tgccgctttt atcaaacagc ctcagtgtgt ttgatcttgt 4380
gtgtacgcgc ttttgcgagt tgctagctgc ttgtgctatt tgcgaatacc acccccagca 4440
tccccttccc tcgtttcata tcgcttgcat cccaaccgca acttatctac gctgtcctgc 4500
tatccctcag cgctgctcct gctcctgctc actgcccctc gcacagcctt ggtttgggct 4560

ccgcctgtat tctcctggta ctgcaacctg taaaccagca ctgcaatgct gatgcacggg 4620
aagtagtggg atgggaacac aaatggaaag ctt 4653

<210> 32
<211> 366
<212> PRT
<213> Artificial Sequence
<220><223> synthetic construct
<400> 32

Met Ala Ile Lys Thr Asn Arg Gln Pro Val Glu Lys Pro Pro Phe Thr
1 5 10 15
Ile Gly Thr Leu Arg Lys Ala Ile Pro Ala His Cys Phe Glu Arg Ser
20 25 30

Ala Leu Arg Gly Arg Ala Gln Leu Pro Asp Trp Ser Arg Leu Leu Thr
35 40 45
Ala Ile Thr Thr Val Phe Val Lys Ser Lys Arg Pro Asp Met His Asp
50 55 60

Arg Lys Ser Lys Arg Pro Asp Met Leu Val Asp Ser Phe Gly Leu Glu
65 70 75 80
Ser Thr Val Gln Asp Gly Leu Val Phe Arg Gln Ser Phe Ser Ile Arg
85 90 95

Ser Tyr Glu Ile Gly Thr Asp Arg Thr Ala Ser Ile Glu Thr Leu Met
100 105 110
Asn His Leu Gln Glu Thr Ser Leu Asn His Cys Lys Ser Thr Gly Ile
115 120 125
Leu Leu Asp Gly Phe Gly Arg Thr Leu Glu Met Cys Lys Arg Asp Leu
130 135 140
Ile Trp Val Val Ile Lys Met Gln Ile Lys Val Asn Arg Tyr Pro Ala
145 150 155 160

Trp Gly Asp Thr Val Glu Ile Asn Thr Arg Phe Ser Arg Leu Gly Lys
165 170 175
Ile Gly Met Gly Arg Asp Trp Leu Ile Ser Asp Cys Asn Thr Gly Glu
180 185 190
Ile Leu Val Arg Ala Thr Ser Ala Tyr Ala Met Met Asn Gln Lys Thr
195 200 205
Arg Arg Leu Ser Lys Leu Pro Tyr Glu Val His Gln Glu Ile Val Pro
210 215 220

Leu Phe Val Asp Ser Pro Val Ile Glu Asp Ser Asp Leu Lys Val His
225 230 235 240
Lys Phe Lys Val Lys Thr Gly Asp Ser Ile Gln Lys Gly Leu Thr Pro
245 250 255
Gly Trp Asn Asp Leu Asp Val Asn Gln His Val Ser Asn Val Lys Tyr
260 265 270
Ile Gly Trp Ile Leu Glu Ser Met Pro Thr Glu Val Leu Glu Thr Gln
275 280 285

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

305 310 315 320

Arg Ser Gln Tyr Gln His Leu Leu Arg Leu Glu Asp Gly Thr Ala Ile

325 330 335

Val Asn Gly Ala Thr Glu Trp Arg Pro Lys Asn Ala Gly Ala Asn Gly

340 345 350

Ala Ile Ser Thr Gly Lys Thr Ser Asn Gly Asn Ser Val Ser

355 360 365

<210> 33

<211> 4668

<212> DNA

<213> Artificial Sequence

<220><223> synthetic construct

<400> 33

ggtacccgcc tgcaacgcaa gggcagccac agccgctccc acccgccgct gaaccgacac 60

gtgcttgggc gcctgccgcc tgcctgccgc atgcttgtgc tggtgaggct gggcagtgct 120

gccatgctga ttgaggcttg gttcatcggg tggaagctta tgtgtgtgct gggcttgcat 180

gccgggcaat gcgcatggtg gcaagagggc ggcagcactt gctggagctg ccgcggtgcc 240

tccaggtggt tcaatcgcgg cagccagagg gatttcagat gatcgcgcgt acaggttgag 300

cagcagtgtc agcaaaggta gcagtttgcc agaatgatcg gttcagctgt taatcaatgc 360

cagcaagaga aggggtcaag tgcaaacacg ggcatgccac agcacgggca ccggggagtg 420

gaatggcacc accaagtgtg tgcgagccag catcgccgcc tggctgtttc agctacaacg 480

gcaggagtca tccaacgtaa ccatgagctg atcaacactg caatcatcgg gcgggcgtga 540

tgcaagcatg cctggcgaag acacatggtg tgcggatgct gccggctgct gcctgctgcg 600

cacgccgttg agttggcagc aggctcagcc atgcactgga tggcagctgg gctgccactg 660

caatgtggtg gataggatgc aagtggagcg aataccaaac cctctggctg cttgctgggt 720

tgcatggcat cgcaccatca gcaggagcgc atgcgaaggg actggcccca tgcacgccat 780

gccaaaccgg agcgcaccga gtgtccacac tgtcaccagg cccgcaagct ttgcagaacc 840

atgctcatgg acgcatgtag cgctgacgtc ccttgacggc gctcctctcg ggtgtgggaa 900

acgcaatgca gcacaggcag cagaggcggc ggcagcagag cggcggcagc agcggcgggg 960

gccacccttc ttgcggggtc gcgccccagc cagcggtgat gcgctgatcc caaacgagtt 1020

cacattcatt tgcatgcctg gagaagcgag gctggggcct ttgggctggt gcagcccgca 1080

atggaatgcg ggaccgccag gctagcagca aaggcgcctc ccctactccg catcgatgtt 1140

ccatagtgca ttggactgca tttgggtggg gcggccggct gtttctttcg tgttgcaaaa 1200

cgcgccagct cagcaacctg tcccgtgggt cccccgtgcc gatgaaatcg tgtgcacgcc 1260

gatcagctga ttgcccggct cgcgaagtag gcgccctcct ttctgctcgc cctctctccg 1320

tcccgcctct agaatatcaa tgatcgagca ggacggcctc cacgccggct cccccgccgc 1380

ctgggtggag cgcctgttcg gctacgactg ggcccagcag accatcggct gctccgacgc 1440

cgccgtgttc cgcctgtccg cccagggccg ccccgtgctg ttcgtgaaga ccgacctgtc 1500

cggcgccctg aacgagctgc aggacgaggc cgcccgcctg tcctggctgg ccaccaccgg 1560

cgtgccctgc gccgccgtgc tggacgtggt gaccgaggcc ggccgcgact ggctgctgct 1620

gggcgaggtg cccggccagg acctgctgtc ctcccacctg gcccccgccg agaaggtgtc 1680

catcatggcc gacgccatgc gccgcctgca caccctggac cccgccacct gccccttcga 1740

ccaccaggcc aagcaccgca tcgagcgcgc ccgcacccgc atggaggccg gcctggtgga 1800

ccaggacgac ctggacgagg agcaccaggg cctggccccc gccgagctgt tcgcccgcct 1860

gaaggcccgc atgcccgacg gcgaggacct ggtggtgacc cacggcgacg cctgcctgcc 1920

caacatcatg gtggagaacg gccgcttctc cggcttcatc gactgcggcc gcctgggcgt 1980

ggccgaccgc taccaggaca tcgccctggc cacccgcgac atcgccgagg agctgggcgg 2040

cgagtgggcc gaccgcttcc tggtgctgta cggcatcgcc gcccccgact cccagcgcat 2100

cgccttctac cgcctgctgg acgagttctt ctgacaattg gcagcagcag ctcggatagt 2160

atcgacacac tctggacgct ggtcgtgtga tggactgttg ccgccacact tgctgccttg 2220

acctgtgaat atccctgccg cttttatcaa acagcctcag tgtgtttgat cttgtgtgta 2280

cgcgcttttg cgagttgcta gctgcttgtg ctatttgcga ataccacccc cagcatcccc 2340

ttccctcgtt tcatatcgct tgcatcccaa ccgcaactta tctacgctgt cctgctatcc 2400

ctcagcgctg ctcctgctcc tgctcactgc ccctcgcaca gccttggttt gggctccgcc 2460

tgtattctcc tggtactgca acctgtaaac cagcactgca atgctgatgc acgggaagta 2520

gtgggatggg aacacaaatg gaggatcccg cgtctcgaac agagcgcgca gaggaacgct 2580

gaaggtctcg cctctgtcgc acctcagcgc ggcatacacc acaataacca cctgacgaat 2640

gcgcttggtt cttcgtccat tagcgaagcg tccggttcac acacgtgcca cgttggcgag 2700

gtggcaggtg acaatgatcg gtggagctga tggtcgaaac gttcacagcc tagggatatc 2760

gaattccttt cttgcgctat gacacttcca gcaaaaggta gggcgggctg cgagacggct 2820

tcccggcgct gcatgcaaca ccgatgatgc ttcgaccccc cgaagctcct tcggggctgc 2880

atgggcgctc cgatgccgct ccagggcgag cgctgtttaa atagccaggc ccccgattgc 2940

aaagacatta tagcgagcta ccaaagccat attcaaacac ctagatcact accacttcta 3000

cacaggccac tcgagcttgt gatcgcactc cgctaagggg gcgcctcttc ctcttcgttt 3060

cagtcacaac ccgcaaacac tagtatggct atcaagacga acaggcagcc tgtggagaag 3120

cctccgttca cgatcgggac gctgcgcaag gccatccccg cgcactgttt cgagcgctcg 3180

gcgcttcgtg ggcgcgccca gctgcccgac tggagccgcc tgctgaccgc catcaccacc 3240

gtgttcgtga agtccaagcg ccccgacatg cacgaccgca agtccaagcg ccccgacatg 3300

ctggtggaca gcttcggcct ggagtccacc gtgcaggacg gcctggtgtt ccgccagtcc 3360

ttctccatcc gctcctacga gatcggcacc gaccgcaccg ccagcatcga gaccctgatg 3420

aaccacctgc aggagacctc cctgaaccac tgcaagagca ccggcatcct gctggacggc 3480

ttcggccgca ccctggagat gtgcaagcgc gacctgatct gggtggtgat caagatgcag 3540

atcaaggtga accgctaccc cgcctggggc gacaccgtgg agatcaacac ccgcttcagc 3600

cgcctgggca agatcggcat gggccgcgac tggctgatct ccgactgcaa caccggcgag 3660

atcctggtgc gcgccaccag cgcctacgcc atgatgaacc agaagacccg ccgcctgtcc 3720

aagctgccct acgaggtgca ccaggagatc gtgcccctgt tcgtggacag ccccgtgatc 3780

gaggactccg acctgaaggt gcacaagttc aaggtgaaga ccggcgacag catccagaag 3840

ggcctgaccc ccggctggaa cgacctggac gtgaaccagc acgtgtccaa cgtgaagtac 3900

atcggctgga tcctggagag catgcccacc gaggtgctgg agacccagga gctgtgctcc 3960

ctggccctgg agtaccgccg cgagtgcggc cgcgactccg tgctggagag cgtgaccgcc 4020

atggacccca gcaaggtggg cgtgcgctcc cagtaccagc acctgctgcg cctggaggac 4080

ggcaccgcca tcgtgaacgg cgccaccgag tggcgcccca agaacgccgg cgccaacggc 4140

gccatctcca ccggcaagac cagcaacggc aactccgtgt ccatggacta caaggaccac 4200

gacggcgact acaaggacca cgacatcgac tacaaggacg acgacgacaa gtgactcgag 4260

gcagcagcag ctcggatagt atcgacacac tctggacgct ggtcgtgtga tggactgttg 4320

ccgccacact tgctgccttg acctgtgaat atccctgccg cttttatcaa acagcctcag 4380

tgtgtttgat cttgtgtgta cgcgcttttg cgagttgcta gctgcttgtg ctatttgcga 4440

ataccacccc cagcatcccc ttccctcgtt tcatatcgct tgcatcccaa ccgcaactta 4500

tctacgctgt cctgctatcc ctcagcgctg ctcctgctcc tgctcactgc ccctcgcaca 4560

gccttggttt gggctccgcc tgtattctcc tggtactgca acctgtaaac cagcactgca 4620

atgctgatgc acgggaagta gtgggatggg aacacaaatg gaaagctt 4668

<210> 34

<211> 360

<212> PRT

<213> Artificial Sequence

<220><223> synthetic construct

<400> 34

Met Ala Ile Lys Thr Asn Arg Gln Pro Val Glu Lys Pro Pro Phe Thr
1 5 10 15

Ile Gly Thr Leu Arg Lys Ala Ile Pro Ala His Cys Phe Glu Arg Ser
20 25 30

Ala Leu Arg Gly Arg Ala Pro Asp Trp Ser Met Leu Phe Ala Val Ile
35 40 45

Thr Thr Ile Phe Ser Ala Ala Glu Lys Gln Trp Thr Asn Leu Glu Trp
50 55 60

Lys Pro Lys Pro Lys Leu Pro Gln Leu Leu Asp Asp His Phe Gly Leu
65 70 75 80

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

Gly Pro Asp Arg Ser Thr Ser Ile Leu Ala Val Met Asn His Met Gln
100 105 110

Glu Ala Thr Leu Asn His Ala Lys Ser Val Gly Ile Leu Gly Asp Gly
115 120 125

Phe Gly Thr Thr Leu Glu Met Ser Lys Arg Asp Leu Met Trp Val Val
130 135 140

Arg Arg Thr His Val Ala Val Glu Arg Tyr Pro Thr Trp Gly Asp Thr
145 150 155 160

Val Glu Val Glu Cys Trp Ile Gly Ala Ser Gly Asn Asn Gly Met Arg
165 170 175

Arg Asp Phe Leu Val Arg Asp Cys Lys Thr Gly Glu Ile Leu Thr Arg
180 185 190

Cys Thr Ser Leu Ser Val Leu Met Asn Thr Arg Thr Arg Arg Leu Ser
195 200 205

Thr Ile Pro Asp Glu Val Arg Gly Glu Ile Gly Pro Ala Phe Ile Asp

210 215 220

Asn Val Ala Val Lys Asp Asp Glu Ile Lys Lys Leu Gln Lys Leu Asn

225 230 235 240

Asp Ser Thr Ala Asp Tyr Ile Gln Gly Gly Leu Thr Pro Arg Trp Asn

245 250 255

Asp Leu Asp Val Asn Gln His Val Asn Asn Leu Lys Tyr Val Ala Trp

260 265 270

Val Phe Glu Thr Val Pro Asp Ser Ile Phe Glu Ser His His Ile Ser

275 280 285

Ser Phe Thr Leu Glu Tyr Arg Arg Glu Cys Thr Arg Asp Ser Val Leu

290 295 300

Arg Ser Leu Thr Thr Val Ser Gly Gly Ser Ser Glu Ala Gly Leu Val

305 310 315 320

Cys Asp His Leu Leu Gln Leu Glu Gly Gly Ser Glu Val Leu Arg Ala

325 330 335

Arg Thr Glu Trp Arg Pro Lys Leu Thr Asp Ser Phe Arg Gly Ile Ser

340 345 350

Val Ile Pro Ala Glu Pro Arg Val

355 360

<210> 35

<211> 4653

<212> DNA

<213> Artificial Sequence

<220><223> synthetic construct

<400> 35

ggtacccgcc tgcaacgcaa gggcagccac agccgctccc acccgccgct gaaccgacac 60

gtgcttgggc gcctgccgcc tgcctgccgc atgcttgtgc tggtgaggct gggcagtgct 120

gccatgctga ttgaggcttg gttcatcggg tggaagctta tgtgtgtgct gggcttgcat 180

gccgggcaat gcgcatggtg gcaagagggc ggcagcactt gctggagctg ccgcggtgcc 240

tccaggtggt tcaatcgcgg cagccagagg gatttcagat gatcgcgcgt acaggttgag 300

cagcagtgtc agcaaaggta gcagtttgcc agaatgatcg gttcagctgt taatcaatgc 360

cagcaagaga aggggtcaag tgcaaacacg ggcatgccac agcacgggca ccggggagtg 420

gaatggcacc accaagtgtg tgcgagccag catcgccgcc tggctgtttc agctacaacg 480

gcaggagtca tccaacgtaa ccatgagctg atcaacactg caatcatcgg gcgggcgtga 540

tgcaagcatg cctggcgaag acacatggtg tgcggatgct gccggctgct gcctgctgcg 600

cacgccgttg agttggcagc aggctcagcc atgcactgga tggcagctgg gctgccactg 660

caatgtggtg gataggatgc aagtggagcg aataccaaac cctctggctg cttgctgggt 720

tgcatggcat cgcaccatca gcaggagcgc atgcgaaggg actggcccca tgcacgccat 780

gccaaaccgg agcgcaccga gtgtccacac tgtcaccagg cccgcaagct ttgcagaacc 840

atgctcatgg acgcatgtag cgctgacgtc ccttgacggc gctcctctcg ggtgtgggaa 900

acgcaatgca gcacaggcag cagaggcggc ggcagcagag cggcggcagc agcggcgggg 960

gccacccttc ttgcggggtc gcgccccagc cagcggtgat gcgctgatcc caaacgagtt 1020

cacattcatt tgcatgcctg gagaagcgag gctggggcct ttgggctggt gcagcccgca 1080

atggaatgcg ggaccgccag gctagcagca aaggcgcctc ccctactccg catcgatgtt 1140

ccatagtgca ttggactgca tttgggtggg gcggccggct gtttctttcg tgttgcaaaa 1200

cgcgccagct cagcaacctg tcccgtgggt cccccgtgcc gatgaaatcg tgtgcacgcc 1260

gatcagctga ttgcccggct cgcgaagtag gcgccctcct ttctgctcgc cctctctccg 1320

tcccgcctct agaatatcaa tgatcgagca ggacggcctc cacgccggct cccccgccgc 1380

ctgggtggag cgcctgttcg gctacgactg ggcccagcag accatcggct gctccgacgc 1440

cgccgtgttc cgcctgtccg cccagggccg ccccgtgctg ttcgtgaaga ccgacctgtc 1500

cggcgccctg aacgagctgc aggacgaggc cgcccgcctg tcctggctgg ccaccaccgg 1560

cgtgccctgc gccgccgtgc tggacgtggt gaccgaggcc ggccgcgact ggctgctgct 1620

gggcgaggtg cccggccagg acctgctgtc ctcccacctg gcccccgccg agaaggtgtc 1680

catcatggcc gacgccatgc gccgcctgca caccctggac cccgccacct gcccccttcga 1740

ccaccaggcc aagcaccgca tcgagcgcgc ccgcacccgc atggaggccg gcctggtgga 1800

ccaggacgac ctggacgagg agcaccaggg cctggccccc gccgagctgt tcgcccgcct 1860

gaaggcccgc atgcccgacg gcgaggacct ggtggtgacc cacggcgacg cctgcctgcc 1920

caacatcatg gtggagaacg gccgcttctc cggcttcatc gactgcggcc gcctgggcgt 1980

ggccgaccgc taccaggaca tcgccctggc cacccgcgac atcgccgagg agctgggcgg 2040

cgagtgggcc gaccgcttcc tggtgctgta cggcatcgcc gcccccgact cccagcgcat 2100

cgccttctac cgcctgctgg acgagttctt ctgacaattg gcagcagcag ctcggatagt 2160

atcgacacac tctggacgct ggtcgtgtga tggactgttg ccgccacact tgctgccttg 2220

acctgtgaat atccctgccg cttttatcaa acagcctcag tgtgtttgat cttgtgtgta 2280

cgcgcttttg cgagttgcta gctgcttgtg ctatttgcga ataccacccc cagcatcccc 2340
ttccctcgtt tcatatcgct tgcatcccaa ccgcaactta tctacgctgt cctgctatcc 2400
ctcagcgctg ctcctgctcc tgctcactgc ccctcgcaca gccttggttt gggctccgcc 2460
tgtattctcc tggtactgca acctgtaaac cagcactgca atgctgatgc acgggaagta 2520
gtgggatggg aacacaaatg gaggatcccg cgtctcgaac agagcgcgca gaggaacgct 2580
gaaggtctcg cctctgtcgc acctcagcgc ggcatacacc acaataacca cctgacgaat 2640
gcgcttggtt cttcgtccat tagcgaagcg tccggttcac acacgtgcca cgttggcgag 2700

gtggcaggtg acaatgatcg gtggagctga tggtcgaaac gttcacagcc tagggatatc 2760
gaattccttt cttgcgctat gacacttcca gcaaaaggta gggcgggctg cgagacggct 2820
tcccggcgct gcatgcaaca ccgatgatgc ttcgaccccc cgaagctcct tcggggctgc 2880
atgggcgctc cgatgccgct ccagggcgag cgctgtttaa atagccaggc ccccgattgc 2940
aaagacatta tagcgagcta ccaaagccat attcaaacac ctagatcact accacttcta 3000
cacaggccac tcgagcttgt gatcgcactc cgctaagggg gcgcctcttc ctcttcgttt 3060
cagtcacaac ccgcaaacac tagtatggct atcaagacga acaggcagcc tgtggagaag 3120

cctccgttca cgatcgggac gctgcgcaag gccatccccg cgcactgttt cgagcgctcg 3180
gcgcttcgtg ggcgcgcccc cgactggtcc atgctgttcg ccgtgatcac caccatcttc 3240
agcgccgccg agaagcagtg gaccaacctg gagtggaagc ccaagcccaa gctgccccag 3300
ctgctggacg accacttcgg cctgcacggc ctggtgttcc gccgcacctt cgccatccgc 3360
tcctacgagg tgggccccga ccgcagcacc tccatcctgg ccgtgatgaa ccacatgcag 3420
gaggccaccc tgaaccacgc caagagcgtg ggcatcctgg gcgacggctt cggcaccacc 3480
ctggagatgt ccaagcgcga cctgatgtgg gtggtgcgcc gcacccacgt ggccgtggag 3540

cgctacccca cctggggcga caccgtggag gtggagtgct ggatcggcgc cagcggcaac 3600
aacggcatgc gccgcgactt cctggtgcgc gactgcaaga ccggcgagat cctgacccgc 3660
tgcacctccc tgagcgtgct gatgaacacc cgcacccgcc gcctgagcac catccccgac 3720
gaggtgcgcg gcgagatcgg ccccgccttc atcgacaacg tggccgtgaa ggacgacgag 3780
atcaagaagc tgcagaagct gaacgactcc accgccgact acatccaggg cggcctgacc 3840
ccccgctgga acgacctgga cgtgaaccag cacgtgaaca acctgaagta cgtggcctgg 3900
gtgttcgaga ccgtgcccga cagcatcttc gagtcccacc acatcagctc cttcaccctg 3960

gagtaccgcc gcgagtgcac ccgcgactcc gtgctgcgca gcctgaccac cgtgagcggc 4020

ggcagctccg aggccggcct ggtgtgcgac cacctgctgc agctggaggg cggcagcgag 4080

gtgctgcgcg cccgcaccga gtggcgcccc aagctgaccg actccttccg cggcatcagc 4140

gtgatccccg ccgagccccg cgtgatggac tacaaggacc acgacggcga ctacaaggac 4200

cacgacatcg actacaagga cgacgacgac aagtgatgac tcgaggcagc agcagctcgg 4260

atagtatcga cacactctgg acgctggtcg tgtgatggac tgttgccgcc acacttgctg 4320

ccttgacctg tgaatatccc tgccgctttt atcaaacagc ctcagtgtgt ttgatcttgt 4380

gtgtacgcgc ttttgcgagt tgctagctgc ttgtgctatt tgcgaatacc acccccagca 4440

tcccccttccc tcgtttcata tcgcttgcat cccaaccgca acttatctac gctgtcctgc 4500

tatccctcag cgctgctcct gctcctgctc actgcccctc gcacagcctt ggtttgggct 4560

ccgcctgtat tctcctggta ctgcaacctg taaaccagca ctgcaatgct gatgcacggg 4620

aagtagtggg atgggaacac aaatggaaag ctt 4653

<210> 36

<211> 393

<212> PRT

<213> Artificial Sequence

<220><223> synthetic construct

<400> 36

Met Ala Ile Lys Thr Asn Arg Gln Pro Val Glu Lys Pro Pro Phe Thr

1 5 10 15

Ile Gly Thr Leu Arg Lys Ala Ile Pro Ala His Cys Phe Glu Arg Ser

20 25 30

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

35 40 45

Ser Gly Ser Leu Asn Thr Gln Glu Asp Thr Leu Ser Ser Ser Pro Pro

50 55 60

Pro Arg Ala Phe Phe Asn Gln Leu Pro Asp Trp Ser Met Leu Leu Thr

65 70 75 80

Ala Ile Thr Thr Val Phe Val Ala Pro Glu Lys Arg Trp Thr Met Phe

85 90 95

Asp Arg Lys Ser Lys Arg Pro Asn Met Leu Met Asp Ser Phe Gly Leu

100 105 110

Glu Arg Val Val Gln Asp Gly Leu Val Phe Arg Gln Ser Phe Ser Ile
115 120 125
Arg Ser Tyr Glu Ile Cys Ala Asp Arg Thr Ala Ser Ile Glu Thr Val

130 135 140
Met Asn His Val Gln Glu Thr Ser Leu Asn Gln Cys Lys Ser Ile Gly
145 150 155 160
Leu Leu Asp Asp Gly Phe Gly Arg Ser Pro Glu Met Cys Lys Arg Asp
165 170 175
Leu Ile Trp Val Val Thr Arg Met Lys Ile Met Val Asn Arg Tyr Pro
180 185 190
Thr Trp Gly Asp Thr Ile Glu Val Ser Thr Trp Leu Ser Gln Ser Gly

195 200 205
Lys Ile Gly Met Gly Arg Asp Trp Leu Ile Ser Asp Cys Asn Thr Gly
210 215 220
Glu Ile Leu Val Arg Ala Thr Ser Val Tyr Ala Met Met Asn Gln Lys
225 230 235 240
Thr Arg Arg Phe Ser Lys Leu Pro His Glu Val Arg Gln Glu Phe Ala
245 250 255
Pro His Phe Leu Asp Ser Pro Pro Ala Ile Glu Asp Asn Asp Gly Lys

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

325 330 335
Arg Asp Ser Val Leu Glu Ser Val Thr Ser Met Asp Pro Ser Lys Val
340 345 350
Gly Asp Arg Phe Gln Tyr Arg His Leu Leu Arg Leu Glu Asp Gly Ala

355 360 365

Asp Ile Met Lys Gly Arg Thr Glu Trp Arg Pro Lys Asn Ala Gly Thr

370 375 380

Asn Gly Ala Ile Ser Thr Gly Lys Thr

385 390

<210> 37

<211> 4686

<212> DNA

<213> Artificial Sequence

<220><223> synthetic construct

<400> 37

ggtacccgcc tgcaacgcaa gggcagccac agccgctccc acccgccgct gaaccgacac 60

gtgcttgggc gcctgccgcc tgcctgccgc atgcttgtgc tggtgaggct gggcagtgct 120

gccatgctga ttgaggcttg gttcatcggg tggaagctta tgtgtgtgct gggcttgcat 180

gccgggcaat gcgcatggtg gcaagagggc ggcagcactt gctggagctg ccgcggtgcc 240

tccaggtggt tcaatcgcgg cagccagagg gatttcagat gatcgcgcgt acaggttgag 300

cagcagtgtc agcaaaggta gcagtttgcc agaatgatcg gttcagctgt taatcaatgc 360

cagcaagaga aggggtcaag tgcaaacacg ggcatgccac agcacgggca ccggggagtg 420

gaatggcacc accaagtgtg tgcgagccag catcgccgcc tggctgtttc agctacaacg 480

gcaggagtca tccaacgtaa ccatgagctg atcaacactg caatcatcgg gcgggcgtga 540

tgcaagcatg cctggcgaag acacatggtg tgcggatgct gccggctgct gcctgctgcg 600

cacgccgttg agttggcagc aggctcagcc atgcactgga tggcagctgg gctgccactg 660

caatgtggtg gataggatgc aagtggagcg aataccaaac cctctggctg cttgctgggt 720

tgcatggcat cgcaccatca gcaggagcgc atgcgaaggg actggcccca tgcacgccat 780

gccaaaccgg agcgcaccga gtgtccacac tgtcaccagg cccgcaagct ttgcagaacc 840

atgctcatgg acgcatgtag cgctgacgtc ccttgacggc gctcctctcg ggtgtgggaa 900

acgcaatgca gcacaggcag cagaggcggc ggcagcagag cggcggcagc agcggcgggg 960

gccacccttc ttgcggggtc gcgccccagc cagcggtgat gcgctgatcc caaacgagtt 1020

cacattcatt tgcatgcctg gagaagcgag gctggggcct ttgggctggt gcagcccgca 1080

atggaatgcg ggaccgccag gctagcagca aaggcgcctc ccctactccg catcgatgtt 1140

ccatagtgca ttggactgca tttgggtggg gcggccggct gtttctttcg tgttgcaaaa 1200

cgcgccagct cagcaacctg tcccgtgggt cccccgtgcc gatgaaatcg tgtgcacgcc 1260

gatcagctga ttgcccggct cgcgaagtag gcgccctcct ttctgctcgc cctctctccg 1320

tcccgcctct agaatatcaa tgatcgagca ggacggcctc cacgccggct cccccgccgc 1380

ctgggtggag cgcctgttcg gctacgactg ggcccagcag accatcggct gctccgacgc 1440

cgccgtgttc cgcctgtccg cccagggccg ccccgtgctg ttcgtgaaga ccgacctgtc 1500

cggcgccctg aacgagctgc aggacgaggc cgcccgcctg tcctggctgg ccaccaccgg 1560

cgtgccctgc gccgccgtgc tggacgtggt gaccgaggcc ggccgcgact ggctgctgct 1620

gggcgaggtg cccggccagg acctgctgtc ctcccacctg gcccccgccg agaaggtgtc 1680

catcatggcc gacgccatgc gccgcctgca caccctggac cccgccacct gccccttcga 1740

ccaccaggcc aagcaccgca tcgagcgcgc ccgcacccgc atggaggccg gcctggtgga 1800

ccaggacgac ctggacgagg agcaccaggg cctggccccc gccgagctgt tcgcccgcct 1860

gaaggcccgc atgcccgacg gcgaggacct ggtggtgacc cacggcgacg cctgcctgcc 1920

caacatcatg gtggagaacg gccgcttctc cggcttcatc gactgcggcc gcctgggcgt 1980

ggccgaccgc taccaggaca tcgccctggc cacccgcgac atcgccgagg agctgggcgg 2040

cgagtgggcc gaccgcttcc tggtgctgta cggcatcgcc gcccccgact cccagcgcat 2100

cgccttctac cgcctgctgg acgagttctt ctgacaattg gcagcagcag ctcggatagt 2160

atcgacacac tctggacgct ggtcgtgtga tggactgttg ccgccacact tgctgccttg 2220

acctgtgaat atccctgccg cttttatcaa acagcctcag tgtgtttgat cttgtgtgta 2280

cgcgcttttg cgagttgcta gctgcttgtg ctatttgcga ataccacccc cagcatcccc 2340

ttccctcgtt tcatatcgct tgcatcccaa ccgcaactta tctacgctgt cctgctatcc 2400

ctcagcgctg ctcctgctcc tgctcactgc ccctcgcaca gccttggttt gggctccgcc 2460

tgtattctcc tggtactgca acctgtaaac cagcactgca atgctgatgc acgggaagta 2520

gtgggatggg aacacaaatg gaggatcccg cgtctcgaac agagcgcgca gaggaacgct 2580

gaaggtctcg cctctgtcgc acctcagcgc ggcatacacc acaataacca cctgacgaat 2640

gcgcttggtt cttcgtccat tagcgaagcg tccggttcac acacgtgcca cgttggcgag 2700

gtggcaggtg acaatgatcg gtggagctga tggtcgaaac gttcacagcc tagggatatc 2760

gaattccttt cttgcgctat gacacttcca gcaaaaggta gggcgggctg cgagacggct 2820

tcccggcgct gcatgcaaca ccgatgatgc ttcgaccccc cgaagctcct tcggggctgc 2880

atgggcgctc cgatgccgct ccagggcgag cgctgtttaa atagccaggc ccccgattgc 2940

aaagacatta tagcgagcta ccaaagccat attcaaacac ctagatcact accacttcta 3000

cacaggccac tcgagcttgt gatcgcactc cgctaagggg gcgcctcttc ctcttcgttt 3060

cagtcacaac ccgcaaacac tagtatggct atcaagacga acaggcagcc tgtggagaag 3120

cctccgttca cgatcgggac gctgcgcaag gccatccccg cgcactgttt cgagcgctcg 3180

gcgcttcgtg ggcgcgcccc cgcgaacggc agcgcggtga ccctgaagtc gggctccctg 3240

aacacccagg aggacacgct gagctcgtcc ccccccccgc gcgcgttctt caaccagctg 3300

cccgactgga gcatgctgct gaccgcgatc accacggtct tcgtggcgcc cgagaagcgc 3360

tggaccatgt tcgaccgcaa gtcgaagcgc cccaacatgc tgatggactc cttcggcctg 3420

gagcgcgtgg tccaggacgg cctggtgttc cgccagagct tctcgatccg ctcctacgag 3480

atctgcgcgg accgcaccgc gagcatcgag acggtgatga accacgtcca ggagacctcg 3540

ctgaaccagt gcaagtccat cggcctgctg gacgacggct tcggccgcag ccccgagatg 3600

tgcaagcgcg acctgatctg ggtggtcacc cgcatgaaga tcatggtgaa ccgctacccc 3660

acgtggggcg acaccatcga ggtctcgacg tggctgtccc agagcggcaa gatcggcatg 3720

ggccgcgact ggctgatctc ggactgcaac accggcgaga tcctggtgcg cgcgacgtcc 3780

gtctacgcga tgatgaacca gaagacccgc cgcttcagca agctgcccca cgaggtgcgc 3840

caggagttcg cgccccactt cctggactcg ccccccgcga tcgaggacaa cgacggcaag 3900

ctgcagaagt tcgacgtcaa gacgggcgac tccatccgca agggcctgac ccccggctgg 3960

tacgacctgg acgtgaacca gcacgtgagc aacgtcaagt acatcggctg gatcctggag 4020

tcgatgccca ccgaggtcct ggagacgcag gagctgtgct ccctgaccct ggagtaccgc 4080

cgcgagtgcg gccgcgactc ggtgctggag agcgtcacca gcatggaccc ctcgaaggtg 4140

ggcgaccgct tccagtaccg ccacctgctg cgcctggagg acggcgcgga catcatgaag 4200

ggccgcaccg agtggcgccc caagaacgcg ggcacgaacg gcgcgatctc caccggcaag 4260

acgtgactcg aggcagcagc agctcggata gtatcgacac actctggacg ctggtcgtgt 4320

gatggactgt tgccgccaca cttgctgcct tgacctgtga atatccctgc cgcttttatc 4380

aaacagcctc agtgtgtttg atcttgtgtg tacgcgcttt tgcgagttgc tagctgcttg 4440

tgctatttgc gaataccacc cccagcatcc ccttccctcg tttcatatcg cttgcatccc 4500

aaccgcaact tatctacgct gtcctgctat ccctcagcgc tgctcctgct cctgctcact 4560

gcccctcgca cagccttggt ttgggctccg cctgtattct cctggtactg caacctgtaa 4620

accagcactg caatgctgat gcacgggaag tagtgggatg ggaacacaaa tggaaagctt 4680

gagctc 4686

<210> 38

<211> 408

<212> PRT

<213> Artificial Sequence

<220><223> synthetic construct

<400> 38

Met Ala Ile Lys Thr Asn Arg Gln Pro Val Glu Lys Pro Pro Phe Thr
1 5 10 15
Ile Gly Thr Leu Arg Lys Ala Ile Pro Ala His Cys Phe Glu Arg Ser

20 25 30
Ala Leu Arg Gly Arg Ala Pro Ala Asn Gly Ser Ala Val Asn Leu Lys
35 40 45
Ser Gly Ser Leu Asn Thr Gln Glu Asp Thr Ser Ser Ser Pro Pro Pro
50 55 60
Arg Ala Phe Leu Asn Gln Leu Pro Asp Trp Ser Met Leu Leu Thr Ala
65 70 75 80
Ile Thr Thr Val Phe Val Ala Ala Glu Lys Gln Trp Thr Met Leu Asp

85 90 95
Arg Lys Ser Lys Arg Pro Asp Met Leu Val Asp Ser Val Gly Leu Lys
100 105 110
Ser Ile Val Arg Asp Gly Leu Val Ser Arg Gln Ser Phe Leu Ile Arg
115 120 125
Ser Tyr Glu Ile Gly Ala Asp Arg Thr Ala Ser Ile Glu Thr Leu Met
130 135 140
Asn His Leu Gln Glu Thr Ser Ile Asn His Cys Lys Ser Leu Gly Leu

145 150 155 160
Leu Asn Asp Gly Phe Gly Arg Thr Pro Gly Met Cys Lys Asn Asp Leu
165 170 175
Ile Trp Val Leu Thr Lys Met Gln Ile Met Val Asn Arg Tyr Pro Thr
180 185 190
Trp Gly Asp Thr Val Glu Ile Asn Thr Trp Phe Ser Gln Ser Gly Lys
195 200 205
Ile Gly Met Ala Ser Asp Trp Leu Ile Ser Asp Cys Asn Thr Gly Glu

210 215 220

Ile Leu Ile Arg Ala Thr Ser Val Trp Ala Met Met Asn Gln Lys Thr
225 230 235 240
Arg Arg Phe Ser Arg Leu Pro Tyr Glu Val Arg Gln Glu Leu Thr Pro
245 250 255
His Phe Val Asp Ser Pro His Val Ile Glu Asp Asn Asp Gln Lys Leu
260 265 270
His Lys Phe Asp Val Lys Thr Gly Asp Ser Ile Arg Lys Gly Leu Thr

275 280 285
Pro Arg Trp Asn Asp Leu Asp Val Asn Gln His Val Ser Asn Val Lys
290 295 300
Tyr Ile Gly Trp Ile Leu Glu Ser Met Pro Ile Glu Val Leu Glu Thr
305 310 315 320
Gln Glu Leu Cys Ser Leu Thr Val Glu Tyr Arg Arg Glu Cys Gly Met
325 330 335
Asp Ser Val Leu Glu Ser Val Thr Ala Val Asp Pro Ser Glu Asn Gly

340 345 350
Gly Arg Ser Gln Tyr Lys His Leu Leu Arg Leu Glu Asp Gly Thr Asp
355 360 365
Ile Val Lys Ser Arg Thr Glu Trp Arg Pro Lys Asn Ala Gly Thr Asn
370 375 380
Gly Ala Ile Ser Thr Ser Thr Ala Lys Thr Ser Asn Gly Asn Ser Ala
385 390 395 400
Ser Asp Asp Asp Asp Lys Leu Gly
405

<210> 39
<211> 1239
<212> DNA
<213> Artificial Sequence
<220><223> synthetic construct
<400> 39
actagtatgg ctatcaagac gaacaggcag cctgtggaga agcctccgtt cacgatcggg 60
acgctgcgca aggccatccc cgcgcactgt ttcgagcgct cggcgcttcg tgggcgcgcc 120

cccgcgaacg gcagcgcggt gaacctgaag tcgggctccc tgaacaccca ggaggacacg 180

agctcgtccc ccccccccg cgcgttcctg aaccagctgc ccgactggag catgctgctg 240

accgcgatca ccaccgtctt cgtggcggcg gagaagcagt ggacgatgct ggaccgcaag 300

tcgaagcgcc ccgacatgct ggtggactcc gtcggcctga agagcatcgt gcgcgacggc 360

ctggtctcgc gccagtcctt cctgatccgc agctacgaga tcggcgcgga ccgcaccgcg 420

tcgatcgaga ccctgatgaa ccacctgcag gagacgtcca tcaaccactg caagagcctg 480

ggcctgctga acgacggctt cggccgcacc cccggcatgt gcaagaacga cctgatctgg 540

gtgctgacca agatgcagat catggtcaac cgctacccca cgtggggcga caccgtcgag 600

atcaacacgt ggttctcgca gtccggcaag atcggcatgg cgagcgactg gctgatctcg 660

gactgcaaca ccggcgagat cctgatccgc gcgacctccg tgtgggcgat gatgaaccag 720

aagacgcgcc gcttcagccg cctgccctac gaggtccgcc aggagctgac cccccacttc 780

gtggactcgc cccacgtcat cgaggacaac gaccagaagc tgcacaagtt cgacgtgaag 840

accggcgact ccatccgcaa gggcctgacg ccccgctgga acgacctgga cgtcaaccag 900

cacgtgtcga acgtgaagta catcggctgg atcctggagt ccatgcccat cgaggtcctg 960

gagacccagg agctgtgctc gctgaccgtg gagtaccgcc gcgagtgcgg catggactcc 1020

gtgctggagt cggtcacggc ggtggacccc agcgagaacg gcggccgcag ccagtacaag 1080

cacctgctgc gcctggagga cggcaccgac atcgtcaagt cgcgcaccga gtggcgcccc 1140

aagaacgcgg gcacgaacgg cgcgatctcc accagcaccg cgaagacgtc gaacggcaac 1200

tccgcgagcg atgacgatga caagctggga tgactcgag 1239

<210> 40

<211> 399

<212> PRT

<213> Artificial Sequence

<220><223> synthetic construct

<400> 40

Met Ala Thr Ala Ser Thr Phe Ser Ala Phe Asn Ala Arg Cys Gly Asp

1 5 10 15

Leu Arg Arg Ser Ala Gly Ser Gly Pro Arg Arg Pro Ala Arg Pro Leu

20 25 30

Pro Val Arg Gly Arg Ala Ala Gln Ala Ala Thr Arg Val Asn Gly Ser

35 40 45

Lys Val Gly Leu Lys Thr Asp Thr Asn Lys Leu Glu Asp Ala Pro Phe
50 55 60
Ile Pro Ser Ser Ala Pro Arg Thr Phe Tyr Asn Gln Leu Pro Asp Trp
65 70 75 80
Ser Val Leu Leu Ala Ala Ile Thr Thr Ile Phe Leu Ala Ala Glu Lys
85 90 95
Gln Trp Thr Leu Ile Asp Trp Lys Arg Gly Gly Pro Asp Met Leu Ser

100 105 110
Asp Ala Phe Gly Leu Pro Lys Ile Ile Glu Asn Gly Leu Leu Tyr Arg
115 120 125
Gln Lys Phe Ser Ile Arg Ser Tyr Glu Ile Gly Ala Asp Gln Thr Ala
130 135 140
Ser Ile Glu Thr Leu Met Asn His Leu Gln Glu Thr Ala Leu Asn His
145 150 155 160
Val Lys Cys Ala Gly Leu Leu Gly Asn Gly Phe Gly Ser Thr Pro Glu

165 170 175
Met Ser Lys Met Asn Leu Ile Trp Val Val Thr Lys Met Gln Val Leu
180 185 190
Val Glu His Tyr Pro Ser Trp Gly Asp Val Ile Glu Val Asp Thr Trp
195 200 205
Ala Ala Ala Ser Gly Lys Asn Gly Met Arg Arg Asp Trp His Val Arg
210 215 220
Asp Trp Gln Thr Gly Gln Thr Ile Met Arg Ala Ser Ser Asn Trp Val

225 230 235 240
Met Met Asn Gln Asn Thr Arg Arg Leu Ser Lys Phe Pro Glu Glu Val
245 250 255
Arg Ala Glu Ile Glu Pro Tyr Phe Met Glu Arg Ala Pro Val Ile Asp
260 265 270
Asp Asp Asn Arg Lys Leu Pro Lys Leu Asp Asp Asp Thr Ala Asp His
275 280 285
Val Arg Asn Gly Leu Thr Pro Arg Trp Ser Asp Leu Asp Val Asn Gln

290 295 300

His Val Lys Asn Val Lys Tyr Ile Gly Trp Ile Leu Glu Ser Ala Pro

305 310 315 320

Ile Ser Ile Leu Glu Ser His Glu Leu Ala Ser Met Thr Leu Glu Tyr

325 330 335

Arg Arg Glu Cys Gly Arg Asp Ser Val Leu Gln Ser Leu Thr Ser Val

340 345 350

Ser Asn Asn Cys Thr Asp Gly Ser Glu Glu Leu Pro Ile Glu Cys Gln

355 360 365

His Leu Leu Arg Asn Glu Gly Gly Ser Glu Ile Val Lys Gly Arg Thr

370 375 380

Glu Trp Arg Pro Lys Lys Cys Gly Pro Phe Gly Ala Gly Arg Pro

385 390 395

<210> 41

<211> 1212

<212> DNA

<213> Artificial Sequence

<220><223> synthetic construct

<400> 41

actagtatgg ccaccgcatc cactttctcg gcgttcaatg cccgctgcgg cgacctgcgt 60

cgctcggcgg gctccgggcc ccggcgccca gcgaggcccc tccccgtgcg cgggcgcgcc 120

gcccaggcgg ccacccgcgt gaacggcagc aaggtgggcc tgaagaccga caccaacaag 180

ctggaggacg cgcccttcat cccctcgtcc gccccccgca ccttctacaa ccagctgccc 240

gactggagcg tcctgctggc ggccatcacc accatcttcc tggcggccga gaagcagtgg 300

accctgatcg actggaagcg cggcggcccc gacatgctgt cggacgcgtt cggcctgccc 360

aagatcatcg agaacggcct gctgtaccgc cagaagttct ccatccgcag ctacgagatc 420

ggcgccgacc agaccgcctc gatcgagacc ctgatgaacc acctgcagga gaccgcgctg 480

aaccacgtca agtgcgccgg cctgctgggc aacggcttcg gctccacccc cgagatgagc 540

aagatgaacc tgatctgggt ggtcaccaag atgcaggtgc tggtcgagca ctacccctcg 600

tggggcgacg tgatcgaggt ggacacctgg gcggccgcgt ccggcaagaa cggcatgcgc 660

cgcgactggc acgtccgcga ctggcagacc ggccagacca tcatgcgcgc cagctcgaac 720

tgggtgatga tgaaccagaa cacccgccgc ctgtccaagt tccccgagga ggtccgcgcc 780

gagatcgagc cctacttcat ggagcgcgcc cccgtgatcg acgacgacaa ccgcaagctg 840

cccaagctgg acgacgacac cgcggaccac gtgcgcaacg gcctgacccc ccgctggagc 900

gacctggacg tgaaccagca cgtcaagaac gtgaagtaca tcggctggat cctggagtcg 960

gcccccatct ccatcctgga gagccacgag ctggcctcga tgaccctgga gtaccgccgc 1020

gagtgcggcc gcgactccgt cctgcagagc ctgacctcgg tgtccaacaa ctgcaccgac 1080

ggcagcgagg agctgcccat cgagtgccag cacctgctgc gcaacgaggg cggctcggag 1140

atcgtcaagg gccgcaccga gtggcgcccc aagaagtgcg gcccccttcgg cgccggccgc 1200

ccctgactcg ag 1212

<210> 42

<211> 415

<212> PRT

<213> Artificial Sequence

<220><223> synthetic construct

<400> 42

Met Ala Ile Lys Thr Asn Arg Gln Pro Val Glu Lys Pro Pro Phe Thr
1 5 10 15

Ile Gly Thr Leu Arg Lys Ala Ile Pro Ala His Cys Phe Glu Arg Ser
20 25 30

Ala Leu Arg Gly Arg Ala Ala Asn Ala His Thr Val Pro Lys Ile Asn
35 40 45

Gly Asn Lys Ala Gly Leu Leu Thr Pro Met Glu Ser Thr Lys Asp Glu
50 55 60

Asp Ile Val Ala Ala Pro Thr Val Ala Pro Lys Arg Thr Phe Ile Asn
65 70 75 80

Gln Leu Pro Asp Trp Ser Met Leu Leu Ala Ala Ile Thr Thr Ile Phe
85 90 95

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

Pro Asp Met Leu Val Asp Phe Asp Pro Phe Ser Leu Gly Arg Phe Val
115 120 125

Gln Asp Gly Leu Ile Phe Arg Gln Asn Phe Ser Ile Arg Ser Tyr Glu

130 135 140

Ile Gly Ala Asp Arg Thr Ala Ser Ile Glu Thr Leu Met Asn His Leu

145 150 155 160

Gln Glu Thr Ala Leu Asn His Val Arg Cys Ile Gly Leu Leu Asp Asp

165 170 175

Gly Phe Gly Ser Thr Pro Glu Met Thr Arg Arg Asp Leu Ile Trp Val

180 185 190

Val Thr Arg Met Gln Val Leu Val Asp Arg Tyr Pro Ser Trp Gly Asp

195 200 205

Val Ile Glu Val Asp Ser Trp Val Thr Pro Ser Gly Lys Asn Gly Met

210 215 220

Lys Arg Glu Trp Phe Leu Arg Asp Cys Lys Thr Gly Glu Ile Leu Thr

225 230 235 240

Arg Ala Thr Ser Val Trp Val Met Met Asn Lys Arg Thr Arg Arg Leu

245 250 255

Ser Lys Ile Pro Glu Glu Val Arg Val Glu Ile Glu Pro Tyr Phe Val

260 265 270

Glu His Gly Val Leu Asp Glu Asp Ser Arg Lys Leu Pro Lys Leu Asn

275 280 285

Asp Asn Thr Ala Asn Tyr Ile Arg Arg Gly Leu Ala Pro Arg Trp Ser

290 295 300

Asp Leu Asp Val Asn Gln His Val Asn Asn Val Lys Tyr Ile Gly Trp

305 310 315 320

Ile Leu Glu Ser Val Pro Ser Ser Leu Leu Glu Ser His Glu Leu Tyr

325 330 335

Gly Met Thr Leu Glu Tyr Arg Lys Glu Cys Gly Lys Asp Gly Leu Leu

340 345 350

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

355 360 365

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

370 375 380

Ile Met Arg Gly Lys Thr Glu Trp Arg Pro Lys Arg Ala Ala Asn Thr
385 390 395 400

Thr Tyr Phe Gly Ser Val Asp Asp Ile Pro Pro Ala Asn Asn Ala
405 410 415

<210> 43

<211> 1260

<212> DNA

<213> Artificial Sequence

<220><223> synthetic construct

<400> 43

actagtatgg ctatcaagac gaacaggcag cctgtggaga agcctccgtt cacgatcggg 60

acgctgcgca aggccatccc cgcgcactgt ttcgagcgct cggcgcttcg tgggcgcgcc 120

gccaacgccc acaccgtgcc caagatcaac ggcaacaagg ccggcctgct gacccccatg 180

gagagcacca aggacgagga catcgtcgcg gcccccaccg tggcgcccaa gcgcaccttc 240

atcaaccagc tgcccgactg gtcgatgctg ctggccgcga tcaccaccat cttcctggcg 300

gccgagaagc agtggaccaa cctggactgg aagccccgcc gccccgacat gctggtcgac 360

ttcgacccct tctccctggg ccgcttcgtg caggacggcc tgatcttccg ccagaacttc 420

agcatccgct cgtacgagat cggcgcggac cgcaccgcct ccatcgagac cctgatgaac 480

cacctgcagg agaccgcgct gaaccacgtc cgctgcatcg gcctgctgga cgacggcttc 540

ggcagcaccc ccgagatgac ccgccgcgac ctgatctggg tggtcacccg catgcaggtc 600

ctggtggacc gctacccctc gtggggcgac gtgatcgagg tcgactcctg ggtgaccccc 660

agcggcaaga acggcatgaa gcgcgagtgg ttcctgcgcg actgcaagac cggcgagatc 720

ctgacccgcg ccacctcggt ctgggtgatg atgaacaagc gcacccgccg cctgtccaag 780

atccccgagg aggtccgcgt ggagatcgag ccctacttcg tcgagcacgg cgtgctggac 840

gaggactcgc gcaagctgcc caagctgaac gacaacaccg ccaactacat ccgccgcggc 900

ctggcgcccc gctggtccga cctggacgtc aaccagcacg tgaacaacgt caagtacatc 960

ggctggatcc tggagagcgt gcccagcagc ctgctggagt cgcacgagct gtacggcatg 1020

accctggagt accgcaagga gtgcggcaag gacggcctgc tgcagtccct gaccgccgtc 1080

gccagcgact acggcggcgg ctcgctggag gccggcgtgg agtgcgacca cctgctgcgc 1140

ctggaggacg gctccgagat catgcgcggc aagaccgagt ggcgccccaa gcgcgccgcg 1200

aacaccacct acttcggcag cgtcgacgac atccccccg ccaacaacgc gtgactcgag 1260

<210> 44

<211> 362

<212> PRT

<213> Artificial Sequence

<220><223> synthetic construct

<400> 44

Met Ala Thr Ala Ser Thr Phe Ser Ala Phe Asn Ala Arg Cys Gly Asp

1 5 10 15

Leu Arg Arg Ser Ala Gly Ser Gly Pro Arg Arg Pro Ala Arg Pro Leu

20 25 30

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

35 40 45

Phe Gly Val Ala Glu Lys Gln Trp Pro Met Leu Asp Arg Lys Ser Lys

50 55 60

Arg Pro Asp Met Leu Val Glu Pro Leu Gly Val Asp Arg Ile Val Tyr

65 70 75 80

Asp Gly Val Ser Phe Arg Gln Ser Phe Ser Ile Arg Ser Tyr Glu Ile

85 90 95

Gly Ala Asp Arg Thr Ala Ser Ile Glu Thr Leu Met Asn Met Phe Gln

100 105 110

Glu Thr Ser Leu Asn His Cys Lys Ile Ile Gly Leu Leu Asn Asp Gly

115 120 125

Phe Gly Arg Thr Pro Glu Met Cys Lys Arg Asp Leu Ile Trp Val Val

130 135 140

Thr Lys Met Gln Ile Glu Val Asn Arg Tyr Pro Thr Trp Gly Asp Thr

145 150 155 160

Ile Glu Val Asn Thr Trp Val Ser Ala Ser Gly Lys His Gly Met Gly

165 170 175

Arg Asp Trp Leu Ile Ser Asp Cys His Thr Gly Glu Ile Leu Ile Arg

180 185 190

Ala Thr Ser Val Trp Ala Met Met Asn Gln Lys Thr Arg Arg Leu Ser

195 200 205

Lys Ile Pro Tyr Glu Val Arg Gln Glu Ile Glu Pro Gln Phe Val Asp
210 215 220
Ser Ala Pro Val Ile Val Asp Asp Arg Lys Phe His Lys Leu Asp Leu
225 230 235 240
Lys Thr Gly Asp Ser Ile Cys Asn Gly Leu Thr Pro Arg Trp Thr Asp
245 250 255
Leu Asp Val Asn Gln His Val Asn Asn Val Lys Tyr Ile Gly Trp Ile

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

325 330 335
Thr Glu Trp Arg Pro Lys Asn Ala Gly Ala Lys Gly Ala Ile Leu Thr
340 345 350
Gly Lys Thr Ser Asn Gly Asn Ser Ile Ser
355 360

<210> 45
<211> 1101
<212> DNA
<213> Artificial Sequence
<220><223> synthetic construct
<400> 45

actagtatgg ccaccgcatc cactttctcg gcgttcaatg cccgctgcgg cgacctgcgt 60
cgctcggcgg gctccgggcc ccggcgccca gcgaggcccc tccccgtgcg cgggcgcgcc 120

agcatgctgc tgtcggcggt gaccacggtc ttcggcgtgg ccgagaagca gtggcccatg 180
ctggaccgca agtccaagcg ccccgacatg ctggtcgagc ccctgggcgt ggaccgcatc 240
gtctacgacg gcgtgagctt ccgccagtcg ttctccatcc gcagctacga gatcggcgcc 300
gaccgcaccg cctcgatcga gacgctgatg aacatgttcc aggagacctc cctgaaccac 360

tgcaagatca tcggcctgct gaacgacggc ttcggccgca cgcccgagat gtgcaagcgc 420

gacctgatct gggtcgtgac caagatgcag atcgaggtga accgctaccc cacgtggggc 480

gacaccatcg aggtcaacac gtgggtgagc gcctcgggca agcacggcat gggccgcgac 540

tggctgatct ccgactgcca caccggcgag atcctgatcc gcgcgacgag cgtctgggcg 600

atgatgaacc agaagacccg ccgcctgtcg aagatcccct acgaggtgcg ccaggagatc 660

gagccccagt tcgtcgactc cgcccccgtg atcgtggacg accgcaagtt ccacaagctg 720

gacctgaaga cgggcgacag catctgcaac ggcctgaccc cccgctggac ggacctggac 780

gtgaaccagc acgtcaacaa cgtgaagtac atcggctgga tcctgcagtc ggtccccacc 840

gaggtgttcg agacgcagga gctgtgcggc ctgaccctgg agtaccgccg cgagtgcggc 900

cgcgactccg tgctggagag cgtcacggcc atggaccccct cgaaggaggg cgaccgctcc 960

ctgtaccagc acctgctgcg cctggaggac ggcgcggaca tcgtgaaggg ccgcaccgag 1020

tggcgcccca agaacgccgg cgccaagggc gccatcctga cgggcaagac cagcaacggc 1080

aactcgatct cctgactcga g 1101

<210> 46

<211> 369

<212> PRT

<213> Artificial Sequence

<220><223> synthetic construct

<400> 46

Met Ala Thr Ala Ser Thr Phe Ser Ala Phe Asn Ala Arg Cys Gly Asp

1 5 10 15

Leu Arg Arg Ser Ala Gly Ser Gly Pro Arg Arg Pro Ala Arg Pro Leu

20 25 30

Pro Val Arg Gly Arg Ala Gln Leu Pro Asp Trp Ser Met Leu Leu Ala

35 40 45

Ala Ile Thr Thr Leu Phe Leu Ala Ala Glu Lys Gln Trp Met Met Leu

50 55 60

Asp Trp Lys Pro Lys Arg Pro Asp Met Leu Val Asp Pro Phe Gly Leu

65 70 75 80

Gly Arg Phe Val Gln Asp Gly Leu Val Phe Arg Asn Asn Phe Ser Ile

85 90 95

Arg Ser Tyr Glu Ile Gly Ala Asp Arg Thr Ala Ser Ile Glu Thr Leu
100 105 110
Met Asn His Leu Gln Glu Thr Ala Leu Asn His Val Lys Ser Val Gly
115 120 125
Leu Leu Glu Asp Gly Leu Gly Ser Thr Arg Glu Met Ser Leu Arg Asn
130 135 140
Leu Ile Trp Val Val Thr Lys Met Gln Val Ala Val Asp Arg Tyr Pro

145 150 155 160
Thr Trp Gly Asp Glu Val Gln Val Ser Ser Trp Ala Thr Ala Ile Gly
165 170 175
Lys Asn Gly Met Arg Arg Glu Trp Ile Val Thr Asp Phe Arg Thr Gly
180 185 190
Glu Thr Leu Leu Arg Ala Thr Ser Val Trp Val Met Met Asn Lys Leu
195 200 205
Thr Arg Arg Ile Ser Lys Ile Pro Glu Glu Val Trp His Glu Ile Gly

210 215 220
Pro Ser Phe Ile Asp Ala Pro Pro Leu Pro Thr Val Glu Asp Asp Gly
225 230 235 240
Arg Lys Leu Thr Arg Phe Asp Glu Ser Ser Ala Asp Phe Ile Arg Lys
245 250 255
Gly Leu Thr Pro Arg Trp Ser Asp Leu Asp Ile Asn Gln His Val Asn
260 265 270
Asn Val Lys Tyr Ile Gly Trp Leu Leu Glu Ser Ala Pro Pro Glu Ile

275 280 285
His Glu Ser His Glu Ile Ala Ser Leu Thr Leu Glu Tyr Arg Arg Glu
290 295 300
Cys Gly Arg Asp Ser Val Leu Asn Ser Ala Thr Lys Val Ser Asp Ser
305 310 315 320
Ser Gln Leu Gly Lys Ser Ala Val Glu Cys Asn His Leu Val Arg Leu
325 330 335
Gln Asn Gly Gly Glu Ile Val Lys Gly Arg Thr Val Trp Arg Pro Lys

340 345 350

Arg Pro Leu Tyr Asn Asp Gly Ala Val Val Asp Val Pro Ala Lys Thr

355 360 365

Ser

<210> 47

<211> 1122

<212> DNA

<213> Artificial Sequence

<220><223> synthetic construct

<400> 47

actagtatgg ccaccgcatc cactttctcg gcgttcaatg cccgctgcgg cgacctgcgt 60

cgctcggcgg gctccgggcc ccggcgccca gcgaggcccc tccccgtgcg cgggcgcgcc 120

cagctgcccg actggagcat gctgctggcc gcgatcacca ccctgttcct ggcggccgag 180

aagcagtgga tgatgctgga ctggaagccc aagcgccccg acatgctggt ggaccccttc 240

ggcctgggcc gcttcgtgca ggacggcctg gtgttccgca acaacttcag catccgcagc 300

tacgagatcg gcgcggaccg caccgccagc atcgagaccc tgatgaacca cctgcaggag 360

accgccctga accacgtgaa gagcgtgggc ctgctggagg acggcctggg cagcacccgc 420

gagatgagcc tgcgcaacct gatctgggtg gtgaccaaga tgcaggtggc ggtggaccgc 480

taccccacct ggggcgacga ggtgcaggtg agcagctggg cgaccgccat cggcaagaac 540

ggcatgcgcc gcgagtggat cgtgaccgac ttccgcaccg gcgagaccct gctgcgcgcc 600

accagcgtgt gggtgatgat gaacaagctg acccgccgca tcagcaagat ccccgaggag 660

gtgtggcacg agatcggccc cagcttcatc gacgcgcccc ccctgcccac cgtggaggac 720

gacggccgca agctgacccg cttcgacgag agcagcgccg acttcatccg caagggcctg 780

acccccccgct ggagcgacct ggacatcaac cagcacgtga acaacgtgaa gtacatcggc 840

tggctgctgg agagcgcgcc ccccgagatc cacgagagcc acgagatcgc cagcctgacc 900

ctggagtacc gccgcgagtg cggccgcgac agcgtgctga acagcgccac caaggtgagc 960

gacagcagcc agctgggcaa gagcgccgtg gagtgcaacc acctggtgcg cctgcagaac 1020

ggcggcgaga tcgtgaaggg ccgcaccgtg tggcgcccca agcgcccccct gtacaacgac 1080

ggcgccgtgg tggacgtgcc cgccaagacc agctgactcg ag 1122

<210> 48

<211> 120

<212> DNA

<213> Artificial Sequence

<220><223> synthetic construct

<400> 48

actagtatgg ctatcaagac gaacaggcag cctgtggaga agcctccgtt cacgatcggg 60

acgctgcgca aggccatccc cgcgcactgt ttcgagcgct cggcgcttcg tgggcgcgcc 120

<210> 49

<211> 120

<212> DNA

<213> Artificial Sequence

<220><223> synthetic construct

<400> 49

actagtatgg ccaccgcatc cactttctcg gcgttcaatg cccgctgcgg cgacctgcgt 60

cgctcggcgg gctccgggcc ccggcgccca gcgaggcccc tccccgtgcg cgggcgcgcc 120

<210> 50

<211> 6207

<212> DNA

<213> Artificial Sequence

<220><223> synthetic construct

<400> 50

gctcttcggc cgccgccact cctgctcgag cgcgcccgac tcgcgctccg cctgcgcccg 60

cgcgtgcgcc gccagcgcct tggccttttc gccgcgctcg tgcgcgtcgc tgatgtccat 120

caccaggtcc atgaggtctg ccttgcgccg gctgagccac tgcttcgtcc gggcggccaa 180

gaggagcatg agggaggact cctggtccag ggtcctgacg tggtcgcggc tctgggagcg 240

ggccagcatc atctggctct gccgcaccga ggccgcctcc aactggtcct ccagcagccg 300

cagtcgccgc cgaccctggc agaggaagac aggtgagggg ggtatgaatt gtacagaaca 360

accacgagcc ttgtctaggc agaatcccta ccagtcatgg ctttacctgg atgacggcct 420

gcgaacagct gtccagcgac cctcgctgcc gccgcttctc ccgcacgctt ctttccagca 480

ccgtgatggc gcgagccagc gccgcacgct ggcgctgcgc ttcgccgatc tgaggacagt 540

cggggaactc tgatcagtct aaaccccctt gcgcgttagt gttgccatcc tttgcagacc 600

ggtgagagcc gacttgttgt gcgccacccc ccacaccacc tcctcccaga ccaattctgt 660

caccttttg gcgaaggcat cggcctcggc ctgcagagag gacagcagtg cccagccgct 720

ggggggttggc ggatgcacgc tcaggtaccc tttcttgcgc tatgacactt ccagcaaaag 780

gtagggcggg ctgcgagacg gcttcccggc gctgcatgca acaccgatga tgcttcgacc 840

ccccgaagct ccttcggggc tgcatgggcg ctccgatgcc gctccagggc gagcgctgtt 900

taaatagcca ggcccccgat tgcaaagaca ttatagcgag ctaccaaagc catattcaaa 960

cacctagatc actaccactt ctacacaggc cactcgagct tgtgatcgca ctccgctaag 1020

ggggcgcctc ttcctcttcg tttcagtcac aacccgcaaa cggcgcgcca tgctgctgca 1080

ggccttcctg ttcctgctgg ccggcttcgc cgccaagatc agcgcctcca tgacgaacga 1140

gacgtccgac cgcccccctgg tgcacttcac ccccaacaag ggctggatga acgaccccaa 1200

cggcctgtgg tacgacgaga aggacgccaa gtggcacctg tacttccagt acaacccgaa 1260

cgacaccgtc tgggggacgc ccttgttctg gggccacgcc acgtccgacg acctgaccaa 1320

ctgggaggac cagcccatcg ccatcgcccc gaagcgcaac gactccggcg ccttctccgg 1380

ctccatggtg gtggactaca acaacacctc cggcttcttc aacgacacca tcgacccgcg 1440

ccagcgctgc gtggccatct ggacctacaa caccccggag tccgaggagc agtacatctc 1500

ctacagcctg gacggcggct acaccttcac cgagtaccag aagaaccccg tgctggccgc 1560

caactccacc cagttccgcg acccgaaggt cttctggtac gagccctccc agaagtggat 1620

catgaccgcg gccaagtccc aggactacaa gatcgagatc tactcctccg acgacctgaa 1680

gtcctggaag ctggagtccg cgttcgccaa cgagggcttc ctcggctacc agtacgagtg 1740

ccccggcctg atcgaggtcc ccaccgagca ggaccccagc aagtcctact gggtgatgtt 1800

catctccatc aaccccggcg ccccggccgg cggctccttc aaccagtact tcgtcggcag 1860

cttcaacggc acccacttcg aggccttcga caaccagtcc cgcgtggtgg acttcggcaa 1920

ggactactac gccctgcaga ccttcttcaa caccgacccg acctacggga gcgccctggg 1980

catcgcgtgg gcctccaact gggagtactc cgccttcgtg cccaccaacc cctggcgctc 2040

ctccatgtcc ctcgtgcgca agttctccct caacaccgag taccaggcca acccggagac 2100

ggagctgatc aacctgaagg ccgagccgat cctgaacatc agcaacgccg gcccctggag 2160

ccggttcgcc accaacacca cgttgacgaa ggccaacagc tacaacgtcg acctgtccaa 2220

cagcaccggc accctggagt tcgagctggt gtacgccgtc aacaccaccc agacgatctc 2280

caagtccgtg ttcgcggacc tctccctctg gttcaagggc ctggaggacc ccgaggagta 2340

cctccgcatg ggcttcgagg tgtccgcgtc ctccttcttc ctggaccgcg ggaacagcaa 2400

ggtgaagttc gtgaaggaga acccctactt caccaaccgc atgagcgtga acaaccagcc 2460

cttcaagagc gagaacgacc tgtcctacta caaggtgtac ggcttgctgg accagaacat 2520

cctggagctg tacttcaacg acggcgacgt cgtgtccacc aacacctact tcatgaccac 2580

cgggaacgcc ctgggctccg tgaacatgac gacgggggtg gacaacctgt tctacatcga 2640

caagttccag gtgcgcgagg tcaagtgaca attggcagca gcagctcgga tagtatcgac 2700

acactctgga cgctggtcgt gtgatggact gttgccgcca cacttgctgc cttgacctgt 2760

gaatatccct gccgctttta tcaaacagcc tcagtgtgtt tgatcttgtg tgtacgcgct 2820

tttgcgagtt gctagctgct tgtgctattt gcgaatacca cccccagcat ccccttccct 2880

cgtttcatat cgcttgcatc ccaaccgcaa cttatctacg ctgtcctgct atccctcagc 2940

gctgctcctg ctcctgctca ctgcccctcg cacagccttg gtttgggctc cgcctgtatt 3000

ctcctggtac tgcaacctgt aaaccagcac tgcaatgctg atgcacggga agtagtggga 3060

tgggaacaca aatggaggat cccgcgtctc gaacagagcg cgcagaggaa cgctgaaggt 3120

ctcgcctctg tcgcacctca gcgcggcata caccacaata accacctgac gaatgcgctt 3180

ggttcttcgt ccattagcga agcgtccggt tcacacacgt gccacgttgg cgaggtggca 3240

ggtgacaatg atcggtggag ctgatggtcg aaacgttcac agcctaggga tatcgaattc 3300

ctttcttgcg ctatgacact tccagcaaaa ggtagggcgg gctgcgagac ggcttcccgg 3360

cgctgcatgc aacaccgatg atgcttcgac cccccgaagc tccttcgggg ctgcatgggc 3420

gctccgatgc cgctccaggg cgagcgctgt ttaaatagcc aggccccccga ttgcaaagac 3480

attatagcga gctaccaaag ccatattcaa acacctagat cactaccact tctacacagg 3540

ccactcgagc ttgtgatcgc actccgctaa gggggcgcct cttcctcttc gtttcagtca 3600

caacccgcaa acactagtat ggccaccgca tccactttct cggcgttcaa tgcccgctgc 3660

ggcgacctgc gtcgctcggc gggctccggg ccccggcgcc cagcgaggcc cctccccgtg 3720

cgcgggcgcg cccagctgcc cgactggagc atgctgctgg ccgcgatcac caccctgttc 3780

ctggcggccg agaagcagtg gatgatgctg gactggaagc ccaagcgccc cgacatgctg 3840

gtggacccct tcggcctggg ccgcttcgtg caggacggcc tggtgttccg caacaacttc 3900

agcatccgca gctacgagat cggcgcggac cgcaccgcca gcatcgagac cctgatgaac 3960

cacctgcagg agaccgccct gaaccacgtg aagagcgtgg gcctgctgga ggacggcctg 4020

ggcagcaccc gcgagatgag cctgcgcaac ctgatctggg tggtgaccaa gatgcaggtg 4080

gcggtggacc gctaccccac ctggggcgac gaggtgcagg tgagcagctg ggcgaccgcc 4140

atcggcaaga acggcatgcg ccgcgagtgg atcgtgaccg acttccgcac cggcgagacc 4200

ctgctgcgcg ccaccagcgt gtgggtgatg atgaacaagc tgacccgccg catcagcaag 4260

atccccgagg aggtgtggca cgagatcggc cccagcttca tcgacgcgcc ccccctgccc 4320

accgtggagg acgacggccg caagctgacc cgcttcgacg agagcagcgc cgacttcatc 4380

cgcaagggcc tgacccccg ctggagcgac ctggacatca accagcacgt gaacaacgtg 4440

aagtacatcg gctggctgct ggagagcgcg ccccccgaga tccacgagag ccacgagatc 4500
gccagcctga ccctggagta ccgccgcgag tgcggccgcg acagcgtgct gaacagcgcc 4560
accaaggtga gcgacagcag ccagctgggc aagagcgccg tggagtgcaa ccacctggtg 4620
cgcctgcaga acggcggcga gatcgtgaag ggccgcaccg tgtggcgccc caagcgcccc 4680
ctgtacaacg acggcgccgt ggtggacgtg cccgccaaga ccagcgatga cgatgacaag 4740
ctgggatgac tcgagttaat taactcgagg cagcagcagc tcggatagta tcgacacact 4800
ctggacgctg gtcgtgtgat ggactgttgc cgccacactt gctgccttga cctgtgaata 4860

tccctgccgc ttttatcaaa cagcctcagt gtgtttgatc ttgtgtgtac gcgcttttgc 4920
gagttgctag ctgcttgtgc tatttgcgaa taccaccccc agcatcccct tccctcgttt 4980
catatcgctt gcatcccaac cgcaacttat ctacgctgtc ctgctatccc tcagcgctgc 5040
tcctgctcct gctcactgcc cctcgcacag ccttggtttg ggctccgcct gtattctcct 5100
ggtactgcaa cctgtaaacc agcactgcaa tgctgatgca cgggaagtag tgggatggga 5160
acacaaatgg aaagctgtag agctccttgt tttccagaag gagttgctcc ttgagccttt 5220
cattctcagc ctcgataacc tccaaagccg ctctaattgt ggaggggttt cgaatttaaa 5280

agcttggaat gttggttcgt gcgtctggaa caagcccaga cttgttgctc actgggaaaa 5340
ggaccatcag ctccaaaaaa cttgccgctc aaaccgcgta cctctgcttt cgcgcaatct 5400
gccctgttga aatcgccacc acattcatat tgtgacgctt gagcagtctg taattgcctc 5460
agaatgtgga atcatctgcc ccctgtgcga gcccatgcca ggcatgtcgc gggcgaggac 5520
acccgccact cgtacagcag accattatgc tacctcacaa tagttcataa cagtgaccat 5580
atttctcgaa gctccccaac gagcacctcc atgctctgag tggccacccc ccggccctgg 5640
tgcttgcgga gggcaggtca accggcatgg ggctaccgaa atccccgacc ggatcccacc 5700

acccccgcga tgggaagaat ctctccccgg gatgtgggcc caccaccagc acaacctgct 5760
ggcccaggcg agcgtcaaac cataccacac aaatatcctt ggcatcggcc ctgaattcct 5820
tctgccgctc tgctacccgg tgcttctgtc cgaagcaggg gttgctaggg atcgctccga 5880
gtccgcaaac ccttgtcgcg tggcggggct tgttcgagct tgttcgagct tgaagagcct 5940
ctagagtcga cctgcaggca tgcaagcttg gcgtaatcat ggtcatagct gtttcctgtg 6000
tgaaattgtt atccgctcac aattccacac aacatacgag ccggaagcat aaagtgtaaa 6060
gcctggggtg cctaatgagt gagctaactc acattaattg cgttgcgctc actgcccgct 6120

ttccagtcgg gaaacctgtc gtgccagctg cattaatgaa tcggccaacg cgcggggaga 6180

ggcggtttgc gtattgggcg ctcttcc 6207

<210> 51

<211> 5576

<212> DNA

<213> Artificial Sequence

<220><223> synthetic construct

<400> 51

gaattcgccc tcccgtgatc acacaggtgc cttgcgagcg tgatcacact attttggggg 60

tcctacagta ctgaaatggt gagaagtcgt actgaaatca aggatgaaca atgaaaatgg 120

tgctgtggtg gcttctcaaa ggtcaagaat cagtcgctcg cgtcaggaaa tcgcggcgtc 180

aaccagcgtg ggcgcggtca gtggccccgc actggtcacc atagcctctc ctgccacagt 240

agcgatcccc tgggcgttca ctctcagcag cggctgtact gcctcccaga ttttcttctt 300

ctggacctgc gggcgtgaga ggatgagcag ggtgggccaa gggctcaatc ctgaacggcc 360

ctcattcggt ttccaatccc acaacacata cccacagcag gtcagaccac gcattccacc 420

atgcgcacca ataacgtgtc cttacctgat tgggtgtggc aggctccgtg gacaggagtg 480

cctcgtcccc cgcccagacc cgctcccccg tcacggcggc gtccgggacc cgcagcggct 540

ccaccgcggt gtgatccgcg ttggcggcgc agagcagcat cccagccgat ttgaccccgc 600

gcatgctccg aggcttgagg ttggccagca ccaccacccg ccggccgaca aggtcctcca 660

gggtcacgtg ccggaccagg ccactcacga tggtgcgagg gccccccctcc tcgccgaggt 720

cgatctgctc gacgtacaga ctgcgacatg cgtggcgagt ggtcatcaga aggaagcagg 780

tgtgcagaag gggcacgtgg ttggtattga gagtagccaa agctttgtgc caatcagaaa 840

gtcaacgcag ctgcctgcct ggctcgcgta caattccttt cttgcgctat gacacttcca 900

gcaaaaggta gggcgggctg cgagacggct tcccggcgct gcatgcaaca ccgatgatgc 960

ttcgaccccc cgaagctcct tcggggctgc atgggcgctc cgatgccgct ccagggcgag 1020

cgctgtttaa atagccaggc ccccgattgc aaagacatta tagcgagcta ccaaagcata 1080

ttcaaacacc tagatcacta ccacttctac acaggccact cgagcttgtg atcgcactcc 1140

gctaaggggg cgcctcttcc cttcgtttca gtcacaaccc gcaaacggcg cgccatgctg 1200

ctgcaggcct tcctgttcct gctggccggc ttcgccgcca agatcagcgc ctccatgacg 1260

aacgagacgt ccgaccgccc cctggtgcac ttcaccccca acaagggctg gatgaacgac 1320

cccaacggcc tgtggtacga cgagaaggac gccaagtggc acctgtactt ccagtacaac 1380

ccgaacgaca ccgtctgggg gacgcccttg ttctggggcc acgccacgtc cgacgacctg 1440

accaactggg aggaccagcc catcgccatc gccccgaagc gcaacgactc cggcgccttc 1500

tccggctcca tggtggtgga ctacaacaac acctccggct tcttcaacga caccatcgac 1560
ccgcgccagc gctgcgtggc catctggacc tacaacaccc cggagtccga ggagcagtac 1620
atctcctaca gcctggacgg cggctacacc ttcaccgagt accagaagaa ccccgtgctg 1680
gccgccaact ccacccagtt ccgcgacccg aaggtcttct ggtacgagcc ctcccagaag 1740
tggatcatga ccgcggccaa gtcccaggac tacaagatcg agatctactc ctccgacgac 1800
ctgaagtcct ggaagctgga gtccgcgttc gccaacgagg gcttcctcgg ctaccagtac 1860
gagtgccccg gcctgatcga ggtccccacc gagcaggacc ccagcaagtc ctactgggtg 1920

atgttcatct ccatcaaccc cggcgccccg gccggcggct ccttcaacca gtacttcgtc 1980
ggcagcttca acggcaccca cttcgaggcc ttcgacaacc agtcccgcgt ggtggacttc 2040
ggcaaggact actacgccct gcagaccttc ttcaacaccg acccgaccta cgggagcgcc 2100
ctgggcatcg cgtgggcctc caactgggag tactccgcct tcgtgcccac caacccctgg 2160
cgctcctcca tgtccctcgt gcgcaagttc tccctcaaca ccgagtacca ggccaacccg 2220
gagacggagc tgatcaacct gaaggccgag ccgatcctga acatcagcaa cgccggcccc 2280
tggagccggt tcgccaccaa caccacgttg acgaaggcca acagctacaa cgtcgacctg 2340

tccaacagca ccggcaccct ggagttcgag ctggtgtacg ccgtcaacac cacccagacg 2400
atctccaagt ccgtgttcgc ggacctctcc ctctggttca agggcctgga ggaccccgag 2460
gagtacctcc gcatgggctt cgaggtgtcc gcgtcctcct tcttcctgga ccgcgggaac 2520
agcaaggtga agttcgtgaa ggagaacccc tacttcacca accgcatgag cgtgaacaac 2580
cagcccttca agagcgagaa cgacctgtcc tactacaagg tgtacggctt gctggaccag 2640
aacatcctgg agctgtactt caacgacggc gacgtcgtgt ccaccaacac ctacttcatg 2700
accaccggga acgccctggg ctccgtgaac atgacgacgg gggtggacaa cctgttctac 2760

atcgacaagt tccaggtgcg cgaggtcaag tgattaatta actcgaggca gcagcagctc 2820
ggatagtatc gacacactct ggacgctggt cgtgtgatgg actgttgccg ccacacttgc 2880
tgccttgacc tgtgaatatc cctgccgctt ttatcaaaca gcctcagtgt gtttgatctt 2940
gtgtgtacgc gcttttgcga gttgctagct gcttgtgcta tttgcgaata ccacccccag 3000
catcccctcc ctcgtttcat atcgcttgca tcccaaccgc aacttatcta cgctgtcctg 3060
ctatccctca gcgctgctcc tgctcctgct cactgcccct cgcacagcct tggtttgggc 3120
tccgcctgta ttctcctggt actgcaacct gtaaaccagc actgcaatgc tgatgcacgg 3180

gaagtagtgg gatgggaaca caaatggaaa gcttgagctc ctttcttgcg ctatgacact 3240

tccagcaaaa ggtagggcgg gctgcgagac ggcttcccgg cgctgcatgc aacaccgatg 3300

atgcttcgac cccccgaagc tccttcgggg ctgcatgggc gctccgatgc cgctccaggg 3360

cgagcgctgt ttaaatagcc aggccccccga ttgcaaagac attatagcga gctaccaaag 3420

ccatattcaa acacctagat cactaccact tctacacagg ccactcgagc ttgtgatcgc 3480

actccgctaa gggggcgcct cttcctcttc gtttcagtca caacccgcaa acactagtat 3540

gacgttcggg gtcgccctcc cggccatggg ccgcggtgtc tcccttcccc ggcccagggt 3600

cgcggtgcgc gcccagtcgg cgagtcaggt tttggagagc gggcgcgccc ccgactggtc 3660

catgctgttc gccgtgatca ccaccatctt ctccgccgcc gagaagcagt ggaccaacct 3720

ggagtggaag cccaagccca acccccccca gctgctggac gaccacttcg gcccccacgg 3780

cctggtgttc cgccgcacct tcgccatccg cagctacgag gtgggccccg accgctccac 3840

cagcatcgtg gccgtgatga accacctgca ggaggccgcc ctgaaccacg ccaagtccgt 3900

gggcatcctg ggcgacggct tcggcaccac cctggagatg tccaagcgcg acctgatctg 3960

ggtggtgaag cgcacccacg tggccgtgga gcgctacccc gcctggggcg acaccgtgga 4020

ggtggagtgc tgggtgggcg cctccggcaa caacggccgc cgccacgact tcctggtgcg 4080

cgactgcaag accggcgaga tcctgacccg ctgcacctcc ctgagcgtga tgatgaacac 4140

ccgcacccgc cgcctgagca agatccccga ggaggtgcgc ggcgagatcg gccccgcctt 4200

catcgacaac gtggccgtga aggacgagga gatcaagaag ccccagaagc tgaacgactc 4260

caccgccgac tacatccagg gcggcctgac cccccgctgg aacgacctgg acatcaacca 4320

gcacgtgaac aacatcaagt acgtggactg gatcctggag accgtgcccg acagcatctt 4380

cgagagccac cacatctcct ccttcaccat cgagtaccgc cgcgagtgca ccatggacag 4440

cgtgctgcag tccctgacca ccgtgagcgg cggctcctcc gaggccggcc tggtgtgcga 4500

gcacctgctg cagctggagg gcggcagcga ggtgctgcgc gccaagaccg agtggcgccc 4560

caagctgacc gactccttcc gcggcatcag cgtgatcccc gccgagtcca gcgtgatgga 4620

ctacaaggac cacgacggcg actacaagga ccacgacatc gactacaagg acgacgacga 4680

caagtgactc gaggcagcag cagctcggat agtatcgaca cactctggac gctggtcgtg 4740

tgatggactg ttgccgccac acttgctgcc ttgacctgtg aatatccctg ccgcttttat 4800

caaacagcct cagtgtgttt gatcttgtgt gtacgcgctt ttgcgagttg ctagctcttg 4860

tgctatttgc gaataccacc cccagcatcc ccttccctcg tttcatatcg cttgcatccc 4920

aaccgcaact tatctacgct gtcctgctat ccctcagcgc tgctcctgct cctgctcact 4980

gcccctcgca cagccttggt ttgggctccg cctgtattct cctggtactg caacctgtaa 5040

accagcactg caatgctgat gcacgggaag tagtgggatg ggaacacaaa tggaaagctg 5100

gtacccgtac ccatcagcat ccgggtgaat cttggcctcc aagatatggc caatcctcac 5160

atccagcttg gcaaaatcga ctagactgtc tgcaagtggg aatgtggagc acaaggttgc 5220

ttgtagcgat cgacagactg gtggggtaca ttgacaggtg ggcagcgccg catccatcgt 5280

gcctgacgcg agcgccgccg gttgctcgcc cgtgcctgcc gtcaaagagc ggcagagaaa 5340

tcgggaaccg aaaacgtcac attgcctgat gttgttacat gctggactag actttcttgg 5400

cgtgggtctg ctcctcgcca ggtgcgcgac gcctcggggc tgggtgcgag ggagccgtgc 5460

ggccacgcat ttgacaagac ccaaagctcg catctcagac ggtcaaccgt tcgtattata 5520

cattcaacat atggtacata cgcaaaaagc atgccaacga tgacataggc gaattc 5576

<210> 52

<211> 4674

<212> DNA

<213> Artificial Sequence

<220><223> synthetic construct

<400> 52

ggtacccgcc tgcaacgcaa gggcagccac agccgctccc acccgccgct gaaccgacac 60

gtgcttgggc gcctgccgcc tgcctgccgc atgcttgtgc tggtgaggct gggcagtgct 120

gccatgctga ttgaggcttg gttcatcggg tggaagctta tgtgtgtgct gggcttgcat 180

gccgggcaat gcgcatggtg gcaagagggc ggcagcactt gctggagctg ccgcggtgcc 240

tccaggtggt tcaatcgcgg cagccagagg gatttcagat gatcgcgcgt acaggttgag 300

cagcagtgtc agcaaaggta gcagtttgcc agaatgatcg gttcagctgt taatcaatgc 360

cagcaagaga aggggtcaag tgcaaacacg ggcatgccac agcacgggca ccggggagtg 420

gaatggcacc accaagtgtg tgcgagccag catcgccgcc tggctgtttc agctacaacg 480

gcaggagtca tccaacgtaa ccatgagctg atcaacactg caatcatcgg gcgggcgtga 540

tgcaagcatg cctggcgaag acacatggtg tgcggatgct gccggctgct gcctgctgcg 600

cacgccgttg agttggcagc aggctcagcc atgcactgga tggcagctgg gctgccactg 660

caatgtggtg gataggatgc aagtggagcg aataccaaac cctctggctg cttgctgggt 720

tgcatggcat cgcaccatca gcaggagcgc atgcgaaggg actggcccca tgcacgccat 780

gccaaaccgg agcgcaccga gtgtccacac tgtcaccagg cccgcaagct ttgcagaacc 840

atgctcatgg acgcatgtag cgctgacgtc ccttgacggc gctcctctcg ggtgtgggaa 900

acgcaatgca gcacaggcag cagaggcggc ggcagcagag cggcggcagc agcggcgggg 960

gccacccttc ttgcggggtc gcgccccagc cagcggtgat gcgctgatcc caaacgagtt 1020

cacattcatt tgcatgcctg gagaagcgag gctggggcct ttgggctggt gcagcccgca 1080

atggaatgcg ggaccgccag gctagcagca aaggcgcctc ccctactccg catcgatgtt 1140

ccatagtgca ttggactgca tttgggtggg gcggccggct gtttctttcg tgttgcaaaa 1200

cgcgccagct cagcaacctg tcccgtgggt cccccgtgcc gatgaaatcg tgtgcacgcc 1260

gatcagctga ttgcccggct cgcgaagtag gcgccctcct ttctgctcgc cctctctccg 1320

tcccgcctct agaatatcaa tgatcgagca ggacggcctc cacgccggct cccccgccgc 1380

ctgggtggag cgcctgttcg gctacgactg ggcccagcag accatcggct gctccgacgc 1440

cgccgtgttc cgcctgtccg cccagggccg ccccgtgctg ttcgtgaaga ccgacctgtc 1500

cggcgccctg aacgagctgc aggacgaggc cgcccgcctg tcctggctgg ccaccaccgg 1560

cgtgccctgc gccgccgtgc tggacgtggt gaccgaggcc ggccgcgact ggctgctgct 1620

gggcgaggtg cccggccagg acctgctgtc ctcccacctg gcccccgccg agaaggtgtc 1680

catcatggcc gacgccatgc gccgcctgca caccctggac cccgccacct gccccttcga 1740

ccaccaggcc aagcaccgca tcgagcgcgc ccgcacccgc atggaggccg gcctggtgga 1800

ccaggacgac ctggacgagg agcaccaggg cctggccccc gccgagctgt tcgcccgcct 1860

gaaggcccgc atgcccgacg gcgaggacct ggtggtgacc cacggcgacg cctgcctgcc 1920

caacatcatg gtggagaacg gccgcttctc cggcttcatc gactgcggcc gcctgggcgt 1980

ggccgaccgc taccaggaca tcgccctggc cacccgcgac atcgccgagg agctgggcgg 2040

cgagtgggcc gaccgcttcc tggtgctgta cggcatcgcc gcccccgact cccagcgcat 2100

cgccttctac cgcctgctgg acgagttctt ctgacaattg gcagcagcag ctcggatagt 2160

atcgacacac tctggacgct ggtcgtgtga tggactgttg ccgccacact tgctgccttg 2220

acctgtgaat atccctgccg cttttatcaa acagcctcag tgtgtttgat cttgtgtgta 2280

cgcgcttttg cgagttgcta gctgcttgtg ctatttgcga ataccacccc cagcatcccc 2340

ttccctcgtt tcatatcgct tgcatcccaa ccgcaactta tctacgctgt cctgctatcc 2400

ctcagcgctg ctcctgctcc tgctcactgc ccctcgcaca gccttggttt gggctccgcc 2460

tgtattctcc tggtactgca acctgtaaac cagcactgca atgctgatgc acgggaagta 2520

gtgggatggg aacacaaatg gaggatcccg cgtctcgaac agagcgcgca gaggaacgct 2580

gaaggtctcg cctctgtcgc acctcagcgc ggcatacacc acaataacca cctgacgaat 2640

gcgcttggtt cttcgtccat tagcgaagcg tccggttcac acacgtgcca cgttggcgag 2700

gtggcaggtg acaatgatcg gtggagctga tggtcgaaac gttcacagcc tagggatatc 2760

gaattccttt cttgcgctat gacacttcca gcaaaaggta gggcgggctg cgagacggct 2820

tcccggcgct gcatgcaaca ccgatgatgc ttcgaccccc cgaagctcct tcggggctgc 2880

atgggcgctc cgatgccgct ccagggcgag cgctgtttaa atagccaggc ccccgattgc 2940

aaagacatta tagcgagcta ccaaagccat attcaaacac ctagatcact accacttcta 3000

cacaggccac tcgagcttgt gatcgcactc cgctaagggg gcgcctcttc ctcttcgttt 3060

cagtcacaac ccgcaaacac tagtatggcc accgcatcca ctttctcggc gttcaatgcc 3120

cgctgcggcg acctgcgtcg ctcggcgggc tccgggcccc ggcgcccagc gaggcccctc 3180

cccgtgcgcg ggcgcgccca gctgcccgac tggagccgcc tgctgaccgc catcaccacc 3240

gtgttcgtga agtccaagcg ccccgacatg cacgaccgca agtccaagcg ccccgacatg 3300

ctggtggaca gcttcggcct ggagtccacc gtgcaggacg gcctggtgtt ccgccagtcc 3360

ttctccatcc gctcctacga gatcggcacc gaccgcaccg ccagcatcga gaccctgatg 3420

aaccacctgc aggagacctc cctgaaccac tgcaagagca ccggcatcct gctggacggc 3480

ttcggccgca ccctggagat gtgcaagcgc gacctgatct gggtggtgat caagatgcag 3540

atcaaggtga accgctaccc cgcctggggc gacaccgtgg agatcaacac ccgcttcagc 3600

cgcctgggca agatcggcat gggccgcgac tggctgatct ccgactgcaa caccggcgag 3660

atcctggtgc gcgccaccag cgcctacgcc atgatgaacc agaagacccg ccgcctgtcc 3720

aagctgccct acgaggtgca ccaggagatc gtgcccctgt tcgtggacag ccccgtgatc 3780

gaggactccg acctgaaggt gcacaagttc aaggtgaaga ccggcgacag catccagaag 3840

ggcctgaccc ccggctggaa cgacctggac gtgaaccagc acgtgtccaa cgtgaagtac 3900

atcggctgga tcctggagag catgcccacc gaggtgctgg agacccagga gctgtgctcc 3960

ctggccctgg agtaccgccg cgagtgcggc cgcgactccg tgctggagag cgtgaccgcc 4020

atggacccca gcaaggtggg cgtgcgctcc cagtaccagc acctgctgcg cctggaggac 4080

ggcaccgcca tcgtgaacgg cgccaccgag tggcgcccca agaacgccgg cgccaacggc 4140

gccatctcca ccggcaagac cagcaacggc aactccgtgt ccatggacta caaggaccac 4200

gacggcgact acaaggacca cgacatcgac tacaaggacg acgacgacaa gtgactcgag 4260

gcagcagcag ctcggatagt atcgacacac tctggacgct ggtcgtgtga tggactgttg 4320

ccgccacact tgctgccttg acctgtgaat atccctgccg cttttatcaa acagcctcag 4380

tgtgtttgat cttgtgtgta cgcgcttttg cgagttgcta gctgcttgtg ctatttgcga 4440

ataccacccc cagcatcccc ttccctcgtt tcatatcgct tgcatcccaa ccgcaactta 4500

tctacgctgt cctgctatcc ctcagcgctg ctcctgctcc tgctcactgc ccctcgcaca 4560

gccttggttt gggctccgcc tgtattctcc tggtactgca acctgtaaac cagcactgca 4620

atgctgatgc acgggaagta gtgggatggg aacacaaatg gaaagcttga gctc 4674

<210> 53

<211> 4656

<212> DNA

<213> Artificial Sequence

<220><223> synthetic construct

<400> 53

ggtacccgcc tgcaacgcaa gggcagccac agccgctccc acccgccgct gaaccgacac 60

gtgcttgggc gcctgccgcc tgcctgccgc atgcttgtgc tggtgaggct gggcagtgct 120

gccatgctga ttgaggcttg gttcatcggg tggaagctta tgtgtgtgct gggcttgcat 180

gccgggcaat gcgcatggtg gcaagagggc ggcagcactt gctggagctg ccgcggtgcc 240

tccaggtggt tcaatcgcgg cagccagagg gatttcagat gatcgcgcgt acaggttgag 300

cagcagtgtc agcaaaggta gcagtttgcc agaatgatcg gttcagctgt taatcaatgc 360

cagcaagaga aggggtcaag tgcaaacacg ggcatgccac agcacgggca ccggggagtg 420

gaatggcacc accaagtgtg tgcgagccag catcgccgcc tggctgtttc agctacaacg 480

gcaggagtca tccaacgtaa ccatgagctg atcaacactg caatcatcgg gcgggcgtga 540

tgcaagcatg cctggcgaag acacatggtg tgcggatgct gccggctgct gcctgctgcg 600

cacgccgttg agttggcagc aggctcagcc atgcactgga tggcagctgg gctgccactg 660

caatgtggtg gataggatgc aagtggagcg aataccaaac cctctggctg cttgctgggt 720

tgcatggcat cgcaccatca gcaggagcgc atgcgaaggg actggcccca tgcacgccat 780

gccaaaccgg agcgcaccga gtgtccacac tgtcaccagg cccgcaagct ttgcagaacc 840

atgctcatgg acgcatgtag cgctgacgtc ccttgacggc gctcctctcg ggtgtgggaa 900

acgcaatgca gcacaggcag cagaggcggc ggcagcagag cggcggcagc agcggcgggg 960

gccacccttc ttgcggggtc gcgccccagc cagcggtgat gcgctgatcc caaacgagtt 1020

cacattcatt tgcatgcctg gagaagcgag gctggggcct ttgggctggt gcagcccgca 1080

atggaatgcg ggaccgccag gctagcagca aaggcgcctc ccctactccg catcgatgtt 1140

ccatagtgca ttggactgca tttgggtggg gcggccggct gtttctttcg tgttgcaaaa 1200

cgcgccagct cagcaacctg tcccgtgggt cccccgtgcc gatgaaatcg tgtgcacgcc 1260

gatcagctga ttgcccggct cgcgaagtag gcgccctcct ttctgctcgc cctctctccg 1320

tcccgcctct agaatatcaa tgatcgagca ggacggcctc cacgccggct cccccgccgc 1380

ctgggtggag cgcctgttcg gctacgactg ggcccagcag accatcggct gctccgacgc 1440

cgccgtgttc cgcctgtccg cccagggccg ccccgtgctg ttcgtgaaga ccgacctgtc 1500

cggcgccctg aacgagctgc aggacgaggc cgcccgcctg tcctggctgg ccaccaccgg 1560

cgtgccctgc gccgccgtgc tggacgtggt gaccgaggcc ggccgcgact ggctgctgct 1620

gggcgaggtg cccggccagg acctgctgtc ctcccacctg gcccccgccg agaaggtgtc 1680

catcatggcc gacgccatgc gccgcctgca caccctggac cccgccacct gccccttcga 1740

ccaccaggcc aagcaccgca tcgagcgcgc ccgcacccgc atggaggccg gcctggtgga 1800

ccaggacgac ctggacgagg agcaccaggg cctggccccc gccgagctgt tcgcccgcct 1860

gaaggcccgc atgcccgacg gcgaggacct ggtggtgacc cacggcgacg cctgcctgcc 1920

caacatcatg gtggagaacg gccgcttctc cggcttcatc gactgcggcc gcctgggcgt 1980

ggccgaccgc taccaggaca tcgccctggc cacccgcgac atcgccgagg agctgggcgg 2040

cgagtgggcc gaccgcttcc tggtgctgta cggcatcgcc gcccccgact cccagcgcat 2100

cgccttctac cgcctgctgg acgagttctt ctgacaattg gcagcagcag ctcggatagt 2160

atcgacacac tctggacgct ggtcgtgtga tggactgttg ccgccacact tgctgccttg 2220

acctgtgaat atccctgccg cttttatcaa acagcctcag tgtgtttgat cttgtgtgta 2280

cgcgcttttg cgagttgcta gctgcttgtg ctatttgcga ataccacccc cagcatcccc 2340

ttccctcgtt tcatatcgct tgcatcccaa ccgcaactta tctacgctgt cctgctatcc 2400

ctcagcgctg ctcctgctcc tgctcactgc ccctcgcaca gccttggttt gggctccgcc 2460

tgtattctcc tggtactgca acctgtaaac cagcactgca atgctgatgc acgggaagta 2520

gtgggatggg aacacaaatg gaggatcccg cgtctcgaac agagcgcgca gaggaacgct 2580

gaaggtctcg cctctgtcgc acctcagcgc ggcatacacc acaataacca cctgacgaat 2640

gcgcttggtt cttcgtccat tagcgaagcg tccggttcac acacgtgcca cgttggcgag 2700

gtggcaggtg acaatgatcg gtggagctga tggtcgaaac gttcacagcc tagggatatc 2760

gaattccttt cttgcgctat gacacttcca gcaaaaggta gggcgggctg cgagacggct 2820

tcccggcgct gcatgcaaca ccgatgatgc ttcgaccccc cgaagctcct tcggggctgc 2880

atgggcgctc cgatgccgct ccagggcgag cgctgtttaa atagccaggc ccccgattgc 2940

aaagacatta tagcgagcta ccaaagccat attcaaacac ctagatcact accacttcta 3000

cacaggccac tcgagcttgt gatcgcactc cgctaagggg gcgcctcttc ctcttcgttt 3060

cagtcacaac ccgcaaacac tagtatggcc accgcatcca ctttctcggc gttcaatgcc 3120

cgctgcggcg acctgcgtcg ctcggcgggc tccgggcccc ggcgcccagc gaggcccctc 3180

cccgtgcgcg ggcgcgcccc cgactggtcc atgctgttcg ccgtgatcac caccatcttc 3240

agcgccgccg agaagcagtg gaccaacctg gagtggaagc ccaagcccaa gctgccccag 3300

ctgctggacg accacttcgg cctgcacggc ctggtgttcc gccgcacctt cgccatccgc 3360
tcctacgagg tgggccccga ccgcagcacc tccatcctgg ccgtgatgaa ccacatgcag 3420
gaggccaccc tgaaccacgc caagagcgtg ggcatcctgg gcgacggctt cggcaccacc 3480
ctggagatgt ccaagcgcga cctgatgtgg gtggtgcgcc gcacccacgt ggccgtggag 3540
cgctacccca cctggggcga caccgtggag gtggagtgct ggatcggcgc cagcggcaac 3600
aacggcatgc gccgcgactt cctggtgcgc gactgcaaga ccggcgagat cctgacccgc 3660
tgcacctccc tgagcgtgct gatgaacacc cgcacccgcc gcctgagcac catccccgac 3720

gaggtgcgcg gcgagatcgg ccccgccttc atcgacaacg tggccgtgaa ggacgacgag 3780
atcaagaagc tgcagaagct gaacgactcc accgccgact acatccaggg cggcctgacc 3840
ccccgctgga acgacctgga cgtgaaccag cacgtgaaca acctgaagta cgtggcctgg 3900
gtgttcgaga ccgtgcccga cagcatcttc gagtcccacc acatcagctc cttcaccctg 3960
gagtaccgcc gcgagtgcac ccgcgactcc gtgctgcgca gcctgaccac cgtgagcggc 4020
ggcagctccg aggccggcct ggtgtgcgac cacctgctgc agctggaggg cggcagcgag 4080
gtgctgcgcg cccgcaccga gtggcgcccc aagctgaccg actccttccg cggcatcagc 4140

gtgatccccg ccgagccccg cgtgatggac tacaaggacc acgacggcga ctacaaggac 4200
cacgacatcg actacaagga cgacgacgac aagtgactcg aggcagcagc agctcggata 4260
gtatcgacac actctggacg ctggtcgtgt gatggactgt tgccgccaca cttgctgcct 4320
tgacctgtga atatccctgc cgcttttatc aaacagcctc agtgtgtttg atcttgtgtg 4380
tacgcgcttt tgcgagttgc tagctgcttg tgctatttgc gaataccacc cccagcatcc 4440
ccttccctcg tttcatatcg cttgcatccc aaccgcaact tatctacgct gtcctgctat 4500
ccctcagcgc tgctcctgct cctgctcact gcccctcgca cagccttggt ttgggctccg 4560

cctgtattct cctggtactg caacctgtaa accagcactg caatgctgat gcacgggaag 4620
tagtgggatg ggaacacaaa tggaaagctt gagctc 4656
<210> 54
<211> 4635
<212> DNA
<213> Artificial Sequence
<220><223> synthetic construct
<400> 54
ggtacccgcc tgcaacgcaa gggcagccac agccgctccc acccgccgct gaaccgacac 60

gtgcttgggc gcctgccgcc tgcctgccgc atgcttgtgc tggtgaggct gggcagtgct 120

gccatgctga ttgaggcttg gttcatcggg tggaagctta tgtgtgtgct gggcttgcat 180

gccgggcaat gcgcatggtg gcaagagggc ggcagcactt gctggagctg ccgcggtgcc 240

tccaggtggt tcaatcgcgg cagccagagg gatttcagat gatcgcgcgt acaggttgag 300

cagcagtgtc agcaaaggta gcagtttgcc agaatgatcg gttcagctgt taatcaatgc 360

cagcaagaga aggggtcaag tgcaaacacg ggcatgccac agcacgggca ccggggagtg 420

gaatggcacc accaagtgtg tgcgagccag catcgccgcc tggctgtttc agctacaacg 480

gcaggagtca tccaacgtaa ccatgagctg atcaacactg caatcatcgg gcgggcgtga 540

tgcaagcatg cctggcgaag acacatggtg tgcggatgct gccggctgct gcctgctgcg 600

cacgccgttg agttggcagc aggctcagcc atgcactgga tggcagctgg gctgccactg 660

caatgtggtg gataggatgc aagtggagcg aataccaaac cctctggctg cttgctgggt 720

tgcatggcat cgcaccatca gcaggagcgc atgcgaaggg actggcccca tgcacgccat 780

gccaaaccgg agcgcaccga gtgtccacac tgtcaccagg cccgcaagct ttgcagaacc 840

atgctcatgg acgcatgtag cgctgacgtc ccttgacggc gctcctctcg ggtgtgggaa 900

acgcaatgca gcacaggcag cagaggcggc ggcagcagag cggcggcagc agcggcgggg 960

gccacccttc ttgcggggtc gcgccccagc cagcggtgat gcgctgatcc caaacgagtt 1020

cacattcatt tgcatgcctg gagaagcgag gctggggcct ttgggctggt gcagcccgca 1080

atggaatgcg ggaccgccag gctagcagca aaggcgcctc ccctactccg catcgatgtt 1140

ccatagtgca ttggactgca tttgggtggg gcggccggct gtttctttcg tgttgcaaaa 1200

cgcgccagct cagcaacctg tcccgtgggt cccccgtgcc gatgaaatcg tgtgcacgcc 1260

gatcagctga ttgcccggct cgcgaagtag gcgccctcct ttctgctcgc cctctctccg 1320

tcccgcctct agaatatcaa tgatcgagca ggacggcctc cacgccggct cccccgccgc 1380

ctgggtggag cgcctgttcg gctacgactg ggcccagcag accatcggct gctccgacgc 1440

cgccgtgttc cgcctgtccg cccagggccg ccccgtgctg ttcgtgaaga ccgacctgtc 1500

cggcgccctg aacgagctgc aggacgaggc cgcccgcctg tcctggctgg ccaccaccgg 1560

cgtgccctgc gccgccgtgc tggacgtggt gaccgaggcc ggccgcgact ggctgctgct 1620

gggcgaggtg cccggccagg acctgctgtc ctcccacctg gcccccgccg agaaggtgtc 1680

catcatggcc gacgccatgc gccgcctgca caccctggac cccgccacct gcccccttcga 1740

ccaccaggcc aagcaccgca tcgagcgcgc ccgcacccgc atggaggccg gcctggtgga 1800

ccaggacgac ctggacgagg agcaccaggg cctggcccccc gccgagctgt tcgcccgcct 1860

gaaggcccgc atgcccgacg gcgaggacct ggtggtgacc cacggcgacg cctgcctgcc 1920

caacatcatg gtggagaacg gccgcttctc cggcttcatc gactgcggcc gcctgggcgt 1980
ggccgaccgc taccaggaca tcgccctggc cacccgcgac atcgccgagg agctgggcgg 2040
cgagtgggcc gaccgcttcc tggtgctgta cggcatcgcc gcccccgact cccagcgcat 2100
cgccttctac cgcctgctgg acgagttctt ctgacaattg gcagcagcag ctcggatagt 2160
atcgacacac tctggacgct ggtcgtgtga tggactgttg ccgccacact tgctgccttg 2220
acctgtgaat atccctgccg cttttatcaa acagcctcag tgtgtttgat cttgtgtgta 2280
cgcgcttttg cgagttgcta gctgcttgtg ctatttgcga ataccacccc cagcatcccc 2340

ttccctcgtt tcatatcgct tgcatcccaa ccgcaactta tctacgctgt cctgctatcc 2400
ctcagcgctg ctcctgctcc tgctcactgc ccctcgcaca gccttggttt gggctccgcc 2460
tgtattctcc tggtactgca acctgtaaac cagcactgca atgctgatgc acgggaagta 2520
gtgggatggg aacacaaatg gaggatcccg cgtctcgaac agagcgcgca gaggaacgct 2580
gaaggtctcg cctctgtcgc acctcagcgc ggcatacacc acaataacca cctgacgaat 2640
gcgcttggtt cttcgtccat tagcgaagcg tccggttcac acacgtgcca cgttggcgag 2700
gtggcaggtg acaatgatcg gtggagctga tggtcgaaac gttcacagcc tagggatatc 2760

gaattccttt cttgcgctat gacacttcca gcaaaaggta gggcgggctg cgagacggct 2820
tcccggcgct gcatgcaaca ccgatgatgc ttcgaccccc cgaagctcct tcggggctgc 2880
atgggcgctc cgatgccgct ccagggcgag cgctgtttaa atagccaggc ccccgattgc 2940
aaagacatta tagcgagcta ccaaagccat attcaaacac ctagatcact accacttcta 3000
cacaggccac tcgagcttgt gatcgcactc cgctaagggg gcgcctcttc ctcttcgttt 3060
cagtcacaac ccgcaaacac tagtatggcc accgcatcca ctttctcggc gttcaatgcc 3120
cgctgcggcg acctgcgtcg ctcggcgggc tccgggcccc ggcgcccagc gaggcccctc 3180

cccgtgcgcg ggcgcgccca gctgcccgac tggagcatgc tgctggccgc gatcaccacc 3240
ctgttcctgg cggccgagaa gcagtggatg atgctggact ggaagcccaa gcgccccgac 3300
atgctggtgg acccccttcgg cctgggccgc ttcgtgcagg acggcctggt gttccgcaac 3360
aacttcagca tccgcagcta cgagatcggc gcggaccgca ccgccagcat cgagaccctg 3420
atgaaccacc tgcaggagac cgccctgaac cacgtgaaga gcgtgggcct gctggaggac 3480
ggcctgggca gcacccgcga gatgagcctg cgcaacctga tctgggtggt gaccaagatg 3540
caggtggcgg tggaccgcta ccccacctgg ggcgacgagg tgcaggtgag cagctgggcg 3600

accgccatcg gcaagaacgg catgcgccgc gagtggatcg tgaccgactt ccgcaccggc 3660

gagaccctgc tgcgcgccac cagcgtgtgg gtgatgatga acaagctgac ccgccgcatc 3720

agcaagatcc ccgaggaggt gtggcacgag atcggcccca gcttcatcga cgcgcccccc 3780

ctgcccaccg tggaggacga cggccgcaag ctgacccgct tcgacgagag cagcgccgac 3840

ttcatccgca agggcctgac cccccgctgg agcgacctgg acatcaacca gcacgtgaac 3900

aacgtgaagt acatcggctg gctgctggag agcgcgcccc ccgagatcca cgagagccac 3960

gagatcgcca gcctgaccct ggagtaccgc cgcgagtgcg gccgcgacag cgtgctgaac 4020

agcgccacca aggtgagcga cagcagccag ctgggcaaga gcgccgtgga gtgcaaccac 4080

ctggtgcgcc tgcagaacgg cggcgagatc gtgaagggcc gcaccgtgtg gcgccccaag 4140

cgccccctgt acaacgacgg cgccgtggtg gacgtgcccg ccaagaccag cgatgacgat 4200

gacaagctgg gatgactcga ggcagcagca gctcggatag tatcgacaca ctctggacgc 4260

tggtcgtgtg atggactgtt gccgccacac ttgctgcctt gacctgtgaa tatccctgcc 4320

gcttttatca aacagcctca gtgtgtttga tcttgtgtgt acgcgctttt gcgagttgct 4380

agctgcttgt gctatttgcg aataccaccc ccagcatccc cttccctcgt ttcatatcgc 4440

ttgcatccca accgcaactt atctacgctg tcctgctatc cctcagcgct gctcctgctc 4500

ctgctcactg cccctcgcac agccttggtt tgggctccgc ctgtattctc ctggtactgc 4560

aacctgtaaa ccagcactgc aatgctgatg cacgggaagt agtgggatgg gaacacaaat 4620

ggaaagcttg agctc 4635

<210> 55

<211> 997

<212> DNA

<213> Prototheca moriformis

<400> 55

cctgtcgatc gaagagaagg agacatgtgt acattattgg tgtgagggcg ctgaatcggc 60

catttttaa aatgatcacg ctcatgccaa tagacgcggc acataacgac gttcaaaccc 120

ccgccaaagc cgcggacaac cccatccctc cacacccccc acacaaagaa cccgccaccg 180

cttaccttgc ccacgaggta ggcctttcgt tgcgcaaaac cggcctcggt gatgaatgca 240

tgcccgttcc tgacgagcgc tgcccgggcc aacacgctct tttgctgcgt ctcctcaggc 300

ttgggggcct ccttgggctt gggtgccgcc atgatctgcg cgcatcagag aaacgttgct 360

ggtaaaaagg agcgcccggc tgcgcaatat atatataggc atgccaacac agcccaacct 420

cactcgggag cccgtcccac cacccccaag tcgcgtgcct tgacggcata ctgctgcaga 480

agcttcatga gaatgatgcc gaacaagagg ggcacgagga cccaatcccg gacatccttg 540

tcgataatga tctcgtgagt ccccatcgtc cgcccgacgc tccggggagc ccgccgatgc 600
tcaagacgag agggccctcg accaggaggg gctggcccgg gcgggcactg gcgtcgaagg 660
tgcgcccgtc gttcgcctgc agtcctatgc cacaaaacaa gtcttctgac ggggtgcgtt 720
tgctcccgtg cgggcaggca acagaggtat tcaccctggt catggggaga tcggcgatcg 780
agctgggata agagatactt ctggcaagca atgacaactt gtcaggaccg gaccgtgcca 840
tatatttctc acctagcgcc gcaaaaccta acaatttggg agtcactgtg ccactgagtt 900
cgactggtag ctgaatggag tcgctgctcc actaaacgaa ttgtcagcac cgccagccgg 960

ccgaggaccc gagtcatagc gagggtagta gcgcgcc 997

<210> 56
<211> 753
<212> DNA
<213> Prototheca moriformis
<400> 56

actaattgca atcgtgcagt aatcatcgat atggtcacaa gtagatcccc tactgacacc 60
ctctcgtaca tgtaggcaat gtcatcggcg ccgtcctgct gaccgatgcc gacgtagcag 120
agcagacccg ggccgatctg ggatacgagc cggccctcca cctgcgctcg aggtggaatc 180
aagtaaataa ccaatacact tttcgacacc acacagagtt gcacggacgg tggcgtacct 240
ctacgctcgc gctcttcacg cgctggacga ccgcacgcat gagcccgggt ggcttggtct 300

gggctgcaaa aatgcacaac aaacaagtat cagacgctca tggatgcaca cgcgctccca 360
agcacgctca gactaaatat tacagtagct cgtatctgat aagatatcga gacataccgc 420
tcaactcacc cgcaaactgc gccccgccag gtgatgcgca cagggcccca ccatgcgatc 480
catcgcatcg ctcctcgagg gcgctatcac gtggccggag agcgttcaca gcgtacgcca 540
ctgtatctgg gcggtatgcg gtccgtcaac atggagacag atacccgcac caccaccttg 600
caagctcttc catattggaa gtagaaaatt gtaattgtat catcgcacga ggggccaact 660
tgccgtcggc gagctgggcg acgaacacca cctggacgtt gtcgagactc gctcgtgccg 720

tgcgccgggc cgctgggtat ccagaccgtc gcc 753

<210> 57
<211> 1122
<212> DNA
<213> Prototheca moriformis

<400> 57

caacgacaac cagcaggcaa ctcggtcagc gacccaacac gcgagtcaaa ttgttgcgtg 60

ttcttgcctt gtctatttac tgtgatagca agactgtcgg tcagtcaata ccgcggtgcg 120

cacgtcgggg tgccaagcct agcagagcac gggacggctg gtgctgtgcg ccagctcagc 180

tcgcttcgcg accaattgta ggaccggcaa agtcaccaaa acatgccagc ggtgcgattc 240

aattggtcat gagctctaca aaattgtttt gtgcgtcgcg caggtatcca acggcgcggc 300

agagaaagtt tgacagctct cgatttcatc tcggaaaaat ggggagaatt tatgacacac 360

aagtgcgcag gcggcccagg cggccagcat attctggcgt gacctgggcc gcccacaaaa 420

tgcttggatg cactctaaaa taattatatt tgccatgaac aagggaagag ttaccgcacc 480

cagccctaga cttgggcgcc cgagcaaggt tacgtcaagc caccttcgcc catcgcccaa 540

ctccgtattc cccgacagcc gcacgtggcc ctcgccggaa tgaaccctga atcggcatca 600

cgccacgcgt tcgccaatcg ttccgctctc tggcttcatc ggcctgcgcc ttcacgtcgt 660

ggtcacgaca gtgcattcat acttccattt gcacctcggc acacactttt acgcatcgcc 720

taccttgct gcggcagtct agggtcactt tgcagccatg ggacagtgct acaccaccgt 780

cggtgcgcaa agctatttca agtgaaccgt gggcggaaaa aaggaatgta cactgtctca 840

accgactcct acaattgttt accatgcaga tcagagctcg acggccatca tcgagcaggt 900

gtggggcctt ggtggcgcgg cgcggggccc cagggcgtcg caggcattga tggcactctg 960

agactttcgc acgcgcatga gggaccccat caagagaaga gtgtgtcttt atgtccccat 1020

tcatgatgat gtatcttgtg attgtcgcag tttggcaagt ttaaccggat cgccgctcca 1080

ggtgtggcgt ggcggatttt tctaggggtg cttgagcagt cg 1122

<210> 58

<211> 574

<212> DNA

<213> Prototheca moriformis

<400> 58

ggcccagggc cctgcggatg gcccacacca gatctagcct ctcttatgcc atgcccgcct 60

cgctgcccgt cgtatccccc cgccgatccg cgcgtagggg accgcggcct gacccacgcc 120

acgaaagagc tttgctcctc aatttctcgc caacagaacc gtatcaaacg ctcaacgcct 180

atcccgaaca atccgtattc acaccaaatc gagtataccg gactggtttg cctagtcttg 240

aaggaaatga tcccgtccat gctcggaagg gggagcgggc ggaggatcct actcatctct 300

gaaatgggat tggtccgaag atgggttggg caagcacgtg ccaaacccca gcgagttgct 360

gacgagcagg ctcatccaat cccccggcga atcctccctc acgccccgca tgcatacaag 420

tccctcccac acgccccctc ccatccattt tcgcctggtc cgaacgcgag cggcgtcgag 480

gcggaccact tgctccgcag cgccgtctgg gtctccaccc cacagcggct ttgctgccag 540

aggcaccccc cttgccccac ctcctcttgc agcc 574

<210> 59

<211> 1096

<212> DNA

<213> Prototheca moriformis

<400> 59

ccaggcaggc ggtagggttg ccgattgctt gagcgaattg gaagatataa ttttttgtgg 60

tgtccctgga cgctgtttgt ggcgctcctt tttggagaag attgcgtggg ggagctttcc 120

atgtaccacg cttccttctg aaaggattct ggccgagtcc tgatgagccc aaagaaaaca 180

cctgcctttc agtgctggca ctctgaaaac gtcaacagat gattatacat gtcacaaaag 240

gcagccgatt aggaacggga gctctggccg ttcgtttggc tgcctgggct gattgaagtg 300

atccaccctg ttcgaatgaa ggcggtcgag tcgaattatc gaccggagct gtcgggaagg 360

cgtccggggc agagtgaggt gctgcggcct ggttgtcgtt caaaaagacc ccggtagccc 420

aacaatcacg aacgaaagga atataattgc ttgcatacta tacattcagt ttctatgtgg 480

cgggtagaca agtctcatgg gcttctaaag gctgtccctt gaaggctact tataaaaact 540

tgctgcgcca tggcacggat cgcgcttgcg caggctgcaa ccctgcgcgc aaggtcaaat 600

acacagcaaa agatactaac agaatttcta aaaacattta aatatttgtt tcgaccagcc 660

aattgtggtc gtaggcacgc aaaagacttt gttttgcgcc caccgagcat ccacgctggc 720

agtcaagcca gtccgatgtg cattgcgtgg cagcatcgag gagcatcaaa aacctcgtgc 780

acgcttttct gtcaatcatc atcaaccact ccaccatgta tacccgatgc atcgcggtgc 840

gcagcgcgcc acgcgtccca gacccgccca aaaacccagc agcggcgaaa gcaaatcttc 900

acttgcccga aaccccgagc agcggcattc acacgtgggc gaaaacccca cttgccctaa 960

caggcgtatg tctgctgtca cgatgcctga caacggtatt atagatatac actgattaat 1020

gtttgagtgt gtgcgagtcg cgaatcagga atgaattgct agtaggcact ccgaccgggc 1080

gggggccgag ggacca 1096

<210> 60

<211> 1075

<212> DNA

<213> Prototheca moriformis

<400> 60

ggccgacagg acgcgcgtca aaggtgctgg gcgtgtatgc cctggtcggc aggtcgttgc 60

tgttgctgcg ctcgtggttc cgcaaccctg attttggcgt cttattctgg cgtggcaagc 120

gctgacgccc gcgagccggg ccggcggcga tgcggtgtct cacggctgcc gagctccaag 180

ggaggcaaga gcgcccggat cagctgaagg gctttacacg caaggtacag ccgctcctgc 240

aaggctgcgt ggtggacttg aacctgtagg tcctctgctg aagttcctcc actacctcac 300

caggcccagc agaccaaagc acaggctttt caggtccgtg tcatccactc taaaacactc 360

gactacgacc tactgatggc cctagattct tcatcaacaa tgcctgagac acttgctcag 420

aattgaaact ccctgaaggg accaccagag gccctgagtt gttccttccc cccgtggcga 480

gctgccagcc aggctgtacc tgtgatcgag gctggcggga aaataggctt cgtgtgctca 540

ggtcatggga ggtgcaggac agctcatgaa acgccaacaa tcgcacaatt catgtcaagc 600

taatcagcta tttcctcttc acgagctgta attgtcccaa aattctggtc taccgggggt 660

gatccttcgt gtacgggccc ttccctcaac cctaggtatg cgcgcatgcg gtcgccgcgc 720

aactcgcgcg agggccgagg gtttgggacg ggccgtcccg aaatgcagtt gcacccggat 780

gcgcggcgcc tttcttgcga taatttatgc aatggactgc tctgcaaatt tctgggtctg 840

tcgccaaccc taggatcagc ggcgtaggat ttcgtaatca ttcgtcctga tggggagcta 900

ccgactaccc taatatcagc ccggctgcct gacgccagcg tccacttttg cgtacacatt 960

ccattcgtgc ccaagacatt tcattgtggt gcgaagcgtc cccagttacg ctcacctgtt 1020

tcccgacctc cttactgttc tgtcgacaga gcgggcccac aggccggtcg cagcc 1075

<210> 61

<211> 772

<212> DNA

<213> Prototheca moriformis

<400> 61

tcaccagcgg acaaagcacc ggtgtatcag gtccgtgtca tccactctaa agagctcgac 60

tacgacctac tgatggccct agattcttca tcaaaaacgc ctgagacact tgcccaggat 120

tgaaactccc tgaagggacc accaggggcc ctgagttgtt ccttcccccc gtggcgagct 180

gccagccagg ctgtacctgt gatcggggct ggcgggaaaa caggcttcgt gtgctcaggt 240

tatgggaggt gcaggacagc tcattaaacg ccaacaatcg cacaattcat ggcaagctaa 300

tcagttattt cccattaacg agctataatt gtcccaaaat tctggtctac cgggggtgat 360

ccttcgtgta cgggcccttc cctcaaccct aggtatgcgc acatgcggtc gccgcgcaac 420

gcgcgcgagg gccgagggtt tgggacgggc cgtcccgaaa tgcagttgca cccggatgcg 480

tggcaccttt tttgcgataa tttatgcaat ggactgctct gcaaaattct ggctctgtcg 540

ccaaccctag gatcagcggt gtaggatttc gtaatcattc gtcctgatgg ggagctaccg 600

actgccctag tatcagcccg actgcctgac gccagcgtcc acttttgtgc acacattcca 660

ttcgtgccca agacatttca ttgtggtgcg aagcgtcccc agttacgctc acctgatccc 720

caacctcctt attgttctgt cgacagagtg ggcccagagg ccggtcgcag cc 772

<210> 62

<211> 991

<212> DNA

<213> Prototheca moriformis

<400> 62

cgaaggggtc tgcatcgatt cgcgcggtct ggaggccagc gtgactgctc gcgaaaatgc 60

tctgccgtgt cgggctctgg ctggggcggc cagagatctc accgtgccac acgcaactgc 120

cgcactctgt gcccgccacc tggcgcgcac atgcgacctc ttccccgtca taccctctcc 180

tcatgtgatc tttccacacg agtgacgcag gtgcgcggag tggagggaat caggacgttt 240

tcaaggtacc tgctcgagcc gtaccaacag ctgccgcccg gcaaggaaga gatcgaggca 300

gagattgccc ggctggaggc ccggataacg gagctcaaga gcaagctgtc cgagtgagac 360

cgcccaggtg cacgtgtcga ctcgctatga catgtactcg acacaacatg aggaattcat 420

cgaatttgta ggaagcgggc attggtacgg gagtgggaaa gcgaaaaaac ctccctccgg 480

cagtgccatc tgccggagtc gaacgttgat agggttctcg tgacagggtg tgacctctca 540

gccttgcatc aattaaacgc tatagacatt atcagtaacc gtgaatcccg cattggatgc 600

cacccgcgcg accattgggg acctgcatta cagatctagg tgagatgaca gcgaggcaac 660

ttcggcccgc ggcccagctt gcggcgcacc aatattggtc acgggaagcc acacaccgac 720

cataaatgaa tacttgtaag ctatgtcaac cgatcaatgg cgtcgaaagt gtgccacgag 780

gatccatctg gcggggcggc gtggcgcaca agcgcagtcg caatttctcg gacccatctg 840

acctaggccc agcgccgcgg gagaaatccc cggcgggtcc tccacgcagt aaccctaatg 900

agtatcgagc gccgaccatt tacaccatcg cccccgaaat ccttccgaca ttattattat 960

cttttagatc ttggaacaga ctctgccaac c 991

<210> 63

<211> 1347

<212> DNA

<213> Prototheca moriformis

<400> 63

agagagcgga ggtggggttg tgaggtgggg ttgctgacca ggagctcgcg tcgccgagcg 60

cgactcgcac acggtccagt tacccccccc tccgcccaaa cgcaagcctc ccatcttgat 120

gcctttccgg ccacctatac tatttcttag ttcgctgtaa catccagacc gtcctgaata 180

ataacaatgc cctgtgtcaa gtgcattcct aaaaaaattc tgtcccaacc aacaatccca 240

cctgaaatac caccagccct gcccagtaca ctcttccaat accatctccc tacctccacg 300

cgcaagcgac ccccatgcgc gaccaggctc gaaagtgatt tatgacttga gacgagcgag 360

tggcggcgcg gtcgactgcc ttttcatcac gtgccgtacg tcggcgaccg ctagggcttt 420

gcacggcaac gcacggcttc gccaacccga ccagccagga cctcgactac tctaccgcga 480

attcgcctca agaagtcgcc aaatgtgcca tacaccattc cttacagcac tgttcaaact 540

tgatgccaat tttgacattc gggttgctcg ttggctgcgc ccacatcggc cgtgagtgca 600

gcaggcggga tcggacacgg aggacgcggc gtcacgcccc gaacgcagcc cgtaactcta 660

catcaacacg acgtgttgcg taatcccgcc cggctgcgca tcgtgccaac ccattcgcga 720

tggatggtcg gaaaatggtg tgccaactgc cctgagggag gctctcgcga aacgggcacg 780

tccctgaaac cgaaactgtg gccttgtcgt cggccacgca agcacgtgga ccctaaacac 840

caagaaaatc agtaaacaag gttgacatcc tctacgggcg aattgtttgc ccaacccttc 900

atcgcacact gccattataa tgcatctagc tcggcgacaa gtttagaaaa ggcaggctgc 960

attgttccat ttcgccgtgg cggcgtgggt gcccatttta cgaggtttgg gctcccgggc 1020

agcgaccgag ccaggtcgag tccctctcgc ccgtcgacaa tgttgcgaac cccacaagcg 1080

gctaacaaca acttgatggt acctgtacac tgccaattcc ttcttccccg gccgaggttt 1140

acacgtgatg gccatggctt cgcattcagg ccgacttccc attccgactt tccagagggt 1200

ccgcggacgc tgggggttgg ctgcctgagg cccacccttt gttccccgcg tcccgacaaa 1260

cacaattgcg ttacataagg gggagccgcc cccgttcaga gtgcagaaat ctttcactat 1320

attttccagt cgtcagcgaa atcaagt 1347

<210> 64

<211> 1180

<212> DNA

<213> Prototheca moriformis

<400> 64

gatggtgggg tgtctgcctt gggctgggtg atggaggctg gtggtgcgcg ggtttcctga 60

tgcattctat ctacgcagtg tcatggtgtc cattccacac accagtacac ccttacacta 120

aggatccatc cctccttccc tcttcaggac tacatggacc ccacgagcta ccgaccgggc 180

tttctcaaaa acgtcaaggt catgtttgac atgcgggacg tggtggacga cgtgcaaggt 240

gcgtccggag tgcgcgcaaa tgagcaagtc gggcaatgtg tcggggtggg caccggggct 300

ggagatccgc gatccccgag aaaacgccgt accacccccc gcgctattcc ctcgattgcg 360

cgcagatgtg gtgaccgaca cgggggacaa cctggcggac atggggcgcc ggacctggaa 420

gcacgccaag tcgcacacgg ggaggctcgt gcagtccccc ccatcgtacc tcaagggtct 480

ctttggtcgc gatccaaagt acgctggtgg catggcatgc ccgaaatgaa catcatgtgt 540

gatctccgat tgccaatggc cacctccacg gaccaccttg caggcggaag cgcaatccag 600

ggcccgagcc tgacgaggac ggagactcct cgtccagcgc ggggtccccg acccgacgca 660

gcagccgacc cctgctaacc cggcaacgat cggaccagca accttgctgt agttccgatc 720

cgtgatgacg ggcattgccg ccgctcgatc cgctttgatg actgtctatt atttgcgcgg 780

agccccctcg gaaccctacc ccgctcttgc aagccccttg catcggagat cctcgtgcgc 840

ccgccatgac cccactggat tgcccaacat ccttctttat cgtgtaaaat gtgattcctc 900

ggctgcaatc gactggcctt cgcttctggc cccaagaggg ctcgaacgtg cggcagcgag 960

ggcgctgaca cacccaagcc ctagggcttt caacgtcggc tgccaggccg gataggggga 1020

tcgcctcctt tccaccaccc acctacgagg gattcgagtc ggcttccagc tcagctattc 1080

ggccgcgccc ccggccctgc agacgtcctc cagtttccga acaggtcgct ctcagaacac 1140

ctgccgcggc tgcgatacgg caggctctca aagcgtcgac 1180

<210> 65

<211> 1263

<212> DNA

<213> Prototheca moriformis

<400> 65

cgcgtggagc ggtgcgtgcg gatgccgcgc gcctgccaag gccttttgta tgcctggcct 60

gggaagtttc ctgactgaag catcttcaag atgctctctc acgaccagcg acaccaacac 120

cgtcactttt tgcccctcct gccgcaggtg ccactttcta ctttgacgtc ttctccaggc 180

ggtacattgc gggactgagc gccaattcgg ccaagaacag cgctgtcgac ttgaggaggc 240

aggggtccgt cgactctgcc gagtgacacg ccttcgaccc gactgtacta cggcctgctg 300

aagagtgggt ctcgccggcc ggcgtgaccg gccctgtgcc cacaatcgac catctattcg 360

ctccttgtca tctggcgccg tcaattgccc gcgacttgac ggcaactggc tcgatcgagt 420

cgtattgaaa aagcacgttt tgtcctacag ggccgcggtc cgttaccaac gtggttctcg 480

ttaggttttc gtcgggcggt ggtgcgcgaa ctgtccgatg ccatcccggc aaaccccagc 540

aaggtcgcca gtctggttct gacgcaatag agtgcgtttt gggccagtct aaaaattcgt 600

ctggcatgac gtggctccac atcgtacccg gagcctgcct tggtaatgtg aggcaccggt 660

gccaactcca ttatggcagg catcgagcgc gcaggtgagt acatgacctt ccgtgaattg 720

ggaaggcgag cttgtgtaac gcctgcgatc gtgccagtga ggcatcgtaa actcaaaata 780

ttttgtagaa agtgtctgat gcctggtgag gctgcgtagg gcaagggcaa gcccttggca 840

gatgggtaat gggtccggac ctcacaacag caaccccgcg tccccttag ggcccctgag 900

gctcgatggc agggccagcg agcccgcggc caaagggcgc catcccacgg tcgcccaacg 960

actccacggg tcctatacct catcttgaat ggcactaaaa actatagaat atcgggcact 1020

ggtgggcgtc tggggtacag ctggccgagc gcagtggcaa accctaggtc ccgcctcaag 1080

ggcgattccc gggtcaatga cacgcaagca agatcacatg gcgcggtccg cctcgcggct 1140

ccacacccag gccctagttt cgcaacccat aaatatcgcc ccgataccat cataagccag 1200

caaataattt tttatcagag ttccaaacct cctcagctgt gggaaaccag cccactctga 1260

acg 1263

<210> 66

<211> 1400

<212> DNA

<213> Prototheca moriformis

<400> 66

ccgagcagtt catggccaag tacaaggact agagaccgga ggtcggtagg ctgaatggag 60

ctggcgtcgt cgtgcgcgac gtgcacgcga tgcgatacta cgaccccaca aacgcatgcc 120

tcccatcttg atgcctttcc ggccatttat actatttctc atttcgctgt aacatcttga 180

ataatagaat tgccctgtgt caagtggatt ccaagaaata ttctgtccca acaaaacaac 240

ccaacctgaa aacaacctca aataccacca gccctgccca cctgcccagt acacttttcc 300

aataccatct ccctaccttc acgcgcaagc ggcacccatg cgcgaccagg ctcgaaagga 360

tttcacgact caggacgagc gagtggcggc gcgaccgcct gcctgttcgt cacgtgccgt 420

acgtcggcga ccgctagagc tttgcctggc aacccccggc ttcgtcaacc cggccagcca 480

ggatctcgac cactctaccg cgaaatcgcc tcaagaagtc gccaaaagtg ccgtacacca 540

tgcttcgcag cgctgttcaa acttgatgcc aatcttgaca atcaggttgc tcgttggctg 600

cgtccacatc ggccgtgatt gcagcaggcg gggatcggac acggaggacg cggcgtcacg 660

ccgcgaacgc agcccgtaac tctacatcaa cgcgatatgt tgcgtaatcc cgcccggctg 720

cgcattgtga caacccattc gcgatggatg gtcggaaaat ggtgtgccaa ctgccctgag 780

ggactctctc gcgaaacggg cacgtccctg tatccgaaac tgtggcatgg ccttgtcgac 840

cacgcaagca cgtggaccct aacaccacga aaataagtaa aaaaggttga catcctctac 900

gagcgaattg tttgctcgac ccttcatcgc acactgtcat tataatgcat ctagctcggc 960

gacaagttta aaaaaggcag gctgcattat tccattttgc cgtggcggca tgggtgccca 1020

ttttatgagg tttgggctct tgggcagcga ccgagccagg ttgagtccct ctcgcccgtc 1080

gacaacgttc caaagcccat aagtggctaa taaacaactt gatggtacct gtacactgcc 1140

agttccttct tccccggccg aggtttacac gtgatggcca tggcttcgcg tttcaggctg 1200

acttcccatt ccgactttcc agagggtccg cggacgccgg gggttggctg cgtgaggccc 1260

acccccttgtt ccccgcgtcc cgacaaacac aattgcgtta cataaggggg aagccgcccc 1320

ccgttcagag tgcaaacatc tttcattata ttttttcagtc gtcagcgaaa tcaagtatgt 1380

cgctgacagg catgaaggcc 1400

<210> 67

<211> 6490

<212> DNA

<213> Artificial Sequence

<220><223> synthetic construct

<400> 67

gctcttcggc cgccgccact cctgctcgag cgcgcccgac tcgcgctccg cctgcgcccg 60

cgcgtgcgcc gccagcgcct tggccttttc gccgcgctcg tgcgcgtcgc tgatgtccat 120

caccaggtcc atgaggtctg ccttgcgccg gctgagccac tgcttcgtcc gggcggccaa 180

gaggagcatg agggaggact cctggtccag ggtcctgacg tggtcgcggc tctgggagcg 240

ggccagcatc atctggctct gccgcaccga ggccgcctcc aactggtcct ccagcagccg 300

cagtcgccgc cgaccctggc agaggaagac aggtgagggg ggtatgaatt gtacagaaca 360

accacgagcc ttgtctaggc agaatcccta ccagtcatgg ctttacctgg atgacggcct 420

gcgaacagct gtccagcgac cctcgctgcc gccgcttctc ccgcacgctt ctttccagca 480

ccgtgatggc gcgagccagc gccgcacgct ggcgctgcgc ttcgccgatc tgaggacagt 540

cggggaactc tgatcagtct aaaccccctt gcgcgttagt gttgccatcc tttgcagacc 600

ggtgagagcc gacttgttgt gcgccacccc ccacaccacc tcctcccaga ccaattctgt 660

cacctttttg gcgaaggcat cggcctcggc ctgcagagag gacagcagtg cccagccgct 720

gggggttggc ggatgcacgc tcaggtaccc tttcttgcgc tatgacactt ccagcaaaag 780

gtagggcggg ctgcgagacg gcttcccggc gctgcatgca acaccgatga tgcttcgacc 840

ccccgaagct ccttcggggc tgcatgggcg ctccgatgcc gctccagggc gagcgctgtt 900

taaatagcca ggcccccgat tgcaaagaca ttatagcgag ctaccaaagc catattcaaa 960

cacctagatc actaccactt ctacacaggc cactcgagct tgtgatcgca ctccgctaag 1020

ggggcgcctc ttcctcttcg tttcagtcac aacccgcaaa cggcgcgcca tgctgctgca 1080

ggccttcctg ttcctgctgg ccggcttcgc cgccaagatc agcgcctcca tgacgaacga 1140

gacgtccgac cgccccctgg tgcacttcac ccccaacaag ggctggatga acgaccccaa 1200

cggcctgtgg tacgacgaga aggacgccaa gtggcacctg tacttccagt acaacccgaa 1260

cgacaccgtc tgggggacgc ccttgttctg gggccacgcc acgtccgacg acctgaccaa 1320

ctgggaggac cagcccatcg ccatcgcccc gaagcgcaac gactccggcg ccttctccgg 1380

ctccatggtg gtggactaca acaacacctc cggcttcttc aacgacacca tcgacccgcg 1440

ccagcgctgc gtggccatct ggacctacaa caccccggag tccgaggagc agtacatctc 1500

ctacagcctg gacggcggct acaccttcac cgagtaccag aagaaccccg tgctggccgc 1560

caactccacc cagttccgcg acccgaaggt cttctggtac gagccctccc agaagtggat 1620

catgaccgcg gccaagtccc aggactacaa gatcgagatc tactcctccg acgacctgaa 1680

gtcctggaag ctggagtccg cgttcgccaa cgagggcttc ctcggctacc agtacgagtg 1740

ccccggcctg atcgaggtcc ccaccgagca ggaccccagc aagtcctact gggtgatgtt 1800

catctccatc aaccccggcg ccccggccgg cggctccttc aaccagtact tcgtcggcag 1860

cttcaacggc acccacttcg aggccttcga caaccagtcc cgcgtggtgg acttcggcaa 1920

ggactactac gccctgcaga ccttcttcaa caccgacccg acctacggga gcgccctggg 1980

catcgcgtgg gcctccaact gggagtactc cgccttcgtg cccaccaacc cctggcgctc 2040

ctccatgtcc ctcgtgcgca agttctccct caacaccgag taccaggcca acccggagac 2100

ggagctgatc aacctgaagg ccgagccgat cctgaacatc agcaacgccg gcccctggag 2160

ccggttcgcc accaacacca cgttgacgaa ggccaacagc tacaacgtcg acctgtccaa 2220

cagcaccggc accctggagt tcgagctggt gtacgccgtc aacaccaccc agacgatctc 2280

caagtccgtg ttcgcggacc tctccctctg gttcaagggc ctggaggacc ccgaggagta 2340

cctccgcatg ggcttcgagg tgtccgcgtc ctccttcttc ctggaccgcg ggaacagcaa 2400

ggtgaagttc gtgaaggaga acccctactt caccaaccgc atgagcgtga acaaccagcc 2460

cttcaagagc gagaacgacc tgtcctacta caaggtgtac ggcttgctgg accagaacat 2520

cctggagctg tacttcaacg acggcgacgt cgtgtccacc aacacctact tcatgaccac 2580

cgggaacgcc ctgggctccg tgaacatgac gacgggggtg gacaacctgt tctacatcga 2640

caagttccag gtgcgcgagg tcaagtgaca attggcagca gcagctcgga tagtatcgac 2700

acactctgga cgctggtcgt gtgatggact gttgccgcca cacttgctgc cttgacctgt 2760

gaatatccct gccgctttta tcaaacagcc tcagtgtgtt tgatcttgtg tgtacgcgct 2820

tttgcgagtt gctagctgct tgtgctattt gcgaatacca cccccagcat ccccttccct 2880

cgtttcatat cgcttgcatc ccaaccgcaa cttatctacg ctgtcctgct atccctcagc 2940

gctgctcctg ctcctgctca ctgcccctcg cacagccttg gtttgggctc cgcctgtatt 3000

ctcctggtac tgcaacctgt aaaccagcac tgcaatgctg atgcacggga agtagtggga 3060

tgggaacaca aatggaggat cccgcgtctc gaacagagcg cgcagaggaa cgctgaaggt 3120

ctcgcctctg tcgcacctca gcgcggcata caccacaata accacctgac gaatgcgctt 3180

ggttcttcgt ccattagcga agcgtccggt tcacacacgt gccacgttgg cgaggtggca 3240

ggtgacaatg atcggtggag ctgatggtcg aaacgttcac agcctaggga tatcgaattc 3300

atcgaatttg taggaagcgg gcattggtac gggagtggga aagcgaaaaa acctccctcc 3360

ggcagtgcca tctgccggag tcgaacgttg atagggttct cgtgacaggg tgtgacctct 3420

cagccttgca tcaattaaac gctatagaca ttatcagtaa ccgtgaatcc cgcattggat 3480

gccacccgcg cgaccattgg ggacctgcat tacagatcta ggtgagatga cagcgaggca 3540

acttcggccc gcggcccagc ttgcggcgca ccaatattgg tcacgggaag ccacacaccg 3600

accataaatg aatacttgta agctatgtca accgatcaat ggcgtcgaaa gtgtgccacg 3660

aggatccatc tggcggggcg gcgtggcgca caagcgcagt cgcaatttct cggacccatc 3720

tgacctaggc ccagcgccgc gggagaaatc cccggcgggt cctccacgca gtaaccctaa 3780

tgagtatcga gcgccgacca tttacaccat cgcccccga aatccttccg acattattat 3840

tatcttttag atcttggaac agactctgcc aaccactagt atggccaccg catccacttt 3900

ctcggcgttc aatgcccgct gcggcgacct gcgtcgctcg gcgggctccg ggccccggcg 3960

cccagcgagg cccctccccg tgcgcgggcg cgcccccgac tggtccatgc tgttcgccgt 4020

gatcaccacc atcttctccg ccgccgagaa gcagtggacc aacctggagt ggaagcccaa 4080

gcccaacccc ccccagctgc tggacgacca cttcggcccc cacggcctgg tgttccgccg 4140

caccttcgcc atccgcagct acgaggtggg ccccgaccgc tccaccagca tcgtggccgt 4200

gatgaaccac ctgcaggagg ccgccctgaa ccacgccaag tccgtgggca tcctgggcga 4260

cggcttcggc accaccctgg agatgtccaa gcgcgacctg atctgggtgg tgaagcgcac 4320

ccacgtggcc gtggagcgct accccgcctg gggcgacacc gtggaggtgg agtgctgggt 4380

gggcgcctcc ggcaacaacg gccgccgcca cgacttcctg gtgcgcgact gcaagaccgg 4440

cgagatcctg acccgctgca cctccctgag cgtgatgatg aacacccgca cccgccgcct 4500

gagcaagatc cccgaggagg tgcgcggcga gatcggcccc gccttcatcg acaacgtggc 4560

cgtgaaggac gaggagatca agaagcccca gaagctgaac gactccaccg ccgactacat 4620
ccagggcggc ctgacccccc gctggaacga cctggacatc aaccagcacg tgaacaacat 4680
caagtacgtg gactggatcc tggagaccgt gcccgacagc atcttcgaga gccaccacat 4740
ctcctccttc accatcgagt accgccgcga gtgcaccatg gacagcgtgc tgcagtccct 4800
gaccaccgtg agcggcggct cctccgaggc cggcctggtg tgcgagcacc tgctgcagct 4860
ggagggcggc agcgaggtgc tgcgcgccaa gaccgagtgg cgccccaagc tgaccgactc 4920
cttccgcggc atcagcgtga tccccgccga gtccagcgtg atggactaca aggaccacga 4980

cggcgactac aaggaccacg acatcgacta caaggacgac gacgacaagt gactcgagtt 5040
aattaactcg aggcagcagc agctcggata gtatcgacac actctggacg ctggtcgtgt 5100
gatggactgt tgccgccaca cttgctgcct tgacctgtga atatccctgc cgcttttatc 5160
aaacagcctc agtgtgtttg atcttgtgtg tacgcgcttt tgcgagttgc tagctgcttg 5220
tgctatttgc gaataccacc cccagcatcc ccttccctcg tttcatatcg cttgcatccc 5280
aaccgcaact tatctacgct gtcctgctat ccctcagcgc tgctcctgct cctgctcact 5340
gcccctcgca cagccttggt ttgggctccg cctgtattct cctggtactg caacctgtaa 5400

accagcactg caatgctgat gcacgggaag tagtgggatg ggaacacaaa tggaaagctg 5460
tagagctcct tgttttccag aaggagttgc tccttgagcc tttcattctc agcctcgata 5520
acctccaaag ccgctctaat tgtggagggg gttcgaattt aaaagcttgg aatgttggtt 5580
cgtgcgtctg gaacaagccc agacttgttg ctcactggga aaaggaccat cagctccaaa 5640
aaacttgccg ctcaaaccgc gtacctctgc tttcgcgcaa tctgccctgt tgaaatcgcc 5700
accacattca tattgtgacg cttgagcagt ctgtaattgc ctcagaatgt ggaatcatct 5760
gccccctgtg cgagcccatg ccaggcatgt cgcgggcgag gacacccgcc actcgtacag 5820

cagaccatta tgctacctca caatagttca taacagtgac catatttctc gaagctcccc 5880
aacgagcacc tccatgctct gagtggccac cccccggccc tggtgcttgc ggagggcagg 5940
tcaaccggca tggggctacc gaaatccccg accggatccc accacccccg cgatgggaag 6000
aatctctccc cgggatgtgg gcccaccacc agcacaacct gctggcccag gcgagcgtca 6060
aaccatacca cacaaatatc cttggcatcg gccctgaatt ccttctgccg ctctgctacc 6120
cggtgcttct gtccgaagca ggggttgcta gggatcgctc cgagtccgca aacccttgtc 6180
gcgtggcggg gcttgttcga gcttgttcga gcttgaagag cctctagagt cgacctgcag 6240

gcatgcaagc ttggcgtaat catggtcata gctgtttcct gtgtgaaatt gttatccgct 6300

cacaattcca cacaacatac gagccggaag cataaagtgt aaagcctggg gtgcctaatg 6360

agtgagctaa ctcacattaa ttgcgttgcg ctcactgccc gctttccagt cgggaaacct 6420

gtcgtgccag ctgcattaat gaatcggcca acgcgcgggg agaggcggtt tgcgtattgg 6480

gcgctcttcc 6490

<210> 68

<211> 20

<212> DNA

<213> Cinnamomum camphora

<400> 68

taccccgcct ggggcgacac 20

<210> 69

<211> 20

<212> DNA

<213> Cinnamomum camphora

<400> 69

cttgctcagg cggcgggtgc 20

<210> 70

<211> 20

<212> DNA

<213> Prototheca moriformis

<400> 70

ccggatctcg gccagggcta 20

<210> 71

<211> 20

<212> DNA

<213> Prototheca moriformis

<400> 71

tcgatgtcgt gcaccgtcgc 20

<210> 72

<211> 1092

<212> DNA

<213>

Prototheca moriformis

<400> 72

gctcttcggg tttgctcacc cgcgaggtcg acgcccagca tggctatcaa gacgaacagg 60

cagcctgtgg agaagcctcc gttcacgatc gggacgctgc gcaaggccat ccccgcgcac 120

tgtttcgagc gctcggcgct tcgtagcagc atgtacctgg cctttgacat cgcggtcatg 180

tccctgctct acgtcgcgtc gacgtacatc gaccctgcgc cggtgcctac gtgggtcaag 240

tatggcgtca tgtggccgct ctactggttc ttccaggtgt gtgtgagggt tgtggttgcc 300

cgtatcgagg tcctggtggc gcgcatgggg gagaaggcgc ctgtcccgct gacccccccg 360

gctaccctcc cggcaccttc cagggcgcct tcggcacggg tgtctgggtg tgcgcgcacg 420

agtgcggcca ccaggccttt tcctccagcc aggccatcaa cgacggcgtg ggcctggtgt 480

tccacagcct gctgctggtg ccctactact cctggaagca ctcgcaccgc cgccaccact 540

ccaacacggg gtgcctggac aaggacgagg tgtttgtgcc gccgcaccgc gcagtggcgc 600

acgagggcct ggagtgggag gagtggctgc ccatccgcat gggcaaggtg ctggtcaccc 660

tgaccctggg ctggccgctg tacctcatgt tcaacgtcgc ctcgcggccg tacccgcgct 720

tcgccaacca ctttgacccg tggtcgccca tcttcagcaa gcgcgaggta ccctttcttg 780

cgctatgaca cttccagcaa aaggtagggc gggctgcgag acggcttccc ggcgctgcat 840

gcaacaccga tgatgcttcg acccccccgaa gctccttcgg ggctgcatgg gcgctccgat 900

gccgctccag ggcgagcgct gtttaaatag ccaggccccc gattgcaaag acattatagc 960

gagctaccaa agccatattc aaacacctag atcactacca cttctacaca ggccactcga 1020

gcttgtgatc gcactccgct aaggggggcgc ctcttcctct tcgtttcagt cacaacccgc 1080

aaacggcgcg cc 1092

<210> 73

<211> 1400

<212> DNA

<213> Prototheca moriformis

<400> 73

caattggcag cagcagctcg gatagtatcg acacactctg gacgctggtc gtgtgatgga 60

ctgttgccgc cacacttgct gccttgacct gtgaatatcc ctgccgcttt tatcaaacag 120

cctcagtgtg tttgatcttg tgtgtacgcg cttttgcgag ttgctagctg cttgtgctat 180

ttgcgaatac caccccccagc atcccccttcc ctcgtttcat atcgcttgca tcccaaccgc 240

aacttatcta cgctgtcctg ctatccctca gcgctgctcc tgctcctgct cactgcccct 300

cgcacagcct tggtttgggc tccgcctgta ttctcctggt actgcaacct gtaaaccagc 360

actgcaatgc tgatgcacgg gaagtagtgg gatgggaaca caaatggagc atcgaggtgg 420

tcatctccga cctcgcgttg gtggcggtgc tcagcgggct cagcgtgctg ggccgcacca 480

tgggctgggc ctggctggtc aagacctacg tggtgcccta catgatcgtg aacatgtggc 540
tggtgctcat cacgctgctc cagcacacgc acccggccct gccgcactac ttcgagaagg 600
actgggactg gctacgcggc gccatggcca ccgtcgaccg ctccatgggc ccgcccttca 660
tggacagcat cctgcaccac atctccgaca cccacgtgct gcaccacctc ttcagcacca 720
tcccgcacta ccacgccgag gaggcctccg ccgccatccg gcccatcctg ggcaagtact 780
accaatccga cagccgctgg gtcggccgcg ccctgtggga ggactggcgc gactgccgct 840
acgtcgtccc cgacgcgccc gaggacgact ccgcgctctg gttccacaag tgagcgcgcc 900

tgcgcgagga cgcagaacaa cgctgccgcc gtgtcttttg cacgcgcgac tccggcgctt 960
cgctggtggc acccccataa agaaaccctc aattctgttt gtggaagaca cggtgtaccc 1020
ccacccaccc acctgcacct ctattattgg tattattgac gcgggagtgg gcgttgtacc 1080
ctacaacgta gcttctctag ttttcagctg gctcccacca ttgtaaagag cctctagagt 1140
cgacctgcag gcatgcaagc ttggcgtaat catggtcata gctgtttcct gtgtgaaatt 1200
gttatccgct cacaattcca cacaacatac gagccggaag cataaagtgt aaagcctggg 1260
gtgcctaatg agtgagctaa ctcacattaa ttgcgttgcg ctcactgccc gctttccagt 1320

cgggaaacct gtcgtgccag ctgcattaat gaatcggcca acgcgcgggg agaggcggtt 1380
tgcgtattgg gcgctcttcc 1400
<210> 74
<211> 4091
<212> DNA
<213> Prototheca moriformis
<400> 74
gctcttcggg tttgctcacc cgcgaggtcg acgcccagca tggctatcaa gacgaacagg 60
cagcctgtgg agaagcctcc gttcacgatc gggacgctgc gcaaggccat ccccgcgcac 120
tgtttcgagc gctcggcgct tcgtagcagc atgtacctgg cctttgacat cgcggtcatg 180
tccctgctct acgtcgcgtc gacgtacatc gaccctgcgc cggtgcctac gtgggtcaag 240

tatggcgtca tgtggccgct ctactggttc ttccaggtgt gtgtgagggt tgtggttgcc 300
cgtatcgagg tcctggtggc gcgcatgggg gagaaggcgc ctgtcccgct gacccccccg 360
gctaccctcc cggcaccttc cagggcgcct tcggcacggg tgtctgggtg tgcgcgcacg 420
agtgcggcca ccaggccttt tcctccagcc aggccatcaa cgacggcgtg ggcctggtgt 480
tccacagcct gctgctggtg ccctactact cctggaagca ctcgcaccgc cgccaccact 540

ccaacacggg gtgcctggac aaggacgagg tgtttgtgcc gccgcaccgc gcagtggcgc 600

acgagggcct ggagtgggag gagtggctgc ccatccgcat gggcaaggtg ctggtcaccc 660

tgaccctggg ctggccgctg tacctcatgt tcaacgtcgc ctcgcggccg tacccgcgct 720

tcgccaacca ctttgacccg tggtcgccca tcttcagcaa gcgcgaggta ccctttcttg 780

cgctatgaca cttccagcaa aaggtagggc gggctgcgag acggcttccc ggcgctgcat 840

gcaacaccga tgatgcttcg acccccccgaa gctccttcgg ggctgcatgg gcgctccgat 900

gccgctccag ggcgagcgct gtttaaatag ccaggccccc gattgcaaag acattatagc 960

gagctaccaa agccatattc aaacacctag atcactacca cttctacaca ggccactcga 1020

gcttgtgatc gcactccgct aaggggggcgc ctcttcctct tcgtttcagt cacaacccgc 1080

aaacggcgcg ccatgctgct gcaggccttc ctgttcctgc tggccggctt cgccgccaag 1140

atcagcgcct ccatgacgaa cgagacgtcc gaccgccccc tggtgcactt cacccccaac 1200

aagggctgga tgaacgaccc caacggcctg tggtacgacg agaaggacgc caagtggcac 1260

ctgtacttcc agtacaaccc gaacgacacc gtctgggggga cgcccttgtt ctggggccac 1320

gccacgtccg acgacctgac caactgggag gaccagccca tcgccatcgc cccgaagcgc 1380

aacgactccg gcgccttctc cggctccatg gtggtggact acaacaacac ctccggcttc 1440

ttcaacgaca ccatcgaccc gcgccagcgc tgcgtggcca tctggaccta caacaccccg 1500

gagtccgagg agcagtacat ctcctacagc ctggacggcg gctacacctt caccgagtac 1560

cagaagaacc ccgtgctggc cgccaactcc acccagttcc gcgacccgaa ggtcttctgg 1620

tacgagccct cccagaagtg gatcatgacc gcggccaagt cccaggacta caagatcgag 1680

atctactcct ccgacgacct gaagtcctgg aagctggagt ccgcgttcgc caacgagggc 1740

ttcctcggct accagtacga gtgccccggc ctgatcgagg tccccaccga gcaggacccc 1800

agcaagtcct actgggtgat gttcatctcc atcaaccccg gcgccccggc cggcggctcc 1860

ttcaaccagt acttcgtcgg cagcttcaac ggcacccact tcgaggcctt cgacaaccag 1920

tcccgcgtgg tggacttcgg caaggactac tacgccctgc agaccttctt caacaccgac 1980

ccgacctacg ggagcgccct gggcatcgcg tgggcctcca actgggagta ctccgccttc 2040

gtgcccacca acccctggcg ctcctccatg tccctcgtgc gcaagttctc cctcaacacc 2100

gagtaccagg ccaacccgga gacggagctg atcaacctga aggccgagcc gatcctgaac 2160

atcagcaacg ccggcccctg gagccggttc gccaccaaca ccacgttgac gaaggccaac 2220

agctacaacg tcgacctgtc caacagcacc ggcaccctgg agttcgagct ggtgtacgcc 2280

gtcaacacca cccagacgat ctccaagtcc gtgttcgcgg acctctccct ctggttcaag 2340

ggcctggagg accccgagga gtacctccgc atgggcttcg aggtgtccgc gtcctccttc 2400
ttcctggacc gcgggaacag caaggtgaag ttcgtgaagg agaaccccta cttcaccaac 2460
cgcatgagcg tgaacaacca gcccttcaag agcgagaacg acctgtccta ctacaaggtg 2520
tacggcttgc tggaccagaa catcctggag ctgtacttca acgacggcga cgtcgtgtcc 2580
accaacacct acttcatgac caccgggaac gccctgggct ccgtgaacat gacgacgggg 2640
gtggacaacc tgttctacat cgacaagttc caggtgcgcg aggtcaagtg acaattggca 2700
gcagcagctc ggatagtatc gacacactct ggacgctggt cgtgtgatgg actgttgccg 2760

ccacacttgc tgccttgacc tgtgaatatc cctgccgctt ttatcaaaca gcctcagtgt 2820
gtttgatctt gtgtgtacgc gcttttgcga gttgctagct gcttgtgcta tttgcgaata 2880
ccacccccag catccccttc cctcgtttca tatcgcttgc atcccaaccg caacttatct 2940
acgctgtcct gctatccctc agcgctgctc ctgctcctgc tcactgcccc tcgcacagcc 3000
ttggtttggg ctccgcctgt attctcctgg tactgcaacc tgtaaaccag cactgcaatg 3060
ctgatgcacg ggaagtagtg ggatgggaac acaaatggag catcgaggtg gtcatctccg 3120
acctcgcgtt ggtggcggtg ctcagcgggc tcagcgtgct gggccgcacc atgggctggg 3180

cctggctggt caagacctac gtggtgccct acatgatcgt gaacatgtgg ctggtgctca 3240
tcacgctgct ccagcacacg cacccggccc tgccgcacta cttcgagaag gactgggact 3300
ggctacgcgg cgccatggcc accgtcgacc gctccatggg cccgcccttc atggacagca 3360
tcctgcacca catctccgac acccacgtgc tgcaccacct cttcagcacc atcccgcact 3420
accacgccga ggaggcctcc gccgccatcc ggcccatcct gggcaagtac taccaatccg 3480
acagccgctg ggtcggccgc gccctgtggg aggactggcg cgactgccgc tacgtcgtcc 3540
ccgacgcgcc cgaggacgac tccgcgctct ggttccacaa gtgagcgcgc ctgcgcgagg 3600

acgcagaaca acgctgccgc cgtgtctttt gcacgcgcga ctccggcgct tcgctggtgg 3660
caccccccata aagaaaccct caattctgtt tgtggaagac acggtgtacc cccacccacc 3720
cacctgcacc tctattattg gtattattga cgcgggagtg ggcgttgtac cctacaacgt 3780
agcttctcta gttttcagct ggctcccacc attgtaaaga gcctctagag tcgacctgca 3840
ggcatgcaag cttggcgtaa tcatggtcat agctgtttcc tgtgtgaaat tgttatccgc 3900
tcacaattcc acacaacata cgagccggaa gcataaagtg taaagcctgg ggtgcctaat 3960
gagtgagcta actcacatta attgcgttgc gctcactgcc cgctttccag tcgggaaacc 4020

tgtcgtgcca gctgcattaa tgaatcggcc aacgcgcggg gagaggcggt ttgcgtattg 4080

ggcgctcttc c 4091

<210> 75

<211> 510

<212> DNA

<213> Prototheca moriformis

<400> 75

gctcttccgc ctggagctgg tgcagagcat gggtcagttt gcggaggaga gggtgctccc 60

cgtgctgcac cccgtggaca agctgtggca gccgcaggac ttcctgcccg accccgagtc 120

gcccgacttc gaggaccagg tggcggagct gcgcgcgcgc gccaaggacc tgcccgacga 180

gtactttgtg gtgctggtgg gcgacatgat cacggaggag gcgctgccga cctacatggc 240

catgctcaac accttggacg gtgtgcgcga cgacacgggc gcggctgacc acccgtgggc 300

gcgctggacg cggcagtggg tggccgagga gaaccggcac ggcgacctgc tgaacaagta 360

ctgttggctg acggggcgcg tcaacatgcg ggccgtggag gtgaccatca acaacctgat 420

caagagcggc atgaacccgc agacggacaa caacccttac ttgggcttcg tctacacctc 480

cttccaggag cgcgccacca agtaggtacc 510

<210> 76

<211> 1186

<212> DNA

<213> Prototheca moriformis

<400> 76

caattggcag cagcagctcg gatagtatcg acacactctg gacgctggtc gtgtgatgga 60

ctgttgccgc cacacttgct gccttgacct gtgaatatcc ctgccgcttt tatcaaacag 120

cctcagtgtg tttgatcttg tgtgtacgcg cttttgcgag ttgctagctg cttgtgctat 180

ttgcgaatac caccccagc atcccttcc ctcgtttcat atcgcttgca tcccaaccgc 240

aacttatcta cgctgtcctg ctatccctca gcgctgctcc tgctcctgct cactgcccct 300

cgcacagcct tggtttgggc tccgcctgta ttctcctggt actgcaacct gtaaaccagc 360

actgcaatgc tgatgcacgg gaagtagtgg gatgggaaca caaatggaag gatcgtagag 420

ctccagccac ggcaacaccg cgcgcctggc ggccgagcac ggcgacaagg gcctgagcaa 480

gatctgcggg ctgatcgcca gcgacgaggg ccggcacgag atcgcctaca cgcgcatcgt 540

ggacgagttc ttccgcctcg accccgaggg cgccgtcgcc gcctacgcca acatgatgcg 600

caagcagatc accatgcccg cgcacctcat ggacgacatg ggccacggcg aggccaaccc 660

gggccgcaac ctcttcgccg acttctccgc cgtcgccgag aagatcgacg tctacgacgc 720

cgaggactac tgccgcatcc tggagcacct caacgcgcgc tggaaggtgg acgagcgcca 780

ggtcagcggc caggccgccg cggaccagga gtacgttctg ggcctgcccc agcgcttccg 840

gaaactcgcc gagaagaccg ccgccaagcg caagcgcgtc gcgcgcaggc ccgtcgcctt 900

ctcctggaga gaagagcctc tagagtcgac ctgcaggcat gcaagcttgg cgtaatcatg 960

gtcatagctg tttcctgtgt gaaattgtta tccgctcaca attccacaca acatacgagc 1020

cggaagcata aagtgtaaag cctggggtgc ctaatgagtg agctaactca cattaattgc 1080

gttgcgctca ctgcccgctt tccagtcggg aaacctgtcg tgccagctgc attaatgaat 1140

cggccaacgc gcggggagag gcggtttgcg tattgggcgc tcttcc 1186

<210> 77

<211> 3615

<212> DNA

<213> Prototheca moriformis

<400> 77

gctcttccgc ctggagctgg tgcagagcat gggtcagttt gcggaggaga gggtgctccc 60

cgtgctgcac cccgtggaca agctgtggca gccgcaggac ttcctgcccg accccgagtc 120

gcccgacttc gaggaccagg tggcggagct gcgcgcgcgc gccaaggacc tgcccgacga 180

gtactttgtg gtgctggtgg gcgacatgat cacggaggag gcgctgccga cctacatggc 240

catgctcaac accttggacg gtgtgcgcga cgacacgggc gcggctgacc acccgtgggc 300

gcgctggacg cggcagtggg tggccgagga gaaccggcac ggcgacctgc tgaacaagta 360

ctgttggctg acggggcgcg tcaacatgcg ggccgtggag gtgaccatca acaacctgat 420

caagagcggc atgaacccgc agacggacaa caacccttac ttgggcttcg tctacacctc 480

cttccaggag cgcgccacca agtaggtacc ctttcttgcg ctatgacact tccagcaaaa 540

ggtagggcgg gctgcgagac ggcttcccgg cgctgcatgc aacaccgatg atgcttcgac 600

cccccgaagc tccttcgggg ctgcatgggc gctccgatgc cgctccaggg cgagcgctgt 660

ttaaatagcc aggccccccga ttgcaaagac attatagcga gctaccaaag ccatattcaa 720

acacctagat cactaccact tctacacagg ccactcgagc ttgtgatcgc actccgctaa 780

gggggcgcct cttcctcttc gtttcagtca caacccgcaa acggcgcgcc atgctgctgc 840

aggccttcct gttcctgctg gccggcttcg ccgccaagat cagcgcctcc atgacgaacg 900

agacgtccga ccgccccctg gtgcacttca cccccaacaa gggctggatg aacgacccca 960

acggcctgtg gtacgacgag aaggacgcca agtggcacct gtacttccag tacaacccga 1020

acgacaccgt ctgggggacg cccttgttct ggggccacgc cacgtccgac gacctgacca 1080

actgggagga ccagcccatc gccatcgccc cgaagcgcaa cgactccggc gccttctccg 1140

gctccatggt ggtggactac aacaacacct ccggcttctt caacgacacc atcgacccgc 1200

gccagcgctg cgtggccatc tggacctaca acaccccgga gtccgaggag cagtacatct 1260

cctacagcct ggacggcggc tacaccttca ccgagtacca gaagaacccc gtgctggccg 1320

ccaactccac ccagttccgc gacccgaagg tcttctggta cgagccctcc cagaagtgga 1380

tcatgaccgc ggccaagtcc caggactaca agatcgagat ctactcctcc gacgacctga 1440

agtcctggaa gctggagtcc gcgttcgcca acgagggctt cctcggctac cagtacgagt 1500

gccccggcct gatcgaggtc cccaccgagc aggaccccag caagtcctac tgggtgatgt 1560

tcatctccat caaccccggc gccccggccg gcggctcctt caaccagtac ttcgtcggca 1620

gcttcaacgg cacccacttc gaggccttcg acaaccagtc ccgcgtggtg gacttcggca 1680

aggactacta cgccctgcag accttcttca acaccgaccc gacctacggg agcgccctgg 1740

gcatcgcgtg ggcctccaac tgggagtact ccgccttcgt gcccaccaac ccctggcgct 1800

cctccatgtc cctcgtgcgc aagttctccc tcaacaccga gtaccaggcc aacccggaga 1860

cggagctgat caacctgaag gccgagccga tcctgaacat cagcaacgcc ggcccctgga 1920

gccggttcgc caccaacacc acgttgacga aggccaacag ctacaacgtc gacctgtcca 1980

acagcaccgg caccctggag ttcgagctgg tgtacgccgt caacaccacc cagacgatct 2040

ccaagtccgt gttcgcggac ctctccctct ggttcaaggg cctggaggac cccgaggagt 2100

acctccgcat gggcttcgag gtgtccgcgt cctccttctt cctggaccgc gggaacagca 2160

aggtgaagtt cgtgaaggag aacccctact tcaccaaccg catgagcgtg aacaaccagc 2220

ccttcaagag cgagaacgac ctgtcctact acaaggtgta cggcttgctg gaccagaaca 2280

tcctggagct gtacttcaac gacggcgacg tcgtgtccac caacacctac ttcatgacca 2340

ccgggaacgc cctgggctcc gtgaacatga cgacgggggt ggacaacctg ttctacatcg 2400

acaagttcca ggtgcgcgag gtcaagtgac aattggcagc agcagctcgg atagtatcga 2460

cacactctgg acgctggtcg tgtgatggac tgttgccgcc acacttgctg ccttgacctg 2520

tgaatatccc tgccgctttt atcaaacagc ctcagtgtgt ttgatcttgt gtgtacgcgc 2580

ttttgcgagt tgctagctgc ttgtgctatt tgcgaatacc acccccagca tcccccttccc 2640

tcgtttcata tcgcttgcat cccaaccgca acttatctac gctgtcctgc tatccctcag 2700

cgctgctcct gctcctgctc actgcccctc gcacagcctt ggtttgggct ccgcctgtat 2760

tctcctggta ctgcaacctg taaaccagca ctgcaatgct gatgcacggg aagtagtggg 2820

atgggaacac aaatggaagg atcgtagagc tccagccacg gcaacaccgc gcgcctggcg 2880

gccgagcacg gcgacaaggg cctgagcaag atctgcgggc tgatcgccag cgacgagggc 2940

cggcacgaga tcgcctacac gcgcatcgtg gacgagttct tccgcctcga ccccgagggc 3000

gccgtcgccg cctacgccaa catgatgcgc aagcagatca ccatgcccgc gcacctcatg 3060

gacgacatgg gccacggcga ggccaacccg ggccgcaacc tcttcgccga cttctccgcc 3120

gtcgccgaga agatcgacgt ctacgacgcc gaggactact gccgcatcct ggagcacctc 3180

aacgcgcgct ggaaggtgga cgagcgccag gtcagcggcc aggccgccgc ggaccaggag 3240

tacgttctgg gcctgcccca gcgcttccgg aaactcgccg agaagaccgc cgccaagcgc 3300

aagcgcgtcg cgcgcaggcc cgtcgccttc tcctggagag aagagcctct agagtcgacc 3360

tgcaggcatg caagcttggc gtaatcatgg tcatagctgt ttcctgtgtg aaattgttat 3420

ccgctcacaa ttccacacaa catacgagcc ggaagcataa agtgtaaagc ctggggtgcc 3480

taatgagtga gctaactcac attaattgcg ttgcgctcac tgcccgcttt ccagtcggga 3540

aacctgtcgt gccagctgca ttaatgaatc ggccaacgcg cggggagagg cggtttgcgt 3600

attgggcgct cttcc 3615

<210> 78

<211> 511

<212> DNA

<213> Prototheca moriformis

<400> 78

gctcttcccg cctggagctg gtgcagagca tggggcagtt tgcggaggag agggtgctcc 60

ccgtgctgca ccccgtggac aagctgtggc agccgcagga cttcctgccc gaccccgagt 120

cgcccgactt cgaggaccag gtggcggagc tgcgcgcgcg cgccaaggac ctgcccgacg 180

agtactttgt ggtgctggtg ggcgacatga tcacggagga ggcgctgccg acctacatgg 240

ccatgctcaa caccttggac ggtgtgcgcg acgacacggg cgcggctgac cacccgtggg 300

cgcgctggac gcggcagtgg gtggccgagg agaaccggca cggcgacctg ctgaacaagt 360

actgttggct gacggggcgc gtcaacatgc gggccgtgga ggtgaccatc aacaacctga 420

tcaagagcgg catgaacccg cagacggaca acaaccctta cttgggcttc gtctacacct 480

ccttccagga gcgcgccacc aagtaggtac c 511

<210> 79

<211> 761

<212> DNA

<213> Prototheca moriformis

<400> 79

cagccacggc aacaccgcgc gccttgcggc cgagcacggc gacaagaacc tgagcaagat 60

ctgcgggctg atcgccagcg acgagggccg gcacgagatc gcctacacgc gcatcgtgga 120

cgagttcttc cgcctcgacc ccgagggcgc cgtcgccgcc tacgccaaca tgatgcgcaa 180

gcagatcacc atgcccgcgc acctcatgga cgacatgggc cacggcgagg ccaacccggg 240

ccgcaacctc ttcgccgact tctccgcggt cgccgagaag atcgacgtct acgacgccga 300

ggactactgc cgcatcctgg agcacctcaa cgcgcgctgg aaggtggacg agcgccaggt 360

cagcggccag gccgccgcgg accaggagta cgtcctgggc ctgccccagc gcttccggaa 420

actcgccgag aagaccgccg ccaagcgcaa gcgcgtcgcg cgcaggcccg tcgccttctc 480

ctggagaaga gcctctagag tcgacctgca ggcatgcaag cttggcgtaa tcatggtcat 540

agctgtttcc tgtgtgaaat tgttatccgc tcacaattcc acacaacata cgagccggaa 600

gcataaagtg taaagcctgg ggtgcctaat gagtgagcta actcacatta attgcgttgc 660

gctcactgcc cgctttccag tcgggaaacc tgtcgtgcca gctgcattaa tgaatcggcc 720

aacgcgcggg gagaggcggt ttgcgtattg ggcgctcttc c 761

<210> 80

<211> 3599

<212> DNA

<213> Prototheca moriformis

<400> 80

gctcttcccg cctggagctg gtgcagagca tggggcagtt tgcggaggag agggtgctcc 60

ccgtgctgca ccccgtggac aagctgtggc agccgcagga cttcctgccc gaccccgagt 120

cgcccgactt cgaggaccag gtggcggagc tgcgcgcgcg cgccaaggac ctgcccgacg 180

agtactttgt ggtgctggtg ggcgacatga tcacggagga ggcgctgccg acctacatgg 240

ccatgctcaa caccttggac ggtgtgcgcg acgacacggg cgcggctgac cacccgtggg 300

cgcgctggac gcggcagtgg gtggccgagg agaaccggca cggcgacctg ctgaacaagt 360

actgttggct gacggggcgc gtcaacatgc gggccgtgga ggtgaccatc aacaacctga 420

tcaagagcgg catgaacccg cagacggaca acaaccctta cttgggcttc gtctacacct 480

ccttccagga gcgcgccacc aagtaggtac cctttcttgc gctatgacac ttccagcaaa 540

aggtagggcg ggctgcgaga cggcttcccg gcgctgcatg caacaccgat gatgcttcga 600

ccccccgaag ctccttcggg gctgcatggg cgctccgatg ccgctccagg gcgagcgctg 660

tttaaatagc caggcccccg attgcaaaga cattatagcg agctaccaaa gccatattca 720

aacacctaga tcactaccac ttctacacag gccactcgag cttgtgatcg cactccgcta 780

aggggcgcc tcttcctctt cgtttcagtc acaacccgca aacggcgcgc catgctgctg 840

caggccttcc tgttcctgct ggccggcttc gccgccaaga tcagcgcctc catgacgaac 900

gagacgtccg accgcccccc ggtgcacttc acccccaaca agggctggat gaacgacccc 960

aacggcctgt ggtacgacga gaaggacgcc aagtggcacc tgtacttcca gtacaacccg 1020

aacgacaccg tctgggggac gcccttgttc tggggccacg ccacgtccga cgacctgacc 1080

aactgggagg accagcccat cgccatcgcc ccgaagcgca acgactccgg cgccttctcc 1140

ggctccatgg tggtggacta caacaacacc tccggcttct tcaacgacac catcgacccg 1200

cgccagcgct gcgtggccat ctggacctac aacaccccgg agtccgagga gcagtacatc 1260

tcctacagcc tggacggcgg ctacaccttc accgagtacc agaagaaccc cgtgctggcc 1320

gccaactcca cccagttccg cgacccgaag gtcttctggt acgagccctc ccagaagtgg 1380

atcatgaccg cggccaagtc ccaggactac aagatcgaga tctactcctc cgacgacctg 1440

aagtcctgga agctggagtc cgcgttcgcc aacgagggct tcctcggcta ccagtacgag 1500

tgccccggcc tgatcgaggt ccccaccgag caggacccca gcaagtccta ctgggtgatg 1560

ttcatctcca tcaaccccgg cgccccggcc ggcggctcct tcaaccagta cttcgtcggc 1620

agcttcaacg gcacccactt cgaggccttc gacaaccagt cccgcgtggt ggacttcggc 1680

aaggactact acgccctgca gaccttcttc aacaccgacc cgacctacgg gagcgccctg 1740

ggcatcgcgt gggcctccaa ctgggagtac tccgccttcg tgcccaccaa cccctggcgc 1800

tcctccatgt ccctcgtgcg caagttctcc ctcaacaccg agtaccaggc caacccggag 1860

acggagctga tcaacctgaa ggccgagccg atcctgaaca tcagcaacgc cggcccctgg 1920

agccggttcg ccaccaacac cacgttgacg aaggccaaca gctacaacgt cgacctgtcc 1980

aacagcaccg gcaccctgga gttcgagctg gtgtacgccg tcaacaccac ccagacgatc 2040

tccaagtccg tgttcgcgga cctctccctc tggttcaagg gcctggagga ccccgaggag 2100

tacctccgca tgggcttcga ggtgtccgcg tcctccttct tcctggaccg cgggaacagc 2160

aaggtgaagt tcgtgaagga gaacccctac ttcaccaacc gcatgagcgt gaacaaccag 2220

cccttcaaga gcgagaacga cctgtcctac tacaaggtgt acggcttgct ggaccagaac 2280

atcctggagc tgtacttcaa cgacggcgac gtcgtgtcca ccaacaccta cttcatgacc 2340

accgggaacg ccctgggctc cgtgaacatg acgacggggg tggacaacct gttctacatc 2400

gacaagttcc aggtgcgcga ggtcaagtga caattggcag cagcagctcg gatagtatcg 2460

acacactctg gacgctggtc gtgtgatgga ctgttgccgc cacacttgct gccttgacct 2520

gtgaatatcc ctgccgcttt tatcaaacag cctcagtgtg tttgatcttg tgtgtacgcg 2580

cttttgcgag ttgctagctg cttgtgctat ttgcgaatac cacccccagc atccccttcc 2640

ctcgtttcat atcgcttgca tcccaaccgc aacttatcta cgctgtcctg ctatccctca 2700

gcgctgctcc tgctcctgct cactgcccct cgcacagcct tggtttgggc tccgcctgta 2760

ttctcctggt actgcaacct gtaaaccagc actgcaatgc tgatgcacgg gaagtagtgg 2820

gatgggaaca caaatggaca gccacggcaa caccgcgcgc cttgcggccg agcacggcga 2880

caagaacctg agcaagatct gcgggctgat cgccagcgac gagggccggc acgagatcgc 2940

ctacacgcgc atcgtggacg agttcttccg cctcgacccc gagggcgccg tcgccgccta 3000

cgccaacatg atgcgcaagc agatcaccat gcccgcgcac ctcatggacg acatgggcca 3060

cggcgaggcc aacccgggcc gcaacctctt cgccgacttc tccgcggtcg ccgagaagat 3120

cgacgtctac gacgccgagg actactgccg catcctggag cacctcaacg cgcgctggaa 3180

ggtggacgag cgccaggtca gcggccaggc cgccgcggac caggagtacg tcctgggcct 3240

gccccagcgc ttccggaaac tcgccgagaa gaccgccgcc aagcgcaagc gcgtcgcgcg 3300

caggcccgtc gccttctcct ggagaagagc ctctagagtc gacctgcagg catgcaagct 3360

tggcgtaatc atggtcatag ctgtttcctg tgtgaaattg ttatccgctc acaattccac 3420

acaacatacg agccggaagc ataaagtgta aagcctgggg tgcctaatga gtgagctaac 3480

tcacattaat tgcgttgcgc tcactgcccg ctttccagtc gggaaacctg tcgtgccagc 3540

tgcattaatg aatcggccaa cgcgcgggga gaggcggttt gcgtattggg cgctcttcc 3599

<210> 81

<211> 2933

<212> DNA

<213> Artificial Sequence

<220><223> synthetic construct

<400> 81

agagagcgga ggtggggttg tgaggtgggg ttgctgacca ggagctcgcg tcgccgagcg 60

cgactcgcac acggtccagt tacccccccc tccgcccaaa cgcaagcctc ccatcttgat 120

gcctttccgg ccacctatac tatttcttag ttcgctgtaa catccagacc gtcctgaata 180

ataacaatgc cctgtgtcaa gtgcattcct aaaaaaattc tgtcccaacc aacaatccca 240

cctgaaatac caccagccct gcccagtaca ctcttccaat accatctccc tacctccacg 300

cgcaagcgac ccccatgcgc gaccaggctc gaaagtgatt tatgacttga gacgagcgag 360

tggcggcgcg gtcgactgcc ttttcatcac gtgccgtacg tcggcgaccg ctagggcttt 420

gcacggcaac gcacggcttc gccaacccga ccagccagga cctcgactac tctaccgcga 480

attcgcctca agaagtcgcc aaatgtgcca tacaccattc cttacagcac tgttcaaact 540
tgatgccaat tttgacattc gggttgctcg ttggctgcgc ccacatcggc cgtgagtgca 600
gcaggcggga tcggacacgg aggacgcggc gtcacgcccc gaacgcagcc cgtaactcta 660
catcaacacg acgtgttgcg taatcccgcc cggctgcgca tcgtgccaac ccattcgcga 720
tggatggtcg gaaaatggtg tgccaactgc cctgagggag gctctcgcga aacgggcacg 780
tccctgaaac cgaaactgtg gccttgtcgt cggccacgca agcacgtgga ccctaaacac 840
caagaaaatc agtaaacaag gttgacatcc tctacgggcg aattgtttgc ccaacccttc 900

atcgcacact gccattataa tgcatctagc tcggcgacaa gtttagaaaa ggcaggctgc 960
attgttccat ttcgccgtgg cggcgtgggt gcccatttta cgaggtttgg gctcccgggc 1020
agcgaccgag ccaggtcgag tccctctcgc ccgtcgacaa tgttgcgaac cccacaagcg 1080
gctaacaaca acttgatggt acctgtacac tgccaattcc ttcttccccg gccgaggttt 1140
acacgtgatg gccatggctt cgcattcagg ccgacttccc attccgactt tccagagggt 1200
ccgcggacgc tgggggttgg ctgcctgagg cccacccttt gttccccgcg tcccgacaaa 1260
cacaattgcg ttacataagg gggagccgcc cccgttcaga gtgcagaaat ctttcactat 1320

attttccagt cgtcagcgaa atcaagtact agtatggcca ccgcatccac tttctcggcg 1380
ttcaatgccc gctgcggcga cctgcgtcgc tcggcgggct ccgggccccg gcgcccagcg 1440
aggcccctcc ccgtgcgcgg gcgcgccccc gactggtcca tgctgttcgc cgtgatcacc 1500
accatcttct ccgccgccga gaagcagtgg accaacctgg agtggaagcc caagcccaac 1560
ccccccagc tgctggacga ccacttcggc ccccacggcc tggtgttccg ccgcaccttc 1620
gccatccgca gctacgaggt gggccccgac cgctccacca gcatcgtggc cgtgatgaac 1680
cacctgcagg aggccgccct gaaccacgcc aagtccgtgg gcatcctggg cgacggcttc 1740

ggcaccaccc tggagatgtc caagcgcgac ctgatctggg tggtgaagcg cacccacgtg 1800
gccgtggagc gctaccccgc ctggggcgac accgtggagg tggagtgctg ggtgggcgcc 1860
tccggcaaca acggccgccg ccacgacttc ctggtgcgcg actgcaagac cggcgagatc 1920
ctgacccgct gcacctccct gagcgtgatg atgaacaccc gcacccgccg cctgagcaag 1980
atccccgagg aggtgcgcgg cgagatcggc cccgccttca tcgacaacgt ggccgtgaag 2040
gacgaggaga tcaagaagcc ccagaagctg aacgactcca ccgccgacta catccagggc 2100
ggcctgaccc cccgctggaa cgacctggac atcaaccagc acgtgaacaa catcaagtac 2160

gtggactgga tcctggagac cgtgcccgac agcatcttcg agagccacca catctcctcc 2220

ttcaccatcg agtaccgccg cgagtgcacc atggacagcg tgctgcagtc cctgaccacc 2280

gtgagcggcg gctcctccga ggccggcctg gtgtgcgagc acctgctgca gctggagggc 2340

ggcagcgagg tgctgcgcgc caagaccgag tggcgcccca agctgaccga ctccttccgc 2400

ggcatcagcg tgatccccgc cgagtccagc gtgatggact acaaggacca cgacggcgac 2460

tacaaggacc acgacatcga ctacaaggac gacgacgaca agtgactcga gttaattaac 2520

tcgaggcagc agcagctcgg atagtatcga cacactctgg acgctggtcg tgtgatggac 2580

tgttgccgcc acacttgctg ccttgacctg tgaatatccc tgccgctttt atcaaacagc 2640

ctcagtgtgt ttgatcttgt gtgtacgcgc ttttgcgagt tgctagctgc ttgtgctatt 2700

tgcgaatacc acccccagca tccccttccc tcgtttcata tcgcttgcat cccaaccgca 2760

acttatctac gctgtcctgc tatccctcag cgctgctcct gctcctgctc actgcccctc 2820

gcacagcctt ggtttgggct ccgcctgtat tctcctggta ctgcaacctg taaaccagca 2880

ctgcaatgct gatgcacggg aagtagtggg atgggaacac aaatggaaag ctt 2933

<210> 82

<211> 726

<212> DNA

<213> Artificial Sequence

<220><223> synthetic construct

<400> 82

gctcttcgcc gccgccactc ctgctcgagc gcgcccgcgc gtgcgccgcc agcgccttgg 60

ccttttcgcc gcgctcgtgc gcgtcgctga tgtccatcac caggtccatg aggtctgcct 120

tgcgccggct gagccactgc ttcgtccggg cggccaagag gagcatgagg gaggactcct 180

ggtccagggt cctgacgtgg tcgcggctct gggagcgggc cagcatcatc tggctctgcc 240

gcaccgaggc cgcctccaac tggtcctcca gcagccgcag tcgccgccga ccctggcaga 300

ggaagacagg tgagggggt atgaattgta cagaacaacc acgagccttg tctaggcaga 360

atccctacca gtcatggctt tacctggatg acggcctgcg aacagctgtc cagcgaccct 420

cgctgccgcc gcttctcccg cacgcttctt tccagcaccg tgatggcgcg agccagcgcc 480

gcacgctggc gctgcgcttc gccgatctga ggacagtcgg ggaactctga tcagtctaaa 540

ccccccttgcg cgttagtgtt gccatccttt gcagaccggt gagagccgac ttgttgtgcg 600

ccacccccca caccacctcc tcccagacca attctgtcac ctttttggcg aaggcatcgg 660

cctcggcctg cagagaggac agcagtgccc agccgctggg ggttggcgga tgcacgctca 720

ggtacc 726

<210> 83

<211> 3635

<212> DNA

<213> Artificial Sequence

<220><223> synthetic construct

<400> 83

ctttcttgcg ctatgacact tccagcaaaa ggtagggcgg gctgcgagac ggcttcccgg 60

cgctgcatgc aacaccgatg atgcttcgac cccccgaagc tccttcgggg ctgcatgggc 120

gctccgatgc cgctccaggg cgagcgctgt ttaaatagcc aggccccccga ttgcaaagac 180

attatagcga gctaccaaag ccatattcaa acacctagat cactaccact tctacacagg 240

ccactcgagc ttgtgatcgc actccgctaa gggggcgcct cttcctcttc gtttcagtca 300

caacccgcaa actctagaat atcaatgatc gagcaggacg gcctccacgc cggctccccc 360

gccgcctggg tggagcgcct gttcggctac gactgggccc agcagaccat cggctgctcc 420

gacgccgccg tgttccgcct gtccgcccag ggccgccccg tgctgttcgt gaagaccgac 480

ctgtccggcg ccctgaacga gctgcaggac gaggccgccc gcctgtcctg gctggccacc 540

accggcgtgc cctgcgccgc cgtgctggac gtggtgaccg aggccggccg cgactggctg 600

ctgctgggcg aggtgcccgg ccaggacctg ctgtcctccc acctggcccc cgccgagaag 660

gtgtccatca tggccgacgc catgcgccgc ctgcacaccc tggaccccgc cacctgcccc 720

ttcgaccacc aggccaagca ccgcatcgag cgcgcccgca cccgcatgga ggccggcctg 780

gtggaccagg acgacctgga cgaggagcac cagggcctgg cccccgccga gctgttcgcc 840

cgcctgaagg cccgcatgcc cgacggcgag gacctggtgg tgacccacgg cgacgcctgc 900

ctgcccaaca tcatggtgga gaacggccgc ttctccggct tcatcgactg cggccgcctg 960

ggcgtggccg accgctacca ggacatcgcc ctggccaccc gcgacatcgc cgaggagctg 1020

ggcggcgagt gggccgaccg cttcctggtg ctgtacggca tcgccgcccc cgactcccag 1080

cgcatcgcct tctaccgcct gctggacgag ttcttctgac aattggcagc agcagctcgg 1140

atagtatcga cacactctgg acgctggtcg tgtgatggac tgttgccgcc acacttgctg 1200

ccttgacctg tgaatatccc tgccgctttt atcaaacagc ctcagtgtgt ttgatcttgt 1260

gtgtacgcgc ttttgcgagt tgctagctgc ttgtgctatt tgcgaatacc acccccagca 1320

tcccccttccc tcgtttcata tcgcttgcat cccaaccgca acttatctac gctgtcctgc 1380

tatccctcag cgctgctcct gctcctgctc actgcccctc gcacagcctt ggtttgggct 1440

ccgcctgtat tctcctggta ctgcaacctg taaaccagca ctgcaatgct gatgcacggg 1500

aagtagtggg atgggaacac aaatggagga tcccgcgtct cgaacagagc gcgcagagga 1560

acgctgaagg tctcgcctct gtcgcacctc agcgcggcat acaccacaat aaccacctga 1620

cgaatgcgct tggttcttcg tccattagcg aagcgtccgg ttcacacacg tgccacgttg 1680

gcgaggtggc aggtgacaat gatcggtgga gctgatggtc gaaacgttca cagcctaggg 1740

atatcgaatt cctttcttgc gctatgacac ttccagcaaa aggtagggcg ggctgcgaga 1800

cggcttcccg gcgctgcatg caacaccgat gatgcttcga ccccccgaag ctccttcggg 1860

gctgcatggg cgctccgatg ccgctccagg gcgagcgctg tttaaatagc caggcccccg 1920

attgcaaaga cattatagcg agctaccaaa gccatattca aacacctaga tcactaccac 1980

ttctacacag gccactcgag cttgtgatcg cactccgcta agggggcgcc tcttcctctt 2040

cgtttcagtc acaacccgca aacactagta tggccaccgc atccactttc tcggcgttca 2100

atgcccgctg cggcgacctg cgtcgctcgg cgggctccgg gccccggcgc ccagcgaggc 2160

ccctccccgt gcgcgggcgc gcccccgact ggtccatgct gttcgccgtg atcaccacca 2220

tcttctccgc cgccgagaag cagtggacca acctggagtg gaagcccaag cccaaccccc 2280

cccagctgct ggacgaccac ttcggccccc acggcctggt gttccgccgc accttcgcca 2340

tccgcagcta cgaggtgggc cccgaccgct ccaccagcat cgtggccgtg atgaaccacc 2400

tgcaggaggc cgccctgaac cacgccaagt ccgtgggcat cctgggcgac ggcttcggca 2460

ccaccctgga gatgtccaag cgcgacctga tctgggtggt gaagcgcacc cacgtggccg 2520

tggagcgcta ccccgcctgg ggcgacaccg tggaggtgga gtgctgggtg ggcgcctccg 2580

gcaacaacgg ccgccgccac gacttcctgg tgcgcgactg caagaccggc gagatcctga 2640

cccgctgcac ctccctgagc gtgatgatga acacccgcac ccgccgcctg agcaagatcc 2700

ccgaggaggt gcgcggcgag atcggccccg ccttcatcga caacgtggcc gtgaaggacg 2760

aggagatcaa gaagccccag aagctgaacg actccaccgc cgactacatc cagggcggcc 2820

tgaccccccg ctggaacgac ctggacatca accagcacgt gaacaacatc aagtacgtgg 2880

actggatcct ggagaccgtg cccgacagca tcttcgagag ccaccacatc tcctccttca 2940

ccatcgagta ccgccgcgag tgcaccatgg acagcgtgct gcagtccctg accaccgtga 3000

gcggcggctc ctccgaggcc ggcctggtgt gcgagcacct gctgcagctg gagggcggca 3060

gcgaggtgct gcgcgccaag accgagtggc gccccaagct gaccgactcc ttccgcggca 3120

tcagcgtgat ccccgccgag tccagcgtga tggactacaa ggaccacgac ggcgactaca 3180

aggaccacga catcgactac aaggacgacg acgacaagtg actcgaggca gcagcagctc 3240

ggatagtatc gacacactct ggacgctggt cgtgtgatgg actgttgccg ccacacttgc 3300

tgccttgacc tgtgaatatc cctgccgctt ttatcaaaca gcctcagtgt gtttgatctt 3360

gtgtgtacgc gcttttgcga gttgctagct gcttgtgcta tttgcgaata ccacccccag 3420

catccccttc cctcgtttca tatcgcttgc atcccaaccg caacttatct acgctgtcct 3480

gctatccctc agcgctgctc ctgctcctgc tcactgcccc tcgcacagcc ttggtttggg 3540

ctccgcctgt attctcctgg tactgcaacc tgtaaaccag cactgcaatg ctgatgcacg 3600

ggaagtagtg ggatgggaac acaaatggaa agctt 3635

<210> 84

<211> 749

<212> DNA

<213> Artificial Sequence

<220><223> synthetic construct

<400> 84

gagctccttg ttttccagaa ggagttgctc cttgagcctt tcattctcag cctcgataac 60

ctccaaagcc gctctaattg tggaggggggt tcgaatttaa aagcttggaa tgttggttcg 120

tgcgtctgga acaagcccag acttgttgct cactgggaaa aggaccatca gctccaaaaa 180

acttgccgct caaaccgcgt acctctgctt tcgcgcaatc tgccctgttg aaatcgccac 240

cacattcata ttgtgacgct tgagcagtct gtaattgcct cagaatgtgg aatcatctgc 300

cccctgtgcg agcccatgcc aggcatgtcg cgggcgagga cacccgccac tcgtacagca 360

gaccattatg ctacctcaca atagttcata acagtgacca tatttctcga agctccccaa 420

cgagcacctc catgctctga gtggccaccc cccggccctg gtgcttgcgg agggcaggtc 480

aaccggcatg gggctaccga aatccccgac cggatcccac cacccccgcg atgggaagaa 540

tctctccccg ggatgtgggc ccaccaccag cacaacctgc tggcccaggc gagcgtcaaa 600

ccataccaca caaatatcct tggcatcggc cctgaattcc ttctgccgct ctgctacccg 660

gtgcttctgt ccgaagcagg ggttgctagg gatcgctccg agtccgcaaa cccttgtcgc 720

gtggcggggc ttgttcgagc ttgaagagc 749

<210> 85

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> synthesized primer

<400> 85

ggaggaattc ggccgacagg acgcgcgtca 30

<210> 86

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> synthesized primer

<400> 86

ggagactagt ggctgcgacc ggcctgtg 28

<210> 87

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> synthesized primer

<400> 87

ggaggaattc tcaccagcgg acaaagcacc g 31

<210> 88

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> synthesized primer

<400> 88

ggagactagt ggctgcgacc ggcctctgg 29

<210> 89

<211> 1065

<212> DNA

<213> Artificial Sequence

<220><223> synthetic promoter

<400> 89

ggccgacagg acgcgcgtca aaggtgctgg tcgtgtatgc cctggccggc aggtcgttgc 60

tgctgctggt tagtgattcc gcaaccctga ttttggcgtc ttattttggc gtggcaaacg 120

ctggcgcccg cgagccgggc cggcggcgat gcggtgcccc acggctgccg gaatccaagg 180

gaggcaagag cgcccgggtc agttgaaggg ctttacgcgc aaggtacagc cgctcctgca 240

aggctgcgtg gtggaattgg acgtgcaggt cctgctgaag ttcctccacc gcctcaccag 300

cggacaaagc accggtgtat caggtccgtg tcatccactc taaagagctc gactacgacc 360

tactgatggc cctagattct tcatcaaaaa cgcctgagac acttgcccag gattgaaact 420

ccctgaaggg accaccaggg gccctgagtt gttccttccc cccgtggcga gctgccagcc 480

aggctgtacc tgtgatcgag gctggcggga aaataggctt cgtgtgctca ggtcatggga 540

ggtgcaggac agctcatgaa acgccaacaa tcgcacaatt catgtcaagc taatcagcta 600

tttcctcttc acgagctgta attgtcccaa aattctggtc taccgggggt gatccttcgt 660

gtacgggccc ttccctcaac cctaggtatg cgcgcatgcg gtcgccgcgc aactcgcgcg 720

agggccgagg gtttgggacg ggccgtcccg aaatgcagtt gcacccggat gcgtggcacc 780

ttttttgcga taatttatgc aatggactgc tctgcaaaat tctggctctg tcgccaaccc 840

taggatcagc ggcgtaggat ttcgtaatca ttcgtcctga tggggagcta ccgactaccc 900

taatatcagc ccgactgcct gacgccagcg tccacttttg tgcacacatt ccattcgtgc 960

ccaagacatt tcattgtggt gcgaagcgtc cccagttacg ctcacctgtt tcccgacctc 1020

cttactgttc tgtcgacaga gcgggcccac aggccggtcg cagcc 1065

<210> 90

<211> 6981

<212> DNA

<213> Cinnamomum camphora

<400> 90

gctcttcggc cgccgccact cctgctcgag cgcgcccgac tcgcgctccg cctgcgcccg 60

cgcgtgcgcc gccagcgcct tggccttttc gccgcgctcg tgcgcgtcgc tgatgtccat 120

caccaggtcc atgaggtctg ccttgcgccg gctgagccac tgcttcgtcc gggcggccaa 180

gaggagcatg agggaggact cctggtccag ggtcctgacg tggtcgcggc tctgggagcg 240

ggccagcatc atctggctct gccgcaccga ggccgcctcc aactggtcct ccagcagccg 300

cagtcgccgc cgaccctggc agaggaagac aggtgagggg ggtatgaatt gtacagaaca 360

accacgagcc ttgtctaggc agaatcccta ccagtcatgg ctttacctgg atgacggcct 420

gcgaacagct gtccagcgac cctcgctgcc gccgcttctc ccgcacgctt ctttccagca 480

ccgtgatggc gcgagccagc gccgcacgct ggcgctgcgc ttcgccgatc tgaggacagt 540

cggggaactc tgatcagtct aaacccccct gcgcgttagt gttgccatcc tttgcagacc 600

ggtgagagcc gacttgttgt gcgccacccc ccacaccacc tcctcccaga ccaattctgt 660

cacctttttg gcgaaggcat cggcctcggc ctgcagagag gacagcagtg cccagccgct 720

gggggttggc ggatgcacgc tcaggtaccc tttcttgcgc tatgacactt ccagcaaaag 780

gtagggcggg ctgcgagacg gcttcccggc gctgcatgca acaccgatga tgcttcgacc 840

ccccgaagct ccttcggggc tgcatgggcg ctccgatgcc gctccagggc gagcgctgtt 900

taaatagcca ggccccccgat tgcaaagaca ttatagcgag ctaccaaagc catattcaaa 960

cacctagatc actaccactt ctacacaggc cactcgagct tgtgatcgca ctccgctaag 1020

ggggcgcctc ttcctcttcg tttcagtcac aacccgcaaa cggcgcgcca tgctgctgca 1080

ggccttcctg ttcctgctgg ccggcttcgc cgccaagatc agcgcctcca tgacgaacga 1140

gacgtccgac cgccccctgg tgcacttcac ccccaacaag ggctggatga acgaccccaa 1200

cggcctgtgg tacgacgaga aggacgccaa gtggcacctg tacttccagt acaacccgaa 1260

cgacaccgtc tgggggacgc ccttgttctg gggccacgcc acgtccgacg acctgaccaa 1320

ctgggaggac cagcccatcg ccatcgcccc gaagcgcaac gactccggcg ccttctccgg 1380

ctccatggtg gtggactaca acaacacctc cggcttcttc aacgacacca tcgacccgcg 1440

ccagcgctgc gtggccatct ggacctacaa caccccggag tccgaggagc agtacatctc 1500

ctacagcctg gacggcggct acaccttcac cgagtaccag aagaaccccg tgctggccgc 1560

caactccacc cagttccgcg acccgaaggt cttctggtac gagccctccc agaagtggat 1620

catgaccgcg gccaagtccc aggactacaa gatcgagatc tactcctccg acgacctgaa 1680

gtcctggaag ctggagtccg cgttcgccaa cgagggcttc ctcggctacc agtacgagtg 1740

ccccggcctg atcgaggtcc ccaccgagca ggaccccagc aagtcctact gggtgatgtt 1800

catctccatc aaccccggcg ccccggccgg cggctccttc aaccagtact tcgtcggcag 1860

cttcaacggc acccacttcg aggccttcga caaccagtcc cgcgtggtgg acttcggcaa 1920

ggactactac gccctgcaga ccttcttcaa caccgacccg acctacggga gcgccctggg 1980

catcgcgtgg gcctccaact gggagtactc cgccttcgtg cccaccaacc cctggcgctc 2040

ctccatgtcc ctcgtgcgca agttctccct caacaccgag taccaggcca acccggagac 2100

ggagctgatc aacctgaagg ccgagccgat cctgaacatc agcaacgccg gcccctggag 2160

ccggttcgcc accaacacca cgttgacgaa ggccaacagc tacaacgtcg acctgtccaa 2220

cagcaccggc accctggagt tcgagctggt gtacgccgtc aacaccaccc agacgatctc 2280

caagtccgtg ttcgcggacc tctccctctg gttcaagggc ctggaggacc ccgaggagta 2340

cctccgcatg ggcttcgagg tgtccgcgtc ctccttcttc ctggaccgcg ggaacagcaa 2400

ggtgaagttc gtgaaggaga acccctactt caccaaccgc atgagcgtga acaaccagcc 2460

cttcaagagc gagaacgacc tgtcctacta caaggtgtac ggcttgctgg accagaacat 2520

cctggagctg tacttcaacg acggcgacgt cgtgtccacc aacacctact tcatgaccac 2580

cgggaacgcc ctgggctccg tgaacatgac gacgggggtg gacaacctgt tctacatcga 2640
caagttccag gtgcgcgagg tcaagtgaca attggcagca gcagctcgga tagtatcgac 2700
acactctgga cgctggtcgt gtgatggact gttgccgcca cacttgctgc cttgacctgt 2760
gaatatccct gccgctttta tcaaacagcc tcagtgtgtt tgatcttgtg tgtacgcgct 2820
tttgcgagtt gctagctgct tgtgctattt gcgaatacca cccccagcat ccccttccct 2880
cgtttcatat cgcttgcatc ccaaccgcaa cttatctacg ctgtcctgct atccctcagc 2940
gctgctcctg ctcctgctca ctgcccctcg cacagccttg gtttgggctc cgcctgtatt 3000

ctcctggtac tgcaacctgt aaaccagcac tgcaatgctg atgcacggga agtagtggga 3060
tgggaacaca aatggaggat cccgcgtctc gaacagagcg cgcagaggaa cgctgaaggt 3120
ctcgcctctg tcgcacctca gcgcggcata caccacaata accacctgac gaatgcgctt 3180
ggttcttcgt ccattagcga agcgtccggt tcacacacgt gccacgttgg cgaggtggca 3240
ggtgacaatg atcggtggag ctgatggtcg aaacgttcac agcctaggga tatcgaattc 3300
ggccgacagg acgcgcgtca aaggtgctgg tcgtgtatgc cctggccggc aggtcgttgc 3360
tgctgctggt tagtgattcc gcaaccctga ttttggcgtc ttattttggc gtggcaaacg 3420

ctggcgcccg cgagccgggc cggcggcgat gcggtgcccc acggctgccg gaatccaagg 3480
gaggcaagag cgcccgggtc agttgaaggg ctttacgcgc aaggtacagc cgctcctgca 3540
aggctgcgtg gtggaattgg acgtgcaggt cctgctgaag ttcctccacc gcctcaccag 3600
cggacaaagc accggtgtat caggtccgtg tcatccactc taaagagctc gactacgacc 3660
tactgatggc cctagattct tcatcaaaaa cgcctgagac acttgcccag gattgaaact 3720
ccctgaaggg accaccaggg gccctgagtt gttccttccc cccgtggcga gctgccagcc 3780
aggctgtacc tgtgatcgag gctggcggga aaataggctt cgtgtgctca ggtcatggga 3840

ggtgcaggac agctcatgaa acgccaacaa tcgcacaatt catgtcaagc taatcagcta 3900
tttcctcttc acgagctgta attgtcccaa aattctggtc taccgggggt gatccttcgt 3960
gtacgggccc ttccctcaac cctaggtatg cgcgcatgcg gtcgccgcgc aactcgcgcg 4020
agggccgagg gtttgggacg ggccgtcccg aaatgcagtt gcacccggat gcgtggcacc 4080
ttttttgcga taatttatgc aatggactgc tctgcaaaat tctggctctg tcgccaaccc 4140
taggatcagc ggcgtaggat ttcgtaatca ttcgtcctga tggggagcta ccgactaccc 4200
taatatcagc ccgactgcct gacgccagcg tccacttttg tgcacacatt ccattcgtgc 4260

ccaagacatt tcattgtggt gcgaagcgtc cccagttacg ctcacctgtt tcccgacctc 4320

cttactgttc tgtcgacaga gcgggcccac aggccggtcg cagccactag tatggccacc 4380

gcatccactt tctcggcgtt caatgcccgc tgcggcgacc tgcgtcgctc ggcgggctcc 4440

gggccccggc gcccagcgag gcccctcccc gtgcgcgggc gcgcccccga ctggtccatg 4500

ctgttcgccg tgatcaccac catcttctcc gccgccgaga agcagtggac caacctggag 4560

tggaagccca agcccaaccc cccccagctg ctggacgacc acttcggccc ccacggcctg 4620

gtgttccgcc gcaccttcgc catccgcagc tacgaggtgg gccccgaccg ctccaccagc 4680

atcgtggccg tgatgaacca cctgcaggag gccgccctga accacgccaa gtccgtgggc 4740

atcctgggcg acggcttcgg caccaccctg gagatgtcca agcgcgacct gatctgggtg 4800

gtgaagcgca cccacgtggc cgtggagcgc taccccgcct ggggcgacac cgtggaggtg 4860

gagtgctggg tgggcgcctc cggcaacaac ggccgccgcc acgacttcct ggtgcgcgac 4920

tgcaagaccg gcgagatcct gacccgctgc acctccctga gcgtgatgat gaacacccgc 4980

acccgccgcc tgagcaagat ccccgaggag gtgcgcggcg agatcggccc cgccttcatc 5040

gacaacgtgg ccgtgaagga cgaggagatc aagaagcccc agaagctgaa cgactccacc 5100

gccgactaca tccagggcgg cctgaccccc cgctggaacg acctggacat caaccagcac 5160

gtgaacaaca tcaagtacgt ggactggatc ctggagaccg tgcccgacag catcttcgag 5220

agccaccaca tctcctcctt caccatcgag taccgccgcg agtgcaccat ggacagcgtg 5280

ctgcagtccc tgaccaccgt gagcggcggc tcctccgagg ccggcctggt gtgcgagcac 5340

ctgctgcagc tggagggcgg cagcgaggtg ctgcgcgcca agaccgagtg gcgccccaag 5400

ctgaccgact ccttccgcgg catcagcgtg atccccgccg agtccagcgt gatggactac 5460

aaggaccacg acggcgacta caaggaccac gacatcgact acaaggacga cgacgacaag 5520

tgactcgagt taattaactc gaggcagcag cagctcggat agtatcgaca cactctggac 5580

gctggtcgtg tgatggactg ttgccgccac acttgctgcc ttgacctgtg aatatccctg 5640

ccgcttttat caaacagcct cagtgtgttt gatcttgtgt gtacgcgctt ttgcgagttg 5700

ctagctgctt gtgctatttg cgaataccac ccccagcatc cccttccctc gtttcatatc 5760

gcttgcatcc caaccgcaac ttatctacgc tgtcctgcta tccctcagcg ctgctcctgc 5820

tcctgctcac tgcccctcgc acagccttgg tttgggctcc gcctgtattc tcctggtact 5880

gcaacctgta aaccagcact gcaatgctga tgcacgggaa gtagtgggat gggaacacaa 5940

atggaaagct gtagagctcc ttgttttcca gaaggagttg ctccttgagc ctttcattct 6000

cagcctcgat aacctccaaa gccgctctaa ttgtggaggg ggttcgaatt taaaagcttg 6060

gaatgttggt tcgtgcgtct ggaacaagcc cagacttgtt gctcactggg aaaaggacca 6120

tcagctccaa aaaacttgcc gctcaaaccg cgtacctctg ctttcgcgca atctgccctg 6180

ttgaaatcgc caccacattc atattgtgac gcttgagcag tctgtaattg cctcagaatg 6240

tggaatcatc tgccccctgt gcgagcccat gccaggcatg tcgcgggcga ggacacccgc 6300

cactcgtaca gcagaccatt atgctacctc acaatagttc ataacagtga ccatatttct 6360

cgaagctccc caacgagcac ctccatgctc tgagtggcca ccccccggcc ctggtgcttg 6420

cggagggcag gtcaaccggc atggggctac cgaaatcccc gaccggatcc caccaccccc 6480

gcgatgggaa gaatctctcc ccgggatgtg ggcccaccac cagcacaacc tgctggccca 6540

ggcgagcgtc aaaccatacc acacaaatat ccttggcatc ggccctgaat tccttctgcc 6600

gctctgctac ccggtgcttc tgtccgaagc aggggttgct agggatcgct ccgagtccgc 6660

aaacccttgt cgcgtggcgg ggcttgttcg agcttgttcg agcttgaaga gcctctagag 6720

tcgacctgca ggcatgcaag cttggcgtaa tcatggtcat agctgtttcc tgtgtgaaat 6780

tgttatccgc tcacaattcc acacaacata cgagccggaa gcataaagtg taaagcctgg 6840

ggtgcctaat gagtgagcta actcacatta attgcgttgc gctcactgcc cgctttccag 6900

tcgggaaacc tgtcgtgcca gctgcattaa tgaatcggcc aacgcgcggg gagaggcggt 6960

ttgcgtattg ggcgctcttc c 6981

<210> 91

<211> 1152

<212> DNA

<213> Cinnamomum camphora

<400> 91

atggccaccg catccacttt ctcggcgttc aatgcccgct gcggcgacct gcgtcgctcg 60

gcgggctccg ggccccggcg cccagcgagg cccctccccg tgcgcgggcg cgccccgac 120

tggtccatgc tgttcgccgt gatcaccacc atcttctccg ccgccgagaa gcagtggacc 180

aacctggagt ggaagcccaa gcccaacccc ccccagctgc tggacgacca cttcggcccc 240

cacggcctgg tgttccgccg caccttcgcc atccgcagct acgaggtggg ccccgaccgc 300

tccaccagca tcgtggccgt gatgaaccac ctgcaggagg ccgccctgaa ccacgccaag 360

tccgtgggca tcctgggcga cggcttcggc accaccctgg agatgtccaa gcgcgacctg 420

atctgggtgg tgaagcgcac ccacgtggcc gtggagcgct accccgcctg gggcgacacc 480

gtggaggtgg agtgctgggt gggcgcctcc ggcaacaacg gccgccgcca cgacttcctg 540

gtgcgcgact gcaagaccgg cgagatcctg acccgctgca cctccctgag cgtgatgatg 600

aacacccgca cccgccgcct gagcaagatc cccgaggagg tgcgcggcga gatcggcccc 660

gccttcatcg acaacgtggc cgtgaaggac gaggagatca agaagcccca gaagctgaac 720

gactccaccg ccgactacat ccagggcggc ctgacccccc gctggaacga cctggacatc 780

aaccagcacg tgaacaacat caagtacgtg gactggatcc tggagaccgt gcccgacagc 840

atcttcgaga gccaccacat ctcctccttc accatcgagt accgccgcga gtgcaccatg 900

gacagcgtgc tgcagtccct gaccaccgtg agcggcggct cctccgaggc cggcctggtg 960

tgcgagcacc tgctgcagct ggagggcggc agcgaggtgc tgcgcgccaa gaccgagtgg 1020

cgccccaagc tgaccgactc cttccgcggc atcagcgtga tccccgccga gtccagcgtg 1080

atggactaca aggaccacga cggcgactac aaggaccacg acatcgacta caaggacgac 1140

gacgacaagt ga 1152

<210> 92

<211> 1052

<212> DNA

<213> Umbellularia californica

<400> 92

ggcgcgcccc cgactggtcc atgctgttcg ccgtgatcac caccatcttc agcgccgccg 60

agaagcagtg gaccaacctg gagtggaagc ccaagcccaa gctgccccag ctgctggacg 120

accacttcgg cctgcacggc ctggtgttcc gccgcacctt cgccatccgc tcctacgagg 180

tgggccccga ccgcagcacc tccatcctgg ccgtgatgaa ccacatgcag gaggccaccc 240

tgaaccacgc caagagcgtg ggcatcctgg gcgacggctt cggcaccacc ctggagatgt 300

ccaagcgcga cctgatgtgg gtggtgcgcc gcacccacgt ggccgtggag cgctacccca 360

cctggggcga caccgtggag gtggagtgct ggatcggcgc cagcggcaac aacggcatgc 420

gccgcgactt cctggtgcgc gactgcaaga ccggcgagat cctgacccgc tgcacctccc 480

tgagcgtgct gatgaacacc cgcacccgcc gcctgagcac catccccgac gaggtgcgcg 540

gcgagatcgg ccccgccttc atcgacaacg tggccgtgaa ggacgacgag atcaagaagc 600

tgcagaagct gaacgactcc accgccgact acatccaggg cggcctgacc ccccgctgga 660

acgacctgga cgtgaaccag cacgtgaaca acctgaagta cgtggcctgg gtgttcgaga 720

ccgtgcccga cagcatcttc gagtcccacc acatcagctc cttcaccctg gagtaccgcc 780

gcgagtgcac ccgcgactcc gtgctgcgca gcctgaccac cgtgagcggc ggcagctccg 840

aggccggcct ggtgtgcgac cacctgctgc agctggaggg cggcagcgag gtgctgcgcg 900

cccgcaccga gtggcgcccc aagctgaccg actccttccg cggcatcagc gtgatccccg 960

ccgagccccg cgtgatggac tacaaggacc acgacggcga ctacaaggac cacgacatcg 1020

actacaagga cgacgacgac aagtgactcg ag 1052

<210> 93

<211> 1031

<212> DNA

<213> Ulmus americana

<400> 93

ggcgcgccca gctgcccgac tggagcatgc tgctggccgc gatcaccacc ctgttcctgg 60

cggccgagaa gcagtggatg atgctggact ggaagcccaa gcgccccgac atgctggtgg 120

acccccttcgg cctgggccgc ttcgtgcagg acggcctggt gttccgcaac aacttcagca 180

tccgcagcta cgagatcggc gcggaccgca ccgccagcat cgagaccctg atgaaccacc 240

tgcaggagac cgccctgaac cacgtgaaga gcgtgggcct gctggaggac ggcctgggca 300

gcacccgcga gatgagcctg cgcaacctga tctgggtggt gaccaagatg caggtggcgg 360

tggaccgcta ccccacctgg ggcgacgagg tgcaggtgag cagctgggcg accgccatcg 420

gcaagaacgg catgcgccgc gagtggatcg tgaccgactt ccgcaccggc gagaccctgc 480

tgcgcgccac cagcgtgtgg gtgatgatga acaagctgac ccgccgcatc agcaagatcc 540

ccgaggaggt gtggcacgag atcggcccca gcttcatcga cgcgcccccc ctgcccaccg 600

tggaggacga cggccgcaag ctgacccgct tcgacgagag cagcgccgac ttcatccgca 660

agggcctgac cccccgctgg agcgacctgg acatcaacca gcacgtgaac aacgtgaagt 720

acatcggctg gctgctggag agcgcgcccc ccgagatcca cgagagccac gagatcgcca 780

gcctgaccct ggagtaccgc cgcgagtgcg gccgcgacag cgtgctgaac agcgccacca 840

aggtgagcga cagcagccag ctgggcaaga gcgccgtgga gtgcaaccac ctggtgcgcc 900

tgcagaacgg cggcgagatc gtgaagggcc gcaccgtgtg gcgccccaag cgccccctgt 960

acaacgacgg cgccgtggtg gacgtgcccg ccaagaccag cgatgacgat gacaagctgg 1020

gatgactcga g 1031

<210> 94

<211> 1182

<212> DNA

<213> Cuphea hookeriana

<400> 94

actagtatgg ctatcaagac gaacaggcag cctgtggaga agcctccgtt cacgatcggg 60

acgctgcgca aggccatccc cgcgcactgt ttcgagcgct cggcgcttcg tgggcgcgcc 120

cagctgcccg actggagccg cctgctgacc gccatcacca ccgtgttcgt gaagtccaag 180

cgccccgaca tgcacgaccg caagtccaag cgccccgaca tgctggtgga cagcttcggc 240

ctggagtcca ccgtgcagga cggcctggtg ttccgccagt ccttctccat ccgctcctac 300

gagatcggca ccgaccgcac cgccagcatc gagaccctga tgaaccacct gcaggagacc 360

tccctgaacc actgcaagag caccggcatc ctgctggacg gcttcggccg caccctggag 420

atgtgcaagc gcgacctgat ctgggtggtg attaagatgc agatcaaggt gaaccgctac 480

cccgcctggg gcgacaccgt ggagatcaac acccgcttca gccgcctggg caagatcggc 540

atgggccgcg actggctgat ctccgactgc aacaccggcg agatcctggt gcgcgccacc 600

agcgcctacg ccatgatgaa ccagaagacc cgccgcctgt ccaagctgcc ctacgaggtg 660

caccaggaga tcgtgccccc gttcgtggac agccccgtga tcgaggactc cgacctgaag 720

gtgcacaagt tcaaggtgaa gaccggcgac agcatccaga agggcctgac ccccggctgg 780

aacgacctgg acgtgaacca gcacgtgtcc aacgtgaagt acatcggctg gatcctggag 840

agcatgccca ccgaggtgct ggagacccag gagctgtgct ccctggccct ggagtaccgc 900

cgcgagtgcg gccgcgactc cgtgctggag agcgtgaccg ccatggaccc cagcaaggtg 960

ggcgtgcgct cccagtacca gcacctgctg cgcctggagg acggcaccgc catcgtgaac 1020

ggcgccaccg agtggcgccc caagaacgcc ggcgccaacg gcgccatctc caccggcaag 1080

accagcaacg gcaactccgt gtccatggac tacaaggacc acgacggcga ctacaaggac 1140

cacgacatcg actacaagga cgacgacgac aagtgactcg ag 1182

<210> 95

<211> 20

<212> DNA

<213> Umbellularia californica

<400> 95

ctgggcgacg gcttcggcac 20

<210> 96

<211> 20

<212> DNA

<213> Umbellularia californica

<400> 96

aagtcgcggc gcatgccgtt 20

<210> 97

<211> 20

<212> DNA

<213> Ulmus americana

<400> 97

cccagctgct cacctgcacc 20

<210> 98

<211> 26

<212> DNA

<213>

Ulmus americana

<400> 98

cacccaaggc caacggcagc gccgtg 26

<210> 99

<211> 20

<212> DNA

<213> Cuphea hookeriana

<400> 99

taccccgcct ggggcgacac 20

<210> 100

<211> 20

<212> DNA

<213> Cuphea hookeriana

<400> 100

agcttggaca ggcggcgggt 20

<210> 101

<211> 20

<212> DNA

<213> Prototheca moriformis

<400> 101

tcacttcatg ccggcggtcc 20

<210> 102

<211> 20

<212> DNA

<213> Prototheca moriformis

<400> 102

gcgctcctgc ttggctcgaa 20

<210> 103

<211> 733

<212> DNA

<213> Artificial Sequence

<220><223> synthetic polynucleotide

<400> 103

gctcttcgag acgtggtctg aatcctccag gcgggtttcc ccgagaaaga aagggtgccg 60

atttcaaagc agagccatgt gccgggccct gtggcctgtg ttggcgccta tgtagtcacc 120

ccccctcacc caattgtcgc cagtttgcgc aatccataaa ctcaaaactg cagcttctga 180

gctgcgctgt tcaagaacac ctctggggtt tgctcacccg cgaggtcgac gcccagcatg 240

gctatcaaga cgaacaggca gcctgtggag aagcctccgt tcacgatcgg gacgctgcgc 300

aaggccatcc ccgcgcactg tttcgagcgc tcggcgcttc gtagcagcat gtacctggcc 360

tttgacatcg cggtcatgtc cctgctctac gtcgcgtcga cgtacatcga ccctgcgccg 420

gtgcctacgt gggtcaagta tggcgtcatg tggccgctct actggttctt ccaggtgtgt 480

gtgagggttg tggttgcccg tatcgaggtc ctggtggcgc gcatgggggа gaaggcgcct 540

gtcccgctga ccccccggc taccctcccg gcaccttcca gggcgccttc ggcacgggtg 600

tctgggtgtg cgcgcacgag tgcggccacc aggccttttc ctccagccag gccatcaacg 660

acggcgtggg cctggtgttc cacagcctgc tgctggtgcc ctactactcc tggaagcact 720

cgcaccgggt acc 733

<210> 104

<211> 739

<212> DNA

<213> Artificial Sequence

<220><223> synthetic polynucleotide

<400> 104

ccgccaccac tccaacacgg ggtgcctgga caaggacgag gtgtttgtgc cgccgcaccg 60

cgcagtggcg cacgagggcc tggagtggga ggagtggctg cccatccgca tgggcaaggt 120

gctggtcacc ctgaccctgg gctggccgct gtacctcatg ttcaacgtcg cctcgcggcc 180

gtacccgcgc ttcgccaacc actttgaccc gtggtcgccc atcttcagca agcgcgagcg 240

catcgaggtg gtcatctccg acctggcgct ggtggcggtg ctcagcgggc tcagcgtgct 300

gggccgcacc atgggctggg cctggctggt caagacctac gtggtgccct acctgatcgt 360

gaacatgtgg ctcgtgctca tcacgctgct ccagcacacg cacccggcgc tgccgcacta 420

cttcgagaag gactgggact ggctgcgcgg cgccatggcc accgtggacc gctccatggg 480

cccgcccttc atggacaaca tcctgcacca catctccgac acccacgtgc tgcaccacct 540

cttcagcacc atcccgcact accacgccga ggaggcctcc gccgccatca ggcccatcct 600

gggcaagtac taccagtccg acagccgctg ggtcggccgc gccctgtggg aggactggcg 660

cgactgccgc tacgtcgtcc cggacgcgcc cgaggacgac tccgcgctct ggttccacaa 720

gtgagtgagt gagaagagc 739

<210> 105

<211> 2327

<212> DNA

<213> Saccharomyces cerevisiae

<400> 105

ctttcttgcg ctatgacact tccagcaaaa ggtagggcgg gctgcgagac ggcttcccgg 60

cgctgcatgc aacaccgatg atgcttcgac cccccgaagc tccttcgggg ctgcatgggc 120

gctccgatgc cgctccaggg cgagcgctgt ttaaatagcc aggccccccga ttgcaaagac 180

attatagcga gctaccaaag ccatattcaa acacctagat cactaccact tctacacagg 240

ccactcgagc ttgtgatcgc actccgctaa gggggcgcct cttcctcttc gtttcagtca 300

caacccgcaa acggcgcgcc atgctgctgc aggccttcct gttcctgctg gccggcttcg 360

ccgccaagat cagcgcctcc atgacgaacg agacgtccga ccgcccccctg gtgcacttca 420

cccccaacaa gggctggatg aacgacccca acggcctgtg gtacgacgag aaggacgcca 480

agtggcacct gtacttccag tacaacccga acgacaccgt ctgggggacg cccttgttct 540

ggggccacgc cacgtccgac gacctgacca actgggagga ccagcccatc gccatcgccc 600

cgaagcgcaa cgactccggc gccttctccg gctccatggt ggtggactac aacaacacct 660

ccggcttctt caacgacacc atcgacccgc gccagcgctg cgtggccatc tggacctaca 720

acaccccgga gtccgaggag cagtacatct cctacagcct ggacggcggc tacaccttca 780

ccgagtacca gaagaacccc gtgctggccg ccaactccac ccagttccgc gacccgaagg 840

tcttctggta cgagccctcc cagaagtgga tcatgaccgc ggccaagtcc caggactaca 900

agatcgagat ctactcctcc gacgacctga agtcctggaa gctggagtcc gcgttcgcca 960

acgagggctt cctcggctac cagtacgagt gccccggcct gatcgaggtc cccaccgagc 1020

aggaccccag caagtcctac tgggtgatgt tcatctccat caaccccggc gccccggccg 1080

gcggctcctt caaccagtac ttcgtcggca gcttcaacgg cacccacttc gaggccttcg 1140

acaaccagtc ccgcgtggtg gacttcggca aggactacta cgccctgcag accttcttca 1200

acaccgaccc gacctacggg agcgccctgg gcatcgcgtg ggcctccaac tgggagtact 1260

ccgccttcgt gcccaccaac ccctggcgct cctccatgtc cctcgtgcgc aagttctccc 1320

tcaacaccga gtaccaggcc aacccggaga cggagctgat caacctgaag gccgagccga 1380

tcctgaacat cagcaacgcc ggcccctgga gccggttcgc caccaacacc acgttgacga 1440

aggccaacag ctacaacgtc gacctgtcca acagcaccgg caccctggag ttcgagctgg 1500

tgtacgccgt caacaccacc cagacgatct ccaagtccgt gttcgcggac ctctccctct 1560

ggttcaaggg cctggaggac cccgaggagt acctccgcat gggcttcgag gtgtccgcgt 1620

cctccttctt cctggaccgc gggaacagca aggtgaagtt cgtgaaggag aacccctact 1680

tcaccaaccg catgagcgtg aacaaccagc ccttcaagag cgagaacgac ctgtcctact 1740

acaaggtgta cggcttgctg gaccagaaca tcctggagct gtacttcaac gacggcgacg 1800

tcgtgtccac caacacctac ttcatgacca ccgggaacgc cctgggctcc gtgaacatga 1860

cgacgggggt ggacaacctg ttctacatcg acaagttcca ggtgcgcgag gtcaagtgac 1920

aattggcagc agcagctcgg atagtatcga cacactctgg acgctggtcg tgtgatggac 1980

tgttgccgcc acacttgctg ccttgacctg tgaatatccc tgccgctttt atcaaacagc 2040

ctcagtgtgt ttgatcttgt gtgtacgcgc ttttgcgagt tgctagctgc ttgtgctatt 2100

tgcgaatacc acccccagca tcccccttccc tcgtttcata tcgcttgcat cccaaccgca 2160

acttatctac gctgtcctgc tatccctcag cgctgctcct gctcctgctc actgcccctc 2220

gcacagcctt ggtttgggct ccgcctgtat tctcctggta ctgcaacctg taaaccagca 2280

ctgcaatgct gatgcacggg aagtagtggg atgggaacac aaatgga 2327

<210> 106

<211> 735

<212> DNA

<213> Artificial Sequence

<220><223> synthetic polynucleotide

<400> 106

gctcttcgag gggctggtct gaatccttca ggcgggtgtt acccgagaaa gaaagggtgc 60

cgatttcaaa gcagacccat gtgccgggcc ctgtggcctg tgttggcgcc tatgtagtca 120

ccccccctca cccaattgtc gccagtttgc gcactccata aactcaaaac agcagcttct 180

gagctgcgct gttcaagaac acctctgggg tttgctcacc cgcgaggtcg acgcccagca 240

tggctatcaa gacgaacagg cagcctgtgg agaagcctcc gttcacgatc gggacgctgc 300

gcaaggccat ccccgcgcac tgtttcgagc gctcggcgct tcgtagcagc atgtacctgg 360

cctttgacat cgcggtcatg tccctgctct acgtcgcgtc gacgtacatc gaccctgcac 420

cggtgcctac gtgggtcaag tacggcatca tgtggccgct ctactggttc ttccaggtgt 480

gtttgagggt tttggttgcc cgtattgagg tcctggtggc gcgcatggag gagaaggcgc 540

ctgtcccgct gacccccccg gctaccctcc cggcaccttc cagggcgcct tcggcacggg 600

tgtctgggtg tgcgcgcacg agtgcggcca ccaggccttt tcctccagcc aggccatcaa 660

cgacggcgtg ggcctggtgt tccacagcct gctgctggtg ccctactact cctggaagca 720

ctcgcaccgg gtacc 735

<210> 107

<211> 739

<212> DNA

<213> Artificial Sequence

<220><223> synthetic polynucleotide

<400> 107

ccgccaccac tccaacacgg ggtgcctgga caaggacgag gtgtttgtgc cgccgcaccg 60

cgcagtggcg cacgagggcc tggagtggga ggagtggctg cccatccgca tgggcaaggt 120

gctggtcacc ctgaccctgg gctggccgct gtacctcatg ttcaacgtcg cctcgcggcc 180

gtacccgcgc ttcgccaacc actttgaccc gtggtcgccc atcttcagca agcgcgagcg 240

catcgaggtg gtcatctccg acctggcgct ggtggcggtg ctcagcgggc tcagcgtgct 300

gggccgcacc atgggctggg cctggctggt caagacctac gtggtgccct acctgatcgt 360

gaacatgtgg ctcgtgctca tcacgctgct ccagcacacg cacccggcgc tgccgcacta 420

cttcgagaag gactgggact ggctgcgcgg cgccatggcc accgtggacc gctccatggg 480

cccgcccttc atggacaaca tcctgcacca catctccgac acccacgtgc tgcaccacct 540

cttcagcacc atcccgcact accacgccga ggaggcctcc gccgccatca ggcccatcct 600

gggcaagtac taccagtccg acagccgctg ggtcggccgc gccctgtggg aggactggcg 660

cgactgccgc tacgtcgtcc cggacgcgcc cgaggacgac tccgcgctct ggttccacaa 720

gtgagtgagt gagaagagc 739

<210> 108

<211> 471

<212> PRT

<213> Saccharomyces carlbergensis

<400> 108

Met Phe Ala Phe Tyr Phe Leu Thr Ala Cys Ile Ser Leu Lys Gly Val

1 5 10 15

Phe Gly Val Ser Pro Ser Tyr Asn Gly Leu Gly Leu Thr Pro Gln Met

20 25 30

Gly Trp Asp Asn Trp Asn Thr Phe Ala Cys Asp Val Ser Glu Gln Leu

35 40 45

Leu Leu Asp Thr Ala Asp Arg Ile Ser Asp Leu Gly Leu Lys Asp Met

50 55 60

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

65 70 75 80

Ser Asp Gly Phe Leu Val Ala Asp Glu Gln Lys Phe Pro Asn Gly Met

85 90 95

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

100 105 110

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

115 120 125

Gly Arg Glu Glu Glu Asp Ala Gln Phe Phe Ala Asn Asn Arg Val Asp

130 135 140

Tyr Leu Lys Tyr Asp Asn Cys Tyr Asn Lys Gly Gln Phe Gly Thr Pro

145 150 155 160

Glu Ile Ser Tyr His Arg Tyr Lys Ala Met Ser Asp Ala Leu Asn Lys

165 170 175

Thr Gly Arg Pro Ile Phe Tyr Ser Leu Cys Asn Trp Gly Gln Asp Leu

180 185 190

Thr Phe Tyr Trp Gly Ser Gly Ile Ala Asn Ser Trp Arg Met Ser Gly

195 200 205

Asp Val Thr Ala Glu Phe Thr Arg Pro Asp Ser Arg Cys Pro Cys Asp

210 215 220

Gly Asp Glu Tyr Asp Cys Lys Tyr Ala Gly Phe His Cys Ser Ile Met

225 230 235 240

Asn Ile Leu Asn Lys Ala Ala Pro Met Gly Gln Asn Ala Gly Val Gly

245 250 255

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

305 310 315 320
Gly Ile Pro Ala Thr Arg Val Trp Arg Tyr Tyr Val Ser Asp Thr Asp
325 330 335
Glu Tyr Gly Gln Gly Glu Ile Gln Met Trp Ser Gly Pro Leu Asp Asn
340 345 350
Gly Asp Gln Val Val Ala Leu Leu Asn Gly Gly Ser Val Ser Arg Pro
355 360 365
Met Asn Thr Thr Leu Glu Glu Ile Phe Phe Asp Ser Asn Leu Gly Ser

370 375 380
Lys Lys Leu Thr Ser Thr Trp Asp Ile Tyr Asp Leu Trp Ala Asn Arg
385 390 395 400
Val Asp Asn Ser Thr Ala Ser Ala Ile Leu Gly Arg Asn Lys Thr Ala
405 410 415
Thr Gly Ile Leu Tyr Asn Ala Thr Glu Gln Ser Tyr Lys Asp Gly Leu
420 425 430
Ser Lys Asn Asp Thr Arg Leu Phe Gly Gln Lys Ile Gly Ser Leu Ser

435 440 445
Pro Asn Ala Ile Leu Asn Thr Thr Val Pro Ala His Gly Ile Ala Phe
450 455 460
Tyr Arg Leu Arg Pro Ser Ser
465 470
<210> 109
<211> 18
<212> PRT
<213> Saccharomyces carlbergensis

<400> 109

Met Phe Ala Phe Tyr Phe Leu Thr Ala Cys Ile Ser Leu Lys Gly Val
1 5 10 15

Phe Gly

<210> 110

<211> 3183

<212> DNA

<213> Saccharomyces carlbergensis

<400> 110

gcggccgcgt ggacgagggc tacaaccccg cctacggcgc gcgcccgctg cgccgcgcca 60

tcatgcgcct gctggaggac gcgctggccg agcgcatgct cgccggcgac gtcaaggagg 120

gcgactcggt catcatggac gtcgatggcg atggcgccat cagcgtcctc aacggagacc 180

gcacccacac caccaccatc gactcctccc cggcgggcat ctcgtagacg cgagagggag 240

ggagggagga ttttcagagg gagatgagac gaggactggg ccgggggcct tggggcgctg 300

ctggagcgtg gtgagagcgc ggcggacgtg ccttttcttc ttccgtgcgc gcgctcttgg 360

ccattgatcc ccgattcgcg cccgcatccc cccactgccc ccatcatctt gcctgttgtc 420

gtggcactga cataaacccc ctgcgctgcg ctgctccgct actattgata taggtctcac 480

gcgccaatct tttttgctcc gggtaaccgt ctggacgcca gaattccttt cttgcgctat 540

gacacttcca gcaaaaggta gggcgggctg cgagacggct tcccggcgct gcatgcaaca 600

ccgatgatgc ttcgaccccc cgaagctcct tcggggctgc atgggcgctc cgatgccgct 660

ccagggcgag cgctgtttaa atagccaggc ccccgattgc aaagacatta tagcgagcta 720

ccaaagccat attcaaacac ctagatcact accacttcta cacaggccac tcgagcttgt 780

gatcgcactc cgctaagggg gcgcctcttc ctcttcgttt cagtcacaac ccgcaaacac 840

tagtatgttc gcgttctact tcctgacggc ctgcatctcc ctgaagggcg tgttcggcgt 900

ctccccctcc tacaacggcc tgggcctgac gccccagatg ggctgggaca actggaacac 960

gttcgcctgc gacgtctccg agcagctgct gctggacacg gccgaccgca tctccgacct 1020

gggcctgaag gacatgggct acaagtacat catcctggac gactgctggt cctccggccg 1080

cgactccgac ggcttcctgg tcgccgacga gcagaagttc cccaacggca tgggccacgt 1140

cgccgaccac ctgcacaaca actccttcct gttcggcatg tactcctccg cgggcgagta 1200

cacgtgcgcc ggctaccccg gctccctggg ccgcgaggag gaggacgccc agttcttcgc 1260

gaacaaccgc gtggactacc tgaagtacga caactgctac aacaagggcc agttcggcac 1320

gcccgagatc tcctaccacc gctacaaggc catgtccgac gccctgaaca agacgggccg 1380

ccccatcttc tactccctgt gcaactgggg ccaggacctg accttctact ggggctccgg 1440

catcgcgaac tcctggcgca tgtccggcga cgtcacggcg gagttcacgc gccccgactc 1500

ccgctgcccc tgcgacggcg acgagtacga ctgcaagtac gccggcttcc actgctccat 1560

catgaacatc ctgaacaagg ccgcccccat gggccagaac gcgggcgtcg gcggctggaa 1620

cgacctggac aacctggagg tcggcgtcgg caacctgacg gacgacgagg agaaggcgca 1680

cttctccatg tgggccatgg tgaagtcccc cctgatcatc ggcgcgaacg tgaacaacct 1740

gaaggcctcc tcctactcca tctactccca ggcgtccgtc atcgccatca accaggactc 1800

caacggcatc cccgccacgc gcgtctggcg ctactacgtg tccgacacgg acgagtacgg 1860

ccagggcgag atccagatgt ggtccggccc cctggacaac ggcgaccagg tcgtggcgct 1920

gctgaacggc ggctccgtgt cccgccccat gaacacgacc ctggaggaga tcttcttcga 1980

ctccaacctg ggctccaaga agctgacctc cacctgggac atctacgacc tgtgggcgaa 2040

ccgcgtcgac aactccacgg cgtccgccat cctgggccgc aacaagaccg ccaccggcat 2100

cctgtacaac gccaccgagc agtcctacaa ggacggcctg tccaagaacg acacccgcct 2160

gttcggccag aagatcggct ccctgtcccc caacgcgatc ctgaacacga ccgtccccgc 2220

ccacggcatc gcgttctacc gcctgcgccc ctcctcctga caattgaagc agcagcagct 2280

cggatagtat cgacacactc tggacgctgg tcgtgtgatg gactgttgcc gccacacttg 2340

ctgccttgac ctgtgaatat ccctgccgct tttatcaaac agcctcagtg tgtttgatct 2400

tgtgtgtacg cgcttttgcg agttgctagc tgcttgtgct atttgcgaat accacccccа 2460

gcatcccctt ccctcgtttc atatcgcttg catcccaacc gcaacttatc tacgctgtcc 2520

tgctatccct cagcgctgct cctgctcctg ctcactgccc ctcgcacagc cttggtttgg 2580

gctccgcctg tattctcctg gtactgcaac ctgtaaacca gcactgcaat gctgatgcac 2640

gggaagtagt gggatgggaa cacaaatgga aagcttgagc tcagaatagt atcgggtgat 2700

gcgaagtcag aaccaggcag ggcctgtcgc ctgaggtggc aacgatggga agcaatcaat 2760

ctgggtacag tcgtccgcac gatcccgtga tctcccccac cgacacctat ccccgcccat 2820

cccggcccac cctttcagtc ccctcagcat gcattgtgca ccgcgacaaa gcatgtctgc 2880

tcgtgcactg gttcaggcca cggcgcaccg agtcctcgcc cttcgcagag tgatcaccct 2940

ccccggaacc agccacgctc gctgctgcgg gccgatcagc cgcgcgcact ccctgcaact 3000

agggacaact caggcaacca cgcgcctcac aagcatggcc gccgtggcat ccaacccgct 3060

cgtgacggtg ggtgcgcaag tgccaggggc ctcgtcgtca cggcgtgcat cctcgaggga 3120

tgcgatccgg caactatatg tcgtttatct ccccaccaat cacaggatga gcccctgtct 3180

aga 3183

<210> 111

<211> 1416

<212> DNA

<213> Saccharomyces carlbergensis

<400> 111

atgttcgcgt tctacttcct gacggcctgc atctccctga agggcgtgtt cggcgtctcc 60

ccctcctaca acggcctggg cctgacgccc cagatgggct gggacaactg gaacacgttc 120

gcctgcgacg tctccgagca gctgctgctg gacacggccg accgcatctc cgacctgggc 180

ctgaaggaca tgggctacaa gtacatcatc ctggacgact gctggtcctc cggccgcgac 240

tccgacggct tcctggtcgc cgacgagcag aagttcccca acggcatggg ccacgtcgcc 300

gaccacctgc acaacaactc cttcctgttc ggcatgtact cctccgcggg cgagtacacg 360

tgcgccggct accccggctc cctgggccgc gaggaggagg acgcccagtt cttcgcgaac 420

aaccgcgtgg actacctgaa gtacgacaac tgctacaaca agggccagtt cggcacgccc 480

gagatctcct accaccgcta caaggccatg tccgacgccc tgaacaagac gggccgcccc 540

atcttctact ccctgtgcaa ctggggccag gacctgacct tctactgggg ctccggcatc 600

gcgaactcct ggcgcatgtc cggcgacgtc acggcggagt tcacgcgccc cgactcccgc 660

tgcccctgcg acggcgacga gtacgactgc aagtacgccg gcttccactg ctccatcatg 720

aacatcctga acaaggccgc ccccatgggc cagaacgcgg gcgtcggcgg ctggaacgac 780

ctggacaacc tggaggtcgg cgtcggcaac ctgacggacg acgaggagaa ggcgcacttc 840

tccatgtggg ccatggtgaa gtccccccctg atcatcggcg cgaacgtgaa caacctgaag 900

gcctcctcct actccatcta ctcccaggcg tccgtcatcg ccatcaacca ggactccaac 960

ggcatccccg ccacgcgcgt ctggcgctac tacgtgtccg acacggacga gtacggccag 1020

ggcgagatcc agatgtggtc cggcccccctg gacaacggcg accaggtcgt ggcgctgctg 1080

aacggcggct ccgtgtcccg ccccatgaac acgaccctgg aggagatctt cttcgactcc 1140

aacctgggct ccaagaagct gacctccacc tgggacatct acgacctgtg ggcgaaccgc 1200

gtcgacaact ccacggcgtc cgccatcctg ggccgcaaca agaccgccac cggcatcctg 1260

tacaacgcca ccgagcagtc ctacaaggac ggcctgtcca agaacgacac ccgcctgttc 1320

ggccagaaga tcggctccct gtcccccaac gcgatcctga acacgaccgt ccccgcccac 1380

ggcatcgcgt tctaccgcct gcgcccctcc tcctga 1416

<210> 112

<211> 526

<212> DNA

<213> Artificial Sequence

<220><223> synthetic construct

<400> 112

gcggccgcgt ggacgagggc tacaaccccg cctacggcgc gcgcccgctg cgccgcgcca 60

tcatgcgcct gctggaggac gcgctggccg agcgcatgct cgccggcgac gtcaaggagg 120

gcgactcggt catcatggac gtcgatggcg atggcgccat cagcgtcctc aacggagacc 180

gcacccacac caccaccatc gactcctccc cggcgggcat ctcgtagacg cgagagggag 240

ggagggagga ttttcagagg gagatgagac gaggactggg ccgggggcct tggggcgctg 300

ctggagcgtg gtgagagcgc ggcggacgtg ccttttcttc ttccgtgcgc gcgctcttgg 360

ccattgatcc ccgattcgcg cccgcatccc cccactgccc ccatcatctt gcctgttgtc 420

gtggcactga cataaacccc ctgcgctgcg ctgctccgct actattgata taggtctcac 480

gcgccaatct tttttgctcc gggtaaccgt ctggacgcca gaattc 526

<210> 113

<211> 507

<212> DNA

<213> Artificial Sequence

<220><223> synthetic construct

<400> 113

gagctcagaa tagtatcggg tgatgcgaag tcagaaccag gcagggcctg tcgcctgagg 60

tggcaacgat gggaagcaat caatctgggt acagtcgtcc gcacgatccc gtgatctccc 120

ccaccgacac ctatccccgc ccatcccggc ccacccttcc agtcccctca gcatgcattg 180

tgcaccgcga caaagcatgt ctgctcgtgc actggttcag gccacggcgc accgagtcct 240

cgcccttcgc agagtgatca ccctccccgg aaccagccac gctcgctgct gcgggccgat 300

cagccgcgcg cactccctgc aactagggac aactcaggca accacgcgcc tcacaagcat 360

ggccgccgtg gcatccaacc cgctcgtgac ggtgggtgcg caagtgccag gggcctcgtc 420

gtcacggcgt gcatcctcga gggatgcgat ccggcaacta tatgtcgttt atctccccac 480

caatcacagg atgagcccct gtctaga 507

<210> 114

<211> 312

<212> DNA

<213> Chlamydomonas reinhardtii

<400> 114

ctttcttgcg ctatgacact tccagcaaaa ggtagggcgg gctgcgagac ggcttcccgg 60

cgctgcatgc aacaccgatg atgcttcgac cccccgaagc tccttcgggg ctgcatgggc 120

gctccgatgc cgctccaggg cgagcgctgt ttaaatagcc aggccccccga ttgcaaagac 180

attatagcga gctaccaaag ccatattcaa acacctagat cactaccact tctacacagg 240

ccactcgagc ttgtgatcgc actccgctaa gggggcgcct cttcctcttc gtttcagtca 300

caacccgcaa ac 312

<210> 115

<211> 408

<212> DNA

<213> Chlorella vulgaris

<400> 115

gcagcagcag ctcggatagt atcgacacac tctggacgct ggtcgtgtga tggactgttg 60

ccgccacact tgctgccttg acctgtgaat atccctgccg cttttatcaa acagcctcag 120

tgtgtttgat cttgtgtgta cgcgcttttg cgagttgcta gctgcttgtg ctatttgcga 180

ataccacccc cagcatcccc ttccctcgtt tcatatcgct tgcatcccaa ccgcaactta 240

tctacgctgt cctgctatcc ctcagcgctg ctcctgctcc tgctcactgc ccctcgcaca 300

gccttggttt gggctccgcc tgtattctcc tggtactgca acctgtaaac cagcactgca 360

atgctgatgc acgggaagta gtgggatggg aacacaaatg gaaagctt 408

<210> 116

<211> 747

<212> PRT

<213> Aspergillus niger

<400> 116

Met Ile Gly Ser Ser His Ala Val Val Ala Leu Gly Leu Phe Thr Leu
1 5 10 15

Tyr Gly His Ser Ala Ala Ala Pro Ala Ile Gly Ala Ser Asn Ser Gln
20 25 30

Thr Ile Val Thr Asn Gly Thr Ser Phe Ala Leu Asn Gly Asp Asn Val
35 40 45

Ser Tyr Arg Phe His Val Asn Ser Ser Thr Gly Asp Leu Ile Ser Asp
50 55 60
His Phe Gly Gly Val Val Ser Gly Thr Ile Pro Ser Pro Val Glu Pro
65 70 75 80
Ala Val Asn Gly Trp Val Gly Met Pro Gly Arg Ile Arg Arg Glu Phe
85 90 95

Pro Asp Gln Gly Arg Gly Asp Phe Arg Ile Pro Ala Val Arg Ile Arg
100 105 110
Glu Ser Ala Gly Tyr Thr Val Ser Asp Leu Gln Tyr Val Ser His Glu
115 120 125
Val Ile Glu Gly Lys Tyr Ala Leu Pro Gly Leu Pro Ala Thr Phe Gly
130 135 140
Asp Ala Gln Asp Ala Thr Thr Leu Val Val His Leu Tyr Asp Asn Tyr
145 150 155 160

Ser Ser Val Ala Ala Asp Leu Ser Tyr Ser Ile Phe Pro Lys Tyr Asp
165 170 175
Ala Ile Val Arg Ser Val Asn Val Thr Asn Gln Gly Pro Gly Asn Ile
180 185 190
Thr Ile Glu Ala Leu Ala Ser Ile Ser Ile Asp Phe Pro Tyr Glu Asp
195 200 205
Leu Asp Met Val Ser Leu Arg Gly Asp Trp Ala Arg Glu Ala Asn Val
210 215 220

Gln Arg Ser Lys Val Gln Tyr Gly Val Gln Gly Phe Gly Ser Ser Thr
225 230 235 240
Gly Tyr Ser Ser His Leu His Asn Pro Phe Leu Ala Ile Val Asp Pro
245 250 255
Ala Thr Thr Glu Ser Gln Gly Glu Ala Trp Gly Phe Asn Leu Val Tyr
260 265 270
Thr Gly Ser Phe Ser Ala Gln Val Glu Lys Gly Ser Gln Gly Phe Thr
275 280 285

Arg Ala Leu Leu Gly Phe Asn Pro Asp Gln Leu Ser Trp Asn Leu Gly

290 295 300

Pro Gly Glu Thr Leu Thr Ser Pro Glu Cys Val Ala Val Tyr Ser Asp
305 310 315 320

Lys Gly Leu Gly Ser Val Ser Arg Lys Phe His Arg Leu Tyr Arg Asn
325 330 335

His Leu Met Lys Ser Lys Phe Ala Thr Ser Asp Arg Pro Val Leu Leu
340 345 350

Asn Ser Trp Glu Gly Val Tyr Phe Asp Tyr Asn Gln Ser Ser Ile Glu
355 360 365

Thr Leu Ala Glu Glu Ser Ala Ala Leu Gly Val His Leu Phe Val Met
370 375 380

Asp Asp Gly Trp Phe Gly Asp Lys Tyr Pro Arg Val Ser Asp Asn Ala
385 390 395 400

Gly Leu Gly Asp Trp Met Pro Asn Pro Ala Arg Phe Pro Asp Gly Leu
405 410 415

Thr Pro Val Val Gln Asp Ile Thr Asn Leu Thr Val Asn Gly Thr Glu
420 425 430

Ser Thr Lys Leu Arg Phe Gly Ile Trp Val Glu Pro Glu Met Val Asn
435 440 445

Pro Asn Ser Thr Leu Tyr His Glu His Pro Glu Trp Ala Leu His Ala
450 455 460

Gly Pro Tyr Pro Arg Thr Glu Arg Arg Asn Gln Leu Val Leu Asn Leu
465 470 475 480

Ala Leu Pro Ala Val Gln Asp Phe Ile Ile Asp Phe Met Thr Asn Leu
485 490 495

Leu Gln Asp Thr Gly Ile Ser Tyr Val Lys Trp Asp Asn Asn Arg Gly
500 505 510

Ile His Glu Thr Pro Ser Pro Ser Thr Asp His Gln Tyr Met Leu Gly
515 520 525

Leu Tyr Arg Val Phe Asp Thr Leu Thr Thr Arg Phe Pro Asp Val Leu
530 535 540

Trp Glu Gly Cys Ala Ser Gly Gly Gly Arg Phe Asp Ala Gly Met Leu
545 550 555 560
Gln Tyr Val Pro Gln Ile Trp Thr Ser Asp Asn Thr Asp Ala Ile Asp
565 570 575
Arg Ile Thr Ile Gln Phe Gly Thr Ser Leu Ala Tyr Pro Pro Ser Ala
580 585 590
Met Gly Ala His Leu Ser Ala Val Pro Asn Ala Gln Thr Gly Arg Thr
595 600 605

Val Pro Phe Thr Phe Arg Ala His Val Ala Met Met Gly Gly Ser Phe
610 615 620
Gly Leu Glu Leu Asp Pro Ala Thr Val Glu Gly Asp Glu Ile Val Pro
625 630 635 640
Glu Leu Leu Ala Leu Ala Glu Lys Val Asn Pro Ile Ile Leu Asn Gly
645 650 655
Asp Leu Tyr Arg Leu Arg Leu Pro Gln Asp Ser Gln Trp Pro Ala Ala
660 665 670

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

Tyr Gly Ser Lys Val Val Phe Leu Glu Arg Gln
740 745

<210> 117
<211> 28
<212> PRT
<213> Aspergillus niger
<400> 117
Met Ile Gly Ser Ser His Ala Val Val Ala Leu Gly Leu Phe Thr Leu

1 5 10 15

Tyr Gly His Ser Ala Ala Ala Pro Ala Ile Gly Ala

20 25

<210> 118

<211> 2244

<212> DNA

<213> Aspergillus niger

<400> 118

atgatcggct cctcccacgc ggtcgtcgcc ctgggcctgt tcaccctgta cggccactcc 60

gccgccgcgc ccgccatcgg cgcctccaac tcccagacca tcgtcacgaa cggcacctcc 120

ttcgccctga acggcgacaa cgtgtcctac cgcttccacg tgaactcctc cacgggcgac 180

ctgatctccg accacttcgg cggcgtggtg tccggcacca tcccctcccc cgtggagccc 240

gcggtcaacg gctgggtggg catgcccggc cgcatccgcc gcgagttccc cgaccagggc 300

cgcggcgact tccgcatccc cgcggtgcgc atccgcgagt ccgccggcta caccgtctcc 360

gacctgcagt acgtgtccca cgaggtgatc gagggcaagt acgcgctgcc cggcctgccc 420

gccacgttcg gcgacgccca ggacgccacc accctggtgg tgcacctgta cgacaactac 480

tcctccgtcg cggccgacct gtcctactcc atcttcccca agtacgacgc gatcgtccgc 540

tccgtgaacg tgaccaacca gggccccggc aacatcacca tcgaggcgct ggcctccatc 600

tccatcgact tcccctacga ggacctggac atggtgtccc tgcgcggcga ctgggcccgc 660

gaggcgaacg tgcagcgctc caaggtccag tacggcgtgc agggcttcgg ctcctccacc 720

ggctactcct cccacctgca caaccccttc ctggcgatcg tcgaccccgc gaccaccgag 780

tcccagggcg aggcctgggg cttcaacctg gtctacaccg gctccttctc cgcccaggtc 840

gagaagggct cccagggctt cacgcgcgcc ctgctgggct tcaaccccga ccagctgtcc 900

tggaacctgg gccccggcga gacgctgacg tcccccgagt gcgtcgccgt ctactccgac 960

aagggcctgg gctccgtctc ccgcaagttc caccgcctgt accgcaacca cctgatgaag 1020

tccaagttcg ccacgtccga ccgccccgtg ctgctgaact cctgggaggg cgtctacttc 1080

gactacaacc agtcctccat cgagacgctg gcggaggagt ccgccgccct gggcgtgcac 1140

ctgttcgtca tggacgacgg ctggttcggc gacaagtacc cccgcgtgtc cgacaacgcc 1200

ggcctgggcg actggatgcc caaccccgcc cgcttccccg acggcctgac gcccgtcgtg 1260

caggacatca ccaacctgac cgtcaacggc accgagtcca ccaagctgcg cttcggcatc 1320

tgggtggagc ccgagatggt gaaccccaac tccaccctgt accacgagca ccccgagtgg 1380

gcgctgcacg cgggccccta cccccgcacc gagcgccgca accagctggt cctgaacctg 1440

gccctgcccg cggtccagga cttcatcatc gacttcatga ccaacctgct gcaggacacc 1500

ggcatctcct acgtcaagtg ggacaacaac cgcggcatcc acgagacgcc ctccccctcc 1560

acggaccacc agtacatgct gggcctgtac cgcgtgttcg acacgctgac cacgcgcttc 1620

cccgacgtcc tgtgggaggg ctgcgcgtcc ggcggcggcc ggttcgacgc cggcatgctg 1680

cagtacgtcc cccagatctg gacgtccgac aacacggacg cgatcgaccg catcaccatc 1740

cagttcggca cctccctggc ctacccccccc tccgccatgg gcgcccacct gtccgcggtg 1800

cccaacgccc agaccggccg cacggtgccc ttcaccttcc gcgcccacgt cgcgatgatg 1860

ggcggctcct tcggcctgga gctggacccc gccaccgtgg agggcgacga gatcgtgccc 1920

gagctgctgg cgctggccga gaaggtgaac cccatcatcc tgaacggcga cctgtaccgc 1980

ctgcgcctgc cccaggactc ccagtggccc gcggccctgt tcgtgtccca ggacggcgcc 2040

caggccgtcc tgttctactt ccaggtgcag cccaacgtca accacgccgt cccctgggtc 2100

cgcctgcagg gcctggaccc caaggccgac tacacggtgg acggcgacca gacgtactcc 2160

ggcgcgaccc tgatgaacct gggcctgcag tactccttcg acaccgagta cggctccaag 2220

gtggtgttcc tggagcgcca gtaa 2244

<210> 119

<211> 4011

<212> DNA

<213> Aspergillus niger

<400> 119

gcggccgcgt ggacgagggc tacaaccccg cctacggcgc gcgcccgctg cgccgcgcca 60

tcatgcgcct gctggaggac gcgctggccg agcgcatgct cgccggcgac gtcaaggagg 120

gcgactcggt catcatggac gtcgatggcg atggcgccat cagcgtcctc aacggagacc 180

gcacccacac caccaccatc gactcctccc cggcgggcat ctcgtagacg cgagagggag 240

ggagggagga ttttcagagg gagatgagac gaggactggg ccgggggcct tggggcgctg 300

ctggagcgtg gtgagagcgc ggcggacgtg ccttttcttc ttccgtgcgc gcgctcttgg 360

ccattgatcc ccgattcgcg cccgcatccc cccactgccc ccatcatctt gcctgttgtc 420

gtggcactga cataaacccc ctgcgctgcg ctgctccgct actattgata taggtctcac 480

gcgccaatct tttttgctcc gggtaaccgt ctggacgcca gaattccttt cttgcgctat 540

gacacttcca gcaaaaggta gggcgggctg cgagacggct tcccggcgct gcatgcaaca 600

ccgatgatgc ttcgaccccc cgaagctcct tcggggctgc atgggcgctc cgatgccgct 660

ccagggcgag cgctgtttaa atagccaggc ccccgattgc aaagacatta tagcgagcta 720

ccaaagccat attcaaacac ctagatcact accacttcta cacaggccac tcgagcttgt 780

gatcgcactc cgctaagggg gcgcctcttc ctcttcgttt cagtcacaac ccgcaaacac 840

tagtatgatc ggctcctccc acgcggtcgt cgccctgggc ctgttcaccc tgtacggcca 900

ctccgccgcc gcgcccgcca tcggcgcctc caactcccag accatcgtca cgaacggcac 960

ctccttcgcc ctgaacggcg acaacgtgtc ctaccgcttc cacgtgaact cctccacggg 1020

cgacctgatc tccgaccact tcggcggcgt ggtgtccggc accatcccct cccccgtgga 1080

gcccgcggtc aacggctggg tgggcatgcc cggccgcatc cgccgcgagt tccccgacca 1140

gggccgcggc gacttccgca tccccgcggt gcgcatccgc gagtccgccg gctacaccgt 1200

ctccgacctg cagtacgtgt cccacgaggt gatcgagggc aagtacgcgc tgcccggcct 1260

gcccgccacg ttcggcgacg cccaggacgc caccaccctg gtggtgcacc tgtacgacaa 1320

ctactcctcc gtcgcggccg acctgtccta ctccatcttc cccaagtacg acgcgatcgt 1380

ccgctccgtg aacgtgacca accagggccc cggcaacatc accatcgagg cgctggcctc 1440

catctccatc gacttccgct acgaggacct ggacatggtg tccctgcgcg gcgactgggc 1500

ccgcgaggcg aacgtgcagc gctccaaggt ccagtacggc gtgcagggct tcggctcctc 1560

caccggctac tcctcccacc tgcacaaccc cttcctggcg atcgtcgacc ccgcgaccac 1620

cgagtcccag ggcgaggcct ggggcttcaa cctggtctac accggctcct tctccgccca 1680

ggtcgagaag ggctcccagg gcttcacgcg cgccctgctg ggcttcaacc ccgaccagct 1740

gtcctggaac ctgggccccg gcgagacgct gacgtccccc gagtgcgtcg ccgtctactc 1800

cgacaagggc ctgggctccg tctcccgcaa gttccaccgc ctgtaccgca accacctgat 1860

gaagtccaag ttcgccacgt ccgaccgccc cgtgctgctg aactcctggg agggcgtcta 1920

cttcgactac aaccagtcct ccatcgagac gctggcggag gagtccgccg ccctgggcgt 1980

gcacctgttc gtcatggacg acggctggtt cggcgacaag taccccgcgg tgtccgacaa 2040

cgccggcctg ggcgactgga tgcccaaccc cgcccgcttc cccgacggcc tgacgcccgt 2100

cgtgcaggac atcaccaacc tgaccgtcaa cggcaccgag tccaccaagc tgcgcttcgg 2160

catctgggtg gagcccgaga tggtgaaccc caactccacc ctgtaccacg agcaccccga 2220

gtgggcgctg cacgcgggcc cctacccccg caccgagcgc cgcaaccagc tggtcctgaa 2280

cctggccctg cccgcggtcc aggacttcat catcgacttc atgaccaacc tgctgcagga 2340

caccggcatc tcctacgtca agtgggacaa caaccgcggc atccacgaga cgccctcccc 2400

ctccacggac caccagtaca tgctgggcct gtaccgcgtg ttcgacacgc tgaccacgcg 2460

cttccccgac gtcctgtggg agggctgcgc gtccggcggc ggccggttcg acgccggcat 2520

gctgcagtac gtcccccaga tctggacgtc cgacaacacg gacgcgatcg accgcatcac 2580

catccagttc ggcacctccc tggcctaccc cccctccgcc atgggcgccc acctgtccgc 2640

ggtgcccaac gcccagaccg gccgcacggt gcccttcacc ttccgcgccc acgtcgcgat 2700

gatgggcggc tccttcggcc tggagctgga ccccgccacc gtggagggcg acgagatcgt 2760

gcccgagctg ctggcgctgg ccgagaaggt gaaccccatc atcctgaacg gcgacctgta 2820

ccgcctgcgc ctgccccagg actcccagtg gcccgcggcc ctgttcgtgt cccaggacgg 2880

cgcccaggcc gtcctgttct acttccaggt gcagcccaac gtcaaccacg ccgtcccctg 2940

ggtccgcctg cagggcctgg accccaaggc cgactacacg gtggacggcg accagacgta 3000

ctccggcgcg accctgatga acctgggcct gcagtactcc ttcgacaccg agtacggctc 3060

caaggtggtg ttcctggagc gccagtaaca attgaagcag cagcagctcg gatagtatcg 3120

acacactctg gacgctggtc gtgtgatgga ctgttgccgc cacacttgct gccttgacct 3180

gtgaatatcc ctgccgcttt tatcaaacag cctcagtgtg tttgatcttg tgtgtacgcg 3240

cttttgcgag ttgctagctg cttgtgctat ttgcgaatac cacccccagc atccccttcc 3300

ctcgtttcat atcgcttgca tcccaaccgc aacttatcta cgctgtcctg ctatccctca 3360

gcgctgctcc tgctcctgct cactgccccc cgcacagcct tggtttgggc tccgcctgta 3420

ttctcctggt actgcaacct gtaaaccagc actgcaatgc tgatgcacgg gaagtagtgg 3480

gatgggaaca caaatggaaa gcttgagctc agaatagtat cgggtgatgc gaagtcagaa 3540

ccaggcaggg cctgtcgcct gaggtggcaa cgatgggaag caatcaatct gggtacagtc 3600

gtccgcacga tcccgtgatc tcccccaccg acacctatcc ccgcccatcc cggcccaccc 3660

tttcagtccc ctcagcatgc attgtgcacc gcgacaaagc atgtctgctc gtgcactggt 3720

tcaggccacg gcgcaccgag tcctcgccct tcgcagagtg atcaccctcc ccggaaccag 3780

ccacgctcgc tgctgcgggc cgatcagccg cgcgcactcc ctgcaactag ggacaactca 3840

ggcaaccacg cgcctcacaa gcatggccgc cgtggcatcc aacccgctcg tgacggtggg 3900

tgcgcaagtg ccaggggcct cgtcgtcacg gcgtgcatcc tcgagggatg cgatccggca 3960

actatatgtc gtttatctcc ccaccaatca caggatgagc ccctgtctag a 4011

<210> 120

<211> 411

<212> PRT

<213> Cyamopsis tetragonoloba

<400> 120

Met Ala Thr His Tyr Ser Ile Ile Gly Gly Met Ile Ile Val Val Leu

1 5 10 15

Leu Met Ile Ile Gly Ser Glu Gly Gly Arg Leu Leu Glu Lys Lys Asn

20 25 30

Arg Thr Ser Ala Glu Ala Glu His Tyr Asn Val Arg Arg Tyr Leu Ala

35 40 45

Glu Asn Gly Leu Gly Gln Thr Pro Pro Met Gly Trp Asn Ser Trp Asn

50 55 60

His Phe Gly Cys Asp Ile Asn Glu Asn Val Val Arg Glu Thr Ala Asp

65 70 75 80

Ala Met Val Ser Thr Gly Leu Ala Ala Leu Gly Tyr Gln Tyr Ile Asn

85 90 95

Leu Asp Asp Cys Trp Ala Glu Leu Asn Arg Asp Ser Glu Gly Asn Met

100 105 110

Val Pro Asn Ala Ala Ala Phe Pro Ser Gly Ile Lys Ala Leu Ala Asp

115 120 125

Tyr Val His Ser Lys Gly Leu Lys Leu Gly Val Tyr Ser Asp Ala Gly

130 135 140

Asn Gln Thr Cys Ser Lys Arg Met Pro Gly Ser Leu Gly His Glu Glu

145 150 155 160

Gln Asp Ala Lys Thr Phe Ala Ser Trp Gly Val Asp Tyr Leu Lys Tyr

165 170 175

Asp Asn Cys Glu Asn Leu Gly Ile Ser Val Lys Glu Arg Tyr Pro Pro

180 185 190

Met Gly Lys Ala Leu Leu Ser Ser Gly Arg Pro Ile Phe Phe Ser Met

195 200 205

Cys Glu Trp Gly Trp Glu Asp Pro Gln Ile Trp Ala Lys Ser Ile Gly

210 215 220

Asn Ser Trp Arg Thr Thr Gly Asp Ile Glu Asp Asn Trp Asn Ser Met

225 230 235 240

Thr Ser Ile Ala Asp Ser Asn Asp Lys Trp Ala Ser Tyr Ala Gly Pro

245 250 255

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

305 310 315 320
Leu Gly Val Gln Gly Lys Lys Val Lys Ser Thr Asn Asp Leu Glu Val
325 330 335
Trp Ala Gly Pro Leu Ser Asp Asn Lys Val Ala Val Ile Leu Trp Asn
340 345 350
Arg Ser Ser Ser Arg Ala Thr Val Thr Ala Ser Trp Ser Asp Ile Gly
355 360 365
Leu Gln Gln Gly Thr Thr Val Asp Ala Arg Asp Leu Trp Glu His Ser

370 375 380
Thr Gln Ser Leu Val Ser Gly Glu Ile Ser Ala Glu Ile Asp Ser His
385 390 395 400
Ala Cys Lys Met Tyr Val Leu Thr Pro Arg Ser
405 410

<210> 121
<211> 1236
<212> DNA
<213> Cyamopsis tetragonoloba
<400> 121
atggccaccc actactccat catcggcggc atgatcatcg tcgtcctgct gatgatcatc 60
ggctccgagg gcggccggct gctggagaag aagaaccgca cctccgccga ggcggagcac 120

tacaacgtgc gccgctacct ggccgagaac ggcctgggcc agacccccccc catgggctgg 180
aactcctgga accacttcgg ctgcgacatc aacgagaacg tcgtccgcga gacggcggac 240
gccatggtgt ccaccggcct ggccgccctg ggctaccagt acatcaacct ggacgactgc 300
tgggcggagc tgaaccgcga ctccgagggc aacatggtgc ccaacgccgc cgcgttcccc 360
tccggcatca aggcgctggc ggactacgtc cactccaagg gcctgaagct gggcgtctac 420

tccgacgcgg gcaaccagac ctgctccaag cgcatgcccg gctccctggg ccacgaggag 480

caggacgcga agacgttcgc ctcctggggc gtcgactacc tgaagtacga caactgcgag 540

aacctgggca tctccgtgaa ggagcgctac ccccccatgg gcaaggcgct gctgtcctcc 600

ggccgcccca tcttcttctc catgtgcgag tggggctggg aggaccccca gatctgggcc 660

aagtccatcg gcaactcctg gcgcaccacc ggcgacatcg aggacaactg gaactccatg 720

acctccatcg ccgactccaa cgacaagtgg gcctcctacg ccggccccgg cggctggaac 780

gaccccgaca tgctggaggt cggcaacggc ggcatgacga cggaggagta ccgctcccac 840

ttctccatct gggccctggc caaggcgccc ctgctggtgg gctgcgacat ccgcgcgatg 900

gacgacacga cgcacgagct gatctccaac gccgaggtga tcgcggtgaa ccaggacaag 960

ctgggcgtgc agggcaagaa ggtcaagtcc acgaacgacc tggaggtctg ggccggcccc 1020

ctgtccgaca acaaggtggc ggtgatcctg tggaaccgct cctcctcccg cgccaccgtc 1080

accgcgtcct ggtccgacat cggcctgcag cagggcacca ccgtcgacgc gcgcgacctg 1140

tgggagcact ccacgcagtc cctggtgtcc ggcgagatct ccgccgagat cgactcccac 1200

gcctgcaaga tgtacgtcct gacgccccgc tcctaa 1236

<210> 122

<211> 3003

<212> DNA

<213> Cyamopsis tetragonoloba

<400> 122

gcggccgcgt ggacgagggc tacaaccccg cctacggcgc gcgcccgctg cgccgcgcca 60

tcatgcgcct gctggaggac gcgctggccg agcgcatgct cgccggcgac gtcaaggagg 120

gcgactcggt catcatggac gtcgatggcg atggcgccat cagcgtcctc aacggagacc 180

gcacccacac caccaccatc gactcctccc cggcgggcat ctcgtagacg cgagagggag 240

ggagggagga ttttcagagg gagatgagac gaggactggg ccgggggcct tggggcgctg 300

ctggagcgtg gtgagagcgc ggcggacgtg ccttttcttc ttccgtgcgc gcgctcttgg 360

ccattgatcc ccgattcgcg cccgcatccc cccactgccc ccatcatctt gcctgttgtc 420

gtggcactga cataaacccc ctgcgctgcg ctgctccgct actattgata taggtctcac 480

gcgccaatct tttttgctcc gggtaaccgt ctggacgcca gaattccttt cttgcgctat 540

gacacttcca gcaaaaggta gggcgggctg cgagacggct tcccggcgct gcatgcaaca 600

ccgatgatgc ttcgaccccc cgaagctcct tcggggctgc atgggcgctc cgatgccgct 660

ccagggcgag cgctgtttaa atagccaggc ccccgattgc aaagacatta tagcgagcta 720

ccaaagccat attcaaacac ctagatcact accacttcta cacaggccac tcgagcttgt 780

gatcgcactc cgctaagggg gcgcctcttc ctcttcgttt cagtcacaac ccgcaaacac 840

tagtatggcc acccactact ccatcatcgg cggcatgatc atcgtcgtcc tgctgatgat 900

catcggctcc gagggcggcc ggctgctgga gaagaagaac cgcacctccg ccgaggcgga 960

gcactacaac gtgcgccgct acctggccga gaacggcctg ggccagaccc cccccatggg 1020

ctggaactcc tggaaccact tcggctgcga catcaacgag aacgtcgtcc gcgagacggc 1080

ggacgccatg gtgtccaccg gcctggccgc cctgggctac cagtacatca acctggacga 1140

ctgctgggcg gagctgaacc gcgactccga gggcaacatg gtgcccaacg ccgccgcgtt 1200

cccctccggc atcaaggcgc tggcggacta cgtccactcc aagggcctga agctgggcgt 1260

ctactccgac gcgggcaacc agacctgctc caagcgcatg cccggctccc tgggccacga 1320

ggagcaggac gcgaagacgt tcgcctcctg gggcgtcgac tacctgaagt acgacaactg 1380

cgagaacctg ggcatctccg tgaaggagcg ctacccccccc atgggcaagg cgctgctgtc 1440

ctccggccgc cccatcttct tctccatgtg cgagtggggc tgggaggacc cccagatctg 1500

ggccaagtcc atcggcaact cctggcgcac caccggcgac atcgaggaca actggaactc 1560

catgacctcc atcgccgact ccaacgacaa gtgggcctcc tacgccggcc ccggcggctg 1620

gaacgacccc gacatgctgg aggtcggcaa cggcggcatg acgacggagg agtaccgctc 1680

ccacttctcc atctgggccc tggccaaggc gcccctgctg gtgggctgcg acatccgcgc 1740

gatggacgac acgacgcacg agctgatctc caacgccgag gtgatcgcgg tgaaccagga 1800

caagctgggc gtgcagggca agaaggtcaa gtccacgaac gacctggagg tctgggccgg 1860

ccccctgtcc gacaacaagg tggcggtgat cctgtggaac cgctcctcct cccgcgccac 1920

cgtcaccgcg tcctggtccg acatcggcct gcagcagggc accaccgtcg acgcgcgcga 1980

cctgtgggag cactccacgc agtccctggt gtccggcgag atctccgccg agatcgactc 2040

ccacgcctgc aagatgtacg tcctgacgcc ccgctcctaa caattgaagc agcagcagct 2100

cggatagtat cgacacactc tggacgctgg tcgtgtgatg gactgttgcc gccacacttg 2160

ctgccttgac ctgtgaatat ccctgccgct tttatcaaac agcctcagtg tgtttgatct 2220

tgtgtgtacg cgcttttgcg agttgctagc tgcttgtgct atttgcgaat accacccccа 2280

gcatcccctt ccctcgtttc atatcgcttg catcccaacc gcaacttatc tacgctgtcc 2340

tgctatccct cagcgctgct cctgctcctg ctcactgccc ctcgcacagc cttggtttgg 2400

gctccgcctg tattctcctg gtactgcaac ctgtaaacca gcactgcaat gctgatgcac 2460

gggaagtagt gggatgggaa cacaaatgga aagcttgagc tcagaatagt atcgggtgat 2520

gcgaagtcag aaccaggcag ggcctgtcgc ctgaggtggc aacgatggga agcaatcaat 2580

ctgggtacag tcgtccgcac gatcccgtga tctcccccac cgacacctat ccccgcccat 2640

cccggcccac cctttcagtc ccctcagcat gcattgtgca ccgcgacaaa gcatgtctgc 2700

tcgtgcactg gttcaggcca cggcgcaccg agtcctcgcc cttcgcagag tgatcaccct 2760

ccccggaacc agccacgctc gctgctgcgg gccgatcagc cgcgcgcact ccctgcaact 2820

agggacaact caggcaacca cgcgcctcac aagcatggcc gccgtggcat ccaacccgct 2880

cgtgacggtg ggtgcgcaag tgccaggggc ctcgtcgtca cggcgtgcat cctcgaggga 2940

tgcgatccgg caactatatg tcgtttatct ccccaccaat cacaggatga gcccctgtct 3000

aga 3003

<210> 123

<211> 22

<212> DNA

<213> Chlorella vulgaris

<400> 123

actgcaatgc tgatgcacgg ga 22

<210> 124

<211> 19

<212> DNA

<213> Chlorella vulgaris

<400> 124

tccaggtcct tttcgcact 19

<210> 125

<211> 632

<212> PRT

<213> Unknown

<220><223> Coccomyxa species

<400> 125

Met Thr Thr Asn Leu Ala Lys Leu Pro Leu Gly Gly Leu Ser Ser Arg

1 5 10 15

Ser Ser Ile Ala Gly Ala Pro Leu Arg Val Asn Ser His Ser Arg Glu

20 25 30

Arg His Leu Gly Ala Lys Thr Thr Ala Ile Ala Ala Pro Glu Arg Leu

35 40 45

Asp Tyr Leu Asp Asn Ala Glu Glu Ala Arg Leu Gln Gln Thr Asp Ala

50 55 60

Phe Ala Glu Leu Lys Ala Leu Ser Ser Arg Gln Ser Val Asn Arg Pro

65 70 75 80

Gln Lys Gly Glu Leu Ser Phe Arg Gln Ser Pro Thr Phe Gln Asp Cys

85 90 95

Phe Pro Gly Ser Glu Lys Cys Tyr Arg Glu Val Glu His Asp Gly Lys

100 105 110

Thr Leu Lys Val Pro Phe Arg Arg Val His Leu Gln Asp Asp Asn Gly

115 120 125

His Phe Asp Leu Tyr Asp Thr Ser Gly Pro Gln Gly Val Asn Pro Arg

130 135 140

Glu Gly Leu Pro Lys Ile Arg Ser Ser Trp Val Glu Pro Arg Glu Ala

145 150 155 160

Arg Gly Asp Lys Val Gln Thr Gln Gln Tyr Tyr Ala Lys Gln Gly Ile

165 170 175

Ile Thr Glu Glu Met Ala Phe Cys Ala Ala Arg Glu Arg Met Asp Pro

180 185 190

Glu Phe Ile Arg Ser Glu Val Ala Arg Gly Arg Ala Ile Ile Pro Ala

195 200 205

Asn Lys Arg His Leu Glu Leu Glu Pro Thr Val Val Gly Arg Asn Phe

210 215 220

Leu Val Lys Val Asn Ala Asn Ile Gly Asn Ser Ala Val Ser Ser Ser

225 230 235 240

Ile Glu Glu Glu Val Glu Lys Leu Gln Trp Ser Thr Ile Trp Gly Ala

245 250 255

Asp Thr Ala Met Asp Leu Ser Thr Gly Asn Asn Ile His Glu Thr Arg

260 265 270

Glu Trp Val Met Arg Asn Ser Pro Val Pro Val Gly Thr Val Pro Ile

275 280 285

```
Tyr Gln Cys Leu Glu Lys Ala Gly Gly Ile Val Glu Asn Ile Thr Trp
    290                 295                 300
Glu Leu Phe Arg Glu Thr Leu Ile Glu Gln Ala Glu Gln Gly Val Asp
305                 310                 315                 320
Tyr Phe Thr Ile His Ala Gly Val Leu Leu Arg Tyr Ile Pro Leu Thr

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

385                 390                 395                 400
Gln Phe Ala Glu Leu Lys Thr Gln Gly Glu Leu Thr Arg Arg Ala Trp
                405                 410                 415
Ala Lys Asp Val Gln Val Met Asn Glu Gly Pro Gly His Val Pro Leu
            420                 425                 430
His Lys Ile Pro Glu Asn Met Glu Lys Gln Leu Asp Trp Cys Ser Glu
        435                 440                 445
Ala Pro Phe Tyr Thr Leu Gly Pro Leu Ala Thr Asp Ile Ala Pro Ala

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

        515                 520                 525
Asp Asn Ala Leu Ser Lys Ala Arg Phe Glu Phe Arg Trp Tyr Asp Gln
```

530 535 540

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

545 550 555 560

Thr Leu Pro Gln Glu Pro Ala Lys Thr Ala His Phe Cys Ser Met Cys

565 570 575

Gly Pro Lys Phe Cys Ser Met Gln Ile Thr Gln Asp Leu Arg Glu Tyr

580 585 590

Ala Lys Asn His Gln Met Glu Glu Asp Glu Ala Ile Gln Thr Gly Met

595 600 605

Ala Glu Met Ser Glu Gln Phe Lys Ser Ile Gly Ala Glu Val Tyr Leu

610 615 620

Asp Glu Ala Val Leu Glu His Ala

625 630

<210> 126

<211> 40

<212> PRT

<213> Unknown

<220><223> Coccomyxa species

<400> 126

Met Thr Thr Asn Leu Ala Lys Leu Pro Leu Gly Gly Leu Ser Ser Arg

1 5 10 15

Ser Ser Ile Ala Gly Ala Pro Leu Arg Val Asn Ser His Ser Arg Glu

20 25 30

Arg His Leu Gly Ala Lys Thr Thr

35 40

<210> 127

<211> 6649

<212> DNA

<213> Unknown

<220><223> Coccomyxa species

<400> 127

gctcttcgcc gccgccactc ctgctcgagc gcgcccgcgc gtgcgccgcc agcgccttgg 60

cctttcgcc gcgctcgtgc gcgtcgctga tgtccatcac caggtccatg aggtctgcct 120

tgcgccggct gagccactgc ttcgtccggg cggccaagag gagcatgagg gaggactcct 180
ggtccagggt cctgacgtgg tcgcggctct gggagcgggc cagcatcatc tggctctgcc 240
gcaccgaggc cgcctccaac tggtcctcca gcagccgcag tcgccgccga ccctggcaga 300
ggaagacagg tgagggggggt atgaattgta cagaacaacc acgagccttg tctaggcaga 360
atccctacca gtcatggctt tacctggatg acggcctgcg aacagctgtc cagcgaccct 420
cgctgccgcc gcttctcccg cacgcttctt tccagcaccg tgatggcgcg agccagcgcc 480
gcacgctggc gctgcgcttc gccgatctga ggacagtcgg ggaactctga tcagtctaaa 540

ccccccttgcg cgttagtgtt gccatccttt gcagaccggt gagagccgac ttgttgtgcg 600
ccacccccca caccacctcc tcccagacca attctgtcac ctttttggcg aaggcatcgg 660
cctcggcctg cagagaggac agcagtgccc agccgctggg ggttggcgga tgcacgctca 720
ggtacccttt cttgcgctat gacacttcca gcaaaaggta gggcgggctg cgagacggct 780
tcccggcgct gcatgcaaca ccgatgatgc ttcgaccccc cgaagctcct tcggggctgc 840
atgggcgctc cgatgccgct ccagggcgag cgctgtttaa atagccaggc ccccgattgc 900
aaagacatta tagcgagcta ccaaagccat attcaaacac ctagatcact accacttcta 960

cacaggccac tcgagcttgt gatcgcactc cgctaagggg gcgcctcttc ctcttcgttt 1020
cagtcacaac ccgcaaacgg cgcgccatgc tgctgcaggc cttcctgttc ctgctggccg 1080
gcttcgccgc caagatcagc gcctccatga cgaacgagac gtccgaccgc cccctggtgc 1140
acttcacccc caacaagggc tggatgaacg accccaacgg cctgtggtac gacgagaagg 1200
acgccaagtg gcacctgtac ttccagtaca acccgaacga caccgtctgg gggacgccct 1260
tgttctgggg ccacgccacg tccgacgacc tgaccaactg ggaggaccag cccatcgcca 1320
tcgccccgaa gcgcaacgac tccggcgcct tctccggctc catggtggtg gactacaaca 1380

acacctccgg cttcttcaac gacaccatcg acccgcgcca gcgctgcgtg gccatctgga 1440
cctacaacac cccggagtcc gaggagcagt acatctccta cagcctggac ggcggctaca 1500
ccttcaccga gtaccagaag aaccccgtgc tggccgccaa ctccacccag ttccgcgacc 1560
cgaaggtctt ctggtacgag ccctcccaga agtggatcat gaccgcggcc aagtcccagg 1620
actacaagat cgagatctac tcctccgacg acctgaagtc ctggaagctg gagtccgcgt 1680
tcgccaacga gggcttcctc ggctaccagt acgagtgccc cggcctgatc gaggtcccca 1740
ccgagcagga ccccagcaag tcctactggg tgatgttcat ctccatcaac cccggcgccc 1800

cggccggcgg ctccttcaac cagtacttcg tcggcagctt caacggcacc cacttcgagg 1860

ccttcgacaa ccagtcccgc gtggtggact tcggcaagga ctactacgcc ctgcagacct 1920

tcttcaacac cgacccgacc tacgggagcg ccctgggcat cgcgtgggcc tccaactggg 1980

agtactccgc cttcgtgccc accaacccct ggcgctcctc catgtccctc gtgcgcaagt 2040

tctccctcaa caccgagtac caggccaacc cggagacgga gctgatcaac ctgaaggccg 2100

agccgatcct gaacatcagc aacgccggcc cctggagccg gttcgccacc aacaccacgt 2160

tgacgaaggc caacagctac aacgtcgacc tgtccaacag caccggcacc ctggagttcg 2220

agctggtgta cgccgtcaac accacccaga cgatctccaa gtccgtgttc gcggacctct 2280

ccctctggtt caagggcctg gaggaccccg aggagtacct ccgcatgggc ttcgaggtgt 2340

ccgcgtcctc cttcttcctg gaccgcggga acagcaaggt gaagttcgtg aaggagaacc 2400

cctacttcac caaccgcatg agcgtgaaca accagccctt caagagcgag aacgacctgt 2460

cctactacaa ggtgtacggc ttgctggacc agaacatcct ggagctgtac ttcaacgacg 2520

gcgacgtcgt gtccaccaac acctacttca tgaccaccgg gaacgccctg ggctccgtga 2580

acatgacgac gggggtggac aacctgttct acatcgacaa gttccaggtg cgcgaggtca 2640

agtgacaatt ggcagcagca gctcggatag tatcgacaca ctctggacgc tggtcgtgtg 2700

atggactgtt gccgccacac ttgctgcctt gacctgtgaa tatccctgcc gcttttatca 2760

aacagcctca gtgtgtttga tcttgtgtgt acgcgctttt gcgagttgct agctgcttgt 2820

gctatttgcg aataccaccc ccagcatccc cttccctcgt ttcatatcgc ttgcatccca 2880

accgcaactt atctacgctg tcctgctatc cctcagcgct gctcctgctc ctgctcactg 2940

cccctcgcac agccttggtt tgggctccgc ctgtattctc ctggtactgc aacctgtaaa 3000

ccagcactgc aatgctgatg cacgggaagt agtgggatgg gaacacaaat ggaggatccc 3060

gcgtctcgaa cagagcgcgc agaggaacgc tgaaggtctc gcctctgtcg cacctcagcg 3120

cggcatacac cacaataacc acctgacgaa tgcgcttggt tcttcgtcca ttagcgaagc 3180

gtccggttca cacacgtgcc acgttggcga ggtggcaggt gacaatgatc ggtggagctg 3240

atggtcgaaa cgttcacagc ctagggatat cgaattcctt tcttgcgcta tgacacttcc 3300

agcaaaaggt agggcgggct gcgagacggc ttcccggcgc tgcatgcaac accgatgatg 3360

cttcgacccc ccgaagctcc ttcggggctg catgggcgct ccgatgccgc tccagggcga 3420

gcgctgttta aatagccagg cccccgattg caaagacatt atagcgagct accaaagcca 3480

tattcaaaca cctagatcac taccacttct acacaggcca ctcgagcttg tgatcgcact 3540

ccgctaaggg ggcgcctctt cctcttcgtt tcagtcacaa cccgcaaaca ctagtatgac 3600

caccaacctg gctaagctgc ctctcggcgg cctctcttct cggtcgtcca tcgccggggc 3660

tcctctgcgc gtgaacagcc acagccgcga gcgccacctg ggcgccaaga ccaccgcgat 3720

cgccgctccc gagcggctgg actacctcga caacgcggag gaggcgcggc tgcagcagac 3780

ggacgccttc gccgagctga aggcgctgtc gagccgccag tccgtgaacc ggccccaaaa 3840

gggcgagctg tcgttccggc agtcgccgac ctttcaggac tgcttccctg gctcggagaa 3900

gtgctaccgc gaggtggagc acgatggcaa gacgctcaag gtgccttttc gccgggtgca 3960

cctgcaggac gacaacggcc atttcgacct gtacgacacc tctggccccc agggcgtgaa 4020

cccccgcgag ggtctgccca agatccgctc ctcctgggtg gagccgcgcg aggcgcgggg 4080

cgacaaggtg caaacgcagc agtactacgc caagcaaggc atcatcaccg aggagatggc 4140

cttctgcgct gcgcgcgagc gcatggaccc cgagttcatc cggtctgagg tggcccgggg 4200

tcgggccatc atccctgcga acaagcgcca tctggagctg gagccgaccg tcgtgggccg 4260

caactttctg gtcaaggtga acgccaacat cggcaactct gcggtcagct cgagcatcga 4320

ggaggaggtg gagaagctgc agtggagcac catctggggt gccgacacgg ccatggacct 4380

gagcaccggc aacaacatcc acgagacgcg cgagtgggtg atgcgcaaca gcccggtccc 4440

tgtgggtacg gtccctatct accagtgcct ggagaaggcg ggcggcatcg tggagaacat 4500

cacctgggag ctgttccggg agacgctgat cgagcaggcc gagcagggcg tcgactactt 4560

taccatccac gcgggggtcc tgctccgcta catccccctg acggcgaacc gcgtcaccgg 4620

catcgtcagc cgcggcggct ctatccatgc caagctctgc ctgctggacc acaccgagaa 4680

cttcgcctac atgcactggg acgagatcct ggacatctgc gcccagtacg acatcaccct 4740

gtcgatcggc gacggcctgc gccccggctg catcgcggac gccaacgacg ctgcccagtt 4800

tgctgagctg aagacccagg gcgagctgac ccgccgcgcc tgggcgaagg acgtgcaggt 4860

gatgaacgag ggtcccggcc atgtgccgct gcacaagatc cccgagaaca tggagaagca 4920

gctggactgg tgcagcgagg cgcccttcta cacgctcggc cccctggcga cggacatcgc 4980

gcctgcgtac gaccacatca cctccgcgat cggcgccgcc acgatcggcg cgctggggac 5040

cgcgctgctg tgttacgtga ccccgaagga gcacctgggc ctccccgatc gcgacgacgt 5100

gaaggccggg gtgatcgcct acaagatcgc ggcgcacgcg gctgacctgg cgaagggtca 5160

cccctacgct caggagtggg ataacgccct ctctaaggcg cgcttcgagt tccggtggta 5220

cgaccagttc cacctgtcgc tggaccccgt caccgcccgc ctgttccatg acgcgaccct 5280

gccgcaggag cccgccaaga ccgcccactt ctgctccatg tgcggcccca agttctgctc 5340

catgcagatc acccaggacc tccgcgagta cgccaagaac catcaaatgg aggaggacga 5400

ggcgatccaa accggcatgg ccgagatgtc cgagcagttc aagagcatcg gcgcggaggt 5460

ctacctcgac gaggcggtgc tggagcacgc gtaagcagca gcagctcgga tagtatcgac 5520

acactctgga cgctggtcgt gtgatggact gttgccgcca cacttgctgc cttgacctgt 5580

gaatatccct gccgctttta tcaaacagcc tcagtgtgtt tgatcttgtg tgtacgcgct 5640
tttgcgagtt gctagctgct tgtgctattt gcgaatacca cccccagcat ccccttccct 5700
cgtttcatat cgcttgcatc ccaaccgcaa cttatctacg ctgtcctgct atccctcagc 5760
gctgctcctg ctcctgctca ctgcccctcg cacagccttg gtttgggctc cgcctgtatt 5820
ctcctggtac tgcaacctgt aaaccagcac tgcaatgctg atgcacggga agtagtggga 5880
tgggaacaca aatggaaagc ttgagctctt gttttccaga aggagttgct ccttgagcct 5940
ttcattctca gcctcgataa cctccaaagc cgctctaatt gtggaggggg ttcgaattta 6000

aaagcttgga atgttggttc gtgcgtctgg aacaagccca gacttgttgc tcactgggaa 6060
aaggaccatc agctccaaaa aacttgccgc tcaaaccgcg tacctctgct ttcgcgcaat 6120
ctgccctgtt gaaatcgcca ccacattcat attgtgacgc ttgagcagtc tgtaattgcc 6180
tcagaatgtg gaatcatctg ccccctgtgc gagcccatgc caggcatgtc gcgggcgagg 6240
acacccgcca ctcgtacagc agaccattat gctacctcac aatagttcat aacagtgacc 6300
atatttctcg aagctcccca acgagcacct ccatgctctg agtggccacc ccccggccct 6360
ggtgcttgcg gagggcaggt caaccggcat ggggctaccg aaatccccga ccggatccca 6420

ccacccccgc gatgggaaga atctctcccc gggatgtggg cccaccacca gcacaacctg 6480
ctggcccagg cgagcgtcaa accataccac acaaatatcc ttggcatcgg ccctgaattc 6540
cttctgccgc tctgctaccc ggtgcttctg tccgaagcag gggttgctag ggatcgctcc 6600
gagtccgcaa acccttgtcg cgtggcgggg cttgttcgag ctgaagagc 6649

<210> 128
<211> 1899
<212> DNA
<213> Unknown
<220><223> Coccomyxa species
<400> 128

atgaccacca acctggctaa gctgcctctc ggcggcctct cttctcggtc gtccatcgcc 60
ggggctcctc tgcgcgtgaa cagccacagc cgcgagcgcc acctgggcgc caagaccacc 120

gcgatcgccg ctcccgagcg gctggactac ctcgacaacg cggaggaggc gcggctgcag 180
cagacggacg ccttcgccga gctgaaggcg ctgtcgagcc gccagtccgt gaaccggccc 240
caaaagggcg agctgtcgtt ccggcagtcg ccgacctttc aggactgctt ccctggctcg 300
gagaagtgct accgcgaggt ggagcacgat ggcaagacgc tcaaggtgcc ttttcgccgg 360

gtgcacctgc aggacgacaa cggccatttc gacctgtacg acacctctgg cccccagggc 420

gtgaaccccc gcgagggtct gcccaagatc cgctcctcct gggtggagcc gcgcgaggcg 480

cggggcgaca aggtgcaaac gcagcagtac tacgccaagc aaggcatcat caccgaggag 540

atggccttct gcgctgcgcg cgagcgcatg gaccccgagt tcatccggtc tgaggtggcc 600

cggggtcggg ccatcatccc tgcgaacaag cgccatctgg agctggagcc gaccgtcgtg 660

ggccgcaact ttctggtcaa ggtgaacgcc aacatcggca actctgcggt cagctcgagc 720

atcgaggagg aggtggagaa gctgcagtgg agcaccatct ggggtgccga cacggccatg 780

gacctgagca ccggcaacaa catccacgag acgcgcgagt gggtgatgcg caacagcccg 840

gtccctgtgg gtacggtccc tatctaccag tgcctggaga aggcgggcgg catcgtggag 900

aacatcacct gggagctgtt ccgggagacg ctgatcgagc aggccgagca gggcgtcgac 960

tactttacca tccacgcggg ggtcctgctc cgctacatcc ccctgacggc gaaccgcgtc 1020

accggcatcg tcagccgcgg cggctctatc catgccaagc tctgcctgct ggaccacacc 1080

gagaacttcg cctacatgca ctgggacgag atcctggaca tctgcgccca gtacgacatc 1140

accctgtcga tcggcgacgg cctgcgcccc ggctgcatcg cggacgccaa cgacgctgcc 1200

cagtttgctg agctgaagac ccagggcgag ctgacccgcc gcgcctgggc gaaggacgtg 1260

caggtgatga acgagggtcc cggccatgtg ccgctgcaca agatccccga gaacatggag 1320

aagcagctgg actggtgcag cgaggcgccc ttctacacgc tcggcccccc ggcgacggac 1380

atcgcgcctg cgtacgacca catcacctcc gcgatcggcg ccgccacgat cggcgcgctg 1440

gggaccgcgc tgctgtgtta cgtgaccccg aaggagcacc tgggcctccc cgatcgcgac 1500

gacgtgaagg ccggggtgat cgcctacaag atcgcggcgc acgcggctga cctggcgaag 1560

ggtcacccct acgctcagga gtgggataac gccctctcta aggcgcgctt cgagttccgg 1620

tggtacgacc agttccacct gtcgctggac cccgtcaccg cccgcctgtt ccatgacgcg 1680

accctgccgc aggagcccgc caagaccgcc cacttctgct ccatgtgcgg ccccaagttc 1740

tgctccatgc agatcaccca ggacctccgc gagtacgcca agaaccatca aatggaggag 1800

gacgaggcga tccaaaccgg catggccgag atgtccgagc agttcaagag catcggcgcg 1860

gaggtctacc tcgacgaggc ggtgctggag cacgcgtaa 1899

<210> 129

<211> 1613

<212> DNA

<213> Saccharomyces cerevisiae

<400> 129

ggcgcgccat gctgctgcag gccttcctgt tcctgctggc cggcttcgcc gccaagatca 60

gcgcctccat gacgaacgag acgtccgacc gccccctggt gcacttcacc cccaacaagg 120

gctggatgaa cgaccccaac ggcctgtggt acgacgagaa ggacgccaag tggcacctgt 180

acttccagta caacccgaac gacaccgtct gggggacgcc cttgttctgg ggccacgcca 240

cgtccgacga cctgaccaac tgggaggacc agcccatcgc catcgccccg aagcgcaacg 300

actccggcgc cttctccggc tccatggtgg tggactacaa caacacctcc ggcttcttca 360

acgacaccat cgacccgcgc cagcgctgcg tggccatctg gacctacaac accccggagt 420

ccgaggagca gtacatctcc tacagcctgg acggcggcta caccttcacc gagtaccaga 480

agaaccccgt gctggccgcc aactccaccc agttccgcga cccgaaggtc ttctggtacg 540

agccctccca gaagtggatc atgaccgcgg ccaagtccca ggactacaag atcgagatct 600

actcctccga cgacctgaag tcctggaagc tggagtccgc gttcgccaac gagggcttcc 660

tcggctacca gtacgagtgc cccggcctga tcgaggtccc caccgagcag gaccccagca 720

agtcctactg ggtgatgttc atctccatca accccggcgc cccggccggc ggctccttca 780

accagtactt cgtcggcagc ttcaacggca cccacttcga ggccttcgac aaccagtccc 840

gcgtggtgga cttcggcaag gactactacg ccctgcagac cttcttcaac accgacccga 900

cctacgggag cgccctgggc atcgcgtggg cctccaactg ggagtactcc gccttcgtgc 960

ccaccaaccc ctggcgctcc tccatgtccc tcgtgcgcaa gttctccctc aacaccgagt 1020

accaggccaa cccggagacg gagctgatca acctgaaggc cgagccgatc ctgaacatca 1080

gcaacgccgg cccctggagc cggttcgcca ccaacaccac gttgacgaag gccaacagct 1140

acaacgtcga cctgtccaac agcaccggca ccctggagtt cgagctggtg tacgccgtca 1200

acaccaccca gacgatctcc aagtccgtgt tcgcggacct ctccctctgg ttcaagggcc 1260

tggaggaccc cgaggagtac ctccgcatgg gcttcgaggt gtccgcgtcc tccttcttcc 1320

tggaccgcgg gaacagcaag gtgaagttcg tgaaggagaa cccctacttc accaaccgca 1380

tgagcgtgaa caaccagccc ttcaagagcg agaacgacct gtcctactac aaggtgtacg 1440

gcttgctgga ccagaacatc ctggagctgt acttcaacga cggcgacgtc gtgtccacca 1500

acacctactt catgaccacc gggaacgccc tgggctccgt gaacatgacg acgggggtgg 1560

acaacctgtt ctacatcgac aagttccagg tgcgcgaggt caagtgacaa ttg 1613

<210> 130

<211> 3660

<212> DNA

<213> Artificial Sequence

<220><223> synthetic nucleotide

<400> 130

gcggccgcgt ggacgagggc tacaaccccg cctacggcgc gcgcccgctg cgccgcgcca 60

tcatgcgcct gctggaggac gcgctggccg agcgcatgct cgccggcgac gtcaaggagg 120

gcgactcggt catcatggac gtcgatggcg atggcgccat cagcgtcctc aacggagacc 180

gcacccacac caccaccatc gactcctccc cggcgggcat ctcgtagacg cgagagggag 240

ggagggagga ttttcagagg gagatgagac gaggactggg ccgggggcct tggggcgctg 300

ctggagcgtg gtgagagcgc ggcggacgtg ccttttcttc ttccgtgcgc gcgctcttgg 360

ccattgatcc ccgattcgcg cccgcatccc cccactgccc ccatcatctt gcctgttgtc 420

gtggcactga cataaacccc ctgcgctgcg ctgctccgct actattgata taggtctcac 480

gcgccaatct tttttgctcc gggtaaccgt ctggacgcca gaattccttt cttgcgctat 540

gacacttcca gcaaaaggta gggcgggctg cgagacggct tcccggcgct gcatgcaaca 600

ccgatgatgc ttcgaccccc cgaagctcct tcggggctgc atgggcgctc cgatgccgct 660

ccagggcgag cgctgtttaa atagccaggc ccccgattgc aaagacatta tagcgagcta 720

ccaaagccat attcaaacac ctagatcact accacttcta cacaggccac tcgagcttgt 780

gatcgcactc cgctaagggg gcgcctcttc ctcttcgttt cagtcacaac ccgcaaacac 840

tagtatggcc accgcatcca ctttctcggc gttcaatgcc cgctgcggcg acctgcgtcg 900

ctcggcgggc tccgggcccc ggcgcccagc gaggcccctc cccgtgcgcg ggcgcgccgc 960

gatcgccgct cccgagcggc tggactacct cgacaacgcg gaggaggcgc ggctgcagca 1020

gacggacgcc ttcgccgagc tgaaggcgct gtcgagccgc cagtccgtga accggcccca 1080

aaagggcgag ctgtcgttcc ggcagtcgcc gacctttcag gactgcttcc ctggctcgga 1140

gaagtgctac cgcgaggtgg agcacgatgg caagacgctc aaggtgcctt ttcgccgggt 1200

gcacctgcag gacgacaacg gccatttcga cctgtacgac acctctggcc cccagggcgt 1260

gaacccccgc gagggtctgc ccaagatccg ctcctcctgg gtggagccgc gcgaggcgcg 1320

gggcgacaag gtgcaaacgc agcagtacta cgccaagcaa ggcatcatca ccgaggagat 1380

ggccttctgc gctgcgcgcg agcgcatgga ccccgagttc atccggtctg aggtggcccg 1440

gggtcgggcc atcatccctg cgaacaagcg ccatctggag ctggagccga ccgtcgtggg 1500

ccgcaacttt ctggtcaagg tgaacgccaa catcggcaac tctgcggtca gctcgagcat 1560

cgaggaggag gtggagaagc tgcagtggag caccatctgg ggtgccgaca cggccatgga 1620

cctgagcacc ggcaacaaca tccacgagac gcgcgagtgg gtgatgcgca acagcccggt 1680

ccctgtgggt acggtcccta tctaccagtg cctggagaag gcgggcggca tcgtggagaa 1740

catcacctgg gagctgttcc gggagacgct gatcgagcag gccgagcagg gcgtcgacta 1800

ctttaccatc cacgcggggg tcctgctccg ctacatcccc ctgacggcga accgcgtcac 1860

cggcatcgtc agccgcggcg gctctatcca tgccaagctc tgcctgctgg accacaccga 1920

gaacttcgcc tacatgcact gggacgagat cctggacatc tgcgcccagt acgacatcac 1980

cctgtcgatc ggcgacggcc tgcgccccgg ctgcatcgcg gacgccaacg acgctgccca 2040

gtttgctgag ctgaagaccc agggcgagct gacccgccgc gcctgggcga aggacgtgca 2100

ggtgatgaac gagggtcccg gccatgtgcc gctgcacaag atccccgaga acatggagaa 2160

gcagctggac tggtgcagcg aggcgccctt ctacacgctc ggcccccctgg cgacggacat 2220

cgcgcctgcg tacgaccaca tcacctccgc gatcggcgcc gccacgatcg gcgcgctggg 2280

gaccgcgctg ctgtgttacg tgaccccgaa ggagcacctg ggcctccccg atcgcgacga 2340

cgtgaaggcc ggggtgatcg cctacaagat cgcggcgcac gcggctgacc tggcgaaggg 2400

tcacccctac gctcaggagt gggataacgc cctctctaag gcgcgcttcg agttccggtg 2460

gtacgaccag ttccacctgt cgctggaccc cgtcaccgcc cgcctgttcc atgacgcgac 2520

cctgccgcag gagcccgcca agaccgccca cttctgctcc atgtgcggcc ccaagttctg 2580

ctccatgcag atcacccagg acctccgcga gtacgccaag aaccatcaaa tggaggagga 2640

cgaggcgatc caaaccggca tggccgagat gtccgagcag ttcaagagca tcggcgcgga 2700

ggtctacctc gacgaggcgg tgctggagca cgcgtaacaa ttgaagcagc agcagctcgg 2760

atagtatcga cacactctgg acgctggtcg tgtgatggac tgttgccgcc acacttgctg 2820

ccttgacctg tgaatatccc tgccgctttt atcaaacagc ctcagtgtgt ttgatcttgt 2880

gtgtacgcgc ttttgcgagt tgctagctgc ttgtgctatt tgcgaatacc acccccagca 2940

tcccccttccc tcgtttcata tcgcttgcat cccaaccgca acttatctac gctgtcctgc 3000

tatccctcag cgctgctcct gctcctgctc actgcccctc gcacagcctt ggtttgggct 3060

ccgcctgtat tctcctggta ctgcaacctg taaaccagca ctgcaatgct gatgcacggg 3120

aagtagtggg atgggaacac aaatggaaag cttgagctca gaatagtatc gggtgatgcg 3180

aagtcagaac caggcagggc ctgtcgcctg aggtggcaac gatgggaagc aatcaatctg 3240

ggtacagtcg tccgcacgat cccgtgatct cccccaccga cacctatccc cgcccatccc 3300

ggcccaccct ttcagtcccc tcagcatgca ttgtgcaccg cgacaaagca tgtctgctcg 3360

tgcactggtt caggccacgg cgcaccgagt cctcgcccctt cgcagagtga tcaccctccc 3420

cggaaccagc cacgctcgct gctgcgggcc gatcagccgc gcgcactccc tgcaactagg 3480

gacaactcag gcaaccacgc gcctcacaag catggccgcc gtggcatcca acccgctcgt 3540

gacggtgggt gcgcaagtgc caggggcctc gtcgtcacgg cgtgcatcct cgagggatgc 3600

gatccggcaa ctatatgtcg tttatctccc caccaatcac aggatgagcc cctgtctaga 3660

<210> 131

<211> 1905

<212> DNA

<213> Artificial Sequence

<220><223> synthetic nucleotide

<400> 131

actagtatgg ccaccgcatc cactttctcg gcgttcaatg cccgctgcgg cgacctgcgt 60

cgctcggcgg gctccgggcc ccggcgccca gcgaggcccc tccccgtgcg cgggcgcgcc 120

gcgatcgccg ctcccgagcg gctggactac ctcgacaacg cggaggaggc gcggctgcag 180

cagacggacg ccttcgccga gctgaaggcg ctgtcgagcc gccagtccgt gaaccggccc 240

caaaagggcg agctgtcgtt ccggcagtcg ccgacctttc aggactgctt ccctggctcg 300

gagaagtgct accgcgaggt ggagcacgat ggcaagacgc tcaaggtgcc ttttcgccgg 360

gtgcacctgc aggacgacaa cggccatttc gacctgtacg acacctctgg cccccagggc 420

gtgaaccccc gcgagggtct gcccaagatc cgctcctcct gggtggagcc gcgcgaggcg 480

cggggcgaca aggtgcaaac gcagcagtac tacgccaagc aaggcatcat caccgaggag 540

atggccttct gcgctgcgcg cgagcgcatg gaccccgagt tcatccggtc tgaggtggcc 600

cggggtcggg ccatcatccc tgcgaacaag cgccatctgg agctggagcc gaccgtcgtg 660

ggccgcaact ttctggtcaa ggtgaacgcc aacatcggca actctgcggt cagctcgagc 720

atcgaggagg aggtggagaa gctgcagtgg agcaccatct ggggtgccga cacggccatg 780

gacctgagca ccggcaacaa catccacgag acgcgcgagt gggtgatgcg caacagcccg 840

gtccctgtgg gtacggtccc tatctaccag tgcctggaga aggcgggcgg catcgtggag 900

aacatcacct gggagctgtt ccgggagacg ctgatcgagc aggccgagca gggcgtcgac 960

tactttacca tccacgcggg ggtcctgctc cgctacatcc ccctgacggc gaaccgcgtc 1020

accggcatcg tcagccgcgg cggctctatc catgccaagc tctgcctgct ggaccacacc 1080

gagaacttcg cctacatgca ctgggacgag atcctggaca tctgcgccca gtacgacatc 1140

accctgtcga tcggcgacgg cctgcgcccc ggctgcatcg cggacgccaa cgacgctgcc 1200

cagtttgctg agctgaagac ccagggcgag ctgacccgcc gcgcctgggc gaaggacgtg 1260

caggtgatga acgagggtcc cggccatgtg ccgctgcaca agatccccga gaacatggag 1320

aagcagctgg actggtgcag cgaggcgccc ttctacacgc tcggcccccc ggcgacggac 1380

atcgcgcctg cgtacgacca catcacctcc gcgatcggcg ccgccacgat cggcgcgctg 1440

gggaccgcgc tgctgtgtta cgtgaccccg aaggagcacc tgggcctccc cgatcgcgac 1500

gacgtgaagg ccggggtgat cgcctacaag atcgcggcgc acgcggctga cctggcgaag 1560

ggtcacccct acgctcagga gtgggataac gccctctcta aggcgcgctt cgagttccgg 1620

tggtacgacc agttccacct gtcgctggac cccgtcaccg cccgcctgtt ccatgacgcg 1680

accctgccgc aggagcccgc caagaccgcc cacttctgct ccatgtgcgg ccccaagttc 1740

tgctccatgc agatcaccca ggacctccgc gagtacgcca agaaccatca aatggaggag 1800

gacgaggcga tccaaaccgg catggccgag atgtccgagc agttcaagag catcggcgcg 1860

gaggtctacc tcgacgaggc ggtgctggag cacgcgtaac aattg 1905

<210> 132

<211> 1182

<212> DNA

<213> Chlorella protothecoides

<400> 132

actagagagt ttaggtccag cgtccgtggg gggggacggg ctgggagctt gggccgggaa 60

gggcaagacg atgcagtccc tctggggagt cacagccgac tgtgtgtgtt gcactgtgcg 120

gcccgcagca ctcacacgca aaatgcctgg ccgacaggca ggccctgtcc agtgcaacat 180

ccacggtccc tctcatcagg ctcaccttgc tcattgacat aacggaatgc gtaccgctct 240

ttcagatctg tccatccaga gaggggagca ggctccccac cgacgctgtc aaacttgctt 300

cctgcccaac cgaaaacatt attgtttgag gggggggggg gggggcaga ttgcatggcg 360

ggatatctcg tgaggaacat cactgggaca ctgtggaaca cagtgagtgc agtatgcaga 420

gcatgtatgc taggggtcag cgcaggaagg gggcctttcc cagtctccca tgccactgca 480

ccgtatccac gactcaccag gaccagcttc ttgatcggct tccgctcccg tggacaccag 540

tgtgtagcct ctggactcca ggtatgcgtg caccgcaaag gccagccgat cgtgccgatt 600

cctggggtgg aggatatgag tcagccaact tggggctcag agtgcacact ggggcacgat 660

acgaaacaac atctacaccg tgtcctccat gctgacacac cacagcttcg ctccacctga 720

atgtgggcgc atgggcccga atcacagcca atgtcgctgc tgccataatg tgatccagac 780

cctctccgcc cagatgccga gcggatcgtg ggcgctgaat agattcctgt ttcgatcact 840

gtttgggtcc tttccttttc gtctcggatg cgcgtctcga aacaggctgc gtcgggcttt 900

cggatccctt ttgctccctc cgtcaccatc ctgcgcgcgg gcaagttgct tgaccctggg 960

ctggtaccag ggttggaggg tattaccgcg tcaggccatt cccagcccgg attcaattca 1020

aagtctgggc caccaccctc cgccgctctg tctgatcact ccacattcgt gcatacacta 1080

cgttcaagtc ctgatccagg cgtgtctcgg gacaaggtgt gcttgagttt gaatctcaag 1140

gacccactcc agcacagctg ctggttgacc ccgccctcgc aa 1182

<210> 133

<211> 161

<212> DNA

<213> Chlorella protothecoides

<400> 133

acggagcgtc gtgcgggagg gagtgtgccg agcggggagt cccggtctgt gcgaggcccg 60

gcagctgacg ctggcgagcc gtacgccccg agggtccccc tcccctgcac cctcttcccc 120

ttccctctga cggccgcgcc tgttcttgca tgttcagcga c 161

<210> 134

<211> 6306

<212> DNA

<213> Arabidopsis thaliana

<400> 134

gctcttcgcc gccgccactc ctgctcgagc gcgcccgcgc gtgcgccgcc agcgccttgg 60

ccttttcgcc gcgctcgtgc gcgtcgctga tgtccatcac caggtccatg aggtctgcct 120

tgcgccggct gagccactgc ttcgtccggg cggccaagag gagcatgagg gaggactcct 180

ggtccagggt cctgacgtgg tcgcggctct gggagcgggc cagcatcatc tggctctgcc 240

gcaccgaggc cgcctccaac tggtcctcca gcagccgcag tcgccgccga ccctggcaga 300

ggaagacagg tgaggggggt atgaattgta cagaacaacc acgagccttg tctaggcaga 360

atccctacca gtcatggctt tacctggatg acggcctgcg aacagctgtc cagcgaccct 420

cgctgccgcc gcttctcccg cacgcttctt tccagcaccg tgatggcgcg agccagcgcc 480

gcacgctggc gctgcgcttc gccgatctga ggacagtcgg ggaactctga tcagtctaaa 540

ccccccttgcg cgttagtgtt gccatccttt gcagaccggt gagagccgac ttgttgtgcg 600

ccacccccca caccacctcc tcccagacca attctgtcac ctttttggcg aaggcatcgg 660

cctcggcctg cagagaggac agcagtgccc agccgctggg ggttggcgga tgcacgctca 720

ggtacccttt cttgcgctat gacacttcca gcaaaaggta gggcgggctg cgagacggct 780

tcccggcgct gcatgcaaca ccgatgatgc ttcgaccccc cgaagctcct tcggggctgc 840

atgggcgctc cgatgccgct ccagggcgag cgctgtttaa atagccaggc ccccgattgc 900

aaagacatta tagcgagcta ccaaagccat attcaaacac ctagatcact accacttcta 960

cacaggccac tcgagcttgt gatcgcactc cgctaagggg gcgcctcttc ctcttcgttt 1020

cagtcacaac ccgcaaactc tagaatatca atgatcgagc aggacggcct ccacgccggc 1080
tcccccgccg cctgggtgga gcgcctgttc ggctacgact gggcccagca gaccatcggc 1140
tgctccgacg ccgccgtgtt ccgcctgtcc gcccagggcc gccccgtgct gttcgtgaag 1200
accgacctgt ccggcgccct gaacgagctg caggacgagg ccgcccgcct gtcctggctg 1260
gccaccaccg gcgtgccctg cgccgccgtg ctggacgtgg tgaccgaggc cggccgcgac 1320
tggctgctgc tgggcgaggt gcccggccag gacctgctgt cctcccacct ggcccccgcc 1380
gagaaggtgt ccatcatggc cgacgccatg cgccgcctgc acaccctgga ccccgccacc 1440

tgcccttcg accaccaggc caagcaccgc atcgagcgcg cccgcacccg catggaggcc 1500
ggcctggtgg accaggacga cctggacgag gagcaccagg gcctggcccc cgccgagctg 1560
ttcgcccgcc tgaaggcccg catgcccgac ggcgaggacc tggtggtgac ccacggcgac 1620
gcctgcctgc ccaacatcat ggtggagaac ggccgcttct ccggcttcat cgactgcggc 1680
cgcctgggcg tggccgaccg ctaccaggac atcgccctgg ccacccgcga catcgccgag 1740
gagctgggcg gcgagtgggc cgaccgcttc ctggtgctgt acggcatcgc cgcccccgac 1800
tcccagcgca tcgccttcta ccgcctgctg gacgagttct tctgacaatt ggcagcagca 1860

gctcggatag tatcgacaca ctctggacgc tggtcgtgtg atggactgtt gccgccacac 1920
ttgctgcctt gacctgtgaa tatccctgcc gcttttatca aacagcctca gtgtgtttga 1980
tcttgtgtgt acgcgctttt gcgagttgct agctgcttgt gctatttgcg aataccaccc 2040
ccagcatccc cttccctcgt ttcatatcgc ttgcatccca accgcaactt atctacgctg 2100
tcctgctatc cctcagcgct gctcctgctc ctgctcactg cccctcgcac agccttggtt 2160
tgggctccgc ctgtattctc ctggtactgc aacctgtaaa ccagcactgc aatgctgatg 2220
cacgggaagt agtgggatgg gaacacaaat ggaggatcca ctagagagtt taggtccagc 2280

gtccgtgggg ggggacgggc tgggagcttg ggccgggaag ggcaagacga tgcagtccct 2340
ctggggagtc acagccgact gtgtgtgttg cactgtgcgg cccgcagcac tcacacgcaa 2400
aatgcctggc cgacaggcag gccctgtcca gtgcaacatc cacggtccct ctcatcaggc 2460
tcaccttgct cattgacata acggaatgcg taccgctctt tcagatctgt ccatccagag 2520
aggggagcag gctccccacc gacgctgtca aacttgcttc ctgcccaacc gaaaacatta 2580
ttgtttgagg gggggggggg gggggcagat tgcatggcgg gatatctcgt gaggaacatc 2640
actgggacac tgtggaacac agtgagtgca gtatgcagag catgtatgct aggggtcagc 2700

gcaggaaggg ggcctttccc agtctcccat gccactgcac cgtatccacg actcaccagg 2760

accagcttct tgatcggctt ccgctcccgt ggacaccagt gtgtagcctc tggactccag 2820

gtatgcgtgc accgcaaagg ccagccgatc gtgccgattc ctggggtgga ggatatgagt 2880

cagccaactt ggggctcaga gtgcacactg gggcacgata cgaaacaaca tctacaccgt 2940

gtcctccatg ctgacacacc acagcttcgc tccacctgaa tgtgggcgca tgggcccgaa 3000

tcacagccaa tgtcgctgct gccataatgt gatccagacc ctctccgccc agatgccgag 3060

cggatcgtgg gcgctgaata gattcctgtt tcgatcactg tttgggtcct ttccttttcg 3120

tctcggatgc gcgtctcgaa acaggctgcg tcgggctttc ggatcccttt tgctccctcc 3180

gtcaccatcc tgcgcgcggg caagttgctt gaccctgggc tggtaccagg gttggagggt 3240

attaccgcgt caggccattc ccagcccgga ttcaattcaa agtctgggcc accaccctcc 3300

gccgctctgt ctgatcactc cacattcgtg catacactac gttcaagtcc tgatccaggc 3360

gtgtctcggg acaaggtgtg cttgagtttg aatctcaagg acccactcca gcacagctgc 3420

tggttgaccc cgccctcgca aactagtatg gccaccgcat ccactttctc ggcgttcaat 3480

gcccgctgcg gcgacctgcg tcgctcggcg ggctccgggc cccggcgccc agcgaggccc 3540

ctccccgtgc gcgggcgcgc cgtccaggcc gcggccaccc gcttcaagaa ggagacgacg 3600

accacccgcg ccacgctgac gttcgacccc cccacgacca actccgagcg cgccaagcag 3660

cgcaagcaca ccatcgaccc ctcctccccc gacttccagc ccatcccctc cttcgaggag 3720

tgcttcccca agtccacgaa ggagcacaag gaggtggtgc acgaggagtc cggccacgtc 3780

ctgaaggtgc ccttccgccg cgtgcacctg tccggcggcg agcccgcctt cgacaactac 3840

gacacgtccg gcccccagaa cgtcaacgcc cacatcggcc tggcgaagct gcgcaaggag 3900

tggatcgacc gccgcgagaa gctgggcacg ccccgctaca cgcagatgta ctacgcgaag 3960

cagggcatca tcacggagga gatgctgtac tgcgcgacgc gcgagaagct ggaccccgag 4020

ttcgtccgct ccgaggtcgc gcggggccgc gccatcatcc cctccaacaa gaagcacctg 4080

gagctggagc ccatgatcgt gggccgcaag ttcctggtga aggtgaacgc gaacatcggc 4140

aactccgccg tggcctcctc catcgaggag gaggtctaca aggtgcagtg ggccaccatg 4200

tggggcgccg acaccatcat ggacctgtcc acgggccgcc acatccacga gacgcgcgag 4260

tggatcctgc gcaactccgc ggtccccgtg ggcaccgtcc ccatctacca ggcgctggag 4320

aaggtggacg gcatcgcgga gaacctgaac tgggaggtgt tccgcgagac gctgatcgag 4380

caggccgagc agggcgtgga ctacttcacg atccacgcgg gcgtgctgct gcgctacatc 4440

cccctgaccg ccaagcgcct gacgggcatc gtgtcccgcg gcggctccat ccacgcgaag 4500

tggtgcctgg cctaccacaa ggagaacttc gcctacgagc actgggacga catcctggac 4560

atctgcaacc agtacgacgt cgccctgtcc atcggcgacg gcctgcgccc cggctccatc 4620

tacgacgcca acgacacggc ccagttcgcc gagctgctga cccagggcga gctgacgcgc 4680

cgcgcgtggg agaaggacgt gcaggtgatg aacgagggcc ccggccacgt gcccatgcac 4740

aagatccccg agaacatgca gaagcagctg gagtggtgca acgaggcgcc cttctacacc 4800

ctgggccccc tgacgaccga catcgcgccc ggctacgacc acatcacctc cgccatcggc 4860

gcggccaaca tcggcgccct gggcaccgcc ctgctgtgct acgtgacgcc caaggagcac 4920

ctgggcctgc ccaaccgcga cgacgtgaag gcgggcgtca tcgcctacaa gatcgccgcc 4980

cacgcggccg acctggccaa gcagcacccc cacgcccagg cgtgggacga cgcgctgtcc 5040

aaggcgcgct tcgagttccg ctggatggac cagttcgcgc tgtccctgga ccccatgacg 5100

gcgatgtcct tccacgacga gacgctgccc gcggacggcg cgaaggtcgc ccacttctgc 5160

tccatgtgcg gccccaagtt ctgctccatg aagatcacgg aggacatccg caagtacgcc 5220

gaggagaacg gctacggctc cgccgaggag gccatccgcc agggcatgga cgccatgtcc 5280

gaggagttca acatcgccaa gaagacgatc tccggcgagc agcacggcga ggtcggcggc 5340

gagatctacc tgcccgagtc ctacgtcaag gccgcgcaga agtgatacgt actcgagacg 5400

gagcgtcgtg cgggagggag tgtgccgagc ggggagtccc ggtctgtgcg aggcccggca 5460

gctgacgctg gcgagccgta cgccccgagg gtccccctcc cctgcaccct cttcccttc 5520

cctctgacgg ccgcgcctgt tcttgcatgt tcagcgacga gctcttgttt tccagaagga 5580

gttgctcctt gagcctttca ttctcagcct cgataacctc caaagccgct ctaattgtgg 5640

agggggttcg aatttaaaag cttggaatgt tggttcgtgc gtctggaaca agcccagact 5700

tgttgctcac tgggaaaagg accatcagct ccaaaaaact tgccgctcaa accgcgtacc 5760

tctgctttcg cgcaatctgc cctgttgaaa tcgccaccac attcatattg tgacgcttga 5820

gcagtctgta attgcctcag aatgtggaat catctgcccc ctgtgcgagc ccatgccagg 5880

catgtcgcgg gcgaggacac ccgccactcg tacagcagac cattatgcta cctcacaata 5940

gttcataaca gtgaccatat ttctcgaagc tccccaacga gcacctccat gctctgagtg 6000

gccacccccc ggccctggtg cttgcggagg gcaggtcaac cggcatgggg ctaccgaaat 6060

ccccgaccgg atcccaccac ccccgcgatg ggaagaatct ctccccggga tgtgggccca 6120

ccaccagcac aacctgctgg cccaggcgag cgtcaaacca taccacacaa atatccttgg 6180

catcggccct gaattccttc tgccgctctg ctacccggtg cttctgtccg aagcaggggt 6240

tgctagggat cgctccgagt ccgcaaaccc ttgtcgcgtg gcggggcttg ttcgagcttg 6300

aagagc 6306

<210> 135

<211> 1938

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 135

atggccaccg catccacttt ctcggcgttc aatgcccgct gcggcgacct gcgtcgctcg 60

gcgggctccg ggccccggcg cccagcgagg cccctccccg tgcgcgggcg cgccgtccag 120

gccgcggcca cccgcttcaa gaaggagacg acgaccaccc gcgccacgct gacgttcgac 180

ccccccacga ccaactccga gcgcgccaag cagcgcaagc acaccatcga cccctcctcc 240

cccgacttcc agcccatccc ctccttcgag gagtgcttcc ccaagtccac gaaggagcac 300

aaggaggtgg tgcacgagga gtccggccac gtcctgaagg tgcccttccg ccgcgtgcac 360

ctgtccggcg gcgagcccgc cttcgacaac tacgacacgt ccggccccca gaacgtcaac 420

gcccacatcg gcctggcgaa gctgcgcaag gagtggatcg accgccgcga gaagctgggc 480

acgccccgct acacgcagat gtactacgcg aagcagggca tcatcacgga ggagatgctg 540

tactgcgcga cgcgcgagaa gctggacccc gagttcgtcc gctccgaggt cgcgcggggc 600

cgcgccatca tcccctccaa caagaagcac ctggagctgg agcccatgat cgtgggccgc 660

aagttcctgg tgaaggtgaa cgcgaacatc ggcaactccg ccgtggcctc ctccatcgag 720

gaggaggtct acaaggtgca gtgggccacc atgtggggcg ccgacaccat catggacctg 780

tccacgggcc gccacatcca cgagacgcgc gagtggatcc tgcgcaactc cgcggtcccc 840

gtgggcaccg tccccatcta ccaggcgctg gagaaggtgg acggcatcgc ggagaacctg 900

aactgggagg tgttccgcga gacgctgatc gagcaggccg agcagggcgt ggactacttc 960

acgatccacg cgggcgtgct gctgcgctac atccccctga ccgccaagcg cctgacgggc 1020

atcgtgtccc gcggcggctc catccacgcg aagtggtgcc tggcctacca caaggagaac 1080

ttcgcctacg agcactggga cgacatcctg gacatctgca accagtacga cgtcgccctg 1140

tccatcggcg acggcctgcg ccccggctcc atctacgacg ccaacgacac ggcccagttc 1200

gccgagctgc tgacccaggg cgagctgacg cgccgcgcgt gggagaagga cgtgcaggtg 1260

atgaacgagg gccccggcca cgtgcccatg cacaagatcc ccgagaacat gcagaagcag 1320

ctggagtggt gcaacgaggc gcccttctac accctgggcc ccctgacgac cgacatcgcg 1380

cccggctacg accacatcac ctccgccatc ggcgcggcca acatcggcgc cctgggcacc 1440

gccctgctgt gctacgtgac gcccaaggag cacctgggcc tgcccaaccg cgacgacgtg 1500

aaggcgggcg tcatcgccta caagatcgcc gcccacgcgg ccgacctggc caagcagcac 1560

ccccacgccc aggcgtggga cgacgcgctg tccaaggcgc gcttcgagtt ccgctggatg 1620

gaccagttcg cgctgtccct ggaccccatg acggcgatgt ccttccacga cgagacgctg 1680

cccgcggacg gcgcgaaggt cgcccacttc tgctccatgt gcggccccaa gttctgctcc 1740

atgaagatca cggaggacat ccgcaagtac gccgaggaga acggctacgg ctccgccgag 1800

gaggccatcc gccagggcat ggacgccatg tccgaggagt tcaacatcgc caagaagacg 1860

atctccggcg agcagcacgg cgaggtcggc ggcgagatct acctgcccga gtcctacgtc 1920

aaggccgcgc agaagtga 1938

<210> 136

<211> 644

<212> PRT

<213> Arabidopsis thaliana

<400> 136

Met Ala Ala Ser Val His Cys Thr Leu Met Ser Val Val Cys Asn Asn
1 5 10 15

Lys Asn His Ser Ala Arg Pro Lys Leu Pro Asn Ser Ser Leu Leu Pro
20 25 30

Gly Phe Asp Val Val Val Gln Ala Ala Ala Thr Arg Phe Lys Lys Glu
35 40 45

Thr Thr Thr Thr Arg Ala Thr Leu Thr Phe Asp Pro Pro Thr Thr Asn
50 55 60

Ser Glu Arg Ala Lys Gln Arg Lys His Thr Ile Asp Pro Ser Ser Pro
65 70 75 80

Asp Phe Gln Pro Ile Pro Ser Phe Glu Glu Cys Phe Pro Lys Ser Thr
85 90 95

Lys Glu His Lys Glu Val Val His Glu Glu Ser Gly His Val Leu Lys
100 105 110

Val Pro Phe Arg Arg Val His Leu Ser Gly Gly Glu Pro Ala Phe Asp
115 120 125

Asn Tyr Asp Thr Ser Gly Pro Gln Asn Val Asn Ala His Ile Gly Leu
130 135 140

Ala Lys Leu Arg Lys Glu Trp Ile Asp Arg Arg Glu Lys Leu Gly Thr

145 150 155 160

Pro Arg Tyr Thr Gln Met Tyr Tyr Ala Lys Gln Gly Ile Ile Thr Glu

165 170 175

Glu Met Leu Tyr Cys Ala Thr Arg Glu Lys Leu Asp Pro Glu Phe Val

180 185 190

Arg Ser Glu Val Ala Arg Gly Arg Ala Ile Ile Pro Ser Asn Lys Lys

195 200 205

His Leu Glu Leu Glu Pro Met Ile Val Gly Arg Lys Phe Leu Val Lys

210 215 220

Val Asn Ala Asn Ile Gly Asn Ser Ala Val Ala Ser Ser Ile Glu Glu

225 230 235 240

Glu Val Tyr Lys Val Gln Trp Ala Thr Met Trp Gly Ala Asp Thr Ile

245 250 255

Met Asp Leu Ser Thr Gly Arg His Ile His Glu Thr Arg Glu Trp Ile

260 265 270

Leu Arg Asn Ser Ala Val Pro Val Gly Thr Val Pro Ile Tyr Gln Ala

275 280 285

Leu Glu Lys Val Asp Gly Ile Ala Glu Asn Leu Asn Trp Glu Val Phe

290 295 300

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

305 310 315 320

Ile His Ala Gly Val Leu Leu Arg Tyr Ile Pro Leu Thr Ala Lys Arg

325 330 335

Leu Thr Gly Ile Val Ser Arg Gly Gly Ser Ile His Ala Lys Trp Cys

340 345 350

Leu Ala Tyr His Lys Glu Asn Phe Ala Tyr Glu His Trp Asp Asp Ile

355 360 365

Leu Asp Ile Cys Asn Gln Tyr Asp Val Ala Leu Ser Ile Gly Asp Gly

370 375 380

Leu Arg Pro Gly Ser Ile Tyr Asp Ala Asn Asp Thr Ala Gln Phe Ala

385 390 395 400

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

Ile Thr Ser Ala Ile Gly Ala Ala Asn Ile Gly Ala Leu Gly Thr Ala
465 470 475 480
Leu Leu Cys Tyr Val Thr Pro Lys Glu His Leu Gly Leu Pro Asn Arg
485 490 495
Asp Asp Val Lys Ala Gly Val Ile Ala Tyr Lys Ile Ala Ala His Ala
500 505 510
Ala Asp Leu Ala Lys Gln His Pro His Ala Gln Ala Trp Asp Asp Ala
515 520 525

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

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

<210> 137

<211> 6130

<212> DNA

<213> Unknown

<220><223> Synechocystis species

<400> 137

gctcttcgcc gccgccactc ctgctcgagc gcgcccgcgc gtgcgccgcc agcgccttgg 60

cctttcgcc gcgctcgtgc gcgtcgctga tgtccatcac caggtccatg aggtctgcct 120

tgcgccggct gagccactgc ttcgtccggg cggccaagag gagcatgagg gaggactcct 180

ggtccagggt cctgacgtgg tcgcggctct gggagcgggc cagcatcatc tggctctgcc 240

gcaccgaggc cgcctccaac tggtcctcca gcagccgcag tcgccgccga ccctggcaga 300

ggaagacagg tgagggggt atgaattgta cagaacaacc acgagccttg tctaggcaga 360

atccctacca gtcatggctt tacctggatg acggcctgcg aacagctgtc cagcgaccct 420

cgctgccgcc gcttctcccg cacgcttctt tccagcaccg tgatggcgcg agccagcgcc 480

gcacgctggc gctgcgcttc gccgatctga ggacagtcgg ggaactctga tcagtctaaa 540

ccccttgcg cgttagtgtt gccatccttt gcagaccggt gagagccgac ttgttgtgcg 600

ccacccccca caccacctcc tcccagacca attctgtcac ctttttggcg aaggcatcgg 660

cctcggcctg cagagaggac agcagtgccc agccgctggg ggttggcgga tgcacgctca 720

ggtacccttt cttgcgctat gacacttcca gcaaaaggta gggcgggctg cgagacggct 780

tcccggcgct gcatgcaaca ccgatgatgc ttcgaccccc cgaagctcct tcggggctgc 840

atgggcgctc cgatgccgct ccagggcgag cgctgtttaa atagccaggc ccccgattgc 900

aaagacatta tagcgagcta ccaaagccat attcaaacac ctagatcact accacttcta 960

cacaggccac tcgagcttgt gatcgcactc cgctaagggg gcgcctcttc ctcttcgttt 1020

cagtcacaac ccgcaaactc tagaatatca atgatcgagc aggacggcct ccacgccggc 1080

tcccccgccg cctgggtgga gcgcctgttc ggctacgact gggcccagca gaccatcggc 1140

tgctccgacg ccgccgtgtt ccgcctgtcc gcccagggcc gccccgtgct gttcgtgaag 1200

accgacctgt ccggcgccct gaacgagctg caggacgagg ccgcccgcct gtcctggctg 1260

gccaccaccg gcgtgccctg cgccgccgtg ctggacgtgg tgaccgaggc cggccgcgac 1320

tggctgctgc tgggcgaggt gcccggccag gacctgctgt cctcccacct ggcccccgcc 1380

gagaaggtgt ccatcatggc cgacgccatg cgccgcctgc acaccctgga ccccgccacc 1440

tgccccttcg accaccaggc caagcaccgc atcgagcgcg cccgcacccg catggaggcc 1500

ggcctggtgg accaggacga cctggacgag gagcaccagg gcctggcccc cgccgagctg 1560

ttcgcccgcc tgaaggcccg catgcccgac ggcgaggacc tggtggtgac ccacggcgac 1620

gcctgcctgc ccaacatcat ggtggagaac ggccgcttct ccggcttcat cgactgcggc 1680

cgcctgggcg tggccgaccg ctaccaggac atcgccctgg ccacccgcga catcgccgag 1740

gagctgggcg gcgagtgggc cgaccgcttc ctggtgctgt acggcatcgc cgcccccgac 1800

tcccagcgca tcgccttcta ccgcctgctg gacgagttct tctgacaatt ggcagcagca 1860

gctcggatag tatcgacaca ctctggacgc tggtcgtgtg atggactgtt gccgccacac 1920

ttgctgcctt gacctgtgaa tatccctgcc gcttttatca aacagcctca gtgtgtttga 1980

tcttgtgtgt acgcgctttt gcgagttgct agctgcttgt gctatttgcg aataccaccc 2040

ccagcatccc cttccctcgt ttcatatcgc ttgcatccca accgcaactt atctacgctg 2100

tcctgctatc cctcagcgct gctcctgctc ctgctcactg cccctcgcac agccttggtt 2160

tgggctccgc ctgtattctc ctggtactgc aacctgtaaa ccagcactgc aatgctgatg 2220

cacgggaagt agtgggatgg gaacacaaat ggaggatcca ctagagagtt taggtccagc 2280

gtccgtgggg ggggacgggc tgggagcttg ggccgggaag ggcaagacga tgcagtccct 2340

ctggggagtc acagccgact gtgtgtgttg cactgtgcgg cccgcagcac tcacacgcaa 2400

aatgcctggc cgacaggcag gccctgtcca gtgcaacatc cacggtccct ctcatcaggc 2460

tcaccttgct cattgacata acggaatgcg taccgctctt tcagatctgt ccatccagag 2520

aggggagcag gctccccacc gacgctgtca aacttgcttc ctgcccaacc gaaaacatta 2580

ttgtttgagg gggggggggg gggggcagat tgcatggcgg gatatctcgt gaggaacatc 2640

actgggacac tgtggaacac agtgagtgca gtatgcagag catgtatgct aggggtcagc 2700

gcaggaaggg ggcctttccc agtctcccat gccactgcac cgtatccacg actcaccagg 2760

accagcttct tgatcggctt ccgctcccgt ggacaccagt gtgtagcctc tggactccag 2820

gtatgcgtgc accgcaaagg ccagccgatc gtgccgattc ctggggtgga ggatatgagt 2880

cagccaactt ggggctcaga gtgcacactg gggcacgata cgaaacaaca tctacaccgt 2940

gtcctccatg ctgacacacc acagcttcgc tccacctgaa tgtgggcgca tgggcccgaa 3000

tcacagccaa tgtcgctgct gccataatgt gatccagacc ctctccgccc agatgccgag 3060

cggatcgtgg gcgctgaata gattcctgtt tcgatcactg tttgggtcct ttccttttcg 3120

tctcggatgc gcgtctcgaa acaggctgcg tcgggctttc ggatcccttt tgctccctcc 3180

gtcaccatcc tgcgcgcggg caagttgctt gaccctgggc tggtaccagg gttggagggt 3240

attaccgcgt caggccattc ccagcccgga ttcaattcaa agtctgggcc accaccctcc 3300

gccgctctgt ctgatcactc cacattcgtg catacactac gttcaagtcc tgatccaggc 3360

gtgtctcggg acaaggtgtg cttgagtttg aatctcaagg acccactcca gcacagctgc 3420
tggttgaccc cgccctcgca aactagtatg gccaccgcat ccactttctc ggcgttcaat 3480
gcccgctgcg gcgacctgcg tcgctcggcg ggctccgggc cccggcgccc agcgaggccc 3540
ctccccgtgc gcgggcgcgc catgcgcacc gcgtgggtcg ccaagcgcca gggccagacg 3600
aacgtctccc agatgcacta cgcgcgcaag ggcgtgatca ccgaggagat ggactacgtg 3660
gcgaagcgcg agaacctgcc cgtggagctg atcaaggacg aggtggcgcg gggccgcatg 3720
atcatccccg ccaacatcaa ccacacgaac ctggagccca tggccatcgg catcgcctcc 3780

aagtgcaagg tcaacgcgaa catcggcgcg tcccccaact cctccaacat cgacgaggag 3840
gtggagaagc tgctgctgtc cgtgaagtac ggcgcggaca cggtgatgga cctgtccacc 3900
ggcggcggcg acctggacgt gatccgcacc gccatcatca acgcctcccc cgtgcccatc 3960
ggcacggtgc ccatctacca ggccctggag tccgtgcacg gctccatcga gaacctgacg 4020
cccgacgact tcctgcacat catcgagaag cacgcccagc agggcgtgga ctacatgacg 4080
atccacgcgg gcctgctgat cgagtacctg cccctggtga agtcccgcat cacgggcatc 4140
gtgtcccgcg gcggcggcat catcgccaag tggatgctgc accaccacaa gcagaacccc 4200

ctgtacacgc acttcgacga gatcatcgag atcttcaaga agtacgacgt gtccttctcc 4260
ctgggcgact ccctgcgccc cggctgcacc cacgacgcct ccgacgacgc ccagctgtcc 4320
gagctgaaga cgctgggcca gctgacgcgc cgcgcctggg agcacgacgt ccaggtcatg 4380
gtggagggcc ccggccacgt ccccatcgac cagatcgagt tcaacgtcaa gaagcagatg 4440
gaggagtgct ccgaggcccc cttctacgtg ctgggccccc tggtcaccga catcgcgccc 4500
ggctacgacc acatcacctc cgcgatcggc gccgccatcg cgggctggca cggcaccgcg 4560
atgctgtgct acgtgacgcc caaggagcac ctgggcctgc ccaacgcgga ggacgtgcgc 4620

aacggcctga tcgcctacaa gatcgccgcg cacgccgcgg acatcgcccg ccaccgcccc 4680
ggcgcgcgcg accgcgacga cgagctgtcc aaggcgcgct acaacttcga ctggaaccgc 4740
cagttcgagc tgtccctgga ccccgagcgc gcgaaggagt accacgacga gacgctgccc 4800
gcggacatct acaagaccgc ggagttctgc tccatgtgcg gccccaagtt ctgccccatg 4860
cagaccaagg tggacgccga gatgctggag gagctggagg tcttcctggc caaggacaag 4920
gagatggtgt cccagcgctg atacgtactc gagacggagc gtcgtgcggg agggagtgtg 4980
ccgagcgggg agtcccggtc tgtgcgaggc ccggcagctg acgctggcga gccgtacgcc 5040

ccgagggtcc ccctcccctg caccctcttc cccttccctc tgacggccgc gcctgttctt 5100

gcatgttcag cgacgagctc ttgttttcca gaaggagttg ctccttgagc ctttcattct 5160

cagcctcgat aacctccaaa gccgctctaa ttgtggaggg ggttcgaatt taaaagcttg 5220

gaatgttggt tcgtgcgtct ggaacaagcc cagacttgtt gctcactggg aaaaggacca 5280

tcagctccaa aaaacttgcc gctcaaaccg cgtacctctg ctttcgcgca atctgccctg 5340

ttgaaatcgc caccacattc atattgtgac gcttgagcag tctgtaattg cctcagaatg 5400

tggaatcatc tgccccctgt gcgagcccat gccaggcatg tcgcgggcga ggacacccgc 5460

cactcgtaca gcagaccatt atgctacctc acaatagttc ataacagtga ccatatttct 5520

cgaagctccc caacgagcac ctccatgctc tgagtggcca ccccccggcc ctggtgcttg 5580

cggagggcag gtcaaccggc atggggctac cgaaatcccc gaccggatcc caccaccccc 5640

gcgatgggaa gaatctctcc ccgggatgtg ggcccaccac cagcacaacc tgctggccca 5700

ggcgagcgtc aaaccatacc acacaaatat ccttggcatc ggccctgaat tccttctgcc 5760

gctctgctac ccggtgcttc tgtccgaagc aggggttgct agggatcgct ccgagtccgc 5820

aaaccccttgt cgcgtggcgg ggcttgttcg agcttgaaga gcctctagag tcgacctgca 5880

ggcatgcaag cttggcgtaa tcatggtcat agctgtttcc tgtgtgaaat tgttatccgc 5940

tcacaattcc acacaacata cgagccggaa gcataaagtg taaagcctgg ggtgcctaat 6000

gagtgagcta actcacatta attgcgttgc gctcactgcc cgctttccag tcgggaaacc 6060

tgtcgtgcca gctgcattaa tgaatcggcc aacgcgcggg gagaggcggt ttgcgtattg 6120

ggcgctcttc 6130

<210> 138

<211> 1494

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 138

atggccaccg catccacttt ctcggcgttc aatgcccgct gcggcgacct gcgtcgctcg 60

gcgggctccg ggccccggcg cccagcgagg cccctccccg tgcgcgggcg cgccatgcgc 120

accgcgtggg tcgccaagcg ccagggccag acgaacgtct cccagatgca ctacgcgcgc 180

aagggcgtga tcaccgagga gatggactac gtggcgaagc gcgagaacct gcccgtggag 240

ctgatcaagg acgaggtggc gcggggccgc atgatcatcc ccgccaacat caaccacacg 300

aacctggagc ccatggccat cggcatcgcc tccaagtgca aggtcaacgc gaacatcggc 360

gcgtccccca actcctccaa catcgacgag gaggtggaga agctgctgct gtccgtgaag 420

tacggcgcgg acacggtgat ggacctgtcc accggcggcg gcgacctgga cgtgatccgc 480

accgccatca tcaacgcctc ccccgtgccc atcggcacgg tgcccatcta ccaggccctg 540
gagtccgtgc acggctccat cgagaacctg acgcccgacg acttcctgca catcatcgag 600
aagcacgccc agcagggcgt ggactacatg acgatccacg cgggcctgct gatcgagtac 660
ctgcccctgg tgaagtcccg catcacgggc atcgtgtccc gcggcggcgg catcatcgcc 720
aagtggatgc tgcaccacca caagcagaac cccctgtaca cgcacttcga cgagatcatc 780
gagatcttca agaagtacga cgtgtccttc tccctgggcg actccctgcg ccccggctgc 840
acccacgacg cctccgacga cgcccagctg tccgagctga agacgctggg ccagctgacg 900

cgccgcgcct gggagcacga cgtccaggtc atggtggagg gccccggcca cgtccccatc 960
gaccagatcg agttcaacgt caagaagcag atggaggagt gctccgaggc ccccttctac 1020
gtgctgggcc ccctggtcac cgacatcgcg cccggctacg accacatcac ctccgcgatc 1080
ggcgccgcca tcgcgggctg gcacggcacc gcgatgctgt gctacgtgac gcccaaggag 1140
cacctgggcc tgcccaacgc ggaggacgtg cgcaacggcc tgatcgccta caagatcgcc 1200
gcgcacgccg cggacatcgc ccgccaccgc cccggcgcgc gcgaccgcga cgacgagctg 1260
tccaaggcgc gctacaactt cgactggaac cgccagttcg agctgtccct ggaccccgag 1320

cgcgcgaagg agtaccacga cgagacgctg cccgcggaca tctacaagac cgcggagttc 1380
tgctccatgt gcggccccaa gttctgcccc atgcagacca aggtggacgc cgagatgctg 1440
gaggagctgg aggtcttcct ggccaaggac aaggagatgg tgtcccagcg ctga 1494

<210> 139
<211> 459
<212> PRT
<213> Unknown
<220><223> Synechosystis species
<400> 139

Met Arg Thr Ala Trp Val Ala Lys Arg Gln Gly Gln Thr Asn Val Ser
1 5 10 15
Gln Met His Tyr Ala Arg Lys Gly Val Ile Thr Glu Glu Met Asp Tyr

20 25 30
Val Ala Lys Arg Glu Asn Leu Pro Val Glu Leu Ile Lys Asp Glu Val
35 40 45
Ala Arg Gly Arg Met Ile Ile Pro Ala Asn Ile Asn His Thr Asn Leu

50 55 60

Glu Pro Met Ala Ile Gly Ile Ala Ser Lys Cys Lys Val Asn Ala Asn

65 70 75 80

Ile Gly Ala Ser Pro Asn Ser Ser Asn Ile Asp Glu Glu Val Glu Lys

85 90 95

Leu Leu Leu Ser Val Lys Tyr Gly Ala Asp Thr Val Met Asp Leu Ser

100 105 110

Thr Gly Gly Gly Asp Leu Asp Val Ile Arg Thr Ala Ile Ile Asn Ala

115 120 125

Ser Pro Val Pro Ile Gly Thr Val Pro Ile Tyr Gln Ala Leu Glu Ser

130 135 140

Val His Gly Ser Ile Glu Asn Leu Thr Pro Asp Asp Phe Leu His Ile

145 150 155 160

Ile Glu Lys His Ala Gln Gln Gly Val Asp Tyr Met Thr Ile His Ala

165 170 175

Gly Leu Leu Ile Glu Tyr Leu Pro Leu Val Lys Ser Arg Ile Thr Gly

180 185 190

Ile Val Ser Arg Gly Gly Gly Ile Ile Ala Lys Trp Met Leu His His

195 200 205

His Lys Gln Asn Pro Leu Tyr Thr His Phe Asp Glu Ile Ile Glu Ile

210 215 220

Phe Lys Lys Tyr Asp Val Ser Phe Ser Leu Gly Asp Ser Leu Arg Pro

225 230 235 240

Gly Cys Thr His Asp Ala Ser Asp Asp Ala Gln Leu Ser Glu Leu Lys

245 250 255

Thr Leu Gly Gln Leu Thr Arg Arg Ala Trp Glu His Asp Val Gln Val

260 265 270

Met Val Glu Gly Pro Gly His Val Pro Ile Asp Gln Ile Glu Phe Asn

275 280 285

Val Lys Lys Gln Met Glu Glu Cys Ser Glu Ala Pro Phe Tyr Val Leu

290 295 300

Gly Pro Leu Val Thr Asp Ile Ala Pro Gly Tyr Asp His Ile Thr Ser
305 310 315 320
Ala Ile Gly Ala Ala Ile Ala Gly Trp His Gly Thr Ala Met Leu Cys
325 330 335
Tyr Val Thr Pro Lys Glu His Leu Gly Leu Pro Asn Ala Glu Asp Val

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

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

<210> 140
<211> 312
<212> DNA
<213> Artificial Sequence
<220><223> synthesized construct

<400> 140
ctttcttgcg ctatgacact tccagcaaaa ggtagggcgg gctgcgagac ggcttcccgg 60
cgctgcatgc aacaccgatg atgcttcgac cccccgaagc tccttcgggg ctgcatgggc 120
gctccgatgc cgctccaggg cgagcgctgt ttaaatagcc aggcccccga ttgcaaagac 180
attatagcga gctaccaaag ccatattcaa acacctagat cactaccact tctacacagg 240
ccactcgagc ttgtgatcgc actccgctaa gggggcgcct cttcctcttc gtttcagtca 300
caacccgcaa ac 312

<210> 141

<211> 18

<212> DNA

<213> Unknown

<220><223> Coccomyxa species

<400> 141

gggtgatcgc ctacaaga 18

<210> 142

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> synthesized primer

<400> 142

acgtcgcgac ccatgcttcc 20

<210> 143

<211> 18

<212> DNA

<213> Arabidopsis thaliana

<400> 143

gcgtcatcgc ctacaaga 18

<210> 144

<211> 18

<

212> DNA

<213> Unknown

<220><223> Synechosystis sp.

<400> 144

cgatgctgtg ctacgtga 18

<210> 145

<211> 1260

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 145

actagtatgg ctatcaagac gaacaggcag cctgtggaga agcctccgtt cacgatcggg 60

acgctgcgca aggccatccc cgcgcactgt ttcgagcgct cggcgcttcg tgggcgcgcc 120

gccaacgccc acaccgtgcc caagatcaac ggcaacaagg ccggcctgct gacccccatg 180

gagagcacca aggacgagga catcgtcgcg gcccccaccg tggcgcccaa gcgcaccttc 240

atcaaccagc tgcccgactg gtcgatgctg ctggccgcga tcaccaccat cttcctggcg 300

gccgagaagc agtggaccaa cctggactgg aagccccgcc gccccgacat gctggtcgac 360

ttcgacccct tctccctggg ccgcttcgtg caggacggcc tgatcttccg ccagaacttc 420

agcatccgct cgtacgagat cggcgcggac cgcaccgcct ccatcgagac cctgatgaac 480

cacctgcagg agaccgcgct gaaccacgtc cgctgcatcg gcctgctgga cgacggcttc 540

ggcagcaccc ccgagatgac ccgccgcgac ctgatctggg tggtcacccg catgcaggtc 600

ctggtggacc gctacccctc gtggggcgac gtgatcgagg tcgactcctg ggtgaccccc 660

agcggcaaga acggcatgaa gcgcgagtgg ttcctgcgcg actgcaagac cggcgagatc 720

ctgacccgcg ccacctcggt ctgggtgatg atgaacaagc gcacccgccg cctgtccaag 780

atccccgagg aggtccgcgt ggagatcgag ccctacttcg tcgagcacgg cgtgctggac 840

gaggactcgc gcaagctgcc caagctgaac gacaacaccg ccaactacat ccgccgcggc 900

ctggcgcccc gctggtccga cctggacgtc aaccagcacg tgaacaacgt caagtacatc 960

ggctggatcc tggagagcgt gcccagcagc ctgctggagt cgcacgagct gtacggcatg 1020

accctggagt accgcaagga gtgcggcaag gacggcctgc tgcagtccct gaccgccgtc 1080

gccagcgact acggcggcgg ctcgctggag gccggcgtgg agtgcgacca cctgctgcgc 1140

ctggaggacg gctccgagat catgcgcggc aagaccgagt ggcgccccaa gcgcgccgcg 1200

aacaccacct acttcggcag cgtcgacgac atccccccg ccaacaacgc gtgactcgag 1260

<210> 146

<211> 429

<212> PRT

<213> Myristica fragrans

<400> 146

Met Val Ala Thr Ser Ala Ala Ser Ala Phe Phe Pro Val Ala Ser Pro

1 5 10 15

Ser Pro Val Lys Pro Ser Met Met Leu Gly Gly Gly Gly Gly Ser Asp

20 25 30

Asn Leu Asp Ala Arg Gly Ile Lys Ser Arg Pro Ala Ser Ser Gly Gly

35 40 45

Leu Gln Val Lys Ala Asn Ala His Thr Val Pro Lys Ile Asn Gly Asn

50 55 60

Lys Ala Gly Leu Leu Thr Pro Met Glu Ser Thr Lys Asp Glu Asp Ile

65 70 75 80

Val Ala Ala Pro Thr Val Ala Pro Lys Arg Thr Phe Ile Asn Gln Leu

85 90 95

Pro Asp Trp Ser Met Leu Leu Ala Ala Ile Thr Thr Ile Phe Leu Ala

100 105 110

Ala Glu Lys Gln Trp Thr Asn Leu Asp Trp Lys Pro Arg Arg Pro Asp

115 120 125

Met Leu Val Asp Phe Asp Pro Phe Ser Leu Gly Arg Phe Val Gln Asp

130 135 140

Gly Leu Ile Phe Arg Gln Asn Phe Ser Ile Arg Ser Tyr Glu Ile Gly

145 150 155 160

Ala Asp Arg Thr Ala Ser Ile Glu Thr Leu Met Asn His Leu Gln Glu

165 170 175

Thr Ala Leu Asn His Val Arg Cys Ile Gly Leu Leu Asp Asp Gly Phe

180 185 190

Gly Ser Thr Pro Glu Met Thr Arg Arg Asp Leu Ile Trp Val Val Thr

195 200 205

Arg Met Gln Val Leu Val Asp Arg Tyr Pro Ser Trp Gly Asp Val Ile

210 215 220

Glu Val Asp Ser Trp Val Thr Pro Ser Gly Lys Asn Gly Met Lys Arg

225 230 235 240

Glu Trp Phe Leu Arg Asp Cys Lys Thr Gly Glu Ile Leu Thr Arg Ala

245 250 255

Thr Ser Val Trp Val Met Met Asn Lys Arg Thr Arg Arg Leu Ser Lys

260 265 270

Ile Pro Glu Glu Val Arg Val Glu Ile Glu Pro Tyr Phe Val Glu His

275 280 285

Gly Val Leu Asp Glu Asp Ser Arg Lys Leu Pro Lys Leu Asn Asp Asn
290 295 300
Thr Ala Asn Tyr Ile Arg Arg Gly Leu Ala Pro Arg Trp Ser Asp Leu
305 310 315 320
Asp Val Asn Gln His Val Asn Asn Val Lys Tyr Ile Gly Trp Ile Leu
325 330 335
Glu Ser Val Pro Ser Ser Leu Leu Glu Ser His Glu Leu Tyr Gly Met
340 345 350

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

Phe Gly Ser Val Asp Asp Ile Pro Pro Ala Asn Asn Ala
420 425

<210> 147
<211> 1269
<212> DNA
<213> Artificial Sequence
<220><223> synthesized construct
<400> 147

actagtatgg ccaccgcatc cactttctcg gcgttcaatg cccgctgcgg cgacctgcgt 60
cgctcggcgg gctccgggcc ccggcgccca gcgaggcccc tccccgtgcg cgggcgcgcc 120
gccaacgccc aggcccccac caagatcaac ggctccaccg acgacgccca gctgcccgcc 180
ccccgcacct tcatcaacca gctgcccgac tggtccatgc tgctggccgc catcaccacc 240

gtgttcctgg ccgccgagaa gcagtggatg atgctggact ggaagccccg ccgccccgac 300
atgctgatcg acaccttcgg cctgggccgc atcgtgcagg acggcctggt gttccgccag 360
aacttctcca tccgctccta cgagatcggc gccgaccgca ccgcctccat cgagaccgtg 420
atgaaccacc tgcaggagac cgccctgaac cacgtgaaga ccgccggcct gctgggcgac 480

ggcttcggct ccacccccga gatgtccaag cgcaacctga tctgggtggt gaccaagatg 540

caggtggagg tggaccgcta ccccacctgg ggcgacgtgg tgcaggtgga cacctgggtg 600

tccgcctccg gcaagaacgg catgcgccgc gactggctgc tgcgcgacgg caacaccggc 660

gagaccctga cccgcgcctc ctccgtgtgg gtgatgatga acaagctgac ccgccgcctg 720

tccaagatcc ccgaggaggt gcgcgaggag atcggctcct acttcgtgaa ctccgacccc 780

gtggtggagg aggacggccg caaggtgacc aagctggacg acaacaccgc cgacttcgtg 840

cgcaagggcc tgacccccaa gtggaacgac ctggacatca accagcacgt gaacaacgtg 900

aagtacatcg gctggatcct ggagtccgcc ccccagccca tcctggagac ccgcgagctg 960

tccgccgtga ccctggagta ccgccgcgag tgcggccgcg actccgtgct gcgctccctg 1020

accgccgtgt ccggcggcgg cgtgggcgac ctgggccacg ccggcaacgt ggagtgccag 1080

cacgtgctgc gcctggagga cggcgccgag atcgtgcgcg gccgcaccga gtggcgcccc 1140

aagtacatca acaacttctc catcatgggc cagatcccca ccgacgcctc catggactac 1200

aaggaccacg acggcgacta caaggaccac gacatcgact acaaggacga cgacgacaag 1260

tgaatcgat 1269

<210> 148

<211> 406

<212> PRT

<213> Garcinia mangostana

<400> 148

Met Val Ala Thr Ala Ala Thr Ser Ser Phe Phe Pro Leu Thr Ser Pro
1 5 10 15

Ser Gly Asp Ala Lys Ser Gly Asn Pro Gly Lys Gly Ser Val Ser Phe
20 25 30

Gly Ser Met Lys Ser Lys Ser Ala Ala Ser Ser Arg Gly Leu Gln Val
35 40 45

Lys Ala Asn Ala Gln Ala Pro Thr Lys Ile Asn Gly Ser Thr Asp Asp
50 55 60

Ala Gln Leu Pro Ala Pro Arg Thr Phe Ile Asn Gln Leu Pro Asp Trp
65 70 75 80

Ser Met Leu Leu Ala Ala Ile Thr Thr Val Phe Leu Ala Ala Glu Lys
85 90 95

Gln Trp Met Met Leu Asp Trp Lys Pro Arg Arg Pro Asp Met Leu Ile
100 105 110
Asp Thr Phe Gly Leu Gly Arg Ile Val Gln Asp Gly Leu Val Phe Arg
115 120 125
Gln Asn Phe Ser Ile Arg Ser Tyr Glu Ile Gly Ala Asp Arg Thr Ala
130 135 140

Ser Ile Glu Thr Val Met Asn His Leu Gln Glu Thr Ala Leu Asn His
145 150 155 160
Val Lys Thr Ala Gly Leu Leu Gly Asp Gly Phe Gly Ser Thr Pro Glu
165 170 175
Met Ser Lys Arg Asn Leu Ile Trp Val Val Thr Lys Met Gln Val Glu
180 185 190
Val Asp Arg Tyr Pro Thr Trp Gly Asp Val Val Gln Val Asp Thr Trp
195 200 205

Val Ser Ala Ser Gly Lys Asn Gly Met Arg Arg Asp Trp Leu Leu Arg
210 215 220
Asp Gly Asn Thr Gly Glu Thr Leu Thr Arg Ala Ser Ser Val Trp Val
225 230 235 240
Met Met Asn Lys Leu Thr Arg Arg Leu Ser Lys Ile Pro Glu Glu Val
245 250 255
Arg Glu Glu Ile Gly Ser Tyr Phe Val Asn Ser Asp Pro Val Val Glu
260 265 270

Glu Asp Gly Arg Lys Val Thr Lys Leu Asp Asp Asn Thr Ala Asp Phe
275 280 285
Val Arg Lys Gly Leu Thr Pro Lys Trp Asn Asp Leu Asp Ile Asn Gln
290 295 300
His Val Asn Asn Val Lys Tyr Ile Gly Trp Ile Leu Glu Ser Ala Pro
305 310 315 320
Gln Pro Ile Leu Glu Thr Arg Glu Leu Ser Ala Val Thr Leu Glu Tyr
325 330 335

Arg Arg Glu Cys Gly Arg Asp Ser Val Leu Arg Ser Leu Thr Ala Val

340 345 350

Ser Gly Gly Gly Val Gly Asp Leu Gly His Ala Gly Asn Val Glu Cys

355 360 365

Gln His Val Leu Arg Leu Glu Asp Gly Ala Glu Ile Val Arg Gly Arg

370 375 380

Thr Glu Trp Arg Pro Lys Tyr Ile Asn Asn Phe Ser Ile Met Gly Gln

385 390 395 400

Ile Pro Thr Asp Ala Ser

405

<210> 149

<211> 1338

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 149

actagtatgg tggccaccgc cgcctcctcc gccttcttcc ccctgccctc cgccgacacc 60

tcctcccgcc ccggcaagct gggcaacaag ccctcctccc tgtccccccc gaagcccaag 120

tccaccccca acggcggcct gcaggtgaag gccaacgcct ccgccccccc caagatcaac 180

ggctcccccg tgggcctgaa gtccggcggc ctgaaggggc gcgccaccca ggaggacgcc 240

cactccgccc ccccccccgg caccttcatc aaccagctgc ccgactggtc catgctgctg 300

gccgccatca ccaccgtgtt cctggccgcc gagaagcagt ggatgatgct ggactggaag 360

cccaagcgcc ccgacatgct ggtggacccc ttcggcctgg gctccatcgt gcaggacggc 420

ctggtgttcc gccagaactt ctccatccgc tcctacgaga tcggcgccga ccgcaccgcc 480

tccatcgaga ccgtgatgaa ccacctgcag gagaccgccc tgaaccacgt gaagatcgcc 540

ggcctgtcca acgacggctt cggccgcacc cccgagatgt acaagcgcga cctgatctgg 600

gtggtggcca agatgcaggt gatggtgaac cgctacccca cctggggcga caccgtggag 660

gtgaacacct gggtggccaa gtccggcaag aacggcatgc gccgcgactg gctgatctcc 720

gactgcaaca ccggcgagat cctgacccgc gcctcctccg tgtgggtgat gatgaaccag 780

aagacccgcc gcctgtccaa gatccccgac gaggtgcgca acgagatcga gccccacttc 840

gtggactccc cccccgtgat cgaggacgac gaccgcaagc tgcccaagct ggacgagaag 900

accgccgact ccatccgcaa gggcctgacc ccccgctgga acgacctgga cgtgaaccag 960

cacgtgaaca acgtgaagta catcggctgg atcctggagt ccaccccccc cgaggtgctg 1020

gagacccagg agctgtgctc cctgaccctg gagtaccgcc gcgagtgcgg ccgcgagtcc 1080

gtgctggagt ccctgaccgc catggacccc tccggcggcg gctacggctc ccagttccag 1140

cacctgctgc gcctggagga cggcggcgag atcgtgaagg gccgcaccga gtggcgcccc 1200

aagaacggcg tgatcaacgg cgtggtgccc accggcgagt cctcccccgg cgactactcc 1260

atggactaca aggaccacga cggcgactac aaggaccacg acatcgacta caaggacgac 1320

gacgacaagt gaatcgat 1338

<210> 150

<211> 415

<212> PRT

<213> Cuphea hookeriana

<400> 150

Met Val Ala Thr Ala Ala Ser Ser Ala Phe Phe Pro Leu Pro Ser Ala
1 5 10 15

Asp Thr Ser Ser Arg Pro Gly Lys Leu Gly Asn Lys Pro Ser Ser Leu
20 25 30

Ser Pro Leu Lys Pro Lys Ser Thr Pro Asn Gly Gly Leu Gln Val Lys
35 40 45

Ala Asn Ala Ser Ala Pro Pro Lys Ile Asn Gly Ser Pro Val Gly Leu
50 55 60

Lys Ser Gly Gly Leu Lys Thr Gln Glu Asp Ala His Ser Ala Pro Pro
65 70 75 80

Pro Arg Thr Phe Ile Asn Gln Leu Pro Asp Trp Ser Met Leu Leu Ala
85 90 95

Ala Ile Thr Thr Val Phe Leu Ala Ala Glu Lys Gln Trp Met Met Leu
100 105 110

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

Gly Ser Ile Val Gln Asp Gly Leu Val Phe Arg Gln Asn Phe Ser Ile
130 135 140

Arg Ser Tyr Glu Ile Gly Ala Asp Arg Thr Ala Ser Ile Glu Thr Val
145 150 155 160

Met Asn His Leu Gln Glu Thr Ala Leu Asn His Val Lys Ile Ala Gly
165 170 175
Leu Ser Asn Asp Gly Phe Gly Arg Thr Pro Glu Met Tyr Lys Arg Asp
180 185 190
Leu Ile Trp Val Val Ala Lys Met Gln Val Met Val Asn Arg Tyr Pro
195 200 205

Thr Trp Gly Asp Thr Val Glu Val Asn Thr Trp Val Ala Lys Ser Gly
210 215 220
Lys Asn Gly Met Arg Arg Asp Trp Leu Ile Ser Asp Cys Asn Thr Gly
225 230 235 240
Glu Ile Leu Thr Arg Ala Ser Ser Val Trp Val Met Met Asn Gln Lys
245 250 255
Thr Arg Arg Leu Ser Lys Ile Pro Asp Glu Val Arg Asn Glu Ile Glu
260 265 270

Pro His Phe Val Asp Ser Pro Pro Val Ile Glu Asp Asp Asp Arg Lys
275 280 285
Leu Pro Lys Leu Asp Glu Lys Thr Ala Asp Ser Ile Arg Lys Gly Leu
290 295 300
Thr Pro Arg Trp Asn Asp Leu Asp Val Asn Gln His Val Asn Asn Val
305 310 315 320
Lys Tyr Ile Gly Trp Ile Leu Glu Ser Thr Pro Pro Glu Val Leu Glu
325 330 335

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

Asn Gly Val Val Pro Thr Gly Glu Ser Ser Pro Gly Asp Tyr Ser

405 410 415

<210> 151

<211> 1344

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 151

actagtatgg tggcctccat cgtggcctgg gccttcttcc ccacccccctc cttctccccc 60

accgggcgcg ccgcctccgc caaggcctcc aagaccatcg gcgagggctc cgagaacctg 120

aacgtgcgcg gcatcatcgc caagcccacc tcctcctccg ccgccaagca gggcaaggtg 180

atggcccagg ccgtgcccaa gatcaacggc gccaaggtgg gcctgaaggc cgagtcccag 240

aaggccgagg aggacgccgc cccctcctcc gccccccgca ccttctacaa ccagctgccc 300

gactggtccg tgctgctggc cgccgtgacc accatcttcc tggccgccga gaagcagtgg 360

accctgctgg actggaagcc ccgccgcccc gacatgctga ccggcgcctt ctccctgggc 420

aagatcgtgc aggacggcct ggtgttccgc cagaacttct ccatccgctc ctacgagatc 480

ggcgccgacc gcaccgcctc catcgagacc ctgatgaacc acctgcagga gaccgccctg 540

aaccacgtgc gcaacgccgg cctgctgggc gacggcttcg gcgccacccc cgagatgtcc 600

aagcgcaacc tgatctgggt ggtgaccaag atgcaggtgc tgatcgagca ctacccctcc 660

tggggcgacg tggtggaggt ggacacctgg gtgggcgcct ccggcaagaa cggcatgcgc 720

cgcgactggc acgtgcgcga ctaccgcacc ggccagacca tcctgcgcgc cacctccatc 780

tgggtgatga tggacaagca cacccgcaag ctgtccaaga tgcccgagga ggtgcgcgcc 840

gagatcggcc cctacttcat ggagcacgcc gccatcgtgg acgaggactc ccgcaagctg 900

cccaagctgg acgacgacac cgccgactac atcaagtggg gcctgacccc ccgctggtcc 960

gacctggacg tgaaccagca cgtgaacaac gtgaagtaca tcggctggat cctggagtcc 1020

gcccccatct ccatcctgga gaaccacgag ctggcctcca tgaccctgga gtaccgccgc 1080

gagtgcggcc gcgactccgt gctgcagtcc ctgaccgccg tggccaacga ctgcaccggc 1140

ggcctgcccg aggcctccat cgagtgccag cacctgctgc agctggagtg cggcgccgag 1200

atcgtgcgcg gccgcaccca gtggcgcccc cgccgcgcct ccggccccac ctccgccggc 1260

tccgccatgg actacaagga ccacgacggc gactacaagg accacgacat cgactacaag 1320

gacgacgacg acaagtgaat cgat 1344

<210> 152

<211> 417

<212> PRT

<213> Elaeis guiniensis

<400> 152

Met Val Ala Ser Ile Val Ala Trp Ala Phe Phe Pro Thr Pro Ser Phe

1 5 10 15

Ser Pro Thr Ala Ser Ala Lys Ala Ser Lys Thr Ile Gly Glu Gly Ser

20 25 30

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

35 40 45

Ala Ala Lys Gln Gly Lys Val Met Ala Gln Ala Val Pro Lys Ile Asn

50 55 60

Gly Ala Lys Val Gly Leu Lys Ala Glu Ser Gln Lys Ala Glu Glu Asp

65 70 75 80

Ala Ala Pro Ser Ser Ala Pro Arg Thr Phe Tyr Asn Gln Leu Pro Asp

85 90 95

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

100 105 110

Lys Gln Trp Thr Leu Leu Asp Trp Lys Pro Arg Arg Pro Asp Met Leu

115 120 125

Thr Gly Ala Phe Ser Leu Gly Lys Ile Val Gln Asp Gly Leu Val Phe

130 135 140

Arg Gln Asn Phe Ser Ile Arg Ser Tyr Glu Ile Gly Ala Asp Arg Thr

145 150 155 160

Ala Ser Ile Glu Thr Leu Met Asn His Leu Gln Glu Thr Ala Leu Asn

165 170 175

His Val Arg Asn Ala Gly Leu Leu Gly Asp Gly Phe Gly Ala Thr Pro

180 185 190

Glu Met Ser Lys Arg Asn Leu Ile Trp Val Val Thr Lys Met Gln Val

195 200 205

Leu Ile Glu His Tyr Pro Ser Trp Gly Asp Val Val Glu Val Asp Thr

210 215 220

Trp Val Gly Ala Ser Gly Lys Asn Gly Met Arg Arg Asp Trp His Val
225 230 235 240
Arg Asp Tyr Arg Thr Gly Gln Thr Ile Leu Arg Ala Thr Ser Ile Trp
245 250 255
Val Met Met Asp Lys His Thr Arg Lys Leu Ser Lys Met Pro Glu Glu

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

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

385 390 395 400
Thr Gln Trp Arg Pro Arg Arg Ala Ser Gly Pro Thr Ser Ala Gly Ser
405 410 415
Ala

<210> 153
<211> 1191
<212> DNA
<213> Artificial Sequence
<220><223> synthesized construct
<400> 153

actagtatgc tgaagctgtc ctgcaacgtg accaacaacc tgcacacctt ctccttcttc 60

tccgactcct ccctgttcat ccccgtgaac cgccgcacca tcgccgtgtc ctccgggcgc 120

gcctcccagc tgcgcaagcc cgccctggac cccctgcgcg ccgtgatctc cgccgaccag 180

ggctccatct cccccgtgaa ctcctgcacc cccgccgacc gcctgcgcgc cggccgcctg 240

atggaggacg gctactccta caaggagaag ttcatcgtgc gctcctacga ggtgggcatc 300

aacaagaccg ccaccgtgga gaccatcgcc aacctgctgc aggaggtggc ctgcaaccac 360

gtgcagaagt gcggcttctc caccgacggc ttcgccacca ccctgaccat gcgcaagctg 420

cacctgatct gggtgaccgc ccgcatgcac atcgagatct acaagtaccc cgcctggtcc 480

gacgtggtgg agatcgagac ctggtgccag tccgagggcc gcatcggcac ccgccgcgac 540

tggatcctgc gcgactccgc caccaacgag gtgatcggcc gcgccacctc caagtgggtg 600

atgatgaacc aggacacccg ccgcctgcag cgcgtgaccg acgaggtgcg cgacgagtac 660

ctggtgttct gcccccgcga gccccgcctg gccttccccg aggagaacaa ctcctccctg 720

aagaagatcc ccaagctgga ggaccccgcc cagtactcca tgctggagct gaagccccgc 780

cgcgccgacc tggacatgaa ccagcacgtg aacaacgtga cctacatcgg ctgggtgctg 840

gagtccatcc cccaggagat catcgacacc cacgagctgc aggtgatcac cctggactac 900

cgccgcgagt gccagcagga cgacatcgtg gactccctga ccacctccga gatccccgac 960

gaccccatct ccaagttcac cggcaccaac ggctccgcca tgtcctccat ccagggccac 1020

aacgagtccc agttcctgca catgctgcgc ctgtccgaga acggccagga gatcaaccgc 1080

ggccgcaccc agtggcgcaa gaagtcctcc cgcatggact acaaggacca cgacggcgac 1140

tacaaggacc acgacatcga ctacaaggac gacgacgaca agtgaatcga t 1191

<210> 154

<211> 366

<212> PRT

<213> Brassica napus

<400> 154

Met Leu Lys Leu Ser Cys Asn Val Thr Asn Asn Leu His Thr Phe Ser
1 5 10 15

Phe Phe Ser Asp Ser Ser Leu Phe Ile Pro Val Asn Arg Arg Thr Ile
20 25 30

Ala Val Ser Ser Ser Gln Leu Arg Lys Pro Ala Leu Asp Pro Leu Arg
35 40 45

Ala Val Ile Ser Ala Asp Gln Gly Ser Ile Ser Pro Val Asn Ser Cys
50 55 60
Thr Pro Ala Asp Arg Leu Arg Ala Gly Arg Leu Met Glu Asp Gly Tyr
65 70 75 80
Ser Tyr Lys Glu Lys Phe Ile Val Arg Ser Tyr Glu Val Gly Ile Asn
85 90 95

Lys Thr Ala Thr Val Glu Thr Ile Ala Asn Leu Leu Gln Glu Val Ala
100 105 110
Cys Asn His Val Gln Lys Cys Gly Phe Ser Thr Asp Gly Phe Ala Thr
115 120 125
Thr Leu Thr Met Arg Lys Leu His Leu Ile Trp Val Thr Ala Arg Met
130 135 140
His Ile Glu Ile Tyr Lys Tyr Pro Ala Trp Ser Asp Val Val Glu Ile
145 150 155 160

Glu Thr Trp Cys Gln Ser Glu Gly Arg Ile Gly Thr Arg Arg Asp Trp
165 170 175
Ile Leu Arg Asp Ser Ala Thr Asn Glu Val Ile Gly Arg Ala Thr Ser
180 185 190
Lys Trp Val Met Met Asn Gln Asp Thr Arg Arg Leu Gln Arg Val Thr
195 200 205
Asp Glu Val Arg Asp Glu Tyr Leu Val Phe Cys Pro Arg Glu Pro Arg
210 215 220

Leu Ala Phe Pro Glu Glu Asn Asn Ser Ser Leu Lys Lys Ile Pro Lys
225 230 235 240
Leu Glu Asp Pro Ala Gln Tyr Ser Met Leu Glu Leu Lys Pro Arg Arg
245 250 255
Ala Asp Leu Asp Met Asn Gln His Val Asn Asn Val Thr Tyr Ile Gly
260 265 270
Trp Val Leu Glu Ser Ile Pro Gln Glu Ile Ile Asp Thr His Glu Leu
275 280 285

Gln Val Ile Thr Leu Asp Tyr Arg Arg Glu Cys Gln Gln Asp Asp Ile

290 295 300

Val Asp Ser Leu Thr Thr Ser Glu Ile Pro Asp Asp Pro Ile Ser Lys

305 310 315 320

Phe Thr Gly Thr Asn Gly Ser Ala Met Ser Ser Ile Gln Gly His Asn

325 330 335

Glu Ser Gln Phe Leu His Met Leu Arg Leu Ser Glu Asn Gly Gln Glu

340 345 350

Ile Asn Arg Gly Arg Thr Gln Trp Arg Lys Lys Ser Ser Arg

355 360 365

<210> 155

<211> 1038

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 155

actagtatgg ccaccgcatc cactttctcg gcgttcaatg cccgctgcgg cgacctgcgt 60

cgctcggcgg gctccgggcc ccggcgccca gcgaggcccc tccccgtgcg cgggcgcgcc 120

cgctcctacg aggtgggcat caacaagacc gccaccgtgg agaccatcgc caacctgctg 180

caggaggtgg gctgcaacca cgcccagtcc gtgggcttct ccaccgacgg cttcgccacc 240

accacctcca tgcgcaagat gcacctgatc tgggtgaccg cccgcatgca catcgagatc 300

tacaagtacc ccgcctggtc cgacgtggtg gaggtggaga cctggtgcca gtccgagggc 360

cgcatcggca cccgccgcga ctggatcctg accgactacg ccaccggcca gatcatcggc 420

cgcgccacct ccaagtgggt gatgatgaac caggacaccc gccgcctgca gaaggtgacc 480

gacgacgtgc gcgaggagta cctggtgttc tgcccccgcg agctgcgcct ggccttcccc 540

gaggagaaca accgctcctc caagaagatc tccaagctgg aggaccccgc ccagtactcc 600

aagctgggcc tggtgccccg ccgcgccgac ctggacatga accagcacgt gaacaacgtg 660

acctacatcg gctgggtgct ggagtccatc ccccaggaga tcatcgacac ccacgagctg 720

cagaccatca ccctggacta ccgccgcgag tgccagcacg acgacatcgt ggactccctg 780

acctccgtgg agccctccga gaacctggag gccgtgtccg agctgcgcgg caccaacggc 840

tccgccacca ccaccgccgg cgacgaggac tgccgcaact tcctgcacct gctgcgcctg 900

tccggcgacg gcctggagat caaccgcggc cgcaccgagt ggcgcaagaa gtccgcccgc 960

atggactaca aggaccacga cggcgactac aaggaccacg acatcgacta caaggacgac 1020

gacgacaagt gaatcgat 1038

<210> 156

<211> 371

<212> PRT

<213> Ricinus communis

<400> 156

Met Leu Lys Val Pro Cys Cys Asn Ala Thr Asp Pro Ile Gln Ser Leu
1 5 10 15

Ser Ser Gln Cys Arg Phe Leu Thr His Phe Asn Asn Arg Pro Tyr Phe
20 25 30

Thr Arg Arg Pro Ser Ile Pro Thr Phe Phe Ser Ser Lys Asn Ser Ser
35 40 45

Ala Ser Leu Gln Ala Val Val Ser Asp Ile Ser Ser Val Glu Ser Ala
50 55 60

Ala Cys Asp Ser Leu Ala Asn Arg Leu Arg Leu Gly Lys Leu Thr Glu
65 70 75 80

Asp Gly Phe Ser Tyr Lys Glu Lys Phe Ile Val Arg Ser Tyr Glu Val
85 90 95

Gly Ile Asn Lys Thr Ala Thr Val Glu Thr Ile Ala Asn Leu Leu Gln
100 105 110

Glu Val Gly Cys Asn His Ala Gln Ser Val Gly Phe Ser Thr Asp Gly
115 120 125

Phe Ala Thr Thr Thr Ser Met Arg Lys Met His Leu Ile Trp Val Thr
130 135 140

Ala Arg Met His Ile Glu Ile Tyr Lys Tyr Pro Ala Trp Ser Asp Val
145 150 155 160

Val Glu Val Glu Thr Trp Cys Gln Ser Glu Gly Arg Ile Gly Thr Arg
165 170 175

Arg Asp Trp Ile Leu Thr Asp Tyr Ala Thr Gly Gln Ile Ile Gly Arg
180 185 190

Ala Thr Ser Lys Trp Val Met Met Asn Gln Asp Thr Arg Arg Leu Gln

195 200 205

Lys Val Thr Asp Asp Val Arg Glu Glu Tyr Leu Val Phe Cys Pro Arg

210 215 220

Glu Leu Arg Leu Ala Phe Pro Glu Glu Asn Asn Arg Ser Ser Lys Lys

225 230 235 240

Ile Ser Lys Leu Glu Asp Pro Ala Gln Tyr Ser Lys Leu Gly Leu Val

245 250 255

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

260 265 270

Tyr Ile Gly Trp Val Leu Glu Ser Ile Pro Gln Glu Ile Ile Asp Thr

275 280 285

His Glu Leu Gln Thr Ile Thr Leu Asp Tyr Arg Arg Glu Cys Gln His

290 295 300

Asp Asp Ile Val Asp Ser Leu Thr Ser Val Glu Pro Ser Glu Asn Leu

305 310 315 320

Glu Ala Val Ser Glu Leu Arg Gly Thr Asn Gly Ser Ala Thr Thr Thr

325 330 335

Ala Gly Asp Glu Asp Cys Arg Asn Phe Leu His Leu Leu Arg Leu Ser

340 345 350

Gly Asp Gly Leu Glu Ile Asn Arg Gly Arg Thr Glu Trp Arg Lys Lys

355 360 365

Ser Ala Arg

370

<210> 157

<211> 1248

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 157

atggccaccg catccacttt ctcggcgttc aatgcccgct gcggcgacct gcgtcgctcg 60

gcgggctccg ggccccggcg cccagcgagg cccctccccg tgcgcgggcg cgccgccaac 120

gcccacaccg tgcccaagat caacggcaac aaggccggcc tgctgacccc catggagagc 180

accaaggacg aggacatcgt cgcggccccc accgtggcgc ccaagcgcac cttcatcaac 240

cagctgcccg actggtcgat gctgctggcc gcgatcacca ccatcttcct ggcggccgag 300

aagcagtgga ccaacctgga ctggaagccc cgccgccccg acatgctggt cgacttcgac 360

cccttctccc tgggccgctt cgtgcaggac ggcctgatct tccgccagaa cttcagcatc 420

cgctcgtacg agatcggcgc ggaccgcacc gcctccatcg agaccctgat gaaccacctg 480

caggagaccg cgctgaacca cgtccgctgc atcggcctgc tggacgacgg cttcggcagc 540

acccccgaga tgacccgccg cgacctgatc tgggtggtca cccgcatgca ggtcctggtg 600

gaccgctacc cctcgtgggg cgacgtgatc gaggtcgact cctgggtgac ccccagcggc 660

aagaacggca tgaagcgcga gtggttcctg cgcgactgca agaccggcga gatcctgacc 720

cgcgccacct cggtctgggt gatgatgaac aagcgcaccc gccgcctgtc caagatcccc 780

gaggaggtcc gcgtggagat cgagccctac ttcgtcgagc acggcgtgct ggacgaggac 840

tcgcgcaagc tgcccaagct gaacgacaac accgccaact acatccgccg cggcctggcg 900

ccccgctggt ccgacctgga cgtcaaccag cacgtgaaca acgtcaagta catcggctgg 960

atcctggaga gcgtgcccag cagcctgctg gagtcgcacg agctgtacgg catgaccctg 1020

gagtaccgca aggagtgcgg caaggacggc ctgctgcagt ccctgaccgc cgtcgccagc 1080

gactacggcg gcggctcgct ggaggccggc gtggagtgcg accacctgct gcgcctggag 1140

gacggctccg agatcatgcg cggcaagacc gagtggcgcc ccaagcgcgc cgcgaacacc 1200

acctacttcg gcagcgtcga cgacatcccc cccgccaaca acgcgtga 1248

<210> 158

<211> 415

<212> PRT

<213> Artificial Sequence

<220><223> synthesized construct

<400> 158

Met Ala Thr Ala Ser Thr Phe Ser Ala Phe Asn Ala Arg Cys Gly Asp

1 5 10 15

Leu Arg Arg Ser Ala Gly Ser Gly Pro Arg Arg Pro Ala Arg Pro Leu

20 25 30

Pro Val Arg Gly Arg Ala Ala Asn Ala His Thr Val Pro Lys Ile Asn

35 40 45

Gly Asn Lys Ala Gly Leu Leu Thr Pro Met Glu Ser Thr Lys Asp Glu

50 55 60

Asp Ile Val Ala Ala Pro Thr Val Ala Pro Lys Arg Thr Phe Ile Asn

65 70 75 80

Gln Leu Pro Asp Trp Ser Met Leu Leu Ala Ala Ile Thr Thr Ile Phe

85 90 95

Leu Ala Ala Glu Lys Gln Trp Thr Asn Leu Asp Trp Lys Pro Arg Arg

100 105 110

Pro Asp Met Leu Val Asp Phe Asp Pro Phe Ser Leu Gly Arg Phe Val

115 120 125

Gln Asp Gly Leu Ile Phe Arg Gln Asn Phe Ser Ile Arg Ser Tyr Glu

130 135 140

Ile Gly Ala Asp Arg Thr Ala Ser Ile Glu Thr Leu Met Asn His Leu

145 150 155 160

Gln Glu Thr Ala Leu Asn His Val Arg Cys Ile Gly Leu Leu Asp Asp

165 170 175

Gly Phe Gly Ser Thr Pro Glu Met Thr Arg Arg Asp Leu Ile Trp Val

180 185 190

Val Thr Arg Met Gln Val Leu Val Asp Arg Tyr Pro Ser Trp Gly Asp

195 200 205

Val Ile Glu Val Asp Ser Trp Val Thr Pro Ser Gly Lys Asn Gly Met

210 215 220

Lys Arg Glu Trp Phe Leu Arg Asp Cys Lys Thr Gly Glu Ile Leu Thr

225 230 235 240

Arg Ala Thr Ser Val Trp Val Met Met Asn Lys Arg Thr Arg Arg Leu

245 250 255

Ser Lys Ile Pro Glu Glu Val Arg Val Glu Ile Glu Pro Tyr Phe Val

260 265 270

Glu His Gly Val Leu Asp Glu Asp Ser Arg Lys Leu Pro Lys Leu Asn

275 280 285

Asp Asn Thr Ala Asn Tyr Ile Arg Arg Gly Leu Ala Pro Arg Trp Ser

290 295 300

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

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

<210> 159

<211> 2333

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 159

ctttcttgcg ctatgacact tccagcaaaa ggtagggcgg gctgcgagac ggcttcccgg 60
cgctgcatgc aacaccgatg atgcttcgac cccccgaagc tccttcgggg ctgcatgggc 120
gctccgatgc cgctccaggg cgagcgctgt ttaaatagcc aggcccccga ttgcaaagac 180
attatagcga gctaccaaag ccatattcaa acacctagat cactaccact tctacacagg 240
ccactcgagc ttgtgatcgc actccgctaa gggggcgcct cttcctcttc gtttcagtca 300
caacccgcaa acggcgcgcc atgctgctgc aggccttcct gttcctgctg gccggcttcg 360
ccgccaagat cagcgcctcc atgacgaacg agacgtccga ccgccccctg gtgcacttca 420

cccccaacaa gggctggatg aacgacccca acggcctgtg gtacgacgag aaggacgcca 480
agtggcacct gtacttccag tacaacccga acgacaccgt ctgggggacg cccttgttct 540
ggggccacgc cacgtccgac gacctgacca actgggagga ccagcccatc gccatcgccc 600
cgaagcgcaa cgactccggc gccttctccg gctccatggt ggtggactac aacaacacct 660
ccggcttctt caacgacacc atcgacccgc gccagcgctg cgtggccatc tggacctaca 720

acaccccgga gtccgaggag cagtacatct cctacagcct ggacggcggc tacaccttca 780

ccgagtacca gaagaacccc gtgctggccg ccaactccac ccagttccgc gacccgaagg 840

tcttctggta cgagccctcc cagaagtgga tcatgaccgc ggccaagtcc caggactaca 900

agatcgagat ctactcctcc gacgacctga agtcctggaa gctggagtcc gcgttcgcca 960

acgagggctt cctcggctac cagtacgagt gccccggcct gatcgaggtc cccaccgagc 1020

aggaccccag caagtcctac tgggtgatgt tcatctccat caaccccggc gccccggccg 1080

gcggctcctt caaccagtac ttcgtcggca gcttcaacgg cacccacttc gaggccttcg 1140

acaaccagtc ccgcgtggtg gacttcggca aggactacta cgccctgcag accttcttca 1200

acaccgaccc gacctacggg agcgccctgg gcatcgcgtg ggcctccaac tgggagtact 1260

ccgccttcgt gcccaccaac ccctggcgct cctccatgtc cctcgtgcgc aagttctccc 1320

tcaacaccga gtaccaggcc aacccggaga cggagctgat caacctgaag gccgagccga 1380

tcctgaacat cagcaacgcc ggcccctgga gccggttcgc caccaacacc acgttgacga 1440

aggccaacag ctacaacgtc gacctgtcca acagcaccgg caccctggag ttcgagctgg 1500

tgtacgccgt caacaccacc cagacgatct ccaagtccgt gttcgcggac ctctccctct 1560

ggttcaaggg cctggaggac cccgaggagt acctccgcat gggcttcgag gtgtccgcgt 1620

cctccttctt cctggaccgc gggaacagca aggtgaagtt cgtgaaggag aacccctact 1680

tcaccaaccg catgagcgtg aacaaccagc ccttcaagag cgagaacgac ctgtcctact 1740

acaaggtgta cggcttgctg gaccagaaca tcctggagct gtacttcaac gacggcgacg 1800

tcgtgtccac caacacctac ttcatgacca ccgggaacgc cctgggctcc gtgaacatga 1860

cgacgggggt ggacaacctg ttctacatcg acaagttcca ggtgcgcgag gtcaagtgac 1920

aattggcagc agcagctcgg atagtatcga cacactctgg acgctggtcg tgtgatggac 1980

tgttgccgcc acacttgctg ccttgacctg tgaatatccc tgccgctttt atcaaacagc 2040

ctcagtgtgt ttgatcttgt gtgtacgcgc ttttgcgagt tgctagctgc ttgtgctatt 2100

tgcgaatacc acccccagca tcccccttccc tcgtttcata tcgcttgcat cccaaccgca 2160

acttatctac gctgtcctgc tatccctcag cgctgctcct gctcctgctc actgcccctc 2220

gcacagcctt ggtttgggct ccgcctgtat tctcctggta ctgcaacctg taaaccagca 2280

ctgcaatgct gatgcacggg aagtagtggg atgggaacac aaatggagga tcc 2333

<210> 160

<211> 1233

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 160

actagtatgg ccaccgcatc cactttctcg gcgttcaatg cccgctgcgg cgacctgcgt 60

cgctcggcgg gctccgggcc ccggcgccca gcgaggcccc tccccgtgcg cgggcgcgcc 120

acccaggagg acgcccactc cgccccccccc ccccgcacct tcatcaacca gctgcccgac 180

tggtccatgc tgctggccgc catcaccacc gtgttcctgg ccgccgagaa gcagtggatg 240

atgctggact ggaagcccaa gcgccccgac atgctggtgg acccccttcgg cctgggctcc 300

atcgtgcagg acggcctggt gttccgccag aacttctcca tccgctccta cgagatcggc 360

gccgaccgca ccgcctccat cgagaccgtg atgaaccacc tgcaggagac cgccctgaac 420

cacgtgaaga tcgccggcct gtccaacgac ggcttcggcc gcaccccccga gatgtacaag 480

cgcgacctga tctgggtggt ggccaagatg caggtgatgg tgaaccgcta ccccacctgg 540

ggcgacaccg tggaggtgaa cacctgggtg gccaagtccg gcaagaacgg catgcgccgc 600

gactggctga tctccgactg caacaccggc gagatcctga cccgcgcctc ctccgtgtgg 660

gtgatgatga accagaagac ccgccgcctg tccaagatcc ccgacgaggt gcgcaacgag 720

atcgagcccc acttcgtgga ctccccccccc gtgatcgagg acgacgaccg caagctgccc 780

aagctggacg agaagaccgc cgactccatc cgcaagggcc tgacccccccg ctggaacgac 840

ctggacgtga accagcacgt gaacaacgtg aagtacatcg gctggatcct ggagtccacc 900

ccccccgagg tgctggagac ccaggagctg tgctccctga ccctggagta ccgccgcgag 960

tgcggccgcg agtccgtgct ggagtccctg accgccatgg acccctccgg cggcggctac 1020

ggctcccagt tccagcacct gctgcgcctg gaggacggcg gcgagatcgt gaagggccgc 1080

accgagtggc gccccaagaa cggcgtgatc aacggcgtgg tgcccaccgg cgagtcctcc 1140

cccggcgact actccatgga ctacaaggac cacgacggcg actacaagga ccacgacatc 1200

gactacaagg acgacgacga caagtgaatc gat 1233

<210> 161

<211> 1344

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 161

actagtatgg tggcctccat cgccgcctcc gccttcttcc ccacccccctc cttctccccc 60

accgggcgcg ccgcctccgc caaggcctcc aagaccatcg gcgagggctc cgagaacctg 120

gacgcccgcg gcatcatcgc caagcccacc tcctcctccg ccgccatgca gggcaaggtg 180

atggcccagg ccgtgtccaa gatcaacggc gccaaggtgg gcctgaaggc cgagtcccag 240

aaggccgagg aggacgccgc cccctcctcc gccccccgca ccttctacaa ccagctgccc 300

gactggtccg tgctgctggc cgccgtgacc accatcttcc tggccgccga gaagcagtgg 360

accctgctgg actggaagcc ccgccgcccc gacatgctga ccgacgcctt ctccctgggc 420

aagatcgtgc aggacggcct ggtgttcaag cagaacttct ccatccgctc ctacgagatc 480

ggcgccgacc gcaccgcctc catcgagacc ctgatgaacc acctgcagga gaccgccctg 540

aaccacgtgc gctccgccgg cctgatgggc gacggcttcg gcgccacccc cgagatgtcc 600

aagcgcaacc tgatctgggt ggtgaccaag gtgcgcgtgc tgatcgagca ctacccctcc 660

tggggcgacg tggtggaggt ggacacctgg gtgggccccg ccggcaagaa cggcatgcgc 720

cgcgactggc acgtgcgcga ccaccgcacc ggccagacca tcctgcgcgc cacccgcgtg 780

tgggtgatga tgaacaagaa cacccgcaag ctgtccaagg tgcccgagga ggtgcgcgcc 840

gagatcggcc cctacttcgt ggagcgcgcc gccatcgtgg acgaggactc ccgcaagctg 900

cccaagctgg acgaggacac caccgactac gtgaagaagg gcctgacccc ccgctggtcc 960

gacctggacg tgaaccagca cgtgaacaac gtgaagtaca tcggctggat cctggagtcc 1020

gcccccatct ccatcctgga gaaccacgag ctggcctcca tgaccctgga gtaccgccgc 1080

gagtgcggcc gcgactccgt gctgcagtcc ctgaccgccg tggccaacga ctgcaccggc 1140

ggcctgcccg aggcctccat cgagtgccag cacctgctgc agctggagtg cggcgccgag 1200

atcgtgcgcg gccgcaccca gtggcgcccc cgccgcgcct ccggccccac ctccgccggc 1260

tccgccatgg actacaagga ccacgacggc gactacaagg accacgacat cgactacaag 1320

gacgacgacg acaagtgaat cgat 1344

<210> 162

<211> 417

<212> PRT

<213> Elaeis guiniensis

<400> 162

Met Val Ala Ser Ile Ala Ala Ser Ala Phe Phe Pro Thr Pro Ser Phe

1 5 10 15

Ser Pro Thr Ala Ser Ala Lys Ala Ser Lys Thr Ile Gly Glu Gly Ser

20 25 30

Glu Asn Leu Asp Ala Arg Gly Ile Ile Ala Lys Pro Thr Ser Ser Ser

35 40 45

Ala Ala Met Gln Gly Lys Val Met Ala Gln Ala Val Ser Lys Ile Asn
50 55 60
Gly Ala Lys Val Gly Leu Lys Ala Glu Ser Gln Lys Ala Glu Glu Asp
65 70 75 80
Ala Ala Pro Ser Ser Ala Pro Arg Thr Phe Tyr Asn Gln Leu Pro Asp
85 90 95
Trp Ser Val Leu Leu Ala Ala Val Thr Thr Ile Phe Leu Ala Ala Glu
100 105 110

Lys Gln Trp Thr Leu Leu Asp Trp Lys Pro Arg Arg Pro Asp Met Leu
115 120 125
Thr Asp Ala Phe Ser Leu Gly Lys Ile Val Gln Asp Gly Leu Val Phe
130 135 140
Lys Gln Asn Phe Ser Ile Arg Ser Tyr Glu Ile Gly Ala Asp Arg Thr
145 150 155 160
Ala Ser Ile Glu Thr Leu Met Asn His Leu Gln Glu Thr Ala Leu Asn
165 170 175

His Val Arg Ser Ala Gly Leu Met Gly Asp Gly Phe Gly Ala Thr Pro
180 185 190
Glu Met Ser Lys Arg Asn Leu Ile Trp Val Val Thr Lys Val Arg Val
195 200 205
Leu Ile Glu His Tyr Pro Ser Trp Gly Asp Val Val Glu Val Asp Thr
210 215 220
Trp Val Gly Pro Ala Gly Lys Asn Gly Met Arg Arg Asp Trp His Val
225 230 235 240

Arg Asp His Arg Thr Gly Gln Thr Ile Leu Arg Ala Thr Arg Val Trp
245 250 255
Val Met Met Asn Lys Asn Thr Arg Lys Leu Ser Lys Val Pro Glu Glu
260 265 270
Val Arg Ala Glu Ile Gly Pro Tyr Phe Val Glu Arg Ala Ala Ile Val
275 280 285

Asp Glu Asp Ser Arg Lys Leu Pro Lys Leu Asp Glu Asp Thr Thr Asp
290 295 300

Tyr Val Lys Lys Gly Leu Thr Pro Arg Trp Ser Asp Leu Asp Val Asn
305 310 315 320
Gln His Val Asn Asn Val Lys Tyr Ile Gly Trp Ile Leu Glu Ser Ala
325 330 335
Pro Ile Ser Ile Leu Glu Asn His Glu Leu Ala Ser Met Thr Leu Glu
340 345 350
Tyr Arg Arg Glu Cys Gly Arg Asp Ser Val Leu Gln Ser Leu Thr Ala
355 360 365

Val Ala Asn Asp Cys Thr Gly Gly Leu Pro Glu Ala Ser Ile Glu Cys
370 375 380
Gln His Leu Leu Gln Leu Glu Cys Gly Ala Glu Ile Val Arg Gly Arg
385 390 395 400
Thr Gln Trp Arg Pro Arg Arg Ala Ser Gly Pro Thr Ser Ala Gly Ser
405 410 415
Ala

<210> 163
<211> 595
<212> DNA
<213> Artificial Sequence
<220><223> synthesized construct

<400> 163
actagtatgg ctatcaagac gaacaggcag cctgtggaga agcctccgtt cacgatcggg 60
acgctgcgca aggccatccc cgcgcactgt ttcgagcgct cggcgcttcg tagcagcatg 120
tacctggcct ttgacatcgc ggtcatgtcc ctgctctacg tcgcgtcgac gtacatcgac 180
cctgcaccgg tgcctacgtg ggtcaagtac ggcatcatgt ggccgctcta ctggttcttc 240
caggtgtgtt tgagggtttt ggttgcccgt attgaggtcc tggtggcgcg catggaggag 300
aaggcgcctg tcccgctgac ccccccggct accctcccgg caccttccag ggcgcgtacg 360
ggaagaacca gtagagcggc cacatgatgc cgtacttgac ccacgtaggc accggtgcag 420

ggtcgatgta cgtcgacgcg acgtagagca gggacatgac cgcgatgtca aaggccaggt 480

acatgctgct acgaagcgcc gagcgctcga aacagtgcgc ggggatggcc ttgcgcagcg 540

tcccgatcgt gaacggaggc ttctccacag gctgcctgtt cgtcttgata gccat 595

<210> 164

<211> 6101

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 164

gctcttcgcc gccgccactc ctgctcgagc gcgcccgcgc gtgcgccgcc agcgccttgg 60

ccttttcgcc gcgctcgtgc gcgtcgctga tgtccatcac caggtccatg aggtctgcct 120

tgcgccggct gagccactgc ttcgtccggg cggccaagag gagcatgagg gaggactcct 180

ggtccagggt cctgacgtgg tcgcggctct gggagcgggc cagcatcatc tggctctgcc 240

gcaccgaggc cgcctccaac tggtcctcca gcagccgcag tcgccgccga ccctggcaga 300

ggaagacagg tgaggggggt atgaattgta cagaacaacc acgagccttg tctaggcaga 360

atccctacca gtcatggctt tacctggatg acggcctgcg aacagctgtc cagcgaccct 420

cgctgccgcc gcttctcccg cacgcttctt tccagcaccg tgatggcgcg agccagcgcc 480

gcacgctggc gctgcgcttc gccgatctga ggacagtcgg ggaactctga tcagtctaaa 540

ccccccttgcg cgttagtgtt gccatccttt gcagaccggt gagagccgac ttgttgtgcg 600

ccacccccca caccacctcc tcccagacca attctgtcac ctttttggcg aaggcatcgg 660

cctcggcctg cagagaggac agcagtgccc agccgctggg ggttggcgga tgcacgctca 720

ggtacccttt cttgcgctat gacacttcca gcaaaaggta gggcgggctg cgagacggct 780

tcccggcgct gcatgcaaca ccgatgatgc ttcgaccccc cgaagctcct tcggggctgc 840

atgggcgctc cgatgccgct ccagggcgag cgctgtttaa atagccaggc ccccgattgc 900

aaagacatta tagcgagcta ccaaagccat attcaaacac ctagatcact accacttcta 960

cacaggccac tcgagcttgt gatcgcactc cgctaagggg gcgcctcttc ctcttcgttt 1020

cagtcacaac ccgcaaacgg cgcgccatgc tgctgcaggc cttcctgttc ctgctggccg 1080

gcttcgccgc caagatcagc gcctccatga cgaacgagac gtccgaccgc cccctggtgc 1140

acttcacccc caacaagggc tggatgaacg accccaacgg cctgtggtac gacgagaagg 1200

acgccaagtg gcacctgtac ttccagtaca acccgaacga caccgtctgg gggacgccct 1260

tgttctgggg ccacgccacg tccgacgacc tgaccaactg ggaggaccag cccatcgcca 1320

tcgccccgaa gcgcaacgac tccggcgcct tctccggctc catggtggtg gactacaaca 1380

acacctccgg cttcttcaac gacaccatcg acccgcgcca gcgctgcgtg gccatctgga 1440

cctacaacac cccggagtcc gaggagcagt acatctccta cagcctggac ggcggctaca 1500
ccttcaccga gtaccagaag aaccccgtgc tggccgccaa ctccacccag ttccgcgacc 1560
cgaaggtctt ctggtacgag ccctcccaga agtggatcat gaccgcggcc aagtcccagg 1620
actacaagat cgagatctac tcctccgacg acctgaagtc ctggaagctg gagtccgcgt 1680
tcgccaacga gggcttcctc ggctaccagt acgagtgccc cggcctgatc gaggtcccca 1740
ccgagcagga ccccagcaag tcctactggg tgatgttcat ctccatcaac cccggcgccc 1800
cggccggcgg ctccttcaac cagtacttcg tcggcagctt caacggcacc cacttcgagg 1860

ccttcgacaa ccagtcccgc gtggtggact tcggcaagga ctactacgcc ctgcagacct 1920
tcttcaacac cgacccgacc tacgggagcg ccctgggcat cgcgtgggcc tccaactggg 1980
agtactccgc cttcgtgccc accaacccct ggcgctcctc catgtccctc gtgcgcaagt 2040
tctccctcaa caccgagtac caggccaacc cggagacgga gctgatcaac ctgaaggccg 2100
agccgatcct gaacatcagc aacgccggcc cctggagccg gttcgccacc aacaccacgt 2160
tgacgaaggc caacagctac aacgtcgacc tgtccaacag caccggcacc ctggagttcg 2220
agctggtgta cgccgtcaac accacccaga cgatctccaa gtccgtgttc gcggacctct 2280

ccctctggtt caagggcctg gaggaccccg aggagtacct ccgcatgggc ttcgaggtgt 2340
ccgcgtcctc cttcttcctg gaccgcggga acagcaaggt gaagttcgtg aaggagaacc 2400
cctacttcac caaccgcatg agcgtgaaca accagccctt caagagcgag aacgacctgt 2460
cctactacaa ggtgtacggc ttgctggacc agaacatcct ggagctgtac ttcaacgacg 2520
gcgacgtcgt gtccaccaac acctacttca tgaccaccgg gaacgccctg ggctccgtga 2580
acatgacgac gggggtggac aacctgttct acatcgacaa gttccaggtg cgcgaggtca 2640
agtgacaatt ggcagcagca gctcggatag tatcgacaca ctctggacgc tggtcgtgtg 2700

atggactgtt gccgccacac ttgctgcctt gacctgtgaa tatccctgcc gcttttatca 2760
aacagcctca gtgtgtttga tcttgtgtgt acgcgctttt gcgagttgct agctgcttgt 2820
gctatttgcg aataccaccc ccagcatccc cttccctcgt ttcatatcgc ttgcatccca 2880
accgcaactt atctacgctg tcctgctatc cctcagcgct gctcctgctc ctgctcactg 2940
cccctcgcac agccttggtt tgggctccgc ctgtattctc ctggtactgc aacctgtaaa 3000
ccagcactgc aatgctgatg cacgggaagt agtgggatgg gaacacaaat ggaggatccc 3060
gcgtctcgaa cagagcgcgc agaggaacgc tgaaggtctc gcctctgtcg cacctcagcg 3120

cggcatacac cacaataacc acctgacgaa tgcgcttggt tcttcgtcca ttagcgaagc 3180

gtccggttca cacacgtgcc acgttggcga ggtggcaggt gacaatgatc ggtggagctg 3240

atggtcgaaa cgttcacagc ctagggatat cgaattcggc cgacaggacg cgcgtcaaag 3300

gtgctggtcg tgtatgccct ggccggcagg tcgttgctgc tgctggttag tgattccgca 3360

accctgattt tggcgtctta ttttggcgtg gcaaacgctg gcgcccgcga gccgggccgg 3420

cggcgatgcg gtgccccacg gctgccggaa tccaagggag gcaagagcgc ccgggtcagt 3480

tgaagggctt tacgcgcaag gtacagccgc tcctgcaagg ctgcgtggtg gaattggacg 3540

tgcaggtcct gctgaagttc ctccaccgcc tcaccagcgg acaaagcacc ggtgtatcag 3600

gtccgtgtca tccactctaa agagctcgac tacgacctac tgatggccct agattcttca 3660

tcaaaaacgc ctgagacact tgcccaggat tgaaactccc tgaagggacc accaggggcc 3720

ctgagttgtt ccttcccccc gtggcgagct gccagccagg ctgtacctgt gatcgaggct 3780

ggcgggaaaa taggcttcgt gtgctcaggt catgggaggt gcaggacagc tcatgaaacg 3840

ccaacaatcg cacaattcat gtcaagctaa tcagctattt cctcttcacg agctgtaatt 3900

gtcccaaaat tctggtctac cgggggtgat ccttcgtgta cgggcccttc cctcaaccct 3960

aggtatgcgc gcatgcggtc gccgcgcaac tcgcgcgagg gccgagggtt tgggacgggc 4020

cgtcccgaaa tgcagttgca cccggatgcg tggcaccttt tttgcgataa tttatgcaat 4080

ggactgctct gcaaaattct ggctctgtcg ccaaccctag gatcagcggc gtaggatttc 4140

gtaatcattc gtcctgatgg ggagctaccg actaccctaa tatcagcccg actgcctgac 4200

gccagcgtcc acttttgtgc acacattcca ttcgtgccca agacatttca ttgtggtgcg 4260

aagcgtcccc agttacgctc acctgtttcc cgacctcctt actgttctgt cgacagagcg 4320

ggcccacagg ccggtcgcag ccactagtat ggctatcaag acgaacaggc agcctgtgga 4380

gaagcctccg ttcacgatcg ggacgctgcg caaggccatc cccgcgcact gtttcgagcg 4440

ctcggcgctt cgtagcagca tgtacctggc ctttgacatc gcggtcatgt ccctgctcta 4500

cgtcgcgtcg acgtacatcg accctgcacc ggtgcctacg tgggtcaagt acggcatcat 4560

gtggccgctc tactggttct tccaggtgtg tttgagggtt ttggttgccc gtattgaggt 4620

cctggtggcg cgcatggagg agaaggcgcc tgtcccgctg acccccccgg ctaccctccc 4680

ggcaccttcc agggcgcgta cgggaagaac cagtagagcg gccacatgat gccgtacttg 4740

acccacgtag gcaccggtgc agggtcgatg tacgtcgacg cgacgtagag cagggacatg 4800

accgcgatgt caaaggccag gtacatgctg ctacgaagcg ccgagcgctc gaaacagtgc 4860

gcggggatgg ccttgcgcag cgtcccgatc gtgaacggag gcttctccac aggctgcctg 4920

ttcgtcttga tagccatctc gaggcagcag cagctcggat agtatcgaca cactctggac 4980

gctggtcgtg tgatggactg ttgccgccac acttgctgcc ttgacctgtg aatatccctg 5040

ccgcttttat caaacagcct cagtgtgttt gatcttgtgt gtacgcgctt ttgcgagttg 5100

ctagctgctt gtgctatttg cgaataccac ccccagcatc cccttccctc gtttcatatc 5160

gcttgcatcc caaccgcaac ttatctacgc tgtcctgcta tccctcagcg ctgctcctgc 5220

tcctgctcac tgcccctcgc acagccttgg tttgggctcc gcctgtattc tcctggtact 5280

gcaacctgta aaccagcact gcaatgctga tgcacgggaa gtagtgggat gggaacacaa 5340

atggaaagct gtagagctct tgttttccag aaggagttgc tccttgagcc tttcattctc 5400

agcctcgata acctccaaag ccgctctaat tgtggagggg gttcgaattt aaaagcttgg 5460

aatgttggtt cgtgcgtctg gaacaagccc agacttgttg ctcactggga aaaggaccat 5520

cagctccaaa aaacttgccg ctcaaaccgc gtacctctgc tttcgcgcaa tctgccctgt 5580

tgaaatcgcc accacattca tattgtgacg cttgagcagt ctgtaattgc ctcagaatgt 5640

ggaatcatct gccccctgtg cgagcccatg ccaggcatgt cgcgggcgag gacacccgcc 5700

actcgtacag cagaccatta tgctacctca caatagttca taacagtgac catatttctc 5760

gaagctcccc aacgagcacc tccatgctct gagtggccac cccccggccc tggtgcttgc 5820

ggagggcagg tcaaccggca tggggctacc gaaatccccg accggatccc accacccccg 5880

cgatgggaag aatctctccc cgggatgtgg gcccaccacc agcacaacct gctggcccag 5940

gcgagcgtca aaccatacca cacaaatatc cttggcatcg gccctgaatt ccttctgccg 6000

ctctgctacc cggtgcttct gtccgaagca ggggttgcta gggatcgctc cgagtccgca 6060

aacccttgtc gcgtggcggg gcttgttcga gcttgaagag c 6101

<210> 165

<211> 595

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 165

actagtatgg ctatcaagac gaacaggcag cctgtggaga agcctccgtt cacgatcggg 60

acgctgcgca aggccatccc cgcgcactgt ttcgagcgct cggcgcttcg tagcagcatg 120

tacctggcct ttgacatcgc ggtcatgtcc ctgctctacg tcgcgtcgac gtacatcgac 180

cctgcaccgg tgcctacgtg ggtcaagtac ggcatcatgt ggccgctcta ctggttcttc 240

caggtgtgtt tgagggtttt ggttgcccgt attgaggtcc tggtggcgcg catggaggag 300

aaggcgcctg tcccgctgac ccccccggct accctcccgg caccttccag ggcgcgtacg 360

ggaagaacca gtagagcggc cacatgatgc cgtacttgac ccacgtaggc accggtgcag 420

ggtcgatgta cgtcgacgcg acgtagagca gggacatgac cgcgatgtca aaggccaggt 480

acatgctgct acgaagcgcc gagcgctcga aacagtgcgc ggggatggcc ttgcgcagcg 540

tcccgatcgt gaacggaggc ttctccacag gctgcctgtt cgtcttgata gccat 595

<210> 166

<211> 5348

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 166

gctcttcgcc gccgccactc ctgctcgagc gcgcccgcgc gtgcgccgcc agcgccttgg 60

ccttttcgcc gcgctcgtgc gcgtcgctga tgtccatcac caggtccatg aggtctgcct 120

tgcgccggct gagccactgc ttcgtccggg cggccaagag gagcatgagg gaggactcct 180

ggtccagggt cctgacgtgg tcgcggctct gggagcgggc cagcatcatc tggctctgcc 240

gcaccgaggc cgcctccaac tggtcctcca gcagccgcag tcgccgccga ccctggcaga 300

ggaagacagg tgaggggggt atgaattgta cagaacaacc acgagccttg tctaggcaga 360

atccctacca gtcatggctt tacctggatg acggcctgcg aacagctgtc cagcgaccct 420

cgctgccgcc gcttctcccg cacgcttctt tccagcaccg tgatggcgcg agccagcgcc 480

gcacgctggc gctgcgcttc gccgatctga ggacagtcgg ggaactctga tcagtctaaa 540

ccccccttgcg cgttagtgtt gccatccttt gcagaccggt gagagccgac ttgttgtgcg 600

ccacccccca caccacctcc tcccagacca attctgtcac ctttttggcg aaggcatcgg 660

cctcggcctg cagagaggac agcagtgccc agccgctggg ggttggcgga tgcacgctca 720

ggtaccctttt cttgcgctat gacacttcca gcaaaaggta gggcgggctg cgagacggct 780

tcccggcgct gcatgcaaca ccgatgatgc ttcgaccccc cgaagctcct tcggggctgc 840

atgggcgctc cgatgccgct ccagggcgag cgctgtttaa atagccaggc ccccgattgc 900

aaagacatta tagcgagcta ccaaagccat attcaaacac ctagatcact accacttcta 960

cacaggccac tcgagcttgt gatcgcactc cgctaagggg gcgcctcttc ctcttcgttt 1020

cagtcacaac ccgcaaacgg cgcgccatgc tgctgcaggc cttcctgttc ctgctggccg 1080

gcttcgccgc caagatcagc gcctccatga cgaacgagac gtccgaccgc cccctggtgc 1140

acttcacccc caacaagggc tggatgaacg accccaacgg cctgtggtac gacgagaagg 1200

acgccaagtg gcacctgtac ttccagtaca acccgaacga caccgtctgg gggacgccct 1260

tgttctgggg ccacgccacg tccgacgacc tgaccaactg ggaggaccag cccatcgcca 1320

tcgccccgaa gcgcaacgac tccggcgcct tctccggctc catggtggtg gactacaaca 1380
acacctccgg cttcttcaac gacaccatcg acccgcgcca gcgctgcgtg gccatctgga 1440
cctacaacac cccggagtcc gaggagcagt acatctccta cagcctggac ggcggctaca 1500
ccttcaccga gtaccagaag aaccccgtgc tggccgccaa ctccacccag ttccgcgacc 1560
cgaaggtctt ctggtacgag ccctcccaga agtggatcat gaccgcggcc aagtcccagg 1620
actacaagat cgagatctac tcctccgacg acctgaagtc ctggaagctg gagtccgcgt 1680
tcgccaacga gggcttcctc ggctaccagt acgagtgccc cggcctgatc gaggtcccca 1740

ccgagcagga ccccagcaag tcctactggg tgatgttcat ctccatcaac cccggcgccc 1800
cggccggcgg ctccttcaac cagtacttcg tcggcagctt caacggcacc cacttcgagg 1860
ccttcgacaa ccagtcccgc gtggtggact tcggcaagga ctactacgcc ctgcagacct 1920
tcttcaacac cgacccgacc tacgggagcg ccctgggcat cgcgtgggcc tccaactggg 1980
agtactccgc cttcgtgccc accaacccct ggcgctcctc catgtccctc gtgcgcaagt 2040
tctccctcaa caccgagtac caggccaacc cggagacgga gctgatcaac ctgaaggccg 2100
agccgatcct gaacatcagc aacgccggcc cctggagccg gttcgccacc aacaccacgt 2160

tgacgaaggc caacagctac aacgtcgacc tgtccaacag caccggcacc ctggagttcg 2220
agctggtgta cgccgtcaac accacccaga cgatctccaa gtccgtgttc gcggacctct 2280
ccctctggtt caagggcctg gaggaccccg aggagtacct ccgcatgggc ttcgaggtgt 2340
ccgcgtcctc cttcttcctg gaccgcggga acagcaaggt gaagttcgtg aaggagaacc 2400
cctacttcac caaccgcatg agcgtgaaca accagccctt caagagcgag aacgacctgt 2460
cctactacaa ggtgtacggc ttgctggacc agaacatcct ggagctgtac ttcaacgacg 2520
gcgacgtcgt gtccaccaac acctacttca tgaccaccgg gaacgccctg ggctccgtga 2580

acatgacgac gggggtggac aacctgttct acatcgacaa gttccaggtg cgcgaggtca 2640
agtgacaatt ggcagcagca gctcggatag tatcgacaca ctctggacgc tggtcgtgtg 2700
atggactgtt gccgccacac ttgctgcctt gacctgtgaa tatccctgcc gcttttatca 2760
aacagcctca gtgtgtttga tcttgtgtgt acgcgctttt gcgagttgct agctgcttgt 2820
gctatttgcg aataccaccc ccagcatccc cttccctcgt ttcatatcgc ttgcatccca 2880
accgcaactt atctacgctg tcctgctatc cctcagcgct gctcctgctc ctgctcactg 2940
cccctcgcac agccttggtt tgggctccgc ctgtattctc ctggtactgc aacctgtaaa 3000

ccagcactgc aatgctgatg cacgggaagt agtgggatgg gaacacaaat ggaggatccc 3060

gcgtctcgaa cagagcgcgc agaggaacgc tgaaggtctc gcctctgtcg cacctcagcg 3120

cggcatacac cacaataacc acctgacgaa tgcgcttggt tcttcgtcca ttagcgaagc 3180

gtccggttca cacacgtgcc acgttggcga ggtggcaggt gacaatgatc ggtggagctg 3240

atggtcgaaa cgttcacagc ctagggatat cgaattcctt tcttgcgcta tgacacttcc 3300

agcaaaaggt agggcgggct gcgagacggc ttcccggcgc tgcatgcaac accgatgatg 3360

cttcgacccc ccgaagctcc ttcggggctg catgggcgct ccgatgccgc tccagggcga 3420

gcgctgttta aatagccagg cccccgattg caaagacatt atagcgagct accaaagcca 3480

tattcaaaca cctagatcac taccacttct acacaggcca ctcgagcttg tgatcgcact 3540

ccgctaaggg ggcgcctctt cctcttcgtt tcagtcacaa cccgcaaaca ctagtatggc 3600

tatcaagacg aacaggcagc ctgtggagaa gcctccgttc acgatcggga cgctgcgcaa 3660

ggccatcccc gcgcactgtt tcgagcgctc ggcgcttcgt agcagcatgt acctggcctt 3720

tgacatcgcg gtcatgtccc tgctctacgt cgcgtcgacg tacatcgacc ctgcaccggt 3780

gcctacgtgg gtcaagtacg gcatcatgtg gccgctctac tggttcttcc aggtgtgttt 3840

gagggttttg gttgcccgta ttgaggtcct ggtggcgcgc atggaggaga aggcgcctgt 3900

cccgctgacc cccccggcta ccctcccggc accttccagg gcgcgtacgg gaagaaccag 3960

tagagcggcc acatgatgcc gtacttgacc cacgtaggca ccggtgcagg gtcgatgtac 4020

gtcgacgcga cgtagagcag ggacatgacc gcgatgtcaa aggccaggta catgctgcta 4080

cgaagcgccg agcgctcgaa acagtgcgcg gggatggcct tgcgcagcgt cccgatcgtg 4140

aacggaggct tctccacagg ctgcctgttc gtcttgatag ccatctcgag gcagcagcag 4200

ctcggatagt atcgacacac tctggacgct ggtcgtgtga tggactgttg ccgccacact 4260

tgctgccttg acctgtgaat atccctgccg cttttatcaa acagcctcag tgtgtttgat 4320

cttgtgtgta cgcgcttttg cgagttgcta gctgcttgtg ctatttgcga ataccacccc 4380

cagcatcccc ttccctcgtt tcatatcgct tgcatcccaa ccgcaactta tctacgctgt 4440

cctgctatcc ctcagcgctg ctcctgctcc tgctcactgc ccctcgcaca gccttggttt 4500

gggctccgcc tgtattctcc tggtactgca acctgtaaac cagcactgca atgctgatgc 4560

acgggaagta gtgggatggg aacacaaatg gaaagctgta gagctcttgt tttccagaag 4620

gagttgctcc ttgagccttt cattctcagc ctcgataacc tccaaagccg ctctaattgt 4680

ggaggggggtt cgaatttaaa agcttggaat gttggttcgt gcgtctggaa caagcccaga 4740

cttgttgctc actgggaaaa ggaccatcag ctccaaaaaa cttgccgctc aaaccgcgta 4800

cctctgcttt cgcgcaatct gccctgttga aatcgccacc acattcatat tgtgacgctt 4860

gagcagtctg taattgcctc agaatgtgga atcatctgcc ccctgtgcga gcccatgcca 4920

ggcatgtcgc gggcgaggac acccgccact cgtacagcag accattatgc tacctcacaa 4980

tagttcataa cagtgaccat atttctcgaa gctccccaac gagcacctcc atgctctgag 5040

tggccacccc ccggccctgg tgcttgcgga gggcaggtca accggcatgg ggctaccgaa 5100

atccccgacc ggatcccacc acccccgcga tgggaagaat ctctccccgg gatgtgggcc 5160

caccaccagc acaacctgct ggcccaggcg agcgtcaaac cataccacac aaatatcctt 5220

ggcatcggcc ctgaattcct tctgccgctc tgctacccgg tgcttctgtc cgaagcaggg 5280

gttgctaggg atcgctccga gtccgcaaac ccttgtcgcg tggcggggct tgttcgagct 5340

tgaagagc 5348

<210> 167

<211> 1220

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 167

actagttcac ttgtggaacc agagcgcgga gtcgtcctcg ggcgcgtccg ggacgacgta 60

gcggcagtcg cgccagtcct cccacagggc gcggccgacc cagcggctgt cggactggta 120

gtacttgccc aggatgggcc tgatggcggc ggaggcctcc tcggcgtggt agtgcgggat 180

ggtgctgaag aggtggtgca gcacgtgggt gtcggagatg tggtgcagga tgttgtccat 240

gaagggcggg cccatggagc ggtccacggt ggccatggcg ccgcgcagcc agtcccagtc 300

cttctcgaag tagtgcggca gcgccgggtg cgtgtgctgg agcagcgtga tgagcacgag 360

ccacatgttc acgatcaggt agggcaccac gtaggtcttg accagccagg cccagcccat 420

ggtgcggccc agcacgctga gcccgctgag caccgccacc agcgccaggt cggagatgac 480

cacctcgatg cgctcgcgct tgctgaagat gggcgaccac gggtcaaagt ggttggcgaa 540

gcgcgggtac ggccgcgagg cgacgttgaa catgaggtac agcggccagc ccagggtcag 600

ggtgaccagc accttgccca tgcggatggg cagccactcc tcccactcca ggccctcgtg 660

cgccactgcg cggtgcggcg gcacaaacac ctcgtccttg tccaggcacc ccgtgttgga 720

gtggtggcgg cggtgcgagt gcttccagga gtagtagggc accagcagca ggctgtggaa 780

caccaggccc acgccgtcgt tgatggcctg gctggaggaa aaggcctggt ggccgcactc 840

gtgcgcgcac acccagacac ccgtgccgaa ggcgccctgg aaggtgccgg gagggtagcc 900

gggggggtca gcgggacagg cgccttctcc tccatgcgcg ccaccaggac ctcaatacgg 960

gcaaccaaaa ccctcaaaca cacctggaag aaccagtaga gcggccacat gatgccgtac 1020

ttgacccacg taggcaccgg tgcagggtcg atgtacgtcg acgcgacgta gagcagggac 1080

atgaccgcga tgtcaaaggc caggtacatg ctgctacgaa gcgccgagcg ctcgaaacag 1140

tgcgcgggga tggccttgcg cagcgtcccg atcgtgaacg gaggcttctc cacaggctgc 1200

ctgttcgtct tgatagccat 1220

<210> 168

<211> 6726

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 168

gctcttcgcc gccgccactc ctgctcgagc gcgcccgcgc gtgcgccgcc agcgccttgg 60

ccttttcgcc gcgctcgtgc gcgtcgctga tgtccatcac caggtccatg aggtctgcct 120

tgcgccggct gagccactgc ttcgtccggg cggccaagag gagcatgagg gaggactcct 180

ggtccagggt cctgacgtgg tcgcggctct gggagcgggc cagcatcatc tggctctgcc 240

gcaccgaggc cgcctccaac tggtcctcca gcagccgcag tcgccgccga ccctggcaga 300

ggaagacagg tgaggggggt atgaattgta cagaacaacc acgagccttg tctaggcaga 360

atccctacca gtcatggctt tacctggatg acggcctgcg aacagctgtc cagcgaccct 420

cgctgccgcc gcttctcccg cacgcttctt tccagcaccg tgatggcgcg agccagcgcc 480

gcacgctggc gctgcgcttc gccgatctga ggacagtcgg ggaactctga tcagtctaaa 540

ccccccttgcg cgttagtgtt gccatccttt gcagaccggt gagagccgac ttgttgtgcg 600

ccacccccca caccacctcc tcccagacca attctgtcac ctttttggcg aaggcatcgg 660

cctcggcctg cagagaggac agcagtgccc agccgctggg ggttggcgga tgcacgctca 720

ggtacccttt cttgcgctat gacacttcca gcaaaaggta gggcgggctg cgagacggct 780

tcccggcgct gcatgcaaca ccgatgatgc ttcgaccccc cgaagctcct tcggggctgc 840

atgggcgctc cgatgccgct ccagggcgag cgctgtttaa atagccaggc ccccgattgc 900

aaagacatta tagcgagcta ccaaagccat attcaaacac ctagatcact accacttcta 960

cacaggccac tcgagcttgt gatcgcactc cgctaagggg gcgcctcttc ctcttcgttt 1020

cagtcacaac ccgcaaacgg cgcgccatgc tgctgcaggc cttcctgttc ctgctggccg 1080

gcttcgccgc caagatcagc gcctccatga cgaacgagac gtccgaccgc cccctggtgc 1140

acttcacccc caacaagggc tggatgaacg accccaacgg cctgtggtac gacgagaagg 1200

acgccaagtg gcacctgtac ttccagtaca acccgaacga caccgtctgg gggacgccct 1260

tgttctgggg ccacgccacg tccgacgacc tgaccaactg ggaggaccag cccatcgcca 1320
tcgccccgaa gcgcaacgac tccggcgcct tctccggctc catggtggtg gactacaaca 1380
acacctccgg cttcttcaac gacaccatcg acccgcgcca gcgctgcgtg gccatctgga 1440
cctacaacac cccggagtcc gaggagcagt acatctccta cagcctggac ggcggctaca 1500
ccttcaccga gtaccagaag aaccccgtgc tggccgccaa ctccacccag ttccgcgacc 1560
cgaaggtctt ctggtacgag ccctcccaga agtggatcat gaccgcggcc aagtcccagg 1620
actacaagat cgagatctac tcctccgacg acctgaagtc ctggaagctg gagtccgcgt 1680

tcgccaacga gggcttcctc ggctaccagt acgagtgccc cggcctgatc gaggtcccca 1740
ccgagcagga ccccagcaag tcctactggg tgatgttcat ctccatcaac cccggcgccc 1800
cggccggcgg ctccttcaac cagtacttcg tcggcagctt caacggcacc cacttcgagg 1860
ccttcgacaa ccagtcccgc gtggtggact tcggcaagga ctactacgcc ctgcagacct 1920
tcttcaacac cgacccgacc tacgggagcg ccctgggcat cgcgtgggcc tccaactggg 1980
agtactccgc cttcgtgccc accaacccct ggcgctcctc catgtccctc gtgcgcaagt 2040
tctccctcaa caccgagtac caggccaacc cggagacgga gctgatcaac ctgaaggccg 2100

agccgatcct gaacatcagc aacgccggcc cctggagccg gttcgccacc aacaccacgt 2160
tgacgaaggc caacagctac aacgtcgacc tgtccaacag caccggcacc ctggagttcg 2220
agctggtgta cgccgtcaac accacccaga cgatctccaa gtccgtgttc gcggacctct 2280
ccctctggtt caagggcctg gaggaccccg aggagtacct ccgcatgggc ttcgaggtgt 2340
ccgcgtcctc cttcttcctg gaccgcggga acagcaaggt gaagttcgtg aaggagaacc 2400
cctacttcac caaccgcatg agcgtgaaca accagccctt caagagcgag aacgacctgt 2460
cctactacaa ggtgtacggc ttgctggacc agaacatcct ggagctgtac ttcaacgacg 2520

gcgacgtcgt gtccaccaac acctacttca tgaccaccgg gaacgccctg ggctccgtga 2580
acatgacgac gggggtggac aacctgttct acatcgacaa gttccaggtg cgcgaggtca 2640
agtgacaatt ggcagcagca gctcggatag tatcgacaca ctctggacgc tggtcgtgtg 2700
atggactgtt gccgccacac ttgctgcctt gacctgtgaa tatccctgcc gcttttatca 2760
aacagcctca gtgtgtttga tcttgtgtgt acgcgctttt gcgagttgct agctgcttgt 2820
gctatttgcg aataccaccc ccagcatccc cttccctcgt ttcatatcgc ttgcatccca 2880
accgcaactt atctacgctg tcctgctatc cctcagcgct gctcctgctc ctgctcactg 2940

cccctcgcac agccttggtt tgggctccgc ctgtattctc ctggtactgc aacctgtaaa 3000

ccagcactgc aatgctgatg cacgggaagt agtgggatgg gaacacaaat ggaggatccc 3060

gcgtctcgaa cagagcgcgc agaggaacgc tgaaggtctc gcctctgtcg cacctcagcg 3120

cggcatacac cacaataacc acctgacgaa tgcgcttggt tcttcgtcca ttagcgaagc 3180

gtccggttca cacacgtgcc acgttggcga ggtggcaggt gacaatgatc ggtggagctg 3240

atggtcgaaa cgttcacagc ctagggatat cgaattcggc cgacaggacg cgcgtcaaag 3300

gtgctggtcg tgtatgccct ggccggcagg tcgttgctgc tgctggttag tgattccgca 3360

accctgattt tggcgtctta ttttggcgtg gcaaacgctg gcgcccgcga gccgggccgg 3420

cggcgatgcg gtgccccacg gctgccggaa tccaagggag gcaagagcgc ccgggtcagt 3480

tgaagggctt tacgcgcaag gtacagccgc tcctgcaagg ctgcgtggtg gaattggacg 3540

tgcaggtcct gctgaagttc ctccaccgcc tcaccagcgg acaaagcacc ggtgtatcag 3600

gtccgtgtca tccactctaa agagctcgac tacgacctac tgatggccct agattcttca 3660

tcaaaaacgc ctgagacact tgcccaggat tgaaactccc tgaagggacc accaggggcc 3720

ctgagttgtt ccttcccccc gtggcgagct gccagccagg ctgtacctgt gatcgaggct 3780

ggcgggaaaa taggcttcgt gtgctcaggt catgggaggt gcaggacagc tcatgaaacg 3840

ccaacaatcg cacaattcat gtcaagctaa tcagctattt cctcttcacg agctgtaatt 3900

gtcccaaaat tctggtctac cgggggtgat ccttcgtgta cgggcccttc cctcaaccct 3960

aggtatgcgc gcatgcggtc gccgcgcaac tcgcgcgagg gccgagggtt tgggacgggc 4020

cgtcccgaaa tgcagttgca cccggatgcg tggcaccttt tttgcgataa tttatgcaat 4080

ggactgctct gcaaaattct ggctctgtcg ccaaccctag gatcagcggc gtaggatttc 4140

gtaatcattc gtcctgatgg ggagctaccg actaccctaa tatcagcccg actgcctgac 4200

gccagcgtcc acttttgtgc acacattcca ttcgtgccca agacatttca ttgtggtgcg 4260

aagcgtcccc agttacgctc acctgtttcc cgacctcctt actgttctgt cgacagagcg 4320

ggcccacagg ccggtcgcag ccactagttc acttgtggaa ccagagcgcg gagtcgtcct 4380

cgggcgcgtc cgggacgacg tagcggcagt cgcgccagtc ctcccacagg gcgcggccga 4440

cccagcggct gtcggactgg tagtacttgc ccaggatggg cctgatggcg gcggaggcct 4500

cctcggcgtg gtagtgcggg atggtgctga agaggtggtg cagcacgtgg gtgtcggaga 4560

tgtggtgcag gatgttgtcc atgaagggcg ggcccatgga gcggtccacg gtggccatgg 4620

cgccgcgcag ccagtcccag tccttctcga agtagtgcgg cagcgccggg tgcgtgtgct 4680

ggagcagcgt gatgagcacg agccacatgt tcacgatcag gtagggcacc acgtaggtct 4740

tgaccagcca ggcccagccc atggtgcggc ccagcacgct gagcccgctg agcaccgcca 4800

ccagcgccag gtcggagatg accacctcga tgcgctcgcg cttgctgaag atgggcgacc 4860

acgggtcaaa gtggttggcg aagcgcgggt acggccgcga ggcgacgttg aacatgaggt 4920

acagcggcca gcccagggtc agggtgacca gcaccttgcc catgcggatg ggcagccact 4980

cctcccactc caggccctcg tgcgccactg cgcggtgcgg cggcacaaac acctcgtcct 5040

tgtccaggca ccccgtgttg gagtggtggc ggcggtgcga gtgcttccag gagtagtagg 5100

gcaccagcag caggctgtgg aacaccaggc ccacgccgtc gttgatggcc tggctggagg 5160

aaaaggcctg gtggccgcac tcgtgcgcgc acacccagac acccgtgccg aaggcgccct 5220

ggaaggtgcc gggagggtag ccgggggggt cagcgggaca ggcgccttct cctccatgcg 5280

cgccaccagg acctcaatac gggcaaccaa aaccctcaaa cacacctgga agaaccagta 5340

gagcggccac atgatgccgt acttgaccca cgtaggcacc ggtgcagggt cgatgtacgt 5400

cgacgcgacg tagagcaggg acatgaccgc gatgtcaaag gccaggtaca tgctgctacg 5460

aagcgccgag cgctcgaaac agtgcgcggg gatggccttg cgcagcgtcc cgatcgtgaa 5520

cggaggcttc tccacaggct gcctgttcgt cttgatagcc atctcgaggc agcagcagct 5580

cggatagtat cgacacactc tggacgctgg tcgtgtgatg gactgttgcc gccacacttg 5640

ctgccttgac ctgtgaatat ccctgccgct tttatcaaac agcctcagtg tgtttgatct 5700

tgtgtgtacg cgcttttgcg agttgctagc tgcttgtgct atttgcgaat accaccccca 5760

gcatcccctt ccctcgtttc atatcgcttg catcccaacc gcaacttatc tacgctgtcc 5820

tgctatccct cagcgctgct cctgctcctg ctcactgccc ctcgcacagc cttggtttgg 5880

gctccgcctg tattctcctg gtactgcaac ctgtaaacca gcactgcaat gctgatgcac 5940

gggaagtagt gggatgggaa cacaaatgga aagctgtaga gctcttgttt tccagaagga 6000

gttgctcctt gagcctttca ttctcagcct cgataacctc caaagccgct ctaattgtgg 6060

agggggttcg aatttaaaag cttggaatgt tggttcgtgc gtctggaaca agcccagact 6120

tgttgctcac tgggaaaagg accatcagct ccaaaaaact tgccgctcaa accgcgtacc 6180

tctgctttcg cgcaatctgc cctgttgaaa tcgccaccac attcatattg tgacgcttga 6240

gcagtctgta attgcctcag aatgtggaat catctgcccc ctgtgcgagc ccatgccagg 6300

catgtcgcgg gcgaggacac ccgccactcg tacagcagac cattatgcta cctcacaata 6360

gttcataaca gtgaccatat ttctcgaagc tccccaacga gcacctccat gctctgagtg 6420

gccacccccc ggccctggtg cttgcggagg gcaggtcaac cggcatgggg ctaccgaaat 6480

ccccgaccgg atcccaccac ccccgcgatg ggaagaatct ctccccggga tgtgggccca 6540

ccaccagcac aacctgctgg cccaggcgag cgtcaaacca taccacacaa atatccttgg 6600

catcggccct gaattccttc tgccgctctg ctacccggtg cttctgtccg aagcaggggt 6660

tgctagggat cgctccgagt ccgcaaaccc ttgtcgcgtg gcggggcttg ttcgagcttg 6720

aagagc 6726

<210> 169

<211> 1220

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 169

actagttcac ttgtggaacc agagcgcgga gtcgtcctcg ggcgcgtccg ggacgacgta 60

gcggcagtcg cgccagtcct cccacagggc gcggccgacc cagcggctgt cggactggta 120

gtacttgccc aggatgggcc tgatggcggc ggaggcctcc tcggcgtggt agtgcgggat 180

ggtgctgaag aggtggtgca gcacgtgggt gtcggagatg tggtgcagga tgttgtccat 240

gaagggcggg cccatggagc ggtccacggt ggccatggcg ccgcgcagcc agtcccagtc 300

cttctcgaag tagtgcggca gcgccgggtg cgtgtgctgg agcagcgtga tgagcacgag 360

ccacatgttc acgatcaggt agggcaccac gtaggtcttg accagccagg cccagcccat 420

ggtgcggccc agcacgctga gcccgctgag caccgccacc agcgccaggt cggagatgac 480

cacctcgatg cgctcgcgct tgctgaagat gggcgaccac gggtcaaagt ggttggcgaa 540

gcgcgggtac ggccgcgagg cgacgttgaa catgaggtac agcggccagc ccagggtcag 600

ggtgaccagc accttgccca tgcggatggg cagccactcc tcccactcca ggccctcgtg 660

cgccactgcg cggtgcggcg gcacaaacac ctcgtccttg tccaggcacc ccgtgttgga 720

gtggtggcgg cggtgcgagt gcttccagga gtagtagggc accagcagca ggctgtggaa 780

caccaggccc acgccgtcgt tgatggcctg gctggaggaa aaggcctggt ggccgcactc 840

gtgcgcgcac acccagacac ccgtgccgaa ggcgccctgg aaggtgccgg gagggtagcc 900

gggggggtca gcgggacagg cgccttctcc tccatgcgcg ccaccaggac ctcaatacgg 960

gcaaccaaaa ccctcaaaca cacctggaag aaccagtaga gcggccacat gatgccgtac 1020

ttgacccacg taggcaccgg tgcagggtcg atgtacgtcg acgcgacgta gagcagggac 1080

atgaccgcga tgtcaaaggc caggtacatg ctgctacgaa gcgccgagcg ctcgaaacag 1140

tgcgcgggga tggccttgcg cagcgtcccg atcgtgaacg gaggcttctc cacaggctgc 1200

ctgttcgtct tgatagccat 1220

<210> 170

<211> 5973

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 170

gctcttcgcc gccgccactc ctgctcgagc gcgcccgcgc gtgcgccgcc agcgccttgg 60

ccttttcgcc gcgctcgtgc gcgtcgctga tgtccatcac caggtccatg aggtctgcct 120

tgcgccggct gagccactgc ttcgtccggg cggccaagag gagcatgagg gaggactcct 180

ggtccagggt cctgacgtgg tcgcggctct gggagcgggc cagcatcatc tggctctgcc 240

gcaccgaggc cgcctccaac tggtcctcca gcagccgcag tcgccgccga ccctggcaga 300

ggaagacagg tgagggggt atgaattgta cagaacaacc acgagccttg tctaggcaga 360

atccctacca gtcatggctt tacctggatg acggcctgcg aacagctgtc cagcgaccct 420

cgctgccgcc gcttctcccg cacgcttctt tccagcaccg tgatggcgcg agccagcgcc 480

gcacgctggc gctgcgcttc gccgatctga ggacagtcgg ggaactctga tcagtctaaa 540

ccccccttgcg cgttagtgtt gccatccttt gcagaccggt gagagccgac ttgttgtgcg 600

ccacccccca caccacctcc tcccagacca attctgtcac ctttttggcg aaggcatcgg 660

cctcggcctg cagagaggac agcagtgccc agccgctggg ggttggcgga tgcacgctca 720

ggtacccttt cttgcgctat gacacttcca gcaaaaggta gggcgggctg cgagacggct 780

tcccggcgct gcatgcaaca ccgatgatgc ttcgaccccc cgaagctcct tcggggctgc 840

atgggcgctc cgatgccgct ccagggcgag cgctgtttaa atagccaggc ccccgattgc 900

aaagacatta tagcgagcta ccaaagccat attcaaacac ctagatcact accacttcta 960

cacaggccac tcgagcttgt gatcgcactc cgctaagggg gcgcctcttc ctcttcgttt 1020

cagtcacaac ccgcaaacgg cgcgccatgc tgctgcaggc cttcctgttc ctgctggccg 1080

gcttcgccgc caagatcagc gcctccatga cgaacgagac gtccgaccgc cccctggtgc 1140

acttcacccc caacaagggc tggatgaacg accccaacgg cctgtggtac gacgagaagg 1200

acgccaagtg gcacctgtac ttccagtaca acccgaacga caccgtctgg gggacgccct 1260

tgttctgggg ccacgccacg tccgacgacc tgaccaactg ggaggaccag cccatcgcca 1320

tcgccccgaa gcgcaacgac tccggcgcct tctccggctc catggtggtg gactacaaca 1380

acacctccgg cttcttcaac gacaccatcg acccgcgcca gcgctgcgtg gccatctgga 1440

cctacaacac cccggagtcc gaggagcagt acatctccta cagcctggac ggcggctaca 1500

ccttcaccga gtaccagaag aaccccgtgc tggccgccaa ctccacccag ttccgcgacc 1560

cgaaggtctt ctggtacgag ccctcccaga agtggatcat gaccgcggcc aagtcccagg 1620

actacaagat cgagatctac tcctccgacg acctgaagtc ctggaagctg gagtccgcgt 1680

tcgccaacga gggcttcctc ggctaccagt acgagtgccc cggcctgatc gaggtcccca 1740

ccgagcagga ccccagcaag tcctactggg tgatgttcat ctccatcaac cccggcgccc 1800

cggccggcgg ctccttcaac cagtacttcg tcggcagctt caacggcacc cacttcgagg 1860

ccttcgacaa ccagtcccgc gtggtggact tcggcaagga ctactacgcc ctgcagacct 1920

tcttcaacac cgacccgacc tacgggagcg ccctgggcat cgcgtgggcc tccaactggg 1980

agtactccgc cttcgtgccc accaacccct ggcgctcctc catgtccctc gtgcgcaagt 2040

tctccctcaa caccgagtac caggccaacc cggagacgga gctgatcaac ctgaaggccg 2100

agccgatcct gaacatcagc aacgccggcc cctggagccg gttcgccacc aacaccacgt 2160

tgacgaaggc caacagctac aacgtcgacc tgtccaacag caccggcacc ctggagttcg 2220

agctggtgta cgccgtcaac accacccaga cgatctccaa gtccgtgttc gcggacctct 2280

ccctctggtt caagggcctg gaggaccccg aggagtacct ccgcatgggc ttcgaggtgt 2340

ccgcgtcctc cttcttcctg gaccgcggga acagcaaggt gaagttcgtg aaggagaacc 2400

cctacttcac caaccgcatg agcgtgaaca accagccctt caagagcgag aacgacctgt 2460

cctactacaa ggtgtacggc ttgctggacc agaacatcct ggagctgtac ttcaacgacg 2520

gcgacgtcgt gtccaccaac acctacttca tgaccaccgg gaacgccctg ggctccgtga 2580

acatgacgac gggggtggac aacctgttct acatcgacaa gttccaggtg cgcgaggtca 2640

agtgacaatt ggcagcagca gctcggatag tatcgacaca ctctggacgc tggtcgtgtg 2700

atggactgtt gccgccacac ttgctgcctt gacctgtgaa tatccctgcc gcttttatca 2760

aacagcctca gtgtgtttga tcttgtgtgt acgcgctttt gcgagttgct agctgcttgt 2820

gctatttgcg aataccaccc ccagcatccc cttccctcgt ttcatatcgc ttgcatccca 2880

accgcaactt atctacgctg tcctgctatc cctcagcgct gctcctgctc ctgctcactg 2940

cccctcgcac agccttggtt tgggctccgc ctgtattctc ctggtactgc aacctgtaaa 3000

ccagcactgc aatgctgatg cacgggaagt agtgggatgg gaacacaaat ggaggatccc 3060

gcgtctcgaa cagagcgcgc agaggaacgc tgaaggtctc gcctctgtcg cacctcagcg 3120

cggcatacac cacaataacc acctgacgaa tgcgcttggt tcttcgtcca ttagcgaagc 3180

gtccggttca cacacgtgcc acgttggcga ggtggcaggt gacaatgatc ggtggagctg 3240

atggtcgaaa cgttcacagc ctagggatat cgaattcctt tcttgcgcta tgacacttcc 3300

agcaaaaggt agggcgggct gcgagacggc ttcccggcgc tgcatgcaac accgatgatg 3360

cttcgacccc ccgaagctcc ttcggggctg catgggcgct ccgatgccgc tccagggcga 3420

gcgctgttta aatagccagg cccccgattg caaagacatt atagcgagct accaaagcca 3480

tattcaaaca cctagatcac taccacttct acacaggcca ctcgagcttg tgatcgcact 3540

ccgctaaggg ggcgcctctt cctcttcgtt tcagtcacaa cccgcaaaca ctagttcact 3600

tgtggaacca gagcgcggag tcgtcctcgg gcgcgtccgg gacgacgtag cggcagtcgc 3660

gccagtcctc ccacagggcg cggccgaccc agcggctgtc ggactggtag tacttgccca 3720

ggatgggcct gatggcggcg gaggcctcct cggcgtggta gtgcgggatg gtgctgaaga 3780

ggtggtgcag cacgtgggtg tcggagatgt ggtgcaggat gttgtccatg aagggcgggc 3840

ccatggagcg gtccacggtg gccatggcgc cgcgcagcca gtcccagtcc ttctcgaagt 3900

agtgcggcag cgccgggtgc gtgtgctgga gcagcgtgat gagcacgagc cacatgttca 3960

cgatcaggta gggcaccacg taggtcttga ccagccaggc ccagcccatg gtgcggccca 4020

gcacgctgag cccgctgagc accgccacca gcgccaggtc ggagatgacc acctcgatgc 4080

gctcgcgctt gctgaagatg ggcgaccacg ggtcaaagtg gttggcgaag cgcgggtacg 4140

gccgcgaggc gacgttgaac atgaggtaca gcggccagcc cagggtcagg gtgaccagca 4200

ccttgcccat gcggatgggc agccactcct cccactccag gccctcgtgc gccactgcgc 4260

ggtgcggcgg cacaaacacc tcgtccttgt ccaggcaccc cgtgttggag tggtggcggc 4320

ggtgcgagtg cttccaggag tagtagggca ccagcagcag gctgtggaac accaggccca 4380

cgccgtcgtt gatggcctgg ctggaggaaa aggcctggtg gccgcactcg tgcgcgcaca 4440

cccagacacc cgtgccgaag gcgccctgga aggtgccggg agggtagccg ggggggtcag 4500

cgggacaggc gccttctcct ccatgcgcgc caccaggacc tcaatacggg caaccaaaac 4560

cctcaaacac acctggaaga accagtagag cggccacatg atgccgtact tgacccacgt 4620

aggcaccggt gcagggtcga tgtacgtcga cgcgacgtag agcagggaca tgaccgcgat 4680

gtcaaaggcc aggtacatgc tgctacgaag cgccgagcgc tcgaaacagt gcgcggggat 4740

ggccttgcgc agcgtcccga tcgtgaacgg aggcttctcc acaggctgcc tgttcgtctt 4800

gatagccatc tcgaggcagc agcagctcgg atagtatcga cacactctgg acgctggtcg 4860

tgtgatggac tgttgccgcc acacttgctg ccttgacctg tgaatatccc tgccgctttt 4920

atcaaacagc ctcagtgtgt ttgatcttgt gtgtacgcgc ttttgcgagt tgctagctgc 4980

ttgtgctatt tgcgaatacc acccccagca tcccccttccc tcgtttcata tcgcttgcat 5040

cccaaccgca acttatctac gctgtcctgc tatccctcag cgctgctcct gctcctgctc 5100

actgcccctc gcacagcctt ggtttgggct ccgcctgtat tctcctggta ctgcaacctg 5160

taaaccagca ctgcaatgct gatgcacggg aagtagtggg atgggaacac aaatggaaag 5220

ctgtagagct cttgttttcc agaaggagtt gctccttgag cctttcattc tcagcctcga 5280

taacctccaa agccgctcta attgtggagg gggttcgaat ttaaaagctt ggaatgttgg 5340
ttcgtgcgtc tggaacaagc ccagacttgt tgctcactgg gaaaaggacc atcagctcca 5400
aaaaacttgc cgctcaaacc gcgtacctct gctttcgcgc aatctgccct gttgaaatcg 5460
ccaccacatt catattgtga cgcttgagca gtctgtaatt gcctcagaat gtggaatcat 5520
ctgccccctg tgcgagccca tgccaggcat gtcgcgggcg aggacacccg ccactcgtac 5580
agcagaccat tatgctacct cacaatagtt cataacagtg accatatttc tcgaagctcc 5640
ccaacgagca cctccatgct ctgagtggcc acccccggc cctggtgctt gcggagggca 5700

ggtcaaccgg catggggcta ccgaaatccc cgaccggatc ccaccacccc cgcgatggga 5760
agaatctctc cccgggatgt gggcccacca ccagcacaac ctgctggccc aggcgagcgt 5820
caaaccatac cacacaaata tccttggcat cggccctgaa ttccttctgc cgctctgcta 5880
cccggtgctt ctgtccgaag caggggttgc tagggatcgc tccgagtccg caaacccttg 5940
tcgcgtggcg gggcttgttc gagcttgaag agc 5973

<210> 171
<211> 1221
<212> DNA
<213> Artificial Sequence
<220><223> synthesized construct
<400> 171

actagtatgg ccaccgcatc cactttctcg gcgttcaatg cccgctgcgg cgacctgcgt 60

cgctcggcgg gctccgggcc ccggcgccca gcgaggcccc tccccgtgcg cgggcgcgcc 120
gaggtgcacg tgcaggtgac ccactccctg gcccccgaga agcgcgagat cttcaactcc 180
ctgaacaact gggcccagga gaacatcctg gtgctgctga aggacgtgga caagtgctgg 240
cagccctccg acttcctgcc cgactccgcc tccgagggct tcgacgagca ggtgatggag 300
ctgcgcaagc gctgcaagga gatccccgac gactacttca tcgtgctggt gggcgacatg 360
atcaccgagg aggccctgcc cacctaccag accatgctga acaccctgga cggcgtgcgc 420
gacgagaccg gcgcctccct gacccccctgg gccatctgga cccgcgcctg gaccgccgag 480

gagaaccgcc acggcgacct gctgaacaag tacctgtacc tgtccggccg cgtggacatg 540
aagcagatcg agaagaccat ccagtacctg atcggctccg gcatggaccc ccgcaccgag 600
aacaacccct acctgggctt catctacacc tccttccagg agcgcgccac cttcatctcc 660
cacggcaaca ccgcccgcct ggccaaggag cacggcgacc tgaagctggc ccagatctgc 720

ggcatcatcg ccgccgacga gaagcgccac gagaccgcct acaccaagat cgtggagaag 780

ctgttcgaga tcgaccccga cggcaccgtg ctggccctgg ccgacatgat gcgcaagaag 840

gtgtccatgc ccgcccacct gatgtacgac ggccaggacg acaacctgtt cgagaacttc 900

tcctccgtgg cccagcgcct gggcgtgtac accgccaagg actacgccga catcctggag 960

ttcctggtgg gccgctggga catcgagaag ctgaccggcc tgtccggcga gggccgcaag 1020

gcccaggact acgtgtgcac cctgcccccc cgcatccgcc gcctggagga gcgcgcccag 1080

tcccgcgtga agaaggcctc cgccaccccc ttctcctgga tcttcggccg cgagatcaac 1140

ctgatggact acaaggacca cgacggcgac tacaaggacc acgacatcga ctacaaggac 1200

gacgacgaca agtgaatcga t 1221

<210> 172

<211> 390

<212> PRT

<213> Olea europaea

<400> 172

Met Ala Leu Lys Leu Cys Phe Pro Pro His Lys Met Pro Ser Phe Pro
1 5 10 15

Asp Ala Arg Ile Arg Ser His Arg Val Phe Met Ala Ser Thr Ile His
20 25 30

Ser Pro Ser Met Glu Val Gly Lys Val Lys Lys Pro Phe Thr Pro Pro
35 40 45

Arg Glu Val His Val Gln Val Thr His Ser Leu Ala Pro Glu Lys Arg
50 55 60

Glu Ile Phe Asn Ser Leu Asn Asn Trp Ala Gln Glu Asn Ile Leu Val
65 70 75 80

Leu Leu Lys Asp Val Asp Lys Cys Trp Gln Pro Ser Asp Phe Leu Pro
85 90 95

Asp Ser Ala Ser Glu Gly Phe Asp Glu Gln Val Met Glu Leu Arg Lys
100 105 110

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

Met Ile Thr Glu Glu Ala Leu Pro Thr Tyr Gln Thr Met Leu Asn Thr

130 135 140

Leu Asp Gly Val Arg Asp Glu Thr Gly Ala Ser Leu Thr Pro Trp Ala
145 150 155 160

Ile Trp Thr Arg Ala Trp Thr Ala Glu Glu Asn Arg His Gly Asp Leu
165 170 175

Leu Asn Lys Tyr Leu Tyr Leu Ser Gly Arg Val Asp Met Lys Gln Ile
180 185 190

Glu Lys Thr Ile Gln Tyr Leu Ile Gly Ser Gly Met Asp Pro Arg Thr
195 200 205

Glu Asn Asn Pro Tyr Leu Gly Phe Ile Tyr Thr Ser Phe Gln Glu Arg
210 215 220

Ala Thr Phe Ile Ser His Gly Asn Thr Ala Arg Leu Ala Lys Glu His
225 230 235 240

Gly Asp Leu Lys Leu Ala Gln Ile Cys Gly Ile Ile Ala Ala Asp Glu
245 250 255

Lys Arg His Glu Thr Ala Tyr Thr Lys Ile Val Glu Lys Leu Phe Glu
260 265 270

Ile Asp Pro Asp Gly Thr Val Leu Ala Leu Ala Asp Met Met Arg Lys
275 280 285

Lys Val Ser Met Pro Ala His Leu Met Tyr Asp Gly Gln Asp Asp Asn
290 295 300

Leu Phe Glu Asn Phe Ser Ser Val Ala Gln Arg Leu Gly Val Tyr Thr
305 310 315 320

Ala Lys Asp Tyr Ala Asp Ile Leu Glu Phe Leu Val Gly Arg Trp Asp
325 330 335

Ile Glu Lys Leu Thr Gly Leu Ser Gly Glu Gly Arg Lys Ala Gln Asp
340 345 350

Tyr Val Cys Thr Leu Pro Pro Arg Ile Arg Arg Leu Glu Glu Arg Ala
355 360 365

Gln Ser Arg Val Lys Lys Ala Ser Ala Thr Pro Phe Ser Trp Ile Phe
370 375 380

Gly Arg Glu Ile Asn Leu

385 390

<210> 173

<211> 1254

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 173

actagtatgg gcggcggcgg ccgcatgtcc accgtgatca cctccaacaa ctccgagaag 60

aagggcggct cctcccacct gaagcgcgcc ccccacacca agccccccct caccctgggc 120

gacctgaagc gcgccatccc cccccactgc ttcgagcgct ccttcgtgcg cgggcgcgcc 180

tccttctcct acgtggccta cgacgtgtgc ctgtccttcc tgttctactc catcgccacc 240

aacttcttcc cctacatctc ctcccccctg tcctacgtgg cctggctggt gtactggctg 300

ttccagggct gcatcctgac cggcctgtgg gtgatcggcc acgagtgcgg ccaccacgcc 360

ttctccgagt accagctggc cgacgacatc gtgggcctga tcgtgcactc cgccctgctg 420

gtgccctact tctcctggaa gtactcccac cgccgccacc actccaacat cggctccctg 480

gagcgcgacg aggtgttcgt gcccaagtcc aagtccaaga tctcctggta ctccaagtac 540

tccaacaacc cccccggccg cgtgctgacc ctggccgcca ccctgctgct gggctggccc 600

ctgtacctgg ccttcaacgt gtccggccgc ccctacgacc gcttcgcctg ccactacgac 660

ccctacggcc ccatcttctc cgagcgcgag cgcctgcaga tctacatcgc cgacctgggc 720

atcttcgcca ccaccttcgt gctgtaccag gccaccatgg ccaagggcct ggcctgggtg 780

atgcgcatct acggcgtgcc cctgctgatc gtgaactgct tcctggtgat gatcacctac 840

ctgcagcaca cccaccccgc catcccccgc tacggctcct ccgagtggga ctggctgcgc 900

ggcgccatgg tgaccgtgga ccgcgactac ggcgtgctga acaaggtgtt ccacaacatc 960

gccgacaccc acgtggccca ccacctgttc gccaccgtgc cccactacca cgccatggag 1020

gccaccaagg ccatcaagcc catcatgggc gagtactacc gctacgacgg caccccctac 1080

tacaaggccc tgtggcgcga ggccaaggag tgcctgttcg tggagcccga cgagggcgcc 1140

cccacccagg gcgtgttctg gtaccgcaac aagtacatgg actacaagga ccacgacggc 1200

gactacaagg accacgacat cgactacaag gacgacgacg acaagtgaat cgat 1254

<210> 174

<211> 387

<212> PRT

<213> Ricinus communis

<400> 174

Met Gly Gly Gly Gly Arg Met Ser Thr Val Ile Thr Ser Asn Asn Ser
1 5 10 15

Glu Lys Lys Gly Gly Ser Ser His Leu Lys Arg Ala Pro His Thr Lys
20 25 30

Pro Pro Phe Thr Leu Gly Asp Leu Lys Arg Ala Ile Pro Pro His Cys
35 40 45

Phe Glu Arg Ser Phe Val Arg Ser Phe Ser Tyr Val Ala Tyr Asp Val
50 55 60

Cys Leu Ser Phe Leu Phe Tyr Ser Ile Ala Thr Asn Phe Phe Pro Tyr
65 70 75 80

Ile Ser Ser Pro Leu Ser Tyr Val Ala Trp Leu Val Tyr Trp Leu Phe
85 90 95

Gln Gly Cys Ile Leu Thr Gly Leu Trp Val Ile Gly His Glu Cys Gly
100 105 110

His His Ala Phe Ser Glu Tyr Gln Leu Ala Asp Asp Ile Val Gly Leu
115 120 125

Ile Val His Ser Ala Leu Leu Val Pro Tyr Phe Ser Trp Lys Tyr Ser
130 135 140

His Arg Arg His His Ser Asn Ile Gly Ser Leu Glu Arg Asp Glu Val
145 150 155 160

Phe Val Pro Lys Ser Lys Ser Lys Ile Ser Trp Tyr Ser Lys Tyr Ser
165 170 175

Asn Asn Pro Pro Gly Arg Val Leu Thr Leu Ala Ala Thr Leu Leu Leu
180 185 190

Gly Trp Pro Leu Tyr Leu Ala Phe Asn Val Ser Gly Arg Pro Tyr Asp
195 200 205

Arg Phe Ala Cys His Tyr Asp Pro Tyr Gly Pro Ile Phe Ser Glu Arg
210 215 220

Glu Arg Leu Gln Ile Tyr Ile Ala Asp Leu Gly Ile Phe Ala Thr Thr

225 230 235 240

Phe Val Leu Tyr Gln Ala Thr Met Ala Lys Gly Leu Ala Trp Val Met

245 250 255

Arg Ile Tyr Gly Val Pro Leu Leu Ile Val Asn Cys Phe Leu Val Met

260 265 270

Ile Thr Tyr Leu Gln His Thr His Pro Ala Ile Pro Arg Tyr Gly Ser

275 280 285

Ser Glu Trp Asp Trp Leu Arg Gly Ala Met Val Thr Val Asp Arg Asp

290 295 300

Tyr Gly Val Leu Asn Lys Val Phe His Asn Ile Ala Asp Thr His Val

305 310 315 320

Ala His His Leu Phe Ala Thr Val Pro His Tyr His Ala Met Glu Ala

325 330 335

Thr Lys Ala Ile Lys Pro Ile Met Gly Glu Tyr Tyr Arg Tyr Asp Gly

340 345 350

Thr Pro Phe Tyr Lys Ala Leu Trp Arg Glu Ala Lys Glu Cys Leu Phe

355 360 365

Val Glu Pro Asp Glu Gly Ala Pro Thr Gln Gly Val Phe Trp Tyr Arg

370 375 380

Asn Lys Tyr

385

<210> 175

<211> 535

<212> PRT

<213> Ricinus communis

<400> 175

Met Ser Val Thr Cys Ser Lys Glu Asn Arg His Ala Phe Phe Ser Ser

1 5 10 15

Ser Thr Pro Gly Thr Thr Ser Ser His Ser Arg Thr Arg Arg Arg Pro

20 25 30

Lys Tyr Asn Ser Ile Ser Thr Pro Ala Ser Gln Ser Phe Phe Asn Ser

35 40 45

Leu Ser Ser Ser Gly Ser Ser Phe Gln Gln Leu Met Ser Ser Cys Leu
50 55 60
Ala Phe Glu Pro Cys Ser His Tyr Tyr Ser Ser Asn Gly Leu Phe Pro
65 70 75 80
Asn Thr Pro Leu Leu Pro Lys Arg His Pro Arg Leu His His Arg Leu
85 90 95

Pro Arg Ser Gly Glu Ala Met Ala Val Ala Val Gln Pro Glu Lys Glu
100 105 110
Val Ala Thr Asn Lys Lys Pro Leu Met Lys Gln Arg Arg Val Val Val
115 120 125
Thr Gly Met Gly Val Val Ser Pro Leu Gly His Asp Ile Asp Val Tyr
130 135 140
Tyr Asn Asn Leu Leu Asp Gly Ser Ser Gly Ile Ser Gln Ile Asp Ser
145 150 155 160

Phe Asp Cys Ala Gln Phe Pro Thr Arg Ile Ala Gly Glu Ile Lys Ser
165 170 175
Phe Ser Thr Asp Gly Trp Val Ala Pro Lys Leu Ser Lys Arg Met Asp
180 185 190
Lys Phe Met Leu Tyr Met Leu Thr Ala Gly Lys Lys Ala Leu Ala Asp
195 200 205
Gly Gly Ile Thr Glu Asp Met Met Asp Glu Leu Asp Lys Ala Arg Cys
210 215 220

Gly Val Leu Ile Gly Ser Ala Met Gly Gly Met Lys Val Phe Asn Asp
225 230 235 240
Ala Ile Glu Ala Leu Arg Ile Ser Tyr Arg Lys Met Asn Pro Phe Cys
245 250 255
Val Pro Phe Ala Thr Thr Asn Met Gly Ser Ala Met Leu Ala Met Asp
260 265 270
Leu Gly Trp Met Gly Pro Asn Tyr Ser Ile Ser Thr Ala Cys Ala Thr
275 280 285

Ser Asn Phe Cys Ile Leu Asn Ala Ala Asn His Ile Ile Arg Gly Glu

290 295 300

Ala Asp Ile Met Leu Cys Gly Gly Ser Asp Ala Ala Ile Ile Pro Ile
305 310 315 320

Gly Leu Gly Gly Phe Val Ala Cys Arg Ala Leu Ser Gln Arg Asn Asp
325 330 335

Asp Pro Thr Lys Ala Ser Arg Pro Trp Asp Met Asn Arg Asp Gly Phe
340 345 350

Val Met Gly Glu Gly Ala Gly Val Leu Leu Leu Glu Glu Leu Glu His
355 360 365

Ala Lys Lys Arg Gly Ala Asn Ile Tyr Ala Glu Phe Leu Gly Gly Ser
370 375 380

Phe Thr Cys Asp Ala Tyr His Met Thr Glu Pro Arg Pro Asp Gly Val
385 390 395 400

Gly Val Ile Leu Cys Ile Glu Lys Ala Leu Ala Arg Ser Gly Val Ser
405 410 415

Lys Glu Glu Val Asn Tyr Ile Asn Ala His Ala Thr Ser Thr Pro Ala
420 425 430

Gly Asp Leu Lys Glu Tyr Glu Ala Leu Met Arg Cys Phe Ser Gln Asn
435 440 445

Pro Asp Leu Arg Val Asn Ser Thr Lys Ser Met Ile Gly His Leu Leu
450 455 460

Gly Ala Ala Gly Ala Val Glu Ala Ile Ala Thr Ile Gln Ala Ile Arg
465 470 475 480

Thr Gly Trp Val His Pro Asn Ile Asn Leu Glu Asn Pro Glu Glu Gly
485 490 495

Val Asp Thr Lys Val Leu Val Gly Pro Lys Lys Glu Arg Leu Asp Ile
500 505 510

Lys Val Ala Leu Ser Asn Ser Phe Gly Phe Gly Gly His Asn Ser Ser
515 520 525

Ile Ile Phe Ala Pro Tyr Lys
530 535

<210> 176

<211> 491

<212> PRT

<213> Helianthus annus

<400> 176

Met Ala Ala Ser Met Val Cys Thr Trp Ile Val Ala Ala Cys Met Ser
1 5 10 15
Val Ser Ser Ser Ser Ser Ser Asn Ser Leu Phe Ser Ser Lys Arg Lys
20 25 30
Arg Arg Leu Asp Arg Phe Ser Leu Ser Ser Arg Cys Arg Arg Gly Gly
35 40 45
Arg Val Ala Met Ala Ile Ala Ile Gln Pro Ser Ser Ile Glu Met Glu
50 55 60

Glu Glu Thr Thr Leu Thr Lys Arg Lys Gln Pro Pro Thr Lys Gln Arg
65 70 75 80
Arg Val Val Val Thr Gly Met Gly Val Glu Thr Pro Ile Gly Asn Asn
85 90 95
Pro Asp Gln Phe Tyr Asn Asn Leu Leu Gln Gly Val Ser Gly Ile Thr
100 105 110
Gln Ile Glu Ala Phe Asp Cys Ser Ser Tyr Pro Thr Arg Ile Ala Gly
115 120 125

Glu Ile Lys Asn Phe Ser Thr Asp Gly Trp Val Ala Pro Lys Leu Ser
130 135 140
Lys Arg Met Asp Arg Phe Met Leu Tyr Met Leu Thr Ala Gly Lys Lys
145 150 155 160
Ala Leu Ala Asp Ala Gly Ile Ser Pro Ser Asp Ser Asp Glu Ile Asp
165 170 175
Lys Ser Arg Cys Gly Val Leu Ile Gly Ser Ala Met Gly Gly Met Lys
180 185 190

Val Phe Asn Asp Ala Ile Glu Ala Leu Arg Val Ser Tyr Arg Lys Met
195 200 205
Asn Pro Phe Cys Val Pro Phe Ala Thr Thr Asn Met Gly Ser Ala Met

210 215 220

Leu Ala Met Asp Leu Gly Trp Met Gly Pro Asn Tyr Ser Ile Ser Thr
225 230 235 240

Ala Cys Ala Thr Ser Asn Phe Cys Ile Leu Asn Ala Ala Asn His Ile
245 250 255

Ile Arg Gly Glu Ala Asp Met Met Leu Cys Gly Gly Ser Asp Ala Val
260 265 270

Ile Ile Pro Ile Gly Leu Gly Gly Phe Val Ala Cys Arg Ala Leu Ser
275 280 285

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

Arg Asp Gly Phe Val Met Gly Glu Gly Ala Gly Val Leu Leu Leu Glu
305 310 315 320

Glu Leu Glu His Ala Lys Lys Arg Gly Ala Lys Ile Tyr Ala Glu Phe
325 330 335

Leu Gly Gly Ser Phe Thr Cys Asp Ala Tyr His Met Thr Glu Pro His
340 345 350

Pro Glu Gly Ala Gly Val Ile Leu Cys Ile Glu Lys Ala Leu Ser Gln
355 360 365

Ala Gly Val Arg Arg Glu Asp Val Asn Tyr Ile Asn Ala His Ala Thr
370 375 380

Ser Thr Pro Ala Gly Asp Leu Lys Glu Tyr His Ala Leu Leu His Cys
385 390 395 400

Phe Gly Asn Asn Gln Glu Leu Arg Val Asn Ser Thr Lys Ser Met Ile
405 410 415

Gly His Leu Leu Gly Ala Ala Gly Ala Val Glu Ala Val Ala Thr Val
420 425 430

Gln Ala Ile Arg Thr Gly Trp Ile His Pro Asn Ile Asn Leu Glu Asn
435 440 445

Pro Asp Gln Gly Val Asp Thr Lys Val Leu Val Gly Ser Lys Lys Glu
450 455 460

Arg Leu Asn Val Lys Val Gly Leu Ser Asn Ser Phe Gly Phe Gly Gly
465 470 475 480
His Asn Ser Ser Ile Leu Phe Ala Pro Phe Gln
485 490

<210> 177
<211> 433
<212> PRT
<213> Brassica napus
<400> 177

Met Ala Val Ala Met Asp Met Glu Lys Glu Ala Lys Val Asp Asn Lys
1 5 10 15

Pro Pro Thr Glu Gln Arg Arg Val Val Val Thr Gly Met Gly Val Glu
20 25 30
Thr Ser Leu Gly His Asp Pro Asp Thr Phe Tyr Glu Asn Leu Leu Gln
35 40 45
Gly Asn Ser Gly Ile Ser Gln Ile Glu Asn Phe Asp Cys Ser Ala Phe
50 55 60
Pro Thr Arg Ile Ala Gly Glu Ile Lys Ser Phe Ser Thr Glu Gly Trp
65 70 75 80

Val Ala Pro Lys Leu Ser Lys Arg Met Asp Lys Phe Met Leu Tyr Leu
85 90 95
Leu Thr Ala Gly Lys Lys Ala Leu Val Asp Gly Gly Val Thr Glu Glu
100 105 110
Val Met Ala Glu Phe Asp Lys Ala Lys Cys Gly Val Leu Ile Gly Ser
115 120 125
Ala Met Gly Gly Met Lys Val Phe Gln Asp Ala Ile Glu Ala Met Lys
130 135 140

Ile Ser Tyr Lys Lys Met Asn Pro Phe Cys Val Pro Phe Ala Thr Thr
145 150 155 160
Asn Met Gly Ser Ala Met Leu Ala Leu Asp Leu Gly Trp Met Gly Pro
165 170 175
Asn Tyr Ser Ile Ser Thr Ala Cys Ala Thr Ser Asn Phe Cys Ile Leu

180 185 190

Asn Ser Ala Asn His Ile Ile Lys Gly Glu Ala Asp Val Met Leu Cys

195 200 205

Gly Gly Ser Asp Ser Val Ile Ile Pro Ile Gly Leu Gly Gly Phe Val

210 215 220

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

225 230 235 240

Arg Pro Trp Asp Ser Asn Arg Asp Gly Phe Val Met Gly Glu Gly Ala

245 250 255

Gly Val Leu Leu Leu Glu Glu Leu Glu His Ala Lys Lys Arg Gly Ala

260 265 270

Thr Ile Tyr Ala Glu Phe Leu Gly Gly Ser Phe Thr Cys Asp Ala Tyr

275 280 285

His Met Thr Glu Pro Arg Pro Asp Gly Ala Gly Val Ile Leu Cys Ile

290 295 300

Glu Arg Ala Leu Ala Asp Ala Gly Ile Ser Lys Glu Gln Ile Asn Tyr

305 310 315 320

Ile Asn Ala His Ala Thr Ser Thr Pro Ala Gly Asp Leu Lys Glu Tyr

325 330 335

Gln Ala Leu Ala His Cys Phe Gly Gln Asn Pro Glu Ile Lys Val Asn

340 345 350

Ser Thr Lys Ser Met Ile Gly His Leu Leu Gly Ala Ala Gly Ala Val

355 360 365

Glu Ala Val Ala Thr Val Gln Ala Ile Arg Thr Gly Trp Val His Pro

370 375 380

Asn Ile Asn Leu Glu Ser Pro Asp Asn Gly Val Asp Thr Asn Leu Leu

385 390 395 400

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

405 410 415

Ser Phe Gly Phe Gly Gly His Asn Ser Ser Ile Ile Phe Ala Pro Tyr

420 425 430

Lys

<210> 178

<211> 488

<212> PRT

<213> Glycine max

<400> 178

Met Ala Ser Thr Thr Thr Ser Ser Leu Cys Thr Trp Leu Val Ala Ala
1 5 10 15
Cys Met Ser Val Thr Cys His Ala Asp Arg Thr Lys Thr Pro His Ala

20 25 30
Met Phe Arg Ser Ser Lys Lys Ser Arg Tyr Ser Gln Phe Asn Val Cys
35 40 45
Arg Ser Thr His Ser Gly Lys Thr Met Ala Val Ala Leu Gln Pro Thr
50 55 60
Gln Glu Ile Thr Thr Ile Lys Lys Pro Pro Thr Lys Gln Arg Arg Val
65 70 75 80
Val Val Thr Gly Leu Gly Val Val Thr Pro Leu Gly His Glu Pro Asp

85 90 95
Ile Phe Tyr Asn Asn Leu Leu Asp Gly Ala Ser Gly Ile Ser Glu Ile
100 105 110
Glu Thr Phe Asp Cys Ala Glu Tyr Pro Thr Arg Ile Ala Gly Glu Ile
115 120 125
Lys Ser Phe Ser Thr Asp Gly Trp Val Ala Pro Lys Leu Ser Lys Arg
130 135 140
Met Asp Lys Phe Met Leu Tyr Met Leu Thr Ala Gly Lys Lys Ala Leu

145 150 155 160
Val Asp Gly Gly Ile Thr Asp Asp Val Met Asp Glu Leu Asn Lys Glu
165 170 175
Lys Cys Gly Val Leu Ile Gly Ser Ala Met Gly Gly Met Lys Val Phe
180 185 190
Asn Asp Ala Ile Glu Ala Leu Arg Ile Ser Tyr Lys Lys Met Asn Pro

195 200 205

Phe Cys Val Pro Phe Ala Thr Thr Asn Met Gly Ser Ala Met Leu Ala

210 215 220

Met Asp Leu Gly Trp Met Gly Pro Asn Tyr Ser Ile Ser Thr Ala Cys

225 230 235 240

Ala Thr Ser Asn Phe Cys Ile Leu Asn Ala Ala Asn His Ile Ile Arg

245 250 255

Gly Glu Ala Asp Val Met Leu Cys Gly Gly Ser Asp Ala Ala Ile Ile

260 265 270

Pro Ile Gly Leu Gly Gly Phe Val Ala Cys Arg Ala Leu Ser Gln Arg

275 280 285

Asn Thr Asp Pro Thr Lys Ala Ser Arg Pro Trp Asp Ile Asn Arg Asp

290 295 300

Gly Phe Val Met Gly Glu Gly Ala Gly Val Leu Leu Leu Glu Glu Leu

305 310 315 320

Glu His Ala Lys Glu Arg Gly Ala Thr Ile Tyr Ala Glu Phe Leu Gly

325 330 335

Gly Ser Phe Thr Cys Asp Ala Tyr His Val Thr Glu Pro Arg Pro Asp

340 345 350

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

355 360 365

Val Ser Lys Glu Asp Val Asn Tyr Ile Asn Ala His Ala Thr Ser Thr

370 375 380

Pro Ala Gly Asp Leu Lys Glu Tyr Gln Ala Leu Met His Cys Phe Gly

385 390 395 400

Gln Asn Pro Glu Leu Arg Val Asn Ser Thr Lys Ser Met Ile Gly His

405 410 415

Leu Leu Gly Ala Ala Gly Gly Val Glu Ala Val Ala Thr Ile Gln Ala

420 425 430

Ile Arg Thr Gly Trp Val His Pro Asn Ile Asn Leu Glu Asn Pro Asp

435 440 445

Asn Gly Val Asp Ala Lys Val Leu Val Gly Ser Lys Lys Glu Arg Leu

450 455 460

Asp Val Lys Ala Ala Leu Ser Asn Ser Phe Gly Phe Gly Gly His Asn

465 470 475 480

Ser Ser Ile Ile Phe Ala Pro Tyr

485

<210> 179

<211> 467

<212> PRT

<213> Prototheca moriformis

<400> 179

Met Gln Thr Ala His Gln Arg Pro Pro Thr Glu Gly His Cys Phe Gly

1 5 10 15

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

20 25 30

Arg Ile Ala Arg Ala Ala Ala Ala Ala Asp Ala Asn Pro Ala Arg Pro

35 40 45

Glu Arg Arg Val Val Ile Thr Gly Gln Gly Val Val Thr Ser Leu Gly

50 55 60

Gln Thr Ile Glu Gln Phe Tyr Ser Ser Leu Leu Glu Gly Val Ser Gly

65 70 75 80

Ile Ser Gln Ile Gln Lys Phe Asp Thr Thr Gly Tyr Thr Thr Thr Ile

85 90 95

Ala Gly Glu Ile Lys Ser Leu Gln Leu Asp Pro Tyr Val Pro Lys Arg

100 105 110

Trp Ala Lys Arg Val Asp Asp Val Ile Lys Tyr Val Tyr Ile Ala Gly

115 120 125

Lys Gln Ala Leu Glu Ser Ala Gly Leu Pro Ile Glu Ala Ala Gly Leu

130 135 140

Ala Gly Ala Gly Leu Asp Pro Ala Leu Cys Gly Val Leu Ile Gly Thr

145 150 155 160

Ala Met Ala Gly Met Thr Ser Phe Ala Ala Gly Val Glu Ala Leu Thr

165 170 175

Arg Gly Gly Val Arg Lys Met Asn Pro Phe Cys Ile Pro Phe Ser Ile

180 185 190

Ser Asn Met Gly Gly Ala Met Leu Ala Met Asp Ile Gly Phe Met Gly

195 200 205

Pro Asn Tyr Ser Ile Ser Thr Ala Cys Ala Thr Gly Asn Tyr Cys Ile

210 215 220

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

225 230 235 240

Ala Gly Gly Ala Asp Ala Ala Ile Ile Pro Ser Gly Ile Gly Gly Phe

245 250 255

Ile Ala Cys Lys Ala Leu Ser Lys Arg Asn Asp Glu Pro Glu Arg Ala

260 265 270

Ser Arg Pro Trp Asp Ala Asp Arg Asp Gly Phe Val Met Gly Glu Gly

275 280 285

Ala Gly Val Leu Val Leu Glu Glu Leu Glu His Ala Lys Arg Arg Gly

290 295 300

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

305 310 315 320

His His Met Thr Glu Pro Asp Pro Gln Gly Arg Gly Val Arg Leu Cys

325 330 335

Leu Glu Arg Ala Leu Glu Arg Ala Arg Leu Ala Pro Glu Arg Val Gly

340 345 350

Tyr Val Asn Ala His Gly Thr Ser Thr Pro Ala Gly Asp Val Ala Glu

355 360 365

Tyr Arg Ala Ile Arg Ala Val Ile Pro Gln Asp Ser Leu Arg Ile Asn

370 375 380

Ser Thr Lys Ser Met Ile Gly His Leu Leu Gly Gly Ala Gly Ala Val

385 390 395 400

Glu Ala Val Ala Ala Ile Gln Ala Leu Arg Thr Gly Trp Leu His Pro

405 410 415

Asn Leu Asn Leu Glu Asn Pro Ala Pro Gly Val Asp Pro Val Val Leu

420 425 430

Val Gly Pro Arg Lys Glu Arg Ala Glu Asp Leu Asp Val Val Leu Ser

435 440 445

Asn Ser Phe Gly Phe Gly Gly His Asn Ser Cys Val Ile Phe Arg Lys

450 455 460

Tyr Asp Glu

465

<210> 180

<211> 400

<212> PRT

<213> Prototheca moriformis

<400> 180

Met Ala Ser Ala Val Thr Phe Ala Cys Ala Pro Pro Arg Gly Ala Val

1 5 10 15

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

20 25 30

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

35 40 45

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

50 55 60

Glu Arg Leu Glu Leu Val Gln Ser Met Gly Gln Phe Ala Glu Glu Arg

65 70 75 80

Val Leu Pro Val Leu His Pro Val Asp Lys Leu Trp Gln Pro Gln Asp

85 90 95

Phe Leu Pro Asp Pro Glu Ser Pro Asp Phe Glu Asp Gln Val Ala Glu

100 105 110

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

115 120 125

Val Gly Asp Met Ile Thr Glu Glu Ala Leu Pro Thr Tyr Met Ala Met

130 135 140

Leu Asn Thr Leu Asp Gly Val Arg Asp Asp Thr Gly Ala Ala Asp His
145 150 155 160
Pro Trp Ala Arg Trp Thr Arg Gln Trp Val Ala Glu Glu Asn Arg His
165 170 175
Gly Asp Leu Leu Asn Lys Tyr Cys Trp Leu Thr Gly Arg Val Asn Met
180 185 190
Arg Ala Val Glu Val Thr Ile Asn Asn Leu Ile Lys Ser Gly Met Asn
195 200 205

Pro Gln Thr Asp Asn Asn Pro Tyr Leu Gly Phe Val Tyr Thr Ser Phe
210 215 220
Gln Glu Arg Ala Thr Lys Tyr Ser His Gly Asn Thr Ala Arg Leu Ala
225 230 235 240
Ala Glu His Gly Asp Lys Asn Leu Ser Lys Ile Cys Gly Leu Ile Ala
245 250 255
Ser Asp Glu Gly Arg His Glu Ile Ala Tyr Thr Arg Ile Val Asp Glu
260 265 270

Phe Phe Arg Leu Asp Pro Glu Gly Ala Val Ala Ala Tyr Ala Asn Met
275 280 285
Met Arg Lys Gln Ile Thr Met Pro Ala His Leu Met Asp Asp Met Gly
290 295 300
His Gly Glu Ala Asn Pro Gly Arg Asn Leu Phe Ala Asp Phe Ser Ala
305 310 315 320
Val Ala Glu Lys Ile Asp Val Tyr Asp Ala Glu Asp Tyr Cys Arg Ile
325 330 335

Leu Glu His Leu Asn Ala Arg Trp Lys Val Asp Glu Arg Gln Val Ser
340 345 350
Gly Gln Ala Ala Ala Asp Gln Glu Tyr Val Leu Gly Leu Pro Gln Arg
355 360 365
Phe Arg Lys Leu Ala Glu Lys Thr Ala Ala Lys Arg Lys Arg Val Ala
370 375 380

Arg Arg Pro Val Ala Phe Ser Trp Ile Ser Gly Arg Glu Ile Met Val
385 390 395 400

<210> 181
<211> 378
<212> PRT
<213> Linus usitatissimum
<400> 181
Met Gly Ala Gly Gly Arg Met Pro Val Pro Pro Ser Ser Lys Pro Met
1 5 10 15
Lys Arg Ser Pro Tyr Ser Lys Pro Pro Phe Thr Leu Gly Glu Leu Lys
20 25 30
Lys Ala Ile Pro Pro His Cys Phe Lys Arg Ser Ile Pro Arg Ser Phe
35 40 45
Ala Tyr Val Ala His Asp Leu Thr Ile Ala Ala Ile Phe Tyr Tyr Ile

50 55 60
Ala Thr Thr Tyr Phe His Leu Leu Pro Ser Pro Leu Asn Tyr Leu Ala
65 70 75 80
Trp Pro Val Tyr Trp Ala Cys Gln Gly Cys Ile Leu Thr Gly Val Trp
85 90 95
Val Leu Ala His Glu Cys Gly His His Ala Phe Ser Asp Tyr Gln Trp
100 105 110
Leu Asp Asp Met Asp Gly Phe Val Leu His Ser Ser Leu Leu Val Pro

115 120 125
Tyr Phe Ser Trp Lys His Ser His Arg Arg His His Ser Asn Thr Gly
130 135 140
Ser Leu Asp Arg Asp Glu Val Phe Val Pro Lys Gln Lys Ala Glu Ile
145 150 155 160
Gly Trp Tyr Ser Lys Tyr Leu Asn Asn Pro Pro Gly Arg Val Ile Thr
165 170 175
Leu Ala Val Thr Leu Thr Leu Gly Trp Pro Leu Tyr Leu Ala Phe Asn

180 185 190

Val Ser Gly Arg Pro Tyr Asp Arg Phe Ala Cys His Phe Asp Pro His
195 200 205
Gly Pro Ile Tyr Asn Asp Arg Glu Arg Met Glu Ile Tyr Leu Ser Asp
210 215 220
Ala Gly Ile Phe Thr Val Cys Tyr Ile Leu Tyr Arg Leu Val Leu Thr
225 230 235 240
Lys Gly Leu Val Trp Val Val Ser Ile Tyr Gly Val Pro Leu Leu Ile

245 250 255
Val Asn Gly Phe Leu Val Leu Ile Thr Phe Leu Gln His Thr His Pro
260 265 270
Ser Leu Pro His Tyr Lys Ser Ser Glu Trp Asp Trp Met Arg Gly Ala
275 280 285
Leu Ser Thr Val Asp Arg Asp Tyr Gly Leu Leu Asn Thr Val Phe His
290 295 300
Asn Ile Thr Asp Thr His Val Ala His His Leu Phe Ser Thr Met Pro

305 310 315 320
His Tyr His Ala Met Glu Ala Thr Lys Ala Ile Lys Pro Val Leu Gly
325 330 335
Glu Tyr Tyr Gln Phe Asp Gly Thr Pro Phe Val Lys Ala Met Trp Arg
340 345 350
Glu Ala Lys Glu Cys Ile Tyr Val Glu Pro Asp Glu Gly Asp Pro Ser
355 360 365
Gln Gly Val Phe Trp Tyr Asn Asn Lys Leu
370 375

<210> 182
<211> 391
<212> PRT
<213> Linus usitatissimum
<400> 182
Met Ser Pro Pro Asn Ser Met Ser Pro Thr Thr Asn Gly Asn Gly Val
1 5 10 15
Ala Met Asn Gly Ala Lys Lys Gln Leu Asp Phe Asp Pro Ser Ala Ala

20 25 30

Pro Pro Phe Lys Ile Ala Asp Ile Arg Ala Ala Ile Pro Pro His Cys

35 40 45

Trp Val Lys Asn Pro Trp Arg Ser Leu Ser Tyr Val Leu Arg Asp Leu

50 55 60

Leu Val Ile Leu Ser Phe Ala Val Ala Ala Ala Lys Leu Asp Ser Trp

65 70 75 80

Thr Phe Trp Pro Leu Tyr Trp Val Ala Gln Gly Thr Met Phe Trp Ala

85 90 95

Val Phe Val Leu Gly His Asp Cys Gly His Gly Ser Phe Ser Asp Ile

100 105 110

Trp Leu Leu Asn Asn Val Met Gly His Ile Leu His Ser Ser Ile Leu

115 120 125

Val Pro Tyr His Gly Trp Arg Ile Ser His Lys Thr His His Gln Asn

130 135 140

His Gly Asn Val Glu Lys Asp Glu Ser Trp Val Pro Leu Pro Glu Lys

145 150 155 160

Val Tyr Lys Ser Leu Asp Thr Ser Thr Lys Phe Met Arg Phe Thr Ile

165 170 175

Pro Leu Pro Met Phe Ala Tyr Pro Ile Tyr Leu Trp Thr Arg Ser Pro

180 185 190

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

195 200 205

Asn Glu Arg Ala Ala Val Leu Ile Ser Thr Leu Cys Trp Thr Ala Met

210 215 220

Ala Leu Leu Leu Cys Tyr Ser Ser Phe Ile Tyr Gly Phe Ala Pro Val

225 230 235 240

Leu Lys Ile Tyr Gly Val Pro Tyr Leu Ile Phe Val Ala Trp Leu Asp

245 250 255

Met Val Thr Tyr Leu His His His Gly Tyr Glu Gln Lys Leu Pro Trp

260 265 270

Tyr Arg Gly Lys Glu Trp Ser Tyr Leu Arg Gly Gly Leu Thr Thr Val
275 280 285
Asp Arg Asp Tyr Gly Val Ile Asn Asn Ile His His Asp Ile Gly Thr
290 295 300
His Val Ile His His Leu Phe Pro Gln Met Pro His Tyr His Leu Val

305 310 315 320
Glu Ala Thr Gln Ala Ala Lys His Val Leu Gly Lys Tyr Tyr Arg Glu
325 330 335
Pro Lys Lys Ser Gly Pro Phe Pro Phe His Leu Phe Gly Tyr Leu Val
340 345 350
Arg Ser Leu Gly Glu Asp His Tyr Val Ser Asp Thr Gly Asp Val Val
355 360 365
Phe Tyr Gln Ser Asp Pro His Ile Pro Lys Phe Arg Thr Ser Ser Ala

370 375 380
Thr Thr Lys Ser Lys Ser Ser
385 390
<210> 183
<211> 392
<212> PRT
<213> Linus usitatissimum
<400> 183
Met Ser Pro Pro Asn Ser Met Ser Pro Ala Thr Asn Gly Ser Thr Asn
1 5 10 15
Gly Val Ala Ile Asn Gly Ala Lys Lys Leu Leu Asp Phe Asp Pro Ser
20 25 30
Ala Ala Pro Pro Phe Lys Ile Ala Asp Ile Arg Ala Ala Ile Pro Pro

35 40 45
His Cys Trp Val Lys Asn Pro Trp Arg Ser Leu Ser Tyr Val Leu Arg
50 55 60
Asp Leu Leu Val Ile Leu Ser Phe Ala Val Ala Ala Thr Lys Leu Asp
65 70 75 80
Ser Trp Thr Val Trp Pro Leu Tyr Trp Ile Ala Gln Gly Thr Met Phe

85 90 95

Trp Ala Val Phe Val Leu Gly His Asp Cys Gly His Gly Ser Phe Ser

100 105 110

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

115 120 125

Ile Leu Val Pro Tyr His Gly Trp Arg Ile Ser His Lys Thr His His

130 135 140

Gln Asn His Gly Asn Val Glu Lys Asp Glu Ser Trp Val Pro Leu Pro

145 150 155 160

Glu Lys Val Tyr Lys Ser Leu Asp Thr Gly Thr Lys Phe Met Arg Phe

165 170 175

Thr Ile Pro Leu Pro Met Phe Ala Tyr Pro Ile Tyr Leu Trp Arg Arg

180 185 190

Ser Pro Gly Lys Lys Gly Ser His Phe Asn Pro Tyr Ser Asp Leu Phe

195 200 205

Ala Pro Asn Glu Arg Thr Ser Val Met Ile Ser Thr Leu Cys Trp Thr

210 215 220

Ala Met Ala Leu Leu Leu Cys Tyr Ser Ser Phe Ile Tyr Gly Phe Leu

225 230 235 240

Pro Val Phe Lys Ile Tyr Gly Val Pro Tyr Leu Ile Phe Val Ala Trp

245 250 255

Leu Asp Met Val Thr Tyr Leu His His His Gly Tyr Glu Gln Lys Leu

260 265 270

Pro Trp Tyr Arg Gly Lys Glu Trp Ser Tyr Leu Arg Gly Gly Leu Thr

275 280 285

Thr Val Asp Arg Asp Tyr Gly Val Ile Asn Asn Ile His His Asp Ile

290 295 300

Gly Thr His Val Ile His His Leu Phe Pro Gln Met Pro His Tyr His

305 310 315 320

Leu Val Glu Ala Thr Gln Ala Ala Lys His Val Leu Gly Lys Tyr Tyr

325 330 335

Arg Glu Pro Lys Lys Ser Gly Pro Phe Pro Phe His Leu Phe Gly Tyr
340 345 350
Leu Val Arg Ser Leu Gly Glu Asp His Tyr Val Ser Asp Thr Gly Asp

355 360 365
Val Val Phe Tyr Gln Ser Asp Pro His Ile Pro Lys Phe Pro Thr Ser
370 375 380
Ala Thr Thr Lys Ser Lys Ser Ser
385 390

<210> 184
<211> 383
<212> PRT
<213> Carthamus tinctorus
<400> 184

Met Gly Ala Gly Gly Arg Met Ser Asn Pro Ser Glu Gly Glu Lys Lys
1 5 10 15
Thr Glu Leu Glu Gly Ile Gln Arg Val Pro Tyr Gln Lys Pro Pro Phe

20 25 30
Thr Val Gly Asp Val Lys Lys Ala Ile Pro Pro His Cys Phe Asn Arg
35 40 45
Ser Val Ile Arg Ser Phe Ser Tyr Val Val Tyr Asp Leu Thr Ile Ala
50 55 60
Ser Ile Leu Tyr Tyr Ile Ala Thr Thr Phe Ile Pro Leu Leu Pro His
65 70 75 80
Pro Leu Ala Tyr Val Ala Trp Pro Ile Tyr Trp Ala Val Gln Gly Cys

85 90 95
Val Met Thr Gly Val Trp Val Ile Ala His Glu Cys Gly His His Ala
100 105 110
Phe Ser Asp Tyr Gln Trp Leu Asp Asp Thr Val Gly Leu Ile Leu His
115 120 125
Ser Val Leu Leu Val Pro Tyr Phe Ser Trp Lys Tyr Ser His Arg Arg
130 135 140

His His Ser Asn Thr Gly Ser Ile Glu His Asp Glu Val Phe Val Pro

145 150 155 160

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

165 170 175

Pro Gly Arg Ile Leu Thr Leu Leu Val Thr Leu Thr Leu Gly Trp Pro

180 185 190

Leu Tyr Leu Met Phe Asn Val Ser Gly Arg Tyr Tyr Asp Arg Phe Ala

195 200 205

Cys His Phe Asp Pro Asn Ser Pro Ile Tyr Ser Asn Arg Glu Arg Ala

210 215 220

Gln Ile Phe Ile Ser Asp Ala Gly Ile Phe Ala Val Leu Tyr Gly Leu

225 230 235 240

Tyr Arg Leu Ala Ala Val Lys Gly Leu Val Trp Val Leu Thr Val Tyr

245 250 255

Ala Gly Pro Leu Leu Val Val Asn Gly Phe Leu Val Leu Ile Thr Phe

260 265 270

Leu Gln His Thr His Pro Ser Leu Pro His Tyr Asp Ser Thr Glu Trp

275 280 285

Asp Trp Leu Arg Gly Ala Leu Ala Thr Ile Asp Arg Asp Tyr Gly Ile

290 295 300

Leu Asn Lys Val Phe His Asn Ile Thr Asp Thr His Val Thr His His

305 310 315 320

Leu Phe Ser Thr Met Pro His Tyr His Ala Met Glu Ala Thr Lys Ala

325 330 335

Ile Ile Pro Ile Leu Gly Asp Tyr Tyr Gln Phe Asp Gly Thr Ser Val

340 345 350

Phe Lys Ala Met Tyr Arg Glu Thr Lys Glu Cys Ile Tyr Val Asp Lys

355 360 365

Asp Glu Glu Val Lys Asp Gly Val Tyr Trp Tyr Arg Asn Lys Ile

370 375 380

<210> 185

<211> 382

<212> PRT

<213> Helianthus annus

<400> 185

Met Gly Ala Gly Gly Arg Met Ser Ser Pro Asn Gly Lys Glu Lys Asp
1 5 10 15

Gly Pro Lys Pro Leu Glu Arg Ala Leu His Glu Lys Pro Pro Phe Thr
20 25 30

Val Gly Asp Ile Lys Lys Val Ile Pro Pro His Cys Phe Lys Arg Ser
35 40 45

Val Ile Arg Ser Phe Ser Tyr Val Val Tyr Asp Leu Thr Ile Ala Ser
50 55 60

Ile Phe Tyr Tyr Leu Ala Asn Asn Tyr Ile Pro Leu Leu Pro Asn Ser
65 70 75 80

Leu Ala Tyr Val Ala Trp Pro Val Tyr Trp Ile Phe Gln Gly Cys Val
85 90 95

Leu Thr Gly Val Trp Val Ile Ala His Glu Cys Gly His His Ala Phe
100 105 110

Ser Asp Tyr Gln Trp Leu Asp Asp Thr Val Gly Leu Ile Leu His Ser
115 120 125

Ala Leu Leu Val Pro Tyr Phe Ser Trp Lys Tyr Ser His Arg Arg His
130 135 140

His Ser Asn Thr Gly Ser Ile Glu His Asp Glu Val Phe Val Pro Lys
145 150 155 160

Leu Lys Ser Ser Val Arg Ser Thr Ala Lys Tyr Leu Asn Asn Pro Pro
165 170 175

Gly Arg Ile Leu Thr Leu Leu Val Thr Leu Thr Met Gly Trp Pro Leu
180 185 190

Tyr Leu Met Phe Asn Val Ser Gly Arg Tyr Tyr Asp Arg Phe Ala Cys
195 200 205

His Phe Asp Pro Asn Ser Pro Ile Tyr Ser Asn Arg Glu Arg Ala Gln

210 215 220

Ile Phe Ile Ser Asp Ala Gly Ile Leu Thr Val Phe Tyr Ile Leu Phe

225 230 235 240

Arg Leu Ala Ser Thr Lys Gly Leu Val Trp Val Leu Thr Met Tyr Gly

245 250 255

Gly Pro Leu Leu Val Val Asn Gly Phe Leu Val Leu Ile Thr Phe Leu

260 265 270

Gln His Thr His Pro Ser Leu Pro His Tyr Asp Ser Thr Glu Trp Asp

275 280 285

Trp Leu Arg Gly Ala Leu Ala Thr Val Asp Arg Asp Tyr Gly Ile Leu

290 295 300

Asn Lys Val Phe His Asn Ile Thr Asp Thr His Val Thr His His Leu

305 310 315 320

Phe Ser Thr Met Pro His Tyr His Ala Met Glu Ala Thr Lys Ala Ile

325 330 335

Lys Pro Ile Leu Gly Asp Tyr Tyr Gln Phe Asp Gly Thr Ser Ile Phe

340 345 350

Lys Ala Met Tyr Arg Glu Thr Lys Glu Cys Ile Tyr Val Asp Lys Asp

355 360 365

Glu Asp Val Lys Asp Gly Val Tyr Trp Tyr Arg Asn Lys Ile

370 375 380

<210> 186

<211> 397

<212> PRT

<213> Cuphea wrightii

<400> 186

Met Val Ala Ala Ala Ala Ser Ser Ala Phe Phe Ser Val Pro Thr Pro

1 5 10 15

Gly Thr Ser Pro Lys Pro Gly Lys Phe Gly Asn Trp Pro Ser Ser Leu

20 25 30

Ser Val Pro Phe Lys Pro Asp Asn Gly Gly Phe Val Lys Ala Asn Ala

35 40 45

Ser Ala His Pro Lys Ala Asn Gly Ser Ala Val Asn Leu Lys Ser Gly
50 55 60
Ser Leu Glu Thr Pro Pro Arg Ser Phe Ile Asn Gln Leu Pro Asp Leu

65 70 75 80
Ser Met Leu Leu Ser Lys Ile Thr Thr Val Phe Gly Ala Ala Glu Lys
85 90 95
Gln Trp Lys Arg Pro Gly Met Leu Val Glu Pro Phe Gly Val Asp Arg
100 105 110
Ile Phe Gln Asp Gly Val Phe Phe Arg Gln Ser Phe Ser Ile Arg Ser
115 120 125
Tyr Glu Ile Gly Val Asp Arg Thr Ala Ser Ile Glu Thr Leu Met Asn

130 135 140
Ile Phe Gln Glu Thr Ser Leu Asn His Cys Lys Ser Ile Gly Leu Leu
145 150 155 160
Asn Asp Gly Phe Gly Arg Thr Pro Glu Met Cys Lys Arg Asp Leu Ile
165 170 175
Trp Val Val Thr Lys Ile Gln Val Glu Val Asn Arg Tyr Pro Thr Trp
180 185 190
Gly Asp Thr Ile Glu Val Asn Thr Trp Val Ser Glu Ser Gly Lys Asn

195 200 205
Gly Met Gly Arg Asp Trp Leu Ile Ser Asp Cys Arg Thr Gly Glu Ile
210 215 220
Leu Ile Arg Ala Thr Ser Val Trp Ala Met Met Asn Gln Asn Thr Arg
225 230 235 240
Arg Leu Ser Lys Phe Pro Tyr Glu Val Arg Gln Glu Ile Ala Pro His
245 250 255
Phe Val Asp Ser Ala Pro Val Ile Glu Asp Asp Arg Lys Leu His Lys

260 265 270
Leu Asp Val Lys Thr Gly Asp Ser Ile Arg Asp Gly Leu Thr Pro Arg
275 280 285
Trp Asn Asp Leu Asp Val Asn Gln His Val Asn Asn Val Lys Tyr Ile

290 295 300

Gly Trp Ile Leu Lys Ser Val Pro Ile Glu Val Phe Glu Thr Gln Glu

305 310 315 320

Leu Cys Gly Val Thr Leu Glu Tyr Arg Arg Glu Cys Gly Arg Asp Ser

325 330 335

Val Leu Glu Ser Val Thr Thr Met Asp Pro Ala Lys Glu Gly Asp Arg

340 345 350

Cys Val Tyr Gln His Leu Leu Arg Leu Glu Asp Gly Ala Asp Ile Thr

355 360 365

Ile Gly Arg Thr Glu Trp Arg Pro Lys Asn Ala Gly Ala Asn Gly Ala

370 375 380

Ile Ser Ser Gly Lys Thr Ser Asn Gly Asn Ser Val Ser

385 390 395

<210> 187

<211> 408

<212> PRT

<213> Cuphea wrightii

<400> 187

Met Val Val Ala Ala Ala Ala Ser Ser Ala Phe Phe Pro Val Pro Ala

1 5 10 15

Pro Arg Pro Thr Pro Lys Pro Gly Lys Phe Gly Asn Trp Pro Ser Ser

20 25 30

Leu Ser Gln Pro Phe Lys Pro Lys Ser Asn Pro Asn Gly Arg Phe Gln

35 40 45

Val Lys Ala Asn Val Ser Pro His Pro Lys Ala Asn Gly Ser Ala Val

50 55 60

Ser Leu Lys Ser Gly Ser Leu Asn Thr Leu Glu Asp Pro Pro Ser Ser

65 70 75 80

Pro Pro Pro Arg Thr Phe Leu Asn Gln Leu Pro Asp Trp Ser Arg Leu

85 90 95

Arg Thr Ala Ile Thr Thr Val Phe Val Ala Ala Glu Lys Gln Phe Thr

100 105 110

Arg Leu Asp Arg Lys Ser Lys Arg Pro Asp Met Leu Val Asp Trp Phe
115 120 125
Gly Ser Glu Thr Ile Val Gln Asp Gly Leu Val Phe Arg Glu Arg Phe
130 135 140
Ser Ile Arg Ser Tyr Glu Ile Gly Ala Asp Arg Thr Ala Ser Ile Glu
145 150 155 160
Thr Leu Met Asn His Leu Gln Asp Thr Ser Leu Asn His Cys Lys Ser
165 170 175

Val Gly Leu Leu Asn Asp Gly Phe Gly Arg Thr Pro Glu Met Cys Thr
180 185 190
Arg Asp Leu Ile Trp Val Leu Thr Lys Met Gln Ile Val Val Asn Arg
195 200 205
Tyr Pro Thr Trp Gly Asp Thr Val Glu Ile Asn Ser Trp Phe Ser Gln
210 215 220
Ser Gly Lys Ile Gly Met Gly Arg Glu Trp Leu Ile Ser Asp Cys Asn
225 230 235 240

Thr Gly Glu Ile Leu Val Arg Ala Thr Ser Ala Trp Ala Met Met Asn
245 250 255
Gln Lys Thr Arg Arg Phe Ser Lys Leu Pro Cys Glu Val Arg Gln Glu
260 265 270
Ile Ala Pro His Phe Val Asp Ala Pro Pro Val Ile Glu Asp Asn Asp
275 280 285
Arg Lys Leu His Lys Phe Asp Val Lys Thr Gly Asp Ser Ile Cys Lys
290 295 300

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

Cys Gly Arg Glu Ser Val Val Glu Ser Val Thr Ser Met Asn Pro Ser
355 360 365

Lys Val Gly Asp Arg Ser Gln Tyr Gln His Leu Leu Arg Leu Glu Asp
370 375 380
Gly Ala Asp Ile Met Lys Gly Arg Thr Glu Trp Arg Pro Lys Asn Ala
385 390 395 400
Gly Thr Asn Arg Ala Ile Ser Thr
405

<210> 188

<211> 5975

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 188

gctcttcgcc gccgccactc ctgctcgagc gcgcccgcgc gtgcgccgcc agcgccttgg 60

cctttcgcc gcgctcgtgc gcgtcgctga tgtccatcac caggtccatg aggtctgcct 120
tgcgccggct gagccactgc ttcgtccggg cggccaagag gagcatgagg gaggactcct 180
ggtccagggt cctgacgtgg tcgcggctct gggagcgggc cagcatcatc tggctctgcc 240
gcaccgaggc cgcctccaac tggtcctcca gcagccgcag tcgccgccga ccctggcaga 300
ggaagacagg tgagggggt atgaattgta cagaacaacc acgagccttg tctaggcaga 360
atccctacca gtcatggctt tacctggatg acggcctgcg aacagctgtc cagcgaccct 420
cgctgccgcc gcttctcccg cacgcttctt tccagcaccg tgatggcgcg agccagcgcc 480

gcacgctggc gctgcgcttc gccgatctga ggacagtcgg ggaactctga tcagtctaaa 540
ccccccttgcg cgttagtgtt gccatccttt gcagaccggt gagagccgac ttgttgtgcg 600
ccacccccca caccacctcc tcccagacca attctgtcac ctttttggcg aaggcatcgg 660
cctcggcctg cagagaggac agcagtgccc agccgctggg ggttggcgga tgcacgctca 720
ggtaccctttt cttgcgctat gacacttcca gcaaaaggta gggcgggctg cgagacggct 780
tcccggcgct gcatgcaaca ccgatgatgc ttcgaccccc cgaagctcct tcggggctgc 840
atgggcgctc cgatgccgct ccagggcgag cgctgtttaa atagccaggc ccccgattgc 900

aaagacatta tagcgagcta ccaaagcata ttcaaacacc tagatcacta ccacttctac 960
acaggccact cgagcttgtg atcgcactcc gctaaggggg cgcctcttcc cttcgtttca 1020

gtcacaaccc gcaaacggcg cgccatatca atgctgctgc aggccttcct gttcctgctg 1080

gccggcttcg ccgccaagat cagcgcctcc atgacgaacg agacgtccga ccgccccctg 1140

gtgcacttca cccccaacaa gggctggatg aacgacccca acggcctgtg gtacgacgag 1200

aaggacgcca agtggcacct gtacttccag tacaacccga acgacaccgt ctgggggacg 1260

cccttgttct ggggccacgc cacgtccgac gacctgacca actgggagga ccagcccatc 1320

gccatcgccc cgaagcgcaa cgactccggc gccttctccg gctccatggt ggtggactac 1380

aacaacacct ccggcttctt caacgacacc atcgacccgc gccagcgctg cgtggccatc 1440

tggacctaca acaccccgga gtccgaggag cagtacatct cctacagcct ggacggcggc 1500

tacaccttca ccgagtacca gaagaacccc gtgctggccg ccaactccac ccagttccgc 1560

gacccgaagg tcttctggta cgagccctcc cagaagtgga tcatgaccgc ggccaagtcc 1620

caggactaca agatcgagat ctactcctcc gacgacctga agtcctggaa gctggagtcc 1680

gcgttcgcca acgagggctt cctcggctac cagtacgagt gccccggcct gatcgaggtc 1740

cccaccgagc aggaccccag caagtcctac tgggtgatgt tcatctccat caaccccggc 1800

gccccggccg gcggctcctt caaccagtac ttcgtcggca gcttcaacgg cacccacttc 1860

gaggccttcg acaaccagtc ccgcgtggtg gacttcggca aggactacta cgccctgcag 1920

accttcttca acaccgaccc gacctacggg agcgccctgg gcatcgcgtg ggcctccaac 1980

tgggagtact ccgccttcgt gcccaccaac ccctggcgct cctccatgtc cctcgtgcgc 2040

aagttctccc tcaacaccga gtaccaggcc aacccggaga cggagctgat caacctgaag 2100

gccgagccga tcctgaacat cagcaacgcc ggcccctgga gccggttcgc caccaacacc 2160

acgttgacga aggccaacag ctacaacgtc gacctgtcca acagcaccgg caccctggag 2220

ttcgagctgg tgtacgccgt caacaccacc cagacgatct ccaagtccgt gttcgcggac 2280

ctctccctct ggttcaaggg cctggaggac cccgaggagt acctccgcat gggcttcgag 2340

gtgtccgcgt cctccttctt cctggaccgc gggaacagca aggtgaagtt cgtgaaggag 2400

aacccctact tcaccaaccg catgagcgtg aacaaccagc ccttcaagag cgagaacgac 2460

ctgtcctact acaaggtgta cggcttgctg gaccagaaca tcctggagct gtacttcaac 2520

gacggcgacg tcgtgtccac caacacctac ttcatgacca ccgggaacgc cctgggctcc 2580

gtgaacatga cgacgggggt ggacaacctg ttctacatcg acaagttcca ggtgcgcgag 2640

gtcaagtgac aattggcagc agcagctcgg atagtatcga cacactctgg acgctggtcg 2700

tgtgatggac tgttgccgcc acacttgctg ccttgacctg tgaatatccc tgccgctttt 2760

atcaaacagc ctcagtgtgt ttgatcttgt gtgtacgcgc ttttgcgagt tgctagctgc 2820

ttgtgctatt tgcgaatacc acccccagca tcccccttccc tcgtttcata tcgcttgcat 2880

cccaaccgca acttatctac gctgtcctgc tatccctcag cgctgctcct gctcctgctc 2940

actgcccctc gcacagcctt ggtttgggct ccgcctgtat tctcctggta ctgcaacctg 3000

taaaccagca ctgcaatgct gatgcacggg aagtagtggg atgggaacac aaatggagga 3060

tcccgcgtct cgaacagagc gcgcagagga acgctgaagg tctcgcctct gtcgcacctc 3120

agcgcggcat acaccacaat aaccacctga cgaatgcgct tggttcttcg tccattagcg 3180

aagcgtccgg ttcacacacg tgccacgttg gcgaggtggc aggtgacaat gatcggtgga 3240

gctgatggtc gaaacgttca cagcctaggg atatcgaatt cctttcttgc gctatgacac 3300

ttccagcaaa aggtagggcg ggctgcgaga cggcttcccg gcgctgcatg caacaccgat 3360

gatgcttcga ccccccgaag ctccttcggg gctgcatggg cgctccgatg ccgctccagg 3420

gcgagcgctg tttaaatagc caggcccccg attgcaaaga cattatagcg agctaccaaa 3480

gccatattca aacacctaga tcactaccac ttctacacag gccactcgag cttgtgatcg 3540

cactccgcta agggggcgcc tcttcctctt cgtttcagtc acaacccgca aacactagta 3600

tggtggccgc cgccgccagc agcgccttct tcagcgtgcc cacccccggc accagcccca 3660

agcccggcaa gttcggcaac tggcccagca gcctgagcgt gcccttcaag cccgacaacg 3720

gcggcttcca cgtgaaggcc aacgccagcg cccacgggcg cgcccccaag gccaacggca 3780

gcgccgtgaa cctgaagtcc ggcagcctgg agacccccccc ccgcagcttc atcaaccagc 3840

tgcccgacct gagcatgctg ctgagcaaga tcaccaccgt gttcggcgcc gccgagaagc 3900

agtggaagcg ccccggcatg ctggtggagc ccttcggcgt ggaccgcatc ttccaggacg 3960

gcgtgttctt ccgccagagc ttcagcatcc gcagctacga gatcggcgtg gaccgcaccg 4020

ccagcatcga gaccctgatg aacatcttcc aggagaccag cctgaaccac tgcaagagca 4080

tcggcctgct gaacgacggc ttcggccgca cccccgagat gtgcaagcgc gacctgatct 4140

gggtggtgac caagatccag gtggaggtga accgctaccc cacctggggc gacaccatcg 4200

aggtgaacac ctgggtgagc gagagcggca agaacggcat gggccgcgac tggctgatca 4260

gcgactgccg caccggcgag atcctgatcc gcgccaccag cgtgtgggcc atgatgaacc 4320

agaacacccg ccgcctgagc aagttccccc acgaggtgcg ccaggagatc gccccccact 4380

tcgtggacag cgcccccgtg atcgaggacg accgcaagct gcacaagctg gacgtgaaga 4440

ccggcgacag catccgcgac ggcctgaccc cccgctggaa cgacctggac gtgaaccagc 4500

acgtgaacaa cgtgaagtac atcggctgga ttctgaagtc cgtgcccatc gaggtgttcg 4560

agacccagga gctgtgcggc gtgaccctgg agtaccgccg cgagtgcggc cgcgacagcg 4620

tgctggagag cgtgaccacc atggaccccg ccaaggaggg cgaccgctgc gtgtaccagc 4680

acctgctgcg cctggaggac ggcgccgaca tcaccatcgg ccgcaccgag tggcgcccca 4740
agaacgccgg cgccaacggc gccatcagca gcggcaagac cagcaacggc aacagcgtga 4800
gctgattaat taactcgagg cagcagcagc tcggatagta tcgacacact ctggacgctg 4860
gtcgtgtgat ggactgttgc cgccacactt gctgccttga cctgtgaata tccctgccgc 4920
ttttatcaaa cagcctcagt gtgtttgatc ttgtgtgtac gcgcttttgc gagttgctag 4980
ctgcttgtgc tatttgcgaa taccaccccc agcatcccct tccctcgttt catatcgctt 5040
gcatcccaac cgcaacttat ctacgctgtc ctgctatccc tcagcgctgc tcctgctcct 5100

gctcactgcc cctcgcacag ccttggtttg ggctccgcct gtattctcct ggtactgcaa 5160
cctgtaaacc agcactgcaa tgctgatgca cgggaagtag tgggatggga acacaaatgg 5220
aaagcttgag ctcttgtttt ccagaaggag ttgctccttg agcctttcat tctcagcctc 5280
gataacctcc aaagccgctc taattgtgga gggggttcga atttaaaagc ttggaatgtt 5340
ggttcgtgcg tctggaacaa gcccagactt gttgctcact gggaaaagga ccatcagctc 5400
caaaaaactt gccgctcaaa ccgcgtacct ctgctttcgc gcaatctgcc ctgttgaaat 5460
cgccaccaca ttcatattgt gacgcttgag cagtctgtaa ttgcctcaga atgtggaatc 5520

atctgccccc tgtgcgagcc catgccaggc atgtcgcggg cgaggacacc cgccactcgt 5580
acagcagacc attatgctac ctcacaatag ttcataacag tgaccatatt tctcgaagct 5640
ccccaacgag cacctccatg ctctgagtgg ccacccccg gccctggtgc ttgcggaggg 5700
caggtcaacc ggcatggggc taccgaaatc cccgaccgga tcccaccacc cccgcgatgg 5760
gaagaatctc tccccgggat gtgggcccac caccagcaca acctgctggc ccaggcgagc 5820
gtcaaaccat accacacaaa tatccttggc atcggccctg aattccttct gccgctctgc 5880
tacccggtgc ttctgtccga agcaggggtt gctagggatc gctccgagtc cgcaaaccct 5940

tgtcgcgtgg cggggcttgt tcgagcttga agagc 5975

<210> 189

<211> 1256

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 189

actagtatgg tggtggccgc cgccgccagc agcgccttct tccccgtgcc cgcccccccgc 60
cccacccccа agcccggcaa gttcggcaac tggcccagca gcctgagcca gcccttcaag 120

cccaagagca accccaacgg ccgcttccag gtgaaggcca acgtgagccc ccacgggcgc 180

gcccccaagg ccaacggcag cgccgtgagc ctgaagtccg gcagcctgaa caccctggag 240

gacccccca gcagcccccc cccccgcacc ttcctgaacc agctgcccga ctggagccgc 300

ctgcgcaccg ccatcaccac cgtgttcgtg gccgccgaga agcagttcac ccgcctggac 360

cgcaagagca agcgccccga catgctggtg gactggttcg gcagcgagac catcgtgcag 420

gacggcctgg tgttccgcga gcgcttcagc atccgcagct acgagatcgg cgccgaccgc 480

accgccagca tcgagaccct gatgaaccac ctgcaggaca ccagcctgaa ccactgcaag 540

agcgtgggcc tgctgaacga cggcttcggc cgcacccccg agatgtgcac ccgcgacctg 600

atctgggtgc tgaccaagat gcagatcgtg gtgaaccgct accccacctg gggcgacacc 660

gtggagatca acagctggtt cagccagagc ggcaagatcg gcatgggccg cgagtggctg 720

atcagcgact gcaacaccgg cgagatcctg gtgcgcgcca ccagcgcctg ggccatgatg 780

aaccagaaga cccgccgctt cagcaagctg ccctgcgagg tgcgccagga gatcgccccc 840

cacttcgtgg acgccccccc cgtgatcgag gacaacgacc gcaagctgca caagttcgac 900

gtgaagaccg gcgacagcat ctgcaagggc ctgacccccg gctggaacga cttcgacgtg 960

aaccagcacg tgagcaacgt gaagtacatc ggctggattc tggagagcat gcccaccgag 1020

gtgctggaga cccaggagct gtgcagcctg accctggagt accgccgcga gtgcggccgc 1080

gagagcgtgg tggagagcgt gaccagcatg aaccccagca aggtgggcga ccgcagccag 1140

taccagcacc tgctgcgcct ggaggacggc gccgacatca tgaagggccg caccgagtgg 1200

cgccccaaga acgccggcac caaccgcgcc atcagcacct gattaattaa ctcgag 1256

<210> 190

<211> 1370

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 190

gaattcggcg ctatgacact tccagcaaaa ggtagggcgg gctgcgagac ggcttcccgg 60

cgctgcatgc aacaccgatg atgcttcgac cccccgaagc tccttcgggg ctgcatgggc 120

gctccgatgc cgctccaggg cgagcgctgt ttaaatagcc aggcccccga ttgcaaagac 180

attatagcga gctaccaaag ccatattcaa acacctagat cactaccact tctacacagg 240

ccactcgagc ttgtgatcgc actccgctaa gggggcgcct cttcctcttc gtttcagtcg 300

gccgacagga cgcgcgtcaa aggtgctggt cgtgtatgcc ctggccggca ggtcgttgct 360

gctgctggtt agtgattccg caaccctgat tttggcgtct tattttggcg tggcaaacgc 420

tggcgcccgc gagccgggcc ggcggcgatg cggtgcccca cggctgccgg aatccaaggg 480

aggcaagagc gcccgggtca gttgaagggc tttacgcgca aggtacagcc gctcctgcaa 540

ggctgcgtgg tggaattgga cgtgcaggtc ctgctgaagt tcctccaccg cctcaccagc 600

ggacaaagca ccggtgtatc aggtccgtgt catccactct aaagagctcg actacgacct 660

actgatggcc ctagattctt catcaaaaac gcctgagaca cttgcccagg attgaaactc 720

cctgaaggga ccaccagggg ccctgagttg ttccttcccc ccgtggcgag ctgccagcca 780

ggctgtacct gtgatcgagg ctggcgggaa aataggcttc gtgtgctcag gtcatgggag 840

gtgcaggaca gctcatgaaa cgccaacaat cgcacaattc atgtcaagct aatcagctat 900

ttcctcttca cgagctgtaa ttgtcccaaa attctggtct accgggggtg atccttcgtg 960

tacgggccct tccctcaacc ctaggtatgc gcgcatgcgg tcgccgcgca actcgcgcga 1020

gggccgaggg tttgggacgg gccgtcccga aatgcagttg cacccggatg cgtggcacct 1080

tttttgcgat aatttatgca atggactgct ctgcaaaatt ctggctctgt cgccaaccct 1140

aggatcagcg gcgtaggatt tcgtaatcat tcgtcctgat ggggagctac cgactaccct 1200

aatatcagcc cgactgcctg acgccagcgt ccacttttgt gcacacattc cattcgtgcc 1260

caagacattt cattgtggtg cgaagcgtcc ccagttacgc tcacctgttt cccgacctcc 1320

ttactgttct gtcgacagag cgggcccaca ggccggtcgc agccactagt 1370

<210> 191

<211> 120

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 191

actagtatgg ccaccgcatc cactttctcg gcgttcaatg cccgctgcgg cgacctgcgt 60

cgctcggcgg gctccgggcc ccggcgccca gcgaggcccc tccccgtgcg cgggcgcgcc 120

<210> 192

<211> 114

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 192

actagtatgg cttccgcggc attcaccatg tcggcgtgcc ccgcgatgac tggcagggcc 60

cctggggcac gtcgctccgg acggccagtc gccacccgcc tgagggggcg cgcc 114

<210> 193

<211> 120

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 193

actagtatgg ctatcaagac gaacaggcag cctgtggaga agcctccgtt cacgatcggg 60

acgctgcgca aggccatccc cgcgcactgt ttcgagcgct cggcgcttcg tgggcgcgcc 120

<210> 194

<211> 117

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 194

actagtatga cgttcggggt cgccctcccg gccatgggcc gcggtgtctc ccttccccgg 60

cccagggtcg cggtgcgcgc ccagtcggcg agtcaggttt tggagagcgg gcgcgcc 117

<210> 195

<211> 385

<212> PRT

<213> Carthamus tinctorius

<400> 195

Met Leu Ser Lys Gly Ala Pro Ala Ala Pro Ala Val Ala Ala Met Tyr
1 5 10 15

Asn Ala Ser Ala Lys Asp Thr Thr Phe Ala Leu Thr His Ser Arg Ser
20 25 30

Ile Gly Ser Val Ser Ile Arg Arg Arg Tyr Asn Val Phe Leu Cys Asn
35 40 45

Ser Ser Ser Ser Ser Arg Lys Val Ser Pro Leu Leu Ala Val Ala Thr
50 55 60

Gly Glu Gln Pro Ser Gly Val Ala Ser Leu Arg Glu Ala Asp Lys Glu
65 70 75 80

Lys Ser Leu Gly Asn Arg Leu Arg Leu Gly Ser Leu Thr Glu Asp Gly
85 90 95
Leu Ser Tyr Lys Glu Lys Phe Val Ile Arg Cys Tyr Glu Val Gly Ile
100 105 110
Asn Lys Thr Ala Thr Ile Glu Thr Ile Ala Asn Leu Leu Gln Glu Val
115 120 125
Gly Gly Asn His Ala Gln Gly Val Gly Phe Ser Thr Asp Gly Phe Ala
130 135 140

Thr Thr Thr Thr Met Arg Lys Leu His Leu Ile Trp Val Thr Ala Arg
145 150 155 160
Met His Ile Glu Ile Tyr Arg Tyr Pro Ala Trp Ser Asp Val Ile Glu
165 170 175
Ile Glu Thr Trp Val Gln Gly Glu Gly Lys Val Gly Thr Arg Arg Asp
180 185 190
Trp Ile Leu Lys Asp Tyr Ala Asn Gly Glu Val Ile Gly Arg Ala Thr
195 200 205

Ser Lys Trp Val Met Met Asn Glu Asp Thr Arg Arg Leu Gln Lys Val
210 215 220
Ser Asp Asp Val Arg Glu Glu Tyr Leu Val Phe Cys Pro Arg Thr Leu
225 230 235 240
Arg Leu Ala Phe Pro Glu Glu Asn Asn Asn Ser Met Lys Lys Ile Pro
245 250 255
Lys Leu Glu Asp Pro Ala Glu Tyr Ser Arg Leu Gly Leu Val Pro Arg
260 265 270

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

325 330 335

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

<210> 196
<211> 395
<212> PRT
<213> Ricinus communis
<400> 196

Met Ala Leu Lys Leu Asn Arg Leu Thr Phe Pro Pro Thr His Lys Leu

1 5 10 15
Pro Ser Gln Ser Pro Phe Gln Arg Ser Pro Asn Arg Val Phe Met Ala
20 25 30
Ala Thr Val Gly Ser Thr Ser Thr Lys Glu Val Glu Asn Leu Lys Lys
35 40 45
Pro Phe Ser Pro Pro Arg Glu Val His Val Gln Val Thr His Ser Met
50 55 60
Pro Pro Gln Lys Met Glu Ile Phe Lys Ser Leu Glu Gly Trp Ala Glu

65 70 75 80
Asp Asn Ile Leu Val His Leu Lys Pro Val Glu Lys Cys Trp Gln Pro
85 90 95
Gln Asp Phe Leu Pro Asp Pro Glu Ser Glu Gly Phe Tyr Glu Gln Val
100 105 110
Lys Glu Leu Arg Glu Arg Ala Arg Glu Leu Pro Asp Asp Tyr Phe Val
115 120 125
Val Leu Val Gly Asp Met Ile Thr Glu Glu Ala Leu Pro Thr Tyr Gln

130 135 140

Thr Met Leu Asn Thr Leu Asp Gly Val Arg Asp Glu Thr Gly Ala Ser

145 150 155 160

Leu Thr Ser Trp Ala Ile Trp Thr Arg Ala Trp Thr Ala Glu Glu Asn

165 170 175

Arg His Gly Asp Leu Leu Asn Lys Tyr Leu Tyr Leu Ser Gly Arg Val

180 185 190

Asp Met Lys His Ile Glu Lys Thr Ile Gln Tyr Leu Ile Gly Ser Gly

195 200 205

Met Asp Pro Lys Thr Glu Asn Asn Pro Tyr Leu Gly Phe Ile Tyr Thr

210 215 220

Ser Phe Gln Glu Arg Ala Thr Phe Ile Ser His Gly Asn Thr Ala Arg

225 230 235 240

Leu Ala Lys Glu His Gly Asp Met Lys Leu Ala Gln Ile Cys Gly Ile

245 250 255

Ile Ala Ala Asp Glu Lys Arg His Glu Thr Ala Tyr Thr Lys Ile Ala

260 265 270

Glu Lys Leu Phe Glu Ile Asp Pro Asp Gly Thr Val Met Ala Leu Ala

275 280 285

Asp Met Met Arg Lys Lys Ile Ser Met Pro Ala His Leu Met Phe Asp

290 295 300

Gly Gln Asp Asp Asn Leu Phe Asp His Tyr Ser Ser Val Ala Gln Arg

305 310 315 320

Ile Gly Val Tyr Thr Ala Lys Asp Tyr Ala Asp Ile Leu Glu Phe Leu

325 330 335

Val Gly Arg Trp Lys Val Glu Asn Leu Ile Gly Leu Ser Gly Glu Gly

340 345 350

Arg Arg Ala Gln Asp Phe Val Cys Gly Leu Pro Gly Arg Ile Arg Arg

355 360 365

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

370 375 380

Phe Ser Trp Ile Phe Gly Arg Glu Leu Lys Leu

385 390 395

<210> 197

<211> 396

<212> PRT

<213> Helianthus annus

<400> 197

Met Ala Ile Arg Ile Asn Thr Ala Thr Phe Gln Ser Asp Leu Tyr Arg
1 5 10 15

Ser Phe Ala Phe Pro Gln Pro Lys Pro Leu Arg Ser Pro Lys Phe Ala
20 25 30

Met Ala Ser Thr Ile Gly Ser Ala Thr Thr Lys Val Glu Ser Thr Lys
35 40 45

Lys Pro Phe Thr Pro Pro Arg Glu Val His Gln Gln Val Leu His Ser

50 55 60

Met Pro Pro Gln Lys Ile Glu Ile Phe Lys Ser Met Glu Gly Trp Ala
65 70 75 80

Glu Asp Asn Ile Leu Val His Leu Lys Pro Val Glu Lys Cys Trp Gln
85 90 95

Ala Gln Asp Phe Leu Pro Asp Pro Ala Ser Asp Gly Phe Met Glu Gln
100 105 110

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

115 120 125

Val Val Leu Val Gly Asp Met Ile Thr Glu Glu Ala Leu Pro Thr Tyr
130 135 140

Gln Thr Met Leu Asn Thr Leu Asp Gly Val Arg Asp Glu Thr Gly Ala
145 150 155 160

Ser Pro Thr Ser Trp Ala Ile Trp Thr Arg Ala Trp Thr Ala Glu Glu
165 170 175

Asn Arg His Gly Asp Leu Leu His Gln Tyr Leu Tyr Leu Ser Gly Arg

180 185 190

Val Asp Met Arg Gln Ile Gln Lys Thr Ile Gln Tyr Leu Ile Gly Ser

195 200 205

Gly Met Asp Pro Arg Thr Glu Asn Ser Pro Tyr Leu Gly Phe Ile Tyr

210 215 220

Thr Ser Phe Gln Glu Arg Ala Thr Phe Ile Ser His Gly Asn Thr Ala

225 230 235 240

Arg His Ala Lys Glu His Gly Asp Val Lys Leu Ala Gln Met Cys Gly

245 250 255

Ile Ile Ala Ala Asp Glu Lys Arg His Glu Thr Ala Tyr Thr Lys Ile

260 265 270

Val Glu Lys Leu Phe Glu Ile Asp Pro Asp Gly Thr Val Leu Ala Phe

275 280 285

Ala Asp Met Met Arg Lys Lys Ile Ser Met Pro Ala His Leu Met Tyr

290 295 300

Asp Gly Arg Asp Asp Asn Leu Phe Glu Asn Phe Ser Ala Val Ala Gln

305 310 315 320

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

325 330 335

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

340 345 350

Gly Arg Lys Ala Gln Asp Tyr Val Cys Gly Leu Ala Pro Arg Ile Arg

355 360 365

Arg Leu Glu Glu Arg Asn Ser Ala Arg Ala Lys Glu Ser Val Asn Val

370 375 380

Pro Phe Ser Trp Ile Phe Asp Arg Glu Val Lys Leu

385 390 395

<210> 198

<211> 401

<212> PRT

<213> Brassica juncea

<400> 198

Met Ala Leu Lys Leu Asn Pro Leu Ala Ser Gln Pro Tyr Asn Phe Pro

1 5 10 15

Ser Ser Ala Arg Pro Pro Ile Ser Thr Phe Arg Ser Pro Lys Phe Leu
20 25 30
Cys Leu Ala Ser Ser Ser Ser Pro Ala Leu Ser Ser Ser Thr Lys Glu

35 40 45
Val Glu Ser Leu Lys Lys Pro Phe Thr Pro Pro Lys Glu Val His Val
50 55 60
Gln Val Leu His Ser Met Pro Pro Gln Lys Ile Glu Ile Phe Lys Ser
65 70 75 80
Met Glu Asp Trp Ala Glu Gln Asn Leu Leu Thr His Leu Lys Asp Val
85 90 95
Glu Lys Ser Trp Gln Pro Gln Asp Phe Leu Pro Asp Pro Ala Ser Asp

100 105 110
Gly Phe Glu Asp Gln Val Lys Glu Leu Arg Glu Arg Ala Arg Glu Leu
115 120 125
Pro Asp Asp Tyr Phe Val Val Leu Val Gly Asp Met Ile Thr Glu Glu
130 135 140
Ala Leu Pro Thr Tyr Gln Thr Met Leu Asn Thr Leu Asp Gly Val Arg
145 150 155 160
Asp Glu Thr Gly Ala Ser Pro Thr Pro Trp Ala Val Trp Thr Arg Ala

165 170 175
Trp Thr Ala Glu Glu Asn Arg His Gly Asp Leu Leu Asn Lys Tyr Leu
180 185 190
Tyr Leu Ser Gly Arg Val Asp Met Arg Gln Ile Glu Lys Thr Ile Gln
195 200 205
Tyr Leu Ile Gly Ser Gly Met Asp Pro Arg Thr Glu Asn Asn Pro Tyr
210 215 220
Leu Gly Phe Ile Tyr Thr Ser Phe Gln Glu Arg Ala Thr Phe Val Ser

225 230 235 240
His Gly Asn Thr Ala Arg Gln Ala Lys Glu His Gly Asp Leu Lys Leu
245 250 255
Ala Gln Ile Cys Gly Thr Ile Ala Ala Asp Glu Lys Arg His Glu Thr

260 265 270

Ala Tyr Thr Lys Ile Val Glu Lys Leu Leu Glu Ile Asp Pro Asp Gly

275 280 285

Thr Val Val Ala Phe Ala Asp Met Met Arg Lys Lys Ile Ser Met Pro

290 295 300

Ala His Leu Met Tyr Asp Gly Arg Asp Asp Asn Leu Phe Asp Asn Phe

305 310 315 320

Ser Ser Val Ala Gln Arg Leu Gly Val Tyr Thr Ala Lys Asp Tyr Ala

325 330 335

Asp Ile Leu Glu Phe Leu Ala Gly Arg Trp Lys Ile Glu Ser Leu Ser

340 345 350

Gly Leu Ser Gly Glu Gly Asn Lys Ala Gln Glu Tyr Leu Cys Gly Leu

355 360 365

Thr Pro Arg Ile Arg Arg Leu Asp Glu Arg Ala Gln Ala Arg Ala Lys

370 375 380

Lys Gly Pro Lys Ile Pro Phe Ser Trp Ile His Asp Arg Glu Val Gln

385 390 395 400

Leu

<210> 199

<211> 391

<212> PRT

<213> Glycine max

<400> 199

Met Ala Leu Arg Leu Asn Pro Ile Pro Thr Gln Thr Phe Ser Leu Pro

1 5 10 15

Gln Met Ala Ser Leu Arg Ser Pro Arg Phe Arg Met Ala Ser Thr Leu

20 25 30

Arg Ser Gly Ser Lys Glu Val Glu Asn Ile Lys Lys Pro Phe Thr Pro

35 40 45

Pro Arg Glu Val His Val Gln Val Thr His Ser Met Pro Pro Gln Lys

50 55 60

Ile Glu Ile Phe Lys Ser Leu Glu Asp Trp Ala Glu Gln Asn Ile Leu
65 70 75 80

Thr His Leu Lys Pro Val Glu Lys Cys Trp Gln Pro Gln Asp Phe Leu
85 90 95

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

Glu Arg Ala Lys Glu Leu Pro Asp Asp Tyr Phe Val Val Leu Val Gly
115 120 125

Asp Met Ile Thr Glu Glu Ala Leu Pro Thr Tyr Gln Thr Met Leu Asn
130 135 140

Thr Leu Asp Gly Val Arg Asp Glu Thr Gly Ala Ser Leu Thr Ser Trp
145 150 155 160

Ala Ile Trp Thr Arg Ala Trp Thr Ala Glu Glu Asn Arg His Gly Asp
165 170 175

Leu Leu Asn Lys Tyr Leu Tyr Leu Ser Gly Arg Val Asp Met Lys Gln
180 185 190

Ile Glu Lys Thr Ile Gln Tyr Leu Ile Gly Ser Gly Met Asp Pro Arg
195 200 205

Thr Glu Asn Ser Pro Tyr Leu Gly Phe Ile Tyr Thr Ser Phe Gln Glu
210 215 220

Arg Ala Thr Phe Ile Ser His Gly Asn Thr Ala Arg Leu Ala Lys Glu
225 230 235 240

His Gly Asp Ile Lys Leu Ala Gln Ile Cys Gly Met Ile Ala Ser Asp
245 250 255

Glu Lys Arg His Glu Thr Ala Tyr Thr Lys Ile Val Glu Lys Leu Phe
260 265 270

Glu Val Asp Pro Asp Gly Thr Val Met Ala Phe Ala Asp Met Met Arg
275 280 285

Lys Lys Ile Ala Met Pro Ala His Leu Met Tyr Asp Gly Arg Asp Asp
290 295 300

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

305 310 315 320

Thr Ala Lys Asp Tyr Ala Asp Ile Leu Glu Phe Leu Val Gly Arg Trp

325 330 335

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

340 345 350

Glu Tyr Ile Cys Gly Leu Pro Pro Arg Ile Arg Arg Leu Glu Glu Arg

355 360 365

Ala Gln Ala Arg Val Lys Glu Ser Ser Thr Leu Lys Phe Ser Trp Ile

370 375 380

His Asp Arg Glu Val Leu Leu

385 390

<210> 200

<211> 390

<212> PRT

<213> Olea europaea

<400> 200

Met Ala Leu Lys Leu Cys Phe Pro Pro His Lys Met Pro Ser Phe Pro

1 5 10 15

Asp Ala Arg Ile Arg Ser His Arg Val Phe Met Ala Ser Thr Ile His

20 25 30

Ser Pro Ser Met Glu Val Gly Lys Val Lys Lys Pro Phe Thr Pro Pro

35 40 45

Arg Glu Val His Val Gln Val Thr His Ser Leu Ala Pro Glu Lys Arg

50 55 60

Glu Ile Phe Asn Ser Leu Asn Asn Trp Ala Gln Glu Asn Ile Leu Val

65 70 75 80

Leu Leu Lys Asp Val Asp Lys Cys Trp Gln Pro Ser Asp Phe Leu Pro

85 90 95

Asp Ser Ala Ser Glu Gly Phe Asp Glu Gln Val Met Glu Leu Arg Lys

100 105 110

Arg Cys Lys Glu Ile Pro Asp Asp Tyr Phe Ile Val Leu Val Gly Asp

115 120 125

Met Ile Thr Glu Glu Ala Leu Pro Thr Tyr Gln Thr Met Leu Asn Thr
130 135 140
Leu Asp Gly Val Arg Asp Glu Thr Gly Ala Ser Leu Thr Pro Trp Ala
145 150 155 160
Ile Trp Thr Arg Ala Trp Thr Ala Glu Glu Asn Arg His Gly Asp Leu
165 170 175
Leu Asn Lys Tyr Leu Tyr Leu Ser Gly Arg Val Asp Met Lys Gln Ile
180 185 190

Glu Lys Thr Ile Gln Tyr Leu Ile Gly Ser Gly Met Asp Pro Arg Thr
195 200 205
Glu Asn Asn Pro Tyr Leu Gly Phe Ile Tyr Thr Ser Phe Gln Glu Arg
210 215 220
Ala Thr Phe Ile Ser His Gly Asn Thr Ala Arg Leu Ala Lys Glu His
225 230 235 240
Gly Asp Leu Lys Leu Ala Gln Ile Cys Gly Ile Ile Ala Ala Asp Glu
245 250 255

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

Ala Lys Asp Tyr Ala Asp Ile Leu Glu Phe Leu Val Gly Arg Trp Asp
325 330 335
Ile Glu Lys Leu Thr Gly Leu Ser Gly Glu Gly Arg Lys Ala Gln Asp
340 345 350
Tyr Val Cys Thr Leu Pro Pro Arg Ile Arg Arg Leu Glu Glu Arg Ala
355 360 365

Gln Ser Arg Val Lys Lys Ala Ser Ala Thr Pro Phe Ser Trp Ile Phe
370 375 380

Gly Arg Glu Ile Asn Leu
385 390

<210> 201

<211> 392

<212> PRT

<213> Vernicia fordii

<400> 201

Met Ala Leu Lys Leu Asn Pro Phe Ile Ser Gln Ser His Lys Phe Pro
1 5 10 15

Thr Phe Ala Leu Pro Pro Met Ala Asn Leu Arg Ser Pro Lys Phe Tyr
20 25 30

Met Ala Ser Thr Leu Arg Glu Ile Glu His Leu Lys Lys Pro Phe Met
35 40 45

Pro Pro Arg Glu Val His Val Gln Val Thr His Ser Met Pro Pro Gln
50 55 60

Lys Ile Glu Ile Phe Lys Ser Leu Glu Asp Trp Ala Glu Gln Asn Ile
65 70 75 80

Leu Val His Leu Lys Pro Val Glu Lys Cys Trp Gln Pro Gln Asp Phe
85 90 95

Leu Pro Asp Pro Ala Ser Asp Asp Phe Asp Glu Gln Val Lys Glu Leu
100 105 110

Arg Glu Arg Ala Lys Glu Ile Pro Asp Asp Tyr Phe Val Val Leu Val
115 120 125

Gly Asp Met Ile Thr Glu Glu Ala Leu Pro Thr Tyr Gln Thr Met Leu
130 135 140

Asn Thr Leu Asp Gly Val Arg Asp Glu Thr Gly Ala Ser Leu Thr Ser
145 150 155 160

Trp Ala Ile Trp Thr Arg Ala Trp Thr Ala Glu Glu Asn Arg His Gly
165 170 175

Asp Leu Leu Asn Lys Tyr Leu Tyr Leu Ser Gly Arg Val Asp Met Arg
180 185 190

Gln Ile Glu Lys Thr Ile Gln Tyr Leu Ile Gly Ser Gly Met Asp Pro
195 200 205

Arg Thr Glu Asn Ser Pro Tyr Leu Gly Phe Ile Tyr Thr Ser Phe Gln
210 215 220

Glu Arg Ala Thr Phe Ile Ser His Gly Asn Thr Ala Arg His Ala Lys
225 230 235 240

Glu His Gly Asp Ile Lys Leu Ala Gln Ile Cys Gly Thr Ile Ala Ser
245 250 255

Asp Glu Lys Arg His Glu Thr Ala Tyr Thr Lys Ile Val Glu Lys Leu
260 265 270

Phe Glu Ile Asp Pro Asn Gly Thr Val Leu Ala Phe Ala Asp Met Met
275 280 285

Arg Lys Lys Ile Ser Met Pro Ala His Leu Met Tyr Asp Gly Cys Asn
290 295 300

Asp Asn Leu Phe Asp His Phe Ser Ala Val Ala Gln Arg Leu Gly Val
305 310 315 320

Tyr Thr Ala Lys Asp Tyr Ala Asp Ile Leu Glu Phe Leu Val Gly Arg
325 330 335

Trp Lys Val Asp Lys Leu Met Gly Leu Ser Ala Glu Gly Gln Lys Ala
340 345 350

Gln Asp Tyr Val Cys Arg Leu Pro Pro Arg Ile Arg Arg Leu Glu Glu
355 360 365

Arg Ala Gln Gly Arg Ala Lys Glu Ala Thr Ile Val Pro Phe Ser Trp
370 375 380

Ile Phe Glu Arg Glu Val Gln Leu
385 390

<210> 202

<211> 365

<212> PRT

<213> Helinathus annus

<400> 202

Met Leu Ser Arg Gly Val Pro Thr Ala Thr Ala Thr Ala Tyr Asn Asp
1 5 10 15

Val Ile Lys Glu Thr Ala Arg Ser Val Thr Asn Ser Arg Ser Ile Asp
20 25 30

Ser Val Ser Ile Arg Arg Arg Asn Asn Gly Phe Val Ser Asn Ser Leu
35 40 45

Cys Arg Arg Val Ala Pro Val Met Ala Val Lys Val Asp Glu Gln Arg
50 55 60

Thr Gly Val Ala Val Asp Val Thr Glu Lys Arg Leu Ala Asp Arg Leu
65 70 75 80

Arg Met Gly Ser Leu Thr Glu Asp Gly Leu Ser Tyr Lys Glu Arg Phe
85 90 95

Ile Ile Arg Cys Tyr Glu Val Gly Ile Asn Lys Thr Ala Thr Val Glu
100 105 110

Thr Ile Ala Asn Leu Leu Gln Glu Val Gly Gly Asn His Ala Gln Ser
115 120 125

Val Gly Phe Ser Thr Asp Gly Phe Ala Thr Thr Thr Thr Met Arg Lys
130 135 140

Leu Asn Leu Ile Trp Val Thr Ser Arg Met His Ile Glu Ile Tyr Arg
145 150 155 160

Tyr Pro Ala Trp Ser Asp Val Val Glu Ile Glu Thr Trp Cys Gln Gly
165 170 175

Glu Gly Arg Ile Gly Thr Arg Arg Asp Trp Ile Ile Lys Asp His Ala
180 185 190

Asn Gly Glu Val Ile Gly Arg Ala Thr Ser Lys Trp Val Met Met Asn
195 200 205

Ser Glu Thr Arg Arg Leu Gln Lys Val Asn Asp Asp Ile Arg Asp Glu
210 215 220

Tyr Leu Ile Phe Cys Pro Lys Thr Leu Arg Leu Ala Phe Pro Glu Glu
225 230 235 240

Asn Asn Asn Ser Leu Lys Lys Ile Ala Lys Leu Glu Asp Pro Ala Glu
245 250 255
Cys Ser Thr Leu Gly Leu Val Pro Arg Arg Ala Asp Leu Asp Met Asn
260 265 270
Lys His Val Asn Asn Val Thr Tyr Ile Gly Trp Val Leu Glu Ser Ile
275 280 285

Pro Gln Glu Val Ile Asp Thr His Glu Leu Gln Thr Ile Thr Leu Asp
290 295 300
Tyr Arg Arg Glu Cys Gln His Asp Asp Val Val Asp Ser Leu Thr Ser
305 310 315 320
Ser Glu Ser Pro Ala Val Asn Gly Cys Ala Ser Gly Glu Asn Leu Ser
325 330 335
Lys Phe Leu His Leu Leu Arg Ser Ser Gly Glu Gly Leu Glu Leu Asn
340 345 350

Arg Gly Arg Thr Glu Trp Arg Lys Lys Pro Ala Lys Lys
355 360 365
<210> 203
<211> 362
<212> PRT
<213> Brassica rapa
<400> 203
Met Leu Lys Leu Ser Cys Asn Ala Thr Asp Lys Leu Gln Thr Leu Phe
1 5 10 15
Ser His Ser His Gln Pro Asp Pro Ala His Arg Arg Thr Val Ser Ser
20 25 30
Val Ser Cys Ser His Leu Arg Lys Pro Val Leu Asp Pro Leu Arg Ala

35 40 45
Ile Val Ser Ala Asp Gln Gly Ser Val Ile Arg Ala Glu Gln Gly Leu
50 55 60
Gly Ser Leu Ala Asp Gln Leu Arg Leu Gly Ser Leu Thr Glu Asp Gly
65 70 75 80
Leu Ser Tyr Lys Glu Lys Phe Ile Val Arg Ser Tyr Glu Val Gly Ser

85 90 95

Asn Lys Thr Ala Thr Val Glu Thr Val Ala Asn Leu Leu Gln Glu Val

100 105 110

Gly Cys Asn His Ala Gln Ser Val Gly Phe Ser Thr Asp Gly Phe Ala

115 120 125

Thr Thr Pro Thr Met Arg Lys Leu His Leu Ile Trp Val Thr Ala Arg

130 135 140

Met His Ile Glu Ile Tyr Lys Tyr Pro Ala Trp Gly Asp Val Val Glu

145 150 155 160

Ile Glu Thr Trp Cys Gln Ser Glu Gly Arg Ile Gly Thr Arg Arg Asp

165 170 175

Trp Ile Leu Lys Asp Val Ala Thr Gly Glu Val Thr Gly Arg Ala Thr

180 185 190

Ser Lys Trp Val Met Met Asn Gln Asp Thr Arg Arg Leu Gln Lys Val

195 200 205

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

210 215 220

Arg Leu Ala Phe Pro Glu Glu Asn Asn Arg Ser Leu Lys Lys Ile Pro

225 230 235 240

Lys Leu Glu Asp Pro Ala Gln Tyr Ser Met Ile Gly Leu Lys Pro Arg

245 250 255

Arg Ala Asp Leu Asp Met Asn Gln His Val Asn Asn Val Thr Tyr Ile

260 265 270

Gly Trp Val Leu Glu Ser Ile Pro Gln Glu Ile Val Asp Thr His Glu

275 280 285

Leu Gln Val Ile Thr Leu Asp Tyr Arg Arg Glu Cys Gln Gln Asp Asp

290 295 300

Val Val Asp Ser Leu Thr Thr Thr Thr Ser Glu Ile Gly Gly Thr Asn

305 310 315 320

Gly Ser Ala Ser Ser Gly Thr Gln Gly Gln Asn Asp Ser Gln Phe Leu

325 330 335

His Leu Leu Arg Leu Ser Gly Asp Gly Gln Glu Ile Asn Arg Gly Thr
340 345 350

Thr Leu Trp Arg Lys Lys Pro Ser Asn Leu
355 360

<210> 204
<211> 369
<212> PRT
<213> Jatropha curcas
<400> 204

Met Met Leu Lys Val Pro Cys Thr Ala Pro Asp Gln Ile Gln Ser Leu
1 5 10 15

Ala Pro Gln Cys Arg Phe Leu Asn Gln His Arg Ser Ser Ser His Phe
20 25 30

Thr Arg Arg Arg Ala Asn Ala Val Val Tyr Ser Ser Pro Val Ala Lys
35 40 45

Ile Ser Thr Gln Pro Val Val Ser Asp Gly Val Thr Lys Val Glu Ser

50 55 60

Gly Cys Glu Ser Leu Ala Asp Arg Leu Arg Leu Gly His Leu Thr Glu
65 70 75 80

Asp Gly Phe Ser Tyr Lys Glu Lys Phe Ile Val Arg Ser Tyr Glu Val
85 90 95

Gly Ile Asn Lys Thr Ala Thr Val Glu Thr Ile Ala Asn Phe Leu Gln
100 105 110

Glu Val Gly Cys Asn His Ala Gln Ser Val Gly Phe Ser Thr Asp Gly

115 120 125

Phe Ala Thr Thr Pro Thr Met Arg Lys Leu His Leu Ile Trp Val Thr
130 135 140

Ala Arg Met His Ile Glu Ile Tyr Lys Tyr Pro Ala Trp Ser Asp Val
145 150 155 160

Val Glu Ile Glu Thr Trp Cys Gln Ser Glu Gly Arg Ile Gly Thr Arg
165 170 175

Arg Asp Trp Ile Leu Lys Asp Tyr Ala Thr Gly Gln Val Ile Gly Arg

180 185 190

Ala Thr Ser Lys Trp Val Met Met Asn Gln Asp Thr Arg Arg Leu Gln

195 200 205

Lys Val Thr Asp Asp Val Arg Asp Glu Tyr Leu Val Phe Cys Pro Arg

210 215 220

Glu Leu Arg Leu Ala Phe Pro Glu Glu Asn Asn Arg Ser Leu Lys Lys

225 230 235 240

Ile Ser Lys Leu Glu Asp Thr Thr Gln Tyr Ser Lys Leu Gly Leu Val

245 250 255

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

260 265 270

Tyr Ile Gly Trp Val Leu Glu Ser Met Pro Gln Glu Ile Ile Asp Thr

275 280 285

His Glu Leu Gln Thr Ile Thr Leu Asp Tyr Arg Arg Glu Cys Gln His

290 295 300

Asp Asp Val Val Asp Ser Leu Thr Ser Ala Glu Ser Leu Glu Gly Ala

305 310 315 320

Gly Leu Asp Leu His Ala Thr Asn Gly Ser Ala Thr Ala Ile Ala Gly

325 330 335

Glu Gln Asp Ser Arg Asn Phe Leu His Leu Leu Arg Phe Ser Ser Asp

340 345 350

Gly Leu Glu Ile Asn Arg Gly Arg Thr Glu Trp Arg Lys Lys Pro Ser

355 360 365

Arg

<210> 205

<211> 229

<212> PRT

<213> Zea mays

<400> 205

Met Arg Lys Leu Gly Leu Ile Trp Val Thr Asn Arg Met His Ile Glu
1 5 10 15
Ile Tyr Lys Tyr Pro Ala Trp Gly Asp Val Val Glu Ile Glu Thr Trp
20 25 30
Cys Gln Glu Asp Gly Lys Ile Gly Thr Arg Arg Asp Trp Ile Leu Lys
35 40 45
Asp Leu Cys Thr Gly Glu Val Thr Gly Arg Ala Thr Ser Lys Trp Val
50 55 60

Met Met Asn Gln Asn Thr Arg Arg Leu Gln Arg Val Ser Asp Asp Val
65 70 75 80
Arg Asp Glu Val Phe Ile His Cys Pro Lys Thr Pro Arg Leu Ala Phe
85 90 95
Pro Glu Glu Asn Asn Gly Ser Leu Lys Lys Ile Pro Asn Leu Ser Asp
100 105 110
Pro Ala Gln Tyr Ser Arg Leu Gly Leu Val Pro Arg Arg Ala Asp Leu
115 120 125

Asp Met Asn Gln His Val Asn Asn Val Thr Tyr Ile Gly Trp Val Leu
130 135 140
Glu Ser Ile Pro Gln Asp Ile Ile Asp Thr His Glu Leu Gln Thr Ile
145 150 155 160
Thr Leu Asp Tyr Arg Arg Glu Cys Gln His Asp Asp Ile Val Asp Ser
165 170 175
Leu Thr Tyr Val Glu Glu Gly Glu Glu Arg Ser Met Asn Gly Ser Ala
180 185 190

Ser Ser Val Pro His Thr Glu Gln Arg Arg Gln Phe Leu His Cys Leu
195 200 205
Arg Phe Ala Ala Asn Gly Asp Glu Ile Asn Arg Gly Arg Thr Val Trp
210 215 220
Arg Lys Leu Ala Arg
225
<210> 206
<211> 366

<212> PRT

<213> Zea mays

<400> 206

Met Leu Arg Cys His Thr Pro Pro Gln Cys Ala Arg Ala Pro Leu Arg
1 5 10 15

His His Gly Arg Trp Glu Ser Pro Pro Ala Ala Ala Pro Ala Val Val
20 25 30

Val Arg Cys Ala Arg Gly Ala Pro Gln Val Ser Gly Ile Glu Ala Ala
35 40 45

Ser Pro Gly His Ala Ala Val Thr Ala Ala Leu Ala Lys Ala Glu Gly
50 55 60

Gly Asp Ala Arg Pro Ser Leu Ala Glu Arg Leu Arg Leu Gly Asn Leu
65 70 75 80

Leu Glu Asp Gly Leu Ser Tyr Lys Glu Ser Phe Ile Val Arg Cys Tyr
85 90 95

Glu Val Gly Ile Asn Lys Thr Ala Thr Val Glu Thr Ile Ala Asn Leu
100 105 110

Leu Gln Glu Val Gly Cys Asn His Ala Gln Ser Val Gly Phe Ser Thr
115 120 125

Asp Gly Phe Ala Thr Thr Thr Thr Met Arg Lys Leu Gly Leu Ile Trp
130 135 140

Val Thr Asn Arg Met His Ile Glu Ile Tyr Lys Tyr Pro Ala Trp Gly
145 150 155 160

Asp Phe Val Glu Ile Glu Thr Trp Cys Gln Glu Asp Gly Arg Ile Gly
165 170 175

Thr Arg Arg Asp Trp Ile Leu Lys Asp Leu Ala Asn Gly Glu Val Ile
180 185 190

Gly Arg Ala Thr Ser Lys Trp Val Met Met Asn Gln Asn Thr Arg Arg
195 200 205

Leu Gln Arg Val Ser Asp Asp Val Arg Asp Glu Val Phe Met His Cys
210 215 220

Pro Lys Ala Pro Arg Leu Ala Phe Pro Glu Glu Asn Asn Gly Ser Leu
225 230 235 240
Lys Lys Ile Pro Asn Leu Ser Asp Pro Ala Glu Tyr Ser Arg Leu Gly
245 250 255
Leu Val Pro Arg Arg Ala Asp Leu Asp Met Asn Gln His Val Asn Asn
260 265 270

Val Thr Tyr Ile Gly Trp Val Leu Glu Ser Ile Pro Gln Asp Ile Ile
275 280 285
Asp Thr His Glu Leu Gln Thr Ile Thr Leu Asp Tyr Arg Arg Glu Cys
290 295 300
Gln Gln Asp Asp Ile Val Asp Ser Leu Thr Cys Ile Glu Glu Gly Glu
305 310 315 320
Glu Lys Ser Met Asn Gly Ser Ala Ser Ala Ala Ala Pro His Lys Glu
325 330 335

Glu Arg Gln Gln Phe Leu His Cys Leu Arg Phe Ala Ala Asn Gly His
340 345 350
Glu Ile Asn Arg Gly Arg Thr Val Trp Arg Lys Leu Ala Arg
355 360 365

<210> 207
<211> 381
<212> PRT
<213> Carthamus tinctorius
<400> 207

Met Gly Ala Gly Gly Arg Met Asn Val Ala Val Ser Gly Ser Glu Lys
1 5 10 15
Pro His Ala Phe Lys Arg Val Pro Val Ser Lys Pro Pro Phe Glu Leu
20 25 30
Ser Asp Leu Lys Lys Ala Val Pro Pro His Cys Phe Lys Arg Ser Leu
35 40 45
Val Arg Ser Phe Ala Ala Leu Phe Arg Asp Ile Ile Ile Val Thr Ala
50 55 60
Leu Tyr Tyr Leu Ala Ala Thr Ile Ile Pro Val Leu Pro Lys Pro Leu

65 70 75 80

Thr Tyr Val Ala Trp Pro Leu Tyr Trp Phe Phe Gln Gly Ala Tyr Leu

85 90 95

Met Gly Leu Trp Val Ile Gly His Glu Cys Gly His His Gly Phe Ser

100 105 110

Glu Tyr Gln Trp Leu Asp Asp Thr Val Gly Phe Ile Val His Ser Phe

115 120 125

Ile Leu Thr Pro Tyr Phe Gly Phe Lys Tyr Ser His Arg Thr His His

130 135 140

Ala Asn Thr Asn Ser Ile Glu Tyr Asp Glu Val Trp Ile Pro Lys Arg

145 150 155 160

Lys Ser Asp Lys Leu Tyr Ser Glu Val Leu Asn Asn Pro Leu Gly Ser

165 170 175

Phe Ile Val Phe Val Phe Lys Ile Val Leu Gly Phe Pro Leu Tyr Phe

180 185 190

Val Phe Asn Leu Tyr Gly Arg Lys Tyr Glu Lys Gly Ile Thr Ser His

195 200 205

Phe Tyr Pro Tyr Ser Pro Ile Phe Asn Asp Ser Glu Arg Phe Gln Ile

210 215 220

Phe Leu Thr Asp Leu Gly Val Phe Gly Thr Leu Tyr Gly Val Tyr Arg

225 230 235 240

Leu Ala Leu Ile Lys Gly Thr Glu Trp Val Ile Asn Phe Tyr Gly Met

245 250 255

Pro Ile Leu Phe Met Ser Gly Ser Phe Ile Leu Leu Thr Tyr Leu His

260 265 270

His Thr His Pro Ser Ile Pro His Tyr Asp Ser Thr Glu Trp Asp Trp

275 280 285

Leu Arg Gly Ala Leu Ala Thr Val Asp Arg Asn Phe Gly Phe Leu Asn

290 295 300

His Ala Phe His Asp Ala Thr Arg Thr His Ala Val His His Leu Phe

305 310 315 320

Pro Thr Ile Pro His Tyr His Thr Phe Glu Ala Arg Gln Ala Val Met
325 330 335
Pro Ile Leu Gly Asp Tyr Tyr Lys Tyr Asp Asp Thr Pro Ile Leu Gln

340 345 350
Ala Ile Trp Arg Glu Thr Lys Asp Cys Thr Phe Ile Glu Pro Glu Glu
355 360 365
Val Asn Gly Glu Lys Lys Gly Ile Tyr Trp Phe Tyr Lys
370 375 380

<210> 208
<211> 383
<212> PRT
<213> Gossypium hirsutum
<400> 208

Met Gly Ala Gly Gly Arg Met Ser Val Pro Thr Ser Pro Lys Lys Pro
1 5 10 15

Glu Phe Asn Ser Leu Lys Arg Val Pro Tyr Ser Lys Pro Pro Phe Thr
20 25 30
Leu Ser Glu Ile Lys Lys Ala Ile Pro Pro His Cys Phe Gln Arg Ser
35 40 45
Val Leu Arg Ser Phe Ser Tyr Leu Leu Tyr Asp Phe Ile Leu Ala Ser
50 55 60
Leu Phe Tyr His Val Ala Thr Asn Tyr Phe Pro Asn Leu Pro Gln Ala
65 70 75 80

Leu Ser Asn Val Ala Trp Pro Leu Tyr Trp Ala Met Gln Gly Cys Ile
85 90 95
Leu Thr Gly Val Trp Val Ile Ala His Glu Cys Gly His His Ala Phe
100 105 110
Ser Asp Tyr Gln Trp Leu Asp Asp Thr Val Gly Leu Ile Leu His Ser
115 120 125
Ser Leu Leu Val Pro Tyr Phe Ser Trp Lys Tyr Ser His Arg Arg His
130 135 140

His Ser Asn Thr Gly Ser Leu Glu Arg Asp Glu Val Phe Val Pro Lys
145 150 155 160
Lys Lys Ser Gly Leu Arg Trp Trp Ala Lys His Phe Asn Asn Pro Pro
165 170 175
Gly Arg Phe Leu Ser Ile Thr Ile Gln Leu Thr Leu Gly Trp Pro Leu
180 185 190
Tyr Leu Ala Phe Asn Val Ala Gly Arg Pro Tyr Asp Arg Phe Ala Cys
195 200 205

His Tyr Asp Pro Tyr Gly Pro Ile Phe Ser Asp Arg Glu Arg Leu Gln
210 215 220
Ile Tyr Ile Ser Asp Ala Gly Val Leu Ala Val Ala Tyr Ala Leu Tyr
225 230 235 240
Arg Leu Val Leu Ala Lys Gly Val Gly Trp Val Ile Ser Val Tyr Gly
245 250 255
Val Pro Leu Leu Val Val Asn Ala Phe Leu Val Met Ile Thr Tyr Leu
260 265 270

Gln His Thr His Pro Ser Leu Pro His Tyr Asp Ser Ser Glu Trp Asp
275 280 285
Trp Met Arg Gly Ala Leu Ser Thr Val Asp Arg Asp Tyr Gly Ile Leu
290 295 300
Asn Lys Val Phe His Asn Ile Thr Asp Thr His Val Ala His His Leu
305 310 315 320
Phe Ser Thr Met Pro His Tyr His Ala Met Val Ala Thr Lys Ala Ile
325 330 335

Lys Pro Ile Leu Gly Glu Tyr Tyr Gln Phe Asp Gly Met Pro Val Tyr
340 345 350
Lys Ala Ile Trp Arg Glu Ala Lys Glu Cys Leu Tyr Val Glu Pro Asp
355 360 365
Glu Gly Asp Lys Asp Lys Gly Val Phe Trp Phe Arg Asn Lys Leu
370 375 380
<210> 209
<211> 383

<212> PRT

<213> Glycine max

<400> 209

Met Gly Ala Gly Gly Arg Thr Asp Val Pro Pro Ala Asn Arg Lys Ser

1 5 10 15

Glu Val Asp Pro Leu Lys Arg Val Pro Phe Glu Lys Pro Pro Phe Ser

20 25 30

Leu Ser Gln Ile Lys Lys Val Ile Pro Pro His Cys Phe Gln Arg Ser

35 40 45

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

50 55 60

Cys Leu Tyr Tyr Val Ala Thr His Tyr Phe His Leu Leu Pro Ser Pro

65 70 75 80

Leu Ser Phe Leu Ala Trp Pro Ile Tyr Trp Ala Val Gln Gly Cys Ile

85 90 95

Leu Thr Gly Val Trp Val Ile Ala His Glu Cys Gly His His Ala Phe

100 105 110

Ser Asp Tyr Gln Leu Leu Asp Asp Ile Val Gly Leu Val Leu His Ser

115 120 125

Gly Leu Leu Val Pro Tyr Phe Ser Trp Lys Tyr Ser His Arg Arg His

130 135 140

His Ser Asn Thr Gly Ser Leu Glu Arg Asp Glu Val Phe Val Pro Lys

145 150 155 160

Gln Lys Ser Cys Ile Lys Trp Tyr Ser Lys Tyr Leu Asn Asn Pro Pro

165 170 175

Gly Arg Val Leu Thr Leu Ala Val Thr Leu Thr Leu Gly Trp Pro Leu

180 185 190

Tyr Leu Ala Leu Asn Val Ser Gly Arg Pro Tyr Asp Arg Phe Ala Cys

195 200 205

His Tyr Asp Pro Tyr Gly Pro Ile Tyr Ser Asp Arg Glu Arg Leu Gln

210 215 220

Ile Tyr Ile Ser Asp Ala Gly Val Leu Ala Val Cys Tyr Gly Leu Phe
225 230 235 240
Arg Leu Ala Met Ala Lys Gly Leu Ala Trp Val Val Cys Val Tyr Gly
245 250 255
Val Pro Leu Leu Val Val Asn Gly Phe Leu Val Leu Ile Thr Phe Leu

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

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

<210> 210
<211> 387
<212> PRT
<213> Zea mays

<400> 210
Met Gly Ala Gly Gly Arg Met Thr Glu Lys Glu Arg Glu Lys Gln Glu
1 5 10 15
Gln Leu Ala Arg Ala Thr Gly Gly Ala Ala Met Gln Arg Ser Pro Val
20 25 30
Glu Lys Pro Pro Phe Thr Leu Gly Gln Ile Lys Lys Ala Ile Pro Pro
35 40 45
His Cys Phe Glu Arg Ser Val Leu Lys Ser Phe Ser Tyr Val Val His

50 55 60

Asp Leu Val Ile Ala Ala Ala Leu Leu Tyr Phe Ala Leu Ala Ile Ile
65 70 75 80
Pro Ala Leu Pro Ser Pro Leu Gly Tyr Ala Ala Trp Pro Leu Tyr Trp
85 90 95
Ile Ala Gln Gly Cys Val Cys Thr Gly Val Trp Val Ile Ala His Glu
100 105 110
Cys Gly His His Ala Phe Ser Asp Tyr Ser Leu Leu Asp Asp Val Val
115 120 125

Gly Leu Val Leu His Ser Ser Leu Met Val Pro Tyr Phe Ser Trp Lys
130 135 140
Tyr Ser His Arg Arg His His Ser Asn Thr Gly Ser Leu Glu Arg Asp
145 150 155 160
Glu Val Phe Val Pro Lys Lys Lys Glu Ala Leu Pro Trp Tyr Thr Pro
165 170 175
Tyr Val Tyr Asn Asn Pro Val Gly Arg Leu Ala His Ile Val Val Gln
180 185 190

Leu Thr Leu Gly Trp Pro Leu Tyr Leu Ala Thr Asn Ala Ser Gly Arg
195 200 205
Pro Tyr Pro Arg Phe Ala Cys His Phe Asp Pro Tyr Gly Pro Ile Tyr
210 215 220
Asn Asp Arg Glu Arg Ala Gln Ile Phe Val Ser Asp Ala Gly Val Ala
225 230 235 240
Ala Val Ala Phe Gly Leu Tyr Lys Leu Ala Ala Ala Phe Gly Val Trp
245 250 255

Trp Val Val Arg Val Tyr Ala Val Pro Leu Leu Ile Val Asn Ala Trp
260 265 270
Leu Val Leu Ile Thr Tyr Leu Gln His Thr His Pro Ser Leu Pro His
275 280 285
Tyr Asp Ser Ser Glu Trp Asp Trp Leu Arg Gly Ala Leu Ala Thr Met
290 295 300

Asp Arg Asp Tyr Gly Ile Leu Asn Arg Val Phe His Asn Ile Thr Asp
305 310 315 320

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

Lys Lys Phe
385

<210> 211
<211> 368
<212> PRT
<213> Prototheca moriformis
<400> 211
Met Ala Ile Lys Thr Asn Arg Gln Pro Val Glu Lys Pro Pro Phe Thr
1 5 10 15
Ile Gly Thr Leu Arg Lys Ala Ile Pro Ala His Cys Phe Glu Arg Ser
20 25 30
Ala Leu Arg Ser Ser Met Tyr Leu Ala Phe Asp Ile Ala Val Met Ser
35 40 45
Leu Leu Tyr Val Ala Ser Thr Tyr Ile Asp Pro Ala Pro Val Pro Thr

50 55 60
Trp Val Lys Tyr Gly Val Met Trp Pro Leu Tyr Trp Phe Phe Gln Gly
65 70 75 80
Ala Phe Gly Thr Gly Val Trp Val Cys Ala His Glu Cys Gly His Gln
85 90 95
Ala Phe Ser Ser Ser Gln Ala Ile Asn Asp Gly Val Gly Leu Val Phe
100 105 110

His Ser Leu Leu Leu Val Pro Tyr Tyr Ser Trp Lys His Ser His Arg

115 120 125

Arg His His Ser Asn Thr Gly Cys Leu Asp Lys Asp Glu Val Phe Val

130 135 140

Pro Pro His Arg Ala Val Ala His Glu Gly Leu Glu Trp Glu Glu Trp

145 150 155 160

Leu Pro Ile Arg Met Gly Lys Val Leu Val Thr Leu Thr Leu Gly Trp

165 170 175

Pro Leu Tyr Leu Met Phe Asn Val Ala Ser Arg Pro Tyr Pro Arg Phe

180 185 190

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

195 200 205

Ile Glu Val Val Ile Ser Asp Leu Ala Leu Val Ala Val Leu Ser Gly

210 215 220

Leu Ser Val Leu Gly Arg Thr Met Gly Trp Ala Trp Leu Val Lys Thr

225 230 235 240

Tyr Val Val Pro Tyr Leu Ile Val Asn Met Trp Leu Val Leu Ile Thr

245 250 255

Leu Leu Gln His Thr His Pro Ala Leu Pro His Tyr Phe Glu Lys Asp

260 265 270

Trp Asp Trp Leu Arg Gly Ala Met Ala Thr Val Asp Arg Ser Met Gly

275 280 285

Pro Pro Phe Met Asp Asn Ile Leu His His Ile Ser Asp Thr His Val

290 295 300

Leu His His Leu Phe Ser Thr Ile Pro His Tyr His Ala Glu Glu Ala

305 310 315 320

Ser Ala Ala Ile Arg Pro Ile Leu Gly Lys Tyr Tyr Gln Ser Asp Ser

325 330 335

Arg Trp Val Gly Arg Ala Leu Trp Glu Asp Trp Arg Asp Cys Arg Tyr

340 345 350

Val Val Pro Asp Ala Pro Glu Asp Asp Ser Ala Leu Trp Phe His Lys

355 360 365

<210> 212

<211> 368

<212> PRT

<213> Prototheca moriformis

<400> 212

Met Ala Ile Lys Thr Asn Arg Gln Pro Val Glu Lys Pro Pro Phe Thr
1 5 10 15

Ile Gly Thr Leu Arg Lys Ala Ile Pro Ala His Cys Phe Glu Arg Ser
20 25 30

Ala Leu Arg Ser Ser Met Tyr Leu Ala Phe Asp Ile Ala Val Met Ser
35 40 45

Leu Leu Tyr Val Ala Ser Thr Tyr Ile Asp Pro Ala Pro Val Pro Thr
50 55 60

Trp Val Lys Tyr Gly Ile Met Trp Pro Leu Tyr Trp Phe Phe Gln Gly
65 70 75 80

Ala Phe Gly Thr Gly Val Trp Val Cys Ala His Glu Cys Gly His Gln
85 90 95

Ala Phe Ser Ser Ser Gln Ala Ile Asn Asp Gly Val Gly Leu Val Phe
100 105 110

His Ser Leu Leu Leu Val Pro Tyr Tyr Ser Trp Lys His Ser His Arg
115 120 125

Arg His His Ser Asn Thr Gly Cys Leu Asp Lys Asp Glu Val Phe Val
130 135 140

Pro Pro His Arg Ala Val Ala His Glu Gly Leu Glu Trp Glu Glu Trp
145 150 155 160

Leu Pro Ile Arg Met Gly Lys Val Leu Val Thr Leu Thr Leu Gly Trp
165 170 175

Pro Leu Tyr Leu Met Phe Asn Val Ala Ser Arg Pro Tyr Pro Arg Phe
180 185 190

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

195 200 205

Ile Glu Val Val Ile Ser Asp Leu Ala Leu Val Ala Val Leu Ser Gly

210 215 220

Leu Ser Val Leu Gly Arg Thr Met Gly Trp Ala Trp Leu Val Lys Thr

225 230 235 240

Tyr Val Val Pro Tyr Met Ile Val Asn Met Trp Leu Val Leu Ile Thr

245 250 255

Leu Leu Gln His Thr His Pro Ala Leu Pro His Tyr Phe Glu Lys Asp

260 265 270

Trp Asp Trp Leu Arg Gly Ala Met Ala Thr Val Asp Arg Ser Met Gly

275 280 285

Pro Pro Phe Met Asp Ser Ile Leu His His Ile Ser Asp Thr His Val

290 295 300

Leu His His Leu Phe Ser Thr Ile Pro His Tyr His Ala Glu Glu Ala

305 310 315 320

Ser Ala Ala Ile Arg Pro Ile Leu Gly Lys Tyr Tyr Gln Ser Asp Ser

325 330 335

Arg Trp Val Gly Arg Ala Leu Trp Glu Asp Trp Arg Asp Cys Arg Tyr

340 345 350

Val Val Pro Asp Ala Pro Glu Asp Asp Ser Ala Leu Trp Phe His Lys

355 360 365

<210> 213

<211> 383

<212> PRT

<213> Brassica napa

<400> 213

Met Val Val Ala Met Asp Gln Arg Ser Asn Val Asn Gly Asp Ser Gly

1 5 10 15

Ala Arg Lys Glu Glu Gly Phe Asp Pro Ser Ala Gln Pro Pro Phe Lys

20 25 30

Ile Gly Asp Ile Arg Ala Ala Ile Pro Lys His Cys Trp Val Lys Ser

35 40 45

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

50 55 60

Ala Leu Ala Met Ala Ala Val Tyr Phe Asp Ser Trp Phe Leu Trp Pro

65 70 75 80

Leu Tyr Trp Val Ala Gln Gly Thr Leu Phe Trp Ala Ile Phe Val Leu

85 90 95

Gly His Asp Cys Gly His Gly Ser Phe Ser Asp Ile Pro Leu Leu Asn

100 105 110

Ser Val Val Gly His Ile Leu His Ser Phe Ile Leu Val Pro Tyr His

115 120 125

Gly Trp Arg Ile Ser His Arg Thr His His Gln Asn His Gly His Val

130 135 140

Glu Asn Asp Glu Ser Trp Val Pro Leu Pro Glu Lys Leu Tyr Lys Asn

145 150 155 160

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

165 170 175

Leu Ala Tyr Pro Ile Tyr Leu Trp Tyr Arg Ser Pro Gly Lys Glu Gly

180 185 190

Ser His Phe Asn Pro Tyr Ser Ser Leu Phe Ala Pro Ser Glu Arg Lys

195 200 205

Leu Ile Ala Thr Ser Thr Thr Cys Trp Ser Ile Met Leu Ala Thr Leu

210 215 220

Val Tyr Leu Ser Phe Leu Val Asp Pro Val Thr Val Leu Lys Val Tyr

225 230 235 240

Gly Val Pro Tyr Ile Ile Phe Val Met Trp Leu Asp Ala Val Thr Tyr

245 250 255

Leu His His His Gly His Asp Glu Lys Leu Pro Trp Tyr Arg Gly Lys

260 265 270

Glu Trp Ser Tyr Leu Arg Gly Gly Leu Thr Thr Ile Asp Arg Asp Tyr

275 280 285

Gly Ile Phe Asn Asn Ile His His Asp Ile Gly Thr His Val Ile His

290 295 300

His Leu Phe Pro Gln Ile Pro His Tyr His Leu Val Asp Ala Thr Arg

305 310 315 320

Ala Ala Lys His Val Leu Gly Arg Tyr Tyr Arg Glu Pro Lys Thr Ser

325 330 335

Gly Ala Ile Pro Ile His Leu Val Glu Ser Leu Val Ala Ser Ile Lys

340 345 350

Lys Asp His Tyr Val Ser Asp Thr Gly Asp Ile Val Phe Tyr Glu Thr

355 360 365

Asp Pro Asp Leu Tyr Val Tyr Ala Ser Asp Lys Ser Lys Ile Asn

370 375 380

<210> 214

<211> 448

<212> PRT

<213> Camelina sativa

<400> 214

Met Ala Asn Leu Val Leu Ser Glu Cys Gly Ile Arg Pro Leu Pro Arg

1 5 10 15

Ile Tyr Thr Thr Pro Arg Ser Asn Phe Val Ser Asn Asn Asn Lys Pro

20 25 30

Ile Phe Lys Phe Arg Pro Phe Thr Ser Tyr Lys Thr Ser Ser Ser Pro

35 40 45

Leu Ala Cys Ser Arg Asp Gly Phe Gly Lys Asn Trp Ser Leu Asn Val

50 55 60

Ser Val Pro Leu Thr Thr Thr Thr Pro Ile Val Asp Glu Ser Pro Leu

65 70 75 80

Lys Glu Glu Glu Glu Glu Lys Gln Arg Phe Asp Pro Gly Ala Pro Pro

85 90 95

Pro Phe Asn Leu Ala Asp Ile Arg Ala Ala Ile Pro Lys His Cys Trp

100 105 110

Val Lys Asn Pro Trp Lys Ser Met Ser Tyr Val Leu Arg Asp Val Ala

115 120 125

Ile Val Phe Ala Leu Ala Ala Gly Ala Ser Tyr Leu Asn Asn Trp Ile
130 135 140
Val Trp Pro Leu Tyr Trp Leu Ala Gln Gly Thr Met Phe Trp Ala Leu
145 150 155 160
Phe Val Leu Gly His Asp Cys Gly His Gly Ser Phe Ser Asn Asn Pro
165 170 175
Arg Leu Asn Asn Val Val Gly His Leu Leu His Ser Ser Ile Leu Val
180 185 190

Pro Tyr His Gly Trp Arg Ile Ser His Arg Thr His His Gln Asn His
195 200 205
Gly His Val Glu Asn Asp Glu Ser Trp His Pro Met Ser Glu Lys Ile
210 215 220
Tyr Gln Ser Leu Asp Lys Pro Thr Arg Phe Phe Arg Phe Thr Leu Pro
225 230 235 240
Leu Val Met Leu Ala Tyr Pro Phe Tyr Leu Trp Ala Arg Ser Pro Gly
245 250 255

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

Val Thr Tyr Leu His His His Gly His Glu Asp Lys Leu Pro Trp Tyr
325 330 335
Arg Gly Lys Glu Trp Ser Tyr Leu Arg Gly Gly Leu Thr Thr Leu Asp
340 345 350
Arg Asp Tyr Gly Val Ile Asn Asn Ile His His Asp Ile Gly Thr His
355 360 365

Val Ile His His Leu Phe Pro Gln Ile Pro His Tyr His Leu Val Glu
370 375 380

Ala Thr Glu Ala Val Lys Pro Val Leu Gly Lys Tyr Tyr Arg Glu Pro
385 390 395 400
Asp Lys Ser Gly Pro Leu Pro Leu His Leu Leu Gly Ile Leu Ala Lys
405 410 415
Ser Ile Lys Glu Asp His Tyr Val Ser Asp Glu Gly Asp Val Val Tyr
420 425 430
Tyr Lys Ala Asp Pro Asn Met Tyr Gly Glu Ile Lys Val Gly Ala Asp
435 440 445

<210> 215
<211> 448
<212> PRT
<213> Camelina sativa
<400> 215
Met Ala Asn Leu Val Leu Ser Glu Cys Gly Ile Arg Pro Leu Pro Arg
1 5 10 15
Ile Tyr Thr Thr Pro Arg Ser Asn Phe Val Ser Asn Asn Asn Lys Pro
20 25 30
Ile Phe Lys Phe Arg Pro Leu Thr Ser Tyr Lys Thr Ser Ser Pro Leu
35 40 45
Phe Cys Ser Arg Asp Gly Phe Gly Arg Asn Trp Ser Leu Asn Val Ser

50 55 60
Val Pro Leu Ala Thr Thr Thr Pro Ile Val Asp Glu Ser Pro Leu Glu
65 70 75 80
Glu Glu Glu Glu Glu Glu Lys Gln Arg Phe Asp Pro Gly Ala Pro Pro
85 90 95
Pro Phe Asn Leu Ala Asp Ile Arg Ala Ala Ile Pro Lys His Cys Trp
100 105 110
Val Lys Asn Pro Trp Lys Ser Met Ser Tyr Val Leu Arg Asp Val Ala

115 120 125

Ile Val Phe Ala Leu Ala Ala Gly Ala Ala Tyr Leu Asn Asn Trp Ile
130 135 140
Val Trp Pro Leu Tyr Trp Leu Ala Gln Gly Thr Met Phe Trp Ala Leu
145 150 155 160
Phe Val Leu Gly His Asp Cys Gly His Gly Ser Phe Ser Asn Asn Pro
165 170 175
Arg Leu Asn Asn Val Val Gly His Leu Leu His Ser Ser Ile Leu Val

180 185 190
Pro Tyr His Gly Trp Arg Ile Ser His Arg Thr His His Gln Asn His
195 200 205
Gly His Val Glu Asn Asp Glu Ser Trp His Pro Met Ser Glu Lys Ile
210 215 220
Tyr Gln Ser Leu Asp Lys Pro Thr Arg Phe Phe Arg Phe Thr Leu Pro
225 230 235 240
Leu Val Met Leu Ala Tyr Pro Phe Tyr Leu Trp Ala Arg Ser Pro Gly

245 250 255
Lys Lys Gly Ser His Tyr His Pro Glu Ser Asp Leu Phe Leu Pro Lys
260 265 270
Glu Lys Thr Asp Val Leu Thr Ser Thr Ala Cys Trp Thr Ala Met Ala
275 280 285
Ala Leu Leu Ile Cys Leu Asn Phe Val Val Gly Pro Val Gln Met Leu
290 295 300
Lys Leu Tyr Gly Ile Pro Tyr Trp Ile Asn Val Met Trp Leu Asp Phe

305 310 315 320
Val Thr Tyr Leu His His His Gly His Glu Asp Lys Leu Pro Trp Tyr
325 330 335
Arg Gly Lys Glu Trp Ser Tyr Leu Arg Gly Gly Leu Thr Thr Leu Asp
340 345 350
Arg Asp Tyr Gly Val Ile Asn Asn Ile His His Asp Ile Gly Thr His
355 360 365
Val Ile His His Leu Phe Pro Gln Ile Pro His Tyr His Leu Val Glu

370 375 380

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

385 390 395 400

Asp Lys Ser Gly Pro Leu Pro Leu His Leu Leu Gly Ile Leu Ala Lys

405 410 415

Ser Ile Lys Glu Asp His Tyr Val Ser Asp Glu Gly Asp Val Val Tyr

420 425 430

Tyr Lys Ala Asp Pro Asn Met Tyr Gly Glu Ile Lys Val Gly Ala Asp

435 440 445

<210> 216

<211> 453

<212> PRT

<213> Glycine max

<400> 216

Met Ala Thr Trp Val Leu Ser Glu Cys Ser Leu Lys Pro Leu Ala Pro

1 5 10 15

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

20 25 30

Val Gly Phe Leu Glu Thr Asn Lys Val Leu Glu Gly Ser Lys Phe Gln

35 40 45

Pro Leu Arg Cys Asn Leu Arg Glu Arg Asn Trp Gly Leu Lys Val Ser

50 55 60

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

65 70 75 80

Asp Val Ile Asn Gly Ser Asn Gly Val Glu His Glu Lys Leu Pro Glu

85 90 95

Phe Asp Pro Gly Ala Pro Pro Pro Phe Asn Leu Ala Asp Ile Arg Ala

100 105 110

Ala Ile Pro Lys His Cys Trp Val Lys Asp Pro Leu Lys Ser Met Ser

115 120 125

Tyr Val Val Arg Asp Val Ile Ala Val Phe Gly Leu Ala Ser Ala Ala

130 135 140

Ala Tyr Leu Asn Asn Trp Leu Val Trp Pro Leu Tyr Trp Ala Ala Gln
145 150 155 160
Gly Thr Met Phe Trp Ala Leu Phe Val Leu Gly His Asp Cys Gly His
165 170 175

Gly Ser Phe Ser Asn Asn Pro Lys Leu Asn Ser Val Val Gly His Leu
180 185 190
Leu His Ser Ser Ile Leu Val Pro Tyr His Gly Trp Arg Ile Ser His
195 200 205
Arg Thr His His Gln His His Gly His Val Glu Asn Asp Glu Ser Trp
210 215 220
His Pro Leu Pro Glu Lys Leu Phe Arg Ser Leu Asp Thr Val Thr Arg
225 230 235 240

Met Leu Arg Phe Thr Ala Pro Phe Pro Leu Leu Ala Tyr Pro Val Tyr
245 250 255
Leu Trp Gly Arg Ser Pro Gly Lys Thr Gly Ser His Phe Asp Pro Ser
260 265 270
Ser Asp Leu Phe Val Pro Asn Glu Arg Lys Asp Val Ile Thr Ser Thr
275 280 285
Ala Cys Trp Ala Ala Met Leu Gly Leu Leu Val Gly Leu Gly Phe Val
290 295 300

Met Gly Pro Ile Gln Leu Leu Lys Leu Tyr Gly Val Pro Tyr Val Ile
305 310 315 320
Phe Val Met Trp Leu Asp Leu Val Thr Tyr Leu His His His Gly His
325 330 335
Glu Asp Lys Leu Pro Trp Tyr Arg Gly Lys Glu Trp Ser Tyr Leu Arg
340 345 350
Gly Gly Leu Thr Thr Leu Asp Arg Asp Tyr Gly Leu Ile Asn Asn Ile
355 360 365

His His Asp Ile Gly Thr His Val Ile His His Leu Phe Pro Gln Ile
370 375 380
Pro His Tyr His Leu Val Glu Ala Thr Glu Ala Ala Lys Pro Val Phe

385 390 395 400

Gly Lys Tyr Tyr Arg Glu Pro Lys Lys Ser Ser Pro Leu Pro Phe His

405 410 415

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

420 425 430

Asp Lys Gly Asp Val Val Tyr Tyr Gln Thr Asp Ser Glu Ile Asn Gly

435 440 445

Ser Ser Lys Ser Glu

450

<210> 217

<211> 451

<212> PRT

<213> Vernicia fordii

<400> 217

Met Ala Val Trp Ala Leu Ser Asp Cys Gly Ile Arg Pro Leu Pro Lys

1 5 10 15

Ile Tyr Ser Lys Pro Arg Leu Ala Phe Thr Ser Asn Asn Pro Gln Pro

20 25 30

Thr Lys Pro Pro Ile Pro Arg Pro Glu Leu Arg Asn Gly Ser Ser Phe

35 40 45

Lys Leu Ser Ser Ala Gly Leu Arg Asp Lys Trp Val Val Asn Leu Ser

50 55 60

Ala Pro Leu Arg Gly Gly Leu Val Glu Glu Asp Glu Asp Asn Phe Glu

65 70 75 80

Gly Asn Arg Val Ile Ser Ile Asp Glu Ser Gly Gly Glu Asp Phe Asp

85 90 95

Pro Asp Ala Pro Pro Pro Asp Thr Leu Ala Asp Ile Arg Ala Ala Ile

100 105 110

Pro Lys His Arg Trp Thr Lys Asn Pro Glu Ile Ser Leu Ser Tyr Val

115 120 125

Val Arg Asn Val Ala Val Val Ser Gly Leu Ala Ala Val Ala Ala Tyr

130 135 140

Phe Asn Asp Trp Ala Phe Trp Ser Pro Tyr Trp Ile Cys Gln Gly Thr
145 150 155 160

Met Leu Ser Ala Leu Phe Val Leu Gly His Asp Cys Gly His Gly Ser
165 170 175
Phe Ser Asn Asn Pro Asn Leu Asp Ser Val Val Gly His Leu Leu His
180 185 190
Tyr Ser Ile Leu Val Pro Tyr His Gly Trp Arg Ile Ser Tyr Arg Thr
195 200 205
His Tyr Arg His His Gln His Ala Glu Asn Asp Glu Ser Trp His Pro
210 215 220

Leu Ser Glu Lys Ile Tyr Lys Gly Leu Asn Asn Val Thr Arg Thr Leu
225 230 235 240
Arg Leu Ser Leu Pro Phe Ser Leu Leu Ala Tyr Pro Ser Tyr Leu Trp
245 250 255
Ser Arg Ser Pro Gly Lys Lys Gly Phe His Phe His Pro Ser Ser Asp
260 265 270
Leu Phe Val Pro Ser Glu Lys Lys Asp Val Ile Thr Ser Thr Val Cys
275 280 285

Trp Thr Ala Met Leu Ala Leu Leu Val Gly Ser Ser Phe Val Met Gly
290 295 300
Leu Val Gln Leu Leu Lys Pro Tyr Gly Ile Pro Tyr Leu Ile Phe Val
305 310 315 320
Met Trp Leu Asp Leu Val Thr Tyr Leu His His His Gly His Glu Asn
325 330 335
Lys Ile Pro Trp Tyr Arg Gly Glu Glu Trp Ser Tyr Leu Arg Gly Arg
340 345 350

Leu Thr Thr Val Gly Arg Asp Tyr Gly Trp Phe Asn Lys Ile His His
355 360 365
Asp Ile Gly Ile His Val Ile His His Leu Phe Pro Gln Ile Pro His
370 375 380
Tyr His Leu Ile Glu Ala Thr Glu Ala Ala Lys Pro Val Leu Gly Lys

385 390 395 400

Tyr Tyr Arg Glu Pro Lys Lys Ser Gly Pro Leu Pro Phe His Leu Leu

405 410 415

Gly Ser Leu Ile Arg Ser Met Lys Glu Asp His Tyr Val Ser Asp Ala

420 425 430

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

435 440 445

Cys Leu Ser

450

<210> 218

<211> 506

<212> PRT

<213> Ricinus communis

<400> 218

Met Ser Ala Val Met Leu Ala Arg Ala Thr Asn Val Ala Ala Asn Asn

1 5 10 15

Asn Val Thr Ser Arg Thr Gly Ser His Glu Arg Pro His Thr Pro Ala

20 25 30

Ala Ala Ala His Phe Asn Leu Ser Leu Cys Leu Leu Ser Pro Phe Lys

35 40 45

Gly Ser Pro Leu Gln Leu Glu Gly Ile Leu Leu Met Lys Lys Ala Asp

50 55 60

Ser Val Asn Ala Phe Asp Ala Leu Asp Gly Asp Ala Asp Phe Asp Pro

65 70 75 80

Ser Asp Pro Pro Pro Phe Lys Ile Ala Glu Ile Arg Ala Ala Ile Pro

85 90 95

Lys His Cys Trp Ile Lys Asn Pro Trp Arg Ser Leu Ser Tyr Val Leu

100 105 110

Arg Asp Ala Phe Val Val Leu Ala Leu Ala Ala Ala Ala Leu Phe Leu

115 120 125

Asp Ser Trp Cys Phe Trp Pro Leu Tyr Trp Ala Ala Gln Gly Thr Met

130 135 140

```
Phe Trp Ala Ile Phe Val Leu Gly His Asp Cys Gly His Gly Ser Phe

145                 150                 155                 160
Ser Asp Ser Thr Leu Leu Asn Asn Val Val Gly His Leu Leu His Ser
                165                 170                 175
Ser Ile Leu Val Pro Tyr His Gly Trp Arg Ile Ser His Arg Thr His
            180                 185                 190
His Gln Asn His Gly Asn Val Asp Lys Asp Glu Ser Trp Val Pro Leu
        195                 200                 205
Pro Glu Lys Ile Tyr Asn Ser Leu Asp Val Ser Thr Arg Ile Met Arg

    210                 215                 220
Phe Thr Val Pro Leu Pro Leu Leu Ala Tyr Pro Ile Tyr Leu Trp Thr
225                 230                 235                 240
Arg Gly Pro Gly Lys Asp Gly Ser His Phe Ser Pro Tyr Ser Asp Leu
                245                 250                 255
Phe Ala Pro Asn Glu Arg Arg Ala Val Leu Thr Ser Thr Val Cys Trp
            260                 265                 270
Ile Leu Met Ile Leu Leu Leu Leu Tyr Ser Cys Phe Val Phe Gly Ser

        275                 280                 285
Val Gln Val Leu Lys Leu Tyr Gly Val Pro Tyr Ile Ile Phe Ile Met
    290                 295                 300
Trp Leu Asp Phe Val Thr Tyr Leu His His His Gly His Glu Gln Lys
305                 310                 315                 320
Leu Pro Trp Tyr Arg Gly Lys Glu Trp Ser Tyr Leu Arg Gly Gly Leu
                325                 330                 335
Thr Thr Val Asp Arg Asp Tyr Gly Trp Ile Asn Asn Ile His His Asp

            340                 345                 350
Ile Gly Thr His Val Ile His His Leu Phe Pro Gln Ile Pro His Tyr
        355                 360                 365
His Leu Val Glu Ala Thr Lys Ala Ala Lys Pro Val Leu Gly Lys Tyr
    370                 375                 380
Tyr Arg Glu Pro Arg Lys Ser Gly Pro Phe Pro Leu His Leu Phe Glu
```

385 390 395 400

Asn Leu Val Lys Ser Ile Ser Gln Asp His Tyr Val Ser Asp Ile Gly

405 410 415

Asp Val Val Tyr Tyr Gln Thr Asp Pro Glu Leu Tyr Lys Lys Leu Ala

420 425 430

Lys Thr Lys Asp Val Leu Glu Asp Leu Glu Ser Cys Asp Ala Leu Gly

435 440 445

Ile Leu Lys Cys Lys Asn Cys Ser Ala Lys Ser Gly Ser Ser Tyr Ser

450 455 460

Ala Leu Gln Phe Leu Glu Leu Lys Gly Arg Ala Ala Gly Ala Glu Lys

465 470 475 480

Gly Arg Glu Trp Lys Ile Leu Val Phe Glu Ile Leu Thr Leu His His

485 490 495

Arg Ile Ile Ala Gln Met Ser Phe Met Gly

500 505

<210> 219

<211> 392

<212> PRT

<213> Linum usitatissimum

<400> 219

Met Ser Pro Pro Asn Ser Met Gly Met Glu Ala Ala His Pro Thr Gly

1 5 10 15

Asn Gly Asn Gly Val Ala Val Met Asn Gly Ala Ser Ala Asn Lys Pro

20 25 30

Asp Phe Asp Pro Ser Glu Ala Pro Pro Phe Lys Ile Ala Asp Ile Arg

35 40 45

Ala Ala Ile Pro Pro His Cys Trp Val Lys Asn Pro Trp Arg Ser Leu

50 55 60

Ser Tyr Val Leu Arg Asp Ala Val Val Ile Leu Ala Phe Ala Ala Ala

65 70 75 80

Ala Leu Lys Leu Asp Leu Trp Ala Val Trp Pro Leu Tyr Trp Ile Ala

85 90 95

Gln Gly Thr Met Phe Trp Ala Val Phe Val Leu Gly His Asp Cys Gly

100 105 110

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

115 120 125

Ile Leu His Ser Ala Ile Leu Val Pro Tyr His Gly Trp Arg Ile Ser

130 135 140

His Lys Thr His His Gln Asn His Gly Asn Val Glu Lys Asp Glu Ser

145 150 155 160

Trp Val Pro Leu Pro Glu Lys Val Tyr Lys Thr Leu Asp Thr Ser Thr

165 170 175

Lys Phe Met Arg Phe Thr Ile Pro Leu Pro Met Phe Ala Tyr Pro Ile

180 185 190

Tyr Leu Trp Thr Arg Ser Pro Gly Lys Lys Gly Ser His Phe Asn Pro

195 200 205

Tyr Ser Asp Leu Phe Ala Pro Gln Glu Arg Lys Ser Val Leu Ile Ser

210 215 220

Thr Ile Ser Trp Ile Ser Met Val Leu Ile Leu Leu Tyr Ala Ser Phe

225 230 235 240

Leu Phe Gly Phe Leu Thr Val Phe Lys Val Tyr Thr Val Pro Tyr Leu

245 250 255

Ile Phe Val Ala Trp Leu Asp Met Val Thr Tyr Leu His His His Gly

260 265 270

His Glu Glu Lys Leu Pro Trp Tyr Arg Gly Gln Glu Trp Ser Tyr Leu

275 280 285

Arg Gly Gly Leu Thr Thr Val Asp Arg Asp Tyr Gly Ile Ile Asn Asn

290 295 300

Ile His His Asp Ile Gly Thr His Val Ile His His Leu Phe Pro Gln

305 310 315 320

Ile Pro His Tyr His Leu Val Glu Ala Thr Lys Ala Ala Lys Ser Val

325 330 335

Leu Gly Lys Tyr Tyr Arg Glu Pro Lys Lys Ser Gly Pro Phe Pro Phe

340 345 350

His Leu Phe Asp Asn Leu Val Arg Ser Leu Gly Glu Asp His Tyr Val

355 360 365

Ser Asp Ala Gly Asp Val Val Phe Tyr Gln Ser Asp Pro Glu Ile Phe

370 375 380

Lys Phe Ser Lys Ser Lys Ser Ala

385 390

<210> 220

<211> 315

<212> PRT

<213> Prototheca moriformis

<400> 220

Met Phe Trp Ala Leu Phe Val Val Gly His Asp Cys Gly His Gln Ser

1 5 10 15

Phe Ser Thr Ser Lys Arg Leu Asn Asp Ala Val Gly Leu Phe Val His

20 25 30

Ser Ile Ile Gly Val Pro Tyr His Gly Trp Arg Ile Ser His Arg Thr

35 40 45

His His Asn Asn His Gly His Val Glu Asn Asp Glu Ser Trp Tyr Pro

50 55 60

Pro Thr Glu Ser Gly Leu Lys Ala Met Thr Asp Met Gly Arg Gln Gly

65 70 75 80

Arg Phe His Phe Pro Ser Met Leu Phe Val Tyr Pro Phe Tyr Leu Phe

85 90 95

Trp Arg Ser Pro Gly Lys Thr Gly Ser His Phe Ser Pro Ala Thr Asp

100 105 110

Leu Phe Ala Leu Trp Glu Ala Pro Leu Ile Arg Thr Ser Asn Ala Cys

115 120 125

Gln Leu Ala Trp Leu Gly Ala Leu Ala Ala Gly Thr Trp Ala Leu Gly

130 135 140

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

195 200 205
Val His His Asp Ile Gly Thr His Val Val His His Leu Phe Pro Gln
210 215 220
Ile Pro His Tyr Asn Leu Cys Arg Ala Thr Lys Ala Ala Lys Lys Val
225 230 235 240
Leu Gly Pro Tyr Tyr Arg Glu Pro Glu Arg Cys Pro Leu Gly Leu Leu
245 250 255
Pro Val His Leu Leu Ala Pro Leu Leu Arg Ser Leu Gly Gln Asp His

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

<210> 221
<211> 452
<212> PRT
<213> Prototheca moriformis
<400> 221
Met Ser Ile Gln Phe Ala Leu Arg Ala Ala Tyr Ile Lys Gly Thr Cys

1 5 10 15
Gln Arg Leu Ser Gly Arg Gly Ala Ala Leu Gly Leu Ser Arg Asp Trp
20 25 30
Thr Pro Gly Trp Thr Leu Pro Arg Cys Trp Pro Ala Ser Ala Ala Ala

35 40 45

Thr Ala Pro Pro Arg Ala Arg His Gln Glu Arg Ala Ile His Leu Thr

50 55 60

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

65 70 75 80

Ala Leu Pro Ser Asn Ala Pro Gly Leu Val Met Ala Ser Gln Ala Asn

85 90 95

Tyr Phe Arg Val Arg Leu Leu Pro Glu Gln Glu Glu Gly Glu Leu Glu

100 105 110

Ser Trp Ser Pro Asn Val Arg His Thr Thr Leu Leu Cys Lys Pro Arg

115 120 125

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

130 135 140

Ile Val Thr Ala Ile Asp Pro Val Asn Met Thr Val His Ala Pro Pro

145 150 155 160

Phe Asp Pro Leu Pro Ala Thr Arg Phe Leu Val Ala Gly Glu Ala Ala

165 170 175

Asp Met Asp Ile Thr Val Val Leu Asn Lys Ala Asp Leu Val Pro Glu

180 185 190

Glu Glu Ser Ala Ala Leu Ala Gln Glu Val Ala Ser Trp Gly Pro Val

195 200 205

Val Leu Thr Ser Thr Leu Thr Gly Arg Gly Leu Gln Glu Leu Glu Arg

210 215 220

Gln Leu Gly Ser Thr Thr Ala Val Leu Ala Gly Pro Ser Gly Ala Gly

225 230 235 240

Lys Ser Ser Ile Ile Asn Ala Leu Ala Arg Ala Ala Arg Glu Arg Pro

245 250 255

Ser Asp Ala Ser Val Ser Asn Val Pro Glu Glu Gln Val Val Gly Glu

260 265 270

Asp Gly Arg Ala Leu Ala Asn Pro Pro Pro Phe Thr Leu Ala Asp Ile

275 280 285

Arg Asn Ala Ile Pro Lys Asp Cys Phe Arg Lys Ser Ala Ala Lys Ser
290 295 300
Leu Ala Tyr Leu Gly Asp Leu Ser Ile Thr Gly Met Ala Val Leu Ala
305 310 315 320
Tyr Lys Ile Asn Ser Pro Trp Leu Trp Pro Leu Tyr Trp Phe Ala Gln

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

385 390 395 400
Leu His His His Gly Pro Ser Asp Pro Arg Glu Glu Met Pro Trp Tyr
405 410 415
Arg Gly Arg Glu Trp Ser Tyr Met Arg Gly Gly Leu Thr Thr Ile Asp
420 425 430
Arg Asp Tyr Gly Leu Phe Asn Lys Val His His Asp Ile Gly Thr His
435 440 445
Val Val His His
450

<210> 222

<211> 6756
<212> DNA
<213> Artificial Sequence
<220><223> synthesized construct
<400> 222

gctcttcgcc gccgccactc ctgctcgagc gcgcccgcgc gtgcgccgcc agcgccttgg 60
ccttttcgcc gcgctcgtgc gcgtcgctga tgtccatcac caggtccatg aggtctgcct 120
tgcgccggct gagccactgc ttcgtccggg cggccaagag gagcatgagg gaggactcct 180
ggtccagggt cctgacgtgg tcgcggctct gggagcgggc cagcatcatc tggctctgcc 240

gcaccgaggc cgcctccaac tggtcctcca gcagccgcag tcgccgccga ccctggcaga 300

ggaagacagg tgaggggggt atgaattgta cagaacaacc acgagccttg tctaggcaga 360

atccctacca gtcatggctt tacctggatg acggcctgcg aacagctgtc cagcgaccct 420

cgctgccgcc gcttctcccg cacgcttctt tccagcaccg tgatggcgcg agccagcgcc 480

gcacgctggc gctgcgcttc gccgatctga ggacagtcgg ggaactctga tcagtctaaa 540

ccccccttgcg cgttagtgtt gccatccttt gcagaccggt gagagccgac ttgttgtgcg 600

ccacccccca caccacctcc tcccagacca attctgtcac ctttttggcg aaggcatcgg 660

cctcggcctg cagagaggac agcagtgccc agccgctggg ggttggcgga tgcacgctca 720

ggtaccctttt cttgcgctat gacacttcca gcaaaaggta gggcgggctg cgagacggct 780

tcccggcgct gcatgcaaca ccgatgatgc ttcgaccccc cgaagctcct tcggggctgc 840

atgggcgctc cgatgccgct ccagggcgag cgctgtttaa atagccaggc ccccgattgc 900

aaagacatta tagcgagcta ccaaagccat attcaaacac ctagatcact accacttcta 960

cacaggccac tcgagcttgt gatcgcactc cgctaagggg gcgcctcttc ctcttcgttt 1020

cagtcacaac ccgcaaactc tagaatatca atgctgctgc aggccttcct gttcctgctg 1080

gccggcttcg ccgccaagat cagcgcctcc atgacgaacg agacgtccga ccgccccctg 1140

gtgcacttca cccccaacaa gggctggatg aacgacccca acggcctgtg gtacgacgag 1200

aaggacgcca agtggcacct gtacttccag tacaacccga acgacaccgt ctgggggacg 1260

cccttgttct ggggccacgc cacgtccgac gacctgacca actgggagga ccagcccatc 1320

gccatcgccc cgaagcgcaa cgactccggc gccttctccg gctccatggt ggtggactac 1380

aacaacacct ccggcttctt caacgacacc atcgacccgc gccagcgctg cgtggccatc 1440

tggacctaca acaccccgga gtccgaggag cagtacatct cctacagcct ggacggcggc 1500

tacaccttca ccgagtacca gaagaacccc gtgctggccg ccaactccac ccagttccgc 1560

gacccgaagg tcttctggta cgagccctcc cagaagtgga tcatgaccgc ggccaagtcc 1620

caggactaca agatcgagat ctactcctcc gacgacctga agtcctggaa gctggagtcc 1680

gcgttcgcca acgagggctt cctcggctac cagtacgagt gccccggcct gatcgaggtc 1740

cccaccgagc aggaccccag caagtcctac tgggtgatgt tcatctccat caaccccggc 1800

gccccggccg gcggctcctt caaccagtac ttcgtcggca gcttcaacgg cacccacttc 1860

gaggccttcg acaaccagtc ccgcgtggtg gacttcggca aggactacta cgccctgcag 1920

accttcttca acaccgaccc gacctacggg agcgccctgg gcatcgcgtg ggcctccaac 1980

tgggagtact ccgccttcgt gcccaccaac ccctggcgct cctccatgtc cctcgtgcgc 2040

aagttctccc tcaacaccga gtaccaggcc aacccggaga cggagctgat caacctgaag 2100
gccgagccga tcctgaacat cagcaacgcc ggcccctgga gccggttcgc caccaacacc 2160
acgttgacga aggccaacag ctacaacgtc gacctgtcca acagcaccgg caccctggag 2220
ttcgagctgg tgtacgccgt caacaccacc cagacgatct ccaagtccgt gttcgcggac 2280
ctctccctct ggttcaaggg cctggaggac cccgaggagt acctccgcat gggcttcgag 2340
gtgtccgcgt cctccttctt cctggaccgc gggaacagca aggtgaagtt cgtgaaggag 2400
aacccctact tcaccaaccg catgagcgtg aacaaccagc ccttcaagag cgagaacgac 2460

ctgtcctact acaaggtgta cggcttgctg gaccagaaca tcctggagct gtacttcaac 2520
gacggcgacg tcgtgtccac caacacctac ttcatgacca ccgggaacgc cctgggctcc 2580
gtgaacatga cgacgggggt ggacaacctg ttctacatcg acaagttcca ggtgcgcgag 2640
gtcaagtgac aattggcagc agcagctcgg atagtatcga cacactctgg acgctggtcg 2700
tgtgatggac tgttgccgcc acacttgctg ccttgacctg tgaatatccc tgccgctttt 2760
atcaaacagc ctcagtgtgt ttgatcttgt gtgtacgcgc ttttgcgagt tgctagctgc 2820
ttgtgctatt tgcgaatacc acccccagca tccccttccc tcgtttcata tcgcttgcat 2880

cccaaccgca acttatctac gctgtcctgc tatccctcag cgctgctcct gctcctgctc 2940
actgcccctc gcacagcctt ggtttgggct ccgcctgtat tctcctggta ctgcaacctg 3000
taaaccagca ctgcaatgct gatgcacggg aagtagtggg atgggaacac aaatggagga 3060
tcccgcgtct cgaacagagc gcgcagagga acgctgaagg tctcgcctct gtcgcacctc 3120
agcgcggcat acaccacaat aaccacctga cgaatgcgct tggttcttcg tccattagcg 3180
aagcgtccgg ttcacacacg tgccacgttg gcgaggtggc aggtgacaat gatcggtgga 3240
gctgatggtc gaaacgttca cagcctaggg atatcgaatt cggccgacag gacgcgcgtc 3300

aaaggtgctg gtcgtgtatg ccctggccgg caggtcgttg ctgctgctgg ttagtgattc 3360
cgcaaccctg attttggcgt cttattttgg cgtggcaaac gctggcgccc gcgagccggg 3420
ccggcggcga tgcggtgccc cacggctgcc ggaatccaag ggaggcaaga gcgcccgggt 3480
cagttgaagg gctttacgcg caaggtacag ccgctcctgc aaggctgcgt ggtggaattg 3540
gacgtgcagg tcctgctgaa gttcctccac cgcctcacca gcggacaaag caccggtgta 3600
tcaggtccgt gtcatccact ctaaagaact cgactacgac ctactgatgg ccctagattc 3660
ttcatcaaaa acgcctgaga cacttgccca ggattgaaac tccctgaagg gaccaccagg 3720

ggccctgagt tgttccttcc ccccgtggcg agctgccagc caggctgtac ctgtgatcga 3780

ggctggcggg aaaataggct tcgtgtgctc aggtcatggg aggtgcagga cagctcatga 3840

aacgccaaca atcgcacaat tcatgtcaag ctaatcagct atttcctctt cacgagctgt 3900

aattgtccca aaattctggt ctaccggggg tgatccttcg tgtacgggcc cttccctcaa 3960

ccctaggtat gcgcgcatgc ggtcgccgcg caactcgcgc gagggccgag ggtttgggac 4020

gggccgtccc gaaatgcagt tgcacccgga tgcgtggcac ctttttgcg ataatttatg 4080

caatggactg ctctgcaaaa ttctggctct gtcgccaacc ctaggatcag cggcgtagga 4140

tttcgtaatc attcgtcctg atggggagct accgactacc ctaatatcag cccgactgcc 4200

tgacgccagc gtccactttt gtgcacacat tccattcgtg cccaagacat ttcattgtgg 4260

tgcgaagcgt ccccagttac gctcacctgt ttcccgacct ccttactgtt ctgtcgacag 4320

agcgggccca caggccggtc gcagccacta gtatggtggt ggccatggac cagcgctcca 4380

acgtgaacgg cgactccggc gcccgcaagg aggagggcgg gcgcgccatg gactacaagg 4440

accacgacgg cgactacaag gaccacgaca tcgactacaa ggacgacgac gacaagttcg 4500

acccctccgc ccagcccccc ttcaagatcg gcgacatccg cgccgccatc cccaagcact 4560

gctgggtgaa gtccccccctg cgctccatgt cctacgtgac ccgcgacatc ttcgccgtgg 4620

ccgccctggc catggccgcc gtgtacttcg actcctggtt cctgtggccc ctgtactggg 4680

tggcccaggg caccctgttc tgggccatct tcgtgctggg ccacgactgc ggccacggct 4740

ccttctccga catcccccctg ctgaactccg tggtgggcca catcctgcac tccttcatcc 4800

tggtgcccta ccacggctgg cgcatctccc accgcaccca ccaccagaac cacggccacg 4860

tggagaacga cgagtcctgg gtgcccctgc ccgagaagct gtacaagaac ctgccccact 4920

ccacccgcat gctgcgctac accgtgcccc tgcccatgct ggcctacccc atctacctgt 4980

ggtaccgctc ccccggcaag gagggctccc acttcaaccc ctactcctcc ctgttcgccc 5040

cctccgagcg caagctgatc gccacctcca ccacctgctg gtccatcatg ctggccaccc 5100

tggtgtacct gtccttcctg gtggaccccg tgaccgtgct gaaggtgtac ggcgtgccct 5160

acatcatctt cgtgatgtgg ctggacgccg tgacctacct gcaccaccac ggccacgacg 5220

agaagctgcc ctggtaccgc ggcaaggagt ggtcctacct gcgcggcggc ctgaccacca 5280

tcgaccgcga ctacggcatc ttcaacaaca tccaccacga catcggcacc cacgtgatcc 5340

accacctgtt cccccagatc ccccactacc acctggtgga cgccacccgc gccgccaagc 5400

acgtgctggg ccgctactac cgcgagccca agacctccgg cgccatcccc atccacctgg 5460

tggagtccct ggtggcctcc atcaagaagg accactacgt gtccgacacc ggcgacatcg 5520

tgttctacga gaccgacccc gacctgtacg tgtacgcctc cgacaagtcc aagatcaact 5580

agatcgatct taaggcagca gcagctcgga tagtatcgac acactctgga cgctggtcgt 5640

gtgatggact gttgccgcca cacttgctgc cttgacctgt gaatatccct gccgctttta 5700

tcaaacagcc tcagtgtgtt tgatcttgtg tgtacgcgct tttgcgagtt gctagctgct 5760

tgtgctattt gcgaatacca cccccagcat ccccttccct cgtttcatat cgcttgcatc 5820

ccaaccgcaa cttatctacg ctgtcctgct atccctcagc gctgctcctg ctcctgctca 5880

ctgcccctcg cacagccttg gtttgggctc cgcctgtatt ctcctggtac tgcaacctgt 5940

aaaccagcac tgcaatgctg atgcacggga agtagtggga tgggaacaca aatggaaagc 6000

ttaattaaga gctcttgttt tccagaagga gttgctcctt gagcctttca ttctcagcct 6060

cgataacctc caaagccgct ctaattgtgg agggggttcg aatttaaaag cttggaatgt 6120

tggttcgtgc gtctggaaca agcccagact tgttgctcac tgggaaaagg accatcagct 6180

ccaaaaaact tgccgctcaa accgcgtacc tctgctttcg cgcaatctgc cctgttgaaa 6240

tcgccaccac attcatattg tgacgcttga gcagtctgta attgcctcag aatgtggaat 6300

catctgcccc ctgtgcgagc ccatgccagg catgtcgcgg gcgaggacac ccgccactcg 6360

tacagcagac cattatgcta cctcacaata gttcataaca gtgaccatat ttctcgaagc 6420

tccccaacga gcacctccat gctctgagtg gccacccccc ggccctggtg cttgcggagg 6480

gcaggtcaac cggcatgggg ctaccgaaat ccccgaccgg atcccaccac ccccgcgatg 6540

ggaagaatct ctccccggga tgtgggccca ccaccagcac aacctgctgg cccaggcgag 6600

cgtcaaacca taccacacaa atatccttgg catcggccct gaattccttc tgccgctctg 6660

ctacccggtg cttctgtccg aagcaggggt tgctagggat cgctccgagt ccgcaaaccc 6720

ttgtcgcgtg gcggggcttg ttcgagcttg aagagc 6756

<210> 223

<211> 1437

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 223

actagtatgg ccaacctggt gctgtccgag tgcggcatcc gccccctgcc ccgcatctac 60

accacccccc gctccaactt cgtgtccaac aacaacaagc ccatcttcaa gttccgcccc 120

ctgacctcct acaagacctc ctccccccctg ttctgctccc gcgacggctt cggccgcaac 180

tggtccctga acgtgtccgt gcccctggcc accaccaccc ccatcgtgga cgagtccccc 240

ctggaggagg aggaggagga ggagaagcag cgcgggcgcg ccatggacta caaggaccac 300

gacggcgact acaaggacca cgacatcgac tacaaggacg acgacgacaa gttcgacccc 360

ggcgcccccc ccccttcaa cctggccgac atccgcgccg ccatccccaa gcactgctgg 420

gtgaagaacc cctggaagtc catgtcctac gtgctgcgcg acgtggccat cgtgttcgcc 480

ctggccgccg gcgccgccta cctgaacaac tggatcgtgt ggcccctgta ctggctggcc 540

cagggcacca tgttctgggc cctgttcgtg ctgggccacg actgcggcca cggctccttc 600

tccaacaacc cccgcctgaa caacgtggtg ggccacctgc tgcactcctc catcctggtg 660

ccctaccacg gctggcgcat ctcccaccgc acccaccacc agaaccacgg ccacgtggag 720

aacgacgagt cctggcaccc catgtccgag aagatctacc agtccctgga caagcccacc 780

cgcttcttcc gcttcaccct gcccctggtg atgctggcct acccccttcta cctgtgggcc 840

cgctcccccg gcaagaaggg ctcccactac caccccgagt ccgacctgtt cctgcccaag 900

gagaagaccg acgtgctgac ctccaccgcc tgctggaccg ccatggccgc cctgctgatc 960

tgcctgaact tcgtggtggg ccccgtgcag atgctgaagc tgtacggcat cccctactgg 1020

atcaacgtga tgtggctgga cttcgtgacc tacctgcacc accacggcca cgaggacaag 1080

ctgccctggt accgcggcaa ggagtggtcc tacctgcgcg gcggcctgac caccctggac 1140

cgcgactacg gcgtgatcaa caacatccac cacgacatcg gcacccacgt gatccaccac 1200

ctgttccccc agatccccca ctaccacctg gtggaggcca ccgaggccgt gaagcccgtg 1260

ctgggcaagt actaccgcga gcccgacaag tccggccccc tgcccctgca cctgctgggc 1320

atcctggcca agtccatcaa ggaggaccac tacgtgtccg acgagggcga cgtggtgtac 1380

tacaaggccg accccaacat gtacggcgag atcaaggtgg gcgccgacta gatcgat 1437

<210> 224

<211> 1269

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 224

actagtatgt ccccccccaa ctccatgtcc cccgccacca acggctccac caacggcgtg 60

gccatcaacg gcgccaagaa gctgctggac gggcgcgcca tggactacaa ggaccacgac 120

ggcgactaca aggaccacga catcgactac aaggacgacg acgacaagtt cgacccctcc 180

gccgcccccc ccttcaagat cgccgacatc cgcgccgcca tccccccccca ctgctgggtg 240

aagaacccct ggcgctccct gtcctacgtg ctgcgcgacc tgctggtgat cctgtccttc 300

gccgtggccg ccaccaagct ggactcctgg accgtgtggc ccctgtactg gatcgcccag 360

ggcaccatgt tctgggccgt gttcgtgctg ggccacgact gcggccacgg ctccttctcc 420

gactcctggc tgctgaacaa cgtgatgggc cacatcctgc actcctccat cctggtgccc 480

taccacggct ggcgcatctc ccacaagacc caccaccaga accacggcaa cgtggagaag 540

gacgagtcct gggtgcccct gcccgagaag gtgtacaagt ccctggacac cggcaccaag 600

ttcatgcgct tcaccatccc cctgcccatg ttcgcctacc ccatctacct gtggcgccgc 660

tcccccggca agaagggctc ccacttcaac ccctactccg acctgttcgc ccccaacgag 720

cgcacctccg tgatgatctc caccctgtgc tggaccgcca tggccctgct gctgtgctac 780

tcctccttca tctacggctt cctgcccgtg ttcaagatct acggcgtgcc ctacctgatc 840

ttcgtggcct ggctggacat ggtgacctac ctgcaccacc acggctacga gcagaagctg 900

ccctggtacc gcggcaagga gtggtcctac ctgcgcggcg gcctgaccac cgtggaccgc 960

gactacggcg tgatcaacaa catccaccac gacatcggca cccacgtgat ccaccacctg 1020

ttcccccaga tgccccacta ccacctggtg gaggccaccc aggccgccaa gcacgtgctg 1080

ggcaagtact accgcgagcc caagaagtcc ggccccttcc ccttccacct gttcggctac 1140

ctggtgcgct ccctgggcga ggaccactac gtgtccgaca ccggcgacgt ggtgttctac 1200

cagtccgacc cccacatccc caagttcccc acctccgcca ccaccaagtc caagtcctcc 1260

tagatcgat 1269

<210> 225

<211> 1266

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 225

actagtatgt ccccccccaa ctccatgtcc cccaccacca acggcaacgg cgtggccatg 60

aacggcgcca agaagcagct ggacgggcgc gccatggact acaaggacca cgacggcgac 120

tacaaggacc acgacatcga ctacaaggac gacgacgaca agttcgaccc ctccgccgcc 180

ccccccttca agatcgccga catccgcgcc gccatccccc cccactgctg ggtgaagaac 240

ccctggcgct ccctgtccta cgtgctgcgc gacctgctgg tgatcctgtc cttcgccgtg 300

gccgccgcca agctggactc ctggaccttc tggcccctgt actgggtggc ccagggcacc 360

atgttctggg ccgtgttcgt gctgggccac gactgcggcc acggctcctt ctccgacatc 420

tggctgctga acaacgtgat gggccacatc ctgcactcct ccatcctggt gccctaccac 480

ggctggcgca tctcccacaa gacccaccac cagaaccacg gcaacgtgga gaaggacgag 540

tcctgggtgc ccctgcccga gaaggtgtac aagtccctgg acacctccac caagttcatg 600

cgcttcacca tccccctgcc catgttcgcc taccccatct acctgtggac ccgctccccc 660

ggcaagaagg gctcccactt caacccctac tccgacctgt tcgcccccaa cgagcgcgcc 720

gccgtgctga tctccaccct gtgctggacc gccatggccc tgctgctgtg ctactcctcc 780

ttcatctacg gcttcgcccc cgtgctgaag atctacggcg tgccctacct gatcttcgtg 840

gcctggctgg acatggtgac ctacctgcac caccacggct acgagcagaa gctgccctgg 900

taccgcggca aggagtggtc ctacctgcgc ggcggcctga ccaccgtgga ccgcgactac 960

ggcgtgatca acaacatcca ccacgacatc ggcacccacg tgatccacca cctgttcccc 1020

cagatgcccc actaccacct ggtggaggcc acccaggccg ccaagcacgt gctgggcaag 1080

tactaccgcg agcccaagaa gtccggcccc ttcccccttcc acctgttcgg ctacctggtg 1140

cgctccctgg gcgaggacca ctacgtgtcc gacaccggcg acgtggtgtt ctaccagtcc 1200

gacccccaca tccccaagtt ccgcacctcc tccgccacca ccaagtccaa gtcctcctag 1260

atcgat 1266

<210> 226

<211> 5246

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 226

gctcttcgcc gccgccactc ctgctcgagc gcgcccgcgc gtgcgccgcc agcgccttgg 60

ccttttcgcc gcgctcgtgc gcgtcgctga tgtccatcac caggtccatg aggtctgcct 120

tgcgccggct gagccactgc ttcgtccggg cggccaagag gagcatgagg gaggactcct 180

ggtccagggt cctgacgtgg tcgcggctct gggagcgggc cagcatcatc tggctctgcc 240

gcaccgaggc cgcctccaac tggtcctcca gcagccgcag tcgccgccga ccctggcaga 300

ggaagacagg tgaggggtgt atgaattgta cagaacaacc acgagccttg tctaggcaga 360

atccctacca gtcatggctt tacctggatg acggcctgcg aacagctgtc cagcgaccct 420

cgctgccgcc gcttctcccg cacgcttctt tccagcaccg tgatggcgcg agccagcgcc 480

gcacgctggc gctgcgcttc gccgatctga ggacagtcgg ggaactctga tcagtctaaa 540

ccccccttgcg cgttagtgtt gccatccttt gcagaccggt gagagccgac ttgttgtgcg 600

ccacccccca caccacctcc tcccagacca attctgtcac ctttttggcg aaggcatcgg 660

cctcggcctg cagagaggac agcagtgccc agccgctggg ggttggcgga tgcacgctca 720

ggtacccttt cttgcgctat gacacttcca gcaaaaggta gggcgggctg cgagacggct 780

tcccggcgct gcatgcaaca ccgatgatgc ttcgaccccc cgaagctcct tcggggctgc 840

atgggcgctc cgatgccgct ccagggcgag cgctgtttaa atagccaggc ccccgattgc 900

aaagacatta tagcgagcta ccaaagccat attcaaacac ctagatcact accacttcta 960

cacaggccac tcgagcttgt gatcgcactc cgctaagggg gcgcctcttc ctcttcgttt 1020

cagtcacaac ccgcaaacgg cgcgccatgc tgctgcaggc cttcctgttc ctgctggccg 1080

gcttcgccgc caagatcagc gcctccatga cgaacgagac gtccgaccgc cccctggtgc 1140

acttcacccc caacaagggc tggatgaacg accccaacgg cctgtggtac gacgagaagg 1200

acgccaagtg gcacctgtac ttccagtaca acccgaacga caccgtctgg gggacgccct 1260

tgttctgggg ccacgccacg tccgacgacc tgaccaactg ggaggaccag cccatcgcca 1320

tcgccccgaa gcgcaacgac tccggcgcct tctccggctc catggtggtg gactacaaca 1380

acacctccgg cttcttcaac gacaccatcg acccgcgcca gcgctgcgtg gccatctgga 1440

cctacaacac cccggagtcc gaggagcagt acatctccta cagcctggac ggcggctaca 1500

ccttcaccga gtaccagaag aaccccgtgc tggccgccaa ctccacccag ttccgcgacc 1560

cgaaggtctt ctggtacgag ccctcccaga agtggatcat gaccgcggcc aagtcccagg 1620

actacaagat cgagatctac tcctccgacg acctgaagtc ctggaagctg gagtccgcgt 1680

tcgccaacga gggcttcctc ggctaccagt acgagtgccc cggcctgatc gaggtcccca 1740

ccgagcagga ccccagcaag tcctactggg tgatgttcat ctccatcaac cccggcgccc 1800

cggccggcgg ctccttcaac cagtacttcg tcggcagctt caacggcacc cacttcgagg 1860

ccttcgacaa ccagtcccgc gtggtggact tcggcaagga ctactacgcc ctgcagacct 1920

tcttcaacac cgacccgacc tacgggagcg ccctgggcat cgcgtgggcc tccaactggg 1980

agtactccgc cttcgtgccc accaacccct ggcgctcctc catgtccctc gtgcgcaagt 2040

tctccctcaa caccgagtac caggccaacc cggagacgga gctgatcaac ctgaaggccg 2100

agccgatcct gaacatcagc aacgccggcc cctggagccg gttcgccacc aacaccacgt 2160

tgacgaaggc caacagctac aacgtcgacc tgtccaacag caccggcacc ctggagttcg 2220

agctggtgta cgccgtcaac accacccaga cgatctccaa gtccgtgttc gcggacctct 2280

ccctctggtt caagggcctg gaggaccccg aggagtacct ccgcatgggc ttcgaggtgt 2340

ccgcgtcctc cttcttcctg gaccgcggga acagcaaggt gaagttcgtg aaggagaacc 2400

cctacttcac caaccgcatg agcgtgaaca accagccctt caagagcgag aacgacctgt 2460

cctactacaa ggtgtacggc ttgctggacc agaacatcct ggagctgtac ttcaacgacg 2520

gcgacgtcgt gtccaccaac acctacttca tgaccaccgg gaacgccctg ggctccgtga 2580

acatgacgac gggggtggac aacctgttct acatcgacaa gttccaggtg cgcgaggtca 2640

agtgacaatt ggcagcagca gctcggatag tatcgacaca ctctggacgc tggtcgtgtg 2700

atggactgtt gccgccacac ttgctgcctt gacctgtgaa tatccctgcc gcttttatca 2760

aacagcctca gtgtgtttga tcttgtgtgt acgcgctttt gcgagttgct agctgcttgt 2820

gctatttgcg aataccaccc ccagcatccc cttccctcgt ttcatatcgc ttgcatccca 2880

accgcaactt atctacgctg tcctgctatc cctcagcgct gctcctgctc ctgctcactg 2940

cccctcgcac agccttggtt tgggctccgc ctgtattctc ctggtactgc aacctgtaaa 3000

ccagcactgc aatgctgatg cacgggaagt agtgggatgg gaacacaaat ggaggatccc 3060

gcgtctcgaa cagagcgcgc agaggaacgc tgaaggtctc gcctctgtcg cacctcagcg 3120

cggcatacac cacaataacc acctgacgaa tgcgcttggt tcttcgtcca ttagcgaagc 3180

gtccggttca cacacgtgcc acgttggcga ggtggcaggt gacaatgatc ggtggagctg 3240

atggtcgaaa cgttcacagc ctagggatat cgaattcctt tcttgcgcta tgacacttcc 3300

agcaaaaggt agggcgggct gcgagacggc ttcccggcgc tgcatgcaac accgatgatg 3360

cttcgacccc ccgaagctcc ttcggggctg catgggcgct ccgatgccgc tccagggcga 3420

gcgctgttta aatagccagg cccccgattg caaagacatt atagcgagct accaaagcca 3480

tattcaaaca cctagatcac taccacttct acacaggcca ctcgagcttg tgatcgcact 3540

ccgctaaggg ggcgcctctt cctcttcgtt tcagtcacaa cccgcaaaca ctagtgggtc 3600

gccgcgctgc ctcgcgtccc ctggtggtgc acgccgtcgc cagcgaggcc ccgctgggcg 3660

tgccgccctc ggtgcagcgc ccctcccccg tggtctactc caagctggac aagcaacacc 3720

gcctgacgcc cgagcgcctg gagctggtgc agagcatggg gcagtttgcg gaggagaggt 3780

gtgtttgagg gttttggttg cccgtatcga ggtcctggtg gcgcgcatgg gggagaaggc 3840

gcctgtcccg ctgacccccc cggctaccct cccggcacct tccagggcgc gtacgggatc 3900

cctctcctcc acaaactgcc ccatgctctg caccagctcc aggcgctcgg gcgtcaggcg 3960

gtgctgcttg tccagcttgg agtagaccac gggggagggg cgctgcaccg agggcggaac 4020

gcccagcggg gcctcgctgg cgaccgcgcg caccaccagg ggacgcgagg cagcgcggcg 4080

acccaagctt gcagcagcag ctcggatagt atcgacacac tctggacgct ggtcgtgtga 4140

tggactgttg ccgccacact tgctgccttg acctgtgaat atccctgccg cttttatcaa 4200

acagcctcag tgtgtttgat cttgtgtgta cgcgcttttg cgagttgcta gctgcttgtg 4260

ctatttgcga ataccacccc cagcatcccc ttccctcgtt tcatatcgct tgcatcccaa 4320

ccgcaactta tctacgctgt cctgctatcc ctcagcgctg ctcctgctcc tgctcactgc 4380

ccctcgcaca gccttggttt gggctccgcc tgtattctcc tggtactgca acctgtaaac 4440
cagcactgca atgctgatgc acgggaagta gtgggatggg aacacaaatg gaaagctgga 4500
gctcttgttt tccagaagga gttgctcctt gagcctttca ttctcagcct cgataacctc 4560
caaagccgct ctaattgtgg agggggttcg aatttaaaag cttggaatgt tggttcgtgc 4620
gtctggaaca agcccagact tgttgctcac tgggaaaagg accatcagct ccaaaaaact 4680
tgccgctcaa accgcgtacc tctgctttcg cgcaatctgc cctgttgaaa tcgccaccac 4740
attcatattg tgacgcttga gcagtctgta attgcctcag aatgtggaat catctgcccc 4800

ctgtgcgagc ccatgccagg catgtcgcgg gcgaggacac ccgccactcg tacagcagac 4860
cattatgcta cctcacaata gttcataaca gtgaccatat ttctcgaagc tccccaacga 4920
gcacctccat gctctgagtg gccacccccc ggccctggtg cttgcggagg gcaggtcaac 4980
cggcatgggg ctaccgaaat ccccgaccgg atcccaccac ccccgcgatg ggaagaatct 5040
ctccccggga tgtgggccca ccaccagcac aacctgctgg cccaggcgag cgtcaaacca 5100
taccacacaa atatccttgg catcggccct gaattccttc tgccgctctg ctacccggtg 5160
cttctgtccg aagcaggggt tgctagggat cgctccgagt ccgcaaaccc ttgtcgcgtg 5220

gcggggcttg ttcgagcttg aagagc 5246

<210> 227

<211> 615

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 227

actagttgcc cgtgctgcac cccgtggaca agctgtggca gccgcaggac ttcctgcccg 60
accccgagtc gcccgacttc gaggaccagg tggcggagct gcgcgcgcgc gccaaggacc 120
tgcccgacga gtactttgtg gtgctggtgg gcgacatgat cacggaggag gcgctgccga 180
cctacatggc catgctcaac accttggacg gtgtgcgcga cgacacgggc gcggccgacc 240
acccgtgtgt gtttgagggt tttggttgcc cgtattgagg tcctggtggc gcgcatggag 300

gagaaggcgc ctgtcccgct gacccccccg gctaccctcc cggcaccttc cagggcgcgt 360
acgggatcca cgggtggtcg gccgcgcccg tgtcgtcgcg cacaccgtcc aaggtgttga 420
gcatggccat gtaggtcggc agcgcctcct ccgtgatcat gtcgcccacc agcaccacaa 480
agtactcgtc gggcaggtcc ttggcgcgcg cgcgcagctc cgccacctgg tcctcgaagt 540

cgggcgactc ggggtcgggc aggaagtcct gcggctgcca cagcttgtcc acggggtgca 600

gcacgggcaa agctt 615

<210> 228

<211> 615

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 228

actagtgcgc tggacgcggc agtgggtggc cgaggagaac cggcacggcg acctgctgaa 60

caagtactgt tggctgacgg ggcgcgtcaa catgcgggcc gtggaggtga ccatcaacaa 120

cctgatcaag agcggcatga acccgcagac ggacaacaac ccttacttgg gcttcgtcta 180

cacctccttc caggagcgcg cgaccaagta cagccacggc aacaccgcgc gccttgcggc 240

cgagcagtgt gtttgagggt tttggttgcc cgtatcgagg tcctggtggc gcgcatgggg 300

gagaaggcgc ctgtcccgct gacccccccg gctaccctcc cggcaccttc cagggcgcgt 360

acgggatcct gctcggccgc aaggcgcgcg gtgttgccgt ggctgtactt ggtcgcgcgc 420

tcctggaagg aggtgtagac gaagcccaag taagggttgt tgtccgtctg cgggttcatg 480

ccgctcttga tcaggttgtt gatggtcacc tccacggccc gcatgttgac gcgccccgtc 540

agccaacagt acttgttcag caggtcgccg tgccggttct cctcggccac ccactgccgc 600

gtccagcgca agctt 615

<210> 229

<211> 613

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 229

actagtgctg atcgccagcg acgagggccg gcacgagatc gcctacacgc gcatcgtgga 60

cgagttcttc cgcctcgacc ccgagggcgc cgtcgccgcc tacgccaaca tgatgcgcaa 120

gcagatcacc atgcccgcgc acctcatgga cgacatgggc cacggcgagg ccaacccggg 180

ccgcaacctc ttcgccgact tctccgcggt cgccgagaag atcgacgtct acgacgccga 240

ggactgtgtg tttgagggtt ttggttgccc gtatcgaggt cctggtggcg cgcatggggg 300

agaaggcgcc tgtcccgctg acccccccgg ctaccctccc ggcaccttcc agggcgcgta 360

cgggatccag tcctcggcgt cgtagacgtc gatcttctcg gcgaccgcgg agaagtcggc 420

gaagaggttg cggcccgggt tggcctcgcc gtggcccatg tcgtccatga ggtgcgcggg 480

catggtgatc tgcttgcgca tcatgttggc gtaggcggcg acggcgccct cggggtcgag 540

gcggaagaac tcgtccacga tgcgcgtgta ggcgatctcg tgccggccct cgtcgctggc 600

gatcagcaag ctt 613

<210> 230

<211> 1503

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 230

atggccaccg catccacttt ctcggcgttc aatgcccgct gcggcgacct gcgtcgctcg 60

gcgggctccg ggccccggcg cccagcgagg cccctccccg tgcgcgggcg cgcctccacc 120

cactccggca agaccatggc cgtggccctg cagcccaccc aggagatcac caccatcaag 180

aagcccccca ccaagcagcg ccgcgtggtg gtgaccggcc tgggcgtggt gacccccctg 240

ggccacgagc ccgacatctt ctacaacaac ctgctggacg gcgcctccgg catctccgag 300

atcgagacct tcgactgcgc cgagtacccc acccgcatcg ccggcgagat caagtccttc 360

tccaccgacg gctgggtggc ccccaagctg tccaagcgca tggacaagtt catgctgtac 420

atgctgaccg ccggcaagaa ggccctggtg gacggcggca tcaccgacga cgtgatggac 480

gagctgaaca aggagaagtg cggcgtgctg atcggctccg ccatgggcgg catgaaggtg 540

ttcaacgacg ccatcgaggc cctgcgcatc tcctacaaga agatgaaccc cttctgcgtg 600

cccttcgcca ccaccaacat gggctccgcc atgctggcca tggacctggg ctggatgggc 660

cccaactact ccatctccac cgcctgcgcc acctccaact tctgcatcct gaacgccgcc 720

aaccacatca tccgcggcga ggccgacgtg atgctgtgcg gcggctccga cgccgccatc 780

atccccatcg gcctgggcgg cttcgtggcc tgccgcgccc tgtcccagcg caacaccgac 840

cccaccaagg cctcccgccc ctgggacatc aaccgcgacg gcttcgtgat gggcgagggc 900

gccggcgtgc tgctgctgga ggagctggag cacgccaagg agcgcggcgc caccatctac 960

gccgagttcc tgggcggctc cttcacctgc gacgcctacc acgtgaccga gccccgcccc 1020

gacggcgccg gcgtgatcct gtgcatcgag aaggccctgg cccagtccgg cgtgtccaag 1080

gaggacgtga actacatcaa cgcccacgcc acctccaccc ccgccggcga cctgaaggag 1140

taccaggccc tgatgcactg cttcggccag aaccccgagc tgcgcgtgaa ctccaccaag 1200

tccatgatcg gccacctgct gggcgccgcc ggcggcgtgg aggccgtggc caccatccag 1260

gccatccgca ccggctgggt gcaccccaac atcaacctgg agaaccccga caacggcgtg 1320

gacgccaagg tgctggtggg ctccaagaag gagcgcctgg acgtgaaggc cgccctgtcc 1380

aactccttcg gcttcggcgg ccacaactcc tccatcatct tcgcccccta catggactac 1440

aaggaccacg acggcgacta caaggaccac gacatcgact acaaggacga cgacgacaag 1500

tga 1503

<210> 231

<211> 1530

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 231

atggccaccg catccacttt ctcggcgttc aatgcccgct gcggcgacct gcgtcgctcg 60

gcgggctccg ggccccggcg cccagcgagg cccctccccg tgcgcgggcg cgcctgccgc 120

cgcggcggcc gcgtggccat ggccatcgcc atccagccct cctccatcga gatggaggag 180

gagaccaccc tgaccaagcg caagcagccc cccaccaagc agcgccgcgt ggtggtgacc 240

ggcatgggcg tggagacccc catcggcaac aaccccgacc agttctacaa caacctgctg 300

cagggcgtgt ccggcatcac ccagatcgag gccttcgact gctcctccta ccccacccgc 360

atcgccggcg agatcaagaa cttctccacc gacggctggg tggcccccaa gctgtccaag 420

cgcatggacc gcttcatgct gtacatgctg accgccggca agaaggccct ggccgacgcc 480

ggcatctccc cctccgactc cgacgagatc gacaagtccc gctgcggcgt gctgatcggc 540

tccgccatgg gcggcatgaa ggtgttcaac gacgccatcg aggccctgcg cgtgtcctac 600

cgcaagatga acccttctg cgtgcccttc gccaccacca acatgggctc cgccatgctg 660

gccatggacc tgggctggat gggccccaac tactccatct ccaccgcctg cgccacctcc 720

aacttctgca tcctgaacgc cgccaaccac atcatccgcg gcgaggccga catgatgctg 780

tgcggcggct ccgacgccgt gatcatcccc atcggcctgg gcggcttcgt ggcctgccgc 840

gccctgtccg agcgcaacac cgaccccgcc aaggcctccc gcccctggga ctccggccgc 900

gacggcttcg tgatgggcga gggcgccggc gtgctgctgc tggaggagct ggagcacgcc 960

aagaagcgcg gcgccaagat ctacgccgag ttcctgggcg gctccttcac ctgcgacgcc 1020

taccacatga ccgagcccca ccccgagggc gccggcgtga tcctgtgcat cgagaaggcc 1080

ctgtcccagg ccggcgtgcg ccgcgaggac gtgaactaca tcaacgccca cgccacctcc 1140

acccccgccg gcgacctgaa ggagtaccac gccctgctgc actgcttcgg caacaaccag 1200

gagctgcgcg tgaactccac caagtccatg atcggccacc tgctgggcgc cgccggcgcc 1260

gtggaggccg tggccaccgt gcaggccatc cgcaccggct ggatccaccc caacatcaac 1320

ctggagaacc ccgaccaggg cgtggacacc aaggtgctgg tgggctccaa gaaggagcgc 1380

ctgaacgtga aggtgggcct gtccaactcc ttcggcttcg gcggccacaa ctcctccatc 1440

ctgttcgccc ccttccagat ggactacaag gaccacgacg gcgactacaa ggaccacgac 1500

atcgactaca aggacgacga cgacaagtga 1530

<210> 232

<211> 1584

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 232

atggccaccg catccacttt ctcggcgttc aatgcccgct gcggcgacct gcgtcgctcg 60

gcgggctccg ggccccggcg cccagcgagg cccctccccg tgcgcgggcg cgcctcccac 120

tactactcct ccaacggcct gttccccaac acccccctgc tgcccaagcg ccacccccgc 180

ctgcaccacc gcctgccccg ctccggcgag gccatggccg tggccgtgca gcccgagaag 240

gaggtggcca ccaacaagaa gcccctgatg aagcagcgcc gcgtggtggt gaccggcatg 300

ggcgtggtgt ccccccctggg ccacgacatc gacgtgtact acaacaacct gctggacggc 360

tcctccggca tctcccagat cgactccttc gactgcgccc agttccccac ccgcatcgcc 420

ggcgagatca agtccttctc caccgacggc tgggtggccc ccaagctgtc caagcgcatg 480

gacaagttca tgctgtacat gctgaccgcc ggcaagaagg ccctggccga cggcggcatc 540

accgaggaca tgatggacga gctggacaag gcccgctgcg gcgtgctgat cggctccgcc 600

atgggcggca tgaaggtgtt caacgacgcc atcgaggccc tgcgcatctc ctaccgcaag 660

atgaacccct tctgcgtgcc cttcgccacc accaacatgg gctccgccat gctggccatg 720

gacctgggct ggatgggccc caactactcc atctccaccg cctgcgccac ctccaacttc 780

tgcatcctga acgccgccaa ccacatcatc cgcggcgagg ccgacatcat gctgtgcggc 840

ggctccgacg ccgccatcat ccccatcggc ctgggcggct tcgtggcctg ccgcgccctg 900

tcccagcgca acgacgaccc caccaaggcc tcccgccccc gggacatgaa ccgcgacggc 960

ttcgtgatgg gcgagggcgc cggcgtgctg ctgctggagg agctggagca cgccaagaag 1020

cgcggcgcca acatctacgc cgagttcctg ggcggctcct tcacctgcga cgcctaccac 1080

atgaccgagc cccgccccga cggcgtgggc gtgatcctgt gcatcgagaa ggccctggcc 1140

cgctccggcg tgtccaagga ggaggtgaac tacatcaacg cccacgccac ctccaccccc 1200

gccggcgacc tgaaggagta cgaggccctg atgcgctgct tctcccagaa ccccgacctg 1260
cgcgtgaact ccaccaagtc catgatcggc cacctgctgg gcgccgccgg cgccgtggag 1320
gccatcgcca ccatccaggc catccgcacc ggctgggtgc accccaacat caacctggag 1380
aaccccgagg agggcgtgga caccaaggtg ctggtgggcc ccaagaagga gcgcctggac 1440
atcaaggtgg ccctgtccaa ctccttcggc ttcggcggcc acaactcctc catcatcttc 1500
gccccctaca agatggacta caaggaccac gacggcgact acaaggacca cgacatcgac 1560
tacaaggacg acgacgacaa gtga 1584

<210> 233
<211> 7008
<212> DNA
<213> Artificial Sequence
<220><223> synthesized construct
<400> 233
gctcttcgcc gccgccactc ctgctcgagc gcgcccgcgc gtgcgccgcc agcgccttgg 60
ccttttcgcc gcgctcgtgc gcgtcgctga tgtccatcac caggtccatg aggtctgcct 120
tgcgccggct gagccactgc ttcgtccggg cggccaagag gagcatgagg gaggactcct 180
ggtccagggt cctgacgtgg tcgcggctct gggagcgggc cagcatcatc tggctctgcc 240
gcaccgaggc cgcctccaac tggtcctcca gcagccgcag tcgccgccga ccctggcaga 300
ggaagacagg tgagggggt atgaattgta cagaacaacc acgagccttg tctaggcaga 360

atccctacca gtcatggctt tacctggatg acggcctgcg aacagctgtc cagcgaccct 420
cgctgccgcc gcttctcccg cacgcttctt tccagcaccg tgatggcgcg agccagcgcc 480
gcacgctggc gctgcgcttc gccgatctga ggacagtcgg ggaactctga tcagtctaaa 540
ccccccttgcg cgttagtgtt gccatccttt gcagaccggt gagagccgac ttgttgtgcg 600
ccacccccca caccacctcc tcccagacca attctgtcac ctttttggcg aaggcatcgg 660
cctcggcctg cagagaggac agcagtgccc agccgctggg ggttggcgga tgcacgctca 720
ggtacccttt cttgcgctat gacacttcca gcaaaaggta gggcgggctg cgagacggct 780

tcccggcgct gcatgcaaca ccgatgatgc ttcgaccccc cgaagctcct tcggggctgc 840
atgggcgctc cgatgccgct ccagggcgag cgctgtttaa atagccaggc ccccgattgc 900
aaagacatta tagcgagcta ccaaagccat attcaaacac ctagatcact accacttcta 960
cacaggccac tcgagcttgt gatcgcactc cgctaagggg gcgcctcttc ctcttcgttt 1020

cagtcacaac ccgcaaactc tagaatatca atgctgctgc aggccttcct gttcctgctg 1080

gccggcttcg ccgccaagat cagcgcctcc atgacgaacg agacgtccga ccgccccctg 1140

gtgcacttca cccccaacaa gggctggatg aacgacccca acggcctgtg gtacgacgag 1200

aaggacgcca agtggcacct gtacttccag tacaacccga acgacaccgt ctgggggacg 1260

cccttgttct ggggccacgc cacgtccgac gacctgacca actgggagga ccagcccatc 1320

gccatcgccc cgaagcgcaa cgactccggc gccttctccg gctccatggt ggtggactac 1380

aacaacacct ccggcttctt caacgacacc atcgacccgc gccagcgctg cgtggccatc 1440

tggacctaca acaccccgga gtccgaggag cagtacatct cctacagcct ggacggcggc 1500

tacaccttca ccgagtacca gaagaacccc gtgctggccg ccaactccac ccagttccgc 1560

gacccgaagg tcttctggta cgagccctcc cagaagtgga tcatgaccgc ggccaagtcc 1620

caggactaca agatcgagat ctactcctcc gacgacctga agtcctggaa gctggagtcc 1680

gcgttcgcca acgagggctt cctcggctac cagtacgagt gccccggcct gatcgaggtc 1740

cccaccgagc aggaccccag caagtcctac tgggtgatgt tcatctccat caaccccggc 1800

gccccggccg gcggctcctt caaccagtac ttcgtcggca gcttcaacgg cacccacttc 1860

gaggccttcg acaaccagtc ccgcgtggtg gacttcggca aggactacta cgccctgcag 1920

accttcttca acaccgaccc gacctacggg agcgccctgg gcatcgcgtg ggcctccaac 1980

tgggagtact ccgccttcgt gcccaccaac ccctggcgct cctccatgtc cctcgtgcgc 2040

aagttctccc tcaacaccga gtaccaggcc aacccggaga cggagctgat caacctgaag 2100

gccgagccga tcctgaacat cagcaacgcc ggcccctgga gccggttcgc caccaacacc 2160

acgttgacga aggccaacag ctacaacgtc gacctgtcca acagcaccgg caccctggag 2220

ttcgagctgg tgtacgccgt caacaccacc cagacgatct ccaagtccgt gttcgcggac 2280

ctctccctct ggttcaaggg cctggaggac cccgaggagt acctccgcat gggcttcgag 2340

gtgtccgcgt cctccttctt cctggaccgc gggaacagca aggtgaagtt cgtgaaggag 2400

aacccctact tcaccaaccg catgagcgtg aacaaccagc ccttcaagag cgagaacgac 2460

ctgtcctact acaaggtgta cggcttgctg gaccagaaca tcctggagct gtacttcaac 2520

gacggcgacg tcgtgtccac caacacctac ttcatgacca ccgggaacgc cctgggctcc 2580

gtgaacatga cgacgggggt ggacaacctg ttctacatcg acaagttcca ggtgcgcgag 2640

gtcaagtgac aattggcagc agcagctcgg atagtatcga cacactctgg acgctggtcg 2700

tgtgatggac tgttgccgcc acacttgctg ccttgacctg tgaatatccc tgccgctttt 2760

atcaaacagc ctcagtgtgt ttgatcttgt gtgtacgcgc ttttgcgagt tgctagctgc 2820

ttgtgctatt tgcgaatacc acccccagca tccccttccc tcgtttcata tcgcttgcat 2880

cccaaccgca acttatctac gctgtcctgc tatccctcag cgctgctcct gctcctgctc 2940
actgcccctc gcacagcctt ggtttgggct ccgcctgtat tctcctggta ctgcaacctg 3000
taaaccagca ctgcaatgct gatgcacggg aagtagtggg atgggaacac aaatggagga 3060
tcccgcgtct cgaacagagc gcgcagagga acgctgaagg tctcgcctct gtcgcacctc 3120
agcgcggcat acaccacaat aaccacctga cgaatgcgct tggttcttcg tccattagcg 3180
aagcgtccgg ttcacacacg tgccacgttg gcgaggtggc aggtgacaat gatcggtgga 3240
gctgatggtc gaaacgttca cagcctaggg atatcgaatt cggccgacag gacgcgcgtc 3300

aaaggtgctg gtcgtgtatg ccctggccgg caggtcgttg ctgctgctgg ttagtgattc 3360
cgcaaccctg attttggcgt cttattttgg cgtggcaaac gctggcgccc gcgagccggg 3420
ccggcggcga tgcggtgccc cacggctgcc ggaatccaag ggaggcaaga gcgcccgggt 3480
cagttgaagg gctttacgcg caaggtacag ccgctcctgc aaggctgcgt ggtggaattg 3540
gacgtgcagg tcctgctgaa gttcctccac cgcctcacca gcggacaaag caccggtgta 3600
tcaggtccgt gtcatccact ctaaagaact cgactacgac ctactgatgg ccctagattc 3660
ttcatcaaaa acgcctgaga cacttgccca ggattgaaac tccctgaagg gaccaccagg 3720

ggccctgagt tgttccttcc ccccgtggcg agctgccagc caggctgtac ctgtgatcga 3780
ggctggcggg aaaataggct tcgtgtgctc aggtcatggg aggtgcagga cagctcatga 3840
aacgccaaca atcgcacaat tcatgtcaag ctaatcagct atttcctctt cacgagctgt 3900
aattgtccca aaattctggt ctaccggggg tgatccttcg tgtacgggcc cttccctcaa 3960
ccctaggtat gcgcgcatgc ggtcgccgcg caactcgcgc gagggccgag ggtttgggac 4020
gggccgtccc gaaatgcagt tgcacccgga tgcgtggcac ctttttgcg ataatttatg 4080
caatggactg ctctgcaaaa ttctggctct gtcgccaacc ctaggatcag cggcgtagga 4140

tttcgtaatc attcgtcctg atggggagct accgactacc ctaatatcag cccgactgcc 4200
tgacgccagc gtccactttt gtgcacacat tccattcgtg cccaagacat ttcattgtgg 4260
tgcgaagcgt ccccagttac gctcacctgt ttcccgacct ccttactgtt ctgtcgacag 4320
agcgggccca caggccggtc gcagccacta gtatggccac cgcatccact ttctcggcgt 4380
tcaatgcccg ctgcggcgac ctgcgtcgct cggcgggctc cgggccccgg cgcccagcga 4440
ggcccctccc cgtgcgcggg cgcgccgccg ccgccgccga cgccaacccc gcccgccccg 4500
agcgccgcgt ggtgatcacc ggccagggcg tggtgacctc cctgggccag accatcgagc 4560

agttctactc ctccctgctg gagggcgtgt ccggcatctc ccagatccag aagttcgaca 4620

ccaccggcta caccaccacc atcgccggcg agatcaagtc cctgcagctg gacccctacg 4680

tgcccaagcg ctgggccaag cgcgtggacg acgtgatcaa gtacgtgtac atcgccggca 4740

agcaggccct ggagtccgcc ggcctgccca tcgaggccgc cggcctggcc ggcgccggcc 4800

tggaccccgc cctgtgcggc gtgctgatcg gcaccgccat ggccggcatg acctccttcg 4860

ccgccggcgt ggaggccctg acccgcggcg gcgtgcgcaa gatgaacccc ttctgcatcc 4920

ccttctccat ctccaacatg ggcggcgcca tgctggccat ggacatcggc ttcatgggcc 4980

ccaactactc catctccacc gcctgcgcca ccggcaacta ctgcatcctg ggcgccgccg 5040

accacatccg ccgcggcgac gccaacgtga tgctggccgg cggcgccgac gccgccatca 5100

tcccctccgg catcggcggc ttcatcgcct gcaaggccct gtccaagcgc aacgacgagc 5160

ccgagcgcgc ctcccgcccc tgggacgccg accgcgacgg cttcgtgatg ggcgagggcg 5220

ccggcgtgct ggtgctggag gagctggagc acgccaagcg ccgcggcgcc accatcctgg 5280

ccgagctggt gggcggcgcc gccacctccg acgcccacca catgaccgag cccgaccccc 5340

agggccgcgg cgtgcgcctg tgcctggagc gcgccctgga gcgcgcccgc ctggcccccg 5400

agcgcgtggg ctacgtgaac gcccacggca cctccacccc cgccggcgac gtggccgagt 5460

accgcgccat ccgcgccgtg atcccccagg actccctgcg catcaactcc accaagtcca 5520

tgatcggcca cctgctgggc ggcgccggcg ccgtggaggc cgtggccgcc atccaggccc 5580

tgcgcaccgg ctggctgcac cccaacctga acctggagaa ccccgccccc ggcgtggacc 5640

ccgtggtgct ggtgggcccc cgcaaggagc gcgccgagga cctggacgtg gtgctgtcca 5700

actccttcgg cttcggcggc cacaactcct gcgtgatctt ccgcaagtac gacgagatgg 5760

actacaagga ccacgacggc gactacaagg accacgacat cgactacaag gacgacgacg 5820

acaagtgaat cgatagatct cttaaggcag cagcagctcg gatagtatcg acacactctg 5880

gacgctggtc gtgtgatgga ctgttgccgc cacacttgct gccttgacct gtgaatatcc 5940

ctgccgcttt tatcaaacag cctcagtgtg tttgatcttg tgtgtacgcg cttttgcgag 6000

ttgctagctg cttgtgctat ttgcgaatac cacccccagc atccccttcc ctcgtttcat 6060

atcgcttgca tcccaaccgc aacttatcta cgctgtcctg ctatccctca gcgctgctcc 6120

tgctcctgct cactgccccc cgcacagcct tggtttgggc tccgcctgta ttctcctggt 6180

actgcaacct gtaaaccagc actgcaatgc tgatgcacgg gaagtagtgg gatgggaaca 6240

caaatggaaa gcttaattaa gagctcttgt tttccagaag gagttgctcc ttgagccttt 6300

cattctcagc ctcgataacc tccaaagccg ctctaattgt ggaggggggtt cgaatttaaa 6360

agcttggaat gttggttcgt gcgtctggaa caagcccaga cttgttgctc actgggaaaa 6420

ggaccatcag ctccaaaaaa cttgccgctc aaaccgcgta cctctgcttt cgcgcaatct 6480

gccctgttga aatcgccacc acattcatat tgtgacgctt gagcagtctg taattgcctc 6540

agaatgtgga atcatctgcc ccctgtgcga gcccatgcca ggcatgtcgc gggcgaggac 6600

acccgccact cgtacagcag accattatgc tacctcacaa tagttcataa cagtgaccat 6660

atttctcgaa gctccccaac gagcacctcc atgctctgag tggccacccc ccggccctgg 6720

tgcttgcgga gggcaggtca accggcatgg ggctaccgaa atccccgacc ggatcccacc 6780

acccccgcga tgggaagaat ctctccccgg gatgtgggcc caccaccagc acaacctgct 6840

ggcccaggcg agcgtcaaac cataccacac aaatatcctt ggcatcggcc ctgaattcct 6900

tctgccgctc tgctacccgg tgcttctgtc cgaagcaggg gttgctaggg atcgctccga 6960

gtccgcaaac ccttgtcgcg tggcggggct tgttcgagct tgaagagc 7008

<210> 234

<211> 7008

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 234

gctcttcgcc gccgccactc ctgctcgagc gcgcccgcgc gtgcgccgcc agcgccttgg 60

ccttttcgcc gcgctcgtgc gcgtcgctga tgtccatcac caggtccatg aggtctgcct 120

tgcgccggct gagccactgc ttcgtccggg cggccaagag gagcatgagg gaggactcct 180

ggtccagggt cctgacgtgg tcgcggctct gggagcgggc cagcatcatc tggctctgcc 240

gcaccgaggc cgcctccaac tggtcctcca gcagccgcag tcgccgccga ccctggcaga 300

ggaagacagg tgaggggggt atgaattgta cagaacaacc acgagccttg tctaggcaga 360

atccctacca gtcatggctt tacctggatg acggcctgcg aacagctgtc cagcgaccct 420

cgctgccgcc gcttctcccg cacgcttctt tccagcaccg tgatggcgcg agccagcgcc 480

gcacgctggc gctgcgcttc gccgatctga ggacagtcgg ggaactctga tcagtctaaa 540

ccccccttgcg cgttagtgtt gccatccttt gcagaccggt gagagccgac ttgttgtgcg 600

ccacccccca caccacctcc tcccagacca attctgtcac ctttttggcg aaggcatcgg 660

cctcggcctg cagagaggac agcagtgccc agccgctggg ggttggcgga tgcacgctca 720

ggtacccttt cttgcgctat gacacttcca gcaaaaggta gggcgggctg cgagacggct 780

tcccggcgct gcatgcaaca ccgatgatgc ttcgaccccc cgaagctcct tcggggctgc 840

atgggcgctc cgatgccgct ccagggcgag cgctgtttaa atagccaggc ccccgattgc 900

aaagacatta tagcgagcta ccaaagccat attcaaacac ctagatcact accacttcta 960

cacaggccac tcgagcttgt gatcgcactc cgctaagggg gcgcctcttc ctcttcgttt 1020

cagtcacaac ccgcaaactc tagaatatca atgctgctgc aggccttcct gttcctgctg 1080

gccggcttcg ccgccaagat cagcgcctcc atgacgaacg agacgtccga ccgccccctg 1140

gtgcacttca cccccaacaa gggctggatg aacgacccca acggcctgtg gtacgacgag 1200

aaggacgcca agtggcacct gtacttccag tacaacccga acgacaccgt ctgggggacg 1260

cccttgttct ggggccacgc cacgtccgac gacctgacca actgggagga ccagcccatc 1320

gccatcgccc cgaagcgcaa cgactccggc gccttctccg gctccatggt ggtggactac 1380

aacaacacct ccggcttctt caacgacacc atcgacccgc gccagcgctg cgtggccatc 1440

tggacctaca acaccccgga gtccgaggag cagtacatct cctacagcct ggacggcggc 1500

tacaccttca ccgagtacca gaagaacccc gtgctggccg ccaactccac ccagttccgc 1560

gacccgaagg tcttctggta cgagccctcc cagaagtgga tcatgaccgc ggccaagtcc 1620

caggactaca agatcgagat ctactcctcc gacgacctga agtcctggaa gctggagtcc 1680

gcgttcgcca acgagggctt cctcggctac cagtacgagt gccccggcct gatcgaggtc 1740

cccaccgagc aggaccccag caagtcctac tgggtgatgt tcatctccat caaccccggc 1800

gccccggccg gcggctcctt caaccagtac ttcgtcggca gcttcaacgg cacccacttc 1860

gaggccttcg acaaccagtc ccgcgtggtg gacttcggca aggactacta cgccctgcag 1920

accttcttca acaccgaccc gacctacggg agcgccctgg gcatcgcgtg ggcctccaac 1980

tgggagtact ccgccttcgt gcccaccaac ccctggcgct cctccatgtc cctcgtgcgc 2040

aagttctccc tcaacaccga gtaccaggcc aacccggaga cggagctgat caacctgaag 2100

gccgagccga tcctgaacat cagcaacgcc ggcccctgga gccggttcgc caccaacacc 2160

acgttgacga aggccaacag ctacaacgtc gacctgtcca acagcaccgg caccctggag 2220

ttcgagctgg tgtacgccgt caacaccacc cagacgatct ccaagtccgt gttcgcggac 2280

ctctccctct ggttcaaggg cctggaggac cccgaggagt acctccgcat gggcttcgag 2340

gtgtccgcgt cctccttctt cctggaccgc gggaacagca aggtgaagtt cgtgaaggag 2400

aacccctact tcaccaaccg catgagcgtg aacaaccagc ccttcaagag cgagaacgac 2460

ctgtcctact acaaggtgta cggcttgctg gaccagaaca tcctggagct gtacttcaac 2520

gacggcgacg tcgtgtccac caacacctac ttcatgacca ccgggaacgc cctgggctcc 2580

gtgaacatga cgacgggggt ggacaacctg ttctacatcg acaagttcca ggtgcgcgag 2640

gtcaagtgac aattggcagc agcagctcgg atagtatcga cacactctgg acgctggtcg 2700

tgtgatggac tgttgccgcc acacttgctg ccttgacctg tgaatatccc tgccgctttt 2760

atcaaacagc ctcagtgtgt ttgatcttgt gtgtacgcgc ttttgcgagt tgctagctgc 2820

ttgtgctatt tgcgaatacc acccccagca tccccttccc tcgtttcata tcgcttgcat 2880

cccaaccgca acttatctac gctgtcctgc tatccctcag cgctgctcct gctcctgctc 2940

actgcccctc gcacagcctt ggtttgggct ccgcctgtat tctcctggta ctgcaacctg 3000

taaaccagca ctgcaatgct gatgcacggg aagtagtggg atgggaacac aaatggagga 3060

tcccgcgtct cgaacagagc gcgcagagga acgctgaagg tctcgcctct gtcgcacctc 3120

agcgcggcat acaccacaat aaccacctga cgaatgcgct tggttcttcg tccattagcg 3180

aagcgtccgg ttcacacacg tgccacgttg gcgaggtggc aggtgacaat gatcggtgga 3240

gctgatggtc gaaacgttca cagcctaggg atatcgaatt cggccgacag gacgcgcgtc 3300

aaaggtgctg gtcgtgtatg ccctggccgg caggtcgttg ctgctgctgg ttagtgattc 3360

cgcaaccctg attttggcgt cttattttgg cgtggcaaac gctggcgccc gcgagccggg 3420

ccggcggcga tgcggtgccc cacggctgcc ggaatccaag ggaggcaaga gcgcccgggt 3480

cagttgaagg gctttacgcg caaggtacag ccgctcctgc aaggctgcgt ggtggaattg 3540

gacgtgcagg tcctgctgaa gttcctccac cgcctcacca gcggacaaag caccggtgta 3600

tcaggtccgt gtcatccact ctaaagaact cgactacgac ctactgatgg ccctagattc 3660

ttcatcaaaa acgcctgaga cacttgccca ggattgaaac tccctgaagg gaccaccagg 3720

ggccctgagt tgttccttcc ccccgtggcg agctgccagc caggctgtac ctgtgatcga 3780

ggctggcggg aaaataggct tcgtgtgctc aggtcatggg aggtgcagga cagctcatga 3840

aacgccaaca atcgcacaat tcatgtcaag ctaatcagct atttcctctt cacgagctgt 3900

aattgtccca aaattctggt ctaccggggg tgatccttcg tgtacgggcc cttccctcaa 3960

ccctaggtat gcgcgcatgc ggtcgccgcg caactcgcgc gagggccgag ggtttgggac 4020

gggccgtccc gaaatgcagt tgcacccgga tgcgtggcac ctttttgcg ataatttatg 4080

caatggactg ctctgcaaaa ttctggctct gtcgccaacc ctaggatcag cggcgtagga 4140

tttcgtaatc attcgtcctg atggggagct accgactacc ctaatatcag cccgactgcc 4200

tgacgccagc gtccactttt gtgcacacat tccattcgtg cccaagacat ttcattgtgg 4260

tgcgaagcgt ccccagttac gctcacctgt ttcccgacct ccttactgtt ctgtcgacag 4320

agcgggccca caggccggtc gcagccacta gtatggccac cgcatccact ttctcggcgt 4380

tcaatgcccg ctgcggcgac ctgcgtcgct cggcgggctc cgggccccgg cgcccagcga 4440

ggcccctccc cgtgcgcggg cgcgccgccg ccgccgccga cgccaacccc gcccgccccg 4500

agcgccgcgt ggtgatcacc ggccagggcg tggtgacctc cctgggccag accatcgagc 4560

agttctactc ctccctgctg gagggcgtgt ccggcatctc ccagatccag aagttcgaca 4620

ccaccggcta caccaccacc atcgccggcg agatcaagtc cctgcagctg gacccctacg 4680
tgcccaagcg ctgggccaag cgcgtggacg acgtgatcaa gtacgtgtac atcgccggca 4740
agcaggccct ggagtccgcc ggcctgccca tcgaggccgc cggcctggcc ggcgccggcc 4800
tggaccccgc cctgtgcggc gtgctgatcg gcaccgccat ggccggcatg acctccttcg 4860
ccgccggcgt ggaggccctg acccgcggcg gcgtgcgcaa gatgaacccc ttctgcatcc 4920
ccttctccat ctccaacatg ggcggcgcca tgctggccat ggacatcggc ttcatgggcc 4980
ccaactactc catctccacc gcctgcgcca ccggcaacta ctgcatcctg ggcgccgccg 5040

accacatccg ccgcggcgac gccaacgtga tgctggccgg cggcgccgac gccgccatca 5100
tcccctccgg catcggcggc ttcatcgcct gcaaggccct gtccaagcgc aacgacgagc 5160
ccgagcgcgc ctcccgcccc tgggacgccg accgcgacgg cttcgtgatg ggcgagggcg 5220
ccggcgtgct ggtgctggag gagctggagc acgccaagcg ccgcggcgcc accatcctgg 5280
ccgagctggt gggcggcgcc gccacctccg acgcccacca catgaccgag cccgaccccc 5340
agggccgcgg cgtgcgcctg tgcctggagc gcgccctgga gcgcgcccgc ctggcccccg 5400
agcgcgtggg ctacgtgaac gcccacggca cctccacccc cgccggcgac gtggccgagt 5460

accgcgccat ccgcgccgtg atcccccagg actccctgcg catcaactcc accaagtcca 5520
tgatcggcca cctgctgggc ggcgccggcg ccgtggaggc cgtggccgcc atccaggccc 5580
tgcgcaccgg ctggctgcac cccaacctga acctggagaa ccccgccccc ggcgtggacc 5640
ccgtggtgct ggtgggcccc cgcaaggagc gcgccgagga cctggacgtg gtgctgtcca 5700
actccttcgg cttcggcggc cacaactcct gcgtgatctt ccgcaagtac gacgagatgg 5760
actacaagga ccacgacggc gactacaagg accacgacat cgactacaag gacgacgacg 5820
acaagtgaat cgatagatct cttaaggcag cagcagctcg gatagtatcg acacactctg 5880

gacgctggtc gtgtgatgga ctgttgccgc cacacttgct gccttgacct gtgaatatcc 5940
ctgccgcttt tatcaaacag cctcagtgtg tttgatcttg tgtgtacgcg cttttgcgag 6000
ttgctagctg cttgtgctat ttgcgaatac cacccccagc atccccttcc ctcgtttcat 6060
atcgcttgca tcccaaccgc aacttatcta cgctgtcctg ctatccctca gcgctgctcc 6120
tgctcctgct cactgccccc cgcacagcct tggtttgggc tccgcctgta ttctcctggt 6180
actgcaacct gtaaaccagc actgcaatgc tgatgcacgg gaagtagtgg gatgggaaca 6240
caaatggaaa gcttaattaa gagctcttgt tttccagaag gagttgctcc ttgagccttt 6300

cattctcagc ctcgataacc tccaaagccg ctctaattgt ggaggggttt cgaatttaaa 6360

agcttggaat gttggttcgt gcgtctggaa caagcccaga cttgttgctc actgggaaaa 6420

ggaccatcag ctccaaaaaa cttgccgctc aaaccgcgta cctctgcttt cgcgcaatct 6480

gccctgttga aatcgccacc acattcatat tgtgacgctt gagcagtctg taattgcctc 6540

agaatgtgga atcatctgcc ccctgtgcga gcccatgcca ggcatgtcgc gggcgaggac 6600

acccgccact cgtacagcag accattatgc tacctcacaa tagttcataa cagtgaccat 6660

atttctcgaa gctccccaac gagcacctcc atgctctgag tggccacccc ccggccctgg 6720

tgcttgcgga gggcaggtca accggcatgg ggctaccgaa atccccgacc ggatcccacc 6780

acccccgcga tgggaagaat ctctccccgg gatgtgggcc caccaccagc acaacctgct 6840

ggcccaggcg agcgtcaaac cataccacac aaatatcctt ggcatcggcc ctgaattcct 6900

tctgccgctc tgctacccgg tgcttctgtc cgaagcaggg gttgctaggg atcgctccga 6960

gtccgcaaac ccttgtcgcg tggcggggct tgttcgagct tgaagagc 7008

<210> 235

<211> 1476

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 235

atggccaccg catccacttt ctcggcgttc aatgcccgct gcggcgacct gcgtcgctcg 60

gcgggctccg ggccccggcg cccagcgagg cccctccccg tgcgcgggcg cgccgccgcc 120

gccgccgacg ccaaccccgc ccgccccgag cgccgcgtgg tgatcaccgg ccagggcgtg 180

gtgacctccc tgggccagac catcgagcag ttctactcct ccctgctgga gggcgtgtcc 240

ggcatctccc agatccagaa gttcgacacc accggctaca ccaccaccat cgccggcgag 300

atcaagtccc tgcagctgga cccctacgtg cccaagcgct gggccaagcg cgtggacgac 360

gtgatcaagt acgtgtacat cgccggcaag caggccctgg agtccgccgg cctgcccatc 420

gaggccgccg gcctggccgg cgccggcctg gaccccgccc tgtgcggcgt gctgatcggc 480

accgccatgg ccggcatgac ctccttcgcc gccggcgtgg aggccctgac ccgcggcggc 540

gtgcgcaaga tgaacccctt ctgcatcccc ttctccatct ccaacatggg cggcgccatg 600

ctggccatgg acatcggctt catgggcccc aactactcca tctccaccgc ctgcgccacc 660

ggcaactact gcatcctggg cgccgccgac cacatccgcc gcggcgacgc caacgtgatg 720

ctggccggcg gcgccgacgc cgccatcatc ccctccggca tcggcggctt catcgcctgc 780

aaggccctgt ccaagcgcaa cgacgagccc gagcgcgcct cccgcccctg ggacgccgac 840

cgcgacggct tcgtgatggg cgagggcgcc ggcgtgctgg tgctggagga gctggagcac 900

gccaagcgcc gcggcgccac catcctggcc gagctggtgg gcggcgccgc cacctccgac 960

gcccaccaca tgaccgagcc cgaccccccag ggccgcggcg tgcgcctgtg cctggagcgc 1020

gccctggagc gcgcccgcct ggcccccgag cgcgtgggct acgtgaacgc ccacggcacc 1080

tccacccccg ccggcgacgt ggccgagtac cgcgccatcc gcgccgtgat cccccaggac 1140

tccctgcgca tcaactccac caagtccatg atcggccacc tgctgggcgg cgccggcgcc 1200

gtggaggccg tggccgccat ccaggccctg cgcaccggct ggctgcaccc caacctgaac 1260

ctggagaacc ccgcccccgg cgtggacccc gtggtgctgg tgggcccccg caaggagcgc 1320

gccgaggacc tggacgtggt gctgtccaac tccttcggct tcggcggcca caactcctgc 1380

gtgatcttcc gcaagtacga cgagatggac tacaaggacc acgacggcga ctacaaggac 1440

cacgacatcg actacaagga cgacgacgac aagtga 1476

<210> 236

<211> 491

<212> PRT

<213> Artificial Sequence

<220><223> synthesized construct

<400> 236

Met Ala Thr Ala Ser Thr Phe Ser Ala Phe Asn Ala Arg Cys Gly Asp
1 5 10 15

Leu Arg Arg Ser Ala Gly Ser Gly Pro Arg Arg Pro Ala Arg Pro Leu

20 25 30

Pro Val Arg Gly Arg Ala Ala Ala Ala Ala Asp Ala Asn Pro Ala Arg
35 40 45

Pro Glu Arg Arg Val Val Ile Thr Gly Gln Gly Val Val Thr Ser Leu
50 55 60

Gly Gln Thr Ile Glu Gln Phe Tyr Ser Ser Leu Leu Glu Gly Val Ser
65 70 75 80

Gly Ile Ser Gln Ile Gln Lys Phe Asp Thr Thr Gly Tyr Thr Thr Thr

85 90 95

Ile Ala Gly Glu Ile Lys Ser Leu Gln Leu Asp Pro Tyr Val Pro Lys
100 105 110

Arg Trp Ala Lys Arg Val Asp Asp Val Ile Lys Tyr Val Tyr Ile Ala
115 120 125
Gly Lys Gln Ala Leu Glu Ser Ala Gly Leu Pro Ile Glu Ala Ala Gly
130 135 140
Leu Ala Gly Ala Gly Leu Asp Pro Ala Leu Cys Gly Val Leu Ile Gly

145 150 155 160
Thr Ala Met Ala Gly Met Thr Ser Phe Ala Ala Gly Val Glu Ala Leu
165 170 175
Thr Arg Gly Gly Val Arg Lys Met Asn Pro Phe Cys Ile Pro Phe Ser
180 185 190
Ile Ser Asn Met Gly Gly Ala Met Leu Ala Met Asp Ile Gly Phe Met
195 200 205
Gly Pro Asn Tyr Ser Ile Ser Thr Ala Cys Ala Thr Gly Asn Tyr Cys

210 215 220
Ile Leu Gly Ala Ala Asp His Ile Arg Arg Gly Asp Ala Asn Val Met
225 230 235 240
Leu Ala Gly Gly Ala Asp Ala Ala Ile Ile Pro Ser Gly Ile Gly Gly
245 250 255
Phe Ile Ala Cys Lys Ala Leu Ser Lys Arg Asn Asp Glu Pro Glu Arg
260 265 270
Ala Ser Arg Pro Trp Asp Ala Asp Arg Asp Gly Phe Val Met Gly Glu

275 280 285
Gly Ala Gly Val Leu Val Leu Glu Glu Leu Glu His Ala Lys Arg Arg
290 295 300
Gly Ala Thr Ile Leu Ala Glu Leu Val Gly Gly Ala Ala Thr Ser Asp
305 310 315 320
Ala His His Met Thr Glu Pro Asp Pro Gln Gly Arg Gly Val Arg Leu
325 330 335
Cys Leu Glu Arg Ala Leu Glu Arg Ala Arg Leu Ala Pro Glu Arg Val

340 345 350
Gly Tyr Val Asn Ala His Gly Thr Ser Thr Pro Ala Gly Asp Val Ala

355 360 365

Glu Tyr Arg Ala Ile Arg Ala Val Ile Pro Gln Asp Ser Leu Arg Ile

370 375 380

Asn Ser Thr Lys Ser Met Ile Gly His Leu Leu Gly Gly Ala Gly Ala

385 390 395 400

Val Glu Ala Val Ala Ala Ile Gln Ala Leu Arg Thr Gly Trp Leu His

405 410 415

Pro Asn Leu Asn Leu Glu Asn Pro Ala Pro Gly Val Asp Pro Val Val

420 425 430

Leu Val Gly Pro Arg Lys Glu Arg Ala Glu Asp Leu Asp Val Val Leu

435 440 445

Ser Asn Ser Phe Gly Phe Gly Gly His Asn Ser Cys Val Ile Phe Arg

450 455 460

Lys Tyr Asp Glu Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp

465 470 475 480

His Asp Ile Asp Tyr Lys Asp Asp Asp Asp Lys

485 490

<210> 237

<211> 3602

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 237

gctcttcgct caccgcgtga attgctgtcc caaacgtaag catcatcgtg gctcggtcac 60

gcgatcctgg atccggggat cctagaccgc tggtggagag cgctgccgtc ggattggtgg 120

caagtaagat tgcgcaggtt ggcgaaggga gagaccaaaa ccggaggctg gaagcgggca 180

caacatcgta ttattgcgta tagtagagca gtggcagtcg catttcgagg tccgcaacgg 240

atctcgcaag ctcgctacgc tcacagtagg agaaagggga ccactgcccc tgccagaatg 300

gtcgcgaccc tctccctcgc cggccccgcc tgcaacacgc agtgcgtatc cggcaagcgg 360

gctgtcgcct tcaaccgccc ccatgttggc gtccgggctc gatcaggtgc gctgaggggg 420

gtttggtgtg cccgcgcctc tgggcccgtg tcggccgtgc ggacgtgggg ccctgggcag 480

tggatcagca gggtttgcgt gcaaatgcct ataccggcga ttgaatagcg atgaacggga 540

tacggttgcg ctcactccat gcccatgcga ccccgtttct gtccgccagc cgtggtcgcc 600

cgggctgcga agcgggaccc cacccagcgc attgtgatca ccggaatggg cgtggggtac 660

cctttcttgc gctatgacac ttccagcaaa aggtagggcg ggctgcgaga cggcttcccg 720

gcgctgcatg caacaccgat gatgcttcga ccccccgaag ctccttcggg gctgcatggg 780

cgctccgatg ccgctccagg gcgagcgctg tttaaatagc caggcccccg attgcaaaga 840

cattatagcg agctaccaaa gccatattca aacacctaga tcactaccac ttctacacag 900

gccactcgag cttgtgatcg cactccgcta agggggcgcc tcttcctctt cgtttcagtc 960

acaacccgca aacggcgcgc catgctgctg caggccttcc tgttcctgct ggccggcttc 1020

gccgccaaga tcagcgcctc catgacgaac gagacgtccg accgcccccct ggtgcacttc 1080

acccccaaca agggctggat gaacgacccc aacggcctgt ggtacgacga gaaggacgcc 1140

aagtggcacc tgtacttcca gtacaacccg aacgacaccg tctgggggac gcccttgttc 1200

tggggccacg ccacgtccga cgacctgacc aactgggagg accagcccat cgccatcgcc 1260

ccgaagcgca acgactccgg cgccttctcc ggctccatgg tggtggacta caacaacacc 1320

tccggcttct tcaacgacac catcgacccg cgccagcgct gcgtggccat ctggacctac 1380

aacaccccgg agtccgagga gcagtacatc tcctacagcc tggacggcgg ctacaccttc 1440

accgagtacc agaagaaccc cgtgctggcc gccaactcca cccagttccg cgacccgaag 1500

gtcttctggt acgagccctc ccagaagtgg atcatgaccg cggccaagtc ccaggactac 1560

aagatcgaga tctactcctc cgacgacctg aagtcctgga agctggagtc cgcgttcgcc 1620

aacgagggct tcctcggcta ccagtacgag tgccccggcc tgatcgaggt ccccaccgag 1680

caggacccca gcaagtccta ctgggtgatg ttcatctcca tcaaccccgg cgccccggcc 1740

ggcggctcct tcaaccagta cttcgtcggc agcttcaacg gcacccactt cgaggccttc 1800

gacaaccagt cccgcgtggt ggacttcggc aaggactact acgccctgca gaccttcttc 1860

aacaccgacc cgacctacgg gagcgccctg ggcatcgcgt gggcctccaa ctgggagtac 1920

tccgccttcg tgcccaccaa cccctggcgc tcctccatgt ccctcgtgcg caagttctcc 1980

ctcaacaccg agtaccaggc caacccggag acggagctga tcaacctgaa ggccgagccg 2040

atcctgaaca tcagcaacgc cggcccctgg agccggttcg ccaccaacac cacgttgacg 2100

aaggccaaca gctacaacgt cgacctgtcc aacagcaccg gcaccctgga gttcgagctg 2160

gtgtacgccg tcaacaccac ccagacgatc tccaagtccg tgttcgcgga cctctccctc 2220

tggttcaagg gcctggagga ccccgaggag tacctccgca tgggcttcga ggtgtccgcg 2280

tcctccttct tcctggaccg cgggaacagc aaggtgaagt tcgtgaagga gaacccctac 2340

ttcaccaacc gcatgagcgt gaacaaccag cccttcaaga gcgagaacga cctgtcctac 2400
tacaaggtgt acggcttgct ggaccagaac atcctggagc tgtacttcaa cgacggcgac 2460
gtcgtgtcca ccaacaccta cttcatgacc accgggaacg ccctgggctc cgtgaacatg 2520
acgacggggg tggacaacct gttctacatc gacaagttcc aggtgcgcga ggtcaagtga 2580
caattggcag cagcagctcg gatagtatcg acacactctg gacgctggtc gtgtgatgga 2640
ctgttgccgc cacacttgct gccttgacct gtgaatatcc ctgccgcttt tatcaaacag 2700
cctcagtgtg tttgatcttg tgtgtacgcg cttttgcgag ttgctagctg cttgtgctat 2760

ttgcgaatac caccccagc atcccttcc ctcgtttcat atcgcttgca tcccaaccgc 2820
aacttatcta cgctgtcctg ctatccctca gcgctgctcc tgctcctgct cactgcccct 2880
cgcacagcct tggtttgggc tccgcctgta ttctcctggt actgcaacct gtaaaccagc 2940
actgcaatgc tgatgcacgg gaagtagtgg gatgggaaca caaatggagg atcgtagagc 3000
tccacctgca tccgcctggc gctcgaggac gccggcgtct cgcccgacga ggtcaactac 3060
gtcaacgcgc acgccacctc caccctggtg ggcgacaagg ccgaggtgcg cgcggtcaag 3120
tcggtctttg gcgacatgaa gggcatcaag atgaacgcca ccaagtccat gatcgggcac 3180

tgcctgggcg ccgccggcgg catggaggcc gtcgccacgc tcatggccat ccgcaccggc 3240
tgggtgcacc ccaccatcaa ccacgacaac cccatcgccg aggtcgacgg cctggacgtc 3300
gtcgccaacg ccaaggccca gcacaaaatc aacgtcgcca tctccaactc cttcggcttc 3360
ggcgggcaca actccgtcgt cgcctttgcg cccttccgcg agtaggcgga gcgagcgcgc 3420
ttggctgagg agggaggcgg ggtgcgagcc ctttggctgc gcgcgatact ctccccgcac 3480
gagcagactc cacgcgcctg aatctacttg tcaacgagca accgtgtgtt ttgtccgtgg 3540
ccattcttat tatttctccg actgtggccg tactctgttt ggctgtgcaa gcaccgaaga 3600

gc 3602

<210> 238

<211> 661

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 238

gctcttcgct caccgcgtga attgctgtcc caaacgtaag catcatcgtg gctcggtcac 60
gcgatcctgg atccggggat cctagaccgc tggtggagag cgctgccgtc ggattggtgg 120

caagtaagat tgcgcaggtt ggcgaaggga gagaccaaaa ccggaggctg gaagcgggca 180

caacatcgta ttattgcgta tagtagagca gtggcagtcg catttcgagg tccgcaacgg 240

atctcgcaag ctcgctacgc tcacagtagg agaaagggga ccactgcccc tgccagaatg 300

gtcgcgaccc tctccctcgc cggccccgcc tgcaacacgc agtgcgtatc cggcaagcgg 360

gctgtcgcct tcaaccgccc ccatgttggc gtccgggctc gatcaggtgc gctgaggggg 420

gtttggtgtg cccgcgcctc tgggcccgtg tcggccgtgc ggacgtgggg ccctgggcag 480

tggatcagca gggtttgcgt gcaaatgcct ataccggcga ttgaatagcg atgaacggga 540

tacggttgcg ctcactccat gcccatgcga ccccgtttct gtccgccagc cgtggtcgcc 600

cgggctgcga agcgggaccc cacccagcgc attgtgatca ccggaatggg cgtggggtac 660

c 661

<210> 239

<211> 605

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 239

gagctccacc tgcatccgcc tggcgctcga ggacgccggc gtctcgcccg acgaggtcaa 60

ctacgtcaac gcgcacgcca cctccaccct ggtgggcgac aaggccgagg tgcgcgcggt 120

caagtcggtc tttggcgaca tgaagggcat caagatgaac gccaccaagt ccatgatcgg 180

gcactgcctg ggcgccgccg gcggcatgga ggccgtcgcc acgctcatgg ccatccgcac 240

cggctgggtg caccccacca tcaaccacga caaccccatc gccgaggtcg acggcctgga 300

cgtcgtcgcc aacgccaagg cccagcacaa aatcaacgtc gccatctcca actccttcgg 360

cttcggcggg cacaactccg tcgtcgcctt tgcgcccttc cgcgagtagg cggagcgagc 420

gcgcttggct gaggagggag gcggggtgcg agccctttgg ctgcgcgcga tactctcccc 480

gcacgagcag actccacgcg cctgaatcta cttgtcaacg agcaaccgtg tgttttgtcc 540

gtggccattc ttattatttc tccgactgtg gccgtactct gtttggctgt gcaagcaccg 600

aagag 605

<210> 240

<211> 3124

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 240

gctcttcgct ggcctgttgc cggacgatcc gtgtcgtcga gactgcattt tgttttgggt 60

gtggggctgg ggtactggat ggcttgaggg catgactttt tctgatggag aagattgcaa 120

tgagatcatt tgggtcgtct atttgtttgc tgtgcaagag ggtttactgg tatctggcac 180

cagcttttgg cccgtgcccg tttgatggac gcgtgacagg caggcgtcct ggaaagcaca 240

gacaccgtac gtacgacctt gacctccccc ccttctccac acggcaggtg cgaggctgcc 300

cacggcgtcg aggcgggcgg tgcgccgggc atggtcccgc atcgcgcgcg cggcggccgc 360

ggccgacgca aaccccgccc gccctgagcg ccgcgtggtc atcacgggcc agggcgtggt 420

gaccagcctg ggccagacga tcgagcagtt ttacagcagc ctgctggagg gcgtgagcgg 480

catctcgcag atacagaagt tcgacaccac gggctacacg acgacgatcg cgggcgagat 540

caagtcgctg cagctggacc cgtacgtgcc caagcgctgg gcgaagcgcg tggacgacgt 600

gataaagtac gtctacatcg cgggcaagca ggcgctggag agcgccggcc tgccgatcga 660

ggcggcgggg ctggcgggcg cggggctgga cccggcgctg tgcggcgtgc tcatcggcac 720

cgccatggcg ggcatgacgt ctttcgcggc gggcgtggag gcgctgacgc gcggtaccct 780

ttcttgcgct atgacacttc cagcaaaagg tagggcgggc tgcgagacgg cttcccggcg 840

ctgcatgcaa caccgatgat gcttcgaccc cccgaagctc cttcggggct gcatgggcgc 900

tccgatgccg ctccagggcg agcgctgttt aaatagccag gcccccgatt gcaaagacat 960

tatagcgagc taccaaagcc atattcaaac acctagatca ctaccacttc tacacaggcc 1020

actcgagctt gtgatcgcac tccgctaagg gggcgcctct tcctcttcgt ttcagtcaca 1080

acccgcaaac tctagaatat caatgatcga gcaggacggc ctccacgccg gctcccccgc 1140

cgcctgggtg gagcgcctgt tcggctacga ctgggcccag cagaccatcg gctgctccga 1200

cgccgccgtg ttccgcctgt ccgcccaggg ccgccccgtg ctgttcgtga agaccgacct 1260

gtccggcgcc ctgaacgagc tgcaggacga ggccgcccgc ctgtcctggc tggccaccac 1320

cggcgtgccc tgcgccgccg tgctggacgt ggtgaccgag gccggccgcg actggctgct 1380

gctgggcgag gtgcccggcc aggacctgct gtcctcccac ctggcccccg ccgagaaggt 1440

gtccatcatg gccgacgcca tgcgccgcct gcacaccctg gaccccgcca cctgcccctt 1500

cgaccaccag gccaagcacc gcatcgagcg cgcccgcacc cgcatggagg ccggcctggt 1560

ggaccaggac gacctggacg aggagcacca gggcctggcc cccgccgagc tgttcgcccg 1620

cctgaaggcc cgcatgcccg acggcgagga cctggtggtg acccacggcg acgcctgcct 1680

gcccaacatc atggtggaga acggccgctt ctccggcttc atcgactgcg gccgcctggg 1740

cgtggccgac cgctaccagg acatcgccct ggccacccgc gacatcgccg aggagctggg 1800

cggcgagtgg gccgaccgct tcctggtgct gtacggcatc gccgcccccg actcccagcg 1860

catcgccttc taccgcctgc tggacgagtt cttctgacaa ttggcagcag cagctcggat 1920

agtatcgaca cactctggac gctggtcgtg tgatggactg ttgccgccac acttgctgcc 1980

ttgacctgtg aatatccctg ccgcttttat caaacagcct cagtgtgttt gatcttgtgt 2040

gtacgcgctt ttgcgagttg ctagctgctt gtgctatttg cgaataccac ccccagcatc 2100

cccttccctc gtttcatatc gcttgcatcc caaccgcaac ttatctacgc tgtcctgcta 2160

tccctcagcg ctgctcctgc tcctgctcac tgcccctcgc acagccttgg tttgggctcc 2220

gcctgtattc tcctggtact gcaacctgta aaccagcact gcaatgctga tgcacgggaa 2280

gtagtgggat gggaacacaa atggaggatc cactagttct agagcggccg ccaccgcggt 2340

ggagctcggc ggcgtgcgca agatgaaccc cttttgcatc cccttctcca tctccaacat 2400

gggcggcgcg atgctggcga tggacatcgg cttcatgggc cccaactact ccatctccac 2460

ggcctgcgcg acgggcaact actgcatcct gggcgcggcg gaccacatcc ggcgcggcga 2520

cgcaaacgtg atgctggccg gcggcgcgga cgcggccatc atcccctcgg gcatcggcgg 2580

cttcatcgcg tgcaaggcgc tgagcaagcg caacgacgag cccgagcgcg cgagccggcc 2640

ctgggacgcc gaccgcgacg gcttcgtcat gggcgagggc gccggcgtgc tggtgctgga 2700

ggagctggag cacgccaagc gccgcggcgc gaccattttg gctgaattag ttggcggcgc 2760

ggccacctcg gacgcgcacc acatgaccga gcccgacccg cagggccgcg gcgtgcgcct 2820

ctgcctcgag cgcgcgctcg agcgcgcgcg cctcgcgccc gagcgcgtcg gctacgtcaa 2880

cgcgcacggc accagcacgc ccgcgggcga cgtggccgag taccgcgcca tccgcgccgt 2940

catcccgcag gactcactac gcatcaactc cacaaagtcc atgatcgggc acctgctcgg 3000

cggcgccggc gcggtcgagg ccgtggccgc catccaggcc ctgcgcaccg gctggctcca 3060

ccccaacttg aacctcgaga accccgcgcc tggcgtcgac cccgtcgtgc tcgtgggctc 3120

ttcc 3124

<210> 241

<211> 1173

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 241

actagtatgg gcgccggcgg ccgcatgtcc aacccctccg agggcgagaa gaagaccgag 60

ctggagggca tccagcgcgt gccctaccag aagccccccc tcaccgtggg cgacgggcgc 120

gccgtgaaga aggccatccc cccccactgc ttcaaccgct ccgtgatccg ctccttctcc 180

tacgtggtgt acgacctgac catcgcctcc atcctgtact acatcgccac caccttcatc 240

cccctgctgc cccacccccc ggcctacgtg gcctggccca tctactgggc cgtgcagggc 300

tgcgtgatga ccggcgtgtg ggtgatcgcc cacgagtgcg gccaccacgc cttctccgac 360

taccagtggc tggacgacac cgtgggcctg atcctgcact ccgtgctgct ggtgccctac 420

ttctcctgga agtactccca ccgccgccac cactccaaca ccggctccat cgagcacgac 480

gaggtgttcg tgcccaagct gaagtccggc gtgcgctcca ccgccaagta cctgaacaac 540

ccccccggcc gcatcctgac cctgctggtg accctgaccc tgggctggcc cctgtacctg 600

atgttcaacg tgtccggccg ctactacgac cgcttcgcct gccacttcga ccccaactcc 660

cccatctact ccaaccgcga gcgcgcccag atcttcatct ccgacgccgg catcttcgcc 720

gtgctgtacg gcctgtaccg cctggccgcc gtgaagggcc tggtgtgggt gctgaccgtg 780

tacgccggcc ccctgctggt ggtgaacggc ttcctggtgc tgatcacctt cctgcagcac 840

acccaccccc ccctgcccca ctacgactcc accgagtggg actggctgcg cggcgccctg 900

gccaccatcg accgcgacta cggcatcctg aacaaggtgt tccacaacat caccgacacc 960

cacgtgaccc accacctgtt ctccaccatg ccccactacc acgccatgga ggccaccaag 1020

gccatcatcc ccatcctggg cgactactac cagttcgacg gcacctccgt gttcaaggcc 1080

atgtaccgcg agaccaagga gtgcatctac gtggacaagg acgaggaggt gaaggacggc 1140

gtgtactggt accgcaacaa gatctagatc gat 1173

<210> 242

<211> 1173

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 242

actagtatgg gcgccggcgg ccgcaccgac gtgccccccg ccaaccgcaa gtccgaggtg 60

gacccccctga agcgcgtgcc cttcgagaag ccccccttct ccctgtccca ggggcgcgcc 120

atcaagaagg tgatcccccc ccactgcttc cagcgctccg tgttccgctc cttctcctac 180

gtggtgtacg acctgaccat cgccttctgc ctgtactacg tggccaccca ctacttccac 240

ctgctgccct ccccccctgtc cttcctggcc tggcccatct actgggccgt gcagggctgc 300

atcctgaccg gcgtgtgggt gatcgcccac gagtgcggcc accacgcctt ctccgactac 360

cagctgctgg acgacatcgt gggcctggtg ctgcactccg gcctgctggt gccctacttc 420

tcctggaagt actcccaccg ccgccaccac tccaacaccg gctccctgga gcgcgacgag 480

gtgttcgtgc ccaagcagaa gtcctgcatc aagtggtact ccaagtacct gaacaacccc 540

cccggccgcg tgctgaccct ggccgtgacc ctgaccctgg gctggcccct gtacctggcc 600

ctgaacgtgt ccggccgccc ctacgaccgc ttcgcctgcc actacgaccc ctacggcccc 660

atctactccg accgcgagcg cctgcagatc tacatctccg acgccggcgt gctggccgtg 720

tgctacggcc tgttccgcct ggccatggcc aagggcctgg cctgggtggt gtgcgtgtac 780

ggcgtgcccc tgctggtggt gaacggcttc ctggtgctga tcaccttcct gcagcacacc 840

caccccgccc tgccccacta cacctcctcc gagtgggact ggctgcgcgg cgccctggcc 900

accgtggacc gcgactacgg catcctggac aaggtgttcc acaacatcac cgacacccac 960

gtggcccacg agctgttctc caccatgccc cactaccacg ccatggaggc caccaaggcc 1020

atcaagccca tcctgggcga gtactaccgc ttcgacgaga ccccccttcgt gaaggccatg 1080

tggcgcgagg cccgcgagtg catctacgtg gagcccgacc agtccaccga gtccaagggc 1140

gtgttctggt acaacaacaa gctgtagatc gat 1173

<210> 243

<211> 1170

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 243

actagtatgg gcgccggcgg ccgcatgtcc tcccccaacg gcaaggagaa ggacggcccc 60

aagcccctgg agcgcgccct gcacgagaag ccccccttca ccgtgggcga cgggcgcgcc 120

atcaagaagg tgatcccccc ccactgcttc aagcgctccg tgatccgctc cttctcctac 180

gtggtgtacg acctgaccat cgcctccatc ttctactacc tggccaacaa ctacatcccc 240

ctgctgccca actccctggc ctacgtggcc tggcccgtgt actggatctt ccagggctgc 300

gtgctgaccg gcgtgtgggt gatcgcccac gagtgcggcc accacgcctt ctccgactac 360

cagtggctgg acgacaccgt gggcctgatc ctgcactccg ccctgctggt gccctacttc 420

tcctggaagt actcccaccg ccgccaccac tccaacaccg gctccatcga gcacgacgag 480

gtgttcgtgc ccaagctgaa gtcctccgtg cgctccaccg ccaagtacct gaacaacccc 540

cccggccgca tcctgaccct gctggtgacc ctgaccatgg gctggcccct gtacctgatg 600

ttcaacgtgt ccggccgcta ctacgaccgc ttcgcctgcc acttcgaccc caactccccc 660

atctactcca accgcgagcg cgcccagatc ttcatctccg acgccggcat cctgaccgtg 720

ttctacatcc tgttccgcct ggcctccacc aagggcctgg tgtgggtgct gaccatgtac 780

ggcggccccc tgctggtggt gaacggcttc ctggtgctga tcaccttcct gcagcacacc 840

cacccctccc tgccccacta cgactccacc gagtgggact ggctgcgcgg cgccctggcc 900

accgtggacc gcgactacgg catcctgaac aaggtgttcc acaacatcac cgacacccac 960

gtgacccacc acctgttctc caccatgccc cactaccacg ccatggaggc caccaaggcc 1020

atcaagccca tcctgggcga ctactaccag ttcgacggca cctccatctt caaggccatg 1080

taccgcgaga ccaaggagtg catctacgtg gacaaggacg aggacgtgaa ggacggcgtg 1140

tactggtacc gcaacaagat ctagatcgat 1170

<210> 244

<211> 1185

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 244

actagtatgg gcgccggcgg ccgcatgacc gagaaggagc gcgagaagca ggagcagctg 60

gcccgcgcca ccggcggcgc cgccatgcag cgctcccccg tggagaagcc ccccttcacc 120

ctgggccagg ggcgcgccat caagaaggcc atccccccccc actgcttcga gcgctccgtg 180

ctgaagtcct tctcctacgt ggtgcacgac ctggtgatcg ccgccgccct gctgtacttc 240

gccctggcca tcatccccgc cctgccctcc cccctgggct acgccgcctg gcccctgtac 300

tggatcgccc agggctgcgt gtgcaccggc gtgtgggtga tcgcccacga gtgcggccac 360

cacgccttct ccgactactc cctgctggac gacgtggtgg gcctggtgct gcactcctcc 420

ctgatggtgc cctacttctc ctggaagtac tcccaccgcc gccaccactc caacaccggc 480

tccctggagc gcgacgaggt gttcgtgccc aagaagaagg aggccctgcc ctggtacacc 540

ccctacgtgt acaacaaccc cgtgggccgc ctggcccaca tcgtggtgca gctgaccctg 600

ggctggcccc tgtacctggc caccaacgcc tccggccgcc cctacccccg cttcgcctgc 660

cacttcgacc cctacggccc catctacaac gaccgcgagc gcgcccagat cttcgtgtcc 720

gacgccggcg tggccgccgt ggccttcggc ctgtacaagc tggccgccgc cttcggcgtg 780

tggtgggtgg tgcgcgtgta cgccgtgccc ctgctgatcg tgaacgcctg gctggtgctg 840

atcacctacc tgcagcacac ccacccctcc ctgccccact acgactcctc cgagtgggac 900

tggctgcgcg gcgccctggc caccatggac cgcgactacg gcatcctgaa ccgcgtgttc 960

cacaacatca ccgacaccca cgtggcccac cacctgttct ccaccatgcc ccactaccac 1020

gccatggagg ccaccaaggc catccgcccc atcctgggcg actactacca cttcgacccc 1080

acccccgtgg ccaaggccac ctggcgcgag gccggcgagt gcatctacgt ggagcccgag 1140

gaccgcaagg gcgtgttctg gtacaacaag aagttctaga tcgat 1185

<210> 245

<211> 6756

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 245

gctcttcgcc gccgccactc ctgctcgagc gcgcccgcgc gtgcgccgcc agcgccttgg 60

ccttttcgcc gcgctcgtgc gcgtcgctga tgtccatcac caggtccatg aggtctgcct 120

tgcgccggct gagccactgc ttcgtccggg cggccaagag gagcatgagg gaggactcct 180

ggtccagggt cctgacgtgg tcgcggctct gggagcgggc cagcatcatc tggctctgcc 240

gcaccgaggc cgcctccaac tggtcctcca gcagccgcag tcgccgccga ccctggcaga 300

ggaagacagg tgaggggggt atgaattgta cagaacaacc acgagccttg tctaggcaga 360

atccctacca gtcatggctt tacctggatg acggcctgcg aacagctgtc cagcgaccct 420

cgctgccgcc gcttctcccg cacgcttctt tccagcaccg tgatggcgcg agccagcgcc 480

gcacgctggc gctgcgcttc gccgatctga ggacagtcgg ggaactctga tcagtctaaa 540

ccccccttgcg cgttagtgtt gccatccttt gcagaccggt gagagccgac ttgttgtgcg 600

ccacccccca caccacctcc tcccagacca attctgtcac ctttttggcg aaggcatcgg 660

cctcggcctg cagagaggac agcagtgccc agccgctggg ggttggcgga tgcacgctca 720

ggtacccttt cttgcgctat gacacttcca gcaaaaggta gggcgggctg cgagacggct 780

tcccggcgct gcatgcaaca ccgatgatgc ttcgaccccc cgaagctcct tcggggctgc 840

atgggcgctc cgatgccgct ccagggcgag cgctgtttaa atagccaggc ccccgattgc 900

aaagacatta tagcgagcta ccaaagccat attcaaacac ctagatcact accacttcta 960

cacaggccac tcgagcttgt gatcgcactc cgctaagggg gcgcctcttc ctcttcgttt 1020

cagtcacaac ccgcaaactc tagaatatca atgctgctgc aggccttcct gttcctgctg 1080

gccggcttcg ccgccaagat cagcgcctcc atgacgaacg agacgtccga ccgccccctg 1140

gtgcacttca cccccaacaa gggctggatg aacgacccca acggcctgtg gtacgacgag 1200

aaggacgcca agtggcacct gtacttccag tacaacccga acgacaccgt ctgggggacg 1260

ccctttgttct ggggccacgc cacgtccgac gacctgacca actgggagga ccagcccatc 1320

gccatcgccc cgaagcgcaa cgactccggc gccttctccg gctccatggt ggtggactac 1380

aacaacacct ccggcttctt caacgacacc atcgacccgc gccagcgctg cgtggccatc 1440

tggacctaca acaccccgga gtccgaggag cagtacatct cctacagcct ggacggcggc 1500

tacaccttca ccgagtacca gaagaacccc gtgctggccg ccaactccac ccagttccgc 1560

gacccgaagg tcttctggta cgagccctcc cagaagtgga tcatgaccgc ggccaagtcc 1620

caggactaca agatcgagat ctactcctcc gacgacctga agtcctggaa gctggagtcc 1680

gcgttcgcca acgagggctt cctcggctac cagtacgagt gccccggcct gatcgaggtc 1740

cccaccgagc aggaccccag caagtcctac tgggtgatgt tcatctccat caaccccggc 1800

gccccggccg gcggctcctt caaccagtac ttcgtcggca gcttcaacgg cacccacttc 1860

gaggccttcg acaaccagtc ccgcgtggtg gacttcggca aggactacta cgccctgcag 1920

accttcttca acaccgaccc gacctacggg agcgccctgg gcatcgcgtg ggcctccaac 1980

tgggagtact ccgccttcgt gcccaccaac ccctggcgct cctccatgtc cctcgtgcgc 2040

aagttctccc tcaacaccga gtaccaggcc aacccggaga cggagctgat caacctgaag 2100

gccgagccga tcctgaacat cagcaacgcc ggcccctgga gccggttcgc caccaacacc 2160

acgttgacga aggccaacag ctacaacgtc gacctgtcca acagcaccgg caccctggag 2220

ttcgagctgg tgtacgccgt caacaccacc cagacgatct ccaagtccgt gttcgcggac 2280

ctctccctct ggttcaaggg cctggaggac cccgaggagt acctccgcat gggcttcgag 2340

gtgtccgcgt cctccttctt cctggaccgc gggaacagca aggtgaagtt cgtgaaggag 2400

aacccctact tcaccaaccg catgagcgtg aacaaccagc ccttcaagag cgagaacgac 2460

ctgtcctact acaaggtgta cggcttgctg gaccagaaca tcctggagct gtacttcaac 2520

gacggcgacg tcgtgtccac caacacctac ttcatgacca ccgggaacgc cctgggctcc 2580

gtgaacatga cgacgggggt ggacaacctg ttctacatcg acaagttcca ggtgcgcgag 2640

gtcaagtgac aattggcagc agcagctcgg atagtatcga cacactctgg acgctggtcg 2700

tgtgatggac tgttgccgcc acacttgctg ccttgacctg tgaatatccc tgccgctttt 2760

atcaaacagc ctcagtgtgt ttgatcttgt gtgtacgcgc ttttgcgagt tgctagctgc 2820

ttgtgctatt tgcgaatacc acccccagca tccccttccc tcgtttcata tcgcttgcat 2880

cccaaccgca acttatctac gctgtcctgc tatccctcag cgctgctcct gctcctgctc 2940

actgcccctc gcacagcctt ggtttgggct ccgcctgtat tctcctggta ctgcaacctg 3000

taaaccagca ctgcaatgct gatgcacggg aagtagtggg atgggaacac aaatggagga 3060

tcccgcgtct cgaacagagc gcgcagagga acgctgaagg tctcgcctct gtcgcacctc 3120

agcgcggcat acaccacaat aaccacctga cgaatgcgct tggttcttcg tccattagcg 3180
aagcgtccgg ttcacacacg tgccacgttg gcgaggtggc aggtgacaat gatcggtgga 3240
gctgatggtc gaaacgttca cagcctaggg atatcgaatt cggccgacag gacgcgcgtc 3300
aaaggtgctg gtcgtgtatg ccctggccgg caggtcgttg ctgctgctgg ttagtgattc 3360
cgcaaccctg attttggcgt cttattttgg cgtggcaaac gctggcgccc gcgagccggg 3420
ccggcggcga tgcggtgccc cacggctgcc ggaatccaag ggaggcaaga gcgcccgggt 3480
cagttgaagg gctttacgcg caaggtacag ccgctcctgc aaggctgcgt ggtggaattg 3540

gacgtgcagg tcctgctgaa gttcctccac cgcctcacca gcggacaaag caccggtgta 3600
tcaggtccgt gtcatccact ctaaagaact cgactacgac ctactgatgg ccctagattc 3660
ttcatcaaaa acgcctgaga cacttgccca ggattgaaac tccctgaagg gaccaccagg 3720
ggccctgagt tgttccttcc ccccgtggcg agctgccagc caggctgtac ctgtgatcga 3780
ggctggcggg aaaataggct tcgtgtgctc aggtcatggg aggtgcagga cagctcatga 3840
aacgccaaca atcgcacaat tcatgtcaag ctaatcagct atttcctctt cacgagctgt 3900
aattgtccca aaattctggt ctaccggggg tgatccttcg tgtacgggcc cttccctcaa 3960

ccctaggtat gcgcgcatgc ggtcgccgcg caactcgcgc gagggccgag ggtttgggac 4020
gggccgtccc gaaatgcagt tgcacccgga tgcgtggcac ctttttgcg ataatttatg 4080
caatggactg ctctgcaaaa ttctggctct gtcgccaacc ctaggatcag cggcgtagga 4140
tttcgtaatc attcgtcctg atggggagct accgactacc ctaatatcag cccgactgcc 4200
tgacgccagc gtccactttt gtgcacacat tccattcgtg cccaagacat ttcattgtgg 4260
tgcgaagcgt ccccagttac gctcacctgt ttcccgacct ccttactgtt ctgtcgacag 4320
agcgggccca caggccggtc gcagccacta gtatgggcgc cggcggccgc atgtccgtgc 4380

ccacctcccc caagaagccc gagttcaact ccctgaagcg cgtgccctac tccaagcccc 4440
ccttcaccct gtccgagggg cgcgccatgg actacaagga ccacgacggc gactacaagg 4500
accacgacat cgactacaag gacgacgacg acaagatcaa gaaggccatc ccccccccact 4560
gcttccagcg ctccgtgctg cgctccttct cctacctgct gtacgacttc atcctggcct 4620
ccctgttcta ccacgtggcc accaactact tccccaacct gccccaggcc ctgtccaacg 4680
tggcctggcc cctgtactgg gccatgcagg gctgcatcct gaccggcgtg tgggtgatcg 4740
cccacgagtg cggccaccac gccttctccg actaccagtg gctggacgac accgtgggcc 4800

tgatcctgca ctcctccctg ctggtgccct acttctcctg gaagtactcc caccgccgcc 4860

accactccaa caccggctcc ctggagcgcg acgaggtgtt cgtgcccaag aagaagtccg 4920

gcctgcgctg gtgggccaag cacttcaaca acccccccgg ccgcttcctg tccatcacca 4980

tccagctgac cctgggctgg cccctgtacc tggccttcaa cgtggccggc cgcccctacg 5040

accgcttcgc ctgccactac gacccctacg gccccatctt ctccgaccgc gagcgcctgc 5100

agatctacat ctccgacgcc ggcgtgctgg ccgtggccta cgccctgtac cgcctggtgc 5160

tggccaaggg cgtgggctgg gtgatctccg tgtacggcgt gcccctgctg gtggtgaacg 5220

ccttcctggt gatgatcacc tacctgcagc acacccaccc ctccctgccc cactacgact 5280

cctccgagtg ggactggatg cgcggcgccc tgtccaccgt ggaccgcgac tacggcatcc 5340

tgaacaaggt gttccacaac atcaccgaca cccacgtggc ccaccacctg ttctccacca 5400

tgccccacta ccacgccatg gtggccacca aggccatcaa gcccatcctg ggcgagtact 5460

accagttcga cggcatgccc gtgtacaagg ccatctggcg cgaggccaag gagtgcctgt 5520

acgtggagcc cgacgagggc gacaaggaca agggcgtgtt ctggttccgc aacaagctgt 5580

agatcgatct taaggcagca gcagctcgga tagtatcgac acactctgga cgctggtcgt 5640

gtgatggact gttgccgcca cacttgctgc cttgacctgt gaatatccct gccgctttta 5700

tcaaacagcc tcagtgtgtt tgatcttgtg tgtacgcgct tttgcgagtt gctagctgct 5760

tgtgctattt gcgaatacca cccccagcat ccccttccct cgtttcatat cgcttgcatc 5820

ccaaccgcaa cttatctacg ctgtcctgct atccctcagc gctgctcctg ctcctgctca 5880

ctgcccctcg cacagccttg gtttgggctc cgcctgtatt ctcctggtac tgcaacctgt 5940

aaaccagcac tgcaatgctg atgcacggga agtagtggga tgggaacaca aatggaaagc 6000

ttaattaaga gctcttgttt tccagaagga gttgctcctt gagcctttca ttctcagcct 6060

cgataacctc caaagccgct ctaattgtgg agggggttcg aatttaaaag cttggaatgt 6120

tggttcgtgc gtctggaaca agcccagact tgttgctcac tgggaaaagg accatcagct 6180

ccaaaaaact tgccgctcaa accgcgtacc tctgctttcg cgcaatctgc cctgttgaaa 6240

tcgccaccac attcatattg tgacgcttga gcagtctgta attgcctcag aatgtggaat 6300

catctgcccc ctgtgcgagc ccatgccagg catgtcgcgg gcgaggacac ccgccactcg 6360

tacagcagac cattatgcta cctcacaata gttcataaca gtgaccatat ttctcgaagc 6420

tccccaacga gcacctccat gctctgagtg gccacccccc ggccctggtg cttgcggagg 6480

gcaggtcaac cggcatgggg ctaccgaaat ccccgaccgg atcccaccac ccccgcgatg 6540

ggaagaatct ctccccggga tgtgggccca ccaccagcac aacctgctgg cccaggcgag 6600

cgtcaaacca taccacacaa atatccttgg catcggccct gaattccttc tgccgctctg 6660

ctacccggtg cttctgtccg aagcaggggt tgctagggat cgctccgagt ccgcaaaccc 6720

ttgtcgcgtg gcggggcttg ttcgagcttg aagagc 6756

<210> 246

<211> 9416

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 246

cactggcagc agccactggt aacaggatta gcagagcgag gtatgtaggc ggtgctacag 60

agttcttgaa gtggtggcct aactacggct acactagaag gacagtattt ggtatctgcg 120

ctctgctgaa gccagttacc ttcggaaaaa gagttggtag ctcttgatcc ggcaaacaaa 180

ccaccgctgg tagcggtggt ttttttgttt gcaagcagca gattacgcgc agaaaaaaag 240

gatctcaaga agatcctttg atcttttcta cggggtctga cgctcagtgg aacgaaaact 300

cacgttaagg gattttggtc atgagattat caaaaaggat cttcacctag atccttttaa 360

attaaaaatg aagttttaaa tcaatctaaa gtatatatga gtaaacttgg tctgacagtt 420

accaatgctt aatcagtgag gcacctatct cagcgatctg tctatttcgt tcatccatag 480

ttgcctgact ccccgtcgtg tagataacta cgatacggga gggcttacca tctggcccca 540

gtgctgcaat gataccgcga gacccacgct caccggctcc agatttatca gcaataaacc 600

agccagccgg aagggccgag cgcagaagtg gtcctgcaac tttatccgcc tccatccagt 660

ctattaattg ttgccgggaa gctagagtaa gtagttcgcc agttaatagt ttgcgcaacg 720

ttgttgccat tgctacaggc atcgtggtgt cacgctcgtc gtttggtatg gcttcattca 780

gctccggttc ccaacgatca aggcgagtta catgatcccc catgttgtgc aaaaaagcgg 840

ttagctcctt cggtcctccg atcgttgtca gaagtaagtt ggccgcagtg ttatcactca 900

tggttatggc agcactgcat aattctctta ctgtcatgcc atccgtaaga tgcttttctg 960

tgactggtga gtactcaacc aagtcattct gagaatagtg tatgcggcga ccgagttgct 1020

cttgcccggc gtcaatacgg gataataccg cgccacatag cagaacttta aaagtgctca 1080

tcattggaaa acgttcttcg gggcgaaaac tctcaaggat cttaccgctg ttgagatcca 1140

gttcgatgta acccactcgt gcacccaact gatcttcagc atcttttact ttcaccagcg 1200

tttctgggtg agcaaaaaca ggaaggcaaa atgccgcaaa aaagggaata agggcgacac 1260

ggaaatgttg aatactcata ctcttccttt ttcaatatta ttgaagcatt tatcagggtt 1320

attgtctcat gagcggatac atatttgaat gtatttagaa aaataaacaa atagggggttc 1380

cgcgcacatt tccccgaaaa gtgccacctg acgtctaaga aaccattatt atcatgacat 1440

taacctataa aaataggcgt atcacgaggc cctttcgtct cgcgcgtttc ggtgatgacg 1500

gtgaaaacct ctgacacatg cagctcccgg agacggtcac agcttgtctg taagcggatg 1560

ccgggagcag acaagcccgt cagggcgcgt cagcgggtgt tggcgggtgt cggggctggc 1620

ttaactatgc ggcatcagag cagattgtac tgagagtgca ccatatgcgg tgtgaaatac 1680

cgcacagatg cgtaaggaga aaataccgca tcaggcgcca ttcgccattc aggctgcgca 1740

actgttggga agggcgatcg gtgcgggcct cttcgctatt acgccagctg gcgaaagggg 1800

gatgtgctgc aaggcgatta agttgggtaa cgccagggtt ttcccagtca cgacgttgta 1860

aaacgacggc cagtgaattg atgcatgctc ttcgggctgg tctgaatcct tcaggcgggt 1920

gttacccgag aaagaaaggg tgccgatttc aaagcagacc catgtgccgg gccctgtggc 1980

ctgtgttggc gcctatgtag tcacccccccc tcacccaatt gtcgccagtt tgcgcactcc 2040

ataaactcaa aacagcagct tctgagctgc gctgttcaag aacacctctg gggtttgctc 2100

acccgcgagg tcgacgccca gcatggctat caagacgaac aggcagcctg tggagaagcc 2160

tccgttcacg atcgggacgc tgcgcaaggc catccccgcg cactgtttcg agcgctcggc 2220

gcttcgtagc agcatgtacc tggcctttga catcgcggtc atgtccctgc tctacgtcgc 2280

gtcgacgtac atcgaccctg caccggtgcc tacgtgggtc aagtacggca tcatgtggcc 2340

gctctactgg ttcttccagg tgtgtttgag ggttttggtt gcccgtattg aggtcctggt 2400

ggcgcgcatg gaggagaagg cgcctgtccc gctgaccccc ccggctaccc tcccggcacc 2460

ttccagggcg ccttcggcac gggtgtctgg gtgtgcgcgc acgagtgcgg ccaccaggcc 2520

ttttcctcca gccaggccat caacgacggc gtgggcctgg tgttccacag cctgctgctg 2580

gtgccctact actcctggaa gcactcgcac cgggtaccct ttcttgcgct atgacacttc 2640

cagcaaaagg tagggcgggc tgcgagacgg cttcccggcg ctgcatgcaa caccgatgat 2700

gcttcgaccc cccgaagctc cttcggggct gcatgggcgc tccgatgccg ctccagggcg 2760

agcgctgttt aaatagccag gcccccgatt gcaaagacat tatagcgagc taccaaagcc 2820

atattcaaac acctagatca ctaccacttc tacacaggcc actcgagctt gtgatcgcac 2880

tccgctaagg gggcgcctct tcctcttcgt ttcagtcaca acccgcaaac tctagaatat 2940

caatgctgct gcaggccttc ctgttcctgc tggccggctt cgccgccaag atcagcgcct 3000

ccatgacgaa cgagacgtcc gaccgccccc tggtgcactt cacccccaac aagggctgga 3060

tgaacgaccc caacggcctg tggtacgacg agaaggacgc caagtggcac ctgtacttcc 3120

agtacaaccc gaacgacacc gtctgggggga cgcccttgtt ctggggccac gccacgtccg 3180

acgacctgac caactgggag gaccagccca tcgccatcgc cccgaagcgc aacgactccg 3240

gcgccttctc cggctccatg gtggtggact acaacaacac ctccggcttc ttcaacgaca 3300

ccatcgaccc gcgccagcgc tgcgtggcca tctggaccta caacaccccg gagtccgagg 3360

agcagtacat ctcctacagc ctggacggcg gctacacctt caccgagtac cagaagaacc 3420

ccgtgctggc cgccaactcc acccagttcc gcgacccgaa ggtcttctgg tacgagccct 3480

cccagaagtg gatcatgacc gcggccaagt cccaggacta caagatcgag atctactcct 3540

ccgacgacct gaagtcctgg aagctggagt ccgcgttcgc caacgagggc ttcctcggct 3600

accagtacga gtgccccggc ctgatcgagg tccccaccga gcaggacccc agcaagtcct 3660

actgggtgat gttcatctcc atcaaccccg gcgccccggc cggcggctcc ttcaaccagt 3720

acttcgtcgg cagcttcaac ggcacccact tcgaggcctt cgacaaccag tcccgcgtgg 3780

tggacttcgg caaggactac tacgccctgc agaccttctt caacaccgac ccgacctacg 3840

ggagcgccct gggcatcgcg tgggcctcca actgggagta ctccgccttc gtgcccacca 3900

acccctggcg ctcctccatg tccctcgtgc gcaagttctc cctcaacacc gagtaccagg 3960

ccaacccgga gacggagctg atcaacctga aggccgagcc gatcctgaac atcagcaacg 4020

ccggcccctg gagccggttc gccaccaaca ccacgttgac gaaggccaac agctacaacg 4080

tcgacctgtc caacagcacc ggcaccctgg agttcgagct ggtgtacgcc gtcaacacca 4140

cccagacgat ctccaagtcc gtgttcgcgg acctctccct ctggttcaag ggcctggagg 4200

accccgagga gtacctccgc atgggcttcg aggtgtccgc gtcctccttc ttcctggacc 4260

gcgggaacag caaggtgaag ttcgtgaagg agaaccccta cttcaccaac cgcatgagcg 4320

tgaacaacca gcccttcaag agcgagaacg acctgtccta ctacaaggtg tacggcttgc 4380

tggaccagaa catcctggag ctgtacttca acgacggcga cgtcgtgtcc accaacacct 4440

acttcatgac caccgggaac gccctgggct ccgtgaacat gacgacgggg gtggacaacc 4500

tgttctacat cgacaagttc caggtgcgcg aggtcaagtg acaattggca gcagcagctc 4560

ggatagtatc gacacactct ggacgctggt cgtgtgatgg actgttgccg ccacacttgc 4620

tgccttgacc tgtgaatatc cctgccgctt ttatcaaaca gcctcagtgt gtttgatctt 4680

gtgtgtacgc gcttttgcga gttgctagct gcttgtgcta tttgcgaata ccacccccag 4740

catccccttc cctcgtttca tatcgcttgc atcccaaccg caacttatct acgctgtcct 4800

gctatccctc agcgctgctc ctgctcctgc tcactgcccc tcgcacagcc ttggtttggg 4860

ctccgcctgt attctcctgg tactgcaacc tgtaaaccag cactgcaatg ctgatgcacg 4920

ggaagtagtg ggatgggaac acaaatggag gatcccgcgt ctcgaacaga gcgcgcagag 4980

gaacgctgaa ggtctcgcct ctgtcgcacc tcagcgcggc atacaccaca ataaccacct 5040

gacgaatgcg cttggttctt cgtccattag cgaagcgtcc ggttcacaca cgtgccacgt 5100

tggcgaggtg gcaggtgaca atgatcggtg gagctgatgg tcgaaacgtt cacagcctag 5160
ggatatcgaa ttcggccgac aggacgcgcg tcaaaggtgc tggtcgtgta tgccctggcc 5220
ggcaggtcgt tgctgctgct ggttagtgat tccgcaaccc tgattttggc gtcttatttt 5280
ggcgtggcaa acgctggcgc ccgcgagccg ggccggcggc gatgcggtgc cccacggctg 5340
ccggaatcca agggaggcaa gagcgcccgg gtcagttgaa gggctttacg cgcaaggtac 5400
agccgctcct gcaaggctgc gtggtggaat tggacgtgca ggtcctgctg aagttcctcc 5460
accgcctcac cagcggacaa agcaccggtg tatcaggtcc gtgtcatcca ctctaaagaa 5520

ctcgactacg acctactgat ggccctagat tcttcatcaa aaacgcctga gacacttgcc 5580
caggattgaa actccctgaa gggaccacca ggggccctga gttgttcctt ccccccgtgg 5640
cgagctgcca gccaggctgt acctgtgatc gaggctggcg ggaaaatagg cttcgtgtgc 5700
tcaggtcatg ggaggtgcag gacagctcat gaaacgccaa caatcgcaca attcatgtca 5760
agctaatcag ctatttcctc ttcacgagct gtaattgtcc caaaattctg gtctaccggg 5820
ggtgatcctt cgtgtacggg cccttccctc aaccctaggt atgcgcgcat gcggtcgccg 5880
cgcaactcgc gcgagggccg agggtttggg acgggccgtc ccgaaatgca gttgcacccg 5940

gatgcgtggc acctttttg cgataattta tgcaatggac tgctctgcaa aattctggct 6000
ctgtcgccaa ccctaggatc agcggcgtag gatttcgtaa tcattcgtcc tgatggggag 6060
ctaccgacta ccctaatatc agcccgactg cctgacgcca gcgtccactt ttgtgcacac 6120
attccattcg tgcccaagac atttcattgt ggtgcgaagc gtccccagtt acgctcacct 6180
gtttcccgac ctccttactg ttctgtcgac agagcgggcc cacaggccgg tcgcagccac 6240
tagtatggcc accgcatcca ctttctcggc gttcaatgcc cgctgcggcg acctgcgtcg 6300
ctcggcgggc tccgggcccc ggcgcccagc gaggcccctc cccgtgcgcg ggcgcgccga 6360

ggtgcacgtg caggtgaccc actccctggc ccccgagaag cgcgagatct tcaactccct 6420
gaacaactgg gcccaggaga acatcctggt gctgctgaag gacgtggaca agtgctggca 6480
gccctccgac ttcctgcccg actccgcctc cgagggcttc gacgagcagg tgatggagct 6540
gcgcaagcgc tgcaaggaga tccccgacga ctacttcatc gtgctggtgg gcgacatgat 6600
caccgaggag gccctgccca cctaccagac catgctgaac accctggacg gcgtgcgcga 6660
cgagaccggc gcctccctga ccccctgggc catctggacc cgcgcctgga ccgccgagga 6720
gaaccgccac ggcgacctgc tgaacaagta cctgtacctg tccggccgcg tggacatgaa 6780

gcagatcgag aagaccatcc agtacctgat cggctccggc atggaccccc gcaccgagaa 6840

caacccctac ctgggcttca tctacacctc cttccaggag cgcgccacct tcatctccca 6900

cggcaacacc gcccgcctgg ccaaggagca cggcgacctg aagctggccc agatctgcgg 6960

catcatcgcc gccgacgaga agcgccacga gaccgcctac accaagatcg tggagaagct 7020

gttcgagatc gaccccgacg gcaccgtgct ggccctggcc gacatgatgc gcaagaaggt 7080

gtccatgccc gcccacctga tgtacgacgg ccaggacgac aacctgttcg agaacttctc 7140

ctccgtggcc cagcgcctgg gcgtgtacac cgccaaggac tacgccgaca tcctggagtt 7200

cctggtgggc cgctgggaca tcgagaagct gaccggcctg tccggcgagg gccgcaaggc 7260

ccaggactac gtgtgcaccc tgccccccgg catccgccgc ctggaggagc gcgcccagtc 7320

ccgcgtgaag aaggcctccg ccacccccct ctcctggatc ttcggccgcg agatcaacct 7380

gatggactac aaggaccacg acggcgacta caaggaccac gacatcgact acaaggacga 7440

cgacgacaag tgaatcgata gatctcttaa ggcagcagca gctcggatag tatcgacaca 7500

ctctggacgc tggtcgtgtg atggactgtt gccgccacac ttgctgcctt gacctgtgaa 7560

tatccctgcc gcttttatca aacagcctca gtgtgtttga tcttgtgtgt acgcgctttt 7620

gcgagttgct agctgcttgt gctatttgcg aataccaccc ccagcatccc cttccctcgt 7680

ttcatatcgc ttgcatccca accgcaactt atctacgctg tcctgctatc cctcagcgct 7740

gctcctgctc ctgctcactg cccctcgcac agccttggtt tgggctccgc ctgtattctc 7800

ctggtactgc aacctgtaaa ccagcactgc aatgctgatg cacgggaagt agtgggatgg 7860

gaacacaaat ggaaagctta attaagagct cccgccacca ctccaacacg gggtgcctgg 7920

acaaggacga ggtgtttgtg ccgccgcacc gcgcagtggc gcacgagggc ctggagtggg 7980

aggagtggct gcccatccgc atgggcaagg tgctggtcac cctgaccctg ggctggccgc 8040

tgtacctcat gttcaacgtc gcctcgcggc cgtacccgcg cttcgccaac cactttgacc 8100

cgtggtcgcc catcttcagc aagcgcgagc gcatcgaggt ggtcatctcc gacctggcgc 8160

tggtggcggt gctcagcggg ctcagcgtgc tgggccgcac catgggctgg gcctggctgg 8220

tcaagaccta cgtggtgccc tacctgatcg tgaacatgtg gctcgtgctc atcacgctgc 8280

tccagcacac gcacccggcg ctgccgcact acttcgagaa ggactgggac tggctgcgcg 8340

gcgccatggc caccgtggac cgctccatgg gcccgccctt catggacaac atcctgcacc 8400

acatctccga cacccacgtg ctgcaccacc tcttcagcac catcccgcac taccacgccg 8460

aggaggcctc cgccgccatc aggcccatcc tgggcaagta ctaccagtcc gacagccgct 8520

gggtcggccg cgccctgtgg gaggactggc gcgactgccg ctacgtcgtc ccggacgcgc 8580

ccgaggacga ctccgcgctc tggttccaca agtgagtgag tgagaagagc ctctagagtc 8640

gacctgcagg catgcaagct tggcgtaatc atggtcatag ctgtttcctg tgtgaaattg 8700

ttatccgctc acaattccac acaacatacg agccggaagc ataaagtgta aagcctgggg 8760

tgcctaatga gtgagctaac tcacattaat tgcgttgcgc tcactgcccg ctttccagtc 8820

gggaaacctg tcgtgccagc tgcattaatg aatcggccaa cgcgcgggga gaggcggttt 8880

gcgtattggg cgctcttccg cttcctcgct cactgactcg ctgcgctcgg tcgttcggct 8940

gcggcgagcg gtatcagctc actcaaaggc ggtaatacgg ttatccacag aatcagggga 9000

taacgcagga aagaacatgt gagcaaaagg ccagcaaaag gccaggaacc gtaaaaaggc 9060

cgcgttgctg gcgtttttcc ataggctccg ccccccctgac gagcatcaca aaaatcgacg 9120

ctcaagtcag aggtggcgaa acccgacagg actataaaga taccaggcgt ttcccccctgg 9180

aagctccctc gtgcgctctc ctgttccgac cctgccgctt accggatacc tgtccgcctt 9240

tctcccttcg ggaagcgtgg cgctttctca tagctcacgc tgtaggtatc tcagttcggt 9300

gtaggtcgtt cgctccaagc tgggctgtgt gcacgaaccc cccgttcagc ccgaccgctg 9360

cgccttatcc ggtaactatc gtcttgagtc caacccggta agacacgact tatcgc 9416

<210> 247

<211> 719

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 247

gggctggtct gaatccttca ggcgggtgtt acccgagaaa gaaagggtgc cgatttcaaa 60

gcagacccat gtgccgggcc ctgtggcctg tgttggcgcc tatgtagtca ccccccctca 120

cccaattgtc gccagtttgc gcactccata aactcaaaac agcagcttct gagctgcgct 180

gttcaagaac acctctgggg tttgctcacc cgcgaggtcg acgcccagca tggctatcaa 240

gacgaacagg cagcctgtgg agaagcctcc gttcacgatc gggacgctgc gcaaggccat 300

ccccgcgcac tgtttcgagc gctcggcgct tcgtagcagc atgtacctgg cctttgacat 360

cgcggtcatg tccctgctct acgtcgcgtc gacgtacatc gaccctgcac cggtgcctac 420

gtgggtcaag tacggcatca tgtggccgct ctactggttc ttccaggtgt gtttgagggt 480

tttggttgcc cgtattgagg tcctggtggc gcgcatggag gagaaggcgc ctgtcccgct 540

gacccccccg gctaccctcc cggcaccttc cagggcgcct tcggcacggg tgtctgggtg 600

tgcgcgcacg agtgcggcca ccaggccttt tcctccagcc aggccatcaa cgacggcgtg 660

ggcctggtgt tccacagcct gctgctggtg ccctactact cctggaagca ctcgcaccg 719

<210> 248

<211> 732

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 248

ccgccaccac tccaacacgg ggtgcctgga caaggacgag gtgtttgtgc cgccgcaccg 60

cgcagtggcg cacgagggcc tggagtggga ggagtggctg cccatccgca tgggcaaggt 120

gctggtcacc ctgaccctgg gctggccgct gtacctcatg ttcaacgtcg cctcgcggcc 180

gtacccgcgc ttcgccaacc actttgaccc gtggtcgccc atcttcagca agcgcgagcg 240

catcgaggtg gtcatctccg acctggcgct ggtggcggtg ctcagcgggc tcagcgtgct 300

gggccgcacc atgggctggg cctggctggt caagacctac gtggtgccct acctgatcgt 360

gaacatgtgg ctcgtgctca tcacgctgct ccagcacacg cacccggcgc tgccgcacta 420

cttcgagaag gactgggact ggctgcgcgg cgccatggcc accgtggacc gctccatggg 480

cccgcccttc atggacaaca tcctgcacca catctccgac acccacgtgc tgcaccacct 540

cttcagcacc atcccgcact accacgccga ggaggcctcc gccgccatca ggcccatcct 600

gggcaagtac taccagtccg acagccgctg ggtcggccgc gccctgtggg aggactggcg 660

cgactgccgc tacgtcgtcc cggacgcgcc cgaggacgac tccgcgctct ggttccacaa 720

gtgagtgagt ga 732

<210> 249

<211> 12342

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 249

tagaaaaact catcgagcat caaatgaaac tgcaatttat tcatatcagg attatcaata 60

ccatattttt gaaaaagccg tttctgtaat gaaggagaaa actcaccgag gcagttccat 120

aggatggcaa gatcctggta tcggtctgcg attccgactc gtccaacatc aatacaacct 180

attaatttcc cctcgtcaaa aataaggtta tcaagtgaga aatcaccatg agtgacgact 240

gaatccggtg agaatggcaa aagtttatgc atttctttcc agacttgttc aacaggccag 300

ccattacgct cgtcatcaaa atcactcgca tcaaccaaac cgttattcat tcgtgattgc 360

gcctgagcga ggcgaaatac gcgatcgctg ttaaaaggac aattacaaac aggaatcgag 420

tgcaaccggc gcaggaacac tgccagcgca tcaacaatat tttcacctga atcaggatat 480

tcttctaata cctggaacgc tgtttttccg gggatcgcag tggtgagtaa ccatgcatca 540

tcaggagtac ggataaaatg cttgatggtc ggaagtggca taaattccgt cagccagttt 600

agtctgacca tctcatctgt aacatcattg gcaacgctac ctttgccatg tttcagaaac 660

aactctggcg catcgggctt cccatacaag cgatagattg tcgcacctga ttgcccgaca 720

ttatcgcgag cccatttata cccatataaa tcagcatcca tgttggaatt taatcgcggc 780

ctcgacgttt cccgttgaat atggctcata ttcttccttt ttcaatatta ttgaagcatt 840

tatcagggtt attgtctcat gagcggatac atatttgaat gtatttagaa aaataaacaa 900

ataggggtca gtgttacaac caattaacca attctgaaca ttatcgcgag cccatttata 960

cctgaatatg gctcataaca ccccttgttt gcctggcggc agtagcgcgg tggtcccacc 1020

tgaccccatg ccgaactcag aagtgaaacg ccgtagcgcc gatggtagtg tggggactcc 1080

ccatgcgaga gtagggaact gccaggcatc aaataaaacg aaaggctcag tcgaaagact 1140

gggcctttcg cccgggctaa ttagggggtg tcgcccttat tcgactctat agtgaagttc 1200

ctattctcta gaaagtatag gaacttctga agtggggaag aatgcatggc tcttccggcc 1260

tgttgccgga cgatccgtgt cgtcgagact gcattttgtt ttgggtgtgg ggctggggta 1320

ctggatggct tgagggcatg actttttctg atggagaaga ttgcaatgag atcatttggg 1380

tcgtctattt gtttgctgtg caagagggtt tactggtatc tggcaccagc ttttggcccg 1440

tgcccgtttg atggacgcgt gacaggcagg cgtcctggaa agcacagaca ccgtacgtac 1500

gaccttgacc tccccccctt ctccacacgg caggtgcgag gctgcccacg gcgtcgaggc 1560

gggcggtgcg ccgggcatgg tcccgcatcg cgcgcgcggc ggccgcggcc gacgcaaacc 1620

ccgcccgccc tgagcgccgc gtggtcatca cgggccaggg cgtggtgacc agcctgggcc 1680

agacgatcga gcagttttac agcagcctgc tggagggcgt gagcggcatc tcgcagatac 1740

agaagttcga caccacgggc tacacgacga cgatcgcggg cgagatcaag tcgctgcagc 1800

tggacccgta cgtgcccaag cgctgggcga agcgcgtgga cgacgtgata aagtacgtct 1860

acatcgcggg caagcaggcg ctggagagcg ccggcctgcc gatcgaggcg gcggggctgg 1920

cgggcgcggg gctggacccg gcgctgtgcg gcgtgctcat cggcaccgcc atggcgggca 1980

tgacgtcttt cgcggcgggc gtggaggcgc tgacgcgcgg taccctttct tgcgctatga 2040

cacttccagc aaaaggtagg gcgggctgcg agacggcttc ccggcgctgc atgcaacacc 2100

gatgatgctt cgaccccccg aagctccttc ggggctgcat gggcgctccg atgccgctcc 2160

agggcgagcg ctgtttaaat agccaggccc ccgattgcaa agacattata gcgagctacc 2220

aaagccatat tcaaacacct agatcactac cacttctaca caggccactc gagcttgtga 2280

tcgcactccg ctaaggggggc gcctcttcct cttcgtttca gtcacaaccc gcaaacggcg 2340

cgccatgctg ctgcaggcct tcctgttcct gctggccggc ttcgccgcca agatcagcgc 2400

ctccatgacg aacgagacgt ccgaccgccc cctggtgcac ttcaccccca acaagggctg 2460

gatgaacgac cccaacggcc tgtggtacga cgagaaggac gccaagtggc acctgtactt 2520

ccagtacaac ccgaacgaca ccgtctgggg gacgcccttg ttctggggcc acgccacgtc 2580

cgacgacctg accaactggg aggaccagcc catcgccatc gccccgaagc gcaacgactc 2640

cggcgccttc tccggctcca tggtggtgga ctacaacaac acctccggct tcttcaacga 2700

caccatcgac ccgcgccagc gctgcgtggc catctggacc tacaacaccc cggagtccga 2760

ggagcagtac atctcctaca gcctggacgg cggctacacc ttcaccgagt accagaagaa 2820

ccccgtgctg gccgccaact ccacccagtt ccgcgacccg aaggtcttct ggtacgagcc 2880

ctcccagaag tggatcatga ccgcggccaa gtcccaggac tacaagatcg agatctactc 2940

ctccgacgac ctgaagtcct ggaagctgga gtccgcgttc gccaacgagg gcttcctcgg 3000

ctaccagtac gagtgccccg gcctgatcga ggtccccacc gagcaggacc ccagcaagtc 3060

ctactgggtg atgttcatct ccatcaaccc cggcgccccg gccggcggct ccttcaacca 3120

gtacttcgtc ggcagcttca acggcaccca cttcgaggcc ttcgacaacc agtcccgcgt 3180

ggtggacttc ggcaaggact actacgccct gcagaccttc ttcaacaccg acccgaccta 3240

cgggagcgcc ctgggcatcg cgtgggcctc caactgggag tactccgcct tcgtgcccac 3300

caacccctgg cgctcctcca tgtccctcgt gcgcaagttc tccctcaaca ccgagtacca 3360

ggccaacccg gagacggagc tgatcaacct gaaggccgag ccgatcctga acatcagcaa 3420

cgccggcccc tggagccggt tcgccaccaa caccacgttg acgaaggcca acagctacaa 3480

cgtcgacctg tccaacagca ccggcaccct ggagttcgag ctggtgtacg ccgtcaacac 3540

cacccagacg atctccaagt ccgtgttcgc ggacctctcc ctctggttca agggcctgga 3600

ggaccccgag gagtacctcc gcatgggctt cgaggtgtcc gcgtcctcct tcttcctgga 3660

ccgcgggaac agcaaggtga agttcgtgaa ggagaacccc tacttcacca accgcatgag 3720

cgtgaacaac cagcccttca agagcgagaa cgacctgtcc tactacaagg tgtacggctt 3780

gctggaccag aacatcctgg agctgtactt caacgacggc gacgtcgtgt ccaccaacac 3840

ctacttcatg accaccggga acgccctggg ctccgtgaac atgacgacgg gggtggacaa 3900

cctgttctac atcgacaagt tccaggtgcg cgaggtcaag tgacaattgg cagcagcagc 3960

tcggatagta tcgacacact ctggacgctg gtcgtgtgat ggactgttgc cgccacactt 4020

gctgccttga cctgtgaata tccctgccgc ttttatcaaa cagcctcagt gtgtttgatc 4080

ttgtgtgtac gcgcttttgc gagttgctag ctgcttgtgc tatttgcgaa taccaccccc 4140
agcatcccct tccctcgttt catatcgctt gcatcccaac cgcaacttat ctacgctgtc 4200
ctgctatccc tcagcgctgc tcctgctcct gctcactgcc cctcgcacag ccttggtttg 4260
ggctccgcct gtattctcct ggtactgcaa cctgtaaacc agcactgcaa tgctgatgca 4320
cgggaagtag tgggatggga acacaaatgg aggatcgtag agctggatcc tctaaagctt 4380
ggccgacagg acgcgcgtca aaggtgctgg tcgtgtatgc cctggccggc aggtcgttgc 4440
tgctgctggt tagtgattcc gcaaccctga ttttggcgtc ttattttggc gtggcaaacg 4500

ctggcgcccg cgagccgggc cggcggcgat gcggtgcccc acggctgccg gaatccaagg 4560
gaggcaagag cgcccgggtc agttgaaggg ctttacgcgc aaggtacagc cgctcctgca 4620
aggctgcgtg gtggaattgg acgtgcaggt cctgctgaag ttcctccacc gcctcaccag 4680
cggacaaagc accggtgtat caggtccgtg tcatccactc taaagaactc gactacgacc 4740
tactgatggc cctagattct tcatcaaaaa cgcctgagac acttgcccag gattgaaact 4800
ccctgaaggg accaccaggg gccctgagtt gttccttccc cccgtggcga gctgccagcc 4860
aggctgtacc tgtgatcgag gctggcggga aaataggctt cgtgtgctca ggtcatggga 4920

ggtgcaggac agctcatgaa acgccaacaa tcgcacaatt catgtcaagc taatcagcta 4980
tttcctcttc acgagctgta attgtcccaa aattctggtc taccgggggt gatccttcgt 5040
gtacgggccc ttccctcaac cctaggtatg cgcgcatgcg gtcgccgcgc aactcgcgcg 5100
agggccgagg gtttgggacg ggccgtcccg aaatgcagtt gcacccggat gcgtggcacc 5160
ttttttgcga taatttatgc aatggactgc tctgcaaaat tctggctctg tcgccaaccc 5220
taggatcagc ggcgtaggat ttcgtaatca ttcgtcctga tggggagcta ccgactaccc 5280
taatatcagc ccgactgcct gacgccagcg tccacttttg tgcacacatt ccattcgtgc 5340

ccaagacatt tcattgtggt gcgaagcgtc cccagttacg ctcacctgtt tcccgacctc 5400
cttactgttc tgtcgacaga gcgggcccac aggccggtcg cagccactag tatggtggcc 5460
accgccgcct cctccgcctt cttccccctg ccctccgccg acacctcctc ccgccccggc 5520
aagctgggca acaagccctc ctccctgtcc cccctgaagc ccaagtccac ccccaacggc 5580
ggcctgcagg tgaaggccaa cgcctccgcc ccccccaaga tcaacggctc ccccgtgggc 5640
ctgaagtccg gcggcctgaa ggggcgcgcc acccaggagg acgcccactc cgcccccccc 5700
ccccgcacct tcatcaacca gctgcccgac tggtccatgc tgctggccgc catcaccacc 5760

gtgttcctgg ccgccgagaa gcagtggatg atgctggact ggaagcccaa gcgccccgac 5820

atgctggtgg acccccttcgg cctgggctcc atcgtgcagg acggcctggt gttccgccag 5880

aacttctcca tccgctccta cgagatcggc gccgaccgca ccgcctccat cgagaccgtg 5940

atgaaccacc tgcaggagac cgccctgaac cacgtgaaga tcgccggcct gtccaacgac 6000

ggcttcggcc gcacccccga gatgtacaag cgcgacctga tctgggtggt ggccaagatg 6060

caggtgatgg tgaaccgcta ccccacctgg ggcgacaccg tggaggtgaa cacctgggtg 6120

gccaagtccg gcaagaacgg catgcgccgc gactggctga tctccgactg caacaccggc 6180

gagatcctga cccgcgcctc ctccgtgtgg gtgatgatga accagaagac ccgccgcctg 6240

tccaagatcc ccgacgaggt gcgcaacgag atcgagcccc acttcgtgga ctccccccc 6300

gtgatcgagg acgacgaccg caagctgccc aagctggacg agaagaccgc cgactccatc 6360

cgcaagggcc tgacccccg ctggaacgac ctggacgtga accagcacgt gaacaacgtg 6420

aagtacatcg gctggatcct ggagtccacc ccccccgagg tgctggagac ccaggagctg 6480

tgctccctga ccctggagta ccgccgcgag tgcggccgcg agtccgtgct ggagtccctg 6540

accgccatgg acccctccgg cggcggctac ggctcccagt tccagcacct gctgcgcctg 6600

gaggacggcg gcgagatcgt gaagggccgc accgagtggc gccccaagaa cggcgtgatc 6660

aacggcgtgg tgcccaccgg cgagtcctcc cccggcgact actccatgga ctacaaggac 6720

cacgacggcg actacaagga ccacgacatc gactacaagg acgacgacga caagtgaatc 6780

gatagatctc ttaaggcagc agcagctcgg atagtatcga cacactctgg acgctggtcg 6840

tgtgatggac tgttgccgcc acacttgctg ccttgacctg tgaatatccc tgccgctttt 6900

atcaaacagc ctcagtgtgt ttgatcttgt gtgtacgcgc ttttgcgagt tgctagctgc 6960

ttgtgctatt tgcgaatacc acccccagca tcccccttccc tcgtttcata tcgcttgcat 7020

cccaaccgca acttatctac gctgtcctgc tatccctcag cgctgctcct gctcctgctc 7080

actgcccctc gcacagcctt ggtttgggct ccgcctgtat tctcctggta ctgcaacctg 7140

taaaccagca ctgcaatgct gatgcacggg aagtagtggg atgggaacac aaatggaaag 7200

ctgtagaatt cggccgacag gacgcgcgtc aaaggtgctg gtcgtgtatg ccctggccgg 7260

caggtcgttg ctgctgctgg ttagtgattc cgcaaccctg attttggcgt cttattttgg 7320

cgtggcaaac gctggcgccc gcgagccggg ccggcggcga tgcggtgccc cacggctgcc 7380

ggaatccaag ggaggcaaga gcgcccgggt cagttgaagg gctttacgcg caaggtacag 7440

ccgctcctgc aaggctgcgt ggtggaattg gacgtgcagg tcctgctgaa gttcctccac 7500

cgcctcacca gcggacaaag caccggtgta tcaggtccgt gtcatccact ctaaagaact 7560

cgactacgac ctactgatgg ccctagattc ttcatcaaaa acgcctgaga cacttgccca 7620

ggattgaaac tccctgaagg gaccaccagg ggccctgagt tgttccttcc ccccgtggcg 7680

agctgccagc caggctgtac ctgtgatcga ggctggcggg aaaataggct tcgtgtgctc 7740

aggtcatggg aggtgcagga cagctcatga aacgccaaca atcgcacaat tcatgtcaag 7800

ctaatcagct atttcctctt cacgagctgt aattgtccca aaattctggt ctaccggggg 7860

tgatccttcg tgtacgggcc cttccctcaa ccctaggtat gcgcgcatgc ggtcgccgcg 7920

caactcgcgc gagggccgag ggtttgggac gggccgtccc gaaatgcagt tgcacccgga 7980

tgcgtggcac ctttttgcg ataatttatg caatggactg ctctgcaaaa ttctggctct 8040

gtcgccaacc ctaggatcag cggcgtagga tttcgtaatc attcgtcctg atggggagct 8100

accgactacc ctaatatcag cccgactgcc tgacgccagc gtccactttt gtgcacacat 8160

tccattcgtg cccaagacat ttcattgtgg tgcgaagcgt ccccagttac gctcacctgt 8220

ttcccgacct ccttactgtt ctgtcgacag agcgggccca caggccggtc gcagccacta 8280

gtatggtggc caccgccgcc tcctccgcct tcttcccccc gccctccgcc gacacctcct 8340

cccgccccgg caagctgggc aacaagccct cctccctgtc ccccctgaag cccaagtcca 8400

cccccaacgg cggcctgcag gtgaaggcca acgcctccgc ccccccccaag atcaacggct 8460

cccccgtggg cctgaagtcc ggcggcctga aggggcgcgc cacccaggag gacgcccact 8520

ccgccccccc cccccgcacc ttcatcaacc agctgcccga ctggtccatg ctgctggccg 8580

ccatcaccac cgtgttcctg gccgccgaga agcagtggat gatgctggac tggaagccca 8640

agcgccccga catgctggtg gacccccttcg gcctgggctc catcgtgcag gacggcctgg 8700

tgttccgcca gaacttctcc atccgctcct acgagatcgg cgccgaccgc accgcctcca 8760

tcgagaccgt gatgaaccac ctgcaggaga ccgccctgaa ccacgtgaag atcgccggcc 8820

tgtccaacga cggcttcggc cgcacccccg agatgtacaa gcgcgacctg atctgggtgg 8880

tggccaagat gcaggtgatg gtgaaccgct accccacctg gggcgacacc gtggaggtga 8940

acacctgggt ggccaagtcc ggcaagaacg gcatgcgccg cgactggctg atctccgact 9000

gcaacaccgg cgagatcctg acccgcgcct cctccgtgtg ggtgatgatg aaccagaaga 9060

cccgccgcct gtccaagatc cccgacgagg tgcgcaacga gatcgagccc cacttcgtgg 9120

actccccccc cgtgatcgag gacgacgacc gcaagctgcc caagctggac gagaagaccg 9180

ccgactccat ccgcaagggc ctgacccccc gctggaacga cctggacgtg aaccagcacg 9240

tgaacaacgt gaagtacatc ggctggatcc tggagtccac ccccccccgag gtgctggaga 9300

cccaggagct gtgctccctg accctggagt accgccgcga gtgcggccgc gagtccgtgc 9360

tggagtccct gaccgccatg gacccctccg gcggcggcta cggctcccag ttccagcacc 9420

tgctgcgcct ggaggacggc ggcgagatcg tgaagggccg caccgagtgg cgccccaaga 9480

acggcgtgat caacggcgtg gtgcccaccg gcgagtcctc ccccggcgac tactccatgg 9540

actacaagga ccacgacggc gactacaagg accacgacat cgactacaag gacgacgacg 9600
acaagtgaat cgatagatct cttaaggcag cagcagctcg gatagtatcg acacactctg 9660
gacgctggtc gtgtgatgga ctgttgccgc cacacttgct gccttgacct gtgaatatcc 9720
ctgccgcttt tatcaaacag cctcagtgtg tttgatcttg tgtgtacgcg cttttgcgag 9780
ttgctagctg cttgtgctat ttgcgaatac cacccccagc atccccttcc ctcgtttcat 9840
atcgcttgca tcccaaccgc aacttatcta cgctgtcctg ctatccctca gcgctgctcc 9900
tgctcctgct cactgcccct cgcacagcct tggtttgggc tccgcctgta ttctcctggt 9960

actgcaacct gtaaaccagc actgcaatgc tgatgcacgg gaagtagtgg gatgggaaca 10020
caaatggaaa gctgtagagc tcggcggcgt gcgcaagatg aacccctttt gcatcccctt 10080
ctccatctcc aacatgggcg gcgcgatgct ggcgatggac atcggcttca tgggccccaa 10140
ctactccatc tccacggcct gcgcgacggg caactactgc atcctgggcg cggcggacca 10200
catccggcgc ggcgacgcaa acgtgatgct ggccggcggc gcggacgcgg ccatcatccc 10260
ctcgggcatc ggcggcttca tcgcgtgcaa ggcgctgagc aagcgcaacg acgagcccga 10320
gcgcgcgagc cggccctggg acgccgaccg cgacggcttc gtcatgggcg agggcgccgg 10380

cgtgctggtg ctggaggagc tggagcacgc caagcgccgc ggcgcgacca ttttggctga 10440
attagttggc ggcgcggcca cctcggacgc gcaccacatg accgagcccg acccgcaggg 10500
ccgcggcgtg cgcctctgcc tcgagcgcgc gctcgagcgc gcgcgcctcg cgcccgagcg 10560
cgtcggctac gtcaacgcgc acggcaccag cacgcccgcg ggcgacgtgg ccgagtaccg 10620
cgccatccgc gccgtcatcc cgcaggactc actacgcatc aactccacaa agtccatgat 10680
cgggcacctg ctcggcggcg ccggcgcggt cgaggccgtg gccgccatcc aggccctgcg 10740
caccggctgg ctccacccca acttgaacct cgagaacccc gcgcctggcg tcgaccccgt 10800

cgtgctcgtg gggaagagcc tctagatctt aaaatgaagt gaagttccta tactttctag 10860
agaataggaa cttctatagt gagtcgaata agggcgacac aaaatttatt ctaaatgcat 10920
aataaatact gataacatct tatagtttgt attatatttt gtattatcgt tgacatgtat 10980
aattttgata tcaaaaactg attttccctt tattattttc gagatttatt ttcttaattc 11040
tctttaacaa actagaaata ttgtatatac aaaaaatcat aaataataga tgaatagttt 11100
aattataggt gttcatcaat cgaaaaagca acgtatctta tttaaagtgc gttgcttttt 11160
tctcatttat aaggttaaat aattctcata tatcaagcaa agtgacaggc gcccttaaat 11220

attctgacaa atgctctttc cctaaactcc ccccataaaa aaacccgccg aagcgggttt 11280

ttacgttatt tgcggattaa cgattactcg ttatcagaac cgcccagggg gcccgagctt 11340

aagactggcc gtcgttttac aacacagaaa gagtttgtag aaacgcaaaa aggccatccg 11400

tcaggggcct tctgcttagt ttgatgcctg gcagttccct actctcgcct tccgcttcct 11460

cgctcactga ctcgctgcgc tcggtcgttc ggctgcggcg agcggtatca gctcactcaa 11520

aggcggtaat acggttatcc acagaatcag gggataacgc aggaaagaac atgtgagcaa 11580

aaggccagca aaaggccagg aaccgtaaaa aggccgcgtt gctggcgttt ttccataggc 11640

tccgcccccc tgacgagcat cacaaaaatc gacgctcaag tcagaggtgg cgaaacccga 11700

caggactata aagataccag gcgtttcccc ctggaagctc cctcgtgcgc tctcctgttc 11760

cgaccctgcc gcttaccgga tacctgtccg cctttctccc ttcgggaagc gtggcgcttt 11820

ctcatagctc acgctgtagg tatctcagtt cggtgtaggt cgttcgctcc aagctgggct 11880

gtgtgcacga acccccgtt cagcccgacc gctgcgcctt atccggtaac tatcgtcttg 11940

agtccaaccc ggtaagacac gacttatcgc cactggcagc agccactggt aacaggatta 12000

gcagagcgag gtatgtaggc ggtgctacag agttcttgaa gtggtgggct aactacggct 12060

acactagaag aacagtattt ggtatctgcg ctctgctgaa gccagttacc ttcggaaaaa 12120

gagttggtag ctcttgatcc ggcaaacaaa ccaccgctgg tagcggtggt ttttttgttt 12180

gcaagcagca gattacgcgc agaaaaaaag gatctcaaga agatcctttg atcttttcta 12240

cggggtctga cgctcagtgg aacgacgcgc gcgtaactca cgttaaggga ttttggtcat 12300

gagcttgcgc cgtcccgtca agtcagcgta atgctctgct tt 12342

<210> 250

<211> 9362

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 250

cactggcagc agccactggt aacaggatta gcagagcgag gtatgtaggc ggtgctacag 60

agttcttgaa gtggtggcct aactacggct acactagaag gacagtattt ggtatctgcg 120

ctctgctgaa gccagttacc ttcggaaaaa gagttggtag ctcttgatcc ggcaaacaaa 180

ccaccgctgg tagcggtggt ttttttgttt gcaagcagca gattacgcgc agaaaaaaag 240

gatctcaaga agatcctttg atcttttcta cggggtctga cgctcagtgg aacgaaaact 300

cacgttaagg gattttggtc atgagattat caaaaaggat cttcacctag atccttttaa 360

attaaaaatg aagttttaaa tcaatctaaa gtatatatga gtaaacttgg tctgacagtt 420

accaatgctt aatcagtgag gcacctatct cagcgatctg tctatttcgt tcatccatag 480

ttgcctgact ccccgtcgtg tagataacta cgatacggga gggcttacca tctggcccca 540
gtgctgcaat gataccgcga gacccacgct caccggctcc agatttatca gcaataaacc 600
agccagccgg aagggccgag cgcagaagtg gtcctgcaac tttatccgcc tccatccagt 660
ctattaattg ttgccgggaa gctagagtaa gtagttcgcc agttaatagt ttgcgcaacg 720
ttgttgccat tgctacaggc atcgtggtgt cacgctcgtc gtttggtatg gcttcattca 780
gctccggttc ccaacgatca aggcgagtta catgatcccc catgttgtgc aaaaaagcgg 840
ttagctcctt cggtcctccg atcgttgtca gaagtaagtt ggccgcagtg ttatcactca 900

tggttatggc agcactgcat aattctctta ctgtcatgcc atccgtaaga tgcttttctg 960
tgactggtga gtactcaacc aagtcattct gagaatagtg tatgcggcga ccgagttgct 1020
cttgcccggc gtcaatacgg gataataccg cgccacatag cagaacttta aaagtgctca 1080
tcattggaaa acgttcttcg gggcgaaaac tctcaaggat cttaccgctg ttgagatcca 1140
gttcgatgta acccactcgt gcacccaact gatcttcagc atcttttact ttcaccagcg 1200
tttctgggtg agcaaaaaca ggaaggcaaa atgccgcaaa aaagggaata agggcgacac 1260
ggaaatgttg aatactcata ctcttccttt ttcaatatta ttgaagcatt tatcagggtt 1320

attgtctcat gagcggatac atatttgaat gtatttagaa aaataaacaa ataggggttc 1380
cgcgcacatt tccccgaaaa gtgccacctg acgtctaaga aaccattatt atcatgacat 1440
taacctataa aaataggcgt atcacgaggc cctttcgtct cgcgcgtttc ggtgatgacg 1500
gtgaaaacct ctgacacatg cagctcccgg agacggtcac agcttgtctg taagcggatg 1560
ccgggagcag acaagcccgt cagggcgcgt cagcgggtgt tggcgggtgt cggggctggc 1620
ttaactatgc ggcatcagag cagattgtac tgagagtgca ccatatgcgg tgtgaaatac 1680
cgcacagatg cgtaaggaga aaataccgca tcaggcgcca ttcgccattc aggctgcgca 1740

actgttggga agggcgatcg gtgcgggcct cttcgctatt acgccagctg gcgaaagggg 1800
gatgtgctgc aaggcgatta agttgggtaa cgccagggtt ttcccagtca cgacgttgta 1860
aaacgacggc cagtgaattg atgcatgctc ttcgggctgg tctgaatcct tcaggcgggt 1920
gttacccgag aaagaaaggg tgccgatttc aaagcagacc catgtgccgg gccctgtggc 1980
ctgtgttggc gcctatgtag tcaccccccc tcacccaatt gtcgccagtt tgcgcactcc 2040
ataaactcaa aacagcagct tctgagctgc gctgttcaag aacacctctg gggtttgctc 2100
acccgcgagg tcgacgccca gcatggctat caagacgaac aggcagcctg tggagaagcc 2160

tccgttcacg atcgggacgc tgcgcaaggc catccccgcg cactgtttcg agcgctcggc 2220

gcttcgtagc agcatgtacc tggcctttga catcgcggtc atgtccctgc tctacgtcgc 2280

gtcgacgtac atcgaccctg caccggtgcc tacgtgggtc aagtacggca tcatgtggcc 2340

gctctactgg ttcttccagg tgtgtttgag ggttttggtt gcccgtattg aggtcctggt 2400

ggcgcgcatg gaggagaagg cgcctgtccc gctgaccccc ccggctaccc tcccggcacc 2460

ttccagggcg ccttcggcac gggtgtctgg gtgtgcgcgc acgagtgcgg ccaccaggcc 2520

ttttcctcca gccaggccat caacgacggc gtgggcctgg tgttccacag cctgctgctg 2580

gtgccctact actcctggaa gcactcgcac cgggtaccct ttcttgcgct atgacacttc 2640

cagcaaaagg tagggcgggc tgcgagacgg cttcccggcg ctgcatgcaa caccgatgat 2700

gcttcgaccc cccgaagctc cttcggggct gcatgggcgc tccgatgccg ctccagggcg 2760

agcgctgttt aaatagccag gcccccgatt gcaaagacat tatagcgagc taccaaagcc 2820

atattcaaac acctagatca ctaccacttc tacacaggcc actcgagctt gtgatcgcac 2880

tccgctaagg gggcgcctct tcctcttcgt ttcagtcaca acccgcaaac tctagaatat 2940

caatgctgct gcaggccttc ctgttcctgc tggccggctt cgccgccaag atcagcgcct 3000

ccatgacgaa cgagacgtcc gaccgccccc tggtgcactt caccccccaac aagggctgga 3060

tgaacgaccc caacggcctg tggtacgacg agaaggacgc caagtggcac ctgtacttcc 3120

agtacaaccc gaacgacacc gtctgggggga cgcccttgtt ctggggccac gccacgtccg 3180

acgacctgac caactgggag gaccagccca tcgccatcgc cccgaagcgc aacgactccg 3240

gcgccttctc cggctccatg gtggtggact acaacaacac ctccggcttc ttcaacgaca 3300

ccatcgaccc gcgccagcgc tgcgtggcca tctggaccta caacaccccg gagtccgagg 3360

agcagtacat ctcctacagc ctggacggcg gctacacctt caccgagtac cagaagaacc 3420

ccgtgctggc cgccaactcc acccagttcc gcgacccgaa ggtcttctgg tacgagccct 3480

cccagaagtg gatcatgacc gcggccaagt cccaggacta caagatcgag atctactcct 3540

ccgacgacct gaagtcctgg aagctggagt ccgcgttcgc caacgagggc ttcctcggct 3600

accagtacga gtgccccggc ctgatcgagg tccccaccga gcaggacccc agcaagtcct 3660

actgggtgat gttcatctcc atcaaccccg gcgccccggc cggcggctcc ttcaaccagt 3720

acttcgtcgg cagcttcaac ggcacccact tcgaggcctt cgacaaccag tcccgcgtgg 3780

tggacttcgg caaggactac tacgccctgc agaccttctt caacaccgac ccgacctacg 3840

ggagcgccct gggcatcgcg tgggcctcca actgggagta ctccgccttc gtgcccacca 3900

acccctggcg ctcctccatg tccctcgtgc gcaagttctc cctcaacacc gagtaccagg 3960

ccaacccgga gacggagctg atcaacctga aggccgagcc gatcctgaac atcagcaacg 4020

ccggcccctg gagccggttc gccaccaaca ccacgttgac gaaggccaac agctacaacg 4080

tcgacctgtc caacagcacc ggcaccctgg agttcgagct ggtgtacgcc gtcaacacca 4140

cccagacgat ctccaagtcc gtgttcgcgg acctctccct ctggttcaag ggcctggagg 4200

accccgagga gtacctccgc atgggcttcg aggtgtccgc gtcctccttc ttcctggacc 4260

gcgggaacag caaggtgaag ttcgtgaagg agaaccccta cttcaccaac cgcatgagcg 4320

tgaacaacca gcccttcaag agcgagaacg acctgtccta ctacaaggtg tacggcttgc 4380

tggaccagaa catcctggag ctgtacttca acgacggcga cgtcgtgtcc accaacacct 4440

acttcatgac caccgggaac gccctgggct ccgtgaacat gacgacgggg gtggacaacc 4500

tgttctacat cgacaagttc caggtgcgcg aggtcaagtg acaattggca gcagcagctc 4560

ggatagtatc gacacactct ggacgctggt cgtgtgatgg actgttgccg ccacacttgc 4620

tgccttgacc tgtgaatatc cctgccgctt ttatcaaaca gcctcagtgt gtttgatctt 4680

gtgtgtacgc gcttttgcga gttgctagct gcttgtgcta tttgcgaata ccacccccag 4740

catccccttc cctcgtttca tatcgcttgc atcccaaccg caacttatct acgctgtcct 4800

gctatccctc agcgctgctc ctgctcctgc tcactgcccc tcgcacagcc ttggtttggg 4860

ctccgcctgt attctcctgg tactgcaacc tgtaaaccag cactgcaatg ctgatgcacg 4920

ggaagtagtg ggatgggaac acaaatggag gatcccgcgt ctcgaacaga gcgcgcagag 4980

gaacgctgaa ggtctcgcct ctgtcgcacc tcagcgcggc atacaccaca ataaccacct 5040

gacgaatgcg cttggttctt cgtccattag cgaagcgtcc ggttcacaca cgtgccacgt 5100

tggcgaggtg gcaggtgaca atgatcggtg gagctgatgg tcgaaacgtt cacagcctag 5160

ggatatcgaa ttcggccgac aggacgcgcg tcaaaggtgc tggtcgtgta tgccctggcc 5220

ggcaggtcgt tgctgctgct ggttagtgat tccgcaaccc tgattttggc gtcttatttt 5280

ggcgtggcaa acgctggcgc ccgcgagccg ggccggcggc gatgcggtgc cccacggctg 5340

ccggaatcca agggaggcaa gagcgcccgg gtcagttgaa gggctttacg cgcaaggtac 5400

agccgctcct gcaaggctgc gtggtggaat tggacgtgca ggtcctgctg aagttcctcc 5460

accgcctcac cagcggacaa agcaccggtg tatcaggtcc gtgtcatcca ctctaaagaa 5520

ctcgactacg acctactgat ggccctagat tcttcatcaa aaacgcctga gacacttgcc 5580

caggattgaa actccctgaa gggaccacca ggggccctga gttgttcctt ccccccgtgg 5640

cgagctgcca gccaggctgt acctgtgatc gaggctggcg ggaaaatagg cttcgtgtgc 5700

tcaggtcatg ggaggtgcag gacagctcat gaaacgccaa caatcgcaca attcatgtca 5760

agctaatcag ctatttcctc ttcacgagct gtaattgtcc caaaattctg gtctaccggg 5820

ggtgatcctt cgtgtacggg cccttccctc aaccctaggt atgcgcgcat gcggtcgccg 5880

cgcaactcgc gcgagggccg agggtttggg acgggccgtc ccgaaatgca gttgcacccg 5940

gatgcgtggc acctttttg cgataattta tgcaatggac tgctctgcaa aattctggct 6000
ctgtcgccaa ccctaggatc agcggcgtag gatttcgtaa tcattcgtcc tgatggggag 6060
ctaccgacta ccctaatatc agcccgactg cctgacgcca gcgtccactt ttgtgcacac 6120
attccattcg tgcccaagac atttcattgt ggtgcgaagc gtccccagtt acgctcacct 6180
gtttcccgac ctccttactg ttctgtcgac agagcgggcc cacaggccgg tcgcagccac 6240
tagtatggcc accgcatcca ctttctcggc gttcaatgcc cgctgcggcg acctgcgtcg 6300
ctcggcgggc tccgggcccc ggcgcccagc gaggcccctc cccgtgcgcg ggcgcgccgc 6360

caccggcgag cagccctccg gcgtggcctc cctgcgcgag gccgacaagg agaagtccct 6420
gggcaaccgc ctgcgcctgg gctccctgac cgaggacggc ctgtcctaca aggagaagtt 6480
cgtgatccgc tgctacgagg tgggcatcaa caagaccgcc accatcgaga ccatcgccaa 6540
cctgctgcag gaggtgggcg gcaaccacgc ccagggcgtg ggcttctcca ccgacggctt 6600
cgccaccacc accaccatgc gcaagctgca cctgatctgg gtgaccgccc gcatgcacat 6660
cgagatctac cgctaccccg cctggtccga cgtgatcgag atcgagacct gggtgcaggg 6720
cgagggcaag gtgggcaccc gccgcgactg gatcctgaag gactacgcca acggcgaggt 6780

gatcggccgc gccacctcca agtgggtgat gatgaacgag gacacccgcc gcctgcagaa 6840
ggtgtccgac gacgtgcgcg aggagtacct ggtgttctgc ccccgcaccc tgcgcctggc 6900
cttccccgag gagaacaaca actccatgaa gaagatcccc aagctggagg accccgccga 6960
gtactcccgc ctgggcctgg tgccccgccg ctccgacctg gacatgaaca agcacgtgaa 7020
caacgtgacc tacatcggct gggccctgga gtccatcccc cccgagatca tcgacaccca 7080
cgagctgcag gccatcaccc tggactaccg ccgcgagtgc cagcgcgacg acatcgtgga 7140
ctccctgacc tcccgcgagc ccctgggcaa cgccgccggc gtgaagttca aggagatcaa 7200

cggctccgtg tcccccaaga aggacgagca ggacctgtcc cgcttcatgc acctgctgcg 7260
ctccgccggc tccggcctgg agatcaaccg ctgccgcacc gagtggcgca agaagcccgc 7320
caagcgcatg gactacaagg accacgacgg cgactacaag gaccacgaca tcgactacaa 7380
ggacgacgac gacaagtgaa tcgatagatc tcttaaggca gcagcagctc ggatagtatc 7440
gacacactct ggacgctggt cgtgtgatgg actgttgccg ccacacttgc tgccttgacc 7500
tgtgaatatc cctgccgctt ttatcaaaca gcctcagtgt gtttgatctt gtgtgtacgc 7560
gcttttgcga gttgctagct gcttgtgcta tttgcgaata ccacccccag catcccccttc 7620

cctcgtttca tatcgcttgc atcccaaccg caacttatct acgctgtcct gctatccctc 7680

agcgctgctc ctgctcctgc tcactgcccc tcgcacagcc ttggtttggg ctccgcctgt 7740

attctcctgg tactgcaacc tgtaaaccag cactgcaatg ctgatgcacg ggaagtagtg 7800

ggatgggaac acaaatggaa agcttaatta agagctcccg ccaccactcc aacacggggt 7860

gcctggacaa ggacgaggtg tttgtgccgc cgcaccgcgc agtggcgcac gagggcctgg 7920

agtgggagga gtggctgccc atccgcatgg gcaaggtgct ggtcaccctg accctgggct 7980

ggccgctgta cctcatgttc aacgtcgcct cgcggccgta cccgcgcttc gccaaccact 8040

ttgacccgtg gtcgcccatc ttcagcaagc gcgagcgcat cgaggtggtc atctccgacc 8100

tggcgctggt ggcggtgctc agcgggctca gcgtgctggg ccgcaccatg ggctgggcct 8160

ggctggtcaa gacctacgtg gtgccctacc tgatcgtgaa catgtggctc gtgctcatca 8220

cgctgctcca gcacacgcac ccggcgctgc cgcactactt cgagaaggac tgggactggc 8280

tgcgcggcgc catggccacc gtggaccgct ccatgggccc gcccttcatg gacaacatcc 8340

tgcaccacat ctccgacacc cacgtgctgc accacctctt cagcaccatc ccgcactacc 8400

acgccgagga ggcctccgcc gccatcaggc ccatcctggg caagtactac cagtccgaca 8460

gccgctgggt cggccgcgcc ctgtgggagg actggcgcga ctgccgctac gtcgtcccgg 8520

acgcgcccga ggacgactcc gcgctctggt tccacaagtg agtgagtgag aagagcctct 8580

agagtcgacc tgcaggcatg caagcttggc gtaatcatgg tcatagctgt ttcctgtgtg 8640

aaattgttat ccgctcacaa ttccacacaa catacgagcc ggaagcataa agtgtaaagc 8700

ctggggtgcc taatgagtga gctaactcac attaattgcg ttgcgctcac tgcccgcttt 8760

ccagtcggga aacctgtcgt gccagctgca ttaatgaatc ggccaacgcg cggggagagg 8820

cggtttgcgt attgggcgct cttccgcttc ctcgctcact gactcgctgc gctcggtcgt 8880

tcggctgcgg cgagcggtat cagctcactc aaaggcggta atacggttat ccacagaatc 8940

aggggataac gcaggaaaga acatgtgagc aaaaggccag caaaaggcca ggaaccgtaa 9000

aaaggccgcg ttgctggcgt ttttccatag gctccgcccc cctgacgagc atcacaaaaa 9060

tcgacgctca agtcagaggt ggcgaaaccc gacaggacta taaagatacc aggcgtttcc 9120

ccctggaagc tccctcgtgc gctctcctgt tccgaccctg ccgcttaccg gatacctgtc 9180

cgcctttctc ccttcgggaa gcgtggcgct ttctcatagc tcacgctgta ggtatctcag 9240

ttcggtgtag gtcgttcgct ccaagctggg ctgtgtgcac gaacccccgg ttcagcccga 9300

ccgctgcgcc ttatccggta actatcgtct tgagtccaac ccggtaagac acgacttatc 9360

gc 9362

<210> 251

<211> 612

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 251

gctcttcgga gtcactgtgc cactgagttc gactggtagc tgaatggagt cgctgctcca 60

ctaaacgaat tgtcagcacc gccagccggc cgaggacccg agtcatagcg agggtagtag 120

cgcgccatgg caccgaccag cctgcttgcc agtactggcg tctcttccgc ttctctgtgg 180

tcctctgcgc gctccagcgc gtgcgctttt ccggtggatc atgcggtccg tggcgcaccg 240

cagcggccgc tgcccatgca gcgccgctgc ttccgaacag tggcggtcag ggccgcaccc 300

gcggtagccg tccgtccgga acccgcccaa gagttttggg agcagcttga gccctgcaag 360

atggcggagg acaagcgcat cttcctggag gagcaccggt gcgtggaggt ccggggctga 420

ccggccgtcg cattcaacgt aatcaatcgc atgatgatca gaggacacga agtcttggtg 480

gcggtggcca gaaacactgt ccattgcaag ggcataggga tgcgttcctt cacctctcat 540

ttctcatttc tgaatccctc cctgctcact ctttctcctc ctccttcccg ttcacgcagc 600

attcggggta cc 612

<210> 252

<211> 528

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 252

gacagggtgg ttggctggat ggggaaacgc tggtcgcggg attcgatcct gctgcttata 60

tcctccctgg aagcacaccc acgactctga agaagaaaac gtgcacacac acaacccaac 120

cggccgaata tttgcttcct tatcccgggt ccaagagaga ctgcgatgcc cccctcaatc 180

agcatcctcc tccctgccgc ttcaatcttc cctgcttgcc tgcgcccgcg gtgcgccgtc 240

tgcccgccca gtcagtcact cctgcacagg ccccttgtgc gcagtgctcc tgtacccttt 300

accgctcctt ccattctgcg aggcccccta ttgaatgtat tcgttgcctg tgtggccaag 360

cgggctgctg ggcgcgccgc cgtcgggcag tgctcggcga ctttggcgga agccgattgt 420

tcttctgtaa gccacgcgct tgctgctttg ggaagagaag gggggggtta ctgaatggat 480

gaggaggaga aggaggggta ttggtattat ctgagttggg tgaagagc 528

<210> 253

<211> 3499

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 253

gctcttcgga gtcactgtgc cactgagttc gactggtagc tgaatggagt cgctgctcca 60

ctaaacgaat tgtcagcacc gccagccggc cgaggacccg agtcatagcg agggtagtag 120

cgcgccatgg caccgaccag cctgcttgcc agtactggcg tctcttccgc ttctctgtgg 180

tcctctgcgc gctccagcgc gtgcgctttt ccggtggatc atgcggtccg tggcgcaccg 240

cagcggccgc tgcccatgca gcgccgctgc ttccgaacag tggcggtcag ggccgcaccc 300

gcggtagccg tccgtccgga acccgcccaa gagttttggg agcagcttga gccctgcaag 360

atggcggagg acaagcgcat cttcctggag gagcaccggt gcgtggaggt ccggggctga 420

ccggccgtcg cattcaacgt aatcaatcgc atgatgatca gaggacacga agtcttggtg 480

gcggtggcca gaaacactgt ccattgcaag ggcataggga tgcgttcctt cacctctcat 540

ttctcatttc tgaatccctc cctgctcact ctttctcctc ctccttcccg ttcacgcagc 600

attcggggta ccctttcttg cgctatgaca cttccagcaa aaggtagggc gggctgcgag 660

acggcttccc ggcgctgcat gcaacaccga tgatgcttcg acccccccgaa gctccttcgg 720

ggctgcatgg gcgctccgat gccgctccag ggcgagcgct gtttaaatag ccaggccccc 780

gattgcaaag acattatagc gagctaccaa agccatattc aaacacctag atcactacca 840

cttctacaca ggccactcga gcttgtgatc gcactccgct aaggggggcgc ctcttcctct 900

tcgtttcagt cacaacccgc aaacggcgcg ccatgctgct gcaggccttc ctgttcctgc 960

tggccggctt cgccgccaag atcagcgcct ccatgacgaa cgagacgtcc gaccgccccc 1020

tggtgcactt cacccccaac aagggctgga tgaacgaccc caacggcctg tggtacgacg 1080

agaaggacgc caagtggcac ctgtacttcc agtacaaccc gaacgacacc gtctgggggga 1140

cgcccttgtt ctggggccac gccacgtccg acgacctgac caactgggag gaccagccca 1200

tcgccatcgc cccgaagcgc aacgactccg gcgccttctc cggctccatg gtggtggact 1260

acaacaacac ctccggcttc ttcaacgaca ccatcgaccc gcgccagcgc tgcgtggcca 1320

tctggaccta caacaccccg gagtccgagg agcagtacat ctcctacagc ctggacggcg 1380

gctacacctt caccgagtac cagaagaacc ccgtgctggc cgccaactcc acccagttcc 1440

gcgacccgaa ggtcttctgg tacgagccct cccagaagtg gatcatgacc gcggccaagt 1500

cccaggacta caagatcgag atctactcct ccgacgacct gaagtcctgg aagctggagt 1560

ccgcgttcgc caacgagggc ttcctcggct accagtacga gtgccccggc ctgatcgagg 1620

tccccaccga gcaggacccc agcaagtcct actgggtgat gttcatctcc atcaaccccg 1680

gcgccccggc cggcggctcc ttcaaccagt acttcgtcgg cagcttcaac ggcacccact 1740

tcgaggcctt cgacaaccag tcccgcgtgg tggacttcgg caaggactac tacgccctgc 1800

agaccttctt caacaccgac ccgacctacg ggagcgccct gggcatcgcg tgggcctcca 1860

actgggagta ctccgccttc gtgcccacca acccctggcg ctcctccatg tccctcgtgc 1920

gcaagttctc cctcaacacc gagtaccagg ccaacccgga gacggagctg atcaacctga 1980

aggccgagcc gatcctgaac atcagcaacg ccggcccctg gagccggttc gccaccaaca 2040

ccacgttgac gaaggccaac agctacaacg tcgacctgtc caacagcacc ggcaccctgg 2100

agttcgagct ggtgtacgcc gtcaacacca cccagacgat ctccaagtcc gtgttcgcgg 2160

acctctccct ctggttcaag ggcctggagg accccgagga gtacctccgc atgggcttcg 2220

aggtgtccgc gtcctccttc ttcctggacc gcgggaacag caaggtgaag ttcgtgaagg 2280

agaaccccta cttcaccaac cgcatgagcg tgaacaacca gcccttcaag agcgagaacg 2340

acctgtccta ctacaaggtg tacggcttgc tggaccagaa catcctggag ctgtacttca 2400

acgacggcga cgtcgtgtcc accaacacct acttcatgac caccgggaac gccctgggct 2460

ccgtgaacat gacgacgggg gtggacaacc tgttctacat cgacaagttc caggtgcgcg 2520

aggtcaagtg acaattggca gcagcagctc ggatagtatc gacacactct ggacgctggt 2580

cgtgtgatgg actgttgccg ccacacttgc tgccttgacc tgtgaatatc cctgccgctt 2640

ttatcaaaca gcctcagtgt gtttgatctt gtgtgtacgc gcttttgcga gttgctagct 2700

gcttgtgcta tttgcgaata ccacccccag catccccttc cctcgtttca tatcgcttgc 2760

atcccaaccg caacttatct acgctgtcct gctatccctc agcgctgctc ctgctcctgc 2820

tcactgcccc tcgcacagcc ttggtttggg ctccgcctgt attctcctgg tactgcaacc 2880

tgtaaaccag cactgcaatg ctgatgcacg ggaagtagtg ggatgggaac acaaatggag 2940

gatcgtagag ctcactagta tcgatttcga agacagggtg gttggctgga tggggaaacg 3000

ctggtcgcgg gattcgatcc tgctgcttat atcctccctg gaagcacacc cacgactctg 3060

aagaagaaaa cgtgcacaca cacaacccaa ccggccgaat atttgcttcc ttatcccggg 3120

tccaagagag actgcgatgc ccccctcaat cagcatcctc ctccctgccg cttcaatctt 3180

ccctgcttgc ctgcgcccgc ggtgcgccgt ctgcccgccc agtcagtcac tcctgcacag 3240

gccccttgtg cgcagtgctc ctgtaccctt taccgctcct tccattctgc gaggcccccc 3300

attgaatgta ttcgttgcct gtgtggccaa gcgggctgct gggcgcgccg ccgtcgggca 3360

gtgctcggcg actttggcgg aagccgattg ttcttctgta agccacgcgc ttgctgcttt 3420

gggaagagaa gggggggggt actgaatgga tgaggaggag aaggaggggt attggtatta 3480

tctgagttgg gtgaagagc 3499

<210> 254

<211> 6514

<212> DNA

<213> Artificial Sequence

<220><223> synthetic construct

<400> 254

gctcttcgga gtcactgtgc cactgagttc gactggtagc tgaatggagt cgctgctcca 60

ctaaacgaat tgtcagcacc gccagccggc cgaggacccg agtcatagcg agggtagtag 120

cgcgccatgg caccgaccag cctgcttgcc agtactggcg tctcttccgc ttctctgtgg 180

tcctctgcgc gctccagcgc gtgcgctttt ccggtggatc atgcggtccg tggcgcaccg 240

cagcggccgc tgcccatgca gcgccgctgc ttccgaacag tggcggtcag ggccgcaccc 300

gcggtagccg tccgtccgga acccgcccaa gagttttggg agcagcttga gccctgcaag 360

atggcggagg acaagcgcat cttcctggag gagcaccggt gcgtggaggt ccggggctga 420

ccggccgtcg cattcaacgt aatcaatcgc atgatgatca gaggacacga agtcttggtg 480

gcggtggcca gaaacactgt ccattgcaag ggcataggga tgcgttcctt cacctctcat 540

ttctcatttc tgaatccctc cctgctcact ctttctcctc ctccttcccg ttcacgcagc 600

attcggggta ccctttcttg cgctatgaca cttccagcaa aaggtagggc gggctgcgag 660

acggcttccc ggcgctgcat gcaacaccga tgatgcttcg acccccccgaa gctccttcgg 720

ggctgcatgg gcgctccgat gccgctccag ggcgagcgct gtttaaatag ccaggccccc 780

gattgcaaag acattatagc gagctaccaa agccatattc aaacacctag atcactacca 840

cttctacaca ggccactcga gcttgtgatc gcactccgct aaggggggcgc ctcttcctct 900

tcgtttcagt cacaacccgc aaacggcgcg ccatgctgct gcaggccttc ctgttcctgc 960

tggccggctt cgccgccaag atcagcgcct ccatgacgaa cgagacgtcc gaccgccccc 1020

tggtgcactt cacccccaac aagggctgga tgaacgaccc caacggcctg tggtacgacg 1080

agaaggacgc caagtggcac ctgtacttcc agtacaaccc gaacgacacc gtctgggggga 1140

cgcccttgtt ctggggccac gccacgtccg acgacctgac caactgggag gaccagccca 1200

tcgccatcgc cccgaagcgc aacgactccg gcgccttctc cggctccatg gtggtggact 1260

acaacaacac ctccggcttc ttcaacgaca ccatcgaccc gcgccagcgc tgcgtggcca 1320

tctggaccta caacaccccg gagtccgagg agcagtacat ctcctacagc ctggacggcg 1380

gctacacctt caccgagtac cagaagaacc ccgtgctggc cgccaactcc acccagttcc 1440

gcgacccgaa ggtcttctgg tacgagccct cccagaagtg gatcatgacc gcggccaagt 1500

cccaggacta caagatcgag atctactcct ccgacgacct gaagtcctgg aagctggagt 1560

ccgcgttcgc caacgagggc ttcctcggct accagtacga gtgccccggc ctgatcgagg 1620

tccccaccga gcaggacccc agcaagtcct actgggtgat gttcatctcc atcaaccccg 1680

gcgccccggc cggcggctcc ttcaaccagt acttcgtcgg cagcttcaac ggcacccact 1740

tcgaggcctt cgacaaccag tcccgcgtgg tggacttcgg caaggactac tacgccctgc 1800

agaccttctt caacaccgac ccgacctacg ggagcgccct gggcatcgcg tgggcctcca 1860

actgggagta ctccgccttc gtgcccacca acccctggcg ctcctccatg tccctcgtgc 1920

gcaagttctc cctcaacacc gagtaccagg ccaacccgga gacggagctg atcaacctga 1980

aggccgagcc gatcctgaac atcagcaacg ccggcccctg gagccggttc gccaccaaca 2040

ccacgttgac gaaggccaac agctacaacg tcgacctgtc caacagcacc ggcaccctgg 2100

agttcgagct ggtgtacgcc gtcaacacca cccagacgat ctccaagtcc gtgttcgcgg 2160

acctctccct ctggttcaag ggcctggagg accccgagga gtacctccgc atgggcttcg 2220

aggtgtccgc gtcctccttc ttcctggacc gcgggaacag caaggtgaag ttcgtgaagg 2280

agaaccccta cttcaccaac cgcatgagcg tgaacaacca gcccttcaag agcgagaacg 2340

acctgtccta ctacaaggtg tacggcttgc tggaccagaa catcctggag ctgtacttca 2400

acgacggcga cgtcgtgtcc accaacacct acttcatgac caccgggaac gccctgggct 2460

ccgtgaacat gacgacgggg gtggacaacc tgttctacat cgacaagttc caggtgcgcg 2520

aggtcaagtg acaattggca gcagcagctc ggatagtatc gacacactct ggacgctggt 2580

cgtgtgatgg actgttgccg ccacacttgc tgccttgacc tgtgaatatc cctgccgctt 2640

ttatcaaaca gcctcagtgt gtttgatctt gtgtgtacgc gcttttgcga gttgctagct 2700

gcttgtgcta tttgcgaata ccacccccag catccccttc cctcgtttca tatcgcttgc 2760

atcccaaccg caacttatct acgctgtcct gctatccctc agcgctgctc ctgctcctgc 2820

tcactgcccc tcgcacagcc ttggtttggg ctccgcctgt attctcctgg tactgcaacc 2880

tgtaaaccag cactgcaatg ctgatgcacg ggaagtagtg ggatgggaac acaaatggag 2940

gatcccgcgt ctcgaacaga gcgcgcagag gaacgctgaa ggtctcgcct ctgtcgcacc 3000

tcagcgcggc atacaccaca ataaccacct gacgaatgcg cttggttctt cgtccattag 3060

cgaagcgtcc ggttcacaca cgtgccacgt tggcgaggtg gcaggtgaca atgatcggtg 3120

gagctgatgg tcgaaacgtt cacagcctag ggatatcgaa ttcggccgac aggacgcgcg 3180

tcaaaggtgc tggtcgtgta tgccctggcc ggcaggtcgt tgctgctgct ggttagtgat 3240

tccgcaaccc tgattttggc gtcttatttt ggcgtggcaa acgctggcgc ccgcgagccg 3300

ggccggcggc gatgcggtgc cccacggctg ccggaatcca agggaggcaa gagcgcccgg 3360
gtcagttgaa gggctttacg cgcaaggtac agccgctcct gcaaggctgc gtggtggaat 3420
tggacgtgca ggtcctgctg aagttcctcc accgcctcac cagcggacaa agcaccggtg 3480
tatcaggtcc gtgtcatcca ctctaaagag ctcgactacg acctactgat ggccctagat 3540
tcttcatcaa aaacgcctga gacacttgcc caggattgaa actccctgaa gggaccacca 3600
ggggccctga gttgttcctt ccccccgtgg cgagctgcca gccaggctgt acctgtgatc 3660
gaggctggcg ggaaaatagg cttcgtgtgc tcaggtcatg ggaggtgcag gacagctcat 3720

gaaacgccaa caatcgcaca attcatgtca agctaatcag ctatttcctc ttcacgagct 3780
gtaattgtcc caaaattctg gtctaccggg ggtgatcctt cgtgtacggg cccttccctc 3840
aaccctaggt atgcgcgcat gcggtcgccg cgcaactcgc gcgagggccg agggtttggg 3900
acgggccgtc ccgaaatgca gttgcacccg gatgcgtggc acctttttg cgataattta 3960
tgcaatggac tgctctgcaa aattctggct ctgtcgccaa ccctaggatc agcggcgtag 4020
gatttcgtaa tcattcgtcc tgatggggag ctaccgacta ccctaatatc agcccgactg 4080
cctgacgcca gcgtccactt ttgtgcacac attccattcg tgcccaagac atttcattgt 4140

ggtgcgaagc gtccccagtt acgctcacct gtttcccgac ctccttactg ttctgtcgac 4200
agagcgggcc cacaggccgg tcgcagccac tagtatggtg gtggccgccg ccgccagcag 4260
cgccttcttc cccgtgcccg ccccccgccc cacccccaag cccggcaagt tcggcaactg 4320
gcccagcagc ctgagccagc ccttcaagcc caagagcaac cccaacggcc gcttccaggt 4380
gaaggccaac gtgagccccc acgggcgcgc ccccaaggcc aacggcagcg ccgtgagcct 4440
gaagtccggc agcctgaaca ccctggagga ccccccccagc agccccccc cccgcacctt 4500
cctgaaccag ctgcccgact ggagccgcct gcgcaccgcc atcaccaccg tgttcgtggc 4560

cgccgagaag cagttcaccc gcctggaccg caagagcaag cgccccgaca tgctggtgga 4620
ctggttcggc agcgagacca tcgtgcagga cggcctggtg ttccgcgagc gcttcagcat 4680
ccgcagctac gagatcggcg ccgaccgcac cgccagcatc gagaccctga tgaaccacct 4740
gcaggacacc agcctgaacc actgcaagag cgtgggcctg ctgaacgacg gcttcggccg 4800
cacccccgag atgtgcaccc gcgacctgat ctgggtgctg accaagatgc agatcgtggt 4860
gaaccgctac cccacctggg gcgacaccgt ggagatcaac agctggttca gccagagcgg 4920
caagatcggc atgggccgcg agtggctgat cagcgactgc aacaccggcg agatcctggt 4980

gcgcgccacc agcgcctggg ccatgatgaa ccagaagacc cgccgcttca gcaagctgcc 5040

ctgcgaggtg cgccaggaga tcgcccccca cttcgtggac gccccccccg tgatcgagga 5100

caacgaccgc aagctgcaca agttcgacgt gaagaccggc gacagcatct gcaagggcct 5160

gacccccggc tggaacgact tcgacgtgaa ccagcacgtg agcaacgtga agtacatcgg 5220

ctggattctg gagagcatgc ccaccgaggt gctggagacc caggagctgt gcagcctgac 5280

cctggagtac cgccgcgagt gcggccgcga gagcgtggtg gagagcgtga ccagcatgaa 5340

ccccagcaag gtgggcgacc gcagccagta ccagcacctg ctgcgcctgg aggacggcgc 5400

cgacatcatg aagggccgca ccgagtggcg ccccaagaac gccggcacca accgcgccat 5460

cagcacctga ttaattaact cgaggcagca gcagctcgga tagtatcgac acactctgga 5520

cgctggtcgt gtgatggact gttgccgcca cacttgctgc cttgacctgt gaatatccct 5580

gccgctttta tcaaacagcc tcagtgtgtt tgatcttgtg tgtacgcgct tttgcgagtt 5640

gctagctgct tgtgctattt gcgaatacca cccccagcat ccccttccct cgtttcatat 5700

cgcttgcatc ccaaccgcaa cttatctacg ctgtcctgct atccctcagc gctgctcctg 5760

ctcctgctca ctgcccctcg cacagccttg gtttgggctc cgcctgtatt ctcctggtac 5820

tgcaacctgt aaaccagcac tgcaatgctg atgcacggga agtagtggga tgggaacaca 5880

aatggaaagc ttgagctctt gttttccaga aggagttgct ccttgagcct ttcattctca 5940

gcctcgataa cctccaaagc cgctctaatt gtggaggggg ttcgaagaca gggtggttgg 6000

ctggatgggg aaacgctggt cgcgggattc gatcctgctg cttatatcct ccctggaagc 6060

acacccacga ctctgaagaa gaaaacgtgc acacacacaa cccaaccggc cgaatatttg 6120

cttccttatc ccgggtccaa gagagactgc gatgcccccc tcaatcagca tcctcctccc 6180

tgccgcttca atcttccctg cttgcctgcg cccgcggtgc gccgtctgcc cgcccagtca 6240

gtcactcctg cacaggcccc ttgtgcgcag tgctcctgta ccctttaccg ctccttccat 6300

tctgcgaggc cccctattga atgtattcgt tgcctgtgtg gccaagcggg ctgctgggcg 6360

cgccgccgtc gggcagtgct cggcgacttt ggcggaagcc gattgttctt ctgtaagcca 6420

cgcgcttgct gctttgggaa gagaaggggg ggggtactga atggatgagg aggagaagga 6480

ggggtattgg tattatctga gttgggtgaa gagc 6514

<210> 255

<211> 767

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 255

atggcaccga ccagcctgct tgccagtact ggcgtctctt ccgcttctct gtggtcctct 60

gcgcgctcca gcgcgtgcgc ttttccggtg gatcatgcgg tccgtggcgc accgcagcgg 120

ccgctgccca tgcagcgccg ctgcttccga acagtggcgg tcagggccgc acccgcggta 180

gccgtccgtc cggaacccgc ccaagagttt tgggagcagc ttgagccctg caagatggcg 240

gaggacaagc gcatcttcct ggaggagcac cggtgcgtgg aggtccgggg ctgaccggcc 300

gtcgcattca acgtaatcaa tcgcatgatg atcagaggac acgaagtctt ggtggcggtg 360

gccagaaaca ctgtccattg caagggcata gggatgcgtt ccttcacctc tcatttctca 420

tttctgaatc cctccctgct cactctttct cctcctcctt cccgttcacg cagcattcgg 480

ggcaacgagg tgggcccgtg ctcctccagg aagatgcgct tgtcctccgc catcttgcag 540

ggctcaagct gctcccaaaa ctcttgggcg ggttccggac ggacggctac cgcgggtgcg 600

gccctgaccg ccactgttcg gaagcagcgg cgctgcatgg gcagcggccg ctgcggtgcg 660

ccacggaccg catgatccac cggaaaagcg cacgcgctgg agcgcgcaga ggaccacaga 720

gaagcggaag agacgccagt actggcaagc aggctggtcg gtgccat 767

<210> 256

<211> 9426

<212> DNA

<213> Artificial Sequence

<220><223> synthesized construct

<400> 256

cactggcagc agccactggt aacaggatta gcagagcgag gtatgtaggc ggtgctacag 60

agttcttgaa gtggtggcct aactacggct acactagaag gacagtattt ggtatctgcg 120

ctctgctgaa gccagttacc ttcggaaaaa gagttggtag ctcttgatcc ggcaaacaaa 180

ccaccgctgg tagcggtggt ttttttgttt gcaagcagca gattacgcgc agaaaaaaag 240

gatctcaaga gatcctttg atcttttcta cggggtctga cgctcagtgg aacgaaaact 300

cacgttaagg gattttggtc atgagattat caaaaaggat cttcacctag atccttttaa 360

attaaaaatg aagttttaaa tcaatctaaa gtatatatga gtaaacttgg tctgacagtt 420

accaatgctt aatcagtgag gcacctatct cagcgatctg tctatttcgt tcatccatag 480

ttgcctgact ccccgtcgtg tagataacta cgatacggga gggcttacca tctggcccca 540

gtgctgcaat gataccgcga gacccacgct caccggctcc agatttatca gcaataaacc 600

agccagccgg aagggccgag cgcagaagtg gtcctgcaac tttatccgcc tccatccagt 660

ctattaattg ttgccgggaa gctagagtaa gtagttcgcc agttaatagt ttgcgcaacg 720

ttgttgccat tgctacaggc atcgtggtgt cacgctcgtc gtttggtatg gcttcattca 780

gctccggttc ccaacgatca aggcgagtta catgatcccc catgttgtgc aaaaaagcgg 840

ttagctcctt cggtcctccg atcgttgtca gaagtaagtt ggccgcagtg ttatcactca 900

tggttatggc agcactgcat aattctctta ctgtcatgcc atccgtaaga tgcttttctg 960

tgactggtga gtactcaacc aagtcattct gagaatagtg tatgcggcga ccgagttgct 1020

cttgcccggc gtcaatacgg gataataccg cgccacatag cagaacttta aaagtgctca 1080

tcattggaaa acgttcttcg gggcgaaaac tctcaaggat cttaccgctg ttgagatcca 1140

gttcgatgta acccactcgt gcacccaact gatcttcagc atcttttact ttcaccagcg 1200

tttctgggtg agcaaaaaca ggaaggcaaa atgccgcaaa aaagggaata agggcgacac 1260

ggaaatgttg aatactcata ctcttccttt ttcaatatta ttgaagcatt tatcagggtt 1320

attgtctcat gagcggatac atatttgaat gtatttagaa aaataaacaa ataggggttc 1380

cgcgcacatt tccccgaaaa gtgccacctg acgtctaaga aaccattatt atcatgacat 1440

taacctataa aaataggcgt atcacgaggc cctttcgtct cgcgcgtttc ggtgatgacg 1500

gtgaaaacct ctgacacatg cagctcccgg agacggtcac agcttgtctg taagcggatg 1560

ccgggagcag acaagcccgt cagggcgcgt cagcgggtgt tggcgggtgt cggggctggc 1620

ttaactatgc ggcatcagag cagattgtac tgagagtgca ccatatgcgg tgtgaaatac 1680

cgcacagatg cgtaaggaga aaataccgca tcaggcgcca ttcgccattc aggctgcgca 1740

actgttggga agggcgatcg gtgcgggcct cttcgctatt acgccagctg gcgaaagggg 1800

gatgtgctgc aaggcgatta agttgggtaa cgccagggtt ttcccagtca cgacgttgta 1860

aaacgacggc cagtgaattg atgcatgctc ttcgccgccg ccactcctgc tcgagcgcgc 1920

ccgcgcgtgc gccgccagcg ccttggcctt ttcgccgcgc tcgtgcgcgt cgctgatgtc 1980

catcaccagg tccatgaggt ctgccttgcg ccggctgagc cactgcttcg tccgggcggc 2040

caagaggagc atgagggagg actcctggtc cagggtcctg acgtggtcgc ggctctggga 2100

gcgggccagc atcatctggc tctgccgcac cgaggccgcc tccaactggt cctccagcag 2160

ccgcagtcgc cgccgaccct ggcagaggaa gacaggtgag gggggtatga attgtacaga 2220

acaaccacga gccttgtcta ggcagaatcc ctaccagtca tggctttacc tggatgacgg 2280

cctgcgaaca gctgtccagc gaccctcgct gccgccgctt ctcccgcacg cttctttcca 2340

gcaccgtgat ggcgcgagcc agcgccgcac gctggcgctg cgcttcgccg atctgaggac 2400

agtcggggaa ctctgatcag tctaaacccc cttgcgcgtt agtgttgcca tcctttgcag 2460

accggtgaga gccgacttgt tgtgcgccac cccccacacc acctcctccc agaccaattc 2520

tgtcaccttt ttggcgaagg catcggcctc ggcctgcaga gaggacagca gtgcccagcc 2580

gctgggggtt ggcggatgca cgctcaggta ccctttcttg cgctatgaca cttccagcaa 2640
aaggtagggc gggctgcgag acggcttccc ggcgctgcat gcaacaccga tgatgcttcg 2700
acccccccgaa gctccttcgg ggctgcatgg gcgctccgat gccgctccag ggcgagcgct 2760
gtttaaatag ccaggccccc gattgcaaag acattatagc gagctaccaa agccatattc 2820
aaacacctag atcactacca cttctacaca ggccactcga gcttgtgatc gcactccgct 2880
aaggggggcgc ctcttcctct tcgtttcagt cacaacccgc aaacggcgcg ccatgctgct 2940
gcaggccttc ctgttcctgc tggccggctt cgccgccaag atcagcgcct ccatgacgaa 3000

cgagacgtcc gaccgccccc tggtgcactt cacccccaac aagggctgga tgaacgaccc 3060
caacggcctg tggtacgacg agaaggacgc caagtggcac ctgtacttcc agtacaaccc 3120
gaacgacacc gtctggggga cgcccttgtt ctggggccac gccacgtccg acgacctgac 3180
caactgggag gaccagccca tcgccatcgc cccgaagcgc aacgactccg gcgccttctc 3240
cggctccatg gtggtggact acaacaacac ctccggcttc ttcaacgaca ccatcgaccc 3300
gcgccagcgc tgcgtggcca tctggaccta caacaccccg gagtccgagg agcagtacat 3360
ctcctacagc ctggacggcg gctacacctt caccgagtac cagaagaacc ccgtgctggc 3420

cgccaactcc acccagttcc gcgacccgaa ggtcttctgg tacgagccct cccagaagtg 3480
gatcatgacc gcggccaagt cccaggacta caagatcgag atctactcct ccgacgacct 3540
gaagtcctgg aagctggagt ccgcgttcgc caacgagggc ttcctcggct accagtacga 3600
gtgccccggc ctgatcgagg tccccaccga gcaggacccc agcaagtcct actgggtgat 3660
gttcatctcc atcaaccccg gcgccccggc cggcggctcc ttcaaccagt acttcgtcgg 3720
cagcttcaac ggcacccact tcgaggcctt cgacaaccag tcccgcgtgg tggacttcgg 3780
caaggactac tacgccctgc agaccttctt caacaccgac ccgacctacg ggagcgccct 3840

gggcatcgcg tgggcctcca actgggagta ctccgccttc gtgcccacca acccctggcg 3900
ctcctccatg tccctcgtgc gcaagttctc cctcaacacc gagtaccagg ccaacccgga 3960
gacggagctg atcaacctga aggccgagcc gatcctgaac atcagcaacg ccggcccctg 4020
gagccggttc gccaccaaca ccacgttgac gaaggccaac agctacaacg tcgacctgtc 4080
caacagcacc ggcaccctgg agttcgagct ggtgtacgcc gtcaacacca cccagacgat 4140
ctccaagtcc gtgttcgcgg acctctccct ctggttcaag ggcctggagg accccgagga 4200
gtacctccgc atgggcttcg aggtgtccgc gtcctccttc ttcctggacc gcgggaacag 4260

caaggtgaag ttcgtgaagg agaaccccta cttcaccaac cgcatgagcg tgaacaacca 4320

gcccttcaag agcgagaacg acctgtccta ctacaaggtg tacggcttgc tggaccagaa 4380

catcctggag ctgtacttca acgacggcga cgtcgtgtcc accaacacct acttcatgac 4440

caccgggaac gccctgggct ccgtgaacat gacgacgggg gtggacaacc tgttctacat 4500

cgacaagttc caggtgcgcg aggtcaagtg acaattggca gcagcagctc ggatagtatc 4560

gacacactct ggacgctggt cgtgtgatgg actgttgccg ccacacttgc tgccttgacc 4620

tgtgaatatc cctgccgctt ttatcaaaca gcctcagtgt gtttgatctt gtgtgtacgc 4680

gcttttgcga gttgctagct gcttgtgcta tttgcgaata ccacccccag catcccccttc 4740

cctcgtttca tatcgcttgc atcccaaccg caacttatct acgctgtcct gctatccctc 4800

agcgctgctc ctgctcctgc tcactgcccc tcgcacagcc ttggtttggg ctccgcctgt 4860

attctcctgg tactgcaacc tgtaaaccag cactgcaatg ctgatgcacg ggaagtagtg 4920

ggatgggaac acaaatggag gatcccgcgt ctcgaacaga gcgcgcagag gaacgctgaa 4980

ggtctcgcct ctgtcgcacc tcagcgcggc atacaccaca ataaccacct gacgaatgcg 5040

cttggttctt cgtccattag cgaagcgtcc ggttcacaca cgtgccacgt tggcgaggtg 5100

gcaggtgaca atgatcggtg gagctgatgg tcgaaacgtt cacagcctag ggatatcgaa 5160

ttcggccgac aggacgcgcg tcaaaggtgc tggtcgtgta tgccctggcc ggcaggtcgt 5220

tgctgctgct ggttagtgat tccgcaaccc tgattttggc gtcttatttt ggcgtggcaa 5280

acgctggcgc ccgcgagccg ggccggcggc gatgcggtgc cccacggctg ccggaatcca 5340

agggaggcaa gagcgcccgg gtcagttgaa gggctttacg cgcaaggtac agccgctcct 5400

gcaaggctgc gtggtggaat tggacgtgca ggtcctgctg aagttcctcc accgcctcac 5460

cagcggacaa agcaccggtg tatcaggtcc gtgtcatcca ctctaaagag ctcgactacg 5520

acctactgat ggccctagat tcttcatcaa aaacgcctga gacacttgcc caggattgaa 5580

actccctgaa gggaccacca ggggccctga gttgttcctt ccccccgtgg cgagctgcca 5640

gccaggctgt acctgtgatc gaggctggcg ggaaaatagg cttcgtgtgc tcaggtcatg 5700

ggaggtgcag gacagctcat gaaacgccaa caatcgcaca attcatgtca agctaatcag 5760

ctatttcctc ttcacgagct gtaattgtcc caaaattctg gtctaccggg ggtgatcctt 5820

cgtgtacggg cccttccctc aaccctaggt atgcgcgcat gcggtcgccg cgcaactcgc 5880

gcgagggccg agggtttggg acgggccgtc ccgaaatgca gttgcacccg gatgcgtggc 5940

acctttttg cgataattta tgcaatggac tgctctgcaa aattctggct ctgtcgccaa 6000

ccctaggatc agcggcgtag gatttcgtaa tcattcgtcc tgatggggag ctaccgacta 6060

ccctaatatc agcccgactg cctgacgcca gcgtccactt ttgtgcacac attccattcg 6120

tgcccaagac atttcattgt ggtgcgaagc gtccccagtt acgctcacct gtttcccgac 6180

ctccttactg ttctgtcgac agagcgggcc cacaggccgg tcgcagccac tagtatggtg 6240

gtggccgccg ccgccagcag cgccttcttc cccgtgcccg ccccccgccc cacccccaag 6300

cccggcaagt tcggcaactg gcccagcagc ctgagccagc ccttcaagcc caagagcaac 6360

cccaacggcc gcttccaggt gaaggccaac gtgagccccc acgggcgcgc ccccaaggcc 6420

aacggcagcg ccgtgagcct gaagtccggc agcctgaaca ccctggagga ccccccccagc 6480

agccccccccc cccgcacctt cctgaaccag ctgcccgact ggagccgcct gcgcaccgcc 6540

atcaccaccg tgttcgtggc cgccgagaag cagttcaccc gcctggaccg caagagcaag 6600

cgccccgaca tgctggtgga ctggttcggc agcgagacca tcgtgcagga cggcctggtg 6660

ttccgcgagc gcttcagcat ccgcagctac gagatcggcg ccgaccgcac cgccagcatc 6720

gagaccctga tgaaccacct gcaggacacc agcctgaacc actgcaagag cgtgggcctg 6780

ctgaacgacg gcttcggccg caccccccgag atgtgcaccc gcgacctgat ctgggtgctg 6840

accaagatgc agatcgtggt gaaccgctac cccacctggg gcgacaccgt ggagatcaac 6900

agctggttca gccagagcgg caagatcggc atgggccgcg agtggctgat cagcgactgc 6960

aacaccggcg agatcctggt gcgcgccacc agcgcctggg ccatgatgaa ccagaagacc 7020

cgccgcttca gcaagctgcc ctgcgaggtg cgccaggaga tcgcccccca cttcgtggac 7080

gcccccccccg tgatcgagga caacgaccgc aagctgcaca agttcgacgt gaagaccggc 7140

gacagcatct gcaagggcct gaccccggc tggaacgact tcgacgtgaa ccagcacgtg 7200

agcaacgtga agtacatcgg ctggattctg gagagcatgc ccaccgaggt gctggagacc 7260

caggagctgt gcagcctgac cctggagtac cgccgcgagt gcggccgcga gagcgtggtg 7320

gagagcgtga ccagcatgaa ccccagcaag gtgggcgacc gcagccagta ccagcacctg 7380

ctgcgcctgg aggacggcgc cgacatcatg aagggccgca ccgagtggcg ccccaagaac 7440

gccggcacca accgcgccat cagcacctga ttaattaact cgaggcagca gcagctcgga 7500

tagtatcgac acactctgga cgctggtcgt gtgatggact gttgccgcca cacttgctgc 7560

cttgacctgt gaatatccct gccgctttta tcaaacagcc tcagtgtgtt tgatcttgtg 7620

tgtacgcgct tttgcgagtt gctagctgct tgtgctattt gcgaatacca cccccagcat 7680

ccccttccct cgtttcatat cgcttgcatc ccaaccgcaa cttatctacg ctgtcctgct 7740

atccctcagc gctgctcctg ctcctgctca ctgcccctcg cacagccttg gtttgggctc 7800

cgcctgtatt ctcctggtac tgcaacctgt aaaccagcac tgcaatgctg atgcacggga 7860

agtagtggga tgggaacaca aatggaaagc ttgagctctt gttttccaga aggagttgct 7920

ccttgagcct ttcattctca gcctcgataa cctccaaagc cgctctaatt gtggaggggg 7980

ttcgaattta aaagcttgga atgttggttc gtgcgtctgg aacaagccca gacttgttgc 8040

tcactgggaa aaggaccatc agctccaaaa aacttgccgc tcaaaccgcg tacctctgct 8100
ttcgcgcaat ctgccctgtt gaaatcgcca ccacattcat attgtgacgc ttgagcagtc 8160
tgtaattgcc tcagaatgtg gaatcatctg ccccctgtgc gagcccatgc caggcatgtc 8220
gcgggcgagg acacccgcca ctcgtacagc agaccattat gctacctcac aatagttcat 8280
aacagtgacc atatttctcg aagctcccca acgagcacct ccatgctctg agtggccacc 8340
ccccggccct ggtgcttgcg gagggcaggt caaccggcat ggggctaccg aaatccccga 8400
ccggatccca ccacccccgc gatgggaaga atctctcccc gggatgtggg cccaccacca 8460

gcacaacctg ctggcccagg cgagcgtcaa accataccac acaaatatcc ttggcatcgg 8520
ccctgaattc cttctgccgc tctgctaccc ggtgcttctg tccgaagcag gggttgctag 8580
ggatcgctcc gagtccgcaa acccttgtcg cgtggcgggg cttgttcgag cttgaagagc 8640
ctctagagtc gacctgcagg catgcaagct tggcgtaatc atggtcatag ctgtttcctg 8700
tgtgaaattg ttatccgctc acaattccac acaacatacg agccggaagc ataaagtgta 8760
aagcctgggg tgcctaatga gtgagctaac tcacattaat tgcgttgcgc tcactgcccg 8820
ctttccagtc gggaaacctg tcgtgccagc tgcattaatg aatcggccaa cgcgcgggga 8880

gaggcggttt gcgtattggg cgctcttccg cttcctcgct cactgactcg ctgcgctcgg 8940
tcgttcggct gcggcgagcg gtatcagctc actcaaaggc ggtaatacgg ttatccacag 9000
aatcagggga taacgcagga aagaacatgt gagcaaaagg ccagcaaaag gccaggaacc 9060
gtaaaaaggc cgcgttgctg gcgtttttcc ataggctccg ccccccctgac gagcatcaca 9120
aaaatcgacg ctcaagtcag aggtggcgaa acccgacagg actataaaga taccaggcgt 9180
ttcccccctgg aagctccctc gtgcgctctc ctgttccgac cctgccgctt accggatacc 9240
tgtccgcctt tctcccttcg ggaagcgtgg cgctttctca tagctcacgc tgtaggtatc 9300

tcagttcggt gtaggtcgtt cgctccaagc tgggctgtgt gcacgaaccc cccgttcagc 9360
ccgaccgctg cgccttatcc ggtaactatc gtcttgagtc caacccggta agacacgact 9420
tatcgc 9426