



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2013년02월19일
(11) 등록번호 10-1234198
(24) 등록일자 2013년02월12일

(51) 국제특허분류(Int. Cl.)
C12N 15/10 (2006.01) C07D 407/12 (2006.01)
(21) 출원번호 10-2005-7018067
(22) 출원일자(국제) 2004년03월26일
심사청구일자 2009년03월20일
(85) 번역문제출일자 2005년09월26일
(65) 공개번호 10-2006-0006014
(43) 공개일자 2006년01월18일
(86) 국제출원번호 PCT/US2004/009323
(87) 국제공개번호 WO 2004/087879
국제공개일자 2004년10월14일
(30) 우선권주장
60/457,979 2003년03월26일 미국(US)
(56) 선행기술조사문헌
W02002083870 A1
W0200042195 A1

(73) 특허권자
마텍 바이오사이언스 코퍼레이션
미합중국 메릴랜드 21045 콜롬비아 도빈로드 6480
(72) 발명자
메츠, 제임스, 지.
미국, 콜로라도 80501 롱몬트, 16번 아베뉴, 542 이.
위버, 크래그, 에이.
미국, 콜로라도 80301, 보울더, 스타보드 드라이브 4542
(74) 대리인
(뒷면에 계속)
이세진, 유미특허법인

전체 청구항 수 : 총 16 항

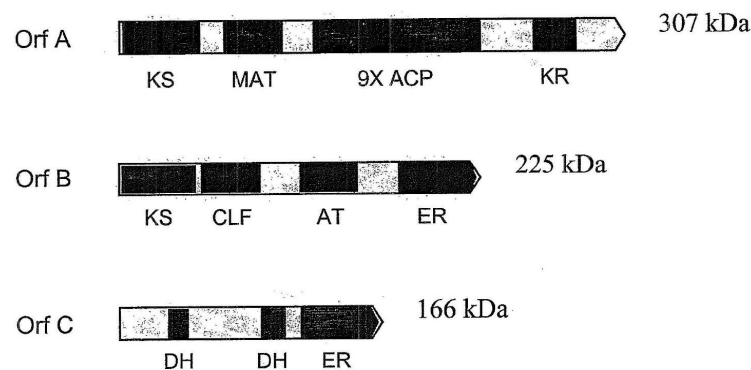
심사관 : 김윤경

(54) 발명의 명칭 P U F A 폴리케타이드 신타제 시스템 및 이의 용도

(57) 요약

본 발명은 다중불포화 지방산(PUFA) 폴리케타이드 신타제(PKS) 시스템, 이의 상동체, 그러한 PUFA PKS 시스템의 생물학적으로 활성인 도메인을 코딩하는 분리된 핵산 분자 및 재조합 핵산 분자, PUFA PKS 시스템을 포함하는 유전적으로 변형된 유기체, 생활성 분자의 제조를 위하여 그러한 시스템을 제조하고 이용하는 방법 및 그러한 PUFA PKS 시스템을 가진 새로운 박테리아 및 비박테리아의 미생물을 동정하는 신규한 방법에 관한 것이다.

대표도



(72) 발명자

바클레이, 윌리엄, 알.

미국, 콜로라도 80303, 보울더, 파노라마 드라이브
7356

플래트, 제임스, 에이치.

미국, 콜로라도 80503, 롱몬트, 콰이어트 리트리트
코트 7024

특허청구의 범위

청구항 1

다음으로 이루어진 그룹으로부터 선택되는 핵산 서열을 포함하는 분리된 핵산 분자:

(a) 서열번호: 62, 서열번호: 64, 서열번호: 66, 및 서열번호: 68로 이루어진 그룹으로부터 선택된 아미노산 서열을 코딩하는 핵산 서열;

(b) 서열번호: 61, 서열번호: 63, 서열번호: 65, 및 서열번호: 67로 이루어진 그룹으로부터 선택된 핵산 서열; 및

(c) 상기 (a) 또는 (b)의 핵산 서열에 완전히 상보적인 핵산 서열.

청구항 2

하나 이상의 전사 조절 서열에 작동가능하게 연결된 제1항에 따른 핵산 분자를 포함하는 재조합 핵산 분자.

청구항 3

제2항에 따른 재조합 핵산 분자로 형질감염된 재조합 세포.

청구항 4

제1항에 따른 핵산 분자 하나 이상에 의한 형질전환에 의해 유전적으로 변형되고, 다중불포화 지방산 (PUFA: polyunsaturated fatty acid) 폴리케타이드 신타제(PKS polyketide synthase) 시스템을 발현하는, 유전적으로 변형된 미생물.

청구항 5

제4항에 있어서,

상기 미생물은 내생적으로 PUFA PKS 시스템을 발현하며, 상기 유전적 변형은, 내생 PUFA PKS 시스템으로부터의 도메인을, 상기 미생물을 형질전환시킨 핵산 분자에 의해 암호화된 해당 도메인으로의 치환을 포함하는 것인, 유전적으로 변형된 미생물.

청구항 6

제4항에 있어서,

상기 미생물은 트라우스토카이트리드(*Thraustochytrid*)인, 유전적으로 변형된 미생물.

청구항 7

제6항에 있어서,

상기 미생물은 스킴조카이트리움(*Schizochytrium*)인, 유전적으로 변형된 미생물.

청구항 8

삭제

청구항 9

삭제

청구항 10

삭제

청구항 11

삭제

청구항 12

삭제

청구항 13

삭제

청구항 14

삭제

청구항 15

삭제

청구항 16

삭제

청구항 17

제4항 내지 제7항 중 어느 한 항에 있어서,

상기 미생물은, 포스포판테테인 트랜스퍼라제(phosphopantetheine transferase)를 코딩하는 재조합 핵산 분자를 발현하도록 유전적으로 더 변형된, 유전적으로 변형된 미생물.

청구항 18

삭제

청구항 19

삭제

청구항 20

삭제

청구항 21

삭제

청구항 22

제1항에 따른 핵산 분자 하나 이상을 재조합적으로 발현하도록 유전적으로 변형된, 유전적으로 변형된 식물.

청구항 23

제 22항에 있어서,

상기 식물은, 포스포판테테인 트랜스퍼라제(phosphopantetheine transferase)를 코딩하는 재조합 핵산 분자를 발현하도록 유전적으로 더 변형된, 유전적으로 변형된 식물.

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

식이 보조제, 식료품, 약제학적 제형, 인간화된 동물의 밀크(humanized animal milk), 및 분유로 이루어진 그룹으로부터 선택되는 조성물의 다중불포화 지방산 함량을 증가시키는 방법으로서, 상기 방법은, 제1항에 따른 핵산 분자 하나 이상을 재조합적으로 발현하도록 유전적으로 변형된 재조합 숙주 세포에 의해 생산된 다중 불포화 지방산을 상기 조성물에 첨가하는 것을 포함하는 방법.

청구항 39

삭제

청구항 40

제38항에 있어서,

상기 약제학적 제형은 항염증 제형, 화학요법제, 활성 부형제, 골다공증 약, 항우울제, 항경련제, 항-헬리코박터 파일로리 약, 신경퇴행성 질환의 치료를 위한 약, 퇴행성 간 질환의 치료를 위한 약, 항생제, 및 콜레스테롤 저해 제형으로 이루어진 그룹으로부터 선택되는 방법.

청구항 41

제 38항에 있어서,

상기 조성물은 만성 염증, 급성 염증, 위장관계 질환, 암, 카렉시아, 심장의 재협착증, 신경퇴행성 질환, 간의 퇴행성 질환, 혈액 리피드 질환, 골다공증, 동맥경화, 자가면역 질환, 자간전증, 조산, 노화와 연관된 황달변증, 호흡기계 질환, 및 폐육시조말 질환으로 이루어진 그룹으로부터 선택된 용태를 치료하는데 이용되는 방법.

청구항 42

밀크 생산 세포가 밀크에 다중 불포화지방산을 생산하게 하기 위하여, 제1항에 따른 핵산 분자 하나 이상을 유전적으로 발현하도록, 밀크 생산 동물의 밀크 생산 세포를 유전적으로 변형하는 것을 포함하는, 다중불포화 지방산을 함유하는 인간화된 동물 밀크의 제조 방법.

청구항 43

유전적으로 변형된 트라우스카이트리드 미생물로서, 상기 미생물은 제1항에 따른 핵산 분자 하나 이상을 포함하는 내인성 다중불포화 지방산 (PUFA) 폴리카타이드 신타제 (PKS) 시스템을 보유하고, 상기 내인성 PUFA PKS 시스템은, 유전적으로 변형되지 않은 트라우스토키타리드 미생물에 비해, 트라우스카이트리드 미생물에 의한 다중불포화 지방산(PUFA)의 생산 프로필을 변경하도록 유전적으로 변형된 것인, 유전적으로 변형된 트라우스토키타리드 미생물.

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

제43항에 있어서,

상기 PUFA 생산 프로필이, 미생물에 의한 도코사헥사엔산(DHA), 에이코사펜타엔산(EPA), 도코사펜타엔산(DPA)의 하나 또는 두 개 모두의 이성체, 또는 아라키돈산(ARA)의 생산을 개시, 증가, 또는 감소시키도록 변경된 것인, 유전적으로 변형된 트라우스토카이트리드 미생물.

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

명세서

기술분야

[0001] 본 발명은 트라우스토카이트리드 미생물과 같은 진핵 유기체를 포함한 미생물로부터 유래된 다중불포화 지방산(polyunsaturated fatty acid, PUFA) 폴리케타이드 신타제(polyketide synthase systems, PKS) 시스템에 관한 것이다. 보다 자세하게는 본 발명은 비박테리아의(non-bacterial) PUFA PKS 시스템을 코딩하는 핵산, 비박테리아의 PUFA PKS 시스템, 비박테리아의 PUFA PKS 시스템을 포함하는 유전적으로 변형된 유기체 및 본 발명에 개시된 비박테리아의 PUFA PKS 시스템을 제조하고 이용하는 방법에 관한 것이다. 또한, 본 발명은 PUFA 폴리케타이드 신타제 (PKS) 시스템의 조작에 의해 다양한 다중불포화 지방산(PUFA), 특히 에이코사펜타엔산(C20:5, ω -3; EPA)이 풍부한 리피드(멤브레인과 관련된 포스포리피드(PL)뿐만 아니라 트리아실글리세롤(TAG))를 효과적으로 제조하는 유전적으로 변형된 미생물 및 방법에 관한 것이다.

배경기술

[0002] 폴리케타이드 신타제(PKS) 시스템은 지방산 신타제(FAS) 시스템에서 유래된 효소 복합체로서 당업계에 공지되어 있지만, 종종 변형되어 지방산과 거의 유사함을 보이지 않는 특별한 생산물을 제조하기도 한다. 그러나 폴리케타이드 신타제 시스템은 말로닐 Co-A로부터 PUFA를 합성할 수 있는 해양 박테리아 및 특정한 미세조류(microalgae)에 존재하는 것으로 보여져왔다. 셰와넬라(*Shewanella*) 및 다른 해양 미생물인 비브리오 마리누스(*Vibrio marinus*)의 PUFA 합성을 위한 PKS 경로는 미국특허 제6,140,486호에 상세히 개시되어 있다. 진핵의 트라우스토카이트리드(Thraustochytrid)인 스킨조카이트리움(*Schizochytrium*)의 PUFA 합성을 위한 PKS 경로는 미국특허 제6,566,583호에 상세히 개시되어 있다. 또한, 스킨조카이트리움에서 PUFA PKS 경로의 완전한 구조적인 묘사를 포함한 트라우스토카이트리알레(Thraustochytriales)의 멤버와 같은 진핵에서 PUFA 합성을 위한 PKS 경로와 트라우스토카이트리움에서 PUFA PKS 경로의 동정이, 이 경로의 이용에 관한 상세 설명을 포함하여 (2002년 4월 16일에 출원된 미국특허출원 제10/124,800호에 대응되는) 2002년 12월 19일에 공개된 미국특허공개 제20020194641호에 상세히 개시되어 있다.

[0003] 연구원들은 상기 문헌에 개시된 폴리케타이드 신타제(PKS) 시스템을 타입 II, 타입 I 및 모듈라(modular)로 언급되는 세 가지 기본 타입 중 어느 하나로 분류함으로써 개발을 시도하였다. 타입 II 시스템은 각각이 개별적인 효소 반응을 수행하는 분리될 수 있는 단백질에 의해 특징지어진다. 상기 효소는 협력하여 최종 생산물을 제조하고 상기 시스템의 각각의 효소는 최종 생산물의 제조에 여러 차례 참여한다. 이러한 타입의 시스템은 식물과 박테리아에서 발견된 지방산 신타제(FAS) 시스템과 유사한 방식으로 작동한다. 타입 I PKS 시스템은 효소가 최종 생산물을 제조하기 위하여 반복적으로 사용된다는 점에서 타입 II 시스템과 유사하다. 타입 I은 효소 활성이 분리될 수 있는 단백질과 연관되어 있는 대신에 더 큰 단백질의 도메인으로서 존재한다는 점에서 타입 II와 상이하다. 상기 시스템은 동물과 진균에서 발견되는 타입 I FAS 시스템과 유사하다.

[0004] 타입 I 및 타입 II 시스템과는 대조적으로 모듈라 PKS 시스템에서 각 효소 도메인은 최종 생산물의 제조에서 단지 한차례만 이용된다. 도메인은 매우 큰 단백질에서 발견되고 각 반응의 산물은 PKS 단백질에 있는 다른 도메인으로 전달된다. 또한, 상술한 모든 PKS 시스템에서 최종 산물에 탄소-탄소 이중 결합이 통합된다면 항상 트랜스 배형(configuration)이다.

[0005] 상술한 타입 I 및 타입 II PKS 시스템에서는 최종 생산물이 수득될 때까지 각 사이클마다 동일한 반응의 세트가 수행된다. 상기 생합성 과정 중에 독특한 반응이 도입될 가능성은 없다. 모듈라 PKS 시스템은 반복적인 반응의 효율을 이용하지 않는 거대한 단백질을 요구한다(즉, 개별적인 도메인이 각 반응마다 요구된다). 또한, 탄소-탄소 이중 결합은 상술한 모든 PKS 시스템에서 트랜스 배형으로 도입된다.

[0006] 다중불포화 지방산(PUFA)은 대부분의 진핵생물에서 멤브레인 리피드의 주요 성분이고(Lauritzen et al. *Prog. Lipid Res.* **40** 1(2001); McConn et al. *Plant J.* **15**, 521 (1998)) 특정 호르몬과 신호 분자의 전구체이다(Heller et al. *Drug* **55**, 487(1998); Creelman et al. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **48**, 355 (1997)). 공지된 PUFA 합성 경로에는 연장(elongation)과 호기성의 불포화(desaturation) 반응에 의한 지방산 신타제(FAS)로부터 유래된 포화 16:0 또는 18:0 지방산(약자 X:Y에서 X 탄소 원소와 (일반적으로 PUFA에서 시스인) Y 이중 결합을 포함한 아실기를 의미한다; PUFA의 이중 결합 위치는 지방산 쇄의 메틸 탄소(ω 3 또는 ω 6)에 대하여 이중 결합의 체계적인 메틸렌 도입으로 나타낸다)의 처리를 포함한다(Sprecher, *Curr. Opin. Clin. Nutr. Metab. Care* **2**, 135 (1999); Parker-Barnes et al. *Proc. Natl. Acad. Sci. USA* **97**, 8284 (2000); Shanklin et al. *Annu Rev. Plant Physiol. Plant Mol. Biol.* **49**, 611 (1998)). 아세틸-CoA로 시작하여 도코사헥사엔산(DHA)은 약 30개의 개별적인 효소 활성화와 지방산 합성 사이클의 네 번의 반복적인 단계를

포함하여 거의 70개의 반응을 요구한다. 폴리케타이드 신타제(PKS)는 FAS와 일부 동일한 반응을 수행하고 (Hopwood et al. *Annu. Rev. Genet.* **24**, 37(1990); Bentley et al. *Annu Rev. Microbiol.* **53**, 411 (1999)) 성장하는 탄소쇄에 공유결합의 부착 위치로서 동일한 작은 단백질(또는 도메인)인 아실 캐리어 단백질(ACP)을 이용한다. 그러나 이러한 효소 시스템에서는 FAS에서 보여지는 환원, 탈수 및 환원의 완전한 사이클은 종종 생략되어 일반적으로 트랜스 배형의 탄소-탄소 이중 결합 뿐만 아니라 많은 케토- 및 하이드록시기를 포함한 과생된 탄소쇄가 제조된다. PKS의 일차 산물은 종종 순환되어 항생제와 많은 다른 이차 산물을 포함한 복잡한 생화학 물질을 형성한다(상기된 Hopwood et al.(1990); Bentley et al. (1999); Keating et al. *Curr. Opin. Chem Biol.* **3**, 598(1999)).

[0007]

도코사헥사엔산(DHA; 22:6 ω 3) 및 에이코사펜타엔산(EPA; 20:5 ω 3)과 같은 장쇄의 PUFA가 *쉐와넬라* sp를 포함한 여러 종의 해양 박테리아로부터 보고되었다(Nichols et al. *Curr. Op. Biotechnol.* **10**, 240(1999); Yazawa. *Lipids* 31, S (1996); Delong et al. *Appl Environ. Microbiol.* **51**, 730 (1986)). *쉐와넬라* sp 균주 SCRC2738 유래의 (플라스미드 pEPA로서 클론된) 게놈 단편의 분석은 대장균에서 EPA 제조에 필수적이고 충분한 총 20Kb인 다섯 개의 오픈 리딩 프레임(Orfs)의 동정으로 이어졌다(상기 Yazawa, (1996)). 추정된 단백질 도메인 중 몇 개는 FAS 효소의 상동체인 반면에 다른 영역은 공지된 기능의 단백질과 상동성을 보이지 않았다. 다섯 개의 Orfs 내에서 적어도 11개의 영역의 추정된 효소 도메인으로서 동정되었다(Metz et al. *Science* 293:290-293(2001) 참조). 유전자 데이터베이스에 있는 서열들과 비교하였을 때 이들 중 일곱 개가 FAS 단백질 보다는 PKS 단백질과 더욱 관련성이 있었다. 상기 그룹 내에는 말로닐-CoA:ACP 아실트랜스퍼라제(MAT), β -케토아실-ACP 신타제(KS), β -케토아실-ACP 환원효소(KR), 아실트랜스퍼라제(AT), 포스포판테테인 트랜스퍼라제, 쉐 길이 (또는 쉐 개시) 인자(CLF) 및 독특한 여섯 개의 ACP 도메인 무리(즉, PKS 또는 FAS 서열에서 종래에는 보고된 바 없는 둘 이상의 무리를 이룬 ACP 도메인의 존재)을 코딩하는 것으로 추정되는 도메인이 포함되어 있다. *쉐와넬라*에서 동정된 PUFA 합성을 위한 PKS 경로가 해양 박테리아에서 일반적인 것으로 보인다. *쉐와넬라* 유전자 무리와 높은 상동성을 가진 유전자가 *포토박테리움 프로펀둠*(*Photobacterium profundum*, Allen et al. *Appl. Environ Microbiol.* 65:1710 (1999)과 *모리텔라 마리나*(비브리오 마린스)(*Moritella Marina*(*Vibrio marinus*, 미국특허 제6,140,486호 및 Tanaka et al. *Biotechnol. Lett.* 21:939 (1999))에서 동정되었다.

[0008]

다중불포화 지방산(PUFA)은 영양학적, 약제학적, 산업적 및 기타 목적에 유용한 것으로 여겨진다. 천연원 및 화학적 합성에서 유래된 PUFA의 확장된 공급은 상업적 수요에는 충분하지 못하다. PUFA의 현재 주요 공급원은 해양의 어류로부터 유래된다; 그러한 어류의 재고는 감소하고 있어 이는 지속 가능한 자원일 수 없다. 또한, 중금속 및 독성의 유기 분자에 의한 오염은 해양 어류로부터 유래된 오일에서 심각한 문제가 되고 있다. 오일 종자 작물에서 유래된 식물의 오일은 상대적으로 저렴하고 어류 오일과 관련된 오염 문제로부터 자유롭다. 그러나 상업적으로 개발된 식물의 오일에서 발견되는 PUFA는 일반적으로 리놀렌산(델타 9 및 12 위치에 2개의 이중 결합을 가진 18개의 탄소 - 18:2 델타 9,12)과 리놀렌산(18:3 델타 9, 12, 15)에만 제한된다. PUFA 합성을 위한 전통적인 경로에서 중쇄 길이의 포화 지방산(지방산 신타제(FAS) 시스템의 생산물)은 일련의 연장과 탈포화 반응에 의해 변형된다. 다수의 개별 탈포화 효소 및 연장 효소들이 리놀렌산 및 리놀렌산으로부터 더욱 포화되고 더욱 긴 쉐의 PUFA를 제조하는 지방산의 합성을 위하여 요구되기 때문에 EPA 및 도코사헥사엔산(DHA)과 같은 PUFA의 발현을 위해 식물 숙주 세포를 조작하는 것은 합성을 달성하기 위하여 여러 개의 개별 효소의 발현을 요구할 것이다. 또한, 그러한 PUFA를 유용한 양으로 제조하기 위해서는 돌연변이 유발 또는 색소체(plastid) 세포 소기관으로 효소의 표적화와 같이 예를 들면 기질에 대하여 경쟁하는 효소를 하향 조정하는 조작, 더 높은 효소 활성으로의 조작과 같은 추가의 조작 노력이 요구된다. 따라서 천연적으로 이러한 지방산을 제조하는 종으로부터 유래된 PUFA 생합성에 관여하는 유전적 물질을 입수하고 이렇게 분리된 물질을 단독으로 또는 상업적인 양으로 PUFA의 제조가 가능하도록 조작될 수 있는 이형성의 시스템과 조합하여 발현시키는 것이 적절하다.

[0009]

쉐와넬라 및 비브리오 마린스와 같은 해양 박테리아에서 PUFA PKS 시스템의 발견(미국특허 제 6,140,486호 참조)은 상업적인 PUFA 제조의 새로운 방법을 위한 자원을 제공한다. 그러나 이러한 해양 박테리아는 상업적인 수준의 그들의 유용성을 제한하는 한계를 가지고 있다. 첫 번째 미국특허 제6,140,486호에는 상기 해양 박테리아의 PUFA PKS 시스템이 식물을 유전적으로 변형시키는데 이용될 수 있다고 개시되어 있으나, 상기 해양 박테리아는 차가운 해양 환경에서 천연적으로 생존하고 성장하며, 상기 박테리아의 효소 시스템은 22℃ 이상에서 제대로 기능을 발휘하지 못한다. 이와 대조적으로 PUFA PKS 시스템을 사용하는 유전적 조작의 매력적인 표적인 많은 작물은 22℃ 이상 40℃ 보다 높은 온도에서 일반적인 성장 조건을 가진다. 따라서 이러한 해양 박테리아 유래의 PUFA PKS 시스템은 일반적인 성장 조건 하에 있는 식물 발현에 용이하게 적용할 수 없을 것으로 예상된다. 또한, 공지된 해양 박테리아의 PUFA PKS 시스템은 트리아실글리세롤(TAG)을 직접적으로 제조하지

않는데, 일반적으로 낮은 수준으로 축적될 수 있는 "구조적" 리피드 생산물(예, 포스포리피드)에 반해 TAG는 리피드 저장 생산물로서, 결과적으로 세포에서 매우 높은 수준으로 축적될 수 있기 때문에 TAG의 직접적인 제조가 적합할 것이다.

[0010] 에이코사펜타엔산(EPA)의 제조와 관련하여 특히 연구원들은 광합성 및 종속영양(heterotrophic) 배양 모두에서 미생물을 성장시킴으로써 미생물에 의해 EPA를 제조하기 위하여 노력해왔다. 그들은 배양 조건하에서 유기체의 생산성을 증가시키기 위한 전통적 및 직접적인 유전적 방식 모두를 사용하였다. 다른 연구원은 다양한 탈포화 및 연장 효소를 코딩하는 유전자의 도입에 의해 오일-종자 식물에서 EPA를 제조하는 것을 시도하였다.

[0011] 연구원들은 적색 미세조류(모노두스, *Monodus*), 규조류[예, 패오닥틸룸(*Phaeodactylum*)], 다른 미세조류 및 진균[예, 낮은 온도에서 배양되는 모티렐라(*Mortierella*)]의 배양을 이용하는 시도를 해보았다. 그러나 모든 경우에서 DHA와 같은 다른 장쇄의 PUFA를 위한 종래의 상업적인 미생물의 제조 시스템과 비교하였을 때 생산성은 낮았다. 많은 경우에 EPA가 일차적으로 트리아실글리세롤(TAG) 보다는 포스포리피드(PL)에서 존재하였다. 미세 조류의 생산성은 종속영양 성장 환경 하에서 광합성 환경 하에서 보다 훨씬 높을 수 있기 때문에 연구원들은 특정한 당 트랜스포터를 코딩하는 유전자를 도입함으로써 영양의 전환을 시도하였고 달성하였다. 그러나 신규하게 수득된 종속 영양의 능력으로도 오일과 관련된 생산성은 상대적으로 낮았다.

[0012] 내인성의 지방산 생합성 경로의 변형에 의한 오일-종자 작물에서 EPA를 제조하기 위한 노력은 오일 중에 매우 낮은 수준의 PUFA를 가진 식물을 수득할 뿐이었다. 상기된 바와 같이 몇 개의 해양 박테리아가 PUFA(DHA 뿐만 아니라 EPA)를 제조하는 것으로 보여졌다. 그러나 상기 박테리아는 TAG를 제조하지 않고 EPA는 PL 멤브레인에서 일차적으로 발견된다. (상기된) 이러한 특정한 해양 박테리아의 성장 특성 뿐만 아니라 제조된 EPA의 수준은 EPA의 상업적인 제조를 위한 이들의 능력을 제한한다.

[0013] 따라서 상업적으로 이용될 수 있도록 더 높은 유동성을 가지는 또 다른 PUFA PKS 시스템과 상업적으로 유용한 제조방법으로 EPA와 같은 목적인 PUFA가 풍부한 양의 리피드를 효과적으로 제조하는 생물학적 시스템에 대한 수요가 있다.

발명의 상세한 설명

[0017] 발명의 요약

[0018] 본 발명의 한 양태는 분리된 핵산 분자에 관한 것이다. 상기 핵산 분자는 (a) 서열번호: 39, 서열번호: 41, 서열번호: 43, 서열번호: 45, 서열번호: 48, 서열번호: 50, 서열번호: 52, 서열번호: 54, 서열번호: 56, 서열번호: 58, 서열번호: 60, 서열번호: 62, 서열번호: 64, 서열번호: 66, 서열번호: 68 및 이의 생물학적으로 활성인 단편으로 이루어진 그룹으로부터 선택된 아미노산 서열을 코딩하는 핵산 분자; (b) 서열번호: 39, 서열번호: 43, 서열번호: 50, 서열번호: 52, 서열번호: 56 및 서열번호: 58로 이루어진 그룹으로부터 선택된 아미노산 서열과 적어도 약 60% 동일한, 보다 바람직하게는 적어도 약 70% 동일한, 보다 바람직하게는 적어도 약 80% 동일한, 보다 바람직하게는 적어도 약 90% 동일하되, 다중불포화 지방산(PUFA) 폴리케타이드 신타제(PKS) 시스템의 적어도 한 도메인의 생물학적 활성을 가지는 아미노산 서열을 코딩하는 핵산 서열; (c) 서열번호: 54의 아미노산 서열과 적어도 약 65% 동일한, 보다 바람직하게는 적어도 약 70% 동일한, 보다 바람직하게는 적어도 약 80% 동일한, 보다 바람직하게는 적어도 약 90% 동일하되, 다중불포화 지방산(PUFA) 폴리케타이드 신타제(PKS) 시스템의 적어도 한 도메인의 생물학적 활성을 가지는 아미노산 서열을 코딩하는 핵산 서열; (d) 서열번호: 45, 서열번호: 48, 서열번호: 60, 서열번호: 62 및 서열번호: 64로 이루어진 그룹으로부터 선택된 아미노산 서열과 적어도 약 70% 동일한, 보다 바람직하게는 적어도 약 80% 동일한, 보다 바람직하게는 적어도 약 90% 동일하되, 다중불포화 지방산(PUFA) 폴리케타이드 신타제(PKS) 시스템의 적어도 한 도메인의 생물학적 활성을 가지는 아미노산 서열을 코딩하는 핵산 서열; (e) 서열번호: 41, 서열번호: 66, 서열번호: 68로 이루어진 그룹으로부터 선택된 아미노산 서열과 적어도 약 80% 동일한, 보다 바람직하게는 적어도 약 90% 동일하되, 다중불포화 지방산(PUFA) 폴리케타이드 신타제(PKS) 시스템 중 적어도 하나의 도메인의 생물학적 활성을 가지는 아미노산 서열을 코딩하는 핵산 서열; 및/또는 (f) (a), (b), (c), (d) 또는 (e)의 핵산 서열과 완전히 상보적인 핵산 서열로부터 선택된 핵산 서열을 포함한다. 한 측면에서, 상기 핵산 서열은 서열번호: 39, 서열번호: 41, 서열번호: 43, 서열번호: 45, 서열번호: 48, 서열번호: 50, 서열번호: 52, 서열번호: 54, 서열번호: 56, 서열번호: 58, 서열번호: 60, 서열번호: 62, 서열번호: 64, 서열번호: 66, 서열번호: 68 및 이들의 생물학적으로 활성인 단편으로부터 선택된 아미노산 서열을 코딩한다. 한 측면에서, 상기 핵산 서열은 서열번호: 38, 서열번호: 40, 서열번호: 42, 서열번호: 44, 서열번호: 47, 서열번호: 49, 서열번호: 51, 서열번호: 53,

서열번호: 55, 서열번호: 57, 서열번호: 59, 서열번호: 61, 서열번호: 63, 서열번호: 65 및 서열번호: 67로 이루어진 그룹으로부터 선택된다.

[0019] 본 발명의 다른 양태는 상술한 핵산 분자 중 어느 하나를 적어도 하나의 전사 조절 서열에 작동가능하게 연결되도록 포함하는 제조합 핵산 분자에 관한 것이다.

[0020] 본 발명의 또 다른 양태는 상술한 제조합 핵산 분자로 형질감염된 제조합 세포에 관한 것이다.

[0021] 본 발명의 다른 양태는 다중불포화 지방산 (PUFA) 폴리케타이드 신타제(PKS) 시스템의 적어도 하나의 생물학적으로 활성인 도메인을 포함하는 PKS 시스템을 발현하도록 유전적으로 변형된 미생물에 관한 것으로서, 상기 PUFA PKS 시스템의 적어도 하나의 도메인은 (a) 서열번호: 39, 서열번호: 41, 서열번호: 43, 서열번호: 45, 서열번호: 48, 서열번호: 50, 서열번호: 52, 서열번호: 54, 서열번호: 56, 서열번호: 58, 서열번호: 60, 서열번호: 62, 서열번호: 64, 서열번호: 66, 서열번호: 68로 이루어진 그룹으로부터 선택된 아미노산 서열 및 이들의 생물학적으로 활성인 단편; (b) 서열번호: 39, 서열번호: 43, 서열번호: 50, 서열번호: 52, 서열번호: 56 및 서열번호: 58로 이루어진 그룹으로부터 선택된 아미노산 서열과 적어도 약 60% 동일한, 보다 바람직하게는 적어도 약 70% 동일한, 보다 바람직하게는 적어도 약 80% 동일한, 보다 바람직하게는 적어도 약 90% 동일하되, 다중불포화 지방산(PUFA) 폴리케타이드 신타제(PKS) 시스템의 적어도 하나의 도메인의 생물학적 활성을 가지는 아미노산 서열; (c) 서열번호: 54의 아미노산 서열과 적어도 약 65% 동일한, 보다 바람직하게는 적어도 약 70% 동일한, 보다 바람직하게는 적어도 약 80% 동일한, 보다 바람직하게는 적어도 약 90% 동일하되, 다중불포화 지방산(PUFA) 폴리케타이드 신타제(PKS) 시스템 중 적어도 하나의 도메인의 생물학적 활성을 가지는 아미노산 서열; (d) 서열번호: 45, 서열번호: 48, 서열번호: 60, 서열번호: 62 및 서열번호: 64로 이루어진 그룹으로부터 선택된 아미노산 서열과 적어도 약 70% 동일한, 보다 바람직하게는 적어도 약 80% 동일한, 보다 바람직하게는 적어도 약 90% 동일하되, 다중불포화 지방산(PUFA) 폴리케타이드 신타제(PKS) 시스템 중 적어도 하나의 도메인의 생물학적 활성을 가지는 아미노산 서열; 및/또는 (e) 서열번호: 41, 서열번호: 66, 서열번호: 68로 이루어진 그룹으로부터 선택된 아미노산 서열과 적어도 약 80% 동일한, 보다 바람직하게는 적어도 약 90% 동일하되, 다중불포화 지방산(PUFA) 폴리케타이드 신타제(PKS) 시스템 중 적어도 하나의 도메인의 생물학적 활성을 가지는 아미노산 서열을 포함한다. 상기 미생물은 PKS 시스템의 활성에 영향을 미치도록 유전적으로 변형된다.

[0022] 한 측면에서, 상기 미생물은 다중불포화 지방산 (PUFA) 폴리케타이드 (PKS) 시스템 중 적어도 하나의 도메인을 코딩하는 제조합 핵산 분자로 형질감염됨으로써 유전적으로 변형된다. 예를 들면, 상기 미생물은 스킨조카이트리움과 같은 트라우스토카이트리드(Thraustochytrid)를 포함할 수 있다. 한 측면에서, 그러한 미생물은 박테리아의 PUFA PKS 시스템, 타입 I PKS 시스템, 타입 II PKS 시스템, 모듈라 PKS 시스템 및 비박테리아의 PUFA PKS 시스템으로 이루어진 그룹으로부터 선택된 PKS 시스템 유래의 적어도 하나의 생물학적으로 활성인 도메인을 코딩하는 적어도 하나의 핵산 분자를 제조합에 의해 발현하도록 추가로 유전적으로 변형되었다. 비박테리아의 PUFA PKS 시스템은 트라우스토카이트리드 PUFA PKS 시스템을, 한 측면으로는 스킨조카이트리움 PUFA PKS 시스템을 포함할 수 있다.

[0023] 다른 측면에서, 상기 미생물은 PUFA PKS 시스템의 적어도 하나의 도메인을 포함하는 PKS 시스템을 내인적으로 발현하고 유전적 변형은 PUFA PKS 시스템의 적어도 하나의 도메인을 코딩하는 핵산 서열에 있다. 다른 측면에서, 상기 미생물은 박테리아의 PUFA PKS 시스템, 타입 I PKS 시스템, 타입 II PKS 시스템, 모듈라 PKS 시스템 및 비박테리아의 PUFA PKS 시스템(예, 스킨조카이트리움 PUFA PKS 시스템과 같은 트라우스토카이트리드 PUFA PKS 시스템)으로 이루어진 그룹으로부터 선택된 PKS 시스템 유래의 적어도 하나의 생물학적으로 활성인 도메인을 코딩하는 적어도 하나의 핵산 분자를 제조합에 의해 발현하도록 추가로 유전적으로 변형되었다.

[0024] 다른 측면에서, 상기 미생물은 PUFA PKS 시스템의 적어도 하나의 생물학적으로 활성인 도메인을 포함하는 PUFA PKS 시스템을 내인적으로 발현하고 유전적 변형은 제 2의 PKS 시스템에서 유래된 적어도 하나의 생물학적으로 활성인 도메인을 코딩하는 제조합 핵산 분자와 내인성의 PUFA PKS 시스템의 활성에 영향을 미치는 단백질을 코딩하는 제조합 핵산 분자로 이루어진 그룹으로부터 선택된 제조합 핵산 분자의 발현을 포함한다. 상기 제 2의 PKS 시스템에서 유래된 적어도 하나의 생물학적으로 활성인 도메인은 이에 제한되는 것은 아니지만, (a) 트라우스토카이트리드 미생물 유래의 다중불포화 지방산 (PUFA) 폴리케타이드 신타제 (PKS) 시스템의 도메인; (b) (i) 적어도 하나의 PUFA를 제조하는 미생물을 선택하는 단계; (ii) 발효 배지 중 약 5% 이상의 용존 산소 포화 농도에서 미생물에 의한 PUFA 생산량과 비교하였을 때 발효 배지 중 약 5% 이하의 용존 산소 포화 농도에서 증가된 PUFA를 제조하는 능력을 가진 미생물을 동정하는 단계에 의해 동정된 미생물로부터 PUFA PKS 시스템의 도메인; (c) 서열번호: 2, 서열번호: 4, 서열번호: 6, 서열번호: 8, 서열번호: 10, 서열번호: 13,

서열번호:18, 서열번호: 20, 서열번호: 22, 서열번호: 24, 서열번호: 26, 서열번호: 28, 서열번호: 30, 서열번호: 32 및 이들의 생물학적으로 활성인 단편으로 이루어진 그룹으로부터 선택된 아미노산 서열을 포함하는 도메인; (d) 상기 (c)의 아미노산 서열과 적어도 약 60% 동일한, 보다 바람직하게는 적어도 약 70% 동일한, 보다 바람직하게는 적어도 약 80% 동일한, 보다 바람직하게는 적어도 약 90% 동일하되, 다중불포화 지방산 (PUFA) 폴리케타이드 신타제 (PKS) 시스템의 적어도 하나의 도메인의 생물학적 활성을 가지는 도메인을 포함할 수 있다. 한 측면에서, 재조합 핵산 분자는 포스포판테테인 트랜스퍼라제를 코딩한다. 한 측면에서 두 번째 PKS 시스템은 박테리아의 PUFA PKS 시스템, 타입 I PKS 시스템, 타입 II PKS 시스템, 모듈라 PKS 시스템 및 비박테리아의 PUFA PKS 시스템(예, 이에 제한되는 것은 아니지만, 스킴조카이트리움 PUFA PKS 시스템을 포함한 트라우스토카이트리드 PUFA PKS 시스템과 진핵의 PUFA PKS 시스템)으로 이루어진 그룹으로부터 선택된다.

[0025]

본 발명의 또 다른 양태는 다중불포화 지방산 (PUFA) 폴리케타이드 신타제(PKS) 시스템 중 적어도 하나의 생물학적으로 활성인 도메인을 포함하는 PKS 시스템을 발현하도록 유전적으로 변형된 식물에 관한 것으로서, 상기 도메인은 (a) 서열번호: 39, 서열번호: 41, 서열번호: 43, 서열번호: 45, 서열번호: 48, 서열번호: 50, 서열번호: 52, 서열번호: 54, 서열번호: 56, 서열번호: 58, 서열번호: 60 서열번호: 62, 서열번호: 64, 서열번호: 66, 서열번호: 68 및 이들의 생물학적으로 활성인 단편으로 이루어진 그룹으로부터 선택된 아미노산 서열; (b) 서열번호: 39, 서열번호: 43, 서열번호: 50, 서열번호: 52, 서열번호: 56 및 서열번호: 58로 이루어진 그룹으로부터 선택된 아미노산 서열과 적어도 약 60% 동일한, 보다 바람직하게는 적어도 약 70% 동일한, 보다 바람직하게는 적어도 약 80% 동일한, 보다 바람직하게는 적어도 약 90% 동일하되, 다중불포화 지방산(PUFA) 폴리케타이드 신타제(PKS) 시스템 중 적어도 하나의 도메인의 생물학적 활성을 가지는 아미노산 서열; (c) 서열번호: 54의 아미노산 서열과 적어도 약 65% 동일한, 보다 바람직하게는 적어도 약 70% 동일한, 보다 바람직하게는 적어도 약 80% 동일한, 보다 바람직하게는 적어도 약 90% 동일하되, 다중불포화 지방산(PUFA) 폴리케타이드 신타제 (PKS) 시스템 중 적어도 하나의 도메인의 생물학적 활성을 가지는 아미노산 서열; (d) 서열번호: 45, 서열번호: 48, 서열번호:60, 서열번호: 62 및 서열번호: 64로 이루어진 그룹으로부터 선택된 아미노산 서열과 적어도 약 70% 동일한, 보다 바람직하게는 적어도 약 80% 동일한, 보다 바람직하게는 적어도 약 90% 동일하되, 다중불포화 지방산(PUFA) 폴리케타이드 신타제(PKS) 시스템 중 적어도 하나의 도메인의 생물학적 활성을 가지는 아미노산 서열; 및/또는 (e) 서열번호: 41, 서열번호: 66, 서열번호: 68로 이루어진 그룹으로부터 선택된 아미노산 서열과 적어도 약 80% 동일한, 보다 바람직하게는 적어도 약 90% 동일하되, 다중불포화 지방산(PUFA) 폴리케타이드 신타제(PKS) 시스템 중 적어도 하나의 도메인의 생물학적 활성을 가지는 아미노산 서열로 이루어진 그룹으로부터 선택된 아미노산 서열을 포함한다. 한 측면에서, PUFA PKS 시스템의 적어도 하나의 도메인은 서열번호: 39, 서열번호: 41, 서열번호: 43, 서열번호: 45, 서열번호: 48, 서열번호: 50, 서열번호: 52, 서열번호: 54, 서열번호: 56, 서열번호: 58, 서열번호: 60, 서열번호: 62, 서열번호: 64, 서열번호: 66 및 서열번호: 68 및 이들의 생물학적으로 활성인 단편으로 이루어진 그룹으로부터 선택된 아미노산 서열을 포함한다. 한 측면에서, 상기 식물은 박테리아의 PUFA PKS 시스템, 타입 I PKS 시스템, 타입 II PKS 시스템, 모듈라 PKS 시스템 및 비박테리아의 PUFA PKS 시스템(예, 스킴조카이트리움 PUFA PKS 시스템과 같은 트라우스토카이트리드 PUFA PKS 시스템)으로 이루어진 그룹으로부터 선택된 PKS 시스템 유래의 적어도 하나의 생물학적으로 활성인 도메인을 코딩하는 적어도 하나의 핵산 분자를 재조합에 의해 발현하도록 추가로 유전적으로 변형되었다.

[0026]

본 발명의 또 다른 양태는 생활성 분자를 제조하기에 효과적인 조건하에서 다중불포화 지방산 (PUFA) 폴리케타이드 신타제(PKS) 시스템의 적어도 하나의 생물학적으로 활성인 도메인을 포함하는 PKS 시스템을 발현하도록 유전적으로 변형된 유기체를 배양하는 단계를 포함하여 폴리케타이드 신타제 시스템에 의해 제조되는 생 활성 분자를 제조하는 방법에 관한 것으로서, 상기 PUFA PKS 시스템의 적어도 하나의 도메인은 (a) 서열번호: 39, 서열번호: 41, 서열번호: 43, 서열번호: 45, 서열번호: 48, 서열번호: 50, 서열번호: 52, 서열번호: 54, 서열번호: 56, 서열번호: 58, 서열번호: 60, 서열번호: 62, 서열번호: 64, 서열번호: 66, 서열번호: 68 및 이들의 생물학적으로 활성인 단편으로 이루어진 그룹으로부터 선택된 아미노산 서열; (b) 서열번호: 39, 서열번호: 43, 서열번호: 50, 서열번호: 52, 서열번호: 56 및 서열번호: 58로 이루어진 그룹으로부터 선택된 아미노산 서열과 적어도 약 60% 동일한, 보다 바람직하게는 적어도 약 70% 동일한, 보다 바람직하게는 적어도 약 80% 동일한, 보다 바람직하게는 적어도 약 90% 동일하되, 다중불포화 지방산(PUFA) 폴리케타이드 신타제(PKS) 시스템 중 적어도 하나의 도메인의 생물학적 활성을 가지는 아미노산 서열; (c) 서열번호: 54의 아미노산 서열과 적어도 약 65% 동일한, 보다 바람직하게는 적어도 약 70% 동일한, 보다 바람직하게는 적어도 약 80% 동일한, 보다 바람직하게는 적어도 약 90% 동일하되, 다중불포화 지방산(PUFA) 폴리케타이드 신타제(PKS) 시스템 중 적어도 하나의 도메인의 생물학적 활성을 가지는 아미노산 서열; (d) 서열번호: 45, 서열번호: 48, 서열번호:60, 서열번호: 62 및 서열번호: 64로 이루어진 그룹으로부터 선택된 아미노산 서열과 적어도 약 70% 동일한, 보다 바람직하게

는 적어도 약 80% 동일한, 보다 바람직하게는 적어도 약 90% 동일하되, 다중불포화 지방산(PUFA) 폴리케타이드 신타제(PKS) 시스템 중 적어도 하나의 도메인의 생물학적 활성을 가지는 아미노산 서열; 및/또는 (e) 서열번호: 41, 서열번호: 66, 서열번호: 68로 이루어진 그룹으로부터 선택된 아미노산 서열과 적어도 약 80% 동일한, 보다 바람직하게는 적어도 약 90% 동일하되, 다중불포화 지방산(PUFA) 폴리케타이드 신타제(PKS) 시스템 중 적어도 하나의 도메인의 생물학적 활성을 가지는 아미노산 서열로 이루어진 그룹으로부터 선택된 아미노산 서열을 포함한다.

[0027] 한 측면에서, 상기 유기체는 PUFA PKS 시스템의 적어도 하나의 도메인을 포함하는 PKS 시스템을 내인적으로 발현하고 유전적 변형은 PUFA PKS 시스템의 적어도 하나의 도메인을 코딩하는 핵산 서열에 있다. 한 측면에서, 상기 유전적 변형은 PUFA PKS 시스템이 유전적으로 변형되지 않은 유기체와 비교하였을 때 내인성의 PKS 시스템에 의해 제조되는 적어도 하나의 생산물을 변화시킨다.

[0028] 다른 측면에서, 상기 유기체는 PUFA PKS 시스템의 적어도 하나의 생물학적으로 활성인 도메인을 포함하는 PKS 시스템을 내인적으로 발현하고 유전적 변형은 제 2의 PKS 시스템에서 유래된 적어도 하나의 생물학적으로 활성인 도메인을 코딩하는 재조합 핵산 분자와 PUFA PKS 시스템의 활성에 영향을 미치는 단백질을 코딩하는 재조합 핵산 분자로 이루어진 그룹으로부터 선택된 재조합 핵산 분자에 의한 유기체의 형질감염을 포함한다. 한 측면에서 유전적 변형은 PUFA 제조에 영향을 미치도록 유전적으로 변형되지 않은 유기체와 비교하였을 때 내인적으로 PKS 시스템에 의해 제조되는 적어도 하나의 생산물을 변화시킨다.

[0029] 다른 측면에서, 상기 유기체는 다중불포화 지방산 (PUFA) 폴리케타이드 신타제 (PKS)의 적어도 하나의 도메인을 코딩하는 재조합 핵산 분자로 형질전환됨으로써 유전적으로 변형된다.

[0030] 다른 측면에서, 상기 유기체는 유전적 변형이 없는 자연적으로 존재하는 유기체와는 상이한 다중불포화 지방산 (PUFA) 프로필을 제조한다.

[0031] 다른 측면에서, 상기 유기체는 비박테리아의 PUFA PKS 시스템을 내인적으로 발현하고, 유전적 변형은 비박테리아의 PUFA PKS 시스템 중 적어도 하나의 도메인을 코딩하는 핵산 서열을 다른 PKS 시스템 유래의 도메인으로 치환하는 것을 포함한다.

[0032] 또 다른 측면에서, 상기 유기체는 PUFA PKS 시스템에 의해 제조되는 지방산의쇄 길이를 조절하는 단백질을 코딩하는 재조합 핵산 분자로 유기체를 형질감염시킴으로써 변형된 비박테리아의 PUFA PKS 시스템을 내인적으로 발현한다.

[0033] 다른 측면에서, 상기 생활성 분자는 항염증 제형, 화학 요법제, 활성 부형제, 골다공증 약, 항우울제, 항경련제, 항-헬리코박터 파이로리 약, 신경퇴행성 질환의 치료를 위한 약, 퇴행성 간 질환의 치료를 위한 약, 항생제 및/또는 콜레스테롤 저하 제형으로부터 선택된다. 한 측면에서, 생활성 분자는 항생제이다. 다른 측면에서, 생활성 분자는 다중불포화 지방산(PUFA)이다. 또 다른 측면에서, 생활성 분자는 시스-배형의 탄소-탄소 이중 결합을 포함하는 분자이다. 한 측면에서 상기 생활성 분자는 세 번째 탄소마다 이중 결합을 포함하는 분자이다. 한 측면에서, 유기체는 미생물이다. 또 다른 측면에서, 유기체는 식물이다.

[0034] 본 발명의 다른 양태는 PUFA PKS 시스템의 적어도 하나의 생물학적으로 활성인 도메인을 코딩하는 핵산 서열을 포함하는 적어도 하나의 재조합 핵산 분자를 포함하는 PKS 시스템을 발현하도록 식물의 세포를 유전적으로 변형시키는 단계를 포함하여 천연적으로 존재하는 식물과는 상이한 다중불포화 지방산(PUFA) 프로필을 가진 식물을 제조하는 방법에 관한 것으로, 상기 PUFA PKS 시스템의 적어도 하나의 도메인은 (a) 서열번호: 39, 서열번호: 41, 서열번호: 43, 서열번호: 45, 서열번호: 48, 서열번호: 50, 서열번호: 52, 서열번호: 54, 서열번호: 56, 서열번호: 58, 서열번호: 60 서열번호: 62, 서열번호: 64, 서열번호: 66, 서열번호: 68 및 이들의 생물학적으로 활성인 단편으로 이루어진 그룹으로부터 선택된 아미노산 서열; (b) 서열번호: 39, 서열번호: 43, 서열번호: 50, 서열번호: 52, 서열번호: 56 및 서열번호: 58로 이루어진 그룹으로부터 선택된 아미노산 서열과 적어도 약 60% 동일한, 보다 바람직하게는 적어도 약 70% 동일한, 보다 바람직하게는 적어도 약 80% 동일한, 보다 바람직하게는 적어도 약 90% 동일하되, 다중불포화 지방산(PUFA) 폴리케타이드 신타제(PKS) 시스템 중 적어도 하나의 도메인의 생물학적 활성을 가지는 아미노산 서열; (c) 서열번호: 54의 아미노산 서열과 적어도 약 65% 동일한, 보다 바람직하게는 적어도 약 70% 동일한, 보다 바람직하게는 적어도 약 80% 동일한, 보다 바람직하게는 적어도 약 90% 동일하되, 다중불포화 지방산(PUFA) 폴리케타이드 신타제(PKS) 시스템 중 적어도 하나의 도메인의 생물학적 활성을 가지는 아미노산 서열; (d) 서열번호: 45, 서열번호: 48, 서열번호: 60, 서열번호: 62 및 서열번호: 64로 이루어진 그룹으로부터 선택된 아미노산 서열과 적어도 약 70% 동일한, 보다 바람직하게는 적어도

약 80% 동일한, 보다 바람직하게는 적어도 약 90% 동일하되, 다중불포화 지방산(PUFA) 폴리케타이드 신타제(PKS) 시스템 중 적어도 하나의 도메인의 생물학적 활성을 가지는 아미노산 서열; 및 (e) 서열번호: 41, 서열번호: 66, 서열번호: 68로 이루어진 그룹으로부터 선택된 아미노산 서열과 적어도 약 80% 동일한, 보다 바람직하게는 적어도 약 90% 동일하되, 다중불포화 지방산(PUFA) 폴리케타이드 신타제(PKS) 시스템 중 적어도 하나의 도메인의 생물학적 활성을 가지는 아미노산 서열로 이루어진 그룹으로부터 선택된 아미노산 서열을 포함한다.

[0035]

본 발명의 다른 양태는 PUFA PKS 시스템의 적어도 하나의 생물학적으로 활성인 도메인을 코딩하는 핵산 서열을 포함하는 적어도 하나의 제조합 핵산 분자를 발현하는 제조합 숙주 세포에 의해 제조된 오일을 최종 산물에 첨가하는 단계를 포함하여, 적어도 하나의 지방산을 함유하는 최종 산물을 변형시키는 방법에 관한 것으로, 상기 PUFA PKS 시스템의 적어도 하나의 도메인은 (a) 서열번호: 39, 서열번호: 41, 서열번호: 43, 서열번호: 45, 서열번호: 48, 서열번호: 50, 서열번호: 52, 서열번호: 54, 서열번호: 56, 서열번호: 58, 서열번호: 60 서열번호: 62, 서열번호: 64, 서열번호: 66, 서열번호: 68 및 이들의 생물학적으로 활성인 단편으로 이루어진 그룹으로부터 선택된 아미노산 서열; (b) 서열번호: 39, 서열번호: 43, 서열번호: 50, 서열번호: 52, 서열번호: 56 및 서열번호: 58로 이루어진 그룹으로부터 선택된 아미노산 서열과 적어도 약 60% 동일한, 보다 바람직하게는 적어도 약 70% 동일한, 보다 바람직하게는 적어도 약 80% 동일한, 보다 바람직하게는 적어도 약 90% 동일하되, 다중불포화 지방산(PUFA) 폴리케타이드 신타제(PKS) 시스템 중 적어도 하나의 도메인의 생물학적 활성을 가지는 아미노산 서열; (c) 서열번호: 54의 아미노산 서열과 적어도 약 65% 동일한, 보다 바람직하게는 적어도 약 70% 동일한, 보다 바람직하게는 적어도 약 80% 동일한, 보다 바람직하게는 적어도 약 90% 동일하되, 다중불포화 지방산(PUFA) 폴리케타이드 신타제(PKS) 시스템 중 적어도 하나의 도메인의 생물학적 활성을 가지는 아미노산 서열; (d) 서열번호: 45, 서열번호: 48, 서열번호: 60, 서열번호: 62 및 서열번호: 64로 이루어진 그룹으로부터 선택된 아미노산 서열과 적어도 약 70% 동일한, 보다 바람직하게는 적어도 약 80% 동일한, 보다 바람직하게는 적어도 약 90% 동일하되, 다중불포화 지방산(PUFA) 폴리케타이드 신타제(PKS) 시스템 중 적어도 하나의 도메인의 생물학적 활성을 가지는 아미노산 서열; 및 (e) 서열번호: 41, 서열번호: 66, 서열번호: 68로 이루어진 그룹으로부터 선택된 아미노산 서열과 적어도 약 80% 동일한, 보다 바람직하게는 적어도 약 90% 동일하되, 다중불포화 지방산(PUFA) 폴리케타이드 신타제(PKS) 시스템 중 적어도 하나의 도메인의 생물학적 활성을 가지는 아미노산 서열로 이루어진 그룹으로부터 선택된 아미노산 서열을 포함한다. 한 측면에서, 상기 최종 산물은 식이 보조제, 식료품, 약제학적 제형, 인간화된 동물의 밀크 및 분유로부터 선택된다. 한 측면에서, 상기 약제학적 제형은 항염증 제형, 화학 요법제, 활성 부형제, 골다공증 약, 항우울제, 항경련제, 항-헬리코박터 파이로리 약, 신경퇴행성 질환의 치료를 위한 약, 퇴행성 간 질환의 치료를 위한 약, 항생제, 콜레스테롤 저해 제형으로부터 선택된다. 한 측면에서, 최종 산물은 만성 염증, 급성 염증, 위장관계 질환, 암, 카캐시아(cachexia), 심장의 재협착증, 신경퇴행성 질환, 간의 퇴행성 질환, 혈액 리피드 질환, 골다공증, 동맥경화, 자가면역 질환, 자간전증, 조산, 노화와 연관된 황반변증, 호흡기계 질환 및 퍼옥시소말(peroxisomal) 질환으로부터 선택된 용태를 치료하는데 이용된다.

[0036]

본 발명의 또 다른 양태는 PUFA PKS 시스템의 적어도 하나의 생물학적으로 활성인 도메인을 코딩하는 핵산 서열을 포함하는 적어도 하나의 핵산 분자로 밀크를 제조하는 동물의 밀크를 제조하는 세포를 유전적으로 변형시키는 단계를 포함하여, 인간화된 동물 밀크를 제조하는 방법에 관한 것으로 상기 PUFA PKS 시스템의 적어도 하나의 도메인은 (a) 서열번호: 39, 서열번호: 41, 서열번호: 43, 서열번호: 45, 서열번호: 48, 서열번호: 50, 서열번호: 52, 서열번호: 54, 서열번호: 56, 서열번호: 58, 서열번호: 60 서열번호: 62, 서열번호: 64, 서열번호: 66, 서열번호: 68 및 이들의 생물학적으로 활성인 단편으로 이루어진 그룹으로부터 선택된 아미노산 서열; (b) 서열번호: 39, 서열번호: 43, 서열번호: 50, 서열번호: 52, 서열번호: 56 및 서열번호: 58로 이루어진 그룹으로부터 선택된 아미노산 서열과 적어도 약 60% 동일한, 보다 바람직하게는 적어도 약 70% 동일한, 보다 바람직하게는 적어도 약 80% 동일한, 보다 바람직하게는 적어도 약 90% 동일하되, 다중불포화 지방산(PUFA) 폴리케타이드 신타제(PKS) 시스템 중 적어도 하나의 도메인의 생물학적 활성을 가지는 아미노산 서열; (c) 서열번호: 54의 아미노산 서열과 적어도 약 65% 동일한, 보다 바람직하게는 적어도 약 70% 동일한, 보다 바람직하게는 적어도 약 80% 동일한, 보다 바람직하게는 적어도 약 90% 동일하되, 다중불포화 지방산(PUFA) 폴리케타이드 신타제(PKS) 시스템 중 적어도 하나의 도메인의 생물학적 활성을 가지는 아미노산 서열; (d) 서열번호: 45, 서열번호: 48, 서열번호: 60, 서열번호: 62 및 서열번호: 64로 이루어진 그룹으로부터 선택된 아미노산 서열과 적어도 약 70% 동일한, 보다 바람직하게는 적어도 약 80% 동일한, 보다 바람직하게는 적어도 약 90% 동일하되, 다중불포화 지방산(PUFA) 폴리케타이드 신타제(PKS) 시스템 중 적어도 하나의 도메인의 생물학적 활성을 가지는 아미노산 서열; 및 (e) 서열번호: 41, 서열번호: 66, 서열번호: 68로 이루어진 그룹으로부터 선택된 아미노산 서열과 적어도 약 80% 동일한, 보다 바람직하게는 적어도 약 90% 동일하되, 다중불포화 지방산(PUFA) 폴리케타이드

드 신타제(PKS) 시스템 중 적어도 하나의 도메인의 생물학적 활성을 가지는 아미노산 서열로 이루어진 그룹으로부터 선택된 아미노산 서열을 포함한다.

[0037] 본 발명의 다른 양태는 내인적으로 다중불포화 지방산 (PUFA) 폴리케타이드 신타제 (PKS) 시스템을 발현하고, 내인성의 PUFA PKS 시스템은 유전적으로 변형되지 않은 트라우스토카이트리드 미생물과 비교했을 때 트라우스카이트리드 미생물에 의한 다중불포화 지방산 (PUFA)의 발현 프로필을 변형시키도록 유전적으로 변형시킨 트라우스토카이트리드 미생물에 관한 것이다.

[0038] 한 측면에서, 상기 내인성의 PUFA PKS 시스템은 내인성의 PUFA PKS 시스템의 적어도 하나의 도메인을 코딩하는 핵산 서열의 돌연변이에 의해 변형되었다. 한 측면에서, 상기 변형은 표적화된 돌연변이에 의해 제조된다. 다른 측면에서, 상기 변형은 전통적인 돌연변이 및 스크리닝에 의해 제조된다.

[0039] 다른 측면에서, 내인성의 PUFA PKS 시스템은 트라우스토카이트리드 미생물의 PUFA 제조 프로필을 변이시키기 위하여 내인성의 PUFA PKS 시스템의 적어도 하나의 도메인을 코딩하는 적어도 하나의 핵산 서열을 결실하고 상기 내인성 도메인의 상동체를 코딩하는 핵산 서열을 삽입함으로써 유전적으로 변형시킨다. 여기서, 상기 상동체는 PKS 시스템의 적어도 하나의 도메인의 생물학적 활성을 가진다. 한 측면에서, 내인성 도메인의 상동체는 내인성 도메인과 비교하였을 때 미생물에 의한 PUFA 제조의 변이를 초래하는 적어도 하나의 결실, 삽입 또는 치환으로 이루어진 그룹으로부터 선택된 변이를 포함한다. 다른 측면에서, 상기 상동체의 아미노산 서열은 내인성 도메인의 아미노산 서열의 적어도 약 60% 동일한, 보다 바람직하게는 적어도 약 70% 동일한, 보다 바람직하게는 적어도 약 80% 동일한, 보다 바람직하게는 적어도 약 90% 동일하다. 한 측면에서, 내인성 도메인의 상동체는 다른 트라우스토카이트리드 미생물의 PUFA PKS 시스템에서 유래된 도메인이다.

[0040] 다른 측면에서, 내인성의 PUFA PKS 시스템은 내인성의 PUFA PKS 시스템의 적어도 하나의 도메인을 코딩하는 적어도 하나의 핵산 서열을 결실하고 다른 미생물 유래 PKS 시스템의 적어도 하나의 도메인을 코딩하는 핵산 서열을 삽입함으로써 변형되었다. 한 측면에서, 다른 미생물 유래 PKS 시스템의 적어도 하나의 도메인을 코딩하는 핵산 서열은 박테리아의 PUFA PKS 시스템으로부터 유래된다. 예를 들어 다른 미생물은 약 25°C 이상의 온도에서 천연적으로 PUFA를 제조하는 PUFA PKS 시스템을 가진 해양의 박테리아일 수 있다. 한 측면에서, 상기 해양의 박테리아는 셰와넬라 올레야나 및 셰와넬라 자포니카로 이루어진 그룹으로부터 선택된다. 한 측면에서, 다른 미생물 유래의 PKS 시스템의 도메인은 타입 I PKS 시스템, 타입 II PKS 시스템, 모듈라 PKS 시스템 및 다른 트라우스토카이트리드 미생물 유래의 PUFA PKS 시스템으로 이루어진 그룹으로부터 선택된 PKS 시스템 유래이다.

[0041] 상기 측면 중 하나에서, 내인성의 PUFA PKS 시스템의 도메인은 이에 제한되는 것은 아니지만, 말로닐-CoA: ACP 아실트랜스퍼라제(MAT), β -케토 아실-ACP 신타제(KS), 케토환원효소(KR), 아실트랜스퍼라제(AT), FabA-유사 β -하이드록시 아실 ACP 탈수소효소(DH), 포스포판테테인 트랜스퍼라제, 쉐갈이 인자(CLF), 아실 캐리어 단백질(ACP), 엔오일 ACP-환원효소(ER), 트랜스-2-아실 ACP의 합성을 촉매하는 효소, 트랜스-2-아실 ACP의 시스-3-아실 ACP로의 가역적인 이성질화를 촉매하는 효소, 시스-3-아실-ACP의 시스-5- β -케토-아실 ACP로의 연장을 촉매하는 효소 중 어느 하나의 생물학적 활성을 가진 도메인을 포함할 수 있다. 상기 측면 중 어느 하나에서, 내인성의 PUFA PKS 시스템의 도메인은 (a) 서열번호: 2, 서열번호: 4, 서열번호: 6, 서열번호: 8, 서열번호: 10, 서열번호: 13, 서열번호: 18, 서열번호: 20, 서열번호: 22, 서열번호: 24, 서열번호: 26, 서열번호: 28, 서열번호: 30, 서열번호: 32, 서열번호: 39, 서열번호: 41, 서열번호: 43, 서열번호: 45, 서열번호: 48, 서열번호: 50, 서열번호: 52, 서열번호: 54, 서열번호: 56, 서열번호: 58, 서열번호: 60, 서열번호: 62, 서열번호: 64, 서열번호: 66, 서열번호: 68 및 이의 생물학적으로 활성인 단편으로 이루어진 그룹으로부터 선택된 아미노산 서열 및 (b) (a)의 아미노산 서열과 적어도 약 60% 동일한, 보다 바람직하게는 적어도 약 70% 동일한, 보다 바람직하게는 적어도 약 80% 동일한, 보다 바람직하게는 적어도 약 90% 동일하되, 다중불포화 지방산 (PUFA) 폴리케타이드 신타제(PKS) 시스템의 적어도 하나의 도메인의 생물학적 활성을 가진 아미노산 서열을 포함할 수 있다.

[0042] 한 측면에서, PUFA 제조 프로필은 미생물에 의한 에이코사펜타엔산(EPA)의 제조를 개시, 증가 또는 감소시키도록 변이된다. 다른 측면에서 PUFA 제조 프로필은 미생물에 의한 도코사헥사엔산(DHA)의 제조를 개시, 증가 또는 감소시키도록 변이된다. 다른 측면에서, PUFA 제조 프로필은 미생물에 의한 도코사펜타엔산(DPA)의 하나 또는 두 개의 이성질체의 제조를 개시, 증가 또는 감소시키도록 변이된다. 다른 측면에서, PUFA 제조 프로필은 미생물에 의한 아라키돈산의 제조를 개시, 증가 또는 감소시키도록 변이된다. 다른 측면에서, 상기 트라우스토카이트리드는 스키노카이트리움, 트라우스토카이트리움 및 자포노카이트리움으로 이루어진 그룹으로부터

선택된 속에서 유래된다. 다른 측면에서, 트라우스토카이트리드는 스킨조카이트리움 속에서 유래된다. 다른 측면에서, 트라우스토카이트리드는 스킨조카이트리움 아그레가툼, 스킨조카이트리움 리마시눔 및 스킨조카이트리움 미누툼으로 이루어진 그룹으로부터 선택된 스킨조카이트리움 중에서 유래된다. 다른 측면에서, 트라우스토카이트리드는 트라우스토카이트리움 중에서 유래된다.

[0043]

본 발명의 다른 양태는 에이코사펜타엔산(EPA)을 제조하는 유전적으로 변형된 스킨조카이트리움에 관한 것으로 상기 스킨조카이트리움은 스킨조카이트리움에 의한 EPA의 제조를 초래하는 내인성의 PUFA PKS 시스템의 적어도 하나의 도메인을 코딩하는 적어도 하나의 핵산 서열에서 유전적 변형을 포함하는 내인성의 다중불포화 지방산(PUFA) 폴리케타이드 신타제 (PKS) 시스템을 가진다. 한 측면에서, 상기 스킨조카이트리움은 말로닐-CoA: ACP 아실트랜스퍼라제(MAT), β -케토 아실-ACP 신타제(KS), 케토환원효소(KR), 아실트랜스퍼라제(AT), FabA-유사 β -하이드록시 아실 ACP 탈수소효소(DH), 포스포판테테인 트랜스퍼라제, 쇠길인자(CLF), 아실 캐리어 단백질(ACP), 엔오일 ACP-환원효소(ER), 트랜스-2-아실 ACP의 합성을 촉매하는 효소, 트랜스-2-아실 ACP의 시스-3-아실 ACP로의 가역적인 이성질화를 촉매하는 효소, 시스-3-아실-ACP의 시스-5- β -케토-아실 ACP로의 연장을 촉매하는 효소 중 어느 하나의 생물학적 활성을 가진 도메인을 포함할 수 있다. 한 측면에서, 스킨조카이트리움은 내인성의 PUFA PKS 시스템의 서열번호 : 2를 코딩하는 오픈 리딩 프레임에서 유래된 적어도 하나의 도메인을 코딩하는 적어도 하나의 핵산 서열에서 유전적 변형을 포함한다. 한 측면에서, 스킨조카이트리움은 내인성의 PUFA PKS 시스템의 서열번호 : 4를 코딩하는 오픈 리딩 프레임에서 유래된 적어도 하나의 도메인을 코딩하는 적어도 하나의 핵산 서열에서 유전적 변형을 포함한다. 한 측면에서, 스킨조카이트리움은 내인성의 PUFA PKS 시스템의 서열번호 : 6을 코딩하는 오픈 리딩 프레임에서 유래된 적어도 하나의 도메인을 코딩하는 적어도 하나의 핵산 서열에서 유전적 변형을 포함한다. 한 측면에서, 스킨조카이트리움은 β -케토 아실-ACP 신타제 (KS), FabA-유사 β -하이드록시 아실 ACP 탈수소효소(DH), 포스포판테테인 트랜스퍼라제, 쇠길인자(CLF), 트랜스-2-아실 ACP의 합성을 촉매하는 효소, 트랜스-2-아실 ACP의 시스-3-아실 ACP로의 가역적인 이성질화를 촉매하는 효소, 시스-3-아실-ACP의 시스-5- β -케토-아실 ACP로의 연장을 촉매하는 효소 중 어느 하나의 생물학적 활성을 가진 적어도 하나의 도메인을 코딩하는 적어도 하나의 핵산 서열에서 유전적 변형을 포함한다. 한 측면에서, 스킨조카이트리움은 내인성 PUFA PKS 시스템의 서열번호 : 20, 서열번호 : 22, 서열번호 : 28 및 서열번호 : 30으로 이루어진 그룹으로부터 선택된 아미노산 서열을 코딩하는 적어도 하나의 핵산 서열에서 유전적 변형을 포함한다. 한 측면에서, 스킨조카이트리움은 내인성의 PUFA PKS 시스템의 적어도 하나의 도메인을 코딩하는 적어도 하나의 핵산 서열을 결실하고 비스킨조카이트리움 미생물 유래의 PKS 시스템의 적어도 하나의 도메인을 코딩하는 핵산 서열을 삽입함으로써 변형된다. 한 측면에서, 비스킨조카이트리움 미생물은 적어도 약 15°C, 보다 바람직하게는 적어도 약 20°C, 보다 바람직하게는 적어도 약 25°C, 보다 바람직하게는 적어도 약 30°C, 보다 바람직하게는 약 20°C 내지 약 40°C의 온도에서 성장하고 PUFA를 제조한다. 한 측면에서 비스킨조카이트리움 미생물 유래의 PKS 시스템의 적어도 하나의 도메인을 코딩하는 핵산 서열은 박테리아의 PUFA PKS 시스템 유래이다. 한 측면에서, 상기 박테리아의 PUFA PKS 시스템은 *쉐와넬라 올레야나* 및 *쉐와넬라 자포니카*로 이루어진 그룹으로부터 선택된 박테리아이다. 다른 측면에서, PKS 시스템의 적어도 하나의 도메인을 코딩하는 핵산 서열은 타입 I PKS 시스템, 타입 II PKS 시스템, 모듈라 PKS 시스템 및 비박테리아의 PUFA PKS 시스템으로 이루어진 그룹으로부터 선택된다(예를 들면, 트라우스토카이트리드 PUFA PKS 시스템과 같은 진핵의 PUFA PKS 시스템).

[0044]

본 발명의 다른 양태는 유전적으로 변형되지 않은 스킨조카이트리움과 비교하였을 때 증가된 양의 도코사헥사엔산(DHA)을 제조하는 유전적으로 변형된 스킨조카이트리움에 관한 것으로 상기 스킨조카이트리움은 스킨조카이트리움에 의한 DHA의 증가된 제조를 초래하는 내인성의 PUFA PKS 시스템의 적어도 하나의 도메인을 코딩하는 적어도 하나의 핵산 서열에서 유전적 변형을 포함하는 내인성의 다중불포화 지방산(PUFA) 폴리케타이드 신타제 (PKS) 시스템을 가진다. 한 측면에서, 내인성의 PUFA PKS 시스템의 적어도 하나의 도메인은 트라우스토카이트리움 유래의 PUFA PKS 시스템의 적어도 하나의 도메인 대신에 치환됨으로써 변형되었다. 한 측면에서, 스킨조카이트리움에 의해 제조된 DHA 대 DPA의 비율은 유전적으로 변형되지 않은 스킨조카이트리움과 비교하였을 때 증가된다.

[0045]

본 발명의 다른 양태는 리피드를 제조하기에 효과적인 조건하에서 상술한 유전적으로 변형된 트라우스토카이트리드 미생물 또는 상술한 유전적으로 변형된 스킨조카이트리움을 배양하는 단계를 포함하여 적어도 하나의 선택된 다중불포화 지방산(PUFA)으로 풍부한 리피드를 제조하는 방법에 관한 것이다. 한 측면에서, 선택된 PUFA는 에이코사펜타엔산(EPA)이다.

[0046]

본 발명의 또 다른 측면에서, EPA가 풍부한 리피드를 제조하기에 효과적인 조건하에서 내인성 다중불포

화 지방산(PUFA) 폴리케타이드 신타제(PKS) 시스템을 가지고, 유전적으로 변형되지 않은 미생물과 비교하였을 때 내인성의 PUFA PKS 시스템이 미생물의 리피드에서 EPA의 제조를 개시 또는 증가시키도록 적어도 하나의 도메인에서 유전적으로 변형된 유전적으로 변형된 트라우스토카이트리드 미생물을 배양하는 단계를 포함하여 에이코사펜타엔산(EPA)이 풍부한 리피드를 제조하는 방법에 관한 것이다.

[0047]

본 발명은 일반적으로 다중불포화 지방산(PUFA) 폴리케타이드 신타제(PKS) 시스템, 그러한 PUFA PKS 시스템을 포함하는 유전적으로 변형된 유기체, 생물질, 특히 DHA, DPA 및 EPA와 같은 PUFA를 포함한 생산물의 제조를 위하여 그러한 시스템을 제조하고 이용하는 방법에 관한 것이다. 본 발명에서 PUFA PKS 시스템은 일반적으로 하기 동정 특징을 가지고 있다: (a) 시스템의 천연 물질로서 PUFA를 제조한다; (2) 상기 시스템은 지방산 쉼의 반복적인 과정 뿐만 아니라 선택된 사이클에서 트랜스-시스 이성질화 및 엔오일 환원 반응을 포함한 비반복적 과정 모두를 수행하는 복합체로 어셈블리되는 여러 개의 다기능성 단백질을 포함한다(예를 들면 도 1 참조). PUFA PKS 시스템은 유기체에서 PUFA를 제조하기 위하여 복합적으로 작용하는 모든 유전자 및 이들의 코딩된 생산물을 집합적으로 의미한다. 따라서 PUFA PKS 시스템은 이의 천연 생산물이 PUFA인 PKS 시스템을 특히 의미한다.

[0048]

보다 자세하게, 첫 번째, 본 발명의 기본을 형성하는 PUFA PKS 시스템은 생산물로서 다중불포화 지방산(PUFA)을 제조한다 (즉, 내인적으로 (천연적으로) PKS 시스템을 이용하여 PUFA를 제조하는 그러한 시스템을 함유하는 유기체) 본 발명에서 PUFA는 적어도 16개의 탄소, 보다 바람직하게는 적어도 18개의 탄소, 보다 바람직하게는 적어도 20개의 탄소, 보다 바람직하게는 22개 이상의 탄소를 가지고, 적어도 3개 이상의 이중 결합, 바람직하게는 4개 이상, 보다 바람직하게는 5개 이상, 보다 더 바람직하게는 6개 이상의 이중 결합을 가지되, 상기 모든 이중 결합은 시스 배형(configuration)인 다중불포화지방산을 의미한다. 본 발명의 목적은 유전적 변형 또는 최종 생산물의 조작을 통하여 목적인 쉼의 길이와 목적인 이중 결합수를 가진 다중불포화지방산을 제조하는 PKS 시스템을 발견하거나 제조하는 것이다. PUFA의 예로는 이에 제한되는 것은 아니지만, DHA(도코사헥사엔산 (C22:6, ω -3)), ARA(에이코사테트라엔산 또는 아라키돈산 (C20:4, n-6)), DPA (도코사펜타엔산 (C22:5, ω -6 또는 ω -3) 및 EPA (에이코사펜타엔산 (C20:5, ω -3))를 포함한다.

[0049]

두 번째, 본 발명의 PUFA PKS 시스템은 반복적이고 비반복적인 반응 모두를 통합하고 있어 상기 시스템은 기존에 개시된 PKS 시스템(예, 타입 I, 타입 II 또는 모듈라)과는 차별화된다. 보다 상세하게 본 발명의 PUFA PKS 시스템은 각 사이클 동안 기능을 발휘하는 도메인 뿐만 아니라 일부 사이클 동안만 기능을 발휘하는 도메인을 모두 포함한다. 이러한 기능성의 중요 측면은 박테리아의 Fab-A 효소와 상동성을 보이는 도메인과 관련될 수 있다. 예를 들어, 대장균의 Fab-A 효소는 두 개의 효소 활성을 보유하는 것으로 보여진다. 이는 하이드록시기를 포함하는 탄소쇄로부터 물 분자(H₂O)를 추출하고 탄소 쉼에 트랜스 이중 결합을 남겨두는 탈수 활성을 보유한다. 또한, 트랜스 이중 결합을 시스 배형으로 전환하는 이소머라제 활성을 가지고 있다. 이러한 이성질화는 근접한 탄소 원소로 이중 결합을 이동시키는 것과 접합하여 달성된다. PKS (및 FAS) 시스템에서 주 탄소 쉼의 개수는 2개 탄소의 증가분으로 연장된다. 따라서 이러한 PKS 시스템의 PUFA 생산물을 제조하기 위하여 요구되는 연장 반응 회수가 예측된다. 예를 들어, DHA(C22:6, 모두 시스)를 제조하기 위하여 10번의 연장 반응이 요구된다. 최종 산물에 단지 6개의 이중 결합이 있기 때문에 일부 반응 사이클 동안에 이중 결합이 (시스 이성질체로서) 유지되며, 나머지에서 이중 결합은 다음 연장 이전에 환원된다.

[0050]

해양 박테리아(미국특허 제6,140,486호 참조)에서의 PUFA PKS 시스템의 발견 이전에 PKS 시스템은 이러한 반복적이고 선택적인 효소 반응의 조합을 보유하지 않는 것으로 알려졌고 시스 배형의 탄소-탄소 이중 결합을 제조할 수 없는 것으로 여겨졌다. 그러나 본 발명의 PUFA PKS 시스템은 시스 이중 결합을 도입하는 능력과 사이클에 있는 반응 순서를 다양화하는 능력을 가지고 있다.

[0051]

본 발명자들은 이전에 개시된(타입 II, 타입 I 및 모듈라) PKS 시스템에서 의해 제조될 수 없는 범위의 생활성 분자를 제조하기 위하여 PUFA PKS 시스템의 이러한 특징을 이용할 것을 제안한다. 이러한 생활성 분자는 이에 제한되는 것은 아니지만, 다중불포화지방산(PUFA), 항생제 또는 다른 생활성 화합물이며, 이들 다수는 하기에서 설명된다. 예를 들어, 본 발명에서 개시된 PUFA PKS 유전자 구조에 대한 지식을 이용하여 다수의 방법 중 어느 하나가 PUFA PKS 유전자를 변이시키거나 다른 PKS 시스템을 포함한 다른 합성 시스템과 이러한 유전자의 일부를 조합하여 사용해서 새로운 생산물을 제조할 수 있다. 반복적이고 선택적인 반응을 모두 수행하는 이러한 특별한 타입의 시스템의 타고난 능력은 상기 시스템으로 하여금 유사한 방법이 PKS 시스템의 다른 타입에 적용된다면 발견될 수 없는 생산물을 제조 가능하게 할 것이다.

[0052]

진핵의 트라우스토카이트리드인 스킨조카이트리움의 PUFA 합성을 위한 PKS 시스템 구조의 대부분이 미국특허 제6,566,583호에 상세히 기술되어 있다. 본 발명자에 의한 스킨조카이트리움의 cDNA 및 계놈 클론의 완전한 서열분석은 OrfA, OrfB 및 OrfC 각각의 전장 계놈 서열의 동정과 웨와넬라에 있는 도메인과 상동성을 가진 스킨조카이트리움 Orf에 있는 특정한 도메인의 완전한 동정을 가능하게 하였다(도 2와 미국특허출원 제 10/124,800호 참조). 미국특허출원 제10/124,800호에서, 발명자들은 트라우스토카이트리움 종이 PUFA PKS 시스템을 가진 기준을 만족시킨다는 것을 동정하였고 서던 블롯 분석에 의해 이러한 유기체가 스킨조카이트리움 PUFA PKS 유전자와 상동성을 가진 유전자를 보유하는 것을 증명하였다. 그러나 상기 유전자의 분리와 구조의 결정 및 상기 유전자의 도메인 조직화는 미국특허출원 제10/124,800호에 개시되어 있지 않다. 본 발명에서 발명자들은 트라우스토카이트리움 속의 상기 트라우스토카이트리드(특히 트라우스토카이트리움 sp.23B(ATCC 20892))에 있는 상동성의 오픈 리딩 프레임(Orf)의 전장 계놈 서열을 클로닝하고 분석하였으며, 상기 트라우스토카이트리움에 있는 PUFA PKS 시스템을 포함한 도메인을 동정하였다. 따라서 본 발명은 상업적 이용을 위하여 유용성을 가지는 추가의 PUFA PKS 시스템을 제공한다 상술한 문제점을 해결할 수 있다. 트라우스토카이트리움 PUFA PKS 시스템은 하기에서 상세히 기술된다.

[0053]

또한, 본 발명은 본 발명자들에 의한 스킨조카이트리움과 트라우스토카이트리움과 같은 트라우스토카이트리알레 목의 멤버를 포함하여 진핵생물에서 PUFA를 제조하는 폴리케타이드 신타제 유사 시스템의 조작에 의해 PUFA가 풍부한 리피드(멤브레인 관련 포스포리피드(PL) 뿐만 아니라 트리아실글리세롤(TAG))을 효과적으로 제조하는 방법과 유전적으로 변형된 미생물의 발달에 의해 EPA와 같은 상업적으로 가치있는 목적인 PUFA가 풍부한 리피드의 제조를 위한 상술한 문제점을 해결한다. 특히 본 발명자들은 PUFA, 일차적으로 도코사헥사엔산(DHA; C22:6, n-3) 및 도코사펜타엔산(DPA; C22:5, n-6)가 풍부한 오일의 상업적인 제조를 위하여 앞서 최적화되었고, 여기서 유기체의 오일 제조 특성을 희생하지 않으면서 EPA(C20:5, n-3) 제조(또는 다른 PUFA 제조)가 DHA 제조를 대체하도록 유전적으로 변형된 스킨조카이트리움의 균주를 예시로서 기술한다. 또한, 본 발명자들은 특히, PUFA PKS 시스템의 변형을 통하여 미생물에 의해 제조되는 DHA 대 DPA의 비율을 변이시킴으로써 스킨조카이트리움 유기체에 의한 DHA 제조를 증진시키기 위하여 트라우스토카이트리움 유래의 PUFA PKS 유전자로 스킨조카이트리움의 유전적 변형을 기술하고 있다. 이는 단지 본 발명에 포함되는 기술의 일부 예에 지나지 않으며, 본 발명의 사상은 하기에서 설명되는 다른 제조 유기체 및 다른 목적인 PUFA에 대하여 용이하게 적용될 수 있다.

[0054]

한 양태에서 본 발명에 따른 PUFA PKS 시스템은 적어도 하나의 하기 생물학적으로 활성인 도메인을 포함한다:(a) 적어도 두 개의 엔오일-ACP 환원효소(ER) 도메인; (b) 적어도 여섯 개의 아실 캐리어 단백질(ACP) 도메인; (c) 적어도 두 개의 β -케토아실-ACP 신타제(KS) 도메인; (d) 적어도 하나의 아실트랜스퍼라제(AT) 도메인; (e) 적어도 하나의 β -케토아실-ACP 환원효소(KR) 도메인; (f) 적어도 두개의 FabA 유사 β -하이드록시아실-ACP 탈수소효소(DH) 도메인; (g) 적어도 하나의 체일인자(chain length factor, CLF) 도메인 및 (h) 적어도 하나의 말로닐-CoA:ACP 아실트랜스퍼라제(MAT) 도메인. 이러한 도메인의 기능은 당업계에 공지되어 있고 본 발명의 PUFA PKS 시스템과 관련하여 하기에서 상세하게 기술될 것이다.

[0055]

다른 양태에서, PUFA PKS 시스템은 적어도 하나의 하기 생물학적 활성인 도메인을 포함한다: (a) 적어도 하나의 엔오일-ACP 환원효소(ER) 도메인; (b) 다수의 아실 캐리어 단백질(ACP) 도메인(적어도 한 개 내지 네 개, 바람직하게는 적어도 다섯 개, 보다 바람직하게는 적어도 여섯 개, 보다 더 바람직하게는 일곱 개, 여덟 개, 아홉 개 이상); (c) 적어도 두 개의 β -케토아실-ACP 신타제(KS) 도메인; (d) 적어도 한 개의 아실트랜스퍼라제(AT) 도메인; (e) 적어도 한 개의 β -케토아실-ACP 환원효소(KR) 도메인; (f) 적어도 두 개의 FabA 유사 β -하이드록시아실-ACP 환원효소(DH) 도메인; (g) 적어도 한 개의 체일인자(CLF) 도메인 및 (h) 적어도 한 개의 말로닐-CoA:ACP 아실트랜스퍼라제(MAT) 도메인. 바람직하게는 그러한 PUFA PKS 시스템은 비박테리아의 PUFA-PKS 시스템이다.

[0056]

한 양태에서, 본 발명의 PUFA PKS 시스템은 비박테리아의 PUFA PKS 시스템이다. 다시 말해 한 양태에서, 본 발명의 PUFA PKS 시스템은 박테리아가 아닌 유기체로부터 분리되거나 진핵생물 또는 고세균(archaeobacterium)과 같은 박테리아가 아닌 유기체 유래의 PUFA PKS 시스템으로부터 유래되거나 이의 상동체이다. 진핵생물은 세포의 분화 정도를 기초로 하여 원핵생물과 구별되는데, 진핵생물은 보다 더 분화된 세포를 가지고 있으며 원핵생물은 덜 분화된 세포를 가지고 있다. 일반적으로 원핵생물은 핵의 멤브레인을 보유하지 않으며, 세포 분열 중에 유사분열을 보이지 않고, 단지 하나의 염색체를 가지고 있으며, 이들의 사이토플라즘은 70S 리보솜을 함유하고, 미토콘드리아, 소포체, 엽록체, 리소솜 또는 골지체를 보유하지 않으며, 이들의 편모는(존재한다면) 단일 원섬유로 구성된다. 대조적으로 진핵생물은 핵의 멤브레인을 가지고 있으며, 세포 분열 중

에 유사 분열을 보이고, 다수의 염색체를 가지며, 이들의 사이토플라즘은 80S 리보솜을 함유하고, 미토콘드리아, 소포체, (조류에서) 엽록체, 리소솜 및 골지체를 보유하고, 이들의 편모는 (존재한다면) 다수의 원섬유로 구성된다. 일반적으로 박테리아는 원핵생물인데 반하여 조류, 진균, 원생생물, 프로토조아 및 보다 고도의 식물이 진핵생물이다.

[0057]

해양 박테리아(예, *쉐와넬라* sp. 군주 SCRC2738 및 비브리오 마리누스)의 PUFA PKS 시스템은 본 발명의 기본을 이루지 않지만, 이러한 박테리아의 PUFA PKS 시스템에서 유래된 도메인을 본 발명에 따른 비박테리아의 PUFA PKS 시스템에서 유래된 도메인과 접합하여 사용하는 것을 고려한다. 또한, 본 발명은 하이브리드 작제물 및 유전적으로 변형된 미생물과 식물을 제조하기 위해서 본 발명에서 언급된 비박테리아의 PUFA PKS 유전자를 변형시키거나 조합하기 위한 PUFA PKS 유전자원으로서 이용하기에 특히 적합할 다른 박테리아(예, *쉐와넬라* *올레아나* 및 *쉐와넬라 자포니카*)로부터 분리된 PUFA PKS 유전자 세트(및 이에 의해 코딩된 단백질 및 도메인)의 분리와 이용을 고려한다. 예를 들어 본 발명에 따른 유전적으로 변형된 유기체는 비박테리아의 PUFA PKS 기능성 도메인을 박테리아의 PUFA PKS 기능성 도메인 뿐만 아니라 다른 PKS 시스템(타입 I, 타입 II, 모듈라) 유래의 PKS 기능성 도메인 또는 단백질 또는 FAS 시스템과 통합하여 제조될 수 있다. 하기에서 상세히 설명되는 *쉐와넬라*의 두 종 즉, *쉐와넬라 올레아나* 또는 *쉐와넬라 자포니카* 유래의 PUFA PKS 유전자는 본 발명에 따른 유전적으로 변형된 미생물, 식물 및 방법에 이용하는데 바람직한 예시적인 박테리아의 유전자이다. 그러한 해양 박테리아(예, *쉐와넬라 올레아나* 또는 *쉐와넬라 자포니카*) 유래의 PUFA PKS 시스템(유전자 및 이에 의해 코딩된 단백질과 도메인)은 신규한 PUFA PKS 서열로서 본 발명에 통합된다.

[0058]

본 발명에서 용어/구 "트라우스토카이트리드", "트라우스토카이트리알레 미생물" 및 "트라우스토카이트리알레 목의 미생물"은 상호 교환적으로 이용될 수 있으며, 트라우스토카이트리알레 목의 멤버를 의미하고, 트라우스토카이트리아시에 과 및 라비린툴라시에 과 모두를 포함한다. 용어 "라비린툴리드" 및 "라비린툴라시에"는 특히 라비린툴라시에 과의 멤버를 의미하기 위하여 사용된다. 트라우스토카이트리아시에 과의 멤버인 트라우스토카이트리드를 특히 의미하기 위하여 용어 "트라우스토카이트리아시에"가 사용된다. 따라서 본 발명에서 라비린툴리드의 멤버는 트라우스토카이트리드에 포함되는 것으로 여겨진다.

[0059]

발달은 트라우스토카이트리드의 분류학에 빈번한 교정을 유발하였다. 분류학의 이론학자들은 트라우스토카이트리드를 조류 또는 조류 유사 원생 생물과 같이 지정한다. 그러나 분류학의 불확실성 때문에 본 발명의 목적에 따라 본 발명에서 트라우스토카이트리드로서 언급되는 군주를 하기 유기체를 포함하는 것이 적합하다: 목: 트라우스토카이트리알레; 과: 트라우스토카이트리아시에(중: 트라우스토카이트리움, 스킴조카이트리움, 자포노카이트리움, 아플라노카이트리움 또는 엘리나) 또는 라비린툴라시에(속: 라비린툴라, 라비린툴로이드 또는 라비린토크사). 또한, 하기 속들이 종종 트라우스토카이트리아시에 또는 라비린툴라시에 둘 중 하나의 군으로 포함되는데: 알도르니카, 코랄로카이트리움, 디플로파이스 및 피르호소러스, 본 발명의 목적에 따라 트라우스토카이트리드 또는 트라우스토카이트리알레 목의 멤버로 포함된다. 본 발명의 시기에 트라우스토카이트리드의 분류학에서의 교정은 라비린툴로이드 종을 라비린툴라시에의 과에 지정하고, 상기 두 개의 트라우스토카이트리아시에 및 라비린툴라시에 과의 위치를 스트라메노필 계통(Stramenopile lineage)으로 확정한다. 라비린툴라시아는 통상적으로 라비린툴리드 또는 라비린툴라 또는 라비린툴로이드라고 불리우며, 트라우스토카이트리아시아는 통상적으로 트라우스토카이트리드로 불리지만, 앞서 설명된 바와 같이 본 발명을 명확하기 위한 목적으로 트라우스토카이트리드는 트라우스토카이트리알레 목의 모든 멤버를 포함하고/거나 트라우스토카이트리아시에 및 라비린툴라시아 모두의 멤버를 포함한다. 최근 분류학에서의 변화를 다음과 같이 정리한다.

[0060]

본 발명에서 개시된 특정 단일 세포의 미생물 군주는 트라우스토카이트리알레 목의 멤버다. 트라우스토카이트리드는 진화한 분류학 역사를 가진 해양의 진핵생물이다. 트라우스토카이트리드의 분류학적 지정과 관련된 문제는 Moss(1986), Bahnweb 및 Jackle(1986) 및 Chamberlain 및 Moss(1988)에 의해 검토되었다.

[0061]

편의를 목적으로 트라우스토카이트리드는 분류학자들에 의하여 다른 무색의 유주자성(zooporphic) 진핵생물과 함께 파이코미세트(Phycomycete, 조류 유사 진균)에 지정되었다. 그러나 파이코미세트라는 이름이 분류학 상태에서부터 떨어지고 트라우스토카이트리드는 우마이세트(Oomycete, 이편모의 유주자성 진균)로 유지되었다. 초기에 우마이세트는 헤테로콘트 조류(heterokont algae)와 관련이 있는 것으로 예상되었고 Barr에 의해 정리된 광범위한 초구조 및 생화학적 연구는 이러한 예상을 지지하였다(Barr, 1981, *Biosystems* 14:359-370). 실제로 우마이세트는 Leedale(Leedale, 1974, *Taxon* 23:261-271)에 의해 수용되었고 다른 조류학자들은 헤테로콘트 조류의 일부로서 수용하였다. 그러나 이들의 종속영양 특징으로부터 야기된 편의상의 문제로 우마이세트와 트라우스토카이트리드는 조류학자(조류를 연구하는 과학자) 보다는 균학자(진균을 연구하는 과학자)에 의해 주로 연

구되었다.

[0062] 또 다른 분류학의 시각으로부터 진화 관련 생물학자들은 진핵생물이 어떻게 진화되었는지에 대한 두 가지 학설을 발달시켜왔다. 하나의 가설은 일련의 내부공생(endosymbioses)을 통한 멤브레인-바운드 소기관의 외인성 유래를 제안한다(Margulis, 1970, *Origin of Eukaryotic Cells*, Yale University Press, New Haven); 예, 미토콘드리아는 박테리아의 내부공생자로부터, 엽록체는 시아노파이트(cyanophyte)로부터, 편모는 스피로헤타(spirochaete)로부터 유래된다. 다른 가설은 자생적 과정(autogenous process)을 통한 원핵생물 조상의 비-멤브레인-바운드 시스템으로부터 멤브레인-바운드 소기관으로의 점진적인 진화를 제안한다(Cavalier-Smith, 1975, *Nature (Lond.)* 256:462-468). 그러나 상기 두 그룹의 진화 관련 생물학자들은 진균으로부터 우마이세트와 트라우스토카이트리드를 제거하고 크로모사이트 조류(chromocyste algae)와 함께 크로모파이트(Chromophyta) 계에(Cavalier-Smith, 1987, *BioSystems* 14:461-481)(상기 계는 최근에 다른 원생생물을 포함하도록 확장되었고 그러한 계의 멤버를 현재 스트라메노필이라고 부른다) 또는 모든 조류와 함께 원생생물 계에 지정하였다(Margulis and Sagen, 1985, *Biosystems* 18:141-147).

[0063] 전자 현미경의 발달로 두 개의 트라우스토카이트리드 속인 트라우스토카이트리움 및 스킨조카이트리움의 유주자낭의 구조에 대한 연구는 트라우스토카이트리아시에는 우마이세트와 관련성이 적다는 증거를 제공하였다(Perkins, 1976, pp.279-312, in "Recent Advances in Aquatic Mycology" (ed.E.B.G. Jones), John Wiley & Sons, New York; Kazama, 1980, *Can.J.Bot.* 58:2324-2446; Barr, 1981, *Biosystems* 14:359-370). 또한, 5-S 리보솜 RNA 서열의 대응되는 분석 (다변량 통계의 형태)을 나타내는 유전적 데이터는 트라우스토카이트리알레는 독특한 진핵 생물의 그룹으로, 진균과는 완전히 구별되며 적갈색의 조류 및 우마이세트의 멤버와 가장 밀접하게 관련 있다는 것을 암시한다(Mannella, et al. 1987, *Mol.Evol.* 24:228-235). 대부분의 분류학자들은 트라우스토카이트리드를 우마이세트로부터 제거하는데 동의하였다(Bartnicki-Garcia, 1987, pp.389-403 in "Evolutionary Biology of the Fungi" (eds. Rayner, A.D.M., Brasier, C.M. & Moore, D. Cambridge University Press, Cambridge).

[0064] 요약하면, Cavalier-Smith(Cavalier-Smith, 1981, *BioSystems* 14:461-481, 1983; Cavalier-Smith, 1993, *Microbiol. Rev.* 57:953-994)의 분류학적 체계를 도입하여 트라우스토카이트리드는 크로모사이트 조류와 함께 크로모파이트 계(스트라메노필)로 분류된다. 이러한 분류학적 지정은 트라우스토카이트리드가 진균이 아닌 크로미스트(chromist)라는 것을 증명한 헤테로콘타의 18S rRNA 신호를 사용하여 Cavalier-Smith에 의해 최근에 재확인되었다(Cavalier-Smith et al. 1994, *Phil. Tran. Roy. Soc. London Series BioSciences* 346:387-397). 이는 트라우스토카이트리드를 진균과는 완전히 상이한 계에 지정한 것이며, 모두 유편자이(Eufungi)계에 지정된다.

[0065] 최근에 71개의 개별적인 진핵생물 유기체의 그룹이 있으며(Patterson 1999) 상기 그룹 내에서 네 개의 주요 계통이 동정되었다.: (1) 알베올레이트(Alveolates), (2) 스트라메노필, (3) a Land Plant-green algae Rhodophyte_Glaucophyte("식물") 클라드 및 (4) Opisthokont 클라드(진균과 동물). 이전에 상기 네 개의 주요 계통은 계로 식별되었으나, 몇몇 연구원들은 "계"라는 개념의 이용을 더 이상 유용하지 않은 것으로 여긴다.

[0066] Armstrong에 의해 주지된 바와 같이 스트라메노필은 삼등분된 관상의 모(hair)와 관련이 있고 상기 계통의 대부분의 멤버는 그러한 모를 보유한 편모를 가지고 있다. 스트라메노필의 운동성 세포들(단인 세포 유기체, 정자 및 유주자)은 두개의 측면으로 삽입된 편모, 즉 편모의 추진을 역전시키는 삼등분 관상 모를 보유한 긴 편모와 짧고 부드러운 편모를 가진 비대칭성이다. 예전에 그룹이 덜 광범위하였을 때 스트라메노필은 이 그룹이 옐로우-그린 조류, 골든-브라운 조류, 유스티그마토파이트(Eustigmatophyte) 및 규조류(Diatom)와 함께 브라운 조류 또는 파오파이트(phaeophyte)를 구성하기 때문에 크로미스타 계 또는 헤테로콘트(=서로 다른 편모) 조류로서 불리었다. 이어서 몇 개의 종속 영양의, 진균 유사 유기체, 물 곰팡이 및 라비린툴리드(점액망 아베마)가 동일한 운동성 세포를 보유한다는 것이 밝혀져서, 광합성 색소 또는 조류를 지칭하는 그룹 명칭이 부적합하게 되었다. 현재, 스트라메노필 계통 내에 있는 두개의 과는 라비린툴라시에 및 트라우스토카이트리시에이다. 역사적으로 이러한 독특한 미생물에 대하여 수많은 분류 전략이 있어왔고 이들은 동일한 목(즉, 트라우스토카이트리알레)으로 분류되었다. 상기 그룹 내에 있는 멤버들 간의 관계는 여전히 변화되고 있다. Porter 와 Leander은 단계적에서 트라우스토카이트리드-라비린툴리드 클라드를 암시하는 18S 스몰 서브유닛 리보솜 DNA를 기반으로 한 데이터를 개발하였다. 그러나, 상기 클라드는 두 가지에 의해 지지된다: 첫 번째는 트라우스토카이트리움 및 울케니아 프로판다를 포함하고 두 번째는 라비린툴라의 세개의 종, 라비린툴로이드의 두 개의 종 및 스킨조카이트리움 아그레가툼을 포함한다.

- [0067] 본 발명에서 이용된 트라우스토카이트리드의 분류학적 지정은 다음과 같이 정리된다:
- [0068] 계: 크로모피타(스트라메노필)
- [0069] 문: 헤테로콘타
- [0070] 목: 트라우스토카이트리알레(트라우스토카이트리드)
- [0071] 과: 트라우스토카이트리아시애 또는 라비린투라시애
- [0072] 속: 트라우스토카이트리움, 스킨조카이트리움, 자포노카이트리움, 아플라노카이트리움, 엘리나, 라비린툴라, 라비린툴로이드 또는 라비린툴로믹사.
- [0073] 초기 일부 계통학자들은 (아메바형의 생존 단계를 가진) 트라우스토카이트리움 속의 본래 멤버 중 일부를 울케니아(*Ulkenia*)로 불리는 별개의 속으로 구별지었다. 그러나 모두는 아니지만, (트라우스토카이트리움과 스킨조카이트리움을 포함한) 대부분의 트라우스토카이트리드는 아메바형의 단계를 보이고, 이로써 울케니아는 일부에 의하여 하나의 확실한 종으로 여겨지지 않는다는 것이 알려져 있다. 본 발명에서 트라우스토카이트리움 속은 울케니아를 포함할 것이다.
- [0074] 목과 계의 더 높은 분류에서 분류학적 지정의 불확실성에도 불구하고 트라우스토카이트리드는 이들의 멤버가 트라우스토카이트리알레 목으로 분류되도록 하는 구별되고 특징적인 그룹을 유지하고 있다.
- [0075] 스킨조카이트리움은 DHA와 도코사헵타엔산(DPA; 22:5 ω -6)이 풍부한; 예를 들면, 건조 중량으로 30% DHA + DPA인 다량의 트리글리세리드를 축적하는 트라우스토카이트리드 해양 미생물이다(Barclay et al. *J. Appl. Phycol.* 6, 123 (1994)). 연장/탈포화 경로에 의해 20- 및 22- 탄소의 PUFA를 합성하는 진핵생물에서 18-, 20- 및 22-탄소 중간체의 풀(pool)은 상대적으로 커서 [^{14} C]-아세테이트를 이용한 생체내 표지 실험은 예측된 중간체에 대한 명백한 전구체-생산물 카이네틱(kinetic)을 밝혀냈다.(Gellerman et al. *Biochim. Biophys. Acta* 573:33 (1979)). 추가로 그러한 유기체로 외인적으로 제공된 방사능 표지된 중간체는 최종 PUFA 생산물로 전환된다. 본 발명자들은 [^{14} C]-아세테이트가 스킨조카이트리움 세포에 의해 신속하게 섭취되고 지방산으로 통합되는 것을 보여주었지만, 가장 짧은 표지 시간(1분)에 DHA가 지방산에서 회수된 표지의 31%를 포함하고 이러한 비율은 [^{14} C]-아세테이트 통합의 10-15분과 연속된 24시간의 배양 성장 동안에 변하지 않고 유지되었다. 유사하게 DPA는 실험 내내 표지의 10%를 보였다. 16- 또는 18- 탄소의 지방산과 22-탄소 다중불포화지방산 간의 전구체-생산물 관계에 대한 증거는 없다. 이러한 결과는 매우 작은(가능한 효소에 결합된) 풀의 중간체를 포함하는 [^{14} C]-아세테이트로부터 DHA의 신속한 합성과 일치한다. 스킨조카이트리움 배양으로부터 유래된 세포 배제 호모게네이트 [^{14} C]-말로닐-CoA를 DHA, DPA 및 포화된 지방산으로 통합하였다. 동일한 생합성 활성이 100,000xg에 의해 상등액 분획물에 유지되지만 멤브레인 침전물에 존재하지는 않았다. 따라서 스킨조카이트리움에서 DHA 및 DPA 합성은 다른 진핵생물에 대하여 언급된 바와 같은 멤브레인-바운드 불포화 효소 또는 지방산 연장 효소를 포함하지 않는다(상기 Parker-Barnes et al. 2000; Shanklin et al. 1998). 상기 분획물 데이터는 스웨넬라 효소로부터 얻은 데이터와는 대조적이며(Metz et al. 2001 참조) 스킨조카이트리움 효소에 의한 CoA와 같은 상이한 (용해되는) 아실 수용체 분자의 이용을 암시한다. 트라우스토카이트리움은 유사한 생화학을 가질 것으로 예상된다.
- [0076] 미국특허 제6,566,583호에서 스킨조카이트리움 유래의 cDNA 라이브러리가 작제되었고 약 8500개의 무작위적 클론(EST)이 서열분석되었다. 도2에 나타난 스웨넬라 PKS 유전자의 11 도메인 중 8에 대한 상동성을 보여준 서열은 0.2-0.5%의 확률로 동정되었다. 미국특허 제6,566,583호에서 스웨넬라 PKS 유전자에 대한 상동성을 보이는 스킨조카이트리움 유래의 수개의 cDNA 클론이 서열분석되었고 다양한 클론이 두개의 부분적인 오픈 리딩 프레임과 하나의 완전한 오픈 리딩 프레임으로 어셈블리되었다.
- [0077] 본 발명자들에 의한 cDNA와 게놈 클론의 추가 서열 분석은 스킨조카이트리움의 OrfA, OrfB 및 OrfC 각각의 전장 게놈 서열의 동정, 스킨조카이트리움에 있는 도메인(도 2 참조)에 대한 상동성을 가진 스킨조카이트리움에 있는 도메인의 완전한 동정을 가능하게 하였다. 미국특허출원 제10/124,800호에 상세히 기술되고 하기에서 상세히 설명된다.
- [0078] 본 발명자들은 트라우스토카이트리움 종의 트라우스토카이트리드(특히, 트라우스토카이트리움 sp.

23B(ATCC20892))에 있는 상동성의 Orf의 전장 게놈 서열을 동정, 클론 및 서열분석하였고 상기 트라우스토카이트리움에서 PUFA PKs 시스템의 도메인을 동정하였다.

[0079] 스웨넬라의 PUFA PKs 시스템의 도메인과 스킨조카이트리움의 PUFA PKs 시스템의 도메인의 비교를 기초로 하여 스킨조카이트리움 게놈은 EPA 합성을 촉매할 수 있는 스웨넬라에 있는 단백질과 상당히 유사한 단백질을 코딩한다. 스킨조카이트리움에 있는 단백질은 DHA 및 DPA 합성을 촉매하는 PUFA PKs를 구성한다. 스웨넬라에 대하여 동정된 반응 전략의 간단한 변형은 스킨조카이트리움에서 DHA 합성을 가능하게 할 것이다. 원핵생물인 스웨넬라와 진핵생물인 스킨조카이트리움간의 상동성은 PUFA PKs가 측면 유전자 전달을 수행한다는 것을 지지한다.

[0080] 트라우스토카이트리움에 대한 유사한 비교가 이루어질 수 있다. 모든 경우에 트라우스토카이트리움 23B(Th.23B) PUFA PKs 단백질 또는 도메인과 다른 공지된 서열의 비교는 미국특허출원 10/124,800호에 기술된 스킨조카이트리움 PUFA PKs 단백질 (OrfA, B 또는 C 또는 이의 도메인)의 어느 하나와 가장 가깝게 매치된다는 것을 보여주었다. 모든 경우에서 그 다음으로 가깝게 매치되는 대상은 해양 박테리아(스웨넬라 SCRC-2738, 스웨넬라 오네이텐시스, 포토박터 프로판툼 및 모리텔라 마리아)로부터 유래된 PUFA PKs 단백질 중 어느 하나 또는 질소고정 시아노박테리아(예, 노스톡 평티포르메 및 노스톡 sp. PCC 7120)에서 발견되는 관련 시스템에서 유래된 것이었다. 시아노박테리아의 효소 시스템의 생산물은 이중 결합이 결여되고 상기 단백질은 시스 이중 결합 형성을 암시하는 DH 도메인과 관련된 도메인(즉, FabA와 관련된 DH 도메인)이 결여된다.

[0081] 본 발명에서 용어 "오픈 리딩 프레임(open reading frame)"은 약자 "Orf"로 표기한다. 오픈 리딩 프레임에 의해 코딩된 단백질은 모두 대문자로 "ORF"라고 표기하고, 오픈 리딩 프레임인 핵산 서열은 모두 소문자 "orf"라고 표기하지만, 일치시키기 위하여 스펠링 "Orf"가 핵산 서열 또는 이에 의해 코딩된 단백질 모두를 정의하는데 바람직하게 사용된다. 단백질 또는 핵산 서열을 정의하든지 간에 상기 용어의 용도는 문맥으로부터 명백해질 것이다.

[0082] 스킨조카이트리움 PUFA PKs

[0083] 도 1은 스킨조카이트리움 PUFA PKs 시스템에서 유래된 세 개의 오픈 리딩 프레임을 도식적으로 나타낸 것이고, 상기 PUFA PKs 시스템의 도메인 구조를 포함한다. 미국특허출원 제10/124,800호에 상세히 기술된 바와 같이 각 오픈 리딩 프레임의 도메인 구조는 다음과 같다:

[0084] 오픈 리딩 프레임 A(OrfA):

[0085] OrfA의 완전한 뉴클레오타이드 서열은 서열번호: 1에 나타내었다. OrfA는 서열번호: 2에 기재된 2910개의 아미노산 서열을 코딩하는 (정지 코돈을 포함하지 않는) 8730개의 뉴클레오타이드 서열이다. OrfA 내에는 12개의 도메인이 있다: (a) 하나의 β -케토아실-ACP 신타제(KS) 도메인; (b) 하나의 말로닐-CoA:ACP 아실트랜스퍼라제 (MAT) 도메인; (c) 아홉 개의 아실 캐리어 단백질 (ACP) 도메인; 및 (d) 하나의 β -케토아실-ACP 환원효소 (KR) 도메인. OrfA의 뉴클레오타이드 서열은 GenBank에 Accession No. AF378327(아미노산 서열 Accession No. AAK728879)로 수탁되었다.

[0086] 스킨조카이트리움 OrfA의 첫 번째 도메인은 β -케토아실-ACP 신타제(KS) 도메인이며, 본 발명에서 OrfA의-KS라고도 한다. 상기 도메인은 서열번호: 1(OrfA)의 약 1 내지 40 위치의 시작점으로부터 서열번호 1의 약 1428 내지 1500 위치의 종결점에 이르는 뉴클레오타이드 서열 내에 포함된다. OrfA-KS 도메인을 코딩하는 서열을 포함하는 뉴클레오타이드 서열은 서열번호: 7에 나타내었다(서열번호: 1의 1-1500 위치). KS 도메인을 포함하는 아미노산 서열은 서열번호: 2(OrfA)의 약 1 내지 14 위치의 시작점으로부터 서열번호: 2의 약 476 내지 500 위치의 종결점까지 이른다. OrfA-KS 도메인을 포함하는 아미노산 서열은 서열번호: 8(서열번호:2의 1-500 위치)에 나타내었다. OrfA-KS 도메인은 활성 부위 모티프:DXAC(*아실 결합 부위 C₂₁₅)를 포함한다.

[0087] 본 발명에서 β -케토아실-ACP 신타제(KS) 생물학적 활성(기능)을 가진 도메인 또는 단백질은 FAS (및 PKs) 연장 반응 사이클의 개시 단계를 수행하는 효소로서 특징지어진다. 용어 " β -케토아실-ACP 신타제"는 "3-케토 아실-ACP 신타제", " β -케토 아실-ACP 신타제" 및 "케토-아실 ACP 신타제" 및 유사한 파생어와 상호교환적으로 이용될 수 있다. 연장을 위하여 예정된 아실기는 티오에스테르 결합에 의하여 효소의 활성 부위에 있는 시스테인 잔기로 결합된다. 단단계 반응으로 상기 아실-효소는 말로닐-ACP와 축합하여 -케토아실-ACP, CO₂와 유리 효소를 형성한다. KS는 연장 사이클에서 중요한 역할을 수행하고, 많은 시스템에서 상기 반응 사이클의

다른 효소 보다 높은 기질 특이성을 보유하는 것으로 보여진다. 예를 들어 대장균은 세 개의 개별적인 KS 효소를 함유하는데 각각은 유기체의 생리학에서 자신만의 특정한 역할을 갖는다(Magnuson et al. *Microbiol. Rev.*, 57, 522(1993)). PUFA-PKS 시스템의 두개의 KS 도메인은 PUFA 생합성 반응 순서에서 개별적인 역할을 가질 수 있다.

[0088] KS는 효소의 한 분류로서 충분히 특징되었다. 많은 확인된 KS 유전자의 서열이 공지되었고, 활성 부위 모티프가 동정되었으며, 몇 개의 결정 구조가 결정되었다. 단백질(또는 단백질의 도메인)은 공지된 KS 서열에 대한 상동성에 의해 KS 군의 효소에 속하는 것으로서 동정될 수 있다.

[0089] OrfA에 있는 두 번째 도메인은 말로닐-CoA:ACP 아실트랜스퍼라제(MAT) 도메인이고, 본 발명에서 OrfA-MAT라고도 한다. 상기 도메인은 서열번호: 1(OrfA)의 약 1723 내지 1798 위치의 시작점으로부터 서열번호: 1의 약 2805 내지 3000 위치의 종결점에 이르는 뉴클레오타이드 서열 내에 포함된다. OrfA-MAT 도메인을 코딩하는 서열을 포함하는 뉴클레오타이드 서열은 서열번호: 9에 나타내었다(서열번호:1의 1723-3000 위치). MAT 도메인을 포함하는 아미노산 서열은 서열번호: 2(OrfA)의 약 575 내지 600 위치의 시작점으로부터 서열번호: 2의 약 935 내지 1000 위치의 종결점까지 이른다. OrfA-MAT 도메인을 포함하는 아미노산 서열은 서열번호: 10(서열번호: 2의 575-1000 위치)에 나타내었다. OrfA-MAT 도메인은 서열번호: 11에 나타낸 바와 같이 활성 부위 모티프:GHS XG(*아실 결합 부위 S₇₀₆)를 포함한다.

[0090] 본 발명에서, 말로닐-CoA:ACP 아실트랜스퍼라제(MAT) 생물학적 활성(기능)을 가진 도메인 또는 단백질은 말로닐-CoA로부터 ACP로 말로닐 부를 전달하는 효소로서 특징지어진다. 용어 "말로닐-CoA:ACP 아실트랜스퍼라제"는 "말로닐 아실트랜스퍼라제" 및 유사한 파생어와 상호교환적으로 이용될 수 있다. 상기 효소는 활성 부위 모티프(GxSxG) 뿐만 아니라 (스키조카이트리움 OrfB의 AT 도메인과는 대조적으로) 상기 효소를 MAT 효소로서 동정하게 한 연장된 모티프(중요 위치에 있는 R 및 Q 아미노산)를 보유한다. (PUFA PKS 도메인을 제외하고) 일부 PKS 시스템에서 MAT 도메인은 메틸- 또는 에틸- 말로네이트를 (상응하는 CoA 에스테르 유래의) ACP기로 로딩할 것이고, 이에 의해 축적을 선행의 탄소쇄로 도입할 것이다. MAT 도메인은 공지된 MAT 서열에 대한 이들의 상동성 및 이들의 연장된 모티프 구조에 의하여 인식될 수 있다.

[0091] OrfA의 3-11번째 도메인은 아홉 개의 직렬 연결된 아실 캐리어 단백질(ACP) 도메인이고, 본 발명에서 OrfA-ACP라고도 한다(서열에서 첫 번째 도메인은 OrfA-ACP1이고, 두 번째 도메인은 OrfA-ACP2이며, 세 번째 도메인은 OrfA-ACP3이다). 첫 번째 ACP 도메인인 OrfA-ACP1은 서열번호: 1 (OrfA)의 약 3343 위치로부터 약 3600위치에 이르는 뉴클레오타이드 서열 내에 포함된다. OrfA-ACP1 도메인을 코딩하는 뉴클레오타이드 서열은 서열번호: 12(서열번호: 1의 3343-3600 위치)에 나타내었다. 첫 번째 ACP 도메인을 포함하는 아미노산 서열은 서열번호: 2의 약 1115 위치로부터 1200 위치까지 이른다. OrfA-ACP1 도메인을 포함하는 아미노산 서열은 서열번호: 13(서열번호: 2의 1115-1200 위치)에 나타내었다. OrfA-ACP1 도메인은 서열번호: 14에 나타낸 바와 같이 활성 부위 모티프: LGIDS* (*판테테인 결합 모티프 S₁₁₅₇)를 포함한다.

[0092] 아홉 개의 ACP 도메인의 뉴클레오타이드 및 아미노산 서열은 매우 보존적이어서 각 도메인의 서열은 개별적인 서열 식별자에 의해 나타내지 않았다. 그러나 본 발명에 개시된 정보를 기반으로 하여 당업자는 다른 여덟 개의 ACP 도메인 각각을 포함하는 서열을 용이하게 결정할 수 있다(하기 설명을 참조).

[0093] 아홉 개의 모든 ACP 도메인은 서열번호: 1의 약 3283 위치로부터 6288 위치까지의 OrfA의 영역에 이르며, 서열번호: 2의 약 1095로부터 2096까지의 아미노산 위치에 대응한다. 아홉 개의 모든 도메인을 포함하는 전체 ACP 영역의 뉴클레오타이드 서열은 서열번호: 16에 나타내었다. 서열번호: 16에 기재된 영역은 각각의 ACP 도메인 사이에 존재하는 링커 절편을 포함한다. 아홉 개 도메인의 반복적인 간격은 서열번호: 16의 약 330 뉴클레오타이드(근접한 활성 부위 세린 간에 측정된 실질적인 아미노산 수가 104 내지 116 아미노산 이내이다)마다이다. 아홉 개의 ACP 도메인 각각은 (서열번호: 14에 나타낸) 판테테인 결합 모티프 LGIDS*을 포함하며, 상기 S*는 판테테인 결합 부위 세린(S)이다. 판테테인 결합 부위 세린은 각 ACP 도메인 서열의 중심 근처에 위치한다. ACP 도메인 영역의 각 말단과 각 ACP 도메인 사이에는 프롤린(P)과 알라닌(A)이 매우 풍부한 영역이 있는데 상기 영역은 링커 영역인 것으로 여겨진다. 예를 들어 ACP 도메인 1과 2 사이에는 서열번호: 15에 나타낸 APAPVKAAAPVAPASAPAPA 서열이 있다. 서열번호: 2의 아미노산 서열에 대한 아홉 개 ACP 도메인 각각의 활성 부위 세린 잔기의 위치(즉, 판테테인 결합 부위)는 다음과 같다: ACP1 = S₁₁₅₇; ACP2 = S₁₂₆₆; ACP3 = S₁₃₇₇; ACP4 = S₁₄₈₈; ACP5 = S₁₆₀₄; ACP6 = S₁₇₁₅; ACP7 = S₁₈₁₉; ACP8 = S₁₉₃₀; 및 ACP9 = S₂₀₃₄. ACP 도메인의 평균 크기는 링커

를 제외하면 약 85개의 아미노산이고, 링커를 포함하면 약 110개의 아미노산이며, 도메인의 중심 근처에 활성 부위 세린이 있다는 것이 주어진 바 당업자라면 OrfA에 있는 아홉 개 ACP 도메인 각각의 위치를 용이하게 결정할 수 있다.

[0094] 본 발명에서 아실 캐리어 단백질(ACP) 생물학적 활성(기능)을 가진 도메인 또는 단백질은 단백질에 공유결합으로 결합된 보조인자로의 티오에스테르 결합을 통하여 성장하는 지방 아실쇄의 캐리어로서 기능을 발휘하는 작은 폴리펩타이드(전형적으로 80 내지 100개의 아미노산)인 것으로 특징지어진다. 이들은 더 큰 단백질 내에서 개별적인 유닛 또는 도메인으로서 존재한다. ACP는 CoA의 포스포판테테닐 부를 ACP의 매우 보존적인 세린 잔기로 전달함으로써 불활성 아포-형태(apo-form)에서 기능성의 홀로-형태(holo-form)로 전환된다. 아실기는 포스포판테테닐 부의 자유 말단에서 티오에스테르 결합에 의해 ACP로 부착된다. ACP는 방사능의 판테테닐로 표지함으로써 그리고 공지된 ACP에 대한 서열 상동성에 의해 동정될 수 있다. 상기된 모티브(LGIDS*)의 다양성의 존재가 또한 ACP의 신호이다.

[0095] OrfA에 있는 12번째 도메인은 β -케토아실-ACP 환원효소(KR) 도메인이며, 본 발명에서 OrfA-KR이라고도 한다. 상기 도메인은 서열번호: 1의 약 6598 위치의 시작점으로부터 서열번호: 1의 약 8730 위치의 종결점에 이르는 뉴클레오타이드 서열 내에 포함된다. OrfA-KR 도메인을 코딩하는 서열을 포함하는 뉴클레오타이드 서열은 서열번호: 17(서열번호: 1의 6598-8730 위치)에 나타내었다. KR 도메인을 포함하는 아미노산 서열은 서열번호: 2(OrfA)의 약 2200 위치의 시작점으로부터 서열번호: 2의 약 2910 위치의 종결점까지 이른다. OrfA-KR 도메인을 포함하는 아미노산 서열은 서열번호: 18(서열번호: 2의 2200-2910 위치)에 나타내었다. KR 도메인 내에는 단쇄 알데하이드-탈수소효소에 상동성을 가진 코어 영역이 있다(KR은 상기 군의 멤버다). 상기 코어 영역은 서열번호: 1의 약 7189로부터 약 7500 위치에 이르고 이는 서열번호:2의 2400-2500 아미노산 위치에 대응된다.

[0096] 본 발명에서 β -케토아실-ACP 환원효소(KR) 활성을 가진 도메인 또는 단백질은 피리딘-뉴클레오타이드 의존적인 ACP의 3-케토아실 형태의 환원을 촉매하는 효소로서 특징지어진다. 용어 " β -케토아실-ACP 환원효소"는 "케토환원효소", "3-케토아실-ACP 환원효소", "케토-아실 ACP 환원효소" 및 상기 용어의 유사한 파생어가 상호교환적으로 이용될 수 있다. 이는 데노보(de novo) 지방산 생합성 연장 사이클과 폴리케타이드 생합성에서 수행되는 반응에서 첫 번째 환원 단계이다. 상당한 서열 유사성이 엔오일-ACP 환원효소(ER)의 일 군, (PUFA PKS 시스템에 존재하는 ER군 이외의) FAS의 다른 환원효소 및 단쇄 알코올 탈수소효소 군에서 확인된다. 상기된 PUFA PKS 영역의 Pfam 분석은 코어 영역에서 단쇄 알코올 탈수소효소의 군과의 상동성을 보여주었다. 동일한 영역의 Blast 분석은 다른 특징지어진 PUFA PKS 시스템 유래의 도메인과 상동성의 연장된 영역 뿐만 아니라 공지된 KR 효소와 코어 범위에서 매치된다.

[0097] 오픈 리딩 프레임 B(OrfB):

[0098] OrfB의 완전한 뉴클레오타이드 서열은 서열번호: 3에 나타내었다. OrfB는 서열번호: 4에 나타낸 2059 개의 아미노산 서열을 코딩하는 (정지 코돈을 포함하지 않는) 6177개의 뉴클레오타이드 서열이다. OrfB 내에는 네 개의 도메인이 있다: (a) 하나의 β -케토아실-ACP 신타제(KS) 도메인; (b) 하나의 쇠길인자(CLF) 도메인; (c) 하나의 아실트랜스퍼라제(AT) 도메인 및 (d) 하나의 엔오일-ACP 환원효소(ER) 도메인. OrfB의 뉴클레오타이드 서열은 GenBank에 Accession No. AF378328(아미노산 서열 Accession No. AAK728880)로 수탁되었다.

[0099] OrfB에 있는 첫 번째 도메인은 β -케토아실-ACP 신타제(KS) 도메인이고, 본 발명에서 OrfB-KS라고도 한다. 상기 도메인은 서열번호: 3(OrfB)의 약 1 내지 43 위치의 시작점으로부터 서열번호: 3의 약 1332 내지 1350 위치의 종결점에 이르는 뉴클레오타이드 서열 내에 포함된다. OrfB-KS 도메인을 코딩하는 서열을 포함하는 뉴클레오타이드 서열은 서열번호: 19(서열번호: 3의 1-1350)에 나타내었다. KS 도메인을 포함하는 아미노산 서열은 서열번호: 4(OrfB)의 약 1 내지 15 위치의 시작점으로부터 서열번호: 4의 약 444 내지 450 위치의 종결점까지 이른다. OrfB-KS 도메인을 포함하는 아미노산 서열은 서열번호: 20에 나타내었다(서열번호: 4의 1-450 위치). OrfB-KS 도메인은 활성 부위 모티브:DXAC*(*아실 결합 부위 C₁₉₆)를 포함한다. KS의 생물학적 활성 및 그러한 활성을 가진 단백질 또는 도메인을 동정하는 방법은 앞서 기술되었다.

[0100] OrfB에 있는 두 번째 도메인은 쇠길인자(CLF) 도메인이고, 본 발명에서 OrfB-CLF라고도 한다. 상기 도메인은 서열번호: 3(OrfB)의 약 1378 내지 1402 위치의 시작점으로부터 서열번호: 3의 약 2682 내지 2700 위치의 종결점에 이르는 뉴클레오타이드 서열 내에 포함된다. OrfB-CLF 도메인을 코딩하는 서열을 포함하는 뉴클레오타이드 서열은 서열번호: 21(서열번호: 3의 1378-2700)에 나타내었다. CLF 도메인을 포함하는 아미노산 서열은 서열번호: 4(OrfB)의 약 460 내지 468 위치의 시작점으로부터 서열번호: 4의 약 894 내지 900 위치의 종결

점까지 이른다. OrfB-CLF 도메인을 포함하는 아미노산 서열은 서열번호: 22(서열번호: 4의 460-900 위치)에 나타내었다. OrfB-CLF 도메인은 아실 결합 시스테인 없는 KS 활성 부위 모티브를 포함한다.

[0101]

본 발명에서 도메인 또는 단백질은 하기 논리를 기본으로 하여 쇠길인자(CLF)로서 정의된다. CLF는 본래 타입 II(분해되는 효소) PKS 시스템의 특징으로서 정의되었고 연장 사이클의 회수, 그 결과 최종 산물의 쇠길이를 결정하는데 중요한 역할을 하는 것으로 가정되었다. CLF 아미노산 서열은 KS 도메인에 상동성을 보이지만 (그리고 KS 단백질과 이중이량체를 형성하는 것으로 여겨지지만) 활성 부위 시스테인이 결여되어 있다. PKS 시스템에서 CLF의 역할은 현재 논쟁중이다. 새로운 증거(C. Bisang et al. *Nature*, 401, 502(1999))는 PKS 시스템을 준비하는(연장될 초기 아실기를 제공하는) 역할을 제안한다. 이러한 역할에서 CLF 도메인은 말로네이트(말로닐-ACP로서)를 탈카르복실화시켜 KS 활성 부위로 전달될 수 있는 아세테이트기를 형성할 것으로 여겨진다. 따라서 이러한 아세테이트는 초기 연장(축합) 반응을 수행할 수 있는 '준비(priming)' 분자로서 작용한다. 타입 II CLF의 상동체가 일부 모듈라 PKS 시스템에서 '로딩' 도메인으로서 동정되었다. CLF의 서열 특징을 가진 도메인이 최근에 동정된 모든 PUFA PKS 시스템에서 확인되고 각각의 경우에 다중도메인 단백질의 일부로서 발견된다.

[0102]

OrfB에 있는 세 번째 도메인은 AT 도메인이고, 본 발명에서 OrfB-AT라고도 한다. 상기 도메인은 서열번호: 3(OrfB)의 약 2701 내지 3598 위치의 시작점으로부터 서열번호: 3의 약 3975 내지 4200 위치의 종결점에 이르는 뉴클레오타이드 서열 내에 포함된다. OrfB-AT 도메인을 코딩하는 서열을 포함하는 뉴클레오타이드 서열은 서열번호: 23(서열번호: 3의 2701-4200 위치)에 나타내었다. AT 도메인을 포함하는 아미노산 서열은 서열번호: 4(OrfB)의 약 901 내지 1200 위치의 시작점으로부터 서열번호: 4의 약 1325 내지 1400 위치의 종결점까지 이른다. OrfB-AT 도메인을 포함하는 아미노산 서열은 서열번호: 24(서열번호: 4의 901-1400 위치)에 나타내었다. OrfB-AT 도메인은 아실트랜스퍼라제(AT) 단백질의 특징인 활성 부위 모티브 GxS^{*}xG(*아실 결합 부위 S₁₁₄₀)을 포함한다.

[0103]

"아실트랜스퍼라제" 또는 "AT"는 다수의 개별적인 아실 전달 반응을 수행할 수 있는 효소의 일반적인 부류를 의미한다. 용어 "아실트랜스퍼라제"는 용어 "아실 트랜스퍼라제"와 상호교환적으로 사용될 수 있다. 스키토카이트리움 도메인은 현재 연구중인 모든 다른 PUFA PKS 시스템에 존재하는 도메인과 상당한 상동성을 보이지만 아실트랜스퍼라제 기능이 동정된 일부 아실트랜스퍼라제(예, 말로닐-CoA:ACP 아실트랜스퍼라제, MAT)에 대해서는 거의 상동성을 보이지 않았다. MAT와의 약한 상동성에도 불구하고 상기 AT 도메인은 MAT 효소의 연장된 모티브 구조 특징을 보유하지 않기 때문에 상기 AT 도메인이 MAT로서 기능을 발휘할 것으로 여겨지지 않는다(상기 MAT 도메인에 대한 설명 참조). 상기된 목적에 따라 PUFA PKS 시스템에 있는 AT 도메인의 기능은 이에 제한되는 것은 아니지만: OrfA ACP 도메인으로부터 물로 지방 아실기를 전달 (즉, 티오에스테라제-유리 지방산으로서 지방 아실기를 방출), CoA와 같은 수용체로 지방 아실기를 전달, 다양한 ACP 도메인간에 아실기를 전달 또는 친지질 수용체 분자(예, 리소포스파드산)로 지방 아실기의 전달을 포함한다.

[0104]

OrfB에 있는 네 번째 도메인은 ER 도메인이고, 본 발명에서 OrfB-ER이라고도 한다. 상기 도메인은 서열번호: 3(OrfB)의 약 4648 위치의 시작점으로부터 서열번호: 3의 약 6177 위치의 종결점에 이르는 뉴클레오타이드 서열 내에 포함된다. OrfB-ER 도메인을 코딩하는 서열을 포함하는 뉴클레오타이드 서열은 서열번호: 25(서열번호: 3의 4648-6177)에 나타내었다. ER 도메인을 포함하는 아미노산 서열은 서열번호: 4(OrfB)의 약 1550 위치의 시작점으로부터 서열번호: 4의 약 2059 위치의 종결점까지 이른다. OrfB-ER 도메인을 포함하는 아미노산 서열은 서열번호: 26(서열번호: 4의 1550-2059 위치)에 나타내었다.

[0105]

본 발명에서 이러한 도메인은 엔오일-ACP 환원효소(ER) 생물학적 활성을 가진다. 본 발명에서 용어"엔오일-ACP 환원효소"는 "엔오일 환원효소", "엔오일 ACP-환원효소" 및 "엔오일 아실-ACP 환원효소"와 상호교환적으로 이용될 수 있다. ER 효소는 지방 아실-ACP에 있는 (DH 활성에 의해 도입된) 트랜스-이중 결합을 환원하여 결과적으로 상기 탄소를 완전히 포화시킨다. PUFA-PKS에 있는 ER 도메인은 새롭게 특정된 ER 효소의 군과 상동성을 보인다(Heath et al. *Nature* **406**, 145 (2000)). Heath 및 Rock는 스트렙토코쿠스 뉴모니애로부터 대상 유전자를 클로닝하고, 그 유전자로부터 발현된 단백질을 정제하고, 실험실내 에세이에서 ER 활성을 가지고 있다는 것을 보여줌으로써 이러한 ER 효소의 새로운 부류를 동정하였다. OrfB의 스키토카이트리움 ER 도메인의 서열은 S. 뉴모니애 ER 단백질과 상동성을 보인다. 현재 연구중인 모든 PUFA PKS 시스템은 스키토카이트리움 ER 도메인과 매우 높은 서열 상동성을 가진 적어도 하나의 도메인을 포함한다. 스키토카이트리움 PUFA PKS 시스템은 두개의 ER 도메인을 포함한다(OrfB 상에 하나와 OrfC 상에 하나).

[0106]

오픈 리딩 프레임 C(OrfC):

[0107]

OrfC의 완전한 뉴클레오타이드 서열은 서열번호: 5에 나타내었다. OrfC는 서열번호:6에 나타낸 1503개의 아미노산 서열을 코딩하는 (정지 코돈을 포함하지 않는) 4509개의 뉴클레오타이드 서열이다. OrfC 내에는 세 개의 도메인이 있다: (a) 두개의 FabA 유사 β -하이드록시아실-ACP 탈수소효소(DH) 도메인; 및 (b) 하나의 엔오일-ACP 환원효소(ER) 도메인. OrfC의 뉴클레오타이드 서열은 GenBank에 Accession No. AF378329(아미노산 서열 Accession No. AAK728881)로 수탁되었다.

[0108]

OrfC에 있는 첫 번째 도메인은 DH 도메인이고, 본 발명에서 OrfC-DH1이라고도 한다. 이는 OrfC에 있는 두개의 DH 도메인 중 하나이며, 그래서 DH1으로 명명된다. 상기 도메인은 서열번호: 5(OrfC)의 약 1 내지 778 위치의 시작점으로부터 서열번호: 5의 약 1233 내지 1350 위치의 종결점에 이르는 뉴클레오타이드 서열 내에 포함된다. OrfC-DH1 도메인을 코딩하는 서열을 포함하는 뉴클레오타이드 서열은 서열번호: 27(서열번호: 5의 1-1350 위치)에 나타내었다. DH1 도메인을 포함하는 아미노산 서열은 서열번호: 6(OrfC)의 약 1 내지 260 위치의 시작점으로부터 서열번호: 6의 약 411 내지 450 위치의 종결점까지 이른다. OrfC-DH1 도메인을 포함하는 아미노산 서열은 서열번호: 28(서열번호: 6의 1-450 위치)에 나타내었다.

[0109]

본 발명에서, 상기 도메인은 FabA 유사 β -하이드록시아실-ACP 탈수소효소(DH) 생물학적 활성을 가진다. 용어 "FabA 유사 β -하이드록시아실-ACP 탈수소효소"는 용어 "FabA 유사 β -하이드록시 아실-ACP 탈수소효소", " β -하이드록시아실-ACP 탈수소효소", "환원효소" 및 유사한 파생어와 상호교환적으로 이용될 수 있다. PUFA PKS 시스템에 있는 두개의 DH 도메인(하기 DH2의 경우 참조)은 앞부분에서 기술되었다. 이러한 부류의 효소는 β -케토아실-ACP로부터 HOH를 제거하고 탄소쇄에 트랜스 이중 결합을 남긴다. PUFA PKS 시스템의 DH 도메인은 (다른 PKS 시스템의 DH 도메인 보다는) FAS 시스템과 관련된 박테리아의 DH 효소에 상동성을 보인다. 박테리아의 DH, FabA 유사 DH의 일부 집합은 시스-트랜스 이소머라제 활성을 보유한다(Heath et al. *J.Biol.Chem.* 271, 27795 (1996)). FabA 유사 DH에 대한 상동성은 두 개의 DH 도메인 중 하나가 PUFA PKS 산물에서 시스 이중 결합의 삽입에 관련이 있다는 것을 암시한다.

[0110]

OrfC에 있는 두 번째 도메인은 DH 도메인이고, 본 발명에서 OrfC-DH2라고도 한다. 이는 OrfC에 있는 두 개의 DH 도메인 중 두 번째이며, 그래서 DH로서 명명된다. 상기 도메인은 서열번호: 5(OrfC)의 약 1351 내지 2437 위치의 시작점으로부터 서열번호: 5의 약 2607 내지 2850 위치의 종결점에 이르는 뉴클레오타이드 서열 내에 포함된다. OrfC-DH2 도메인을 코딩하는 서열을 포함하는 뉴클레오타이드 서열은 서열번호: 29(서열번호: 5의 1351-2850)에 나타내었다. DH2 도메인을 포함하는 아미노산 서열은 서열번호: 6(OrfC)의 약 451 내지 813 위치의 시작점으로부터 서열번호: 6의 약 869 내지 950 위치의 종결점까지 이른다. OrfC-DH2 도메인을 포함하는 아미노산 서열은 서열번호: 30(서열번호: 6의 451-950 위치)에 나타내었다. DH 생물학적 활성은 앞서 기술되었다.

[0111]

OrfC에 있는 세 번째 도메인은 ER 도메인이고, 본 발명에서 OrfC-ER이라고도 한다. 상기 도메인은 서열번호: 5(OrfC)의 약 2998 위치의 시작점으로부터 서열번호: 5의 약 4509 위치의 종결점에 이르는 뉴클레오타이드 서열 내에 포함된다. OrfC-ER 도메인을 코딩하는 서열을 포함하는 뉴클레오타이드 서열은 서열번호: 31(서열번호: 5의 2998-4509)에 나타내었다. ER 도메인을 포함하는 아미노산 서열은 서열번호: 6(OrfC)의 약 1000 위치의 시작점으로부터 서열번호: 6의 약 1502 위치의 종결점까지 이른다. OrfC-ER 도메인을 포함하는 아미노산 서열은 서열번호: 32(서열번호: 6의 1000-1502 위치)에 나타내었다. ER 생물학적 활성은 앞서 기술되었다.

[0112]

트라우스토카이트리움 23B PUFA PKS

[0113]

Th. 23B 오픈 리딩 프레임 A(OrfA):

[0114]

Th. 23B OrfA의 완전한 뉴클레오타이드 서열은 서열번호: 38에 나타내었다. 서열번호: 38은 Th. 23B OrfA에 있는 하기 도메인을 코딩한다: (a) 하나의 β -케토아실-ACP 신타제(KS) 도메인; (b) 하나의 말로닐-CoA:ACP 아실트랜스퍼라제(MAT) 도메인; (c) 여덟 개의 아실 캐리어 단백질(ACP)도메인; 및 (d) 하나의 β -케토아실-ACP 환원효소(KR) 도메인. 이러한 도메인 조직화는 Th. 23B OrfA는 8개의 근접한 ACP 도메인을 가진 반면에, 스킴조카이트리움 OrfA는 9개의 근접한 ACP 도메인을 가지고 있다는 점을 제외하고는 스킴조카이트리움 OrfA(서열번호: 1)에 존재하는 바와 동일하다. Th. 23B OrfA는 서열번호: 39에 나타낸 2811개의 아미노산 서열을 코딩하는 (정지 코돈을 포함하지 않는) 8433개의 뉴클레오타이드 서열이다. Th. 23B OrfA 아미노산 서열(서

열번호: 39)은 표준 BLAST 조사로 공지된 서열과 비교되었다(BLAST 변수: Blastp, 낮은 복잡성 필터 오프, 프로그램- BLOSUM62, 캡-cost-존재:11, 신장:1)(Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res. 25: 3389-3402; 이의 전문은 본 발명에서 참조로 인용된다). 아미노산 레베에서 Th. 23B OrfA에 가장 높은 상동성을 가진 서열은 스킨조카이트리움 OrfA(gb AAK72879.1)(서열번호: 2)이었다. 정렬은 (ACP 반복 수의 차이로 인하여) 2 조각으로 분리된 전체 질의(query)에 걸쳐서 신장된다. 서열번호: 39는 먼저 서열번호: 2와 (8개의 ACP 도메인을 포함하는) 6 내지 1985 위치에서 정렬되고 2017개의 아미노산에 걸쳐서 서열번호: 2와 54%의 서열 동일성을 보인다. 또한 서열번호: 39는 서열번호: 2와 980 내지 2811 위치에서 정렬되고 1861개의 아미노산에 걸쳐서 서열번호: 2와 43%의 서열 동일성을 보인다. 상기 두 번째 정렬에서, 매치는 Th. 23B 8 X ACP의 보존적인 판테테인 부착 부위 모티브의 영역에서 명백하지만, 첫 번째 스킨조카이트리움 ACP 도메인에 걸쳐서 매우 적었다(즉, Th. 23B 질의 서열에는 9번째 ACP 도메인이 없지만 이러한 조건 하에서 Blastp 출력은 이들의 정렬을 시도한다). 서열번호: 39는 세와넬라 오네이텐시스 (Accession No. NP_717214) 및 포토박터 프로판툼(Accession No. AAL_01060) 유래의 서열과 그 다음으로 가장 가까운 동일성을 보인다.

[0115] Th. 23B OrfA에 있는 첫 번째 도메인은 KS 도메인이고, 본 발명에서 Th. 23B OrfA-KS라고도 한다. KS 도메인의 기능을 앞서 기술되었다. 상기 도메인은 서열번호: 40으로 나타낸 서열번호: 38의 약 1 위치로부터 약 1500 위치까지 이르는 뉴클레오타이드 서열 내에 포함된다. Th. 23B KS 도메인을 포함하는 아미노산 서열은 서열번호: 41로 나타낸 서열번호: 39의 약 1 위치로부터 약 500 위치까지 이르는 서열번호: 39의 영역이다. 서열번호: 39의 이러한 영역은 서열번호: 39의 약 1 위치로부터 약 450 위치까지(또한 서열번호: 41의 1 위치로부터 약 450 위치까지) 이르는 FabB(β -케토아실-ACP 신타제)에 Pfam 매치를 가지고 있다. Th. 23B OrfA-KS 도메인은 DXAC^{*}(^{*}아실 결합 부위 C₂₀₇)을 포함한다. 또한, Th. 23B KS 영역의 말단에 있는 특징적인 모티브인 GFGG는 서열번호: 39의 453-456 위치(또한, 서열번호 41의 453-456 위치)에 존재한다. 서열번호: 39의 1-500위치에 이르는 아미노산 서열은 496개 아미노산에 걸쳐서 스킨조카이트리움 OrfA(서열번호: 2)과 약 79% 동일하다. 서열번호: 39의 1-450 위치에 이르는 아미노산 서열은 446개 아미노산에 걸쳐서 스킨조카이트리움 OrfA(서열번호: 2)와 약 81% 동일하다.

[0116] Th. 23B OrfA에 있는 두 번째 도메인은 MAT 도메인이고, 본 발명에서 Th. 23B OrfA-MAT라고도 한다. MAT 도메인 기능은 앞서 상세히 기술되었다. 상기 도메인은 서열번호: 42로 나타낸 서열번호: 38의 약 1503 위치로부터 약 3000 위치까지 이르는 뉴클레오타이드 서열 내에 포함된다. Th. 23B MAT 도메인을 포함하는 아미노산 서열은 서열번호: 43으로 나타낸, 약 501 위치로부터 약 1000 위치까지 이르는 서열번호: 39의 한 영역이다. 서열번호: 39의 이러한 영역은 서열번호: 39의 약 580 위치로부터 약 900 위치까지(서열번호: 43의 80-400) 이르는 FabD (말로닐-CoA:ACP 아실트랜스퍼라제)에 Pfam 매치를 가지고 있다. Th. 23B OrfA-MAT 도메인은 서열번호: 39의 695-699 위치에 나타낸 활성 부위 모티브: GHS^{*}XG(^{*}아실 결합 부위 S₆₉₇)을 포함한다. 서열번호: 39의 501-1000 위치에 이르는 아미노산 서열은 481개 아미노산에 걸쳐서 스킨조카이트리움 OrfA(서열번호: 2)과 약 46% 동일하다. 서열번호: 39의 580-900 위치에 이르는 아미노산 서열은 333개 아미노산에 걸쳐서 스킨조카이트리움 OrfA(서열번호: 2)와 약 50% 동일하다.

[0117] Th. 23B OrfA에 있는 3-10번째 도메인은 여덟 개의 직렬 연결된 ACP 도메인이고, 또한 Th. 23B OrfA-ACP라고도 한다(서열에서 첫 번째 도메인은 OrfA-ACP1, 두 번째 도메인은 OrfA-ACP2, 세 번째 도메인은 OrfA-ACP3이다). ACP 도메인 기능은 앞서 상세히 기술되었다. 첫 번째 Th. 23B ACP 도메인인 Th. 23B OrfA-ACP1은 서열번호: 44로 나타낸 서열번호: 38(OrfA)의 약 3205 위치로부터 약 3555 위치까지 이르는 뉴클레오타이드 서열 내에 포함된다. Th. 23B ACP 도메인을 포함하는 아미노산 서열은 서열번호: 45로 나타낸, 서열번호: 39의 약 1069 위치로부터 약 1185 위치까지 이르는 서열번호: 39의 한 영역이다. 서열번호: 39의 1069-1185 위치에 이르는 아미노산 서열은 85개 아미노산에 걸쳐서 스킨조카이트리움 OrfA(서열번호: 2)와 약 65% 동일하다. Th. 23B OrfA-ACP1은 스킨조카이트리움 OrfA에 있는 아홉 개의 ACP 도메인과 유사한 동일성을 가지고 있다.

[0118] Th. 23B OrfA에 있는 여덟 개의 ACP 도메인은 서로 인접해있고 (서열번호: 46으로 나타낸) 포스포판테테인 결합 부위 모티브 LGXDS^{*}의 존재에 의해 동정될 수 있다. 여기서, S^{*}는 포스포판테테인 부착 부위이다. 서열번호: 39에 대하여 여덟 개의 S^{*} 부위 각각의 아미노산 위치는 1128(ACP1), 1244(ACP2), 1360(ACP3), 1476(ACP4), 1592(ACP5), 1708(ACP6), 1824(ACP7) 및 1940(ACP8)이다. 여덟 개의 모든 Th. 23B ACP 도메인의

뉴클레오타이드 및 아미노산 서열은 매우 보존적이고, 따라서 각 도메인의 서열은 개별적인 서열 식별자에 의해 나타내지 않았다. 그러나 본 발명에서 개시된 정보를 기반으로 당업자라면 서열번호: 38 및 서열번호: 39에 있는 다른 일곱 개의 ACP 도메인 각각을 포함하는 서열을 용이하게 결정할 수 있다.

[0119] 여덟 개의 모든 Th. 23B ACP 도메인은 서열번호: 38의 약 3205 위치로부터 약 5994 위치까지 이르는 Th. 23B OrfA의 영역에 이르며, 이는 서열번호: 39의 약 1069로부터 약 1998 위치까지의 아미노산에 대응된다. 여덟 개의 모든 도메인을 포함하는 전체 ACP 영역의 뉴클레오타이드 서열은 서열번호: 47에 나타내었다. 서열번호: 47은 서열번호: 48로 나타낸 아미노산 서열을 코딩한다. 서열번호: 48은 각각의 ACP 도메인 간의 링커 절편을 포함한다. 여덟 개 도메인의 반복적인 간격은 서열번호: 48의 약 116개의 아미노산마다이고 각 도메인은 (상기된) 활성 부위 모티프를 중심으로 약 116개의 아미노산으로 구성된 것으로 여겨진다. 스킨조카이트리움의 OrfA에 있는 아홉 개의 근접한 ACP 도메인 사이에 있는 링커 영역은 프롤린과 알라닌 잔기가 매우 풍부한 반면에 트라우스토카이트리움의 OrfA에 있는 여덟 개의 근접한 ACP 영역 사이에 있는 링커 영역은 세린 잔기(프롤린 또는 알라닌 잔기가 아님)가 풍부하다.

[0120] Th. 23B OrfA에 있는 마지막 도메인은 KR 도메인이고, 본 발명에서 Th. 23B OrfA-KR이라고 한다. KR 도메인 기능은 앞서 상세히 기술되었다. 상기 도메인은 서열번호: 49로 나타낸 서열번호: 38의 약 6001 위치 내지 약 8433 위치까지 이르는 뉴클레오타이드 서열 내에 포함된다. Th. 23B KR 도메인을 포함하는 아미노산 서열은 서열번호: 50으로 나타낸, 서열번호: 39의 약 2001 위치로부터 약 2811 위치까지 이르는 서열번호: 39의 한 영역이다. 서열번호: 39의 이러한 영역은 서열번호: 39의 약 2300 위치로부터 약 2550 위치까지(서열번호: 50의 300-550 위치) 이르는 FabG (β -케토아실-ACP 환원효소)에 Pfam 매치를 가지고 있다. 서열번호: 39의 2001-2811 위치에 이르는 아미노산 서열은 831개 아미노산에 걸쳐서 스킨조카이트리움 OrfA(서열번호: 2)와 약 40% 동일하다. 서열번호: 39의 2300-2550 위치에 이르는 아미노산 서열은 235개 아미노산에 걸쳐서 스킨조카이트리움 OrfA(서열번호: 2)와 약 51% 동일하다.

[0121] Th. 23B 오픈 리딩 프레임 B(OrfB):

[0122] Th. 23B OrfB의 완전한 뉴클레오타이드 서열은 서열번호: 51에 나타내었다. 서열번호: 51은 Th. 23B OrfB에 있는 다음 도메인을 코딩한다: (a) 하나의 β -케토아실-ACP 신타제 (KS) 도메인; (b) 하나의 쇠길이인자 (CLF) 도메인; (c) 하나의 아실트랜스퍼라제 (AT) 도메인 및 (d) 하나의 엔오일-ACP 환원효소 (ER) 도메인. 상기 도메인 조직화는 스킨조카이트리움 단백질의 AT와 ER 사이에 있는 링커 영역이 Th.23B 보다 약 50-60개 아미노산 정도 더 길다는 점을 제외하고는 스킨조카이트리움 OrfB(서열번호: 3)과 동일하다. 또한, 스킨조카이트리움에 있는 이러한 링커 영역은 세린 잔기가 매우 풍부한 특정한 영역을 가지고 있는데 반하여(15개의 근접한 세린 잔기 뿐만 아니라 상기 영역 내에 있는 다른 세린도 포함된다) Th. 23B OrfB에 있는 상응하는 링커 영역은 세린 잔기가 풍부하지 않다. AT/ER 링커 영역에 있는 이러한 차이점은 상기 영역의 시작부에서 스킨조카이트리움 OrfB와 Th.23B OrfB 간의 정렬에서 분리를 설명할 것이다.

[0123] Th. 23B OrfB는 서열번호: 52에 나타낸 1935 아미노산 서열을 코딩하는 (정지 코돈을 포함하지 않는) 5805 뉴클레오타이드 서열이다. Th. 23B OrfB 아미노산 서열(서열번호: 52)은 표준 BLAST 조사로 공지 서열과 비교되었다(BLAST 변수: Blastp, 낮은 복잡성 필터 오프, 프로그램-BLOSUM62, 캡-cost-존재:11, 신장:1)(Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res. 25: 3389-3402; 이의 전문은 본 발명에서 참조로 인용된다). 아미노산 레벨에서, Th. 23B OrfB에 대하여 가장 높은 상동성을 가진 서열은 대부분의 OrfB에 걸쳐서 스킨조카이트리움 OrfB(gb AAK72880.1)(서열번호: 4)와 마지막 도메인에 걸쳐서 스킨조카이트리움 OrfC(gb AAK72881.1)(서열번호: 6)(정렬은 상기된 바와 같이 2조각으로 분리된다)이었다. 서열번호: 52는 먼저 서열번호: 4와 (KS, CLF 및 AT 도메인을 포함한) 10 내지 1479 위치에서 정렬되고 1483개의 아미노산에 걸쳐서 서열번호: 4에 대하여 52%의 서열 동일성을 보인다. 서열번호: 52는 또한 서열번호: 6과 (ER 도메인을 포함한) 1491 내지 1935 위치에서 정렬되고 448개의 아미노산에 걸쳐서 서열번호: 4에 대하여 64%의 서열 동일성을 보인다.

[0124] Th.23B OrfB에 있는 첫 번째 도메인은 KS 도메인이고, 본 발명에서 Th.23B OrfB-KS라고도 한다. KS 도메인 기능은 앞서 상세히 기술되었다. 상기 도메인은 서열번호: 53으로 나타낸 서열번호: 51의 약 1 위치로부터 약 1500 위치까지 이르는 뉴클레오타이드 서열 내에 포함된다. Th. 23B KS 도메인을 포함하는 아미노산 서열은 서열번호: 54로 나타낸, 서열번호: 52의 약 1 위치로부터 약 500 위치까지 이르는 서열번호: 52의 한 영역이다. 서열번호: 52의 이러한 영역은 약 1 위치로부터 약 450 위치까지(서열번호: 54의 1-450) 이르는 FabB

(β -케토아실-ACP 신타제)에 Pfam 매치를 가지고 있다. Th.23B OrfB-KS 도메인은 활성 부위 모티브: DXAC*를 포함한다. 여기서, C*는 아실기 부착 부위이고 C*는 서열번호: 52의 201 위치에 있다. KS 영역의 말단에 있는 특징적인 모티브인 GFGG는 서열번호: 52의 434-437 위치의 아미노산에 존재한다. 서열번호: 52의 1-500 위치에 이르는 아미노산 서열은 500개 아미노산에 걸쳐서 스킨조카이트리움 OrfB(서열번호: 4)와 약 64% 동일하다. 서열번호: 52의 1-450 위치에 이르는 아미노산 서열은 442개 아미노산에 걸쳐서 스킨조카이트리움 OrfB(서열번호: 4)와 약 67% 동일하다.

[0125] Th.23B OrfB에 있는 두 번째 도메인은 CLF 도메인이고, 본 발명에서 Th.23B OrfB-CLF라고도 한다. CLF 도메인 기능은 앞서 상세히 기술되었다. 상기 도메인은 서열번호: 55로 나타낸 서열번호: 51(OrfB)의 약 1501 위치로부터 약 3000 위치까지 이르는 뉴클레오타이드 서열 내에 포함된다. CLF 도메인을 포함하는 아미노산 서열은 서열번호: 56으로 나타낸, 서열번호: 52의 약 501 위치로부터 약 1000 위치까지 이르는 서열번호: 52의 한 영역이다. 서열번호: 52의 이러한 영역은 약 550 위치로부터 약 910 위치까지(서열번호: 56의 50-410) 이르는 FabB (β -케토아실-ACP 신타제)에 Pfam 매치를 가지고 있다. CLF 도메인은 KS 단백질에 상동성을 가짐에도 불구하고 KS 단백질에서 아실기가 부착되는 활성 부위 시스템이 결여되었다. 서열번호: 52의 501-1000 위치에 이르는 아미노산 서열은 517개 아미노산에 걸쳐서 스킨조카이트리움 OrfB(서열번호: 4)와 약 49% 동일하다. 서열번호: 52의 550-910 위치에 이르는 아미노산 서열은 360개 아미노산에 걸쳐서 스킨조카이트리움 OrfB(서열번호: 4)와 약 54% 동일하다.

[0126] Th.23B OrfB에 있는 세 번째 도메인은 AT 도메인이고, 본 발명에서 Th.23B OrfB-AT라고 한다. AT 도메인 기능은 앞서 기술되었다. 상기 도메인은 서열번호: 57로 나타낸, 서열번호: 51(Th.23B OrfB)의 약 3001 위치로부터 약 4500 위치까지 이르는 뉴클레오타이드 서열 내에 포함된다. Th.23B AT 도메인을 포함하는 아미노산 서열은 서열번호: 58로 나타낸, 서열번호: 52의 약 1001 위치로부터 약 1500 위치까지 이르는 서열번호: 52의 한 영역이다. 서열번호: 52의 이러한 영역은 약 1100 위치로부터 약 1375 위치까지(서열번호: 58의 100-375) 이르는 FabD (말로닐-CoA:ACP 아실트랜스퍼라제)에 Pfam 매치를 가지고 있다. PUFA 신타제의 이러한 AT 도메인은 MAT 단백질과 상동성을 가짐에도 불구하고, MAT의 연장된 모티브(중요 아르기닌과 글루타민 잔기)가 부족하며 말로닐-CoA 전달에 관여하지 않을 것으로 여겨진다. 아실트랜스퍼라제의 GX^{*}XG 모티브가 존재하되, S^{*}는 아실 부착 부위이고 서열번호: 52에 대하여 1123 위치에 존재한다. 서열번호: 52의 1001-1500 위치에 이르는 아미노산 서열은 459개 아미노산에 걸쳐서 스킨조카이트리움 OrfB(서열번호: 4)와 약 44% 동일하다. 서열번호: 52의 1100-1375 위치에 이르는 아미노산 서열은 283개 아미노산에 걸쳐서 스킨조카이트리움 OrfB(서열번호: 4)와 약 45% 동일하다.

[0127] Th.23B OrfB에 있는 네 번째 도메인은 ER 도메인이고, 본 발명에서 Th.23B OrfB-ER이라고 한다. ER 도메인 기능은 앞서 상세히 기술되었다. 상기 도메인은 서열번호: 59로 나타낸 서열번호: 51(OrfB)의 약 4501 위치로부터 약 5805 위치까지 이르는 뉴클레오타이드 서열 내에 포함된다. Th.23B ER 도메인을 포함하는 아미노산 서열은 서열번호: 60으로 나타낸, 서열번호: 52의 약 1501 위치로부터 약 1935 위치까지 이르는 서열번호: 52의 한 영역이다. 서열번호: 52의 이러한 영역은 약 1501 위치로부터 약 1810 위치까지(서열번호: 60의 1-310 위치) 이르는 2-니트로프로판 디옥시게나제에 관련된 디옥시게나제 군에 Pfam 매치를 가지고 있다. 상기 도메인이 ER로서 기능을 발휘한다는 것은 스트렙토코쿠스 뉴모니에 유래의 새롭게 특정된 ER 효소에 대한 상동성 때문인 것으로 추정된다. 서열번호: 52의 1501-1935 위치에 이르는 아미노산 서열은 433개 아미노산에 걸쳐서 스킨조카이트리움 OrfB(서열번호: 4)와 약 66% 동일하다. 서열번호: 52의 1501-1810 위치에 이르는 아미노산 서열은 305개 아미노산에 걸쳐서 스킨조카이트리움 OrfB(서열번호: 4)와 약 70% 동일하다.

[0128] Th. 23B 오픈 리딩 프레임 C(OrfC):

[0129] Th. 23B OrfC의 완전한 뉴클레오타이드 서열을 서열번호: 61에 나타내었다. 서열번호: 61은 Th. 23B OrfC에 있는 다음 도메인을 코딩한다: (a) 두 개의 FabA 유사 β -하이드록시아실-ACP 탈수소효소(DH) 도메인, 둘 모두는 FabA 단백질(트랜스-2-데케노일-ACP의 합성과 상기 생산물을 시스-3-데케노일-ACP로 가역적인 이성질화를 촉매하는 효소)에 대하여 상동성을 가진다; (b) 스킨조카이트리움 OrfB의 ER 도메인에 대하여 높은 상동성을 가진 하나의 엔오일-ACP 환원효소 (ER) 도메인. 상기 도메인 조직화는 스킨조카이트리움 OrfC(서열번호: 5)에 있는 바와 동일하다.

[0130] Th. 23B OrfC는 서열번호: 62에 나타낸 1470 개의 아미노산 서열을 코딩하는 (정지 코돈을 포함하지 않는) 4410개의 뉴클레오타이드 서열이다. Th. 23B OrfC(서열번호: 62)는 표준 BLAST 조사로 공지된 서열과 비

교되었다(BLAST 변수: Blastp, 낮은 복잡성 필터 오프, 프로그램- BLOSUM62, 갭-cost-존재:11, 신장:1)(Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res. 25: 3389-3402; 이의 전문은 본 발명에서 참조로 인용된다). 아미노산 레벨에서 Th. 23B OrfC에 대하여 가장 높은 상동성을 가진 서열은 스킨조카이트리움 OrfC(gb AAK728881.1)(서열번호: 6)이다. 서열번호: 62는 스킨조카이트리움 OrfC(서열번호: 6)과 66% 동일하다.

[0131]

Th. 23B OrfC에 있는 첫 번째 도메인은 DH 도메인이고, 본 발명에서 Th. 23B OrfC-DH1이라고 한다. DH 도메인 기능은 앞서 상세히 기술되었다. 상기 도메인은 서열번호: 63으로 나타낸 서열번호: 61(OrfC)의 1 위치로부터 약 1500 위치까지 이르는 뉴클레오타이드 서열 내에 포함된다. Th.23B DH1 도메인을 포함하는 아미노산 서열은 서열번호: 64로 나타낸, 서열번호: 62의 약 1 위치로부터 약 500 위치까지 이르는 서열번호: 62의 한 영역이다. 서열번호: 62의 이러한 영역은 약 275 위치로부터 약 400 위치까지(서열번호: 64의 275-400 위치) 이르는 FabA에 Pfam 매치를 가지고 있다. 서열번호: 62의 1-500 위치에 이르는 아미노산 서열은 526개 아미노산에 걸쳐서 스킨조카이트리움 OrfC(서열번호: 6)와 약 66% 동일하다. 서열번호: 62의 275-400 위치에 이르는 아미노산 서열은 126개 아미노산에 걸쳐서 스킨조카이트리움 OrfC(서열번호: 6)와 약 81% 동일하다.

[0132]

Th. 23B OrfC에 있는 두 번째 도메인은 DH 도메인이고, 본 발명에서 Th. 23B OrfC-DH2라고도 한다. 이는 OrfC에 있는 두 개의 DH 도메인 중 두 번째이고 따라서 DH2라고 명명된다. 상기 도메인은 서열번호: 65로 나타낸 서열번호: 61(OrfC)의 약 1501로부터 내지 약 3000 위치까지 이르는 뉴클레오타이드 서열내에 포함된다. Th. 23B DH2 도메인을 포함하는 아미노산 서열은 서열번호: 66에 나타낸, 서열번호 62의 약 501 위치로부터 약 1000 위치까지 이르는 서열번호: 62의 영역이다. 서열번호: 62의 이러한 영역은 상기된 바와 같이 약 800 내지 약 925 위치에 이르는 FabA에 대한 Pfam 매치를 가지고 있다(서열번호: 66의 300-425 위치). 서열번호: 62의 501-1000 위치에 이르는 아미노산 서열은 518개의 아미노산에 걸쳐서 스킨조카이트리움 OrfC(서열번호: 6)와 약 56% 동일하다. 서열번호: 62의 800-925에 이르는 아미노산 서열은 124개의 아미노산에 걸쳐서 스킨조카이트리움 OrfC(서열번호: 6)와 약 58% 동일하다.

[0133]

Th. 23B OrfC에 있는 세 번째 도메인은 ER 도메인이고, 본 발명에서 Th. 23B OrfC-ER이라고 한다. ER 도메인 기능은 앞서 상세히 기술되었다. 상기 도메인은 서열번호: 67에 나타낸 서열번호: 61(OrfC)의 약 3001 위치로부터 약 4410 위치까지 이르는 뉴클레오타이드 서열내에 포함된다. Th. 23B ER 도메인을 포함하는 아미노산 서열은 서열번호 68에 나타낸, 서열번호 62의 약 1001 위치로부터 약 1470 위치까지 이르는 서열번호: 62의 한 영역이다. 서열번호: 62의 이러한 영역은 상기된 바와 같이 약 1025 위치로부터 약 1320 위치(서열번호: 68의 25-320 위치)에 이르는 2-니트로프로판 디옥시게나제에 관련된 디옥시게나제에 대한 Pfam 매치를 가지고 있다. ER로서의 이러한 도메인의 기능은 스트렙토코쿠스 뉴모니에 유래의 새롭게 특정된 ER 효소에 대한 상동성으로 인하여 추정된다. 서열번호: 62의 1001-1470 위치에 이르는 아미노산 서열은 474개 아미노산에 걸쳐서 스킨조카이트리움 OrfB(서열번호: 4)와 75% 동일하다. 서열번호: 62의 1025-1320 위치에 이르는 아미노산 서열은 296개 아미노산에 걸쳐서 스킨조카이트리움 OrfB(서열번호: 4)와 약 81% 동일하다.

[0134]

본 발명의 한 양태는 비박테리아의 PUFA PKS 시스템 유래의 분리된 단백질 또는 도메인, 이의 상동체 및/또는 이의 단편에 관한 것이다. 또한, 본 발명은 본 발명에서 언급된(하기에서 상세히 설명되는) 단백질, 도메인 또는 펩타이드 중 어느 하나를 코딩하는 분리된 핵산 분자를 포함한다. 본 발명에서 PUFA PKS 시스템 유래의 단백질 또는 펩타이드와 같은 분리된 단백질 또는 펩타이드는 천연 환경으로부터 제거되어진(즉, 인간의 조작이 이루어진) 단백질 또는 (폴리펩타이드 또는 펩타이드를 포함한) 이의 단편이고 예를 들면 정제된, 부분적으로 정제된, 재조합으로 제조된 단백질 및 합성으로 제조된 단백질을 포함한다. 이에 "분리된"은 단백질이 정제된 정도를 반영하지 않는다. 바람직하게는 본 발명에 따른 분리된 단백질은 재조합으로 제조된 것이다. 분리된 펩타이드는 합성(예, 펩타이드 합성에 의한 것과 같이 화학적으로) 또는 재조합으로 제조될 수 있다. 또한, 예를 들어 "트라우스토카이트리움 PUFA PKS 단백질"은 트라우스토카이트리움 미생물 유래의 (일반적으로 PUFA PKS 단백질의 상동체를 포함하는) PUFA PKS 단백질 또는 트라우스토카이트리움 유래의 천연적으로 존재하는 PUFA PKS 단백질의 공지된 구조(예, 서열)와 혹은 기능으로부터 제조된 PUFA PKS 단백질을 의미한다. 즉, 트라우스토카이트리움 PUFA PKS 단백질은 트라우스토카이트리움 유래의 천연적으로 존재하는 PUFA PKS 단백질의 실질적으로 유사한 구조와 기능을 가진 PUFA PKDS 단백질 또는 상기된 트라우스토카이트리움 유래의 천연적으로 존재하는 PUFA PKS 단백질의 생물학적으로 활성인(즉, 생물학적 활성을 가지는) 상동체를 포함한다. 이에, 트라우스토카이트리움 PUFA PKS 단백질은 정제된, 부분적으로 정제된, 재조합, 돌연변이된/변형된 및 합성된 단백질을 포함할 수 있다. 동일한 정의가 스킨조카이트리움 또는 다른 미생물 유래의 PUFA PKS 단백질 및 도메인과

같은 본 발명에서 언급된 다른 단백질 또는 펩타이드에 적용될 수 있다.

[0135]

본 발명에서 용어 "변형" 및 "돌연변이"는 본 발명에 언급된 단백질 또는 펩타이드의 일차 아미노산 서열 (또는 핵산 서열)에 대한 변형/돌연변이의 경우에는 상호 교환적으로 사용될 수 있다. 용어 "변형"은 이에 제한되는 것은 아니지만, 메틸화, 파르네실화, 카르복시메틸화, 게라닐 게라닐화, 당쇄화, 인산화, 아세틸화, 미리스틸화, 프레닐화, 팔미토일화 및/또는 아미드화를 포함한 단백질 또는 펩타이드에 대한 해독후 변형을 정의하는데 사용될 수 있다. 변형은 또한 예를 들면 단백질을 다른 화합물과 복합체화 (complexing)하는 것을 포함할 수 있다. 그러한 변형은 예를 들면 천연적인 야생형 단백질 또는 펩타이드에 존재하는 해독후 변형과 상이하다면 돌연변이로서 여겨질 수 있다.

[0136]

본 발명에서 용어 "상동체"는 천연적으로 존재하는 단백질 또는 펩타이드에 대한 하나 이상의 미미한 변형 또는 돌연변이에 의해 상기 천연적으로 존재하는 단백질 또는 펩타이드 (즉, "프로토타입" 또는 "야생형" 단백질)와 상이하지만, 천연적으로 존재하는 형태의 측쇄(side chain) 구조와 전반적인 기본 단백질을 유지하는 (즉, 그로 인하여 상동체는 야생형 단백질과 관련된 것으로 동정될 수 있다) 단백질을 의미하는 데 이용된다. 그러한 변화는 이에 제한되는 것은 아니지만, 하나 또는 수개의 아미노산 측쇄에서 변화; 결실 (예를 들면 단백질 또는 펩타이드의 절단된 형태), 삽입 및/또는 치환을 포함한 하나 또는 몇 개 아미노산의 변화; 하나 또는 몇 개 원소의 입체화학에서의 변화; 및/또는 이에 제한되는 것은 아니지만, 메틸화, 파르네실화, 게라닐 게라닐화, 당쇄화, 카르복시메틸화, 인산화, 아세틸화, 미리스틸화, 프레닐화, 팔미토일화 및/또는 아미드화를 포함한 미미한 파생화를 포함한다. 상동체는 천연적으로 존재하는 단백질 또는 펩타이드와 비교하였을 때 증진된, 감소된 또는 실질적으로 유사한 성질을 가진다. PUFA PKS 단백질 또는 도메인의 바람직한 상동체가 하기에서 상세히 설명된다. 상동체는 합성으로 제조된 상동체, 주어진 단백질 또는 도메인의 천연적으로 존재하는 대립형질 변이체, 또는 대상 서열이 유래된 유기체 이외의 유기체로부터 유래된 상동성의 서열을 포함할 수 있다.

[0137]

보존적인 치환은 전형적으로 하기 그룹 내에서의 치환을 포함한다: 글리신 및 알라닌; 발린, 이소루신 및 루신; 아스파르트산, 글루탐산, 아스파라긴 및 글루타민; 세린 및 트레오닌; 리신 및 아르기닌; 그리고 페닐알라닌 및 티로신. 치환은 보존적인 소수성 또는 친수성을 기반으로 (Kyte and Doolittle, *J. Mol. Biol.* (1998) 157: 105-132) 또는 유사한 폴리펩타이드 이차 구조를 예상하는 능력을 기반으로 (Chou and Fasman, *Adv. Enzymol.* (1978) 47: 45-148, 1978) 이루어질 수 있다.

[0138]

상동체는 천연의 대립형질 변이 또는 천연의 돌연변이의 결과일 수 있다. 단백질을 코딩하는 핵산의 천연적으로 발생하는 대립형질 변이체는 상기 단백질을 코딩하는 유전자와 게놈상에서 실질적으로 동일한 위치에서 존재하는 유전자이지만 예를 들면 돌연변이 또는 재조합에 의해 유발된 천연의 변이에 의해 유사하지만 동일하지 않은 서열을 갖게 된다. 대립형질 변이체는 비교 대상이 되는 유전자에 의해 코딩된 단백질의 활성과 유사한 활성을 갖는다. 한 부류의 대립형질 변이체는 유전자의 코드의 축퇴성으로 인하여 상이한 핵산 서열을 갖지만 동일한 단백질을 코딩할 수도 있다. 대립형질 변이체는 또한 유전자의 5' 또는 3' 비해독 영역(예, 조절 영역)에서의 변이를 포함할 수도 있다. 대립형질 변이체는 또한 당업계에 공지되어 있다.

[0139]

상동체는 이에 제한되는 것은 아니지만 분리된, 천연적으로 발생한 단백질에 대한 직접적인 변형, 직접 단백질 합성 또는 무작위적 또는 표적화 돌연변이를 일으킬 수 있는 예를 들면 전통적인 또는 재조합 DNA 기술을 사용하여 단백질을 코딩하는 핵산 서열에 변형을 가하는 방법을 포함한 단백질의 제조 방법과 관련하여 공지 기술을 사용하여 제조할 수 있다.

[0140]

야생형 단백질과 비교했을 때 단백질 상동체에서의 변형은 천연적으로 발생한(야생형) 단백질과 비교했을 때 상동체의 기본적인 생물학적 활성을 증가, 감소 또는 실질적으로 변화를 일으키지 않을 수 있다. 일반적으로 단백질의 생물학적 활성 또는 생물학적 작용은 생체내(즉, 단백질의 천연의 생리학적 환경) 또는 시험관내(즉, 실험실 조건 하에서)에서 측정되거나 관측된 천연적으로 발생하는 형태의 단백질에 의해 보여지고 수행되는 기능을 의미한다. PUFA PKS 시스템과 PUFA PKS 시스템을 이루는 각각의 단백질/도메인의 생물학적 활성이 본 발명에서 상세히 언급되어 있다. 상동체 또는 (후술하는) 모방체와 같은 단백질의 변형은 천연적으로 발생하는 단백질과 동일한 생물학적 활성을 갖는 단백질 또는 천연적으로 발생하는 단백질과 비교했을 때 감소된 또는 증가된 생물학적 활성을 갖는 단백질을 생성시킬 수 있다. 단백질의 발현을 감소시키거나 단백질의 활성을 감소시키는 변형은 단백질의 (완전한 또는 부분적인) 불활성화, 하향 조정 또는 단백질의 감소된 작용 (또는 활성)을 의미한다. 유사하게 단백질 발현을 증가시키거나 단백질의 활성을 증가시키는 변형은 증폭, 과잉생산, 활성화, 향상, 상향 조정 또는 단백질의 증가된 작용 (또는 활성)을 의미한다. 야생형 단백질의 생물학적 활성

을 가지는 상동체란 야생형 단백질과, 특히 생물학적 활성의 정도에 있어서 동일하다는 것을 필수적으로 의미하지 않는다. 그보다 상동체는 야생형 단백질과 비교했을 때 감소 또는 증가된 정도로 야생형 단백질과 동일한 생물학적 활성을 수행할 수 있다. PUFA PKS 시스템의 기능적 도메인은 생물학적 기능을 수행할 수 있는 (즉, 생물학적 활성을 가지는) 도메인 (즉, 도메인은 단백질의 일부분일 수 있다)이다.

[0141]

PUFA PKS 단백질 또는 도메인의 생물학적 활성을 검출하고 측정하는 방법은 이에 제한되는 것은 아니지만 PUFA PKS 단백질 또는 도메인의 전사량의 측정, PUFA PKS 단백질 또는 도메인의 해독량의 측정, PUFA PKS 단백질 또는 도메인의 해독 후 변형의 측정, PUFA PKS 단백질 또는 도메인의 효소 활성의 측정 및/또는 PUFA PKS 시스템의 하나 이상의 생산물(예, PUFA 생산량)의 생산량 측정을 포함한다. (상동체를 포함한) 본 발명에 따른 분리된 단백질은 야생형 단백질의 생물학적 활성을 가질 것을 반드시 요구받지 않는다. 예를 들어 PUFA PKS 단백질 또는 도메인은 절단된, 돌연변이된 또는 불활성화 단백질일 수 있다. 그러한 단백질을 스크리닝 에세이 또는 예를 들어 항체 제조와 같은 다른 목적에 유용하다. 바람직한 양태에서 본 발명의 분리된 단백질은 야생형 단백질과 (상기된 바와 같이 반드시 동등하지 않을지라도) 유사한 생물학적 활성을 가진다.

[0142]

단백질 발현을 측정하는 방법은 이에 제한되는 것은 아니지만 웨스턴 블롯, 면역블롯, 효소결합면역흡착 분석법(ELISA), 방사면역측정법(RIA), 면역침전법, 표면 플라즈몬 공명, 화학발광, 형광 편광, 야광(phosphorescence), 면역조직화학 분석법, 매트릭스 보조된 레이저 탈락/이온화 타임 오프 플레이트(MALDI-TOF) 거대질량분석기, 미세포분류, 마이크로에레이, 현미경, 형광 활성화된 유세포 분석(FACS) 및 유세포 분류 뿐만 아니라 이에 제한되는 것은 아니지만 효소의 활성 또는 다른 단백질 파트너와의 결합을 포함한 단백질의 성질을 기반으로 한 분석법을 포함할 수 있다. 결합 분석법이 또한 당업계에 공지되어 있다. 예를 들어, BIAcore 기계가 두 단백질간 복합체의 결합 상수를 측정하는데 사용될 수 있다. 상기 복합체에 대한 분해 상수는 완충용액이 칩 상에 통과하는 시간에 대하여 굴절율에서의 변화를 모니터링함으로써 측정될 수 있다(O'Shannessy et al. *Anal. Biochem.* 212:457-468 (1993), Schuster et al. *Nature* 365:343-347 (1993)). 한 단백질과 다른 단백질 간의 결합을 측정하는 적합한 다른 분석법은 예를 들면 효소결합면역흡착(ELISA) 및 방사면역측정법(RIA)과 같은 면역분석법; 형광, UV 흡수, 원편광분광분석 또는 핵자기공명(NMR)을 통한 단백질의 분광학적 및 광학적 성질에서의 변화를 모니터링함으로써 결합의 측정을 포함한다.

[0143]

한 양태에서, 본 발명은 (a) 서열번호: 39, 서열번호: 52, 서열번호: 62 및 이의 생물학적으로 활성인 단편으로 이루어진 그룹으로부터 선택된 아미노산 서열; (b) 서열번호: 41, 서열번호: 43, 서열번호: 45, 서열번호: 48, 서열번호: 50, 서열번호: 54, 서열번호: 56, 서열번호: 58, 서열번호: 60, 서열번호: 64, 서열번호: 66, 서열번호: 68 및 이의 생물학적으로 활성인 단편으로 이루어진 그룹으로부터 선택된 아미노산 서열; (c) (a)의 아미노산 서열의 적어도 500개의 연속적인 아미노산 서열과 적어도 약 60% 동일하되, 다중불포화 지방산(PUFA) 폴리케타이드 신타제(PKS) 시스템의 적어도 하나의 도메인의 생물학적 활성을 가지는 아미노산 서열; 및/또는 (d) (b)의 아미노산 서열과 적어도 약 60% 동일하되, 다중불포화 지방산(PUFA) 폴리케타이드 신타제(PKS) 시스템의 적어도 하나의 도메인의 생물학적 활성을 가지는 아미노산 서열로 이루어진 그룹으로부터 선택된 아미노산 서열을 포함하는 분리된 단백질에 관한 것이다. 추가의 양태에서, PUFA PKS 도메인 중 몇 개와 관련하여 상기된 활성 부위 도메인 또는 다른 기능적 모티프를 포함한 아미노산 서열이 본 발명에 의해 포함된다. 한 양태에서, 상기된 아미노산 서열은 서열번호: 2, 서열번호: 4, 서열번호: 6, 서열번호: 8, 서열번호: 10, 서열번호: 13, 서열번호: 18, 서열번호: 20, 서열번호: 22, 서열번호: 24, 서열번호: 26, 서열번호: 28, 서열번호: 30, 서열번호: 32 중 어느 하나의 아미노산 서열을 포함하지 않는다.

[0144]

본 발명의 한 측면에서, 본 발명에서 언급된 특정한 PUFA PKS 단백질 또는 도메인의 상동체를 포함하여, 본 발명에 포함되는 PUFA PKS 단백질 또는 도메인은 서열번호: 39, 서열번호: 52 또는 서열번호: 62로부터 선택된 아미노산 서열의 적어도 500개의 연속적인 아미노산과 적어도 약 60% 동일한 아미노산 서열을 포함하되, 상기 아미노산 서열은 PUFA PKS 시스템의 적어도 하나의 도메인의 생물학적 활성을 가지는 아미노산 서열을 포함한다. 추가 측면에서, 상기 단백질의 아미노산 서열은 서열번호: 39, 서열번호: 52 또는 서열번호: 62의 적어도 약 600개의 연속적인 아미노산, 보다 바람직하게는 적어도 약 700개의 연속적인 아미노산, 보다 바람직하게는 적어도 약 800개의 연속적인 아미노산, 보다 바람직하게는 적어도 약 900개의 연속적인 아미노산, 보다 바람직하게는 적어도 약 1000개의 연속적인 아미노산, 보다 바람직하게는 적어도 약 1100개의 연속적인 아미노산, 보다 바람직하게는 적어도 약 1200개의 연속적인 아미노산, 보다 바람직하게는 적어도 약 1300개의 연속적인 아미노산, 보다 바람직하게는 적어도 약 1400개의 연속적인 아미노산 또는 서열번호: 62의 전장에 대하여 적어도 약 60% 동일하다. 추가의 측면에서, 상기 단백질의 아미노산 서열은 서열번호: 39 또는 서열번호: 52의 적어도 약 1500개의 연속적인 아미노산, 보다 바람직하게는 적어도 약 1600개의 연속적인 아미노산, 보다

바람직하게는 적어도 약 1700개의 연속적인 아미노산, 보다 바람직하게는 적어도 약 1800개의 연속적인 아미노산, 보다 바람직하게는 적어도 약 1900개의 연속적인 아미노산 또는 서열번호: 52의 전장에 대하여 적어도 약 60% 동일하다. 추가의 측면에서, 상기 단백질의 아미노산 서열은 서열번호: 39의 적어도 약 2000개의 연속적인 아미노산, 바람직하게는 적어도 약 2100개의 연속적인 아미노산, 보다 바람직하게는 적어도 약 2200개의 연속적인 아미노산, 보다 바람직하게는 적어도 약 2300개의 연속적인 아미노산, 보다 바람직하게는 적어도 약 2400개의 연속적인 아미노산, 보다 바람직하게는 적어도 약 2500개의 연속적인 아미노산, 보다 바람직하게는 적어도 약 2600개의 연속적인 아미노산, 보다 바람직하게는 적어도 약 2700개의 연속적인 아미노산, 보다 바람직하게는 적어도 약 2800개의 연속적인 아미노산, 보다 더 바람직하게는 서열번호 39의 전장에 대하여 적어도 약 60% 동일하다. 한 양태에서, 상기 아미노산 서열은 서열번호: 2, 서열번호: 4, 서열번호: 6, 서열번호: 8, 서열번호: 10, 서열번호: 13, 서열번호: 18, 서열번호: 20, 서열번호: 22, 서열번호: 24, 서열번호: 26, 서열번호: 28, 서열번호: 30, 서열번호: 32 중 하나의 아미노산 서열을 포함하지 않는다.

[0145] 다른 측면에서, 상기된 상동체를 포함하여, 본 발명에 의해 포함되는 PUFA PKS 단백질 또는 도메인은 상기 단락에서 기술된 연속적인 아미노산 길이에 걸쳐서 서열번호: 39, 서열번호: 52 또는 서열번호: 62로부터 선택된 아미노산 서열에 대하여 적어도 약 65% 동일한, 보다 바람직하게는 적어도 약 70% 동일한, 보다 바람직하게는 적어도 약 75% 동일한, 보다 바람직하게는 적어도 약 80% 동일한, 보다 바람직하게는 적어도 85% 동일한, 보다 바람직하게는 적어도 약 90% 동일한, 보다 바람직하게는 적어도 약 95% 동일한, 보다 바람직하게는 적어도 약 96% 동일한, 보다 바람직하게는 적어도 약 97% 동일한, 보다 바람직하게는 적어도 약 98% 동일한, 보다 바람직하게는 적어도 약 99% 동일하되, PUFA PKS 시스템의 적어도 하나의 생물학적 활성을 가지는 아미노산 서열을 포함한다. 한 양태에서, 상기 아미노산 서열은 서열번호: 2, 서열번호: 4, 서열번호: 6, 서열번호: 8, 서열번호: 10, 서열번호: 13, 서열번호: 18, 서열번호: 20, 서열번호: 22, 서열번호: 24, 서열번호: 26, 서열번호: 28, 서열번호: 30, 서열번호: 32 중 하나의 아미노산 서열을 포함하지 않는다.

[0146] 본 발명의 한 측면에서, 상기된 상동체를 포함하여, 본 발명에 의해 포함되는 PUFA PKS 단백질 또는 도메인은 서열번호: 39, 서열번호: 41, 서열번호: 43, 서열번호: 45, 서열번호: 48, 서열번호: 50, 서열번호: 52, 서열번호: 54, 서열번호: 56, 서열번호: 58, 서열번호: 60, 서열번호: 62, 서열번호: 64, 서열번호: 66, 서열번호: 68로부터 선택된 아미노산 서열에 대하여 적어도 약 60% 동일하되, PUFA PKS 시스템의 적어도 하나의 도메인의 생물학적 활성을 가지는 아미노산 서열을 포함한다. 추가의 측면에서, 상기 단백질의 아미노산 서열은 서열번호: 39, 서열번호: 41, 서열번호: 43, 서열번호: 45, 서열번호: 48, 서열번호: 50, 서열번호: 52, 서열번호: 54, 서열번호: 56, 서열번호: 58, 서열번호: 60, 서열번호: 62, 서열번호: 64, 서열번호: 66, 서열번호: 68로부터 선택된 아미노산 서열에 대하여 적어도 약 65% 동일한, 보다 바람직하게는 적어도 약 70% 동일한, 보다 바람직하게는 적어도 약 75% 동일한, 보다 바람직하게는 적어도 약 80% 동일한, 보다 바람직하게는 적어도 85% 동일한, 보다 바람직하게는 적어도 약 90% 동일한, 보다 바람직하게는 적어도 약 95% 동일한, 보다 바람직하게는 적어도 약 96% 동일한, 보다 바람직하게는 적어도 약 97% 동일한, 보다 바람직하게는 적어도 약 98% 동일한, 보다 바람직하게는 적어도 약 99% 동일하되, PUFA PKS 시스템의 적어도 하나의 생물학적 활성을 가진다. 한 양태에서, 상기 아미노산 서열은 서열번호: 2, 서열번호: 4, 서열번호: 6, 서열번호: 8, 서열번호: 10, 서열번호: 13, 서열번호: 18, 서열번호: 20, 서열번호: 22, 서열번호: 24, 서열번호: 26, 서열번호: 28, 서열번호: 30, 서열번호: 32 중 하나의 아미노산 서열을 포함하지 않는다.

[0147] 다른 측면에서, 상기된 상동체를 포함하여, 본 발명에 의해 포함되는 PUFA PKS 단백질 또는 도메인은 서열번호: 39, 서열번호: 43, 서열번호: 50, 서열번호: 52 및 서열번호: 58로부터 선택된 아미노산 서열에 대하여 적어도 약 50% 동일하되, PUFA PKS 시스템의 적어도 하나의 도메인의 생물학적 활성을 가지는 아미노산 서열을 포함한다. 다른 측면에서, 상기 단백질의 아미노산 서열은 서열번호: 39, 서열번호: 43, 서열번호: 50, 서열번호: 52, 서열번호: 56 및 서열번호: 58로부터 선택된 아미노산 서열에 대하여 적어도 약 55%, 보다 바람직하게는 적어도 약 60% 동일하되, PUFA PKS 시스템의 적어도 하나의 도메인의 생물학적 활성을 가진다. 추가의 측면에서 상기 단백질의 아미노산 서열은 서열번호: 39, 서열번호: 43, 서열번호: 50, 서열번호: 52, 서열번호: 54, 서열번호: 56 및 서열번호: 58로부터 선택된 아미노산 서열에 대하여 적어도 약 65% 동일하되, PUFA PKS 시스템의 적어도 하나의 도메인의 생물학적 활성을 가진다. 다른 측면에서, 상기 단백질의 아미노산 서열은 서열번호: 39, 서열번호: 43, 서열번호: 45, 서열번호: 48, 서열번호: 50, 서열번호: 52, 서열번호: 54, 서열번호: 56, 서열번호: 58, 서열번호: 60, 서열번호: 62 및 서열번호: 64로부터 선택된 아미노산 서열에 대하여 적어도 약 70% 동일한, 보다 바람직하게는 적어도 약 75% 동일하되, PUFA PKS 시스템의 적어도 하나의 도메인의 생물학적 활성을 가진다. 다른 측면에서, 상기 단백질의 아미노산 서열은 서열번호: 39, 서열번호: 41, 서열번호: 43, 서열번호: 45, 서열번호: 48, 서열번호: 50, 서열번호: 52, 서열번호: 54, 서열번호: 56, 서열번호: 58,

서열번호: 60, 서열번호: 62, 서열번호: 64, 서열번호: 66, 서열번호: 68로부터 선택된 아미노산 서열에 대하여 적어도 약 80% 동일한, 보다 바람직하게는 적어도 약 85% 동일한, 보다 바람직하게는 적어도 약 90% 동일한, 보다 바람직하게는 적어도 약 95% 동일한, 보다 바람직하게는 적어도 약 96% 동일한, 보다 바람직하게는 적어도 약 97% 동일한, 보다 바람직하게는 적어도 약 98% 동일한, 보다 바람직하게는 적어도 약 99% 동일하되, PUFA PKS 시스템의 적어도 하나의 도메인의 생물학적 활성을 가진다. 한 양태에서, 상기 아미노산 서열은 서열번호: 2, 서열번호: 4, 서열번호: 6, 서열번호: 8, 서열번호: 10, 서열번호: 13, 서열번호: 18, 서열번호: 20, 서열번호: 22, 서열번호: 24, 서열번호: 26, 서열번호: 28, 서열번호: 30, 서열번호: 32 중 하나의 아미노산 서열을 포함하지 않는다.

[0148] 바람직한 양태에서, 본 발명에 따른 분리된 단백질 또는 도메인은 서열번호: 39, 서열번호: 41, 서열번호: 43, 서열번호: 45, 서열번호: 48, 서열번호: 50, 서열번호: 52, 서열번호: 54, 서열번호: 56, 서열번호: 58, 서열번호: 60, 서열번호: 62, 서열번호: 64, 서열번호: 66, 서열번호: 68 또는 PUFA PKS 시스템의 적어도 하나의 생물학적 활성을 가진 임의의 단편을 포함하여, 상기의 생물학적으로 활성인 단편으로부터 선택된 아미노산 서열을 포함하거나, 필수적으로 구성하거나, 구성한다.

[0149] 본 발명의 한 측면에서, 하기 스킴조카이트리움 단백질과 도메인이 본 발명의 하나 이상의 양태에서 유용하며, 이들 모두는 상기된 미국특허출원 제10/124,800호에 개시되어져 있다. 본 발명의 한 측면에서, 본 발명에서 유용한 PUFA PKS 단백질 또는 도메인은 서열번호: 2, 서열번호: 4 및 서열번호: 6으로부터 선택된 아미노산의 적어도 약 500개의 연속적인 아미노산 서열에 대하여 적어도 약 60% 동일하되, PUFA PKS 시스템의 적어도 하나의 도메인의 생물학적 활성을 가지는 아미노산 서열을 포함한다. 추가의 측면에서, 서열번호: 2, 서열번호: 4 및 서열번호: 6의 적어도 약 600개의 연속적인 아미노산, 보다 바람직하게는 적어도 약 700개의 연속적인 아미노산, 보다 바람직하게는 적어도 약 800개의 연속적인 아미노산, 보다 바람직하게는 적어도 약 900개의 연속적인 아미노산, 보다 바람직하게는 적어도 약 1000개의 연속적인 아미노산, 보다 바람직하게는 적어도 약 1100개의 연속적인 아미노산, 보다 바람직하게는 적어도 약 1200개의 연속적인 아미노산, 보다 바람직하게는 적어도 약 1300개의 연속적인 아미노산, 보다 바람직하게는 적어도 약 1400개의 연속적인 아미노산, 보다 바람직하게는 적어도 약 1500개의 연속적인 아미노산 또는 서열번호: 6의 전장에 대하여 적어도 약 60% 동일하다. 추가의 측면에서, 상기 단백질의 아미노산 서열은 서열번호: 2 또는 서열번호: 4의 적어도 약 1600개의 연속적인 아미노산, 보다 바람직하게는 적어도 약 1700개의 연속적인 아미노산, 보다 바람직하게는 적어도 약 1800개의 연속적인 아미노산, 보다 바람직하게는 적어도 약 1900개의 연속적인 아미노산, 보다 바람직하게는 적어도 약 2000개의 연속적인 아미노산 또는 서열번호: 4의 전장에 대하여 적어도 약 60% 동일하다. 추가의 측면에서, 상기 단백질의 아미노산은 서열번호: 2의 적어도 약 2100개의 연속적인 아미노산, 보다 바람직하게는 적어도 약 2200개의 연속적인 아미노산, 보다 바람직하게는 적어도 약 2300개의 연속적인 아미노산, 보다 바람직하게는 적어도 약 2400개의 연속적인 아미노산, 보다 바람직하게는 적어도 약 2500개의 연속적인 아미노산, 보다 바람직하게는 적어도 약 2600개의 연속적인 아미노산, 보다 바람직하게는 적어도 약 2700개의 연속적인 아미노산, 보다 바람직하게는 적어도 약 2800개의 연속적인 아미노산, 보다 더 바람직하게는 서열번호: 2 전장에 대하여 적어도 약 60% 동일하다.

[0150] 다른 측면에서, 본 발명에 따른 하나 이상의 양태에서 유용한 PUFA PKS 단백질 또는 도메인은 상기 단락에 언급된 연속적인 아미노산 길이에 걸쳐서 서열번호: 2, 서열번호: 4 또는 서열번호: 6으로부터 선택된 아미노산 서열에 대하여 적어도 약 65% 동일한, 보다 바람직하게는 적어도 약 70% 동일한, 보다 바람직하게는 적어도 약 75% 동일한, 보다 바람직하게는 적어도 약 80% 동일한, 보다 바람직하게는 적어도 약 85% 동일한, 보다 바람직하게는 적어도 약 90% 동일한, 보다 바람직하게는 적어도 약 95% 동일한, 보다 바람직하게는 적어도 약 96% 동일한, 보다 바람직하게는 적어도 약 97% 동일한, 보다 바람직하게는 적어도 약 98% 동일한, 보다 바람직하게는 적어도 약 99% 동일하되, PUFA PKS 시스템의 적어도 하나의 도메인의 생물학적 활성을 가지는 아미노산 서열을 포함한다.

[0151] 본 발명의 다른 측면에서, 본 발명의 하나 이상의 양태에서 유용한 PUFA PKS 단백질 또는 도메인은 서열번호: 8, 서열번호: 10, 서열번호: 13, 서열번호: 18, 서열번호: 20, 서열번호: 22, 서열번호: 24, 서열번호: 26, 서열번호: 28, 서열번호: 30, 서열번호: 32로부터 선택된 아미노산 서열과 적어도 약 60% 동일하되, PUFA PKS 시스템의 적어도 하나의 도메인의 생물학적 활성을 가지는 아미노산 서열을 포함한다. 추가의 측면에서, 서열번호: 8, 서열번호: 10, 서열번호: 13, 서열번호: 18, 서열번호: 20, 서열번호: 22, 서열번호: 24, 서열번호: 26, 서열번호: 28, 서열번호: 30, 서열번호: 32로부터 선택된 아미노산 서열과 적어도 약 65% 동일한, 보다 바람직하게는 적어도 약 70% 동일한, 보다 바람직하게는 적어도 약 75% 동일한, 보다 바람직하게는 적어도 약

80% 동일한, 보다 바람직하게는 적어도 약 85% 동일한, 보다 바람직하게는 적어도 약 90% 동일한, 보다 바람직하게는 적어도 약 95% 동일한, 보다 바람직하게는 적어도 약 96% 동일한, 보다 바람직하게는 적어도 약 97% 동일한, 보다 바람직하게는 적어도 약 98% 동일한, 보다 바람직하게는 적어도 약 99% 동일하되, PUFA PKS 시스템의 적어도 하나의 도메인의 생물학적 활성을 가진다.

[0152] 본 발명의 또 다른 측면에서, 본 발명에 따른 하나 이상의 양태에서 유용한 PUFA PKS 단백질 또는 도메인은 서열번호: 2, 서열번호: 4, 서열번호: 6, 서열번호: 8, 서열번호: 10, 서열번호: 13, 서열번호: 18, 서열번호: 20, 서열번호: 22, 서열번호: 24, 서열번호: 26, 서열번호: 28, 서열번호: 30, 서열번호: 32 또는 PUFA PKS 시스템의 적어도 하나의 도메인의 생물학적 활성을 가지는 임의의 도메인을 포함하여, 상기의 생물학적으로 활성인 단편을 포함하거나, 필수적으로 구성하거나, 구성한다.

[0153] 본 발명에 있어서, 핵산 또는 아미노산 서열과 관련하여 이용된 용어 "인접한" 또는 "연속적인"은 중단되지 않은 서열에서 연결된 것을 의미한다. 예를 들어 제 1 서열이 제 2 서열의 30개의 인접한(또는 연속적인) 아미노산을 포함한다는 것은 상기 제 1 서열이 제 2 서열의 중단되지 않은 30개 아미노산 잔기의 서열과 100% 동일한 중단되지 않은 30개 아미노산 잔기의 서열을 포함하는 것을 의미한다. 유사하게 제 1 서열이 제 2 서열과 "100% 동일성"을 갖는다는 것은 상기 제 1 서열과 제 2 서열이 뉴클레오타이드 또는 아미노산 간의 갭(gap)없이 정확하게 일치한다는 것을 의미한다.

[0154] 본 발명에서 사용되는 바와 같이, 달리 명시하지 않는 한 동일성 %에 대한 언급은 (1) 아미노산 조사용 blastp 및 핵산 조사용 blastn, 핵산 조사와 6개의 오픈 리딩 프레임내 해독된 아미노산 조사용 blastX를 모두 디폴트 변수로 사용해서 BLAST 2.0 Basic BLAST 상동성 조사; 여기서, 질의(query) 서열은 디폴트에 의해 복잡성이 낮은 영역에 대해 필터링된다 (참조: Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res. 25: 3389-3402; 이의 전문은 본 발명에서 참조로 인용된다); (2) BLAST 2 정렬(하기 기재한 변수를 사용함); (3) 및/또는 표준 디폴트 변수를 사용한 PSI-BLAST(Position Specific Iterated BLAST)를 사용하여 수행하는 상동성 평가를 의미한다. BLAST 2.0 Basic BLAST와 BLAST 2 간에 표준 변수의 몇 가지 차이가 있기 때문에, BLAST 2 프로그램을 사용하는 경우에 2개의 특이적 서열이 유의적 상동성을 갖는 것으로 인정될 수 있는 반면에, 질의 서열로서 서열들 중 하나를 사용하여 BLAST 2.0 Basic BLAST으로 수행한 조사는 제 2 서열을 최상의 매치로서 식별하지 않을 수 있음이 주지된다. 또한, PSI-BLAST는 서열 상동체를 고찰하기 위한 민감한 방법으로서, 사용이 용이한 자동화 형태의 "프로필" 조사를 제공한다. 상기 프로그램은 먼저, 갭이 있는(gapped) BLAST 데이터베이스 조사를 수행한다. PSI-BLAST 프로그램은 다음 라운드의 데이터베이스 조사를 위해 질의 서열을 대신하게 되는 위치-특이적 스코어 행렬을 구축하기 위해 복귀된 임의의 유의적 정렬로부터의 정보를 사용한다. 따라서, 동일성%는 이들 프로그램 중 어느 하나를 사용함으로써 측정할 수 있는 것으로 이해될 것이다.

[0155] 2개의 특이적 서열을 문헌[참조: Tatusova and Madden, (1999), "Blast 2 sequences-a new tool for comparing protein and nucleotide sequences", FEMS Microbiol Lett. 174: 247; 이의 전문은 본 발명에서 참조로 인용된다]에 기재된 바와 같이 BLAST2 서열을 사용하여 서로 정렬할 수 있다. 생성된 정렬에 갭(결실 및 삽입)의 도입을 허용하면서 상기 두 서열 간에 갭이 있는(Gapped) BLAST 조사(BLAST 2.0)를 수행하기 위해 BLAST 2 서열 정렬은 BLAST 2.0 알고리즘을 사용하여 blastp 또는 blastn으로 수행한다. 본 발명에서 명료성의 목적상, BLAST 2 서열 정렬은 다음과 같은 표준 디폴트 변수를 사용하여 수행한다.

[0156] blastn의 경우, 다음의 0 BLOSUM62 행렬을 사용한다:

[0157] 매치에 대한 보상값=1

[0158] 미스매치에 대한 패널티= -2

[0159] 오픈 갭(5) 및 신장 갭(2) 패널티

[0160] 갭 x_탈락값(50) 기대값(10) 단어 길이(11) 필터(적용)

[0161] blastp의 경우, 다음의 0 BLOSUM62 행렬을 사용한다:

[0162] 오픈 갭(11) 및 신장 갭(1) 패널티

[0163] 갭 x_탈락값(50) 기대값(10) 단어 길이(3) 필터(적용).

[0164] 본 발명에서 PUFA PKS 시스템의 적어도 하나의 도메인의 생물학적 활성을 가지는 아미노산 서열은 앞서 스킴 카이트리움 PUFA PKS 시스템에 의해 예시되고 추가로 트라우스토카이트리움 PUFA PKS 시스템에 의해 예시된 바와 같이, 본 발명에서 언급된 PUFA PKS 시스템의 적어도 하나의 도메인의 생물학적 활성을 가지는 아미노산 서열이다. 스킴 카이트리움 또는 트라우스토카이트리움 PUFA PKS 시스템내에 있는 다양한 도메인의 생물학적 활성은 앞서 기술되었다. 따라서, 본 발명에서 유용한 분리된 단백질은 PUFA PKS 오픈 리딩 프레임의 해독 생산물, PUFA PKS 도메인, 이의 생물학적으로 활성인 단편 또는 생물학적 활성을 가지는 천연적으로 존재하는 PUFA PKS 오픈 리딩 프레임 생산물 또는 도메인의 상동체를 포함할 수 있다.

[0165] 본 발명의 다른 양태에서, 본 발명에 따른 PUFA PKS 시스템의 적어도 하나의 도메인의 생물학적 활성을 가지는 아미노산 서열은 천연적으로 존재하는 PUFA PKS 단백질 또는 폴리펩타이드와 충분히 유사한 아미노산 서열을 갖는 단백질을 포함하는 것으로 상기 아미노산 서열을 코딩하는 핵산 서열은 (이하에서 언급되는) 온화한, 높은 또는 매우 높은 엄격한 조건에서 천연적으로 존재하는 PUFA PKS 단백질 또는 폴리펩타이드를 코딩하는 핵산 분자(즉, 천연적으로 존재하는 PUFA PKS 단백질 또는 폴리펩타이드를 코딩하는 핵산 가닥의 상보체)에 (즉, 와) 혼성화될 수 있다. 바람직하게는 본 발명에 따른 PUFA PKS 시스템의 적어도 하나의 생물학적 활성을 가지는 아미노산 서열은 온화한, 높은 또는 매우 높은 엄격한 조건에서 PUFA PKS 단백질 또는 도메인에 대하여 전술된 아미노산 서열을 코딩하는 핵산 서열의 상보체에 혼성화하는 핵산 서열에 의해 코딩된다. 상보적 서열을 유추하는 방법은 당업계에 알려져 있다. 아미노산 서열 분석 및 핵산 서열 분석 기술은 오류로부터 완전히 자유롭지 않기 때문에 본 발명에 나타난 서열은 최대한 본 발명에 따른 PUFA PKS 도메인 또는 단백질의 분명한 서열을 나타낸 것이다.

[0166] 본 발명에서 혼성화 조건은 임의의 핵산 분자가 유사한 핵산 분자를 동정하기 위하여 이용될 수 있는 표준 혼성화 조건을 의미한다. 그러한 표준 조건은 예를 들어, Sambrook et al. *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Labs Press, 1989에 개시되어 있다. 상기 문헌 Sambrook et al.은 전문이 본 발명에서 참조로 인용된다(특히 페이지 9.31-9.62 참조). 또한, 뉴클레오타이드의 부정합(mismatch) 정도를 다양화하는 혼성화를 달성하기에 적합한 혼성화 및 세정 조건을 계산하는 공식은 예를 들어 Meinkoth et al. 1984, *Anal. Biochem.* 138, 267-284에 개시되어 있다. 상기 문헌 Meinkoth et al.은 전문이 본 발명에서 참조로 인용된다.

[0167] 보다 자세하게는 본 발명에서 기술된 온화한 엄격한 혼성화 및 세정 조건은 혼성화 반응에서 프로브로 사용될 핵산 분자와 적어도 약 70% 이상의 핵산 서열 동일성을 갖는 핵산 분자의 분리를 가능하게 하는 조건을 의미한다(즉, 약 30% 이하의 뉴클레오타이드의 부정합을 가능하게 하는 조건). 본 발명에서 높은 엄격한 혼성화 및 세정 조건은 혼성화 반응에서 프로브로 사용될 핵산 분자와 적어도 약 80% 이상의 핵산 서열 동일성을 갖는 핵산 분자의 분리를 가능하게 하는 조건을 말한다(즉, 뉴클레오타이드의 약 20% 이하의 부정합을 가능하게 하는 조건). 본 발명에서 매우 높은 엄격한 혼성화 및 세정 조건은 혼성화 반응에서 프로브로 사용될 핵산 분자와 적어도 약 90% 이상의 핵산 서열 동일성을 갖는 핵산 분자의 분리를 가능하게 하는 조건을 말한다(즉, 뉴클레오타이드의 약 10% 이하의 부정합을 가능하게 하는 조건). 앞서 기재된 바와 같이 당업자는 이러한 특정한 뉴클레오타이드 부정합 정도를 달성할 수 있는 적합한 혼성화 및 세정 조건을 계산하는데 Meinkoth et al.에 기술된 공식을 사용할 수 있다. 그러한 조건은 DNA:RNA 또는 DNA:DNA 혼성체가 형성되는지에 따라 달라질 것이다. DNA:DNA 혼성체에 대하여 계산된 용융 온도는 DNA:RNA 혼성체 보다 10°C 이하이다. 특정한 양태에서, DNA:DNA 혼성체에 대한 엄격한 혼성화 조건은 적합한 세정 조건을 포함하여 6X SSC(0.9M Na⁺)의 이온 강도로 약 20°C 내지 약 35°C(낮은 엄격성), 보다 바람직하게는 약 28°C 내지 약 40°C(높은 엄격성), 보다 더 바람직하게는 약 35°C 내지 45°C의 온도(보다 높은 엄격성)에서의 혼성화를 포함한다. 특정 양태에서, DNA:RNA 혼성체에 대한 엄격한 혼성화 조건은 유사하게 엄격한 세정 조건을 포함하여 6X SSC(0.9M Na⁺)의 이온 강도로 약 30°C 내지 약 45°C, 보다 바람직하게는 약 38°C 내지 약 50°C, 보다 더 바람직하게는 약 45°C 내지 55°C의 온도에서의 혼성화를 포함한다. 이러한 수치는 약 100개 이상의 뉴클레오타이드, 0% 포름아미드 및 약 40%의 G+C 함량의 분자에 대한 용융 온도의 계산에 기반을 둔다. 선택적으로 T_m은 위에서 인용된 Sambrook et al. 페이지 9.31 내지 9.62에 기술된 바에 따라 실험적으로 계산될 수도 있다. 일반적으로 세정 조건은 가능한 엄격해야 하고 선택된 혼성화 조건에 적합해야 할 것이다. 예를 들어, 혼성화 조건은 염과 특정한 혼성체에 대하여 계산된 T_m

보다 약 20-25℃ 낮은 온도 조건의 조합을 포함할 수 있으며, 세정 조건은 염과 특정한 혼성체에 대하여 계산된 T_m 보다 약 12-20℃ 낮은 온도 조건의 조합을 포함한다. DNA:DNA 혼성체에 이용하기에 적합한 혼성화 조건의 한 예는 약 42℃에서의 6X SSC(50% 포름아미드)에서 2-24시간 혼성화하고 실온에서 약 2X SSC로 한 차례 이상 세정하고 이어서 더 높은 온도와 더 낮은 이온 강도에서 추가로 세정(예, 약 37℃에서 약 0.1X-0.5X SSC로 적어도 한 차례 이상 세정한 후 약 68℃에서 약 0.1X 내지 0.5X SSC로 적어도 한 차례 이상 세정)하는 세정 단계를 포함한다.

[0168] 본 발명은 하나 이상의 융합 절편에 부착된 PUFA PKS 단백질 또는 도메인 또는 상동체 또는 이들의 단편을 포함하는 융합 단백질을 포함한다. 본 발명에서 이용하기에 적합한 융합 절편은 이에 제한되는 것은 아니지만 단백질의 안정성을 증진시키고; 다른 목적한 생물학적 활성을 제공하고/거나 단백질의 (예, 친화 크로마토그래피에 의한) 정제를 용이하게 하는 절편을 포함한다. 적합한 융합 절편은 목적한 기능(예, 증가된 안정성, 용해성, 생물학적 활성을 부여고/거나 단백질의 정제를 간편화하는)을 가지는 임의의 크기의 도메인일 수 있다. 융합 절편은 단백질의 아미노 및/또는 카르복시 말단에 결합될 수 있으며, 목적한 단백질의 간편한 회수가 가능하도록 절단에 민감할 수 있다. 융합 단백질은 본 발명에 따른 단백질의 카르복시 및/또는 아미노 말단에 부착된 융합 절편을 포함한 단백질을 코딩하는 융합 핵산 분자로 형질전환된 재조합 세포를 배양함으로써 제조된다.

[0169] 본 발명의 한 양태에서, 상기된 PUFA PKS 아미노산 서열 뿐만 아니라 상기 서열의 상동체는 특정된 아미노산 서열의 C- 및/또는 N-말단 각각의 측면에 적어도 하나 내지 적어도 약 20개까지 추가 이형성 아미노산 서열이 위치하도록 제조될 수 있다. 이렇게 형성된 단백질 또는 폴리펩타이드는 특정한 아미노산 서열로 "필수적으로 구성되는"이라고 기술될 수 있다. 본 발명에서 상기 이형성 아미노산 서열은 특정한 아미노산 서열의 측면에 천연적으로 위치하지 않거나(즉, 천연 생체내에서 발견되지 않는), 천연적으로 발생하는 서열에서 그러한 뉴클레오타이드가 주어진 아미노산 서열이 유래되는 유기체의 표준 코돈 이용(usage)을 사용하여 획득되었다면 유전자에서 발생함에 따라 특정한 아미노산 서열을 코딩하는 천연적으로 발생하는 핵산 서열 측면에 위치하는 뉴클레오타이드에 의해 코딩되지 않는 아미노산 서열이다. 유사하게, "필수적으로 구성되는"이라는 문구가 본 발명에서 핵산 서열과 관련되어서 사용되면 특정한 아미노산 서열을 코딩하는 핵산 서열의 5'- 및/또는 3'-말단 각각에 적어도 하나 내지 약 60개까지 추가 이형성 뉴클레오타이드가 측면에 위치하는 특정한 아미노산 서열을 코딩하는 핵산 서열을 의미한다. 이형성 뉴클레오타이드는 천연 유전자에서 발생함에 따라 특정한 아미노산 서열을 코딩하는 핵산 서열의 측면에 천연적으로 발견되지 않는다(즉, 천연 생체내에서 발견되지 않는다).

[0170] 한 측면에서 본 발명에 따른 단백질 또는 도메인 및/또는 상동체 또는 이의 단편의 최소 크기는 필요한 생물학적 활성을 가지거나 항체의 생성을 위한 항원으로서 또는 실험실내 에세이에서 표적으로서 작용하기에 충분한 크기이다. 한 양태에서, 본 발명의 단백질은 본 발명에 따른 단백질 또는 도메인의 (예, 항체 에피토프 또는 에세이에서 검출할 수 있는 펩타이드로 적합한) 적어도 약 8개의 아미노산, 적어도 약 25개의 아미노산, 적어도 약 50개의 아미노산, 적어도 약 100개의 아미노산, 적어도 약 150개의 아미노산, 적어도 약 200개의 아미노산, 적어도 약 250개의 아미노산, 적어도 약 300개의 아미노산, 적어도 약 350개의 아미노산, 적어도 약 400개의 아미노산, 적어도 약 450개의 아미노산, 적어도 약 500개의 아미노산, 적어도 약 750개의 아미노산, 정수(예, 8, 9, 10,...25, 26,...500, 501,...1234, 1235...)로 8개 아미노산 내지 전장 중 임의의 길이 또는 더 길 수 있다. 단백질의 최대 크기에 대해서는 통상적인 제한 이외의 특별한 제한은 없어서, 단백질은 PUFA PKS 단백질의 일부, 도메인 또는 이의 생물학적으로 활성이거나 유용한 단편 또는 전장의 PUFA PKS 단백질 또는 단편, 목적한다면 더해진 추가 서열(예, 융합 단백질 서열)을 포함할 수 있다.

[0171] 본 발명의 추가 양태는 상동체 또는 이의 단편 포함하여, 상기된 단백질 또는 도메인을 코딩하는 핵산 서열 뿐만 아니라 이에 완전히 상보적인 핵산 서열을 포함하는, 필수적으로 구성되는 또는 구성되는 분리된 핵산 분자를 포함한다. 본 발명에서, 분리된 핵산 분자는 이의 천연 환경 즉, 상기 핵산 분자가 천연에서 위치하는 게놈 또는 염색체인 그것의 천연 환경으로부터 제거된(즉, 인간의 조작이 가해진) 핵산 분자이다. 이에 "분리된"은 핵산 분자가 정제된 정도를 필수적으로 반영하지는 않지만, 상기 분자는 상기 분자가 천연적으로 존재하는 전체 게놈 또는 전체 염색체를 포함하지 않는다는 것을 의미한다. 분리된 핵산 분자는 유전자를 포함할 수 있다. 유전자를 포함하는 분리된 핵산 분자는 그 유전자를 포함하는 염색체의 단편이 아니라 동일한 염색체에 천연적으로 존재하는 추가 유전자 없이 상기 유전자와 관련된 코딩 영역과 조절 영역을 포함한다. 분리된 핵산 분자는 천연에서 특정된 핵산 서열의 측면에 정상적으로 위치하는 않는(즉, 이형성의 서열) 추가 핵산이 측면(즉, 서열의 5' 및/또는 3'의 말단)에 위치하는 특정한 핵산 서열을 또한 포함할 수 있다. 분리된 핵산 분자는 DNA, RNA(예, mRNA), 또는 DNA 또는 RNA 중 어느 하나의 유도체(예, cDNA)를 포함할 수 있다. "핵산 분자"는 일차적으로 물리적인 핵산 분자를 의미하는 것이고 "핵산 서열은"은 일차적으로 핵산 분자의 뉴클레오타

이드 서열을 의미하는 것이지만, 상기 두 문구는 특히 단백질 또는 단백질의 도메인을 코딩할 수 있는 핵산 분자 또는 핵산 서열이라는 관점에서 상호교환적으로 사용될 수 있다.

[0172] 바람직하게는 본 발명에 따른 분리된 핵산 분자는 재조합 DNA 기술(예, 중합효소연쇄반응(PCR) 증폭, 클로닝) 또는 화학적 합성법을 이용하여 제조할 수 있다. 분리된 핵산 분자는 천연의 핵산 분자 및 이의 상동체로 이에 제한되는 것은 아니지만 천연 대립형질 변이체 및 PUFA PKS 단백질의 상기된 바와 같은 생물학적 활성에 목적인 영향을 제공할 수 있는 방식으로 뉴클레오타이드가 삽입, 결실, 치환 및/또는 역위된 변형된 핵산 분자를 포함한 상동체를 포함한다. 단백질 상동체(예, 핵산 분자 상동체에 의해 코딩된 단백질)는 앞에서 기술되었다.

[0173] 핵산 분자 상동체는 당업계에서 공지된 다양한 방법을 이용하여 제조될 수 있다(예를 들어 상기 문헌 Sambrook et al. Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Labs Press, 1989). 예를 들어 핵산 분자는 이에 제한되는 것은 아니지만, 위치 지정 돌연변이, 핵산 분자에 돌연변이를 유발할 수 있는 화합물을 처리, 핵산 단편의 효소 처리에 의한 절단, 핵산 단편의 라이게이션, PCR 증폭 및/또는 핵산 서열의 선별된 영역의 돌연변이 유발, 올리고뉴클레오타이드 혼합물의 합성과 핵산 분자의 혼합체를 "세우기"위한 혼합군의 라이게이션 및 이들의 조합과 같은 전통적인 돌연변이 유발 기술 및 재조합 DNA 기술을 포함한 다양한 기술을 사용하여 변형될 수 있다. 핵산 분자 상동체는 그 핵산 분자에 의해 코딩된 단백질의 기능에 대하여 스크리닝 및/또는 야생형 유전자와의 혼성화에 의해 변형된 핵산의 혼합물로부터 선택할 수 있다.

[0174] 본 발명에 따른 핵산 분자의 최소의 크기는 본 발명에서 유용한 핵산 분자의 상보적인 서열과 (예, 온화한, 높은 또는 매우 높은 엄격한 조건에서) 안정적 혼성체를 형성할 수 있는 올리고뉴클레오타이드 프라이머 또는 프로브를 형성하기에 충분하거나 본 발명에 따른 PUFA PKS 시스템의 적어도 하나의 도메인의 생물학적 활성을 가진 아미노산 서열을 코딩하기에 충분한 크기이다. 이와 같이 그러한 단백질을 코딩하는 핵산 분자의 크기는 혼성화 조건 그 자체(예, 온도, 염농도 및 포름아미드 농도) 뿐만 아니라 핵산 분자의 조성 및 핵산 분자와 상보적 서열간의 백분율 상동성 또는 동일성에 따라 달라질 수 있다. 올리고뉴클레오타이드 프라이머 또는 프로브로서 이용되는 핵산 분자의 최소 크기는 일반적으로 GC가 풍부하다면 적어도 약 12 내지 약 15 뉴클레오타이드 이상이며 AT가 풍부하다면 적어도 약 15 so지 약 18 염기이다. 본 발명에 따른 핵산 분자의 최대 크기에 대하여 통상적인 제한 이외의 제한은 없지만, 핵산 분자는 PUFA PKS 시스템의 도메인의 생물학적으로 활성인 단편, PUFA PKS 시스템의 전체 도메인, PUFA PKS 시스템의 오픈 리딩 프레임 (Orf) 내에 있는 여러 개의 도메인, PUFA PKS 시스템의 전체 Orf 또는 PUFA PKS 시스템의 하나 이상의 Orf를 코딩하기에 충분한 서열을 포함할 수 있다.

[0175] 본 발명의 한 양태에서, 분리된 핵산 분자는 본 발명에서 언급된 스킴조카이트리움 또는 트라우스토카이트리움 유래의 아미노산 서열 또는 이의 상동체를 포함한 상술된 아미노산 서열 중 어느 하나를 코딩하는 핵산 서열을 포함하거나, 필수적으로 구성하거나, 구성한다. 한 측면에서, 상기 핵산 서열은 서열번호: 1, 서열번호: 3, 서열번호: 5, 서열번호: 7, 서열번호: 9, 서열번호: 12, 서열번호: 17, 서열번호: 19, 서열번호: 21, 서열번호: 23, 서열번호: 25, 서열번호: 27, 서열번호: 29, 서열번호: 31, 서열번호: 38, 서열번호: 40, 서열번호: 42, 서열번호: 44, 서열번호: 47, 서열번호: 49, 서열번호: 51, 서열번호: 53, 서열번호: 55, 서열번호: 57, 서열번호: 59, 서열번호: 61, 서열번호: 63, 서열번호: 65 또는 서열번호: 67 또는 (상기 서열에 대하여 적어도 약 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% 또는 99% 동일한 서열을 포함한) 상동체 또는 이의 단편 또는 이의 상보적인 서열로 이루어진 그룹으로부터 선택된다.

[0176] 본 발명의 다른 양태는 재조합 벡터와 상기된 PUFA PKS 시스템의 적어도 하나의 도메인의 생물학적 활성을 가진 단백질 또는 펩타이드 (또는 상동체 또는 이들의 단편)를 코딩하는 핵산 서열을 포함하는 재조합 핵산 분자를 포함한다. 상기 핵산 서열은 앞서 상세히 기술되었다. 본 발명에 있어서, 재조합 벡터는 선택된 핵산 서열을 조작하고 그러한 핵산 서열을 숙주 세포로 도입하는 도구로서 이용되도록 설계된(즉, 인위적으로 제조된) 핵산 분자이다. 따라서 재조합 벡터는 선택된 핵산 서열을 클로닝, 서열분석 및/또는 선택된 핵산 서열의 발현 및/또는 그러한 핵산 서열의 숙주 세포로 전달하여 재조합 세포를 형성하는 것과 같이 다른 조작에 이용하는데 적합하다. 그러한 벡터는 일반적으로 클로닝되고 전달되는 핵산 서열에 천연적으로 근접하여 위치하지 않는 핵산 서열인 이형성의 핵산 서열을 포함하지만, 상기 벡터는 또한 본 발명의 핵산 분자에 천연적으로 근접하여 위치하거나 (이하에서 상세히 기술되는) 본 발명의 핵산 분자의 발현에 유용한 조절 핵산 서열(예, 프로모터, 비핵독 영역)을 포함할 수 있다. 벡터는 RNA 또는 DNA일 수 있으며 원핵생물 또는 진핵생물 유래일 수 있으며 일반적으로는 플라스미드이다. 벡터는 염색체의 성분(예, 플라스미드)으로서 유지되거나 재조합 유기체(예, 미생물 또는 식물)의 염색체로 통합될 수 있다. 전체 벡터는 숙주 세포내에 위치하거나 특정 조건에서는

본 발명의 핵산 분자를 남겨두면서 플라스미드 DNA가 결실될 수 있다. 통합된 핵산 분자는 염색체 프로모터의 조절하에 천연 또는 플라스미드 프로모터 조절하에 또는 여러 프로모터 조합의 조절하에 위치시킬 수 있다. 핵산 분자의 단일 또는 다수의 카피가 염색체내로 통합될 수 있다. 본 발명의 재조합 벡터는 적어도 하나 이상의 선별 마커를 포함할 수 있다.

[0177] 한 양태에서, 본 발명에 따른 재조합 핵산 분자로 이용되는 재조합 벡터는 발현 벡터이다. 본 발명에서 용어 "발현 벡터"는 코딩된 생산물(예, 대상 단백질)의 제조에 적합한 벡터를 의미한다. 이러한 양태에서, 제조될 생산물(예, PUFA PKS 도메인)을 코딩하는 핵산 서열은 재조합 벡터에 삽입되어 재조합 핵산 분자로 제조된다. 제조될 단백질을 코딩하는 핵산 서열은 재조합 숙주 세포내에서 상기 핵산 서열의 전사 및 해독을 가능하게 하는 벡터의 조절 서열에 상기 핵산 서열을 작동가능하게 연결시키는 방식으로 벡터에 삽입할 수 있다.

[0178] 다른 양태에서, 본 발명에 따른 재조합 핵산 분자로 이용되는 재조합 벡터는 표적화 벡터이다. 본 발명에서 용어 "표적화 벡터"는 재조합 숙주 세포 내로 특정 핵산 분자를 전달하는데 이용되는 벡터를 의미하는 것으로 상기 핵산 분자는 숙주 세포 또는 미생물 내에서 내인성 유전자를 결실 또는 불활성화시키는데 이용된다(즉, 표적 유전자의 파괴 또는 녹아웃 기술에서 이용되는). 이러한 벡터는 당업계에 또한 "녹-아웃"벡터로서 알려져 있다. 이러한 양태의 한 측면에서, 벡터의 일부분, 보다 일반적으로는 벡터에 삽입된 핵산 분자(즉, 삽입물)는 숙주 세포내 표적 유전자(즉, 결실 또는 불활성화되도록 표적화된 유전자)의 핵산 서열과 상동성인 핵산서열을 갖는다. 벡터 삽입물의 핵산 서열은 표적 유전자와 결합하도록 설계되어 표적 유전자와 삽입물간의 상동성 재조합이 일어나서 그에 의하여(즉, 돌연변이 또는 결실된 내인성 표적 유전자의 적어도 일부분에 의해) 내인성 표적 유전자가 결실, 불활성 또는 약화된다. 내인성의 스킴조카이트리움 유전자를 재조합 유전자로 치환하기 위하여 상기 타입의 재조합 벡터를 이용한 것이 실시예 부분에 기술되어 있으며, 트라우스토카이트리드의 유전적 형질전환을 위한 일반적인 기술은 2003년 9월 4일자 미국특허공개 제20030166207호로 공개된 미국특허출원 제 10/124,807호에 상세히 개시되어 있다.

[0179] 일반적으로 재조합 핵산 분자는 하나 이상의 발현 조절 서열과 작동가능하게 연결된 적어도 하나 이상의 본 발명에 따른 핵산 분자를 포함한다. 본 발명에서 용어 "재조합 분자"또는 "재조합 핵산 분자"는 일차적으로 발현 조절 서열에 작동가능하게 연결된 핵산 분자 또는 핵산 서열을 의미하는 것이지만, 그러한 핵산 분자가 본 발명에서 기술된 바와 같이 재조합 분자일 때에는 용어 "핵산 분자"와 상호교환적으로 이용될 수 있다. 본 발명에 있어서, 용어 "작동가능하게 연결된"은 핵산 분자가 숙주 세포로 형질감염(즉, 형질전환, 형질도입, 형질감염, 접합 또는 전도)되었을 때 발현될 수 있는 방식으로 발현 조절 서열(예, 전사 조절 서열 및/또는 해독 조절 서열)에 연결된 것을 의미한다. 전사 조절 서열은 전사의 개시, 연장 또는 종결을 조절하는 서열이다. 특히 중요한 전사 조절 서열은 프로모터, 인핸서, 오퍼레이터 및 리프레서 서열과 같은 전사 개시를 조절하는 것이다. 적합한 전사 조절 서열은 재조합 핵산 분자가 도입된 숙주 세포 또는 유기체에서 기능을 발휘할 수 있는 임의의 전사 조절 서열을 포함한다.

[0180] 본 발명에 따른 재조합 핵산 분자는 재조합 세포에 적합한 해독 조절 서열, 복제 개시점 및 다른 조절 서열과 같은 추가의 조절 서열을 또한 포함할 수 있다. 한 양태에서, 숙주 세포 염색체로 통합된 재조합 분자를 포함하여 본 발명에 따른 재조합 분자는 발현된 단백질을 제조한 세포로부터 분리되도록 하는 분비 신호(즉, 신호 단편 핵산 서열)를 또한 포함한다. 적합한 신호 절편은 발현된 단백질과 천연적으로 연관된 신호 절편 또는 본 발명에 따른 단백질의 분비를 유도할 수 있는 임의의 이형성 신호 절편을 포함한다. 다른 양태에서, 본 발명의 재조합 분자는 발현된 단백질이 숙주 세포막에 전달 및 삽입되도록 하는 리더 서열을 포함한다. 적합한 리더 서열은 단백질과 천연적으로 연관된 리더 서열 또는 세포막으로 단백질의 전달 및 삽입을 유도할 수 있는 임의의 이형성 리더 서열을 포함한다.

[0181] 본 발명자들은 스킴조카이트리움 PUFA PKS OrfA 및 B가 게놈 상에서 가깝게 연결되어 있다는 것을 발견하였고, Orfs 간의 영역을 서열분석하였다. Orf는 반대 방향으로 배향되어 있으며, 4244개의 염기쌍이 개시(ATG) 코돈을 이격시킨다(즉, 이들은 다음과 같이 배열된다:3'OrfA5'-4244bp-5'OrfB3'). 4244bp의 유전자간 영역의 연구는 어떠한 분명한 Orf를 밝혀내지 못하였다(BlastX 조사에서 상당한 매치가 발견되지 않았다). OrfA 및 B 모두가 스킴조카이트리움에서 적어도 오일 제조 중에 발현되었는데, 이는 활성 프로모터 요소가 상기 유전자간 영역에 포함되어 있다는 것을 암시한다. 이러한 유전적 요소는 유전자 도입용으로 적용함에 있어서 이방향성(bi-directional) 프로모터로서의 유용성을 갖는 것으로 여겨진다. 예를 들어 바람직한 양태에서, 상기 영역을 클로닝하고 대상 유전자를 각 말단에 위치하고 그 작제물을 스킴조카이트리움 (또는 상기 프로모터가 기능을 발휘하는 것으로 보이는 일부 다른 숙주)으로 도입할 수 있다. 적합한 조건하에서 조절 요소는 조합 작용하여 두 개의 도입된 유전자의 높은 발현율을 제공할 것이다. 스킴조카이트리움 PUFA PKS 조절 요소(예, 프

로모터)를 포함한 조절 영역의 완전한 뉴클레오타이드 서열은 서열번호: 36에 나타내었다.

[0182] 유사한 방식으로 OrfC도 오일 제조 중에 스킴조카이트리움에서 발견되고 조절 요소는 이의 개시 코돈의 업스트림 영역에 존재할 것으로 예상된다. OrfC의 게놈 DNA 업스트림의 영역은 클로닝되고 서열분석되어 본 발명에서 나타내었다(서열번호: 37). 상기 서열은 OrfC 개시 코돈의 3886 nt의 바로 업스트림을 포함한다. 이러한 영역의 연구는 어떠한 분명한 Orf를 밝혀내지 못하였다(BlastX 조사에서 상당한 매치가 발견되지 않았다). 상기 영역내에 포함된 조절 요소는 적합한 조건 하에서 이들의 뒤에 위치하는 유전자의 높은 발현을 가능하게 할 것이다. 또한, 적합한 환경 하에서 유전자의 발현율은 A-B 유전자간 영역(서열번호: 36)의 조절 하에 있는 유전자와 조합 조절될 수 있다.

[0183] 따라서 한 양태에서, 상기된 본 발명에서 유용한 재조합 핵산 분자는 서열번호: 36 및/또는 서열번호: 37 내에 포함된 PUFA PKS 조절 영역을 포함할 수 있다. 상기 조절 영역은 적어도 기본적인(basal) PUFA PKS 전사 활성을 가진 서열번호: 36 및/또는 서열번호: 37의 일부 (단편)를 포함할 수 있다.

[0184] 본 발명에 따른 하나 이상의 재조합 분자가 본 발명의 코딩된 생산물(예, PUFA PKS 도메인, 단백질 또는 시스템)을 제조하는데 이용될 수 있다. 한 양태에서, 코딩된 생산물은 단백질을 제조하기에 효과적인 조건 하에서 상기 핵산 분자를 발현함으로써 제조된다. 코딩된 단백질을 제조하기에 바람직한 방법은 하나 이상의 재조합 분자로 숙주 세포를 형질감염하여 재조합 세포를 형성하는 것이다. 형질감염하기에 적합한 숙주 세포는 이에 제한되는 것은 아니지만, 형질감염될 수 있는 박테리아, 진균(예, 효모), 곤충, 식물 또는 동물의 세포를 포함한다. 본 발명의 한 양태에서, 바람직한 숙주 세포는 (하기에서 상세히 설명되는) 트라우스토카이트리드 숙주 세포 또는 식물 숙주 세포이다. 숙주세포는 형질감염되지 않은 세포 또는 적어도 하나의 다른 재조합 핵산 분자로 형질감염된 세포일 수 있다.

[0185] 본 발명에서 용어 "형질감염"은 외인성 핵산 분자(즉, 재조합 핵산 분자)를 세포내로 삽입할 수 있는 임의의 방법을 의미한다. 용어 "형질전환"은 이 용어가 핵산 분자를 조류, 박테리아 및 효소와 같은 미생물 또는 식물 세포로 도입하는데 이용할 경우 용어 "형질감염"과 상호교환적으로 이용될 수 있다. 미생물 시스템에서 용어 "형질전환"은 미생물 또는 식물에 의한 외인성 유전자의 습득으로 인한 유전되는 변화를 설명하는데 이용되며 "형질감염"과 본질적으로 같은 의미이다. 그러나 동물 세포에서 형질전환은 배양 중 세포의 성장 특성에서의 변화로 이들이 예를 들어 종양화된(cancerous) 후를 말하는 제 2의 의미를 갖고 있다. 따라서 혼동을 피하기 위하여 동물 세포로의 외래 핵산의 도입에 관련해서는 용어 "형질감염"을 사용하는 것이 바람직하고, 본 발명에서 용어 "형질감염"은 상기 용어들이 외래 핵산 분자를 세포내로 도입하는 것과 연관된 것에 한해서는 동물 세포의 형질감염과 식물 세포와 미생물 세포의 형질전환을 포함한다. 따라서 형질감염 기술은 이에 제한되는 것은 아니지만, 형질전환, 입자 사격(particle bombardment), 확산(diffusion), 활성 트랜스포트, 수조내 초음파처리(bath sonication), 전기천공, 미세주입, 리포펙션(lipofection), 흡수, 감염 및 원형질체 융합을 포함한다.

[0186] 재조합 DNA 기술의 이용은 예를 들면 숙주 세포 내에서 핵산 분자의 카피수, 이러한 핵산 분자가 전사되는 효율, 이렇게 형성된 전사체가 해독되는 효율, 해독후 변형의 효율을 조작함으로써 형질감염된 핵산 분자의 발현의 조절을 개선시킬 수 있다는 것이 당업자에게는 자명하다. 또한, 프로모터 서열은 본래(active) 프로모터와 비교하였을 때 발현율을 증진시키도록 유전적으로 조작될 수 있다. 핵산 분자의 발현을 조절하는데 유용한 재조합 기술은 이에 제한되는 것은 아니지만, 하나 이상의 숙주 세포 염색체로 핵산 분자의 통합, 플라스미드로 벡터 안정화 서열의 추가, 전사 조절 신호(예, 프로모터, 오퍼레이터, 인핸서)의 치환 또는 변형, 해독 조절 서열(예, 리보솜 결합 부위, Shine-Dalgarno 서열)의 치환 또는 변형, 숙주 세포의 코돈 이용에 상응하도록 핵산 분자의 변형, 전사체를 불안정하게 하는 서열의 결실을 포함한다.

[0187] 재조합 핵산 분자와 숙주 세포의 형질감염과 관련된 상기 설명은 PUFA PKS 유래의 적어도 하나의 도메인의 생물학적 활성을 가지는 아미노산 서열을 코딩하는, 다른 PKS 시스템 유래의 아미노산 서열을 코딩하는, 다른 단백질 또는 도메인을 코딩하는 재조합 핵산 분자를 포함하여 본 발명에서 기술된 임의의 재조합 핵산 분자에 적용하고자 한 것이다.

[0188] 다중불포화 지방산(PUFA)는 보다 고등 진핵생물의 필수적인 멤브레인 구성요소이고 많은 리피드 유래의 신호 분자의 전구체이다. 본 발명에 따른 PUFA PKS 시스템은 포화 지방산의 탈포화 및 연장을 요구하지 않는 PUFA 합성을 위한 경로를 이용한다. PUFA PKS에 의해 촉매되는 경로는 구조와 메커니즘에 있어서, 기존에 알려진 PKS과는 구별된다.시스 이중 결합의 생성은 위치 특이적인 이소머라제를 포함할 것으로 제안된다; 이러한 효소는 새로운 군의 항생제를 제조하는데 유용할 것으로 여겨진다.

- [0189] 하나 이상의 목적인 다중불포화 지방산 또는 다른 생활성 분자를 상당히 높은 수율로 제조하기 위하여 유기체, 바람직하게는 미생물 또는 식물, 가장 바람직하게는 트라우스토카이트리드 미생물을 유전적으로 변형시켜 미생물 또는 식물에서의 PUFA PKS 시스템의 활성화, 특히 최종 생산물을 변이시킬 수 있다.
- [0190] 따라서 본 발명의 한 양태는 다중불포화 지방산(PUFA) 폴리케타이드 신타제(PKS) 시스템의 적어도 하나의 생물학적으로 활성인 도메인을 포함하는 PKS 시스템을 발현하는 유전적으로 변형된 미생물에 관한 것이다. PUFA PKS 시스템의 한 도메인은 상기된(예, 스킴조카이트리움과 트라우스토카이트리움에 있어서) PUFA PKS 시스템과 관련된 상동체를 포함하여, 도메인은 포함할 수 있으며, 트라우스토카이트리드 미생물을 포함하여 진핵생물의 미생물을 포함한 다른 비박테리아의 미생물에서 유래된 PUFA PKS 시스템의 도메인을 포함할 수 있다. 상기 유기체에서 유전적 변형은 PKS 시스템의 활성화에 영향을 미친다. 미국특허출원 제10/124,800호에 개시된 스크리닝 방법에 의해 동정된 미생물 유래의 PUFA PKS 시스템의 도메인을 포함할 수 있다. 상기 유기체에서 유전적 변형은 PKS 시스템의 활성화에 영향을 미친다. 미국특허출원 제10/124,800호에 개시된 스크리닝 방법은 (a) 적어도 하나의 PUFA를 제조하는 미생물을 선택하는 단계 및 (b) 발효 배지 중 약 5%, 바람직하게는 약 10%, 보다 바람직하게는 약 15%, 보다 바람직하게는 약 20% 이상의 용존 산소 포화 농도에서 미생물에 의한 PUFA 생산량과 비교하였을 때 발효 배지 중 약 5% 이하의 용존 산소 포화 농도에서 증가된 PUFA를 제조하는 능력을 가진 미생물을 동정하는 단계를 포함한다.
- [0191] 한 측면에서, 그러한 유기체는 내인적으로 PUFA PKS 시스템을 포함하고 발현할 수 있으며, 유전적 변형은 상기 내인성의 PUFA PKS 시스템의 적어도 하나 이상의 기능성 도메인의 유전적 변형일 수 있으며, 이로써 상기 변형은 PUFA PKS 시스템의 활성화에 영향을 미친다. 다른 측면에서, 그러한 유기체는 내인적으로 PUFA PKS 시스템을 포함하고 발현할 수 있고, 유전적 변형은 적어도 하나의 외래성 핵산 서열(예, 재조합 핵산 분자)의 도입일 수 있으며, 상기 외래성 핵산 서열은 제 2의 PKS 시스템 유래의 적어도 하나의 생물학적으로 활성인 도메인 또는 단백질 및/또는 PUFA PKS 시스템의 활성화에 영향을 미치는 단백질(예, 하기에서 설명되는 포스포판테테인 트랜스퍼라제(PPTase))을 코딩한다. 또 다른 측면에서, 유기체는 반드시 내인적으로(천연적으로) PUFA PKS 시스템을 포함하지 않지만, PUFA PKS 시스템의 적어도 하나의 도메인의 생물학적 활성을 가진 아미노산 서열을 코딩하는 적어도 하나의 재조합 핵산 분자를 도입하여 유전적으로 변형된다. 이러한 측면에서, PUFA PKS 활성화는 유기체에 PUFA PKS 활성을 도입하거나 증가시킴으로써 영향을 미친다. 이러한 측면과 관련된 다양한 양태가 하기에서 더욱 상세히 설명될 것이다.
- [0192] PUFA PKS 시스템 또는 PUFA PKS 시스템을 포함하는 유기체의 유전적 변형은 (도메인의 일부를 포함한) PUFA PKS 시스템의 적어도 하나의 도메인, (PUFA PKS 시스템에서 근접한 도메인, 비인접한 도메인 또는 상이한 단백질의 도메인을 포함한) PUFA PKS 시스템의 하나 이상 또는 여러 개의 도메인, PUFA PKS 시스템의 전체 단백질, 전체 PUFA PKS 시스템(예, PUFA PKS에 의해 코딩된 모든 단백질)의 변형을 포함할 수 있다. 이에, 변형은 내인성 PUFA PKS 시스템의 단일 도메인에 대한 미미한 변형; 주어진 PUFA PKS 시스템의 적어도 하나 이상의 도메인 또는 단백질에 대한 치환, 결실 또는 추가; 한 유기체의 전체 PUFA PKS 시스템을 다른 유기체 유래의 PUFA PKS 시스템으로 치환하는 것을 포함할 수 있다. 당업자에게는 PUFA PKS 시스템에 대한 유전적 변형이 본 발명에 의해 포함된다는 것이 자명하다.
- [0193] 본 발명에 개시된 유전적으로 변형된 미생물은 유전적으로 변형된 박테리아, 원생생물, 미세조류, 진균류, 또는 다른 세균과 특히 본문에서 서술한 트라우스토카이트리알레(예, 트라우스토카이트리드)목의 모든 속(예를 들어 스킴조카이트리움, 트라우스토카이트리움, 자포노카이트리움, 라비린툴라, 라비린툴로이데스 등)을 포함한다. 상기의 유전적으로 변형된 미생물은 정상적인(즉, 야생형 또는 천연적으로 존재하는) 형태로부터 목적한 결과를 수득하기(PUFA PKS의 활성화 및/또는 PKS 시스템을 사용하여 목적인 생산물의 제조를 증가 또는 변형시키기) 위해 변형된(즉, 돌연변이 또는 변화된) 계통을 가지고 있다. 미생물의 유전적인 변형은 전통적인 균주 개발 및/또는 분자 유전학적 기술을 이용하여 달성될 수 있다. 당업계에 공지된 이와 같은 기술은 예를 들어 Sambrook et al., 1989, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Labs Press에 일반적으로 미생물에 관하여 개시되어 있다. 이의 전문은 본 발명에서 참조로 인용된다. 유전적으로 변형된 미생물은 상기 변형이 미생물 내에서 목적한 효과를 제공하는 방식으로 핵산 분자가 삽입, 결실 또는 변형(즉, 돌연변이된; 예를 들면 뉴클레오티드의 삽입, 결실, 치환 및/또는 역위에 의해)된 미생물을 포함한다.
- [0194] 본 발명에 따라 변형할 바람직한 미생물 숙주세포는 이에 제한되는 것은 아니지만, 박테리아, 원생생물, 미세조류, 진균류 또는 프로토조아를 포함한다. 한 측면에서, 유전적으로 변형시키기에 바람직한 미생물은 이에 제한되는 것은 아니지만, 트라우스토카이트리아세 및 라비린툴라시아에 있는 임의의 미생물을 포함한 트라우스토카이트리얼 목의 미생물을 포함한다. 본 발명에서 이용하기에 특히 바람직한 숙주 세포는 이에 제한되는 것은 아니지만 트라우스토카이트리움, 자포노카이트리움, 아플라노카이트리움, 트라우스토카이트리아

시에 및 라비린툴라 내의 엘리나 및 스킨조카이트리움, 라비린툴로이드 및 라비린툴라시에 내의 라비린토크사를 포함한 속 유래의 미생물을 포함할 수 있다. 상기 속 내의 바람직한 종은 이에 제한되는 것은 아니지만: 라비린툴라 sp., 라비린툴라 알게리엔시스, 라비린툴라 시엔콥스키, 라비린툴라 샤토니, 라비린툴라 코에노시스티스, 라비린툴라 마크로시스티스, 라비린툴라 마크로시스티스 아틀란티카, 라비린툴라 마크로시스티스 마크로시스티스, 라비린툴라 마그니피카, 라비린툴라 미누타, 라비린툴라 로스코펜시스, 라비린툴라 발카노비, 라비린툴라 비텔리나, 라비린툴라 비텔리나 파시피카, 라비린툴라 비텔리나 비텔리나, 라비린툴라 조피를 포함한 라비린툴라의 모든 종; 라비린툴로이드 sp., 라비린툴로이드 미누타, 라비린툴로이드 스킨조카이트리움을 포함한 라비린툴로이드의 모든 종; 라비린토크사 sp., 라비린토크사 폴리아, 라비린토크사 사바게이를 포함한 모든 라비린토크사 종; 아플라노카이트리움 sp. 과 아플라노카이트리움 케르구엘렌시스를 포함한 모든 아플라노카이트리움종; 엘리나 sp., 엘리나 마리살바, 엘리나 시노리피카를 포함한 모든 엘리나종; 자파노카이트리움 sp., 자파노카이트리움 마리눔을 포함한 모든 자파노카이트리움종; 스킨조카이트리움 sp., 스킨조카이트리움 아그레가툼, 스킨조카이트리움 리마시눔, 스킨조카이트리움 미누툼, 스킨조카이트리움 옥토포룸을 포함한 모든 스킨조카이트리움종; 트라우스토카이트리움 sp., 트라우스토카이트리움 아그레가툼, 트라우스토카이트리움 아루디멘탈레, 트라우스토카이트리움 오레움, 트라우스토카이트리움 벤티콜라, 트라우스토카이트리움 글로버섬, 트라우스토카이트리움 킨네이, 트라우스토카이트리움 모티툼, 트라우스토카이트리움 파카이더문, 트라우스토카이트리움 프로리페툼, 트라우스토카이트리움 로세움, 트라우스토카이트리움 스타리아툼, 울케니아 sp., 울케니아 미누타, 울케니아 프로판다, 울케니아 라디에이트, 울케니아 사카리아나, 울케니아 비수르젠시스를 포함한 모든 트라우스토카이트리움 종. 특히 이들 속에서도 바람직한 종은 이에 제한되는 것은 아니지만, 스킨조카이트리움 아그레가툼, 스킨조카이트리움 리마시눔, 스킨조카이트리움 미누툼을 포함한 모든 스킨조카이트리움 종; (U. 비수르젠시스, U. 아메보이다, U. 사카리아나, U. 프로판다, U. 라디에이트, U. 미누타 및 울케니아 sp. BP-5601과 같은 이전의 울케니아 종을 포함한) 그리고 트라우스토카이트리움 스트리아툼, 트라우스토카이트리움 오레움, 트라우스토카이트리움 로세움을 포함한 모든 트라우스토카이트리움 종 및 모든 자파노카이트리움 종을 포함한다. 특히 트라우스토카이트리움의 바람직한 균주는 이에 제한되는 것은 아니지만, 다음과 같다: 스킨조카이트리움 sp. (S31)(ATCC 20888); 스킨조카이트리움 sp. (S8)(ATCC 20889); 스킨조카이트리움 sp. (LC-RM)(ATCC 18915); 스킨조카이트리움 sp. (SR21); 스킨조카이트리움 아그레가툼 (Goldstein et Belsky)(ATCC 28209); 스킨조카이트리움 리마시눔(Honda et Yokochi)(IFO 32693); 트라우스토카이트리움 sp. (23B)(ATCC 28210); 트라우스토카이트리움 스트라이툼(Schneider)(ATCC 24473); 트라우스토카이트리움 오레움 (Goldstein)(ATCC 34304); 트라우스토카이트리움 로세움 (Goldstein)(ATCC 28210); 및 자파노카이트리움 sp. (L1)(ATCC 28207)을 포함한다. 이외의 유전적인 변형을 위한 숙주 미생물로 적합한 예로는 이에 제한되는 것은 아니지만, 사카로미세스 세레비시애, 사카로미세스 칼스버젠시스 또는 칸디다, 클루베로미세스와 같은 다른 효모를 포함한 효모 또는 아스퍼질러스, 뉴로스포라, 페니실리움 등과 같은 섬사상의 진균류를 포함한다. 박테리아 세포도 숙주로 사용될 수 있다. 이에 제한되는 것은 아니지만, 발효 과정에 유용할 수 있는 대장균을 포함한다. 일례로 락토바실러스 종 또는 바실러스 종과 같은 숙주를 사용할 수도 있다.

[0195] 본 발명의 또 다른 한 양태는 유전적으로 변형된 식물에 관한 것으로서, 상기 식물은 다중불포화 지방산(PUFA) 폴리케타이드 신타제 (PKS) 시스템의 적어도 하나의 생물학적으로 활성인 도메인을 포함한 PKS 시스템을 재조으로 발현하기 위해 유전적으로 변형되었다. 상기 PUFA PKS 시스템의 도메인은 전술한(예, 스킨조카이트리움 및/또는 트라우스토카이트리움에 있어서) PUFA PKS 시스템과 관련된 상동체를 포함하여, 도메인을 포함하고, 또한 (진핵의 미생물 또는 다른 트라우스토카이트리드 미생물을 포함한) 비박테리아 미생물 유래의 PUFA PKS 시스템의 도메인 또는 상기된 미국특허출원 제10/124,800호에 개시된 스크리닝 방법으로 동정된 미생물 유래의 PUFA PKS 시스템의 도메인을 포함한다. 상기 식물은 또한 이에 제한되는 것은 아니지만, 박테리아의 PUFA PKS 또는 PKS 시스템, 타입 I PKS 시스템, 타입 II PKS 시스템, 모듈라 PKS 시스템, 및/또는 모든 비박테리아의 PUFA PKS 시스템(예로 들면 진핵의, 트라우스토카이트리드, 트라우스토카이트리아세 혹은 라비린툴라시에, 스킨조카이트리움 등)을 포함한 다른 PKS 시스템의 적어도 하나의 도메인 또는 생물학적으로 활성인 단편으로 추가로 변형될 수 있다.

[0196] 본 발명에서 유전적으로 변형된 식물은 고등식물, 특히 소모할 수 있는 식물 또는 본 발명의 목적인 생 활성 분자를 제조하는데 유용한 식물을 포함하여 유전적으로 변형된 식물을 포함할 수 있다. 상기 유전적으로 변형된 식물은 목적인 결과가 달성되도록(즉, PUFA PKS의 활성 및/또는 PKS 시스템을 사용하여 목적인 생산물의 제조를 증가 또는 변형시키기) 정상(야생형 또는 천연적으로 존재하는) 형태로부터 변형된(돌연변이 또는 변화시킨) 계통을 가지고 있다. 식물의 유전적인 변형은 전통적인 균주의 개발 및/또는 분자 유전학적 기술을 이용하여 달성될 수 있다. 목적인 아미노산 서열을 코딩하는 재조합 핵산 분자가 식물의 계통에 통합된 유전자 도

입 식물의 제조 방법이 당업계에 알려져 있다. 본 발명에 따라 유전적으로 변형시킬 식물은 사람을 포함한 동물에 의해 소비되는 식물이 바람직하다.

[0197] 본 발명에 따라 유전적으로 변형시킬 바람직한 식물(즉, 식물의 숙주세포)은 이에 제한되는 것은 아니지만, 고등식물, 특히 농작물, 특히 이의 기름이 이용되는 식물을 포함하여 소비가 가능한 식물을 포함한다. 상기 식물의 예는 다음을 포함한다: 카놀라, 콩, 평지씨, 아마씨, 옥수수, 홍화, 해바라기와 담배. 이외의 바람직한 식물은 약제학적 에이전트, 향발산 에이전트, 약효식품 에이전트, 기능성 식품의 성분 또는 화장용으로 활성인 에이전트로 이용되는 화합물을 제조하는 것으로 공지된 식물 또는 상기 화합물/에이전트를 제조하도록 유전적으로 변형된 식물을 포함한다.

[0198] 본 발명에 따른 유전적으로 변형된 미생물 혹은 식물은 재조합 기술을 사용하거나 전통적인 돌연변이 유발과 스크리닝 기술에 의해 변형된 미생물 또는 식물을 포함한다. 본 발명에서 유전자 발현, 유전자의 기능 또는 유전자 생산물(즉, 유전자에 의해 코딩된 단백질)의 기능의 감소를 야기하는 유전적인 변형은 유전자의 (완전한 또는 부분적인) 불활성화, 결실, 파괴(interruption), 차단 또는 하향 조절로 기재될 수 있다. 예로 들어 유전자에 의해 코딩된 단백질의 기능 감소를 야기하는 유전자에서의 유전적 변형은 유전자의 완전한 결실(즉, 유전자가 존재하지 않으므로 단백질이 존재하지 않는다), 단백질의 불완전하거나 전혀 해독되지 않도록 하는(즉, 단백질이 발현되지 않는다) 유전자에서의 돌연변이 또는 단백질의 천연 기능을 감소시키거나 파괴시키는(예로 들어, 효소의 활성 또는 작용이 감소되거나 전혀 없는 단백질이 발현된 경우) 돌연변이의 결과일 수 있다. 유전자의 발현 또는 기능에서의 증가를 초래하는 유전적 변형은 유전자의 증폭, 과잉 생산, 과잉 발현, 활성화, 개선, 첨가 또는 상향 조절로서 기재될 수 있다.

[0199] 본 발명에 따른 미생물 또는 식물의 유전적인 변형은 바람직하게는 PKS 시스템이 내인성이고 유전적으로 변형되었는지, 유기체(내생적인 시스템의 변형 여부에 따라)로의 재조합 핵산 분자의 도입이 내인성인지 또는 재조합 기술에 의해 완전히 제공되었는지 여부에 상관없이 미생물 또는 식물에 의해 발현된 PKS 시스템의 활성에 영향을 미친다. 상기 시스템을 발현하는 PUFA PKS 시스템 또는 유기체의 PUFA 생산물 프로필을 변경시키는 것은 유전적인 변형이 부재된(즉, 변형되지 않은 야생형 미생물 또는 식물, 또는 적어도 PUFA 합성에 관련해서는 변형되지 않은 미생물 또는 식물 - 즉, 유기체는 PUFA 합성과 관련되지 않은 다른 변형을 가질 수 있다) 경우와 비교하였을 때 숙주 미생물 또는 식물에 의한 하나 이상의 PUFA 제조에서의 검출 또는 측정할 수 있는 변화를 유발하는 것을 포함한다. PKS 시스템의 활성에 영향을 미치기 위해서는 유전적 변형이 부재된 경우와 비교하였을 때 유기체에 의해 발현되는 PKS 시스템에서의 검출하거나 측정할 수 있는 변화 또는 변형을 유발하는 유전적 변형을 포함한다. PKS 시스템 내에서의 검출할 수 있는 변화 또는 변형은 이에 제한되는 것은 아니지만 다음을 포함할 수 있다: 유전적 변형이 부재된 내인성 PUFA PKS 시스템과 비교하였을 때 변형된 PUFA PKS 시스템내 하나 이상의 도메인의 발현 및/또는 생물학적 활성의 변화 또는 변형(도입, 증가 또는 감소), 유기체가 검출/측정할 수 있는 PKS 시스템 활성을 가지도록(즉, 유기체가 유전적 변형 이전에는 PKS 시스템을 보유하고 있지 않다) 유기체로의 PKS 시스템을 도입, PKS 시스템 활동이 변형되도록 유기체에 의해 내인적으로 발현되는 PKS 시스템과는 상이한 PKS 시스템으로부터 유래된 기능적인 도메인을 유기체로 도입(예를 들면, 내생적으로 비 박테리아의 PUFA PKS 시스템을 발현하는 유기체로 박테리아의 PUFA PKS 도메인 또는 타입 I PKS의 도메인이 도입된다), PKS 시스템(예를 들면, 유전적 변형이 부재된 경우와 비교하였을 때 시스템이 더 많은(증가된 양) 또는 더 적은(감소된 양)을 제조한다)에 의해 제조되는 생활성 분자(예, PUFA)의 양에서의 변화, PKS 시스템에 의해 제조된 생활성 분자의 타입(예, PUFA 타입의 변화)에서의 변화(예를 들면, 시스템은 추가의 또는 상이한 PUFA, 신규하거나 상이한 생산물 또는 상기 시스템에 의해 천연적으로 제조되는 PUFA의 변이체 또는 다른 생산물을 제조한다), PKS 시스템에 의해 제조되는 다수의 생활성 분자의 비율에서의 변화(예를 들면, 시스템이 한 PUFA에 대한 다른 PUFA를 상이한 비율로 제조하거나, 유전적 변형이 부재된 경우와 비교했을 때 완전히 상이한 리피드 프로필을 제조하거나, 천연의 배형과 비교하였을 때 트리아실글리세롤 상에서 상이한 위치에 다양한 PUFA를 배치시킨다). 상기의 유전적 변형은 모든 타입의 유전적인 변형, 특히 재조합 기술과 전통적인 돌연변이에 의한 변형을 포함한다.

[0200] PUFA PKS 시스템의 기능성 도메인 또는 단백질의 활성을 증가시킨다는 것은 유기체의 다음과 같은 유전적 변형을 의미한다. 이 유기체는 도메인이나 단백질 시스템의 기능성의 증가를 가져오는 도메인이나 단백질(혹은 도메인이나 단백질이 도입되는)을 보유하고 도메인이나 단백질의 상위의 활동(예로 들면 특정한 활동 혹은 생체 안에서의 효소에 의한 활동), 도메인이나 단백질 시스템의 감소된 억제 혹은 퇴화, 그리고 도메인이나 단백질의 과잉 발현을 포함한다. 예로 들면, 유전자의 카피 수가 증가되거나, 본래의 프로모터에 비하여 높은 수준의 발현을 제공하는 프로모터의 사용으로 발현 수준을 증가시키거나, 유전자는 그 유전자에 의해 코딩된 단백

질 또는 도메인의 활동을 증가시키도록 유전자 공학 또는 전통적인 돌연변이 유발에 의해 변이될 수 있다.

[0201]

유사하게 PUFA PKS 시스템에 있는 기능성 도메인 또는 단백질의 활성을 감소시킨다는 것은 도메인 또는 단백질의 감소된 기능을 유발하는 도메인 또는 단백질을 포함한 (또는 도메인 또는 단백질이 도입될) 유기체에서의 임의의 유전적 변형을 의미하고, 도메인 또는 단백질의 감소된 활성, 도메인 또는 단백질의 증가된 억제 또는 분해, 및 도메인 또는 단백질의 발현의 감소 또는 제거를 포함한다. 예를 들어 본 발명에 따른 도메인 또는 단백질의 작용은 도메인 또는 단백질의 제조를 차단 또는 감소시키거나, 도메인 또는 단백질을 코딩하는 유전자 또는 이의 일부분을 "녹아웃(knocking out)"하거나, 도메인 또는 단백질 활성을 감소시키거나, 도메인 또는 단백질의 활성을 감소시킴으로써 감소될 수 있다. 도메인 또는 단백질의 제조를 차단 또는 감소시키는 것은 도메인 또는 단백질을 코딩하는 유전자를, 성장 배지 중 유도하는 화합물의 존재를 필요로 하는 프로모터의 조절 하에 위치시키는 것을 포함한다. 유도인자가 배지로부터 고갈되도록 조건을 설정함으로써 도메인 또는 단백질을 코딩하는 유전자의 발현이 턴오프(turn off)될 것이다. 본 발명자들은 실시예 부분에서 트라우스토카이트리드 미생물의 표적 유전자를 결실시키는 능력을 증명하였다. 도메인 또는 단백질의 활성을 차단 또는 감소시키는 것은 미국특허 제4,743,546호에 개시된 바와 유사한 절제(excision) 기술 방식을 사용하는 것을 또한 포함한다. 이러한 방식을 이용하기 위하여 대상 단백질을 코딩하는 유전자가 게놈으로부터 그 유전자의 특이적이고 조절된 절제를 가능하게 하는 특정 유전자 서열 사이에 클로닝한다. 절제는 예를 들어 미국특허 4,743,546호에서처럼 배양시에 배양 온도의 전환 또는 몇 가지 다른 물리적 또는 영양소의 신호에 의해 촉진될 수 있다.

[0202]

본 발명의 한 양태에서 유전적 변형은 본 발명에서 기술된 비박테리아의 PUFA PKS 시스템의 적어도 한 도메인(예로 들면 트라우스토카이트리드 숙주의 내인성 PUFA PKS 시스템의 한 도메인, 하나 이상의 도메인, 단백질, 전체 PUFA PKS 시스템)의 생물학적 활성을 가진 아미노산 서열을 코딩하는 핵산 서열의 변형을 포함한다. 이와 같은 변형은 내인적으로(자연적으로) 발현되는 비박테리아의 PUFA PKS 시스템 내의 아미노산 서열에 실행되어져, 그러한 시스템을 천연적으로 포함한 미생물은, 예를 들면 전통적인 돌연변이 및 선택 기술 및/또는 유전 공학 기술을 포함한 분자 유전학 기술에 의해 유전적으로 변형될 수 있다. 유전 공학 기술은 예로 들면, (실시예에서 보여지는) 내인성 유전자의 일부분을 결실하기 위해 표적의 재조합 벡터를 사용하거나 (실시예에서 보여지는) 이중의 서열로 내인성 유전자의 일부분을 대체하는 것을 포함한다. 숙주 게놈에 도입할 수 있는 이중의 서열의 예로는 다른 비박테리아의 PUFA PKS 시스템(예로 들면 다른 트라우스토카이트리드를 포함한 진핵세포 유래), 박테리아의 PUFA PKS 시스템, 타입 I PKS 시스템, 타입 II PKS 시스템 또는 모듈라 PKS 시스템과 같은 다른 PKS 시스템의 적어도 하나의 기능적 도메인을 코딩한 서열을 포함한다. 또한 이중의 서열은 숙주에 있는 전체 내인성 PUFA PKS 시스템(예, 내인성 PUFA PKS 시스템의 모든 유전자)을 대체하기 위해서 이용되는 전체 PUFA PKS 시스템(예, PUFA PKS 시스템과 관련된 모든 유전자)을 포함할 수 있다. 또한, 이중의 서열은 숙주 트라우스토카이트리드의 PUFA PKS 시스템에서 유래된 천연의 도메인의 변형된 기능적 도메인(상동체)을 코딩하는 서열(예를 들어, 스킴조카이트리움의 OrfB 유래의 변형된 도메인을 코딩한 핵산 서열, 여기서, 변형된 도메인은 스킴조카이트리움에서 발현되는 천연적으로 존재하는 도메인을 대체하기 위해 사용되었을 때 스킴조카이트리움에 의한 PUFA 산물 프로파일 변이시킬 것이다)을 포함한다. 숙주의 게놈에 도입되는 다른 이중의 서열은 PKS 시스템의 도메인은 아니지만 내인성 PKS 시스템의 성동에 영향을 끼치는, 기능적 도메인 또는 단백질을 코딩한 서열을 포함한다. 예를 들어, 숙주 게놈에 포스포판테테인 트랜스퍼라제를 코딩한 핵산 분자를 도입할 수 있다 (이하에서 설명됨). 내인성 PUFA PKS 시스템에 실행할 수 있는 특별한 변형을 이하에서 상세히 설명된다.

[0203]

본 발명의 다른 양태에서, 유전적 변형은 (1)PUFA PKS 시스템의 적어도 하나의 도메인의 생물학적인 활동을 가진 아미노산 서열을 코딩한 재조합 핵산 분자의 도입; 및/또는 (2)PUFA PKS 시스템의 활성에 영향을 미치는 단백질 또는 기능적 도메인을 코딩한 재조합 핵산 분자의 숙주로의 도입을 포함한다. 숙주는 다음을 포함할 수 있다: (1) PKS 시스템을 발현하지 않되, PKS 시스템의 모든 기능적 도메인이 숙주 세포로 도입된, 여기서 적어도 하나의 기능적 도메인은 비박테리아의 PUFA PKS 시스템으로부터 유래된 것을 특징으로 하는 숙주; (2) (내인성 또는 재조합으로) 비박테리아의 PUFA PKS 시스템의 적어도 하나의 기능적 도메인을 가진 PKS 시스템을 발현하되, 도입된 재조합 핵산 분자가 적어도 하나의 추가 비박테리아 PUFA PKS 도메인 또는 숙주의 PKS 시스템의 활성에 영향을 미치는 다른 단백질을 코딩하는 것을 특징으로 하는 숙주 세포; (3) (내인성 또는 재조합으로) 비박테리아 유래의 도메인을 기능을 필수적으로 포함하지 않은 PKS 시스템을 발현하되, 도입된 핵산 분자는 비박테리아 PUFA PKS 시스템의 적어도 하나의 기능적 도메인을 코딩하는 핵산 서열을 포함하는 것을 특징으로 하는 숙주 세포. 즉, 본 발명은 유전적으로 변형된 유기체(예, 미생물 또는 식물)를 포함하되, 상기 유기체는 (내인적 또는 재조합 변형에 의해 도입된) 적어도 하나의 비박테리아의 PUFA PKS 도메인 기능을 포함하며, 상기 유전적 변형은 유기체가 기능적 PKS 시스템을 포함할 때 비박테리아의 PUFA PKS 도메인 기능 또는

PKS 시스템에 측정할 수 있는 영향을 미친다.

[0204]

본 발명은 본 발명에서 기술된 PUFA PKS 시스템을 위한 가능한 숙주 세포로서 및/또는 본 발명에서 기술된 유전적 변형 및 방법에 이용되는 PUFA PKS 시스템 단백질 및 도메인을 코딩하는 추가의 유전적 물질을 위한 원천으로서 가능한 많은 비박테리아 및 박테리아의 미생물을 포함한다. 예를 들어, 실시예 1에서 설명되고 본 발명자들에 의해 발견된 트라우스토카이트리움 미생물과 같은 스킨조카이트리움에서 발견된 것과 유사한 PUFA PKS 시스템을 가진 미생물 유기체가 (미국특허출원 제10/124,800에 대응되는) 미국특허공개 제20020194641 호에 자세히 기술된, 다중불포화 지방산(PUFA) 폴리케타이드 신타제(PKS) 시스템을 가진 다른 비박테리아의 미생물을 동정하기 위한 방법에 따라 용이하게 동정/분리/스크리닝될 수 있다.

[0205]

본 발명에 따른 PUFA PKS 시스템을 스크리닝하기에 바람직한 타입의 미생물을 수집할 수 있는 장소로는 다음을 포함한다: 저산소 환경((필터 통과 유기체를 포함한)미생물 또는 미생물을 포함한 음식물을 소비하는 무척추동물의 위장을 포함한, 저산소 환경에 가까운 장소, (민물, 염수와 해수를 포함한) 무산소 또는 저산소의 수상 서식지 및 대양의 저산소 또는 저산소에 가까운 환경(영역). 상기 미생물 군주는 무기성 생물로만 적합하기보다는 호기성과 저산소 또는 산소가 결핍된 환경 모두에서 적응해 생존하는 것이 바람직하다. 호기성과 저산소 또는 산소가 결핍된 환경을 모두 가진 토양 환경은, 특히 수상 서식지 또는 일시적 수상 서식지에 있는 토양은 상기 유기체를 발견할 수 있는 탁월한 환경일 것이다.

[0206]

본 발명에 따라 숙주로서 및/또는 PUFA PKS 유전자의 원천으로 사용하기 위하여 스크리닝될 특히 바람직한 비박테리아의 미생물 군주는 생존 주기 중의 일부 동안에 파고사이토시스, 파고트로피시스(phagotrophic) 또는 엔도사이토시스 능력과 같은 메커니즘에 의해 전체 박테리아 세포를 소비(bacterivory)할 수 있고/거나 생존 주기 중에 아메바 단계 또는 프로토플라스타로서 존재하는 단계를 가지는(조류, 진균(효모), 프로토조아 또는 원생생물로 이루어진 그룹으로부터 선택된) 군주일 것이다. 이러한 영양물 섭취 방법은 실수가 발생한다면, 박테리아의 PKS 시스템을 진행 세포로 전달할 가능성을 증가시키고 박테리아의 세포(또는 이의 DNA)는 소화되지 않고 기능적으로 진행 세포로 통합된다.

[0207]

본 발명에서는 PUFA PKS 유전자(이에 의해 코딩된 단백질과 도메인)의 공급원으로서 본 발명에서 기술된 것 이외의 PUFA PKS 시스템을 포함하는 임의의 트라우스토카이트리드를 포함한다. 그러한 트라우스토카이트리드에는 이에 제한되는 것은 아니지만, 트라우스토카이트리움, 자파노카이트리움, 아플라노카이트리움, 트라우스토카이트리아시애 및 라비린톨라 내의 엘리나 및 스킨조카이트리움, 라비린톨로이드 및 라비린톨라시애 내의 라비린토크사를 포함한 속을 포함한 트라우스토카이트리아세 및 라비린톨라시애 과에 있는 임의의 미생물을 포함한 트라우스토카이트리얼 목의 미생물을 포함한다. 상기 속 내의 바람직한 종은 이에 제한되는 것은 아니지만, 라비린톨라 sp., 라비린톨라 알게리엔시스, 라비린톨라 시엔콥스키, 라비린톨라 샤토니, 라비린톨라 코엔카 이시스티스, 라비린톨라 마크로시스티스, 라비린톨라 마크로시스티스 아틀란티카, 라비린톨라 마크로시스티스 마크로시스티스, 라비린톨라 마그니피카, 라비린톨라 미누타, 라비린톨라 로스코벤시스, 라비린톨라 발카노비, 라비린톨라 비텔리나, 라비린톨라 비텔리나 파시피카, 라비린톨라 비텔리나 비텔리나, 라비린톨라 줍피를 포함한 라비린톨라 내의 모든 종들; 라비린톨로이데스 sp., 라비린톨로이데스 미누타, 라비린톨로이데스 스킨조카이트로프스를 포함한 모든 라비린톨로이데스 종; 라비린토크사 sp., 라비린토크사 폴리아, 라비린토크사 사우바게 유를 포함한 모든 라비린토크사 종, 아플라노카이트리움 sp. 와 아플라노카이트리움 케르구엘렌시스를 포함한 모든 아플라노카이트리움 종; 엘리나 sp., 엘리나 마리살바, 엘리나 시노리피카를 포함한 모든 엘리나 종; 자파노카이트리움 sp., 자파노카이트리움 마리눔을 포함한 모든 자파노카이트리움 종; 스킨조카이트리움 sp., 스킨조카이트리움 아그레가툼, 스킨조카이트리움 리마시눔, 스킨조카이트리움 미누툼, 스킨조카이트리움 옥토스포룸을 포함한 모든 스킨조카이트리움종; 트라우스토카이트리움 sp., 트라우스토카이트리움 아그레가툼, 트라우스토카이트리움 아루디멘탈레, 트라우스토카이트리움 오레움, 트라우스토카이트리움 벤티콜라, 트라우스토카이트리움 글로버섬, 트라우스토카이트리움 킨네이, 트라우스토카이트리움 모티툼, 트라우스토카이트리움 파카이더문, 트라우스토카이트리움 프로리페움, 트라우스토카이트리움 로제움, 트라우스토카이트리움 스트리아툼, 울케니아 sp., 울케니아 미누타, 울케니아 프로판다, 울케니아 라디에이트, 울케니아 사카리아나 및 울케니아 비수르젠시스를 포함한 트라우스토카이트리움 종.

[0208]

이론과 상관없이 본 발명자들은 라비린톨라시애가 PUFA PKS 유전자를 보유하는 것으로 공지된 트라우스토카이트리아세와 밀접하게 관련되어 있고, 라비린톨라시애는 박테리오포로스/파고사이토시스 하는 것으로 공지되어 있고, 라비린톨라시애의 일부 멤버는 PUFA PKS 시스템 가진 것과 일관된 지방산/PUFA 프로필을 가지고 있기 때문에 라비린톨라와 다른 라비린톨라시애를 PUFA PKS 유전자의 공급원으로 간주한다.

- [0209] (특히, 파괴사이토시스 또는 엔도사이토시스에 의해) 박테리오포리를 할 수 있는 (트라우스토카이트리드의 멤버가 아닌) 미생물 균주는 (예시의 속에 포함되지만 이에 제한적이지는 않은) 다음 미생물의 부류에서 발견될 수 있다:
- [0210] (Stramenopiles를 포함한) 조류 및 이와 유사한 미생물에서: 유글레노파이시에 부류 (예, 유글레나속과 페라네마속), 크라이소파이시에 부류(예, 오크로모나스속), 디노브라이아시에 부류 (예, 디노브라이온, 플라타이크라이시스와 크라이소크로몰리나속), (크라이데코디늄, 짐노디늄, 페리디늄, 세라티움, 자이로디늄 및 옥시르피스속을 포함한) 디노파이시에 부류, 크립토파이시에 (예, 크립토모나스와 로도모나스속), 크산토파이시에 부류 (예, 올리스토티스쿠스속)(그리고 아메바의 단계가 편모충 리조클로리다시에서처럼 존재하는 조류의 형태 및 아파노캐트 파스케리, 부밀레리아스티게오클로늄과 바우케리아 제미나타의 유주자/배우자(gamete)을 포함한), 유스티가토파이시에 부류, (프라임네슘과 디아크로네마속을 포함한) 프라임네시오파이시에 부류.
- [0211] Stramenopiles에서 다음을 포함한다: 프로테로모나드, 오팔린, 데벨로파엘라, 디프로포리, 라빌린톨로이드, 트라우스토카이트리드, 비코세키즈, 우마이세트, 하이포크라이크리디오마이세트, 코마티온, 레티쿨로스테라, 펠라고모나스, 펠라코포쿠스, 올리콜라, 아우레코쿠스, 파라레스, 라피디오파이트, 사이누리드, 리조크로몰리날레스, 페디넬알레, 디티오칼레스, 크라이소머리달레스, 사르시노크라이시달레스, 하이드루랄레스, 히버디알레스 및 크로몰리날레스.
- [0212] 진균류에서: 믹소마이세트 부류 (점액 아메바 형태) -- 점균류, 아크라시애목(예, 사피니아 속)을 포함한 아크라시애 부류, 구틀리나시애(예, 구틀리노프시스와 구틀리나속) 부류, 디탁이스텔리아시애(예, 아크라시스, 디타이오스텔룸, 폴리스폰딜룸 및 코예노니아 속) 부류, 크라이트리디알레스, 안크라일리스탈레스, 블라스토클라디알레스, 모노블레파리달레스, 사프로레그니알레스, 페로노스포탈레스, 무코랄레스 및 엔토모프톨라레스목을 포함한 피코마이세스 부류.
- [0213] 프로토조아에서: (파괴사이토시스를 포함한) 박테리오포리를 할 수 있는 생존 단계를 가진 프로토조아 균주가 섬모충, 편모충, 아메바로 분류된 타입으로부터 선택될 수 있다. 섬모충 프로토조아는 코노트리크, 콜포디드, 카이트토포레스, 합토리드, 카리오렐릭트, 올리고하이메노포, 폴리하이메노포(스피로트리크), 프로스토메스와 숙토리아 그룹을 포함한다. 편모충 프로토조아는 비오숙시드, 보도니스, 세르코모나드, 크라이소파이트(예를 들면 안소파이사, 크립사아메바, 크립토파엘라, 덴드로모나스, 디노브라이온, 말로모나스, 오크로모나스, 파라파이스모나스, 포테리우크로모나스, 스푸멜라, 신크립타, 시누라 및 우로글레나 속), 콜라 플라겔레이트, 크립토파이트(예로 들면 킬로모나스, 크립토모나스, 시아노모나스 및 고니오모나스 속), 디노플라겔레이트, 디플로모나드, 유글레노이드, 헤테로로보새, 페디넬리드, 펠로비온트, 팔란스테리드, 슈도덴드로모나드, 스포고모나드 및 볼보칼레스 (그리고 아토디스쿠스, 클라우트리아비아, 헬케시마스틱스, 카사블레파리스 및 멀티실라의 지정되지 않은 편모충 속을 포함한 다른 편모충)를 포함한다. 아메바 프로토조아는 악티노파이라드, 센트로헬리드, 데스모토리시드, 디플로파이라드, 움아메바, 헤테로보새, 랩토믹시드, 뉴클레아리드 필로세 아메바, 펠레비온트, 테스타테 아메바 및 밤파이렐리드를 포함한다(그리고 짐노프리스, 비오믹사, 마이크로코미드, 레티쿨로믹사, 벨로노시스티스, 엘레오르하니스, 알레오그로미아, 그로미아 또는 라이버쿠니아의 지정되지 않은 아메바 속을 포함한다). 프로토조아 속은 퍼콜로모나데, 헤테로로보새, 라이로모나데, 슈도실리아타, 트리코모나데, 하이퍼마스티게, 헤테로미테아, 텔로네매, 카이아토보도내, 에브리데, 피토믹새, 오팔리데, 카네토모나데, 헤미나스티게, 프로토슬레래, 믹사가스트래, 디토이오스텔래, 코아노모나데, 아피코모나데, 예오그레가리내, 네오그레가리내, 코엘로트룰레, 유코키데, 해모스포래, 피로플라스매, 스피록트리케, 프로스토마테, 리토스토마테, 필로파린게, 나소포래, 올리고하이메노포래, 콜포데, 카이로렐릭타, 뉴클레오헬, 센트로헬레, 아칸타래야, 스티코론케, 폴리시스티내, 패오다래, 로보새, 필로새, 아탈라매, 모노탈라매, 폴리탈라매, 섹노파이오포래, 스킴조클라데, 홀로새, 엔트아메바, 믹스포래, 악티노믹새, 할로스포래, 파라믹새, 롬보조아 및 오르소넥테를 포함한다.
- [0214] 본 발명의 바람직한 양태는 앞서 나열된 바람직한 서식지 중 하나로부터 수집된 상기된 미생물의 균주를 포함한다.
- [0215] 본 발명의 방법의 몇 가지 양태에서 박테리아의 PUFA PKS 시스템은 (전체 PUFA PKS 시스템, 단백질 및/또는 도메인을 코딩하는) 그의 유전자 및 일부분을 포함하여, 다른 PUFA PKS 시스템(예, 비박테리아의 PUFA PKS 시스템) 및/또는 본 발명의 양태에서 동일한 (또는 그와는 다른) 것을 포함한 미생물을 유전적으로 변형시키는 데 사용할 수 있다. 한 측면에서, 유전적으로 변형된 미생물(예, 유전적으로 변형된 트라우스토카이트리드)을 생성하는데 특히 유용할 것으로 예상되는 신규한 PUFA PKS 시스템 및/또는 본 발명에 따른 유전적으로 변형된 미생물과 식물 및 방법에서 유용한 PUFA PKS 시스템을 코딩하는 신규한 하이브리드 작제물을 박테리아에서 동정

할 수 있다. 한 측면에서, 본 발명에 따른 양태에서 특히 유용한 박테리아는 PUFA PKS 시스템을 가지며, 상기 PUFA PKS 시스템은 약 20℃ 이상, 바람직하게는 25℃ 이상, 보다 더 바람직하게는 30℃ 이상에서 PUFA를 제조할 수 있다. 전술된 미국특허 제 6,140,486호에서 개시된 해양 박테리아인 *쉐와넬라*와 *비브리오 마리아누스*는 더 높은 온도에서 PUFA를 제조하지 않아 이러한 박테리아로부터 유래된 PUFA PKS 시스템의 실용성은, 특히 필드(field)의 조건하의 식물에 적용할 때, 제한된다. 따라서 한 양태에서, 본 발명의 스크리닝 방법은 박테리아가 더 높은 온도(예로 들면 약 15℃, 20℃, 25℃ 또는 30℃ 또는 그 이상)에서 성장과 PUFA 제조할 수 있는 PUFA PKS 시스템을 가진 박테리아를 동정하는데 이용될 수 있다. 그러나 박테리아 원천이 상기 높은 온도에서 잘 성장되지 않고/PUFA를 제조하지 못할지라도, 본 발명은 상기 박테리아 유래의 PUFA PKS 시스템이 약 15℃, 20℃, 25℃ 또는 30℃ 또는 그 이상의 온도에서 효소/생물학적 활성을 가지고 있는 한, 상기 PUFA PKS 시스템(유전자 및 이에 의해 코딩된 단백질/도메인)의 동정, 분리 및 이용을 포함한다. 본 양태의 다른 측면에서 (항진균성의) 나이스틴(nystatin), 또는 사이클로헥시미드(cycloheximide: 진핵세포의 단백질 합성 억제제)와 같은 진핵생물의 성장의 억제제가 해양 또는 하구와 같은 타입의 서식지/적소 또는 상기 박테리아가 발견될 수 있는 다른 서식지에서 수집된 물 샘플/토양 샘플 유래의 초기 균주를 배양하고/선택하는데 이용되는 아가 플레이트에 첨가될 수 있다. 이러한 과정은 진핵의 균주의 오염 없이 (또는 최소화된 오염으로) 박테리아 균주를 풍부하게 선택할 수 있게 한다. 상기 선택 과정은 상승된 온도 (예, 30℃)에서 플레이트를 배양하고 적어도 하나의 PUFA를 제조하는 균주를 선택하는 것과 조합되어져, (약 20℃ 이하, 보다 바람직하게는 약 5℃ 이하의 낮은 온도에서만 PUFA 제조를 보여주는 이전의 박테리아의 균주와는 달리) 상승된 온도에서 작동하는 PUFA PKS 시스템을 가진 후보 박테리아 균주를 동정할 것이다.

[0216]

그러나 높은 온도에서 잘 자라지 않는(또는 전혀 자라지 않는) 박테리아 또는 높은 온도에서 저어도 하나의 PUFA를 제조하지 않는 박테리아에서, 상기 균주는 임의의 조건에서 PUFA를 제조하는 박테리아의 능력을 동정하고/거나 박테리아 또는 비박테리아의 유기체에서 유래된 다른 공지된 PUFA PKS 유전자와 상동성을 가진 유전자에 대하여 상기 박테리아의 게놈을 스크리닝함으로써 PUFA PKS 시스템을 포함하는 것으로서 동정되고 선택될 수 있다(실시예 7 참조). 임의의 박테리아의 원천에서 유래된 PUFA PKS 기능을 평가하기 위하여 세포가 제거된 추출물을 제조하고 다양한 온도에서 PUFA 제조 여부에 대하여 테스트한 후, 이어서 높은 온도(예, 15℃, 20℃, 25℃ 또는 30℃ 또는 그 이상)에서 효소의/생물학적 활성을 가진 PUFA PKS 유전자를 포함한 미생물을 선택할 수 있다.

[0217]

유전적 변형을 위한 숙주로 사용하기에 적합한 박테리아는 앞서 기술된 모든 박테리아의 균주를 포함한다. 특히 본 발명에 따른 유전적으로 변형된 서열과 유기체의 제조를 위한 PUFA PKS 유전자(그리고 이에 코딩된 단백질과 도메인)의 원천으로 사용하기에 적합한 박테리아는 PUFA PKS 시스템을 포함하는 박테리아를 포함한다. 이러한 박테리아는 전형적으로 바다 또는 하구의 서식지로부터 분리되고, PUFA를 제조하는 능력에 의해 및/또는 상기 유기체내에 공지된 PUFA PKS 유전자와 상동성을 가진 적어도 하나 이상의 유전자의 존재에 동정될 수 있다. 이러한 박테리아는 이에 제한되는 것은 아니지만, *쉐와넬라* 및 *비브리오* 속의 박테리아를 포함한다. 본 발명에 사용하기 바람직한 박테리아는 높은 온도(예, 15℃, 20℃, 25℃ 또는 30℃ 또는 그 이상)에서 생물학적으로 활성인 PUFA PKS 시스템을 가진 것을 포함한다. 본 발명자들은 PUFA PKS 유전자의 원천으로 사용하기에 특히 적합할 두 개의 예시적인 박테리아(예, *쉐와넬라 올레아나* 및 *쉐와넬라 자포니카*; 실시예 7 및 8 참조)를 동정하였고, 다른 박테리아가 용이하게 동정될 수 있고 PUFA PKS 유전자를 포함한 것으로 공지되어 있으며 본 발명의 한 양태(예, *쉐와넬라 겔리디마리나*)에서 유용할 것이다.

[0218]

추가로, 모든 박테리아 또는 비박테리아의 미생물이 천연의 서식지에서 용이하게 배양될 수 있는 것은 아니다. 그러나 이와 같이 배양할 수 없는 미생물의 유전자적인 특징은, 혼합된 또는 천연 그대로의 샘플로부터 엔 매스(en mass) 준비한 DNA로부터 유전자를 분리함으로써 평가될 수 있다. 본 발명에 특히 적합한 배양할 수 없는 미생물로부터 유래된 PUFA PKS 유전자는 공지된 PUFA PKS 유전자에 있는 높은 상동성의 영역을 매치하도록 일반적으로 고안된 프라이머를 사용하여 디제너러티브(degenerative) PCR함으로써 환경적(environmental) DNA 샘플로부터 분리될 수 있다(예, 실시예 7 참조). 대안으로는, DNA 단편의 전체가 당업계에 공지된 몇 가지 방법에 의해 정제된 환경적 DNA로부터 클로닝될 수 있다. 이렇게 입수된 DNA 단편의 서열은 PUFA PKS 유전자와 같은 공지된 유전자에 상동성을 밝혀줄 수 있다. OrfB 와 OrfC (예를 들면, 스킴조카이트리움 과 트라우스토카이트리움의 도메인 구조를 참조하여)의 상동체가 PUFA PKS 최종 생산물을 정의하는데 특히 유용할 것이다. PUFA PKS 유전자의 전체 코딩 영역들은 상호간 또는 공지된 PUFA PKS 유전자와 함께 또는 유전자 단편의 조합물과 조합으로 숙주 유기체(예, 대장균 또는 효모)에서 발현되어 PUFA 제조에 미치는 이들의 영향을 평가할 수 있다. 전술된 바와 같이, 세포가 제거된 추출물에서의 활성이 목적인 온도에서 기능을 결정하는데 사용될 수 있다. 또한, 분리된 PUFA PKS 유전자가 적절한 스킴조카이트리움 또는 기능을 측정하기 위해 적합한 균주로 직접적

으로 형질전환될 수 있다. 이러한 방식으로 동정 또는 제조된 PUFA PKS 시스템을 코딩하는 작제물은 하이브리드 작제물을 포함하여, 식물과 같은 다른 유기체를 형질전환시키는데 이용할 수 있다.

[0219] 따라서, 예를 들어 트라우스토카이트리드 PUFA PKS 시스템 뿐만 아니라 박테리아의 PUFA PKS 시스템과 PKS 시스템을 사용하는 것과 같이, 본 발명에 따른 비박테리아의 PUFA PKS 시스템을 사용한 유전자의 혼합은 (하기에서 상세히 설명되는) EPA, DHA, ARA, GLA, SDA 등을 포함한 PUFA 생산물 범위를 확장할 뿐만 아니라 항생제, 다른 약제학적 화합물 및 다른 적합한 생산물을 포함한 광범위한 생활성 분자를 제조하는데 이용될 수 있다. 이와 같은 생활성 분자를 수득하기 위한 방법은 다양한 유기체로부터의 유전자를 혼합하는 것 뿐만 아니라 본 발명에서 기술된 비박테리아의 PUFA PKS 유전자를 유전적으로 변형하는 다양한 방법도 포함한다. 본 발명에 따른 비박테리아의 PUFA PKS 시스템의 유전적 기초와 도메인 구조에 관한 지식은 다양한 생활성 분자를 제조하는 신규한 유전적으로 변형된 유기체를 고안하는 기초를 제공한다. PKS 도메인 및 관련된 유전자의 혼합과 변형이 본 발명자들에 의해 고려되었지만, 예시를 통해, PUFA-PKS 시스템의 다양한 가능한 조각이 유전적 변형과 대생물 분자 제조와 관련하여 하기에서 상세히 설명된다.

[0220] 따라서 본 발명은 본 발명에 따른 비박테리아의 PUFA PKS 시스템의 적어도 하나의 기능적 도메인의 생물학적 활성을 가진 아미노산 서열을 코딩하는 유기체에서 적어도 하나의 핵산 서열을 유전적으로 변형시키고/거나 그러한 아미노산 서열을 코딩하는 핵산 서열을 포함하는 적어도 하나의 재조합 핵산 분자를 발현시킴으로써 미생물 또는 식물의 세포를 유전적으로 변형시키는 방법을 포함한다. 그러한 서열의 다양한 양태, 유기체를 유전적으로 변형시키는 방법 및 특정한 변형이 앞서 기술되었다. 전형적으로 상기 방법은 특히 생활성 분자를 제조하는 유전적으로 변형된 유기체를 제조하는데 이용된다.

[0221] 본 발명의 한 양태는 유전적으로 변형된 트라우스토카이트리드 미생물에 관한 것으로, 상기 미생물은 내인성 다중불포화 지방산 (PUFA) 폴리케티드 신타제(PKS) 시스템을 가지며, 상기 내인성 PUFA PKS 시스템은 변형되지 않은 트라우스토카이트리드 미생물과 비교하였을 때 미생물에 의한 다중불포화 지방산(PUFA)의 발현 프로필을 변이시키기 위해 유전적으로 변형된 것을 특징으로 한다. 본 발명에서 숙주 유기체로서 유용한 트라우스토카이트리드 미생물은 내생적으로 PUFA PKS 시스템을 가지며 발현한다. 유전적 변형은 내인성 PUFA PKS 시스템의 하나 이상의 기능적인 도메인의 유전적 변형일 수 있으며, 이로써 상기 변형은 내인성 PUFA PKS 시스템의 PUFA 제조 프로필을 변이시킨다. 추가로 또는 대안으로 유전적 변형은 미생물로의, 적어도 하나의 외인성 핵산 서열(예, 재조합 핵산 분자)의 도입일 수 있는데, 여기서 상기 외인성 핵산 서열은 제 2의 PKS 시스템에서 유래된 적어도 하나의 생물학적으로 활성인 도메인 또는 단백질 및/또는 PUFA PKS 시스템의 활성에 영향을 미치는 단백질(예, 포스포판테테인 트랜스퍼라제(PPTase))을 코딩한다. 상기 제 2의 PKS 시스템은 다른 PUFA PKS 시스템을 포함한 그리고 유전적으로 변형될 트라우스토카이트리드 PUFA PKS 시스템 유래의 유전자의 상동체를 포함한 임의의 PKS 시스템일 수 있다.

[0222] 본 발명의 이러한 양태는 본 발명자들에 의한 PUFA가 풍부한 리피드(멤브레인 관련 포스포리피드(PL) 뿐만 아니라 트리아실글리세롤(TAG))를 효과적으로 제조하는 유전적으로 변형된 미생물 및 방법의 개발을 통하여 EPA와 같은 목적인 PUFA가 풍부한 상업적으로 가치있는 리피드의 제조에 특히 유용하다.

[0223] 본 발명의 이러한 특정 양태는 다음 지식으로부터 일부 유추된 것이다:(1) 선택된 미생물, 특히 본 발명에서 기술된 상업적으로 개발된 스킨조카이트리움 균주와 트라우스토카이트리드의 내재된 TAG 제조 능력의 활용; (2) 진핵세포, 트라우스토카이트리얼 목의 멤버에서의 PUFA PKS 생합성 경로(즉, PUFA PKS 시스템)에 대한 본 발명자들의 상세한 이해; (3)스킨조카이트리움에서의 상동성의 유전적 재조합 시스템의 활용. 관련된 본 발명자들의 시스템에 대한 지식을 기반으로 하여, 동일한 일반적 방식이 EPA가 아닌 PUFA를 제조하기 위하여 개발될 수 있다.

[0224] 본 발명의 한 양태에서, 정상적으로 DHA 및 DPA를 제조하는 PUFA PKS 효소를 코딩하는 스킨조카이트리움 유전자와 같은 내인성 트라우스토카이트리드 PUFA PKS 유전자가 무작위적 또는 표적 돌연변이에 의해 변형되거나, 다른 유기체(예, 박테리아 또는 다른 원천)에서 유래된 상동성의 PKS 단백질을 코딩하는 유전자로 대체되거나, 유전적으로 변형된 스킨조카이트리움, 트라우스토카이트리움 또는 다른 트라우스토카이트리드 PUFA PKS 유전자로 대체된다. 이렇게 도입 및/또는 변형된 유전자에 의해 코딩된 효소의 생산물은 예를 들면 EPA이거나, 다른 PUFA를 포함한 다른 관련된 분자일 수 있다. 이러한 방법의 한 특징은 PUFA의 효율적인 제조 및 PUFA의 PL과 TAG로로의 통합에 필수적인 트라우스토카이트리드 PUFA 합성의 내인성 요소와 축적 기구들을 활용하는 것이다. 특히, 본 발명의 이러한 양태는 유기체에 의해 제조되는 PUFA의 타입에 변형을 지시하지만, 부모 균주에 의한 높은 오일 제조를 유지한다.

[0225] 하기 일부 설명은 예시적인 숙주 유기체로서 유기체 스킨조카이트리움을 사용하지만, 트라우스토카이트리움, 라비린톨로이드스 및 자포포노카이트리움 속의 멤버를 포함한 임의의 트라우스토카이트리드가 본 발명에 따라 변형될 수 있다. 예를 들어, 트라우스토카이트리움 PKS 유전자가 스킨조카이트리움과 같은 다른 트라우스토카이트리드의 내인성 PUFA PKS 유전자를 변형시키는데 이용될 가능성이 높지만, 트라우스토카이트리움 종과 관련하여 PUFA PKS 시스템을 코딩한 유전자가 동정되었고(실시예 6 참조), 상기 유기체는 본 발명에서 기술된 방법을 사용하여 유전적 변형을 위한 숙주 유기체로 사용될 수 있다. 더 나아가, 미국특허출원 제10/124,800호에서 개시된 바와 같이 유기체를 스크리닝하는 방법을 사용하여 본 발명에 유용한 유기체를 동정할 수 있고, 그러한 유기체는 모두 본 발명에 포함된다.

[0226] 본 발명의 한 양태는 다음과 같이 예시될 수 있다. 일례로 스킨조카이트리움에서의 PUFA 합성과 축적에 대한 본 발명자들의 현재의 이해를 기반으로 하여 전반적인 생화학 과정은 세 부분으로 나뉠 수 있다.

[0227] 첫 번째, 스킨조카이트리움 오일 (DHA 와 DPA)로 축적한 PUFA는 상기된 PUFA PKS 시스템의 생산물이다. 스킨조카이트리움에서 PUFA PKS 시스템은 말로닐-CoA를 상당량의 중간체 화합물의 방출 없이 최종 생산물인 PUFA로 전환시킨다. 스킨조카이트리움에서 PUFA의 실제적인 합성에 필요한 것으로 알려진 효소의 도메인 모두를 코딩하는 세 개의 유전자(Orfs A, B 및 C; 또한 서열번호: 1, 서열번호: 3 및 서열번호: 5로 각각 나타내었다)가 동정되었다. (효소 도메인의 상동성 세트를 포함한 단백질을 코딩하는) 유사한 유전자 세트가 클로닝되었고 PUFA를 제조하는 다른 비진핵의 유기체, 즉 몇 개의 해양 박테리아 균주로부터 확인되었다. 또한, 본 발명자들은 (이하에서 상세히 설명되는) 적어도 하나의 다른 해양 원생생물(트라우스토카이트리움 계통 23B)의 PUFA PKS 유전자를 동정하고 서열분석하였다.

[0228] 해양 박테리아의 PUFA 생산물은 (미국특허 제6,140,486호; 미국특허 제6,566,583호에 개시된)(지금은 모리텔라 마리나로 알려진 비브리오 마리누스에 의해 제조되는) DHA 뿐만 아니라 (예를 들어, 쉼와넬라 SRC2738과 포토박터 프로판툼에 의해 제조되는) EPA를 포함한다. 발효 조건에서 제조 유기체(예, 스킨조카이트리움)의 생리학적 성장 요구조건이 충족되는 한, 임의의 PUFA PKS 유전자 세트는 본 발명의 실시예에서 기술된 스킨조카이트리움 유전자를 대신하도록 구상될 수 있다는 것이 본 발명의 한 양태이다. 특히, 상기된 PUFA-제조 박테리아의 균주가 비교적 낮은 온도(일반적으로 20℃이하)에서만 성장하며, 이는 추가로 이들의 PUFA PKS 유전자 생산물이 스킨조카이트리움의 표준 성장 온도(25-30℃)에서는 기능을 하지 않을 것임을 시사한다. 그러나 본 발명자들은 최근에 스킨조카이트리움 및 다른 트라우스토카이트리드(실시예 7 참조)의 표준 성장 온도에서 성장하고 EPA를 제조하는 적어도 두 개의 다른 해양 박테리아를 확인하였다. 상기 대체 해양 박테리아는 본 발명에서 기술된 방법에 의해 스킨조카이트리움 및 다른 트라우스토카이트리드의 변형을 위한 재료로 사용될 수 있는 PUFA-PKS 유사 유전자를 보유한 것으로 여겨진다. 따라서 현재 연구되지 않거나 동정되지 않은 PUFA-제조 박테리아가 트라우스토카이트리드의 변형에 유용한 PUFA PKS 유전자를 또한 포함할 수 있음이 당업자에는 자명할 것이다.

[0229] 두 번째, PUFA 합성에 직접적으로 관련된 효소를 코딩하는 유전자 외에 "보조" 효소가 필요하다. 유전자는 포스포판테테인 트랜스퍼라제(PPTase)를 코딩하고 이는 PUFA PKS 복합체에 존재하는 아실-캐리어 단백질(ACP) 도메인을 활성화한다. 이러한 보조인자의 추가에 의한 ACP 도메인의 활성화는 PUFA PKS 효소 복합체가 기능을 발휘하는데 필요하다. 지금까지 확인된 PUFA PKS 시스템의 모든 ACP 도메인이 높은 수준의 아미노산 서열 보존성을 보임에 따라 이론에 관계없이, 본 발명자들은 스킨조카이트리움과 다른 트라우스토카이트리드의 PPTase가 다른 PUFA PKS 시스템의 ACP 도메인을 인식하고 활성화할 것으로 예상한다. 이중의 PPTases와 PUFA PKS 유전자가 PUFA 생산물을 제조하기 위해 같이 기능한다는 원칙의 증거로, 본 발명자들은 스킨조카이트리움 유래의 PUFA PKS 유전자와 두 가지 상이한 이중성 PPTase를 이용하여 박테리아의 숙주세포에서 PUFA를 제조하는 것을 증명한다.

[0230] 세 번째로, 스킨조카이트리움에서 PUFA PKS 시스템의 생산물은 효과적으로 포스포리피드(PL)와 트리아실글리세롤(TAG) 모두로 보내진다. 본 발명자들의 데이터는 PUFA가 PKS 복합체의 ACP 도메인으로부터 조효소 A(CoA)로 전달되었다는 것을 보여준다. 다른 진핵의 유기체에서, 상기 아실-CoA는 PL 과 TAG 분자를 형성하는 다양한 아실-트랜스퍼라제의 기질로서 작용한다. 대조적으로, 상기 데이터는 박테리아에서는 CoA로의 전달이 일어나지 않으며; 오히려 PKS 복합체의 ACP도메인으로부터, PL을 형성하는 아실-트랜스퍼라제의 직접적인 전달이 있었다. PUFA를 ACP로부터 CoA로 전달하는 스킨조카이트리움의 효소 시스템은 DHA와 DPA 모두를 인식할 수 있고, 그래서 본 발명자들은 (PUFA PKS ACP 도메인에 부착된) PUFA PKS 시스템의 PUFA 생산물이 기질로서 작용하는 것은 예측할 수 있다고 여긴다.

- [0231] 따라서, 본 발명의 한 양태에서, 본 발명자들은 PUFA PKS 효소 복합체의 성분(첫번째 부분)를 코딩하는 유전자를 변이시키는 한편, 스킨조카이트리움 또는 다른 트라우스토카이트리드 숙주 유래의 내인성 PPTase(두번째 부분) 및 PUFA-ACP에서 PUFA-CoA로의 트랜스퍼라제 활성 및 TAG/PL 합성 시스템(또는 내인성 PUFA ACP 내지 TAG/PL 메카니즘)(세번째 부분)을 활용할 것을 제안한다. 본 발명에 따른 이러한 방법은 실험 데이터에 의해 지지되고, 이의 일부는 실시예 부분에서 자세히 설명된다.
- [0232] 첫 번째, 본 발명자들은 PUFA PKS 시스템이 유기체간에 전달될 수 있고, 일부는 상호교환적임을 발견하였다. 특히, 앞에서 기술한 것 같이 웨와넬라 SCR2738 (Yazawa, 1996, *lipids* 31:S297-300)와 비브리오 마리누스(웨와넬라의 PPTase와 더불어)(미국특허 제6,140,486호) 같은 해양 박테리아의 PUFA PKS 경로가 이종의 숙주(즉, 대장균)에 성공적으로 전달될 수 있다는 것을 앞서 보여주었다. 또한, 상기 두 가지 유기체(웨와넬라 SCR2738과 비브리오 마리누스)에서 유래된 PUFA PKS 효소의 서브유닛 간의 구조적 상동성의 정도는 두 시스템의 유전자를 혼합하고 매치하는 것이 가능할 정도이다(미국특허 제6,140,486호). 혼합된 유전자의 세트의 PUFA 최종 생산물은 특정한 유전자 상동체의 기원에 따라 달라진다. 적어도 하나의 오픈 리딩 프레임(웨와넬라의 Orf7 및 이의 비브리오 마리누스 상동체; 미국특허 제6,140,486호의 도 13 참조; 상기 Orf의 명명법이 달라졌음을 주목; 상기 특허에서는 Orf8로 표기되었지만, Genbank에 Orf7로 제출되어, GenBank에 의해 그와 같이 지칭된다)이 DHA 또는 EPA가 복합적인 시스템의 생산물이 될지를 결정하는 것과 연관될 수 있다. 지금까지 동정된 모든 PUFA PKS 효소의 기능성 도메인은 서로 서열 상동을 보여준다. 이와 유사하게, 상기 데이터는 해양 박테리아에서 유래된 PUFA PKS 시스템을 포함한 시스템이 스킨조카이트리움과 다른 트라우스토카이트리드로 전달되고, 기능할 수 있음을 보여준다.
- [0233] 본 발명자들은 대장균에서 스킨조카이트리움 유래의 PUFA PKS 유전자 (Orfs A, B 및 C)를 발현하고, 상기 세포는 스킨조카이트리움의 이러한 PUFA의 내인성 제조와 거의 동일한 비율로 DHA와 DPA를 만든다는 것을 증명했다(실시예 2 참조). 따라서 재조합 스킨조카이트리움 PUFA PKS 유전자가 기능성의 PUFA 합성 시스템을 코딩한다는 것이 증명되었다. 또한, 트라우스토카이트리움 23B OrfA 및 OrfC 유전자의 전부 혹은 일부가 스킨조카이트리움에서 기능을 발휘한다는 것을 보여주었다(실시예 6 참조).
- [0234] 두 번째, 본 발명자들은 PPTase가 이종의 PUFA PKS ACP 도메인을 활성화시킬 수 있는 것을 이전에 발견했다. 비브리오 마리누스의 PUFA PKS 유전자로 형질전환된 대장균에서의 DHA의 제조는 적절한 (이 경우에는 웨와넬라 SCRC2738 유래의) PPTase 유전자가 존재할 때에만 일어난다(미국특허 제6,140,486호). 이는 웨와넬라 PPTase가 비브리오 PUFA PKS ACP 도메인을 활성화시킬 수 있다는 것을 증명하였다. 또한, 본 발명자들은 바실루스 썩틸리스의 PPTase(sfp)를 사용한 스킨조카이트리움 OrfA 유래 ACP 도메인의 활성화(판테테닐화)를 증명하였다(실시예 2 참조). 본 발명자들은 또한 노스톡에서 유래된 Het I이라 불리는 PPTase에 의한 스킨조카이트리움 OrfA 유래 ACP 도메인의 활성화(판테테닐화)를 증명하였다(실시예 2참조). HetI효소는 재조합 스킨조카이트리움 PUFA PKS 유전자를 사용하여 대장균에서 DHA와 DPA를 제조하기 위하여 앞서 언급된 실험에서 PPTase로서 사용되었다(실시예 2).
- [0235] 세 번째, 데이터는 DHA-CoA와 DPA-CoA가 스킨조카이트리움 TAG 및 PL 합성 경로에서 대사 중간체임을 보여준다. 공개된 생화학 데이터는 박테리아의 경우에 새로 합성된 PUFA는 PUFA PKS ACP 도메인으로부터 직접 포스포리피드 합성 효소로 전달된다는 것을 지지한다. 대조적으로, 본 발명자들의 데이터는 스킨조카이트리움, 즉 진핵 유기체의 경우에 PUFA PKS ACP 도메인의 PUFA와 표적 TAG와 PL 분자 사이에 중간체가 있을 수 있다는 것을 암시한다. 진핵의 사이토플라즘에서 지방산의 전형적인 캐리어는 CoA이다. 본 발명자들은 스킨조카이트리움 세포의 추출물을 검사하여 DHA-CoA, DPA-CoA, 16:0-CoA와 18:1-CoA의 인증된 기준으로 HPLC 분석 중에 함께 이동하는 수준을 발견했다. DHA-CoA와 DPA-CoA로 추정되는 피크의 동정이 질량 분석기를 사용하여 확인되었다. 이와는 대조적으로 발명자들은 비브리오 마리누스의 추출물에서 DHA-CoA를 검출할 수 없었는데, 이는 박테리아에 최종 목적지로서의 PUFA 이동을 위한 다른 메카니즘이 존재한다는 것을 보여준다(예로 들면 PL로의 직접 전달). 상기 데이터는 새로 합성된 PUFA가 (아마도 ACP로부터 CoA로의 직접 전달을 통하여) CoA로 전달되는 메카니즘이 스킨조카이트리움에 존재할 것이라는 것을 암시한다. TAG와 PL 합성 효소 모두는 그런 다음 PUFA-CoA에 접근할 수 있다. DHA와 DPA CoA 모두 제조된다는 관측은 효소의 전달 기구가 PUFA의 범위를 인식함을 지지한다.
- [0236] 네 번째, 본 발명자들은 스킨조카이트리움에서 OrfA, OrfB 및 OrfC의 녹아웃을 생성하였다(실시예 3 참조). 녹아웃 전략은 스킨조카이트리움에 일어나는 것으로 증명된 상동성 재조합에 기초한다(미국특허출원 제 10/124,807호). 몇 가지 전략이 녹아웃 작제물의 설계에 도입될 수 있다. 상기 세 가지 유전자를 비활성화하는데 이용되는 특별한 전략은 튜블린 프로모터에(pMON50000에서 유래된, 미국특허출원 제10/124,807호 참조)와

작을 이룬 Zeocin™ 저항 유전자를 Orf의 클로닝 부분에 삽입하는 것을 활용하는 것이다. 파괴된 코딩 영역을 포함한 새로운 작제물이 입자 사격(particle bombardment)을 통한 야생형 스킨조카이트리움 세포의 형질전환에 사용하였다(미국특허출원 제10/124,807호). 이렇게 사격된 세포는 Zeocin™ 과 PUFA 공급을 포함한 플레이트에 도말되었고(하기 참조). 상기 플레이트에서 성장한 콜로니들은 PUFA로 보충되지 않은 Zeocin™ 플레이트 상으로 스트리킹(streaking)되었다. 성장을 위해 PUFA 보충을 필요로 하는 상기 콜로니들은 상동성 재조합을 통하여 PUFA PKS Orf가 비활성화된 후보자이다. 상기된 세 가지 경우에서, 이러한 전제는 각각의 스킨조카이트리움 Orf의 전장 게놈의 DNA 클론으로 세포를 형질전환함으로써 녹아웃을 복구함으로써 확인되었다. 추가로, 일부 경우에는 Zeocin™ 저항 유전자가 제거된다는 것이 발견되었는데(실시에 5 참조), 이는 도입된 기능성 유전자가 이중의 상동성 재조합에 의해 원래의 위치로 통합된다는 것(즉, 저항 마커를 결실함)을 의미한다. 이러한 전략의 성공을 위한 핵심은 PUFA로 성장 배지를 보충하는 것이다. 이러한 경우에, 효율적인 보충 수단은 PUFA를 성장 배지에 첨가하기 전에 부분적으로 메틸화된 베타-사이클로텍스트린을 혼합함으로써 격리되게 하는 것임을 발견하였다(실시에 5 참조). 동시에, 상기 실험들은 본 발명에서 제공된 가이드라인에 의해 당업자라면 스킨조카이트리움과 같은 PUFA PKS를 포함하는 미생물의 하나 이상의 PUFA PKS 유전자를 불활성화시키고, 본 발명에 따라 추가 유전적 변형을 위해 이용될 수 있는 PUFA 영양요구체를 생성할 수 있는 원리를 증명한다.

[0237]

본 발명의 유기체의 유전적 변형에 있어서 중요한 요소는 트라우스토카이트리드 게놈을 직접적으로 형질전환시키는 능력이다. 미국특허출원 제 10/124,807호에 단일 크로소버 상동성 재조합을 통한 스킨조카이트리움의 형질전환과 이중 크로소버 상동성 재조합을 통한 표적화된 유전자 치환이 증명되었다. 상기된 바, 본 발명자들은 스킨조카이트리움에 있는 PUFA PKA 시스템의 OrfA, OrfB 및 OrfC를 비활성화시킨 상동성 재조합에서 상기 기술을 이용하였다. 결과적으로 얻는 돌연변이체는 PUFA를 가진 배지의 보충물에 의존적이 되었다. 몇 개의 형질전환 마커, 도입된 유전자의 높은 발현율을 위한 프로모터 요소 및 외인성 유전자 물질의 전달을 위한 방법이 개발되었고 이용 가능하다. 따라서 상기 기술은 트라우스토카이트리드 및 유사한 PUFA PKS 시스템을 가진 다른 진핵에서 내인성 PUFA PKS 유전자를 녹아웃하고 이를 다른 유기체에서 유래된 유전자로 (또는 변형된 스킨조카이트리움 유전자로) 대체하는 위치에 이르렀다.

[0238]

EPA로 풍부한 TAG의 제조를 위한 한 방식에서, 스킨조카이트리움의 PUFA PKS 시스템은 EPA 생산물을 제조하는 PUFA PKS를 코딩하는 이중의 유전자를 추가함으로써 치환될 수 있다. 내인성의 PPTase가 이중의 PUFA PKS 시스템의 ACP 도메인을 활성화시킬 것으로 예상된다. 추가로, EPA가 EPA-CoA로 전환되어 스킨조카이트리움 TAG와 PL 맴브레인으로 통합될 것으로 예상된다. 이러한 방식의 한 변형에서 기술은 최종 생산물이 DHA와 DPA 보다는 EPA이도록 (이중 유전자의 특정 영역의 도입에 의해 또는 스킨조카이트리움 유전자 자체의 돌연변이에 의해) 내인성 스킨조카이트리움 시스템의 관련 도메인을 변형하는데 사용될 수 있다. 이는 예시적인 방식으로, 상기 기술은 다른 PUFA의 최종 생산물의 제조 및 PUFA PKS 시스템을 포함하고 PUFA PKS 시스템의 생산물을 포스포리피드(PL)와 트리아실글리세롤(TAG)의 두 가지 모두로 효율적으로 보내는 능력이 있는 진핵의 미생물에 적용될 수 있다. 특히, 본 발명은 하기에서 예시로서 설명되는 트라우스토카이트리드 미생물 또는 내인성 PUFA PKS 시스템을 가진 다른 진핵 생물에 적용될 수 있다. 또한, 본 발명은 본 발명에서 기술된 다양한 PUFA 프로필의 제조를 위하여 변형된 유전 물질로 형질전환될 수 있는 임의의 적합한 숙주 유기체에 적용될 수 있다. 예를 들어, 실시예에서는 스킨조카이트리움의 PUFA PKS 시스템이 대장균으로 형질전환되었다. 이렇게 형질전환된 유기체는 앞서 기술된 방법을 사용하여 PUFA 제조 프로필을 변이시키기 위하여 추가로 변형될 수 있다.

[0239]

본 발명은 유전자와 본 발명과 미국특허출원 제10/124,800호에 기술된 PUFA PKS 시스템 이외의 PKS 시스템 유래의 단백질 또는 도메인을 코딩하는 유전자 및 핵산서열을 이용할 수 있게 하며, 앞서 기술된 타입 II, 타입 I 및 모듈라 PKS 시스템을 포함한 박테리아 및 비박테리아의 PKS 시스템 유래의 유전자 및 핵산 서열을 포함한다. PKS 시스템의 이러한 타입의 각각을 발현하는 유기체는 당업계에 잘 알려져 있으며, 본 발명의 유전적 변형 과정에서 유용한 핵산의 원천으로 사용될 수 있다.

[0240]

바람직한 양태에서, PUFA PKS 시스템 이외의 PKS 시스템 또는 다른 PUFA PKS 시스템으로부터 유래된 단백질 또는 도메인을 코딩하는 유전자와 핵산 서열은 PUFA의 제조에 바람직한 성장 특징을 가진 유기체로부터 분리되거나 추출한 것이다. 특히, 유전적으로 변형된 트라우스토카이트리드 미생물을 약 15℃ 이상, 20℃ 이상, 25℃ 이상, 30℃ 이상, 35℃ 이상, 40℃ 이상에서 또는 한 양태에서는 약 20℃ 내지 40℃ 사이의 임의의 온도에서 배양할 수 있는 것이 바람직하다. 따라서, 상기 온도에서 기능성 효소 활성을 가진 PKS 단백질 또는 도메인이 바람직하다. 예로 들어, 본 발명자들은 각각 30℃ 및 35℃까지의 온도에서 천연적으로 EPA를 제조하고 성장하는 해양 박테리아인 *쉐와넬라 올레야나* 또는 *쉐와넬라 자포니카*에서 유래된 PKS 유전자의 사용을 설명한다(실시에 7). 상기 유기체로부터 유래된 PKS 단백질 또는 도메인은 특별히 고안 또는 변형된 PUFA 제조 프로필을

가진 유전적으로 변형된 유기체를 제조하기 위하여 본 발명에서 기술된 트라우스토카이트리드 PUFA PKS 단백질 및 도메인과 혼합될 수 있는 단백질 및 도메인의 예시이다.

[0241]

다른 바람직한 양태에서, 하나의 지방산 프로필을 제조하는 PUFA PKS 시스템에서 유래된 단백질 또는 도메인을 코딩하는 유전자와 핵산 서열은 다른 PUFA PKS 시스템을 변형시키는데 이용되어, 이에 의해 숙주의 지방산 프로필을 변이시킬 수 있다. 예를 들어, 트라우스토카이트리움 23B(ATCC 20892)는 스킨조카이트리움 sp. (ATCC 20888)과 지방산 프로필에 있어서 현격히 다르다. 트라우스토카이트리움 23B는 DHA:DPA($n-6$) 비율이 스킨조카이트리움(ATCC 20888)의 2-3:1과 비교하였을 때, 40:1로 상당히 높다. 트라우스토카이트리움 23B는 C20:5($n-3$)로 역시 높은 수준을 보인다. 그러나 스킨조카이트리움 (ATCC 20888)은 트라우스토카이트리움 23B에 비교하였을 때 탁월한 오일 생산자이다. 스킨조카이트리움은 도코사헵타엔산과 DHA(DPA; 22:5 ω 6)가 풍부한, 예를 들어 건조 중량으로 30% DHA + DPA인, 다량의 트리아실글리세롤을 축적한다. 따라서 본 발명자들은 트라우스토카이트리움 23B와 보다 유사한 DHA:DPA 프로파일을 가진 유전적으로 변형된 스킨조카이트리움을 생성하기 위해서 트라우스토카이트리움 23B PUFA PKS 유전자로 스킨조카이트리움 내인성 PUFA PKS 시스템을 변형시키는 것을 설명한다(즉, "초-DHA-생산자" 스킨조카이트리움으로, 여기서 스킨조카이트리움의 제조 능력이 트라우스토카이트리움의 DHA:DPA 비율과 조합된다).

[0242]

따라서 본 발명은 트라우스토카이트리드 PUFA PKS 시스템에서 유래된 유전자를 이용하게 하고, 추가로 EPA, DHA, DPA, ARA, GLA, SDA 및 기타를 포함한 PUFA 생산물의 범위를 연장 및/또는 변이시키기 위하여 유전자 혼합을 활용한다. 이와 같이 변이된 PUFA 제조 프로필을 수득하기 위한 방법은 다양한 유기체 유래의 유전자를 트라우스토카이트리드 PUFA PKS 유전자로 혼합하는 것뿐만 아니라, 본 발명에서 기술된 내인성 트라우스토카이트리드 PUFA PKS 유전자를 유전적으로 변형시키는 다양한 방법을 포함한다. 본 발명의 트라우스토카이트리드 PUFA PKS 시스템의 (예를 들어 상기 스킨조카이트리움에 대하여 상세히 기술된) 유전적 기초와 도메인의 구조에 관한 지식은 다양한 PUFA 프로필을 제조하는 신규한 유전적으로 변형된 유기체를 고안하는데 기초를 제공한다. 트라우스토카이트리드와 같은 미생물에서 제조된 신규한 PUFA PKS 작제물은 분리될 수 있으며, 식물에 유사한 PUFA 제조 성질을 부여하기 위하여 식물을 형질전환시키는데 이용될 수 있다.

[0243]

변형이 목적인 결과(즉, 미생물의 PUFA 제조 프로필의 변이)를 가져온다면, 본 발명에 따라 하나 이상의 내인성 트라우스토카이트리드 PUFA PKS 도메인이 변이되거나 대체될 수 있다. 변이 또는 대체될 특히 바람직한 도메인은 이에 제한되는 것은 아니지만, 스킨조카이트리움 OrfB 또는 OrfC에 있는 도메인(β -케토 아실-ACP 신타제(KS), 아실트랜스퍼라제(AT), FabA-유사 β -하이드록시 아실-ACP 탈수소효소(DH), 쇠 길이 인자(CLF), 엔오일 ACP-환원효소(ER), 트랜스-2-아실-ACP의 합성을 촉진시키는 효소, 트랜스-2-아실-ACP의 시스-3-아실-ACP로의 가역적인 이성질화를 촉매하는 효소, 시스-3-아실-ACP의 시스-5- β -케토-아실-ACP로의 연장을 촉매하는 효소)에 대응되는 임의의 도메인을 포함한다. 한 양태에서, 변이 또는 대체하기에 바람직한 도메인은 이에 제한되는 것은 아니지만 β -케토 아실-ACP 신타제(KS), FabA-유사 β -하이드록시 아실-ACP 탈수소효소(DH) 및 쇠 길이 인자(CLF)를 포함한다.

[0244]

본 발명의 한 측면에서, 트라우스토카이트리드 PUFA-PKS PUFA 제조는 CLF(쇠 길이 인자) 도메인을 변형 시킴으로 변이된다. 상기 도메인은 타입II(분리되는 효소) PKS 시스템의 특징이다. 이의 아미노산 서열은 KS(케토 신타제 쌍) 도메인에 상동성을 보이지만, 활성 부위 시스테인이 결여되어 있다. CLF는 연장 사이클의 회수를, 그리하여 최종 생산물의 쇠 길이를 결정하는 기능을 한다. 본 발명의 한 양태에서 FAS 및 PKS 합성과 관련된 현재의 지식을 이용하여, 비박테리아의 PUFA-PKS 시스템의 지정된 변형에 의한 ARA 제조의 합리적인 전략이 제공된다. PKS 시스템에서 CLF 기능과 관련된 문헌(Bisang et al. *Nature* **401**, 502(1999); Yi et al. *J. Am. Chem. Soc.* **125**, 12708(2003))에는 여전히 논란이 있는데, 이는 다른 도메인이 최종 생산물의 쇠 길이 결정에 관여할 수 있다는 것을 보여주었다. 그러나 스킨조카이트리움이 DHA(C22:6, ω -3)와 DPA(C22:5, ω -6) 두 가지 모두를 제조한다는 것이 중요하다. PUFA-PKS 시스템에서 시스 이중결합은 성장하는 탄소쇄의 합성 중에 도입된다. ω -3과 ω -6의 이중결합의 배치는 분자의 합성 초기에 일어나므로 최종 생산물의 쇠길이 결정에 영향을 미칠 것이라고 예상되지 않는다. 따라서 이론에 얽매이지 않고, 본 발명자들은 스킨조카이트리움 PUFA-PKS로(C22 단위 대신에) C20 단위의 합성을 지시하는 인자(예, CLF)를 도입하는 것은 EPA(C20:5, ω -3)와 ARA(C20:4, ω -6)의 제조로 이어질 것으로 예상한다. 예를 들어, 이중의 시스템에서, (포도박테리움 또는 바람직하게는 이하에서 기술되는 바람직한 성장 필요조건을 가진 미생물에서 유래된 것과 같은) EPA 제조 시스템 유래의 CLF를 스킨조카이트리움 유전자 세트와 직접적으로 치환함으로써 CLF를 개발할 수 있다. 이렇게 얻은 형질전환체의 지방산은 EPA 및/또는 ARA를 제조하는 형질전환체를 동정하기 위하여 프로파일에서의 변경에 대하여 분석될 수 있다.

[0245]

예시로 발명의 한 측면은 C20 PUFA-PKS 시스템 유래의 CLF로 대체된 OrfB의 CLF를 가진 클론을 작제할 수 있다. 마커 유전자가 코딩 영역의 다운 스트림에 삽입될 수 있다. 보다 자세하게는 본 발명에서 기술된 그리고 미국특허출원 제10/124,807호에 상세히 개시된 바에 따라 트라우스토카이트리드의 형질전환을 위하여 상동성 재조합 시스템을 사용할 수 있다. 야생형 트라우스토카이트리드 세포(예, 스킨조카이트리움 세포)를 형질전환하고, 마커 표현형에 대하여 선택하고, 새로운 CLF를 병합시킨 이들을 스크리닝할 수 있다. EPA 및/또는 ARA를 제조하는 형질전환체를 동정하기 위하여 지방산 프로파일에 대한 영향과 관련해서 상기 형질전환체를 분석할 것이다. CLF와 관련된 것 이외의 인자가 최종 생산물의 쇠 길이에 영향을 미친다는 것이 발견되면, 유사한 전략이 상기 인자를 변이시키기 위하여 도입될 수 있다.

[0246]

본 발명의 다른 측면에서, β -하이드록시 아실-ACP 탈수소효소/케토 신타제 쌍의 변형 또는 치환이 고려된다. 대장균에서 시스-바센산(C18:1, Δ 11) 합성 중에 시스 이중결합의 생성은 특정한 DH 효소, 즉 *fabA* 유전자의 생산물인 β -하이드록시 아실-ACP 환원효소에 따라 결정되는 것으로 여겨진다. 상기 효소는 β -케토 아실-ACP로부터 HOH를 제거하고 탄소쇄에 트랜스 이중결합을 남긴다. FabA-유사 DH의 소집합은 시스-트랜스 이성화 활성을 가진다(Heath et al. 1996). 박테리아 및 비박테리아의 PUFA-PKS 시스템의 새로운 측면은 두 가지 FabA-유사 DH 도메인의 존재이다. 이론에 얽매이지 않고, 본 발명자들은 하나 또는 둘 모두의 DH 도메인이 시스-트랜스 이성질화 활성을 보유할 것이라고 여긴다(DH 도메인의 조작은 이하에서 상세히 설명될 것이다).

[0247]

대장균에서 불포화 지방산 합성의 다른 측면은 특정한 KS 효소, 즉 *fabB* 유전자의 생산물인 β -케토아실-ACP 신타제에 대한 필요이다. 이는 말로닐-ACP와, (티오-에스테르에 의해) 활성 부위에 있는 시스테인 잔기에 연결된 지방산의 축합을 수행하는 효소이다. 단단계 반응으로 CO₂가 방출되면서 선형의 쇠는 2개의 탄소로 확장된다. 상기 KS만이 이중 결합을 포함한 탄소 쇠를 연장시킬 수 있는 것으로 여겨진다. 이러한 확장은 이중 결합이 시스 배형일 때에만 일어난다; 트랜스 배형이라면 연장되기 전에 이중결합은 엔오일-ACP 환원효소(ER)에 의해 환원된다(Heath et al. 1996). 지금까지 확인된 모든 PUFA-PKS 시스템은 두 가지 KS 도메인을 가지고 있으며, 이중의 하나가 대장균의 FabB-유사 KS에 더 많은 상동을 보인다. 또한, 이론과 상관없이, 본 발명자들은 PUFA-PKS 시스템에서 DH(FabA-유사)와 KS(FabB-유사)의 효소의 도메인의 상호작용과 특이성이 최종 생산물에서 시스 이중결합의 수와 위치를 결정한다고 여긴다. 2-탄소 연장 반응의 회수가 PUFA-PKS 최종 생산물에 존재하는 이중결합의 수보다 크기 때문에, 어떤 연장 사이클에서 완전한 환원이 발생하는지 결정할 수 있다. 따라서, DH 와 KS도메인은 DHA/DPA 비율 또는 다른 장쇄 지방산의 비율의 변이를 위한 표적으로서 이용될 수 있다. 이는 다른 시스템에서 유래된 상동성 도메인의 도입 또는 이러한 유전자의 단편의 돌연변이 의해 변형되고/거나 평가될 수 있다.

[0248]

다른 양태에서, ER(엔오일-ACP 환원효소-지방 아실-ACP에서의 트랜스-이중 결합을 환원시켜 완전히 포화된 탄소를 만드는 효소) 도메인은 PKS 시스템에 의해 만들어진 생산물의 타입을 변화시키기 위하여 변형되거나 치환될 수 있다. 예를 들어 본 발명자들은 스킨조카이트리움 PUFA-PKS 시스템이 (하나가 아닌) 두 개의 ER 도메인을 가지고 있다는 점에서 앞서 개시된 박테리아의 시스템과는 상이하다는 것을 인지한다. 이론과 관계없이 본 발명자들은 이러한 ER 도메인이 결과적인 PKS 제조 생산물에 강한 영향을 미친다고 여긴다. 결과적인 PKS 생산물은 각각의 도메인을 개별적으로 녹아웃시키거나 도메인의 뉴클레오타이드 서열을 변형시키거나 다른 유기체로부터 유래된 ER 도메인을 치환시킴으로써 변화될 수 있다.

[0249]

본 발명의 다른 측면에서 PUFA-PKS 시스템의 DH 도메인(FabA 유사) 중 하나를 이성질화 활성을 보유하지 않은 DH 도메인으로 치환하는 것을 고려하는데, 이는 시스-와 트랜스- 이중결합의 혼합을 가진 분자를 잠정적으로 생성한다. 스킨조카이트리움 PUFA PKS 시스템의 현재 생산물은 DHA와 DPA(C22:5, ω 6)이다. C20 지방산을 제조하기 위해 시스템을 조작한다면, EPA와 ARA(C20:4, ω 6)를 생산물로 예상할 수 있다. 이는 ARA의 새로운 원천을 제공한다. 다른 DHA 대 DPA 비율을 제조하는, 예를 들면 트라우스토카이트리움 23B(미국특허출원 제10/124,800호에서 동정된 PUFA PKS 시스템)로부터의 유전자를 사용하여 관련된 PUFA-PKS 시스템으로부터의 도메인을 대체할 수 있다.

[0250]

추가로, 한 양태에서 ER 도메인 중 하나가 최종 생산물의 프로파일을 변이시키기 위해 (예를 들어, 제거 또는 비활성화함으로써) 트라우스토카이트리드 PUFA PKS 시스템에서 변이된다. 유사한 전략이 다소 정교한 방식을 이용하여 PUFA-PKS 단백질의 개별 도메인의 각각에 대하여 직접적인 방법으로 시도될 수 있다. 물론 단일 도메인에 대한 조작으로 한정되는 것은 아니다. 또한, 새로운 PUFA 최종 생산물의 전체 범위를 생성하기 위하여 PUFA-PKS 시스템과 다른 PKS 또는 FAS 시스템(예를 들면 타입 I, 타입 II, 모듈라) 유래의 도메인을 혼합함으로써 방식을 확장시킬 수 있다.

[0251]

본래의(내인성, 천연의) PKS 시스템으로 도입되는 무작위적 또는 지정된 많은 유전적 변형은 효소 기능의 비활성화를 초래할 수도 있다. 따라서 제어된 환경에서 트라우스토카이트리드 PUFA PKS 시스템의 유전적 조작의 영향에 대하여 테스트하기 위하여 대장균과 같은 다른 숙주의 재조합 시스템을 먼저 사용하여 시스템의 다양한 측면을 조작하고 그 결과를 평가할 수 있다. 예를 들어, 대장균의 FabB-균주는 불포화 지방산을 합성할 수 없으며, 성장하기 위해서는 정상적인 불포화 지방산을 치환할 수 있는 지방산으로 배지를 보충할 필요가 있다(Metz et al. 2001). 그러나, (배지의 보충이라는) 상기 필요는 균주가 기능성 PUFA-PKS 시스템(즉, 대장균 숙주에서 PUFA 생산물을 제조하는 시스템 - Metz et al. 2001, 도 2A)으로 형질전환되었을 때 제거될 수 있다. 형질전환된 FabB-균주는 보충 없이 성장을 위한(불포화 지방산을 제조하기 위한) 기능성 PUFA-PKS 시스템을 필요로 한다. 이러한 예시에서 중요한 요소는 (분지된 쇠 지방산과 같은 다중불포화지방산 치환물까지) 다양한 다중불포화 지방산의 제조가 충족될 것이라는 것이다. 따라서 본 발명의 다른 바람직한 양태는 본 발명에서 언급된 하나 이상의 PUFA PKS 유전자에서 많은 변형을 생성시키고, 그 다음 적합하게 변형된 FabB-균주를 형질전환시켜(예를 들어, ER 도메인을 포함하는 발현 작제물에서 변형을 생성시키고 개별적인 플라스미드에 필수적인 도메인을 가지고 있거나 염색체상에 통합된 FabB-균주를 형질전환시켜서) 배지의 보충 없이도 성장하는 (즉, FabB-결점을 보완하는 분자를 제조하는 능력을 보유한) 형질전환체만을 선택할 수 있다.

[0252]

PUFA PKS의 유전적 변형을 위한 테스트 시스템이 실시예 부분에서 예시된다. 간략히 설명하면, 대장균과 같은 숙주 미생물이 트라우스토카이트리드 PUFA PKS 시스템(예를 들면, 스키토카이트리움의 OrfsA, B 및 C)의 모두 혹은 일부분을 포함하는 PUFA PKS 시스템을 코딩하는 유전자와, PKS 시스템에서 활성의 홀로-ACP를 제조하기 해서는 포스포판테테인 보조인자의 부착이 요구되는 포스포판테테인 트랜스퍼라제(PPTase)를 코딩하는 유전자로 형질전환된다. PKS 시스템을 코딩하는 유전자는 트라우스토카이트리드 PUFA PKS 유전자에 적어도 하나 이상의 변형을 도입 및/또는 다른 PKS 시스템에서 유래된 도메인을 코딩하는 핵산 분자를 트라우스토카이트리드 유전자(비-트라우스토카이트리드 미생물로부터의 유전자와 다른 트라우스토카이트리드 미생물로부터의 유전자를 포함한)로 도입하기 위해 유전적으로 조작될 수 있다. PUFA PKS 시스템은 대장균에서 발현될 수 있고 PUFA 제조 프로필이 측정될 수 있다. 이러한 방식으로 강력한 유전적 변형이 트라우스토카이트리드 PUFA 제조 유기체의 조작 이전에 평가될 수 있다.

[0253]

본 발명은 내인성 핵산 분자의 조작 및/또는 트라우스토카이트리드 PUFA PKS 시스템 유래의 핵산 서열 또는 이의 상동체를 포함한 분리된 핵산 분자의 이용을 포함한다. 한 측면에서 말로닐-CoA:ACP 아실트랜스퍼라제(MAT), β -케토 아실-ACP 신타제(KS), 케토환원효소(KR), 아실트랜스퍼라제(AT), FabA-유사 β -하이드록시-ACP 탈수소효소(DH), 포스포판테테인 트랜스퍼라제, 쇠길인자(CLF), 아실 캐리어 단백질(ACP), 엔오일 ACP-환원효소(ER), 트랜스-2-아실-ACP의 합성을 촉매하는 효소, 트랜스-2-아실-AC의 시스-3-아실-ACP로 가역적인 이성질화를 촉매하는 효소 및/또는 시스-3-아실-ACP의 시스-5- β -케토-아실-ACP로의 연장을 촉매하는 효소 중 적어도 하나의 생물학적 활성을 가지는 PUFA PKS 시스템 유래의 도메인을 코딩하는 핵산 서열을 포함한 핵산 분자의 조작 및/또는 이용에 관한 것이다. 숙주 트라우스토카이트리드의 PUFA 제조 프로필을 변이시키기 위하여 변형될 바람직한 도메인은 앞서 설명되었다.

[0254]

본 발명에 따른 트라우스토카이트리드 미생물의 유전자적인 변형은 바람직하게는 내인성 PUFA PKS 시스템이 유전적으로 변형되었는지 및/또는 재조합 핵산 분자가 유기체에 도입되었는지 여부에 상관없이 미생물에 의해 제조되는 PUFA의 타입, 양, 및/또는 활성에 영향을 미친다. 본 발명에서 PUFA 제조 프로필에 영향을 미치는 것과 같이, PUFA PKS 시스템의 활성에 영향을 미친다는 것은 유전적 변형이 없는 경우와 비교하였을 때 미생물에 의해 발현되는 PUFA PKS 시스템의 생물학적 활성에서의 임의의 검출 또는 측정할 수 있는 변화 또는 변형을 일으키는 PUFA PKS 시스템과 관련된 PUFA PKS 시스템 또는 유전자에서의 임의의 변형을 포함한다. 본 발명에서 용어 "PUFA 프로필", "PUFA 발현 프로필" 및 "PUFA 제조 프로필"은 상호교환적으로 사용될 수 있으며, 미생물에 의해 발현/제조되는 PUFA의 전반적인 프로필을 설명할 수 있다. PUFA 발현 프로필은 미생물에 의해 발현되는 PUFA의 타입 뿐만 아니라 제조되는 PUFA의 절대적인 및 상대적인 양을 포함한다. 따라서 PUFA 프로파일은 미생물에 의해 제조되는 한 PUFA 대 다른 것의 비율의 관점, 미생물에 의해 제조된 PUFA의 타입의 관점, 및/또는 미생물에 의해 제조되는 PUFA의 절대적 또는 상대적인 양의 관점에서 설명될 수 있다.

[0255]

상기에서 보여준 것 같이, 숙주 미생물은 내생적인 PUFA PKS 시스템을 가진 모든 진핵 미생물과 PUFA PKS 시스템의 산물을 포스포리피드(PS)과 트리아실글리세롤(TAG) 2가지 모두에게 효율적으로 흘려보내는 능력을 포함하는 한편, 바람직한 숙주 미생물은 트라우스토카이트리아세 와 라비린톨라시아과를 포함한 트라우스토카이트리아레목의 일원은 어느 것이라도 무방하다. 본 발명에 쓰일 특히 선호되는 숙주세포는, 제한적이지는 않지만, 다음의 속으로부터의 미생물을 포함한다: 트라우스토카이트리아세과내의 트라우스토카이트리움, 자포노

카이트리움, 아플라노카이트리움, 엘리나, 스킨조카이트리움과 라비린톨라시에과내의 라비린톨라, 라비린톨로이데스, 라비린토크사. 위의 속 내에서의 바람직한 종은, 제한적이지는 않지만, 다음과 같다: 라비린톨라 sp., 라비린톨라 알게리엔시스, 라비린톨라 시엔콥스키, 라비린톨라 샤토니, 라비린톨라 코에노시스티스, 라비린톨라 마크로시스티스, 라비린톨라 마크로시스티스 아틀란티카, 라비린톨라 마크로시스티스 마크로시스티스, 라비린톨라 마그니피카, 라비린톨라 미누타, 라비린톨라 로스코젠시스, 라비린톨라 발카노비, 라비린톨라 비텔리나, 라비린톨라 비텔리나 파시피카, 라비린톨라 비텔리나 비텔리나, 라비린톨라 조피를 포함한 라비린톨라의 모든 종; 라비린톨로이데스 sp., 라비린톨로이데스 미누타, 라비린톨로이데스 스킨조카이트로프를 포함한 라비린톨로이데스의 모든 종; 라비린토크사 sp., 라비린토크사 폴리아, 라비린토크사 사우바게우이를 포함한 모든 라비린토크사종; 아플라노카이트리움 sp. 과 아플라노카이트리움 케르구엘렌시스를 포함한 모든 아플라노카이트리움종; 엘리나 sp., 엘리나 마리살바, 엘리나 시노리피카를 포함한 모든 엘리나종; 야파노카이트리움 sp., 야파노카이트리움 마리눔을 포함한 모든 야파노카이트리움종; 스킨조카이트리움 sp., 스킨조카이트리움 아그레가툼, 스킨조카이트리움 리마시눔, 스킨조카이트리움 미누툼, 스킨조카이트리움 옥토포룸을 포함한 모든 스킨조카이트리움종; 트라우스토카이트리움 sp., 트라우스토카이트리움 아그레가툼, 트라우스토카이트리움 아루디멘탈레, 트라우스토카이트리움 오레움, 트라우스토카이트리움 벤티콜라, 트라우스토카이트리움 글로버섬, 트라우스토카이트리움 킨네이, 트라우스토카이트리움 모티툼, 트라우스토카이트리움 파카이더툼, 트라우스토카이트리움 프로리페툼, 트라우스토카이트리움 로세움, 트라우스토카이트리움 스트리아툼, 울케니아 sp., 울케니아 미누타, 울케니아 프로판다, 울케니아 라디에이트, 울케니아 사카리아나, 울케니아 비수르젠시스를 포함한 모든 트라우스토카이트리움종. 특히 이들 속에서도 바람직한 종은, 이에 제한되는 것은 아니지만: 스킨조카이트리움 아그레가툼, 스킨조카이트리움 리마시눔, 스킨조카이트리움 미누툼을 포함한 모든 스킨조카이트리움 종; 트라우스토카이트리움 스트리아툼, 트라우스토카이트리움 오레움, 트라우스토카이트리움 로세움을 포함한 모든 트라우스토카이트리움 종(U. 비수르젠시스, U. 아메보이다, U. 사카리아나, U. 프로판다, U. 라디아타, U. 미누타 와 울케니아 sp. BP-5601과 같은 이전의 울케니아 종을 포함한); 모든 자포노카이트리움 종. 특히 트라우스토카이트리알레의 바람직한 균주는 이에 제한되는 것은 아니지만 다음과 같다: 스킨조카이트리움 sp. (S31)(ATCC 20888); 스킨조카이트리움 sp. (S8)(ATCC 20889); 스킨조카이트리움 sp. (LC-RM)(ATCC 18915); 스킨조카이트리움 sp. (SR21); 스킨조카이트리움 아그레가툼 (Goldstein et Belsky)(ATCC 28209); 스킨조카이트리움 리마시눔(Honda et Yokochi)(IFO 32693); 트라우스토카이트리움 sp. (23B)(ATCC 28210); 트라우스토카이트리움 striatum (Schneider)(ATCC 24473); 트라우스토카이트리움 오레움 (Goldstein)(ATCC 34304); 트라우스토카이트리움 roseum (Goldstein)(ATCC 28210); 그리고 자포노카이트리움 sp. (L1)(ATCC 28207).

[0256] 본 발명의 한 양태에서는 돌연변이 프로그램이 대상 PUFA 제조 프로필을 가진 트라우스토카이트리드 미생물을 획득하기 위한 선택적인 스크리닝 과정과 조합될 수 있다는 것을 고려할 수 있다. 돌연변이 방법은 이에 제한되는 것은 아니지만, 화합물에 의한 돌연변이, 유전자 섞기(gene shuffling), 특정 효소의 도메인을 코딩하는 유전자의 영역을 스위치 또는 이러한 유전자의 특정 영역에 제한된 돌연변이 뿐만 아니라 다른 방법을 포함할 수 있다.

[0257] 예를 들어, 고효율 돌연변이 방법(high throughput mutagenesis)이 목적인 PUFA 프로필의 제조를 최적화하거나 영향을 미치는데 이용될 수 있다. 일단 효율적인 모델 시스템이 개발되면, 고효율 방식으로 유전자들을 변형시킬 수 있다. 이러한 기술의 활용화는 두 단계로 나누어 볼 수 있다. 첫째로, 만일 대상 생산물(예, EPA)의 제조를 위하여 충분히 선택적인 스크리닝이 고안된다면, 이를 이러한 생산물을 제조하는 시스템을 변이시킬 시도에 이용될 수 있다(예를 들어 상기에서 설명한 다른 전략들 대신에 또는 협의하여). 추가로, 앞서 요약된 전략이 대상 PUFA 프로필을 제조하는 유전자 세트를 나오게 한다면, 고효율 기술이 상기 시스템을 최적화하는데 이용될 수 있다. 예를 들어, 도입된 도메인이 비교적 낮은 온도에서만 기능하다면, 선택 방법은 그러한 제한의 제거를 허용하도록 고안될 수 있다.

[0258] 본 발명의 한 양태에서, 유전적으로 변형된 트라우스토카이트리드 미생물은 목적인 PUFA를 합성하는데 개선했던 능력을 가지고/거나 상이한 프로필의 PUFA를 합성할 수 있는 새롭게 도입된 능력을 가진다. 본 발명에서 생산물을 "합성하는데 개선된 능력"이란 상기 미생물이 동일한 조건하에서 배양 또는 성장된 야생형 미생물과 비교하였을 때 증가된 양의 (전혀 제조하지 않은 생산물의 제조도 포함해서) 생산물을 제조할 수 있도록 상기 생산물의 합성과 관련된 경로에서의 증진 또는 상향 조절을 의미한다. 그러한 유전적으로 변형된 유기체를 제조하기 위한 방법은 위에서 상세히 다루어졌다.

[0259] 본 발명의 한 양태에서 유전적으로 변형된 미생물 또는 식물은 목적인 생활성 분자를(생산물) 합성하는데 개선된 능력을 가지고/거나 특정한 생산물을 합성할 수 있는(예를 들면, 특정한 항생제를 합성하는) 새롭게 도

입된 능력을 가진 미생물 또는 식물을 포함한다. 본 발명에서 생산물을 "합성하는데 개선된 능력"이란 미생물 또는 식물이 동일한 조건하에서 배양 또는 성장된 야생형 미생물 또는 식물과 비교하였을 때 증가된 양의 (전혀 제조하지 않은 생산물의 제조도 포함해서) 생산물을 제조할 수 있도록 상기 생산물의 합성과 관련된 경로에서의 증진 또는 상향 조절을 의미한다. 그러한 유전적으로 변형된 유기체를 제조하기 위한 방법은 위에서 상세히 다루어졌다.

[0260] 본 발명의 한 양태는 (앞서 설명된) 본 발명에 따른 유전적으로 변형된 미생물 또는 식물을 성장시키거나 배양함으로써 원하는 목적인 생활성 분자(또는 생산물 또는 화합물이라고 함)를 제조하는 방법이다. 이러한 방법은 앞서 설명된 바와 같이 그리고 본 발명에 따라 유전적 변형을 가진 미생물 또는 식물을 발효 배지에서 배양하거나 흙과 같은 적합한 환경에서 성장시키는 단계를 포함한다. 본 발명의 PUFA PKS 시스템과 관련된 유전적 변형을 위해 바람직한 숙주 세포는 앞서 설명되어졌다.

[0261] 본 발명의 한 양태는 (상기에서 상세히 설명된) 본 발명의 유전적으로 변형된 트라우스토카이트리드 미생물을 배양함으로써 목적인 PUFA를 제조하는 방법이다. 이러한 방법은 상기에서 설명된 바에 따라 그리고 본 발명에 따라 유전적 변형을 가진 트라우스토카이트리드 미생물을 PUFA를 제조하는데 효과적인 조건 및 발효 배지에서 배양하는 단계를 포함한다. 상기 적절한 또는 효과적인 배지는 트라우스토카이트리드와 다른 미생물을 포함하여 본 발명에 따른 유전적으로 변형된 미생물이 배양되었을 때 목적인 PUFA 생산물을 제조하는 배지를 의미한다. 이러한 배지는 일반적으로 동화할 수 있는 탄소, 질소 및 포스페이트 원을 포함한 수용성의 배지이다. 상기 배지는 적합한 염, 미네랄, 금속 및 다른 영양분을 또한 포함할 수 있다. 본 발명의 모든 미생물은 전통적인 발효 생물반응기에서 배양될 수 있다. 미생물은 이에 제한되는 것은 아니지만 회분식, 유가식, 세포 재순환 및 연속식 발효를 포함한 임의의 발효 공정에 의해 배양될 수 있다. 본 발명에 따른 트라우스토카이트리드 미생물의 바람직한 성장 조건은 당업계에 알려져 있으며, 예를 들면, 미국특허 제5,130,242호, 미국특허 제5,340,742호 및 미국특허 제5,698,244호에 상세히 기술되어 있으며, 이들 각각의 전문은 본 발명에서 참조로서 인용된다.

[0262] 한 양태에서 유전적으로 변형된 미생물은 약 15℃ 이상, 다른 양태에서는 약 20℃ 이상, 다른 양태에서는 약 25℃ 이상, 다른 양태에서는 약 30℃ 이상, 다른 양태에서는 약 35℃ 이상, 다른 양태에서는 약 40℃ 이상, 그리고 한 양태에서는 20℃에서 40℃사이의 임의의 온도에서 배양된다.

[0263] 유전적으로 변형된 미생물에 의해 제조되는, 목적인 PUFA 및/또는 다른 생활성 분자는 전통적인 분리와 정제의 기술을 이용하여 발효 배지로부터 회수될 수 있다. 예를 들어 발효 배지는 미생물, 세포 파편 및 다른 미립자를 제거하기 위해 여과되거나 원심분리될 수 있으며, 생산물은 예를 들면, 이온 교환, 크로마토그래피, 추출, 용매추출, 상 분리, 멤브레인 분리, 전기투석, 역삼투, 증류, 화학적 유도(chemical derivatization) 및 결정화와 같은 전통적인 방법에 의해 세포가 제거된 상층액으로부터 회수될 수 있다. 선택적으로 PUFA를 제조하는 미생물 또는 추출물 및 이들의 다양한 분획물이 생산물로부터 미생물의 제거 없이 이용될 수 있다.

[0264] 바람직하게, 본 발명의 유전적으로 변형된 트라우스토카이트리드 미생물은 이에 제한되는 것은 아니지만: EPA(C20:5, ω -3), DHA(C22:6, ω -3), DPA(C22:5, ω -6), ARA(C20:4, ω -6), GLA(C18:3, n-6) 및 SDA(C18:4, n-3)를 포함한 적어도 하나 이상의 다중불포화 지방산을 제조한다. 하나의 바람직한 양태에서 높은 수준의 DHA와 DPA를 제조하는 야생형의 스킴조카이트리움이 본 발명에 따라 유전적으로 변형되어 높은 수준의 EPA를 제조한다. PUFA를 제조하기 위해 유전적으로 변형된 트라우스토카이트리드 미생물을 사용하는 경우의 장점은 PUFA가 포스포글리세르(PL)와 트리아실글리세리드(TAG) 모두에 직접적으로 통합된다는 것이다.

[0265] 바람직하게, PUFA는 미생물 건조 중량의 약 5% 이상, 한 측면에서는 6% 이상, 다른 측면에서는 7% 이상, 다른 측면에서는 8% 이상, 다른 측면에서는 9% 이상, 다른 측면에서는 10% 이상으로 제조되거나, 미생물의 건조 중량의 90% 이상이 될 때까지 정수의 백분율로 제조된다(예를 들어 15%, 20%, 30%, 40%, 50% 및 그 사이의 모든 백분율).

[0266] 본 발명에 따른 목적인 생활성 화합물을 제조하는 방법에서 유전적으로 변형된 식물은 발효 배지에서 배양되거나 흙과 같이 적합한 배지에서 성장된다. 적합한 또는 효율적인 발효 배지는 앞서 상세히 설명되어졌다. 고등식물을 위한 적합한 성장 배지는 이에 제한되는 것은 아니지만, 흙, 모래, 뿌리 성장(예를 들어 질석, 진주암, 등) 또는 수경 재배를 지지하는 미립자 매개물 뿐만 아니라 적합한 광, 물 및 고등식물의 성장을 최적화하는 영양학적 보충물이 포함된다. 본 발명에 따른 유전적으로 변형된 식물은 본 발명에 따라 유전적으로 변형된 PKS 시스템의 활성을 통하여 상당한 양의 목적인 생산물을 제조하도록 조작된다. 상기 합성물은 식물로부터 화합물을 추출하는 정제 공정을 통해 회수될 수 있다. 바람직한 양태에서 화합물은 식물을 수확함으로써 회

수된다. 이러한 양태에서 식물은 천연의 상태로 소비될 수 있거나 소비되는 생산물로 추가로 가공된다.

[0267] 생활성 분자를 제조하는데 유용한 많은 유전적 변형은 본 발명에서 주어진 바와 같이 당업자에는 분명하고, 이외의 다양한 변형이 앞서 설명되었다. 본 발명은 목적인 생활성 분자의 제조를 유도하는 본 발명에서 언급된 PUFA PKS 시스템과 관련된 유전적 변형을 고려한다.

[0268] 본 발명에 따른 생활성 분자는 생물학적 활성을 가진 분자(화합물, 생산물 등)를 포함하며, 이들은 본 발명에서 기술된 비박테리아의 PUFA PKS 시스템의 적어도 하나의 기능성 도메인의 생물학적 활성을 가진 적어도 하나의 아미노산 서열을 포함하는 PKS 시스템에 의해 제조될 수 있다. 그러한 생활성 분자는 이에 제한되는 것은 아니지만, 다중불포화 지방산(PUFA), 항염증 제형, 화학 요법제, 활성 부형제, 골다공증 약, 항우울제, 항경련제, 항-헬리코박터 파이로리 약, 신경퇴행성 질환의 치료를 위한 약, 퇴행성 간질환의 치료를 위한 약, 항생제 및 콜레스테롤 저해 제형을 포함할 수 있다. 본 발명에 따른 비박테리아의 PUFA PKS 시스템의 한 장점은 시스템 배형의 탄소-탄소 이중결합과 세 번째 탄소마다 이중결합을 포함하는 분자를 도입하는 상기 시스템의 능력이다. 상기 능력은 다양한 화합물을 제조하는데 활용될 수 있다.

[0269] 생활성 화합물은 미생물의 건조중량의 약 0.05%이상, 바람직하게는 약 0.1%이상, 보다 바람직하게는 약 0.25%이상, 보다 바람직하게는 약 0.5%이상, 보다 바람직하게는 약 0.75%이상, 보다 바람직하게는 약 1%이상, 보다 바람직하게는 약 2.5%이상, 보다 바람직하게는 약 5%이상, 보다 바람직하게는 약 10%이상, 보다 바람직하게는 약 15%이상, 보다 더 바람직하게는 약 20% 이상의 양으로 유전적으로 변형된 미생물에 의해 제조되는 것이 바람직하다. 지질 화합물의 경우, 그러한 화합물은 미생물의 건조 중량의 약 5%이상 제조되는 것이 바람직하다. 항생제 또는 보다 적은 양으로 합성되는 화합물과 같은 다른 생활성 화합물의 경우에는 이러한 합성물을 미생물의 건조중량의 무게만큼 소지한 계통은 위에서 설명한 타입의 새로운 PKS 시스템을 예상대로 가진 것으로 확인할 수 있다. 일부 양태에서, 특정 생활성 분자(화합물)는 축적되기 보다는 미생물에 의해 분비된다. 따라서, 이러한 생활성 분자는 작용의 분자는 배양 배지로부터 회수되며, 제조된 분자의 농도는 미생물과 배양의 크기에 따라 달라질 것이다.

[0270] 본 발명의 한 양태는 PUFA PKS 시스템의 적어도 하나의 생물학적으로 활성인 도메인을 코딩하는 핵산 서열을 포함한 적어도 하나의 재조합 핵산 분자를 발현하는 재조합 숙주 세포에 의해 제조되는 오일을 최종 생산물에 첨가하는 단계를 포함한 적어도 하나의 지방산을 포함하는 최종 생산물을 변형하는 방법에 관한 것이다. PUFA PKS 시스템은 트라우스토카이트리움과 스킨조카이트리움에서 유래된 PUFA PKS 시스템 또는 쉐와넬라 올레야나 또는 쉐와넬라 자포니카와 같이 22°C 이상의 온도(즉, 정상 또는 천연의 조건)에서 정상적으로 성장하고 PUFA를 제조하는 박테리아에서 유래된 PUFA PKS 시스템을 포함한, 본 발명에서 개시된 적합한 박테리아 또는 비박테리아의 PUFA PKS 시스템을 포함한다.

[0271] 바람직하게 최종 생산물은 음식,식이 보조제, 약제학적 제형, 인간화된 동물 밀크와 분유로부터 선택된다. 적합한 약제학적 제형은 이에 제한되는 것은 아니지만, 항염증 제형, 화학 요법제, 활성 부형제, 골다공증약, 항우울제, 항경련제, 항-헬리코박터파이로리약, 신경퇴행성 질환의 치료를 위한 약, 퇴행성 간질환의 치료를 위한 약, 항생제, 콜레스테롤 저해 제형을 포함한다. 한 양태에서 최종 생산물은 만성 염증, 급성 염증, 위장관계 질환, 암, 카킵시아, 심장의 재협착증, 신경퇴행성 질환, 간의 퇴행성 질환, 혈액 리피드 질환, 골다공증, 동맥경화, 자가면역 질환, 자간전증, 조산, 노화와 연관된 황발변증, 호흡기계 질환 및 폐옥시조말 질환으로부터 선택된 용태를 치료하는데 이용된다.

[0272] 적합한 음식물은 이에 제한되는 것은 아니지만, 제과점 제품, 빵과 롤, 아침식사용 시리얼, 가공된 및 가공되지 않은 치즈, 조미료(케첩, 마요네즈 등), 유제품(밀크, 요구르트), 푸딩과 젤라틴 후식, 탄산음료, 차, 분말화된 음료 믹스, 가공된 어류, 과일을 원료로 한 음료, 껌, 딱딱한 사탕, 냉동 유제품, 가공된 육류, 콩과 콩을 원료로 한 스프레드(spread), 파스타, 가공된 조류, 그레이비와 소스, 포테이토칩과 다른 칩 또는 비스킷, 초코렛과 다른 사탕, 스프와 스프 믹스, 콩을 원료로 한 제품(밀크, 음료, 크림, 표백제), 식물성 유지를 원료로 스프레드, 야채를 원료로 한 음료를 포함한다.

[0273] 본 발명의 다른 양태는 인간화된 동물 밀크를 제조하는 방법과 연관되어 있다. 상기 방법은 본 발명에서 개시된 PUFA PKS시스템의 적어도 하나의 생물학적으로 활성인 도메인을 코딩하는 핵산 서열을 포함한 적어도 하나의 재조합 핵산 분자로 밀크-제조 동물의 밀크-제조 세포를 유전적으로 변형하는 단계를 포함한다.

[0274] 숙주세포를 유전적으로 변형하는 방법과 유전적으로 변형된 비인간의 밀크 제조 동물을 제조하는 방법은 당업계에 공지되어 있다. 변형시킬 숙주 동물의 예는 소, 양, 돼지, 염소, 야크 등을 포함하며, 이들은 유

전적 조작과 트랜스 유전자를 발현하는 개체군의 급속한 위한 클로닝에 수정될 수 있다. 예를 들어, PKS-유사 트랜스 유전자는 유전자 조절 영역의 변형을 통하여 표적 세포내 소기관, 조직 및 체액에서의 발현을 위해 순응될 수 있다. 숙주 동물의 모유에서 PUFA를 제조하는 것이 적절하다.

[0275] 하기의 실시예는 오로지 본 발명을 구체적으로 설명하기 위한 것으로 이들 실시예에 의해 본 발명의 범위를 한정하는 것은 아니다.

실시예

[0276] <실시예1>

[0277] 미국특허출원 제10/124,800호로부터의 하기의 실시예는 본 발명에 따른 PUFA PKS 시스템을 포함한 다른 비박테리아의 유기체를 동정하기 위한 본 발명의 스크리닝 과정의 이용을 보여준다.

[0278] 트라우스토카이트리움 sp. 23B(ATCC 20892)은 본 발명에 하기에 상세히 개시된 것처럼 배양되어졌다. 트라우스토카이트리움 sp. 23B(ATCC 20892)의 냉동된 바이얼이 50mL의 RCA 배지를 담고 있는 250mL셰이크 플라스크를 접종하는데 사용되어졌다. 배양액은 셰이커 테이블에서 (200rpm) 72시간 동안 25°C에서 진탕시켰다. RCA 배지는 다음을 함유하고 있다:

[0279] RCA 배지

[0280] 탈이온수 1000mL

[0281] Reef Crystals®

바다 소금 40g/L

[0282] 글루코스 20g/L

[0283] 모노소듐 글루타메이트(MSG) 20g/L

[0284] 이스트 추출물 1g/L

[0285] PII 금속* 5mL/L

[0286] 비타민 믹스* 1mL/L

[0287] pH 7.0

[0288] *PII 금속 믹스와 비타민 믹스는 본 발명에 본문 전체를 참고문헌으로 통합한 미국특허 제5,130,742호에 개시된 것과 동일하다.

[0289] 25mL의 72시간된 배양액이 50mL의 낮은 질소 RCA 매개물(20g/L 대신10g/L MSG)을 함유한 또 다른 250mL의 셰이크 플라스크를 접종하는데 사용하였고, 다른 25mL의 배양액은 175mL의 낮은 질소 RCA 배지를 함유한 250mL의 셰이크 플라스크를 접종하는데 사용하였다. 두 플라스크는 72시간 동안 25°C에서 셰이크 테이블에 놓여졌다. 세포는 원심분리를 통해 채취되었고 동결건조에 의해 건조되었다. 건조된 세포는 표준 가스 크로마토그래피 과정을 이용해서 지방 함유량과 지방산 프로파일과 함량이 분석되었다.

[0290] 저산소의 조건하에 있을 때와 고산소 조건하에 있을 때 비교하는 트라우스토카이트리움 23B의 스크리닝 결과는 다음과 같다:

[0291] DHA의 %FAME은 증가했는가? 예(38->44%)

[0292] C14:0 + C16:0+C16:1은 40%의 TFA보다 큰가? 예(44%)

[0293] No C18:3(n-3) 혹은 C18:3(n-6)? 예(0%)

[0294] 지방 함량은 증가했는가? 예(2배 증가)

- [0295] DHA(혹은 다른 HUFA 함량이 증가했는가?) 예(2.3배 증가)
- [0296] 위의 결과는, 특히 저산소의 조건하에서 DHA 함량(%FAME)의 현저한 증가는 트라우스토카이트리움 균주에서 PKS를 제조하는 PUFA의 존재를 강력하게 보여주는 것이다.
- [0297] PUFA PKS의 존재를 확인하는 추가의 정보를 제공하기 위해 PUFA를 제조하는 PKS 시스템(즉, 전술한 서열번호: 1-32)을 갖고 있는 것으로 이미 결정된 균주인 스킨조카이트리움 계통 20888에서 유래한 PKS 프로브를 이용한 트라우스토카이트리움 23B의 서던 블롯이 수행되어졌다. 트라우스토카이트리움 23B 계통의 DNA 단편은 PKS PUFA 합성 유전자에서 유래된 혼성 프로브와 상동성이 있는지, 이는 서던 블롯 기술을 통해 감지되었다. 트라우스토카이트리움 23B 계통의 DNA는 ClaI 혹은 KpnI 제한 엔도뉴클레아제로 소화되어지고, 아가로스 겔 전기영동(0.7% 아가로스, 표준의 트리스아세트이트-EDTA 완충용액)에 의해 분리되고, 모세관의 전이에 의해 Schleicher & Schuell Nytran Supercharge 멤브레인으로 블롯되었다. 혼성 프로브로 표시한 두 가지의 디옥시제닌이 이용되어졌다. 하나는 스킨조카이트리움 PKS OrfB(OrfB의 뉴클레오티드 5012-5511; 서열번호: 3)의 엔오일-ACP 환원효소(ER) 부위에 특이적이고, 다른 하나는 스킨조카이트리움 PKS OrfC(OrfC의 뉴클레오티드 76-549; 서열번호: 5)의 개시점인 보존된 부위에 특이적이다.
- [0298] OrfB-ER 프로브는 약 13kb ClaI 단편과 약 3.6kb KpnI 단편을 트라우스토카이트리움 23B 계통의 DNA에서 감지했다. OrfC 프로브는 약 7.5kb ClaI 단편과 약 4.6kb KpnI 단편을 트라우스토카이트리움 23B 계통의 DNA에서 감지했다.
- [0299] 마지막으로 벡터 람다 FIX II(Stratagene)으로 삽입된 트라우스토카이트리움 23B 계통의 DNA 유래의 DNA 단편으로 구성된 제조합 계통의 라이브러리는 스킨조카이트리움 20888 PUFA-PKS 유전자의 다음의 절편에 상응하는 디옥시제닌으로 표시된 프로브를 이용하여 스크리닝되어졌다: OrfA의 뉴클레오티드 7385-7879(서열번호: 1), OrfB의 뉴클레오티드 5012-5511(서열번호: 3), OrfC의 뉴클레오티드 76-549(서열번호: 5). 이러한 프로브는 각각 트라우스토카이트리움 23B 라이브러리에서 양성의 플라크를 감지했는데, 이는 스킨조카이트리움 PUFA-PKS 유전자와 트라우스토카이트리움 23B의 유전자 사이에 상당한 상동성을 의미한다.
- [0300] 위의 결과는 트라우스토카이트리움 23B 계통의 DNA가 스킨조카이트리움 20888의 PKS 유전자와 상동성 있는 서열을 갖고 있음을 보여주었다.
- [0301] <실시예 2>
- [0302] 다음의 실시예는 스킨조카이트리움 Orfs A, B, C가 대장균에서 기능적인 발현을 통해 기능적인 DHA/DPA 합성 효소를 코딩한다는 사실을 보여준다.
- [0303] 대장균 형질전환체의 일반적인 제조
- [0304] 스킨조카이트리움에서 DHA와 DPA를 제조(Orfs A, B & C; 서열번호: 1, 서열번호: 3, 서열번호: 5, 각각)하는 스킨조카이트리움 PUFA PKS 시스템을 코딩하는 세 개의 유전자는 단일 대장균 발현 벡터 pET21c(Novagen)에 클론되었다. 유전자는 단일의 메시지로 (T7RNA-폴리메라제에 의해) 전사되고, 각각의 유전자 앞에서 클론을 만든 리보솜 결합 부위는 해독을 개시한다. OrfB 코딩 서열의 변형은 대장균에서 (아래에 서술한) OrfB 단백질 전장의 제조를 수득하기 위해 필요하였다. PPTase(아래에 서술한)를 코딩한 보조의 (accessory) 유전자는 두 번째의 플라스미드(pACY184, New England Biolabs에서 유래한)에 클로닝되었다.
- [0305] OrfB
- [0306] OrfB 유전자는 ~224 kDa의 질량의 단백질을 코딩하는 것으로 예측할 수 있다. 대장균에서 유전자의 발현에의 최초의 시도는 겔보기 분자량 ~165kDa의 단백질(SDS-PAGE동안 공지된 질량의 단백질에 비교하여 판단했을 때)을 축적하는 것으로 귀착한다. OrfB 뉴클레오티드 서열의 조사는 15 연속의 세린 코돈-모두가 TCT 코돈-을 포함하고 있는 부분을 드러낸다. 유전자 코드는 6개의 다른 세린 코돈을 갖고 있고, 이중 3개는 대장균에서 자주 쓰인다. 본 발명자들은 4개의 겹치는 올리고뉴클레오티드를 폴리메라제 쉐 반응 프로토콜과 조합해서 사용하여 세린 코돈 반복 영역을 갖고 있는 OrfB 유전자의 작은 부분(~195 base pair, BspHI에서 SacII제한 효소 단편)을 재합성한다. 합성된 OrfB 단편에서, 대장균에서 흔히 사용되는, 3 세린 코돈의 임의적인 혼합이 이용되고, 잠정적으로 문제가 있는 다른 코돈(즉, 대장균에 의해 거의 이용되지 않은 기타 코돈)도 역시 교체된다. 원래의 OrfB에 존재하는 BspHI에서 SacII단편은 재합성된 단편(OrfB*를 얻는)에 의해 대체되고 변형된 유전자는

관련된 발현 벡터로 클로닝된다. 변형된 OrfB^{*}는 여전히 서열번호: 4의 아미노산 서열을 코딩한다. 대장균에서 변형된 OrfB^{*} 클론의 발현은 ~224 kDa 단백질의 출현으로 이어지는데, 이는 OrfB의 전장 산물이 제조되었음을 나타내는 것이다. OrfB^{*} BspHI에서 SacII 단편의 재합성된 서열은 서열번호: 80에서 나타내진다. 서열번호: 80에 언급하여, 재합성된 OrfB의 BspHI에서 SacII 영역의 뉴클레오티드 서열이 나타내진다. BspHI 제한 부위와 SacII 제한 부위 동정된다. BspHI 부위는 OrfB CDS의 뉴클레오티드 4415에서 시작된다(서열번호: 3)(주의: SacII 부위가 독특한 반면, OrfB CDS에 총 3개의 BspHI 부위가 있다). 변형되지 않은 OrfB CDS의 서열이 GenBank Accession No. AF378328와 서열번호: 3으로 주어진다.

[0307] PPTase

[0308] OrfA 단백질(스키조카이트리움에서 서열번호: 2)의 ACP 도메인은 기능을 발휘하기 위해 포스판테테인 그룹의 첨가에 의해 활성화되어야 한다. 이러한 반응의 일반적인 유형을 촉매하는 효소는 포스판테테인 트랜스퍼라제라 불린다(PPTase). 대장균은 2 가지의 내인성의 PPTase를 갖고 있는데, 이 PPTase는 스키타카이트리움에서 유래하는 OrfA ACP 도메인을 인식하지 못할 것으로 예상되었다. 이는 대장균에서 OrfA, B^{*}(상기 참조) 및 C를 추가의 PPTase없이 발현하는 것으로 입증되었다. 이러한 형질전환체에서 DHA 제조는 감지되지 않았다. 발명자들은 대장균 PUFA PKS 발현 시스템에서 두 개의 (이형성의)-PPTase를 시험해봤다: (1)sfp (바실러스 섭틸리스 유래) 그리고 (2) Het I (시아노박테리아 노스톡 균주 7120 유래)

[0309] sfp PPTase는 잘 특징지어졌고 다양한 영역의 기질을 인지하는 능력 때문에 널리 사용되고 있다. 발표된 서열 정보를 기초로(Nakana, et al. 1992, *Molecular and General Genetics* 232:313-321), 한정된 업스트림 또는 다운스트림 측면에 위치한 DNA 서열과 함께, sfp를 위한 발현 벡터가 코딩한 영역을 클로닝함으로써 구축되어 pACYC-184 클로닝 벡터가 되었다. 다음의 올리고뉴클레오티드가 계놈의 B. 섭틸리스 DNA로부터의 관심 부분을 확장하기 위해 쓰였다.

[0310] CGGGGTACCCGGGAGCCGCTTGGCTTTGT(정방향; 서열번호: 73)

[0311] 그리고

[0312] AAACTGCAGCCCGGTCCAGCTGGCAGGCACCCTG(역방향; 서열번호: 74)

[0313] 편리한 제한 효소 부위는 올리고뉴클레오티드에 포함되어 있어 중간 정도로 높은 카피수의 벡터에 클로닝을 용이하게 하였다. 그리고 마침내 플라스미드: pBR 301을 제조하기 위해 pACYC184의 EcoRV 부위에 포함되었다. 플라스미드와 함께 형질전환된 대장균 추출물의 조사는 sfp에 기대되는 이동성을 지닌 신규한 단백질의 존재를 드러냈다. 특정한 조건하에, Orf A, B^{*}, C 단백질을 발현하는, 세포에서의 sfp 작제물의 공동발현은 DHA 제조를 결과로 가져온다. 이 실험은 sfp가 스키타카이트리움 Orf A ACP 도메인을 활성화시킬 수 있음을 입증한다. 뿐만 아니라, sfp 유전자와 관련된 조절 요소는 다른 유전자를 삽입할 수 있는 발현 카세트를 제조하는데 사용되어졌다. 특히, pBR 301에서 sfp 코딩 영역(ATG 바로 업스트림의 3개 뉴클레오타이드)은 여러 독특한(이 작제물에 관한) 제한 효소 부위를 갖도록 설계된 DNA의 53 염기쌍 섹션으로 대체되었다. 이 영역의 초기 제한 효소 부위는 NdeI(CATATG; 서열번호: 79)이다. 이 부위에 삽입된 ATG 서열은 도입된 유전자의 개시 메티오닌 코돈으로 사용되어질 수 있다. 추가의 제한 부위(BglII, NotI, SmaI, PmeI, HindIII, SpeI 그리고 XhoI)는 클로닝 과정을 촉진하게 하는데 포함된다. 이 발현 벡터 카세트의 기능성은 5' 말단의 NdeI 부위와 3' 말단의 XhoI 부위로 sfp 버전을 생성시키기 위해 PCR을 사용하여 시험되었다. 이 단편은 발현 카세트로 클로닝되고 Orf A, B^{*} 및 C 발현 벡터와 함께 대장균으로 전달되었다. DHA로 축적된 이 세포들은 적절한 조건 하에 기능적인 sfp가 제조되었음을 입증하였다.

[0314] 본 발명자들은 아는 한도 내에서는, Het I는 이형성의 상황에서는 이전에 시험되어보지 않았다. Het I는 장쇄 하이드록시 지방산의 합성에 책임 있는 것으로 알려진 노스톡에 있는 유전자 무리(cluster)에 존재한다. 상기 장쇄 하이드록시 지방산은 유기체의 헤테로사이트(heterocyte)에 존재하는 글리코 리피드 층의 구성요소이다. 본 발명자들은, 이론과 상관없이, Het I가 무리에 존재하는 단백질, Hgl E의 ACP 도메인을 활성화시키는 것으로 믿는다. Hgl E의 두 ACP 도메인은 스키타카이트리움 OrfA에서 볼 수 있는 ACP 도메인에 고도의 서열 상동성을 갖는다. Het I의 내인성 개시 코돈은 동정되지 않았다(추정의 단백질에 메티오닌이 존재하지 않는다). 오픈 리딩 프레임의 5'말단 근방에 여러 잠재적인 대안의 개시 코돈(예를 들어 TTG와 ATT)이 존재한다. HetI 유전자를 코딩한 노스톡 DNA 영역의 서열은 서열번호: 81에서 볼 수 있다. 서열번호: 82는

서열번호: 81에 의해 코딩된 아미노산 서열을 나타낸다. 서열번호: 81을 참조하면, 구조의 의미없는 세 개 (TAA)로 나타내지는 업스트림 코딩 영역을 서열번호:81의 위치 1-3으로 제한하고, 정지 코돈(TGA)으로 서열번호: 81의 위치 715-717에서 종결한다. 이 서열에는 메티오닌 코돈(ATG)은 존재하지 않는다. 잠재적인 대안의 개시 코돈은: TTG 코돈(서열번호: 81의 위치 4-6, 7-9 그리고 49-51), ATT(서열번호: 81의 위치 76-78) 그리고 GTG(서열번호: 81의 위치 235-237). Het I 발현 작제물은 PCR을 이용해 가장 먼 5' 잠재적인 대행 개시 코돈(TTG)을 메티오닌 코돈(ATG, 상기에서 설명한 NdeI 제한 효소 인식 부위의 일부)으로 대체하고 코딩 서열의 3' 말단에 XhoI 부위를 도입함으로 만들어졌다. 변형된 HetI 코딩 서열은 sfp 조절 요소를 포함하고 있는 pACYC 184 벡터 작제물의 NdeI와 XhoI 부위로 삽입되었다. 대장균에 있는 이 HetI 작제물의 발현은 서열 데이터로부터 기대했던 크기의 신규한 단백질의 출현을 결과로 가져온다. 여러 조건 하에 대장균에서 스킴조카이트리움 Orfs A, B*, C와 함께 HetI의 공동발현은 이러한 세포들에 DHA와 DPA의 축적을 결과로 가져온다. sfp와 Het I 가 비교된 모든 실험에서, HetI 작제물을 포함하는 경우, sfp 작제물을 포함하는 경우보다 세포에 DHA 와 DPA가 많이 축적되었다.

[0315] 대장균 형질전환체에서 DHA와 DPA의 제조

[0316] 코딩하는 두개의 플라스미드: (1)스키조카이트리움 PUFA PKS 유전자 (Orfs A, B* 및 C) 그리고 (2) 유도할 수 있는 T7 RNA 폴리머라제 유전자를 포함하고 있는 대장균 균주 BL21로 PPTase(sfp 혹은 HetI으로부터)가 형질전환되었다. PPTase의 발현이 별도의 조절하는 요소에 의해 통제되는 동안, 스킴조카이트리움 단백질의 합성은 IPTG의 배지로의 첨가에 의해 유도되었다(위를 참조). 두 가지 이형성의 PPTase 유전자 중에서 하나를 이용하면서 세포는 다양한 한정된 조건하에서 성장하였다. 세포는 채취되었고, 지방산은 메틸에스테르(FAME)로 전환되었고 기체-액체 크로마토그래피를 사용해서 분석되어졌다.

[0317] 여러 가지 조건 하에, DHA와 DPA가 스킴조카이트리움 PUFA PKS 유전자와 두 가지 이형성의 PPTase 둘 중 하나를 발현하는 대장균 세포에서 감지되었다. DHA 혹은 DPA가 대조군 세포로부터 만들어진 FAME에서는 감지되지 않았다(즉, Orf 중 하나가 부족한 플라스미드로 형질전환된 세포). 대장균에서 관찰된 DHA대 DPA의 비율은 스킴조카이트리움에서 관찰되는 내인성의 DHA와 DPA 제조 비율에 근접했다. 전체 FAME의 ~17%를 나타내는, PUFA(DHA 더하기 DPA)의 가장 높은 수준이 10%(중량) 글리세린으로 보충된 765 배지(American Type Culture Collection에서 방법을 입수할 수 있는)에서 32℃에서 성장한 세포에서 발견되었다. PUFA 축적은 또한 세포가 5 내지 10% 글리세린으로 보충된 루리아 브로스(Luria Broth)에서 20℃에서 성장하였을 때 관찰될 수 있음을 주목한다. 각각의 플라스미드의 존재를 위한 선별은 성장하는 동안 적절한 항생제를 포함시킴으로써 유지되었고 IPTG(0.5mM의 최종 농도)는 Orf A, B* 및 C의 발현을 유도하게 위해 이용되어졌다.

[0318] 도 4는 대조군 세포와 스킴조카이트리움 PUFA PKS 유전자 더하기 PPTase(이 경우 Het I)를 발현하는 세포 및 대조군 세포로부터 추출한 FAME의 기체-액체 크로마토그래피에 의한 예로서의 크로마토그램을 보여준다. 표지된 FAME의 동정은 질량 분석기를 이용해 확인되었다.

[0319] 실시예3

[0320] 다음의 실시예는 스킴조카이트리움 PUFA PKS 효소 복합체를 코딩한 유전자가 선별적으로 비활성화(녹아웃)될 수 있음을 보여주는데, 이는 배지가 다중불포화지방산으로 보충되지 않는 한 치명적인 표현형이다.

[0321] 상동성의 제조합은 스킴조카이트리움(미국특허출원 제10/124,807호, 인용문으로 전문이 여기 실렸다)에서 보여주었다. 스킴조카이트리움 OrfA(서열번호: 1)를 비활성화하도록 설계된 플라스미드가 Zeocin™ 저항 마커를 Orf A코딩 서열을 포함하는 클론의 Sma I 부위로 삽입함으로써 만들어졌다. Zeocin™ 저항 마커는 플라스미드 pMON50000으로부터 얻었고 - Zeocin™저항 유전자의 발현은 튜블린 프로모터 요소로부터 추출된 스킴조카이트리움에 의해 추진되었다(미국특허출원 제10/124,807호 참조). 녹아웃 작제물은 다음과 같이 구성된다: 5' 스킴조카이트리움 OrfA 코딩 서열, tub-Zeocin™ 저항 요소 그리고 3'스키조카이트리움 OrfA 코딩 서열, 모두 pBluescript II SK (+)벡터(유전자층)으로 클론화.

[0322] 플라스미드는 입자 사격(particle bombardment)에 의해 스킴조카이트리움에 도입되었고, 형질전환체는 Zeocin™을 함유한 플레이트에 선별되어지고, 다중불포화지방산(PUFA)으로 보충되어졌다(실시예 4 참조). Zeocin™ 더하기 PUFA 플레이트에서 성장하는 콜로니는 PUFA 보충물 없이도 플레이트에서 성장하는 능력을 테스트 받았고 여러 PUFA를 필요로 함을 발견하였다. 이러한 PUFA 영양 요구체는 추정되고 있는 OrfA 녹아웃이다.

이러한 몇몇의 돌연변이체로부터 추출한 RNA의 노던 블롯 분석은 전장의 OrfA 메시지가 이러한 돌연변이체에서는 제조되지 않았음을 확증해주었다. 이러한 실험은 스킨조카이트리움 유전자(예를 들면, OrfA)가 상동성의 재조합을 통해 비활성화될 수 있고, OrfA의 비활성화는 치명적인 표현형의 결과를 낳고, 이러한 돌연변이체는 PUFA를 가지고 있는 배지의 보충물을 통해 제조될 수 있다는 사실을 보여준다.

[0323] 스킨조카이트리움 OrfB(서열번호: 3)과 OrfC(서열번호: 5)의 비활성화를 지시한 유사한 세트의 실험은 유사한 결과를 가져왔다. 즉 OrfB와 OrfC는 상동성의 재조합에 의해 개별적으로 비활성화될 수 있고 이러한 세포는 성장을 위해 PUFA 보충을 필요로 한다.

[0324] 실시예 4

[0325] 다음의 실시예는 PUFA 영양요구체가 EPA로 보충된 배지에서 유지될 수 있음을 보여주는데, 이는 EPA가 스킨조카이트리움에서 DHA를 대체할 수 있음을 입증하는 것이다.

[0326] 실시예 3에서 설명했듯이, PUFA PKS 복합체가 비활성화된 스킨조카이트리움 세포는 PUFA가 생존하기 위해서는 보충물이 필요하다. 스킨조카이트리움이 성장을 위해 본 시스템의 생산물에 의존한다는 것을 입증하는 것 외에, 이 실험적인 시스템은 이 돌연변이체를 구제할 수 있는 다양한 지방산의 능력을 시험해보는 것을 허용한다. 돌연변이 세포(세 유전자 중 어느 것이든 비활성화된)가 DHA로 보충한 배지에서 잘 성장하였듯이 EPA로 보충한 배지에서도 잘 성장함을 발견하였다. 이러한 결과는 DHA를 제조하는 내인성의 PUFA PKS 복합물이 EPA를 제조하는 물질로 대체되었을 때, 세포가 생존 가능함을 의미한다. 또한, 이러한 돌연변이 세포는 ARA 혹은 GLA의 보충에 의해 구제될 수 있는데, 이는 이러한 생산물을 제조하는 유전적으로 변형된 스킨조카이트리움을 제조하는 실험적 여부를 입증하는 것이다. PUFA로 보충하는 바람직한 방법은 PUFA를 배지에 첨가하기 전에 유리 지방산을 부분적으로 메틸화된 베타-사이클로덱스트린과 합치는 것에 관계한다.

[0327] 실시예 5

[0328] 다음의 실시예는 비활성화된 PUFA 유전자가, PUFA 합성을 복원하기 위해 같은 부위에서 유전자의 활성화된 형태로 대체될 수 있음을 보여준다.

[0329] 아세트락테이트 신타제 유전자 부위에서의 이중의 상동성의 재조합은 스킨조카이트리움에서 입증되었다(미국특허출원 제10/124,807호 참조). 본 발명자들은 스킨조카이트리움 OrfA 녹아웃 균주의 형질전환에 의한 스킨조카이트리움 PUFA 유전자의 대체를 전장의 스킨조카이트리움 OrfA 게놈 클론으로 하는 개념을 테스트해봤다. 형질전환체가 보충하는 PUFA 없이 배지에서 성장하는 능력에 따라 선별되어졌다. 이러한 PUFA 원시영양체(prototroph)는 Zeocin™에 대한 내성에 대해 테스트 받고 항생제에 민감한 몇 가지를 찾았다. 이러한 결과는 도입된 스킨조카이트리움 OrfA가 Zeocin™ 저항 유전자를 녹아웃 균주에서 이중의 상동성의 재조합을 통해 대체하였음을 보여준다. 이러한 실험은 PUFA PKS 유전자내에서 유전자 대체에 대한 개념의 입증을 보여준다. 스킨조카이트리움 OrfB와 OrfC의 녹아웃에 대한 유사한 실험도 같은 결과를 가져왔다.

[0330] 실시예 6

[0331] 이 실시예는 트라우스토카이트리움 23B PUFA PKS 유전자 전부 혹은 일부가 스킨조카이트리움에서 기능할 수 있음을 보인다.

[0332] 미국특허출원 10/124,800호에 개시된 것처럼, DHA를 제조하는 원생생물 트라우스토카이트리움 23B(Th.23B)는 orfA, orfB 및 orfC의 상동체를 포함하고 있음을 보여준다. 3개의 Th.23B 유전자의 완전한 게놈의 클론이 동족의 orf "녹-아웃"을 포함한 스킨조카이트리움 균주를 형질전환하는데 사용하곤 하였다. 보완해주는 형질전환체에 대한 직접적인 선별이 PUFA 보충의 부재중에 행해졌다. 이 방법에 의해 Th. 23B orfA와 orfC 유전자는 PUFA 원시영양으로의, 각각, 스킨조카이트리움 orfA, orfC 녹아웃 균주를 보완할 수 있음을 보인다. 보완의 형질전환체가 Zeocin™ 내성을 보유하거나 잃어버린 것으로 발견되었다(스킨조카이트리움 유전자에 삽입된 마커는 그래서 녹아웃을 규정한다). Zeocin™ 내성 보완의 형질전환체는 각각의 orf 영역의 스킨조카이트리움 게놈 외부로 전체 트라우스토카이트리움 유전자의 단독의 크로스오버 통합에 의해 발생하는 경향이 있다. 이러한 결과는 전체의 트라우스토카이트리움 유전자가 스킨조카이트리움에 기능하고 있음을 보인다.

Zeocin™ -민감한 보완의 형질전환체는 트라우스토카이트리움 유전자의 부분(혹은 인지할 수 있는 전부)이 중단된 Zeocin™ 내성 마커를 포함하는 스킨조카이트리움 유전자의 동족 영역을 기능적으로 대체하는 이중의 크로스 오버 이벤트에 의해 발생하는 경향이 있다. 이러한 결과는 트라우스토카이트리움 유전자의 분획이 스킨조카이트리움에서 기능함을 보인다.

[0333] 실시예 7

[0334] 다음의 실시예는 PUFA PKS 유사한 유전자를 포함한 일정한 EPA를 제조하는 박테리아가 스킨조카이트리움의 변형에 적합함을 보여준다.

[0335] 쉼와넬라 속의 두 가지의 EPA-제조하는 해양 박테리아 균주는 스킨조카이트리움 발효의 전형적인 온도에서 성장하고 PUFA PKS와 유사한 유전자를 갖고 있음을 보여준다. 쉼와넬라 올레야나(Australian Collection of Antarctic Microorganism (ACAM) 균주번호 644; SKerratt et al. *Int. J. Syst. Evol. Microbiol.* 52,2101(2002))는 EPA를 제조하고 30℃에서까지 성장한다. 쉼와넬라 자포니카(American Type Culture Collection(ATCC) 균주번호 BAA-316; Ivanova et al. *Int. J. Syst. Evol. Microbiol.* 51, 1027(2001))는 EPA를 제조하고 35℃에서까지 성장한다.

[0336] 이러한 박테리아 균주로부터 PUFA-PKS 유전자를 동정하고 분리시키기 위해서는, orf5/pfaA 유전자의 KS-MAT 영역과 박테리아의 orf7/pfaC 유전자의 DH-DH 영역을 위한 PCR 프라이머 쌍을 축퇴시키고, 이 유전자들은 쉼와넬라 SCRC-2738, 쉼와넬라 오네이텐시스 MR-1; 쉼와넬라 sp. GA-22; 포토박터 프로판둠, 그리고 모리텔라 마리나 (상기의 설명 참조)를 위한 공개된 유전자 서열에 기초한다. 특히 프라이머 및 PCR 조건은 다음과 같이 설계되었다.

[0337] KS/AT 영역을 위한 프라이머; 이하의 공개된 서열을 기초하여: 쉼와넬라 sp. SCRC-2738; 쉼와넬라 오네이텐시스 MR-1; 포토박터 프로판둠; 모리텔라 마리나:

[0338] prRZ23 GGYATGMTGRTTGGTGAAGG (정방향; 서열번호: 69)

[0339] prRZ24 TRTTSASRTAYTGYGAACCTTG (역방향; 서열번호: 70)

[0340] DH 영역을 위한 프라이머: 이하의 공개된 서열을 기초하여: 쉼와넬라 sp. GA-22; 쉼와넬라 sp. SCRC-2738; 포토박터 프로판둠; 모리텔라 마리나:

[0341] prRZ28 ATGKCNAGAAGTTGTGGCCA (정방향; 서열번호: 71)

[0342] prRZ29 CCWGARATRAAGCCRTTDGGTTG (역방향; 서열번호: 72)

[0343] PCR 조건(주형으로서의 박테리아 염색체 DNA)은 다음과 같았다:

[0344] 반응 혼합물:

[0345] 0.2 μM dNTPs

[0346] 0.1 μM 각 프라이머

[0347] 8% DMSO

[0348] 250ng 염색체의 DNA

[0349] 2.5U Herculase[®] DNA polymerase (Stratagene)

[0350] 1X Herculase[®] 완충액

[0351] 50 μL 총 부피

[0352] PCR 프로토콜: (1) 98℃에서 3분; (2) 98℃에서 40초; (3) 56℃에서 30초; (4) 72℃에서 90초; (5) 2-4단계를 29 사이클로 반복; (6) 72℃에서 10분; (7) 6℃에서 보존.

[0353] 양 프라이머 쌍을 위해, PCR은 쉼와넬라 올레야나 혹은 쉼와넬라 자포니카 둘중 하나로부터 염색체의 DNA 주형을 이용하여 예상의 크기를 가진 별개의 생산물을 주었다. 4가지 각기의 PCR 산물은 pCR-BLUNT II-

TOPO(Invitrogen)으로 클로닝되고 서열을 삽입하는 것은 M13 정방향과 역방향의 프라이머를 이용해 결정되었다. 이 모든 경우, 수득한 DNA 서열은 공지된 박테리아의 PUFA PKS 유전자 영역에 매우 상동성이 있었다.

[0354] 박테리아의 PCR 산물로부터 수득한 DNA 서열은 공지된 서열과, 그리고 표준의 Blastx 조사(BLAST 파라미터: 낮은 복잡성 필터; 온 ; 매트릭스; BLOSUM62; 단어 크기 3; 캡-cost: 존재 11, 연장 1(Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J.(1997) "Gapped BLAST and PSI-BLAST: "A new generation of protein database search programs." Nucleic Acids Res. 25:3389-3402, 전문을 인용문으로 실었다))에서 스킴조카이트리움 ATCC 20888에서 유래한 PUFA PKS 유전자와 비교되었다.

[0355] 아미노산 수준에서, 서열번호: 76에 의해 코딩된 아미노산 서열을 추론한 쉼벨라 올레야나 ACAM644 케토아실 신타제/아실 트랜스페라제(KS-AT)에 고도로 상동성 있는 서열은: 포토박터 프로판둠 pfaA(동일성 = 70%; 파지티브=81%); 쉼벨라 오네이텐시스 MR-1 "복수 도메인 α -케토아실 신타제"(동일성=66%; 파지티브=77%); 모리텔라 마리나 ORF8(동일성=56%; 파지티브=71%). 스킴조카이트리움 sp. ATCC 20888 orfA는 서열번호: 76에 의해 코딩된 추론된 아미노산 서열에 41% 동일하고 56% 파지티브했다.

[0356] 아미노산 수준에서 서열번호: 78에 의해 코딩된 아미노산 서열을 추론한 쉼벨라 자포니카 ATCCBAA-316 케토아실 신타제/아실 트랜스페라제(KS-AT)에 고도로 상동성 있는 서열은: 쉼벨라 오네이텐시스 MR-1 "복수 도메인 β -케토아실 신타제"(동일성=67%; 파지티브=79%); 쉼벨라 sp. SCRC-2738 orf5(동일성=69%; 파지티브=79%); 모리텔라 마리나 ORF8(동일성=56%; 파지티브=70%). 스킴조카이트리움 sp. ATCC 20888 orfA는 서열번호: 78에 의해 코딩된 추론된 아미노산 서열에 41% 동일하고 55% 파지티브했다.

[0357] 아미노산 수준에서, 서열번호: 75에 의해 코딩된 아미노산 서열을 추론한 쉼벨라 올레야나 ACAM644 탈수소효소(DH)에 고도로 상동성 있는 서열은: 쉼벨라 sp. SCRC-2738 orf7(동일성=77%; 파지티브=86%); 포토박터 프로판둠 pfaC(동일성=72%; 파지티브=81%); 쉼벨라 오네이텐시스 MR-1 "복수 도메인 α -케토아실 신타제"(동일성=75%; 파지티브=83%). 스킴조카이트리움 sp. ATCC 20888 orfC는 서열번호: 75에 의해 코딩된 추론된 아미노산 서열에 26% 동일하고 42% 파지티브했다.

[0358] 아미노산 수준에서, 서열번호: 77에 의해 코딩된 아미노산 서열을 추론한 쉼벨라 자포니카 ATCC BAA-316 탈수소효소(DH)에 고도로 상동성 있는 서열은: 쉼벨라 sp. SCRC-2738 orf7(동일성=77%; 파지티브=86%); 포토박터 프로판둠 pfaC(동일성=73%; 파지티브=83%); 쉼벨라 오네이텐시스 MR-1 "복수 도메인 α -케토아실 신타제"(동일성=74%; 파지티브=81%). 스킴조카이트리움 sp. ATCC 20888 orfC는 서열번호: 77에 의해 코딩된 추론된 아미노산 서열에 27% 동일하고 42% 파지티브했다.

[0359] 상기 두 가지 쉼벨라 균주의 PUFA PKS 유전자 세트는 스킴조카이트리움 PUFA 제조의 변형을 위한 개별적인 도메인이나 전체 유전자의 유익한 원천을 제공할 것이다. 그 때문에 쉼벨라 올레야나 혹은 쉼벨라 자포니카에 의해 코딩된 PUFA PKS 유전자와 단백질과 도메인은 본 발명에 명백하게 포함될 것이다.

[0360] 실시예8

[0361] 본 실시예는 실시예 7에서 설명한 박테리아의 PUFA PKS 유전자 단편이 스킴조카이트리움에서 PUFA 제조를 변형하는데 어떻게 이용될지를 보여준다.

[0362] 현재까지 알려진 박테리아의 PUFA PKS 유전자 예는 3개의 유전자로 된 스킴조카이트리움과 같은 도메인을 포함한 4개의 가깝게 연결된 유전자로 존재한다. 쉼벨라 올레야나 그리고 쉼벨라 자포니카에서의 PUFA PKS 유전자는 뾰뾰하게 밀집된 배열에서 찾기가 쉽다. 실시예 7에서 동정한 상동성의 영역은 쉼벨라 올레야나와 쉼벨라 자포니카 DNA의 클론 은행으로부터 PUFA PKS 유전자 무리를 분리시키는데 쓰인다. 클론 은행은 박테리오파지 람다 벡터, 코스미드 벡터, 박테리아의 인공의 염색체 ("BAC")벡터, 혹은 당업계에 공지된 다른 방법으로 작제될 수 있다. 박테리아의 PUFA PKS 유전자를 포함한 바람직한 클론은 유전자 서열로부터 설계된 프라이머를 사용하여 실시예 7의 부분적인 유전자 서열의 PCR에서 생성된 프로브를 이용하여 콜로니나 플라크 혼성화(실시예 1에서 설명한 것 처럼)에 의해 동정될 수 있다. 신규한 박테리아의 PUFA PKS 유전자 세트의 전체의 DNA 서열은 내인성의 PUFA PKS 유전자(실시예 3, 5 및 6을 참조)에 결합된 스킴조카이트리움 균주의 형질전환을 위하여 벡터를 설계하는데 이용된다. 전체의 박테리아 유전자(서열을 코딩하는)는 스킴조카이트리움 유전자 전체를 대체하는데 이용될 수 있다. 그래서 스킴조카이트리움 유전자 발현 영역을 활용하고, 4번째의 박테리아 유전자는 게놈내에서 다른 위치로 표적화할 수도 있다. 대안으로 개별적인 박테리아의 PUFA PKS 기능적인 도메인은 "스왑(swapped)" 또는 상동성의 재조합 기술과 유사한, 유사한 스킴조카이트리움 도메인과 교환될 수

도 있다. 박테리아의 PUFA PKS 유전자 또는 도메인의 서열은 스킴조카이트리움 코돈 사용의 세세한 내용을 수용하기 위해서는 변형되어야 할지 모르지만, 이는 당업계의 기술능력 내에서이다.

[0363] 인용되거나 설명되어진 각각의 공개적인 논문은 전문을 인용문으로 실었다. 본 출원은 다음의 출원을 참조하여 병합한다. 2003년 3월 26일에 출원한 미국 가출원 제60/457,979호; 2002년 4월 16일에 출원한 미국특허출원 제10/124,800호; 2001년 6월 15일에 출원한 미국 가출원 제60/298,796호; 2001년 9월 18일에 출원한 미국 가출원 제60/323,269호; 현재 미국특허출원 제6,566,583호이고, 1999년 1월 14일에 출원한 미국특허출원 제09/231,899호, 미국특허출원 제6,140,486호이고, 1998년 6월4일에 출원한 미국특허출원 제09/090,793호.

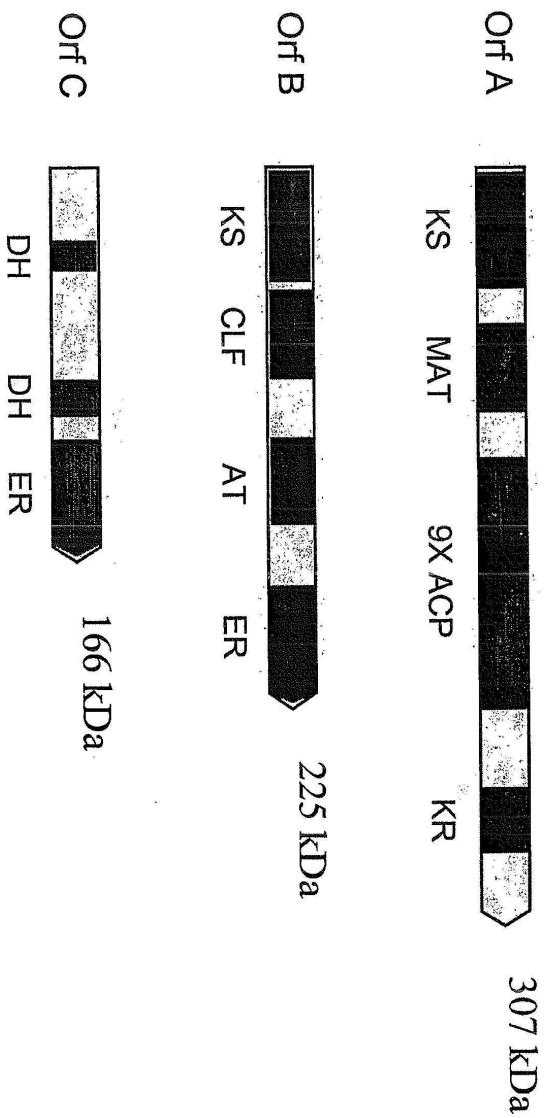
[0364] 이상에서 설명한 바와 같이, 본 발명이 속하는 기술 분야의 당업계에서는 본 발명 양태의 변형이나 각색이 일어날 수 있음은 자명한 일이다. 그러므로 모든 변경 또는 변형은 본 발명의 범위에 포함되는 것으로 해석되어야 하고, 이는 후술하는 특허청구범위에 설명되어있다.

도면의 간단한 설명

[0014] 도 1은 스킴조카이트리움 PUFA PKS 시스템의 도메인 구조를 도식적으로 나타낸 것이다.

[0015] 도 2는 스킴조카이트리움과 웨와넬라에서 유래된 PUFA PKS 시스템의 도메인을 비교하여 나타낸 것이다.

[0016] 도 3은 스킴조카이트리움 유래의 PUFA PKS 시스템과 이중 결합을 함유하지 않은 장쇄의 지방산을 제조하는 노스톡 유래의 PKS 시스템의 도메인을 비교하여 나타낸 것이다.

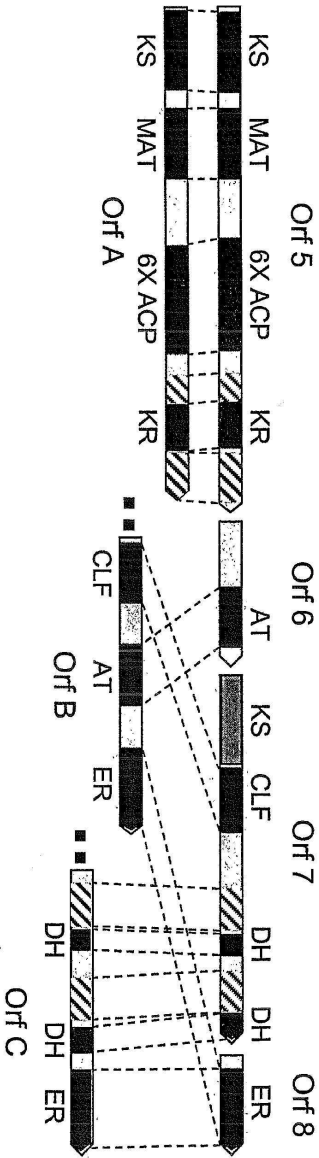


도면

도면1

PKS Orfs/도메인의 비교:

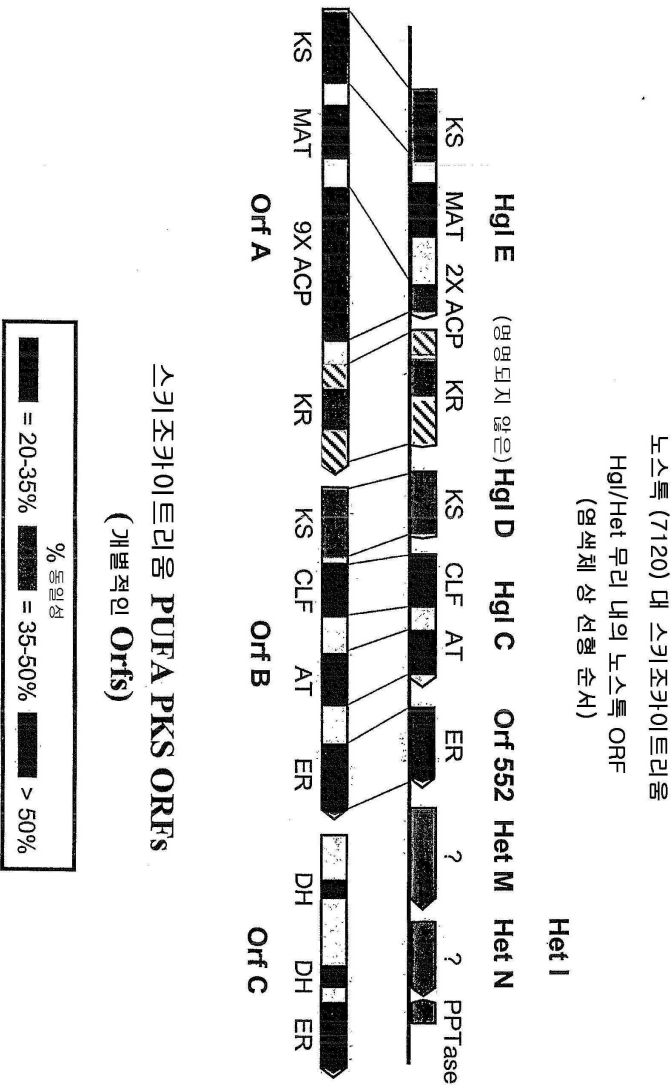
스키조카이트리움 대 쉐와넬라



아미노산 서열 상동성



도면3



서열 목록

- <110> MARTEK BIOSCIENCES CORPORATION
- <120> PUFA polyketide synthase systems and uses thereof
- <130> 2997-49-pct
- <150> US 60/457,979
- <151> 2003-03-26
- <160> 82
- <170> KopatentIn 1.71
- <210> 1

<211> 8730
 <212> DNA
 <213> Schizochytrium sp.

<220>
 <221> CDS
 <222> (1)..(8730)

<400> 1
 atg gcg gcc cgt ctg cag gag caa aag gga ggc gag atg gat acc cgc 48
 Met Ala Ala Arg Leu Gln Glu Gln Lys Gly Gly Glu Met Asp Thr Arg
 1 5 10 15

 att gcc atc atc ggc atg tcg gcc atc ctc ccc tgc ggc acg acc gtg 96
 Ile Ala Ile Ile Gly Met Ser Ala Ile Leu Pro Cys Gly Thr Thr Val
 20 25 30

 cgc gag tcg tgg gag acc atc cgc gcc ggc atc gac tgc ctg tcg gat 144
 Arg Glu Ser Trp Glu Thr Ile Arg Ala Gly Ile Asp Cys Leu Ser Asp
 35 40 45

 ctc ccc gag gac cgc gtc gac gtg acg gcg tac ttt gac ccc gtc aag 192
 Leu Pro Glu Asp Arg Val Asp Val Thr Ala Tyr Phe Asp Pro Val Lys
 50 55 60

 acc acc aag gac aag atc tac tgc aag cgc ggt ggc ttc att ccc gag 240
 Thr Thr Lys Asp Lys Ile Tyr Cys Lys Arg Gly Gly Phe Ile Pro Glu
 65 70 75 80

 tac gac ttt gac gcc cgc gag ttc gga ctc aac atg ttc cag atg gag 288
 Tyr Asp Phe Asp Ala Arg Glu Phe Gly Leu Asn Met Phe Gln Met Glu
 85 90 95

 gac tcg gac gca aac cag acc atc tcg ctt ctc aag gtc aag gag gcc 336
 Asp Ser Asp Ala Asn Gln Thr Ile Ser Leu Leu Lys Val Lys Glu Ala
 100 105 110

 ctc cag gac gcc ggc atc gac gcc ctc ggc aag gaa aag aag aac atc 384
 Leu Gln Asp Ala Gly Ile Asp Ala Leu Gly Lys Glu Lys Lys Asn Ile
 115 120 125

 ggc tgc gtg ctc ggc att ggc ggc ggc caa aag tcc agc cac gag ttc 432
 Gly Cys Val Leu Gly Ile Gly Gly Gly Gln Lys Ser Ser His Glu Phe
 130 135 140

tac tcg cgc ctt aat tat gtt gtc gtg gag aag gtc ctc cgc aag atg	480
Tyr Ser Arg Leu Asn Tyr Val Val Val Glu Lys Val Leu Arg Lys Met	
145 150 155 160	
ggc atg ccc gag gag gac gtc aag gtc gcc gtc gaa aag tac aag gcc	528
Gly Met Pro Glu Glu Asp Val Lys Val Ala Val Glu Lys Tyr Lys Ala	
165 170 175	
aac ttc ccc gag tgg cgc ctc gac tcc ttc cct ggc ttc ctc ggc aac	576
Asn Phe Pro Glu Trp Arg Leu Asp Ser Phe Pro Gly Phe Leu Gly Asn	
180 185 190	
gtc acc gcc ggt cgc tgc acc aac acc ttc aac ctc gac ggc atg aac	624
Val Thr Ala Gly Arg Cys Thr Asn Thr Phe Asn Leu Asp Gly Met Asn	
195 200 205	
tgc gtt gtc gac gcc gca tgc gcc tcg tcc ctc atc gcc gtc aag gtc	672
Cys Val Val Asp Ala Ala Cys Ala Ser Ser Leu Ile Ala Val Lys Val	
210 215 220	
gcc atc gac gag ctg ctc tac ggt gac tgc gac atg atg gtc acc ggt	720
Ala Ile Asp Glu Leu Leu Tyr Gly Asp Cys Asp Met Met Val Thr Gly	
225 230 235 240	
gcc acc tgc acg gat aac tcc atc ggc atg tac atg gcc ttc tcc aag	768
Ala Thr Cys Thr Asp Asn Ser Ile Gly Met Tyr Met Ala Phe Ser Lys	
245 250 255	
acc ccc gtg ttc tcc acg gac ccc agc gtg cgc gcc tac gac gaa aag	816
Thr Pro Val Phe Ser Thr Asp Pro Ser Val Arg Ala Tyr Asp Glu Lys	
260 265 270	
aca aag ggc atg ctc atc ggc gag ggc tcc gcc atg ctc gtc ctc aag	864
Thr Lys Gly Met Leu Ile Gly Glu Gly Ser Ala Met Leu Val Leu Lys	
275 280 285	
cgc tac gcc gac gcc gtc cgc gac ggc gat gag atc cac gct gtt att	912
Arg Tyr Ala Asp Ala Val Arg Asp Gly Asp Glu Ile His Ala Val Ile	
290 295 300	
cgc ggc tgc gcc tcc tcc agt gat ggc aag gcc gcc ggc atc tac acg	960
Arg Gly Cys Ala Ser Ser Ser Asp Gly Lys Ala Ala Gly Ile Tyr Thr	
305 310 315 320	

ccc acc att tcg ggc cag gag gag gcc ctc cgc cgc gcc tac aac cgc Pro Thr Ile Ser Gly Gln Glu Glu Ala Leu Arg Arg Ala Tyr Asn Arg 325 330 335	1008
gcc tgt gtc gac ccg gcc acc gtc act ctc gtc gag ggt cac ggc acc Ala Cys Val Asp Pro Ala Thr Val Thr Leu Val Glu Gly His Gly Thr 340 345 350	1056
ggc act ccc gtt ggc gac cgc atc gag ctc acc gcc ttg cgc aac ctc Gly Thr Pro Val Gly Asp Arg Ile Glu Leu Thr Ala Leu Arg Asn Leu 355 360 365	1104
ttt gac aag gcc tac ggc gag ggc aac acc gaa aag gtc gct gtg ggc Phe Asp Lys Ala Tyr Gly Glu Gly Asn Thr Glu Lys Val Ala Val Gly 370 375 380	1152
agc atc aag tcc agc atc ggc cat ctc aag gcc gtc gcc ggt ctc gcc Ser Ile Lys Ser Ser Ile Gly His Leu Lys Ala Val Ala Gly Leu Ala 385 390 395 400	1200
ggc atg atc aag gtc atc atg ggc ctc aag cac aag act ctc ccg ggc Gly Met Ile Lys Val Ile Met Ala Leu Lys His Lys Thr Leu Pro Gly 405 410 415	1248
acc atc aac gtc gac aac cca ccc aac ctc tac gac aac acg ccc atc Thr Ile Asn Val Asp Asn Pro Pro Asn Leu Tyr Asp Asn Thr Pro Ile 420 425 430	1296
aac gag tcc tcg ctc tac att aac acc atg aac cgc ccc tgg ttc ccg Asn Glu Ser Ser Leu Tyr Ile Asn Thr Met Asn Arg Pro Trp Phe Pro 435 440 445	1344
ccc cct ggt gtg ccc cgc cgc gcc ggc att tcg agc ttt ggc ttt ggt Pro Pro Gly Val Pro Arg Arg Ala Gly Ile Ser Ser Phe Gly Phe Gly 450 455 460	1392
ggc gcc aac tac cac gcc gtc ctc gag gag gcc gag ccc gag cac acg Gly Ala Asn Tyr His Ala Val Leu Glu Glu Ala Glu Pro Glu His Thr 465 470 475 480	1440
acc gcg tac cgc ctc aac aag cgc ccg cag ccc gtg ctc atg atg gcc Thr Ala Tyr Arg Leu Asn Lys Arg Pro Gln Pro Val Leu Met Met Ala 485 490 495	1488

gcc acg ccc gcg gcc ctc cag tgc ctc tgc gag gcc cag ctc aag gag Ala Thr Pro Ala Ala Leu Gln Ser Leu Cys Glu Ala Gln Leu Lys Glu 500 505 510	1536
ttc gag gcc gcc atc aag gag aac gag acc gtc aag aac acc gcc tac Phe Glu Ala Ala Ile Lys Glu Asn Glu Thr Val Lys Asn Thr Ala Tyr 515 520 525	1584
atc aag tgc gtc aag ttc ggc gag cag ttc aaa ttc cct ggc tcc atc Ile Lys Cys Val Lys Phe Gly Glu Gln Phe Lys Phe Pro Gly Ser Ile 530 535 540	1632
ccg gcc aca aac gcg cgc ctc ggc ttc ctc gtc aag gat gct gag gat Pro Ala Thr Asn Ala Arg Leu Gly Phe Leu Val Lys Asp Ala Glu Asp 545 550 555 560	1680
gcc tgc tcc acc ctc cgt gcc atc tgc gcc caa ttc gcc aag gat gtc Ala Cys Ser Thr Leu Arg Ala Ile Cys Ala Gln Phe Ala Lys Asp Val 565 570 575	1728
acc aag gag gcc tgg cgc ctc ccc cgc gag ggc gtc agc ttc cgc gcc Thr Lys Glu Ala Trp Arg Leu Pro Arg Glu Gly Val Ser Phe Arg Ala 580 585 590	1776
aag ggc atc gcc acc aac ggc gct gtc gcc gcg ctc ttc tcc ggc cag Lys Gly Ile Ala Thr Asn Gly Ala Val Ala Ala Leu Phe Ser Gly Gln 595 600 605	1824
ggc gcg cag tac acg cac atg ttt agc gag gtg gcc atg aac tgg ccc Gly Ala Gln Tyr Thr His Met Phe Ser Glu Val Ala Met Asn Trp Pro 610 615 620	1872
cag ttc cgc cag agc att gcc gcc atg gac gcc gcc cag tcc aag gtc Gln Phe Arg Gln Ser Ile Ala Ala Met Asp Ala Ala Gln Ser Lys Val 625 630 635 640	1920
gct gga agc gac aag gac ttt gag cgc gtc tcc cag gtc ctc tac ccg Ala Gly Ser Asp Lys Asp Phe Glu Arg Val Ser Gln Val Leu Tyr Pro 645 650 655	1968
cgc aag ccg tac gag cgt gag ccc gag cag aac ccc aag aag atc tcc Arg Lys Pro Tyr Glu Arg Glu Pro Glu Gln Asn Pro Lys Lys Ile Ser 660 665 670	2016

ctc acc gcc tac tcg cag ccc tcg acc ctg gcc tgc gct ctc ggt gcc Leu Thr Ala Tyr Ser Gln Pro Ser Thr Leu Ala Cys Ala Leu Gly Ala 675 680 685	2064
ttt gag atc ttc aag gag gcc ggc ttc acc ccg gac ttt gcc gcc ggc Phe Glu Ile Phe Lys Glu Ala Gly Phe Thr Pro Asp Phe Ala Ala Gly 690 695 700	2112
cat tcg ctc ggt gag ttc gcc gcc ctc tac gcc gcg ggc tgc gtc gac His Ser Leu Gly Glu Phe Ala Ala Leu Tyr Ala Ala Gly Cys Val Asp 705 710 715 720	2160
cgc gac gag ctc ttt gag ctt gtc tgc cgc cgc gcc cgc atc atg ggc Arg Asp Glu Leu Phe Glu Leu Val Cys Arg Arg Ala Arg Ile Met Gly 725 730 735	2208
ggc aag gac gca ccg gcc acc ccc aag gga tgc atg gcc gcc gtc att Gly Lys Asp Ala Pro Ala Thr Pro Lys Gly Cys Met Ala Ala Val Ile 740 745 750	2256
ggc ccc aac gcc gag aac atc aag gtc cag gcc gcc aac gtc tgg ctc Gly Pro Asn Ala Glu Asn Ile Lys Val Gln Ala Ala Asn Val Trp Leu 755 760 765	2304
ggc aac tcc aac tcg cct tcg cag acc gtc atc acc ggc tcc gtc gaa Gly Asn Ser Asn Ser Pro Ser Gln Thr Val Ile Thr Gly Ser Val Glu 770 775 780	2352
ggt atc cag gcc gag agc gcc cgc ctc cag aag gag ggc ttc cgc gtc Gly Ile Gln Ala Glu Ser Ala Arg Leu Gln Lys Glu Gly Phe Arg Val 785 790 795 800	2400
gtg cct ctt gcc tgc gag agc gcc ttc cac tcg ccc cag atg gag aac Val Pro Leu Ala Cys Glu Ser Ala Phe His Ser Pro Gln Met Glu Asn 805 810 815	2448
gcc tcg tcg gcc ttc aag gac gtc atc tcc aag gtc tcc ttc cgc acc Ala Ser Ser Ala Phe Lys Asp Val Ile Ser Lys Val Ser Phe Arg Thr 820 825 830	2496
ccc aag gcc gag acc aag ctc ttc agc aac gtc tct ggc gag acc tac Pro Lys Ala Glu Thr Lys Leu Phe Ser Asn Val Ser Gly Glu Thr Tyr 835 840 845	2544

ccc acg gac gcc cgc gag atg ctt acg cag cac atg acc agc agc gtc Pro Thr Asp Ala Arg Glu Met Leu Thr Gln His Met Thr Ser Ser Val 850 855 860	2592
aag ttc ctc acc cag gtc cgc aac atg cac cag gcc ggt gcg cgc atc Lys Phe Leu Thr Gln Val Arg Asn Met His Gln Ala Gly Ala Arg Ile 865 870 875 880	2640
ttt gtc gag ttc gga ccc aag cag gtg ctc tcc aag ctt gtc tcc gag Phe Val Glu Phe Gly Pro Lys Gln Val Leu Ser Lys Leu Val Ser Glu 885 890 895	2688
acc ctc aag gat gac ccc tcg gtt gtc acc gtc tct gtc aac ccg gcc Thr Leu Lys Asp Asp Pro Ser Val Val Thr Val Ser Val Asn Pro Ala 900 905 910	2736
tcg ggc acg gat tcg gac atc cag ctc cgc gac gcg gcc gtc cag ctc Ser Gly Thr Asp Ser Asp Ile Gln Leu Arg Asp Ala Ala Val Gln Leu 915 920 925	2784
gtt gtc gct ggc gtc aac ctt cag ggc ttt gac aag tgg gac gcc ccc Val Val Ala Gly Val Asn Leu Gln Gly Phe Asp Lys Trp Asp Ala Pro 930 935 940	2832
gat gcc acc cgc atg cag gcc atc aag aag aag cgc act acc ctc cgc Asp Ala Thr Arg Met Gln Ala Ile Lys Lys Lys Arg Thr Thr Leu Arg 945 950 955 960	2880
ctt tcg gcc gcc acc tac gtc tcg gac aag acc aag aag gtc cgc gac Leu Ser Ala Ala Thr Tyr Val Ser Asp Lys Thr Lys Lys Val Arg Asp 965 970 975	2928
gcc gcc atg aac gat ggc cgc tgc gtc acc tac ctc aag ggc gcc gca Ala Ala Met Asn Asp Gly Arg Cys Val Thr Tyr Leu Lys Gly Ala Ala 980 985 990	2976
ccg ctc atc aag gcc ccg gag ccc gtt gtc gac gag gcc gcc aag cgc Pro Leu Ile Lys Ala Pro Glu Pro Val Val Asp Glu Ala Ala Lys Arg 995 1000 1005	3024
gag gcc gag cgt ctc cag aag gag ctt cag gat gcc cag cgc cag Glu Ala Glu Arg Leu Gln Lys Glu Leu Gln Asp Ala Gln Arg Gln 1010 1015 1020	3069

ctc gac gac gcc aag cgc gcc gcc gcc gag gcc aac tcc aag ctc Leu Asp Asp Ala Lys Arg Ala Ala Ala Glu Ala Asn Ser Lys Leu 1025 1030 1035	3114
gcc gct gcc aag gag gag gcc aag acc gcc gct gct tcg gcc aag Ala Ala Ala Lys Glu Glu Ala Lys Thr Ala Ala Ala Ser Ala Lys 1040 1045 1050	3159
ccc gca gtt gac act gct gtt gtc gaa aag cat cgt gcc atc ctc Pro Ala Val Asp Thr Ala Val Val Glu Lys His Arg Ala Ile Leu 1055 1060 1065	3204
aag tcc atg ctc gcg gag ctc gat ggc tac gga tcg gtc gac gct Lys Ser Met Leu Ala Glu Leu Asp Gly Tyr Gly Ser Val Asp Ala 1070 1075 1080	3249
tct tcc ctc cag cag cag cag cag cag cag acg gcc ccc gcc ccg Ser Ser Leu Gln Gln Gln Gln Gln Gln Gln Thr Ala Pro Ala Pro 1085 1090 1095	3294
gtc aag gct gct gcg cct gcc gcc ccc gtt gcc tcg gcc cct gcc Val Lys Ala Ala Ala Pro Ala Ala Pro Val Ala Ser Ala Pro Ala 1100 1105 1110	3339
ccg gct gtc tcg aac gag ctt ctt gag aag gcc gag act gtc gtc Pro Ala Val Ser Asn Glu Leu Leu Glu Lys Ala Glu Thr Val Val 1115 1120 1125	3384
atg gag gtc ctc gcc gcc aag acc ggc tac gag acc gac atg atc Met Glu Val Leu Ala Ala Lys Thr Gly Tyr Glu Thr Asp Met Ile 1130 1135 1140	3429
gag gct gac atg gag ctc gag acc gag ctc ggc att gac tcc atc Glu Ala Asp Met Glu Leu Glu Thr Glu Leu Gly Ile Asp Ser Ile 1145 1150 1155	3474
aag cgt gtc gag atc ctc tcc gag gtc cag gcc atg ctc aat gtc Lys Arg Val Glu Ile Leu Ser Glu Val Gln Ala Met Leu Asn Val 1160 1165 1170	3519
gag gcc aag gat gtc gat gcc ctc agc cgc act cgc act gtt ggt Glu Ala Lys Asp Val Asp Ala Leu Ser Arg Thr Arg Thr Val Gly 1175 1180 1185	3564

gag gtt gtc aac gcc atg aag gcc gag atc gct ggc agc tct gcc Glu Val Val Asn Ala Met Lys Ala Glu Ile Ala Gly Ser Ser Ala 1190 1195 1200	3609
ccg gcg cct gct gcc gct gct ccg gct ccg gcc aag gct gcc cct Pro Ala Pro Ala Ala Ala Ala Pro Ala Pro Ala Lys Ala Ala Pro 1205 1210 1215	3654
gcc gcc gct gcg cct gct gtc tcg aac gag ctt ctc gag aag gcc Ala Ala Ala Ala Pro Ala Val Ser Asn Glu Leu Leu Glu Lys Ala 1220 1225 1230	3699
gag acc gtc gtc atg gag gtc ctc gcc gcc aag act ggc tac gag Glu Thr Val Val Met Glu Val Leu Ala Ala Lys Thr Gly Tyr Glu 1235 1240 1245	3744
act gac atg atc gag tcc gac atg gag ctc gag act gag ctc ggc Thr Asp Met Ile Glu Ser Asp Met Glu Leu Glu Thr Glu Leu Gly 1250 1255 1260	3789
att gac tcc atc aag cgt gtc gag atc ctc tcc gag gtt cag gcc Ile Asp Ser Ile Lys Arg Val Glu Ile Leu Ser Glu Val Gln Ala 1265 1270 1275	3834
atg ctc aac gtc gag gcc aag gac gtc gac gct ctc agc cgc act Met Leu Asn Val Glu Ala Lys Asp Val Asp Ala Leu Ser Arg Thr 1280 1285 1290	3879
cgc act gtg ggt gag gtc gtc aac gcc atg aag gct gag atc gct Arg Thr Val Gly Glu Val Val Asn Ala Met Lys Ala Glu Ile Ala 1295 1300 1305	3924
ggt ggc tct gcc ccg gcg cct gcc gcc gct gcc cca ggt ccg gct Gly Gly Ser Ala Pro Ala Pro Ala Ala Ala Ala Pro Gly Pro Ala 1310 1315 1320	3969
gct gcc gcc cct gcg cct gcc gcc gcc gcc cct gct gtc tcg aac Ala Ala Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Val Ser Asn 1325 1330 1335	4014
gag ctt ctt gag aag gcc gag acc gtc gtc atg gag gtc ctc gcc Glu Leu Leu Glu Lys Ala Glu Thr Val Val Met Glu Val Leu Ala 1340 1345 1350	4059

gcc aag act ggc tac gag act gac atg atc gag tcc gac atg gag Ala Lys Thr Gly Tyr Glu Thr Asp Met Ile Glu Ser Asp Met Glu 1355 1360 1365	4104
ctc gag acc gag ctc ggc att gac tcc atc aag cgt gtc gag att Leu Glu Thr Glu Leu Gly Ile Asp Ser Ile Lys Arg Val Glu Ile 1370 1375 1380	4149
ctc tcc gag gtc cag gcc atg ctc aac gtc gag gcc aag gac gtc Leu Ser Glu Val Gln Ala Met Leu Asn Val Glu Ala Lys Asp Val 1385 1390 1395	4194
gac gct ctc agc cgc acc cgc act gtt ggc gag gtc gtc gat gcc Asp Ala Leu Ser Arg Thr Arg Thr Val Gly Glu Val Val Asp Ala 1400 1405 1410	4239
atg aag gcc gag atc gct ggt ggc tct gcc ccg gcg cct gcc gcc Met Lys Ala Glu Ile Ala Gly Gly Ser Ala Pro Ala Pro Ala Ala 1415 1420 1425	4284
gct gct cct gct ccg gct gct gcc gcc cct gcg cct gcc gcc cct Ala Ala Pro Ala Pro Ala Ala Ala Pro Ala Pro Ala Ala Pro 1430 1435 1440	4329
gcg cct gct gtc tcg agc gag ctt ctc gag aag gcc gag act gtc Ala Pro Ala Val Ser Ser Glu Leu Leu Glu Lys Ala Glu Thr Val 1445 1450 1455	4374
gtc atg gag gtc ctc gcc gcc aag act ggc tac gag act gac atg Val Met Glu Val Leu Ala Ala Lys Thr Gly Tyr Glu Thr Asp Met 1460 1465 1470	4419
atc gag tcc gac atg gag ctc gag acc gag ctc ggc att gac tcc Ile Glu Ser Asp Met Glu Leu Glu Thr Glu Leu Gly Ile Asp Ser 1475 1480 1485	4464
atc aag cgt gtc gag att ctc tcc gag gtc cag gcc atg ctc aac Ile Lys Arg Val Glu Ile Leu Ser Glu Val Gln Ala Met Leu Asn 1490 1495 1500	4509
gtc gag gcc aag gac gtc gac gct ctc agc cgc acc cgc act gtt Val Glu Ala Lys Asp Val Asp Ala Leu Ser Arg Thr Arg Thr Val 1505 1510 1515	4554

ggc gag gtc gtc gat gcc atg aag gcc gag atc gct ggt ggc tct Gly Glu Val Val Asp Ala Met Lys Ala Glu Ile Ala Gly Gly Ser 1520 1525 1530	4599
gcc ccg gcg cct gcc gcc gct gct cct gct ccg gct gct gcc gcc Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Pro Ala Ala Ala Ala 1535 1540 1545	4644
cct gcg cct gcc gcc cct gcg cct gcc gcc cct gcg cct gct gtc Pro Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Pro Ala Val 1550 1555 1560	4689
tcg agc gag ctt ctc gag aag gcc gag act gtc gtc atg gag gtc Ser Ser Glu Leu Leu Glu Lys Ala Glu Thr Val Val Met Glu Val 1565 1570 1575	4734
ctc gcc gcc aag act gcc tac gag act gac atg att gag tcc gac Leu Ala Ala Lys Thr Gly Tyr Glu Thr Asp Met Ile Glu Ser Asp 1580 1585 1590	4779
atg gag ctc gag acc gag ctc gcc att gac tcc atc aag cgt gtc Met Glu Leu Glu Thr Glu Leu Gly Ile Asp Ser Ile Lys Arg Val 1595 1600 1605	4824
gag att ctc tcc gag gtt cag gcc atg ctc aac gtc gag gcc aag Glu Ile Leu Ser Glu Val Gln Ala Met Leu Asn Val Glu Ala Lys 1610 1615 1620	4869
gac gtc gac gct ctc agc cgc act cgc act gtt ggt gag gtc gtc Asp Val Asp Ala Leu Ser Arg Thr Arg Thr Val Gly Glu Val Val 1625 1630 1635	4914
gat gcc atg aag gct gag atc gct gcc agc tcc gcc tcg gcg cct Asp Ala Met Lys Ala Glu Ile Ala Gly Ser Ser Ala Ser Ala Pro 1640 1645 1650	4959
gcc gcc gct gct cct gct ccg gct gct gcc gct cct gcg ccc gct Ala Ala Ala Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Pro Ala 1655 1660 1665	5004
gcc gcc gcc cct gct gtc tcg aac gag ctt ctc gag aaa gcc gag Ala Ala Ala Pro Ala Val Ser Asn Glu Leu Leu Glu Lys Ala Glu 1670 1675 1680	5049

act gtc gtc atg gag gtc ctc gcc gcc aag act ggc tac gag act Thr Val Val Met Glu Val Leu Ala Ala Lys Thr Gly Tyr Glu Thr 1685 1690 1695	5094
gac atg atc gag tcc gac atg gag ctc gag act gag ctc ggc att Asp Met Ile Glu Ser Asp Met Glu Leu Glu Thr Glu Leu Gly Ile 1700 1705 1710	5139
gac tcc atc aag cgt gtc gag atc ctc tcc gag gtt cag gcc atg Asp Ser Ile Lys Arg Val Glu Ile Leu Ser Glu Val Gln Ala Met 1715 1720 1725	5184
ctc aac gtc gag gcc aag gac gtc gat gcc ctc agc cgc acc cgc Leu Asn Val Glu Ala Lys Asp Val Asp Ala Leu Ser Arg Thr Arg 1730 1735 1740	5229
act gtt ggc gag gtt gtc gat gcc atg aag gcc gag atc gct ggt Thr Val Gly Glu Val Val Asp Ala Met Lys Ala Glu Ile Ala Gly 1745 1750 1755	5274
ggc tct gcc ccg gcg cct gcc gcc gct gcc cct gct ccg gct gcc Gly Ser Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Pro Ala Ala 1760 1765 1770	5319
gcc gcc cct gct gtc tcg aac gag ctt ctc gag aag gcc gag act Ala Ala Pro Ala Val Ser Asn Glu Leu Leu Glu Lys Ala Glu Thr 1775 1780 1785	5364
gtc gtc atg gag gtc ctc gcc gcc aag act ggc tac gag acc gac Val Val Met Glu Val Leu Ala Ala Lys Thr Gly Tyr Glu Thr Asp 1790 1795 1800	5409
atg atc gag tcc gac atg gag ctc gag acc gag ctc ggc att gac Met Ile Glu Ser Asp Met Glu Leu Glu Thr Glu Leu Gly Ile Asp 1805 1810 1815	5454
tcc atc aag cgt gtc gag att ctc tcc gag gtt cag gcc atg ctc Ser Ile Lys Arg Val Glu Ile Leu Ser Glu Val Gln Ala Met Leu 1820 1825 1830	5499
aac gtc gag gcc aag gac gtc gat gct ctc agc cgc act cgc act Asn Val Glu Ala Lys Asp Val Asp Ala Leu Ser Arg Thr Arg Thr 1835 1840 1845	5544

gtt ggc gag gtc gtc gat gcc atg aag gct gag atc gcc ggc agc Val Gly Glu Val Val Asp Ala Met Lys Ala Glu Ile Ala Gly Ser 1850 1855 1860	5589
tcc gcc ccg gcg cct gcc gcc gct gct cct gct ccg gct gct gcc Ser Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Pro Ala Ala Ala 1865 1870 1875	5634
gct cct gcg ccc gct gcc gct gcc cct gct gtc tcg agc gag ctt Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Val Ser Ser Glu Leu 1880 1885 1890	5679
ctc gag aag gcc gag acc gtc gtc atg gag gtc ctc gcc gcc aag Leu Glu Lys Ala Glu Thr Val Val Met Glu Val Leu Ala Ala Lys 1895 1900 1905	5724
act ggc tac gag act gac atg att gag tcc gac atg gag ctc gag Thr Gly Tyr Glu Thr Asp Met Ile Glu Ser Asp Met Glu Leu Glu 1910 1915 1920	5769
act gag ctc gcc att gac tcc atc aag cgt gtc gag atc ctc tcc Thr Glu Leu Gly Ile Asp Ser Ile Lys Arg Val Glu Ile Leu Ser 1925 1930 1935	5814
gag gtt cag gcc atg ctc aac gtc gag gcc aag gac gtc gat gcc Glu Val Gln Ala Met Leu Asn Val Glu Ala Lys Asp Val Asp Ala 1940 1945 1950	5859
ctc agc cgc acc cgc act gtt gcc gag gtt gtc gat gcc atg aag Leu Ser Arg Thr Arg Thr Val Gly Glu Val Val Asp Ala Met Lys 1955 1960 1965	5904
gcc gag atc gct ggt gcc tct gcc ccg gcg cct gcc gcc gct gcc Ala Glu Ile Ala Gly Gly Ser Ala Pro Ala Pro Ala Ala Ala Ala 1970 1975 1980	5949
cct gct ccg gct gcc gcc gcc cct gct gtc tcg aac gag ctt ctt Pro Ala Pro Ala Ala Ala Ala Pro Ala Val Ser Asn Glu Leu Leu 1985 1990 1995	5994
gag aag gcc gag acc gtc gtc atg gag gtc ctc gcc gcc aag act Glu Lys Ala Glu Thr Val Val Met Glu Val Leu Ala Ala Lys Thr 2000 2005 2010	6039

ggc tac gag acc gac atg atc gag tcc gac atg gag ctc gag acc Gly Tyr Glu Thr Asp Met Ile Glu Ser Asp Met Glu Leu Glu Thr 2015 2020 2025	6084
gag ctc ggc att gac tcc atc aag cgt gtc gag att ctc tcc gag Glu Leu Gly Ile Asp Ser Ile Lys Arg Val Glu Ile Leu Ser Glu 2030 2035 2040	6129
gtt cag gcc atg ctc aac gtc gag gcc aag gac gtc gac gct ctc Val Gln Ala Met Leu Asn Val Glu Ala Lys Asp Val Asp Ala Leu 2045 2050 2055	6174
agc cgc act cgc act gtt ggc gag gtc gtc gat gcc atg aag gct Ser Arg Thr Arg Thr Val Gly Glu Val Val Asp Ala Met Lys Ala 2060 2065 2070	6219
gag atc gct ggt ggc tct gcc ccg gcg cct gcc gcc gct gct cct Glu Ile Ala Gly Gly Ser Ala Pro Ala Pro Ala Ala Ala Ala Pro 2075 2080 2085	6264
gcc tcg gct ggc gcc gcg cct gcg gtc aag att gac tcg gtc cac Ala Ser Ala Gly Ala Ala Pro Ala Val Lys Ile Asp Ser Val His 2090 2095 2100	6309
ggc gct gac tgt gat gat ctt tcc ctg atg cac gcc aag gtg gtt Gly Ala Asp Cys Asp Asp Leu Ser Leu Met His Ala Lys Val Val 2105 2110 2115	6354
gac atc cgc cgc ccg gac gag ctc atc ctg gag cgc ccc gag aac Asp Ile Arg Arg Pro Asp Glu Leu Ile Leu Glu Arg Pro Glu Asn 2120 2125 2130	6399
cgc ccc gtt ctc gtt gtc gat gac ggc agc gag ctc acc ctc gcc Arg Pro Val Leu Val Val Asp Asp Gly Ser Glu Leu Thr Leu Ala 2135 2140 2145	6444
ctg gtc cgc gtc ctc ggc gcc tgc gcc gtt gtc ctg acc ttt gag Leu Val Arg Val Leu Gly Ala Cys Ala Val Val Leu Thr Phe Glu 2150 2155 2160	6489
ggt ctc cag ctc gct cag cgc gct ggt gcc gct gcc atc cgc cac Gly Leu Gln Leu Ala Gln Arg Ala Gly Ala Ala Ala Ile Arg His 2165 2170 2175	6534

gtg ctc gcc aag gat ctt tcc gcg gag agc gcc gag aag gcc atc Val Leu Ala Lys Asp Leu Ser Ala Glu Ser Ala Glu Lys Ala Ile 2180 2185 2190	6579
aag gag gcc gag cag cgc ttt ggc gct ctc ggc ggc ttc atc tcg Lys Glu Ala Glu Gln Arg Phe Gly Ala Leu Gly Gly Phe Ile Ser 2195 2200 2205	6624
cag cag gcg gag cgc ttc gag ccc gcc gaa atc ctc ggc ttc acg Gln Gln Ala Glu Arg Phe Glu Pro Ala Glu Ile Leu Gly Phe Thr 2210 2215 2220	6669
ctc atg tgc gcc aag ttc gcc aag gct tcc ctc tgc acg gct gtg Leu Met Cys Ala Lys Phe Ala Lys Ala Ser Leu Cys Thr Ala Val 2225 2230 2235	6714
gct ggc ggc cgc ccg gcc ttt atc ggt gtg gcg cgc ctt gac ggc Ala Gly Gly Arg Pro Ala Phe Ile Gly Val Ala Arg Leu Asp Gly 2240 2245 2250	6759
cgc ctc gga ttc act tcg cag ggc act tct gac gcg ctc aag cgt Arg Leu Gly Phe Thr Ser Gln Gly Thr Ser Asp Ala Leu Lys Arg 2255 2260 2265	6804
gcc cag cgt ggt gcc atc ttt ggc ctc tgc aag acc atc ggc ctc Ala Gln Arg Gly Ala Ile Phe Gly Leu Cys Lys Thr Ile Gly Leu 2270 2275 2280	6849
gag tgg tcc gag tct gac gtc ttt tcc cgc ggc gtg gac att gct Glu Trp Ser Glu Ser Asp Val Phe Ser Arg Gly Val Asp Ile Ala 2285 2290 2295	6894
cag ggc atg cac ccc gag gat gcc gcc gtg gcg att gtg cgc gag Gln Gly Met His Pro Glu Asp Ala Ala Val Ala Ile Val Arg Glu 2300 2305 2310	6939
atg gcg tgc gct gac att cgc att cgc gag gtc ggc att ggc gca Met Ala Cys Ala Asp Ile Arg Ile Arg Glu Val Gly Ile Gly Ala 2315 2320 2325	6984
aac cag cag cgc tgc acg atc cgt gcc gcc aag ctc gag acc ggc Asn Gln Gln Arg Cys Thr Ile Arg Ala Ala Lys Leu Glu Thr Gly 2330 2335 2340	7029

aac ccg cag cgc cag atc gcc aag gac gac gtg ctg ctc gtt tct Asn Pro Gln Arg Gln Ile Ala Lys Asp Asp Val Leu Leu Val Ser 2345 2350 2355	7074
ggc ggc gct cgc ggc atc acg cct ctt tgc atc cgg gag atc acg Gly Gly Ala Arg Gly Ile Thr Pro Leu Cys Ile Arg Glu Ile Thr 2360 2365 2370	7119
cgc cag atc gcg ggc ggc aag tac att ctg ctt ggc cgc agc aag Arg Gln Ile Ala Gly Gly Lys Tyr Ile Leu Leu Gly Arg Ser Lys 2375 2380 2385	7164
gtc tct gcg agc gaa ccg gca tgg tgc gct ggc atc act gac gag Val Ser Ala Ser Glu Pro Ala Trp Cys Ala Gly Ile Thr Asp Glu 2390 2395 2400	7209
aag gct gtg caa aag gct gct acc cag gag ctc aag cgc gcc ttt Lys Ala Val Gln Lys Ala Ala Thr Gln Glu Leu Lys Arg Ala Phe 2405 2410 2415	7254
agc gct ggc gag ggc ccc aag ccc acg ccc cgc gct gtc act aag Ser Ala Gly Glu Gly Pro Lys Pro Thr Pro Arg Ala Val Thr Lys 2420 2425 2430	7299
ctt gtg ggc tct gtt ctt ggc gct cgc gag gtg cgc agc tct att Leu Val Gly Ser Val Leu Gly Ala Arg Glu Val Arg Ser Ser Ile 2435 2440 2445	7344
gct gcg att gaa gcg ctc ggc ggc aag gcc atc tac tcg tcg tgc Ala Ala Ile Glu Ala Leu Gly Gly Lys Ala Ile Tyr Ser Ser Cys 2450 2455 2460	7389
gac gtg aac tct gcc gcc gac gtg gcc aag gcc gtg cgc gat gcc Asp Val Asn Ser Ala Ala Asp Val Ala Lys Ala Val Arg Asp Ala 2465 2470 2475	7434
gag tcc cag ctc ggt gcc cgc gtc tcg ggc atc gtt cat gcc tcg Glu Ser Gln Leu Gly Ala Arg Val Ser Gly Ile Val His Ala Ser 2480 2485 2490	7479
ggc gtg ctc cgc gac cgt ctc atc gag aag aag ctc ccc gac gag Gly Val Leu Arg Asp Arg Leu Ile Glu Lys Lys Leu Pro Asp Glu 2495 2500 2505	7524

ttc gac gcc gtc ttt ggc acc aag gtc acc ggt ctc gag aac ctc Phe Asp Ala Val Phe Gly Thr Lys Val Thr Gly Leu Glu Asn Leu 2510 2515 2520	7569
ctc gcc gcc gtc gac cgc gcc aac ctc aag cac atg gtc ctc ttc Leu Ala Ala Val Asp Arg Ala Asn Leu Lys His Met Val Leu Phe 2525 2530 2535	7614
agc tcg ctc gcc ggc ttc cac ggc aac gtc ggc cag tct gac tac Ser Ser Leu Ala Gly Phe His Gly Asn Val Gly Gln Ser Asp Tyr 2540 2545 2550	7659
gcc atg gcc aac gag gcc ctt aac aag atg ggc ctc gag ctc gcc Ala Met Ala Asn Glu Ala Leu Asn Lys Met Gly Leu Glu Leu Ala 2555 2560 2565	7704
aag gac gtc tcg gtc aag tcg atc tgc ttc ggt ccc tgg gac ggt Lys Asp Val Ser Val Lys Ser Ile Cys Phe Gly Pro Trp Asp Gly 2570 2575 2580	7749
ggc atg gtg acg ccg cag ctc aag aag cag ttc cag gag atg ggc Gly Met Val Thr Pro Gln Leu Lys Lys Gln Phe Gln Glu Met Gly 2585 2590 2595	7794
gtg cag atc atc ccc cgc gag ggc ggc gct gat acc gtg gcg cgc Val Gln Ile Ile Pro Arg Glu Gly Gly Ala Asp Thr Val Ala Arg 2600 2605 2610	7839
atc gtg ctc ggc tcc tcg ccg gct gag atc ctt gtc ggc aac tgg Ile Val Leu Gly Ser Ser Pro Ala Glu Ile Leu Val Gly Asn Trp 2615 2620 2625	7884
cgc acc ccg tcc aag aag gtc ggc tcg gac acc atc acc ctg cac Arg Thr Pro Ser Lys Lys Val Gly Ser Asp Thr Ile Thr Leu His 2630 2635 2640	7929
cgc aag att tcc gcc aag tcc aac ccc ttc ctc gag gac cac gtc Arg Lys Ile Ser Ala Lys Ser Asn Pro Phe Leu Glu Asp His Val 2645 2650 2655	7974
atc cag ggc cgc cgc gtg ctg ccc atg acg ctg gcc att ggc tcg Ile Gln Gly Arg Arg Val Leu Pro Met Thr Leu Ala Ile Gly Ser 2660 2665 2670	8019

ctc gcg gag acc tgc ctc ggc	ctc ttc ccc ggc tac	tcg ctc tgg	8064
Leu Ala Glu Thr Cys Leu Gly	Leu Phe Pro Gly Tyr	Ser Leu Trp	
2675	2680	2685	
gcc att gac gac gcc cag ctc	ttc aag ggt gtc act	gtc gac ggc	8109
Ala Ile Asp Asp Ala Gln Leu	Phe Lys Gly Val Thr	Val Asp Gly	
2690	2695	2700	
gac gtc aac tgc gag gtg acc	ctc acc ccg tcg acg	gcg ccc tcg	8154
Asp Val Asn Cys Glu Val Thr	Leu Thr Pro Ser Thr	Ala Pro Ser	
2705	2710	2715	
ggc cgc gtc aac gtc cag gcc	acg ctc aag acc ttt	tcc agc ggc	8199
Gly Arg Val Asn Val Gln Ala	Thr Leu Lys Thr Phe	Ser Ser Gly	
2720	2725	2730	
aag ctg gtc ccg gcc tac cgc	gcc gtc atc gtg ctc	tcc aac cag	8244
Lys Leu Val Pro Ala Tyr Arg	Ala Val Ile Val Leu	Ser Asn Gln	
2735	2740	2745	
ggc gcg ccc ccg gcc aac gcc	acc atg cag ccg ccc	tcg ctc gat	8289
Gly Ala Pro Pro Ala Asn Ala	Thr Met Gln Pro Pro	Ser Leu Asp	
2750	2755	2760	
gcc gat ccg gcg ctc cag gcc	tcc gtc tac gac ggc	aag acc ctc	8334
Ala Asp Pro Ala Leu Gln Gly	Ser Val Tyr Asp Gly	Lys Thr Leu	
2765	2770	2775	
ttc cac ggc ccg gcc ttc cgc	ggc atc gat gac gtg	ctc tcg tgc	8379
Phe His Gly Pro Ala Phe Arg	Gly Ile Asp Asp Val	Leu Ser Cys	
2780	2785	2790	
acc aag agc cag ctt gtg gcc	aag tgc agc gct gtc	ccc ggc tcc	8424
Thr Lys Ser Gln Leu Val Ala	Lys Cys Ser Ala Val	Pro Gly Ser	
2795	2800	2805	
gac gcc gct cgc ggc gag ttt	gcc acg gac act gac	gcc cat gac	8469
Asp Ala Ala Arg Gly Glu Phe	Ala Thr Asp Thr Asp	Ala His Asp	
2810	2815	2820	
ccc ttc gtg aac gac ctg gcc	ttt cag gcc atg ctc	gtc tgg gtg	8514
Pro Phe Val Asn Asp Leu Ala	Phe Gln Ala Met Leu	Val Trp Val	
2825	2830	2835	

cgc cgc acg ctc ggc cag gct gcg ctc ccc aac tcg atc cag cgc 8559
 Arg Arg Thr Leu Gly Gln Ala Ala Leu Pro Asn Ser Ile Gln Arg
 2840 2845 2850

atc gtc cag cac cgc ccg gtc ccg cag gac aag ccc ttc tac att 8604
 Ile Val Gln His Arg Pro Val Pro Gln Asp Lys Pro Phe Tyr Ile
 2855 2860 2865

acc ctc cgc tcc aac cag tcg ggc ggt cac tcc cag cac aag cac 8649
 Thr Leu Arg Ser Asn Gln Ser Gly Gly His Ser Gln His Lys His
 2870 2875 2880

gcc ctt cag ttc cac aac gag cag ggc gat ctc ttc att gat gtc 8694
 Ala Leu Gln Phe His Asn Glu Gln Gly Asp Leu Phe Ile Asp Val
 2885 2890 2895

cag gct tcg gtc atc gcc acg gac agc ctt gcc ttc 8730
 Gln Ala Ser Val Ile Ala Thr Asp Ser Leu Ala Phe
 2900 2905 2910

<210> 2
 <211> 2910
 <212> PRT
 <213> Schizochytrium sp.

<400> 2

Met Ala Ala Arg Leu Gln Glu Gln Lys Gly Gly Glu Met Asp Thr Arg
 1 5 10 15

Ile Ala Ile Ile Gly Met Ser Ala Ile Leu Pro Cys Gly Thr Thr Val
 20 25 30

Arg Glu Ser Trp Glu Thr Ile Arg Ala Gly Ile Asp Cys Leu Ser Asp
 35 40 45

Leu Pro Glu Asp Arg Val Asp Val Thr Ala Tyr Phe Asp Pro Val Lys
 50 55 60

Thr Thr Lys Asp Lys Ile Tyr Cys Lys Arg Gly Gly Phe Ile Pro Glu
65 70 75 80

Tyr Asp Phe Asp Ala Arg Glu Phe Gly Leu Asn Met Phe Gln Met Glu
85 90 95

Asp Ser Asp Ala Asn Gln Thr Ile Ser Leu Leu Lys Val Lys Glu Ala
100 105 110

Leu Gln Asp Ala Gly Ile Asp Ala Leu Gly Lys Glu Lys Lys Asn Ile
115 120 125

Gly Cys Val Leu Gly Ile Gly Gly Gly Gln Lys Ser Ser His Glu Phe
130 135 140

Tyr Ser Arg Leu Asn Tyr Val Val Val Glu Lys Val Leu Arg Lys Met
145 150 155 160

Gly Met Pro Glu Glu Asp Val Lys Val Ala Val Glu Lys Tyr Lys Ala
165 170 175

Asn Phe Pro Glu Trp Arg Leu Asp Ser Phe Pro Gly Phe Leu Gly Asn
180 185 190

Val Thr Ala Gly Arg Cys Thr Asn Thr Phe Asn Leu Asp Gly Met Asn
195 200 205

Cys Val Val Asp Ala Ala Cys Ala Ser Ser Leu Ile Ala Val Lys Val
210 215 220

Ala Ile Asp Glu Leu Leu Tyr Gly Asp Cys Asp Met Met Val Thr Gly
225 230 235 240

Ala Thr Cys Thr Asp Asn Ser Ile Gly Met Tyr Met Ala Phe Ser Lys
245 250 255

Thr Pro Val Phe Ser Thr Asp Pro Ser Val Arg Ala Tyr Asp Glu Lys

260 265 270

Thr Lys Gly Met Leu Ile Gly Glu Gly Ser Ala Met Leu Val Leu Lys
275 280 285

Arg Tyr Ala Asp Ala Val Arg Asp Gly Asp Glu Ile His Ala Val Ile
290 295 300

Arg Gly Cys Ala Ser Ser Ser Asp Gly Lys Ala Ala Gly Ile Tyr Thr
305 310 315 320

Pro Thr Ile Ser Gly Gln Glu Glu Ala Leu Arg Arg Ala Tyr Asn Arg
325 330 335

Ala Cys Val Asp Pro Ala Thr Val Thr Leu Val Glu Gly His Gly Thr
340 345 350

Gly Thr Pro Val Gly Asp Arg Ile Glu Leu Thr Ala Leu Arg Asn Leu
355 360 365

Phe Asp Lys Ala Tyr Gly Glu Gly Asn Thr Glu Lys Val Ala Val Gly
370 375 380

Ser Ile Lys Ser Ser Ile Gly His Leu Lys Ala Val Ala Gly Leu Ala
385 390 395 400

Gly Met Ile Lys Val Ile Met Ala Leu Lys His Lys Thr Leu Pro Gly
405 410 415

Thr Ile Asn Val Asp Asn Pro Pro Asn Leu Tyr Asp Asn Thr Pro Ile
420 425 430

Asn Glu Ser Ser Leu Tyr Ile Asn Thr Met Asn Arg Pro Trp Phe Pro
435 440 445

Pro Pro Gly Val Pro Arg Arg Ala Gly Ile Ser Ser Phe Gly Phe Gly

450

455

460

Gly Ala Asn Tyr His Ala Val Leu Glu Glu Ala Glu Pro Glu His Thr
465 470 475 480

Thr Ala Tyr Arg Leu Asn Lys Arg Pro Gln Pro Val Leu Met Met Ala
485 490 495

Ala Thr Pro Ala Ala Leu Gln Ser Leu Cys Glu Ala Gln Leu Lys Glu
500 505 510

Phe Glu Ala Ala Ile Lys Glu Asn Glu Thr Val Lys Asn Thr Ala Tyr
515 520 525

Ile Lys Cys Val Lys Phe Gly Glu Gln Phe Lys Phe Pro Gly Ser Ile
530 535 540

Pro Ala Thr Asn Ala Arg Leu Gly Phe Leu Val Lys Asp Ala Glu Asp
545 550 555 560

Ala Cys Ser Thr Leu Arg Ala Ile Cys Ala Gln Phe Ala Lys Asp Val
565 570 575

Thr Lys Glu Ala Trp Arg Leu Pro Arg Glu Gly Val Ser Phe Arg Ala
580 585 590

Lys Gly Ile Ala Thr Asn Gly Ala Val Ala Ala Leu Phe Ser Gly Gln
595 600 605

Gly Ala Gln Tyr Thr His Met Phe Ser Glu Val Ala Met Asn Trp Pro
610 615 620

Gln Phe Arg Gln Ser Ile Ala Ala Met Asp Ala Ala Gln Ser Lys Val
625 630 635 640

Ala Gly Ser Asp Lys Asp Phe Glu Arg Val Ser Gln Val Leu Tyr Pro
645 650 655

Arg Lys Pro Tyr Glu Arg Glu Pro Glu Gln Asn Pro Lys Lys Ile Ser
660 665 670

Leu Thr Ala Tyr Ser Gln Pro Ser Thr Leu Ala Cys Ala Leu Gly Ala
675 680 685

Phe Glu Ile Phe Lys Glu Ala Gly Phe Thr Pro Asp Phe Ala Ala Gly
690 695 700

His Ser Leu Gly Glu Phe Ala Ala Leu Tyr Ala Ala Gly Cys Val Asp
705 710 715 720

Arg Asp Glu Leu Phe Glu Leu Val Cys Arg Arg Ala Arg Ile Met Gly
725 730 735

Gly Lys Asp Ala Pro Ala Thr Pro Lys Gly Cys Met Ala Ala Val Ile
740 745 750

Gly Pro Asn Ala Glu Asn Ile Lys Val Gln Ala Ala Asn Val Trp Leu
755 760 765

Gly Asn Ser Asn Ser Pro Ser Gln Thr Val Ile Thr Gly Ser Val Glu
770 775 780

Gly Ile Gln Ala Glu Ser Ala Arg Leu Gln Lys Glu Gly Phe Arg Val
785 790 795 800

Val Pro Leu Ala Cys Glu Ser Ala Phe His Ser Pro Gln Met Glu Asn
805 810 815

Ala Ser Ser Ala Phe Lys Asp Val Ile Ser Lys Val Ser Phe Arg Thr
820 825 830

Pro Lys Ala Glu Thr Lys Leu Phe Ser Asn Val Ser Gly Glu Thr Tyr
835 840 845

Pro Thr Asp Ala Arg Glu Met Leu Thr Gln His Met Thr Ser Ser Val
850 855 860

Lys Phe Leu Thr Gln Val Arg Asn Met His Gln Ala Gly Ala Arg Ile
865 870 875 880

Phe Val Glu Phe Gly Pro Lys Gln Val Leu Ser Lys Leu Val Ser Glu
885 890 895

Thr Leu Lys Asp Asp Pro Ser Val Val Thr Val Ser Val Asn Pro Ala
900 905 910

Ser Gly Thr Asp Ser Asp Ile Gln Leu Arg Asp Ala Ala Val Gln Leu
915 920 925

Val Val Ala Gly Val Asn Leu Gln Gly Phe Asp Lys Trp Asp Ala Pro
930 935 940

Asp Ala Thr Arg Met Gln Ala Ile Lys Lys Lys Arg Thr Thr Leu Arg
945 950 955 960

Leu Ser Ala Ala Thr Tyr Val Ser Asp Lys Thr Lys Lys Val Arg Asp
965 970 975

Ala Ala Met Asn Asp Gly Arg Cys Val Thr Tyr Leu Lys Gly Ala Ala
980 985 990

Pro Leu Ile Lys Ala Pro Glu Pro Val Val Asp Glu Ala Ala Lys Arg
995 1000 1005

Glu Ala Glu Arg Leu Gln Lys Glu Leu Gln Asp Ala Gln Arg Gln
1010 1015 1020

Leu Asp Asp Ala Lys Arg Ala Ala Ala Glu Ala Asn Ser Lys Leu
1025 1030 1035

Ala Ala Ala Lys Glu Glu Ala Lys Thr Ala Ala Ala Ser Ala Lys
1040 1045 1050

Pro Ala Val Asp Thr Ala Val Val Glu Lys His Arg Ala Ile Leu
1055 1060 1065

Lys Ser Met Leu Ala Glu Leu Asp Gly Tyr Gly Ser Val Asp Ala
1070 1075 1080

Ser Ser Leu Gln Gln Gln Gln Gln Gln Thr Ala Pro Ala Pro
1085 1090 1095

Val Lys Ala Ala Ala Pro Ala Ala Pro Val Ala Ser Ala Pro Ala
1100 1105 1110

Pro Ala Val Ser Asn Glu Leu Leu Glu Lys Ala Glu Thr Val Val
1115 1120 1125

Met Glu Val Leu Ala Ala Lys Thr Gly Tyr Glu Thr Asp Met Ile
1130 1135 1140

Glu Ala Asp Met Glu Leu Glu Thr Glu Leu Gly Ile Asp Ser Ile
1145 1150 1155

Lys Arg Val Glu Ile Leu Ser Glu Val Gln Ala Met Leu Asn Val
1160 1165 1170

Glu Ala Lys Asp Val Asp Ala Leu Ser Arg Thr Arg Thr Val Gly
1175 1180 1185

Glu Val Val Asn Ala Met Lys Ala Glu Ile Ala Gly Ser Ser Ala
1190 1195 1200

Pro Ala Pro Ala Ala Ala Ala Pro Ala Pro Ala Lys Ala Ala Pro

1205

1210

1215

Ala Ala Ala Ala Pro Ala Val Ser Asn Glu Leu Leu Glu Lys Ala
1220 1225 1230

Glu Thr Val Val Met Glu Val Leu Ala Ala Lys Thr Gly Tyr Glu
1235 1240 1245

Thr Asp Met Ile Glu Ser Asp Met Glu Leu Glu Thr Glu Leu Gly
1250 1255 1260

Ile Asp Ser Ile Lys Arg Val Glu Ile Leu Ser Glu Val Gln Ala
1265 1270 1275

Met Leu Asn Val Glu Ala Lys Asp Val Asp Ala Leu Ser Arg Thr
1280 1285 1290

Arg Thr Val Gly Glu Val Val Asn Ala Met Lys Ala Glu Ile Ala
1295 1300 1305

Gly Gly Ser Ala Pro Ala Pro Ala Ala Ala Ala Pro Gly Pro Ala
1310 1315 1320

Ala Ala Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Val Ser Asn
1325 1330 1335

Glu Leu Leu Glu Lys Ala Glu Thr Val Val Met Glu Val Leu Ala
1340 1345 1350

Ala Lys Thr Gly Tyr Glu Thr Asp Met Ile Glu Ser Asp Met Glu
1355 1360 1365

Leu Glu Thr Glu Leu Gly Ile Asp Ser Ile Lys Arg Val Glu Ile
1370 1375 1380

Leu Ser Glu Val Gln Ala Met Leu Asn Val Glu Ala Lys Asp Val
1385 1390 1395

Asp Ala Leu Ser Arg Thr Arg Thr Val Gly Glu Val Val Asp Ala
1400 1405 1410

Met Lys Ala Glu Ile Ala Gly Gly Ser Ala Pro Ala Pro Ala Ala
1415 1420 1425

Ala Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Pro Ala Ala Pro
1430 1435 1440

Ala Pro Ala Val Ser Ser Glu Leu Leu Glu Lys Ala Glu Thr Val
1445 1450 1455

Val Met Glu Val Leu Ala Ala Lys Thr Gly Tyr Glu Thr Asp Met
1460 1465 1470

Ile Glu Ser Asp Met Glu Leu Glu Thr Glu Leu Gly Ile Asp Ser
1475 1480 1485

Ile Lys Arg Val Glu Ile Leu Ser Glu Val Gln Ala Met Leu Asn
1490 1495 1500

Val Glu Ala Lys Asp Val Asp Ala Leu Ser Arg Thr Arg Thr Val
1505 1510 1515

Gly Glu Val Val Asp Ala Met Lys Ala Glu Ile Ala Gly Gly Ser
1520 1525 1530

Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Pro Ala Ala Ala Ala
1535 1540 1545

Pro Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Pro Ala Val
1550 1555 1560

Ser Ser Glu Leu Leu Glu Lys Ala Glu Thr Val Val Met Glu Val
1565 1570 1575

Leu Ala Ala Lys Thr Gly Tyr Glu Thr Asp Met Ile Glu Ser Asp
1580 1585 1590

Met Glu Leu Glu Thr Glu Leu Gly Ile Asp Ser Ile Lys Arg Val
1595 1600 1605

Glu Ile Leu Ser Glu Val Gln Ala Met Leu Asn Val Glu Ala Lys
1610 1615 1620

Asp Val Asp Ala Leu Ser Arg Thr Arg Thr Val Gly Glu Val Val
1625 1630 1635

Asp Ala Met Lys Ala Glu Ile Ala Gly Ser Ser Ala Ser Ala Pro
1640 1645 1650

Ala Ala Ala Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Pro Ala
1655 1660 1665

Ala Ala Ala Pro Ala Val Ser Asn Glu Leu Leu Glu Lys Ala Glu
1670 1675 1680

Thr Val Val Met Glu Val Leu Ala Ala Lys Thr Gly Tyr Glu Thr
1685 1690 1695

Asp Met Ile Glu Ser Asp Met Glu Leu Glu Thr Glu Leu Gly Ile
1700 1705 1710

Asp Ser Ile Lys Arg Val Glu Ile Leu Ser Glu Val Gln Ala Met
1715 1720 1725

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

1730

1735

1740

Thr Val Gly Glu Val Val Asp Ala Met Lys Ala Glu Ile Ala Gly
1745 1750 1755

Gly Ser Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Pro Ala Ala
1760 1765 1770

Ala Ala Pro Ala Val Ser Asn Glu Leu Leu Glu Lys Ala Glu Thr
1775 1780 1785

Val Val Met Glu Val Leu Ala Ala Lys Thr Gly Tyr Glu Thr Asp
1790 1795 1800

Met Ile Glu Ser Asp Met Glu Leu Glu Thr Glu Leu Gly Ile Asp
1805 1810 1815

Ser Ile Lys Arg Val Glu Ile Leu Ser Glu Val Gln Ala Met Leu
1820 1825 1830

Asn Val Glu Ala Lys Asp Val Asp Ala Leu Ser Arg Thr Arg Thr
1835 1840 1845

Val Gly Glu Val Val Asp Ala Met Lys Ala Glu Ile Ala Gly Ser
1850 1855 1860

Ser Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Pro Ala Ala Ala
1865 1870 1875

Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Val Ser Ser Glu Leu
1880 1885 1890

Leu Glu Lys Ala Glu Thr Val Val Met Glu Val Leu Ala Ala Lys
1895 1900 1905

Thr Gly Tyr Glu Thr Asp Met Ile Glu Ser Asp Met Glu Leu Glu
1910 1915 1920

Thr Glu Leu Gly Ile Asp Ser Ile Lys Arg Val Glu Ile Leu Ser
1925 1930 1935

Glu Val Gln Ala Met Leu Asn Val Glu Ala Lys Asp Val Asp Ala
1940 1945 1950

Leu Ser Arg Thr Arg Thr Val Gly Glu Val Val Asp Ala Met Lys
1955 1960 1965

Ala Glu Ile Ala Gly Gly Ser Ala Pro Ala Pro Ala Ala Ala Ala
1970 1975 1980

Pro Ala Pro Ala Ala Ala Ala Pro Ala Val Ser Asn Glu Leu Leu
1985 1990 1995

Glu Lys Ala Glu Thr Val Val Met Glu Val Leu Ala Ala Lys Thr
2000 2005 2010

Gly Tyr Glu Thr Asp Met Ile Glu Ser Asp Met Glu Leu Glu Thr
2015 2020 2025

Glu Leu Gly Ile Asp Ser Ile Lys Arg Val Glu Ile Leu Ser Glu
2030 2035 2040

Val Gln Ala Met Leu Asn Val Glu Ala Lys Asp Val Asp Ala Leu
2045 2050 2055

Ser Arg Thr Arg Thr Val Gly Glu Val Val Asp Ala Met Lys Ala
2060 2065 2070

Glu Ile Ala Gly Gly Ser Ala Pro Ala Pro Ala Ala Ala Ala Pro
2075 2080 2085

Ala Ser Ala Gly Ala Ala Pro Ala Val Lys Ile Asp Ser Val His
2090 2095 2100

Gly Ala Asp Cys Asp Asp Leu Ser Leu Met His Ala Lys Val Val
2105 2110 2115

Asp Ile Arg Arg Pro Asp Glu Leu Ile Leu Glu Arg Pro Glu Asn
2120 2125 2130

Arg Pro Val Leu Val Val Asp Asp Gly Ser Glu Leu Thr Leu Ala
2135 2140 2145

Leu Val Arg Val Leu Gly Ala Cys Ala Val Val Leu Thr Phe Glu
2150 2155 2160

Gly Leu Gln Leu Ala Gln Arg Ala Gly Ala Ala Ala Ile Arg His
2165 2170 2175

Val Leu Ala Lys Asp Leu Ser Ala Glu Ser Ala Glu Lys Ala Ile
2180 2185 2190

Lys Glu Ala Glu Gln Arg Phe Gly Ala Leu Gly Gly Phe Ile Ser
2195 2200 2205

Gln Gln Ala Glu Arg Phe Glu Pro Ala Glu Ile Leu Gly Phe Thr
2210 2215 2220

Leu Met Cys Ala Lys Phe Ala Lys Ala Ser Leu Cys Thr Ala Val
2225 2230 2235

Ala Gly Gly Arg Pro Ala Phe Ile Gly Val Ala Arg Leu Asp Gly
2240 2245 2250

Arg Leu Gly Phe Thr Ser Gln Gly Thr Ser Asp Ala Leu Lys Arg

2255	2260	2265
Ala Gln Arg Gly Ala Ile Phe Gly Leu Cys Lys Thr Ile Gly Leu		
2270	2275	2280
Glu Trp Ser Glu Ser Asp Val Phe Ser Arg Gly Val Asp Ile Ala		
2285	2290	2295
Gln Gly Met His Pro Glu Asp Ala Ala Val Ala Ile Val Arg Glu		
2300	2305	2310
Met Ala Cys Ala Asp Ile Arg Ile Arg Glu Val Gly Ile Gly Ala		
2315	2320	2325
Asn Gln Gln Arg Cys Thr Ile Arg Ala Ala Lys Leu Glu Thr Gly		
2330	2335	2340
Asn Pro Gln Arg Gln Ile Ala Lys Asp Asp Val Leu Leu Val Ser		
2345	2350	2355
Gly Gly Ala Arg Gly Ile Thr Pro Leu Cys Ile Arg Glu Ile Thr		
2360	2365	2370
Arg Gln Ile Ala Gly Gly Lys Tyr Ile Leu Leu Gly Arg Ser Lys		
2375	2380	2385
Val Ser Ala Ser Glu Pro Ala Trp Cys Ala Gly Ile Thr Asp Glu		
2390	2395	2400
Lys Ala Val Gln Lys Ala Ala Thr Gln Glu Leu Lys Arg Ala Phe		
2405	2410	2415
Ser Ala Gly Glu Gly Pro Lys Pro Thr Pro Arg Ala Val Thr Lys		
2420	2425	2430

Leu Val Gly Ser Val Leu Gly Ala Arg Glu Val Arg Ser Ser Ile
2435 2440 2445

Ala Ala Ile Glu Ala Leu Gly Gly Lys Ala Ile Tyr Ser Ser Cys
2450 2455 2460

Asp Val Asn Ser Ala Ala Asp Val Ala Lys Ala Val Arg Asp Ala
2465 2470 2475

Glu Ser Gln Leu Gly Ala Arg Val Ser Gly Ile Val His Ala Ser
2480 2485 2490

Gly Val Leu Arg Asp Arg Leu Ile Glu Lys Lys Leu Pro Asp Glu
2495 2500 2505

Phe Asp Ala Val Phe Gly Thr Lys Val Thr Gly Leu Glu Asn Leu
2510 2515 2520

Leu Ala Ala Val Asp Arg Ala Asn Leu Lys His Met Val Leu Phe
2525 2530 2535

Ser Ser Leu Ala Gly Phe His Gly Asn Val Gly Gln Ser Asp Tyr
2540 2545 2550

Ala Met Ala Asn Glu Ala Leu Asn Lys Met Gly Leu Glu Leu Ala
2555 2560 2565

Lys Asp Val Ser Val Lys Ser Ile Cys Phe Gly Pro Trp Asp Gly
2570 2575 2580

Gly Met Val Thr Pro Gln Leu Lys Lys Gln Phe Gln Glu Met Gly
2585 2590 2595

Val Gln Ile Ile Pro Arg Glu Gly Gly Ala Asp Thr Val Ala Arg
2600 2605 2610

Ile Val Leu Gly Ser Ser Pro Ala Glu Ile Leu Val Gly Asn Trp
2615 2620 2625

Arg Thr Pro Ser Lys Lys Val Gly Ser Asp Thr Ile Thr Leu His
2630 2635 2640

Arg Lys Ile Ser Ala Lys Ser Asn Pro Phe Leu Glu Asp His Val
2645 2650 2655

Ile Gln Gly Arg Arg Val Leu Pro Met Thr Leu Ala Ile Gly Ser
2660 2665 2670

Leu Ala Glu Thr Cys Leu Gly Leu Phe Pro Gly Tyr Ser Leu Trp
2675 2680 2685

Ala Ile Asp Asp Ala Gln Leu Phe Lys Gly Val Thr Val Asp Gly
2690 2695 2700

Asp Val Asn Cys Glu Val Thr Leu Thr Pro Ser Thr Ala Pro Ser
2705 2710 2715

Gly Arg Val Asn Val Gln Ala Thr Leu Lys Thr Phe Ser Ser Gly
2720 2725 2730

Lys Leu Val Pro Ala Tyr Arg Ala Val Ile Val Leu Ser Asn Gln
2735 2740 2745

Gly Ala Pro Pro Ala Asn Ala Thr Met Gln Pro Pro Ser Leu Asp
2750 2755 2760

Ala Asp Pro Ala Leu Gln Gly Ser Val Tyr Asp Gly Lys Thr Leu
2765 2770 2775

Phe His Gly Pro Ala Phe Arg Gly Ile Asp Asp Val Leu Ser Cys

Met	Ala	Ala	Arg	Asn	Val	Ser	Ala	Ala	His	Glu	Met	His	Asp	Glu	Lys	
1				5					10					15		
cgc	atc	gcc	gtc	gtc	ggc	atg	gcc	gtc	cag	tac	gcc	gga	tgc	aaa	acc	96
Arg	Ile	Ala	Val	Val	Gly	Met	Ala	Val	Gln	Tyr	Ala	Gly	Cys	Lys	Thr	
			20					25					30			
aag	gac	gag	ttc	tgg	gag	gtg	ctc	atg	aac	ggc	aag	gtc	gag	tcc	aag	144
Lys	Asp	Glu	Phe	Trp	Glu	Val	Leu	Met	Asn	Gly	Lys	Val	Glu	Ser	Lys	
		35					40				45					
gtg	atc	agc	gac	aaa	cga	ctc	ggc	tcc	aac	tac	cgc	gcc	gag	cac	tac	192
Val	Ile	Ser	Asp	Lys	Arg	Leu	Gly	Ser	Asn	Tyr	Arg	Ala	Glu	His	Tyr	
	50					55				60						
aaa	gca	gag	cgc	agc	aag	tat	gcc	gac	acc	ttt	tgc	aac	gaa	acg	tac	240
Lys	Ala	Glu	Arg	Ser	Lys	Tyr	Ala	Asp	Thr	Phe	Cys	Asn	Glu	Thr	Tyr	
65					70				75					80		
ggc	acc	ctt	gac	gag	aac	gag	atc	gac	aac	gag	cac	gaa	ctc	ctc	ctc	288
Gly	Thr	Leu	Asp	Glu	Asn	Glu	Ile	Asp	Asn	Glu	His	Glu	Leu	Leu	Leu	
			85					90					95			
aac	ctc	gcc	aag	cag	gca	ctc	gca	gag	aca	tcc	gtc	aaa	gac	tcg	aca	336
Asn	Leu	Ala	Lys	Gln	Ala	Leu	Ala	Glu	Thr	Ser	Val	Lys	Asp	Ser	Thr	
		100					105					110				
cgc	tgc	ggc	atc	gtc	agc	ggc	tgc	ctc	tcg	ttc	ccc	atg	gac	aac	ctc	384
Arg	Cys	Gly	Ile	Val	Ser	Gly	Cys	Leu	Ser	Phe	Pro	Met	Asp	Asn	Leu	
		115				120				125						
cag	ggt	gaa	ctc	ctc	aac	gtg	tac	caa	aac	cat	gtc	gag	aaa	aag	ctc	432
Gln	Gly	Glu	Leu	Leu	Asn	Val	Tyr	Gln	Asn	His	Val	Glu	Lys	Lys	Leu	
	130					135				140						
ggg	gcc	cgc	gtc	ttc	aag	gac	gcc	tcc	cat	tgg	tcc	gaa	cgc	gag	cag	480
Gly	Ala	Arg	Val	Phe	Lys	Asp	Ala	Ser	His	Trp	Ser	Glu	Arg	Glu	Gln	
145					150				155					160		
tcc	aac	aaa	ccc	gag	gcc	ggt	gac	cgc	cgc	atc	ttc	atg	gac	ccg	gcc	528
Ser	Asn	Lys	Pro	Glu	Ala	Gly	Asp	Arg	Arg	Ile	Phe	Met	Asp	Pro	Ala	
			165				170						175			
tcc	ttc	gtc	gcc	gaa	gaa	ctc	aac	ctc	ggc	gcc	ctt	cac	tac	tcc	gtc	576
Ser	Phe	Val	Ala	Glu	Glu	Leu	Asn	Leu	Gly	Ala	Leu	His	Tyr	Ser	Val	

180	185	190	
gac gca gca tgc gcc acg gcg ctc tac gtg ctc cgc ctc gcg cag gat			624
Asp Ala Ala Cys Ala Thr Ala Leu Tyr Val Leu Arg Leu Ala Gln Asp			
195	200	205	
cat ctc gtc tcc ggc gcc gcc gac gtc atg ctc tgc ggt gcc acc tgc			672
His Leu Val Ser Gly Ala Ala Asp Val Met Leu Cys Gly Ala Thr Cys			
210	215	220	
ctg ccg gag ccc ttt ttc atc ctt tgc ggc ttt tcc acc ttc cag gcc			720
Leu Pro Glu Pro Phe Phe Ile Leu Ser Gly Phe Ser Thr Phe Gln Ala			
225	230	235	240
atg ccc gtc ggc acg ggc cag aac gtg tcc atg ccg ctg cac aag gac			768
Met Pro Val Gly Thr Gly Gln Asn Val Ser Met Pro Leu His Lys Asp			
245	250	255	
agc cag ggc ctc acc ccg ggt gag ggc ggc tcc atc atg gtc ctc aag			816
Ser Gln Gly Leu Thr Pro Gly Glu Gly Gly Ser Ile Met Val Leu Lys			
260	265	270	
cgt ctc gat gat gcc atc cgc gac ggc gac cac att tac ggc acc ctt			864
Arg Leu Asp Asp Ala Ile Arg Asp Gly Asp His Ile Tyr Gly Thr Leu			
275	280	285	
ctc ggc gcc aat gtc agc aac tcc ggc aca ggt ctg ccc ctc aag ccc			912
Leu Gly Ala Asn Val Ser Asn Ser Gly Thr Gly Leu Pro Leu Lys Pro			
290	295	300	
ctt ctc ccc agc gag aaa aag tgc ctc atg gac acc tac acg cgc att			960
Leu Leu Pro Ser Glu Lys Lys Cys Leu Met Asp Thr Tyr Thr Arg Ile			
305	310	315	320
aac gtg cac ccg cac aag att cag tac gtc gag tgc cac gcc acc ggc			1008
Asn Val His Pro His Lys Ile Gln Tyr Val Glu Cys His Ala Thr Gly			
325	330	335	
acg ccc cag ggt gat cgt gtg gaa atc gac gcc gtc aag gcc tgc ttt			1056
Thr Pro Gln Gly Asp Arg Val Glu Ile Asp Ala Val Lys Ala Cys Phe			
340	345	350	
gaa ggc aag gtc ccc cgt ttc ggt acc aca aag ggc aac ttt gga cac			1104
Glu Gly Lys Val Pro Arg Phe Gly Thr Thr Lys Gly Asn Phe Gly His			

355	360	365	
acc cts gyc gca gcc ggc ttt gcc ggt atg tgc aag gtc ctc ctc tcc			1152
Thr Xaa Xaa Ala Ala Gly Phe Ala Gly Met Cys Lys Val Leu Leu Ser			
370	375	380	
atg aag cat ggc atc atc ccg ccc acc ccg ggt atc gat gac gag acc			1200
Met Lys His Gly Ile Ile Pro Pro Thr Pro Gly Ile Asp Asp Glu Thr			
385	390	395	400
aag atg gac cct ctc gtc gtc tcc ggt gag gcc atc cca tgg cca gag			1248
Lys Met Asp Pro Leu Val Val Ser Gly Glu Ala Ile Pro Trp Pro Glu			
405	410	415	
acc aac ggc gag ccc aag cgc gcc ggt ctc tcg gcc ttt ggc ttt ggt			1296
Thr Asn Gly Glu Pro Lys Arg Ala Gly Leu Ser Ala Phe Gly Phe Gly			
420	425	430	
ggc acc aac gcc cat gcc gtc ttt gag gag cat gac ccc tcc aac gcc			1344
Gly Thr Asn Ala His Ala Val Phe Glu Glu His Asp Pro Ser Asn Ala			
435	440	445	
gcc tgc acg ggc cac gac tcc att tct gcg ctc tcg gcc cgc tgc ggc			1392
Ala Cys Thr Gly His Asp Ser Ile Ser Ala Leu Ser Ala Arg Cys Gly			
450	455	460	
ggt gaa agc aac atg cgc atc gcc atc act ggt atg gac gcc acc ttt			1440
Gly Glu Ser Asn Met Arg Ile Ala Ile Thr Gly Met Asp Ala Thr Phe			
465	470	475	480
ggc gct ctc aag gga ctc gac gcc ttc gag cgc gcc att tac acc ggc			1488
Gly Ala Leu Lys Gly Leu Asp Ala Phe Glu Arg Ala Ile Tyr Thr Gly			
485	490	495	
gct cac ggt gcc atc cca ctc cca gaa aag cgc tgg cgc ttt ctc ggc			1536
Ala His Gly Ala Ile Pro Leu Pro Glu Lys Arg Trp Arg Phe Leu Gly			
500	505	510	
aag gac aag gac ttt ctt gac ctc tgc ggc gtc aag gcc acc ccg cac			1584
Lys Asp Lys Asp Phe Leu Asp Leu Cys Gly Val Lys Ala Thr Pro His			
515	520	525	
ggc tgc tac att gaa gat gtt gag gtc gac ttc cag cgc ctc cgc acg			1632
Gly Cys Tyr Ile Glu Asp Val Glu Val Asp Phe Gln Arg Leu Arg Thr			

530	535	540	
ccc atg acc cct gaa gac atg ctc ctc cct cag cag ctt ctg gcc gtc			1680
Pro Met Thr Pro Glu Asp Met Leu Leu Pro Gln Gln Leu Leu Ala Val			
545	550	555	560
acc acc att gac cgc gcc atc ctc gac tcg gga atg aaa aag ggt ggc			1728
Thr Thr Ile Asp Arg Ala Ile Leu Asp Ser Gly Met Lys Lys Gly Gly			
	565	570	575
aat gtc gcc gtc ttt gtc ggc ctc ggc acc gac ctc gag ctc tac cgt			1776
Asn Val Ala Val Phe Val Gly Leu Gly Thr Asp Leu Glu Leu Tyr Arg			
	580	585	590
cac cgt gct cgc gtc gct ctc aag gag cgc gtc cgc cct gaa gcc tcc			1824
His Arg Ala Arg Val Ala Leu Lys Glu Arg Val Arg Pro Glu Ala Ser			
	595	600	605
aag aag ctc aat gac atg atg cag tac att aac gac tgc ggc aca tcc			1872
Lys Lys Leu Asn Asp Met Met Gln Tyr Ile Asn Asp Cys Gly Thr Ser			
	610	615	620
aca tcg tac acc tcg tac att ggc aac ctc gtc gcc acg cgc gtc tcg			1920
Thr Ser Tyr Thr Ser Tyr Ile Gly Asn Leu Val Ala Thr Arg Val Ser			
	625	630	635
tcg cag tgg ggc ttc acg ggc ccc tcc ttt acg atc acc gag ggc aac			1968
Ser Gln Trp Gly Phe Thr Gly Pro Ser Phe Thr Ile Thr Glu Gly Asn			
	645	650	655
aac tcc gtc tac cgc tgc gcc gag ctc ggc aag tac ctc ctc gag acc			2016
Asn Ser Val Tyr Arg Cys Ala Glu Leu Gly Lys Tyr Leu Leu Glu Thr			
	660	665	670
ggc gag gtc gat ggc gtc gtc gtt gcg ggt gtc gat ctc tgc ggc agt			2064
Gly Glu Val Asp Gly Val Val Val Ala Gly Val Asp Leu Cys Gly Ser			
	675	680	685
gcc gaa aac ctt tac gtc aag tct cgc cgc ttc aag gtg tcc acc tcc			2112
Ala Glu Asn Leu Tyr Val Lys Ser Arg Arg Phe Lys Val Ser Thr Ser			
	690	695	700
gat acc ccg cgc gcc agc ttt gac gcc gcc gcc gat ggc tac ttt gtc			2160
Asp Thr Pro Arg Ala Ser Phe Asp Ala Ala Ala Asp Gly Tyr Phe Val			

705	710	715	720
ggc gag ggc tgc ggt gcc ttt gtg ctc aag cgt gag act agc tgc acc Gly Glu Gly Cys Gly Ala Phe Val Leu Lys Arg Glu Thr Ser Cys Thr 725 730 735	2208		
aag gac gac cgt atc tac gct tgc atg gat gcc atc gtc cct ggc aac Lys Asp Asp Arg Ile Tyr Ala Cys Met Asp Ala Ile Val Pro Gly Asn 740 745 750	2256		
gtc cct agc gcc tgc ttg cgc gag gcc ctc gac cag gcg cgc gtc aag Val Pro Ser Ala Cys Leu Arg Glu Ala Leu Asp Gln Ala Arg Val Lys 755 760 765	2304		
ccg ggc gat atc gag atg ctc gag ctc agc gcc gac tcc gcc cgc cac Pro Gly Asp Ile Glu Met Leu Glu Leu Ser Ala Asp Ser Ala Arg His 770 775 780	2352		
ctc aag gac ccg tcc gtc ctg ccc aag gag ctc act gcc gag gag gaa Leu Lys Asp Pro Ser Val Leu Pro Lys Glu Leu Thr Ala Glu Glu Glu 785 790 795 800	2400		
atc ggc ggc ctt cag acg atc ctt cgt gac gat gac aag ctc ccg cgc Ile Gly Gly Leu Gln Thr Ile Leu Arg Asp Asp Asp Lys Leu Pro Arg 805 810 815	2448		
aac gtc gca acg ggc agt gtc aag gcc acc gtc ggt gac acc ggt tat Asn Val Ala Thr Gly Ser Val Lys Ala Thr Val Gly Asp Thr Gly Tyr 820 825 830	2496		
gcc tct ggt gct gcc agc ctc atc aag gct gcg ctt tgc atc tac aac Ala Ser Gly Ala Ala Ser Leu Ile Lys Ala Ala Leu Cys Ile Tyr Asn 835 840 845	2544		
cgc tac ctg ccc agc aac ggc gac gac tgg gat gaa ccc gcc cct gag Arg Tyr Leu Pro Ser Asn Gly Asp Asp Trp Asp Glu Pro Ala Pro Glu 850 855 860	2592		
gcg ccc tgg gac agc acc ctc ttt gcg tgc cag acc tcg cgc gct tgg Ala Pro Trp Asp Ser Thr Leu Phe Ala Cys Gln Thr Ser Arg Ala Trp 865 870 875 880	2640		
ctc aag aac cct ggc gag cgt cgc tat gcg gcc gtc tcg ggc gtc tcc Leu Lys Asn Pro Gly Glu Arg Arg Tyr Ala Ala Val Ser Gly Val Ser	2688		

885	890	895	
gag acg cgc tgc tgc tat tcc gtg ctc ctc tcc gaa gcc gag ggc cac			2736
Glu Thr Arg Ser Cys Tyr Ser Val Leu Leu Ser Glu Ala Glu Gly His			
900	905	910	
tac gag cgc gag aac cgc atc tgc ctc gac gag gag gcg ccc aag ctc			2784
Tyr Glu Arg Glu Asn Arg Ile Ser Leu Asp Glu Glu Ala Pro Lys Leu			
915	920	925	
att gtg ctt cgc gcc gac tcc cac gag gag atc ctt ggt cgc ctc gac			2832
Ile Val Leu Arg Ala Asp Ser His Glu Glu Ile Leu Gly Arg Leu Asp			
930	935	940	
aag atc cgc gag cgc ttc ttg cag ccc acg ggc gcc gcc ccg cgc gag			2880
Lys Ile Arg Glu Arg Phe Leu Gln Pro Thr Gly Ala Ala Pro Arg Glu			
945	950	955	960
tcc gag ctc aag gcg cag gcc cgc cgc atc ttc ctc gag ctc ctc ggc			2928
Ser Glu Leu Lys Ala Gln Ala Arg Arg Ile Phe Leu Glu Leu Leu Gly			
965	970	975	
gag acc ctt gcc cag gat gcc gct tct tca ggc tcg caa aag ccc ctc			2976
Glu Thr Leu Ala Gln Asp Ala Ala Ser Ser Gly Ser Gln Lys Pro Leu			
980	985	990	
gct ctc agc ctc gtc tcc acg ccc tcc aag ctc cag cgc gag gtc gag			3024
Ala Leu Ser Leu Val Ser Thr Pro Ser Lys Leu Gln Arg Glu Val Glu			
995	1000	1005	
ctc gcg gcc aag ggt atc ccg cgc tgc ctc aag atg cgc cgc gat			3069
Leu Ala Ala Lys Gly Ile Pro Arg Cys Leu Lys Met Arg Arg Asp			
1010	1015	1020	
tgg agc tcc cct gct ggc agc cgc tac gcg cct gag ccg ctc gcc			3114
Trp Ser Ser Pro Ala Gly Ser Arg Tyr Ala Pro Glu Pro Leu Ala			
1025	1030	1035	
agc gac cgc gtc gcc ttc atg tac ggc gaa ggt cgc agc cct tac			3159
Ser Asp Arg Val Ala Phe Met Tyr Gly Glu Gly Arg Ser Pro Tyr			
1040	1045	1050	
tac ggc atc acc caa gac att cac cgc att tgg ccc gaa ctc cac			3204
Tyr Gly Ile Thr Gln Asp Ile His Arg Ile Trp Pro Glu Leu His			

1055	1060	1065	
gag gtc atc aac gaa aag acg aac cgt ctc tgg gcc gaa ggc gac			3249
Glu Val Ile Asn Glu Lys Thr Asn Arg Leu Trp Ala Glu Gly Asp			
1070	1075	1080	
cgc tgg gtc atg ccg cgc gcc agc ttc aag tcg gag ctc gag agc			3294
Arg Trp Val Met Pro Arg Ala Ser Phe Lys Ser Glu Leu Glu Ser			
1085	1090	1095	
cag cag caa gag ttt gat cgc aac atg att gaa atg ttc cgt ctt			3339
Gln Gln Gln Glu Phe Asp Arg Asn Met Ile Glu Met Phe Arg Leu			
1100	1105	1110	
gga atc ctc acc tca att gcc ttc acc aat ctg gcg cgc gac gtt			3384
Gly Ile Leu Thr Ser Ile Ala Phe Thr Asn Leu Ala Arg Asp Val			
1115	1120	1125	
ctc aac atc acg ccc aag gcc gcc ttt ggc ctc agt ctt ggc gag			3429
Leu Asn Ile Thr Pro Lys Ala Ala Phe Gly Leu Ser Leu Gly Glu			
1130	1135	1140	
att tcc atg att ttt gcc ttt tcc aag aag aac ggt ctc atc tcc			3474
Ile Ser Met Ile Phe Ala Phe Ser Lys Lys Asn Gly Leu Ile Ser			
1145	1150	1155	
gac cag ctc acc aag gat ctt cgc gag tcc gac gtg tgg aac aag			3519
Asp Gln Leu Thr Lys Asp Leu Arg Glu Ser Asp Val Trp Asn Lys			
1160	1165	1170	
gct ctg gcc gtt gaa ttt aat gcg ctg cgc gag gcc tgg ggc att			3564
Ala Leu Ala Val Glu Phe Asn Ala Leu Arg Glu Ala Trp Gly Ile			
1175	1180	1185	
cca cag agt gtc ccc aag gac gag ttc tgg caa ggc tac att gtg			3609
Pro Gln Ser Val Pro Lys Asp Glu Phe Trp Gln Gly Tyr Ile Val			
1190	1195	1200	
cgc ggc acc aag cag gat atc gag gcg gcc atc gcc ccg gac agc			3654
Arg Gly Thr Lys Gln Asp Ile Glu Ala Ala Ile Ala Pro Asp Ser			
1205	1210	1215	
aag tac gtg cgc ctc acc atc atc aat gat gcc aac acc gcc ctc			3699
Lys Tyr Val Arg Leu Thr Ile Ile Asn Asp Ala Asn Thr Ala Leu			

1220	1225	1230	
att agc ggc aag ccc gac gcc tgc aag gct gcg atc gcg cgt ctc Ile Ser Gly Lys Pro Asp Ala Cys Lys Ala Ala Ile Ala Arg Leu 1235 1240 1245			3744
ggt ggc aac att cct gcg ctt ccc gtg acc cag ggc atg tgc ggc Gly Gly Asn Ile Pro Ala Leu Pro Val Thr Gln Gly Met Cys Gly 1250 1255 1260			3789
cac tgc ccc gag gtg gga cct tat acc aag gat atc gcc aag atc His Cys Pro Glu Val Gly Pro Tyr Thr Lys Asp Ile Ala Lys Ile 1265 1270 1275			3834
cat gcc aac ctt gag ttc ccc gtt gtc gac ggc ctt gac ctc tgg His Ala Asn Leu Glu Phe Pro Val Val Asp Gly Leu Asp Leu Trp 1280 1285 1290			3879
acc aca atc aac cag aag cgc ctc gtg cca cgc gcc acg ggc gcc Thr Thr Ile Asn Gln Lys Arg Leu Val Pro Arg Ala Thr Gly Ala 1295 1300 1305			3924
aag gac gaa tgg gcc cct tct tcc ttt ggc gag tac gcc ggc cag Lys Asp Glu Trp Ala Pro Ser Ser Phe Gly Glu Tyr Ala Gly Gln 1310 1315 1320			3969
ctc tac gag aag cag gct aac ttc ccc caa atc gtc gag acc att Leu Tyr Glu Lys Gln Ala Asn Phe Pro Gln Ile Val Glu Thr Ile 1325 1330 1335			4014
tac aag caa aac tac gac gtc ttt gtc gag gtt ggg ccc aac aac Tyr Lys Gln Asn Tyr Asp Val Phe Val Glu Val Gly Pro Asn Asn 1340 1345 1350			4059
cac cgt agc acc gca gtg cgc acc acg ctt ggt ccc cag cgc aac His Arg Ser Thr Ala Val Arg Thr Thr Leu Gly Pro Gln Arg Asn 1355 1360 1365			4104
cac ctt gct ggc gcc atc gac aag cag aac gag gat gct tgg acg His Leu Ala Gly Ala Ile Asp Lys Gln Asn Glu Asp Ala Trp Thr 1370 1375 1380			4149
acc atc gtc aag ctt gtg gct tcg ctc aag gcc cac ctt gtt cct Thr Ile Val Lys Leu Val Ala Ser Leu Lys Ala His Leu Val Pro 1385 1390 1395			4194

1385	1390	1395	
ggc gtc acg atc tcg ccg ctg	tac cac tcc aag ctt	gtg gcg gag	4239
Gly Val Thr Ile Ser Pro Leu	Tyr His Ser Lys Leu	Val Ala Glu	
1400	1405	1410	
gct cag gct tgc tac gct gcg	ctc tgc aag ggt gaa	aag ccc aag	4284
Ala Gln Ala Cys Tyr Ala Ala	Leu Cys Lys Gly Glu	Lys Pro Lys	
1415	1420	1425	
aag aac aag ttt gtg cgc aag	att cag ctc aac ggt	cgc ttc aac	4329
Lys Asn Lys Phe Val Arg Lys	Ile Gln Leu Asn Gly	Arg Phe Asn	
1430	1435	1440	
agc aag gcg gac ccc atc tcc	tcg gcc gat ctt gcc	agc ttt ccg	4374
Ser Lys Ala Asp Pro Ile Ser	Ser Ala Asp Leu Ala	Ser Phe Pro	
1445	1450	1455	
cct gcg gac cct gcc att gaa	gcc gcc atc tcg agc	cgc atc atg	4419
Pro Ala Asp Pro Ala Ile Glu	Ala Ala Ile Ser Ser	Arg Ile Met	
1460	1465	1470	
aag cct gtc gct ccc aag ttc	tac gcg cgt ctc aac	att gac gag	4464
Lys Pro Val Ala Pro Lys Phe	Tyr Ala Arg Leu Asn	Ile Asp Glu	
1475	1480	1485	
cag gac gag acc cga gat ccg	atc ctc aac aag gac	aac gcg ccg	4509
Gln Asp Glu Thr Arg Asp Pro	Ile Leu Asn Lys Asp	Asn Ala Pro	
1490	1495	1500	
tct tct tct tct tct tct	tct tct tct tct tct	tct tct tct	4554
Ser Ser Ser Ser Ser Ser	Ser Ser Ser Ser Ser	Ser Ser Ser	
1505	1510	1515	
ccg tcg cct gct cct tcg gcc	ccc gtg caa aag aag	gct gct ccc	4599
Pro Ser Pro Ala Pro Ser Ala	Pro Val Gln Lys Lys	Ala Ala Pro	
1520	1525	1530	
gcc gcg gag acc aag gct gtt	gct tcg gct gac gca	ctt cgc agt	4644
Ala Ala Glu Thr Lys Ala Val	Ala Ser Ala Asp Ala	Leu Arg Ser	
1535	1540	1545	
gcc ctg ctc gat ctc gac agt	atg ctt gcg ctg agc	tct gcc agt	4689
Ala Leu Leu Asp Leu Asp Ser	Met Leu Ala Leu Ser	Ser Ala Ser	

1550	1555	1560	
gcc tcc ggc aac ctt gtt gag act gcg cct agc gac gcc tcg gtc			4734
Ala Ser Gly Asn Leu Val Glu Thr Ala Pro Ser Asp Ala Ser Val			
1565	1570	1575	
att gtg ccg ccc tgc aac att gcg gat ctc ggc agc cgc gcc ttc			4779
Ile Val Pro Pro Cys Asn Ile Ala Asp Leu Gly Ser Arg Ala Phe			
1580	1585	1590	
atg aaa acg tac ggt gtt tcg gcg cct ctg tac acg ggc gcc atg			4824
Met Lys Thr Tyr Gly Val Ser Ala Pro Leu Tyr Thr Gly Ala Met			
1595	1600	1605	
gcc aag ggc att gcc tct gcg gac ctc gtc att gcc gcc ggc cgc			4869
Ala Lys Gly Ile Ala Ser Ala Asp Leu Val Ile Ala Ala Gly Arg			
1610	1615	1620	
cag ggc atc ctt gcg tcc ttt ggc gcc ggc gga ctt ccc atg cag			4914
Gln Gly Ile Leu Ala Ser Phe Gly Ala Gly Gly Leu Pro Met Gln			
1625	1630	1635	
gtt gtg cgt gag tcc atc gaa aag att cag gcc gcc ctg ccc aat			4959
Val Val Arg Glu Ser Ile Glu Lys Ile Gln Ala Ala Leu Pro Asn			
1640	1645	1650	
ggc ccg tac gct gtc aac ctt atc cat tct ccc ttt gac agc aac			5004
Gly Pro Tyr Ala Val Asn Leu Ile His Ser Pro Phe Asp Ser Asn			
1655	1660	1665	
ctc gaa aag ggc aat gtc gat ctc ttc ctc gag aag ggt gtc acc			5049
Leu Glu Lys Gly Asn Val Asp Leu Phe Leu Glu Lys Gly Val Thr			
1670	1675	1680	
ttt gtc gag gcc tcg gcc ttt atg acg ctc acc ccg cag gtc gtg			5094
Phe Val Glu Ala Ser Ala Phe Met Thr Leu Thr Pro Gln Val Val			
1685	1690	1695	
cgg tac cgc gcg gct ggc ctc acg cgc aac gcc gac ggc tcg gtc			5139
Arg Tyr Arg Ala Ala Gly Leu Thr Arg Asn Ala Asp Gly Ser Val			
1700	1705	1710	
aac atc cgc aac cgt atc att ggc aag gtc tcg cgc acc gag ctc			5184
Asn Ile Arg Asn Arg Ile Ile Gly Lys Val Ser Arg Thr Glu Leu			

1715	1720	1725	
gcc gag atg ttc atg cgt cct	gcg ccc gag cac ctt	ctt cag aag	5229
Ala Glu Met Phe Met Arg Pro	Ala Pro Glu His Leu	Leu Gln Lys	
1730	1735	1740	
ctc att gct tcc ggc gag atc	aac cag gag cag gcc	gag ctc gcc	5274
Leu Ile Ala Ser Gly Glu Ile	Asn Gln Glu Gln Ala	Glu Leu Ala	
1745	1750	1755	
cgc cgt gtt ccc gtc gct gac	gac atc gcg gtc gaa	gct gac tcg	5319
Arg Arg Val Pro Val Ala Asp	Asp Ile Ala Val Glu	Ala Asp Ser	
1760	1765	1770	
ggt ggc cac acc gac aac cgc	ccc atc cac gtc att	ctg ccc ctc	5364
Gly Gly His Thr Asp Asn Arg	Pro Ile His Val Ile	Leu Pro Leu	
1775	1780	1785	
atc atc aac ctt cgc gac cgc	ctt cac cgc gag tgc	ggc tac ccg	5409
Ile Ile Asn Leu Arg Asp Arg	Leu His Arg Glu Cys	Gly Tyr Pro	
1790	1795	1800	
gcc aac ctt cgc gtc cgt gtg	ggc gcc ggc ggt ggc	att ggg tgc	5454
Ala Asn Leu Arg Val Arg Val	Gly Ala Gly Gly Gly	Ile Gly Cys	
1805	1810	1815	
ccc cag gcg gcg ctg gcc acc	ttc aac atg ggt gcc	tcc ttt att	5499
Pro Gln Ala Ala Leu Ala Thr	Phe Asn Met Gly Ala	Ser Phe Ile	
1820	1825	1830	
gtc acc ggc acc gtg aac cag	gtc gcc aag cag tcg	ggc acg tgc	5544
Val Thr Gly Thr Val Asn Gln	Val Ala Lys Gln Ser	Gly Thr Cys	
1835	1840	1845	
gac aat gtg cgc aag cag ctc	gcg aag gcc act tac	tcg gac gta	5589
Asp Asn Val Arg Lys Gln Leu	Ala Lys Ala Thr Tyr	Ser Asp Val	
1850	1855	1860	
tgc atg gcc ccg gct gcc gac	atg ttc gag gaa ggc	gtc aag ctt	5634
Cys Met Ala Pro Ala Ala Asp	Met Phe Glu Glu Gly	Val Lys Leu	
1865	1870	1875	
cag gtc ctc aag aag gga acc	atg ttt ccc tcg cgc	gcc aac aag	5679
Gln Val Leu Lys Lys Gly Thr	Met Phe Pro Ser Arg	Ala Asn Lys	

1880	1885	1890	
ctc tac gag ctc ttt tgc aag tac gac tcg ttc gag tcc atg ccc			5724
Leu Tyr Glu Leu Phe Cys Lys Tyr Asp Ser Phe Glu Ser Met Pro			
1895	1900	1905	
ccc gca gag ctt gcg cgc gtc gag aag cgc atc ttc agc cgc gcg			5769
Pro Ala Glu Leu Ala Arg Val Glu Lys Arg Ile Phe Ser Arg Ala			
1910	1915	1920	
ctc gaa gag gtc tgg gac gag acc aaa aac ttt tac att aac cgt			5814
Leu Glu Glu Val Trp Asp Glu Thr Lys Asn Phe Tyr Ile Asn Arg			
1925	1930	1935	
ctt cac aac ccg gag aag atc cag cgc gcc gag cgc gac ccc aag			5859
Leu His Asn Pro Glu Lys Ile Gln Arg Ala Glu Arg Asp Pro Lys			
1940	1945	1950	
ctc aag atg tcg ctg tgc ttt cgc tgg tac ctg agc ctg gcg agc			5904
Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Ser Leu Ala Ser			
1955	1960	1965	
cgc tgg gcc aac act gga gct tcc gat cgc gtc atg gac tac cag			5949
Arg Trp Ala Asn Thr Gly Ala Ser Asp Arg Val Met Asp Tyr Gln			
1970	1975	1980	
gtc tgg tgc ggt cct gcc att ggt tcc ttc aac gat ttc atc aag			5994
Val Trp Cys Gly Pro Ala Ile Gly Ser Phe Asn Asp Phe Ile Lys			
1985	1990	1995	
gga act tac ctt gat ccg gcc gtc gca aac gag tac ccg tgc gtc			6039
Gly Thr Tyr Leu Asp Pro Ala Val Ala Asn Glu Tyr Pro Cys Val			
2000	2005	2010	
gtt cag att aac aag cag atc ctt cgt gga gcg tgc ttc ttg cgc			6084
Val Gln Ile Asn Lys Gln Ile Leu Arg Gly Ala Cys Phe Leu Arg			
2015	2020	2025	
cgt ctc gaa att ctg cgc aac gca cgc ctt tcc gat ggc gct gcc			6129
Arg Leu Glu Ile Leu Arg Asn Ala Arg Leu Ser Asp Gly Ala Ala			
2030	2035	2040	
gct ctt gtg gcc agc atc gat gac aca tac gtc ccg gcc gag aag			6174
Ala Leu Val Ala Ser Ile Asp Asp Thr Tyr Val Pro Ala Glu Lys			

2045 2050 2055

ctg 6177
Leu

<210> 4
<211> 2059
<212> PRT
<213> Schizochytrium sp.

<220>
<221> misc_feature
<222> (370)..(370)
<223> The 'Xaa' at location 370 stands for Leu.

<220>
<221> misc_feature
<222> (371)..(371)
<223> The 'Xaa' at location 371 stands for Ala, or Val.

<400> 4

Met Ala Ala Arg Asn Val Ser Ala Ala His Glu Met His Asp Glu Lys
1 5 10 15

Arg Ile Ala Val Val Gly Met Ala Val Gln Tyr Ala Gly Cys Lys Thr
 20 25 30

Lys Asp Glu Phe Trp Glu Val Leu Met Asn Gly Lys Val Glu Ser Lys
 35 40 45

Val Ile Ser Asp Lys Arg Leu Gly Ser Asn Tyr Arg Ala Glu His Tyr
 50 55 60

Lys Ala Glu Arg Ser Lys Tyr Ala Asp Thr Phe Cys Asn Glu Thr Tyr
65 70 75 80

Gly Thr Leu Asp Glu Asn Glu Ile Asp Asn Glu His Glu Leu Leu Leu

85

90

95

Asn Leu Ala Lys Gln Ala Leu Ala Glu Thr Ser Val Lys Asp Ser Thr
100 105 110

Arg Cys Gly Ile Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu
115 120 125

Gln Gly Glu Leu Leu Asn Val Tyr Gln Asn His Val Glu Lys Lys Leu
130 135 140

Gly Ala Arg Val Phe Lys Asp Ala Ser His Trp Ser Glu Arg Glu Gln
145 150 155 160

Ser Asn Lys Pro Glu Ala Gly Asp Arg Arg Ile Phe Met Asp Pro Ala
165 170 175

Ser Phe Val Ala Glu Glu Leu Asn Leu Gly Ala Leu His Tyr Ser Val
180 185 190

Asp Ala Ala Cys Ala Thr Ala Leu Tyr Val Leu Arg Leu Ala Gln Asp
195 200 205

His Leu Val Ser Gly Ala Ala Asp Val Met Leu Cys Gly Ala Thr Cys
210 215 220

Leu Pro Glu Pro Phe Phe Ile Leu Ser Gly Phe Ser Thr Phe Gln Ala
225 230 235 240

Met Pro Val Gly Thr Gly Gln Asn Val Ser Met Pro Leu His Lys Asp
245 250 255

Ser Gln Gly Leu Thr Pro Gly Glu Gly Gly Ser Ile Met Val Leu Lys
260 265 270

Arg Leu Asp Asp Ala Ile Arg Asp Gly Asp His Ile Tyr Gly Thr Leu

275

280

285

Leu Gly Ala Asn Val Ser Asn Ser Gly Thr Gly Leu Pro Leu Lys Pro
290 295 300

Leu Leu Pro Ser Glu Lys Lys Cys Leu Met Asp Thr Tyr Thr Arg Ile
305 310 315 320

Asn Val His Pro His Lys Ile Gln Tyr Val Glu Cys His Ala Thr Gly
325 330 335

Thr Pro Gln Gly Asp Arg Val Glu Ile Asp Ala Val Lys Ala Cys Phe
340 345 350

Glu Gly Lys Val Pro Arg Phe Gly Thr Thr Lys Gly Asn Phe Gly His
355 360 365

Thr Xaa Xaa Ala Ala Gly Phe Ala Gly Met Cys Lys Val Leu Leu Ser
370 375 380

Met Lys His Gly Ile Ile Pro Pro Thr Pro Gly Ile Asp Asp Glu Thr
385 390 395 400

Lys Met Asp Pro Leu Val Val Ser Gly Glu Ala Ile Pro Trp Pro Glu
405 410 415

Thr Asn Gly Glu Pro Lys Arg Ala Gly Leu Ser Ala Phe Gly Phe Gly
420 425 430

Gly Thr Asn Ala His Ala Val Phe Glu Glu His Asp Pro Ser Asn Ala
435 440 445

Ala Cys Thr Gly His Asp Ser Ile Ser Ala Leu Ser Ala Arg Cys Gly
450 455 460

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

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

Gly Glu Val Asp Gly Val Val Val Ala Gly Val Asp Leu Cys Gly Ser
675 680 685

Ala Glu Asn Leu Tyr Val Lys Ser Arg Arg Phe Lys Val Ser Thr Ser
690 695 700

Asp Thr Pro Arg Ala Ser Phe Asp Ala Ala Ala Asp Gly Tyr Phe Val
705 710 715 720

Gly Glu Gly Cys Gly Ala Phe Val Leu Lys Arg Glu Thr Ser Cys Thr
725 730 735

Lys Asp Asp Arg Ile Tyr Ala Cys Met Asp Ala Ile Val Pro Gly Asn
740 745 750

Val Pro Ser Ala Cys Leu Arg Glu Ala Leu Asp Gln Ala Arg Val Lys
755 760 765

Pro Gly Asp Ile Glu Met Leu Glu Leu Ser Ala Asp Ser Ala Arg His
770 775 780

Leu Lys Asp Pro Ser Val Leu Pro Lys Glu Leu Thr Ala Glu Glu Glu
785 790 795 800

Ile Gly Gly Leu Gln Thr Ile Leu Arg Asp Asp Asp Lys Leu Pro Arg
805 810 815

Asn Val Ala Thr Gly Ser Val Lys Ala Thr Val Gly Asp Thr Gly Tyr
820 825 830

Ala Ser Gly Ala Ala Ser Leu Ile Lys Ala Ala Leu Cys Ile Tyr Asn
835 840 845

Arg Tyr Leu Pro Ser Asn Gly Asp Asp Trp Asp Glu Pro Ala Pro Glu
850 855 860

Ala Pro Trp Asp Ser Thr Leu Phe Ala Cys Gln Thr Ser Arg Ala Trp
865 870 875 880

Leu Lys Asn Pro Gly Glu Arg Arg Tyr Ala Ala Val Ser Gly Val Ser
885 890 895

Glu Thr Arg Ser Cys Tyr Ser Val Leu Leu Ser Glu Ala Glu Gly His
900 905 910

Tyr Glu Arg Glu Asn Arg Ile Ser Leu Asp Glu Glu Ala Pro Lys Leu
915 920 925

Ile Val Leu Arg Ala Asp Ser His Glu Glu Ile Leu Gly Arg Leu Asp
930 935 940

Lys Ile Arg Glu Arg Phe Leu Gln Pro Thr Gly Ala Ala Pro Arg Glu
945 950 955 960

Ser Glu Leu Lys Ala Gln Ala Arg Arg Ile Phe Leu Glu Leu Leu Gly
965 970 975

Glu Thr Leu Ala Gln Asp Ala Ala Ser Ser Gly Ser Gln Lys Pro Leu
980 985 990

Ala Leu Ser Leu Val Ser Thr Pro Ser Lys Leu Gln Arg Glu Val Glu
995 1000 1005

Leu Ala Ala Lys Gly Ile Pro Arg Cys Leu Lys Met Arg Arg Asp
1010 1015 1020

Trp Ser Ser Pro Ala Gly Ser Arg Tyr Ala Pro Glu Pro Leu Ala
1025 1030 1035

Ser Asp Arg Val Ala Phe Met Tyr Gly Glu Gly Arg Ser Pro Tyr
1040 1045 1050

Tyr Gly Ile Thr Gln Asp Ile His Arg Ile Trp Pro Glu Leu His
1055 1060 1065

Glu Val Ile Asn Glu Lys Thr Asn Arg Leu Trp Ala Glu Gly Asp
1070 1075 1080

Arg Trp Val Met Pro Arg Ala Ser Phe Lys Ser Glu Leu Glu Ser
1085 1090 1095

Gln Gln Gln Glu Phe Asp Arg Asn Met Ile Glu Met Phe Arg Leu
1100 1105 1110

Gly Ile Leu Thr Ser Ile Ala Phe Thr Asn Leu Ala Arg Asp Val
1115 1120 1125

Leu Asn Ile Thr Pro Lys Ala Ala Phe Gly Leu Ser Leu Gly Glu
1130 1135 1140

Ile Ser Met Ile Phe Ala Phe Ser Lys Lys Asn Gly Leu Ile Ser
1145 1150 1155

Asp Gln Leu Thr Lys Asp Leu Arg Glu Ser Asp Val Trp Asn Lys
1160 1165 1170

Ala Leu Ala Val Glu Phe Asn Ala Leu Arg Glu Ala Trp Gly Ile
1175 1180 1185

Pro Gln Ser Val Pro Lys Asp Glu Phe Trp Gln Gly Tyr Ile Val
1190 1195 1200

Arg Gly Thr Lys Gln Asp Ile Glu Ala Ala Ile Ala Pro Asp Ser
1205 1210 1215

Lys Tyr Val Arg Leu Thr Ile Ile Asn Asp Ala Asn Thr Ala Leu

1220

1225

1230

Ile Ser Gly Lys Pro Asp Ala Cys Lys Ala Ala Ile Ala Arg Leu
1235 1240 1245

Gly Gly Asn Ile Pro Ala Leu Pro Val Thr Gln Gly Met Cys Gly
1250 1255 1260

His Cys Pro Glu Val Gly Pro Tyr Thr Lys Asp Ile Ala Lys Ile
1265 1270 1275

His Ala Asn Leu Glu Phe Pro Val Val Asp Gly Leu Asp Leu Trp
1280 1285 1290

Thr Thr Ile Asn Gln Lys Arg Leu Val Pro Arg Ala Thr Gly Ala
1295 1300 1305

Lys Asp Glu Trp Ala Pro Ser Ser Phe Gly Glu Tyr Ala Gly Gln
1310 1315 1320

Leu Tyr Glu Lys Gln Ala Asn Phe Pro Gln Ile Val Glu Thr Ile
1325 1330 1335

Tyr Lys Gln Asn Tyr Asp Val Phe Val Glu Val Gly Pro Asn Asn
1340 1345 1350

His Arg Ser Thr Ala Val Arg Thr Thr Leu Gly Pro Gln Arg Asn
1355 1360 1365

His Leu Ala Gly Ala Ile Asp Lys Gln Asn Glu Asp Ala Trp Thr
1370 1375 1380

Thr Ile Val Lys Leu Val Ala Ser Leu Lys Ala His Leu Val Pro
1385 1390 1395

Gly Val Thr Ile Ser Pro Leu Tyr His Ser Lys Leu Val Ala Glu
1400 1405 1410

Ala Gln Ala Cys Tyr Ala Ala Leu Cys Lys Gly Glu Lys Pro Lys
1415 1420 1425

Lys Asn Lys Phe Val Arg Lys Ile Gln Leu Asn Gly Arg Phe Asn
1430 1435 1440

Ser Lys Ala Asp Pro Ile Ser Ser Ala Asp Leu Ala Ser Phe Pro
1445 1450 1455

Pro Ala Asp Pro Ala Ile Glu Ala Ala Ile Ser Ser Arg Ile Met
1460 1465 1470

Lys Pro Val Ala Pro Lys Phe Tyr Ala Arg Leu Asn Ile Asp Glu
1475 1480 1485

Gln Asp Glu Thr Arg Asp Pro Ile Leu Asn Lys Asp Asn Ala Pro
1490 1495 1500

Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser
1505 1510 1515

Pro Ser Pro Ala Pro Ser Ala Pro Val Gln Lys Lys Ala Ala Pro
1520 1525 1530

Ala Ala Glu Thr Lys Ala Val Ala Ser Ala Asp Ala Leu Arg Ser
1535 1540 1545

Ala Leu Leu Asp Leu Asp Ser Met Leu Ala Leu Ser Ser Ala Ser
1550 1555 1560

Ala Ser Gly Asn Leu Val Glu Thr Ala Pro Ser Asp Ala Ser Val
1565 1570 1575

Ile Val Pro Pro Cys Asn Ile Ala Asp Leu Gly Ser Arg Ala Phe
1580 1585 1590

Met Lys Thr Tyr Gly Val Ser Ala Pro Leu Tyr Thr Gly Ala Met
1595 1600 1605

Ala Lys Gly Ile Ala Ser Ala Asp Leu Val Ile Ala Ala Gly Arg
1610 1615 1620

Gln Gly Ile Leu Ala Ser Phe Gly Ala Gly Gly Leu Pro Met Gln
1625 1630 1635

Val Val Arg Glu Ser Ile Glu Lys Ile Gln Ala Ala Leu Pro Asn
1640 1645 1650

Gly Pro Tyr Ala Val Asn Leu Ile His Ser Pro Phe Asp Ser Asn
1655 1660 1665

Leu Glu Lys Gly Asn Val Asp Leu Phe Leu Glu Lys Gly Val Thr
1670 1675 1680

Phe Val Glu Ala Ser Ala Phe Met Thr Leu Thr Pro Gln Val Val
1685 1690 1695

Arg Tyr Arg Ala Ala Gly Leu Thr Arg Asn Ala Asp Gly Ser Val
1700 1705 1710

Asn Ile Arg Asn Arg Ile Ile Gly Lys Val Ser Arg Thr Glu Leu
1715 1720 1725

Ala Glu Met Phe Met Arg Pro Ala Pro Glu His Leu Leu Gln Lys
1730 1735 1740

Leu Ile Ala Ser Gly Glu Ile Asn Gln Glu Gln Ala Glu Leu Ala

1745 1750 1755

Arg Arg Val Pro Val Ala Asp Asp Ile Ala Val Glu Ala Asp Ser
1760 1765 1770

Gly Gly His Thr Asp Asn Arg Pro Ile His Val Ile Leu Pro Leu
1775 1780 1785

Ile Ile Asn Leu Arg Asp Arg Leu His Arg Glu Cys Gly Tyr Pro
1790 1795 1800

Ala Asn Leu Arg Val Arg Val Gly Ala Gly Gly Gly Ile Gly Cys
1805 1810 1815

Pro Gln Ala Ala Leu Ala Thr Phe Asn Met Gly Ala Ser Phe Ile
1820 1825 1830

Val Thr Gly Thr Val Asn Gln Val Ala Lys Gln Ser Gly Thr Cys
1835 1840 1845

Asp Asn Val Arg Lys Gln Leu Ala Lys Ala Thr Tyr Ser Asp Val
1850 1855 1860

Cys Met Ala Pro Ala Ala Asp Met Phe Glu Glu Gly Val Lys Leu
1865 1870 1875

Gln Val Leu Lys Lys Gly Thr Met Phe Pro Ser Arg Ala Asn Lys
1880 1885 1890

Leu Tyr Glu Leu Phe Cys Lys Tyr Asp Ser Phe Glu Ser Met Pro
1895 1900 1905

Pro Ala Glu Leu Ala Arg Val Glu Lys Arg Ile Phe Ser Arg Ala
1910 1915 1920

Leu Glu Glu Val Trp Asp Glu Thr Lys Asn Phe Tyr Ile Asn Arg
1925 1930 1935

Leu His Asn Pro Glu Lys Ile Gln Arg Ala Glu Arg Asp Pro Lys
1940 1945 1950

Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Ser Leu Ala Ser
1955 1960 1965

Arg Trp Ala Asn Thr Gly Ala Ser Asp Arg Val Met Asp Tyr Gln
1970 1975 1980

Val Trp Cys Gly Pro Ala Ile Gly Ser Phe Asn Asp Phe Ile Lys
1985 1990 1995

Gly Thr Tyr Leu Asp Pro Ala Val Ala Asn Glu Tyr Pro Cys Val
2000 2005 2010

Val Gln Ile Asn Lys Gln Ile Leu Arg Gly Ala Cys Phe Leu Arg
2015 2020 2025

Arg Leu Glu Ile Leu Arg Asn Ala Arg Leu Ser Asp Gly Ala Ala
2030 2035 2040

Ala Leu Val Ala Ser Ile Asp Asp Thr Tyr Val Pro Ala Glu Lys
2045 2050 2055

Leu

<210> 5

<211> 4509

<212> DNA

<213> Schizochytrium sp.

<220>

<221> CDS

<222> (1)..(4509)

<400> 5

atg gcg ctc cgt gtc aag acg aac aag aag cca tgc tgg gag atg acc	48
Met Ala Leu Arg Val Lys Thr Asn Lys Lys Pro Cys Trp Glu Met Thr	
1 5 10 15	

aag gag gag ctg acc agc ggc aag acc gag gtg ttc aac tat gag gaa	96
Lys Glu Glu Leu Thr Ser Gly Lys Thr Glu Val Phe Asn Tyr Glu Glu	
20 25 30	

ctc ctc gag ttc gca gag ggc gac atc gcc aag gtc ttc gga ccc gag	144
Leu Leu Glu Phe Ala Glu Gly Asp Ile Ala Lys Val Phe Gly Pro Glu	
35 40 45	

ttc gcc gtc atc gac aag tac ccg cgc cgc gtg cgc ctg ccc gcc cgc	192
Phe Ala Val Ile Asp Lys Tyr Pro Arg Arg Val Arg Leu Pro Ala Arg	
50 55 60	

gag tac ctg ctc gtg acc cgc gtc acc ctc atg gac gcc gag gtc aac	240
Glu Tyr Leu Leu Val Thr Arg Val Thr Leu Met Asp Ala Glu Val Asn	
65 70 75 80	

aac tac cgc gtc ggc gcc cgc atg gtc acc gag tac gat ctc ccc gtc	288
Asn Tyr Arg Val Gly Ala Arg Met Val Thr Glu Tyr Asp Leu Pro Val	
85 90 95	

aac gga gag ctc tcc gag ggc gga gac tgc ccc tgg gcc gtc ctg gtc	336
Asn Gly Glu Leu Ser Glu Gly Gly Asp Cys Pro Trp Ala Val Leu Val	
100 105 110	

gag agt ggc cag tgc gat ctc atg ctc atc tcc tac atg ggc att gac	384
Glu Ser Gly Gln Cys Asp Leu Met Leu Ile Ser Tyr Met Gly Ile Asp	
115 120 125	

ttc cag aac cag ggc gac cgc gtc tac cgc ctg ctc aac acc acg ctc	432
Phe Gln Asn Gln Gly Asp Arg Val Tyr Arg Leu Leu Asn Thr Thr Leu	
130 135 140	

acc ttt tac ggc gtg gcc cac gag ggc gag acc ctc gag tac gac att	480
Thr Phe Tyr Gly Val Ala His Glu Gly Glu Thr Leu Glu Tyr Asp Ile	
145 150 155 160	

cgc gtc acc ggc ttc gcc aag cgt ctc gac ggc ggc atc tcc atg ttc	528
Arg Val Thr Gly Phe Ala Lys Arg Leu Asp Gly Gly Ile Ser Met Phe	

165	170	175	
ttc ttc gag tac gac tgc tac gtc aac ggc cgc ctc ctc atc gag atg Phe Phe Glu Tyr Asp Cys Tyr Val Asn Gly Arg Leu Leu Ile Glu Met 180 185 190			576
cgc gat ggc tgc gcc ggc ttc ttc acc aac gag gag ctc gac gcc ggc Arg Asp Gly Cys Ala Gly Phe Phe Thr Asn Glu Glu Leu Asp Ala Gly 195 200 205			624
aag ggc gtc gtc ttc acc cgc ggc gac ctc gcc gcc cgc gcc aag atc Lys Gly Val Val Phe Thr Arg Gly Asp Leu Ala Ala Arg Ala Lys Ile 210 215 220			672
cca aag cag gac gtc tcc ccc tac gcc gtc gcc ccc tgc ctc cac aag Pro Lys Gln Asp Val Ser Pro Tyr Ala Val Ala Pro Cys Leu His Lys 225 230 235 240			720
acc aag ctc aac gaa aag gag atg cag acc ctc gtc gac aag gac tgg Thr Lys Leu Asn Glu Lys Glu Met Gln Thr Leu Val Asp Lys Asp Trp 245 250 255			768
gca tcc gtc ttt ggc tcc aag aac ggc atg ccg gaa atc aac tac aaa Ala Ser Val Phe Gly Ser Lys Asn Gly Met Pro Glu Ile Asn Tyr Lys 260 265 270			816
ctc tgc gcg cgt aag atg ctc atg att gac cgc gtc acc agc att gac Leu Cys Ala Arg Lys Met Leu Met Ile Asp Arg Val Thr Ser Ile Asp 275 280 285			864
cac aag ggc ggt gtc tac ggc ctc ggt cag ctc gtc ggt gaa aag atc His Lys Gly Gly Val Tyr Gly Leu Gly Gln Leu Val Gly Glu Lys Ile 290 295 300			912
ctc gag cgc gac cac tgg tac ttt ccc tgc cac ttt gtc aag gat cag Leu Glu Arg Asp His Trp Tyr Phe Pro Cys His Phe Val Lys Asp Gln 305 310 315 320			960
gtc atg gcc gga tcc ctc gtc tcc gac ggc tgc agc cag atg ctc aag Val Met Ala Gly Ser Leu Val Ser Asp Gly Cys Ser Gln Met Leu Lys 325 330 335			1008
atg tac atg atc tgg ctc ggc ctc cac ctc acc acc gga ccc ttt gac Met Tyr Met Ile Trp Leu Gly Leu His Leu Thr Thr Gly Pro Phe Asp			1056

340	345	350	
ttc cgc ccg gtc aac ggc cac ccc aac aag gtc cgc tgc cgc ggc caa Phe Arg Pro Val Asn Gly His Pro Asn Lys Val Arg Cys Arg Gly Gln 355 360 365			1104
atc tcc ccg cac aag ggc aag ctc gtc tac gtc atg gag atc aag gag Ile Ser Pro His Lys Gly Lys Leu Val Tyr Val Met Glu Ile Lys Glu 370 375 380			1152
atg ggc ttc gac gag gac aac gac ccg tac gcc att gcc gac gtc aac Met Gly Phe Asp Glu Asp Asn Asp Pro Tyr Ala Ile Ala Asp Val Asn 385 390 395 400			1200
atc att gat gtc gac ttc gaa aag ggc cag gac ttt agc ctc gac cgc Ile Ile Asp Val Asp Phe Glu Lys Gly Gln Asp Phe Ser Leu Asp Arg 405 410 415			1248
atc agc gac tac ggc aag ggc gac ctc aac aag aag atc gtc gtc gac Ile Ser Asp Tyr Gly Lys Gly Asp Leu Asn Lys Lys Ile Val Val Asp 420 425 430			1296
ttt aag ggc atc gct ctc aag atg cag aag cgc tcc acc aac aag aac Phe Lys Gly Ile Ala Leu Lys Met Gln Lys Arg Ser Thr Asn Lys Asn 435 440 445			1344
ccc tcc aag gtt cag ccc gtc ttt gcc aac ggc gcc gcc act gtc ggc Pro Ser Lys Val Gln Pro Val Phe Ala Asn Gly Ala Ala Thr Val Gly 450 455 460			1392
ccc gag gcc tcc aag gct tcc tcc ggc gcc agc gcc agc gcc agc gcc Pro Glu Ala Ser Lys Ala Ser Ser Gly Ala Ser Ala Ser Ala Ser Ala 465 470 475 480			1440
gcc ccg gcc aag cct gcc ttc agc gcc gat gtt ctt gcg ccc aag ccc Ala Pro Ala Lys Pro Ala Phe Ser Ala Asp Val Leu Ala Pro Lys Pro 485 490 495			1488
gtt gcc ctt ccc gag cac atc ctc aag ggc gac gcc ctc gcc ccc aag Val Ala Leu Pro Glu His Ile Leu Lys Gly Asp Ala Leu Ala Pro Lys 500 505 510			1536
gag atg tcc tgg cac ccc atg gcc cgc atc ccg ggc aac ccg acg ccc Glu Met Ser Trp His Pro Met Ala Arg Ile Pro Gly Asn Pro Thr Pro			1584

515	520	525	
tct ttt gcg ccc tcg gcc tac aag ccg cgc aac atc gcc ttt acg ccc			1632
Ser Phe Ala Pro Ser Ala Tyr Lys Pro Arg Asn Ile Ala Phe Thr Pro			
530	535	540	
ttc ccc ggc aac ccc aac gat aac gac cac acc ccg ggc aag atg ccg			1680
Phe Pro Gly Asn Pro Asn Asp Asn Asp His Thr Pro Gly Lys Met Pro			
545	550	555	560
ctc acc tgg ttc aac atg gcc gag ttc atg gcc ggc aag gtc agc atg			1728
Leu Thr Trp Phe Asn Met Ala Glu Phe Met Ala Gly Lys Val Ser Met			
565	570	575	
tgc ctc ggc ccc gag ttc gcc aag ttc gac gac tcg aac acc agc cgc			1776
Cys Leu Gly Pro Glu Phe Ala Lys Phe Asp Asp Ser Asn Thr Ser Arg			
580	585	590	
agc ccc gct tgg gac ctc gct ctc gtc acc cgc gcc gtg tct gtg tct			1824
Ser Pro Ala Trp Asp Leu Ala Leu Val Thr Arg Ala Val Ser Val Ser			
595	600	605	
gac ctc aag cac gtc aac tac cgc aac atc gac ctc gac ccc tcc aag			1872
Asp Leu Lys His Val Asn Tyr Arg Asn Ile Asp Leu Asp Pro Ser Lys			
610	615	620	
ggc acc atg gtc ggc gag ttc gac tgc ccc gcg gac gcc tgg ttc tac			1920
Gly Thr Met Val Gly Glu Phe Asp Cys Pro Ala Asp Ala Trp Phe Tyr			
625	630	635	640
aag ggc gcc tgc aac gat gcc cac atg ccg tac tcg atc ctc atg gag			1968
Lys Gly Ala Cys Asn Asp Ala His Met Pro Tyr Ser Ile Leu Met Glu			
645	650	655	
atc gcc ctc cag acc tcg ggt gtg ctc acc tcg gtg ctc aag gcg ccc			2016
Ile Ala Leu Gln Thr Ser Gly Val Leu Thr Ser Val Leu Lys Ala Pro			
660	665	670	
ctg acc atg gag aag gac gac atc ctc ttc cgc aac ctc gac gcc aac			2064
Leu Thr Met Glu Lys Asp Asp Ile Leu Phe Arg Asn Leu Asp Ala Asn			
675	680	685	
gcc gag ttc gtg cgc gcc gac ctc gac tac cgc ggc aag act atc cgc			2112
Ala Glu Phe Val Arg Ala Asp Leu Asp Tyr Arg Gly Lys Thr Ile Arg			

690	695	700	
aac gtc acc aag tgc act ggc tac agc atg ctc ggc gag atg ggc gtc			2160
Asn Val Thr Lys Cys Thr Gly Tyr Ser Met Leu Gly Glu Met Gly Val			
705	710	715	720
cac cgc ttc acc ttt gag ctc tac gtc gat gat gtg ctc ttt tac aag			2208
His Arg Phe Thr Phe Glu Leu Tyr Val Asp Asp Val Leu Phe Tyr Lys			
	725	730	735
ggc tcg acc tcg ttc ggc tgg ttc gtg ccc gag gtc ttt gcc gcc cag			2256
Gly Ser Thr Ser Phe Gly Trp Phe Val Pro Glu Val Phe Ala Ala Gln			
	740	745	750
gcc ggc ctc gac aac ggc cgc aag tcg gag ccc tgg ttc att gag aac			2304
Ala Gly Leu Asp Asn Gly Arg Lys Ser Glu Pro Trp Phe Ile Glu Asn			
	755	760	765
aag gtt ccg gcc tcg cag gtc tcc tcc ttt gac gtg cgc ccc aac ggc			2352
Lys Val Pro Ala Ser Gln Val Ser Ser Phe Asp Val Arg Pro Asn Gly			
	770	775	780
agc ggc cgc acc gcc atc ttc gcc aac gcc ccc agc ggc gcc cag ctc			2400
Ser Gly Arg Thr Ala Ile Phe Ala Asn Ala Pro Ser Gly Ala Gln Leu			
	785	790	800
aac cgc cgc acg gac cag ggc cag tac ctc gac gcc gtc gac att gtc			2448
Asn Arg Arg Thr Asp Gln Gly Gln Tyr Leu Asp Ala Val Asp Ile Val			
	805	810	815
tcc ggc agc ggc aag aag agc ctc ggc tac gcc cac ggt tcc aag acg			2496
Ser Gly Ser Gly Lys Lys Ser Leu Gly Tyr Ala His Gly Ser Lys Thr			
	820	825	830
gtc aac ccg aac gac tgg ttc ttc tcg tgc cac ttt tgg ttt gac tcg			2544
Val Asn Pro Asn Asp Trp Phe Phe Ser Cys His Phe Trp Phe Asp Ser			
	835	840	845
gtc atg ccc gga agt ctc ggt gtc gag tcc atg ttc cag ctc gtc gag			2592
Val Met Pro Gly Ser Leu Gly Val Glu Ser Met Phe Gln Leu Val Glu			
	850	855	860
gcc atc gcc gcc cac gag gat ctc gct ggc aaa gca cgg cat tgc caa			2640
Ala Ile Ala Ala His Glu Asp Leu Ala Gly Lys Ala Arg His Cys Gln			

865	870	875	880	
ccc cac ctt tgt gca cgc ccc cgg gca aga tca agc tgg aag tac cgc				2688
Pro His Leu Cys Ala Arg Pro Arg Ala Arg Ser Ser Trp Lys Tyr Arg				
885	890	895		
ggc cag ctc acg ccc aag agc aag aag atg gac tcg gag gtc cac atc				2736
Gly Gln Leu Thr Pro Lys Ser Lys Lys Met Asp Ser Glu Val His Ile				
900	905	910		
gtg tcc gtg gac gcc cac gac ggc gtt gtc gac ctc gtc gcc gac ggc				2784
Val Ser Val Asp Ala His Asp Gly Val Val Asp Leu Val Ala Asp Gly				
915	920	925		
ttc ctc tgg gcc gac agc ctc cgc gtc tac tcg gtg agc aac att cgc				2832
Phe Leu Trp Ala Asp Ser Leu Arg Val Tyr Ser Val Ser Asn Ile Arg				
930	935	940		
gtg cgc atc gcc tcc ggt gag gcc cct gcc gcc gcc tcc tcc gcc gcc				2880
Val Arg Ile Ala Ser Gly Glu Ala Pro Ala Ala Ala Ser Ser Ala Ala				
945	950	955	960	
tct gtg ggc tcc tcg gct tcg tcc gtc gag cgc acg cgc tcg agc ccc				2928
Ser Val Gly Ser Ser Ala Ser Ser Val Glu Arg Thr Arg Ser Ser Pro				
965	970	975		
gct gtc gcc tcc ggc ccg gcc cag acc atc gac ctc aag cag ctc aag				2976
Ala Val Ala Ser Gly Pro Ala Gln Thr Ile Asp Leu Lys Gln Leu Lys				
980	985	990		
acc gag ctc ctc gag ctc gat gcc ccg ctc tac ctc tcg cag gac ccg				3024
Thr Glu Leu Leu Glu Leu Asp Ala Pro Leu Tyr Leu Ser Gln Asp Pro				
995	1000	1005		
acc agc ggc cag ctc aag aag cac acc gac gtg gcc tcc ggc cag				3069
Thr Ser Gly Gln Leu Lys Lys His Thr Asp Val Ala Ser Gly Gln				
1010	1015	1020		
gcc acc atc gtg cag ccc tgc acg ctc ggc gac ctc ggt gac cgc				3114
Ala Thr Ile Val Gln Pro Cys Thr Leu Gly Asp Leu Gly Asp Arg				
1025	1030	1035		
tcc ttc atg gag acc tac ggc gtc gtc gcc ccg ctg tac acg ggc				3159
Ser Phe Met Glu Thr Tyr Gly Val Val Ala Pro Leu Tyr Thr Gly				

1040	1045	1050	
gcc atg gcc aag ggc att gcc tcg gcg gac ctc gtc atc gcc gcc			3204
Ala Met Ala Lys Gly Ile Ala Ser Ala Asp Leu Val Ile Ala Ala			
1055	1060	1065	
ggc aag cgc aag atc ctc ggc tcc ttt ggc gcc ggc ggc ctc ccc			3249
Gly Lys Arg Lys Ile Leu Gly Ser Phe Gly Ala Gly Gly Leu Pro			
1070	1075	1080	
atg cac cac gtg cgc gcc gcc ctc gag aag atc cag gcc gcc ctg			3294
Met His His Val Arg Ala Ala Leu Glu Lys Ile Gln Ala Ala Leu			
1085	1090	1095	
cct cag ggc ccc tac gcc gtc aac ctc atc cac tcg cct ttt gac			3339
Pro Gln Gly Pro Tyr Ala Val Asn Leu Ile His Ser Pro Phe Asp			
1100	1105	1110	
agc aac ctc gag aag ggc aac gtc gat ctc ttc ctc gag aag ggc			3384
Ser Asn Leu Glu Lys Gly Asn Val Asp Leu Phe Leu Glu Lys Gly			
1115	1120	1125	
gtc act gtg gtg gag gcc tcg gca ttc atg acc ctc acc ccg cag			3429
Val Thr Val Val Glu Ala Ser Ala Phe Met Thr Leu Thr Pro Gln			
1130	1135	1140	
gtc gtg cgc tac cgc gcc gcc ggc ctc tcg cgc aac gcc gac ggt			3474
Val Val Arg Tyr Arg Ala Ala Gly Leu Ser Arg Asn Ala Asp Gly			
1145	1150	1155	
tcg gtc aac atc cgc aac cgc atc atc ggc aag gtc tcg cgc acc			3519
Ser Val Asn Ile Arg Asn Arg Ile Ile Gly Lys Val Ser Arg Thr			
1160	1165	1170	
gag ctc gcc gag atg ttc atc cgc ccg gcc ccg gag cac ctc ctc			3564
Glu Leu Ala Glu Met Phe Ile Arg Pro Ala Pro Glu His Leu Leu			
1175	1180	1185	
gag aag ctc atc gcc tcg ggc gag atc acc cag gag cag gcc gag			3609
Glu Lys Leu Ile Ala Ser Gly Glu Ile Thr Gln Glu Gln Ala Glu			
1190	1195	1200	
ctc gcg cgc cgc gtt ccc gtc gcc gac gat atc gct gtc gag gct			3654
Leu Ala Arg Arg Val Pro Val Ala Asp Asp Ile Ala Val Glu Ala			

1205	1210	1215	
gac tcg ggc ggc cac acc gac aac cgc ccc atc cac gtc atc ctc			3699
Asp Ser Gly Gly His Thr Asp Asn Arg Pro Ile His Val Ile Leu			
1220	1225	1230	
ccg ctc atc atc aac ctc cgc aac cgc ctg cac cgc gag tgc ggc			3744
Pro Leu Ile Ile Asn Leu Arg Asn Arg Leu His Arg Glu Cys Gly			
1235	1240	1245	
tac ccc gcg cac ctc cgc gtc cgc gtt ggc gcc ggc ggt ggc gtc			3789
Tyr Pro Ala His Leu Arg Val Arg Val Gly Ala Gly Gly Gly Val			
1250	1255	1260	
ggc tgc ccg cag gcc gcc gcc gcc gcg ctc acc atg ggc gcc gcc			3834
Gly Cys Pro Gln Ala Ala Ala Ala Ala Leu Thr Met Gly Ala Ala			
1265	1270	1275	
ttc atc gtc acc ggc act gtc aac cag gtc gcc aag cag tcc ggc			3879
Phe Ile Val Thr Gly Thr Val Asn Gln Val Ala Lys Gln Ser Gly			
1280	1285	1290	
acc tgc gac aac gtg cgc aag cag ctc tcg cag gcc acc tac tcg			3924
Thr Cys Asp Asn Val Arg Lys Gln Leu Ser Gln Ala Thr Tyr Ser			
1295	1300	1305	
gat atc tgc atg gcc ccg gcc gcc gac atg ttc gag gag ggc gtc			3969
Asp Ile Cys Met Ala Pro Ala Ala Asp Met Phe Glu Glu Gly Val			
1310	1315	1320	
aag ctc cag gtc ctc aag aag gga acc atg ttc ccc tcg cgc gcc			4014
Lys Leu Gln Val Leu Lys Lys Gly Thr Met Phe Pro Ser Arg Ala			
1325	1330	1335	
aac aag ctc tac gag ctc ttt tgc aag tac gac tcc ttc gac tcc			4059
Asn Lys Leu Tyr Glu Leu Phe Cys Lys Tyr Asp Ser Phe Asp Ser			
1340	1345	1350	
atg cct cct gcc gag ctc gag cgc atc gag aag cgt atc ttc aag			4104
Met Pro Pro Ala Glu Leu Glu Arg Ile Glu Lys Arg Ile Phe Lys			
1355	1360	1365	
cgc gca ctc cag gag gtc tgg gag gag acc aag gac ttt tac att			4149
Arg Ala Leu Gln Glu Val Trp Glu Glu Thr Lys Asp Phe Tyr Ile			

1370	1375	1380	
aac ggt ctc aag aac ccg gag	aag atc cag cgc gcc	gag cac gac	4194
Asn Gly Leu Lys Asn Pro Glu	Lys Ile Gln Arg Ala	Glu His Asp	
1385	1390	1395	
ccc aag ctc aag atg tcg ctc	tgc ttc cgc tgg tac	ctt ggt ctt	4239
Pro Lys Leu Lys Met Ser Leu	Cys Phe Arg Trp Tyr	Leu Gly Leu	
1400	1405	1410	
gcc agc cgc tgg gcc aac atg	ggc gcc ccg gac cgc	gtc atg gac	4284
Ala Ser Arg Trp Ala Asn Met	Gly Ala Pro Asp Arg	Val Met Asp	
1415	1420	1425	
tac cag gtc tgg tgt ggc ccg	gcc att ggc gcc ttc	aac gac ttc	4329
Tyr Gln Val Trp Cys Gly Pro	Ala Ile Gly Ala Phe	Asn Asp Phe	
1430	1435	1440	
atc aag ggc acc tac ctc gac	ccc gct gtc tcc aac	gag tac ccc	4374
Ile Lys Gly Thr Tyr Leu Asp	Pro Ala Val Ser Asn	Glu Tyr Pro	
1445	1450	1455	
tgt gtc gtc cag atc aac ctg	caa atc ctc cgt ggt	gcc tgc tac	4419
Cys Val Val Gln Ile Asn Leu	Gln Ile Leu Arg Gly	Ala Cys Tyr	
1460	1465	1470	
ctg cgc cgt ctc aac gcc ctg	cgc aac gac ccg cgc	att gac ctc	4464
Leu Arg Arg Leu Asn Ala Leu	Arg Asn Asp Pro Arg	Ile Asp Leu	
1475	1480	1485	
gag acc gag gat gct gcc ttt	gtc tac gag ccc acc	aac gcg ctc	4509
Glu Thr Glu Asp Ala Ala Phe	Val Tyr Glu Pro Thr	Asn Ala Leu	
1490	1495	1500	

<210> 6
 <211> 1503
 <212> PRT
 <213> Schizochytrium sp.

<400> 6

Met	Ala	Leu	Arg	Val	Lys	Thr	Asn	Lys	Lys	Pro	Cys	Trp	Glu	Met	Thr
1				5						10				15	

Lys Glu Glu Leu Thr Ser Gly Lys Thr Glu Val Phe Asn Tyr Glu Glu
20 25 30

Leu Leu Glu Phe Ala Glu Gly Asp Ile Ala Lys Val Phe Gly Pro Glu
35 40 45

Phe Ala Val Ile Asp Lys Tyr Pro Arg Arg Val Arg Leu Pro Ala Arg
50 55 60

Glu Tyr Leu Leu Val Thr Arg Val Thr Leu Met Asp Ala Glu Val Asn
65 70 75 80

Asn Tyr Arg Val Gly Ala Arg Met Val Thr Glu Tyr Asp Leu Pro Val
85 90 95

Asn Gly Glu Leu Ser Glu Gly Gly Asp Cys Pro Trp Ala Val Leu Val
100 105 110

Glu Ser Gly Gln Cys Asp Leu Met Leu Ile Ser Tyr Met Gly Ile Asp
115 120 125

Phe Gln Asn Gln Gly Asp Arg Val Tyr Arg Leu Leu Asn Thr Thr Leu
130 135 140

Thr Phe Tyr Gly Val Ala His Glu Gly Glu Thr Leu Glu Tyr Asp Ile
145 150 155 160

Arg Val Thr Gly Phe Ala Lys Arg Leu Asp Gly Gly Ile Ser Met Phe
165 170 175

Phe Phe Glu Tyr Asp Cys Tyr Val Asn Gly Arg Leu Leu Ile Glu Met
180 185 190

Arg Asp Gly Cys Ala Gly Phe Phe Thr Asn Glu Glu Leu Asp Ala Gly

195

200

205

Lys Gly Val Val Phe Thr Arg Gly Asp Leu Ala Ala Arg Ala Lys Ile
210 215 220

Pro Lys Gln Asp Val Ser Pro Tyr Ala Val Ala Pro Cys Leu His Lys
225 230 235 240

Thr Lys Leu Asn Glu Lys Glu Met Gln Thr Leu Val Asp Lys Asp Trp
245 250 255

Ala Ser Val Phe Gly Ser Lys Asn Gly Met Pro Glu Ile Asn Tyr Lys
260 265 270

Leu Cys Ala Arg Lys Met Leu Met Ile Asp Arg Val Thr Ser Ile Asp
275 280 285

His Lys Gly Gly Val Tyr Gly Leu Gly Gln Leu Val Gly Glu Lys Ile
290 295 300

Leu Glu Arg Asp His Trp Tyr Phe Pro Cys His Phe Val Lys Asp Gln
305 310 315 320

Val Met Ala Gly Ser Leu Val Ser Asp Gly Cys Ser Gln Met Leu Lys
325 330 335

Met Tyr Met Ile Trp Leu Gly Leu His Leu Thr Thr Gly Pro Phe Asp
340 345 350

Phe Arg Pro Val Asn Gly His Pro Asn Lys Val Arg Cys Arg Gly Gln
355 360 365

Ile Ser Pro His Lys Gly Lys Leu Val Tyr Val Met Glu Ile Lys Glu
370 375 380

Met Gly Phe Asp Glu Asp Asn Asp Pro Tyr Ala Ile Ala Asp Val Asn

385	390	395	400
Ile Ile Asp Val	Asp Phe Glu Lys Gly Gln Asp Phe Ser Leu Asp Arg		
	405	410	415
Ile Ser Asp Tyr	Gly Lys Gly Asp Leu Asn Lys Lys Ile Val Val Asp		
	420	425	430
Phe Lys Gly Ile	Ala Leu Lys Met Gln Lys Arg Ser Thr Asn Lys Asn		
	435	440	445
Pro Ser Lys Val	Gln Pro Val Phe Ala Asn Gly Ala Ala Thr Val Gly		
	450	455	460
Pro Glu Ala Ser	Lys Ala Ser Ser Gly Ala Ser Ala Ser Ala Ser Ala		
	465	470	475
Ala Pro Ala Lys	Pro Ala Phe Ser Ala Asp Val Leu Ala Pro Lys Pro		
	485	490	495
Val Ala Leu Pro	Glu His Ile Leu Lys Gly Asp Ala Leu Ala Pro Lys		
	500	505	510
Glu Met Ser Trp	His Pro Met Ala Arg Ile Pro Gly Asn Pro Thr Pro		
	515	520	525
Ser Phe Ala Pro	Ser Ala Tyr Lys Pro Arg Asn Ile Ala Phe Thr Pro		
	530	535	540
Phe Pro Gly Asn	Pro Asn Asp Asn Asp His Thr Pro Gly Lys Met Pro		
	545	550	555
Leu Thr Trp Phe	Asn Met Ala Glu Phe Met Ala Gly Lys Val Ser Met		
	565	570	575
Cys Leu Gly Pro	Glu Phe Ala Lys Phe Asp Asp Ser Asn Thr Ser Arg		
	580	585	590

Ser Pro Ala Trp Asp Leu Ala Leu Val Thr Arg Ala Val Ser Val Ser
595 600 605

Asp Leu Lys His Val Asn Tyr Arg Asn Ile Asp Leu Asp Pro Ser Lys
610 615 620

Gly Thr Met Val Gly Glu Phe Asp Cys Pro Ala Asp Ala Trp Phe Tyr
625 630 635 640

Lys Gly Ala Cys Asn Asp Ala His Met Pro Tyr Ser Ile Leu Met Glu
645 650 655

Ile Ala Leu Gln Thr Ser Gly Val Leu Thr Ser Val Leu Lys Ala Pro
660 665 670

Leu Thr Met Glu Lys Asp Asp Ile Leu Phe Arg Asn Leu Asp Ala Asn
675 680 685

Ala Glu Phe Val Arg Ala Asp Leu Asp Tyr Arg Gly Lys Thr Ile Arg
690 695 700

Asn Val Thr Lys Cys Thr Gly Tyr Ser Met Leu Gly Glu Met Gly Val
705 710 715 720

His Arg Phe Thr Phe Glu Leu Tyr Val Asp Asp Val Leu Phe Tyr Lys
725 730 735

Gly Ser Thr Ser Phe Gly Trp Phe Val Pro Glu Val Phe Ala Ala Gln
740 745 750

Ala Gly Leu Asp Asn Gly Arg Lys Ser Glu Pro Trp Phe Ile Glu Asn
755 760 765

Lys Val Pro Ala Ser Gln Val Ser Ser Phe Asp Val Arg Pro Asn Gly
770 775 780

Ser Gly Arg Thr Ala Ile Phe Ala Asn Ala Pro Ser Gly Ala Gln Leu
785 790 795 800

Asn Arg Arg Thr Asp Gln Gly Gln Tyr Leu Asp Ala Val Asp Ile Val
805 810 815

Ser Gly Ser Gly Lys Lys Ser Leu Gly Tyr Ala His Gly Ser Lys Thr
820 825 830

Val Asn Pro Asn Asp Trp Phe Phe Ser Cys His Phe Trp Phe Asp Ser
835 840 845

Val Met Pro Gly Ser Leu Gly Val Glu Ser Met Phe Gln Leu Val Glu
850 855 860

Ala Ile Ala Ala His Glu Asp Leu Ala Gly Lys Ala Arg His Cys Gln
865 870 875 880

Pro His Leu Cys Ala Arg Pro Arg Ala Arg Ser Ser Trp Lys Tyr Arg
885 890 895

Gly Gln Leu Thr Pro Lys Ser Lys Lys Met Asp Ser Glu Val His Ile
900 905 910

Val Ser Val Asp Ala His Asp Gly Val Val Asp Leu Val Ala Asp Gly
915 920 925

Phe Leu Trp Ala Asp Ser Leu Arg Val Tyr Ser Val Ser Asn Ile Arg
930 935 940

Val Arg Ile Ala Ser Gly Glu Ala Pro Ala Ala Ala Ser Ser Ala Ala
945 950 955 960

Ser Val Gly Ser Ser Ala Ser Ser Val Glu Arg Thr Arg Ser Ser Pro
965 970 975

Ala Val Ala Ser Gly Pro Ala Gln Thr Ile Asp Leu Lys Gln Leu Lys
980 985 990

Thr Glu Leu Leu Glu Leu Asp Ala Pro Leu Tyr Leu Ser Gln Asp Pro
995 1000 1005

Thr Ser Gly Gln Leu Lys Lys His Thr Asp Val Ala Ser Gly Gln
1010 1015 1020

Ala Thr Ile Val Gln Pro Cys Thr Leu Gly Asp Leu Gly Asp Arg
1025 1030 1035

Ser Phe Met Glu Thr Tyr Gly Val Val Ala Pro Leu Tyr Thr Gly
1040 1045 1050

Ala Met Ala Lys Gly Ile Ala Ser Ala Asp Leu Val Ile Ala Ala
1055 1060 1065

Gly Lys Arg Lys Ile Leu Gly Ser Phe Gly Ala Gly Gly Leu Pro
1070 1075 1080

Met His His Val Arg Ala Ala Leu Glu Lys Ile Gln Ala Ala Leu
1085 1090 1095

Pro Gln Gly Pro Tyr Ala Val Asn Leu Ile His Ser Pro Phe Asp
1100 1105 1110

Ser Asn Leu Glu Lys Gly Asn Val Asp Leu Phe Leu Glu Lys Gly
1115 1120 1125

Val Thr Val Val Glu Ala Ser Ala Phe Met Thr Leu Thr Pro Gln
1130 1135 1140

Val Val Arg Tyr Arg Ala Ala Gly Leu Ser Arg Asn Ala Asp Gly

1145

1150

1155

Ser Val Asn Ile Arg Asn Arg Ile Ile Gly Lys Val Ser Arg Thr
1160 1165 1170

Glu Leu Ala Glu Met Phe Ile Arg Pro Ala Pro Glu His Leu Leu
1175 1180 1185

Glu Lys Leu Ile Ala Ser Gly Glu Ile Thr Gln Glu Gln Ala Glu
1190 1195 1200

Leu Ala Arg Arg Val Pro Val Ala Asp Asp Ile Ala Val Glu Ala
1205 1210 1215

Asp Ser Gly Gly His Thr Asp Asn Arg Pro Ile His Val Ile Leu
1220 1225 1230

Pro Leu Ile Ile Asn Leu Arg Asn Arg Leu His Arg Glu Cys Gly
1235 1240 1245

Tyr Pro Ala His Leu Arg Val Arg Val Gly Ala Gly Gly Gly Val
1250 1255 1260

Gly Cys Pro Gln Ala Ala Ala Ala Ala Leu Thr Met Gly Ala Ala
1265 1270 1275

Phe Ile Val Thr Gly Thr Val Asn Gln Val Ala Lys Gln Ser Gly
1280 1285 1290

Thr Cys Asp Asn Val Arg Lys Gln Leu Ser Gln Ala Thr Tyr Ser
1295 1300 1305

Asp Ile Cys Met Ala Pro Ala Ala Asp Met Phe Glu Glu Gly Val
1310 1315 1320

Lys Leu Gln Val Leu Lys Lys Gly Thr Met Phe Pro Ser Arg Ala
1325 1330 1335

Asn Lys Leu Tyr Glu Leu Phe Cys Lys Tyr Asp Ser Phe Asp Ser
1340 1345 1350

Met Pro Pro Ala Glu Leu Glu Arg Ile Glu Lys Arg Ile Phe Lys
1355 1360 1365

Arg Ala Leu Gln Glu Val Trp Glu Glu Thr Lys Asp Phe Tyr Ile
1370 1375 1380

Asn Gly Leu Lys Asn Pro Glu Lys Ile Gln Arg Ala Glu His Asp
1385 1390 1395

Pro Lys Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Gly Leu
1400 1405 1410

Ala Ser Arg Trp Ala Asn Met Gly Ala Pro Asp Arg Val Met Asp
1415 1420 1425

Tyr Gln Val Trp Cys Gly Pro Ala Ile Gly Ala Phe Asn Asp Phe
1430 1435 1440

Ile Lys Gly Thr Tyr Leu Asp Pro Ala Val Ser Asn Glu Tyr Pro
1445 1450 1455

Cys Val Val Gln Ile Asn Leu Gln Ile Leu Arg Gly Ala Cys Tyr
1460 1465 1470

Leu Arg Arg Leu Asn Ala Leu Arg Asn Asp Pro Arg Ile Asp Leu
1475 1480 1485

Glu Thr Glu Asp Ala Ala Phe Val Tyr Glu Pro Thr Asn Ala Leu
1490 1495 1500

<210> 7
 <211> 600
 <212> DNA
 <213> Schizochytrium sp.

<220>
 <221> CDS
 <222> (1)..(600)

<400> 7
 atg gcg gcc cgt ctg cag gag caa aag gga ggc gag atg gat acc cgc 48
 Met Ala Ala Arg Leu Gln Glu Gln Lys Gly Gly Glu Met Asp Thr Arg
 1 5 10 15

att gcc atc atc ggc atg tcg gcc atc ctc ccc tgc ggc acg acc gtg 96
 Ile Ala Ile Ile Gly Met Ser Ala Ile Leu Pro Cys Gly Thr Thr Val
 20 25 30

cgc gag tcg tgg gag acc atc cgc gcc ggc atc gac tgc ctg tcg gat 144
 Arg Glu Ser Trp Glu Thr Ile Arg Ala Gly Ile Asp Cys Leu Ser Asp
 35 40 45

ctc ccc gag gac cgc gtc gac gtg acg gcg tac ttt gac ccc gtc aag 192
 Leu Pro Glu Asp Arg Val Asp Val Thr Ala Tyr Phe Asp Pro Val Lys
 50 55 60

acc acc aag gac aag atc tac tgc aag cgc ggt ggc ttc att ccc gag 240
 Thr Thr Lys Asp Lys Ile Tyr Cys Lys Arg Gly Gly Phe Ile Pro Glu
 65 70 75 80

tac gac ttt gac gcc cgc gag ttc gga ctc aac atg ttc cag atg gag 288
 Tyr Asp Phe Asp Ala Arg Glu Phe Gly Leu Asn Met Phe Gln Met Glu
 85 90 95

gac tcg gac gca aac cag acc atc tcg ctt ctc aag gtc aag gag gcc 336
 Asp Ser Asp Ala Asn Gln Thr Ile Ser Leu Leu Lys Val Lys Glu Ala
 100 105 110

ctc cag gac gcc ggc atc gac gcc ctc ggc aag gaa aag aag aac atc 384
 Leu Gln Asp Ala Gly Ile Asp Ala Leu Gly Lys Glu Lys Lys Asn Ile
 115 120 125

ggc tgc gtg ctc ggc att ggc ggc ggc caa aag tcc agc cac gag ttc 432

Gly Cys Val Leu Gly Ile Gly Gly Gly Gln Lys Ser Ser His Glu Phe
130 135 140

tac tcg cgc ctt aat tat gtt gtc gtg gag aag gtc ctc cgc aag atg 480
Tyr Ser Arg Leu Asn Tyr Val Val Val Glu Lys Val Leu Arg Lys Met
145 150 155 160

ggc atg ccc gag gag gac gtc aag gtc gcc gtc gaa aag tac aag gcc 528
Gly Met Pro Glu Glu Asp Val Lys Val Ala Val Glu Lys Tyr Lys Ala
165 170 175

aac ttc ccc gag tgg cgc ctc gac tcc ttc cct ggc ttc ctc ggc aac 576
Asn Phe Pro Glu Trp Arg Leu Asp Ser Phe Pro Gly Phe Leu Gly Asn
180 185 190

gtc acc gcc ggt cgc tgc acc aac 600
Val Thr Ala Gly Arg Cys Thr Asn
195 200

<210> 8
<211> 200
<212> PRT
<213> Schizochytrium sp.

<400> 8

Met Ala Ala Arg Leu Gln Glu Gln Lys Gly Gly Glu Met Asp Thr Arg
1 5 10 15

Ile Ala Ile Ile Gly Met Ser Ala Ile Leu Pro Cys Gly Thr Thr Val
20 25 30

Arg Glu Ser Trp Glu Thr Ile Arg Ala Gly Ile Asp Cys Leu Ser Asp
35 40 45

Leu Pro Glu Asp Arg Val Asp Val Thr Ala Tyr Phe Asp Pro Val Lys
50 55 60

Thr Thr Lys Asp Lys Ile Tyr Cys Lys Arg Gly Gly Phe Ile Pro Glu
65 70 75 80

Tyr Asp Phe Asp Ala Arg Glu Phe Gly Leu Asn Met Phe Gln Met Glu
85 90 95

Asp Ser Asp Ala Asn Gln Thr Ile Ser Leu Leu Lys Val Lys Glu Ala
100 105 110

Leu Gln Asp Ala Gly Ile Asp Ala Leu Gly Lys Glu Lys Lys Asn Ile
115 120 125

Gly Cys Val Leu Gly Ile Gly Gly Gly Gln Lys Ser Ser His Glu Phe
130 135 140

Tyr Ser Arg Leu Asn Tyr Val Val Val Glu Lys Val Leu Arg Lys Met
145 150 155 160

Gly Met Pro Glu Glu Asp Val Lys Val Ala Val Glu Lys Tyr Lys Ala
165 170 175

Asn Phe Pro Glu Trp Arg Leu Asp Ser Phe Pro Gly Phe Leu Gly Asn
180 185 190

Val Thr Ala Gly Arg Cys Thr Asn
195 200

<210> 9
<211> 1278
<212> DNA
<213> Schizochytrium sp.

<220>
<221> CDS
<222> (1)..(1278)

<400> 9
gat gtc acc aag gag gcc tgg cgc ctc ccc cgc gag ggc gtc agc ttc 48
Asp Val Thr Lys Glu Ala Trp Arg Leu Pro Arg Glu Gly Val Ser Phe
1 5 10 15

cgc gcc aag ggc atc gcc acc aac ggc gct gtc gcc gcg ctc ttc tcc Arg Ala Lys Gly Ile Ala Thr Asn Gly Ala Val Ala Ala Leu Phe Ser 20 25 30	96
ggc cag ggc gcg cag tac acg cac atg ttt agc gag gtg gcc atg aac Gly Gln Gly Ala Gln Tyr Thr His Met Phe Ser Glu Val Ala Met Asn 35 40 45	144
tgg ccc cag ttc cgc cag agc att gcc gcc atg gac gcc gcc cag tcc Trp Pro Gln Phe Arg Gln Ser Ile Ala Ala Met Asp Ala Ala Gln Ser 50 55 60	192
aag gtc gct gga agc gac aag gac ttt gag cgc gtc tcc cag gtc ctc Lys Val Ala Gly Ser Asp Lys Asp Phe Glu Arg Val Ser Gln Val Leu 65 70 75 80	240
tac ccg cgc aag ccg tac gag cgt gag ccc gag cag aac ccc aag aag Tyr Pro Arg Lys Pro Tyr Glu Arg Glu Pro Glu Gln Asn Pro Lys Lys 85 90 95	288
atc tcc ctc acc gcc tac tcg cag ccc tcg acc ctg gcc tgc gct ctc Ile Ser Leu Thr Ala Tyr Ser Gln Pro Ser Thr Leu Ala Cys Ala Leu 100 105 110	336
ggt gcc ttt gag atc ttc aag gag gcc ggc ttc acc ccg gac ttt gcc Gly Ala Phe Glu Ile Phe Lys Glu Ala Gly Phe Thr Pro Asp Phe Ala 115 120 125	384
gcc gcc cat tcg ctc ggt gag ttc gcc gcc ctc tac gcc gcg ggc tgc Ala Gly His Ser Leu Gly Glu Phe Ala Ala Leu Tyr Ala Ala Gly Cys 130 135 140	432
gtc gac cgc gac gag ctc ttt gag ctt gtc tgc cgc cgc gcc cgc atc Val Asp Arg Asp Glu Leu Phe Glu Leu Val Cys Arg Arg Ala Arg Ile 145 150 155 160	480
atg ggc ggc aag gac gca ccg gcc acc ccc aag gga tgc atg gcc gcc Met Gly Gly Lys Asp Ala Pro Ala Thr Pro Lys Gly Cys Met Ala Ala 165 170 175	528
gtc att ggc ccc aac gcc gag aac atc aag gtc cag gcc gcc aac gtc Val Ile Gly Pro Asn Ala Glu Asn Ile Lys Val Gln Ala Ala Asn Val 180 185 190	576

tgg ctc ggc aac tcc aac tcg cct tcg cag acc gtc atc acc ggc tcc Trp Leu Gly Asn Ser Asn Ser Pro Ser Gln Thr Val Ile Thr Gly Ser 195 200 205	624
gtc gaa ggt atc cag gcc gag agc gcc cgc ctc cag aag gag ggc ttc Val Glu Gly Ile Gln Ala Glu Ser Ala Arg Leu Gln Lys Glu Gly Phe 210 215 220	672
cgc gtc gtg cct ctt gcc tgc gag agc gcc ttc cac tcg ccc cag atg Arg Val Val Pro Leu Ala Cys Glu Ser Ala Phe His Ser Pro Gln Met 225 230 235 240	720
gag aac gcc tcg tcg gcc ttc aag gac gtc atc tcc aag gtc tcc ttc Glu Asn Ala Ser Ser Ala Phe Lys Asp Val Ile Ser Lys Val Ser Phe 245 250 255	768
cgc acc ccc aag gcc gag acc aag ctc ttc agc aac gtc tct ggc gag Arg Thr Pro Lys Ala Glu Thr Lys Leu Phe Ser Asn Val Ser Gly Glu 260 265 270	816
acc tac ccc acg gac gcc cgc gag atg ctt acg cag cac atg acc agc Thr Tyr Pro Thr Asp Ala Arg Glu Met Leu Thr Gln His Met Thr Ser 275 280 285	864
agc gtc aag ttc ctc acc cag gtc cgc aac atg cac cag gcc ggt gcg Ser Val Lys Phe Leu Thr Gln Val Arg Asn Met His Gln Ala Gly Ala 290 295 300	912
cgc atc ttt gtc gag ttc gga ccc aag cag gtg ctc tcc aag ctt gtc Arg Ile Phe Val Glu Phe Gly Pro Lys Gln Val Leu Ser Lys Leu Val 305 310 315 320	960
tcc gag acc ctc aag gat gac ccc tcg gtt gtc acc gtc tct gtc aac Ser Glu Thr Leu Lys Asp Asp Pro Ser Val Val Thr Val Ser Val Asn 325 330 335	1008
ccg gcc tcg ggc acg gat tcg gac atc cag ctc cgc gac gcg gcc gtc Pro Ala Ser Gly Thr Asp Ser Asp Ile Gln Leu Arg Asp Ala Ala Val 340 345 350	1056
cag ctc gtt gtc gct ggc gtc aac ctt cag ggc ttt gac aag tgg gac Gln Leu Val Val Ala Gly Val Asn Leu Gln Gly Phe Asp Lys Trp Asp 355 360 365	1104

gcc ccc gat gcc acc cgc atg cag gcc atc aag aag aag cgc act acc 1152
Ala Pro Asp Ala Thr Arg Met Gln Ala Ile Lys Lys Lys Arg Thr Thr
370 375 380

ctc cgc ctt tcg gcc gcc acc tac gtc tcg gac aag acc aag aag gtc 1200
Leu Arg Leu Ser Ala Ala Thr Tyr Val Ser Asp Lys Thr Lys Lys Val
385 390 395 400

cgc gac gcc gcc atg aac gat ggc cgc tgc gtc acc tac ctc aag ggc 1248
Arg Asp Ala Ala Met Asn Asp Gly Arg Cys Val Thr Tyr Leu Lys Gly
405 410 415

gcc gca ccg ctc atc aag gcc ccg gag ccc 1278
Ala Ala Pro Leu Ile Lys Ala Pro Glu Pro
420 425

<210> 10
<211> 426
<212> PRT
<213> Schizochytrium sp.

<400> 10

Asp Val Thr Lys Glu Ala Trp Arg Leu Pro Arg Glu Gly Val Ser Phe
1 5 10 15

Arg Ala Lys Gly Ile Ala Thr Asn Gly Ala Val Ala Ala Leu Phe Ser
20 25 30

Gly Gln Gly Ala Gln Tyr Thr His Met Phe Ser Glu Val Ala Met Asn
35 40 45

Trp Pro Gln Phe Arg Gln Ser Ile Ala Ala Met Asp Ala Ala Gln Ser
50 55 60

Lys Val Ala Gly Ser Asp Lys Asp Phe Glu Arg Val Ser Gln Val Leu
65 70 75 80

Tyr Pro Arg Lys Pro Tyr Glu Arg Glu Pro Glu Gln Asn Pro Lys Lys
 85 90 95

Ile Ser Leu Thr Ala Tyr Ser Gln Pro Ser Thr Leu Ala Cys Ala Leu
 100 105 110

Gly Ala Phe Glu Ile Phe Lys Glu Ala Gly Phe Thr Pro Asp Phe Ala
 115 120 125

Ala Gly His Ser Leu Gly Glu Phe Ala Ala Leu Tyr Ala Ala Gly Cys
 130 135 140

Val Asp Arg Asp Glu Leu Phe Glu Leu Val Cys Arg Arg Ala Arg Ile
 145 150 155 160

Met Gly Gly Lys Asp Ala Pro Ala Thr Pro Lys Gly Cys Met Ala Ala
 165 170 175

Val Ile Gly Pro Asn Ala Glu Asn Ile Lys Val Gln Ala Ala Asn Val
 180 185 190

Trp Leu Gly Asn Ser Asn Ser Pro Ser Gln Thr Val Ile Thr Gly Ser
 195 200 205

Val Glu Gly Ile Gln Ala Glu Ser Ala Arg Leu Gln Lys Glu Gly Phe
 210 215 220

Arg Val Val Pro Leu Ala Cys Glu Ser Ala Phe His Ser Pro Gln Met
 225 230 235 240

Glu Asn Ala Ser Ser Ala Phe Lys Asp Val Ile Ser Lys Val Ser Phe
 245 250 255

Arg Thr Pro Lys Ala Glu Thr Lys Leu Phe Ser Asn Val Ser Gly Glu
 260 265 270

Thr Tyr Pro Thr Asp Ala Arg Glu Met Leu Thr Gln His Met Thr Ser

275	280	285
Ser Val Lys Phe Leu Thr Gln Val Arg Asn Met His Gln Ala Gly Ala		
290	295	300
Arg Ile Phe Val Glu Phe Gly Pro Lys Gln Val Leu Ser Lys Leu Val		
305	310	315
320		
Ser Glu Thr Leu Lys Asp Asp Pro Ser Val Val Thr Val Ser Val Asn		
325	330	335
340		
Pro Ala Ser Gly Thr Asp Ser Asp Ile Gln Leu Arg Asp Ala Ala Val		
345	350	
355		
Gln Leu Val Val Ala Gly Val Asn Leu Gln Gly Phe Asp Lys Trp Asp		
360	365	
370		
Ala Pro Asp Ala Thr Arg Met Gln Ala Ile Lys Lys Lys Arg Thr Thr		
375	380	
385		
Leu Arg Leu Ser Ala Ala Thr Tyr Val Ser Asp Lys Thr Lys Lys Val		
390	395	400
405		
Arg Asp Ala Ala Met Asn Asp Gly Arg Cys Val Thr Tyr Leu Lys Gly		
410	415	
420		
Ala Ala Pro Leu Ile Lys Ala Pro Glu Pro		
425		

<210>	11
<211>	5
<212>	PRT
<213>	Schizochytrium sp.

```
<220>
<221> MISC_FEATURE
<222> (1)..(5)
<223> Xaa = any amino acid
```

<400> 11

Gly His Ser Xaa Gly
1 5

<210> 12

<211> 258

<212> DNA

<213> Schizochytrium sp.

<220>

<221> CDS

<222> (1)..(258)

<400> 12

gct gtc tcg aac gag ctt ctt gag aag gcc gag act gtc gtc atg gag 48
Ala Val Ser Asn Glu Leu Leu Glu Lys Ala Glu Thr Val Val Met Glu
1 5 10 15

gtc ctc gcc gcc aag acc ggc tac gag acc gac atg atc gag gct gac 96
Val Leu Ala Ala Lys Thr Gly Tyr Glu Thr Asp Met Ile Glu Ala Asp
20 25 30

atg gag ctc gag acc gag ctc ggc att gac tcc atc aag cgt gtc gag 144
Met Glu Leu Glu Thr Glu Leu Gly Ile Asp Ser Ile Lys Arg Val Glu
35 40 45

atc ctc tcc gag gtc cag gcc atg ctc aat gtc gag gcc aag gat gtc 192
Ile Leu Ser Glu Val Gln Ala Met Leu Asn Val Glu Ala Lys Asp Val
50 55 60

gat gcc ctc agc cgc act cgc act gtt ggt gag gtt gtc aac gcc atg 240
Asp Ala Leu Ser Arg Thr Arg Thr Val Gly Glu Val Val Asn Ala Met
65 70 75 80

aag gcc gag atc gct ggc 258
Lys Ala Glu Ile Ala Gly
85

<210> 13

<211> 86

<212> PRT

<213> Schizochytrium sp.

<400> 13

Ala	Val	Ser	Asn	Glu	Leu	Leu	Glu	Lys	Ala	Glu	Thr	Val	Val	Met	Glu
1				5					10					15	

Val	Leu	Ala	Ala	Lys	Thr	Gly	Tyr	Glu	Thr	Asp	Met	Ile	Glu	Ala	Asp
			20					25					30		

Met	Glu	Leu	Glu	Thr	Glu	Leu	Gly	Ile	Asp	Ser	Ile	Lys	Arg	Val	Glu
		35					40					45			

Ile	Leu	Ser	Glu	Val	Gln	Ala	Met	Leu	Asn	Val	Glu	Ala	Lys	Asp	Val
	50						55					60			

Asp	Ala	Leu	Ser	Arg	Thr	Arg	Thr	Val	Gly	Glu	Val	Val	Asn	Ala	Met
65					70					75					80

Lys	Ala	Glu	Ile	Ala	Gly
				85	

<210> 14

<211> 5

<212> PRT

<213> Schizochytrium sp.

<400> 14

Leu	Gly	Ile	Asp	Ser
1				5

<210> 15

<211> 21

<212> PRT

<213> Schizochytrium sp.

<400> 15

Ala Pro Ala Pro Val Lys Ala Ala Ala Pro Ala Ala Pro Val Ala Ser
1 5 10 15

Ala Pro Ala Pro Ala
20

<210> 16

<211> 3006

<212> DNA

<213> Schizochytrium sp.

<400> 16

gcccccgccc cggtaaggcc tgctgcgcct gccgcccccg ttgcctcggc cctgccccg	60
gtgtgtctga acgagcttct tgagaaggcc gagactgtcg tcatggaggt cctcgccgcc	120
aagaccggct acgagaccga catgatcgag gctgacatgg agctcgagac cgagctcggc	180
attgactcca tcaagcgtgt cgagatcctc tccgaggtcc aggccatgct caatgtcgag	240
gccaaaggatg tcgatgcctt cagccgcact cgcactgttg gtgaggttgt caacgccatg	300
aaggccgaga tcgctggcag ctctgccccg ggcctgtctg ccgctgtccc ggctccggcc	360
aaggctgccc ctgccgccgc tgcgctgtct gtctcgaacg agcttctcga gaaggccgag	420
accgtcgtca tggaggctct cgccgccaag actggctacg agactgacat gatcgagtcc	480
gacatggagc tcgagactga gctcggcatt gactccatca agcgtgtcga gatcctctcc	540
gaggttcagg ccatgtctaa cgtcgaggcc aaggacgtcg acgctctcag ccgcactcgc	600
actgtgggtg aggtcgtcaa cgccatgaag gctgagatcg ctggtggctc tgccccggcg	660
cctgcccgcg ctgccccagg tccggtgtct gccgccccctg cgctcgccgc cgccgcccct	720
gtgtgtctga acgagcttct tgagaaggcc gagaccgtcg tcatggaggt cctcgccgcc	780

aagactggct acgagactga catgatcgag tccgacatgg agctcgagac cgagctcggc 840

attgactcca tcaagcgtgt cgagattctc tccgaggtcc aggccatgct caacgtcgag 900

gccaaaggacg tcgacgctct cagccgcacc cgcactgttg gcgaggctgt cgatgccatg 960

aaggccgaga tcgctggtgg ctctgccccg ggcctgccg ccgctgctcc tgctccggct 1020

gctgccgcc ctgcgcctgc cgccttgcg cctgctgtct cgagcgagct tctcgagaag 1080

gccgagactg tcgtcatgga ggtcctgcc gccaaactg gctacgagac tgacatgatc 1140

gagtccgaca tggagctcga gaccgagctc ggcatgact ccatcaagcg tctcgagatt 1200

ctctccgagg tccaggccat gctcaacgtc gagccaagg acgtcgacgc tctcagccgc 1260

acccgactg ttggcgaggt cgtcgatgcc atgaaggccg agatcgctgg tggctctgcc 1320

ccggcgctg cgcgcctgc tctgctccg gctgctgccg ccctgcgcc tgccgccct 1380

gcgcctgccg ccctgcgcc tgctgtctcg agcgagcttc tcgagaaggc cgagactgtc 1440

gtcatggagg tctcgcgc caagactggc tacgagactg acatgattga gtccgacatg 1500

gagctcgaga ccgagctcgg cattgactcc atcaagcgtg tcgagattct ctccgaggtt 1560

caggccatgc tcaacgtcga ggccaaggac gtcgacgtc tcagccgcac tcgactgtt 1620

ggtgaggctg tcgatccat gaaggctgag atcgctggca gctccgctc ggcgcctgcc 1680

gccgctgtc ctgctccggc tgctgccgt cctgcgccg ctgccgcgc ccctgtgtc 1740

tcgaacgagc ttctcgagaa agccgagact gtcgtcatgg aggtcctgc gcccaagact 1800

ggctacgaga ctgacatgat cgagtccgac atggagctcg agactgagct cggcattgac 1860

tccatcaagc gtgtcgagat cctctccgag gttcaggcca tgctcaacgt cgaggccaag 1920

gacgtcgatg ccctcagccg caccgcgact gttggcgagg ttgtcgatgc catgaaggcc 1980

gagatcgctg gtggctctgc cccggcgccct gccgccgctg cccctgctcc ggctgccgcc 2040

gccccctgctg tctcgaacga gcttctcgag aaggccgaga ctgtcgatcat ggaggtcctc 2100

gccgccaaga ctggctacga gaccgacatg atcgagtcgg acatggagct cgagaccgag 2160

ctcggcattg actccatcaa gcgtgtcgag attctctccg aggttcaggc catgctcaac 2220

gtcgaggcca aggacgtcga tgctctcagc cgactcgca ctgttggcga ggtcgtcgat 2280

gccatgaagg ctgagatcgc cggcagctcc gccccggcgc ctgccgccgc tgctcctgct 2340

ccggctgctg ccgtctctgc gcccgctgcc gctgccccctg ctgtctcgag cgagcttctc 2400

gagaaggccg agaccgtcgt catggaggtc ctgccgccca agactggcta cgagactgac 2460

atgattgagt ccgacatgga gctcgagact gagctcggca ttgactccat caagcgtgtc 2520

gagatcctct ccgaggttca ggccatgctc aacgtcgagg ccaaggacgt cgatgccctc 2580

agccgcaccc gcactgttgg cgaggttgct gatgcatga aggccgagat cgctggtggc 2640

tctgccccgg cgcttgccgc cgctgcccct gctccggctg ccgccgcccc tgctgtctcg 2700

aacgagcttc ttgagaaggc cgagaccgtc gtcattggagg tcctcgccgc caagactggc 2760

tacgagaccg acatgatcga gtccgacatg gagctcgaga ccgagctcgg cattgactcc 2820

atcaagcgtg tcgagattct ctccgaggtt caggccatgc tcaacgtcga ggccaaggac 2880

gtcgacgctc tcagccgcac tcgcactgtt ggcgaggctg tcgatgcat gaaggctgag 2940

atcgctggtg gctctgcccc ggcgccctgcc gccgctgctc ctgcctcggc tggcgccgcg 3000

cctgcg 3006

<211> 2133
 <212> DNA
 <213> Schizochytrium sp.

<220>
 <221> CDS
 <222> (1)..(2133)

<400> 17
 ttt ggc gct ctc ggc ggc ttc atc tcg cag cag gcg gag cgc ttc gag 48
 Phe Gly Ala Leu Gly Gly Phe Ile Ser Gln Gln Ala Glu Arg Phe Glu
 1 5 10 15

ccc gcc gaa atc ctc ggc ttc acg ctc atg tgc gcc aag ttc gcc aag 96
 Pro Ala Glu Ile Leu Gly Phe Thr Leu Met Cys Ala Lys Phe Ala Lys
 20 25 30

gct tcc ctc tgc acg gct gtg gct ggc ggc cgc ccg gcc ttt atc ggt 144
 Ala Ser Leu Cys Thr Ala Val Ala Gly Gly Arg Pro Ala Phe Ile Gly
 35 40 45

gtg gcg cgc ctt gac ggc cgc ctc gga ttc act tcg cag ggc act tct 192
 Val Ala Arg Leu Asp Gly Arg Leu Gly Phe Thr Ser Gln Gly Thr Ser
 50 55 60

gac gcg ctc aag cgt gcc cag cgt ggt gcc atc ttt ggc ctc tgc aag 240
 Asp Ala Leu Lys Arg Ala Gln Arg Gly Ala Ile Phe Gly Leu Cys Lys
 65 70 75 80

acc atc ggc ctc gag tgg tcc gag tct gac gtc ttt tcc cgc ggc gtg 288
 Thr Ile Gly Leu Glu Trp Ser Glu Ser Asp Val Phe Ser Arg Gly Val
 85 90 95

gac att gct cag ggc atg cac ccc gag gat gcc gcc gtg gcg att gtg 336
 Asp Ile Ala Gln Gly Met His Pro Glu Asp Ala Ala Val Ala Ile Val
 100 105 110

cgc gag atg gcg tgc gct gac att cgc att cgc gag gtc ggc att ggc 384
 Arg Glu Met Ala Cys Ala Asp Ile Arg Ile Arg Glu Val Gly Ile Gly
 115 120 125

gca aac cag cag cgc tgc acg atc cgt gcc gcc aag ctc gag acc ggc 432
 Ala Asn Gln Gln Arg Cys Thr Ile Arg Ala Ala Lys Leu Glu Thr Gly
 130 135 140

aac ccg cag cgc cag atc gcc aag gac gac gtg ctg ctc gtt tct ggc Asn Pro Gln Arg Gln Ile Ala Lys Asp Asp Val Leu Leu Val Ser Gly 145 150 155 160	480
ggc gct cgc ggc atc acg cct ctt tgc atc cgg gag atc acg cgc cag Gly Ala Arg Gly Ile Thr Pro Leu Cys Ile Arg Glu Ile Thr Arg Gln 165 170 175	528
atc gcg ggc ggc aag tac att ctg ctt ggc cgc agc aag gtc tct gcg Ile Ala Gly Gly Lys Tyr Ile Leu Leu Gly Arg Ser Lys Val Ser Ala 180 185 190	576
agc gaa ccg gca tgg tgc gct ggc atc act gac gag aag gct gtg caa Ser Glu Pro Ala Trp Cys Ala Gly Ile Thr Asp Glu Lys Ala Val Gln 195 200 205	624
aag gct gct acc cag gag ctc aag cgc gcc ttt agc gct ggc gag ggc Lys Ala Ala Thr Gln Glu Leu Lys Arg Ala Phe Ser Ala Gly Glu Gly 210 215 220	672
ccc aag ccc acg ccc cgc gct gtc act aag ctt gtg ggc tct gtt ctt Pro Lys Pro Thr Pro Arg Ala Val Thr Lys Leu Val Gly Ser Val Leu 225 230 235 240	720
ggc gct cgc gag gtg cgc agc tct att gct gcg att gaa gcg ctc ggc Gly Ala Arg Glu Val Arg Ser Ser Ile Ala Ala Ile Glu Ala Leu Gly 245 250 255	768
ggc aag gcc atc tac tcg tcg tgc gac gtg aac tct gcc gcc gac gtg Gly Lys Ala Ile Tyr Ser Ser Cys Asp Val Asn Ser Ala Ala Asp Val 260 265 270	816
gcc aag gcc gtg cgc gat gcc gag tcc cag ctc ggt gcc cgc gtc tcg Ala Lys Ala Val Arg Asp Ala Glu Ser Gln Leu Gly Ala Arg Val Ser 275 280 285	864
ggc atc gtt cat gcc tcg ggc gtg ctc cgc gac cgt ctc atc gag aag Gly Ile Val His Ala Ser Gly Val Leu Arg Asp Arg Leu Ile Glu Lys 290 295 300	912
aag ctc ccc gac gag ttc gac gcc gtc ttt ggc acc aag gtc acc ggt Lys Leu Pro Asp Glu Phe Asp Ala Val Phe Gly Thr Lys Val Thr Gly 305 310 315 320	960

ctc gag aac ctc ctc gcc gcc gtc gac cgc gcc aac ctc aag cac atg Leu Glu Asn Leu Leu Ala Ala Val Asp Arg Ala Asn Leu Lys His Met 325 330 335	1008
gtc ctc ttc agc tcg ctc gcc ggc ttc cac ggc aac gtc ggc cag tct Val Leu Phe Ser Ser Leu Ala Gly Phe His Gly Asn Val Gly Gln Ser 340 345 350	1056
gac tac gcc atg gcc aac gag gcc ctt aac aag atg ggc ctc gag ctc Asp Tyr Ala Met Ala Asn Glu Ala Leu Asn Lys Met Gly Leu Glu Leu 355 360 365	1104
gcc aag gac gtc tcg gtc aag tcg atc tgc ttc ggt ccc tgg gac ggt Ala Lys Asp Val Ser Val Lys Ser Ile Cys Phe Gly Pro Trp Asp Gly 370 375 380	1152
ggc atg gtg acg ccg cag ctc aag aag cag ttc cag gag atg ggc gtg Gly Met Val Thr Pro Gln Leu Lys Lys Gln Phe Gln Glu Met Gly Val 385 390 395 400	1200
cag atc atc ccc cgc gag ggc ggc gct gat acc gtg gcg cgc atc gtg Gln Ile Ile Pro Arg Glu Gly Gly Ala Asp Thr Val Ala Arg Ile Val 405 410 415	1248
ctc ggc tcc tcg ccg gct gag atc ctt gtc ggc aac tgg cgc acc ccg Leu Gly Ser Ser Pro Ala Glu Ile Leu Val Gly Asn Trp Arg Thr Pro 420 425 430	1296
tcc aag aag gtc ggc tcg gac acc atc acc ctg cac cgc aag att tcc Ser Lys Lys Val Gly Ser Asp Thr Ile Thr Leu His Arg Lys Ile Ser 435 440 445	1344
gcc aag tcc aac ccc ttc ctc gag gac cac gtc atc cag ggc cgc cgc Ala Lys Ser Asn Pro Phe Leu Glu Asp His Val Ile Gln Gly Arg Arg 450 455 460	1392
gtg ctg ccc atg acg ctg gcc att ggc tcg ctc gcg gag acc tgc ctc Val Leu Pro Met Thr Leu Ala Ile Gly Ser Leu Ala Glu Thr Cys Leu 465 470 475 480	1440
ggc ctc ttc ccc ggc tac tcg ctc tgg gcc att gac gac gcc cag ctc Gly Leu Phe Pro Gly Tyr Ser Leu Trp Ala Ile Asp Asp Ala Gln Leu 485 490 495	1488

ttc aag ggt gtc act gtc gac ggc gac gtc aac tgc gag gtg acc ctc Phe Lys Gly Val Thr Val Asp Gly Asp Val Asn Cys Glu Val Thr Leu 500 505 510	1536
acc ccg tcg acg gcg ccc tcg ggc cgc gtc aac gtc cag gcc acg ctc Thr Pro Ser Thr Ala Pro Ser Gly Arg Val Asn Val Gln Ala Thr Leu 515 520 525	1584
aag acc ttt tcc agc ggc aag ctg gtc ccg gcc tac cgc gcc gtc atc Lys Thr Phe Ser Ser Gly Lys Leu Val Pro Ala Tyr Arg Ala Val Ile 530 535 540	1632
gtg ctc tcc aac cag ggc gcg ccc ccg gcc aac gcc acc atg cag ccg Val Leu Ser Asn Gln Gly Ala Pro Pro Ala Asn Ala Thr Met Gln Pro 545 550 555 560	1680
ccc tcg ctc gat gcc gat ccg gcg ctc cag ggc tcc gtc tac gac ggc Pro Ser Leu Asp Ala Asp Pro Ala Leu Gln Gly Ser Val Tyr Asp Gly 565 570 575	1728
aag acc ctc ttc cac ggc ccg gcc ttc cgc ggc atc gat gac gtg ctc Lys Thr Leu Phe His Gly Pro Ala Phe Arg Gly Ile Asp Asp Val Leu 580 585 590	1776
tcg tgc acc aag agc cag ctt gtg gcc aag tgc agc gct gtc ccc ggc Ser Cys Thr Lys Ser Gln Leu Val Ala Lys Cys Ser Ala Val Pro Gly 595 600 605	1824
tcc gac gcc gct cgc ggc gag ttt gcc acg gac act gac gcc cat gac Ser Asp Ala Ala Arg Gly Glu Phe Ala Thr Asp Thr Asp Ala His Asp 610 615 620	1872
ccc ttc gtg aac gac ctg gcc ttt cag gcc atg ctc gtc tgg gtg cgc Pro Phe Val Asn Asp Leu Ala Phe Gln Ala Met Leu Val Trp Val Arg 625 630 635 640	1920
cgc acg ctc ggc cag gct gcg ctc ccc aac tcg atc cag cgc atc gtc Arg Thr Leu Gly Gln Ala Ala Leu Pro Asn Ser Ile Gln Arg Ile Val 645 650 655	1968
cag cac cgc ccg gtc ccg cag gac aag ccc ttc tac att acc ctc cgc Gln His Arg Pro Val Pro Gln Asp Lys Pro Phe Tyr Ile Thr Leu Arg 660 665 670	2016

tcc aac cag tgc ggc ggt cac tcc cag cac aag cac gcc ctt cag ttc 2064
Ser Asn Gln Ser Gly Gly His Ser Gln His Lys His Ala Leu Gln Phe
675 680 685

cac aac gag cag ggc gat ctc ttc att gat gtc cag gct tgc gtc atc 2112
His Asn Glu Gln Gly Asp Leu Phe Ile Asp Val Gln Ala Ser Val Ile
690 695 700

gcc acg gac agc ctt gcc ttc 2133
Ala Thr Asp Ser Leu Ala Phe
705 710

<210> 18
<211> 711
<212> PRT
<213> Schizochytrium sp.

<400> 18

Phe Gly Ala Leu Gly Gly Phe Ile Ser Gln Gln Ala Glu Arg Phe Glu
1 5 10 15

Pro Ala Glu Ile Leu Gly Phe Thr Leu Met Cys Ala Lys Phe Ala Lys
20 25 30

Ala Ser Leu Cys Thr Ala Val Ala Gly Gly Arg Pro Ala Phe Ile Gly
35 40 45

Val Ala Arg Leu Asp Gly Arg Leu Gly Phe Thr Ser Gln Gly Thr Ser
50 55 60

Asp Ala Leu Lys Arg Ala Gln Arg Gly Ala Ile Phe Gly Leu Cys Lys
65 70 75 80

Thr Ile Gly Leu Glu Trp Ser Glu Ser Asp Val Phe Ser Arg Gly Val
85 90 95

Asp Ile Ala Gln Gly Met His Pro Glu Asp Ala Ala Val Ala Ile Val
100 105 110

Arg Glu Met Ala Cys Ala Asp Ile Arg Ile Arg Glu Val Gly Ile Gly
115 120 125

Ala Asn Gln Gln Arg Cys Thr Ile Arg Ala Ala Lys Leu Glu Thr Gly
130 135 140

Asn Pro Gln Arg Gln Ile Ala Lys Asp Asp Val Leu Leu Val Ser Gly
145 150 155 160

Gly Ala Arg Gly Ile Thr Pro Leu Cys Ile Arg Glu Ile Thr Arg Gln
165 170 175

Ile Ala Gly Gly Lys Tyr Ile Leu Leu Gly Arg Ser Lys Val Ser Ala
180 185 190

Ser Glu Pro Ala Trp Cys Ala Gly Ile Thr Asp Glu Lys Ala Val Gln
195 200 205

Lys Ala Ala Thr Gln Glu Leu Lys Arg Ala Phe Ser Ala Gly Glu Gly
210 215 220

Pro Lys Pro Thr Pro Arg Ala Val Thr Lys Leu Val Gly Ser Val Leu
225 230 235 240

Gly Ala Arg Glu Val Arg Ser Ser Ile Ala Ala Ile Glu Ala Leu Gly
245 250 255

Gly Lys Ala Ile Tyr Ser Ser Cys Asp Val Asn Ser Ala Ala Asp Val
260 265 270

Ala Lys Ala Val Arg Asp Ala Glu Ser Gln Leu Gly Ala Arg Val Ser
275 280 285

Gly Ile Val His Ala Ser Gly Val Leu Arg Asp Arg Leu Ile Glu Lys

290

295

300

Lys Leu Pro Asp Glu Phe Asp Ala Val Phe Gly Thr Lys Val Thr Gly
305 310 315 320

Leu Glu Asn Leu Leu Ala Ala Val Asp Arg Ala Asn Leu Lys His Met
325 330 335

Val Leu Phe Ser Ser Leu Ala Gly Phe His Gly Asn Val Gly Gln Ser
340 345 350

Asp Tyr Ala Met Ala Asn Glu Ala Leu Asn Lys Met Gly Leu Glu Leu
355 360 365

Ala Lys Asp Val Ser Val Lys Ser Ile Cys Phe Gly Pro Trp Asp Gly
370 375 380

Gly Met Val Thr Pro Gln Leu Lys Lys Gln Phe Gln Glu Met Gly Val
385 390 395 400

Gln Ile Ile Pro Arg Glu Gly Gly Ala Asp Thr Val Ala Arg Ile Val
405 410 415

Leu Gly Ser Ser Pro Ala Glu Ile Leu Val Gly Asn Trp Arg Thr Pro
420 425 430

Ser Lys Lys Val Gly Ser Asp Thr Ile Thr Leu His Arg Lys Ile Ser
435 440 445

Ala Lys Ser Asn Pro Phe Leu Glu Asp His Val Ile Gln Gly Arg Arg
450 455 460

Val Leu Pro Met Thr Leu Ala Ile Gly Ser Leu Ala Glu Thr Cys Leu
465 470 475 480

Gly Leu Phe Pro Gly Tyr Ser Leu Trp Ala Ile Asp Asp Ala Gln Leu
485 490 495

Phe Lys Gly Val Thr Val Asp Gly Asp Val Asn Cys Glu Val Thr Leu
500 505 510

Thr Pro Ser Thr Ala Pro Ser Gly Arg Val Asn Val Gln Ala Thr Leu
515 520 525

Lys Thr Phe Ser Ser Gly Lys Leu Val Pro Ala Tyr Arg Ala Val Ile
530 535 540

Val Leu Ser Asn Gln Gly Ala Pro Pro Ala Asn Ala Thr Met Gln Pro
545 550 555 560

Pro Ser Leu Asp Ala Asp Pro Ala Leu Gln Gly Ser Val Tyr Asp Gly
565 570 575

Lys Thr Leu Phe His Gly Pro Ala Phe Arg Gly Ile Asp Asp Val Leu
580 585 590

Ser Cys Thr Lys Ser Gln Leu Val Ala Lys Cys Ser Ala Val Pro Gly
595 600 605

Ser Asp Ala Ala Arg Gly Glu Phe Ala Thr Asp Thr Asp Ala His Asp
610 615 620

Pro Phe Val Asn Asp Leu Ala Phe Gln Ala Met Leu Val Trp Val Arg
625 630 635 640

Arg Thr Leu Gly Gln Ala Ala Leu Pro Asn Ser Ile Gln Arg Ile Val
645 650 655

Gln His Arg Pro Val Pro Gln Asp Lys Pro Phe Tyr Ile Thr Leu Arg
660 665 670

Ser Asn Gln Ser Gly Gly His Ser Gln His Lys His Ala Leu Gln Phe
675 680 685

His Asn Glu Gln Gly Asp Leu Phe Ile Asp Val Gln Ala Ser Val Ile
690 695 700

Ala Thr Asp Ser Leu Ala Phe
705 710

<210> 19
<211> 1350
<212> DNA
<213> Schizochytrium sp.

<220>
<221> CDS
<222> (1)..(1350)

<400> 19
atg gcc gct cgg aat gtg agc gcc gcg cat gag atg cac gat gaa aag 48
Met Ala Ala Arg Asn Val Ser Ala Ala His Glu Met His Asp Glu Lys
1 5 10 15

cgc atc gcc gtc gtc ggc atg gcc gtc cag tac gcc gga tgc aaa acc 96
Arg Ile Ala Val Val Gly Met Ala Val Gln Tyr Ala Gly Cys Lys Thr
20 25 30

aag gac gag ttc tgg gag gtg ctc atg aac ggc aag gtc gag tcc aag 144
Lys Asp Glu Phe Trp Glu Val Leu Met Asn Gly Lys Val Glu Ser Lys
35 40 45

gtg atc agc gac aaa cga ctc ggc tcc aac tac cgc gcc gag cac tac 192
Val Ile Ser Asp Lys Arg Leu Gly Ser Asn Tyr Arg Ala Glu His Tyr
50 55 60

aaa gca gag cgc agc aag tat gcc gac acc ttt tgc aac gaa acg tac 240
Lys Ala Glu Arg Ser Lys Tyr Ala Asp Thr Phe Cys Asn Glu Thr Tyr
65 70 75 80

ggc acc ctt gac gag aac gag atc gac aac gag cac gaa ctc ctc ctc 288
Gly Thr Leu Asp Glu Asn Glu Ile Asp Asn Glu His Glu Leu Leu Leu
85 90 95

aac ctc gcc aag cag gca ctc gca gag aca tcc gtc aaa gac tcg aca 336

Asn Leu Ala Lys Gln Ala Leu Ala Glu Thr Ser Val Lys Asp Ser Thr	
100 105 110	
cgc tgc ggc atc gtc agc ggc tgc ctc tcg ttc ccc atg gac aac ctc	384
Arg Cys Gly Ile Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	
115 120 125	
cag ggt gaa ctc ctc aac gtg tac caa aac cat gtc gag aaa aag ctc	432
Gln Gly Glu Leu Leu Asn Val Tyr Gln Asn His Val Glu Lys Lys Leu	
130 135 140	
ggg gcc cgc gtc ttc aag gac gcc tcc cat tgg tcc gaa cgc gag cag	480
Gly Ala Arg Val Phe Lys Asp Ala Ser His Trp Ser Glu Arg Glu Gln	
145 150 155 160	
tcc aac aaa ccc gag gcc ggt gac cgc cgc atc ttc atg gac ccg gcc	528
Ser Asn Lys Pro Glu Ala Gly Asp Arg Arg Ile Phe Met Asp Pro Ala	
165 170 175	
tcc ttc gtc gcc gaa gaa ctc aac ctc ggc gcc ctt cac tac tcc gtc	576
Ser Phe Val Ala Glu Glu Leu Asn Leu Gly Ala Leu His Tyr Ser Val	
180 185 190	
gac gca gca tgc gcc acg gcg ctc tac gtg ctc cgc ctc gcg cag gat	624
Asp Ala Ala Cys Ala Thr Ala Leu Tyr Val Leu Arg Leu Ala Gln Asp	
195 200 205	
cat ctc gtc tcc ggc gcc gcc gac gtc atg ctc tgc ggt gcc acc tgc	672
His Leu Val Ser Gly Ala Ala Asp Val Met Leu Cys Gly Ala Thr Cys	
210 215 220	
ctg ccg gag ccc ttt ttc atc ctt tcg ggc ttt tcc acc ttc cag gcc	720
Leu Pro Glu Pro Phe Phe Ile Leu Ser Gly Phe Ser Thr Phe Gln Ala	
225 230 235 240	
atg ccc gtc ggc acg ggc cag aac gtg tcc atg ccg ctg cac aag gac	768
Met Pro Val Gly Thr Gly Gln Asn Val Ser Met Pro Leu His Lys Asp	
245 250 255	
agc cag ggc ctc acc ccg ggt gag ggc ggc tcc atc atg gtc ctc aag	816
Ser Gln Gly Leu Thr Pro Gly Glu Gly Gly Ser Ile Met Val Leu Lys	
260 265 270	
cgt ctc gat gat gcc atc cgc gac ggc gac cac att tac ggc acc ctt	864
Arg Leu Asp Asp Ala Ile Arg Asp Gly Asp His Ile Tyr Gly Thr Leu	

275	280	285	
ctc ggc gcc aat gtc agc aac tcc ggc aca ggt ctg ccc ctc aag ccc Leu Gly Ala Asn Val Ser Asn Ser Gly Thr Gly Leu Pro Leu Lys Pro 290 295 300			912
ctt ctc ccc agc gag aaa aag tgc ctc atg gac acc tac acg cgc att Leu Leu Pro Ser Glu Lys Lys Cys Leu Met Asp Thr Tyr Thr Arg Ile 305 310 315 320			960
aac gtg cac ccg cac aag att cag tac gtc gag tgc cac gcc acc ggc Asn Val His Pro His Lys Ile Gln Tyr Val Glu Cys His Ala Thr Gly 325 330 335			1008
acg ccc cag ggt gat cgt gtg gaa atc gac gcc gtc aag gcc tgc ttt Thr Pro Gln Gly Asp Arg Val Glu Ile Asp Ala Val Lys Ala Cys Phe 340 345 350			1056
gaa ggc aag gtc ccc cgt ttc ggt acc aca aag ggc aac ttt gga cac Glu Gly Lys Val Pro Arg Phe Gly Thr Thr Lys Gly Asn Phe Gly His 355 360 365			1104
acc cts gyc gca gcc ggc ttt gcc ggt atg tgc aag gtc ctc ctc tcc Thr Xaa Xaa Ala Ala Gly Phe Ala Gly Met Cys Lys Val Leu Leu Ser 370 375 380			1152
atg aag cat ggc atc atc ccg ccc acc ccg ggt atc gat gac gag acc Met Lys His Gly Ile Ile Pro Pro Thr Pro Gly Ile Asp Asp Glu Thr 385 390 395 400			1200
aag atg gac cct ctc gtc gtc tcc ggt gag gcc atc cca tgg cca gag Lys Met Asp Pro Leu Val Val Ser Gly Glu Ala Ile Pro Trp Pro Glu 405 410 415			1248
acc aac ggc gag ccc aag cgc gcc ggt ctc tcg gcc ttt ggc ttt ggt Thr Asn Gly Glu Pro Lys Arg Ala Gly Leu Ser Ala Phe Gly Phe Gly 420 425 430			1296
ggc acc aac gcc cat gcc gtc ttt gag gag cat gac ccc tcc aac gcc Gly Thr Asn Ala His Ala Val Phe Glu Glu His Asp Pro Ser Asn Ala 435 440 445			1344
gcc tgc Ala Cys			1350

450

<210> 20
 <211> 450
 <212> PRT
 <213> Schizochytrium sp.

<220>
 <221> misc_feature
 <222> (370)..(370)
 <223> The 'Xaa' at location 370 stands for Leu.

<220>
 <221> misc_feature
 <222> (371)..(371)
 <223> The 'Xaa' at location 371 stands for Ala, or Val.

<400> 20

Met Ala Ala Arg Asn Val Ser Ala Ala His Glu Met His Asp Glu Lys
 1 5 10 15

Arg Ile Ala Val Val Gly Met Ala Val Gln Tyr Ala Gly Cys Lys Thr
 20 25 30

Lys Asp Glu Phe Trp Glu Val Leu Met Asn Gly Lys Val Glu Ser Lys
 35 40 45

Val Ile Ser Asp Lys Arg Leu Gly Ser Asn Tyr Arg Ala Glu His Tyr
 50 55 60

Lys Ala Glu Arg Ser Lys Tyr Ala Asp Thr Phe Cys Asn Glu Thr Tyr
 65 70 75 80

Gly Thr Leu Asp Glu Asn Glu Ile Asp Asn Glu His Glu Leu Leu Leu
 85 90 95

Asn Leu Ala Lys Gln Ala Leu Ala Glu Thr Ser Val Lys Asp Ser Thr
 100 105 110

Arg Cys Gly Ile Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu
115 120 125

Gln Gly Glu Leu Leu Asn Val Tyr Gln Asn His Val Glu Lys Lys Leu
130 135 140

Gly Ala Arg Val Phe Lys Asp Ala Ser His Trp Ser Glu Arg Glu Gln
145 150 155 160

Ser Asn Lys Pro Glu Ala Gly Asp Arg Arg Ile Phe Met Asp Pro Ala
165 170 175

Ser Phe Val Ala Glu Glu Leu Asn Leu Gly Ala Leu His Tyr Ser Val
180 185 190

Asp Ala Ala Cys Ala Thr Ala Leu Tyr Val Leu Arg Leu Ala Gln Asp
195 200 205

His Leu Val Ser Gly Ala Ala Asp Val Met Leu Cys Gly Ala Thr Cys
210 215 220

Leu Pro Glu Pro Phe Phe Ile Leu Ser Gly Phe Ser Thr Phe Gln Ala
225 230 235 240

Met Pro Val Gly Thr Gly Gln Asn Val Ser Met Pro Leu His Lys Asp
245 250 255

Ser Gln Gly Leu Thr Pro Gly Glu Gly Gly Ser Ile Met Val Leu Lys
260 265 270

Arg Leu Asp Asp Ala Ile Arg Asp Gly Asp His Ile Tyr Gly Thr Leu
275 280 285

Leu Gly Ala Asn Val Ser Asn Ser Gly Thr Gly Leu Pro Leu Lys Pro
290 295 300

Leu Leu Pro Ser Glu Lys Lys Cys Leu Met Asp Thr Tyr Thr Arg Ile
305 310 315 320

Asn Val His Pro His Lys Ile Gln Tyr Val Glu Cys His Ala Thr Gly
325 330 335

Thr Pro Gln Gly Asp Arg Val Glu Ile Asp Ala Val Lys Ala Cys Phe
340 345 350

Glu Gly Lys Val Pro Arg Phe Gly Thr Thr Lys Gly Asn Phe Gly His
355 360 365

Thr Xaa Xaa Ala Ala Gly Phe Ala Gly Met Cys Lys Val Leu Leu Ser
370 375 380

Met Lys His Gly Ile Ile Pro Pro Thr Pro Gly Ile Asp Asp Glu Thr
385 390 395 400

Lys Met Asp Pro Leu Val Val Ser Gly Glu Ala Ile Pro Trp Pro Glu
405 410 415

Thr Asn Gly Glu Pro Lys Arg Ala Gly Leu Ser Ala Phe Gly Phe Gly
420 425 430

Gly Thr Asn Ala His Ala Val Phe Glu Glu His Asp Pro Ser Asn Ala
435 440 445

Ala Cys
450

<210> 21
<211> 1323
<212> DNA
<213> Schizochytrium sp.

<220>
<221> CDS

<222> (1)..(1323)

<400> 21

tgc gcc cgc tgc ggc ggt gaa agc aac atg cgc atc gcc atc act ggt	48
Ser Ala Arg Cys Gly Gly Glu Ser Asn Met Arg Ile Ala Ile Thr Gly	
1 5 10 15	

atg gac gcc acc ttt ggc gct ctc aag gga ctc gac gcc ttc gag cgc	96
Met Asp Ala Thr Phe Gly Ala Leu Lys Gly Leu Asp Ala Phe Glu Arg	
20 25 30	

gcc att tac acc ggc gct cac ggt gcc atc cca ctc cca gaa aag cgc	144
Ala Ile Tyr Thr Gly Ala His Gly Ala Ile Pro Leu Pro Glu Lys Arg	
35 40 45	

tgg cgc ttt ctc ggc aag gac aag gac ttt ctt gac ctc tgc ggc gtc	192
Trp Arg Phe Leu Gly Lys Asp Lys Asp Phe Leu Asp Leu Cys Gly Val	
50 55 60	

aag gcc acc ccg cac ggc tgc tac att gaa gat gtt gag gtc gac ttc	240
Lys Ala Thr Pro His Gly Cys Tyr Ile Glu Asp Val Glu Val Asp Phe	
65 70 75 80	

cag cgc ctc cgc acg ccc atg acc cct gaa gac atg ctc ctc cct cag	288
Gln Arg Leu Arg Thr Pro Met Thr Pro Glu Asp Met Leu Leu Pro Gln	
85 90 95	

cag ctt ctg gcc gtc acc acc att gac cgc gcc atc ctc gac tgc gga	336
Gln Leu Leu Ala Val Thr Thr Ile Asp Arg Ala Ile Leu Asp Ser Gly	
100 105 110	

atg aaa aag ggt ggc aat gtc gcc gtc ttt gtc ggc ctc ggc acc gac	384
Met Lys Lys Gly Gly Asn Val Ala Val Phe Val Gly Leu Gly Thr Asp	
115 120 125	

ctc gag ctc tac cgt cac cgt gct cgc gtc gct ctc aag gag cgc gtc	432
Leu Glu Leu Tyr Arg His Arg Ala Arg Val Ala Leu Lys Glu Arg Val	
130 135 140	

cgc cct gaa gcc tcc aag aag ctc aat gac atg atg cag tac att aac	480
Arg Pro Glu Ala Ser Lys Lys Leu Asn Asp Met Met Gln Tyr Ile Asn	
145 150 155 160	

gac tgc ggc aca tcc aca tcg tac acc tcg tac att ggc aac ctc gtc	528
Asp Cys Gly Thr Ser Thr Ser Tyr Thr Ser Tyr Ile Gly Asn Leu Val	

165	170	175	
gcc acg cgc gtc tcg tcg cag tgg ggc ttc acg ggc ccc tcc ttt acg Ala Thr Arg Val Ser Ser Gln Trp Gly Phe Thr Gly Pro Ser Phe Thr 180 185 190			576
atc acc gag ggc aac aac tcc gtc tac cgc tgc gcc gag ctc ggc aag Ile Thr Glu Gly Asn Asn Ser Val Tyr Arg Cys Ala Glu Leu Gly Lys 195 200 205			624
tac ctc ctc gag acc ggc gag gtc gat ggc gtc gtc gtt gcg ggt gtc Tyr Leu Leu Glu Thr Gly Glu Val Asp Gly Val Val Val Ala Gly Val 210 215 220			672
gat ctc tgc ggc agt gcc gaa aac ctt tac gtc aag tct cgc cgc ttc Asp Leu Cys Gly Ser Ala Glu Asn Leu Tyr Val Lys Ser Arg Arg Phe 225 230 235 240			720
aag gtg tcc acc tcc gat acc ccg cgc gcc agc ttt gac gcc gcc gcc Lys Val Ser Thr Ser Asp Thr Pro Arg Ala Ser Phe Asp Ala Ala Ala 245 250 255			768
gat ggc tac ttt gtc ggc gag ggc tgc ggt gcc ttt gtg ctc aag cgt Asp Gly Tyr Phe Val Gly Glu Gly Cys Gly Ala Phe Val Leu Lys Arg 260 265 270			816
gag act agc tgc acc aag gac gac cgt atc tac gct tgc atg gat gcc Glu Thr Ser Cys Thr Lys Asp Asp Arg Ile Tyr Ala Cys Met Asp Ala 275 280 285			864
atc gtc cct ggc aac gtc cct agc gcc tgc ttg cgc gag gcc ctc gac Ile Val Pro Gly Asn Val Pro Ser Ala Cys Leu Arg Glu Ala Leu Asp 290 295 300			912
cag gcg cgc gtc aag ccg ggc gat atc gag atg ctc gag ctc agc gcc Gln Ala Arg Val Lys Pro Gly Asp Ile Glu Met Leu Glu Leu Ser Ala 305 310 315 320			960
gac tcc gcc cgc cac ctc aag gac ccg tcc gtc ctg ccc aag gag ctc Asp Ser Ala Arg His Leu Lys Asp Pro Ser Val Leu Pro Lys Glu Leu 325 330 335			1008
act gcc gag gag gaa atc ggc ggc ctt cag acg atc ctt cgt gac gat Thr Ala Glu Glu Glu Ile Gly Gly Leu Gln Thr Ile Leu Arg Asp Asp 1056			

340

345

350

gac aag ctc ccg cgc aac gtc gca acg ggc agt gtc aag gcc acc gtc 1104
Asp Lys Leu Pro Arg Asn Val Ala Thr Gly Ser Val Lys Ala Thr Val
355 360 365

ggt gac acc ggt tat gcc tct ggt gct gcc agc ctc atc aag gct gcg 1152
Gly Asp Thr Gly Tyr Ala Ser Gly Ala Ala Ser Leu Ile Lys Ala Ala
370 375 380

ctt tgc atc tac aac cgc tac ctg ccc agc aac ggc gac gac tgg gat 1200
Leu Cys Ile Tyr Asn Arg Tyr Leu Pro Ser Asn Gly Asp Asp Trp Asp
385 390 395 400

gaa ccc gcc cct gag gcg ccc tgg gac agc acc ctc ttt gcg tgc cag 1248
Glu Pro Ala Pro Glu Ala Pro Trp Asp Ser Thr Leu Phe Ala Cys Gln
405 410 415

acc tcg cgc gct tgg ctc aag aac cct ggc gag cgt cgc tat gcg gcc 1296
Thr Ser Arg Ala Trp Leu Lys Asn Pro Gly Glu Arg Arg Tyr Ala Ala
420 425 430

gtc tcg ggc gtc tcc gag acg cgc tcg 1323
Val Ser Gly Val Ser Glu Thr Arg Ser
435 440

<210> 22

<211> 441

<212> PRT

<213> Schizochytrium sp.

<400> 22

Ser Ala Arg Cys Gly Gly Glu Ser Asn Met Arg Ile Ala Ile Thr Gly
1 5 10 15

Met Asp Ala Thr Phe Gly Ala Leu Lys Gly Leu Asp Ala Phe Glu Arg
20 25 30

Ala Ile Tyr Thr Gly Ala His Gly Ala Ile Pro Leu Pro Glu Lys Arg
35 40 45

Trp Arg Phe Leu Gly Lys Asp Lys Asp Phe Leu Asp Leu Cys Gly Val
50 55 60

Lys Ala Thr Pro His Gly Cys Tyr Ile Glu Asp Val Glu Val Asp Phe
65 70 75 80

Gln Arg Leu Arg Thr Pro Met Thr Pro Glu Asp Met Leu Leu Pro Gln
85 90 95

Gln Leu Leu Ala Val Thr Thr Ile Asp Arg Ala Ile Leu Asp Ser Gly
100 105 110

Met Lys Lys Gly Gly Asn Val Ala Val Phe Val Gly Leu Gly Thr Asp
115 120 125

Leu Glu Leu Tyr Arg His Arg Ala Arg Val Ala Leu Lys Glu Arg Val
130 135 140

Arg Pro Glu Ala Ser Lys Lys Leu Asn Asp Met Met Gln Tyr Ile Asn
145 150 155 160

Asp Cys Gly Thr Ser Thr Ser Tyr Thr Ser Tyr Ile Gly Asn Leu Val
165 170 175

Ala Thr Arg Val Ser Ser Gln Trp Gly Phe Thr Gly Pro Ser Phe Thr
180 185 190

Ile Thr Glu Gly Asn Asn Ser Val Tyr Arg Cys Ala Glu Leu Gly Lys
195 200 205

Tyr Leu Leu Glu Thr Gly Glu Val Asp Gly Val Val Val Ala Gly Val
210 215 220

Asp Leu Cys Gly Ser Ala Glu Asn Leu Tyr Val Lys Ser Arg Arg Phe
225 230 235 240

Lys Val Ser Thr Ser Asp Thr Pro Arg Ala Ser Phe Asp Ala Ala Ala
245 250 255

Asp Gly Tyr Phe Val Gly Glu Gly Cys Gly Ala Phe Val Leu Lys Arg
260 265 270

Glu Thr Ser Cys Thr Lys Asp Asp Arg Ile Tyr Ala Cys Met Asp Ala
275 280 285

Ile Val Pro Gly Asn Val Pro Ser Ala Cys Leu Arg Glu Ala Leu Asp
290 295 300

Gln Ala Arg Val Lys Pro Gly Asp Ile Glu Met Leu Glu Leu Ser Ala
305 310 315 320

Asp Ser Ala Arg His Leu Lys Asp Pro Ser Val Leu Pro Lys Glu Leu
325 330 335

Thr Ala Glu Glu Glu Ile Gly Gly Leu Gln Thr Ile Leu Arg Asp Asp
340 345 350

Asp Lys Leu Pro Arg Asn Val Ala Thr Gly Ser Val Lys Ala Thr Val
355 360 365

Gly Asp Thr Gly Tyr Ala Ser Gly Ala Ala Ser Leu Ile Lys Ala Ala
370 375 380

Leu Cys Ile Tyr Asn Arg Tyr Leu Pro Ser Asn Gly Asp Asp Trp Asp
385 390 395 400

Glu Pro Ala Pro Glu Ala Pro Trp Asp Ser Thr Leu Phe Ala Cys Gln
405 410 415

Thr Ser Arg Ala Trp Leu Lys Asn Pro Gly Glu Arg Arg Tyr Ala Ala
420 425 430

Val Ser Gly Val Ser Glu Thr Arg Ser
435 440

<210> 23
<211> 1500
<212> DNA
<213> Schizochytrium sp.

<220>
<221> CDS
<222> (1)..(1500)

<400> 23
tgc tat tcc gtg ctc ctc tcc gaa gcc gag ggc cac tac gag cgc gag 48
Cys Tyr Ser Val Leu Leu Ser Glu Ala Glu Gly His Tyr Glu Arg Glu
1 5 10 15

aac cgc atc tcg ctc gac gag gag gcg ccc aag ctc att gtg ctt cgc 96
Asn Arg Ile Ser Leu Asp Glu Glu Ala Pro Lys Leu Ile Val Leu Arg
20 25 30

gcc gac tcc cac gag gag atc ctt ggt cgc ctc gac aag atc cgc gag 144
Ala Asp Ser His Glu Glu Ile Leu Gly Arg Leu Asp Lys Ile Arg Glu
35 40 45

cgc ttc ttg cag ccc acg ggc gcc gcc ccg cgc gag tcc gag ctc aag 192
Arg Phe Leu Gln Pro Thr Gly Ala Ala Pro Arg Glu Ser Glu Leu Lys
50 55 60

gcg cag gcc cgc cgc atc ttc ctc gag ctc ctc ggc gag acc ctt gcc 240
Ala Gln Ala Arg Arg Ile Phe Leu Glu Leu Gly Glu Thr Leu Ala
65 70 75 80

cag gat gcc gct tct tca ggc tcg caa aag ccc ctc gct ctc agc ctc 288
Gln Asp Ala Ala Ser Ser Gly Ser Gln Lys Pro Leu Ala Leu Ser Leu
85 90 95

gtc tcc acg ccc tcc aag ctc cag cgc gag gtc gag ctc gcg gcc aag 336
Val Ser Thr Pro Ser Lys Leu Gln Arg Glu Val Glu Leu Ala Ala Lys
100 105 110

ggt atc ccg cgc tgc ctc aag atg cgc cgc gat tgg agc tcc cct gct 384

Gly Ile Pro Arg Cys Leu Lys Met Arg Arg Asp Trp Ser Ser Pro Ala	
115 120 125	
ggc agc cgc tac gcg cct gag ccg ctc gcc agc gac cgc gtc gcc ttc	432
Gly Ser Arg Tyr Ala Pro Glu Pro Leu Ala Ser Asp Arg Val Ala Phe	
130 135 140	
atg tac ggc gaa ggt cgc agc cct tac tac ggc atc acc caa gac att	480
Met Tyr Gly Glu Gly Arg Ser Pro Tyr Tyr Gly Ile Thr Gln Asp Ile	
145 150 155 160	
cac cgc att tgg ccc gaa ctc cac gag gtc atc aac gaa aag acg aac	528
His Arg Ile Trp Pro Glu Leu His Glu Val Ile Asn Glu Lys Thr Asn	
165 170 175	
cgt ctc tgg gcc gaa ggc gac cgc tgg gtc atg ccg cgc gcc agc ttc	576
Arg Leu Trp Ala Glu Gly Asp Arg Trp Val Met Pro Arg Ala Ser Phe	
180 185 190	
aag tcg gag ctc gag agc cag cag caa gag ttt gat cgc aac atg att	624
Lys Ser Glu Leu Glu Ser Gln Gln Gln Glu Phe Asp Arg Asn Met Ile	
195 200 205	
gaa atg ttc cgt ctt gga atc ctc acc tca att gcc ttc acc aat ctg	672
Glu Met Phe Arg Leu Gly Ile Leu Thr Ser Ile Ala Phe Thr Asn Leu	
210 215 220	
gcg cgc gac gtt ctc aac atc acg ccc aag gcc gcc ttt ggc ctc agt	720
Ala Arg Asp Val Leu Asn Ile Thr Pro Lys Ala Ala Phe Gly Leu Ser	
225 230 235 240	
ctt ggc gag att tcc atg att ttt gcc ttt tcc aag aag aac ggt ctc	768
Leu Gly Glu Ile Ser Met Ile Phe Ala Phe Ser Lys Lys Asn Gly Leu	
245 250 255	
atc tcc gac cag ctc acc aag gat ctt cgc gag tcc gac gtg tgg aac	816
Ile Ser Asp Gln Leu Thr Lys Asp Leu Arg Glu Ser Asp Val Trp Asn	
260 265 270	
aag gct ctg gcc gtt gaa ttt aat gcg ctg cgc gag gcc tgg ggc att	864
Lys Ala Leu Ala Val Glu Phe Asn Ala Leu Arg Glu Ala Trp Gly Ile	
275 280 285	
cca cag agt gtc ccc aag gac gag ttc tgg caa ggc tac att gtg cgc	912
Pro Gln Ser Val Pro Lys Asp Glu Phe Trp Gln Gly Tyr Ile Val Arg	

290	295	300	
ggc acc aag cag gat atc gag gcg gcc atc gcc ccg gac agc aag tac			960
Gly Thr Lys Gln Asp Ile Glu Ala Ala Ile Ala Pro Asp Ser Lys Tyr			
305	310	315	320
gtg cgc ctc acc atc atc aat gat gcc aac acc gcc ctc att agc ggc			1008
Val Arg Leu Thr Ile Ile Asn Asp Ala Asn Thr Ala Leu Ile Ser Gly			
	325	330	335
aag ccc gac gcc tgc aag gct gcg atc gcg cgt ctc ggt ggc aac att			1056
Lys Pro Asp Ala Cys Lys Ala Ala Ile Ala Arg Leu Gly Gly Asn Ile			
	340	345	350
cct gcg ctt ccc gtg acc cag ggc atg tgc ggc cac tgc ccc gag gtg			1104
Pro Ala Leu Pro Val Thr Gln Gly Met Cys Gly His Cys Pro Glu Val			
	355	360	365
gga cct tat acc aag gat atc gcc aag atc cat gcc aac ctt gag ttc			1152
Gly Pro Tyr Thr Lys Asp Ile Ala Lys Ile His Ala Asn Leu Glu Phe			
	370	375	380
ccc gtt gtc gac ggc ctt gac ctc tgg acc aca atc aac cag aag cgc			1200
Pro Val Val Asp Gly Leu Asp Leu Trp Thr Thr Ile Asn Gln Lys Arg			
	385	390	400
ctc gtg cca cgc gcc acg ggc gcc aag gac gaa tgg gcc cct tct tcc			1248
Leu Val Pro Arg Ala Thr Gly Ala Lys Asp Glu Trp Ala Pro Ser Ser			
	405	410	415
ttt ggc gag tac gcc ggc cag ctc tac gag aag cag gct aac ttc ccc			1296
Phe Gly Glu Tyr Ala Gly Gln Leu Tyr Glu Lys Gln Ala Asn Phe Pro			
	420	425	430
caa atc gtc gag acc att tac aag caa aac tac gac gtc ttt gtc gag			1344
Gln Ile Val Glu Thr Ile Tyr Lys Gln Asn Tyr Asp Val Phe Val Glu			
	435	440	445
gtt ggg ccc aac aac cac cgt agc acc gca gtg cgc acc acg ctt ggt			1392
Val Gly Pro Asn Asn His Arg Ser Thr Ala Val Arg Thr Thr Leu Gly			
	450	455	460
ccc cag cgc aac cac ctt gct ggc gcc atc gac aag cag aac gag gat			1440
Pro Gln Arg Asn His Leu Ala Gly Ala Ile Asp Lys Gln Asn Glu Asp			

465 470 475 480

gct tgg acg acc atc gtc aag ctt gtg gct tcg ctc aag gcc cac ctt 1488
 Ala Trp Thr Thr Ile Val Lys Leu Val Ala Ser Leu Lys Ala His Leu
 485 490 495

gtt cct ggc gtc 1500
 Val Pro Gly Val
 500

<210> 24
 <211> 500
 <212> PRT
 <213> Schizochytrium sp.

<400> 24

Cys Tyr Ser Val Leu Leu Ser Glu Ala Glu Gly His Tyr Glu Arg Glu
 1 5 10 15

Asn Arg Ile Ser Leu Asp Glu Glu Ala Pro Lys Leu Ile Val Leu Arg
 20 25 30

Ala Asp Ser His Glu Glu Ile Leu Gly Arg Leu Asp Lys Ile Arg Glu
 35 40 45

Arg Phe Leu Gln Pro Thr Gly Ala Ala Pro Arg Glu Ser Glu Leu Lys
 50 55 60

Ala Gln Ala Arg Arg Ile Phe Leu Glu Leu Leu Gly Glu Thr Leu Ala
 65 70 75 80

Gln Asp Ala Ala Ser Ser Gly Ser Gln Lys Pro Leu Ala Leu Ser Leu
 85 90 95

Val Ser Thr Pro Ser Lys Leu Gln Arg Glu Val Glu Leu Ala Ala Lys
 100 105 110

Gly Ile Pro Arg Cys Leu Lys Met Arg Arg Asp Trp Ser Ser Pro Ala
115 120 125

Gly Ser Arg Tyr Ala Pro Glu Pro Leu Ala Ser Asp Arg Val Ala Phe
130 135 140

Met Tyr Gly Glu Gly Arg Ser Pro Tyr Tyr Gly Ile Thr Gln Asp Ile
145 150 155 160

His Arg Ile Trp Pro Glu Leu His Glu Val Ile Asn Glu Lys Thr Asn
165 170 175

Arg Leu Trp Ala Glu Gly Asp Arg Trp Val Met Pro Arg Ala Ser Phe
180 185 190

Lys Ser Glu Leu Glu Ser Gln Gln Gln Glu Phe Asp Arg Asn Met Ile
195 200 205

Glu Met Phe Arg Leu Gly Ile Leu Thr Ser Ile Ala Phe Thr Asn Leu
210 215 220

Ala Arg Asp Val Leu Asn Ile Thr Pro Lys Ala Ala Phe Gly Leu Ser
225 230 235 240

Leu Gly Glu Ile Ser Met Ile Phe Ala Phe Ser Lys Lys Asn Gly Leu
245 250 255

Ile Ser Asp Gln Leu Thr Lys Asp Leu Arg Glu Ser Asp Val Trp Asn
260 265 270

Lys Ala Leu Ala Val Glu Phe Asn Ala Leu Arg Glu Ala Trp Gly Ile
275 280 285

Pro Gln Ser Val Pro Lys Asp Glu Phe Trp Gln Gly Tyr Ile Val Arg
290 295 300

Gly Thr Lys Gln Asp Ile Glu Ala Ala Ile Ala Pro Asp Ser Lys Tyr

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

<210> 25
 <211> 1530
 <212> DNA
 <213> Schizochytrium sp.

<220>
 <221> CDS
 <222> (1)..(1530)

<400> 25
 ctg ctc gat ctc gac agt atg ctt gcg ctg agc tct gcc agt gcc tcc 48
 Leu Leu Asp Leu Asp Ser Met Leu Ala Leu Ser Ser Ala Ser Ala Ser
 1 5 10 15

ggc aac ctt gtt gag act gcg cct agc gac gcc tcg gtc att gtg ccg 96
 Gly Asn Leu Val Glu Thr Ala Pro Ser Asp Ala Ser Val Ile Val Pro
 20 25 30

ccc tgc aac att gcg gat ctc ggc agc cgc gcc ttc atg aaa acg tac 144
 Pro Cys Asn Ile Ala Asp Leu Gly Ser Arg Ala Phe Met Lys Thr Tyr
 35 40 45

ggt gtt tcg gcg cct ctg tac acg ggc gcc atg gcc aag ggc att gcc 192
 Gly Val Ser Ala Pro Leu Tyr Thr Gly Ala Met Ala Lys Gly Ile Ala
 50 55 60

tct gcg gac ctc gtc att gcc gcc ggc cgc cag ggc atc ctt gcg tcc 240
 Ser Ala Asp Leu Val Ile Ala Ala Gly Arg Gln Gly Ile Leu Ala Ser
 65 70 75 80

ttt ggc gcc ggc gga ctt ccc atg cag gtt gtg cgt gag tcc atc gaa 288
 Phe Gly Ala Gly Gly Leu Pro Met Gln Val Val Arg Glu Ser Ile Glu
 85 90 95

aag att cag gcc gcc ctg ccc aat ggc ccg tac gct gtc aac ctt atc 336
 Lys Ile Gln Ala Ala Leu Pro Asn Gly Pro Tyr Ala Val Asn Leu Ile
 100 105 110

cat tct ccc ttt gac agc aac ctc gaa aag ggc aat gtc gat ctc ttc 384
 His Ser Pro Phe Asp Ser Asn Leu Glu Lys Gly Asn Val Asp Leu Phe
 115 120 125

ctc gag aag ggt gtc acc ttt gtc gag gcc tcg gcc ttt atg acg ctc 432

Leu Glu Lys Gly Val Thr Phe Val Glu Ala Ser Ala Phe Met Thr Leu	
130 135 140	
acc ccg cag gtc gtg cgg tac cgc gcg gct ggc ctc acg cgc aac gcc	480
Thr Pro Gln Val Val Arg Tyr Arg Ala Ala Gly Leu Thr Arg Asn Ala	
145 150 155 160	
gac ggc tcg gtc aac atc cgc aac cgt atc att ggc aag gtc tcg cgc	528
Asp Gly Ser Val Asn Ile Arg Asn Arg Ile Ile Gly Lys Val Ser Arg	
165 170 175	
acc gag ctc gcc gag atg ttc atg cgt cct gcg ccc gag cac ctt ctt	576
Thr Glu Leu Ala Glu Met Phe Met Arg Pro Ala Pro Glu His Leu Leu	
180 185 190	
cag aag ctc att gct tcc ggc gag atc aac cag gag cag gcc gag ctc	624
Gln Lys Leu Ile Ala Ser Gly Glu Ile Asn Gln Glu Gln Ala Glu Leu	
195 200 205	
gcc cgc cgt gtt ccc gtc gct gac gac atc gcg gtc gaa gct gac tcg	672
Ala Arg Arg Val Pro Val Ala Asp Asp Ile Ala Val Glu Ala Asp Ser	
210 215 220	
ggt ggc cac acc gac aac cgc ccc atc cac gtc att ctg ccc ctc atc	720
Gly Gly His Thr Asp Asn Arg Pro Ile His Val Ile Leu Pro Leu Ile	
225 230 235 240	
atc aac ctt cgc gac cgc ctt cac cgc gag tgc ggc tac ccg gcc aac	768
Ile Asn Leu Arg Asp Arg Leu His Arg Glu Cys Gly Tyr Pro Ala Asn	
245 250 255	
ctt cgc gtc cgt gtg ggc gcc ggc ggt ggc att ggg tgc ccc cag gcg	816
Leu Arg Val Arg Val Gly Ala Gly Gly Gly Ile Gly Cys Pro Gln Ala	
260 265 270	
gcg ctg gcc acc ttc aac atg ggt gcc tcc ttt att gtc acc ggc acc	864
Ala Leu Ala Thr Phe Asn Met Gly Ala Ser Phe Ile Val Thr Gly Thr	
275 280 285	
gtg aac cag gtc gcc aag cag tcg ggc acg tgc gac aat gtg cgc aag	912
Val Asn Gln Val Ala Lys Gln Ser Gly Thr Cys Asp Asn Val Arg Lys	
290 295 300	
cag ctc gcg aag gcc act tac tcg gac gta tgc atg gcc ccg gct gcc	960
Gln Leu Ala Lys Ala Thr Tyr Ser Asp Val Cys Met Ala Pro Ala Ala	

305	310	315	320	
gac atg ttc gag gaa ggc gtc aag ctt cag gtc ctc aag aag gga acc				1008
Asp Met Phe Glu Glu Gly Val Lys Leu Gln Val Leu Lys Lys Gly Thr				
	325	330	335	
atg ttt ccc tcg cgc gcc aac aag ctc tac gag ctc ttt tgc aag tac				1056
Met Phe Pro Ser Arg Ala Asn Lys Leu Tyr Glu Leu Phe Cys Lys Tyr				
	340	345	350	
gac tcg ttc gag tcc atg ccc ccc gca gag ctt gcg cgc gtc gag aag				1104
Asp Ser Phe Glu Ser Met Pro Pro Ala Glu Leu Ala Arg Val Glu Lys				
	355	360	365	
cgc atc ttc agc cgc gcg ctc gaa gag gtc tgg gac gag acc aaa aac				1152
Arg Ile Phe Ser Arg Ala Leu Glu Glu Val Trp Asp Glu Thr Lys Asn				
	370	375	380	
ttt tac att aac cgt ctt cac aac ccg gag aag atc cag cgc gcc gag				1200
Phe Tyr Ile Asn Arg Leu His Asn Pro Glu Lys Ile Gln Arg Ala Glu				
	385	390	395	400
cgc gac ccc aag ctc aag atg tcg ctg tgc ttt cgc tgg tac ctg agc				1248
Arg Asp Pro Lys Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Ser				
	405	410	415	
ctg gcg agc cgc tgg gcc aac act gga gct tcc gat cgc gtc atg gac				1296
Leu Ala Ser Arg Trp Ala Asn Thr Gly Ala Ser Asp Arg Val Met Asp				
	420	425	430	
tac cag gtc tgg tgc ggt cct gcc att ggt tcc ttc aac gat ttc atc				1344
Tyr Gln Val Trp Cys Gly Pro Ala Ile Gly Ser Phe Asn Asp Phe Ile				
	435	440	445	
aag gga act tac ctt gat ccg gcc gtc gca aac gag tac ccg tgc gtc				1392
Lys Gly Thr Tyr Leu Asp Pro Ala Val Ala Asn Glu Tyr Pro Cys Val				
	450	455	460	
gtt cag att aac aag cag atc ctt cgt gga gcg tgc ttc ttg cgc cgt				1440
Val Gln Ile Asn Lys Gln Ile Leu Arg Gly Ala Cys Phe Leu Arg Arg				
	465	470	475	480
ctc gaa att ctg cgc aac gca cgc ctt tcc gat ggc gct gcc gct ctt				1488
Leu Glu Ile Leu Arg Asn Ala Arg Leu Ser Asp Gly Ala Ala Ala Leu				

485

490

495

gtg gcc agc atc gat gac aca tac gtc ccg gcc gag aag ctg 1530
Val Ala Ser Ile Asp Asp Thr Tyr Val Pro Ala Glu Lys Leu
500 505 510

<210> 26

<211> 510

<212> PRT

<213> Schizochytrium sp.

<400> 26

Leu Leu Asp Leu Asp Ser Met Leu Ala Leu Ser Ser Ala Ser Ala Ser
1 5 10 15

Gly Asn Leu Val Glu Thr Ala Pro Ser Asp Ala Ser Val Ile Val Pro
20 25 30

Pro Cys Asn Ile Ala Asp Leu Gly Ser Arg Ala Phe Met Lys Thr Tyr
35 40 45

Gly Val Ser Ala Pro Leu Tyr Thr Gly Ala Met Ala Lys Gly Ile Ala
50 55 60

Ser Ala Asp Leu Val Ile Ala Ala Gly Arg Gln Gly Ile Leu Ala Ser
65 70 75 80

Phe Gly Ala Gly Gly Leu Pro Met Gln Val Val Arg Glu Ser Ile Glu
85 90 95

Lys Ile Gln Ala Ala Leu Pro Asn Gly Pro Tyr Ala Val Asn Leu Ile
100 105 110

His Ser Pro Phe Asp Ser Asn Leu Glu Lys Gly Asn Val Asp Leu Phe
115 120 125

Leu Glu Lys Gly Val Thr Phe Val Glu Ala Ser Ala Phe Met Thr Leu
130 135 140

Thr Pro Gln Val Val Arg Tyr Arg Ala Ala Gly Leu Thr Arg Asn Ala
145 150 155 160

Asp Gly Ser Val Asn Ile Arg Asn Arg Ile Ile Gly Lys Val Ser Arg
165 170 175

Thr Glu Leu Ala Glu Met Phe Met Arg Pro Ala Pro Glu His Leu Leu
180 185 190

Gln Lys Leu Ile Ala Ser Gly Glu Ile Asn Gln Glu Gln Ala Glu Leu
195 200 205

Ala Arg Arg Val Pro Val Ala Asp Asp Ile Ala Val Glu Ala Asp Ser
210 215 220

Gly Gly His Thr Asp Asn Arg Pro Ile His Val Ile Leu Pro Leu Ile
225 230 235 240

Ile Asn Leu Arg Asp Arg Leu His Arg Glu Cys Gly Tyr Pro Ala Asn
245 250 255

Leu Arg Val Arg Val Gly Ala Gly Gly Gly Ile Gly Cys Pro Gln Ala
260 265 270

Ala Leu Ala Thr Phe Asn Met Gly Ala Ser Phe Ile Val Thr Gly Thr
275 280 285

Val Asn Gln Val Ala Lys Gln Ser Gly Thr Cys Asp Asn Val Arg Lys
290 295 300

Gln Leu Ala Lys Ala Thr Tyr Ser Asp Val Cys Met Ala Pro Ala Ala
305 310 315 320

Asp Met Phe Glu Glu Gly Val Lys Leu Gln Val Leu Lys Lys Gly Thr

325

330

335

Met Phe Pro Ser Arg Ala Asn Lys Leu Tyr Glu Leu Phe Cys Lys Tyr
340 345 350

Asp Ser Phe Glu Ser Met Pro Pro Ala Glu Leu Ala Arg Val Glu Lys
355 360 365

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

Phe Tyr Ile Asn Arg Leu His Asn Pro Glu Lys Ile Gln Arg Ala Glu
385 390 395 400

Arg Asp Pro Lys Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Ser
405 410 415

Leu Ala Ser Arg Trp Ala Asn Thr Gly Ala Ser Asp Arg Val Met Asp
420 425 430

Tyr Gln Val Trp Cys Gly Pro Ala Ile Gly Ser Phe Asn Asp Phe Ile
435 440 445

Lys Gly Thr Tyr Leu Asp Pro Ala Val Ala Asn Glu Tyr Pro Cys Val
450 455 460

Val Gln Ile Asn Lys Gln Ile Leu Arg Gly Ala Cys Phe Leu Arg Arg
465 470 475 480

Leu Glu Ile Leu Arg Asn Ala Arg Leu Ser Asp Gly Ala Ala Ala Leu
485 490 495

Val Ala Ser Ile Asp Asp Thr Tyr Val Pro Ala Glu Lys Leu
500 505 510

<210> 27
<211> 4512
<212> DNA

<213> Schizochytrium sp.

 $\langle 220 \rangle$

<221> CDS

$\langle 222 \rangle$ (1).. (4512)

<400> 27

atg gcg ctc cgt gtc aag acg aac aag aag cca tgc tgg gag atg acc 48
Met Ala Leu Arg Val Lys Thr Asn Lys Lys Pro Cys Trp Glu Met Thr
1 5 10 15

aag gag gag ctg acc agc ggc aag acc gag gtg ttc aac tat gag gaa 96
Lys Glu Glu Leu Thr Ser Gly Lys Thr Glu Val Phe Asn Tyr Glu Glu
20 25 30

ctc ctc gag ttc gca gag ggc gac atc gcc aag gtc ttc gga ccc gag 144
Leu Leu Glu Phe Ala Glu Gly Asp Ile Ala Lys Val Phe Gly Pro Glu
35 40 45

ttc gcc gtc atc gac aag tac ccg cgc cgc gtg cgc ctg ccc gcc cgc 192
Phe Ala Val Ile Asp Lys Tyr Pro Arg Arg Val Arg Leu Pro Ala Arg
50 55 60

gag tac ctg ctc gtg acc cgc gtc acc ctc atg gac gcc gag gtc aac 240
Glu Tyr Leu Leu Val Thr Arg Val Thr Leu Met Asp Ala Glu Val Asn
65 70 75 80

aac tac cgc gtc gcc gcc cgc atg gtc acc gag tac gat ctc ccc gtc 288
Asn Tyr Arg Val Gly Ala Arg Met Val Thr Glu Tyr Asp Leu Pro Val
85 90 95

aac gga gag ctc tcc gag ggc gga gac tgc ccc tgg gcc gtc ctg gtc 336
Asn Gly Glu Leu Ser Glu Gly Gly Asp Cys Pro Trp Ala Val Leu Val
100 105 110

gag agt ggc cag tgc gat ctc atg ctc atc tcc tac atg ggc att gac 384
Glu Ser Gly Gln Cys Asp Leu Met Leu Ile Ser Tyr Met Gly Ile Asp
115 120 125

t t c c a g a a c c a g g g c g a c c g c g t c t a c c g c c t g c t c a a c a c c a c g c t c 432
Phe Gln Asn Gln Gly Asp Arg Val Tyr Arg Leu Leu Asn Thr Thr Leu
130 135 140

acc ttt tac ggc gtg gcc cac gag ggc gag acc ctc gag tac gac att 480

Thr Phe Tyr Gly Val Ala His Glu Gly Glu Thr Leu Glu Tyr Asp Ile	
145 150 155 160	
cgc gtc acc ggc ttc gcc aag cgt ctc gac ggc ggc atc tcc atg ttc	528
Arg Val Thr Gly Phe Ala Lys Arg Leu Asp Gly Gly Ile Ser Met Phe	
165 170 175	
ttc ttc gag tac gac tgc tac gtc aac ggc cgc ctc ctc atc gag atg	576
Phe Phe Glu Tyr Asp Cys Tyr Val Asn Gly Arg Leu Leu Ile Glu Met	
180 185 190	
cgc gat ggc tgc gcc ggc ttc ttc acc aac gag gag ctc gac gcc ggc	624
Arg Asp Gly Cys Ala Gly Phe Phe Thr Asn Glu Glu Leu Asp Ala Gly	
195 200 205	
aag ggc gtc gtc ttc acc cgc ggc gac ctc gcc gcc cgc gcc aag atc	672
Lys Gly Val Val Phe Thr Arg Gly Asp Leu Ala Ala Arg Ala Lys Ile	
210 215 220	
cca aag cag gac gtc tcc ccc tac gcc gtc gcc ccc tgc ctc cac aag	720
Pro Lys Gln Asp Val Ser Pro Tyr Ala Val Ala Pro Cys Leu His Lys	
225 230 235 240	
acc aag ctc aac gaa aag gag atg cag acc ctc gtc gac aag gac tgg	768
Thr Lys Leu Asn Glu Lys Glu Met Gln Thr Leu Val Asp Lys Asp Trp	
245 250 255	
gca tcc gtc ttt ggc tcc aag aac ggc atg ccg gaa atc aac tac aaa	816
Ala Ser Val Phe Gly Ser Lys Asn Gly Met Pro Glu Ile Asn Tyr Lys	
260 265 270	
ctc tgc gcg cgt aag atg ctc atg att gac cgc gtc acc agc att gac	864
Leu Cys Ala Arg Lys Met Leu Met Ile Asp Arg Val Thr Ser Ile Asp	
275 280 285	
cac aag ggc ggt gtc tac ggc ctc ggt cag ctc gtc ggt gaa aag atc	912
His Lys Gly Gly Val Tyr Gly Leu Gly Gln Leu Val Gly Glu Lys Ile	
290 295 300	
ctc gag cgc gac cac tgg tac ttt ccc tgc cac ttt gtc aag gat cag	960
Leu Glu Arg Asp His Trp Tyr Phe Pro Cys His Phe Val Lys Asp Gln	
305 310 315 320	
gtc atg gcc gga tcc ctc gtc tcc gac ggc tgc agc cag atg ctc aag	1008
Val Met Ala Gly Ser Leu Val Ser Asp Gly Cys Ser Gln Met Leu Lys	

325	330	335	
atg tac atg atc tgg ctc ggc ctc cac ctc acc acc gga ccc ttt gac Met Tyr Met Ile Trp Leu Gly Leu His Leu Thr Thr Gly Pro Phe Asp 340 345 350			1056
ttc cgc ccg gtc aac ggc cac ccc aac aag gtc cgc tgc cgc ggc caa Phe Arg Pro Val Asn Gly His Pro Asn Lys Val Arg Cys Arg Gly Gln 355 360 365			1104
atc tcc ccg cac aag ggc aag ctc gtc tac gtc atg gag atc aag gag Ile Ser Pro His Lys Gly Lys Leu Val Tyr Val Met Glu Ile Lys Glu 370 375 380			1152
atg ggc ttc gac gag gac aac gac ccg tac gcc att gcc gac gtc aac Met Gly Phe Asp Glu Asp Asn Asp Pro Tyr Ala Ile Ala Asp Val Asn 385 390 395 400			1200
atc att gat gtc gac ttc gaa aag ggc cag gac ttt agc ctc gac cgc Ile Ile Asp Val Asp Phe Glu Lys Gly Gln Asp Phe Ser Leu Asp Arg 405 410 415			1248
atc agc gac tac ggc aag ggc gac ctc aac aag aag atc gtc gtc gac Ile Ser Asp Tyr Gly Lys Gly Asp Leu Asn Lys Lys Ile Val Val Asp 420 425 430			1296
ttt aag ggc atc gct ctc aag atg cag aag cgc tcc acc aac aag aac Phe Lys Gly Ile Ala Leu Lys Met Gln Lys Arg Ser Thr Asn Lys Asn 435 440 445			1344
ccc tcc aag gtt cag ccc gtc ttt gcc aac ggc gcc gcc act gtc ggc Pro Ser Lys Val Gln Pro Val Phe Ala Asn Gly Ala Ala Thr Val Gly 450 455 460			1392
ccc gag gcc tcc aag gct tcc tcc ggc gcc agc gcc agc gcc agc gcc Pro Glu Ala Ser Lys Ala Ser Ser Gly Ala Ser Ala Ser Ala Ser Ala 465 470 475 480			1440
gcc ccg gcc aag cct gcc ttc agc gcc gat gtt ctt gcg ccc aag ccc Ala Pro Ala Lys Pro Ala Phe Ser Ala Asp Val Leu Ala Pro Lys Pro 485 490 495			1488
gtt gcc ctt ccc gag cac atc ctc aag ggc gac gcc ctc gcc ccc aag Val Ala Leu Pro Glu His Ile Leu Lys Gly Asp Ala Leu Ala Pro Lys 1536			

500	505	510	
gag atg tcc tgg cac ccc atg gcc cgc atc ccg ggc aac ccg acg ccc Glu Met Ser Trp His Pro Met Ala Arg Ile Pro Gly Asn Pro Thr Pro 515 520 525			1584
tct ttt gcg ccc tcg gcc tac aag ccg cgc aac atc gcc ttt acg ccc Ser Phe Ala Pro Ser Ala Tyr Lys Pro Arg Asn Ile Ala Phe Thr Pro 530 535 540			1632
ttc ccc ggc aac ccc aac gat aac gac cac acc ccg ggc aag atg ccg Phe Pro Gly Asn Pro Asn Asp Asn Asp His Thr Pro Gly Lys Met Pro 545 550 555 560			1680
ctc acc tgg ttc aac atg gcc gag ttc atg gcc ggc aag gtc agc atg Leu Thr Trp Phe Asn Met Ala Glu Phe Met Ala Gly Lys Val Ser Met 565 570 575			1728
tgc ctc ggc ccc gag ttc gcc aag ttc gac gac tcg aac acc agc cgc Cys Leu Gly Pro Glu Phe Ala Lys Phe Asp Asp Ser Asn Thr Ser Arg 580 585 590			1776
agc ccc gct tgg gac ctc gct ctc gtc acc cgc gcc gtg tct gtg tct Ser Pro Ala Trp Asp Leu Ala Leu Val Thr Arg Ala Val Ser Val Ser 595 600 605			1824
gac ctc aag cac gtc aac tac cgc aac atc gac ctc gac ccc tcc aag Asp Leu Lys His Val Asn Tyr Arg Asn Ile Asp Leu Asp Pro Ser Lys 610 615 620			1872
ggt acc atg gtc ggc gag ttc gac tgc ccc gcg gac gcc tgg ttc tac Gly Thr Met Val Gly Glu Phe Asp Cys Pro Ala Asp Ala Trp Phe Tyr 625 630 635 640			1920
aag ggc gcc tgc aac gat gcc cac atg ccg tac tcg atc ctc atg gag Lys Gly Ala Cys Asn Asp Ala His Met Pro Tyr Ser Ile Leu Met Glu 645 650 655			1968
atc gcc ctc cag acc tcg ggt gtg ctc acc tcg gtg ctc aag gcg ccc Ile Ala Leu Gln Thr Ser Gly Val Leu Thr Ser Val Leu Lys Ala Pro 660 665 670			2016
ctg acc atg gag aag gac gac atc ctc ttc cgc aac ctc gac gcc aac Leu Thr Met Glu Lys Asp Asp Ile Leu Phe Arg Asn Leu Asp Ala Asn			2064

675	680	685	
gcc gag ttc gtg cgc gcc gac ctc gac tac cgc ggc aag act atc cgc Ala Glu Phe Val Arg Ala Asp Leu Asp Tyr Arg Gly Lys Thr Ile Arg 690 695 700			2112
aac gtc acc aag tgc act ggc tac agc atg ctc ggc gag atg ggc gtc Asn Val Thr Lys Cys Thr Gly Tyr Ser Met Leu Gly Glu Met Gly Val 705 710 715 720			2160
cac cgc ttc acc ttt gag ctc tac gtc gat gat gtg ctc ttt tac aag His Arg Phe Thr Phe Glu Leu Tyr Val Asp Asp Val Leu Phe Tyr Lys 725 730 735			2208
ggc tcg acc tcg ttc ggc tgg ttc gtg ccc gag gtc ttt gcc gcc cag Gly Ser Thr Ser Phe Gly Trp Phe Val Pro Glu Val Phe Ala Ala Gln 740 745 750			2256
gcc ggc ctc gac aac ggc cgc aag tcg gag ccc tgg ttc att gag aac Ala Gly Leu Asp Asn Gly Arg Lys Ser Glu Pro Trp Phe Ile Glu Asn 755 760 765			2304
aag gtt ccg gcc tcg cag gtc tcc tcc ttt gac gtg cgc ccc aac ggc Lys Val Pro Ala Ser Gln Val Ser Ser Phe Asp Val Arg Pro Asn Gly 770 775 780			2352
agc ggc cgc acc gcc atc ttc gcc aac gcc ccc agc ggc gcc cag ctc Ser Gly Arg Thr Ala Ile Phe Ala Asn Ala Pro Ser Gly Ala Gln Leu 785 790 795 800			2400
aac cgc cgc acg gac cag ggc cag tac ctc gac gcc gtc gac att gtc Asn Arg Arg Thr Asp Gln Gly Gln Tyr Leu Asp Ala Val Asp Ile Val 805 810 815			2448
tcc ggc agc ggc aag aag agc ctc ggc tac gcc cac ggt tcc aag acg Ser Gly Ser Gly Lys Lys Ser Leu Gly Tyr Ala His Gly Ser Lys Thr 820 825 830			2496
gtc aac ccg aac gac tgg ttc ttc tcg tgc cac ttt tgg ttt gac tcg Val Asn Pro Asn Asp Trp Phe Phe Ser Cys His Phe Trp Phe Asp Ser 835 840 845			2544
gtc atg ccc gga agt ctc ggt gtc gag tcc atg ttc cag ctc gtc gag Val Met Pro Gly Ser Leu Gly Val Glu Ser Met Phe Gln Leu Val Glu			2592

850	855	860	
gcc atc gcc gcc cac gag gat ctc gct ggc aaa gca cgg cat tgc caa			2640
Ala Ile Ala Ala His Glu Asp Leu Ala Gly Lys Ala Arg His Cys Gln			
865	870	875	880
ccc cac ctt tgt gca cgc ccc cgg gca aga tca agc tgg aag tac cgc			2688
Pro His Leu Cys Ala Arg Pro Arg Ala Arg Ser Ser Trp Lys Tyr Arg			
	885	890	895
ggc cag ctc acg ccc aag agc aag aag atg gac tcg gag gtc cac atc			2736
Gly Gln Leu Thr Pro Lys Ser Lys Lys Met Asp Ser Glu Val His Ile			
	900	905	910
gtg tcc gtg gac gcc cac gac ggc gtt gtc gac ctc gtc gcc gac ggc			2784
Val Ser Val Asp Ala His Asp Gly Val Val Asp Leu Val Ala Asp Gly			
	915	920	925
ttc ctc tgg gcc gac agc ctc cgc gtc tac tcg gtg agc aac att cgc			2832
Phe Leu Trp Ala Asp Ser Leu Arg Val Tyr Ser Val Ser Asn Ile Arg			
	930	935	940
gtg cgc atc gcc tcc ggt gag gcc cct gcc gcc gcc tcc tcc gcc gcc			2880
Val Arg Ile Ala Ser Gly Glu Ala Pro Ala Ala Ala Ser Ser Ala Ala			
945	950	955	960
tct gtg ggc tcc tcg gct tcg tcc gtc gag cgc acg cgc tcg agc ccc			2928
Ser Val Gly Ser Ser Ala Ser Ser Val Glu Arg Thr Arg Ser Ser Pro			
	965	970	975
gct gtc gcc tcc ggc ccg gcc cag acc atc gac ctc aag cag ctc aag			2976
Ala Val Ala Ser Gly Pro Ala Gln Thr Ile Asp Leu Lys Gln Leu Lys			
	980	985	990
acc gag ctc ctc gag ctc gat gcc ccg ctc tac ctc tcg cag gac ccg			3024
Thr Glu Leu Leu Glu Leu Asp Ala Pro Leu Tyr Leu Ser Gln Asp Pro			
	995	1000	1005
acc agc ggc cag ctc aag aag cac acc gac gtg gcc tcc ggc cag			3069
Thr Ser Gly Gln Leu Lys Lys His Thr Asp Val Ala Ser Gly Gln			
1010	1015	1020	
gcc acc atc gtg cag ccc tgc acg ctc ggc gac ctc ggt gac cgc			3114
Ala Thr Ile Val Gln Pro Cys Thr Leu Gly Asp Leu Gly Asp Arg			

1025	1030	1035	
tcc ttc atg gag acc tac ggc gtc gtc gcc ccg ctg tac acg ggc			3159
Ser Phe Met Glu Thr Tyr Gly Val Val Ala Pro Leu Tyr Thr Gly			
1040	1045	1050	
gcc atg gcc aag ggc att gcc tcg gcg gac ctc gtc atc gcc gcc			3204
Ala Met Ala Lys Gly Ile Ala Ser Ala Asp Leu Val Ile Ala Ala			
1055	1060	1065	
ggc aag cgc aag atc ctc ggc tcc ttt ggc gcc ggc ggc ctc ccc			3249
Gly Lys Arg Lys Ile Leu Gly Ser Phe Gly Ala Gly Gly Leu Pro			
1070	1075	1080	
atg cac cac gtg cgc gcc gcc ctc gag aag atc cag gcc gcc ctg			3294
Met His His Val Arg Ala Ala Leu Glu Lys Ile Gln Ala Ala Leu			
1085	1090	1095	
cct cag ggc ccc tac gcc gtc aac ctc atc cac tcg cct ttt gac			3339
Pro Gln Gly Pro Tyr Ala Val Asn Leu Ile His Ser Pro Phe Asp			
1100	1105	1110	
agc aac ctc gag aag ggc aac gtc gat ctc ttc ctc gag aag ggc			3384
Ser Asn Leu Glu Lys Gly Asn Val Asp Leu Phe Leu Glu Lys Gly			
1115	1120	1125	
gtc act gtg gtg gag gcc tcg gca ttc atg acc ctc acc ccg cag			3429
Val Thr Val Val Glu Ala Ser Ala Phe Met Thr Leu Thr Pro Gln			
1130	1135	1140	
gtc gtg cgc tac cgc gcc gcc ggc ctc tcg cgc aac gcc gac ggt			3474
Val Val Arg Tyr Arg Ala Ala Gly Leu Ser Arg Asn Ala Asp Gly			
1145	1150	1155	
tcg gtc aac atc cgc aac cgc atc atc ggc aag gtc tcg cgc acc			3519
Ser Val Asn Ile Arg Asn Arg Ile Ile Gly Lys Val Ser Arg Thr			
1160	1165	1170	
gag ctc gcc gag atg ttc atc cgc ccg gcc ccg gag cac ctc ctc			3564
Glu Leu Ala Glu Met Phe Ile Arg Pro Ala Pro Glu His Leu Leu			
1175	1180	1185	
gag aag ctc atc gcc tcg ggc gag atc acc cag gag cag gcc gag			3609
Glu Lys Leu Ile Ala Ser Gly Glu Ile Thr Gln Glu Gln Ala Glu			

1190	1195	1200	
ctc gcg cgc cgc gtt ccc gtc gcc gac gat atc gct gtc gag gct			3654
Leu Ala Arg Arg Val Pro Val Ala Asp Asp Ile Ala Val Glu Ala			
1205	1210	1215	
gac tcg ggc ggc cac acc gac aac cgc ccc atc cac gtc atc ctc			3699
Asp Ser Gly Gly His Thr Asp Asn Arg Pro Ile His Val Ile Leu			
1220	1225	1230	
ccg ctc atc atc aac ctc cgc aac cgc ctg cac cgc gag tgc ggc			3744
Pro Leu Ile Ile Asn Leu Arg Asn Arg Leu His Arg Glu Cys Gly			
1235	1240	1245	
tac ccc gcg cac ctc cgc gtc cgc gtt ggc gcc ggc ggt ggc gtc			3789
Tyr Pro Ala His Leu Arg Val Arg Val Gly Ala Gly Gly Gly Val			
1250	1255	1260	
ggc tgc ccg cag gcc gcc gcc gcc gcc ctc acc atg ggc gcc gcc			3834
Gly Cys Pro Gln Ala Ala Ala Ala Ala Leu Thr Met Gly Ala Ala			
1265	1270	1275	
ttc atc gtc acc ggc act gtc aac cag gtc gcc aag cag tcc ggc			3879
Phe Ile Val Thr Gly Thr Val Asn Gln Val Ala Lys Gln Ser Gly			
1280	1285	1290	
acc tgc gac aac gtg cgc aag cag ctc tcg cag gcc acc tac tcg			3924
Thr Cys Asp Asn Val Arg Lys Gln Leu Ser Gln Ala Thr Tyr Ser			
1295	1300	1305	
gat atc tgc atg gcc ccg gcc gcc gac atg ttc gag gag ggc gtc			3969
Asp Ile Cys Met Ala Pro Ala Ala Asp Met Phe Glu Glu Gly Val			
1310	1315	1320	
aag ctc cag gtc ctc aag aag gga acc atg ttc ccc tcg cgc gcc			4014
Lys Leu Gln Val Leu Lys Lys Gly Thr Met Phe Pro Ser Arg Ala			
1325	1330	1335	
aac aag ctc tac gag ctc ttt tgc aag tac gac tcc ttc gac tcc			4059
Asn Lys Leu Tyr Glu Leu Phe Cys Lys Tyr Asp Ser Phe Asp Ser			
1340	1345	1350	
atg cct cct gcc gag ctc gag cgc atc gag aag cgt atc ttc aag			4104
Met Pro Pro Ala Glu Leu Glu Arg Ile Glu Lys Arg Ile Phe Lys			

1355	1360	1365	
cgc gca ctc cag gag gtc tgg gag gag acc aag gac ttt tac att			4149
Arg Ala Leu Gln Glu Val Trp Glu Glu Thr Lys Asp Phe Tyr Ile			
1370	1375	1380	
aac ggt ctc aag aac ccg gag aag atc cag cgc gcc gag cac gac			4194
Asn Gly Leu Lys Asn Pro Glu Lys Ile Gln Arg Ala Glu His Asp			
1385	1390	1395	
ccc aag ctc aag atg tcg ctc tgc ttc cgc tgg tac ctt ggt ctt			4239
Pro Lys Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Gly Leu			
1400	1405	1410	
gcc agc cgc tgg gcc aac atg ggc gcc ccg gac cgc gtc atg gac			4284
Ala Ser Arg Trp Ala Asn Met Gly Ala Pro Asp Arg Val Met Asp			
1415	1420	1425	
tac cag gtc tgg tgt ggc ccg gcc att ggc gcc ttc aac gac ttc			4329
Tyr Gln Val Trp Cys Gly Pro Ala Ile Gly Ala Phe Asn Asp Phe			
1430	1435	1440	
atc aag ggc acc tac ctc gac ccc gct gtc tcc aac gag tac ccc			4374
Ile Lys Gly Thr Tyr Leu Asp Pro Ala Val Ser Asn Glu Tyr Pro			
1445	1450	1455	
tgt gtc gtc cag atc aac ctg caa atc ctc cgt ggt gcc tgc tac			4419
Cys Val Val Gln Ile Asn Leu Gln Ile Leu Arg Gly Ala Cys Tyr			
1460	1465	1470	
ctg cgc cgt ctc aac gcc ctg cgc aac gac ccg cgc att gac ctc			4464
Leu Arg Arg Leu Asn Ala Leu Arg Asn Asp Pro Arg Ile Asp Leu			
1475	1480	1485	
gag acc gag gat gct gcc ttt gtc tac gag ccc acc aac gcg ctc			4509
Glu Thr Glu Asp Ala Ala Phe Val Tyr Glu Pro Thr Asn Ala Leu			
1490	1495	1500	
taa			4512
<210> 28			
<211> 1503			
<212> PRT			

<213> Schizochytrium sp.

<400> 28

Met Ala Leu Arg Val Lys Thr Asn Lys Lys Pro Cys Trp Glu Met Thr
1 5 10 15

Lys Glu Glu Leu Thr Ser Gly Lys Thr Glu Val Phe Asn Tyr Glu Glu
20 25 30

Leu Leu Glu Phe Ala Glu Gly Asp Ile Ala Lys Val Phe Gly Pro Glu
35 40 45

Phe Ala Val Ile Asp Lys Tyr Pro Arg Arg Val Arg Leu Pro Ala Arg
50 55 60

Glu Tyr Leu Leu Val Thr Arg Val Thr Leu Met Asp Ala Glu Val Asn
65 70 75 80

Asn Tyr Arg Val Gly Ala Arg Met Val Thr Glu Tyr Asp Leu Pro Val
85 90 95

Asn Gly Glu Leu Ser Glu Gly Gly Asp Cys Pro Trp Ala Val Leu Val
100 105 110

Glu Ser Gly Gln Cys Asp Leu Met Leu Ile Ser Tyr Met Gly Ile Asp
115 120 125

Phe Gln Asn Gln Gly Asp Arg Val Tyr Arg Leu Leu Asn Thr Thr Leu
130 135 140

Thr Phe Tyr Gly Val Ala His Glu Gly Glu Thr Leu Glu Tyr Asp Ile
145 150 155 160

Arg Val Thr Gly Phe Ala Lys Arg Leu Asp Gly Gly Ile Ser Met Phe
165 170 175

Phe Phe Glu Tyr Asp Cys Tyr Val Asn Gly Arg Leu Leu Ile Glu Met
180 185 190

Arg Asp Gly Cys Ala Gly Phe Phe Thr Asn Glu Glu Leu Asp Ala Gly
195 200 205

Lys Gly Val Val Phe Thr Arg Gly Asp Leu Ala Ala Arg Ala Lys Ile
210 215 220

Pro Lys Gln Asp Val Ser Pro Tyr Ala Val Ala Pro Cys Leu His Lys
225 230 235 240

Thr Lys Leu Asn Glu Lys Glu Met Gln Thr Leu Val Asp Lys Asp Trp
245 250 255

Ala Ser Val Phe Gly Ser Lys Asn Gly Met Pro Glu Ile Asn Tyr Lys
260 265 270

Leu Cys Ala Arg Lys Met Leu Met Ile Asp Arg Val Thr Ser Ile Asp
275 280 285

His Lys Gly Gly Val Tyr Gly Leu Gly Gln Leu Val Gly Glu Lys Ile
290 295 300

Leu Glu Arg Asp His Trp Tyr Phe Pro Cys His Phe Val Lys Asp Gln
305 310 315 320

Val Met Ala Gly Ser Leu Val Ser Asp Gly Cys Ser Gln Met Leu Lys
325 330 335

Met Tyr Met Ile Trp Leu Gly Leu His Leu Thr Thr Gly Pro Phe Asp
340 345 350

Phe Arg Pro Val Asn Gly His Pro Asn Lys Val Arg Cys Arg Gly Gln
355 360 365

Ile Ser Pro His Lys Gly Lys Leu Val Tyr Val Met Glu Ile Lys Glu
370 375 380

Met Gly Phe Asp Glu Asp Asn Asp Pro Tyr Ala Ile Ala Asp Val Asn
385 390 395 400

Ile Ile Asp Val Asp Phe Glu Lys Gly Gln Asp Phe Ser Leu Asp Arg
405 410 415

Ile Ser Asp Tyr Gly Lys Gly Asp Leu Asn Lys Lys Ile Val Val Asp
420 425 430

Phe Lys Gly Ile Ala Leu Lys Met Gln Lys Arg Ser Thr Asn Lys Asn
435 440 445

Pro Ser Lys Val Gln Pro Val Phe Ala Asn Gly Ala Ala Thr Val Gly
450 455 460

Pro Glu Ala Ser Lys Ala Ser Ser Gly Ala Ser Ala Ser Ala
465 470 475 480

Ala Pro Ala Lys Pro Ala Phe Ser Ala Asp Val Leu Ala Pro Lys Pro
485 490 495

Val Ala Leu Pro Glu His Ile Leu Lys Gly Asp Ala Leu Ala Pro Lys
500 505 510

Glu Met Ser Trp His Pro Met Ala Arg Ile Pro Gly Asn Pro Thr Pro
515 520 525

Ser Phe Ala Pro Ser Ala Tyr Lys Pro Arg Asn Ile Ala Phe Thr Pro
530 535 540

Phe Pro Gly Asn Pro Asn Asp Asn Asp His Thr Pro Gly Lys Met Pro
545 550 555 560

Leu Thr Trp Phe Asn Met Ala Glu Phe Met Ala Gly Lys Val Ser Met
565 570 575

Cys Leu Gly Pro Glu Phe Ala Lys Phe Asp Asp Ser Asn Thr Ser Arg
580 585 590

Ser Pro Ala Trp Asp Leu Ala Leu Val Thr Arg Ala Val Ser Val Ser
595 600 605

Asp Leu Lys His Val Asn Tyr Arg Asn Ile Asp Leu Asp Pro Ser Lys
610 615 620

Gly Thr Met Val Gly Glu Phe Asp Cys Pro Ala Asp Ala Trp Phe Tyr
625 630 635 640

Lys Gly Ala Cys Asn Asp Ala His Met Pro Tyr Ser Ile Leu Met Glu
645 650 655

Ile Ala Leu Gln Thr Ser Gly Val Leu Thr Ser Val Leu Lys Ala Pro
660 665 670

Leu Thr Met Glu Lys Asp Asp Ile Leu Phe Arg Asn Leu Asp Ala Asn
675 680 685

Ala Glu Phe Val Arg Ala Asp Leu Asp Tyr Arg Gly Lys Thr Ile Arg
690 695 700

Asn Val Thr Lys Cys Thr Gly Tyr Ser Met Leu Gly Glu Met Gly Val
705 710 715 720

His Arg Phe Thr Phe Glu Leu Tyr Val Asp Asp Val Leu Phe Tyr Lys
725 730 735

Gly Ser Thr Ser Phe Gly Trp Phe Val Pro Glu Val Phe Ala Ala Gln
740 745 750

Ala Gly Leu Asp Asn Gly Arg Lys Ser Glu Pro Trp Phe Ile Glu Asn

755

760

765

Lys Val Pro Ala Ser Gln Val Ser Ser Phe Asp Val Arg Pro Asn Gly
770 775 780

Ser Gly Arg Thr Ala Ile Phe Ala Asn Ala Pro Ser Gly Ala Gln Leu
785 790 795 800

Asn Arg Arg Thr Asp Gln Gly Gln Tyr Leu Asp Ala Val Asp Ile Val
805 810 815

Ser Gly Ser Gly Lys Lys Ser Leu Gly Tyr Ala His Gly Ser Lys Thr
820 825 830

Val Asn Pro Asn Asp Trp Phe Phe Ser Cys His Phe Trp Phe Asp Ser
835 840 845

Val Met Pro Gly Ser Leu Gly Val Glu Ser Met Phe Gln Leu Val Glu
850 855 860

Ala Ile Ala Ala His Glu Asp Leu Ala Gly Lys Ala Arg His Cys Gln
865 870 875 880

Pro His Leu Cys Ala Arg Pro Arg Ala Arg Ser Ser Trp Lys Tyr Arg
885 890 895

Gly Gln Leu Thr Pro Lys Ser Lys Lys Met Asp Ser Glu Val His Ile
900 905 910

Val Ser Val Asp Ala His Asp Gly Val Val Asp Leu Val Ala Asp Gly
915 920 925

Phe Leu Trp Ala Asp Ser Leu Arg Val Tyr Ser Val Ser Asn Ile Arg
930 935 940

Val Arg Ile Ala Ser Gly Glu Ala Pro Ala Ala Ala Ser Ser Ala Ala

945	950	955	960
Ser Val Gly Ser Ser Ala Ser Ser Val Glu Arg Thr Arg Ser Ser Pro	965	970	975
Ala Val Ala Ser Gly Pro Ala Gln Thr Ile Asp Leu Lys Gln Leu Lys	980	985	990
Thr Glu Leu Leu Glu Leu Asp Ala Pro Leu Tyr Leu Ser Gln Asp Pro	995	1000	1005
Thr Ser Gly Gln Leu Lys Lys His Thr Asp Val Ala Ser Gly Gln	1010	1015	1020
Ala Thr Ile Val Gln Pro Cys Thr Leu Gly Asp Leu Gly Asp Arg	1025	1030	1035
Ser Phe Met Glu Thr Tyr Gly Val Val Ala Pro Leu Tyr Thr Gly	1040	1045	1050
Ala Met Ala Lys Gly Ile Ala Ser Ala Asp Leu Val Ile Ala Ala	1055	1060	1065
Gly Lys Arg Lys Ile Leu Gly Ser Phe Gly Ala Gly Gly Leu Pro	1070	1075	1080
Met His His Val Arg Ala Ala Leu Glu Lys Ile Gln Ala Ala Leu	1085	1090	1095
Pro Gln Gly Pro Tyr Ala Val Asn Leu Ile His Ser Pro Phe Asp	1100	1105	1110
Ser Asn Leu Glu Lys Gly Asn Val Asp Leu Phe Leu Glu Lys Gly	1115	1120	1125
Val Thr Val Val Glu Ala Ser Ala Phe Met Thr Leu Thr Pro Gln			

1130

1135

1140

Val Val Arg Tyr Arg Ala Ala Gly Leu Ser Arg Asn Ala Asp Gly
1145 1150 1155

Ser Val Asn Ile Arg Asn Arg Ile Ile Gly Lys Val Ser Arg Thr
1160 1165 1170

Glu Leu Ala Glu Met Phe Ile Arg Pro Ala Pro Glu His Leu Leu
1175 1180 1185

Glu Lys Leu Ile Ala Ser Gly Glu Ile Thr Gln Glu Gln Ala Glu
1190 1195 1200

Leu Ala Arg Arg Val Pro Val Ala Asp Asp Ile Ala Val Glu Ala
1205 1210 1215

Asp Ser Gly Gly His Thr Asp Asn Arg Pro Ile His Val Ile Leu
1220 1225 1230

Pro Leu Ile Ile Asn Leu Arg Asn Arg Leu His Arg Glu Cys Gly
1235 1240 1245

Tyr Pro Ala His Leu Arg Val Arg Val Gly Ala Gly Gly Gly Val
1250 1255 1260

Gly Cys Pro Gln Ala Ala Ala Ala Ala Leu Thr Met Gly Ala Ala
1265 1270 1275

Phe Ile Val Thr Gly Thr Val Asn Gln Val Ala Lys Gln Ser Gly
1280 1285 1290

Thr Cys Asp Asn Val Arg Lys Gln Leu Ser Gln Ala Thr Tyr Ser
1295 1300 1305

Asp Ile Cys Met Ala Pro Ala Ala Asp Met Phe Glu Glu Gly Val
1310 1315 1320

Lys Leu Gln Val Leu Lys Lys Gly Thr Met Phe Pro Ser Arg Ala
1325 1330 1335

Asn Lys Leu Tyr Glu Leu Phe Cys Lys Tyr Asp Ser Phe Asp Ser
1340 1345 1350

Met Pro Pro Ala Glu Leu Glu Arg Ile Glu Lys Arg Ile Phe Lys
1355 1360 1365

Arg Ala Leu Gln Glu Val Trp Glu Glu Thr Lys Asp Phe Tyr Ile
1370 1375 1380

Asn Gly Leu Lys Asn Pro Glu Lys Ile Gln Arg Ala Glu His Asp
1385 1390 1395

Pro Lys Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Gly Leu
1400 1405 1410

Ala Ser Arg Trp Ala Asn Met Gly Ala Pro Asp Arg Val Met Asp
1415 1420 1425

Tyr Gln Val Trp Cys Gly Pro Ala Ile Gly Ala Phe Asn Asp Phe
1430 1435 1440

Ile Lys Gly Thr Tyr Leu Asp Pro Ala Val Ser Asn Glu Tyr Pro
1445 1450 1455

Cys Val Val Gln Ile Asn Leu Gln Ile Leu Arg Gly Ala Cys Tyr
1460 1465 1470

Leu Arg Arg Leu Asn Ala Leu Arg Asn Asp Pro Arg Ile Asp Leu
1475 1480 1485

Glu Thr Glu Asp Ala Ala Phe Val Tyr Glu Pro Thr Asn Ala Leu
1490 1495 1500

<210> 29
<211> 1500
<212> DNA
<213> Schizochytrium sp.

<220>
<221> CDS
<222> (1)..(1500)

<400> 29
aag gtt cag ccc gtc ttt gcc aac ggc gcc gcc act gtc ggc ccc gag 48
Lys Val Gln Pro Val Phe Ala Asn Gly Ala Ala Thr Val Gly Pro Glu
1 5 10 15

gcc tcc aag gct tcc tcc ggc gcc agc gcc agc gcc agc gcc gcc ccg 96
Ala Ser Lys Ala Ser Ser Gly Ala Ser Ala Ser Ala Ser Ala Ala Pro
20 25 30

gcc aag cct gcc ttc agc gcc gat gtt ctt gcg ccc aag ccc gtt gcc 144
Ala Lys Pro Ala Phe Ser Ala Asp Val Leu Ala Pro Lys Pro Val Ala
35 40 45

ctt ccc gag cac atc ctc aag ggc gac gcc ctc gcc ccc aag gag atg 192
Leu Pro Glu His Ile Leu Lys Gly Asp Ala Leu Ala Pro Lys Glu Met
50 55 60

tcc tgg cac ccc atg gcc cgc atc ccg ggc aac ccg acg ccc tct ttt 240
Ser Trp His Pro Met Ala Arg Ile Pro Gly Asn Pro Thr Pro Ser Phe
65 70 75 80

gcg ccc tcg gcc tac aag ccg cgc aac atc gcc ttt acg ccc ttc ccc 288
Ala Pro Ser Ala Tyr Lys Pro Arg Asn Ile Ala Phe Thr Pro Phe Pro
85 90 95

ggc aac ccc aac gat aac gac cac acc ccg ggc aag atg ccg ctc acc 336
Gly Asn Pro Asn Asp Asn Asp His Thr Pro Gly Lys Met Pro Leu Thr
100 105 110

tgg ttc aac atg gcc gag ttc atg gcc ggc aag gtc agc atg tgc ctc 384

Trp Phe Asn Met Ala Glu Phe Met Ala Gly Lys Val Ser Met Cys Leu	
115 120 125	
ggc ccc gag ttc gcc aag ttc gac gac tcg aac acc agc cgc agc ccc	432
Gly Pro Glu Phe Ala Lys Phe Asp Asp Ser Asn Thr Ser Arg Ser Pro	
130 135 140	
gct tgg gac ctc gct ctc gtc acc cgc gcc gtg tct gtg tct gac ctc	480
Ala Trp Asp Leu Ala Leu Val Thr Arg Ala Val Ser Val Ser Asp Leu	
145 150 155 160	
aag cac gtc aac tac cgc aac atc gac ctc gac ccc tcc aag ggt acc	528
Lys His Val Asn Tyr Arg Asn Ile Asp Leu Asp Pro Ser Lys Gly Thr	
165 170 175	
atg gtc ggc gag ttc gac tgc ccc gcg gac gcc tgg ttc tac aag ggc	576
Met Val Gly Glu Phe Asp Cys Pro Ala Asp Ala Trp Phe Tyr Lys Gly	
180 185 190	
gcc tgc aac gat gcc cac atg ccg tac tcg atc ctc atg gag atc gcc	624
Ala Cys Asn Asp Ala His Met Pro Tyr Ser Ile Leu Met Glu Ile Ala	
195 200 205	
ctc cag acc tcg ggt gtg ctc acc tcg gtg ctc aag gcg ccc ctg acc	672
Leu Gln Thr Ser Gly Val Leu Thr Ser Val Leu Lys Ala Pro Leu Thr	
210 215 220	
atg gag aag gac gac atc ctc ttc cgc aac ctc gac gcc aac gcc gag	720
Met Glu Lys Asp Asp Ile Leu Phe Arg Asn Leu Asp Ala Asn Ala Glu	
225 230 235 240	
ttc gtg cgc gcc gac ctc gac tac cgc ggc aag act atc cgc aac gtc	768
Phe Val Arg Ala Asp Leu Asp Tyr Arg Gly Lys Thr Ile Arg Asn Val	
245 250 255	
acc aag tgc act ggc tac agc atg ctc ggc gag atg ggc gtc cac cgc	816
Thr Lys Cys Thr Gly Tyr Ser Met Leu Gly Glu Met Gly Val His Arg	
260 265 270	
ttc acc ttt gag ctc tac gtc gat gat gtg ctc ttt tac aag ggc tcg	864
Phe Thr Phe Glu Leu Tyr Val Asp Asp Val Leu Phe Tyr Lys Gly Ser	
275 280 285	
acc tcg ttc ggc tgg ttc gtg ccc gag gtc ttt gcc gcc cag gcc ggc	912
Thr Ser Phe Gly Trp Phe Val Pro Glu Val Phe Ala Ala Gln Ala Gly	

290	295	300	
ctc gac aac ggc cgc aag tcg gag ccc tgg ttc att gag aac aag gtt			960
Leu Asp Asn Gly Arg Lys Ser Glu Pro Trp Phe Ile Glu Asn Lys Val			
305	310	315	320
ccg gcc tcg cag gtc tcc tcc ttt gac gtg cgc ccc aac ggc agc ggc			1008
Pro Ala Ser Gln Val Ser Ser Phe Asp Val Arg Pro Asn Gly Ser Gly			
	325	330	335
cgc acc gcc atc ttc gcc aac gcc ccc agc ggc gcc cag ctc aac cgc			1056
Arg Thr Ala Ile Phe Ala Asn Ala Pro Ser Gly Ala Gln Leu Asn Arg			
	340	345	350
cgc acg gac cag ggc cag tac ctc gac gcc gtc gac att gtc tcc ggc			1104
Arg Thr Asp Gln Gly Gln Tyr Leu Asp Ala Val Asp Ile Val Ser Gly			
	355	360	365
agc ggc aag aag agc ctc ggc tac gcc cac ggt tcc aag acg gtc aac			1152
Ser Gly Lys Lys Ser Leu Gly Tyr Ala His Gly Ser Lys Thr Val Asn			
	370	375	380
ccg aac gac tgg ttc ttc tcg tgc cac ttt tgg ttt gac tcg gtc atg			1200
Pro Asn Asp Trp Phe Phe Ser Cys His Phe Trp Phe Asp Ser Val Met			
385	390	395	400
ccc gga agt ctc ggt gtc gag tcc atg ttc cag ctc gtc gag gcc atc			1248
Pro Gly Ser Leu Gly Val Glu Ser Met Phe Gln Leu Val Glu Ala Ile			
	405	410	415
gcc gcc cac gag gat ctc gct ggc aaa gca cgg cat tgc caa ccc cac			1296
Ala Ala His Glu Asp Leu Ala Gly Lys Ala Arg His Cys Gln Pro His			
	420	425	430
ctt tgt gca cgc ccc cgg gca aga tca agc tgg aag tac cgc ggc cag			1344
Leu Cys Ala Arg Pro Arg Ala Arg Ser Ser Trp Lys Tyr Arg Gly Gln			
	435	440	445
ctc acg ccc aag agc aag aag atg gac tcg gag gtc cac atc gtg tcc			1392
Leu Thr Pro Lys Ser Lys Lys Met Asp Ser Glu Val His Ile Val Ser			
	450	455	460
gtg gac gcc cac gac ggc gtt gtc gac ctc gtc gcc gac ggc ttc ctc			1440
Val Asp Ala His Asp Gly Val Val Asp Leu Val Ala Asp Gly Phe Leu			

465	470	475	480	
tgg gcc gac agc ctc cgc gtc tac tcg gtg agc aac att cgc gtg cgc				1488
Trp Ala Asp Ser Leu Arg Val Tyr Ser Val Ser Asn Ile Arg Val Arg				
	485	490	495	

atc gcc tcc ggt	1500
Ile Ala Ser Gly	
500	

<210>	30
<211>	500
<212>	PRT
<213>	Schizochytrium sp.

<400> 30

Lys Val Gln Pro Val Phe Ala Asn Gly Ala Ala Thr Val Gly Pro Glu
1 5 10 15

Ala Ser Lys Ala Ser Ser Gly Ala Ser Ala Ser Ala Ser Ala Ala Pro
20 25 30

Ala Lys Pro Ala Phe Ser Ala Asp Val Leu Ala Pro Lys Pro Val Ala
35 40 45

Leu Pro Glu His Ile Leu Lys Gly Asp Ala Leu Ala Pro Lys Glu Met
50 55 60

Ser Trp His Pro Met Ala Arg Ile Pro Gly Asn Pro Thr Pro Ser Phe
65 70 75 80

Ala Pro Ser Ala Tyr Lys Pro Arg Asn Ile Ala Phe Thr Pro Phe Pro
85 90 95

Gly Asn Pro Asn Asp Asn Asp His Thr Pro Gly Lys Met Pro Leu Thr
100 105 110

Trp Phe Asn Met Ala Glu Phe Met Ala Gly Lys Val Ser Met Cys Leu
115 120 125

Gly Pro Glu Phe Ala Lys Phe Asp Asp Ser Asn Thr Ser Arg Ser Pro
130 135 140

Ala Trp Asp Leu Ala Leu Val Thr Arg Ala Val Ser Val Ser Asp Leu
145 150 155 160

Lys His Val Asn Tyr Arg Asn Ile Asp Leu Asp Pro Ser Lys Gly Thr
165 170 175

Met Val Gly Glu Phe Asp Cys Pro Ala Asp Ala Trp Phe Tyr Lys Gly
180 185 190

Ala Cys Asn Asp Ala His Met Pro Tyr Ser Ile Leu Met Glu Ile Ala
195 200 205

Leu Gln Thr Ser Gly Val Leu Thr Ser Val Leu Lys Ala Pro Leu Thr
210 215 220

Met Glu Lys Asp Asp Ile Leu Phe Arg Asn Leu Asp Ala Asn Ala Glu
225 230 235 240

Phe Val Arg Ala Asp Leu Asp Tyr Arg Gly Lys Thr Ile Arg Asn Val
245 250 255

Thr Lys Cys Thr Gly Tyr Ser Met Leu Gly Glu Met Gly Val His Arg
260 265 270

Phe Thr Phe Glu Leu Tyr Val Asp Asp Val Leu Phe Tyr Lys Gly Ser
275 280 285

Thr Ser Phe Gly Trp Phe Val Pro Glu Val Phe Ala Ala Gln Ala Gly
290 295 300

Leu Asp Asn Gly Arg Lys Ser Glu Pro Trp Phe Ile Glu Asn Lys Val

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

<210> 31
 <211> 1512
 <212> DNA
 <213> Schizochytrium sp.

<220>
 <221> CDS
 <222> (1)..(1512)

<400> 31
 gcc ccg ctc tac ctc tcg cag gac ccg acc agc ggc cag ctc aag aag 48
 Ala Pro Leu Tyr Leu Ser Gln Asp Pro Thr Ser Gly Gln Leu Lys Lys
 1 5 10 15

cac acc gac gtg gcc tcc ggc cag gcc acc atc gtg cag ccc tgc acg 96
 His Thr Asp Val Ala Ser Gly Gln Ala Thr Ile Val Gln Pro Cys Thr
 20 25 30

ctc ggc gac ctc ggt gac cgc tcc ttc atg gag acc tac ggc gtc gtc 144
 Leu Gly Asp Leu Gly Asp Arg Ser Phe Met Glu Thr Tyr Gly Val Val
 35 40 45

gcc ccg ctg tac acg ggc gcc atg gcc aag ggc att gcc tcg gcg gac 192
 Ala Pro Leu Tyr Thr Gly Ala Met Ala Lys Gly Ile Ala Ser Ala Asp
 50 55 60

ctc gtc atc gcc gcc ggc aag cgc aag atc ctc ggc tcc ttt ggc gcc 240
 Leu Val Ile Ala Ala Gly Lys Arg Lys Ile Leu Gly Ser Phe Gly Ala
 65 70 75 80

ggc ggc ctc ccc atg cac cac gtg cgc gcc gcc ctc gag aag atc cag 288
 Gly Gly Leu Pro Met His His Val Arg Ala Ala Leu Glu Lys Ile Gln
 85 90 95

gcc gcc ctg cct cag ggc ccc tac gcc gtc aac ctc atc cac tcg cct 336
 Ala Ala Leu Pro Gln Gly Pro Tyr Ala Val Asn Leu Ile His Ser Pro
 100 105 110

ttt gac agc aac ctc gag aag ggc aac gtc gat ctc ttc ctc gag aag 384
 Phe Asp Ser Asn Leu Glu Lys Gly Asn Val Asp Leu Phe Leu Glu Lys
 115 120 125

ggc gtc act gtg gtg gag gcc tcg gca ttc atg acc ctc acc ccg cag 432

Gly Val Thr Val Val Glu Ala Ser Ala Phe Met Thr Leu Thr Pro Gln	
130 135 140	
gtc gtg cgc tac cgc gcc gcc ggc ctc tcg cgc aac gcc gac ggt tcg	480
Val Val Arg Tyr Arg Ala Ala Gly Leu Ser Arg Asn Ala Asp Gly Ser	
145 150 155 160	
gtc aac atc cgc aac cgc atc atc ggc aag gtc tcg cgc acc gag ctc	528
Val Asn Ile Arg Asn Arg Ile Ile Gly Lys Val Ser Arg Thr Glu Leu	
165 170 175	
gcc gag atg ttc atc cgc ccg gcc ccg gag cac ctc ctc gag aag ctc	576
Ala Glu Met Phe Ile Arg Pro Ala Pro Glu His Leu Leu Glu Lys Leu	
180 185 190	
atc gcc tcg ggc gag atc acc cag gag cag gcc gag ctc gcg cgc cgc	624
Ile Ala Ser Gly Glu Ile Thr Gln Glu Gln Ala Glu Leu Ala Arg Arg	
195 200 205	
gtt ccc gtc gcc gac gat atc gct gtc gag gct gac tcg ggc ggc cac	672
Val Pro Val Ala Asp Asp Ile Ala Val Glu Ala Asp Ser Gly Gly His	
210 215 220	
acc gac aac cgc ccc atc cac gtc atc ctc ccg ctc atc atc aac ctc	720
Thr Asp Asn Arg Pro Ile His Val Ile Leu Pro Leu Ile Ile Asn Leu	
225 230 235 240	
cgc aac cgc ctg cac cgc gag tgc ggc tac ccc gcg cac ctc cgc gtc	768
Arg Asn Arg Leu His Arg Glu Cys Gly Tyr Pro Ala His Leu Arg Val	
245 250 255	
cgc gtt ggc gcc ggc ggt ggc gtc ggc tgc ccg cag gcc gcc gcc gcc	816
Arg Val Gly Ala Gly Gly Gly Val Gly Cys Pro Gln Ala Ala Ala Ala	
260 265 270	
gcg ctc acc atg ggc gcc gcc ttc atc gtc acc ggc act gtc aac cag	864
Ala Leu Thr Met Gly Ala Ala Phe Ile Val Thr Gly Thr Val Asn Gln	
275 280 285	
gtc gcc aag cag tcc ggc acc tgc gac aac gtg cgc aag cag ctc tcg	912
Val Ala Lys Gln Ser Gly Thr Cys Asp Asn Val Arg Lys Gln Leu Ser	
290 295 300	
cag gcc acc tac tcg gat atc tgc atg gcc ccg gcc gcc gac atg ttc	960
Gln Ala Thr Tyr Ser Asp Ile Cys Met Ala Pro Ala Ala Asp Met Phe	

305	310	315	320	
gag gag ggc gtc aag ctc cag gtc ctc aag aag gga acc atg ttc ccc				1008
Glu Glu Gly Val Lys Leu Gln Val Leu Lys Lys Gly Thr Met Phe Pro				
	325	330	335	
tcg cgc gcc aac aag ctc tac gag ctc ttt tgc aag tac gac tcc ttc				1056
Ser Arg Ala Asn Lys Leu Tyr Glu Leu Phe Cys Lys Tyr Asp Ser Phe				
	340	345	350	
gac tcc atg cct cct gcc gag ctc gag cgc atc gag aag cgt atc ttc				1104
Asp Ser Met Pro Pro Ala Glu Leu Glu Arg Ile Glu Lys Arg Ile Phe				
	355	360	365	
aag cgc gca ctc cag gag gtc tgg gag gag acc aag gac ttt tac att				1152
Lys Arg Ala Leu Gln Glu Val Trp Glu Glu Thr Lys Asp Phe Tyr Ile				
	370	375	380	
aac ggt ctc aag aac ccg gag aag atc cag cgc gcc gag cac gac ccc				1200
Asn Gly Leu Lys Asn Pro Glu Lys Ile Gln Arg Ala Glu His Asp Pro				
	385	390	395	400
aag ctc aag atg tcg ctc tgc ttc cgc tgg tac ctt ggt ctt gcc agc				1248
Lys Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Gly Leu Ala Ser				
	405	410	415	
cgc tgg gcc aac atg ggc gcc ccg gac cgc gtc atg gac tac cag gtc				1296
Arg Trp Ala Asn Met Gly Ala Pro Asp Arg Val Met Asp Tyr Gln Val				
	420	425	430	
tgg tgt ggc ccg gcc att ggc gcc ttc aac gac ttc atc aag ggc acc				1344
Trp Cys Gly Pro Ala Ile Gly Ala Phe Asn Asp Phe Ile Lys Gly Thr				
	435	440	445	
tac ctc gac ccc gct gtc tcc aac gag tac ccc tgt gtc gtc cag atc				1392
Tyr Leu Asp Pro Ala Val Ser Asn Glu Tyr Pro Cys Val Val Gln Ile				
	450	455	460	
aac ctg caa atc ctc cgt ggt gcc tgc tac ctg cgc cgt ctc aac gcc				1440
Asn Leu Gln Ile Leu Arg Gly Ala Cys Tyr Leu Arg Arg Leu Asn Ala				
	465	470	475	480
ctg cgc aac gac ccg cgc att gac ctc gag acc gag gat gct gcc ttt				1488
Leu Arg Asn Asp Pro Arg Ile Asp Leu Glu Thr Glu Asp Ala Ala Phe				

485

490

495

gtc tac gag ccc acc aac gcg ctc
Val Tyr Glu Pro Thr Asn Ala Leu
500

1512

<210> 32
<211> 504
<212> PRT
<213> Schizochytrium sp.

<400> 32

Ala Pro Leu Tyr Leu Ser Gln Asp Pro Thr Ser Gly Gln Leu Lys Lys
1 5 10 15

His Thr Asp Val Ala Ser Gly Gln Ala Thr Ile Val Gln Pro Cys Thr
20 25 30

Leu Gly Asp Leu Gly Asp Arg Ser Phe Met Glu Thr Tyr Gly Val Val
35 40 45

Ala Pro Leu Tyr Thr Gly Ala Met Ala Lys Gly Ile Ala Ser Ala Asp
50 55 60

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

Gly Gly Leu Pro Met His His Val Arg Ala Ala Leu Glu Lys Ile Gln
85 90 95

Ala Ala Leu Pro Gln Gly Pro Tyr Ala Val Asn Leu Ile His Ser Pro
100 105 110

Phe Asp Ser Asn Leu Glu Lys Gly Asn Val Asp Leu Phe Leu Glu Lys
115 120 125

Gly Val Thr Val Val Glu Ala Ser Ala Phe Met Thr Leu Thr Pro Gln
130 135 140

Val Val Arg Tyr Arg Ala Ala Gly Leu Ser Arg Asn Ala Asp Gly Ser
145 150 155 160

Val Asn Ile Arg Asn Arg Ile Ile Gly Lys Val Ser Arg Thr Glu Leu
165 170 175

Ala Glu Met Phe Ile Arg Pro Ala Pro Glu His Leu Leu Glu Lys Leu
180 185 190

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

Val Pro Val Ala Asp Asp Ile Ala Val Glu Ala Asp Ser Gly Gly His
210 215 220

Thr Asp Asn Arg Pro Ile His Val Ile Leu Pro Leu Ile Ile Asn Leu
225 230 235 240

Arg Asn Arg Leu His Arg Glu Cys Gly Tyr Pro Ala His Leu Arg Val
245 250 255

Arg Val Gly Ala Gly Gly Gly Val Gly Cys Pro Gln Ala Ala Ala Ala
260 265 270

Ala Leu Thr Met Gly Ala Ala Phe Ile Val Thr Gly Thr Val Asn Gln
275 280 285

Val Ala Lys Gln Ser Gly Thr Cys Asp Asn Val Arg Lys Gln Leu Ser
290 295 300

Gln Ala Thr Tyr Ser Asp Ile Cys Met Ala Pro Ala Ala Asp Met Phe
305 310 315 320

Glu Glu Gly Val Lys Leu Gln Val Leu Lys Lys Gly Thr Met Phe Pro

325

330

335

Ser Arg Ala Asn Lys Leu Tyr Glu Leu Phe Cys Lys Tyr Asp Ser Phe
340 345 350

Asp Ser Met Pro Pro Ala Glu Leu Glu Arg Ile Glu Lys Arg Ile Phe
355 360 365

Lys Arg Ala Leu Gln Glu Val Trp Glu Glu Thr Lys Asp Phe Tyr Ile
370 375 380

Asn Gly Leu Lys Asn Pro Glu Lys Ile Gln Arg Ala Glu His Asp Pro
385 390 395 400

Lys Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Gly Leu Ala Ser
405 410 415

Arg Trp Ala Asn Met Gly Ala Pro Asp Arg Val Met Asp Tyr Gln Val
420 425 430

Trp Cys Gly Pro Ala Ile Gly Ala Phe Asn Asp Phe Ile Lys Gly Thr
435 440 445

Tyr Leu Asp Pro Ala Val Ser Asn Glu Tyr Pro Cys Val Val Gln Ile
450 455 460

Asn Leu Gln Ile Leu Arg Gly Ala Cys Tyr Leu Arg Arg Leu Asn Ala
465 470 475 480

Leu Arg Asn Asp Pro Arg Ile Asp Leu Glu Thr Glu Asp Ala Ala Phe
485 490 495

Val Tyr Glu Pro Thr Asn Ala Leu
500

<210> 33

<211> 9

<212> PRT
<213> Artificial sequence

<220>
<223> motif

<220>
<221> MISC_FEATURE
<222> (2)..(3)
<223> Xaa = any amino acid

<220>
<221> MISC_FEATURE
<222> (6)..(6)
<223> Xaa = Ala or Ser

<220>
<221> MISC_FEATURE
<222> (7)..(8)
<223> Xaa = any amino acid

<400> 33

Trp Xaa Xaa Lys Glu Xaa Xaa Xaa Lys
1 5

<210> 34
<211> 6
<212> PRT
<213> Artificial sequence

<220>
<223> motif

<220>
<221> MISC_FEATURE
<222> (1)..(6)
<223> Xaa = Ile or Leu or Val

<400> 34

Phe Asn Xaa Ser His Ser

1 5

<210> 35
<211> 5
<212> PRT
<213> Artificial sequence

<220>
<223> motif

<220>
<221> MISC_FEATURE
<222> (1)..(5)
<223> Xaa = Ile or Leu or Val

<400> 35

Xaa Gly Xaa Asp Xaa
1 5

<210> 36
<211> 4244
<212> DNA
<213> Schizochytrium sp.

<400> 36
tttctctctc tcgagctgtt gctgctgctg ctgctgctgc tgcctccttg ctggttctca 60

cgtccgttcg atcaagcgtc gcctcgctcg accgatcggg gcgtgcgtgc gtgcgtgagt 120

cttgttgcca ggcagccgca ggctgtctgt ctgtttgtgt agttttaccc tcgggggttcg 180

gggtctgcct gcctcccgtc cccgcccgcc gccgccgta tccaccccgc tcgcctccgc 240

ccatcggggc tcgcctcttc gcgcgcacg catcgcgccg atcgcatgca tcattgtgcc 300

acgcacgggg ggacgcgcgc cccgcgtccc ccgccgccgc cgtcgtcgtc tggcgatgcc 360

gtcgcgcgcc tcttctcttc cctgccttcc tcttctctcc gagccccct gtcttcttcc 420

gccccgcag cgccgcgcag gaagcgagga gagcggggag gagagaagaa aagaaaagaa 480

aagaaaagaa aataacacgc ccgtctcgcg cagacgcgcg cgcccgctg cgaggcggcg 540

tgatggggct tctcgtggcg cggctgcggc ctggcccggc ctgcctttg aggtgcaggc 600

tttgggagag aagagtggga cgcggagaag ataagatggt gccatggcgc aggacggaga 660

ggttgctgaa acttcttcga gcggcacagg cgatggcgag agaccgacag ctgccggcgc 720

ggaggggatg gataacctcc gaggtggca tggacgagct ggccgcgcgg atctggctgg 780

ccgcgcggcg gtgggtccgg aggcgcgagg ttggttttct tcatactga taccatacgg 840

tattcattct tcctctccag gaaggaagca agtcacatag agtatcacta gcctaattgat 900

ggactctatg ttttagggca cgtcggagca gaaggcgcga gcgattcgaa tgcgagcgat 960

agatacagca cagagacctt gccggcgacg cgatgcagg cgagcacgca cgcaccgcac 1020

gcacggcagc ggtgcacgcg ctctcggca gatgcacggt tctgcgccgc gcctttacat 1080

tttttgattt taggtgggtg gcctgccact ttgaacatca tccacaagtc aacgcagcat 1140

caagaggcaa gcaagtacat acatccattc gaattcaagt tcaagagacg cagcaacagc 1200

cgccgctccg ctcaagctgc agctagctgg ctgacagggc tcgctggctg tagtggaana 1260

ttccattcac ttttctgcat ccgcggccag caggcccgtg cgcacgttct ctgctttgtt 1320

tgttcgttcg tgcgtgcgtg cgtgcgtccc agctgcctgt ctaatctgcc gcgcgatcca 1380

acgacctcgc gtcgtgcgcg caagcgaaac ccgacgccga cctggccaat gccgcaagaa 1440

tgctaagcgc gcagcaatgc tgagagtaat cttcagccca ccaagtcatt atcgtgccc 1500

aagtctccat cgcagccaca ttacggcttt ctctctctct cctccctct ctttctgccg 1560

ggagagaagg aaagaccgcg cgcccgccc tctgcgctg tgacgggctg tccgttgtaa 1620

gccctcttag acagtcccta ggtgccgggc gccgccgcgc ctccgtcgca ggcacacgta 1680

ggcggccacg ggttcccccc gcacattcca cacctttctc ccccgagcc ggaccgcgcg 1740

ccgtctgctt acgcacttcg cgcggccgcc gcccggaac ccgagcgct gctgtgggcg 1800

ccgtcttcg gccgcgtcg aggtcgctcc cgcgccgcgc tactccgggt cctgtgcgtt 1860

acgtacttaa tattaacagt gggacctcgc acaggacctg acggcagcac agacgtcgcc 1920

gcctcgcatc gctggggacg caggcgaggc atccccgcgc gggccgcac cggggaggct 1980

gcggggcggc ctcttcggc cggcgccgc atcaggcgga tgacgaaga gccctcgag 2040

tcgtcgtc gcgggagcgc agcgggcgc cagcgtggcc aagctccgc cccttctggc 2100

tggtgcatg cctgcctgcc tgctgcctg cgtgcgtgcg tgctgcgtg ccttcgtgcg 2160

tgctgcctt cgtgcgtgcg tgctgagtg cggcggaaga gggatcatgc gaggatcaat 2220

caccgcgc accctgactt ttgaagaagc cgcgatgcga tgcatgcga tgcatgcga 2280

cgcgataccg tgcgaggcta cgaagcgagt ctggccggcc gtcataaac gcacgtttc 2340

gagaaggagg gctggcgag gcgtgcatgc cggcgacat tgcaacgcg gcgtctcgtg 2400

gctggcgaag gtgcctggag gatctaaca tcgtgctat gatgctatag ctgtgctgat 2460

ccccgtcca ttccaccag tctgtgcctg ccgctgacc tgcgcttgcc ttctctcaa 2520

gttctctcc gccgggcctt caggaccgag acgagacctg cagctgcagc tagactcgcg 2580

ctcgtcgcg gaggattcgc cggccgccgg gccggacggg actcgcgagg tcacacggcc 2640

gccggcgatc gcgatggctg tgctgacgta ctcgtgcgtg gcagccgtac gtcagcgacg 2700

ccgctccgt attgtggatt cgtagttgg ttgttggtg attgttgat taatttttt 2760

gttcgtaggc ttggttatag ctaatagttt agtttatact ggtgctcttc ggtgctgatt 2820

tagctcgact tgggtccaca ccactgcccc tctactgtga atggatcaat ggacgcacga 2880

cgggccgacg aaagtgcgcg agtgaggtaa cctaagcaac ggcggtcttc agaggggacg 2940

cacgcctcc gtcgcagtca gtccagacag gcagaaaagc gtcttaggga ccacgcacgc 3000

acgcacgcac gcacgcacgc ccgcacgcac gtcctctccc tcgctgcct attttttag 3060

gcttccttc gcacgggcct acctctgct cctcgcctc gccgcaccag gcggcagcag 3120

cgatactgc cggtgccgcc tccgtcacgc gctcagccgc agctcagccc agccgcgagc 3180

tagggtttgt tcgtctgaa ttgtttgatt tgatttgatt tgatttgatc cgatccgac 3240

cgatctgac tgatttgctt tgccttgctt tctctccctc ccggcgcgga ccaagctcc 3300

gtctgcgcgc cgcagctcc ctctctctcc cagccctctt tctgtcccg ctctcgcgc 3360

aagcacgcag ctccgccgcc gcatccggtc ggtcggtcgg tcgacgacc cgcctgccgc 3420

tgctgctgtg gccgggcttt tctccatcgg cgactctttc tctccatac gtctactac 3480

gtacatacat acigccggtc tctctctctt ccagcgcggc gacggcgga ggctgcgacg 3540

tcgtcggcg cgcgggcgcc gcgcgcggcg ccgcccgcg ccgcgtcgca ggccctcgtc 3600

gccgccggcg ctccgtccg ctccgaggcc gcgagaggcg cgcggcgcg cgatggatgg 3660

atggatggat ggatggatgg atggattttg ttgatcgatg gcggcgcatg gcgggagatg 3720

agcgaggacg agcgcgcgag cgcggcagcc ggattcgcag ggcctcgtc gcctcgcgc 3780

cgtgcgcgc cccgccttgc gagcctgcgc cgcgagcgag cgagcgagcg agcggggctt 3840

tctttgtctc gcgcgccgt tggcctcgtg tctcttgc ttgcgtagcg ggccgcgg 3900

tggaagatgg ctattcaat cgaccattc acgcacgcac tccggcgcg agagaaggcc 3960

gaggaggagc agcaagcaaa ccaaaagctc tcgcgctcgc ggtctcgggc tcgagcggtc 4020

tcggagagag agtcttgccg cgaccaccgg cagcagcagc agcagcagca gcgctgtcga 4080

gcacgagcac gagcacgagc acgagcacga gcattcgagc aagaggacag acacggttgt 4140

cagcgcttag ctgcctcgat acagaaagag gcgggttggg cgtaaaaaa aaggagcacg 4200

caagccgcca gccagccagc tagctagcca gcctgcctgc caaa 4244

<210> 37
 <211> 3886
 <212> DNA
 <213> Schizochytrium sp.

<220>
 <221> misc_feature
 <222> (1)..(3886)
 <223> n = a, c, g, or t

<400> 37
 gatcttgatt gccaaagctct ggattgtcga ttccgatgaa tcgagctctt tgttgtcgag 60

ctctggcttg ccgagctttc agaaatagac aaaattgccg agttcctgat tgcggggctc 120

tcgattgcc aaggtctggtg gatttcgaa ctctcgattg tcaaaatctt ggtcgtctcg 180

tcggattctt tcttgatttg ttttgtcaag accttgagat tgtgcaaaac ctgtatcggt 240

gacaaacct tgatcgacag cagcctttca tcacgtcag ctcttgtcat tgattatatt 300

ccccctgaca gccaacacct tgatgcaggg tctcaacctt gatTTTTGGA ggccatcatc 360

agcatcacgc cccggcactc acctcaaca ttcgacagcc aacgcttttt tttcttcgac 420

taggatctga gaataaaagc aggtcaccac gaccgtaggc caacgcgaca accatggaaa 480

taaagtgaca acgaacgact tgcaagttaa aatgtaaaga gcagcaattg cccgcccaca 540

gacaaatgaa agcaggcgcc gagtcttatt tgaggagggtg ggcctgtggc aatgggcgaa 600

agaaaaatcaa ggacaaggag agcaggttac gtaccggtat actggtatac gtacatggat 660

ggttcttggc aagttgacgg gatgtgtgcg agtgaccgtg gtagttaacg aaagagccgc 720

aagggaagg aaagcaagag aatgcagact tttccacagg atggatgggt ccgcagcttg 780

ccgcatgatg aaacgtgta tttcacctgg cacgtggtgg cgcacgcgcc cacatatgat 840

cgcgcgccg ggtgtattat acattttccc cctcaggtct actgccatcc ctccatgcgt 900

cgctcgtgcg aacgacgcaa gcctttcgca tcgtgcagcc tctttctggt aaggcaagag 960

ctaaacccaa acctaaacga aagaacattt ttacctctct ctctctccca ttggtcgcgt 1020

gcgtccgcc gctcgtcct cctcctgcc a gtgcgcgcc ctaacttccc cctcctccc 1080

ctccctccc cctcctctct cctcctgcc cgcctctctc tccgcgtgc gtgcggtgct 1140

gccctggacc aatggcatgc tgctgcagc tcggcggtat acgcaagccg cttcgcaatt 1200

tccggtacag atctcgccgg ggcgtgcgcc gcgggggtcac tgcggacctg ccgcggcccc 1260

tgcttctttc acatccatca tgctctccaa acctccgct cctccacga cgtacgcacg 1320

cccgtctgca cgcgcgact gccgtgcga aagcaagcg ccgcccgcg cccggcgacg 1380

ggaaggcggc cgcggtctcc ctccgcggtt gcctcgtcc cgcgcggggc tgggcgggca 1440

gcagaaggcg ggtggcgcg gcggttccg tcttcgtcag cggcctacgt cggcgcgcg 1500

gcgcgagact acgcatgcc ttgcgtcatg cgtcgcagg tagccgccg ggcctagcg 1560

tttccgtgg cgcgcgctt aagccccgg cgcgcaggt attgccgga taccgtacgg 1620

ccaagaccg cgcagacgt gccctctcg cgccagcca gccagcagc cagcggagga 1680

agagcgcgca ggcgcgggcg gagggcggcc gcggagcagc gcagagcggg gcggagcagc 1740

gcggagcaga acgggcagac tcggagcggg cagggcgggc agagctttgg ggtttaagga 1800

ccgggttacc ggcaagtga gcggctgcgg ggagcggctg tgggagggtt ggttacgcaa 1860

gcacgatgcg agcgagagag agacgtgcc gcgaatcaag aagtaggcg cgctgcgagg 1920

cgcgcgggcg gagcggagcg agggagaggg agaggagag agaggagggg agacgtgcc 1980

gcggcggggc ctggcctggc ctggtttggc ttggtcagcg cggccttgc cgagcgtgca 2040

gctggagtgg ggtggattca tttgatttt cttttgttt tgttttttc tctttcccgg 2100

aaagtgttg ccgncggtg ttctttgttt tgatttttc aaaagtttg gtggttggtt 2160

ctctctcttg gctctctgc aggcggtcg gtccacccc cggcctctcc tctctctcc 2220

tctctctcc tctccgtgc tatacgtacg tacgtttgta tacgtacata catcccggc 2280

gccgtgccgg cgagggtttg ctacgcctgg agcaatgcga tgcgatgcga tgcgatgcga 2340

cgcgacgcga cgcgagtcac tggttcgcg tgtggctgtg gcttgcttgc ttacttgctt 2400

tcgagctctc ccgttttctt ctttcttct cacgccacca ccaacgaaag aagatcggcc 2460

ccggcacgcc gctgagaagg gctggcggcg atgacggcac gcgcggccgc tgccacgttg 2520

gcgctcgctg ctgctgctg tgctgctgct gctgctgctg ctgctgctg tgctgcttct 2580

gcgcgcaggc ttgcccaga ggccggcgtg ctggccgctg ccgcttcag tccgctgga 2640

gagatcgaat gagagataaa ctggatggat tcacgaggg atgaatgaac gatggttga 2700

tgctttttc ctttttcagg tccacagcgg gaagcaggag cgcgtgaatc tgccgccatc 2760

gcatacgtc tgcatcgcgt gcacgcgtc gcacgcgtc ctcgccggga gccacagacg 2820

ggcgacaggg cgccagcca gccaggcagc cagccaggca ggcaccagag ggccagagag 2880

cgcgctcac gcacgcgcg cagtgcgcgc atcgctcgca gtgcagacct tgattccccg 2940

cgcggtatctc cgcgagcccc aaacgaagag cgccgtacgg gcccatccta gcgtcgcttc 3000

gcaccgcatac gcatcgcatc gcgttcccta gagagtagta ctgcacgaag gcaccatttc 3060

cgcgctcctc ttcggcgcga tcgaggcccc cgcgcccgcg acgatcgcg cgcccgcgcg 3120

gctggcgcg gccttgcg ctcgctggc ggcccgcg ggctctggc cctggcgcg 3180

gcggcgccg caggaggagc ggcagcggt gctcgccgc agagaagagc gcgccgggc 3240

cggggaggga cgggaggag aaggagaagg cgcgcaaggc ggccccgaaa gagaagacc 3300

tggacttgaa cggaagaag aagaagaagg agaagaagt gaagaagaag aagaagaagg 3360

agaggaagt gaagaagac aggagcaggc gcgttccaag gcgcgttctc ttccggaggc 3420

gcgttccagc tgcggcgcg gggcggtg cgggcgggc gcggcgcg gtgcgggcag 3480

aggggacgc cgcgcgagg cgagggggc cgagcgggag cccctgctgc tgcggggcg 3540

ccgggcccga ggtgtggcg gcgcgacgac ggagcgacg acgccagcg ccgcgacgac 3600

aaggccggcg gcgtcgcg gcggaaggc ccgcgggag caggggcgg agcaggacaa 3660

ggcgaggag caggagcagg gccgggagc ggagcgggag cgggcgcg agcccgaggc 3720

agaaccaat cgagatccag agcgagcaga ggccggccgc gagcccgag ccgcgccga 3780

gatcactagt accgctgcg aatcacagca gcagcagcag cagcagcagc agcagcagca 3840

gcagcagcag ccacgagagg gagataaaga aaaagcggca gagacg 3886

<210> 38
 <211> 8436
 <212> DNA
 <213> *Thraustochytrium* sp.

<220>

<221> CDS

<222> (1)..(8433)

<400> 38

atg aag gac atg gaa gat aga cgg gtc gct att gtg ggc atg tca gct	48
Met Lys Asp Met Glu Asp Arg Arg Val Ala Ile Val Gly Met Ser Ala	
1 5 10 15	

cac ttg cct tgt ggg aca gat gtg aag gaa tca tgg cag gct att cgc	96
His Leu Pro Cys Gly Thr Asp Val Lys Glu Ser Trp Gln Ala Ile Arg	
20 25 30	

gat gga atc gac tgt cta agt gac cta ccc gcg gat cgt ctc gac gtt	144
Asp Gly Ile Asp Cys Leu Ser Asp Leu Pro Ala Asp Arg Leu Asp Val	
35 40 45	

aca gct tac tac aat ccc aac aaa gcc acg aaa gac aag atc tac tgc	192
Thr Ala Tyr Tyr Asn Pro Asn Lys Ala Thr Lys Asp Lys Ile Tyr Cys	
50 55 60	

aaa cgg ggt ggc ttc atc ccg aac tat gac ttc gac ccc cgc gaa ttt	240
Lys Arg Gly Gly Phe Ile Pro Asn Tyr Asp Phe Asp Pro Arg Glu Phe	
65 70 75 80	

ggg ctc aac atg ttt caa atg gaa gac tct gat gcg aat cag aca ctt	288
Gly Leu Asn Met Phe Gln Met Glu Asp Ser Asp Ala Asn Gln Thr Leu	
85 90 95	

acc ttg ctc aaa gtc aaa caa gct ctc gaa gat gca agc ata gag cct	336
Thr Leu Leu Lys Val Lys Gln Ala Leu Glu Asp Ala Ser Ile Glu Pro	
100 105 110	

ttc acc aag gag aag aag aac att gga tgt gtt tta ggt att ggt ggg	384
Phe Thr Lys Glu Lys Lys Asn Ile Gly Cys Val Leu Gly Ile Gly Gly	
115 120 125	

ggc caa aag gcg agt cat gag ttc tac tct cgt ctc aac tac gtt gtc	432
Gly Gln Lys Ala Ser His Glu Phe Tyr Ser Arg Leu Asn Tyr Val Val	
130 135 140	

gtt gaa aag gta ctt cgg aaa atg ggt tta cca gat gct gat gtt gaa	480
Val Glu Lys Val Leu Arg Lys Met Gly Leu Pro Asp Ala Asp Val Glu	

145	150	155	160	
gaa gct gtg gag aaa tac aag gca aat ttt ccc gag tgg cgc cta gac				
Glu Ala Val Glu Lys Tyr Lys Ala Asn Phe Pro Glu Trp Arg Leu Asp				
	165	170	175	528
tct ttc cct ggg ttt ctt ggg aat gta acg gct ggt cgg tgc agt aac				
Ser Phe Pro Gly Phe Leu Gly Asn Val Thr Ala Gly Arg Cys Ser Asn				
	180	185	190	576
acc ttc aac atg gaa ggt atg aac tgc gtt gtg gat gct gca tgt gcc				
Thr Phe Asn Met Glu Gly Met Asn Cys Val Val Asp Ala Ala Cys Ala				
	195	200	205	624
agt tct cta att gca atc aag gtt gca gtt gaa gag cta ctc ttt ggt				
Ser Ser Leu Ile Ala Ile Lys Val Ala Val Glu Glu Leu Leu Phe Gly				
	210	215	220	672
gac tgt gac acc atg att gca ggt gcc acc tgc acg gac aat tca ctt				
Asp Cys Asp Thr Met Ile Ala Gly Ala Thr Cys Thr Asp Asn Ser Leu				
	225	230	235	720
ggc atg tac atg gcc ttc tct aaa acg cca gtt ttt tct act gac cca				
Gly Met Tyr Met Ala Phe Ser Lys Thr Pro Val Phe Ser Thr Asp Pro				
	245	250	255	768
agt gtc cgc gcg tat gat gag aaa aca aaa ggg atg cta att gga gaa				
Ser Val Arg Ala Tyr Asp Glu Lys Thr Lys Gly Met Leu Ile Gly Glu				
	260	265	270	816
ggt tca gca atg ttc gtt ctt aaa cgc tat gcg gat gcc gta cgt gat				
Gly Ser Ala Met Phe Val Leu Lys Arg Tyr Ala Asp Ala Val Arg Asp				
	275	280	285	864
ggc gac aca att cac gcg gtt ctg cgt tct tgc tct tcg tct agt gat				
Gly Asp Thr Ile His Ala Val Leu Arg Ser Cys Ser Ser Ser Ser Asp				
	290	295	300	912
gga aaa gcg gca gga att tat act cct act ata tct gga caa gaa gaa				
Gly Lys Ala Ala Gly Ile Tyr Thr Pro Thr Ile Ser Gly Gln Glu Glu				
	305	310	315	960
gct ttg cgt cga gcg tat gcc cgt gcg ggg gta tgt cca tct acg atc				
Ala Leu Arg Arg Ala Tyr Ala Arg Ala Gly Val Cys Pro Ser Thr Ile				
				1008

325	330	335	
ggg ctt gtt gag ggt cac ggg aca ggg acc cct gtt gga gat cgc att			1056
Gly Leu Val Glu Gly His Gly Thr Gly Thr Pro Val Gly Asp Arg Ile			
340	345	350	
gag tta aca gct ctg cgg aac ttg ttt gac aaa gct ttt ggt agc aag			1104
Glu Leu Thr Ala Leu Arg Asn Leu Phe Asp Lys Ala Phe Gly Ser Lys			
355	360	365	
aag gaa caa ata gca gtt ggc agc ata aag tct cag ata ggt cac ctg			1152
Lys Glu Gln Ile Ala Val Gly Ser Ile Lys Ser Gln Ile Gly His Leu			
370	375	380	
aaa tct gtt gcc ggc ttt gcc ggc ttg gtc aaa gct gtg ctt gcg ctt			1200
Lys Ser Val Ala Gly Phe Ala Gly Leu Val Lys Ala Val Leu Ala Leu			
385	390	395	400
aaa cac aaa acg ctc cca ggt tcg att aat gtc gac cag cca cct ttg			1248
Lys His Lys Thr Leu Pro Gly Ser Ile Asn Val Asp Gln Pro Pro Leu			
405	410	415	
ttg tat gac ggt act caa att caa gac tct tct tta tat atc aac aag			1296
Leu Tyr Asp Gly Thr Gln Ile Gln Asp Ser Ser Leu Tyr Ile Asn Lys			
420	425	430	
aca aat aga cca tgg ttt acg caa aac aag ctt ccg cgt cgg gct ggt			1344
Thr Asn Arg Pro Trp Phe Thr Gln Asn Lys Leu Pro Arg Arg Ala Gly			
435	440	445	
gtc tca agt ttt gga ttt gga ggt gca aac tac cac gcg gtt ctg gaa			1392
Val Ser Ser Phe Gly Phe Gly Gly Ala Asn Tyr His Ala Val Leu Glu			
450	455	460	
gaa ttc gag ccc gag cat gaa aaa cca tac cgc ctc aat act gtt gga			1440
Glu Phe Glu Pro Glu His Glu Lys Pro Tyr Arg Leu Asn Thr Val Gly			
465	470	475	480
cat cct gtc ctc ttg tac gct ccg tct gtg gaa gcc ctc aaa gta ctt			1488
His Pro Val Leu Leu Tyr Ala Pro Ser Val Glu Ala Leu Lys Val Leu			
485	490	495	
tgc aac gac cag ctt gcg gag ctc aca att gca ttg gaa gag gca aaa			1536
Cys Asn Asp Gln Leu Ala Glu Leu Thr Ile Ala Leu Glu Glu Ala Lys			

500	505	510	
aca cat aaa aat gtt gac aaa gtt tgt ggc tac aag ttt att gac gaa			1584
Thr His Lys Asn Val Asp Lys Val Cys Gly Tyr Lys Phe Ile Asp Glu			
515	520	525	
ttt cag ctc caa gga agc tgt cct cca gaa aat ccg aga gta gga ttt			1632
Phe Gln Leu Gln Gly Ser Cys Pro Pro Glu Asn Pro Arg Val Gly Phe			
530	535	540	
tta gca aca ctg cct act tca aat atc att gtc gcg ctt aag gca att			1680
Leu Ala Thr Leu Pro Thr Ser Asn Ile Ile Val Ala Leu Lys Ala Ile			
545	550	555	560
ctc gcg cag ctt gat gca aaa cca gat gcg aag aaa tgg gat ttg cct			1728
Leu Ala Gln Leu Asp Ala Lys Pro Asp Ala Lys Lys Trp Asp Leu Pro			
565	570	575	
cat aaa aag gct ttt ggg gct acc ttc gca tcg tct tca gtg aaa ggc			1776
His Lys Lys Ala Phe Gly Ala Thr Phe Ala Ser Ser Ser Val Lys Gly			
580	585	590	
tct gtt gct gcg ctc ttc gca gga cag ggt acc cag tac tta aac atg			1824
Ser Val Ala Ala Leu Phe Ala Gly Gln Gly Thr Gln Tyr Leu Asn Met			
595	600	605	
ttc tct gat gtg gca atg aac tgg cca ccg ttc cgt gac agc att gtc			1872
Phe Ser Asp Val Ala Met Asn Trp Pro Pro Phe Arg Asp Ser Ile Val			
610	615	620	
gca atg gaa gaa gct caa act gag gta ttt gag ggc caa gtt gaa cca			1920
Ala Met Glu Glu Ala Gln Thr Glu Val Phe Glu Gly Gln Val Glu Pro			
625	630	635	640
att agc aaa gtt ctg ttt cca cga gag cgc tat gca tcc gaa agt gaa			1968
Ile Ser Lys Val Leu Phe Pro Arg Glu Arg Tyr Ala Ser Glu Ser Glu			
645	650	655	
cag ggg aat gaa ctt ctt tgc tta aca gag tac tct cag cca act acg			2016
Gln Gly Asn Glu Leu Leu Cys Leu Thr Glu Tyr Ser Gln Pro Thr Thr			
660	665	670	
ata gca gcc gca gta ggg gcc ttc gat att ttc aaa gcg gct ggc ttt			2064
Ile Ala Ala Ala Val Gly Ala Phe Asp Ile Phe Lys Ala Ala Gly Phe			

675	680	685	
aag cca gac atg gtt gga ggg cat tca ctt ggc gaa ttt gct gct ttg			2112
Lys Pro Asp Met Val Gly Gly His Ser Leu Gly Glu Phe Ala Ala Leu			
690	695	700	
tac gcg gct ggg tcc att tcg cgt gac gac ctg tac aag ctt gtg tgc			2160
Tyr Ala Ala Gly Ser Ile Ser Arg Asp Asp Leu Tyr Lys Leu Val Cys			
705	710	715	720
aaa cgg gca aag gca atg gcg aac gct agt gac gga gct atg gca gca			2208
Lys Arg Ala Lys Ala Met Ala Asn Ala Ser Asp Gly Ala Met Ala Ala			
725	730	735	
gtg att ggc cca gat gca cgt cta gtt acg cca caa aat agt gac gtt			2256
Val Ile Gly Pro Asp Ala Arg Leu Val Thr Pro Gln Asn Ser Asp Val			
740	745	750	
tat gtc gca aac ttc aac tcc gca act caa gta gtc atc agt ggc act			2304
Tyr Val Ala Asn Phe Asn Ser Ala Thr Gln Val Val Ile Ser Gly Thr			
755	760	765	
gtt caa ggt gtg aaa gaa gag tcg aaa ttg ctc att tca aag ggg ttc			2352
Val Gln Gly Val Lys Glu Glu Ser Lys Leu Leu Ile Ser Lys Gly Phe			
770	775	780	
cgc gta ctg cca ctt aaa tgc cag ggc gcc ttc cat tct cct ttg atg			2400
Arg Val Leu Pro Leu Lys Cys Gln Gly Ala Phe His Ser Pro Leu Met			
785	790	795	800
ggg cct tct gag gat agt ttc aaa tca ctt gtg gag act tgt acc atc			2448
Gly Pro Ser Glu Asp Ser Phe Lys Ser Leu Val Glu Thr Cys Thr Ile			
805	810	815	
tcg ccg cca aaa aat gtg aaa ttc ttt tgc aat gtt agt ggc aag gaa			2496
Ser Pro Pro Lys Asn Val Lys Phe Phe Cys Asn Val Ser Gly Lys Glu			
820	825	830	
agc cca aac cca aaa cag acc ctc aag tca cac atg acg tct agc gtt			2544
Ser Pro Asn Pro Lys Gln Thr Leu Lys Ser His Met Thr Ser Ser Val			
835	840	845	
cag ttc gag gag cag att cgt aac atg tac gat gcc gga gca cgt gtt			2592
Gln Phe Glu Glu Gln Ile Arg Asn Met Tyr Asp Ala Gly Ala Arg Val			

850	855	860	
ttt ctg gag ttt gga ccc cgc caa gtc ctt gca aag ctt atc gcg gaa			2640
Phe Leu Glu Phe Gly Pro Arg Gln Val Leu Ala Lys Leu Ile Ala Glu			
865	870	875	880
atg ttt ccc tcg tgt aca gct atc agc gtt aac ccc gcg agc agt ggt			2688
Met Phe Pro Ser Cys Thr Ala Ile Ser Val Asn Pro Ala Ser Ser Gly			
	885	890	895
gac agt gac gtg caa ctc cgc ctc gcc gcc gta aaa ttc gcg gtc tcg			2736
Asp Ser Asp Val Gln Leu Arg Leu Ala Ala Val Lys Phe Ala Val Ser			
	900	905	910
ggt gca gcc ctt agc acc ttt gat cca tgg gag tat cgc aag cca caa			2784
Gly Ala Ala Leu Ser Thr Phe Asp Pro Trp Glu Tyr Arg Lys Pro Gln			
	915	920	925
gat ctt ctt att cga aaa cca cga aaa act gcc ctt gtt cta tca gca			2832
Asp Leu Leu Ile Arg Lys Pro Arg Lys Thr Ala Leu Val Leu Ser Ala			
	930	935	940
gca aca tat gtt tcc cca aag act ctt gca gaa cgt aaa aag gct atg			2880
Ala Thr Tyr Val Ser Pro Lys Thr Leu Ala Glu Arg Lys Lys Ala Met			
945	950	955	960
gaa gat atc aag cta gta tcc att aca cca aga gat agt atg gta tca			2928
Glu Asp Ile Lys Leu Val Ser Ile Thr Pro Arg Asp Ser Met Val Ser			
	965	970	975
att gga aaa atc gcg caa gaa gta cgg aca gct aaa cag cct tta gaa			2976
Ile Gly Lys Ile Ala Gln Glu Val Arg Thr Ala Lys Gln Pro Leu Glu			
	980	985	990
acc gaa att cga aga ctc aac aaa gaa tta gaa cat ctc aag aga gag			3024
Thr Glu Ile Arg Arg Leu Asn Lys Glu Leu Glu His Leu Lys Arg Glu			
	995	1000	1005
cta gca gca gcc aaa gcg agt gtc aag tct gca tca aaa agc tct			3069
Leu Ala Ala Ala Lys Ala Ser Val Lys Ser Ala Ser Lys Ser Ser			
1010	1015	1020	
aaa gag cga tct gtc cta tca aag cac cgc gct ttg ctt caa aac			3114
Lys Glu Arg Ser Val Leu Ser Lys His Arg Ala Leu Leu Gln Asn			

1025	1030	1035	
att ttg caa gac tac gat gat	ctt cgt gtg gtg cca	ttc gct gtt	3159
Ile Leu Gln Asp Tyr Asp Asp	Leu Arg Val Val Pro	Phe Ala Val	
1040	1045	1050	
cgt tct gtt gca gtg gac aac	acc gcg ccg tat gct	gac caa gtt	3204
Arg Ser Val Ala Val Asp Asn	Thr Ala Pro Tyr Ala	Asp Gln Val	
1055	1060	1065	
tcg acc cca gcg tca gag ccg	tcg gct tca ccg ctt	ttc gag aaa	3249
Ser Thr Pro Ala Ser Glu Arg	Ser Ala Ser Pro Leu	Phe Glu Lys	
1070	1075	1080	
cgc agt tcg gtt tcg tca gca	cgc ctc gct gaa gct	gaa gcc gcg	3294
Arg Ser Ser Val Ser Ser Ala	Arg Leu Ala Glu Ala	Glu Ala Ala	
1085	1090	1095	
gta ctg agc gtt ctc gca gac	aag aca ggc tac gac	agc tca atg	3339
Val Leu Ser Val Leu Ala Asp	Lys Thr Gly Tyr Asp	Ser Ser Met	
1100	1105	1110	
atc gag atg gac atg gac ctg	gag agt gag ctt ggc	gtt gat agc	3384
Ile Glu Met Asp Met Asp Leu	Glu Ser Glu Leu Gly	Val Asp Ser	
1115	1120	1125	
atc aaa cgc gtg gag atc atg	agc gag gtt caa acg	ctg ctc agc	3429
Ile Lys Arg Val Glu Ile Met	Ser Glu Val Gln Thr	Leu Leu Ser	
1130	1135	1140	
gtg gaa gtc tcc gac gtt gac	gct ctg tca aga acc	aag act gtt	3474
Val Glu Val Ser Asp Val Asp	Ala Leu Ser Arg Thr	Lys Thr Val	
1145	1150	1155	
ggc gac gtc atc gag gcg atg	aag ctg gaa ctc ggt	gga ccc caa	3519
Gly Asp Val Ile Glu Ala Met	Lys Leu Glu Leu Gly	Gly Pro Gln	
1160	1165	1170	
ggc cag act ttg acc gcg gaa	tcg atc cgt cag cca	ccg gtg tcc	3564
Gly Gln Thr Leu Thr Ala Glu	Ser Ile Arg Gln Pro	Pro Val Ser	
1175	1180	1185	
gag cct gct gta ccg acc tca	tcg tca agc agt att	gct aat gtt	3609
Glu Pro Ala Val Pro Thr Ser	Ser Ser Ser Ser Ile	Ala Asn Val	

1190	1195	1200	
tcg tca gca cgc ctc gct gaa gct gaa gct gcg gta ctg agc gtt			3654
Ser Ser Ala Arg Leu Ala Glu Ala Glu Ala Ala Val Leu Ser Val			
1205	1210	1215	
ctc gca gac aag aca ggc tac gac agc tca atg atc gag atg gac			3699
Leu Ala Asp Lys Thr Gly Tyr Asp Ser Ser Met Ile Glu Met Asp			
1220	1225	1230	
atg gac ctg gag agc gag ctt ggc gtt gat agc atc aaa cgc gtg			3744
Met Asp Leu Glu Ser Glu Leu Gly Val Asp Ser Ile Lys Arg Val			
1235	1240	1245	
gag atc atg agc gag gtt caa acg ctg ctc agc gtg gaa gtc tcc			3789
Glu Ile Met Ser Glu Val Gln Thr Leu Leu Ser Val Glu Val Ser			
1250	1255	1260	
gac gtt gac gct ctg tca aga act aag act gtt ggc gac gtc atc			3834
Asp Val Asp Ala Leu Ser Arg Thr Lys Thr Val Gly Asp Val Ile			
1265	1270	1275	
gag gcg atg aag ctg gaa ctc ggt gga ccc caa ggc cag act ttg			3879
Glu Ala Met Lys Leu Glu Leu Gly Gly Pro Gln Gly Gln Thr Leu			
1280	1285	1290	
acc gcg gaa tcg atc cgt cag cca ccg gtg tct gag cct gct gta			3924
Thr Ala Glu Ser Ile Arg Gln Pro Pro Val Ser Glu Pro Ala Val			
1295	1300	1305	
ccg acc tca tcg tca agc agt att gct aat gtt tcg tca gca cgc			3969
Pro Thr Ser Ser Ser Ser Ser Ile Ala Asn Val Ser Ser Ala Arg			
1310	1315	1320	
ctc gct gaa gct gaa gcg gcg gta ctg agc gtt ctc gca gac aag			4014
Leu Ala Glu Ala Glu Ala Ala Val Leu Ser Val Leu Ala Asp Lys			
1325	1330	1335	
aca ggc tac gac agc tca atg atc gag atg gac atg gac ctg gag			4059
Thr Gly Tyr Asp Ser Ser Met Ile Glu Met Asp Met Asp Leu Glu			
1340	1345	1350	
agc gag ctt ggc gtc gac agc atc aaa cgc gtg gag atc atg agc			4104
Ser Glu Leu Gly Val Asp Ser Ile Lys Arg Val Glu Ile Met Ser			

1355	1360	1365	
gag gtt caa acg ctg ctc agc	gtg gaa gtc tcc gac	gtt gac gct	4149
Glu Val Gln Thr Leu Leu Ser	Val Glu Val Ser Asp	Val Asp Ala	
1370	1375	1380	
ctg tca aga acc aag act gtt	ggc gac gtc atc gag	gcg atg aag	4194
Leu Ser Arg Thr Lys Thr Val	Gly Asp Val Ile Glu	Ala Met Lys	
1385	1390	1395	
ctg gaa ctc ggt gga ccc caa	ggc cag act ttg acc	gcg gaa tcg	4239
Leu Glu Leu Gly Gly Pro Gln	Gly Gln Thr Leu Thr	Ala Glu Ser	
1400	1405	1410	
atc cgt cag cca ccg gtg tcc	gag cct gct gta ccg	acc tca tcg	4284
Ile Arg Gln Pro Pro Val Ser	Glu Pro Ala Val Pro	Thr Ser Ser	
1415	1420	1425	
tca agc agt att gct aat gtt	ttg tca gca cgc ctc	gct gaa gct	4329
Ser Ser Ser Ile Ala Asn Val	Leu Ser Ala Arg Leu	Ala Glu Ala	
1430	1435	1440	
gaa gcc gcg gta ctg agc gtt	ctc gca gac aag aca	ggc tac gac	4374
Glu Ala Ala Val Leu Ser Val	Leu Ala Asp Lys Thr	Gly Tyr Asp	
1445	1450	1455	
agc tca atg atc gag atg gac	atg gac ctg gag agc	gag ctt ggc	4419
Ser Ser Met Ile Glu Met Asp	Met Asp Leu Glu Ser	Glu Leu Gly	
1460	1465	1470	
gtt gat agc atc aaa cgc gtg	gag atc atg agc gag	gtt caa acg	4464
Val Asp Ser Ile Lys Arg Val	Glu Ile Met Ser Glu	Val Gln Thr	
1475	1480	1485	
ttg ctc agc gtg gaa gtc tcc	gac gtt gac gct ctg	tca aga acc	4509
Leu Leu Ser Val Glu Val Ser	Asp Val Asp Ala Leu	Ser Arg Thr	
1490	1495	1500	
aag act gtt ggc gac gtc atc	gag gcg atg aag ctg	gaa ctc ggt	4554
Lys Thr Val Gly Asp Val Ile	Glu Ala Met Lys Leu	Glu Leu Gly	
1505	1510	1515	
gga ccc caa ggc cag act ttg	acc gcg gaa tcg atc	cgt cag cca	4599
Gly Pro Gln Gly Gln Thr Leu	Thr Ala Glu Ser Ile	Arg Gln Pro	

1520	1525	1530	
ccg gtg tct gag cct gct gta	ccg acc tca tcg tca	agc agt att	4644
Pro Val Ser Glu Pro Ala Val	Pro Thr Ser Ser Ser	Ser Ser Ile	
1535	1540	1545	
gct aat gtt tcg tca gca cgc	ctc gct gaa gct gaa	gcc gcg gta	4689
Ala Asn Val Ser Ser Ala Arg	Leu Ala Glu Ala Glu	Ala Ala Val	
1550	1555	1560	
ctg agc gtt ctc gca gac aag	aca ggc tac gac agc	tca atg atc	4734
Leu Ser Val Leu Ala Asp Lys	Thr Gly Tyr Asp Ser	Ser Met Ile	
1565	1570	1575	
gag atg gac atg gac ctg gag	agt gag ctt ggc gtc	gac agc atc	4779
Glu Met Asp Met Asp Leu Glu	Ser Glu Leu Gly Val	Asp Ser Ile	
1580	1585	1590	
aaa cgc gtg gag atc atg agc	gag gtt caa acg ctg	ctc agc gtg	4824
Lys Arg Val Glu Ile Met Ser	Glu Val Gln Thr Leu	Leu Ser Val	
1595	1600	1605	
gaa gtc tcc gac gtt gac gct	ctg tca aga acc aag	act gtt ggc	4869
Glu Val Ser Asp Val Asp Ala	Leu Ser Arg Thr Lys	Thr Val Gly	
1610	1615	1620	
gac gtc atc gag gcg atg aag	ctg gaa ctc ggt gga	ccc caa ggc	4914
Asp Val Ile Glu Ala Met Lys	Leu Glu Leu Gly Gly	Pro Gln Gly	
1625	1630	1635	
cag act ttg acc tct gaa ccg	atc cat cag cca cca	gtg tcc gag	4959
Gln Thr Leu Thr Ser Glu Pro	Ile His Gln Pro Pro	Val Ser Glu	
1640	1645	1650	
cct gct gta ccg acc tca tcg	tca agc agt att gct	aat gtt tct	5004
Pro Ala Val Pro Thr Ser Ser	Ser Ser Ser Ile Ala	Asn Val Ser	
1655	1660	1665	
tca gca cgc ctc gct gaa gct	gaa gcc gcg gta ctg	agc gtt ctc	5049
Ser Ala Arg Leu Ala Glu Ala	Glu Ala Ala Val Leu	Ser Val Leu	
1670	1675	1680	
gca gac aag aca ggc tac gac	agc tca atg atc gag	atg gac atg	5094
Ala Asp Lys Thr Gly Tyr Asp	Ser Ser Met Ile Glu	Met Asp Met	

1685	1690	1695	
gac ctg gag agc gag ctt ggc gtt gat agc atc aaa cgc gtg gaa			5139
Asp Leu Glu Ser Glu Leu Gly Val Asp Ser Ile Lys Arg Val Glu			
1700	1705	1710	
atc atg agc gag gtt caa acg ctg ctc agc gtg gaa gtc tcc gac			5184
Ile Met Ser Glu Val Gln Thr Leu Leu Ser Val Glu Val Ser Asp			
1715	1720	1725	
gtt gac gct ctg tca aga acc aag act gtt ggc gac gtc atc gag			5229
Val Asp Ala Leu Ser Arg Thr Lys Thr Val Gly Asp Val Ile Glu			
1730	1735	1740	
gcg atg aag atg gaa ctc ggt gga ccc caa ggc cag act ttg acc			5274
Ala Met Lys Met Glu Leu Gly Gly Pro Gln Gly Gln Thr Leu Thr			
1745	1750	1755	
gcg gaa tcg atc cgt cag cca ccg gtg tct gag cct gct gta ccg			5319
Ala Glu Ser Ile Arg Gln Pro Pro Val Ser Glu Pro Ala Val Pro			
1760	1765	1770	
acc tca tcg tca agc agt att gct aat gtt tcg tca gca cgc ctc			5364
Thr Ser Ser Ser Ser Ile Ala Asn Val Ser Ser Ala Arg Leu			
1775	1780	1785	
gct gaa gct gaa gcg gcg gta ctg agc gtt ctc gca gac aag aca			5409
Ala Glu Ala Glu Ala Ala Val Leu Ser Val Leu Ala Asp Lys Thr			
1790	1795	1800	
ggc tac gac agc tca atg atc gag atg gac atg gac ctg gag agc			5454
Gly Tyr Asp Ser Ser Met Ile Glu Met Asp Met Asp Leu Glu Ser			
1805	1810	1815	
gag ctt ggc gtt gat agc atc aaa cgc gtg gag atc atg agc gag			5499
Glu Leu Gly Val Asp Ser Ile Lys Arg Val Glu Ile Met Ser Glu			
1820	1825	1830	
gtt caa gcg ctg ctc agc gtg gaa gtc tcc gac gtt gac gct ctg			5544
Val Gln Ala Leu Leu Ser Val Glu Val Ser Asp Val Asp Ala Leu			
1835	1840	1845	
tca aga acc aag act gtt ggc gac gtc atc gag gcg atg aag atg			5589
Ser Arg Thr Lys Thr Val Gly Asp Val Ile Glu Ala Met Lys Met			

1850	1855	1860	
gaa ctc ggt gga ccc caa ggc	cag act ttg acc gca	gaa tcg atc	5634
Glu Leu Gly Gly Pro Gln Gly	Gln Thr Leu Thr Ala	Glu Ser Ile	
1865	1870	1875	
cgt gag cca ccg gtg tct gag	cct gct gta ccg acc	tca tcg tca	5679
Arg Glu Pro Pro Val Ser Glu	Pro Ala Val Pro Thr	Ser Ser Ser	
1880	1885	1890	
agt agt atc gct aat gtt tct	tca gct cgc ctc gct	gaa gct gaa	5724
Ser Ser Ile Ala Asn Val Ser	Ser Ala Arg Leu Ala	Glu Ala Glu	
1895	1900	1905	
gcc gcg gta ctg agc gtt ctc	gca gac aag aca ggc	tac gac agc	5769
Ala Ala Val Leu Ser Val Leu	Ala Asp Lys Thr Gly	Tyr Asp Ser	
1910	1915	1920	
tca atg atc gag atg gac atg	gac ctg gag agt gag	ctt ggc gtc	5814
Ser Met Ile Glu Met Asp Met	Asp Leu Glu Ser Glu	Leu Gly Val	
1925	1930	1935	
gac agc atc aaa cgc gtg gag	atc atg agc gag gtt	caa acg ttg	5859
Asp Ser Ile Lys Arg Val Glu	Ile Met Ser Glu Val	Gln Thr Leu	
1940	1945	1950	
ctc agc gtg gaa gtc tcc gac	gtt gac gct ctg tca	aga acc aag	5904
Leu Ser Val Glu Val Ser Asp	Val Asp Ala Leu Ser	Arg Thr Lys	
1955	1960	1965	
act gtt ggc gac gtc atc gag	gcg atg aag ctg gaa	ctt ggg gaa	5949
Thr Val Gly Asp Val Ile Glu	Ala Met Lys Leu Glu	Leu Gly Glu	
1970	1975	1980	
tca tca agt att gag act ctc	aat tgt acc gag gtt	gag cac acg	5994
Ser Ser Ser Ile Glu Thr Leu	Asn Cys Thr Glu Val	Glu His Thr	
1985	1990	1995	
agc tac aaa agt gtc aag gct	tca ggg tgt gag aat	gta gat acc	6039
Ser Tyr Lys Ser Val Lys Ala	Ser Gly Cys Glu Asn	Val Asp Thr	
2000	2005	2010	
cgt ttc gct aag gtt gta caa	atc tcg ctt cct agc	aag ctg aaa	6084
Arg Phe Ala Lys Val Val Gln	Ile Ser Leu Pro Ser	Lys Leu Lys	

2015	2020	2025	
tcc act gtg tcg cac gat cga Ser Thr Val Ser His Asp Arg 2030	cct gta att gtt gta gat gat gga Pro Val Ile Val Val Asp Asp Gly 2035		6129
acg ccc tta acc acg gag ctt Thr Pro Leu Thr Thr Glu Leu 2045	tgt aaa att ctt ggg ggt aat att Cys Lys Ile Leu Gly Gly Asn Ile 2050		6174
gtg gtt ctc tct tat caa ggg Val Val Leu Ser Tyr Gln Gly 2060	aag ccc gct ggt cca cgg gga gtc Lys Pro Ala Gly Pro Arg Gly Val 2065		6219
gag gtg cca gat ctt tcc gag Glu Val Pro Asp Leu Ser Glu 2075	gaa gcc cta att caa gct ctt gca Glu Ala Leu Ile Gln Ala Leu Ala 2080		6264
ttg att cgg tct aca tat gga Leu Ile Arg Ser Thr Tyr Gly 2090	gtt cca att ggt ttt att tgt cag Val Pro Ile Gly Phe Ile Cys Gln 2095		6309
caa gtg tct aat gtg agc acc Gln Val Ser Asn Val Ser Thr 2105	aag gca cag ctt tgt tgg gca ctc Lys Ala Gln Leu Cys Trp Ala Leu 2110		6354
ctc gca gcg aag cat ctc aag Leu Ala Ala Lys His Leu Lys 2120	aag gat ttg aat gct gtc tta ccc Lys Asp Leu Asn Ala Val Leu Pro 2125		6399
gat tca aga tcc ttc ttc gtc Asp Ser Arg Ser Phe Phe Val 2135	gga gtt gta cgc ttg aac ggg aaa Gly Val Val Arg Leu Asn Gly Lys 2140		6444
ctt gga act ttc gaa aac atc Leu Gly Thr Phe Glu Asn Ile 2150	agc gac ttc tct aaa ttt gat ttg Ser Asp Phe Ser Lys Phe Asp Leu 2155		6489
acg aaa gcc cta gat tac gga Thr Lys Ala Leu Asp Tyr Gly 2165	cag cgt ggt tct ctc tta ggc ctg Gln Arg Gly Ser Leu Leu Gly Leu 2170		6534
tgc aag tca cta gac tta gaa Cys Lys Ser Leu Asp Leu Glu 2180	tgg gaa cag gtg ttt tgc cgt gga Trp Glu Gln Val Phe Cys Arg Gly 2185		6579

2180	2185	2190	
ata gat ctt gcg tgt gat ctt atg cca ctc cag gcc gca agg ata			6624
Ile Asp Leu Ala Cys Asp Leu Met Pro Leu Gln Ala Ala Arg Ile			
2195	2200	2205	
ctc aga aat gag ctt cag tgt ccc aat atg cgc ctt cgc gag gtt			6669
Leu Arg Asn Glu Leu Gln Cys Pro Asn Met Arg Leu Arg Glu Val			
2210	2215	2220	
ggg tac gat att tct ggc gcc agg tac acc att tca acc gat gac			6714
Gly Tyr Asp Ile Ser Gly Ala Arg Tyr Thr Ile Ser Thr Asp Asp			
2225	2230	2235	
ctg cta tgt gga ccc tcg aag gct aaa gta gag gcc gca gac ttg			6759
Leu Leu Cys Gly Pro Ser Lys Ala Lys Val Glu Ala Ala Asp Leu			
2240	2245	2250	
ttt ctt gtg aca ggt ggc gca cga ggt att aca cct cat tgt gtt			6804
Phe Leu Val Thr Gly Gly Ala Arg Gly Ile Thr Pro His Cys Val			
2255	2260	2265	
cgt gag att gca agt cga tcc ccc gga acc aca ttt gtg ctg gtt			6849
Arg Glu Ile Ala Ser Arg Ser Pro Gly Thr Thr Phe Val Leu Val			
2270	2275	2280	
gga aga agc gaa atg tcc gac gag cct gac tgg gct gtt ggc cac			6894
Gly Arg Ser Glu Met Ser Asp Glu Pro Asp Trp Ala Val Gly His			
2285	2290	2295	
tac aat aaa gac ctg gac caa agc aca atg aaa cac ttg aaa gca			6939
Tyr Asn Lys Asp Leu Asp Gln Ser Thr Met Lys His Leu Lys Ala			
2300	2305	2310	
acg cat gct gct gga ggg gta aaa cct acg cct aaa gca cat cgt			6984
Thr His Ala Ala Gly Gly Val Lys Pro Thr Pro Lys Ala His Arg			
2315	2320	2325	
gca ctt gtg aac agg gtc act ggc tca cgg gag gta cga gaa tct			7029
Ala Leu Val Asn Arg Val Thr Gly Ser Arg Glu Val Arg Glu Ser			
2330	2335	2340	
ctt aga gca atc cag gag gca ggg gca aat gtc gaa tat atc gcc			7074
Leu Arg Ala Ile Gln Glu Ala Gly Ala Asn Val Glu Tyr Ile Ala			

2345	2350	2355	
tgt gat gtt tcg gat gaa aac aag gtc cgc caa ctt gtg caa aga			7119
Cys Asp Val Ser Asp Glu Asn Lys Val Arg Gln Leu Val Gln Arg			
2360	2365	2370	
gtg gag caa aag tat ggc tgt gaa ata act ggg att tgg cat gca			7164
Val Glu Gln Lys Tyr Gly Cys Glu Ile Thr Gly Ile Trp His Ala			
2375	2380	2385	
agc ggg gtt ctt cgt gac aaa ctt gtc gag caa aag act aca gac			7209
Ser Gly Val Leu Arg Asp Lys Leu Val Glu Gln Lys Thr Thr Asp			
2390	2395	2400	
gac ttt gag gca gtt ttt ggg acc aag gtg act ggc ctt gta aac			7254
Asp Phe Glu Ala Val Phe Gly Thr Lys Val Thr Gly Leu Val Asn			
2405	2410	2415	
atc gtg tca caa gtc aat atg tct aag cta cga cac ttc atc ctc			7299
Ile Val Ser Gln Val Asn Met Ser Lys Leu Arg His Phe Ile Leu			
2420	2425	2430	
ttc agt tct ttg gct gga ttt cat ggg aac aag ggc caa acg gat			7344
Phe Ser Ser Leu Ala Gly Phe His Gly Asn Lys Gly Gln Thr Asp			
2435	2440	2445	
tat gca att gct aat gaa gcc ttg aac aaa atc gcg cat act ctc			7389
Tyr Ala Ile Ala Asn Glu Ala Leu Asn Lys Ile Ala His Thr Leu			
2450	2455	2460	
tca gcg ttt ttg ccc aaa ctg aat gca aag gtg cta gac ttc ggt			7434
Ser Ala Phe Leu Pro Lys Leu Asn Ala Lys Val Leu Asp Phe Gly			
2465	2470	2475	
ccg tgg gta ggt tca gga atg gta acc gaa aca ctt gag aag cat			7479
Pro Trp Val Gly Ser Gly Met Val Thr Glu Thr Leu Glu Lys His			
2480	2485	2490	
ttt aaa gct atg ggg gtt cag act att cct ctc gag cca gga gca			7524
Phe Lys Ala Met Gly Val Gln Thr Ile Pro Leu Glu Pro Gly Ala			
2495	2500	2505	
cgg act gtt gcg caa atc att ttg gca agt tcg cca ccg caa tcg			7569
Arg Thr Val Ala Gln Ile Ile Leu Ala Ser Ser Pro Pro Gln Ser			

2510	2515	2520	
ctt ttg ggg aac tgg ggc ttt	cca gcc acc aaa ccg	cta caa cgc	7614
Leu Leu Gly Asn Trp Gly Phe	Pro Ala Thr Lys Pro	Leu Gln Arg	
2525	2530	2535	
tct aat gta gtc acg ggc aca	ctc tct ccg gaa gag	ata gaa ttc	7659
Ser Asn Val Val Thr Gly Thr	Leu Ser Pro Glu Glu	Ile Glu Phe	
2540	2545	2550	
atc gca gac cac aaa att caa	ggc cgc aag gtg ctt	ccc atg atg	7704
Ile Ala Asp His Lys Ile Gln	Gly Arg Lys Val Leu	Pro Met Met	
2555	2560	2565	
gct gca atc ggg ttc atg gcc	tct att gcg gaa gga	ctc tac ccg	7749
Ala Ala Ile Gly Phe Met Ala	Ser Ile Ala Glu Gly	Leu Tyr Pro	
2570	2575	2580	
ggg tac aat ctg caa ggc gtg	gaa aat gct cag ctc	ttt caa ggc	7794
Gly Tyr Asn Leu Gln Gly Val	Glu Asn Ala Gln Leu	Phe Gln Gly	
2585	2590	2595	
ttg act atc aac caa gag aca	aaa ttt caa atc act	ctc att gag	7839
Leu Thr Ile Asn Gln Glu Thr	Lys Phe Gln Ile Thr	Leu Ile Glu	
2600	2605	2610	
gag cac aac tct gag gaa aac	ctg gat gtc ctg aca	tcc ctt ggt	7884
Glu His Asn Ser Glu Glu Asn	Leu Asp Val Leu Thr	Ser Leu Gly	
2615	2620	2625	
gta atg ttg gaa agc ggg aag	gtg ctt ccc gct tac	cga tgt gtt	7929
Val Met Leu Glu Ser Gly Lys	Val Leu Pro Ala Tyr	Arg Cys Val	
2630	2635	2640	
gta tgc ttg aat aca acc cag	cag cag ccc aag cta	tct cca aaa	7974
Val Cys Leu Asn Thr Thr Gln	Gln Gln Pro Lys Leu	Ser Pro Lys	
2645	2650	2655	
att ctt aac ttg gaa gtt gac	cct gca tgc gag gtt	aac ccc tat	8019
Ile Leu Asn Leu Glu Val Asp	Pro Ala Cys Glu Val	Asn Pro Tyr	
2660	2665	2670	
gat gga aag tcg ttg ttc cac	ggt ccg ctt ttg caa	ttc gtt caa	8064
Asp Gly Lys Ser Leu Phe His	Gly Pro Leu Leu Gln	Phe Val Gln	

2675	2680	2685	
caa gtg ttg cac tca agt acc	aaa ggc ctc gtt gcc	aag tgc cgc	8109
Gln Val Leu His Ser Ser Thr	Lys Gly Leu Val Ala	Lys Cys Arg	
2690	2695	2700	
gcg ctt cca atc aaa gaa gcc	atc cga ggg cca ttt	atc aag caa	8154
Ala Leu Pro Ile Lys Glu Ala	Ile Arg Gly Pro Phe	Ile Lys Gln	
2705	2710	2715	
aca ctc cat gat cca att cta	gac gac gtc att ttt	cag cta atg	8199
Thr Leu His Asp Pro Ile Leu	Asp Asp Val Ile Phe	Gln Leu Met	
2720	2725	2730	
ctc gtg tgg tgt cgt aat gct	cta gga agt gca tcg	cta ccc aac	8244
Leu Val Trp Cys Arg Asn Ala	Leu Gly Ser Ala Ser	Leu Pro Asn	
2735	2740	2745	
aga att gaa aag atg tca tac	ttt ggg aat gtc tca	gaa ggt agc	8289
Arg Ile Glu Lys Met Ser Tyr	Phe Gly Asn Val Ser	Glu Gly Ser	
2750	2755	2760	
act ttc ttt gcc tca gtt aca	cct gtg gga cca aga	gta cca aag	8334
Thr Phe Phe Ala Ser Val Thr	Pro Val Gly Pro Arg	Val Pro Lys	
2765	2770	2775	
gat ccc gtg atc aaa atg cag	ttt ctt ctc caa gat	gaa tcc ggc	8379
Asp Pro Val Ile Lys Met Gln	Phe Leu Leu Gln Asp	Glu Ser Gly	
2780	2785	2790	
aac aca ttt tca tcg ggg gag	ggc tcg gtt gtg ctt	agt gac gaa	8424
Asn Thr Phe Ser Ser Gly Glu	Gly Ser Val Val Leu	Ser Asp Glu	
2795	2800	2805	
ctc gtc ttt tga			8436
Leu Val Phe			
2810			
<210> 39			
<211> 2811			
<212> PRT			
<213> Thraustochytrium sp.			

<400> 39

Met Lys Asp Met Glu Asp Arg Arg Val Ala Ile Val Gly Met Ser Ala
1 5 10 15

His Leu Pro Cys Gly Thr Asp Val Lys Glu Ser Trp Gln Ala Ile Arg
20 25 30

Asp Gly Ile Asp Cys Leu Ser Asp Leu Pro Ala Asp Arg Leu Asp Val
35 40 45

Thr Ala Tyr Tyr Asn Pro Asn Lys Ala Thr Lys Asp Lys Ile Tyr Cys
50 55 60

Lys Arg Gly Gly Phe Ile Pro Asn Tyr Asp Phe Asp Pro Arg Glu Phe
65 70 75 80

Gly Leu Asn Met Phe Gln Met Glu Asp Ser Asp Ala Asn Gln Thr Leu
85 90 95

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

Phe Thr Lys Glu Lys Lys Asn Ile Gly Cys Val Leu Gly Ile Gly Gly
115 120 125

Gly Gln Lys Ala Ser His Glu Phe Tyr Ser Arg Leu Asn Tyr Val Val
130 135 140

Val Glu Lys Val Leu Arg Lys Met Gly Leu Pro Asp Ala Asp Val Glu
145 150 155 160

Glu Ala Val Glu Lys Tyr Lys Ala Asn Phe Pro Glu Trp Arg Leu Asp
165 170 175

Ser Phe Pro Gly Phe Leu Gly Asn Val Thr Ala Gly Arg Cys Ser Asn

	180		185		190
Thr Phe Asn Met Glu Gly Met Asn Cys Val Val Asp Ala Ala Cys Ala					
195		200		205	
Ser Ser Leu Ile Ala Ile Lys Val Ala Val Glu Glu Leu Leu Phe Gly					
210		215		220	
Asp Cys Asp Thr Met Ile Ala Gly Ala Thr Cys Thr Asp Asn Ser Leu					
225		230		235	240
Gly Met Tyr Met Ala Phe Ser Lys Thr Pro Val Phe Ser Thr Asp Pro					
	245		250		255
Ser Val Arg Ala Tyr Asp Glu Lys Thr Lys Gly Met Leu Ile Gly Glu					
	260		265		270
Gly Ser Ala Met Phe Val Leu Lys Arg Tyr Ala Asp Ala Val Arg Asp					
	275		280		285
Gly Asp Thr Ile His Ala Val Leu Arg Ser Cys Ser Ser Ser Ser Asp					
	290		295		300
Gly Lys Ala Ala Gly Ile Tyr Thr Pro Thr Ile Ser Gly Gln Glu Glu					
305		310		315	320
Ala Leu Arg Arg Ala Tyr Ala Arg Ala Gly Val Cys Pro Ser Thr Ile					
	325		330		335
Gly Leu Val Glu Gly His Gly Thr Gly Thr Pro Val Gly Asp Arg Ile					
	340		345		350
Glu Leu Thr Ala Leu Arg Asn Leu Phe Asp Lys Ala Phe Gly Ser Lys					
	355		360		365
Lys Glu Gln Ile Ala Val Gly Ser Ile Lys Ser Gln Ile Gly His Leu					

370

375

380

Lys Ser Val Ala Gly Phe Ala Gly Leu Val Lys Ala Val Leu Ala Leu
385 390 395 400

Lys His Lys Thr Leu Pro Gly Ser Ile Asn Val Asp Gln Pro Pro Leu
405 410 415

Leu Tyr Asp Gly Thr Gln Ile Gln Asp Ser Ser Leu Tyr Ile Asn Lys
420 425 430

Thr Asn Arg Pro Trp Phe Thr Gln Asn Lys Leu Pro Arg Arg Ala Gly
435 440 445

Val Ser Ser Phe Gly Phe Gly Gly Ala Asn Tyr His Ala Val Leu Glu
450 455 460

Glu Phe Glu Pro Glu His Glu Lys Pro Tyr Arg Leu Asn Thr Val Gly
465 470 475 480

His Pro Val Leu Leu Tyr Ala Pro Ser Val Glu Ala Leu Lys Val Leu
485 490 495

Cys Asn Asp Gln Leu Ala Glu Leu Thr Ile Ala Leu Glu Glu Ala Lys
500 505 510

Thr His Lys Asn Val Asp Lys Val Cys Gly Tyr Lys Phe Ile Asp Glu
515 520 525

Phe Gln Leu Gln Gly Ser Cys Pro Pro Glu Asn Pro Arg Val Gly Phe
530 535 540

Leu Ala Thr Leu Pro Thr Ser Asn Ile Ile Val Ala Leu Lys Ala Ile
545 550 555 560

Leu Ala Gln Leu Asp Ala Lys Pro Asp Ala Lys Lys Trp Asp Leu Pro
565 570 575

His Lys Lys Ala Phe Gly Ala Thr Phe Ala Ser Ser Ser Val Lys Gly
580 585 590

Ser Val Ala Ala Leu Phe Ala Gly Gln Gly Thr Gln Tyr Leu Asn Met
595 600 605

Phe Ser Asp Val Ala Met Asn Trp Pro Pro Phe Arg Asp Ser Ile Val
610 615 620

Ala Met Glu Glu Ala Gln Thr Glu Val Phe Glu Gly Gln Val Glu Pro
625 630 635 640

Ile Ser Lys Val Leu Phe Pro Arg Glu Arg Tyr Ala Ser Glu Ser Glu
645 650 655

Gln Gly Asn Glu Leu Leu Cys Leu Thr Glu Tyr Ser Gln Pro Thr Thr
660 665 670

Ile Ala Ala Ala Val Gly Ala Phe Asp Ile Phe Lys Ala Ala Gly Phe
675 680 685

Lys Pro Asp Met Val Gly Gly His Ser Leu Gly Glu Phe Ala Ala Leu
690 695 700

Tyr Ala Ala Gly Ser Ile Ser Arg Asp Asp Leu Tyr Lys Leu Val Cys
705 710 715 720

Lys Arg Ala Lys Ala Met Ala Asn Ala Ser Asp Gly Ala Met Ala Ala
725 730 735

Val Ile Gly Pro Asp Ala Arg Leu Val Thr Pro Gln Asn Ser Asp Val
740 745 750

Tyr Val Ala Asn Phe Asn Ser Ala Thr Gln Val Val Ile Ser Gly Thr
755 760 765

Val Gln Gly Val Lys Glu Glu Ser Lys Leu Leu Ile Ser Lys Gly Phe
770 775 780

Arg Val Leu Pro Leu Lys Cys Gln Gly Ala Phe His Ser Pro Leu Met
785 790 795 800

Gly Pro Ser Glu Asp Ser Phe Lys Ser Leu Val Glu Thr Cys Thr Ile
805 810 815

Ser Pro Pro Lys Asn Val Lys Phe Phe Cys Asn Val Ser Gly Lys Glu
820 825 830

Ser Pro Asn Pro Lys Gln Thr Leu Lys Ser His Met Thr Ser Ser Val
835 840 845

Gln Phe Glu Glu Gln Ile Arg Asn Met Tyr Asp Ala Gly Ala Arg Val
850 855 860

Phe Leu Glu Phe Gly Pro Arg Gln Val Leu Ala Lys Leu Ile Ala Glu
865 870 875 880

Met Phe Pro Ser Cys Thr Ala Ile Ser Val Asn Pro Ala Ser Ser Gly
885 890 895

Asp Ser Asp Val Gln Leu Arg Leu Ala Ala Val Lys Phe Ala Val Ser
900 905 910

Gly Ala Ala Leu Ser Thr Phe Asp Pro Trp Glu Tyr Arg Lys Pro Gln
915 920 925

Asp Leu Leu Ile Arg Lys Pro Arg Lys Thr Ala Leu Val Leu Ser Ala
930 935 940

Ala Thr Tyr Val Ser Pro Lys Thr Leu Ala Glu Arg Lys Lys Ala Met
945 950 955 960

Glu Asp Ile Lys Leu Val Ser Ile Thr Pro Arg Asp Ser Met Val Ser
965 970 975

Ile Gly Lys Ile Ala Gln Glu Val Arg Thr Ala Lys Gln Pro Leu Glu
980 985 990

Thr Glu Ile Arg Arg Leu Asn Lys Glu Leu Glu His Leu Lys Arg Glu
995 1000 1005

Leu Ala Ala Ala Lys Ala Ser Val Lys Ser Ala Ser Lys Ser Ser
1010 1015 1020

Lys Glu Arg Ser Val Leu Ser Lys His Arg Ala Leu Leu Gln Asn
1025 1030 1035

Ile Leu Gln Asp Tyr Asp Asp Leu Arg Val Val Pro Phe Ala Val
1040 1045 1050

Arg Ser Val Ala Val Asp Asn Thr Ala Pro Tyr Ala Asp Gln Val
1055 1060 1065

Ser Thr Pro Ala Ser Glu Arg Ser Ala Ser Pro Leu Phe Glu Lys
1070 1075 1080

Arg Ser Ser Val Ser Ser Ala Arg Leu Ala Glu Ala Glu Ala Ala
1085 1090 1095

Val Leu Ser Val Leu Ala Asp Lys Thr Gly Tyr Asp Ser Ser Met
1100 1105 1110

Ile Glu Met Asp Met Asp Leu Glu Ser Glu Leu Gly Val Asp Ser
1115 1120 1125

Ile Lys Arg Val Glu Ile Met Ser Glu Val Gln Thr Leu Leu Ser

1130

1135

1140

Val Glu Val Ser Asp Val Asp Ala Leu Ser Arg Thr Lys Thr Val
1145 1150 1155

Gly Asp Val Ile Glu Ala Met Lys Leu Glu Leu Gly Gly Pro Gln
1160 1165 1170

Gly Gln Thr Leu Thr Ala Glu Ser Ile Arg Gln Pro Pro Val Ser
1175 1180 1185

Glu Pro Ala Val Pro Thr Ser Ser Ser Ser Ile Ala Asn Val
1190 1195 1200

Ser Ser Ala Arg Leu Ala Glu Ala Glu Ala Ala Val Leu Ser Val
1205 1210 1215

Leu Ala Asp Lys Thr Gly Tyr Asp Ser Ser Met Ile Glu Met Asp
1220 1225 1230

Met Asp Leu Glu Ser Glu Leu Gly Val Asp Ser Ile Lys Arg Val
1235 1240 1245

Glu Ile Met Ser Glu Val Gln Thr Leu Leu Ser Val Glu Val Ser
1250 1255 1260

Asp Val Asp Ala Leu Ser Arg Thr Lys Thr Val Gly Asp Val Ile
1265 1270 1275

Glu Ala Met Lys Leu Glu Leu Gly Gly Pro Gln Gly Gln Thr Leu
1280 1285 1290

Thr Ala Glu Ser Ile Arg Gln Pro Pro Val Ser Glu Pro Ala Val
1295 1300 1305

Pro Thr Ser Ser Ser Ser Ser Ile Ala Asn Val Ser Ser Ala Arg
1310 1315 1320

Leu Ala Glu Ala Glu Ala Ala Val Leu Ser Val Leu Ala Asp Lys
1325 1330 1335

Thr Gly Tyr Asp Ser Ser Met Ile Glu Met Asp Met Asp Leu Glu
1340 1345 1350

Ser Glu Leu Gly Val Asp Ser Ile Lys Arg Val Glu Ile Met Ser
1355 1360 1365

Glu Val Gln Thr Leu Leu Ser Val Glu Val Ser Asp Val Asp Ala
1370 1375 1380

Leu Ser Arg Thr Lys Thr Val Gly Asp Val Ile Glu Ala Met Lys
1385 1390 1395

Leu Glu Leu Gly Gly Pro Gln Gly Gln Thr Leu Thr Ala Glu Ser
1400 1405 1410

Ile Arg Gln Pro Pro Val Ser Glu Pro Ala Val Pro Thr Ser Ser
1415 1420 1425

Ser Ser Ser Ile Ala Asn Val Leu Ser Ala Arg Leu Ala Glu Ala
1430 1435 1440

Glu Ala Ala Val Leu Ser Val Leu Ala Asp Lys Thr Gly Tyr Asp
1445 1450 1455

Ser Ser Met Ile Glu Met Asp Met Asp Leu Glu Ser Glu Leu Gly
1460 1465 1470

Val Asp Ser Ile Lys Arg Val Glu Ile Met Ser Glu Val Gln Thr
1475 1480 1485

Leu Leu Ser Val Glu Val Ser Asp Val Asp Ala Leu Ser Arg Thr
1490 1495 1500

Lys Thr Val Gly Asp Val Ile Glu Ala Met Lys Leu Glu Leu Gly
1505 1510 1515

Gly Pro Gln Gly Gln Thr Leu Thr Ala Glu Ser Ile Arg Gln Pro
1520 1525 1530

Pro Val Ser Glu Pro Ala Val Pro Thr Ser Ser Ser Ser Ser Ile
1535 1540 1545

Ala Asn Val Ser Ser Ala Arg Leu Ala Glu Ala Glu Ala Ala Val
1550 1555 1560

Leu Ser Val Leu Ala Asp Lys Thr Gly Tyr Asp Ser Ser Met Ile
1565 1570 1575

Glu Met Asp Met Asp Leu Glu Ser Glu Leu Gly Val Asp Ser Ile
1580 1585 1590

Lys Arg Val Glu Ile Met Ser Glu Val Gln Thr Leu Leu Ser Val
1595 1600 1605

Glu Val Ser Asp Val Asp Ala Leu Ser Arg Thr Lys Thr Val Gly
1610 1615 1620

Asp Val Ile Glu Ala Met Lys Leu Glu Leu Gly Gly Pro Gln Gly
1625 1630 1635

Gln Thr Leu Thr Ser Glu Pro Ile His Gln Pro Pro Val Ser Glu
1640 1645 1650

Pro Ala Val Pro Thr Ser Ser Ser Ser Ser Ile Ala Asn Val Ser

Val Gln Ala Leu Leu Ser Val Glu Val Ser Asp Val Asp Ala Leu
1835 1840 1845

Ser Arg Thr Lys Thr Val Gly Asp Val Ile Glu Ala Met Lys Met
1850 1855 1860

Glu Leu Gly Gly Pro Gln Gly Gln Thr Leu Thr Ala Glu Ser Ile
1865 1870 1875

Arg Glu Pro Pro Val Ser Glu Pro Ala Val Pro Thr Ser Ser Ser
1880 1885 1890

Ser Ser Ile Ala Asn Val Ser Ser Ala Arg Leu Ala Glu Ala Glu
1895 1900 1905

Ala Ala Val Leu Ser Val Leu Ala Asp Lys Thr Gly Tyr Asp Ser
1910 1915 1920

Ser Met Ile Glu Met Asp Met Asp Leu Glu Ser Glu Leu Gly Val
1925 1930 1935

Asp Ser Ile Lys Arg Val Glu Ile Met Ser Glu Val Gln Thr Leu
1940 1945 1950

Leu Ser Val Glu Val Ser Asp Val Asp Ala Leu Ser Arg Thr Lys
1955 1960 1965

Thr Val Gly Asp Val Ile Glu Ala Met Lys Leu Glu Leu Gly Glu
1970 1975 1980

Ser Ser Ser Ile Glu Thr Leu Asn Cys Thr Glu Val Glu His Thr
1985 1990 1995

Ser Tyr Lys Ser Val Lys Ala Ser Gly Cys Glu Asn Val Asp Thr
2000 2005 2010

Arg Phe Ala Lys Val Val Gln Ile Ser Leu Pro Ser Lys Leu Lys
2015 2020 2025

Ser Thr Val Ser His Asp Arg Pro Val Ile Val Val Asp Asp Gly
2030 2035 2040

Thr Pro Leu Thr Thr Glu Leu Cys Lys Ile Leu Gly Gly Asn Ile
2045 2050 2055

Val Val Leu Ser Tyr Gln Gly Lys Pro Ala Gly Pro Arg Gly Val
2060 2065 2070

Glu Val Pro Asp Leu Ser Glu Glu Ala Leu Ile Gln Ala Leu Ala
2075 2080 2085

Leu Ile Arg Ser Thr Tyr Gly Val Pro Ile Gly Phe Ile Cys Gln
2090 2095 2100

Gln Val Ser Asn Val Ser Thr Lys Ala Gln Leu Cys Trp Ala Leu
2105 2110 2115

Leu Ala Ala Lys His Leu Lys Lys Asp Leu Asn Ala Val Leu Pro
2120 2125 2130

Asp Ser Arg Ser Phe Phe Val Gly Val Val Arg Leu Asn Gly Lys
2135 2140 2145

Leu Gly Thr Phe Glu Asn Ile Ser Asp Phe Ser Lys Phe Asp Leu
2150 2155 2160

Thr Lys Ala Leu Asp Tyr Gly Gln Arg Gly Ser Leu Leu Gly Leu
2165 2170 2175

Cys Lys Ser Leu Asp Leu Glu Trp Glu Gln Val Phe Cys Arg Gly

Cys Asp Val Ser Asp Glu Asn Lys Val Arg Gln Leu Val Gln Arg
2360 2365 2370

Val Glu Gln Lys Tyr Gly Cys Glu Ile Thr Gly Ile Trp His Ala
2375 2380 2385

Ser Gly Val Leu Arg Asp Lys Leu Val Glu Gln Lys Thr Thr Asp
2390 2395 2400

Asp Phe Glu Ala Val Phe Gly Thr Lys Val Thr Gly Leu Val Asn
2405 2410 2415

Ile Val Ser Gln Val Asn Met Ser Lys Leu Arg His Phe Ile Leu
2420 2425 2430

Phe Ser Ser Leu Ala Gly Phe His Gly Asn Lys Gly Gln Thr Asp
2435 2440 2445

Tyr Ala Ile Ala Asn Glu Ala Leu Asn Lys Ile Ala His Thr Leu
2450 2455 2460

Ser Ala Phe Leu Pro Lys Leu Asn Ala Lys Val Leu Asp Phe Gly
2465 2470 2475

Pro Trp Val Gly Ser Gly Met Val Thr Glu Thr Leu Glu Lys His
2480 2485 2490

Phe Lys Ala Met Gly Val Gln Thr Ile Pro Leu Glu Pro Gly Ala
2495 2500 2505

Arg Thr Val Ala Gln Ile Ile Leu Ala Ser Ser Pro Pro Gln Ser
2510 2515 2520

Leu Leu Gly Asn Trp Gly Phe Pro Ala Thr Lys Pro Leu Gln Arg
2525 2530 2535

Ser Asn Val Val Thr Gly Thr Leu Ser Pro Glu Glu Ile Glu Phe
2540 2545 2550

Ile Ala Asp His Lys Ile Gln Gly Arg Lys Val Leu Pro Met Met
2555 2560 2565

Ala Ala Ile Gly Phe Met Ala Ser Ile Ala Glu Gly Leu Tyr Pro
2570 2575 2580

Gly Tyr Asn Leu Gln Gly Val Glu Asn Ala Gln Leu Phe Gln Gly
2585 2590 2595

Leu Thr Ile Asn Gln Glu Thr Lys Phe Gln Ile Thr Leu Ile Glu
2600 2605 2610

Glu His Asn Ser Glu Glu Asn Leu Asp Val Leu Thr Ser Leu Gly
2615 2620 2625

Val Met Leu Glu Ser Gly Lys Val Leu Pro Ala Tyr Arg Cys Val
2630 2635 2640

Val Cys Leu Asn Thr Thr Gln Gln Gln Pro Lys Leu Ser Pro Lys
2645 2650 2655

Ile Leu Asn Leu Glu Val Asp Pro Ala Cys Glu Val Asn Pro Tyr
2660 2665 2670

Asp Gly Lys Ser Leu Phe His Gly Pro Leu Leu Gln Phe Val Gln
2675 2680 2685

Gln Val Leu His Ser Ser Thr Lys Gly Leu Val Ala Lys Cys Arg
2690 2695 2700

Ala Leu Pro Ile Lys Glu Ala Ile Arg Gly Pro Phe Ile Lys Gln

2705 2710 2715

Thr Leu His Asp Pro Ile Leu Asp Asp Val Ile Phe Gln Leu Met
2720 2725 2730

Leu Val Trp Cys Arg Asn Ala Leu Gly Ser Ala Ser Leu Pro Asn
2735 2740 2745

Arg Ile Glu Lys Met Ser Tyr Phe Gly Asn Val Ser Glu Gly Ser
2750 2755 2760

Thr Phe Phe Ala Ser Val Thr Pro Val Gly Pro Arg Val Pro Lys
2765 2770 2775

Asp Pro Val Ile Lys Met Gln Phe Leu Leu Gln Asp Glu Ser Gly
2780 2785 2790

Asn Thr Phe Ser Ser Gly Glu Gly Ser Val Val Leu Ser Asp Glu
2795 2800 2805

Leu Val Phe
2810

<210> 40
<211> 1500
<212> DNA
<213> Thraustochytrium sp.

<220>
<221> CDS
<222> (1)..(1500)

<400> 40
atg aag gac atg gaa gat aga cgg gtc gct att gtg ggc atg tca gct 48
Met Lys Asp Met Glu Asp Arg Arg Val Ala Ile Val Gly Met Ser Ala
1 5 10 15

cac ttg cct tgt ggg aca gat gtg aag gaa tca tgg cag gct att cgc 96

His	Leu	Pro	Cys	Gly	Thr	Asp	Val	Lys	Glu	Ser	Trp	Gln	Ala	Ile	Arg	
		20						25					30			
gat	gga	atc	gac	tgt	cta	agt	gac	cta	ccc	gcg	gat	cgt	ctc	gac	gtt	144
Asp	Gly	Ile	Asp	Cys	Leu	Ser	Asp	Leu	Pro	Ala	Asp	Arg	Leu	Asp	Val	
	35						40					45				
aca	gct	tac	tac	aat	ccc	aac	aaa	gcc	acg	aaa	gac	aag	atc	tac	tgc	192
Thr	Ala	Tyr	Tyr	Asn	Pro	Asn	Lys	Ala	Thr	Lys	Asp	Lys	Ile	Tyr	Cys	
	50					55					60					
aaa	cgg	ggt	ggc	ttc	atc	ccg	aac	tat	gac	ttc	gac	ccc	cgc	gaa	ttt	240
Lys	Arg	Gly	Gly	Phe	Ile	Pro	Asn	Tyr	Asp	Phe	Asp	Pro	Arg	Glu	Phe	
65				70					75					80		
ggg	ctc	aac	atg	ttt	caa	atg	gaa	gac	tct	gat	gcg	aat	cag	aca	ctt	288
Gly	Leu	Asn	Met	Phe	Gln	Met	Glu	Asp	Ser	Asp	Ala	Asn	Gln	Thr	Leu	
			85					90					95			
acc	ttg	ctc	aaa	gtc	aaa	caa	gct	ctc	gaa	gat	gca	agc	ata	gag	cct	336
Thr	Leu	Leu	Lys	Val	Lys	Gln	Ala	Leu	Glu	Asp	Ala	Ser	Ile	Glu	Pro	
		100					105						110			
ttc	acc	aag	gag	aag	aag	aac	att	gga	tgt	gtt	tta	ggg	att	ggg	ggg	384
Phe	Thr	Lys	Glu	Lys	Lys	Asn	Ile	Gly	Cys	Val	Leu	Gly	Ile	Gly	Gly	
	115					120					125					
ggc	caa	aag	gcg	agt	cat	gag	ttc	tac	tct	cgt	ctc	aac	tac	gtt	gtc	432
Gly	Gln	Lys	Ala	Ser	His	Glu	Phe	Tyr	Ser	Arg	Leu	Asn	Tyr	Val	Val	
	130					135					140					
gtt	gaa	aag	gta	ctt	cgg	aaa	atg	ggg	tta	cca	gat	gct	gat	gtt	gaa	480
Val	Glu	Lys	Val	Leu	Arg	Lys	Met	Gly	Leu	Pro	Asp	Ala	Asp	Val	Glu	
145				150					155					160		
gaa	gct	gtg	gag	aaa	tac	aag	gca	aat	ttt	ccc	gag	tgg	cgc	cta	gac	528
Glu	Ala	Val	Glu	Lys	Tyr	Lys	Ala	Asn	Phe	Pro	Glu	Trp	Arg	Leu	Asp	
		165						170					175			
tct	ttc	cct	ggg	ttt	ctt	ggg	aat	gta	acg	gct	ggg	tgc	agt	aac		576
Ser	Phe	Pro	Gly	Phe	Leu	Gly	Asn	Val	Thr	Ala	Gly	Arg	Cys	Ser	Asn	
		180					185					190				
acc	ttc	aac	atg	gaa	ggg	atg	aac	tgc	gtt	gtg	gat	gct	gca	tgt	gcc	624
Thr	Phe	Asn	Met	Glu	Gly	Met	Asn	Cys	Val	Val	Asp	Ala	Ala	Cys	Ala	

195	200	205	
agt tct cta att gca atc aag gtt gca gtt gaa gag cta ctc ttt ggt Ser Ser Leu Ile Ala Ile Lys Val Ala Val Glu Glu Leu Leu Phe Gly 210 215 220			672
gac tgt gac acc atg att gca ggt gcc acc tgc acg gac aat tca ctt Asp Cys Asp Thr Met Ile Ala Gly Ala Thr Cys Thr Asp Asn Ser Leu 225 230 235 240			720
ggc atg tac atg gcc ttc tct aaa acg cca gtt ttt tct act gac cca Gly Met Tyr Met Ala Phe Ser Lys Thr Pro Val Phe Ser Thr Asp Pro 245 250 255			768
agt gtc cgc gcg tat gat gag aaa aca aaa ggg atg cta att gga gaa Ser Val Arg Ala Tyr Asp Glu Lys Thr Lys Gly Met Leu Ile Gly Glu 260 265 270			816
ggt tca gca atg ttc gtt ctt aaa cgc tat gcg gat gcc gta cgt gat Gly Ser Ala Met Phe Val Leu Lys Arg Tyr Ala Asp Ala Val Arg Asp 275 280 285			864
ggc gac aca att cac gcg gtt ctg cgt tct tgc tct tcg tct agt gat Gly Asp Thr Ile His Ala Val Leu Arg Ser Cys Ser Ser Ser Asp 290 295 300			912
gga aaa gcg gca gga att tat act cct act ata tct gga caa gaa gaa Gly Lys Ala Ala Gly Ile Tyr Thr Pro Thr Ile Ser Gly Gln Glu Glu 305 310 315 320			960
gct ttg cgt cga gcg tat gcc cgt gcg ggg gta tgt cca tct acg atc Ala Leu Arg Arg Ala Tyr Ala Arg Ala Gly Val Cys Pro Ser Thr Ile 325 330 335			1008
ggg ctt gtt gag ggt cac ggg aca ggg acc cct gtt gga gat cgc att Gly Leu Val Glu Gly His Gly Thr Gly Thr Pro Val Gly Asp Arg Ile 340 345 350			1056
gag tta aca gct ctg cgg aac ttg ttt gac aaa gct ttt ggt agc aag Glu Leu Thr Ala Leu Arg Asn Leu Phe Asp Lys Ala Phe Gly Ser Lys 355 360 365			1104
aag gaa caa ata gca gtt ggc agc ata aag tct cag ata ggt cac ctg Lys Glu Gln Ile Ala Val Gly Ser Ile Lys Ser Gln Ile Gly His Leu			1152

370

375

380

aaa tct gtt gcc ggc ttt gcc ggc ttg gtc aaa gct gtg ctt gcg ctt 1200
Lys Ser Val Ala Gly Phe Ala Gly Leu Val Lys Ala Val Leu Ala Leu
385 390 395 400

aaa cac aaa acg ctc cca ggt tgc att aat gtc gac cag cca cct ttg 1248
Lys His Lys Thr Leu Pro Gly Ser Ile Asn Val Asp Gln Pro Pro Leu
405 410 415

ttg tat gac ggt act caa att caa gac tct tct tta tat atc aac aag 1296
Leu Tyr Asp Gly Thr Gln Ile Gln Asp Ser Ser Leu Tyr Ile Asn Lys
420 425 430

aca aat aga cca tgg ttt acg caa aac aag ctt ccg cgt cgg gct ggt 1344
Thr Asn Arg Pro Trp Phe Thr Gln Asn Lys Leu Pro Arg Arg Ala Gly
435 440 445

gtc tca agt ttt gga ttt gga ggt gca aac tac cac gcg gtt ctg gaa 1392
Val Ser Ser Phe Gly Phe Gly Gly Ala Asn Tyr His Ala Val Leu Glu
450 455 460

gaa ttc gag ccc gag cat gaa aaa cca tac cgc ctc aat act gtt gga 1440
Glu Phe Glu Pro Glu His Glu Lys Pro Tyr Arg Leu Asn Thr Val Gly
465 470 475 480

cat cct gtc ctc ttg tac gct ccg tct gtg gaa gcc ctc aaa gta ctt 1488
His Pro Val Leu Leu Tyr Ala Pro Ser Val Glu Ala Leu Lys Val Leu
485 490 495

tgc aac gac cag 1500
Cys Asn Asp Gln
500

<210> 41
<211> 500
<212> PRT
<213> Thraustochytrium sp.

<400> 41

Met Lys Asp Met Glu Asp Arg Arg Val Ala Ile Val Gly Met Ser Ala
1 5 10 15

His Leu Pro Cys Gly Thr Asp Val Lys Glu Ser Trp Gln Ala Ile Arg
20 25 30

Asp Gly Ile Asp Cys Leu Ser Asp Leu Pro Ala Asp Arg Leu Asp Val
35 40 45

Thr Ala Tyr Tyr Asn Pro Asn Lys Ala Thr Lys Asp Lys Ile Tyr Cys
50 55 60

Lys Arg Gly Gly Phe Ile Pro Asn Tyr Asp Phe Asp Pro Arg Glu Phe
65 70 75 80

Gly Leu Asn Met Phe Gln Met Glu Asp Ser Asp Ala Asn Gln Thr Leu
85 90 95

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

Phe Thr Lys Glu Lys Lys Asn Ile Gly Cys Val Leu Gly Ile Gly Gly
115 120 125

Gly Gln Lys Ala Ser His Glu Phe Tyr Ser Arg Leu Asn Tyr Val Val
130 135 140

Val Glu Lys Val Leu Arg Lys Met Gly Leu Pro Asp Ala Asp Val Glu
145 150 155 160

Glu Ala Val Glu Lys Tyr Lys Ala Asn Phe Pro Glu Trp Arg Leu Asp
165 170 175

Ser Phe Pro Gly Phe Leu Gly Asn Val Thr Ala Gly Arg Cys Ser Asn
180 185 190

Thr Phe Asn Met Glu Gly Met Asn Cys Val Val Asp Ala Ala Cys Ala

195

200

205

Ser Ser Leu Ile Ala Ile Lys Val Ala Val Glu Glu Leu Leu Phe Gly
210 215 220

Asp Cys Asp Thr Met Ile Ala Gly Ala Thr Cys Thr Asp Asn Ser Leu
225 230 235 240

Gly Met Tyr Met Ala Phe Ser Lys Thr Pro Val Phe Ser Thr Asp Pro
245 250 255

Ser Val Arg Ala Tyr Asp Glu Lys Thr Lys Gly Met Leu Ile Gly Glu
260 265 270

Gly Ser Ala Met Phe Val Leu Lys Arg Tyr Ala Asp Ala Val Arg Asp
275 280 285

Gly Asp Thr Ile His Ala Val Leu Arg Ser Cys Ser Ser Ser Asp
290 295 300

Gly Lys Ala Ala Gly Ile Tyr Thr Pro Thr Ile Ser Gly Gln Glu Glu
305 310 315 320

Ala Leu Arg Arg Ala Tyr Ala Arg Ala Gly Val Cys Pro Ser Thr Ile
325 330 335

Gly Leu Val Glu Gly His Gly Thr Gly Thr Pro Val Gly Asp Arg Ile
340 345 350

Glu Leu Thr Ala Leu Arg Asn Leu Phe Asp Lys Ala Phe Gly Ser Lys
355 360 365

Lys Glu Gln Ile Ala Val Gly Ser Ile Lys Ser Gln Ile Gly His Leu
370 375 380

Lys Ser Val Ala Gly Phe Ala Gly Leu Val Lys Ala Val Leu Ala Leu

Lys His Lys Thr Leu Pro Gly Ser Ile Asn Val Asp Gln Pro Pro Leu
405 410 415

Leu Tyr Asp Gly Thr Gln Ile Gln Asp Ser Ser Leu Tyr Ile Asn Lys
420 425 430

Thr Asn Arg Pro Trp Phe Thr Gln Asn Lys Leu Pro Arg Arg Ala Gly
435 440 445

Val Ser Ser Phe Gly Phe Gly Gly Ala Asn Tyr His Ala Val Leu Glu
450 455 460

Glu Phe Glu Pro Glu His Glu Lys Pro Tyr Arg Leu Asn Thr Val Gly
465 470 475 480

His Pro Val Leu Leu Tyr Ala Pro Ser Val Glu Ala Leu Lys Val Leu
485 490 495

Cys Asn Asp Gln
500

<210>	42
<211>	1500
<212>	DNA
<213>	Thraustochytrium sp.

<220>
<221> CDS
<222> (1)..(1500)

<400> 42
 ctt gcg gag ctc aca att gca ttg gaa gag gca aaa aca cat aaa aat 48
 Leu Ala Glu Leu Thr Ile Ala Leu Glu Glu Ala Lys Thr His Lys Asn
 1 5 10 15

gtt gac aaa gtt tgt ggc tac aag ttt att gac gaa ttt cag ctc caa 96
Val Asp Lys Val Cys Gly Tyr Lys Phe Ile Asp Glu Phe Gln Leu Gln
20 25 30

gga agc tgt cct cca gaa aat ccg aga gta gga ttt tta gca aca ctg Gly Ser Cys Pro Pro Glu Asn Pro Arg Val Gly Phe Leu Ala Thr Leu 35 40 45	144
cct act tca aat atc att gtc gcg ctt aag gca att ctc gcg cag ctt Pro Thr Ser Asn Ile Ile Val Ala Leu Lys Ala Ile Leu Ala Gln Leu 50 55 60	192
gat gca aaa cca gat gcg aag aaa tgg gat ttg cct cat aaa aag gct Asp Ala Lys Pro Asp Ala Lys Lys Trp Asp Leu Pro His Lys Lys Ala 65 70 75 80	240
ttt ggg gct acc ttc gca tcg tct tca gtg aaa ggc tct gtt gct gcg Phe Gly Ala Thr Phe Ala Ser Ser Ser Val Lys Gly Ser Val Ala Ala 85 90 95	288
ctc ttc gca gga cag ggt acc cag tac tta aac atg ttc tct gat gtg Leu Phe Ala Gly Gln Gly Thr Gln Tyr Leu Asn Met Phe Ser Asp Val 100 105 110	336
gca atg aac tgg cca ccg ttc cgt gac agc att gtc gca atg gaa gaa Ala Met Asn Trp Pro Pro Phe Arg Asp Ser Ile Val Ala Met Glu Glu 115 120 125	384
gct caa act gag gta ttt gag ggc caa gtt gaa cca att agc aaa gtt Ala Gln Thr Glu Val Phe Glu Gly Gln Val Glu Pro Ile Ser Lys Val 130 135 140	432
ctg ttt cca cga gag cgc tat gca tcc gaa agt gaa cag ggg aat gaa Leu Phe Pro Arg Glu Arg Tyr Ala Ser Glu Ser Glu Gln Gly Asn Glu 145 150 155 160	480
ctt ctt tgc tta aca gag tac tct cag cca act acg ata gca gcc gca Leu Leu Cys Leu Thr Glu Tyr Ser Gln Pro Thr Thr Ile Ala Ala Ala 165 170 175	528
gta ggg gcc ttc gat att ttc aaa gcg gct ggc ttt aag cca gac atg Val Gly Ala Phe Asp Ile Phe Lys Ala Ala Gly Phe Lys Pro Asp Met 180 185 190	576
gtt gga ggg cat tca ctt ggc gaa ttt gct gct ttg tac gcg gct ggg Val Gly Gly His Ser Leu Gly Glu Phe Ala Ala Leu Tyr Ala Ala Gly 195 200 205	624

tcc att tcg cgt gac gac ctg tac aag ctt gtg tgc aaa cgg gca aag Ser Ile Ser Arg Asp Asp Leu Tyr Lys Leu Val Cys Lys Arg Ala Lys 210 215 220	672
gca atg gcg aac gct agt gac gga gct atg gca gca gtg att ggc cca Ala Met Ala Asn Ala Ser Asp Gly Ala Met Ala Ala Val Ile Gly Pro 225 230 235 240	720
gat gca cgt cta gtt acg cca caa aat agt gac gtt tat gtc gca aac Asp Ala Arg Leu Val Thr Pro Gln Asn Ser Asp Val Tyr Val Ala Asn 245 250 255	768
ttc aac tcc gca act caa gta gtc atc agt ggc act gtt caa ggt gtg Phe Asn Ser Ala Thr Gln Val Val Ile Ser Gly Thr Val Gln Gly Val 260 265 270	816
aaa gaa gag tcg aaa ttg ctc att tca aag ggg ttc cgc gta ctg cca Lys Glu Glu Ser Lys Leu Leu Ile Ser Lys Gly Phe Arg Val Leu Pro 275 280 285	864
ctt aaa tgc cag ggc gcc ttc cat tct cct ttg atg ggg cct tct gag Leu Lys Cys Gln Gly Ala Phe His Ser Pro Leu Met Gly Pro Ser Glu 290 295 300	912
gat agt ttc aaa tca ctt gtg gag act tgt acc atc tcg ccg cca aaa Asp Ser Phe Lys Ser Leu Val Glu Thr Cys Thr Ile Ser Pro Pro Lys 305 310 315 320	960
aat gtg aaa ttc ttt tgc aat gtt agt ggc aag gaa agc cca aac cca Asn Val Lys Phe Phe Cys Asn Val Ser Gly Lys Glu Ser Pro Asn Pro 325 330 335	1008
aaa cag acc ctc aag tca cac atg acg tct agc gtt cag ttc gag gag Lys Gln Thr Leu Lys Ser His Met Thr Ser Ser Val Gln Phe Glu Glu 340 345 350	1056
cag att cgt aac atg tac gat gcc gga gca cgt gtt ttt ctg gag ttt Gln Ile Arg Asn Met Tyr Asp Ala Gly Ala Arg Val Phe Leu Glu Phe 355 360 365	1104
gga ccc cgc caa gtc ctt gca aag ctt atc gcg gaa atg ttt ccc tcg Gly Pro Arg Gln Val Leu Ala Lys Leu Ile Ala Glu Met Phe Pro Ser 370 375 380	1152

tgt aca gct atc agc gtt aac ccc gcg agc agt ggt gac agt gac gtg 1200
Cys Thr Ala Ile Ser Val Asn Pro Ala Ser Ser Gly Asp Ser Asp Val
385 390 395 400

caa ctc cgc ctc gcc gcc gta aaa ttc gcg gtc tcg ggt gca gcc ctt 1248
Gln Leu Arg Leu Ala Ala Val Lys Phe Ala Val Ser Gly Ala Ala Leu
405 410 415

agc acc ttt gat cca tgg gag tat cgc aag cca caa gat ctt ctt att 1296
Ser Thr Phe Asp Pro Trp Glu Tyr Arg Lys Pro Gln Asp Leu Leu Ile
420 425 430

cga aaa cca cga aaa act gcc ctt gtt cta tca gca gca aca tat gtt 1344
Arg Lys Pro Arg Lys Thr Ala Leu Val Leu Ser Ala Ala Thr Tyr Val
435 440 445

tcc cca aag act ctt gca gaa cgt aaa aag gct atg gaa gat atc aag 1392
Ser Pro Lys Thr Leu Ala Glu Arg Lys Lys Ala Met Glu Asp Ile Lys
450 455 460

cta gta tcc att aca cca aga gat agt atg gta tca att gga aaa atc 1440
Leu Val Ser Ile Thr Pro Arg Asp Ser Met Val Ser Ile Gly Lys Ile
465 470 475 480

gcg caa gaa gta cgg aca gct aaa cag cct tta gaa acc gaa att cga 1488
Ala Gln Glu Val Arg Thr Ala Lys Gln Pro Leu Glu Thr Glu Ile Arg
485 490 495

aga ctc aac aaa 1500
Arg Leu Asn Lys
500

<210> 43
<211> 500
<212> PRT
<213> Thraustochytrium sp.

<400> 43

Leu Ala Glu Leu Thr Ile Ala Leu Glu Glu Ala Lys Thr His Lys Asn
1 5 10 15

Val Asp Lys Val Cys Gly Tyr Lys Phe Ile Asp Glu Phe Gln Leu Gln
20 25 30

Gly Ser Cys Pro Pro Glu Asn Pro Arg Val Gly Phe Leu Ala Thr Leu
35 40 45

Pro Thr Ser Asn Ile Ile Val Ala Leu Lys Ala Ile Leu Ala Gln Leu
50 55 60

Asp Ala Lys Pro Asp Ala Lys Lys Trp Asp Leu Pro His Lys Lys Ala
65 70 75 80

Phe Gly Ala Thr Phe Ala Ser Ser Ser Val Lys Gly Ser Val Ala Ala
85 90 95

Leu Phe Ala Gly Gln Gly Thr Gln Tyr Leu Asn Met Phe Ser Asp Val
100 105 110

Ala Met Asn Trp Pro Pro Phe Arg Asp Ser Ile Val Ala Met Glu Glu
115 120 125

Ala Gln Thr Glu Val Phe Glu Gly Gln Val Glu Pro Ile Ser Lys Val
130 135 140

Leu Phe Pro Arg Glu Arg Tyr Ala Ser Glu Ser Glu Gln Gly Asn Glu
145 150 155 160

Leu Leu Cys Leu Thr Glu Tyr Ser Gln Pro Thr Thr Ile Ala Ala Ala
165 170 175

Val Gly Ala Phe Asp Ile Phe Lys Ala Ala Gly Phe Lys Pro Asp Met
180 185 190

Val Gly Gly His Ser Leu Gly Glu Phe Ala Ala Leu Tyr Ala Ala Gly
195 200 205

Ser Ile Ser Arg Asp Asp Leu Tyr Lys Leu Val Cys Lys Arg Ala Lys
210 215 220

Ala Met Ala Asn Ala Ser Asp Gly Ala Met Ala Ala Val Ile Gly Pro
225 230 235 240

Asp Ala Arg Leu Val Thr Pro Gln Asn Ser Asp Val Tyr Val Ala Asn
245 250 255

Phe Asn Ser Ala Thr Gln Val Val Ile Ser Gly Thr Val Gln Gly Val
260 265 270

Lys Glu Glu Ser Lys Leu Leu Ile Ser Lys Gly Phe Arg Val Leu Pro
275 280 285

Leu Lys Cys Gln Gly Ala Phe His Ser Pro Leu Met Gly Pro Ser Glu
290 295 300

Asp Ser Phe Lys Ser Leu Val Glu Thr Cys Thr Ile Ser Pro Pro Lys
305 310 315 320

Asn Val Lys Phe Phe Cys Asn Val Ser Gly Lys Glu Ser Pro Asn Pro
325 330 335

Lys Gln Thr Leu Lys Ser His Met Thr Ser Ser Val Gln Phe Glu Glu
340 345 350

Gln Ile Arg Asn Met Tyr Asp Ala Gly Ala Arg Val Phe Leu Glu Phe
355 360 365

Gly Pro Arg Gln Val Leu Ala Lys Leu Ile Ala Glu Met Phe Pro Ser
370 375 380

Cys Thr Ala Ile Ser Val Asn Pro Ala Ser Ser Gly Asp Ser Asp Val
385 390 395 400

Gln Leu Arg Leu Ala Ala Val Lys Phe Ala Val Ser Gly Ala Ala Leu
405 410 415

Ser Thr Phe Asp Pro Trp Glu Tyr Arg Lys Pro Gln Asp Leu Leu Ile
420 425 430

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

Ser Pro Lys Thr Leu Ala Glu Arg Lys Lys Ala Met Glu Asp Ile Lys
450 455 460

Leu Val Ser Ile Thr Pro Arg Asp Ser Met Val Ser Ile Gly Lys Ile
465 470 475 480

Ala Gln Glu Val Arg Thr Ala Lys Gln Pro Leu Glu Thr Glu Ile Arg
485 490 495

Arg Leu Asn Lys
500

<210> 44
<211> 351
<212> DNA
<213> Thraustochytrium sp.

<220>
<221> CDS
<222> (1)..(351)

<400> 44
tcg acc cca gcg tca gag cgg tcg gct tca ccg ctt ttc gag aaa cgc 48
Ser Thr Pro Ala Ser Glu Arg Ser Ala Ser Pro Leu Phe Glu Lys Arg
1 5 10 15

agt tcg gtt tcg tca gca cgc ctc gct gaa gct gaa gcc gcg gta ctg 96
Ser Ser Val Ser Ser Ala Arg Leu Ala Glu Ala Glu Ala Ala Val Leu
20 25 30

agc gtt ctc gca gac aag aca ggc tac gac agc tca atg atc gag atg 144
Ser Val Leu Ala Asp Lys Thr Gly Tyr Asp Ser Ser Met Ile Glu Met
35 40 45

gac atg gac ctg gag agt gag ctt ggc gtt gat agc atc aaa cgc gtg 192
Asp Met Asp Leu Glu Ser Glu Leu Gly Val Asp Ser Ile Lys Arg Val
50 55 60

gag atc atg agc gag gtt caa acg ctg ctc agc gtg gaa gtc tcc gac 240
Glu Ile Met Ser Glu Val Gln Thr Leu Leu Ser Val Glu Val Ser Asp
65 70 75 80

gtt gac gct ctg tca aga acc aag act gtt ggc gac gtc atc gag gcg 288
Val Asp Ala Leu Ser Arg Thr Lys Thr Val Gly Asp Val Ile Glu Ala
85 90 95

atg aag ctg gaa ctc ggt gga ccc caa ggc cag act ttg acc gcg gaa 336
Met Lys Leu Glu Leu Gly Gly Pro Gln Gly Gln Thr Leu Thr Ala Glu
100 105 110

tcg atc cgt cag cca 351
Ser Ile Arg Gln Pro
115

<210> 45
<211> 117
<212> PRT
<213> Thraustochytrium sp.

<400> 45

Ser Thr Pro Ala Ser Glu Arg Ser Ala Ser Pro Leu Phe Glu Lys Arg
1 5 10 15

Ser Ser Val Ser Ser Ala Arg Leu Ala Glu Ala Glu Ala Ala Val Leu
20 25 30

Ser Val Leu Ala Asp Lys Thr Gly Tyr Asp Ser Ser Met Ile Glu Met
35 40 45

Asp Met Asp Leu Glu Ser Glu Leu Gly Val Asp Ser Ile Lys Arg Val

50

55

60

Glu Ile Met Ser Glu Val Gln Thr Leu Leu Ser Val Glu Val Ser Asp
65 70 75 80

Val Asp Ala Leu Ser Arg Thr Lys Thr Val Gly Asp Val Ile Glu Ala
85 90 95

Met Lys Leu Glu Leu Gly Gly Pro Gln Gly Gln Thr Leu Thr Ala Glu
100 105 110

Ser Ile Arg Gln Pro
115

<210> 46
<211> 5
<212> PRT
<213> Thraustochytrium sp.

<220>
<221> MISC_FEATURE
<222> (1)..(5)
<223> Xaa = any amino acid

<400> 46

Leu Gly Xaa Asp Ser
1 5

<210> 47
<211> 2790
<212> DNA
<213> Thraustochytrium sp.

<220>
<221> CDS
<222> (1)..(2790)

<400> 47

tcg acc cca gcg tca gag cgg tcg gct tca ccg ctt ttc gag aaa cgc	48
Ser Thr Pro Ala Ser Glu Arg Ser Ala Ser Pro Leu Phe Glu Lys Arg	
1 5 10 15	
agt tcg gtt tcg tca gca cgc ctc gct gaa gct gaa gcc gcg gta ctg	96
Ser Ser Val Ser Ser Ala Arg Leu Ala Glu Ala Glu Ala Val Leu	
20 25 30	
agc gtt ctc gca gac aag aca ggc tac gac agc tca atg atc gag atg	144
Ser Val Leu Ala Asp Lys Thr Gly Tyr Asp Ser Ser Met Ile Glu Met	
35 40 45	
gac atg gac ctg gag agt gag ctt ggc gtt gat agc atc aaa cgc gtg	192
Asp Met Asp Leu Glu Ser Glu Leu Gly Val Asp Ser Ile Lys Arg Val	
50 55 60	
gag atc atg agc gag gtt caa acg ctg ctc agc gtg gaa gtc tcc gac	240
Glu Ile Met Ser Glu Val Gln Thr Leu Leu Ser Val Glu Val Ser Asp	
65 70 75 80	
gtt gac gct ctg tca aga acc aag act gtt ggc gac gtc atc gag gcg	288
Val Asp Ala Leu Ser Arg Thr Lys Thr Val Gly Asp Val Ile Glu Ala	
85 90 95	
atg aag ctg gaa ctc ggt gga ccc caa ggc cag act ttg acc gcg gaa	336
Met Lys Leu Glu Leu Gly Gly Pro Gln Gly Gln Thr Leu Thr Ala Glu	
100 105 110	
tcg atc cgt cag cca ccg gtg tcc gag cct gct gta ccg acc tca tcg	384
Ser Ile Arg Gln Pro Pro Val Ser Glu Pro Ala Val Pro Thr Ser Ser	
115 120 125	
tca agc agt att gct aat gtt tcg tca gca cgc ctc gct gaa gct gaa	432
Ser Ser Ser Ile Ala Asn Val Ser Ser Ala Arg Leu Ala Glu Ala Glu	
130 135 140	
gct gcg gta ctg agc gtt ctc gca gac aag aca ggc tac gac agc tca	480
Ala Ala Val Leu Ser Val Leu Ala Asp Lys Thr Gly Tyr Asp Ser Ser	
145 150 155 160	
atg atc gag atg gac atg gac ctg gag agc gag ctt ggc gtt gat agc	528
Met Ile Glu Met Asp Met Asp Leu Glu Ser Glu Leu Gly Val Asp Ser	
165 170 175	
atc aaa cgc gtg gag atc atg agc gag gtt caa acg ctg ctc agc gtg	576

Ile Lys Arg Val Glu Ile Met Ser Glu Val Gln Thr Leu Leu Ser Val	
180 185 190	
gaa gtc tcc gac gtt gac gct ctg tca aga act aag act gtt ggc gac	624
Glu Val Ser Asp Val Asp Ala Leu Ser Arg Thr Lys Thr Val Gly Asp	
195 200 205	
gtc atc gag gcg atg aag ctg gaa ctc ggt gga ccc caa ggc cag act	672
Val Ile Glu Ala Met Lys Leu Glu Leu Gly Gly Pro Gln Gly Gln Thr	
210 215 220	
ttg acc gcg gaa tcg atc cgt cag cca ccg gtg tct gag cct gct gta	720
Leu Thr Ala Glu Ser Ile Arg Gln Pro Pro Val Ser Glu Pro Ala Val	
225 230 235 240	
ccg acc tca tcg tca agc agt att gct aat gtt tcg tca gca cgc ctc	768
Pro Thr Ser Ser Ser Ser Ser Ile Ala Asn Val Ser Ser Ala Arg Leu	
245 250 255	
gct gaa gct gaa gcg gcg gta ctg agc gtt ctc gca gac aag aca ggc	816
Ala Glu Ala Glu Ala Ala Val Leu Ser Val Leu Ala Asp Lys Thr Gly	
260 265 270	
tac gac agc tca atg atc gag atg gac atg gac ctg gag agc gag ctt	864
Tyr Asp Ser Ser Met Ile Glu Met Asp Met Asp Leu Glu Ser Glu Leu	
275 280 285	
ggc gtc gac agc atc aaa cgc gtg gag atc atg agc gag gtt caa acg	912
Gly Val Asp Ser Ile Lys Arg Val Glu Ile Met Ser Glu Val Gln Thr	
290 295 300	
ctg ctc agc gtg gaa gtc tcc gac gtt gac gct ctg tca aga acc aag	960
Leu Leu Ser Val Glu Val Ser Asp Val Asp Ala Leu Ser Arg Thr Lys	
305 310 315 320	
act gtt ggc gac gtc atc gag gcg atg aag ctg gaa ctc ggt gga ccc	1008
Thr Val Gly Asp Val Ile Glu Ala Met Lys Leu Glu Leu Gly Gly Pro	
325 330 335	
caa ggc cag act ttg acc gcg gaa tcg atc cgt cag cca ccg gtg tcc	1056
Gln Gly Gln Thr Leu Thr Ala Glu Ser Ile Arg Gln Pro Pro Val Ser	
340 345 350	
gag cct gct gta ccg acc tca tcg tca agc agt att gct aat gtt ttg	1104
Glu Pro Ala Val Pro Thr Ser Ser Ser Ser Ser Ile Ala Asn Val Leu	

355	360	365	
tca gca cgc ctc gct gaa gct gaa gcc gcg gta ctg agc gtt ctc gca			1152
Ser Ala Arg Leu Ala Glu Ala Glu Ala Ala Val Leu Ser Val Leu Ala			
370	375	380	
gac aag aca ggc tac gac agc tca atg atc gag atg gac atg gac ctg			1200
Asp Lys Thr Gly Tyr Asp Ser Ser Met Ile Glu Met Asp Met Asp Leu			
385	390	395	400
gag agc gag ctt ggc gtt gat agc atc aaa cgc gtg gag atc atg agc			1248
Glu Ser Glu Leu Gly Val Asp Ser Ile Lys Arg Val Glu Ile Met Ser			
405	410	415	
gag gtt caa acg ttg ctc agc gtg gaa gtc tcc gac gtt gac gct ctg			1296
Glu Val Gln Thr Leu Leu Ser Val Glu Val Ser Asp Val Asp Ala Leu			
420	425	430	
tca aga acc aag act gtt ggc gac gtc atc gag gcg atg aag ctg gaa			1344
Ser Arg Thr Lys Thr Val Gly Asp Val Ile Glu Ala Met Lys Leu Glu			
435	440	445	
ctc ggt gga ccc caa ggc cag act ttg acc gcg gaa tcg atc cgt cag			1392
Leu Gly Gly Pro Gln Gly Gln Thr Leu Thr Ala Glu Ser Ile Arg Gln			
450	455	460	
cca ccg gtg tct gag cct gct gta ccg acc tca tcg tca agc agt att			1440
Pro Pro Val Ser Glu Pro Ala Val Pro Thr Ser Ser Ser Ser Ile			
465	470	475	480
gct aat gtt tcg tca gca cgc ctc gct gaa gct gaa gcc gcg gta ctg			1488
Ala Asn Val Ser Ser Ala Arg Leu Ala Glu Ala Glu Ala Ala Val Leu			
485	490	495	
agc gtt ctc gca gac aag aca ggc tac gac agc tca atg atc gag atg			1536
Ser Val Leu Ala Asp Lys Thr Gly Tyr Asp Ser Ser Met Ile Glu Met			
500	505	510	
gac atg gac ctg gag agt gag ctt ggc gtc gac agc atc aaa cgc gtg			1584
Asp Met Asp Leu Glu Ser Glu Leu Gly Val Asp Ser Ile Lys Arg Val			
515	520	525	
gag atc atg agc gag gtt caa acg ctg ctc agc gtg gaa gtc tcc gac			1632
Glu Ile Met Ser Glu Val Gln Thr Leu Leu Ser Val Glu Val Ser Asp			

530	535	540	
gtt gac gct ctg tca aga acc aag act gtt ggc gac gtc atc gag gcg			1680
Val Asp Ala Leu Ser Arg Thr Lys Thr Val Gly Asp Val Ile Glu Ala			
545	550	555	560
atg aag ctg gaa ctc ggt gga ccc caa ggc cag act ttg acc tct gaa			1728
Met Lys Leu Glu Leu Gly Gly Pro Gln Gly Gln Thr Leu Thr Ser Glu			
	565	570	575
ccg atc cat cag cca cca gtg tcc gag cct gct gta ccg acc tca tcg			1776
Pro Ile His Gln Pro Pro Val Ser Glu Pro Ala Val Pro Thr Ser Ser			
	580	585	590
tca agc agt att gct aat gtt tct tca gca cgc ctc gct gaa gct gaa			1824
Ser Ser Ser Ile Ala Asn Val Ser Ser Ala Arg Leu Ala Glu Ala Glu			
	595	600	605
gcc gcg gta ctg agc gtt ctc gca gac aag aca ggc tac gac agc tca			1872
Ala Ala Val Leu Ser Val Leu Ala Asp Lys Thr Gly Tyr Asp Ser Ser			
	610	615	620
atg atc gag atg gac atg gac ctg gag agc gag ctt ggc gtt gat agc			1920
Met Ile Glu Met Asp Met Asp Leu Glu Ser Glu Leu Gly Val Asp Ser			
625	630	635	640
atc aaa cgc gtg gaa atc atg agc gag gtt caa acg ctg ctc agc gtg			1968
Ile Lys Arg Val Glu Ile Met Ser Glu Val Gln Thr Leu Leu Ser Val			
	645	650	655
gaa gtc tcc gac gtt gac gct ctg tca aga acc aag act gtt ggc gac			2016
Glu Val Ser Asp Val Asp Ala Leu Ser Arg Thr Lys Thr Val Gly Asp			
	660	665	670
gtc atc gag gcg atg aag atg gaa ctc ggt gga ccc caa ggc cag act			2064
Val Ile Glu Ala Met Lys Met Glu Leu Gly Gly Pro Gln Gly Gln Thr			
	675	680	685
ttg acc gcg gaa tcg atc cgt cag cca ccg gtg tct gag cct gct gta			2112
Leu Thr Ala Glu Ser Ile Arg Gln Pro Pro Val Ser Glu Pro Ala Val			
	690	695	700
ccg acc tca tcg tca agc agt att gct aat gtt tcg tca gca cgc ctc			2160
Pro Thr Ser Ser Ser Ser Ser Ile Ala Asn Val Ser Ser Ala Arg Leu			

705	710	715	720
gct gaa gct gaa gcg gcg gta ctg agc gtt ctc gca gac aag aca ggc 2208			
Ala Glu Ala Glu Ala Ala Val Leu Ser Val Leu Ala Asp Lys Thr Gly			
	725	730	735
tac gac agc tca atg atc gag atg gac atg gac ctg gag agc gag ctt 2256			
Tyr Asp Ser Ser Met Ile Glu Met Asp Met Asp Leu Glu Ser Glu Leu			
	740	745	750
ggc gtt gat agc atc aaa cgc gtg gag atc atg agc gag gtt caa gcg 2304			
Gly Val Asp Ser Ile Lys Arg Val Glu Ile Met Ser Glu Val Gln Ala			
	755	760	765
ctg ctc agc gtg gaa gtc tcc gac gtt gac gct ctg tca aga acc aag 2352			
Leu Leu Ser Val Glu Val Ser Asp Val Asp Ala Leu Ser Arg Thr Lys			
	770	775	780
act gtt ggc gac gtc atc gag gcg atg aag atg gaa ctc ggt gga ccc 2400			
Thr Val Gly Asp Val Ile Glu Ala Met Lys Met Glu Leu Gly Gly Pro			
	785	790	800
caa ggc cag act ttg acc gca gaa tcg atc cgt gag cca ccg gtg tct 2448			
Gln Gly Gln Thr Leu Thr Ala Glu Ser Ile Arg Glu Pro Pro Val Ser			
	805	810	815
gag cct gct gta ccg acc tca tcg tca agt agt atc gct aat gtt tct 2496			
Glu Pro Ala Val Pro Thr Ser Ser Ser Ser Ser Ile Ala Asn Val Ser			
	820	825	830
tca gct cgc ctc gct gaa gct gaa gcc gcg gta ctg agc gtt ctc gca 2544			
Ser Ala Arg Leu Ala Glu Ala Glu Ala Ala Val Leu Ser Val Leu Ala			
	835	840	845
gac aag aca ggc tac gac agc tca atg atc gag atg gac atg gac ctg 2592			
Asp Lys Thr Gly Tyr Asp Ser Ser Met Ile Glu Met Asp Met Asp Leu			
	850	855	860
gag agt gag ctt ggc gtc gac agc atc aaa cgc gtg gag atc atg agc 2640			
Glu Ser Glu Leu Gly Val Asp Ser Ile Lys Arg Val Glu Ile Met Ser			
	865	870	880
gag gtt caa acg ttg ctc agc gtg gaa gtc tcc gac gtt gac gct ctg 2688			
Glu Val Gln Thr Leu Leu Ser Val Glu Val Ser Asp Val Asp Ala Leu			

885

890

895

tca aga acc aag act gtt ggc gac gtc atc gag gcg atg aag ctg gaa 2736
Ser Arg Thr Lys Thr Val Gly Asp Val Ile Glu Ala Met Lys Leu Glu
900 905 910

ctt ggg gaa tca tca agt att gag act ctc aat tgt acc gag gtt gag 2784
Leu Gly Glu Ser Ser Ser Ile Glu Thr Leu Asn Cys Thr Glu Val Glu
915 920 925

cac acg 2790
His Thr
930

<210> 48
<211> 930
<212> PRT
<213> Thraustochytrium sp.

<400> 48

Ser Thr Pro Ala Ser Glu Arg Ser Ala Ser Pro Leu Phe Glu Lys Arg
1 5 10 15

Ser Ser Val Ser Ser Ala Arg Leu Ala Glu Ala Glu Ala Ala Val Leu
20 25 30

Ser Val Leu Ala Asp Lys Thr Gly Tyr Asp Ser Ser Met Ile Glu Met
35 40 45

Asp Met Asp Leu Glu Ser Glu Leu Gly Val Asp Ser Ile Lys Arg Val
50 55 60

Glu Ile Met Ser Glu Val Gln Thr Leu Leu Ser Val Glu Val Ser Asp
65 70 75 80

Val Asp Ala Leu Ser Arg Thr Lys Thr Val Gly Asp Val Ile Glu Ala
85 90 95

Met Lys Leu Glu Leu Gly Gly Pro Gln Gly Gln Thr Leu Thr Ala Glu
100 105 110

Ser Ile Arg Gln Pro Pro Val Ser Glu Pro Ala Val Pro Thr Ser Ser
115 120 125

Ser Ser Ser Ile Ala Asn Val Ser Ser Ala Arg Leu Ala Glu Ala Glu
130 135 140

Ala Ala Val Leu Ser Val Leu Ala Asp Lys Thr Gly Tyr Asp Ser Ser
145 150 155 160

Met Ile Glu Met Asp Met Asp Leu Glu Ser Glu Leu Gly Val Asp Ser
165 170 175

Ile Lys Arg Val Glu Ile Met Ser Glu Val Gln Thr Leu Leu Ser Val
180 185 190

Glu Val Ser Asp Val Asp Ala Leu Ser Arg Thr Lys Thr Val Gly Asp
195 200 205

Val Ile Glu Ala Met Lys Leu Glu Leu Gly Gly Pro Gln Gly Gln Thr
210 215 220

Leu Thr Ala Glu Ser Ile Arg Gln Pro Pro Val Ser Glu Pro Ala Val
225 230 235 240

Pro Thr Ser Ser Ser Ser Ser Ile Ala Asn Val Ser Ser Ala Arg Leu
245 250 255

Ala Glu Ala Glu Ala Ala Val Leu Ser Val Leu Ala Asp Lys Thr Gly
260 265 270

Tyr Asp Ser Ser Met Ile Glu Met Asp Met Asp Leu Glu Ser Glu Leu
275 280 285

Gly Val Asp Ser Ile Lys Arg Val Glu Ile Met Ser Glu Val Gln Thr
290 295 300

Leu Leu Ser Val Glu Val Ser Asp Val Asp Ala Leu Ser Arg Thr Lys
305 310 315 320

Thr Val Gly Asp Val Ile Glu Ala Met Lys Leu Glu Leu Gly Gly Pro
325 330 335

Gln Gly Gln Thr Leu Thr Ala Glu Ser Ile Arg Gln Pro Pro Val Ser
340 345 350

Glu Pro Ala Val Pro Thr Ser Ser Ser Ser Ile Ala Asn Val Leu
355 360 365

Ser Ala Arg Leu Ala Glu Ala Glu Ala Val Leu Ser Val Leu Ala
370 375 380

Asp Lys Thr Gly Tyr Asp Ser Ser Met Ile Glu Met Asp Met Asp Leu
385 390 395 400

Glu Ser Glu Leu Gly Val Asp Ser Ile Lys Arg Val Glu Ile Met Ser
405 410 415

Glu Val Gln Thr Leu Leu Ser Val Glu Val Ser Asp Val Asp Ala Leu
420 425 430

Ser Arg Thr Lys Thr Val Gly Asp Val Ile Glu Ala Met Lys Leu Glu
435 440 445

Leu Gly Gly Pro Gln Gly Gln Thr Leu Thr Ala Glu Ser Ile Arg Gln
450 455 460

Pro Pro Val Ser Glu Pro Ala Val Pro Thr Ser Ser Ser Ser Ile
465 470 475 480

Ala Asn Val Ser Ser Ala Arg Leu Ala Glu Ala Glu Ala Ala Val Leu
485 490 495

Ser Val Leu Ala Asp Lys Thr Gly Tyr Asp Ser Ser Met Ile Glu Met
500 505 510

Asp Met Asp Leu Glu Ser Glu Leu Gly Val Asp Ser Ile Lys Arg Val
515 520 525

Glu Ile Met Ser Glu Val Gln Thr Leu Leu Ser Val Glu Val Ser Asp
530 535 540

Val Asp Ala Leu Ser Arg Thr Lys Thr Val Gly Asp Val Ile Glu Ala
545 550 555 560

Met Lys Leu Glu Leu Gly Gly Pro Gln Gly Gln Thr Leu Thr Ser Glu
565 570 575

Pro Ile His Gln Pro Pro Val Ser Glu Pro Ala Val Pro Thr Ser Ser
580 585 590

Ser Ser Ser Ile Ala Asn Val Ser Ser Ala Arg Leu Ala Glu Ala Glu
595 600 605

Ala Ala Val Leu Ser Val Leu Ala Asp Lys Thr Gly Tyr Asp Ser Ser
610 615 620

Met Ile Glu Met Asp Met Asp Leu Glu Ser Glu Leu Gly Val Asp Ser
625 630 635 640

Ile Lys Arg Val Glu Ile Met Ser Glu Val Gln Thr Leu Leu Ser Val
645 650 655

Glu Val Ser Asp Val Asp Ala Leu Ser Arg Thr Lys Thr Val Gly Asp
660 665 670

Val Ile Glu Ala Met Lys Met Glu Leu Gly Gly Pro Gln Gly Gln Thr

675

680

685

Leu Thr Ala Glu Ser Ile Arg Gln Pro Pro Val Ser Glu Pro Ala Val
690 695 700

Pro Thr Ser Ser Ser Ser Ser Ile Ala Asn Val Ser Ser Ala Arg Leu
705 710 715 720

Ala Glu Ala Glu Ala Ala Val Leu Ser Val Leu Ala Asp Lys Thr Gly
725 730 735

Tyr Asp Ser Ser Met Ile Glu Met Asp Met Asp Leu Glu Ser Glu Leu
740 745 750

Gly Val Asp Ser Ile Lys Arg Val Glu Ile Met Ser Glu Val Gln Ala
755 760 765

Leu Leu Ser Val Glu Val Ser Asp Val Asp Ala Leu Ser Arg Thr Lys
770 775 780

Thr Val Gly Asp Val Ile Glu Ala Met Lys Met Glu Leu Gly Gly Pro
785 790 795 800

Gln Gly Gln Thr Leu Thr Ala Glu Ser Ile Arg Glu Pro Pro Val Ser
805 810 815

Glu Pro Ala Val Pro Thr Ser Ser Ser Ser Ser Ile Ala Asn Val Ser
820 825 830

Ser Ala Arg Leu Ala Glu Ala Glu Ala Ala Val Leu Ser Val Leu Ala
835 840 845

Asp Lys Thr Gly Tyr Asp Ser Ser Met Ile Glu Met Asp Met Asp Leu
850 855 860

Glu Ser Glu Leu Gly Val Asp Ser Ile Lys Arg Val Glu Ile Met Ser

865 870 875 880

Glu Val Gln Thr Leu Leu Ser Val Glu Val Ser Asp Val Asp Ala Leu
885 890 895

Ser Arg Thr Lys Thr Val Gly Asp Val Ile Glu Ala Met Lys Leu Glu
900 905 910

Leu Gly Glu Ser Ser Ser Ile Glu Thr Leu Asn Cys Thr Glu Val Glu
915 920 925

His Thr
930

<210>	49
<211>	2433
<212>	DNA
<213>	Thraustochytrium sp.

<220>
<221> CDS
<222> (1)..(2433)

<400> 49
aaa agt gtc aag gct tca ggg tgt gag aat gta gat acc cgt ttc gct 48
Lys Ser Val Lys Ala Ser Gly Cys Glu Asn Val Asp Thr Arg Phe Ala
1 5 10 15

aag gtt gta caa atc tcg ctt cct agc aag ctg aaa tcc act gtg tcg 96
Lys Val Val Gln Ile Ser Leu Pro Ser Lys Leu Lys Ser Thr Val Ser
20 25 30

cac gat cga cct gta att gtt gta gat gat gga acg ccc tta acc acg 144
 His Asp Arg Pro Val Ile Val Val Asp Asp Gly Thr Pro Leu Thr Thr
 35 40 45

gag ctt tgt aaa att ctt ggg ggt aat att gtg gtt ctc tct tat caa 192
Glu Leu Cys Lys Ile Leu Gly Gly Asn Ile Val Val Leu Ser Tyr Gln
50 55 60

ggg aag ccc gct ggt cca cgg gga gtc gag gtg cca gat ctt tcc gag 240
Gly Lys Pro Ala Gly Pro Arg Gly Val Glu Val Pro Asp Leu Ser Glu

65	70	75	80	
gaa gcc cta att caa gct ctt gca ttg att cgg tct aca tat gga gtt				288
Glu Ala Leu Ile Gln Ala Leu Ala Leu Ile Arg Ser Thr Tyr Gly Val				
	85	90	95	
cca att ggt ttt att tgt cag caa gtg tct aat gtg agc acc aag gca				336
Pro Ile Gly Phe Ile Cys Gln Gln Val Ser Asn Val Ser Thr Lys Ala				
	100	105	110	
cag ctt tgt tgg gca ctc ctc gca gcg aag cat ctc aag aag gat ttg				384
Gln Leu Cys Trp Ala Leu Leu Ala Ala Lys His Leu Lys Lys Asp Leu				
	115	120	125	
aat gct gtc tta ccc gat tca aga tcc ttc ttc gtc gga gtt gta cgc				432
Asn Ala Val Leu Pro Asp Ser Arg Ser Phe Phe Val Gly Val Val Arg				
	130	135	140	
ttg aac ggg aaa ctt gga act ttc gaa aac atc agc gac ttc tct aaa				480
Leu Asn Gly Lys Leu Gly Thr Phe Glu Asn Ile Ser Asp Phe Ser Lys				
	145	150	155	160
ttt gat ttg acg aaa gcc cta gat tac gga cag cgt ggt tct ctc tta				528
Phe Asp Leu Thr Lys Ala Leu Asp Tyr Gly Gln Arg Gly Ser Leu Leu				
	165	170	175	
ggc ctg tgc aag tca cta gac tta gaa tgg gaa cag gtg ttt tgc cgt				576
Gly Leu Cys Lys Ser Leu Asp Leu Glu Trp Glu Gln Val Phe Cys Arg				
	180	185	190	
gga ata gat ctt gcg tgt gat ctt atg cca ctc cag gcc gca agg ata				624
Gly Ile Asp Leu Ala Cys Asp Leu Met Pro Leu Gln Ala Ala Arg Ile				
	195	200	205	
ctc aga aat gag ctt cag tgt ccc aat atg cgc ctt cgc gag gtt ggg				672
Leu Arg Asn Glu Leu Gln Cys Pro Asn Met Arg Leu Arg Glu Val Gly				
	210	215	220	
tac gat att tct ggc gcc agg tac acc att tca acc gat gac ctg cta				720
Tyr Asp Ile Ser Gly Ala Arg Tyr Thr Ile Ser Thr Asp Asp Leu Leu				
	225	230	235	240
tgt gga ccc tcg aag gct aaa gta gag gcc gca gac ttg ttt ctt gtg				768
Cys Gly Pro Ser Lys Ala Lys Val Glu Ala Ala Asp Leu Phe Leu Val				

245	250	255	
aca ggt ggc gca cga ggt att aca cct cat tgt gtt cgt gag att gca Thr Gly Gly Ala Arg Gly Ile Thr Pro His Cys Val Arg Glu Ile Ala 260 265 270			816
agt cga tcc ccc gga acc aca ttt gtg ctg gtt gga aga agc gaa atg Ser Arg Ser Pro Gly Thr Thr Phe Val Leu Val Gly Arg Ser Glu Met 275 280 285			864
tcc gac gag cct gac tgg gct gtt ggc cac tac aat aaa gac ctg gac Ser Asp Glu Pro Asp Trp Ala Val Gly His Tyr Asn Lys Asp Leu Asp 290 295 300			912
caa agc aca atg aaa cac ttg aaa gca acg cat gct gct gga ggg gta Gln Ser Thr Met Lys His Leu Lys Ala Thr His Ala Ala Gly Gly Val 305 310 315 320			960
aaa cct acg cct aaa gca cat cgt gca ctt gtg aac agg gtc act ggc Lys Pro Thr Pro Lys Ala His Arg Ala Leu Val Asn Arg Val Thr Gly 325 330 335			1008
tca cgg gag gta cga gaa tct ctt aga gca atc cag gag gca ggg gca Ser Arg Glu Val Arg Glu Ser Leu Arg Ala Ile Gln Glu Ala Gly Ala 340 345 350			1056
aat gtc gaa tat atc gcc tgt gat gtt tcg gat gaa aac aag gtc cgc Asn Val Glu Tyr Ile Ala Cys Asp Val Ser Asp Glu Asn Lys Val Arg 355 360 365			1104
caa ctt gtg caa aga gtg gag caa aag tat ggc tgt gaa ata act ggg Gln Leu Val Gln Arg Val Glu Gln Lys Tyr Gly Cys Glu Ile Thr Gly 370 375 380			1152
att tgg cat gca agc ggg gtt ctt cgt gac aaa ctt gtc gag caa aag Ile Trp His Ala Ser Gly Val Leu Arg Asp Lys Leu Val Glu Gln Lys 385 390 395 400			1200
act aca gac gac ttt gag gca gtt ttt ggg acc aag gtg act ggc ctt Thr Thr Asp Asp Phe Glu Ala Val Phe Gly Thr Lys Val Thr Gly Leu 405 410 415			1248
gta aac atc gtg tca caa gtc aat atg tct aag cta cga cac ttc atc Val Asn Ile Val Ser Gln Val Asn Met Ser Lys Leu Arg His Phe Ile			1296

420	425	430	
ctc ttc agt tct ttg gct gga ttt cat ggg aac aag ggc caa acg gat			1344
Leu Phe Ser Ser Leu Ala Gly Phe His Gly Asn Lys Gly Gln Thr Asp			
435	440	445	
tat gca att gct aat gaa gcc ttg aac aaa atc gcg cat act ctc tca			1392
Tyr Ala Ile Ala Asn Glu Ala Leu Asn Lys Ile Ala His Thr Leu Ser			
450	455	460	
gcg ttt ttg ccc aaa ctg aat gca aag gtg cta gac ttc ggt ccg tgg			1440
Ala Phe Leu Pro Lys Leu Asn Ala Lys Val Leu Asp Phe Gly Pro Trp			
465	470	475	480
gta ggt tca gga atg gta acc gaa aca ctt gag aag cat ttt aaa gct			1488
Val Gly Ser Gly Met Val Thr Glu Thr Leu Glu Lys His Phe Lys Ala			
485	490	495	
atg ggg gtt cag act att cct ctc gag cca gga gca cgg act gtt gcg			1536
Met Gly Val Gln Thr Ile Pro Leu Glu Pro Gly Ala Arg Thr Val Ala			
500	505	510	
caa atc att ttg gca agt tcg cca ccg caa tcg ctt ttg ggg aac tgg			1584
Gln Ile Ile Leu Ala Ser Ser Pro Pro Gln Ser Leu Leu Gly Asn Trp			
515	520	525	
ggc ttt cca gcc acc aaa ccg cta caa cgc tct aat gta gtc acg ggc			1632
Gly Phe Pro Ala Thr Lys Pro Leu Gln Arg Ser Asn Val Val Thr Gly			
530	535	540	
aca ctc tct ccg gaa gag ata gaa ttc atc gca gac cac aaa att caa			1680
Thr Leu Ser Pro Glu Glu Ile Glu Phe Ile Ala Asp His Lys Ile Gln			
545	550	555	560
ggc cgc aag gtg ctt ccc atg atg gct gca atc ggg ttc atg gcc tct			1728
Gly Arg Lys Val Leu Pro Met Met Ala Ala Ile Gly Phe Met Ala Ser			
565	570	575	
att gcg gaa gga ctc tac ccg ggg tac aat ctg caa ggc gtg gaa aat			1776
Ile Ala Glu Gly Leu Tyr Pro Gly Tyr Asn Leu Gln Gly Val Glu Asn			
580	585	590	
gct cag ctc ttt caa ggc ttg act atc aac caa gag aca aaa ttt caa			1824
Ala Gln Leu Phe Gln Gly Leu Thr Ile Asn Gln Glu Thr Lys Phe Gln			

595	600	605	
atc act ctc att gag gag cac aac tct gag gaa aac ctg gat gtc ctg			1872
Ile Thr Leu Ile Glu Glu His Asn Ser Glu Glu Asn Leu Asp Val Leu			
610	615	620	
aca tcc ctt ggt gta atg ttg gaa agc ggg aag gtg ctt ccc gct tac			1920
Thr Ser Leu Gly Val Met Leu Glu Ser Gly Lys Val Leu Pro Ala Tyr			
625	630	635	640
cga tgt gtt gta tgc ttg aat aca acc cag cag cag ccc aag cta tct			1968
Arg Cys Val Val Cys Leu Asn Thr Thr Gln Gln Gln Pro Lys Leu Ser			
645	650	655	
cca aaa att ctt aac ttg gaa gtt gac cct gca tgc gag gtt aac ccc			2016
Pro Lys Ile Leu Asn Leu Glu Val Asp Pro Ala Cys Glu Val Asn Pro			
660	665	670	
tat gat gga aag tcg ttg ttc cac ggt ccg ctt ttg caa ttc gtt caa			2064
Tyr Asp Gly Lys Ser Leu Phe His Gly Pro Leu Leu Gln Phe Val Gln			
675	680	685	
caa gtg ttg cac tca agt acc aaa ggc ctc gtt gcc aag tgc cgc gcg			2112
Gln Val Leu His Ser Ser Thr Lys Gly Leu Val Ala Lys Cys Arg Ala			
690	695	700	
ctt cca atc aaa gaa gcc atc cga ggg cca ttt atc aag caa aca ctc			2160
Leu Pro Ile Lys Glu Ala Ile Arg Gly Pro Phe Ile Lys Gln Thr Leu			
705	710	715	720
cat gat cca att cta gac gac gtc att ttt cag cta atg ctc gtg tgg			2208
His Asp Pro Ile Leu Asp Asp Val Ile Phe Gln Leu Met Leu Val Trp			
725	730	735	
tgt cgt aat gct cta gga agt gca tcg cta ccc aac aga att gaa aag			2256
Cys Arg Asn Ala Leu Gly Ser Ala Ser Leu Pro Asn Arg Ile Glu Lys			
740	745	750	
atg tca tac ttt ggg aat gtc tca gaa ggt agc act ttc ttt gcc tca			2304
Met Ser Tyr Phe Gly Asn Val Ser Glu Gly Ser Thr Phe Phe Ala Ser			
755	760	765	
gtt aca cct gtg gga cca aga gta cca aag gat ccc gtg atc aaa atg			2352
Val Thr Pro Val Gly Pro Arg Val Pro Lys Asp Pro Val Ile Lys Met			

770

775

780

cag ttt ctt ctc caa gat gaa tcc ggc aac aca ttt tca tcg ggg gag 2400
Gln Phe Leu Leu Gln Asp Glu Ser Gly Asn Thr Phe Ser Ser Gly Glu
785 790 795 800

ggc tcg gtt gtg ctt agt gac gaa ctc gtc ttt 2433
Gly Ser Val Val Leu Ser Asp Glu Leu Val Phe
805 810

<210> 50
<211> 811
<212> PRT
<213> Thraustochytrium sp.

<400> 50

Lys Ser Val Lys Ala Ser Gly Cys Glu Asn Val Asp Thr Arg Phe Ala
1 5 10 15

Lys Val Val Gln Ile Ser Leu Pro Ser Lys Leu Lys Ser Thr Val Ser
20 25 30

His Asp Arg Pro Val Ile Val Val Asp Asp Gly Thr Pro Leu Thr Thr
35 40 45

Glu Leu Cys Lys Ile Leu Gly Gly Asn Ile Val Val Leu Ser Tyr Gln
50 55 60

Gly Lys Pro Ala Gly Pro Arg Gly Val Glu Val Pro Asp Leu Ser Glu
65 70 75 80

Glu Ala Leu Ile Gln Ala Leu Ala Leu Ile Arg Ser Thr Tyr Gly Val
85 90 95

Pro Ile Gly Phe Ile Cys Gln Gln Val Ser Asn Val Ser Thr Lys Ala
100 105 110

Gln Leu Cys Trp Ala Leu Leu Ala Ala Lys His Leu Lys Lys Asp Leu
115 120 125

Asn Ala Val Leu Pro Asp Ser Arg Ser Phe Phe Val Gly Val Val Arg
130 135 140

Leu Asn Gly Lys Leu Gly Thr Phe Glu Asn Ile Ser Asp Phe Ser Lys
145 150 155 160

Phe Asp Leu Thr Lys Ala Leu Asp Tyr Gly Gln Arg Gly Ser Leu Leu
165 170 175

Gly Leu Cys Lys Ser Leu Asp Leu Glu Trp Glu Gln Val Phe Cys Arg
180 185 190

Gly Ile Asp Leu Ala Cys Asp Leu Met Pro Leu Gln Ala Ala Arg Ile
195 200 205

Leu Arg Asn Glu Leu Gln Cys Pro Asn Met Arg Leu Arg Glu Val Gly
210 215 220

Tyr Asp Ile Ser Gly Ala Arg Tyr Thr Ile Ser Thr Asp Asp Leu Leu
225 230 235 240

Cys Gly Pro Ser Lys Ala Lys Val Glu Ala Ala Asp Leu Phe Leu Val
245 250 255

Thr Gly Gly Ala Arg Gly Ile Thr Pro His Cys Val Arg Glu Ile Ala
260 265 270

Ser Arg Ser Pro Gly Thr Thr Phe Val Leu Val Gly Arg Ser Glu Met
275 280 285

Ser Asp Glu Pro Asp Trp Ala Val Gly His Tyr Asn Lys Asp Leu Asp
290 295 300

Gln Ser Thr Met Lys His Leu Lys Ala Thr His Ala Ala Gly Gly Val

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

Gln Ile Ile Leu Ala Ser Ser Pro Pro Gln Ser Leu Leu Gly Asn Trp
515 520 525

Gly Phe Pro Ala Thr Lys Pro Leu Gln Arg Ser Asn Val Val Thr Gly
530 535 540

Thr Leu Ser Pro Glu Glu Ile Glu Phe Ile Ala Asp His Lys Ile Gln
545 550 555 560

Gly Arg Lys Val Leu Pro Met Met Ala Ala Ile Gly Phe Met Ala Ser
565 570 575

Ile Ala Glu Gly Leu Tyr Pro Gly Tyr Asn Leu Gln Gly Val Glu Asn
580 585 590

Ala Gln Leu Phe Gln Gly Leu Thr Ile Asn Gln Glu Thr Lys Phe Gln
595 600 605

Ile Thr Leu Ile Glu Glu His Asn Ser Glu Glu Asn Leu Asp Val Leu
610 615 620

Thr Ser Leu Gly Val Met Leu Glu Ser Gly Lys Val Leu Pro Ala Tyr
625 630 635 640

Arg Cys Val Val Cys Leu Asn Thr Thr Gln Gln Gln Pro Lys Leu Ser
645 650 655

Pro Lys Ile Leu Asn Leu Glu Val Asp Pro Ala Cys Glu Val Asn Pro
660 665 670

Tyr Asp Gly Lys Ser Leu Phe His Gly Pro Leu Leu Gln Phe Val Gln
675 680 685

Gln Val Leu His Ser Ser Thr Lys Gly Leu Val Ala Lys Cys Arg Ala
690 695 700

Leu Pro Ile Lys Glu Ala Ile Arg Gly Pro Phe Ile Lys Gln Thr Leu
705 710 715 720

His Asp Pro Ile Leu Asp Asp Val Ile Phe Gln Leu Met Leu Val Trp
725 730 735

Cys Arg Asn Ala Leu Gly Ser Ala Ser Leu Pro Asn Arg Ile Glu Lys
740 745 750

Met Ser Tyr Phe Gly Asn Val Ser Glu Gly Ser Thr Phe Phe Ala Ser
755 760 765

Val Thr Pro Val Gly Pro Arg Val Pro Lys Asp Pro Val Ile Lys Met
770 775 780

Gln Phe Leu Leu Gln Asp Glu Ser Gly Asn Thr Phe Ser Ser Gly Glu
785 790 795 800

Gly Ser Val Val Leu Ser Asp Glu Leu Val Phe
805 810

<210> 51
<211> 5808
<212> DNA
<213> Thraustochytrium sp.

<220>
<221> CDS
<222> (1)..(5805)

<220>
<221> misc_feature
<222> (1)..(5808)
<223> n = a c t o r g

<400> 51
atg caa ctt cct cca gcg cat tct gcc gat gag aat cgc atc gcg gtc
Met Gln Leu Pro Pro Ala His Ser Ala Asp Glu Asn Arg Ile Ala Val

48

1	5	10	15	
gtg ggc atg gcc gtc aaa tat gcg ggc tgt gac aat aaa gaa gag ttt				96
Val Gly Met Ala Val Lys Tyr Ala Gly Cys Asp Asn Lys Glu Glu Phe				
	20	25	30	
tgg aag act ttg atg aat ggt agt atc aat acc aag tcg att tcg gca				144
Trp Lys Thr Leu Met Asn Gly Ser Ile Asn Thr Lys Ser Ile Ser Ala				
	35	40	45	
gca agg ttg ggc agc aat aag cgt gac gaa cac tat gtt cct gaa cga				192
Ala Arg Leu Gly Ser Asn Lys Arg Asp Glu His Tyr Val Pro Glu Arg				
	50	55	60	
tcg aaa tat gca gat acg ttc tgt aac gaa agg tac ggt tgt atc cag				240
Ser Lys Tyr Ala Asp Thr Phe Cys Asn Glu Arg Tyr Gly Cys Ile Gln				
65	70	75	80	
caa ggt acg gat aat gag cat gac ctc ctc cta ggt ctt gct caa gaa				288
Gln Gly Thr Asp Asn Glu His Asp Leu Leu Leu Gly Leu Ala Gln Glu				
	85	90	95	
gct ctc gct gac gct gcc ggg cgg atg gag aaa caa cct tcg gag gcg				336
Ala Leu Ala Asp Ala Ala Gly Arg Met Glu Lys Gln Pro Ser Glu Ala				
	100	105	110	
ttc gat ctg gaa aat act ggc atc gtg agt ggg tgc tta tct ttt cca				384
Phe Asp Leu Glu Asn Thr Gly Ile Val Ser Gly Cys Leu Ser Phe Pro				
	115	120	125	
atg gat aac ctg caa gga gag ttg ttg aac ttg tat caa agc cat gtg				432
Met Asp Asn Leu Gln Gly Glu Leu Leu Asn Leu Tyr Gln Ser His Val				
	130	135	140	
gag aaa caa ctt cca cct agt gcc ttg gta gaa gcc gtg aag ctt tgg				480
Glu Lys Gln Leu Pro Ser Ala Leu Val Glu Ala Val Lys Leu Trp				
	145	150	155	160
tct gag cga cag aaa tct acg aaa gca cat gca ggg gac aag cgc cgg				528
Ser Glu Arg Gln Lys Ser Thr Lys Ala His Ala Gly Asp Lys Arg Arg				
	165	170	175	
ttc att gac cca gct tct ttt gta gct gat aaa ctg aac cta ggc cca				576
Phe Ile Asp Pro Ala Ser Phe Val Ala Asp Lys Leu Asn Leu Gly Pro				

180	185	190	
cta cat tat gcg atc gat gca gca tgc gct tct gca ttg tac gtg tta			624
Leu His Tyr Ala Ile Asp Ala Ala Cys Ala Ser Ala Leu Tyr Val Leu			
195	200	205	
aaa tta gct caa gac cac ctt gtt tca ggt gcc gtt gat atg atg tta			672
Lys Leu Ala Gln Asp His Leu Val Ser Gly Ala Val Asp Met Met Leu			
210	215	220	
tgt gga gcg acg tgc ttc cca gaa cca ttc ttc atc ttg tct ggg ttc			720
Cys Gly Ala Thr Cys Phe Pro Glu Pro Phe Phe Ile Leu Ser Gly Phe			
225	230	235	240
tcg act ttt caa gcg atg cct gnt ggg gca gat gga gtc tca cta cct			768
Ser Thr Phe Gln Ala Met Pro Xaa Gly Ala Asp Gly Val Ser Leu Pro			
245	250	255	
ctc cat aaa acg agt gct ggg ctc act cca ggt gaa ggg ggg tcc att			816
Leu His Lys Thr Ser Ala Gly Leu Thr Pro Gly Glu Gly Gly Ser Ile			
260	265	270	
atg gtg ctc aag cga ctg aaa gac gct atc aga gat gga aat cac att			864
Met Val Leu Lys Arg Leu Lys Asp Ala Ile Arg Asp Gly Asn His Ile			
275	280	285	
tat ggt gtg ctc ctt gaa gca aat tta agt aac gca ggt tgt ggg ctt			912
Tyr Gly Val Leu Leu Glu Ala Asn Leu Ser Asn Ala Gly Cys Gly Leu			
290	295	300	
cca ctc agc ccg cac tta ccg agc gaa gaa tca tgt att cgt gat acc			960
Pro Leu Ser Pro His Leu Pro Ser Glu Glu Ser Cys Ile Arg Asp Thr			
305	310	315	320
tac cgc cgt gct gga gtt gct gca gat caa agt att cag tat att gag			1008
Tyr Arg Arg Ala Gly Val Ala Ala Asp Gln Ser Ile Gln Tyr Ile Glu			
325	330	335	
tgc cac gct acg gga acc cct cga ggg gat gtc gtg gaa att gag gcg			1056
Cys His Ala Thr Gly Thr Pro Arg Gly Asp Val Val Glu Ile Glu Ala			
340	345	350	
gtt gaa aga gtt ttc aag aaa aac gtt cca cgc tta ggc tcg acg aaa			1104
Val Glu Arg Val Phe Lys Lys Asn Val Pro Arg Leu Gly Ser Thr Lys			

355	360	365	
gga aat ttt ggt cac tcg tta gtt gcg gct ggt ttc gca ggt atg gca Gly Asn Phe Gly His Ser Leu Val Ala Ala Gly Phe Ala Gly Met Ala 370 375 380			1152
aag ctt ctt ctt gca atg gaa cat gga gtg att cct ccc aca cca ggt Lys Leu Leu Leu Ala Met Glu His Gly Val Ile Pro Pro Thr Pro Gly 385 390 395 400			1200
ctt gat gct tcg aac cag gca agt gag cac gtt gtg aca aag gct atc Leu Asp Ala Ser Asn Gln Ala Ser Glu His Val Val Thr Lys Ala Ile 405 410 415			1248
act tgg cct gag aca cat ggg gct cca aaa cga gct ggc ctt tca gca Thr Trp Pro Glu Thr His Gly Ala Pro Lys Arg Ala Gly Leu Ser Ala 420 425 430			1296
ttt gga ttt ggt ggg act aat gcg cat gca ctc ttc gaa gag ttt aat Phe Gly Phe Gly Gly Thr Asn Ala His Ala Leu Phe Glu Glu Phe Asn 435 440 445			1344
gcc gag ggc ata agt tat cgc cct gga aag cct cca gtc gaa tcg aat Ala Glu Gly Ile Ser Tyr Arg Pro Gly Lys Pro Pro Val Glu Ser Asn 450 455 460			1392
acc cgt cct tcc gtc gta ata act ggg atg gac tgt acc ttt ggg agc Thr Arg Pro Ser Val Val Ile Thr Gly Met Asp Cys Thr Phe Gly Ser 465 470 475 480			1440
ctt gaa ggg att gat gcg ttc gag act gcc ctg tac gag ggg cgt gac Leu Glu Gly Ile Asp Ala Phe Glu Thr Ala Leu Tyr Glu Gly Arg Asp 485 490 495			1488
gca gct cgt gac tta ccc gcc aaa cgt tgg agg ttc cta ggt gag gac Ala Ala Arg Asp Leu Pro Ala Lys Arg Trp Arg Phe Leu Gly Glu Asp 500 505 510			1536
ttg gag ttt ctc cga gcc atc agg ctc aag gaa aag cct agg ggt tgt Leu Glu Phe Leu Arg Ala Ile Arg Leu Lys Glu Lys Pro Arg Gly Cys 515 520 525			1584
ttt gtg gag agt gtt gac gtt aac ttt aga cgg ctg aaa acg ccc ttg Phe Val Glu Ser Val Asp Val Asn Phe Arg Arg Leu Lys Thr Pro Leu			1632

530	535	540	
aca cca gaa gat atg ttg cgg ccc caa caa ctc ttg gcg gtt tct acg			1680
Thr Pro Glu Asp Met Leu Arg Pro Gln Gln Leu Ala Val Ser Thr			
545	550	555	560
atg gac cga gca att atc gat gca ggt cta aag aag ggc caa cat gta			1728
Met Asp Arg Ala Ile Ile Asp Ala Gly Leu Lys Lys Gly Gln His Val			
565	570	575	
gca gtt ctt gtt ggc cta gga act gac ctg gaa ctt tac cgt cat cga			1776
Ala Val Leu Val Gly Leu Gly Thr Asp Leu Glu Leu Tyr Arg His Arg			
580	585	590	
gca aga gtc gcg ctt aaa gag gtt ttg cac ccg agc tta aag tca gac			1824
Ala Arg Val Ala Leu Lys Glu Val Leu His Pro Ser Leu Lys Ser Asp			
595	600	605	
act gca att ctc cag aaa ata atg caa tat gtg aat gat gca gga act			1872
Thr Ala Ile Leu Gln Lys Ile Met Gln Tyr Val Asn Asp Ala Gly Thr			
610	615	620	
tcg act tca tac aca tct tac att gga aac ctc gtt gcc acg cgt att			1920
Ser Thr Ser Tyr Thr Ser Tyr Ile Gly Asn Leu Val Ala Thr Arg Ile			
625	630	635	640
tcg tct cag tgg gga ttc aca ggg ccg tcc ttt act gtc aca gaa gga			1968
Ser Ser Gln Trp Gly Phe Thr Gly Pro Ser Phe Thr Val Thr Glu Gly			
645	650	655	
aat aat tcc gtg tac aga tgt gca caa cta gcc aaa gat atg ctt cag			2016
Asn Asn Ser Val Tyr Arg Cys Ala Gln Leu Ala Lys Asp Met Leu Gln			
660	665	670	
gtt aac cga gtt gat gct gtc gtc atc gca ggc gtt gat ctc aac gga			2064
Val Asn Arg Val Asp Ala Val Val Ile Ala Gly Val Asp Leu Asn Gly			
675	680	685	
agc gcc gaa agt ttt ttt gtc cga gca aat cgt caa aag ata tcc aag			2112
Ser Ala Glu Ser Phe Phe Val Arg Ala Asn Arg Gln Lys Ile Ser Lys			
690	695	700	
cta agt cat cca tgt gca agc ttc gac aga gat gca gat gga ttt ttc			2160
Leu Ser His Pro Cys Ala Ser Phe Asp Arg Asp Ala Asp Gly Phe Phe			

705	710	715	720
gca ggt gag ggc tgt ggt gcc cta gtt ttc aag agg tta gaa gac tgt			
Ala Gly Glu Gly Cys Gly Ala Leu Val Phe Lys Arg Leu Glu Asp Cys			
	725	730	735
gct cct cag gaa aaa att tat gct agt ata gac tct atc gca ata gat			
Ala Pro Gln Glu Lys Ile Tyr Ala Ser Ile Asp Ser Ile Ala Ile Asp			
	740	745	750
aaa gag cct act agc tca gct gtg aaa gct gtc tac caa agt gat tcg			
Lys Glu Pro Thr Ser Ser Ala Val Lys Ala Val Tyr Gln Ser Asp Ser			
	755	760	765
agt ctc tcc gat att gag ctg tta gaa atc agt gga gac tcc aaa cgg			
Ser Leu Ser Asp Ile Glu Leu Leu Glu Ile Ser Gly Asp Ser Lys Arg			
	770	775	780
ttt gca gca ttc gaa ggc gct gtg gaa att caa tca agt gtg gaa gcc			
Phe Ala Ala Phe Glu Gly Ala Val Glu Ile Gln Ser Ser Val Glu Ala			
	785	790	800
cag cta aaa gga ctt tcc aaa gtc ctt gaa cct gca aaa ggc caa ggc			
Gln Leu Lys Gly Leu Ser Lys Val Leu Glu Pro Ala Lys Gly Gln Gly			
	805	810	815
gta gcg gtg gga agt act cga gca acc gtt ggg gat ata ggg tat gct			
Val Ala Val Gly Ser Thr Arg Ala Thr Val Gly Asp Ile Gly Tyr Ala			
	820	825	830
aca gga gcg gca agc ctg att aaa act gca ctc tgc tta tat aat cgc			
Thr Gly Ala Ala Ser Leu Ile Lys Thr Ala Leu Cys Leu Tyr Asn Arg			
	835	840	845
tac ctt ccg gca tta gca aac tgg agt ggc cca tgt gaa cag tcc gcc			
Tyr Leu Pro Ala Leu Ala Asn Trp Ser Gly Pro Cys Glu Gln Ser Ala			
	850	855	860
tgg ggc tca aac atg ttc gtt tgc cat gaa aca cgg ccg tgg atg aaa			
Trp Gly Ser Asn Met Phe Val Cys His Glu Thr Arg Pro Trp Met Lys			
	865	870	875
aac cag aat gaa aag aga tgt gcc ctc att tct gga aca gat cca tct			
Asn Gln Asn Glu Lys Arg Cys Ala Leu Ile Ser Gly Thr Asp Pro Ser			

2208

2256

2304

2352

2400

2448

2496

2544

2592

2640

2688

885	890	895	
cat aca tgc ttt tcc ctc gta cta tcg gat act ggg tgt tat gaa gag			2736
His Thr Cys Phe Ser Leu Val Leu Ser Asp Thr Gly Cys Tyr Glu Glu			
900	905	910	
cac aat cga acg tgc ttt gat gtg caa gcg cca cag cta gtt ctg ata			2784
His Asn Arg Thr Cys Phe Asp Val Gln Ala Pro Gln Leu Val Leu Ile			
915	920	925	
cac gga ttc gat gga aaa act att gtg cgg cga ctt gaa gga tat ctc			2832
His Gly Phe Asp Gly Lys Thr Ile Val Arg Arg Leu Glu Gly Tyr Leu			
930	935	940	
ctt gaa ctt gtt gaa ggg cat gca agc cct tca gag tat ttc cac aaa			2880
Leu Glu Leu Val Glu Gly His Ala Ser Pro Ser Glu Tyr Phe His Lys			
945	950	955	960
ctg att gga caa agt cta ctt gag aac tcg aaa gaa agt aaa ctc aca			2928
Leu Ile Gly Gln Ser Leu Leu Glu Asn Ser Lys Glu Ser Lys Leu Thr			
965	970	975	
ctt tcg ctt gtg tgc aat ccg aac cag ctc caa aag gag ctc atg ctt			2976
Leu Ser Leu Val Cys Asn Pro Asn Gln Leu Gln Lys Glu Leu Met Leu			
980	985	990	
gct atc aaa gga gta caa cga agc atg tta aca ggg aag gat tgg gtc			3024
Ala Ile Lys Gly Val Gln Arg Ser Met Leu Thr Gly Lys Asp Trp Val			
995	1000	1005	
agt cca tca gga agt tgt ttt gcc cca aat ccg tta tca agc gca			3069
Ser Pro Ser Gly Ser Cys Phe Ala Pro Asn Pro Leu Ser Ser Ala			
1010	1015	1020	
aaa gtg gca ttc atg tac gga gaa ggc cga agc ccg tac tgt ggt			3114
Lys Val Ala Phe Met Tyr Gly Glu Gly Arg Ser Pro Tyr Cys Gly			
1025	1030	1035	
gta ggc ttg ggt cta cat cgt ttg tgg ccc ggt ctc cat gaa aat			3159
Val Gly Leu Gly Leu His Arg Leu Trp Pro Gly Leu His Glu Asn			
1040	1045	1050	
gtg aac aat aag aca gtc gat tta tgg acg gaa gga gat ggt tgg			3204
Val Asn Asn Lys Thr Val Asp Leu Trp Thr Glu Gly Asp Gly Trp			

1055	1060	1065	
tta tat cct cga acg ttg aca	cga gaa gag cat aca	aaa gcc atc	3249
Leu Tyr Pro Arg Thr Leu Thr	Arg Glu Glu His Thr	Lys Ala Ile	
1070	1075	1080	
gaa tct ttc aac gca aat caa	att gaa atg ttt cgc	gct ggg att	3294
Glu Ser Phe Asn Ala Asn Gln	Ile Glu Met Phe Arg	Ala Gly Ile	
1085	1090	1095	
ttc atc tca atg tgt cag aca	gac tat gtc atg aat	gtt ctc ggt	3339
Phe Ile Ser Met Cys Gln Thr	Asp Tyr Val Met Asn	Val Leu Gly	
1100	1105	1110	
gtc cag cct aag gcc gga ttt	ggg ctg agc ttg gga	gaa att tca	3384
Val Gln Pro Lys Ala Gly Phe	Gly Leu Ser Leu Gly	Glu Ile Ser	
1115	1120	1125	
atg ctc ttt gcg atg tca aag	gag aac tgc agg cag	tca cag gaa	3429
Met Leu Phe Ala Met Ser Lys	Glu Asn Cys Arg Gln	Ser Gln Glu	
1130	1135	1140	
atg acc aat cgt ttg cgc ggt	tct cca gtg tgg tct	aac gag ctt	3474
Met Thr Asn Arg Leu Arg Gly	Ser Pro Val Trp Ser	Asn Glu Leu	
1145	1150	1155	
gct atc aac ttc aat gca att	cgc aag tta tgg aaa	atc ccc cga	3519
Ala Ile Asn Phe Asn Ala Ile	Arg Lys Leu Trp Lys	Ile Pro Arg	
1160	1165	1170	
gga gct ccc tta gaa tcc ttt	tgg caa gga tac ttg	gtt cac ggc	3564
Gly Ala Pro Leu Glu Ser Phe	Trp Gln Gly Tyr Leu	Val His Gly	
1175	1180	1185	
aca aga gaa gaa gta gag cat	gct att ggt ctt tct	gag cct tat	3609
Thr Arg Glu Glu Val Glu His	Ala Ile Gly Leu Ser	Glu Pro Tyr	
1190	1195	1200	
gta cgt ctg ctt att gtg aac	gat tca agg agt gcc	ttg att gct	3654
Val Arg Leu Leu Ile Val Asn	Asp Ser Arg Ser Ala	Leu Ile Ala	
1205	1210	1215	
gga aaa cca gac gcc tgt cag	gca gta atc agt aga	cta aac tcc	3699
Gly Lys Pro Asp Ala Cys Gln	Ala Val Ile Ser Arg	Leu Asn Ser	

1220	1225	1230	
aag ttc cct tct ctg ccg gta	aag caa gga atg att	ggt cat tgc	3744
Lys Phe Pro Ser Leu Pro Val	Lys Gln Gly Met Ile	Gly His Cys	
1235	1240	1245	
cca gaa gtt cgt gcg ttc atc	aaa gat att ggg tac	atc cat gaa	3789
Pro Glu Val Arg Ala Phe Ile	Lys Asp Ile Gly Tyr	Ile His Glu	
1250	1255	1260	
aca ctc cga att tcc aat gac	tat tcg gat tgt cag	ctt ttc tca	3834
Thr Leu Arg Ile Ser Asn Asp	Tyr Ser Asp Cys Gln	Leu Phe Ser	
1265	1270	1275	
gcg gta acc aag ggc gca ctt	gac agc tcc aca atg	gaa atc aaa	3879
Ala Val Thr Lys Gly Ala Leu	Asp Ser Ser Thr Met	Glu Ile Lys	
1280	1285	1290	
cac ttt gtg gga gag gtc tac	tcc cgg atc gca gac	ttt cct caa	3924
His Phe Val Gly Glu Val Tyr	Ser Arg Ile Ala Asp	Phe Pro Gln	
1295	1300	1305	
atc gtc aac acg gtg cat tcg	gct ggt tat gac gta	ttt ctt gag	3969
Ile Val Asn Thr Val His Ser	Ala Gly Tyr Asp Val	Phe Leu Glu	
1310	1315	1320	
ctt ggc tgt gat gct tct aga	tct gca gca gtt caa	aac att ctt	4014
Leu Gly Cys Asp Ala Ser Arg	Ser Ala Ala Val Gln	Asn Ile Leu	
1325	1330	1335	
ggt ggt caa gga aag ttc ttg	tct aca gct att gac	aaa aaa gga	4059
Gly Gly Gln Gly Lys Phe Leu	Ser Thr Ala Ile Asp	Lys Lys Gly	
1340	1345	1350	
cac tcc gcc tgg tca caa gta	ctt cgg gct acc gca	tca tta gct	4104
His Ser Ala Trp Ser Gln Val	Leu Arg Ala Thr Ala	Ser Leu Ala	
1355	1360	1365	
gca cat cga gta ccg gga atc	tca att ttg gat ttg	ttt cac cca	4149
Ala His Arg Val Pro Gly Ile	Ser Ile Leu Asp Leu	Phe His Pro	
1370	1375	1380	
aat ttc cga gaa atg tgc tgt	aca atg gca acc aca	cct aaa gtg	4194
Asn Phe Arg Glu Met Cys Cys	Thr Met Ala Thr Thr	Pro Lys Val	

1385	1390	1395	
gaa gat aag ttc ctg cgc acg att caa atc aat ggt cgg ttt gaa			4239
Glu Asp Lys Phe Leu Arg Thr Ile Gln Ile Asn Gly Arg Phe Glu			
1400	1405	1410	
aaa gaa atg att cac cta gaa gat aca aca tta agt tgc tta ccc			4284
Lys Glu Met Ile His Leu Glu Asp Thr Thr Leu Ser Cys Leu Pro			
1415	1420	1425	
gct cca agt gaa gca aat atc gca gct att caa tct cgg tca att			4329
Ala Pro Ser Glu Ala Asn Ile Ala Ala Ile Gln Ser Arg Ser Ile			
1430	1435	1440	
cga tct gct gcg gcg cgt tct gga caa tcc cat gat tgt gca tcc			4374
Arg Ser Ala Ala Ala Arg Ser Gly Gln Ser His Asp Cys Ala Ser			
1445	1450	1455	
cat agc cat gaa gaa aat aag gat tca tgc cct gaa aag ctg aag			4419
His Ser His Glu Glu Asn Lys Asp Ser Cys Pro Glu Lys Leu Lys			
1460	1465	1470	
ctt gat tct gtg tcc gtc gcc ata aat ttc gac aat gat gac cgc			4464
Leu Asp Ser Val Ser Val Ala Ile Asn Phe Asp Asn Asp Asp Arg			
1475	1480	1485	
att cag ctt ggg cac gcg ggt ttt cgg gag atg tac aat aca aga			4509
Ile Gln Leu Gly His Ala Gly Phe Arg Glu Met Tyr Asn Thr Arg			
1490	1495	1500	
tat agc ttg tac aca ggg gcg atg gca aag gga att gca tct gca			4554
Tyr Ser Leu Tyr Thr Gly Ala Met Ala Lys Gly Ile Ala Ser Ala			
1505	1510	1515	
gat ctt gtc att gcc gct ggg aaa gag ggc atc cta gct tcc tat			4599
Asp Leu Val Ile Ala Ala Gly Lys Glu Gly Ile Leu Ala Ser Tyr			
1520	1525	1530	
gga gct gga gga cta cct ctt gct act gtt cga aag gga ata gac			4644
Gly Ala Gly Gly Leu Pro Leu Ala Thr Val Arg Lys Gly Ile Asp			
1535	1540	1545	
aaa att caa caa gcc ttg cca agt ggc cca tat gct gta aat ctt			4689
Lys Ile Gln Gln Ala Leu Pro Ser Gly Pro Tyr Ala Val Asn Leu			

1550	1555	1560	
att cac tct ccc ttt gac ggc aac ttg gag cag gga aac gtc gat			4734
Ile His Ser Pro Phe Asp Gly Asn Leu Glu Gln Gly Asn Val Asp			
1565	1570	1575	
ttg ttc ttg gaa aag aac gtc cgc gtg gcg gaa tgt tcc gcg ttt			4779
Leu Phe Leu Glu Lys Asn Val Arg Val Ala Glu Cys Ser Ala Phe			
1580	1585	1590	
aca acg cta aca gtg cca gta gta cac tat cgt gct gca ggg ctt			4824
Thr Thr Leu Thr Val Pro Val Val His Tyr Arg Ala Ala Gly Leu			
1595	1600	1605	
gtt cgg cgc caa gat gga agc att ttg atc aag aac cga atc att			4869
Val Arg Arg Gln Asp Gly Ser Ile Leu Ile Lys Asn Arg Ile Ile			
1610	1615	1620	
gct aaa gta tct agg aca gaa ctc gct gag atg ttc ctt cgt ccg			4914
Ala Lys Val Ser Arg Thr Glu Leu Ala Glu Met Phe Leu Arg Pro			
1625	1630	1635	
gca cct caa atc atc ctc gaa aaa ctg gta gca gca gaa atc att			4959
Ala Pro Gln Ile Ile Leu Glu Lys Leu Val Ala Ala Glu Ile Ile			
1640	1645	1650	
tca tct gac caa gcg cgt atg gca gcc aaa gtt ccc atg gcg gac			5004
Ser Ser Asp Gln Ala Arg Met Ala Ala Lys Val Pro Met Ala Asp			
1655	1660	1665	
gac atc gca gtc gaa gcc gac tct ggt ggg cac acg gat aat cgg			5049
Asp Ile Ala Val Glu Ala Asp Ser Gly Gly His Thr Asp Asn Arg			
1670	1675	1680	
cct atg cac gtc att ttg ccc ctg ata att caa ctc cgc aat act			5094
Pro Met His Val Ile Leu Pro Leu Ile Ile Gln Leu Arg Asn Thr			
1685	1690	1695	
ata ctt gca gag tat ggc tgt gcc acg gct ttt cgt acc cgt ata			5139
Ile Leu Ala Glu Tyr Gly Cys Ala Thr Ala Phe Arg Thr Arg Ile			
1700	1705	1710	
ggc gct gga gga ggc att ggt tgt cct tca gcg gcc ctc gca gcc			5184
Gly Ala Gly Gly Gly Ile Gly Cys Pro Ser Ala Ala Leu Ala Ala			

1715	1720	1725	
ttt gat atg ggt gcg agt ttt gtc gtg act gga agc ata aat caa			5229
Phe Asp Met Gly Ala Ser Phe Val Val Thr Gly Ser Ile Asn Gln			
1730	1735	1740	
att tgc cgc gag gca ggg act tgc gat act gtt cgg gag cta ctt			5274
Ile Cys Arg Glu Ala Gly Thr Cys Asp Thr Val Arg Glu Leu Leu			
1745	1750	1755	
gcc aac tca agc tac tcg gac gtg acg atg gcg cca gca gca gac			5319
Ala Asn Ser Ser Tyr Ser Asp Val Thr Met Ala Pro Ala Ala Asp			
1760	1765	1770	
atg ttt gac caa ggt gtg aaa ctc caa gtc tta aaa cga gga acg			5364
Met Phe Asp Gln Gly Val Lys Leu Gln Val Leu Lys Arg Gly Thr			
1775	1780	1785	
atg ttt cca agc aga gca aat aaa ctc cgg aag ctc ttt gtg aac			5409
Met Phe Pro Ser Arg Ala Asn Lys Leu Arg Lys Leu Phe Val Asn			
1790	1795	1800	
tac gaa tct cta gaa aca ctc ccg tcg aaa gag ttg aaa tac ctg			5454
Tyr Glu Ser Leu Glu Thr Leu Pro Ser Lys Glu Leu Lys Tyr Leu			
1805	1810	1815	
gaa aac atc ata ttc aag caa gca gta gac cag gtg tgg gag gaa			5499
Glu Asn Ile Ile Phe Lys Gln Ala Val Asp Gln Val Trp Glu Glu			
1820	1825	1830	
aca aag cgc ttt tac tgt gaa aaa ctg aac aat cca gat aaa att			5544
Thr Lys Arg Phe Tyr Cys Glu Lys Leu Asn Asn Pro Asp Lys Ile			
1835	1840	1845	
gca agg gcc atg aaa gat cct aaa ttg aag atg tcg ctt tgc ttt			5589
Ala Arg Ala Met Lys Asp Pro Lys Leu Lys Met Ser Leu Cys Phe			
1850	1855	1860	
cgg tgg tat ctc tcc aag agc tct ggg tgg gcc aac gca gga att			5634
Arg Trp Tyr Leu Ser Lys Ser Ser Gly Trp Ala Asn Ala Gly Ile			
1865	1870	1875	
aaa tct cgt gca ctc gac tac cag atc tgg tgt ggc ccg gca atg			5679
Lys Ser Arg Ala Leu Asp Tyr Gln Ile Trp Cys Gly Pro Ala Met			

1880 1885 1890

ggc tgc ttc aac aat ttc gcc agc ggc aca tcc ctc gat tgg aaa 5724
Gly Ser Phe Asn Asn Phe Ala Ser Gly Thr Ser Leu Asp Trp Lys
1895 1900 1905

gtg act ggg gtt ttc cct ggc gtt gcg gaa gta aac atg gcc att 5769
Val Thr Gly Val Phe Pro Gly Val Ala Glu Val Asn Met Ala Ile
1910 1915 1920

tta gat ggc gcg cga gaa cta gct gct aaa cga aat taa 5808
Leu Asp Gly Ala Arg Glu Leu Ala Ala Lys Arg Asn
1925 1930 1935

<210> 52
<211> 1935
<212> PRT
<213> Thraustochytrium sp.

<220>
<221> misc_feature
<222> (248)..(248)
<223> The 'Xaa' at location 248 stands for Asp, Gly, Ala, or Val.

<400> 52

Met Gln Leu Pro Pro Ala His Ser Ala Asp Glu Asn Arg Ile Ala Val
1 5 10 15

Val Gly Met Ala Val Lys Tyr Ala Gly Cys Asp Asn Lys Glu Glu Phe
20 25 30

Trp Lys Thr Leu Met Asn Gly Ser Ile Asn Thr Lys Ser Ile Ser Ala
35 40 45

Ala Arg Leu Gly Ser Asn Lys Arg Asp Glu His Tyr Val Pro Glu Arg
50 55 60

Ser Lys Tyr Ala Asp Thr Phe Cys Asn Glu Arg Tyr Gly Cys Ile Gln
65 70 75 80

Gln Gly Thr Asp Asn Glu His Asp Leu Leu Leu Gly Leu Ala Gln Glu
85 90 95

Ala Leu Ala Asp Ala Ala Gly Arg Met Glu Lys Gln Pro Ser Glu Ala
100 105 110

Phe Asp Leu Glu Asn Thr Gly Ile Val Ser Gly Cys Leu Ser Phe Pro
115 120 125

Met Asp Asn Leu Gln Gly Glu Leu Leu Asn Leu Tyr Gln Ser His Val
130 135 140

Glu Lys Gln Leu Pro Pro Ser Ala Leu Val Glu Ala Val Lys Leu Trp
145 150 155 160

Ser Glu Arg Gln Lys Ser Thr Lys Ala His Ala Gly Asp Lys Arg Arg
165 170 175

Phe Ile Asp Pro Ala Ser Phe Val Ala Asp Lys Leu Asn Leu Gly Pro
180 185 190

Leu His Tyr Ala Ile Asp Ala Ala Cys Ala Ser Ala Leu Tyr Val Leu
195 200 205

Lys Leu Ala Gln Asp His Leu Val Ser Gly Ala Val Asp Met Met Leu
210 215 220

Cys Gly Ala Thr Cys Phe Pro Glu Pro Phe Phe Ile Leu Ser Gly Phe
225 230 235 240

Ser Thr Phe Gln Ala Met Pro Xaa Gly Ala Asp Gly Val Ser Leu Pro
245 250 255

Leu His Lys Thr Ser Ala Gly Leu Thr Pro Gly Glu Gly Gly Ser Ile
260 265 270

Met Val Leu Lys Arg Leu Lys Asp Ala Ile Arg Asp Gly Asn His Ile
275 280 285

Tyr Gly Val Leu Leu Glu Ala Asn Leu Ser Asn Ala Gly Cys Gly Leu
290 295 300

Pro Leu Ser Pro His Leu Pro Ser Glu Glu Ser Cys Ile Arg Asp Thr
305 310 315 320

Tyr Arg Arg Ala Gly Val Ala Ala Asp Gln Ser Ile Gln Tyr Ile Glu
325 330 335

Cys His Ala Thr Gly Thr Pro Arg Gly Asp Val Val Glu Ile Glu Ala
340 345 350

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

Gly Asn Phe Gly His Ser Leu Val Ala Ala Gly Phe Ala Gly Met Ala
370 375 380

Lys Leu Leu Leu Ala Met Glu His Gly Val Ile Pro Pro Thr Pro Gly
385 390 395 400

Leu Asp Ala Ser Asn Gln Ala Ser Glu His Val Val Thr Lys Ala Ile
405 410 415

Thr Trp Pro Glu Thr His Gly Ala Pro Lys Arg Ala Gly Leu Ser Ala
420 425 430

Phe Gly Phe Gly Gly Thr Asn Ala His Ala Leu Phe Glu Glu Phe Asn
435 440 445

Ala Glu Gly Ile Ser Tyr Arg Pro Gly Lys Pro Pro Val Glu Ser Asn
450 455 460

Thr Arg Pro Ser Val Val Ile Thr Gly Met Asp Cys Thr Phe Gly Ser
465 470 475 480

Leu Glu Gly Ile Asp Ala Phe Glu Thr Ala Leu Tyr Glu Gly Arg Asp
485 490 495

Ala Ala Arg Asp Leu Pro Ala Lys Arg Trp Arg Phe Leu Gly Glu Asp
500 505 510

Leu Glu Phe Leu Arg Ala Ile Arg Leu Lys Glu Lys Pro Arg Gly Cys
515 520 525

Phe Val Glu Ser Val Asp Val Asn Phe Arg Arg Leu Lys Thr Pro Leu
530 535 540

Thr Pro Glu Asp Met Leu Arg Pro Gln Gln Leu Leu Ala Val Ser Thr
545 550 555 560

Met Asp Arg Ala Ile Ile Asp Ala Gly Leu Lys Lys Gly Gln His Val
565 570 575

Ala Val Leu Val Gly Leu Gly Thr Asp Leu Glu Leu Tyr Arg His Arg
580 585 590

Ala Arg Val Ala Leu Lys Glu Val Leu His Pro Ser Leu Lys Ser Asp
595 600 605

Thr Ala Ile Leu Gln Lys Ile Met Gln Tyr Val Asn Asp Ala Gly Thr
610 615 620

Ser Thr Ser Tyr Thr Ser Tyr Ile Gly Asn Leu Val Ala Thr Arg Ile
625 630 635 640

Ser Ser Gln Trp Gly Phe Thr Gly Pro Ser Phe Thr Val Thr Glu Gly
645 650 655

Asn Asn Ser Val Tyr Arg Cys Ala Gln Leu Ala Lys Asp Met Leu Gln
660 665 670

Val Asn Arg Val Asp Ala Val Val Ile Ala Gly Val Asp Leu Asn Gly
675 680 685

Ser Ala Glu Ser Phe Phe Val Arg Ala Asn Arg Gln Lys Ile Ser Lys
690 695 700

Leu Ser His Pro Cys Ala Ser Phe Asp Arg Asp Ala Asp Gly Phe Phe
705 710 715 720

Ala Gly Glu Gly Cys Gly Ala Leu Val Phe Lys Arg Leu Glu Asp Cys
725 730 735

Ala Pro Gln Glu Lys Ile Tyr Ala Ser Ile Asp Ser Ile Ala Ile Asp
740 745 750

Lys Glu Pro Thr Ser Ser Ala Val Lys Ala Val Tyr Gln Ser Asp Ser
755 760 765

Ser Leu Ser Asp Ile Glu Leu Leu Glu Ile Ser Gly Asp Ser Lys Arg
770 775 780

Phe Ala Ala Phe Glu Gly Ala Val Glu Ile Gln Ser Ser Val Glu Ala
785 790 795 800

Gln Leu Lys Gly Leu Ser Lys Val Leu Glu Pro Ala Lys Gly Gln Gly
805 810 815

Val Ala Val Gly Ser Thr Arg Ala Thr Val Gly Asp Ile Gly Tyr Ala
820 825 830

Thr Gly Ala Ala Ser Leu Ile Lys Thr Ala Leu Cys Leu Tyr Asn Arg
835 840 845

Tyr Leu Pro Ala Leu Ala Asn Trp Ser Gly Pro Cys Glu Gln Ser Ala
850 855 860

Trp Gly Ser Asn Met Phe Val Cys His Glu Thr Arg Pro Trp Met Lys
865 870 875 880

Asn Gln Asn Glu Lys Arg Cys Ala Leu Ile Ser Gly Thr Asp Pro Ser
885 890 895

His Thr Cys Phe Ser Leu Val Leu Ser Asp Thr Gly Cys Tyr Glu Glu
900 905 910

His Asn Arg Thr Cys Phe Asp Val Gln Ala Pro Gln Leu Val Leu Ile
915 920 925

His Gly Phe Asp Gly Lys Thr Ile Val Arg Arg Leu Glu Gly Tyr Leu
930 935 940

Leu Glu Leu Val Glu Gly His Ala Ser Pro Ser Glu Tyr Phe His Lys
945 950 955 960

Leu Ile Gly Gln Ser Leu Leu Glu Asn Ser Lys Glu Ser Lys Leu Thr
965 970 975

Leu Ser Leu Val Cys Asn Pro Asn Gln Leu Gln Lys Glu Leu Met Leu
980 985 990

Ala Ile Lys Gly Val Gln Arg Ser Met Leu Thr Gly Lys Asp Trp Val
995 1000 1005

Ser Pro Ser Gly Ser Cys Phe Ala Pro Asn Pro Leu Ser Ser Ala
1010 1015 1020

Lys Val Ala Phe Met Tyr Gly Glu Gly Arg Ser Pro Tyr Cys Gly
1025 1030 1035

Val Gly Leu Gly Leu His Arg Leu Trp Pro Gly Leu His Glu Asn
1040 1045 1050

Val Asn Asn Lys Thr Val Asp Leu Trp Thr Glu Gly Asp Gly Trp
1055 1060 1065

Leu Tyr Pro Arg Thr Leu Thr Arg Glu Glu His Thr Lys Ala Ile
1070 1075 1080

Glu Ser Phe Asn Ala Asn Gln Ile Glu Met Phe Arg Ala Gly Ile
1085 1090 1095

Phe Ile Ser Met Cys Gln Thr Asp Tyr Val Met Asn Val Leu Gly
1100 1105 1110

Val Gln Pro Lys Ala Gly Phe Gly Leu Ser Leu Gly Glu Ile Ser
1115 1120 1125

Met Leu Phe Ala Met Ser Lys Glu Asn Cys Arg Gln Ser Gln Glu
1130 1135 1140

Met Thr Asn Arg Leu Arg Gly Ser Pro Val Trp Ser Asn Glu Leu
1145 1150 1155

Ala Ile Asn Phe Asn Ala Ile Arg Lys Leu Trp Lys Ile Pro Arg
1160 1165 1170

Gly Ala Pro Leu Glu Ser Phe Trp Gln Gly Tyr Leu Val His Gly
1175 1180 1185

Thr Arg Glu Glu Val Glu His Ala Ile Gly Leu Ser Glu Pro Tyr
1190 1195 1200

Val Arg Leu Leu Ile Val Asn Asp Ser Arg Ser Ala Leu Ile Ala

1205

1210

1215

Gly Lys Pro Asp Ala Cys Gln Ala Val Ile Ser Arg Leu Asn Ser
1220 1225 1230

Lys Phe Pro Ser Leu Pro Val Lys Gln Gly Met Ile Gly His Cys
1235 1240 1245

Pro Glu Val Arg Ala Phe Ile Lys Asp Ile Gly Tyr Ile His Glu
1250 1255 1260

Thr Leu Arg Ile Ser Asn Asp Tyr Ser Asp Cys Gln Leu Phe Ser
1265 1270 1275

Ala Val Thr Lys Gly Ala Leu Asp Ser Ser Thr Met Glu Ile Lys
1280 1285 1290

His Phe Val Gly Glu Val Tyr Ser Arg Ile Ala Asp Phe Pro Gln
1295 1300 1305

Ile Val Asn Thr Val His Ser Ala Gly Tyr Asp Val Phe Leu Glu
1310 1315 1320

Leu Gly Cys Asp Ala Ser Arg Ser Ala Ala Val Gln Asn Ile Leu
1325 1330 1335

Gly Gly Gln Gly Lys Phe Leu Ser Thr Ala Ile Asp Lys Lys Gly
1340 1345 1350

His Ser Ala Trp Ser Gln Val Leu Arg Ala Thr Ala Ser Leu Ala
1355 1360 1365

Ala His Arg Val Pro Gly Ile Ser Ile Leu Asp Leu Phe His Pro
1370 1375 1380

Asn Phe Arg Glu Met Cys Cys Thr Met Ala Thr Thr Pro Lys Val
1385 1390 1395

Glu Asp Lys Phe Leu Arg Thr Ile Gln Ile Asn Gly Arg Phe Glu
1400 1405 1410

Lys Glu Met Ile His Leu Glu Asp Thr Thr Leu Ser Cys Leu Pro
1415 1420 1425

Ala Pro Ser Glu Ala Asn Ile Ala Ala Ile Gln Ser Arg Ser Ile
1430 1435 1440

Arg Ser Ala Ala Ala Arg Ser Gly Gln Ser His Asp Cys Ala Ser
1445 1450 1455

His Ser His Glu Glu Asn Lys Asp Ser Cys Pro Glu Lys Leu Lys
1460 1465 1470

Leu Asp Ser Val Ser Val Ala Ile Asn Phe Asp Asn Asp Asp Arg
1475 1480 1485

Ile Gln Leu Gly His Ala Gly Phe Arg Glu Met Tyr Asn Thr Arg
1490 1495 1500

Tyr Ser Leu Tyr Thr Gly Ala Met Ala Lys Gly Ile Ala Ser Ala
1505 1510 1515

Asp Leu Val Ile Ala Ala Gly Lys Glu Gly Ile Leu Ala Ser Tyr
1520 1525 1530

Gly Ala Gly Gly Leu Pro Leu Ala Thr Val Arg Lys Gly Ile Asp
1535 1540 1545

Lys Ile Gln Gln Ala Leu Pro Ser Gly Pro Tyr Ala Val Asn Leu
1550 1555 1560

Ile His Ser Pro Phe Asp Gly Asn Leu Glu Gln Gly Asn Val Asp
1565 1570 1575

Leu Phe Leu Glu Lys Asn Val Arg Val Ala Glu Cys Ser Ala Phe
1580 1585 1590

Thr Thr Leu Thr Val Pro Val Val His Tyr Arg Ala Ala Gly Leu
1595 1600 1605

Val Arg Arg Gln Asp Gly Ser Ile Leu Ile Lys Asn Arg Ile Ile
1610 1615 1620

Ala Lys Val Ser Arg Thr Glu Leu Ala Glu Met Phe Leu Arg Pro
1625 1630 1635

Ala Pro Gln Ile Ile Leu Glu Lys Leu Val Ala Ala Glu Ile Ile
1640 1645 1650

Ser Ser Asp Gln Ala Arg Met Ala Ala Lys Val Pro Met Ala Asp
1655 1660 1665

Asp Ile Ala Val Glu Ala Asp Ser Gly Gly His Thr Asp Asn Arg
1670 1675 1680

Pro Met His Val Ile Leu Pro Leu Ile Ile Gln Leu Arg Asn Thr
1685 1690 1695

Ile Leu Ala Glu Tyr Gly Cys Ala Thr Ala Phe Arg Thr Arg Ile
1700 1705 1710

Gly Ala Gly Gly Gly Ile Gly Cys Pro Ser Ala Ala Leu Ala Ala
1715 1720 1725

Phe Asp Met Gly Ala Ser Phe Val Val Thr Gly Ser Ile Asn Gln

Val Thr Gly Val Phe Pro Gly Val Ala Glu Val Asn Met Ala Ile
1910 1915 1920

Leu Asp Gly Ala Arg Glu Leu Ala Ala Lys Arg Asn
1925 1930 1935

<210> 53
<211> 1500
<212> DNA
<213> Thraustochytrium sp.

<220>
<221> CDS
<222> (1)..(1500)

<220>
<221> misc_feature
<222> (1)..(1500)
<223> n = a c t o r g

<400> 53
atg caa ctt cct cca gcg cat tct gcc gat gag aat cgc atc gcg gtc 48
Met Gln Leu Pro Pro Ala His Ser Ala Asp Glu Asn Arg Ile Ala Val
1 5 10 15

gtg ggc atg gcc gtc aaa tat gcg ggc tgt gac aat aaa gaa gag ttt 96
Val Gly Met Ala Val Lys Tyr Ala Gly Cys Asp Asn Lys Glu Glu Phe
20 25 30

tgg aag act ttg atg aat ggt agt atc aat acc aag tcg att tcg gca 144
Trp Lys Thr Leu Met Asn Gly Ser Ile Asn Thr Lys Ser Ile Ser Ala
35 40 45

gca agg ttg ggc agc aat aag cgt gac gaa cac tat gtt cct gaa cga 192
Ala Arg Leu Gly Ser Asn Lys Arg Asp Glu His Tyr Val Pro Glu Arg
50 55 60

tcg aaa tat gca gat acg ttc tgt aac gaa agg tac ggt tgt atc cag 240
Ser Lys Tyr Ala Asp Thr Phe Cys Asn Glu Arg Tyr Gly Cys Ile Gln
65 70 75 80

caa ggt acg gat aat gag cat gac ctc ctc cta ggt ctt gct caa gaa 288
Gln Gly Thr Asp Asn Glu His Asp Leu Leu Leu Gly Leu Ala Gln Glu

85	90	95	
gct ctc gct gac gct gcc ggg cgg atg gag aaa caa cct tcg gag gcg			336
Ala Leu Ala Asp Ala Ala Gly Arg Met Glu Lys Gln Pro Ser Glu Ala			
100	105	110	
ttc gat ctg gaa aat act ggc atc gtg agt ggg tgc tta tct ttt cca			384
Phe Asp Leu Glu Asn Thr Gly Ile Val Ser Gly Cys Leu Ser Phe Pro			
115	120	125	
atg gat aac ctg caa gga gag ttg ttg aac ttg tat caa agc cat gtg			432
Met Asp Asn Leu Gln Gly Glu Leu Leu Asn Leu Tyr Gln Ser His Val			
130	135	140	
gag aaa caa ctt cca cct agt gcc ttg gta gaa gcc gtg aag ctt tgg			480
Glu Lys Gln Leu Pro Pro Ser Ala Leu Val Glu Ala Val Lys Leu Trp			
145	150	155	160
tct gag cga cag aaa tct acg aaa gca cat gca ggg gac aag cgc cgg			528
Ser Glu Arg Gln Lys Ser Thr Lys Ala His Ala Gly Asp Lys Arg Arg			
165	170	175	
ttc att gac cca gct tct ttt gta gct gat aaa ctg aac cta ggc cca			576
Phe Ile Asp Pro Ala Ser Phe Val Ala Asp Lys Leu Asn Leu Gly Pro			
180	185	190	
cta cat tat gcg atc gat gca gca tgc gct tct gca ttg tac gtg tta			624
Leu His Tyr Ala Ile Asp Ala Ala Cys Ala Ser Ala Leu Tyr Val Leu			
195	200	205	
aaa tta gct caa gac cac ctt gtt tca ggt gcc gtt gat atg atg tta			672
Lys Leu Ala Gln Asp His Leu Val Ser Gly Ala Val Asp Met Met Leu			
210	215	220	
tgt gga gcg acg tgc ttc cca gaa cca ttc ttc atc ttg tct ggg ttc			720
Cys Gly Ala Thr Cys Phe Pro Glu Pro Phe Phe Ile Leu Ser Gly Phe			
225	230	235	240
tcg act ttt caa gcg atg cct gnt ggg gca gat gga gtc tca cta cct			768
Ser Thr Phe Gln Ala Met Pro Xaa Gly Ala Asp Gly Val Ser Leu Pro			
245	250	255	
ctc cat aaa acg agt gct ggg ctc act cca ggt gaa ggg ggg tcc att			816
Leu His Lys Thr Ser Ala Gly Leu Thr Pro Gly Glu Gly Gly Ser Ile			

260	265	270	
atg gtg ctc aag cga ctg aaa gac gct atc aga gat gga aat cac att			864
Met Val Leu Lys Arg Leu Lys Asp Ala Ile Arg Asp Gly Asn His Ile			
275	280	285	
tat ggt gtg ctc ctt gaa gca aat tta agt aac gca ggt tgt ggg ctt			912
Tyr Gly Val Leu Leu Glu Ala Asn Leu Ser Asn Ala Gly Cys Gly Leu			
290	295	300	
cca ctc agc ccg cac tta ccg agc gaa gaa tca tgt att cgt gat acc			960
Pro Leu Ser Pro His Leu Pro Ser Glu Glu Ser Cys Ile Arg Asp Thr			
305	310	315	320
tac cgc cgt gct gga gtt gct gca gat caa agt att cag tat att gag			1008
Tyr Arg Arg Ala Gly Val Ala Ala Asp Gln Ser Ile Gln Tyr Ile Glu			
325	330	335	
tgc cac gct acg gga acc cct cga ggg gat gtc gtg gaa att gag gcg			1056
Cys His Ala Thr Gly Thr Pro Arg Gly Asp Val Val Glu Ile Glu Ala			
340	345	350	
gtt gaa aga gtt ttc aag aaa aac gtt cca cgc tta ggc tcg acg aaa			1104
Val Glu Arg Val Phe Lys Lys Asn Val Pro Arg Leu Gly Ser Thr Lys			
355	360	365	
gga aat ttt ggt cac tcg tta gtt gcg gct ggt ttc gca ggt atg gca			1152
Gly Asn Phe Gly His Ser Leu Val Ala Ala Gly Phe Ala Gly Met Ala			
370	375	380	
aag ctt ctt ctt gca atg gaa cat gga gtg att cct ccc aca cca ggt			1200
Lys Leu Leu Leu Ala Met Glu His Gly Val Ile Pro Pro Thr Pro Gly			
385	390	395	400
ctt gat gct tcg aac cag gca agt gag cac gtt gtg aca aag gct atc			1248
Leu Asp Ala Ser Asn Gln Ala Ser Glu His Val Val Thr Lys Ala Ile			
405	410	415	
act tgg cct gag aca cat ggg gct cca aaa cga gct ggc ctt tca gca			1296
Thr Trp Pro Glu Thr His Gly Ala Pro Lys Arg Ala Gly Leu Ser Ala			
420	425	430	
ttt gga ttt ggt ggg act aat gcg cat gca ctc ttc gaa gag ttt aat			1344
Phe Gly Phe Gly Gly Thr Asn Ala His Ala Leu Phe Glu Glu Phe Asn			

435 440 445

gcc gag ggc ata agt tat cgc cct gga aag cct cca gtc gaa tcg aat 1392
 Ala Glu Gly Ile Ser Tyr Arg Pro Gly Lys Pro Pro Val Glu Ser Asn
 450 455 460

acc cgt cct tcc gtc gta ata act ggg atg gac tgt acc ttt ggg agc 1440
 Thr Arg Pro Ser Val Val Ile Thr Gly Met Asp Cys Thr Phe Gly Ser
 465 470 475 480

ctt gaa ggg att gat gcg ttc gag act gcc ctg tac gag ggg cgt gac 1488
 Leu Glu Gly Ile Asp Ala Phe Glu Thr Ala Leu Tyr Glu Gly Arg Asp
 485 490 495

gca gct cgt gac 1500
 Ala Ala Arg Asp
 500

<210> 54
 <211> 500
 <212> PRT
 <213> Thraustochytrium sp.

<220>
 <221> misc_feature
 <222> (248)..(248)
 <223> The 'Xaa' at location 248 stands for Asp, Gly, Ala, or Val.

<400> 54

Met Gln Leu Pro Pro Ala His Ser Ala Asp Glu Asn Arg Ile Ala Val
 1 5 10 15

Val Gly Met Ala Val Lys Tyr Ala Gly Cys Asp Asn Lys Glu Glu Phe
 20 25 30

Trp Lys Thr Leu Met Asn Gly Ser Ile Asn Thr Lys Ser Ile Ser Ala
 35 40 45

Ala Arg Leu Gly Ser Asn Lys Arg Asp Glu His Tyr Val Pro Glu Arg
 50 55 60

Ser Lys Tyr Ala Asp Thr Phe Cys Asn Glu Arg Tyr Gly Cys Ile Gln
65 70 75 80

Gln Gly Thr Asp Asn Glu His Asp Leu Leu Leu Gly Leu Ala Gln Glu
85 90 95

Ala Leu Ala Asp Ala Ala Gly Arg Met Glu Lys Gln Pro Ser Glu Ala
100 105 110

Phe Asp Leu Glu Asn Thr Gly Ile Val Ser Gly Cys Leu Ser Phe Pro
115 120 125

Met Asp Asn Leu Gln Gly Glu Leu Leu Asn Leu Tyr Gln Ser His Val
130 135 140

Glu Lys Gln Leu Pro Pro Ser Ala Leu Val Glu Ala Val Lys Leu Trp
145 150 155 160

Ser Glu Arg Gln Lys Ser Thr Lys Ala His Ala Gly Asp Lys Arg Arg
165 170 175

Phe Ile Asp Pro Ala Ser Phe Val Ala Asp Lys Leu Asn Leu Gly Pro
180 185 190

Leu His Tyr Ala Ile Asp Ala Ala Cys Ala Ser Ala Leu Tyr Val Leu
195 200 205

Lys Leu Ala Gln Asp His Leu Val Ser Gly Ala Val Asp Met Met Leu
210 215 220

Cys Gly Ala Thr Cys Phe Pro Glu Pro Phe Phe Ile Leu Ser Gly Phe
225 230 235 240

Ser Thr Phe Gln Ala Met Pro Xaa Gly Ala Asp Gly Val Ser Leu Pro
245 250 255

Leu His Lys Thr Ser Ala Gly Leu Thr Pro Gly Glu Gly Gly Ser Ile
260 265 270

Met Val Leu Lys Arg Leu Lys Asp Ala Ile Arg Asp Gly Asn His Ile
275 280 285

Tyr Gly Val Leu Leu Glu Ala Asn Leu Ser Asn Ala Gly Cys Gly Leu
290 295 300

Pro Leu Ser Pro His Leu Pro Ser Glu Glu Ser Cys Ile Arg Asp Thr
305 310 315 320

Tyr Arg Arg Ala Gly Val Ala Ala Asp Gln Ser Ile Gln Tyr Ile Glu
325 330 335

Cys His Ala Thr Gly Thr Pro Arg Gly Asp Val Val Glu Ile Glu Ala
340 345 350

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

Gly Asn Phe Gly His Ser Leu Val Ala Ala Gly Phe Ala Gly Met Ala
370 375 380

Lys Leu Leu Leu Ala Met Glu His Gly Val Ile Pro Pro Thr Pro Gly
385 390 395 400

Leu Asp Ala Ser Asn Gln Ala Ser Glu His Val Val Thr Lys Ala Ile
405 410 415

Thr Trp Pro Glu Thr His Gly Ala Pro Lys Arg Ala Gly Leu Ser Ala
420 425 430

Phe Gly Phe Gly Gly Thr Asn Ala His Ala Leu Phe Glu Glu Phe Asn
435 440 445

Ala Glu Gly Ile Ser Tyr Arg Pro Gly Lys Pro Pro Val Glu Ser Asn
450 455 460

Thr Arg Pro Ser Val Val Ile Thr Gly Met Asp Cys Thr Phe Gly Ser
465 470 475 480

Leu Glu Gly Ile Asp Ala Phe Glu Thr Ala Leu Tyr Glu Gly Arg Asp
485 490 495

Ala Ala Arg Asp
500

<210> 55
<211> 1500
<212> DNA
<213> Thraustochytrium sp.

<220>
<221> CDS
<222> (1)..(1500)

<400> 55
tta ccc gcc aaa cgt tgg agg ttc cta ggt gag gac ttg gag ttt ctc 48
Leu Pro Ala Lys Arg Trp Arg Phe Leu Gly Glu Asp Leu Glu Phe Leu
1 5 10 15

cga gcc atc agg ctc aag gaa aag cct agg ggt tgt ttt gtg gag agt 96
Arg Ala Ile Arg Leu Lys Glu Lys Pro Arg Gly Cys Phe Val Glu Ser
20 25 30

gtt gac gtt aac ttt aga cgg ctg aaa acg ccc ttg aca cca gaa gat 144
Val Asp Val Asn Phe Arg Arg Leu Lys Thr Pro Leu Thr Pro Glu Asp
35 40 45

atg ttg cgg ccc caa caa ctc ttg gcg gtt tct acg atg gac cga gca 192
Met Leu Arg Pro Gln Gln Leu Ala Val Ser Thr Met Asp Arg Ala
50 55 60

att atc gat gca ggt cta aag aag ggc caa cat gta gca gtt ctt gtt 240
Ile Ile Asp Ala Gly Leu Lys Lys Gly Gln His Val Ala Val Leu Val

65	70	75	80	
ggc cta gga act gac ctg gaa ctt tac cgt cat cga gca aga gtc gcg				288
Gly Leu Gly Thr Asp Leu Glu Leu Tyr Arg His Arg Ala Arg Val Ala				
	85	90	95	
ctt aaa gag gtt ttg cac ccg agc tta aag tca gac act gca att ctc				336
Leu Lys Glu Val Leu His Pro Ser Leu Lys Ser Asp Thr Ala Ile Leu				
	100	105	110	
cag aaa ata atg caa tat gtg aat gat gca gga act tcg act tca tac				384
Gln Lys Ile Met Gln Tyr Val Asn Asp Ala Gly Thr Ser Thr Ser Tyr				
	115	120	125	
aca tct tac att gga aac ctc gtt gcc acg cgt att tcg tct cag tgg				432
Thr Ser Tyr Ile Gly Asn Leu Val Ala Thr Arg Ile Ser Ser Gln Trp				
	130	135	140	
gga ttc aca ggg ccg tcc ttt act gtc aca gaa gga aat aat tcc gtg				480
Gly Phe Thr Gly Pro Ser Phe Thr Val Thr Glu Gly Asn Asn Ser Val				
	145	150	155	160
tac aga tgt gca caa cta gcc aaa gat atg ctt cag gtt aac cga gtt				528
Tyr Arg Cys Ala Gln Leu Ala Lys Asp Met Leu Gln Val Asn Arg Val				
	165	170	175	
gat gct gtc gtc atc gca ggc gtt gat ctc aac gga agc gcc gaa agt				576
Asp Ala Val Val Ile Ala Gly Val Asp Leu Asn Gly Ser Ala Glu Ser				
	180	185	190	
ttt ttt gtc cga gca aat cgt caa aag ata tcc aag cta agt cat cca				624
Phe Phe Val Arg Ala Asn Arg Gln Lys Ile Ser Lys Leu Ser His Pro				
	195	200	205	
tgt gca agc ttc gac aga gat gca gat gga ttt ttc gca ggt gag ggc				672
Cys Ala Ser Phe Asp Arg Asp Ala Asp Gly Phe Phe Ala Gly Glu Gly				
	210	215	220	
tgt ggt gcc cta gtt ttc aag agg tta gaa gac tgt gct cct cag gaa				720
Cys Gly Ala Leu Val Phe Lys Arg Leu Glu Asp Cys Ala Pro Gln Glu				
	225	230	235	240
aaa att tat gct agt ata gac tct atc gca ata gat aaa gag cct act				768
Lys Ile Tyr Ala Ser Ile Asp Ser Ile Ala Ile Asp Lys Glu Pro Thr				

245	250	255	
agc tca gct gtg aaa gct gtc tac caa agt gat tcg agt ctc tcc gat			816
Ser Ser Ala Val Lys Ala Val Tyr Gln Ser Asp Ser Ser Leu Ser Asp			
260	265	270	
att gag ctg tta gaa atc agt gga gac tcc aaa cgg ttt gca gca ttc			864
Ile Glu Leu Leu Glu Ile Ser Gly Asp Ser Lys Arg Phe Ala Ala Phe			
275	280	285	
gaa ggc gct gtg gaa att caa tca agt gtg gaa gcc cag cta aaa gga			912
Glu Gly Ala Val Glu Ile Gln Ser Ser Val Glu Ala Gln Leu Lys Gly			
290	295	300	
ctt tcc aaa gtc ctt gaa cct gca aaa ggc caa ggc gta gcg gtg gga			960
Leu Ser Lys Val Leu Glu Pro Ala Lys Gly Gln Gly Val Ala Val Gly			
305	310	315	320
agt act cga gca acc gtt ggg gat ata ggg tat gct aca gga gcg gca			1008
Ser Thr Arg Ala Thr Val Gly Asp Ile Gly Tyr Ala Thr Gly Ala Ala			
325	330	335	
agc ctg att aaa act gca ctc tgc tta tat aat cgc tac ctt ccg gca			1056
Ser Leu Ile Lys Thr Ala Leu Cys Leu Tyr Asn Arg Tyr Leu Pro Ala			
340	345	350	
tta gca aac tgg agt ggc cca tgt gaa cag tcc gcc tgg ggc tca aac			1104
Leu Ala Asn Trp Ser Gly Pro Cys Glu Gln Ser Ala Trp Gly Ser Asn			
355	360	365	
atg ttc gtt tgc cat gaa aca cgg ccg tgg atg aaa aac cag aat gaa			1152
Met Phe Val Cys His Glu Thr Arg Pro Trp Met Lys Asn Gln Asn Glu			
370	375	380	
aag aga tgt gcc ctc att tct gga aca gat cca tct cat aca tgc ttt			1200
Lys Arg Cys Ala Leu Ile Ser Gly Thr Asp Pro Ser His Thr Cys Phe			
385	390	395	400
tcc ctc gta cta tcg gat act ggg tgt tat gaa gag cac aat cga acg			1248
Ser Leu Val Leu Ser Asp Thr Gly Cys Tyr Glu Glu His Asn Arg Thr			
405	410	415	
tgc ttt gat gtg caa gcg cca cag cta gtt ctg ata cac gga ttc gat			1296
Cys Phe Asp Val Gln Ala Pro Gln Leu Val Leu Ile His Gly Phe Asp			

420	425	430	
gga aaa act att gtg cgg cga ctt gaa gga tat ctc ctt gaa ctt gtt Gly Lys Thr Ile Val Arg Arg Leu Glu Gly Tyr Leu Leu Glu Leu Val 435 440 445			1344
gaa ggg cat gca agc cct tca gag tat ttc cac aaa ctg att gga caa Glu Gly His Ala Ser Pro Ser Glu Tyr Phe His Lys Leu Ile Gly Gln 450 455 460			1392
agt cta ctt gag aac tcg aaa gaa agt aaa ctc aca ctt tcg ctt gtg Ser Leu Leu Glu Asn Ser Lys Glu Ser Lys Leu Thr Leu Ser Leu Val 465 470 475 480			1440
tgc aat ccg aac cag ctc caa aag gag ctc atg ctt gct atc aaa gga Cys Asn Pro Asn Gln Leu Gln Lys Glu Leu Met Leu Ala Ile Lys Gly 485 490 495			1488
gta caa cga agc Val Gln Arg Ser 500			1500

<210> 56
<211> 500
<212> PRT
<213> Thraustochytrium sp.

<400> 56

Leu	Pro	Ala	Lys	Arg	Trp	Arg	Phe	Leu	Gly	Glu	Asp	Leu	Glu	Phe	Leu
1				5					10					15	

Arg	Ala	Ile	Arg	Leu	Lys	Glu	Lys	Pro	Arg	Gly	Cys	Phe	Val	Glu	Ser
			20					25					30		

Val	Asp	Val	Asn	Phe	Arg	Arg	Leu	Lys	Thr	Pro	Leu	Thr	Pro	Glu	Asp
		35					40					45			

Met	Leu	Arg	Pro	Gln	Gln	Leu	Leu	Ala	Val	Ser	Thr	Met	Asp	Arg	Ala
						55						60			

Ile Ile Asp Ala Gly Leu Lys Lys Gly Gln His Val Ala Val Leu Val
65 70 75 80

Gly Leu Gly Thr Asp Leu Glu Leu Tyr Arg His Arg Ala Arg Val Ala
85 90 95

Leu Lys Glu Val Leu His Pro Ser Leu Lys Ser Asp Thr Ala Ile Leu
100 105 110

Gln Lys Ile Met Gln Tyr Val Asn Asp Ala Gly Thr Ser Thr Ser Tyr
115 120 125

Thr Ser Tyr Ile Gly Asn Leu Val Ala Thr Arg Ile Ser Ser Gln Trp
130 135 140

Gly Phe Thr Gly Pro Ser Phe Thr Val Thr Glu Gly Asn Asn Ser Val
145 150 155 160

Tyr Arg Cys Ala Gln Leu Ala Lys Asp Met Leu Gln Val Asn Arg Val
165 170 175

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

Phe Phe Val Arg Ala Asn Arg Gln Lys Ile Ser Lys Leu Ser His Pro
195 200 205

Cys Ala Ser Phe Asp Arg Asp Ala Asp Gly Phe Phe Ala Gly Glu Gly
210 215 220

Cys Gly Ala Leu Val Phe Lys Arg Leu Glu Asp Cys Ala Pro Gln Glu
225 230 235 240

Lys Ile Tyr Ala Ser Ile Asp Ser Ile Ala Ile Asp Lys Glu Pro Thr
245 250 255

Ser Ser Ala Val Lys Ala Val Tyr Gln Ser Asp Ser Ser Leu Ser Asp
260 265 270

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

Glu Gly Ala Val Glu Ile Gln Ser Ser Val Glu Ala Gln Leu Lys Gly
290 295 300

Leu Ser Lys Val Leu Glu Pro Ala Lys Gly Gln Gly Val Ala Val Gly
305 310 315 320

Ser Thr Arg Ala Thr Val Gly Asp Ile Gly Tyr Ala Thr Gly Ala Ala
325 330 335

Ser Leu Ile Lys Thr Ala Leu Cys Leu Tyr Asn Arg Tyr Leu Pro Ala
340 345 350

Leu Ala Asn Trp Ser Gly Pro Cys Glu Gln Ser Ala Trp Gly Ser Asn
355 360 365

Met Phe Val Cys His Glu Thr Arg Pro Trp Met Lys Asn Gln Asn Glu
370 375 380

Lys Arg Cys Ala Leu Ile Ser Gly Thr Asp Pro Ser His Thr Cys Phe
385 390 395 400

Ser Leu Val Leu Ser Asp Thr Gly Cys Tyr Glu Glu His Asn Arg Thr
405 410 415

Cys Phe Asp Val Gln Ala Pro Gln Leu Val Leu Ile His Gly Phe Asp
420 425 430

Gly Lys Thr Ile Val Arg Arg Leu Glu Gly Tyr Leu Leu Glu Leu Val
435 440 445

Glu Gly His Ala Ser Pro Ser Glu Tyr Phe His Lys Leu Ile Gly Gln
450 455 460

Ser Leu Leu Glu Asn Ser Lys Glu Ser Lys Leu Thr Leu Ser Leu Val
465 470 475 480

Cys Asn Pro Asn Gln Leu Gln Lys Glu Leu Met Leu Ala Ile Lys Gly
485 490 495

Val Gln Arg Ser
500

<210> 57
<211> 1500
<212> DNA
<213> Thraustochytrium sp.

<220>
<221> CDS
<222> (1)..(1500)

<400> 57
atg tta aca ggg aag gat tgg gtc agt cca tca gga agt tgt ttt gcc 48
Met Leu Thr Gly Lys Asp Trp Val Ser Pro Ser Gly Ser Cys Phe Ala
1 5 10 15

cca aat ccg tta tca agc gca aaa gtg gca ttc atg tac gga gaa ggc 96
Pro Asn Pro Leu Ser Ser Ala Lys Val Ala Phe Met Tyr Gly Glu Gly
20 25 30

cga agc ccg tac tgt ggt gta ggc ttg ggt cta cat cgt ttg tgg ccc 144
Arg Ser Pro Tyr Cys Gly Val Gly Leu Gly Leu His Arg Leu Trp Pro
35 40 45

ggt ctc cat gaa aat gtg aac aat aag aca gtc gat tta tgg acg gaa 192
Gly Leu His Glu Asn Val Asn Asn Lys Thr Val Asp Leu Trp Thr Glu
50 55 60

gga gat ggt tgg tta tat cct cga acg ttg aca cga gaa gag cat aca 240
Gly Asp Gly Trp Leu Tyr Pro Arg Thr Leu Thr Arg Glu Glu His Thr
65 70 75 80

aaa gcc atc gaa tct ttc aac gca aat caa att gaa atg ttt cgc gct Lys Ala Ile Glu Ser Phe Asn Ala Asn Gln Ile Glu Met Phe Arg Ala 85 90 95	288
ggg att ttc atc tca atg tgt cag aca gac tat gtc atg aat gtt ctc Gly Ile Phe Ile Ser Met Cys Gln Thr Asp Tyr Val Met Asn Val Leu 100 105 110	336
ggg gtc cag cct aag gcc gga ttt ggg ctg agc ttg gga gaa att tca Gly Val Gln Pro Lys Ala Gly Phe Gly Leu Ser Leu Gly Glu Ile Ser 115 120 125	384
atg ctc ttt gcg atg tca aag gag aac tgc agg cag tca cag gaa atg Met Leu Phe Ala Met Ser Lys Glu Asn Cys Arg Gln Ser Gln Glu Met 130 135 140	432
acc aat cgt ttg cgc ggt tct cca gtg tgg tct aac gag ctt gct atc Thr Asn Arg Leu Arg Gly Ser Pro Val Trp Ser Asn Glu Leu Ala Ile 145 150 155 160	480
aac ttc aat gca att cgc aag tta tgg aaa atc ccc cga gga gct ccc Asn Phe Asn Ala Ile Arg Lys Leu Trp Lys Ile Pro Arg Gly Ala Pro 165 170 175	528
tta gaa tcc ttt tgg caa gga tac ttg gtt cac ggc aca aga gaa gaa Leu Glu Ser Phe Trp Gln Gly Tyr Leu Val His Gly Thr Arg Glu Glu 180 185 190	576
gta gag cat gct att ggt ctt tct gag cct tat gta cgt ctg ctt att Val Glu His Ala Ile Gly Leu Ser Glu Pro Tyr Val Arg Leu Leu Ile 195 200 205	624
gtg aac gat tca agg agt gcc ttg att gct gga aaa cca gac gcc tgt Val Asn Asp Ser Arg Ser Ala Leu Ile Ala Gly Lys Pro Asp Ala Cys 210 215 220	672
cag gca gta atc agt aga cta aac tcc aag ttc cct tct ctg ccg gta Gln Ala Val Ile Ser Arg Leu Asn Ser Lys Phe Pro Ser Leu Pro Val 225 230 235 240	720
aag caa gga atg att ggt cat tgc cca gaa gtt cgt gcg ttc atc aaa Lys Gln Gly Met Ile Gly His Cys Pro Glu Val Arg Ala Phe Ile Lys 245 250 255	768

gat att ggg tac atc cat gaa aca ctc cga att tcc aat gac tat tcg Asp Ile Gly Tyr Ile His Glu Thr Leu Arg Ile Ser Asn Asp Tyr Ser 260 265 270	816
gat tgt cag ctt ttc tca gcg gta acc aag ggc gca ctt gac agc tcc Asp Cys Gln Leu Phe Ser Ala Val Thr Lys Gly Ala Leu Asp Ser Ser 275 280 285	864
aca atg gaa atc aaa cac ttt gtg gga gag gtc tac tcc cgg atc gca Thr Met Glu Ile Lys His Phe Val Gly Glu Val Tyr Ser Arg Ile Ala 290 295 300	912
gac ttt cct caa atc gtc aac acg gtg cat tcg gct ggt tat gac gta Asp Phe Pro Gln Ile Val Asn Thr Val His Ser Ala Gly Tyr Asp Val 305 310 315 320	960
ttt ctt gag ctt ggc tgt gat gct tct aga tct gca gca gtt caa aac Phe Leu Glu Leu Gly Cys Asp Ala Ser Arg Ser Ala Ala Val Gln Asn 325 330 335	1008
att ctt ggt ggt caa gga aag ttc ttg tct aca gct att gac aaa aaa Ile Leu Gly Gly Gln Gly Lys Phe Leu Ser Thr Ala Ile Asp Lys Lys 340 345 350	1056
gga cac tcc gcc tgg tca caa gta ctt cgg gct acc gca tca tta gct Gly His Ser Ala Trp Ser Gln Val Leu Arg Ala Thr Ala Ser Leu Ala 355 360 365	1104
gca cat cga gta ccg gga atc tca att ttg gat ttg ttt cac cca aat Ala His Arg Val Pro Gly Ile Ser Ile Leu Asp Leu Phe His Pro Asn 370 375 380	1152
ttc cga gaa atg tgc tgt aca atg gca acc aca cct aaa gtg gaa gat Phe Arg Glu Met Cys Cys Thr Met Ala Thr Thr Pro Lys Val Glu Asp 385 390 395 400	1200
aag ttc ctg cgc acg att caa atc aat ggt cgg ttt gaa aaa gaa atg Lys Phe Leu Arg Thr Ile Gln Ile Asn Gly Arg Phe Glu Lys Glu Met 405 410 415	1248
att cac cta gaa gat aca aca tta agt tgc tta ccc gct cca agt gaa Ile His Leu Glu Asp Thr Thr Leu Ser Cys Leu Pro Ala Pro Ser Glu 420 425 430	1296

gca aat atc gca gct att caa tct cgg tca att cga tct gct gcg gcg 1344
Ala Asn Ile Ala Ala Ile Gln Ser Arg Ser Ile Arg Ser Ala Ala Ala
435 440 445

cgt tct gga caa tcc cat gat tgt gca tcc cat agc cat gaa gaa aat 1392
Arg Ser Gly Gln Ser His Asp Cys Ala Ser His Ser His Glu Glu Asn
450 455 460

aag gat tca tgc cct gaa aag ctg aag ctt gat tct gtg tcc gtc gcc 1440
Lys Asp Ser Cys Pro Glu Lys Leu Lys Leu Asp Ser Val Ser Val Ala
465 470 475 480

ata aat ttc gac aat gat gac cgc att cag ctt ggg cac gcg ggt ttt 1488
Ile Asn Phe Asp Asn Asp Asp Arg Ile Gln Leu Gly His Ala Gly Phe
485 490 495

cgg gag atg tac 1500
Arg Glu Met Tyr
500

<210> 58
<211> 500
<212> PRT
<213> Thraustochytrium sp.

<400> 58

Met Leu Thr Gly Lys Asp Trp Val Ser Pro Ser Gly Ser Cys Phe Ala
1 5 10 15

Pro Asn Pro Leu Ser Ser Ala Lys Val Ala Phe Met Tyr Gly Glu Gly
20 25 30

Arg Ser Pro Tyr Cys Gly Val Gly Leu Gly Leu His Arg Leu Trp Pro
35 40 45

Gly Leu His Glu Asn Val Asn Asn Lys Thr Val Asp Leu Trp Thr Glu
50 55 60

Gly Asp Gly Trp Leu Tyr Pro Arg Thr Leu Thr Arg Glu Glu His Thr
65 70 75 80

Lys Ala Ile Glu Ser Phe Asn Ala Asn Gln Ile Glu Met Phe Arg Ala
85 90 95

Gly Ile Phe Ile Ser Met Cys Gln Thr Asp Tyr Val Met Asn Val Leu
100 105 110

Gly Val Gln Pro Lys Ala Gly Phe Gly Leu Ser Leu Gly Glu Ile Ser
115 120 125

Met Leu Phe Ala Met Ser Lys Glu Asn Cys Arg Gln Ser Gln Glu Met
130 135 140

Thr Asn Arg Leu Arg Gly Ser Pro Val Trp Ser Asn Glu Leu Ala Ile
145 150 155 160

Asn Phe Asn Ala Ile Arg Lys Leu Trp Lys Ile Pro Arg Gly Ala Pro
165 170 175

Leu Glu Ser Phe Trp Gln Gly Tyr Leu Val His Gly Thr Arg Glu Glu
180 185 190

Val Glu His Ala Ile Gly Leu Ser Glu Pro Tyr Val Arg Leu Leu Ile
195 200 205

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

Gln Ala Val Ile Ser Arg Leu Asn Ser Lys Phe Pro Ser Leu Pro Val
225 230 235 240

Lys Gln Gly Met Ile Gly His Cys Pro Glu Val Arg Ala Phe Ile Lys
245 250 255

Asp Ile Gly Tyr Ile His Glu Thr Leu Arg Ile Ser Asn Asp Tyr Ser

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

450

455

460

Lys Asp Ser Cys Pro Glu Lys Leu Lys Leu Asp Ser Val Ser Val Ala
465 470 475 480

Ile Asn Phe Asp Asn Asp Asp Arg Ile Gln Leu Gly His Ala Gly Phe
485 490 495

Arg Glu Met Tyr
500

<210> 59

<211> 1305

<212> DNA

<213> Thraustochytrium sp.

<220>

<221> CDS

<222> (1)..(1305)

<400> 59

aat aca aga tat agc ttg tac aca ggg gcg atg gca aag gga att gca 48
Asn Thr Arg Tyr Ser Leu Tyr Thr Gly Ala Met Ala Lys Gly Ile Ala
1 5 10 15

tct gca gat ctt gtc att gcc gct ggg aaa gag ggc atc cta gct tcc 96
Ser Ala Asp Leu Val Ile Ala Ala Gly Lys Glu Gly Ile Leu Ala Ser
20 25 30

tat gga gct gga gga cta cct ctt gct act gtt cga aag gga ata gac 144
Tyr Gly Ala Gly Gly Leu Pro Leu Ala Thr Val Arg Lys Gly Ile Asp
35 40 45

aaa att caa caa gcc ttg cca agt ggc cca tat gct gta aat ctt att 192
Lys Ile Gln Gln Ala Leu Pro Ser Gly Pro Tyr Ala Val Asn Leu Ile
50 55 60

cac tct ccc ttt gac ggc aac ttg gag cag gga aac gtc gat ttg ttc 240
His Ser Pro Phe Asp Gly Asn Leu Glu Gln Gly Asn Val Asp Leu Phe
65 70 75 80

ttg gaa aag aac gtc cgc gtg gcg gaa tgt tcc gcg ttt aca acg cta 288
Leu Glu Lys Asn Val Arg Val Ala Glu Cys Ser Ala Phe Thr Thr Leu

85	90	95	
aca gtg cca gta gta cac tat cgt gct gca ggg ctt gtt cgg cgc caa Thr Val Pro Val Val His Tyr Arg Ala Ala Gly Leu Val Arg Arg Gln 100 105 110			336
gat gga agc att ttg atc aag aac cga atc att gct aaa gta tct agg Asp Gly Ser Ile Leu Ile Lys Asn Arg Ile Ile Ala Lys Val Ser Arg 115 120 125			384
aca gaa ctc gct gag atg ttc ctt cgt ccg gca cct caa atc atc ctc Thr Glu Leu Ala Glu Met Phe Leu Arg Pro Ala Pro Gln Ile Ile Leu 130 135 140			432
gaa aaa ctg gta gca gca gaa atc att tca tct gac caa gcg cgt atg Glu Lys Leu Val Ala Ala Glu Ile Ile Ser Ser Asp Gln Ala Arg Met 145 150 155 160			480
gca gcc aaa gtt ccc atg gcg gac gac atc gca gtc gaa gcc gac tct Ala Ala Lys Val Pro Met Ala Asp Asp Ile Ala Val Glu Ala Asp Ser 165 170 175			528
ggg ggc cac acg gat aat cgg cct atg cac gtc att ttg ccc ctg ata Gly Gly His Thr Asp Asn Arg Pro Met His Val Ile Leu Pro Leu Ile 180 185 190			576
att caa ctc cgc aat act ata ctt gca gag tat ggc tgt gcc acg gct Ile Gln Leu Arg Asn Thr Ile Leu Ala Glu Tyr Gly Cys Ala Thr Ala 195 200 205			624
ttt cgt acc cgt ata ggc gct gga gga ggc att ggt tgt cct tca gcg Phe Arg Thr Arg Ile Gly Ala Gly Gly Gly Ile Gly Cys Pro Ser Ala 210 215 220			672
gcc ctc gca gcc ttt gat atg ggt gcg agt ttt gtc gtg act gga agc Ala Leu Ala Ala Phe Asp Met Gly Ala Ser Phe Val Val Thr Gly Ser 225 230 235 240			720
ata aat caa att tgc cgc gag gca ggg act tgc gat act gtt cgg gag Ile Asn Gln Ile Cys Arg Glu Ala Gly Thr Cys Asp Thr Val Arg Glu 245 250 255			768
cta ctt gcc aac tca agc tac tcg gac gtg acg atg gcg cca gca gca Leu Leu Ala Asn Ser Ser Tyr Ser Asp Val Thr Met Ala Pro Ala Ala			816

260	265	270	
gac atg ttt gac caa ggt gtg aaa ctc caa gtc tta aaa cga gga acg			864
Asp Met Phe Asp Gln Gly Val Lys Leu Gln Val Leu Lys Arg Gly Thr			
275	280	285	
atg ttt cca agc aga gca aat aaa ctc cgg aag ctc ttt gtg aac tac			912
Met Phe Pro Ser Arg Ala Asn Lys Leu Arg Lys Leu Phe Val Asn Tyr			
290	295	300	
gaa tct cta gaa aca ctc ccg tcg aaa gag ttg aaa tac ctg gaa aac			960
Glu Ser Leu Glu Thr Leu Pro Ser Lys Glu Leu Lys Tyr Leu Glu Asn			
305	310	315	320
atc ata ttc aag caa gca gta gac cag gtg tgg gag gaa aca aag cgc			1008
Ile Ile Phe Lys Gln Ala Val Asp Gln Val Trp Glu Glu Thr Lys Arg			
325	330	335	
ttt tac tgt gaa aaa ctg aac aat cca gat aaa att gca agg gcc atg			1056
Phe Tyr Cys Glu Lys Leu Asn Asn Pro Asp Lys Ile Ala Arg Ala Met			
340	345	350	
aaa gat cct aaa ttg aag atg tcg ctt tgc ttt cgg tgg tat ctc tcc			1104
Lys Asp Pro Lys Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Ser			
355	360	365	
aag agc tct ggg tgg gcc aac gca gga att aaa tct cgt gca ctc gac			1152
Lys Ser Ser Gly Trp Ala Asn Ala Gly Ile Lys Ser Arg Ala Leu Asp			
370	375	380	
tac cag atc tgg tgt ggc ccg gca atg ggc tcg ttc aac aat ttc gcc			1200
Tyr Gln Ile Trp Cys Gly Pro Ala Met Gly Ser Phe Asn Asn Phe Ala			
385	390	395	400
agc ggc aca tcc ctc gat tgg aaa gtg act ggg gtt ttc cct ggc gtt			1248
Ser Gly Thr Ser Leu Asp Trp Lys Val Thr Gly Val Phe Pro Gly Val			
405	410	415	
gcg gaa gta aac atg gcc att tta gat ggc gcg cga gaa cta gct gct			1296
Ala Glu Val Asn Met Ala Ile Leu Asp Gly Ala Arg Glu Leu Ala Ala			
420	425	430	
aaa cga aat			1305
Lys Arg Asn			

435

<210> 60
 <211> 435
 <212> PRT
 <213> Thraustochytrium sp.

<400> 60

Asn Thr Arg Tyr Ser Leu Tyr Thr Gly Ala Met Ala Lys Gly Ile Ala
 1 5 10 15

Ser Ala Asp Leu Val Ile Ala Ala Gly Lys Glu Gly Ile Leu Ala Ser
 20 25 30

Tyr Gly Ala Gly Gly Leu Pro Leu Ala Thr Val Arg Lys Gly Ile Asp
 35 40 45

Lys Ile Gln Gln Ala Leu Pro Ser Gly Pro Tyr Ala Val Asn Leu Ile
 50 55 60

His Ser Pro Phe Asp Gly Asn Leu Glu Gln Gly Asn Val Asp Leu Phe
 65 70 75 80

Leu Glu Lys Asn Val Arg Val Ala Glu Cys Ser Ala Phe Thr Thr Leu
 85 90 95

Thr Val Pro Val Val His Tyr Arg Ala Ala Gly Leu Val Arg Arg Gln
 100 105 110

Asp Gly Ser Ile Leu Ile Lys Asn Arg Ile Ile Ala Lys Val Ser Arg
 115 120 125

Thr Glu Leu Ala Glu Met Phe Leu Arg Pro Ala Pro Gln Ile Ile Leu
 130 135 140

Glu Lys Leu Val Ala Ala Glu Ile Ile Ser Ser Asp Gln Ala Arg Met
145 150 155 160

Ala Ala Lys Val Pro Met Ala Asp Asp Ile Ala Val Glu Ala Asp Ser
165 170 175

Gly Gly His Thr Asp Asn Arg Pro Met His Val Ile Leu Pro Leu Ile
180 185 190

Ile Gln Leu Arg Asn Thr Ile Leu Ala Glu Tyr Gly Cys Ala Thr Ala
195 200 205

Phe Arg Thr Arg Ile Gly Ala Gly Gly Gly Ile Gly Cys Pro Ser Ala
210 215 220

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

Ile Asn Gln Ile Cys Arg Glu Ala Gly Thr Cys Asp Thr Val Arg Glu
245 250 255

Leu Leu Ala Asn Ser Ser Tyr Ser Asp Val Thr Met Ala Pro Ala Ala
260 265 270

Asp Met Phe Asp Gln Gly Val Lys Leu Gln Val Leu Lys Arg Gly Thr
275 280 285

Met Phe Pro Ser Arg Ala Asn Lys Leu Arg Lys Leu Phe Val Asn Tyr
290 295 300

Glu Ser Leu Glu Thr Leu Pro Ser Lys Glu Leu Lys Tyr Leu Glu Asn
305 310 315 320

Ile Ile Phe Lys Gln Ala Val Asp Gln Val Trp Glu Glu Thr Lys Arg
325 330 335

Phe Tyr Cys Glu Lys Leu Asn Asn Pro Asp Lys Ile Ala Arg Ala Met

340 345 350

Lys Asp Pro Lys Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Ser
355 360 365

Lys Ser Ser Gly Trp Ala Asn Ala Gly Ile Lys Ser Arg Ala Leu Asp
370 375 380

Tyr Gln Ile Trp Cys Gly Pro Ala Met Gly Ser Phe Asn Asn Phe Ala
385 390 395 400

Ser Gly Thr Ser Leu Asp Trp Lys Val Thr Gly Val Phe Pro Gly Val
405 410 415

Ala Glu Val Asn Met Ala Ile Leu Asp Gly Ala Arg Glu Leu Ala Ala
420 425 430

Lys Arg Asn
435

<210> 61
<211> 4410
<212> DNA
<213> Thraustochytrium sp.

<220>
<221> CDS
<222> (1)..(4410)

<400> 61
atg ggc ccg cga gtg gcg tca ggc aag gtg ccg gct tgg gag atg agc 48
Met Gly Pro Arg Val Ala Ser Gly Lys Val Pro Ala Trp Glu Met Ser
1 5 10 15

aag tcc gag ctg tgt gat gac cgc acg gta gtc ttt gac tat gag gag 96
Lys Ser Glu Leu Cys Asp Asp Arg Thr Val Val Phe Asp Tyr Glu Glu
20 25 30

ctg ctg gag ttc gct gag ggc gat atc agt aag gtt ttt ggg ccg gag 144
Leu Leu Glu Phe Ala Glu Gly Asp Ile Ser Lys Val Phe Gly Pro Glu

35	40	45	
ttc aaa gtg gtg gac ggg ttt agg cgc agg gtg agg ttg ccc gct cga			192
Phe Lys Val Val Asp Gly Phe Arg Arg Arg Val Arg Leu Pro Ala Arg			
50	55	60	
gag tac ctg ctg gtg acc cgg gtt acg ctg atg gat gcc gag gtg ggc			240
Glu Tyr Leu Leu Val Thr Arg Val Thr Leu Met Asp Ala Glu Val Gly			
65	70	75	80
aac ttt cga gtg gga gca cgt atg gtg aca gag tat gac gta cct gtg			288
Asn Phe Arg Val Gly Ala Arg Met Val Thr Glu Tyr Asp Val Pro Val			
	85	90	95
aac gga gag ctc tcg gaa ggg gga gat gtg ccg tgg gct gtg ttg gtg			336
Asn Gly Glu Leu Ser Glu Gly Gly Asp Val Pro Trp Ala Val Leu Val			
	100	105	110
gaa gcc ggg cag tgc gac ttg ctg cta att tct tac atg ggc atc gat			384
Glu Ala Gly Gln Cys Asp Leu Leu Leu Ile Ser Tyr Met Gly Ile Asp			
	115	120	125
ttc cag tgc aaa gga gag cgg gtc tac cgg ctg ctg aac acc acc ttg			432
Phe Gln Cys Lys Gly Glu Arg Val Tyr Arg Leu Leu Asn Thr Thr Leu			
130	135	140	
acg ttt ttt ggc gtc gcg aaa gaa ggg gaa acg ctt gtg tac gat att			480
Thr Phe Phe Gly Val Ala Lys Glu Gly Glu Thr Leu Val Tyr Asp Ile			
145	150	155	160
cgc gtc acg ggt ttc gcc aag agg ccg gac gga gat atc tcc atg ttc			528
Arg Val Thr Gly Phe Ala Lys Arg Pro Asp Gly Asp Ile Ser Met Phe			
	165	170	175
ttt ttc gaa tat gat tgc tac tgc aat ggc aag ctt ctc atc gaa atg			576
Phe Phe Glu Tyr Asp Cys Tyr Cys Asn Gly Lys Leu Leu Ile Glu Met			
	180	185	190
cga gat ggc tct gca ggc ttc ttc acg gac gaa gag ctc gct gcc ggc			624
Arg Asp Gly Ser Ala Gly Phe Phe Thr Asp Glu Glu Leu Ala Ala Gly			
195	200	205	
aaa gga gtg gtc gtc act cgt gca cag caa aac atg cgg gac aaa att			672
Lys Gly Val Val Val Thr Arg Ala Gln Gln Asn Met Arg Asp Lys Ile			

210	215	220	
gta cgg cag tcc att gag cct ttt gca ctg gcg gct tgc acg cac aaa			720
Val Arg Gln Ser Ile Glu Pro Phe Ala Leu Ala Ala Cys Thr His Lys			
225	230	235	240
acg act ctg aac gag agt gac atg cag tcc ctt gtg gag cga aac tgg			768
Thr Thr Leu Asn Glu Ser Asp Met Gln Ser Leu Val Glu Arg Asn Trp			
245	250	255	
gca aac gtt ttt ggc acc agt aac aag atg gcg gag ctc aac tat aaa			816
Ala Asn Val Phe Gly Thr Ser Asn Lys Met Ala Glu Leu Asn Tyr Lys			
260	265	270	
att tgc gcc agg aaa atg ctc atg atc gac agg gtt acc cac att gac			864
Ile Cys Ala Arg Lys Met Leu Met Ile Asp Arg Val Thr His Ile Asp			
275	280	285	
cac cac ggt ggg gcg tat ggc ctc gga cta ctt gtt gga gag aag atc			912
His His Gly Gly Ala Tyr Gly Leu Gly Leu Leu Val Gly Glu Lys Ile			
290	295	300	
ttg gat cga aac cat tgg tac ttt cct tgt cac ttt gtc aat gat caa			960
Leu Asp Arg Asn His Trp Tyr Phe Pro Cys His Phe Val Asn Asp Gln			
305	310	315	320
gtc atg gca ggg tca ctg gtc agc gat ggt tgc agc cag ctc tta aaa			1008
Val Met Ala Gly Ser Leu Val Ser Asp Gly Cys Ser Gln Leu Leu Lys			
325	330	335	
ctc tat atg atc tgg ctt ggc ctc cac ctg aaa atg gag gaa ttt gat			1056
Leu Tyr Met Ile Trp Leu Gly Leu His Leu Lys Met Glu Glu Phe Asp			
340	345	350	
ttt ctc cca gtt agc ggc cac aaa aac aag gtg cga tgc agg gga caa			1104
Phe Leu Pro Val Ser Gly His Lys Asn Lys Val Arg Cys Arg Gly Gln			
355	360	365	
att tca ccg cat aaa ggc aag ctt gtc tac gtc atg gaa atc aaa aag			1152
Ile Ser Pro His Lys Gly Lys Leu Val Tyr Val Met Glu Ile Lys Lys			
370	375	380	
atg ggt tac gat caa gca tct gga agc cca tac gcc atc gcg gac gtt			1200
Met Gly Tyr Asp Gln Ala Ser Gly Ser Pro Tyr Ala Ile Ala Asp Val			

385	390	395	400	
gat atc att gac gtc aac gaa gag ctg ggt caa agt ttt gac atc aac				1248
Asp Ile Ile Asp Val Asn Glu Glu Leu Gly Gln Ser Phe Asp Ile Asn				
	405	410	415	
gac ctt gcg agc tac gga aaa ggt gac ctg agc aaa aaa atc gtg gtt				1296
Asp Leu Ala Ser Tyr Gly Lys Gly Asp Leu Ser Lys Lys Ile Val Val				
	420	425	430	
gac ttc aaa gga att gct ttg cag ctc aaa ggc cgc gct ttt tca cgc				1344
Asp Phe Lys Gly Ile Ala Leu Gln Leu Lys Gly Arg Ala Phe Ser Arg				
	435	440	445	
atg agt tcc agc tcg tcc ttg aac gaa gga tgg caa tgt gtt cca aaa				1392
Met Ser Ser Ser Ser Ser Leu Asn Glu Gly Trp Gln Cys Val Pro Lys				
	450	455	460	
cca agc cag aga atg gaa cac gaa cag ccc cct gct cac tgc ctt gca				1440
Pro Ser Gln Arg Met Glu His Glu Gln Pro Pro Ala His Cys Leu Ala				
	465	470	475	480
agc gac ccc gaa gcc cct tca act gtg acc tgg cac cca atg tca aag				1488
Ser Asp Pro Glu Ala Pro Ser Thr Val Thr Trp His Pro Met Ser Lys				
	485	490	495	
ctt cct ggc aac cct acg ccg ttc ttc tcc cct tca tct tac cct ccg				1536
Leu Pro Gly Asn Pro Thr Pro Phe Phe Ser Pro Ser Ser Tyr Pro Pro				
	500	505	510	
agg gca att tgc ttc atc cct ttc ccg ggc aat ccc ctt gac aac aac				1584
Arg Ala Ile Cys Phe Ile Pro Phe Pro Gly Asn Pro Leu Asp Asn Asn				
	515	520	525	
tgc aag gct gga gaa atg ccc ctg aac tgg tac aac atg tca gag ttc				1632
Cys Lys Ala Gly Glu Met Pro Leu Asn Trp Tyr Asn Met Ser Glu Phe				
	530	535	540	
atg tgt ggc aag gtt tct aac tgc ttg ggc cca gaa ttc gca cgc ttt				1680
Met Cys Gly Lys Val Ser Asn Cys Leu Gly Pro Glu Phe Ala Arg Phe				
	545	550	555	560
gac aag tcg aac acc agc cgg agc cct gct ttt gac ttg gct ctg gtg				1728
Asp Lys Ser Asn Thr Ser Arg Ser Pro Ala Phe Asp Leu Ala Leu Val				

565	570	575	
acc cga gtt gtt gaa gtc aca aac atg gaa cac ggc aag ttt cta aac			1776
Thr Arg Val Val Glu Val Thr Asn Met Glu His Gly Lys Phe Leu Asn			
580	585	590	
gtt gat tgc aat cca agc aaa ggc aca atg gtg ggg gag ttt gac tgt			1824
Val Asp Cys Asn Pro Ser Lys Gly Thr Met Val Gly Glu Phe Asp Cys			
595	600	605	
ccc caa gac gcg tgg ttc ttt gat ggt tgc tgc aac gac ggc cat atg			1872
Pro Gln Asp Ala Trp Phe Phe Asp Gly Ser Cys Asn Asp Gly His Met			
610	615	620	
ccg tat tcc att atc atg gaa atc gga ctg caa acc tca ggt gtt ctc			1920
Pro Tyr Ser Ile Ile Met Glu Ile Gly Leu Gln Thr Ser Gly Val Leu			
625	630	635	640
acc tcg gtg ttg aag gca ccg ctg act atg gac aag gat gac att ctc			1968
Thr Ser Val Leu Lys Ala Pro Leu Thr Met Asp Lys Asp Asp Ile Leu			
645	650	655	
ttt cga aac ctc gat gca agt gct gaa atg gtg cgt cca gac gtg gat			2016
Phe Arg Asn Leu Asp Ala Ser Ala Glu Met Val Arg Pro Asp Val Asp			
660	665	670	
gtt cgc ggc aaa acg att cga aac gtg acc aag tgt acc ggc tat gca			2064
Val Arg Gly Lys Thr Ile Arg Asn Val Thr Lys Cys Thr Gly Tyr Ala			
675	680	685	
atg ttg gga aag atg ggg att cac cgg ttc acg ttt gag ttg agc gtt			2112
Met Leu Gly Lys Met Gly Ile His Arg Phe Thr Phe Glu Leu Ser Val			
690	695	700	
gac ggc gtg gta ttt tat aaa gga tcc act tcc ttt gga tgg ttc act			2160
Asp Gly Val Val Phe Tyr Lys Gly Ser Thr Ser Phe Gly Trp Phe Thr			
705	710	715	720
ccc gag gtg ttt gct cag caa gct gga ctc gac aac ggg aaa aag acg			2208
Pro Glu Val Phe Ala Gln Gln Ala Gly Leu Asp Asn Gly Lys Lys Thr			
725	730	735	
gag ccc tgg tgc aag act aac aac acc tcg gtt cga aga gtt gaa atc			2256
Glu Pro Trp Cys Lys Thr Asn Asn Thr Ser Val Arg Arg Val Glu Ile			

740	745	750	
gca tcc gcc aaa gga aaa gag cag ctg act gag aag ctt ccc gac gca Ala Ser Ala Lys Gly Lys Glu Gln Leu Thr Glu Lys Leu Pro Asp Ala 755 760 765			2304
act aat gct caa gtt ctt cgg cgt tca gag cag tgt gaa tac ctc gat Thr Asn Ala Gln Val Leu Arg Arg Ser Glu Gln Cys Glu Tyr Leu Asp 770 775 780			2352
tac ctc aat att gcc cct gac tct ggg ctg cat ggg aag ggc tac gcc Tyr Leu Asn Ile Ala Pro Asp Ser Gly Leu His Gly Lys Gly Tyr Ala 785 790 795 800			2400
cac gga cac aaa gac gtt aac ccg caa gac tgg ttc ttc tct tgc cac His Gly His Lys Asp Val Asn Pro Gln Asp Trp Phe Phe Ser Cys His 805 810 815			2448
ttt tgg ttc gat cct gta atg cca gga tct tta gga att gaa tca atg Phe Trp Phe Asp Pro Val Met Pro Gly Ser Leu Gly Ile Glu Ser Met 820 825 830			2496
ttc cag ctt atc gag gcc ttt gcg gtg gac caa aac att cct gga gag Phe Gln Leu Ile Glu Ala Phe Ala Val Asp Gln Asn Ile Pro Gly Glu 835 840 845			2544
tac aac gta tcc aat ccg acc ttt gcc cat gca cca ggc aaa acg gcg Tyr Asn Val Ser Asn Pro Thr Phe Ala His Ala Pro Gly Lys Thr Ala 850 855 860			2592
tgg aaa tac cga ggc cag ctc aca cca aag aac cgt gcg atg gac tgc Trp Lys Tyr Arg Gly Gln Leu Thr Pro Lys Asn Arg Ala Met Asp Cys 865 870 875 880			2640
gag gtg cat atc gtt tca att acc gcc tcc ccc gag aac ggg ggc tac Glu Val His Ile Val Ser Ile Thr Ala Ser Pro Glu Asn Gly Gly Tyr 885 890 895			2688
gtt gac atc gtg gcc gat gga gcg ctt tgg gta gat gga ctt cgc gtg Val Asp Ile Val Ala Asp Gly Ala Leu Trp Val Asp Gly Leu Arg Val 900 905 910			2736
tac gaa gcc aaa gag ctt cga gtt cgt gtc gtt tcg gca aaa cct caa Tyr Glu Ala Lys Glu Leu Arg Val Arg Val Val Ser Ala Lys Pro Gln			2784

915	920	925	
gca att ccg gat gta caa caa cag cca cct agc gca aag gcg gac ccg Ala Ile Pro Asp Val Gln Gln Gln Pro Pro Ser Ala Lys Ala Asp Pro 930 935 940			2832
ggg aaa aca gga gtt gca ctt tcg ccc act cag cta cgc gac gtc ctg Gly Lys Thr Gly Val Ala Leu Ser Pro Thr Gln Leu Arg Asp Val Leu 945 950 955 960			2880
ctt gaa gtg gac aat cca ttg tat ctt ggt gta gag aac tcc aat ttg Leu Glu Val Asp Asn Pro Leu Tyr Leu Gly Val Glu Asn Ser Asn Leu 965 970 975			2928
gtg cag ttt gag tcg aaa cct gca act tct tca cgt atc gtt tcg atc Val Gln Phe Glu Ser Lys Pro Ala Thr Ser Ser Arg Ile Val Ser Ile 980 985 990			2976
aaa ccg tgc tcg att agt gac ctt ggc gat aag tct ttt atg gaa acg Lys Pro Cys Ser Ile Ser Asp Leu Gly Asp Lys Ser Phe Met Glu Thr 995 1000 1005			3024
tac aac gtg tca gca cct ctg tat act gga gca atg gcc aag ggc Tyr Asn Val Ser Ala Pro Leu Tyr Thr Gly Ala Met Ala Lys Gly 1010 1015 1020			3069
att gca tcc gcc gac ttg gtc att gct gct ggg aaa cgc aag ata Ile Ala Ser Ala Asp Leu Val Ile Ala Ala Gly Lys Arg Lys Ile 1025 1030 1035			3114
ctt gga tcg ttt ggt gcg gga ggg ctg cct att tcc ata gtc cgt Leu Gly Ser Phe Gly Ala Gly Gly Leu Pro Ile Ser Ile Val Arg 1040 1045 1050			3159
gaa gca ctg gag aaa att caa caa cac ctg ccc cac ggc ccc tac Glu Ala Leu Glu Lys Ile Gln Gln His Leu Pro His Gly Pro Tyr 1055 1060 1065			3204
gct gtt aac ctc att cac tcg cct ttc gac agc aac ttg gaa aag Ala Val Asn Leu Ile His Ser Pro Phe Asp Ser Asn Leu Glu Lys 1070 1075 1080			3249
ggc aac gtt gac ctc ttt ctc gag atg ggc gtg aca gtg gta gaa Gly Asn Val Asp Leu Phe Leu Glu Met Gly Val Thr Val Val Glu			3294

1085	1090	1095	
tgc agc gcg ttc atg gaa ctc	acg gcc cag gtt gtc	cgg tac cgc	3339
Cys Ser Ala Phe Met Glu Leu	Thr Ala Gln Val Val	Arg Tyr Arg	
1100	1105	1110	
gcg tct ggt cta agc aaa agt	gcg gac ggt tcg att	cgc att gct	3384
Ala Ser Gly Leu Ser Lys Ser	Ala Asp Gly Ser Ile	Arg Ile Ala	
1115	1120	1125	
cac cgt att att ggc aag gtt	tcc aga acc gag ctg	gca gaa atg	3429
His Arg Ile Ile Gly Lys Val	Ser Arg Thr Glu Leu	Ala Glu Met	
1130	1135	1140	
ttt att cgt cca gca cca cag	cac ctc ctc caa aaa	ctc gta gcc	3474
Phe Ile Arg Pro Ala Pro Gln	His Leu Leu Gln Lys	Leu Val Ala	
1145	1150	1155	
tcc ggc gag ctg aca gct gag	caa gcc gag ctt gca	aca cag gtt	3519
Ser Gly Glu Leu Thr Ala Glu	Gln Ala Glu Leu Ala	Thr Gln Val	
1160	1165	1170	
ccg gtg gcg gat gac att gcg	gtc gaa gcc gac tcg	ggg ggg cat	3564
Pro Val Ala Asp Asp Ile Ala	Val Glu Ala Asp Ser	Gly Gly His	
1175	1180	1185	
acc gac aac agg cct att cac	gtc att ctt cct cta	atc atc aac	3609
Thr Asp Asn Arg Pro Ile His	Val Ile Leu Pro Leu	Ile Ile Asn	
1190	1195	1200	
cta cgc aac cgt ttg cat aaa	gag ctt gac tac cct	tcg cat ctc	3654
Leu Arg Asn Arg Leu His Lys	Glu Leu Asp Tyr Pro	Ser His Leu	
1205	1210	1215	
cgg gta cgt gtg ggt gct ggt	ggt ggt att gga tgt	cct caa gcc	3699
Arg Val Arg Val Gly Ala Gly	Gly Gly Ile Gly Cys	Pro Gln Ala	
1220	1225	1230	
gct ctt gca gca ttt caa atg	ggg gca gcg ttt tta	atc act gga	3744
Ala Leu Ala Ala Phe Gln Met	Gly Ala Ala Phe Leu	Ile Thr Gly	
1235	1240	1245	
acg gtg aac cag ctt gct cgt	gaa agt ggc act tgt	gac aac gtc	3789
Thr Val Asn Gln Leu Ala Arg	Glu Ser Gly Thr Cys	Asp Asn Val	

1250	1255	1260	
cgg tta cag ctc tca aag gcc acg tat agc gac gtg tgt atg gct			3834
Arg Leu Gln Leu Ser Lys Ala Thr Tyr Ser Asp Val Cys Met Ala			
1265	1270	1275	
cct gct gcc gat atg ttt gac caa ggc gtg gag ctg caa gta ttg			3879
Pro Ala Ala Asp Met Phe Asp Gln Gly Val Glu Leu Gln Val Leu			
1280	1285	1290	
aag aaa ggc acg ctg ttc cca agt cgt gct aag aag ctg tac gag			3924
Lys Lys Gly Thr Leu Phe Pro Ser Arg Ala Lys Lys Leu Tyr Glu			
1295	1300	1305	
ctg ttc tgc aag tat gac tcg ttt gag gca atg ccg gct gaa gaa			3969
Leu Phe Cys Lys Tyr Asp Ser Phe Glu Ala Met Pro Ala Glu Glu			
1310	1315	1320	
ttg caa cgg gtt gaa aag cgg att ttt caa aag tcg ctt gct gaa			4014
Leu Gln Arg Val Glu Lys Arg Ile Phe Gln Lys Ser Leu Ala Glu			
1325	1330	1335	
gtt tgg cag gag acc agt gac ttt tac att cat cgt atc aag aac			4059
Val Trp Gln Glu Thr Ser Asp Phe Tyr Ile His Arg Ile Lys Asn			
1340	1345	1350	
cct gag aaa atc aat cgt gct gca agc gat ggc aaa ctg aaa atg			4104
Pro Glu Lys Ile Asn Arg Ala Ala Ser Asp Gly Lys Leu Lys Met			
1355	1360	1365	
tcg ctt tgc ttt cgc tgg tac ctt ggg ctt tcc tca ttt tgg gcc			4149
Ser Leu Cys Phe Arg Trp Tyr Leu Gly Leu Ser Ser Phe Trp Ala			
1370	1375	1380	
aac tct ggg gca caa gat cgc gtc atg gac tat caa att tgg tgt			4194
Asn Ser Gly Ala Gln Asp Arg Val Met Asp Tyr Gln Ile Trp Cys			
1385	1390	1395	
ggc cct gct att ggc gct ttc aat gat ttt acc aag ggc acg tac			4239
Gly Pro Ala Ile Gly Ala Phe Asn Asp Phe Thr Lys Gly Thr Tyr			
1400	1405	1410	
ctt gac gtg act gtt gca aag agt tac cct tgt gtg gca cag atc			4284
Leu Asp Val Thr Val Ala Lys Ser Tyr Pro Cys Val Ala Gln Ile			

1415

1420

1425

aat ttg caa att ttg caa gga gct gcg tat ctg aaa cgc ctt ggt 4329
Asn Leu Gln Ile Leu Gln Gly Ala Ala Tyr Leu Lys Arg Leu Gly
1430 1435 1440

gtc att cgt ttt gac cgc atg ctg ctg cag gcc gtc gat atc gac 4374
Val Ile Arg Phe Asp Arg Met Leu Leu Gln Ala Val Asp Ile Asp
1445 1450 1455

gat cct gta ttt act tac gtg ccg acc cag cca ctt 4410
Asp Pro Val Phe Thr Tyr Val Pro Thr Gln Pro Leu
1460 1465 1470

<210> 62
<211> 1470
<212> PRT
<213> Thraustochytrium sp.

<400> 62

Met Gly Pro Arg Val Ala Ser Gly Lys Val Pro Ala Trp Glu Met Ser
1 5 10 15

Lys Ser Glu Leu Cys Asp Asp Arg Thr Val Val Phe Asp Tyr Glu Glu
20 25 30

Leu Leu Glu Phe Ala Glu Gly Asp Ile Ser Lys Val Phe Gly Pro Glu
35 40 45

Phe Lys Val Val Asp Gly Phe Arg Arg Arg Val Arg Leu Pro Ala Arg
50 55 60

Glu Tyr Leu Leu Val Thr Arg Val Thr Leu Met Asp Ala Glu Val Gly
65 70 75 80

Asn Phe Arg Val Gly Ala Arg Met Val Thr Glu Tyr Asp Val Pro Val
85 90 95

Asn Gly Glu Leu Ser Glu Gly Gly Asp Val Pro Trp Ala Val Leu Val
100 105 110

Glu Ala Gly Gln Cys Asp Leu Leu Leu Ile Ser Tyr Met Gly Ile Asp
115 120 125

Phe Gln Cys Lys Gly Glu Arg Val Tyr Arg Leu Leu Asn Thr Thr Leu
130 135 140

Thr Phe Phe Gly Val Ala Lys Glu Gly Glu Thr Leu Val Tyr Asp Ile
145 150 155 160

Arg Val Thr Gly Phe Ala Lys Arg Pro Asp Gly Asp Ile Ser Met Phe
165 170 175

Phe Phe Glu Tyr Asp Cys Tyr Cys Asn Gly Lys Leu Leu Ile Glu Met
180 185 190

Arg Asp Gly Ser Ala Gly Phe Phe Thr Asp Glu Glu Leu Ala Ala Gly
195 200 205

Lys Gly Val Val Val Thr Arg Ala Gln Gln Asn Met Arg Asp Lys Ile
210 215 220

Val Arg Gln Ser Ile Glu Pro Phe Ala Leu Ala Ala Cys Thr His Lys
225 230 235 240

Thr Thr Leu Asn Glu Ser Asp Met Gln Ser Leu Val Glu Arg Asn Trp
245 250 255

Ala Asn Val Phe Gly Thr Ser Asn Lys Met Ala Glu Leu Asn Tyr Lys
260 265 270

Ile Cys Ala Arg Lys Met Leu Met Ile Asp Arg Val Thr His Ile Asp
275 280 285

His His Gly Gly Ala Tyr Gly Leu Gly Leu Leu Val Gly Glu Lys Ile
290 295 300

Leu Asp Arg Asn His Trp Tyr Phe Pro Cys His Phe Val Asn Asp Gln
305 310 315 320

Val Met Ala Gly Ser Leu Val Ser Asp Gly Cys Ser Gln Leu Leu Lys
325 330 335

Leu Tyr Met Ile Trp Leu Gly Leu His Leu Lys Met Glu Glu Phe Asp
340 345 350

Phe Leu Pro Val Ser Gly His Lys Asn Lys Val Arg Cys Arg Gly Gln
355 360 365

Ile Ser Pro His Lys Gly Lys Leu Val Tyr Val Met Glu Ile Lys Lys
370 375 380

Met Gly Tyr Asp Gln Ala Ser Gly Ser Pro Tyr Ala Ile Ala Asp Val
385 390 395 400

Asp Ile Ile Asp Val Asn Glu Glu Leu Gly Gln Ser Phe Asp Ile Asn
405 410 415

Asp Leu Ala Ser Tyr Gly Lys Gly Asp Leu Ser Lys Lys Ile Val Val
420 425 430

Asp Phe Lys Gly Ile Ala Leu Gln Leu Lys Gly Arg Ala Phe Ser Arg
435 440 445

Met Ser Ser Ser Ser Ser Leu Asn Glu Gly Trp Gln Cys Val Pro Lys
450 455 460

Pro Ser Gln Arg Met Glu His Glu Gln Pro Pro Ala His Cys Leu Ala
465 470 475 480

Ser Asp Pro Glu Ala Pro Ser Thr Val Thr Trp His Pro Met Ser Lys
485 490 495

Leu Pro Gly Asn Pro Thr Pro Phe Phe Ser Pro Ser Ser Tyr Pro Pro
500 505 510

Arg Ala Ile Cys Phe Ile Pro Phe Pro Gly Asn Pro Leu Asp Asn Asn
515 520 525

Cys Lys Ala Gly Glu Met Pro Leu Asn Trp Tyr Asn Met Ser Glu Phe
530 535 540

Met Cys Gly Lys Val Ser Asn Cys Leu Gly Pro Glu Phe Ala Arg Phe
545 550 555 560

Asp Lys Ser Asn Thr Ser Arg Ser Pro Ala Phe Asp Leu Ala Leu Val
565 570 575

Thr Arg Val Val Glu Val Thr Asn Met Glu His Gly Lys Phe Leu Asn
580 585 590

Val Asp Cys Asn Pro Ser Lys Gly Thr Met Val Gly Glu Phe Asp Cys
595 600 605

Pro Gln Asp Ala Trp Phe Phe Asp Gly Ser Cys Asn Asp Gly His Met
610 615 620

Pro Tyr Ser Ile Ile Met Glu Ile Gly Leu Gln Thr Ser Gly Val Leu
625 630 635 640

Thr Ser Val Leu Lys Ala Pro Leu Thr Met Asp Lys Asp Asp Ile Leu
645 650 655

Phe Arg Asn Leu Asp Ala Ser Ala Glu Met Val Arg Pro Asp Val Asp
660 665 670

Val Arg Gly Lys Thr Ile Arg Asn Val Thr Lys Cys Thr Gly Tyr Ala

675	680	685
Met Leu Gly Lys Met Gly Ile His Arg Phe Thr Phe Glu Leu Ser Val		
690	695	700
Asp Gly Val Val Phe Tyr Lys Gly Ser Thr Ser Phe Gly Trp Phe Thr		
705	710	715
Pro Glu Val Phe Ala Gln Gln Ala Gly Leu Asp Asn Gly Lys Lys Thr		
725	730	735
Glu Pro Trp Cys Lys Thr Asn Asn Thr Ser Val Arg Arg Val Glu Ile		
740	745	750
Ala Ser Ala Lys Gly Lys Glu Gln Leu Thr Glu Lys Leu Pro Asp Ala		
755	760	765
Thr Asn Ala Gln Val Leu Arg Arg Ser Glu Gln Cys Glu Tyr Leu Asp		
770	775	780
Tyr Leu Asn Ile Ala Pro Asp Ser Gly Leu His Gly Lys Gly Tyr Ala		
785	790	795
His Gly His Lys Asp Val Asn Pro Gln Asp Trp Phe Phe Ser Cys His		
805	810	815
Phe Trp Phe Asp Pro Val Met Pro Gly Ser Leu Gly Ile Glu Ser Met		
820	825	830
Phe Gln Leu Ile Glu Ala Phe Ala Val Asp Gln Asn Ile Pro Gly Glu		
835	840	845
Tyr Asn Val Ser Asn Pro Thr Phe Ala His Ala Pro Gly Lys Thr Ala		
850	855	860
Trp Lys Tyr Arg Gly Gln Leu Thr Pro Lys Asn Arg Ala Met Asp Cys		

865											870											875											880
Glu	Val	His	Ile	Val	Ser	Ile	Thr	Ala	Ser	Pro	Glu	Asn	Gly	Gly	Tyr																		
				885					890								895																
Val	Asp	Ile	Val	Ala	Asp	Gly	Ala	Leu	Trp	Val	Asp	Gly	Leu	Arg	Val																		
				900					905								910																
Tyr	Glu	Ala	Lys	Glu	Leu	Arg	Val	Arg	Val	Val	Ser	Ala	Lys	Pro	Gln																		
				915					920								925																
Ala	Ile	Pro	Asp	Val	Gln	Gln	Gln	Pro	Pro	Ser	Ala	Lys	Ala	Asp	Pro																		
				930					935								940																
Gly	Lys	Thr	Gly	Val	Ala	Leu	Ser	Pro	Thr	Gln	Leu	Arg	Asp	Val	Leu																		
945								950								955								960									
Leu	Glu	Val	Asp	Asn	Pro	Leu	Tyr	Leu	Gly	Val	Glu	Asn	Ser	Asn	Leu																		
				965								970								975													
Val	Gln	Phe	Glu	Ser	Lys	Pro	Ala	Thr	Ser	Ser	Arg	Ile	Val	Ser	Ile																		
				980								985								990													
Lys	Pro	Cys	Ser	Ile	Ser	Asp	Leu	Gly	Asp	Lys	Ser	Phe	Met	Glu	Thr																		
				995				1000								1005																	
Tyr	Asn	Val	Ser	Ala	Pro	Leu	Tyr	Thr	Gly	Ala	Met	Ala	Lys	Gly																			
1010								1015								1020																	
Ile	Ala	Ser	Ala	Asp	Leu	Val	Ile	Ala	Ala	Gly	Lys	Arg	Lys	Ile																			
1025								1030								1035																	
Leu	Gly	Ser	Phe	Gly	Ala	Gly	Gly	Leu	Pro	Ile	Ser	Ile	Val	Arg																			
1040								1045								1050																	
Glu	Ala	Leu	Glu	Lys	Ile	Gln	Gln	His	Leu	Pro	His	Gly	Pro	Tyr																			

1055	1060	1065
Ala Val Asn Leu Ile His Ser Pro Phe Asp Ser Asn Leu Glu Lys		
1070	1075	1080
Gly Asn Val Asp Leu Phe Leu Glu Met Gly Val Thr Val Val Glu		
1085	1090	1095
Cys Ser Ala Phe Met Glu Leu Thr Ala Gln Val Val Arg Tyr Arg		
1100	1105	1110
Ala Ser Gly Leu Ser Lys Ser Ala Asp Gly Ser Ile Arg Ile Ala		
1115	1120	1125
His Arg Ile Ile Gly Lys Val Ser Arg Thr Glu Leu Ala Glu Met		
1130	1135	1140
Phe Ile Arg Pro Ala Pro Gln His Leu Leu Gln Lys Leu Val Ala		
1145	1150	1155
Ser Gly Glu Leu Thr Ala Glu Gln Ala Glu Leu Ala Thr Gln Val		
1160	1165	1170
Pro Val Ala Asp Asp Ile Ala Val Glu Ala Asp Ser Gly Gly His		
1175	1180	1185
Thr Asp Asn Arg Pro Ile His Val Ile Leu Pro Leu Ile Ile Asn		
1190	1195	1200
Leu Arg Asn Arg Leu His Lys Glu Leu Asp Tyr Pro Ser His Leu		
1205	1210	1215
Arg Val Arg Val Gly Ala Gly Gly Gly Ile Gly Cys Pro Gln Ala		
1220	1225	1230

Ala Leu Ala Ala Phe Gln Met Gly Ala Ala Phe Leu Ile Thr Gly
1235 1240 1245

Thr Val Asn Gln Leu Ala Arg Glu Ser Gly Thr Cys Asp Asn Val
1250 1255 1260

Arg Leu Gln Leu Ser Lys Ala Thr Tyr Ser Asp Val Cys Met Ala
1265 1270 1275

Pro Ala Ala Asp Met Phe Asp Gln Gly Val Glu Leu Gln Val Leu
1280 1285 1290

Lys Lys Gly Thr Leu Phe Pro Ser Arg Ala Lys Lys Leu Tyr Glu
1295 1300 1305

Leu Phe Cys Lys Tyr Asp Ser Phe Glu Ala Met Pro Ala Glu Glu
1310 1315 1320

Leu Gln Arg Val Glu Lys Arg Ile Phe Gln Lys Ser Leu Ala Glu
1325 1330 1335

Val Trp Gln Glu Thr Ser Asp Phe Tyr Ile His Arg Ile Lys Asn
1340 1345 1350

Pro Glu Lys Ile Asn Arg Ala Ala Ser Asp Gly Lys Leu Lys Met
1355 1360 1365

Ser Leu Cys Phe Arg Trp Tyr Leu Gly Leu Ser Ser Phe Trp Ala
1370 1375 1380

Asn Ser Gly Ala Gln Asp Arg Val Met Asp Tyr Gln Ile Trp Cys
1385 1390 1395

Gly Pro Ala Ile Gly Ala Phe Asn Asp Phe Thr Lys Gly Thr Tyr
1400 1405 1410

Leu Asp Val Thr Val Ala Lys Ser Tyr Pro Cys Val Ala Gln Ile
1415 1420 1425

Asn Leu Gln Ile Leu Gln Gly Ala Ala Tyr Leu Lys Arg Leu Gly
1430 1435 1440

Val Ile Arg Phe Asp Arg Met Leu Leu Gln Ala Val Asp Ile Asp
1445 1450 1455

Asp Pro Val Phe Thr Tyr Val Pro Thr Gln Pro Leu
1460 1465 1470

<210> 63
<211> 1500
<212> DNA
<213> Thraustochytrium sp.

<220>
<221> CDS
<222> (1)..(1500)

<400> 63
atg ggc ccg cga gtg gcg tca ggc aag gtg ccg gct tgg gag atg agc 48
Met Gly Pro Arg Val Ala Ser Gly Lys Val Pro Ala Trp Glu Met Ser
1 5 10 15

aag tcc gag ctg tgt gat gac cgc acg gta gtc ttt gac tat gag gag 96
Lys Ser Glu Leu Cys Asp Asp Arg Thr Val Val Phe Asp Tyr Glu Glu
20 25 30

ctg ctg gag ttc gct gag ggc gat atc agt aag gtt ttt ggg ccg gag 144
Leu Leu Glu Phe Ala Glu Gly Asp Ile Ser Lys Val Phe Gly Pro Glu
35 40 45

ttc aaa gtg gtg gac ggg ttt agg cgc agg gtg agg ttg ccc gct cga 192
Phe Lys Val Val Asp Gly Phe Arg Arg Arg Val Arg Leu Pro Ala Arg
50 55 60

gag tac ctg ctg gtg acc cgg gtt acg ctg atg gat gcc gag gtg ggc 240
Glu Tyr Leu Leu Val Thr Arg Val Thr Leu Met Asp Ala Glu Val Gly

65	70	75	80	
aac ttt cga gtg gga gca cgt atg gtg aca gag tat gac gta cct gtg				288
Asn Phe Arg Val Gly Ala Arg Met Val Thr Glu Tyr Asp Val Pro Val				
85	90	95		
aac gga gag ctc tcg gaa ggg gga gat gtg ccg tgg gct gtg ttg gtg				336
Asn Gly Glu Leu Ser Glu Gly Gly Asp Val Pro Trp Ala Val Leu Val				
100	105	110		
gaa gcc ggg cag tgc gac ttg ctg cta att tct tac atg ggc atc gat				384
Glu Ala Gly Gln Cys Asp Leu Leu Leu Ile Ser Tyr Met Gly Ile Asp				
115	120	125		
ttc cag tgc aaa gga gag cgg gtc tac cgg ctg ctg aac acc acc ttg				432
Phe Gln Cys Lys Gly Glu Arg Val Tyr Arg Leu Leu Asn Thr Thr Leu				
130	135	140		
acg ttt ttt ggc gtc gcg aaa gaa ggg gaa acg ctt gtg tac gat att				480
Thr Phe Phe Gly Val Ala Lys Glu Gly Glu Thr Leu Val Tyr Asp Ile				
145	150	155	160	
cgc gtc acg ggt ttc gcc aag agg ccg gac gga gat atc tcc atg ttc				528
Arg Val Thr Gly Phe Ala Lys Arg Pro Asp Gly Asp Ile Ser Met Phe				
165	170	175		
ttt ttc gaa tat gat tgc tac tgc aat ggc aag ctt ctc atc gaa atg				576
Phe Phe Glu Tyr Asp Cys Tyr Cys Asn Gly Lys Leu Leu Ile Glu Met				
180	185	190		
cga gat ggc tct gca ggc ttc ttc acg gac gaa gag ctc gct gcc ggc				624
Arg Asp Gly Ser Ala Gly Phe Phe Thr Asp Glu Glu Leu Ala Ala Gly				
195	200	205		
aaa gga gtg gtc gtc act cgt gca cag caa aac atg cgg gac aaa att				672
Lys Gly Val Val Val Thr Arg Ala Gln Gln Asn Met Arg Asp Lys Ile				
210	215	220		
gta cgg cag tcc att gag cct ttt gca ctg gcg gct tgc acg cac aaa				720
Val Arg Gln Ser Ile Glu Pro Phe Ala Leu Ala Ala Cys Thr His Lys				
225	230	235	240	
acg act ctg aac gag agt gac atg cag tcc ctt gtg gag cga aac tgg				768
Thr Thr Leu Asn Glu Ser Asp Met Gln Ser Leu Val Glu Arg Asn Trp				

245	250	255	
gca aac gtt ttt ggc acc agt aac aag atg gcg gag ctc aac tat aaa			816
Ala Asn Val Phe Gly Thr Ser Asn Lys Met Ala Glu Leu Asn Tyr Lys			
260	265	270	
att tgc gcc agg aaa atg ctc atg atc gac agg gtt acc cac att gac			864
Ile Cys Ala Arg Lys Met Leu Met Ile Asp Arg Val Thr His Ile Asp			
275	280	285	
cac cac ggt ggg gcg tat ggc ctc gga cta ctt gtt gga gag aag atc			912
His His Gly Gly Ala Tyr Gly Leu Gly Leu Leu Val Gly Glu Lys Ile			
290	295	300	
ttg gat cga aac cat tgg tac ttt cct tgt cac ttt gtc aat gat caa			960
Leu Asp Arg Asn His Trp Tyr Phe Pro Cys His Phe Val Asn Asp Gln			
305	310	315	320
gtc atg gca ggg tca ctg gtc agc gat ggt tgc agc cag ctc tta aaa			1008
Val Met Ala Gly Ser Leu Val Ser Asp Gly Cys Ser Gln Leu Leu Lys			
325	330	335	
ctc tat atg atc tgg ctt ggc ctc cac ctg aaa atg gag gaa ttt gat			1056
Leu Tyr Met Ile Trp Leu Gly Leu His Leu Lys Met Glu Glu Phe Asp			
340	345	350	
ttt ctc cca gtt agc ggc cac aaa aac aag gtg cga tgc agg gga caa			1104
Phe Leu Pro Val Ser Gly His Lys Asn Lys Val Arg Cys Arg Gly Gln			
355	360	365	
att tca ccg cat aaa ggc aag ctt gtc tac gtc atg gaa atc aaa aag			1152
Ile Ser Pro His Lys Gly Lys Leu Val Tyr Val Met Glu Ile Lys Lys			
370	375	380	
atg ggt tac gat caa gca tct gga agc cca tac gcc atc gcg gac gtt			1200
Met Gly Tyr Asp Gln Ala Ser Gly Ser Pro Tyr Ala Ile Ala Asp Val			
385	390	395	400
gat atc att gac gtc aac gaa gag ctg ggt caa agt ttt gac atc aac			1248
Asp Ile Ile Asp Val Asn Glu Glu Leu Gly Gln Ser Phe Asp Ile Asn			
405	410	415	
gac ctt gcg agc tac gga aaa ggt gac ctg agc aaa aaa atc gtg gtt			1296
Asp Leu Ala Ser Tyr Gly Lys Gly Asp Leu Ser Lys Lys Ile Val Val			

[illegible]

<210>	64
<211>	500
<212>	PRT
<213>	Thraustochytrium sp.

<400> 64

Met Gly Pro Arg Val Ala Ser Gly Lys Val Pro Ala Trp Glu Met Ser
1 5 10 15

Lys Ser Glu Leu Cys Asp Asp Arg Thr Val Val Phe Asp Tyr Glu Glu
20 25 30

Leu Leu Glu Phe Ala Glu Gly Asp Ile Ser Lys Val Phe Gly Pro Glu
35 40 45

Phe Lys Val Val Asp Gly Phe Arg Arg Arg Val Arg Leu Pro Ala Arg
50 55 60

Glu Tyr Leu Leu Val Thr Arg Val Thr Leu Met Asp Ala Glu Val Gly
65 70 75 80

Asn Phe Arg Val Gly Ala Arg Met Val Thr Glu Tyr Asp Val Pro Val
85 90 95

Asn Gly Glu Leu Ser Glu Gly Gly Asp Val Pro Trp Ala Val Leu Val
100 105 110

Glu Ala Gly Gln Cys Asp Leu Leu Leu Ile Ser Tyr Met Gly Ile Asp
115 120 125

Phe Gln Cys Lys Gly Glu Arg Val Tyr Arg Leu Leu Asn Thr Thr Leu
130 135 140

Thr Phe Phe Gly Val Ala Lys Glu Gly Glu Thr Leu Val Tyr Asp Ile
145 150 155 160

Arg Val Thr Gly Phe Ala Lys Arg Pro Asp Gly Asp Ile Ser Met Phe
165 170 175

Phe Phe Glu Tyr Asp Cys Tyr Cys Asn Gly Lys Leu Leu Ile Glu Met
180 185 190

Arg Asp Gly Ser Ala Gly Phe Phe Thr Asp Glu Glu Leu Ala Ala Gly
195 200 205

Lys Gly Val Val Val Thr Arg Ala Gln Gln Asn Met Arg Asp Lys Ile
210 215 220

Val Arg Gln Ser Ile Glu Pro Phe Ala Leu Ala Ala Cys Thr His Lys
225 230 235 240

Thr Thr Leu Asn Glu Ser Asp Met Gln Ser Leu Val Glu Arg Asn Trp
245 250 255

Ala Asn Val Phe Gly Thr Ser Asn Lys Met Ala Glu Leu Asn Tyr Lys
260 265 270

Ile Cys Ala Arg Lys Met Leu Met Ile Asp Arg Val Thr His Ile Asp
275 280 285

His His Gly Gly Ala Tyr Gly Leu Gly Leu Leu Val Gly Glu Lys Ile
290 295 300

Leu Asp Arg Asn His Trp Tyr Phe Pro Cys His Phe Val Asn Asp Gln
305 310 315 320

Val Met Ala Gly Ser Leu Val Ser Asp Gly Cys Ser Gln Leu Leu Lys
325 330 335

Leu Tyr Met Ile Trp Leu Gly Leu His Leu Lys Met Glu Glu Phe Asp
340 345 350

Phe Leu Pro Val Ser Gly His Lys Asn Lys Val Arg Cys Arg Gly Gln
355 360 365

Ile Ser Pro His Lys Gly Lys Leu Val Tyr Val Met Glu Ile Lys Lys
370 375 380

Met Gly Tyr Asp Gln Ala Ser Gly Ser Pro Tyr Ala Ile Ala Asp Val
385 390 395 400

Asp Ile Ile Asp Val Asn Glu Glu Leu Gly Gln Ser Phe Asp Ile Asn
405 410 415

Asp Leu Ala Ser Tyr Gly Lys Gly Asp Leu Ser Lys Lys Ile Val Val
420 425 430

Asp Phe Lys Gly Ile Ala Leu Gln Leu Lys Gly Arg Ala Phe Ser Arg
435 440 445

Met Ser Ser Ser Ser Ser Leu Asn Glu Gly Trp Gln Cys Val Pro Lys
450 455 460

Pro Ser Gln Arg Met Glu His Glu Gln Pro Pro Ala His Cys Leu Ala
465 470 475 480

Ser Asp Pro Glu Ala Pro Ser Thr Val Thr Trp His Pro Met Ser Lys
485 490 495

Leu Pro Gly Asn
500

<210> 65
<211> 1500
<212> DNA
<213> Thraustochytrium sp.

<220>
<221> CDS
<222> (1)..(1500)

<400> 65
cct acg ccg ttc ttc tcc cct tca tct tac cct ccg agg gca att tgc 48
Pro Thr Pro Phe Phe Ser Pro Ser Ser Tyr Pro Pro Arg Ala Ile Cys
1 5 10 15

ttc atc cct ttc ccg ggc aat ccc ctt gac aac aac tgc aag gct gga 96
Phe Ile Pro Phe Pro Gly Asn Pro Leu Asp Asn Asn Cys Lys Ala Gly
20 25 30

gaa atg ccc ctg aac tgg tac aac atg tca gag ttc atg tgt ggc aag 144
Glu Met Pro Leu Asn Trp Tyr Asn Met Ser Glu Phe Met Cys Gly Lys
35 40 45

gtt tct aac tgc ttg ggc cca gaa ttc gca cgc ttt gac aag tcg aac 192
Val Ser Asn Cys Leu Gly Pro Glu Phe Ala Arg Phe Asp Lys Ser Asn
50 55 60

acc agc cgg agc cct gct ttt gac ttg gct ctg gtg acc cga gtt gtt 240
Thr Ser Arg Ser Pro Ala Phe Asp Leu Ala Leu Val Thr Arg Val Val
65 70 75 80

gaa gtc aca aac atg gaa cac ggc aag ttt cta aac gtt gat tgc aat Glu Val Thr Asn Met Glu His Gly Lys Phe Leu Asn Val Asp Cys Asn 85 90 95	288
cca agc aaa ggc aca atg gtg ggg gag ttt gac tgt ccc caa gac gcg Pro Ser Lys Gly Thr Met Val Gly Glu Phe Asp Cys Pro Gln Asp Ala 100 105 110	336
tgg ttc ttt gat ggt tgc tgc aac gac ggc cat atg ccg tat tcc att Trp Phe Phe Asp Gly Ser Cys Asn Asp Gly His Met Pro Tyr Ser Ile 115 120 125	384
atc atg gaa atc gga ctg caa acc tca ggt gtt ctc acc tcg gtg ttg Ile Met Glu Ile Gly Leu Gln Thr Ser Gly Val Leu Thr Ser Val Leu 130 135 140	432
aag gca ccg ctg act atg gac aag gat gac att ctc ttt cga aac ctc Lys Ala Pro Leu Thr Met Asp Lys Asp Asp Ile Leu Phe Arg Asn Leu 145 150 155 160	480
gat gca agt gct gaa atg gtg cgt cca gac gtg gat gtt cgc ggc aaa Asp Ala Ser Ala Glu Met Val Arg Pro Asp Val Asp Val Arg Gly Lys 165 170 175	528
acg att cga aac gtg acc aag tgt acc ggc tat gca atg ttg gga aag Thr Ile Arg Asn Val Thr Lys Cys Thr Gly Tyr Ala Met Leu Gly Lys 180 185 190	576
atg ggg att cac cgg ttc acg ttt gag ttg agc gtt gac ggc gtg gta Met Gly Ile His Arg Phe Thr Phe Glu Leu Ser Val Asp Gly Val Val 195 200 205	624
ttt tat aaa gga tcc act tcc ttt gga tgg ttc act ccc gag gtg ttt Phe Tyr Lys Gly Ser Thr Ser Phe Gly Trp Phe Thr Pro Glu Val Phe 210 215 220	672
gct cag caa gct gga ctc gac aac ggg aaa aag acg gag ccc tgg tgc Ala Gln Gln Ala Gly Leu Asp Asn Gly Lys Lys Thr Glu Pro Trp Cys 225 230 235 240	720
aag act aac aac acc tcg gtt cga aga gtt gaa atc gca tcc gcc aaa Lys Thr Asn Asn Thr Ser Val Arg Arg Val Glu Ile Ala Ser Ala Lys 245 250 255	768

gga aaa gag cag ctg act gag aag ctt ccc gac gca act aat gct caa Gly Lys Glu Gln Leu Thr Glu Lys Leu Pro Asp Ala Thr Asn Ala Gln 260 265 270	816
gtt ctt cgg cgt tca gag cag tgt gaa tac ctc gat tac ctc aat att Val Leu Arg Arg Ser Glu Gln Cys Glu Tyr Leu Asp Tyr Leu Asn Ile 275 280 285	864
gcc cct gac tct ggg ctg cat ggg aag ggc tac gcc cac gga cac aaa Ala Pro Asp Ser Gly Leu His Gly Lys Gly Tyr Ala His Gly His Lys 290 295 300	912
gac gtt aac ccg caa gac tgg ttc ttc tct tgc cac ttt tgg ttc gat Asp Val Asn Pro Gln Asp Trp Phe Phe Ser Cys His Phe Trp Phe Asp 305 310 315 320	960
cct gta atg cca gga tct tta gga att gaa tca atg ttc cag ctt atc Pro Val Met Pro Gly Ser Leu Gly Ile Glu Ser Met Phe Gln Leu Ile 325 330 335	1008
gag gcc ttt gcg gtg gac caa aac att cct gga gag tac aac gta tcc Glu Ala Phe Ala Val Asp Gln Asn Ile Pro Gly Glu Tyr Asn Val Ser 340 345 350	1056
aat ccg acc ttt gcc cat gca cca ggc aaa acg gcg tgg aaa tac cga Asn Pro Thr Phe Ala His Ala Pro Gly Lys Thr Ala Trp Lys Tyr Arg 355 360 365	1104
ggc cag ctc aca cca aag aac cgt gcg atg gac tgc gag gtg cat atc Gly Gln Leu Thr Pro Lys Asn Arg Ala Met Asp Cys Glu Val His Ile 370 375 380	1152
gtt tca att acc gcc tcc ccc gag aac ggg ggc tac gtt gac atc gtg Val Ser Ile Thr Ala Ser Pro Glu Asn Gly Gly Tyr Val Asp Ile Val 385 390 395 400	1200
gcc gat gga gcg ctt tgg gta gat gga ctt cgc gtg tac gaa gcc aaa Ala Asp Gly Ala Leu Trp Val Asp Gly Leu Arg Val Tyr Glu Ala Lys 405 410 415	1248
gag ctt cga gtt cgt gtc gtt tcg gca aaa cct caa gca att ccg gat Glu Leu Arg Val Arg Val Val Ser Ala Lys Pro Gln Ala Ile Pro Asp 420 425 430	1296

gta caa caa cag cca cct agc gca aag gcg gac ccg ggg aaa aca gga 1344
Val Gln Gln Gln Pro Pro Ser Ala Lys Ala Asp Pro Gly Lys Thr Gly
435 440 445

gtt gca ctt tcg ccc act cag cta cgc gac gtc ctg ctt gaa gtg gac 1392
Val Ala Leu Ser Pro Thr Gln Leu Arg Asp Val Leu Leu Glu Val Asp
450 455 460

aat cca ttg tat ctt ggt gta gag aac tcc aat ttg gtg cag ttt gag 1440
Asn Pro Leu Tyr Leu Gly Val Glu Asn Ser Asn Leu Val Gln Phe Glu
465 470 475 480

tcg aaa cct gca act tct tca cgt atc gtt tcg atc aaa ccg tgc tcg 1488
Ser Lys Pro Ala Thr Ser Ser Arg Ile Val Ser Ile Lys Pro Cys Ser
485 490 495

att agt gac ctt 1500
Ile Ser Asp Leu
500

<210> 66
<211> 500
<212> PRT
<213> Thraustochytrium sp.

<400> 66

Pro Thr Pro Phe Phe Ser Pro Ser Ser Tyr Pro Pro Arg Ala Ile Cys
1 5 10 15

Phe Ile Pro Phe Pro Gly Asn Pro Leu Asp Asn Asn Cys Lys Ala Gly
20 25 30

Glu Met Pro Leu Asn Trp Tyr Asn Met Ser Glu Phe Met Cys Gly Lys
35 40 45

Val Ser Asn Cys Leu Gly Pro Glu Phe Ala Arg Phe Asp Lys Ser Asn
50 55 60

Thr Ser Arg Ser Pro Ala Phe Asp Leu Ala Leu Val Thr Arg Val Val
65 70 75 80

Glu Val Thr Asn Met Glu His Gly Lys Phe Leu Asn Val Asp Cys Asn
85 90 95

Pro Ser Lys Gly Thr Met Val Gly Glu Phe Asp Cys Pro Gln Asp Ala
100 105 110

Trp Phe Phe Asp Gly Ser Cys Asn Asp Gly His Met Pro Tyr Ser Ile
115 120 125

Ile Met Glu Ile Gly Leu Gln Thr Ser Gly Val Leu Thr Ser Val Leu
130 135 140

Lys Ala Pro Leu Thr Met Asp Lys Asp Asp Ile Leu Phe Arg Asn Leu
145 150 155 160

Asp Ala Ser Ala Glu Met Val Arg Pro Asp Val Asp Val Arg Gly Lys
165 170 175

Thr Ile Arg Asn Val Thr Lys Cys Thr Gly Tyr Ala Met Leu Gly Lys
180 185 190

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

Phe Tyr Lys Gly Ser Thr Ser Phe Gly Trp Phe Thr Pro Glu Val Phe
210 215 220

Ala Gln Gln Ala Gly Leu Asp Asn Gly Lys Lys Thr Glu Pro Trp Cys
225 230 235 240

Lys Thr Asn Asn Thr Ser Val Arg Arg Val Glu Ile Ala Ser Ala Lys
245 250 255

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

260 265 270

Val Leu Arg Arg Ser Glu Gln Cys Glu Tyr Leu Asp Tyr Leu Asn Ile
275 280 285

Ala Pro Asp Ser Gly Leu His Gly Lys Gly Tyr Ala His Gly His Lys
290 295 300

Asp Val Asn Pro Gln Asp Trp Phe Phe Ser Cys His Phe Trp Phe Asp
305 310 315 320

Pro Val Met Pro Gly Ser Leu Gly Ile Glu Ser Met Phe Gln Leu Ile
325 330 335

Glu Ala Phe Ala Val Asp Gln Asn Ile Pro Gly Glu Tyr Asn Val Ser
340 345 350

Asn Pro Thr Phe Ala His Ala Pro Gly Lys Thr Ala Trp Lys Tyr Arg
355 360 365

Gly Gln Leu Thr Pro Lys Asn Arg Ala Met Asp Cys Glu Val His Ile
370 375 380

Val Ser Ile Thr Ala Ser Pro Glu Asn Gly Gly Tyr Val Asp Ile Val
385 390 395 400

Ala Asp Gly Ala Leu Trp Val Asp Gly Leu Arg Val Tyr Glu Ala Lys
405 410 415

Glu Leu Arg Val Arg Val Val Ser Ala Lys Pro Gln Ala Ile Pro Asp
420 425 430

Val Gln Gln Gln Pro Pro Ser Ala Lys Ala Asp Pro Gly Lys Thr Gly
435 440 445

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

450

455

460

Asn Pro Leu Tyr Leu Gly Val Glu Asn Ser Asn Leu Val Gln Phe Glu
465 470 475 480

Ser Lys Pro Ala Thr Ser Ser Arg Ile Val Ser Ile Lys Pro Cys Ser
485 490 495

Ile Ser Asp Leu
500

<210> 67

<211> 1410

<212> DNA

<213> Thraustochytrium sp.

<220>

<221> CDS

<222> (1)..(1410)

<400> 67

ggc gat aag tct ttt atg gaa acg tac aac gtg tca gca cct ctg tat 48
Gly Asp Lys Ser Phe Met Glu Thr Tyr Asn Val Ser Ala Pro Leu Tyr
1 5 10 15

act gga gca atg gcc aag ggc att gca tcc gcc gac ttg gtc att gct 96
Thr Gly Ala Met Ala Lys Gly Ile Ala Ser Ala Asp Leu Val Ile Ala
20 25 30

gct ggg aaa cgc aag ata ctt gga tgc ttt ggt gcg gga ggg ctg cct 144
Ala Gly Lys Arg Lys Ile Leu Gly Ser Phe Gly Ala Gly Gly Leu Pro
35 40 45

att tcc ata gtc cgt gaa gca ctg gag aaa att caa caa cac ctg ccc 192
Ile Ser Ile Val Arg Glu Ala Leu Glu Lys Ile Gln Gln His Leu Pro
50 55 60

cac ggc ccc tac gct gtt aac ctc att cac tgc cct ttc gac agc aac 240
His Gly Pro Tyr Ala Val Asn Leu Ile His Ser Pro Phe Asp Ser Asn
65 70 75 80

ttg gaa aag ggc aac gtt gac ctc ttt ctc gag atg ggc gtg aca gtg 288
Leu Glu Lys Gly Asn Val Asp Leu Phe Leu Glu Met Gly Val Thr Val

85	90	95	
gta gaa tgc agc gcg ttc atg gaa ctc acg gcc cag gtt gtc cgg tac			336
Val Glu Cys Ser Ala Phe Met Glu Leu Thr Ala Gln Val Val Arg Tyr			
100	105	110	
cgc gcg tct ggt cta agc aaa agt gcg gac ggt tcg att cgc att gct			384
Arg Ala Ser Gly Leu Ser Lys Ser Ala Asp Gly Ser Ile Arg Ile Ala			
115	120	125	
cac cgt att att ggc aag gtt tcc aga acc gag ctg gca gaa atg ttt			432
His Arg Ile Ile Gly Lys Val Ser Arg Thr Glu Leu Ala Glu Met Phe			
130	135	140	
att cgt cca gca cca cag cac ctc ctc caa aaa ctc gta gcc tcc ggc			480
Ile Arg Pro Ala Pro Gln His Leu Leu Gln Lys Leu Val Ala Ser Gly			
145	150	155	160
gag ctg aca gct gag caa gcc gag ctt gca aca cag gtt ccg gtg gcg			528
Glu Leu Thr Ala Glu Gln Ala Glu Leu Ala Thr Gln Val Pro Val Ala			
165	170	175	
gat gac att gcg gtc gaa gcc gac tcg ggg ggg cat acc gac aac agg			576
Asp Asp Ile Ala Val Glu Ala Asp Ser Gly Gly His Thr Asp Asn Arg			
180	185	190	
cct att cac gtc att ctt cct cta atc atc aac cta cgc aac cgt ttg			624
Pro Ile His Val Ile Leu Pro Leu Ile Ile Asn Leu Arg Asn Arg Leu			
195	200	205	
cat aaa gag ctt gac tac cct tcg cat ctc cgg gta cgt gtg ggt gct			672
His Lys Glu Leu Asp Tyr Pro Ser His Leu Arg Val Arg Val Gly Ala			
210	215	220	
ggt ggt ggt att gga tgt cct caa gcc gct ctt gca gca ttt caa atg			720
Gly Gly Gly Ile Gly Cys Pro Gln Ala Ala Leu Ala Ala Phe Gln Met			
225	230	235	240
ggg gca gcg ttt tta atc act gga acg gtg aac cag ctt gct cgt gaa			768
Gly Ala Ala Phe Leu Ile Thr Gly Thr Val Asn Gln Leu Ala Arg Glu			
245	250	255	
agt ggc act tgt gac aac gtc cgg tta cag ctc tca aag gcc acg tat			816
Ser Gly Thr Cys Asp Asn Val Arg Leu Gln Leu Ser Lys Ala Thr Tyr			

260	265	270	
agc gac gtg tgt atg gct cct gct gcc gat atg ttt gac caa ggc gtg			864
Ser Asp Val Cys Met Ala Pro Ala Ala Asp Met Phe Asp Gln Gly Val			
275	280	285	
gag ctg caa gta ttg aag aaa ggc acg ctg ttc cca agt cgt gct aag			912
Glu Leu Gln Val Leu Lys Lys Gly Thr Leu Phe Pro Ser Arg Ala Lys			
290	295	300	
aag ctg tac gag ctg ttc tgc aag tat gac tcg ttt gag gca atg ccg			960
Lys Leu Tyr Glu Leu Phe Cys Lys Tyr Asp Ser Phe Glu Ala Met Pro			
305	310	315	320
gct gaa gaa ttg caa cgg gtt gaa aag cgg att ttt caa aag tcg ctt			1008
Ala Glu Glu Leu Gln Arg Val Glu Lys Arg Ile Phe Gln Lys Ser Leu			
325	330	335	
gct gaa gtt tgg cag gag acc agt gac ttt tac att cat cgt atc aag			1056
Ala Glu Val Trp Gln Glu Thr Ser Asp Phe Tyr Ile His Arg Ile Lys			
340	345	350	
aac cct gag aaa atc aat cgt gct gca agc gat ggc aaa ctg aaa atg			1104
Asn Pro Glu Lys Ile Asn Arg Ala Ala Ser Asp Gly Lys Leu Lys Met			
355	360	365	
tcg ctt tgc ttt cgc tgg tac ctt ggg ctt tcc tca ttt tgg gcc aac			1152
Ser Leu Cys Phe Arg Trp Tyr Leu Gly Leu Ser Ser Phe Trp Ala Asn			
370	375	380	
tct ggg gca caa gat cgc gtc atg gac tat caa att tgg tgt ggc cct			1200
Ser Gly Ala Gln Asp Arg Val Met Asp Tyr Gln Ile Trp Cys Gly Pro			
385	390	395	400
gct att ggc gct ttc aat gat ttt acc aag ggc acg tac ctt gac gtg			1248
Ala Ile Gly Ala Phe Asn Asp Phe Thr Lys Gly Thr Tyr Leu Asp Val			
405	410	415	
act gtt gca aag agt tac cct tgt gtg gca cag atc aat ttg caa att			1296
Thr Val Ala Lys Ser Tyr Pro Cys Val Ala Gln Ile Asn Leu Gln Ile			
420	425	430	
ttg caa gga gct gcg tat ctg aaa cgc ctt ggt gtc att cgt ttt gac			1344
Leu Gln Gly Ala Ala Tyr Leu Lys Arg Leu Gly Val Ile Arg Phe Asp			

435

440

445

cgc atg ctg ctg cag gcc gtc gat atc gac gat cct gta ttt act tac 1392
Arg Met Leu Leu Gln Ala Val Asp Ile Asp Asp Pro Val Phe Thr Tyr
450 455 460

gtg ccg acc cag cca ctt 1410
Val Pro Thr Gln Pro Leu
465 470

<210> 68

<211> 470

<212> PRT

<213> Thraustochytrium sp.

<400> 68

Gly Asp Lys Ser Phe Met Glu Thr Tyr Asn Val Ser Ala Pro Leu Tyr
1 5 10 15

Thr Gly Ala Met Ala Lys Gly Ile Ala Ser Ala Asp Leu Val Ile Ala
20 25 30

Ala Gly Lys Arg Lys Ile Leu Gly Ser Phe Gly Ala Gly Gly Leu Pro
35 40 45

Ile Ser Ile Val Arg Glu Ala Leu Glu Lys Ile Gln Gln His Leu Pro
50 55 60

His Gly Pro Tyr Ala Val Asn Leu Ile His Ser Pro Phe Asp Ser Asn
65 70 75 80

Leu Glu Lys Gly Asn Val Asp Leu Phe Leu Glu Met Gly Val Thr Val
85 90 95

Val Glu Cys Ser Ala Phe Met Glu Leu Thr Ala Gln Val Val Arg Tyr
100 105 110

Arg Ala Ser Gly Leu Ser Lys Ser Ala Asp Gly Ser Ile Arg Ile Ala
115 120 125

His Arg Ile Ile Gly Lys Val Ser Arg Thr Glu Leu Ala Glu Met Phe
130 135 140

Ile Arg Pro Ala Pro Gln His Leu Leu Gln Lys Leu Val Ala Ser Gly
145 150 155 160

Glu Leu Thr Ala Glu Gln Ala Glu Leu Ala Thr Gln Val Pro Val Ala
165 170 175

Asp Asp Ile Ala Val Glu Ala Asp Ser Gly Gly His Thr Asp Asn Arg
180 185 190

Pro Ile His Val Ile Leu Pro Leu Ile Ile Asn Leu Arg Asn Arg Leu
195 200 205

His Lys Glu Leu Asp Tyr Pro Ser His Leu Arg Val Arg Val Gly Ala
210 215 220

Gly Gly Gly Ile Gly Cys Pro Gln Ala Ala Leu Ala Ala Phe Gln Met
225 230 235 240

Gly Ala Ala Phe Leu Ile Thr Gly Thr Val Asn Gln Leu Ala Arg Glu
245 250 255

Ser Gly Thr Cys Asp Asn Val Arg Leu Gln Leu Ser Lys Ala Thr Tyr
260 265 270

Ser Asp Val Cys Met Ala Pro Ala Ala Asp Met Phe Asp Gln Gly Val
275 280 285

Glu Leu Gln Val Leu Lys Lys Gly Thr Leu Phe Pro Ser Arg Ala Lys
290 295 300

Lys Leu Tyr Glu Leu Phe Cys Lys Tyr Asp Ser Phe Glu Ala Met Pro

305 310 315 320

Ala Glu Glu Leu Gln Arg Val Glu Lys Arg Ile Phe Gln Lys Ser Leu
325 330 335

Ala Glu Val Trp Gln Glu Thr Ser Asp Phe Tyr Ile His Arg Ile Lys
340 345 350

Asn Pro Glu Lys Ile Asn Arg Ala Ala Ser Asp Gly Lys Leu Lys Met
355 360 365

Ser Leu Cys Phe Arg Trp Tyr Leu Gly Leu Ser Ser Phe Trp Ala Asn
370 375 380

Ser Gly Ala Gln Asp Arg Val Met Asp Tyr Gln Ile Trp Cys Gly Pro
385 390 395 400

Ala Ile Gly Ala Phe Asn Asp Phe Thr Lys Gly Thr Tyr Leu Asp Val
405 410 415

Thr Val Ala Lys Ser Tyr Pro Cys Val Ala Gln Ile Asn Leu Gln Ile
420 425 430

Leu Gln Gly Ala Ala Tyr Leu Lys Arg Leu Gly Val Ile Arg Phe Asp
435 440 445

Arg Met Leu Leu Gln Ala Val Asp Ile Asp Asp Pro Val Phe Thr Tyr
450 455 460

Val Pro Thr Gln Pro Leu
465 470

<210> 69

<211> 20

<212> DNA

<213> Artificial sequence

 $\langle 220 \rangle$

<223> primer

<400> 69

ggyatgmtgr ttggtgaagg

20

<210> 70

<211> 22

<212> DNA

<213> Artificial sequence

<220>

<223> primer

<400> 70

trttsasrta ytggyaacct tg

22

<210> 71

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> primer

<220>

<221> misc_feature

<222> (1)..(20)

<223> n = a c g or t

<400> 71

atgkcngaag gttgtggcca

20

<210> 72

<211> 23

<212> DNA

<213> Artificial sequence

<220>

<223> primer

<400> 72

ccwgaratra agccrttdgg ttg	23
<210> 73	
<211> 30	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> primer	
<400> 73	
cggggtaccc gggagccgcc ttggctttgt	30
<210> 74	
<211> 35	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> primer	
<400> 74	
aaactgcagc ccgggtccag ctggcaggca ccctg	35
<210> 75	
<211> 798	
<212> DNA	
<213> Shewanella olleyana	
<400> 75	
atggcagaag gttgtggcca attattgcag ttcttcatgc tgcacattgg tatgcacacc	60
ttagttgaaa acggacgttt ccagccttta gaaaatgctt cacaaaaagt acgttgtcgt	120
ggccaagtac tgccacaaca tggatgaactg acgtaccgca tggaagtcac agaaattggt	180
actcacctc gcccatagc caaagccaat attgaaatat tgctcaatgg taaagcggtc	240
gtggacttcc aaaatcttgg ggtgatgatt aaagaagaag gtgaatgtac tcgttacact	300

gccgactcta ctgaaacaca tacaacctca ggacacgtcc aaaaaaaca cagccacaac 360

acaccagcat cattaaatgc accgttaatg gcacaagtgc cagacttaag tgaaccagcc 420

aataaaggcg ttatcccgt gcaacatgtt gaagcgccta tgctgccaga ctacccaaat 480

cgaaccctg atacgtgcc gttcaccgcg taccatatgt ttgagtttgc aacaggtgac 540

atcgaaaact gttttggacc tgactttagt atttaccggg gctttattcc gccgcgcacg 600

ccatgtggtg acttacagct aacaaccgt gttgttgata ttcaaggtaa acgtggcgag 660

cttaaaaaac cgtcatcgtg tatcgtgaa tatgaagtgc caaccgatgc gtggtatfff 720

gctaaaaaca gtcacgttc agtgcgtcct tactcggtat taatggaaat atcactgcaa 780

cccaacggct tcatttca 798

<210> 76
 <211> 1042
 <212> DNA
 <213> Shewanella olleyana

<400> 76

ggcatgatgg ttgtgaagg tatcggcatg gtagcactta agcgcctaga agatgctgag 60

cgtgatggcg accgtattta ttctgtgatt aaaggtgtcg gcgcttcac agacggtaaa 120

tttaagagta ttatgcacc gcgccctgaa ggccaagcaa aagcattaaa acgagcttat 180

gatgacgtg gttttgcccc tgaaacagtt ggcttaatcg aagctcacgg tacgggtact 240

gctgcagggtg atgtagccga atttaacggc cttaaactctg tatttgggtga aaacgatcca 300

actaagcaac acatcgtttt aggttcagtg aaatcacaag tgggtcacac gaaatcaacc 360

gtggttactg ctggcgtgat taaagctgcc cttgccctgc accataaagt attgccaccg 420

accattaacg tccttaagcc aaaccttaag cttaatgttg aggattcacc gtttttcgtt 480

aataccgaaa cacgcccattg gatgcctcgc cctgacggca ctctcgccg tgctggtatt 540
 agctcgttcg gttttgggtg aactaacttc cacttagtat tagaagaata caccctgag 600
 cacagccatg atgagaaata ccgtcaacgc caagtggctc aaagcttatt aatgagtgc 660
 gataataaag cagccttgat tgcagaagtg aataagctaa ctgcagacat cagcgcgtt 720
 aaaggcacag ataacagcag cattgaacaa gctgaacttg ctgcattgc taaactatat 780
 gctgttcgca ccatagatac ttcagcagcc cgtttaggtc ttgtggtatc aagcctaat 840
 gaattaacca ctgagcttgg tttagcgtta aagcagctta ataatgatgt tgatgcatgg 900
 caactgccat cagggactag ctaccgctct tcagcactca tcagattaa tgcaaaccaa 960
 aaggcgacta aaggtaaaaa agcgactaac gcaccgaaag ttgcagcatt gtttcaggt 1020
 caaggttcac aatacctcaa ca 1042

<210> 77
 <211> 806
 <212> DNA
 <213> Shewanella japonica

<400> 77
 atggcggaag gttgtggcca attactgcaa ttctttatgc tgcacattgg tatgcacacg 60
 ctggttgaaa atggccgttt ccaaccactt gaaaatgctt cacaaaaagt gcgttgctgt 120
 ggtcaagttc tgccgcagca cggatgaactg acttaccgga tggaatcac tgaaattggc 180
 attcaccctc gcccatatgc caaagcgaat attgatattt tgcttaacgg taaagcggtt 240
 gtgcacttcc aaaacttagg tgtcatgac aaagaagaaa gcgaatgtac gcgctacctt 300
 aatgatacgc ccgtgtcga tgcctcagct gatcgaatta attcagcaac caataatatt 360

ctataccag cggttcaac caatgcgcca ctcatggctc aactgcctga tttgaatgcc 420

ccaacgaata aaggcgttat ccactgcaa catgttgaag cgccgataat tccagattat 480

ccaaatcgta ctctgatac cctgccattc acggcgtatc acatgttcga atttgccact 540

ggcaatattg aaaactgctt tggaccggac tttagtattt accgtggttt cattccaccg 600

cgcacaccat gtggcgactt acagctaacg actcgtattg ttgatattca aggtaaacgt 660

ggcgaattga aaaagccatc atcgtgtatc gcagaatatg aagtgccaac tgatgcatgg 720

tatttcgcta aaaacagcca cgcctcggtc ataccttatt cagtgttgat ggaaatttca 780

ctgcaacca atggcttcat ctgagg 806

<210> 78
 <211> 1042
 <212> DNA
 <213> Shewanella japonica

<400> 78
 ggtatgctgg ttgtgaagg cattggcatg gtggcattaa aacgtcttga agatgctgag 60

cgtgacggtg accgtattta ctcatgcatt aaaggggtcg gcgcttcac tgatggtaag 120

ttcaaatcaa tttatgcacc tcgacctgaa ggccaagcta aagcgtgaa gcgtgcttat 180

gatgacgccg gctttgcacc tgaaaccgtt ggcttaattg aagctcacgg aacaggcact 240

gcagcgggtg atgtggcaga atttaatggt cttaaactcg tatttgggtga gaatgactca 300

acaaagcaac acattgcttt aggttcagtt aagtcacaag tgggccatac taaatcaact 360

gcgggaaccg cgggtgtgat taaagcggcg ttagcactgc atcataaagt gctgccgcca 420

accatcaacg tctctaagcc taaccctaag cttaatgttg aggattcacc gtttttcatt 480

aacactgaaa ctgcaccttg gatgcctcgc cctgatggca caccagccg agctggtata 540

agttcgttcg gttttggtgg cacaaacttc cacttagtac tagaagaata cagcccagag 600

cacagccgtg atgagaaata tcgtcagcgc caagtagcac aaagcttatt gattagcgct 660

gacaataaag ctgagctcat tgcagaaatc aacaagctta acgctgacat cagcgcgctt 720

aaaggcacag ataacagcag catcgaacaa gctgaacttg cccgcattgc taaactatat 780

gctgttcgca ctttagatac ttcagcagtc cgtttgggtc ttgtggtctc aagccttaat 840

gaattaacca ctcaacttgg tttagcgtta aagcagctaa gtaacgacgc tgaagcatgg 900

caattacat caggtacgag ctatcgctca tctgcgctca tcacgattaa tgccaaccaa 960

aagacgacta aaggtaaaaa agcagctaac acaccgaaag tagcagcatt atttcaggt 1020

caaggttcgc aatatctcaa ca 1042

<210> 79
 <211> 6
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> restriction enzyme site

<400> 79
 catatg 6

<210> 80
 <211> 192
 <212> DNA
 <213> Schizochytrium sp.

<220>
 <221> misc_feature
 <222> (1)..(7)
 <223> BspHI restriction site

<220>
 <221> misc_feature
 <222> (188)..(192)
 <223> SacII restriction site

<400> 80
 tcatgaagcc ggttgctccg aagttctacg cgcgtctcaa cattgacgag caggacgaga 60

 cccgtgatcc gatcctcaac aaggacaacg cgccgtcttc cagctctagc tcctcttcca 120

 gctcttccag ctcttccagc ccgtcgccag ctccgtccgc cccagtgcaa aagaaggctg 180

 ctccggccgc gg 192

<210> 81
 <211> 717
 <212> DNA
 <213> Nostoc sp.

<220>
 <221> CDS
 <222> (4)..(714)

<400> 81
 taa ttg ttg cag cat act tgg cta cca aaa ccc cca aat tta acc tta 48
 Leu Leu Gln His Thr Trp Leu Pro Lys Pro Pro Asn Leu Thr Leu
 1 5 10 15

 ttg tca gat gaa gtt cat ctc tgg cgc att ccc ctt gac caa cca gaa 96
 Leu Ser Asp Glu Val His Leu Trp Arg Ile Pro Leu Asp Gln Pro Glu
 20 25 30

 tca cag cta cag gat tta gcc gct acc tta tct agt gac gaa tta gcc 144
 Ser Gln Leu Gln Asp Leu Ala Ala Thr Leu Ser Ser Asp Glu Leu Ala
 35 40 45

 cgt gca aac aga ttt tat ttt ccc gaa cat cgc cgg cgt ttt act gct 192
 Arg Ala Asn Arg Phe Tyr Phe Pro Glu His Arg Arg Arg Phe Thr Ala
 50 55 60

 ggt cgt ggt att ctc cgc agt atc ttg ggg ggc tat ttg ggt gtg gaa 240
 Gly Arg Gly Ile Leu Arg Ser Ile Leu Gly Gly Tyr Leu Gly Val Glu
 65 70 75

cca ggg caa gtt aaa ttt gat tat gaa tcc cgt ggt aaa cca ata tta 288
Pro Gly Gln Val Lys Phe Asp Tyr Glu Ser Arg Gly Lys Pro Ile Leu
80 85 90 95

ggc gat cgc ttt gcc gag agt ggt tta tta ttt aac ttg tca cac tcc 336
Gly Asp Arg Phe Ala Glu Ser Gly Leu Leu Phe Asn Leu Ser His Ser
100 105 110

cag aac ttg gcc ttg tgt gca gtc aat tac acg cgc caa atc ggc atc 384
Gln Asn Leu Ala Leu Cys Ala Val Asn Tyr Thr Arg Gln Ile Gly Ile
115 120 125

gat tta gaa tat ctc cgc ccc aca tct gat tta gaa tcc ctt gcc aaa 432
Asp Leu Glu Tyr Leu Arg Pro Thr Ser Asp Leu Glu Ser Leu Ala Lys
130 135 140

agg ttc ttt tta ccg cga gaa tat gaa tta ttg cga tcg cta ccc gat 480
Arg Phe Phe Leu Pro Arg Glu Tyr Glu Leu Leu Arg Ser Leu Pro Asp
145 150 155

gag caa aaa caa aaa att ttc ttt cgt tac tgg act tgt aaa gag gct 528
Glu Gln Lys Gln Lys Ile Phe Phe Arg Tyr Trp Thr Cys Lys Glu Ala
160 165 170 175

tat ctt aaa gca acg ggt gac ggc atc gct aaa tta gag gaa att gaa 576
Tyr Leu Lys Ala Thr Gly Asp Gly Ile Ala Lys Leu Glu Glu Ile Glu
180 185 190

ata gca cta act ccc aca gaa cca gct aag tta cag aca gct cca gcg 624
Ile Ala Leu Thr Pro Thr Glu Pro Ala Lys Leu Gln Thr Ala Pro Ala
195 200 205

tgg agt ctc cta gag cta gtg cca gat gat aat tgt gtt gct gct gtt 672
Trp Ser Leu Leu Glu Leu Val Pro Asp Asp Asn Cys Val Ala Ala Val
210 215 220

gcc gtg gcg ggt ttt ggc tgg cag cca aaa ttc tgg cat tat tga 717
Ala Val Ala Gly Phe Gly Trp Gln Pro Lys Phe Trp His Tyr
225 230 235

<210> 82
<211> 237
<212> PRT

<213> Nostoc sp.

<400> 82

Leu Leu Gln His Thr Trp Leu Pro Lys Pro Pro Asn Leu Thr Leu Leu
1 5 10 15

Ser Asp Glu Val His Leu Trp Arg Ile Pro Leu Asp Gln Pro Glu Ser
20 25 30

Gln Leu Gln Asp Leu Ala Ala Thr Leu Ser Ser Asp Glu Leu Ala Arg
35 40 45

Ala Asn Arg Phe Tyr Phe Pro Glu His Arg Arg Arg Phe Thr Ala Gly
50 55 60

Arg Gly Ile Leu Arg Ser Ile Leu Gly Gly Tyr Leu Gly Val Glu Pro
65 70 75 80

Gly Gln Val Lys Phe Asp Tyr Glu Ser Arg Gly Lys Pro Ile Leu Gly
85 90 95

Asp Arg Phe Ala Glu Ser Gly Leu Leu Phe Asn Leu Ser His Ser Gln
100 105 110

Asn Leu Ala Leu Cys Ala Val Asn Tyr Thr Arg Gln Ile Gly Ile Asp
115 120 125

Leu Glu Tyr Leu Arg Pro Thr Ser Asp Leu Glu Ser Leu Ala Lys Arg
130 135 140

Phe Phe Leu Pro Arg Glu Tyr Glu Leu Leu Arg Ser Leu Pro Asp Glu
145 150 155 160

Gln Lys Gln Lys Ile Phe Phe Arg Tyr Trp Thr Cys Lys Glu Ala Tyr
165 170 175

Leu Lys Ala Thr Gly Asp Gly Ile Ala Lys Leu Glu Glu Ile Glu Ile
 180 185 190

Ala Leu Thr Pro Thr Glu Pro Ala Lys Leu Gln Thr Ala Pro Ala Trp
 195 200 205

Ser Leu Leu Glu Leu Val Pro Asp Asp Asn Cys Val Ala Ala Val Ala
 210 215 220

Val Ala Gly Phe Gly Trp Gln Pro Lys Phe Trp His Tyr
 225 230 235