



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0072147
(43) 공개일자 2021년06월16일

(51) 국제특허분류(Int. Cl.)
C12N 9/02 (2006.01) C12P 7/04 (2006.01)
C12P 7/24 (2006.01)
(52) CPC특허분류
C12N 9/0008 (2013.01)
C12P 7/04 (2013.01)
(21) 출원번호 10-2021-7017654(분할)
(22) 출원일자(국제) 2014년01월16일
심사청구일자 없음
(62) 원출원 특허 10-2020-7004969
원출원일자(국제) 2014년01월16일
심사청구일자 2020년03월20일
(85) 번역문제출일자 2021년06월09일
(86) 국제출원번호 PCT/US2014/011859
(87) 국제공개번호 WO 2014/113571
국제공개일자 2014년07월24일
(30) 우선권주장
61/753,273 2013년01월16일 미국(US)

(71) 출원인
게노마티카 인코포레이티드
미국 92121 캘리포니아주 샌디에고 텍서스 센터
드라이브 4757
(72) 발명자
루드, 매튜
미국 캘리포니아 94114 샌프란시스코 834-14 스트
리트 아파트먼트 1
트린, 나
미국 캘리포니아 94065 레드우드 시티 레드우드
쇼어스 파크웨이 274 #331
(74) 대리인
(뒷면에 계속)
특허법인(유)화우

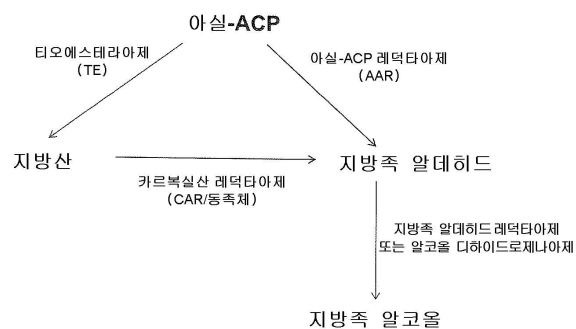
전체 청구항 수 : 총 82 항

(54) 발명의 명칭 개선된 특성들을 갖는 아실-ACP 레덕타아제

(57) 요약

본 발명은 재조합 숙주 세포들에서 발현될 때 개선된 지방족 알데히드 및 지방족 알코올 생성을 유도하는 아실-ACP 레덕타아제(AAR) 효소 변이체들에 관한 것이다. 또한, 본 발명은 특정한 특성들을 갖는 지방족 알코올 조성물들의 생성을 위해 이러한 AAR 변이체들을 만들고 이용하는 방법에 관한 것이다.

대표도



(52) CPC특허분류

C12P 7/24 (2013.01)

C12Y 102/0108 (2013.01)

Y02P 20/52 (2015.11)

(72) 발명자

셔머, 안드레아스

미국 캘리포니아 94080 사우스 샌프란시스코 게이
트웨이 블러바드 600

가노, 제이콥

미국 캘리포니아 94002 벨몬트 뉴랜즈 애비뉴 2802

명세서

청구범위

청구항 1

SEQ ID NO: 57에 대해 적어도 90 %의 서열 동일성(sequence identity)을 포함하는 변이 아실-ACP 레덕타아제(AAR) 폴리펩티드에 있어서,

상기 AAR 폴리펩티드는 아실-ACP의 지방족 알데히드로의 전환을 촉진시키는 변이 AAR 폴리펩티드.

청구항 2

제 1 항에 있어서,

제조합 숙주 세포에서의 상기 변이 AAR 폴리펩티드의 발현은 대응하는 야생형 숙주 세포에서의 야생형 AAR 폴리펩티드의 발현에 의해 생성된 지방족 알데히드 또는 지방족 알코올 조성물의 역가(titer)에 비해 지방족 알데히드 또는 지방족 알코올 조성물의 더 높은 역가를 유도하는 변이 AAR 폴리펩티드.

청구항 3

제 2 항에 있어서,

상기 지방족 알코올 조성물은 C12, C14 또는 C16 지방족 알코올 조성물 또는 이의 조합인 변이 AAR 폴리펩티드.

청구항 4

제 1 항에 있어서,

상기 변이 AAR 폴리펩티드는 아미노산 위치 18에서 돌연변이를 포함하는 변이 AAR 폴리펩티드.

청구항 5

제 4 항에 있어서,

상기 돌연변이는 S18W인 변이 AAR 폴리펩티드.

청구항 6

제 4 항에 있어서,

상기 변이 AAR 폴리펩티드는 아미노산들 8, 16, 21, 24, 31, 34, 35, 43, 50, 63, 86, 112, 113, 116, 118, 120, 135, 148, 153, 154, 155, 157, 159, 168, 172, 187, 188, 191, 209, 210, 211, 236, 277, 281, 283, 285, 291, 324, 328, 335, 337 및 338로 구성된 그룹으로부터 선택된 아미노산 위치에서 돌연변이를 더 포함하는 변이 AAR 폴리펩티드.

청구항 7

제 6 항에 있어서,

상기 돌연변이는 L8A, D16L, M21L, D24E, D24Y, D24V, D24P, L31V, L31M, W34F, W35F, D43E, A50Q, C63A, C63G, C63Y, S86G, A112R, S113K, Q116G, R118Q, T120S, A135S, T148C, T148E, T148V, I153P, T154A, Q155C, Q155L, T157V, A159V, I168V, C172L, T187V, T188H, T188V, Q191A, L209R, E210Y, A211W, T236C, Q277V, A281L, E283G, E283S, A285V, M291V, A324T, A328S, Q335N, L337V 및 L338W로 구성된 그룹으로부터 선택되는 변이 AAR 폴리펩티드.

청구항 8

제 7 항에 있어서,

상기 변이 AAR 폴리펩티드는 M21L 돌연변이를 포함하는 변이 AAR 폴리펩티드.

청구항 9

제 7 항에 있어서,

상기 변이 AAR 폴리펩티드는 C63GA 돌연변이를 포함하는 변이 AAR 폴리펩티드.

청구항 10

제 7 항에 있어서,

상기 변이 AAR 폴리펩티드는 S113K 돌연변이를 포함하는 변이 AAR 폴리펩티드.

청구항 11

제 7 항에 있어서,

상기 변이 AAR 폴리펩티드는 T154A 돌연변이를 포함하는 변이 AAR 폴리펩티드.

청구항 12

제 7 항에 있어서,

상기 변이 AAR 폴리펩티드는 A281L 돌연변이를 포함하는 변이 AAR 폴리펩티드.

청구항 13

제 7 항에 있어서,

상기 변이 AAR 폴리펩티드는 L8A 돌연변이를 포함하는 변이 AAR 폴리펩티드.

청구항 14

제 5 항에 있어서,

상기 변이 AAR 폴리펩티드는 M21L 돌연변이, C63G 돌연변이, S113K 돌연변이, T154A, 및 A281L 돌연변이를 더 포함하는 변이 AAR 폴리펩티드.

청구항 15

제 14 항에 있어서,

SEQ ID NO: 58을 포함하는 변이 AAR 폴리펩티드.

청구항 16

제 5 항에 있어서,

상기 변이 AAR 폴리펩티드는 L8A 돌연변이, M21L 돌연변이, C63G 돌연변이, S113K 돌연변이, T154A 돌연변이, 및 A281L 돌연변이를 더 포함하는 변이 AAR 폴리펩티드.

청구항 17

제 16 항에 있어서,

SEQ ID NO: 59를 포함하는 변이 AAR 폴리펩티드.

청구항 18

제 5 항에 있어서,

상기 변이 AAR 폴리펩티드는 D16L 돌연변이, M21L 돌연변이, C63G 돌연변이, S113K 돌연변이, T154A, 및 A281L 돌연변이를 더 포함하는 변이 AAR 폴리펩티드.

청구항 19

제 18 항에 있어서,

SEQ ID NO: 60을 포함하는 변이 AAR 폴리펩티드.

청구항 20

제 5 항에 있어서,

상기 변이 AAR 폴리펩티드는 L8A 돌연변이, D24V 돌연변이, C63G 돌연변이, S113K 돌연변이, Q155L 돌연변이, 및 A281L 돌연변이를 더 포함하는 변이 AAR 폴리펩티드.

청구항 21

제 20 항에 있어서,

SEQ ID NO: 61을 포함하는 변이 AAR 폴리펩티드.

청구항 22

제 5 항에 있어서,

상기 변이 AAR 폴리펩티드는 D24P 돌연변이, L31M 돌연변이, C63G 돌연변이, S113K 돌연변이, T154A 돌연변이, 및 A281L 돌연변이를 더 포함하는 변이 AAR 폴리펩티드.

청구항 23

제 22 항에 있어서,

SEQ ID NO: 62를 포함하는 변이 AAR 폴리펩티드.

청구항 24

제 5 항에 있어서,

상기 변이 AAR 폴리펩티드는 L8A 돌연변이, D16L 돌연변이, D24V 돌연변이, C63G 돌연변이, S113K 돌연변이, T154A 돌연변이, 및 A281L 돌연변이를 더 포함하는 변이 AAR 폴리펩티드.

청구항 25

제 24 항에 있어서,

SEQ ID NO: 63을 포함하는 변이 AAR 폴리펩티드.

청구항 26

제 5 항에 있어서,

상기 변이 AAR 폴리펩티드는 D24E 돌연변이, C63G 돌연변이, S113K 돌연변이, T154A 돌연변이, 및 A281L 돌연변이를 더 포함하는 변이 AAR 폴리펩티드.

청구항 27

제 26 항에 있어서,

SEQ ID NO: 64를 포함하는 변이 AAR 폴리펩티드.

청구항 28

제 1 항 내지 제 27 항 중 어느 한 항의 변이 AAR 폴리펩티드를 발현시키는 재조합 숙주 세포.

청구항 29

제 28 항에 있어서,

상기 재조합 숙주 세포는, 변이 AAR 폴리펩티드를 발현시키기에 효과적인 조건들 하에서 탄소원을 함유한 배지

에서 배양될 때, 대응하는 야생형 AAR 폴리펩티드를 발현시키는 숙주 세포에 의해 생성된 지방족 알데히드 또는 지방족 알코올 조성물의 역가보다 적어도 10 % 더 큰, 적어도 15 % 더 큰, 적어도 20 % 더 큰, 적어도 25 % 더 큰, 또는 적어도 30 % 더 큰 역가를 갖는 지방족 알데히드 또는 지방족 알코올 조성물을 생성하는 재조합 숙주 세포.

청구항 30

제 29 항에 있어서,

상기 지방족 알데히드 또는 지방족 알코올 조성물은 약 30 g/L 내지 250 g/L의 역가에서 생성되는 재조합 숙주 세포.

청구항 31

제 29 항에 있어서,

상기 지방족 알데히드 또는 지방족 알코올 조성물은 세포외에서(extracellularly) 생성되는 재조합 숙주 세포.

청구항 32

제 28 항 내지 제 31 항 중 어느 한 항의 재조합 숙주 세포를 포함하는 세포 배양물.

청구항 33

제 32 항에 있어서,

상기 지방족 알코올 조성물은 C6, C8, C10, C12, C13, C14, C15, C16, C17, 및 C18 지방족 알코올 중 1 이상을 포함하는 세포 배양물.

청구항 34

제 33 항에 있어서,

상기 지방족 알코올 조성물은 C10:1, C12:1, C14:1, C16:1, 및 C18:1 불포화 지방족 알코올 중 1 이상을 포함하는 세포 배양물.

청구항 35

제 33 항에 있어서,

상기 지방족 알코올 조성물은 불포화 지방족 알코올을 포함하는 세포 배양물.

청구항 36

제 33 항에 있어서,

상기 지방족 알코올 조성물은 지방족 알코올의 환원 말단으로부터 C₇과 C₈ 사이의 탄소 사슬의 위치 7에서 이중 결합을 갖는 지방족 알코올을 포함하는 세포 배양물.

청구항 37

제 33 항에 있어서,

상기 지방족 알코올 조성물은 포화 지방족 알코올을 포함하는 세포 배양물.

청구항 38

역가의 증가를 갖는 지방족 알코올 조성물을 생성하는 방법에 있어서,

- i. 탄소원을 갖는 제 27 항의 숙주 세포를 배양하는 단계; 및
- ii. 지방족 알코올 조성물을 수득(harvest)하는 단계를 포함하는 지방족 알코올 조성물 생성 방법.

청구항 39

제 38 항에 있어서,

지방족 알코올의 역가는 야생형 AAR-발현 숙주 세포에 의해 생성된 지방족 알코올 조성물의 역가보다 적어도 20 % 내지 30 % 더 큰 지방족 알코올 조성물 생성 방법.

청구항 40

SEQ ID NO: 65에 대해 적어도 90 %의 서열 동일성을 포함하는 변이 아실-ACP 레덕타아제(AAR) 폴리펩티드에 있어서,

상기 폴리펩티드는 아실-ACP의 지방족 알데히드로의 전환을 촉진시키는 변이 AAR 폴리펩티드.

청구항 41

제 40 항에 있어서,

제조합 숙주 세포에서의 상기 변이 AAR 폴리펩티드의 발현은 대응하는 야생형 숙주 세포에서의 야생형 AAR 폴리펩티드의 발현에 의해 생성된 지방족 알데히드 또는 지방족 알코올 조성물의 역가에 비해 지방족 알데히드 또는 지방족 알코올 조성물의 더 높은 역가를 유도하는 변이 AAR 폴리펩티드.

청구항 42

제 41 항에 있어서,

상기 지방족 알코올 조성물은 C12, C14 또는 C16 지방족 알코올 조성물 또는 이의 조합인 변이 AAR 폴리펩티드.

청구항 43

제 40 항에 있어서,

상기 변이 AAR 폴리펩티드는 아미노산 위치 61에서 돌연변이를 포함하는 변이 AAR 폴리펩티드.

청구항 44

제 43 항에 있어서,

상기 돌연변이는 D61E인 변이 AAR 폴리펩티드.

청구항 45

SEQ ID NO: 34에 대해 적어도 90 %의 서열 동일성을 포함하는 변이 아실-ACP 레덕타아제(AAR) 폴리펩티드에 있어서,

제조합 숙주 세포에서의 상기 변이 AAR 폴리펩티드의 발현은 대응하는 야생형 숙주 세포에서의 야생형 AAR 폴리펩티드의 발현에 의해 생성된 지방족 알코올 조성물의 역가에 비해 C12, C14 또는 C16 지방족 알코올 조성물의 더 높은 역가 또는 지방족 알데히드 또는 지방족 알코올 조성물의 더 높은 역가를 유도하며, 상기 변이 AAR 폴리펩티드는 아미노산 40, 52, 61, 273, 303, 340, 344, 345, 및 346으로 구성된 그룹으로부터 선택된 아미노산 위치에서 돌연변이를 포함하는 변이 AAR 폴리펩티드.

청구항 46

제 45 항에 있어서,

상기 변이 AAR 폴리펩티드는 Q40V, G52V, D61E, G273E, K303G, H340P, L344A, L344D, L344S, L344T, L345R, V346P, V346G, 및 A345*로 구성된 그룹으로부터 선택된 돌연변이를 포함하는 변이 AAR 폴리펩티드.

청구항 47

제 46 항에 있어서,

상기 변이 AAR 폴리펩티드는 V346P에서 돌연변이를 포함하는 변이 AAR 폴리펩티드.

청구항 48

제 47 항에 있어서,

SEQ ID NO: 66을 포함하는 변이 AAR 폴리펩티드.

청구항 49

제 46 항에 있어서,

상기 변이 AAR 폴리펩티드는 Q40V에서 돌연변이를 포함하는 변이 AAR 폴리펩티드.

청구항 50

제 49 항에 있어서,

SEQ ID NO: 67을 포함하는 변이 AAR 폴리펩티드.

청구항 51

제 46 항에 있어서,

상기 변이 AAR 폴리펩티드는 A345R에서 돌연변이를 포함하는 변이 AAR 폴리펩티드.

청구항 52

제 51 항에 있어서,

SEQ ID NO: 68을 포함하는 변이 AAR 폴리펩티드.

청구항 53

제 46 항에 있어서,

상기 변이 AAR 폴리펩티드는 L344S에서 돌연변이를 포함하는 변이 AAR 폴리펩티드.

청구항 54

제 53 항에 있어서,

SEQ ID NO: 69를 포함하는 변이 AAR 폴리펩티드.

청구항 55

제 46 항에 있어서,

상기 변이 AAR 폴리펩티드는 V346G에서 돌연변이를 포함하는 변이 AAR 폴리펩티드.

청구항 56

제 55 항에 있어서,

SEQ ID NO: 70을 포함하는 변이 AAR 폴리펩티드.

청구항 57

제 46 항에 있어서,

상기 변이 AAR 폴리펩티드는 L344D에서 돌연변이를 포함하는 변이 AAR 폴리펩티드.

청구항 58

제 57 항에 있어서,

SEQ ID NO: 71을 포함하는 변이 AAR 폴리펩티드.

청구항 59

제 46 항에 있어서,

상기 변이 AAR 폴리펩티드는 G52V에서 돌연변이를 포함하는 변이 AAR 폴리펩티드.

청구항 60

제 59 항에 있어서,

SEQ ID NO: 72를 포함하는 변이 AAR 폴리펩티드.

청구항 61

제 46 항에 있어서,

상기 변이 AAR 폴리펩티드는 L344T에서 돌연변이를 포함하는 변이 AAR 폴리펩티드.

청구항 62

제 61 항에 있어서,

SEQ ID NO: 73을 포함하는 변이 AAR 폴리펩티드.

청구항 63

제 46 항에 있어서,

상기 변이 AAR 폴리펩티드는 K303G에서 돌연변이를 포함하는 변이 AAR 폴리펩티드.

청구항 64

제 63 항에 있어서,

SEQ ID NO: 74를 포함하는 변이 AAR 폴리펩티드.

청구항 65

제 46 항에 있어서,

상기 변이 AAR 폴리펩티드는 L344A에서 돌연변이를 포함하는 변이 AAR 폴리펩티드.

청구항 66

제 65 항에 있어서,

SEQ ID NO: 75를 포함하는 변이 AAR 폴리펩티드.

청구항 67

제 46 항에 있어서,

상기 변이 AAR 폴리펩티드는 H340P에서 돌연변이를 포함하는 변이 AAR 폴리펩티드.

청구항 68

제 67 항에 있어서,

SEQ ID NO: 76을 포함하는 변이 AAR 폴리펩티드.

청구항 69

제 46 항에 있어서,

상기 변이 AAR 폴리펩티드는 G273E에서 돌연변이를 포함하는 변이 AAR 폴리펩티드.

청구항 70

제 69 항에 있어서,

SEQ ID NO: 77을 포함하는 변이 AAR 폴리펩티드.

청구항 71

제 42 항 내지 제 70 항 중 어느 한 항의 변이 AAR 폴리펩티드를 발현시키는 재조합 숙주 세포.

청구항 72

제 71 항에 있어서,

상기 재조합 숙주 세포는, 변이 AAR 폴리펩티드를 발현시키기에 효과적인 조건들 하에서 탄소원을 함유한 배지에서 배양될 때, 대응하는 야생형 AAR 폴리펩티드를 발현시키는 숙주 세포에 의해 생성된 지방족 알데히드 또는 알코올 조성물의 역가보다 적어도 10 % 더 큰, 적어도 15 % 더 큰, 적어도 20 % 더 큰, 적어도 25 % 더 큰, 또는 적어도 30 % 더 큰 역가를 갖는 지방족 알데히드 또는 지방족 알코올 조성물을 생성하는 재조합 숙주 세포.

청구항 73

제 72 항에 있어서,

상기 지방족 알코올 조성물은 약 30 g/L 내지 250 g/L의 역가에서 생성되는 재조합 숙주 세포.

청구항 74

제 73 항에 있어서,

상기 지방족 알코올 조성물은 세포외에서 생성되는 재조합 숙주 세포.

청구항 75

제 73 항 내지 제 74 항 중 어느 한 항의 재조합 숙주 세포를 포함하는 세포 배양물.

청구항 76

제 75 항에 있어서,

상기 지방족 알코올 조성물은 C6, C8, C10, C12, C13, C14, C15, C16, C17, 및 C18 지방족 알코올 중 1 이상을 포함하는 세포 배양물.

청구항 77

제 76 항에 있어서,

상기 지방족 알코올 조성물은 C10:1, C12:1, C14:1, C16:1, 및 C18:1 불포화 지방족 알코올 중 1 이상을 포함하는 세포 배양물.

청구항 78

제 76 항에 있어서,

상기 지방족 알코올 조성물은 불포화 지방족 알코올을 포함하는 세포 배양물.

청구항 79

제 76 항에 있어서,

상기 지방족 알코올 조성물은 지방족 알코올의 환원 말단으로부터 C₇과 C₈ 사이의 탄소 사슬의 위치 7에서 이중 결합을 갖는 지방족 알코올을 포함하는 세포 배양물.

청구항 80

제 76 항에 있어서,

상기 지방족 알코올 조성물은 포화 지방족 알코올을 포함하는 세포 배양물.

청구항 81

역가의 증가를 갖는 지방족 알코올 조성물을 생성하는 방법에 있어서,

- i. 탄소원을 갖는 제 73 항의 숙주 세포를 배양하는 단계; 및
- ii. 지방족 알코올 조성물을 수득하는 단계를 포함하는 지방족 알코올 조성물 생성 방법.

청구항 82

제 81 항에 있어서,

지방족 알코올의 역가는 야생형 AAR-발현 숙주 세포에 의해 생성된 지방족 알코올 조성물의 역가보다 적어도 20 % 내지 30 % 더 큰 지방족 알코올 조성물 생성 방법.

발명의 설명

기술 분야

- [0001] 본 출원은 2013년 1월 16일에 출원된 미국 가출원 61/753,273의 이익을 주장하며, 이의 내용은 본 명세서에서 전문이 인용 참조된다.
- [0002] 본 출원은 ASCII 포맷으로 전자 제출된 서열 목록을 포함하며, 본 명세서에서 전문이 인용 참조된다. 2014년 1월 13일에 작성된 상기의 ASCII의 복사본은 파일명을 LS00046PCT_SL.txt라고 하였고, 크기는 155,731 바이트이다.
- [0003] 본 발명은, 재조합 숙주 세포들에서 발현될 때 개선된 지방족 알데히드 및/또는 지방족 알코올 생성을 유도하는 아실-ACP 레덕타아제(AAR) 효소 변이체에 관한 것이다. 또한, 본 발명은 특정한 특성들을 갖는 지방족 알코올 조성물의 생성을 위해 이러한 AAR 변이체를 만들고 이용하는 방법에 관한 것이다.

배경 기술

- [0004] 지방족 알코올은 산업 생화학의 중요한 카테고리를 나타낸다. 예를 들어, 지방족 알코올 및 이의 유도체의 세계적인 연간 매출은 US 달러로 10 억 달러가 넘는다. 이러한 분자들 및 이의 유도체들은, 계면활성제, 윤활제, 가소제, 용제, 유화제, 연화제(emollients), 증점제(thickeners), 향미제(flavors), 향료(fragrances) 및 연료로서의 이용을 포함하여, 다수에 적용들을 갖는다. 이들의 양친매성 성질(amphiphilic nature)로 인해, 지방족 알코올은 비이온 계면활성제처럼 행동하며, 이는 개인용품(personal care) 및 가정용품, 예를 들어 세제에 유용하다. 짧은 사슬형 지방족 알코올은 유화제, 연화제 및 증점제로서 화장품 및 식품 산업에 이용된다.
- [0005] 본질적으로, 지방족 알코올은 다양한 아실-ACP 또는 아실-CoA 분자들을 대응하는 1차 알코올들로 환원시킬 수 있는 효소들에 의해 만들어진(예를 들어, 미국 특허 8,323,924; 8,268,599 및 8,097,439; 및 미국 특허 공보 20120282663 및 20100105963, 본 명세서에서 인용 참조됨). 하지만, 현재의 기술들은 주로 지방산의 대응하는 1차 알코올로의 무기 촉매-매개 환원(inorganic catalyst-mediated reduction)을 수반한다. 이러한 지방족 알코올은, 석유화학 공급원료(petrochemical feedstock)로부터 생성된 알파-올레핀의 화학적 수화에 의해, 또는 코코넛유, 팜유, 팜핵유, 우지 및 돈지와 같은 천연 자원으로부터 생성된 지방산의 촉매 수소화를 통해 생성된다. 천연 자원으로부터 유도된 지방족 알코올은 다양한 사슬 길이를 갖는데, 이는 특정 적용들과 관련이 있으며 특수하다(specific). 지방족 알코올의 알파-올레핀으로의 탈수가 화학 촉매에 의해 달성될 수 있다.
- [0006] 지방족 알데히드는 공업용 특수 화학제들을 생성하는데 사용될 수 있다. 예를 들어, 알데히드는 통상적으로 중합체, 수지, 염료, 착향료, 가소제, 향수, 약제를 생성하는데 사용된다. 또한, 알데히드는 용제, 보존제 및 소독제로서 사용될 수 있다. 비타민 및 호르몬과 같은 몇몇 천연 및 합성 화합물이 알데히드이며, 많은 당류가 알데히드기를 함유한다. 지방족 알데히드는 화학적 또는 효소적 환원에 의해 지방족 알코올로 전환될 수 있다.
- [0007] 지방족 알데히드 및 지방족 알코올의 생성에 대한 더 환경친화적이고 더 깨끗한 대안(greener and cleaner alternative)은 발효성 당 및/또는 바이오매스를 통하는 것이다. 하지만, 발효성 당 또는 바이오매스로부터의 지방족 알데히드 및 지방족 알코올의 생성이 상업적으로 실현가능하게 하기 위해서는, 산업 공정들이 최종 생성물의 효율적인 전환 및 회수(recovery)를 위해 최적화되어야 한다. 본 발명은 생체촉매로서 조작 숙주 세포

(engineered host cell)를 이용하여 지방족 알데히드 및 지방족 알코올의 개선된 생성을 위한 방법들 및 조성물들을 제공함으로써 이러한 필요성에 대처한다(address).

발명의 내용

해결하려는 과제

[0008] 본 발명은, 정제된 지방산의 촉매 전환이 필요하지 않도록 특정 사슬 길이의 지방족 알코올 및/또는 지방족 알데히드(fmf) 바로 생성하는 광합성 및 종속영양 숙주 세포(heterotrophic host cell)를 제공한다. 이 생물학적 경로는 더 고품질의 생성물, 상당한 비용 감소 및 환경에 대한 더 적은 영향을 제공한다. 보다 구체적으로, 본 발명은 지방족 알데히드 및/또는 지방족 알코올 및 이의 조성물을 생성하는 신규한 아실-ACP 레덕타아제(AAR) 효소 변이체들을 제공한다. 또한, 특수한 AAR 변이 핵산 및 단백질 서열들 및 신규한 재조합 숙주 세포들 및 세포 배양물들 - 이는 이러한 조작 AAR 효소 변이체들을 포괄함 - 이 제공된다. 또한, 본 발명은 특정한 특성들을 갖는 지방족 알데히드 및/또는 지방족 알코올 조성물을 만들기 위해 재조합 AAR 변이-발현 숙주 세포들을 이용하는 방법들을 제공한다.

과제의 해결 수단

[0009] 본 발명의 일 실시형태는, 아실-ACP의 지방족 알데히드로의 전환을 촉진시키는 변이 아실-ACP 레덕타아제(AAR) 폴리펩티드들 - AAR 폴리펩티드는 SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 42 또는 SEQ ID NO: 44로서 제시된 대응하는 야생형 AAR 폴리펩티드 서열에 대해 적어도 90 %의 서열 동일성(sequence identity)을 가짐 -, 및 대응하는 숙주 세포에서의 야생형 AAR 폴리펩티드의 발현에 의해 생성된 지방족 알데히드 및/또는 지방족 알코올 조성물의 역가(titer)에 비해, 지방족 알데히드 및/또는 지방족 알코올 조성물의 더 높은 역가를 유도하는 재조합 숙주 세포에서 변이 AAR 폴리펩티드들을 발현시키는 방법들을 제공한다. 일 실시예에서, 유전적 조작 변이 AAR 폴리펩티드는 SEQ ID NO: 28로서 제시된 대응하는 야생형 AAR 폴리펩티드 서열에 대해 적어도 90 %의 서열 동일성을 가지며, 재조합 숙주 세포에서의 변이 AAR 폴리펩티드의 발현은 대응하는 숙주 세포에서의 야생형 AAR 폴리펩티드의 발현에 의해 생성된 역가에 비해 C12, C14 또는 C16 지방족 알코올의 더 높은 역가 또는 지방족 알데히드 및/또는 지방족 알코올 조성물의 더 높은 역가를 유도한다.

[0010] 일 실시형태에서, 변이 AAR 폴리펩티드는 SEQ ID NO: 28의 아미노산들 18, 24, 31, 34, 35, 43, 50, 63, 86, 112, 113, 116, 118, 120, 135, 148, 153, 155, 157, 159, 168, 172, 187, 188, 191, 209, 210, 211, 236, 277, 283, 285, 291, 324, 328, 335, 337 및 338 중 1 이상의 아미노산 위치들에서 돌연변이를 갖는다. 바람직한 일 실시예에서, 유전적 조작 변이 AAR 폴리펩티드는 S18W 돌연변이를 갖는다. 또 다른 바람직한 실시예에서, 유전적 조작 변이 AAR 폴리펩티드는 S18W 돌연변이를 가지며, 또한 M21L, D24E, D24Y, L31V, W34F, W35F, D43E, A50Q, C63A, C63G, C63Y, S86G, A112R, S113K, Q116G, R118Q, T120S, A135S, T148C, T148E, T148V, I153P, Q155C, Q155L, T157V, A159V, I168V, C172L, T187V, T188H, T188V, Q191A, L209R, E210Y, A211W, T236C, Q277V, E283G, E283S, A285V, M291V, A324T, A328S, Q335N, L337V 및/또는 L338W와 같은 돌연변이를 포함한다.

[0011] 또 다른 실시형태에서, SEQ ID NO: 34로서 제시된 대응하는 야생형 AAR 폴리펩티드 서열에 대해 적어도 90 %의 서열 동일성을 갖는 변이 AAR 폴리펩티드 및 재조합 숙주 세포에서 변이 AAR 폴리펩티드의 발현은 대응하는 숙주 세포에서의 야생형 AAR 폴리펩티드의 발현에 의해 생성된 역가에 비해 C12 지방족 알코올의 더 높은 역가 또는 지방족 알데히드 및/또는 지방족 알코올의 더 높은 역가를 유도한다. 변이 AAR 폴리펩티드는 SEQ ID NO: 34의 아미노산 40, 52, 58, 61, 273, 303, 339, 340, 344, 345, 346 및 588을 포함하는 아미노산 위치에서 돌연변이를 갖는다. 바람직한 일 실시예에서, 변이 AAR 폴리펩티드는 아미노산 위치에서 돌연변이 Q40V, G52V, S58V, D61E, G273E, K303G, K339L, H340P, L344A, L344D, L344S, L344T, L345R, V346P, V346G, 및/또는 S588V를 갖는다.

[0012] 본 발명의 또 다른 실시형태는, 숙주 세포가 SEQ ID NO: 28 또는 SEQ ID NO: 34의 변이 AAR 폴리펩티드를 발현시키도록 조작될 때 앞서 설명된 바와 같은 1 이상의 돌연변이들을 갖는 재조합 숙주 세포를 제공한다. 이 재조합 숙주 세포는, 변이 AAR 폴리펩티드를 발현시키기에 효과적인 조건들 하에서 탄소원을 함유한 배지에서 배양될 때, 대응하는 야생형 AAR 폴리펩티드를 발현시키는 숙주 세포에 의해 생성된 지방족 알데히드 및/또는 지방족 알코올 조성물의 역가보다 적어도 10 % 더 큰, 적어도 15 % 더 큰, 적어도 20 % 더 큰, 적어도 25 % 더

큰, 또는 적어도 30 % 더 큰 역가를 갖는 지방족 알데히드 및/또는 지방족 알코올 조성물을 생성한다. 일 실시예에서, 지방족 알데히드 및/또는 지방족 알코올 조성물은 30 g/L 내지 250 g/L의 역가에서 생성된다. 또 다른 실시예에서, 지방족 알코올 조성물은 세포외에서(extracellularly) 생성된다.

[0013] 또한, 본 발명은 앞서 설명된 바와 같은 제조합 숙주 세포를 포함하는 세포 배양물을 포함하며, 지방족 알코올 조성물은 C6, C8, C10, C12, C13, C14, C15, C16, C17, 및 C18 지방족 알코올, 예를 들어 C10:1, C12:1, C14:1, C16:1, 또는 C18:1 불포화 지방족 알코올 중 1 이상을 포함한다. 또 다른 실시예에서, 지방족 알코올 조성물은 포화 지방족 알코올을 포함한다.

[0014] 본 발명의 또 다른 실시형태는 SEQ ID NO: 57에 대해 적어도 90 %의 서열 동일성을 갖는 변이 아실-ACP 레덕타아제(AAR) 폴리펩티드를 제공하며, AAR 폴리펩티드는 아실-ACP의 지방족 알데히드로의 전환을 촉진시킨다. 일 실시예에서, 제조합 숙주 세포에서의 변이 AAR 폴리펩티드의 발현은 대응하는 야생형 숙주 세포에서의 야생형 AAR 폴리펩티드의 발현에 의해 생성된 지방족 알데히드 또는 지방족 알코올 조성물의 역가에 비해 지방족 알데히드 또는 지방족 알코올 조성물의 더 높은 역가를 유도한다. 또 다른 실시예에서, 제조합 숙주 세포에서의 변이 AAR 폴리펩티드의 발현은 대응하는 야생형 숙주 세포에서의 야생형 AAR 폴리펩티드의 발현에 의해 생성된 지방족 알데히드 또는 지방족 알코올 조성물의 역가에 비해 C12, C14 및/또는 C16 지방족 알코올 조성물인 지방족 알데히드 또는 지방족 알코올 조성물의 더 높은 역가를 유도한다. 상기 SEQ ID NO: 57의 서열은 첫번째 아미노산인 메티오닌(Met)이 아미노산 위치 '0'으로 시작한다. 또 다른 실시예에서, 변이 AAR 폴리펩티드는 아미노산 위치 18에서 돌연변이를 가지며, 돌연변이는 S18W이다.

[0015] 본 발명의 또 다른 실시형태는 SEQ ID NO: 57에 대해 적어도 90 %의 서열 동일성을 갖는 변이 아실-ACP 레덕타아제(AAR) 폴리펩티드를 제공하며, 변이 AAR 폴리펩티드는 아미노산 8, 16, 21, 24, 31, 34, 35, 43, 50, 63, 86, 112, 113, 116, 118, 120, 135, 148, 153, 154, 155, 157, 159, 168, 172, 187, 188, 191, 209, 210, 211, 236, 277, 281, 283, 285, 291, 324, 328, 335, 337 및/또는 338의 아미노산 위치에서 또 다른 돌연변이를 갖는다. 상기 SEQ ID NO: 57의 서열은 첫번째 아미노산인 메티오닌(Met)이 아미노산 위치 '0'으로 시작한다. 일 실시예에서, 돌연변이는 L8A, D16L, M21L, D24E, D24Y, D24V, D24P, L31V, L31M, W34F, W35F, D43E, A50Q, C63A, C63G, C63Y, S86G, A112R, S113K, Q116G, R118Q, T120S, A135S, T148C, T148E, T148V, I153P, T154A, Q155C, Q155L, T157V, A159V, I168V, C172L, T187V, T188H, T188V, Q191A, L209R, E210Y, A211W, T236C, Q277V, A281L, E283G, E283S, A285V, M291V, A324T, A328S, Q335N, L337V 및/또는 L338W로부터 선택된다. 바람직한 일 실시예에서, 변이 AAR 폴리펩티드는 M21L 돌연변이, C63G 돌연변이, S113K 돌연변이, T154A, 및 A281L 돌연변이(SEQ ID NO: 58)를 갖는다. 또 다른 바람직한 실시예에서, 변이 AAR 폴리펩티드는 L8A 돌연변이, M21L 돌연변이, C63G 돌연변이, S113K 돌연변이, T154A 돌연변이, 및 A281L 돌연변이(SEQ ID NO: 59)를 갖는다. 또 다른 바람직한 실시예에서, 변이 AAR 폴리펩티드는 D16L 돌연변이, M21L 돌연변이, C63G 돌연변이, S113K 돌연변이, T154A, 및 A281L 돌연변이(SEQ ID NO: 60)를 갖는다. 또 다른 바람직한 실시예에서, 변이 AAR 폴리펩티드는 L8A 돌연변이, D24V 돌연변이, C63G 돌연변이, S113K 돌연변이, Q155L 돌연변이, 및 A281L 돌연변이(SEQ ID NO: 61)를 갖는다. 또 다른 바람직한 실시예에서, 변이 AAR 폴리펩티드는 D24P 돌연변이, L31M 돌연변이, C63G 돌연변이, S113K 돌연변이, T154A 돌연변이, 및 A281L 돌연변이(SEQ ID NO: 62)를 갖는다. 또 다른 바람직한 실시예에서, 변이 AAR 폴리펩티드는 L8A 돌연변이, D16L 돌연변이, D24V 돌연변이, C63G 돌연변이, S113K 돌연변이, T154A 돌연변이, 및 A281L 돌연변이(SEQ ID NO: 63)를 갖는다. 또 다른 바람직한 실시예에서, 변이 AAR 폴리펩티드는 D24E 돌연변이, C63G 돌연변이, S113K 돌연변이, T154A 돌연변이, 및 A281L 돌연변이(SEQ ID NO: 64)를 갖는다. 상기 SEQ ID NO: 58 내지 64의 서열들은 첫번째 아미노산인 메티오닌(Met)이 아미노산 위치 '0'으로 시작한다.

[0016] 본 발명의 또 다른 실시형태는 앞서 설명된 바와 같은 변이 AAR 폴리펩티드들을 발현시키는 제조합 숙주 세포를 제공한다(위 참조). 일 실시예에서, 제조합 숙주 세포는, 변이 AAR 폴리펩티드를 발현시키기에 효과적인 조건들 하에서 탄소원을 함유한 배지에서 배양될 때, 대응하는 야생형 AAR 폴리펩티드를 발현시키는 숙주 세포에 의해 생성된 지방족 알데히드 또는 지방족 알코올 조성물의 역가보다 적어도 10 % 더 큰, 적어도 15 % 더 큰, 적어도 20 % 더 큰, 적어도 25 % 더 큰, 또는 적어도 30 % 더 큰 역가를 갖는 지방족 알데히드 또는 지방족 알코올 조성물을 생성한다. 일 실시예에서, 제조합 숙주 세포에 의해 생성된 지방족 알데히드 또는 지방족 알코올 조성물은 약 30 g/L 내지 약 250 g/L의 역가로 생성된다. 또 다른 실시예에서, 지방족 알데히드 또는 지방족 알코올 조성물은 세포외에서 생성된다.

[0017] 또한, 본 발명은 앞서 설명된 바와 같은 변이 AAR 폴리펩티드들을 발현시키는 제조합 숙주 세포를 포함하는 세포 배양물을 고려한다(위 참조). 일 실시예에서, 지방족 알코올 조성물은 포화 및/또는 불포화 지방족 알코올

을 포함한다. 일 실시예에서, 세포 배양물은 C6, C8, C10, C12, C13, C14, C15, C16, C17, 및 C18 지방족 알코올 중 1 이상을 포함하는 지방족 알코올 조성물을 포함한다. 또 다른 실시예에서, 지방족 알코올 조성물은 C10:1, C12:1, C14:1, C16:1, 및 C18:1 불포화 지방족 알코올 중 1 이상을 포함한다. 또 다른 실시예에서, 지방족 알코올 조성물은 지방족 알코올의 환원 말단으로부터 C7과 C8 사이의 탄소 사슬의 위치 7에서 이중 결합을 갖는 지방족 알코올을 포함한다.

[0018] 또한, 본 발명은 역가의 증가를 갖는 지방족 알코올 조성물을 생성하는 방법을 포함하며, 상기 방법은 탄소원을 갖는 (앞서 설명된 바와 같은) 변이 AAR-발현 숙주 세포를 배양하는 단계; 및 지방족 알코올 조성물을 수득 (harvest)하는 단계를 포함한다. 일 실시예에서, 지방족 알코올의 역가는 야생형 AAR-발현 숙주 세포에 의해 생성된 지방족 알코올 조성물의 역가보다 적어도 20 % 내지 30 % 더 크다.

[0019] 본 발명의 또 다른 실시형태는 SEQ ID NO: 65에 대해 적어도 90 %의 서열 동일성을 갖는 변이 아실-ACP 레덕타아제(AAR) 폴리펩티드를 제공하며, 폴리펩티드는 아실-ACP의 지방족 알데히드로의 전환을 촉진시킨다. 일 실시예에서, 재조합 숙주 세포에서의 변이 AAR 폴리펩티드의 발현은 대응하는 야생형 숙주 세포에서의 야생형 AAR 폴리펩티드의 발현에 의해 생성된 지방족 알데히드 또는 지방족 알코올 조성물의 역가에 비해 지방족 알데히드 또는 지방족 알코올 조성물의 더 높은 역가를 유도한다. 또 다른 실시예에서, 재조합 숙주 세포에서의 변이 AAR 폴리펩티드의 발현은 대응하는 야생형 숙주 세포에서의 야생형 AAR 폴리펩티드의 발현에 의해 생성된 지방족 알데히드 또는 지방족 알코올 조성물의 역가에 비해 C12, C14 및/또는 C16 지방족 알코올 조성물의 더 높은 역가를 유도한다. 일 특정 실시형태에서, 본 발명은 SEQ ID NO: 65에 대해 적어도 90 %의 서열 동일성을 갖는 변이 아실-ACP 레덕타아제(AAR) 폴리펩티드를 제공하며, 폴리펩티드는 아미노산 위치 61에서 돌연변이를 갖는다. 바람직한 일 실시예에서, 돌연변이는 D61E이다.

[0020] 또한, 본 발명은 SEQ ID NO: 34에 대해 적어도 90 %의 서열 동일성을 갖는 변이 아실-ACP 레덕타아제(AAR) 폴리펩티드를 포함하며, 재조합 숙주 세포에서의 변이 AAR 폴리펩티드의 발현은 대응하는 야생형 숙주 세포에서의 야생형 AAR 폴리펩티드의 발현에 의해 생성된 지방족 알코올 조성물의 역가에 비해 C12, C14 및/또는 C16 지방족 알코올 조성물의 더 높은 역가 또는 지방족 알데히드 또는 지방족 알코올 조성물의 더 높은 역가를 유도하며, AAR 폴리펩티드는 아미노산 위치 40, 52, 273, 303, 340, 344, 345, 또는 346에서 돌연변이를 갖는다. 일 실시예에서, 변이 AAR 폴리펩티드는 Q40V, G52V, G273E, K303G, H340P, L344A, L344D, L344S, L344T, L345R, V346P, 및 V346G로부터 선택된 돌연변이를 갖는다. 바람직한 일 실시예에서, 변이 AAR 폴리펩티드는 V346P(SEQ ID NO: 66)에서 돌연변이를 갖는다. 또 다른 바람직한 실시예에서, 변이 AAR 폴리펩티드는 Q40V(SEQ ID NO: 67)에서 돌연변이를 갖는다. 또 다른 바람직한 실시예에서, 변이 AAR 폴리펩티드는 돌연변이 A345R(SEQ ID NO: 68)을 갖는다. 또 다른 바람직한 실시예에서, 변이 AAR 폴리펩티드는 L344S(SEQ ID NO: 69)에서 돌연변이를 갖는다. 또 다른 바람직한 실시예에서, 변이 AAR 폴리펩티드는 V346G(SEQ ID NO: 70)에서 돌연변이를 갖는다. 또 다른 바람직한 실시예에서, 변이 AAR 폴리펩티드는 L344D(SEQ ID NO: 71)에서 돌연변이를 갖는다. 또 다른 바람직한 실시예에서, 변이 AAR 폴리펩티드는 G52V(SEQ ID NO: 72)에서 돌연변이를 갖는다. 또 다른 바람직한 실시예에서, 변이 AAR 폴리펩티드는 L344T(SEQ ID NO: 73)에서 돌연변이를 갖는다. 또 다른 바람직한 실시예에서, 변이 AAR 폴리펩티드는 K303G(SEQ ID NO: 74)에서 돌연변이를 갖는다. 또 다른 바람직한 실시예에서, 변이 AAR 폴리펩티드는 L344A(SEQ ID NO: 75)에서 돌연변이를 갖는다. 또 다른 바람직한 실시예에서, 변이 AAR 폴리펩티드는 H340P(SEQ ID NO: 76)에서 돌연변이를 갖는다. 또 다른 바람직한 실시예에서, 변이 AAR 폴리펩티드는 G273E(SEQ ID NO: 77)에서 돌연변이를 갖는다.

[0021] 본 발명의 또 다른 실시형태는 앞서 설명된 바와 같은 변이 AAR 폴리펩티드를 발현시키는 재조합 숙주 세포를 제공한다(위 참조). 일 실시예에서, 재조합 숙주 세포는, 변이 AAR 폴리펩티드를 발현시키기에 효과적인 조건들 하에서 탄소원을 함유한 배지에서 배양될 때, 대응하는 야생형 AAR 폴리펩티드를 발현시키는 숙주 세포에 의해 생성된 지방족 알데히드 또는 지방족 알코올 조성물의 역가보다 적어도 10 % 더 큰, 적어도 15 % 더 큰, 적어도 20 % 더 큰, 적어도 25 % 더 큰, 또는 적어도 30 % 더 큰 역가를 갖는 지방족 알데히드 또는 지방족 알코올 조성물을 생성한다. 또 다른 실시예에서, 지방족 알코올 조성물은 약 30 g/L 내지 약 250 g/L의 역가로 생성된다. 또 다른 실시예에서, 지방족 알코올 조성물은 세포외에서 생성된다.

[0022] 또한, 본 발명은 앞서 설명된 바와 같은 변이 AAR 폴리펩티드들을 발현시키는 재조합 숙주 세포를 갖는 세포 배양물을 고려한다(위 참조). 일 실시예에서, 지방족 알코올 조성물은 C6, C8, C10, C12, C13, C14, C15, C16, C17, 및 C18 지방족 알코올 중 1 이상을 포함한다. 또 다른 실시예에서, 지방족 알코올 조성물은 포화 또는 불포화 지방족 알코올을 포함한다. 또 다른 실시예에서, 지방족 알코올 조성물은 C10:1, C12:1, C14:1, C16:1, 및 C18:1 불포화 지방족 알코올 중 1 이상을 포함한다. 또 다른 실시예에서, 지방족 알코올 조성물은 지방족

알코올의 환원 말단으로부터 C7과 C8 사이의 탄소 사슬의 위치 7에서 이중 결합을 갖는 지방족 알코올을 포함한다.

[0023] 본 발명의 또 다른 실시형태는 역가의 증가를 갖는 지방족 알코올 조성물을 생성하는 방법을 포함하며, 상기 방법은 탄소원을 갖는 (앞서 설명된 바와 같은) AAR을 발현시키는 숙주 세포를 배양하는 단계; 및 지방족 알코올 조성물을 수득하는 단계를 포함한다. 일 실시예에서, 지방족 알코올은 야생형 AAR-발현 숙주 세포에 의해 생성된 지방족 알코올 조성물의 역가보다 적어도 20 % 내지 30 % 더 크다.

도면의 간단한 설명

[0024] 본 발명은 바람직한 실시예들을 예시하는데 도움을 주는 첨부한 도면들과 연계하여 읽을 때 가장 잘 이해될 것이다. 하지만, 본 발명은 도면들에 개시된 특정 실시예들로 제한되지 않음을 이해할 것이다.

도 1은 재조합 숙주 세포에서 지방산 유도체들에 대한 전구체로서 아실-CoA의 생성에 사용하기 위한 예시적인 생합성 경로의 개략도이다. 사이클은 말로닐-ACP 및 아세틸-CoA의 축합(condensation)에 의해 개시된다.

도 2는 예시적인 지방산 생합성 사이클의 개략도로, 신장 주기(elongation cycles)는 β -케토아실-ACP 신타아제 I(fabB) 및 β -케토아실-ACP 신타아제 II(fabF)에 의해 촉진된 아실-ACP 및 말로닐-ACP의 축합으로 시작하여 β -케토-아실-ACP를 생성하고, 이후 β -케토-아실-ACP가 NADPH-의존적 β -케토아실-ACP 레덕타아제(fabG)에 의해 환원되어 β -하이드록시-아실-ACP를 생성하며, 이는 β -하이드록시아실-ACP 탈수효소(fabA 또는 fabZ)에 의해 트랜스-2-엔오일-아실-ACP로 탈수된다. 또한, FabA는 트랜스-2-엔오일-아실-ACP를 시스-3-엔오일-아실-ACP로 이성질화(isomerize)할 수 있고, 이는 fabI를 바이패스(bypass)할 수 있으며, (통상적으로, C16의 지방족 사슬 길이 이하에 대하여) fabB에 의해 이용될 수 있어 β -케토-아실-ACP를 생성한다. 각 사이클의 최종 단계는 트랜스-2-엔오일-아실-ACP를 아실-ACP로 전환하는 NADH 또는 NADHPH-의존적 엔오일-ACP 레덕타아제(fabI)에 의해 촉진된다. 본 명세서에 설명된 방법들에서, 지방산 합성의 종결은 아실-ACP로부터 아실기의 티오에스테라아제 제거에 의해 일어나 유리 지방산(FFA)을 방출한다. 티오에스테라아제(예를 들어, tesA)는 티오에스테르 결합을 가수분해하며, 이는 설프하이드릴 결합을 통해 아실 사슬과 ACP 사이에 일어난다.

도 3은 아세틸-CoA 카르복실라아제(accABCD) 효소 복합체의 구조 및 기능을 예시한다. 비오틴 카르복실라아제는 accC 유전자에 의해 인코딩되는 한편, 비오틴 카르복실 운반 단백질(biotin carboxyl carrier protein: BCCP)은 accB 유전자에 의해 인코딩된다. 카르복실 트랜스페라아제 활성에 관련된 2 개의 서브유닛(subunit)은 accA 및 accD 유전자들에 의해 인코딩된다. BCCP의 공유 결합된 비오틴은 카르복실레이트 모이어티(carboxylate moiety)를 갖는다(carries). birA 유전자는 홀로-accB를 비오틴화한다.

도 4는 아실-ACP로 시작하는 지방족 알코올의 생성을 위한 예시적인 생합성 경로의 개략도이며, 지방족 알데히드의 생성은 아실-ACP 레덕타아제(AAR) 또는 티오에스테라아제 및 카르복실산 레덕타아제(Car)의 효소 활성에 의해 촉진된다. 지방족 알데히드는 알데히드 레덕타아제(또한, 알코올 디하이드로제나아제라고도 함)에 의해 지방족 알코올로 전환된다. 이 경로는 지방족 아실-CoA 신타아제(fadD)를 포함하지 않는다.

도 5는 시네토크쿠스 엘롱가투스(*Synechococcus elongatus*) 아실-ACP 레덕타아제(AAR_7942)를 발현시키고, 다양한 시아노박테리아 아실기 운반 단백질(ACP)을 공동발현시키는(coexpressing) 대장균 DV2의 지방족 알코올 생성을 나타낸다.

도 6은 시네토크쿠스 엘롱카투스 아실-ACP 레덕타아제(AAR_7942)를 발현시키고, 코리네박테리움 글루탐미쿰(*Corynebacterium glutamicum*)(pSL9-185-D+)으로부터 아세틸 카르복실라아제 복합체를 동시발현시키는 대장균 DV2의 지방족 알코올 생성을 나타낸다(3 개의 개별 균주들이 도시되어 있음, D+1, D+2 및 D+3 참조).

도 7은 ifab 및 ifadR 과발현에 의한 AAR-매개에 의존하는 재조합 숙주 세포들의 개선된 지방족 알코올 생성을 예시하는 결과들을 나타낸다.

도 8은 ACP, A1rA[알코올 디하이드로제나아제(ADH)] 및 균주 Shu2에서의 합성 acc 오페론으로 공동발현되는 조합 라이브러리(AAR_Com 2a-d)로부터 유도된 AAR_7942 변이체들을 발현시키는 재조합 숙주 세포들에서의 상승된 지방족 알코올 수준을 나타낸 결과들을 나타낸다. 다수의 알코올 디하이드로제나아제 폴리펩티드는 본 발명에 따라 유용하며, *아시네토크박터* 종 M-1(SEQ ID NO: 52)의 A1rA 및 A1rAadp1(SEQ ID NO: 53)와 같은 A1rA 동족체를 포함한다(이로 제한되지 않음).

도 9a 및 도 9b는 AAR_7942 변이 S18W의 발현이 변경된 활성 및 사슬 길이 분포를 유도함을 나타내는 FALC 역가

(도 9a) 및 글루코오스의 산출량(도 9b)을 예시하는 AAR_7942 변이 S18W의 탱크 발효로부터 얻어진 결과들을 나타낸다.

도 10은 MED4_AAR의 D61E 변이체가 재조합 숙주 세포들에서 발현되었을 때 C16 내지 C14의 FALC에 대한 사슬 길이 분포의 시프트(shift)를 예시하는 결과들을 나타낸다.

발명을 실시하기 위한 구체적인 내용

[0025] 개관(General Overview)

[0026] 석유화학제품에 대한 의존성을 없애는 한가지 방법은, 소형 생성 숙주(miniature production host)로서 역할하는 환경-친화적인 미생물들을 통해 지방족 알데히드 및 지방족 알코올과 같은 지방산 유도체를 생성하는 것이다. 이러한 세포 숙주들(즉, 재조합 숙주 세포들 또는 생성 균주들)은 재생가능한 공급원료(예를 들어, 발효성 당, 탄수화물, 바이오매스, 셀룰로오스, 글리세롤, CO, CO₂, 등)와 같은 재생가능한 공급원으로부터 지방족 알데히드 및/또는 지방족 알코올을 생성하도록 조작되었다. 이러한 지방족 알데히드 및 지방족 알코올은 세제 및 연료를 포함하는 다수의 산업 제품들에 대한 원재료(raw material)이다.

[0027] 본 발명은, 재조합 숙주 세포들에서 발현될 때, 지방족 알데히드 및/또는 지방족 알코올 조성물의 개선된 역가, 수율 및/또는 생산성을 유도하는 아실-ACP 레덕타아제(AAR) 효소 변이체에 관한 것이다. 본 명세서에서, 향상된 지방족 알데히드 및/또는 지방족 알코올 생합성은, 변이 아실-ACP 레덕타아제(AAR) 단백질을 발현시키도록 숙주 세포들을 형질전환시킴으로써 달성되며, 이는 아실-ACP의 지방족 알데히드 및/또는 지방족 알코올로의 반응을 촉진시킨다. 또한, 본 발명은 AAR 효소 변이체들을 발현시키는 재조합 숙주 세포들 또는 생성 균주들에 관한 것이다.

[0028] 정의

[0029] 본 명세서 및 첨부된 청구항에서 사용되는 바와 같이, 단수 형태 "a", "an" 및 "the"는 본문에 명확히 다르게 명시되지 않는다면 복수의 지시대상(referent)을 포함한다. 따라서, 예를 들어 "숙주 세포"에 대한 언급은 이러한 2 이상의 숙주 세포를 포함하고, "지방족 에스테르"에 대한 언급은 1 이상의 지방족 에스테르 또는 에스테르의 혼합물을 포함하며, "핵산 서열"에 대한 언급은 1 이상의 핵산 서열을 포함하고, "효소"에 대한 언급은 1 이상의 효소를 포함한다.

[0030] 본 명세서 전반에 걸쳐 수탁 번호는 (여기에서 "NCBI 수탁 번호" 또는 대안적으로 "GenBank 수탁 번호"로 식별되는) 미국 국립보건원에 의해 유지되는 NCBI(미국 국립 생물공학 정보센터)에 의해 제공되는데이터베이스로부터 그리고 (여기에서 "UniProtKB 수탁 번호"로 식별되는) 스위스 생물정보학 연구소에 의해 제공되는 UniProt 지식베이스(UniProtKB) 및 Swiss-Prot 데이터베이스로부터 얻어졌다.

[0031] EC(Enzyme Classification) 번호는 생화학 및 분자생물학 국제 연합(International Union of Biochemistry and Molecular Biology: IUBMB)의 명명 위원회(Nomenclature Committee)에 의해 제정되며, 이의 설명은 월드 와이드 웹의 IUBMB 효소 명명 웹사이트에서 이용가능하다. EC 번호는 촉진된 반응에 따라 효소들을 분류한다. 예를 들어, 아실-ACP 레덕타아제(AAR) 효소 활성은 E.C. 1.2.1.80(또한, 긴-사슬형 아실-[아실기-운반-단백질] 레덕타아제라고도 알려짐) 또는 EC 1.2.1.42로 분류된다. AAR의 기능은 하나의 종으로부터 다음 종으로 대부분의 원핵생물들에 보존된다. 따라서, 상이한 미생물 종이 E.C. 1.2.1.80 또는 EC 1.2.1.42로 분류된 동일한 AAR 효소 활성을 수행할 수 있다.

[0032] 본 명세서에서 사용되는 바와 같이, "뉴클레오타이드"라는 용어는 헤테로사이클릭 염기, 당 및 1 이상의 인산기로 이루어진 폴리뉴클레오타이드의 단위체 단위를 지칭한다. 자연적으로 발생하는 염기들[구아닌(G), 아데닌(A), 사이토신(C), 티민(T) 및 우라실(U)]은 통상적으로 퓨린 또는 피리미딘의 유도체들이지만, 자연적으로 및 비-자연적으로 발생하는 염기 유사체(base analog)들도 포함되는 것으로 이해하여야 한다. 자연적으로 발생하는 당은 펜토오스(5-탄당)의 (DNA를 형성하는) 디옥시리보오스 또는 (RNA를 형성하는) 리보오스이지만, 자연적으로 및 비-자연적으로 발생하는 당 유사체들도 포함되는 것으로 이해하여야 한다. 핵산은 통상적으로 핵산 또는 폴리뉴클레오타이드를 형성하는 인산 결합을 통해 연결되지만, 많은 다른 연결들(예를 들어, 포스포로티오에이트, 보라노포스페이트 등)이 해당 기술분야에 알려져 있다.

[0033] "폴리뉴클레오타이드"라는 용어는 리보뉴클레오타이드(RNA) 또는 디옥시리보뉴클레오타이드(DNA)의 중합체를 지칭하고, 이는 외가닥 또는 이중 가닥일 수 있으며, 비-자연적 또는 변경된 뉴클레오타이드들을 함유할 수 있다. 본 명세서에서, "폴리뉴클레오타이드", "핵산 서열" 및 "뉴클레오타이드 서열"이라는 용어는 여하한 길이의 뉴클레오타이드

드의 중합체 형태, RNA 또는 DNA 중 하나를 지칭하는데 교환가능하게 사용된다. 이러한 용어는 분자의 일차 구조를 지칭하고, 따라서 이중 및 외가닥의 DNA, 및 이중 및 외가닥의 RNA를 포함한다. 이 용어는 메틸화된 및/또는 캡핑된(capped) 폴리뉴클레오타이드(단, 이로 제한되지 않음)와 같은 변형된 폴리뉴클레오타이드 및 뉴클레오타이드 유사체로부터 만들어지는 RNA 또는 DNA 중 어느 하나의 유사체를 등가물로서 포함한다. 폴리뉴클레오타이드는 플라스미드, 바이러스성, 염색체의, EST, cDNA, mRNA 및 rRNA를 포함하는(단, 이로 제한되지 않음) 여하한 형태로 존재할 수 있다.

[0034] 본 명세서에서 사용되는 바와 같이, "폴리펩티드" 및 "단백질"이라는 용어는 아미노산 잔기의 중합체를 지칭하는데 교환가능하게 사용된다. "재조합 폴리펩티드"라는 용어는 재조합 기술들에 의해 생성된 폴리펩티드를 지칭하며, 일반적으로 발현된 단백질을 인코딩하는 DNA 또는 RNA는 숙주 세포를 형질전환하여 폴리펩티드를 생성하는데 사용되는 적합한 발현 벡터(expression vector) 내로 삽입된다. 유사하게, "재조합 폴리뉴클레오타이드" 또는 "재조합 핵산" 또는 "재조합 DNA"라는 용어는 당업자에게 알려진 재조합 기술들에 의해 생성된다.

[0035] 본 명세서에서 사용되는 바와 같이, "동족체" 및 "상동(homologous)"이라는 용어는 대응하는 폴리뉴클레오타이드 또는 폴리펩티드 서열과 적어도 약 50 % 동일한 서열을 포함하는 폴리뉴클레오타이드 또는 폴리펩티드를 지칭한다. 바람직하게, 상동 폴리뉴클레오타이드들 또는 폴리펩티드들은 대응하는 아미노산 서열 또는 폴리뉴클레오타이드 서열과 적어도 약 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % 또는 적어도 약 99 % 상동성(homology)을 갖는 폴리뉴클레오타이드 서열들 또는 아미노산 서열들을 갖는다. 본 명세서에서 사용되는 바와 같이, 서열 "상동성" 및 서열 "동일성"은 교환가능하게 사용된다. 해당 기술분야의 당업자는 2 이상의 서열들 간의 상동성을 결정하는 방법들을 잘 알고 있다. 간명하게, 두 서열들 간의 "상동성"의 계산은 다음과 같이 수행될 수 있다. 이 서열들은 최적의 비교를 위해 정렬된다[예를 들어, 최적의 정렬을 위해 제 1 및 제 2 아미노산 또는 핵산 서열 중 하나 또는 둘 모두에 갭(gap)이 도입될 수 있으며, 비교를 위해 비-상동 서열들은 무시될 수 있다]. 바람직한 일 실시예에서, 비교를 위해 정렬되는 제 1 서열의 길이는 제 2 서열의 길이의 적어도 약 30 %, 바람직하게는 적어도 약 40 %, 더 바람직하게는 적어도 약 50 %, 훨씬 더 바람직하게는 적어도 약 60 %, 그리고 훨씬 더 바람직하게는 적어도 약 70 %, 적어도 약 80 %, 적어도 약 85 %, 적어도 약 90 %, 적어도 약 95 %, 적어도 약 98 %, 또는 약 100 %이다. 이후, 제 1 및 제 2 서열들의 대응하는 아미노산 위치들 또는 뉴클레오타이드 위치들에서의 아미노산 잔기들 또는 뉴클레오타이드들이 비교된다. 제 1 서열의 위치가 제 2 서열의 대응하는 위치와 동일한 아미노산 잔기 또는 뉴클레오타이드에 의해 점유될 때, 분자들은 그 위치에서 동일하다. 두 서열들 간의 퍼센트 상동성은 두 서열들의 최적의 정렬을 위해 도입될 필요가 있는, 갭의 개수 및 각 갭의 길이를 고려한, 서열들에 의해 공유되는 동일한 위치들의 개수의 함수이다. 두 서열들 간의 퍼센트 상동성의 결정 및 서열의 비교는 BLAST와 같은 수학적 알고리즘을 이용하여 달성될 수 있다[Altschul 외(1990) *J. Mol. Biol.* 215(3):403-410]. 또한, 두 아미노산 서열들 간의 퍼센트 상동성은 Blossum 62 매트릭스 또는 PAM250 매트릭스 중 어느 하나, 그리고 16, 14, 12, 10, 8, 6 또는 4의 갭 중량(gap weight) 및 1, 2, 3, 4, 5 또는 6의 길이 중량(length weight)을 이용하는, GCG 소프트웨어 패키지의 GAP 프로그램에 통합된 Needleman 및 Wunsch 알고리즘을 사용하여 결정될 수 있다[Needleman 및 Wunsch(1970), *J. Mol. Biol.* 48:444-453]. 또한, 두 뉴클레오타이드 서열들 간의 퍼센트 상동성은 NWSgapdna.CMP 매트릭스 그리고 40, 50, 60, 70 또는 80의 갭 중량 및 1, 2, 3, 4, 5 또는 6의 길이 중량을 이용하는, GCG 소프트웨어 패키지의 GAP 프로그램을 이용하여 결정될 수 있다. 해당 기술분야의 당업자는 초기 상동성 계산을 수행할 수 있으며, 이에 따라 알고리즘 파라미터들을 조정할 수 있다. 바람직한 파라미터들의 세트(및, 당업자가, 분자가 청구항의 상동성 제한 내에 있는지 여부를 결정하기 위해 어떤 파라미터들이 적용되어야 하는지에 대한 확신이 없는 경우에 사용되어야 하는 파라미터들의 세트)는 12의 갭 페널티(gap penalty), 4의 갭 확장 페널티 및 5의 프레임시프트 갭 페널티(frameshift gap penalty)를 갖는 Blossum 62 스코어링 매트릭스(scoring matrix)이다. 서열 정렬의 추가적인 방법들은 생물공학 분야에 알려져 있다[예를 들어, Rosenberg(2005) *BMC Bioinformatics* 6:278; Altschul 외(2005) *FEBS J.* 272(20):5101-5109].

[0036] "낮은 엄격성, 중간 엄격성, 높은 엄격성 또는 매우 높은 엄격성 조건들 하에서 혼성화한다(hybridize)"라는 용어는 혼성화 및 세정에 대한 조건들을 설명한다. 혼성화 반응을 수행하기 위한 안내는 Current Protocols in Molecular Biology[John Wiley & Sons, N.Y.(1989), 6.3.1-6.3.6]에서 찾을 수 있다. 수성 및 비-수성 방법들이 상기 참고문헌에 설명되며, 어느 하나의 방법이 이용될 수 있다. 여기에 언급된 특정 혼성화 조건들은 다음과 같다: (1) 낮은 엄격성 혼성화 조건들 -- 약 45 °C에서 6X 소듐 클로라이드/소듐 시트레이트(SSC), 이후에 적어도 50 °C에서 0.2X SSC, 0.1 % SDS로 두 번 세정(세정의 온도는 낮은 엄격성 조건들에 대해 55 °C로 증가될 수 있음); (2) 중간 엄격성 혼성화 조건들 -- 약 45 °C에서 6X SSC, 이후에 60 °C에서 0.2X SSC, 0.1 % SDS로 한번 이상 세정; (3) 높은 엄격성 혼성화 조건들 -- 약 45 °C에서 6X SSC, 이후에 65 °C에서 0.2X SSC, 0.1 %

SDS로 한번 이상 세정; 및 (4) 매우 높은 엄격성 혼성화 조건들 -- 65 °C에서 0.5 M 소듐 포스페이트, 7 % SDS, 이후에 65 °C에서 0.2X SSC, 1 % SDS로 한번 이상 세정. 다르게 명시하지 않는다면, 매우 높은 엄격성 조건들 (4)가 바람직한 조건들이다.

[0037] "내인성" 폴리펩티드는 모 세포(parental cell)(또는 숙주 세포)의 게놈(genome)에 의해 인코딩되는 폴리펩티드를 지칭한다. "외인성" 폴리펩티드는 모 세포의 게놈에 의해 인코딩되지 않은 폴리펩티드를 지칭한다. 변이 또는 돌연변이 폴리펩티드는 외인성 폴리펩티드의 일 예시이다. 따라서, 비-자연적으로-발생하는 핵산 분자가, 세포 내로 도입되면, 그 세포에 대해 외인성인 것으로 고려된다. 또한, 자연적으로-발생하는 핵산 분자는 특정 세포에 대해 외인성일 수 있다. 예를 들어, 세포 X로부터 격리된(isolated) 전체 코딩 서열은, 그 코딩 서열이 세포 Y 내로 도입되면, X 및 Y가 동일한 세포 타입이더라도, 세포 Y에 대해 외인성 핵산이다.

[0038] "과발현"이라는 용어는, 유전자가 그 유전자에 대한 내인성 전사율(endogenous transcription rate)에 비해 상승된 비율로 전사되도록 유도됨을 의미한다. 몇몇 예시들에서, 과발현은 추가적으로 유전자에 대한 내인성 번역율(endogenous translation rate)에 비해 유전자의 상승된 번역율을 포함한다. 과발현에 대한 테스트 방법들은 당업계에 잘 알려져 있으며, 예를 들어 전사된 RNA 수준이 rtPCR을 이용하여 평가되고, 단백질 수준이 SDS page 겔 분석을 이용하여 평가될 수 있다.

[0039] "이종(heterologous)"이라는 용어는 상이한 유기체, 상이한 세포 유형 또는 상이한 종으로부터 유도됨을 의미한다. 본 명세서에서 사용되는 바와 같이, 이는 주어진 유기체에 자연적으로 존재하지 않는 뉴클레오티드-, 폴리뉴클레오티드-, 폴리펩티드- 또는 단백질 서열을 지칭한다. 예를 들어, 시아노박테리아에 대해 원시(native)인 폴리뉴클레오티드 서열이 재조합 방법들에 의해 대장균의 숙주 세포로 도입될 수 있으며, 그 후 시아노박테리아로부터의 폴리뉴클레오티드는 대장균 세포(예를 들어, 재조합 세포)에 대해 이종이다. 또한, "이종"이라는 용어는 비-원시 상태에서 재조합 숙주 세포에 존재하는 뉴클레오티드-, 폴리뉴클레오티드-, 폴리펩티드- 또는 단백질 서열에 대한 언급에 사용될 수 있다. 예를 들어, "이종" 뉴클레오티드, 폴리뉴클레오티드, 폴리펩티드 또는 단백질 서열은 대응하는 야생형 숙주 세포에 자연적으로 존재하는 야생형 서열에 대해 변형될 수 있으며, 이는 예를 들어 발현의 수준, 또는 뉴클레오티드, 폴리뉴클레오티드, 폴리펩티드 또는 단백질의 서열의 변형이다.

[0040] 본 명세서에서 사용되는 바와 같이, 폴리펩티드의 "분절(fragment)"이라는 용어는 2 개의 아미노산 잔기에서부터 1 개의 아미노산 잔기를 뺀 전체 아미노산 서열에 이르는 크기 범위를 갖는 전체-길이 폴리펩티드 또는 단백질의 더 짧은 부분을 지칭한다. 본 발명의 특정 실시예들에서, 분절은 폴리펩티드 또는 단백질의 도메인[예를 들어, 기질 결합 도메인 또는 촉매 도메인]의 전체 아미노산 서열을 지칭한다.

[0041] "돌연변이유발(mutagenesis)"이라는 용어는 유기체의 유전 정보가 안정한 방식으로 변화되는 과정을 지칭한다. 단백질 코딩 핵산 서열의 돌연변이유발은 돌연변이 단백질을 생성한다. 또한, 돌연변이유발은 변형된 단백질 활성을 유도하는 비-코딩 핵산 서열의 변화를 지칭한다.

[0042] 본 명세서에서 사용되는 바와 같은 "돌연변이"는 유전자의 핵산 위치 또는 폴리펩티드 또는 단백질의 아미노산 위치의 영구적인 변화를 지칭한다. 돌연변이는 치환, 추가, 삽입 및 결실을 포함한다. 예를 들어, 아미노산 위치의 돌연변이는 일 유형의 아미노산을 다른 유형의 아미노산으로 치환한 것일 수 있다[예컨대, 세린(S)이 알라닌(A)으로 치환될 수 있고; 리신(L)이 T(트레오닌)로 치환될 수 있음]. 이와 같이, 폴리펩티드 또는 단백질은 1 이상의 돌연변이를 가질 수 있으며, 일 아미노산은 다른 아미노산으로 치환된다.

[0043] "아실-ACP 레덕타아제(AAR) 변이체" 및 "변이 아실-ACP 레덕타아제(AAR)"라는 용어는 본 명세서에서 교환가능하게 사용되며, 아미노산 서열에서 1 이상의 돌연변이들을 갖는 AAR 관련 폴리펩티드 또는 단백질을 의미한다. AAR은 아실-ACP의 지방족 알데히드 및/또는 지방족 알코올로의 환원을 촉진시키는 효소를 지칭한다. AAR 변이체는 그 폴리펩티드 서열의 1 이상의 아미노산에서의 돌연변이를 포괄할 수 있다. 세포가 AAR 변이체로 형질전환되었을 때, 이는 AAR 변이체(예를 들어, 재조합 세포)를 발현시키는 세포이다. 일 실시예에서, AAR 변이체를 발현시키는 세포에 의해 생성된 지방족 알코올의 역가 및/또는 수율은 대응하는 야생형 세포(즉, AAR 변이체를 발현시키지 않는 대응 세포)보다 적어도 2 배이다. 대장균과 같은 이종 숙주에서, 지방족 알데히드는 내인성 알코올 디하이드로제나아제에 의해 지방족 알코올로 전환될 수 있다. 또 다른 실시예에서, AAR 변이체를 발현시키는 세포에 의해 생성된 지방족 알코올의 역가 및/또는 수율은, 대응하는 야생형 세포의 역가 및/또는 수율보다 적어도 약 1 배, 적어도 약 2 배, 적어도 약 3 배, 적어도 약 4 배, 적어도 약 5 배, 적어도 약 6 배, 적어도 약 7 배, 적어도 약 8 배, 적어도 약 9 배, 또는 적어도 약 10 배 더 크다. 일 실시예에서, AAR 변이체를 발현시키는 세포에 의해 생성된 지방족 알코올의 역가 및/또는 수율은, 대응하는 야생형 세포의 역가 및/또는 수율보다 적어도 약 1 퍼센트, 적어도 약 2 퍼센트, 적어도 약 3 퍼센트, 적어도 약 4 퍼센트, 적어도 약 5 퍼

센트, 적어도 약 6 퍼센트, 적어도 약 7 퍼센트, 적어도 약 8 퍼센트, 적어도 약 9 퍼센트, 또는 적어도 약 10 퍼센트 더 크다. 또 다른 실시예에서, AAR 변이체의 발현으로 인해 재조합 세포에 생성된 지방족 알코올들의 역가 및/또는 수율은 야생형 세포의 역가 및/또는 수율보다 적어도 약 20 퍼센트 내지 적어도 약 100 퍼센트 더 크다. 특정 실시예들에서, 세포에 의해 생성된 지방족 알코올의 역가 및/또는 수율은 대응하는 야생형 세포의 역가 및/또는 수율보다 적어도 약 20 퍼센트, 적어도 약 25 퍼센트, 적어도 약 30 퍼센트, 적어도 약 35 퍼센트, 적어도 약 40 퍼센트, 적어도 약 45 퍼센트, 적어도 약 50 퍼센트, 적어도 약 55 퍼센트, 적어도 약 60 퍼센트, 적어도 약 65 퍼센트, 적어도 약 70 퍼센트, 적어도 약 75 퍼센트, 적어도 약 80 퍼센트, 적어도 약 85 퍼센트, 적어도 약 90 퍼센트, 적어도 약 95 퍼센트, 적어도 약 97 퍼센트, 적어도 약 98 퍼센트, 또는 적어도 약 100 퍼센트 더 크다.

[0044] 본 명세서에서 사용되는 바와 같이, "유전자"라는 용어는 RNA 생성물 또는 단백질 생성물 중 어느 하나를 인코딩하는 핵산 서열뿐만 아니라, RNA 또는 단백질의 발현에 영향을 주는 작동가능하게-연결된 핵산 서열들(operably-linked nucleic acid sequences)[예를 들어, 이러한 서열들은 프로모터 또는 증폭자(enhancer) 서열들을 포함함(단, 이로 제한되지 않음)] 또는 RNA 또는 단백질의 발현에 영향을 주는 서열들을 인코딩하는 작동가능하게-연결된 핵산 서열들[예를 들어, 이러한 서열들은 리보솜 결합 부위 또는 번역 조절 서열(translational control sequence)들을 포함함(단, 이로 제한되지 않음)]을 지칭한다.

[0045] 발현 조절 서열(expression control sequence)들은 해당 기술분야에 알려져 있으며, 예를 들어 숙주 세포에서 폴리뉴클레오티드 서열의 발현을 위하여 제공되는 프로모터, 증폭자, 아데닐산중합반응 신호(polyadenylation signal), 전사 종결자(transcription terminator), 내부 리보솜 유입점(internal ribosome entry sites: IRES) 등을 포함한다. 발현 조절 서열들은 특히 전사에 관련된 세포성 단백질들과 상호작용한다[Maniatis 외(1987) *Science* 236:1237-1245]. 예시적인 발현 조절 서열들은, 예를 들어 Goeddel, *Gene Expression Technology: Methods in Enzymology*, Vol. 185, Academic Press, San Diego, Calif.(1990)에 개시되어 있다. 본 발명의 방법들에서, 발현 조절 서열은 폴리뉴클레오티드 서열에 작동가능하게 연결된다. "작동가능하게 연결된"이라는 것은, 적절한 분자들(예를 들어, 전사 활성화인자 단백질들)이 발현 조절 서열에 결합될 때, 폴리뉴클레오티드 서열 및 발현 조절 서열이 유전자 발현을 허용하는 방식으로 연결되는 것을 의미한다. 작동가능하게 연결된 프로모터는, 전사 및 번역의 방향에 관하여, 선택된 폴리뉴클레오티드 서열의 상류(upstream)에 위치된다. 작동가능하게 연결된 증폭자는 선택된 폴리뉴클레오티드의 상류, 내부 또는 하류(downstream)에 위치될 수 있다.

[0046] 본 명세서에서 사용되는 바와 같이, "벡터"라는 용어는 벡터가 연결된 다른 핵산, 즉 폴리뉴클레오티드 서열을 수송할 수 있는 핵산 분자를 지칭한다. 일 유형의 유용한 벡터는 에피솜(episome)[즉, 염색체의 복제(extra-chromosomal replication)가 가능한 핵산]이다. 유용한 벡터는 벡터가 연결된 핵산의 자동 복제 및/또는 발현이 가능한 벡터이다. 작동가능하게 연결된 유전자의 발현을 지향할 수 있는 벡터는 본 명세서에서 "발현 벡터"로 칭해진다. 일반적으로, 재조합 DNA 기술들에 유용한 발현 벡터는 흔히 "플라스미드"의 형태로 되어 있으며, 이는 일반적으로 벡터 형태로 염색체에 결합되지 않는 원형 이중 가닥 DNA 루프를 지칭한다. 다른 유용한 발현 벡터가 선형 형태로 제공된다. 또한, 증가의 기능들을 제공하고, 이후에 해당 기술분야에 알려지는 이러한 다른 형태들의 발현 벡터도 포함된다. 몇몇 실시예들에서, 재조합 벡터는 폴리뉴클레오티드 서열에 작동가능하게 연결된 프로모터를 더 포함한다. 몇몇 실시예들에서, 프로모터는 발달-조절된 프로모터(developmentally-regulated promoter), 세포소기관-특이적 프로모터(organelle-specific promoter), 조직-특이적 프로모터(tissue-specific promoter), 유도성 프로모터(inducible promoter), 구성적 프로모터(constitutive promoter), 또는 세포-특이적 프로모터이다. 재조합 벡터는 통상적으로 폴리뉴클레오티드 서열에 작동가능하게 연결된 발현 조절 서열; 폴리뉴클레오티드 서열에 작동가능하게 연결된 선택 마커; 폴리뉴클레오티드 서열에 작동가능하게 연결된 마커 서열; 폴리뉴클레오티드 서열에 작동가능하게 연결된 정제 모이어티(purification moiety); 폴리뉴클레오티드 서열에 작동가능하게 연결된 분비 서열(secretion sequence); 및 폴리뉴클레오티드 서열에 작동가능하게 연결된 표적화 서열(targeting sequence)로부터 선택된 1 이상의 서열을 포함한다. 특정 실시예들에서, 뉴클레오티드 서열은 숙주 세포의 게놈 DNA 내로 안정하게 통합되며, 뉴클레오티드 서열의 발현은 조절된 프로모터 구역의 제어를 받는다. 본 명세서에 사용되는 바와 같은 발현 벡터는 숙주 세포에서의 폴리뉴클레오티드 서열의 발현에 적합한 형태로 된 본 명세서에 설명된 바와 같은 특정 폴리뉴클레오티드 서열을 포함한다. 해당 기술분야의 당업자라면, 발현 벡터의 설계는 형질전환될 숙주 세포의 선택, 원하는 폴리펩티드의 발현 수준 등과 같은 인자들에 의존할 수 있음을 이해할 것이다. 본 명세서에 설명되는 발현 벡터는 본 명세서에 설명되는 바와 같은 폴리뉴클레오티드 서열들에 의해 인코딩되는, 융합 폴리펩티드들을 포함하는 폴리펩티드들을 생성하기 위하여 숙주 세포들 내로 도입될 수 있다. 원핵생물, 예를 들어 대장균에서 폴리펩티드들을 인코딩하는 유전자의 발현은 대부분 융합 및 비-융합 폴리펩티드들 중 어느 하나의 발현을

지향하는 구성적 또는 유도성 프로모터를 포함하는 벡터들로 수행된다. 융합 벡터들은 그 안에 인코딩된 폴리펩티드에, 일반적으로는 재조합 폴리펩티드의 아미노- 또는 카르복시-말단에 다수의 아미노산들을 첨가한다. 이러한 융합 벡터들은 통상적으로 다음의 세 가지 목적: 재조합 폴리펩티드의 발현을 증가시키고; 재조합 폴리펩티드의 용해도를 증가시키며; 친화성 정제에서 리간드로서 작용함으로써 재조합 폴리펩티드의 정제를 돕는 것 중 1 이상을 제공한다. 흔히, 융합 발현 벡터에서, 융합 모이어티와 재조합 폴리펩티드의 접합점에 단백질을 분해 절단(proteolytic cleavage) 부위가 도입된다. 이는 융합 폴리펩티드의 정제 이후에 융합 모이어티로부터 재조합 폴리펩티드의 분리를 가능하게 한다. 특정 실시예들에서, 본 발명의 폴리뉴클레오티드 서열은 박테리오파지 T5로부터 유도된 프로모터에 작동가능하게 연결된다.

[0047] 특정 실시예들에서, 숙주 세포는 효모 세포이고, 발현 벡터는 효모 발현 벡터이다. 효모인 *백주효모균*(*S. cerevisiae*)에서의 발현을 위한 벡터들의 예시들은 pYepSec1[Baldari 외(1987) *EMBO J.* 6:229-234]; pMFa[Kurjan 외(1982) *Cell* 30:933-943]; pJRY88[Schultz 외(1987) *Gene* 54:113-123]; pYES2(Invitrogen Corp., San Diego, CA), 및 picZ(Invitrogen Corp., San Diego, CA)를 포함한다. 다른 실시예들에서, 숙주 세포는 곤충 세포이고, 발현 벡터는 배칼로바이러스 발현 벡터이다. 배양된 곤충 세포들(예를 들어, Sf9 세포들)에서의 단백질들의 발현에 이용가능한 배칼로바이러스 벡터들은, 예를 들어 pAc 계열(series)[Smith 외(1983) *Mol. Cell Biol.* 3:2156-2165] 및 pVL 계열[Lucklow 외(1989) *Virology* 170:31-39]을 포함한다. 또 다른 실시예에서, 본 명세서에 설명된 폴리뉴클레오티드 서열들은 포유동물 발현 벡터를 이용하여 포유동물 세포들에서 발현될 수 있다. 원핵 및 진핵 세포들 둘 모두에 대한 다른 적절한 발현계들이 당업계에 잘 알려져 있다; 예를 들어, Sambrook 외, "Molecular Cloning: A Laboratory Manual", 제 2 판, Cold Spring Harbor Laboratory(1989)를 참조한다.

[0048] 본 명세서에서 사용되는 바와 같이, "CoA" 또는 "아실-CoA"는 조효소 A(CoA)의 4'-포스포판테티오닐 모이어티의 설프하이드릴기와 알킬 사슬의 카르보닐 탄소 사이에 형성되는 아실 티오에스테르를 지칭하고, 이는 화학식 R-C(O)S-CoA를 가지며, 여기서 R은 적어도 4 개의 탄소 원자들을 갖는 여하한의 알킬기이다.

[0049] "ACP"라는 용어는 아실기 운반 단백질을 의미한다. ACP는 지방산 생합성 동안의 아실 중간체(intermediates)의 고보존성 운반체(highly conserved carrier)이며, 성장 사슬(growing chain)은 4'-포스포판테티오닐 모이어티의 말단 티올에서 티올 에스테르로서 합성 동안 결합된다. 단백질은 두 가지 형태, 즉 아포-ACP(지방산 생합성에서 비활성)및 ACP 또는 홀로(holo)-ACP(지방산 생합성에서 활성)로 존재한다. "ACP" 및 "홀로-ACP"라는 용어들은 본 명세서에서 교환가능하게 사용되며, 단백질의 활성 형태를 지칭한다. 포스포판테티오닐트랜스페라아제라고 하는 효소가 비활성 아포-ACP의 활성 홀로-ACP로의 전환에 관련된다. 더 구체적으로, ACP는 비활성 아포-ACP 형태로 발현되고, 4'-포스포판테티오닐 모이어티가 홀로-ACP를 생성하기 위해 홀로-아실기 운반 단백질 신타아제(ACPS), 포스포판테티오닐트랜스페라아제의 작용에 의해 ACP 상의 보존된 세린 잔기에 번역-후(post-translationally) 부착되어야 한다.

[0050] 본 명세서에서 사용되는 바와 같이, "아실-ACP"라는 용어는 아실기 운반 단백질(ACP)의 포스포판테티오닐 모이어티의 설프하이드릴기와 알킬 사슬의 카르보닐 탄소 사이에 형성되는 아실 티오에스테르를 지칭한다. 몇몇 실시예들에서, ACP는 완전히 포화된 아실-ACP의 합성에 있어서 중간체이다. 다른 실시예들에서, ACP는 불포화된 아실-ACP의 합성에 있어서 중간체이다. 몇몇 실시예들에서, 탄소 사슬은 약 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 또는 26 개의 탄소들을 가질 것이다.

[0051] 본 명세서에서 사용되는 바와 같이, "지방산 유도체"라는 용어는 "지방산" 또는 "지방산 유도체"를 의미하고, 이는 "지방산 또는 이의 유도체"로 칭해질 수 있다. "지방산"이라는 용어는 화학식 RCOOH를 갖는 카르복실산을 의미한다. R은 지방족기(aliphatic group), 바람직하게는 알킬기를 나타낸다. R은 약 4 내지 약 22 개의 탄소 원자들을 포함할 수 있다. 지방산들은 포화, 단일불포화 또는 다중불포화될 수 있다. "지방산 유도체"는 부분적으로 생성 숙주 유기체의 지방산 생합성 경로로부터 만들어지는 생성물이다. "지방산 유도체들"은 부분적으로 ACP, 아실-ACP 또는 아실-ACP 유도체들로부터 만들어지는 생성물들을 포함한다. 예시적인 지방산 유도체들은, 예를 들어 아실-CoA, 지방산, 지방족 알데히드, 짧은 사슬형 및 긴 사슬형 알코올, 지방족 알코올, 탄화수소, 에스테르(예를 들어, 왁스, 지방산 에스테르 또는 지방족 에스테르), 말단 올레핀, 내부 올레핀 및 케톤을 포함한다.

[0052] 본 명세서에서 사용되는 바와 같이, "지방산 생합성 경로"라는 용어는 지방산들 및 이의 유도체들을 생성하는 생합성 경로를 의미한다. 지방산 생합성 경로는 원하는 특성들을 갖는 지방산 유도체들을 생성하도록 추가 효소들을 포함할 수 있다.

- [0053] 본 명세서에서 사용되는 바와 같이, "지방족 알데히드"는 카르보닐기($C=O$)에 의해 특성화되는 화학식 $RCHO$ 를 갖는 알데히드를 의미한다. 몇몇 실시예들에서, 지방족 알데히드는 지방족 알코올로부터 만들어지는 여하한의 알데히드이다. 특정 실시예들에서, R기는 길이가 적어도 5, 적어도 6, 적어도 7, 적어도 8, 적어도 9, 적어도 10, 적어도 11, 적어도 12, 적어도 13, 적어도 14, 적어도 15, 적어도 16, 적어도 17, 적어도 18 또는 적어도 19 개의 탄소들이다. 대안적으로 또는 추가적으로, R기는 길이가 20 이하, 19 이하, 18 이하, 17 이하, 16 이하, 15 이하, 14 이하, 13 이하, 12 이하, 11 이하, 10 이하, 9 이하, 8 이하, 7 이하 또는 6 이하의 탄소들이다. 따라서, R기는 상기의 종단점들 중 어느 두 개로 한정되는 R기를 가질 수 있다. 예를 들어, R기는 길이가 6 내지 16 개의 탄소들, 길이가 10 내지 14 개의 탄소들, 또는 길이가 12 내지 18 개의 탄소들일 수 있다. 몇몇 실시예들에서, 지방족 알데히드는 C_6 , C_7 , C_8 , C_9 , C_{10} , C_{11} , C_{12} , C_{13} , C_{14} , C_{15} , C_{16} , C_{17} , C_{18} , C_{19} , C_{20} , C_{21} , C_{22} , C_{23} , C_{24} , C_{25} 또는 C_{26} 지방족 알데히드이다. 특정 실시예들에서, 지방족 알데히드는 C_6 , C_8 , C_{10} , C_{12} , C_{13} , C_{14} , C_{15} , C_{16} , C_{17} 또는 C_{18} 지방족 알데히드이다.
- [0054] 본 명세서에서 사용되는 바와 같이, "지방족 알코올"은 화학식 ROH 를 갖는 알코올을 의미한다. 몇몇 실시예들에서, R기는 길이가 적어도 5, 적어도 6, 적어도 7, 적어도 8, 적어도 9, 적어도 10, 적어도 11, 적어도 12, 적어도 13, 적어도 14, 적어도 15, 적어도 16, 적어도 17, 적어도 18 또는 적어도 19 개의 탄소들이다. 대안적으로 또는 추가적으로, R기는 길이가 20 이하, 19 이하, 18 이하, 17 이하, 16 이하, 15 이하, 14 이하, 13 이하, 12 이하, 11 이하, 10 이하, 9 이하, 8 이하, 7 이하 또는 6 이하의 탄소들이다. 따라서, R기는 상기의 종단점들 중 어느 두 개로 한정되는 R기를 가질 수 있다. 예를 들어, R기는 길이가 6 내지 16 개의 탄소들, 길이가 10 내지 14 개의 탄소들, 또는 길이가 12 내지 18 개의 탄소들일 수 있다. 몇몇 실시예들에서, 지방족 알코올은 C_6 , C_7 , C_8 , C_9 , C_{10} , C_{11} , C_{12} , C_{13} , C_{14} , C_{15} , C_{16} , C_{17} , C_{18} , C_{19} , C_{20} , C_{21} , C_{22} , C_{23} , C_{24} , C_{25} 또는 C_{26} 지방족 알코올이다. 특정 실시예들에서, 지방족 알코올은 C_6 , C_8 , C_{10} , C_{12} , C_{13} , C_{14} , C_{15} , C_{16} , C_{17} 또는 C_{18} 지방족 알코올이다.
- [0055] 본 명세서에서 언급된 바와 같은 "지방족 알코올 조성물"은 재조합 숙주 세포에 의해 생성되며, 통상적으로 지방족 알코올의 혼합물을 포함한다. 몇몇 경우, 혼합물은 1 이상의 유형의 생성물(예를 들어, 지방족 알코올 및 지방산)을 포함한다. 다른 경우, 지방산 유도체 조성물은, 예를 들어 다양한 사슬 길이 및 포화 또는 분지 특성들을 갖는 지방족 알코올의 혼합물을 포함할 수 있다. 또 다른 경우, 지방족 알코올 조성물은 다양한 사슬 길이 및 포화 또는 분지 특성들을 갖는 생성물과 1 이상의 유형의 생성물의 혼합물을 둘 모두 포함한다.
- [0056] 지방족 알데히드를 생성하도록 조작된 숙주 세포는 통상적으로 지방족 알데히드의 일부를 지방족 알코올로 전환할 것이다. 일 예시적인 실시예에서, 아실-ACP는 AAR의 작용을 통해 지방족 알데히드로 전환된다. 지방족 알데히드의 지방족 알코올로의 전환은, 예를 들어 지방족 알코올 생합성 폴리펩티드를 통해 더 용이해질 수 있다. 몇몇 실시예들에서, 지방족 알코올 생합성 폴리펩티드를 인코딩하는 유전자는 숙주 세포에서 발현 또는 과발현된다. 특정 실시예들에서, 지방족 알코올 생합성 폴리펩티드는 알데히드 레덕타아제 또는 알코올 디하이드로제나아제 활성을 갖는다.
- [0057] 본 발명에 따라 유용한 알코올 디하이드로제나아제 폴리펩티드들의 예시들은 아시네토박터 종 M-1(SEQ ID NO: 52)의 AlrA 또는 AlrAadpl(SEQ ID NO: 53)와 같은 AlrA 동족체, 그리고 내인성 대장균 알코올 디하이드로제나아제, 예컨대 YjgB(AAC77226), DkgA(NP_417485), DkgB(NP_414743), YdjL(AAC74846), YdjJ(NP_416288), AdhP(NP_415995), YhdH(NP_417719), YahK(NP_414859), YphC(AAC75598), YqhD(446856) 및 YbbO[AAC73595.1]을 포함한다(단, 이로 제한되지 않음). 추가적인 예시들은 국제 특허 출원 공개공보 WO 2007/136762, WO2008/119082 및 WO 2010/062480에 개시되어 있다. 특정 실시예들에서, 지방족 알코올 생합성 폴리펩티드는 알데히드 레덕타아제 또는 알코올 디하이드로제나아제 활성(EC 1.1.1.1)을 갖는다.
- [0058] 지방산, 지방족 알데히드 또는 지방족 알코올의 R기는 직쇄형 또는 분지쇄형일 수 있다. 분지쇄형은 1 이상의 분지점을 가질 수 있으며, 사이클릭형 분지들을 포함할 수 있다. 몇몇 실시예들에서, 분지형 지방산, 분지형 지방족 알데히드 또는 분지형 지방족 알코올은 C_6 , C_7 , C_8 , C_9 , C_{10} , C_{11} , C_{12} , C_{13} , C_{14} , C_{15} , C_{16} , C_{17} , C_{18} , C_{19} , C_{20} , C_{21} , C_{22} , C_{23} , C_{24} , C_{25} 또는 C_{26} 분지형 지방산, 분지형 지방족 알데히드 또는 분지형 지방족 알코올이다. 특정 실시예들에서, 분지형 지방산, 분지형 지방족 알데히드 또는 분지형 지방족 알코올은 C_6 , C_8 , C_{10} , C_{12} , C_{13} , C_{14} , C_{15} , C_{16} , C_{17} 또는 C_{18} 분지형 지방산, 분지형 지방족 알데히드 또는 분지형 지방족 알코올이다. 특정 실시예들에서, 분지형 지방산, 분지형 지방족 알데히드 또는 분지형 지방족 알코올의 하이드록실기는 첫번째(C_1) 위

치에 있다.

- [0059] 특정 실시예들에서, 분지형 지방산, 분지형 지방족 알데히드 또는 분지형 지방족 알코올은 이소(iso)-지방산, 이소-지방족 알데히드 또는 이소-지방족 알코올, 또는 안테이소(anteiso)-지방산, 안테이소-지방족 알데히드 또는 안테이소-지방족 알코올이다. 예시적인 실시예들에서, 분지형 지방산, 분지형 지방족 알데히드 또는 분지형 지방족 알코올은 이소-C_{7:0}, 이소-C_{8:0}, 이소-C_{9:0}, 이소-C_{10:0}, 이소-C_{11:0}, 이소-C_{12:0}, 이소-C_{13:0}, 이소-C_{14:0}, 이소-C_{15:0}, 이소-C_{16:0}, 이소-C_{17:0}, 이소-C_{18:0}, 이소-C_{19:0}, 안테이소-C_{7:0}, 안테이소-C_{8:0}, 안테이소-C_{9:0}, 안테이소-C_{10:0}, 안테이소-C_{11:0}, 안테이소-C_{12:0}, 안테이소-C_{13:0}, 안테이소-C_{14:0}, 안테이소-C_{15:0}, 안테이소-C_{16:0}, 안테이소-C_{17:0}, 안테이소-C_{18:0} 및 안테이소-C_{19:0} 분지형 지방산, 분지형 지방족 알데히드 또는 분지형 지방족 알코올로부터 선택된다.
- [0060] 분지형 또는 비분지형 지방산, 분지형 또는 비분지형 지방족 알데히드, 또는 분지형 또는 비분지형 지방족 알코올의 R기는 포화 또는 불포화될 수 있다. 불포화된 경우, R기는 1 이상의 불포화점을 가질 수 있다. 몇몇 실시예들에서, 불포화 지방산, 불포화 지방족 알데히드, 또는 불포화 지방족 알코올은 단일불포화 지방산, 단일불포화 지방족 알데히드, 또는 단일불포화 지방족 알코올이다. 특정 실시예들에서, 불포화 지방산, 불포화 지방족 알데히드, 또는 불포화 지방족 알코올은 C6:1, C7:1, C8:1, C9:1, C10:1, C11:1, C12:1, C13:1, C14:1, C15:1, C16:1, C17:1, C18:1, C19:1, C20:1, C21:1, C22:1, C23:1, C24:1, C25:1 또는 C26:1 불포화 지방산, 불포화 지방족 알데히드, 또는 불포화 지방족 알코올이다. 특정한 바람직한 실시예에서, 불포화 지방산, 불포화 지방족 알데히드, 또는 불포화 지방족 알코올은 C10:1, C12:1, C14:1, C16:1 또는 C18:1이다. 다른 실시예들에서, 불포화 지방산, 불포화 지방족 알데히드, 또는 불포화 지방족 알코올은 오메가-7 위치에서 불포화된다. 특정 실시예들에서, 불포화 지방산, 불포화 지방족 알데히드, 또는 불포화 지방족 알코올은 cis 이중 결합을 포함한다.
- [0061] 본 명세서에서 사용되는 바와 같이, "재조합 숙주 세포" 또는 "조작된 숙주 세포"는 숙주 세포, 예를 들어 지방족 알코올들을 생성하도록 변형된 미생물이다. 몇몇 실시예들에서, 재조합 숙주 세포는 1 이상의 폴리뉴클레오티드들을 포함하고, 각각의 폴리뉴클레오티드는 지방족 알데히드 및/또는 지방족 알코올 생합성 효소 활성을 갖는 폴리펩티드를 인코딩하며, 재조합 숙주 세포는 폴리뉴클레오티드들을 발현시키기에 효과적인 조건들 하에서 탄소원의 존재 시에 배양될 때 지방족 알코올 조성물을 생성한다.
- [0062] 본 명세서에서 사용되는 바와 같이, "클론"이라는 용어는 통상적으로 단일 공통 선조와 본질적으로 유전적으로 동일하고 이의 자손인 세포 또는 세포들의 그룹, 예를 들어 단일 박테리아성 세포에서 발생하는 클로닝된 박테리아성 콜로니(cloned bacterial colony)의 박테리아를 지칭한다.
- [0063] 본 명세서에서 사용되는 바와 같이, "배양물"이라는 용어는 통상적으로 생세포(viable cell)들을 포함하는 액체 배지를 지칭한다. 일 실시예에서, 배양은 제어된 조건들 하에서 사전설정된 배양 배지에서 번식하는 세포들, 예를 들어 선택된 탄소원 및/또는 질소를 포함하는 액체 배지에서 성장되는 재조합 숙주 세포들의 배양을 포함한다.
- [0064] "배양하는" 또는 "배양"이라는 용어는 액체 또는 고체 배지의 적절한 조건들 하에서 세포들의 개체군을 성장시키는 것을 지칭한다. 특정 실시예들에서, 배양은 최종 생성물로의 기질의 발효성 생물전환(bioconversion)을 지칭한다. 배양 배지는 잘 알려져 있으며, 이러한 배양 배지의 각각의 성분들은 예를 들어 DIFCO 배지 및 BBL 배지라는 이름의 상용 공급원(commercial source)들로부터 이용가능하다. 비-제한적인 일 예시에서, 수성 영양 배지(aqueous nutrient medium)는 YP 배지와 같이 질소, 염 및 탄소의 복합원을 포함하는 "풍부한 배지(rich medium)"이며, 이는 이러한 배지의 10 g/L의 펩톤 및 10 g/L 효모 추출물을 포함한다.
- [0065] 추가적으로, 숙주 세포는 미국 특허 5,000,000; 5,028,539; 5,424,202; 5,482,846; 5,602,030; 및 특허 출원 공개공보 WO 2010127318에 설명된 방법들에 따라, 탄소를 효율적으로 동화시키고 탄소원으로서 셀룰로오스 물질(cellulosic material)들을 사용하도록 조작될 수 있다. 또한, 숙주 세포는 인베르타아제를 발현시키도록 조작될 수 있어, 수크로오스가 탄소원으로서 사용될 수 있도록 한다.
- [0066] 본 명세서에서 사용되는 바와 같이, "상기 이중 뉴클레오티드 서열들을 발현시키기에 효과적인 조건들 하에서"라는 용어는 숙주 세포가 원하는 지방족 알데히드 또는 지방족 알코올을 생성하게 하는 여하한 조건들을 의미한다. 적합한 조건들은, 예를 들어 발효 조건들을 포함한다.
- [0067] 본 명세서에서 사용되는 바와 같이, 재조합 숙주 세포 내에서 단백질을, 예를 들어 효소의 "변형된" 또는 "변경된

수준"의 활성은 모(parent) 또는 원시(native) 숙주 세포에 대해 결정된 활성의 1 이상의 특성의 차이를 지칭한다. 통상적으로, 활성의 차이는 변형된 활성을 갖는 재조합 숙주 세포와 대응하는 야생형 숙주 세포 사이에서 결정된다(예를 들어, 야생형 숙주 세포에 대한 재조합 숙주 세포의 배양물의 비교). 예를 들어, 변형된 활성은 [예를 들어, 단백질을 인코딩하는 DNA 서열의 증가 또는 감소된 수의 복제, 단백질을 인코딩하는 증가 또는 감소된 수의 mRNA 전사체, 및/또는 mRNA로부터의 단백질의 증가 또는 감소된 양의 단백질 번역(protein translation)의 결과로서] 재조합 숙주 세포에 의해 발현되는 변형된 양의 단백질; 단백질의 구조 변화[예를 들어, 기질 특수성의 변화, 관찰되는 운동 파라미터들(kinetic parameters)의 변화를 유도하는 단백질 코딩 서열에 대한 변화와 같은 일차 구조에 대한 변화]; 및 단백질 안정성의 변화(예를 들어, 단백질의 증가 또는 감소된 분해)의 결과일 수 있다. 몇몇 실시예들에서, 폴리펩티드는 본 명세서에 설명된 폴리펩티드 중 어느 하나의 돌연변이체 또는 변이체이다. 특정 경우에서, 본 명세서에 설명된 바와 같은 폴리펩티드들에 대한 코딩 서열은 특정 숙주 세포에서의 발현에 최적화된 코돈(codon)이다. 예를 들어, 대장균의 발현을 위해, 1 이상의 코돈이 최적화될 수 있다[Grosjean 외(1982) Gene 18:199-209].

[0068] 통상적으로, 본 명세서에서 사용되는 바와 같은 "조절 서열(regulatory sequences)"이라는 용어는 궁극적으로 단백질의 발현을 제어하는 단백질을 인코딩하는 DNA 서열들에 작동가능하게-연결된, DNA의 염기 서열을 지칭한다. 조절 서열의 예시들은 RNA 프로모터 서열, 전사 인자 결합 서열(transcription factor binding sequences), 전사 종결 서열, (증폭자 요소와 같은) 전사의 조절인자(modulators), RNA 안정성에 영향을 주는 뉴클레오티드 서열, 및 번역 조절 서열[예컨대, 리보솜 결합 부위(예를 들어, 원핵생물의 샤인-달가노 서열(Shine-Dalgarno sequences) 또는 진핵생물의 코작 서열(Kozak sequences), 개시 코돈, 종결 코돈)을 포함한다(이로 제한되지 않음)].

[0069] 본 명세서에 사용되는 바와 같이, 상기 뉴클레오티드 서열의 발현은 야생형 뉴클레오티드 서열에 대해 변형된다"라는 어구는, 내인성 뉴클레오티드 서열의 발현 및/또는 활성 또는 이중 혹은 비-원시 폴리펩티드-인 코딩 뉴클레오티드 서열의 발현 및/또는 활성의 수준의 증가 또는 감소를 의미한다.

[0070] 본 명세서에 사용되는 바와 같이, 폴리뉴클레오티드에 대한 "발현"이라는 용어는 이것이 기능하게 한다는 것이다. 폴리펩티드(또는 단백질)를 인코딩하는 폴리뉴클레오티드는, 발현될 때, 그 폴리펩티드(또는 단백질)를 생성하기 위해 전사 및 번역될 것이다. 본 명세서에 사용되는 바와 같이, "과발현"이라는 용어는 동일한 조건들 하에서 대응하는 야생형 세포에서 정상적으로 발현되는 것보다 더 높은 농도로 세포에서 폴리뉴클레오티드 또는 폴리펩티드를 발현시키는(또는 발현되게 하는) 것을 의미한다.

[0071] "변경된 수준의 발현" 및 "변형된 수준의 발현"이라는 용어는 교환가능하게 사용되며, 폴리뉴클레오티드, 폴리펩티드 또는 탄화수소가 동일한 조건들 하에서 대응하는 야생형 세포의 농도에 비해 조작된 숙주 세포에서 상이한 농도로 존재함을 의미한다.

[0072] 본 명세서에서 사용되는 바와 같이, "역가"라는 용어는 숙주 세포 배양의 단위 부피당 생성된 지방족 알데히드 또는 지방족 알코올의 양(quantity)을 지칭한다. 본 명세서에 설명된 조성물들 및 방법들의 여하한 실시형태에서, 지방족 알코올은 약 25 mg/L, 약 50 mg/L, 약 75 mg/L, 약 100 mg/L, 약 125 mg/L, 약 150 mg/L, 약 175 mg/L, 약 200 mg/L, 약 225 mg/L, 약 250 mg/L, 약 275 mg/L, 약 300 mg/L, 약 325 mg/L, 약 350 mg/L, 약 375 mg/L, 약 400 mg/L, 약 425 mg/L, 약 450 mg/L, 약 475 mg/L, 약 500 mg/L, 약 525 mg/L, 약 550 mg/L, 약 575 mg/L, 약 600 mg/L, 약 625 mg/L, 약 650 mg/L, 약 675 mg/L, 약 700 mg/L, 약 725 mg/L, 약 750 mg/L, 약 775 mg/L, 약 800 mg/L, 약 825 mg/L, 약 850 mg/L, 약 875 mg/L, 약 900 mg/L, 약 925 mg/L, 약 950 mg/L, 약 975 mg/L, 약 1000 mg/L, 약 1050 mg/L, 약 1075 mg/L, 약 1100 mg/L, 약 1125 mg/L, 약 1150 mg/L, 약 1175 mg/L, 약 1200 mg/L, 약 1225 mg/L, 약 1250 mg/L, 약 1275 mg/L, 약 1300 mg/L, 약 1325 mg/L, 약 1350 mg/L, 약 1375 mg/L, 약 1400 mg/L, 약 1425 mg/L, 약 1450 mg/L, 약 1475 mg/L, 약 1500 mg/L, 약 1525 mg/L, 약 1550 mg/L, 약 1575 mg/L, 약 1600 mg/L, 약 1625 mg/L, 약 1650 mg/L, 약 1675 mg/L, 약 1700 mg/L, 약 1725 mg/L, 약 1750 mg/L, 약 1775 mg/L, 약 1800 mg/L, 약 1825 mg/L, 약 1850 mg/L, 약 1875 mg/L, 약 1900 mg/L, 약 1925 mg/L, 약 1950 mg/L, 약 1975 mg/L, 약 2000 mg/L(2 g/L), 3 g/L, 5 g/L, 10 g/L, 20 g/L, 30 g/L, 40 g/L, 50 g/L, 60 g/L, 70 g/L, 80 g/L, 90 g/L, 100 g/L 또는 상기 값들 중 어느 두 개의 값들에 의해 한정되는 범위의 역가로 생성된다. 다른 실시예들에서, 지방족 알데히드 또는 지방족 알코올은 100 g/L 초과, 200 g/L 초과 또는 300 g/L 초과, 또는 500 g/L, 700 g/L, 1000 g/L, 1200 g/L, 1500 g/L 또는 2000 g/L과 같은 더 높은 역가로 생성된다. 본 발명의 방법들에 따라 재조합 숙주 세포에 의해 생성된 지방족 알데히드 또는 지방족 알코올의 바람직한 역가는 5 g/L 내지 200g/L, 10 g/L 내지 150 g/L,

20 g/L 내지 120 g/L, 및 30 g/L 내지 100 g/L이다.

[0073] 본 명세서에서 사용되는 바와 같이, "숙주 세포에 의해 생성된 지방족 알데히드 또는 지방족 알코올의 수율"은 투입된 탄소원이 숙주 세포의 생성물(즉, 지방족 알코올 또는 지방족 알데히드)로 전환되는 효율을 지칭한다. 본 발명의 방법들에 따라 지방족 알코올 및/또는 지방족 알데히드를 생성하도록 조작된 숙주 세포는 적어도 약 3 %, 적어도 약 4 %, 적어도 약 5 %, 적어도 약 6 %, 적어도 약 7 %, 적어도 약 8 %, 적어도 약 9 %, 적어도 약 10 %, 적어도 약 11 %, 적어도 약 12 %, 적어도 약 13 %, 적어도 약 14 %, 적어도 약 15 %, 적어도 약 16 %, 적어도 약 17 %, 적어도 약 18 %, 적어도 약 19 %, 적어도 약 20 %, 적어도 약 21 %, 적어도 약 22 %, 적어도 약 23 %, 적어도 약 24 %, 적어도 약 25 %, 적어도 약 26 %, 적어도 약 27 %, 적어도 약 28 %, 적어도 약 29 % 또는 적어도 약 30 %, 또는 상기 값들 중 어느 두 개의 값들에 한정되는 한정되는 범위의 수율을 갖는다. 다른 실시예들에서, 지방족 알데히드 또는 지방족 알코올은 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 % 또는 그 이상을 초과하는 수율로 생성된다. 대안적으로 또는 추가적으로, 수율은 약 30 % 이하, 약 27 % 이하, 약 25 % 이하 또는 약 22 % 이하이다. 따라서, 수율은 상기 종단점들 중 어느 두 개의 값들에 의해 한정될 수 있다. 예를 들어, 본 발명의 방법들에 따라 재조합 숙주 세포에 의해 생성된 지방족 알코올 또는 지방족 알데히드의 수율은 5 % 내지 15 %, 10 % 내지 25 %, 10 % 내지 22 %, 15 % 내지 27 %, 18 % 내지 22 %, 20 % 내지 28 %, 또는 20 % 내지 30 %일 수 있다. 본 발명의 방법들에 따라 재조합 숙주 세포에 의해 생성된 지방족 알코올의 바람직한 수율은 10 % 내지 30 %이다.

[0074] 본 명세서에서 사용되는 바와 같이, "생산성"이라는 용어는 단위 시간당 숙주 세포 배양의 단위 부피당 생성된 지방족 알데히드 또는 지방족 알코올의 양을 지칭한다. 본 명세서에 설명된 조성물들 및 방법들의 여하한의 실시형태에서, 재조합 숙주 세포에 의해 생성된 지방족 알데히드 또는 지방족 알코올의 생산성은 적어도 100 mg/L/hour, 적어도 200 mg/L/hour, 적어도 300 mg/L/hour, 적어도 400 mg/L/hour, 적어도 500 mg/L/hour, 적어도 600 mg/L/hour, 적어도 700 mg/L/hour, 적어도 800 mg/L/hour, 적어도 900 mg/L/hour, 적어도 1000 mg/L/hour, 적어도 1100 mg/L/hour, 적어도 1200 mg/L/hour, 적어도 1300 mg/L/hour, 적어도 1400 mg/L/hour, 적어도 1500 mg/L/hour, 적어도 1600 mg/L/hour, 적어도 1700 mg/L/hour, 적어도 1800 mg/L/hour, 적어도 1900 mg/L/hour, 적어도 2000 mg/L/hour, 적어도 2100 mg/L/hour, 적어도 2200 mg/L/hour, 적어도 2300 mg/L/hour, 적어도 2400 mg/L/hour, 또는 적어도 2500 mg/L/hour이다. 대안적으로, 또는 추가적으로, 생산성은 2500 mg/L/hour 이하, 2000 mg/L/hour 이하, 1500 mg/L/hour 이하, 120 mg/L/hour 이하, 1000 mg/L/hour 이하, 800 mg/L/hour 이하, 또는 600 mg/L/hour 이하이다. 따라서, 생산성은 상기 종단점들 중 어느 두 개의 값들에 의해 한정될 수 있다. 예를 들어, 본 발명의 방법들에 따라 재조합 숙주 세포에 의해 생성된 지방족 알데히드 또는 지방족 알코올의 바람직한 생산성은 500 mg/L/hour 내지 2500 mg/L/hour, 또는 700 mg/L/hour 내지 2000 mg/L/hour에서 선택된다.

[0075] "총 지방 중" 및 "총 지방산 생성물"이라는 용어는 국제 특허 출원 공개공보 WO 2008/119082에 개시된 GC-FID에 의해 평가되는 바와 같이 샘플에 존재하는 지방족 알코올, 지방족 알데히드, 유리 지방산 및 지방족 에스테르의 전체 양과 관련하여 본 명세서에서 교환가능하게 사용될 수 있다. 샘플들은 상황에 따라 이러한 화합물들 중 1 개, 2 개, 3 개, 또는 4 개를 함유할 수 있다.

[0076] 본 명세서에서 사용되는 바와 같이, "글루코오스 이용률"이라는 용어는 그램/리터/시간(g/L/hr)으로 기록되는 단위 시간당 배양에 사용되는 글루코오스의 양을 의미한다.

[0077] 본 명세서에서 사용되는 바와 같이, "탄소원"이라는 용어는 원핵 또는 단순 진핵 세포 성장을 위한 탄소원으로 사용되는데 적합한 기질 또는 화합물을 지칭한다. 탄소원은 중합체, 탄수화물, 산, 알코올, 알데히드, 케톤, 아미노산, 펩티드 및 기체(예를 들어, CO 및 CO₂)를 포함하는(단, 이로 제한되지 않음) 다양한 형태로 존재할 수 있다. 예시적인 탄소원은 글루코오스, 프럭토오스, 만노오스, 갈락토오스, 자일로오스 및 아라비노오스와 같은 단당류; 프럭토-올리고당 및 갈락토-올리고당과 같은 올리고당류; 녹말, 셀룰로오스, 펙틴 및 자일란과 같은 다당류; 수크로오스, 말토오스, 셀로비오스 및 투라노오스(turanose)와 같은 이당류; 헤미셀룰로오스(hemicellulose), 메틸 셀룰로오스 및 소듐 카르복시메틸 셀룰로오스와 같은 셀룰로오스 물질 및 변이체들; 속시네이트, 락테이트 및 아세테이트와 같은 포화 또는 불포화 지방산류; 에탄올, 메탄올 및 글리세롤과 같은 알코올류, 또는 이의 혼합물들을 포함한다(단, 이로 제한되지 않음). 또한, 탄소원은 글루코오스와 같은 광합성의 산물일 수 있다. 특정 실시예들에서, 탄소원은 플루 가스(flu gas)로부터 나오는 CO를 함유한 기체 혼합물이다. 또 다른 실시예에서, 탄소원은 바이오매스, 석탄, 또는 천연 가스와 같은 탄소 함유 물질의 개량(reformation)으로부터 나오는 CO를 함유한 기체 혼합물이다. 다른 실시예들에서, 탄소원은 합성가스, 메탄,

또는 천연 가스이다. 바람직한 특정 실시예들에서, 탄소원은 바이오매스이다. 다른 바람직한 실시예들에서, 탄소원은 글루코오스이다. 다른 바람직한 실시예들에서, 탄소원은 수크로오스이다. 다른 실시예들에서, 탄소원은 글리세롤이다. 다른 바람직한 실시예들에서, 탄소원은 사탕수수 주스, 사탕수수 시럽, 또는 옥수수 시럽이다. 다른 바람직한 실시예들에서, 탄소원은 CO₂, CO, 글루코오스, 수크로오스, 자일로오스, 아라비노오스, 글리세롤, 마노오스, 또는 이의 혼합물들과 같은 재생가능한 공급원료로부터 유도된다. 다른 실시예들에서, 탄소원은 녹말, 셀룰로오스 바이오매스, 당밀, 및 셀룰로오스 바이오매스의 가수분해로부터 유도된 탄수화물 혼합물들 또는 식물이나 천연 오일 처리로부터 유도된 폐기물을 포함한 다른 탄수화물 공급원들을 포함하는 재생가능한 공급원료로부터 유도된다.

[0078] 본 명세서에서 사용되는 바와 같이, "바이오매스"라는 용어는 탄소원이 유도되는 여하한 생물학적 물질을 지칭한다. 몇몇 실시예들에서, 바이오매스는 탄소원으로 처리되며, 이는 생물전환에 적합하다. 다른 실시예들에서, 바이오매스는 탄소원으로의 추가적인 처리를 요구하지 않는다. 예시적인 바이오매스의 공급원은 옥수수, 사탕수수 또는 스위치그래스(switchgrass)와 같은 식물성 물질 또는 식생(vegetation)이다. 또 다른 예시적인 바이오매스의 공급원은 동물성 물질[예를 들어, 우분(cow manure)]과 같은 신진대사 노폐물(metabolic waste products)이다. 또 다른 예시적인 바이오매스의 공급원은 조류 및 다른 해양 식물을 포함한다. 또한, 바이오매스는 글리세롤, 발효 찌꺼기, 목초, 짚, 목재, 오수, 쓰레기, 셀룰로오스 도시 폐기물 및 음식 쓰레기(예를 들어, 비누, 오일 및 지방산)를 포함하는(단, 이로 제한되지 않음) 산업, 농업, 임업 및 가정으로부터의 폐기물을 포함한다. 또한, "바이오매스"라는 용어는 탄수화물(예를 들어, 단당류, 이당류 또는 다당류)과 같은 탄소원을 지칭할 수 있다.

[0079] 본 명세서에서 사용되는 바와 같이, (지방산 및 이의 유도체와 같은) 생성물들에 대하여 "격리된(isolated)"이라는 용어는 세포 성분, 세포 배양 배지, 또는 화학적 또는 합성 전구체로부터 분리된(separated) 생성물들을 지칭한다. 본 명세서에 설명된 방법들에 의해 생성된 지방산 및 이의 유도체는 세포질뿐만 아니라 발효액에서 비교적 혼합되지 않을 수 있다. 그러므로, 지방산 및 이의 유도체는 세포내 또는 세포외에서 유기 상(organic phase)으로 수집될 수 있다.

[0080] 본 명세서에서 사용되는 바와 같이, "정제한다", "정제된" 또는 "정제"라는 용어는, 예를 들어 격리 또는 분리 의해 그 환경으로부터 분자의 제거 또는 격리를 의미한다. "실질적으로 정제된" 분자는 연계된 다른 성분들로부터 적어도 약 60 % 유리(예를 들어, 적어도 약 70 % 유리, 적어도 약 75 % 유리, 적어도 약 85 % 유리, 적어도 약 90 % 유리, 적어도 약 95 % 유리, 적어도 약 97 % 유리, 적어도 약 99 % 유리)된다. 본 명세서에서 사용되는 바와 같이, 또한 이러한 용어는 샘플로부터 오염물의 제거를 지칭한다. 예를 들어, 오염물의 제거는 샘플에서의 지방족 알코올과 같은 지방산 유도체의 비율의 증가를 유도할 수 있다. 예를 들어, 지방산 유도체가 재조합 숙주 세포에서 생성될 때, 지방산 유도체는 숙주 세포 단백질 또는 다른 숙주 세포 물질의 제거에 의해 정제될 수 있다. 정제 후, 샘플에서의 지방산 유도체의 비율이 증가된다. "정제한다", "정제된" 및 "정제"라는 용어는 절대 순도를 필요로 하지 않는 상대적인 용어이다. 따라서, 예를 들어 지방산 유도체가 재조합 숙주 세포에 생성될 때, 정제된 지방산 유도체는 다른 세포 성분(예를 들어, 핵산, 폴리펩티드, 지질, 탄수화물 또는 다른 탄화수소)로부터 실질적으로 분리되는 지방산 유도체이다.

[0081] 증가된 지방족 알코올 생성

[0082] 본 발명은 숙주 세포에서의 아실-ACP 레덕타아제(AAR) 유전자의 변형된 발현의 결과로서 향상된 지방족 알코올 조성물의 생성을 제공한다. AAR은 지방족 알데히드 및 지방족 알코올의 생성을 위한 생합성 경로와 관련된다. 변이 AAR은 지방족 알데히드를 지방족 알코올로 전환하는 생합성 경로와 관련된 또 다른 유전자의 변형된 발현과 조합하여 또는 단독으로 사용된다. 본 명세서에서, 본 발명은 재조합 숙주 세포들을 제공하며, 이는 야생형 AAR 또는 AAR과 동일한 기능을 갖는 다른 지방족 알코올 생합성 폴리펩티드들을 발현시키는 비-조작된 또는 원시 또는 야생형 숙주 세포들에 대해 향상된 지방족 알코올 생합성을 제공하기 위해 변이 AAR을 발현시키도록 조작되었다. 본 발명은 재조합 숙주 세포들에 유용한 AAR 관련 폴리뉴클레오티드들 및 폴리펩티드들을 식별한다. 하지만, AAR 관련 폴리뉴클레오티드들에 대한 절대적 서열 동일성이 필수적인 것은 아님을 이해할 것이다. 예를 들어, 특정 폴리뉴클레오티드 서열의 변화가 이루어질 수 있으며, 인코딩된 폴리펩티드가 활성에 대해 스크리닝될 수 있다. 이러한 변화는 통상적으로, 이를테면 코돈 최적화를 통해 보존적 돌연변이들 및 잠재적 돌연변이들을 포함한다. 변형된 또는 돌연변이된(즉, 돌연변이) 폴리뉴클레오티드들 및 인코딩된 변이 폴리펩티드들은, 해당 기술분야에 알려진 방법들을 이용하여, 증가된 촉매 활성, 증가된 안정성 또는 감소된 억제(예를 들어, 감소된 피드백 억제)를 포함하는(단, 이로 제한되지 않음) 모 폴리펩티드에 비해 개선된 기능과 같은 원하

는 기능에 대해 스크리닝될 수 있다. 본 발명은 효소 분류(Enzyme Classification: EC) 번호에 따라 본 명세서에 설명된 지방산 생합성 경로들의 다양한 단계들(즉, 반응들)과 관련된 효소 활성을 식별하며, 이러한 EC번호에 의해 분류된 예시적인 폴리펩티드들(효소들) 및 이러한 폴리펩티드들을 인코딩하는 예시적인 폴리뉴클레오티드들을 제공한다. 수탁 번호 및/또는 서열 식별 번호(Sequence Identifier Number: SEQ ID NO)에 의해 본 명세서에서 식별되는 이러한 예시적인 폴리펩티드들 및 폴리뉴클레오티드들은 본 명세서에 설명된 재조합 숙주 세포들을 얻기 위해 모 숙주 세포들의 지방산 경로들을 조작하는데 유용하다. 하지만, 본 명세서에 설명된 폴리펩티드들 및 폴리뉴클레오티드들은 예시이며, 따라서 비-제한적임을 이해하여야 한다. 본 명세서에 설명된 예시적인 폴리펩티드들의 동족체들의 서열들은, 데이터베이스들[예를 들어, 미국 국립 생물공학 정보센터(NCBI)에 의해 제공되는 Entrez 데이터베이스, 스위스 생물정보학 연구소(Swiss Institute of Bioinformatics)에 의해 제공되는 ExPasy 데이터베이스, 브라운슈바이크 기술대학(Technical University of Braunschweig)에 의해 제공되는 BRENDA 데이터베이스, 및 교토대 및 동경대의 생물정보학 센터(Bioinformatics Center of Kyoto University and University of Tokyo)에 의해 제공되는 KEGG 데이터베이스, 이 모두는 월드 와이드 웹에서 이용가능함]을 이용하여 해당 기술분야의 당업자들이 이용가능하다. 지방족 알코올 조성물들의 향상된 생성에 적합한 재조합 숙주 세포들을 유도하는 본 명세서에 설명된 효소들과 같은 변이 AAR 지방족 알코올 생화학 효소들을 발현시키기 위해 여러 상이한 숙주 세포들이 변형될 수 있다. 다양한 세포들이 본 명세서에 설명된 바와 같은 재조합 숙주 세포에서의 사용에 적합한 폴리펩티드들을 인코딩하는 폴리뉴클레오티드 서열들을 포함한 유전 물질원을 제공할 수 있음을 이해할 것이다.

[0083] **아실-ACP 레덕타아제(AAR) 폴리펩티드 및 이의 변이체**

[0084] 일 실시형태에서, 본 발명은 원시 또는 비-원시 아실-ACP 레덕타아제(AAR) 단백질을 발현시키기 위해 숙주 세포를 조작함으로써 지방족 알데히드 및/또는 지방족 알코올과 같은 지방산 유도체들의 개선된 생성에 관한 것이다. AAR 단백질은 아실-ACP의 지방족 알데히드로의 환원을 촉진시키며, 또한 지방족 알데히드의 지방족 알코올로의 전환을 촉진시킬 수 있다(본 명세서에서 인용 참조되는 미국 특허 공개공보 20120282663을 참조). AAR 폴리펩티드 또는 AAR 폴리펩티드를 인코딩하는 폴리뉴클레오티드 서열은 원시(예를 들어, 내인성) 또는 비-원시(예를 들어, 외인성, 이종 등)일 수 있으며, 즉 이는 대응하는 야생형 숙주 세포에 자연적으로 존재하는 야생형 서열 및 이의 발현과 상이할 수 있다. 예시들은 변이 AAR(예를 들어, 돌연변이체) 및/또는 이의 발현의 수준을 유도하는 AAR 폴리뉴클레오티드, 폴리펩티드 또는 단백질의 서열의 변형을 포함한다. 본 발명은 AAR 폴리펩티드들, 동족체들 및 변이체들을 포함한다.

[0085] 일 실시예에서, 본 발명을 구현하는데 사용하기 위한 AAR 폴리펩티드는 SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 42 또는 SEQ ID NO: 44의 야생형 AAR 폴리펩티드 서열에 대해 적어도 90 %의 서열 동일성을 갖는다. 몇몇 실시예들에서, AAR은 *시네크코쿠스* 종 또는 *프로클로로코쿠스* 종으로부터 유도된다. 다른 실시예들에서, 본 발명을 구현하는데 사용하기 위한 AAR 폴리펩티드는 SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 42 또는 SEQ ID NO: 44의 야생형 AAR 폴리펩티드 서열에 대해 적어도 75 %(예를 들어, 적어도 76 %, 77 %, 78 %, 79 %, 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % 또는 적어도 99 %)의 서열 동일성을 가지며, 또한 본 명세서에 설명된 바와 같은 유용한 특징들 및/또는 특성들을 유도하는 1 이상의 치환체들을 포함할 수 있다. 일 실시예에서, 본 발명을 구현하는데 사용하기 위한 AAR 폴리펩티드는 SEQ ID NO: 28 또는 SEQ ID NO: 34의 야생형 AAR 폴리펩티드 서열에 대해 적어도 75 %(예를 들어, 적어도 76 %, 77 %, 78 %, 79 %, 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % 또는 적어도 99 %)의 서열 동일성을 갖는다. 다른 실시예들에서, 본 발명을 구현하는데 사용하기 위한 AAR 폴리펩티드는 SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 42 또는 SEQ ID NO: 44에 대해 100 %의 서열 동일성을 갖는다. 또 다른 실시예들에서, 개선된 또는 변이 AAR 폴리펩티드 서열은 *시네크코쿠스* 종 또는 *프로클로로코쿠스* 종 이외의 종으로부터 유도된다.

[0086] 관련된 실시예에서, 본 발명은 SEQ ID NO: 27, SEQ ID NO: 29, SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 35, SEQ ID NO: 37, SEQ ID NO: 39, SEQ ID NO: 41 또는 SEQ ID NO: 43에 대해 적어도 75 %(예를 들어, 적어도 76 %, 77 %, 78 %, 79 %, 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % 또는 적어도 99 %)의 서열 동일성을 갖는 핵산 서열에 의해 인코딩된 아미노산 서열을 갖는 AAR 폴리펩티드들을 포함한다. 몇몇 실시예들에서, 핵산 서열은 본 명세서에 설명된 바와 같은

개선된 특징들 및/또는 특성들을 유도하는 1 이상의 치환체들을 갖는 AAR 변이체를 인코딩한다. 또 다른 관련된 실시예에서, 본 발명을 구현하는데 사용하기 위한 AAR 폴리펩티드는 SEQ ID NO: 27, SEQ ID NO: 29, SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 35, SEQ ID NO: 37, SEQ ID NO: 39, SEQ ID NO: 41 또는 SEQ ID NO: 43에 대해 100 %의 서열 동일성을 갖는 뉴클레오티드 서열에 의해 인코딩된다. 또 다른 실시형태에서, 본 발명은 SEQ ID NO: 27, SEQ ID NO: 29, SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 35, SEQ ID NO: 37, SEQ ID NO: 39, SEQ ID NO: 41 또는 SEQ ID NO: 43에 대응하는 핵산의 실질적인 전체 길이에 걸쳐 엄격한 조건들 하에서 혼성화하는 핵산에 의해 인코딩된 아미노산 서열을 포함하는 AAR 폴리펩티드에 관한 것이다.

[0087] 바람직한 일 실시예에서, 본 발명은 SEQ ID NO: 57 내지 SEQ ID NO: 78 중 어느 하나의 변이 AAR 폴리펩티드 서열에 대해 적어도 90 %의 서열 동일성을 갖는 본 발명을 구현하는데 사용하기 위한 AAR 폴리펩티드를 제공한다. 몇몇 실시예들에서, AAR은 *시네키크코쿠스* 종 또는 *프로클로로코쿠스* 종으로부터 유도된다. 다른 실시예들에서, 본 발명을 구현하는데 사용하기 위한 AAR 폴리펩티드는 SEQ ID NO: 57 내지 SEQ ID NO: 78 중 어느 하나의 변이 AAR 폴리펩티드 서열에 대해 적어도 75 % (예를 들어, 적어도 76 %, 77 %, 78 %, 79 %, 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % 또는 적어도 99 %)의 서열 동일성을 갖는다. 변이 AAR 폴리펩티드는 본 명세서에 설명된 바와 같은 유용한 특징들 및/또는 특성들을 유도하는 1 이상의 치환체(들)를 포함할 수 있다. 또 다른 바람직한 실시예에서, 본 발명을 구현하는데 사용하기 위한 AAR 폴리펩티드는 SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 59 또는 SEQ ID NO: 65의 변이 AAR 폴리펩티드 서열에 대해 적어도 75 % (예를 들어, 적어도 76 %, 77 %, 78 %, 79 %, 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % 또는 적어도 99 %)의 서열 동일성을 갖는다. 다른 실시예들에서, 본 발명을 구현하는데 사용하기 위한 AAR 폴리펩티드는 SEQ ID NO: 57 내지 SEQ ID NO: 78 중 어느 하나에 대해 100 %의 서열 동일성을 갖는다. 또 다른 실시예들에서, 개선된 또는 변이 AAR 폴리펩티드 서열은 *시네키크코쿠스* 종 또는 *프로클로로코쿠스* 종 이외의 종으로부터 유도된다.

[0088] 본 발명자들은 야생형 AAR₇₉₄₂를 넘어서는 개선점들을 갖는 변이체들을 스크리닝하기 위해 *시네키크코쿠스 엘롱가투스* PCC7942(AAR₇₉₄₂)로부터 아실-ACP 레딕타아제의 오류 유발 라이브러리(error prone library)를 만들었다(아래의 예시 3을 참조). 이러한 개선점들은, 전반적으로 지방족 알코올 역가를 개선시키거나, 역가에 실질적인 영향을 주지 않고 C10, C12, C14 또는 C16 지방족 알코올들의 분율을 증가시키는 것으로서 분류될 수 있다. 오류 유발 라이브러리는 아미노산 위치 18을 포함하는 다양한 아미노산 위치들을 식별한다. 이러한 위치들을 더 테스트하기 위해 포화 및 조합 라이브러리들이 제조(prepare)되었다. SEQ ID NO: 57은 아미노산 18의 돌연변이를 갖는 AAAR 변이체(돌연변이체)에 대한 아미노산 서열에 대해 코딩한다. 세린이 트립토판으로 대체된 돌연변이 S18W는 세포들에서 발현될 때 지방족 알코올 생성의 상당한 증가를 유도한다(아래의 예시 3 및 4 참조). 더 구체적으로, S18W 돌연변이는 대조군으로서 사용된 야생형 AAR에 비해 총 지방족 알코올(FALC) 생성물의 227 % 증가 및 C14 지방족 알코올의 324 % 증가를 야기하였다(아래의 표 4A 및 4B 참조).

[0089] 본 발명자들은 C12 지방족 알코올들의 분율 또는 전반적인 FALC 역가를 더 증가시킨 변이체들(돌연변이체들)을 식별하기 위해 S18W 돌연변이에 기초하여 포화 라이브러리들을 만들었다(아래의 예시 4 및 표 5 참조). 주형(template)으로서 S18W 돌연변이(SEQ ID NO: 57)를 이용한 조합 라이브러리들은 총 FALC 역가 및/또는 C12 지방족 알코올 생성물의 더 상당한 증가를 나타낸 7 개의 돌연변이체들을 생성하였다(아래의 표 6B 참조). 7 개의 조합 돌연변이체들은 돌연변이 S18W(SEQ ID NO: 57)를 갖는 AAR; 돌연변이들 M21L, C63G, S113K, T154A, A281L(SEQ ID NO: 58)을 갖는 AAR; 돌연변이들 L8A, M21L, C63G, A77A(GCC 내지 GCA 잠재성 코돈 돌연변이), S113K, T154A, A281L(SEQ ID NO: 59)을 갖는 AAR; 돌연변이들 D16L, M21L, C63G, S113K, T154A, A281L(SEQ ID NO: 60)을 갖는 AAR; 돌연변이들 L8A, D24V, C63G, S113K, Q155L, A281L(SEQ ID NO: 61)을 갖는 AAR; 돌연변이들 D24P, L31M, C63G, S113K, T154A, A281L(SEQ ID NO: 62)을 갖는 AAR; 돌연변이들 L8A, D16L, D24V, C63G, S113K, T154A, A281L(SEQ ID NO: 63)을 갖는 AAR; 돌연변이들 D24E, C63G, S113K, T154A, A281L(SEQ ID NO: 64)을 갖는 AAR을 포함한다. 특히, SEQ ID NO: 58은 C12 지방족 알코올들의 최고 분율을 나타낸 한편, SEQ ID NO: 59는 조합 돌연변이체들의 최고 역가를 나타내었다(아래의 표 6B 참조).

[0090] 또한, 본 발명자들은 야생형 MED4_AAR을 넘어서는 개선점들을 나타낸 변이체들을 스크리닝하기 위해 *프로클로로코쿠스 마리누스* MED4_AAR로부터 아실-ACP 레딕타아제의 전체 포화 라이브러리(full saturation library)를 만들었다(아래의 예시 7을 참조). AAR 변이체들은 야생형 AAR 효소보다 더 많은 지방족 알코올들의 생성 또는 변경된 사슬 길이 프로파일, 예를 들어 C12, C14 또는 C16 지방족 알코올들의 증가된 분율을 갖는 지방족 알코올들을 생성할 수 있는 능력에 기초하여 선택되었다. 표 8(아래의 예시 7 참조)은 야생형 MED4_AAR을 넘어서는

1.4-배(fold) 내지 2.2 배 범위의 최고 FALC 역가를 생성한 16 개의 AAR 변이체들로부터 대표적인 데이터를 나타낸다. 또한, 본 발명자들은 변경된 사슬 길이 프로파일을 갖는 AAR 변이체들을 스크리닝하였으며, C16보다 짧은 사슬 길이를 갖는 FALC 종의 증가된 비율이 관심거리이다(of interest). FALC의 양의 2-3-배 증가를 야기하는 2 개의 변이 클론들이 도 10에 나타나 있다. 제조합 숙주 세포에서의 이러한 AAR 변이체들 중 하나, 즉 D61E 돌연변이체(SEQ ID NO: 65)의 발현은 더 짧은 탄소 사슬들 쪽으로 지방족 알코올 종의 사슬 길이 분포를 치우치게 했다(skewed). 모든 변이체들이 증가된 역가를 가짐에 따라 더 높은 양의 지방족 알코올을 야기하였지만, SEQ ID NO: 65만이 C14의 증가된 분율(및 더 높은 역가)을 가졌다. 표 9(아래의 예시 7 참조)는 변경된 사슬 길이를 갖는 생성물들을 생성하지 않은 MED4_AAR의 V346P 변이체 그리고 야생형(WT) MED4_AAR에 비해 MED4_AAR의 D61E 변이체를 발현시키는 제조합 숙주 세포들에 의해 생성된 FALC 사슬 길이 분포를 예시한다.

[0091] **아실기 운반 단백질(ACP)**

[0092] 대장균(*E. coli*)과 같은 숙주 세포들에서 지방산 생합성을 제한할 수 있는 인자들에 관하여 문헌들에 상반된 보고가 존재한다. 아실기 운반 단백질(ACP)들은 모든 유기체에서 어느 정도까지 보존되지만, 이들의 일차 서열(primary sequence)이 상이할 수 있다. 지방족 아실-ACP를 생성물로 전환시키기 위해 대장균 이외의 공급원들로부터의 말단 경로 효소들이 대장균에서 발현될 때, 예를 들어 지방족 아실-ACP를 향한 제조합 경로 효소의 인식, 친화성, 및/또는 대사회전(turnover)에 있어서 한계들이 존재할 수 있음이 제안되었다[Suh 외(1999) *The Plant Journal* 17(6):679-688; Salas 외(2002) *Archives of Biochemistry and Biophysics* 403:25-34 참조]. 한 가지 제안은, 지방산 생합성을 위한 주요 전구체들, 예를 들어 아세틸-CoA 및 말로닐-CoA의 한계가 지방산 유도체들의 감소된 합성을 유도할 수 있다는 점이다. 지방산 생합성을 통해 플럭스(flux)를 증가시키는 한가지 접근법은 경로에서 다양한 효소들을 조정하는 것이다(도 1 내지 도 3 참조). 아세틸-CoA 카르복실라아제(acc) 복합체 및 지방산 생합성(fab) 경로를 통한 아세틸-CoA로부터의 아실-ACP의 공급은 지방산 유도체 생성물에 영향을 줄 수 있다(도 2 참조). 예시들(이하 참조)에 자세히 설명된 바와 같이, 예시를 위해 지방족 알코올들의 생성에 관한 ACP의 과발현의 효과가 평가되었다.

[0093] ACP를 발현시키도록 조작된 숙주 세포는 지방족 알데히드 및/또는 지방족 알코올 조성물 또는 특정 지방족 알데히드 및/또는 지방족 알코올 조성물의 역가의 증가를 나타낼 수 있으며, 이러한 증가는 동일한 조건들 하에서 배양될 때 ACP를 발현시키지 않는 대응하는 숙주 세포에 의해 생성된 지방족 알데히드 및/또는 지방족 알코올 조성물의 역가보다 적어도 3 %, 적어도 4 %, 적어도 5 %, 적어도 6 %, 적어도 7 %, 적어도 8 %, 적어도 9 %, 적어도 10 %, 적어도 11 %, 적어도 12 %, 적어도 13 %, 적어도 14 %, 적어도 15 %, 적어도 16 %, 적어도 17 %, 적어도 18 %, 적어도 19 %, 적어도 20 %, 적어도 21 %, 적어도 22 %, 적어도 23 %, 적어도 24 %, 적어도 25 %, 적어도 26 %, 적어도 27 %, 적어도 28 %, 적어도 29 %, 또는 적어도 30 % 더 크다. 일 실시형태에서, 본 발명은 원시(예를 들어, 내인성) 또는 비-원시(예를 들어, 외인성, 이종 등) ACP 단백질을 발현시키기 위해 숙주 세포를 조작함으로써 지방족 알데히드 및/또는 지방족 알코올 조성물의 개선된 생성에 관한 것이다. ACP 폴리펩티드 또는 ACP 폴리펩티드를 인코딩하는 폴리뉴클레오티드 서열은 비-원시일 수 있으며, 즉 이는 대응하는 야생형 숙주 세포에 자연적으로 존재하는 야생형 서열과 상이할 수 있다. 예시들은 뉴클레오티드, 폴리펩티드 또는 단백질의 서열 또는 이의 발현의 수준의 변형을 포함한다. 본 발명은 ACP 폴리펩티드들 및 이의 동족체들을 포함한다.

[0094] 일 실시예에서, 본 발명을 구현하는데 사용하기 위한 ACP 폴리펩티드는 SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8 또는 SEQ ID NO: 10에 대해 적어도 70 %의 서열 동일성을 갖는다. 몇몇 실시예들에서, ACP는 *마리노박터 하이드로카보노클라스티쿠스*(*Marinobacter hydrocarbonoclasticus*) 또는 대장균으로부터 유도된다. 다른 실시예에서, 본 발명을 구현하는데 사용하기 위한 ACP 폴리펩티드는 SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, 또는 SEQ ID NO: 10의 야생형 ACP 폴리펩티드 서열에 대해 적어도 75 %(예를 들어, 적어도 76 %, 77 %, 78 %, 79 %, 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % 또는 적어도 99 %)의 서열 동일성을 가지며, 또한 본 명세서에 설명된 바와 같은 유용한 특징들 및/또는 특성들을 유도하는 1 이상의 치환체들을 포함할 수 있다. 일 실시형태에서, 본 발명을 구현하는데 사용하기 위한 ACP 폴리펩티드는 SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, 또는 SEQ ID NO: 10에 대해 적어도 100 %의 서열 동일성을 갖는다. 다른 실시예들에서, 개선된 또는 변이 ACP 폴리펩티드 서열은 *M. 하이드로카보노클라스티쿠스* 또는 대장균 이외의 종으로부터 유도된다. 관련 실시형태에서, 본 발명을 구현하는데 사용하기 위한 ACP 폴리펩티드는 SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, 또는 SEQ ID NO: 9에 대해 100 %의 서열 동일성을 갖는 뉴클레오티드에 의해 인코딩된다. 또 다른 관련 실시형태에서, 본 발명은 SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, 또

는 SEQ ID NO: 9에 대해 적어도 75 %(예를 들어, 적어도 76 %, 77 %, 78 %, 79 %, 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % 또는 적어도 99 %)의 서열 동일성을 갖는 핵산 서열에 의해 인코딩되는 아미노산 서열을 포함하는 ACP 폴리펩티드에 관한 것이다. 몇몇 실시예들에서, 핵산 서열은 본 명세서에 설명된 바와 같은 개선된 특징들 및/또는 특성들을 유도하는 1 이상의 치환체들을 갖는 ACP 변이체를 인코딩한다. 다른 실시예들에서, 개선된 또는 변이 ACP 핵산 서열은 *마리노박터* 종 또는 *대장균* 이외의 종으로부터 유도된다. 또 다른 실시형태에서, 본 발명은 SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, 또는 SEQ ID NO: 9에 대응하는 핵산의 실질적인 전체 길이에 걸쳐 엄격한 조건들 하에서 혼성화되는 핵산에 의해 인코딩되는 아미노산 서열을 갖는 ACP 폴리펩티드들에 관한 것이다.

[0095] 변이체 및 돌연변이체

[0096] 몇몇 실시예들에서, AAR 또는 ACP 폴리펩티드는 본 명세서에서 설명된 폴리펩티드들 중 어느 하나의 돌연변이체 또는 변이체이다. 본 명세서에서 사용되는 바와 같은 변이 또는 돌연변이 폴리펩티드는 적어도 하나의 아미노산에 의해 야생형 폴리펩티드와 상이한 아미노산 서열을 갖는 폴리펩티드를 지칭한다. 예를 들어, 돌연변이체는 다음의 보존적 아미노산 치환체들: 지방족 아미노산, 예컨대 알라닌, 발린, 류신, 및 이소류신의 또 다른 지방족 아미노산으로의 대체; 세린의 트레오닌으로의 대체; 트레오닌의 세린으로의 대체; 산성 잔기, 예컨대 아스파르트산 및 글루탐산의 또 다른 산성 잔기로의 대체; 아미드기를 지닌 잔기, 예컨대 아스파라긴 및 글루타민의 또 다른 아미드기를 지닌 잔기로의 대체; 염기성 잔기, 예컨대 리신 및 아르기닌의 또 다른 염기성 잔기로의 교체; 및 방향족 잔기, 예컨대 페닐알라닌 및 티로신의 또 다른 방향족 잔기로의 대체 중 1 이상을 포함할 수 있다(단, 이로 제한되지 않음). 몇몇 실시예들에서, 변이 또는 돌연변이 폴리펩티드는 약 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100 또는 그 이상의 아미노산 치환, 추가, 삽입 또는 결실을 갖는다. 변이체 또는 돌연변이체로서 기능하는 폴리펩티드의 몇몇 바람직한 분절은 대응하는 야생형 폴리펩티드의 생물학적 기능(예를 들어, 효소 활성)의 일부 또는 전부를 보유한다. 몇몇 실시예들에서, 분절은 대응하는 야생형 폴리펩티드의 생물학적 기능의 적어도 75 %, 적어도 80 %, 적어도 90 %, 적어도 95 %, 또는 적어도 98 % 이상을 보유한다. 다른 실시예들에서, 분절 또는 돌연변이체는 대응하는 야생형 폴리펩티드의 생물학적 기능의 약 100 %를 보유한다. 어떤 아미노산 잔기들이 생물학적 활성에 영향을 주지 않고 치환, 삽입, 또는 결실될 수 있는지를 결정하는 가이드선(guidance)는 해당 기술분야에 잘 알려진 컴퓨터 프로그램들, 예를 들어 LASERGENE 소프트웨어(DNASTAR, Inc., Madison, WI)를 이용하여 찾을 수 있다. 다른 실시예들에서, 분절은 대응하는 야생형 폴리펩티드에 비해 증가된 생물학적 기능을 나타낸다. 예를 들어, 분절은 대응하는 야생형 폴리펩티드에 비해 효소 활성에 있어서 적어도 10 %, 적어도 25 %, 적어도 50 %, 적어도 75 %, 또는 적어도 90 %의 개선을 나타낼 수 있다. 다른 실시예들에서, 분절은 대응하는 야생형 폴리펩티드에 비해 효소 활성에 있어서 적어도 100 % 또는 적어도 200 % 또는 적어도 500 %의 개선을 나타낸다.

[0097] 본 명세서에 설명된 폴리펩티드들은 추가적인 보존적 또는 비-필수 아미노산 치환체들을 가질 수 있으며, 이는 폴리펩티드 기능에 실질적인 영향을 주지 않는다는 것을 이해한다. 특정 치환이 용인되는지(즉, 아실-ACP 레딕 타아제 활성과 같은 요구되는 생물학적 기능에 부정적인 영향을 주지 않는지)의 여부가 해당 기술분야에서 설명되는 바와 같이 결정될 수 있다[Bowie 외(1990) *Science* 247:1306-1310 참조]. 보존적 아미노산 치환은 아미노산 잔기가 유사한 곁사슬(side chain)을 갖는 아미노산 잔기로 대체되는 것이다. 유사한 곁사슬들을 갖는 아미노산 잔기들의 계열(family)들이 해당 기술분야에 정의되어 있다. 이 계열들은 염기성 곁사슬(예를 들어, 리신, 아르기닌, 히스티딘), 산성 곁사슬(예를 들어, 아스파르트산, 글루탐산), 비전하 극성 곁사슬(uncharged polar side chain)(예를 들어, 글리신, 아스파라긴, 글루타민, 세린, 트레오닌, 티로신, 시스테인), 무극성 곁사슬(예를 들어, 알라닌, 발린, 류신, 이소류신, 프롤린, 페닐알라닌, 메티오닌, 트립토판), 베타-분지형 곁사슬(예를 들어, 트레오닌, 발린, 이소류신), 및 방향족 곁사슬(예를 들어, 티로신, 페닐알라닌, 트립토판, 히스티딘)을 갖는 아미노산들을 포함한다.

[0098] 변이체들은 자연적으로 발생하거나 시험관 내에서(*in vitro*) 생성될 수 있다. 특히, 이러한 변이체들은 위치 지정 돌연변이, 무작위 화학적 돌연변이유발(random chemical mutagenesis), 엑소뉴클레아제 III 결실 절차, 또는 표준 클로닝 기술들과 같은 유전자 조작 기술들을 이용하여 생성될 수 있다. 대안적으로, 이러한 변이체, 돌연변이체, 분절, 유사체, 또는 유도체는 화학 합성 또는 변형 절차들을 이용하여 생성될 수 있다. 변이체를 만드는 방법들이 해당 기술분야에 잘 알려져 있다. 예를 들어, 변이체들은 무작위 및 위치 지정 돌연변이를 이용함으로써 제조될 수 있다. 무작위 및 위치 지정 돌연변이는 일반적으로 당업계에 알려져 있다[예를 들어, Arnold(1993) *Curr. Opin. Biotech.* 4:450-455 참조]. 무작위 돌연변이유발은 오류 유발 PCR[예를 들어,

Leung 외(1989) *Technique* 1:11-15; 및 Caldwell 외(1992) *PCR Methods Applic.* 2:28-33 참조]을 이용하여 달성될 수 있다. 오류 유발 PCR에서, 실제 PCR은 DNA 폴리메라아제의 복사 정확도(copying fidelity)가 낮은 조건들 하에서 수행되어, PCR 생성물의 전체 길이를 따라 높은 점 돌연변이율(high rate of point mutations)이 얻어지도록 한다. 간명하게, 이러한 절차들에서, 돌연변이유발될 핵산(AAR 효소를 인코딩하는 폴리뉴클레오티드 서열)은 PCR 생성물의 전체 길이를 따라 높은 점 돌연변이율을 달성하기 위하여 PCR 프라이머, 반응 완충제, MgCl₂, MnCl₂, Taq 폴리메라아제, 및 적절한 농도의 dNTP와 혼합된다. 예를 들어, 이러한 반응은 돌연변이유발될 20 fmole의 핵산, 30 pmole의 각 PCR 프라이머, 50 mM KCl, 10 mM Tris HCl(pH 8.3) 및 0.01 % 젤라틴을 포함하는 반응 완충제, 7 mM MgCl₂, 0.5 mM MnCl₂, 5 단위(unit)의 Taq 폴리메라아제, 0.2 mM dGTP, 0.2 mM dATP, 1 mM dCTP, 그리고 1 mM dTTP를 이용하여 수행될 수 있다. PCR은 94 °C에서 1 분, 45 °C에서 1 분, 그리고 72 °C에서 1 분의 30 회 주기 동안 수행될 수 있다. 하지만, 이러한 파라미터들은 적절하게 변동될 수 있음을 이해할 것이다. 이후, 돌연변이유발된 핵산은 적절한 벡터 내로 클로닝되며, 돌연변이유발된 핵산에 의해 인코딩되는 폴리펩티드들의 활성이 평가된다. 위치 지정 돌연변이가 관심 있는 여하한 클로닝된 DNA의 부위-특이적 돌연변이를 생성하기 위해 올리고뉴클레오티드-지정 돌연변이유발을 이용하여 달성될 수 있다. 올리고뉴클레오티드 돌연변이유발은 당업계에서 기술되어 있다[예를 들어, Reidhaar-Olson 외(1988) *Science* 241:53-57 참조]. 간명하게, 이러한 절차들에서, 클로닝된 DNA 내로 도입될 1 이상의 돌연변이를 지닌 복수의 이중 가닥 올리고뉴클레오티드가 합성되고, 돌연변이유발될 클로닝된 DNA(예를 들어, AAR 폴리펩티드를 인코딩하는 폴리뉴클레오티드 서열) 내로 삽입된다. 돌연변이유발된 DNA를 함유한 클론들이 회수되고, 이들이 인코딩한 폴리펩티드의 활성이 평가된다.

[0099]

변이체를 생성하는 또 다른 방법은 조립 PCR이다. 조립 PCR은 작은 DNA 분절들의 혼합으로부터 PCR 생성물의 조립을 수반한다. 다수의 상이한 PCR 반응이 동일한 바이알(vial)에서 병행하여 일어나며, 일 반응의 생성물이 또 다른 반응의 생성물을 프라이밍한다(priming)(미국 특허 5,965,408 참조). 변이체를 생성하는 또 다른 방법은 유성(sexual) PCR 돌연변이유발이다. 유성 PCR 돌연변이유발에서는, 강제된(forced) 상동성 재조합이 서열 상동성에 기초한 DNA 분자의 무작위 분절화(random fragmentation)의 결과로서 시험관 내에서 상이하지만 고도로 관련된 DNA 서열의 DNA 분자들 사이에 일어난다. 이후, PCR 반응의 프라이머 신장법에 의한 크로스오버(crossover)의 고정화 후속된다. 유성 PCR 돌연변이유발은 당업계에 알려진 출간물에 기술되어 있다[예를 들어, Stemmer(1994) *Proc. Natl. Acad. Sci., U.S.A.* 91:10747-10751]. 또한, 변이체들은 생체 내 돌연변이유발(*in vivo* mutagenesis)에 의해 생성될 수 있다. 몇몇 실시예들에서, 핵산 서열에서의 무작위 돌연변이는 DNA 복원 경로(repair pathway) 중 1 이상에서 돌연변이를 갖는(carries) 박테리아 균주, 예컨대 대장균 균주에서의 서열을 증식시킴으로써 생성된다. 이러한 돌연변이유발 유전자 균주(mutator strain)는 야생형 균주보다 높은 무작위 돌연변이율을 갖는다. 이러한 균주들 중 하나에서의 DNA 서열(예를 들어, AAR 폴리펩티드를 인코딩하는 폴리뉴클레오티드 서열)의 증식이 결국 DNA 내에서 무작위 돌연변이를 생성할 것이다. 생체 내 돌연변이유발에 사용하기 적절한 돌연변이유발 유전자 균주는 해당 기술분야의 공보(예를 들어, 국제 특허 출원 공개 공보 WO 1991/016427)에 기술되어 있다. 또한, 변이체들은 카세트 돌연변이유발(cassette mutagenesis)을 이용하여 생성될 수 있다. 카세트 돌연변이유발에서는, 이중 가닥 DNA 분자의 작은 영역이 원시 서열과 상이한 합성 올리고뉴클레오티드 카세트로 대체된다. 올리고뉴클레오티드는 흔히 완전히 및/또는 부분적으로 무작위화된 원시 서열을 포함한다. 또한, 반복 양상돌 돌연변이유발(Recursive ensemble mutagenesis)이 변이체들을 생성하는데 사용될 수 있다. 반복 양상돌 돌연변이유발은 표현형 관련 돌연변이체들(phenotypically related mutants)(이의 멤버들은 아미노산 서열에 있어서 상이함)의 다양한 개체군들을 생성하기 위해 개발된 단백질 조작(즉, 단백질 돌연변이유발)을 위한 알고리즘이다. 이 방법은 조합 카세트 돌연변이유발의 연속 라운드(successive rounds)를 제어하기 위해 피드백 메커니즘을 이용한다[예를 들어, Arkin 외(1992) *Proc. Natl. Acad. Sci., U.S.A.* 89:7811-7815 참조]. 몇몇 실시예들에서, 변이체들은 지수 양상돌 돌연변이유발(exponential ensemble mutagenesis)을 이용하여 생성된다. 지수 양상돌 돌연변이유발은 높은 비율의 독특한 기능성 돌연변이체를 갖는 조합 라이브러리를 생성하는 프로세스이며, 잔기들의 작은 그룹들이 기능성 단백질들을 야기하는 아미노산을 각각의 변경된 위치에서 식별하기 위해 병행하여 무작위화된다[예를 들어, Delegrave 외(1993) *Biotech. Res.* 11:1548-1552 참조]. 몇몇 실시예들에서, 변이체들은 별개의 폴리펩티드들을 인코딩하는 복수의 핵산들의 부분들이 함께 융합되어 키메라 폴리펩티드를 인코딩하는 키메라 핵산 서열들을 생성하도록 셔플링 절차(shuffling procedure)들을 이용하여 생성된다(예를 들어, 미국 특허 5,965,408 및 5,939,250에 개시됨).

[0100]

지방족 알데히드 및 지방족 알코올의 생성

[0101] 원시 또는 재조합 숙주 세포가 지방족 알데히드 생합성 활성을 갖는 효소(또한, 본 명세서에서 지방족 알데히드 생합성 폴리펩티드 또는 지방족 알데히드 생합성 폴리펩티드 또는 효소라고도 함)를 인코딩하는 폴리뉴클레오티드를 포함할 수 있다. 지방족 알데히드 생합성 효소가 숙주 세포에서 발현되거나 과발현될 때 지방족 알데히드가 생성된다. 재조합 숙주 세포에서의 아실-ACP 레덕타아제(AAR) 폴리펩티드의 발현 또는 과발현은 재조합 숙주 세포에 의한 지방족 알데히드의 생성을 유도할 수 있다. 일 실시예에서, 재조합 숙주 세포는 지방족 알데히드를 생성한다. 몇몇 실시예들에서, 재조합 숙주 세포에 의해 생성된 지방족 알데히드가 지방족 알코올로 전환된다. 몇몇 실시예들에서, 알데히드 레덕타아제와 같은 원시(내인성) 지방족 알데히드 생합성 폴리펩티드들이 숙주 세포(예를 들어, *대장균*) 내에 존재하며, 지방족 알데히드를 지방족 알코올로 전환하기에 효과적이다. 다른 실시예들에서, 원시(내인성) 지방족 알데히드 생합성 폴리펩티드가 과발현된다. 다른 실시예들에서, 외인성 지방족 알데히드 생합성 폴리펩티드가 재조합 숙주 세포로 도입되며, 발현되거나 과발현된다. 지방족 알데히드는 아실-ACP 레덕타아제(AAR) 활성을 갖는 폴리펩티드와 같은 지방족 알데히드 생합성 폴리펩티드를 인코딩하는 폴리뉴클레오티드를 재조합 숙주 세포에서 발현시키거나 과발현시킴으로써 생성된다. 재조합 숙주 세포에서의 AAR의 발현은 지방족 알데히드 및 지방족 알코올의 생성을 유도한다(도 4). 예시적인 AAR 폴리펩티드들은 본 명세서에 그리고 PCT 공개공보 WO2009/140695 및 WO2009/140696에 개시되어 있으며, 이는 둘 다 본 명세서에서 인용 참조된다.

[0102] 지방족 알데히드를 포함하는 조성물(지방족 알데히드 조성물)이 지방족 알데히드 생합성 효소, 예를 들어 AAR을 발현시키기에 효과적인 조건들 하에서 탄소원의 존재 시에 숙주 세포를 배양함으로써 생성된다. 지방족 알데히드를 생성하도록 조작된 재조합 숙주 세포는 통상적으로 지방족 알데히드의 일부를 지방족 알코올로 전환시킬 것이다. 몇몇 실시예들에서, 지방족 알데히드 조성물은 지방족 알데히드 및 지방족 알코올을 포함한다. 통상적으로, 지방족 알데히드 조성물은 재조합 숙주 세포들의 세포외 환경, 즉 세포 배양 배지로부터 회수된다. 몇몇 실시예들에서, 재조합 숙주 세포는 지방족 알코올 생합성 활성을 갖는 폴리펩티드(효소)(또한, 본 명세서에서 지방족 알코올 생합성 폴리펩티드 또는 지방족 알코올 생합성 효소라고도 함)를 인코딩하는 폴리뉴클레오티드를 포함하며, 지방족 알코올이 재조합 숙주 세포에 의해 생성된다. 지방족 알코올을 포함하는 조성물(즉, 지방족 알코올 조성물)이 지방족 알코올 생합성 효소를 발현시키기에 효과적인 조건들 하에서 탄소원의 존재 시에 재조합 숙주 세포를 배양함으로써 생성될 수 있다. 재조합 숙주 세포(예를 들어, *대장균*)에 존재하는 원시(예를 들어, 내인성) 알데히드 레덕타아제는 지방족 알데히드를 지방족 알코올로 전환시킬 것이다. 몇몇 실시예들에서, 숙주 세포에 존재하는 알데히드 레덕타아제와 같은 원시(예를 들어, 내인성) 지방족 알데히드 생합성 폴리펩티드는 지방족 알데히드를 지방족 알코올로 전환시키기에 충분하다. 하지만, 다른 실시예들에서 지방족 알코올은 지방족 알데히드를 지방족 알코올로 전환시키는 지방족 알코올 생합성 활성을 갖는 폴리펩티드를 인코딩하는 폴리뉴클레오티드를 재조합 숙주 세포에서 발현시키거나 과발현시킴으로써 생성된다. 예를 들어, 알코올 디하이드로제나아제(또한, 본 명세서에서 알데히드 레덕타아제라고도 함, 예를 들어 EC 1.1.1.1)가 본 발명을 구현하는데 사용될 수 있다. 본 명세서에서 사용되는 바와 같이, 알코올 디하이드로제나아제라는 용어는 지방족 알데히드의 지방족 알코올로의 전환을 촉진시킬 수 있는 폴리펩티드를 지칭한다. 당업자라면, 특정 알코올 디하이드로제나아제들이 다른 반응들도 촉진시킬 수 있고, 이러한 비-특이적 알코올 디하이드로제나아제들도 알코올 디하이드로제나아제라는 용어에 포괄된다는 것을 이해할 것이다. 본 발명에 따라 유용한 알코올 디하이드로제나아제 폴리펩티드의 예시들은 *아시네토박터* 종 *M-I*(SEQ ID NO: 52)의 AlrA 또는 AlrAadp1(SEQ ID NO: 53)과 같은 AlrA 동족체들, YjgB(AAC77226)(SEQ ID NO: 5), DkgA(NP_417485), DkgB(NP_414743), YdjL(AAC74846), YdjJ(NP_416288), AdhP(NP_415995), YhdH(NP_417719), YahK(NP_414859), YphC(AAC75598), YqhD(446856) 및 YbbO[AAC73595.1]와 같은 내인성 *대장균* 알코올 디하이드로제나아제들을 포함한다(단, 이로 제한되지 않음). 추가적인 예시들이 국제 특허 출원 공개공보 WO 2007/136762, WO 2008/119082 및 WO 2010/062480에 개시되어 있으며, 이 각각은 본 명세서에서 인용 참조된다. 특정 실시예들에서, 지방족 알코올 생합성 폴리펩티드는 알데히드 레덕타아제 또는 알코올 디하이드로제나아제 활성(EC 1.1.1.1)을 갖는다. 몇몇 실시예들에서, 원시(예를 들어, 내인성) 지방족 알코올 생합성 폴리펩티드가 과발현되며, 다른 실시예들에서는 외인성 지방족 알코올 생합성 폴리펩티드가 재조합 숙주 세포로 도입되고 발현 또는 과발현된다.

[0103] 지방족 알코올은 지방족 아실-ACP 및 지방족 아실-CoA 중간체들을 이용하는 아실-CoA 의존적 경로, 및 지방족 아실-ACP 중간체들을 이용하지만 지방족 아실-CoA 중간체는 이용하지 않는 아실-CoA 독립적 경로를 통해 생성될 수 있다. 특정 실시예들에서, 과발현된 유전자에 의해 인코딩되는 효소는 지방산 신타아제, 아실-ACP 티오에스테라아제, 지방족 아실-CoA 신타아제 및 아세틸-CoA 카르복실라아제로부터 선택된다. 또한, 지방족 알코올은 다양한 아실-ACP 또는 아실-CoA 분자들을 대응하는 일차 알코올들로 환원시킬 수 있는 효소들에 의해 자연적으로 만들어진다(본 명세서에서 명시적으로 인용 참조되는 미국 특허 공개 20100105963 및 20110206630; 및 미국

특허 8,097,439 참조). 지방족 알코올 조성물은 흔히 다른 지방산 유도체들, 예를 들어 지방족 알데히드 및/또는 지방산과 함께 지방족 알코올을 포함한다. 통상적으로, 지방족 알코올 조성물은 재조합 숙주 세포의 세포외 환경, 즉 세포 배양 배지로부터 회수된다. 몇몇 실시예들에서, 지방산 생합성과 직접적 또는 간접적으로 관련된 폴리펩티드, 예를 들어 효소의 발현이 조절(예를 들어, 발현, 과발현 또는 감쇠)되며, 이러한 조절은 지방족 알코올과 같은 관심 지방산 유도체의 더 높은 수율, 더 높은 역가 또는 더 높은 생산성을 유도한다. 효소는, 외인성 또는 이중인 지방산 생합성 폴리뉴클레오티드[예를 들어, 모 숙주 세포 이외의 유기체로부터 발생하는 폴리펩티드, 또는 모 미생물 세포에 고유한(native) 폴리펩티드의 변이체] 또는 내인성 폴리펩티드(예를 들어, 모 숙주 세포에 고유한 폴리펩티드)에 의해 인코딩될 수 있으며, 내인성 폴리펩티드는 재조합 숙주 세포에서 과발현된다. 표 1은 특정 지방족 알코올 조성물의 생성을 용이하게 하도록 재조합 숙주 세포들에서 발현될 수 있는 예시적인 단백질들의 목록을 제공한다.

표 1: 유전자 명칭(Gene Designations)

유전자 명칭	공급 유기체	효소명	수탁 번호	EC 번호	예시적인 용도
지방산 생성 증가 / 생성물 생성 증가					
<i>accA</i>	대장균, 락토코커스 (<i>Lactococci</i>)	아세틸-CoA 카복실라아제, 서브유닛 A (카복실트랜스페라아제 알파)	AAC73296, NP_414727	6.4.1.2	말로닐-CoA 생성 증가
<i>accB</i>	대장균, 락토코커스	아세틸-CoA 카복실라아제, 서브유닛 B(BCCP: biotin carboxyl carrier protein)	NP_417721	6.4.1.2	말로닐-CoA 생성 증가
<i>accC</i>	대장균, 락토코커스	아세틸-CoA 카복실라아제, 서브유닛 C(비오틴 카복실라아제)	NP_417722	6.4.1.2, 6.3.4.14	말로닐-CoA 생성 증가
<i>accD</i>	대장균, 락토코커스	아세틸-CoA 카복실라아제, 서브유닛 D (카복실트랜스페라아제 베타)	NP_416819	6.4.1.2	말로닐-CoA 생성 증가
<i>fadD</i>	대장균 W3110	아실-CoA 신타아제	AP_002424	2.3.1.86, 6.2.1.3	지방산 생성 증가
<i>fabA</i>	대장균 K12	β -하이드록시데카노일 티오에스테르 탈수효소/ 이성질화효소	NP_415474	4.2.1.60	지방족 아실-ACP/CoA 생성 증가
<i>fabB</i>	대장균	3-옥소아실-[아실기-운반-단백질] 신타아제 I	BAA16180	2.3.1.41	지방족 아실-ACP/CoA 생성 증가
<i>fabD</i>	대장균 K12	[아실기-운반-단백질] S-말로닐 트랜스페라아제	AAC74176	2.3.1.39	지방족 아실-ACP/CoA 생성 증가
<i>fabF</i>	대장균 K12	3-옥소아실-[아실기-운반-단백질] 신타아제 II	AAC74179	2.3.1.179	지방족 아실-ACP/CoA 생성 증가
<i>fabG</i>	대장균 K12	3-옥소아실-[아실기-운반-단백질] 레덕타아제	AAC74177	1.1.1.100	지방족 아실-ACP/CoA 생성 증가
<i>fabH</i>	대장균 K12	3-옥소아실-[아실기-운반-단백질]	AAC74175	2.3.1.180	지방족 아실-ACP/CoA 생성

유전자 명칭	공급 유기체	효소명	수탁 번호	EC 번호	예시적인 용도
		신타아제 III			증가
<i>fabI</i>	대장균 K12	엔오일-[아실기-운반-단백질] 레덕타아제	NP_415804	1.3.1.9	지방족 아실-ACP/CoA 생성 증가
<i>fabR</i>	대장균 K12	전사 리프레서 (transcriptional Repressor)	NP_418398	none	불포화 지방산 생성 조절
<i>fabV</i>	콜레라균	엔오일-[아실기-운반-단백질] 레덕타아제	YP_001217283	1.3.1.9	지방족 아실-ACP/CoA 생성 증가
<i>fabZ</i>	대장균 K12	(3R)-하이드록시미리스톨 아실기 운반 단백질 탈수효소	NP_414722	4.2.1.-	지방족 아실-ACP/CoA 생성 증가
<i>fadE</i>	대장균 K13	아실-CoA 디하이드로제나아제	AAC73325	1.3.99.3, 1.3.99.-	지방산 분해 감소
<i>fadR</i>	대장균	전사 조절 단백질 (transcriptional regulatory protein)	NP_415705	none	지방산 분해 차단 또는 역전(reverse)
사슬 길이 제어					
(선도 서열을 포함하거나 포함하지 않는) <i>tesA</i>	대장균	티오에스테라아제 - 선도 서열은 아미노산 1-26임	P0ADA1	3.1.2.-, 3.1.1.5	C18 사슬 길이
(선도 서열을 포함하지 않는) <i>tesA</i>	대장균	티오에스테라아제	AAC73596, NP_415027	3.1.2.-, 3.1.1.5	C18:1 사슬 길이
<i>tesA</i> (옥탄산과 복합체를 형성하는 대장균 티오에스테라아제 I 의 돌연변이체)	대장균	티오에스테라아제	L109P	3.1.2.-, 3.1.1.5	<C18 사슬 길이
<i>fatB1</i>	월계수(<i>Umbellularia californica</i>)	티오에스테라아제	Q41635	3.1.2.14	C12:0 사슬 길이
<i>fatB2</i>	쿠페아 후커리아나(<i>Cuphea hookeriana</i>)	티오에스테라아제	AAC49269	3.1.2.14	C8:0 - C10:0 사슬 길이
<i>fatB3</i>	쿠페아 후커리아나	티오에스테라아제	AAC72881	3.1.2.14	C14:0 - C16:0 사슬 길이
<i>fatB</i>	녹나무(<i>Cinnamomum camphora</i>)	티오에스테라아제	Q39473	3.1.2.14	C14:0 사슬 길이
<i>fatB</i>	애기장대	티오에스테라아제	CAA85388	3.1.2.14	C16:1 사슬 길이
<i>fatA1</i>	해바라기(<i>Helianthus annuus</i>)	티오에스테라아제	AAL79361	3.1.2.14	C18:1 사슬 길이
<i>atfata</i>	애기장대	티오에스테라아제	NP_189147, NP_193041	3.1.2.14	C18:1 사슬 길이

[0106]

유전자 명칭	공급 유기체	효소명	수탁 번호	EC 번호	예시적인 용도
<i>fatA</i>	브라씨카 준세아	티오에스테라아제	CAC39106	3.1.2.14	C18:1 사슬 길이
<i>fatA</i>	쿠파아 후커리아나	티오에스테라아제	AAC72883	3.1.2.14	C18:1 사슬 길이
<i>tesA</i>	포도박테리움 프로판둠	티오에스테라아제	YP_130990	3.1.2.14	사슬 길이
<i>tesB</i>	대장균	티오에스테라아제	NP_414986	3.1.2.14	사슬 길이
<i>fadM</i>	대장균	티오에스테라아제	NP_414977	3.1.2.14	사슬 길이
<i>yciA</i>	대장균	티오에스테라아제	NP_415769	3.1.2.14	사슬 길이
<i>ybgC</i>	대장균	티오에스테라아제	NP_415264	3.1.2.14	사슬 길이
포화 수준 제어*					
<i>Sfa</i>	대장균	<i>fabA</i> 의 억제유전자 (suppressor)	AAN79592, AAC44390	없음	단일불포화 지방산 증가
<i>fabA</i>	대장균 K12	β -하이드록시테카노일티오에스테르 탈수효소/이성질화효소	NP_415474	4.2.1.60	불포화 지방산 생성
<i>GnsA</i>	대장균	<i>secG</i> 삭제 돌연변이의 억제유전자	ABD18647.1	없음	불포화 지방산 에스테르 증가
<i>GnsB</i>	대장균	<i>secG</i> 삭제 돌연변이의 억제유전자	AAC74076.1	없음	불포화 지방산 에스테르 증가
<i>fabB</i>	대장균	3-옥소아실-[아실기-운반-단백질]신타아제 I	BAA16180	2.3.1.41	불포화 지방산 생성 조절
<i>des</i>	바실러스 서브틸러스	D5 지방족 아실 불포화효소 (desaturase)	O34653	1.14.19	불포화 지방산 생성 조절
생성물 산출: 에스테르 생성					
AT3G51970	애기장대	긴 사슬형 알코올 0-지방족아실트랜스페라아제	NP_190765	2.3.1.26	왁스 생성
<i>ELO1</i>	피치아앙구스타 (<i>Pichia angusta</i>)	지방산 연장효소 (elongase)	BAD98251	2.3.1.-	매우 긴 사슬 길이의 지방산 생성
<i>plsC</i>	맥주효모균	아실트랜스페라아제	AAA16514	2.3.1.51	왁스 생성
<i>DAGAT/DGAT</i>	애기장대	디아실글리세롤 아실트랜스페라아제	AAF19262	2.3.1.20	왁스 생성
<i>hWS</i>	호모 사피엔스 (<i>Homo sapiens</i>)	아실-CoA 왁스 알코올 아실트랜스페라아제	AAX48018	2.3.1.20	왁스 생성
<i>af11</i>	아시네토박터 중 ADP1	이기능성(bifunctional) 왁스 에스테르 신타아제/아실-CoA:디아실글리세롤아실트랜스페라아제	AAO17391	2.3.1.20	왁스 생성
<i>ES9</i>	마리노박터 하이드로카보노클라스티쿠스	왁스 에스테르 신타아제	ABO21021	2.3.1.20	왁스 생성
<i>mWS</i>	호호바(<i>Simmondsi chinensis</i>)	왁스 에스테르 신타아제	AAD38041	2.3.1.-	왁스 생성
<i>acr1</i>	아시네토박터 중 ADP1	아실-CoA 레덕타아제	YP_047869	1.2.1.42	산출 조절
<i>yqhD</i>	대장균 K12	알코올 디하이드로제나아제	AP_003562	1.1.-.-	산출 조절
<i>AAT</i>	프레저리아	알코올 0-	AAG13130	2.3.1.84	산출 조절

[0107]

유전자 명칭	공급 유기체	효소명	수탁 번호	EC 번호	예시적인 용도
	자나나사	아세틸트랜스페라아제			
생성물 산출: 지방족 알코올 산출					
		티오에스테라아제 (상기 참조)			지방산/지방족 알코올 생성 증가
BmFAR	누에나방 (<i>Bombyx mori</i>)	FAR(지방족 알코올 형성 아실-CoA 레덕타아제)	BAC79425	1.1.1.-	아실-CoA의 지방족 알코올로의 전환
acr1	아시네토박터 중 ADP1	아실-CoA 레덕타아제	YP_047869	1.2.1.42	지방족 아실-CoA의 지방족 알데히드로의 환원
yqhD	대장균 W3110	알코올 디하이드로제나아제	AP_003562	1.1.-.-	지방족 알데히드의 지방족 알코올로의 환원; 지방족 알코올 생성 증가
alrA	아시네토박터 중 ADP1	알코올 디하이드로제나아제	CAG70252	1.1.-.-	지방족 알데히드의 지방족 알코올로의 환원
BmFAR	누에나방	FAR(지방족 알코올 형성 아실-CoA 레덕타아제)	BAC79425	1.1.1.-	지방족 아실-CoA의 지방족 알코올로의 환원
GTNG_1865	지오바실러스씨 모디나트라피 칸스 NG80-2	긴 사슬형 알데히드 디하이드로제나아제	YP_001125970	1.2.1.3	지방족 알데히드의 지방족 알코올로의 환원
AAR	시네토코쿠스 엘롱가투스	아실-ACP 레덕타아제	YP_400611	1.2.1.80 1.2.1.42	지방족 아실-ACP/CoA의 지방족 알데히드로의 환원
carB	미코박테리움 스메그마티스	카르복실산 레덕타아제 (CAR) 단백질	YP_889972	6.2.1.3, 1.2.1.42	지방산의 지방족 알데히드로의 환원
FadD	대장균 K12	아실-CoA 산티아아제	NP_416319	6.2.1.3	지방산의 지방족 아실-CoA로의 활성화
atoB	에르위니아 카로도보라	아세틸-CoA 아세틸트랜스페라아제	YP_049388	2.3.1.9	부탄올의 생성
hbd	부티리비브리오 피브리올렌스	베타-하이드록시부티릴-CoA 디하이드로제나아제	BAD51424	1.1.1.157	부탄올의 생성
CPE0095	클로스트리움 퍼프린젠스	크로토네이스부티릴-CoA 디하이드로제나아제	BAB79801	4.2.1.55	부탄올의 생성
bcd	클로스트리움 베이저랑키	부티릴-CoA 디하이드로제나아제	AAM14583	1.3.99.2	부탄올의 생성
ALDH	클로스트리움 베이저랑키	조효소 A - 알데히드를 아실화하는 디하이드로제나아제	AAT66436	1.2.1.3	부탄올의 생성
AdhE	대장균 CFT073	알데히드-알코올 디하이드로제나아제	AAN80172	1.1.1.1 1.2.1.10	부탄올의 생성
생성물 이출(Product Export)					
AtMRP5	애기장대	애기장대 다중약물 내성-관련	NP_171908	없음	생성물 이출량 조절
AmiS2	로도코쿠스종	ABC 트랜스포터 (transpoter) AmiS2	JC5491	없음	생성물 이출량 조절

[0108]

유전자 명칭	공급 유기체	효소명	수탁 번호	EC 번호	예시적인 용도
<i>ATPGP1</i>	애기장대	애기장대 p 당단백질 1	NP_181228	없음	생성물 이출량 조절
<i>AcrA</i>	칸디다투스프 로도클라미디 아이모에보플라 UWE25	잠재적인 다중약물 방출 수송 단백질 (putative multidrug-efflux transport protein) <i>acrA</i>	CAF23274	없음	생성물 이출량 조절
<i>AcrB</i>	칸디다투스프 로도클라미디 아이모에보플라 UWE25	가능한 다중약물 방출 수송 단백질, <i>acrB</i>	CAF23275	없음	생성물 이출량 조절
<i>TolC</i>	야토병균 아종 노비시다	외막 단백질/세포 외피 생체내 합성	ABD59001	없음	생성물 이출량 조절
<i>AcrE</i>	시겔라소네이균 (<i>Shigella sonnei</i>) Ss046	막관통 단백질은 격막(septum) 형성 및 세포 막 투과성에 영향을 미침	YP_312213	없음	생성물 이출량 조절
<i>AcrF</i>	대장균	아크리플라빈 저항 단백질 F	P24181	없음	생성물 이출량 조절
<i>dl1619</i>	써모시네코코 쿠스앨롱게이트 [BP-1]	다중약물 방출 트랜스포터	NP_682409.1	없음	생성물 이출량 조절
<i>dl10139</i>	써모시네코코 쿠스앨롱게이트 [BP-1]	다중약물 방출 트랜스포터	NP_680930.1	없음	생성물 이출량 조절
발효					
복제 검사점 유전자					산출 효율 증가
<i>umuD</i>	시겔라소네이균 Ss046	DNA 중합효소 V, 서브유닛	YP_310132	3.4.21.-	산출 효율 증가
<i>umuC</i>	대장균	DNA 중합효소 V, 서브유닛	ABC42261	2.7.7.7	산출 효율 증가
<i>pntA, pntB</i>	시겔라플렉스 네라(<i>Shigella flexneri</i>)	NADH:NADPH 트랜스하이드로제나 아제(알파 및 베타 서브유닛)	P07001, P0AB70	1.6.1.2	산출 효율 증가
기타					
<i>fabK</i>	폐렴구균 (<i>Streptococcus pneumoniae</i>)	트랜스-2-엔오일-ACP 레덕타아제 II	AAF98273	1.3.1.9	지방산 생합성에 기여
<i>fabL</i>	바실러스 리케니포르미스 DSM 13	엔오일-(아실기 운반 단백질) 레덕타아제	AAU39821	1.3.1.9	지방산 생합성에 기여
<i>fabM</i>	스트렙토코쿠스 무탄스 (<i>Streptococcus mutans</i>)	트랜스-2, 시스-3-데세노일-ACP 이성질화효소	DAA05501	4.2.1.17	지방산 생합성에 기여

재조합 숙주 세포 및 세포 배양물

재조합 숙주 세포에 의해 지방족 알데히드 또는 지방족 알코올 조성물의 생성을 증가시키기 위한 전략들은 생성 숙주 내의 상이한 유기체들로부터의 외인성 지방족 알데히드 및 지방족 알코올 생합성 유전자들의 발현 및 원시 지방족 알데히드 또는 지방족 알코올 생합성 유전자들의 과발현에 의한 지방산 생합성 경로를 통한 증가된 플럭스를 포함한다. 본 명세서에서 사용되는 바와 같이, 재조합 숙주 세포 또는 조작된 숙주 세포라는 용어는 유전적 구성(genetic makeup)이, 예를 들어 새로운 유전적 요소들의 의도적인 도입 및/또는 숙주 세포에 자연적으로 존재하는 유전적 요소들의 의도적인 변형에 의해 대응하는 야생형 숙주 세포에 대해 변경된 숙주 세포를 지칭한다. 또한, 이러한 재조합 숙주 세포의 자손은 이러한 새로운 및/또는 변형된 유전적 요소들을 함유한다. 본 명세서에 설명된 본 발명의 실시형태들 중 어느 것에서, 숙주 세포는 식물 세포, 곤충 세포, 균류 세포[예를 들어, 칸디다 종(*Candida* sp.)과 같은 사상균류, 사카로미세스 종(*Saccharomyces* sp.)과 같은 출아 효모], 조류 세포(algal cell) 및 박테리아 세포로부터 선택될 수 있다. 바람직한 일 실시예에서, 재조합 숙주 세포는 재조합 미생물이다. 미생물인 숙주 세포의 예시들은 *에스체리치아*(*Escherichia*) 속, *바실루스*(*Bacillus*) 속, *락토바실루스*(*Lactobacillus*) 속, *자이모모나스*(*Zymomonas*) 속, *로도코쿠스*(*Rhodococcus*) 속, *슈도모나스*(*Pseudomonas*) 속, *아스페르길루스*(*Aspergillus*) 속, *트리코테르마*(*Trichoderma*) 속, *뉴로스포라*(*Neurospora*) 속, *푸사리움*(*Fusarium*) 속, *후미콜라*(*Humicola*) 속, *리조무코르*(*Rhizomucor*) 속, *클루이베로미세스*

(*Kluyveromyces*) 속, *피치아(Pichia)* 속, *무코르(Mucor)* 속, *미셀리오프토라(Myceliophthora)* 속, *페니실리움(Penicillium)* 속, *파네로카에테(Phanerochaete)* 속, *느타리(Pleurotus)* 속, *트라메테스(Trametes)* 속, *크리소스포리움(Chrysosporium)* 속, *사카로미세스(Saccharomyces)* 속, *스테노트로파모나스(Stenotrophomonas)* 속, *스키조사카로미세스(Schizosaccharomyces)* 속, *야로위아(Yarrowia)* 속, 또는 *스트렙토미세스(Streptomyces)* 속으로부터의 세포들을 포함한다(단, 이로 제한되지 않음). 몇몇 실시예들에서, 숙주 세포는 그람-양성(Gram-positive) 박테리아 세포이다. 다른 실시예들에서, 숙주 세포는 그람-음성(Gram-negative) 박테리아 세포이다. 몇몇 실시예들에서, 숙주 세포는 *대장균* 세포이다. 다른 실시예들에서, 숙주 세포는 *바실루스 렌투스(Bacillus lentus)* 세포, *바실루스 브레비스(Bacillus brevis)* 세포, *바실루스 스테아로써모필루스(Bacillus stearothermophilus)* 세포, *바실루스 리케니포르미스(Bacillus licheniformis)* 세포, *바실루스 알칼로필루스(Bacillus alkalophilus)* 세포, *바실루스 코아굴란스(Bacillus coagulans)* 세포, *바실루스 키르쿨란스(Bacillus circulans)* 세포, *바실루스 푸밀리스(Bacillus pumilis)* 세포, *바실루스 투링지엔시스(Bacillus thuringiensis)* 세포, *바실루스 클라우시(Bacillus clausii)* 세포, *바실루스 메가테리움(Bacillus megaterium)* 세포, *바실루스 서브틸리스(Bacillus subtilis)* 세포, 또는 *바실루스 아밀로리퀴파시엔스(Bacillus amyloliquefaciens)* 세포이다. 다른 실시예들에서, 숙주 세포는 *트리코테르마 코닝지(Trichoderma koningii)* 세포, *트리코테르마 비리데(Trichoderma viride)* 세포, *트리코테르마 르에세이(Trichoderma reesei)* 세포, *트리코테르마 롱기브라키아툼(Trichoderma longibrachiatum)* 세포, *아스페르길루스 아와모리(Aspergillus awamori)* 세포, *아스페르길루스 푸미가테스(Aspergillus fumigates)* 세포, *아스페르길루스 포에티두스(Aspergillus foetidus)* 세포, *아스페르길루스 니들란스(Aspergillus nidulans)* 세포, *아스페르길루스 니게르(Aspergillus niger)* 세포, *아스페르길루스 오리제(Aspergillus oryzae)* 세포, *후미콜라 인솔렌스(Humicola insolens)* 세포, *후미콜라 라누기노세(Humicola lanuginosa)* 세포, *로도코쿠스 오파쿠스(Rhodococcus opacus)* 세포, *리조무코르미에헤이(Rhizomucormiehei)* 세포, 또는 *무코르미에헤이(Mucormiehei)* 세포이다. 또 다른 실시예들에서, 숙주 세포는 *스트렙토미세스 리비단스(Streptomyces lividans)* 세포 또는 *스트렙토미세스 무리누스(Streptomyces murinus)* 세포이다. 또 다른 실시예들에서, 숙주 세포는 *방선균(Actinomycetes)* 세포이다. 몇몇 실시예들에서, 숙주 세포는 *맥주효모균(Saccharomyces cerevisiae)* 세포이다.

[0112] 다른 실시예들에서, 숙주 세포는 진핵 식물(eukaryotic plant) 세포, 조류 세포, 남세균(cyanobacterium) 세포, 녹색-황 세균 세포, 녹색 비-황 세균(green non-sulfur bacterium) 세포, 자색 황세균 세포, 자색 비-황 세균 세포, 극한 생물(extremophile) 세포, 효모 세포, 균류 세포, 본 명세서에 설명된 종 중 어느 것의 조작된 세포, 또는 합성 유기체이다. 몇몇 실시예들에서, 숙주 세포는 광 의존적이거나 탄소를 고정시킨다. 몇몇 실시예들에서, 숙주 세포는 독립영양적 활성을 갖는다. 몇몇 실시예들에서, 숙주 세포는 빛의 존재 하에서와 같이, 광독립영양적 활성(photoautotrophic activity)을 갖는다. 몇몇 실시예들에서, 숙주 세포는 빛의 부재 하에서 종속영양적 또는 혼합영양적이다. 특정 실시예들에서, 숙주 세포는 *애기장대, 파니쿰 비르가툼(Panicum virgatum)*, *미스칸투스 기간테우스(Miscanthus giganteus)*, *제아 메이스(Zea mays)*, *보트리오크쿠스 브라우니(Botryococcuse braunii)*, *녹조류(Chlamydomonas reinhardtii)*, *두나리엘라 살리나(Dunaliella salina)*, *시네코코쿠스 종 PCC 7002*, *시네코코쿠스 종 PCC 7942*, *시네코시스티스 종(Synechocystis Sp.) PCC 6803*, *서모시네코코쿠스 엘롱가투스(Thermosynechococcus elongates) BP-1*, *클로로비움 테피둠(Chlorobium tepidum)*, *클로로프렉수스 아우란티쿠스(Chlorojlexus auranticus)*, *크로마티움 비노숨(Chromatium vinosum)*, *로도스피릴룸 루브룸(Rhodospirillum rubrum)*, *로도박터 캡슐라투스(Rhodobacter capsulatus)*, *로도슈도모나스 파루스리스(Rhodopseudomonas palusris)*, *클로스트리디움 융달리(Clostridium ljungdahlii)*, *클로스트리디움 서모셀룸(Clostridium thermocellum)*, *페니실리움 크리소게눔(Penicillium chrysogenum)*, *피치아파스토리스(Pichiapastoris)*, *맥주효모균(Saccharomyces cerevisiae)*, *스키조사카로미세스 폼베(Schizosaccharomyces pombe)*, *슈도모나스 플루오레센스(Pseudomonas fluorescens)*, 또는 *자이모모나스 모빌리스(Zymomonas mobilis)*로부터의 세포이다.

[0113] **숙주 세포의 조작**

[0114] 몇몇 실시예들에서, 폴리뉴클레오티드(또는 유전자) 서열이 폴리뉴클레오티드 서열에 작동가능하게 연결된 프로모터를 포함하는 재조합 벡터에 의해 숙주 세포에 제공된다. 특정 실시예들에서, 프로모터는 발달-조절된, 세포소기관-특이적, 조직-특이적, 유도성, 구성적, 또는 세포-특이적 프로모터이다. 몇몇 실시예들에서, 재조합 벡터는 폴리뉴클레오티드 서열에 작동가능하게 연결된 발현 조절 서열; 폴리뉴클레오티드 서열에 작동가능하게 연결된 선택 마커; 폴리뉴클레오티드 서열에 작동가능하게 연결된 마커 서열; 폴리뉴클레오티드 서열에 작동가능하게 연결된 정제 모이어티; 폴리뉴클레오티드 서열에 작동가능하게 연결된 분비 서열; 및 폴리뉴클레오티드 서열에 작동가능하게 연결된 표적화 서열로부터 선택되는 적어도 하나의 서열을 포함한다. 본 명세서에 설명된

발현 벡터들은 숙주 세포 내에서의 폴리뉴클레오티드 서열의 발현에 적절한 형태인 폴리뉴클레오티드 서열을 포함한다. 당업자라면, 발현 벡터의 설계가 형질전환되는 숙주 세포의 선택, 원하는 폴리펩티드의 발현 수준 등과 같은 요인들에 의존할 수 있다는 것을 이해할 것이다. 본 명세서에 설명된 발현 벡터들은 앞서 설명된 바와 같은 폴리뉴클레오티드 서열들에 의해 인코딩되는, 융합 폴리펩티드들을 포함하는 폴리펩티드들을 생성하도록 숙주 세포들에 도입될 수 있다(위 참조). 원핵생물, 예를 들어 *대장균*에서 폴리펩티드들을 인코딩하는 유전자의 발현은 대부분 융합 또는 비-융합 폴리펩티드들의 발현을 지시하는 구성적 또는 유도성 프로모터를 함유하는 벡터들로 수행된다. 융합 벡터들은 그 안에 인코딩된 폴리펩티드에, 일반적으로는 재조합 폴리펩티드의 아미노 - 또는 카복시-말단에 다수의 아미노산들을 첨가한다. 이러한 융합 벡터들은 통상적으로 다음의 세 가지 목적, 즉 재조합 폴리펩티드의 발현을 증가시키고; 재조합 폴리펩티드의 용해도를 증가시키며; 친화성 정제에서 리간드로서 작용함으로써 재조합 폴리펩티드의 정제를 돕는 것 중 1 이상을 수행한다. 흔히, 융합 발현 벡터에서, 융합 모이어티와 재조합 폴리펩티드의 접합점에 단백질분해 절단 부위가 도입된다. 이는 융합 폴리펩티드의 정제 이후에 융합 모이어티로부터 재조합 폴리펩티드의 분리를 가능하게 한다. 이러한 효소들 및 이들의 동족 인식 서열(cognate recognition sequence)들의 예시들은 Xa 인자(Factor Xa), 트롬빈 및 엔테로키나아제를 포함한다. 예시적인 융합 발현 벡터는 pGEX 벡터[Pharmacia Biotech, Inc., Piscataway, NJ; Smith 외(1988) *Gene* 67:31-40], pMAL 벡터(New England Biolabs, Beverly, MA), 및 pRITS 벡터(Pharmacia Biotech, Inc., Piscataway, N.J.)를 포함하고, 이들은 표적 재조합 폴리펩티드에 각각 글루타티온 S-트랜스페라아제(GST), 말토스 E 결합 단백질, 또는 단백질 A를 융합시킨다.

[0115] 유도성, 비-융합 *대장균* 발현 벡터의 예시들은 pTrc 벡터[Amann 외(1988) *Gene* 69:301-315] 및 pET 11d 벡터[Studier 외, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, Calif.(1990) 60-89]를 포함한다. pTrc 벡터로부터의 표적 유전자 발현은 하이브리드 trp-lac 융합 프로모터로부터의 숙주 RNA 폴리메라아제 전사에 의존한다. pET 11d 벡터로부터의 표적 유전자 발현은 공동발현되는 바이러스 RNA 폴리메라아제(T7 gn1)에 의해 매개되는 T7 gn10-lac 융합 프로모터로부터의 전사에 의존한다. 이 바이러스 폴리메라아제는 lacUV 5 프로모터의 전사 제어 하에서 T7 gn1 유전자를 보유하는 내재성 λ 프로파지(resident λ prophage)로부터 BL21(DE3) 또는 HMS174(DE3)와 같은 숙주 균주들에 의해 공급된다. 원핵 세포 및 진핵 세포 모두에 적절한 발현계들이 해당 기술분야에서 잘 알려져 있다[예를 들어, Sambrook 외(1989) *Molecular Cloning: A Laboratory Manual*, second edition, Cold Spring Harbor Laboratory 참조]. 유도성, 비-융합 *대장균* 발현 벡터의 예시들은 pTrc 벡터[Amann 외(1988) *Gene* 69:301-315] 및 PET 11d 벡터[Studier 외(1990) *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA, pp. 60-89]를 포함한다. 특정 실시예들에서, 본 발명의 폴리뉴클레오티드 서열은 박테리오파지 T5로부터 유도된 프로모터에 작동가능하게 연결된다. 일 실시예에서, 숙주 세포는 효모 세포이다. 이 실시예에서, 발현 벡터는 효모 발현 벡터이다. 벡터들은 외래성(foreign) 핵산(예를 들어, DNA)을 숙주 세포 내로 도입하기 위하여 본 기술분야에서 인정되는 다양한 기술들을 통해 원핵 세포 또는 진핵 세포 내로 도입될 수 있다. 숙주 세포를 형질전환 또는 형질감염(transfecting)시키는 적절한 방법들은, 예를 들어 Sambrook 외에서 찾아볼 수 있다(위 참조). 박테리아 세포들의 안정적인 형질전환을 위하여, (사용되는 발현 벡터 및 형질전환 기술에 따라) 세포들의 특정 부분이 발현 벡터를 흡수하고 복제하는 것으로 알려져 있다. 이 형질전환체들을 식별하고 선택하기 위하여, 선택가능한 마커(예를 들어, 항생제에 대한 내성)를 인코딩하는 유전자가 관심 유전자와 함께 숙주 세포 내로 도입될 수 있다. 선택가능한 마커들은 암피실린, 카나마이신, 클로람페니콜, 또는 테트라사이클린과 같은(단, 이로 제한되지 않음) 약물들에 대한 내성을 제공하는 것들을 포함한다. 선택가능한 마커를 인코딩하는 핵산들은 본 명세서에서 설명되는 폴리펩티드를 인코딩하는 벡터와 동일한 벡터에서 숙주 세포 내로 도입될 수 있거나, 또는 별개의 벡터에서 도입될 수 있다. 도입된 핵산으로 안정적으로 형질전환된 세포들은 적절한 선택 약물의 존재 하에서 성장에 의해 식별될 수 있다. 본 명세서에서(위 참조) 설명된 바와 같은 조작된 또는 재조합 숙주 세포는 지방산 알데히드 또는 지방산 알코올과 같은 지방산 유도체 조성물을 생성하는데 사용되는 세포이다. 본 명세서에서 설명된 본 발명의 실시형태들 중 어느 것에서, 숙주 세포는 진핵 식물, 박테리아, 조류, 남세균, 녹색-황 세균, 녹색 비-황 세균, 자색 황 세균, 자색 비-황 세균, 극한 생물, 효모, 균류, 이들의 조작된 유기체, 또는 합성 유기체로부터 선택될 수 있다. 몇몇 실시예들에서, 숙주 세포는 광 의존적이거나 탄소를 고정시킨다. 몇몇 실시예들에서, 숙주 세포는 독립영양적 활성을 갖는다. 다양한 숙주 세포들이 본 명세서에 설명된 바와 같은 지방산 유도체들을 생성하는데 사용될 수 있다.

[0116] 본 발명의 숙주 세포 또는 미생물은, 효소 활성들에 대한 특정 돌연변이의 효율성을 테스트하기 위해 변경들을 포함하도록 유전적으로 조작되거나 변형된 숙주 균주 또는 숙주 세포(즉, 재조합 세포 또는 미생물)를 포함한다. 다양한 선택적인 유전적 조정 및 변경이, 어떠한 원시 효소 경로들이 원래 숙주 세포에 존재하는지

에 따라 한 숙주 세포로부터 다른 세포로 교환가능하게 사용될 수 있다. 일 실시예에서, 숙주 균주는 다른 생합성 폴리펩티드(예를 들어, 효소)와 조합하여 AAP 폴리펩티드의 발현을 테스트하는데 사용될 수 있다. 숙주 균주는, 발효 성분들, 탄소원(예를 들어, 공급원료), 온도, 압력, 감소된 배양 오염 상태들, 및 산소 수준들을 포함한 배양 조건들을 포함하는(단, 이로 제한되지 않음) 특정 변수들을 테스트하기 위해 다수의 유전적 변형들을 포함한다.

[0117] 일 실시예에서, 숙주 균주는 선택적인 *fadE* 및 *fhuA* 결실을 포함한다. 아실-CoA 디하이드로제나아제(FadE)는 지방산을 대사시키는데 중요한 효소이다. 이는 지방산 이용(베타-산화)에서 두번째 단계를 촉진시키며, 이는 지방산(아실-CoA)의 긴 사슬들을 아세틸-CoA 분자로 쪼개는 프로세스이다. 더 구체적으로, 박테리아의 지방산 분해의 β -산화 사이클의 두번째 단계는 아실-CoA의 2-엔오일-CoA로의 산화이며, 이는 FadE에 의해 촉진된다. 대장균에 FadE가 없는 경우, 이는 탄소원으로서 지방산에서 성장할 수 없지만, 아세테이트에서 성장할 수 있다. 여하한 사슬 길이의 지방산 이용의 불능(inability)은 *fadE* 균주들, 즉 FadE 기능이 파괴되는 *fadE* 돌연변이 균주들의 보고된 표현형과 일치한다. *fadE* 유전자는 선택적으로 녹아웃(knock out)되거나 감소되어, 지방산 유도체 경로에서 중간체일 수 있는 아실-CoA가 모두 지방산 유도체로 효율적으로 전환될 수 있도록 세포에 누적될 것을 보장할 수 있다. 하지만, *fadE* 감소는 탄소원으로서 당(sugar)이 사용될 때 선택적인데, 이는 이러한 조건 하에서 FadE의 발현이 억제될 가능성이 있고, 이에 따라 FadE가 소량만 존재할 수 있으며 아실-CoA 기질에 대해 에스테르 신타아제 또는 다른 효소들과 효율적으로 경쟁할 수 없기 때문이다. FadE는 분해대사물 억제(catabolite repression)로 인해 억제된다. 대장균 및 많은 다른 미생물들은 지방산보다 당의 소모를 선호하므로, 두 공급원들이 이용가능한 경우 *fad* 레귤론을 억제함으로써 당이 우선 소모된다[D. Clark, *J Bacteriol.* (1981) 148(2):521-6 참조]. 또한, 당의 부재 및 지방산의 존재는 FadE 발현을 포함한다. (FadE를 포함한) *fad* 레귤론에 의해 발현되는 단백질들이 상향-조절되고 아실-CoA에 대해 효율적으로 경쟁할 것이므로, 아실-CoA 중간체들이 베타 산화 경로에 대해 손실될 수 있다. 따라서, 녹아웃 또는 감소되는 *fadE* 유전자를 갖는 것이 유리할 수 있다. 대부분의 탄소원들이 주로 당 기반이기 때문에, FadE를 감소시키는 것은 선택적이다. 유전자 *fhuA*는 TonA 단백질을 코딩하며, 이는 대장균의 외막에서의 에너지-관련 운반체(energy-coupled transporter) 및 수용체이다[V. Braun(2009) *J Bacteriol.* 191(11):3431-3436 참조]. 이것의 결실은 선택적이다. *fhuA* 결실은 세포로 하여금 특정 발효 조건들에 유리할 수 있는 파지 공격(phage attack)에 대해 더 내성이 되게 한다. 따라서, 발효 진행 시 잠재적 오염의 가능성이 있는 숙주 세포에서 *fhuA*를 결실시키는 것이 바람직할 수 있다.

[0118] 또 다른 실시예에서, 숙주 세포(위 참조)는 *fadR*, *fabA*, *fabD*, *fabG*, *fabH*, *fabV*, 및/또는 *fabF*를 포함하는 유전자들 중 1 이상의 선택적인 과발현을 포함한다. 이러한 유전자들의 예시들은 대장균으로부터의 *fadR*, 켄티푸스균(*Salmonella typhimurium*)으로부터의 *fabA*(NP_460041), 켄티푸스균으로부터의 *fabD*(NP_460164), 켄티푸스균으로부터의 *fabG*(NP_460165), 켄티푸스균으로부터의 *fabH*(NP_460163), 콜레라균(*Vibrio cholera*)으로부터의 *fabV*(YP_001217283), 및 클로스트리디움 아세트부틸리쿰(*Clostridium acetobutylicum*)으로부터의 *fabF*(NP_350156)이다. 지방산 생합성에서 효소들 및 조절제를 코딩하는 이러한 유전자들 중 1 이상의 과발현은 다양한 배양 조건들 하에서 지방족 알데히드 및 지방족 알코올을 포함하는 지방산 유도체 화합물들의 역가를 증가시키는 역할을 할 수 있다.

[0119] 또 다른 실시예에서, 지방족 알데히드 및/또는 지방산 알코올과 같은 지방산 유도체들의 생성을 위해 숙주 세포로서 대장균 균주가 사용된다. 이와 유사하게, 이러한 숙주 세포들은 *fadR*, *fabA*, *fabD*, *fabG*, *fabH*, *fabV*, 및/또는 *fabF*를 포함하는(단, 이로 제한되지 않음) 다양한 배양 조건들 하에서 지방산 유도체(예를 들어, 지방족 알코올, 지방족 알데히드 등)와 같은 지방산 유도체 화합물의 역가를 더 증가시키거나 증대시킬 수 있는 1 이상의 생합성 유전자들(즉, 지방산 생합성의 효소들 및 조절제들을 코딩하는 유전자들)의 선택적인 과발현을 제공한다. 유전적 변형의 예시들은 대장균으로부터의 *fadR*, 켄티푸스균으로부터의 *fabA*(NP_460041), 켄티푸스균으로부터의 *fabD*(NP_460164), 켄티푸스균으로부터의 *fabG*(NP_460165), 켄티푸스균으로부터의 *fabH*(NP_460163), 콜레라균으로부터의 *fabV*(YP_001217283), 및 클로스트리디움 아세트부틸리쿰으로부터의 *fabF*(NP_350156)를 포함한다. 몇몇 실시예들에서, 이러한 생합성 유전자들을 운반하는 합성 오페론들이, 다양한 배양 조건들 하에서 지방족 알데히드 및/또는 지방족 알코올 과발현을 테스트하고, 및/또는 지방족 알데히드 및/또는 지방족 알코올 생성을 더 증대시키기 위해 세포 내에서 조작되고 발현될 수 있다. 이러한 합성 오페론들은 1 이상의 생합성 유전자를 함유한다. ifab138 오페론은, 예를 들어 특정한 배양 조건들을 테스트하기 위해 지방산 유도체들의 과발현을 용이하게 하는데 사용될 수 있는 콜레라균으로부터의 *fabV*, 켄티푸스균으로부터의 *fabH*, 켄티푸스균으로부터의 *fabD*, 켄티푸스균으로부터의 *fabG*, 켄티푸스균으로부터의 *fabA*, 및/또는 클로스트리디움 아세트부틸리쿰으로부터의 *fabF*를 포함한 선택적인 지방산 생합성 유전자들을 함유하는 조작된 오페론

이다. 이러한 합성 오페론들의 한가지 장점은, 지방산 유도체 생성률이 더 증가되거나 증대될 수 있다는 것이다.

[0120] 몇몇 실시예들에서, ACP 및 생합성 효소들(예를 들어, TE, ES, CAR, AAR, ADC 등)을 발현시키는데 사용되는 숙주 세포 또는 미생물은, 지방족 알코올, 지방족 알데히드, 지방족 에스테르, 지방족 아민, 이작용기 지방산 유도체, 디에시드(diacid) 등과 같은 1 이상의 특정 지방산 유도체(들)에 대한 생성을 증가시킬 수 있는 특정 효소 활성들을 포괄하는 유전자들을 더 발현시킬 것이다. 일 실시예에서, 숙주 세포는 유전자를 과발현시킴으로써 증가될 수 있는 지방산의 생성을 위해 티오에스테라아제 활성(E.C. 3.1.2.* 또는 E.C. 3.1.2.14 또는 E.C. 3.1.1.5)을 갖는다. 또 다른 실시예에서, 숙주 세포는 지방족 에스테르의 생성을 위해 에스테르 신타아제 활성(E.C. 2.3.1.75)을 갖는다. 또 다른 실시예에서, 숙주 세포는 지방족 알코올의 생성을 위해 아실-ACP 레덕타아제(AAR)(E.C. 1.2.1.80) 활성 및/또는 알코올 디하이드로제나아제 활성(E.C. 1.1.1.1) 및/또는 지방족 알코올 아실-CoA 레덕타아제(FAR)(E.C. 1.1.1.*) 활성 및/또는 카르복실산 레덕타아제(CAR)(E.C. 1.2.99.6) 활성을 갖는다. 또 다른 실시예에서, 숙주 세포는 지방족 알데히드의 생성을 위해 아실-ACP 레덕타아제(AAR)(E.C. 1.2.1.80) 활성을 갖는다. 또 다른 실시예에서, 숙주 세포는 알칸 및 알켄의 생성을 위해 아실-ACP 레덕타아제(AAR)(E.C. 1.2.1.80) 활성 및 탈카르복실라아제(ADC) 활성을 갖는다. 또 다른 실시예에서, 숙주 세포는 지방족 알코올의 생성을 위해 아실-CoA 레덕타아제(E.C. 1.2.1.50) 활성, 아실-CoA 신타아제(FadD)(E.C. 2.3.1.86) 활성, 및 티오에스테라아제(E.C. 3.1.2.* 또는 E.C. 3.1.2.14 또는 E.C. 3.1.1.5) 활성을 갖는다. 또 다른 실시예에서, 숙주 세포는 지방족 에스테르의 생성을 위해 에스테르 신타아제 활성(E.C. 2.3.1.75), 아실-CoA 신타아제(FadD)(E.C. 2.3.1.86) 활성, 및 티오에스테라아제(E.C. 3.1.2.* 또는 E.C. 3.1.2.14 또는 E.C. 3.1.1.5) 활성을 갖는다. 또 다른 실시예에서, 숙주 세포는 케톤의 생성을 위해 OleA 활성을 갖는다. 또 다른 실시예에서, 숙주 세포는 내부 올레핀의 생성을 위해 OleBCD 활성을 갖는다. 또 다른 실시예에서, 숙주 세포는 지방족 알코올의 생성을 위해 아실-ACP 레덕타아제(AAR)(E.C. 1.2.1.80) 활성 및 알코올 디하이드로제나아제 활성(E.C. 1.1.1.1)을 갖는다. 또 다른 실시예에서, 숙주 세포는 말단 올레핀을 만들기 위해 티오에스테라아제(E.C. 3.1.2.* 또는 E.C. 3.1.2.14 또는 E.C. 3.1.1.5) 활성 및 탈카르복실라아제 활성을 갖는다. 미생물 및 미생물 세포에서의 효소 활성들의 발현은 미국 특허 8,097,439; 8,110,093; 8,110,670; 8,183,028; 8,268,559; 8,283,143; 8,232,924; 8,372,610; 및 8,530,221에 개시되어 있으며, 이들은 본 명세서에서 인용 참조된다. 다른 실시예들에서, ACP 및 다른 생합성 효소들을 발현시키는데 사용되는 숙주 세포 또는 미생물은 지방족 알데히드 및/또는 지방족 알코올과 같은 1 이상의 특정 지방산 유도체(들)를 생성하기 위해 상향조절되거나 과발현되는 특정 원시 효소 활성들을 포함할 것이다. 일 실시예에서, 숙주 세포는 티오에스테라아제 유전자를 과발현시킴으로써 증가될 수 있는 지방산의 생성을 위해 원시 티오에스테라아제(E.C. 3.1.2.* 또는 E.C. 3.1.2.14 또는 E.C. 3.1.1.5) 활성을 갖는다.

[0121] 본 발명은 AAR 및 다른 생합성 효소들을 코딩하는 유전자들을 발현시키는 숙주 균주 또는 미생물을 포함한다(위 참조). 제조된 숙주 세포들은 지방족 알데히드 및 지방족 알코올과 같은 지방산 유도체들 및 이들의 조성물들 및 블렌드(blend)들을 생성한다. 지방산 유도체들은 통상적으로 배양 배지로부터 회수되고, 및/또는 숙주 세포로부터 격리된다. 일 실시예에서, 지방족 알데히드 및 지방족 알코올과 같은 지방산 유도체들은 배양 배지(세포외)로부터 회수된다. 또 다른 실시예에서, 지방족 알데히드 및 지방족 알코올과 같은 지방산 유도체들은 숙주 세포(세포내)로부터 격리된다. 또 다른 실시예에서, 지방족 알데히드 및 지방족 알코올과 같은 지방산 유도체들은 배양 배지로부터 회수되고, 숙주 세포로부터 격리된다. 숙주 세포에 의해 생성되는 지방산 유도체 조성물은, 지방족 알데히드 및 지방족 알코올 조성물들과 같은 지방산 유도체 조성물의 성분들의 포화도 및 사슬 길이뿐만 아니라 특정 지방산 유도체들의 분포를 결정하기 위해 당업계에서 알려진 방법들, 예를 들어 GC-FID를 이용하여 분석될 수 있다.

[0122] 미생물(예를 들어, 미생물 세포)로서 기능하는 숙주 세포의 예시들은 *에스체리치아* 속, *바실루스* 속, *락토바실루스* 속, *자이모모나스* 속, *로도코쿠스* 속, *슈도모나스* 속, *아스페르길루스* 속, *트리코테르마* 속, *뉴로스포라* 속, *푸사리움* 속, *후미콜라* 속, *리조무코르* 속, *클루이베로미세스* 속, *피치아* 속, *무코르* 속, *미셀리오프토라* 속, *페니실리움* 속, *파넬로카에테* 속, *느타리* 속, *트라메테스* 속, *크리소스포리움* 속, *사카로미세스* 속, *스테노트로파모나스* 속, *스키조사카로미세스* 속, *야로위아* 속, 또는 *스트렙토미세스* 속으로부터의 세포들을 포함한다(단, 이로 제한되지 않음). 몇몇 실시예들에서, 숙주 세포는 그람-양성 박테리아 세포이다. 다른 실시예들에서, 숙주 세포는 그람-음성 박테리아 세포이다. 몇몇 실시예들에서, 숙주 세포는 *대장균* 세포이다. 몇몇 실시예들에서, 숙주 세포는 *대장균* B 세포, *대장균* C 세포, *대장균* K 세포, 또는 *대장균* W 세포이다. 다른 실시예들에서, 숙주 세포는 *바실루스* *렌투스* 세포, *바실루스* *브레비스* 세포, *바실루스* *스테아로서모필루스* 세포, *바실루스* *리케니포르미스* 세포, *바실루스* *알칼로필루스* 세포, *바실루스* *코아굴란스* 세포, *바실루스* *키르쿨란스*

세포, 바실루스 푸밀리스 세포, 바실루스 투린지엔시스 세포, 바실루스 클라우시 세포, 바실루스 메가테리움 세포, 바실루스 서브틸리스 세포, 또는 바실루스 아밀로리퀴파시엔스 세포이다. 다른 실시예들에서, 숙주 세포는 트리코테르마 코닌지 세포, 트리코테르마 비리데 세포, 트리코테르마 르에세이 세포, 트리코테르마 롱기브라키 아툼 세포, 아스페르길루스 아와모리 세포, 아스페르길루스 푸미가테스 세포, 아스페르길루스 포에티두스 세포, 아스페르길루스 니들란스 세포, 아스페르길루스 니게르 세포, 아스페르길루스 오리제 세포, 후미콜라 인솔랜스 세포, 후미콜라 라누기노세 세포, 로도코쿠스 오파쿠스 세포, 리조무코르 미에헤이 세포, 또는 무코르 미에헤이 세포이다. 다른 실시예들에서, 숙주 세포는 스트렙토미세스 리비단스 세포 또는 스트렙토미세스 무리누스 세포이다. 다른 실시예들에서, 숙주 세포는 방선균 세포이다. 몇몇 실시예들에서, 숙주 세포는 맥주효모균 세포이다. 다른 실시예들에서, 숙주 세포는 진핵 식물, 조류, 남세균, 녹색-황 세균, 녹색 비-황 세균, 자색 황 세균, 자색 비-황 세균, 극한 생물, 효모, 균류, 이들의 조작된 유기체, 또는 합성 유기체로부터의 세포이다. 몇몇 실시예들에서, 숙주 세포는 광 의존적이거나 탄소를 고정시킨다. 몇몇 실시예들에서, 숙주 세포는 독립영양적 활성을 갖는다. 몇몇 실시예들에서, 숙주 세포는 빛의 존재 하에서와 같이, 광독립영양적 활성을 갖는다. 몇몇 실시예들에서, 숙주 세포는 빛의 부재 하에서 중속영양적 또는 혼합영양적이다. 특정 실시예들에서, 숙주 세포는 애기장대, 파니쿰 비르가툼, 미스칸투스 기간테우스, 제아 메이스, 보트리오크쿠스 브라우니, 녹조류, 두나리엘라 살리나, 시네코코쿠스 종 PCC 7002, 시네코코쿠스 종 PCC 7942, 시네코시스티스 종 PCC 6803, 서모 시네코코쿠스 엘롱가투스 BP-1, 클로로비움 테피둠, 클로로프렉투스 아우란티쿠스, 크로마티움 비노숨, 로도스 피렐룸 루브룸, 로도박터 캡슐라투스, 로도슈도모나스 파루스리시, 클로스트리디움 용달리, 클로스트리디움 서모셀룸, 페니실리움 크리스게눔, 피치아 파스토리스, 맥주효모균, 스킨조사카로미세스 폼베, 슈도모나스 플루오레센스, 또는 자이모모나스 모빌리스로부터의 세포이다. 일 특정 실시예에서, 미생물 세포는 프로클로로코쿠스 (*Prochlorococcus*), 시네코코쿠스, 시네코시스티스, 시아노테세(*Cyanotheca*), 및 노스톡 펑크티포르메(*Nostoc punctiforme*)를 포함하는 시아노박테리아로부터의 세포이다(단, 이로 제한되지 않음). 또 다른 실시예에서, 미생물 세포는 시네코코쿠스 엘롱가투스 PCC7942, 시네코시스티스 종 PCC6803, 및 시네코코쿠스 종 PCC7001을 포함하는 특정한 시아노박테리아 종으로부터의 세포이다(단, 이로 제한되지 않음).

[0123] 재조합 숙주 세포의 배양 및 발효

[0124] 본 명세서에서 사용되는 바와 같이, 발효라는 용어는 광범위하게 숙주 세포에 의한 유기 물질의 표적 물질로의 전환, 예를 들어 탄소원을 포함하는 배지에서 재조합 숙주 세포의 배양물을 증식시킴으로써 재조합 숙주 세포에 의한 탄소원의 지방산 또는 그 유도체로의 전환을 지칭한다. 생성을 위하여 허용되는 조건들은 숙주 세포로 하여금 원하는 생성물, 예컨대 지방족 알데히드 또는 지방족 알코올을 생성하게 하는 여하한의 조건들을 지칭한다. 이와 유사하게, 박터의 폴리뉴클레오티드 서열이 발현되는 조건 또는 조건들은 숙주 세포로 하여금 폴리펩티드를 합성하게 하는 여하한의 조건들을 의미한다. 적절한 조건들은, 예를 들어 발효 조건들을 포함한다. 발효 조건들은 온도 범위, 통기 수준, 공급량(feed rate), 및 배지 조성물을 포함하는 다수 파라미터를 포함할 수 있다(단, 이로 제한되지 않음). 이러한 조건들의 각각은 개별적으로 그리고 조합하여 숙주 세포가 성장하게 한다. 발효는 호기성, 혐기성, 또는 [미-호기성(micro-aerobic)과 같은] 이의 변이성(variations)일 수 있다. 예시적인 배양 배지는 배양액(broths) 또는 겔(gels)을 포함한다. 일반적으로, 배지는 숙주 세포에 의해 직접적으로 대사될 수 있는 탄소원을 포함한다. 또한, 효소들이 탄소원의 가동화(mobilization)[예를 들어, 전분 또는 셀룰로오스의 발효성 당들의 해중합(depolymerization)] 및 탄소원의 후속한 대사를 용이하게 하도록 배지에서 사용될 수 있다.

[0125] 소규모 생성을 위해, 조작된 숙주 세포들은 예를 들어 약 100 μ L, 200 μ L, 300 μ L, 400 μ L, 500 μ L, 1 mL, 5 mL, 10 mL, 15 mL, 25 mL, 50 mL, 75 mL, 100 mL, 500 mL, 1 L, 2 L, 5 L 또는 10 L의 배치(batches)에서 성장될 수 있고; 발효될 수 있으며; 원하는 폴리뉴클레오티드 서열, 예컨대 ACP 및/또는 생합성 폴리펩티드를 인코딩하는 폴리뉴클레오티드 서열을 발현시키도록 유도될 수 있다. 대규모 생성을 위해, 조작된 숙주 세포들은 약 10 L, 100 L, 1000 L, 10,000 L, 100,000 L 및 1,000,000 L 또는 그 이상의 배치에서 성장될 수 있고; 발효될 수 있으며; 원하는 폴리뉴클레오티드 서열을 발현시키도록 유도될 수 있다. 대안적으로, 대규모 유가배양식 발효(fed-batch fermentation)가 수행될 수 있다. 본 명세서에서 설명되는 지방산 유도체 조성물은 재조합 숙주 세포 배양의 세포의 환경에서 발견되고, 배양 배지로부터 쉽게 격리될 수 있다. 지방족 알데히드 또는 지방족 알코올과 같은 지방산 유도체는 재조합 숙주 세포에 의해 분비될 수 있고, 세포의 환경으로 수송될 수 있거나 또는 수동적으로 재조합 숙주 세포 배양의 세포의 환경으로 전달될 수 있다. 지방산 유도체는 해당 기술분야에 알려진 통상의 방법들을 이용하여 재조합 숙주 세포 배양물로부터 격리된다.

[0126] 재조합 숙주 세포로부터 유도된 생성물

[0127]

본 명세서에서 사용되는 바와 같이, 현대 탄소의 분율 또는 fM은 각각 옥살산 표준 H0xI 및 H0xII로 알려져 있는 미국 국립표준기술연구소(National Institute of Standards and Technology: NIST) 표준 참고 물질(Standard Reference Materials: SRMs) 4990B 및 4990C에 의해 정의된 바와 동일한 의미를 갖는다. 이러한 기본적인 정의는 (AD 1950을 기준으로) 0.95 배의 $^{14}\text{C}/^{12}\text{C}$ 동위원소 비율 H0xI에 관한 것이다. 이는 붕괴-보정된 산업혁명-전 목재(decay-corrected pre-Industrial Revolution wood)에 거의 등가이다. 현재 생존 생물권(living biosphere)(식물 재료)에 대하여, fM은 약 1.1이다. 바이오생성물(Bioproduction)(예를 들어, 본 발명에 따라 생성되는 지방족 알데히드 및/또는 지방족 알코올을 포함하는 지방산 유도체들)은 생물학적으로 생성되는 유기 화합물을 포함한다. 특히, 본 명세서의 지방산 생합성 경로를 이용하여 생성되는 지방산 유도체들은 재생 가능한 공급원들로부터 생성되지 않았으며, 이를테면 새로운 물질의 조성물이다. 이러한 새로운 바이오생성물은 이중 탄소-동위원소 핑거프린팅(dual carbon-isotopic fingerprinting) 또는 ^{14}C 연대측정(dating)에 기초하여 석유화학의 탄소로부터 유도되는 유기 화합물과 구별될 수 있다. 추가적으로, 생물자원 탄소(biosourced carbon)의 특이적 공급원(예를 들어, 글루코오스 대 글리세롤)은 이중 탄소-동위원소 핑거프린팅에 의해 결정될 수 있다(예를 들어, 미국 특허 7,169,588 참조). 석유 기반 유기 화합물과 바이오생성물을 구별하는 능력은 상업적으로(in commerce) 이러한 물질들을 추적하는데 유익하다. 예를 들어, 생물학적 기반 및 석유 기반 탄소 동위원소 프로파일 둘 모두를 포함하는 유기 화합물 또는 화합물질은 석유 기반 물질들로부터 만들어진 유기 화합물 및 화합물질과 구별될 수 있다. 따라서, 본 명세서의 바이오생성물은 이들의 특유한 탄소 동위원소 프로파일에 기초하여 상업적으로 후속되거나 추적될 수 있다. 바이오생성물은 각 샘플의 안정적인 탄소 동위원소 비율($^{13}\text{C}/^{12}\text{C}$)을 비교함으로써 석유 기반 유기 화합물과 구별될 수 있다. 주어진 바이오생성물의 $^{13}\text{C}/^{12}\text{C}$ 비율은, 이산화탄소가 고정된 시간에 대기 중의 이산화탄소에서 $^{13}\text{C}/^{12}\text{C}$ 비율의 결과이다. 또한, 이는 정확한 대사 경로를 반영한다. 또한, 국부적인 변이들도 일어난다. 석유, C3 식물(활엽), C4 식물(목초), 및 해양 탄산염(marine carbonate)이 모두 $^{13}\text{C}/^{12}\text{C}$ 및 대응하는 $\delta^{13}\text{C}$ 값들에서 상당한 차이를 나타낸다. 또한, C3 및 C4 식물들의 지질 물질은 대사 경로의 결과로서 동일한 식물들의 탄수화물 성분들로부터 유도된 물질들과 상이하게 분석한다. 측정의 정확도 내에서, ^{13}C 은 동위원소 분별 효과들로 인하여 큰 변화들을 나타내며, 바이오생성물들에 대한 이의 가장 중요한 점은 광합성 메카니즘이다. 식물들에서의 탄소 동위원소 비율 차이의 주요한 원인은 식물들에서의 광합성 탄소 대사의 경로, 특히 일차 카르복실화(primary carboxylation)(즉, 대기 중 CO_2 의 초기 고정) 동안 발생하는 반응의 차이들과 밀접하게 관련된다. 식생의 크게 두 가지 종류들은 C3[또는 캘빈-벤슨(Calvin-Benson)] 광합성 사이클을 병합하는 것들 및 C4[또는 헤치-슬랙(Hatch-Slack)] 광합성 사이클을 병합하는 것들이다. C3 식물들에서, 일차 CO_2 고정 또는 카르복실화 반응은 효소인 리불로오스-1,5-디포스페이트 카르복실라아제를 수반하며, 첫번째 안정적인 생성물은 3-탄소 화합물이다. 경목(hardwood)들 및 침엽수들과 같은 C3 식물들은 온대 기후 지역들에서 우세하다. C4 식물들에서, 또 다른 효소인 포스포엔올-피루베이트 카르복실라아제를 수반하는 부가적인 카르복실화 반응은 일차 카르복실화 반응이다. 첫번째 안정적인 탄소 화합물은 이후에 탈카르복실화되는 4-탄소산(carbon acid)이다. 이에 따라 방출된 CO_2 는 C3 사이클에 의해 재고정된다. C4 식물의 예시들은 열대형 목초들, 옥수수, 및 사탕수수이다. C4 및 C3 식물들이 모두 $^{13}\text{C}/^{12}\text{C}$ 동위원소 비율의 범위를 나타내지만, 통상적인 값들은 C4 식물에 대해 약 -7 내지 약 -13 퍼밀(per mil)이고, C3 식물에 대해 약 -19 내지 약 -27 퍼밀이다[예를 들어, Stuiver 외(1977) *Radiocarbon* 19:355 참조]. 석탄 및 석유는 일반적으로 이 후자의 범위에 속한다. ^{13}C 측정 척도는 본래 Pee Dee Belemnite(PDB) 석회암에 의한 제로 세트(zero set)에 의해 정의되며, 여기서 값들은 이 재료로부터 천분율 편차로 주어진다. $\delta^{13}\text{C}$ 값들은 천분율(퍼밀), 약어로는 ‰로 표현되고, 아래와 같이 계산된다:

[0128]

$$\delta^{13}\text{C} (\text{‰}) = [({}^{13}\text{C}/{}^{12}\text{C})_{\text{샘플}} - ({}^{13}\text{C}/{}^{12}\text{C})_{\text{표준물질}}] / ({}^{13}\text{C}/{}^{12}\text{C})_{\text{표준물질}} \times 1000$$

[0129]

PDB 표준 물질(RM)이 고갈되었으므로, 일련의 대안적인 RM들이 IAEA, USGS, NIST 및 다른 선택된 국제 동위원소 실험실들과 협력하여 개발되었다. PDB로부터의 퍼밀 편차에 대한 표기는 $\delta^{13}\text{C}$ 이다. 질량 44, 45 및 46의 분자 이온들에 대한 고정밀 안정 비율 질량 분석[high precision stable ratio mass spectrometry](IRMS)에 의해 CO_2 상에서 측정이 행해진다. 본 명세서에서 설명되는 조성물들은 본 명세서에서 설명되는 방법들 중 어느 하나에 의해 생성되는 바이오생성물을 포함하며, 예를 들어 지방산 유도체 생성물을 포함한다. 구체적으로, 바이오생성물은 약 -28 이상, 약 -27 이상, -20 이상, -18 이상, -15 이상, -13 이상, -10 이상, 또는 -8 이상의 δ

^{13}C 를 가질 수 있다. 예를 들어, 바이오생성물은 약 -30 내지 약 -15, 약 -27 내지 약 -19, 약 -25 내지 약 -21, 약 -15 내지 약 -5, 약 -13 내지 약 -7, 또는 약 -13 내지 약 -10의 $\delta^{13}\text{C}$ 를 가질 수 있다. 다른 경우들에서, 바이오생성물은 약 -10, -11, -12 또는 -12.3의 $\delta^{13}\text{C}$ 를 가질 수 있다. 또한, 본 명세서의 기재내용에 따라 생성되는 바이오생성물은 각 화합물의 ^{14}C 의 양을 비교함으로써 석유 기반 유기 화합물과 구별될 수 있다. ^{14}C 는 핵 반감기가 5730년이기 때문에, 더 오래된 탄소를 함유한 석유 기반 연료가 더 새로운 탄소를 함유한 바이오생성물과 구별될 수 있다[예를 들어, Currie, Source Apportionment of Atmospheric Particles, Characterization of Environmental Particles, J. Buffle and H. P. van Leeuwen, Eds., 1 of Vol. I of the IUPAC Environmental Analytical Chemistry Series(Lewis Publishers, Inc.) 3-74, (1992) 참조]. 방사성탄소 연대측정법(radiocarbon dating)의 기본적인 가정은 대기 중의 ^{14}C 농도의 항상성(constancy)이 생물(living organism)의 ^{14}C 의 항상성을 유도한다는 것이다. 하지만, 1950년 이후부터의 대기권 핵실험 및 1850년 이후부터의 화석 연료의 연소로 인하여, ^{14}C 는 제 2의, 지구화화적인 시간 특성을 얻었다. 대기 CO_2 및 이에 따른 생물권(living biosphere)의 그 농도는 1960년대 중반의 핵실험 피크의 거의 두 배였다. 이후, 7년 내지 10년의 근사적 이완 "반-감기"(approximate relaxation "half-life")를 갖는 약 1.2×10^{-12} 의 정상-상태 우주기원(steady-state cosmogenic) (대기) 기준 동위원소 비율($^{14}\text{C}/^{12}\text{C}$)로 점진적으로 복귀되었다. 이 후자의 반감기가 문자 그대로 취해져야 하는 것이 아니라; 그보다는 핵무기 시대의 시작 이후로 대기권 및 생물권의 ^{14}C 의 변이성을 추적하기 위해 상세한 대기 핵 투입/붕괴의 함수(detailed atmospheric nuclear input/decay function)를 사용하여야 한다. 이는 최근 생물권 탄소의 매년 연대측정의 가능성(promise of annual dating)을 지속하는 후자의 생물권 ^{14}C 시간 특성이다. ^{14}C 는 현대 탄소의 분율(fm)의 단위로 주어지는 결과들을 갖는 가속기 질량 분광 분석(accelerator mass spectrometry: AMS)에 의해 측정될 수 있다. fm은 미국 국립표준기술연구소(NIST) 표준 참고 물질(SRMs) 4990B 및 4990C에 의해 정의된다. 본 명세서에서 사용되는 바와 같이, 현대 탄소의 분율 또는 fm은 각각 옥살산 표준 HOxI 및 HOxII로 알려져 있는 미국 국립표준기술연구소(NIST) 표준 참고 물질들(SRMs) 4990B 및 4990C에 의해 정의되는 바와 동일한 의미를 갖는다. 기본적인 정의는 (AD 1950을 기준으로) 0.95 배의 $^{14}\text{C}/^{12}\text{C}$ 동위원소 비율 HOxI에 관한 것이다. 이는 붕괴-보정된 산업혁명-전 목재에 거의 등가이다. 현재 생물권(식물 재료)에 대하여, fm은 약 1.1이다. 본 명세서에서 설명되는 조성물들은 적어도 약 1의 fm ^{14}C 를 가질 수 있는 바이오생성물을 포함한다. 예를 들어, 본 발명의 바이오생성물은 적어도 약 1.01의 fm ^{14}C , 약 1 내지 약 1.5의 fm ^{14}C , 약 1.04 내지 약 1.18의 fm ^{14}C , 또는 약 1.111 내지 약 1.124의 fm ^{14}C 를 가질 수 있다.

[0130]

^{14}C 의 또 다른 측정은 pMC(percent of modern carbon)로 알려져 있다. ^{14}C 연대를 이용하는 고고학자 또는 지질학자에 대하여, AD 1950년은 0의 해(zero years old)와 같다. 또한, 이는 100 pMC를 나타낸다. 대기 중의 핵무기 탄소(Bomb carbon)는 열(thermo)-핵무기의 피크에서 1963년의 통상 수준의 거의 두 배에 달하였다. 대기권 내의 이의 분포는 이의 출현 이후로 근사화되었으며, AD 1950년 이후로 살아 있는 식물들 및 동물들에 대하여 100 pMC보다 더 큰 값을 나타낸다. 이는 시간이 지나면서 107.5 pMC 부근인 현재의 값으로 점차 감소하였다. 이는 옥수수과 같은 신선한(fresh) 바이오매스 재료가 107.5 pMC 부근의 ^{14}C 시그니처(signature)를 제공한다는 것을 의미한다. 석유 기반 화합물들은 0의 pMC 값을 가질 것이다. 오늘날의 탄소와 화석 탄소의 조합은 오늘날의 pMC 함량의 희석을 유도할 것이다. 107.5 pMC가 오늘날의 바이오매스 재료의 ^{14}C 함량을 나타내고 0 pMC가 석유 기반 생성물의 ^{14}C 함량을 나타낸다고 가정함으로써, 재료에 대해 측정된 pMC 값은 두 성분 유형의 비율을 반영할 것이다. 예를 들어, 오늘날의 콩으로부터 100 % 유도된 재료는 107.5 pMC 부근의 방사성탄소 시그니처를 제공할 것이다. 그 재료가 석유 기반 생성물로 50 % 희석되었다면, 이는 약 54 pMC의 방사성탄소 시그니처를 제공할 것이다. 생물학적 기반 탄소 함량은 107.5 pMC와 같은 100 %, 및 0 pMC와 같은 0 %를 할당함으로써 유도된다. 예를 들어, 99 pMC로 측정된 샘플은 93 %의 등가의 생물학적 기반 탄소 함량을 제공할 것이다. 이 값은 평균 생물학적 기반 탄소 결과로 지칭되며, 오늘날의 생물학적 재료 또는 석유 기반 재료 중 어느 하나로부터 비롯된 분석된 재료 내의 모든 성분을 가정한다. 본 명세서에서 설명되는 바와 같은 1 이상의 지방산 유도체를 포함하는 바이오생성물은 적어도 약 50, 60, 70, 75, 80, 85, 90, 95, 96, 97, 98, 99

또는 100의 pMC를 가질 수 있다. 다른 경우에, 본 명세서에서 설명되는 지방산 유도체는 약 50 내지 약 100; 약 60 내지 약 100; 약 70 내지 약 100; 약 80 내지 약 100; 약 85 내지 약 100; 약 87 내지 약 98; 또는 약 90 내지 약 95의 pMC를 가질 수 있다. 또 다른 경우에, 본 명세서에서 설명되는 지방산 유도체는 약 90, 91, 92, 93, 94, 또는 94.2의 pMC를 가질 수 있다.

[0131] 지방족 알데히드 및 지방족 알코올 조성물들 및 이들의 이용

[0132] 알데히드는 많은 특수 화학물질을 생성하는데 사용된다. 예를 들어, 알데히드는 중합체, 수지[예를 들어, 베이클라이트(BAKELITE) 수지], 염료, 착향료, 가소제, 향수, 약제 및 다른 화학물질들을 생성하는데 사용되고, 이들 중 일부는 용제, 보존제 또는 소독제로서 사용될 수 있다. 또한, 비타민 및 호르몬과 같은 특정 천연 및 합성 화합물들이 알데히드들이고, 많은 당이 알데히드기들을 함유한다. 지방족 알데히드들은 화학적 또는 효소적 환원에 의해 지방족 알코올들로 전환될 수 있다. 지방족 알코올 또한 많은 상업적 용도들을 갖는다. 지방족 알코올 및 이들의 유도체들의 전 세계적인 연간 판매액은 10억 달러를 초과한다. 짧은 사슬형 지방족 알코올들은 유화제, 연화제 및 증점제로서 화장품 산업 및 식품 산업에 사용된다. 이들의 양친매성 성질로 인해, 지방족 알코올은 비이온성 계면활성제처럼 행동하고, 이는 예를 들어 세제와 같은 개인 위생 용품 및 가정 용품에 유용하다. 또한, 지방족 알코올은 왁스, 검, 수지, 약제학적 연고 및 로션, 윤활유 첨가제, 직물 정전기 방지제 및 마무리제, 가소제, 화장품, 공업 용제 및 지방용 용제에 사용된다.

[0133] 또한, 본 발명은 본 명세서에서 설명되는 방법들 중 어느 것에 의해 생성되는 지방족 알코올을 포함하는 계면활성제 조성물 또는 세제 조성물을 제공한다. 본 기술분야의 당업자라면, 계면활성제 또는 세제 조성물의 의도된 목적에 따라, 상이한 지방족 알코올이 생성되고 사용될 수 있다는 것을 이해할 것이다. 예를 들어, 본 명세서에서 설명되는 지방족 알코올이 계면활성제 또는 세제 생성을 위한 공급원료로서 사용되는 경우, 본 기술분야의 당업자라면 지방족 알코올 공급원료의 특성들이 생성되는 계면활성제 또는 세제 조성물의 특성들에 영향을 미칠 것임을 이해할 것이다. 따라서, 계면활성제 또는 세제 조성물의 특성들은 공급원료로서 사용되는 특정 지방족 알코올을 생성함으로써 선택될 수 있다. 본 명세서에서 설명되는 지방족 알코올-기반 계면활성제 및/또는 세제 조성물은 본 기술분야에 잘 알려진 다른 계면활성제 및/또는 세제와 혼합될 수 있다. 몇몇 실시예들에서, 혼합물은 지방족 알코올의 중량의 적어도 약 10 %, 적어도 약 15 %, 적어도 약 20 %, 적어도 약 30 %, 적어도 약 40 %, 적어도 약 50 %, 적어도 약 60 %, 또는 상기 값들 중 어느 두 개의 값들에 의해 한정되는 범위를 포함할 수 있다. 다른 예시들에서, 계면활성제 또는 세제 조성물은 길이가 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 또는 22 개의 탄소들인 탄소 사슬을 포함하는 지방족 알코올의 중량의 적어도 약 5 %, 적어도 약 10 %, 적어도 약 20 %, 적어도 약 30 %, 적어도 약 40 %, 적어도 약 50 %, 적어도 약 60 %, 적어도 약 70 %, 적어도 약 80 %, 적어도 약 85 %, 적어도 약 90 %, 적어도 약 95 %, 또는 상기 값들 중 어느 두 개의 값들에 의해 한정되는 범위를 포함하도록 만들어질 수 있다. 또한, 이러한 계면활성제 또는 세제 조성물은 식물성유 또는 석유와 같은 비-미생물 공급원(nonmicrobial sources)으로부터의 계면활성제 또는 세제 또는 마이크로에멀션(microemulsion)과 같은 적어도 하나의 첨가제를 포함할 수 있고, 이는 지방족 알코올의 중량의 적어도 약 5 %, 적어도 약 10 %, 적어도 약 15 %, 적어도 약 20 %, 적어도 약 30 %, 적어도 약 40 %, 적어도 약 50 %, 적어도 약 60 %, 적어도 약 70 %, 적어도 약 80 %, 적어도 약 85 %, 적어도 약 90 %, 적어도 약 95 %, 또는 상기 값들 중 어느 두 개의 값들에 의해 한정되는 범위의 양으로 존재할 수 있다.

[0134] 예시들

[0135] 다음의 특정 예시들은 본 발명을 예시하기 위해 의도된 것이며, 청구항의 범위를 제한하는 것으로 해석되어서는 안 된다.

[0136] 프로토콜들 및 방법들

[0137] 라이브러리 스크리닝

[0138] 본 명세서에 설명된 모든 프로토콜들은 배양물을 성장시키기 위한 96 웰 플레이트(96 well plates)-마스터 블록-2mL 시스템(Greiner Bio-One, Monroe, NC 또는 Corning, 암스테르담, 네덜란드), 및 배양액으로부터 지방산종을 추출하기 위한 플레이트들(Costar, Inc.)에 의존한다. 아래에 제공된 프로토콜들은 발효 조건들의 예시이다. 지방산 종의 생성을 평가하기 위해 대안적인 프로토콜들이 사용될 수 있다.

[0139] 32 °C 배양 프로토콜(4NBT)

[0140] (96 웰 플레이트에서 성장된 LB 배양물로부터) 20 μ L LB 배양물이 400 μ L 2NBT 배지(표 2)를 주입(inoculate)하는데 사용되었으며, 이는 이후 32 °C 진탕으로 약 16 시간 동안 배양되었다. 하룻밤이 지난 종자

(overnight seed)의 20 μ L가 400 μ L NBT에 1 또는 2 g/L 질소(NBT_1N 또는 NBT_2N)를 주입하는데 사용되었다. 6 시간 동안 32 $^{\circ}$ C에서 성장한 후, 배양물은 IPTG(최종 농도 1 mM)로 유도되었다(표 2). 이후, 배양물은 18 시간 동안 32 $^{\circ}$ C에서 진탕으로 배양되었으며, 이후 배양물은 아래에 자세히 설명된 표준 추출 프로토콜에 따라 추출되었다.

표 2: 배지명 및 제제(Formulations)

배지명	제제		
2NBT	1	X	NH ₄ Cl을 갖는 5x 염 용액
	1	g/L	100 g/L NH ₄ Cl
	1	mg/L	10mg/mL 티아민
	1	mM	1M MgSO ₄
	0.1	mM	1M CaCl ₂
	20	g/L	500g/L 글루코오스
	1	X	1000x TM2
	10	mg/L	10g/L Fe 시트레이트
	100	μ g/mL	100 mg/mL 스펙티노마이신
	100	mM	2M 비스 트리스 (pH7.0)
4NBT_1N	1	X	NH ₄ Cl을 갖는 5x 염 용액
	1	mg/L	10mg/mL 티아민
	1	mM	1M MgSO ₄
	0.1	mM	1M CaCl ₂
	40	g/L	500g/L 글루코오스
	1	X	1000x TM2
	10	mg/L	10g/L Fe 시트레이트
	100	μ g/mL	100 mg/mL 스펙티노마이신
	100	mM	2M 비스 트리스 (pH7.0)
	100	mM	2M 비스 트리스 (pH7.0)
4NBT_2N	1	X	NH ₄ Cl을 갖는 5x 염 용액
	1	g/L	100 g/L NH ₄ Cl
	1	mg/L	10mg/mL 티아민
	1	mM	1M MgSO ₄
	0.1	mM	1M CaCl ₂
	40	g/L	500g/L 글루코오스
	1	X	1000x TM2
	10	mg/L	10g/L Fe 시트레이트
	100	μ g/mL	100 mg/mL 스펙티노마이신
	100	mM	2M 비스 트리스 (pH7.0)

지방산 중 표준 추출 프로토콜

각각의 웰에 40 μ L의 1M HCl이 추출되게 하기 위해, 이후 (내부 표준으로서 500 mg/L C11-FAME을 갖는) 300 μ L의 부틸 아세테이트가 첨가되었다. 96 웰 플레이트는 이후 플레이트 실러(plate sealer)(ALPS-300 가열기; Abgene, ThermoScientific, Rockford, IL)를 이용하여 열-용착되었고(heat-sealed), MIXMATE 혼합기(Eppendorf, Hamburg, Germany)를 이용하여 2000 rpm으로 15 분 동안 진탕되었다. 진탕 후, 상기 플레이트는 실온에서 4500 rpm으로 10 분 동안 원심분리되어(Allegra X-15R, rotor SX4750A, Beckman Coulter, Brea, CA), 수성 및 유기 층들을 분리하였다. 유기 층의 50 μ L가 96 웰 플레이트(폴리프로필렌, Corning, Amsterdam, The Netherlands)로 전달되었으며, 이후 이는 열-용착되었고, FALC_Broth.met 방법을 이용하여 GC-FID에 의해 평가 될 때까지 -20 $^{\circ}$ C로 저장되었다. FALC_Broth.met 방법은 다음과 같이 수행되었다: 1 μ L의 샘플이 불꽃 이온화 검출기(FID)를 갖는 Agilent 7890A GC Ultra 장치(Agilent, Santa Clara, CA)에서 분석 컬럼(DB-1, 10 m x 180 μ m x 0.2 μ M 막 두께, JW 121-101A로부터 이용가능함) 상으로 주입되었다. 이 기구는 C6 내지 C18 지방족 알코올을 검출 및 정량화하도록 셋업되었다. 앞서 자세히 설명된 프로토콜은 표준 조건들을 나타내며, 이는 분석 결과들을 최적화하기 위해 필요에 따라 변형될 수 있다.

- [0145] 지방산 중 - 표준 나일 레드 어세이 프로토콜(Standard Nile Red Assay Protocol)
- [0146] 24 시간 발효 후, 나일 레드 어세이는 Greiner MicroLonFluotrac 200 플레이트에서 1 $\mu\text{g/mL}$ 나일 레드의 최종 어세이 농도에 대해 84.6 %의 물 및 15.4 %의 아세트니트릴 용액의 1.54 $\mu\text{g/mL}$ 나일 레드의 130 μL 에 70 μL 의 발효액을 첨가함으로써 수행되었으며, 위아래로 피펫팅(pipetting)함으로써 혼합되었다. 상대 형광 단위(Relative fluorescence units)는 SpectraMax M2 유닛을 이용하여 630 nm의 방출(emission) 및 540 nm의 여기(excitation)에서 측정되었다.
- [0147] 오류 유발 라이브러리 구축
- [0148] 해당 기술분야의 당업자에게 알려진 표준 기술들은 오류 유발 라이브러리를 제조하는데 사용되었다. 일 예시에서, 벡터 골격(vector backbone)은 벡터의 제한 엔도뉴클레아제(restriction endonucleases)를 이용함으로써 제조된 한편, DNA 삽입체의 다양성(diversity)의 생성은 미스매치된 뉴클레오타이드의 통합에 우호적인 조건들 하에서 DNA 주형으로부터 PCR 증폭에 의해 생성되었다. 일 접근법에서, 다양성을 갖는 DNA 삽입체 및 벡터 골격의 클로닝은 제조업체의 프로토콜에 따라 INFUSION Cloning System(Clontech Laboratories, Inc., Mountain View, CA)을 이용하여 수행되었다.
- [0149] 포화 라이브러리 구축
- [0150] 해당 기술분야의 당업자에게 알려진 표준 기술들은 포화 라이브러리를 제조하는데 사용되었다. 일 예시에서, 벡터 골격은 벡터의 제한 엔도뉴클레아제를 이용함으로써 제조된 한편, DNA 삽입체의 다양성의 생성은 변성 프라이머(degenerate primers)를 이용하여 생성되었다. 일 접근법에서, 다양성을 갖는 DNA 삽입체 및 벡터 골격의 클로닝은 제조업체의 프로토콜에 따라 INFUSION Cloning System(Clontech Laboratories, Inc., Mountain View, CA)을 이용하여 수행되었다.
- [0151] 조합 라이브러리 구축
- [0152] 유익한 것으로 식별된 돌연변이들이 지방족 알코올 종의 생성 시 AAR 변이체에 추가 개선들을 제공하도록 조합되었다. 해당 기술분야의 당업자에게 알려진 표준 기술들은 조합 라이브러리를 제조하는데 사용되었다. 일 예시에서, 벡터 골격은 벡터의 제한 엔도뉴클레아제를 이용함으로써 제조된 한편, DNA 삽입체의 다양성의 생성은 원하는 돌연변이들을 도입하기 위한 프라이머를 이용하여 생성되었다. 앞서 설명된 바와 같이, 일 접근법에서, 다양성을 갖는 DNA 삽입체 및 벡터 골격의 클로닝은 제조업체의 프로토콜에 따라 INFUSION Cloning System(Clontech Laboratories, Inc., Mountain View, CA)을 이용하여 수행되었다. 조합 라이브러리는 tPCR(transfer PCR) 프로토콜을 이용하여 생성될 수 있다[Erijman 외(2011) *J. Structural Bio.*175:171-177].
- [0153] 라이브러리 스크리닝
- [0154] 라이브러리 다양성이 오류-유발, 포화 라이브러리 또는 조합 라이브러리에 생성되었으면, 이는 앞서 설명된 방법들 중 하나를 이용하여 스크리닝되었다. 두 가지 유형의 히트(hit)들이 식별되었다: (1) 증가된 양의 지방족 알코올(FALC 역가); 및/또는 (2) 도데칸올(C12) 또는 테트라데칸올(C14)과 같은 증가된 양의 중간-사슬 FALC. 또한, 헥사데칸올(C16) 및 옥타데칸올(C18)이 식별되었다. 각 히트 내의 AAR 변이체의 돌연변이들은 해당 기술분야의 당업자에 의해 통상적으로 이용되는 표준 기술들을 이용하여 서열화함으로써 식별되었다. 표 4, 표 5 및 표 6은 포화 라이브러리 및 조합 라이브러리에 유익한 것으로 식별된 돌연변이(히트)들을 목록화한 것이다.
- [0155] 예시 1: 지방산 합성 경로를 통한 아실기 운반 단백질(ACP) 매개되는 증가된 플럭스에 의한 AAR_7942를 이용하여 개선된 지방족 알코올 생성
- [0156] 지방족 아실-ACP를 생성물로 전환시키기 위해 *대장균* 이외의 공급원들로부터의 말단 경로 효소들이 이중 숙주로서 *대장균*에서 발현될 때, *대장균* 지방족 아실-ACP를 향한 재조합 경로 효소의 인식, 친화성, 및/또는 대사회전에 있어서 한계들이 존재할 수 있다. ACP 단백질들이 모든 유기체들에서 어느 정도 보존되더라도, 이들의 일차 서열(primary sequence)은 주어진 종 내에서조차 상이할 수 있다. 이러한 가설을 테스트하기 위해, 수 개의 시아노박테리아로부터의 *acp* 유전자들이 플라스미드 pLS9-185에 존재하는 *시네키크쿠스 엘롱가투스* PCC7942 아실-ACP 레덕타아제(AAR_7942)로부터 하류에 클로닝되었으며, 이는 pCL1920 유도체(3-5 복사체/세포)이다. 또한, 폭넓은 기질 특이성을 갖는 포스포판테이닐 트랜스페라아제를 인코딩하는 *바실루스 서브틸리스*로부터의 *sfp* 유전자(수탁 번호 X63158; SEQ ID NO: 11)는 각각의 *acp* 유전자들의 하류에 클로닝되었다. 이 효소는 비활성 아포-ACP의 활성 홀로-ACP로의 전환에 관련된다. 구성된 플라스미드들이 표 3에 설명되어 있다.
- [0157] 표 3: *S. 엘롱가투스* PCC7942 AAR로부터의 하류에 *B. 서브틸리스* *sfp*를 포함 및 불포함하는 시아노박테리아 ACP

를 공동발현시키는 플라스미드들

기본 플라스미드	ACP 공급원	ACP - SEQ ID NO. (NA/AA*)	sfp 불포함	sfp 포함
pLS9-185	시네코코쿠스 엘롱가투스 7942	7/8	pDS168	pDS168S
pLS9-185	시네코시스투스 종 6803	3/4	pDS169	이용불가능
pLS9-185	프로클로로코쿠스 마리누스 MED4	5/6	pDS170	pDS170S
pLS9-185	노스톡 펙크티포르미 73102	1/2	pDS171	pDS171S
pLS9-185	노스톡 종 7120	9/10	pDS172	pDS172S

*NA = 핵산 서열; AA = 아미노산 서열/폴리펩티드 서열

acp 유전자들은 모두 INFUSION 기술을 이용하여 pLS9-185에서 aar 유전자의 하류에 가까운 EcoRI 부위로 합성 RBS와 클로닝되었다. EcoRI 부위는 acp 유전자의 하류에 재구성되었다. 이와 유사하게, B. 서브틸리스 sfp 유전자는 합성 RBS와 함께 이 EcoRI 부위로 INFUSION 클로닝되었다. 모든 플라스미드들은 대장균 MG1655 DV2로 형질전환되었다. 이러한 실험들에 대한 대조군은 AAR 단독(pLS9-185) 발현이다.

표준 진탕 플라스크 발효 실험들로부터 얻어진 결과들이 도 5에 나타나 있다. 지방족 알코올 역가의 상당한 개선이 플라스미드들 pDS171S, pDS172S, pDS168 및 pDS169를 함유하는 균주들에서 관찰되었으며, 이는 ACP 과발현이 지방족 알코올 생성에 유익할 수 있음을 입증한다. 이론에 얽매이고자 하는 것은 아니지만, ACP 과발현이 이종의 말단 경로 효소에 의해 아실-ACP의 인식, 친화성 및/또는 대사회전에 도움을 줌으로써 지방족 알코올 생성에 유익할 수 있다는 가설을 세웠다[sfp의 유무 및 ACP의 공급원에 대한 표 3(위) 참조].

예시 2: 지방산 합성을 통한 아세틸-CoA 카르복실라아제(ACC) 매개되는 증가된 플럭스에 의한 AAR_7942를 이용하는 개선된 지방족 알코올 생성

지방산 생합성의 주요한 전구체들은 말로닐-CoA 및 아세틸-CoA이다. 이러한 전구체들은 대장균에서 지방산 생합성의 속도를 제한하는 것으로 여겨졌다. 이 예시에서, 합성 아세틸-CoA 카르복실라아제, acc, 오페론[코리네 박테리움 글루타미쿰 accDCAB+birA; SEQ ID NO: 45 또는 46, 48, 및 50, D+라고도 지칭함]이 시네코코쿠스 엘롱가투스 PCC7942(AAR_7942)로부터의 아실-ACP 레덕타아제(AAR)과 함께 발현되었다. accD+ 오페론은 플라스미드 pLS9-185에서 AAR_7942의 하류에 클로닝되었다. 결과적인 플라스미드 및 pLS9-185 대조군 플라스미드는 대장균 DV2로 형질전환되었다. 균주들은 표준 진탕 플라스크 프로토콜에서 지방족 알코올 생성에 대해 평가되었다. 도 6에 도시된 바와 같이, 합성 코리네박테리움 글루타미쿰 acc 오페론의 공동발현은 증가된 지방족 알코올 생성을 야기하였다.

예시 3: 주형으로서 AAR_7942를 이용하여 제조된 오류 유발 라이브러리, 조합 및 제한된 포화 라이브러리들

A. 오류 유발 라이브러리

시네코코쿠스 엘롱가투스 PCC7942(AAR_7942)로부터의 아실-ACP 레덕타아제의 오류 유발 라이브러리는 야생형 AAR_7942를 넘어서는 개선들을 나타낸 변이체들에 대해 구축 및 스크리닝되었다. 오류 유발 라이브러리를 만드는 데 사용된 플라스미드는 "pDS171S"(표 3 참조)로 지정되었다. 오류 유발 라이브러리는 앞서 설명된 표준 프로토콜들 중 하나를 이용하여 스크리닝되었다. 이러한 개선들은 역가를 개선하거나, 역가에 실질적인 영향을 주지 않고 생성된 C10 내지 C14의 분율을 증가시키는 것으로서 분류되었다(결과들이 나타나 있지 않음).

B. 조합 및 제한된 포화 라이브러리들

해당 기술분야의 당업자에게 알려진 표준 기술들은 위치들 17, 18 및 19에 기초하여 조합 라이브러리 및 포화 라이브러리를 제조하는데 사용되었다. 조합 라이브러리 및 포화 라이브러리에서 테스트된 돌연변이들(표 4A 및 4B)은 앞서 설명된 AAR_7942의 오류 유발 라이브러리에서 근원적으로 식별되었다. 사용된 플라스미드들, 균주들 및 스크리닝 프로토콜들은 예시 1에 설명된 것과 동일하였다. 오류 유발 라이브러리의 스크리닝으로부터 얻어진 결과들이 표 4A 및 4B에 나타나 있다. 표 4A는 증가된 지방족 알코올 역가를 야기한 AAR_7942 돌연변이들을 나타내고, 표 4B는 전체 지방족 알코올 역가에 실질적인 영향을 주지 않고 C14 지방족 알코올의 증가된 분율을 야기한 돌연변이들을 나타낸다.

표 4A: 개선된 지방족 알코올 역가와 상호관련된 AAR 7942 조합 및 제한된 포화 라이브러리들로부터의 돌연변이

를

돌연변이	콤보 α 라이브러리			
	FALC 역가 (총계*)	C14-FALC (% 분율)	정상화 FALC (% 대조군 증가)	정상화 C14 (% 대조군 증가)
S18T, C63Y, S96T, L177H, A281V	2473	48%	251%	294%
S18T, C63Y, S96T, L177H, A281V E284K	2202	50%	200%	229%
S18T, C63Y, S96T, L177H	2088	43%	189%	197%
S18T, C63Y	1798	39%	182%	241%
C63H, S967T, L177H, A281V	2045	53%	180%	237%
S18T, C63Y, A281V, E283K	1696	45%	172%	278%
S18T, C63Y	1640	43%	166%	260%
S18T, L177H, A281V, E283K	1618	44%	164%	270%
S18T, L177H, A281V, E283K	1617	43%	164%	264%
S18T, S96T, L177H, E283K	1814	43%	160%	194%
S18T, C63H, L177H	1811	39%	160%	173%
S18T, L177H, A281V, E283K	1503	39%	152%	237%
L65F, S96T, A281V, E283K	1638	43%	144%	191%
C63Y, A281V, A282T	1622	38%	143%	168%
L11F, C63Y, S96T, L177H, A281V A282T, E283K	1579	41%	143%	186%
L65F, S96T, L177H, A281V, E283K	1585	42%	140%	189%
G22S, C63H, S96T, L177H, E283K	1229	39%	125%	237%
C63H, L65F, S96T, A281V	1178	41%	119%	252%
S18T, L65F, S96T, L177H, A281V	1136	33%	115%	203%
C63R, L177H, A281V, E283K	1275	44%	112%	196%

포화 라이브러리				
돌연변이	FALC 역가 (총계*)	C14-FALC (% 분율)	정상화 FALC (% 대조군 증가)	정상화 C14 (% 대조군 증가)
S18M	2809	40%	243%	212%
S18F	2564	54%	222%	285%
S18Y	2550	45%	221%	236%
S18W	2530	55%	227%	314%
S18M	2495	41%	224%	231%
S18Y	2039	44%	188%	252%
S18Y	1997	44%	179%	251%

[0170]

포화 라이브러리				
S18T	1729	31%	155%	177%
S18T	1690	31%	152%	176%
S18T	1683	31%	151%	176%
S18T	1599	30%	143%	172%
S18L	1568	48%	141%	271%
S18C	1381	28%	120%	149%
S18C	1372	26%	123%	147%
S18C	1368	28%	118%	148%
S18L	1359	47%	122%	266%
S18C	1355	26%	121%	148%
S18V	1344	30%	121%	170%
S18C	1340	26%	120%	147%
S18L	1326	47%	119%	268%
S18L	1315	47%	118%	266%
S18C	1313	26%	118%	149%
S18V	1306	30%	121%	170%
S18C	1305	26%	117%	148%
S18L	1302	47%	117%	267%
V17L	1188	23%	110%	130%
V17L	1185	33%	103%	173%
V17L	1175	32%	102%	171%
콤보 β 라이브러리				
	FALC 역가 (총계*)	C14-FALC (% 분율)	정상화 FALC (% 대조군 능가)	정상화 C14 (% 대조군 능가)
S18W,S96T,E283K,R286Q,A324V	3437	49%	246%	255%
S18W	3286	53%	235%	275%
S18W,S96T,E283K	3199	46%	229%	237%
S18W,C63Y,S96T	3186	46%	228%	241%
S18W,C68Y,E2834K	2949	50%	211%	260%
C63Y,E283K	2680	36%	192%	189%
S18W,C63Y	2675	54%	191%	280%
C63Y,S96T	2583	38%	185%	196%
C63Y,E283K,Q316K	2420	36%	173%	188%
S96T	1605	23%	115%	121%
E283K	1532	25%	110%	131%

[0171]

[0172]

표 4B: 개선된 C14 지방족 알코올 분율과 상호관련된 AAR_7942 조합 및 제한된 포화 라이브러리들로부터의 돌연변이들

콤보 α 라이브러리				
	FALC 역가 (총계*)	C14-FALC (% 분율)	정상화 FALC (% 대조군 능가)	정상화 C14 (% 대조군 능가)
C63H, S96T, L177H, A281V	2045.36	53%	180%	237%
G22S, C63H, F64V, S96T, L177H, A281V	152.47	51%	14%	233%

[0173]

S18T, C63H, S96T, L177H, A281V E283K	2202.46	50%	200%	229%
S18T, C63Y, S96T, L177H, A281V	2472.80	48%	251%	294%
S18T, C63Y, A281V, E283K	1695.59	45%	172%	278%
S18T, L177H, A281V, E283K	1618.20	44%	164%	270%
C63R, L177H, A281V, E283K	1275.38	44%	112%	196%
S18T, C63H, L65F, S96T, L177H A281V, E283K	912.50	44%	83%	199%
S18T, S96T, L177H, E283K	1814.34	43%	160%	194%
S18T, L179H, A281V, E283K	1616.51	43%	164%	264%
S18T, C63H, S96T, L177H	2088.12	43%	189%	197%
L65F, S96T, A281V, E283K	1638.28	43%	144%	191%
C63H, L65F, A281V, E283K	632.66	43%	56%	190%
S18T, C63Y	1640.33	43%	166%	260%
L65F, S96T, L177H, A281V, E283K	1585.32	42%	140%	189%
C63Y, L65F, L177H	639.46	41%	56%	184%
C63H, L65F, S96T, A281V	1177.76	41%	119%	252%
L11F, C63Y, S96T, L177H, A281V A282T, E283K	1578.54	41%	143%	186%
S18T, C63Y	1797.99	39%	182%	241%
S18T, L178H, A282V, E284K	1502.59	39%	152%	237%
G23S, C63H, S96T, L177H, E283K	1229.34	39%	125%	237%
S18T, C63H, L177H	1811.40	39%	160%	173%
C63Y, A281V, A282T	1622.23	38%	143%	168%
S18T, C63Y, L65F, S96T, L177H A281V, E283K, M284I	454.42	35%	41%	162%
S18T, L65F, S96T, L177H, A281V	1135.95	33%	115%	203%
L65F, L178H, G180S, E283K	151.40	33%	13%	146%
L66F, L177H, G180S, E283K	143.83	31%	13%	138%
S18T, C63H, L65F, L177H	103.21	30%	9%	133%
	포화 라이브리			
	FALC 역가 (총계*)	C14-FALC (% 분율)	정상화 FALC (% 대조군 증가)	정상화 C14 (% 대조군 증가)
S18W	2530	55%	227%	314%
S18F	2564	54%	222%	285%
V17N	571	49%	49%	260%
S18L	1568	48%	141%	271%
S18L	1024	47%	92%	268%
S18L	1326	47%	119%	268%
S18L	1302	47%	117%	267%
S18L	1106	47%	99%	267%
S18L	1315	47%	118%	266%
S18L	1359	47%	122%	266%
S18Y	2550	45%	221%	236%
S18Y	2039	44%	188%	252%
S18Y	1997	44%	179%	251%
S18M	2495	41%	224%	231%
S18M	2809	40%	243%	212%
V17L	1127	33%	97%	175%

[0174]

V17L	1185	33%	103%	173%
V17L	1091	32%	94%	171%
V17L	1175	32%	102%	171%
S18T	1729	31%	155%	177%
S18T	1690	31%	152%	176%
S18T	1683	31%	151%	176%
S18T	1599	30%	143%	172%
S18V	1344	30%	121%	170%
S18V	1306	30%	121%	170%
S18V	1015	30%	91%	169%
S18C	1381	28%	120%	149%
S18C	1368	28%	118%	148%
R19H, R58S	873	27%	81%	152%
V17C	927	27%	80%	140%
V17C	882	26%	76%	139%
S18C	1313	26%	118%	149%
S18C	1355	26%	121%	148%
S18C	1305	26%	117%	148%
S18C	1372	26%	123%	147%
S18C	1340	26%	120%	147%
V17W	849	25%	73%	134%
V17W	815	25%	71%	133%
V17W	831	25%	72%	132%
V17W	840	25%	73%	131%
R19V	844	24%	78%	139%
R19V	813	24%	75%	136%
R19V	821	24%	76%	135%
R19V	862	24%	80%	135%
R19V	814	24%	75%	135%
R19K, V117L	1188	23%	110%	130%
R19S	803	22%	74%	128%
R19T	735	22%	68%	127%
R19S	783	22%	72%	127%
R19S	764	22%	75%	129%
R19I	862	22%	80%	123%
R19A	901	21%	83%	122%
R19A	705	21%	65%	119%
R19M	937	20%	86%	114%
	콤보 β 라이브리			
	FALC 역가 (총계*)	C14-FALC (% 분율)	정상화 FALC (% 대조군 증가)	정상화 C14 (% 대조군 증가)
S18W	3184	57%	228%	296%
S18W,C63Y	2675	54%	191%	280%
S18W,C63Y,S95T	2832	54%	203%	279%
S18W,S96T,E283K	2874	54%	206%	278%
S18W,C63Y,E283K	2915	51%	209%	263%
S18W,S96T,E283K,R286Q,A324V	3437	49%	246%	255%
C63Y,S96T	2583	38%	185%	196%

[0175]

C63Y,E283K	2176	38%	156%	195%
C63Y,E283K,Q316K	2420	36%	173%	188%
E283K	1474	25%	105%	132%
S96T	1605	23%	115%	121%
콤보 3 라이브러리				
	FALC 역가 (총계*)	C14-FALC (% 분율)	정상화 FALC (% 대조군 증가)	정상화 C14 (% 대조군 증가)
S18W	2853	217%	51%	274%
Y23N	954	72%	35%	191%
Y23N, Q238H	892	68%	35%	190%
G150D, I274N	766	62%	30%	162%
I274N	1359	107%	29%	161%
I274N, A285V	1317	103%	28%	155%
P38T, I274N, A285V	587	46%	26%	143%
Q238H, M291V	1383	109%	22%	119%
M291V	1618	123%	22%	118%
G151D, M291V	1302	105%	21%	114%
T135M, M291V	1265	100%	18%	101%

예시 4: 주형으로서 AAR(S18W)을 이용하여 제조된 포화 라이브러리

시네코코쿠스 엘롱가투스 PCC7942("AAR(S18W)_7942")로부터의 아실-ACP 레덕타아제 변이체의 전체 포화 라이브러리가 구축되었으며, 스크리닝의 제 1 라운드에서 상당히 개선된 AAR 변이체로서 식별된, AAR(S18W)을 넘어서는 개선들을 나타낸 변이체들에 대해 스크리닝되었다(예시 2 및 3). 선택 기준은 C12의 퍼센트 분율의 증가 또는 FALC 역가의 증가였다. 높은 역가를 산출하기 위해 ACP 과발현에 대한 AAR의 의존성을 완하시키는 것에 조작 노력이 집중되었다. 이론에 얽매이고자 하는 것은 아니지만, ACP를 과발현시킨 균주들에서 관찰된 장점은 AAR에 의한 분할(cleavage)에 이용가능한 지방산 생합성 중간체들의 더 높은 농도에 기인한다는 가설을 세웠다. (FAS 중간체들의 더 낮은 농도를 갖는) 균주 결여 ACP 과발현에서 AAR에 구축된 포화 및 조합 라이브러리들을 스크리닝함으로써, FAS 중간체들에 대해 더 높은 친화성을 갖는 변이체들이 선택될 수 있다.

전체 포화 라이브러리를 만드는데 사용된 플라스미드는 pAAR-1로 지정되었다. 이는 *아시네토박터균* (*Acinetobacter baylyi*)(SEQ ID NO: 54)으로부터 알데히드 레덕타아제 유전자, AlrA에 의해 후속되는 (S18W) 변이체를 인코딩하는 AAR 유전자를 포함하는(harboring) pLS9-185의 유도체이다. AlrA는 AAR에 의해 생성되고 대장균의 내인성 지방족 알데히드 레덕타아제에 의해 완전히 환원되지 않은 지방족 알데히드 중간체들을 완전히 환원시키도록 첨가되었다. 전체 포화 라이브러리는 균주 Shu.002에서 스크리닝되었다. 균주 Shu.002는 DV2 PT5-ifab138 PT5_ifadR이다(표 7, 아래 참조). ifab138의 경우, 표 10의 SEQ ID NO: 55를 참조한다. 라이브러리들은 앞서 설명된 표준 프로토콜들 중 하나를 이용하여 스크리닝되었다. 개선들은 역가를 개선하거나, 역가에 실질적인 영향을 주지 않고 아실-ACP 레덕타아제에 의해 생성된 C12 지방족 알코올의 분율을 증가시키는 것으로 분류되었다. 포화 라이브러리의 스크리닝으로부터 얻어진 결과들이 아래의 표 5에 나타나 있다.

표 5: 증가된 지방족 알코올 역가 및/또는 증가된 C12 지방족 알코올 분율과 상호관련된 AAR(S18W) 7942 전체 포화 라이브러리로부터의 돌연변이들

AAR 돌연변이 (S18W 추가)	FALC 역가*		C12	
	총 계*	FIOC**	분율*	FIOC**
T148V	747	1.21	7.9%	0.99
A159V	702	1.14	7.0%	0.87
T157V	674	1.10	7.2%	0.90
A135S	803	1.31	7.8%	0.98
A328S	712	1.16	8.1%	1.02
Q191A	780	1.27	7.7%	0.96
A285V, M291V	837	1.36	8.0%	1.00
Q277V	854	1.39	7.9%	0.99
Q155C	626	1.02	8.3%	1.04
E210Y	676	1.10	8.5%	1.06
T120S	623	1.01	8.1%	1.02
T236C	691	1.12	7.6%	0.95
Q335N	763	1.24	6.8%	0.86
C172L	700	1.14	7.7%	0.97
E283S	858	1.39	8.0%	1.00
L209R	837	1.36	7.6%	0.96
I153P	606	0.99	8.0%	1.00
A211W	797	1.30	6.9%	0.86
A324T	909	1.48	4.0%	0.50
W34F	817	1.22	7.5%	0.98
T187V	825	1.23	6.9%	0.91
D24E	737	1.10	7.1%	0.93
T188H	862	1.29	8.6%	1.13
V18A	798	1.19	6.4%	0.84
V18A, L338W	732	1.27	7.6%	0.97
T188V	775	1.34	7.4%	0.95
I168V	731	1.27	8.1%	1.03
W35F	666	1.15	7.5%	0.96
T148V	836	1.45	6.2%	0.80

Q155L	740	1.28	8.5%	1.08
T148C	731	1.26	8.6%	1.10
T148E	694	1.20	8.3%	1.06
A50Q, L337V	800	1.38	7.2%	0.92
R118Q	724	1.25	7.3%	0.94
L31V	901	1.56	6.8%	0.87
A135S	775	1.34	8.7%	1.12
D24Y	890	0.93	19.1%	2.05
C63A	895	0.94	16.1%	1.73
S113K	812	0.85	13.3%	1.43
L31V	645	0.68	15.7%	1.69
E283G	917	0.96	14.3%	1.54
A112R	918	0.96	10.7%	1.15
D43E	1002	1.05	13.0%	1.39
Q116G	894	0.94	14.0%	1.50
S86G	849	0.89	13.4%	1.44

* 4회 복제로부터 얻어진 평균(Average from 4 replicates)

** FIOC = AAR(S18W) 대조군을 넘어서는 배 증가[fold increase over AAR(S18W) control]

예시 5: 주형으로서 AAR(S18W)을 이용하여 제조된 조합 라이브러리

해당 기술분야의 당업자에게 알려진 표준 기술들은 조합 라이브러리를 제조하는데 사용되었다. 조합 라이브러리에서 테스트된 돌연변이들(표 6A, 6B 및 6C)은 전체 포화 라이브러리에서 근원적으로 식별되었다(예시 4). 조합 라이브러리는 예시 3에 설명된 것과 동일한 플라스미드에서 구축되었고 동일한 균주에서 스크리닝되었다.

이 라이브러리는 앞서 설명된 표준 프로토콜들 중 하나를 이용하여 스크리닝되었다. 개선들은 역가를 개선하거나, 역가에 실질적인 영향을 주지 않고 아실-ACP 레덕타아제에 의해 생성된 C12 지방족 알코올의 분율을 증가시키는 것으로 분류되었다. AAR 조합 라이브러리의 스크리닝으로부터 얻어진 결과들이 아래의 표 6A, 6B 및 6C에 나타나 있다.

표 6A: 증가된 지방족 알코올 역가와 상호관련된 제 1 AAR(S18W)_7942 조합 라이브러리로부터의 돌연변이들

AAR 돌연변이 (S18W 추가)	FALC 역가*		C12	
	mg/L*	FIOC**	분율*	FIOC**
A281L	1478	1.29	14.1%	1.11
A281Y	1527	1.28	12.1%	0.96
T154A, A281L	1550	1.28	14.3%	1.13
T154A, A281Y	1384	1.19	13.3%	1.05

C63G, A281F	1472	1.24	12.9%	1.02
C63G, A281L	1432	1.19	18.2%	1.44
N83G, A281Y	1428	1.23	10.7%	0.84
C63G, A281Y	1419	1.21	15.6%	1.24
S10G, A281L	1419	1.20	12.6%	1.00
S113K, Q155G	1305	1.15	12.3%	0.97
D43E, A50Q, A281L	1517	1.32	9.1%	0.72
M21L, N83G, A281Y	1513	1.29	10.2%	0.81
M21L, S113K, A281L	1392	1.20	20.4%	1.61
C63G, T154A, A281F	1369	1.24	12.4%	0.98
C63G, N83G, A281Y	1330	1.32	14.6%	1.16
Q155G, A281L, E283G	1327	1.16	22.2%	1.76
A50Q, C63G, A281F	1286	1.15	17.9%	1.41
S18W, D43E, A50Q, A281L	1635	1.32	9.4%	0.77
M21L, C63G, S113K, T154A, A281L	1518	1.24	28.3%	2.23
M21L, L31V, N83G, Q116G, A281L	1353	1.18	15.3%	1.21

* 4회 복제의 평균

** FIOC = (S18W) 대조군을 넘어서는 배 증가

표 6B: 증가된 지방족 알코올 역가와 상호관련된 제 2 AAR(S18W)_7942 조합 라이브러리로부터의 돌연변이들

AAR 돌연변이 (S18W 추가)	FALC 역가*		C12	
	mg/L*	FIOC**	분율*	FIOC**
S18W (대조군)	1034	1.00	10.0%	1.00
M21L, C63G, S113K, T154A, A281L	1179	1.14	29.4%	2.94
D16L, M21L, C63G, S113K, T154A, A281L	1221	1.18	19.4%	1.94
L8A, D24V, C63G, S113K, Q155L, A281L	1414	1.37	10.4%	1.04
L8A, M21L, C63G, A77A***, S113K, T154A, A281L	1310	1.27	14.2%	1.42
D24P, L31M, C63G, S113K, T154A, A281L	1437	1.39	13.0%	1.30
L8A, D16L, D24V, C63G, S113K, T154A, A281L	1342	1.30	12.0%	1.20
D24E, C63G, S113K, T154A, A281L	1425	1.38	14.3%	1.43

* 4회 복제로부터 얻어진 평균

** FIOC = (S18W) 대조군을 넘어서는 배 증가

*** A77A 돌연변이는 gcc 내지 gca 잠재적 코돈 돌연변이임

표 6C: C12 지방족 알코올의 증가된 분율과 상호관련된 AAR(S18W)_7942 조합 라이브러리로부터의 돌연변이들

AAR 돌연변이 (S18W 추가)	C12		FALC 역가*
	분율*	FIOC**	FIOC**
S18W (대조군)	11.0%	1.00	1.00
M21L, C63G, S113K, T154A, A281L	28.3%	2.57	1.24
M21L, Q155G, A281L	24.8%	2.25	1.10
Q155G, A281L, E283G	22.2%	2.02	1.16
M21L, S113K, A281L	20.4%	1.85	1.20
C63G, A281L	18.1%	1.65	1.18
A50Q, C63G, A281F	17.9%	1.63	1.15
C63G	17.7%	1.61	1.00
S10G, C63G, Q155G, S253N, A281F	16.1%	1.46	1.03
D43E, C63G, S113K, A281L	16.0%	1.46	0.98
A281L	15.6%	1.42	1.20
M21L, L31V, N83G, Q116G, A281L	15.3%	1.39	1.18
C63G, N83G, A281Y	14.6%	1.33	1.32
A281L	14.1%	1.28	1.29

* 4회 복제로부터 얻어진 평균

** FIOC = (S18W) 대조군을 넘어서는 배 증가

예시 6: 지방산 합성 경로를 통해 증가된 플럭스 - AAR을 이용하여 iFab 및 iFadR 매개되는 지방족 알코올 생성 이 예시에서는, 지방산 대사의 조절, FadR 단백질(ifad)의 과발현 및/또는 수 개의 FAB 단백질(ifab138)을 포함하는 합성 오페론의 과발현에 의해 매개되는 지방산 합성 경로를 통해 플럭스를 증가시킴으로써, AAR을 이용하는 개선된 지방족 알코올 생성을 나타내었다. iFAB138(SEQ ID NO:55)은 다음의 순서로 유전자들 *콜레라균*으로부터의 *fabV*, *퀴터푸스균*으로부터의 *FabH*, *fabD*, *fabG* 및 *fabA*, 그리고 *클로스트리디움 아세토부틸리쿰*으로부터의 *FabF*를 포함하며, 이는 PlacUV5 또는 PT5 프로모터에 의해 제어되는 합성 오페론으로서 통합된다. iFadR은 T5 프로모터에 의해 제어되는 *대장균*(SEQ ID NO: 56)으로부터의 *fadR* 유전자를 포함한다. 이 예시에서 평가된 *대장균* 균주에 존재하는 성분들은 아래의 표 7에 나타나 있다.

표 7: iFAB138 또는 fadR을 갖는 *대장균* 균주

성분	DV2	BD061	BD064	Shu002
i(PlacUV5-fab138)	-	+	+	-
i(PT5-fab138)	-	-	-	+
i(PT5-fadR)	-	-	+	+

AAR(S18W)은 AAR 코돈 18이 세린 (pCL-AAR(S18W)+ACP-sfp) 대신 트립토판을 특정하는 플라스미드 pDS171S의 변이체인 플라스미드 pDS311로부터 발현되었고, *아시네토박터균*(SEQ ID NO: 54)으로부터의 알데히드 레덕타아제 유전자 *alrA*가 *B. 서브틸리스*로부터의 *sfp* 유전자의 하류에 클로닝되었다. pDS311은 균주들 DV2, BD061, BD064 및 Shu002로 형질전환되었다. 균주들은 4NBT_2N 프로토콜에서 평가되었다(상기 참조). 도 7에 도시된 바와 같이, 지방족 알코올 생성은 ifab138 및 ifadR의 존재의 결과로서 상당히 증가하였으며, 균주 Shu002는 가장 높은 지방족 알코올 역가를 나타낸다.

예시 7: 지방산 합성 경로를 통해 증가된 플럭스를 갖는 균주에서의 개선된 AAR 변이체들

이 예시에서는, 조합 라이브러리로부터의 AAR 변이체로 형질전환된 재조합 숙주 세포들에서의 개선된 지방족 알코올 생성이 FadR 단백질(위 참조)의 과발현 및 ifab138의 과발현에 의해 매개되는 지방산 생합성 경로를 통해 증가된 플럭스를 갖는 *대장균* 균주인 Shu2에서 입증되었다. 제 2 조합 라이브러리(표 6B)로부터의 4 개의 AAR 변이체가 평가되었다. 이러한 변이체들은 다음의 돌연변이들: Com2a: S18W, D24P, L31M, C63G, S113K, T154A, A281L; Comb2b: S18W, D16L, M21L, C63G, S113K, T154A, A281L; Com2c: S18W, L8A, M21L, C63G, A77A, S113K, T154A, A281L; Com2d: D24E, C63G, S113K, T154A, A281L을 포함한다(도 8 참조). 이러한 변이체들은 플라스미드 pJL104의 골격 내로 클로닝되었다. pJL104는 pDS311의 *alrA* 유전자의 하류에 *C. 글루타미쿰*(예시 2에 설명됨, 위 참조)으로부터의 합성 accD+ 오페론을 클로닝함으로써 생성되었다. 결과적인 플라스미드는 균주 Shu002 내로 형질전환되었고, 균주들은 4NBT_1N 프로토콜(위 참조)에서 평가되었다. 도 8에 도시된 바와 같이, 상승된

지방족 알코올 생성이 모든 균주들에서 관찰되었다. 또한, AAR_7942의 S18W 변이체를 발현시키는 플라스미드 pDS311을 포함하는 균주 BD064가 탱크 발효로 평가되었다. 균주는 도 9에 도시된 바와 같이 42 g/L의 최대 역가, (글루코오스 상에서의) 12.6 %의 수율 및 0.62 g/L/h의 생산성을 갖는 지방족 알코올을 생성하였다. 지방족 알코올의 사슬 길이 분포는 다음과 같다: 1.2 % C8, 7.7 % C10, 26.9 % C12, 44.7 % C14, 16.0 % C16 및 1.9 % C18. 포화 및 불포화 지방족 알코올의 분율은 각각 72.6 % 및 27.4 %이었다.

예시 8: C14 지방족 알코올 생성을 위한 사슬 길이 선택성의 재분포 및 증가된 MED4 아실 알데히드 레덕타아제 (AAR) 활성

AAR은 시아노박테리아 알칸 생합성에 필수적인 성분들 중 하나이며, 다른 필수적인 성분은 알데히드 탈카르보닐 라아제(aldehyde decarbonylase: ADC)이다. 본 발명자들은 MED4_ADC가 존재하지 않으면 *프로클로로코쿠스 마 리누스* MED4_AAR가 촉매 비활성임을, 즉 MED4_AAR만이 *대장균*에서 발현될 때에는 생성물들이 검출되지 않으며, MED4_AAR 및 ADC가 *대장균*에서 공동발현될 때에만 검출된 생성물들이 알칸임을 발견하였다. 또한, 본 발명자들은, MED4_AAR이, 히스티딘 156이 아르기닌[이후, MED4_ADC(H156R)라고도 함], 지방족 알코올에 의해 대체되는 MED4_ADC의 피상적 촉매 비활성 변이체(apparently catalytically inactive variant)를 갖는 *대장균*에서 공동 발현될 때, 알칸이 검출되지 않음을 발견하였다(도 10 참조). 이 데이터로부터, MED4_AAR은 촉매 활성화될 MED4_ADC와 물리적인 상호작용을 필요로 하는 것으로 결론을 내렸다. 본 발명자들은 FALC 생성 목적으로 증가 된 활성 및/또는 변경된 기질 특이성을 갖는 MED4_AAR을 식별하기 위해 이 시스템을 사용하였다.

MED4_ADC(H156R)는 MED4_AAR의 전체 포화 라이브러리와 함께 발현되었다. MED4_AAR 포화 라이브러리는 플라스 미드 pCL1920-유도체(pLS9-195)에서 제조되었고, (ADC 유전자에 H156R을 포함하는 플라스미드 pACYC-Ptrc-MED4_ADC인) 플라스미드 pGLAK-043을 갖는 생성 균주 내로 도입되었다. 클론들이 유도되었으며, 야생형 AAR 효 소보다 더 많은 지방족 알코올의 생성 또는 변경된 사슬 길이 프로파일, 예를 들어 C14 지방족 알코올의 증가된 분율을 갖는 지방족 알코올을 생성할 수 있는 능력에 기초하여 AAR 변이체들이 선택되었다. 이후, 선택된 클론 들은 검증 라운드(validation round)에서 다시 테스트되었다. 1차 및 2차 발효 중에 일관적인 FALC 역가를 나 타낸 모든 변이체들이 재성장되었고, 플라스미드 DNA가 격리되었으며, 서열화되었고, 추가 테스트를 위해 모 생 성 균주 내로 재도입되었다. 이후, 이러한 새로운 형질전환체들은 또 다른 확증적 발효 및 분석을 거쳤다. 아 래의 표 8은 가장 높은 FALC 역가를 생성한 16 개의 AAR 변이체들로부터의 대표적인 데이터를 나타낸다(위에서 부터 아래 순서로 활성 감소됨). 이러한 변이체들은 야생형보다 1.4-배 내지 2.2 배 범위를 가졌다. 이러한 변이체들은 방향성 진화 기술(directed evolution technique)들을 이용하여 MED4_AAR 활성을 증가시키고, 추가 적인 개선들에 대한 토대를 더욱 형성할 수 있는 능력을 나타내었다.

표 8: 야생형(WT) MED4 AAR에 대한 MED4 AAR 돌연변이체들의 FALC 생산성

AAR 돌연변이	야생형보다 FALC 배 증가
V346P	2.2
Q40V	2.2
A345R	2.1
L344S	2.1
D61E	2.1
V346G	2.0
L344D	1.9
G52V	1.9
A345*	1.9
L344T	1.8
K303G	1.8
L344A	1.8
H340P	1.6
S588V	1.5
K339L	1.4
G273E	1.4

* 잘린 변이체(Truncated variant)는 마지막 두 개의 아미노산이 빠져 있다.

또한, 데이터 세트는 변경된 사슬 길이 프로파일을 나타낸 AAR 변이체들에 대해 스캐닝되었다. 야생형 AAR에 의해 생성된 대다수의 종은 C16의 사슬 길이를 갖는다. C16보다 짧은 사슬 길이를 갖는 FALC 종의 증가된 비율

이 관심거리이다. C14의 양에서 거의 3-배 증가를 나타낸 2 개의 변이 클론들이 식별되었다(도 10). 이러한 클론들의 서열화는 이러한 클론들이 동일한 뉴클레오타이드 코돈 서열을 갖는 동일 D61E 돌연변이체들임을 나타내었다. D61E 변이체에 대한 플라스미드 DNA는 H156R ADC를 함유하는 모 균주 내로 재도입되었다. 결과들은 재조합 숙주 세포에서의 AAR의 D61E 변이체의 발현이 더 짧은 탄소 사슬들 쪽으로 FALC 종의 사슬 길이 분포를 치우치게 했다. 아래의 표 9는 변경된 사슬 길이를 갖는 어떤 생성물들도 생성하지 않은 MED4_AAR의 V346P 변이체 그리고 야생형(WT) MED4_AAR에 비해 MED4_AAR의 D61E 변이체를 발현시키는 재조합 숙주 세포들에 의해 생성된 FALC 사슬 길이를 예시한다.

[0215] 표 9: AAR 변이체 및 야생형 AAR에 대한 사슬 길이 분포

MED4 AAR/ADC	지방족 알코올		
	C14%	C16%	C18%
AAR WT/ADC(H156R)	6	92	2
AAR(D61E)/ADC(H156R)	14*	85	0
AAR(V346P)/ADC(H156R)	4	92	4
AAR WT/ADC WT	ND	ND	ND
AAR(D61E)/ADC WT	ND	ND	ND
AAR(V346P)/ADC WT	ND	ND	ND

[0216] ND = 검출되지 않음

[0217] * 평균으로부터의 편차에 기인한 1 % 분산(variance)

[0218] 이러한 변이체들은 더욱 조합될 수 있으며, MED4_ADC (H156R) 백그라운드(background)의 개선들에 대해 스크리닝될 수 있다. MED4_AAR (D61E) 변이체 및 추가 돌연변이된 자손이 지방족 알코올과 알칸 둘 모두의 평균 사슬 길이를 감소시키는데 유용할 수 있다. MED4_AAR (D61E) 변이체는, MED4_AAR 사슬-길이 특이성이 가변성적(malleable)이고, 추가 단백질 조작 노력을 통해 이 활성의 추가 개선에 대한 가능성을 제시함을 입증한다. 설명된 모든 변이체들은 pLS9-195의 자손으로부터 서열화되었으며, 목록화된 아미노산 치환체들에 대응하는 코돈 돌연변이들을 함유하였다.

[0219] 표 10: 서열 목록과 관련된 명칭

SEQ ID NO	유형	명칭
1	핵산서열	노스톡 펑크티포르멜 PCC 73102_acp 수탁번호 YP_001867863
2	아미노산 서열	노스톡 펑크티포르멜 PCC 73102_acp 수탁번호 YP_001867863
3	핵산서열	시네키크시스티스 종 PCC 6803_acp 수탁번호 NP_440632.1
4	아미노산 서열	시네키크시스티스 종 PCC 6803_acp 수탁번호 NP_440632.1
5	핵산서열	프로클로로코쿠스마리누스 야종 파스토리스 str. CCMP1986_acp 수탁번호 NP_893725.1
6	아미노산 서열	프로클로로코쿠스마리누스 야종 파스토리스 str. CCMP1986_acp 수탁번호 NP_893725.1
7	핵산서열	시네키크시스티스 종 PCC 7942_acp 수탁번호 YP_399555
8	아미노산 서열	시네키크시스티스 종 PCC 7942_acp 수탁번호 YP_399555
9	핵산서열	노스톡 종 PCC 7120_acp 수탁번호 NP_487382.1
10	아미노산 서열	노스톡 종 PCC 7120_acp 수탁번호 NP_487382.1
11	핵산서열	(합성된) B. 서브틸리스 sfp 수탁번호 X63158.1
12	아미노산 서열	(합성된) B. 서브틸리스 sfp 수탁번호 X63158.1
13	프라이머 서열	168IFF
14	프라이머 서열	168IFR
15	프라이머 서열	169IFF
16	프라이머 서열	169IFR
17	프라이머 서열	170IFF
18	프라이머 서열	170IFR
19	프라이머 서열	171IFF

[0220]

20	프라이머 서열	171IFR
21	프라이머 서열	172IFF
22	프라이머 서열	172IFR
23	프라이머 서열	168SIF
24	프라이머 서열	170SIF
25	프라이머 서열	171SIF
26	프라이머 서열	168SIFR
27	핵산서열	시네코코쿠스 엘롱가투스 PCC7942 YP_400611 (Synpcc7942_1594) 아실-CoA 레덕타아제 (AAR)
28	핵산서열	시네코코쿠스 엘롱가투스 PCC7942 YP_400611 (Synpcc7942_1594) 아실-CoA 레덕타아제 (AAR)
29	핵산서열	시네코시스티스 종 PCC6803 s110209 (NP_442146) AAR
30	아미노산 서열	시네코시스티스 종 PCC6803 s110209 (NP_442146) AAR
31	핵산서열	시아노테세 종 ATCC51142 cce_1430 (YP_001802846) AAR
32	아미노산 서열	시아노테세 종 ATCC51142 cce_1430 (YP_001802846) AAR
33	핵산서열	프로클로로코쿠스 마리누스 CCMP1986 PMM0533 (NP_892651) AAR
34	아미노산 서열	프로클로로코쿠스 마리누스 CCMP1986 PMM0533 (NP_892651) AAR
35	핵산서열	글레오박터 비올라세우스 PCC7421 NP_96091 (g113145) AAR
36	아미노산 서열	글레오박터 비올라세우스 PCC7421 NP_96091 (g113145) AAR
37	핵산서열	노스톡 펑크티포르메 PCC73102 ZP_00108837 (Npun02004176) AAR
38	아미노산 서열	노스톡 펑크티포르메 PCC73102 ZP_00108837 (Npun02004176) AAR
39	핵산서열	아나베나 바리아발리스 ATCC29413 YP_323044 (Ava_2534) AAR
40	아미노산 서열	아나베나 바리아발리스 ATCC29413 YP_323044 (Ava_2534) AAR
41	핵산서열	시네코코쿠스 엘롱가투스 PCC6301 YP_170761 (syc0051_d) AAR
42	아미노산 서열	시네코코쿠스 엘롱가투스 PCC6301 YP_170761 (syc0051_d) AAR
43	핵산서열	노스톡 종 PCC7120a1r5284 (NP_489324) AAR
44	아미노산 서열	노스톡 종 PCC7120a1r5284 (NP_489324) AAR
45	핵산서열	코리네파테리움 글루타미쿰 (YP_224991)으로부터의 <i>birA</i>
46	합성 DNA	코리네파테리움 글루타미쿰 (YP_224991)으로부터의 <i>birA</i>
47	아미노산 서열	코리네파테리움 글루타미쿰 (YP_224991)으로부터의 <i>birA</i>
48	핵산서열	코리네파테리움 글루타미쿰 (YP_224991)으로부터의 <i>accDA1 (dfsR)</i>
49	아미노산 서열	코리네파테리움 글루타미쿰 (YP_224991)으로부터의 <i>accDA1 (dfsR)</i>
50	핵산서열	코리네파테리움 글루타미쿰 (YP_224991)으로부터의 <i>accCB</i>
51	아미노산 서열	코리네파테리움 글루타미쿰 (YP_224991)으로부터의 <i>accCB</i>
52	아미노산 서열	AlrA 아시네토박터 종 M-1
53	아미노산 서열	AlrAadp1
54	아미노산 서열	alrAadp1 아시네토박터 ADP1-WT 단백질
55	핵산서열	iFAB138
56	핵산서열	대장균 MG1655 (NP_415705)으로부터의 FadR
57	아미노산 서열	돌연변이 S18W를 갖는 AAR 돌연변이체(시네코코쿠스엘롱가투스 PCC7942 YP_400611 (Synpcc7942_1594) 아실-CoA 레덕타아제(AAR)로부터 만들어짐)

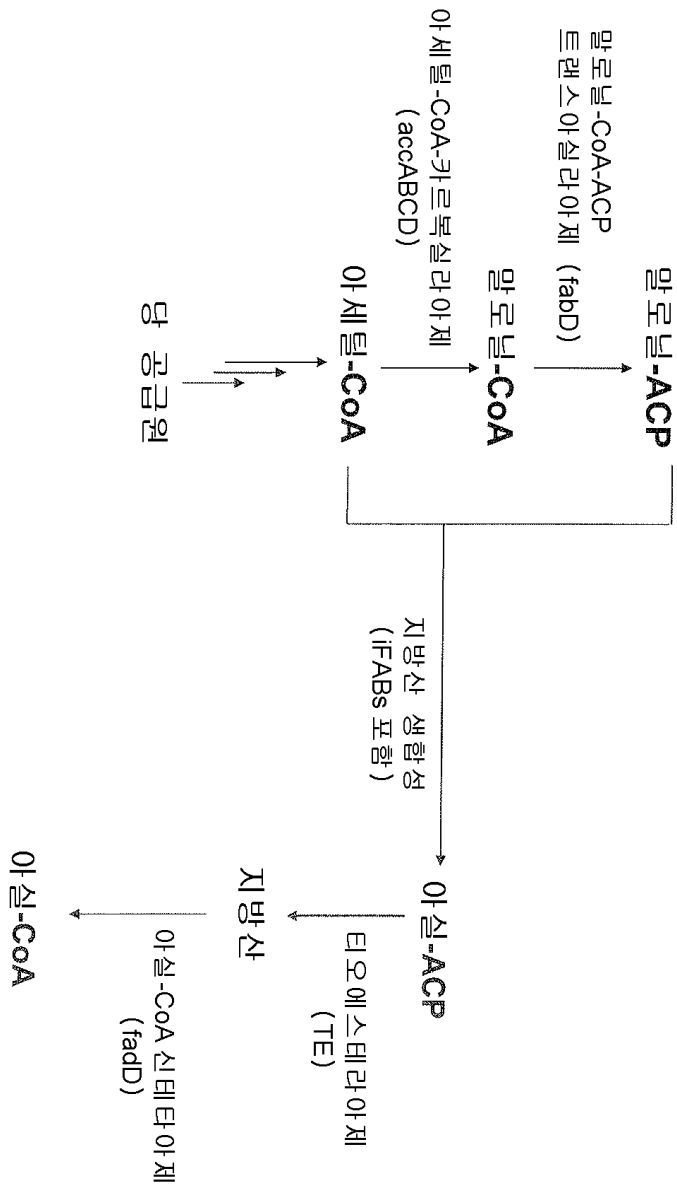
[0221]

[0222]

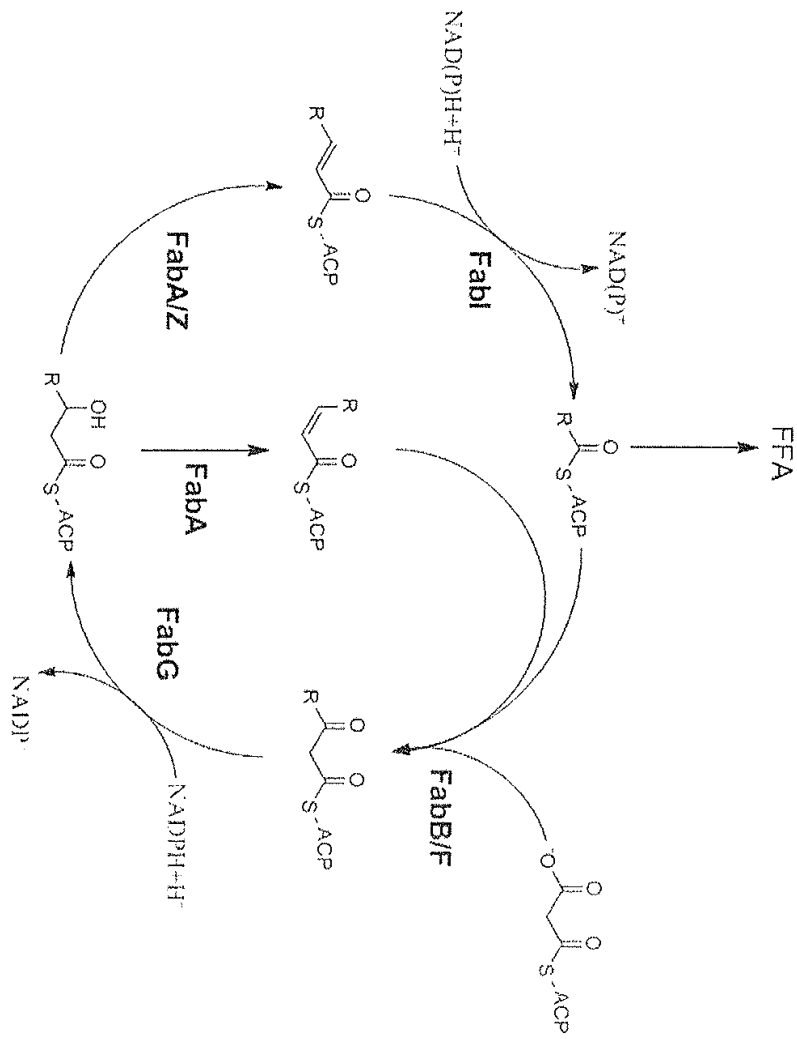
당업자라면 이해할 수 있는 바와 같이, 앞선 실시형태들 및 실시예들의 다양한 변형들 및 변경들이 본 발명의 기술사상 및 범위를 벗어나지 않고 행해질 수 있다. 이러한 변형들 및 변경들은 본 발명의 범위 내에 있다.

도면

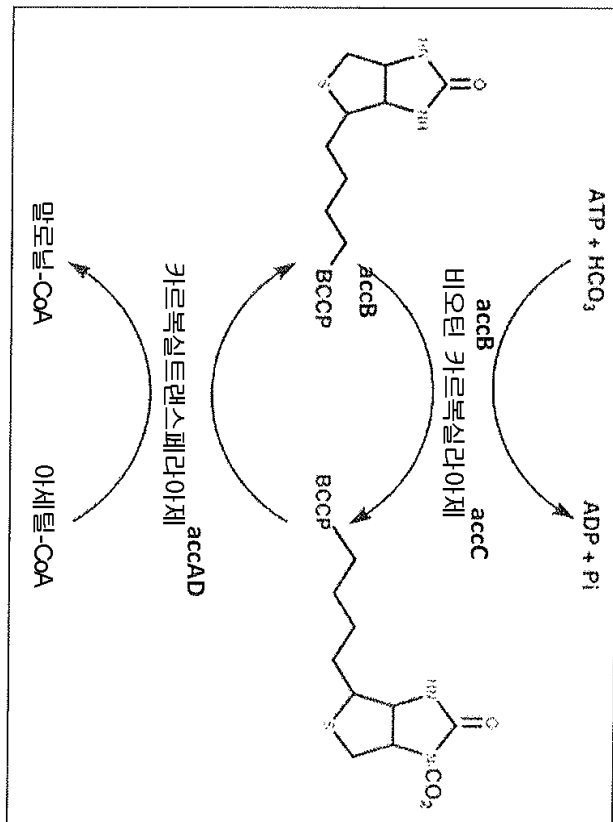
도면1



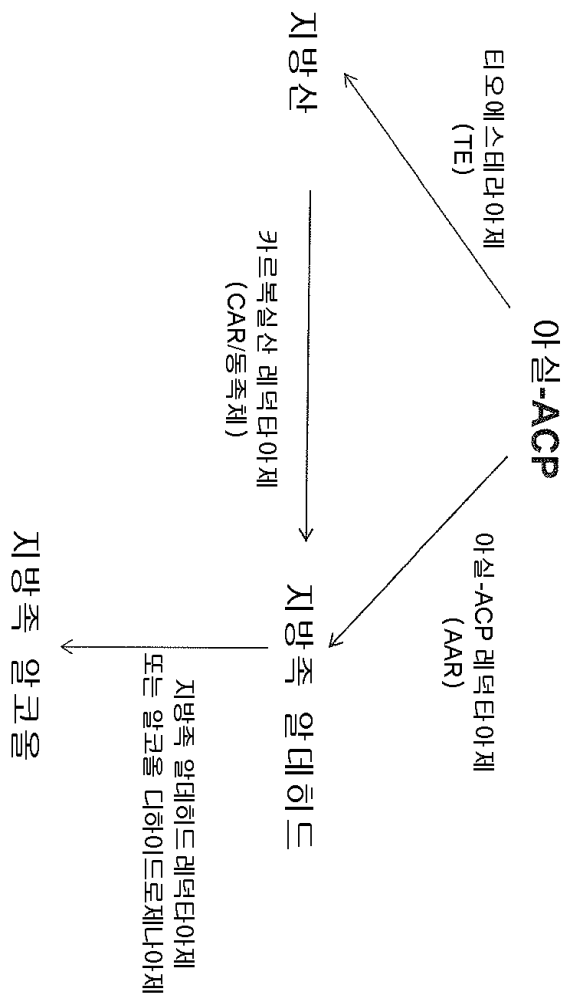
도면2



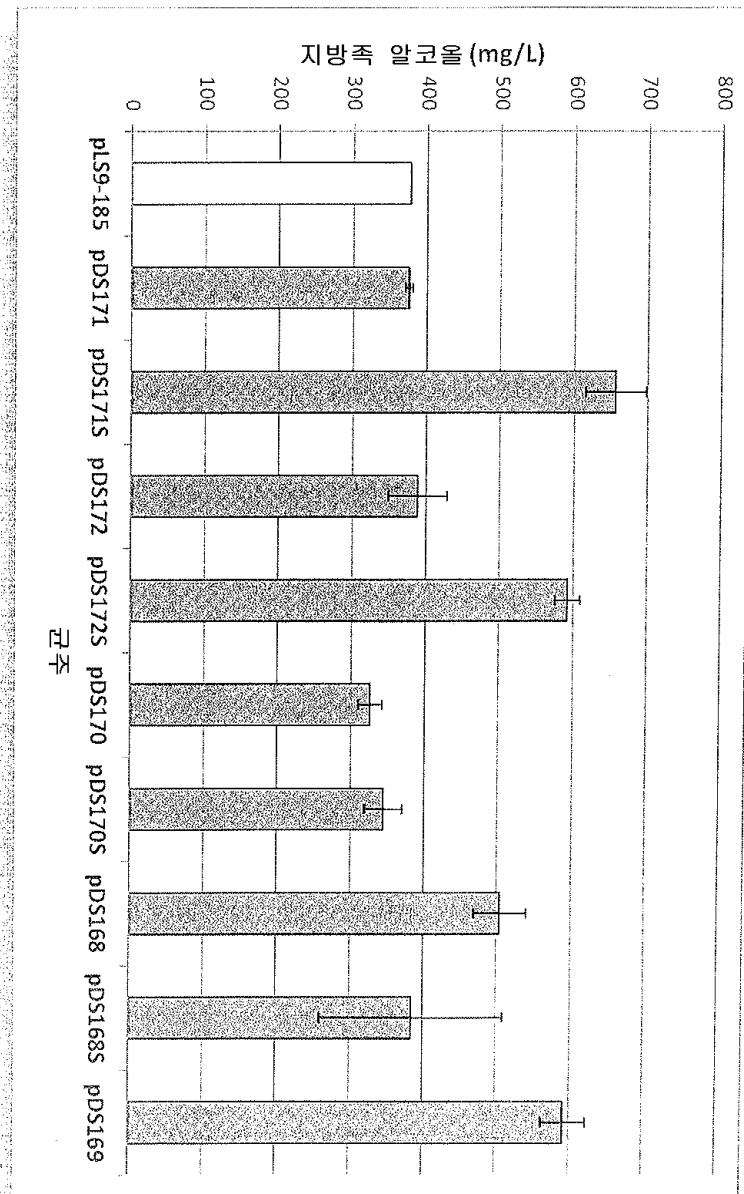
도면3



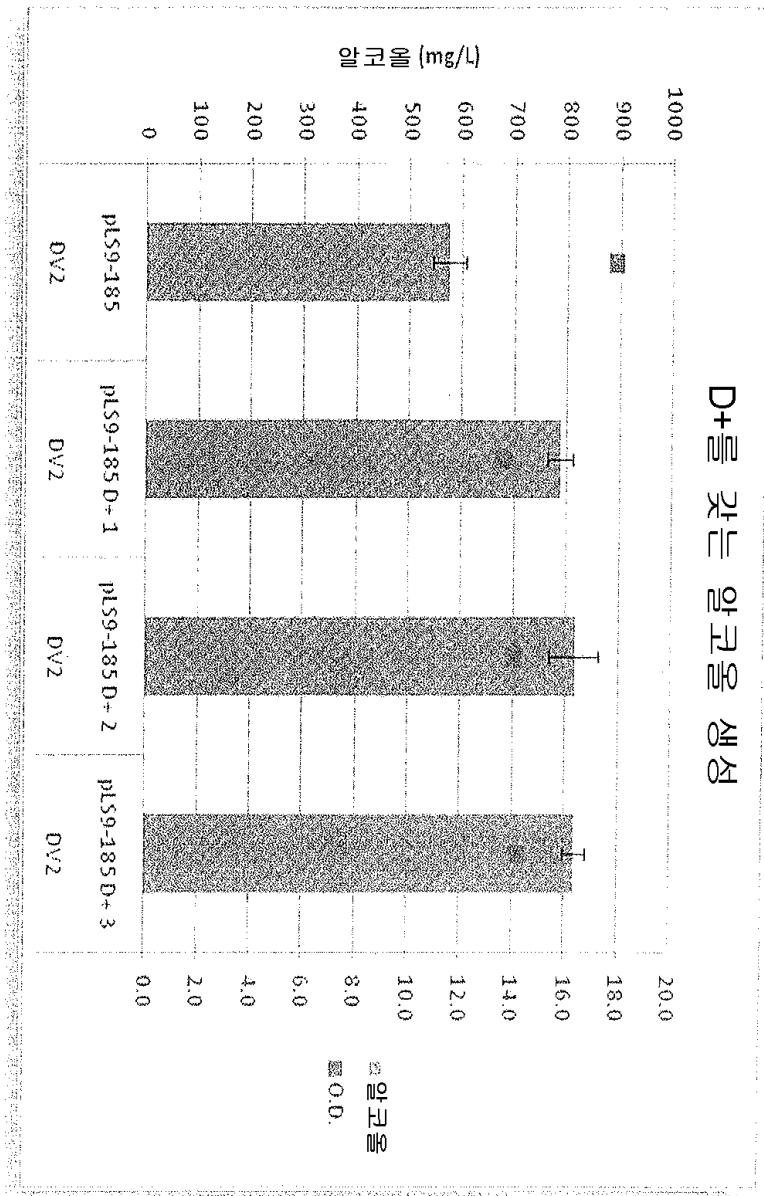
도면4



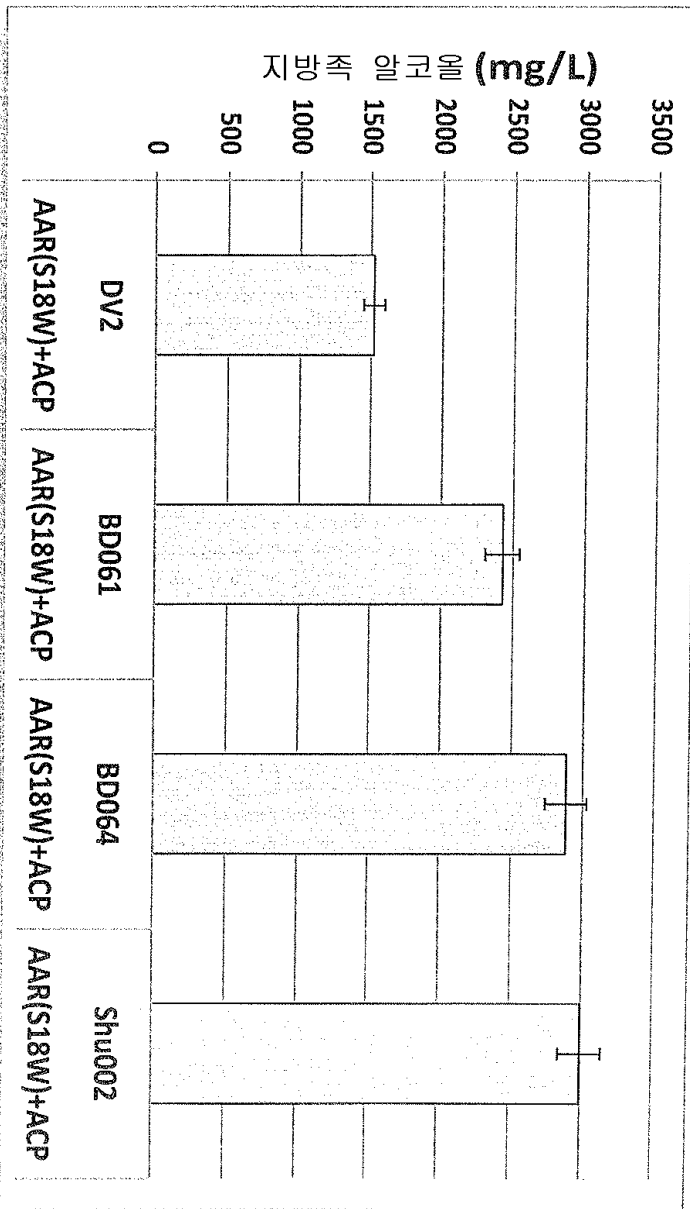
도면5



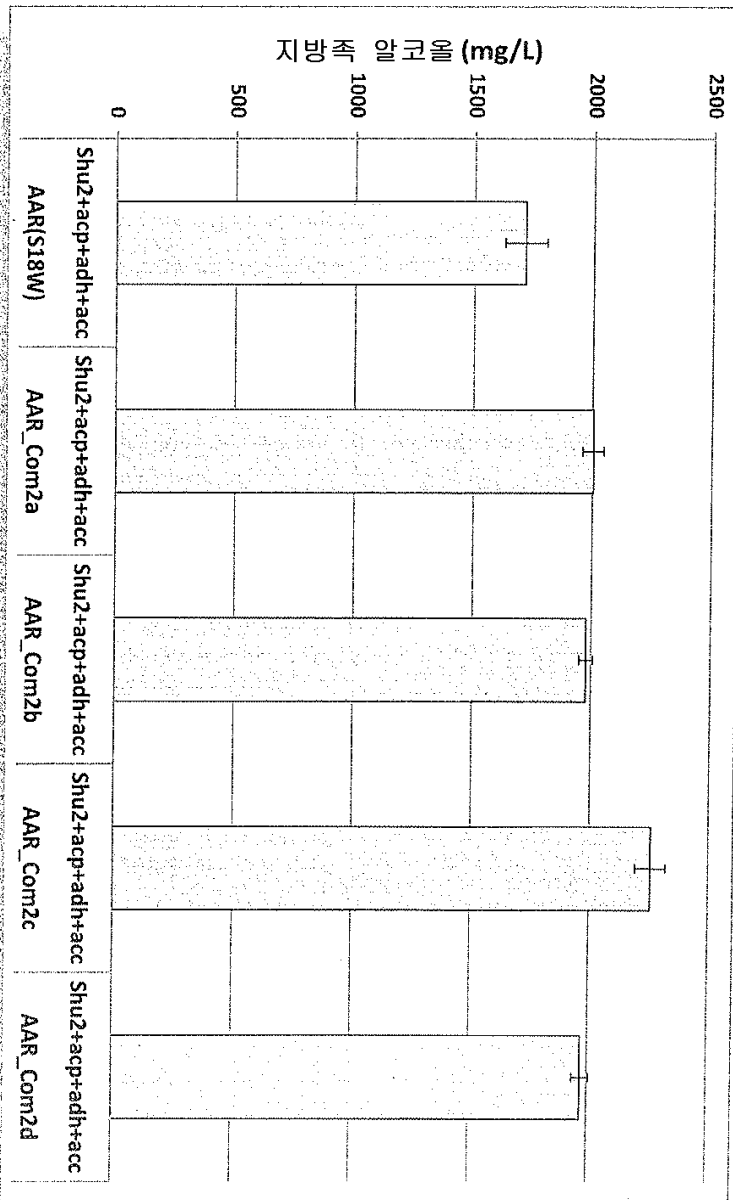
도면6



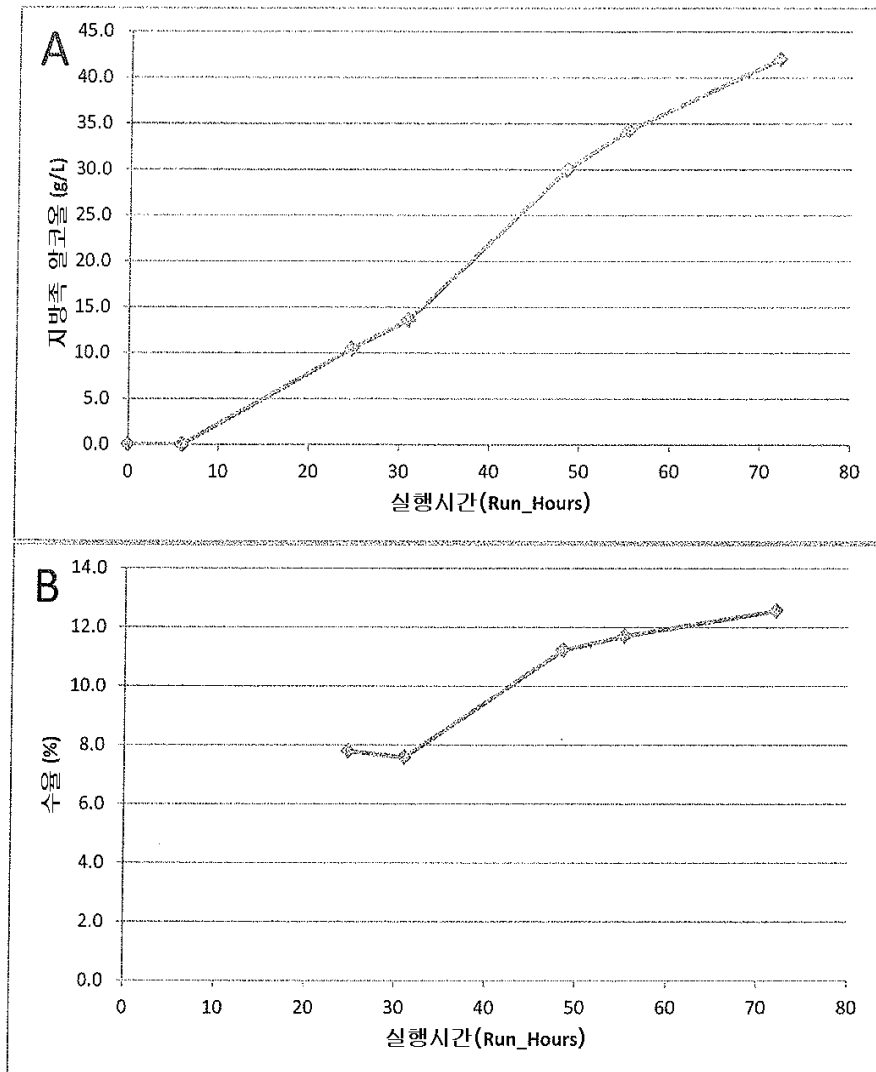
도면7



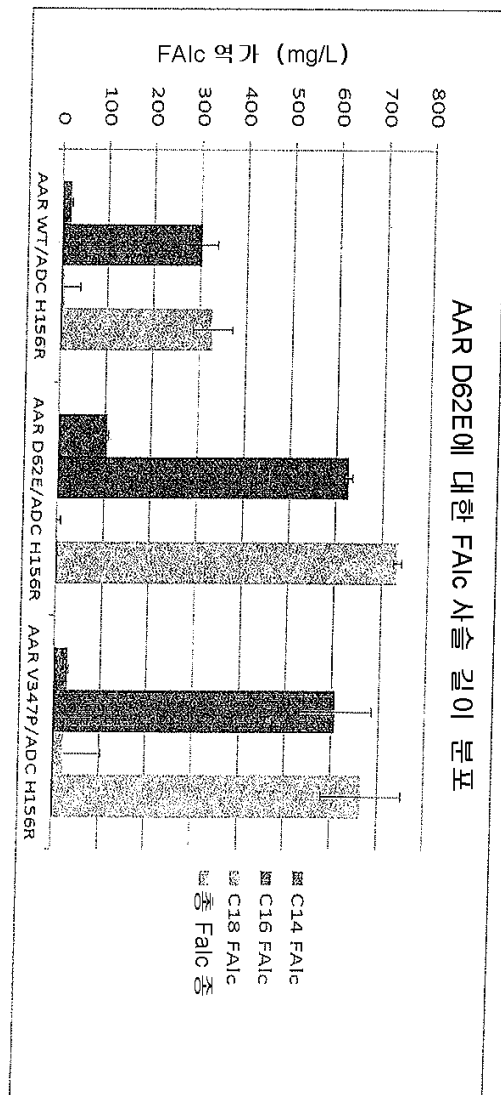
도면8



도면9



도면10



서열 목록

SEQUENCE LISTING

<110> GENOMATICA, INC.

<120> ACYL-ACP REDUCTASE WITH IMPROVED PROPERTIES

<130> FPA11964-D03/C

<150> PCT/US2014/011859

<151> 2014-01-16

<150> 61/753,273

<151> 2013-01-16

<160> 78

<170> PatentIn version 3.5

<210> 1

<211> 255

<212> DNA

<213> Nostoc punctiforme

<400> 1

atgagccaaa cggaactttt tgaaaaggtc aagaaaatcg tcatcgaaca actgagtgtt	60
gaagatgctt ccaaaatcac tccacaagct aagtttatgg aagatttagg agctgattcc	120
ctggatactg ttgaactcgt gatggctttg gaagaagaat ttgatatcga aattcccgac	180
gaagctgccg agcagattgt atcggttcaa gacgcagtag attacatcaa taacaaagtt	240
gctgcatcag cttaa	255

<210> 2

<211> 84

<212> PRT

<213> Nostoc punctiforme

<400> 2

Met Ser Gln Thr Glu Leu Phe Glu Lys Val Lys Lys Ile Val Ile Glu	
1 5 10 15	
Gln Leu Ser Val Glu Asp Ala Ser Lys Ile Thr Pro Gln Ala Lys Phe	
20 25 30	
Met Glu Asp Leu Gly Ala Asp Ser Leu Asp Thr Val Glu Leu Val Met	
35 40 45	
Ala Leu Glu Glu Glu Phe Asp Ile Glu Ile Pro Asp Glu Ala Ala Glu	
50 55 60	
Gln Ile Val Ser Val Gln Asp Ala Val Asp Tyr Ile Asn Asn Lys Val	
65 70 75 80	
Ala Ala Ser Ala	

<210> 3

<211> 234

<212> DNA

<213> Synechocystis sp.

<400> 3

atgaatcagg aaatttttga aaaagtaaaa aaaatcgtcg tggaacagtt ggaagtggat	60
---	----

cctgacaaag tgacccccga tgccaccttt gccgaagatt taggggtga ttcctcgat 120
acagtggaat tggcatggc cctggaagaa gagtttgata ttgaaattcc cgatgaagtg 180
gcggaacca ttgataccgt gggcaaagcc gttgagcata tcgaaagtaa ataa 234

<210> 4

<211> 77

<212> PRT

<213> Synechocystis sp.

<400> 4

Met Asn Gln Glu Ile Phe Glu Lys Val Lys Lys Ile Val Val Glu Gln

1 5 10 15

Leu Glu Val Asp Pro Asp Lys Val Thr Pro Asp Ala Thr Phe Ala Glu

20 25 30

Asp Leu Gly Ala Asp Ser Leu Asp Thr Val Glu Leu Val Met Ala Leu

35 40 45

Glu Glu Glu Phe Asp Ile Glu Ile Pro Asp Glu Val Ala Glu Thr Ile

50 55 60

Asp Thr Val Gly Lys Ala Val Glu His Ile Glu Ser Lys

65 70 75

<210> 5

<211> 243

<212> DNA

<213> Prochlorococcus marinus

<400> 5

atgtcacaag aagaaatcct tcaaaaagta tgctctattg tttctgagca actaagtgtt 60

gaatcagccg aagtaaaatc tgattcaaac tttcaaaatg atttaggtgc agactcccta 120

gacaccgtag agctagttag ggctcttgaa gaagcatttg atatcgagat acctgatgaa 180

gcagctgaag gtatcgcaac agtaggagat gctgttaaata tcatcgaaga aaaaaaaggt 240

taa 243

<210> 6

<211> 80

<212> PRT

<213> Prochlorococcus marinus

<400> 6

Met Ser Gln Glu Glu Ile Leu Gln Lys Val Cys Ser Ile Val Ser Glu
1 5 10 15

Gln Leu Ser Val Glu Ser Ala Glu Val Lys Ser Asp Ser Asn Phe Gln
20 25 30
Asn Asp Leu Gly Ala Asp Ser Leu Asp Thr Val Glu Leu Val Met Ala
35 40 45
Leu Glu Glu Ala Phe Asp Ile Glu Ile Pro Asp Glu Ala Ala Glu Gly
50 55 60
Ile Ala Thr Val Gly Asp Ala Val Lys Phe Ile Glu Glu Lys Lys Gly
65 70 75 80

<210> 7

<211> 243

<212> DNA

<213> *Synechococcus elongatus*

<400> 7

atgagccaag aagacatctt cagcaaagtc aaagacattg tggctgagca gctgagtgtg 60
gatgtggctg aagtcaagcc agaatccagc ttccaaaacg atctgggagc ggactcgctg 120
gacaccgtgg aactggtgat ggctctggaa gaggctttcg atatcgaaat ccccgatgaa 180
gccgctgaag gcattgcgac cgttcaagac gccgtcgatt tcatcgctag caaagctgcc 240
tag 243

<210> 8

<211> 80

<212> PRT

<213> *Synechococcus elongatus*

<400> 8

Met Ser Gln Glu Asp Ile Phe Ser Lys Val Lys Asp Ile Val Ala Glu
1 5 10 15
Gln Leu Ser Val Asp Val Ala Glu Val Lys Pro Glu Ser Ser Phe Gln
20 25 30
Asn Asp Leu Gly Ala Asp Ser Leu Asp Thr Val Glu Leu Val Met Ala
35 40 45
Leu Glu Glu Ala Phe Asp Ile Glu Ile Pro Asp Glu Ala Ala Glu Gly

50

55

60

Ile Ala Thr Val Gln Asp Ala Val Asp Phe Ile Ala Ser Lys Ala Ala

65

70

75

80

<210> 9

<211> 255

<212> DNA

<213> Nostoc sp.

<400> 9

atgagccaat cagaaacttt tgaaaaagtc aaaaaaattg ttatcgaaca actaagtgtg	60
gagaaccctg acacagtaac tccagaagct agttttgcc aacgatttaca ggctgattcc	120
ctcgatacag tagaactagt aatggctttg gaagaagaat ttgatatcga aattcccgat	180
gaagccgcag agaaaattac cactgttcaa gaagcgggtgg attacatcaa taaccaagtt	240
gccgcatcag cttaa	255

<210> 10

<211> 84

<212> PRT

<213> Nostoc sp.

<400> 10

Met Ser Gln Ser Glu Thr Phe Glu Lys Val Lys Lys Ile Val Ile Glu	
1 5 10 15	
Gln Leu Ser Val Glu Asn Pro Asp Thr Val Thr Pro Glu Ala Ser Phe	
20 25 30	
Ala Asn Asp Leu Gln Ala Asp Ser Leu Asp Thr Val Glu Leu Val Met	
35 40 45	
Ala Leu Glu Glu Glu Phe Asp Ile Glu Ile Pro Asp Glu Ala Ala Glu	

50

55

60

Lys Ile Thr Thr Val Gln Glu Ala Val Asp Tyr Ile Asn Asn Gln Val

65

70

75

80

Ala Ala Ser Ala

<210> 11

<211> 674

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"

<400> 11

atgaagattt acggaattta tatggaccgc ccgctttcac aggaagaaaa tgaacggttc 60
atgactttca tatcacctga aaaacgggag aaatgccgga gattttatca taaagaagat 120

gtcaccgca ccctgctggg agatgtgctc gttcgctcag tcataagcag gcagtatcag 180
ttggacaaat ccgatatccg ctttagcacg caggaatacg ggaagccgtg catccctgat 240
cttcccgacg ctcatattcaa catttctcac tccggccgct ggggtcattgg tgcgtttgat 300
tcacagccga tcggcataga tatcgaaaaa acgaaaccga tcagccttga gatcgccaag 360
cgcttctttt caaaaacaga gtacagcgac cttttagcaa aagacaagga cgagcagaca 420
gactattttt atcatctatg gtcaatgaaa gaaagcttta tcaaacagga aggcaaaggc 480
ttatcgcttc cgcttgattc cttttcagtg cgcctgcatc aggacggaca agtatccatt 540

gagcttccgg acagccattc cccatgctat atcaaaacgt atgaggtcga tcccggctac 600
aaaatggctg tatgcgccgc acacctgtt tccccgagga tatcacaatg gtctcgtacg 660
aagagctttt ataa 674

<210> 12

<211> 224

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
polypeptide"

<400> 12

Met Lys Ile Tyr Gly Ile Tyr Met Asp Arg Pro Leu Ser Gln Glu Glu

1 5 10 15

Asn Glu Arg Phe Met Thr Phe Ile Ser Pro Glu Lys Arg Glu Lys Cys

20 25 30

Arg Arg Phe Tyr His Lys Glu Asp Ala His Arg Thr Leu Leu Gly Asp

35 40 45

Val Leu Val Arg Ser Val Ile Ser Arg Gln Tyr Gln Leu Asp Lys Ser
50 55 60

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

Leu Pro Asp Ala His Phe Asn Ile Ser His Ser Gly Arg Trp Val Ile
85 90 95

Gly Ala Phe Asp Ser Gln Pro Ile Gly Ile Asp Ile Glu Lys Thr Lys
100 105 110

Pro Ile Ser Leu Glu Ile Ala Lys Arg Phe Phe Ser Lys Thr Glu Tyr
115 120 125

Ser Asp Leu Leu Ala Lys Asp Lys Asp Glu Gln Thr Asp Tyr Phe Tyr
130 135 140

His Leu Trp Ser Met Lys Glu Ser Phe Ile Lys Gln Glu Gly Lys Gly
145 150 155 160

Leu Ser Leu Pro Leu Asp Ser Phe Ser Val Arg Leu His Gln Asp Gly
165 170 175

Gln Val Ser Ile Glu Leu Pro Asp Ser His Ser Pro Cys Tyr Ile Lys
180 185 190

Thr Tyr Glu Val Asp Pro Gly Tyr Lys Met Ala Val Cys Ala Ala His
195 200 205

Pro Asp Phe Pro Glu Asp Ile Thr Met Val Ser Tyr Glu Glu Leu Leu
210 215 220

<210> 13

<211> 52

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
primer"

<400> 13

ggcaatttga gaatttaagg aggaaaacaa aatgagccaa gaagacatct tc

52

<210> 14

<211> 33

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
primer"

<400> 14

cccaagcttc gaattcctag gcagctttgc tag 33

<210> 15

<211> 56

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
primer"

<400> 15

ggcaatttga gaatttaagg aggaaaacaa aatgaatcag gaaatttttg aaaaag 56

<210> 16

<211> 41

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
primer"

<400>

> 16

cccaagcttc gaattcttat ttactttcga tatgctcaac g 41

<210> 17

<211> 53

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
primer"

<400> 17
 ggcaatttga gaatttaagg aggaaaacaa aatgtcaca gaagaaatcc ttc 53
 <210> 18
 <211> 46
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 primer"
 <400> 18
 cccaagcttc gaattcttaa cctttttttt ctctgatgaa tttaac 46
 <210> 19
 <211> 53
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 primer"
 <400> 19
 ggcaatttga gaatttaagg aggaaaacaa aatgagccaa acggaacttt ttg 53
 <210> 20
 <211> 38
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 primer"
 <400> 20
 cccaagcttc gaattcttaa gctgatgcag caactttg 38
 <210> 21
 <211> 53
 <212> DNA
 <213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic primer"

<400> 21

ggcaatttga gaatttaagg aggaaaacaa aatgagccaa tcagaaactt ttg 53

<210> 22

<211> 34

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic primer"

<400> 22

cccaagcttc gaattcttaa gctgatgcgg caac 34

<210> 23

<211> 43

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic primer"

<400> 23

agctgcctag gaatttaagg aggaataaac catgaagatt tac 43

<210> 24

<211> 43

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic primer"

<400> 24

aaaagggttaa gaatttaagg aggaataaac catgaagatt tac 43

<210> 25

<211> 43

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
primer"

<400> 25

atcagcttaa gaatttaagg aggaataaac catgaagatt tac 43

<210> 26

<211> 41

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
primer"

<400> 26

cccaagcttc gaattcttat aaaagctctt cgtacgagac c 41

<210> 27

<211> 1029

<212> DNA

<213> Synechococcus elongatus

<400> 27

atggcattcg gtcttatcgg tcatctcacc agtttgagc aggcccgca cgtttctcgc 60

aggatgggct acgacgaata cgccgatcaa ggattggagt tttggagtag cgctcctcct 120

caaatcggtg atgaaatcac agtcaccagt gccacaggca aggtgattca cggtcgctac 180

atcgaatcgt gtttcttgcc ggaaatgctg gcggcgcgcc gcttcaaac agccacgcgc 240

aaagttctca atgccatgtc ccatgcccac aaacacggca tcgacatctc ggccttgggg 300

ggctttacct cgattatctt cgagaatttc gatttgccca gtttcggca agtgcgcgac 360

actaccttgg agtttgaacg gttcaccacc ggcaatactc acacggccta cgtaatctgt 420

agacaggtgg aagccgctgc taaaacgctg ggcatcgaca ttaccaagc gacagtagcg 480

gttgtcggcg cgactggcga tatcggtagc gctgtctgcc gctggctcga cctcaaaactg 540

ggtgtcgggtg atttgatcct gacggcgcgc aatcaggagc gtttgataa cctgcaggct 600

gaactcggcc ggggcaagat tctgcccttg gaagccgctc tgccggaagc tgactttatc 660

gtgtgggtcg ccagtatgcc tcagggcgta gtgatcgacc cagcaaccct gaagcaaccc 720
 tgcgtcctaa tcgacggggg ctaccccaaa aacttgggca gcaaagtcca aggtgagggc 780
 atctatgtcc tcaatggcgg ggtagttgaa cattgcttcg acatcgactg gcagatcatg 840
 tccgtctcag agatggcgcg gcccgagcgc cagatgtttg cctgctttgc cgaggcgatg 900

ctcttgaat ttgaaggctg gcatactaac ttctcctggg gccgcaacca aatcacgac 960
 gagaagatgg aagcgatcgg tgaggcatcg gtgcgccacg gcttccaacc ctggcattg 1020
 gcaatttga 1029

<210> 28

<211> 342

<212> PRT

<213> Synechococcus elongatus

<400> 28

Met Ala Phe Gly Leu Ile Gly His Leu Thr Ser Leu Glu Gln Ala Arg

1 5 10 15

Asp Val Ser Arg Arg Met Gly Tyr Asp Glu Tyr Ala Asp Gln Gly Leu

20 25 30

Glu Phe Trp Ser Ser Ala Pro Pro Gln Ile Val Asp Glu Ile Thr Val

35 40 45

Thr Ser Ala Thr Gly Lys Val Ile His Gly Arg Tyr Ile Glu Ser Cys

50 55 60

Phe Leu Pro Glu Met Leu Ala Ala Arg Arg Phe Lys Thr Ala Thr Arg

65 70 75 80

Lys Val Leu Asn Ala Met Ser His Ala Gln Lys His Gly Ile Asp Ile

85 90 95

Ser Ala Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asp Leu

100 105 110

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

115 120 125

Thr Thr Gly Asn Thr His Thr Ala Tyr Val Ile Cys Arg Gln Val Glu

130 135 140

Ala Ala Ala Lys Thr Leu Gly Ile Asp Ile Thr Gln Ala Thr Val Ala

145 150 155 160

Val Val Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu

 165 170 175

Asp Leu Lys Leu Gly Val Gly Asp Leu Ile Leu Thr Ala Arg Asn Gln

 180 185 190

Glu Arg Leu Asp Asn Leu Gln Ala Glu Leu Gly Arg Gly Lys Ile Leu

 195 200 205

Pro Leu Glu Ala Ala Leu Pro Glu Ala Asp Phe Ile Val Trp Val Ala

 210 215 220

Ser Met Pro Gln Gly Val Val Ile Asp Pro Ala Thr Leu Lys Gln Pro

225 230 235 240

Cys Val Leu Ile Asp Gly Gly Tyr Pro Lys Asn Leu Gly Ser Lys Val

 245 250 255

Gln Gly Glu Gly Ile Tyr Val Leu Asn Gly Gly Val Val Glu His Cys

 260 265 270

Phe Asp Ile Asp Trp Gln Ile Met Ser Ala Ala Glu Met Ala Arg Pro

 275 280 285

Glu Arg Gln Met Phe Ala Cys Phe Ala Glu Ala Met Leu Leu Glu Phe

 290 295 300

Glu Gly Trp His Thr Asn Phe Ser Trp Gly Arg Asn Gln Ile Thr Ile

305 310 315 320

Glu Lys Met Glu Ala Ile Gly Glu Ala Ser Val Arg His Gly Phe Gln

 325 330 335

Pro Leu Ala Leu Ala Ile

 340

<210> 29

<211> 1023

<212> DNA

<213> Synechocystis sp.

<400> 29

atgtttggtc ttattgggtca tctcacgagt ttagaacacg cccaagcggg tgctgaagat

60

ttaggtatc ctgagtacgc caaccaaggc ctggattttt ggtgttcggc tcctcccaa 120
 gtggttgata attttcaggt gaaaagtgtg acggggcagg tgattgaagg caaatatgtg 180
 gagtcttgct ttttgccgga aatgttaacc caacggcgga tcaaagcggc cattcgtaaa 240
 atcctcaatg ctatggccct ggcccaaaag gtgggcttgg atattacggc cctgggaggc 300
 ttttcttcaa tcgtatttga agaatttaac ctcaagcaaa ataatcaagt ccgcaatgtg 360
 gaactagatt ttcagcggtt caccactggt aataccaca ccgcttatgt gatctgccgt 420

caggtcgagt ctggagctaa acagttgggt attgatctaa gtcaggcaac ggtagcggtt 480
 tgtggcgcca cgggagatat tggtagcgcc gtatgtcgtt ggtagatag caaacatcaa 540
 gtaaggaat tattgctaatt tgcccgtaac cgccaaagat tggaaaatct ccaagaggaa 600
 ttgggtcggg gcaaaattat ggatttggaa acagccctgc ccaggcaga tattattgtt 660
 tgggtggcta gtatgcccga gggggtagaa attgcggggg aaatgctgaa aaagccctgt 720
 ttgattgtgg atgggggcta tccaagaat ttagacacca gggtgaaagc ggatggggtg 780
 catattctca agggggggat tgtagaacat tccttgata ttacctgga aattatgaag 840

attgtggaga tggatattcc ctcccgcaa atgttcgcct gttttgcgga ggccattttg 900
 ctagagtttg agggctggcg cactaatttt tcttggggcc gcaaccaaatt tccgttaat 960
 aaaatggagg cgatttgtga agcttctgtc aagcatggct tttgcccttt agtagctctt 1020
 tag 1023

<210> 30

<211> 340

<212> PRT

<213> Synechocystis sp.

<400> 30

Met Phe Gly Leu Ile Gly His Leu Thr Ser Leu Glu His Ala Gln Ala

1 5 10 15

Val Ala Glu Asp Leu Gly Tyr Pro Glu Tyr Ala Asn Gln Gly Leu Asp

20 25 30

Phe Trp Cys Ser Ala Pro Pro Gln Val Val Asp Asn Phe Gln Val Lys

35 40 45

Ser Val Thr Gly Gln Val Ile Glu Gly Lys Tyr Val Glu Ser Cys Phe

50 55 60

Leu Pro Glu Met Leu Thr Gln Arg Arg Ile Lys Ala Ala Ile Arg Lys

65	70	75	80
Ile Leu Asn Ala Met Ala Leu Ala Gln Lys Val Gly Leu Asp Ile Thr			
	85	90	95
Ala Leu Gly Gly Phe Ser Ser Ile Val Phe Glu Glu Phe Asn Leu Lys			
	100	105	110
Gln Asn Asn Gln Val Arg Asn Val Glu Leu Asp Phe Gln Arg Phe Thr			
	115	120	125
Thr Gly Asn Thr His Thr Ala Tyr Val Ile Cys Arg Gln Val Glu Ser			
	130	135	140
Gly Ala Lys Gln Leu Gly Ile Asp Leu Ser Gln Ala Thr Val Ala Val			
	145	150	155
Cys Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Asp			
	165	170	175
Ser Lys His Gln Val Lys Glu Leu Leu Leu Ile Ala Arg Asn Arg Gln			
	180	185	190
Arg Leu Glu Asn Leu Gln Glu Glu Leu Gly Arg Gly Lys Ile Met Asp			
	195	200	205
Leu Glu Thr Ala Leu Pro Gln Ala Asp Ile Ile Val Trp Val Ala Ser			
	210	215	220
Met Pro Lys Gly Val Glu Ile Ala Gly Glu Met Leu Lys Lys Pro Cys			
	225	230	235
Leu Ile Val Asp Gly Gly Tyr Pro Lys Asn Leu Asp Thr Arg Val Lys			
	245	250	255
Ala Asp Gly Val His Ile Leu Lys Gly Gly Ile Val Glu His Ser Leu			
	260	265	270
Asp Ile Thr Trp Glu Ile Met Lys Ile Val Glu Met Asp Ile Pro Ser			
	275	280	285
Arg Gln Met Phe Ala Cys Phe Ala Glu Ala Ile Leu Leu Glu Phe Glu			
	290	295	300
Gly Trp Arg Thr Asn Phe Ser Trp Gly Arg Asn Gln Ile Ser Val Asn			
	305	310	315
			320

Lys Met Glu Ala Ile Gly Glu Ala Ser Val Lys His Gly Phe Cys Pro

325

330

335

Leu Val Ala Leu

340

<210> 31

<211> 1023

<212> DNA

<213> Cyanothec sp.

<400> 31

atgtttggtt taattgggtca tcttacaagt ttagaacacg cccactccgt tgctgatgcc	60
tttggctatg gcccatacgc cactcaggga cttgatttgt ggtgttctgc tccaccccaa	120
ttcgtcgagc attttcatgt tactagcatc acaggacaaa ccatcgaagg aaagtatata	180
gaatccgctt tcttaccaga aatgctgata aagcgacgga ttaaagcagc aattcgcaaa	240
atactgaatg cgatggcctt tgctcagaaa aataacctta acatcacagc attagggggc	300
ttttcttcga ttatttttga agaatttaat ctcaaagaga atagacaagt tcgtaatgtc	360

tcttttagagt ttgatcgctt caccaccgga aacaccata ctgcttatat catttgctgt	420
caagttgaac aggcacccgc taaactaggg attgacttat cccaagcaac ggttgctatt	480
tgccggggcaa ccggagatat tggcagtgca gtgtgtcgtt ggtagatag aaaaaccgat	540
accaggaac tattcttaat tgctcgcaat aaagaacgat tacaacgact gcaagatgag	600
ttgggacggg gtaaaattat gggattggag gaggctttac ccgaagcaga tattatcgtt	660
tgggtggcga gtatgcccaa aggagtggaa attaatgccg aaactctcaa aaaaccctgt	720
ttaattatcg atggtggtta tcctaagaat ttagacacaa aaattaaaca tcctgatgtc	780

catatcctga aagggggaat tgtagaacaat tctctagata ttgactggaa gattatggaa	840
actgtcaata tggatgttcc ttctcgtaaa atgtttgctt gttttgccga agccatttta	900
ttagagtttg aacaatggca cactaatttt tcttggggac gcaatcaaat tacagtgact	960
aaaatggaac aaataggaga agcttctgtc aaacatgggt tacaaccgtt gttgagttgg	1020
taa	1023

<210> 32

<211> 340

<212> PRT

<213> Cyanothec sp.

<400> 32

Met Phe Gly Leu Ile Gly His Leu Thr Ser Leu Glu His Ala His Ser

1 5 10 15
Val Ala Asp Ala Phe Gly Tyr Gly Pro Tyr Ala Thr Gln Gly Leu Asp
 20 25 30
Leu Trp Cys Ser Ala Pro Pro Gln Phe Val Glu His Phe His Val Thr
 35 40 45
Ser Ile Thr Gly Gln Thr Ile Glu Gly Lys Tyr Ile Glu Ser Ala Phe
 50 55 60
Leu Pro Glu Met Leu Ile Lys Arg Arg Ile Lys Ala Ala Ile Arg Lys

65 70 75 80
Ile Leu Asn Ala Met Ala Phe Ala Gln Lys Asn Asn Leu Asn Ile Thr
 85 90 95
Ala Leu Gly Gly Phe Ser Ser Ile Ile Phe Glu Glu Phe Asn Leu Lys
 100 105 110
Glu Asn Arg Gln Val Arg Asn Val Ser Leu Glu Phe Asp Arg Phe Thr
 115 120 125
Thr Gly Asn Thr His Thr Ala Tyr Ile Ile Cys Arg Gln Val Glu Gln

130 135 140
Ala Ser Ala Lys Leu Gly Ile Asp Leu Ser Gln Ala Thr Val Ala Ile
145 150 155 160
Cys Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Asp
 165 170 175
Arg Lys Thr Asp Thr Gln Glu Leu Phe Leu Ile Ala Arg Asn Lys Glu
 180 185 190
Arg Leu Gln Arg Leu Gln Asp Glu Leu Gly Arg Gly Lys Ile Met Gly

195 200 205
Leu Glu Glu Ala Leu Pro Glu Ala Asp Ile Ile Val Trp Val Ala Ser
 210 215 220
Met Pro Lys Gly Val Glu Ile Asn Ala Glu Thr Leu Lys Lys Pro Cys
225 230 235 240
Leu Ile Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Thr Lys Ile Lys

245 250 255
His Pro Asp Val His Ile Leu Lys Gly Gly Ile Val Glu His Ser Leu

260 265 270
Asp Ile Asp Trp Lys Ile Met Glu Thr Val Asn Met Asp Val Pro Ser

275 280 285
Arg Gln Met Phe Ala Cys Phe Ala Glu Ala Ile Leu Leu Glu Phe Glu

290 295 300
Gln Trp His Thr Asn Phe Ser Trp Gly Arg Asn Gln Ile Thr Val Thr

305 310 315 320
Lys Met Glu Gln Ile Gly Glu Ala Ser Val Lys His Gly Leu Gln Pro

325 330 335
Leu Leu Ser Trp

340

<210> 33

<211> 1044

<212> DNA

<213> *Prochlorococcus marinus*

<400> 33

atggcatttg ggcttatagg tcattcaact agttttgaag atgcaaaaag aaaggcttca 60
ttattgggct ttgatcatat tgcggatggt gatttagatg tttgggtgcac agctccacct 120
caactagttag aaaaatgtaga gggttaaaagt gctataggta tatcaattga aggttcttat 180
attgattcat gtttcgttcc tgaaatgctt tcaagattta aaacggcaag aagaaaagta 240
ttaaatgcaa tggaattagc tcaaaaaaaaa ggtattaata ttaccgcttt ggggggggttc 300

acttctatca tctttgaaaa ttttaatctc ctccaacata agcagattag aaacacttca 360
ctagagtggg aaaggtttac aactggtaat actcactctg cgtgggttat ttgcaggcaa 420
ttagagatga atgctcctaa aataggtatt gatcttaaaa gcgcaacagt tgctgtagtt 480
gggtgctactg gagatatagg cagtgcgtgt tgcgatggt taatcaataa aacaggtatt 540
ggggaacttc ttttggtagc taggcaaaaag gaacccttgg attctttgca aaaggaatta 600
gatggtggaa ctatcaaaaa tctagatgaa gcattgcctg aagcagatat tgttgtatgg 660
gtagcaagta tgccaaagac aatggaaatc gatgctaata atcttaaaaca accatgttta 720

atgattgatg gaggttatcc aaagaatcta gatgaaaaat ttcaaggaaa taatatatcat 780

gttgtaaaag gaggtatagt aagattcttc aatgatatag gttggaatat gatggaacta 840
gctgaaatgc aaaatcccca gagagaaatg ttgcatgct ttgcagaagc aatgatttta 900
gaatttgaaa aatgtcatac aaacttttagc tggggaagaa ataatatatc tctcgagaaa 960
atggagttaa ttggagctgc ttctgtaaag catggcttct ctgcaattgg cctagataag 1020
catccaaaag tactagcagt ttga 1044

<210> 34

<211> 347

<212> PRT

<213> *Prochlorococcus marinus*

<400> 34

Met Ala Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Asp Ala Lys

1 5 10 15

Arg Lys Ala Ser Leu Leu Gly Phe Asp His Ile Ala Asp Gly Asp Leu

20 25 30

Asp Val Trp Cys Thr Ala Pro Pro Gln Leu Val Glu Asn Val Glu Val

35 40 45

Lys Ser Ala Ile Gly Ile Ser Ile Glu Gly Ser Tyr Ile Asp Ser Cys

50 55 60

Phe Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val

65 70 75 80

Leu Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala

85 90 95

Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln

100 105 110

His Lys Gln Ile Arg Asn Thr Ser Leu Glu Trp Glu Arg Phe Thr Thr

115 120 125

Gly Asn Thr His Thr Ala Trp Val Ile Cys Arg Gln Leu Glu Met Asn

130 135 140

Ala Pro Lys Ile Gly Ile Asp Leu Lys Ser Ala Thr Val Ala Val Val

145 150 155 160

Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ile Asn

165 170 175

Lys Thr Gly Ile Gly Glu Leu Leu Leu Val Ala Arg Gln Lys Glu Pro
180 185 190

Leu Asp Ser Leu Gln Lys Glu Leu Asp Gly Gly Thr Ile Lys Asn Leu
195 200 205

Asp Glu Ala Leu Pro Glu Ala Asp Ile Val Val Trp Val Ala Ser Met
210 215 220

Pro Lys Thr Met Glu Ile Asp Ala Asn Asn Leu Lys Gln Pro Cys Leu
225 230 235 240

Met Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Glu Lys Phe Gln Gly
245 250 255

Asn Asn Ile His Val Val Lys Gly Gly Ile Val Arg Phe Phe Asn Asp
260 265 270

Ile Gly Trp Asn Met Met Glu Leu Ala Glu Met Gln Asn Pro Gln Arg
275 280 285

Glu Met Phe Ala Cys Phe Ala Glu Ala Met Ile Leu Glu Phe Glu Lys
290 295 300

Cys His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Ser Leu Glu Lys
305 310 315 320

Met Glu Phe Ile Gly Ala Ala Ser Val Lys His Gly Phe Ser Ala Ile
325 330 335

Gly Leu Asp Lys His Pro Lys Val Leu Ala Val
340 345

<210> 35

<211> 1053

<212> DNA

<213> *Gloeobacter violaceus*

<400> 35

atgtttggcc tgatcggaca cttgaccaat ctttcccatg cccagcgggt cgcccgcgac 60

ctgggctacg acgagtatgc aagccacgac ctgcaattct ggtgcatggc ccctccccag 120

gcggtcgatg aaatcacgat caccagcgtc accggtcagg tgatccacgg tcagtacgtc 180

gaatcgtgct ttctgccgga gatgctcgcc cagggccgct tcaagaccgc catgcgcaag 240

atcctcaatg ccatggccct ggtccagaag cgcggcatcg acattacggc cctgggaggc 300
 ttctctcgca tcatcttcga gaatttcagc ctcgataaat tgctcaacgt ccgcgacatc 360
 accctcgaca tccagcgctt caccaccggc aacaccacaca cggcctacat cctttgtcag 420
 caggtcgagc aggggtcggg acgctacggc atcgatccgg ccaaagcgac cgtggcggta 480
 gtcggggcca ccggcgacat cggtagcgcc gtctgccgat ggctcaccga ccgcgccggc 540
 atccacgaac tcttgctggt ggcccgcgac gccgaaagge tgcaccggt gcagcaggaa 600

ctcggcaccg gtcggatcct gccggtcgaa gaagcacttc ccaaagccga catcgtcgtc 660
 tgggtcgctt cgatgaacca gggcatggcc atcgaccccg ccggcctgcg caccctctgc 720
 ctgctcatcg acggcggcta cccaagaac atggccggca ccctgcagcg cccgggcatc 780
 catatcctcg acggcggcat ggtcgagcac tcgctcgaca tcgactggca gatcatgtcg 840
 tttctaaatg tgcccaaccc cgcccgcag tttctgcct gtttcgccga gtcgatgtcg 900
 ctggaattcg aagggttca cttcaatttt tcttggggcc gcaaccacat caccgtcgag 960
 aagatggccc agatcggtc gctgtctaaa aaacatggct ttcgtccct gcttgaaccc 1020

agtcagcgca gcggcgaact cgtacacgga taa 1053

<210> 36

<211> 350

<212> PRT

<213> *Gloeobacter violaceus*

<400> 36

Met Phe Gly Leu Ile Gly His Leu Thr Asn Leu Ser His Ala Gln Arg

1 5 10 15

Val Ala Arg Asp Leu Gly Tyr Asp Glu Tyr Ala Ser His Asp Leu Glu

20 25 30

Phe Trp Cys Met Ala Pro Pro Gln Ala Val Asp Glu Ile Thr Ile Thr

35 40 45

Ser Val Thr Gly Gln Val Ile His Gly Gln Tyr Val Glu Ser Cys Phe

50 55 60

Leu Pro Glu Met Leu Ala Gln Gly Arg Phe Lys Thr Ala Met Arg Lys

65 70 75 80

Ile Leu Asn Ala Met Ala Leu Val Gln Lys Arg Gly Ile Asp Ile Thr

85 90 95

Ala Leu Gly Gly Phe Ser Ser Ile Ile Phe Glu Asn Phe Ser Leu Asp

100	105	110	
Lys Leu Leu Asn Val Arg Asp Ile Thr Leu Asp Ile Gln Arg Phe Thr			
115	120	125	
Thr Gly Asn Thr His Thr Ala Tyr Ile Leu Cys Gln Gln Val Glu Gln			
130	135	140	
Gly Ala Val Arg Tyr Gly Ile Asp Pro Ala Lys Ala Thr Val Ala Val			
145	150	155	160
Val Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Thr			
165	170	175	
Asp Arg Ala Gly Ile His Glu Leu Leu Leu Val Ala Arg Asp Ala Glu			
180	185	190	
Arg Leu Asp Arg Leu Gln Gln Glu Leu Gly Thr Gly Arg Ile Leu Pro			
195	200	205	
Val Glu Glu Ala Leu Pro Lys Ala Asp Ile Val Val Trp Val Ala Ser			
210	215	220	
Met Asn Gln Gly Met Ala Ile Asp Pro Ala Gly Leu Arg Thr Pro Cys			
225	230	235	240
Leu Leu Ile Asp Gly Gly Tyr Pro Lys Asn Met Ala Gly Thr Leu Gln			
245	250	255	
Arg Pro Gly Ile His Ile Leu Asp Gly Gly Met Val Glu His Ser Leu			
260	265	270	
Asp Ile Asp Trp Gln Ile Met Ser Phe Leu Asn Val Pro Asn Pro Ala			
275	280	285	
Arg Gln Phe Phe Ala Cys Phe Ala Glu Ser Met Leu Leu Glu Phe Glu			
290	295	300	
Gly Leu His Phe Asn Phe Ser Trp Gly Arg Asn His Ile Thr Val Glu			
305	310	315	320
Lys Met Ala Gln Ile Gly Ser Leu Ser Lys Lys His Gly Phe Arg Pro			
325	330	335	
Leu Leu Glu Pro Ser Gln Arg Ser Gly Glu Leu Val His Gly			
340	345	350	

<210> 37

<211> 1020

<212> DNA

<213> Nostoc punctiforme

<400> 37

```

atgtttggtc taattggaca tctgactagt ttagaacacg ctcaagccgt agcccaagaa      60

ttgggatacc cagaatatgc cgatcaaggg ctagactttt ggtgcagcgc cccgccgcaa      120
attgtcgata gtattattgt caccagtgtt actgggcaac aaattgaagg acgatatgta      180
gaatcttgct ttttgccgga aatgctagct agtcgccgca tcaaagccgc aacacggaaa      240
atcctcaacg ctatggccca tgcacagaag caccgcatta acatcacagc tttaggcgga      300
ttttcctcga ttatttttga aaactttaag ttagagcagt ttagccaagt ccgaaatatic      360
aagctagagt ttgaacgctt caccacagga aacacgcata ctgcctacat tttttgtaag      420
caggtggaag aagcatccaa acaactggga attaattctat caaacgcgac tgttgcggtta      480

tgtggagcaa ctggggatat tggtagtgcc gttacacgct ggctagatgc gagaacagat      540
gtccaagaac tcctgcta atcgcccgcat caagaacgct tcaaagagtt gcaaggcgaa      600
ctggggcggg ggaaaatcat gggtttgaca gaagcactac cccaagccga tgtttagttt      660
tgggttgcta gtatgccag aggcgtagaa attgacccca ccactttgaa acaaccctgt      720
ttgttgattg atggtggcta tcctaaaaac ttagcaacaa aaattcaata tcctggcgta      780
cacgtgttaa atggtgggat ttagagcat tccctggata ttgactggaa aattatgaaa      840
atagtcaata tggacgtgcc agcccgtagc ttgtttgcct gttttgccga atcaatgcta      900

ctggaatttg agaagttata cacgaacttt tcgtggggac ggaatcagat taccgtagat      960
aaaatggagc agattggccg ggtgtcagta aaacatggat ttagaccgtt gttggttttag      1020

```

<210> 38

<211> 339

<212> PRT

<213> Nostoc punctiforme

<400> 38

```

Met Phe Gly Leu Ile Gly His Leu Thr Ser Leu Glu His Ala Gln Ala
1           5           10           15

Val Ala Gln Glu Leu Gly Tyr Pro Glu Tyr Ala Asp Gln Gly Leu Asp
          20           25           30

```

Phe Trp Cys Ser Ala Pro Pro Gln Ile Val Asp Ser Ile Ile Val Thr

35 40 45

Ser Val Thr Gly Gln Gln Ile Glu Gly Arg Tyr Val Glu Ser Cys Phe

50 55 60

Leu Pro Glu Met Leu Ala Ser Arg Arg Ile Lys Ala Ala Thr Arg Lys

65 70 75 80

Ile Leu Asn Ala Met Ala His Ala Gln Lys His Gly Ile Asn Ile Thr

85 90 95

Ala Leu Gly Gly Phe Ser Ser Ile Ile Phe Glu Asn Phe Lys Leu Glu

100 105 110

Gln Phe Ser Gln Val Arg Asn Ile Lys Leu Glu Phe Glu Arg Phe Thr

115 120 125

Thr Gly Asn Thr His Thr Ala Tyr Ile Ile Cys Lys Gln Val Glu Glu

130 135 140

Ala Ser Lys Gln Leu Gly Ile Asn Leu Ser Asn Ala Thr Val Ala Val

145 150 155 160

Cys Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Thr Arg Trp Leu Asp

165 170 175

Ala Arg Thr Asp Val Gln Glu Leu Leu Leu Ile Ala Arg Asp Gln Glu

180 185 190

Arg Leu Lys Glu Leu Gln Gly Glu Leu Gly Arg Gly Lys Ile Met Gly

195 200 205

Leu Thr Glu Ala Leu Pro Gln Ala Asp Val Val Val Trp Val Ala Ser

210 215 220

Met Pro Arg Gly Val Glu Ile Asp Pro Thr Thr Leu Lys Gln Pro Cys

225 230 235 240

Leu Leu Ile Asp Gly Gly Tyr Pro Lys Asn Leu Ala Thr Lys Ile Gln

245 250 255

Tyr Pro Gly Val His Val Leu Asn Gly Gly Ile Val Glu His Ser Leu

260 265 270

Asp Ile Asp Trp Lys Ile Met Lys Ile Val Asn Met Asp Val Pro Ala

275 280 285

Arg Gln Leu Phe Ala Cys Phe Ala Glu Ser Met Leu Leu Glu Phe Glu

[illegible]

<210> 39

<211> 1020

<212> DNA

<213> *Anabaena variabilis*

<400> 39

atgtttggtc taattggaca tctgacaagt ttagaacacg ctcaagecgt agctcaagaa 60

ctgggatacc cagaatacgc cgaccaaggg ctagatTTTT ggtgcagcgc tccaccgcaa 120

atagttgacc acattaaagt tactagcatt actggtgaaa taattgaagg gaggtatgta 180

gaatcttgct ttttaccaga aatgctagcc agccgtagga tttaaagccgc aacccgcaaa 240

gtcctcaatg ctatggctca tgctcaaaaa catggcattg acatcacgcg tttgggtggt 300

ttctcctcca ttatttttga aaacttcaaa ttggaacagt ttagccaagt tcgtaatgtc 360

acactagagt ttgaacgctt cactacaggc aacactcaca cagcttatat catttgcgg 420

caggtagaac aagcatcaca acaactcggc attgaactct cccaagcaac agtagctata 480

tgtggggcta ctggtgacat tggtagtgca gttactcgct ggctggatgc caaaacagac 540

gtaaaagaat tactgttaat cgcccgaat caagaacgtc tccaagagtt gcaaagcgag 600

ttgggacgcg gtaaaatcat gaggctagat gaagcattgc ctcaagctga tattgtagtt 660

tggttagcta gtatgcctaa aggcgtggaa attaatcctc aagttttgaa acaaccctgt 720

ttattgattg atggtggta tccgaaaaac ttgggtacaa aagttcagta tcctgggtgtt 780

tatgtactga acggaggtat cgtcgaacat tccctagata ttgactggaa aatcatgaaa 840

atagtcaata tggatgtacc tgcacgcaa ttatttgctt gttttgcgga atctatgctc 900

ttggaatttg agaagttgta cacgaacttt tcttgggggc gcaatcagat taccgtagac 960

aaaatggagc agattggtca agcatcagtg aaacatgggt ttagaccact gctggtttag 1020

<210> 40

<211> 339

<212> PRT

<213> *Anabaena variabilis*

<400> 40

Met Phe Gly Leu Ile Gly His Leu Thr Ser Leu Glu His Ala Gln Ala

1 5 10 15

Val Ala Gln Glu Leu Gly Tyr Pro Glu Tyr Ala Asp Gln Gly Leu Asp

20 25 30

Phe Trp Cys Ser Ala Pro Pro Gln Ile Val Asp His Ile Lys Val Thr

35 40 45

Ser Ile Thr Gly Glu Ile Ile Glu Gly Arg Tyr Val Glu Ser Cys Phe

50 55 60

Leu Pro Glu Met Leu Ala Ser Arg Arg Ile Lys Ala Ala Thr Arg Lys

65 70 75 80

Val Leu Asn Ala Met Ala His Ala Gln Lys His Gly Ile Asp Ile Thr

85 90 95

Ala Leu Gly Gly Phe Ser Ser Ile Ile Phe Glu Asn Phe Lys Leu Glu

100 105 110

Gln Phe Ser Gln Val Arg Asn Val Thr Leu Glu Phe Glu Arg Phe Thr

115 120 125

Thr Gly Asn Thr His Thr Ala Tyr Ile Ile Cys Arg Gln Val Glu Gln

130 135 140

Ala Ser Gln Gln Leu Gly Ile Glu Leu Ser Gln Ala Thr Val Ala Ile

145 150 155 160

Cys Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Thr Arg Trp Leu Asp

165 170 175

Ala Lys Thr Asp Val Lys Glu Leu Leu Leu Ile Ala Arg Asn Gln Glu

180 185 190

Arg Leu Gln Glu Leu Gln Ser Glu Leu Gly Arg Gly Lys Ile Met Ser

195 200 205

Leu Asp Glu Ala Leu Pro Gln Ala Asp Ile Val Val Trp Val Ala Ser

210 215 220
Met Pro Lys Gly Val Glu Ile Asn Pro Gln Val Leu Lys Gln Pro Cys

225 230 235 240
Leu Leu Ile Asp Gly Gly Tyr Pro Lys Asn Leu Gly Thr Lys Val Gln

245 250 255
Tyr Pro Gly Val Tyr Val Leu Asn Gly Gly Ile Val Glu His Ser Leu

260 265 270
Asp Ile Asp Trp Lys Ile Met Lys Ile Val Asn Met Asp Val Pro Ala

275 280 285
Arg Gln Leu Phe Ala Cys Phe Ala Glu Ser Met Leu Leu Glu Phe Glu

290 295 300
Lys Leu Tyr Thr Asn Phe Ser Trp Gly Arg Asn Gln Ile Thr Val Asp

305 310 315 320
Lys Met Glu Gln Ile Gly Gln Ala Ser Val Lys His Gly Phe Arg Pro

325 330 335
Leu Leu Val

<210> 41

<211> 1026

<212> DNA

<213> Synechococcus elongatus

<400> 41

atgttcggtc ttatcgggtca tctcaccagt ttggagcagg cccgcgacgt ttctcgcagg 60

atgggctacg acgaatacgc cgatcaagga ttggagtttt ggagtagcgc tcctcctcaa 120

atcgttgatg aaatcacagt caccagtgcc acaggcaagg tgattcacgg tcgctacatc 180

gaatcgtgtt tcttgccgga aatgctggcg gcgcgccgct tcaaacagc cacgcgcaaa 240

gttctcaatg ccatgtccca tgcccaaaaa caggcatcg acatctcggc ctggtgggggc 300

tttacctcga ttattttcga gaatttcgat ttggccagtt tgcggcaagt gcgcgacact 360

accttggagt ttgaacggtt caccaccggc aatactcaca cggcctacgt aatctgtaga 420

caggtggaag ccgctgctaa aacgctgggc atcgacatta cccaagcgac agtagcggtt 480

gtcggcgcca ctggcgatat cggtagcgct gtctgccgct ggctcgacct caaactgggt 540

gtcgggtgatt tgatcctgac ggcgcgcaat caggagcggtt tggataacct gcaggctgaa 600
 ctccggccggg gcaagattct gcccttggaa gccgctctgc cggaagctga ctttatcgtg 660
 tgggtcgcca gtatgcctca gggcgtagtg atcgaccag caaccctgaa gcaaccctgc 720
 gtccaatcg acgggggcta ccccaaaaac ttgggcagca aagtccaagg tgagggcac 780
 tatgtcctca atggcggggt agttgaacat tgcttcgaca tcgactggca gatcatgtcc 840
 gctgcagaga tggcgcggcc cgagcgccag atgtttgcct gctttgccga ggcgatgctc 900

ttggaatttg aaggctggca tactaacttc tcctggggcc gcaaccaaat cacgatcgag 960
 aagatggaag cgatcggtag ggcacgggtg cgccacggct tccaaccctt ggcattggca 1020
 atttga 1026

<210> 42

<211> 340

<212> PRT

<213> Synechococcus elongatus

<400> 42

Met Phe Gly Leu Ile Gly His Leu Thr Ser Leu Glu Gln Ala Arg Asp

1 5 10 15

Val Ser Arg Arg Met Gly Tyr Asp Glu Tyr Ala Asp Gln Gly Leu Glu

20 25 30

Phe Trp Ser Ser Ala Pro Pro Gln Ile Val Asp Glu Ile Thr Val Thr

35 40 45

Ser Ala Thr Gly Lys Val Ile His Gly Arg Tyr Ile Glu Ser Cys Phe

50 55 60

Leu Pro Glu Met Leu Ala Ala Arg Arg Phe Lys Thr Ala Thr Arg Lys

65 70 75 80

Val Leu Asn Ala Met Ser His Ala Gln Lys His Gly Ile Asp Ile Ser

85 90 95

Ala Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asp Leu Ala

100 105 110

Ser Leu Arg Gln Val Arg Asp Thr Thr Leu Glu Phe Glu Arg Phe Thr

115 120 125

Thr Gly Asn Thr His Thr Ala Tyr Val Ile Cys Arg Gln Val Glu Ala

130 135 140

Ala Ala Lys Thr Leu Gly Ile Asp Ile Thr Gln Ala Thr Val Ala Val
145 150 155 160

Val Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Asp
165 170 175

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

Arg Leu Asp Asn Leu Gln Ala Glu Leu Gly Arg Gly Lys Ile Leu Pro
195 200 205

Leu Glu Ala Ala Leu Pro Glu Ala Asp Phe Ile Val Trp Val Ala Ser
210 215 220

Met Pro Gln Gly Val Val Ile Asp Pro Ala Thr Leu Lys Gln Pro Cys
225 230 235 240

Val Leu Ile Asp Gly Gly Tyr Pro Lys Asn Leu Gly Ser Lys Val Gln
245 250 255

Gly Glu Gly Ile Tyr Val Leu Asn Gly Gly Val Val Glu His Cys Phe
260 265 270

Asp Ile Asp Trp Gln Ile Met Ser Ala Ala Glu Met Ala Arg Pro Glu
275 280 285

Arg Gln Met Phe Ala Cys Phe Ala Glu Ala Met Leu Leu Glu Phe Glu
290 295 300

Gly Trp His Thr Asn Phe Ser Trp Gly Arg Asn Gln Ile Thr Ile Glu
305 310 315 320

Lys Met Glu Ala Ile Gly Glu Ala Ser Val Arg His Gly Phe Gln Pro
325 330 335

Leu Ala Leu Ala
340

<210> 43

<211> 1020

<212> DNA

<213> Nostoc sp.

<400> 43

atgtttggtc taattggaca tctgacaagt ttagaacacg ctcaagcggg agctcaagaa 60

ctgggatacc cagaatacgc cgaccaaggg ctagattttt ggtgtagcgc tccaccgcaa 120

atagttgacc acattaaagt tactagtatt actggtgaaa taattgaagg gaggtatgta 180

gaatcttgct ttttaccgga gatgctagcc agtcgtcgga ttaaagccgc aaccgcgcaa 240

gtcctcaatg ctatggctca tgcctaaaag aatggcattg atatcacagc ttgggtgggt 300

ttctcctcca ttatttttga aaactttaaa ttggagcagt ttagccaagt tcgtaatgtg 360

acactagagt ttgaacgctt cactacaggc aacactcaca cagcatatat ttttgtcgg 420

caggtagaac aagcatcaca acaactcggc attgaactct cccaagcaac agtagctata 480

tgtggggcta ctggtgatat tggtagtgca gttactcgct ggctggatgc taaaacagac 540

gtgaaagaat tgctgttaat cgcccgtaat caagaacgct tccaagagtt gcaaagcgag 600

ctgggacgcg gtaaaatcat gagccttgat gaagcactgc cccaagctga tatcgtagtt 660

tgggtagcca gtatgcctaa aggtgtggaa attaatcctc aagttttgaa gcaaccctgt 720

ttgctgattg atgggggtta tccgaaaaac ttgggtacaa aagttcagta tcctggtgtt 780

tatgtactga acggcgggtat cgtcgaacat tcgctggata ttgactggaa aatcatgaaa 840

atagtcaata tggatgtacc tgcacgcaa ttatttgctt gttttgcgga atctatgctc 900

ttggaatttg agaagttgta cacgaacttt tcttgggggc gcaatcagat taccgtagac 960

aaaatggagc agatttgtca agcatcagtg aaacatgggt ttagaccact gctggtttag 1020

<210> 44

<211> 339

<212> PRT

<213> Nostoc sp.

<400> 44

Met Phe Gly Leu Ile Gly His Leu Thr Ser Leu Glu His Ala Gln Ala

1 5 10 15

Val Ala Gln Glu Leu Gly Tyr Pro Glu Tyr Ala Asp Gln Gly Leu Asp

20 25 30

Phe Trp Cys Ser Ala Pro Pro Gln Ile Val Asp His Ile Lys Val Thr

35 40 45

Ser Ile Thr Gly Glu Ile Ile Glu Gly Arg Tyr Val Glu Ser Cys Phe

50 55 60

Leu Pro Glu Met Leu Ala Ser Arg Arg Ile Lys Ala Ala Thr Arg Lys

65 70 75 80
 Val Leu Asn Ala Met Ala His Ala Gln Lys Asn Gly Ile Asp Ile Thr
 85 90 95
 Ala Leu Gly Gly Phe Ser Ser Ile Ile Phe Glu Asn Phe Lys Leu Glu

 100 105 110
 Gln Phe Ser Gln Val Arg Asn Val Thr Leu Glu Phe Glu Arg Phe Thr
 115 120 125
 Thr Gly Asn Thr His Thr Ala Tyr Ile Ile Cys Arg Gln Val Glu Gln
 130 135 140
 Ala Ser Gln Gln Leu Gly Ile Glu Leu Ser Gln Ala Thr Val Ala Ile
 145 150 155 160
 Cys Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Thr Arg Trp Leu Asp

 165 170 175
 Ala Lys Thr Asp Val Lys Glu Leu Leu Leu Ile Ala Arg Asn Gln Glu
 180 185 190
 Arg Leu Gln Glu Leu Gln Ser Glu Leu Gly Arg Gly Lys Ile Met Ser
 195 200 205
 Leu Asp Glu Ala Leu Pro Gln Ala Asp Ile Val Val Trp Val Ala Ser
 210 215 220
 Met Pro Lys Gly Val Glu Ile Asn Pro Gln Val Leu Lys Gln Pro Cys

 225 230 235 240
 Leu Leu Ile Asp Gly Gly Tyr Pro Lys Asn Leu Gly Thr Lys Val Gln
 245 250 255
 Tyr Pro Gly Val Tyr Val Leu Asn Gly Gly Ile Val Glu His Ser Leu
 260 265 270
 Asp Ile Asp Trp Lys Ile Met Lys Ile Val Asn Met Asp Val Pro Ala
 275 280 285
 Arg Gln Leu Phe Ala Cys Phe Ala Glu Ser Met Leu Leu Glu Phe Glu

 290 295 300
 Lys Leu Tyr Thr Asn Phe Ser Trp Gly Arg Asn Gln Ile Thr Val Asp
 305 310 315 320

Lys Met Glu Gln Ile Gly Gln Ala Ser Val Lys His Gly Phe Arg Pro

325

330

335

Leu Leu Val

<210> 45

<211> 867

<212> DNA

<213> *Corynebacterium glutamicum*

<400> 45

ttgggcgtgt cgcccttaaa gcgcgctttt cgacgcgacc ccactacatt ggcttccatg 60

aacgttgaca ttccacgata cagagagccg ctaaactgtg agctcctgaa ggaaaaattg 120

ctccaaaacg gtgactttgg ccaggtcatt tacgaaaaag tgacaggctc cactaatgct 180

gacttgctgg cacttgcagg ttctggcgct ccaaactgga cggtgaaaac tgcgagttt 240

caagatcatg cgcgtgggag actcggccgc ccgtgggtctg cccctgaggg ttcccaaaca 300

atcgtgtctg tgctcgttca actatctatt gatcaagtgg accggattgg cactattcca 360

ctcgcggcgg gactcgtgt catggatgcg ttgaatgacc tcggtgtgga aggtgccgga 420

ctgaaatggc ccaacgatgt tcaaatccac ggcaagaaac tctgcggcat cctggtggaa 480

gccaccggct ttgattccac cccaacagtt gtcacgggtt ggggcactaa taccagcctg 540

actaaagagg agcttctgt tctcatgca acttcctcg cattggaagg tgttgaagtc 600

gacagaacca cattccttat taatatgtc acacatctgc atactcgact ggaccagtgg 660

caggttccaa gtgtggattg gctcgtatg tacctgcgg tatgttcag tattggccaa 720

gatgttcgag tgcttctacc tggggataaa gaactcttag gtgaagcgat cgggtgtcgcg 780

actggcggag aaattcgtgt tcgcgatgct tcgggcaccg ttcacaccct caacgccggt 840

gaaattacgc accttcgct gcagtaa 867

<210> 46

<211> 810

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"

<400> 46

atgaatgttg acattagccg ctctcgtgaa ccgttgaacg tggaactgtt gaaagaaaaa 60
ctgctgcaga acggtgattt cggtaacgtg atctacgaga aggtcaccgg ctctaccaat 120
gcggacctgc tggtcttggc gggcagcggc gctccaaact ggaccgtcaa gactgttgaa 180
tttcaggacc acgcccgtgg ccgtctgggt cgtccgtgga gcgcaccgga gggttcccaa 240
accatcgtca gcgttctggt ccaactgagc attgatcagg tggaccgtat tggtagatc 300

ccgtctggccg caggcttggc tggttatggat gcgctgaatg atctgggcgt ggagggtgca 360
ggcctgaaat ggccgaacga tggtcagatc cacggtaaga agttgtgcgg tattctggtt 420
gaagcaaccg gcttcgactc cactccgacc gtggttatcg gttggggtac gaatatctcg 480
ttgacgaaag aagagctgcc ggtcccgac gcgaccagcc tggccctgga gggtgttgaa 540
gttgaccgta cgacgttctt gattaacatg ctgaccatc tgcatacccg tctggatcag 600
tggcagggtc cgtctgtgga ctggctggat gactatcgcg cggttttag cagcattggc 660
caagatgtgc gtgtcctgct gcctggtgac aaagagctgc tgggcgaggc gattggcgtg 720

gcgaccggtg gtgagatccg tgtgcgcgac gccagcggca cggccacac gctgaatgcg 780
ggtgaaatca cgcatctgcg ttgcaataa 810

<210> 47

<211> 269

<212> PRT

<213> Corynebacterium glutamicum

<400> 47

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

1 5 10 15

Leu Lys Glu Lys Leu Leu Gln Asn Gly Asp Phe Gly Gln Val Ile Tyr

20 25 30

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

35 40 45

Ser Gly Ala Pro Asn Trp Thr Val Lys Thr Val Glu Phe Gln Asp His

50 55 60

Ala Arg Gly Arg Leu Gly Arg Pro Trp Ser Ala Pro Glu Gly Ser Gln

65 70 75 80

Thr Ile Val Ser Val Leu Val Gln Leu Ser Ile Asp Gln Val Asp Arg

85 90 95

Ile Gly Thr Ile Pro Leu Ala Ala Gly Leu Ala Val Met Asp Ala Leu

100 105 110
Asn Asp Leu Gly Val Glu Gly Ala Gly Leu Lys Trp Pro Asn Asp Val
115 120 125
Gln Ile His Gly Lys Lys Leu Cys Gly Ile Leu Val Glu Ala Thr Gly
130 135 140
Phe Asp Ser Thr Pro Thr Val Val Ile Gly Trp Gly Thr Asn Ile Ser
145 150 155 160
Leu Thr Lys Glu Glu Leu Pro Val Pro His Ala Thr Ser Leu Ala Leu

165 170 175
Glu Gly Val Glu Val Asp Arg Thr Thr Phe Leu Ile Asn Met Leu Thr
180 185 190
His Leu His Thr Arg Leu Asp Gln Trp Gln Gly Pro Ser Val Asp Trp
195 200 205
Leu Asp Asp Tyr Arg Ala Val Cys Ser Ser Ile Gly Gln Asp Val Arg
210 215 220
Val Leu Leu Pro Gly Asp Lys Glu Leu Leu Gly Glu Ala Ile Gly Val

225 230 235 240
Ala Thr Gly Gly Glu Ile Arg Val Arg Asp Ala Ser Gly Thr Val His
245 250 255
Thr Leu Asn Ala Gly Glu Ile Thr His Leu Arg Leu Gln
260 265

<210> 48

<211> 1632

<212> DNA

<213> *Corynebacterium glutamicum*

<400> 48

atgaccattt cctcaccttt gattgacgtc gccaaccttc cagacatcaa caccactgcc 60

ggcaagatcg ccgaccttaa ggctcgccgc gcggaagccc atttcccat ggggtgaaaag 120

gcagtagaga aggtccacgc tgctggacgc ctcaactgccc gtgagcgctt ggattactta 180

ctcgatgagg gctccttcac cgagaccgat cagctggctc gccaccgcac caccgctttc 240

ggcctgggcg ctaagcgtcc tgcaaccgac ggcatcgtga ccggctgggg caccattgat 300
 ggacgcgaag tctgcatctt ctgcgaggac ggacacgtat tcggtggcgc gcttggtgag 360
 gtgtacggcg aaaagatgat caagatcatg gagctggcaa tcgacaccgg ccgcccattg 420
 atcggctctt acgaaggcgc tggcgctcgt attcaggacg gcgctgtctc cctggacttc 480
 atttcccaga cttctacca aaacattcag gcttctggcg ttatcccaca gatctccgtc 540

atcatgggcg catgtgcagg tggcaacgct tacggcccag ctctgaccga cttcgtggtc 600
 atggtggaca agacctcaa gatgttcgtt accggcccag acgtgatcaa gaccgtcacc 660
 ggcgaggaaa tcacccagga agagcttggc ggagcaacca cccacatggt gaccgtggt 720
 aactcccact acaccgtgc gaccgatgag gaagcactgg attgggtaca ggacctggtg 780
 tccttctcc catccaaca tcgctctac gcaccgatgg aagacttca cgaggaagaa 840
 ggcggcgttg aaaaaacat caccgtgac gatctgaagc tcgacgagat catccagat 900
 tccgcgaccg ttccttacga cgtccgcat gtcacgaat gcctcaccga cgatggcgaa 960

tacctggaaa tccaggcaga ccgcgcagaa aacgttgta ttgcattcgg ccgcatcgaa 1020
 ggccagtcgg ttggctttgt tgccaaccag ccaaccagt tcgctggctg cctggacatc 1080
 gactcctctg agaaggcgc tcgcttcgtc cgcacctgcg acgcgttcaa catccaatc 1140
 gtcatgcttg tcgacgtccc cggcttctc ccaggcgag gccaggagta cgggtgcatt 1200
 ctgcgtcgtg gcgcaaagct gctctacgca tacggcgaag caaccgttcc aaagatcacc 1260
 gtcacatgc gtaaggctta cggcgagcgc tactgcgtga tgggttcaa gggcttgggc 1320
 tctgacatca accttgcatg gccaacgca cagatcgccg tcattggcgc tgctggcgca 1380

gttggattca tciaccgcaa ggagctcatg gcagctgatg ccaagggcct cgataccgta 1440
 gctctggcta agtcttca ggcgagat gaagaccaca tgctaaccc gtaccacgt 1500
 gcagaacgtg gcctgatga cgccgtgatc ctgccaagcg aaaccgcgg acagatttc 1560
 cgcaaccttc gcctgtcaa gcacaagaac gtcactgcc ctgctcgcaa gcacggcaac 1620
 atgccactgt aa 1632

<210> 49

<211> 543

<212> PRT

<213> Corynebacterium glutamicum

<400> 49

Met Thr Ile Ser Ser Pro Leu Ile Asp Val Ala Asn Leu Pro Asp Ile

Asn Thr Thr Ala Gly Lys Ile Ala Asp Leu Lys Ala Arg Arg Ala Glu
 20 25 30
 Ala His Phe Pro Met Gly Glu Lys Ala Val Glu Lys Val His Ala Ala
 35 40 45
 Gly Arg Leu Thr Ala Arg Glu Arg Leu Asp Tyr Leu Leu Asp Glu Gly
 50 55 60
 Ser Phe Ile Glu Thr Asp Gln Leu Ala Arg His Arg Thr Thr Ala Phe

 65 70 75 80
 Gly Leu Gly Ala Lys Arg Pro Ala Thr Asp Gly Ile Val Thr Gly Trp
 85 90 95
 Gly Thr Ile Asp Gly Arg Glu Val Cys Ile Phe Ser Gln Asp Gly Thr
 100 105 110
 Val Phe Gly Gly Ala Leu Gly Glu Val Tyr Gly Glu Lys Met Ile Lys
 115 120 125
 Ile Met Glu Leu Ala Ile Asp Thr Gly Arg Pro Leu Ile Gly Leu Tyr

 130 135 140
 Glu Gly Ala Gly Ala Arg Ile Gln Asp Gly Ala Val Ser Leu Asp Phe
 145 150 155 160
 Ile Ser Gln Thr Phe Tyr Gln Asn Ile Gln Ala Ser Gly Val Ile Pro
 165 170 175
 Gln Ile Ser Val Ile Met Gly Ala Cys Ala Gly Gly Asn Ala Tyr Gly
 180 185 190
 Pro Ala Leu Thr Asp Phe Val Val Met Val Asp Lys Thr Ser Lys Met

 195 200 205
 Phe Val Thr Gly Pro Asp Val Ile Lys Thr Val Thr Gly Glu Glu Ile
 210 215 220
 Thr Gln Glu Glu Leu Gly Gly Ala Thr Thr His Met Val Thr Ala Gly
 225 230 235 240
 Asn Ser His Tyr Thr Ala Ala Thr Asp Glu Glu Ala Leu Asp Trp Val
 245 250 255
 Gln Asp Leu Val Ser Phe Leu Pro Ser Asn Asn Arg Ser Tyr Ala Pro

260 265 270
 Met Glu Asp Phe Asp Glu Glu Glu Gly Gly Val Glu Glu Asn Ile Thr
 275 280 285
 Ala Asp Asp Leu Lys Leu Asp Glu Ile Ile Pro Asp Ser Ala Thr Val
 290 295 300
 Pro Tyr Asp Val Arg Asp Val Ile Glu Cys Leu Thr Asp Asp Gly Glu
 305 310 315 320
 Tyr Leu Glu Ile Gln Ala Asp Arg Ala Glu Asn Val Val Ile Ala Phe
 325 330 335
 Gly Arg Ile Glu Gly Gln Ser Val Gly Phe Val Ala Asn Gln Pro Thr
 340 345 350
 Gln Phe Ala Gly Cys Leu Asp Ile Asp Ser Ser Glu Lys Ala Ala Arg
 355 360 365
 Phe Val Arg Thr Cys Asp Ala Phe Asn Ile Pro Ile Val Met Leu Val
 370 375 380
 Asp Val Pro Gly Phe Leu Pro Gly Ala Gly Gln Glu Tyr Gly Gly Ile
 385 390 395 400
 Leu Arg Arg Gly Ala Lys Leu Leu Tyr Ala Tyr Gly Glu Ala Thr Val
 405 410 415
 Pro Lys Ile Thr Val Thr Met Arg Lys Ala Tyr Gly Gly Ala Tyr Cys
 420 425 430
 Val Met Gly Ser Lys Gly Leu Gly Ser Asp Ile Asn Leu Ala Trp Pro
 435 440 445
 Thr Ala Gln Ile Ala Val Met Gly Ala Ala Gly Ala Val Gly Phe Ile
 450 455 460
 Tyr Arg Lys Glu Leu Met Ala Ala Asp Ala Lys Gly Leu Asp Thr Val
 465 470 475 480
 Ala Leu Ala Lys Ser Phe Glu Arg Glu Tyr Glu Asp His Met Leu Asn
 485 490 495
 Pro Tyr His Ala Ala Glu Arg Gly Leu Ile Asp Ala Val Ile Leu Pro
 500 505 510

Ser Glu Thr Arg Gly Gln Ile Ser Arg Asn Leu Arg Leu Leu Lys His

515	520	525	
Lys Asn Val Thr Arg Pro Ala Arg Lys His Gly Asn Met Pro Leu			
530	535	540	
<210> 50			
<211> 1776			
<212> DNA			
<213> Corynebacterium glutamicum			
<400> 50			
atgtcagtcg agactcgcaa gatcaccaag gttcttgtcg ctaaccgtgg tgagattgca	60		
atccgcgtgt tccgtgcagc tcgagatgaa ggcatcggat ctgtcgccgt ctacgcagag	120		
ccagatgcag atgcaccatt cgtgtcatat gcagacgagg cttttgccct cgggtggccaa	180		
acatccgctg agtcctacct tgtcattgac aagatcatcg atgcggcccg caagtccggc	240		
gccgacgcca tcaccccgg ctacggcttc ctgcagaaa acgtgactt cgcagaagca	300		
gtcatcaacg aaggcctgat ctggattgga ccttcacctg agtccatccg ctccctcggc	360		
gacaaggtea ccgctcgcca catgcagat accgccaagg ctccaatggc tcctggcacc	420		
aaggaaccag taaaagacgc agcagaagtt gtggctttcg ctgaagaatt cgggtctccca	480		
atcgccatca aggcagcttt cgggtggcggc ggacgtggca tgaaggttgc ctacaagatg	540		
gaagaagtcg ctgacctctt cgagtccgca acccgtgaag caaccgcagc gttcggccgc	600		
ggcgagtgtc tcgtggagcg ctacctggac aaggcacgcc acgttgaggc tcaggtcac	660		
gccgataagc acggcaacgt tgttgtcgcc ggaaccctg actgtccct gcagcgccgt	720		
ttccagaagc tcgtcgaaga agcaccagca ccattcctca ccgatgacca gcgcgagcgt	780		
ctccactcct ccgcaaggc tatctgtaag gaagctggct actacggtgc aggcaccgtt	840		
gagtacctcg ttggctccga cggcctgac tccttctcgc aggtcaacac ccgcctccag	900		
gtggaacacc cagtcaccga agagaccacc ggcatcgacc tggtcgcga aatgttccgc	960		
atcgcagaag gccacgagct ctccatcaag gaagatccag ctccacgcgg ccacgcattc	1020		
gagttccga tcaacggcga agacgtggc tccaacttca tgctgcacc aggcaagatc	1080		
accagctacc gcgagccaca gggcccaggc gtccgcatgg actccggtgt cgttgaaggt	1140		
tccgaaatct ccggacagtt cgactccatg ctggcaaagc tgatcgtttg gggcgacacc	1200		
cgcgagcagg ctctccagcg ctcccgccgt gcacttgcag agtacgttgt cgagggcag	1260		
ccaaccgtta tccattcca ccagcacatc gtggaaaacc cagcattcgt gggcaacgac	1320		

gaaggttcg agatctacac caagtggatc gaagaggttt gggataaccc aatcgcacct 1380
 tacgttgacg ctccgagct cgacgaagat gaggacaaga ccccgacaca gaaggttggt 1440
 gtggagatca acggccgctg cgttgaggtt gcactcccag gcgatctggc actcggtggc 1500

accgctggtc ctaagaagaa ggccaagaag cgtcgcgcag gtggtgcaaa ggctggcgta 1560
 tccggcgatg cagtggcagc tccaatgcag ggactgtca tcaaggtcaa cgtcgaagaa 1620
 ggcgctgaag tcaacgaagg cgacaccgtt gttgtcctcg aggctatgaa gatggaaaac 1680
 cctgtgaagg ctcataagtc cggaaccgta accggcctta ctgtcgtgc aggcgagggt 1740
 gtcaacaagg gcgttgttct cctcgagatc aagtaa 1776

<210> 51

<211> 591

<212> PRT

<213> Corynebacterium glutamicum

<400> 51

Met Ser Val Glu Thr Arg Lys Ile Thr Lys Val Leu Val Ala Asn Arg

1 5 10 15
 Gly Glu Ile Ala Ile Arg Val Phe Arg Ala Ala Arg Asp Glu Gly Ile
 20 25 30
 Gly Ser Val Ala Val Tyr Ala Glu Pro Asp Ala Asp Ala Pro Phe Val
 35 40 45
 Ser Tyr Ala Asp Glu Ala Phe Ala Leu Gly Gly Gln Thr Ser Ala Glu
 50 55 60
 Ser Tyr Leu Val Ile Asp Lys Ile Ile Asp Ala Ala Arg Lys Ser Gly

65 70 75 80
 Ala Asp Ala Ile His Pro Gly Tyr Gly Phe Leu Ala Glu Asn Ala Asp
 85 90 95
 Phe Ala Glu Ala Val Ile Asn Glu Gly Leu Ile Trp Ile Gly Pro Ser
 100 105 110
 Pro Glu Ser Ile Arg Ser Leu Gly Asp Lys Val Thr Ala Arg His Ile
 115 120 125
 Ala Asp Thr Ala Lys Ala Pro Met Ala Pro Gly Thr Lys Glu Pro Val

130

135

140

Lys Asp Ala Ala Glu Val Val Ala Phe Ala Glu Glu Phe Gly Leu Pro
 145 150 155 160
 Ile Ala Ile Lys Ala Ala Phe Gly Gly Gly Gly Arg Gly Met Lys Val
 165 170 175
 Ala Tyr Lys Met Glu Glu Val Ala Asp Leu Phe Glu Ser Ala Thr Arg
 180 185 190
 Glu Ala Thr Ala Ala Phe Gly Arg Gly Glu Cys Phe Val Glu Arg Tyr

 195 200 205
 Leu Asp Lys Ala Arg His Val Glu Ala Gln Val Ile Ala Asp Lys His
 210 215 220
 Gly Asn Val Val Val Ala Gly Thr Arg Asp Cys Ser Leu Gln Arg Arg
 225 230 235 240
 Phe Gln Lys Leu Val Glu Glu Ala Pro Ala Pro Phe Leu Thr Asp Asp
 245 250 255
 Gln Arg Glu Arg Leu His Ser Ser Ala Lys Ala Ile Cys Lys Glu Ala

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

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

385 390 395 400
 Arg Glu Gln Ala Leu Gln Arg Ser Arg Arg Ala Leu Ala Glu Tyr Val
 405 410 415
 Val Glu Gly Met Pro Thr Val Ile Pro Phe His Gln His Ile Val Glu
 420 425 430
 Asn Pro Ala Phe Val Gly Asn Asp Glu Gly Phe Glu Ile Tyr Thr Lys
 435 440 445
 Trp Ile Glu Glu Val Trp Asp Asn Pro Ile Ala Pro Tyr Val Asp Ala

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

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

 580 585 590
 <210> 52
 <211> 340
 <212> PRT
 <213> Acinetobacter sp.
 <400> 52
 Met Ser Asn His Gln Ile Arg Ala Tyr Ala Ala Met Gln Ala Gly Glu
 1 5 10 15

Gln Val Val Pro Tyr Gln Phe Asp Ala Gly Glu Leu Lys Ala His Gln
20 25 30

Val Glu Val Lys Val Glu Tyr Cys Gly Leu Cys His Ser Asp Leu Ser
35 40 45

Val Ile Asn Asn Glu Trp Gln Ser Ser Val Tyr Pro Ala Val Ala Gly
50 55 60

His Glu Ile Ile Gly Thr Ile Ile Ala Leu Gly Ser Glu Ala Lys Gly
65 70 75 80

Leu Lys Leu Gly Gln Arg Val Gly Ile Gly Trp Thr Ala Glu Thr Cys
85 90 95

Gln Ala Cys Asp Pro Cys Ile Gly Gly Asn Gln Val Leu Cys Thr Gly
100 105 110

Glu Lys Lys Ala Thr Ile Ile Gly His Ala Gly Gly Phe Ala Asp Lys
115 120 125

Val Arg Ala Gly Trp Gln Trp Val Ile Pro Leu Pro Asp Asp Leu Asp
130 135 140

Pro Glu Ser Ala Gly Pro Leu Leu Cys Gly Gly Ile Thr Val Leu Asp
145 150 155 160

Pro Leu Leu Lys His Lys Ile Gln Ala Thr His His Val Gly Val Ile
165 170 175

Gly Ile Gly Gly Leu Gly His Ile Ala Ile Lys Leu Leu Lys Ala Trp
180 185 190

Gly Cys Glu Ile Thr Ala Phe Ser Ser Asn Pro Asp Lys Thr Glu Glu
195 200 205

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

Ala Ile Lys Gly Thr Arg Trp Lys Leu Ile Ile Leu Ser Thr Ala Asn
225 230 235 240

Gly Thr Leu Asn Val Lys Ala Tyr Leu Asn Thr Leu Ala Pro Lys Gly
245 250 255

Ser Leu His Phe Leu Gly Val Thr Leu Glu Pro Ile Pro Val Ser Val

260 265 270
 Gly Ala Ile Met Gly Gly Ala Lys Ser Val Thr Ser Ser Pro Thr Gly
 275 280 285
 Ser Pro Leu Ala Leu Arg Gln Leu Leu Gln Phe Ala Ala Arg Lys Asn
 290 295 300

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

<210> 53

<211> 314

<212> PRT

<213> Acinetobacter baylyi

<400> 53

Met Ala Thr Thr Asn Val Ile His Ala Tyr Ala Ala Met Gln Ala Gly
 1 5 10 15
 Glu Ala Leu Val Pro Tyr Ser Phe Asp Ala Gly Glu Leu Gln Pro His

20 25 30
 Gln Val Glu Val Lys Val Glu Tyr Cys Gly Leu Cys His Ser Asp Val
 35 40 45
 Ser Val Leu Asn Asn Glu Trp His Ser Ser Val Tyr Pro Val Val Ala
 50 55 60
 Gly His Glu Val Ile Gly Thr Ile Thr Gln Leu Gly Ser Glu Ala Lys
 65 70 75 80
 Gly Leu Lys Ile Gly Gln Arg Val Gly Ile Gly Trp Thr Ala Glu Ser

85 90 95
 Cys Gln Ala Cys Asp Gln Cys Ile Ser Gly Gln Gln Val Leu Cys Thr
 100 105 110
 Gly Glu Asn Thr Ala Thr Ile Ile Gly His Ala Gly Gly Phe Ala Asp
 115 120 125

Lys Val Arg Ala Gly Trp Gln Trp Val Ile Pro Leu Pro Asp Glu Leu
 130 135 140
 Asp Pro Thr Ser Ala Gly Pro Leu Leu Cys Gly Gly Ile Thr Val Phe
 145 150 155 160
 Asp Pro Ile Leu Lys His Gln Ile Gln Ala Ile His His Val Ala Val
 165 170 175
 Ile Gly Ile Gly Gly Leu Gly His Met Ala Ile Lys Leu Leu Lys Ala
 180 185 190
 Trp Gly Cys Glu Ile Thr Ala Phe Ser Ser Asn Pro Asn Lys Thr Asp
 195 200 205
 Glu Leu Lys Ala Met Gly Ala Asp His Val Val Asn Ser Arg Asp Asp
 210 215 220
 Ala Glu Ile Lys Ser Gln Gln Gly Lys Phe Asp Leu Leu Leu Ser Thr
 225 230 235 240
 Val Asn Val Pro Leu Asn Trp Asn Ala Tyr Leu Asn Thr Leu Ala Pro
 245 250 255
 Asn Gly Thr Phe His Phe Leu Gly Val Val Met Glu Pro Ile Pro Val
 260 265 270
 Pro Val Gly Ala Leu Leu Gly Gly Ala Lys Ser Leu Thr Ala Ser Pro
 275 280 285
 Thr Gly Ser Pro Ala Ala Leu Arg Lys Leu Leu Glu Phe Ala Ala Arg
 290 295 300
 Lys Asn Ile Ala Pro Gln Ile Glu Met Tyr
 305 310
 <210> 54
 <211> 341
 <212> PRT
 <213> Acinetobacter baylyi
 <400> 54
 Met Ala Thr Thr Asn Val Ile His Ala Tyr Ala Ala Met Gln Ala Gly
 1 5 10 15

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

20 25 30

Gln Val Glu Val Lys Val Glu Tyr Cys Gly Leu Cys His Ser Asp Val

35 40 45

Ser Val Leu Asn Asn Glu Trp His Ser Ser Val Tyr Pro Val Val Ala

50 55 60

Gly His Glu Val Ile Gly Thr Ile Thr Gln Leu Gly Ser Glu Ala Lys

65 70 75 80

Gly Leu Lys Ile Gly Gln Arg Val Gly Ile Gly Trp Thr Ala Glu Ser

85 90 95

Cys Gln Ala Cys Asp Gln Cys Ile Ser Gly Gln Gln Val Leu Cys Thr

100 105 110

Gly Glu Asn Thr Ala Thr Ile Ile Gly His Ala Gly Gly Phe Ala Asp

115 120 125

Lys Val Arg Ala Gly Trp Gln Trp Val Ile Pro Leu Pro Asp Glu Leu

130 135 140

Asp Pro Thr Ser Ala Gly Pro Leu Leu Cys Gly Gly Ile Thr Val Phe

145 150 155 160

Asp Pro Ile Leu Lys His Gln Ile Gln Ala Ile His His Val Ala Val

165 170 175

Ile Gly Ile Gly Gly Leu Gly His Met Ala Ile Lys Leu Leu Lys Ala

180 185 190

Trp Gly Cys Glu Ile Thr Ala Phe Ser Ser Asn Pro Asn Lys Thr Asp

195 200 205

Glu Leu Lys Ala Met Gly Ala Asp His Val Val Asn Ser Arg Asp Asp

210 215 220

Ala Glu Ile Lys Ser Gln Gln Gly Lys Phe Asp Leu Leu Leu Ser Thr

225 230 235 240

Val Asn Val Pro Leu Asn Trp Asn Ala Tyr Leu Asn Thr Leu Ala Pro

245 250 255

Asn Gly Thr Phe His Phe Leu Gly Val Val Met Glu Pro Ile Pro Val

260 265 270
Pro Val Gly Ala Leu Leu Gly Gly Ala Lys Ser Leu Thr Ala Ser Pro

275 280 285
Thr Gly Ser Pro Ala Ala Leu Arg Lys Leu Leu Glu Phe Ala Ala Arg

290 295 300
Lys Asn Ile Ala Pro Gln Ile Glu Met Tyr Pro Met Ser Glu Leu Asn

305 310 315 320
Glu Ala Ile Glu Arg Leu His Ser Gly Gln Ala Arg Tyr Arg Ile Val

325 330 335
Leu Lys Ala Asp Phe

340

<210

> 55

<211> 4805

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"

<400> 55

tgtaggctgg agctgcttcg aagttcctat actttctaga gaataggaac ttcggaatag 60

gaacttcgaa ctgcaggctcg acggatcccc ggaatatatta aatcatttgt actttttgaa 120

cagcagagtc gcattatggc caccgaagcc caggctgttg gacagaacgt agttgacttc 180

tgcattacgg cctcgttag gaacgtaatc caggctcgc at tccggatccg cctctttgta 240

gccgatggtc ggcggaatga aacctcttc aatagctttg gcacagataa tcgcttcgac 300

tgcaccgcca gcgccagca ggtggccggt catgctcttg gtgctagaca ccggcacttt 360

glaggcgtat tcaccagga ccgtcttgat cgcttgggtt tcgaagctgt cattgtacgc 420

cgtgctcgta ccgtgcgcgt tgatatagga aatgtcctct gggcggacat tatcttcttc 480

cattgccagt ttattgcac gtgcaccacc ttcaccattc ggcgctgggc tcgtgatatg 540

atatgcgtcg caggtcgcac catagccaac gatctcggca tagattttgg caccacgctt 600

cagcgcgtgc tccaactctt ccaagataac gataccgtg cctcgcacca tcacaaaacc 660

gctgcgatcc ttatcgaacg ggatgctggc gcgcttcggg tcctcagatt tggtcacggc 720

cttcatcgag gcaaaacccg ccaggctcaa cggggtgata cctgcttcgc taccaccaga 780
gatcataacg tcgctataac caaacttaat gttacggaag gactaccaa tgctgttgtt 840
cgcgctcgca catgcggtga caatggctgt gcaaatacct ttagcgccat aacgaatcgc 900
cagattaccg ctigccatat tcgcaatgat catcggaata gtcatagggc tcacacgacc 960
cggacctttg gtaatcagct tttcatcctg cttctcaatg gtgccgatgc cgccaatgcc 1020
gctaccaaca atgacgccga aacgattctt atcaatcgac tccagggtcca gtttgcgtgc 1080
cttgattgcc tcatccgccg caacgatcgc aaactggcta aaacgggtcca tacggttcgc 1140

ctcacgcttg tcgataaagt cctccggggt gaagtccctc acttcggcag ccagcttaac 1200
tttgaatcgc gttgcgtcaa acgctttgat cttgtcaatg ccacatttac cctctttgat 1260
gctgcaccag aagctatcag cgttggttacc caccggcgtc actgcaccaa taccgtaat 1320
gacaacgcgg cgattcattt tgttgccctc ttttagaacg cggaagtatc ctggaacaaa 1380
ccgactttca aatcgtgtgc ggtatagatc aggcgaccat ccaccagaac ctacacgtcc 1440
gccaggccca tgatcaggcg acggtttacg atacgtttga aatgaatacg atagggtgact 1500
ttcttggtg tcggcagaac ctggccggta aatttcactt cgcccacgcc cagagcgcg 1560

cctttgcctt cgcgcgcccc ccagcccagg tagaatccca ccaattgcc catagcatcc 1620
agaccagac aaccgggcat caccggatcg ccgataaagt ggcatccga gaaccataga 1680
tccggattga tatccagctc ggcttcgaca tagcctttgt cgaaattgcc gcccgtttcg 1740
gtcatcttaa cgacgcggtc catcatcagc atgttcggtg cagggagtgt cggcccttta 1800
gcgccaacaa gttcaccacg accagaggca agaaggctct cttttgtata ggattcgcgt 1860
ttatctacca tgttttatgt aaaccttaaa attaaacat gtacattccg ccgttgacgt 1920
gcagagtctc accagtgatg taactcgtt cgtcagaggc taaaaatgca accgcactgg 1980

cgatttctg agcgcgcgcg aggcgacccg caggcacctg cgccaggata cccgcacgt 2040
gatcgtcaga cagcgcacgc gtcatgtccg tttcaataaa acccgagacc acaacattga 2100
cagtaatacc acgggacgca acttcacgcg ccagtgttt actgaaaccg atcaggcccc 2160
ctttcgccgc agcgtagtgt gcctgacctg catttccat ggtaccaacc acagaaccaa 2220
tagtgataat gcgaccacaa cgctttttca tcatagcgcg cattaccgct tttagacggc 2280
ggaaaacgga tgataagttg gtttcgataa tatcgttcca ctcatcatct ttcattcgca 2340
tcaacagatt atcacgagt ataccggcat tattaaccag gatatccact tcaccaaatt 2400

ctgcgcgaat attttcaga acagattcaa tagatgcagg atcggtcaca ttcaacatca 2460
aacctttccc gttagacct aaatagtcgc taatgttctt cgcaccattt tcaactggtcg 2520
cagtcgccat aactttcgcg ccgcgggcaa cgagagtctc tgcaattgcg cggcctatgc 2580

cacggcttgc accagtcacc agcgcaatct ttccttcaaa gctcatgggtt ttcctctttt	2640
attgcgtaag tgccgcagac agcgccgccc gctcgttcag cgccgacgct gtcagggtgt	2700
cgacaatacg tttcgtcaga ccagtggagga ctttacctgg acccacttca taaagatgtt	2760
caacgccctg cgccgcgata aattccacgc tcttcgtcca ctgtaccgga ttgtacaact	2820
ggcgaaccag cgcatcgcg atagcgggcg catcggtttc acatttcacg tcaacgttgt	2880
tcactaccgg caccgttggc gcgctaaagg taattttggc taattcaacc gccagcttat	2940
ctgccgttgg tttcatcagc gcgcagtgcg acggtacgct caccggcagc ggcagcgcg	3000
gtttcgcgcc agcggtttta caggctgcgc ccgcacgttc taccgctct tttatgcccg	3060
cgataaccac ctgtcccgcc gagttaaagt taaccggcga aacaacctgc ctttcggcag	3120
attcttcaca ggcttttagc atagaggcat catccagccc gatgatcgca gacatgccgc	3180
cagtgccttc cggaaccgct tctgcatga atttaccgcg catttcacc agacgaacgg	3240
catcagcaaa gttgatgacg ccagcgcaaa ccagcgcgga atattcgccc aggtgtgac	3300
ctgccattaa cgcaggcatt ttaccgcctt gctgctgcca aacgcgcaa agcgcgacgg	3360
aagcggttaa taacgccgcg tgcgtctgcc aggtttttatt cagttcttcc gctggacctt	3420
gctgggttag cgcacacaga tcatatcca gagccgcaga agcttcagca aacgtttctt	3480
ctacgatagg gtaatttgcc gccatctcgg ccaacatccc aacgctctga gaacctgac	3540
cggggaacac aaatgcaaat tgcgtcatgt ttaaatcctt atactagaaa cgaatcagcg	3600
cggagcccca ggtgaatcca ccccggaagg cttcaagcaa taccagctga ccggctttaa	3660
ttcgcccgtc acgcacggct tcatccagcg cgcacggcac agaagccgcg gaggtattgc	3720
cgtgcctgtc cagcgtgacg acgacattgt ccatcgacat gccgagtttt ttcgctgtcg	3780
cgctaatgat acgcaggtta gcctgatgcg gcaccagcca atcgagttct gagcgatcca	3840
ggttattagc cgccagcgtc tcatcgacaa tatgcgccag ttcagtgacc gccactttaa	3900
agacttcatt gcccgccatt gtcaggtaaa tcgggttata cggatttacg cgatcggcac	3960
tcggcagggt cagtaattca ccgtaacggc catcggcacg aagatgagtg gagataatac	4020
ccggttcttc agaagcgctc agtacggccg cgcctgcgcc atcgccgaaa ataatgatcg	4080
taccgcgac gccaggatcg caagtgcggg ctaatacatc ggaaccgacc accagcgct	4140
gtttaaccgc gccggattta acgtactggt cggcgatgct taacgcgtag gtgaaacctg	4200
cgcacgctgc cgcgacatca aacgccgggc aacctttaat accgagcata ctttgaatct	4260
gacatgccgc gcttggaat gcacgcgttg ctgatgtggt agccaccaca atcaagccaa	4320
tttggtcttt atcgatcccc gccatctcaa tcgcgcgatt cgcagcggt aagcccatcg	4380
tcgcgacagt ttatttcggc gcggcgatat ggcggtttacg aatacctgta cgagtgacaa	4440

tccactcgtc agaggtctca accatTTTTT ccagatcggc gttagtccgc acttggtcgg 4500

gcagatagct gccagtacca ataattctcg tatacatgta cgctcagtca ctaaattact 4560

cgatatcaat cacatcaaat tcgacttctg gattgacgtc agcatcgtaa tcaatgcctt 4620

caatgccaaa gccaaacagc ttgatgaact cttctttgta catgtcgtaa tcggtcagct 4680

cacgcaggtt ctctgtggtg atttgtggcc acagatcacg gcagtgtgc tgaatgtcat 4740

cacgcagttc ccagtcaccc aaacgcagac gattgtgate atccacttcc ggcgtgaac 4800

catct 4805

<210> 56

<211> 239

<212> PRT

<213> Escherichia coli

<400> 56

Met Val Ile Lys Ala Gln Ser Pro Ala Gly Phe Ala Glu Glu Tyr Ile

1 5 10 15

Ile Glu Ser Ile Trp Asn Asn Arg Phe Pro Pro Gly Thr Ile Leu Pro

20 25 30

Ala Glu Arg Glu Leu Ser Glu Leu Ile Gly Val Thr Arg Thr Thr Leu

35 40 45

Arg Glu Val Leu Gln Arg Leu Ala Arg Asp Gly Trp Leu Thr Ile Gln

50 55 60

His Gly Lys Pro Thr Lys Val Asn Asn Phe Trp Glu Thr Ser Gly Leu

65 70 75 80

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

85 90 95

Leu Ile Asp Asn Leu Leu Ser Val Arg Thr Asn Ile Ser Thr Ile Phe

100 105 110

Ile Arg Thr Ala Phe Arg Gln His Pro Asp Lys Ala Gln Glu Val Leu

115 120 125

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

130 135 140

Asp Tyr Asn Ile Phe Arg Gly Leu Ala Phe Ala Ser Gly Asn Pro Ile

145 150 155 160
 Tyr Gly Leu Ile Leu Asn Gly Met Lys Gly Leu Tyr Thr Arg Ile Gly
 165 170 175
 Arg His Tyr Phe Ala Asn Pro Glu Ala Arg Ser Leu Ala Leu Gly Phe
 180 185 190

 Tyr His Lys Leu Ser Ala Leu Cys Ser Glu Gly Ala His Asp Gln Val
 195 200 205
 Tyr Glu Thr Val Arg Arg Tyr Gly His Glu Ser Gly Glu Ile Trp His
 210 215 220
 Arg Met Gln Lys Asn Leu Pro Gly Asp Leu Ala Ile Gln Gly Arg
 225 230 235
 <210> 57
 <211> 342
 <212> PRT
 <213> Synechococcus elongatus
 <400> 57
 Met Ala Phe Gly Leu Ile Gly His Leu Thr Ser Leu Glu Gln Ala Arg

 1 5 10 15
 Asp Val Trp Arg Arg Met Gly Tyr Asp Glu Tyr Ala Asp Gln Gly Leu
 20 25 30
 Glu Phe Trp Ser Ser Ala Pro Pro Gln Ile Val Asp Glu Ile Thr Val
 35 40 45
 Thr Ser Ala Thr Gly Lys Val Ile His Gly Arg Tyr Ile Glu Ser Cys
 50 55 60
 Phe Leu Pro Glu Met Leu Ala Ala Arg Arg Phe Lys Thr Ala Thr Arg

 65 70 75 80
 Lys Val Leu Asn Ala Met Ser His Ala Gln Lys His Gly Ile Asp Ile
 85 90 95
 Ser Ala Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asp Leu
 100 105 110
 Ala Ser Leu Arg Gln Val Arg Asp Thr Thr Leu Glu Phe Glu Arg Phe
 115 120 125

Thr Thr Gly Asn Thr His Thr Ala Tyr Val Ile Cys Arg Gln Val Glu

130 135 140

Ala Ala Ala Lys Thr Leu Gly Ile Asp Ile Thr Gln Ala Thr Val Ala

145 150 155 160

Val Val Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu

165 170 175

Asp Leu Lys Leu Gly Val Gly Asp Leu Ile Leu Thr Ala Arg Asn Gln

180 185 190

Glu Arg Leu Asp Asn Leu Gln Ala Glu Leu Gly Arg Gly Lys Ile Leu

195 200 205

Pro Leu Glu Ala Ala Leu Pro Glu Ala Asp Phe Ile Val Trp Val Ala

210 215 220

Ser Met Pro Gln Gly Val Val Ile Asp Pro Ala Thr Leu Lys Gln Pro

225 230 235 240

Cys Val Leu Ile Asp Gly Gly Tyr Pro Lys Asn Leu Gly Ser Lys Val

245 250 255

Gln Gly Glu Gly Ile Tyr Val Leu Asn Gly Gly Val Val Glu His Cys

260 265 270

Phe Asp Ile Asp Trp Gln Ile Met Ser Ala Ala Glu Met Ala Arg Pro

275 280 285

Glu Arg Gln Met Phe Ala Cys Phe Ala Glu Ala Met Leu Leu Glu Phe

290 295 300

Glu Gly Trp His Thr Asn Phe Ser Trp Gly Arg Asn Gln Ile Thr Ile

305 310 315 320

Glu Lys Met Glu Ala Ile Gly Glu Ala Ser Val Arg His Gly Phe Gln

325 330 335

Pro Leu Ala Leu Ala Ile

340

<210> 58

<211> 342

<212> PRT

<213> Synechococcus elongatus

<400> 58

Met Ala Phe Gly Leu Ile Gly His Leu Thr Ser Leu Glu Gln Ala Arg

1 5 10 15

Asp Val Trp Arg Arg Leu Gly Tyr Asp Glu Tyr Ala Asp Gln Gly Leu

20 25 30

Glu Phe Trp Ser Ser Ala Pro Pro Gln Ile Val Asp Glu Ile Thr Val

35 40 45

Thr Ser Ala Thr Gly Lys Val Ile His Gly Arg Tyr Ile Glu Ser Gly

50 55 60

Phe Leu Pro Glu Met Leu Ala Ala Arg Arg Phe Lys Thr Ala Thr Arg

65 70 75 80

Lys Val Leu Asn Ala Met Ser His Ala Gln Lys His Gly Ile Asp Ile

85 90 95

Ser Ala Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asp Leu

100 105 110

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

115 120 125

Thr Thr Gly Asn Thr His Thr Ala Tyr Val Ile Cys Arg Gln Val Glu

130 135 140

Ala Ala Ala Lys Thr Leu Gly Ile Asp Ile Ala Gln Ala Thr Val Ala

145 150 155 160

Val Val Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu

165 170 175

Asp Leu Lys Leu Gly Val Gly Asp Leu Ile Leu Thr Ala Arg Asn Gln

180 185 190

Glu Arg Leu Asp Asn Leu Gln Ala Glu Leu Gly Arg Gly Lys Ile Leu

195 200 205

Pro Leu Glu Ala Ala Leu Pro Glu Ala Asp Phe Ile Val Trp Val Ala

210 215 220

Ser Met Pro Gln Gly Val Val Ile Asp Pro Ala Thr Leu Lys Gln Pro

225 230 235 240
 Cys Val Leu Ile Asp Gly Gly Tyr Pro Lys Asn Leu Gly Ser Lys Val
 245 250 255
 Gln Gly Glu Gly Ile Tyr Val Leu Asn Gly Gly Val Val Glu His Cys
 260 265 270
 Phe Asp Ile Asp Trp Gln Ile Met Ser Leu Ala Glu Met Ala Arg Pro
 275 280 285
 Glu Arg Gln Met Phe Ala Cys Phe Ala Glu Ala Met Leu Leu Glu Phe

 290 295 300
 Glu Gly Trp His Thr Asn Phe Ser Trp Gly Arg Asn Gln Ile Thr Ile
 305 310 315 320
 Glu Lys Met Glu Ala Ile Gly Glu Ala Ser Val Arg His Gly Phe Gln
 325 330 335
 Pro Leu Ala Leu Ala Ile
 340

<210> 59

<211> 342

<212> PRT

<213> Synechococcus elongatus

<400> 59

Met Ala Phe Gly Leu Ile Gly His Ala Thr Ser Leu Glu Gln Ala Arg

1 5 10 15
 Asp Val Trp Arg Arg Leu Gly Tyr Asp Glu Tyr Ala Asp Gln Gly Leu
 20 25 30
 Glu Phe Trp Ser Ser Ala Pro Pro Gln Ile Val Asp Glu Ile Thr Val
 35 40 45
 Thr Ser Ala Thr Gly Lys Val Ile His Gly Arg Tyr Ile Glu Ser Gly
 50 55 60
 Phe Leu Pro Glu Met Leu Ala Ala Arg Arg Phe Lys Thr Ala Thr Arg

65 70 75 80
 Lys Val Leu Asn Ala Met Ser His Ala Gln Lys His Gly Ile Asp Ile
 85 90 95

Ser Ala Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asp Leu
100 105 110

Ala Lys Leu Arg Gln Val Arg Asp Thr Thr Leu Glu Phe Glu Arg Phe
115 120 125

Thr Thr Gly Asn Thr His Thr Ala Tyr Val Ile Cys Arg Gln Val Glu
130 135 140

Ala Ala Ala Lys Thr Leu Gly Ile Asp Ile Ala Gln Ala Thr Val Ala
145 150 155 160

Val Val Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu
165 170 175

Asp Leu Lys Leu Gly Val Gly Asp Leu Ile Leu Thr Ala Arg Asn Gln
180 185 190

Glu Arg Leu Asp Asn Leu Gln Ala Glu Leu Gly Arg Gly Lys Ile Leu
195 200 205

Pro Leu Glu Ala Ala Leu Pro Glu Ala Asp Phe Ile Val Trp Val Ala
210 215 220

Ser Met Pro Gln Gly Val Val Ile Asp Pro Ala Thr Leu Lys Gln Pro
225 230 235 240

Cys Val Leu Ile Asp Gly Gly Tyr Pro Lys Asn Leu Gly Ser Lys Val
245 250 255

Gln Gly Glu Gly Ile Tyr Val Leu Asn Gly Gly Val Val Glu His Cys
260 265 270

Phe Asp Ile Asp Trp Gln Ile Met Ser Leu Ala Glu Met Ala Arg Pro
275 280 285

Glu Arg Gln Met Phe Ala Cys Phe Ala Glu Ala Met Leu Leu Glu Phe
290 295 300

Glu Gly Trp His Thr Asn Phe Ser Trp Gly Arg Asn Gln Ile Thr Ile
305 310 315 320

Glu Lys Met Glu Ala Ile Gly Glu Ala Ser Val Arg His Gly Phe Gln
325 330 335

Pro Leu Ala Leu Ala Ile

340

<210> 60

<211> 342

<212> PRT

<213> *Synechococcus elongatus*

<400> 60

Met Ala Phe Gly Leu Ile Gly His Leu Thr Ser Leu Glu Gln Ala Arg

1 5 10 15

Leu Val Trp Arg Arg Leu Gly Tyr Asp Glu Tyr Ala Asp Gln Gly Leu

20 25 30

Glu Phe Trp Ser Ser Ala Pro Pro Gln Ile Val Asp Glu Ile Thr Val

35 40 45

Thr Ser Ala Thr Gly Lys Val Ile His Gly Arg Tyr Ile Glu Ser Gly

50 55 60

Phe Leu Pro Glu Met Leu Ala Ala Arg Arg Phe Lys Thr Ala Thr Arg

65 70 75 80

Lys Val Leu Asn Ala Met Ser His Ala Gln Lys His Gly Ile Asp Ile

85 90 95

Ser Ala Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asp Leu

100 105 110

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

115 120 125

Thr Thr Gly Asn Thr His Thr Ala Tyr Val Ile Cys Arg Gln Val Glu

130 135 140

Ala Ala Ala Lys Thr Leu Gly Ile Asp Ile Ala Gln Ala Thr Val Ala

145 150 155 160

Val Val Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu

165 170 175

Asp Leu Lys Leu Gly Val Gly Asp Leu Ile Leu Thr Ala Arg Asn Gln

180 185 190

Glu Arg Leu Asp Asn Leu Gln Ala Glu Leu Gly Arg Gly Lys Ile Leu

195 200 205

Pro Leu Glu Ala Ala Leu Pro Glu Ala Asp Phe Ile Val Trp Val Ala
210 215 220

Ser Met Pro Gln Gly Val Val Ile Asp Pro Ala Thr Leu Lys Gln Pro

225 230 235 240

Cys Val Leu Ile Asp Gly Gly Tyr Pro Lys Asn Leu Gly Ser Lys Val
245 250 255

Gln Gly Glu Gly Ile Tyr Val Leu Asn Gly Gly Val Val Glu His Cys
260 265 270

Phe Asp Ile Asp Trp Gln Ile Met Ser Leu Ala Glu Met Ala Arg Pro
275 280 285

Glu Arg Gln Met Phe Ala Cys Phe Ala Glu Ala Met Leu Leu Glu Phe

290 295 300

Glu Gly Trp His Thr Asn Phe Ser Trp Gly Arg Asn Gln Ile Thr Ile
305 310 315 320

Glu Lys Met Glu Ala Ile Gly Glu Ala Ser Val Arg His Gly Phe Gln
325 330 335

Pro Leu Ala Leu Ala Ile
340

<210> 61

<211> 342

<212> PRT

<213> Synechococcus elongatus

<400> 61

Met Ala Phe Gly Leu Ile Gly His Ala Thr Ser Leu Glu Gln Ala Arg

1 5 10 15

Asp Val Trp Arg Arg Met Gly Tyr Val Glu Tyr Ala Asp Gln Gly Leu
20 25 30

Glu Phe Trp Ser Ser Ala Pro Pro Gln Ile Val Asp Glu Ile Thr Val
35 40 45

Thr Ser Ala Thr Gly Lys Val Ile His Gly Arg Tyr Ile Glu Ser Gly
50 55 60

Phe Leu Pro Glu Met Leu Ala Ala Arg Arg Phe Lys Thr Ala Thr Arg

65 70 75 80

Lys Val Leu Asn Ala Met Ser His Ala Gln Lys His Gly Ile Asp Ile

85 90 95

Ser Ala Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asp Leu

100 105 110

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

115 120 125

Thr Thr Gly Asn Thr His Thr Ala Tyr Val Ile Cys Arg Gln Val Glu

130 135 140

Ala Ala Ala Lys Thr Leu Gly Ile Asp Ile Thr Leu Ala Thr Val Ala

145 150 155 160

Val Val Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu

165 170 175

Asp Leu Lys Leu Gly Val Gly Asp Leu Ile Leu Thr Ala Arg Asn Gln

180 185 190

Glu Arg Leu Asp Asn Leu Gln Ala Glu Leu Gly Arg Gly Lys Ile Leu

195 200 205

Pro Leu Glu Ala Ala Leu Pro Glu Ala Asp Phe Ile Val Trp Val Ala

210 215 220

Ser Met Pro Gln Gly Val Val Ile Asp Pro Ala Thr Leu Lys Gln Pro

225 230 235 240

Cys Val Leu Ile Asp Gly Gly Tyr Pro Lys Asn Leu Gly Ser Lys Val

245 250 255

Gln Gly Glu Gly Ile Tyr Val Leu Asn Gly Gly Val Val Glu His Cys

260 265 270

Phe Asp Ile Asp Trp Gln Ile Met Ser Leu Ala Glu Met Ala Arg Pro

275 280 285

Glu Arg Gln Met Phe Ala Cys Phe Ala Glu Ala Met Leu Leu Glu Phe

290 295 300

Glu Gly Trp His Thr Asn Phe Ser Trp Gly Arg Asn Gln Ile Thr Ile

305 310 315 320
 Glu Lys Met Glu Ala Ile Gly Glu Ala Ser Val Arg His Gly Phe Gln

 325 330 335
 Pro Leu Ala Leu Ala Ile
 340
 <210> 62
 <211> 342
 <212> PRT
 <213> Synechococcus elongatus
 <400> 62
 Met Ala Phe Gly Leu Ile Gly His Leu Thr Ser Leu Glu Gln Ala Arg
 1 5 10 15
 Asp Val Trp Arg Arg Met Gly Tyr Pro Glu Tyr Ala Asp Gln Gly Met
 20 25 30
 Glu Phe Trp Ser Ser Ala Pro Pro Gln Ile Val Asp Glu Ile Thr Val

 35 40 45
 Thr Ser Ala Thr Gly Lys Val Ile His Gly Arg Tyr Ile Glu Ser Gly
 50 55 60
 Phe Leu Pro Glu Met Leu Ala Ala Arg Arg Phe Lys Thr Ala Thr Arg
 65 70 75 80
 Lys Val Leu Asn Ala Met Ser His Ala Gln Lys His Gly Ile Asp Ile
 85 90 95
 Ser Ala Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asp Leu

 100 105 110
 Ala Lys Leu Arg Gln Val Arg Asp Thr Thr Leu Glu Phe Glu Arg Phe
 115 120 125
 Thr Thr Gly Asn Thr His Thr Ala Tyr Val Ile Cys Arg Gln Val Glu
 130 135 140
 Ala Ala Ala Lys Thr Leu Gly Ile Asp Ile Ala Gln Ala Thr Val Ala
 145 150 155 160
 Val Val Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu

165 170 175
 Asp Leu Lys Leu Gly Val Gly Asp Leu Ile Leu Thr Ala Arg Asn Gln
 180 185 190
 Glu Arg Leu Asp Asn Leu Gln Ala Glu Leu Gly Arg Gly Lys Ile Leu
 195 200 205
 Pro Leu Glu Ala Ala Leu Pro Glu Ala Asp Phe Ile Val Trp Val Ala
 210 215 220
 Ser Met Pro Gln Gly Val Val Ile Asp Pro Ala Thr Leu Lys Gln Pro

225 230 235 240
 Cys Val Leu Ile Asp Gly Gly Tyr Pro Lys Asn Leu Gly Ser Lys Val
 245 250 255
 Gln Gly Glu Gly Ile Tyr Val Leu Asn Gly Gly Val Val Glu His Cys
 260 265 270
 Phe Asp Ile Asp Trp Gln Ile Met Ser Leu Ala Glu Met Ala Arg Pro
 275 280 285
 Glu Arg Gln Met Phe Ala Cys Phe Ala Glu Ala Met Leu Leu Glu Phe

290 295 300
 Glu Gly Trp His Thr Asn Phe Ser Trp Gly Arg Asn Gln Ile Thr Ile
 305 310 315 320
 Glu Lys Met Glu Ala Ile Gly Glu Ala Ser Val Arg His Gly Phe Gln
 325 330 335
 Pro Leu Ala Leu Ala Ile
 340

<210> 63

<211> 342

<212> PRT

<213> *Synechococcus elongatus*

<400> 63

Met Ala Phe Gly Leu Ile Gly His Ala Thr Ser Leu Glu Gln Ala Arg

1 5 10 15
 Leu Val Trp Arg Arg Met Gly Tyr Val Glu Tyr Ala Asp Gln Gly Leu
 20 25 30

Glu Phe Trp Ser Ser Ala Pro Pro Gln Ile Val Asp Glu Ile Thr Val
 35 40 45
 Thr Ser Ala Thr Gly Lys Val Ile His Gly Arg Tyr Ile Glu Ser Gly
 50 55 60
 Phe Leu Pro Glu Met Leu Ala Ala Arg Arg Phe Lys Thr Ala Thr Arg

 65 70 75 80
 Lys Val Leu Asn Ala Met Ser His Ala Gln Lys His Gly Ile Asp Ile
 85 90 95
 Ser Ala Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asp Leu
 100 105 110
 Ala Lys Leu Arg Gln Val Arg Asp Thr Thr Leu Glu Phe Glu Arg Phe
 115 120 125
 Thr Thr Gly Asn Thr His Thr Ala Tyr Val Ile Cys Arg Gln Val Glu

 130 135 140
 Ala Ala Ala Lys Thr Leu Gly Ile Asp Ile Ala Gln Ala Thr Val Ala
 145 150 155 160
 Val Val Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu
 165 170 175
 Asp Leu Lys Leu Gly Val Gly Asp Leu Ile Leu Thr Ala Arg Asn Gln
 180 185 190
 Glu Arg Leu Asp Asn Leu Gln Ala Glu Leu Gly Arg Gly Lys Ile Leu

 195 200 205
 Pro Leu Glu Ala Ala Leu Pro Glu Ala Asp Phe Ile Val Trp Val Ala
 210 215 220
 Ser Met Pro Gln Gly Val Val Ile Asp Pro Ala Thr Leu Lys Gln Pro
 225 230 235 240
 Cys Val Leu Ile Asp Gly Gly Tyr Pro Lys Asn Leu Gly Ser Lys Val
 245 250 255
 Gln Gly Glu Gly Ile Tyr Val Leu Asn Gly Gly Val Val Glu His Cys

 260 265 270
 Phe Asp Ile Asp Trp Gln Ile Met Ser Leu Ala Glu Met Ala Arg Pro

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

325 330 335
 Pro Leu Ala Leu Ala Ile

340

<210> 64

<211> 342

<212> PRT

<213> Synechococcus elongatus

<400> 64

Met Ala Phe Gly Leu Ile Gly His Leu Thr Ser Leu Glu Gln Ala Arg
 1 5 10 15
 Asp Val Trp Arg Arg Met Gly Tyr Glu Glu Tyr Ala Asp Gln Gly Leu
 20 25 30
 Glu Phe Trp Ser Ser Ala Pro Pro Gln Ile Val Asp Glu Ile Thr Val

35 40 45
 Thr Ser Ala Thr Gly Lys Val Ile His Gly Arg Tyr Ile Glu Ser Gly
 50 55 60
 Phe Leu Pro Glu Met Leu Ala Ala Arg Arg Phe Lys Thr Ala Thr Arg
 65 70 75 80
 Lys Val Leu Asn Ala Met Ser His Ala Gln Lys His Gly Ile Asp Ile
 85 90 95
 Ser Ala Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asp Leu

100 105 110
 Ala Lys Leu Arg Gln Val Arg Asp Thr Thr Leu Glu Phe Glu Arg Phe
 115 120 125
 Thr Thr Gly Asn Thr His Thr Ala Tyr Val Ile Cys Arg Gln Val Glu
 130 135 140

Ala Ala Ala Lys Thr Leu Gly Ile Asp Ile Ala Gln Ala Thr Val Ala
145 150 155 160
Val Val Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu
165 170 175
Asp Leu Lys Leu Gly Val Gly Asp Leu Ile Leu Thr Ala Arg Asn Gln
180 185 190
Glu Arg Leu Asp Asn Leu Gln Ala Glu Leu Gly Arg Gly Lys Ile Leu
195 200 205
Pro Leu Glu Ala Ala Leu Pro Glu Ala Asp Phe Ile Val Trp Val Ala
210 215 220
Ser Met Pro Gln Gly Val Val Ile Asp Pro Ala Thr Leu Lys Gln Pro
225 230 235 240
Cys Val Leu Ile Asp Gly Gly Tyr Pro Lys Asn Leu Gly Ser Lys Val
245 250 255
Gln Gly Glu Gly Ile Tyr Val Leu Asn Gly Gly Val Val Glu His Cys
260 265 270
Phe Asp Ile Asp Trp Gln Ile Met Ser Leu Ala Glu Met Ala Arg Pro
275 280 285
Glu Arg Gln Met Phe Ala Cys Phe Ala Glu Ala Met Leu Leu Glu Phe
290 295 300
Glu Gly Trp His Thr Asn Phe Ser Trp Gly Arg Asn Gln Ile Thr Ile
305 310 315 320
Glu Lys Met Glu Ala Ile Gly Glu Ala Ser Val Arg His Gly Phe Gln
325 330 335
Pro Leu Ala Leu Ala Ile
340
<210> 65
<211> 347
<212> PRT
<213> Prochlorococcus marinus
<400> 65

Met Ala Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Asp Ala Lys

1 5 10 15
Arg Lys Ala Ser Leu Leu Gly Phe Asp His Ile Ala Asp Gly Asp Leu
20 25 30
Asp Val Trp Cys Thr Ala Pro Pro Gln Leu Val Glu Asn Val Glu Val
35 40 45
Lys Ser Ala Ile Gly Ile Ser Ile Glu Gly Ser Tyr Ile Glu Ser Cys
50 55 60
Phe Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val

65 70 75 80
Leu Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala
85 90 95
Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln
100 105 110
His Lys Gln Ile Arg Asn Thr Ser Leu Glu Trp Glu Arg Phe Thr Thr
115 120 125
Gly Asn Thr His Thr Ala Trp Val Ile Cys Arg Gln Leu Glu Met Asn

130 135 140
Ala Pro Lys Ile Gly Ile Asp Leu Lys Ser Ala Thr Val Ala Val Val
145 150 155 160
Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ile Asn
165 170 175
Lys Thr Gly Ile Gly Glu Leu Leu Leu Val Ala Arg Gln Lys Glu Pro
180 185 190
Leu Asp Ser Leu Gln Lys Glu Leu Asp Gly Gly Thr Ile Lys Asn Leu

195 200 205
Asp Glu Ala Leu Pro Glu Ala Asp Ile Val Val Trp Val Ala Ser Met
210 215 220
Pro Lys Thr Met Glu Ile Asp Ala Asn Asn Leu Lys Gln Pro Cys Leu
225 230 235 240
Met Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Glu Lys Phe Gln Gly

245 250 255
Asn Asn Ile His Val Val Lys Gly Gly Ile Val Arg Phe Phe Asn Asp

260 265 270
Ile Gly Trp Asn Met Met Glu Leu Ala Glu Met Gln Asn Pro Gln Arg

275 280 285
Glu Met Phe Ala Cys Phe Ala Glu Ala Met Ile Leu Glu Phe Glu Lys

290 295 300
Cys His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Ser Leu Glu Lys

305 310 315 320
Met Glu Phe Ile Gly Ala Ala Ser Val Lys His Gly Phe Ser Ala Ile

325 330 335
Gly Leu Asp Lys His Pro Lys Val Leu Ala Val

340 345

<210> 66

<211> 347

<212> PRT

<213> Prochlorococcus marinus

<400> 66

Met Ala Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Asp Ala Lys

1 5 10 15

Arg Lys Ala Ser Leu Leu Gly Phe Asp His Ile Ala Asp Gly Asp Leu

20 25 30

Asp Val Trp Cys Thr Ala Pro Pro Gln Leu Val Glu Asn Val Glu Val

35 40 45
Lys Ser Ala Ile Gly Ile Ser Ile Glu Gly Ser Tyr Ile Asp Ser Cys

50 55 60

Phe Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val

65 70 75 80

Leu Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala

85 90 95

Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln

100 105 110
 His Lys Gln Ile Arg Asn Thr Ser Leu Glu Trp Glu Arg Phe Thr Thr
 115 120 125
 Gly Asn Thr His Thr Ala Trp Val Ile Cys Arg Gln Leu Glu Met Asn
 130 135 140
 Ala Pro Lys Ile Gly Ile Asp Leu Lys Ser Ala Thr Val Ala Val Val
 145 150 155 160
 Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ile Asn

 165 170 175
 Lys Thr Gly Ile Gly Glu Leu Leu Leu Val Ala Arg Gln Lys Glu Pro
 180 185 190
 Leu Asp Ser Leu Gln Lys Glu Leu Asp Gly Gly Thr Ile Lys Asn Leu
 195 200 205
 Asp Glu Ala Leu Pro Glu Ala Asp Ile Val Val Trp Val Ala Ser Met
 210 215 220
 Pro Lys Thr Met Glu Ile Asp Ala Asn Asn Leu Lys Gln Pro Cys Leu

 225 230 235 240
 Met Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Glu Lys Phe Gln Gly
 245 250 255
 Asn Asn Ile His Val Val Lys Gly Gly Ile Val Arg Phe Phe Asn Asp
 260 265 270
 Ile Gly Trp Asn Met Met Glu Leu Ala Glu Met Gln Asn Pro Gln Arg
 275 280 285
 Glu Met Phe Ala Cys Phe Ala Glu Ala Met Ile Leu Glu Phe Glu Lys

 290 295 300
 Cys His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Ser Leu Glu Lys
 305 310 315 320
 Met Glu Phe Ile Gly Ala Ala Ser Val Lys His Gly Phe Ser Ala Ile
 325 330 335
 Gly Leu Asp Lys His Pro Lys Val Leu Ala Pro
 340 345

<210> 67

<211> 347

<212> PRT

<213> Prochlorococcus marinus

<400> 67

Met Ala Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Asp Ala Lys

1 5 10 15
Arg Lys Ala Ser Leu Leu Gly Phe Asp His Ile Ala Asp Gly Asp Leu

20 25 30
Asp Val Trp Cys Thr Ala Pro Pro Val Leu Val Glu Asn Val Glu Val

35 40 45
Lys Ser Ala Ile Gly Ile Ser Ile Glu Gly Ser Tyr Ile Asp Ser Cys

50 55 60
Phe Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val

65 70 75 80
Leu Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala

85 90 95
Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln

100 105 110
His Lys Gln Ile Arg Asn Thr Ser Leu Glu Trp Glu Arg Phe Thr Thr

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

130 135 140
Ala Pro Lys Ile Gly Ile Asp Leu Lys Ser Ala Thr Val Ala Val Val

145 150 155 160
Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ile Asn

165 170 175
Lys Thr Gly Ile Gly Glu Leu Leu Leu Val Ala Arg Gln Lys Glu Pro

180 185 190
Leu Asp Ser Leu Gln Lys Glu Leu Asp Gly Gly Thr Ile Lys Asn Leu

195 200 205
Asp Glu Ala Leu Pro Glu Ala Asp Ile Val Val Trp Val Ala Ser Met

210 215 220
 Pro Lys Thr Met Glu Ile Asp Ala Asn Asn Leu Lys Gln Pro Cys Leu
 225 230 235 240
 Met Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Glu Lys Phe Gln Gly
 245 250 255
 Asn Asn Ile His Val Val Lys Gly Gly Ile Val Arg Phe Phe Asn Asp

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

 325 330 335
 Gly Leu Asp Lys His Pro Lys Val Leu Ala Val
 340 345
 <210> 68
 <211> 347
 <212> PRT
 <213> Prochlorococcus marinus
 <400> 68
 Met Ala Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Asp Ala Lys
 1 5 10 15
 Arg Lys Ala Ser Leu Leu Gly Phe Asp His Ile Ala Asp Gly Asp Leu
 20 25 30
 Asp Val Trp Cys Thr Ala Pro Pro Gln Leu Val Glu Asn Val Glu Val

 35 40 45
 Lys Ser Ala Ile Gly Ile Ser Ile Glu Gly Ser Tyr Ile Asp Ser Cys
 50 55 60
 Phe Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val
 65 70 75 80

Leu Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala
 85 90 95
 Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln
 100 105 110
 His Lys Gln Ile Arg Asn Thr Ser Leu Glu Trp Glu Arg Phe Thr Thr
 115 120 125
 Gly Asn Thr His Thr Ala Trp Val Ile Cys Arg Gln Leu Glu Met Asn
 130 135 140
 Ala Pro Lys Ile Gly Ile Asp Leu Lys Ser Ala Thr Val Ala Val Val
 145 150 155 160
 Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ile Asn
 165 170 175
 Lys Thr Gly Ile Gly Glu Leu Leu Leu Val Ala Arg Gln Lys Glu Pro
 180 185 190
 Leu Asp Ser Leu Gln Lys Glu Leu Asp Gly Gly Thr Ile Lys Asn Leu
 195 200 205
 Asp Glu Ala Leu Pro Glu Ala Asp Ile Val Val Trp Val Ala Ser Met
 210 215 220
 Pro Lys Thr Met Glu Ile Asp Ala Asn Asn Leu Lys Gln Pro Cys Leu
 225 230 235 240
 Met Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Glu Lys Phe Gln Gly
 245 250 255
 Asn Asn Ile His Val Val Lys Gly Gly Ile Val Arg Phe Phe Asn Asp
 260 265 270
 Ile Gly Trp Asn Met Met Glu Leu Ala Glu Met Gln Asn Pro Gln Arg
 275 280 285
 Glu Met Phe Ala Cys Phe Ala Glu Ala Met Ile Leu Glu Phe Glu Lys
 290 295 300
 Cys His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Ser Leu Glu Lys
 305 310 315 320
 Met Glu Phe Ile Gly Ala Ala Ser Val Lys His Gly Phe Ser Ala Ile

325 330 335
 Gly Leu Asp Lys His Pro Lys Val Leu Arg Val
 340 345
 <210> 69
 <211> 347
 <212> PRT
 <213> *Prochlorococcus marinus*
 <400> 69
 Met Ala Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Asp Ala Lys

 1 5 10 15
 Arg Lys Ala Ser Leu Leu Gly Phe Asp His Ile Ala Asp Gly Asp Leu
 20 25 30
 Asp Val Trp Cys Thr Ala Pro Pro Gln Leu Val Glu Asn Val Glu Val
 35 40 45
 Lys Ser Ala Ile Gly Ile Ser Ile Glu Gly Ser Tyr Ile Asp Ser Cys
 50 55 60
 Phe Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val

 65 70 75 80
 Leu Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala
 85 90 95
 Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln
 100 105 110
 His Lys Gln Ile Arg Asn Thr Ser Leu Glu Trp Glu Arg Phe Thr Thr
 115 120 125
 Gly Asn Thr His Thr Ala Trp Val Ile Cys Arg Gln Leu Glu Met Asn

 130 135 140
 Ala Pro Lys Ile Gly Ile Asp Leu Lys Ser Ala Thr Val Ala Val Val
 145 150 155 160
 Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ile Asn
 165 170 175
 Lys Thr Gly Ile Gly Glu Leu Leu Leu Val Ala Arg Gln Lys Glu Pro
 180 185 190

Leu Asp Ser Leu Gln Lys Glu Leu Asp Gly Gly Thr Ile Lys Asn Leu

195 200 205
Asp Glu Ala Leu Pro Glu Ala Asp Ile Val Val Trp Val Ala Ser Met

210 215 220
Pro Lys Thr Met Glu Ile Asp Ala Asn Asn Leu Lys Gln Pro Cys Leu

225 230 235 240
Met Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Glu Lys Phe Gln Gly

245 250 255
Asn Asn Ile His Val Val Lys Gly Gly Ile Val Arg Phe Phe Asn Asp

260 265 270
Ile Gly Trp Asn Met Met Glu Leu Ala Glu Met Gln Asn Pro Gln Arg

275 280 285
Glu Met Phe Ala Cys Phe Ala Glu Ala Met Ile Leu Glu Phe Glu Lys

290 295 300
Cys His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Ser Leu Glu Lys

305 310 315 320
Met Glu Phe Ile Gly Ala Ala Ser Val Lys His Gly Phe Ser Ala Ile

325 330 335
Gly Leu Asp Lys His Pro Lys Val Ser Ala Val

340 345

<210> 70

<211> 347

<212> PRT

<213> Prochlorococcus marinus

<400> 70

Met Ala Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Asp Ala Lys
1 5 10 15

Arg Lys Ala Ser Leu Leu Gly Phe Asp His Ile Ala Asp Gly Asp Leu
20 25 30

Asp Val Trp Cys Thr Ala Pro Pro Gln Leu Val Glu Asn Val Glu Val

35 40 45

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

290 295 300
 Cys His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Ser Leu Glu Lys
 305 310 315 320
 Met Glu Phe Ile Gly Ala Ala Ser Val Lys His Gly Phe Ser Ala Ile
 325 330 335
 Gly Leu Asp Lys His Pro Lys Val Leu Ala Gly
 340 345
 <210> 71
 <211> 347
 <212> PRT
 <213> Prochlorococcus marinus
 <400> 71
 Met Ala Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Asp Ala Lys

 1 5 10 15
 Arg Lys Ala Ser Leu Leu Gly Phe Asp His Ile Ala Asp Gly Asp Leu
 20 25 30
 Asp Val Trp Cys Thr Ala Pro Pro Gln Leu Val Glu Asn Val Glu Val
 35 40 45
 Lys Ser Ala Ile Gly Ile Ser Ile Glu Gly Ser Tyr Ile Asp Ser Cys
 50 55 60
 Phe Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val

 65 70 75 80
 Leu Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala
 85 90 95
 Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln
 100 105 110
 His Lys Gln Ile Arg Asn Thr Ser Leu Glu Trp Glu Arg Phe Thr Thr
 115 120 125
 Gly Asn Thr His Thr Ala Trp Val Ile Cys Arg Gln Leu Glu Met Asn

 130 135 140
 Ala Pro Lys Ile Gly Ile Asp Leu Lys Ser Ala Thr Val Ala Val Val
 145 150 155 160

Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ile Asn
165 170 175
Lys Thr Gly Ile Gly Glu Leu Leu Leu Val Ala Arg Gln Lys Glu Pro
180 185 190
Leu Asp Ser Leu Gln Lys Glu Leu Asp Gly Gly Thr Ile Lys Asn Leu
195 200 205
Asp Glu Ala Leu Pro Glu Ala Asp Ile Val Val Trp Val Ala Ser Met
210 215 220
Pro Lys Thr Met Glu Ile Asp Ala Asn Asn Leu Lys Gln Pro Cys Leu
225 230 235 240
Met Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Glu Lys Phe Gln Gly
245 250 255
Asn Asn Ile His Val Val Lys Gly Gly Ile Val Arg Phe Phe Asn Asp
260 265 270
Ile Gly Trp Asn Met Met Glu Leu Ala Glu Met Gln Asn Pro Gln Arg
275 280 285
Glu Met Phe Ala Cys Phe Ala Glu Ala Met Ile Leu Glu Phe Glu Lys
290 295 300
Cys His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Ser Leu Glu Lys
305 310 315 320
Met Glu Phe Ile Gly Ala Ala Ser Val Lys His Gly Phe Ser Ala Ile
325 330 335
Gly Leu Asp Lys His Pro Lys Val Asp Ala Val
340 345
<210> 72
<211> 347
<212> PRT
<213> Prochlorococcus marinus
<400> 72
Met Ala Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Asp Ala Lys
1 5 10 15
Arg Lys Ala Ser Leu Leu Gly Phe Asp His Ile Ala Asp Gly Asp Leu

20 25 30
 Asp Val Trp Cys Thr Ala Pro Pro Gln Leu Val Glu Asn Val Glu Val

 35 40 45
 Lys Ser Ala Ile Val Ile Ser Ile Glu Gly Ser Tyr Ile Asp Ser Cys
 50 55 60
 Phe Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val
 65 70 75 80
 Leu Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala
 85 90 95
 Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln

 100 105 110
 His Lys Gln Ile Arg Asn Thr Ser Leu Glu Trp Glu Arg Phe Thr Thr
 115 120 125
 Gly Asn Thr His Thr Ala Trp Val Ile Cys Arg Gln Leu Glu Met Asn
 130 135 140
 Ala Pro Lys Ile Gly Ile Asp Leu Lys Ser Ala Thr Val Ala Val Val
 145 150 155 160
 Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ile Asn

 165 170 175
 Lys Thr Gly Ile Gly Glu Leu Leu Leu Val Ala Arg Gln Lys Glu Pro
 180 185 190
 Leu Asp Ser Leu Gln Lys Glu Leu Asp Gly Gly Thr Ile Lys Asn Leu
 195 200 205
 Asp Glu Ala Leu Pro Glu Ala Asp Ile Val Val Trp Val Ala Ser Met
 210 215 220
 Pro Lys Thr Met Glu Ile Asp Ala Asn Asn Leu Lys Gln Pro Cys Leu

 225 230 235 240
 Met Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Glu Lys Phe Gln Gly
 245 250 255
 Asn Asn Ile His Val Val Lys Gly Gly Ile Val Arg Phe Phe Asn Asp
 260 265 270

Ile Gly Trp Asn Met Met Glu Leu Ala Glu Met Gln Asn Pro Gln Arg
275 280 285
Glu Met Phe Ala Cys Phe Ala Glu Ala Met Ile Leu Glu Phe Glu Lys
290 295 300
Cys His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Ser Leu Glu Lys
305 310 315 320
Met Glu Phe Ile Gly Ala Ala Ser Val Lys His Gly Phe Ser Ala Ile
325 330 335
Gly Leu Asp Lys His Pro Lys Val Leu Ala Val
340 345
<210> 73
<211> 347
<212> PRT
<213> Prochlorococcus marinus
<400> 73
Met Ala Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Asp Ala Lys
1 5 10 15
Arg Lys Ala Ser Leu Leu Gly Phe Asp His Ile Ala Asp Gly Asp Leu
20 25 30
Asp Val Trp Cys Thr Ala Pro Pro Gln Leu Val Glu Asn Val Glu Val
35 40 45
Lys Ser Ala Ile Gly Ile Ser Ile Glu Gly Ser Tyr Ile Asp Ser Cys
50 55 60
Phe Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val
65 70 75 80
Leu Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala
85 90 95
Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln
100 105 110
His Lys Gln Ile Arg Asn Thr Ser Leu Glu Trp Glu Arg Phe Thr Thr
115 120 125

Gly Asn Thr His Thr Ala Trp Val Ile Cys Arg Gln Leu Glu Met Asn

130 135 140

Ala Pro Lys Ile Gly Ile Asp Leu Lys Ser Ala Thr Val Ala Val Val

145 150 155 160

Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ile Asn

165 170 175

Lys Thr Gly Ile Gly Glu Leu Leu Leu Val Ala Arg Gln Lys Glu Pro

180 185 190

Leu Asp Ser Leu Gln Lys Glu Leu Asp Gly Gly Thr Ile Lys Asn Leu

195 200 205

Asp Glu Ala Leu Pro Glu Ala Asp Ile Val Val Trp Val Ala Ser Met

210 215 220

Pro Lys Thr Met Glu Ile Asp Ala Asn Asn Leu Lys Gln Pro Cys Leu

225 230 235 240

Met Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Glu Lys Phe Gln Gly

245 250 255

Asn Asn Ile His Val Val Lys Gly Gly Ile Val Arg Phe Phe Asn Asp

260 265 270

Ile Gly Trp Asn Met Met Glu Leu Ala Glu Met Gln Asn Pro Gln Arg

275 280 285

Glu Met Phe Ala Cys Phe Ala Glu Ala Met Ile Leu Glu Phe Glu Lys

290 295 300

Cys His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Ser Leu Glu Lys

305 310 315 320

Met Glu Phe Ile Gly Ala Ala Ser Val Lys His Gly Phe Ser Ala Ile

325 330 335

Gly Leu Asp Lys His Pro Lys Val Thr Ala Val

340 345

<210> 74

<211> 347

<212> PRT

<213> *Prochlorococcus marinus*

<400> 74

Met Ala Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Asp Ala Lys

1 5 10 15

Arg Lys Ala Ser Leu Leu Gly Phe Asp His Ile Ala Asp Gly Asp Leu

20 25 30

Asp Val Trp Cys Thr Ala Pro Pro Gln Leu Val Glu Asn Val Glu Val

35 40 45

Lys Ser Ala Ile Gly Ile Ser Ile Glu Gly Ser Tyr Ile Asp Ser Cys

50 55 60

Phe Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val

65 70 75 80

Leu Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala

85 90 95

Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln

100 105 110

His Lys Gln Ile Arg Asn Thr Ser Leu Glu Trp Glu Arg Phe Thr Thr

115 120 125

Gly Asn Thr His Thr Ala Trp Val Ile Cys Arg Gln Leu Glu Met Asn

130 135 140

Ala Pro Lys Ile Gly Ile Asp Leu Lys Ser Ala Thr Val Ala Val Val

145 150 155 160

Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ile Asn

165 170 175

Lys Thr Gly Ile Gly Glu Leu Leu Leu Val Ala Arg Gln Lys Glu Pro

180 185 190

Leu Asp Ser Leu Gln Lys Glu Leu Asp Gly Gly Thr Ile Lys Asn Leu

195 200 205

Asp Glu Ala Leu Pro Glu Ala Asp Ile Val Val Trp Val Ala Ser Met

210 215 220

Pro Lys Thr Met Glu Ile Asp Ala Asn Asn Leu Lys Gln Pro Cys Leu

225 230 235 240
Met Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Glu Lys Phe Gln Gly
 245 250 255
Asn Asn Ile His Val Val Lys Gly Gly Ile Val Arg Phe Phe Asn Asp
 260 265 270
Ile Gly Trp Asn Met Met Glu Leu Ala Glu Met Gln Asn Pro Gln Arg
 275 280 285
Glu Met Phe Ala Cys Phe Ala Glu Ala Met Ile Leu Glu Phe Glu Gly

 290 295 300
Cys His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Ser Leu Glu Lys
305 310 315 320
Met Glu Phe Ile Gly Ala Ala Ser Val Lys His Gly Phe Ser Ala Ile
 325 330 335
Gly Leu Asp Lys His Pro Lys Val Leu Ala Val
 340 345

<210> 75

<211> 347

<212> PRT

<213> Prochlorococcus marinus

<400> 75

Met Ala Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Asp Ala Lys

1 5 10 15
Arg Lys Ala Ser Leu Leu Gly Phe Asp His Ile Ala Asp Gly Asp Leu
 20 25 30
Asp Val Trp Cys Thr Ala Pro Pro Gln Leu Val Glu Asn Val Glu Val
 35 40 45
Lys Ser Ala Ile Gly Ile Ser Ile Glu Gly Ser Tyr Ile Asp Ser Cys
 50 55 60
Phe Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val

65 70 75 80
Leu Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala
 85 90 95

Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln
 100 105 110
 His Lys Gln Ile Arg Asn Thr Ser Leu Glu Trp Glu Arg Phe Thr Thr
 115 120 125
 Gly Asn Thr His Thr Ala Trp Val Ile Cys Arg Gln Leu Glu Met Asn
 130 135 140
 Ala Pro Lys Ile Gly Ile Asp Leu Lys Ser Ala Thr Val Ala Val Val
 145 150 155 160
 Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ile Asn
 165 170 175
 Lys Thr Gly Ile Gly Glu Leu Leu Leu Val Ala Arg Gln Lys Glu Pro
 180 185 190
 Leu Asp Ser Leu Gln Lys Glu Leu Asp Gly Gly Thr Ile Lys Asn Leu
 195 200 205
 Asp Glu Ala Leu Pro Glu Ala Asp Ile Val Val Trp Val Ala Ser Met
 210 215 220
 Pro Lys Thr Met Glu Ile Asp Ala Asn Asn Leu Lys Gln Pro Cys Leu
 225 230 235 240
 Met Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Glu Lys Phe Gln Gly
 245 250 255
 Asn Asn Ile His Val Val Lys Gly Gly Ile Val Arg Phe Phe Asn Asp
 260 265 270
 Ile Gly Trp Asn Met Met Glu Leu Ala Glu Met Gln Asn Pro Gln Arg
 275 280 285
 Glu Met Phe Ala Cys Phe Ala Glu Ala Met Ile Leu Glu Phe Glu Lys
 290 295 300
 Cys His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Ser Leu Glu Lys
 305 310 315 320
 Met Glu Phe Ile Gly Ala Ala Ser Val Lys His Gly Phe Ser Ala Ile
 325 330 335
 Gly Leu Asp Lys His Pro Lys Val Ala Ala Val

340 345

<210> 76

<211> 347

<212> PRT

<213> *Prochlorococcus marinus*

<400> 76

Met Ala Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Asp Ala Lys

1 5 10 15

Arg Lys Ala Ser Leu Leu Gly Phe Asp His Ile Ala Asp Gly Asp Leu

20 25 30

Asp Val Trp Cys Thr Ala Pro Pro Gln Leu Val Glu Asn Val Glu Val

35 40 45

Lys Ser Ala Ile Gly Ile Ser Ile Glu Gly Ser Tyr Ile Asp Ser Cys

50 55 60

Phe Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val

65 70 75 80

Leu Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala

85 90 95

Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln

100 105 110

His Lys Gln Ile Arg Asn Thr Ser Leu Glu Trp Glu Arg Phe Thr Thr

115 120 125

Gly Asn Thr His Thr Ala Trp Val Ile Cys Arg Gln Leu Glu Met Asn

130 135 140

Ala Pro Lys Ile Gly Ile Asp Leu Lys Ser Ala Thr Val Ala Val Val

145 150 155 160

Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ile Asn

165 170 175

Lys Thr Gly Ile Gly Glu Leu Leu Leu Val Ala Arg Gln Lys Glu Pro

180 185 190

Leu Asp Ser Leu Gln Lys Glu Leu Asp Gly Gly Thr Ile Lys Asn Leu

195 200 205

Asp Glu Ala Leu Pro Glu Ala Asp Ile Val Val Trp Val Ala Ser Met
 210 215 220
 Pro Lys Thr Met Glu Ile Asp Ala Asn Asn Leu Lys Gln Pro Cys Leu

 225 230 235 240
 Met Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Glu Lys Phe Gln Gly
 245 250 255
 Asn Asn Ile His Val Val Lys Gly Gly Ile Val Arg Phe Phe Asn Asp
 260 265 270
 Ile Gly Trp Asn Met Met Glu Leu Ala Glu Met Gln Asn Pro Gln Arg
 275 280 285
 Glu Met Phe Ala Cys Phe Ala Glu Ala Met Ile Leu Glu Phe Glu Lys

 290 295 300
 Cys His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Ser Leu Glu Lys
 305 310 315 320
 Met Glu Phe Ile Gly Ala Ala Ser Val Lys His Gly Phe Ser Ala Ile
 325 330 335
 Gly Leu Asp Lys Pro Pro Lys Val Leu Ala Val
 340 345
 <210> 77
 <211> 347
 <212> PRT
 <213> Prochlorococcus marinus
 <400> 77
 Met Ala Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Asp Ala Lys

 1 5 10 15
 Arg Lys Ala Ser Leu Leu Gly Phe Asp His Ile Ala Asp Gly Asp Leu
 20 25 30
 Asp Val Trp Cys Thr Ala Pro Pro Gln Leu Val Glu Asn Val Glu Val
 35 40 45
 Lys Ser Ala Ile Gly Ile Ser Ile Glu Gly Ser Tyr Ile Asp Ser Cys
 50 55 60

Phe Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val

65 70 75 80

Leu Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala

85 90 95

Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln

100 105 110

His Lys Gln Ile Arg Asn Thr Ser Leu Glu Trp Glu Arg Phe Thr Thr

115 120 125

Gly Asn Thr His Thr Ala Trp Val Ile Cys Arg Gln Leu Glu Met Asn

130 135 140

Ala Pro Lys Ile Gly Ile Asp Leu Lys Ser Ala Thr Val Ala Val Val

145 150 155 160

Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ile Asn

165 170 175

Lys Thr Gly Ile Gly Glu Leu Leu Leu Val Ala Arg Gln Lys Glu Pro

180 185 190

Leu Asp Ser Leu Gln Lys Glu Leu Asp Gly Gly Thr Ile Lys Asn Leu

195 200 205

Asp Glu Ala Leu Pro Glu Ala Asp Ile Val Val Trp Val Ala Ser Met

210 215 220

Pro Lys Thr Met Glu Ile Asp Ala Asn Asn Leu Lys Gln Pro Cys Leu

225 230 235 240

Met Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Glu Lys Phe Gln Gly

245 250 255

Asn Asn Ile His Val Val Lys Gly Gly Ile Val Arg Phe Phe Asn Asp

260 265 270

Ile Glu Trp Asn Met Met Glu Leu Ala Glu Met Gln Asn Pro Gln Arg

275 280 285

Glu Met Phe Ala Cys Phe Ala Glu Ala Met Ile Leu Glu Phe Glu Lys

290 295 300

Cys His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Ser Leu Glu Lys

Met Glu Phe Ile Gly Ala Ala Ser Val Lys His Gly Phe Ser Ala Ile

325 330 335

Gly Leu Asp Lys His Pro Lys Val Leu Ala Val

340 345

<210> 78

<211> 345

<212> PRT

<213> Prochlorococcus marinus

<400> 78

Met Ala Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Asp Ala Lys

1 5 10 15

Arg Lys Ala Ser Leu Leu Gly Phe Asp His Ile Ala Asp Gly Asp Leu

20 25 30

Asp Val Trp Cys Thr Ala Pro Pro Gln Leu Val Glu Asn Val Glu Val

Lys Ser Ala Ile Gly Ile Ser Ile Glu Gly Ser Tyr Ile Asp Ser Cys

50 55 60

Phe Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val

65 70 75 80

Leu Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala

85 90 95

Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln

100 105 110

His Lys Gln Ile Arg Asn Thr Ser Leu Glu Trp Glu Arg Phe Thr Thr

115 120 125

Gly Asn Thr His Thr Ala Trp Val Ile Cys Arg Gln Leu Glu Met Asn

130 135 140

Ala Pro Lys Ile Gly Ile Asp Leu Lys Ser Ala Thr Val Ala Val Val

145 150 155 160

Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ile Asn

165	170	175	
Lys Thr Gly Ile Gly Glu Leu Leu Leu Val Ala Arg Gln Lys Glu Pro			
180	185	190	
Leu Asp Ser Leu Gln Lys Glu Leu Asp Gly Gly Thr Ile Lys Asn Leu			
195	200	205	
Asp Glu Ala Leu Pro Glu Ala Asp Ile Val Val Trp Val Ala Ser Met			
210	215	220	
Pro Lys Thr Met Glu Ile Asp Ala Asn Asn Leu Lys Gln Pro Cys Leu			
225	230	235	240
Met Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Glu Lys Phe Gln Gly			
245	250	255	
Asn Asn Ile His Val Val Lys Gly Gly Ile Val Arg Phe Phe Asn Asp			
260	265	270	
Ile Gly Trp Asn Met Met Glu Leu Ala Glu Met Gln Asn Pro Gln Arg			
275	280	285	
Glu Met Phe Ala Cys Phe Ala Glu Ala Met Ile Leu Glu Phe Glu Lys			
290	295	300	
Cys His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Ser Leu Glu Lys			
305	310	315	320
Met Glu Phe Ile Gly Ala Ala Ser Val Lys His Gly Phe Ser Ala Ile			
325	330	335	
Gly Leu Asp Lys His Pro Lys Val Leu			
340	345		