



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0095226
(43) 공개일자 2021년07월30일

(51) 국제특허분류(Int. Cl.)
C12N 9/02 (2006.01) C12N 15/63 (2006.01)
C12N 9/16 (2006.01) C12P 7/04 (2006.01)
C12P 7/64 (2006.01)
(52) CPC특허분류
C12N 9/0008 (2013.01)
C12N 15/63 (2013.01)
(21) 출원번호 10-2021-7023277 (분할)
(22) 출원일자(국제) 2013년04월02일
심사청구일자 없음
(62) 원출원 특허 10-2020-7009760
원출원일자(국제) 2013년04월02일
심사청구일자 2020년05월04일
(85) 번역문제출일자 2021년07월22일
(86) 국제출원번호 PCT/US2013/035040
(87) 국제공개번호 WO 2013/152052
국제공개일자 2013년10월10일
(30) 우선권주장
61/619,309 2012년04월02일 미국(US)

(71) 출원인
게노마티카 인코포레이티드
미국 92121 캘리포니아주 샌디에고 텍서스 센터
드라이브 4757
(72) 발명자
그린필드, 데렉 엘.
미국 캘리포니아 94080 사우스 샌프란시스코 게이
트웨이 블러바드 600
클라크, 엘리자베스, 제이.
미국 캘리포니아 94080 사우스 샌프란시스코 게이
트웨이 블러바드 600
(74) 대리인
(뒷면에 계속)
특허법인(유)화우

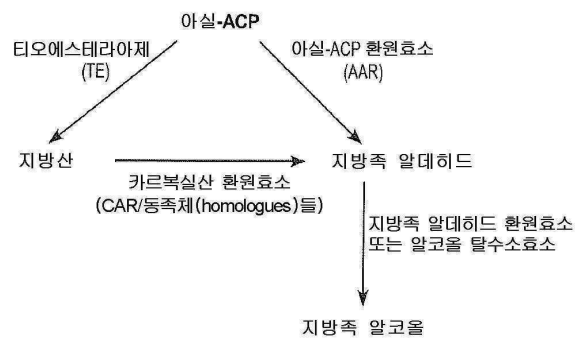
전체 청구항 수 : 총 57 항

(54) 발명의 명칭 C A R 효소들 및 지방족 알코올들의 개선된 생산

(57) 요약

본 발명은 재조합 숙주 세포들 내에서 지방족 알코올의 개선된 생산을 위한 변형 카르복실산 환원효소(CAR)에 관한 것이다.

대표도 - 도4



(52) CPC특허분류

C12N 9/16 (2013.01)

C12P 7/04 (2013.01)

C12P 7/6409 (2013.01)

C12Y 102/99006 (2013.01)

C12P 2203/00 (2013.01)

Y02P 20/52 (2015.11)

(72) 발명자

그로반, 엘리, 에스.

미국 캘리포니아 94080 사우스 샌프란시스코 게이
트웨이 블러바드 600

알라가다, 비크란트

미국 캘리포니아 94080 사우스 샌프란시스코 게이
트웨이 블러바드 600

이, 성원

미국 캘리포니아 94080 사우스 샌프란시스코 게이
트웨이 블러바드 600

리, 웨지

미국 캘리포니아 94080 사우스 샌프란시스코 게이
트웨이 블러바드 600

호, 지하오

미국 캘리포니아 94080 사우스 샌프란시스코 게이
트웨이 블러바드 600

주, 바오롱

미국 캘리포니아 94080 사우스 샌프란시스코 게이
트웨이 블러바드 600

명세서

청구범위

청구항 1

SEQ ID NO: 7에 대하여 적어도 약 90 % 서열 동일성을 가지는 아미노산 서열을 포함하는 변형(variant) 카르복실산 환원효소(carboxylic acid reductase: CAR) 폴리펩티드에 있어서,

여기서 상기 변형 CAR 폴리펩티드는 아미노산 위치들 3, 18, 20, 22, 80, 87, 191, 288, 473, 535, 750, 827, 870, 873, 926, 927, 930 및 1128로 이루어진 그룹으로부터 선택되는 아미노산 위치에 적어도 하나의 돌연변이(mutation)를 갖도록 유전적으로 조작되는(genetically engineered), 변형 CAR 폴리펩티드.

청구항 2

제 1 항에 있어서,

제조합 숙주 세포 내에서 상기 변형 CAR 폴리펩티드는 대응하는 야생형 폴리펩티드를 발현시키는 제조합 숙주 세포와 비교하여 지방족 알코올 조성물들의 더 높은 역가(titer)를 야기하는, 변형 CAR 폴리펩티드.

청구항 3

제 1 항에 있어서,

상기 CAR 폴리펩티드는 CarB 폴리펩티드인, 변형 CAR 폴리펩티드.

청구항 4

제 1 항에 있어서,

상기 변형 CAR 폴리펩티드는 S3R, D18R, D18L, D18T, D18P, E20V, E20S, E20R, S22R, S22N, S22G, L80R, R87G, R87E, V191S, F288R, F288S, F288G, Q473L, Q473W, Q473Y, Q473I, Q473H, A535S, D750A, R827C, R827A, I870L, R873S, V926A, V926E, S927K, S927G, M930K, M930R 및 L1128W로 이루어진 그룹으로부터 선택되는 돌연변이를 포함하는, 변형 CAR 폴리펩티드.

청구항 5

제 4 항에 있어서,

상기 변형 CAR 폴리펩티드는 돌연변이 A535S를 포함하는, 변형 CAR 폴리펩티드.

청구항 6

제 4 항에 있어서,

상기 변형 폴리펩티드는 돌연변이들 E20R, F288G, Q473I 및 A535S를 포함하는, 변형 CAR 폴리펩티드.

청구항 7

제 4 항에 있어서,

상기 변형 폴리펩티드는 돌연변이들 E20R, F288G, Q473H, A535S, R827A 및 S927G를 포함하는, 변형 CAR 폴리펩티드.

청구항 8

제 4 항에 있어서,

상기 변형 폴리펩티드는 돌연변이들 E20R, S22R, F288G, Q473H, A535S, R827A 및 S927G를 포함하는, 변형 CAR 폴리펩티드.

청구항 9

제 4 항에 있어서,

상기 변형 폴리펩티드는 돌연변이들 S3R, E20R, S22R, F288G, Q473H, A535S, R873S, S927G, M930R 및 L1128W를 포함하는, 변형 CAR 폴리펩티드.

청구항 10

제 4 항에 있어서,

상기 변형 폴리펩티드는 돌연변이들 E20R, S22R, F288G, Q473H, A535S, R873S, S927G, M930R 및 L1128W를 포함하는, 변형 CAR 폴리펩티드.

청구항 11

제 4 항에 있어서,

상기 변형 폴리펩티드는 돌연변이들 D18R, E20R, S22R, F288G, Q473I, A535S, S927G, M930K 및 L1128W를 포함하는, 변형 CAR 폴리펩티드.

청구항 12

제 4 항에 있어서,

상기 변형 폴리펩티드는 돌연변이들 E20R, S22R, F288G, Q473I, A535S, R827C, V926E, S927K 및 M930R을 포함하는, 변형 CAR 폴리펩티드.

청구항 13

제 4 항에 있어서,

상기 변형 폴리펩티드는 돌연변이들 D18R, E20R, F288G, Q473I, A535S, R827C, V926E, M930K 및 L1128W를 포함하는, 변형 CAR 폴리펩티드.

청구항 14

제 4 항에 있어서,

상기 변형 폴리펩티드는 돌연변이들 E20R, S22R, F288G, Q473H, A535S, R827C, V926A, S927K 및 M930R을 포함하는, 변형 CAR 폴리펩티드.

청구항 15

제 4 항에 있어서,

상기 변형 폴리펩티드는 돌연변이들 E20R, S22R, F288G, Q473H, A535S 및 R827C를 포함하는, 변형 CAR 폴리펩티드.

청구항 16

제 4 항에 있어서,

상기 변형 폴리펩티드는 돌연변이들 E20R, S22R, F288G, Q473I, A535S, R827C 및 M930R을 포함하는, 변형 CAR 폴리펩티드.

청구항 17

제 4 항에 있어서,

상기 변형 폴리펩티드는 돌연변이들 E20R, S22R, F288G, Q473I, A535S, I870L, S927G 및 M930R을 포함하는, 변형 CAR 폴리펩티드.

청구항 18

제 4 항에 있어서,

상기 변형 폴리펩티드는 돌연변이들 E20R, S22R, F288G, Q473I, A535S, I870L 및 S927G를 포함하는, 변형 CAR 폴리펩티드.

청구항 19

제 4 항에 있어서,

상기 변형 폴리펩티드는 돌연변이들 D18R, E20R, S22R, F288G, Q473I, A535S, R827C, I870L, V926A 및 S927G를 포함하는, 변형 CAR 폴리펩티드.

청구항 20

제 4 항에 있어서,

상기 변형 폴리펩티드는 돌연변이들 E20R, S22R, F288G, Q473H, A535S, R827C, I870L 및 L1128W를 포함하는, 변형 CAR 폴리펩티드.

청구항 21

제 4 항에 있어서,

상기 변형 폴리펩티드는 돌연변이들 D18R, E20R, S22R, F288G, Q473H, A535S, R827C, I870L, S927G 및 L1128W를 포함하는, 변형 CAR 폴리펩티드.

청구항 22

제 4 항에 있어서,

상기 변형 폴리펩티드는 돌연변이들 E20R, S22R, F288G, Q473I, A535S, R827C, I870L, S927G 및 L1128W를 포함하는, 변형 CAR 폴리펩티드.

청구항 23

제 4 항에 있어서,

상기 변형 폴리펩티드는 돌연변이들 E20R, S22R, F288G, Q473I, A535S, R827C, I870L, S927G, M930K 및 L1128W를 포함하는, 변형 CAR 폴리펩티드.

청구항 24

제 4 항에 있어서,

상기 변형 폴리펩티드는 돌연변이들 E20R, S22R, F288G, Q473H, A535S, I870L, S927G 및 M930K를 포함하는, 변형 CAR 폴리펩티드.

청구항 25

제 4 항에 있어서,

상기 변형 폴리펩티드는 돌연변이들 E20R, F288G, Q473I, A535S, I870L, M930K를 포함하는, 변형 CAR 폴리펩티드.

청구항 26

제 4 항에 있어서,

상기 변형 폴리펩티드는 돌연변이들 E20R, S22R, F288G, Q473H, A535S, S927G, M930K 및 L1128W를 포함하는, 변형 CAR 폴리펩티드.

청구항 27

제 4 항에 있어서,

상기 변형 폴리펩티드는 돌연변이들 D18R, E20R, S22R, F288G, Q473I, A535S, S927G 및 L1128W를 포함하는, 변형 CAR 폴리펩티드.

청구항 28

제 4 항에 있어서,

상기 변형 폴리펩티드는 돌연변이들 E20R, S22R, F288G, Q473I, A535S, R827C, I870L 및 S927G를 포함하는, 변형 CAR 폴리펩티드.

청구항 29

제 4 항에 있어서,

상기 변형 폴리펩티드는 돌연변이들 D18R, E20R, S22R, F288G, Q473I, A535S, R827C, I870L, S927G 및 L1128W를 포함하는, 변형 CAR 폴리펩티드.

청구항 30

제 4 항에 있어서,

상기 변형 폴리펩티드는 돌연변이들 D18R, E20R, S22R, F288G, Q473I, A535S, S927G, M930R 및 L1128W를 포함하는, 변형 CAR 폴리펩티드.

청구항 31

제 4 항에 있어서,

상기 변형 폴리펩티드는 돌연변이들 E20R, S22R, F288G, Q473H, A535S, V926E, S927G 및 M930R을 포함하는, 변형 CAR 폴리펩티드.

청구항 32

제 4 항에 있어서,

상기 변형 폴리펩티드는 돌연변이들 E20R, S22R, F288G, Q473H, A535S, R827C, I870L, V926A 및 L1128W를 포함하는, 변형 CAR 폴리펩티드.

청구항 33

아미노산 위치들 3, 18, 20, 22, 80, 87, 191, 288, 473, 535, 750, 827, 870, 873, 926, 927, 930 및 1128로 이루어진 그룹으로부터 선택되는 아미노산 위치에서 적어도 하나의 돌연변이를 가지고 SEQ ID NO: 7에 대하여 적어도 90 % 서열 동일성을 가지는 변형 카르복실산 환원효소(CAR) 폴리펩티드를 인코딩(encoding)하는 폴리뉴클레오티드를 포함하는 재조합 숙주 세포를 제공하고,

여기서 상기 유전적으로 조작된 숙주 세포는 상기 변형 CAR 폴리펩티드를 발현시키는데 효과적인 조건들 하에서 탄소 공급원(carbon source)을 함유하는 배지 내에서 배양되는 경우에 대응하는 야생형 CAR 폴리펩티드를 발현하는 숙주 세포보다 더 높은 역가 또는 수율(yield)로 지방족 알코올 조성물을 생산하는, 재조합 숙주 세포.

청구항 34

제 33 항에 있어서,

상기 SEQ ID NO: 7은 대응하는 야생형 CAR 폴리펩티드인, 재조합 숙주 세포.

청구항 35

제 33 항에 있어서,

티오에스테라아제(thioesterase) 폴리펩티드를 인코딩하는 폴리뉴클레오티드를 더 포함하는, 재조합 숙주 세포.

청구항 36

제 35 항에 있어서,

FabB 폴리펩티드 및 FadR 폴리펩티드를 인코딩하는 폴리뉴클레오티드를 더 포함하는, 재조합 숙주 세포.

청구항 37

제 33 항 내지 제 36 항 중 어느 한 항에 있어서,

상기 유전적으로 조작된 숙주 세포는 유전적으로 조작된 숙주 세포와 동일한 조건들 하에서 배양되는 경우에 대응하는 야생형 CAR 폴리펩티드를 발현시키는 숙주 세포의 역가보다 적어도 3 배 더 큰 역가를 갖는, 재조합 숙주 세포.

청구항 38

제 37 항에 있어서,

상기 재조합 숙주 세포는 약 30 g/L 내지 약 250 g/L의 역가를 갖는, 재조합 숙주 세포.

청구항 39

제 38 항에 있어서,

상기 재조합 숙주 세포는 약 90 g/L 내지 약 120 g/L의 역가를 갖는, 재조합 숙주 세포.

청구항 40

제 33 항 내지 제 36 항 중 어느 한 항에 있어서,

상기 유전적으로 조작된 숙주 세포는 유전적으로 조작된 숙주 세포와 동일한 조건들 하에서 배양되는 경우에 대응하는 야생형 CAR 폴리펩티드를 발현시키는 숙주 세포의 수율보다 적어도 3 배 더 큰 수율을 갖는, 재조합 숙주 세포.

청구항 41

제 40 항에 있어서,

상기 유전적으로 조작된 숙주 세포는 약 10 % 내지 약 40 %의 수율을 갖는, 재조합 숙주 세포.

청구항 42

제 33 항 내지 제 36 항 중 어느 한 항의 재조합 숙주 세포를 포함하는 세포 배양.

청구항 43

제 42 항에 있어서,

상기 세포 배양은 대응하는 야생형 CAR 폴리펩티드를 발현시키는 세포 배양의 생산물(productivity)보다 적어도 3 배 더 큰 생산물을 갖는, 세포 배양.

청구항 44

제 43 항에 있어서,

상기 생산물은 약 0.7 mg/L/hr 내지 약 3 g/L/hr의 범위인, 세포 배양.

청구항 45

제 44 항에 있어서,

배양 배지는 지방족 알코올 조성물을 포함하는, 세포 배양.

청구항 46

제 42 항 내지 제 45 항 중 어느 한 항에 있어서,

상기 지방족 알코올 조성물은 세포 외에서(extracellularly) 생산되는, 세포 배양.

청구항 47

제 46 항에 있어서,

상기 지방족 알코올 조성물은 C6, C8, C10, C12, C13, C14, C15, C16, C17 또는 C18 지방족 알코올 중 하나 이상을 포함하는, 세포 배양.

청구항 48

제 46 항에 있어서,

상기 지방족 알코올 조성물은 C10:1, C12:1, C14:1, C16:1 또는 C18:1 불포화 지방족 알코올을 포함하는, 세포 배양.

청구항 49

제 46 항에 있어서,

상기 지방족 알코올 조성물은 C₁₂ 및 C₁₄ 지방족 알코올들을 포함하는, 세포 배양.

청구항 50

제 49 항에 있어서,

상기 지방족 알코올 조성물은 C₁₂ 및 C₁₄ 지방족 알코올들을 약 3:1의 비율로 포함하는, 세포 배양.

청구항 51

제 46 항에 있어서,

상기 지방족 알코올 조성물은 불포화 지방족 알코올들을 포함하는, 세포 배양.

청구항 52

제 51 항에 있어서,

상기 지방족 알코올 조성물은 상기 지방족 알코올의 환원되는 단부(reduced end)로부터 탄소 사슬의 C₇과 C₈ 사이 7 위치에 이중 결합을 갖는 지방족 알코올을 포함하는, 세포 배양.

청구항 53

제 46 항에 있어서,

상기 지방족 알코올 조성물은 포화 지방족 알코올들을 포함하는, 세포 배양.

청구항 54

제 46 항에 있어서,

상기 지방족 알코올 조성물은 분지쇄형 지방족 알코올들을 포함하는, 세포 배양.

청구항 55

제 35 항에 있어서,

지방족 알데히드 환원효소를 인코딩하는 폴리뉴클레오티드(AlrA)를 더 포함하는, 제조함 숙주 세포.

청구항 56

제 55 항의 재조합 숙주 세포를 포함하는 세포 배양.

청구항 57

높은 역가, 수율 또는 생산물로 지방족 알코올 조성물을 만드는 방법에 있어서,

(a) 제 33 항 내지 제 41 항 중 어느 한 항의 재조합 숙주 세포를 조작(engineering)하는 단계;

(b) 상기 재조합 숙주 세포를 탄소 공급원을 포함하는 배지 내에서 배양하는 단계; 및

(c) 선택적으로 지방족 알코올 조성물을 상기 배지에서 분리시키는 단계를 포함하는, 높은 역가, 수율 또는 생산물로 지방족 알코올 조성물을 만드는 방법.

발명의 설명

기술 분야

[0001] 관련 출원들에 대한 상호 참조

[0002] 본 출원은 2012년 4월 2일에 출원된 미국 가출원 61/619,309의 이익을 주장하며, 이는 본 명세서에서 전문이 인용참조된다.

[0003] 서열 목록(SEQUENCE LISTING)

[0004] 본 출원은 EFS-Web을 통하여 ASCII 포맷으로 제출된 서열 목록을 포함하며, 이는 본 명세서에서 전문이 인용참조된다. 2013년 4월 2일에 작성된 상기 ASCII 복사본(copy)은 LS00039PCT_SL.txt로 파일명을 지정(named)하였고, 크기는 89,038 바이트(bytes)이다.

[0005] 본 발명의 분야

[0006] 본 발명은 재조합 숙주 세포들 내에서 지방족 알코올의 개선된 생산을 위한 변형(variant) 카르복실산 환원효소(carboxylic acid reductase: CAR)에 관한 것이다. 또한, 본 발명은 재조합 숙주 세포들 및 세포 배양들뿐만 아니라 CAR 핵산들 및 폴리펩티드들에 관한 것이다. 지방족 알코올 조성물들을 만드는 방법들도 더 포괄된다.

배경 기술

[0007] 지방족 알코올은 생화학 산업의 중요한 부분을 구성한다. 이러한 분자들 및 이들의 유도체들은 계면활성제, 윤활제, 가스제, 용제, 유화제, 연화제(emollients), 증점제, 향미제(flavors), 향료(fragrances) 및 연료로서의 이용을 포함하여 다수의 용도를 갖는다. 산업분야에서, 지방족 알코올은 야자유, 팜유(palm oil), 팜 핵유(palm kernel oil), 텔로(tallow) 및 라드(lard)와 같은 천연 원료들로부터 생산되는 지방산들의 촉매 수소화에 의해, 또는 석유화학 공급원료(feedstock)로부터 생산되는 알파-올레핀들의 화학적 수소화에 의해 생산된다. 천연 원료들로부터 유래되는 지방족 알코올들은 다양한 사슬 길이를 갖는다. 지방족 알코올들의 사슬 길이는 특정 적용에 대하여 중요하다. 사실상, 지방족 알코올들은 아실-ACP 또는 아실-CoA 분자들을 대용하는 일차 알코올로 환원시킬 수 있는 효소들에 의해서도 만들어진다(예를 들어, 본 명세서에서 인용참조되는 미국 특허 출원 Nos. 20100105955, 20100105963 및 20110250663을 참조).

[0008] 지방족 알코올을 생산하는 현재 기술들은 무기 촉매를 매개로 하는 지방산의 대용하는 알코올로의 환원을 포함하고, 이는 많은 비용이 들고, 시간 소모가 크며, 번거롭다. 이 과정에 사용되는 지방산들은 천연 원료들(예를 들어, 상기의 식물성유 및 동물성유 및 지방)로부터 유래된다. 지방족 알코올의 알파-올레핀으로의 탈수는 화학 촉매작용에 의해 달성될 수도 있다. 그러나, 이 기술은 재생 불가능하고, 높은 작동 비용 및 환경적으로 유해한 화학 폐기물과 관련된다. 따라서, 지방족 알코올 생산의 개선된 방법들이 필요하고, 본 발명은 이러한 필요성에 대해 다룬다.

발명의 내용

해결하려는 과제

[0009] 본 발명의 일 측면에서는 SEQ ID NO: 7에 대하여 적어도 약 90 % 서열 동일성을 가지는 아미노산 서열을 포함하

는 변형 카르복실산 환원효소(CAR) 폴리펩티드를 제공하고, 여기서 변형 CAR 폴리펩티드는 아미노산 위치들 3, 18, 20, 22, 80, 87, 191, 288, 473, 535, 750, 827, 870, 873, 926, 927, 930 및 1128의 그룹으로부터 선택되는 아미노산 위치에 적어도 하나의 돌연변이(mutation)를 갖도록 유전적으로 조작된다(genetically engineered). 본 명세서에서, 재조합 숙주 세포 내에서 변형 CAR 폴리펩티드는 대응하는 야생형 폴리펩티드를 발현시키는 재조합 숙주 세포와 비교하여 지방족 알코올 조성물들의 더 높은 역가(titer)를 야기한다. 관련된 측면에서, CAR 폴리펩티드는 CarB 폴리펩티드이다. 또 다른 관련된 측면에서, 변형 CAR 폴리펩티드는 위치들 S3R, D18R, D18L, D18T, D18P, E20V, E20S, E20R, S22R, S22N, S22G, L80R, R87G, R87E, V191S, F288R, F288S, F288G, Q473L, Q473W, Q473Y, Q473I, Q473H, A535S, D750A, R827C, R827A, I870L, R873S, V926A, V926E, S927K, S927G, M930K, M930R 및/또는 L1128W의 돌연변이를 포함한다. 관련된 측면에서, CAR 폴리펩티드는 돌연변이 A535S; 또는 돌연변이들 E20R, F288G, Q473I 및 A535S; 또는 돌연변이들 E20R, F288G, Q473H, A535S, R827A 및 S927G; 또는 돌연변이들 E20R, S22R, F288G, Q473H, A535S, R827A 및 S927G; 또는 돌연변이들 S3R, E20R, S22R, F288G, Q473H, A535S, R873S, S927G, M930R 및 L1128W; 또는 E20R, S22R, F288G, Q473H, A535S, R873S, S927G, M930R 및 L1128W; 또는 돌연변이들 D18R, E20R, S22R, F288G, Q473I, A535S, S927G, M930K 및 L1128W; 또는 돌연변이들 E20R, S22R, F288G, Q473I, A535S, R827C, V926E, S927K 및 M930R; 또는 돌연변이들 D18R, E20R, F288G, Q473I, A535S, R827C, V926E, M930K 및 L1128W; 또는 돌연변이들 E20R, S22R, F288G, Q473H, A535S, S927K 및 M930R; 또는 돌연변이들 E20R, S22R, F288G, Q473H, A535S 및 R827C; 또는 돌연변이들 E20R, S22R, F288G, Q473I, A535S, R827C 및 M930R; 또는 돌연변이들 E20R, S22R, F288G, Q473I, A535S, I870L, S927G 및 M930R; 또는 돌연변이들 E20R, S22R, F288G, Q473I, A535S, I870L 및 S927G; 또는 돌연변이들 D18R, E20R, S22R, F288G, Q473I, A535S, R827C, I870L, V926A 및 S927G; 또는 돌연변이들 E20R, S22R, F288G, Q473H, A535S, R827C, I870L 및 L1128W; 또는 돌연변이들 D18R, E20R, S22R, F288G, Q473H, A535S, R827C, I870L, S927G 및 L1128W; 또는 돌연변이들 E20R, S22R, F288G, Q473I, A535S, R827C, I870L, S927G 및 L1128W; 또는 돌연변이들 E20R, S22R, F288G, Q473I, A535S, R827C, I870L, S927G, M930K 및 L1128W; 또는 돌연변이들 E20R, S22R, F288G, Q473H, A535S, I870L, S927G 및 M930K; 또는 돌연변이들 E20R, F288G, Q473I, A535S, I870L, M930K; 또는 돌연변이들 E20R, S22R, F288G, Q473H, A535S, S927G, M930K 및 L1128W; 또는 돌연변이들 D18R, E20R, S22R, F288G, Q473I, A535S, S927G 및 L1128W; 또는 돌연변이들 E20R, S22R, F288G, Q473I, A535S, R827C, I870L 및 S927G; 또는 돌연변이들 D18R, E20R, S22R, F288G, Q473I, A535S, R827C, I870L, S927G 및 L1128W; 또는 돌연변이들 D18R, E20R, S22R, F288G, Q473I, A535S, S927G, M930R 및 L1128W; 또는 돌연변이들 E20R, S22R, F288G, Q473H, A535S, V926E, S927G 및 M930R; 또는 돌연변이들 E20R, S22R, F288G, Q473H, A535S, R827C, I870L, V926A 및 L1128W; 또는 이들의 조합들을 포함한다.

과제의 해결 수단

- [0010] 본 발명의 또 다른 측면에서는 아미노산 위치들 3, 18, 20, 22, 80, 87, 191, 288, 473, 535, 750, 827, 870, 873, 926, 927, 930 및 1128을 포함하는 아미노산 위치에서 적어도 하나의 돌연변이를 가지고 SEQ ID NO: 7에 대하여 적어도 90 % 서열 동일성을 가지는 변형 카르복실산 환원효소(CAR) 폴리펩티드를 인코딩(encoding)하는 폴리뉴클레오티드를 포함하는 숙주 세포를 제공하고, 여기서 유전적으로 조작된 숙주 세포는 변형 CAR 폴리펩티드를 발현시키는데 효과적인 조건들 하에서 탄소 공급원(carbon source)을 함유하는 배지 내에서 배양되는 경우에 대응하는 야생형 CAR 폴리펩티드를 발현하는 숙주 세포보다 더 높은 역가 또는 수율(yield)로 지방족 알코올 조성물을 생산하며, 여기서 SEQ ID NO: 7은 대응하는 야생형 CAR 폴리펩티드이다. 관련된 측면에서, 재조합 숙주 세포는 티오에스테라아제(thioesterase) 폴리펩티드를 인코딩하는 폴리뉴클레오티드를 더 포함한다. 또 다른 관련된 측면에서, 재조합 숙주 세포는 FabB 폴리펩티드 및 FadR 폴리펩티드를 인코딩하는 폴리뉴클레오티드를 더 포함한다. 또 다른 관련된 측면에서, 본 발명은 지방족 알데히드 환원효소를 인코딩하는 폴리뉴클레오티드(AlrA)를 포함하는 재조합 숙주 세포 및 이를 함유하는 세포 배양을 제공한다.
- [0011] 본 발명의 또 다른 측면에서는 재조합 숙주 세포를 제공하고, 여기서 유전적으로 조작된 숙주 세포는 유전적으로 조작된 숙주 세포와 동일한 조건들 하에서 배양되는 경우에 대응하는 야생형 CAR 폴리펩티드를 발현시키는 숙주 세포의 역가보다 적어도 3 배 더 큰 역가를 갖는다. 일 관련된 측면에서, 유전적으로 조작된 숙주 세포는 약 30 g/L 내지 약 250 g/L의 역가를 갖는다. 또 다른 관련된 측면에서, 유전적으로 조작된 숙주 세포는 약 90 g/L 내지 약 120 g/L의 역가를 갖는다.
- [0012] 본 발명의 또 다른 측면에서는 재조합 숙주 세포를 제공하고, 여기서 유전적으로 조작된 숙주 세포는 유전적으로 조작된 숙주 세포와 동일한 조건들 하에서 배양되는 경우에 대응하는 야생형 CAR 폴리펩티드를 발현시키는 숙주 세포의 수율보다 적어도 3 배 더 큰 수율을 갖는다. 일 관련된 측면에서, 유전적으로 조작된 숙주 세포는

약 10 % 내지 약 40 %의 수율을 갖는다.

[0013] 본 발명은 본 명세서에 설명된 바와 같은 재조합 숙주 세포를 포함하는 세포 배양을 더 포괄한다. 관련된 측면에서, 세포 배양은 대응하는 야생형 CAR 폴리펩티드를 발현시키는 세포 배양의 생산률(productivity)보다 적어도 약 3 배 더 큰 생산률을 갖는다. 또 다른 관련된 측면에서, 생산률은 약 0.7 mg/L/hr 내지 약 3 g/L/hr의 범위이다. 또 다른 관련된 측면에서, 배양 배지는 지방족 알코올 조성물을 포함한다. 지방족 알코올 조성물은 세포 외에서(extracellularly) 생산된다. 지방족 알코올 조성물은 C6, C8, C10, C12, C13, C14, C15, C16, C17 또는 C18 지방족 알코올 중 하나 이상; 또는 C10:1, C12:1, C14:1, C16:1 또는 C18:1 불포화 지방족 알코올을 포함할 수 있다. 또 다른 관련된 측면에서, 지방족 알코올 조성물은 C12 및 C14 지방족 알코올들을 포함한다. 또 다른 관련된 측면에서, 지방족 알코올 조성물은 C12 및 C14 지방족 알코올들을 약 3:1의 비율로 포함한다. 또 다른 관련된 측면에서, 지방족 알코올 조성물은 불포화 지방족 알코올들을 포함한다. 또한, 지방족 알코올 조성물은 지방족 알코올의 환원되는 단부(reduced end)로부터 탄소 사슬의 C7과 C8 사이 7 위치에 이중 결합을 갖는 지방족 알코올을 포함할 수 있다. 또 다른 측면에서, 지방족 알코올 조성물은 포화 지방족 알코올들을 포함한다. 또 다른 측면에서, 지방족 알코올 조성물은 분지쇄형 지방족 알코올들을 포함한다.

[0014] 추가적으로, 본 발명은 높은 역가, 수율 또는 생산물로 지방족 알코올 조성물을 만드는 방법을 고려하며, 상기 방법은 재조합 숙주 세포를 조작(engineering)하는 단계; 상기 재조합 숙주 세포를 탄소 공급원을 포함하는 배지 내에서 배양하는 단계; 및 선택적으로 지방족 알코올 조성물을 상기 배지에서 분리시키는 단계를 포함한다.

도면의 간단한 설명

[0015] 본원 발명은 바람직한 실시예들을 도시하는데 도움이 되는, 첨부된 도면들과 함께 읽을 때에 가장 잘 이해된다. 그러나, 본 발명이 도면들에 개시된 특정 실시예들로 제한되지 않는 것으로 이해한다.

도 1은 재조합 숙주 세포 내에서 지방산 유도체들에 전구체로서 아실 CoA의 생산에 이용하기 위한 예시적인 생합성 경로의 개략적인 개요이다. 주기(cycle)는 말로닐-ACP 및 아세틸-CoA의 축합에 의해 개시된다.

도 2는 예시적인 지방산 생합성 주기의 개략적인 개요이고, 여기서 말로닐-ACP는 (말로닐-CoA:ACP 아실전달효소; fabD에 의해 촉진되는) 말로닐-CoA의 말로닐-ACP로의 아실기전이(transacylation)에 의해 생산되며, 이후에 β -케토아실-ACP 생성효소 III(fabH)는 아세틸-CoA과 말로닐-ACP의 축합을 개시한다. 신장 주기(Elongation cycle)들은 β -케토아실-ACP 생성효소 I(fabB) 및 β -케토아실-ACP 생성효소 II(fabF)에 의해 촉진되는 말로닐-ACP 및 아실-ACP의 축합으로 시작되어 β -케토-아실-ACP를 생산하고, 이후에 상기 β -케토-아실-ACP는 NADPH-의존성 β -케토아실-ACP 환원효소(fabG)에 의해 환원되어 β -하이드록시-아실-ACP를 생산하며, 이는 β -하이드록시아실-ACP 탈수효소(fabA 또는 fabZ)에 의해 *trans*-2-엔오일-아실-ACP로 탈수된다. 또한, FabA는 *trans*-2-엔오일-아실-ACP를 *cis*-3-엔오일-아실-ACP로 이성질화시킬 수 있고, 이는 fabI를 우회할 수 있으며, (전형적으로 C16의 지방족 사슬 길이까지) fabB에 의해 이용되어 β -케토-아실-ACP를 생산할 수 있다. 각각의 주기에서 최종 단계는 *trans*-2-엔오일-아실-ACP를 아실-ACP로 전환시키는 NADH 또는 NADPH-의존성 엔오일-ACP 환원효소(fabI)에 의해 촉진된다. 본 명세서에 설명된 방법들에서, 지방산 합성의 종료는 유리 지방산(free fatty acid: FFA)을 방출하도록 아실-ACP에서 아실기의 티오에스테라아제(thioesterase)를 제거함으로써 일어난다. 티오에스테라아제들(예를 들어, tesA)은 티오에스테르 결합들을 가수분해하고, 이는 술폰드릴(sulfydryl) 결합들을 통하여 아실 사슬들 및 ACP 사이에서 일어난다.

도 3은 아세틸-CoA 카르복실라아제(accABCD) 효소 복합체(complex)의 구조 및 작용기(function)를 도시한다. 비오틴 카르복실라아제(Biotin carboxylase)는 *accC* 유전자에 의해 인코딩되는 반면에, 비오틴 카르복실 운반 단백질(biotin carboxyl carrier protein: BCCP)은 *accB* 유전자에 의해 인코딩된다. 카르복실전달효소 활성화에 관련된 두 개의 서브유닛(subunit)들은 *accA* 및 *accD* 유전자들에 의해 인코딩된다. BCCP의 공유 결합된 비오틴은 카르복실레이트 잔기(carboxylate moiety)를 수반(carry)한다. (나타내지 않은) *birA* 유전자는 홀로(holo)-*accB*를 비오틴화(biotinylate)한다.

도 4는 아실-ACP와 함께 시작하는 지방족 알코올의 생산에 대한 예시적인 생합성 경로의 개략적인 개요를 나타내고, 여기서 지방족 알데히드의 생산은 아실-ACP 환원효소(AAR) 또는 티오에스테라아제 및 카르복실산 환원효소(Car)의 효소 활성화에 의해 촉진된다. 지방족 알데히드는 (알코올 탈수소효소로도 칭하는) 알데히드 환원효소에 의해 지방족 알코올로 전환된다. 이 경로는 지방족 아실 CoA 합성효소(fadD)를 포함하지 않는다.

도 5는 대장균(*E. coli*) MG1655에 의한 지방산 유도체 생산과 비교하여 감쇠된(attenuated)[즉, 결실된] *fadE* 유전자를 갖는 MG1655 대장균 균주에 의한 지방산 유도체[총 지방 종류(Total Fatty Species)] 생산을

도시한다. 도 5에 나타낸 데이터는 fadE 유전자의 감쇠가 지방산 유도체 생산에 영향을 미치지 않는다는 것을 나타낸다.

도 6a 및 도 6b는 플라스미드 pCL-WT TRC WT TesA가 상기 도면들에 나타낸 각각의 균주들로 형질전환되고, 발효가 32 °C(도 6a) 및 37 °C(도 6b) 둘 모두에서 얻어지도록 유도(induction)로부터 20시간 FA2 배지 내에서 처리되는 경우에, 이중의 플레이트 스크린(duplicate plate screen)들로부터 "총 지방 종류"의 생산에 대한 데이터를 나타낸다.

도 7a 및 도 7b는 FAB138의 앞쪽에 통합된(integrated) cat-loxP-T5 프로모터의 도식(diagram)[도 7a]; 및 iT5_138의 도식(도 7b)을 포함하는 iFAB138 유전자 자리(locus)의 도식적인 묘사이다. cat-loxP-T5 프로모터 영역의 각각의 측면(side)에 나타낸 50 개의 염기쌍의 상동성(homology)을 갖는 FAB138의 앞쪽에 통합된 cat-loxP-T5 프로모터의 서열은 SEQ ID NO:1로 제공되고, 각각의 측면에 50 개의 염기쌍의 상동성을 갖는 iT5_138 프로모터 영역의 서열은 SEQ ID NO: 2로 제공된다.

도 8은 rph 및 ilvG 유전자들의 수정(correcting) 효과를 나타낸다. EG149(rph- ilvg-) 및 V668(EG149 rph+ ilvg+)은 D191로부터 얻어지는 pCL-tesA(P_{TRC}-tesA를 함유하는 pCL1920 플라스미드)와 함께 형질전환된다. 상기 도는 EG149 균주 내의 rph 및 ilvG 유전자들을 수정하는 것이 rph 및 ilvG 유전자들이 수정되지 않은 V668 균주보다 더 높은 수준(level)의 FFA 생산을 가능하게 한다는 점을 나타낸다.

도 9는 균주 LC535의 yijP 유전자내에 트랜스포존 카세트 삽입(transposon cassette insertion)[트랜스포존 히트(transposon hit) 68F11]의 도식적인 묘사이다. 상기 트랜스포존 카세트 내부의 프로모터들을 나타내고, 이들은 인접한 유전자 발현에 영향을 미칠 수 있다.

도 10은 균주 V324 내에서 CarB60에 의한 유리 지방산(free fatty acids)을 지방족 알코올로 전환을 나타낸다. 상기 도는 염색체로부터 CarB60을 발현시키는 세포들(검은색 막대)이 CarB(밝은색 막대)와 비교하여 더 많은 비율(greater fraction)의 C12 및 C14 유리 지방산을 지방족 알코올로 전환시킨다는 것을 나타낸다.

도 11은 염색체로부터 CarB60을 발현시키는 세포들이 CarB와 비교하여 더 많은 비율의 C12 및 C14 유리 지방산을 지방족 알코올로 전환시킨다는 것을 나타낸다.

도 12는 조합 라이브러리 돌연변이체(combination library mutant)들의 발효에 따른 지방족 알코올 생산을 나타낸다.

도 13은 진탕 플라스크(shake-flask) 발효에 따른 생산 플라스미드 내의 carB 변이체들(carB1 및 CarB2)에 의한 지방족 알코올 생산을 나타낸다.

도 14는 진탕 플라스크 발효에 따른 단일-복제(single-copy) 통합된 carB 변이체들(icarB1, icarB2, icarB3 및 icarB4)에 의한 지방족 알코올 생산을 나타낸다.

도 15는 진탕 플라스크 발효에 의해 입증된 바와 같이 개선된 CarB 변이체들에 대한 이중-플라스미드 스크리닝 시스템(dual-plasmid screening system)의 결과들을 나타낸다.

도 16은 생물반응기(bioreactors) 내에서 개선된 지방족 알코올의 생산을 위한 신규한 CarB 변이체들을 나타낸다.

발명을 실시하기 위한 구체적인 내용

개관(General Overview)

본원 발명은 신규한 변형 카르복실산 환원효소(CAR)의 효소들뿐만 아니라 이들의 핵산 및 단백질 서열들을 제공한다. 본 발명에 의해 지방족 알코올의 생산을 위한 변형 CAR 효소들을 포함하는 세포 배양물 및 재조합 숙주 세포들을 더 포괄된다. 상업적으로 실행 가능하도록 발효성 당 또는 바이오매스로부터 지방족 알코올을 생산하기 위하여, 공정은 생성물의 회수율(recovery) 및 효율적인 전환을 위해 최적화되어야 한다. 본원 발명은 조작된 변형 효소들 및 조작된 재조합 숙주 세포들을 사용하는 지방족 알코올의 개선된 생산을 위한 방법들 및 조성물들을 제공함으로써 이러한 필요성을 처리한다. 숙주 세포들은 발효 과정에 사용되는 지방족 알코올의 높은 역가의 생산을 야기하는 생촉매(biocatalysts)의 역할을 한다. 이처럼, 본 발명은 광합성 및 종속영양성(heterotrophic) 숙주 세포들을 창출하는 방법들을 더 제공하고, 상기 숙주 세포들은 직접적으로 특정 사슬 길이의 지방족 알코올 및 알파-올레핀을 생산하여 정제된 지방산의 촉매 전환이 필요하지 않도록 한다.

- [0018] 좀 더 구체적으로, 원하는 지방족 알코올 조성물의 생산은 지방족 알코올의 생산, 분해(degradation) 및/또는 분비에 대한 생합성 경로에 관련된 하나 이상의 유전자들의 발현을 변형(modifying)시킴으로써 향상될 수 있다. 본 발명은 재조합 숙주 세포들을 제공하고, 상기 숙주 세포들은 비-조작된 또는 본래의(native) 숙주 세포들에 관하여 향상된 지방족 알코올 생합성을 제공하도록 조작된다[예를 들어, 균주 개량(strain improvements)]. 또한, 본 발명은 본 발명의 재조합 숙주 세포들, 방법들 및 조성물들에 유용한 폴리뉴클레오티드들을 제공한다. 그러나, 이러한 폴리뉴클레오티드들은 실제 서열(absolute sequence)과 동일할 필요는 없다는 것을 인식할 것이다. 예를 들어, 특정 폴리뉴클레오티드 서열이 변화될 수 있고, 인코딩된 폴리펩티드는 활성에 대하여 평가된다. 전형적으로, 이러한 변화들은 보존성 돌연변이(conservative mutations) 및 침묵성 돌연변이(silent mutations)[예를 들어, 코돈 최적화(codon optimization)]를 포함한다. 변형된 또는 돌연변이된(mutated) 폴리뉴클레오티드들(즉, 돌연변이체들) 및 인코딩된 변형 폴리펩티드들은 본 기술분야에 알려진 방법들을 이용하여, 증가된 촉매 활성, 증가된 안정성 또는 감소된 저해(inhibition)[예를 들어, 감소된 피드백 저해]를 포함하나 이로 제한되지 않는 원하는 작용에 대하여 스크리닝될 수 있다.
- [0019] 본 발명은 효소 분류(Enzyme Classification: EC) 번호에 따라서 본 명세서에 설명된 지방산 생합성 경로들의 다양한 단계들(즉, 반응들)에 관련된 효소 활성들을 식별하고, 이러한 EC 번호들로 분류되는 예시적인 폴리펩티드들(예를 들어, 효소들) 및 이러한 폴리펩티드들을 인코딩하는 예시적인 폴리뉴클레오티드들을 제공한다. 본 명세서에서 수탁 번호(Accession Numbers) 및/또는 서열 식별 번호(Sequence Identifier Numbers: SEQ ID NOs)로 식별되는 이러한 예시적인 폴리펩티드들 및 폴리뉴클레오티드들은 본 명세서에 설명된 재조합 숙주 세포들을 얻기 위하여 모(parental) 숙주 세포들 내에서의 지방산 경로들을 조작하는데 유용하다. 그러나, 본 명세서에 설명된 폴리펩티드들 및 폴리뉴클레오티드들은 예시적이며, 이로 제한되지 않는 것으로 이해한다. 본 명세서에 설명된 예시적인 폴리펩티드들의 동족체(homologue)들의 서열들은 다양한 데이터베이스들[예를 들어, 미국 국립 생물공학 정보센터(NCBI)에 의해 제공되는 rhw 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, 이들 모두는 월드 와이드 웹(World Wide Web)에서 이용가능함]를 사용하여 본 기술분야의 당업자들이 이용가능하다.
- [0020] 여러 가지의 숙주 세포들은 본 명세서에 설명된 것들과 같은 지방족 알코올 생합성 효소들을 함유하도록 변형될 수 있고, 지방족 알코올 조성물들의 생산에 적합한 재조합 숙주 세포들을 야기할 수 있다. 여러 가지의 세포들이 본 명세서에 제공되는 재조합 숙주 세포에 사용하기 적합한 폴리펩티드들을 인코딩하는 폴리뉴클레오티드 서열들을 포함하는, 유전 물질(genetic material)의 공급원들을 제공할 수 있다고 이해한다.
- [0021] 정의
- [0022] 달리 정의되지 않는다면, 본 명세서에서 사용되는 모든 과학 기술 용어들은 본 발명과 관련된 기술 분야의 당업자들이 일반적으로 이해하는 것과 동일한 의미를 갖는다. 본 명세서에서 설명되는 것들과 유사하거나 동등한 다른 방법들 및 물질들이 본원 발명의 실시예에 사용될 수 있으나, 바람직한 물질들 및 방법들은 본 명세서에 기재된다. 본원 발명을 설명하고 주장하는데 있어서, 다음의 용어들은 하기에 규정되는 정의들에 따라서 사용될 것이다.
- [0023] 수탁 번호: 상세한 설명 전반의 서열 수탁 번호는 (데이터베이스 내에서 "NCBI 수탁 번호" 또는 대안적으로 "GenBank 수탁 번호"로 식별되는) 미국 국립 보건원에 의해 유지되는 NCBI(미국 국립 생물공학 정보센터)에서 제공한 데이터베이스, 및 (데이터베이스 내에서 "UniProtKB 수탁 번호"로 식별되는) 스위스 생물정보학 연구소에서 제공하는 Swiss-Prot 데이터베이스 및 UniProt 지식베이스(UniProtKB)로부터 얻어진다.
- [0024] 효소 분류(EC) 번호: EC 번호는 국제 생화학 및 분자생물학 연합(International Union of Biochemistry and Molecular Biology: IUBMB)의 명명 위원회에 의해서 제정되고, 이의 상세한 설명은 월드 와이드 웹의 IUBMB 효소 명명 웹사이트에서 이용가능하다. EC 번호는 효소들이 촉진시키는 반응에 따라서 효소들을 분류한다.
- [0025] 본 명세서에서 사용되는 바와 같이, "뉴클레오티드"라는 용어는 헤테로사이클릭 염기, 당 및 하나 이상의 인산기들로 이루어진 폴리뉴클레오티드의 단위체 단위를 칭한다. 자연적으로 및 비-자연적으로 발생하는 염기 유사체(analog)들도 포함된다고 이해해야 하지만, 자연적으로 발생하는 염기들[구아닌(G), 아데닌(A), 사이토신(C), 티민(T) 및 우라실(U)]은 전형적으로 퓨린 또는 피리미딘의 유도체들이다. 자연적으로 및 비-자연적으로 발생하는 당 유사체들도 포함된다고 이해해야 하지만, 자연적으로 발생하는 당은 펜토오스(5-탄당)의 (DNA를 형성하

는) 디옥시리보오스 또는 (RNA를 형성하는) 리보오스이다. 많은 다른 연결들(예를 들어, 포스포로티오에이트, 보라노포스페이트 및 이와 유사한 것들)이 본 기술분야에 알려져 있으나, 핵산들은 전형적으로 핵산들 또는 폴리뉴클레오티드들을 형성하는 인산 결합들을 통하여 연결된다.

[0026] 본 명세서에서 사용되는 바와 같이, "폴리뉴클레오티드"라는 용어는 리보뉴클레오티드(RNA) 또는 디옥시리보뉴클레오티드(DNA)의 중합체를 칭하고, 이는 외가닥 또는 이중 가닥일 수 있으며, 비-천연 또는 변화된(altered) 뉴클레오티드들을 함유할 수 있다. "폴리뉴클레오티드", "핵산 서열" 및 "뉴클레오티드 서열"이라는 용어들은 여하한 길이의 뉴클레오티드들의 중합체 형태, RNA 또는 DNA 중 하나를 칭한다. 이러한 용어들은 분자의 일차 구조를 지칭하고, 따라서 이중 가닥 및 외가닥의 DNA, 및 이중 가닥 및 외가닥의 RNA를 포함한다. 상기 용어들은 메틸화된 및/또는 캡핑된(capped) 폴리뉴클레오티드들로 제한되지 않으나, 이들과 같은 변형된 폴리뉴클레오티드들 및 뉴클레오티드 유사체들로부터 만들어지는 RNA 또는 DNA 중 하나의 유사체들을 등가물(equivalent)들로서 포함한다. 폴리뉴클레오티드는 플라스미드, 바이러스(viral), 염색체(chromosomal), EST, cDNA, mRNA 및 rRNA를 포함하나 이로 제한되지 않는 여하한 형태로 존재할 수 있다.

[0027] 본 명세서에서 사용되는 바와 같이, "폴리펩티드" 및 "단백질"이라는 용어들은 아미노산 잔기들의 중합체를 칭하도록 통용되어 사용된다. "재조합 폴리펩티드"라는 용어는 재조합 기술들에 의해 생산된 폴리펩티드를 칭하고, 여기서 일반적으로 발현되는 단백질을 인코딩하는 DNA 또는 RNA는 결국 폴리펩티드를 생산하기 위해 숙주 세포를 형질전환시키는데 사용되는 적합한 발현 벡터 내로 삽입된다.

[0028] 본 명세서에서 사용되는 바와 같이, "동족체(homolog)" 및 "상동"이라는 용어들은 대응하는 폴리뉴클레오티드 또는 폴리펩티드 서열과 적어도 약 50 % 동일한 서열을 포함하는 폴리뉴클레오티드 또는 폴리펩티드를 칭한다. 바람직하게, 상동 폴리뉴클레오티드들 또는 폴리펩티드들은 대응하는 아미노산 서열 또는 폴리뉴클레오티드 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 % 상동성을 갖는 폴리뉴클레오티드 서열들 또는 아미노산 서열들을 갖는다. 본 명세서에서 사용되는 바와 같이, 서열 "상동성" 및 서열 "동일성"은 통용되어 사용된다.

[0029] 본 기술분야의 당업자는 둘 이상의 서열들 간의 상동성을 결정하는 방법들을 잘 알고 있을 것이다. 간단히 말해서, 두 서열들 간의 "상동성"의 계산은 다음과 같이 실시될 수 있다. 상기 서열들은 최적의 비교 목적들을 위하여 정렬된다[예를 들어, 갭(gap)들은 최적의 정렬을 위하여 핵산 서열, 또는 제 1 및 제 2 아미노산 중 하나 또는 둘 모두에 도입될 수 있고, 그리고 비-상동 서열들은 비교 목적을 위해 무시될 수 있다]. 바람직한 실시예에서, 비교 목적을 위하여 정렬되는 제 1 서열의 길이는 제 2 서열의 길이의 적어도 약 30 %, 바람직하게는 적어도 약 40 %, 더욱 바람직하게는 적어도 약 50 %, 훨씬 바람직하게는 적어도 약 60 %, 그리고 훨씬 더 바람직하게는 적어도 약 70 %, 적어도 약 80 %, 적어도 약 90 % 또는 약 100 %이다. 이후에, 제 1 및 제 2 서열들의 대응하는 아미노산 위치들 또는 뉴클레오티드 위치들에서의 아미노산 잔기들 또는 뉴클레오티드들이 비교된다. 제 1 서열 내의 위치가 제 2 서열 내의 대응하는 위치로서 동일한 아미노산 잔기 또는 뉴클레오티드에 의해 점유(occupied)되는 경우, 이후에 분자들은 이 위치에서 동일하다. 두 서열들 간의 퍼센트 상동성(percent homology)은 두 서열들의 최적의 정렬을 위하여 도입될 필요가 있는, 갭들의 개수 및 각각의 갭의 길이를 고려하여, 서열들에 의해 공유되는 동일한 위치들의 개수의 함수이다.

[0030] 두 서열들 간의 퍼센트 상동성의 결정 및 서열들의 비교는 BLAST와 같은 수학적 알고리즘을 이용하여 완수될 수 있다[Altschul et al., J. Mol. Biol., 215(3):403-410(1990)]. 또한, 두 아미노산 서열들 간의 퍼센트 상동성은 Blossum 62 매트릭스 또는 PAM250 매트릭스 중 하나, 및 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 이용하여, GCG 소프트웨어 패키지의 GAP 프로그램에 통합되는 Needleman 및 Wunsch 알고리즘을 사용해 결정될 수 있다[Needleman and Wunsch, J. Mol. Biol., 48:444-453 (1970)]. 또한, 두 뉴클레오티드 서열들 간의 퍼센트 상동성은 NWSgapdna.CMP 매트릭스 및 40, 50, 60, 70 또는 80의 갭 중량 및 1, 2, 3, 4, 5 또는 6의 길이 중량을 이용하여, GCG 소프트웨어 패키지의 GAP 프로그램을 사용해 결정될 수 있다. 본 기술분야의 당업자는 초기 상동성 계산을 실시할 수 있고, 이에 맞춰 알고리즘 파라미터들을 조절할 수 있다. 바람직한 파라미터들의 세트(set) [그리고 만약 당업자(practitioner)는 분자가 청구항들의 상동성 제한 내에 있는지의 여부를 결정하는데 적용되어야 하는 파라미터들에 대하여 확신이 없다면 사용되어야 하는 파라미터들의 세트]는 12의 갭 페널티(gap penalty), 4의 갭 확장 페널티(gap extend penalty) 및 5의 프레임시프트 갭 페널티(frameshift gap penalty)를 갖는 Blossum 62 스코어링(scoring) 매트릭스이다. 서열 정렬의 추가적인 방법들은 생물공학 분야에 알려져 있다[예를 들어, Rosenberg, BMC Bioinformatics, 6:278(2005); Altschul, et al., FEBS J., 272(20):5101-5109(2005)].

- [0031] 본 명세서에서 사용되는 바와 같이, 덜 엄격한(low stringency), 엄격한(medium stringency), 매우 엄격한(high stringency) 또는 아주 매우 엄격한(very high stringency) 조건들 하에서 교배하다"라는 용어는 교배 및 세정에 대한 조건들을 설명한다. 혼성화 반응(hybridization reaction)들을 실시하기 위한 안내는 *olecular Biology*, John Wiley & Sons, N.Y.(1989), 6.3.1-6.3.6 내에 현재 프로토콜(Protocol)들에서 찾을 수 있다. 수성(Aqueous) 및 비-수성 방법들은 상기 참고문헌에 설명되어 있고, 상기 방법들 중 하나가 사용될 수 있다. 본 명세서에 언급된 특정한 교배 조건들은 다음과 같다: 1) 덜 엄격한 교배 조건 - 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 엄격성의 조건의 경우 55 °C까지 증가될 수 있음); 2) 엄격한 교배 조건 - 약 45 °C에서 6X SSC, 이후에 60 °C에서 0.2X SSC, 0.1 % SDS로 한번 이상 세정; 3) 매우 엄격한 교배 조건 - 약 45 °C에서 6X SSC, 이후에 65 °C에서 0.2X SSC, 0.1 % SDS로 한번 이상 세정; 및 4) 아주 매우 엄격한 교배 조건 - 65 °C에서 0.5 M 소듐 포스페이트, 7 % SDS, 이후에 65 °C에서 0.2X SSC, 1 % SDS로 한번 이상 세정. 달리 명시하지 않는 한, 아주 매우 엄격한 조건(4)이 바람직한 조건이다.
- [0032] "내인성" 폴리펩티드는 재조합 세포가 조작되는 (또는 "유래되는") 모 미생물 세포("숙주 세포"라고도 칭함)의 게놈(genome)에 의해 인코딩되는 폴리펩티드를 칭한다.
- [0033] "외인성" 폴리펩티드는 모 미생물 세포의 게놈에 의해 인코딩되지 않는 폴리펩티드를 칭한다. 변형(즉, 변종) 폴리펩티드는 외인성 폴리펩티드의 일 예이다.
- [0034] 일반적으로, "이종(heterologous)"이라는 용어는 상이한 종으로부터 유래되거나, 또는 상이한 유기체로부터 유래되는 것을 의미한다. 본 명세서에서 사용되는 바와 같이, 특정한 유기체 내에 자연적으로 존재하지 않는 뉴클레오티드 서열 또는 폴리펩티드 서열을 칭한다. 이종 발현은 단백질 또는 폴리펩티드가 보통 상기 단백질을 발현하지 않는 세포에 실험적으로 첨가되는 것을 의미한다. 이처럼, 이종은 도입된 단백질(transferred protein)이 처음에 수용체(recipient)와는 상이한 세포 타입 또는 상이한 종으로부터 유래된다는 사실을 칭한다. 예를 들어, 식물 세포에 내인성 폴리뉴클레오티드 서열은 재조합 방법들에 의해 박테리아성 숙주 세포 내로 도입될 수 있고, 이후에 상기 식물의 폴리뉴클레오티드가 재조합형 박테리아성 숙주 세포 내의 이종 폴리뉴클레오티드이다.
- [0035] 본 명세서에서 사용되는 바와 같이, 폴리펩티드의 "분절(fragment)"이라는 용어는 네 개의 아미노산 잔기 내지 전체 아미노산 서열에서 하나의 아미노산 잔기를 제외한 크기에 이르는 전체 길이의(full-length) 폴리펩티드 또는 단백질의 짧은 부분을 칭한다. 본 발명의 여하한 실시예에서, 분절은 폴리펩티드 또는 단백질의 도메인(domain)[예를 들어, 기질 결합 도메인 또는 촉매 도메인]의 전체 아미노산 서열을 칭한다.
- [0036] 본 명세서에서 사용되는 바와 같이, "돌연변이 유발(mutagenesis)"이라는 용어는 유기체의 유전 정보를 안정된 방식으로 변화시키는 과정을 칭한다. 단백질 코딩 핵산 서열의 돌연변이 유발은 돌연변이 단백질을 생산한다. 또한, 돌연변이 유발은 변형된 단백질 활성을 야기하는 비-코딩 핵산 서열들의 변화를 칭한다.
- [0037] 본 명세서에서 사용되는 바와 같이, "유전자"라는 용어는 RNA 또는 단백질의 발현에 영향을 미치는 작동가능하게 연결된(operably-linked) 핵산 서열들[예를 들어, 프로모터 또는 증폭자(enhancer) 서열들을 포함하나 이로 제한되지 않는 이러한 서열들] 또는 RNA 또는 단백질의 발현에 영향을 미치는 서열들을 인코딩하는 작동가능하게 연결된 핵산 서열들[예를 들어, 리보솜 결합 부위들 또는 번역 조절(translational control) 서열들을 포함하나 이로 제한되지 않는 이러한 서열들]뿐만 아니라, RNA 생성물과 단백질 생성물 중 하나를 인코딩하는 핵산 서열들을 칭한다.
- [0038] 발현 조절(Expression control) 서열들은 본 기술분야에 알려져 있고, 예를 들어 숙주 세포 내에서 폴리뉴클레오티드 서열의 발현을 위하여 제공되는 프로모터, 증폭자, 아데닐산중합반응 신호(polyadenylation signal), 전사 종결자(transcription terminator), 내부 리보솜 유입점(internal ribosome entry sites: IRES) 및 이와 유사한 것들을 포함한다. 발현 조절 서열들은 특히 전사에 관련된 세포성 단백질들과 상호작용을 한다[Maniatis et al., *Science*, 236:1237-1245(1987)]. 예시적인 발현 조절 서열들은, 예를 들어 Goeddel, *Gene Expression Technology: Methods in Enzymology*, Vol. 185, Academic Press, San Diego, Calif.(1990)에 설명된다.
- [0039] 본 발명의 방법들에서, 발현 조절 서열은 폴리뉴클레오티드 서열에 작동가능하게 연결된다. "작동가능하게 연결된"이란 적절한 분자들(예를 들어, 전사 활성화인자 단백질들)이 발현 조절 서열(들)에 결합되는 경우에, 폴리뉴클레오티드 서열 및 발현 조절 서열(들)이 유전자 발현을 가능하게 하는 방식으로 연결되는 것을 의미한다. 작동가능하게 연결된 프로모터들은 전사 및 번역의 방향에 관하여 선택된 폴리뉴클레오티드 서열의 상류

(upstream)에 위치된다. 작동가능하게 연결된 증폭자들은 선택된 폴리뉴클레오티드의 상류, 내부(within) 또는 하류(downstream)에 위치될 수 있다.

[0040] 본 명세서에서 사용되는 바와 같이, "벡터"라는 용어는 상기 벡터와 연결되는 또 다른 핵산, 즉 폴리뉴클레오티드 서열을 수송할 수 있는 핵산 분자를 칭한다. 유용한 벡터 중 한 가지 타입은 에피솜(episome)[즉, 염색체의 복제(extra-chromosomal replication)가 가능한 핵산]이다. 유용한 벡터는 벡터와 연결되는 핵산들의 자동 복제 및/또는 발현할 수 있는 벡터이다. 작동가능하게 연결되는 유전자들의 발현을 지시할 수 있는 벡터는 본 명세서에서 "발현 벡터"로 칭한다. 일반적으로, 재조합 DNA 기술들에서 유용한 발현 벡터는 종종 "플라스미드"의 형태인데, 이는 일반적으로 벡터 형태에서 염색체에 결합되지 않는 원형 이중 가닥 DNA 루프(loops)를 칭한다. "플라스미드" 및 "벡터"라는 용어는 본 명세서에서 통용되어 사용되는데, 그 이유는 플라스미드가 벡터의 가장 일반적으로 사용되는 형태이기 때문이다. 그러나, 등가적인 기능들을 수행하고 이후에 본 기술분야에 알려지는 다른 형태들의 발현 벡터도 포함된다. 일부 실시예들에서, 재조합 벡터는 (a) 폴리뉴클레오티드 서열에 작동가능하게(operatively) 연결된 발현 조절 서열; (b) 폴리뉴클레오티드 서열에 작동가능하게 연결된 선택 마커; (c) 폴리뉴클레오티드 서열에 작동가능하게 연결된 마커 서열; (d) 폴리뉴클레오티드 서열에 작동가능하게 연결된 정제 모이어티(purification moiety); (e) 폴리뉴클레오티드 서열에 작동가능하게 연결된 분비 서열(secretion sequence); 및 (f) 폴리뉴클레오티드 서열에 작동가능하게 연결된 표적화 서열(targeting sequence)을 포함하는 서열들 중 적어도 하나의 서열을 포함한다. 본 명세서에 설명된 발현 벡터들은 숙주 세포 내에서 폴리뉴클레오티드 서열의 발현에 적합한 형태의 본 명세서에 설명된 폴리뉴클레오티드 서열을 포함한다. 발현 벡터의 설계는 형질전환되는 숙주 세포의 선택, 원하는 폴리펩티드의 발현 수준 등과 같은 요인들에 의존할 수 있다는 것을 본 기술분야의 당업자들은 인지할 것이다. 본 명세서에 설명된 발현 벡터들은 본 명세서에 설명되는 바와 같이 폴리뉴클레오티드 서열들에 의해 인코딩되는, 융합 폴리펩티드들을 포함하는 폴리펩티드들을 생산하기 위하여 숙주 세포들에 도입될 수 있다.

[0041] 원핵생물, 예를 들어 대장균(*E. coli*)에서 폴리펩티드들을 인코딩하는 유전자의 발현은 대부분 융합 및 비-융합 폴리펩티드들 중 하나의 발현을 지시하는 구성적 또는 유도성 프로모터를 함유하는 벡터들로 수행된다. 융합 벡터들은 그 속에 인코딩된 폴리펩티드에, 일반적으로 재조합 폴리펩티드의 아미노 또는 카르복시 말단에 다수의 아미노산들을 첨가한다. 전형적으로, 이러한 융합 벡터들은 다음의 3가지 목적: (1) 재조합 폴리펩티드의 발현을 증가시키고; (2) 재조합 폴리펩티드의 용해도를 증가시키며; 그리고 (3) 친화성 정제(affinity purification)에서 리간드로서 작용함으로써 재조합 폴리펩티드의 정제를 돕는 것을 수행한다. 종종, 융합 발현 벡터에서, 융합 모이어티와 재조합 폴리펩티드의 접합점에 단백분해 절단(proteolytic cleavage) 부위가 도입된다. 이는 융합 폴리펩티드의 정제 이후에 융합 모이어티로부터 재조합 폴리펩티드의 분리를 가능하게 한다. 여하한 실시예들에서, 본 발명의 폴리뉴클레오티드 서열은 박테리오파지 T5로부터 유래된 프로모터에 작동가능하게 연결된다. 여하한 실시예들에서, 숙주 세포는 효모 세포이고, 발현 벡터는 효모 발현 벡터이다. 효모인 맥주효모균(*S. cerevisiae*) 내에서의 발현을 위한 벡터들의 예시들은 pYepSec1[Baldari et al., *EMBO J.*, 6:229-234(1987)], pMFa[Kurjan et al., *Cell*, 30:933-943(1982)], pJRY88[Schultz et al., *Gene*, 54:113-123(1987)], pYES2(Invitrogen Corp., San Diego, CA) 및 picZ(Invitrogen Corp., San Diego, CA)를 포함한다. 다른 실시예들에서, 숙주 세포는 곤충 세포이고, 발현 벡터는 배콜로바이러스(baculovirus) 발현 벡터이다. 배양된 곤충 세포들 내에서 단백질들의 발현을 위하여 이용가능한 배콜로바이러스 벡터들은, 예를 들어 pAc 계열(series)[(Smith et al., *Mol. Cell Biol.*, 3:2156-2165(1983))] 및 pVL 계열[Lucklow et al., *Virology*, 170: 31-39(1989)]을 포함한다. 또 다른 실시예에서, 본 명세서에 설명된 폴리뉴클레오티드 서열들은 포유동물 발현 벡터를 이용하여 포유동물 세포들 내에서 발현될 수 있다. 원핵 및 진핵 세포들 둘 모두에 대한 다른 적합한 발현 시스템들은 본 기술분야에 잘 알려져 있다; 예를 들어, Sambrook et al., "Molecular Cloning: A Laboratory Manual", second edition, Cold Spring Harbor Laboratory, (1989)를 참조한다.

[0042] 본 명세서에서 사용되는 바와 같이, "아실-CoA"는 조효소 A(CoA)의 4'-포스포판테티오닐 잔기(moiety)의 술포드 릴기와 알킬 사슬의 카르보닐 탄소 사이에서 형성되는 아실 티오에스테르를 칭하고, 이는 화학식 R-C(O)S-CoA를 가지며, 여기서 R은 적어도 4개의 탄소 원자들을 갖는 여하한 알킬기이다.

[0043] 본 명세서에서 사용되는 바와 같이, "아실-ACP"는 아실기 운반 단백질(ACP)의 포스포판테티오닐 잔기의 술포드 릴기와 알킬 사슬의 카르보닐 탄소 사이에서 형성되는 아실 티오에스테르를 칭한다. 포스포판테티오닐 잔기는 홀로(holo)-아실기 운반 단백질 생성효소(ACPS)인 포스포판테티오닐 전달효소의 작용에 의해 ACP 상의 보존된 세린 잔기에 이후에-번역되도록(post-translationally) 부착된다. 일부 실시예들에서, 아실-ACP는 완전히 포화된 아실-ACP들의 합성에 있어서 중간체이다. 다른 실시예들에서, 아실-ACP는 불포화된 아실-ACP들의 합성에 있

어서 중간체이다. 일부 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. 이러한 아실-ACP들의 각각은 이들을 지방산 유도체들로 전환시키는 효소들에 대한 기질들이다.

[0044] 본 명세서에서 사용되는 바와 같이, "지방산 또는 이의 유도체"라는 용어는 "지방산" 또는 "지방산 유도체"를 의미한다. "지방산"이란 용어는 화학식 RCOOH 를 갖는 카르복실산을 의미한다. R은 지방족기, 바람직하게는 알킬기를 나타낸다. R은 약 4 내지 22 개의 탄소 원자들을 포함할 수 있다. 지방산들은 포화, 단일불포화 또는 다중불포화될 수 있다. 바람직한 실시예에서, 지방산은 지방산 생합성 경로로부터 만들어진다. "지방산 유도체"라는 용어는 생산 숙주 유기체의 지방산 생합성 경로의 일부에서 만들어지는 생성물을 의미한다. 또한, "지방산 유도체들"은 아실-ACP 또는 아실-ACP 유도체들의 일부에서 만들어지는 생성물들을 포함한다. 예를 들어, 예시적인 지방산 유도체들은 아실-CoA, 지방족 알데히드, 짧은 사슬형 및 긴 사슬형 알코올, 탄화수소 및 에스테르(예를 들어, 왁스, 지방산 에스테르 또는 지방족 에스테르)를 포함한다.

[0045] 본 명세서에서 사용되는 바와 같이, "지방산 생합성 경로"라는 용어는 지방산 유도체들, 예를 들어 지방족 알코올들을 생산하는 생합성 경로를 의미한다. 지방산 생합성 경로는 지방산들을 생산하도록 조작될 수 있고, 일부 실시예들에서는 원하는 특성들을 갖는 지방족 알코올들과 같은 지방산 유도체들을 생산하도록 추가적인 효소들과 함께 발현될 수 있는 지방산 합성효소를 포함한다.

[0046] 본 명세서에서 사용되는 바와 같이, "지방족 알데히드"는 카르보닐기(C=O)에 의해 특징되는 화학식 RCHO 를 갖는 알데히드를 의미한다. 일부 실시예들에서, 지방족 알데히드는 지방족 알코올로부터 만들어지는 여하한 알데히드이다. 여하한 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 종단점(endpoint)들 중 여하한 두 개로 제한되는(bounded by) R기를 가질 수 있다. 예를 들어, R기는 길이가 6 내지 16 개의 탄소들, 10 내지 14 개의 탄소들, 또는 12 내지 18 개의 탄소들일 수 있다. 일부 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. 여하한 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.

[0047] 본 명세서에서 사용되는 바와 같이, "지방족 알코올"은 화학식 ROH 를 가지는 알코올을 의미한다. 일부 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 종단점(endpoint)들 중 여하한 두 개로 제한되는 R기를 가질 수 있다. 예를 들어, R기는 길이가 6 내지 16 개의 탄소들, 10 내지 14 개의 탄소들, 또는 12 내지 18 개의 탄소들일 수 있다. 일부 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. 여하한 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.

[0048] 본 명세서에서 언급된 바와 같이 "지방족 알코올 조성물"은 재조합 숙주 세포에 의해 생산되고, 전형적으로 지방족 알코올들의 혼합물을 포함한다. 일부 경우에, 상기 혼합물은 하나 이상의 타입의 생성물(예를 들어, 지방산 및 지방족 알코올)을 포함한다. 다른 경우에, 지방산 유도체 조성물들은 다양한 사슬 길이 및 포화도 또는 분지 특성(branching characteristics)을 갖는 지방족 알코올들의 혼합물을 포함할 수 있다. 다른 경우에, 지방족 알코올 조성물은 다양한 사슬 길이 및 포화도 또는 분지 특성을 갖는 하나 이상의 타입의 생성물 및 생성물들 둘 모두의 혼합물을 포함한다.

[0049] 전형적으로, 지방족 알데히드를 생산하도록 조작된 숙주 세포는 지방족 알데히드 일부를 지방족 알코올로 전환시킬 것이다. 지방족 알코올들을 생산하는 숙주 세포가 에스테르 합성효소를 인코딩하는 폴리뉴클레오타이드를 발현시키도록 조작되는 경우, 왁스 에스테르(wax esters)가 생산된다. 일 실시예에서, 지방족 알코올들은 지방산 생합성 경로로부터 만들어진다. 일 예시로서, 아실-ACP는 티오에스테라아제(예를 들어, *대장균* *TesA*)의 작용에 의하여 지방산으로 전환될 수 있고, 이는 카르복실산 환원효소(예를 들어, *대장균* *CarB*)의 작용에 의하여

지방족 알데히드 및 지방족 알코올로 전환된다. 예를 들어, 지방족 알데히드의 지방족 알코올로의 전환은 지방족 알코올 생합성 폴리펩티드의 작용에 의하여 더욱 용이하게 될 수 있다. 일부 실시예들에서, 지방족 알코올 생합성 폴리펩티드를 인코딩하는 유전자는 숙주 세포 내에서 발현 또는 과발현된다. 여하한 실시예들에서, 지방족 알코올 생합성 폴리펩티드는 알데히드 환원효소 또는 알코올 탈수소효소 활성을 갖는다. 본 발명에 따라서 유용한 알코올 탈수소효소 폴리펩티드의 예시들은 *아시네토박터*속(*Acinetobacter sp.*) M-1(SEQ ID NO: 3)의 AlrA 또는 AlrAadp1(SEQ ID NO: 4)과 같은 AlrA 동족체들; 그리고 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 탈수소효소들을 포함하나, 이로 제한되지 않는다. 추가적인 예시들은 국제 특허 출원 공개 Nos. WO 2007/136762, WO 2008/119082 및 WO 2010/062480에 기재되고, 이들의 각각은 특히 본 명세서에서 인용참조된다. 여하한 실시예들에서, 지방족 알코올 생합성 폴리펩티드는 알데히드 환원효소 또는 알코올 탈수소효소 활성(EC 1.1.1.1)을 갖는다.

[0050]

본 명세서에서 사용되는 바와 같이, "알코올 탈수소효소"라는 용어는 지방족 알데히드의 알코올(예를 들어, 지방족 알코올)로의 전환을 촉진시킬 수 있는 폴리펩티드를 칭한다. 본 기술분야의 당업자는 여하한 알코올 탈수소효소들이 또한 다른 반응들을 촉진시킬 수 있고, 이러한 비-특이적 알코올 탈수소효소들도 "알코올 탈수소효소"라는 용어에 포괄된다는 것을 인식할 것이다. 지방산, 지방족 알데히드 또는 지방족 알코올의 R기는 직쇄형 또는 분지쇄형일 수 있다. 분지쇄형은 하나 이상의 분지점(point of branching)들을 가질 수 있고, 사이클릭형 분지들을 포함할 수 있다. 일부 실시예들에서, 분지형 지방산, 분지형 지방족 알데히드 또는 분지형 지방족 알코올은 C₆, C₇, C₈, C₉, C₁₀, C₁₁, C₁₂, C₁₃, C₁₄, C₁₅, C₁₆, C₁₇, C₁₈, C₁₉, C₂₀, C₂₁, C₂₂, C₂₃, C₂₄, C₂₅ 또는 C₂₆ 분지형 지방산, 분지형 지방족 알데히드 또는 분지형 지방족 알코올이다. 특정 실시예들에서, 분지형 지방산, 분지형 지방족 알데히드 또는 분지형 지방족 알코올은 C₆, C₈, C₁₀, C₁₂, C₁₃, C₁₄, C₁₅, C₁₆, C₁₇ 또는 C₁₈ 분지형 지방산, 분지형 지방족 알데히드 또는 분지형 지방족 알코올이다. 여하한 실시예들에서, 분지형 지방산, 분지형 지방족 알데히드 또는 분지형 지방족 알코올의 하이드록시기는 첫번째(C₁) 위치에 존재한다. 여하한 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. 분지형 또는 비분지형 지방산, 분지형 또는 비분지형 지방족 알데히드, 또는 분지형 또는 비분지형 지방족 알코올의 R기는 포화되거나 불포화될 수 있다. 만약 불포화된다면, R기는 하나 이상의 불포화 지점들을 가질 수 있다. 일부 실시예들에서, 불포화 지방산, 불포화 지방족 알데히드 또는 불포화 지방족 알코올은 단일불포화 지방산, 단일불포화 지방족 알데히드 또는 단일불포화 지방족 알코올이다. 여하한 실시예들에서, 불포화 지방산, 불포화 지방족 알데히드 또는 불포화 지방족 알코올은 C_{6:1}, C_{7:1}, C_{8:1}, C_{9:1}, C_{10:1}, C_{11:1}, C_{12:1}, C_{13:1}, C_{14:1}, C_{15:1}, C_{16:1}, C_{17:1}, C_{18:1}, C_{19:1}, C_{20:1}, C_{21:1}, C_{22:1}, C_{23:1}, C_{24:1}, C_{25:1} 또는 C_{26:1} 불포화 지방산, 불포화 지방족 알데히드 또는 불포화 지방족 알코올이다. 여하한 바람직한 실시예들에서, 불포화 지방산, 불포화 지방족 알데히드 또는 불포화 지방족 알코올은 C_{10:1}, C_{12:1}, C_{14:1}, C_{16:1} 또는 C_{18:1}이다. 다른 실시예들에서, 불포화 지방산, 불포화 지방족 알데히드 또는 불포화 지방족 알코올은 오메가-7 위치에서 불포화된다. 여하한 실시예들에서, 불포화 지방산, 불포화 지방족 알데히드 또는 불포화 지방족 알코올은 cis 이중 결합을 포함한다.

[0051]

본 명세서에서 사용되는 바와 같이, 재조합 또는 조작된 "숙주 세포"는 숙주 세포, 예를 들어 지방족 알코올을 생산하도록 변형된 미생물이다. 일부 실시예들에서, 재조합 숙주 세포는 하나 이상의 폴리뉴클레오티드들을 포함하고, 각각의 폴리뉴클레오티드는 지방족 알데히드 및/또는 지방족 알코올 생합성 효소 활성을 갖는 폴리펩티드를 인코딩하며, 여기서 폴리펩티드를 발현시키는데 효과적인 조건들 하에서 탄소 공급원의 존재 하에서 배양되는 경우에, 재조합 숙주 세포는 지방족 알코올 조성물을 생산한다.

[0052]

본 명세서에서 사용되는 바와 같이, 전형적으로 "클론"이라는 용어는 본질적으로 유전적으로 동일한 단일 공통 선조의 자손인 세포들의 그룹 또는 세포, 예를 들어 단일 박테리아성 세포에서 발생하는 복제된(cloned) 박테리

아성 콜로니의 박테리아를 칭한다.

[0053] 본 명세서에서 사용되는 바와 같이, "배양"이라는 용어는 생육가능한(viable) 세포들을 포함하는 액체 배지를 칭한다. 일 실시예에서, 배양은 제어된 조건들 하에서 미리 결정된 배양 배지 내에서 번식하는 세포들, 예를 들어 선택된 탄소 공급원 및 질소를 포함하는 액체 배지 내에서 성장하는 재조합 숙주 세포들의 배양물(culture)을 포함한다. "배양하는(culturing)" 또는 "배양(cultivation)"은 액체 또는 고체 배지 내에서 적합한 조건들 하에서 미생물 세포들의 개체군(population)을 성장시키는 것을 칭한다. 특정 실시예들에서, "배양하는"은 기질의 최종 생성물로써 발효성 생물전환(bioconversion)을 칭한다. 배양 배지는 잘 알려져 있고, 이러한 배양 배지의 각각의 성분들은, 예를 들어 상표 Difco™ 및 BBL™라는 이름의 상용 공급원(commercial source)들로부터 이용가능하다. 비-제한적인 일 실시예에서, 수성 영양소를 함유하는 배지(aqueous nutrient medium)는 질소, 염(salt) 및 탄소의 복합적인 공급원들을 포함하는 "영양이 풍부한 배지(rich medium)", 이를테면 배지 중 10 g/L의 펩톤 및 10 g/L 효모 추출물을 포함하는 YP 배지이다. 숙주 세포는 미국 특허 5,000,000; 5,028,539; 5,424,202; 5,482,846; 5,602,030; WO 2010127318에 설명된 방법들에 따라서 탄소 공급원으로서 셀룰로오스성 물질(cellulosic material)들을 사용하고 효과적으로 탄소를 흡수하도록 추가적으로 조작될 수 있고, 상기 각각의 특허들은 특히 본 명세서에서 인용참조된다. 또한, 숙주 세포는 수크로오스를 탄소 공급원으로서 사용할 수 있도록 자당 분해효소(invertase)를 발현하도록 조작될 수 있다.

[0054] 본 명세서에서 사용되는 바와 같이, "상기 이중 뉴클레오티드 서열을 발현시키는데 효과적인 조건들 하에서"라는 용어는 숙주 세포가 원하는 지방족 알데히드 또는 지방족 알코올을 생산할 수 있게 하는 여하한 조건을 의미한다. 예를 들어, 적합한 조건들은 발효 조건들을 포함한다.

[0055] 본 명세서에서 사용되는 바와 같이, 재조합 숙주 세포 내에서 단백질, 예를 들어 효소의 활성의 "변형된" 또는 "변화된 수준(level)"은 모(parent) 또는 본래의(native) 숙주 세포에 대하여 결정된 활성에 있어서 한 가지 이상의 특성들의 차이를 칭한다. 전형적으로, 활성에 대한 차이는 변형된 활성을 갖는 재조합 숙주 세포와 대응하는 야생형 숙주 세포 사이(예를 들어, 야생형 숙주 세포에 대하여 재조합 숙주 세포의 배양물의 비교)에서 결정된다. 예를 들어, 변형된 활성들은 [예를 들어, 단백질을 인코딩하는 DNA 서열들의 증가되거나 감소된 복제 숫자(number of copies), 단백질을 인코딩하는 mRNA 전사체들의 증가되거나 감소된 숫자, 및/또는 mRNA로부터의 단백질의 단백질 번역(protein translation)의 증가되거나 감소된 양의 결과로서] 재조합 숙주 세포에 의해 발현되는 단백질의 변형된 양; 단백질의 구조 변화(예를 들어, 관측되는 동력학적 파라미터들의 변화, 기질 특이성의 변화를 야기하는 단백질 코딩 서열의 변화와 같은 일차 구조의 변화); 및 단백질 안정성의 변화(예를 들어, 단백질의 증가되거나 감소된 분해)의 결과일 수 있다. 일부 실시예들에서, 폴리펩티드는 본 명세서에 설명된 여하한 폴리펩티드들의 변이체 또는 돌연변이체이다. 여하한 실시예들에서, 본 명세서에 설명된 폴리펩티드들에 대한 코딩 서열들은 특정 숙주 세포 내에서의 발현에 최적화된 코돈이다. 예를 들어, 대장균 내에서의 발현에 대하여, 하나 이상의 코돈들은, 예를 들어 Grosjean et al., Gene 18:199-209(1982)에 설명된 바와 같이 최적화될 수 있다.

[0056] 전형적으로, 본 명세서에서 사용되는 바와 같이 "조절 서열(regulatory sequences)"이라는 용어는 궁극적으로 단백질의 발현을 제어하는 단백질을 인코딩하는 DNA 서열들에 작동가능하게 연결된, DNA 내의 염기들의 서열을 칭한다. 조절 서열의 예시들은 RNA 프로모터 서열, 전사 인자 결합 서열(transcription factor binding sequences), 전사 종결 서열, [증폭자 요소(enhancer elements)와 같은] 전사의 조절인자(modulators), RNA 안정성에 영향을 미치는 뉴클레오티드 서열, 및 번역 조절 서열[이를테면, 리보솜 결합 부위(예를 들어, Shine-Dalgarno sequences in prokaryotes or Kozak sequences in eukaryotes), 개시 코돈, 종결 코돈]을 포함하나, 이로 제한되지 않는다.

[0057] 본 명세서에서 사용되는 바와 같이, "상기 뉴클레오티드 서열의 발현은 야생형 뉴클레오티드 서열에 관하여 변형된다"라는 구는, 내인성 뉴클레오티드 서열의 발현 및/또는 활성, 또는 이중 또는 비-고유한(non-native) 폴리펩티드-인코딩 뉴클레오티드 서열의 발현 및/또는 활성의 수준(level)에 대한 증가 또는 감소를 의미한다. 본 명세서에서 사용되는 바와 같이, "과발현하다"라는 용어는 동일한 조건들 하에서 대응하는 야생형 세포 내에서 정상적으로 발현되는 것보다 더 큰 농도로 세포 내에서 폴리펩티드 또는 폴리뉴클레오티드가 발현되도록 야기시키거나 또는 발현하는 것을 의미한다.

[0058] "발현의 변화된 수준" 및 "발현의 변형된 수준"이라는 용어들은 통용되어 사용되고, 폴리뉴클레오티드, 폴리펩티드 또는 탄화수소가 동일한 조건들 하에서 대응하는 야생형 세포 내에서의 이들의 농도와 비교하여 조작된 숙주 세포 내에서 상이한 농도로 존재한다는 것을 의미한다.

[0059] 본 명세서에서 사용되는 바와 같이, "역가"라는 용어는 숙주 세포 배양의 단위 부피 당 생성된 지방족 알데히드 또는 지방족 알코올의 질량을 칭한다. 본 명세서에 설명된 조성물들 및 방법들의 여하한 측면에서, 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 여하한 두 개의 값들로 제한되는 범위의 역가로 생산된다. 다른 실시예들에서, 지방족 알데히드 또는 지방족 알코올은 100 g/L 초과, 200 g/L 초과, 300 g/L 초과 또는 그 이상, 이를 테면 500 g/L, 700 g/L, 1000 g/L, 1200 g/L, 1500 g/L 또는 2000 g/L의 역가로 생산된다. 본 발명의 방법들에 따라서 재조합 숙주 세포에 의해 생산되는 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.

[0060] 본 명세서에서 사용되는 바와 같이, "숙주 세포에 의해 생산되는 지방족 알데히드 또는 지방족 알코올의 수율"이라는 용어는 숙주 세포 내에서 투입된(input) 탄소 공급원이 생성물(즉, 지방족 알코올 또는 지방족 알데히드)로 전환되는 효율을 칭한다. 본 발명의 방법들에 따라서 지방족 알코올들 및/또는 지방족 알데히드들을 생산하도록 조작된 숙주 세포들은 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 %, 또는 상기 값들 중 여하한 두 개의 값들로 제한되는 범위의 수율을 갖는다. 다른 실시예들에서, 지방족 알데히드 또는 지방족 알코올은 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 % 또는 그 이상을 초과하는 수율로 생산된다. 대안적으로 또는 추가적으로, 수율은 약 30 % 이하, 약 27 % 이하, 약 25 % 이하 또는 약 22 % 이하이다. 따라서, 수율은 상기 종단점(endpoint)들 중 여하한 두 개 값들로 제한될 수 있다. 예를 들어, 본 발명의 방법들에 따라서 재조합 숙주 세포에 의해서 생산되는 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 따라서 재조합 숙주 세포에 의해서 생산되는 지방족 알코올의 바람직한 수율은 10 % 내지 30 %이다.

[0061] 본 명세서에서 사용되는 바와 같이, "생산률"이라는 용어는 숙주 세포 세양의 단위 부피 당 단위 시간 당 생산된 지방족 알데히드 또는 지방족 알코올의 질량을 칭한다. 본 발명에 설명된 조성물들 및 방법들의 여하한 측면에서, 재조합 숙주 세포에 의해 생산된 지방족 알데히드 또는 지방족 알코올의 생산률은 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, 생산률은 2500 mg/L/hour 이하, 2000 mg/L/OD₆₀₀ 이하, 1500 mg/L/OD₆₀₀ 이하, 120 mg/L/hour 이하, 1000 mg/L/hour 이하, 800 mg/L/hour 이하 또는 600 mg/L/hour 이하이다. 따라서, 생산률은 상기 종단점(endpoint)들 중 여하한 두 개 값들로 제한될 수 있다. 예를 들어, 생산률은 3 내지 30 mg/L/hour, 6 내지 20 mg/L/hour, 또는 15 내지 30 mg/L/hour일 수 있다. 본 발명의 방법들에 따라서 재조합 숙주 세포에 의해서 생산되는 지방족 알데히드 또는 지방족 알코올의 바람직한 생산률은 500 mg/L/hour 내지 2500 mg/L/hour, 또는 700 mg/L/hour 내지 2000 mg/L/hour로부터 선택된다.

- [0062] "총 지방 종류" 및 "총 지방산 생성물"이라는 용어들은, 국제 특허 출원 공개 WO 2008/119082에 설명된 것처럼 GC-FID에 의해 평가되는 바와 같이, 샘플 내에 존재하는 지방족 알코올, 지방족 알데히드, 유리 지방산 및 지방족 에스테르의 총량과 관련하여 본 명세서에서 통용되어 사용될 수 있다. 샘플들은 문맥에 따라서 한 가지, 두 가지, 세 가지 또는 네 가지의 이러한 화합물들을 포함할 수 있다.
- [0063] 본 명세서에서 사용되는 바와 같이, "글루코오스 이용률"이라는 용어는 그램/리터/시간(g/L/hr)으로 보고되는, 단위 시간 당 배양에 사용되는 글루코오스의 양을 의미한다.
- [0064] 본 명세서에서 사용되는 바와 같이, "탄소 공급원"이라는 용어는 원핵 또는 단순한 진핵 세포 성장을 위한 탄소 공급원으로서 사용되는데 적합한 화합물 또는 기질을 칭한다. 탄소 공급원은 중합체, 탄수화물, 산(acid), 알코올, 알데히드, 케톤, 아미노산, 펩티드 및 기체(예를 들어, CO 및 CO₂)를 포함하나 이로 제한되지 않는 다양한 형태로 존재할 수 있다. 예시적인 탄소 공급원은 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; 숙시네이트, 락테이트 및 아세테이트와 같은 포화 또는 불포화 지방산류; 에탄올, 메탄올 및 글리세롤과 같은 알코올류, 또는 이들의 혼합물들을 포함한다. 또한, 탄소 공급원은 글루코오스와 같은, 광합성의 산물일 수 있다. 여하한 바람직한 실시예들에서, 탄소 공급원은 바이오매스이다. 다른 바람직한 실시예들에서, 탄소 공급원은 수크로오스이다.
- [0065] 본 명세서에서 사용되는 바와 같이, "바이오매스"라는 용어는 탄소 공급원으로부터 유래되는 여하한 생물학적 물질을 칭한다. 일부 실시예들에서, 바이오매스는 탄소 공급원으로 처리(process)되고, 이는 생물전환에 적합하다. 다른 실시예들에서, 바이오매스는 탄소 공급원으로서의 처리를 더 요구하지 않는다. 탄소 공급원은 바이오연료로 전환될 수 있다. 예시적인 바이오매스의 공급원은 옥수수, 사탕수수 또는 스위치그래스(switchgrass)와 같은 식물성 물질 또는 식물(vegetation)이다. 또 다른 예시적인 바이오매스의 공급원은 동물성 물질[예를 들어, 우분(cow manure)]과 같은 신진대사 노폐물(metabolic waste products)이다. 추가 예시적인 바이오매스의 공급원은 조류 및 여타 수산 식물을 포함한다. 또한, 바이오매스는 발효 찌꺼기, 목초(ensilage), 짚, 재목, 오수(sewage), 쓰레기, 셀룰로오스성 대도시 폐기물(cellulosic urban waste) 및 남은 음식들(food leftovers)을 포함하나 이로 제한되지 않는 공업, 농업, 임업 및 가정으로부터의 폐기물(waste products)을 포함한다. 또한, "바이오매스"라는 용어는 탄수화물(예를 들어, 단당류, 이당류 또는 다당류)들과 같은, 탄소 공급원들을 칭할 수 있다.
- [0066] 본 명세서에서 사용되는 바와 같이, (지방산 및 이의 유도체와 같은) 생성물들에 대하여 "분리된"이라는 용어는 세포 성분(cellular components), 세포 배양 배지, 또는 화학적 또는 합성 전구체로부터 분리되는 생성물들을 칭한다. 본 명세서에 설명된 방법들에 의해 생산되는 지방산 및 이의 유도체는 세포질뿐만 아니라 발효액(발효액) 내에서 상대적으로 혼합되지 않을 수 있다. 따라서, 지방산 및 이의 유도체는 세포 내 또는 세포 외의 유기상(organic phase)에 수집될 수 있다.
- [0067] 본 명세서에서 사용되는 바와 같이, "정제한다", "정제된" 또는 "정제"라는 용어들은, 예를 들어 분리(isolation) 또는 구분(separation)에 의해 주변 환경으로부터 분자의 이동(removal) 또는 분리를 의미한다. "실질적으로 정제된" 분자는 결합되는 다른 성분들로부터 적어도 약 60 % (예를 들어, 적어도 약 70 %, 적어도 약 75 %, 적어도 약 85 %, 적어도 약 90 %, 적어도 약 95 %, 적어도 약 97 %, 적어도 약 99 %) 유리된다(free). 본 명세서에서 사용되는 바와 같이, 또한 이러한 용어들은 샘플로부터 오염물질들의 제거를 칭한다. 예를 들어, 오염물질들의 제거는 샘플 내에서 지방족 알데히드 또는 지방족 알코올의 비율(percentage)의 증가를 유발할 수 있다. 예를 들어, 지방족 알데히드 또는 지방족 알코올이 재조합 숙주 세포에서 생산되는 경우에, 상기 지방족 알데히드 또는 지방족 알코올은 재조합 숙주 세포 단백질의 제거에 의해 정제될 수 있다. 정제 후, 샘플 내에서 지방족 알데히드 또는 지방족 알코올의 비율이 증가된다. "정제한다", "정제된" 및 "정제"라는 용어들은 절대 순도(absolute purity)를 요구하지 않는 상대적인 용어들이다. 따라서, 예를 들어 지방족 알데히드 또는 지방족 알코올이 재조합 숙주 세포에서 생산되는 경우에, 정제된 지방족 알데히드 또는 지방족 알코올은 다른 세포 성분(예를 들어, 핵산, 폴리펩티드, 지질, 탄수화물 또는 기타 탄화수소)로부터 실질적으로 분리되는 지방족 알데히드 또는 지방족 알코올이다.

[0068] **균주 개량**

- [0069] 지방족 알코올들의 역가, 수율 및/또는 생산물에 대한 매우 높은 목표를 충족시키기 위하여, 생산 숙주 세포에 다수의 변형들 가해졌다. FadR는 지방산 분해 및 지방산 생합성 경로들에 관련된 주요한 조절 인자이다[Cronan et al., *Mol. Microbiol.*, 29(4):937-943(1998)]. *대장균* ACS 효소인 FadD 및 지방산 수송 단백질인 FadL은 지방산 흡수 시스템의 필수적인 구성요소들이다. FadL는 지방산의 박테리아성 세포로의 수송을 매개(mediate)하고, FadD는 아실-CoA 에스테르의 형성을 매개한다. 다른 탄소 공급원이 이용가능하지 않는 경우, 외인성 지방산이 박테리아에 의해 흡수(taken up)되고, 아실-CoA 에스테르로 전환되며, 이는 전사 인자인 FadR에 결합할 수 있고, 지방산 수송(FadL), 활성화(FadD) 및 β -산화(FadA, FadB, FadE 및 FadH)에 책임이 있는 단백질들을 인코딩하는 *fad* 유전자들의 발현을 저하(depress)시킬 수 있다. 대안적인 탄소 공급원을 이용가능한 경우, 박테리아는 아실-ACP로서 지방산을 합성하고, 이는 β -산화에 대한 기질이 아니라, 인지질 합성에 사용된다. 따라서, 아실-CoA 및 아실-ACP 둘 모두는 상이한 최종 생성물을 야기할 수 있는 지방산들의 독립적인 공급원들이다[Caviglia et al., *J. Biol. Chem.*, 279(12):1163-1169(2004)]. 미국 가출원 No. 61/470,989는 숙주 세포 내에서 지방산 유도체들을 생산하는 개선된 방법들을 기재하며, 상기 숙주 세포는 대응하는 야생형 숙주 세포 내에서의 FadR 폴리펩티드의 발현 수준과 비교해서 FadR 폴리펩티드의 변화된 발현 수준을 갖도록 유전적으로 조작된다.
- [0070] *대장균*과 같은 숙주 세포에서 지방산 생합성을 제한시킬 수 있는 인자들에 관하여 본 기술분야에는 상충하는 공론(speculation)들이 존재한다. 지방산 생합성을 통해 플럭스(flux)를 증가시키는 한가지 접근법은 상기 지방산 생합성 경로에 대한 다양한 효소들을 조작(manipulate)하는 것이다(도 1 및 도 2 참조). 지방산 생합성(fab) 경로 및 아세틸-CoA 카르복실라아제(acc) 복합체(도 3)를 통하여 아세틸-CoA로부터의 아실-ACPs의 공급은, 지방족 알코올 생산 속도를 제한할 수 있다. 실시예 2에 상세하게 설명된 일 예시적인 접근법에서, *코리네박테리움 글루타미쿰*(*Corynebacterium glutamicum*) accABCD ($\pm$ birA)의 과발현 효과는 이러한 유전자 변형들이 *대장균*에서 아세틸-CoA 및 말로닐-CoA의 증가로 이어질 수 있다는 것을 증명한다. 지방산 생합성을 통한 플럭스의 낮은 속도(low rate)에 대하여 가능성 있는 한가지 이유는 지방산 생합성을 위한 주요한 전구체들, 그리고 전구체들, 즉 아세틸-CoA 및, 특히 말로닐-CoA의 제한된 공급이다. 실시예 3은 *대장균* 숙주 세포의 염색체 내로의 통합(integration) 및 말로닐-CoA의 아실-ACPs로의 전환을 위한 생합성 경로에 대한 효소들을 인코딩하는 fab 오페론들의 구성(construction)을 설명한다. 실시예 4에 상세하게 설명된 또 다른 접근법에서, *대장균* 숙주 세포 내에서의 rph 및 ilvG 유전자들의 돌연변이는 더 많은(higher) 유리 지방산(FFA) 생산을 야기하는 것으로 나타났고, 이는 고급(higher) 지방족 알코올의 생산으로 번역된다. 또 다른 접근법에서, 트랜스포존 돌연변이 유발 및 고속-대량 스크리닝(high-throughput screening)은 역가 또는 수율을 증가시키는 유익한 돌연변이들을 찾기 위하여 실시된다. 실시예 5는 어떻게 yijP 유전자 내의 트랜스포존 삽입이 진탕 플라스크(shake flask) 및 유가 배양식 발효(fed-batch fermentations)에서 지방족 알코올 수율을 개선시킬 수 있는지를 설명한다.
- [0071] **카르복실산 환원효소(CAR)**
- [0072] 제조합 숙주 세포들은 아실-ACPs의 유리 지방산(FFAs)으로의 전환을 촉진시키는 티오에스테라아제, 및 유리 지방산을 지방족 알데히드로 전환시키는 카르복실산 환원효소(CAR)를 발현시킴으로써 지방족 알코올을 생산하도록 조작된다. 숙주 세포(예를 들어, *대장균*) 내에 존재하는 본래의 (내인성) 알데히드 환원효소는 지방족 알데히드를 지방족 알코올로 전환시킬 수 있다. 예를 들어, 예시적인 티오에스테라아제는 미국 특허 출원 No. 20100154293에 기재되고, 이는 특허 본 명세서에서 인용참조된다. CarB는 예시적인 카르복실산 환원효소이며, 지방족 알코올 생산 경로에서 주요한 효소이다. WO 2010/062480은 쿼리 서열(query sequence)로서 NRRL 5646 CAR 아미노산 서열(Genpept accession AAR91681)(SEQ ID NO: 6)을 사용하는 BLAST 검색(search), 및 대략 20 개의 상동성 서열들의 식별을 위한 상기 검색의 사용을 설명한다.
- [0073] "카르복실산 환원효소", "CAR" 및 "지방족 알데히드 생합성 폴리펩티드"라는 용어들은 본 명세서에서 통용되어 사용된다. 본 발명을 실시함에 있어서, 카르복실산 환원효소 폴리펩티드를 인코딩하는 유전자는 숙주 세포 내에서 발현 또는 과발현된다. 일부 실시예들에서, CarB 폴리펩티드는 SEQ ID NO: 7의 아미노산 서열을 갖는다. 다른 실시예들에서, CarB 폴리펩티드는 SEQ ID NO: 7의 변이체 또는 돌연변이체이다. 여하한 실시예들에서, CarB 폴리펩티드는 포유동물 세포, 식물 세포, 곤충 세포, 효모 세포, 균류 세포, 곰팡이 세포, 박테리아 세포, 또는 다른 유기체로부터 파생된다. 일부 실시예들에서, 박테리아 세포는 *마이코박테리움 스메그마티스*(*Mycobacterium smegmatis*), *마이코박테리움 엡서수스*(*Mycobacterium abscessus*), *마이코박테리움 아비움*(*Mycobacterium avium*), *마이코박테리움 보비스*(*Mycobacterium bovis*), *마이코박테리움 투베르쿨로시스*(*Mycobacterium tuberculosis*), *마이코박테리움 레프래*(*Mycobacterium leprae*), *마이코박테리움 마리오*

(*Mycobacterium marinum*), 및 마이코박테리움 울세란스(*Mycobacterium ulcerans*)로 이루어진 그룹으로부터 선택되는 마이코박테리움(*mycobacterium*)이다. 다른 실시예들에서, 박테리아 세포는 노카르디아(*Nocardia*) 종, 예를 들어 노카르디아 NRRL 5646, 노카르디아 파르시니카(*Nocardia farcinica*), 스트렙토마이시스 그리세우스(*Streptomyces griseus*), 사리니스포라 아레니콜라(*Salinispora arenicola*), 또는 클라라비박테르 미치가네네시스(*Clavibacter michiganensis*)로부터 파생된다. 다른 실시예들에서, CarB 폴리펩티드는 SEQ ID NO: 7의 아미노산 서열에 대하여 적어도 약 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % 또는 99 % 동일한 아미노산 서열을 갖는 CarB의 동족체(homologue)이다. SEQ ID NO: 7의 아미노산 서열에 대하여 적어도 약 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % 또는 99 % 동일성을 갖는 CarB 폴리펩티드의 동일성은 특히 제한되지 않고, 본 기술분야의 당업자들은 대장균 MG1655 유래된-CarB의 동족체들을 쉽게 식별할 수 있으며, 본 명세서에 설명되는 방법들을 이용하여 이들의 기능을 결정할 수 있다. 다른 실시예들에서, CarB 폴리펩티드는 SEQ ID NO: 7의 아미노산 번호(amino acid number) 3, 12, 20, 28, 46, 74, 103, 191, 288, 473, 827, 926, 927, 930 또는 1128에서 돌연변이를 함유한다. 예시적인 돌연변이들은 표 10에 상세하게 설명된다. 폴리펩티드의 바람직한 분절들 또는 돌연변이체들은 대응하는 야생형 폴리펩티드의 생물학적 기능(예를 들어, 효소 활성) 중 일부 또는 전부를 유지한다. 일부 실시예들에서, 분절 또는 돌연변이체는 대응하는 야생형 폴리펩티드의 생물학적 기능 중 적어도 약 75 %, 적어도 약 80 %, 적어도 약 90 %, 적어도 약 95 % 또는 적어도 약 98 % 또는 그 이상을 유지한다. 다른 실시예들에서, 분절 또는 돌연변이체는 대응하는 야생형 폴리펩티드의 생물학적 기능의 약 100 %를 유지한다. 아미노산 잔기들이 생물학적 활성에 영향을 미치지 않고 치환, 삽입 또는 제거(deleted)될 수 있는지를 결정하기 위한 지침(Guidance)은, 본 기술분야에 잘 알려져 있는 컴퓨터 프로그램, 예를 들어 LASERGENE™ 소프트웨어(DNASTAR, Inc., Madison, WI)를 사용하여 알아낼 수 있다.

[0074] 또 다른 실시예들에서, 분절 또는 돌연변이체는 대응하는 야생형 폴리펩티드와 비교하여 증가된 생물학적 기능을 나타낸다. 예를 들어, 분절 또는 돌연변이체는 대응하는 야생형 폴리펩티드와 비교하여 효소 활성에 있어서 적어도 약 10 %, 적어도 약 25 %, 적어도 약 50 %, 적어도 약 75 % 또는 적어도 약 90 % 개선을 나타낼 수 있다. 다른 실시예들에서, 분절 또는 돌연변이체는 대응하는 야생형 폴리펩티드와 비교하여 효소 활성에 있어서 적어도 약 100 %(예를 들어, 적어도 약 200 % 또는 적어도 약 500 %) 개선을 나타낸다. 본 명세서에서 설명되는 폴리펩티드는 폴리펩티드 기능에 실질적인 영향을 주지 않는 추가적인 보존성 또는 비-필수 아미노산 치환들을 가질 수 있는 것으로 이해된다. 특정 치환이 허용되는지(즉, DNA 결합 또는 효소 활성과 같은 원하는 생물학적 기능에 악영향을 주지 않는지)의 여부는 Bowie et al.[*Science*, 247:1306-1310(1990)]에 기재된 바와 같이 결정될 수 있다.

[0075] 본원 발명의 변형 효소들 및 방법들의 결과로서, CarB 폴리펩티드의 변화된 발현 수준을 갖는 조작된 숙주 세포에 의해 생산되는 지방산 또는 이의 유도체의 역가, 수율 및/또는 생산물 중 하나 이상은 대응하는 야생형 숙주 세포의 역가, 수율 및/또는 생산물에 상대적으로 증가된다. C12 및 C14 지방산들의 지방족 알코올들로의 최대 전환을 위해 감안한다면, CarB는 충분한 활성으로 발현되어야 한다. 개선된 재조합 숙주 세포는, 예를 들어 대장균 염색체로부터 발현되는 CAR 효소를 갖는다. 실시예 6에 나타난 바와 같이, 상기 염색체로부터 CarB 효소를 발현시키는 세포들은 원래(original) CarB에 상대적으로 더 큰 카르복실산 환원효소 활성을 가지며, 더 많은 C12 및 C14 지방산들을 지방족 알코올들로 전환시킬 수 있다. CarB는 큰(large) 유전자(3.5 kb)이고 플라스미드 크기를 상당히 증가시켜, 균주 개발을 하는 동안 새로운 유전자들을 테스트하는데 pCL 플라스미드를 사용하는 것을 어렵게 만든다. CarB의 활성을 증가시키는 접근법들은 이의 용해도, 안정성, 발현 및/또는 기능(functionality)을 증가시키는 것을 포함한다. 일 예시적인 접근법에서, CarB의 N-말단(terminus)에 트롬빈 절단 부위(cleavage site) 및 6 히스티딘들을 함유하는 융합 단백질(fusion protein)이 생산된다. 이 효소는 N-말단에 추가적인 60 개의 뉴클레오티드들에 의해 CarB와 상이하며, CarB60으로 명명된다. CarB 또는 CarB60이 pTRC 프로모터의 제어 하에서 대장균 염색체로부터 발현되는 경우, CarB60을 함유하는 세포들은 총 세포의 카르복실산 환원효소 활성을 증가시켰고, 더 많은 C12 및 C14 유리 지방산(FFAs)들을 지방족 알코올들로 전환시켰다. 본 기술분야의 당업자는 (상기예) 실시예 6에 설명되는 바와 같이 더 많은 C12 및 C14 유리 지방산(FFAs)들의 지방족 알코올들로의 전환을 달성하기 위하여 조작되는 분자의 일 예시라는 점을 인식할 것이다. 유사한 접근법들이 본 명세서에 포함된다(실시예 7를 참조).

[0076] 포스포판테테인 전달효소(Phosphopantetheine transferases: PPTases)[E.C. 2.7.8.7]는 4'-포스포판테테인을 CoA에서 기질로의 이동(transfer)을 촉진시킨다. 노카르디아 Car, CarB 및 이들의 몇몇 동족체들은 4'-포스포판테테인(PPT)에 대한 추정되는(putative) 부착 부위를 포함한다[He et al., *Appl. Environ. Microbiol.*,

70(3):1874-1881(2004)]. 본 발명의 일부 실시예들에서, PPTase는 조작된 숙주 세포 내에서 발현 또는 과발현된다. 여하한 실시예들에서, PPTase는 *대장균* MG1655로부터의 EntD(SEQ ID NO: 8)이다. 일부 실시예들에서, 티오에스테라아제 및 카르복실산 환원효소는 조작된 숙주 세포 내에서 발현 또는 과발현된다. 여하한 실시예들에서, 티오에스테라아제는 tesA이고, 카르복실산 환원효소는 carB이다. 다른 실시예들에서, 티오에스테라아제, 카르복실산 환원효소 및 알코올 탈수소효소는 조작된 숙주 세포 내에서 발현 또는 과발현된다. 여하한 실시예들에서, 티오에스테라아제는 tesA이고, 카르복실산 환원효소는 carB이며, 알코올 탈수소효소는 *아시네토박터 바일리*(*Acinetobacter baylyi*) ADP1로부터의 alrAadp1(GenPept 수탁 번호 CAG70248.1)(SEQ ID NO: 4)이다. 또 다른 실시예들에서, 티오에스테라아제, 카르복실산 환원효소, PPTase 및 알코올 탈수소효소는 조작된 숙주 세포 내에서 발현 또는 과발현된다. 여하한 실시예들에서, 티오에스테라아제는 tesA이고, 카르복실산 환원효소는 carB이며, PPTase는 entD이고, 알코올 탈수소효소는 alrAadp1이다. 또 다른 실시예들에서, 티오에스테라아제, CAR, PPTase 및 알코올 탈수소효소 중 하나 이상을 발현시키는 변형된 숙주 세포도 하나 이상의 균주 개선들을 갖는다. 예시적인 균주 개선들은 아세틸-CoA 카르복실라아제 폴리펩티드의 발현 또는 과발현, FadR 폴리펩티드의 과발현, 이중 iFAB 오페론(operon)의 발현 또는 과발현, 또는 yijP 유전자 또는 또 다른 유전자 내의 트랜스포존 삽입(transposon insertion), 또는 유사한 접근법들을 포함하나, 이로 제한되지 않는다. 또한, 본 발명은 본 명세서에서 설명되는 여하한 방법들에 의해 생산되는 지방족 알코올 조성물을 제공한다. 본 명세서에서 설명되는 여하한 방법들에 의해 생산되는 지방족 알코올 조성물은 다른 화합물들(예를 들어, 중합체, 계면활성제, 플라스틱, 직물, 용제, 접착제 등), 또는 개인 위생 용품의 첨가물(personal care additive)의 생산을 위한 출발 물질(starting materials)로서 직접적으로 사용될 수 있다. 또한, 이러한 화합물들은 다른 생성물들을 만들기 위하여 후속 반응들, 예를 들어 수소화, [예를 들어, 수소화, 열분해(pyrolysis), 또는 둘 모두를 통한] 촉매 크래킹(catalytic cracking)을 위한 공급원료로서 사용될 수 있다.

[0077] 돌연변이체 또는 변이체

[0078] 일부 실시예들에서, 재조합 숙주 세포 내에서 발현되는 폴리펩티드는 본 명세서에서 설명되는 여하한 폴리펩티드들의 돌연변이체 또는 변이체이다. 본 명세서에서 사용되는 바와 같은 "돌연변이체" 및 "변이체"라는 용어들은 적어도 하나의 아미노산에 의해 야생형 폴리펩티드와 상이한 아미노산 서열을 갖는 폴리펩티드를 칭한다. 예를 들어, 돌연변이체는 하기의 보존성 아미노산 치환체들 중에서 하나 이상을 포함한다: 알라닌, 발린, 류신 및 이소류신과 같은 지방성(aliphatic) 아미노산의 또 다른 지방성(aliphatic) 아미노산으로의 대체물(replacement); 세린의 트레오닌으로의 대체물; 트레오닌의 세린으로의 대체물; 아스파르트산 및 글루탐산과 같은 산성 잔기의 또 다른 산성 잔기로의 대체물; 아스파라긴 및 글루타민과 같은 아미드기를 보유하는 잔기의 아미드기를 보유하는 또 다른 잔기로의 대체물; 리신 및 아르기닌과 같은 염기성 잔기의 또 다른 염기성 잔기로의 교환물(exchange); 그리고 페닐알라닌 및 티로신과 같은 방향족 잔기의 또 다른 방향족 잔기로의 대체물. 일부 실시예들에서, 돌연변이형 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 치환들, 부가들, 삽입들 또는 결실들을 갖는다. 폴리펩티드의 바람직한 분절들 또는 돌연변이체들은 대응하는 야생형 폴리펩티드의 생물학적 기능(예를 들어, 효소 활성) 중 일부 또는 전부를 유지한다. 일부 실시예들에서, 분절 또는 돌연변이체는 대응하는 야생형 폴리펩티드의 생물학적 기능 중 적어도 약 75 %, 적어도 약 80 %, 적어도 약 90 %, 적어도 약 95 % 또는 적어도 약 98 % 또는 그 이상을 유지한다. 다른 실시예들에서, 분절 또는 돌연변이체는 대응하는 야생형 폴리펩티드의 생물학적 기능의 약 100 %를 유지한다. 아미노산 잔기들이 생물학적 활성에 영향을 미치지 않고 치환, 삽입 또는 결실될 수 있는지를 결정하기 위한 지침(Guidance)은, 본 기술분야에 잘 알려져 있는 컴퓨터 프로그램, 예를 들어 LASERGENE™ 소프트웨어(DNASTAR, Inc., Madison, WI)를 사용하여 알아낼 수 있다.

[0079] 또 다른 실시예들에서, 분절 또는 돌연변이체는 대응하는 야생형 폴리펩티드와 비교하여 증가된 생물학적 기능을 나타낸다. 예를 들어, 분절 또는 돌연변이체는 대응하는 야생형 폴리펩티드와 비교하여 효소 활성에 있어서 적어도 10 %, 적어도 25 %, 적어도 50 %, 적어도 75 % 또는 적어도 90 % 개선을 나타낼 수 있다. 다른 실시예들에서, 분절 또는 돌연변이체는 대응하는 야생형 폴리펩티드와 비교하여 효소 활성에 있어서 적어도 100 %(예를 들어, 적어도 200 % 또는 적어도 500 %) 개선을 나타낸다. 본 명세서에서 설명되는 폴리펩티드는 폴리펩티드 기능에 실질적인 영향을 주지 않는 추가적인 보존성 또는 비-필수 아미노산 치환들을 가질 수 있는 것으로 이해된다. 특정 치환이 허용되는지(즉, 카르복실산 환원효소 활성과 같은 원하는 생물학적 기능에 부정적인 영향을 주지 않는지)의 여부는 Bowie et al.[*Science*, 247:1306-1310(1990)]에 기재된 바와 같이 결정될 수 있다. 보존성 아미노산 치환은 아미노산 잔기가 유사한 측쇄(side chain)를 가지는 아미노산 잔기로 대체되는 것이다. 유사한 측쇄들을 가지는 아미노산 잔기의 집단(Family)들은 본 기술분야에서 정의되었다. 이러한 집단들은 염기성 측쇄들을 갖는 아미노산류(예를 들어, 리신, 아르기닌, 히스티딘), 산성 측쇄들을 갖는 아미노산

류(예를 들어, 아스파르트산, 글루탐산), 하전되지 않은(uncharged) 극성 측쇄들을 갖는 아미노산류(예를 들어, 글리신, 아스파라긴, 글루타민, 세린, 트레오닌, 티로신, 시스테인), 비극성 측쇄들을 갖는 아미노산류(예를 들어, 알라닌, 발린, 류신, 이소류신, 프롤린, 페닐알라닌, 메티오닌, 트립토판), 베타-분지형 측쇄들을 갖는 아미노산류(예를 들어, 트레오닌, 발린, 이소류신) 및 방향족 측쇄들을 갖는 아미노산류(예를 들어, 티로신, 페닐알라닌, 트립토판, 히스티딘)를 포함한다. 변이체들은 자연적으로 발생되거나 실험관 내에서(*in vitro*) 생성될 수 있다. 특히, 이러한 변이체들은 특정부위 돌연변이 유발(site directed mutagenesis), 무작위 화학적 돌연변이 유발(random chemical mutagenesis), 핵산말단가수분해효소(Exonuclease) III 결실 절차들 또는 표준 클로닝 기술들과 같은 유전 공학 기술들을 사용하여 생성될 수 있다. 대안적으로, 이러한 변이체들, 분절들, 유사체들 또는 유도체들은 화학적 합성 또는 변형 절차들을 사용하여 생성될 수 있다.

[0080]

변이체들의 제조 방법들은 본 기술분야에 잘 알려져 있다. 상기 방법들은 산업적 또는 실험적 적용들에 있어서 이들의 가치를 증대시키는 특성들을 가지는 폴리펩티드들을 인코딩하는 핵산들을 산출하기 위하여, 자연 분리물들(natural isolates)로부터 얻어진 핵산 서열들을 변형시키는 절차들이 포함된다. 이러한 절차들에서, 자연 분리물들로부터 얻어진 서열에 관하여 하나 이상의 뉴클레오티드 차이를 갖는 다수의 변이체 서열들이 산출되고 특성화된다. 전형적으로, 이러한 뉴클레오티드 차이는 자연 분리물들로부터의 핵산들에 의해 인코딩되는 폴리펩티드들에 관하여 아미노산 변화들을 야기한다. 예를 들어, 변이체들은 무작위 및 특정부위 돌연변이 유발을 사용함으로써 제조될 수 있다. 무작위 및 특정부위 돌연변이 유발은, 예를 들어 Arnold, *Curr. Opin. Biotech.*, 4:450-455(1993)에 기재되어 있다. 무작위 돌연변이 유발은 오류 빈발(error prone) PCR을 사용하여 달성될 수 있다[Leung et al., *Technique*, 1:11-15(1989); 및 Caldwell et al., *PCR Methods Applic.*, 2:28-33(1992) 참조]. 오류 빈발 PCR에서, PCR은 DNA 중합효소의 복제 정확도(copying fidelity)가 낮은 조건 하에 수행되고, 높은 비율의 점 돌연변이(point mutation)가 PCR 생성물의 전장(entire length)을 따라서 얻어진다. 간단히 말해, 이러한 절차들에서, 돌연변이 유발되는 핵산들(예를 들어, 카르복실릭 환원효소 효소를 인코딩하는 폴리뉴클레오티드 서열)은 PCR 생성물의 전장을 따라서 높은 비율의 점 돌연변이를 달성하기 위하여 PCR 프라이머들, 반응 완충액, MgCl₂, MnCl₂, Taq 중합효소 및 적절한 농도의 dNTP들과 혼합된다. 예를 들어, 반응이 20 fmole의 돌연변이 유발되는 핵산, 30 pmole의 각각의 PCR 프라이머, 50 mM KCl, 10 mM Tris HCl(pH 8.3) 및 0.01 % 젤라틴을 포함하는 반응 완충액, 7 mM MgCl₂, 0.5 mM MnCl₂, 5 단위(unit)의 Taq 중합효소, 0.2 mM dGTP, 0.2 mM dATP, 1 mM dCTP, 그리고 1 mM dTTP를 사용하여 수행될 수 있다. PCR은 94 °C에서 1분, 45 °C에서 1분, 그리고 72 °C에서 1분의 30회 주기 동안 수행될 수 있다. 그러나, 이러한 파라미터들은 적절하게 변화될 수 있는 것으로 인지될 것이다. 이후에, 돌연변이 유발된 핵산들은 적절한 벡터 내로 클로닝되고, 상기 돌연변이 유발된 핵산들에 의해 인코딩되는 폴리펩티드들의 활성이 평가된다(실시예 7 참조). 특정부위 돌연변이 유발은 여하한 클로닝되는 목적 DNA에서 부위-특이적인 돌연변이들을 발생시키는 올리고뉴클레오티드 지향된 돌연변이 유발을 사용하여 달성될 수 있다. 올리고뉴클레오티드 돌연변이 유발은, 예를 들어 Reidhaar-Olson et al., *Science*, 241:53-57(1988)에 기재된다. 간단히 말해서, 이러한 절차들에서 클로닝된 DNA 내로 도입되는 하나 이상의 돌연변이들을 보유하는 복수의 이중 가닥 올리고뉴클레오티드들이 합성되고, 돌연변이 유발되는 클로닝된 DNA(예를 들어, CAR 폴리펩티드를 인코딩하는 폴리뉴클레오티드 서열) 내로 삽입된다. 돌연변이 유발된 DNA를 함유하는 클론들은 회수되고, 이들이 인코딩하는 폴리펩티드들의 활성이 평가된다. 변이체들을 발생시키기 위한 또 다른 방법은 조립(assembly) PCR이다. 조립 PCR은 소형 DNA 분절들의 혼합물로부터 PCR 생성물의 조립을 포함한다. 다수의 상이한 PCR 반응들이 동일한 바이알 내에서 병행하여 발생하는데, 한 반응의 생성물들이 또 다른 반응의 생성물들을 기폭시킨다. 예를 들어, 조립 PCR은 미국 특허 5,965,408에 기재된다. 변이체들을 발생시키는 또 다른 방법은 유성(sexual) PCR 돌연변이 유발이다. 유성 PCR 돌연변이 유발에서, 강제된 상동성 재조합(forced homologous recombination)은 서열 상동성을 기반으로 하는 DNA 분자의 무작위 분절화(random fragmentation)의 결과로써, 실험관 내에서 상이하지만 고도로 관련된 DNA 서열들의 DNA 분자들 사이에서 발생한다. 그 이후에, PCR 반응에서 프라이머 신장에 의한 교차(crossover)의 고정이 수행된다. 유성 PCR 돌연변이 유발은, 예를 들어 Stemmer, *Proc. Natl. Acad. Sci., U.S.A.*, 91:10747-10751(1994)에 기재된다.

[0081]

또한, 변이체들은 생체 내에서 돌연변이 유발에 의해 생성될 수 있다. 일부 실시예들에서, 핵산 서열 내의 무작위 돌연변이는 하나 이상의 DNA 복원 경로(repair pathway)들에서 돌연변이들을 수송(carry)하는 박테리아 균주, 이를 테면 대장균 균주에서 서열을 증식시킴으로써 발생된다. 이러한 "돌연변이 유발 유전자(mutator)" 균주들은 야생-형 균주에서보다 높은 무작위 돌연변이율을 갖는다. 궁극적으로, 이러한 균주들 중의 한 가지에서 DNA 서열(예를 들어, CAR 폴리펩티드를 인코딩하는 폴리뉴클레오티드 서열)의 증식은 DNA 내에서 무작위 돌연변이

이들을 발생시킬 것이다. *생체 내에서의* 돌연변이 유발에 사용하기 적합한 돌연변이 유발 유전자 균주들은, 예를 들어 국제 특허 출원 공개 No. WO 1991/16427에 기재된다. 또한, 변이체들은 카세트 돌연변이 유발(cassette mutagenesis)을 사용하여 발생될 수 있다. 카세트 돌연변이 유발에서, 이중 가닥 DNA 분자의 작은 영역이 본래의 서열과 상이한 합성 올리고뉴클레오타이드 "카세트"로 대체된다. 종종, 상기 올리고뉴클레오타이드는 완전하게 맞/또는 부분적으로 무작위화된 본래의 서열을 포함한다. 또한, 순환 앙상블 돌연변이 유발(Recursive ensemble mutagenesis)은 변이체들을 발생시키는데 사용될 수 있다. 순환 앙상블 돌연변이 유발은 아미노산 서열과 상이하고 표현형적으로 관련된 돌연변이체들의 다양한 개체군들을 생산하기 위하여 개발된, 단백질 조작(즉, 단백질 돌연변이 유발)을 위한 알고리즘이다. 이러한 방법은 조합 카세트 돌연변이 유발의 연속 라운드들(successive rounds)을 제어하기 위하여 피드백 기전을 사용한다. 순환 앙상블 돌연변이 유발은, 예를 들어 Arkin et al., *Proc. Natl. Acad. Sci., U.S.A.*, 89:7811-7815(1992)에 기재된다. 일부 실시예들에서, 변이체들은 지수 앙상블(exponential ensemble) 돌연변이 유발을 사용하여 생성된다. 지수 앙상블 돌연변이 유발은 높은 비율의 독특한 기능성 돌연변이체들을 포함하는 조합 라이브러리들(combinatorial libraries)을 발생시키기 위한 과정이며, 여기서 잔기들의 작은 기들은 기능성 단백질들을 야기하는 각각의 변화된 위치에서 아미노산을 확인하기 위하여 동시에 무작위화 된다. 지수 앙상블 돌연변이 유발은, 예를 들어 Delegrave et al., *Biotech. Res.*, 11:1548-1552(1993)에 기재된다. 일부 실시예들에서, 변이체들은 재편성 절차(shuffling procedure)들을 사용하여 생성되며, 여기서 별개의 폴리펩티드들을 인코딩하는 복수의 핵산들의 일부분들이, 예를 들어 미국 특허 5,965,408 및 5,939,250에 개시된 바와 같이 키메라 폴리펩티드들을 인코딩하는 키메라 핵산 서열들과 함께 융합된다.

[0082] 삽입 돌연변이 유발(Insertional mutagenesis)은 하나 이상의 염기들의 삽입에 의한 DNA의 돌연변이 유발이다. 삽입 돌연변이는 자연적으로 발생할 수 있거나, 바이러스 또는 트랜스포존에 의해 매개될 수 있거나, 또는 실험실에서 연구 목적을 위하여, 예를 들어 트랜스포존 돌연변이 유발에 의해 인공적으로 생성시킬 수 있다. 외인성 DNA가 숙주의 DNA 내로 통합되는 경우, 여하한지 계속되는 돌연변이의 중대도(severity)는 전적으로 DNA가 삽입되는 숙주의 게놈 내의 위치에 달려있다. 예를 들어, 트랜스포존이 필수 유전자의 중간, 프로모터 영역, 또는 리프레서(repressor) 또는 증폭자 영역(enhancer region) 내에 삽입된다면, 상당한 영향이 분명 있을 수 있다. 트랜스포존 돌연변이 유발 및 고속-대량 스크리닝은 지방족 알코올의 역가 또는 수율을 증가시키는 유익한 돌연변이체들을 찾아내기 위하여 실시된다. 본 발명은 (a) SEQ ID NO: 7의 아미노산 서열에 대하여 적어도 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % 또는 99 % 서열 동일성을 갖는 아미노산 서열을 포함하는 카르복실산 환원효소를 인코딩하는 폴리뉴클레오타이드 서열, 및 (b) 카르복실산 환원효소 활성을 갖는 폴리펩티드를 인코딩하는 폴리뉴클레오타이드를 포함하는 재조합 숙주 세포들을 제공하고, 여기서 상기 재조합 숙주 세포는 지방족 알데히드 또는 지방족 알코올을 생산할 수 있다.

[0083] **숙주 세포의 조작**

[0084] 일부 실시예들에서, 폴리뉴클레오타이드 (또는 유전자) 서열은 폴리뉴클레오타이드 서열에 작동가능하게 연결된 프로모터를 포함하는, 재조합 벡터를 거쳐 숙주 세포에 제공된다. 여하한 실시예들에서, 프로모터는 발달-조절된(developmentally-regulated), 세포소기관(organelle)-특이적, 조직-특이적, 유도적, 구조적 또는 세포-특이적 프로모터이다. 일부 실시예들에서, 재조합 벡터는 (a) 폴리뉴클레오타이드 서열에 작동가능하게(operatively) 연결된 발현 조절 서열; (b) 폴리뉴클레오타이드 서열에 작동가능하게 연결된 선택 마커; (c) 폴리뉴클레오타이드 서열에 작동가능하게 연결된 마커 서열; (d) 폴리뉴클레오타이드 서열에 작동가능하게 연결된 정제 모이어티(purification moiety); (e) 폴리뉴클레오타이드 서열에 작동가능하게 연결된 분비 서열; 및 (f) 폴리뉴클레오타이드 서열에 작동가능하게 연결된 표적화 서열(targeting sequence)을 포함한다. 본 명세서에 설명된 발현 벡터들은 숙주 세포 내에서 폴리뉴클레오타이드 서열의 발현에 적합한 형태의 본 명세서에 설명된 폴리뉴클레오타이드 서열을 포함한다. 발현 벡터의 설계는 형질전환되는 숙주 세포의 선택, 원하는 폴리펩티드의 발현 수준 등과 같은 이러한 요인들에 의존할 수 있다는 것을 본 기술분야의 당업자들은 인지할 것이다. 본 명세서에 설명된 발현 벡터들은 본 명세서에 설명되는 폴리뉴클레오타이드 서열들에 의해 인코딩되는, 융합 폴리펩티드들을 포함하는 폴리펩티드들을 생산하기 위하여 숙주 세포들에 도입될 수 있다. 원핵생물, 예를 들어 *대장균(E. coli)*에서 폴리펩티드들을 인코딩하는 유전자의 발현은 대부분 융합 및 비-융합 폴리펩티드들 중 하나의 발현을 지시하는 구성적 또는 유도성 프로모터를 함유하는 벡터들로 수행된다. 융합 벡터들은 그 속에 인코딩된 폴리펩티드에, 일반적으로 재조합 폴리펩티드의 아미노 또는 카르복시 말단에 다수의 아미노산들을 첨가한다. 전형적으로, 이러한 융합 벡터들은 다음의 3가지 목적: (1) 재조합 폴리펩티드의 발현을 증가시키고; (2) 재조합 폴리펩티드의 용해도를 증가시키며; 그리고 (3) 친화성 정제(affinity purification)에서 리간드로서 작용함으로써 재조합 폴

리펩티드의 정제를 돕는 것을 수행한다. 종종, 융합 발현 벡터에서, 융합 모이어티와 재조합 폴리펩티드의 접합점에 단백분해 절단(proteolytic cleavage) 부위가 도입된다. 이는 융합 폴리펩티드의 정제 이후에 융합 모이어티로부터 재조합 폴리펩티드의 분리를 가능하게 한다. 이러한 효소들 및 이의 동족 인식 서열(cognate recognition sequence)들의 예시들은 Xa 인자(Factor Xa), 트롬빈 및 엔테로키나아제(enterokinase)를 포함한다. 예시적인 융합 발현 벡터는 pGEX[Pharmacia Biotech, Inc., Piscataway, NJ; Smith et al., *Gene*, 67:31-40(1988)], pMAL(New England Biolabs, Beverly, MA) 및 pRITS(Pharmacia Biotech, Inc., Piscataway, N.J.)를 포함하고, 이들은 표적 재조합 폴리펩티드에 각각, 글루타티온 S-전달효소(glutathione S-transferase: GST), 말토오스 E 결합 단백질 또는 단백질 A를 융합시킨다.

[0085] 유도성, 비-융합 대장균 발현 벡터의 예시들은 pTrc[Amann et al., *Gene*(1988) 69:301-315] 및 pET 11d[Studier et al., *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 중합효소 전사에 의존한다. pET 11d 벡터로부터의 표적 유전자 발현은 동시에 발현되는 바이러스 RNA 중합효소(T7 gn1)에 의해 매개되는 T7 gn10-lac 융합 프로모터로부터의 전사에 의존한다. 이러한 바이러스 중합효소는 lacUV 5 프로모터의 전사 제어 하에서 T7 gn1 유전자를 보유하는 내재성 λ 프로파지(resident λ prophage)로부터 숙주 균주들 BL21(DE3) 또는 HMS174(DE3)에 의해 공급된다. 원핵 세포 및 진핵 세포 모두에 적합한 발현 시스템들은 본 기술분야에 잘 알려져 있고; 예를 들어, Sambrook et al., "Molecular Cloning: A Laboratory Manual", second edition, Cold Spring Harbor Laboratory(1989)를 참조한다. 유도성, 비-융합 대장균 발현 벡터의 예시들은 pTrc[Amann et al., *Gene* 69:301-315(1988)] 및 pET 11d[Studier et al., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA, pp.60-89(1990)]를 포함한다. 여하한 실시예들에서, 본 발명의 폴리뉴클레오티드 서열은 박테리오파지 T5로부터 유래된 프로모터에 작동가능하게 연결된다. 일 실시예에서, 숙주 세포는 효모 세포이다. 일 실시예에서, 발현 벡터는 효모 발현 벡터이다. 벡터들은 외래(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. 예를 들어, 숙주 세포를 형질전환 또는 형질감염(transfecting)시키는데 적합한 방법들은 Sambrook et al.(*supra*)에서 찾아볼 수 있다. 박테리아 세포들의 안정한 형질전환을 위하여, 사용되는 발현 벡터 및 형질전환 기술에 따라, 소량(small fraction)의 세포들만 발현 벡터를 흡수하고 복제하는 것으로 알려져 있다. 일부 실시예들에서, 이러한 형질전환체들을 확인하고 선택하기 위하여, 선택가능한 마커(예를 들어, 항생제에 대한 내성)를 인코딩하는 유전자는 목적 유전자와 함께 숙주 세포 내로 도입된다. 선택가능한 마커들은 암피실린, 카나마이신, 클로람페니콜 또는 테트라사이클린과 같은 약물들에 대한 내성을 제공하는 것들이 포함하나, 이로 제한되지 않는다. 선택가능한 마커를 인코딩하는 핵산들은 본 명세서에서 설명되는 폴리펩티드를 인코딩하는 벡터와 동일한 벡터로 숙주 세포 내로 도입되거나, 별개의 벡터로 도입될 수 있다. 도입된 핵산에 의해 안정적으로 형질감염된 세포들은 적절하게 선택된 약물의 존재 하에서 성장에 의해 확인될 수 있다.

[0086] **재조합 숙주 세포에 의한 지방족 알코올 조성물의 생산**

[0087] 재조합 숙주 세포에 의해 지방족 알코올의 생산을 증가시키기 위한 전략들은, 생산 숙주 내의 상이한 유기체들로부터 외인성 지방산 생합성 유전자들의 발현 및 본래의 지방산 생합성 유전자들의 과발현에 의해 지방산 생합성 경로를 통한 증가된 플럭스를 포함한다. 또한, 숙주 세포의 성장 및 생산물을 최적화시키기 위한 다른 전략들뿐만 아니라 지방족 알코올 생합성 경로에 있어서 관련된 효소들, 예를 들어 CAR의 증강된(Enhanced) 활성은 생산을 최대화하는데 이용될 수 있다. 일부 실시예들에서, 재조합 숙주 세포는 지방족 알코올 생합성 활성을 갖는 폴리펩티드(효소)[즉, 지방족 알코올 생합성 폴리펩티드 또는 지방족 알코올 생합성 효소]를 인코딩하는 폴리뉴클레오티드, 및 재조합 숙주 세포에 의해서 생산되는 지방족 알코올을 포함한다. 지방족 알코올을 포함하는 조성물(지방족 알코올 조성물)은 지방족 알코올 생합성 효소를 발현시키는데 효과적인 조건들 하에서 탄소 공급원의 존재 하에서 재조합 숙주 세포를 배양시킴으로써 생산될 수 있다. 일부 실시예들에서, 지방족 알코올 조성물은 지방족 알코올을 포함하나, 지방족 알코올 조성물은 다른 지방산 유도체들을 포함할 수 있다. 전형적으로, 지방족 알코올 조성물은 재조합 숙주 세포의 세포 외 환경, 즉 세포 배양 배지에서 회수된다. 한가지 접근법에서, 재조합 숙주 세포는 아실-ACP의 유리 지방산(FFAs)으로의 전환을 촉진시키는 티오에스테라아제, 및 유리 지방산을 지방족 알데히드로 전환시키는 카르복실산 환원효소(CAR)를 발현시킴으로써 지방족 알코올을 생산하도록 조작된다. 숙주 세포(예를 들어, 대장균) 내에 존재하는 본래의 (내인성) 알데히드 환원효소는 지방족 알데히드를 지방족 알코올로 전환시킬 수 있다. 일부 실시예들에서, 지방족 알코올은 지방족 알데히드를 지방족 알코올로 전환시키는 지방족 알코올 생합성 활성을 갖는 폴리펩티드를 인코딩하는 폴리뉴클레오티드를 재조합 숙주 세포 내에서 발현 또는 과발현시킴으로써 생산된다. 예를 들어, 알코올 탈수소효소(본 명세서에서

알데히드 환원효소라고도 칭함, 예를 들어, EC 1.1.1.1)는 본 발명을 실시하는데 사용될 수 있다. 본 명세서에서 사용되는 바와 같이, "알코올 탈수소효소"라는 용어는 지방족 알데히드의 알코올(예를 들어, 지방족 알코올)로의 전환을 촉진시킬 수 있는 폴리펩티드를 칭한다. 본 기술분야의 당업자는 여하한 알코올 탈수소효소들이 또한 다른 반응들을 촉진시킬 수 있고, 이러한 비-특이적 알코올 탈수소효소들도 "알코올 탈수소효소"라는 용어에 포괄된다는 것을 인식할 것이다. 본 발명에 따라서 유용한 알코올 탈수소효소 폴리펩티드의 예시들은 AlrAadp1(SEQ ID NO: 4) 또는 AlrA 동족체들; 그리고 YjgB(AAC77226)(SEQ ID NO: 5), DkgA(NP_417485), DkgB(NP_417473), 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 탈수소효소들을 포함하나, 이로 제한되지 않는다. 추가적인 예시들은 국제 특허 출원 공개 Nos. WO 2007/136762, WO 2008/119082 및 WO 2010/062480에 기재되고, 이들의 각각은 특히 본 명세서에서 인용참조된다. 여하한 실시예들에서, 지방족 알코올 생합성 폴리펩티드는 알데히드 환원효소 또는 알코올 탈수소효소 활성(EC 1.1.1.1)을 갖는다. 또 다른 접근법에서, 지방족 아실-티오에스테르 기질(예를 들어, 지방족 아실-CoA 또는 지방족 아실-ACP)을 지방족 알코올로 전환시키는 지방족 알코올 형성 아실-CoA 환원효소 또는 지방족 아실 환원효소(FARs)를 발현시킴으로써 지방족 알코올을 생산하도록 조작된다. 일부 실시예들에서, 지방족 알코올은 재조합 숙주 세포 내에서 지방족 알코올 형성 아실-CoA 환원효소(FAR) 활성을 갖는 폴리펩티드를 인코딩하는 폴리뉴클레오티드를 발현 또는 과발현시킴으로써 생산된다. 본 실시예에 따라서 유용한 FAR 폴리펩티드의 예시들은 PCT 공개 WO 2010/062480에 기재되고, 이는 특히 본 명세서에서 인용참조된다.

[0088]

지방족 알코올은 지방족 아실-ACP 및 지방족 아실-CoA 중간체들을 이용하는 아실-CoA 의존적 경로, 및 지방족 아실-CoA 중간체가 아닌 지방족 아실-ACP 중간체들을 이용하는 아실-CoA 독립적 경로를 거쳐 생산될 수 있다. 특정 실시예들에서, 과발현된 유전자에 의해 인코딩되는 효소는 지방산 생성효소, 아실-ACP 티오에스테라아제, 지방족 아실-CoA 생성효소 및 아세틸-CoA 카르복실라아제로부터 선택된다. 일부 실시예들에서, 과발현된 유전자에 의해 인코딩되는 단백질은 숙주 세포에 내인성이다. 다른 실시예들에서, 과발현된 유전자에 의해 인코딩되는 단백질은 숙주 세포와 이종이다. 또한, 지방족 알코올은 다양한 아실-ACP 또는 아실-CoA 분자들을 대응하는 일차 알코올들로 환원시킬 수 있는 효소들에 의해 자연적으로 만들어진다. 또한, 미국 특허 공개 Nos. 20100105963 및 20110206630, 그리고 미국 특허 No. 8097439를 참조하며, 이들은 특히 본 명세서에서 인용참조된다. 본 명세서에서 사용되는 바와 같이, 재조합 숙주 세포 또는 조작된 숙주 세포는, 유전적 구성(genetic makeup)이 대응하는 야생형 숙주 세포에 대하여, 예를 들어 새로운 유전적 요소(genetic element)들의 의도적인 도입(deliberate introduction) 및/또는 숙주 세포 내에 본래 존재하는 유전적 요소들의 의도적인 변형에 의해 변화된 숙주 세포를 칭한다. 또한, 이러한 재조합 숙주 세포의 오프스프링(offspring)은 이러한 새로운 및/또는 변형된 유전적 요소들을 함유한다. 본 명세서에 설명된 본 발명의 여하한 측면들에서, 숙주 세포는 식물 세포, 곤충 세포, 균류 세포[예를 들어, 칸다디아 균(*Candida* sp.)과 같은 곰팡이, 사카로미세스 균(*Saccharomyces* sp.)과 같은 출아 효모(budding yeast)], 조류 세포(algal cell) 및 박테리아성 세포로 이루어진 그룹에서 선택될 수 있다. 바람직한 일 실시예에서, 재조합 숙주 세포는 재조합 미생물 세포이다. 미생물 세포인 숙주 세포의 예시들은 *에스체리키아*(*Escherichia*) 속, *바실러스*(*Bacillus*) 속, *락토바실러스*(*Lactobacillus*) 속, *자이모모나스*(*Zymomonas*) 속, *로도코코스*(*Rhodococcus*) 속, *슈도모나스*(*Pseudomonas*) 속, *아스페르길러스*(*Aspergillus*) 속, *트리코더마*(*Trichoderma*) 속, *뉴로스포라*(*Neurospora*) 속, *푸사리움*(*Fusarium*) 속, *후미콜라*(*Humicola*) 속, *리조무코*(*Rhizomucor*) 속, *클루이베로마이세스*(*Kluyveromyces*) 속, *피치아*(*Pichia*) 속, *무코*(*Mucor*) 속, *미셀리오프토라*(*Myceliophthora*) 속, *페니실리움*(*Penicillium*) 속, *파네로차에테*(*Phanerochaete*) 속, *느타리*(*Pleurotus*) 속, *트라메테스*(*Trametes*) 속, *크리소스포리움*(*Chrysosporium*) 속, *사카로마이세스*(*Saccharomyces*) 속, *스테노트로파모나스*(*Stenotrophomonas*) 속, *스키조사카로마이세스*(*Schizosaccharomyces*) 속, *야로위아*(*Yarrowia*) 속 또는 *스트렙토마이세스*(*Streptomyces*) 속으로부터의 세포들을 포함하나, 이로 제한되지 않는다. 일부 실시예들에서, 숙주 세포는 그람-양성(Gram-positive) 박테리아성 세포이다. 여타 실시예들에서, 숙주 세포는 그람-음성(Gram-negative) 박테리아성 세포이다. 일부 실시예들에서, 숙주 세포는 대장균 세포이다. 다른 실시예들에서, 숙주 세포는 *바실러스 렌투스*(*Bacillus lentus*) 세포, *바실러스 브레비스*(*Bacillus brevis*) 세포, *바실러스 스테아로써모필러스*(*Bacillus stearothermophilus*) 세포, *바실러스 리케니포르미스*(*Bacillus licheniformis*) 세포, *바실러스 알칼로필루스*(*Bacillus alkalophilus*) 세포, *바실러스 코아굴란스*(*Bacillus coagulans*) 세포, *바실러스 키르쿨란스*(*Bacillus circulans*) 세포, *바실러스 푸밀리스*(*Bacillus pumilis*) 세포, *바실러스 튜링기엔시스*(*Bacillus thuringiensis*) 세포, *바실러스 클라우지*(*Bacillus clausii*) 세포, *바실러스 메가테리움*(*Bacillus megaterium*) 세포, *바실러스 서브틸리스*(*Bacillus subtilis*) 세포 또는 *바실러스 아밀로리퀴파시엔스*(*Bacillus amyloliquefaciens*) 세포이다. 다른

실시예들에서, 숙주 세포는 *트리코더마 코닝기*(*Trichoderma koningii*) 세포, *트리코더마 비리데*(*Trichoderma viride*) 세포, *트리코더마 르에세이*(*Trichoderma reesei*) 세포, *트리코더마 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, *로도코코스 오파쿠스*(*Rhodococcus opacus*) 세포, *리조무코 미에헤이*(*Rhizomucor miehei*) 세포 또는 *무코 미에헤이*(*Mucor michei*) 세포이다.

[0089] 또 다른 실시예들에서, 숙주 세포는 *스트렙토마이세스 리비단스*(*Streptomyces lividans*) 세포 또는 *스트렙토마이세스 무리누스*(*Streptomyces murinus*) 세포이다. 또 다른 실시예들에서, 숙주 세포는 방선균(*Actinomycetes*) 세포이다. 일부 실시예들에서, 숙주 세포는 *맥주효모균*(*Saccharomyces cerevisiae*) 세포이다. 일부 실시예들에서, 숙주 세포는 *맥주효모균*(*Saccharomyces cerevisiae*) 세포이다. 다른 실시예들에서, 숙주 세포는 진핵 식물(eukaryotic plant), 조류, 남세균(cyanobacterium), 녹색-황 세균, 녹색 비-황(green non-sulfur) 세균, 자색 황세균, 자색 비-황 세균, 극한성 생물(extremophile), 효모, 균류, 이들의 조합된 유기체 또는 합성 유기체로부터의 세포이다. 일부 실시예들에서, 숙주 세포는 광 의존성이거나 탄소를 고정시킨다. 일부 실시예들에서, 숙주 세포는 광 의존성이거나 탄소를 고정시킨다. 일부 실시예들에서, 숙주 세포는 독립영양적 활성을 갖는다. 일부 실시예들에서, 숙주 세포는 빛의 존재 하에서와 같이, 광독립영양적 활성(photoautotrophic activity)을 갖는다. 일부 실시예들에서, 숙주 세포는 빛의 부재 하에서 종속영양적 또는 혼합영양적이다. 여하한 실시예들에서, 숙주 세포는 *애기장대*(*Arabidopsis thaliana*), *파니쿰 비르가툼*(*Panicum virgatum*), *미스칸투스 기간테우스*(*Miscanthus giganteus*), *제아 메이즈*(*Zea mays*), *보트리오크코브라우니*(*Botryococcuse braunii*), *클라미도모나스 레인하르트티*(*Chlamydomonas reinhardtii*), *두나리엘라 살리나*(*Dunaliella salina*), *시네코코커스 쉼*(*Synechococcus Sp.*) PCC 7002, *시네코코커스 쉼* PCC 7942, *시네코키스티스 쉼*(*Synechocystis Sp.*) PCC 6803, *써모시네코코커스 엘롱게이트*(*Thermosynechococcus elongates*) BP-1, *클로로비움 테피둠*(*Chlorobium tepidum*), *클로로프렉서스 아우란티쿠스*(*Chloroflexus auranticus*), *크로마티움 비노썩*(*Chromatium vinosum*), *로도스피리움 루브룸*(*Rhodospirillum rubrum*), *로도박터 캡슐라투스*(*Rhodobacter capsulatus*), *로도슈도모나스 파루스리스*(*Rhodopseudomonas palustris*), *클로스트리디움 융달리*(*Clostridium ljungdahlii*), *클로스트리디움썩셀룸*(*Clostridiuthermocellum*), *페니실리움 크리소게눔*(*Penicillium chrysogenum*), *피치아 파스토리스*(*Pichia pastoris*), *맥주효모균*(*Saccharomyces cerevisiae*), *스키조사카로마이세스 포롬브*(*Schizosaccharomyces pombe*), *슈도모나스 플루오르센스*(*Pseudomonas fluorescens*) 또는 *자이모모나스 모빌리스*(*Zymomonas mobilis*)로부터의 세포이다.

[0090] 조작된 숙주 세포의 배양 및 발효

[0091] 본 명세서에서 사용되는 바와 같이, 발효는 광범위하게 숙주 세포에 의한 유기 물질의 표적 물질(target substance)로의 전환, 예를 들어 탄소 공급원을 포함하는 배지 내에서 재조합 숙주 세포의 배양물(culture)을 증식시킴으로써 재조합 숙주 세포에 의한 탄소 공급원의 지방산 또는 이의 유도체들로의 전환을 칭한다. 본 명세서에서 사용되는 바와 같이, 생산을 위하여 허용되는 조건들은 숙주 세포가 원하는 생성물, 이를 테면 지방산 또는 지방산 유도체를 생산하도록 하는 여하한 조건들을 의미한다. 유사하게, 벡터의 폴리뉴클레오티드 서열이 발현되는 조건들은 숙주 세포가 폴리펩티드를 합성하도록 하는 여하한 조건들을 의미한다. 예를 들어, 적합한 조건들은 발효 조건들을 포함한다. 발효 조건들은 온도 범위들, 통기 수준들 및 배지 조성과 같은 많은 파라미터들을 포함할 수 있다. 이러한 조건들의 각각은 개별적으로 그리고 조합적으로, 숙주 세포가 성장할 수 있도록 한다. 발효는 호기성, 혐기성 또는 [미-호기성(micro-aerobic)과 같은] 이의 변화(variation)일 수 있다. 예시적인 배양 배지는 액체 배지(broths) 또는 겔(gels)을 포함한다. 일반적으로, 배지는 숙주 세포에 의해 직접적으로 대사작용될 수 있는 탄소 공급원을 포함한다. 또한, 효소들은 탄소 공급원의 기동(mobilization)[예를 들어, 전분 또는 셀룰로오스의 발효가능 당들로의 해중합반응(depolymerization)] 및 차후 신진대사를 촉진시키기 위하여 배지에서 사용될 수 있다. 소규모 생산의 경우에, 조작된 숙주 세포들은 예를 들어, 약 100 mL, 500 mL, 1 L, 2 L, 5 L 또는 10 L의 배치(batchs)에서 성장될 수 있고; 발효될 수 있으며; 원하는 폴리뉴클레오티드 서열, 이를 테면 CAR 폴리펩티드를 인코딩하는 폴리뉴클레오티드 서열을 발현시키도록 유도될 수 있다. 대규모 생산의 경우에, 조작된 숙주 세포들은 약 10 L, 100 L, 1000 L, 10,000 L, 100,000 L 및 1,000,000 L 또는 그 이상의 배치(batchs)에서 성장될 수 있고; 발효될 수 있으며; 원하는 폴리뉴클레오티드 서열을 발현시키도록 유도될 수 있다. 대안적으로, 대규모의 유가 배양식 발효가 실시될 수 있다.

[0092] 지방족 알코올 조성물

[0093] 본 명세서에서 설명되는 지방산 유도체 조성물들은 재조합 숙주 세포 배양의 세포 외 환경에서 발견되고, 배양 배지로부터 손쉽게 분리될 수 있다. 지방산 유도체는 재조합 숙주 세포에 의해 분비될 수 있고, 세포 외 환경으로 수송될 수 있거나 또는 수동적으로 재조합 숙주 세포 배양의 세포 외 환경으로 전달될 수 있다. 지방산 유도체는 본 기술분야에 알려진 정례적인(routine) 방법들을 사용하여 재조합 숙주 세포 배양물(culture)로부터 분리된다. 본 발명은 하나 이상의 지방족 알데히드 및/또는 지방족 알코올을 포함하는 조작된 또는 재조합 숙주 세포들에 의해 생산되는 조성물들[바이오생산물들(Bioproductions)]을 제공한다. 특정 사슬 길이 및 포화도를 갖는 지방족 알코올 요소가 배양되는 조작된 또는 재조합 숙주 세포에 의해 생산되는 다수의 바이오생산물을 구성할지라도, 조성물은 전형적으로 사슬 길이 및/또는 포화도에 따라 달라지는 지방족 알데히드 및/또는 지방족 알코올의 혼합물을 포함한다. 본 명세서에서 사용되는 바와 같이, 현대 탄소(modern carbon)의 분율 또는 fM은 각각, 옥살산 기준 HOxI 및 HOxII로 알려져 있는, 미국 국립표준기술연구소(National Institute of Standards and Technology: NIST) 표준 참고 물질들(Standard Reference Materials: SRMs) 4990B 및 4990C에 의해 정의된 바와 동일한 의미를 갖는다. 이러한 기초적인 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 동등하다. 현재 생존 생물권(living biosphere)[식물 물질]의 경우에, f_M 은 대략 1.1이다.

[0094] 본 명세서의 지방산 생합성 경로를 사용하여 생물학적으로 생산되는 유기 화합물들, 특히 생물학적으로 생산되는 지방족 알데히드들 및 알코올들을 포함하는 바이오생산물들(Bioproductions)[예를 들어, 본원 발명에 따라서 생산되는 지방족 알데히드들 및 알코올들]은, 재생가능한 공급원들로부터 생산되지 않으며, 이로써, 새로운 물질의 조성물이다. 이러한 새로운 바이오생산물들은 이중 탄소-동위원소 지문기술(dual carbon-isotopic fingerprinting) 또는 ^{14}C 연대측정(dating)을 기초로 석유화학의 탄소로부터 유래되는 유기 화합물들과 구별될 수 있다. 부가적으로, 생물자원 탄소(biosourced carbon)의 특이적인 공급원(예를 들어, 글루코오스 대 글리세롤)은 이중 탄소-동위원소 지문기술에 의해 결정될 수 있다(본 명세서에 인용참조되는 미국 특허 No. 7,169,588 참조). 유기 화합물들을 기반으로 하는 석유로부터 바이오생산물을 구별하는 능력은 거래 중에(in commerce) 이러한 물질들을 추적하는데 유익하다. 예를 들어, 생물학적 기반의 탄소 동위원소 프로파일들 및 석유 기반의 탄소 동위원소 프로파일들 모두를 포함하는 유기 화합물들 또는 화학제품들은 석유 기반 물질들로부터 만들어진 유기 화합물들 및 화학제품들과 구별될 수 있다. 따라서, 본 명세서의 바이오생산물들은 이들의 독특한 탄소 동위원소 프로파일들 기반으로 하여 거래 중에 후속되거나 추적될 수 있다. 바이오생산물들은 각각의 연료에서 안정한 탄소 동위원소 비율($^{13}\text{C}/^{12}\text{C}$)을 비교함으로써 유기 화합물들을 기반으로 하는 석유와 구별될 수 있다. 주어진 바이오생산물의 $^{13}\text{C}/^{12}\text{C}$ 비율은 이산화탄소가 고정된 시간에서 대기 중의 이산화탄소에 대한 $^{13}\text{C}/^{12}\text{C}$ 비율의 결과이다. 또한, 이는 정확한 대사작용 경로를 반영한다. 국부적인 변이들도 발생한다. 석유, C_3 식물들[활엽(the broadleaf)], C_4 식물들[목초(the grasses)] 및 해양의 탄산염(marine carbonate)들 모두는 $^{13}\text{C}/^{12}\text{C}$ 및 대응하는 $\delta^{13}\text{C}$ 수치들에서 상당한 차이들을 나타낸다. 더욱이, C_3 및 C_4 식물들의 지질 물질은 대사작용 경로의 결과로서 동일한 식물들의 탄수화물 성분들로부터 유래된 재료들에 비하여 상이하게 분석한다. 측정의 정확도 내에서, ^{13}C 은 동위원소 분별 효과들로 인하여 큰 변화들을 나타내며, 바이오생산물들에 대한 이의 가장 중요한 점은 광합성 메카니즘이다. 식물에서 탄소 동위원소 비율에 있어서 차이의 주요한 원인은 식물에서 광합성 탄소 대사작용의 경로, 특히 일차 카르복실화(primary carboxylation)[즉, 대기 중 CO_2 의 초기 고정] 동안 발생하는 반응에서의 차이들과 밀접하게 관련된다. 식생(vegetation)의 두 가지 종류들은 C_3 [또는 캘빈-벤슨(Calvin-Benson)] 생합성 회로를 병합하는 것들 및 C_4 [또는 해치-슬랙(Hatch-Slack)] 생합성 회로를 병합하는 것들이다. C_3 식물들에서, 일차 CO_2 고정 또는 카르복실화 반응은 효소인 리불로오스-1,5-디포스페이트 카르복실라아제를 포함하고, 첫번째의 안정한 생성물은 3-탄소 화합물이다. 경목(hardwood)들 및 침엽수(conifer)들과 같은 C_3 식물들은 온대 기후 지역들에서 우세하다. C_4 식물들에서, 또 다른 효소인 포스포엔올-피루베이트 카르복실라아제를 포함하는 부가적인 카르복실화 반응은 일차 카르복실화 반응이다. 첫번째의 안정한 탄소 화합물은 이후에 탈카르복실화되는 4-탄소산(carbon acid)이다. 따라서, 방출된 CO_2 는 C_3 회로에 의해 재고정된다. C_4 식물의 예시들은 열대형 목초(tropical grass)들, 옥수수 및 사탕수수이다. C_4 식물들 및 C_3

식물들 모두는 $^{13}\text{C}/^{12}\text{C}$ 동위원소 비율의 범위를 나타내나, 전형적인 수치들은 C_4 식물들에 대하여 mil 당 약 -7 내지 약 -13이며, C_3 식물들에 대하여 mil 당 약 -19 내지 약 -27이다[예를 들어, Stuiver *et al.*, *Radiocarbon* 19:355(1977) 참조]. 석탄 및 석유는 일반적으로 이의 후자 범위에서 떨어져 있다. ^{13}C 측정 척도(measurement scale)는 본래 Pee Dee Belemnite(PDB) 석회암에 의해 제로 세트(zero set)로 정의되며, 여기서 수치들은 이러한 재료로부터 천 분의 일 편차로 주어진다. " $\delta^{13}\text{C}$ " 수치는 천 분의 일[퍼밀(per mil)], 약어로는 ‰로 표현되고, 하기와 같이 계산된다:

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

PDB 표준 물질(RM)이 고갈되고 있으므로, 일련의 대안적인 RM들이 IAEA, USGS, NIST 및 다른 선택된 국제 동위원소 실험실들과 협력하여 개발되고 있다. PDB로부터 퍼밀(per mil) 편차들에 대한 표기는 $\delta^{13}\text{C}$ 이다. 측정은 질량 44, 45 및 46의 분자 이온들에 대한 매우 정확한 신뢰성 있는 비율의 질량 분석기[high precision stable ratio mass spectrometry](IRMS)에 의해 CO_2 상에서 이뤄진다. 본 명세서에서 설명되는 조성물들은 본 명세서에서 설명되는 여하한 방법들에 의해 생산되는 바이오생성물들, 예를 들어 지방족 알데히드 및 알코올 생성물들을 포함한다. 특히, 바이오생성물은 약 -28 이상, 약 -27 이상, -20 이상, -18 이상, -15 이상, -13 이상, -10 이상, 또는 -8 이상의 $\delta^{13}\text{C}$ 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 예시들에서, 바이오생성물은 약 -10, -11, -12 또는 -12.3의 $\delta^{13}\text{C}$ 를 가질 수 있다. 또한, 본 명세서의 상세한 설명에 따라서 생성되는 바이오생성물들을 포함하는 바이오생성물들은 각각의 화합물에서 ^{14}C 의 양을 비교함으로써 유기 화합물들을 기반으로 하는 석유와 구별될 수 있다. ^{14}C 는 핵의 반감기가 5730년이므로, "더 오래된(older)" 탄소를 함유하는 석유 기반 연료는 "최근의(newer)" 탄소를 함유하는 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].

방사성 탄소 연대측정법(radiocarbon dating)에서 기본적인 가정은 대기 중에 ^{14}C 농도의 항상성(constancy)이 생물(living organism)들 내에 ^{14}C 의 항상성에 이르게 한다는 것이다. 그러나, 1950년 이후부터의 대기권의 핵 실험 및 1850년 이후부터의 화석 연료의 연소로 인하여, ^{14}C 는 제 2의, 지구화학적인 시간 특성을 얻었다. 대기 중 CO_2 , 및 그에 따른 생물권(living biosphere)에 있어서 이의 농도는 1960년대 중반의 핵실험의 정점(peak)에 대략 두 배였다. 이후에, 대략 7 내지 10년간의 "반-감기" 이완(relaxation "half-life")을 가지는, 약 1.2×10^{-12} 의 정상-상태의 우주기원(대기) 기준 동위원소 비율($^{14}\text{C}/^{12}\text{C}$)로 서서히 되돌아 온다[이런 후자의 반감기는 문자 그대로 취해지는 것이 아니라; 오히려 핵무기 시대(nuclear age)의 시작 이후로 대기 및 생물권 중에 ^{14}C 의 변화를 추적하기 위하여 자세한 대기권 핵 투입/붕괴의 기능(atmospheric nuclear input/decay function)을 사용해야 한다]. 이는 최근의 생물권 탄소에 대한 매년의 연대측정(annual dating)의 전망을 제공하는 후자의 생물권 ^{14}C 시간 특성이다. ^{14}C 는 "현대 탄소의 분율"(f_M)의 단위로 주어지는 결과들을 갖는, 가속기 질량 분광 분석(accelerator mass spectrometry: AMS)에 의해 측정될 수 있다. f_M 은 미국표준기술연구소(NIST)의 표준 기준 물질들(SRMs) 4990B 및 4990C에 의해 정의된다. 본 명세서에서 사용되는 바와 같이, 현대 탄소의 분율(f_M)은 각각 옥살산 기준 H0xI 및 H0xII로 알려진, 미국표준기술연구소(NIST)의 표준 기준 물질들(SRMs) 4990B 및 4990C에 의해 정의되는 바와 같은 동일한 의미를 갖는다. 기초적인 정의는 (AD 1950을 기준으로) 0.95 배의 $^{14}\text{C}/^{12}\text{C}$ 동위원소 비율 H0xI에 관한 것이다. 이는 감소-수정된(decay-corrected) 산업 혁명 이전의 목재와 거의 동등하다. 현재의 생물권(식물 재료)에 대한, f_M 은 대략 1.1이다. 이는 감소-수정된(decay-corrected) 산업 혁명 이전의 목재와 거의 동등하다. 현재의 생물권(식물 재료)에 대한, f_M 은 대략 1.1이다.

본 명세서에서 설명되는 조성물들은 적어도 약 1 이상의 $f_M^{14}\text{C}$ 를 가질 수 있는 바이오생성물들을 포함한다. 예

를 들어, 바이오생성물은 적어도 약 1.01의 $f_M^{14}C$, 약 1 내지 약 1.5의 $f_M^{14}C$, 약 1.04 내지 약 1.18의 $f_M^{14}C$, 또는 약 1.111 내지 약 1.124의 $f_M^{14}C$ 을 가질 수 있다. ^{14}C 의 또 다른 측정은 현대 탄소의 퍼센트(pMC)로 알려져 있다. ^{14}C 연대를 사용하는 고고학자 또는 지질학자에 경우, AD 1950년은 "측정의 기점 연도(zero years old)"와 같다. 또한, 이는 100 pMC로 나타난다. 대기 중의 "핵무기 탄소(Bomb carbon)"는 열-핵무기들의 정점인 1963년에 정상 수준에 거의 두 배에 달하였다. 대기권 내에서 이의 분포는, 핵무기 탄소의 등장 이후에, AD 1950년 이후로 살아 있는 식물들 및 동물들에 대하여 100 pMC보다 더 큰 수치들을 나타내는 것으로 추정되고 있다. 107.5 pMC에 가까운 오늘날의 수치는 시간이 지나면서 점차 감소하고 있다. 이는 옥수수과 같은 신선한 바이오매스 재료는 107.5 pMC에 가까운 ^{14}C 특징(signature)을 부여하는 것을 의미한다. 석유 기반 화합물들은 0의 pMC 수치를 가질 것이다. 현대 탄소와 화석 탄소의 결합은 현대 pMC 함유량(content)의 희석을 야기할 것이다. 107.5 pMC가 현대 바이오매스 재료들의 ^{14}C 함유량을 나타내며 0 pMC가 석유 기반 생물물들의 ^{14}C 함유량을 나타낸다고 가정함으로써, 재료에 대한 측정된 pMC 수치는 두 개의 성분 타입에 대한 비율을 반영할 것이다. 예를 들어, 오늘날의 콩으로부터 100 % 유래된 재료는 107.5 pMC에 가까운 방사성 탄소 특징을 제공할 것이다. 만약 재료가 석유 기반 생물물로 50 % 희석되었다면, 이는 대략 54 pMC의 방사성 탄소 특징을 제공할 것이다. 생물학적 기반의 탄소 함유량은 107.5 pMC와 같으면 "100 %"이고, 0 pMC와 같으면 "0 %"라고 지정함으로써 유래된다. 예를 들어, 99 pMC로 측정된 샘플은 93 %의 동등한 생물학적 기반의 탄소 함유량을 제공할 것이다. 이 수치는 생물학적 기반의 탄소 결과를 의미하며, 현대의 생물학적 재료 또는 석유 기반 재료로부터 유래되는 분석된 재료 내에서 모든 성분들을 추측하는 것으로 해석된다. 본 명세서에서 설명되는 바와 같은 지방족 알데히드 또는 알코올은 적어도 약 50, 60, 70, 75, 80, 85, 90, 95, 96, 97, 98, 99 또는 100의 pMC를 가질 수 있다. 다른 예시들에서, 본 명세서에서 설명되는 바이오생성물은 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 예시들에서, 본 명세서에서 설명되는 바이오생성물은 약 90, 91, 92, 93, 94 또는 94.2의 pMC를 가질 수 있다.

[0099] **제조합 숙주 세포에 의해 생산되는 지방족 알코올 조성물들의 스크리닝**

[0100] 발현을 가능하게 할 만큼 조건들이 충분한 지를 결정하기 위하여, 이중 유전자 또는 변형된 본래의 유전자를 포함하는 제조합 숙주 세포는, 예를 들어 약 4, 8, 12, 24, 36 또는 48시간 동안 배양될 수 있다. 배양 동안 및/또는 배양 이후, 샘플이 얻어지고, 지방족 알코올 생산 수준(역가, 수율 또는 생산물)이 변형되지 않은 대응하는 야생형 모 세포의 지방족 알코올 생산 수준과 다른지를 결정하기 위하여 분석될 수 있다. 예를 들어, 숙주 세포들이 성장된 배지는 원하는 생성물의 존재에 대하여 테스트될 수 있다. 생성물의 존재에 대하여 테스트하는 경우, 이를 테면 TLC, HPLC, GC/FID, GC/MS, LC/MS, MS가 사용될 수 있지만 이들로 제한되지 않는 분석법(assays)이 사용될 수 있다. 제조합 숙주 세포의 균주들은 변화된 지방족 알코올 또는 지방 종류(fatty species)의 생산 수준에 대하여 스크리닝(screen)하기 위하여 소량(small volumes)[0.001 L 내지 1 L]의 플레이트 또는 진탕 플라스크 내에서 배양될 수 있다. 일단 후보(candidate) 균주들 또는 "히트(hit)들"이 소규모(small scale)에서 확인되면, 이러한 균주들은 정확한 지방족 알코올 또는 지방 종류의 생산 수준을 결정하기 위하여 생물반응기, 탱크(tanks) 및 파일럿 플랜트(pilot plants) 내의 대량(larger volumes)[1 L 내지 1000 L]의 배지 내에서 배양된다. 이러한 대량 배양 조건들은 원하는 지방족 알코올 또는 지방 종류 생산을 얻기 위한 배양 조건들을 최적화하기 위하여 본 기술분야의 당업자들에 의해 사용된다.

[0101] **지방족 알데히드 및 지방족 알코올 조성물들의 이용(Utility)**

[0102] 알데히드는 많은 특수 화합물을 생산하는데 사용된다. 예를 들어, 알데히드는 중합체, 수지(예를 들어, 베이클라이트), 염료, 착향료, 가소제, 향수, 약제 및 다른 화학물질들을 생산하는데 사용되고, 이들 중 일부는 용제, 보존제 또는 소독제로 사용될 수 있다. 또한, 비타민 및 호르몬과 같은 어떤 천연 및 합성 화합물들은 알데히드들이고, 많은 당류는 알데히드기들을 함유한다. 지방족 알데히드들은 화학적 또는 효소적 환원에 의해 지방족 알코올들로 전환될 수 있다. 지방족 알코올들은 많은 상업적 용도들을 갖는다. 지방족 알코올들 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(amphiphilic) 성질로 인해, 지방족 알코올들은 비이온성 계면활성제처럼 거동하고, 상기 비이온성 계면활성제는 예를 들어 세제와 같은 개인 위생용품 및 가정 용품에 유용하다. 또한, 지방족 알코올들은 왁스, 검(gums), 수지, 약제학적 연고 및 로션, 윤활유 첨가제, 식물 정전기 방지제 및 가공제, 가소제, 화장품, 공업 용제 및 지방용 용제에 사용된다. 또한, 본 발

명은 본 명세서에서 설명되는 여하한 방법들에 의해 생산되는 지방족 알코올을 포함하는 계면활성제 조성물 또는 세제 조성물을 제공한다. 본 기술분야의 당업자는 계면활성제 또는 세제 조성물의 의도하는 목적에 따라서, 상이한 지방족 알코올들을 생산하고 사용할 수 있다. 예를 들어, 본 명세서에서 설명되는 지방족 알코올들이 계면활성제 또는 세제 생산을 위하여 공급원료로서 사용되는 경우에, 본 기술분야의 당업자는 지방족 알코올 공급원료의 특성들이 생산되는 계면활성제 또는 세제 조성물의 특성들에 영향을 미칠 것이라는 점을 인식할 것이다. 따라서, 계면활성제 또는 세제 조성물의 특성들은 공급원료로서 사용되는 특정 지방족 알코올들을 생산함으로써 선택될 수 있다. 본 명세서에서 설명되는 지방족 알코올 기반의 계면활성제 및/또는 세제 조성물은 본 기술분야에 잘 알려진 다른 계면활성제 및/또는 세제와 혼합될 수 있다. 일부 실시예들에서, 혼합물은 지방족 알코올의 중량 당 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 여하한 두 개의 값들로 제한되는 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 중량 당 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 여하한 두 개의 값들로 제한되는 범위를 포함하도록 만들어질 수 있다. 또한, 이러한 계면활성제 또는 세제 조성물은 마이크로에멀전(microemulsion)과 같은 적어도 하나의 첨가제 또는 식물성유(plant oils) 또는 석유와 같은 비-미생물적인 원료(non-microbial sources)로부터의 계면활성제 또는 세제를 포함할 수 있고, 이는 지방족 알코올의 중량 당 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 여하한 두 개의 값들로 제한되는 범위의 양으로 존재할 수 있다. 본 발명은 추가적으로 하기의 실시예들로 설명된다. 예시들은 설명의 목적으로만 제공된다. 이들은 어떤 식으로든 본 발명의 범위 또는 내용을 제한하는 것으로 구성되지 않는다.

[0104] 실시예들

[0105] 실시예 1

[0106] 생산 숙주 변형 - 아실-CoA 탈수소효소의 감쇠

[0107] 이 실시예는 지방산 분해 효소의 발현이 감쇠되는, 유전적으로 조작된 숙주 세포의 구조(construction)를 설명한다. *대장균* MG1655(*대장균* K 균주)의 *fadE* 유전자는 하기의 변형과 함께, Datsenko et al., Proc. Natl. Acad. Sci. USA 97:6640-6645(2000)에 기재된 Lambda Red(Red-Driven Integration로도 알려짐) 시스템을 사용하여 결실되었다:

[0108] 하기의 두 개의 프라이머들은 *fadE* 결실을 일으키는데 사용되었다:

Del-*fadE*-

F5'-AAAAACAGCAACAATGTGAGCTTTGTTGTAATTATATTGTAAACATATTGATTCC
GGGGATCCGTCGACC (SEQ ID NO: 9); 및

Del-*fadE*-

R5'-AAACGGAGCCTTTTCGGCTCCGTTATTCATTTACGCGGCTTCAACTTTCCTGTA
GGCTGGAGCTGCTTC (SEQ ID NO: 10)

[0109]

[0110] Del-*fadE*-F 및 Del-*fadE*-R 프라이머들은 PCR에 의해 (상기의 Datsenko et al.,에 기재된) 플라스미드 pKD13로부터 카나마이신 저항성(kanamycin resistance: KmR) 카세트를 증폭시키는데 사용되었다. 이후에, PCR 생성물은 사전에 3 내지 4 시간 동안 아라비노오스로 유도되는 (상기의 Datsenko et al.,에 기재된) pKD46을 함유하는 전기 적격(electrocompetent) *대장균* MG1655 세포들을 형질전환시키는데 사용되었다. 37 °C에서 이화대사물 억제(catabolite repression)를 포함하는 가장 최적화된 배양액(super optimal broth with catabolite repression: SOC) 배지 내에서 3 시간 동안 성장 이후, 세포들은 50 µg/mL의 카나마이신을 함유하는 Luria agar 플레이트 상에 평판배양(plated)되었다. 저항성 콜로니(Resistant colony)들이 식별되었고, 37 °C에서 하룻밤 배양된 후 분리되었다. *fadE* 유전자의 파괴(Disruption)는 *대장균* *fadE* 유전자 측면에 위치하도록 제작된, *fadE*-L2 및 *fadE*-R1 프라이머들을 사용하여 PCR 증폭에 의해 확인되었다.

[0111] *fadE* 결실 확인 프라이머들은:

fadE-L2 5'-CGGGCAGGTGCTATGACCAGGAC (SEQ ID NO: 11); 및
fadE-R1 5'-CGCGGCGTTGACCGGCAGCCTGG (SEQ ID NO: 12) 이다.

fadE 결실을 확인한 후, 단일 콜로니가 상기의 Datsenko et al.,에 기재된 바와 같은 pCP20 플라스미드를 사용하여 KmR 마커를 제거하는데 사용되었다. fadE 유전자가 결실되고 KmR 마커가 제거된 결과적으로 얻어진 (resulting) MG1655 대장균 균주는 대장균 MG1655 ΔfadE 또는 대장균 MG 1655 D1이다. fadE 유전자가 결실된 MG1655 대장균 균주에 의한 지방산 유도체("총 지방 종류") 생산은 대장균 MG1655에 의한 지방산 유도체 생산과 비교되었다. 세포들은 생산 플라스미드 pDG109(pCL1920_P_{TRC}_carBopt_12H08_alrAadp1_fabB[A329G]_fadR)으로 형질전환되었고, 글루코오스 최소(glucose minimal) 배지 내에서 발효되었다. 도 5에 나타난 데이터는 fadE 유전자가 지방산 유도체 생산에 영향을 미치지 않았다는 것을 나타낸다.

실시예 2

아세틸 CoA 카르복실라아제 매개되는 - 지방산 합성 경로를 통한 플럭스(Flux) 증가

지방산 생합성의 주요한 전구체들은 말로닐-CoA 및 아세틸-CoA이다(도 1). 이러한 전구체들은 대장균 내에서 지방산 생합성의 속도를 제한하는 것으로 제안된다. 이 실시예에서, 합성 acc 오페론들[코리네박테리움 글루타미쿰(*Corynebacterium glutamicum*) accABCD (±birA)]은 과발현되었고, 대장균 내에서 아세틸-CoA 및 말로닐 CoA 생산 증가로 이어졌다. 한가지 접근법에서, 말로닐-CoA 수치를 증가시키기 위하여, 코리네박테리움 글루타미쿰(*C. 글루타미쿰*)로부터의 아세틸-CoA 카르복실라아제 효소 복합체는 대장균 내에서 과발현되었다. 아세틸-CoA 카르복실라아제(acc)는 네 개의 개별적인 서브유닛(subunit)들인 accA, accB, accC 및 accD로 이루어진다(도 3). *C. 글루타미쿰* acc의 이점은 두 개의 서브유닛들이 각각 accCB 및 accDA인 융합 단백질로 발현된다는 점이고, 이는 이의 균형적인 발현을 용이하게 한다. 추가적으로, accB 서브유닛을 비오틴화시키는 *C. 글루타미쿰* birA(도 3)가 과발현되었다. 실시예 3은 전체 fab 오페론들과 함께 acc 유전자들의 공동 발현을 설명한다.

실시예 3

iFABs - 지방산 합성 경로를 통한 플럭스 증가

지방산 유도체 생산:

제조합 숙주 세포 내에서 지방산 합성 경로를 통한 플럭스를 증가시키기 위한 전략은 대장균 내에서 상이한 유기체들로부터 외인성 지방산 생합성 유전자들의 발현 및 본래의 대장균 지방산 생합성 유전자들의 과발현 둘 모두를 포함한다. 이 연구에서, 상이한 유기체들로부터의 지방산 생합성 유전자들은 대장균 DV2의 계놈에 조합되었다. iFABs 130-145를 함유하는 16 개의 균주들이 평가되었다. iFABs 130-145의 상세한 구조는 하기에 iFABs 표 1에 나타낸다.

표 1. iFABs 130-145에서 발견되는 성분

축약형	전체 명(Full Description)
St_fabD	쥐티푸스균 fabD 유전자
nSt_fabH	음성 RBS를 포함하는 쥐티푸스균 fabH 유전자
sSt_fabH	합성 RBS를 포함하는 쥐티푸스균 fabH 유전자
Cac_fabF	클로스트리듐 아세토부틸리쿰 (ATCC824) fabF 유전자
St_fabG	쥐티푸스균 fabG 유전자
St_fabA	쥐티푸스균 fabA 유전자
St_fabZ	쥐티푸스균 fabZ 유전자
BS_fabI	바실러스 서브틸러스 fabI 유전자
BS_FabL	바실러스 서브틸러스 fabL 유전자
Vc_FabV	콜레라균 fabV 유전자
Ec_FabI	대장균 fabI 유전자

각각의 "iFAB"는 다음과 같은 순서의 다양한 fab 유전자들을 포함한다: 1) 엔오일-ACP 환원효소(BS_fabI, BS_FabL, Vc_FabV 또는 Ec_FabI); 2) b-케토아실-ACP 합성효소 III(St_fabH); 3) 말로닐-CoA-ACP 아실전달효소(St_fabD); 4) b-케토아실-ACP 환원효소(St_fabG); 5) 3-하이드록시-아실-ACP 탈수효소(St_fabA 또는 St_fabZ); 6) b-케토아실-ACP 합성효소 II(Cac_fabF). St_fabA는 또한 trans-2, cis-3-테세노일-ACP 이성질화 효소 활성(ref)을 가지고, Cac_fabF는 b-케토아실-ACP 합성효소 II 활성 및 b-케토아실-ACP 합성효소 I 활성을

갖는다는 점에 주목한다[Zhu et al., BMC Microbiology 9:119(2009)]. iFABs 130-145의 특정 조성에 대하여, 하기의 표 5를 참조한다. FAB138의 앞쪽에 통합된 cat-loxP-T5 프로모터의 도식(도 7a); 및 iT5_138의 도식(도 7b)을 포함하는 iFAB138 유전자 자리(locus)의 도식적인 묘사를 제공하는 도 7a 및 도 7b를 참조한다.

도 2. iFABs 130-145의 조성

iFab	BS fabI	BS fabI	Vc fabV	Ec fabI	nSt fabH	sSt fabH	St fabD	St fabG	St fabA	St fabZ	Cac fabF
Ifab130	1	0	0	0	1	0	1	1	1	0	1
Ifab131	1	0	0	0	1	0	1	1	0	1	1
Ifab132	1	0	0	0	0	1	1	1	1	0	1
Ifab133	1	0	0	0	0	1	1	1	0	1	1
Ifab134	0	1	0	0	1	0	1	1	1	0	1
Ifab135	0	1	0	0	1	0	1	1	0	1	1
Ifab136	0	1	0	0	0	1	1	1	1	0	1
Ifab137	0	1	0	0	0	1	1	1	0	1	1
Ifab138	0	0	1	0	1	0	1	1	1	0	1
Ifab139	0	0	1	0	1	0	1	1	0	1	1
Ifab140	0	0	1	0	0	1	1	1	1	0	1
Ifab141	0	0	1	0	0	1	1	1	0	1	1
Ifab142	0	0	0	1	1	0	1	1	1	0	1
Ifab143	0	0	0	1	1	0	1	1	0	1	1
Ifab144	0	0	0	1	0	1	1	1	1	0	1
Ifab145	0	0	0	1	0	1	1	1	0	1	1

플라스미드 pCL_{P_{trc}}tesA는 각각의 균주들로 형질전환되었고, 발효는 32 °C 및 37 °C 둘 모두에서 얻어지도록 유도(induction)로부터 20시간 FA2 배지 내에서 처리되었다. 이 중의 플레이트 스크린으로부터 총 지방 종류의 생산에 대한 데이터는 도 6a 및 도 6b에 나타난다. 이러한 라이브러리 스크린(library screen)으로부터, 가장 좋은 구조체는 iFAB 138을 포함하는 DV2로 결정되었다. iFAB138 구조체는 균주 EG149를 만들기 위해서 균주 D178 내로 이동되었다. 이 균주는 추가적인 조작에 사용되었다. EG149의 게놈 내의 iFAB138의 서열은 SEQ ID NO: 13으로 나타난다. 표 3은 다수의 대장균 균주들의 유전적 특성을 나타내고, 본 명세서에서 설명되는 발현 구조체들을 함유하는 플라스미드들이 상기 균주들 내로 하기에 설명되는 바와 같이 도입된다. 이러한 균주들 및 플라스미드들은 본원 발명의 여하한 실시예들에 대한 재조합 숙주 세포들, 배양(culture)들 및 방법들을 입증하는데 사용되었다. 표 3의 유전 명칭(genetic designation)들은 본 기술분야의 당업자들에게 알려진 표준 명칭들이다.

표 3: 대장균 균주들의 유전적 특성

균주	유전적 특성
DV2	MG1655 F ⁻ , λ ⁻ , ilvG ⁻ , rfb-50, rph-1, ΔfhuA::FRT, ΔfadE::FRT
DV2.1	DV2 fabB::fabB[A329V]
D178	DV2.1 entD::FRT P _{T5} entD
EG149	D178 ΔinsH-11::P _{LACUV5} -iFAB138
V642	EG149 rph ⁺
SL313	V642 lacIZ::P _{A1} tesA/pDG109
V668	V642 ilvG ⁺

균주	유전적 특성
LC397	V668 lacIZ::P _{TRC} tesA(var) kan
SL571	V668 lacIZ::P _{TRC} tesA(var) FRT
LC942	SL571 attTn7::P _{TRC} tesA(var)
DG16	LC942/pLC56
V940	LC397/pV171.1
D851	SL571 yjiP::Tn5-cat/pV171.1

플라스미드들: pDG109, pLC56 및 pV171.1은 carB 및 tesA의 다양한 발현을 갖는 pCL_{P_{trc}}carB_{tesA}alrA_{fabB}fadR 오픈론이다. iFAB138은 SEQ ID NO:13이다.

실시예 4

rph 및 ilvG 돌연변이를 회복(Repairing)시킴으로써 유리 지방산(FFA) 생성물의 양 증가

ilvG 및 rph 돌연변이는 이러한 균주 내에서 수정(corrected)되어, 더 많은 FFA 생산이 야기되었다. 균주들 D178, EG149 및 V668(표 3)은 pCL_{P_{trc}}tesA로 형질전환되었다. pCL_{P_{trc}}tesA를 포함하는 균주들 D178, EG149 및 V668의 FFA 생산을 비교하기 위하여 40 시간 동안 FA2 배지 내에서 32 °C에서 발효가 진행되었다. rph 및 ilvG 돌연변이의 수정은 pCL_{P_{trc}}tesA를 포함하는 염기성(base) 균주의 FFA 생산에 있어서 116 % 증가를 야기하였다. 도 8에 나타난 바와 같이, V668/pCL_{P_{trc}}tesA는 D178/pCL_{P_{trc}}tesA 또는 EG149/pCL_{P_{trc}}tesA 대조구보

다 더 많은 FFA를 생산하였다. FFA는 LS9 생성물들의 전구체이므로, 더 많은 FFA 생산은 새로운 균주가 더 높은 수치(higher levels)로 LS9 생성물을 생산할 수 있다는 좋은 지표이다. 발효 및 추출은 하기에 예시되는 표준 FALC 발효 프로토콜에 따라서 실행되었다.

[0134]

선택되는 *대장균* 균주의 얼려진 셀 बैं크 바이알(cell bank vial)은 115 $\mu\text{g/mL}$ 의 농도로 스펙티노마이신 항생물질(spectinomycin antibiotic)을 함유하는 125 mL 격벽형(baffled) 진탕 플라스크 내의 20 mL의 LB 배양액(broth)를 주입(inoculate)하기 위하여 사용되었다. 상기 진탕 플라스크는 대략 6 시간 동안 32 $^{\circ}\text{C}$ 에서 회전형 진탕기(orbital shaker) 내에서 배양되었고, 이후에 1.25 mL의 배양액은 500 mL 격벽형 삼각 진탕 플라스크(baffled Erlenmeyer shake flask) 내의 125 mL의 적은 PFA2 접종 배지(low PFA2 seed media)[2 g/L NH_4Cl , 0.5 g/L NaCl, 3 g/L KH_2PO_4 , 0.25 g/L $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 0.015 g/L mM $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, 30 g/L 글루코오스, 1 mL/L의 미량 무기질 용액(trace minerals solution)(2 g/L의 $\text{ZnCl}_2 \cdot 4\text{H}_2\text{O}$, 2 g/L의 $\text{CaCl}_2 \cdot 6\text{H}_2\text{O}$, 2 g/L의 $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$, 1.9 g/L의 $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$, 0.5 g/L의 H_3BO_3 , 및 10 mL/L의 농축 HCl), 10 mg/L의 구연산 철(ferric citrate), 100 mM의 Bis-Tris 완충액(pH 7.0), 및 115 $\mu\text{g/mL}$ 의 스펙티노마이신] 내로 이동되었고, 32 $^{\circ}\text{C}$ 에서 하룻밤 진탕기 상에서 배양되었다. 100 mL의 적은 PFA2 접종 배양물(low P FA2 seed culture)은 처음에 1.9 L의 멸균된 F1 생물반응기 발효 배지를 포함하는, 5 L의 Biostat Aplus 생물반응기(Sartorius BBI)에 주입하는데 사용되었다. 상기 배지는 처음에 3.5 g/L의 KH_2PO_4 , 0.5 g/L의 $(\text{NH}_4)_2\text{SO}_4$, 0.5 g/L의 MgSO_4 7수화물(heptahydrate), 10 g/L의 멸균 여과된 글루코오스, 80 mg/L 구연산 철, 5 g/L 카사미노산(Casamino acids), 10 mL/L의 멸균 여과된 미량 무기질 용액, 1.25 mL/L의 멸균 여과된 비타민 용액[0.42 g/L의 리보플라빈, 5.4 g/L의 판토텐산, 6 g/L의 나이아신(niacin), 1.4 g/L의 피리독신, 0.06 g/L의 비오틴 및 0.04 g/L의 엽산(folic acid)], 및 상기 접종 배지에 이용된 것과 동일한 농도의 스펙티노마이신으로 이루어져 있다. 배양물의 pH는 28 % w/v 암모니아수를 이용하여 6.9로, 온도는 33 $^{\circ}\text{C}$ 로, 공기 공급량(aeration rate)은 1 lpm(0.5 v/v/m)로, 그리고 산소 보충(oxygen supplementation) 및 DO 제어기에 종속 접속(cascaded)된 교반 루프(agitation loop)를 이용하여 용존 산소 분압(dissolved oxygen tension)은 30 %의 포화도로 유지하였다. 거품 발생(Foaming)은 거품억제제(antifoam)[Dow Corning 1410] 기반 실리콘 에멀전의 자동적인 첨가에 의해 제어되었다.

[0135]

초기 배지 내의 글루코오스가 (주입 후 대략 4 내지 6 시간 후에) 거의 고갈된 경우에 3.9 g/L MgSO_4 7수화물 및 600 g/L 글루코오스로 이루어진 영양 공급(nutrient feed)이 0.3 hr^{-1} 의 기하급수적인 공급 속도(exponential feed rate)가 못 되는(under) 10-12 g/L/hr의 일정한 최대 글루코오스 공급 속도(constant maximal glucose feed rate)로 시작되었고, 2 L의 명목상의(nominal) 발효 부피를 기반으로 한다. 배양물이 (주입 후 대략 3 내지 4 시간 후에) 1M IPTG 저장 용액(stock solution)을 1 mM의 최종 농도에 첨가함으로써 5 AU의 OD에 이르게 될 때, 생물반응기 내에서 지방족 알코올 생산이 유도되었다. 그 후에 생물반응기는 하루에 두 번씩 샘플링(sampled)되었고, 주입 후 대략 72 시간 후에 수확되었다. 잘-혼합된 발효액(fermentation broth) 중 0.5 mL의 샘플은 15 mL 원뿔형 튜브(conical tube)[VWR 사] 내로 옮겨졌고, 5 mL의 부틸 아세테이트와 완전히 혼합되었다. 혼합하기 위하여 튜브를 몇 차례 뒤집었고, 그 후에 대략 2 분 동안 힘차게 소용돌이가 일도록 흔들었다(vortexed). 이후에, 튜브는 유기층과 수성층(organic and aqueous layers)을 분리하기 위하여 5 분 동안 원심분리되었고, 유기층의 일부(portion)는 가스 크로마토그래피 분석을 위하여 유리 바이알(glass vial) 내로 옮겨졌다.

[0136]

실시예 5

[0137]

트랜스포존 돌연변이 유발 - yijP에 의한 지방족 알코올의 생산 증가

[0138]

*대장균*에 의해 생산되는 지방족 알코올의 역가, 수율, 생산률을 개선시키기 위하여, 트랜스포존 돌연변이 유발 및 고속-대량 스크리닝이 실시되었고, 유익한 돌연변이들이 시퀀싱(sequenced)되었다. yijP 균주 내 트랜스포존 삽입은 진탕 플라스크 및 유가 배양식 발효 둘 모두에서 균주의 지방족 알코올 수율을 개선시킨 것으로 나타났다. SL313 균주는 지방족 알코올을 생산한다. 이러한 균주의 유전자형은 표 3에 제공된다. 이후에, 지방족 알코올의 생산을 측정하기 위하여 트랜스포존 클론들은 고속-대량 스크리닝을 거친다. 간단히 말해, 콜로니들은 LB를 함유하는 깊은-홈(deep-well) 플레이트에서 골라내었고, 하룻밤 성장시켰으며, 새로운 LB 내에 주입(inoculated)되었고 3시간 동안 성장시켰으며, 새로운 FA2.1 배지 내에 주입되었고, 16시간 동안 성장시켰으며, 이후에 부틸 아세테이트를 사용하여 추출되었다. 미정제 추출물은 BSTFA(N,O-bis[트리메틸실릴]트리플루오로아세트아미드)로 유도체화(derivatized)되었고, GC/FID를 사용하여 분석되었다. pDG109 플라스미드의 선택을 유지하기 위하여 모든 배지에 스펙티노마이신(100 mg/L)이 포함되었다. 대조군 균주 SL313처럼 유사한 총 지방

종류를 생산하나, 대조구보다 더 높은 퍼센트로 지방족 알코올 종류를 갖고 더 낮은 퍼센트로 유리 지방산을 갖는 클론들을 선택함으로써 히트(Hit)들이 선택되었다. 균주 68F11이 히트(hit)로서 식별되었고, FA2.1 배지를 사용하여 진탕 플라스크 발효에서 입증되었다. 대조구 균주 SL313와 트랜스포존 히트(hit) 68F11의 비교는, 두 균주 모두 유사한 역가의 총 지방 종류를 생산하나, 68F11은 대조구보다 더 높은 퍼센트로 지방족 알코올 종류를 생산한다는 것을 나타내었다. LC535로 명명한, 히트(hit) 68F11의 단일 콜로니는 트랜스포존 삽입의 위치를 확인하기 위하여 시퀀싱되었다. 간단히 말해, 게놈 DNA는 제조자의 지시에 따라서 ZR Fungal/Bacterial DNA MiniPrep™(Zymo Research Corporation, Irvine, CA) 키트(Kit)를 사용하여 10 mL 하룻밤이 지난(overnight) LB 배양물로부터 정제되었다. 정제된 게놈 DNA는 하기의 트랜스포존 내부의 프라이머들을 사용하여 트랜스포존 밖에서 시퀀싱되었다:

DG150 5'-GCAGTTATTGGTGCCCTTAAACGCCTGGTTGCTACGCCTG-3'

(SEQ ID NO:14)

DG131 5'-GAGCCAATATGCGAGAACACCCGAGAA-3' (SEQ ID NO:15)

균주 LC535는 yijP 유전자 내에 트랜스포존 삽입을 갖도록 결정되었다(도 9). yijP는 기능이 불분명한 보존된 내부 막 단백질(conserved inner membrane protein)을 인코딩한다. yijP 유전자는 오페론 내에 존재하고, 포스포에놀피루브산 카르복실라아제를 인코딩하는 ppc 유전자, 및 알려지지 않은 기능의 예상되는 DNA-결합 전사 제어인자(predicted DNA-binding transcriptional regulator)를 인코딩하는 yijO 유전자와 함께 동시에 전사(co-transcribed)된다. 아마도, 트랜스포존 내부의 프로모터들은 yijP, ppc 및 yijO의 전사 시기(timing) 및 수준(level)에 영향을 미치고, 인접한 유전자들 frwD, pflC, pflD 및 argE에도 영향을 미칠 수 있다. 트랜스포존 카세트 내부의 프로모터들은 도 9에 나타내고, 인접한 유전자 발현에 영향을 미칠 수 있다. 균주 LC535는 두 개의 상이한 날짜로 유가 배양식 발효에서 평가되었다. 두 발효 모두에서 LC535가 대조구 SL313보다 더 높은 수율을 갖는 지방족 알코올을 생산하였고, 탄소 투입량(carbon input)에 근거하여 1.3 내지 1.9 % 절대 수율(absolute yield)이 개선되었다는 점이 입증되었다. yijP 트랜스포존 카세트는 추가적으로 상이한 균주 V940에서 평가되었고, 이는 균주 SL313보다 더 높은 수율로 지방족 알코올을 생산한다. yijP::Tn5-cat 카세트는 하기의 프라이머들을 사용하여 균주 LC535로부터 증폭되었다:

LC277 5'-

CGCTGAACGTATTGCAGGCCGAGTTGCTGCACCGCTCCCGCCAGGCAG-3'

(SEQ ID NO:16)

LC278 5'-

GGAATTGCCACGGTGCGGCAGGCTCCATACGCGAGGCCAGGTTATCCAACG-3'

(SEQ ID NO:17)

선형 DNA는 균주 SL571 내로 전기천공(electroporated)되었고, lambda red 재조합 시스템을 사용하여 염색체 내로 통합되었다. 콜로니들은 트랜스포존 영역의 바깥쪽에서 하기의 프라이머들을 사용하여 스크리닝되었다:

DG407 5'-AATCACCAGCACTAAAGTGCAGCGGTTTCGTTACCCG-3' (SEQ ID NO: 18)

DG408 5'-ATCTGCCGTGGATTGCAGAGTCTATTACAGCTACG-3' (SEQ ID NO: 19)

결점 없는(correct) yijP 트랜스포존 카세트(도 9)를 포함하는 콜로니는 균주 D851을 생산하도록 생산 플라스미드 pV171.1로 형질전환되었다. D851(V940 yijP::Tn5-cat)은 yijP 트랜스포존 카세트를 함유하지 않는 동질 유전자형(isogenic) 균주 V940와 대조하여 진탕 플라스크 발효에서 테스트되었다. 이 발효의 결과는 yijP 트랜스포존 카세트가 V940 균주와 비교하여 D851 균주에 의해 더 높은 퍼센트의 지방족 알코올 생산을 부여하고, V940 대조구 균주처럼 유사한 역가의 총 지방 종류를 생산하는 것으로 나타났다. 균주 D851은 두 개의 상이한 날짜로 유가 배양식 발효에서 평가되었다. 이러한 발효들에 대한 데이터는 표 9에 나타냈고, 상기 표 4는 5-리터의 유가 배양식 발효에 있어서, yijP::Tn5-cat 트랜스포존 삽입을 갖는 균주들이 총 지방 종류("FAS") 수율의 증가 및 퍼센트 지방족 알코올("FALC")의 증가를 갖는다는 것을 분명히 보여준다. "지방 종류"는 FALC 및 FFA를 포함한다.

표 4: FAS 및 FALC의 역가 및 수율에 대한 yijP 트랜스포존 삽입의 효과

균주	FAS 역가	FAS 수율	퍼센트 FALC	FALC 수율
V940	68 g/L	18.70%	95.00%	17.80%
D851	70 g/L	19.40%	96.10%	18.60%
V940	64 g/L	18.40%	91.90%	16.90%
D851	67 g/L	19.00%	94.00%	17.80%

탱크 발효법(Tank Fermentation Method):

탱크 내에서의 지방산 에스테르들의 생산을 평가하기 위하여, 원하는 균주의 글리세롤 바이알(glycerol vial)은 진탕 플라스크 내에 20 mL LB+스펙티노마이신을 주입하기 위하여 사용되었고, 대략 여섯 시간 동안 32 °C에서 배양되었다. 4 mL의 LB 배양물은 125 mL의 적은 PFA 접종 배지(Low PFA Seed Media)[하기]에 주입하기 위하여 사용되었고, 이후에 32 °C 진탕기에서 하룻밤 배양되었다. 50 mL의 하룻밤이 지난 배양물은 1 L의 탱크 배지에 주입하기 위하여 사용되었다. 탱크는 16 g/L/hr (글루코오스 또는 메탄올)의 최대 공급량(maximum feed rate)을 가지며 pH 조절(stat) 조건 하에서 pH 7.2 및 30.5 °C으로 작동되었다.

표 5: 적은 PFA 접종 배지(Low PFA Seed Media)

성분	농도
NH ₄ Cl	2 g/L
NaCl	0.5 g/L
KH ₂ PO ₄	1 g/L
MgSO ₄ ·7H ₂ O	0.25 g/L
CaCl ₂ ·2H ₂ O	0.015 g/L
글루코오스	20 g/L
TM2 미량 무기질 용액	1 mL/L
구연산 철	10 mg/L
Bis Tris 완충액 (pH 7.0)	100 mM
스펙티노마이신	115 mg/L

표 6: 탱크 배지

성분	농도
(NH ₄) ₂ SO ₄	0.5 g/L
KH ₂ PO ₄	3.0 g/L
구연산 철	0.034 g/L
TM2 미량 무기질 용액	10 mL/L
카사미노산(Casamino acids)	5 g/L
멸균 후 첨가물(Post sterile additions)	
MgSO ₄ ·7H ₂ O	2.2 g/L
미량 비타민 용액	1.25 mL/L
글루코오스	5 g/L
접종원(Inoculum)	50 mL/L

실시예 6

CarB에 N-말단 60bp 융합 태그(Fusion Tag)의 첨가(CarB60)

단백질의 용해도, 안정성, 발현 또는 기능을 증가시키기 위한 많은 방법들이 존재한다. CarB의 용해도를 증가시키기 위한 일 접근법에서, 융합 태그가 유전자 전에 클로닝될 수 있다. CarB의 발현을 증가시키기 위한 또 다른 접근법에서, 유전자의 리보솜 결합 부위(RBS) 또는 프로모터가 변화될 수 있다. 이러한 연구에서, carB(SEQ ID NO: 7)는 N-말단 60bp 융합 태그를 첨가함으로써 변형되었다. (본 명세서에서 "CarB60"으로 칭하는) 변형된 단백질을 발생시키기 위하여, carB는 맨 먼저 하기의 프라이머들을 사용하여 pET15b 벡터 내로 클로닝되었다:

5'-GCAATTCATATGACGAGCGATGTTTCACGA-3' (SEQ ID NO:20); 및

5'-CCGCTCGAGTAAATCAGACCGAACTCGCG (SEQ ID NO:21).

pET15b - carB 구조체는 carB 유전자의 바로 상류에(directly upstream) 60 개의 뉴클레오타이드들을

함유하였다:

5'-
ATGGGCAGCAGCCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAG
CCAT (SEQ ID NO:22)

[0158]

[0159]

carB의 융합 태그 버전은 carB60으로 재명명하였다. 이후에, pET15b_carB60은 제한 효소(restriction enzyme)들 NcoI 및 HindIII을 사용하여 분해(digested)되었고, 동일한 효소들에 의해 절단(cut)되는 pCL1920-유래되는 벡터 OP80 내로 서브클로닝(subcloned)되었다. 이러한 플라스미드는 FALC 생산을 평가하기 위하여 균주 V324(MG1655 ΔfadE::FRT ΔfhuA::FRT fabB::A329V entD::T5-entD lacIZ::P_{TRC}-'TesA)로 형질전환되었다. 균주들은 (하기에 요약되는) 표준 절차에 따라서 발효되었고, 총 지방 종류의 역가 및 총 지방족 알코올의 역가가 정량화(quantified)되었다. 도 10은 CarB60이 지방족 알코올의 역가를 증가시키므로, 다중복제(multicopy) 플라스미드로부터 발현되는 경우에 CarB60 효소가 CarB보다 더 높은 총 세포 활성을 갖는다는 것을 나타낸다.

[0160]

생산 균주들의 지방족 알코올의 생산을 평가하기 위하여, 형질전환체들은 37 °C에서 항생물질(antibiotics)로 보충된 2 ml의 LB 배양액 내에서 성장되었다. 하룻밤 성장한 후에, 40 ul의 배양물은 항생물질로 보충된 2 ml의 새로운 LB 내로 옮겨졌다. 3 시간 성장한 후에, 2 ml의 배양물은 20 μl 미량 무기질 용액, 10 μg/L 구연산철(iron citrate), 1 μg/L 티아민 및 항생물질(FA2 배지)로 보충된 3 % 글루코오스를 포함하는 20 ml의 M9 배지를 함유하는 125 mL 플라스크 내로 옮겨졌다. 배양물의 OD₆₀₀이 1.0에 도달하는 경우에, 1 mM의 IPTG는 각각의 플라스크에 첨가되었다. 37 °C에서 20 시간 동안 성장한 후에, 각각의 플라스크로부터 400 μl의 샘플들을 떼어냈고, 지방족 알코올들이 400 μl의 부틸 아세테이트로 추출되었다. 향상된 CarB 활성의 메커니즘을 더욱 이해하기 위하여, CarB60은 'TesA를 함유하지 않는 균주 D178(MG1655 ΔfadE::FRT ΔfhuA::FRT fabB::A329V entD::P_{T5}-entD)로부터 정제되었다. 간단히 말해서, pCL1920_carB60은 균주 D178로 형질전환되었고, 이는 지방족 알코올 생산을 위하여 조작되며, 발효는 스펙티노마이신(100 μg/ml)로 보충된 FA-2 배지에서 37 °C에서 실시되었다. 배양물의 OD₆₀₀이 1.6에 도달하는 경우에, 세포들은 1 mM 이소프로필-β-D-티오갈락토피라노사이드(isopropyl-β-D-thiogalactopyranoside: IPTG)로 유도되었고, 37 °C에서 추가적으로 23 시간 동안 배양되었다. CarB60의 정제를 위하여, 세포들은 4,500 rpm으로 4 °C에서 20 분 동안 원심분리에 의해 수확되었다. 세포 페이스트(cell paste)[10 g]는 프로테아제 억제제인자 콕테일 용액(protease inhibitor cocktail solution) 및 BugBuster MasterMix[Novagen 사]의 12 ml 내에 현탁되었다. 세포들은 프렌치 프레스(French Press)에 의해 분쇄되었고, 결과적으로 얻어진 세포분쇄액(homogenate)은 세포 파괴물(cellular debris)을 제거하기 위하여 10,000 rpm으로 원심분리되었다. Ni-NTA가 결과적으로 얻어진 혼합물에 첨가되었고, 현탁액(suspension)은 회전식 진탕기(rotary shaker) 상에서 1 시간 동안 100 rpm으로 4 °C에서 빙빙 돌렸다(swirled). 슬러리(slurry)는 컬럼 내로 부어졌고, 통과액(flow-through)이 수집되었다. Ni-NTA 수지는 300 mM NaCl을 함유하는 pH 8.0의 50 mM 인산나트륨 완충액 내의 10 mM 이미다졸로 세정되었고, 300 mM NaCl을 함유하는 pH 8.0의 50 mM 인산나트륨 완충액 내의 20 mM 이미다졸로 추가적으로 세정되었다. CarB60 단백질은 300 mM NaCl을 함유하는 pH 8.0의 50 mM 인산나트륨 완충액 내의 250 mM 이미다졸로 용리(eluted)되었고, SDS-PAGE에 의해 분석되었다. 단백질은 pH 7.5의 50 mM인산나트륨 완충액 내의 20 %(v/v) 글리세롤에 의해 투석되어, 배양물의 리터(liter) 당 대략 10 mg의 CarB60이 산출되었다. 단백질은 급속 냉동(flash frozen)되었고, 필요할 때까지 -80 °C로 저장되었다.

[0161]

CarB60 단백질은 다중복제 플라스미드로부터 충분히(abundantly) 발현되었다. 추가적인 SDS-PAGE 분석은 CarB60의 발현이 CarB 보다 더 높다는 것을 나타낸다. CarB60의 더 높은 발현 수준은 대장균 염색체 내로 통합되는 carB60 유전자가 동일한 위치의 carB 유전자보다 더 많은 단백질을 생산한다는 것을 시사한다. 이러한 가설을 테스트하기 위하여, carB60 유전자는 대장균 염색체 내로 통합되었다. 간단히 말해서, carB60 유전자는 정방향 프라이머(forward primer):

[0162]

5'-ACGGATCCCCGGAATGCGCAACGCAATTAATGTaAGTTAGCGC-3' (SEQ ID NO:23) ; 및 역방향 프라이머(reverse primer):

[0163]

5'-TGCGTCATCGCCATTGAATTCCTAAATCAGACCGAACTCGCGCAGG-3' (SEQ ID NO:24) 를 사용하여 pCL_carB60로 맨 먼저 증폭되었다.

[0164]

제 2 PCR 생성물은 정방향 프라이머:

- [0165] 5'-ATTCCGGGGATCCGTCGACC-3' (SEQ ID NO:25); 및 역방향 프라이머:
- [0166] 5'-AATGGCGATGACGCATCCTCACG-3' (SEQ ID NO:26) 를 사용하여 벡터 pAH56으로 증폭되었다.
- [0167] 이러한 분절(fragment)은 카나마이신 저항성 카세트(kanamycin resistance cassette), λ attP 부위 및 γ R6k 복제 기점(origin of replication)을 함유한다. 두 개의 PCR 생성물들은 플라스미드 pSL116-126을 생성하기 위하여 InFusion 키트(kit)(Clontech 사)를 사용하여 접합(joined)되었다. 헬퍼 플라스미드(helper plasmid) pINT 및 'TesA12H08의 통합 형태(integrated form)를 함유하는 지방족 알코올 생산 균주는 carB60 유전자를 함유하는 pSL116-126 또는 carB 유전자를 함유하는 플라스미드 F27 중 어느 하나로 형질전환되었다. 이러한 균주들은 상기에 설명된 바와 같은, 진탕 플라스크 발효에 대한 표준 절차들에 따라서 FA2 배지 내에서 발효되었다. 지방족 알코올들 및 지방산 에스테르들을 특징지우고 정량화하기 위하여, 불꽃 이온화 검출기(flame ionization detection: "FID")에 연결된 가스 크로마토그래피("GC")가 사용되었다. 미정제 추출물은 BSTFA(N,O-bis[트리메틸실릴]트리플루오로아세트아미드)로 유도체화(derivatized)되었고, GC/FID를 사용하여 분석되었다. 정량화는 가스 크로마토그래피(GC), 질량 분석기(mass spectroscopy: MS), 박막 크로마토그래피(thin layer chromatography: TLC), 고-성능 액체 크로마토그래피(HPLC), 액체 크로마토그래피(LC), 불꽃 이온화 검출기에 연결된 GC(GC-FID), GC-MS 및 LC-MS를 포함하나 이로 제한되지 않는 사용될 수 있는 분석법뿐만 아니라, 상기에 설명된 GC 방법을 사용하여 다양한 농도의 적절한 인증된 표준 물질(references)을 주입함으로써 실시되었다. 폴리펩티드의 발현에 대하여 테스트하는 경우에, 웨스턴 블로팅(Western blotting) 및 점 블로팅(dot blotting)과 같은 기술들이 사용될 수 있다.
- [0168] 발효 20 시간 후의 결과들을 도 11에 나타낸다. 두 개의 균주들의 총 지방 생성물 역가는 유사(2.4 g/L 총 지방 종류)하지만, 통합된 CarB60은 N-말단 태그가 없는 CarB와 비교하여 더 많은 비율(greater fraction)의 C12 및 C14 사슬 길이의 유리 지방산들을 지방족 알코올들로 전환시킨다. 이러한 데이터는 CarB60을 발현시키는 세포들이 더 높은 세포의 카르복실산 환원효소 활성을 가지고, 더 많은 FFA를 지방족 알코올들로 전환시킬 수 있다는 것을 시사한다. 따라서, 염색체 내에 통합되는 경우에 carB60은 개선된 carB 템플릿(template)이고, 상기 개선된 carB 템플릿은 개선된 carB 변이체들을 식별하기 위하여 carB 유전자를 발달(evolver)시키기 위한 원하는 활성을 제공한다.
- [0169] 실시예 7
- [0170] CarB 돌연변이체의 발생(Generation)
- [0171] CarB 효소는 여하한 처리 조건들 하에서 지방족 알코올의 생산에 있어서 속도-제한 단계(rate-limiting step)에 존재한다. 효율적으로 지방족 알코올을 생산하기 위하여, CarB 효소의 활성을 증가시키기 위한 노력들이 이뤄졌다.
- [0172] 오류 빈발 PCR 라이브러리 스크린(Error prone PCR library screen):
- [0173] 오류 빈발 PCR을 사용한 무작위 돌연변이 유발(Random mutagenesis)은 DNA 중합효소의 복제 정확도(copying fidelity)가 낮은 조건들 하에서 수행되었다. 돌연변이 유발되는 핵산들은 벡터로 클로닝되었고, 오류 빈발 PCR에 뒤이어 고속-대량 스크리닝이 (하기에 상세하게 설명되는 바와 같이) 유리 지방산들의 지방족 알코올들로의 전환을 증가시키는 유익한 돌연변이들을 찾기 위하여 실시되었다. 중요한 잔기들은 추가적으로 다른 아미노산들로 돌연변이시켰다. 다수의 단일 아미노산 돌연변이들 및 돌연변이들의 조합들은 지방족 알코올들로 전환되는 지방 종류의 비율(fraction)을 증가시켰다. 간단히 말해서, 무작위 돌연변이는 Genemorph II 키트(Stratagene 사)를 사용하여 오류 빈발 PCR에 의해 carB60opt 유전자에서 발생되었다. 돌연변이는 클로닝이 용이하도록 개별적으로 carB60opt 두 개의 도메인(domain)들 중 오직 한 도메인에서 발생되었다. 라이브러리 1은 carB60opt의 첫번째 759 잔기들을 함유하였고, 하기의 프라이머들을 사용하여 오류 빈발 PCR에 의해 발생되었다:
- HZ117 5'-ACGGAAAGGAGCTAGCACATGGGCAGCAGCCATCATCAT-3'
(SEQ ID NO:27); 및
- [0174] DG264 5'-GTAAAGGATGGACGGCGGTCACCCGCC-3' (SEQ ID NO:28). 라이브러리 1에 대한 벡터는 효소들 NheI 및 PshAI에 의해 분해(digested)되는 플라스미드 pDG115이다. 라이브러리 2는 carB60opt의 마지막 435 잔기들을 함유하였고, 하기의 프라이머들을 사용하여 오류 빈발 PCR에 의해 발생되었다:

DG263 5'-CACGGCGGGTGACCGCCGTCCATCC-3'(SEQ ID NO:29); 및
HZ118 5'-TTAATTCCGGGGATCCCTAAATCAGACCGAACTCGCGCAGGTC-3'
(SEQ ID NO:30).

라이브리 2에 대한 벡터는 PshAI 및 BamHI에 의해 분해(digested)되는 플라스미드 pDG115이다. 오류 빈발 삽입체(error-prone insert)들은 InFusion Advantage(Clontech 사)를 사용하여 벡터들로 클로닝되었고, 클로닝 균주 NEB Turbo(New England Biolabs 사)를 통하여 통과되었다. 이후에, 라이브러리들은 균주 EG442(EG149 Tn7::P_{TRC}-ABR lacIZ::P_{T50}-ABR)로 형질전환되었다. 이후에, 오류 빈발 carB60opt 클론들은 지방족 알코올의 생산을 측정하기 위하여 고속-대량 스크리닝되었다. 간단히 말해서, 콜로니들은 LB를 함유하는 깊은-홈(deep-well) 플레이트들에서 골라내었고, 하룻밤 성장시켰으며, 새로운 LB 내에 주입(inoculated)되었고 3시간 동안 성장시켰으며, 새로운 FA-2.1 배지 내에 주입되었고, 16시간 동안 성장시켰으며, 이후에 부틸 아세테이트를 사용하여 추출되었다. 미정제 추출물은 BSTFA(N,O-bis[트리메틸실릴]트리플루오로아세트아미드)로 유도체화(derivatized)되었고, 표준 GC/FID 방법을 사용하여 분석되었다. pDG115 플라스미드의 선택을 유지하기 위하여 모든 배지에 스펙티노마이신(100 mg/L)이 포함되었다. 대조군 균주와 비교하여 더 적은 총 유리 지방산의 역가 및 더 큰 총 지방족 알코올의 역가를 생산하는 클론들을 선택함으로써 히트(Hit)들이 선택되었다. 상이한 발효 스크린들로부터의 히트(hit)들을 비교하여 위하여, 유리 지방산의 지방족 알코올로의 전환은 정규화된 유리 지방산 NORM FFA = 돌연변이 퍼센트 FFA / 대조군 퍼센트 FFA의 백분율을 계산함으로써 정규화되었고, 여기서 "퍼센트 FFA"는 총 유리 지방산 종류의 역가 나누기 총 지방 종류의 역가이다. 하기에 설명되는 바와 같이, 진탕 플라스크 발효를 사용하여 히트(Hit)들이 추가적으로 검증되었다.

히트(Hit)들은 유익한 돌연변이들을 식별하기 위하여 시퀀싱되었다. 시퀀싱은 하기의 프라이머들을 사용하여 전체 carB60opt 유전자의 콜로니 PCR에 의해 수행되었고:

SL59 5'-CAGCCGTTTATTGCCGACTGGATG-3'(SEQ ID NO:31); 및

EG479 5'-CTGTTTTATCAGACCGCTTCTGCGTTC-3' (SEQ ID NO:32), 프라이머들을 사용하여 carB60opt 효소 내부에 시퀀싱되었다. CarB60opt 효소를 개선시키는 유익한 돌연변이들은 표 7에 나타났다. 정규화된 유리 지방산(NORM FFA) 란(column)은 효소들의 개선을 나타내고, 수치가 작을수록 가장 좋은 개선을 나타낸다. "홈(Well) #"는 이러한 돌연변이가 발견되는 최초(primary) 스크리닝 홈을 나타낸다. 모든 잔기 번호(residue numbers)는 CarB 단백질 서열을 칭하고, 60 bp 태그를 포함하지 않는다. 돌연변이들은 접두부(prefix) "태그(Tag)": 60 bp/20 잔기 N-말단 태그 내의 표시 돌연변이(indication mutations)와 같이 나타낸다.

표 7: 오류 빈발 스크리닝 동안 식별된 CarB 효소의 유익한 돌연변이들(TAG 돌연변이들 제거)

홈(Well) #	Norm FFA	미스센스 돌연변이(Missense Mutations)	침묵성 돌연변이(Silent Mutations)
131B08	70.50%	L799M V810F S927R M1062L A1158V F1170I	CCG1116CCT
20C07	71.80%	A535S	
65B02	74.70%	M930R	ACC867ACA
54B10	76.30%	L80Q T231M F288L A418T V530M A541V G677D P712A	
67E1	78.20%	D750G R827C D986G G1026D P1149S	GCA1031GCT GTC1073GTT
65C03	78.90%	V926A	ATT941ATA
12C10	80.30%	V46I	
66E08	80.10%	V926A	
70F02	80.90%	D750G R827C D986G G1026D P1149S	GCA1031GCT GTC1073GTT
07D01	82.40%	E20K V191A	
66G09	82.40%	R827C L1128S	ACG780ACA CTG923TTG
25H02	83.50%	F288S	
06C01	85.10%	V46I	06C01
05D02	85.20%	T396S	CCG477CCT
124E03	86.00%	R827C L1128S	ACG780ACA CTG923TTG
17A04	86.20%	A574T	GCA237GCT ACC676ACT GCC529GCT
132C08	87.00%	M1062T R1080H	TTG830ITA TAC834TAT
72C09	87.30%	P809L M1062V	
10F02	87.70%	E636K	
71H03	88.10%	R827C L1128S	ACG780ACA CTG923TTG
38G04	88.90%	D143E A612T	GCA181GCG
42F08	90.20%	T90M	CTG186CTT
66C04	90.30%	L1128S	
18C03	90.40%	Q473L	
12E02	90.60%	D19N S22N R87H L416S	CCG167CCA
28B09	91.10%	E28K H212N Q473L	CCG122CCA ACG178ACA CTG283TTG CTG340CTA ACC401ACT GCA681GCG
103E09	92.20%	E936K P1134R	CGT829CGG CTG1007CTA
03E09	93.20%	M259I	
74G11	93.80%	I870V S927I S985I I1164F	GTG1000GTC
46C01	95.60%	D18V D292N	

[0181] 포화 돌연변이 유발(Saturation Mutagenesis)[발생되는 콤보(Combo) 1 및 2 라이브러리]:

[0182] 오류 빈발 PCR 후에 지방족 알코올 생산에 유익하다고 여겨지는 아미노산 위치들은 추가적으로 돌연변이 유발되었다. 축퇴된(degenerate) 뉴클레오타이드들 NNK 또는 NNS를 함유하는 프라이머들은 이러한 위치들을 다른 아미노산들로 돌연변이 시키는데 사용되었다. 결과적으로 얻어진 "포화 돌연변이 유발 라이브러리들"은 오류 빈발 라이브러리들에 대하여 상기에 설명된 바와 같이 스크리닝되었고, 히트(hit)들은 지방족 알코올 전환을 더욱 개선시켰다(모 "대조구" 균주와 비교하여 더 적은 총 유리 지방산의 역가 및 더 큰 총 지방족 알코올의 역가)는 것으로 확인되었다. 지방족 알코올의 생산을 개선시키는 9 개의 상이한 위치들에 있어서 단일 아미노산/코돈 변화는 표 8에 나타낸다. 본 명세서에 설명되는 바와 같이, 진탕 플라스크 발효를 사용하여 히트(Hit)들이 추가적으로 검증되었다.

[0183] 표 8: 아미노산 포화 돌연변이 유발 동안 식별된 CarB 효소의 유익한 돌연변이들

WT 아미노산	WT 코돈	돌연변이형 아미노산	돌연변이형 코돈	Norm FFA
E20	GAG	F	TTC	92.20%
		L	CTG	94.50%
		L	TTG	96.20%
		R	CGC	86.50%
		S	TCG	87.40%
		V	GTG	86.00%
V191	GTC	V	GTC	85.30%
		Y	TAC	88.80%
		A	GCC	88.70%
		S	AGT	98.00%
		G	GGG	70.30%
		R	AGG	77.20%
F288	TTT	S	TCT	85.60%
		S	AGC	79.60%
		A	GCG	89.50%
		F	TTC	89.10%
		H	CAC	84.10%
		I	ATC	77.20%
Q473	CAA	K	AAG	90.30%
		L	CTA	90.10%
		M	ATG	89.00%
		R	AGG	88.00%
		V	GTG	89.20%
		W	TGG	84.50%
A535	GCC	Y	TAC	86.00%
		A	TCC	71.80%
		A	GCC	93.20%
		C	TGT	87.90%
		C	TGC	83.20%
		A	GCT	78.10%
R827	CGC	A	GCG	66.30%
		A	GCC	69.50%
		E	GAG	65.80%
		G	GGC	78.60%
		G	GGG	77.60%
		G	GGT	79.30%
V926	GTT	I	ATC	90.80%
		K	AAG	70.70%
		V	GTG	87.90%
		K	AAG	82.30%
		R	CGG	73.80%
		R	AGG	69.80%
S927	AGC	A	GCG	92.70%
		G	GGG	89.70%
		K	AAG	94.80%
		M	ATG	95.80%
		P	CCG	98.40%
		R	AGG	90.90%
M930	ATG	R	CGG	88.50%
		S	TCG	88.90%
		T	ACG	96.30%
		V	GTG	93.90%
		W	TGG	78.80%
		Y	TAC	87.90%
L1128	TTG	Y	TAC	87.90%

[0184]

[0185] 지방족 알코올 생산에 유익하다고 여겨지는 아미노산 치환(substitution)들이 다음에 조합되었다. PCR은 다양한 원하는 돌연변이들을 함유하는 carBopt 유전자의 일부분들을 증폭시키기 위하여 사용되었고, PCR-기반 방법을 사용하여 이러한 일부분들이 서로 접합되었다(Horton, R.M., Hunt, H.D., Ho, S.N., Pullen, J.K. and Pease, L.R. 1989). carBopt 유전자는 60 bp N-말단 태그 없이 스크리닝되었다. 이러한 조합 라이브러리에 조합된 돌연변이들을 표 9에 나타낸다.

표 9: 첫번째 조합 라이브러리로부터의 CarB 돌연변이들

돌연변이	코돈
E20V	GTG
E20S	TCG
E20R	CGC
V191S	AGT
F288R	AGG
F288S	AGC
F288G	GGG
Q473L	CTG
Q473W	TGG
Q473Y	TAC
Q473I	ATC
Q473H	CAC
A535S	TCC

스크리닝을 용이하게 하기 위하여, 이후에 결과적으로 얻어진 CarB 조합 라이브러리는 lacZ 유전자 자리(locus)에 균주 V668의 염색체 내로 통합되었다. 이러한 유전자 자리에 carBopt 유전자의 서열은 SEQ ID NO: 7로 나타낸다. 균주 V668의 유전자형은 (표 3 및 도 16에 나타낸 바와 같은) MG1655(Δ fadE::FRT Δ fhuA::FRT Δ fabB::A329V Δ entD::T5-entD Δ insH-11::P_{lacUV5}fab138 rph+ ilvG+)이다. 이후에, 균주들은 플라스미드 pVA3로 형질전환되었고, 이는 TesA, 포스포판테테인 부착 부위를 파괴하는 촉매 비활성(catalytically inactive) CarB 효소인 CarB[S693A], 및 유리 지방산의 생산을 증가시키는 다른 유전자들을 함유한다. 조합 라이브러리는 상기에 설명된 바와 같이 오류 빈발 라이브러리에 대하여 스크리닝되었다. pVA3을 함유하고 lacZ 영역 내에 통합된 carB opt(A535S)를 갖는 V668은 대조구로서 사용되었다. 지방족 알코올의 생산을 증가시키는 히트(Hit)들이 선택되었고, 실시예 5에 설명된 바와 같이, 진탕 플라스크 발효를 사용하여 추가적으로 검증되었다. CarB 조합 돌연변이체를 발현시키는 제조법 숙주 세포들의 진탕 플라스크 발효 후의 지방족 알코올 생산의 향상된 백분율을 도 12에 나타냈다.

통합된 CarB 조합 돌연변이체들은 하기의 프라이머들을 사용하여 PCR에 의해 통합된 carB 히트(hit)들로부터 증폭되었다:

EG58 5'- GCACTCGACCGGAATTATCG (SEQ ID NO:33); 및

EG626 5'- GCACTACGCGTACTGTGAGCCAGAG (SEQ ID NO:34).

이러한 삽입체(insert)들은 하기의 프라이머들을 사용하여 재-증폭되었다:

DG243 5'- GAGGAATAAACCATGACGAGCGATGTTTCACGACGCGACCGACGGC (SEQ ID NO:35); 및

DG210 5'- CTAAATCAGACCGAACTCGCGCAGG (SEQ ID NO:36).

주입 클로닝(InFusion cloning)을 사용하여, 풀링된(pooled) carB 돌연변이체들은 생산 플라스미드인 pV869로 클로닝되었고, 이는 하기의 프라이머들을 사용하여 증폭된 PCR이었다:

DG228 5'- CATGGTTTATTCCTCCTTATTTAATCGATAC(SEQ ID NO:37); 및

DG318 5'- TGACCTGCGCGAGTTCGGTCTGATTTAG (SEQ ID NO:38).

진탕 플라스크 발효 플라스미드 스크린에서 가장 잘 수행된 carB 돌연변이(carB2; 표 11)는 VA101로 표기하였고, carBopt를 수송(carrying)하는 대조구 균주[A535S]는 VA82로 표기하였다. 도 13을 참조한다.

지방족 알코올 생산에 유익하다고 여겨지는 carB의 환원성 도메인(reduction domain) 내에 아미노산 치환들은 가장 좋은 carB-L 조합 라이브러리 히트(hit)들 중 하나와 조합되었다, "carB3"(표 11). PCR은 환원성 도메인 내에 다양한 원하는 돌연변이를 함유하는 carBopt 유전자의 일부분들을 증폭시키기 위하여 사용되었고, SOE PCR을 사용하여 이러한 일부분들이 서로 접합되었다. 이러한 조합 라이브러리에 조합된 돌연변이들을 표 10에 나타낸다.

[0198] 표 10: 두번째 조합 라이브러리로부터의 CarB 돌연변이들

돌연변이	코돈
R827C	TGC
R827A	GCA
V926A	GCG
V926E	GAG
S927K	AAG
S927G	GGG
M930K	AAG
M930R	AGG
L1128W	TGG

[0199]

[0200]

조합 라이브러리는 상기에 설명된 바와 같이 오류 빈발 라이브러리에 대하여 스크리닝되었다. pVA3를 함유하고 lacZ 영역 내에 통합된 carB3를 갖는 V668은 대조구로서 사용되었다. 지방족 알코올의 생산 증가를 나타내는 히트(Hit)들이 선택되었고, 상기에 설명된 바와 같이, 진탕 플라스크 발효를 사용하여 추가적으로 검증되었다. 추가적인 CarB 조합 돌연변이(carB4)를 사용하여 지방족 알코올 생산의 향상된 백분율을 나타내는 진탕 플라스크 발효의 결과들을 표 11에 나타낸다. 적은 복제(low copy) CarB 변이체들의 상대적인 전환 효율에 대한 도식 표현은 도 14에 나타낸다. 표 11에 기록된 결과들은 동일한 조건들 하에서 실시되는 생물반응기 작동(runs)에서 얻어졌다.

[0201]

표 11: CAR 변이체들

명칭	돌연변이(들)	균주	탱크 데이터(Tank data)	비고(Notes)
carB	None = WT (E20 V191 F288 Q473)			단백질은 SEQ ID NO:7
carB60	None + tag	V324		
carB1	A535S	V940	83% FALC; C12/C14=3.4	12H08 염색체의 TE의 하나의 복제(copy)를 가짐
carB2	E20R, F288G, Q473I, A535S	LH375	97% FALC; C12/C14=3.6	12H08 염색체의 TE의 두 개의 복제(copy)를 가짐
carB2	E20R, F288G, Q473I, A535S	LH346	96% FALC; C12/C14=3.7	12H08 염색체의 TE의 하나의 복제(copy)를 가짐
carB3	E20R, F288G, Q473H, A535S	L 클로 라이브러리	생물반응기 내에 데이터가 될 샘플들이 없음	
carB4	E20R, F288G, Q473H, A535S, R827A, S927G	R 클로 라이브러리 (VA-219)	97% FALC; C12/C14=3.9	12H08 염색체의 TE의 두 개의 복제(copy)를 가짐
carA	None	미국 특허 공개 No. 20100105963 참조		단백질은 SEQ ID NO:39
FadD9	None	미국 특허 공개 No. 20100105963 참조		단백질은 SEQ ID NO:40

[0202]

CarA, FadD9, CarB 및 CarB60의 DNA 서열들은 각각 본 명세서에서 SEQ ID NO: 41, 42, 43 및 44로 나타낸다.

[0204]

포화 돌연변이 유발에 의한 CarB 효소 내의 추가적인 유익한 돌연변이들의 식별:

[0205]

이중-플라스미드 스크리닝 시스템이 후에 개발되었고, FALC 생산에 대한 CarB4에 비하여 개선된 CarB 변이체들을 식별하기 위하여 검증되었다. 이중-플라스미드 시스템은 다음의 기준을 충족하였다: 1) 돌연변이형 클론들은 CarB 활성을 초과하여 지방산 플럭스를 제공하기 위하여 높은 FA 역가를 제공한다. 이는 유리 지방산의 생산을 향상시키기 위하여 촉매 비활성(catalytically inactive) CarB 효소인 CarB[S693A]와 함께 FALC 오페론을 수송하는 플라스미드(pLYC4, pCL1920_P_{TRC}-carDead_tesA_alrAadp1_fabB[A329G]_fadR)로 염기성(base) 균주(염색체의 TE의 두 개의 복제(copy)를 갖는 V668)를 형질전환시킴으로써 달성된다; 2) 바람직하게 9-kb보다 더 작은 carB 돌연변이형 템플릿을 갖는 스크리닝 플라스미드는 포화 돌연변이 유발 절차에 처리될 수 있고, 발현을 위하여 pLYC4와 양립할 수 있다; 3) CarB 활성의 동적 범위(dynamic range)는 조정할 수 있다. 이는 CarB4 발현 수준을 조정하기 위하여 약한(weaker) 프로모터(P_{TRC1}) 및 대안적인 개시 코돈들(GTG 또는 TTG)을 조합시킴으로써 달성된다. 3) 좋은 플라스미드 안정성, 독소/항독소 모듈(toxin/antitoxin module)[ccdAB 오페론]은 플라스미드 안정성을 유지하기 위하여 도입되었다.

[0206]

간단히 말해서, 스크리닝 플라스미드 pBZ1(pACYCDuet-1_P_{TRC1}-carB4GTG_rnBter_ccdAB)는 In-Fusion HD 클로닝 방법(Clontech 사)을 사용하여 동일한 몰 비(molar ratios)의 네 개의 부분들(P_{TRC1}, ATG/TTG/GTG 개시 코돈들을 갖는 carB4, ccdAB를 갖는 rnB T1T2 종결자, 및 pACYCDuet-1 벡터)을 혼합함으로써 네 개의 부분들로 구성되었다. 상기 부분들(1 내지 4)은 하기의 프라이머 쌍(primer pair)들에 의해 증폭되는 PCR이었다: (1) P_{TRC1} - 템플릿으로서 pVA232(pCL1920_P_{TRC}-carB4_tesA_alrAadp1_fabB[A329G]_fadR) 플라스미드를 사용하는, 정방향 프

5'
CGGTTCTGGCAAATATTCTGAAATGAGCTGTTGACAATTAATCAAATCCGGCTCGTA
라이머 TAATGTGTG-3' (SEQ ID NO:45)

및 역방향 프라이머

5'-GGTTTATTCCTCCTTATTTAATCGATACAT-3' (SEQ ID NO:46) . (2) ATG/TTG/GTG 개시 코돈들을 갖는 carB4 - 템플릿으로서 pVA232 플라스미드를 사용하는, 정방향 프라이머

carB4 ATG
5'ATGTATCGATTAAATAAGGAGGAATAAACCATGGGCACGAGCGATGTTACACGACG
CGAC-3' (SEQ ID NO:47); carB4 GTG

5'ATGTATCGATTAAATAAGGAGGAATAAACCGTGGGCACGAGCGATGTTACACGACG
CGAC-3' (SEQ ID NO:48); 및 carB4 TTG 5'-

ATGTATCGATTAAATAAGGAGGAATAAACCTTGGGCACGAGCGATGTTACACGACG

GAC-3' (SEQ ID NO:49);

및 역방향 프라이머

carB4 rev 5'-TTCTAAATCAGACCGAACTCGCGCAG-3' (SEQ ID NO:50) . (3) ccdAB를 갖는 rrnB T1T2 종결자 - 플라스미드 pAH008(pV171_ccdBA 오페론)을 사용하는, 정방향 프라이머 rrnB T1T2 term 5'-

CTGCGCGAGTTCGGTCTGATTAGAAATTCCTCGAGGATGGTAGTGTGG-3' (SEQ ID NO:51)

및 역방향 프라이머

ccdAB rev 5'-CAGTCGACATACGAAACGGGAATGCGG-3' (SEQ ID NO:52) . (4) pACYCDuet-1 벡터 골격(backbone) -

5'
CCGCATTCCCGTTTCGTATGTCGACTGAAACCTCAGGCATTGAGAAGCACACGGTC-
3' (SEQ ID NO:53)

에 대한 정방향 프라이머 pACYC 벡터 및 역

방향 프라이머 pACYC 벡터

rev 5'-
CTCATTTTCAGAATATTTGCCAGAACCGTTAATTTCTAATGCAGGAGTCGCATAAG-3'
(SEQ ID NO:54).

pBZ1 플라스미드는 상기에 설명된 균주 내에서 pLYC4와 동시에 발현되었고, 진탕 플라스크 발효 및 깊은-홈(deep-well) 플레이트 발효에 의해 검증되었다. 발효 조건들은 최적화되어, CarB4_GTG 템플릿은 실시예 5에 설명된 바와 같이 둘 모두의 발효 플랫폼에서 재현가능하게(reproducibly) ~65 % FALC 전환을 갖도록 하였다. 진탕 플라스크 발효에 대한 결과들을 도 15에 나타냈다.

개선된 CarB 변이체들 내에 돌연변이들(표 7)을 함유하는 추가적인 부위들(18, 19, 22, 28, 80, 87, 90, 143, 212, 231, 259, 292, 396, 416, 418, 530, 541, 574, 612, 636, 677, 712, 750, 799, 809, 810, 870, 936, 985, 986, 1026, 1062, 1080, 1134, 1149, 1158, 1161, 1170)은 전체(full) 포화 돌연변이 유발되었다. 축퇴된(degenerate) 뉴클레오타이드들 NNK 또는 NNS를 함유하는 프라이머들은 PCR-기반 방법에 의해 이러한 위치들을 다른 아미노산들로 돌연변이 시키는데 사용되었다(Sawano and Miyawaki 2000, Nucl. Acids Res. 28: e78). 포화 라이브러리는 pBZ1(pACYCDuet-1_P_{TRC1}-carB4GTG_rrnBter_ccdAB) 플라스미드 템플릿을 사용하여 구성되었다. 돌연변이형 클론들은 NEB Turbo(New England Biolabs 사) 클로닝 균주들로 형질전환되었고, 플라스미드들은 분리되었고 풀링(pooled)되었다. 이후에, 풀링(pooled)된 플라스미드들은 플라스미드 pLYC4를 수반하는(carrying) V668 기반 균주로 형질전환되었고, 형질전환체들은 항생물질(100 mg/L 스펙티노마이신 및 34 mg/L 클로람페니콜)로 보충된 LB 한천(agar) 플레이트들에서 선택되었다.

이후에, 포화 라이브러리로부터의 CarB 변이체들은 지방족 알코올 생산에 대하여 스크리닝되었다. 단일 콜로니들은 실시예 5에 설명된 바와 같은 변형된 깊은-홈(deep-well) 플레이트 발효 프로토콜에 따라서 96-홈 플레이트들에서 직접 골라내었다. 대조군 균주와 비교하여 더 작은 총 유리 지방산의 역가 및 더 큰 총 지방족 알코올의 역가를 생산하는 클론들을 선택함으로써 히트(Hit)들이 선택되었다. 상이한 발효 배치(batch)들로부터의 히트(hit)들을 비교하기 위하여, 유리 지방산의 지방족 알코올로의 전환은 정규화된 유리 지방산의 백분율을 계산함으로써 정규화되었다. 또한, NORM FFA(%)는 실시예 5에 설명된 바와 같이 히트 확인(hits validation)에 사용되었다. NORM FFA(%) = 돌연변이 퍼센트 FFA / 대조군 퍼센트 FFA; 여기서 "퍼센트 FFA"는 총 유리 지방산 종류의 역가 나누기 총 지방 종류의 역가이다. 실시예 5에 설명된 바와 같이, 진탕 플라스크 발효를 사용하여 히트(Hit)들이 추가적으로 검증되었다. 정규화된 유리 지방산(NORM FFA) 란(column)은 효소들의 개선을 나타내고, 수치가 작을수록 가장 좋은 개선을 나타낸다. "히트(Hit) ID"는 최초(primary) 스크리닝 플레이트 홈 위치를 나타내고, 여기서 더 적은 NORM FFA 표현형이 발견되었다. 히트 돌연변이들(Hits mutations)은 돌연변이형 carB 유전자-특이적 프라이머들 BZ1 5'-GGATCTCGACGCTCTCCCTT-3' (SEQ ID NO:55) 및 BZ12_ccdAB 특유의(unique)

프라이머 5'-TCAAAAACGCCATTAACCTGATGTTCTG-3' (SEQ ID NO:56) 를 사용하여 pBZ1 플라스미드들을 함유하는 "히트 (Hit)"로부터 증폭된 PCR 생성물들을 시퀀싱함으로써 식별되었다. 확인된 히트(hit)들 내에 식별된 돌연변이들 및 NORM FFA 값들은 표 12에 요약된다.

표 12: 아미노산 포화 돌연변이 유발 동안 식별된 CarB4 효소의 유의한 돌연변이들

WT 아미노산	WT 코돈	히트(Hit) ID(아미노산)	돌연변이형 코돈	NORM FFA(%)
D18	GAT	P10H5(R)	AGG	75.5
		P6B4(L)	CTG	83.6
		P4H11(T)	ACG	80.8
		P8D11(P)	CCG	81.8
S22	AGC	P1F3(R)	AGG	57.7
		P2G9(R)	AGG	55.7
		P2A7(N)	AAC	90
		P8D7(G)	GGG	82.1
L80	CTG	P8H11(R)	AGG	87.4
R87	CGT	P7D7(G)	GGG	85.2
		P5D12(E)	GAG	89.4
D750	GAT	P8F11(A)	GCG	87.6
I870	ATT	P3A12(L)	CTG	76.6

전체 조합 돌연변이 유발(full combinatorial mutagenesis)에 의한 CarB 효소의 신규 변이체들의 식별:

전체 조합 라이브러리는 하기의 아미노산 잔기들: 18D, 18R, 22S, 22R, 473H, 473I, 827R, 827C, 870I, 870L, 926V, 926A, 926E, 927S, 927K, 927G, 930M, 930K, 930R, 1128L 및 1128W를 포함하도록 구성되었다. 모든 위치들에서 본래의 코돈 및 돌연변이형 코돈을 함유하는 프라이머들은 PCR-기반 방법에 의해 라이브러리 구성을 의도하여 만들어졌다(Horton, R.M., Hunt, H.D., Ho, S.N., Pullen, J.K. and Pease, L.R. 1989). CarB2, CarB3 및 CarB4(20R, 288G 및 535S) 내에 보존된 유의한 돌연변이들은 변화되지 않았으므로, pBZ1(변형된 pBZ1_P_{TRC1}_carB2GTG_ccdAB)로 클로닝된 carB2GTG는 PCR 템플릿으로서 사용되었다. 라이브러리 구성은 PCR 분절들을 조립하여 상기 조합 돌연변이(combinatorial mutation)들을 함유하는 CarB ORFs로 만듦으로써 완성되었다. 이후에, 돌연변이형 CarB ORFs는 In-Fusion 방법(Clontech 사)에 의해 pBZ1 골격(backbone)으로 클로닝되었다. In-Fusion 생성물은 침전(precipitated)되었고, 플라스미드 pLYC4를 수반(carrying)하는 스크리닝 균주 내로 직접 전기천공(electroporated)되었다. 라이브러리 스크리닝, 깊은-홈(deep-well) 플레이트 및 진탕 플라스크 발효는 실시예 5에 설명된 바와 같이 실시되었다. 특정한 조합 돌연변이들을 갖는 CarB 돌연변이체들의 활성들 (CarB2에 의해 정규화된 NORM FFA, 100 %)은 표 13에 요약되었다. CarB2, CarB4 및 CarB5(CarB4-S22R)는 대조구로서 포함되었다. NORM FFA 란(column)은 CarB 효소들의 개선을 나타내고, 수치가 작을수록 가장 좋은 개선을 나타낸다. 대조구(CarB2)의 접합 개선(fold improvement)(X-FIOC)도 나타났다. 열거된 모든 돌연변이들은 CarB wt의 폴리펩티드 서열(SEQ ID NO:7)과 관련되어 있다. 예를 들어, CarB1은 A535S 돌연변이를 가지며, CarBDead(촉매 비활성 CarB 효소)는 포스포판테테인 부착 부위를 파괴하는 S693A 돌연변이를 수반한다.

생물반응기 내에서의 개선된 지방족 알코올 생산을 위한 신규 CarB 변이체들:

표 13에 열거된 신규 CarB 변이체들을 식별하는 목적은 개선된 지방족 알코올 생산을 위하여 상기 변이체들을 사용하기 위함이다. 표 13에서 맨 위의 CarB 변이체(P06B6 - S3R, E20R, S22R, F288G, Q473H, A535S, R873S, S927G, M930R, L1128W)는 [야생형 AGC에서 AGA로] 3 위치에 자연 돌연변이(spontaneous mutation)를 수반한다. 두 개의 P06B6 CarB 변이체들 모두, 즉 CarB7(3 위치에 AGA에 의한 아미노산 R - S3R, E20R, S22R, F288G, Q473H, A535S, R873S, S927G, M930R, L1128W) 및 CarB8(3 위치에 AGC에 의한 야생형 아미노산 S - E20R, S22R, F288G, Q473H, A535S, R873S, S927G, M930R, L1128W)가 만들어졌고, CarB에 대해서만 상이한 하기의 지방족 알코올 오페론들을 발생시키기 위하여 적은 복제수(low copy number)의 지방족 알코올 생산 플라스미드 골격 pCL1920으로 클로닝되었다. 모든 CarB 변이체들(CarB2, CarB7 및 CarB8)을 위한 번역 개시 코돈(translation initiation codon)[GTG]은 발현을 최대화하기 위하여 ATG로 되돌렸다(reverted).

pCL1920_P_{TRC}_carB2_tesA_alrAadp1_fabB[A329G]_fadR

pCL1920_P_{TRC}_carB7_tesA_alrAadp1_fabB[A329G]_fadR

pCL1920_P_{TRC}_carB8_tesA_alrAadp1_fabB[A329G]_fadR

[0220]

상기에 설명된 플라스미드들은 염색체의 TE_{dm1} 하나의 복제(copy)를 갖는 V668 기반 균주로 형질전환되었고, 결과적으로 얻어진 균주들은 실시예 4에 설명된 바와 같이 생물반응기 내에서 스크리닝되었다. CarB2에 비하여 CarB7 및 CarB8의 (생물반응기 발효 생성물 내의 % 지방족 알코올로 측정되는) 개선을 도 16에 나타낸다. 활성 순서는 CarB7 > CarB8 > CarB2이다. [AGC에서 AGA R 회귀 코돈(rare codon)으로] CarB7의 3 위치의 돌연변이는 CarB8보다 더 높은 활성을 수여하였고, 또한, 총 가용성 단백질(total soluble proteins)의 SDS-PAGE 분석은 CarB8 및 CarB2보다 CarB7의 더 높은 발현을 나타냈다. CarB2 및 CarB8의 발현 수준은 유사하였다. 이는 실시예 6에 설명된 CarB60 데이터와 일치하고, 3 위치의 AGA R 회귀 코돈 돌연변이 및 이의 N-말단에 CarB60 태그 둘 모두는 CarB 발현을 개선시킬 수 있다. CarB7 및 CarB8이 숙주 균주들을 조작하거나 및/또는 지방족 알코올 생산 오페론의 다른 성분들을 조작함으로써 유리 지방산의 증가에 의해 균주들 내에서 CarB2보다 더 잘 수행할 것이라고 이해된다.

[0221]

표 13: 이중-플라스미드 시스템 내의 조합 라이브러리로부터 식별된 CarB 변이체들의 요약

돌연변이체	NORM FFA (%)	X-FIOC	돌연변이들(Mutations)
P06B6	16.5	6.06	S3R, E20R, S22R, F288G, Q473H, A535S, R873S, S927G, M930R, L1128W
P13A3	23.9	4.18	D18R, E20R, S22R, F288G, Q473I, A535S, S927G, M930K, L1128W
P02A2	26.5	3.77	E20R, S22R, F288G, Q473I, A535S, R827C, V926E, S927K, M930R
P05H3	26.7	3.75	D18R, E20R, S288G, Q473I, A535S, R827C, V926E, M930K, L1128W
P10F10	31.9	3.13	E20R, S22R, F288G, Q473H, A535S, R827C, V926A, S927K, M930R
P01C12	34.2	2.92	E20R, S22R, F288G, Q473H, A535S, R827C
P03B1	36.9	2.71	E20R, S22R, F288G, Q473I, A535S, R827C, M930R
P06E4	36.9	2.71	E20R, S22R, F288G, Q473I, A535S, I870L, S927G, M930R
P14C6	37.4	2.67	E20R, S22R, F288G, Q473I, A535S, I870L, S927G
P05F10	40.4	2.48	D18R, E20R, S22R, F288G, Q473I, A535S, R827C, I870L, V926A, S927G
P06C8	40.8	2.45	E20R, S22R, F288G, Q473H, A535S, R827C, I870L, L1128W
P15E4	40.8	2.45	D18R, E20R, S22R, F288G, Q473H, A535S, R827C, I870L, S927G, L1128W
P05H7	40.9	2.44	E20R, S22R, F288G, Q473I, A535S, R827C, I870L, S927G, L1128W
P15A6	41	2.44	E20R, S22R, F288G, Q473I, A535S, R827C, I870L, S927G, M930K, L1128W
P08F5	41.2	2.43	E20R, S22R, F288G, Q473H, A535S, I870L, S927G, M930K
P14C7	41.3	2.42	E20R, F288G, Q473I, A535S, I870L, M930K
P16H10	42.1	2.38	E20R, S22R, F288G, Q473H, A535S, S927G, M930K, L1128W
P16A1	44.1	2.27	D18R, E20R, S22R, F288G, Q473I, A535S, S927G, L1128W
P14H4	44.2	2.26	E20R, S22R, F288G, Q473I, A535S, R827C, I870L, S927G
P15C1	46.5	2.15	D18R, E20R, S22R, F288G, Q473I, A535S, R827C, I870L, S927G, L1128W
P16E5	47.2	2.12	D18R, E20R, S22R, F288G, Q473I, A535S, S927G, M930R, L1128W
P15A3	47.2	2.12	E20R, S22R, F288G, Q473H, A535S, V926E, S927G, M930R
P05A2	52.4	1.91	E20R, S22R, F288G, Q473H, A535S, R827C, I870L, V926A, L1128W
CarB2	100	1	E20R, F288G, Q473I, A535S
CarB4	77.8	1.29	E20R, F288G, Q473H, A535S, R827A, S927G
CarB5	48.9	2.04	E20R, S22R, F288G, Q473H, A535S, R827A, S927G
CarB1	ND		A535S
CarB wt	ND		SEQ ID NO:7
CarB Dead	ND		S693A

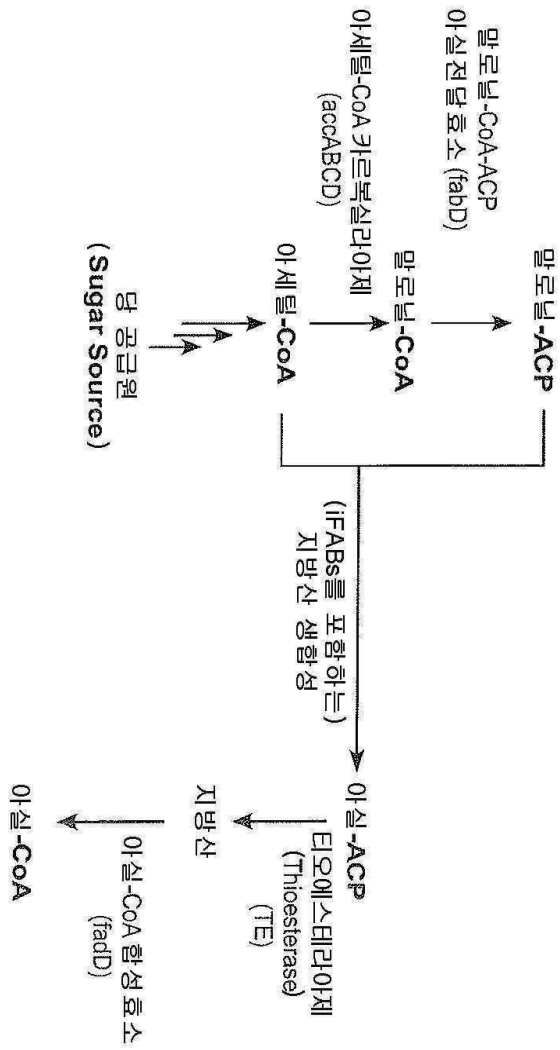
[0222]

[0223]

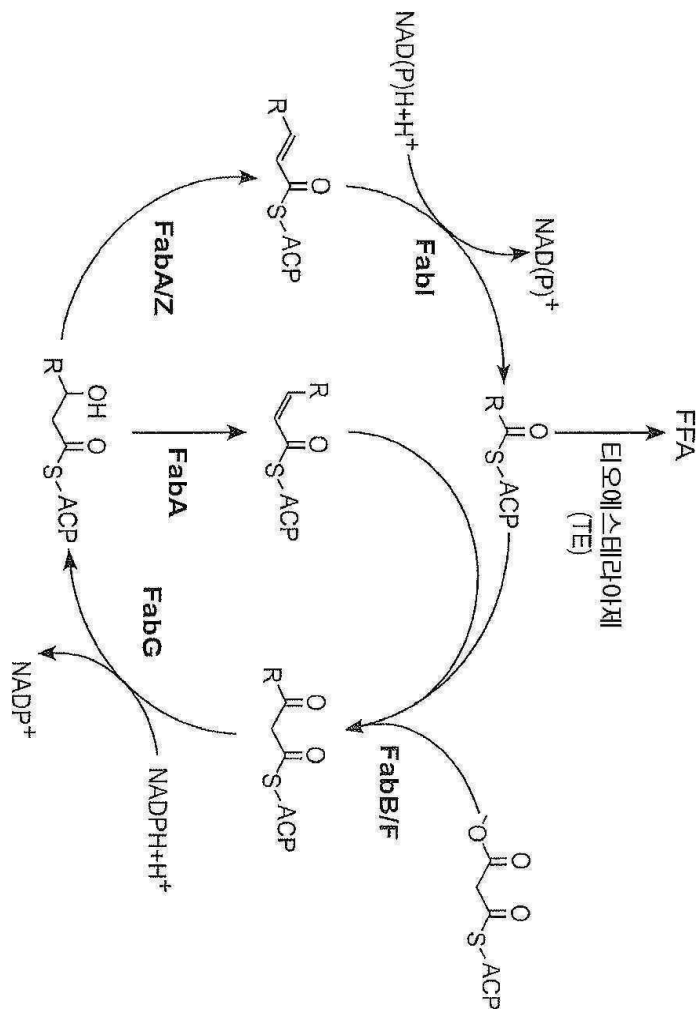
본 명세서에 설명된 모든 방법들은 본 명세서에 달리 지시되거나 문맥상 명백하게 부정되지 않는다면 여하한 적합한 순서로 실시될 수 있다. 본 명세서에 제어되는 여하한 및 모든 예시들 또는 예시어(exemplary language) [예를 들어, "이를 테면"]의 사용은 그저 본 발명을 더 잘 설명하기 위함이며, 달리 주장하지 않는다면 본 발명의 범위를 제한하려는 것이 아니다. 본 발명의 실시예에 있어서 필수적인 것이나 주장되지 않은 여하한 요소를 나타내는 것으로 해석되어야 하는 본 명세서의 표현(language)은 없다. 본 명세서에서 사용되는 전문 용어는 특정 실시예들을 설명하려는 목적만을 위한 것이며, 제한하려고 의도한 것이 아니라고 이해한다. 본 발명의 바람직한 실시예들은 본 명세서에 설명된다. 이러한 바람직한 실시예들의 변형들은 상기의 설명을 읽으므로써 본 기술분야의 당업자들에게는 명백할 수 있다. 발명자들은 당업자들이 적절하게 이러한 변형들을 적용시키는 것을 예상하며, 발명자들은 본 발명이 본 명세서에 상세하게 설명된 바와 다르게 실시되도록 의도한다. 따라서, 본 발명은 적용할 수 있는 법에 의해 허용되는 만큼 본 명세서에 첨부된 청구항들에 기술된 주제의 모든 변형들 및 등가물(equivalent)들을 포함한다. 또한, 본 명세서에서 달리 지시되거나 문맥상 명백하게 부정되지 않는 한, 모든 가능한 변화에 있어서 상기에 설명된 요소들의 임의의 조합이 본 발명에 포괄된다.

도면

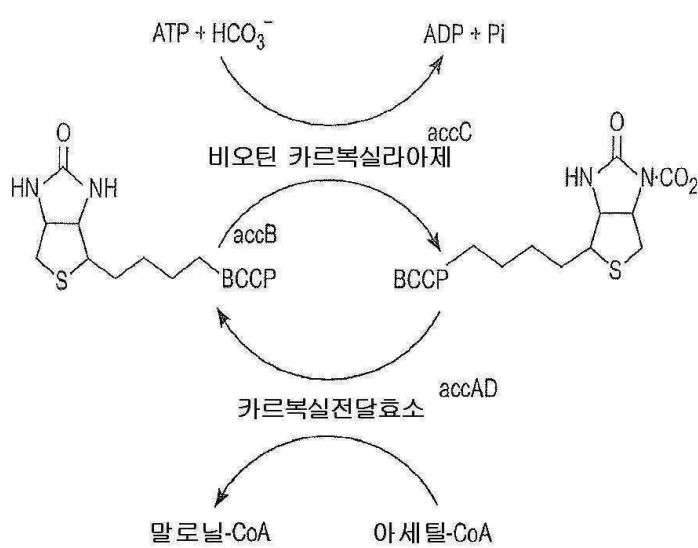
도면1



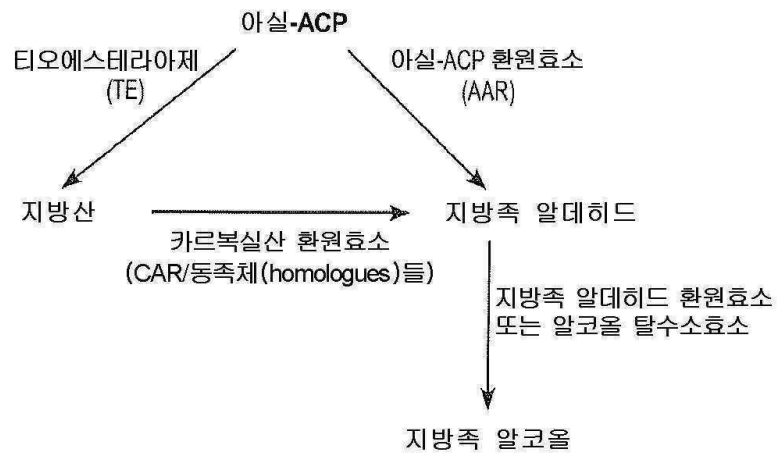
도면2



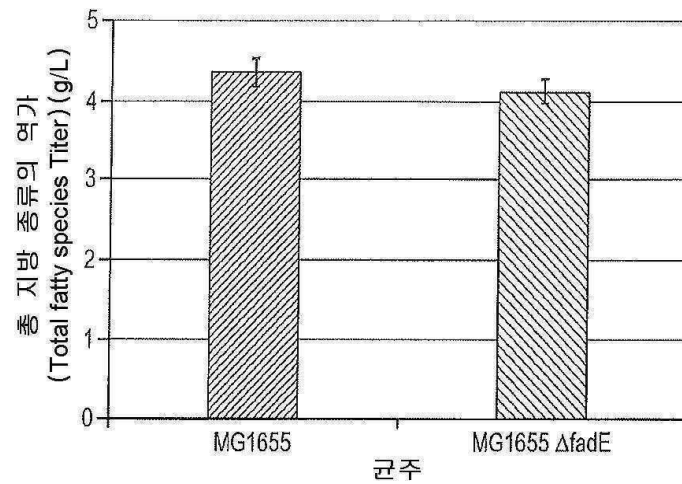
도면3



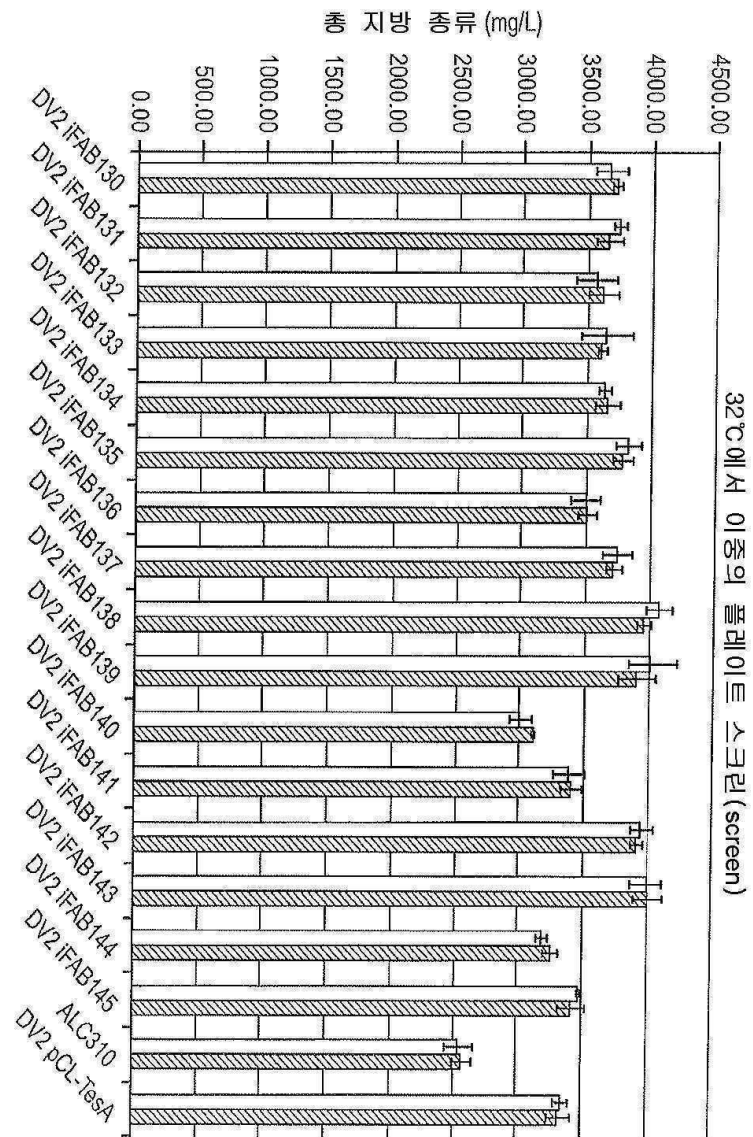
도면4



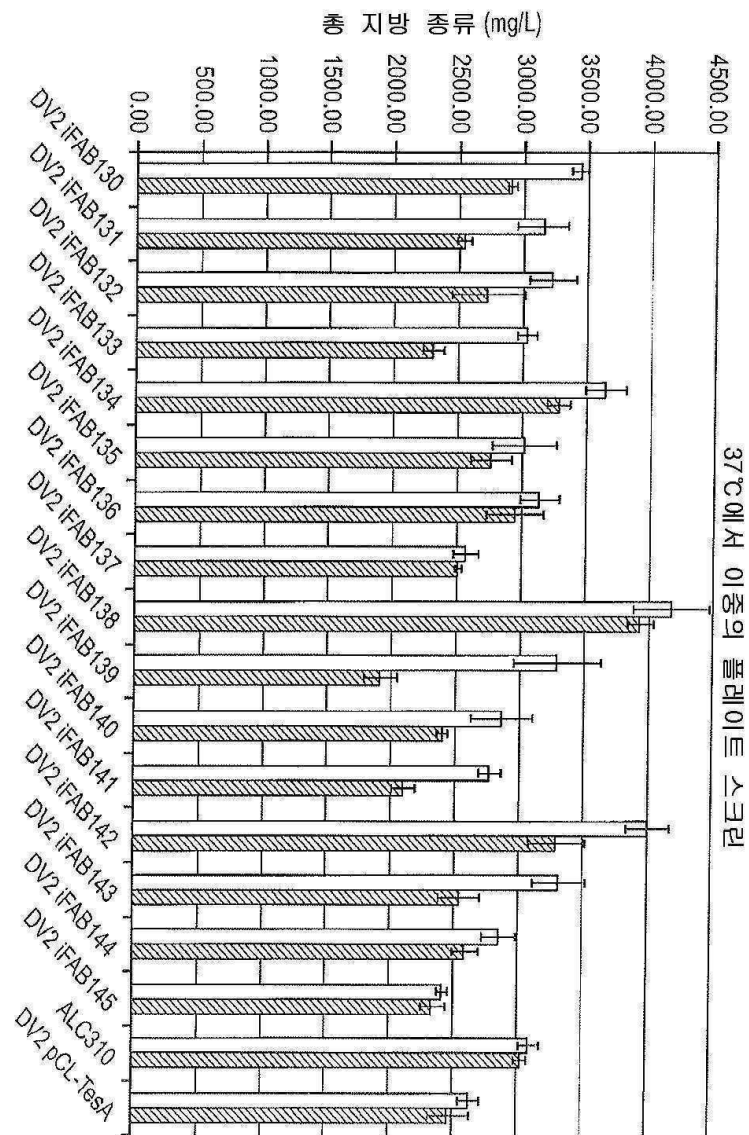
도면5



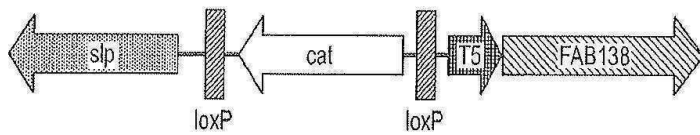
도면6a



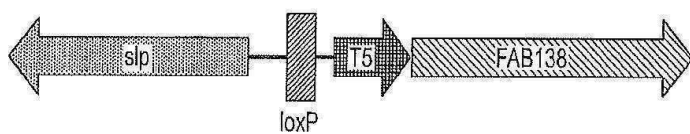
도면6b



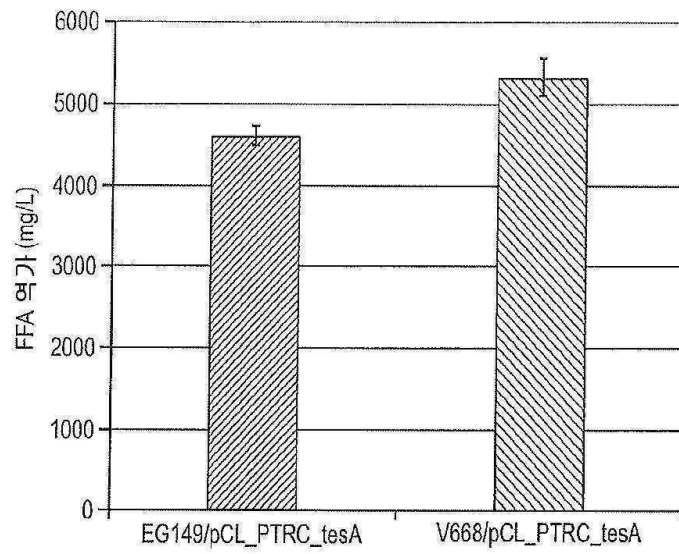
도면7a



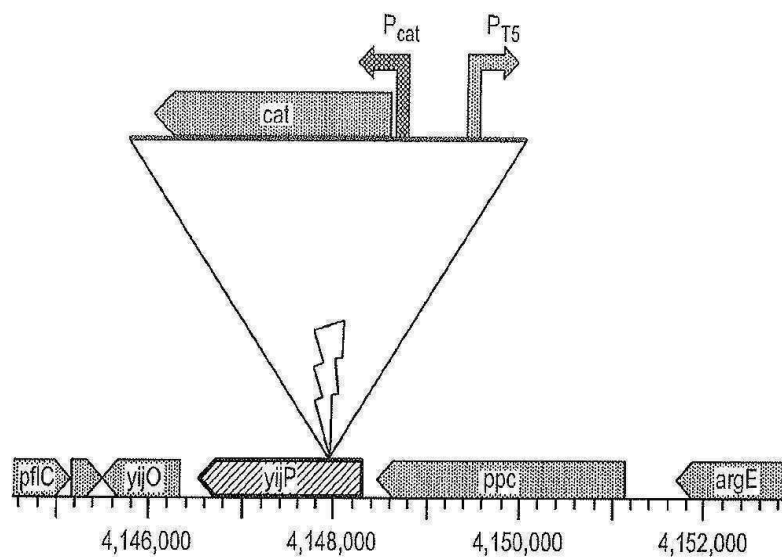
도면7b



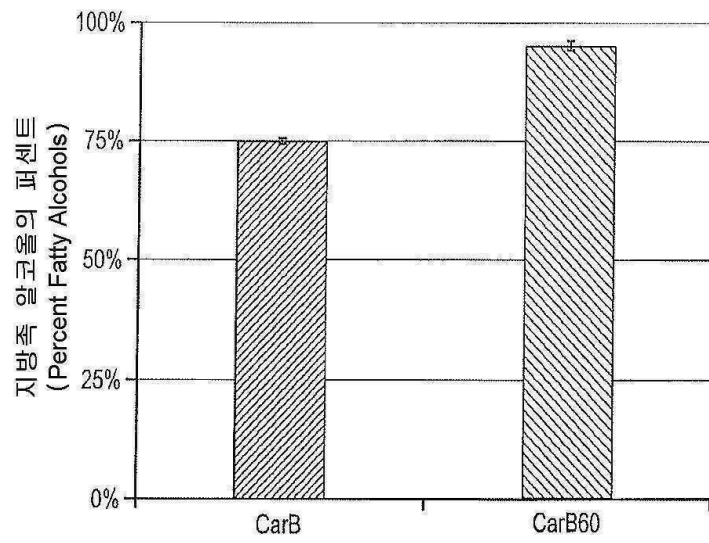
도면8



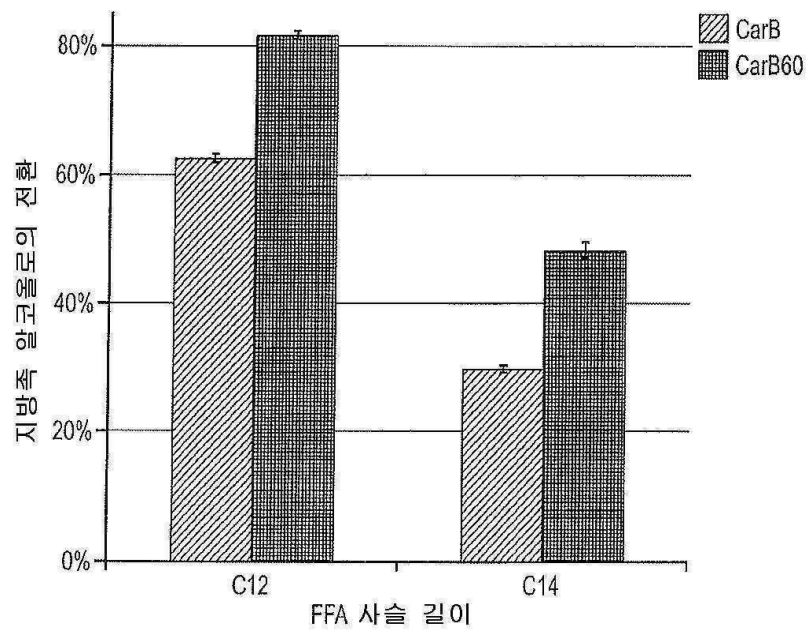
도면9



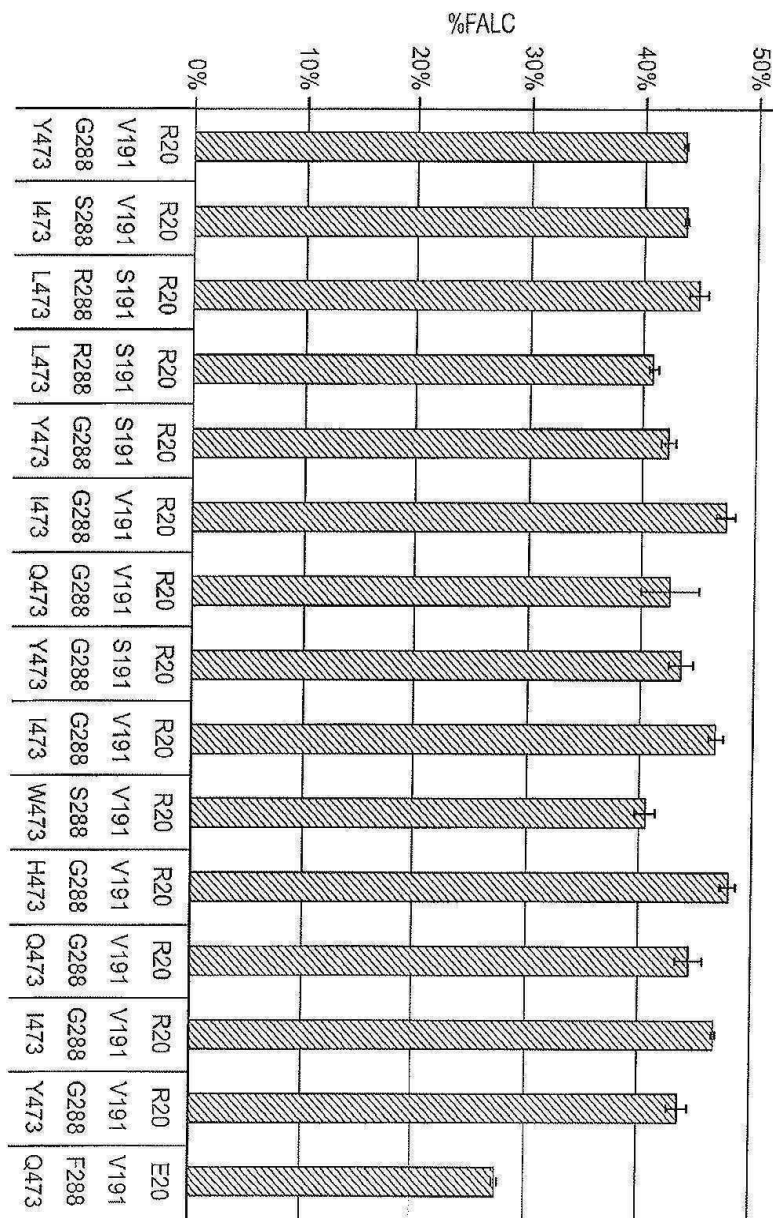
도면10



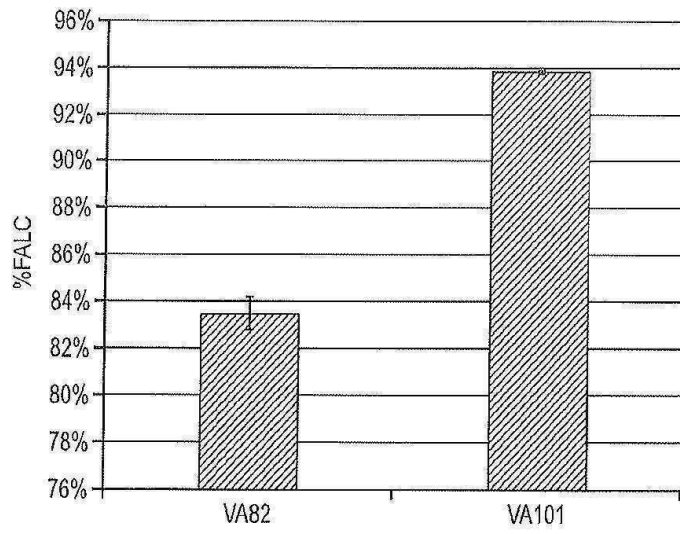
도면11



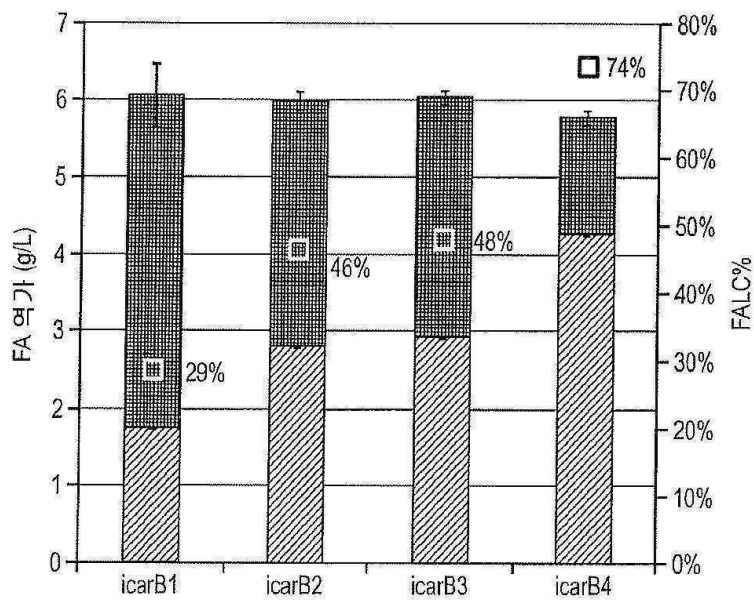
도면12



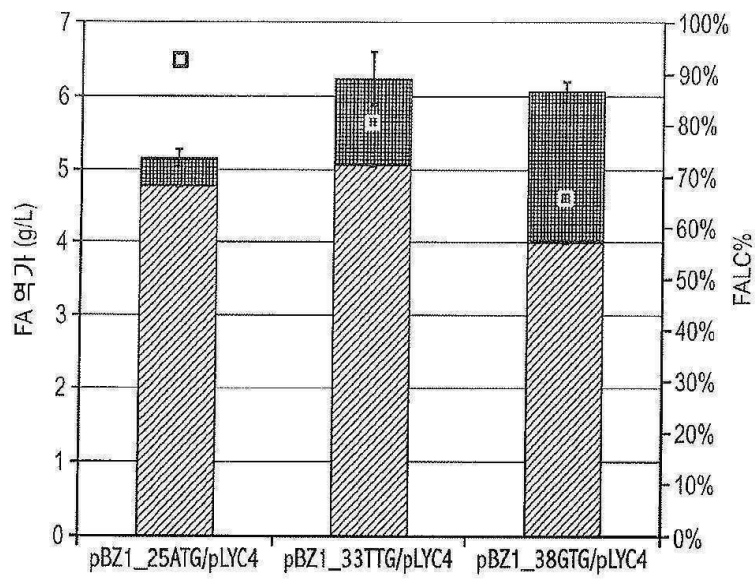
도면13



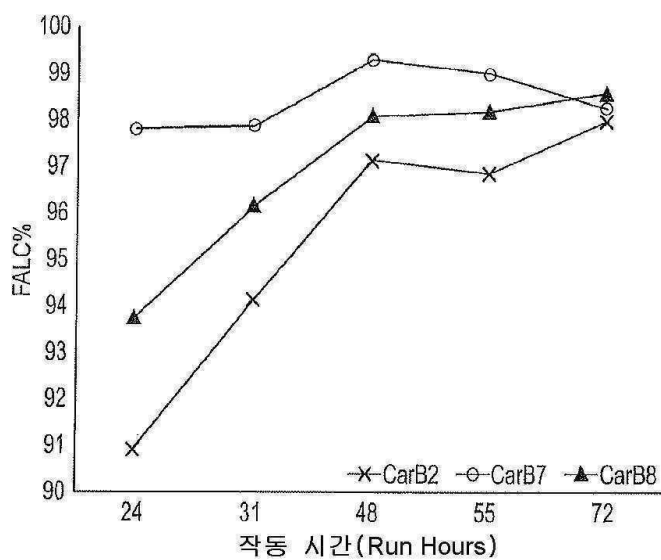
도면14



도면15



도면16



서열 목록

SEQUENCE LISTING

<110> GENOMATICA, INC.

<120> IMPROVED PRODUCTION OF FATTY ALCOHOLS

<130> FPA11749-D05/C

<150> 61/619,309

<151> 2012-04-02

<160> 56

<170> PatentIn version 3.5

<210> 1

<211> 1232

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 1

ttgtccatct ttatataatt tgggggtagg gtgttcttta tgtaaaaaaa acgttttagg	60
atgcatatgg cggccgcata acttcgtata gcatacatta tacgaagtta tctagagttg	120
catgcctgca ggtccgctta ttatcactta ttcaggcgta gcaaccaggc gtttaagggc	180
accaataact gccttaaaaa aattacgccc cgccttgcca ctcatcgag tactgttgta	240
attcattaag cattctgccg acatggaagc catcacaac ggcatgatga acctgaatcg	300
ccagcggcat cagcaccttg tcgcttgccg tataatatat gcccatggcg aaaacggggg	360
cgaagaagtt gtccatatgg gccacgttta aatcaaaact ggtgaaactc acccagggat	420
tggctgagac gaaaaacata ttctcaataa accctttagg gaaataggcc aggttttcac	480
cgtaacacgc cacatcttgc gaatatatgt gtagaaactg ccggaatcg tcgtggtatt	540
cactccagag cgatgaaaac gtttcagttt gtcctatggaa aacggtgtaa caagggtgaa	600
cactatccca tatcaccage tcacctctt tcattgccat acggaattcc ggatgagcat	660
tcctcaggcg ggcaagaatg tgaataaagg ccgataaaa cttgtgctta ttttcttta	720
cggctcttaa aaaggccgta atatccagct gaacggctcg gttataggtta cattgagcaa	780
ctgactgaaa tgctcaaaa tgttctttac gatgccattg ggatatatca acggtggtat	840
atccagtgat tttttctcc attttagctt ccttagctcc tgaaaatctc gataactcaa	900
aaaatacgcc cggtagtgat cttatttcat tatggtgaaa gttggaacct cttacgtgcc	960
gatcaacgtc tcattttcgc caaaagttgg ccaggggctt cccggtatca acagggacac	1020
caggatttat ttattctgcg aagtgatctt ccgtcacagg tatttattcg actctagata	1080
acttcgtata gcatacatta tacgaagtta tggatccagc ttatcgatac cgtcaaacia	1140
atcataaaaa atttatttgc tttcaggaaa atttttctgt ataataagatt caattgcgat	1200
gacgacgaac acgcacctgc aggaggagac ca	1232

<210> 2

<211> 232

<212> DNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 2

ttgtccatct ttatataatt tgggggtagg gtgttcttta tgtaaaaaaa acgttttagg	60
atgcatatgg cggccgcata acttcgtata gcatacatta tacgaagtta tggatccagc	120
ttatcgatac cgtcaaacaa atcataaaaa atttatttgc tticaggaaa atttttctgt	180
ataatagatt caattgcgat gacgacgaac acgcacctgc aggaggagac ca	232

<210> 3

<211> 340

<212> PRT

<213> Acinetobacter sp.

<400> 3

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> 4
 <211> 314
 <212> PRT
 <213> Acinetobacter sp.

<400> 4

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> 5

<211> 1020

<212> DNA

<213> Escherichia coli

<400> 5

atgtcgtatga taaaaagcta tgccgcaaaa gaagcgggcg gcgaactgga agtttatgag	60
tacgatcccc gtgagctgag gccacaagat gttgaagtgc aggtggatta ctgcgggatac	120
tgccattccg atctgtcgtat gatcgataac gaatggggat tttcacaata tccgctggtt	180
gccgggcatg aggtgattgg gcgcgtggtg gcaactcggga gcgccgcgca ggataaaggt	240
ttgcaggctcg gtcagcgtgt cggtgattggc tggacggcgc gtagctgtgg tcaactgcgac	300
gcctgtatta gcggtaatca gatcaactgc gagcaagggt cggtgccgac gattatgaat	360
cgcggtggct ttgccgagaa gttgcgtgcg gactggcaat ggggtgattcc actgccagaa	420
aataattgata tcgagtcgcg cgggcccgtg tttgcggcg gtatcacggt ctttaacca	480
ctgttgatgc accatatcac tgctaccagc cgcgttgggg taattggtat tggcgggctg	540
gggcatatcg ctataaaact tctgcacgca atgggatgcg aggtgacagc ctttagttct	600
aatccggcga aagagcagga agtgctggcg atgggtgccg ataaagtgt gaatagccgc	660
gatccgagg cactgaaagc actggcgggg cagtttgatc tcattatcaa caccgtcaac	720
gtcagcctcg actggcagcc ctattttgag gcgctgacct atggcggtaa ttccatacg	780
gtcgggtgcgg ttctcacgcc gctgtctgtt ccggccttta cgtaattgc gggcgatcgc	840
agcgtctctg gttctgtac cggcacgcct tatgagctgc gtaagctgat gcgttttgcc	900
gccccgacga aggttgccg gaccaccgaa ctgttcccga tgtcgaaaat taacgacgcc	960

atccagcatg tgcgcgacgg taaggcgcgt taccgcgtgg tgttgaaagc cgatttttga 1020

<210> 6

<211> 1174

<212> PRT

<213> Artificial Sequence

<220><221> source

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

<400> 6

Met Ala Val Asp Ser Pro Asp Glu Arg Leu Gln Arg Arg Ile Ala Gln

1 5 10 15
Leu Phe Ala Glu Asp Glu Gln Val Lys Ala Ala Arg Pro Leu Glu Ala

20 25 30
Val Ser Ala Ala Val Ser Ala Pro Gly Met Arg Leu Ala Gln Ile Ala

35 40 45
Ala Thr Val Met Ala Gly Tyr Ala Asp Arg Pro Ala Ala Gly Gln Arg

50 55 60
Ala Phe Glu Leu Asn Thr Asp Asp Ala Thr Gly Arg Thr Ser Leu Arg

65 70 75 80
Leu Leu Pro Arg Phe Glu Thr Ile Thr Tyr Arg Glu Leu Trp Gln Arg

85 90 95
Val Gly Glu Val Ala Ala Ala Trp His His Asp Pro Glu Asn Pro Leu

100 105 110
Arg Ala Gly Asp Phe Val Ala Leu Leu Gly Phe Thr Ser Ile Asp Tyr

115 120 125
Ala Thr Leu Asp Leu Ala Asp Ile His Leu Gly Ala Val Thr Val Pro

130 135 140
Leu Gln Ala Ser Ala Ala Val Ser Gln Leu Ile Ala Ile Leu Thr Glu

145 150 155 160
Thr Ser Pro Arg Leu Leu Ala Ser Thr Pro Glu His Leu Asp Ala Ala

165 170 175
Val Glu Cys Leu Leu Ala Gly Thr Thr Pro Glu Arg Leu Val Val Phe

180 185 190
 Asp Tyr His Pro Glu Asp Asp Asp Gln Arg Ala Ala Phe Glu Ser Ala

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

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

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

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

Ile Gln Arg Pro Pro Val Leu Asp Tyr Lys Leu Val Asp Val Pro Glu
435 440 445
Leu Gly Tyr Phe Arg Thr Asp Arg Pro His Pro Arg Gly Glu Leu Leu
450 455 460
Leu Lys Ala Glu Thr Thr Ile Pro Gly Tyr Tyr Lys Arg Pro Glu Val
465 470 475 480
Thr Ala Glu Ile Phe Asp Glu Asp Gly Phe Tyr Lys Thr Gly Asp Ile
485 490 495
Val Ala Glu Leu Glu His Asp Arg Leu Val Tyr Val Asp Arg Arg Asn
500 505 510
Asn Val Leu Lys Leu Ser Gln Gly Glu Phe Val Thr Val Ala His Leu
515 520 525
Glu Ala Val Phe Ala Ser Ser Pro Leu Ile Arg Gln Ile Phe Ile Tyr
530 535 540
Gly Ser Ser Glu Arg Ser Tyr Leu Leu Ala Val Ile Val Pro Thr Asp
545 550 555 560
Asp Ala Leu Arg Gly Arg Asp Thr Ala Thr Leu Lys Ser Ala Leu Ala
565 570 575
Glu Ser Ile Gln Arg Ile Ala Lys Asp Ala Asn Leu Gln Pro Tyr Glu
580 585 590
Ile Pro Arg Asp Phe Leu Ile Glu Thr Glu Pro Phe Thr Ile Ala Asn
595 600 605
Gly Leu Leu Ser Gly Ile Ala Lys Leu Leu Arg Pro Asn Leu Lys Glu
610 615 620
Arg Tyr Gly Ala Gln Leu Glu Gln Met Tyr Thr Asp Leu Ala Thr Gly
625 630 635 640
Gln Ala Asp Glu Leu Leu Ala Leu Arg Arg Glu Ala Ala Asp Leu Pro
645 650 655
Val Leu Glu Thr Val Ser Arg Ala Ala Lys Ala Met Leu Gly Val Ala
660 665 670
Ser Ala Asp Met Arg Pro Asp Ala His Phe Thr Asp Leu Gly Gly Asp

675 680 685
 Ser Leu Ser Ala Leu Ser Phe Ser Asn Leu Leu His Glu Ile Phe Gly
 690 695 700
 Val Glu Val Pro Val Gly Val Val Val Ser Pro Ala Asn Glu Leu Arg

 705 710 715 720
 Asp Leu Ala Asn Tyr Ile Glu Ala Glu Arg Asn Ser Gly Ala Lys Arg
 725 730 735
 Pro Thr Phe Thr Ser Val His Gly Gly Gly Ser Glu Ile Arg Ala Ala
 740 745 750
 Asp Leu Thr Leu Asp Lys Phe Ile Asp Ala Arg Thr Leu Ala Ala Ala
 755 760 765
 Asp Ser Ile Pro His Ala Pro Val Pro Ala Gln Thr Val Leu Leu Thr

 770 775 780
 Gly Ala Asn Gly Tyr Leu Gly Arg Phe Leu Cys Leu Glu Trp Leu Glu
 785 790 795 800
 Arg Leu Asp Lys Thr Gly Gly Thr Leu Ile Cys Val Val Arg Gly Ser
 805 810 815
 Asp Ala Ala Ala Ala Arg Lys Arg Leu Asp Ser Ala Phe Asp Ser Gly
 820 825 830
 Asp Pro Gly Leu Leu Glu His Tyr Gln Gln Leu Ala Ala Arg Thr Leu

 835 840 845
 Glu Val Leu Ala Gly Asp Ile Gly Asp Pro Asn Leu Gly Leu Asp Asp
 850 855 860
 Ala Thr Trp Gln Arg Leu Ala Glu Thr Val Asp Leu Ile Val His Pro
 865 870 875 880
 Ala Ala Leu Val Asn His Val Leu Pro Tyr Thr Gln Leu Phe Gly Pro
 885 890 895
 Asn Val Val Gly Thr Ala Glu Ile Val Arg Leu Ala Ile Thr Ala Arg

 900 905 910
 Arg Lys Pro Val Thr Tyr Leu Ser Thr Val Gly Val Ala Asp Gln Val
 915 920 925

Asp Pro Ala Glu Tyr Gln Glu Asp Ser Asp Val Arg Glu Met Ser Ala
 930 935 940
 Val Arg Val Val Arg Glu Ser Tyr Ala Asn Gly Tyr Gly Asn Ser Lys
 945 950 955 960
 Trp Ala Gly Glu Val Leu Leu Arg Glu Ala His Asp Leu Cys Gly Leu
 965 970 975
 Pro Val Ala Val Phe Arg Ser Asp Met Ile Leu Ala His Ser Arg Tyr
 980 985 990
 Ala Gly Gln Leu Asn Val Gln Asp Val Phe Thr Arg Leu Ile Leu Ser
 995 1000 1005
 Leu Val Ala Thr Gly Ile Ala Pro Tyr Ser Phe Tyr Arg Thr Asp
 1010 1015 1020
 Ala Asp Gly Asn Arg Gln Arg Ala His Tyr Asp Gly Leu Pro Ala
 1025 1030 1035
 Asp Phe Thr Ala Ala Ala Ile Thr Ala Leu Gly Ile Gln Ala Thr
 1040 1045 1050
 Glu Gly Phe Arg Thr Tyr Asp Val Leu Asn Pro Tyr Asp Asp Gly
 1055 1060 1065
 Ile Ser Leu Asp Glu Phe Val Asp Trp Leu Val Glu Ser Gly His
 1070 1075 1080
 Pro Ile Gln Arg Ile Thr Asp Tyr Ser Asp Trp Phe His Arg Phe
 1085 1090 1095
 Glu Thr Ala Ile Arg Ala Leu Pro Glu Lys Gln Arg Gln Ala Ser
 1100 1105 1110
 Val Leu Pro Leu Leu Asp Ala Tyr Arg Asn Pro Cys Pro Ala Val
 1115 1120 1125
 Arg Gly Ala Ile Leu Pro Ala Lys Glu Phe Gln Ala Ala Val Gln
 1130 1135 1140
 Thr Ala Lys Ile Gly Pro Glu Gln Asp Ile Pro His Leu Ser Ala
 1145 1150 1155
 Pro Leu Ile Asp Lys Tyr Val Ser Asp Leu Glu Leu Leu Gln Leu

1160 1165 1170

Leu

<210> 7

<211> 1173

<212> PRT

<213> *Mycobacterium smegmatis*

<400> 7

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

1 5 10 15

Leu Asp Asp Glu Gln Ser Thr Arg Arg Ile Ala Glu Leu Tyr Ala Thr

20 25 30

Asp Pro Glu Phe Ala Ala Ala Ala Pro Leu Pro Ala Val Val Asp Ala

35 40 45

Ala His Lys Pro Gly Leu Arg Leu Ala Glu Ile Leu Gln Thr Leu Phe

50 55 60

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

65 70 75 80

Ala Thr Asp Glu Gly Gly Arg Thr Val Thr Arg Leu Leu Pro Arg Phe

85 90 95

Asp Thr Leu Thr Tyr Ala Gln Val Trp Ser Arg Val Gln Ala Val Ala

100 105 110

Ala Ala Leu Arg His Asn Phe Ala Gln Pro Ile Tyr Pro Gly Asp Ala

115 120 125

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

130 135 140

Val Cys Ala Tyr Leu Gly Leu Val Ser Val Pro Leu Gln His Asn Ala

145 150 155 160

Pro Val Ser Arg Leu Ala Pro Ile Leu Ala Glu Val Glu Pro Arg Ile

165 170 175

Leu Thr Val Ser Ala Glu Tyr Leu Asp Leu Ala Val Glu Ser Val Arg

180 185 190

Asp Val Asn Ser Val Ser Gln Leu Val Val Phe Asp His His Pro Glu
 195 200 205
 Val Asp Asp His Arg Asp Ala Leu Ala Arg Ala Arg Glu Gln Leu Ala
 210 215 220

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

 Ile Thr Gly Asp Pro Thr Pro Val Ile Asn Val Asn Phe Met Pro Leu
 290 295 300
 Asn His Leu Gly Gly Arg Ile Pro Ile Ser Thr Ala Val Gln Asn Gly
 305 310 315 320
 Gly Thr Ser Tyr Phe Val Pro Glu Ser Asp Met Ser Thr Leu Phe Glu
 325 330 335
 Asp Leu Ala Leu Val Arg Pro Thr Glu Leu Gly Leu Val Pro Arg Val
 340 345 350

 Ala Asp Met Leu Tyr Gln His His Leu Ala Thr Val Asp Arg Leu Val
 355 360 365
 Thr Gln Gly Ala Asp Glu Leu Thr Ala Glu Lys Gln Ala Gly Ala Glu
 370 375 380
 Leu Arg Glu Gln Val Leu Gly Gly Arg Val Ile Thr Gly Phe Val Ser
 385 390 395 400
 Thr Ala Pro Leu Ala Ala Glu Met Arg Ala Phe Leu Asp Ile Thr Leu
 405 410 415

 Gly Ala His Ile Val Asp Gly Tyr Gly Leu Thr Glu Thr Gly Ala Val
 420 425 430
 Thr Arg Asp Gly Val Ile Val Arg Pro Pro Val Ile Asp Tyr Lys Leu

435 440 445
 Ile Asp Val Pro Glu Leu Gly Tyr Phe Ser Thr Asp Lys Pro Tyr Pro
 450 455 460
 Arg Gly Glu Leu Leu Val Arg Ser Gln Thr Leu Thr Pro Gly Tyr Tyr
 465 470 475 480

 Lys Arg Pro Glu Val Thr Ala Ser Val Phe Asp Arg Asp Gly Tyr Tyr
 485 490 495
 His Thr Gly Asp Val Met Ala Glu Thr Ala Pro Asp His Leu Val Tyr
 500 505 510
 Val Asp Arg Arg Asn Asn Val Leu Lys Leu Ala Gln Gly Glu Phe Val
 515 520 525
 Ala Val Ala Asn Leu Glu Ala Val Phe Ser Gly Ala Ala Leu Val Arg
 530 535 540

 Gln Ile Phe Val Tyr Gly Asn Ser Glu Arg Ser Phe Leu Leu Ala Val
 545 550 555 560
 Val Val Pro Thr Pro Glu Ala Leu Glu Gln Tyr Asp Pro Ala Ala Leu
 565 570 575
 Lys Ala Ala Leu Ala Asp Ser Leu Gln Arg Thr Ala Arg Asp Ala Glu
 580 585 590
 Leu Gln Ser Tyr Glu Val Pro Ala Asp Phe Ile Val Glu Thr Glu Pro
 595 600 605

 Phe Ser Ala Ala Asn Gly Leu Leu Ser Gly Val Gly Lys Leu Leu Arg
 610 615 620
 Pro Asn Leu Lys Asp Arg Tyr Gly Gln Arg Leu Glu Gln Met Tyr Ala
 625 630 635 640
 Asp Ile Ala Ala Thr Gln Ala Asn Gln Leu Arg Glu Leu Arg Arg Ala
 645 650 655
 Ala Ala Thr Gln Pro Val Ile Asp Thr Leu Thr Gln Ala Ala Ala Thr
 660 665 670

 Ile Leu Gly Thr Gly Ser Glu Val Ala Ser Asp Ala His Phe Thr Asp
 675 680 685

Leu Gly Gly Asp Ser Leu Ser Ala Leu Thr Leu Ser Asn Leu Leu Ser
 690 695 700
 Asp Phe Phe Gly Phe Glu Val Pro Val Gly Thr Ile Val Asn Pro Ala
 705 710 715 720
 Thr Asn Leu Ala Gln Leu Ala Gln His Ile Glu Ala Gln Arg Thr Ala
 725 730 735

 Gly Asp Arg Arg Pro Ser Phe Thr Thr Val His Gly Ala Asp Ala Thr
 740 745 750
 Glu Ile Arg Ala Ser Glu Leu Thr Leu Asp Lys Phe Ile Asp Ala Glu
 755 760 765
 Thr Leu Arg Ala Ala Pro Gly Leu Pro Lys Val Thr Thr Glu Pro Arg
 770 775 780
 Thr Val Leu Leu Ser Gly Ala Asn Gly Trp Leu Gly Arg Phe Leu Thr
 785 790 795 800

 Leu Gln Trp Leu Glu Arg Leu Ala Pro Val Gly Gly Thr Leu Ile Thr
 805 810 815
 Ile Val Arg Gly Arg Asp Asp Ala Ala Ala Arg Ala Arg Leu Thr Gln
 820 825 830
 Ala Tyr Asp Thr Asp Pro Glu Leu Ser Arg Arg Phe Ala Glu Leu Ala
 835 840 845
 Asp Arg His Leu Arg Val Val Ala Gly Asp Ile Gly Asp Pro Asn Leu
 850 855 860

 Gly Leu Thr Pro Glu Ile Trp His Arg Leu Ala Ala Glu Val Asp Leu
 865 870 875 880
 Val Val His Pro Ala Ala Leu Val Asn His Val Leu Pro Tyr Arg Gln
 885 890 895
 Leu Phe Gly Pro Asn Val Val Gly Thr Ala Glu Val Ile Lys Leu Ala
 900 905 910
 Leu Thr Glu Arg Ile Lys Pro Val Thr Tyr Leu Ser Thr Val Ser Val
 915 920 925

 Ala Met Gly Ile Pro Asp Phe Glu Glu Asp Gly Asp Ile Arg Thr Val

930	935	940	
Ser Pro Val Arg Pro Leu Asp Gly Gly Tyr Ala Asn Gly Tyr Gly Asn			
945	950	955	960
Ser Lys Trp Ala Gly Glu Val Leu Leu Arg Glu Ala His Asp Leu Cys			
	965	970	975
Gly Leu Pro Val Ala Thr Phe Arg Ser Asp Met Ile Leu Ala His Pro			
	980	985	990
Arg Tyr Arg Gly Gln Val Asn Val Pro Asp Met Phe Thr Arg Leu Leu			
	995	1000	1005
Leu Ser Leu Leu Ile Thr Gly Val Ala Pro Arg Ser Phe Tyr Ile			
	1010	1015	1020
Gly Asp Gly Glu Arg Pro Arg Ala His Tyr Pro Gly Leu Thr Val			
	1025	1030	1035
Asp Phe Val Ala Glu Ala Val Thr Thr Leu Gly Ala Gln Gln Arg			
	1040	1045	1050
Glu Gly Tyr Val Ser Tyr Asp Val Met Asn Pro His Asp Asp Gly			
	1055	1060	1065
Ile Ser Leu Asp Val Phe Val Asp Trp Leu Ile Arg Ala Gly His			
	1070	1075	1080
Pro Ile Asp Arg Val Asp Asp Tyr Asp Asp Trp Val Arg Arg Phe			
	1085	1090	1095
Glu Thr Ala Leu Thr Ala Leu Pro Glu Lys Arg Arg Ala Gln Thr			
	1100	1105	1110
Val Leu Pro Leu Leu His Ala Phe Arg Ala Pro Gln Ala Pro Leu			
	1115	1120	1125
Arg Gly Ala Pro Glu Pro Thr Glu Val Phe His Ala Ala Val Arg			
	1130	1135	1140
Thr Ala Lys Val Gly Pro Gly Asp Ile Pro His Leu Asp Glu Ala			
	1145	1150	1155
Leu Ile Asp Lys Tyr Ile Arg Asp Leu Arg Glu Phe Gly Leu Ile			
	1160	1165	1170

<210> 8

<211> 209

<212> PRT

<213> Escherichia coli

<400> 8

Met Val Asp Met Lys Thr Thr His Thr Ser Leu Pro Phe Ala Gly His

1 5 10 15

Thr Leu His Phe Val Glu Phe Asp Pro Ala Asn Phe Cys Glu Gln Asp

20 25 30

Leu Leu Trp Leu Pro His Tyr Ala Gln Leu Gln His Ala Gly Arg Lys

35 40 45

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

50 55 60

Arg Glu Tyr Gly Tyr Lys Cys Val Pro Ala Ile Gly Glu Leu Arg Gln

65 70 75 80

Pro Val Trp Pro Ala Glu Val Tyr Gly Ser Ile Ser His Cys Gly Thr

85 90 95

Thr Ala Leu Ala Val Val Ser Arg Gln Pro Ile Gly Ile Asp Ile Glu

100 105 110

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

115 120 125

Thr Pro Ala Glu His Glu Arg Leu Ala Asp Cys Gly Leu Ala Phe Ser

130 135 140

Leu Ala Leu Thr Leu Ala Phe Ser Ala Lys Glu Ser Ala Phe Lys Ala

145 150 155 160

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

165 170 175

Trp Asn Lys Gln Gln Val Ile Ile His Arg Glu Asn Glu Met Phe Ala

180 185 190

Val His Trp Gln Ile Lys Glu Lys Ile Val Ile Thr Leu Cys Gln His

195 200 205

Asp

<210> 9
 <211> 70
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 9
 aaaaacagca acaatgtgag ctttgttgta attatatgt aaacatattg attccgggga 60
 tccgtcgacc 70

<210> 10
 <211> 68
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 10
 aaacggagcc ttteggctec gttattcatt tacgcggctt caactttcct gtaggctgga 60
 gctgcttc 68

<210> 11
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 11
 cgggcaggtg ctatgaccag gac 23

<210> 12
 <211> 23
 <212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 12

cgcggcgttg accggcagcc tgg 23

<210> 13

<211> 5659

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 13

atgatcatca aacctaaaat tcgtggattt atctgtacaa caacgcaccc agtgggttgt 60

gaagcgaacg taaaagaaca aattgcctac acaaaagcac aaggtccgat caaaaacgca 120

cctaagcgcg tgttggttgt cggatcgtct agcggctatg gtctgtcatc acgcatcgct 180

gcggcggttg gcggtgggtgc ggcgacgatc ggcgatattt tcgaaaagcc gggcactgac 240

aaaaaaccag gtactgcggg tttctacaat gcagcagcgt ttgacaagct agcgcatagaa 300

gcgggcttgt acgcaaaaaa cctgaacggc gatgcgttct cgaacgaagc gaagcaaaaa 360

gcgattgagc tgattaagca agacctcggc cagattgatt tgggtggttta ctctgttggt 420

tctccagtgc gtaagatgcc agacacgggt gagctagtgc gctctgcact aaaaccgatc 480

ggcgaaacgt acacctctac cgcggtagat accaataaag atgtgatcat tgaagccagt 540

gttgaacctg cgaccgagca agaaatcgct gacactgtca ccgtgatggg cggtaaat 600

tgggaactgt ggatccaagc actggaagag gcgggtgttc ttgtgaagg ttgcaaaacc 660

gtggcgtaca gctacatcgg tactgaattg acttggccaa tttactggga tggcgcttta 720

ggccgtgcc aagatggacct agatcgcgca gcgacagcgc tgaacgaaaa gctggcagcg 780

aaagggtgta ccgcaacgt tgcagttttg aatcagtgg tgactcaagc aagctctgcg 840

attcctgtga tgccgctcta catcgcaatg gtgttcaaga agatgcgtga acagggcgtg 900

catgaaggct gtatggagca gatctaccgc atgttcagtc aacgtctgta caaagaagat 960

ggttcagcgc cggaagtgga tgatcacaat cgtctgcgtt tggatgactg ggaactgcgt 1020

gatgacattc agcagcactg ccgtgatctg tggccacaaa tcaccacaga gaacctgcgt 1080

gagctgaccg attacgacat gtacaaagaa gagttcatca agctgtttgg ctttggcatt	1140
gaaggcattg attacgatgc tgacgtcaat ccagaagtcg aatttgatgt gattgatatc	1200
gagtaattta gtgactgagc gtacatgtat acgaagatta ttggtactgg cagctatctg	1260
cccgaacaag tgcggactaa cgccgatctg gaaaaaatgg ttgagacctc tgacgagtgg	1320
attgtcactc gtacaggat tctgtaaagc catatcgccg cgccgaatga aactgtcgcg	1380
acgatgggct ttaccgctgc gaatcgcgcg attgagatgg cggggatcga taaagaccaa	1440
attggcttga ttgtggtggc taccacatca gcaacgcatg catttccaag cgcgcatgt	1500
cagattcaaa gtatgctcgg tattaaaggt tgcccgcgct ttgatgtcgc ggcagcgtgc	1560
gcaggtttca cctacgcgtt aagcatcgcc gaccagtacg ttaaatccgg cgcggttaaa	1620
cacgcgttgg tggtcggttc cgatgtatta gcccgcactt gcgatcctgg cgatcgcggt	1680
acgatcatta ttttcggcga tggcgcaggc gcggccgtac tgagcgcttc tgaagaaccg	1740
ggtattatct ccactcatct tcatgccgat ggccgttacg gtgaattact gacctgccg	1800
aatgccgatc gcgtaaatcc ggataaccg atttacctga caatggcggg caatgaagtc	1860
tttaaagtgg cggtcactga actggcgcat attgtcgatg agacgctggc ggctaataac	1920
ctggatcgct cagaactcga ttggctggtg ccgcatcagg ctaacctgcg tatcattagc	1980
gcgacagcga aaaaactcgg catgtcgatg gacaatgtcg tcgtcacgct ggacaggcac	2040
ggcaataacct ccgcggcttc tgtgccgtgc gcgctggatg aagccgtgcg tgacgggcga	2100
attaaagccg gtcagctggt attgcttgaa gccttcgggg gtggattcac ctggggctcc	2160
gcgctgattc gtttctagta taaggattta aacatgacgc aatttgcat tgtgttcccc	2220
ggtcagggtt ctgagagcgt tgggatgttg gccgagatgg cggcaaatta ccctatcgta	2280
gaagaaacgt ttgtgaagc ttctcgggct ctgggatatg atctgtgggc gtcaccag	2340
caaggtccag cggaagaact gaataaaacc tggcagacgc agccggcggtt attaacgct	2400
tccgtcgcgc ttggcgcgct ttggcagcag cagggcggtg aaatgcctgc gttaatggca	2460
ggtcacagcc tgggcgaata ttccgcgtg gtttgcgtg gcgtcatcaa ctttctgat	2520
gccgttcgtc tgggtgaaat gcgcggtaaa ttcatgcagg aagcggttcc ggaaggcact	2580
ggcggcatgt ctgcgatcat cgggctggat gatgcctcta ttgctaaagc ctgtgaagaa	2640
tctgccgaag ggcaggttgt ttcccggtt aactttaact cgccgggaca ggtggttatac	2700
gccgggcata aagaggcgtt agaactgtcg ggcgcagcct gtaagccgc tggcgcaaaa	2760
cgcgcgctgc cgctgccgtt gagcgtaccg tcgcaactgc cgctgatgaa accagcggca	2820
gataagctgg cggttgaatt agccaaaatt acctttagcg cgccaacggt gccggtagt	2880
aacaacgttg acgtgaaatg tgaaaccgat gccgccgcta tccgcgatgc gctggttcgc	2940

cagttgtaca atccggtaca gtggacgaag agcgtggaat ttatcgcggc gcagggcgtt	3000
gaacatcttt atgaagtggg tccaggtaaa gtcctcactg gtctgacgaa acgtattgtc	3060
gacaccctga cagcgtcggc gctgaacgag ccggcggcgc tgtctcggc acttacgcaa	3120
taaaagagga aaacatgag ctttgaagga aagattgcgc tggtagactgg tgcaagccgt	3180
ggcataggcc gcgcaattgc agagactctc gttgcccgcg gcgcgaaagt tatcgggact	3240
gcgaccagtg aaaatggtag gaagaacatt agcgactatt taggtgctaa cgggaaaggt	3300
ttgatgttga atgtgaccga tctgcatct attgaatctg ttctggaaaa tattcgcgca	3360
gaatttggtg aagtggatat cctggttaat aatgccggtg tcaactgtga taatctgttg	3420
atgcgaatga aagatgatga gtggaacgat attatcgaac ccaacttata atccgttttc	3480
cgctgtcaa aagcggtaat gcgcgctatg atgaaaaagc gttgtggtcg cattatcact	3540
attggttctg tgggttggtac catgggaaat gcaggtcagg caaactacgc tgcggcgaaa	3600
gcgggcctga tgcgtttcag taaatcacat gcgcgtgaag ttgcgtcccg tggattact	3660
gtcaatgttg tggctccggg ttttattgaa acggacatga cgcgtgcgt gtctgacgat	3720
cagcgtgcgg gtatcctggc gcaggtgcct gcgggtcgcc tcggcggcgc tcaggaaatc	3780
gccagtgcgg ttgcattttt agcctctgac gaagcgagtt acatcactgg tgagactctg	3840
cacgtcaacg gcggaatgta catggtttta ttttaagggt tacataaaac atggtagata	3900
aacgcgaatc ctatacaaaa gaagaccttc ttgcctctgg tcgtggtgaa ctgtttggcg	3960
ctaaagggcc gcaactccct gcaccgaaca tgctgatgat ggaccgcgtc gttaatga	4020
ccgaaacggg cggcaatttc gacaaaggct atgtcgaagc cgagctggat atcaatccgg	4080
atctatggtt cticggatgc cactttatcg gcgatccggt gatgcccggt tgtctgggtc	4140
tggatgctat gtggcaattg gtgggattct acctgggctg gttgggcggc gaaggcaaag	4200
gccgcgtctt gggcgtgggc gaagtgaat ttaccggcca ggttctgcgc acagccagga	4260
aagtcaccta tcgtattcat ttcaaagta tcgtaaacg tcgcctgac atgggcctgg	4320
cggacggtga ggttctggtg gatggtcgcc tgatctatac cgcacacgat ttgaaagtcg	4380
gtttgttcca ggatacttcc gcgttctaaa aggaggcaac aaaatgaatc gccgcgttgt	4440
cattacgggt attggtgcag tgacgccggt gggtacaac gctgatagct tctggtgcag	4500
catcaaagag ggtaaatgtg gcattgacaa gatcaaagcg tttgacgcaa ccgatttcaa	4560
agttaagctg gctgccgaag tgaaggactt caccgccgag gactttatcg acaagcgtga	4620
ggcgaaccgt atggaccgtt ttagccagtt tgcgatcgtt gcggcggatg aggcaatcaa	4680

ggacagcaaa ctggacctgg agtcgattga taagaatcgt ttcggcgtca ttgttggtag 4740
 cggcattggc ggcatcgga ccatgagaa gcaggatgaa aagctgatta ccaaaggtcc 4800
 gggctcgtgtg agccctatga ctattccgat gatcattgcg aatatggcaa gcggtaatct 4860
 ggcgattcgt tatggcgcta aaggtatttg caccaccatt gtcaccgcat gtgcgagcgc 4920
 gaacaacagc attgggtagt ctttcgtaa cattaagttt gggtatagcg acgttatgat 4980
 ctctggtggt agcgaagcag gtatcacccc gttgagcctg gcgggttttg cctcgaatgaa 5040
 ggccgtgacc aaatctgagg acccgaagcg cgccagcatc ccgttcgata aggatcgag 5100

cggttttgtg atgggcgagg gcagcggtat cgttatcttg gaagagttgg agcacgcgct 5160
 gaagcgtggt gccaaaatct atgccgagat cgttggtat ggtgcgacct gcgacgcata 5220
 tcatatcacg agcccagcgc cgaatggtga aggtggtgca cgtgcaatga aactggcaat 5280
 ggaagaagat aatgtccgc cagaggacat ttcctatc aacgcgcacg gtacgagcac 5340
 ggcgtaaat gacagcttcg aaaccaagc gatcaagacg gtcttggtg aatacgccta 5400
 caaagtgccg gtgtctagca ccaagagcat gaccggccac ctgctgggcg ctggcggtgc 5460
 agtcgaagcg attatctgtg ccaaagctat tgaagagggt ttcattccgc cgaccatcgg 5520

ctacaaagag gcgatccgg aatgcgacct ggattacgtt cctaacgagg gccgtaatgc 5580
 agaagtcaac tacgttctgt ccaacagcct gggtctcggt ggccataatg cgactctgct 5640
 gttcaaaaag tacaaatga 5659

<210> 14

<211> 40

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 14

gcagttattg gtgcccttaa acgcctggtt gctacgcctg 40

<210> 15

<211> 27

<212>

> DNA

<213> Artificial Sequence

<220><221> source

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

oligonucleotide"

<400> 15

gagccaatat gcgagaacac ccgagaa 27

<210> 16

<211> 48

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 16

cgctgaacgt attgcaggcc gagttgctgc accgtcccg ccaggcag 48

<210> 17

<211> 51

<212>

DNA

<213> Artificial Sequence

<220><221> source

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

<400> 17

ggaattgcc a cgggtcggca ggctccatac gcgaggccag gttatccaac g 51

<210> 18

<211> 35

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 18

aatcaccagc actaaagtgc gcggttcgtt acccg 35

<210> 19

<211> 34

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 19

atctgccgtg gattgcagag tctattcagc tacg 34

<210> 20

<211> 30

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 20

gcaattccat atgacgagcg atgttcacga 30

<210> 21

<211> 29

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 21

ccgctcgagt aaatcagacc gaactcgcg 29

<210> 22

<211> 60

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 22

atgggcagca gccatcatca tcatcatcac agcagcggcc tgggtccgcg cggcagccat 60

<210> 23

```

<211> 43
<212> DNA
<213> Artificial Sequence
<220><221
> source
<223> /note="Description of Artificial Sequence: Synthetic
        oligonucleotide"
<400> 23
acggatcccc ggaatgcgca acgcaattaa tgtaagttag cgc
43
<210> 24
<211> 46
<212> DNA
<213> Artificial Sequence
<220><221> source
<223> /note="Description of Artificial Sequence: Synthetic
        oligonucleotide"
<400> 24
tgcgtcatcg ccattgaatt cctaaatcag accgaactcg cgcagg
46
<210> 25
<211> 20
<212> DNA
<213> Artificial Sequence
<220><221>
source
<223> /note="Description of Artificial Sequence: Synthetic
        oligonucleotide"
<400> 25
attccgggga tccgtcgacc
20
<210> 26
<211> 23
<212> DNA
<213> Artificial Sequence
<220><221> source
<223> /note="Description of Artificial Sequence: Synthetic
        oligonucleotide"

```

<400> 26
aatggcgatg acgcatcctc acg 23
<210> 27
<211> 39
<212> DNA
<213> Artificial Sequence
<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
primer"
<400> 27
acggaaagga gctagcacat gggcagcagc catcatcat 39
<210> 28
<211> 27
<212> DNA
<213> Artificial Sequence
<220><221> source
<223> /note="Description of Artificial Sequence: Synthetic
primer"
<400> 28
gtaaaggatg gacggcggtc acccgcc 27
<210> 29
<211> 25
<212> DNA
<213> Artificial Sequence
<220><221> source
<223> /note="Description of Artificial Sequence: Synthetic
primer"
<400> 29
cacggcgggt gaccgccgtc catcc 25
<210> 30
<211> 43
<212> DNA
<213> Artificial Sequence

<220><221> source

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

<400> 30

ttaattccgg ggatccctaa atcagaccga actcgcgag gtc 43

<210> 31

<211> 24

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400>

> 31

cagccgttta ttgccgactg gatg 24

<210> 32

<211> 27

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 32

ctgttttata agaccgcttc tgcgttc 27

<210> 33

<211> 20

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 33

gcactcgacc ggaattatcg 20

<210> 34

<211> 25

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 34

gcactacgcg tactgtgagc cagag 25

<210> 35

<211> 45

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 35

gaggaataaa ccatgacgag cgatgttcac gacgcgaccg acggc 45

<210> 36

<211> 25

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 36

ctaaatcaga ccgaactcgc gcagg 25

<210> 37

<211> 31

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 37

catggtttat tcctccttat ttaatcgata c

31

<210> 38

<211> 28

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 38

tgacctgctg gagttcggtc tgatttag

28

<210> 39

<211> 1168

<212> PRT

<213> Mycobacterium smegmatis

<400> 39

Met Thr Ile Glu Thr Arg Glu Asp Arg Phe Asn Arg Arg Ile Asp His

1 5 10 15

Leu Phe Glu Thr Asp Pro Gln Phe Ala Ala Ala Arg Pro Asp Glu Ala

20 25 30

Ile Ser Ala Ala Ala Ala Asp Pro Glu Leu Arg Leu Pro Ala Ala Val

35 40 45

Lys Gln Ile Leu Ala Gly Tyr Ala Asp Arg Pro Ala Leu Gly Lys Arg

50 55 60

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

65 70 75 80

Leu Pro Arg Phe Asp Thr Ile Thr Tyr Arg Gln Leu Ala Gly Arg Ile

85 90 95

Gln Ala Val Thr Asn Ala Trp His Asn His Pro Val Asn Ala Gly Asp

100 105 110

Arg Val Ala Ile Leu Gly Phe Thr Ser Val Asp Tyr Thr Thr Ile Asp

115 120 125

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

130 135 140
 Ala Pro Val Ala Gln Leu Gln Pro Ile Val Ala Glu Thr Glu Pro Lys
 145 150 155 160

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

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

 Leu Asn Phe Met Pro Met Ser His Val Met Gly Arg Gly Ile Leu Cys
 290 295 300
 Ser Thr Leu Ala Ser Gly Gly Thr Ala Tyr Phe Ala Ala Arg Ser Asp
 305 310 315 320
 Leu Ser Thr Phe Leu Glu Asp Leu Ala Leu Val Arg Pro Thr Gln Leu
 325 330 335
 Asn Phe Val Pro Arg Ile Trp Asp Met Leu Phe Gln Glu Tyr Gln Ser
 340 345 350

 Arg Leu Asp Asn Arg Arg Ala Glu Gly Ser Glu Asp Arg Ala Glu Ala
 355 360 365
 Ala Val Leu Glu Glu Val Arg Thr Gln Leu Leu Gly Gly Arg Phe Val
 370 375 380

Ser Ala Leu Thr Gly Ser Ala Pro Ile Ser Ala Glu Met Lys Ser Trp
385 390 395 400

Val Glu Asp Leu Leu Asp Met His Leu Leu Glu Gly Tyr Gly Ser Thr
405 410 415

Glu Ala Gly Ala Val Phe Ile Asp Gly Gln Ile Gln Arg Pro Pro Val
420 425 430

Ile Asp Tyr Lys Leu Val Asp Val Pro Asp Leu Gly Tyr Phe Ala Thr
435 440 445

Asp Arg Pro Tyr Pro Arg Gly Glu Leu Leu Val Lys Ser Glu Gln Met
450 455 460

Phe Pro Gly Tyr Tyr Lys Arg Pro Glu Ile Thr Ala Glu Met Phe Asp
465 470 475 480

Glu Asp Gly Tyr Tyr Arg Thr Gly Asp Ile Val Ala Glu Leu Gly Pro
485 490 495

Asp His Leu Glu Tyr Leu Asp Arg Arg Asn Asn Val Leu Lys Leu Ser
500 505 510

Gln Gly Glu Phe Val Thr Val Ser Lys Leu Glu Ala Val Phe Gly Asp
515 520 525

Ser Pro Leu Val Arg Gln Ile Tyr Val Tyr Gly Asn Ser Ala Arg Ser
530 535 540

Tyr Leu Leu Ala Val Val Val Pro Thr Glu Glu Ala Leu Ser Arg Trp
545 550 555 560

Asp Gly Asp Glu Leu Lys Ser Arg Ile Ser Asp Ser Leu Gln Asp Ala
565 570 575

Ala Arg Ala Ala Gly Leu Gln Ser Tyr Glu Ile Pro Arg Asp Phe Leu
580 585 590

Val Glu Thr Thr Pro Phe Thr Leu Glu Asn Gly Leu Leu Thr Gly Ile
595 600 605

Arg Lys Leu Ala Arg Pro Lys Leu Lys Ala His Tyr Gly Glu Arg Leu
610 615 620

Glu Gln Leu Tyr Thr Asp Leu Ala Glu Gly Gln Ala Asn Glu Leu Arg

625 630 635 640
 Glu Leu Arg Arg Asn Gly Ala Asp Arg Pro Val Val Glu Thr Val Ser
 645 650 655
 Arg Ala Ala Val Ala Leu Leu Gly Ala Ser Val Thr Asp Leu Arg Ser
 660 665 670

 Asp Ala His Phe Thr Asp Leu Gly Gly Asp Ser Leu Ser Ala Leu Ser
 675 680 685
 Phe Ser Asn Leu Leu His Glu Ile Phe Asp Val Asp Val Pro Val Gly
 690 695 700
 Val Ile Val Ser Pro Ala Thr Asp Leu Ala Gly Val Ala Ala Tyr Ile
 705 710 715 720
 Glu Gly Glu Leu Arg Gly Ser Lys Arg Pro Thr Tyr Ala Ser Val His
 725 730 735

 Gly Arg Asp Ala Thr Glu Val Arg Ala Arg Asp Leu Ala Leu Gly Lys
 740 745 750
 Phe Ile Asp Ala Lys Thr Leu Ser Ala Ala Pro Gly Leu Pro Arg Ser
 755 760 765
 Gly Thr Glu Ile Arg Thr Val Leu Leu Thr Gly Ala Thr Gly Phe Leu
 770 775 780
 Gly Arg Tyr Leu Ala Leu Glu Trp Leu Glu Arg Met Asp Leu Val Asp
 785 790 795 800

 Gly Lys Val Ile Cys Leu Val Arg Ala Arg Ser Asp Asp Glu Ala Arg
 805 810 815
 Ala Arg Leu Asp Ala Thr Phe Asp Thr Gly Asp Ala Thr Leu Leu Glu
 820 825 830
 His Tyr Arg Ala Leu Ala Ala Asp His Leu Glu Val Ile Ala Gly Asp
 835 840 845
 Lys Gly Glu Ala Asp Leu Gly Leu Asp His Asp Thr Trp Gln Arg Leu
 850 855 860

 Ala Asp Thr Val Asp Leu Ile Val Asp Pro Ala Ala Leu Val Asn His
 865 870 875 880

Val Leu Pro Tyr Ser Gln Met Phe Gly Pro Asn Ala Leu Gly Thr Ala

885 890 895

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

900 905 910

Val Ser Thr Ile Gly Val Gly Gln Gly Ile Ser Pro Glu Ala Phe Val

915 920 925

Glu Asp Ala Asp Ile Arg Glu Ile Ser Ala Thr Arg Arg Val Asp Asp

930 935 940

Ser Tyr Ala Asn Gly Tyr Gly Asn Ser Lys Trp Ala Gly Glu Val Leu

945 950 955 960

Leu Arg Glu Ala His Asp Trp Cys Gly Leu Pro Val Ser Val Phe Arg

965 970 975

Cys Asp Met Ile Leu Ala Asp Thr Thr Tyr Ser Gly Gln Leu Asn Leu

980 985 990

Pro Asp Met Phe Thr Arg Leu Met Leu Ser Leu Val Ala Thr Gly Ile

995 1000 1005

Ala Pro Gly Ser Phe Tyr Glu Leu Asp Ala Asp Gly Asn Arg Gln

1010 1015 1020

Arg Ala His Tyr Asp Gly Leu Pro Val Glu Phe Ile Ala Glu Ala

1025 1030 1035

Ile Ser Thr Ile Gly Ser Gln Val Thr Asp Gly Phe Glu Thr Phe

1040 1045 1050

His Val Met Asn Pro Tyr Asp Asp Gly Ile Gly Leu Asp Glu Tyr

1055 1060 1065

Val Asp Trp Leu Ile Glu Ala Gly Tyr Pro Val His Arg Val Asp

1070 1075 1080

Asp Tyr Ala Thr Trp Leu Ser Arg Phe Glu Thr Ala Leu Arg Ala

1085 1090 1095

Leu Pro Glu Arg Gln Arg Gln Ala Ser Leu Leu Pro Leu Leu His

1100 1105 1110

Asn Tyr Gln Gln Pro Ser Pro Pro Val Cys Gly Ala Met Ala Pro

1115 1120 1125
 Thr Asp Arg Phe Arg Ala Ala Val Gln Asp Ala Lys Ile Gly Pro
 1130 1135 1140
 Asp Lys Asp Ile Pro His Val Thr Ala Asp Val Ile Val Lys Tyr
 1145 1150 1155
 Ile Ser Asn Leu Gln Met Leu Gly Leu Leu
 1160 1165
 <210> 40
 <211> 1168
 <212> PRT

 <213> Mycobacterium tuberculosis
 <400> 40
 Met Ser Ile Asn Asp Gln Arg Leu Thr Arg Arg Val Glu Asp Leu Tyr
 1 5 10 15
 Ala Ser Asp Ala Gln Phe Ala Ala Ala Ser Pro Asn Glu Ala Ile Thr
 20 25 30
 Gln Ala Ile Asp Gln Pro Gly Val Ala Leu Pro Gln Leu Ile Arg Met
 35 40 45
 Val Met Glu Gly Tyr Ala Asp Arg Pro Ala Leu Gly Gln Arg Ala Leu
 50 55 60

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

 Ile Ala Leu Ile Arg Leu Gly Ala Val Ser Val Pro Leu Gln Thr Ser
 130 135 140
 Ala Pro Val Thr Gly Leu Arg Pro Ile Val Thr Glu Thr Glu Pro Thr
 145 150 155 160

Met Ile Ala Thr Ser Ile Asp Asn Leu Gly Asp Ala Val Glu Val Leu
165 170 175

Ala Gly His Ala Pro Ala Arg Leu Val Val Phe Asp Tyr His Gly Lys
180 185 190

Val Asp Thr His Arg Glu Ala Val Glu Ala Ala Arg Ala Arg Leu Ala
195 200 205

Gly Ser Val Thr Ile Asp Thr Leu Ala Glu Leu Ile Glu Arg Gly Arg
210 215 220

Ala Leu Pro Ala Thr Pro Ile Ala Asp Ser Ala Asp Asp Ala Leu Ala
225 230 235 240

Leu Leu Ile Tyr Thr Ser Gly Ser Thr Gly Ala Pro Lys Gly Ala Met
245 250 255

Tyr Arg Glu Ser Gln Val Met Ser Phe Trp Arg Lys Ser Ser Gly Trp
260 265 270

Phe Glu Pro Ser Gly Tyr Pro Ser Ile Thr Leu Asn Phe Met Pro Met
275 280 285

Ser His Val Gly Gly Arg Gln Val Leu Tyr Gly Thr Leu Ser Asn Gly
290 295 300

Gly Thr Ala Tyr Phe Val Ala Lys Ser Asp Leu Ser Thr Leu Phe Glu
305 310 315 320

Asp Leu Ala Leu Val Arg Pro Thr Glu Leu Cys Phe Val Pro Arg Ile
325 330 335

Trp Asp Met Val Phe Ala Glu Phe His Ser Glu Val Asp Arg Arg Leu
340 345 350

Val Asp Gly Ala Asp Arg Ala Ala Leu Glu Ala Gln Val Lys Ala Glu
355 360 365

Leu Arg Glu Asn Val Leu Gly Gly Arg Phe Val Met Ala Leu Thr Gly
370 375 380

Ser Ala Pro Ile Ser Ala Glu Met Thr Ala Trp Val Glu Ser Leu Leu
385 390 395 400

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

405 410 415
 Val Leu Asn Asp Gly Met Val Arg Arg Pro Ala Val Ile Asp Tyr Lys
 420 425 430
 Leu Val Asp Val Pro Glu Leu Gly Tyr Phe Gly Thr Asp Gln Pro Tyr
 435 440 445

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

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

 Gly Leu Gln Ser Tyr Glu Ile Pro Arg Asp Phe Ile Ile Glu Thr Thr
 580 585 590
 Pro Phe Thr Leu Glu Asn Gly Leu Leu Thr Gly Ile Arg Lys Leu Ala
 595 600 605
 Arg Pro Gln Leu Lys Lys Phe Tyr Gly Glu Arg Leu Glu Arg Leu Tyr
 610 615 620
 Thr Glu Leu Ala Asp Ser Gln Ser Asn Glu Leu Arg Glu Leu Arg Gln
 625 630 635 640

 Ser Gly Pro Asp Ala Pro Val Leu Pro Thr Leu Cys Arg Ala Ala Ala
 645 650 655

Ala Leu Leu Gly Ser Thr Ala Ala Asp Val Arg Pro Asp Ala His Phe
660 665 670

Ala Asp Leu Gly Gly Asp Ser Leu Ser Ala Leu Ser Leu Ala Asn Leu
675 680 685

Leu His Glu Ile Phe Gly Val Asp Val Pro Val Gly Val Ile Val Ser
690 695 700

Pro Ala Ser Asp Leu Arg Ala Leu Ala Asp His Ile Glu Ala Ala Arg
705 710 715 720

Thr Gly Val Arg Arg Pro Ser Phe Ala Ser Ile His Gly Arg Ser Ala
725 730 735

Thr Glu Val His Ala Ser Asp Leu Thr Leu Asp Lys Phe Ile Asp Ala
740 745 750

Ala Thr Leu Ala Ala Ala Pro Asn Leu Pro Ala Pro Ser Ala Gln Val
755 760 765

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

Ala Leu Glu Trp Leu Asp Arg Met Asp Leu Val Asn Gly Lys Leu Ile
785 790 795 800

Cys Leu Val Arg Ala Arg Ser Asp Glu Glu Ala Gln Ala Arg Leu Asp
805 810 815

Ala Thr Phe Asp Ser Gly Asp Pro Tyr Leu Val Arg His Tyr Arg Glu
820 825 830

Leu Gly Ala Gly Arg Leu Glu Val Leu Ala Gly Asp Lys Gly Glu Ala
835 840 845

Asp Leu Gly Leu Asp Arg Val Thr Trp Gln Arg Leu Ala Asp Thr Val
850 855 860

Asp Leu Ile Val Asp Pro Ala Ala Leu Val Asn His Val Leu Pro Tyr
865 870 875 880

Ser Gln Leu Phe Gly Pro Asn Ala Ala Gly Thr Ala Glu Leu Leu Arg
885 890 895

Leu Ala Leu Thr Gly Lys Arg Lys Pro Tyr Ile Tyr Thr Ser Thr Ile

900	905	910	
Ala Val Gly Glu Gln Ile Pro Pro Glu Ala Phe Thr Glu Asp Ala Asp			
915	920	925	
Ile Arg Ala Ile Ser Pro Thr Arg Arg Ile Asp Asp Ser Tyr Ala Asn			
930	935	940	
Gly Tyr Ala Asn Ser Lys Trp Ala Gly Glu Val Leu Leu Arg Glu Ala			
945	950	955	960
His Glu Gln Cys Gly Leu Pro Val Thr Val Phe Arg Cys Asp Met Ile			
965	970	975	
Leu Ala Asp Thr Ser Tyr Thr Gly Gln Leu Asn Leu Pro Asp Met Phe			
980	985	990	
Thr Arg Leu Met Leu Ser Leu Ala Ala Thr Gly Ile Ala Pro Gly Ser			
995	1000	1005	
Phe Tyr Glu Leu Asp Ala His Gly Asn Arg Gln Arg Ala His Tyr			
1010	1015	1020	
Asp Gly Leu Pro Val Glu Phe Val Ala Glu Ala Ile Cys Thr Leu			
1025	1030	1035	
Gly Thr His Ser Pro Asp Arg Phe Val Thr Tyr His Val Met Asn			
1040	1045	1050	
Pro Tyr Asp Asp Gly Ile Gly Leu Asp Glu Phe Val Asp Trp Leu			
1055	1060	1065	
Asn Ser Pro Thr Ser Gly Ser Gly Cys Thr Ile Gln Arg Ile Ala			
1070	1075	1080	
Asp Tyr Gly Glu Trp Leu Gln Arg Phe Glu Thr Ser Leu Arg Ala			
1085	1090	1095	
Leu Pro Asp Arg Gln Arg His Ala Ser Leu Leu Pro Leu Leu His			
1100	1105	1110	
Asn Tyr Arg Glu Pro Ala Lys Pro Ile Cys Gly Ser Ile Ala Pro			
1115	1120	1125	
Thr Asp Gln Phe Arg Ala Ala Val Gln Glu Ala Lys Ile Gly Pro			
1130	1135	1140	

Asp Lys Asp Ile Pro His Leu Thr Ala Ala Ile Ile Ala Lys Tyr

1145

1150

1155

Ile Ser Asn Leu Arg Leu Leu Gly Leu Leu

1160

1165

<210> 41

<211> 3507

<212> DNA

<213> Mycobacterium smegmatis

<400> 41

atgacgatcg aaacgcgcga agaccgcttc aaccggcgca ttgaccactt gttcgaaacc 60
gaccgcagtg tcgccgccgc ccgtcccgac gaggcgatca gcgcggctgc cgccgatccg 120
gagttgcgcc ttcttgcgcg ggtcaaacag attctggccg gctatgcgga ccgcctcgcg 180

ctgggcaagc gcgccgtcga gttcgtcacc gacgaagaag gccgcaccac cgcaagctc 240
ctgccccgct tcgacacat cactaccgt cagctcgag gccggatcca ggccgtgacc 300
aatgcctggc acaaccatcc ggtgaatgcc ggtgaccgcg tggccatcct gggtttcacc 360
agtgtcgact acacgacgat cgacatcgcc ctgctcgaac tcggcgccgt gtccgtaccg 420
ctgcagacca gtgcgccgtt ggcccaactg cagccgatcg tcgccgagac cgagcccaag 480
gtgatcgctg cgagcgtcga ctctctcgcc gacgcagtcg ctctcgtcga gtccgggccc 540
gcgccgtcgc gactggtggt gttcgactac agccacgagg tcgacgatca gcgtgaggcg 600

ttcgaggcgg ccaagggcaa gctcgcaggc accggcgtcg tcgtcgagac gatcaccgac 660
gcactggacc gcggggcggtc actcgccgac gcaccgctct acgtgcccga cgaggccgac 720
ccgctgacct ttctcatcta cactccggc agcaccggca ctcccaaggg cgcgatgtac 780
cccagatcca agaccgccac gatgtggcag gccgggtcca agggccggtg ggacgagacc 840
ctcggcgtga tgccgtcgat caccctgaac ttcatgcca tgagtcacgt catggggcgc 900
ggcatcctgt gcagcacact cgccagcggc ggaaccgctg acttcgccgc acgcagcgac 960
ctgtccacct tctggagga cctcgccctc gtgcggccca cgcagctcaa ctctgttctt 1020

cgcatctggg acatgctgtt ccaggagtac cagagccgcc tcgacaaccg ccgcgccgag 1080
ggatccgagg accgagccga agccgcagtc ctggaagagg tccgcacca actgctcggc 1140
gggcgattcg ttctggccct gaccggatcg gctcccatct cggcggagat gaagagctgg 1200
gtcgaggacc tgctcgacat gcatctgctg gagggctacg gctccaccga ggccggcgcg 1260
gtgttcatcg acgggcagat ccagcgcccc ccggtcatcg actacaagct ggtcgacgtg 1320
cccgatctcg gctacttcgc cacggaccgg ccctaccgcg gcggcgaact tctggtcaag 1380

tccgagcaga tgttccccgg ctactacaag cgtccggaga tcaccgccga gatgttcgac	1440
gaggacgggt actaccgcac cggcgacatc gtcgccgagc tcggggcccga ccatctcgaa	1500
tacctcgacc gccgcaacaa cgtgctgaaa ctgtcgagg gcgaattcgt cacggctcc	1560
aagctggagg cgggtgttcgg cgacagcccc ctggtacgcc agatctacgt ctacggcaac	1620
agcgcgcggt cctatctgct ggccgtcgtg gtcccaccg aagaggcact gtcacgttgg	1680
gacggtgacg aactcaagtc gcgcacacg gactcactgc aggacgcggc acgagccgcc	1740
ggattgcagt cgtatgagat cccgcgtgac ttctctgctg agacaacacc ttacacgtg	1800
gagaacggcc tgcgaccgg tatccgcaag ctggcccgcc cgaaactgaa ggcgcactac	1860
ggcgaacgcc tcgaacagct ctacaccgac ctggccgagg ggcaggccaa cgagtgcgc	1920
gagttgcgcc gcaacggagc cgaccggccc gtggtcgaga ccgtcagccg cgccgcggtc	1980
gcactgctcg gtgcctcgt cacggatctg cggtcgatg cgcacttcac cgatctgggt	2040
ggagattcgt tgcggcctt gagcttctcg aacctgttc acgagatctt cgatgtcgac	2100
gtgccggtcg gcgtcatcgt cagcccgccc accgacctgg caggcgtcgc ggcctacatc	2160
gagggcgaac tgcgcggctc caagcgcccc acatacgcgt cgggtgcacgg gcgcgacgcc	2220
accgaggtgc gcgcgcgtga tctcgccctg ggcaagttca tcgacgcaa gacctgtcc	2280
gccgcgccgg gtctgcccg ttcgggcacc gagatccga ccgtgctgct gaccggcgcc	2340
accgggttcc tgggccgcta tctggcgctg gaatggctgg agcgcacgga cctggtggac	2400
ggcaaggtga tctgcctggt gcgcgcccc agcgacgacg agggccgggc gcgtctggac	2460
gccacgttcg acaccgggga cgcgacactg ctcgagcact accgcgcgt ggcagccgat	2520
cacctcgagg tgatcgccgg tgacaagggc gaggccgatc tgggtctcga ccacgacacg	2580
tggcagcgac tggccgacac cgtcgatctg atcgtcgatc cggccgcct ggtcaatcac	2640
gtctgccgt acagccagat gttcggacc aatgcgtcg gcaccgccga actcatccgg	2700
atcgcgtga ccaccacgat caagccgtac gtgtacgtct cgacgatcgg tgtgggacag	2760
ggcatctccc ccaggcggtt cgtcgaggac gccgacatcc gcgagatcag cgcgacgcgc	2820
cgggtcgacg actcgtacg caacggctac ggcaacagca agtgggccgg cgaggctctg	2880
ctgcgggagg cgcacgactg gtgtggtctg ccggtctcgg tgttccgtg cgacatgatc	2940
ctggccgaca cgacctactc gggtcagctg aacctgccgg acatgttcac ccgctgatg	3000
ctgagcctcg tggcgaccgg catcgcccc ggttcgttct acgaactcga tgcggacggc	3060
aaccggcagc gcgccacta cgacgggctg cccgtggagt tcacgccga ggcgatctcc	3120

accatcggct cgcaggcac cgacggattc gagacgttcc acgtgatgaa cccgtacgac 3180
gacggcatcg gcctcgacga gtacgtggac tggctgatcg aggccggcta ccccgtagac 3240
cgcgtcgacg actacgccac ctggctgagc cggttcgaaa ccgcactgcg ggcctgccc 3300
gaacggcaac gtcaggcctc gctgctgccg ctgctgcaca actatcagca gccctcaccg 3360
cccgtgtgcg gtgcatggc acccaccgac cggttccgtg ccgcggtgca ggacgcgaag 3420
atcgccccg acaaggacat tccgcacgtc acggccgacg tgatcgtcaa gtacatcagc 3480
aacctgcaga tgctcggatt gctgttaa 3507

<210> 42

<211> 3507

<212> DNA

<213> Mycobacterium tuberculosis

<400> 42

atgtcgatca acgatcagcg actgacacgc cgcgtcgagg acctatacgc cagcgacgcc 60
cagttcgccg ccgccagtcc caacgaggcg atcaccagg cgatcgacca gcccggggtc 120
gcgcttcac agtcatccg tatggtcatg gagggctacg ccgatcgcc ggcactcggc 180
cagcgtgcgc tccgcttctg caccgacccc gacagcggcc gcaccatggt cgagctactg 240
ccgcggttcg agaccatcac ctaccgcgaa ctgtgggccc gcgccggcac attggccacc 300
gcgttgagcg ctgagccccg gatcgggccg ggcgaccggg ttgctgtgtt gggttcaac 360

agcgtcgact acacaacat cgacatcgcg ctgatccggt tgggcgccgt gtcggttcca 420
ctgcagacca gtgcgccgtt caccgggttg cgcctgatcg tcaccgagac cgagccgacg 480
atgatcgcca ccagcatcga caatcttggc gacgccgtcg aagtgtgtgc cggtcacgcc 540
ccggccccgc tggctgtatt cgattaccac ggcaaggttg acaccaccg cgaggccgtc 600
gaagccgccc gagctcggtt ggccggctcg gtgacatcg acacattgc cgaactgatc 660
gaacgcggca gggcgtgcc ggccacacc attgccgaca gcgccgacga cgcgtgtgcg 720
ctgctgattt acacctcggg tagtaccggc gcacccaaag gcgccatgta tcgcgagagc 780

caggtgatga gcttctggcg caagtcgagt ggctgggtcg agccgagcgg ttaccctcgc 840
atcacgtga acttcatgcc gatgagccac gtcgggggcc gtcaggtgct ctacgggacg 900
ctttccaacg gcggtaccgc ctacttcgtc gccaaagagc acctgtcgac gctgttcgag 960
gacctcgccc tgggtcggcc cacagaattg tgcttcgtgc cgcgcatctg ggacatggtg 1020
ttcgagagt tccacagcga ggtcgaccgc cgttgggtgg acggcgccga tcgagcggcg 1080
ctggaagcgc aggtgaaggc cgagctgcgg gagaacgtgc tcggcggacg gtttgcattg 1140

gcgctgaccg gttccgcgcc gatctccgct gagatgacgg cgtgggtcga gtccctgctg	1200
gccgacgtgc atttgggtga gggttacggc tccaccgagg ccgggatggt cctgaacgac	1260
ggcatgggtgc ggcgccccgc ggtgatcgac tacaagctgg tcgacgtgcc cgagctgggc	1320
tacttcggca ccgatcagcc ctacccccgg ggcgagctgc tggtaagac gcaaaccatg	1380
ttccccggct actaccagcg cccgatgtc accgccgagg tgttcgaccc cgacggcttc	1440
taccggaccg gggacatcat ggccaaagta ggccccgacc agttcgtcta cctcgaccgc	1500
cgcaacaacg tgctaaagct ctcccagggc gagttcatcg ccgtgtcgaa gctcgaggcg	1560
gtgttcggcg acagcccgtt ggtccgacag atcttcatct acggcaacag tgcccgggcc	1620
taccgctgg cgggtggtgt cccgtccggg gacgcgcttt ctgccatgg catcgagaat	1680
ctcaagcccc tgatcagcga gtccctgcag gaggtagcga gggcggccgg cctgcaatcc	1740
tacgagattc cacgcgactt catcatcgaa accacgccgt tcaccctgga gaacggcctg	1800
ctcacggca tccgaagct ggcacgcccc cagttgaaga agttctatgg cgaacgtctc	1860
gagcggtctt ataccgagct ggccgatagc caatccaacg agctgcgcga gctgcggcaa	1920
agcgggtccc atgcgccgtt gcttcgcagc ctgtgccgtg ccgcggctgc gttgctgggc	1980
tctaccgctg cggatgtgcg gccggacgcg cacttcgccg acctgggtgg tgactcgctc	2040
tcggcgtgt cgttggccaa cctgtgcac gagatcttcg gcgtcgacgt gccggtgggt	2100
gtcattgtca gcccggaag cgacctgcgg gcctggccg accacatcga agcagcgcg	2160
accggcgtca ggcgaccag ctctgcctcg atacacggtc gctccgcgac ggaagtgcac	2220
gccagcgacc tcacgtgga caagttcatc gacgtgcca cctggccgc agccccgaac	2280
ctgccggcac cgagcgccca agtgcgcacc gtactgctga ccggcgccac cggctttttg	2340
ggtcgtacc tggcgtgga atggctcgac cgcattggacc tggtaacgg caagctgac	2400
tgcttggtcc gcgccagatc cgacaggaa gcacaagccc ggctggacgc gacgttcgat	2460
agcggcgacc cgtatttggg gcggcactac cggaattgg gcgccggccg cctcgaggtg	2520
ctgcggcgcg acaaggcgga ggccgacctg ggctggacc gggcacctg gcagcggtta	2580
gccgacacgg tggacctgat cgtggacccc gcggccctgg tcaaccacgt gctgccgtat	2640
agccagctgt tcggcccaaa cgcggcgggc accgccgagt tgcttcggct ggcgctgacc	2700
ggcaagcgca agccatacat ctacacctcg acgatcgccg tgggcgagca gatcccgccg	2760
gaggcgttca ccaggacgc cgacatccgg gccatcagcc cgaccgcag gatcgacgac	2820
agctacgcca acggctacgc gaacagcaag tgggccggcg aggtgctgct gcgcgaagct	2880

cacgagcagt gcggcctgcc ggtgacggtc ttccgctgcg acatgatcct ggccgacacc 2940
agctataccg gtcagctcaa cctgccggac atgttcaccc ggctgatgct gaggctggcc 3000
gctaccggca tcgaccccg ttcgttctat gagctggatg cgcacggcaa tcggcaacgc 3060
gccactatg acggcttgcc ggtcgaattc gtcgcagaag ccatttgac ccttgggaca 3120
catagcccg accgttttgt caccaccac gtgatgaacc cctacgacga cggcatcggg 3180
ctggacgagt tcgtcactg gctcaactcc ccaactagcg ggtccggtt cagatccag 3240
cggatcgccg actacggcga gtggctgcag cggttcgaga cttcgtgcg tgccttgccg 3300

gatcgccagc gccacgcctc gctgtgccc ttgtgcaca actaccgaga gcctgcaaag 3360
ccgatatgcg ggtaaatgc gccaccgac cagttccgcg ctgccgtcca agaagcgaag 3420
atcggtcggg acaaagacat tccgcacctc acggcggcga tcatcgcgaa gtacatcagc 3480
aacctgcgac tgctcgggct gctgtga 3507

<210> 43
<211> 3522
<212> DNA
<213> Mycobacterium smegmatis
<400> 43

atgaccagcg atgttcacga cgccacagac ggcgtcaccg aaaccgcact cgacgacgag 60
cagtcgaccc gccgcatcgc cgagctgtac gccaccgatc ccgagttcgc cgccgcgca 120

ccgttgcccc ccgtggtcga cgcggcgac aaacccgggc tcgggctggc agagatcctg 180
cagaccctgt tcaccggcta cggtgaccgc ccggcgctgg gataccgacg ccgtgaactg 240
gccaccgacg agggcgggag caccgtgacg cgtctgtgac cgcggttcga caccctcacc 300
tacgcccagg tgtggtcgcg cgtgcaagcg gtccgcccgg ccctgcgcca caacttcgcg 360
cagccgatct accccggcga cgccgtcgcg acgatcgggt tcgagagtcg cgattacctg 420
acgtggatc tcgtatgcgc ctacctgggc ctggtgagtg ttccgctgca gcacaacgca 480
ccggtcagcc ggctgcgcc gatcctggcc gaggtcgaac cgcggatcct caccgtgagc 540

gccgaatacc tcgacctgc agtcgaatcc gtgcgggacg tcaactcggg gtcgcagctc 600
gtggtgttcg accatcaccc cgaggtcgac gaccaccgag acgactggc ccgcgcgct 660
gaacaactcg ccggcaaggg catcggcgtc accacctgg acgcatcgc cgacgagggc 720
gccgggctgc cggccgaacc gatctacac gccgacctg atcagcgctc cgcgatgatc 780
ctgtacacct cgggttcac cggcgaccc aagggtcga tgtacaccga ggcgatggtg 840
gcgcggctgt ggacctgtc gttcatcacg ggtgaccca cgccggtcat caacgtcaac 900

ttcattgccgc tcaaccacct gggcggggcgc atccccattt ccaccgccgt gcagaacggt	960
ggaaccagtt acttcgtacc ggaatccgac atgtccacgc tgttcgagga tctcgcgctg	1020
gtgcgcccga ccgaactcgg cctggttccg cgcgtcgccg acatgctcta ccagcaccac	1080
ctcgccaccg tcgaccgcct ggtcacgcag ggcgccgacg aactgaccgc cgagaagcag	1140
gccggtgccg aactgcgtga gcaggtgctc ggcggacgcg tgatcaccgg attcgtcagc	1200
accgcaccgc tggccgcgga gatgagggcg ttcctcgaca tcaccctggg cgcacacatc	1260
gtcgacggct acgggtcac cgagaccggc gccgtgacac gcgacggtgt gatcgtgcgg	1320
ccaccggtga tcgactacaa gctgatcgac gttcccgaac tcggctactt cagcaccgac	1380
aagccctacc cgcgtggcga actgctggtc aggtcgcaaa cgcigactcc cgggtactac	1440
aagcgccccg aggtcaccgc gacgtcttc gaccgggacg gctactacca caccggcgac	1500
gtcatggcgg agaccgcacc cgaccacctg gtgtacgtgg accgtcgcaa caacgtctc	1560
aaactcgcg agggcgagtt cgtggcggtc gccaacctgg aggcggtgtt ctccggcgcg	1620
gcgtggtgc gccagatctt cgtgtacggc aacagcgagc gcagtttctt tctggccgtg	1680
gtggtccga cgccggaggc gtcgagcag tacgatccgg ccgcgctcaa ggccgcgctg	1740
gccgactcgc tgcagcgac cgacgcgac gccgaactgc aatctacga ggtgccggcc	1800
gatttcacg tcgagaccga gccgttcagc gccgccaacg ggctgctgtc gggtgtcgga	1860
aaactgctgc ggcccaacct caaagaccgc tacgggcagc gcctggagca gatgtacgcc	1920
gatatcgcg ccacgcaggc caaccagtgc cggaactgc ggcgcgcggc cgccacacaa	1980
ccggtgatcg acacctcac ccaggccgt gccacgatcc tcggcaccgg gagcgaggtg	2040
gcatcgacg ccacttcac cgacctgggc ggggattccc tgtcggcgt gacatttcg	2100
aacctgctga gcgatttctt cggtttcgaa gttcccgctg gcaccatcgt gaaccggcc	2160
accaacctcg cccaactcgc ccagcacatc gaggcgcagc gcaccgcggg tgaccgcagg	2220
ccgagtttca ccaccgtgca cggcgcggac gccaccgaga tccgggcgag tgagctgacc	2280
ctggacaagt tcacgacgc cgaaacgctc cgggccgcac cgggtctgcc caaggtcacc	2340
accgagccac ggacggtgtt gctctcgggc gccaacggct ggctgggccc gttcctcacg	2400
ttgcagtggc tggaacgcct ggcacctgtc ggcggcacc tcatcacgat cgtgcggggc	2460
cgcgacgac cgccggcccg cgcacggctg acccaggcct acgacaccga tcccagttg	2520
tcccgcgct tcgcccagct ggccgaccgc cacctgcggg tggtcgccgg tgacatcggc	2580
gaccggaatc tgggcctcac acccgagatc tggcaccggc tcgccgccga ggtcgacctg	2640

gtggtgcatc cggcagcgct ggtcaaccac gtgctcccct accggcagct gttcggcccc 2700
aacgtcgtgg gcacggccga ggtgatcaag ctggccctca ccgaacggat caagcccgtc 2760
acgtacctgt ccaccgtgtc ggtggccatg gggatccccg acttcgagga ggacggcgac 2820
atccggaccg tgagcccgtt gcgcccgtc gacggcggat acgccaacgg ctacggcaac 2880
agcaagtggg ccggcgaggt gctgctgcgg gaggcccacg atctgtgcgg gctgcccgtg 2940
gcgacgttcc gctcggacat gatcctggcg catccgcgt accgcgttca ggtcaacgtg 3000
ccagacatgt tcacgcgact cctgttgagc ctcttgatca ccggcgtcgc gccgcggtcg 3060

ttctacatcg gagacggtga gcgcccgcgg gcgcactacc ccggcctgac ggtcgatttc 3120
gtggccgagg cggtcacgac gctcggcgcg cagcagcgcg agggatacgt gtcctacgac 3180
gtgatgaacc cgcacgacga cgggatctcc ctggatgtgt tcgtggactg gctgatccgg 3240
gcggggccatc cgatcgaccg ggtcgacgac tacgacgact ggggtcgtcg gttcgagacc 3300
gcgttgaccg cgcttccga gaagcgccgc gcacagaccg tactgccgt gctgcacgcg 3360
ttccgcgtc cgcaggcacc gttgcgcggc gcacccgaac ccacggaggt gttccacgcc 3420
gcggtgcga ccgcgaaggt gggcccggga gacatcccgc acctcgacga ggcgctgac 3480

gacaagtaca tacgcgatct gcgtgagttc ggtctgatct ga 3522

<210> 44

<211> 3582

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 44

atgggcagca gccatcatca tcatcatcac agcagcggcc tgggtccgcg cggcagccat 60
atgacgagcg atgttcagca cgcgaccgac ggcgttaccg agactgcact ggatgatgag 120
cagagcactc gtcgtattgc agaactgtac gcaacggacc cagagttcgc agcagcagct 180
cctctgccgg ccgttgtcga tgcggcgcac aaaccgggcc tgcgtctggc ggaaatcctg 240

cagaccctgt tcaccggcta cggcgatcgt ccggcgttgg gctatcgtgc acgtgagctg 300
gcgacggacg aaggcggctg tacggtcacg cgtctgtgc cgcgcttcga taccctgacc 360
tatgcacagg tgtggagccg gtttcaagca gtggctgcag cgttgcgtca caatttcga 420
caaccgattt acccgggcga cgcggtcgcg actatcggct ttgcgagccc ggactatttg 480
acgtggatc tgggtgtgcg gtatctgggc ctggtcagcg ttcctttgca gcataacgt 540

ccggtgtctc gcctggcccc gattctggcc gaggtggaac cgctattct gacggtgagc	600
gcagaatacc tggacctggc ggttgaatcc gtcctgatg tgaactccgt cagccagctg	660
gttgttttcg accatcatcc ggaagtggac gatcacctg acgcactggc tcgcgcacgc	720
gagcagctgg ccggcaaagg tatcgagtt acgacctgg atcgatcgc agacgaaggc	780
gcaggtttgc cggctgagcc gatttacacg gcggatcacg atcagctct ggccatgatt	840
ctgtatacca gcggctctac ggggtctccg aaaggcgca tgtacaccga agcgatggtg	900
gctcgctgt ggactatgag ctttatcacg ggcgaccga ccccggttat caacgtgaac	960
ttcatgccg tgaacctct gggcggtcgt atcccatta gcaccgccgt gcagaatggc	1020
ggtaccagct acttcgttcc ggaaagcgac atgagcacgc tgtttgagga tctggccctg	1080
gtccgcccta ccgaactggg tctggtgccg cgtgttcgg acatgctgta ccagcatcat	1140
ctggcgaccg tggatcgct ggtgacctg ggcgcgacg aactgactgc ggaagcag	1200
gccggtgcgg aactgcgtga acaggtcttg ggcggtcgtg ttatcaccgg tttgtttcc	1260
accgcgccgt tggcggcaga gatgcgtgct tttctggata tcacctggg tgcacacatc	1320
gttgacggtt acggtctgac cgaaaccggt gcggtcacc gtgatggtg gattgttcgt	1380
cctccggtca ttgattaca gctgatcat gtcccgagc tgggttactt ctccaccgac	1440
aaaccgtacc cgcgtggca gctgctggt cgtagccaaa cgttgactcc gggttactac	1500
aagcgcccag aagtcaccgc gtccgttttc gatcgcgacg gctattacca caccggcgac	1560
gtgatggcag aaaccgcgcc agaccactg gtgtatgtgg accgcgcaa caatgttctg	1620
aagctggcgc aaggtgaatt tgtcgccgtg gctaacctgg aggcggttt cagcgcgct	1680
gctctggtcc gccagatttt cgtgtatggt aacagcgagc gcagctttct gttggtgtt	1740
gttgtcccta ccccgaggc gctggagcaa tacgacctg ccgattgaa agcagccctg	1800
gcggattcgc tgcagcgtac ggcgctgat gccgagctgc agagctatga agtgccggcg	1860
gacttcattg ttgagactga gccttttagc gctgcgaac gctgctgag cgggtttggc	1920
aagttgctgc gtccgaattt gaaggatcgc tacggtcagc gtttgagca gatgtacgcg	1980
gacatcgcg ctacgcaggc gaaccaattg cgtgaactgc gccgtgctgc ggctactcaa	2040
ccggtgatcg acacgtgac gcaagctgcg gcgacctcc tgggtaccgg cagcgaggtt	2100
gcaagcgac cacactttac tgatttgggc ggtgattctc tgagcgctg gacgttgagc	2160
aacttgctgt ctgacttctt tggctttgaa gtcccggttg gcacgattgt taaccagcg	2220
actaatctgg cacagctggc gcaacatatc gaggcgcagc gcacggcggg tgaccgccgt	2280

ccatccttta cgacggtcca cggtagcgat gctacggaaa tccgtgcaag cgaactgact 2340

ctggacaaat tcatcgacgc tgagactctg cgcgcagcac ctggtttgcc gaaggttacg 2400

actgagccgc gtacggtcct gttgagcggg gccaatgggt ggtagggcgc ctctctgacc 2460

ctgcagtggc tggaacgttt ggcaccggtt ggcggtagcc tgatcacat tgtgcgcggt 2520

cgtgacgatg cagcggcacg tgcacgtttg actcaggctt acgatacga cccagagctg 2580

tcccgcgct tcgtgagtt ggcggatcgc cacttgctg tggtagcagg tgatatcggc 2640

gatccgaatc tgggcctgac cccggagatt tggcaccgtc tggcagcaga ggtcgatctg 2700

gtcgttcac cagcggccct ggtcaaccac gtctgcccgt accgccagct gtttggtccg 2760

aatgttggtg gcaccgccga agttatcaag ttggctctga ccgagcgcac caagcctgtt 2820

acctacctgt ccacggttag cgtcgcgatg ggtattcctg attttagga ggacggtgac 2880

attcgtaccg tcagcccgtt tcgtccgtg gatggtggct atgcaaatgg ctatggcaac 2940

agcaagtggg ctggcggagt gctgctgccc gaggcacatg acctgtgtgg cctgccggtt 3000

gcgacgtttc gtagcgacat gattctggcc caccgcgct accgtggcca agtgaatgtg 3060

ccggacatgt tcacccgtc gctgctgtcc ctgctgatca cgggtgtggc accgcgttcc 3120

ttctacattg gtgatggcga gcgtccgct gcacactacc cgggcctgac cgtcgatttt 3180

gttgcggaag cggttactac cctgggtgct cagcaacgtg agggttatgt ctcgatgac 3240

gttatgaatc cgcacgatga cggatttagc ttggatgtct ttgtggactg gctgattcgt 3300

gcgggccacc caattgaccg tgttgacgac tatgatgact gggtagctcg ttttgaacc 3360

gcgttgaccg ccttgccgga gaaacgtcgt gcgcagaccg ttctgccgt gctgcatgcc 3420

tttcgcgcgc cacaggcgcg gttgcgtggc gccctgaac cgaccgaagt gtttcatgca 3480

gcggtgcgta ccgctaaagt cggtaggggt gatattccgc acctggatga agccctgac 3540

gacaagtaca tccgtgacct gcgcgagttc ggtctgattt ag 3582

<210> 45

<211> 66

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 45

cggttctggc aaatattctg aaatgagctg ttgacaatta atcaaatccg gctcgtataa 60

tgtgtg 66

<210> 46

<211> 30

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 46

ggtttattcc tccttattta atcgatacat 30

<210> 47

<211> 59

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 47

atgtatcgat taaataagga ggaataaacc atgggcacga gcatgttca cgacgcgac 59

<210> 48

<211> 59

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 48

atgtatcgat taaataagga ggaataaacc gtgggcacga gcatgttca cgacgcgac 59

<210> 49

<211> 59

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
primer"
<400> 49
atgtatcgat taaataagga ggaataaacc ttgggcacga gcgatgttca cgacgcgac 59
<210> 50
<211> 26
<212> DNA
<213> Artificial Sequence
<220><221> source
<223> /note="Description of Artificial Sequence: Synthetic
primer"
<400> 50
ttctaaatca gaccgaactc gcgcag 26
<210> 51
<211> 48
<212> DNA
<213> Artificial Sequence

<220><221> source
<223> /note="Description of Artificial Sequence: Synthetic
primer"
<400> 51
ctgcgcgagt tcggtctgat ttagaattcc tcgaggatgg tagtgtgg 48
<210> 52
<211> 27
<212> DNA
<213> Artificial Sequence
<220><221> source
<223> /note="Description of Artificial Sequence: Synthetic
primer"
<400> 52
cagtcgacat acgaaacggg aatgcgg 27
<210> 53
<211> 56
<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 53

ccgcattccc gtttcgtatg tcgactgaaa cctcaggcat tgagaagcac acggtc 56

<210> 54

<211> 56

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 54

ctcatttcag aatatttgcc agaaccgtta atttcctaatt gcaggagtcg cataag 56

<210> 55

<211> 20

<212> DNA

<213> Artificial Sequence

<220><221> source

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

primer"

<400> 55

ggatctcgac gctctcctt 20

<210> 56

<211> 28

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 56

tcaaaaacgc cattaacctg atgttctg 28