



(12)发明专利

(10)授权公告号 CN 104704113 B

(45)授权公告日 2018.01.23

(21)申请号 201380026546.8

X·李 Z·胡 B·朱

(22)申请日 2013.04.02

(74)专利代理机构 中国国际贸易促进委员会专利商标事务所 11038

(65)同一申请的已公布的文献号

申请公布号 CN 104704113 A

代理人 袁泉

(43)申请公布日 2015.06.10

(51)Int.Cl.

C12N 9/02(2006.01)

(30)优先权数据

C12P 7/04(2006.01)

61/619,309 2012.04.02 US

C12P 7/64(2006.01)

(85)PCT国际申请进入国家阶段日

2014.11.21

(56)对比文件

CN 102264910 A,2011.11.30,

(86)PCT国际申请的申请数据

WO 2010/135624 A2,2010.11.25,

PCT/US2013/035040 2013.04.02

CN 102325864 A,2012.01.18,

(87)PCT国际申请的公布数据

W02013/152052 EN 2013.10.10

殷福珊.天然脂肪醇的新型生产技术.《日用化学品科学》.2006,第29卷(第12期),第4-5页.

D. Grahame HARDIE et al..THE

(73)专利权人 REG生命科学有限责任公司

地址 美国加利福尼亚

REGULATION OF FATTY ACID BIOSYNTHESIS.

《FEBS LETTERS》.1978,第91卷(第1期),第1-7页.

(72)发明人 德里克·L·格林菲尔德

伊利莎白·J·克拉克

伊莱·S·格罗班

维克朗斯·阿拉盖达 S·李

审查员 曹梦超

权利要求书3页 说明书44页

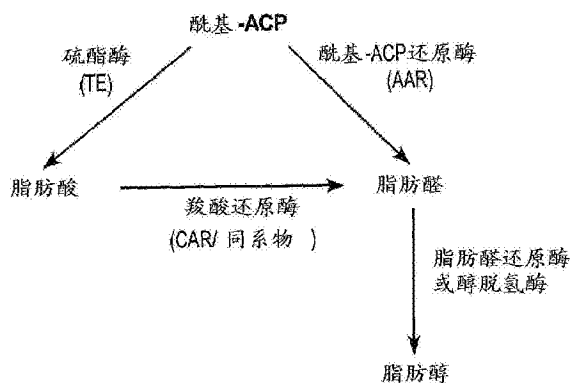
序列表38页 附图12页

(54)发明名称

CAR酶及脂肪醇的改良生产

(57)摘要

本发明公开涉及用于在重组宿主细胞中改良生产脂肪醇的变体羧酸还原酶(CAR)。



1. 一种包括如下氨基酸序列的变体羧酸还原酶 (CAR) 多肽, 其中所述的氨基酸序列与 SEQ ID NO:7 的野生型 CAR 多肽具有至少 99% 的序列一致性, 其中所述的变体 CAR 多肽经基因改造, 从而在氨基酸位置处具有至少一个突变, 其中所述的突变选自 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、V926A、V926E、S927K、S927G、M930K、M930R、L1128W, 以及如下组合:

- (a) 突变 A535S;
- (b) 突变 E20R、F288G、Q473I 和 A535S;
- (c) 突变 E20R、F288G、Q473H、A535S、R827A 和 S927G;
- (d) 突变 E20R、S22R、F288G、Q473H、A535S、R827A 和 S927G;
- (e) 突变 S3R、E20R、S22R、F288G、Q473H、A535S、R873S、S927G、M930R 和 L1128W; (f) 突变 D18R、E20R、S22R、F288G、Q473I、A535S、S927G、M930K 和 L1128W;
- (g) 突变 E20R、S22R、F288G、Q473I、A535S、R827C、V926E、S927K 和 M930R;
- (h) 突变 D18R、E20R、F288G、Q473I、A535S、R827C、V926E、M930K 和 L1128W;
- (i) 突变 E20R、S22R、F288G、Q473H、A535S、R827C、V926A、S927K 和 M930R;
- (j) 突变 E20R、S22R、F288G、Q473H、A535S 和 R827C;
- (k) 突变 E20R、S22R、F288G、Q473I、A535S、R827C 和 M930R;
- (l) 突变 E20R、S22R、F288G、Q473I、A535S、I870L、S927G 和 M930R;
- (m) 突变 E20R、S22R、F288G、Q473I、A535S、I870L 和 S927G;
- (n) 突变 D18R、E20R、S22R、F288G、Q473I、A535S、R827C、I870L、V926A 和 S927G;
- (o) 突变 E20R、S22R、F288G、Q473H、A535S、R827C、I870L 和 L1128W;
- (p) 突变 D18R、E20R、S22R、F288G、Q473H、A535S、R827C、I870L、S927G 和 L1128W;
- (q) 突变 E20R、S22R、F288G、Q473I、A535S、R827C、I870L、S927G 和 L1128W;
- (r) 突变 E20R、S22R、F288G、Q473I、A535S、R827C、I870L、S927G、M930K 和 L1128W;
- (s) 突变 E20R、S22R、F288G、Q473H、A535S、I870L、S927G 和 M930K;
- (t) 突变 E20R、F288G、Q473I、A535S、I870L、M930K;
- (u) 突变 E20R、S22R、F288G、Q473H、A535S、S927G、M930K 和 L1128W;
- (v) 突变 D18R、E20R、S22R、F288G、Q473I、A535S、S927G 和 L1128W;
- (w) 突变 E20R、S22R、F288G、Q473I、A535S、R827C、I870L 和 S927G;
- (x) 突变 D18R、E20R、S22R、F288G、Q473I、A535S、R827C、I870L、S927G 和 L1128W;
- (y) 突变 D18R、E20R、S22R、F288G、Q473I、A535S、S927G、M930R 和 L1128W;
- (z) 突变 E20R、S22R、F288G、Q473H、A535S、V926E、S927G 和 M930R; 或
- (zz) 突变 E20R、S22R、F288G、Q473H、A535S、R827C、I870L、V926A 和 L1128W;

其中与表达 SEQ ID NO:7 的野生型 CAR 多肽的重组宿主细胞相比, 所述的变体 CAR 多肽在重组宿主细胞中的表达得到更高的脂肪醇组合物的效价。

2. 一种重组宿主细胞, 其包括如下多核苷酸序列, 所述的多核苷酸序列编码与 SEQ ID NO:7 的野生型 CAR 多肽具有至少 99% 的序列一致性的变体羧酸还原酶 (CAR) 多肽并且在氨基酸位置处具有至少一个突变, 其中所述的突变选自 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、V926A、V926E、S927K、S927G、M930K、M930R、L1128W,以及如下组合:

- (a) 突变A535S;
- (b) 突变E20R、F288G、Q473I和A535S;
- (c) 突变E20R、F288G、Q473H、A535S、R827A和S927G;
- (d) 突变E20R、S22R、F288G、Q473H、A535S、R827A和S927G;
- (e) 突变S3R、E20R、S22R、F288G、Q473H、A535S、R873S、S927G、M930R和L1128W; (f) 突变D18R、E20R、S22R、F288G、Q473I、A535S、S927G、M930K和L1128W;
- (g) 突变E20R、S22R、F288G、Q473I、A535S、R827C、V926E、S927K和M930R;
- (h) 突变D18R、E20R、F288G、Q473I、A535S、R827C、V926E、M930K和L1128W;
- (i) 突变E20R、S22R、F288G、Q473H、A535S、R827C、V926A、S927K和M930R;
- (j) 突变E20R、S22R、F288G、Q473H、A535S和R827C;
- (k) 突变E20R、S22R、F288G、Q473I、A535S、R827C和M930R;
- (l) 突变E20R、S22R、F288G、Q473I、A535S、I870L、S927G和M930R;
- (m) 突变E20R、S22R、F288G、Q473I、A535S、I870L和S927G;
- (n) 突变D18R、E20R、S22R、F288G、Q473I、A535S、R827C、I870L、V926A和S927G;
- (o) 突变E20R、S22R、F288G、Q473H、A535S、R827C、I870L和L1128W;
- (p) 突变D18R、E20R、S22R、F288G、Q473H、A535S、R827C、I870L、S927G和L1128W;
- (q) 突变E20R、S22R、F288G、Q473I、A535S、R827C、I870L、S927G和L1128W;
- (r) 突变E20R、S22R、F288G、Q473I、A535S、R827C、I870L、S927G、M930K和L1128W;
- (s) 突变E20R、S22R、F288G、Q473H、A535S、I870L、S927G和M930K;
- (t) 突变E20R、F288G、Q473I、A535S、I870L、M930K;
- (u) 突变E20R、S22R、F288G、Q473H、A535S、S927G、M930K和L1128W;
- (v) 突变D18R、E20R、S22R、F288G、Q473I、A535S、S927G和L1128W;
- (w) 突变E20R、S22R、F288G、Q473I、A535S、R827C、I870L和S927G;
- (x) 突变D18R、E20R、S22R、F288G、Q473I、A535S、R827C、I870L、S927G和L1128W;
- (y) 突变D18R、E20R、S22R、F288G、Q473I、A535S、S927G、M930R和L1128W;
- (z) 突变E20R、S22R、F288G、Q473H、A535S、V926E、S927G和M930R;或
- (zz) 突变E20R、S22R、F288G、Q473H、A535S、R827C、I870L、V926A和L1128W;

其中当所述的宿主细胞在有效表达所述的变体CAR多肽的条件下在包括碳源的培养基中培养时,比表达所述的SEQ ID NO:7的野生型CAR多肽的宿主细胞以更高的效价或产率来生产脂肪醇组合物。

3. 权利要求2所述的重组宿主细胞,其进一步包括编码硫酯酶多肽的多核苷酸。

4. 权利要求3所述的重组宿主细胞,其进一步包括编码如下的多核苷酸:

- (i) FabB多肽和FadR多肽,或
- (ii) 脂肪醛还原酶。

5. 根据权利要求2所述的重组宿主细胞,其中当表达相应的野生型CAR多肽的宿主细胞在与所述的重组宿主细胞相同的条件下培养时,所述的重组宿主细胞的效价或产率为所述的表达相应的野生型CAR多肽的宿主细胞的效价或产率的至少3倍。

6. 权利要求5所述的重组宿主细胞,其中所述的重组宿主细胞的效价为30g/L至250g/L。
7. 权利要求5所述的重组宿主细胞,其中所述的重组宿主细胞的效价为90g/L至120g/L。
8. 权利要求5所述的重组宿主细胞,其中所述的重组宿主细胞的产率为10%至40%。
9. 一种包括权利要求2-4的任意一项所述的重组宿主细胞的细胞培养物。
10. 权利要求9所述的细胞培养物,其中所述的细胞培养物的生产率为表达所述的相应的野生型CAR多肽的细胞培养物的生产率的至少3倍。
11. 权利要求10所述的细胞培养物,其中所述的生产率范围为0.7mg/L/hr至3g/L/hr。
12. 权利要求11所述的细胞培养物,其中所述的培养物的培养基包括脂肪醇组合物。
13. 根据权利要求2-8的任意一项所述的重组宿主细胞或权利要求9-12的任意一项所述的细胞培养物,其中所述的脂肪醇组合物细胞外收集于有机相中。
14. 权利要求13所述的重组宿主细胞或细胞培养物,其中所述的脂肪醇组合物包括C6、C8、C10、C12、C13、C14、C15、C16、C17或C18脂肪醇中的一种或多种。
15. 权利要求13所述的重组宿主细胞或细胞培养物,其中所述的脂肪醇组合物包括C10:1、C12:1、C14:1、C16:1或C18:1不饱和脂肪醇。
16. 权利要求13所述的重组宿主细胞或细胞培养物,其中所述的脂肪醇组合物包括C12和C14脂肪醇。
17. 权利要求16所述的重组宿主细胞或细胞培养物,其中所述的脂肪醇组合物包括比例为3:1的C12和C14脂肪醇。
18. 权利要求13所述的重组宿主细胞或细胞培养物,其中所述的脂肪醇组合物包括不饱和脂肪醇或饱和脂肪醇。
19. 权利要求18所述的重组宿主细胞或细胞培养物,其中所述的脂肪醇组合物包括在由所述的脂肪醇的还原末端的C7和C8之间、在碳链的位置7处具有双键的不饱和脂肪醇。
20. 权利要求13所述的重组宿主细胞或细胞培养物,其中所述的脂肪醇组合物包括支链脂肪醇。
21. 一种以高效价、产率或生产率来制备脂肪醇组合物的方法,其包括以下步骤:
 - (a) 改造根据权利要求2-8的任意一项所述的重组宿主细胞;
 - (b) 在包括碳源的培养基中培养所述的重组宿主细胞;以及
 - (c) 可任选地由所述的培养基中分离所述的脂肪醇组合物。

CAR酶及脂肪醇的改良生产

[0001] 相关申请的交叉引用

[0002] 本申请要求2012年4月2日提交的美国临时申请号61/619,309的权益,该申请以引用方式并入本文。

[0003] 序列表

[0004] 本发明包括序列表,该序列表通过EFS-Web以ASCII形式提交,并且全文以引用方式并入本文。在2013年4月2日创建的所述的ASCII拷贝命名为LS00039PCT_SL.txt,并且大小为89,038字节。

技术领域

[0005] 本发明公开涉及在重组宿主细胞中改良生产脂肪醇的变体羧酸还原酶(CAR)。本发明公开进一步涉及变体CAR核酸和多肽以及重组宿主细胞和细胞培养物。本发明公开进一步涵盖了制作脂肪醇组合物的方法。

背景技术

[0006] 脂肪醇构成了工业生化制品的重要类别。这些分子及其衍生物具有多种用途,包括表面活性剂、润滑剂、增塑剂、溶剂、乳化剂、软化剂、增稠剂、风味剂、香料和燃料。在工业上,脂肪醇是通过脂肪酸(其由天然来源产生,该天然来源例如椰子油、棕榈油、棕榈坚果油、牛脂和猪油)的催化氢化或者通过 α -烯烃(由石油化学品供料产生)的化学水合而生产。由天然来源衍生的脂肪醇具有可变的链长。脂肪醇的链长对于具体应用而言是重要的。实际上,脂肪醇还通过酶制得,其中所述的酶能够将酰基-ACP或酰基-CoA分子还原成相应的伯醇(例如参见美国专利公开号20100105955、20100105963和20110250663,这些文献以引用方式并入本文)。

[0007] 目前用于生产脂肪醇的技术涉及脂肪酸的无机催化剂介导的还原成相应的伯醇,这是昂贵、费时且繁冗的。在这种方法中使用的脂肪酸衍生自天然来源(例如上述植物和动物油、以及脂肪)。此外,还可以通过化学催化来将脂肪醇脱水形成 α -烯烃。然而,该技术是不可再生的,并且涉及高的操作成本和具有环境危害的化学废品。因此,需要改良的脂肪醇的生产方法,并且本发明公开解决了该需要。

[0008] 发明概述

[0009] 本发明公开的一个方面提供了包括如下氨基酸序列的变体羧酸还原酶(CAR)多肽,其中所述的氨基酸序列与SEQ ID NO:7具有至少大约90%的序列一致性,其中变体CAR多肽是经过基因改造的,从而在如下氨基酸位置处具有至少一个突变,其中所述的氨基酸位置选自氨基酸位置3、18、20、22、80、87、191、288、473、535、750、827、870、873、926、927、930和1128。因此,变体CAR多肽在重组宿主细胞中的表达得到比表达相应的野生型多肽的重组宿主细胞更高的脂肪醇组合物的效价。在相关的方面中,CAR多肽为CarB多肽。在另一个相关的方面中,变体CAR多肽包括在位置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和/或L1128W中的突变。在相关的方面中，CAR多肽包括：突变A535S；或突变E20R、F288G、Q473I和A535S；或突变E20R、F288G、Q473H、A535S、R827A和S927G；或突变E20R、S22R、F288G、Q473H、A535S、R827A和S927G；或突变S3R、E20R、S22R、F288G、Q473H、A535S、R873S、S927G、M930R和L1128W；或E20R、S22R、F288G、Q473H、A535S、R873S、S927G、M930R和L1128W；或突变D18R、E20R、S22R、F288G、Q473I、A535S、S927G、M930K和L1128W；或突变E20R、S22R、F288G、Q473I、A535S、R827C、V926E、S927K和M930R；或突变D18R、E20R、288G、Q473I、A535S、R827C、V926E、M930K和L1128W；或突变E20R、S22R、F288G、Q473H、A535S、R827C、V926A、S927K和M930R；或突变E20R、S22R、F288G、Q473H、A535S和R827C；或突变E20R、S22R、F288G、Q473I、A535S、R827C和M930R；或突变E20R、S22R、F288G、Q473I、A535S、I870L、S927G和M930R；或突变E20R、S22R、F288G、Q473I、A535S、I870L和S927G；或突变D18R、E20R、S22R、F288G、Q473I、A535S、R827C、I870L、V926A和S927G；或突变E20R、S22R、F288G、Q473H、A535S、R827C、I870L和L1128W；或突变D18R、E20R、S22R、F288G、Q473H、A535S、R827C、I870L、S927G和L1128W；或突变E20R、S22R、F288G、Q473I、A535S、R827C、I870L、S927G和L1128W；或突变E20R、S22R、F288G、Q473I、A535S、R827C、I870L、S927G、M930K和L1128W；或突变E20R、S22R、F288G、Q473H、A535S、I870L、S927G和M930K；或突变E20R、F288G、Q473I、A535S、I870L、M930K；或突变E20R、S22R、F288G、Q473H、A535S、S927G、M930K和L1128W；或突变D18R、E20R、S22R、F288G、Q473I、A535S、S927G和L1128W；或突变E20R、S22R、F288G、Q473I、A535S、、R827C、I870L和S927G；或突变D18R、E20R、S22R、F288G、Q473I、A535S、R827C、I870L、S927G和L1128W；或突变D18R、E20R、S22R、F288G、Q473I、A535S、S927G、M930R和L1128W；或突变E20R、S22R、F288G、Q473H、A535S、V926E、S927G和M930R；或突变E20R、S22R、F288G、Q473H、A535S、R827C、I870L、V926A和L1128W；或它们的组合。

[0010] 本发明公开的另一个方面提供了包括如下多核苷酸序列的宿主细胞，其中所述的多核苷酸序列编码变体羧酸还原酶 (CAR) 多肽，该变体羧酸还原酶 (CAR) 多肽与SEQ ID NO: 7具有至少大约90%的序列一致性，并且在如下氨基酸位置处具有至少一个突变，其中所述的氨基酸位置包括氨基酸位置3、18、20、22、80、87、191、288、473、535、750、827、870、873、926、927、930和1128，其中当在有效表达变体CAR多肽的条件下在包括碳源的培养基中培养时，所述的基因改造的宿主细胞以比表达相应的野生型CAR多肽的宿主细胞更高的效价或产率来生产脂肪醇组合物，并且其中所述的SEQ ID NO: 7为相应的野生型CAR多肽。在相关的方面中，重组宿主细胞进一步包括编码硫酯酶多肽的多核苷酸。在另一个相关的方面中，重组宿主细胞进一步包括编码FabB多肽和FadR多肽的多核苷酸。在另一个相关的方面中，本发明公开提供了包括编码脂肪醛还原酶 (A1rA) 的多核苷酸的重组宿主细胞及包括该细胞的细胞培养物。

[0011] 本发明公开的另一个方面提供了重组宿主细胞，其中当表达相应的野生型CAR多肽的宿主细胞在与所述的基因改造的宿主细胞相同的条件下培养时，所述的基因改造的宿主细胞的效价是表达相应的野生型CAR多肽的宿主细胞效价的至少3倍。在一个相关的方面中，基因改造的宿主细胞的效价为大约30g/L至大约250g/L。在另一个相关的方面中，基因改造的宿主细胞的效价为大约90g/L至大约120g/L。

[0012] 本发明公开的另一个方面提供了重组宿主细胞，其中当表达相应的野生型CAR多

肽的宿主细胞在与所述的基因改造的宿主细胞相同的条件下培养时,所述的基因改造的宿主细胞的产率是表达相应的野生型CAR多肽的宿主细胞产率的至少3倍。在一个相关的方面中,基因改造的宿主细胞的产率为大约10%至大约40%。

[0013] 本发明公开进一步涵盖包括本发明所述的重组宿主细胞的细胞培养物。在相关的方面中,细胞培养物的生产率是表达相应的野生型CAR多肽的细胞培养物的生产率的至少大约3倍。在另一个相关的方面中,生产率为大约0.7mg/L/hr至大约3g/L/hr。在另一个相关的方面中,培养物培养基包括脂肪醇组合物。脂肪醇组合物是在细胞外生产的。脂肪醇组合物可以包括C6、C8、C10、C12、C13、C14、C15、C16、C17或C18脂肪醇;或者C10:1、C12:1、C14:1、C16:1和C18:1不饱和脂肪醇中的一种或多种。在另一个相关的方面中,脂肪醇组合物包括C12和C14脂肪醇。在另一个相关的方面中,脂肪醇组合物包括比例为大约3:1的C12和C14脂肪醇。在另一个相关的方面中,脂肪醇组合物涵盖了不饱和的脂肪醇。此外,脂肪醇组合物可以包括在由脂肪醇的还原末端的C7和C8之间、在碳链的位置7处具有双键的脂肪醇。在另一个方面中,脂肪醇组合物包括饱和的脂肪醇。在另一个方面中,脂肪醇组合物包括支链脂肪醇。

[0014] 本发明公开进一步考虑以更高的效价、产率或生产率来制作脂肪醇组合物的方法,其包括以下步骤:改造重组宿主细胞;在包括碳源的培养基中培养重组宿主细胞;以及可任选地由培养基中分离脂肪醇组合物。

[0015] 附图简述

[0016] 当联合附图一起阅读本发明公开时,可以更好地理解,其中所述的附图起到示意优选的实施方案的作用。此外,应该理解的是本发明公开不限于附图中公开的具体的实施方案。

[0017] 图1为在重组宿主细胞中以酰基CoA作为前体来生产脂肪酸衍生物的示例性生物合成途径的概况。该循环由丙二酰基-ACP和乙酰基-CoA的缩合来引发。

[0018] 图2为示例性脂肪酸生物合成循环的概况,其中通过将丙二酰基-CoA转酰基而形成丙二酰基-ACP来生产丙二酰基-ACP(通过丙二酰基-CoA:ACP酰基转移酶(fabD)的催化);然后, β -酮脂酰基-ACP合成酶III(fabH)引发丙二酰基-ACP和乙酰基-CoA的缩合。延长循环开始于 β -酮脂酰基-ACP合成酶I(fabB)和 β -酮脂酰基-ACP合成酶II(fabF)催化的丙二酰基-ACP与酰基-ACP的缩合,从而生产 β -酮-酰基-ACP,该 β -酮-酰基-ACP接着被NADPH依赖的 β -酮脂酰基-ACP还原酶(fabG)还原,从而生产 β -羟基-酰基-ACP,其通过 β -羟基酰基-ACP脱水酶(fabA或fabZ)脱水形成反-2-烯酰基-酰基-ACP。FabA还可以使反-2-烯酰基-酰基-ACP异构化形成顺-3-烯酰基-酰基-ACP,其可以绕过fabI,并可以被fabB利用(脂肪族链长通常至多为C16),从而生产 β -酮-酰基-ACP。各循环中的最后步骤是由NADH或NADPH依赖的烯酰基-ACP还原酶(fabI)催化的,其将反-2-烯酰基-酰基-ACP转化为酰基-ACP。在上文所述的方法中,脂肪酸合成的终止是通过硫酯酶将酰基由酰基-ACP上移除,从而释放游离的脂肪酸(FFA)来进行的。硫酯酶(例如tesA)水解硫酯键,这是通过硫氢基键在酰基链与ACP之间进行的。

[0019] 图3示出了乙酰基-CoA羧化酶(accABCD)的酶复合物的结构和功能。生物素羧化酶由accC基因编码,而生物素羧基载体蛋白质(BCCP)由accB基因编码。这两个与羧基转移酶活性有关的亚基由accA和accD基因编码。BCCP共价键合的生物素携带羧酸部分。birA基因

(图中未示出)将全-accB生物素化。

[0020] 图4提出了以酰基-ACP为起始来生产脂肪醇的示例性生物合成途径的概况,其中脂肪醛的生产是由酰基-ACP还原酶(AAR)或硫酯酶、以及羧酸还原酶(Car)的酶活性来催化的。脂肪醛通过醛还原酶(也称为醇脱氢酶)转化为脂肪醇。该途径不包括脂肪酰基CoA合成酶(fadD)。

[0021] 图5示出了与由大肠杆菌MG1655所进行的脂肪酸衍生物的生产相比,具有fadE基因衰减(即,删除)的MG1655大肠杆菌菌株所进行的脂肪酸衍生物(总脂肪物质)的生产。图5提供的数据显示fadE基因衰减不会影响脂肪酸衍生物的生产。

[0022] 图6A和6B示出了当质粒pCL-WT TRC WT TesA转化至如图所示的各菌株中时,并在FA2培养基中发酵20小时,从而在32℃(图6A)和37℃(图6B)完成诱导至收获,由2个重复的平板筛选得到的“总脂肪物质”的生产数据。

[0023] 图7A和7B提供了iFAB138基因座的图示,其包括整合在FAB138前面的cat-loxP-T5启动子的示意图(图7A);以及iT5_138的示意图(图7B)。整合在FAB138前面的cat-loxP-T5启动子区的序列(其在cat-loxP-T5启动子区的每一侧上显示出50个碱基对的同源性)在SEQ ID N0:1中提供,而iT5_138启动子区的序列(其在每一侧上具有50个碱基对的同源性)在SEQ ID N0:2中提供。

[0024] 图8示出了校正rph和ilvG基因的作用。使用由D191获得的pCL-tesA(包括PTRC-tesA的pCL1920质粒)转化EG149(rph-ilvG-)和V668(EG149rph+ilvG+)。附图示出了在EG149菌株中校正rph和ilvG基因允许以比V668菌株更高水平地生产FFA,其中所述的V668菌株的rph和ilvG基因是未校正的。

[0025] 图9为转座子盒插入到菌株LC535的yijP基因(转座子匹配物68F11)中的示意图。其中示出位于转座子盒内部的启动子,并且该启动子可以影响相邻的基因的表达。

[0026] 图10示出了在菌株V324中通过CarB60将游离脂肪酸转化为脂肪醇。附图示出了与CarB(浅柱)相比,表达CarB60(得自染色体)的细胞(深柱)将更高级份的C12和C14游离脂肪酸转化成脂肪醇。

[0027] 图11示出了与CarB相比,表达CarB60(得自染色体)的细胞将更高级份的C12和C14游离脂肪酸转化成脂肪醇。

[0028] 图12示出了组合文库的突变体在发酵之后的脂肪醇生产。

[0029] 图13示出了carB变体在生产型质粒(carB1和CarB2)中在摇瓶发酵之后的脂肪醇生产。

[0030] 图14示出了单一拷贝的整合的carB变体(icarB1、icarB2、icarB3和icarB4)在摇瓶发酵之后的脂肪醇生产。

[0031] 图15示出了双质粒筛选系统的结果,其用于通过摇瓶发酵证实被改良的CarB变体。

[0032] 图16示出了用于在生物反应器中改良地生产脂肪醇的新的CarB变体。

[0033] 发明详述

[0034] 综述

[0035] 本发明公开提供了新的变体羧酸还原酶(CAR)及其核酸和蛋白质序列。本发明公开进一步涵盖了重组宿主细胞和细胞培养物,其包括用于生产脂肪醇的变体CAR酶。为了商

业上可利用地由可发酵糖或生物质生产脂肪醇,必须优化所述的方法以用于有效地转化和回收产物。本发明公开通过使用改造的变体酶和改造的重组宿主细胞来提供改良地生产脂肪醇的组合物和方法而解决了这一需要。宿主细胞作为生物催化剂,从而使用发酵方法高效价地生产脂肪醇。依此,本发明公开进一步提供了创建光合异养宿主细胞的方法,其中所述的光合异养宿主细胞直接生产特定链长的脂肪醇和 α -烯烃,这样可不必催化转化纯化的脂肪酸。这种新方法提供了产品质量和成本的优点。

[0036] 更具体而言,可以通过修饰一种或多种基因(与脂肪醇的生产、降解和/或分泌的生物合成途径有关)的表达来增强所需的脂肪醇组合物的生产。本发明公开提供了重组宿主细胞,其经过改造从而提供相对未改造的或天然的宿主细胞而言增强的脂肪醇生物合成(例如菌株改良)。此外,本发明公开提供了用于本发明公开的重组宿主细胞、方法和组合物的多核苷酸。然而,公认的是,与此类多核苷酸绝对一致的序列不是必须的。例如可以对特定多核苷酸序列进行改变,并针对活性评估编码的多肽。这种改变通常包括保守突变和沉默突变(例如密码子优化)。可以使用本领域已知的方法,针对所需的功能筛选修饰的或突变的多核苷酸(即,突变体)、以及编码的变体多肽,其中所述的功能例如为相比亲代多肽的改良的功能,包括但不限于增强的催化活性、增强的稳定性或降低的抑制作用(例如降低的反馈抑制作用)。

[0037] 本发明公开鉴定了根据酶分类(EC)号、与本发明所述的脂肪酸生物合成途径的多个步骤(即,反应)有关的酶活性,并提供通过所述的EC号分类的示例性多肽(即,酶)、以及编码此类多肽的示例性多核苷酸。此类示例性多肽和多核苷酸(其通过编号和/或序列识别号(SEQ ID NO)识别)可用于改造亲代宿主细胞中的脂肪酸途径,从而获得本发明所述的重组宿主细胞。而且,应该理解的是本发明所述的多肽和多核苷酸是示例性的并且是非限定的。本领域的那些技术人员可以使用多个数据库获得本发明所述的示例性多肽的同源序列(例如the National Center for Biotechnology Information (NCBI)提供的Entrez数据库,the Swiss Institute of Bioinformatics提供的ExPasy数据库,the Technical University of Braunschweig提供的BRENDA数据库,以及the Bioinformatics Center of Kyoto University and University of Tokyo提供的KEGG数据库,所有这些均可在the World Wide Web上获得)。

[0038] 多种宿主细胞可以被修饰从而包括脂肪醇生物合成酶(例如本发明所述的那些),从而使重组宿主细胞适于生产脂肪醇组合物。应该理解的是多种细胞可以提供遗传物质的来源,包括编码适用于本发明提供的重组宿主细胞中的多肽的多核苷酸序列。

[0039] 定义

[0040] 除非另外的指示,本发明使用的所有技术和科学术语都具有本发明公开所属技术领域的任一普通技术人员通常所理解的相同的含义。尽管与本发明所述的那些相似的或等价的其他方法和材料可以用于实施本发明公开,但是优选的材料和方法在本发明中进行了描述。在描述和对本发明公开提出要求时,以下术语根据下文列出的定义来使用

[0041] 编号:在整个说明书中使用的序列编号得自由NCBI (National Center for Biotechnology Information)提供的数据库(其由the National Institutes of Health, U.S.A.维护)(该编号在本发明中确认为“NCBI编号”或备选地确认为“GenBank编号”),以及得自the Swiss Institute of Bioinformatics提供的UniProt Knowledgebase

(UniProtKB) and Swiss-Prot数据库(该编号在本发明中确认为“UniProtKB编号”)。

[0042] 酶分类(EC)号:EC号由the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology(IUBMB)建立,其描述可以在国际互联网的IUBMB Enzyme Nomenclature网站上获得。EC号是根据所催化的反应将酶分类。

[0043] 如本文所用,术语“核苷酸”是指多核苷酸的单体单元,其由杂环碱基、糖和一个或多个磷酸基构成。天然形成的碱基(鸟嘌呤(G)、腺嘌呤(A)、胞嘧啶(C)、胸腺嘧啶(T)和尿嘧啶(U))通常为嘌呤或嘧啶的衍生物,但是应该理解的是天然和非天然形成的碱基类似物也被包括在内。天然形成的糖为戊糖(五碳糖)脱氧核糖(其形成DNA)或核糖(其形成RNA),但是应该理解的是天然和非天然形成的糖类类似物也被包括在内。核酸通常通过磷酸键连接,从而形成核酸或多核苷酸,但是本领域已知许多其他的键(例如磷硫酰键、硼烷磷酸键等)。

[0044] 如本文所用,术语“多核苷酸”是指核苷酸(RNA)和脱氧核苷酸(DNA)的聚合物,其可以是单链的或双链的,并且其可以包括非天然或改变的核苷酸。术语“多核苷酸”、“核酸序列”和“核苷酸序列”在本发明中可以交换使用,其是指任何长度的核苷酸(RNA或DNA)的聚合形式。这些术语是指分子的一级结构,因此包括双链和单链DNA、以及双链和单链RNA。该术语包括由核苷酸类似物和修饰多核苷酸(例如但不限于甲基化的和/或加帽的多核苷酸)形成的RNA或DNA的类似物作为等价物。多核苷酸可以为任何形式,包括但不限于质粒、病毒、染色体、EST、cDNA、mRNA和rRNA。

[0045] 如本文所用,术语“多肽”和“蛋白质”可以交换使用,其是指氨基酸残基的聚合物。术语“重组多肽”是指通过重组技术生产的多肽,其中通常,将编码所表达蛋白质的DNA或RNA插入到合适的表达载体中,该载体进而用于转化宿主细胞,从而生产多肽。

[0046] 如本文所用,术语“同源”和“同源的”是指包括如下序列的多核苷酸或多肽,其中所述的序列与相应的多核苷酸或多肽序列具有至少大约50%的一致性。优选的是,同源的多核苷酸或多肽具有如下多核苷酸序列或氨基酸序列,该多核苷酸序列或氨基酸序列与相应的氨基酸序列或核苷酸序列具有至少大约80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或者至少大约99%的同源性。如本文所用,术语序列“同源性”和序列“一致性”可以交换使用。

[0047] 本领域的任一普通技术人员很好地知道测定两个或多个序列之间的同源性的方法。简言之,计算两个序列之间的“同源性”可以按以下方式进行。为了达到最佳比较的目的,对序列进行比对(例如可以在第一和第二氨基酸或核酸序列的一者或二者中引入空位以达到最佳比对的目的,并且可以忽视非同源的序列以达到比较目的)。在优选的实施方案中,用于比较目的的、比对的第一序列的长度为第二序列的长度的至少大约30%、优选至少大约40%、更优选地至少大约50%、甚至更优选地至少大约60%、甚至更优选地至少大约70%、至少大约80%、至少大约90%或者大约100%。然后,比较在第一和第二序列的相应的氨基酸位置或核苷酸位置处的氨基酸残基或核苷酸。当第一序列的位置被第二序列中相应位置处的相同氨基酸残基或核苷酸占据时,则分子在该位置处是相同的。两个序列之间的百分同源性为序列所共有的相同位置的数量的函数,其考虑了空位的数量和各空位的长度,必须引入空位的数量和各空位的长度以达到两个序列的最佳比对的目的。

[0048] 序列的比较以及两个序列之间的百分同源性的确定可以使用数学算法来完成,例如BLAST(Altschul et al., J. Mol. Biol, 215 (3):403-410 (1990))。此外,两个氨基酸序列

之间的百分同源性还可以使用Needleman和Wunsch算法来确定,其中所述的算法被引入到GCG软件包的GAP程序中,其中使用了Blossum 62矩阵或PAM250矩阵,以及16、14、12、10、8、6或4的空位权重和1、2、3、4、5或6的长度权重(Needleman and Wunsch, J. Mol. Biol., 48: 444-453 (1970))。此外,两个核苷酸序列之间的百分同源性还可以使用GCG软件包中的GAP程序来确定,其中使用了NWSgapdna.CMP矩阵以及40、50、60、70或80的空位权重和1、2、3、4、5或6的长度权重。本领域的任一普通技术人员可以进行最初的同源性计算,并相应地调整算法参数。优选的参数组(和如果操作者不确定应当应用哪些参数来确定分子是否在所要求的同源性限制内而应该使用的参数组)是具有12的空位罚分、4的空位延伸罚分和5的移码空位罚分的Blossum 62评分矩阵。序列比对的其他方法是生物技术领域已知的(例如参见Rosenberg, BMC Bioinformatics, 6:278 (2005); Altschul, et al., FEBS J., 272 (20): 5101-5109 (2005))。

[0049] 如本文所用,术语“在低严谨性、中严谨性、高严谨性或极高严谨性的条件下杂交”描述了用于杂交和洗涤的条件。实施杂交反应的指导可以见于Current Protocols in Molecular Biology, John Wiley&Sons, N.Y. (1989), 6.3.1-6.3.6。该参考文献中描述了水性和非水性方法并且可以使用任一种方法。本发明所指的具体的杂交条件如下:1) 低严谨性杂交条件—在大约45℃下、在6x氯化钠/柠檬酸钠(SSC)中,然后在至少50℃下、在0.2xSSC、0.1%SDS中洗涤两次(对于低严谨性条件,洗涤温度可以升至55℃);2) 中严谨性杂交条件—在大约45℃下、在6xSSC中,然后在60℃的0.2xSSC、0.1%SDS中洗涤一次或多次;3) 高严谨性杂交条件—在大约45℃的6xSSC中,然后在65℃下、在0.2xSSC、0.1%SDS中洗涤一次或多次;以及4) 极高严谨性杂交条件—在65℃的0.5M磷酸钠、7%SDS中,然后在65℃下、在0.2xSSC、1%SDS中洗涤一次或多次。除非另外指示,极高的严谨性条件(4)是优选的条件。

[0050] “内源的”多肽是指由亲代微生物细胞(也称为宿主细胞)的基因组编码的多肽,其中重组细胞由所述的亲代微生物细胞改造(或“衍生”)得到。

[0051] “外源的”多肽是指并非由亲代微生物细胞的基因组编码的多肽。变体(即,突变体)多肽为外源多肽的实例。

[0052] 术语“异源的”通常是指由不同的物种衍生得到或有不同的有机体衍生得到。如本文所用,异源的是指并非天然存在于特定的有机体中的核苷酸序列或多肽序列。异源表达是指将蛋白质或多肽凭经验加入到通常不表达该蛋白质的细胞中。依此,异源的是指这样的事实:被转移的蛋白质最初由与受体不同类型的细胞或不同的物种衍生得到。例如可以通过重组方法将与植物细胞内源的多核苷酸序列引入到细菌宿主细胞中,然后该植物多核苷酸在重组细菌宿主细胞中为异源多核苷酸。

[0053] 如本文所用,术语多肽的“片段”是指全长多肽或蛋白质的较短的部分,其大小为4个氨基酸残基甚至整个氨基酸序列减去1个氨基酸残基。在本发明公开的某些实施方案中,片段是指多肽或蛋白质的结构域(例如底物结合结构域或催化结构域)的整个氨基酸序列。

[0054] 如本文所用,术语“诱变”是指一种方法,通过该方法有机体的遗传信息以稳定的方式被改变。编码蛋白质的核酸序列的诱变产生突变体蛋白质。此外,诱变还指非编码核酸序列的改变,而这种改变会导致修饰的蛋白质活性。

[0055] 如本文所用,术语“基因”是指编码RNA产物或蛋白质产物的核酸序列;以及影响

RNA或蛋白质的表达的可操作地连接的核酸序列(例如此类序列包括但不限于启动子或增强子序列)或编码影响RNA或蛋白质的表达的序列的、可操作地连接的核酸序列(例如此类序列包括但不限于核糖体结合位点或翻译控制序列)。

[0056] 表达控制序列是本领域已知的,并且包括例如启动子、增强子、多聚腺苷酸信号、转录终止子、内部核糖体进入位点(IRES)等,它们提供了多核苷酸序列在宿主细胞中的表达。表达控制序列特异性地与转录有关的细胞蛋白质相互作用(Maniatis et al., Science, 236:1237-1245 (1987))。示例性的表达控制序列如例如Goeddel, Gene Expression Technology: Methods in Enzymology, Vol. 185, Academic Press, San Diego, Calif. (1990)中所述。

[0057] 在本发明公开的方法中,表达控制序列与多核苷酸序列可操作地连接。“可操作地连接”是指多核苷酸序列和表达控制序列以这样的方式连接,使得在合适的分子(例如转录活化蛋白质)与表达控制序列结合时,允许基因表达。根据转录和翻译的方向,可操作地连接的启动子位于所选的多核苷酸序列的上游。可操作地连接的增强子可以位于所选的多核苷酸的上游、内部或下游。

[0058] 如本文所用,术语“载体”是指能够运送所连接的另一种核酸(即,多核苷酸序列)的核酸分子。一种类型的有用的载体为附加体(即,能够在染色体外复制的核酸)。有用的载体为能够使所连接的核酸自主复制和/或表达的那些。能够使可操作地连接的基因定向表达的载体在本发明中还称为“表达载体”。总体而言,用于重组DNA技术中的表达载体通常为“质粒”形式,其通常是指环状双链的DNA环,该DNA环作为载体形式不能与染色体结合。术语“质粒”和“载体”在本发明中可以交换使用,因为质粒为最常用的载体形式。然而,还包括此类其他形式的表达载体,这些载体发挥着等价的功能并且到目前为止是本领域已知的。在一些实施方案中,重组载体包括至少一个序列,其包括(a)与多核苷酸序列可操作地缀合的表达控制序列;(b)与多核苷酸序列可操作地缀合的选择标记;(c)与多核苷酸序列可操作地缀合的标记序列;(d)与多核苷酸序列可操作地缀合的纯化部分;(e)与多核苷酸序列可操作地缀合的分泌序列;以及(f)与多核苷酸序列可操作地缀合的靶向序列。本发明所述的表达载体包括以适用于在宿主细胞中表达多核苷酸序列的形式的、本发明所述的多核苷酸序列。本领域的技术人员应该理解的是表达载体的设计可以取决于诸如待转化的宿主细胞的选择、所需多肽的表达水平之类的因素。本发明所述的表达载体可以被引入到宿主细胞中,从而生产由本发明所述的多核苷酸序列编码的多肽,包括融合多肽。

[0059] 在原核生物(例如大肠杆菌)中表达编码多肽的基因最通常的是使用包括构成型或诱导型启动子的载体来实施,其中所述的启动子指导融合或非融合多肽的表达。融合载体将一定数量的氨基酸加入到其中编码的多肽中,通常加入到重组多肽的氨基末端或羧基末端。这种融合载体通常起到以下3种目的中的一种或多种:(1)增加重组多肽的表达;(2)增加重组多肽的溶解性;以及(3)通过在亲和纯化中起到配体的作用而有助于重组多肽的纯化。通常,在融合表达的载体中,在融合部分与重组多肽的连接处引入蛋白水解的切割位点。这使得在纯化融合多肽后,重组多肽与融合部分分离。在某些实施方案中,本发明公开的多核苷酸序列与由噬菌体T5衍生的启动子可操作地连接。在某些实施方案中,宿主细胞为酵母细胞,表达载体为酵母表达载体。用于在酿酒酵母(*S. cerevisiae*)中表达的载体的实例包括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) 和picZ (Invitrogen Corp., San Diego, CA)。在其他的实施方案中, 宿主细胞为昆虫细胞, 表达载体为杆状病毒表达载体。可用于在培养的昆虫细胞 (例如Sf9细胞) 中表达蛋白质的杆状病毒载体包括例如pAc系列 (Smith et al., Mol. Cell Biol., 3:2156-2165 (1983)) 和pVL系列 (Lucklow et al., Virology, 170:31-39 (1989))。在另一个实施方案中, 本发明所述的多核苷酸序列可以使用哺乳动物表达载体在哺乳动物细胞中表达。用于原核和真核细胞的其他合适的表达系统是本领域公知的, 例如参见Sambrook et al., "Molecular Cloning: A Laboratory Manual," 再版, Cold Spring Harbor Laboratory, (1989)。

[0060] 如本文所用, “酰基-CoA”是指在烷基链的羰基碳与辅酶A (CoA) 的4'-磷酸泛酰亚硫酰基部分的硫氢基之间形成酰基硫酯, 其具有下式 $R-C(=O)-S-CoA$, 其中R为具有至少4个碳原子的任何烷基。

[0061] 如本文所用, “酰基-ACP”是指在烷基链的羰基碳与酰基载体蛋白质 (ACP) 的磷酸泛酰硫基乙胺基部分的硫氢基之间形成酰基硫酯。磷酸泛酰硫基乙胺基部分在翻译后通过全-酰基载体蛋白质合成酶 (ACPS) (一种磷酸泛酰硫基乙胺基转移酶) 的作用与ACP上的保守丝氨酸残基连接。在一些实施方案中, 酰基-ACP为完全饱和的酰基-ACP的合成中的中间体。在其他的实施方案中, 酰基-ACP为不饱和的酰基-ACP的合成中的中间体。在一些实施方案中, 碳链具有大约5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25或26个碳。这些酰基-ACP的每一种都是将它们转化为脂肪酸衍生物的酶的底物。

[0062] 如本文所用, 术语“脂肪酸或其衍生物”是指“脂肪酸”或“脂肪酸衍生物”。术语“脂肪酸”是指具有式 $RCOOH$ 的羧酸。R表示脂肪族基团, 优选为烷基。R可以包括大约4至大约22个碳原子。脂肪酸可以是饱和的、单不饱和的或多不饱和的。在优选的实施方案中, 脂肪酸由脂肪酸生物合成途径制得。“脂肪酸衍生物”是指部分由生产型宿主有机体的脂肪酸生物合成途径形成的产物。此外, “脂肪酸衍生物”还包括部分由酰基-ACP或酰基-ACP衍生物形成的产物。示例性的脂肪酸衍生物包括例如酰基-CoA、脂肪醛、短链和长链的醇、碳氢化合物、酯 (例如蜡、脂肪酸酯或脂肪酯)。

[0063] 如本文所用, 术语“脂肪酸生物合成途径”是指生产脂肪酸衍生物 (例如脂肪醇) 的生物合成途径。脂肪酸生物合成途径包括可以被改造从而生产脂肪酸的脂肪酸合成酶, 并且在一些实施方案中, 这种脂肪酸合成酶可以与其他酶一起表达, 从而生产脂肪酸衍生物, 例如具有所需特征的脂肪醇。

[0064] 如本文所用, “脂肪醛”是指具有式 $RCHO$ 的醛, 其特征在于羰基 ($C=O$)。在一些实施方案中, 脂肪醛为由脂肪醇形成的任何的醛。在某些实施方案中, R基的长度为至少5、至少6、至少7、至少8、至少9、至少10、至少11、至少12、至少13、至少14、至少15、至少16、至少17、至少18或者至少19个碳。备选地或此外, R基的长度为20或更少、19或更少、18或更少、17或更少、16或更少、15或更少、14或更少、13或更少、12或更少、11或更少、10或更少、9或更少、8或更少、7或更少、6或更少的碳。因此, R基可以具有上述任何两个端点限定的R基。例如R基可以为6-16个碳的长度、10-14个碳的长度或者12-18个碳的长度。在一些实施方案中, 脂肪醛为 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} 或 C_{26} 脂肪醛。在某些实施方案中, 脂肪醛为 C_6 、 C_7 、 C_8 、 C_9 、 C_{10} 、 C_{11} 、 C_{12} 、 C_{13} 、 C_{14} 、 C_{15} 、 C_{16} 、 C_{17} 或 C_{18} 脂肪醛。

[0065] 如本文所用,“脂肪醇”是指具有式ROH的醇。在一些实施方案中,R基的长度为至少5、至少6、至少7、至少8、至少9、至少10、至少11、至少12、至少13、至少14、至少15、至少16、至少17、至少18或者至少19个碳。备选地或此外,R基的长度为20或更少、19或更少、18或更少、17或更少、16或更少、15或更少、14或更少、13或更少、12或更少、11或更少、10或更少、9或更少、8或更少、7或更少、6或更少的碳。因此,R基可以具有上述任何两个端点限定的R基。例如R基可以为6-16个碳的长度、10-14个碳的长度或者12-18个碳的长度。在一些实施方案中,脂肪醇为C₆、C₇、C₈、C₉、C₁₀、C₁₁、C₁₂、C₁₃、C₁₄、C₁₅、C₁₆、C₁₇、C₁₈、C₁₉、C₂₀、C₂₁、C₂₂、C₂₃、C₂₄、C₂₅或C₂₆脂肪醇。在某些实施方案中,脂肪醇为C₆、C₈、C₁₀、C₁₂、C₁₃、C₁₄、C₁₅、C₁₆、C₁₇或C₁₈脂肪醇。

[0066] “脂肪醇组合物”在本发明中是由重组宿主细胞生产的,并且通常包括脂肪醇的混合物。在一些情况下,所述的混合物包括多于一种类型的产物(例如脂肪醇和脂肪酸)。在其他的情况下,脂肪酸衍生物组合物可以包括例如具有不同链长以及饱和或支化特征的脂肪醇的混合物,在其他的情况下,脂肪醇组合物包括多于一种类型的产物以及具有多种链长及饱和或支化特征的产物的混合物。

[0067] 经改造从而生产脂肪醛的宿主细胞通常将一些脂肪醛转化成脂肪醇。当生产脂肪醇的宿主细胞经改造从而表达编码酯合成酶的多核苷酸时,蜡酯被生产。在一个实施方案中,脂肪醇由脂肪酸生物合成途径制得。一个实例为酰基-ACP可以通过硫酯酶(例如大肠杆菌TesA)的作用而被转化成脂肪酸,该脂肪酸通过羧酸还原酶(例如大肠杆菌CarB)的作用而被转化成脂肪醛和脂肪醇。脂肪醛向脂肪醇的转化可以通过例如脂肪醇生物合成多肽的作用而进一步被促进。在一些实施方案中,编码脂肪醇生物合成多肽的基因在宿主细胞中被表达或共表达。在某些实施方案中,脂肪醇生物合成多肽具有醛还原酶或醇脱氢酶活性。根据本发明公开使用的醇脱氢酶多肽的实例包括但不限于不动杆菌(*Acinetobacter*)菌种M-1 (SEQ ID NO:3)的AlrA或AlrA同源物(例如AlrAadp1 (SEQ ID NO:4)),以及内源大肠杆菌醇脱氢酶,例如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)和Ybb0 [AAC73595.1]。其他的实例在国际专利申请公开号W0 2007/136762、W02008/119082和W02010/062480中有所描述,这些文献均明确地以引用方式并入本文。在某些实施方案中,脂肪醇生物合成多肽具有醛还原酶或醇脱氢酶活性(EC 1.1.1.1)。

[0068] 如本文所用,术语“醇脱氢酶”是指能够催化脂肪醛转化成醇(例如脂肪醇)的多肽。本领域的任一普通技术人员将理解的是某些醇脱氢酶还能够催化其他的反应,并且这些非特异性的醇脱氢酶也被涵盖在“醇脱氢酶”的范围内。脂肪酸、脂肪醛或脂肪醇的R基可以是直连或支链。支链可以具有多于一个的分支点,并且可以包括环状支。在一些实施方案中,支化的脂肪酸、支化的脂肪醛或支化的脂肪醇为C₆、C₇、C₈、C₉、C₁₀、C₁₁、C₁₂、C₁₃、C₁₄、C₁₅、C₁₆、C₁₇、C₁₈、C₁₉、C₂₀、C₂₁、C₂₂、C₂₃、C₂₄、C₂₅或C₂₆的支化的脂肪酸、支化的脂肪醛或支化的脂肪醇。在具体的实施方案在,支化的脂肪酸、支化的脂肪醛或支化的脂肪醇为C₆、C₈、C₁₀、C₁₂、C₁₃、C₁₄、C₁₅、C₁₆、C₁₇或C₁₈的支化的脂肪酸、支化的脂肪醛或支化的脂肪醇。在某些实施方案中,支化的脂肪酸、支化的脂肪醛或支化的脂肪醇的羟基在第一(C₁)的位置中。在某些实施方案中,支化的脂肪酸、支化的脂肪醛或支化的脂肪醇为异脂肪酸、异脂肪醛、异脂肪醇、反异脂肪酸、反异脂肪醛或反异脂肪醇。在示例性的实施方案中,支化的脂肪酸、支化的脂肪

醛或支化的脂肪醇选自异C_{7:0}、异C_{8:0}、异C_{9:0}、异C_{10:0}、异C_{11:0}、异C_{12:0}、异C_{13:0}、异C_{14:0}、异C_{15:0}、异C_{16:0}、异C_{17:0}、异C_{18:0}、异C_{19:0}、反异C_{7:0}、反异C_{8:0}、反异C_{9:0}、反异C_{10:0}、反异C_{11:0}、反异C_{12:0}、反异C_{13:0}、反异C_{14:0}、反异C_{15:0}、反异C_{16:0}、反异C_{17:0}、反异C_{18:0}和反异C_{19:0}的支化的脂肪酸、支化的脂肪醛或支化的脂肪醇。支化的或非支化的脂肪酸、支化的或非支化的脂肪醛、或者支化的或非支化的脂肪醇的R基可以是饱和的或不饱和的。如果是不饱和的,则R基可以具有一个或多于一个的不饱和点。在一些实施方案中,不饱和的脂肪酸、不饱和的脂肪醛、或者不饱和的脂肪醇是单不饱和的脂肪酸、单不饱和的脂肪醛、或者单不饱和的脂肪醇。在某些实施方案中,不饱和的脂肪酸、不饱和的脂肪醛、或者不饱和的脂肪醇为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}或C_{26:1}的不饱和的脂肪酸、不饱和的脂肪醛、或者不饱和的脂肪醇。在某些实施方案中,不饱和的脂肪酸、不饱和的脂肪醛、或者不饱和的脂肪醇为C_{10:1}、C_{12:1}、C_{14:1}、C_{16:1}或C_{18:1}。在其他的实施方案中,不饱和的脂肪酸、不饱和的脂肪醛、或者不饱和的脂肪醇在 ω -7位置处是不饱和的。在某些实施方案中,不饱和的脂肪酸、不饱和的脂肪醛、或者不饱和的脂肪醇包括顺式双键。

[0069] 如本文所用,重组的或改造的“宿主细胞”为这样的宿主细胞,例如经过修饰从而可以生产脂肪醇的微生物有机体。在一些实施方案中,重组宿主细胞包括一种或多种多核苷酸,各个多核苷酸编码具有脂肪醛和/或脂肪醇生物合成酶活性的多肽,其中当重组宿主细胞在有效表达多核苷酸的条件下在碳源存在下培养时,生产脂肪醇组合物。

[0070] 如本文所用,术语“克隆”通常是指起源于单一的共同祖先并与该祖先基本上是遗传一致的细胞或一组细胞,例如由单一的细菌细胞形成的克隆细菌菌落的细菌。

[0071] 如本文所用,术语“培养物”通常是指包括活细胞的液体培养基。在一个实施方案中,培养物包括在受控的条件下在预定的培养物培养基中再生产的细胞,例如在包括所选的碳源和氮的液体培养基中生长的重组宿主细胞培养物。动词形式的“培养”或名词形式的“培养”是指使微生物细胞群体在液体或固体培养基中在合适的条件下生长。在具体的实施方案中,培养是指底物以发酵方式生物转化为终产物。培养用的培养基是公知的,并且此类培养物培养基的单个成分可得自商业来源,例如Difco™和BBL™商标。在一个非限定性的实例中,水性营养培养基为包括氮、盐和碳的复合来源的“富集培养基”,例如YP培养基,其包括此类培养基的10g/L蛋白胨以及10g/L酵母提取物。根据美国专利5,000,000;5,028,539;5,424,202;5,482,846;5,602,030和W0 2010127318中所述的方法,宿主细胞可以另外改造,从而高效地同化碳并使用纤维素材料作为碳源,其中所述的每一份文献均明确地以引用方式并入本文。此外,宿主细胞可以被改造从而表达转化酶,这样蔗糖可以被用作碳源。

[0072] 如本文所用,术语“在有效表达所述的异源核苷酸序列的条件下”是指允许宿主细胞生产所需的脂肪醛或脂肪醇的任何条件。合适的条件包括例如发酵条件。

[0073] 如本文所用,在重组宿主细胞中,“修饰的”或“改变水平”的蛋白质(例如酶)活性是指在相对于亲代或天然宿主细胞测定的活性中存在的一个或多个特征的差异。通常,活性的差异是在重组宿主细胞(具有修饰的活性)与相应的野生型宿主细胞之间测定的(例如重组宿主细胞培养物相对于野生型宿主细胞的比较)。修饰的活性可以是以下情况的结果,例如:重组宿主细胞所表达的蛋白质的量被修饰(例如编码蛋白质的DNA序列的拷贝数量升高或降低、编码蛋白质的mRNA转录子的数量升高或降低、和/或由mRNA形成蛋白质的蛋白质

翻译的量升高或降低的结果);蛋白质结构的改变(例如一级结构改变,如对蛋白质的编码序列的改变,其导致底物特异性的改变,所见的动力学参数改变);以及蛋白质的稳定性改变(例如蛋白质降解的升高或降低)。在一些实施方案中,所述的多肽为本发明所述的任一种多肽的突变体或变体。在某些情况下,用于本发明所述的多肽的编码序列为就在特定的宿主细胞中的表达而优化的密码子。例如就在大肠杆菌中的表达而言,可以根据例如 Grosjean et al., Gene 18:199-209 (1982) 中所述优化一个或多个密码子。

[0074] 如本文所用,术语“调节序列”通常是指DNA中碱基的序列,该序列与编码蛋白质的DNA序列可操作地连接,其中所述的碱基序列最终控制蛋白质的表达。调节序列的实例包括但不限于RNA启动子序列、转录因子结合序列、转录终止序列、转录调控因子(例如增强子元件)、影响RNA稳定性的核苷酸序列以及翻译调节序列(例如核糖体结合位点(例如原核生物中的Shine-Dalgarno序列或真核生物中的Kozak序列)、起始密码子、终止密码子)。

[0075] 如本文所用,短语“所述的核苷酸序列的表达相对于野生型核苷酸序列被修饰”是指内源核苷酸序列的表达和/或活性水平、或者异源的或编码多肽的非天然核苷酸序列的表达和/或活性水平的升高或降低。如本文所用,术语“过表达”是指在细胞中表达一定浓度的多核苷酸或多肽或者使多核苷酸或多肽在细胞中以一定的浓度被表达,其中所述的浓度高于在相同条件下由相应的野生型细胞正常表达的浓度。

[0076] 术语“改变的表达水平”和“修饰的表达水平”可以交换使用,并且是指在相同的条件下,与多核苷酸、多肽或碳氢化合物在相应的野生型细胞中的浓度相比,多核苷酸、多肽或碳氢化合物在改造的宿主细胞中以不同的浓度存在。

[0077] 如本文所用,术语“效价”是指每单位体积的宿主细胞培养物生产的脂肪醛或脂肪醇的量。在本发明所述的组合物和方法的任何方面中,脂肪醇以大约25mg/L、大约50mg/L、大约75mg/L、大约100mg/L、大约125mg/L、大约150mg/L、大约175mg/L、大约200mg/L、大约225mg/L、大约250mg/L、大约275mg/L、大约300mg/L、大约325mg/L、大约350mg/L、大约375mg/L、大约400mg/L、大约425mg/L、大约450mg/L、大约475mg/L、大约500mg/L、大约525mg/L、大约550mg/L、大约575mg/L、大约600mg/L、大约625mg/L、大约650mg/L、大约675mg/L、大约700mg/L、大约725mg/L、大约750mg/L、大约775mg/L、大约800mg/L、大约825mg/L、大约850mg/L、大约875mg/L、大约900mg/L、大约925mg/L、大约950mg/L、大约975mg/L、大约1000mg/L、大约1050mg/L、大约1075mg/L、大约1100mg/L、大约1125mg/L、大约1150mg/L、大约1175mg/L、大约1200mg/L、大约1225mg/L、大约1250mg/L、大约1275mg/L、大约1300mg/L、大约1325mg/L、大约1350mg/L、大约1375mg/L、大约1400mg/L、大约1425mg/L、大约1450mg/L、大约1475mg/L、大约1500mg/L、大约1525mg/L、大约1550mg/L、大约1575mg/L、大约1600mg/L、大约1625mg/L、大约1650mg/L、大约1675mg/L、大约1700mg/L、大约1725mg/L、大约1750mg/L、大约1775mg/L、大约1800mg/L、大约1825mg/L、大约1850mg/L、大约1875mg/L、大约1900mg/L、大约1925mg/L、大约1950mg/L、大约1975mg/L、大约2000mg/L (2g/L)、3g/L、5g/L、10g/L、20g/L、30g/L、40g/L、50g/L、60g/L、70g/L、80g/L、90g/L、100g/L或上述任意两个值所限定的范围的效价生产。在其他的实施方案中,脂肪醛或脂肪醇以高于100g/L、高于200g/L、高于300g/L或更高,例如500g/L、700g/L、1000g/L、1200g/L、1500g/L或2000g/L的效价生产。根据本发明公开的方法由重组宿主细胞生产的脂肪醛或脂肪醇的优选效价为5g/L至200g/L、10g/L至150g/L、20g/L至120g/L、以及30g/L至100g/L。

[0078] 如本文所用,“由宿主细胞生产的脂肪醛或脂肪醇的产率”是指输入的碳源在宿主细胞中被转化为产物(即,脂肪醇或脂肪醛)的效率。根据本发明公开的方法经改造从而生产脂肪醇和/或脂肪醛的宿主细胞的产率为至少3%、至少4%、至少5%、至少6%、至少7%、至少8%、至少9%、至少10%、至少11%、至少12%、至少13%、至少14%、至少15%、至少16%、至少17%、至少18%、至少19%、至少20%、至少21%、至少22%、至少23%、至少24%、至少25%、至少26%、至少27%、至少28%、至少29%、至少30%或者由上述任意两个值所限定的范围。在其他的实施方案中,脂肪醛或脂肪醇以高于30%、40%、50%、60%、70%、80%、90%或更高的产率生产。备选地或此外,产率为大约30%或更低、大约27%或更低、大约25%或更低、大约22%或更低。因此,产率可以由上述任意两个端点来限定。例如根据本发明公开的方法由重组宿主细胞生产的脂肪醇或脂肪醛的产率可以为5%至15%、10%至25%、10%至22%、15%至27%、18%至22%、20%至28%、或20%至30%。根据本发明公开的方法由重组宿主细胞生产的脂肪醇的优选产率为10%至30%。

[0079] 如本文所用,术语“生产率”是指每单位时间内每单位体积的宿主细胞培养物生产脂肪醛或脂肪醇的量。在本发明所述的组合物和方法的任何方面中,由重组宿主细胞生产的脂肪醛或脂肪醇的生产率为至少100mg/L/小时、至少200mg/L/小时、至少300mg/L/小时、至少400mg/L/小时、至少500mg/L/小时、至少600mg/L/小时、至少700mg/L/小时、至少800mg/L/小时、至少900mg/L/小时、至少1000mg/L/小时、至少1100mg/L/小时、至少1200mg/L/小时、至少1300mg/L/小时、至少1400mg/L/小时、至少1500mg/L/小时、至少1600mg/L/小时、至少1700mg/L/小时、至少1800mg/L/小时、至少1900mg/L/小时、至少2000mg/L/小时、至少2100mg/L/小时、至少2200mg/L/小时、至少2300mg/L/小时、至少2400mg/L/小时、或至少2500mg/L/小时。备选地或此外,生产率为2500mg/L/小时或更低、2000mg/L/OD₆₀₀或更低、1500mg/L/OD₆₀₀或更低、120mg/L/小时或更低、1000mg/L/小时或更低、800mg/L/小时或更低、或者600mg/L/小时或更低。因此,生产率可以由上述任何两个端点来限定。例如,生产率可以为3至30mg/L/小时、6至20mg/L/小时或15-30mg/L/小时。根据本发明公开所述的方法由重组宿主细胞生产的脂肪醛或脂肪醇的优选生产率选自500mg/L/小时至2500mg/L/小时、或700mg/L/小时至2000mg/L/小时。

[0080] 术语“总脂肪物质”和“总脂肪酸产物”在本发明中可以交换使用,其是指在样品中存在的脂肪醇、脂肪醛、游离脂肪酸和脂肪酯的总量,其可以通过GC-FID来评价,如国际专利申请公开W02008/119082中所述。根据所述的内容,样品可以包括这些化合物中的一种、两种、三种或四种。

[0081] 如本文所用,术语“葡萄糖利用率”是指每单位时间内培养物使用的葡萄糖的量,其报告为克/升/小时(g/L/hr)。

[0082] 如本文所用,术语“碳源”是指其适于用作原核或简单的真核细胞生长所需碳的来源的底物或化合物。碳源可以为多种形式,包括但不限于聚合物、碳水化合物、酸、醇、醛、酮、氨基酸、肽和气体(例如CO和CO₂)。示例性的碳源包括但不限于:单糖,例如葡萄糖、果糖、甘露糖、半乳糖、木糖和阿拉伯糖;寡糖,例如低聚果糖和低聚半乳糖;多糖,例如淀粉、纤维素、胶质和木聚糖;二糖,例如蔗糖、麦芽糖、纤维二糖和松二糖;纤维素材料和变体,例如半纤维素、甲基纤维素和羧甲基纤维素钠;饱和的或不饱和的脂肪酸、琥珀酸盐或酯、乳酸盐或酯和醋酸盐或酯;醇,例如乙醇、甲醇和甘油;或它们的混合物。此外,碳源还可以为

光合成产物,例如葡萄糖。在某些优选的实施方案中,碳源为生物质。在其他优选的实施方案中,碳源为葡萄糖。在其他优选的实施方案中,碳源为蔗糖。

[0083] 如本文所用,术语“生物质”是指可以衍生得到碳源的任何生物材料。在一些实施方案中,适于生物转化的生物质被加工成碳源。在其他的实施方案中,生物质不需要进一步加工成碳源。碳源可以转化为生物燃料。示例性的生物质来源为植物物质或蔬菜,例如谷物、甘蔗或柳枝稷。另一种示例性的生物质来源为代谢废物,例如动物物质(例如牛粪便)。其他的示例性生物质来源包括藻类和其他海生植物。此外,生物质还包括由工业、农业、林业和家庭得到的废物,包括但不限于发酵废物、未干的秣草、稻草、木材、污物、垃圾、纤维素城市废物和食物剩余物。此外,术语“生物质”还指碳源,例如碳水化合物(例如单糖、二糖或多糖)。

[0084] 如本文所用,对产物(例如脂肪酸及其衍生物)而言的术语“分离的”是指由细胞成分、细胞培养物培养基、或者化学或合成的前体分离得到的产物。通过本发明所述的方法生产的脂肪酸及其衍生物在发酵液及细胞质中相对而言可以是不可混合的。因此,所述的脂肪酸及其衍生物可以在细胞内或细胞外收集于有机相中。

[0085] 如本文所用,术语动词性的“纯化”、形容词性的“纯化的”或名词性的“纯化”是指通过例如分离或隔离将分子从其环境中移出或分离。“基本上纯化的”分子为至少大约60%游离于(例如至少大约70%游离于、至少大约75%游离于、至少大约85%游离于、至少大约90%游离于、至少大约95%游离于、至少大约97%游离于、至少大约99%游离于)它们所关联的其他组分。如本文所用,这些术语还指从样品中去除污染物。例如污染物的去除可以使样品中的脂肪醛或脂肪醇的百分率增加。例如当在重组宿主细胞中产生脂肪醛或脂肪醇时,可以通过去除重组宿主细胞蛋白质来纯化脂肪醛或脂肪醇。纯化后,样品中脂肪醛或脂肪醇的百分率增加。术语动词性的“纯化”、形容词性的“纯化的”或名词性的“纯化”是相对的术语,其不需要绝对的纯度。因此,例如当在重组宿主细胞中产生脂肪醛或脂肪醇时,纯化的脂肪醛或纯化的脂肪醇是基本与其他细胞成分(例如核酸、多肽、脂质、碳水化合物或其他碳氢化合物)隔离的脂肪醛或脂肪醇。

[0086] 菌株改良

[0087] 为了满足脂肪醇的效价、产率和/或生产率的极高目标,对生产型宿主细胞进行了多种修饰。FadR是与脂肪酸降解和脂肪酸生物合成途径中的重要调节因子(Cronan et al, Mol Microbiol, 29 (4):937-943 (1998))。大肠杆菌ACS酶FadD和脂肪酸运送蛋白质FadL为脂肪酸吸收系统的成分。FadL介导了脂肪酸运送至细菌细胞,而FadD介导了酰基-CoA酯的形成。当没有其他碳源可以利用时,外源脂肪酸被细菌吸收,并转化为酰基-CoA酯,其可以与运送因子FadR结合,并抑制fad基因的表达,其中所述的fad基因编码负责脂肪酸运送(FadL)、激活(FadD)和 β -氧化的蛋白质(FadA、FadB、FadE和FadH)。当有备选的碳源可以利用时,细菌合成脂肪酸作为酰基-ACP,其用于磷脂的合成,但是不是 β -氧化的底物。因此,酰基-CoA和酰基-ACP为脂肪酸的独立来源,这可以得到不同的终产物(Caviglia et al, J Biol. Chem., 279 (12):1163-1169 (2004))。美国临时申请号61/470,989描述了在宿主细胞中生产脂肪酸衍生物的改良的方法,其中所述的宿主细胞经基因改造从而具有与在相应的野生型宿主细胞中FadR多肽的表达水平相比,改变的FadR多肽的表达水平。

[0088] 在本领域中,关于宿主细胞(例如大肠杆菌)中的脂肪酸生物合成的限定因素具有

矛盾的推断。一种用于增加通过脂肪酸生物合成的流量的方法是操纵该途径中的各种酶(图1和2)。通过乙酰基-CoA羧化酶(acc)复合物(图3)和脂肪酸生物合成(fab)途径由乙酰基-CoA供应酰基-ACP可以限制脂肪醇生产的速率。在实施例2中详细描述的一种示例性方法中,谷氨酸棒状杆菌(*Corynebacterium glutamicum*) accABCD($\pm$ birA)的共表达效果证明基因修饰可以使大肠杆菌中的乙酰基-CoA和丙二酰基-CoA增加。造成通过脂肪酸生物合成的流量的低速率的一种可能原因是前体的限定供应,即,乙酰基-CoA,特别是丙二酰基-CoA,其为脂肪酸生物合成的主要前体。实施例3描述了fab操纵子的构建,其中所述的fab操纵子编码在生物合成途径中将丙二酰基-CoA转化成酰基-ACP并整合到大肠杆菌宿主细胞染色体中的酶。在实施例4中详细描述的另一方法中,显示大肠杆菌宿主细胞中rph和ilvG基因的突变使得更高地生产游离的脂肪酸(FFA),这可以解释为脂肪醇的更高生产。在另一个方法中,进行转座子诱变和高通量筛选,从而找到增加效价或产率的有利突变。实施例5描述了转座子插入到yijP基因中会如何改良摇瓶和分批供料发酵中的脂肪醇的产率。

[0089] 羧酸还原酶(CAR)

[0090] 重组宿主细胞经改造从而通过表达硫酯酶(其催化酰基-ACP转化成游离的脂肪酸(FFA))和羧酸还原酶(CAR)(其将游离脂肪酸转化成脂肪醛))来生产脂肪醇。在宿主细胞(例如大肠杆菌)中存在的天然(内源的)醛还原酶可以将脂肪醛转化成脂肪醇。示例性的硫酯酶在例如美国专利公开号20100154293中有所描述,其中所述的文献明确地以引用方式并入本文。CarB为一种示例性的羧酸还原酶,其为脂肪醇生产途径中的重要酶。WO2010/062480描述了使用NRRL 5646 CAR氨基酸序列(Genpept编号AAR91681)(SEQ ID NO:6)作为查询序列来进行BLAST检索,及其在鉴定大约20个同源序列中的用途。

[0091] 术语“羧酸还原酶”、“CAR”和“脂肪醛生物合成多肽”在本发明中可以交换使用。在实施本发明公开的过程中,编码羧酸还原酶多肽的基因在宿主细胞中表达或过表达。在一些实施方案中,CarB多肽具有氨基酸序列SEQ ID NO:7。在其他的实施方案中,CarB多肽为SEQ ID NO:7的变体或突变体。在某些实施方案中,CarB多肽由哺乳动物细胞、植物细胞、昆虫细胞、酵母细胞、真菌细胞、丝状真菌细胞、细菌细胞或任何其他有机体得到。在一些实施方案中,细菌细胞为选自耻垢分枝杆菌(*Mycobacterium smegmatis*)、脓肿分枝杆菌(*Mycobacterium abscessus*)、鸟型结核分枝杆菌(*Mycobacterium avium*)、牛型结核分枝杆菌(*Mycobacterium bovis*)、结核分枝杆菌(*Mycobacterium tuberculosis*)、麻风分枝杆菌(*Mycobacterium leprae*)、海洋分枝杆菌(*Mycobacterium marinum*)和溃疡分枝杆菌(*Mycobacterium ulcerans*)中的分枝杆菌。在其他的实施方案中,细菌细胞得自诺卡菌属(*Nocardia*)物种,例如诺卡菌NRRL 5646、鼻疽诺卡菌(*Nocardia farcinica*)、灰色链霉菌(*Streptomyces griseus*)、砂嗜盐产孢菌(*Salinispora arenicola*)或通用型密执安棒状杆菌(*Clavibacter michiganensis*)。在其他的实施方案中,CarB多肽为其氨基酸序列与SEQ ID NO:7的氨基酸序列具有至少大约80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%一致性的CarB的同源物。与SEQ ID NO:7的氨基酸序列具有至少大约80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%一致性的CarB多肽一致性不是特别限定的,并且本领域的任一普通技术人员都可以容易地鉴定大肠杆菌MG1655衍生的-CarB的同源物,并使用本发明所述的方法来测定其功能。在其他的实施方案

中,CarB多肽在SEQ ID NO:7的氨基酸编号3、12、20、28、46、74、103、191、288、473、827、926、927、930或1128处包括突变。示例性的突变在表10中详述。多肽的优选片段或突变体保留了相应的野生型多肽的一些或全部的生物学功能(例如酶活性)。在一些实施方案中,所述的片段或突变体保留了相应的野生型多肽的至少大约75%、至少大约80%、至少大约90%、至少大约95%、至少大约98%或者更高的生物学功能。在其他的实施方案中,所述的片段或突变体保留了相应的野生型多肽的大约100%的生物学功能。关于确定哪种氨基酸残基可以被取代、插入或删除而不会影响生物学活性的指导可以使用本领域公知的计算机程序来找到,例如LASERGENE™软件(DNASTAR, Inc., Madison, WI)。

[0092] 在其他的实施方案中,所述的片段或突变体与相应的野生型多肽相比表现出增强的生物学功能。例如,片段或突变体与相应的野生型多肽相比可以展示酶活性至少大约10%、至少大约25%、至少大约50%、至少大约75%、或者至少大约90%的改良。在其他的实施方案中,所述的片段或突变体的酶活性与相应的野生型多肽相比可以展示至少大约100%(例如至少大约200%或至少大约500%)的改良。应该理解的是本发明所述的多肽可以具有额外的保守型或非必需氨基酸的取代,这基本不会影响多肽的功能。一种特定的取代是否被耐受(即,不会不利地影响所需的生物学功能,例如DNA结合或酶活性)可以如Bowie et al. (Science, 247:1306-1310 (1990))中所述来确定。

[0093] 本发明公开的方法和变体酶的结果是相对于相应的野生型宿主细胞,增加由改造的宿主细胞所生产的脂肪酸或其衍生物的效价、产率和/或生产率中的一种或多种,其中所述的改造的宿主细胞具有表达水平改变的CarB多肽。为了可以将C12和C14脂肪酸最大程度地转化成脂肪醇,必须表达足够活性的CarB。改良的重组宿主细胞可以具有由例如大肠杆菌染色体表达的CAR酶。如实施例6所示,由染色体表达CarB酶的细胞比原始CarB具有更高的羧酸还原酶活性,并且能够将更多的C12和C14脂肪酸转化成脂肪醇。CarB为大基因(3.5kb)并显著地增加了质粒的大小,从而难以在菌株发育过程中使用pCL质粒来测试新基因。增加CarB活性的方法包括增加其溶解度、稳定性、表达和/或功能性。在一个示例性的方法中,生产在CarB的N末端包括6个组氨酸和凝血酶切割位点的融合蛋白质。这种酶在N末端具有额外60个氨基酸而与CarB不同,并且命名为CarB60。当CarB和CarB60在pTRC启动子的控制下由大肠杆菌染色体表达时,包括CarB60的细胞具有增加的总细胞羧酸还原酶活性,并将更多的C12和C14游离的脂肪酸(FFA)转化成脂肪醇。本领域的任一技术人员应该理解的是这是为了将C12和C14游离的脂肪酸(FFA)更多地转变成脂肪醇而进行分子改造的一个实例,如实施例6所示(上文)。相似的方法涵盖在本发明中(参见实施例7)。

[0094] 磷酸泛酰巯基乙胺基转移酶(PPTase) (EC 2.7.8.7) 催化4'-磷酸泛酰巯基乙胺由CoA转移至底物。诺卡菌属Car、CarB及其多种同源物包括假定的用于4'-磷酸泛酰巯基乙胺(PPT)的连接位点(He et al., Appl. Environ. Microbiol., 70 (3):1874-1881 (2004))。在本发明公开的一些实施方案中,PPTase在改造的宿主细胞中表达或过表达。在某些实施方案中,PPTase为得自大肠杆菌MG1655的EntD (SEQ ID NO:8)。在一些实施方案中,硫酯酶和羧酸还原酶在改造的宿主细胞中表达或过表达。在某些实施方案中,硫酯酶为tesA,而羧酸还原酶为carB。在其他的实施方案中,硫酯酶、羧酸还原酶和醇脱氢酶在改造的宿主细胞中表达或过表达。在某些实施方案中,硫酯酶为tesA,羧酸还原酶为carB,而醇脱氢酶为得自贝氏不动杆菌(Acinetobacter baylyi) ADP1的alrAadp1 (GenPept编号CAG70248.1) (SEQ ID

NO:4)。在其他的实施方案中,硫酯酶、羧酸还原酶、PPTase和醇脱氢酶在改造的宿主细胞中表达或过表达。在某些实施方案中,硫酯酶为tesA,羧酸还原酶为carB,PPTase为entD,而醇脱氢酶为alrAadpl。此外,在另一个实施方案中,表达硫酯酶、CAR、PPTase和醇脱氢酶中的一种或多种的改造宿主细胞还具有一种或多种的菌株改良。示例性的菌株改良包括但不限于乙酰基-CoA羧化酶多肽的表达或过表达、FadR多肽的过表达、异源iFAB操纵子的表达或过表达、转座子插入到yijP基因或另一种基因中或类似的方法。此外,本发明公开还提供了通过本发明所述的任一种方法生产的脂肪醇组合物。通过本发明所述的任一种方法生产的脂肪醇组合物可以直接用于生产其他化学品化合物(例如聚合物、表面活性剂、塑料、纺织品、溶剂、粘合剂等)或个人护理添加剂的起始材料。此外,这些化合物还可以用于随后反应(例如氢化作用、催化裂化(例如通过氢化作用和/或高温分解))的供料,从而制作其他的产物。

[0095] 突变体或变体

[0096] 在一些实施方案中,在重组宿主细胞中表达的多肽为本发明所述的任一种多肽的突变体或变体。如本文所用,术语“突变体”和“变体”是指具有如下氨基酸序列的多肽,其中所述的氨基酸序列与野生型多肽具有至少1个氨基酸的差异。例如突变体可以包括以下保守氨基酸取代的一个或多个:脂肪族氨基酸(例如丙氨酸、缬氨酸、亮氨酸和异亮氨酸)被另一种脂肪族氨基酸替代;丝氨酸被苏氨酸替代;苏氨酸被丝氨酸替代;酸性残基(例如天冬氨酸和谷氨酸)被另一种酸性残基替代;带有氨基的残基(例如天冬酰胺和谷氨酰胺)被另一种带有氨基的残基替代;碱性残基(赖氨酸和精氨酸)与另一种碱性残基互换;以及芳香族残基(例如苯丙氨酸和酪氨酸)被另一种芳香族残基替代。在一些实施方案中,突变体多肽具有大约1、2、3、4、5、6、7、8、9、10、15、20、30、40、50、60、70、80、90、100或更多的氨基酸取代、添加、插入或删除。优选的多肽片段或突变体保留了相应的野生型多肽的一些或全部的生物学功能(例如酶活性)。在一些实施方案中,所述的片段或突变体保留了相应的野生型多肽的至少大约75%、至少大约80%、至少大约90%、至少大约95%、至少大约98%或者更高的生物学功能。在其他的实施方案中,所述的片段或突变体保留了相应的野生型多肽的大约100%的生物学功能。关于确定哪种氨基酸残基可以被取代、插入或删除而不会影响生物学活性的指导可以使用本领域公知的计算机程序来找到,例如LASERGENETM软件(DNASTAR, Inc., Madison, WI)。

[0097] 在其他的实施方案中,所述的片段或突变体与相应的野生型多肽相比表现出增强的生物学功能。例如,片段或突变体与相应的野生型多肽相比可以展示酶活性至少10%、至少25%、至少50%、至少75%、或者至少90%的改良。在其他的实施方案中,所述的片段或突变体的酶活性与相应的野生型多肽相比可以展示至少100%(例如至少200%或至少500%)的改良。应该理解的是本发明所述的多肽可以具有额外的保守型或非必需氨基酸的取代,这基本不会影响多肽的功能。一种特定的取代是否被耐受(即,不会不利地影响所需的生物学功能,例如羧酸还原酶活性)可以如Bowie et al. (Science, 247:1306-1310 (1990)) 中所述来确定。保守型氨基酸取代为其中氨基酸残基被具有相似侧链的氨基酸残基替代的取代。本领域已经对具有相似侧链的氨基酸残基家族进行定义。这些家族包括具有碱性侧链的氨基酸(例如赖氨酸、精氨酸、组氨酸)、酸性侧链的氨基酸(例如天冬氨酸、谷氨酸)、不带电荷的极性侧链的氨基酸(例如甘氨酸、天冬酰胺、谷氨酰胺、丝氨酸、苏氨酸、酪氨酸、半胱

氨酸)、非极性侧链的氨基酸(例如丙氨酸、缬氨酸、亮氨酸、异亮氨酸、脯氨酸、苯丙氨酸、甲硫氨酸、色氨酸)、 β -支化侧链的氨基酸(例如苏氨酸、缬氨酸、异亮氨酸)以及芳香族侧链的氨基酸(例如酪氨酸、苯丙氨酸、色氨酸、组氨酸)。变体可以天然形成或在体外创建。具体而言,可以使用基因改造技术(例如定点诱变、随机化学诱变、核酸外切酶III的删除方法或标准的克隆技术)来创建此类变体。备选地,可以使用化学合成或修饰方法来创建此类变体、片段、类似物或衍生物。

[0098] 制备变体的方法是本领域公知的。这些方法包括以下方法:将由天然分离物得到的核酸序列进行修饰,从而生成编码多肽的核酸,其中所述的多肽具有增强它们在工业或实验室应用中的价值的特征。在此类方法中,大量的变体序列被生成并表征,其中所述的变体序列与由天然分离物得到的序列具有一个或多个核苷酸差异。通常,这些核苷酸差异使得相对于由天然分离物得到的核酸所编码的多肽而言发生氨基酸改变。例如可以通过使用随机和定点诱变来制备变体。随机和定点诱变在例如Arnold, Curr. Opin. Biotech., 4:450-455 (1993) 中有所描述。可以使用易错PCR来取得随机诱变(例如参见Leung et al., Technique, 1:11-15 (1989); 和Caldwell et al., PCR Methods Applic, 2:28-33 (1992))。在易错PCR中,在其中DNA聚合酶的复制忠实性低的条件下实施PCR,这样在PCR产物的全长上获得高比率的点突变。简言之,在此类方法中,待诱变的核酸(例如编码羧基还原酶的多核苷酸序列)与PCR引物、反应缓冲剂、MgCl₂、MnCl₂、Taq聚合酶及合适浓度的dNTP混合,以便在PCR产物的全长上获得高比率的点突变。例如可以使用待诱变的20毫微摩尔的核酸,30皮摩尔的每种PCR引物,包含50mM KCl、10mM Tris HCl (pH 8.3)、0.01%凝胶、7mM MgCl₂、0.5mM MnCl₂的反应缓冲剂,5单位的Taq聚合酶,0.2mM dGTP,0.2mM dATP,1mM dCTP和1mM dTTP来进行反应。PCR可以进行30个循环:94°C, 1min; 45°C, 1min; 以及72°C, 1min。然而,应该理解的是这些参数可以适当地改变。然后将诱变核酸克隆到合适的载体中,并评价由诱变的核酸编码的多肽的活性(参见实施例7)。可以使用寡核苷酸定点诱变在目的任何克隆的DNA中生成位点特异性突变,从而完成定点诱变。寡核苷酸诱变在例如Reidhaar-Olson et al., Science, 241:53-57 (1988) 中有所描述。简言之,在此类方法中,许多双链寡核苷酸被合成并插入到待诱变的克隆DNA(例如编码CAR多肽的多核苷酸序列)中,其中所述的双链寡核苷酸带有待引入到克隆DNA中的一个或多个突变。回收包括诱变的DNA的克隆,并评估它们所编码的多肽的活性。用于生成变体的另一种方法为拼装PCR。拼装PCR涉及由小DNA片段的混合物拼装得到PCR产物。大量不同的PCR反应在相同的瓶中平行进行,其中一个反应的产物引发另一个反应的产物。拼装PCR在例如美国专利5,965,408中有所描述。生成变体的另一种方法是有性PCR诱变。在有性PCR诱变中,DNA分子的随机发酵(基于序列同源性)的结果是,在体外在不同的但高度相关的DNA序列的DNA分子之间形成被迫的同源重组。其后为在PCR反应中,通过引物的延伸而形成交叉固定。有性PCR诱变在例如Stemmer, Proc. Natl. Acad. Sci., U.S.A., 91:10747-10751 (1994) 中有所描述。

[0099] 此外,还可以通过体内诱变来创建变体。在一些实施方案中,通过在细菌菌株(例如大肠杆菌菌株)中增殖核酸序列来形成该核酸序列中的随机突变,其中所述的核酸序列在一种或多种DNA修复途径中携带突变。此类“增变基因”菌株具有比野生型菌株更高的随机突变。在这些菌株中的一种菌株中增殖DNA序列(例如编码CAR多肽的多核苷酸序列)最终将在DNA内形成随机突变。适用于体内诱变的增变基因菌株在例如国际专利申请公开号

WO1991/016427中有所描述。此外,还可以使用盒式诱变来生成变体。在盒式诱变中,小区域的双链DNA分子被不同于天然序列的合成寡核苷酸“盒”替代。寡核苷酸通常包括完全和/或部分随机的天然序列。此外,还可以使用递归整体诱变来生成变体。递归整体诱变是用于蛋白质改造(即,蛋白质诱变)的算法,研发该方法用于生产其成员具有氨基酸序列的差异、但表现型相关的突变体的多样化群体。该方法使用反馈机制来控制一连串的组合盒式诱变。递归整体诱变在例如Arkin et al, Proc. Natl. Acad. Sci., U.S.A., 89:7811-7815 (1992) 中有所描述。在一些实施方案中,使用指数整体诱变来创建变体。指数整体诱变为用于生成组合文库的方法,其中所述的组合文库具有高百分率的独特的及功能突变体,其中多个残余物小组被平行随机化,从而鉴定在各改变的位置处会得到功能蛋白质的氨基酸。指数整体诱变在例如Delegrave et al., Biotech. Res, 11:1548-1552 (1993) 中有所描述。在一些实施方案中,使用改组方法来创建变体,其中编码不同多肽的多个核酸的部分融合在一起从而创建编码嵌合多肽的嵌合核酸序列,例如美国专利5,965,408和5,939,250中所述。

[0100] 插入诱变为通过插入一个或多个碱基来进行DNA诱变。插入突变可以天然形成、通过病毒或转座子介导、或者可以用于实验室研究目的的人工创建(例如通过转座子诱变)。当外源DNA被整合到宿主DNA中时,任何随后突变的严重性都完全取决于DNA插入到宿主基因组中的位置。例如如果转座子插入到必需基因的中间、启动子区域中或插入到抑制因子或增强子区域中,那么有意义的效果可以是明显的。实施转座子诱变和高生产量筛选以找到能够增加脂肪醇的效价或产率的有利突变。本发明公开提供了重组宿主细胞,其包括:(a) 编码羧酸还原酶的多核苷酸序列,其中所述的羧酸还原酶包括如下氨基酸序列,该氨基酸序列与SEQ ID NO:7的氨基酸序列具有至少80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%的序列一致性;以及(b) 编码具有羧酸还原酶活性的多肽的多核苷酸,其中所述的重组宿主细胞能够生产脂肪醛或脂肪醇。

[0101] 改造宿主细胞

[0102] 在一些实施方案中,通过重组载体的方式将多核苷酸(或基因)序列提供给宿主细胞,其中所述的重组载体包括与多核苷酸序列可操作地连接的启动子。在某些实施方案中,启动子是发育调节的、细胞器特异性的、组织特异性的、诱导型的、构成型的或者是细胞特异性的启动子。在一些实施方案中,重组载体包括:(a) 与多核苷酸序列可操作地缀合的表达控制序列;(b) 与多核苷酸序列可操作地缀合的选择标记;(c) 与多核苷酸序列可操作地缀合的标记序列;(d) 与多核苷酸序列可操作地缀合的纯化部分;(e) 与多核苷酸序列可操作地缀合的分泌序列;以及(f) 与多核苷酸序列可操作地缀合的靶向序列。本发明所述的表达载体包括本发明所述的、以适用于在宿主细胞中表达多核苷酸序列的形式多核苷酸序列。本领域的那些技术人员将理解的是表达载体的设计可以取决于诸如待转化的宿主细胞的选择、所需多肽的表达水平之类的因素。本发明所述的表达载体可以被引入到宿主细胞中,从而生产由本发明所述的多核苷酸序列编码的多肽,包括融合多肽。在原核生物(例如大肠杆菌)中,表达编码多肽的基因大部分是使用包括构成型或诱导型启动子的载体来实施的,其中所述的启动子指导融合或非融合多肽的表达。融合载体将一定数量的氨基酸加入到其中编码的多肽中,通常加入到重组多肽的氨基末端或羧基末端。这种融合载体通常起到以下3种目的中的一种或多种:(1) 增加重组多肽的表达;(2) 增加重组多肽的溶解性;

以及(3)通过在亲和纯化中起到配体的作用而有助于重组多肽的纯化。通常,在融合表达的载体中,在融合部分与重组多肽的连接处引入蛋白水解的切割位点。这使得在纯化融合多肽后,重组多肽与融合部分分离。此类酶及它们的同种识别序列的实例包括因子Xa、凝血酶和肠激酶。融合表达载体的实例包括pGEX (Pharmacia Biotech, Inc., Piscataway, NJ; Smith et al., Gene, 67:31-40 (1988))、pMAL (New England Biolabs, Beverly, MA) 和 pRITS (Pharmacia Biotech, Inc., Piscataway, N.J.), 它们分别将谷胱甘肽S-转移酶 (GST), 麦芽糖E结合蛋白质或蛋白质A融合到目标重组多肽中。

[0103] 诱导型非融合大肠杆菌表达载体的实例包括pTrc (Amann et al, Gene (1988) 69: 301-315) 和pET 11d (Studier et al, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990) 60-89)。pTrc载体的目标基因表达依赖于由杂交的trp-lac融合启动子引起的宿主RNA聚合酶的转录。pET 11d载体的目标基因表达依赖于由T7gn10-lac融合启动子引起的转录,其中所述的启动子是由共表达的病毒RNA聚合酶 (T7gn1) 介导的。这种病毒聚合酶是通过宿主菌株BL21 (DE3) 和HMS174 (DE3), 由容留T7gn1基因的居民 λ 噬菌体在lacUV5启动子的转录控制之下供应的。用于原核和真核细胞的合适的表达系统是本领域公知的,例如参见Sambrook et al, "Molecular Cloning: A Laboratory Manual," second edition, Cold Spring Harbor Laboratory, (1989)。诱导型非融合大肠杆菌表达载体的实例包括pTrc (Amann et al, Gene, 69:301-315 (1988)) 和PET 11d (Studier et al, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA, pp.60-89 (1990))。在某些实施方案中,本发明公开的多核苷酸序列与由噬菌体T5衍生的启动子可操作地连接。在一个实施方案中,宿主细胞为酵母细胞。在这种实施方案中,表达载体为酵母表达载体。可以通过用于将外来核酸(例如DNA)引入到宿主细胞中的多种本领域公认的技术将载体引入到原核和真核细胞中。用于转化或转染宿主细胞的合适的方法可以在例如Sambrook et al. (上文) 中找到。已知稳定地转化细菌细胞取决于所使用的表达载体和转化技术,仅有一小部分细胞接受并复制表达载体。在一些实施方案中,为了鉴定和选择这些转化子,编码选择标记(例如抗生素抗性)的基因可以与目的基因一起引入到宿主细胞中。选择标记包括赋予药品抗性的那些,例如但不限于氨苄青霉素、卡那霉素、氯霉素或四环素。编码选择标记的核酸可以在与编码本发明所述多肽的相同的载体上引入到宿主细胞中,或者可以在分开的载体上引入到宿主细胞中。稳定地转化有所引入核酸的细胞可以通过在合适的选择药品存在下的生长来鉴定。

[0104] 由重组宿主细胞生产脂肪醇衍生物

[0105] 增加重组宿主细胞生产脂肪醇的策略包括通过在改造的生产型宿主中过表达天然脂肪酸生物合成基因并表达由不同有机体得到的外源脂肪酸生物合成基因来增加通过脂肪酸生物合成途径的流量。此外,还可以使用在脂肪醇生物合成途径中相关酶(例如CAR)的增强的活性以及优化宿主细胞的生长和生产率的其他策略来使生产最大化。在一些实施方案中,重组宿主细胞包括编码多肽(酶)的多核苷酸,其中所述的多肽(酶)具有脂肪醇生物合成活性(即,脂肪醇生物合成多肽或脂肪醇生物合成酶),并且重组宿主细胞生产脂肪醇。可以通过在有效表达脂肪醇生物合成酶的条件下在碳源存在下培养重组宿主细胞来生产包括脂肪醇的组合物(脂肪醇组合物)。在一些实施方案中,脂肪醇组合物包括脂肪醇,然

而,脂肪醇组合物可以包括其他的脂肪酸衍生物。通常,由重组宿主细胞的细胞外环境(即,细胞培养物培养基)中回收脂肪醇组合物。在一种方法中,重组宿主细胞经改造,从而通过表达硫酯酶和羧酸还原酶(CAR)来生产脂肪醇,其中所述的硫酯酶将酰基-ACP转化为游离的脂肪酸(FFA),其中所述的羧酸还原酶将游离的脂肪酸转化为脂肪醛。存在于宿主细胞(例如大肠杆菌)中天然的(内源的)醛还原酶可以将脂肪醛转化为脂肪醇。在一些实施方案中,通过在重组宿主细胞中表达或过表达编码如下多肽的多核苷酸来生产脂肪醇,其中所述的多肽具有脂肪醇生物合成活性,其将脂肪醛转化成脂肪醇。例如醇脱氢酶(本发明也称为醛还原酶,例如EC 1.1.1.1)可以用于实施本发明公开。如本文所用,术语“醇脱氢酶”是指能够催化将脂肪醛转化为醇(例如脂肪醇)的多肽。本领域的任一普通技术人员将理解的是某些醇脱氢酶能够催化其他的反应,并且这些非特异性的醇脱氢酶也被涵盖在术语“醇脱氢酶”的范围内。可以根据本发明公开使用的醇脱氢酶多肽的实例包括但不限于AlrAadp1(SEQ ID NO:4)或AlrA同源物,以及内源大肠杆菌醇脱氢酶,例如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) 和Ybb0 [AAC73595.1]。其他的实例在国际专利申请公开号WO 2007/136762、WO2008/119082和WO2010/062480中有所描述,这些文献均明确地以引用方式并入本文。在某些实施方案中,脂肪醇生物合成多肽具有醛还原酶或醇脱氢酶活性(EC 1.1.1.1)。在另一种方法中,重组宿主细胞经改造,从而通过表达形成脂肪醇的酰基-CoA还原酶或脂肪酰基还原酶(FAR)来生产脂肪醇,其中所述的形成脂肪醇的酰基-CoA还原酶或脂肪酰基还原酶(FAR)将脂肪酰基硫酯底物(例如脂肪酰基-CoA或脂肪酰基-ACP)转化为脂肪醇。在一些实施方案中,通过在重组宿主细胞中表达或过表达编码具有形成脂肪醇的酰基-CoA还原酶(FAR)活性的多肽的多核苷酸来生产脂肪醇。根据该实施方案可以使用的FAR多肽的实例在PCT公开号WO2010/062480中有所描述,该文献明确地以引用方式并入本文。

[0106] 可以通过酰基-CoA依赖途径(利用了脂肪酰基-ACP和脂肪酰基-CoA中间体)和酰基-CoA独立途径(利用了脂肪酰基-ACP中间体,但不利用脂肪酰基-CoA中间体)来生产脂肪醇。在特定的实施方案中,由过表达的基因所编码的酶选自脂肪酸合成酶、酰基-ACP硫酯酶、脂肪酰基-CoA合成酶和乙酰基-CoA羧化酶。在一些实施方案中,由过表达的基因所编码的蛋白质对于宿主细胞而言是内源的。在其他的实施方案中,由过表达的基因所编码的蛋白质对于宿主细胞而言是异源的。此外,脂肪醇还可以通过能够将多种酰基-ACP或酰基-CoA分子还原成相应的伯醇的酶而天然地形成。此外,参见美国专利公开号20100105963、20110206630和美国专利号8097439,这些文献明确地以引用方式并入本文。如本文所用,重组宿主细胞或改造的宿主细胞是指这样的宿主细胞,例如通过蓄意引入新的基因元件和/或蓄意修饰在宿主细胞中天然存在的基因元件而相对于相应的野生型宿主细胞而言使其基因构造被改变。此类重组宿主细胞的后代还包括这些新的和/或修饰的基因元件。在本发明所述的发明公开中的任一方面中,宿主细胞可以选自植物细胞、昆虫细胞、真菌细胞(例如丝状真菌,如假丝酵母菌种;或者芽殖酵母,例如酵母菌种)、海藻细胞和细菌细胞。在一个优选的实施方案中,重组宿主细胞为重组微生物细胞。作为微生物细胞的宿主细胞的实例包括但不限于得自大肠杆菌属(*Escherichia*)、芽孢杆菌属(*Bacillus*)、乳酸杆菌属(*Lactobacillus*)、发酵单胞菌属(*Zymomonas*)、红球菌属(*Rhodococcus*)、假单胞菌属

(*Pseudomonas*)、曲霉菌属(*Aspergillus*)、木霉属(*Trichoderma*)、链孢霉属(*Neurospora*)、镰孢属(*Fusarium*)、腐殖霉属(*Humicola*)、根毛霉属(*Rhizomucor*)、克鲁维氏酵母属(*Kluyveromyces*)、毕赤酵母属(*Pichia*)、毛霉属(*Mucor*)、蚀丝霉属(*Myceliophthora*)、青霉菌属(*Penicillium*)、原毛平革菌属(*Phanerochaete*)、侧耳属(*Pleurotus*)、栓菌属(*Trametes*)、金孢子菌属(*Chrysosporium*)、酵母菌属(*Saccharomyces*)、寡养单胞菌属(*Stenotrophomonas*)、裂殖酵母属(*Schizosaccharomyces*)、耶氏酵母属(*Yarrowia*)或链霉菌属(*Streptomyces*)的细胞。在一些实施方案中,宿主细胞为革兰氏阳性细菌细胞。在其他的实施方案中,宿主细胞为革兰氏阴性细菌细胞。在一些实施方案中,宿主细胞为大肠杆菌细胞。在其他的实施方案中,宿主细胞为迟缓芽孢杆菌(*Bacillus lentus*)细胞、短芽孢杆菌(*Bacillus brevis*)细胞、嗜热脂肪芽孢杆菌(*Bacillus stearothermophilus*)细胞、地衣芽孢杆菌(*Bacillus licheniformis*)细胞、嗜碱芽孢杆菌(*Bacillus alkalophilus*)细胞、凝结芽孢杆菌(*Bacillus coagulans*)细胞、环状芽孢杆菌(*Bacillus circulans*)细胞、短小芽孢杆菌(*Bacillus pumilis*)细胞、苏云金芽孢杆菌(*Bacillus thuringiensis*)细胞、克劳氏芽孢杆菌(*Bacillus clausii*)细胞、巨大芽孢杆菌(*Bacillus megaterium*)细胞、枯草芽孢杆菌(*Bacillus subtilis*)细胞、或解淀粉芽孢杆菌(*Bacillus amyloliquefaciens*)细胞。在其他的实施方案中,宿主细胞为康氏木霉(*Trichoderma koningii*)细胞、绿色木霉(*Trichoderma viride*)细胞、里氏木霉(*Trichoderma reesei*)细胞、长梗木霉(*Trichoderma longibrachiatum*)细胞、泡盛曲霉(*Aspergillus awamori*)细胞、烟曲霉(*Aspergillus fumigatus*)细胞、臭曲霉(*Aspergillus foetidus*)细胞、构巢曲霉(*Aspergillus nidulans*)细胞、黑曲霉(*Aspergillus niger*)细胞、米曲霉(*Aspergillus oryzae*)细胞、特异腐质霉(*Humicola insolens*)细胞、棉毛状腐质霉(*Humicola lanuginosa*)细胞、不透明红球菌细胞、米黑根毛霉(*Rhizomucor miehei*)细胞或米黑毛霉(*Mucor miehei*)细胞。

[0107] 在其他的实施方案中,宿主细胞为变铅青链霉菌(*Streptomyces lividans*)细胞或灰链霉菌(*Streptomyces murinus*)细胞。在其他的实施方案中,宿主细胞为放线菌细胞。在一些实施方案中,宿主细胞为酿酒酵母细胞。在一些实施方案中,宿主细胞为酿酒酵母细胞。在其他的实施方案中,宿主细胞为得自真核植物、海藻、蓝细菌、绿硫细菌、绿色非硫细菌、紫硫细菌、紫色非硫细菌、嗜极生物、酵母、真菌、它们经改造的有机体或合成有机体的细胞。在一些实施方案中,宿主细胞为光依赖的或固定碳。在一些实施方案中,宿主细胞为光依赖的或固定碳。在一些实施方案中,宿主细胞具有自养活性。在一些实施方案中,例如在光缺乏下,宿主细胞具有光能自养活性。在一些实施方案中,例如在光缺乏下,宿主细胞为异养的或兼养的。在某些实施方案中,宿主细胞为得自拟南芥(*Arabidopsis thaliana*)、柳枝稷(*Panicum virgatum*)、巨芒(*Miscanthus giganteus*)、玉米(*Zea mays*)、葡萄藻(*Botryococcus braunii*)、莱茵衣藻(*Chlamydomonas reinhardtii*)、杜氏盐藻(*Dunaliella salina*)、聚球菌属(*Synechococcus* Sp.)菌种PCC 7002、聚球菌属菌种PCC 7942、集胞藻(*Synechocystis* Sp.)菌种PCC 6803、长嗜热聚球蓝细菌(*Thermosynechococcus elongatus*) BP-1、绿硫细菌(*Chlorobium tepidum*)、橙色绿屈挠菌(*Chloroflexus auranticus*)、酒色着色菌(*Chromatium vinosum*)、深红红螺菌(*Rhodospirillum rubrum*)、荚膜红细菌(*Rhodobacter capsulatus*)、沼泽红假单胞菌

(*Rhodopseudomonas palusris*)、杨氏梭菌(*Clostridium ljungdahlii*)、热纤梭菌(*Clostridiuthermocellum*)、产黄青霉(*Penicillium chrysogenum*)、巴斯德毕赤酵母(*Pichia pastoris*)、酿酒酵母、粟酒裂殖(*Schizosaccharomyces pombe*)、萤光假单孢菌(*Pseudomonas fluorescens*)或运动发酵单胞菌(*Zymomonas mobilis*)的细胞。

[0108] 改造宿主细胞的培养和发酵

[0109] 如本文所用,发酵在广义上是指宿主细胞将有机材料转化为目标物质,例如通过在包括碳源的培养基中增殖重组宿主细胞的培养物,使重组宿主细胞将碳源转化为脂肪酸及其衍生物。如本文所用,允许生产的条件是指允许宿主细胞生产所需产物(例如脂肪酸或脂肪酸衍生物)的任何条件。类似地,其中载体的多核苷酸序列被表达的条件是指允许宿主细胞合成多肽的任何条件。合适的条件包括例如发酵条件。发酵条件可以包括例如许多参数,包括但不限于温度范围、通风水平、供料速率和培养基组成。这些条件(单独及组合)均允许宿主细胞生长。发酵可以是需氧的、厌氧的或它们的变体(例如微需氧的)。示例性的培养物培养基包括培养液或凝胶。通常,培养基包括可以被宿主细胞直接代谢的碳源。此外,培养基中可以使用酶,从而有利于碳源的动员(例如淀粉或纤维素解聚成可发酵的糖)和随后的代谢。就小规模的生产而言,改造的宿主细胞可以以例如大约100mL、500mL、1L、2L、5L或10L的批次生长;发酵;并诱导表达所需的多核苷酸序列,例如编码CAR多肽的多核苷酸序列。就大规模的生产而言,改造的宿主细胞可以以例如大约10L、100L、1000L、10,000L、100,000L、1,000,000L或更大的批次生产;发酵;并诱导表达所需的多核苷酸序列。备选地,可以实施大规模分批供料的发酵。

[0110] 脂肪醇组合物

[0111] 本发明所述的脂肪醇组合物可以在重组宿主细胞培养物的细胞外环境中找到,并且可以容易地由培养物培养基中分离。脂肪醇组合物可以由重组宿主细胞分泌、运送至细胞外环境中或者主动转运至重组宿主细胞培养物的细胞外环境中。使用本领域已知的常规方法由重组宿主细胞分离脂肪醇组合物。本发明公开提供了由改造的或重组的宿主细胞生产的组合物(生物产物),其包括一种或多种脂肪醛和/或脂肪醇。尽管具有特定链长和饱和度的脂肪醇成分可以构成由培养的改造或重组宿主细胞所生产的生物产物的大部分,但是所述的组合物通常包括链长和/或饱和度可变的脂肪醛和/或脂肪醇的混合物。如本文所用,现代碳级份或fM具有与National Institute of Standards and Technology (NIST) Standard Reference Materials SRMs4990B和4990C(其分别称为草酸标准品H0xI和HoxII)定义的相同的含义。基本定义是指0.95乘以 $^{14}\text{C}/^{12}\text{C}$ 同位素比H0xI(称为AD1950)。这大致相当于衰变校正过的工业革命前的木材。对于当前的活生物圈(植物材料),fM为大约1.1。

[0112] 包括生物学方法生产的有机化合物,特别是使用本发明的脂肪酸生物合成途径生产的脂肪醛和醇的生物产物(例如根据本发明公开生产的脂肪醛和醇)不是使用可再生资源生产的,因此是新的物质的组合物。基于双重碳同位素指纹法(dual carbon-isotopic fingerprinting)或 ^{14}C 定年法可以将这些新的生物产物与衍生于石油化学碳的有机化合物区别开。此外,可以通过双重碳同位素指纹法(例如参见美国专利No.7,169,588,该文献以引用方式并入本文)来确定生物来源碳的具体来源(例如葡萄糖与甘油)。在商业中,区别生物产物与石油基有机化合物的能力在追踪这些材料方面是有益的。例如,包括生物学基的

和石油基的碳同位素谱二者的有机化合物或化学品可以区别于仅由石油基材料制成的有机化合物和化学品。因此,在商业中,可以基于本发明的生物产物的独特的碳同位素谱来追踪或跟踪它们。通过比较各燃料中稳定碳的同位素比 ($^{13}\text{C}/^{12}\text{C}$),可以将生物产物与石油基有机化合物区分开。给定的生物产物中 $^{13}\text{C}/^{12}\text{C}$ 比是二氧化碳被固定时大气二氧化碳中的 $^{13}\text{C}/^{12}\text{C}$ 比的结果。它还反映了准确的代谢途径。此外,还存在区域性差异。石油、 C_3 植物(阔叶植物)、 C_4 植物(草本植物)以及海洋碳酸盐类在 $^{13}\text{C}/^{12}\text{C}$ 以及相应的 $\delta^{13}\text{C}$ 值中全部表现出显著差异。进一步地,由于代谢途径的原因, C_3 和 C_4 植物的脂物质的分析情况与衍生自相同植物的碳水化合物组分材料不同。在测量精度之内,由于同位素分馏效应, ^{13}C 表现出大的差异,对于生物产物而言最显著的效应是光合机制。植物中碳同位素比中差异的主要原因与植物中光合碳代谢途径的差异紧密相关,特别是在初级羧化(即大气 CO_2 的初始固定)期间发生的反应。两大类植物是涉及“ C_3 ”(或Calvin-Benson)光合循环以及涉及“ C_4 ”(或Hatch-Slack)光合循环的植物。在 C_3 植物中,主要 CO_2 固定或羧化反应涉及核酮糖-1,5-二磷酸羧化酶,并且第一稳定产物是3-碳化合物。诸如硬木树和针叶树之类的 C_3 植物在温带气候地带中是主要的。在 C_4 植物中,涉及另一种酶(磷酸烯醇丙酮酸羧化酶)的其它羧化反应是主要羧化反应。第一稳定的碳化合物是4-碳的酸,该4-碳的酸随后被脱羧基。如此释放的 CO_2 被 C_3 循环再固定。 C_4 植物的实例是热带牧草、玉米和甘蔗。 C_4 和 C_3 植物都表现出一定范围的 $^{13}\text{C}/^{12}\text{C}$ 同位素比,但是对于 C_4 植物而言,通常的值是大约-7至大约-13,而对于 C_3 植物而言,是大约-19至大约-27/密耳(例如参见Stuiver et al, Radiocarbon 19:355 (1977))。煤和石油通常落在后面的范围内。 ^{13}C 测量尺度最初由Pee Dee Belemnite (PDB) 石灰石的零集所定义,其中给出的值是与该材料的偏差的千分比。“ $\delta^{13}\text{C}$ ”值被表达为千分比(每密耳),缩写为‰,并且按照下式进行计算:

$$[0113] \quad \delta^{13}\text{C}(\text{‰}) = [(^{13}\text{C}/^{12}\text{C})_{\text{样品}} - (^{13}\text{C}/^{12}\text{C})_{\text{标准}}] / (^{13}\text{C}/^{12}\text{C})_{\text{标准}} \times 1000$$

[0114] 由于PDB参考材料(RM)已经被耗尽,已经与IAEA、USGS、NIST和其它选定的国际同位素实验室合作开发了一系列备选RM。与PDB的每密耳偏差的标记为 $\delta^{13}\text{C}$ 。通过高精度的稳定比率质谱(IRMS)在质量44、45以及46的分子离子上对 CO_2 进行测量。本发明所述的组合物包括通过本发明所述的任何方法生产的生物产物,包括例如脂肪醛和醇产物。具体而言,所述的生物产物的 $\delta^{13}\text{C}$ 可以为大约-28或更大、大约-27或更大、-20或更大、-18或更大、-15或更大、-13或更大、-10或更大或-8或更大。例如所述的生物产物的 $\delta^{13}\text{C}$ 可以为大约-30至大约-15、大约-27至大约-19、大约-25至大约-21、大约-15至大约-5、大约-13至大约-7或大约-13至大约-10。在其它的情况下,所述的生物产物的 $\delta^{13}\text{C}$ 可以为大约-10、-11、-12或-12.3。此外,还可以通过比较每种化合物中 ^{14}C 的量,将生物产物(包括根据本发明公开生产的生物产物)区别于石油基有机化合物。因为 ^{14}C 具有5730年的核半衰期,所以包括“较老的”碳的石油基燃料可以区别于包括“较新的”碳的生物产物(参见例如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))。

[0115] 放射性碳定年法的基本假设是大气中 ^{14}C 浓度的恒定性导致活有机体中 ^{14}C 的恒定性。然而,由于1950年以来的大气核试验以及1850年以来的化石燃料的燃烧, ^{14}C 已经获得了次要的地球化学时间特征(geochemical time characteristic)。其在大气 CO_2 中以及由此

在活生物圈中的浓度在20世纪60年代中期核试验的高峰时几乎翻倍。此后,其逐渐地返回至大约 1.2×10^{-12} 的稳态宇宙发生(大气)基线同位素率($^{14}\text{C}/^{12}\text{C}$),并且大致的弛豫“半衰期”(relaxation “half-life”)为7至10年。(后者所述的这种半衰期不必按照字面理解;而是必须使用详细的大气核输入/衰变函数来追踪核时代开始以来的大气和生物圈 ^{14}C 的变化。)后者所述的这种生物圈 ^{14}C 时间特征支持每年定年法测定近代生物圈的碳的前景。可以通过加速器质谱(AMS)测量 ^{14}C ,结果以“现代碳的级份”(f_M)为单位给出。f_M由National Institute of Standards and Technology (NIST) Standard Reference Materials (SRM) 4990B和4990C定义。如本文所用,现代碳的级份(f_M)具有与National Institute of Standards and Technology (NIST) Standard Reference Materials (SRMs) 4990B和4990C(其分别称为草酸标准品H0xI和H0xII)定义的相同的含义。基本的定义涉及0.95乘以 $^{14}\text{C}/^{12}\text{C}$ 同位素比H0xI(称为AD 1950)。这大致相当于衰变校正的工业革命前的木材。对于当前的活生物圈(植物材料),f_M为大约1.1。这大致相当于衰变校正的工业革命前的木材。对于当前的活生物圈(植物材料),f_M为大约1.1。

[0116] 本发明所述的组合物包括f_M ^{14}C 可以为至少大约1的生物产物。例如本发明公开的生物产物可以具有至少大约1.01的f_M ^{14}C 、大约1至大约1.5的f_M ^{14}C 、大约1.04至大约1.18的f_M ^{14}C 、或者大约1.111至大约1.124的f_M ^{14}C 。已知 ^{14}C 的另一种量度为现代碳的百分率(pMC)。对于使用 ^{14}C 年代的考古学家或地质学家而言,AD 1950等于“旧的零年”。其还表示100pMC。在1963年热核武器的顶峰时,大气中的“炸弹碳”达到正常水平的几乎两倍。其在大气中的分布自其出现以来是近似的,对于自从AD 1950生活的植物和动物表现出大于100pMC的值。随着时间的流逝,该值逐渐降低,当今的值接近107.5pMC。这意味着诸如谷物之类的新生物质材料将给出接近107.5pMC的 ^{14}C 标记。石油基化合物将具有为零的PMC值。化石碳与现代碳的组合导致现代pMC含量被稀释。通过假定107.5pMC代表现代生物质材料的 ^{14}C 含量并且0pMC代表石油基产物的 ^{14}C 含量,则该材料的测量pMC值将反映两种类型的成分的比例。例如100%来源于现代大豆的材料将给出接近107.5pMC的放射性碳标记。如果使用石油基产物将材料稀释50%,则它将给出大约54pMC的放射性碳标记。通过指定“100%”等于107.5pMC并且指定“0%”等于0pMC而衍生得到基于生物学的碳含量。例如测量为99pMC的样品将给出93%的等同的生物学的基的碳含量。该值被称为基于生物学的平均碳结果并且假定在所分析的材料中的所有成分均起源于现代的生物材料或石油基材料。包括本发明所述的一种或多种脂肪醛或醇的生物产物可以具有至少大约50、60、70、75、80、85、90、95、96、97、98、99或100的pMC。在其它的情况下,本发明所述的生物产物可以具有大约50至大约100、大约60至大约100、大约70至大约100、大约80至大约100、大约85至大约100、大约87至大约98、或者大约90至大约95的pMC。在其它的情况下,本发明所述的生物产物可以具有大约90、91、92、93、94或94.2的pMC。

[0117] 筛选由重组宿主细胞生产的脂肪醇组合物

[0118] 为了测定所述的这些条件是否足以允许表达,可以将包括异源基因或修饰的天然基因的重组宿主细胞培养,例如大约4、8、12、24、36或48小时。在培养过程中和/或培养后,可以获得样品并分析,以测定脂肪醇的生产水平(效价、产率或生产率)是否与相应的野生型亲代细胞(未经修饰的)的生产水平不同。例如可以针对所需产物的存在情况来测定其中生长有宿主细胞的培养基。当测定产物的存在情况时,可以使用例如但不限于TLC、HPLC、

GC/FID、GC/MS、LC/MS和MS的测试。可以在平板或摇瓶的少量(0.001L至1L)培养基中培养重组宿主细胞菌株,以便筛选生产水平改变的脂肪醇或脂肪物质。一旦在小规模内鉴定候选菌株或“匹配物”,这些菌株在生物反应器、罐和中试工厂的大量(1L至1000L)培养基中培养,以测定脂肪醇或脂肪物质的准确的生产水平。本领域的那些技术人员使用这些大量培养条件以优化培养条件,从而获得所需的脂肪醇或脂肪物质的生产。

[0119] 脂肪醛和脂肪醇组合物的用途

[0120] 醛用于生产许多专门的化学品。例如醛用于生产聚合物、树脂(例如酚醛树脂)、染料、风味剂、增塑剂、香精、药物和其他化学品,其中的一些物质可以用作溶剂、防腐剂或消毒剂。此外,某些天然和合成的化合物(例如维生素和激素)为醛,并且许多糖包括醛基。脂肪醛可以通过化学或酶还原而转化为脂肪醇。脂肪醇具有许多商业用途。全世界,脂肪醇及其衍生物的每年销售量超过10亿美元。短链脂肪醇用于化妆品和食品工业,例如乳化剂、软化剂和增稠剂。由于脂肪醇的两性性质,其可作为非离子表面活性剂,可用于个人护理和家庭产品,例如洗涤剂。此外,脂肪醇用于蜡、树胶、树脂、药物软膏和洗剂、润滑油添加剂、纺织品防静电和整理剂、增塑剂、化妆品、工业溶剂和用于脂肪的溶剂。此外,本发明公开还提供了表面活性剂组合物或洗涤剂组合物,其包括通过本发明所述的任一种方法生产的脂肪醇。本领域的任一普通技术人员将理解的是根据表面活性剂或洗涤剂组合物的预期目的,可以生产和使用不同的脂肪醇。例如当本发明所述的脂肪醇用作生产表面活性剂或洗涤剂的供料时,本领域的任一普通技术人员将理解的是脂肪醇供料的特征将影响所生产的表面活性剂或洗涤剂组合物的特征。因此,可以通过生产特定的脂肪醇作为供料来选择表面活性剂或洗涤剂组合物的特征。本发明所述的脂肪醇基表面活性剂和/或洗涤剂组合物可以与本发明公知的其他表面活性剂和/或洗涤剂混合。在一些实施方案中,混合物可以包括脂肪醇按重量计至少大约10%、至少大约15%、至少大约20%、至少大约30%、至少大约40%、至少大约50%、至少大约60%或者由上述任意两个值所限定范围。在其他的实例中,可以制备表面活性剂或洗涤剂组合物,其包括脂肪醇的按重量计至少大约5%、至少大约10%、至少大约20%、至少大约30%、至少大约40%、至少大约50%、至少大约60%、至少大约70%、至少大约80%、至少大约85%、至少大约90%、至少大约95%或者由上述任意两个值所限定范围,该脂肪醇包括长度为8、9、10、11、12、13、14、15、16、17、18、19、20、21或22个碳的碳链。此外,此类表面活性剂或洗涤剂组合物还可以包括至少一种添加剂,例如由非微生物来源(例如植物油或石油)得到的微乳剂、表面活性剂或洗涤剂,该添加剂可以以占脂肪醇的按重量计至少大约5%、至少大约10%、至少大约15%、至少大约20%、至少大约30%、至少大约40%、至少大约50%、至少大约60%、至少大约70%、至少大约80%、至少大约85%、至少大约90%、至少大约95%或者由上述任意两个值所限定范围的量存在。通过以下实施例进一步说明本发明公开。提供这些实施例仅为了说明的目的。它们不应该被解释成以任何方式限定了本发明公开的范围或内容。

实施例

[0121] 实施例1

[0122] 生产型宿主的修饰—酰基-CoA脱氢酶的衰减

[0123] 本实施例描述了基因改造的宿主细胞的构建,其中脂肪酸降解酶的表达被衰减。

使用Datsenko et al, Proc. Natl. Acad. Sci. USA 97:6640-6645 (2000)所述的Lambda Red (还称为Red驱动整合)来选择大肠埃希氏杆菌MG1655 (大肠杆菌K菌株)的fadE基因,其具有以下修饰:

[0124] 使用以下两个引物创建fadE的删除:

[0125] Del-fadE-F

[0126] 5'-AAAAACAGCAACAATGTGAGCTTTGTTGTAATTATATTGTAAACATATTGATTCCGGGGATCCGTC GACC (SEQ ID NO:9); 和

[0127] Del-fadE-R

[0128] 5'-AAACGGAGCCTTTCGGCTCCGTTATTCATTTACGCGGCTTCAACTTTCCTGTAGGCTGGAGCTGCT TC (SEQ ID NO:10)

[0129] 使用Del-fadE-F和Del-fadE-R引物通过PCR来扩增由质粒pKD13 (如Datsenko et al, 如上所述)得到的卡那霉素抗性(KmR)盒。然后,使用PCR产物转化包括pKD46 (如Datsenko et al, 如上所述)的电感受态大肠杆菌MG1655细胞,其中所述的细胞事先使用阿拉伯糖诱导3至4小时。在37°C下在具有代谢物抑制(SOC)培养基的特级优化肉汤中生长3小时后,将细胞平板接种于包括50µg/mL卡那霉素的Luria琼脂平板上。在37°C下温育过夜后,鉴定并分离抗性菌落。使用引物fadE-L2和fadE-R1,通过PCR扩增来证明fadE基因的破坏,其中所述的引物被设计位于大肠杆菌fadE基因的侧翼。

[0130] 证明fadE被删除的引物为:

[0131] fadE-L2 5'-CGGGCAGGTGCTATGACCAGGAC (SEQ ID NO:11); 和

[0132] fadE-R1 5'-CGCGGCGTTGACCGGCAGCCTGG (SEQ ID NO:12)

[0133] 在证明fadE被删除之后,使用Datsenko et al., 如上所述的pCP20质粒除去单一菌落的KmR标记。fadE基因被删除并且KmR标记被除去的所得MG1655大肠杆菌菌株命名为大肠杆菌MG1655 Δ fadE或大肠杆菌MG 1655D1。将由fadE基因被删除的MG1655大肠杆菌菌株得到的脂肪酸衍生物(“总脂肪物质”)生产与由大肠杆菌MG1655得到的脂肪酸衍生物生产比较。使用生产型质粒pDG109 (pCL1920_P_{TRC}_carBopt_12H08_alrAadp1_fabB[A329G]_fadR)转化细胞,并在葡萄糖基本培养基中发酵。图5中提供的数据显示fadE基因的删除不影响脂肪酸衍生物的生产。

[0134] 实施例2

[0135] 通过脂肪酸合成途径的流量增加—乙酰基CoA羧化酶所介导

[0136] 用于脂肪酸生物合成的主要前体为丙二酰基-CoA和乙酰基-CoA (图1)。已经表明这些前体限制了在大肠杆菌中脂肪酸生物合成的速率 (图2)。在本实施例中,合成acc操纵子[谷氨酸棒状杆菌accABCD (± birA)]是过表达的,并且基因修饰导致在大肠杆菌中乙酰基-CoA和丙二酰基-CoA的生产增加。在一种方法中,为了增加丙二酰基-CoA的水平,由谷氨酸棒状杆菌得到的乙酰基-CoA羧化酶复合物在大肠杆菌中是过表达的。乙酰基-CoA羧化酶(acc)由4个独立的亚基accA、accB、accC和accD构成 (图3)。谷氨酸棒状杆菌acc的优点在于2个亚基作为融合蛋白质表达,分别为accCB和accDA,这有利于其平衡的表达。此外,谷氨酸棒状杆菌birA (其将accB亚基生物素化)是过表达的 (图3)。实施例3描述了acc基因与完整的fab操纵子的共表达。

[0137] 实施例3

[0138] 通过脂肪酸合成途径的流量增加—iFAB

[0139] 脂肪酸衍生物的生产：

[0140] 增加重组宿主细胞中通过脂肪酸合成途径的流量的策略包括在大肠杆菌中过表达天然大肠杆菌脂肪酸生物合成基因和表达由不同有机体得到的外源脂肪酸生物合成基因。在本研究中，将由不同有机体得到的脂肪酸生物合成基因与大肠杆菌DV2的基因组相组合。评价包括iFAB130-145的16种菌株。iFAB130-145的详细结构示于以下iFAB表1中。

[0141] 表1：在iFAB130-145中发现的成分

[0142]

缩写	完整描述
St_fabD	鼠伤寒沙门氏杆菌fabD基因
nSt_fabH	具有天然RBS的鼠伤寒沙门氏杆菌fabH基因
sSt_fabH	具有合成RBS的鼠伤寒沙门氏杆菌fabH基因
Cac_fabF	丙酮丁醇梭菌 (ATCC824) fabF基因
St_fabG	鼠伤寒沙门氏杆菌fabG基因
St_fabA	鼠伤寒沙门氏杆菌fabA基因
St_fabZ	鼠伤寒沙门氏杆菌fabZ基因
BS_fabI	枯草芽孢杆菌fabI基因
BS_FabL	枯草芽孢杆菌fabL基因
Vc_FabV	霍乱弧菌fabV基因
Ec_FabI	大肠埃希氏杆菌fabI基因

[0143] 各“iFAB”包括多个fab基因，其顺序如下：1) 烯酰基-ACP还原酶 (BS_fabI、BS_FabL、Vc_FabV或Ec_FabI)；2) b-酮脂酰基-ACP合成酶III (St_fabH)；3) 丙二酰基-CoA-ACP酰基转移酶 (St_fabD)；4) b-酮脂酰基-ACP还原酶 (St_fabG)；5) 3-羟基-酰基-ACP脱水酶 (St_fabA或St_fabZ)；6) b-酮脂酰基-ACP合成酶II (Cac_fabF)。注意St_fabA还具有反-2，顺-3-癸烯酰基-ACP异构酶活性 (ref)，并且Cac_fabF具有b-酮脂酰基-ACP合成酶II和b-酮脂酰基-ACP合成酶I活性 (Zhu et al., BMC Microbiology 9:119 (2009))。关于iFAB130-145的具体组成见下表2。参见图7A和B，其提供了iFAB138基因座的示意图，所述的iFAB138基因座包括整合在FAB138前面的cat-loxP-T5启动子的图 (7A)；以及iT5_138的图 (7B)。

[0144] 表2：iFAB130-145的组成

[0145]

ifab	BS fabI	BS fabL	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

[0146] 将质粒pCL_P_{trc}_tesA转化到各菌株中,并在20小时内FA2培养基中运行发酵,从而在32℃和37℃下完成诱导至发酵。由两个重复的平板筛选得到的总脂肪物质的生产数据示于图6A和6B中。通过该文库筛选,确定最佳的构建体为具有iFAB138的DV2。将iFAB138构建体转移至菌株D178中,从而制得菌株EG149。该菌株用于进一步的改造。EG149的基因组中,iFAB138的序列示于SEQ ID NO:13中。表3列出了多种大肠杆菌菌株的基因特征,其中在所述的菌株中,按照以下所述引入了包括本发明所述的表达构建体的质粒。这些菌株和质粒用于证明本发明公开的某些实施方案中的重组宿主细胞、培养物和方法。表3中的基因名称是本领域的那些普通技术人员已知的标准名称。

[0147] 表3:大肠杆菌菌株的基因特征

[0148]

菌株	基因特征
DV2	MG1655F ⁻ ,λ ⁻ ,ilvG ⁻ ,rph-50,rph-1,ΔfhuA::FRT,ΔfadE::FRT
DV2.1	DV2fabB::fabB[A329V]
D178	DV2.lentD::FRT_P _{T5} _entD
EG149	D178insH-11::P _{LACUV5} -iFAB138
V642	EG149rph ⁺
SL313	V642lacIZ::P _{AI} '_tesA/pDG109
V668	V642ilvG ⁺

[0149]

菌株	基因特征
LC397	V668lacIZ::P _{TRC} '_tesA(var)_kan
SL571	V668lacIZ::P _{TRC} '_tesA(var)_FRT
LC942	SL571attTn7::P _{TRC} '_tesA(var)
DG16	LC942/pLC56
V940	LC397/pV171.1
D851	SL571yijP::Tn5-cat/pV171.1

[0150] 质粒:pDG109、pLC56和pV171.1为具有carB和tesA的可变表达的pCL_P_{trc}_carB_tesA_alrA_fabB_fadR操纵子。iFAB138为SEQ ID NO:13。

[0151] 实施例4

[0152] 通过修复rph和ilvG突变来增加游离脂肪酸(FFA)产物的量

[0153] 在所述的菌株中校正ilvG和rph突变,从而更多地生产FFA。使用pCL_P_{trc}_tesA转化菌株D178、EG149和V668(表3)。在FA2培养基中在32℃下将发酵运行40小时,以比较具有pCL_P_{trc}_tesA的菌株的D178、EG149和V668的FFA生产。校正rph和ilvG突变使得具有pCL_P_{trc}_tesA的基础菌株的FFA生产增加116%。如图8所示,V668/pCL_P_{trc}_tesA比D178/pCL_P_{trc}_tesA或EG149/pCL_P_{trc}_tesA对照生产更多的FFA。由于FFA是LS9产物的前体,更多的FFA生产是新菌株可以生产更高水平的LS9产物的良好的指示剂。根据以下举例的标准的

FALC发酵方案运行发酵和提取。

[0154] 使用所选大肠杆菌菌株的冷冻细胞库的小瓶来接种125mL长颈振荡摇瓶(包括浓度为115 μ g/mL的壮观霉素抗生素)中的20mL LB培养液。将该摇瓶在32 $^{\circ}$ C下在定轨振荡器中温育大约6小时,然后将1.25mL培养液转移至500mL Erlenmeyer长颈振荡摇瓶的125mL低P的P FA2种子培养基(2g/L NH_4Cl 、0.5g/L NaCl 、3g/L KH_2PO_4 、0.25g/L $\text{MgSO}_4\cdot 7\text{H}_2\text{O}$ 、0.015g/L mM $\text{CaCl}_2\cdot 2\text{H}_2\text{O}$ 、30g/L葡萄糖、1mL/L微量矿物溶液(2g/L $\text{ZnCl}_2\cdot 4\text{H}_2\text{O}$ 、2g/L $\text{CaCl}_2\cdot 6\text{H}_2\text{O}$ 、2g/L $\text{Na}_2\text{MoO}_4\cdot 2\text{H}_2\text{O}$ 、1.9g/L $\text{CuSO}_4\cdot 5\text{H}_2\text{O}$ 、0.5g/L H_3BO_3 、和10mL/L浓HCl)、10mg/L柠檬酸铁、100mM Bis-Tris缓冲剂(pH 7.0)、和115 μ g/mL壮观霉素)中,并在32 $^{\circ}$ C下在振荡器上过夜温育。使用100mL这种低P的FA2种子培养基来接种5L Biostat Aplus生物反应器(Sartorius BBI),其中最初包括1.9L无菌的F1生物反应器发酵培养基。该培养基最初由3.5g/L KH_2PO_4 、0.5g/L $(\text{NH}_4)_2\text{SO}_4$ 、0.5g/L MgSO_4 的7水合物、10g/L无菌过滤的葡萄糖、80mg/L柠檬酸铁、5g/L酪蛋白氨基酸、10mL/L无菌过滤的痕量矿物溶液、1.25mL/L无菌过滤的维生素溶液(0.42g/L核黄素、5.4g/L泛酸、6g/L烟酸、1.4g/L吡哆醇、0.06g/L生物素和0.04g/L叶酸)和与种子培养基中使用的相同浓度的壮观霉素构成。在温度为33 $^{\circ}$ C、通风速率为11pm(0.5v/v/m)及溶解氧浓度为30%饱和度下,使用28%w/v氨水将培养物的pH保持在6.9,其中使用与DO控制和氧复苏级联的搅拌环。通过自动加入硅乳液基防沫剂(Dow Corning 1410)来控制起泡。

[0155] 基于2L的正常发酵体积,当在指数供料速率为0.3hr⁻¹至恒定的最大葡萄糖供料速率10-12g/L/hr下,初始培养基中的葡萄糖被几乎耗尽时(在接种后的大约4-6小时),开始供入由3.9g/L MgSO_4 的7水合物和600g/L葡萄糖构成的营养供料。当在培养物达到5AU(在接种后的大约3-4小时)的OD时,通过加入1M IPTG原液至最终浓度为1mM来诱导生物反应器中脂肪醇的生产。此后,每天两次对生物反应器取样,并在接种后大约72小时收获。将0.5mL良好混合的发酵液样品转移至15mL圆锥管(VWR)中,并与5mL醋酸丁酯充分混合。将该管多次倒转以便混合,然后进行剧烈的涡旋处理大约2分钟。接着,将所述的管离心5分钟,从而分离有机层和水层,并将有机层部分转移至用于气相色谱分析的玻璃瓶中。

[0156] 实施例5

[0157] 通过转座子诱变使得脂肪醇的生产增加—yijP

[0158] 为了改善大肠杆菌脂肪醇生产的效价、产率和生产率,实施转座子诱变和高通量的筛选,并对有利的突变测序。显示将转座子插入到yijP菌株中会改良菌株在摇瓶和分批供料发酵中的脂肪醇产率。SL313菌株生产脂肪醇。表3中提供该菌株的基因型。然后,对转座子克隆进行高通量的筛选,以测量脂肪醇的生产。简言之,将菌落挑选至包括LB的深孔平板中,过夜生长,接种至新鲜的LB中,并生长3小时,接种至新鲜的FA2.1培养基中,生长16小时,然后使用醋酸丁酯提取。使用BSTFA(N, O-双[三甲基甲硅烷基]三氟乙酰胺)衍生粗提取物,并使用GC/FID分析。所有培养基中均包括壮观霉素(100mg/L),从而保持对pDG109质粒的选择。通过挑选生产与对照菌株SL313相似的总脂肪物质的克隆来选择匹配物(hit),但所述的克隆比对照菌株具有更高百分率的脂肪醇物质和更低百分率的游离脂肪酸。菌株68F11被鉴定为匹配物,并使用FA2.1培养基在摇瓶发酵中证实。转座子匹配物68F11与对照菌株SL313的比较表明68F11比对照菌株生产更高百分率的脂肪醇物质,而两种菌株均生产相似效价的总脂肪物质。对匹配物68F11的单个菌落(称为LC535)测序,从而鉴定转座子插

入的位置。简言之,使用试剂盒ZR Fungal/Bacterial DNA MiniPrep™(Zymo Research Corporation,Irvine,CA),根据制造商提供的说明由10mL过夜LB培养物来纯化基因组DNA。使用位于转座子内部的引物,由转座子向外测序纯化的基因组DNA。

[0159] DG150

[0160] 5'-GCAGTTATTGGTGGCCCTTAAACGCCTGGTTGCTACGCCTG-3' (SEQ ID NO:14)

[0161] DG131 5'-GAGCCAATATGCGAGAACACCCGAGAA-3' (SEQ ID NO:15)

[0162] 据测定,菌株LC535在*yijP*基因中具有转座子插入(图18)。*yijP*编码保守的内部膜蛋白,其功能尚不清楚。*yijP*基因在操纵子内并与编码磷酸烯醇式丙酮酸羧化酶的*ppc*基因和编码功能未知的预测DNA结合转录调节因子的*yij0*基因共转录。转座子内部的启动子可能影响*yijP*、*ppc*和*yij0*的转录水平和定时,并且还可以影响相邻的基因*frwD*、*pf1C*、*pf1d*和*argE*。转座子盒内的启动子示于图18中,并且可以影响相邻基因的表达。在2个不同的日期在分批供料的发酵中评价菌株LC535。2次发酵证明LC535与对照SL313相比以更高的产率生产脂肪醇,并且根据碳输出,绝对产率改良1.3-1.9%。在不同的菌株V940中进一步评价*yijP*转座子盒,其中所述的菌株与菌株SL313相比以更高的产率生产脂肪醇。使用以下引物,由菌株LC535扩增*yijP::Tn5-cat*盒:

[0163] LC277

[0164] 5'-CGCTGAACGTATTGCAGGCCGAGTTGCTGCACCGCTCCCGCCAGGCAG-3' (SEQ ID NO:16)

[0165] LC278

[0166] 5'-GGAATTGCCACGGTGCGGCAGGCTCCATACGCGAGGCCAG

[0167] GTTATCCAACG-3' (SEQ ID NO:17)

[0168] 将该线性DNA电穿孔至菌株SL571中,并使用lambda red重组系统整合至染色体中。使用转座子区外的引物筛选菌落:

[0169] DG407

[0170] 5'-AATCACCAGCACTAAAGTGCGCGGTTGTTACCCG-3' (SEQ ID NO:18)

[0171] DG408

[0172] 5'-ATCTGCCGTGGATTGCAGAGTCTATTTCAGCTACG-3' (SEQ ID NO:19)

[0173] 使用生产型质粒pV171.1转化具有校正*yijP*转座子盒(图9)的菌落,从而产生菌株D851。在摇瓶发酵中,针对不包括*yijP*转座子盒的等基因系菌株V940,测试D851(V940*yijP::Tn5-cat*)。发酵结果显示*yijP*转座子盒赋予D851菌株比V940菌株生产更高百分率的脂肪醇,并生产与V940对照菌株相似效价的总脂肪物质。在2个不同的日期,在分批供料发酵中评价菌株D851。由这些发酵得到的数据示于表4中,其表明在5升分批供料发酵中,具有*yijP::Tn5-cat*转座子插入的菌株具有增高的总脂肪物质(“FAS”)的产率,以及增高的脂肪醇百分率(“FALC”)。“脂肪物质”包括FALC和FFA。

[0174] 表4:*yijP*转座子插入对FAS和FALC的效价和产率的影响

[0175]

菌株	FAS效价	FAS产率	FALC百分率	FALC产率
V940	68g/L	18.70%	95.00%	17.80%
D851	70g/L	19.40%	96.10%	18.60%
V940	64g/L	18.40%	91.90%	16.90%

D851	67g/L	19.00%	94.00%	17.80%
------	-------	--------	--------	--------

[0176] 罐式发酵方法:

[0177] 为了评估脂肪酸酯在发酵罐中的生产,使用所需菌株的甘油瓶将20mL LB+壮观霉素接种于摇瓶中,并在32℃下温育大约6小时。使用4mL LB培养物接种125mL Low PFA种子培养基(参见下文),然后将其在32℃的振荡器中温育过夜。使用50mL过夜培养物来接种1L发酵罐培养基中。在pH7.2和30.5℃下,在最大供料速率为16g/L/hr(葡萄糖或甲醇)的条件下,在pH稳态条件下运行发酵罐。

[0178] 表5:Low P FA种子培养基

[0179]

成分	浓度
NH ₄ Cl	2g/L
NaCl	0.5g/L
KH ₂ PO ₄	1g/L
MgSO ₄ ·7H ₂ O	0.25g/L
CaCl ₂ ·2H ₂ O	0.015g/L
葡萄糖	20g/L
TM2微量矿物溶液	1mL/L
柠檬酸铁	10mg/L
Bis Tris缓冲剂(pH 7.0)	100mM
壮观霉素	115mg/L

[0180] 表6:发酵罐培养基

[0181]

成分	浓度
(NH ₄) ₂ SO ₄	0.5g/L
KH ₂ PO ₄	3.0g/L
柠檬酸铁	0.034g/L
TM2痕量矿物溶液	10mL/L
酪蛋白氨基酸	5g/L
灭菌后加入	
MgSO ₄ ·7H ₂ O	2.2g/L
衡量维生素溶液	1.25mL/L
葡萄糖	5g/L
接种物	50mL/L

[0182] 实施例6

[0183] 将N末端60bp的融合标签加入到CarB(CarB60)上

[0184] 有多种方式来增加蛋白质的溶解性、稳定性、表达或功能。在增加CarB的溶解性的一种方法中,可以将融合标签克隆在所述的基因之前。在增加CarB的表达的另一种方法中,可以改变所述的基因的启动子或核糖体结合位点(RBS)。在本研究中,通过加入N末端60bp的融合标签来修饰carB(SEQ ID NO:7)。为了生成修饰的蛋白质(本发明也称为“CarB60”),

首先使用以下引物将carB克隆到pET15b载体中:

[0185] 5'-GCAATTCATATGACGAGCGATGTTTACGA-3' (SEQ ID NO:20);和

[0186] 5'-CCGCTCGAGTAAATCAGACCGAACTCGCG (SEQ ID NO:21)

[0187] pET15b—carB构建体包括直接在carB基因的上游的60个核苷酸:

[0188] 5'-ATGGGCAGCAGCCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCAT (SEQ ID NO:22)

[0189] 将融合标签形式的carB重新命名为carB60。然后,使用限制性酶NcoI和HindIII消化pET15b_carB60,并亚克隆到pCL1920衍生的载体OP80中,该载体使用相同的酶切割。将OP80质粒转化至菌株V324 (MG1655 Δ fadE::FRT Δ fhuA::FRT fabB::A329V entD::T5-entD lacIZ::PTRC-⁻TesA)中,从而评价FALC的生产。根据标准的过程(下文中概括)发酵菌株,并对总脂肪物质的效价和总脂肪醇的效价定量。图10示出,CarB60增加脂肪醇的效价,因此CarB60酶由多拷贝的质粒表达时具有比CarB更高的总细胞活性。

[0190] 为了评估脂肪醇在生产型菌株中的生产,使转化子在37℃下在补充有抗生素(100mg/L)的2ml LB发酵液中生长。在过夜生长之后,将40 μ l培养物转移至补充有抗生素的2ml新鲜LB中。在生长3小时后,将2ml培养物转移至包括20ml具有3%葡萄糖的M9培养基的125mL烧瓶中,其中所述的培养基补充有20 μ l痕量矿物溶液、10 μ g/L柠檬酸铁铵、1 μ g/L硫胺和抗生素(FA2培养基)。当培养物的OD₆₀₀达到1.0时,向各烧瓶中加入1mM IPTG。在37℃下生长20小时后,由各烧瓶中移除400 μ l样品,并使用400 μ l醋酸丁酯提取脂肪醇。为了进一步理解修饰的CarB活性的机制,由不包括⁻TesA (MG1655 Δ fadE::FRT Δ fhuA::FRT fabB::A329V entD::P_{T5}-entD)的菌株D178纯化CarB60。简言之,将pCL1920_carB60转化至菌株D178中,所述的质粒经改造用于脂肪醇的生产,并在37℃下在补充有壮观霉素(100 μ g/ml)的FA-2培养基中进行发酵。当培养物的OD₆₀₀达到1.6时,使用1mM异丙基- β -D-硫代半乳糖苷(IPTG)诱导细胞,并在37℃下温育另外的23h。就CarB60的纯化而言,通过在4℃下在4500rpm下离心20min来收获细胞。将细胞团(10g)悬浮于12ml BugBuster MasterMix (Novagen)和蛋白酶抑制剂混合物溶液中。通过法式压滤使细胞破裂,并将所得的匀浆在10000rpm下离心,从而除去细胞残渣。将Ni-NTA加入到所得的混合物中,并将悬浮液在4℃下在旋转振荡器上在100rpm下旋转1小时。将浆料倒入柱中,并收集流过物。使用处于50mM磷酸钠缓冲剂pH8.0(包括300mM NaCl)的10mM咪唑来洗涤Ni-NTA树脂,并使用处于50mM磷酸钠缓冲剂pH8.0(包括300mM NaCl)的20mM咪唑进一步洗涤。使用处于50mM磷酸钠缓冲剂pH8.0(包括300mM NaCl)的250mM咪唑洗脱CarB60蛋白质,并通过SDS-PAGE分析。将蛋白质针对处于50mM磷酸钠缓冲剂pH7.5中的20% (v/v) 甘油进行透析,得到的产率为大约10mg CarB60/升培养物。将蛋白质速冻并储存在-80℃下直至需要为止。

[0191] 多拷贝的质粒丰富地表达CarB蛋白质。其他的SDS-PAGE分析显示CarB60的表达比CarB更高。CarB60的较高的表达水平表明整合至大肠杆菌染色体中的carB60基因比相同位置处的carB基因生产更多的蛋白质。为了测试这种假说,将carB60基因整合至大肠杆菌染色体中。简言之,首先使用如下正向引物和反向引物由pCL_carB60扩增carB60基因,

[0192] 正向引物:

[0193] 5'-ACGGATCCCCGAATGCGCAACGCAATTAATGTaAGTTAGCGC-3' (SEQ ID NO:23);和

[0194] 反向引物:

[0195] 5'-TGCATCATCGCCATTGAATTCCTAAATCAGACCGAACTCGCGCAGG-3' (SEQ ID NO:24)。

[0196] 使用如下正向引物和反向引物由载体pAH56扩增第二PCR产物,正向引物:

[0197] 5'-ATTCCGGGGATCCGTCGACC-3' (SEQ ID NO:25);和

[0198] 反向引物:

[0199] 5'-AATGGCGATGACGCATCCTCACG-3' (SEQ ID NO:26)。

[0200] 所述的片段包括卡那霉素抗性盒、 λ attP位点和 γ R6k复制起点。使用InFusion试剂盒(Clontech)连接两种PCR产物,从而创建质粒pSL116-126。使用包括carB60基因的pSL116-126或包括carB基因的质粒F27转化包括整合形式的'TesA12H08和辅助质粒pINT的脂肪醇生产型菌株。如上文所述,根据用于摇瓶发酵的标准过程在FA2培养基中发酵这些菌株。为了表征和定量脂肪醇和脂肪酸酯,使用与火焰电离("FID")检测缀合的气相色谱("GC")。使用BSTFA(N,O-双[三甲基甲硅烷基]三氟乙酰胺)衍生粗提物,并使用GC/FID分析。使用上文所述的GC方法,通过注射各种浓度的合适可靠的参照物来进行定量,并且可以使用多种测试,包括但不限于气相色谱(GC)、质谱(MS)、薄层色谱(TLC)、高效液相色谱(HPLC)、液相色谱(LC)、与火焰电离检测器缀合的GC(GC-FID)、GC-MS和LC-MS。当针对多肽的表达进行测试时,可以使用诸如Western印迹和斑点印记之类的技术。

[0201] 20小时后的发酵结果示于图11中。两种菌株的总脂肪产物效价是相似的(2.4g/L总脂肪物质),然而,与不具有N末端标签的CarB相比,整合的CarB60将大部分C12和C14链长的游离脂肪酸转化成脂肪醇。这些数据表明表达CarB60的细胞具有更高的总细胞羧酸还原酶活性,并且可以将更多的FFA转化成脂肪醇。因此,当carB60整合到染色体中时,其为改良的carB模板,其提供了用于进化carB基因所需的活性,从而鉴定改良的carB变体。

[0202] 实施例7

[0203] CarB突变体的生成

[0204] 在某些处理条件下,CarB酶是脂肪醇生产中的限速步骤。为了经济地生产脂肪醇,努力增加CarB酶的活性。

[0205] 易错PCR文库筛选:

[0206] 在其中DNA聚合酶的复制忠实性较低的条件使用易错PCR进行随机诱变。将诱变的核酸克隆到载体中,在易错PCR之后进行高生产量筛选,从而发现增加游离脂肪酸向脂肪醇的转化的有利突变(下文将详细描述)。重要的残基进一步突变成其他的氨基酸。一定数量的单一氨基酸突变及突变的组合增加了将被转化成脂肪醇的脂肪物质的级份。简言之,使用Genemorph II试剂盒(Stratagene),通过易错PCR,在carB60opt基因中生产随机突变。分别在carB60opt的两个结构域中的一个结构中生成突变,以促进克隆。文库1包括carB60opt的最初759个残基,并使用以下引物通过易错PCR生成该文库:

[0207] HZ1175'-ACGGAAAGGAGCTAGCACATGGGCAGCAGCCATCATCAT-3'

[0208] (SEQ ID NO:27);和

[0209] DG2645'-GTAAAGGATGGACGGCGGTACCCGCC-3' (SEQ ID NO:28)。

[0210] 文库1的载体为质粒pDG115,该质粒使用酶NheI和PshAI消化。文库2包括carB60opt的最后435个残基,并使用以下引物通过易错PCR生成该文库:

[0211] DG2635'-CACGGCGGTGACCGCGTCCATCC-3' (SEQ ID NO:29);和

[0212] HZ1185'-TTAATTCGGGGATCCCTAAATCAGACCGAACTCGCGCAG GTC-3' (SEQ ID NO:

30)。

[0213] 文库2的载体为使用酶PshAI和BamHI消化的质粒pDG115。使用InFusion Advantage (Clontech) 将易错插入物克隆到载体中,并通过克隆菌株NEB Turbo (New England Biolabs) 传代。然后,将所述的文库转化至菌株EG442 (EG149Tn7::P_{TRC}-ABR lacIZ::P_{T50}-ABR) 中。然后,对易错carB60opt克隆进行高通量筛选,从而测量脂肪醇的生产。简言之,将菌落挑选至包括LB的深孔平板中,过夜生长,接种至新鲜的LB中,并生长3小时,接种至新鲜的FA-2.1培养基中,生长16小时,然后使用醋酸丁酯提取。使用BSTFA (N, O-双[三甲基甲硅烷基]三氟乙酰胺) 衍生粗提物,并使用标准的GC/FID方法进行分析。所有培养基中均包括壮观霉素 (100mg/L),从而保持对pDG115质粒的选择。通过挑选生产比对照菌株更低的总游离脂肪酸效价和更高的总脂肪醇效价来选择匹配物。为了比较由不同发酵筛选物得到的匹配物,通过计算归一化的游离脂肪酸百分率来归一化游离脂肪酸向脂肪醇的转化 (归一化的FFA=突变体FFA百分率/对照FFA百分率,其中“FFA百分率”为总的游离脂肪酸物质的效价除以总的脂肪物质的效价)。如下文所述,使用摇瓶发酵进一步证实匹配物。

[0214] 对匹配物测序以鉴定有利的突变。使用以下引物,通过完整carB60opt基因的菌落PCR来进行测序:

[0215] SL595' -CAGCCGTTTATTGCCGACTGGATG-3' (SEQ ID NO:31); 和EG4795' -CTGTTTTATCAGACCGCTTCTGCGTTC-3' (SEQ ID NO:32),并且使用carB60opt酶内部的引物进行测序。

[0216] 改良CarB60opt酶的有利突变示于表7中。归一化的游离脂肪酸 (归一化的FFA) 的栏表明酶的改良,并且更低的值表示最好的改良。“孔#”表示在其中发现这种突变的最初的筛选孔。所有的残基编号是指不包括60bp标签的CarB蛋白质序列。使用前缀“标签:”表示的突变表明在60bp/20个残基的N末端标签中具有突变。

[0217] 表7:在易错筛选过程中鉴定的CarB酶中的有利突变 (TAG突变被去除)

[0218]

孔#	归一化 FFA	错义突变	沉默突变
131B08	70.50%	L799M V810F S927R M1062L A1158V F1170I	CCG1116CCT
20C07	71.80%	A535S	
65B02	74.70%	M930R	ACC867ACA
54B10	76.30%	L200L I231M F288L A418I V530M A541V G677D P112A	
67E1	78.20%	D750G R827C D986G G1026D P1149S	GCA1081GCT GTC1073GTT
65C03	78.90%	V926A	ATT941ATA
12C10	80.30%	V46I	
66E08	80.30%	V926A	
70F02	80.90%	D750G R827C D986G G1026D P1149S	GCA1081GCT GTC1073GTT
07D01	82.40%	E20R V191A	
66G09	82.40%	R827C L1128S	ACG780ACA CTG923TTG
25H02	83.50%	F288S	
06C01	85.30%	V46I	08C01
05D02	85.20%	T396S	CCG477CCT
124E03	86.00%	R827C L1128S	ACG780ACA CTG923TTG
17A04	86.20%	A574T	GCA237GCT ACC676A CT GCC529GCT
132C08	87.00%	M1062T R1080H	TTG890TTA TAC834TAT
72C09	87.30%	P809L M1062V	
10F02	87.70%	E636K	
71H03	88.30%	R827C L1128S	ACG780ACA CTG923TTG
38G04	88.90%	D143E A612T	GCA181GCC
42F08	90.20%	T908M	CTG186CTT
66C04	90.30%	L1128S	
18C03	90.40%	Q473L	
12E02	90.60%	D19N S22N R87H L416S	CCG167CCA
28B09	91.30%	E28K H212N Q473L	CCG122CCA ACG178ACA CTG283TTG CTG340CTA ACC401ACT GCA881GCC
108E09	92.20%	E936K P1134R	CGT829CGG CTG1007CTA
08E09	93.20%	M259I	
74G11	93.80%	R70V S927I S985I I1164F	GTG1000GTC
46C01	95.60%	D18V D292N	

[0219] 饱和诱变(生成的结合1和2文库):

[0220] 在易错PCR之后被认为有利于脂肪醇生产的氨基酸位置经历进一步的诱变。包括兼并核苷酸NNK或NNS的引物用于将这些位置突变成其他的氨基酸。针对易错文库,如上文所述来筛选所得的“饱和诱变文库”,并鉴定进一步改良了脂肪醇转化(与亲代“对照”菌株相比具有更低的总游离脂肪酸效价和更高的总脂肪醇效价)的匹配物。在改良了脂肪醇生产的9个不同位置中的单一氨基酸/密码子变化示于表8中。如本发明所述,使用摇瓶发酵对匹配物进行证实。

[0221] 表8:在氨基酸饱和和诱变过程中鉴定的CarB酶中的有利突变

[0222]

WT氨基酸	WT密码子	突变体氨基酸	突变体密码子	归一化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%
		V	GTC	85.30%
		Y	TAC	88.80%
V191	GTC	A	GCC	88.70%

		S	AGT	98.00%
F288	TTT	G	GGG	70.30%
		R	AGG	77.20%
		S	TCT	85.60%
		S	AGC	79.60%
Q473	CAA	A	GCG	89.50%
		F	TTC	89.10%
		H	CAC	84.10%
		I	ATC	77.20%
		K	AAG	90.30%
		L	CTA	90.10%
		M	ATG	89.00%
		R	AGG	88.00%
		V	GTG	89.20%
		W	TGG	84.50%
		Y	TAC	86.00%
A535	GCC	A	TCC	71.80%
R827	CGC	A	GCC	93.20%
		C	TGT	87.90%
		C	TGC	83.20%
V926	GTT	A	GCT	78.10%
		A	GCG	66.30%
		A	GCC	69.50%
		E	GAG	65.80%
		G	GGC	78.60%
S927	AGC	G	GGG	77.60%
		G	GGT	79.30%
		I	ATC	90.80%
		K	AAG	70.70%
		V	GTG	87.90%
M930	ATG	K	AAG	82.30%
		R	CGG	73.80%
		R	AGG	69.80%
L1128	TTG	A	GCG	92.70%
		G	GGG	89.70%
		K	AAG	94.80%
		M	ATG	95.80%
		P	CCG	98.40%
		R	AGG	90.90%

		R	CGG	88.50%
		S	TCG	88.90%
		T	ACG	96.30%
		V	GTG	93.90%
		W	TGG	78.80%
		Y	TAC	87.90%

[0223] 接着,将被认为有利于脂肪醇生产的氨基酸取代组合。使用PCR来扩增包括多种所需突变的carBopt基因的部分,并使用PCR基方法将这些部分连接在一起(Horton, R.M., Hunt, H.D., Ho, S.N., Pullen, J.K. and Pease, L.R. 1989)。筛选不具有60bp的N末端标签的carBopt基因。在这种组合文库中组合的突变示于表9中。

[0224] 表9:由第一组合文库得到的CarB突变

突变	密码子
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

[0226] 为了促进筛选,将所得的CarB组合文库随后整合至菌株V668的染色体lacZ基因座中。该基因座中的carBopt基因序列示于SEQ ID NO:7中。菌株V668的基因型为MG1655 (Δ fadE::FRT Δ fhuA::FRT Δ fabB::A329V Δ entD::T5-entD Δ insH-11::P_{lacUV5} fabI38rph+ i1vG+) (如表3和图16所示)。然后使用质粒pVA3转化菌株,其中所述的质粒包括TesA(一种催化惰性的CarB酶—CarB[S693A],其破坏磷酸泛酰巯基乙胺的连接位点)和其他基因(其增加游离脂肪酸的生产)。针对易错文库,按照上文所述来筛选组合文库。在lacZ区中具有整合的carB opt (A535S) 并且包括pVA3的V668用作对照。如实施例5所述,选择增加脂肪醇生产的匹配物,并使用摇瓶发酵对匹配物进一步证实。在表达CarB组合突变体的重组宿主细胞经历摇瓶发酵之后,改良的脂肪醇生产的百分率示于图12中。

[0227] 使用以下引物,通过PCR,由整合的carB匹配物来扩增整合的CarB结合突变体:

[0228] EG585' -GACTCGACCGGAATTATCG (SEQ ID NO:33); 和

[0229] EG6265' -GACTACGCGTACTGTGAGCCAGAG (SEQ ID NO:34)。

[0230] 使用以下引物再扩增这些插入物:

[0231] DG2435' -GAGGAATAAACCATGACGAGCGATGTTACGACGCGACCGACGGC (SEQ ID NO:35); 和

[0232] DG2105' -CTAAATCAGACCGAACTCGCGCAGG (SEQ ID NO:36)。

[0233] 使用InFusion克隆,将合并的carB突变体克隆到生产型质粒pV869中,使用以下引物将其进行PCR扩增:

[0234] DG2285' -CATGGTTTATTCTCCTTATTTAATCGATAC (SEQ ID NO:37); 和

[0235] DG3185' -TGACCTGCGCGAGTTCGGTCTGATTAG (SEQ ID NO:38)。

[0236] 将在摇瓶发酵质粒筛选中表现最佳的carB突变体(carB2;表11)命名为VA101,并将携带carBopt[A535S]的对照菌株命名为VA82。参见图13。

[0237] 被认为对脂肪酸生产有利的carB还原结构域中的氨基酸取代与最佳的carB-L组合文库匹配物“carB3”之一组合(表11)。使用PCR来扩增在还原结构域中包括多种所需突变的carBopt基因的部分,并使用SOE PCR将这些部分连接在一起。在该组合文库中结合的突变示于表10中。

[0238] 表10:由第二组合文库得到的CarB突变

[0239]

突变	密码子
R827C	TGC
R827A	GCA
V926A	GCG
V926E	GAG
S927K	AAG
S927G	GGG
M930K	AAG
M930R	AGG
L1128W	TGG

[0240] 按照上文所述,针对易错文库来筛选结合文库。在lacZ区具有整合的carB3并包括pVA3的V668用作对照。按照上文所述,选择展示出脂肪醇生产增加的匹配物,并使用摇瓶发酵对其进一步证实。其中使用了进一步的CarB组合突变(carB4)而显示出脂肪酸生产百分率得到改良的摇瓶发酵结果示于表11中。低拷贝CarB变体的相对转化效率的图示示于图14中。表11中报告的结果得自在相同的条件下实施的生物反应器运行。

[0241] 表11:CAR变体

[0242]	名称	突变	菌株	发酵罐数据	注释
	carB	无=WT (E20V191F288Q473)			蛋白质为 SEQ ID NO: 7
	carB60	无+标签	V324		
	carB1	A535S	V940	83%FALC; C12/C14=3.4	具有一个拷贝的 12H08 染色体 TE
	carB2	E20R, F288G, Q473L, A535S	LH375	97%FALC; C12/C14=3.6	具有两个拷贝的 12H08 染色体 TE
	carB2	E20R, F288G, Q473L, A535S	LH346	96%FALC; C12/C14=3.7	具有以个拷贝的 12H08 染色体 TE
	carB3	E20R, F288G, Q473H, A535S	L 结合文库	到目前为止无在生物 反应器中运行的实例	
	carB4	E20R, F288G, Q473H, A535S, R827A, S927G	R 结合文库 (VA-219)	97%FALC; C12/C14=3.9	具有两个拷贝的 12H08 染色体 TE
	carA	无	参见美国专利公开 号 20100105963		蛋白质为 SEQ ID No. 39
	FadD9	无	参见美国专利公开 号 20100105963		蛋白质为 SEQ ID No. 40

[0243] CarA、FadD9、CarB和CarB60的DNA序列在本发明中分别示于SEQ ID NO: 41, 42, 43和44中。

[0244] 鉴定通过饱和诱变在CarB酶中得到的其他有利突变:

[0245] 后来研发出双质粒筛选系统,并且证实就FALC生产而言,该系统鉴定出超过CarB4的改良的CarB变体。双质粒系统满足以下标准:1) 突变体克隆生产高的FA效价,从而提供超过CarB活性的脂肪酸流量。这是通过使用质粒(pLYC4, pCL1920_P_{TRC}_carDead_tesA_alrAadp1_fabB[A329G]_fadR)转化基础菌株(具有两个拷贝的染色体TE的V668)来实现的,其中所述的质粒携带具有催化惰性的CarB酶(CarB[S693A])的FALC操纵子,从而增强游离脂肪酸的生产;2) 筛选具有carB突变体模板的质粒(优选小于9kb)能够经受饱和诱变过程,并与使用pLYC4的表达是相容的;3) CarB的动态量程是可调的。可以通过将较弱的启动子(P_{TRC1})与备选的起始密码子(GTG或TTG)组合来完成,从而协调CarB4的表达水平;4) 良好的质粒稳定性,引入毒素/抗毒素模块(ccdBA操纵子),从而保持质粒的稳定性。

[0246] 简言之,使用In-Fusion HD克隆方法(Clontech),通过将等摩尔比的4个部分(P_{TRC1}, 具有ATG/TTG/GTG起始密码子的carB4, 具有ccdAB的rrnB T1T2终止子,和pACYCDuet-1载体)混合来由这4个部分构建筛选质粒pBZ1(pACYCDuet-1_P_{TRC1}-carB4GTG_rrnBter_ccdAB)。通过以下引物对来PCR扩增所述的部分(1至4):

[0247] (1) P_{TRC1}-

[0248] 正向引物

[0249] 5'CGGTTCTGGCAAATATTCTGAAATGAGCTGTTGACAATTA

[0250] ATCAAATCCGGCTCGTATAATGTGTG-3' (SEQ ID NO: 45), 和

[0251] 反向引物5'-GGTTTATTCCTCCTTATTTAATCGATACAT-3' (SEQ ID NO: 46), 其中使用pVA232(pCL1920_P_{TRC}_carB4_tesA_alrAadp1_fabB[A329G])

- [0252] fadR) 质粒作为模板;
- [0253] (2) 具有ATG/TTG/GTG起始密码子的carB4-
- [0254] 正向引物carB4ATG
- [0255] 5' ATGTATCGATTAAATAAGGAGGAATAAACCATGGGCACGAGCGATGTTACGACGCGAC-3' (SEQ ID NO:47),
- [0256] carB4GTG
- [0257] 5' ATGTATCGATTAAATAAGGAGGAATAAACCGTGGGCACG
- [0258] AGCGATGTTACGACGCGAC-3' (SEQ ID NO:48), 和
- [0259] carB4TTG
- [0260] 5'-ATGTATCGATTAAATAAGGAGGAATAAACCTTGGGCACG
- [0261] AGCGATGTTACGACGCGAC-3' (SEQ ID NO:49), 以及
- [0262] 反向引物carB4rev5'-TTCTAAATCAGACCGAACTCGCGCAG-3' (SEQ ID NO:50), 其中使用pVA232质粒作为模板;
- [0263] (3) 具有ccdAB的rrnB T1T2终止子
- [0264] 正向引物rrnB T1T2项
- [0265] 5'-CTGCGCGAGTTCGGTCTGATTTAGAATTCCTC
- [0266] GAGGATGGTAGTGTGG-3' (SEQ ID NO:51), 和
- [0267] 反向引物ccdAB rev5'-CAGTCGACATACGAAACGGGAATGCGG-3' (SEQ ID NO:52), 其中使用质粒pAH008 (pV171_ccdBA操纵子);
- [0268] (4) pACYCDuet-1载体骨架-
- [0269] 正向引物pACYC载体, 用于
- [0270] 5' CCGCATTCCCGTTTCGTATGTCGACTGA
- [0271] AACCTCAGGCATTGAGAAGCACACGGTC-3' (SEQ ID NO:53), 和反向引物pACYC载体rev
- [0272] 5'-CTCATTTTCAGAATATTTGCCAGAACCGTTA
- [0273] ATTTCTAATGCAGGAGTCGCATAAG-3' (SEQ ID NO:54)。
- [0274] pBZ1质粒与pLYC4在上文所述的菌株中共表达, 并通过摇瓶和深孔平板发酵证实。优化发酵条件, 使得CarB4_GTG模板在两种发酵平台中可重复地具有~65%的FALC转化, 如实施例5所述。摇瓶发酵的结果示于图15中。
- [0275] 在改良的CarB变体(表7) 中包括突变的其他位点(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) 经历完全的饱和诱变。通过基于PCR的方法将包括兼并核苷酸NNK或NNS的引物用于将这些位置突变成其他的氨基酸(Sawano and Miyawaki 2000, Nucl. Acids Res. 28:e78)。使用pBZ1 (pACYCDuet-1_P_{TRC1}-carB4GTG_rrnB_{ter}-ccdAB) 质粒模板来构建饱和文库。将突变体克隆转化至NEB Turbo (New England Biolab) 克隆菌株中, 并且分离和合并质粒。然后, 将合并的质粒转化至携带质粒pLYC4的基于V668的菌株中, 在补充抗生素(100mg/L壮观霉素和34mg/L氯霉素)的LB琼脂平板上选择转化子。
- [0276] 然后, 针对脂肪醇的生产来筛选由饱和文库得到的CarB变体。根据实施例5中所述的修饰深孔平板发酵方案将单一菌落直接挑选至96孔板上。通过挑选生产比对照菌株更低

的总游离脂肪酸效价和更高的总脂肪醇效价来选择匹配物。为了比较由不同发酵批次得到的匹配物,通过计算归一化的游离脂肪酸百分率来归一化游离脂肪酸向脂肪醇的转化。此外,如实施例5所述,在匹配物的证实中还使用归一化的FFA(%)。归一化的FFA(%) = 突变体FFA百分率/对照FFA百分率,其中“FFA百分率”为总的游离脂肪酸物质的效价除以总的脂肪物质的效价。使用如实施例5所述的摇瓶发酵对匹配物进一步证实。归一化的游离脂肪酸(归一化的FFA)的栏表明酶的改良,并且更低值表示最好的改良。“匹配物ID”表示在其中发现更低的归一化FFA表现型的、最初的筛选平板孔。通过测序PCR产物来鉴定匹配物的突变,其中所述的产物是使用突变体carB基因特异性的引物(BZ1,用于5'-GGATCTCGACGCTCTCCTT-3'(SEQ ID NO:55))和BZ12_ccdAB独特的引物(5'-TCAAAAACGCCATTAACCTGATGTTCTG-3'(SEQ ID NO:56))由包括pBZ1质粒的“匹配物”扩增得到的。在证实的匹配物中鉴定的归一化FFA值和突变概括于表12中。

[0277] 表12:在氨基酸饱和诱变过程中鉴定的CarB4酶中的有利突变

[0278]

WT氨基酸	WT密码子	匹配物ID(氨基酸)	突变体密码子	归一化FFA(%)
D1.8	GAT	p10H5 (R)	AGG	75.5
		p684 (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

[0279] 通过完全的组合诱变来鉴定CarB酶的新变体:

[0280] 构建完全的结合文库,其包括以下氨基酸残基:18D、18R、22S、22R、473H、473I、827R、827C、870I、870L、926V、926A、926E、927S、927K、927G、930M、930K、930R、1128L和1128W。通过基于PCR的方法设计用于文库构建的、在所有位置处包括天然和突变体密码子的引物(Horton,R.M.,Hunt,H.D.,Ho,S.N.,Pullen,J.K.and Pease,L.R.1989)。在CarB2、CarB3和CarB4(20R、288G和535S)中保守的有利突变未改变,因此将克隆至pBZ1(修饰的pBZ1_P_{TRC1}_carB2GTG_ccdAB)的carB2GTG用作PCR模板。通过将PCR片段拼装至包括上述组合突变的CarB ORF中而完成文库的构建。然后,通过In-Fusion方法将突变体CarB ORF克隆至pBZ1骨架中(Clontech)。将In-Fusion产物沉淀,并直接电穿孔至携带质粒pLYC4的筛选菌株中。按照实施例5所述进行文库筛选、深孔平板和摇瓶发酵。具有特定组合突变的CarB突变体的活性(归一化的FFA通过CarB2归一化,100%)概括于表13中。包括CarB2、CarB4和CarB5(CarB4-S22R)作为对照。归一化的FFA栏表明CarB酶的改良,并且更低值表示最好

的改良。此外,还示出相对于对照 (CarB2) 的改良倍数 (X-FI0C)。所列出的所有突变都是相对于CarB wt的多肽序列 (SEQ ID NO:7) 而言的。例如CarB1具有A535S突变,并且CarBDead (催化惰性的CarB酶) 携带S693A突变,其破坏了磷酸泛酰巯基乙胺的连接位点。

[0281] 用于在生物反应器中改良地生产脂肪醇的新CarB变体:

[0282] 鉴定表13中列出的新CarB变体的目的是将它们用于改良的脂肪醇生产。表13上部的CarB变体 (P06B6-S3R、E20R、S22R、F288G、Q473H、A535S、R873S、S927G、M930R、L1128W) 在位置3处携带自发的突变 (野生型AGC突变为AGA)。形成P06B6CarB变体 (即,CarB7 (在位置3处由AGA编码氨基酸R——S3R、E20R、S22R、F288G、Q473H、A535S、R873S、S927G、M930R、L1128W)) 和CarB8 (在位置3处由AGC编码野生型氨基酸S——E20R、S22R、F288G、Q473H、A535S、R873S、S927G、M930R、L1128W), 并将其克隆到低拷贝数的脂肪醇生产型质粒骨架 pCL1920中,从而生成仅在CarB中不同的后面的脂肪醇操纵子。用于所有CarB变体 (CarB2、CarB7和CarB8) 的翻译起始密码子 (GTG) 恢复成ATG,从而使表达最大化。

[0283] pCL1920_P_{TRC}_carB2_tesA_alrAadp1_fabB[A329G]_fadR

[0284] pCL1920_P_{TRC}_carB7_tesA_alrAadp1_fabB[A329G]_fadR

[0285] pCL1920_P_{TRC}_carB8_tesA_alrAadp1_fabB[A329G]_fadR

[0286] 将上述质粒转化至具有一个拷贝的染色体TE的V668基菌株中,并如实施例4所述将所得的菌株在生物反应器中筛选。CarB7和CarB8超过CarB2的改良 (通过生物反应器发酵产物中的脂肪醇%来测量) 示于图16中。活性的顺序为CarB7>CarB8>CarB2。CarB7的位置3突变 (AGC突变成AGA R稀有密码子) 赋予比CarB8更高的活性,此外,总的可溶性蛋白质的SDS-PAGE分析表明CarB7的表达比CarB8和CarB2更高。CarB2和CarB8的表达水平是相似的。这符合在实施例6中所述的CarB60数据,位置3的AGA R稀有密码子突变和在其N末端的CarB60标签都可以改良CarB的表达。应该理解的是通过改造宿主菌株和/或改造脂肪醇生产操纵子的其他成分使得CarB7和CarB8在游离脂肪酸流量升高的菌株中比CarB2更好地发挥作用。

[0287] 表13:在双质粒系统中由组合文库鉴定的CarB变体的概括

[0288]

突变体	归一化FFA (%)	X-FI0C	突变
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,288G,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
CarBDead	ND		S693A

[0289] 除非本发明中另作说明,或者内容中另外清楚地作出反驳,否则本发明所述的所有方法以任何合适的顺序实施。除非另外作出要求,否则本发明提供的任何和所有的实例、或者示例性语言(“例如”)的使用仅是为了更好地说明本发明公开,而未对本发明公开的范围提出限定。在说明书中,没有语言被解释成将任何未要求的元件指示成实施本发明公开所必需的。应该理解的是本发明使用的术语仅是为了描述具体的实施方案,并且无意于进行限定。本发明公开的优选实施方案在本发明中有所描述。通过阅读上述说明,这些优选实施方案的变体对于本领域的那些普通技术人员而言是显而易见的。本发明人希望技术人员能够适当地使用这些变体,并且本发明人希望除了本发明具体描述的以外的那些公开都被实施。因此,本发明公开包括所附权利要求书中所述的主题事件的所有修改和等价物,这是适用法律所允许的。此外,除非本发明中另作说明,或者内容中另外清楚地作出反驳,否则上述元件在所有可能变体中的任何结合都被涵盖在本发明公开的范围內。

序列表

<110> LS9, INC.

<120> CAR酶及脂肪酶的改良生产

<130> LS00039PCT

<140>

<141>

<150> 61/619,309

<151> 2012-04-02

<160> 56

<170> PatentIn version 3.5

<210> 1

<211> 1232

<212> DNA

<213> 人工序列

<220>

<221> 来源

<223> /注="人工序列的描述: 合成的多核苷酸"

[0001]

```

<400> 1
ttgtccatct ttatataatt tgggggtagg gtgttcttta tgtaaaaaa acgttttagg      60
atgcatafgg cggecgcata acttcgtata gcatacatta tacgaagtta tctagagtig      120
catgcctgea ggiccgtlla ttatcactta ttcaggcgta gcaaccagge gtllaaggge      180
accaataact gccttaaaaa aattacgccc cgccctgcca ctcatcgag tactgttgta      240
attcattaag cattctgccg acaatggaagc catcacaac ggcatgatga acctgaatcg      300
ccagcggeat cagcaccttg tcgcttgccg tataatatatt gcccatgggtg aaaacggggg      360
cgaagaagtt gtccataatt gccacgttta aatcaaaact ggtgaaactc acccagggat      420
tggctgagac gaaaaacata ttctcaataa accctttagg gaaataggcc aggttttcac      480
cgtaacacgc cacatcttgc gaatatatgt gtagaaactg ccggaaatcg tcgtggtatt      540
cactccagag cgaatgaaac gtttcagttt gctcatggaa aacggtgtaa caagggtgaa      600
cactatccca taccaccagc tcaccgtctt tcattgceat acggaattcc ggatgagcat      660
tcacagcgcg ggcaagaatg tgaataaagg ccggataaaa cttgtgctta tttttcttta      720
cggctctttaa aaaggccgta atatccagct gaacggtctg gttataggta catigagcaa      780
ctgactgaaa tgcctcaaaa tgttctttac gatgccattg ggatatatca acggtggtat      840
atccagtgat tttttctcc attttagctt ccttagctcc tgaaaatctc gataactcaa      900
aaaatacgcc cggtagtgat cttatttcat tatggigaaa gitggaacct cttacgtgcc      960
gatcaacgtc tcatttttgc caaaagttag cccagggtct cccggtatca acagggaacac      1020
caggatttat ttattctgcg aagtgtctt ccgtcacagg tattttatcg actetagata      1080
acttcgtata gcatacatta tacgaagtta tggatccagc ttatcgatac cgtcaaaaa      1140
atcataaaaa atttatttgc tttcaggaaa atttttctgt ataatagatt caattgcgat      1200
gacgacgaac acgcacctgc aggaggagac ca                                  1232

```

<210> 2

<211> 232

<212> DNA

<213> 人工序列

<220>
 <221> 来源
 <223> /注="人工序列的描述: 合成的多核苷酸"

<400> 2
 ttgtccatct ttatataatt tgggggtagg gtgttcctta tgtaaaaaaa acgttttagg 60
 atgcataatgg cggccgcata acttcgtata gcatacatta tacgaagtta tggatccagc 120
 ttatcgatac cgteaacaac atcataaaaa atttatattgc tticaggaaa attttctgt 180
 ataataagatt caattgcgat gacgacgaac acgcaectgc aggaggagac ca 232

<210> 3
 <211> 340
 <212> PRT
 <213> 不动杆菌菌种

<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

[0002]

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> 不动杆菌菌种

[0003]

<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

[0004] Lys Asn Ile Ala Pro Gln Ile Glu Met Tyr
 305 310

<210> 5

<211> 1020

<212> DNA

<213> 大肠杆菌

<400> 5

atgtcgaatga taaaaagcta tgccgcaaaa gaagcgggcg gcgaactgga agtttatgag	60
tacgatcccg gtgagctgag gccacaagat gttgaagtgc aggtggatta ctgcgggac	120
tgccattccg atctgtcgat gatcgataac gaatgggat tttcacaata tccgtggtt	180
gccgggcatg aggtgattgg gcgcgtggtg gcaactcgga gcgccgcgca ggataaaggt	240
ttgcaggteg gtcagcgtgt cgggattggc tggacggcgc gtagctgtgg tcactgcgac	300
gectgtatta gcggtaatca gatcaactgc gagcaagggt cggtgccgac gattatgaat	360
cgcggtggct ttgccgagaa gttgcgtgcg gactggcaat gggtgattcc actgccagaa	420
aatattgata tcgagtcgcc cgggccgctg ttgtgcggcg gtatcacggt ctttaaacca	480
ctgttgatgc accatatcac tgctaccagc cgcgttgggg taattggtat tggcgggctg	540
gggcatatcg ctataaaact tctgcacgca atgggatgcg aggtgacagc ctttagttct	600
aatccggcga aagagcagga agtgcgtgcg atgggtgccg ataaagtggg gaatagccgc	660
gatccgcagg cactgaaagc actggcgggg cagtttgatc tcattatcaa caccgtcaac	720
gtcagcctcg actggcagcc ctattttgag gcgtgacct atggcggtaa ttccatacg	780
gtcgggtcgg ttctcagccc gctgtctgtt ccggccitta cgttaattgc gggcgatcgc	840

agcgtctctg gttctgctac cggcaccgct tatgagctgc gtaagctgat gcgttttggc 900
 gcccgacgca aggttgcgcc gaccaccgaa ctgttccga tgcgaaaat taacgacgcc 960
 atccagcatg tgcgcgacgg taaggecgct taccgcgtgg tgttgaaagc cgatttttga 1020

<210> 6

<211> 1174

<212> PRT

<213> 人工序列

<220>

<221> 来源

<223> /注="人工序列的描述：合成的多肽"

<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

[0005]

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		
[0006]	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			
[0007]	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
[0008]	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> 耻垢分枝杆菌

<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

[0009]

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

[0010]

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
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
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
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 545	Gly Asn Ser Glu Arg 550	Ser Phe Leu Leu Ala Val 555 560
Val Val Pro Thr 565	Pro Glu Ala Leu Glu Gln Tyr Asp Pro Ala Ala Leu 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 595	Pro Ala Asp Phe Ile Val Glu Thr Glu Pro 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 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		

[0011]

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
[0012]		
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> 大肠杆菌		
	<400> 8		
	Met Val Asp Met Lys Thr Thr His Thr Ser Leu Pro Phe Ala Gly His		
	1	5	10
	Thr Leu His Phe Val Glu Phe Asp Pro Ala Asn Phe Cys Glu Gln Asp		
		20	25
	Leu Leu Trp Leu Pro His Tyr Ala Gln Leu Gln His Ala Gly Arg Lys		
		35	40
	Arg Lys Thr Glu His Leu Ala Gly Arg Ile Ala Ala Val Tyr Ala Leu		
		50	55
	Arg Glu Tyr Gly Tyr Lys Cys Val Pro Ala Ile Gly Glu Leu Arg Gln		
	65	70	75
[0013]	Pro Val Trp Pro Ala Glu Val Tyr Gly Ser Ile Ser His Cys Gly Thr		
		85	90
	Thr Ala Leu Ala Val Val Ser Arg Gln Pro Ile Gly Ile Asp Ile Glu		
		100	105
	Glu Ile Phe Ser Val Gln Thr Ala Arg Glu Leu Thr Asp Asn Ile Ile		
		115	120
	Thr Pro Ala Glu His Glu Arg Leu Ala Asp Cys Gly Leu Ala Phe Ser		
		130	135
	Leu Ala Leu Thr Leu Ala Phe Ser Ala Lys Glu Ser Ala Phe Lys Ala		
		145	150
	Ser Glu Ile Gln Thr Asp Ala Gly Phe Leu Asp Tyr Gln Ile Ile Ser		
		165	170
	Trp Asn Lys Gln Gln Val Ile Ile His Arg Glu Asn Glu Met Phe Ala		
		180	185
	Val His Trp Gln Ile Lys Glu Lys Ile Val Ile Thr Leu Cys Gln His		
		195	200
	Asp		
	<210> 9		
	<211> 70		
	<212> DNA		

[0014]	<213> 人工序列	
	<220>	
	<221> 来源	
	<223> /注="人工序列的描述: 合成的寡核苷酸"	
	<400> 9	
	aaaaacagca acaatgtgag ctttgttga attatattgt aaacatattg atfccggga	60
	tccgtcgacc	70
	<210> 10	
	<211> 68	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<221> 来源	
	<223> /注="人工序列的描述: 合成的寡核苷酸"	
	<400> 10	
	aaacggagcc ttteggctcc gttattcaat tacggcgctt caactttcct gtaggctgga	60
	gctgcttc	68
	<210> 11	
	<211> 23	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<221> 来源	
	<223> /注="人工序列的描述: 合成的寡核苷酸"	
	<400> 11	
	cgggcaggtg ctatgaccag gac	23
	<210> 12	
	<211> 23	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<221> 来源	
	<223> /注="人工序列的描述: 合成的寡核苷酸"	
	<400> 12	
	cgcggcggtg accggcagcc tgg	23
	<210> 13	
	<211> 5659	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<221> 来源	
	<223> /注="人工序列的描述: 合成的多核苷酸"	
	<400> 13	
	atgatcatca aacctaaaat tcgtggattt atctgtacaa caacgcaccc agtgggttgt	60
	gaagcgaacg taaaagaaca aattgcctac acaaaagcac aaggtccgat caaaaacga	120
	cctaagcgcg tgttggttgt cggatcgtct agcggctatg gtctgtcctc acgcatcgct	180
	gcggcggttg gcggtggtgc ggcgacgac gccgtatitt tcgaaaagcc gggcactgac	240
	aaaaaaccag gtactgcggg ttctacaat gcagcagcgt ttgacaagct agcgcataaa	300
	gcgggcttgt acgcaaaaag cctgaacggc gatgcgttct cgaacgaagc gaagcaaaaa	360
	gcgattgagc tgattaagca agacctcgcc cagattgatt tgggtggtta ctctgttgct	420

tctccagtgc gtaagatgcc agacacgggt gagctagtgc gctctgcaact aaaaccgatac	480
ggcgaaacgt acacctctac cgcggtagat accaataaag atgtgatcat tgaagccagt	540
gttgaacctg cgaccgagca agaaatcgct gacactgtca ccgtgatggg cggtaagat	600
tgggaactgt ggaaccaagc actggaagag gcgggtgttc ttgctgaagg ttgcaaaacc	660
gtggcgtaca gctacatcgg tactgaattg acttggccaa ttacttgga tgccgcttta	720
ggccgtgcc aagatggacct agatcgccga gcgacagcgc tgaacgaaaa gctggcagcg	780
aaaggtggtt ccgcgaacgt tgcagttttg aaatcagttg tgactcaagc aagctctcgc	840
attcctgtga tgcgctcta catcgcaatg gtgtcaaga agatcggtga acaggcggtg	900
catgaaggct gtatggagca gatctaccgc atgttcagtc aacgtctgta caaagaagat	960
ggttcagcgc cggaagtgga tgatcacaat cgtctgcgtt tggatgactg ggaactgcgt	1020
gatgacattc agcagcactg ccgtgatctg tggccacaaa tcaccacaga gaacctgcgt	1080
gagctgacgg attacgacat gtacaaagaa gatttcata agctgtttgg ctttggcatt	1140
gaaggcattg attacgatgc tgacgtcaat ccagaagtcg aatttgatgt gattgatata	1200
gagtaattta gtagcagc gtacatgta acgaagatta ttggtaicgg cagctatcgc	1260
cccgaacaag tgcgactaa cgcgactctg gaaaaaatgg ttgagacctc tgacgagtgg	1320
attgtcactc gtacaggtat tegttaacgc catatcgccg cgcgcaatga aactgtcgcg	1380
acgatgggct ttaccgtgc gaatcgcgcg attgagatgg cggggatcga taaagaccaa	1440
attggcttga ttgtgttggc taccacatca gcaacgatg catttccaag cgcggcatgt	1500
cagattcaaa gtatgctcgg tattaaaggt tgcgcggcgt ttgatgtcgc ggcagcgtgc	1560
gcaggtttca cctacgcgtt aagcatcgcc gaccagtacg ttaaatccgg cgcggttaaa	1620
caecgcgtgg ttgtcggctt cgatgtatta gccegcactt gegatcctgg cgatcgcggt	1680
acgatcatia ttitcgccga tggcgcagcg gcggccgtac tgagcgttc tgaagaaccg	1740
ggtattatct ccactatct tcatgccgat ggccgttacg gtgaattact gacctgcgcg	1800
aatgccgata gcgtaaatcc ggataaccg attacciga caatggcggg caatgaagtc	1860
tttaaatggt cgtcactga actggcgcat attgtcatg agacgctggc ggctaataac	1920
ctggatcgct cagaactcga ttggctggtg ccgcacagc ctaacctgcg tatcattagc	1980
gcgacagcga aaaaactcgg catgtcatg gacaatgtcg tcgtcacgct ggacaggcac	2040
ggcaataacct ccgcgcttc tctgccgtgc gcgttgatg aagccgtgcg tgacgggcga	2100
attaaagccg gtcagctggt attgcttga gccttcgggg gtggattcac ctggggctcc	2160
gcgtgatcc gtttctagta taaggattta aacatgacgc aatttgcatt tgtgttcccc	2220
ggtcagggtt ctcagagcgt tgggatgtg gccgagatgg cggcaaatia cctatcgtta	2280
gaagaaacgt ttgctgaagc ttctgcggct ctgggatatg atctgtgggc gctcaccag	2340
caaggtccag cggaagaact gaataaaacc tggcagacgc agccggcgtt attaacgct	2400
tccgtcgcgc ttggcgcggt ttggcagcag caggcggtta aaatgcctgc gttantgga	2460
ggtcacagcc tgggcgaata ttccgcgctg gtttgcgctg gcgtcataaa ctttctgat	2520
gccgttcgtc ttgttgaaat gcgcggtaaa ttcatgcagg aagcgggttc ggaaggcact	2580
ggcggaalgt ctgcgcatc cgggctggat gatgcctcta ttgctaaagc ctgtgaagaa	2640
tctgccgaag ggcaggttgt ttgcgcggtt aactttaact cgcgggaca ggtggttatc	2700

[0015]

[0016]

```

gccgggcata aagaggcggg agaactgtcg ggccagccct gtaaagccgc tggcgcgaaa 2760
cgcgcgctgc cgctgccggg gagcgtaccg tcgcactgcg cgctgatgaa accagcggca 2820
gataagctgg cggttgaatt agccaaaatt acctttagcg cgccaacggg gccggtagtg 2880
aacaacgttg acgtgaaatg tgaaacgat gccgccgcta tcgcgatgc gcctgttcgc 2940
cagtgtgaca atccggtaca gtggacgaag agcgtggaat ttatcgcggc gcaggcgctt 3000
gaacatcitt atgaagtggg tccaggtaaa gtcttcacig gtctgacgaa acgtaitgtc 3060
gacaccctga cagcgctggc gtgaacgag ccggcgcgcg tctctcggc acitacgcaa 3120
taaaagagga aaacctgag cttgaagga aagattgcg tggtagctgg tgcaagccgt 3180
ggcataggcc gcgaattgc agagactctc gtgcccgcg gcgcgaaagt tatcgggact 3240
gcgaccagtg aaaatggcgc gaagaacatt agcgactatt taggtgctaa cgggaaaggt 3300
ttgatgttga atgtgaccga tctgcatct attgaatctg ttctggaana taticgcgca 3360
gaatttgggt aagtgatata cctagttaat aatgccggtg tcaactcgtg taatctgttg 3420
atgcgaatga aagatgatga gtggaacgat attatcgaaa ccaacttate atccgttttc 3480
cgctgtcaca aagcggttaa gccgcctatg atgaaaaagc gttggttcgc callatcact 3540
attggttctg tggttgttac catgggaaat gcaggteagg caaactacgc tgcggcgaaa 3600
gcgggcctga tcggtttcag taaatcactg gcgcgtgaag ttgcgtcccg tggtaattact 3660
gtcaatgttg tggtccggg tttattgaa acggacatga cgcgtgcgt gtctgacgat 3720
cagcgtgcgg gtatcctggc gcaggtgcct gcgggtcgcc tcggcgcgcg tcaggaaatc 3780
gccagtgcgg ttgcattttt agcctctgac gaagcgagtt acaicactgg tgagactctg 3840
cacgtcaacg gcggaatgta catgttttaa ttttaaggtt tacataaaac atggtagata 3900
aacgcgaate ctatacaaaa gaagaccttc ttgcctctgg tcgtggtgaa ctgtttggcg 3960
ctaaagggcc gcaactccct gcaccgaaca tgcgatgat ggaccgcgtc gttaaagatga 4020
ccgaaacggg cgcaatttc gacaaaggct atgtcgaagc cgagctggat atcaatccgg 4080
atctatggtt ctccgatgc cactttatcg gcgacccgt gatgcccggt tctctgggtc 4140
tggatgtat gtggcaattg gtggatctt acctggcgtg gttggcgcg gcaggcaaaag 4200
gccgcgtctt ggcgtgggc gaagtgaat ttaccggcca ggttctgcg acagccagga 4260
aagtcacctg tcgtattcat ttcaaacgta tcgtaacgg tcgcctgac atgggectgg 4320
cggacggtag ggttctgggt gatggtcgcc tgatctatac cgcacacgat ttgaaagtcg 4380
gtttgttcca ggatacttc gcgttctaaa aggaggcaac aaaatgaate gccgcgttgt 4440
cattacgggt attggtgcag tgacgccggg gggtacaac gctgatagct tctggtgcag 4500
catcaaagag ggtaaatgtg gcattgacaa gatcaaagcg ttgacgcaa ccgatttcaa 4560
agttaagctg gctgccgaag tgaaggactt caccgccgag gactttatcg acaagcgtga 4620
ggcgaaccgt atggaccgtt ttaggcagtt tgcgatcgtt gcggcgcatg aggcfaatca 4680
ggacagcaaa ctggacctgg agtcgatga taagaatcgt ttccgctga ttgttgtag 4740
cggcattggc ggcacggca ccattgagaa gcaggatgaa aagctgatta ccaaaggtcc 4800
gggtcgtgtg agccctatga ctattccgat gatcattgcg aatatggcaa gcgtaatct 4860
ggcgtatcgt tatggcgcta aaggtattg cagcaccatt gtcaccgat gtgcgagcgc 4920
gaacaacagc attggtgagt ccttcgtaa cattaaattt ggttatagcg acgttatgat 4980

```

ctctggtggt agcgaagcag gtatcacccc gttgagcctg gcgggttttg cctcgatgaa 5040
ggccgtgacc aaatctgagg acccgaagcg cgccagcatc ccgttcgata aggatcgcag 5100
cggtttttgtg atgggcgagg gcagcggiat cgttatcttg gaagagttgg agcacgcgct 5160
gaagcgtggt gccaaaatct atgccgagat cgttggtat ggtgcgacct gcgacgcata 5220
tcatatcacg agcccagcgc cgaatggtga aggtggtgca cgtgcaatga aactggcaat 5280
ggaagaagat aatgtccgcc cagaggacat ttctataic aacgcgcacg gtacgagcac 5340
ggcgtacaat gacagcttg aaaccaagc gatcaagacg gtcttggtg aatacgcta 5400
caaagtgcg gtgtetagca ccaagagcat gaccggccac ctgctggcg ctggcggtgc 5460
agtcgaagcg attatcttg ccaaagctat tgaagagggt ttcattecg cgaccatcg 5520
ctacaaagag gcggatccgg aatgcgacct ggattacgtt cctaacgagg gccgtaatgc 5580
agnagtcaac tacgttctgt ccacagcct gggcttcggt ggccataatg cgactctgct 5640
gttcaaaaag tacaaatga 5659

<210> 14
<211> 40
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注="人工序列的描述: 合成的寡核苷酸"

<400> 14
gcagttattg gtgcccttaa acgcttggt gctacgcctg 40

[0017]

<210> 15
<211> 27
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注="人工序列的描述: 合成的寡核苷酸"

<400> 15
gagccaatat gcgagaacac ccgagaa 27

<210> 16
<211> 48
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注="人工序列的描述: 合成的寡核苷酸"

<400> 16
cgctgaacgt attgcaggcc gagttgctgc accgctcccg ccaggcag 48

<210> 17
<211> 51
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> /注="人工序列的描述: 合成的寡核苷酸"

<400> 17
ggaattgcca cggcgcggca ggctccatac gcgagccag gttatccaac g 51

	<210> 18 <211> 35 <212> DNA <213> 人工序列 <220> <221> 来源 <223> /注="人工序列的描述: 合成的寡核苷酸" <400> 18 aatcaccagc actaaagtgc gcggttcggtt acccg	35
	<210> 19 <211> 34 <212> DNA <213> 人工序列 <220> <221> 来源 <223> /注="人工序列的描述: 合成的寡核苷酸" <400> 19 atctgccgtg gattgcagag tctattcagc tacg	34
	<210> 20 <211> 30 <212> DNA <213> 人工序列 <220> <221> 来源 <223> /注="人工序列的描述: 合成的引物" <400> 20 gcaattccat atgacgagcg atgttcacga	30
[0018]	<210> 21 <211> 29 <212> DNA <213> 人工序列 <220> <221> 来源 <223> /注="人工序列的描述: 合成的引物" <400> 21 ccgctcgagt aaatcagacc gaactcgcg	29
	<210> 22 <211> 60 <212> DNA <213> 人工序列 <220> <221> 来源 <223> /注="人工序列的描述: 合成的寡核苷酸" <400> 22 atgggcagca gccatcatca tcatcatcac agcagcggcc tggtagccgc cggcagccat	60
	<210> 23 <211> 43 <212> DNA <213> 人工序列 <220> <221> 来源 <223> /注="人工序列的描述: 合成的寡核苷酸"	

[0019]	<400> 23 acggatcccc ggaatgcgca acgcaattaa tgtaagttag cgc	43
	<210> 24 <211> 46 <212> DNA <213> 人工序列	
	<220> <221> 来源 <223> /注="人工序列的描述: 合成的寡核苷酸"	
	<400> 24 tgcgtcatcg ccattgaatt cctaaatcag accgaactcg cgcagg	46
	<210> 25 <211> 20 <212> DNA <213> 人工序列	
	<220> <221> 来源 <223> /注="人工序列的描述: 合成的寡核苷酸"	
	<400> 25 attccgggga tccgtcgacc	20
	<210> 26 <211> 23 <212> DNA <213> 人工序列	
	<220> <221> 来源 <223> /注="人工序列的描述: 合成的寡核苷酸"	
	<400> 26 aatggcgatg acgcacccfc acg	23
	<210> 27 <211> 39 <212> DNA <213> 人工序列	
	<220> <221> 来源 <223> /注="人工序列的描述: 合成的引物"	
	<400> 27 acggaaagga gctagcacat gggcagcage catcatcat	39
	<210> 28 <211> 27 <212> DNA <213> 人工序列	
	<220> <221> 来源 <223> /注="人工序列的描述: 合成的引物"	
	<400> 28 gtaaaggatg gacggcggtc acccgcc	27
	<210> 29 <211> 25 <212> DNA <213> 人工序列	
	<220> <221> 来源	

	<223> /注="人工序列的描述: 合成的引物"	
	<400> 29 caaggcgggt gaccgcgctc catcc	25
	<210> 30 <211> 43 <212> DNA <213> 人工序列	
	<220> <221> 来源 <223> /注="人工序列的描述: 合成的引物"	
	<400> 30 ttaattccgg ggatecctaa atcagaccga actcgcgcag gtc	43
	<210> 31 <211> 24 <212> DNA <213> 人工序列	
	<220> <221> 来源 <223> /注="人工序列的描述: 合成的引物"	
	<400> 31 cagccgttta ttgcgcactg gatg	24
	<210> 32 <211> 27 <212> DNA <213> 人工序列	
[0020]	<220> <221> 来源 <223> /注="人工序列的描述: 合成的引物"	
	<400> 32 ctgtttatc agaccgcttc tgcgttc	27
	<210> 33 <211> 20 <212> DNA <213> 人工序列	
	<220> <221> 来源 <223> /注="人工序列的描述: 合成的引物"	
	<400> 33 gcactcgacc ggaattatcg	20
	<210> 34 <211> 25 <212> DNA <213> 人工序列	
	<220> <221> 来源 <223> /注="人工序列的描述: 合成的引物"	
	<400> 34 gcactacgcg tactgtgagc cagag	25
	<210> 35 <211> 45 <212> DNA <213> 人工序列	

<220>
 <221> 来源
 <223> /注="人工序列的描述: 合成的引物"

 <400> 35
 gaggaataaa ccatgacgag cgatgttcac gacgcgaccg acggc

 <210> 36
 <211> 25
 <212> DNA
 <213> 人工序列

 <220>
 <221> 来源
 <223> /注="人工序列的描述: 合成的引物"

 <400> 36
 ctaaaticaga ccgaactcgc gcagg

 <210> 37
 <211> 31
 <212> DNA
 <213> 人工序列

 <220>
 <221> 来源
 <223> /注="人工序列的描述: 合成的引物"

 <400> 37
 catggtttat tcctecttat ttaatcgata c

 <210> 38
 <211> 28
 <212> DNA
 <213> 人工序列

 <220>
 <221> 来源
 <223> /注="人工序列的描述: 合成的引物"

 <400> 38
 tgacctgcgc gagttcggtc tgatttag

 <210> 39
 <211> 1168
 <212> PRT
 <213> 耻垢分枝杆菌

 <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	
[0022]	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					
[0023]	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				
[0024]	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
 [0025] 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> 结核分支杆菌
 <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
[0026]	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		
[0027]	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		

[0028]

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
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
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
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	

[0029]

<210> 41

<211> 3507

<212> DNA

<213> 耻垢分枝杆菌

<400> 41

atgacgatcg aaacgcgcga agaccgcttc aaccggcgca ttgaccactt gttcgaaacc	60
gaaccgcagt tcgccgcgcg ccgtcccgac gagcgatca gcgcggetgc cgcgatccg	120
gagtgcgcgc ttctgcgcgc ggtcaaacag attctggcgc gctatgcgga ccgccctgcg	180
ctgggcaage gcgccgtcga gttcgtcacc gacgaagaag gccgcaccac cgcgaagctc	240
ctgccccgct tcgacaccat cacctaccgt cagctcgcag gccggatcca ggccgtgacc	300
aatgcctggc acaaccatcc ggtgaatgcc ggtgaccgcg tggecatcct gggtticacc	360
agtgtcgact acacgacgat cgacatcgcc ctgctcgaac tcggcgccgt gtcctgaccg	420
ctgcagacca gtgcgcgcgt ggcccaactg cagccgatcg tcgccgagac cgagcccaag	480
gtgatcgctg cgagcgtcga ctctctcgcc gacgcagtcg ctctcgtcga gtccgggccc	540
gcgcgcgtcg gactggtggt gttcgactac agccacgagg tcgacgatca gcgtgaggcg	600
ttcgaggcgg ccaaggcga gctcgcaggc accggcgtcg tcgtcgagac gatcaccgac	660
gcactggacc gggggcggtc actcgcgcgc gcaccgtctt acgtgcccga cgaggccgac	720
ccgctgacct ttctcatcta cactccggc agcaccggca ctcccaaggc cgegatgtac	780
cccagtgcca agaccgccac gatgtggcag gccgggtcca aggcccggtg ggacgagacc	840

[0030]

ctccggcgtga tgccgtcgat caccctgaac ttcatgccca tgagtcacgt catggggcgc 900
 ggeatcctgt gcagcacact cgcagcggc ggaaccgct acttcgcgc acgcagcgac 960
 ctgtccacct tectggagga cctcgccctc gtccggccca cgcagctcaa ctctgttctt 1020
 cgcattctggg acatgtctgt ccaggagta cagagccgc tcgacaaccg ccgcgcgag 1080
 ggcattcgagg accgagccga agccgcagtc ctgaagagg tcgcaccca actgctcggc 1140
 gggcgattcg ttctggccct gaccggaicg gtcccatct cggcggagat gaagagctgg 1200
 gtcgaggacc tctcgacat gcatctctg gaggctacg gctccaccga ggccggcgcg 1260
 gtgttcateg accggcagat ccagcgcgc ccggtcctg actacaagct ggtcgacgtg 1320
 cccgattctg gctacttcg caccgaccg cctaccgc gcggcgaact tctggtcaag 1380
 tccgagcaga tgttccccg ctactacaag cgtccggaga tcaccgcga gatgttcgac 1440
 gaggacgggt actaccgac cgcgcacatc gtcgccgagc tcgggcccca ccatctcgaa 1500
 taactcgacc gccgcaacaa cgtgctgaaa ctgtcgagg gcgaattcgt caccgtctcc 1560
 aagctggagg cgggtgttcg cagacgccc ctggtacgc agatctacgt ctacggcaac 1620
 agcgcgggt cclalcgtt ggccgctg gtcccgacc aagaggaact gtcacgtlgg 1680
 gaccgtgacg aactcaagtc gcgcacagc gactcaetg aggaacgggc acgagccgc 1740
 ggattgcagt cgtatgagat ccgcgtgac ttctctctg agacaacac ttacacgtg 1800
 gagaacggcc tctgaccgg taccgcaag ctggcccggc cgaaactgaa ggcgactac 1860
 ggcaacgcc tcgaacagt ctacaccgac ctggccgagg ggcaggccaa cgagttcgc 1920
 gaggttgcgc gcaacggagc cgaacggccc gtggtcgaga ccgicagccg cgcgcggtc 1980
 gcactgtctg gtgcctccgt caccgaictg cgttcctatg ccacttcac cgtctgggt 2040
 ggagattcgt tctcgccctt gagcttctg aacctgttc acgagatctt cgtgtctgac 2100
 gtcccgctg gcgtcatct cagcccgccc accgacctg caggcgtcgc ggctacatc 2160
 gaggcgcaac tgcgcggctc caagcgcgc acatacctg cgttcacgg gcgcgacgc 2220
 accgaggtgc gcgcgctga tctcgccctg ggcaagtca tcgacgcca gacctgtcc 2280
 gccgcgcgg gtctgcgcg ttcgggcacc gagatccga ccgtctgtg gaccggcgc 2340
 accgggttcc tgggcgcta tctggcgtg gaatggctg agcgcctgga cctggtggac 2400
 ggcaagtgta tctgcctggt gcgcgcgcgc agcgacgac aggccgggc gctctggac 2460
 gccacgttc acaccggga cgcgacactc ctcgagcact accgcgcgt ggcagccgat 2520
 caactcgagg tgatcgccg tgacaagggc gaggccgac tgggtctcga ccacgacag 2580
 tggcagcagc tggccgacac cgtcgatctg atctcgatc cggccgccc ggtcaatcac 2640
 gtctcgcgt acagccagat gtccggacc aatgcctcgc gcaacgcga actcatccg 2700
 atcgcctga ccaccagat caagccgtac gtgtacgtc cgacgatcg tgtggacag 2760
 ggcattctcc ccaggcgtt cgtcgaggac gccgaatcc gcgagatcag cgcgacgc 2820
 cgggtcgac actcgtacg caacggctac ggcaacgca agtgggcgg cgaggtctg 2880
 ctccgggagg cgcacgact gtgtggtct ccggtctcg tgttcgctg cgacatgatc 2940
 ctggccgaca cgacctact gggtcagct aacctgccg acatgttcac ccgctgatg 3000
 ctgagctcg tggcgaccg catcgccccc ggltcgtct acgaacgca tgggacggc 3060
 aaccggcagc gcgccacta cagcgggct ccggtggagt tcatcgcca ggcgatctcc 3120

	accatcggct cgcaggteac cgacggattc gagacgttcc acgtgatgaa cccgtacgac	3180
	gacggcatcg gcctcgacga gtacgtggac tggctgatcg aggccggcta ccccgtgcac	3240
	cgcgtcgaag actacgccac ctggctgagc cggttcgaaa ccgcaactgcg ggccctgccc	3300
	gaacggcaac gtcaggcctc gctgctgccc ctgctgcaca actatcagea gcccctaccg	3360
	ccgtgtgccc gtgccaatggc acccaccgac cggttccgtg ccgcgggtgca ggacgcgaag	3420
	atcgccccc acaaggacat tccgcacgtc acggccgacg tgatcgtaaa gtacatcagc	3480
	aaactgcaga tgctcggatt gctgttaa	3507
	<210> 42	
	<211> 3507	
	<212> DNA	
	<213> 结核分支杆菌	
	<400> 42	
	atgtcgatca acgatcagcg actgacacgc cgcgtcgagg acctatacgc cagcgacgcc	60
	cagttcgccc ccgccagtc caacgaggcg atcaccacgg cgatcgacca gcccggggtc	120
	ggctttccac agctcatccg tatggctcag gagggctacg ccgatcggcc ggcaactcggc	180
	cagcgtgcgc tccgtttcgt caccgacccc gacagcggcc gcaccaatgt cgagctactg	240
	ccgcggttcg agaccatcac ctaccgcgaa ctgtgggccc gcgccggcac attgccacc	300
	gcgttgagcg ctgagcccg gatccggccc ggccgaccgg ttgtcgtgct gggtttcaac	360
	agcgtcgact acacaacat cgacatcgcg ctgatccgtt tgggcgccgt gtcggttcca	420
	ctgcagacca gtgcgcgggt caccgggttg ccgccgatcg tcaccgagac cgagccgacg	480
[0031]	atgatcgcca ccagcatcga caatcttggc gacgccgtcg aagtgttggc cggtcacgcc	540
	ccggcccggc tggctcgtatt cgattaccac ggcaagggtg acaccaccg cgaggccgtc	600
	gaagccgccc gagctcgggt ggccggctcg gtgaaccatg acacacttgc cgaactgac	660
	gaacgcggca ggccgctgcc ggccacaccc attgccgaca gcgccgacga cgcgtggcg	720
	ctgtctgatt acacctcggg tagtaccggc gcacccaaag gcgccatgta tcgcgagagc	780
	cagggtgatga gcttctggcg caagtcgagt ggctgggttc agccgagcgg ttaccctcgc	840
	atcacgtcga acttcatgcc gatgagccac gtcggggccc gtcagggtct ctacgggacg	900
	ctttccaaag gcggtaccgc ctacttcgtc gccaaagacg acctgtcgac gctgttcgag	960
	gacctcgccc tgggtcgccc cacagaattg tgcttcgtgc ccgcgatctg ggacatggtg	1020
	ttcgagaggt tccacagcga ggtcgaccgc cgtttggtgg accggcccca tcgagcggcg	1080
	ctggaagcgc aggtgaaggc cgagctgcgg gagaacgtgc tcggcggacg gtttgtcatg	1140
	gcgtgacgc gttccgcgcc gatctcgtt gagatgacgg cgtgggtcga gtccctgtg	1200
	gccgaegtgc atttggtgga gggttaccgc tccaccgagg ccgggatggt cctgaacgac	1260
	ggcatggtgc ggcccccgc ggtgatcgac tacaagctgg tcgacgtgcc cgagctgggc	1320
	tacttcgcca ccgatcagcc ctacccccgg ggccgagctg tggtaagac gcaaaccatg	1380
	ttccccggct actaccagc cccggatgtc accgccgagg tgttcgacce cgacggcttc	1440
	taccggaccg gggacatcat ggccaaagta ggccccgacc agttcgtcta cctcgaccgc	1500
	cgcaacaagc tgctaaagct ctcccagggc gagttcatcg ccgtgtcgaa gctcgaggcg	1560
	gtgttcggcg acagcccgtt ggtccgacag atcttcatct accggcaacg tgccccggcc	1620

taccgcgtgg cggtggttgi cccgtccggg gacgcgttt ctcgcatgg catcgagaat	1680
ctcaagcccg tgatcagcga gtccctgcag gaggtagcga ggccggcccg cctgcaatcc	1740
tacgagatte caccgcactt catcatcgaa accacgccgt tcacctgga gaacggcctg	1800
ctcaccggca tccgcaagct ggcacgcccg cagttgaaga agttctatgg cgaacgtctc	1860
gagcggtctt ataccagact ggcgatagc caatccaac agctgcgcga gctgcggcaa	1920
agcggtcceg atgcgcgggt gcttcgcagc ctgtgccgtg ccgcggtgc gttgctgggc	1980
tctaccgtcg cggatgtgcg gccggacgcg cacttcgccg acctgggtgg tgactcgtc	2040
tcggcgctgt cgttgccaa cctgtgcac gagatcttcg gcgtcgacgt gccggtgggt	2100
gtcatgttca gcccggaag cgacctgcgg gccctggccg accacatcga agcagcgcgc	2160
accggcgta ggcgaccag cttegcctcg atacacggtc gctcgcgcac ggaagtgac	2220
gccagcgacc tcacgttga caagttcatc gacgtgccg cccctggccg agccccgaac	2280
ctgcggcaac cgagcgcga agtgcgcacc gtactgttga ccggcgccac cggtttttg	2340
ggtcgtacc tggcgctga atggctgcg cgcattggacc tggtaacgg caagctgac	2400
tgcctgttcc gcccgagac cgacgaggaa gcacaagccc ggctggacgc gacgttcgat	2460
agcggcgacc cgtatttggg gcggcactac cgcgaattgg gcgcggccg cctcgaggtg	2520
ctgcggcgcg acaaggcgga gcccgacctg gccctggacc gggtcacctg gcagcggcta	2580
gcgcacacgg ttgacctgat cgtggacccc gcggccctgg tcaaccacgt gctgccttat	2640
agccagctgt tcggcccaaa cgcggcgggc accgccgagt tgcttcggtt ggcgtgacc	2700
ggcaagcgca agccatacat ctacacctcg acgatgcgcg tggcgagca gatccgccg	2760
[0032] gaggcgttca ccgaggacgc cgacctcgg gccatcagcc cgaccgcag gatcgacgac	2820
agctacgcca accgctacgc gaacagcaag tggcgccgcg aggtgctgtt gcgcgaagct	2880
cacgagcagt gcggcctgcc ggtgacggtc ttccgctgcg acatgatcct ggccgacacc	2940
agctataccg gtcagctcaa cctgcgggac atgttcaacc ggctgatgtt gagcctggcc	3000
gtaccggca tcgcaccggg ttctttctat gagctggatg cgcacggcaa tcggcaacgc	3060
gcccaactatg accgcttgcc ggtcgaattc gtcgcagaag ccatttgac ccttgggaca	3120
catagcccg accgttttgi caccaccac gtgatgaacc cctacgacga cggcatcggg	3180
ctggacgagt tcgtcactg gtcaactcc ccaactagcg ggtccggtt caccatccag	3240
cggatcgccg actacggcga gtggctgcag cggttcgaga cttecgctgc tgccttgcg	3300
gatcgccagc gccacgcctc gctgctgccc ttgctgcaca actaccgaga gcctgcaaa	3360
ccgatatgcg ggtcaatgc gccacacgac cagttccgag ctgcggtcca agaacgaaa	3420
atcggtccgg aaaaagacat tccgcacctc accgcggcga tcacgcgaa gtacatcagc	3480
aacctgcgac tgctcgggct gctgtga	3507
<210> 43	
<211> 3522	
<212> DNA	
<213> 耻垢分枝杆菌	
<400> 43	
atgaccagcg atgttcagc gccacagac ggcgtcaccg aaaccgcact cgacgacgag	60
cagtcgaccc gccgcacgc cgagctgtac gccaccgac cagagttcgc cgcgcgcga	120
ccgttgcccc ccgtggctga cgcggcgac aaaccgggc tgcggctggc agagatctg	180

[0033]	cagaccctgt tcaccggcta cggtagccgc ccggcgctgg gataccgcgc ccgtgaactg	240
	gccaccgacg agggcggggc caccgtgacg cgtctgtctg cgcggttcga caccctcacc	300
	tacgcccagg tgtggctcgc cgtgcaagcg gtccgccggg ccttgcgcca caacttcgcg	360
	cagccgatct accccggcga cggcgctcgc acgatcgggt tcgcgagtec cgattacctg	420
	acgctggate tcgtatgcgc ctacctgggc ctctgtagtg ttcgctgca gcacaacga	480
	ccggtcagcc ggcctgcccc gatcctggcc gaggtcgaa cgcggatcct caccgtgagc	540
	gccgaatacc tcgacctcgc agtcgaatcc gtccgggacg tcaactcggg gtccgagctc	600
	gtggtgttcg accatcaccc cagggtcgac gaccaccgcg acgcactggc ccgcgcgcgt	660
	gaacaactcg ccggcaaggg catcgccgtc accaccctgg acgcgacgcg cgacgagggc	720
	gccgggctgc cggccgaacc gatctacacc gccgacctg atcagcgctt cgcgatgate	780
	ctgtacacct cgggttcacc cggcgacccc aagggtgcga tgtacaccga gccgatggtg	840
	gcgcggctgt ggaccaitgc gttcatcacg ggtgacecca cgcgggtcat caacgtcaac	900
	ttcatgcgcg tcaaccacct gggcggggcg atccccattt ccaccgccgt gcagaacggg	960
	ggaaccagtt acttcgtacc ggaatccgac atgtccacgc tgttcgagga tctcgcctg	1020
	gtgcgccga ccgaactcgg cctggttccg cgcgtcccg acatgctcta ccagcaccac	1080
	ctcgccaccg tcgaccgctt ggtcacgcag ggcgccgacg aactgaccgc cgagangcag	1140
	gccgggtccg aactgcgiga gcagggtctc ggcggacgcg tgatcaccgg attcgtcagc	1200
	accgcaccgc tggcccgga gatgaggggc ttcttcgaca tcacctggg cgcacacatc	1260
	gtcgacggct acgggctcac cgagaccggc gccgtgacac gcgacgggtg gatcgtcgg	1320
	ccaccggiga tcgactacaa gctgatcgac gttcccgaac tcggctactt cagcaccgac	1380
	aagccctacc cgcgtggcga actgtctgtc aggtcgaaa cgtgactcc cgggtactac	1440
	aagcgcgccg aggtcaccgc gagcgtcttc gaccgggacg gctactacca caccggcgac	1500
	gtcatggccg agaccgcacc cgaccacctg gtgtacgtgg accgtcgaa caacgtctc	1560
	anaactgcgc agggcgagtt cgtggcggtc gccnacctgg aggcgggtgt ctcggcgcg	1620
	gcgctggtgc gccagatctt cgtgtacggc aacagcgagc gcagtttctt tctggccgtg	1680
	gtggtccga cgcggagggc gctcgagcag tacgatccgg ccgcgctcaa ggcgcgctg	1740
	gccgactcgc tcgagcgcac cgcacgcgac gccgaactgc aatcctacga ggtgccggcc	1800
	gatttcacg tcgagaccga gccgttcagc gccgccaaag ggtgtgtgtc ggggtgtcga	1860
	aaactgctgc ggcccaacct caaagaccgc tacgggcagc gccgtgagca gatgtacgcc	1920
	gatategcgg caacgcaggc caaccagtig cgcgaactgc ggcgcgcggc cgcacacaaa	1980
	ccggtgatcg acaccctcac ccaggccgtt gccacgatec tcggcaaccg gagegaggtg	2040
	geatccgacg cccacttcac cgacctgggc ggggattccc tgtcggcgct gacatttcg	2100
	aacctgctga gcgatttctt cggtttcgaa gtccccgtcg gcaccatcgt gaaccggcc	2160
	accaacctcg cccaactcgc ccagcacatc gagggcgagc gcaccgcggg tgaccgcagg	2220
	ccgagtttca ccaccgtgca cggcgccgac gccaccgaga tccgggcgag tgagctgacc	2280
	ctggacaagt tcacgacgc cgaacgcctc cgggcgcgac cgggtctgcc caaggtcacc	2340
	accgagccac ggacgggtgt gctctcgggc gccaacggct ggtggggcgc gttcctcag	2400
	ttgcagtggc tggaaccctt ggcacctgct ggcggcacc ccatcacgat cgtcgggggc	2460

```

cgcgacgacg ccggggcccg cgcacggctg acccaggcct acgacaccga tcccagattg 2520
tcccgcgcgt tcgcccagct ggcgcaccgc cacctgcggg tggtcgcccg tgacatcggc 2580
gaccggaate tgggcctcac acccgagatc tggcaccggc tcgccgccga ggctgacctg 2640
gtggctgcatc cggcagcgct ggtcaaccac gtgctccctt accggcagct gticggcccc 2700
aacgtcgtgg gcacggccga ggtgatcaag ctggccctca ccgaacggat caagcccgic 2760
acgtacctgt ccaccgtgtc ggtggccatg gggatcccg accctcgagga ggacggcgac 2820
atccggacgc tgagcccggt gcgcccgcgc gacggcggat acgccaacgg ctacggcaac 2880
agcaagtggg ccggcgaggt gctgctgcgg gagcccacg atctgtgcgg gctgcccgtg 2940
gcgacgttcc gctcggacat gatcctggcg catccgcgtt accgcggtca ggtcaacgtg 3000
ccagacaigt tcacgcgact cctgttgagc ctcttgatca ccggcgctgc gccgcggctg 3060
ttctacatcg gagacgggtg gcgcccgcgg gcgcactacc ccggcctgac ggctgatttc 3120
gtggccgagg cggcacgac gctcggcgcg cagcagcgcg agggatacgt gicctacgac 3180
gtgatgaacc cgcacgacga cgggatctcc ctggatgtgt tcgtggaact gctgacccgg 3240
gcgggccatc cgatcgaccg ggctgacgac tacgacgact gggctcgctc gttcgagacc 3300
gcgttgaccg cgttcccgga gaagcgcgcg gcacagaccg tactgccgtt gctgcacgcg 3360
ttccgcgctc cgcaggcacc gttgcgcggc gcacccgaac ccacggaggt gttccacgcc 3420
gcggtgcgca ccgcgaaggt gggcccgagg gacatccgcg acctcgacga ggcgctgac 3480
gacaagtaca tacgcgatct gcgtgagttc ggtctgatct ga 3522

```

[0034]

<210> 44
 <211> 3582
 <212> DNA
 <213> 人工序列
 <220>
 <221> 来源
 <223> /注="人工序列的描述: 合成的多核苷酸"

```

<400> 44
atgggcagca gccatcatca tcatcatcac agcagcggcc tggtgcccgc cggcagccat 60
atgacgagcg atgttcacga cgcgaccgac ggcgttacgc agactgcaat ggaatgatgag 120
cagagcaetc gtctatttgc agaactgtac gcaacggacc cagagttcgc agcagcagct 180
cctctgcggc ccgttgtcga tgcggcgcac aaaccgggcc tgcgtctgga ggaaatcctg 240
cagacctgtt tcaccggeta cggcgatcgt ccggcgctgg gctatcgtgc acgtgagctg 300
gcgacggaag aaggcggctc tacggtcacg cgtctgtctc cgcgctfcga taccctgacc 360
tatgcacagg tgtggagccg tggtcaagca gtggctgcag cgttgcttca caatttcgca 420
caaccgattt acccgggcga cgggtcgcg actatcggtt ttgcgagccc ggactatttg 480
acgttgatc tgggtgtcgc gtatctgggc ctggtcagcg ttcttttgca gcataacgct 540
ccggtgtctc gccitggccc gattctggcc gaggtggaac cgcgtattct gacggtgagc 600
gcagaatacc tggacctggc ggttgaatcc gtccgtgatg tgaactccgt cagccagctg 660
gtgtttttcg accatcatcc ggaagtggac gatcaccgtg acgcactggc tcgcracgc 720
gagcagctgg ccggcaaagg tatcgagtt acgacctgg atgcgatcgc agacgaaggc 780
gcaggtttgc cggtgagcc gatttacacg gcggatcacg atcagcgtct ggccatgatt 840

```

[0035]

ctgtatacca gcggtctctac gggtgtctccg aaaggcgcga tgtacaccca agcgatggtg 900
 gctcgccctgt ggactatgag ctttateacg ggcgaccga ccccggttat caacgtgaac 960
 ttcatgcgcg igaacctctt ggcggttcgt atcccgatta gcaccgccgt gcagaatggc 1020
 ggtaccagct acttcgttcc ggaagcgac atgagcacgc lgtttgagga tctggccctg 1080
 gtccgcccta ccgaactggg tctgggtccg cgtgttcggg acatgctgta ccagcatcat 1140
 ctggcgaccg tggatcgctt ggtgaccag ggcgcggacg aactgactgc ggaaaagcag 1200
 gccggtgcgg aactgcgtga acaggctctg ggcggtcgtg ttatcacccg tttgtttcc 1260
 accgcgccgt tggcggcaga gatgcgtgct ttcttgata tcaccttggg tgcacacatc 1320
 gttgacggtt acggtctgac cgaaccgggt ggcgtacccc gtgatggtgt gattgttcgt 1380
 cctccggtca ttgattacaa gctgategat gtcccgagc tgggttactt ctcaccgac 1440
 aaaccgtacc cgcgtggcga gctgctggtt cgtagccaaa cgttgactcc gggttactac 1500
 aagcgccagc aagtcaccgc gtccgttttc gatcgcgacg gctattacca caaccggcag 1560
 gtgatgcgag aaaccgcgcc agaccacctg gtgtatgtgg accgcgcgaa caatgttctg 1620
 aagctgcgcg aagggtgaat tctgcgcgtg gctaacctgg aggcggtttt cagcggcgct 1680
 gctctggtcc gccagatttt cgtgtatggt aacagcgagc gcagctttct gttggctggt 1740
 gttgtcccta ccccgaggc gctggagcaa tacgacctg ccgcattgaa agcagccctg 1800
 gcggattcgc tgcagcgta cggcgctgat gccgagctgc agagctatga agtgccggcg 1860
 gaattcattg ttgagactga gcccttttagc gctgcgaacg gtctgctgag cgggtgttgc 1920
 aagtgtctgc gtccgaattt gaaggatcgc tacggtcagc gtttggagca gatgtacgcg 1980
 gacatcgccg ctaccgagc gaaccaattg cgtgaactgc gccgtgctgc ggetactcaa 2040
 ccggtgatcg acacgcigac gcaagctgcg gcgaccatcc tgggtaccgg cagcgaggtt 2100
 gcaagcgacg cacactttac tgatttgggc ggtagttctc tgagcgcgct gaagttagc 2160
 aacttgcgtc ctgacttctt tggctttgaa gtcccggttg gcacgallgt taaccagcg 2220
 actaatctgg cacagctggc gcaacatate gaggcgcagc gcacggcggg tgaccgccgt 2280
 ccatctttta cgaeggtcca cgggtcggtt gtaeggaaa tccgtgcaag cgaactgact 2340
 ctggacaaat tcatcgacgc tgagactctg cgcgcagcac ctggtttgcc gaaggttacg 2400
 actgagccgc gtacggctct gttgagcgtt gccaatggtt ggttgggccc cttectgacc 2460
 ctgcagtggc tggaaacttt ggcaccggtt ggcgggtacc tgatcaccat tgtgcgcggt 2520
 cgtgacgatg cagcggcacg tgcacgtttg actcaggctt acgatacgga cccagagctg 2580
 tcccgcgcgt tgcgtgagtt ggcggatcgc caattgcgtg tgggtggcagg tgatatcgcc 2640
 gatccgaatc tgggcctgac cccgagatt tggcacgctc tggcagcaga ggtcgatctg 2700
 gtcttctatc cagcgccctt ggtcaaccac gtctgcctt accgccagct gtttgggtccg 2760
 aatgttgttg gaaccgcga agttatcaag ttggctctga ccgagcgeat caagcctgtt 2820
 acctacctgt ccacggttag cgtcgcatg ggtattcctg attttgagga ggacggtgac 2880
 attcgtaccg tcagcccggg tcttccgctg gatggtggtt atgcaaatgg ctatggcaac 2940
 agcaagtggg ctggcgaggt gctgctgcgc gaggcacatg acctgtgtgg cctgccggtt 3000
 gcgacgtttc gtagcgacat gattctggcc caccgcgctt accgtggcca agtgaatgtg 3060
 ccggacatgt tcaccgctct gctgctgtcc ctgctgatea cgggtgttgc accgcgttcc 3120

ttctacattg gtgatggcga gctccgcgt gcacactacc cgggcctgac cgtcgatttt 3180
 gttgcggaag cggttactac cctgggtgct cagcaacgtg agggttatgt ctctatgac 3240
 gttatgaate cgcacgatga cggatattagc ttggatgtct ttgtggactg gctgattcgt 3300
 gggggccacc caattgaccg tgttgacgac taigatgact ggggtgcgtcg ttttgaaacc 3360
 gcggtgaccg ccttgccgga gaaacgtcgt gcgcagaccg ttctgccgtg gctgcatgcc 3420
 tttcgcgcgc cacaggcgcc gttgcgtggc gccctgaac cgaccgaagt gtttcattgca 3480
 gcggtgcgta ccgctaaagt cggtcgggt gatattccgc acctggatga agccctgac 3540
 gacaagtaca tccgtgacct gcgcgagttc ggcttgattt ag 3582

<210> 45
 <211> 66
 <212> DNA
 <213> 人工序列

<220>
 <221> 来源
 <223> /注="人工序列的描述: 合成的引物"

<400> 45
 cggttctggc aaatatcttg aaatgagctg ttgacaatta atcaaattccg gctcgtataa 60
 tgtgtg 66

<210> 46
 <211> 30
 <212> DNA
 <213> 人工序列

[0036]

<220>
 <221> 来源
 <223> /注="人工序列的描述: 合成的引物"

<400> 46
 ggttttattcc tccattattta atcgatacat 30

<210> 47
 <211> 59
 <212> DNA
 <213> 人工序列

<220>
 <221> 来源
 <223> /注="人工序列的描述: 合成的引物"

<400> 47
 atgtatcgat taaataagga ggaataaacc atgggcacga gcgatgttca cgacgcgac 59

<210> 48
 <211> 59
 <212> DNA
 <213> 人工序列

<220>
 <221> 来源
 <223> /注="人工序列的描述: 合成的引物"

<400> 48
 atgtatcgat taaataagga ggaataaacc gtgggcacga gcgatgttca cgacgcgac 59

<210> 49
 <211> 59
 <212> DNA
 <213> 人工序列

[0037]	<220>		
	<221> 来源		
	<223> /注="人工序列的描述：合成的引物"		
	<400> 49		
	atgtatcgat taaataagga ggaataaacc ttgggcacga gcgatgttca cgacgcgac	59	
	<210> 50		
	<211> 26		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<221> 来源		
	<223> /注="人工序列的描述：合成的引物"		
	<400> 50		
	ttctaaatca gaccgaactc gcgcag	26	
	<210> 51		
	<211> 48		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<221> 来源		
	<223> /注="人工序列的描述：合成的引物"		
	<400> 51		
	ctgcgcgagt tcggtctgat ttagaattcc tcgaggatgg tagtgttg	48	
	<210> 52		
	<211> 27		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<221> 来源		
	<223> /注="人工序列的描述：合成的引物"		
	<400> 52		
	cagtcgacat acgaaacggg aatgcgg	27	
	<210> 53		
	<211> 56		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<221> 来源		
	<223> /注="人工序列的描述：合成的引物"		
	<400> 53		
	cgcattccc gtttcgtatg tcgactgaaa cctcaggcat tgagaagcac acggtc	56	
	<210> 54		
	<211> 56		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<221> 来源		
	<223> /注="人工序列的描述：合成的引物"		
	<400> 54		
	ctcatttcag aatatttgcc agaaccgtta atttcctaaf gcaggagtgc cataag	56	
	<210> 55		
	<211> 20		
	<212> DNA		

	<213> 人工序列	
	<220>	
	<221> 来源	
	<223> /注="人工序列的描述: 合成的引物"	
	<400> 55	
	ggatctcgac gctctccctt	20
[0038]	<210> 56	
	<211> 28	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<221> 来源	
	<223> /注="人工序列的描述: 合成的引物"	
	<400> 56	
	tcaaaaacgc cattaacctg atgttctg	28

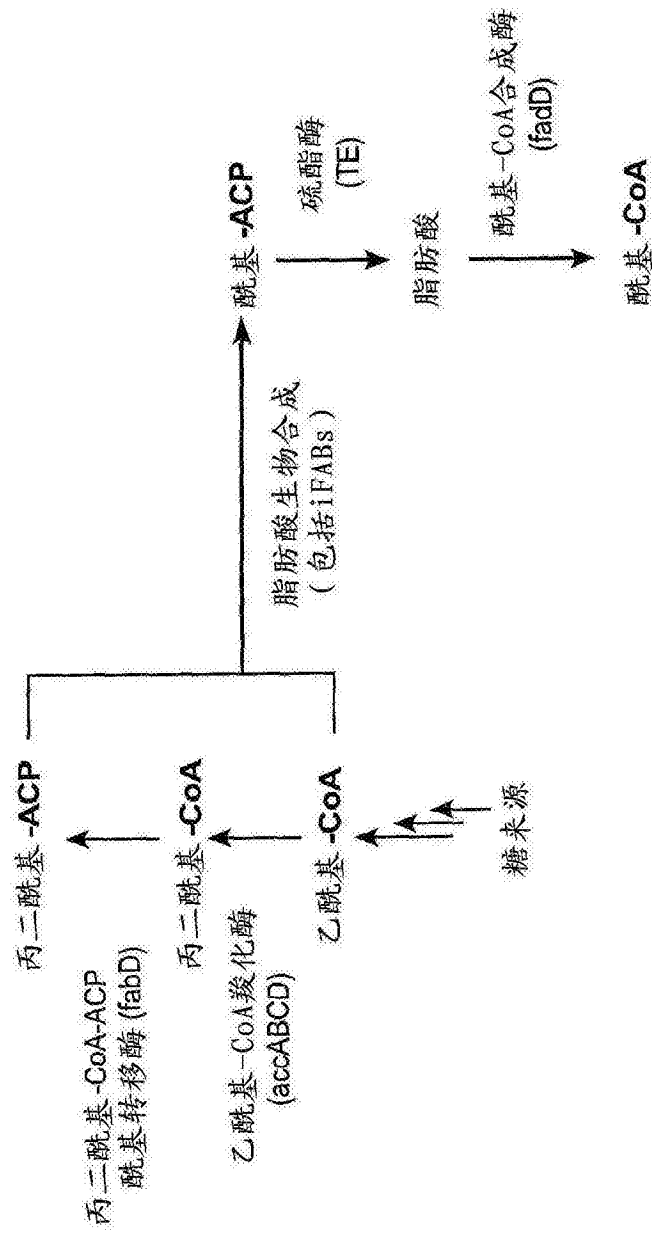


图1

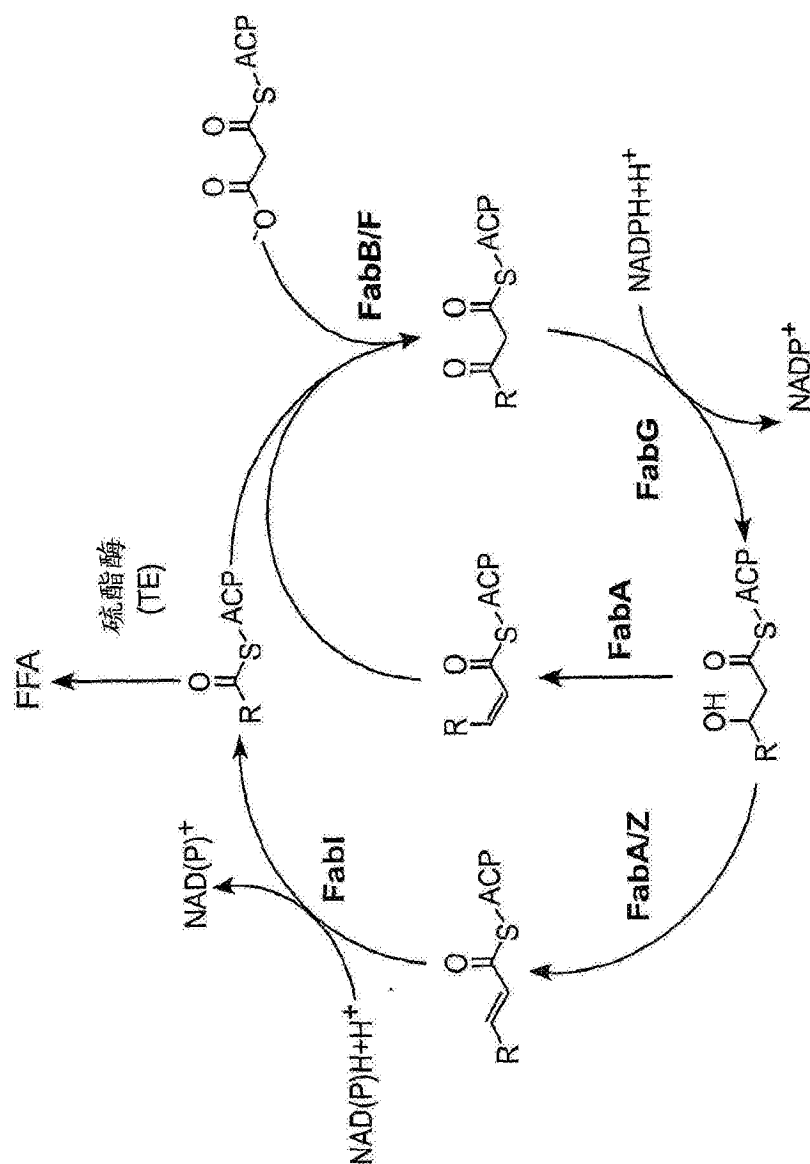


图2

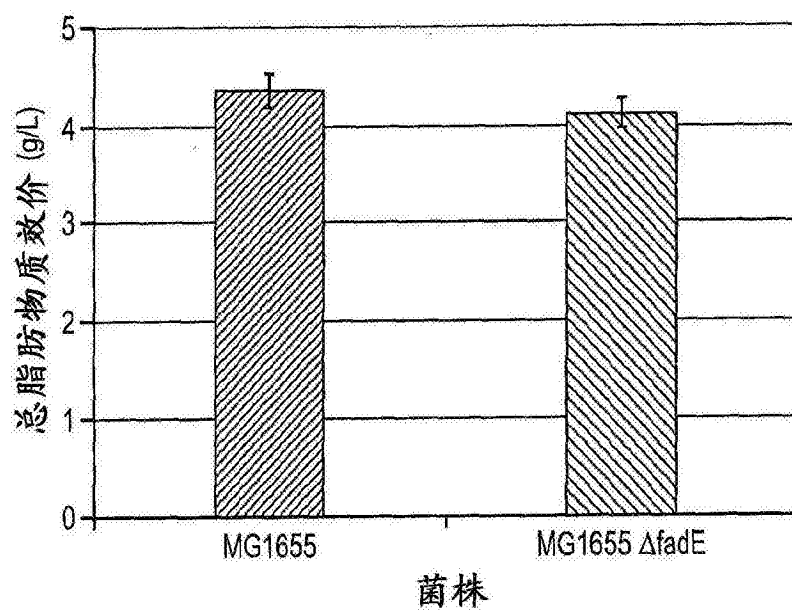


图5

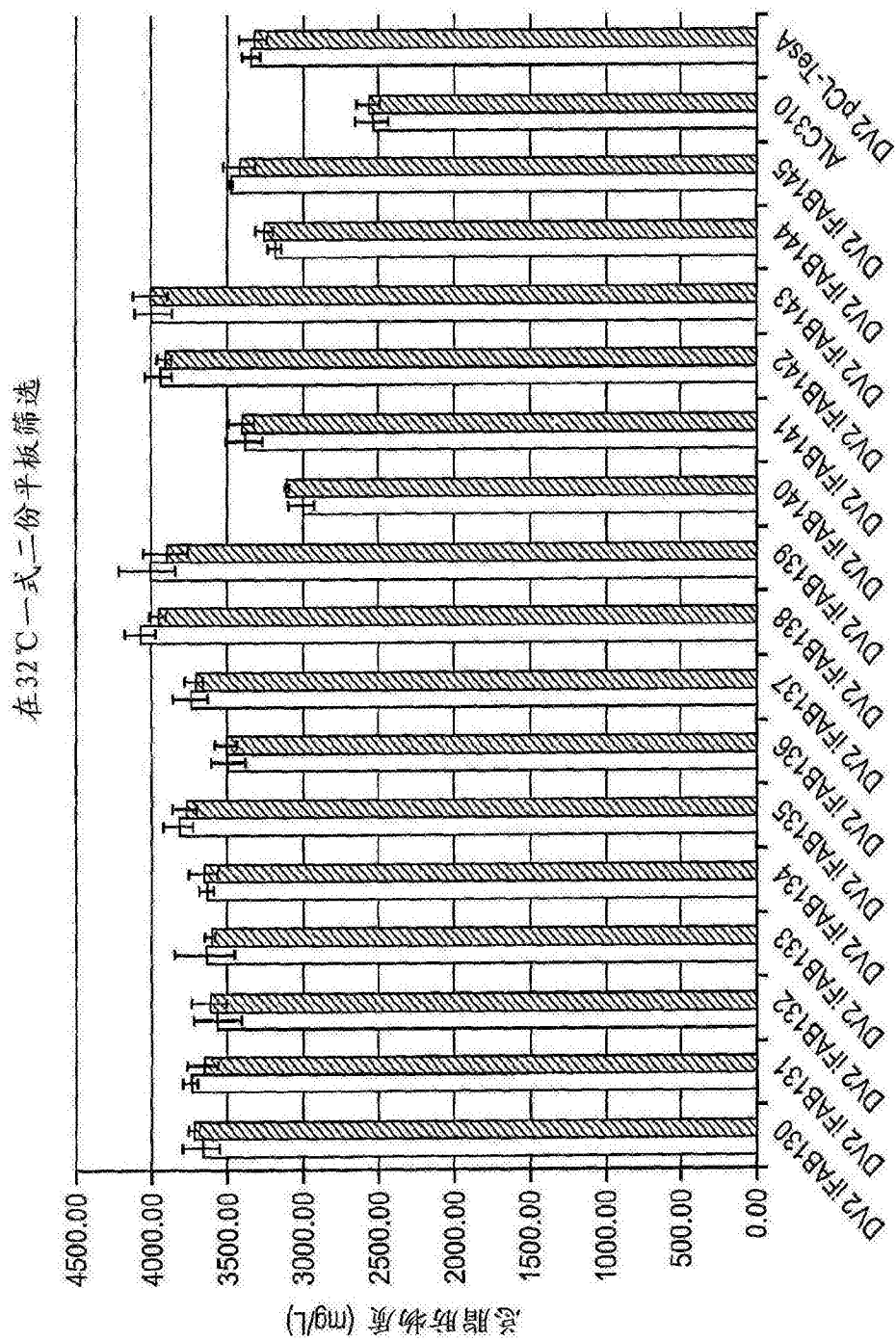


图6A

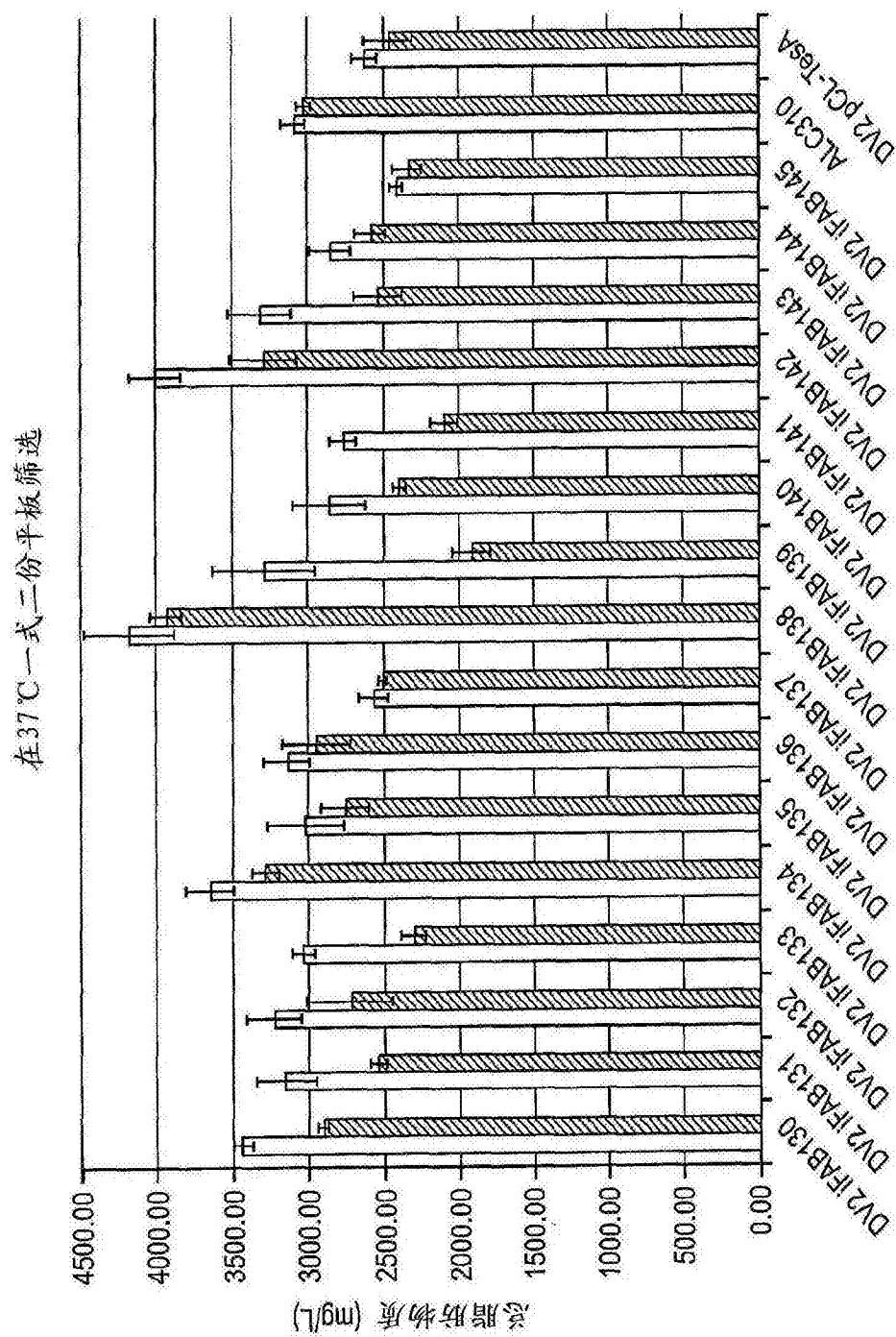


图6B

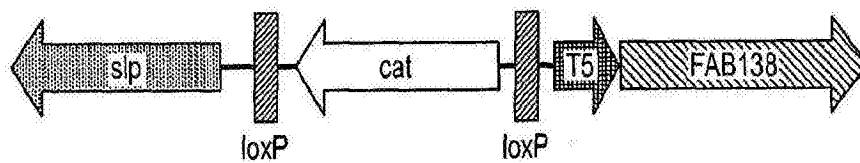


图7A

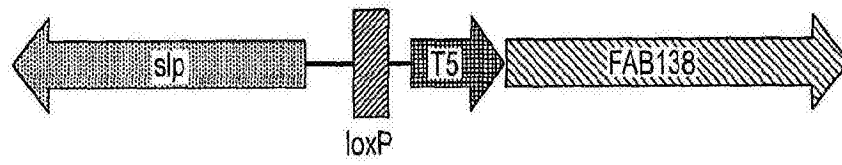


图7B

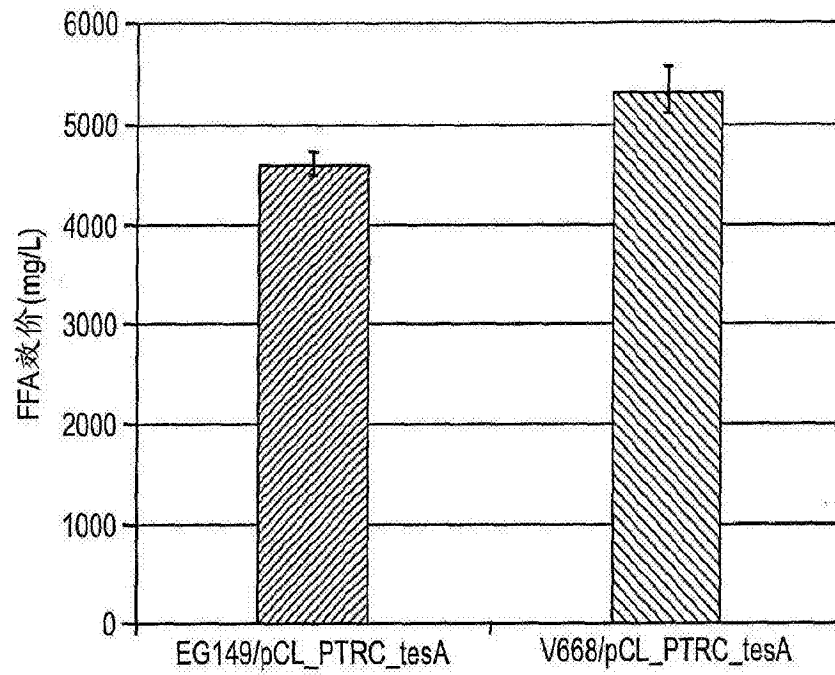


图8

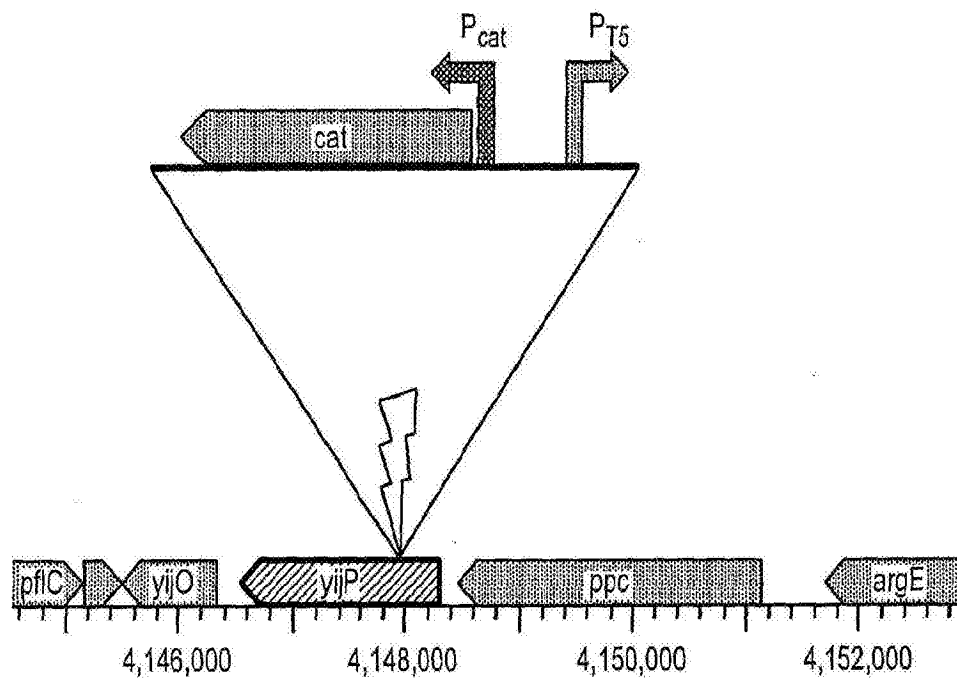


图9

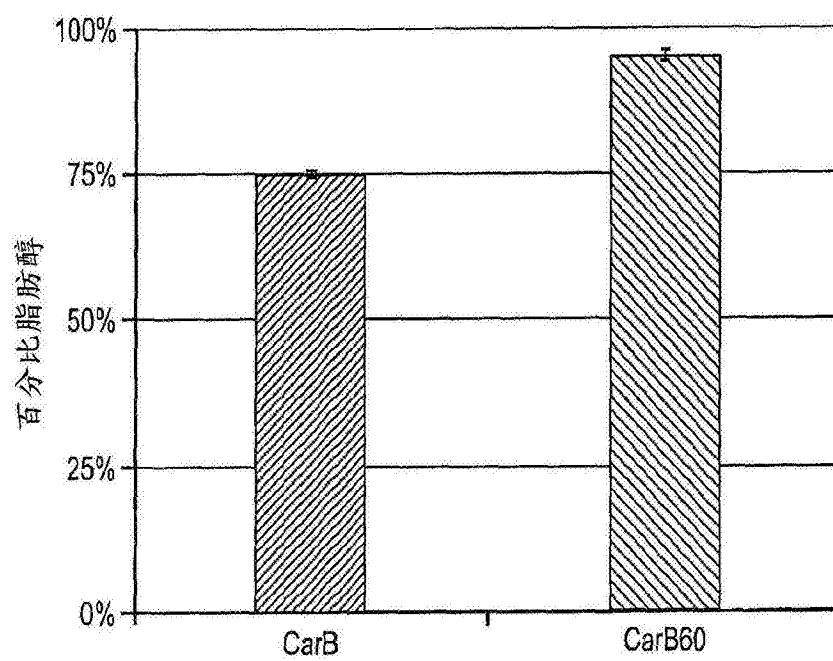


图10

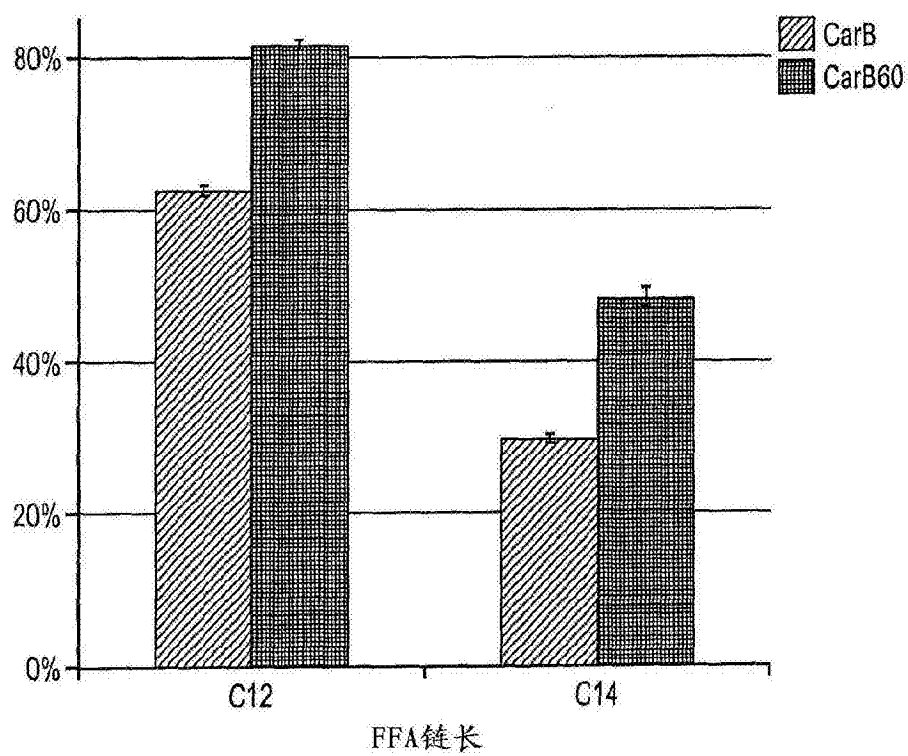


图11

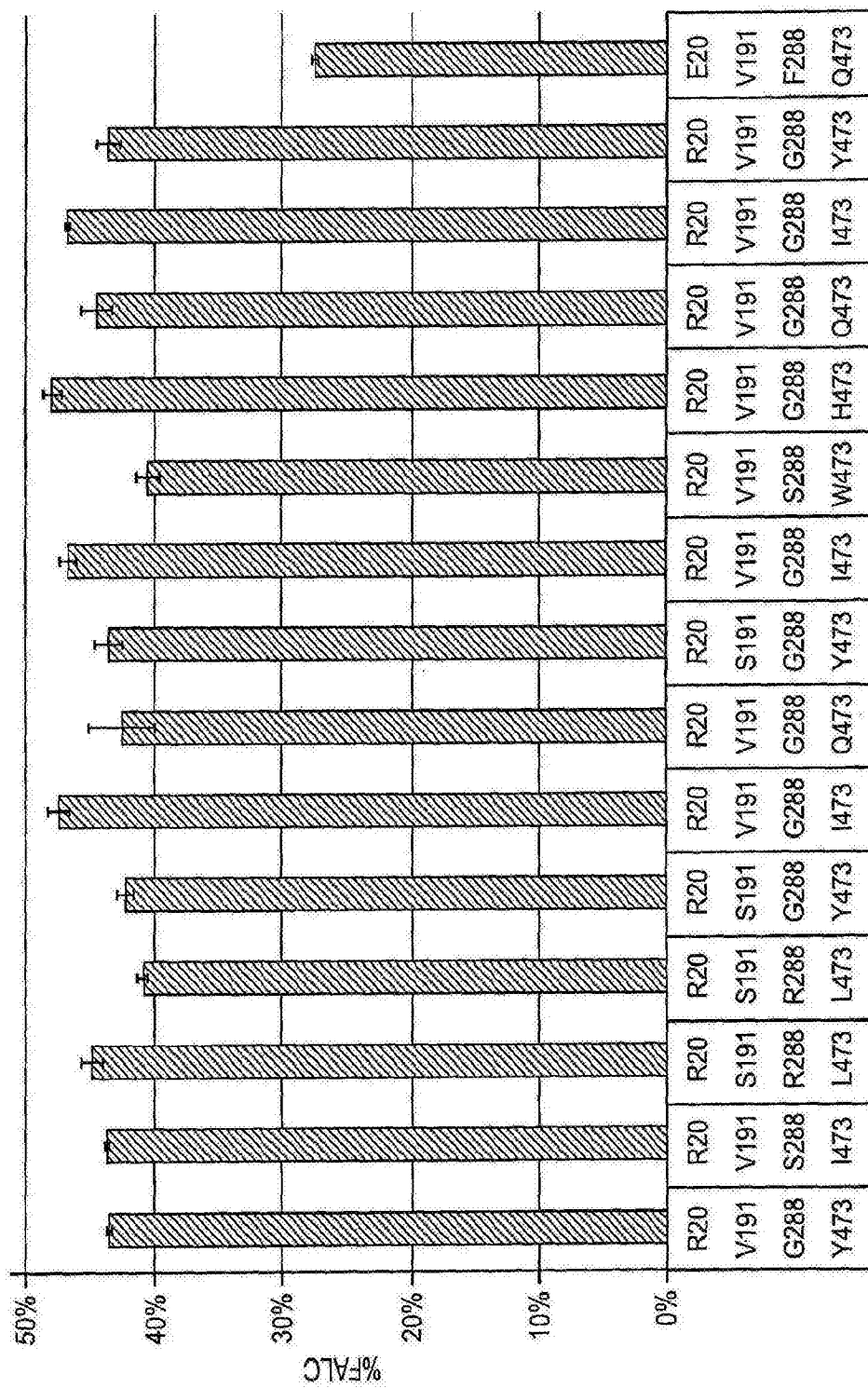


图12

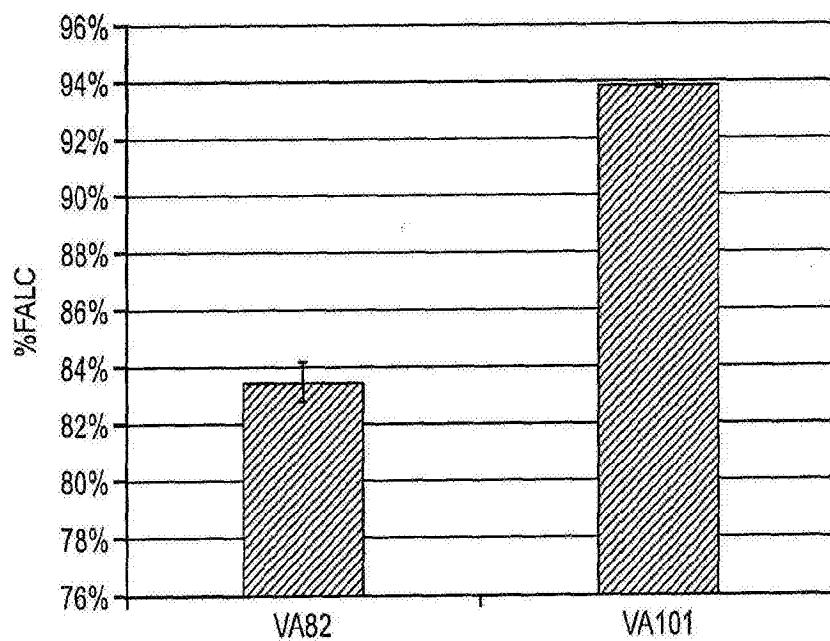


图13

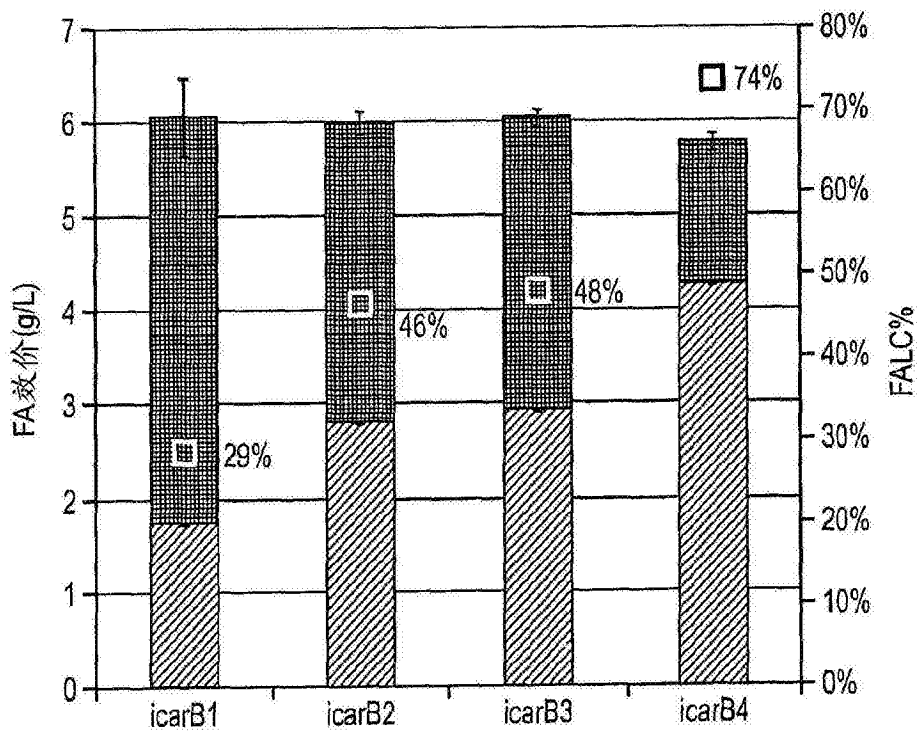


图14

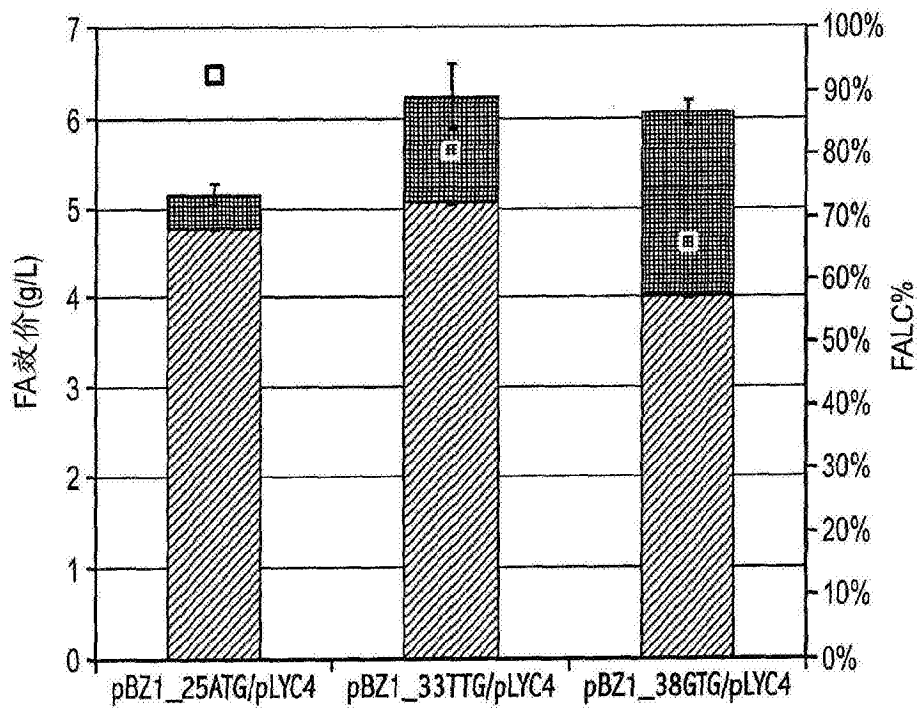


图15

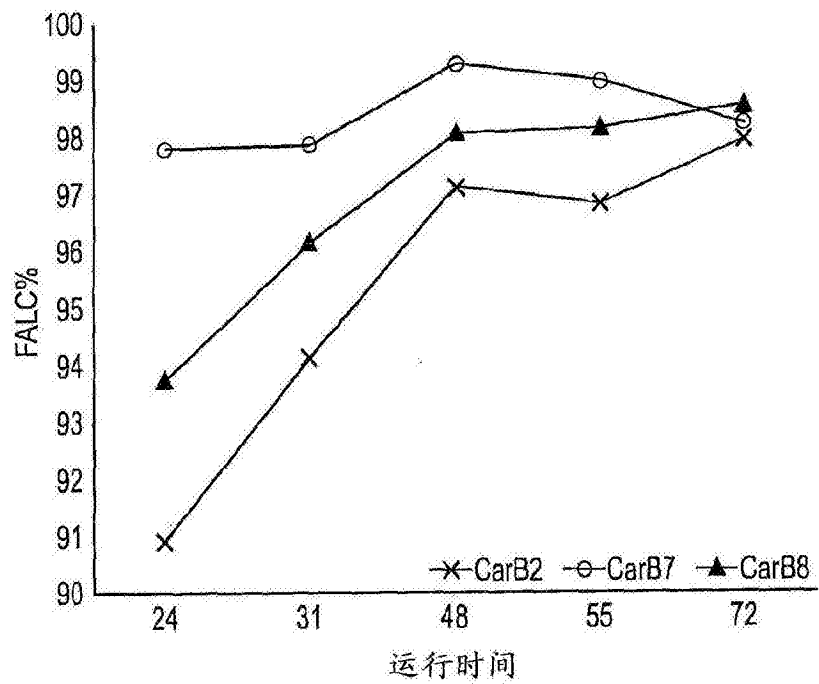


图16