



(12) 发明专利

(10) 授权公告号 CN 104395464 B

(45) 授权公告日 2022. 03. 04

(21) 申请号 201380026304.9
(22) 申请日 2013.04.02
(65) 同一申请的已公布的文献号
申请公布号 CN 104395464 A

(43) 申请公布日 2015.03.04

(30) 优先权数据
61/619,324 2012.04.02 US

(85) PCT国际申请进入国家阶段日
2014.11.20

(86) PCT国际申请的申请数据
PCT/US2013/035037 2013.04.02

(87) PCT国际申请的公布数据
W02013/152051 EN 2013.10.10

(73) 专利权人 基因组股份公司
地址 美国加利福尼亚

(72) 发明人 德里克·L·格林菲尔德
安德烈亚斯·W·席尔默

伊利莎白·J·克拉克
伊莱·S·格罗班
伯纳多·M·达科斯塔 Z·胡
凯文·霍尔登 诺厄·赫尔曼

(74) 专利代理机构 中国贸促会专利商标事务所
有限公司 11038

代理人 袁泉

(51) Int.Cl.
C12N 9/04 (2006.01)
C12N 9/02 (2006.01)
C12N 9/10 (2006.01)
C12N 9/12 (2006.01)
C12N 9/88 (2006.01)
C12N 9/90 (2006.01)
C12N 9/00 (2006.01)
C12P 7/64 (2006.01)

审查员 李有朝

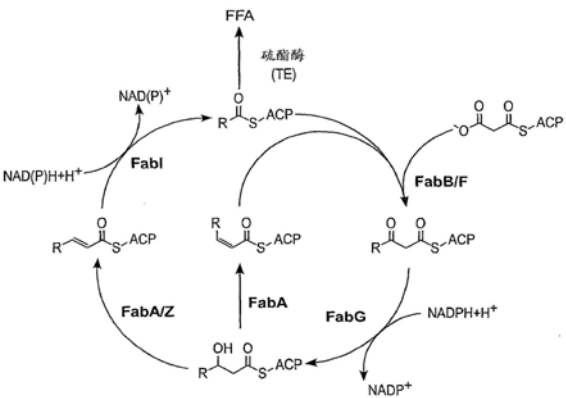
权利要求书1页 说明书52页
序列表57页 附图16页

(54) 发明名称

改善的脂肪酸衍生物的生产

(57) 摘要

本发明公开涉及重组宿主细胞,其包括有效改善了脂肪酸衍生物的效价、产率和/或生产率的菌株修饰。本发明公开进一步涉及包括重组宿主细胞的细胞培养物,其中所述的重组宿主细胞用于发酵生产脂肪酸衍生物及其组合物。



1. 一种包括基因改造的多核苷酸序列的重组细菌宿主细胞,其中在有效减少磷酸烯醇式丙酮酸羧化酶 (ppc) 蛋白质的表达低于野生型水平的条件下在包括碳源的培养基中培养时,所述的重组细菌宿主细胞以比相应的野生型宿主细胞更高的效价、产率或生产率来生产脂肪醇,其中所述的重组细菌宿主细胞是根据表3所述的大肠杆菌菌株D851。

2. 一种包括根据权利要求1所述的重组细菌宿主细胞的细胞培养物。

3. 权利要求2所述的细胞培养物,其中所述的培养基包括脂肪醇组合物。

4. 权利要求2所述的细胞培养物,其中所述的脂肪醇为

a) C6、C8、C10、C12、C13、C14、C15、C16、C17或C18脂肪醇,或

b) C10:1、C12:1、C14:1、C16:1或C18:1不饱和脂肪醇。

5. 权利要求3所述的细胞培养物,其中所述的组合物包括脂肪醇,该脂肪醇在由所述的脂肪醇的还原末端开始的C7和C8之间,在所述的碳链的位置7处具有双键。

6. 权利要求3所述的细胞培养物,其中所述的组合物包括不饱和的脂肪醇。

7. 权利要求3所述的细胞培养物,其中所述的组合物包括饱和的脂肪醇。

8. 权利要求3所述的细胞培养物,其中所述的组合物包括支链的脂肪醇。

9. 权利要求2所述的细胞培养物,其中所述的重组宿主细胞具有64g/L至70g/L的效价。

10. 权利要求2所述的细胞培养物,其中所述的重组宿主细胞具有16.9%至19.4%的产率。

改善的脂肪酸衍生物的生产

[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61/619,324的权益,并且该申请以引用方式全文并入本文。

[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,并以引用方式全文并入本文。在2013年4月2日创建的所述的ASCII拷贝命名为LS00042PCT_SL.txt,并且大小为143,098字节。

技术领域

[0005] 本发明公开涉及包括细胞株修饰的重组宿主细胞,其有效地改善了脂肪酸衍生物的效价、产率和/或生产率。本发明公开进一步涉及细胞培养物,其包括用于发酵生产脂肪酸衍生物及其组合物的重组宿主细胞。

背景技术

[0006] 包括脂肪醛、脂肪醇、碳氢化合物(烷烃和烯烃)、脂肪酯(例如蜡、脂肪酸酯或脂肪酯)和酮在内的脂肪酸衍生物代表了工业化学品和燃料的重要类别。这些分子及其衍生物具有多种用途,包括但不限于用作表面活性剂、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。原油是目前用于生产石油化学品和燃料的原材料的主要来源。由石油衍生的两类主要原材料为短链烯烃(例如乙烯和丙烯)和芳香族(例如苯和二甲苯异构体)。这些原材料通过使用多种方法(例如催化裂化、蒸汽裂化或催化重整)以相当多的花费从原油中的长链碳氢化合物断裂而衍生得到。这些原材料可以用于制造石油化学品,例如单体、溶剂、洗涤剂 and 粘合剂,而这些化学品不能以其他方式由原油直接精炼。由此,石油化学品可以用于制造专门的化学品,例如塑料、树脂、纤维、弹性体、药物、润滑剂、凝胶等。可以由石油化学品原材料生产的具体的专门化学品包括但不限于脂肪酸、碳氢化合物、脂肪醛、脂肪醇、酯和酮。

[0007] 碳氢化合物例如具有许多商业用途。依此,短链烷烃和烯烃用于运输燃料。长链烯烃用于塑料、润滑剂和合成润滑剂。此外,烯烃用作生产醇、酯、增塑剂、表面活性剂、叔胺、增强型石油采收剂、脂肪酸、硫醇、烯基琥珀酸酐、环氧化物、氯化烷烃、氯化烯烃、蜡、燃料添加剂和阻流剂的供料。类似地,酯具有多种商业用途。例如生物柴油(备选燃料)由酯(例如脂肪酸甲基酯、脂肪酸乙基酯等)制成。一些低分子量的酯是具有怡人气味的挥发物,这使它们得以用作香料或风味剂。此外,酯用作漆、涂料和清漆的溶剂。此外,一些天然形成的物质(例如蜡、脂肪和油)还由酯制成。酯进一步用作树脂和塑料中的软化剂、增塑剂、阻燃剂以及汽油和油中的添加剂。此外,酯可以用于制造聚合物、膜、纺织品、染料和药物。

[0008] 醛用于生产多种专门的化学品。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,其中一部分可以用作溶剂、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)和用作激素的化合物为醛。此外,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。相似地,脂肪醇也具有许多商业用途。更短链的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。由于脂肪醇的两性本性,脂肪醇起到非离子表面活性剂的作用,其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、药品软膏和洗涤剂、润滑油添加剂、纺织品抗静电和整理剂、增塑剂、化妆品、工业溶剂和用于脂肪的溶剂。脂肪醇(例如脂肪族的醇)包括8至22个碳原子的链。脂肪醇通常具有偶数的碳原子、以及与末端碳连接的单个醇基(-OH)。一些脂肪醇是不饱和的,一些脂肪醇是支化的。它们广泛用于工业化学中。大多数天然的脂肪醇以蜡的形式被发现,其为脂肪酸与脂肪醇形成的酯。它们由细菌、植物和动物产生。目前,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且特异性的。此外,脂肪醇脱水形成 α -烯烃还可以通过化学催化作用来完成。

[0009] 由于勘探、提取、运输和精炼用于化学和燃料产品的石油提出了固有的挑战,本领域需要备选的来源,其可以经济且高效地产生以用于化学和燃料生产。此外,特别考虑到居住在这个星球上的人口从过去以来一直在增长,所以燃烧石油基燃料已经成为环境的严重危害。因此,需要石油替代物,其不能与产生勘探、提取、运输和精炼石油所造成的相同类型的环境破坏。

[0010] 生产可再生石油的一种选择是通过改造宿主细胞来生产可再生的石油。生物衍生的燃料和化学品提供了超过石油基燃料的益处。诸如碳氢化合物(例如烷烃、烯烃或炔烃)、脂肪醇、酯、脂肪酸、脂肪醛和酮之类的生物衍生的化学品由生物质直接转化成所需的化学产品。然而,为了由商业上切实可行的可发酵的糖或生物质得到生物衍生的脂肪酸衍生物用于生产可再生化学品和燃料的来源,必须优化所述的方法以便高效地转化和回收产品。近年来,在生物衍生的燃料和化学品的研发中,有一个研究和研发的焦点。然而,所述的相关方法和产品仍非常需要改善,以使得生物衍生的燃料和化学品成为商业上切实可行的选择。需要改善的领域包括生产方法的能量效率、以及终产品的产率。本发明公开解决了该需要。

[0011] 发明概述

[0012] 本发明公开的一个方面提供了具有基因改造的多核苷酸序列的重组宿主细胞,其中所述的多核苷酸序列编码一种或多种具有特异性酶活性的多肽。所述的多核苷酸序列对于宿主细胞是外源的或内源的。依此,本发明公开提供了具有基因改造的多核苷酸序列的重组宿主细胞,其中所述的基因改造的多核苷酸序列编码一种或多种如下多肽,其中所述的多肽具有选自以下的活性,包括但不限于3-羟基癸酰基-[acp]脱水酶(E.C.4.2.1.60)活性; β -酮脂酰基-ACP合酶I(E.C.2.3.1.41)活性; β -酮脂酰基-ACP合酶II(E.C.2.3.1.179)活性;[acp]S-丙二酰基转移酶{丙二酰基-CoA-ACP酰基转移酶}(E.C.2.3.1.39)活性;3-氧酰基-{ β -酮脂酰基}-ACP还原酶(E.C.1.1.1.100)活性; β -酮脂酰基-ACP合酶III(E.C.2.3.1.180)活性;烯酰基-ACP还原酶(NADH)(E.C.1.3.1.9)活性;烯酰基-ACP还原酶(NADPH)(E.C.1.3.1.10)活性;3-羟基-酰基-[acp]脱水酶(E.C.4.2.1.59)活性;以及反-2,顺-3-癸烯酰基-ACP异构酶(E.C.5.3.3.14)活性,其中当在有效表达多核苷酸的条件在包括碳源的培养基中培养时,所述的重组宿主细胞以比相应的野生型宿主细胞更高的效

价、产率和生产率来生产脂肪酸衍生物。在相关的方面中,当所述的多肽与至少一种其他的具有酶活性的多肽组合表达时,重组宿主细胞以更高的效价、产率和/或生产率生产脂肪酸衍生物组合物。在另一个方面中,当所述的多肽与至少5种其他的具有酶活性的多肽组合表达时,重组宿主细胞以更高的效价、产率和/或生产率生产脂肪酸衍生物组合物。在另一个方面中,当所述的多肽与至少2种、3种、4种、5种、6种或更多的具有酶活性的多肽组合表达时,重组宿主细胞以更高的效价、产率和/或生产率生产脂肪酸衍生物组合物。在另一个相关的方面中,重组宿主细胞包括进一步编码为酰基载体蛋白质(ACP)的多肽的一种或多种基因改造的多核苷酸序列。ACP可以与一种或多种如下多肽组合表达,其中所述的多肽编码任一种酶活性,其中当在合适的条件下培养时,ACP进一步增加重组宿主细胞的效价、产率和/或生产率。在另一个相关的方面中,基因改造的多核苷酸序列进一步编码具有accABCD活性(E.C.6.4.1.2)的多肽。accABCD可以与一种或多种编码任一种酶活性的多肽组合表达,其中当在合适的条件下培养时,accABCD进一步增加重组宿主细胞的效价、产率和/或生产率。

[0013] 本发明公开的另一个方面提供了具有基因改造的多核苷酸的重组宿主细胞,其中所述的基因改造的多核苷酸序列编码一种或多种多肽,其中所述的多肽具有酶活性,包括但不限于反-2,顺-3-癸烯酰基-ACP异构酶活性(fabA或fabM); β -酮脂酰基-ACP合酶I(fabB);丙二酰基-CoA-ACP酰基转移酶(fabD); β -酮脂酰基-ACP合酶I(fabF或fabB); β -酮脂酰基-ACP还原酶(fabG); β -酮脂酰基-ACP合酶III(fabH);烯酰基-ACP还原酶(fabI、fabL、fabV或fabK);3-羟基-酰基-[acp]脱水酶(fabA或fabZ);和反-2-烯酰基-ACP还原酶II(fabK)。在相关的方面中,多肽选自fabA、fabB、fabD、fabF、fabG、fabH、fabI、fabL、fabV、fabZ、fabM、fabK和/或它们的组合。在另一个相关的方面中,多肽选自:鼠伤寒沙门氏杆菌(*Salmonella typhimurium*) (NP_460041)的FabA;得自大肠埃希氏菌(*Escherichia coli*) (NP_416826)的FabB;得自鼠伤寒沙门氏杆菌(NP_460164)的FabD;得自鼠伤寒沙门氏杆菌(NP_460165)的FabG;得自鼠伤寒沙门氏杆菌(NP_460163)的FabH;得自鼠伤寒沙门氏杆菌(NP_459232)的FabZ;得自变性链球菌(*Streptococcus mutans*) (AAN59379)的FabM;得自肺炎链球菌(*Streptococcus pneumoniae*) (AAF98273)的FabK;得自霍乱弧菌(*Vibrio cholera*) (YP_001217283)的FabV;得自丙酮丁醇梭菌(*Clostridium acetobutylicum*) (NP_350156)的FabF;得自枯草芽孢杆菌亚种枯草菌株168(*Bacillus subtilis* subsp. *subtilis* str.168) (NP_389054)的FabL;得自枯草芽孢杆菌亚种枯草菌株168(NP_388745)的FabI;得自不动杆菌菌种(*Acinetobacter* sp.) ADP1 (YP_047630)的FabI;得自 *Marinobacter aquaeoli* VT8 (YP_958813)的FabI;得自不透明红球菌(*Rhodococcus opacus*) B4 (YP_002784194)的FabI;得自不动杆菌菌种ADP1 (YP_046731)的FabH;得自 *Marinobacter aquaeoli* VT8 (YP_958649)的FabH;得自不透明红球菌B4 (YP_00278448)的FabH或它们的组合。

[0014] 本发明公开进一步考虑了具有基因改造的多核苷酸序列的重组宿主细胞,其中所述的基因改造的多核苷酸序列编码ACP多肽,其中当在有效表达ACP多肽的条件下在包括碳源的培养基中培养时,重组宿主细胞以比相应的野生型宿主细胞更高的效价、产率或生产率生产脂肪酸衍生物。在相关的方面中,基因改造的多核苷酸序列进一步编码具有磷酸泛酰巯基乙胺基转移酶(E.C.2.7.8.7)活性的多肽。此处,基因改造的多核苷酸序列包括编码

磷酸泛酰巯基乙胺基转移酶 (E.C.2.7.8.7) 的sfp基因。在相关的方面中,基因改造的多核苷酸序列进一步编码具有accABCD活性 (E.C.6.4.1.2) 的多肽。ACP可以与accABCD和/或磷酸泛酰巯基乙胺基转移酶组合表达,其中在合适的条件下培养时,任何表达多肽的组合都会进一步使重组宿主细胞的效价、产率和/或生产率增加。在另一个相关的方面中,ACP作为重组宿主细胞中表达的终末途径的酶由同一有机体衍生得到,其中终末途径的酶切割任何的酰基-ACP物质,其为脂肪酸生物合成途径的一部分。ACP对于宿主细胞而言是外源的或内源的。

[0015] 本发明公开进一步涵盖了包括如下基因改造的多核苷酸序列的重组宿主细胞,其中所述的基因改造的多核苷酸序列包括转座子,其中将转座子插入到yijP基因中影响位于yijP基因侧翼的第二基因,其中第二基因编码受到上调的或下调的多核苷酸,并且其中上调的或下调的多核苷酸编码如下多肽,当所述宿主细胞在有效表达所述多肽的条件下在包括碳源的培养基中培养时,所述上调的或下调的多核苷酸编码的多肽影响脂肪酸衍生物的生产。yijP基因的任一侧的侧翼上可以具有基因。在相关的方面中,转座子插入到yijP基因中使得yijP基因或其多核苷酸失活,这影响位于yijP基因侧翼的一个或多个基因,其中一个或多个侧翼基因编码如下多肽,当所述的宿主细胞在有效表达所述多肽的条件下在包括碳源的培养基中培养时,所述一个或多个侧翼基因编码的多肽影响脂肪酸衍生物组合物的生产。在一个相关的方面中,侧翼基因包括多核苷酸,其包括但不限于ppc,yij0,frwD,pf1C pf1D或argE。

[0016] 本发明公开的另一个方面提供了包括基因改造的多核苷酸序列的重组宿主细胞,其中所述的基因改造的多核苷酸序列编码磷酸烯醇式丙酮酸羧化酶 (ppc) 多肽,其中当在有效表达ppc多肽的条件下在包括碳源的培养基中培养时,所述的重组宿主细胞以比相应的野生型宿主细胞更高的效价、产率或生产率产生脂肪酸衍生物组合物。

[0017] 此外,本发明公开的另一个方面提供了细胞培养物,其包括本发明提出的任一种重组宿主细胞(上文所述)。将重组宿主细胞在培养基中培养,这样根据本发明提出的基因改造方法,重组宿主细胞生产脂肪酸衍生物组合物(上文所述)。在相关的方面中,由本发明公开的重组宿主细胞生产的脂肪酸衍生物组合物包括但不限于脂肪酸、脂肪酯、脂肪醇、脂肪醛、烷烃、端烯烃、内烯烃和酮。在另一个相关的方面中,脂肪酸衍生物为C6、C8、C10、C12、C13、C14、C15、C16、C17或C18脂肪酸衍生物。在另一个相关的方面中,脂肪酸衍生物为C10:1、C12:1、C14:1、C16:1或C18:1的不饱和脂肪酸衍生物。在另一个相关的方面中,脂肪酸衍生物组合物包括C8、C10、C12、C14、C16和C18脂肪酸衍生物中的一种或多种。由包括本发明公开的重组宿主细胞的细胞培养物生产的脂肪酸衍生物组合物包括脂肪酸、脂肪醛、脂肪醇、脂肪酯、烷烃、端烯烃、内烯烃和酮。本发明公开进一步涵盖了:包括脂肪酸衍生物的脂肪酸衍生物组合物,其中在所述的脂肪酸衍生物中,在距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;包括不饱和脂肪酸衍生物的脂肪酸衍生物组合物;包括饱和脂肪酸衍生物的脂肪酸衍生物组合物;以及包括支链脂肪酸衍生物的脂肪酸衍生物组合物。

[0018] 本发明公开进一步考虑了包括任一种本发明提出的重组宿主细胞的细胞培养物,其中当相应的野生型宿主细胞在与重组宿主细胞相同的培养条件下时,重组宿主细胞的效价比相应的野生型宿主细胞的效价高至少大约5%。在此,重组宿主细胞具有大约1g/L至大约250g/L的效价,更具体而言,大约90g/L至大约120g/L的效价。在相关的方面中,重组宿主

细胞的产率为至少大约10%至大约40%。在一个方面中,重组宿主细胞的产率为大约25%。此外,本发明还涵盖了包括任一种本发明提出的重组宿主细胞的细胞培养物,其中细胞培养物的生产率为大约0.7mg/L/hr至大约3g/L/hr或更高。

[0019] 本发明公开的另一个方面提出了制备重组宿主细胞的方法,包括对重组宿主细胞进行基因改造,使得该细胞在特定的培养条件下表达由一种或多种多核苷酸序列编码的多肽序列,其中所述的多核苷酸序列编码具有特异酶活性的一种或多种多肽。

[0020] 附图简述

[0021] 当组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。

[0022] 图1提出了在重组微生物有机体中以酰基CoA作为前体来生产脂肪酸衍生物的示例性生物合成途径。该循环由丙二酰基-ACP和乙酰基-CoA的缩合来引发。

[0023] 图2提出了示例性脂肪酸的生物合成循环,其中通过丙二酰基-CoA的转酰基作用形成丙二酰基-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合酶I(fabB)和 β -酮脂酰基-ACP合酶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 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。

[0024] 图3示出了乙酰基-CoA羧化酶(accABCD)的酶复合物的结构和功能。BirA使accB(其为生物素羧基载体蛋白质)生物素化,其为乙酰基-CoA羧化酶的酶复合物的一部分。

[0025] 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(TE)、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。

[0026] 图5提出了以酰基-ACP为起始来生产脂肪酯的示例性生物合成途径的概况,其中脂肪酯的生产是通过单酶系统或三酶系统来完成的。

[0027] 图6提出了以酰基-ACP为起始来生产碳氢化合物的示例性生物合成途径的概况;内烯烃的生产是由OleABCD的酶活性来催化的;烷烃的生产是通过醛脱羧基酶(ADC)将脂肪醛酶转化形成烷烃来催化的;端烯烃的生产是通过羧化酶将脂肪酸酶转化形成端烯烃来催化的。

[0028] 图7示出了由具有fadE基因衰减(即,删除)的MG1655大肠杆菌菌株生产的脂肪酸衍生物(总的脂肪物质)与大肠杆菌MG1655生产的脂肪酸衍生物的比较。图7提出的数据显示fadE基因衰减不会影响脂肪酸衍生物的生产。

[0029] 图8示出了在对数期的DAM1_i377中的丙二酰基-CoA水平,其中所述的DAM1_i377

表达8种不同的谷氨酸棒状杆菌 (*C. glutamicum*) 乙酰基-CoA羧化酶 (Acc) 操纵子构建体。

[0030] 图9示出了在对数期的大肠杆菌DAM1_i377中的细胞内短链-CoA水平,其中所述的大肠杆菌DAM1_i377表达ptrc1/3_accDACB-birA±panK操纵子构建体。accDACB+birA在本发明中还称为accD+。

[0031] 图10示出了大肠杆菌菌株DV2中的脂肪酸甲基酯 (FAME) 的生产,其中所述的大肠杆菌菌株DV2表达由烃海杆菌 (*M. hydrocarbonoclasticus*) 得到的酯合酶9和由谷氨酸棒状杆菌得到的乙酰基-CoA羧化酶复合物的成分。

[0032] 图11示出了大肠杆菌的脂肪醇生产,其中所述的大肠杆菌在pCL质粒上表达细长聚球藻 (*Synechococcus elongatus*) PCC7942AAR以及由谷氨酸棒状杆菌得到的accD+操纵子。其中显示了accD+菌株的3个重复样品。

[0033] 图12A和12B示出当质粒pCL_P_{trc}-tesA转化至如图所示的各包括iFAB的菌株中时,由2个重复的平板筛选得到的总脂肪物质的生产数据 (mg/L),并在FA2培养基中发酵20小时,从而在32℃ (图12A) 和37℃ (图12B) 完成诱导至收获。

[0034] 图13示出具有质粒pDS57和整合的fabHI操纵子的大肠杆菌DAM1的FAME生产。fabH/I基因得自Marinobacter aquaeoli VT8或得自贝氏不动杆菌 (*Acinetobacter baylyi*) ADP1。关于这些菌株的fabH/I操纵子的更多细节参见表7。

[0035] 图14示出具有质粒pDS57、以及不同构造的谷氨酸棒状杆菌acc基因及整合的fabHI操纵子的大肠杆菌DAM1的FAME生产。该菌株包括得自不透明红球菌或贝氏不动杆菌ADP1的fabH/I基因。关于fabH/I和acc操纵子的更多细节参见表7。

[0036] 图15示出2种大肠杆菌DAM1pDS57菌株与对照菌株大肠杆菌DAM1pDS57相比的FAME和FFA效价,其中所述的大肠杆菌DAM1 pDS57菌株选自图13中的整合的fabH/I基因菌株。

[0037] 图16为iFAB138基因座的图示,其包括整合在iFAB138前面的cat-loxP-P_{T5}盒的示意图 (图16A);以及P_{T5}-iFAB138区 (图16B) 的示意图。

[0038] 图17示出了菌株V668,其具有修复的rph和ilvG基因,生产比EG149更高水平的FFA,其中所述的EG149不具有任一修复的rph和ilvG基因。

[0039] 图18为插入在菌株LC535的yijP基因中的转座子盒 (转座子匹配物68F11) 的图示。其中示出位于转座子盒内的启动子,并且该启动子可以影响相邻基因的表达。

[0040] 图19示出表达细长聚球藻酰基-ACP还原酶 (AAR) 并共表达多种蓝细菌酰基载体蛋白质 (ACP) 的大肠杆菌DV2的脂肪醇的生产。关于ACP来源的细节在表12中提供。

[0041] 图20示出表达无前导序列的大肠杆菌硫酯酶'tesA并共表达蓝细菌酰基载体蛋白质 (cACP) 和枯草芽孢杆菌sfp的大肠杆菌DV2的脂肪酸的生产。

[0042] 发明详述

[0043] 综述

[0044] 本发明公开至少部分基于这样的发现:在重组宿主细胞中,脂肪酸生物合成途径的多个方面的修饰有利于增强该宿主细胞生产脂肪酸衍生物。本发明公开涉及具有所需特征的脂肪酸衍生物的组合物及生产该脂肪酸衍生物的方法。此外,本发明公开涉及重组宿主细胞 (例如微生物有机体)、重组宿主细胞的培养物、制备和使用重组宿主细胞的方法,例如在培养的重组宿主细胞在发酵生产具有所需特性的脂肪酸衍生物中的用途。

[0045] 更具体而言,所需的脂肪酸衍生物组合物 (例如酰基-CoA、脂肪酸、脂肪醛、短链和

长链的醇、碳氢化合物、脂肪醇、酯(例如蜡、脂肪酸酯或脂肪酯)、端烯烃、内烯烃和酮)的生产通过修饰一种或多种基因(与脂肪酸、脂肪酯、烷烃、烯烃、炔烃(olefin)或脂肪醇的生物合成途径、生产、降解和/或分泌有关)的表达而增强。本发明公开提供了重组宿主细胞,其经改造,从而提供相对于未改造的或天然的宿主细胞(例如作为对照细胞的野生型宿主细胞)而言增强的脂肪酸生物合成,这种增强是通过例如菌株的改善来完成的。依此,本发明公开鉴定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。通常公认的是,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,并针对活性筛选编码的多肽。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基因改造的或修饰的多核苷酸、以及编码的变体多肽,其中所述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)。

[0046] 本发明公开鉴定根据酶分类(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(例如酶)、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(其通过本发明的编号和/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。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通过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rhw 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)。

[0047] 定义

[0048] 除非明确地作出另外的指示,如本说明书和所附的权利要求书所用,单数形式“一”、“一个”和“一种”包括复数指示物。因此,例如“重组宿主细胞”包括两种或更多种此类重组宿主细胞,“脂肪醇”包括一种或多种脂肪醇或脂肪醇的混合物,“编码序列的核酸”包括一种或多种编码序列的核酸;“酶”包括一种或多种酶等。

[0049] 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”),以及得自由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”)。

[0050] 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号是根据酶催化的反应将酶分类。

[0051] 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）。

[0052] 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。

[0053] 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，该载体依此转化宿主细胞，从而生产多肽。

[0054] 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。

[0055] 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，将序列进行比对（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75%、至少大约80%、至少大约85%、至少大约90%、至少大约95%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。

[0056] 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))。

[0057] 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,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。

[0058] “内源的”多肽是指由宿主细胞(例如亲代微生物细胞)的基因组编码的多肽,其中重组细胞由所述的宿主细胞改造或衍生得到。

[0059] “外源的”多肽是指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。

[0060] 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异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。

[0061] 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。

[0062] 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。

[0063] 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)或编码会影响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)。

[0064] 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例如Goeddel, Gene Expression

Technology:Methods in Enzymology,Vol.185,Academic Press,San Diego,Calif.(1990)中所述。

[0065] 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。

[0066] 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,重组载体进一步包括与多核苷酸序列可操作地连接的启动子。在一些实施方案中,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。重组载体通常包括至少一个序列,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。在某些实施方案中,核苷酸序列被稳定地引入到宿主细胞的基因组DNA中,并且核苷酸序列的表达处于受调节的启动子区的控制之下。本发明所述的表达载体包括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。

[0067] 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。

[0068] 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.,细胞,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)。

[0069] 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. 细胞 Biol., 3:2156-2165 (1983)) 和pVL系列(Lucklow et al., Virology, 170:31-39 (1989))。

[0070] 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)。

[0071] 本发明所指的术语“相应的野生型宿主细胞”是指作为对照细胞的细胞。例如如果重组宿主细胞中的多肽受到上调,则相同的多肽在对照细胞中以更低的水平存在。相反,如果重组宿主细胞中的多肽受到下调,则相同的多肽在对照细胞中以更高的水平存在。此外,“重组的或改造的宿主细胞”为用于生产一种或多种脂肪酸衍生物的微生物有机体,其中所述的脂肪酸衍生物包括例如酰基-CoA、脂肪酸、脂肪醛、短链和长链的醇、碳氢化合物、脂肪醇、酯(例如蜡、脂肪酸酯或脂肪酯)、端烯烃、内烯烃和酮。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,各种多核苷酸均编码具有脂肪酸生物合成酶活性的多肽。

[0072] 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。

[0073] 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)的磷酸泛酰巯基乙胺基部分的硫氢基之间形成酰基硫酯。磷酸泛酰巯基乙胺基部分在翻译后通过全-酰基载体蛋白质合酶(ACPS) (一种磷酸泛酰巯基乙胺基转移酶) 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。

[0074] 如本文所用,术语“脂肪酸衍生物”是指“脂肪酸”或“脂肪酸衍生物”,其可以称为“脂肪酸或其衍生物”。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各碳原子。脂肪酸可以是饱和的、单不饱和的或多不饱和的。“脂肪酸衍生物”为部分由生产宿主有机体的脂肪酸生物合成途径形成的产物。“脂肪酸衍生物”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、脂肪酸、脂肪醛、短链和长链的醇、脂肪醇、碳氢化合物、酯(例如蜡、脂肪酸酯或脂肪酯)、端烯烃、内烯烃和酮。

[0075] 本发明所指的“脂肪酸衍生物组合物”是由重组宿主细胞生产的,并且通常包括脂肪酸衍生物的混合物。在一些情况下,所述的混合物包括多于一种类型的产物(例如脂肪酸和脂肪醇、脂肪酸和脂肪酸酯、或者烷烃或烯烃)。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、且饱和或支化特征的脂肪醇(或者另一种脂肪酸衍生物)的混合物。在其他的情况下,脂肪酸衍生物组合物包括多于一种类型的产物、以及具有不同链长且饱和或支化特征的产物的混合物。

[0076] 如本文所用,术语“脂肪酸生物合成途径”是指生产脂肪酸及其衍生物的生物合成

途径。脂肪酸生物合成途径除了包括本发明所讨论的生产具有所需特征的脂肪酸衍生物的酶或多肽以外,还可以包括其他的具有酶活性的酶或多肽。

[0077] 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C6、C7、C8、C9、C10、C11、C12、C13、C14、C15、C16、C17、C18、C19、C20、C21、C22、C23、C24、C25或C26脂肪醛。在某些实施方案中,脂肪醛为C6、C7、C8、C9、C10、C11、C12、C13、C14、C15、C16、C17或C18脂肪醛。

[0078] 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、C7、C8、C9、C10、C11、C12、C13、C14、C15、C16、C17或C18脂肪醇。

[0079] 脂肪酸衍生物(例如脂肪醇)的R基可以为直连或支链。支链可以具有多于一个的支化点,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位置(C1)处。

[0080] 在某些实施方案中,支化的脂肪酸衍生物为异脂肪酸衍生物,例如异脂肪醛、异脂肪醇或反异脂肪酸衍生物、反异脂肪醛或反异脂肪醇。在示例性的实施方案中,支化的脂肪酸衍生物选自异C7:0、异C8:0、异C9:0、异C10:0、异C11:0、异C12:0、异C13:0、异C14:0、异C15:0、异C16:0、异C17:0、异C18:0、异C19:0、反异C7:0、反异C8:0、反异C9:0、反异C10:0、反异C11:0、反异C12:0、反异C13:0、反异C14:0、反异C15:0、反异C16:0、反异C17:0、反异C18:0和反异C19:0的支化的脂肪醇。

[0081] 支化的或非支化的脂肪酸衍生物的R基可以是饱和的或不饱和的。如果是不饱和的,则R基可以具有多于一个的不饱和点。在一些实施方案中,不饱和的脂肪酸衍生物是单不饱和的脂肪酸衍生物。在某些实施方案中,不饱和的脂肪酸衍生物为C6:1、C7:1、C8:1、C9:1、C10:1、C11:1、C12:1、C13:1、C14:1、C15:1、C16:1、C17:1、C18:1、C19:1、C20:1、C21:1、C22:1、C23:1、C24:1、C25:1或C26:1的不饱和的脂肪酸衍生物。在某些实施方案中,不饱和

的脂肪酸衍生物为C10:1、C12:1、C14:1、C16:1或C18:1的不饱和脂肪酸衍生物。在其他的实施方案中,不饱和的脂肪酸衍生物在 ω -7位置处是不饱和的。在某些实施方案中,不饱和的脂肪酸衍生物包括顺式双键。

[0082] 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。

[0083] 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”是指使重组细胞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DifcoTM和BBLTM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;WO 2010127318中所述的方法,培养物的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。此外,在一些实施方案中,宿主细胞被改造从而表达转化酶,这样蔗糖可以被用作碳源。

[0084] 如本文所用,术语“在有效表达基因改造的多核苷酸序列的条件下”是指允许宿主细胞生产所需的脂肪酸衍生物的任何条件。合适的条件包括例如发酵条件。

[0085] 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相应的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改变(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是为了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。

[0086] 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)。

[0087] 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。

[0088] 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酸衍生物

的量。在本发明所述的组合物和方法的任何方面中,脂肪酸衍生物以大约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,脂肪酸衍生物以高于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酸衍生物的效价为5g/L至200g/L、10g/L至150g/L、20g/L至120g/L、以及30g/L至100g/L。在一个实施方案中,根据本发明公开的方法由重组宿主细胞生产的脂肪酸衍生物的效价为大约1g/L至大约250g/L、并且更具体为90g/L至大约120g/L。所述的效价可以指由给定的重组宿主细胞培养物生产的特定的脂肪酸衍生物或脂肪酸衍生物的组合。

[0089] 如本文所用,“由宿主细胞生产的脂肪酸衍生物的产率”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酸衍生物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,一种或多种脂肪酸衍生物以高于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一种或多种脂肪酸衍生物的产率可以为5%至15%、10%至25%、10%至22%、15%至27%、18%至22%、20%至28%、或20%至30%。在具体的实施方案中,由重组宿主细胞生产的一种或多种脂肪酸衍生物的产率为大约10%至大约40%。在另一个具体的实施方案中,由重组宿主细胞生产的一种或多种脂肪酸衍生物的产率为大约25%。所述的产率可以指由给定的重组宿主细胞培养物生产的特定的脂肪酸衍生物或脂肪酸衍生物的组合。

[0090] 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一种或多种脂肪酸衍生物的量。在本发明所述的组合物和方法的任何方面中,由重组宿主

细胞生产的一种或多种脂肪酸衍生物的生产率为至少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。例如根据本发明所述的方法由重组宿主细胞生产的一种或多种脂肪酸衍生物的生产率可以为500mg/L/小时至2500mg/L/小时、或700mg/L/小时至2000mg/L/小时。在一个具体的实施方案中,产率为大约0.7mg/L/h至大约3g/L/h。所述的生产率可以指由给定的重组宿主细胞培养物生产的特定的脂肪酸衍生物或脂肪酸衍生物的组合。

[0091] 如本文所用,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,其是指脂肪醇、脂肪醛和脂肪酸的量,其可以通过GC-FID来评价,如国际专利申请公开W02008/119082中所述。相同的术语用于指脂肪酯分析时,可以用于指脂肪酯和游离脂肪酸。

[0092] 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)。如本文所用,术语“碳源”是指适于用作原核或简单的真核细胞生长所需的碳源的底物或化合物。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、乳酸盐和醋酸盐;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。

[0093] 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)。

[0094] 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。

[0095] 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脂肪酸衍生物的百分率增加。例如当在重组宿主细胞中产生脂肪酸衍生物时,可以通过去除宿主细胞蛋白质来纯化脂肪酸衍生物。纯化后,样品中脂肪酸衍生物的百分率增加。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酸衍生物时,纯化的脂肪酸衍生物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)隔离的脂肪酸衍生物。

[0096] 菌株改良

[0097] 为了形成脂肪酸衍生物的高效价、产率和/或生产率,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))。

[0098] 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。一种建议是对用于脂肪酸生物合成的主要前体的限定,例如乙酰基-CoA和丙二酰基-CoA可以使得脂肪酸衍生物的合成减少。增加通过脂肪酸生物合成的流量的一种方法是操纵该途径中的多种酶(参见图1和2)。实施例3描述了多个研究,该研究显示fab操纵子的构建以及向大肠杆菌宿主细胞的染色体中的整合,其中所述的fab操纵子编码生物合成途径中的酶,这些酶用于将丙二酰基-CoA转化为酰基-ACP。不想被理论所束缚,这可以增加脂肪酸生物合成的流量。通过乙酰基-CoA羧化酶(acc)复合物和脂肪酸生物合成(fab)途径由乙酰基-CoA供应酰基-ACP是限定脂肪酸衍生物的生产速率的另一个步骤(参见图3)。实施例2显示出优化形式的大肠杆菌谷氨酸棒状杆菌accABCD($\pm$ birA)过表达的效果,其证明此类基因修饰可以使大肠杆菌中的乙酰基-coA和丙二酰基-CoA的生产增加。

[0099] 在另一个方法中,显示大肠杆菌宿主细胞中的rph和ilvG基因突变使得游离的脂肪酸(FFA)的产更高,这转化为脂肪醇的生产更高,如实施例4所示。在另一个方法中,获得转座子诱变和高通量的筛选,从而发现可以增加效价或产率的有利突变。如实施例5所示,将转座子插入到yijP基因中可以改善脂肪醇中摇瓶和分批供料的发酵中的产率。

[0100] 由重组宿主细胞生成脂肪酸衍生物

[0101] 本发明公开提供了多肽(例如酶)的多个实例,其中所述的多肽(例如酶)具有适用于本发明所述的脂肪酸生物合成途径的活性。此类多肽在本发明中统称为“脂肪酸生物合成多肽”或“脂肪酸生物合成酶”。本发明提供了脂肪酸途径的多肽的非限定性实例,其中所述的多肽适用于本发明公开的重组宿主细胞。在一些实施方案中,本发明公开包括含有如

下多核苷酸序列的重组宿主细胞,其中所述的多核苷酸序列编码脂肪酸生物合成的多肽。所述的多核苷酸序列可以被整合到重组宿主细胞的染色体中,引入到存在于重组宿主细胞中的一个或多个质粒表达系统中,或两者,其中所述的多核苷酸序列包括编码脂肪酸生物合成的多肽的开放阅读框以及可操作地连接的调节序列。在一个实施方案中,脂肪酸生物合成的多核苷酸序列编码多肽,该多肽对于被改造的重组宿主细胞的亲代宿主细胞(即,对照细胞)而言是外源的。一些此类的外源多肽在重组宿主细胞中是过表达的。在另一个实施方案中,脂肪酸生物合成的多核苷酸序列编码外源或异源的多肽。换言之,由多核苷酸编码的多肽对于亲代宿主细胞而言是外源的。在另一个实施方案中,遗传修饰的宿主细胞过表达编码如下多肽(蛋白质)的基因,其中所述的多肽(蛋白质)增加宿主细胞生产脂肪酸生物合成酶的底物(即,脂肪酰基硫酯底物)的速率。在某些实施方案中,由所表达的基因编码的酶与脂肪酸的生物合成直接有关。此类重组宿主细胞可以被进一步改造,从而包括编码一种或多种脂肪酸生物合成的多肽(即,与脂肪酸生物合成有关的酶)的多核苷酸序列。此类多肽的实例为具有硫酯酶活性的多肽或蛋白质,其中所述的重组宿主细胞合成脂肪酸;或者具有硫酯酶活性和羧酸还原酶(CAR)活性的多肽或蛋白质,其中重组宿主细胞合成脂肪醛和脂肪醇;或者具有硫酯酶活性、羧酸还原酶活性和醇脱氢酶活性的多肽或蛋白质,其中重组宿主细胞合成脂肪醇;或者具有酰基-CoA还原酶(AAR)活性的多肽或蛋白质,其中重组宿主细胞合成脂肪醛和脂肪醇;或者具有酰基-CoA还原酶(AAR)活性和醇脱氢酶活性的多肽或蛋白质,其中重组宿主细胞合成脂肪醇;或者具有形成脂肪醇的酰基CoA还原酶(FAR)活性的多肽或蛋白质,其中重组宿主细胞合成脂肪醇;或者具有硫酯酶活性、羧酸还原酶活性和醛脱羧基酶活性的多肽或蛋白质,其中重组宿主细胞合成烷烃;或者具有酰基-CoA还原酶活性和醛脱羧基酶活性的多肽和蛋白质,其中重组宿主细胞合成烷烃;或者具有酯合酶活性的多肽和蛋白质,其中重组宿主细胞合成脂肪酯(例如单酶系统;参见图5);或者具有硫酯酶活性、酰基-CoA合酶活性和酯合酶活性的多肽和蛋白质,其中重组宿主细胞合成脂肪酯(例如三酶系统;参见图5);或者具有OleA活性的多肽和蛋白质,其中重组宿主细胞合成脂肪族酮;或者具有OleABCD活性的多肽和蛋白质,其中重组宿主细胞合成内烯烃;或者具有硫酯酶活性和脱羧酶活性的多肽和蛋白质,其中重组宿主细胞合成端烯烃;或者它们的组合。在一些实施方案中,由脂肪酸生物合成的多核苷酸编码的一种多肽是外源(或异源)多肽(例如由除了亲代宿主细胞以外的有机体起源的多肽,或者原产于亲代微生物细胞的多肽的变体),或者内源多肽(即,原产于亲代宿主细胞的多肽),其中内源多肽在重组宿主细胞中是过表达的。

[0102] 下表1提供了示例性蛋白质的列表,其中所述的蛋白质可以在重组宿主细胞中表达从而有利于特定脂肪酸衍生物的生产。

[0103] 表1:基因名称

[0104]

基因名称	来源有机体	酶的名称	编号 No.	EC 号	示例性用途
1.脂肪酸生产的增加/产物生产的增加					
accA	大肠杆菌, 乳酸菌	乙酰基-CoA 羧化酶, 亚基 A(α -羧基转移酶)	AAC73296, NP_414727	6.4.1.2	增加丙二酰基-CoA 的生产
accB	大肠杆菌, 乳酸菌	乙酰基-CoA 羧化酶, 亚基 B(BCCP: 生物素羧基载体蛋白质)	NP_417721	6.4.1.2	增加丙二酰基-CoA 的生产
accC	大肠杆菌, 乳酸菌	乙酰基-CoA 羧化酶, 亚基 C(生物素羧化酶)	NP_417722	6.4.1.2, 6.3.4.14	增加丙二酰基-CoA 的生产

[0105]

基因名称	来源有机体	酶的名称	编号 No.	EC 号	示例性用途
accD	大肠杆菌, 乳酸菌	乙酰基-CoA 羧化酶, 亚基 D(β -羧基转移酶)	NP_416819	6.4.1.2	增加丙二酰基-CoA 的生产
fadD	大肠杆菌 <i>W3110</i>	酰基-CoA 合酶	AP_002424	2.3.1.86, 6.2.1.3	增加脂肪酸的生产
fabA	大肠杆菌 <i>K12</i>	β -羟基癸酰基硫酯脱水酶/异构酶	NP_415474	4.2.1.60	增加脂肪酰基-ACP/CoA 的生产
fabB	大肠杆菌	3-氧酰基-[酰基载体蛋白质]合酶 I	BAA16180	2.3.1.41	增加脂肪酰基-ACP/CoA 的生产
fabD	大肠杆菌 <i>K12</i>	[酰基载体蛋白质] S-丙二酰基转移酶	AAC74176	2.3.1.39	增加脂肪酰基-ACP/CoA 的生产
fabF	大肠杆菌 <i>K12</i>	3-氧酰基-[酰基载体蛋白质]合酶 II	AAC74179	2.3.1.179	增加脂肪酰基-ACP/CoA 的生产
fabG	大肠杆菌 <i>K12</i>	3-氧酰基-[酰基载体蛋白质]还原酶	AAC74177	1.1.1.100	增加脂肪酰基-ACP/CoA 的生产
fabH	大肠杆菌 <i>K12</i>	3-氧酰基-[酰基载体蛋白质]合酶 III	AAC74175	2.3.1.180	增加脂肪酰基-ACP/CoA 的生产
fabI	大肠杆菌 <i>K12</i>	烯酰基-[酰基载体蛋白质]还原酶	NP_415804	1.3.1.9	增加脂肪酰基-ACP/CoA 的生产
fabR	大肠杆菌 <i>K12</i>	转录抑制因子	NP_418398	无	调控不饱和脂肪酸的生产
fabV	霍乱弧菌	烯酰基-[酰基载体蛋白质]还原酶	YP_001217283	1.3.1.9	增加脂肪酰基-ACP/CoA 的生产
fabZ	大肠杆菌 <i>K12</i>	(3R)-羟基肉豆蔻酰基酰基载体蛋白质脱水酶	NP_414722	4.2.1.-	增加脂肪酰基-ACP/CoA 的生产
fadE	大肠杆菌 <i>K13</i>	酰基-CoA 脱氢酶	AAC73325	1.3.99.3, 1.3.99.-	减少脂肪酸的降解

基因名称	来源有机体	酶的名称	编号 No.	EC 号	示例性用途
fadR	大肠杆菌	转录调节蛋白质	NP_415705	无	阻断或逆转脂肪酸的降解
2. 链长度的控制					
tesA(具有或不具有前导序列)	大肠杆菌	硫酯酶 -前导序列为氨基酸 1-26	P0ADA1	3.1.2.-, 3.1.1.5	C18 链长
tesA(不具有前导序列)	大肠杆菌	硫酯酶	AAC73596, NP_415027	3.1.2.-, 3.1.1.5	C18:0 链长
tesA(与辛酸络合的大肠杆菌硫酯酶 I 突变体)	大肠杆菌	硫酯酶	L109P	3.1.2.-, 3.1.1.5	< C18 链长
fatB1	加州月桂 (<i>Umbellularia californica</i>)	硫酯酶	Q41635	3.1.2.14	C12:0 链长
fatB2	萼距花(<i>Cuphea hookeriana</i>)	硫酯酶	AAC49269	3.1.2.14	C8:0 -C10:0 链长
fatB3	萼距花	硫酯酶	AAC72881	3.1.2.14	C14:0 -C16:0 链长
fatB	香樟 (<i>Cinnamomum camphora</i>)	硫酯酶	Q39473	3.1.2.14	C14:0 链长
fatB	拟南芥 (<i>Arabidopsis thaliana</i>)	硫酯酶	CAA85388	3.1.2.14	C16:1 链长
fatA1	向日葵 (<i>Helianthus annuus</i>)	硫酯酶	AAL79361	3.1.2.14	C18:1 链长
atfata	拟南芥	硫酯酶	NP_189147, NP_193041	3.1.2.14	C18:1 链长
fatA	芥菜 (<i>Brassica juncea</i>)	硫酯酶	CAC39106	3.1.2.14	C18:1 链长
fatA	萼距花	硫酯酶	AAC72883	3.1.2.14	C18:1 链长
tes	<i>Photobacterium</i>	硫酯酶	YP_130990	3.1.2.14	链长

[0106]

基因名称	来源有机体	酶的名称	编号 No.	EC 号	示例性用途
	<i>profundum</i>				
tesB	大肠杆菌	硫酯酶	NP_414986	3.1.2.14	链长
fadM	大肠杆菌	硫酯酶	NP_414977	3.1.2.14	链长
yciA	大肠杆菌	硫酯酶	NP_415769	3.1.2.14	链长
ybgC	大肠杆菌	硫酯酶	NP_415264	3.1.2.14	链长
3. 饱和水平的控制*					
Sfa	大肠杆菌	fabA 的抑制因子	AAN79592, AAC44390	无	增加单不饱和脂肪酸
fabA	大肠杆菌 K12	β -羟基癸酰基硫酯脱水酶/异构酶	NP_415474	4.2.1.60	生产不饱和脂肪酸
GnsA	大肠杆菌	secG 沉默突变的抑制因子	ABD18647.1	无	增加不饱和脂肪酸酯
GnsB	大肠杆菌	secG 沉默突变的抑制因子	AAC74076.1	无	增加不饱和脂肪酸酯
fabB	大肠杆菌	3-氧酰基-[酰基载体蛋白质]合酶 I	BAA16180	2.3.1.41	调控不饱和脂肪酸的生产
des	枯草芽孢杆菌	D5 脂肪酰基去饱和酶	O34653	1.14.19	调控不饱和脂肪酸的生产
4. 产物输出：蜡的生产					
AT3G51970	拟南芥	长链醇 O-脂肪-酰基转移酶	NP_190765	2.3.1.26	蜡的生产
ELO1	安格斯毕赤酵母 (<i>Pichia angusta</i>)	脂肪酸延长酶	BAD98251	2.3.1.-	生产极长链长的脂肪酸
plsC	酿酒酵母	酰基转移酶	AAA16514	2.3.1.51	蜡的生产
DAGAT/DGAT	拟南芥	二酰基甘油酰基转移酶	AAF19262	2.3.1.20	蜡的生产
hWS	智人 (<i>Homo sapiens</i>)	酰基-CoA 蜡醇酰基转移酶	AAX48018	2.3.1.20	蜡的生产
aft1	不动杆菌菌种 ADP1	双功能蜡酯合酶/酰基-CoA:二酰基甘油酰基转移酶	AAO17391	2.3.1.20	蜡的生产
ES9	除烃海杆菌	蜡酯合酶	ABO21021	2.3.1.20	蜡的生产

[0107]

基因名称	来源有机体	酶的名称	编号 No.	EC 号	示例性用途
	(<i>Marinobacter hydrocarbonoclasticus</i>)				
mWS	霍霍巴 (<i>Simmondsia chinensis</i>)	蜡酯合酶	AAD38041	2.3.1.-	蜡的生产
5. 产物输出：脂肪醇的输出					
		硫酯酶(参见上文)			增加脂肪酸/脂肪醇的生产
BmFAR	家蚕	FAR(形成脂肪醇的酰基-CoA还原酶)	BAC79425	1.1.1.-	将酰基-CoA 转化为脂肪醇
acr1	不动杆菌菌种 ADPI	酰基-CoA 还原酶	YP_047869	1.2.1.42	将脂肪酰基-CoA 还原成脂肪醇
yqhD	大肠杆菌 W3110	醇脱氢酶	AP_003562	1.1.-.-	将脂肪醛还原为脂肪醇；增加脂肪醇的生产
alrA	不动杆菌菌种 ADPI	醇脱氢酶	CAG70252	1.1.-.-	将脂肪醛还原为脂肪醇
BmFAR	家蚕	FAR(形成脂肪醇的酰基-CoA还原酶)	BAC79425	1.1.1.-	将脂肪酰基-CoA 还原成脂肪醇
GTNG_1865	嗜热脱氮芽孢杆菌 (<i>Geobacillus thermodenitrificans</i>) NG80-2	长链醛脱氢酶	YP_001125970	1.2.1.3	将脂肪醛还原为脂肪醇
AAR	细长聚球藻	酰基-ACP 还原酶	YP_400611	1.2.1.42	将脂肪酰基-ACP/CoA 还原为脂肪醛
carB	耻垢分枝杆菌 (<i>Mycobacterium smegmatis</i>)	羧酸还原酶蛋白	YP_889972	6.2.1.3, 1.2.1.42	将脂肪酸还原为脂肪醛
FadD	大肠杆菌 K12	酰基-CoA 合成酶	NP_416319	6.2.1.3	将脂肪酸活化为脂肪酰基-CoA
atoB	欧文氏菌	乙酰基-CoA 乙	YP_049388	2.3.1.9	丁醇的生产

[0108]

基因名称	来源有机体	酶的名称	编号 No.	EC 号	示例性用途
		酰基转移酶			
hbd	溶纤维丁酸弧菌 (<i>Butyrivibrio</i> <i>bristolvens</i>)	β -羟基丁酰基 -CoA 脱氢酶	BAD51424	1.1.1.157	丁醇的生产
CPE0095	产气荚膜梭菌 (<i>Clostridium</i> <i>perfringens</i>)	巴豆酸酶丁酰基 -CoA 脱氢酶	BAB79801	4.2.1.55	丁醇的生产
bcd	钩形梭状芽胞 杆菌 (<i>Clostridium</i> <i>beijerinckii</i>)	丁酰基-CoA 脱 氢酶	AAM14583	1.3.99.2	丁醇的生产
ALDH	钩形梭状芽胞 杆菌	辅酶 A-酰化醛脱 氢酶	AAT66436	1.2.1.3	丁醇的生产
AdhE	大肠杆菌 <i>CFT073</i>	醛-醇脱氢酶	AAN80172	1.1.1.1 1.2.1.10	丁醇的生产
6. 脂肪醇乙酰基酯的输出					
		硫酯酶(参见上 文)			修饰输出
acr1	不动杆菌菌种 <i>ADPI</i>	酰基-CoA 还原 酶	YP_047869	1.2.1.42	修饰输出
yqhD	大肠杆菌 <i>K12</i>	醇脱氢酶	AP_003562	1.1.-.-	修饰输出
AAT	草莓 (<i>Fragaria</i> <i>x ananassa</i>)	醇 O-乙酰基转移 酶	AAG13130	2.3.1.84	修饰输出
7. 产物的出口					
AtMRP5	拟南芥	与拟南芥多药耐 药相关的	NP_171908	无	修饰产物出口的量
AmiS2	红球菌菌种	ABC 运送因子 AmiS2	JC5491	无	修饰产物出口的量
AtPGP1	拟南芥	拟南芥 p 糖蛋白 1	NP_181228	无	修饰产物出口的量
AcrA	<i>CandidatusProt</i> <i>ochlamydiaamo</i> <i>ebophila</i> <i>UWE25</i>	推断的多药外排 运送蛋白质 acrA	CAF23274	无	修饰产物出口的量

[0109]

[0110]

基因名称	来源有机体	酶的名称	编号 No.	EC 号	示例性用途
AcrB	<i>CandidatusProt ochlamydiaamo ebophila UWE25</i>	可能的多药外排 运送蛋白质 acrB	CAF23275	无	修饰产物出口的量
TolC	土拉弗朗西丝 菌 novicida 亚 种 (<i>Francisellatu larensis subsp. novicida</i>)	外膜蛋白质 (细 胞包膜的生物发 生)	ABD59001	无	修饰产物出口的量
AcrE	宋内志贺菌 Ss046	跨膜蛋白质影响 隔膜形成和细胞 膜的渗透性	YP_312213	无	修饰产物出口的量
AcrF	大肠杆菌	吡啶黄抗性蛋白 质 F	P24181	无	修饰产物出口的量
tll1619	细长嗜热聚球 蓝细菌 (<i>Thermosynec hococcus elongatus</i>) [BP-1]	多药外排运送蛋 白质	NP_682409. 1	无	修饰产物出口的量
tll0139	细长嗜热聚球 蓝细菌 [BP-1]	多药外排运送蛋 白质	NP_680930. 1	无	修饰产物出口的量
8. 发酵					
复制的检 查关口基 因					增加输出效率
umuD	宋内志贺菌 (<i>Shigellasonnei</i>) Ss046	DNA 聚合酶 V, 亚基	YP_310132	3.4.21.-	增加输出效率
umuC	大肠杆菌	DNA 聚合酶 V, 亚基	ABC42261	2.7.7.7	增加输出效率
pntA, pntB	福氏志贺氏菌 (<i>Shigella flexner</i> <i>i</i>)	NADH:NADPH 转氢酶(α 和 β 亚 基)	P07001, P0AB70	1.6.1.2	增加输出效率
9. 其他					

[0111]	基因名称	来源有机体	酶的名称	编号 No.	EC 号	示例性用途
	fabK	肺炎链球菌	反-2-烯酰基-ACP 还原酶 II	AAF98273	1.3.1.9	有助于脂肪酸的生物合成
	fabL	地衣芽孢杆菌 (<i>Bacillus licheniformis</i>) D SM 13	烯酰基-(酰基载体蛋白质)还原酶	AAU39821	1.3.1.9	有助于脂肪酸的生物合成
	fabM	链球菌突变体	反-2,顺-3-癸烯酰基-ACP 异构酶	DAA05501	4.2.1.17	有助于脂肪酸的生物合成

[0112] 脂肪酸的生产

[0113] 重组宿主细胞可以包括一种或多种多核苷酸序列,该多核苷酸序列包括编码硫酯酶(例如Enzyme Commission编号为EC 3.1.1.5或EC 3.1.2(例如EC 3.1.2.14))的开放阅读框,以及有助于蛋白质在重组宿主细胞中的表达的可操作地连接的调节序列。在重组宿主细胞中,相对于编码硫酯酶的相应的野生型基因而言,编码序列的开放阅读框和/或调节序列被修饰。相对于由相应的野生型基因在重组宿主细胞中表达的硫酯酶的活性而言,所述的硫酯酶在相应的宿主细胞中的活性被修饰。在一些实施方案中,包括脂肪酸的脂肪酸衍生物组合物通过在有效表达硫酯酶的条件下在碳源存在下培养重组细胞来生产。在相关的实施方案中,重组宿主细胞包括编码具有硫酯酶活性的多肽的多核苷酸,以及编码具有其他的脂肪酸生物合成酶活性的多肽的一种或多种其他的多核苷酸。在一些此类情况下,通过具有不同的脂肪酸生物合成酶活性的一种或多种酶将在硫酯酶作用下生产的脂肪酸转化为另一种脂肪酸衍生物,例如脂肪酯、脂肪醛、脂肪醇或碳氢化合物。

[0114] 可以通过修饰特定硫酯酶的表达来选择由此形成的脂肪酸或脂肪酸衍生物的链长。硫酯酶将影响所生产的脂肪酸衍生物的链长。可以通过修饰所选的硫酯酶(EC 3.1.2.14或EC 3.1.1.5)的表达来选择脂肪酸衍生物底物的链长。因此,可以改造宿主细胞,从而表达、过表达、衰减表达或不表达一种或多种所选的硫酯酶,其用于增加优选的脂肪酸衍生物底物的生产。例如可以通过表达偏爱生产C₁₀脂肪酸的硫酯酶,并衰减偏爱生产除了C₁₀脂肪酸以外的脂肪酸(例如偏爱生产C₁₄脂肪酸的硫酯酶)的硫酯酶来生产C₁₀脂肪酸。这将得到相对均匀的碳链长为10的脂肪酸群体。在其他的情况下,可以通过衰减生产非C₁₄脂肪酸的内源硫酯酶,并表达利用C₁₄-ACP的硫酯酶来生产C₁₄脂肪酸。在一些情况下,可以通过表达利用C₁₂-ACP的硫酯酶,并衰减生产非C₁₂脂肪酸的硫酯酶来生产C₁₂脂肪酸。例如可以通过表达偏爱生产C12脂肪酸的硫酯酶,并衰减偏爱生产除了C12脂肪酸以外的脂肪酸的硫酯酶来生产C12脂肪酸。这将得到相对均匀的碳链长为12的脂肪酸群体。由具有细胞外生产的几乎所有的脂肪酸衍生物的培养物培养基回收脂肪酸衍生物。可以使用本领域已知的方法(例如GC-FID)来分析由重组宿主细胞生产的脂肪酸衍生物组合物,以便测定特定的脂肪酸衍生物的分布,以及脂肪酸衍生物组合物的成分的链长和饱和度。可以使用本领域已知的方法(例如通过使用放射性前体、HPLC或在细胞溶解后的GC-MS)来证明乙酰基-CoA、丙二酰基-CoA和脂肪酸的过度生产。用于脂肪酸途径的硫酯酶及编码该硫酯酶的多核苷酸的其他非限定性实例在PCT公开申请号W02010/075483中提供,该文献明确地以引用方式并

入本文。

[0115] 脂肪醛的生产

[0116] 在一个实施方案中,重组宿主细胞生产脂肪醛。在一些实施方案中,由重组宿主细胞生产的脂肪酸被转化为脂肪醛。在一些实施方案中,由重组宿主细胞生产的脂肪醛接着被转化为脂肪醇或碳氢化合物。在一些实施方案中,天然的(内源的)脂肪醛生物合成多肽(例如醛还原酶)存在于宿主细胞(例如大肠杆菌)中,并有效地将脂肪醛转化为脂肪醇。在其他的实施方案中,天然的(内源的)脂肪醛生物合成多肽是过表达的。在其他的实施方案中,外源脂肪醛生物合成多肽被引入到重组宿主细胞中,并被表达或过表达。天然的或重组的宿主细胞可以包括编码如下酶的多核苷酸,其中所述的酶具有脂肪醛生物合成活性(例如脂肪醛生物合成多肽或脂肪醛生物合成多肽或酶)。在宿主细胞中表达或过表达脂肪醛生物合成酶时,脂肪醛被生产。经改造从而生产脂肪醛的重组宿主细胞通常将一些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而生产脂肪醛,其中所述的多肽具有脂肪醛生物合成活性,例如羧酸还原酶(CAR)活性。CarB为示例性的羧酸还原酶。在实施本发明公开的过程中,编码羧酸还原酶多肽的基因可以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SEQ ID NO:7的氨基酸序列。在其他的实施方案中,CarB多肽为SEQ ID NO:7的变体或突变体。羧酸还原酶(CAR)多肽及编码该羧酸还原酶(CAR)多肽的多核苷酸的实例包括但不限于FadD9 (EC6.2.1.-, UniProtKB Q50631, GenBank NP_217106, SEQ ID NO:34)、CarA (GenBank ABK75684)、CarB (GenBank YP889972; SEQ ID NO:33)以及在PCT公开号W02010/042664和美国专利号8,097,439中所述的相关多肽,这些文献均明确地以引用方式并入本文。在一些实施方案中,重组宿主细胞进一步包括编码硫酯酶的多核苷酸。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脂肪醛生物合成多肽(例如具有酰基-ACP还原酶(AAR)活性的多肽)的多核苷酸来生产脂肪醛。酰基-ACP还原酶在重组宿主细胞中的表达导致脂肪醛和脂肪醇的生产(参见图4)。在重组宿主细胞(例如大肠杆菌)中存在的天然的(内源的)醛还原酶可以将脂肪醛转化为脂肪醇。示例性的酰基-ACP还原酶多肽在PCT公开号W02009/140695和W0/2009/140696中有所描述,这两份文献均明确地以引用方式并入本文。通过在有效表达脂肪醛生物合成酶的条件下在碳源存在下培养宿主细胞来生产包括脂肪醛的组合物(例如脂肪醛组合物)。在一些实施方案中,脂肪醛组合物包括脂肪醛和脂肪醇。通常,由重组宿主细胞的细胞外环境(即,细胞培养物培养基)中回收脂肪醛组合物。

[0117] 脂肪醇的生产

[0118] 在一些实施方案中,重组宿主细胞包括编码如下多肽(例如酶)的多核苷酸,其中所述的多肽(例如酶)具有脂肪醇生物合成活性(脂肪醇生物合成多肽或脂肪醇生物合成酶),并且所述的脂肪醇由重组宿主细胞生产。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,而且,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)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,存在于宿主细胞中天然的(内源的)脂肪醛生物合成多肽(例如醛还原酶)足可以将脂肪醛转化为脂肪醇。而且,在其他的实施方案中,天然的(内源的)脂肪醛生物合成多肽是过表达的,并且在其他的实施方案中,外源的脂肪醛生物合成多肽被引入到重组宿主细胞中,并且被表达或过表达。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将脂肪醛转化为脂肪醇的脂肪醇生物合成活性。例如醇脱氢酶(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,醇脱氢酶是指能够将脂肪醛转化为醇(例如脂肪醇)的多肽。本领域的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。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不动杆菌菌种M-1 (SEQ ID NO:3)的AlrA或AlrA同源物(例如AlrAadpl (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WO 2007/136762、WO2008/119082和WO 2010/062480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)。

[0119] 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W02010/062480中有所描述,该文献明确地以引用方式并入本文。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。增加重组宿主细胞生产脂肪醇的策略包括通过在生产型宿主中过表达天然的脂肪酸生物合成基因和/或表达由不同的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,这样增加脂肪醇的生物合成。

[0120] 酯的生产

[0121] 如本文所用,术语“脂肪酯”可以用于指酯。本发明所指的脂肪酯可以由脂肪酸制得的任何酯,例如脂肪酸酯。在一些实施方案中,脂肪酯包括A侧链和B侧链。如本文所用,酯的“A侧链”是指与酯的羧酸氧连接的碳链。如本文所用,酯的“B侧链”是指包括酯的亲代羧酸的碳链。在其中脂肪酯衍生自脂肪酸生物合成途径的实施方案中,A侧链由醇提供,而B侧链由脂肪酸提供。任何醇都可以用于形成脂肪酯的A侧链。例如所述的醇可以衍生自脂肪

酸生物合成途径。备选地,所述的醇可以通过非脂肪酸的生物合成途径来生产。此外,外源可以提供醇。例如在脂肪酯由有机体生产的情况下,发酵液中可以供应醇。备选地,在脂肪酯由还可以生产醇的有机体生产的情况下,可以外源地供应羧酸,例如脂肪酸或醋酸。包括A侧链或B侧链的碳链可以为任何长度。在一个实施方案中,酯的A侧链为至少大约1、2、3、4、5、6、7、8、10、12、14、16或18个碳的长度。当脂肪酯为脂肪酸甲基酯时,酯的A侧链为1个碳的长度。当脂肪酯为脂肪酸乙基酯时,酯的A侧链为2个碳的长度。酯的B侧链可以为至少大约4、6、8、10、12、14、16、18、20、22、24或26个碳的长度。A侧链和/或B侧链可以是直链或支链的。支链可以具有一个或多个分支点。此外,支链可以包括环状支。此外,A侧链和/或B侧链可以是饱和的或不饱和的。如果是不饱和的,A侧链和/或B侧链则可以具有一个或多个不饱和点。在一个实施方案中,脂肪酯以生物合成的方式生产。在这种实施方案中,首先,脂肪酸被活化。“活化的”脂肪酸的非限定性实例为酰基-CoA、酰基ACP和酰基磷酸。酰基-CoA可以为脂肪酸生物合成或降解的直接产物。此外,酰基-CoA可以由游离的脂肪酸、CoA、和腺苷核苷酸三磷酸(ATP)合成。生产酰基-CoA的酶的实例为酰基-CoA合酶。在一些实施方案中,重组宿主细胞包括编码多肽的多核苷酸,所述多肽例如具有酯合酶活性的酶(酯合酶多肽或酯合酶)。

[0122] 脂肪酯通过酯合酶多肽催化的反应来生产,其中所述的酯合酶多肽在重组宿主细胞中表达或过表达。在一些实施方案中,通过在有效表达酯合酶的条件下在碳源存在下培养重组细胞来生产包括脂肪酯的组合物。在一些实施方案中,由细胞培养物回收脂肪酯组合物。酯合酶多肽包括例如分类为EC 2.3.1.75的酯合酶多肽,或催化酰基硫酯转化为脂肪酯的任何其他多肽,其包括但不限于硫酯酶、酯合酶、酰基-CoA:醇酰基转移酶、酰基转移酶、或脂肪酰基-CoA:脂肪醇酰基转移酶。例如多核苷酸可以编码蜡/dgat,其为得自霍霍巴、不动杆菌菌种菌株ADP、泊库岛食烷菌(*Alcanivorax borkumensis*)、绿脓杆菌(*Pseudomonas aeruginosa*)、*Fundibacter jadensis*、拟南芥、或*Alkaligenes eutrophus*中的双功能酯合酶/酰基-CoA:二酰基甘油酰基转移酶。在具体的实施方案中,酯合酶多肽为不动杆菌菌种的二酰基甘油O-酰基转移酶(蜡-dgaT;UniProtKB Q8GGG1,GenBank AA017391)或霍霍巴的蜡合酶(UniProtKB Q9XGY6,GenBank AAD38041)。在另一个实施方案中,酯合酶多肽为例ES9(得自除烃海杆菌DSM 8798,UniProtKB A3RE51(SEQ ID NO:6)的蜡酯合酶);除烃海杆菌DSM8789的ES8(GenBank Accession No.AB021021;SEQ ID NO:7);由ws2基因编码的GenBank AB021021;或者ES376(由除烃海杆菌DSM 8798,UniProtKB A3RE50,GenBank AB021020衍生的、ws1基因编码的另一种蜡质合酶)。在特定的实施方案中,编码酯合酶多肽的多核苷酸在重组宿主细胞中是过表达的。在一些实施方案中,脂肪酸酯由重组宿主细胞生产,其中所述的重组宿主细胞经改造,从而表达3种生物合成酶:硫酯酶、酰基-CoA合成酶(fadD)和酯合酶(例如三酶系统;参见图5)。在其他的实施方案中,脂肪酸酯由重组宿主细胞生产,其中所述的重组宿主细胞经改造,从而表达一种脂肪酸生物合成酶,即,酯合酶(例如单酶系统;参见图5)。适用于这些实施方案的酯合酶多肽及编码该酯合酶多肽的多核苷酸的非限定性实例包括在PCT公开号W02007/136762、W02008/119082和W0/2011/038134(三酶系统),以及W0/2011/038132(单酶系统)中所述的那些,这些文献均明确地以引用方式并入本文。重组宿主细胞可以在宿主细胞的细胞外环境中生产脂肪酯,例如脂肪酸甲基酯、脂肪酸乙基酯或蜡酯。

[0123] 碳氢化合物的生产

[0124] 本发明公开的这一方面至少部分基于这样的发现：改变脂肪醛生物合成多肽（例如酰基-ACP还原酶多肽（EC 6.4.1.2））和碳氢化合物生物合成多肽（例如脱羧基酶）在重组宿主细胞中的表达水平促进重组宿主细胞以增强方式生产碳氢化合物。在一个实施方案中，重组宿主细胞生产碳氢化合物，例如烷烃、烯烃（例如端烯烃或内烯烃）或酮。在一些实施方案中，通过脱羧基作用转化由重组宿主细胞生产的脂肪醛，从而除去碳原子而形成碳氢化合物。在其他的实施方案中，通过脱羧基作用转化由重组宿主细胞生产的脂肪酸，从而除去碳原子而形成端烯烃。在一些实施方案中，通过脱羧基作用转化酰基-ACP中间体，从而除去碳原子而形成内烯烃或酮（参见图6）。在一些实施方案中，重组宿主细胞包括编码如下多肽（酶）的多核苷酸，其中所述的多肽（酶）具有碳氢化合物生物合成活性（碳氢化合物生物合成多肽或碳氢化合物生物合成酶），并且通过在重组宿主细胞中表达或过表达碳氢化合物生物合成酶来生产碳氢化合物。蓝细菌的烷烃生物合成途径已经用于改造重组宿主细胞，以用于生产碳氢化合物（图6），其中所述的烷烃生物合成途径由酰基-酰基载体蛋白质还原酶（AAR）和醛脱羧基酶（ADC）组成，这两种酶一起将脂肪酸代谢的中间体转化为烷烃和烯烃。上述途径（通过该途径，在蓝细菌中饱和的酰基-ACP被转化为烷烃）中两个反应的第二个反应涉及由醛脱羧基酶（ADC）（其为铁样蛋白质，具有组成未知的双核金属辅因子）切断脂肪醛中间体的C1-C2键。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碳氢化合物，其中所述的多肽具有碳氢化合物生物合成活性，例如醛脱羧基酶（ADC）活性（例如EC 4.1.99.5）。可以编码醛脱羧基酶（用于本实施方案）的示例性多核苷酸包括但不限于在PCT公开号W02008/119082和W02009/140695中所述的那些以及在下表2中所示的那些序列，所述的这些文献明确地以引用方式并入本文。在一些实施方案中，重组宿主细胞进一步包括编码脂肪醛生物合成多肽的多核苷酸。在一些实施方案中，重组宿主细胞进一步包括编码酰基-ACP还原酶的多核苷酸。例如参见下表2。

[0125] 表2：示例性碳氢化合物生物合成的多核苷酸及多肽

蛋白质名称	多肽序列	核苷酸序列	序列
脱羧基酶(ADC)	SEQ ID NO: 35	SEQ ID NO: 36	细长聚球藻 PCC7942 YP.sub.--400610(Synpcc7942.sub.--1593)
Acyl-ACP 还原酶 (AAR)	SEQ ID NO: 37	SEQ ID NO: 38	细长聚球藻 PCC7942 YP_400611(Synpcc7942_1594)
脱羧基酶(ADC)	SEQ ID NO: 39	SEQ ID NO: 40	Prochlorococcus mariunus CCMP1986 PMM0532
Acyl-ACP 还原酶 (AAR)	SEQ ID NO: 41	SEQ ID NO: 42	原绿球藻 (<i>Prochlorococcus marinus</i>) CCMP1986 PMM0533(NP_892651)

[0127] 在一些实施方案中，通过在有效地表达酰基-CoA还原酶和脱羧基酶多核苷酸的条件在碳源存在下培养重组细胞来生产包括碳氢化合物的组合物。在一些实施方案中，碳氢化合物组合物包括饱和的或不饱和的碳氢化合物。而且，碳氢化合物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）中回收碳

氢化合物组合物。如本文所用,烷烃是指仅由碳(C)和氢(H)组成的饱和的碳氢化合物或化合物,其中这些原子通过单键连接在一起(即,它们是饱和的化合物)。烯烃和烯炔是指包括至少一个碳-碳双键的碳氢化合物(即,它们是不饱和的化合物)。就 α -烯炔或烯炔而言,端烯炔、 α -烯炔、端烯炔和1-烯炔是指相同的化合物,其具有化学式 C_xH_{2x} ,其与具有相似分子式的其他烯炔的不同之处在于碳氢化合物链的线性以及双键在第一或 α -位置处的位置。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碳氢化合物生物合成多肽(例如具有脱羧酶活性的多肽)的多核苷酸来生产端烯炔,例如在PCT公开号W02009/085278中所述,该文献明确地以引用方式并入本文。在一些实施方案中,重组宿主细胞进一步包括编码硫酯酶的多核苷酸。在其他的实施方案中,通过在重组宿主细胞中表达或过表达编码碳氢化合物生物合成多肽(例如具有OleA活性的多肽)的多核苷酸来生产酮,例如在PCT公开号W02008/147781中所述,该文献明确地以引用方式并入本文。在相关的实施方案中,通过在重组宿主细胞中表达或过表达编码碳氢化合物生物合成多肽(例如具有OleCD或OleBCD的多肽)的多核苷酸来生产内烯炔,例如在PCT公开号W02008/147781中所述,该文献明确地以引用方式并入本文。

[0128] 重组宿主细胞和细胞培养物

[0129] 用于增加重组宿主细胞生产脂肪酸衍生物的策略包括通过在生产型宿主中过表达天然的脂肪酸生物合成基因和/或表达由不同的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。如本文所用,重组宿主细胞或经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天然存在的基因元件使基因构成相对于相应的野生型宿主细胞而言被改变。此外,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有机体。作为微生物有机体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。在另一个实施方案中, 宿主细胞为变铅青链霉菌 (*Streptomyces lividans*) 细胞或灰链霉菌 (*Streptomyces murinus*) 细胞。在另一个实施方案中, 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、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, 例如在光存在下, 宿主细胞具有自养活性。在一些实施方案中, 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、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 (*Thermosynechococcus elongatus*) BP-1、绿硫细菌 (*Chlorobium tepidum*)、*Chlorojflexus auranticus*、*Chromatium vinosum*、深红红螺菌 (*Rhodospirillum rubrum*)、荚膜红细菌 (*Rhodobacter capsulatus*)、沼泽红假单胞菌 (*Rhodopseudomonas palustris*)、杨氏梭菌 (*Clostridium ljungdahlii*)、热纤梭菌 (*Clostridium thermocellum*)、产黄青霉菌 (*Penicillium chrysogenum*)、毕赤酵母菌 (*Pichia pastoris*)、酿酒酵母、粟酒酵母 (*Schizosaccharomyces pombe*)、萤光假单胞菌 (*Pseudomonas fluorescens*) 或运动发酵单胞菌 (*Zymomonas mobilis*) 的细胞。

[0130] 由重组宿主细胞生产脂肪酸衍生物

[0131] 可以通过包括本发明所述的菌株改善的重组宿主细胞来生产多种脂肪酸衍生物, 其包括但不限于脂肪酸、酰基-CoA、脂肪醛、短链和长链的醇、碳氢化合物 (例如烷烃、烯烃或炔烃, 例如端烯烃或内烯烃)、脂肪醇、酯 (例如蜡酯、脂肪酸酯 (例如甲基酯或乙基酯)) 和酮。在本发明公开的一些实施方案中, 相对于由相应的野生型宿主细胞的对照培养物所生产的相同脂肪酸衍生物的效价而言, 在特定的组合物中, 更高效价的脂肪酸衍生物为由重组宿主细胞生产的具有更高效价的特定类型的脂肪酸衍生物 (例如脂肪醇、脂肪酸酯或碳氢化合物)。在这种情况下, 脂肪酸衍生物组合物可以包括例如具有各种链长以及饱和或支化特征的脂肪醇的混合物。在本发明公开的其他实施方案中, 相对于由相应的野生型宿主细胞的对照培养物所生产的相同脂肪酸衍生物的效价而言, 在特定的组合物中, 更高效价的脂肪酸衍生物为具有更高效价的不同脂肪酸衍生物 (例如脂肪醛和脂肪醇, 或者脂肪酸和酯) 的组合。

[0132] 改造宿主细胞

[0133] 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,重组载体包括至少一个如下序列,其包括但不限于(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,其中所述的启动子定向于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。此类及它们的同种鉴定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分别融合到目标重组多肽中。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pET11d (Studier et al., Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990) 60-89)。由pTrc载体使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。由pET11d载体使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。这种病毒聚合酶通过宿主菌株BL21 (DE3)和HMS174 (DE3),由容留在T7gn1基因中的居民 λ 噬菌体在lacUV5启动子的转录控制之下来供应的。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PET11d (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. (supra)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。

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)的基因可以与所关注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。稳定地转化有诱导型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。

[0134] 宿主细胞

[0135] 如本文所用,改造的或重组的宿主细胞为用于生产本发明进一步所述的脂肪酸衍生物组合物的细胞。如果在宿主细胞中表达的一种或多种多核苷酸在相同的条件下相对于其在相应的野生型宿主细胞(例如对照细胞)中的表达被改变或修饰,则该宿主细胞称为改造的宿主细胞或重组宿主细胞。在本发明所述的发明公开的任一方面中,可以由真核植物、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中选择宿主细胞。在一些实施方案中,宿主细胞为光依赖的或固定碳的。在一些实施方案中,宿主细胞具有自养活性。如本文所述,多种宿主细胞可以用于生产脂肪酸衍生物。

[0136] 突变体或变体

[0137] 在一些实施方案中,所述的多肽为本发明所述任一种多肽的突变体或变体。如本文所用,术语突变体或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)被另一种碱性残基替代;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、加入、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所有的生物学功能(例如酶活性)。在一些实施方案中,所述的片段或变体保留了相应的野生型多肽的至少75%、至少80%、至少90%、至少95%、至少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)。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10%、至少25%、至少50%、至少75%、或者至少90%的改善。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)。

[0138] 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。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); and 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、 $MgCl_2$ 、 $MnCl_2$ 、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20毫微摩尔的待诱变的核酸,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 $MgCl_2$ 、0.5mM $MnCl_2$ 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,并评价由诱变核酸编码的多肽的活性。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有待引入到克隆DNA中的一个或多个突变。回收包括诱变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片段的混合物得到的PCR产物的拼装。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启动另一个反应的产物。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,作为DNA分子的随机发酵(基于序列同源性)的结果,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。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W01991/016427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大量核酸的多个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。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通量筛选以找到能够增加一种或多种脂肪酸衍生物的效价或产率的有利突变。

[0139] 培养重组宿主细胞和细胞培养物/发酵

[0140] 如本文所用,术语“发酵”广义而言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,术语“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,术语“其中载体的多核苷酸被表达的条件”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、供料速率和培养基成分。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。本发明所述的脂肪酸衍生物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。脂肪酸衍生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酸衍生物。

[0141] 由重组宿主细胞衍生的产物

[0142] 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 (参见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。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酸衍生物的生物产物 (例如根据本发明公开生产的脂肪酸衍生物) 不是使用可再生来源生产的, 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 (例如参见美国专利号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。通过比较各样品中稳定碳同位素比 ($^{13}\text{C}/^{12}\text{C}$), 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 $^{13}\text{C}/^{12}\text{C}$ 比是大气二氧化碳中该二氧化碳被固定时的 $^{13}\text{C}/^{12}\text{C}$ 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、C3植物 (阔叶植物)、C4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“C3” (或Calvin-Benson) 光合循环以及涉及“C4” (或Hatch-Slack) 光合循环的植物。在C3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。诸如阔叶树和针叶树之类的C3植物在温带气候地带中是主要的。在C4植物中, 涉及另一种酶 (磷酸烯醇式丙酮酸羧化酶) 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C3循环再固定。C4植物的实例是热带牧草、玉米和甘蔗。C4和C3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C4植物而言, 通常的值是大约-7至大约-13/密耳, 而对于C3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:

$$[0143] \quad \delta^{13}\text{C} (\text{‰}) = [(^{13}\text{C}/^{12}\text{C})_{\text{样品}} - (^{13}\text{C}/^{12}\text{C})_{\text{标准品}}] / (^{13}\text{C}/^{12}\text{C})_{\text{标准品}} \times 1000$$

[0144] 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的每单位密耳偏差的标记为 $\delta^{13}\text{C}$ 。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,将根据本发明公开所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.,1of Vol.I of the IUPAC Environmental Analytical Chemistry Series(Lewis Publishers,Inc.)3-74,(1992))。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)。 ^{14}C 在大气 CO_2 中以及由此在生物圈中的浓度在19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承诺。可以通过加速器质谱(AMS)测量 ^{14}C ,结果以“现代碳的比例”(fM)的单位给出。fM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”或“fM”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HoxI和HoxII)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HoxI(参考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),fM为大约1.1。本发明所述的组合物包括可以fM ^{14}C 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fM ^{14}C 、大约1至大约1.5的fM ^{14}C 、大约1.04至大约1.18的fM ^{14}C 、或者大约1.111至大约1.124的fM ^{14}C 。

[0145] 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。这意味着诸如玉米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该材料稀释50%,则它将给出约54pMC的放射性碳标记。通过指定“100%”等于107.5pMC并且指定为“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酸衍生物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,本发明所述的脂肪酸衍生物可以具有大约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,本发明所述的脂肪酸衍生物可以具有大约90、91、92、93、94或94.2的pMC。

[0146] 筛选由重组宿主细胞生产的脂肪酸衍生物组合物

[0147] 为了测定所述的这些条件是否足以允许表达,可以将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,以测定这些条件是否允许表达。例如可以针对所需产物的存在情况,测定在其中生长有宿主细胞的样品或培养基中的宿主细胞。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。使用用于“总脂肪物质”的GC/FID测定方法在96孔板中、1升和5升滑石以及1000L中试工厂中筛选重组宿主细胞培养物。

[0148] 脂肪酸衍生物组合物的用途

[0149] 脂肪酸为具有长的脂肪族尾部(链)的羧酸,其可以是饱和的或不饱和的。大部分天然形成的脂肪酸都具有偶数个碳原子(4至28)的链。脂肪酸通常衍生自甘油三酸酯。当脂肪酸未与其他分子连接时,它们称为“游离的”脂肪酸。工业上,通常通过水解甘油三酸酯来水解脂肪酸,并除去甘油。棕榈、大豆、油菜籽、椰子油和向日葵油是目前最普通的脂肪酸来源。由这些油衍生的大部分脂肪酸用于人类食品。椰子油和棕榈坚果油主要由12和14个碳的脂肪酸构成。这些油特别适用于进一步加工成用于洗涤剂和清洁剂、以及化妆品的表面活性剂。棕榈、大豆、油菜籽和向日葵油、以及诸如牛脂之类的动物脂肪主要包括长链脂肪酸(例如C18,其为饱和的或不饱和的),其用作聚合物用途和润滑剂的原材料。生态学和毒物学研究表明基于可再生来源的脂肪酸衍生产物比基于石油化学品的物质具有更有利的性质。脂肪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的范围。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

个值所限定的范围,该脂肪醇包括碳链,其长度为8、9、10、11、12、13、14、15、16、17、18、19、20、21或22个碳。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的范围的量存在。酯具有许多商业用途。例如生物柴油(备选燃料)由酯(例如脂肪酸甲基酯、脂肪酸乙基酯等)构成。一些低分子量的酯是挥发性的,并具有令人愉快的香味,这使得它们可以用作香料或风味剂。此外,酯作为漆、颜料和清漆的溶剂。此外,一些天然形成的物质(例如蜡、脂肪和油)由酯构成。此外,酯还可以用作树脂和增塑剂中的软化剂、阻燃剂、以及汽油和油中的添加剂。此外,酯可以用于制造聚合物、膜、纺织品、染料和药物。碳氢化合物具有许多商业用途。例如短链烷烃被用作燃料。长链烷烃(例如5个碳至16个碳)可以用作运输燃料(例如汽油、柴油或航空燃料)。具有多于16个碳原子的烷烃是燃油和润滑油的重要成分。甚至更长的烷烃(其在室温下为固态)可以用作例如石蜡。此外,长链烷烃可以断裂,从而生产具有商业价值的短链碳氢化合物。与短链烷烃相似,短链烯烃用于运输燃料。长链烯烃用于塑料、润滑剂和合成润滑剂。此外,烯烃可以用作生产醇、酯、增塑剂、表面活性剂、叔胺、增强型石油采收剂、脂肪酸、硫醇、烯基琥珀酸酐、环氧化物、氯化烷烃、氯化烯烃、蜡、燃料添加剂和阻流剂的供料。酮在商业上用作溶剂。例如丙酮通常用作溶剂,而且其还是用于制造聚合物的原材料。此外,酮还可以用于漆、颜料、爆炸物、香精和纺织品加工。此外,酮用于生产醇、烯烃、烷烃、亚胺和烯胺。润滑剂通常由烯烃构成,特别是聚烯烃和 α -烯烃。润滑剂可以由原油精炼,或者使用由原油精炼的原材料来制造。由原油获得这些专门的化学品需要高额的金融投资以及大量的能量。此外,由于原油中的长链碳氢化合物经常断裂而产生较小的单体,所以它们还是无效的方法。随后,这些单体被用作制造更复杂的专门化学品的原材料。通过以下实例进一步说明本发明公开。提供这些实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

[0150] 实施例1

[0151] 生产型宿主的修饰—酰基-CoA脱氢酶的衰减

[0152] 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。

[0153] 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:

[0154] 使用以下两个引物形成fadE的删除:

[0155] Del-fadE-F

[0156] 5'-AAAAACAGCAACAATGTGAGCTTTGTTGTAATTATATTGTAAACATATTGATTCCGGGGATCCGTCGACC (SEQ ID NO:9); 和

[0157] Del-fadE-R

[0158] 5'-AAACGGAGCCTTTCGGCTCCGTTATTCATTTACGCGGCTTCAACTTTCCTGTAGGCTGGAGCTGC

TTC (SEQ ID NO:10)

[0159] 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℃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℃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基因的两侧。

[0160] 证明fadE被删除的引物为:

[0161] fadE-L2 5'-CGGGCAGGTGCTATGACCAGGAC (SEQ ID NO:11); 和

[0162] fadE-R1 5'-CGCGGCGTTGACCGGCAGCCTGG (SEQ ID NO:12)

[0163] 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。fadE基因的删除不影响脂肪酸衍生物的生产 (图7)。本发明描述了多种示例性宿主细胞菌株, 其实例如下表3所述。

[0164] 表3: 大肠杆菌菌株的基因特征

菌株	基因特征
DV2	MG1655 F-, λ-, ilvG-, rfb-50, rph-1, ΔfhuA::FRT, ΔfadE::FRT
DV2.1	DV2 fabB::fabB[A329V]
D178	DV2.1 entD::FRT_P _{T5} _entD
EG149	D178 ΔinsH-11:: P _{LACUV5} -iFAB138
V642	EG149 rph+
SL313	V642 lacIZ::P _{A1} 'tesA/pDG109
V668	V642 ilvG ⁺
LC397	V668 lacIZ::P _{TRC} 'tesA(var)kan
SL571	V668 lacIZ:: P _{TRC} 'tesA(var)_FRT
LC942	SL571 attTn7:: P _{TRC} 'tesA(var)
DG16	LC942/pLC56
V940	LC397/pV171.1
D851	SL571 yijP::Tn5-cat/pV171.1
BD64	DV2 ΔinsH-11:: P _{LACUV5} -iFAB138 loxP_P _{T5} _fadR
DAM1	DV2 attTn7:: P _{TRC} _tesA_fadD
Shu.002	DV2 ΔinsH-11:: P _{T5} -iFAB138 loxP_P _{T5} _fadR

[0167] 质粒: pDG109、pLC56和pV171.1均为具有carB和tesA可变表达的pCL_P_{trc}-carB_

tesA_alrA_fabB_fadR操纵子。iFAB138为SEQ ID NO:19。

[0168] 实施例2

[0169] 通过脂肪酸合成途径的流量增加—乙酰基CoA羧化酶所介导

[0170] 脂肪酯的生产:

[0171] 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。在本实施例中,合成acc操纵子[谷氨酸棒状杆菌accABCD($\pm$ 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。在一个方法中,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)。示例性的谷氨酸棒状杆菌birA DNA序列在SEQ ID NO:55和SEQ ID NO:56中示出。谷氨酸棒状杆菌birA蛋白质序列在SEQ ID NO:57中示出。

[0172] 谷氨酸棒状杆菌acc基因的合成操纵子在OP80(参见W02008/119082,其以引用方式并入本文)的Ptrc1-accDACB、Ptrc3-accDACB、Ptrc1-accCBDA和Ptrc3-CBDA中以以下方式克隆。Ptrc1和Ptrc3为常用的Ptrc启动子的衍生物,其使得目标基因的表达衰减。注意由染色体DNA扩增天然序列,因为其显示出了有利的密码子使用(accCB中,仅Arg6的密码子被改变)。谷氨酸棒状杆菌birA基因为通过基因合成优化并获得的密码子。在所有4种操纵子构建体中,谷氨酸棒状杆菌birA基因克隆到acc基因的下游中。下面,我们将操纵子构造accDACB称为accD⁻,而将操纵子构造accDACB+birA称为accD⁺。将所得的质粒转化到大肠杆菌DAM 1_i377中,所述的质粒包括由大肠杆菌得到的无前导序列的硫酯酶'tesA和酰基-CoA合成酶fadD以及由除烃海杆菌得到的酯合酶9(ES9)的完整拷贝(SEQ ID NO:6)。所有基因都在Ptrc启动子的控制之下。将所述的菌株在摇瓶中在5NBT培养基(如下文所述)中生长,并针对丙二酰基-CoA,使用下文所述的短链-CoA测定来分析所述的菌株。图8显示8种谷氨酸棒状杆菌acc $\pm$ birA构建体中的6种在对数期中显示出丙二酰基-CoA水平升高,证明了它们在大肠杆菌中的功能性。注意birA的共表达进一步增加了丙二酰基-CoA在ptrc1/3_accDACB菌株中的水平,特别是具有包括Ptrc3-accDACB-birA操纵子构造(质粒pAS119.50D;SEQ ID NO:62)的质粒时更是如此。

[0173] 为了测试组合panK与acc-birA共表达的作用,将优化的panK基因克隆到ptrc1/3_accDACB-birA中birA的下游。泛酸酯激酶panK(或CoaA)在辅酶A的生物合成中催化第一步骤,其中所述的辅酶A是与许多反应(例如乙酰基-CoA(乙酰基-CoA羧化酶的底物)的形成)有关的必需辅助子。将所得的质粒转化到DAM1_i377中,使DAM1_i377在摇瓶中在5NBT(+TVS1)培养基中生长,并使用以下所述的方法针对短链CoA来分析所述的菌株。如图9所示,在对数期,panK共表达进一步增加丙二酰基-CoA的水平,并且还增加乙酰基-CoA的水平,证明panK可以进一步增加丙二酰基-CoA的水平。在另一个大肠杆菌生产型宿主中,在具有acc基因和不具有acc基因的条件下,通过表达酯合酶9(SEQ ID NO:6)来评价共表达乙酰基-CoA羧化酶复合物对脂肪酯生产的影响。更具体而言,将质粒OP80(载体对照)、pDS57(具有ES9)、pDS57-accD⁻(具有ES9和accDACB)或pDS57-accD⁺(具有ES9和accDACB-birA;SEQ ID

NO:63) 转化到大肠杆菌菌株DV2中,并在补充有100mg/L壮观霉素的LB平板上选择相应的转化子。

[0174] 将各质粒的2个转化子独立地接种到补充有100mg/L壮观霉素的LB培养基上,并在32℃下生长5至8小时。将培养物以30倍稀释至具有以下组成的基本培养基中:0.5g/L NaCl、1mM $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ 、0.1mM CaCl_2 、2g/L NH_4Cl 、3g/L KH_2PO_4 、6g/L Na_2HPO_4 、1mg/L硫胺、1x微量金属溶液、10mg/L柠檬酸铁、100mM Bis-Tris (pH7.0)、30g/L葡萄糖和100mg/L壮观霉素。在32℃下过夜生长后,将培养物以4个重复10倍稀释至相同组成的基本培养基中,其中所述的培养基与之前的基本培养基的不同之处在于该培养基包括1g/L而非2g/L NH_4Cl ,并且补充有1mM IPTG和2% (v/v) 甲醇。然后,使所得的培养物在摇瓶中在32℃下生长。通过具有火焰离子化检测器的气相色谱法 (GC-FID) 来分析脂肪酸甲基酯 (FAME) 的生产。使用醋酸丁酯以1:1体积/体积的比例提取样品。在涡流处理后,将样品离心,并通过气相色谱法 (GC) 分析有机相。分析条件如下,仪器:Trace GC Ultra (Thermo Electron Corporation),并装配有火焰离子检测器 (FID);柱:得自Thermo Electron Corporation的DB-1 (1%二苯基硅氧烷;99%二甲基硅氧烷) CO1UFM 1/0.1/5 01DET,相pH5,FT:0.4 μm ,长5m,id:0.1mm;进入条件:250℃无分流,根据样品浓度(分导流为75mL/m)使用的3.8m 1/25分流方法;载气,流速:氦气,3.0mL/m;模块温度:330℃;烤箱温度:0.5m,保持在50℃,100℃/m至330℃,0.5m保持在330℃;检测器温度:300℃;注射体积:2 μL ;运行时间/流速:6.3m/3.0mL/m(无分流方法),3.8m/1.5mL/m(分流1/25方法),3.04m/1.2mL/m(分流1/50方法)。生产的FAME示于图10中。ES9本身在大肠杆菌DV2中的表达使得FAME的生产高于对照DV2OP80。谷氨酸棒状杆菌乙酰基-CoA羧化酶复合物的共表达使得FAME增加为大约1.5倍,而谷氨酸棒状杆菌生物素蛋白质连接酶的额外表达使得FAME增加为大约5倍。这些结果表明丙二酰基-CoA的供应增加会改善ES9在大肠杆菌中将脂肪酸生物合成机的中间体转化为脂肪酸甲基酯的能力。

[0175] 短链CoA测试:使用0.467mL 10%TCA制备15mL Falcon管(用巴豆酰基-CoA作为内部标准品),并覆盖2mL硅油。将管在冰上冷冻,并将相当于1mL ($\text{OD}_{600}=31.2$) 的发酵液小心地分层放置在硅油的顶上。在4℃下,以11400g将样品离心4个4分钟的周期。对各样品而言,移出400mL TCA/细胞提取物的等分液,并放置在新的Eppendorf管中,使用1mL辛酸(处于CHC13)中和。在涡流处理后,将样品在13000g下离心30秒。使用0.2 μm PTFE注射器式过滤器过滤200mL顶层,然后进行LC-MS/MS分析。

[0176] 试验中使用的培养基的描述:

[0177]

培养基 ID						
4N-BT	5N-BT	FA2	FA2.1	FA2.3	浓度	组分
0.5	0.5	0.5	0.5	0.5	g/L	NaCl
2	2	2	2	2	g/L	NH ₄ Cl
3	3	3	3	3	g/L	KH ₂ PO ₄
6	6	6	6	6	g/L	Na ₂ PO ₄
1	1	1	1	1	mM	MgSO ₄
0.1	0.1	0.1	0.1	0.1	mM	CaCl ₂
1	1	1	1	1	mg/L	硫胺
0.2	0.2	0.1	0.1	0.1	M	Bis-Tris pH7
0.1	0.1	0.05	0.1	0.1	%	Triton X-100
1	1	1	1	1	x	微量矿物质
27	27	10	10	10	mg/L	FeCl ₂ ·6H ₂ O
40	50	30	30	35	g/L	葡萄糖

[0178] 1000倍浓缩的微量维生素溶液

[0179] 0.06g/L 核黄素

[0180] 6g/L 烟酸

[0181] 5.4g/L 泛酸

[0182] 1.4g/L 维生素B6

[0183] 0.06g/L 生物素

[0184] 0.01g/L 叶酸

[0185] 1000倍浓缩的微量金属溶液

[0186] 2mL/L 浓盐酸

[0187] 0.5g/L 硼酸

[0188] 1.9g/L 硫酸铜,五水合物,USP

[0189] 1g/L 无水氯化锌

[0190] 2g/L 无水钼酸钠

[0191] 2g/L 无水氯化钙

[0192] 脂肪醇的生产

[0193] 通过在大肠杆菌DV2中在具有和不具有acc的条件下表达由细长聚球藻得到的酰基-ACP还原酶(AAR) (SEQ ID NO:38) 来评价共表达乙酰基-CoA羧化酶复合物对脂肪醇生产的影响。由于当accD+操纵子构造与酯合酶共表达时会给出最佳的结果,所以选择该构造(参见之前的实施例)。使用融合技术(Clontech Laboratories, Inc., Mountain View, CA), 将accDABC-birA操纵子克隆到pLS9-185 (pCL1920衍生物) 中aar基因的下游。将所得的质粒转化大肠杆菌DV2中,并在补充有100mg/L壮观霉素的LB平板上选择相应的转化子。所生产的脂肪醇示于图11中。AAR和accD+共表达使得脂肪醇效价与仅具有AAR的对照(pLS9-185)相比增加为大约1.5倍。数据是可重复的(由三个重复的样品显示)。这些结果证明当使用这种酰基-ACP还原酶时,丙二酰基-CoA水平增加使得脂肪酸的生产得到改善。此外,实施例3描述了acc基因与完整的fab操纵子共表达。

[0194] 实施例3

[0195] 通过脂肪酸合成途径的流量增加—iFAB

[0196] 脂肪酸衍生物的生产：

[0197] 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共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 (表3) 的基因组相组合，在lacUV5启动子的控制下，并整合到IS5-11位点中。评价包括iFAB130-145的16个菌株。iFAB130-145的详细结构示于表4和5中。

[0198] 表4：在iFAB130-145中使用得自不同物种的成分

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

[0200] 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，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))。iFAB130-145的具体组成见下表5。

[0201] 表5：iFAB130-145的组成

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

[0203] 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12A和12B中。通过该筛选,确定最佳的构建体为具有iFAB138的DV2。iFAB138在EG149的基因组中的序列示于SEQ ID NO:19中。

[0204] 脂肪酯的生产

[0205] 将全部合成fab操纵子整合到的大肠杆菌染色体中,并通过在大肠杆菌DAM1 pDS57中表达,针对增加的FAME生产来评价所述的操纵子。此外,由谷氨酸棒状杆菌得到的4种合成acc操纵子被共表达,并针对改善的FAME生产率评价这些操纵子。获得以更快的速率和更高的效价生产FAME的多种菌株。将16种不同的iFAB操纵子(表5)置于lacUV5启动子的控制之下,并整合到大肠杆菌DAM1的IS5-11位点中。这些菌株命名为DAM1 ifab130至145。使用pDS57(包括酯合酶377)或共表达不同类型的acc操纵子的pDS57(参见上文)转化这些菌株,以用于评价FAME的生产。示例性质粒在表6中有所描述。

[0206] 表6:包括酯合酶ES9(得自除烃海杆菌)和合成acc操纵子(得自谷氨酸棒状杆菌)的质粒

质粒	基因
pTB.071	pDS57-accCBDA
pTB.072	pDS57-accCBDA-birA
pTB.073	pDS57-accDACB
pTB.074	pDS57-accDACB-birA

[0208] pDS57=pCL_{ptrc}-ES9

[0209] 在32℃下,在96孔板(4NBT培养基)、摇瓶(5NBT培养基)(关于培养基的描述参见上文)和发酵罐中分析DAM1ifab菌株。在96孔板和摇瓶中得到最佳结果,其中具有pDS57-acc-birA质粒的多种DAM1ifab菌株显示出更高的FAME效价。具体而言,具有pDS57-accDACB-birA的DAM1ifab131、ifab135、ifab137、ifab138和ifab143显示出效价改善20-40%,表明在这些菌株中,取得了通过脂肪酸途径的更高流量,这得到了更好的产物形成速率(这些结果在多个独立的试验中是可重复的)。

[0210] 过表达fabH和fabI对脂肪酸甲基酯(FAME)生产的影响

[0211] 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过表达天然的脂肪酸生物合成基因,和表达异源脂肪酸生物合成基因。fabH和fabI是已经显示出反馈抑制的2种脂肪酸生物合成酶(Heath and Rock, JBC 271:1833-1836(1996))。进行研究以测定FabH和FabI是否限制FAME生产的速率。fabH和fabI同源物(得自大肠杆菌、枯草芽孢杆菌、贝氏不动杆菌ADP1、Marinobacter aquaeoli VT8和不透明红球菌)作为合成的操纵子进行过表达,并在大肠杆菌DAM1pDS57(据观察,该菌株为良好的FAME生产者)中进行评价。在一种方法中,由累积蜡(贝氏不动杆菌、Marinobacter aquaeoli)或三酰基甘油酯(不透明红球菌)的有机体构建fabHfabI操纵子,并将其整合到大肠杆菌DAM1 pDS57的染色体中。在相关的方法中,由谷氨酸棒杆菌得到的合成acc操纵子被共表达(如上文实施例2中所述)。构建11

种不同的fabHI操纵子(体外拼装),如表7中概括的那样。将fabHI操纵子置于IPTG诱导的lacUV5启动子的控制之下,并整合到大肠杆菌DAM1的IS5-11位点中。使用pDS57(包括酯合酶377)或共表达不同类型的acc操纵子的pDS57转化这些菌株,以用于评价FAME的生产。

[0212] 表7:整合的fabHI操纵子的基因型

菌株	其他fab操纵子的基因型	质粒
stEP117	DAM1 Δ insH::P _{LACUV5} (snyRBS) EcfabH (snyRBS) BsfabI::kan	pDS57
stEP118	DAM1 Δ insH::P _{LACUV5} (snyRBS) EcfabH (snyRBS) BsfabL::kan	pDS57
stEP127	DAM1 Δ insH::P _{LACUV5} (EcRBS) EcfabH (EcRBS) BsfabI::kan	pDS57
stEP128	DAM1 Δ insH::P _{LACUV5} (EcRBS) EcfabH (EcRBS) BsfabL::kan	pDS57
stEP129	DAM1 Δ insH::P _{LACUV5} (EcRBS) ADP1fabH (EcRBS) ADP1fabI::kan	pDS57
stEP130	DAM1 Δ insH::P _{LACUV5} (snyRBS) ADP1fabH (synRBS) ADP1fabI::kan	pDS57
stEP131	DAM1 Δ insH::P _{LACUV5} (snyRBS) VT8fabH1 (synRBS) VT8fabI::kan	pDS57
stEP132	DAM1 Δ insH::P _{LACUV5} (snyRBS) VT8fabH2 (synRBS) VT8fabI::kan	pDS57
stEP133	DAM1 Δ insH::P _{LACUV5} (EcRBS) VT8fabH1 (synRBS) VT8fabI::kan	pDS57
stEP134	DAM1 Δ insH::P _{LACUV5} (EcRBS) VT8fabH2 (synRBS) VT8fabI::kan	pDS57
stEP151	DAM1 Δ insH::P _{LACUV5} (snyRBS) RofabI (synRBS) RofabH::kan	pDS57
stEP153	DAM1 Δ insH::P _{LACUV5} (EcRBS) ADP1fabH (EcRBS) ADP1fabI::kan	pDS57-accCBDA
stEP154	DAM1 Δ insH::P _{LACUV5} (EcRBS) ADP1fabH (EcRBS) ADP1fabI::kan	pDS57-accDACB
stEP155	DAM1 Δ insH::P _{LACUV5} (EcRBS) ADP1fabH (EcRBS) ADP1fabI::kan	pDS57-accCBDA-birA
stEP156	DAM1 Δ insH::P _{LACUV5} (EcRBS) ADP1fabH (EcRBS) ADP1fabI::kan	pDS57-accDACB-birA
stEP157	DAM1 Δ insH::P _{LACUV5} (snyRBS) EcfabH (synRBS) BsfabI::kan	pDS57-accCBDA
stEP158	DAM1 Δ insH::P _{LACUV5} (snyRBS) EcfabH (synRBS) BsfabI::kan	pDS57-accCBDA-bvirA
stEP159	DAM1 Δ insH::P _{LACUV5} (EcRBS) EcfabH (synRBS) BsfabI::kan	pDS57-accCBDA
stEP160	DAM1 Δ insH::P _{LACUV5} (EcRBS) EcfabH (synRBS) BsfabI::kan	pDS57-accCBDA-birA
stEP161	DAM1 Δ insH::P _{LACUV5} (EcRBS) VT8fabH1 (synRBS) VT8fabI::kan	pDS57-accCBDA
stEP162	DAM1 Δ insH::P _{LACUV5} (EcRBS) VT8fabH1 (synRBS) VT8fabI::kan	pDS57-accCBDA-birA
stEP163	DAM1 Δ insH::P _{LACUV5} (EcRBS) VT8fabH2 (synRBS) VT8fabI::kan	pDS57-accCBDA
stEP164	DAM1 Δ insH::P _{LACUV5} (EcRBS) VT8fabH2 (synRBS) VT8fabI::kan	pDS57-accCBDA-birA

[0214] Bs: 枯草芽孢杆菌; Ec: 大肠埃希氏杆菌; ADP1: 不动杆菌菌株ADP1; VT8: Marinobacter aquaeolei VT8; Ro: 不透明红球菌B4

[0215] 在32℃下,在96孔板(4NBT培养基)、摇瓶(5NBT培养基)、和发酵罐中分析DAM1ifabHI菌株。在摇瓶中,携带pDS57质粒的多种ifabHI菌株比对照DAM1 pDS57菌株工作的更好,其达到高出10至15%的FAME效价(图13)。当使用pDS57-acc-birA质粒转化ifabHI菌株时,得到另外增加的FAME效价,特别是在菌株StEP156(DAM1 IS5-11::lacUV5(ecRBS)ADP1fabH(ecRBS)ADP1fabI pDS57-accDACB-birA)中观察到FAME效价增加50%(图14)。

[0216] 在发酵罐中运行具有ifabHI的一些菌株,其中还观察到FAME效价、特定生产率和产率增加(图15),表明在这些菌株中,获得通过脂肪酸途径的更高流量,这会得到更好的产率形成速率。具体而言,stEP129(DAM1 IS5-11::lacUV5(ecRBS)ADP1fabH(ecRBS)ADP1fabI pDS57)在多个独立的发酵运行中显示更高的FAME效价和产率。fabH和fabI的其他组合可以用于取得相似的效果。尽管此处举例FAME,改变脂肪酸生物合成基因的这种方法为增加任何脂肪酸衍生物的生产的有用方法。

[0217] 在操纵子FAB 138之前插入强启动子对脂肪酸甲基酯(FAME)生产的影响

[0218] 通过T5启动子(SEQ ID NO:2)替代iFAB 138的lacUV5启动子,使得iFAB138表达更高水平,这通过mRNA分析来证明。由T5启动子使iFAB138表达得到更高的脂肪酯效价、产率和生产率。菌株shu.002(表3)与菌株BD64(表3)是同基因的,不同之处在于其包括控制iFAB138操纵子表达的T5启动子(SEQ ID NO:19)。

[0219] 表8:用于形成iT5_138盒的引物并证明其插入到新菌株中

[0220]	引物名称	SEQ ID No	序列
	DG405	20	TTGTCCATCTTTATATAATTTGGGGGTAGGGTGTTCCTTA TGTAACAAAAAACggttTAGGATGCATATGGCGGCC
	DG406	21	GATAAATCCACGAATTTTAGGTTTGATGATCATTGGTCT CCTCCTGCAGGTG CGTGTTTCGTCGTCATCGCAATTG
	DG422	22	ACTCACCGCATTGGTGTAGTAAGGCGCACC
	DG423	23	TGAATGTCATCACGCAGTTCCCAGTCATCC
	EG744	24	CCATCTTCTTTGTACAGACGTTGACTGAACATG
	EG749	24	GCACCATAGCCGTAATCCCACAGGTTATAG
	OTREE047	26	TGTCATTAATGGTTAATAATGTTGA

[0221] 引物DG405和DG406(表8)用于扩增cat-loxP和T5启动子盒,从而将50bp同源物加入到PCR产物的各个末端,这样其能够整合到任何菌株中,从而替代lacUV5启动子来调节iFAB138操纵子的表达。将cat-loxP-T5启动子转化到BD64/pKD46菌株中。在37℃过夜,在LB+氯霉素平板上回收转化子,补到新的LB+氯霉素平板上,并使用引物DG422和DG423通过菌落PCR来证明。将质粒pJW168(Palmeros et al, Gene 247:255-264(2000))转化到菌株BD64i-cat-loxP-T5_138中,并在32℃下在LB+羧苄青霉素平板上进行选择。为了除去cat标志,通过IPTG诱导cre重组酶的表达。通过使培养物在42℃下生长而除去质粒pJW168。将菌落补到LB+氯霉素和LB+羧苄青霉素上,从而分别证明pJW168的失去和cat标志的除去。此外,将菌落补到LB上作为阳性对照,所有补片的平板都在32℃下温育。使用引物DG422和DG423,通过菌落PCR证明cat标志的除去。使用引物EG744、EG749和oTREE047,通过测序来证明所得的PCR产物,该菌株称为shu.002。图16示出了iFAB138基因座:cat-loxP-P_{T5}盒整合到FAB138前面的图(图16A),和P_{T5}-iFAB 138区的图(图16B)。整合到iFAB138前面的并与整合位点具有同源性的cat-loxP-T5启动子序列示于SEQ ID NO:1中,并且与整合位点具有同源性的iT5_FAB138启动子区序列示于SEQ ID NO:2中。有多种条件可以使脂肪酸流量增加。在本实施例中,脂肪酸流量增加是通过改变操纵子iFAB138的启动子强度来实现的。由T5启动子的iFAB138表达是有利的,然而,当该启动子改变与插入yi jP::Tn5盒结合时,观察到脂肪酯和其他的脂肪酸衍生物的效价、产率和生产率的进一步改善(数据未示出)。

[0222] 实施例4

[0223] 通过修复rph和ilvG突变来增加游离脂肪酸(FFA)产率的量

[0224] 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17所示,V668/pCL_P_{trc}-tesA比EG149/pCL_P_{trc}-tesA对照生产更多的FFA。由于FFA是LS9产物的前体,较多的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。

[0225] 实施例5:通过转座子诱变使得脂肪酸衍生物的生产增加—*yijP*

[0226] 脂肪醇的生产

[0227] 为了改善大肠杆菌生产脂肪醇的效价、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*菌株中会改善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),其中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。

[0228] DG150

[0229] 5'-GCAGTTATTGGTGCCCTTAAACGCCTGGTTGCTACGCCTG-3'(SEQ ID NO:27)

[0230] DG131 5'-GAGCCAATATGCGAGAACACCCGAGAA-3'(SEQ ID NO:28)

[0231] 据测定,菌株LC535在*yijP*基因中具有转座子插入(图18),*yijP*编码保守的内在膜蛋白,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,并且还可以影响对相邻的基因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,绝对产率改善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。

[0232] LC277

[0233] 5'-CGCTGAACGTATTGCAGGCCGAGTTGCTGCACCGCTCCCGCCAGGCAG-3'(SEQ ID NO:29)

[0234] LC278

[0235] 5'-GGAATTGCCACGGTGCGGCAGGCTCCATACGCGAGGCCAGGTTATCCAACG-3'(SEQ ID NO:30)

[0236] 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:

[0237] DG407

[0238] 5'-AATCACCAGCACTAAAGTGCGGGTTCGTTACCCG-3'(SEQ ID NO:31)

[0239] DG408

[0240] 5'-ATCTGCCGTGGATTGCAGAGTCTATTTCAGCTACG-3' (SEQ ID NO:32)

[0241] 使用生产型质粒pV171.1转化具有校正yijP转座子盒的菌落,从而生产菌株D851。在摇瓶发酵中,针对不包括yijP转座子盒的菌株V940,测试D851。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9中,其标明在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")。术语"总脂肪物质"和"总脂肪酸产物"可以交换使用,其是指脂肪醇、脂肪醛和游离脂肪酸的量,其可按照国际专利申请公开W0 2008/119082中所述通过GC-FID来评价。当相同的术语用于指脂肪酯分析时,相同的术语可以用于指脂肪酯和游离脂肪酸。如本文所用,术语"脂肪酯"包括β羟基酯。

[0242] 表9:yijP转座子插入对FAS和FALC的效价和产率的影响

菌株	FAS效价	FAS产率	FALC百分率	FALC产率
V940	68g/L	18.7%	95.0%	17.8%
D851	70g/L	19.4%	96.1%	18.6%
V940	64g/L	18.4%	91.9%	16.9%
D851	67g/L	19.0%	94.0%	17.8%

[0244] 罐式发酵方法

[0245] 为了评价脂肪酸和脂肪酸衍生物在发酵罐中的生产,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葡萄糖的条件下,在pH稳态条件下运行发酵罐。

[0246] 表10:Low P FA种子培养基

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

[0248]	壮观霉素	115 mg/L
--------	------	----------

[0249] 表11:发酵罐培养基

[0250]	成分	浓度
--------	----	----

$(\text{NH}_4)_2\text{SO}_4$	0.5g/L
KH_2PO_4	3.0g/L
柠檬酸铁	0.034g/L
TM2微量矿物溶液	10mL/L
酪蛋白氨基酸	5g/L
灭菌后加入	
$\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$	2.2g/L
微量维生素溶液	1.25mL/L
葡萄糖	5g/L
接种物	50mL/L

[0251] 进一步研究表明FAS和FALC的效价和产率在具有yi jP转座子插入的菌株中有所改善是由于磷酸烯醇式丙酮酸羧化酶 (ppc) 活性的降低。在以下菌株中体外进行ppc酶测试，以评价这种假说。

[0252] 1) $\Delta \text{ppc} = \text{DG14 (LC942 } \Delta \text{ppc}::\text{cat-sacB/pLC56)}$

[0253] 2) $\text{wt-ppc} = \text{DG16 (LC942/pLC56)}$

[0254] 3) $\text{yi jP}::\text{Tn5} = \text{DG18 (LC942yi jP}::\text{Tn5-cat/pLC56)}$

[0255] 使用标准的摇瓶方法，在FA2.3培养基(如上文所述)中，在摇瓶发酵中生长的细胞中测量Ppc活性，并在诱导后的12至16小时收获。离心大约5mL细胞，并将细胞团悬浮在具有蛋白酶抑制剂混合物溶液的BugBuster Protein Extraction Reagent (Novagen)中。在振荡器上，在温和摇动的条件下将细胞悬浮物温育20min。通过在4℃下以16000x g离心20min，然后将上清液转移至新管中，从而除去不溶的细胞残渣。使用以下反应混合物，通过使用柠檬酸合酶的缀合反应来测定细胞裂解物中的Ppc活性：0.4mM乙酰基-CoA、10mM磷酸烯醇式丙酮酸盐、0.5mM一溴二胺、5mM MgCl_2 、10mM NaHCO_3 和由猪心得到的10单位的柠檬酸合酶，处于100mM Tris-HCl (pH 8.0)中。使用一溴二胺对CoA进行荧光衍生，用光度计监测使用柠檬酸合酶反应(使用草酰乙酸和乙酰基-CoA)中的CoA的形成。Ppc测试结果显示与野生型细胞相比，yi jP::Tn5-cat转座子盒将所述细胞中的Ppc活性降低2.7倍。ppc删除的细胞生长不好，并且活性比野生型细胞低大约10倍。此外，结果还表明产率最高的脂肪醇的生产需要Ppc表达水平低于野生型的水平。此外，还收集蛋白质组学的数据来评估Ppc蛋白质在具有和不具有yi jP::Tn5-cat转座子盒的2种菌株中的丰度。使用标准的脂肪醇生产条件，由生物反应器中生长的菌株V940和D851收集蛋白质样品(如上文所述)。在2个不同的时间点(32小时和48小时)取得样品，并制备该样品以用于分析。

[0256] 按照如下方式实施样品的收集和蛋白质分离：

[0257] 在各时间点，由各生物反应器收集20ml发酵液。使用冰冷的PBS淬灭样品，并通过在4℃下离心(4500rpm/10min)来收获样品。使用冰冷的PBS洗涤细胞团，并离心一次以上，然后储存在-80℃下以用于进一步加工。

[0258] 使用法式压滤方法进行总蛋白质提取。简言之，将细胞团再次悬浮于7ml冰冷的PBS中，并在2000psi下法式压滤2次，以确保细菌完全溶解。将样品在10000rpm下在4℃下离心20min，从而使未裂解的细胞和细胞残渣与蛋白质级份分离。使用BCA Protein Assay Reagent测定澄清裂解物的总蛋白质浓度。将样品稀释至2mg蛋白质/ml浓度，并冷冻在-80

℃下。

[0259] 将样品再次悬浮于合适的缓冲剂中,并在37℃下在胰蛋白酶作用下过夜,并冻干。将片段化的蛋白质样品用同位素富集的甲基哌嗪乙酸在室温下标记30min。使用阳离子交换液相色谱分离标记的样品,并使用离子阱质谱仪进行质谱分析。使用背景消除和偏畸值校正,将原始数据归一化。

[0260] 蛋白质组学数据显示,当在32小时和48小时与V940比较时,Ppc蛋白质在D851菌株中的相对丰度明显降低。在32小时,D851具有V940的Ppc水平的大约15%,而在48小时,具有V940的Ppc水平的大约35%。这些数据显示*yi jP::Tn5-cat*转座子盒使得细胞中的Ppc丰度显著降低。这表明由容留*yi jP::Tn5-cat*转座子匹配物的菌株生产脂肪醇的可见益处是由于减少了Ppc蛋白质的量。

[0261] 这些结果表明改变ppc活性可以改善脂肪酸衍生物的产率。有多种方式可以改变ppc基因的表达,*yi jP*转座子插入是完成上述改变的一种方式。不想被理论所束缚,如果降低磷酸烯醇式丙酮酸羧化酶活性的效果是限制了碳通过TCA循环的流动,则可以通过降低柠檬酸合酶(*gl tA*)的活性或者减慢TCA循环(通过降低与TCA循环有关的任一种酶的活性)来取得相似的结果。

[0262] 实施例6

[0263] 通过脂肪酸合成途径的流量增加一酰基载体蛋白质(ACP)介导的脂肪醇生产

[0264] 当由除了大肠杆菌以外的来源获得终末途径的酶在大肠杆菌(作为异源宿主)中表达、从而将脂肪酰基-ACP转化为产物时,局限可以存在于重组途径的酶对大肠杆菌脂肪酰基-ACP的识别、亲和和/或转换中。注意尽管ACP蛋白质在所有有机体中在一定程度上是保守的,但是它们的初级序列可能显著不同。为了检验该假说,将由多种蓝细菌得到的*acp*基因克隆到在pLS9-185(其为pCL1920衍生物)中存在的细长聚球藻PCC7942酰基-ACP还原酶(AAR)的下游。此外,由枯草芽孢杆菌得到的*sfp*基因(编号No.X63158;SEQ ID NO:53)编码具有广泛的底物特异性的磷酸泛酰巯基乙胺基转移酶,将其克隆到各*acp*基因的下游。其中所述的酶与非活性的apo-ACP向活性全-ACP的转化有关。所构建的质粒在表12中描述。

[0265] 表12:在细长聚球藻PCC7942AAR下游在具有和不具有枯草芽孢杆菌*sfp*下共表达蓝细菌ACP的质粒

	基础质粒	ACP 来源	ACP SEQ ID NO.(DNA/多肽)	不具有 <i>sfp</i>	具有 <i>sfp</i>
[0266]	pLS9-185	细长聚球藻 7942	49/50	pDS168	pDS168S
	pLS9-185	集胞藻物种 6803	45/46	pDS169	未利用
	pLS9-185	原 绿 球 藻 (<i>Prochlorococcus</i>	47/48	pDS170	pDS170S
[0267]		<i>marinus</i>)MED4			
	pLS9-185	念 球 藻 (<i>Nostoc punctiforme</i>)73102	43/44	pDS171	pDS171S
	pLS9-185	念珠藻属物种 7120	51/52	pDS172	pDS172S

[0268] 使用InFusion技术(Clontech Laboratories, Inc., Mountain View, CA),将所有

的 acp 基因与合成的RBS一起克隆到pLS9-185中 aar 基因下游紧邻的EcoRI位点中。在 acp 基因的下游重建EcoRI位点。相似地,使用InFusion技术将枯草芽孢杆菌 sfp 基因与合成的RBS一起克隆到所述的EcoRI位点中。所有质粒均转化至大肠杆菌MG1655DV2(表3)中。这些试验的对照单独表达AAR(pLS9-185)。由标准的摇瓶发酵试验的结果示于图19中。在包括质粒pDS171S、pDS172S、pDS168和pDS169的菌株中,观察到脂肪醇效价的显著升高,证明ACP过表达对于脂肪醇生产是有利的,推测这种情况是由于有助于酰基-ACP被异源终末途径的酶所识别、亲和和/或转换(关于ACP的来源以及 sfp 的存在或缺乏参见表12)。

[0269] 脂肪酸的生产

[0270] 为了评价ACP的过表达是否还可以增加游离脂肪酸的生产,由pDS171扩增具有 sfp 的一种蓝细菌ACP基因(表12),并克隆到 $tesA$ 的下游而进入pCL载体中。所得的操纵子处于Ptrc3启动子的控制之下,该启动子提供比Ptrc野生型启动子更低的转录水平。构建体被克隆到大肠杆菌DV2中,并针对脂肪酸的生产进行评价。对照菌株包括相同的质粒,但不具有蓝细菌ACP和枯草芽孢杆菌 sfp 。由标准微量滴定板发酵试验得到的结果示于图20中。在共表达异源ACP的菌株中观察到脂肪酸效价显著改善,证明ACP共表达对于脂肪酸生产可以是有利的,推测这种情况是由于增加了通过脂肪酸生物合成途径的流量。

序列表

<110> LS9, INC.

<120> 改善的脂肪酸衍生物的生产

<130> LS00042PCT

<140>

<141>

<150> 61/619,324

<151> 2012-04-02

<160> 64

<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"

<400> 1

ttgtccatct ttatataatt tgggggtagg gtgttcttta tgtaaaaaaa acgttttagg	60
atgcatatgg cgcccgata acttcgtata gcatacatta tacgaagtta tctagagttg	120
catgcctgca ggtccgctta ttatcactta ttcaggcgta gcaaccaggc gtttaagggc	180
accaataact gccttaaaaa aattacgccc cgccctgcc ctcacgcag tactgttgta	240
attcattaag cattctgccg acatggaagc catcacaac ggcatgatga acctgaatcg	300
ccagcgcat cagcaccttg tcgccttgcg tataatatTTT gcccatggtg aaaacggggg	360
cgaagaagtt gtccatattg gccacgttta aatcaaaact ggtgaaactc acccagggat	420
tggttgagac gaaaaacata ttctcaataa acccttttagg gaaataggcc aggttttcac	480
cgtaacacgc cacatcttgc gaatatatgt gtagaaactg ccggaaatcg tcgtggtatt	540
cactccagag cgatgaaac gtttcagttt gtcacatgaa aacggtgtaa caagggtgaa	600
cactatccca tatcaccagc tcaccgtctt tcattgccat acggaattcc ggatgagcat	660
tcacagcgcg ggcaagaatg tgaataaagg ccggataaaa cttgtgctta tttttcttta	720
cggctcttta aaaggccgta atatccagct gaacggtctg gttataggta cattgagcaa	780
ctgactgaaa tgccctcaaaa tgttctttac gatgccattg ggatatatca acggtggtat	840
atccagtgat tttttctcc atttttagctt cttagctcc tgaaaatctc gataactcaa	900
aaaaacgccc cggtagtgat cttatttcat tatggtgaaa gttggaacct cttacgtgcc	960
gatcaacgtc tcatTTTtgc caaaagtTgg ccagggtctt cccggtatca acagggacac	1020
caggatttat ttattctgcg aagtgatctt ccgtcacagg tatttattcg actctagata	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

[0001]

<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 gtgttcttta tgtaaaaaaa acgttttagg 60
 atgcatatgg cggccgcata acttcgtata gcatacatta tacgaagtta tggatccagc 120
 ttatcgatac cgtcaaacia atcataaaaa atttatttgc tttcaggaaa atttttctgt 180
 ataatagatt caattgcgat gacgacgaac acgcacctgc 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
 atgtcgatga taaaaagcta tgccgcaaaa gaagcgggcg gcgaactgga agtttatgag 60
 tacgatcccc gtgagctgag gccacaagat gttgaagtgc aggtggatta ctgcgggcatc 120
 tgccattccg atctgtcgat gatcgataac gaatggggat ttccacaata tccgctggtt 180
 gccgggcatg aggtgattgg gcgcgtggtg gactcggga gcgccgcgca ggataaaggt 240
 ttgcaggtcg gtcagcgtgt cgggattggc tggacggcgc gtagctgtgg tactgcgac 300
 gcctgtatta gcggtaatca gatcaactgc gagcaaggtg cggtgccgac gattatgaat 360
 cgcggtggct ttgccgagaa gttgcgtgcg gactggcaat gggtgattcc actgccagaa 420
 aatattgata tcgagtcgag cgggcccgtg ttgtgcggcg gtatcacggt ctttaaacca 480
 ctgttgatgc accatatcac tgctaccagc cgcgttgggg taattggtat tggcgggctg 540
 gggcatatcg ctataaaact tctgcacgca atgggatgag aggtgacagc ctttagttct 600
 aatccggcga aagagcagga agtgctggcg atgggtgccg ataaagtgtt gaatagccgc 660
 gatccgcagg cactgaaagc actggcgggg cagtttgatc tcattatcaa caccgtcaac 720
 gtcagctcg actggcagcc ctattttgag gcgctgacct atggcggtaa tttccatacg 780
 gtcggtgcgg ttctcacgcc gctgtctgtt ccggccttta cgttaattgc gggcgatcgc 840

```

agcgtctctg gttctgctac cggcacgcct tatgagctgc gtaagctgat gcgttttgcc      900
gcccgcagca aggttgcgcc gaccaccgaa ctgttcccga tgcgaaaaat taacgacgcc      960
atccagcatg tgcgcgacgg taaggcgcgt taccgcgtgg tgttgaaagc cgatttttga    1020

```

<210> 6
 <211> 473
 <212> PRT
 <213> 除烃海杆菌

```

<400> 6
Met Lys Arg Leu Gly Thr Leu Asp Ala Ser Trp Leu Ala Val Glu Ser
1           5           10           15

```

```

Glu Asp Thr Pro Met His Val Gly Thr Leu Gln Ile Phe Ser Leu Pro
                20           25           30

```

```

Glu Gly Ala Pro Glu Thr Phe Leu Arg Asp Met Val Thr Arg Met Lys
          35           40           45

```

```

Glu Ala Gly Asp Val Ala Pro Pro Trp Gly Tyr Lys Leu Ala Trp Ser
          50           55           60

```

```

Gly Phe Leu Gly Arg Val Ile Ala Pro Ala Trp Lys Val Asp Lys Asp
65           70           75           80

```

```

Ile Asp Leu Asp Tyr His Val Arg His Ser Ala Leu Pro Arg Pro Gly
          85           90           95

```

```

[0005] Gly Glu Arg Glu Leu Gly Ile Leu Val Ser Arg Leu His Ser Asn Pro
          100           105           110

```

```

Leu Asp Phe Ser Arg Pro Leu Trp Glu Cys His Val Ile Glu Gly Leu
          115           120           125

```

```

Glu Asn Asn Arg Phe Ala Leu Tyr Thr Lys Met His His Ser Met Ile
          130           135           140

```

```

Asp Gly Ile Ser Gly Val Arg Leu Met Gln Arg Val Leu Thr Thr Asp
145           150           155           160

```

```

Pro Glu Arg Cys Asn Met Pro Pro Pro Trp Thr Val Arg Pro His Gln
          165           170           175

```

```

Arg Arg Gly Ala Lys Thr Asp Lys Glu Ala Ser Val Pro Ala Ala Val
          180           185           190

```

```

Ser Gln Ala Met Asp Ala Leu Lys Leu Gln Ala Asp Met Ala Pro Arg
          195           200           205

```

```

Leu Trp Gln Ala Gly Asn Arg Leu Val His Ser Val Arg His Pro Glu
          210           215           220

```

```

Asp Gly Leu Thr Ala Pro Phe Thr Gly Pro Val Ser Val Leu Asn His
225           230           235           240

```

```

Arg Val Thr Ala Gln Arg Arg Phe Ala Thr Gln His Tyr Gln Leu Asp
          245           250           255

```

Arg Leu Lys Asn Leu Ala His Ala Ser Gly Gly Ser Leu Asn Asp Ile
 260 265 270
 Val Leu Tyr Leu Cys Gly Thr Ala Leu Arg Arg Phe Leu Ala Glu Gln
 275 280 285
 Asn Asn Leu Pro Asp Thr Pro Leu Thr Ala Gly Ile Pro Val Asn Ile
 290 295 300
 Arg Pro Ala Asp Asp Glu Gly Thr Gly Thr Gln Ile Ser Phe Met Ile
 305 310 315 320
 Ala Ser Leu Ala Thr Asp Glu Ala Asp Pro Leu Asn Arg Leu Gln Gln
 325 330 335
 Ile Lys Thr Ser Thr Arg Arg Ala Lys Glu His Leu Gln Lys Leu Pro
 340 345 350
 Lys Ser Ala Leu Thr Gln Tyr Thr Met Leu Leu Met Ser Pro Tyr Ile
 355 360 365
 Leu Gln Leu Met Ser Gly Leu Gly Gly Arg Met Arg Pro Val Phe Asn
 370 375 380
 Val Thr Ile Ser Asn Val Pro Gly Pro Glu Gly Thr Leu Tyr Tyr Glu
 385 390 395 400
 Gly Ala Arg Leu Glu Ala Met Tyr Pro Val Ser Leu Ile Ala His Gly
 405 410 415
 Gly Ala Leu Asn Ile Thr Cys Leu Ser Tyr Ala Gly Ser Leu Asn Phe
 420 425 430
 Gly Phe Thr Gly Cys Arg Asp Thr Leu Pro Ser Met Gln Lys Leu Ala
 435 440 445
 Val Tyr Thr Gly Glu Ala Leu Asp Glu Leu Glu Ser Leu Ile Leu Pro
 450 455 460
 Pro Lys Lys Arg Ala Arg Thr Arg Lys
 465 470
 <210> 7
 <211> 455
 <212> PRT
 <213> 除烃海杆菌
 <400> 7
 Met Thr Pro Leu Asn Pro Thr Asp Gln Leu Phe Leu Trp Leu Glu Lys
 1 5 10 15
 Arg Gln Gln Pro Met His Val Gly Gly Leu Gln Leu Phe Ser Phe Pro
 20 25 30
 Glu Gly Ala Pro Asp Asp Tyr Val Ala Gln Leu Ala Asp Gln Leu Arg
 35 40 45
 Gln Lys Thr Glu Val Thr Ala Pro Phe Asn Gln Arg Leu Ser Tyr Arg

[0006]

[0007]

50	55	60
Leu Gly Gln Pro Val	Trp Val Glu Asp Glu His	Leu Asp Leu Glu His
65	70	75
His Phe Arg Phe Glu	Ala Leu Pro Thr Pro Gly Arg Ile Arg	Glu Leu
	85	90
Leu Ser Phe Val Ser	Ala Glu His Ser His	Leu Met Asp Arg Glu Arg
	100	105
Pro Met Trp Glu Val	His Leu Ile Glu Gly Leu Lys	Asp Arg Gln Phe
	115	120
Ala Leu Tyr Thr Lys	Val His His Ser Leu Val	Asp Gly Val Ser Ala
	130	135
Met Arg Met Ala Thr	Arg Met Leu Ser Glu Asn	Pro Asp Glu His Gly
	145	150
Met Pro Pro Ile Trp	Asp Leu Pro Cys Leu Ser	Arg Asp Arg Gly Glu
	165	170
Ser Asp Gly His Ser	Leu Trp Arg Ser Val Thr His	Leu Leu Gly Leu
	180	185
Ser Asp Arg Gln Leu	Gly Thr Ile Pro Thr Val Ala	Lys Glu Leu Leu
	195	200
Lys Thr Ile Asn Gln	Ala Arg Lys Asp Pro Ala Tyr	Asp Ser Ile Phe
	210	215
His Ala Pro Arg Cys	Met Leu Asn Gln Lys Ile	Thr Gly Ser Arg Arg
	225	230
Phe Ala Ala Gln Ser	Trp Cys Leu Lys Arg Ile	Arg Ala Val Cys Glu
	245	250
Ala Tyr Gly Thr Thr	Val Asn Asp Val Val Thr	Ala Met Cys Ala Ala
	260	265
Ala Leu Arg Thr Tyr	Leu Met Asn Gln Asp Ala	Leu Pro Glu Lys Pro
	275	280
Leu Val Ala Phe Val	Pro Val Ser Leu Arg Arg	Asp Asp Ser Ser Gly
	290	295
Gly Asn Gln Val Gly	Val Ile Leu Ala Ser Leu	His Thr Asp Val Gln
	305	310
Asp Ala Gly Glu Arg	Leu Leu Lys Ile His His	Gly Met Glu Glu Ala
	325	330
Lys Gln Arg Tyr Arg	His Met Ser Pro Glu Glu	Ile Val Asn Tyr Thr
	340	345
Ala Leu Thr Leu Ala	Pro Ala Ala Phe His	Leu Leu Thr Gly Leu Ala

	355		360		365
	Pro Lys Trp Gln Thr Phe Asn Val Val Ile Ser Asn Val Pro Gly Pro				
	370		375		380
	Ser Arg Pro Leu Tyr Trp Asn Gly Ala Lys Leu Glu Gly Met Tyr Pro				
	385		390		395 400
	Val Ser Ile Asp Met Asp Arg Leu Ala Leu Asn Met Thr Leu Thr Ser				
		405		410	415
	Tyr Asn Asp Gln Val Glu Phe Gly Leu Ile Gly Cys Arg Arg Thr Leu				
		420		425	430
	Pro Ser Leu Gln Arg Met Leu Asp Tyr Leu Glu Gln Gly Leu Ala Glu				
		435		440	445
	Leu Glu Leu Asn Ala Gly Leu				
	450		455		
	<210> 8				
	<211> 457				
	<212> PRT				
	<213> 泊库岛食烷菌				
	<400> 8				
	Met Lys Ala Leu Ser Pro Val Asp Gln Leu Phe Leu Trp Leu Glu Lys				
	1		5		10 15
[0008]	Arg Gln Gln Pro Met His Val Gly Gly Leu Gln Leu Phe Ser Phe Pro				
		20		25	30
	Glu Gly Ala Gly Pro Lys Tyr Val Ser Glu Leu Ala Gln Gln Met Arg				
		35		40	45
	Asp Tyr Cys His Pro Val Ala Pro Phe Asn Gln Arg Leu Thr Arg Arg				
		50		55	60
	Leu Gly Gln Tyr Tyr Trp Thr Arg Asp Lys Gln Phe Asp Ile Asp His				
	65		70		75 80
	His Phe Arg His Glu Ala Leu Pro Lys Pro Gly Arg Ile Arg Glu Leu				
		85		90	95
	Leu Ser Leu Val Ser Ala Glu His Ser Asn Leu Leu Asp Arg Glu Arg				
		100		105	110
	Pro Met Trp Glu Ala His Leu Ile Glu Gly Ile Arg Gly Arg Gln Phe				
		115		120	125
	Ala Leu Tyr Tyr Lys Ile His His Ser Val Met Asp Gly Ile Ser Ala				
		130		135	140
	Met Arg Ile Ala Ser Lys Thr Leu Ser Thr Asp Pro Ser Glu Arg Glu				
	145		150		155 160
	Met Ala Pro Ala Trp Ala Phe Asn Thr Lys Lys Arg Ser Arg Ser Leu				
		165		170	175

[0009]

Pro Ser Asn Pro Val Asp Met Ala Ser Ser Met Ala Arg Leu Thr Ala
 180 185 190

Ser Ile Ser Lys Gln Ala Ala Thr Val Pro Gly Leu Ala Arg Glu Val
 195 200 205

Tyr Lys Val Thr Gln Lys Ala Lys Lys Asp Glu Asn Tyr Val Ser Ile
 210 215 220

Phe Gln Ala Pro Asp Thr Ile Leu Asn Asn Thr Ile Thr Gly Ser Arg
 225 230 235 240

Arg Phe Ala Ala Gln Ser Phe Pro Leu Pro Arg Leu Lys Val Ile Ala
 245 250 255

Lys Ala Tyr Asn Cys Thr Ile Asn Thr Val Val Leu Ser Met Cys Gly
 260 265 270

His Ala Leu Arg Glu Tyr Leu Ile Ser Gln His Ala Leu Pro Asp Glu
 275 280 285

Pro Leu Ile Ala Met Val Pro Met Ser Leu Arg Gln Asp Asp Ser Thr
 290 295 300

Gly Gly Asn Gln Ile Gly Met Ile Leu Ala Asn Leu Gly Thr His Ile
 305 310 315 320

Cys Asp Pro Ala Asn Arg Leu Arg Val Ile His Asp Ser Val Glu Glu
 325 330 335

Ala Lys Ser Arg Phe Ser Gln Met Ser Pro Glu Glu Ile Leu Asn Phe
 340 345 350

Thr Ala Leu Thr Met Ala Pro Thr Gly Leu Asn Leu Leu Thr Gly Leu
 355 360 365

Ala Pro Lys Trp Arg Ala Phe Asn Val Val Ile Ser Asn Ile Pro Gly
 370 375 380

Pro Lys Glu Pro Leu Tyr Trp Asn Gly Ala Gln Leu Gln Gly Val Tyr
 385 390 395 400

Pro Val Ser Ile Ala Leu Asp Arg Ile Ala Leu Asn Ile Thr Leu Thr
 405 410 415

Ser Tyr Val Asp Gln Met Glu Phe Gly Leu Ile Ala Cys Arg Arg Thr
 420 425 430

Leu Pro Ser Met Gln Arg Leu Leu Asp Tyr Leu Glu Gln Ser Ile Arg
 435 440 445

Glu Leu Glu Ile Gly Ala Gly Ile Lys
 450 455

<210> 9
 <211> 70
 <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> 9	
	aaaaacagca acaatgtgag ctttgttgta attatattgt aaacatattg attccgggga	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 gttattcatt tacgcggctt 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"	
[0010]	<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	
	cgcgcgcttg accggcagcc tgg	23
	<210> 13	
	<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> 13	
	ccttggcatt ggcaatttga gaattcgagg aggaaaacta aatgaccatt tcctcacctt	60
	<210> 14	
	<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> 14	

	ttttgttcgg gcccaagctt ttattgcaaa cgcagatgcg tgatttcacc cgcattcagc	60
	<210> 15 <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> 15 cgggcccaag cttcgaattc ttattgcaaa cgcagatgcg tgatttcacc cgcattcagc	60
	<210> 16 <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> 16 gaatagcgcc gtcgacgagg aggaaaaacta aatgaccatt tcctcacctt tgattgacgt	60
	<210> 17 <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"	
[0011]	<400> 17 tgatgatgat gatggtcgac ttattgcaaa cgcagatgcg tgatttcacc cgcattcagc	60
	<210> 18 <211> 4250 <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> 18 atgacattt cctcacctt gattgacgtc gccaaccttc cagacatcaa caccactgcc	60
	ggcaagatcg ccgaccttaa ggctcgccgc gcggaagccc atttcccat gggtgaaaag	120
	gcagtagaga aggtccacgc tgcctggacgc ctactgccc gtgagcgctt ggattactta	180
	ctcgatgagg gctccttcat cgagaccgat cagctggctc gccaccgcac caccgctttc	240
	ggcctgggcg ctaagcgtec tgcaaccgac ggcacgtga ccgctgggg caccattgat	300
	ggacgcgaag tctgcatctt ctgcaggac ggcaccgtat tcggtggcgc gcttggtgag	360
	gtgtacggcg aaaagatgat caagatcatg gagctggcaa tcgacaccgg ccgccattg	420
	atcggtcttt acgaaggcgc tggcgctcgt attcaggacg gcgctgtctc cctggacttc	480
	atttccaga ccttctacca aaacattcag gcttctggcg ttatccaca gatctccgtc	540
	atcatgggag catgtgcagg tggcaacgt tacggcccag ctctgaccga cttcgtggtc	600
	atggtggaca agacctcaa gatgttcgtt accggcccag acgtgatcaa gaccgtcacc	660
	ggcgaggaaa tcaccagga agagcttggc ggagcaacca cccacatggt gaccgctggt	720

	aactcccact acaccgctgc gaccgatgag gaagcactgg attgggtaca ggacctggtg	780
	tccttcctec catccaacaa tcgctcctac gcaccgatgg aagacttcga cgaggaagaa	840
	ggcggcggtg aagaaaacat caccgctgac gatctgaagc tcgacgagat catcccagat	900
	tcgcgacceg ttcttiacga cgtccgcat gtcacgaat gcctcaccga cgatggcgaa	960
	tacctggaaa tccaggcaga ccgcgcagaa aacgttgtaa ttgcattcgg ccgcatcgaa	1020
	ggccagtcgg ttggctttgt tgccaaccag ccaaccagc tcgctggctg cctggacatc	1080
	gactcctctg agaaggcagc tcgcttcgtc cgcacctgcg acgcgttcaa catccaatc	1140
	gtcatgcttg tcgacgtccc cggttcctc ccaggcgag gccaggagta cgggtgcatt	1200
	ctgcgtcgtg gcgcaaagct gctctacga tacggcgaag caaccgttc aaagatcacc	1260
	gtcacatgc gtaaggctta cggcggagcg tactgcgtga tgggttcaa gggcttgggc	1320
	tttgacatca accttgcatg gccaaaccga cagatcgccg tcatggcgcg tctggcgca	1380
	gttgattca tctaccgcaa ggagctcatg gcagctgatg ccaaggcct cgataccgta	1440
	gctctggcta agtccttcga gcgcgagtat gaagaccaca tgcacaacc gtaccacgt	1500
	gcagaacgtg gcctgatcga cgcctgatc ctgccaagcg aaaccgcgg acagatttc	1560
	cgcaaccttc gcctgtcaa gcacaagaac gtcactgcc ctgctcgaa gcacggcaac	1620
	atgccactgt aaggaggaaa actaatgtc agtcgagact cgcaagatca ccaagttct	1680
	tgtcgtaac cgtggtgaga ttgcaatccg cgtgtccgt gcagctcgag atgaaggcat	1740
	cggatctgtc gccgtctacg cagagccaga tcgagatgca ccattcgtgt catatgcaga	1800
[0012]	cgaggctttt gccctcggtg gccaaacatc cgtgagtc taccttgta ttgacaagat	1860
	catcgatcgc gcccgcaagt ccggcgccga cgcctccac cccggctacg gcttcctcgc	1920
	agaaaacgt gacttcgcag aagcagtcac caacgaaggc ctgatctgga ttggaccttc	1980
	acctgagtc atccgtccc tcggcgacaa ggtcacgct cgcacatcg cagataccgc	2040
	caaggctcca atggctcctg gcaccaagga accagtaaaa gacgcagcag aagtgtggc	2100
	tttcgtgaa gaattcggc tcccaatcg catcaaggca gcttcggtg gcggcgacg	2160
	tgcatgaag gttgectaca agatgaaga agtcgtgac ctcttcgagt ccgcaaccg	2220
	tgaagcaacc gcagcgttcg gccgcggcga gtgcttcgt gagcgtacc tggacaaggc	2280
	acgccaggtt gaggtcagg tcacgcgca taagcacggc aacgttggtg tcgccggaac	2340
	ccgtgactgc tccctgcagc gccgtttcca gaagctcgt gaagaagcac cagcaccatt	2400
	cctcaccgat gaccagcgcg agcgttcca ctctccgcg aaggctatct gtaaggaagc	2460
	tggtactac ggtgcaggca ccgttgagta cctcgttgcc tcgacggcc tgatctcctt	2520
	cctcagggtc aacaccgcc tccagtgga acaccagtc accgaagaga ccaccgcat	2580
	cgacctggtc cgcgaaatgt tccgcatcg agaaggccac gagcttcca tcaaggaaga	2640
	tccagctcca cgcggccac cattcgagtt ccgcatcaac ggcaagacg ctggtccaa	2700
	cttcagcct gcaccaggca agatcaccag ctaccgcgag ccacaggcc caggcgtccg	2760
	catggactcc ggtgtcgtt aaggttccga aatctccgga cagttcgact ccatgctggc	2820
	aaagctgac gtttggggcg acaccgcga gcaggctctc cagcgtccc gccgtgcact	2880
	tcgaggtac gttgtcagg gcatccaac cgttatccca ttccaccagc acatcgtgga	2940
	aaaccagca ttctgggca acgacgaagg ctctgagatc tacaccaagt gcatcgaaga	3000

	ggtttgggat aacccaatcg caccttacgt tgacgttcc gagctcgacg aagatgagga	3060
	caagacccca gcacagaagg ttgtttgga gatcaacggc cgtcgcgttg aggttgact	3120
	cccaggcgat ctggcactcg gtggcaccgc tggctctaag aagaaggcca agaagcgtcg	3180
	cgcaggtggt gcaaaggctg gcgtatccgg cgatgcagtg gcagctccaa tgcagggcac	3240
	tgtcatcaag gtcaacgtcg aagaaggcgc tgaagtcaac gaaggcgaca ccgttgtgt	3300
	cctcagggt atgaagatgg aaaacctgt gaaggctcat aagtccggaa ccgtaaccgg	3360
	ccttactgtc gctgcaggcg aggggtgtcaa caaggcggtt gttctcctcg agatcaagta	3420
	atctagagga ggaaaactaa atgaatgttg acattagccg ctctcgtgaa ccgttgaacg	3480
	tggaaactgt gaaagaaaaa ctgctgcaga acggtgattt cgggtcaagt atctacgaga	3540
	aggtcaccgg ctctaccaat gcggacctgc tggctctggc gggcagcggc gctccaaact	3600
	ggaccgtcaa gactgttgaa tttcaggacc acgcccgtgg ccgtctgggt cgtccgtgga	3660
	gcgcaccgga gggttcccaa accatcgtca gcgttctggt ccaactgagc attgatcagg	3720
	tggaccgtat tggtagatc ccgtggccg caggcttgcc tgttatggat gcgctgaatg	3780
	atctggcggt ggagggtgca ggcctgaaat ggccgaacga tgttcagatc cacgtaaga	3840
	agttgtcggg tattctggtt gaagcaaccg gcttcgactc cactccgacc gtggttatcg	3900
	gtttgggtac gaatactcgt ttgacgaaag aagagctgcc ggtcccgac gcgaccagcc	3960
	tggccctgga gggtgttgaa gttgaccgta cgacgttctt gattaacatg ctgaccatc	4020
	tgcatacccg tctggatcag tggcagggtc cgtctgtgga ctggctggat gactatcgcg	4080
	cggttttagt cagcattggc caagatgtgc gtgtcctgct gcctgggtgac aaagagctgc	4140
[0013]	tgggcgaggc gattggcgtg gcgaccggtg gtgagatccg tgtgcgcgac gccagcggca	4200
	cggtcacac gctgaatcgc ggtgaaatca cgcactctcg tttgcaataa	4250
	<210> 19	
	<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> 19	
	atgatcatca aacctaaaat tcgtggattt atctgtacaa caacgcaccc agtgggttgt	60
	gaagcgaacg taaaagaaca aattgcctac acaaaagcac aaggtccgat caaaaacgca	120
	cctaagcgcg tgttggttgt cggatcgtct agcggctatg gtctgtcatc acgcatcgt	180
	gcggcgtttg gcggtggtgc ggcgacgac ggcgatattt tcgaaaagcc gggcactgac	240
	aaaaaaccag gtactcggg tttctacaat gcagcagcgt ttgacaagct agcgcataaa	300
	gcgggcttgt acgcaaaaag cctgaacggc gatgcgttct cgaacgaagc gaagcaaaaa	360
	gcgattgagc tgattaagca agacctcggc cagattgatt tgggtggtta ctcgttggt	420
	tctccagtgc gtaagatgcc agacacgggt gagctagtgc gctctgact aaaaccgatc	480
	ggcgaacgt acacctctac cgcggtagat accaataaag atgtgatcat tgaagccagt	540
	gttgaacctg cgaccgagca agaaatcgt gacactgtca ccgtgatggg cgggtcaagat	600
	tgggaactgt ggtaccaagc actggaagag gcgggtgttc ttgctgaagg ttgcaaaacc	660
	gtggcgatca gctacatcgg tactgaattg acttggccaa tttactggga tggcgcttta	720

[0014]

```

ggccgtgcc agatggacct agatcgcga ggcagcgc tgaacgaaa gctggcagcg 780
aaaggtggta ccgcgaacgt tgcagttttg aaatcagtg tgactcaagc aagctctgcg 840
attcctgtga tgcgcctcta catcgcaatg gtgttcaaga agatcgtga acagggcgtg 900
catgaagget gtatggagca gatctaccgc atgttcagtc aacgtctgta caaagaagat 960
ggttcagcgc cggaagtga tgatcacaat cgtctgcgtt tggatgactg ggaactgcgt 1020
gatgacattc agcagcactg ccgtgatctg tggccacaaa tcaccacaga gaacctgcgt 1080
gagctgaccg attacgacat gtacaaagaa gatttcacat agctgtttgg ctttggcatt 1140
gaaggcattg attacgatgc tgacgtcaat ccagaagtcg aatttgatgt gattgatatc 1200
gagtaattta gtgactgagc gtacatgtat acgaagatta ttgtactgg cagctatctg 1260
ccgaacaag tgcggactaa cgcgatctg gaaaaatgg ttgagacctc tgacgagtgg 1320
attgtcactc gtacaggtat tcgtaaacgc catatcgccg cgccgaatga aactgtcgcg 1380
acgatgggct ttaccgctgc gaatcgcgcg attgagatgg cggggatcga taaagaccaa 1440
attggttga ttgtgtggc taccacatca gcaacgcatg catttccaag cgcggcatgt 1500
cagattcaaa gtatgctcgg tattaaggt tgcgcgcgt ttgatgtgc ggcagcgtgc 1560
gcaggtttca cctacgcgtt aagcatcgcc gaccagtacg ttaaaccgg cgcggttaaa 1620
cacgcgtgg tggcgttc cgatgtatta gcccgcactt gcgatcctgg cgatcgcgtt 1680
acgatcatta tttcggcga tggcgcaggc gcggccgtac tgagcgcttc tgaagaaccg 1740
ggtattatct cactcatct tcatgccgat ggccgttacg gtgaattact gaccctgccg 1800
aatgccgatc gcgtaaatcc ggataaccgc atttacctga caatggcggg caatgaagtc 1860
tttaaagtgg cggctactga actggcgcgt attgtcatg agacgctggc ggctaataac 1920
ctggatcgtc cagaactcga ttggctgggt ccgcacatcagg ctaacctgcg tatcattagc 1980
gcgacagcga aaaaactcgg catgtcatg gacaatgtcg tcgtcacgct ggacaggcac 2040
ggcaataacct ccgcggttc tgtgccgtgc gcgttgatg aagccgtgcg tgacgggcga 2100
attaaagccg gtcagctggt attgcttgaa gccttcgggg gtggattcac ctggggctcc 2160
gcgtgatc gtttctagta taaggattta aacatgacgc aatttgcat tgtgttcccc 2220
gttcagggtt ctgagcgt tgggatgtg gccgagatgg cggcaaatta ccctatcgta 2280
gaagaacgt ttgtgaagc ttctgcggct ctgggatatg atctgtgggc gtcaccag 2340
caaggtccag cggaagaact gaataaaacc tggcagacgc agccggcgtt attaacgct 2400
tcctgcgcg tttggcgcgt ttggcagcag caggcggtta aaatgcctgc gttaatggca 2460
ggtcacagcc tgggcgaata ttccgcgtg gtttgcgtg gcgtcatcaa ctttgcgtg 2520
gccgttcgtc tgggtgaaat gcgcggtaaa ttcatgcagg aagcgttcc ggaaggcact 2580
ggcgcatgt ctgcgatcat cgggctggat gatgcctcta ttgctaaagc ctgtgaagaa 2640
tctgccgaag ggcaggttgt ttgcgcgtt aactttaact cgcgggaca ggtggttacc 2700
gccgggcata aagaggcgtt agaacgtgc ggcgcagcct gtaaagcgc tggcgcgaaa 2760
cgcgctgc cgctgccgt gagcgtaccg tcgactgcg cgctgatgaa accagcgga 2820
gataagctgg cgttgaatt agccaaaatt accttagcg cgccaacgtt gccggtagt 2880
aacaacgtt acgtgaaat tgaaccgat gccgccgcta tccgcgatg gctggttcgc 2940
cagttgtaca atccgtaca gtggacgaag agcgtggaat ttatcgcggc gcaggcggtt 3000

```

[0015]

```

gaacatcttt atgaagtggg tccaggtaaa gtctcactg gtctgacgaa acgtattgtc 3060
gacaccctga cagcgctggc gctgaacgag ccggcggcgc tgtctgcggc acttacgcaa 3120
taaaagagga aaacatgag ctttgaagga aagattgcgc tggtagctgg tgcaagccgt 3180
ggcataggcc gcgcaattgc agagactctc gttgcccgcg gcgcgaaagt tatcgggact 3240
gcgaccagtg aaaaatgtgc gaagaacatt agcgactatt tagtgctaa cgggaaaggt 3300
ttgatgtga atgtgaccga tctgcatct attgaatctg ttctggaaaa tattcgcgca 3360
gaatttggtg aagtggatat cctggttaat aatgccggtg tcactcgtga taatctgttg 3420
atgcgaatga aagatgatga gtggaacgat attatcgaac ccaacttata atccgttttc 3480
cgctgtcaa aagcggtaat gcgcgctatg atgaaaaagc gttgtggtcg cattatcact 3540
attggttctg tggttgtgac catgggaaat gcaggtcagg caaactacgc tgcggcgaaa 3600
gcgggcctga tcggtttcag taaatcactg gcgcgtgaag ttgcgtcccg tggattact 3660
gtcaatgttg tggctccggg tttattgaa acggacatga cgcgtgcgt gtctgacgat 3720
cagcgtcggg gtatcctggc gcaggtgcct gcgggtcgcc tcggcggcgc tcaggaaatc 3780
gccagtggg ttgcattttt agcctctgac gaagcgagtt acatcactgg tgagactctg 3840
cacgtcaacg gcggaatgta catggtttaa ttttaaggtt tacataaaac atggtagata 3900
aacgcgaatc ctatacaaaa gaagaccttc ttgcctctgg tcgtggtgaa ctgtttggcg 3960
ctaaagggcc gcaactccct gcaccgaaca tgctgatgat ggaccgcgtc gttaatgatga 4020
ccgaaacggg cggcaatttc gacaaaggct atgtcgaagc cgagctggat atcaatccgg 4080
atctatggtt ctteggatgc cactttatcg gcgatccggt gatgcccggt tgtctgggtc 4140
tggatgctat gtggcaattg gtgggattct acctgggctg gttggcgcg gcgaaggcaaag 4200
gccgcgtctt gggcgtgggc gaagtgaat ttaccggcca ggttctgccg acagccagga 4260
aagtcaccta tcgtattcat ttcaaacgta tcgtaaacgc tcgcctgac atgggcctgg 4320
cggacggtga ggttctggtg gatggtcgcc tgatctatac cgcacacgat ttgaaagtcg 4380
gtttgttcca ggatacttcc gcgttctaaa aggaggcaac aaaaatgaat gccgcgttgt 4440
cattacgggt attggtgcag tgacgccggt ggtaacaac gctgatagt tctggtgcag 4500
catcaaagag ggtaaatgtg gcattgacaa gatcaaagcg ttgacgcaa ccgatttcaa 4560
agttaagctg gctgccgaag tgaaggactt caccggag gactttatcg acaagcgtga 4620
ggcgaaccgt atggaccgtt ttagccagtt tgcgatcgtt gcggcggtg aggcaatcaa 4680
ggacagcaaa ctggacctgg agtcgattga taagaatcgt ttcggcgta ttgttgtag 4740
cggcattggc ggcatcgga ccatgagaa gcaggtgaa aagctgatta ccaaaggtcc 4800
gggtcgtgtg agccctatga ctattccgat gatcattgcg aatatggcaa gcgtaatct 4860
ggcgattcgt tatggcgcta aaggtatttg cagaccatt gtcaccgat gtgcgagcgc 4920
gaacaacagc attggtgagt cttccgtaa cattaaagtt gttatagcg acgttatgat 4980
ctctggtggt agcgaagcag gtatcacccc gttgagcctg gcgggttttg cctcgatgaa 5040
ggccgtgacc aaatctgagg acccgaagcg gccagcatc ccgttcgata aggatcgag 5100
cggttttgtg atgggcgagg gcagcggtat cgttatcttg gaagagtgg agcacgcgt 5160
gaagcgtggt gccaaaatct atgccgagat cgttggtctat ggtgcgacct gcgacgcata 5220
tcatatcacg agcccagcgc cgaatggtga aggtggtgca cgtgcaatga aactggcaat 5280

```

	ggaagaagat aatgtccgcc cagaggacat ttcttatatc aacgcgcacg gtacgagcac	5340
	ggcgtacaat gacagcttcg aaaccaagc gatcaagacg gtcctgggtg aatacgccta	5400
	caaagtgccg gtgtctagca ccaagagcat gaccggccac ctgctggcg ctggcgtgc	5460
	agtcgaagcg attatctgtg ccaaagctat tgaagagggt ttcatccgc cgaccatcgg	5520
	ctacaaagag gcggtccgg aatgcgacct ggattacgtt cctaacgagg gccgtaatgc	5580
	agaagtcaac tacgttctgt ccaacagcct gggttcggt ggccataatg cgactctgct	5640
	gttcaaaaag tacaatatga	5659
	<210> 20	
	<211> 75	
	<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> 20	
	ttgtccatct ttatataatt tggggtagg gtgttcttta tgtaaaaaa acgttttagg	60
	atgcatatgg cggcc	75
	<210> 21	
	<211> 75	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<221> 来源	
[0016]	<223> /注="人工序列的描述: 合成的寡核苷酸"	
	<400> 21	
	gataaatcca cgaatttttag gtttgatgat cattggcttc ctctgcagg tgcgtgttcg	60
	tcgtcatcgc aattg	75
	<210> 22	
	<211> 30	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<221> 来源	
	<223> /注="人工序列的描述: 合成的寡核苷酸"	
	<400> 22	
	actcaccgca ttggtgtagt aaggcgacc	30
	<210> 23	
	<211> 30	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<221> 来源	
	<223> /注="人工序列的描述: 合成的寡核苷酸"	
	<400> 23	
	tgaatgtcat cacgcagttc ccagtcaccc	30
	<210> 24	
	<211> 33	
	<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 ccatcttctt tgtacagacg ttgactgaac atg	33
	<210> 25 <211> 30 <212> DNA <213> 人工序列	
	<220> <221> 来源 <223> /注="人工序列的描述: 合成的寡核苷酸"	
	<400> 25 gcaccatagc cgtaatccca caggttatag	30
	<210> 26 <211> 25 <212> DNA <213> 人工序列	
	<220> <221> 来源 <223> /注="人工序列的描述: 合成的寡核苷酸"	
	<400> 26 tgtcattaat ggtaataat gttga	25
[0017]	<210> 27 <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> 27 gcagttattg gtgcccttaa acgcctggtt gctacgcctg	40
	<210> 28 <211> 27 <212> DNA <213> 人工序列	
	<220> <221> 来源 <223> /注="人工序列的描述: 合成的寡核苷酸"	
	<400> 28 gagccaatat gcgagaacac ccgagaa	27
	<210> 29 <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> 29 cgctgaacgt attgcaggcc gagttgctgc accgctcccg ccaggcag	48
	<210> 30 <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> 30 ggaattgccg cgggtgcggca ggtccatac gcgaggccag gttatccaac g	51
	<210> 31 <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> 31 aatcaccagc actaaagtgc gcggttcggt acccg	35
	<210> 32 <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> 32 atctgccgtg gattgcagag tctattcagc tacg	34
[0018]	<210> 33 <211> 1168 <212> PRT <213> 耻垢分枝杆菌	
	<400> 33 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	

[0019]

130	135	140
Ala Pro Val Ala Gln Leu Gln Pro Ile Val Ala Glu Thr Glu Pro Lys 145	150	155 160
Val Ile Ala Ser Ser Val Asp Phe Leu Ala Asp Ala Val Ala Leu Val 165	170	175
Glu Ser Gly Pro Ala Pro Ser Arg Leu Val Val Phe Asp Tyr Ser His 180	185	190
Glu Val Asp Asp Gln Arg Glu Ala Phe Glu Ala Ala Lys Gly Lys Leu 195	200	205
Ala Gly Thr Gly Val Val Val Glu Thr Ile Thr Asp Ala Leu Asp Arg 210	215	220
Gly Arg Ser Leu Ala Asp Ala Pro Leu Tyr Val Pro Asp Glu Ala Asp 225	230	235 240
Pro Leu Thr Leu Leu Ile Tyr Thr Ser Gly Ser Thr Gly Thr Pro Lys 245	250	255
Gly Ala Met Tyr Pro Glu Ser Lys Thr Ala Thr Met Trp Gln Ala Gly 260	265	270
Ser Lys Ala Arg Trp Asp Glu Thr Leu Gly Val Met Pro Ser Ile Thr 275	280	285
Leu Asn Phe Met Pro Met Ser His Val Met Gly Arg Gly Ile Leu Cys 290	295	300
Ser Thr Leu Ala Ser Gly Gly Thr Ala Tyr Phe Ala Ala Arg Ser Asp 305	310	315 320
Leu Ser Thr Phe Leu Glu Asp Leu Ala Leu Val Arg Pro Thr Gln Leu 325	330	335
Asn Phe Val Pro Arg Ile Trp Asp Met Leu Phe Gln Glu Tyr Gln Ser 340	345	350
Arg Leu Asp Asn Arg Arg Ala Glu Gly Ser Glu Asp Arg Ala Glu Ala 355	360	365
Ala Val Leu Glu Glu Val Arg Thr Gln Leu Leu Gly Gly Arg Phe Val 370	375	380
Ser Ala Leu Thr Gly Ser Ala Pro Ile Ser Ala Glu Met Lys Ser Trp 385	390	395 400
Val Glu Asp Leu Leu Asp Met His Leu Leu Glu Gly Tyr Gly Ser Thr 405	410	415
Glu Ala Gly Ala Val Phe Ile Asp Gly Gln Ile Gln Arg Pro Pro Val 420	425	430
Ile Asp Tyr Lys Leu Val Asp Val Pro Asp Leu Gly Tyr Phe Ala Thr		

[0020]

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
Glu Asp Gly Tyr Tyr Arg Thr Gly Asp Ile Val Ala Glu Leu Gly Pro		
	485	490
Asp His Leu Glu Tyr Leu Asp Arg Arg Asn Asn Val Leu Lys Leu Ser		
	500	505
Gln Gly Glu Phe Val Thr Val Ser Lys Leu Glu Ala Val Phe Gly Asp		
	515	520
Ser Pro Leu Val Arg Gln Ile Tyr Val Tyr Gly Asn Ser Ala Arg Ser		
	530	535
Tyr Leu Leu Ala Val Val Val Pro Thr Glu Glu Ala Leu Ser Arg Trp		
545	550	555
Asp Gly Asp Glu Leu Lys Ser Arg Ile Ser Asp Ser Leu Gln Asp Ala		
	565	570
Ala Arg Ala Ala Gly Leu Gln Ser Tyr Glu Ile Pro Arg Asp Phe Leu		
	580	585
Val Glu Thr Thr Pro Phe Thr Leu Glu Asn Gly Leu Leu Thr Gly Ile		
	595	600
Arg Lys Leu Ala Arg Pro Lys Leu Lys Ala His Tyr Gly Glu Arg Leu		
	610	615
Glu Gln Leu Tyr Thr Asp Leu Ala Glu Gly Gln Ala Asn Glu Leu Arg		
625	630	635
Glu Leu Arg Arg Asn Gly Ala Asp Arg Pro Val Val Glu Thr Val Ser		
	645	650
Arg Ala Ala Val Ala Leu Leu Gly Ala Ser Val Thr Asp Leu Arg Ser		
	660	665
Asp Ala His Phe Thr Asp Leu Gly Gly Asp Ser Leu Ser Ala Leu Ser		
	675	680
Phe Ser Asn Leu Leu His Glu Ile Phe Asp Val Asp Val Pro Val Gly		
	690	695
Val Ile Val Ser Pro Ala Thr Asp Leu Ala Gly Val Ala Ala Tyr Ile		
705	710	715
Glu Gly Glu Leu Arg Gly Ser Lys Arg Pro Thr Tyr Ala Ser Val His		
	725	730
Gly Arg Asp Ala Thr Glu Val Arg Ala Arg Asp Leu Ala Leu Gly Lys		

	740	745	750
	Phe Ile Asp Ala Lys Thr Leu Ser Ala Ala Pro Gly Leu Pro Arg Ser 755 760 765		
	Gly Thr Glu Ile Arg Thr Val Leu Leu Thr Gly Ala Thr Gly Phe Leu 770 775 780		
	Gly Arg Tyr Leu Ala Leu Glu Trp Leu Glu Arg Met Asp Leu Val Asp 785 790 795 800		
	Gly Lys Val Ile Cys Leu Val Arg Ala Arg Ser Asp Asp Glu Ala Arg 805 810 815		
	Ala Arg Leu Asp Ala Thr Phe Asp Thr Gly Asp Ala Thr Leu Leu Glu 820 825 830		
	His Tyr Arg Ala Leu Ala Ala Asp His Leu Glu Val Ile Ala Gly Asp 835 840 845		
	Lys Gly Glu Ala Asp Leu Gly Leu Asp His Asp Thr Trp Gln Arg Leu 850 855 860		
	Ala Asp Thr Val Asp Leu Ile Val Asp Pro Ala Ala Leu Val Asn His 865 870 875 880		
	Val Leu Pro Tyr Ser Gln Met Phe Gly Pro Asn Ala Leu Gly Thr Ala 885 890 895		
[0021]	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 1055	Asp Gly Ile Gly Leu Asp Glu Tyr 1060	
	Val Asp Trp Leu Ile Glu Ala 1070	Gly Tyr Pro Val His Arg Val Asp 1080	
	Asp Tyr Ala Thr Trp Leu Ser 1085	Arg Phe Glu Thr Ala Leu Arg Ala 1095	
	Leu Pro Glu Arg Gln Arg Gln 1100	Ala Ser Leu Leu Pro Leu Leu His 1110	
	Asn Tyr Gln Gln Pro Ser Pro 1115	Pro Val Cys Gly Ala Met Ala Pro 1125	
	Thr Asp Arg Phe Arg Ala Ala 1130	Val Gln Asp Ala Lys Ile Gly Pro 1140	
	Asp Lys Asp Ile Pro His Val 1145	Thr Ala Asp Val Ile Val Lys Tyr 1155	
	Ile Ser Asn Leu Gln Met Leu 1160	Gly Leu Leu 1165	
[0022]	<210> 34		
	<211> 1168		
	<212> PRT		
	<213> 结核分支杆菌		
	<400> 34		
	Met Ser Ile Asn Asp Gln Arg Leu Thr Arg Arg Val Glu Asp Leu Tyr 1 5 10 15		
	Ala Ser Asp Ala Gln Phe Ala Ala Ala Ser Pro Asn Glu Ala Ile Thr 20 25 30		
	Gln Ala Ile Asp Gln Pro Gly Val Ala Leu Pro Gln Leu Ile Arg Met 35 40 45		
	Val Met Glu Gly Tyr Ala Asp Arg Pro Ala Leu Gly Gln Arg Ala Leu 50 55 60		
	Arg Phe Val Thr Asp Pro Asp Ser Gly Arg Thr Met Val Glu Leu Leu 65 70 75 80		
	Pro Arg Phe Glu Thr Ile Thr Tyr Arg Glu Leu Trp Ala Arg Ala Gly 85 90 95		
	Thr Leu Ala Thr Ala Leu Ser Ala Glu Pro Ala Ile Arg Pro Gly Asp 100 105 110		
	Arg Val Cys Val Leu Gly Phe Asn Ser Val Asp Tyr Thr Thr Ile Asp 115 120 125		
	Ile Ala Leu Ile Arg Leu Gly Ala Val Ser Val Pro Leu Gln Thr Ser 130 135 140		

	Ala	Pro	Val	Thr	Gly	Leu	Arg	Pro	Ile	Val	Thr	Glu	Thr	Glu	Pro	Thr	
	145					150					155					160	
	Met	Ile	Ala	Thr	Ser	Ile	Asp	Asn	Leu	Gly	Asp	Ala	Val	Glu	Val	Leu	
					165					170					175		
	Ala	Gly	His	Ala	Pro	Ala	Arg	Leu	Val	Val	Phe	Asp	Tyr	His	Gly	Lys	
				180					185					190			
	Val	Asp	Thr	His	Arg	Glu	Ala	Val	Glu	Ala	Ala	Arg	Ala	Arg	Leu	Ala	
			195					200					205				
	Gly	Ser	Val	Thr	Ile	Asp	Thr	Leu	Ala	Glu	Leu	Ile	Glu	Arg	Gly	Arg	
		210					215					220					
	Ala	Leu	Pro	Ala	Thr	Pro	Ile	Ala	Asp	Ser	Ala	Asp	Asp	Ala	Leu	Ala	
	225					230					235					240	
	Leu	Leu	Ile	Tyr	Thr	Ser	Gly	Ser	Thr	Gly	Ala	Pro	Lys	Gly	Ala	Met	
				245						250					255		
	Tyr	Arg	Glu	Ser	Gln	Val	Met	Ser	Phe	Trp	Arg	Lys	Ser	Ser	Gly	Trp	
				260					265					270			
	Phe	Glu	Pro	Ser	Gly	Tyr	Pro	Ser	Ile	Thr	Leu	Asn	Phe	Met	Pro	Met	
			275					280					285				
[0023]	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				

[0024]

Pro Arg Gly Glu Leu Leu Val Lys Thr Gln Thr Met Phe Pro Gly Tyr
 450 455 460

Tyr Gln Arg Pro Asp Val Thr Ala Glu Val Phe Asp Pro Asp Gly Phe
 465 470 475 480

Tyr Arg Thr Gly Asp Ile Met Ala Lys Val Gly Pro Asp Gln Phe Val
 485 490 495

Tyr Leu Asp Arg Arg Asn Asn Val Leu Lys Leu Ser Gln Gly Glu Phe
 500 505 510

Ile Ala Val Ser Lys Leu Glu Ala Val Phe Gly Asp Ser Pro Leu Val
 515 520 525

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

Val Val Val Pro Ser Gly Asp Ala Leu Ser Arg His Gly Ile Glu Asn
 545 550 555 560

Leu Lys Pro Val Ile Ser Glu Ser Leu Gln Glu Val Ala Arg Ala Ala
 565 570 575

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

Pro Phe Thr Leu Glu Asn Gly Leu Leu Thr Gly Ile Arg Lys Leu Ala
 595 600 605

Arg Pro Gln Leu Lys Lys Phe Tyr Gly Glu Arg Leu Glu Arg Leu Tyr
 610 615 620

Thr Glu Leu Ala Asp Ser Gln Ser Asn Glu Leu Arg Glu Leu Arg Gln
 625 630 635 640

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

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

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

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

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

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

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

	Ala Thr Leu Ala Ala Ala Pro Asn Leu Pro Ala Pro Ser Ala Gln Val	755	760	765
	Arg Thr Val Leu Leu Thr Gly Ala Thr Gly Phe Leu Gly Arg Tyr Leu	770	775	780
	Ala Leu Glu Trp Leu Asp Arg Met Asp Leu Val Asn Gly Lys Leu Ile	785	790	795
	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
[0025]	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

<210> 35
<211> 696
<212> DNA
<213> 细长聚球藻

[0026]

<400> 35
atgccgcagc ttgaagccag ccttgaactg gactttcaaa gcgagtccta caaagacgct 60
tacagccgca tcaacgcgat cgtgattgaa ggcaacaag aggcgttcga caactacaat 120
cgccttgctg agatgctgcc cgaccagcgg gatgagcttc acaagctagc caagatggaa 180
cagcgccaca tgaaggett tatggcctgt ggcaaaaate tctccgtcac tctgacatg 240
ggttttgccc agaaattttt cgagcgcttg cagagaact tcaaagcggc ggctgcggaa 300
ggcaaggtcg tcacctgect actgattcaa tcgctaatac tcgagtgttt tgcgatcgcg 360
gcttacaaca tctacatccc agtggcggat gcttttgccc gcaaaatcac ggagggggtc 420
gtgcgcgacg aatacctgca ccgcaacttc ggtgaagagt ggctgaaggc gaattttgat 480
gcttccaaag ccgaactgga agaagccaat cgtcagaacc tgcccttggt ttggctaata 540
ctcaacgaag tggccgatga tgctcgcgaa ctcgggatgg agcgtgagtc gctcgtcgag 600
gactttatga ttgcctacgg tgaagctctg gaaaacatcg gcttcacaac gcgcgaaatc 660
atgcgtatgt ccgcctatgg ccttgcggcc gtttga 696

<210> 36
<211> 231
<212> PRT
<213> 细长聚球藻

<400> 36
Met Pro Gln Leu Glu Ala Ser Leu Glu Leu Asp Phe Gln Ser Glu Ser
1 5 10 15
Tyr Lys Asp Ala Tyr Ser Arg Ile Asn Ala Ile Val Ile Glu Gly Glu
20 25 30

	Gln Glu Ala Phe Asp Asn Tyr Asn Arg Leu Ala Glu Met Leu Pro Asp	
	35 40 45	
	Gln Arg Asp Glu Leu His Lys Leu Ala Lys Met Glu Gln Arg His Met	
	50 55 60	
	Lys Gly Phe Met Ala Cys Gly Lys Asn Leu Ser Val Thr Pro Asp Met	
	65 70 75 80	
	Gly Phe Ala Gln Lys Phe Phe Glu Arg Leu His Glu Asn Phe Lys Ala	
	85 90 95	
	Ala Ala Ala Glu Gly Lys Val Val Thr Cys Leu Leu Ile Gln Ser Leu	
	100 105 110	
	Ile Ile Glu Cys Phe Ala Ile Ala Ala Tyr Asn Ile Tyr Ile Pro Val	
	115 120 125	
	Ala Asp Ala Phe Ala Arg Lys Ile Thr Glu Gly Val Val Arg Asp Glu	
	130 135 140	
	Tyr Leu His Arg Asn Phe Gly Glu Glu Trp Leu Lys Ala Asn Phe Asp	
	145 150 155 160	
	Ala Ser Lys Ala Glu Leu Glu Glu Ala Asn Arg Gln Asn Leu Pro Leu	
	165 170 175	
[0027]	Val Trp Leu Met Leu Asn Glu Val Ala Asp Asp Ala Arg Glu Leu Gly	
	180 185 190	
	Met Glu Arg Glu Ser Leu Val Glu Asp Phe Met Ile Ala Tyr Gly Glu	
	195 200 205	
	Ala Leu Glu Asn Ile Gly Phe Thr Thr Arg Glu Ile Met Arg Met Ser	
	210 215 220	
	Ala Tyr Gly Leu Ala Ala Val	
	225 230	
	<210> 37	
	<211> 1029	
	<212> DNA	
	<213> 细长聚球藻	
	<400> 37	
	atggcattcg gtcttatcgg tcatttcacc agtttggagc aggcccgca cgtttctcgc	60
	aggatgggct acgacgaata cgccgatcaa ggattggagt tttggagtag cgctcctcct	120
	caaatcgttg atgaaatcac agtcaccagt gccacaggca aggtgattca cggtcgctac	180
	atcgaatcgt gtttcttgcc ggaaatgctg gcggcgcgcc gcttcaaac agccacgcgc	240
	aaagtctca atgccatgtc ccatgcccaa aaacacggca tcgacatctc ggccttgagg	300
	ggctttacct cgattatttt cgagaatttc gatitggcca gtttgcggca agtgcgcgac	360
	actaccttg agtttgaacg gttcaccacc ggcaatactc acacggccta cgtaatctgt	420
	agacaggtgg aagccgctgc taaaacgctg ggcatcgaca ttaccaagc gacagtagcg	480

```

gtgtcgcgcg cgactggcga tateggtagc gctgtctgcc gctggctcga cctcaaactg      540
gggtcgcgtg atttgatcct gacggcgcg c aatcaggagc gtttgataa cctgcaggct      600
gaactcggcc ggggcaagat tctgcccttg gaagccgctc tgccggaagc tgactttatc      660
gtgtgggtcg ccagtatgcc tcaggcgcta gtgatcgacc cagcaaccct gaagcaaccc      720
tgctcctaa tcgacggggg ctacccaaa aacttgggca gcaaagtcca aggtgagggc      780
atctatgtcc tcaatggcgg ggtagttaa cattgcttcg acatcgactg gcagatcatg      840
tccgtcgtag agatggcgcg gcccgagcgc cagatgtttg cctgctttgc cgaggcgatg      900
ctcttgaat ttgaaggctg gcatactaac ttctcctggg gccgcaacca aatcacgac      960
gagaagatgg aagcgatcgg tgaggcatcg gtgcgccacg gcttccaacc ctgggcattg     1020
gcaatttga                                     1029

```

<210> 38
 <211> 342
 <212> PRT
 <213> 细长聚球藻

<400> 38
 Met Ala Phe Gly Leu Ile Gly His Leu Thr Ser Leu Glu Gln Ala Arg
 1 5 10 15

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

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

[0028]

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

Phe Leu Pro Glu Met Leu Ala Ala Arg Arg Phe Lys Thr Ala Thr Arg
 65 70 75 80

Lys Val Leu Asn Ala Met Ser His Ala Gln Lys His Gly Ile Asp Ile
 85 90 95

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

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

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

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

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

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

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

195	200	205
Pro Leu Glu Ala Ala Leu 210	Pro Glu Ala Asp Phe 215	Ile Val Trp Val Ala 220
Ser Met Pro Gln Gly Val Val 225	Ile Asp Pro Ala Thr 230	Leu Lys Gln Pro 240
Cys Val Leu Ile Asp Gly Gly Tyr Pro 245	Lys Asn Leu Gly Ser 250	Lys Val 255
Gln Gly Glu Gly Ile Tyr Val Leu 260	Asn Gly Gly Val Val 265	Glu His Cys 270
Phe Asp Ile Asp Trp Gln Ile 275	Met Ser Ala Ala Glu 280	Met Ala Arg Pro 285
Glu Arg Gln Met Phe Ala Cys 290	Phe Ala Glu Ala Met 295	Leu Leu Glu Phe 300
Glu Gly Trp His Thr Asn Phe Ser Trp 305	Gly Arg Asn Gln Ile Thr 310	Ile 320
Glu Lys Met Glu Ala Ile Gly Glu Ala 325	Ser Val Arg His Gly 330	Phe Gln 335
Pro Leu Ala Leu Ala Ile 340		

[0029]

<210> 39
 <211> 717
 <212> DNA
 <213> *Prochlorococcus mariunus*

 <400> 39
 atgcaaacac tcgaatctaa taaaaaaact aatctagaaa attctattga ttacccgat 60
 ttactactg attcttacaa agacgcttat agcaggataa atgcaatagt tattgaaggt 120
 gaacaagagg ctcattgataa ttacatttcc ttagcaacat taattcctaa cgaattagaa 180
 gagttaacta aattagcgaa aatggagctt aagcacaaaa gaggctttac tgcattgtga 240
 agaaatctag gtgttcaagc tgacatgatt ttgctaaag aattcttttc caaattacat 300
 ggtaattttc aggttgcgtt atctaattgc aagacaacta catgcctatt aatacaggca 360
 attttaattg aagctttttg tatatccgcg taccagttt acataagagt tgctgacct 420
 ttgcgaaaa aaattacca aggtgttgtt aaagatgaat atcttcattt aaattatgga 480
 caagaatggc taaaagaaaa tttagcgact tgtaaatgag agctaatgga agcaataaag 540
 gttacacctt cattaatcaa gaagatgta gatcaagtct cggaagatgc ttcagtacta 600
 gctatggata gggaagaatt aatggaagaa ttcattgatt cctatcagga cactctcctt 660
 gaaataggtt tagataatag agaaattgca agaattggca tggtgctat agtttaa 717

 <210> 40
 <211> 238
 <212> PRT
 <213> *Prochlorococcus mariunus*

 <400> 40

	Met	Gln	Thr	Leu	Glu	Ser	Asn	Lys	Lys	Thr	Asn	Leu	Glu	Asn	Ser	Ile	
	1				5					10					15		
	Asp	Leu	Pro	Asp	Phe	Thr	Thr	Asp	Ser	Tyr	Lys	Asp	Ala	Tyr	Ser	Arg	
				20					25					30			
	Ile	Asn	Ala	Ile	Val	Ile	Glu	Gly	Glu	Gln	Glu	Ala	His	Asp	Asn	Tyr	
			35					40					45				
	Ile	Ser	Leu	Ala	Thr	Leu	Ile	Pro	Asn	Glu	Leu	Glu	Glu	Leu	Thr	Lys	
		50					55				60						
	Leu	Ala	Lys	Met	Glu	Leu	Lys	His	Lys	Arg	Gly	Phe	Thr	Ala	Cys	Gly	
	65					70				75						80	
	Arg	Asn	Leu	Gly	Val	Gln	Ala	Asp	Met	Ile	Phe	Ala	Lys	Glu	Phe	Phe	
					85					90					95		
	Ser	Lys	Leu	His	Gly	Asn	Phe	Gln	Val	Ala	Leu	Ser	Asn	Gly	Lys	Thr	
			100						105					110			
	Thr	Thr	Cys	Leu	Leu	Ile	Gln	Ala	Ile	Leu	Ile	Glu	Ala	Phe	Ala	Ile	
			115				120						125				
	Ser	Ala	Tyr	His	Val	Tyr	Ile	Arg	Val	Ala	Asp	Pro	Phe	Ala	Lys	Lys	
		130					135					140					
[0030]	Ile	Thr	Gln	Gly	Val	Val	Lys	Asp	Glu	Tyr	Leu	His	Leu	Asn	Tyr	Gly	
	145				150						155					160	
	Gln	Glu	Trp	Leu	Lys	Glu	Asn	Leu	Ala	Thr	Cys	Lys	Asp	Glu	Leu	Met	
				165						170					175		
	Glu	Ala	Asn	Lys	Val	Asn	Leu	Pro	Leu	Ile	Lys	Lys	Met	Leu	Asp	Gln	
				180				185						190			
	Val	Ser	Glu	Asp	Ala	Ser	Val	Leu	Ala	Met	Asp	Arg	Glu	Glu	Leu	Met	
			195					200				205					
	Glu	Glu	Phe	Met	Ile	Ala	Tyr	Gln	Asp	Thr	Leu	Leu	Glu	Ile	Gly	Leu	
		210					215					220					
	Asp	Asn	Arg	Glu	Ile	Ala	Arg	Met	Ala	Met	Ala	Ala	Ile	Val			
	225					230					235						
	<210>	41															
	<211>	1044															
	<212>	DNA															
	<213>	Prochlorococcus mariunus															
	<400>	41															
	atggcatttg	ggcttatagg	tcattcaact	agttttgaag	atgcaaaaag	aaaggcttca										60	
	ttattgggct	ttgatcatat	tcgggatggt	gatttagatg	tttggtgcac	agctccacct										120	
	caactagt	ttgaaaatgtaga	ggttaaaagt	gctataggta	tatcaattga	aggttcttat										180	
	attgattcat	gtttcggtcc	tgaaatgctt	tcaagattta	aaacggcaag	aagaaaagta										240	
	ttaaatagcaa	tggaattagc	tcaaaaaaaa	ggtattaata	ttaccgcttt	gggggggttc										300	

```

acttctatca tctttgaaaa ttttaatctc cttcaacata agcagattag aaacacttca    360
ctagagtggg aaaggtttac aactggtaat actcatactg cgtgggttat ttgcaggcaa    420
ttagagatga atgctcctaa aataggtatt gatcttaaaa gcgcaacagt tgctgtagtt    480
ggtgctactg gagatatagg cagtgcigtg ttgctgatgg taatcaataa aacaggtatt    540
ggggaacttc ttttgtagc taggcaaaag gaacccttgg attctttgca aaaggaatta    600
gatggtggaa ctatcaaaaa tctagatgaa gcattgcctg aagcagatat tgttgtaggg    660
gtagcaagta tgccaaagac aatggaaatc gatgctaata atcttaaca accatgttta    720
atgattgatg gaggttatcc aaagaatcta gatgaaaaat ttcaaggaaa taatatacat    780
gttgtaaaag gaggtatagt aagattcttc aatgatatag gttggaatat gatggaacta    840
gctgaaatgc aaaatcccca gagagaaatg tttgcatgct ttgcagaagc aatgatttta    900
gaatttgaaa aatgtcatac aaacttttagc tggggaagaa ataatatatc tctcgagaaa    960
atggagtitta ttggagctgc ttctgtaaag catggcttct ctgcaattgg cctagataag   1020
catccaaaag tactagcagt ttga                                           1044

```

<210> 42
 <211> 347
 <212> PRT
 <213> *Prochlorococcus marinus*

<400> 42
 Met Ala Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Asp Ala Lys
 1 5 10 15

[0031]

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

Asp Val Trp Cys Thr Ala Pro Pro Gln Leu Val Glu Asn Val Glu Val
 35 40 45

Lys Ser Ala Ile Gly Ile Ser Ile Glu Gly Ser Tyr Ile Asp Ser Cys
 50 55 60

Phe Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val
 65 70 75 80

Leu Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala
 85 90 95

Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln
 100 105 110

His Lys Gln Ile Arg Asn Thr Ser Leu Glu Trp Glu Arg Phe Thr Thr
 115 120 125

Gly Asn Thr His Thr Ala Trp Val Ile Cys Arg Gln Leu Glu Met Asn
 130 135 140

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

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

Lys Thr Gly Ile Gly Glu Leu Leu Val Ala Arg Gln Lys Glu Pro	
180 185 190	
Leu Asp Ser Leu Gln Lys Glu Leu Asp Gly Gly Thr Ile Lys Asn Leu	
195 200 205	
Asp Glu Ala Leu Pro Glu Ala Asp Ile Val Val Trp Val Ala Ser Met	
210 215 220	
Pro Lys Thr Met Glu Ile Asp Ala Asn Asn Leu Lys Gln Pro Cys Leu	
225 230 235 240	
Met Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Glu Lys Phe Gln Gly	
245 250 255	
Asn Asn Ile His Val Val Lys Gly Gly Ile Val Arg Phe Phe Asn Asp	
260 265 270	
Ile Gly Trp Asn Met Met Glu Leu Ala Glu Met Gln Asn Pro Gln Arg	
275 280 285	
Glu Met Phe Ala Cys Phe Ala Glu Ala Met Ile Leu Glu Phe Glu Lys	
290 295 300	
Cys His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Ser Leu Glu Lys	
305 310 315 320	
[0032] Met Glu Phe Ile Gly Ala Ala Ser Val Lys His Gly Phe Ser Ala Ile	
325 330 335	
Gly Leu Asp Lys His Pro Lys Val Leu Ala Val	
340 345	
<210> 43	
<211> 255	
<212> DNA	
<213> 念球藻	
<400> 43	
atgagccaaa cggaactttt tgaaaaggtc aagaaaatcg tcatcgaaca actgagtgtt	60
gaagatgctt ccaaaatcac tccacaagct aagtttatgg aagatttagg agctgattcc	120
ctggatactg ttgaactcgt gatggctttg gaagaagaat ttgatatcga aattcccgac	180
gaagctgccg agcagattgt atcggttcaa gacgcagtag attacatcaa taacaaagtt	240
gctgcatcag cttaa	255
<210> 44	
<211> 84	
<212> PRT	
<213> 念球藻	
<400> 44	
Met Ser Gln Thr Glu Leu Phe Glu Lys Val Lys Lys Ile Val Ile Glu	
1 5 10 15	
Gln Leu Ser Val Glu Asp Ala Ser Lys Ile Thr Pro Gln Ala Lys Phe	
20 25 30	

Met Glu Asp Leu Gly Ala Asp Ser Leu Asp Thr Val Glu Leu Val Met
35 40 45

Ala Leu Glu Glu Glu Phe Asp Ile Glu Ile Pro Asp Glu Ala Ala Glu
50 55 60

Gln Ile Val Ser Val Gln Asp Ala Val Asp Tyr Ile Asn Asn Lys Val
65 70 75 80

Ala Ala Ser Ala

<210> 45
<211> 234
<212> DNA
<213> 集胞藻

<400> 45
atgaatcagg aaatttttga aaaagtaaaa aaaatcgctg tggaacagtt ggaagtggat 60
cctgacaaaag tgacccccga tgccaccttt gccgaagatt tagggctga ttccctcgat 120
acagtggaaat tggatcatggc cctggaagaa gagtttgata ttgaaattcc cgatgaagtg 180
gcggaacca ttgataccgt gggcaaagcc gttgagcata tcgaaagtaa ataa 234

<210> 46
<211> 77
<212> PRT
<213> 集胞藻

[0033]

<400> 46
Met Asn Gln Glu Ile Phe Glu Lys Val Lys Lys Ile Val Val Glu Gln
1 5 10 15

Leu Glu Val Asp Pro Asp Lys Val Thr Pro Asp Ala Thr Phe Ala Glu
20 25 30

Asp Leu Gly Ala Asp Ser Leu Asp Thr Val Glu Leu Val Met Ala Leu
35 40 45

Glu Glu Glu Phe Asp Ile Glu Ile Pro Asp Glu Val Ala Glu Thr Ile
50 55 60

Asp Thr Val Gly Lys Ala Val Glu His Ile Glu Ser Lys
65 70 75

<210> 47
<211> 243
<212> DNA
<213> 原绿球藻

<400> 47
atgtcacaag aagaaatcct tcaaaaagta tgctctattg ttctgagca actaagtgtt 60
gaatcagccg aagtaaaatc tgattcaaac ttcaaaatg atttaggtgc agactcccta 120
gacaccgtag agctagttat ggctcttgaa gaagcatttg atatcgagat acctgatgaa 180
gcagctgaag gtatcgcaac agtaggagat gctgttaaatt tcatcgaaga aaaaaaaggt 240
taa 243

<210> 48
<211> 80

	<212> PRT	
	<213> 原绿球藻	
	<400> 48	
	Met Ser Gln Glu Glu Ile Leu Gln Lys Val Cys Ser Ile Val Ser Glu	
	1 5 10 15	
	Gln Leu Ser Val Glu Ser Ala Glu Val Lys Ser Asp Ser Asn Phe Gln	
	20 25 30	
	Asn Asp Leu Gly Ala Asp Ser Leu Asp Thr Val Glu Leu Val Met Ala	
	35 40 45	
	Leu Glu Glu Ala Phe Asp Ile Glu Ile Pro Asp Glu Ala Ala Glu Gly	
	50 55 60	
	Ile Ala Thr Val Gly Asp Ala Val Lys Phe Ile Glu Glu Lys Lys Gly	
	65 70 75 80	
	<210> 49	
	<211> 243	
	<212> DNA	
	<213> 细长聚球藻	
	<400> 49	
	atgagccaag aagacatctt cagcaaagtc aaagacattg tggctgagca gctgagtgtg	60
	gatgtggctg aagtcaagcc agaatccagc ttccaaaacg atctgggagc ggactcgctg	120
	gacaccgtgg aactggtgat ggctctggaa gaggctttcg atatcgaaat ccccgatgaa	180
	gccgtgaag gcattgcgac cgttcaagac gccgtcgatt tcatcgctag caaagctgcc	240
[0034]	tag	243
	<210> 50	
	<211> 80	
	<212> PRT	
	<213> 细长聚球藻	
	<400> 50	
	Met Ser Gln Glu Asp Ile Phe Ser Lys Val Lys Asp Ile Val Ala Glu	
	1 5 10 15	
	Gln Leu Ser Val Asp Val Ala Glu Val Lys Pro Glu Ser Ser Phe Gln	
	20 25 30	
	Asn Asp Leu Gly Ala Asp Ser Leu Asp Thr Val Glu Leu Val Met Ala	
	35 40 45	
	Leu Glu Glu Ala Phe Asp Ile Glu Ile Pro Asp Glu Ala Ala Glu Gly	
	50 55 60	
	Ile Ala Thr Val Gln Asp Ala Val Asp Phe Ile Ala Ser Lys Ala Ala	
	65 70 75 80	
	<210> 51	
	<211> 255	
	<212> DNA	
	<213> 念珠菌菌种	
	<400> 51	
	atgagccaat cagaaacttt tgaaaaagtc aaaaaaattg ttatcgaaca actaagtgtg	60
	gagaaccctg acacagtaac tccagaagct agttttgcc aagatttaca ggctgattcc	120

ctcgatacag tagaactagt aatggctttg gaagaagaat ttgatatcga aattcccgat 180
 gaagccgcag agaaaattac cactgttcaa gaagcgttg attacatcaa taaccaagtt 240
 gccgcatcag cttaa 255

<210> 52
 <211> 84
 <212> PRT
 <213> 念珠菌菌种

<400> 52
 Met Ser Gln Ser Glu Thr Phe Glu Lys Val Lys Lys Ile Val Ile Glu
 1 5 10 15

Gln Leu Ser Val Glu Asn Pro Asp Thr Val Thr Pro Glu Ala Ser Phe
 20 25 30

Ala Asn Asp Leu Gln Ala Asp Ser Leu Asp Thr Val Glu Leu Val Met
 35 40 45

Ala Leu Glu Glu Glu Phe Asp Ile Glu Ile Pro Asp Glu Ala Ala Glu
 50 55 60

Lys Ile Thr Thr Val Gln Glu Ala Val Asp Tyr Ile Asn Asn Gln Val
 65 70 75 80

Ala Ala Ser Ala

[0035]

<210> 53
 <211> 675
 <212> DNA
 <213> 枯草芽孢杆菌

<400> 53
 atgaagattt acggaattta tatggaccgc ccgctttcac aggaagaaaa tgaacggttc 60
 atgactttca tatcacctga aaaacgggag aaatgccgga gattttatca taaagaagat 120
 gctcaccgca ccctgctggg agatgtgctc gttcgtcag tcataagcag gcagtatcag 180
 ttggacaaat ccgatatccg ctttagcacg caggaatagc ggaagccgtg catccctgat 240
 ctccccgacg ctcatattcaa cattttcac tccggccgct gggtcattgg tgcgtttgat 300
 tcacagccga tcggcataga tatcgaaaaa acgaaaccga tcagccttga gatcgccaag 360
 cgctttcttt caaaaacaga gtacagcgac cttttagcaa aagacaagga cgagcagaca 420
 gactattttt atcatctatg gtcaatgaaa gaaagcttta tcaaacagga aggcaaaggc 480
 ttatcgcttc cgcttgattc cttttcagtg cgctgcac aggcaggaca agtatccatt 540
 gagcttccgg acagccattc cccatgctat atcaaaacgt atgaggtcga tcccggctac 600
 aaaatggctg tatgcgccgc acaccctgat ttccccgagg atatcacaat ggtctcgtac 660
 gaagagcttt tataa 675

<210> 54
 <211> 224
 <212> PRT
 <213> 枯草芽孢杆菌

<400> 54
 Met Lys Ile Tyr Gly Ile Tyr Met Asp Arg Pro Leu Ser Gln Glu Glu

1	5	10	15
Asn Glu Arg Phe Met Thr Phe Ile Ser Pro Glu Lys Arg Glu Lys Cys	20	25	30
Arg Arg Phe Tyr His Lys Glu Asp Ala His Arg Thr Leu Leu Gly Asp	35	40	45
Val Leu Val Arg Ser Val Ile Ser Arg Gln Tyr Gln Leu Asp Lys Ser	50	55	60
Asp Ile Arg Phe Ser Thr Gln Glu Tyr Gly Lys Pro Cys Ile Pro Asp	65	70	75
Leu Pro Asp Ala His Phe Asn Ile Ser His Ser Gly Arg Trp Val Ile	85	90	95
Gly Ala Phe Asp Ser Gln Pro Ile Gly Ile Asp Ile Glu Lys Thr Lys	100	105	110
Pro Ile Ser Leu Glu Ile Ala Lys Arg Phe Phe Ser Lys Thr Glu Tyr	115	120	125
Ser Asp Leu Leu Ala Lys Asp Lys Asp Glu Gln Thr Asp Tyr Phe Tyr	130	135	140
His Leu Trp Ser Met Lys Glu Ser Phe Ile Lys Gln Glu Gly Lys Gly	145	150	155
Leu Ser Leu Pro Leu Asp Ser Phe Ser Val Arg Leu His Gln Asp Gly	165	170	175
Gln Val Ser Ile Glu Leu Pro Asp Ser His Ser Pro Cys Tyr Ile Lys	180	185	190
Thr Tyr Glu Val Asp Pro Gly Tyr Lys Met Ala Val Cys Ala Ala His	195	200	205
Pro Asp Phe Pro Glu Asp Ile Thr Met Val Ser Tyr Glu Glu Leu Leu	210	215	220
[0036]			
<210> 55			
<211> 867			
<212> DNA			
<213> 谷氨酸棒状杆菌			
<400> 55			
ttgggcgtgt cgcccttaaa gcgcgctttt cgacgcgacc ccactacatt ggcttccatg			60
aacgttgaca ttacacgac cagagagccg ctaaacgttg agctcctgaa ggaaaaattg			120
ctccaaaacg gtgactttgg ccaggtcatt tacgaaaaag tgacaggctc cactaatgct			180
gacttgctgg cacttgccagg ttctggcgct ccaaactgga cggtgaaaac tgtcgagttt			240
caagatcatg cgcgtgggag actcgccgcg ccgtggtctg cccctgaggg ttcccaaaca			300
atcgtgtctg tgctcgttca actatctatt gatcaagtgg accgatttgg cactattcca			360
ctcgccggcg gactcgctgt catggatgcg ttgaatgacc tcggtgtgga aggtgccgga			420
ctgaaatggc ccaacgatgt tcaaatccac ggcaagaaac tctgcggcat cctggtggaa			480

	gccaccggct ttgattccac cccaacagtt gtcacggtt ggggcactaa taccagcctg	540
	actaaagagg agcttcctgt tcctcatgca acttcctcgc cattggaagg tgttgaagtc	600
	gacagaacca cattccttat taatatgctc acacatctgc atactcgact ggaccagtgg	660
	cagggtccaa gtgtggattg gctcgaatg tacctgctgg tatgttccag tattggccaa	720
	gatgttcgag tgcttctacc tggggataaa gaactcttag gtgaagcgat cgggtgctcg	780
	actggcggag aaattcgtgt tcgcatgct tcgggcaccg ttcacaccct caacgccggt	840
	gaaattacgc accttcgcct gcagtaa	867
	<210> 56	
	<211> 810	
	<212> DNA	
	<213> 谷氨酸棒状杆菌	
	<400> 56	
	atgaatgttg acattagccg ctctcgtgaa ccgttgaacg tggaaactgtt gaaagaaaaa	60
	ctgtcgcaga acggtgattt cgggtcaagt atctacgaga aggtcaccgg ctctaccaat	120
	gcggacctgc tggtctctggc gggcagcggc gctccaaact ggaccgtcaa gactgttgaa	180
	tttcaggacc acgcccgtgg ccgtctgggt cgctcgtgga gcgcaccgga gggttcccaa	240
	accatcgta gcgttcttgt ccaactgagc attgatcagg tggaccgtat tggtagatc	300
	ccgtggccg caggtctggc tgttatggat gcgctgaatg atctgggcgt ggagggtgca	360
	ggcctgaaat ggccgaacga tgttcagatc cacgtaaga agttgtgcgg tattctggtt	420
	gaagcaaccg gcttcgactc cactccgacc gtggttatcg gttgggttac gaatatctcg	480
[0037]	ttgacgaaag aagagctgcc ggtccgcac gcgaccagcc tggccctgga ggtgttgaa	540
	gttgaccgta cgacgttctt gattaacatg ctgaccatc tgcatacccg tctggatcag	600
	tggcagggtc cgtctgtgga ctggctggat gactatcgcg cggttttag cagcattggc	660
	caagatgtgc gtgtcctgct gcctgggtgac aaagagctgc tgggcgaggc gattggcgtg	720
	gcgaccggtg gtgagatccg tgtgcgcgac gccagcggca cggtcacac gctgaatgcg	780
	ggtgaaatca cgcactctcg tttgcaataa	810
	<210> 57	
	<211> 269	
	<212> PRT	
	<213> 谷氨酸棒状杆菌	
	<400> 57	
	Met Asn Val Asp Ile Ser Arg Ser Arg Glu Pro Leu Asn Val Glu Leu	
	1 5 10 15	
	Leu Lys Glu Lys Leu Leu Gln Asn Gly Asp Phe Gly Gln Val Ile Tyr	
	20 25 30	
	Glu Lys Val Thr Gly Ser Thr Asn Ala Asp Leu Leu Ala Leu Ala Gly	
	35 40 45	
	Ser Gly Ala Pro Asn Trp Thr Val Lys Thr Val Glu Phe Gln Asp His	
	50 55 60	
	Ala Arg Gly Arg Leu Gly Arg Pro Trp Ser Ala Pro Glu Gly Ser Gln	
	65 70 75 80	

	Thr Ile Val Ser Val Leu Val Gln Leu Ser Ile Asp Gln Val Asp Arg	
	85 90 95	
	Ile Gly Thr Ile Pro Leu Ala Ala Gly Leu Ala Val Met Asp Ala Leu	
	100 105 110	
	Asn Asp Leu Gly Val Glu Gly Ala Gly Leu Lys Trp Pro Asn Asp Val	
	115 120 125	
	Gln Ile His Gly Lys Lys Leu Cys Gly Ile Leu Val Glu Ala Thr Gly	
	130 135 140	
	Phe Asp Ser Thr Pro Thr Val Val Ile Gly Trp Gly Thr Asn Ile Ser	
	145 150 155 160	
	Leu Thr Lys Glu Glu Leu Pro Val Pro His Ala Thr Ser Leu Ala Leu	
	165 170 175	
	Glu Gly Val Glu Val Asp Arg Thr Thr Phe Leu Ile Asn Met Leu Thr	
	180 185 190	
	His Leu His Thr Arg Leu Asp Gln Trp Gln Gly Pro Ser Val Asp Trp	
	195 200 205	
	Leu Asp Asp Tyr Arg Ala Val Cys Ser Ser Ile Gly Gln Asp Val Arg	
	210 215 220	
[0038]	Val Leu Leu Pro Gly Asp Lys Glu Leu Leu Gly Glu Ala Ile Gly Val	
	225 230 235 240	
	Ala Thr Gly Gly Glu Ile Arg Val Arg Asp Ala Ser Gly Thr Val His	
	245 250 255	
	Thr Leu Asn Ala Gly Glu Ile Thr His Leu Arg Leu Gln	
	260 265	
	<210> 58	
	<211> 1632	
	<212> DNA	
	<213> 谷氨酸棒状杆菌	
	<400> 58	
	atgaccattt cctcaccttt gattgacgtc gccaaccttc cagacatcaa caccactgcc	60
	ggcaagatcg ccgaccttaa ggctcgccgc gcggaagccc atttcccat gggtgaaaag	120
	gcagtagaga aggtccacgc tgctggacgc ctactgccc gtgagcgctt ggattactta	180
	ctcgatgagg gtccttcat cgagaccgat cagctggctc gccaccgcac caccgctttc	240
	ggcctgggcg ctaagcgtec tgcaaccgac ggcacgtga ccgctgggg caccattgat	300
	ggacgcgaag tctgcatctt ctgcaggac ggcaccgtat tcggtggcgc gcttggtgag	360
	gtgtacggcg aaaagatgat caagatcatg gagctggcaa tcgacaccgg ccgcccattg	420
	atcggtcttt acgaaggcgc tggcgctcgt attcaggacg gcgctgtctc cctggacttc	480
	atttccaga ccttctacca aaacattcag gcttctggcg ttatccaca gatctccgtc	540
	atcatgggcg catgtgcagg tggcaacgct tacggcccag ctctgaccga ctctgtggtc	600
	atggtggaca agacctcaa gatgttcgtt accggcccag acgtgatcaa gaccgtcacc	660

```

ggcgaggaaa tcaccagga agagcttggc ggagcaacca cccacatggt gaccgctggt 720
aactcccact acaccgtgc gaccgatgag gaagcactgg attgggtaca ggacctggtg 780
tccttcctec catccaacaa tcgtctctac gcaccgatgg aagacttcga cgaggaagaa 840
ggcgcgcttg aaaaaacat caccgctgac gatctgaagc tcgacgagat catcccagat 900
tcgcgcacgg ttctttacga cgtccgcatg gtcacgaat gcctcaccga cgatggcgaa 960
tacctggaaa tccaggcaga ccgcgcagaa aacgttgtaa ttgcattcgg ccgcatcgaa 1020
ggccagtcgg ttggctttgt tgccaaccag ccaaccagtg tcgttgctg cctggacatc 1080
gactcctctg agaaggcagc tcgttcgtc cgcacctgag acgcgttcaa catcccaatc 1140
gtcatgcttg tcgacgtccc cggttcctc ccaggcgag gccaggagta cgttggcatt 1200
ctgcgtcgtg gcgcaaagct gctctacga tacggcgaag caaccgttcc aaagatcacc 1260
gtcacatgc gtaaggctta cggcggagcg tactgcgtga tgggttccaa gggcttgggc 1320
tctgacatca accttgcatg gccaacgcga cagatcgccg tcatgggcgc tgctggcgca 1380
gttgattca tctaccgaa ggagctcatg gcagctgatg ccaagggcct cgataaccga 1440
gctctggcta agtccttga gcgcgagtat gaagaccaca tgctcaaccc gtaccacgt 1500
gcagaacgtg gctgatcga cgcctgatc ctgccaagcg aaaccgcgg acagatttcc 1560
cgcaaccttc gctgctcaa gcacaagaac gtcactgcc ctgctcgaa gcacggcaac 1620
atgccactgt aa 1632

```

[0039]

<210> 59
 <211> 543
 <212> PRT
 <213> 谷氨酸棒状杆菌

```

<400> 59
Met Thr Ile Ser Ser Pro Leu Ile Asp Val Ala Asn Leu Pro Asp Ile
1          5          10          15

Asn Thr Thr Ala Gly Lys Ile Ala Asp Leu Lys Ala Arg Arg Ala Glu
20          25          30

Ala His Phe Pro Met Gly Glu Lys Ala Val Glu Lys Val His Ala Ala
35          40          45

Gly Arg Leu Thr Ala Arg Glu Arg Leu Asp Tyr Leu Leu Asp Glu Gly
50          55          60

Ser Phe Ile Glu Thr Asp Gln Leu Ala Arg His Arg Thr Thr Ala Phe
65          70          75          80

Gly Leu Gly Ala Lys Arg Pro Ala Thr Asp Gly Ile Val Thr Gly Trp
85          90          95

Gly Thr Ile Asp Gly Arg Glu Val Cys Ile Phe Ser Gln Asp Gly Thr
100         105         110

Val Phe Gly Gly Ala Leu Gly Glu Val Tyr Gly Glu Lys Met Ile Lys
115         120         125

Ile Met Glu Leu Ala Ile Asp Thr Gly Arg Pro Leu Ile Gly Leu Tyr
130         135         140

```

	Glu	Gly	Ala	Gly	Ala	Arg	Ile	Gln	Asp	Gly	Ala	Val	Ser	Leu	Asp	Phe	
	145					150					155					160	
	Ile	Ser	Gln	Thr	Phe	Tyr	Gln	Asn	Ile	Gln	Ala	Ser	Gly	Val	Ile	Pro	
					165					170					175		
	Gln	Ile	Ser	Val	Ile	Met	Gly	Ala	Cys	Ala	Gly	Gly	Asn	Ala	Tyr	Gly	
				180					185						190		
	Pro	Ala	Leu	Thr	Asp	Phe	Val	Val	Met	Val	Asp	Lys	Thr	Ser	Lys	Met	
			195					200					205				
	Phe	Val	Thr	Gly	Pro	Asp	Val	Ile	Lys	Thr	Val	Thr	Gly	Glu	Glu	Ile	
	210						215					220					
	Thr	Gln	Glu	Glu	Leu	Gly	Gly	Ala	Thr	Thr	His	Met	Val	Thr	Ala	Gly	
	225					230					235					240	
	Asn	Ser	His	Tyr	Thr	Ala	Ala	Thr	Asp	Glu	Glu	Ala	Leu	Asp	Trp	Val	
					245					250					255		
	Gln	Asp	Leu	Val	Ser	Phe	Leu	Pro	Ser	Asn	Asn	Arg	Ser	Tyr	Ala	Pro	
				260					265					270			
	Met	Glu	Asp	Phe	Asp	Glu	Glu	Glu	Gly	Gly	Val	Glu	Glu	Asn	Ile	Thr	
			275					280					285				
[0040]	Ala	Asp	Asp	Leu	Lys	Leu	Asp	Glu	Ile	Ile	Pro	Asp	Ser	Ala	Thr	Val	
	290						295					300					
	Pro	Tyr	Asp	Val	Arg	Asp	Val	Ile	Glu	Cys	Leu	Thr	Asp	Asp	Gly	Glu	
	305					310					315					320	
	Tyr	Leu	Glu	Ile	Gln	Ala	Asp	Arg	Ala	Glu	Asn	Val	Val	Ile	Ala	Phe	
					325					330					335		
	Gly	Arg	Ile	Glu	Gly	Gln	Ser	Val	Gly	Phe	Val	Ala	Asn	Gln	Pro	Thr	
					340				345					350			
	Gln	Phe	Ala	Gly	Cys	Leu	Asp	Ile	Asp	Ser	Ser	Glu	Lys	Ala	Ala	Arg	
			355					360					365				
	Phe	Val	Arg	Thr	Cys	Asp	Ala	Phe	Asn	Ile	Pro	Ile	Val	Met	Leu	Val	
		370					375					380					
	Asp	Val	Pro	Gly	Phe	Leu	Pro	Gly	Ala	Gly	Gln	Glu	Tyr	Gly	Gly	Ile	
	385					390					395					400	
	Leu	Arg	Arg	Gly	Ala	Lys	Leu	Leu	Tyr	Ala	Tyr	Gly	Glu	Ala	Thr	Val	
					405					410					415		
	Pro	Lys	Ile	Thr	Val	Thr	Met	Arg	Lys	Ala	Tyr	Gly	Gly	Ala	Tyr	Cys	
					420				425					430			
	Val	Met	Gly	Ser	Lys	Gly	Leu	Gly	Ser	Asp	Ile	Asn	Leu	Ala	Trp	Pro	
			435					440					445				

Thr Ala Gln Ile Ala Val Met Gly Ala Ala Gly Ala Val Gly Phe Ile
 450 455 460
 Tyr Arg Lys Glu Leu Met Ala Ala Asp Ala Lys Gly Leu Asp Thr Val
 465 470 475 480
 Ala Leu Ala Lys Ser Phe Glu Arg Glu Tyr Glu Asp His Met Leu Asn
 485 490 495
 Pro Tyr His Ala Ala Glu Arg Gly Leu Ile Asp Ala Val Ile Leu Pro
 500 505 510
 Ser Glu Thr Arg Gly Gln Ile Ser Arg Asn Leu Arg Leu Leu Lys His
 515 520 525
 Lys Asn Val Thr Arg Pro Ala Arg Lys His Gly Asn Met Pro Leu
 530 535 540

<210> 60
 <211> 1776
 <212> DNA
 <213> 谷氨酸棒状杆菌

[0041]

<400> 60
 atgtcagtcg agactcgcaa gatcaccaag gttcttgtcg ctaaccgtgg tgagattgca 60
 atccgcgtgt tccgtgcagc tcgagatgaa ggcacgcgat ctgtcgccgt ctacgcagag 120
 ccagatgcag atgcaccatt cgtgtcatat gcagacgagg cttttgcct cgggtggccaa 180
 acatccgctg agtcctacct tgtcattgac aagatcatcg atgcggcccg caagtccggc 240
 gccgacgcca tccaccccggt ctacggcttc ctgcagaaa acgtgactt cgcagaagca 300
 gtcatcaacg aaggcctgat ctggattgga cttcacctg agtccatccg ctccctcggc 360
 gacaaggtca ccgctcgcca catgcagat accccaagg ctccaatggc tectggcacc 420
 aaggaaccag taaaagacgc agcagaagtt gtggctttcg ctgaagaatt cgtctccca 480
 atgccatca aggcagcttt cgggtggcggc ggacgtggca tgaaggttgc ctacaagatg 540
 gaagaagtcg ctgaccttct cgagtccgca acccgtgaag caaccgcagc gttcggccgc 600
 ggcgagtgtc tcgtggagcg ctacctggac aaggcacgcc acgttgaggc tcaggtcatc 660
 gccgataagc acggcaacgt tgttgtcgcc ggaaccctg actgctccct gcagcgccgt 720
 ttccagaagc tcgtcgaaga agcaccagca ccattctca ccgatgacca gcgcgagcgt 780
 ctccactcct ccgcaaggc tatctgtaag gaagctggct actacggtgc aggcacggtt 840
 gagtacctcg ttggctccga cggcctgac tccttctcg aggtcaacac ccgcctccag 900
 gtggaacacc cagtcaccga agagaccacc ggcacgcacc tgggccgca aatgttccgc 960
 atcgagaag gccacgagct ctccatcaag gaagatccag ctccacgcgg ccacgcattc 1020
 gatttccgca tcaacggcga agacgctggc tccaacttca tgcctgcacc aggcaagatc 1080
 accagctacc gcgagccaca gggcccaggc gtccgcatgg actccggtgt cgttgaaggt 1140
 tccgaaatct ccgacagtt cgactccatg ctggcaaagc tgatcgtttg gggcgacacc 1200
 cgcgagcagg ctctccagcg ctcccgccgt gcacttgcag agtacgttgt cgagggcag 1260
 ccaaccgtta tccattcca ccagcacatc gtggaaaacc cagcattcgt gggcaacgac 1320

```

gaaggcttcg agatctacac caagtggatc gaagaggttt gggataaccc aatcgcacct    1380
tacgttgacg ctccgagct cgacgaagat gaggacaaga ccccagcaca gaaggttggt    1440
gtggagatca acggccgtcg cgttgagggt gcactcccag gcgatctggc actcgggtggc    1500
accgctggtc ctaagaagaa ggccaagaag cgtcgcgcag gtggtgcaaa ggctggcgta    1560
tccggcgatg cagtggcagc tccaatgcag ggcactgtca tcaaggtaa cgtcgaagaa    1620
ggcgctgaag tcaacgaagg cgacaccgtt gttgtcctcg aggctatgaa gatgaaaac    1680
cctgtgaagg ctcataagtc cggaaccgta accggcctta ctgtcgtgc aggcgagggt    1740
gtcaacaagg gcgttgttct cctcgagatc aagtaa                                1776

```

<210> 61

<211> 591

<212> PRT

<213> 谷氨酸棒状杆菌

<400> 61

```

Met Ser Val Glu Thr Arg Lys Ile Thr Lys Val Leu Val Ala Asn Arg
1          5          10          15

```

```

Gly Glu Ile Ala Ile Arg Val Phe Arg Ala Ala Arg Asp Glu Gly Ile
          20          25          30

```

```

Gly Ser Val Ala Val Tyr Ala Glu Pro Asp Ala Asp Ala Pro Phe Val
          35          40          45

```

```

Ser Tyr Ala Asp Glu Ala Phe Ala Leu Gly Gly Gln Thr Ser Ala Glu
          50          55          60

```

[0042]

```

Ser Tyr Leu Val Ile Asp Lys Ile Ile Asp Ala Ala Arg Lys Ser Gly
65          70          75          80

```

```

Ala Asp Ala Ile His Pro Gly Tyr Gly Phe Leu Ala Glu Asn Ala Asp
          85          90          95

```

```

Phe Ala Glu Ala Val Ile Asn Glu Gly Leu Ile Trp Ile Gly Pro Ser
          100          105          110

```

```

Pro Glu Ser Ile Arg Ser Leu Gly Asp Lys Val Thr Ala Arg His Ile
          115          120          125

```

```

Ala Asp Thr Ala Lys Ala Pro Met Ala Pro Gly Thr Lys Glu Pro Val
          130          135          140

```

```

Lys Asp Ala Ala Glu Val Val Ala Phe Ala Glu Glu Phe Gly Leu Pro
          145          150          155          160

```

```

Ile Ala Ile Lys Ala Ala Phe Gly Gly Gly Gly Arg Gly Met Lys Val
          165          170          175

```

```

Ala Tyr Lys Met Glu Glu Val Ala Asp Leu Phe Glu Ser Ala Thr Arg
          180          185          190

```

```

Glu Ala Thr Ala Ala Phe Gly Arg Gly Glu Cys Phe Val Glu Arg Tyr
          195          200          205

```

```

Leu Asp Lys Ala Arg His Val Glu Ala Gln Val Ile Ala Asp Lys His

```

[0043]

210	215	220
Gly Asn Val Val Val 225	Ala Gly Thr Arg Asp 230	Cys Ser Leu Gln Arg Arg 235 240
Phe Gln Lys Leu Val 245	Glu Glu Ala Pro 250	Ala Pro Phe Leu Thr Asp Asp 255
Gln Arg Glu Arg Leu His Ser Ser 260	Ala Lys Ala Ile Cys 265 270	Lys Glu Ala
Gly Tyr Tyr Gly Ala Gly Thr 275	Val Glu Tyr Leu Val 280	Gly Ser Asp Gly 285
Leu Ile Ser Phe Leu Glu Val 290 295	Asn Thr Arg Leu Gln Val 300	Glu His Pro
Val Thr Glu Glu Thr Thr Gly Ile Asp Leu Val Arg Glu Met Phe Arg 305 310 315 320		
Ile Ala Glu Gly His Glu Leu Ser Ile Lys Glu Asp Pro Ala Pro Arg 325 330 335		
Gly His Ala Phe Glu Phe Arg Ile Asn Gly Glu Asp Ala Gly Ser Asn 340 345 350		
Phe Met Pro Ala Pro Gly Lys Ile Thr Ser Tyr Arg Glu Pro Gln Gly 355 360 365		
Pro Gly Val Arg Met Asp Ser Gly Val Val Glu Gly Ser Glu Ile Ser 370 375 380		
Gly Gln Phe Asp Ser Met Leu Ala Lys Leu Ile Val Trp Gly Asp Thr 385 390 395 400		
Arg Glu Gln Ala Leu Gln Arg Ser Arg Arg Ala Leu Ala Glu Tyr Val 405 410 415		
Val Glu Gly Met Pro Thr Val Ile Pro Phe His Gln His Ile Val Glu 420 425 430		
Asn Pro Ala Phe Val Gly Asn Asp Glu Gly Phe Glu Ile Tyr Thr Lys 435 440 445		
Trp Ile Glu Glu Val Trp Asp Asn Pro Ile Ala Pro Tyr Val Asp Ala 450 455 460		
Ser Glu Leu Asp Glu Asp Glu Asp Lys Thr Pro Ala Gln Lys Val Val 465 470 475 480		
Val Glu Ile Asn Gly Arg Arg Val Glu Val Ala Leu Pro Gly Asp Leu 485 490 495		
Ala Leu Gly Gly Thr Ala Gly Pro Lys Lys Lys Ala Lys Lys Arg Arg 500 505 510		
Ala Gly Gly Ala Lys Ala Gly Val Ser Gly Asp Ala Val Ala Ala Pro		

	515	520	525	
	Met Gln Gly Thr Val Ile Lys Val Asn Val Glu Glu Gly Ala Glu Val			
	530	535	540	
	Asn Glu Gly Asp Thr Val Val Val Leu Glu Ala Met Lys Met Glu Asn			
	545	550	555	560
	Pro Val Lys Ala His Lys Ser Gly Thr Val Thr Gly Leu Thr Val Ala			
		565	570	575
	Ala Gly Glu Gly Val Asn Lys Gly Val Val Leu Leu Glu Ile Lys			
		580	585	590
	<210> 62			
	<211> 10025			
	<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> 62			
[0044]	cactatacca attgagatgg gctagtcaat gataattact agtccttttc ctttgagttg			60
	tgggtatctg taaattctgc tagacctttg ctggaaaact tgtaaattct gctagaccct			120
	ctgtaaattc cgctagacct ttgtgtgttt tttttgttta tattcaagt gttataattt			180
	atagaataaa gaaagaataa aaaaagataa aaagaataga tcccagccct gtgtataact			240
	cactacttta gtcagttccg cagtattaca aaaggatgic gcaaacgtg tttgctcctc			300
	tacaaaacag accttaaaac cctaaaggcg tcggcatccg cttacagaca agctgtgacc			360
	gtctccggga gctgcatgtg tcagaggttt tcaccgtcat caccgaaacg cgcgaggcag			420
	cagatcaatt cgcgcgcgaa ggcaagcgg catgcattta cgttgacacc atcgaatggt			480
	gcaaaacctt tcgcggtatg gcatgatagc gcccgaaga gagtcaattc aggggtgtga			540
	atgtgaaacc agtaacgtta tacgatgtcg cagagtatgc cgggtgtctt tatcagaccg			600
	tttcccgctt ggtaaccag gccagccacg tttctgcgaa aacgcgggaa aaagtgggaag			660
	cggcgatggc ggagctgaat tacattccca accgcgtggc acaacaactg gcgggcaaac			720
	agtcgttgcg gattggcggt gccacctcca gtctggccct gcacgcgcgc tcgcaaattg			780
	tcgcggcgat taaatctcgc gccgatcaac tgggtgccag cgtggtggtg tcgatggtag			840
	aacgaagcgg cgtcgaagcc tgtaaagcgg cgggtgcaca tcttctcgcg caacgcgtca			900
	gtgggctgat cattaactat ccgctggatg accaggatgc cattgctgtg gaagctgcct			960
	gcactaatgt tccggcggtta tttcttgatg tctctgacca gacacccatc aacagtatta			1020
	ttttctccca tgaagacggt acgcgactgg gcgtggagca tctggtcgca ttgggtcacc			1080
	agcaaategc gctgttagcg ggccatttaa gttctgtctc ggcggtctg cgtctggctg			1140
	gctggcataa atatctcact cgcaatcaa ttcagccgat agcggaaacgg gaagcgact			1200
	ggagtgccat gtccggtttt caacaaacca tgcaaatgct gaatgagggc atcgttccca			1260
	ctgcgatgct ggttgccaac gatcagatgg cgtggggcgc aatgcgcgcc attaccgagt			1320
	ccgggctgcg cgttggtgcg gatattctcg tagtgggata cgacgatacc gaagacagct			1380
	catgttatat ccgcgcgtta accaccatca aacaggattt tcgcctgctg gggcaaacca			1440

[0045]	gcgtggaccg cttgtgcaa ctctctcagg gccaggcggg gaagggaat cagctgttgc	1500
	ccgtctcact ggtgaaaaga aaaaccaccc tggcgcccaa tacgcaaacc gcctctcccc	1560
	gcgcgttggc cgattcatta atgcagctgg cagcacaggt ttcccgactg gaaagcgggc	1620
	agtgagcgca acgcaattaa tgtaagttag cgcgaattga tctggtttga cagcttatca	1680
	tcgactgcac ggtgcaccaa tgcttctggc gtcaggcagc catcggaagc tgtggtatgg	1740
	ctgtgcaggt cgtaaatcac tgcataattc gtgtcgtca aggcgcactc ccgttctgga	1800
	taatgttttt tgcgccgaca tcataacggt tctggcaaat attctgttga caattaatca	1860
	tccggctcgt ataaagtgtg gaattgtgag cggataacaa tttcacacag gaaacagcgc	1920
	cgctgagaaa aagcgaagcg gcaactgctt ttaacaattt atcagacaat ctgtgtgggc	1980
	actcgaccgg aattatcgat taactttatt attaaaaatt aaagaggtat atattaatgt	2040
	atcgattaaa taaggaggaa taaacatga ccatttcctc accttgatt gacgtcgcca	2100
	accttcaga catcaacacc actgccggca agatcgccga ccttaaggct cgccgcgcgg	2160
	aagccattt ccccatgggt gaaaaggcag tagagaaggt ccacgctgct ggacgcctca	2220
	ctgccgtga gcgcttgat tacttactcg atgagggtc cttcatcgag accgatcagc	2280
	tggctcgcca ccgaccacc gcttccggc tgggcgctaa gcgtcctgca accgacggca	2340
	tcgtgaccgg ctggggcacc attgatggac gcgaagtctg catcttctcg caggacggca	2400
	ccgtattcgg tggcgcgctt ggtgaggtgt acggcgaaaa gatgatcaag atcatggagc	2460
	tggcaatcga caccggccgc ccattgatcg gtctttacga aggcgctggc gctcgtatc	2520
	aggacggcgc tgctccctg gacttcattt ccagacctt ctacaaaaac attcaggctt	2580
	ctggcggtat cccacagatc tccgtcatca tggcgcatg tgcaggtggc aacgcttacg	2640
	gcccagctct gaccgacttc gtggtcatgg tggacaagac ctccaagatg ttcgttaccg	2700
	gcccagacgt gatcaagacc gtcaccggcg aggaaatcac ccaggaagag cttggcggag	2760
	caaccacca catggtgacc gctggttaact ccactacac cgctgcgacc gatgaggaag	2820
	cactggattg ggtacaggac ctggtgtcct tcttcccatc caacaatcgc tctacgcac	2880
	cgatggaaga cttcgacgag gaagaaggcg gcgttgaaga aaacatcacc gctgacgac	2940
	tgaagctcga cgagatcatc ccagattccg cgaccgttcc ttacgacgtc cgcgatgtca	3000
	tcgaatgcct caccgacgat ggcaaatacc tggaaatcca ggcagaccgc gcagaaaacg	3060
	ttgttattgc attcgccgc atcgaaggcc agtccgttgg ctttgttgcc aaccagccaa	3120
	cccagttcgc tggctgcctg gacatcgact cctctgagaa ggcagctcgc ttcgtccgca	3180
	cctgcgacgc gttcaacatc ccaatcgta tgcctgtcga cgtccccggc ttcctccag	3240
	gcgcaggcca ggagtacggt ggcattctgc gtcgtggcgc aaagctgctc tacgcatacg	3300
	gcgaagcaac cgttccaaag atcaccgtca ccatgcgtaa ggcttacggc ggagcgtact	3360
	gcgtgatggg ttccaaggcg ttgggctctg acatcaacct tgcattggcca accgcacaga	3420
	tcgccgtcat ggcgctgct ggcgcagttg gattcatcta ccgcaaggag ctcatggcag	3480
	ctgatgcaa gggcctcgat accgtagctc tggctaagtc cttcgagcgc gagtatgaag	3540
	accacatgct caaccctgac cagctgcag aacgtggcct gatcgacgcc gtgatcctgc	3600
	caagcgaac ccgcgacag atttcccga accttcgctt gctcaagcac aagaacgtca	3660
	ctgcacctgc tcgaagcac ggcaacatgc cactgtaagg aggaaaacta aatgtcagtc	3720

	gagactcgca agatcaccaa ggttcttgtc gctaaccgtg gtgagattgc aatccgctg	3780
	ttccgtgcag ctcgagatga aggcacgcga tctgtcgccg tctacgcaga gccagatgca	3840
	gatgcaccat tctgttcata tgcagacgag gcttttgccc tcggtggcca aacatccgt	3900
	gagtcttacc ttgtcattga caagatcatc gatgcggccc gcaagtccgg cgccgacgcc	3960
	atccaccccg gctacggctt cctcgagaa aacgtgact tcgcagaagc agtcatcaac	4020
	gaaggcctga tctggattgg accttcacct gattccatcc gtcctcctgg cgacaaggtc	4080
	accgtctgcc acatcgaga taccgccaag gctccaatgg ctcttgccac caaggaacca	4140
	gtaaaagacg cagcagaagt tgtggctttc gctgaagaat tcggtctccc aatcgccatc	4200
	aaggcagctt tcggtggcgg cggacgtggc atgaaggttg cctacaagat ggaagaagtc	4260
	gctgacctct tcgagtcgcg aaccctgaa gcaaccgcag cgttcggccg cggcgagtgc	4320
	ttctgtggagc gctacctgga caaggcacgc cactgtgagg ctgagtcac cgccgataag	4380
	cacggcaacg ttgttgcgc cggaaccgt gactgctccc tgcagcgccg ttccagaag	4440
	ctcgtcgaag aagcaccagc accattcctc accgatgacc agcgcgagcg tctccactcc	4500
	tccgcaag ctatctgtaa ggaagctggc tactacggtg caggcaccgt tgagtacctc	4560
	gttggtccg acggcctgat ctcttctc gaggtcaaca cccgctcca ggtggaacac	4620
	ccagtcaccg aagagaccac cggcatcgac ctggtccgag aaatgttccg catcgagaa	4680
	ggccacgagc tctccatcaa ggaagatcca gctccacgcg gccacgcatt cgagttccgc	4740
	atcaacggcg aagacgtgg ctccaacttc atgcctgcac caggcaagat caccagctac	4800
[0046]	cgcgagccac agggcccagg cgtccgcatg gactccggtg tcgttgaagg ttccgaaatc	4860
	tccgacagt tcgactccat gctggcaaa ctgatcttt gggcgacac ccgcgagcag	4920
	gctctccagc gctcccgccg tgcacttgca gactacgttg tcgagggcat gccaaccgtt	4980
	atccattcc accagcacat cgtggaaaac ccagattcg tgggcaacga cgaaggcttc	5040
	gagatctaca ccaagtggat cgaagaggtt tgggataacc caatcgacc ttacgttgac	5100
	gcttccgagc tcgacgaaga taggacaag accccagcac agaaggttgt tgtggagatc	5160
	aacggccgctc gcttgagggt tgcactccca ggcgatctgg cactcggtgg caccgtggt	5220
	cctaagaaga agccaagaa gcgtcgcga ggtggtgcaa aggctggcgt atccgcgat	5280
	gcagtggcag ctccaatgca gggcactgtc atcaaggta acgtcgaaga aggcgtgaa	5340
	gtcaacgaag gcgacaccgt tttgtctc gaggctatga agatggaaaa cctgtgaag	5400
	gctcataagt ccggaaccgt aaccggcctt actgtcctg caggcgaggg tgtcaacaag	5460
	ggcgttgctt tctcgagat caagtaatct agaggaggaa aactaaatga atgttgacat	5520
	tagccgtct cgtgaaccgt tgaacgtgga actgttgaaa gaaaaactgc tgcagaacgg	5580
	tgatttcggt caagtgatct acgagaaggt caccggtct accaatcgcg acctgctggc	5640
	tctggcgggc agcggcgtc caaactggac cgtcaagact gttgaatttc aggaccacgc	5700
	ccgtggccgt ctgggtcgtc cgtggagcgc accggagggt tcccaacca tcgtcagct	5760
	tctggtccaa ctgagcattg atcagggtga ccgtattggt acgatcccgc tggccgagg	5820
	cttggtgtt atggatgcgc tgaatgatct gggcgtggag ggtgcaggcc tgaatggcc	5880
	gaacgatgtt cagatccacg gtaagaagtt gtgcggtatt ctggttgaag caaccggctt	5940
	cgactccact ccgaccgtgg ttatcggtg ggtacgaat atctcgtga cgaagaaga	6000

[0047]

```

gctgccggtc ccgcacgcga ccagcctggc cctggagggt gttgaagttg accgtacgac 6060
gttcctgatt aacatgctga cccatctgca taccctctg gatcagtggc agggtcctgc 6120
tgtggactgg ctggatgact atcgcgcggt ttgtagcagc attggccaag atgtgcgtgt 6180
cctgtgcctt ggtgacaaag agctgtggg cgaggcgatt ggcgtggcga ccggtggtga 6240
gatccgtgtg cgcgacgcca gcggcacggt ccacacgctg aatgcgggtg aaatcacgca 6300
tctgcgtttg caataaaagc ttgtttaaac ggtctccagc ttggctgttt tggcggatga 6360
gagaagattt tcagcctgat acagattaaa tcagaacgca gaagcgtctt gataaaacag 6420
aatgtgctg gcggcagtag cgcggtggc ccacctgacc ccatgccgaa ctgagaagt 6480
aaacgccga gcgccatgg tagtgtggg tctcccatg cgagagtagg gaactgccag 6540
gcatcaaata aaacgaaagg ctgagtcgaa agactgggccc ttctgtttta tctgttgttt 6600
gtcgggtgaac gctctctga cgctgatgc ggtattttct ccttacgcat ctgtgcggta 6660
tttcacaccg catatgggac actctcagta caatctgctc tgatgccgca tagttaagcc 6720
agccccgaca cccgccaaca cccgtgacg agcttagtaa agccctcgt agattttaat 6780
gcggatgttg cgattacttc gccaaactatt gcgataacaa gaaaagcca gcctttcatg 6840
atatatctcc caatttgtgt agggcttatt atgcacgctt aaaaataata aaagcagact 6900
tgacctgata gtttggctgt gagcaattat gtgcttagtg catctaacgc ttgagttaag 6960
ccgcgccgag aagcggcgtc ggcttgaacg aattgttaga cattatttgc cgactacctt 7020
ggtgatctcg cctttcacgt agtgacaaa ttcttcaac tgatctgcgc gcgaggccaa 7080
gcgactttct tctgttccaa gataagcctg tctagcttca agtatgacgg gctgatactg 7140
ggccggcagg cgctccattg cccagtcggc agcgacatcc ttcggcgcga ttttgccggt 7200
tactgcgctg taccaaatgc gggacaacgt aagcactaca ttctgctcat cgccagccca 7260
gtcggggcgc gatttccata gcgttaaggt ttcathtagc gcctcaaata gatcctgttc 7320
aggaaccgga tcaaagagtt cctccgcgcg tggacctacc aaggcaacgc tatgttctct 7380
tgcttttgtc agcaagatag ccagatcaat gtcgatcgtg gctggctcga agatacctgc 7440
aagaatgtca ttgcgtgccc attctccaaa ttgcagttcg cgcttagctg gataacgcca 7500
cggaatgatg tcgtcgtgca caacaatggt gacttctaca gcgcggagaa tctcgtctc 7560
tccaggggaa gccgaagttt ccaaagggtc gttgatcaaa gctcgccgag ttgtttcatc 7620
aagccttacg gtcaccgtaa ccagcaaatc aatatcactg tgggcttca ggccgccatc 7680
cactgcggag ccgtacaaat gtacggccag caacgtcggg tcgagatggc gctcgatgac 7740
gccaaactacc tctgatagtt gattcgatac ttcggcgac accgcttccc tcatgatgtt 7800
taactttgtt ttagggcgac tgccctgctg cgtaacatcg ttgctgtccc ataacatcaa 7860
acatcgacc accggcgtac gcgcttgctg cttggatgcc cgaggcatag actgtacccc 7920
aaaaaaacag tcataacaag ccatgaaaac cgccactgag ccgttaccac cgtgcgttc 7980
ggtcaaggtt ctggaccagt tgcgtgagcg catacgtac ttgcattaca gttacgaac 8040
cgaacaggct tatgtccact gggttcgtgc cttcatccgt ttccacggtg tgcgtacccc 8100
ggcaaccttg ggcagcagcg aagtcgagc atttctgtcc tggctggcga acgagcga 8160
ggtttcggtc tccacgcatc gtcaggcatt ggcgcccttg ctgttcttct acggcaaggt 8220
gctgtgcacg gatctgccct ggcttcagga gatcggaaga cctcggccgt cgcggcgctt 8280

```

	gccggtggtg ctgaccccg atgaagtgg tgcacccctc ggttttctgg aaggcgagca	8340
	tcgtttgttc gccagcttc tgtatggaac gggcatgcgg atcagtgagg gtttgaact	8400
	gcgggtcaag gatctggatt tcgacacgg cagcatcatc gtgcgggagg gcaagggtc	8460
	caaggatcgg gccttgatgt taccgagag cttggcacc agcctgcgcg agcaggggaa	8520
	ttaattccca cgggttttgc tgcccgaac cgggctgttc tgggtttgct agtttgttat	8580
	cagaatcgca gatccggctt cagccggttt gccggctgaa agcgtatatt ctccagaat	8640
	tgccatgatt ttttccccc gggaggcgtc actggctccc gtgtgtcgg cagcttgat	8700
	tcgataagca gcacgcctg tttcaggctg tctatgtgtg actgttgagc tgtaacaagt	8760
	tgctcagggt gttcaatttc atgttctagt tgccttgttt tactgtttc acctgttcta	8820
	ttaggtgta catgctgttc atctgttaca ttgtcgatct gttcatggtg aacagctttg	8880
	aatgcaccaa aaactcgtaa aagctctgat gtatctatct tttttacacc gttttcatct	8940
	gtgcatatgg acagttttcc ctttgatatg taacggtgaa cagttgttct acttttgttt	9000
	gttagtcttg atgcttact gatagataca agagccataa gaacctcaga tccttcgta	9060
	tttagccagt atgttctcta gtgtggttcg ttgtttttgc gtgagccatg agaacgaacc	9120
	attgagatca tacttacttt gcatgtcact caaaaatttt gcctcaaac tggtagctg	9180
	aatttttga gttaaagcat cgtgtagtgt ttttcttagt ccgttatgta ggtaggaatc	9240
	tgatgtaatg gttgttgga tttgtgcacc attcattttt atctggttgt tctcaagttc	9300
	ggttacgaga tccatttgc tatctagttc aacttgaaa atcaacgtat cagtcgggcg	9360
[0048]	gcctcgctta tcaaccacca atttcatatt gctgtaagtg tttaaatctt tacttattgg	9420
	tttcaaaacc cattggttaa gcctttttaa ctcatggtag ttattttcaa gcattaacat	9480
	gaacttaaat tcatcaaggc taatctctat atttgcttg tgagttttct ttgtgttag	9540
	ttcttttaat aaccactcat aaatctcat agagtatttg ttttcaaaag acttaacatg	9600
	ttccagatta tattttatga atttttttaa ctggaaaaga taaggcaata tctcttact	9660
	aaaaactaat tctaattttt cgcttgagaa cttggcatag tttgtccact ggaaaatctc	9720
	aaagccttta accaaaggat tctgatttc cacagtctc gtcacagct ctctggttgc	9780
	tttagctaat acaccataag cattttccct actgatgttc atcatctgag cgtattggtt	9840
	ataagtgaac gataccgtcc gttctttcct tgtagggttt tcaatcgtgg ggttgagtag	9900
	tgccacacag cataaaatta gcttggttc atgtccgtt aagtcatagc gactaatcgc	9960
	tagttcattt gctttgaaaa caactaatc agacatacat ctcaattggt ctaggtgatt	10020
	ttaat	10025
	<210> 63	
	<211> 11469	
	<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> 63	
	cactatacca attgagatgg gctagtcaat gataattact agtccttttc ctttgagttg	60
	tggttatctg taaattctgc tagaccttg ctggaaaact tgtaaattct getagacct	120

[0049]	ctgtaaattc cgctagacct ttgtgtgttt tttttgttta tattcaagtg gttataattt	180
	atagaataaa gaaagaataa aaaaagataa aaagaataga tcccagccct gtgtataact	240
	cactacttta gtcatgtccg cagtattaca aaaggatgtc gcaaacgctg ttigtctctc	300
	tacaaaacag accttaaaac cctaaaggcg tcggcatccg cttacagaca agctgtgacc	360
	gtctccggga gtgcatgtg tcagaggttt tcaccgtcat caccgaaacg cgcgaggcag	420
	cagatcaatt cgcgcgcgaa ggcgagcgg catgcattta cgttgacacc atcgaatggt	480
	gcaaaacctt tcgcggtatg gcatgatagc gcccggaaga gagtcaattc aggggtgtga	540
	atgtgaaacc agtaacgtta tacgatgtcg cagagtatgc cgggtgtctct tatcagaccg	600
	tttcccgcgt ggtgaaccag gccagccacg tttctgcgaa aacgcgggaa aaagtggaag	660
	cggcgatggc ggagctgaat tacattccca accgcgtggc acaacaactg gcgggcaaac	720
	agtcgttgc tttgtgcgtt gccacctcca gtctggccct gcacgcgccg tcgcaaattg	780
	tcgcggcgat taaatctcgc gccgatcaac tgggtgccag cgtgggtgtg tcgatggtag	840
	aacgaagcgg cgtcgaagcc tgtaaagcgg cgggtgcaca tcttctcgcg caacgcgtca	900
	gtgggctgat cattaactat ccgctggatg accaggatgc cattgctgtg gaagctgcct	960
	gcactaatgt tccggcgta tttcttgatg tctctgacca gacaccatc aacagtatta	1020
	ttttctccca tgaagacggt acgcgactgg gcgtggagca tctggtcgca ttgggtcacc	1080
	agcaaatcgc gctgttagcg ggccattaa gttctgtctc ggcgctctg cgtctggctg	1140
	gctggcataa atatctcact cgcaatcaaa ttcagccgat agcggaacgg gaaggcgact	1200
	ggagtgccat gtccggtttt caacaaacca tgcaaatgct gaatgagggc atcgttccca	1260
	ctgcgatgct ggttgccaac gatcagatgg cgtgggcgc aatgcgcgcc attaccgagt	1320
	ccgggctcgc cgttggtcgc gatatctcgg tagtgggata cgacgatacc gaagacagct	1380
	catgttatat cccgccgtta accaccatca aacaggattt tcgctgctg gggcaaacca	1440
	gcgtggaccg cttgctgcaa ctctctcagg gccaggcggg gaagggaat cagctgttgc	1500
	ccgtctcact ggtgaaaaga aaaaccaccc tggcgcccaa tacgcaaacc gcctctcccc	1560
	gcgcgttgcc cgattcatta atgcagctgg cagcacaggt ttcccactg gaaagcgggc	1620
	agtgagcgca acgcaattaa tgtaagttag cggaattga tctggtttga cagcttatca	1680
	tcgactgcac ggtgcaccaa tgcttctggc gtcaggcagc catcggaagc tgtggtatgg	1740
	ctgtgcaggt cgtaaatcac tcataattc gtgtcgtca aggcgcactc ccgttctgga	1800
	taatgttttt tgcgccgaca tcataacggt tctggcaaat attctgaaat gagctgttga	1860
	caattaatca tccggtcgt ataattgtgt gaattgtgag cggataacaa ttacacacag	1920
	gaaacagcgc cgtgagaaa aagcgaagcg gcaactgtct ttaacaattt atcagacaat	1980
	ctgtgtgggc actcgaccgg aattatcgat taactttatt attaaaaatt aaagaggtat	2040
	atattaatgt atcgattaaa taaggaggaa taaacatga aacgtctcgg aaccctggac	2100
	gcctctctgc tggcggttga atctgaagac accccgatgc atgtgggtac gttcagatt	2160
	ttctcactgc cggaaggcgc accagaaacc ttctgcgtg acatggtcac tcgaatgaaa	2220
	gaggeccggc atgtggcacc accctgggga tacaaactgg cctgggtctg tttcctcggg	2280
	cgcgtgatgc ccccgccctg gaaagtcgat aaggatatcg atctggatta tcacgtccgg	2340
	cactcagccc tgctcgcgcc cggcggggag cgcgaactgg gtattctggt atcccactg	2400

	cactctaacc ccctggattt ttcccgcctt ctttggaat gccacgttat tgaaggcctg	2460
	gagaataaacc gttttgccct ttacacaaa atgcaccact cgatgattga cggcatcagc	2520
	ggcgtgcgac tgatgcagag ggtgctcacc accgatcccg aacgctgcaa tatgccaccg	2580
	ccctggacgg tacgcccaca ccaacgccgt ggtgcaaaaa ccgacaaaga ggccagcgtg	2640
	cccgacggg tttccaggc aatggacgcc ctgaagctcc aggagacat ggccccagg	2700
	ctgtggcagg ccggcaatcg cctggtgcat tcggttcgac acccggaaga cggactgacc	2760
	gcgcccttca ctggaccggt ttcggtgctc aatcaccggg ttaccgcgca gcgacgtttt	2820
	gccaccagc attatcaact ggaccggctg aaaaacctgg cccatgcttc cggcggttcc	2880
	ttgaacgaca tcgtgcttta cctgtgtggc accgcattgc ggcgctttct ggctgagcag	2940
	aacaatctgc cagacacccc gctgacggct ggtataccgg tgaatatccg gccggcagac	3000
	gacgagggtg cgggcaccca gatcagtttt atgattgcct cgctggccac cgacgaagct	3060
	gatccgttga accgcctgca acagatcaaa acctcgacc gacgggcca ggagcacctg	3120
	cagaaacttc caaaaagtgc cctgaccag tacacatgc tgcctgatgc accctacatt	3180
	ctgcaattga tgcaggtct cggggggagg atcgaccag tcttcaactg gaccatttcc	3240
	aacgtcccc gcccggaagg cacgctgtat tatgaaggag cccggttga ggccatgtat	3300
	ccggtatcgc taatcgtca cggcggcgcc ctgaacatca cctgcctgag ctatgccga	3360
	tcgtgaatt tcggttttac cgctgtcgg gatacgtgc cgagcatgca gaaactggcg	3420
	gtttataccg gtgaagctct ggatgagctg gaatcgtga ttctgccacc caagaagcgc	3480
[0050]	gcccgaacc gcaagtaact cgaggaggaa aactaatga ccatttctc acctttgatt	3540
	gacgtcgcca accttcaga catcaacacc actgccgga agatcgcca ccttaaggct	3600
	cgcgcgcgg aagcccattt cccatgggt gaaaaggcag tagagaaggt ccacgtgct	3660
	ggacgcctca ctgcccgtga gcgcttgat tacttactcg atgagggtc cttcatcgag	3720
	accgatcagc tggtcgcca ccgaccacc gtttcggcc tgggcgctaa gcgtcctgca	3780
	accgacgga tcgtgaccgg ctggggcacc attgatggac gcgaagtctg catcttctcg	3840
	caggacgga ccgtattcgg tggcgcgtt ggtgaggtgt acggcgaata gatgatcaag	3900
	atcatggagc tgcaatcga caccggccgc ccattgatcg gtctttacga aggcgctggc	3960
	gtcgtattc aggacggcgc tgtctcctg gacttcattt cccagacctt ctacaaaac	4020
	attcaggctt ctggcggtat cccacagatc tccgtcatca tgggcgcatg tgcagtggc	4080
	aacgcttacg gccagctct gaccgacttc gtggtcatgg tggacaagac ctccaagatg	4140
	ttcgttacg gccagacgt gatcaagacc gtcaccggcg aggaaatcac ccaggaagag	4200
	cttgccggag caaccaccca catggtgacc gctggttaact cccactacac cgtgacgacc	4260
	gatgaggaag cactggattg ggtacaggac ctggtgtcct tcttccatc caacaatcgc	4320
	tcctacgcac cgatgaaga cttegacgag gaagaaggcg gcgttgaaga aaacatcacc	4380
	gctgacgac tgaagctcga cgagatcatc ccagattccg cgaccgttcc ttacgacgtc	4440
	cgcgatgtca tcgaatgcct caccgacgat ggcaataacc tggaaatcca ggcagaccgc	4500
	gcagaaaacg ttgttattgc attcggccgc atcgaaggcc agtccgttgg ctttgttgcc	4560
	aaccagccaa cccagttcgc tggctgcctg gacatcgact cctctgagaa ggcagctcgc	4620
	ttcgtccgca cctgcgacgc gttcaacatc ccaatcgtca tgcctgtcga cgtccccggc	4680

[0051]

```

ttctcccag ggcagggcca ggagtacggt ggcatctctc gtcgtggcgc aaagctgctc 4740
tacgcatacg gcgaagcaac cgttccaaag atcacctgca ccatgcgtaa ggcttacggc 4800
ggagcgtact gcgtgatggg ttccaagggc ttgggctctg acatcaacct tgcattggcca 4860
accgcacaga tcgccgtcat gggcgctgct ggcgagttg gattcatcta ccgaaggag 4920
ctcatggcag ctgatgccaa gggcctcgat accgtagctc tggctaagtc cttcgagcgc 4980
gagtatgaag accacatgct caaccgtac cacgtgcag aacgtggcct gatcgacgcc 5040
gtgatctgc caagcgaac ccgcggacag atttcccgca accttcgctt gctcaagcac 5100
aagaacgtca ctgcctctgc tcgaagcac ggcaacatgc cactgtaagg aggaaaacta 5160
aatgtcagtc gagactcga agatcaccaa ggttcttctc gtaaccgtg gtgagattgc 5220
aatccgcgtg ttccgtgcag ctcgagatga aggcacgga tctgtcgccg tctacgcaga 5280
gccagatgca gatgcacat tcgtgtcata tgcagacgag gcttttgccc tcggtggcca 5340
aacatccgct gagtcttacc ttgtcattga caagatcatc gatgcggccc gcaagtccgg 5400
cgccgacgcc atccaccccg gctacggctt cctcgcagaa aacgtgact tcgcagaagc 5460
agtcataac gaaggcctga tctggattgg accttcacct gattccatcc gctccctcgg 5520
cgacaagtc accgctcgcc acatcgaga taccgcaag gctccaatgg ctcttggcac 5580
caaggaacca gtaaaagacg cagcagaagt tgtggctttc gctgaagaat tcggtctccc 5640
aatcgccatc aaggcagctt tcggtggcgg cggacgtggc atgaaggttg cctacaagat 5700
ggaagaagtc gctgacctct tcgagtcgcg aaccctgtaa gcaaccgcag cgttcggccc 5760
cgcgagtgct ttcttgagc gctacctgga caaggcacgc cacgttgagg ctgaggtcat 5820
cgccgataag cagggcaacg ttgttctcgc cggaaccctg gactgctccc tgcagcgccc 5880
tttcagaag ctctcgaaag aagcaccagc accattcttc accgatgacc agcgcgagcg 5940
tctccactcc tccggaagg ctatctgtaa ggaagctggc tactacggtg caggcaccgt 6000
tgagtacctc gttggctcgc acggcctgat ctcttcttc gaggtcaaca cccgctcca 6060
ggtggaacac ccagtcaccg aagagaccac cggcatcgac ctggtccgcg aaatgttccg 6120
catcgagaa ggccacgagc tctccatcaa ggaagatcca gctccacgcg gccacgcatt 6180
cgagttccgc atcaacggcg aagacgtgg ctccaacttc atgcctgcac caggcaagat 6240
caccagctac cgcgagccac agggcccagg cgtccgcatg gactccggtg tcgttgaagg 6300
ttccgaaatc tccgacagt tcgactccat gctggcaaag ctgatcgttt gggcgacac 6360
ccgcgagcag gctctccagc gctcccgcg tgcacttgca gattacgttg tcgagggcat 6420
gccaacctgt atccatttc accagcacat cgtggaaaac ccagcattcg tgggcaacga 6480
cgaaggttc gagatctaca ccaagtggat cgaagaggtt tgggataacc caatcgacc 6540
ttacgttgac gcttccgagc tcgacgaaga tggagacaag accccagcac agaaggttgt 6600
tgtggagatc aacggccgtc gcgttgaggt tgcaactcca ggcgatctgg cactcggtgg 6660
caccgctggt cctaagaaga aggccaagaa gcgtcgcgca ggtggtgcaa aggttgcgct 6720
atccggcgat gcagtggcag ctccaatgca gggcactgtc atcaaggtca acgtcgaaga 6780
aggcgctgaa gtcaacgaag gcgacaccgt tgttgcctc gaggttatga agatggaaaa 6840
ccctgtgaag gctcataagt ccggaaccgt aaccggcctt actgtcgctg caggcgaggg 6900
tgtcaacaag ggcgttgttc tctcgagat caagtaatct agaggaggaa aactaaatga 6960

```

	atgttgacat tagccgctct cgtgaaccgt tgaacgtgga actgttgaaa gaaaaactgc	7020
	tgcagaacgg tgatttcggt caagtgatct acgagaaggt caccggctct accaatcgcg	7080
	acctgctggc tctggcgggc agcggcgctc caaactggac cgtaagact gttgaatttc	7140
	aggaccacgc ccgtggccgt ctgggtcgtc cgtggagcgc accggagggt tcccaaacca	7200
	tcgtcagcgt tctggtccaa ctgagcattg atcaggtgga ccgtattggt acgatcccgc	7260
	tggccgcagg cttggctgtt atggatgcgc tgaatgatct gggcgtggag ggtgcaggcc	7320
	tgaatggcc gaacgatgtt cagatccacg gtaagaagtt gtgcgttatt ctggttgaag	7380
	caaccggctt cgactccact ccgaccgtgg ttatcggttg ggttacgaat atctcgttga	7440
	cgaagaaga gctgccggtc ccgcacgcga ccagcctggc cctggagggt gttgaagttg	7500
	accgtacgac gttcctgatt aacatgctga cccatctgca taccgctctg gatcagtggc	7560
	agggtccgtc tgtggactgg ctggaigact atcgcgcggt ttgtagcagc attggccaag	7620
	atgtgcgtgt cctgctgcct ggtgacaaag agctgctggg cgaggcgatt ggcgtggcga	7680
	ccggtggtga gatccgtgtg cgcgacgcca gcggcacggt ccacacgctg aatgcgggtg	7740
	aaatcacgca tctgcgtttg caataagttt aaacggtctc cagcttggtt gttttggcgg	7800
	atgagagaag attttcagcc tgatacagat taaatcagaa cgcagaagcg gtctgataaa	7860
	acagaatttg cctggcgga gtagcgggt ggtcccacct gaccccatgc cgaactcaga	7920
	agtgaacgc cgtagcgccg atggtagtgt ggggtctccc catgcgagag tagggaactg	7980
	ccaggcatca aataaaacga aaggctcagt cgaagactg ggcctttcgt ttatctgtt	8040
[0052]	gtttgtcgtt gaacgctctc ctgacgcctg atgcggtatt ttctccttac gcatctgtgc	8100
	ggtatttcac accgcatatg gtgcactctc agtacaatct gctctgatgc cgcatagtta	8160
	agccagcccc gacaccgcgc aacaccgct gacgagctta gtaaagccct cgtagatatt	8220
	taatgcggat gttgcgatta cttcgccaac tattgcgata acaagaaaaa gccagccttt	8280
	catgatatat ctcccaattt gtgtagggtt tattatgcac gcttaaaaat aataaaagca	8340
	gacttgacct gatagtgttg ctgtgagcaa ttatgtgctt agtgcacta acgcttgagt	8400
	taagccgcgc cgcaagcgg cgtcggttg aacgaattgt tagacattat ttgccgacta	8460
	ccttggtgat ctgcctttc acgtagtgga caaattcttc caactgatct gcgcgcgagg	8520
	ccaagcgate ttcttctgtt ccaagataag cctgtctagc ttcaagtatg acgggctgat	8580
	actgggccgg caggcgctcc attgccagt cggcagcgac atccttcggc gcgattttgc	8640
	cggttactgc gctgtacaa atgcgggaca acgtaagcac tacatttcgc tcatcgccag	8700
	cccagtcggg cggcgagttc catagcgtaa aggtttcatt tagcgctca aatagatcct	8760
	gttcaggaac cggatcaaag agttctctcg ccgtggacc taccaaggca acgctatgtt	8820
	ctcttgcttt tgtcagcaag atagccagat caatgtcgat cgtggctggc tcgaagatac	8880
	ctgcaagaat gtcattgcgc tgccattctc caaattgcag ttcgcgctta gctggataac	8940
	gccacggaat gatgtcgtcg tgcacaacaa tggtgacttc tacagcgcg agaattctgc	9000
	tcctccagg ggaagccgaa gtttccaaa ggtcgttgat caaagctcgc cgcgttgttt	9060
	catcaagcct tacggtcacc gtaaccagca aatcaatata actgtgtggc ttcaggccgc	9120
	catccactgc ggagccgtac aaatgtacgg ccagcaacgt cggttcgaga tggcgctcga	9180
	tgacccaac tacctctgat agttgagtcg atacttcggc gatcacgct tcctcatga	9240

	tgtttaactt tgtttttagg cgactgccct gctgcgtaac atcgttgctg ctccataaca	9300
	tcaaacatcg acccacggcg taacgcgctt gctgcttgga tgcccaggcg atagactgta	9360
	ccccaaaaa acagtcataa caagccatga aaaccgccac tgcgccgtta ccaccgctgc	9420
	gttcggtcaa ggttctggac cagttgcgtg agcgcatacg ctacttgcat tacagcttac	9480
	gaaccgaaca ggcttatgtc cactgggttc gtgccttcat ccgtttccac ggtgtgcgtc	9540
	accgcgaac cttgggcagc agcgaagtcg aggcatttct gtccctggctg gcgaacgagc	9600
	gcaaggttcc ggtctccacg catcgtcagg cattggcggc cttgctgttc ttctacggca	9660
	aggtgctgtg cacggatctg ccttggttc aggagatcgg aagacctcgg ccgtcgcggc	9720
	gcttgccggt ggtgctgacc ccgatgaag tggttcgcac cctcggtttt ctggaaggcg	9780
	agcatcgttt gtccgccag ctctgtatg gaacgggcat gcggatcagt gagggtttgc	9840
	aactgcgggt caagatctg gatttcgac acggcacgat catcgtgcgg gagggcaagg	9900
	gtccaagga tcgggccttg atgttaccgc agagcttggc acccagcctg cgcgagcagg	9960
	ggaattaatt cccacgggtt ttgctgccc caaacgggct gttctggtgt tgctagtgtt	10020
	ttatcagaat cgcagatccg gttcagccg gtttgccgcg tgaagcgct atttcttcca	10080
	gaattgcat gattttttcc ccacgggagg cgtcactggc tcccgtgttg tcggcagctt	10140
	tgattcgata agcagcatcg cctgtttcag gctgtctatg tgtgactgtt gagctgtaac	10200
	aagttgtctc aggtgttcaa ttcatgttc tagttgcttt gttttactgg ttccacctgt	10260
	tctattaggt gttacatgct gttcatctgt tacattgtcg atctgttcat ggtgaacagc	10320
[0053]	tttgaatgca caaaaactc gtaaaagctc tgatgtatct atctttttta caccgttttc	10380
	atctgtgcat atggacagtt ttccctttga tatgtaacgg tgaacagttg ttctactttt	10440
	gtttgttagt cttagtgctt cactgataga tacaagagcc ataagaacct cagatccttc	10500
	cgtatttagc cagtatgttc tctagtgtgg ttcggtgttt ttgcgtgagc catgagaacg	10560
	aaccattgag atcatactta ctttgcatgt cactcaaaaa ttttgctca aaactggtga	10620
	gctgaatttt tgcagttaaa gcatcgtgta gtgtttttct tagtccgta ttaggttagg	10680
	aatctgatgt aatggttgtt ggtatttgtt caccattcat ttttatctgg ttgttctcaa	10740
	gttcggttac gagatccatt tgtctatcta gtccaacttg gaaaatcaac gtatcagtcg	10800
	ggcggcctcg ctatcaacc accaatttca tattgctgta agtggttaaa tctttactta	10860
	ttggtttcaa aaccatttgg ttaagccttt taaactcatg gtagttattt tcaagcatta	10920
	acatgaactt aaattcatca aggctaactt ctatatttgc cttgtgagtt ttcttttgtg	10980
	ttagttcttt taataaccac tcataaatcc tcatagagta ttgtttttca aaagacttaa	11040
	catgttccag attatatatt atgaattttt ttaactggaa aagataaggc aatatctctt	11100
	cactaaaaac taattctaatt ttttcgcttg agaacttggc atagtttgtc cactggaaaa	11160
	tctcaaagcc ttaacaaaa ggattcctga ttccacagt tctcgtcacc agctctctgg	11220
	ttgcttttag taatacacca taagcatttt ccctactgat gttcatcacc tgagcgtatt	11280
	ggttataagt gaacgatacc gtccgttctt tccttgtagg gttttcaatc gtggggttga	11340
	gtagtgccac acagcataaa attagcttgg ttcatgctc cgtaaagtca tagcgactaa	11400
	tcgctagttc atttgctttg aaaacaacta attcagacat acatctcaat tggcttaggt	11460
	gattttaat	11469

<210> 64
 <211> 1173
 <212> PRT
 <213> 耻垢分枝杆菌

<400> 64

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 Pro Leu Pro Ala Val Val Asp Ala
 35 40 45

Ala His Lys Pro Gly Leu Arg Leu Ala Glu Ile Leu Gln Thr Leu Phe
 50 55 60

Thr Gly Tyr Gly Asp Arg Pro Ala Leu Gly Tyr Arg Ala Arg Glu Leu
 65 70 75 80

Ala Thr Asp Glu Gly Gly Arg Thr Val Thr Arg Leu Leu Pro Arg Phe
 85 90 95

Asp Thr Leu Thr Tyr Ala Gln Val Trp Ser Arg Val Gln Ala Val Ala
 100 105 110

Ala Ala Leu Arg His Asn Phe Ala Gln Pro Ile Tyr Pro Gly Asp Ala
 115 120 125

[0054]

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

Val Cys Ala Tyr Leu Gly Leu Val Ser Val Pro Leu Gln His Asn Ala
 145 150 155 160

Pro Val Ser Arg Leu Ala Pro Ile Leu Ala Glu Val Glu Pro Arg Ile
 165 170 175

Leu Thr Val Ser Ala Glu Tyr Leu Asp Leu Ala Val Glu Ser Val Arg
 180 185 190

Asp Val Asn Ser Val Ser Gln Leu Val Val Phe Asp His His Pro Glu
 195 200 205

Val Asp Asp His Arg Asp Ala Leu Ala Arg Ala Arg Glu Gln Leu Ala
 210 215 220

Gly Lys Gly Ile Ala Val Thr Thr Leu Asp Ala Ile Ala Asp Glu Gly
 225 230 235 240

Ala Gly Leu Pro Ala Glu Pro Ile Tyr Thr Ala Asp His Asp Gln Arg
 245 250 255

Leu Ala Met Ile Leu Tyr Thr Ser Gly Ser Thr Gly Ala Pro Lys Gly
 260 265 270

Ala Met Tyr Thr Glu Ala Met Val Ala Arg Leu Trp Thr Met Ser Phe

[0055]

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 Gly	Asn Ser Glu Arg Ser	Phe Leu Leu Ala Val
545	550	555
Val Val Pro Thr Pro Glu	Ala Leu Glu Gln Tyr	Asp Pro Ala Ala Leu
565	570	575
Lys Ala Ala Leu Ala Asp	Ser Leu Gln Arg Thr	Ala Arg Asp Ala Glu

	580	585	590
	Leu Gln Ser Tyr Glu Val Pro Ala Asp Phe Ile Val Glu Thr Glu Pro 595 600 605		
	Phe Ser Ala Ala Asn Gly Leu Leu Ser Gly Val Gly Lys Leu Leu Arg 610 615 620		
	Pro Asn Leu Lys Asp Arg Tyr Gly Gln Arg Leu Glu Gln Met Tyr Ala 625 630 635 640		
	Asp Ile Ala Ala Thr Gln Ala Asn Gln Leu Arg Glu Leu Arg Arg Ala 645 650 655		
	Ala Ala Thr Gln Pro Val Ile Asp Thr Leu Thr Gln Ala Ala Ala Thr 660 665 670		
	Ile Leu Gly Thr Gly Ser Glu Val Ala Ser Asp Ala His Phe Thr Asp 675 680 685		
	Leu Gly Gly Asp Ser Leu Ser Ala Leu Thr Leu Ser Asn Leu Leu Ser 690 695 700		
	Asp Phe Phe Gly Phe Glu Val Pro Val Gly Thr Ile Val Asn Pro Ala 705 710 715 720		
	Thr Asn Leu Ala Gln Leu Ala Gln His Ile Glu Ala Gln Arg Thr Ala 725 730 735		
[0056]	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 835 840 845		
	Asp Arg His Leu Arg Val Val Ala Gly Asp Ile Gly Asp Pro Asn Leu 850 855 860		
	Gly Leu Thr Pro Glu Ile Trp His Arg Leu Ala Ala Glu Val Asp Leu 865 870 875 880		
	Val Val His Pro Ala Ala Leu Val Asn His Val Leu Pro Tyr Arg Gln		

	885	890	895
	Leu Phe Gly Pro Asn Val Val Gly Thr Ala Glu Val Ile Lys Leu Ala 900 905 910		
	Leu Thr Glu Arg Ile Lys Pro Val Thr Tyr Leu Ser Thr Val Ser Val 915 920 925		
	Ala Met Gly Ile Pro Asp Phe Glu Glu Asp Gly Asp Ile Arg Thr Val 930 935 940		
	Ser Pro Val Arg Pro Leu Asp Gly Gly Tyr Ala Asn Gly Tyr Gly Asn 945 950 955 960		
	Ser Lys Trp Ala Gly Glu Val Leu Leu Arg Glu Ala His Asp Leu Cys 965 970 975		
	Gly Leu Pro Val Ala Thr Phe Arg Ser Asp Met Ile Leu Ala His Pro 980 985 990		
	Arg Tyr Arg Gly Gln Val Asn Val Pro Asp Met Phe Thr Arg Leu Leu 995 1000 1005		
	Leu Ser Leu Leu Ile Thr Gly Val Ala Pro Arg Ser Phe Tyr Ile 1010 1015 1020		
[0057]	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		

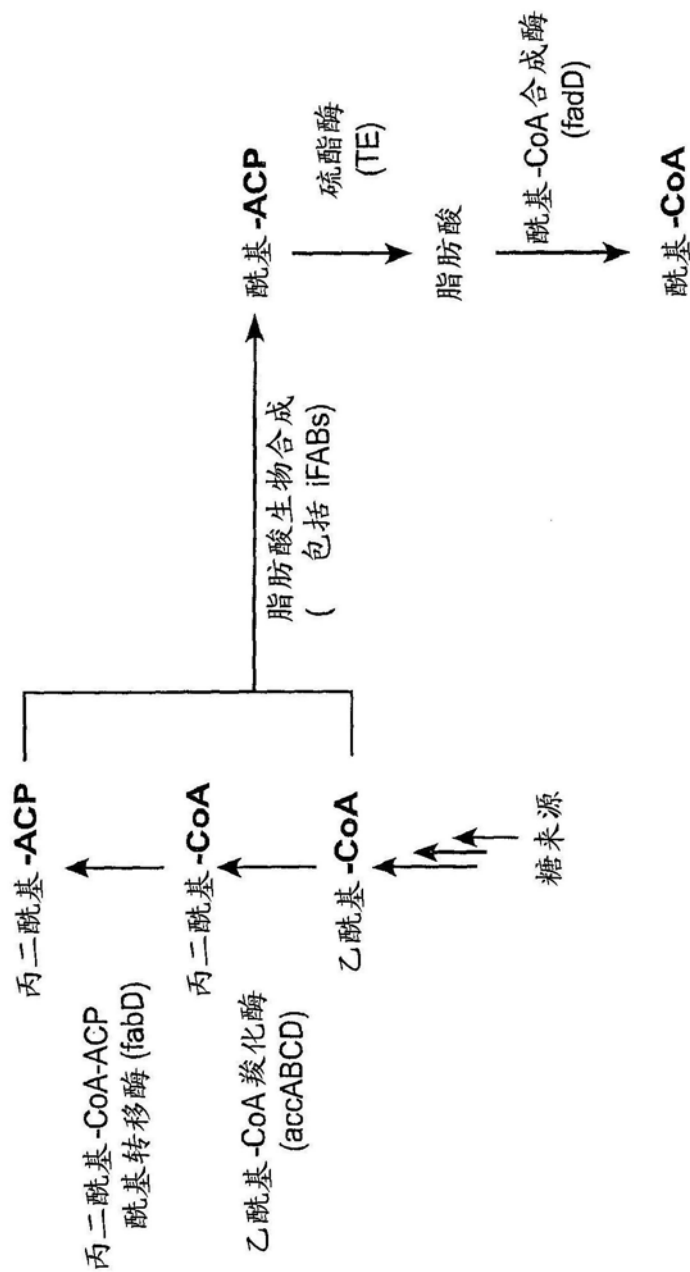


图1

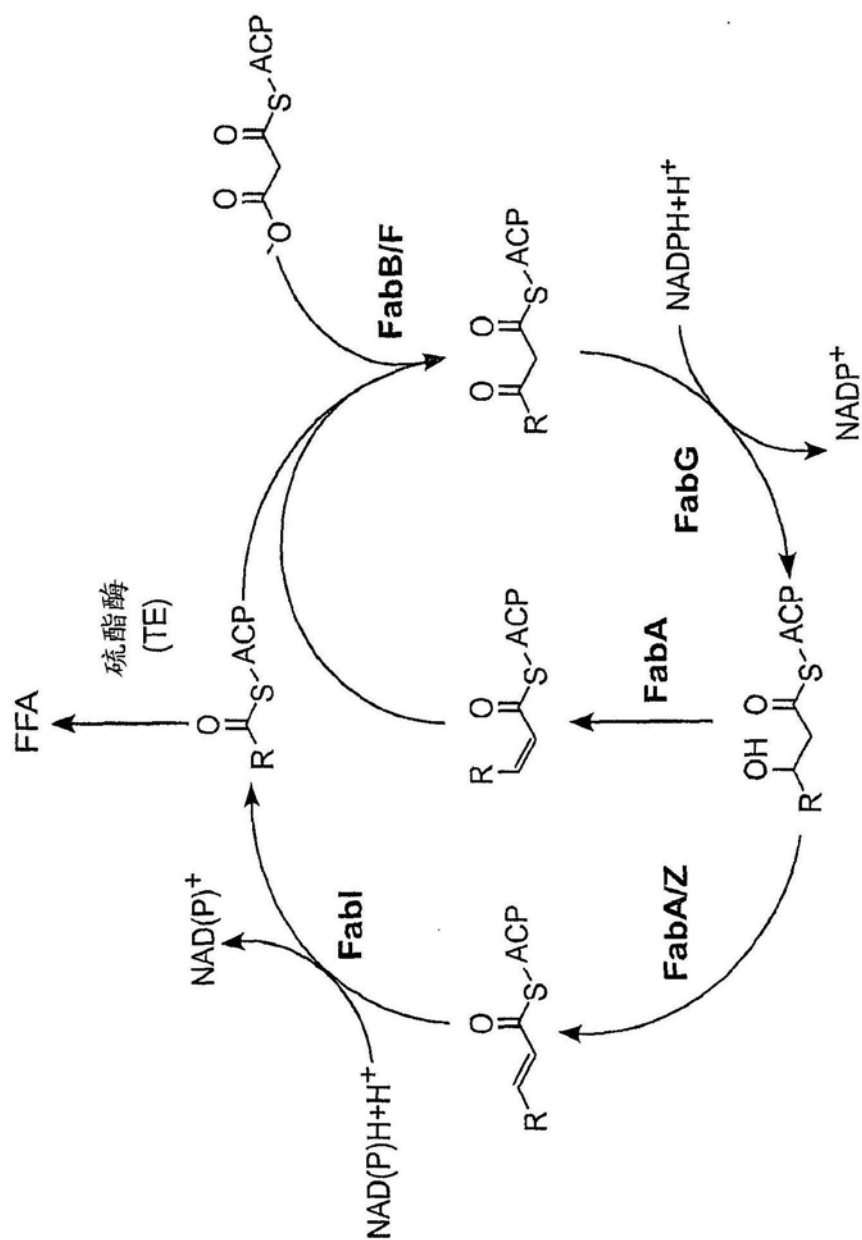


图2

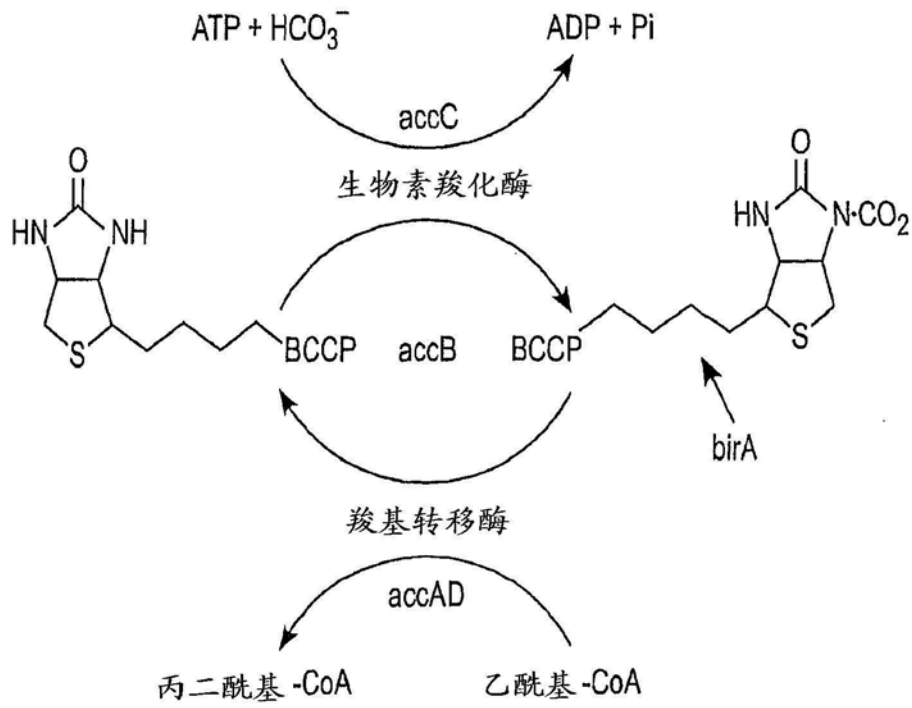


图3

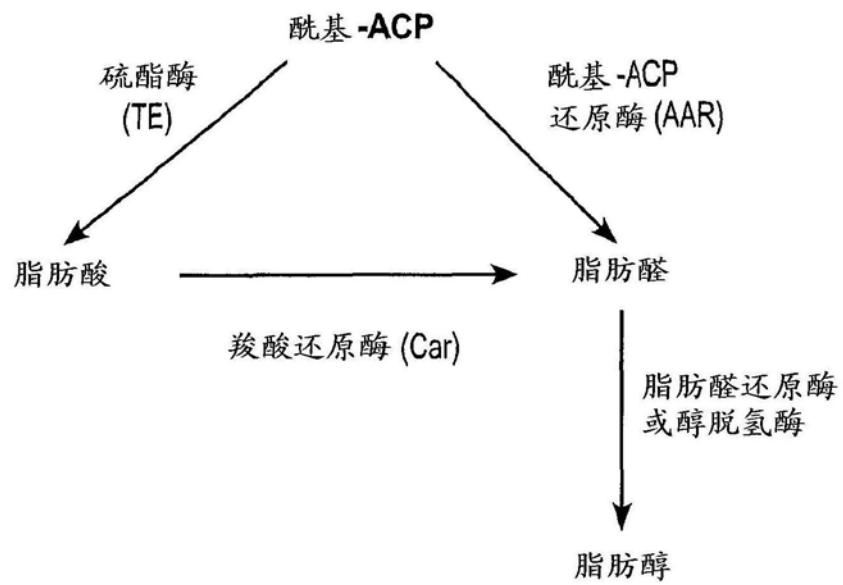


图4

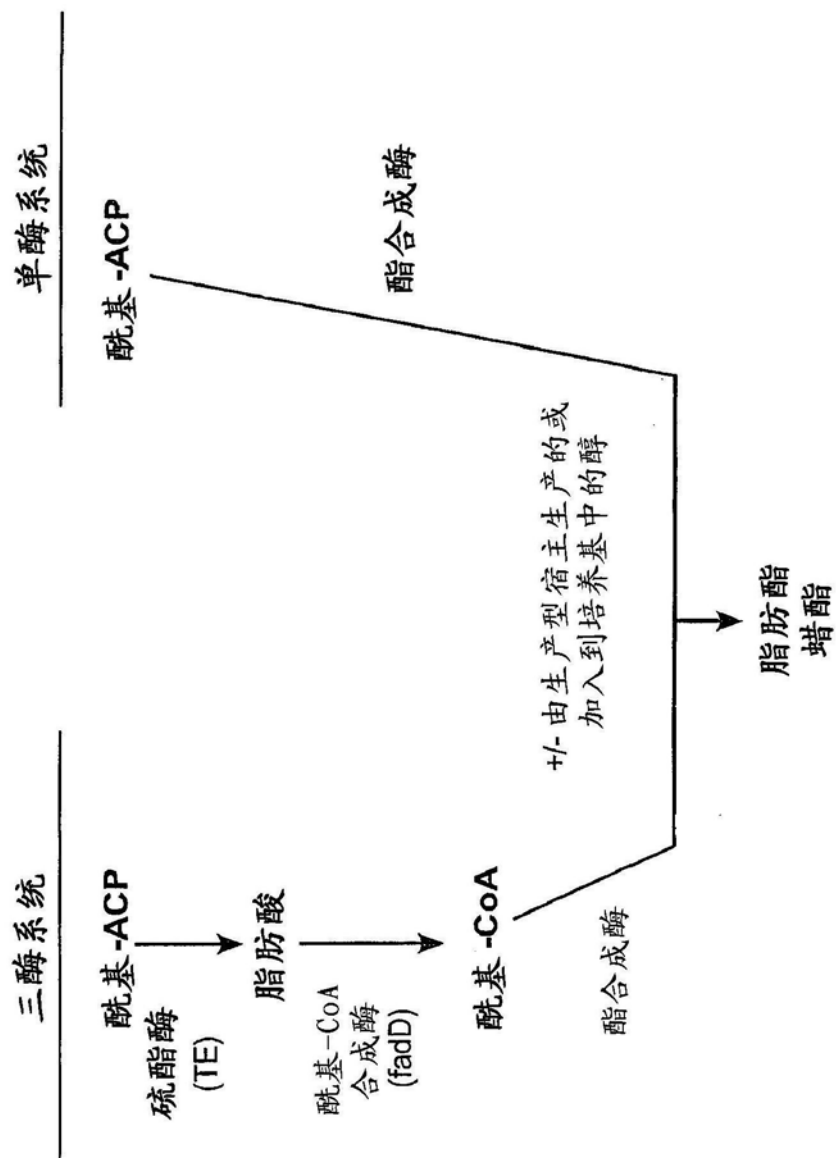


图5

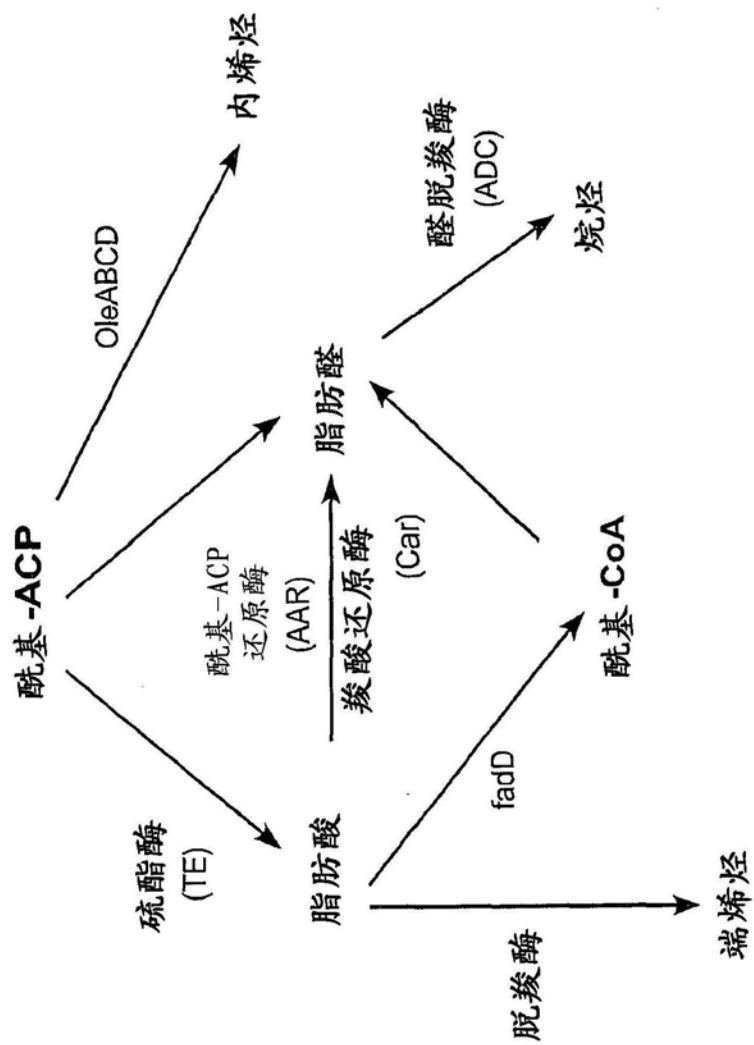


图6

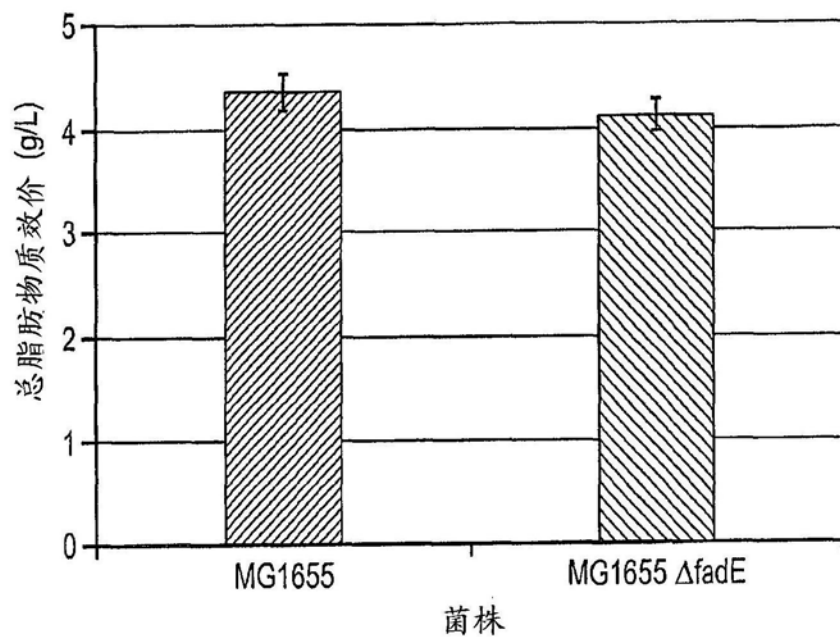


图7

具有谷氨酸棒状杆菌Acc和BirA表达的
DAM1-377对数期的Mal-CoA

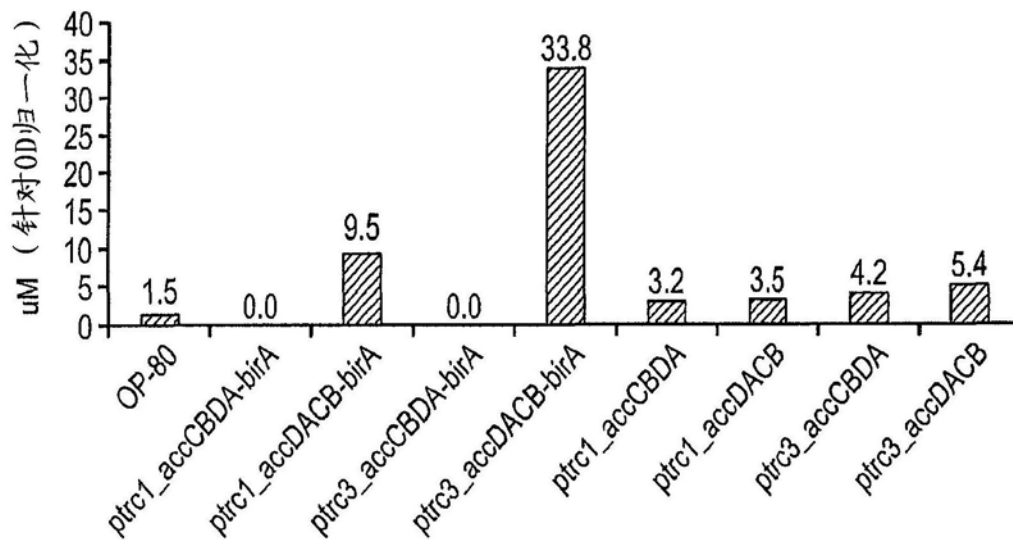


图8

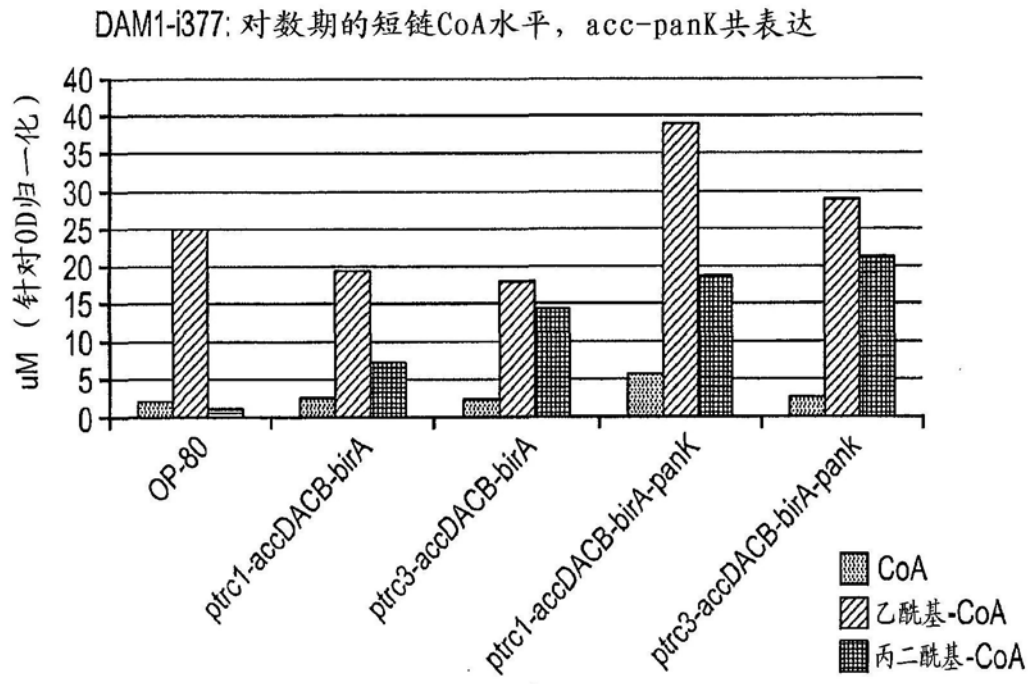


图9

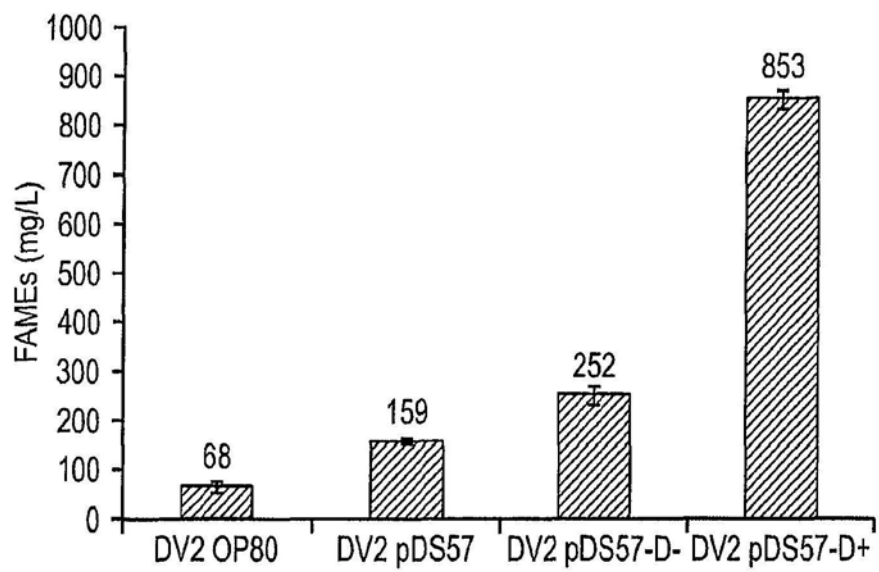


图10

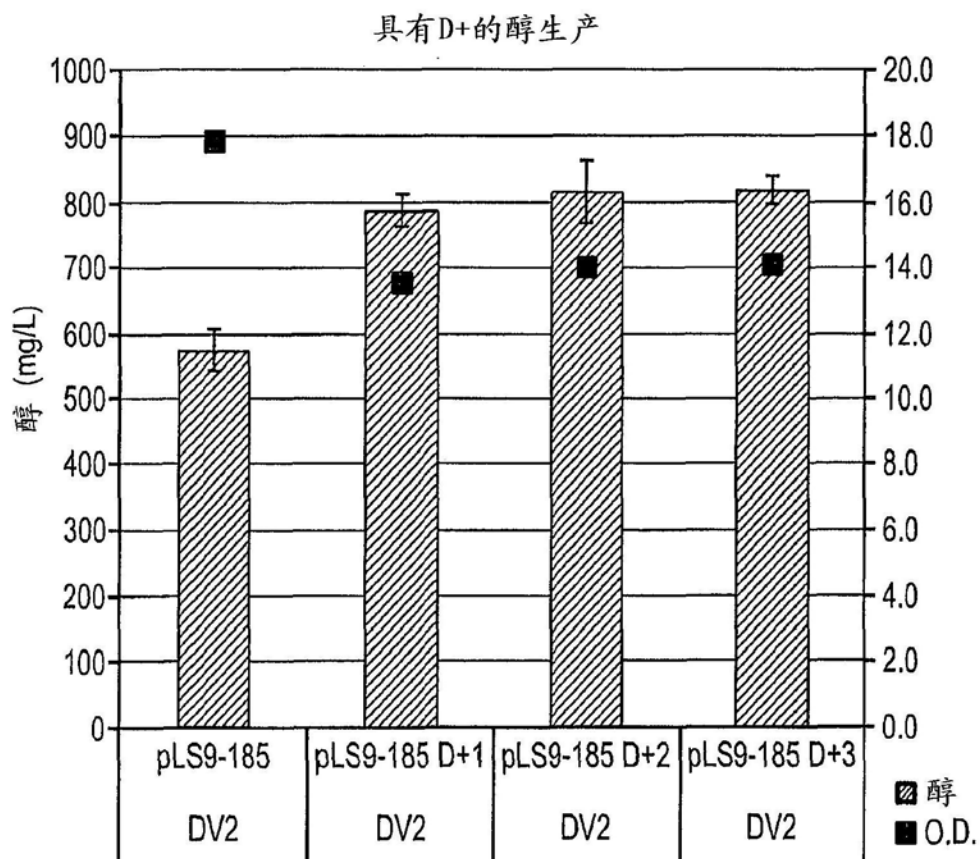


图11

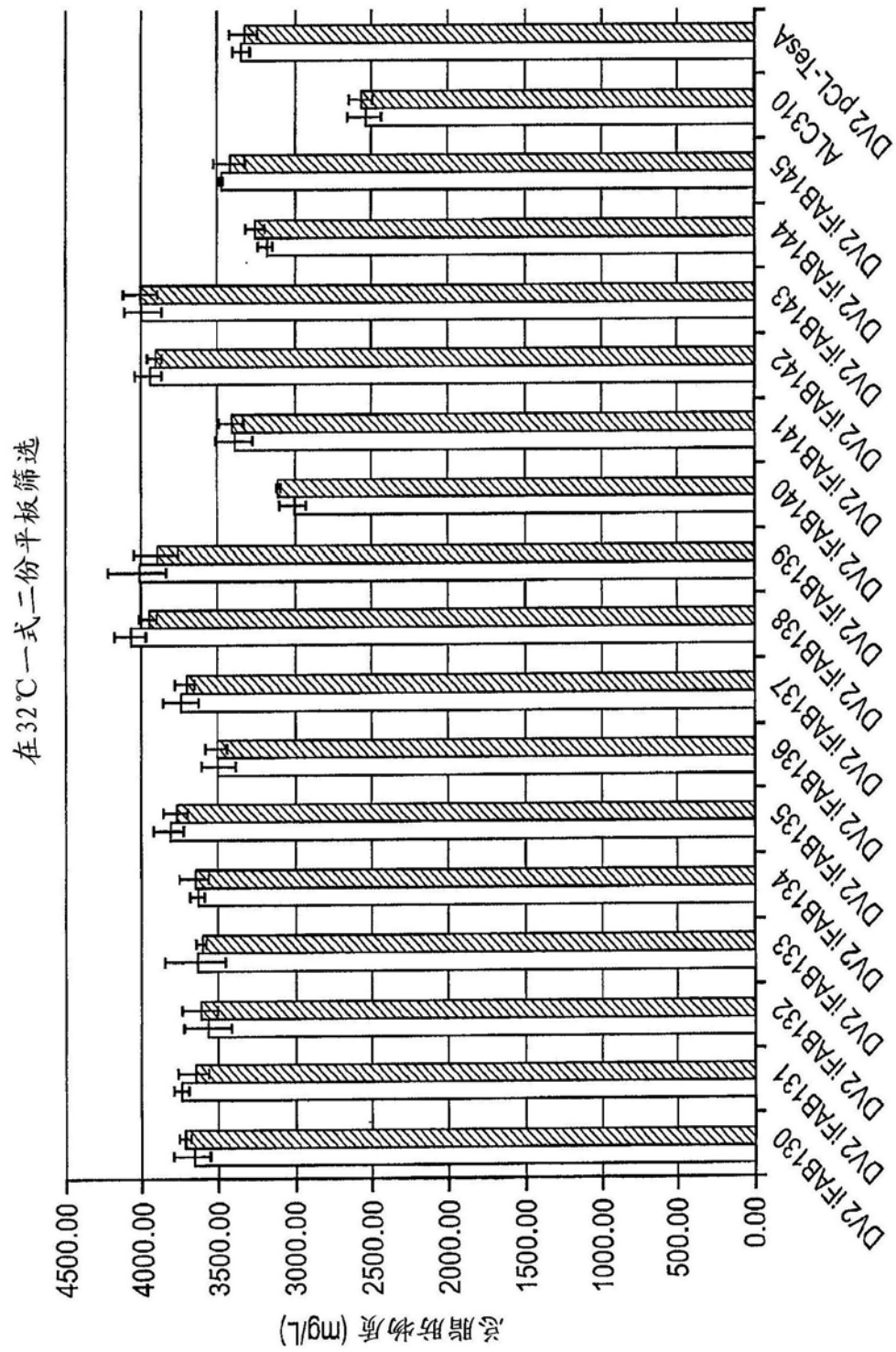


图12A

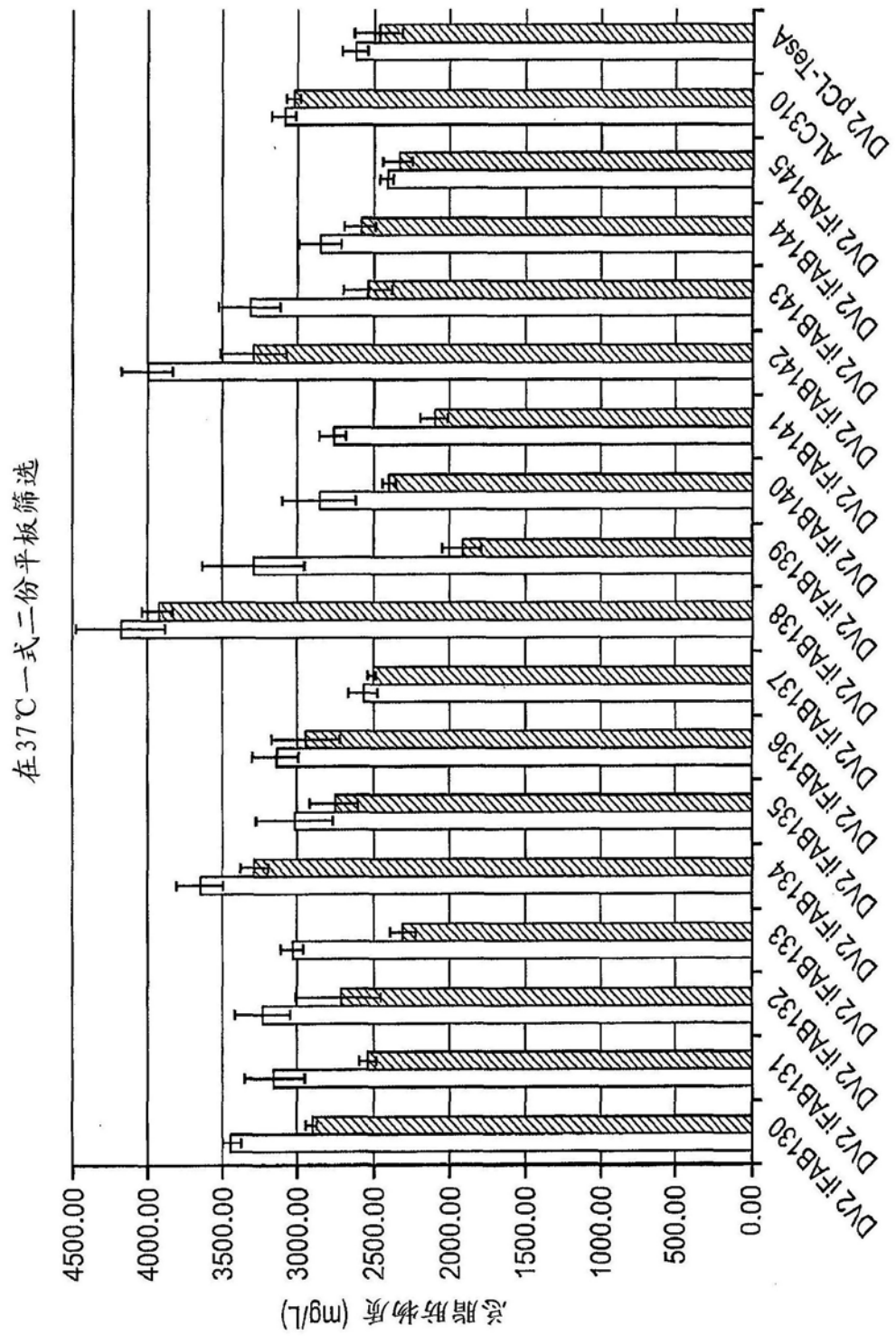


图12B

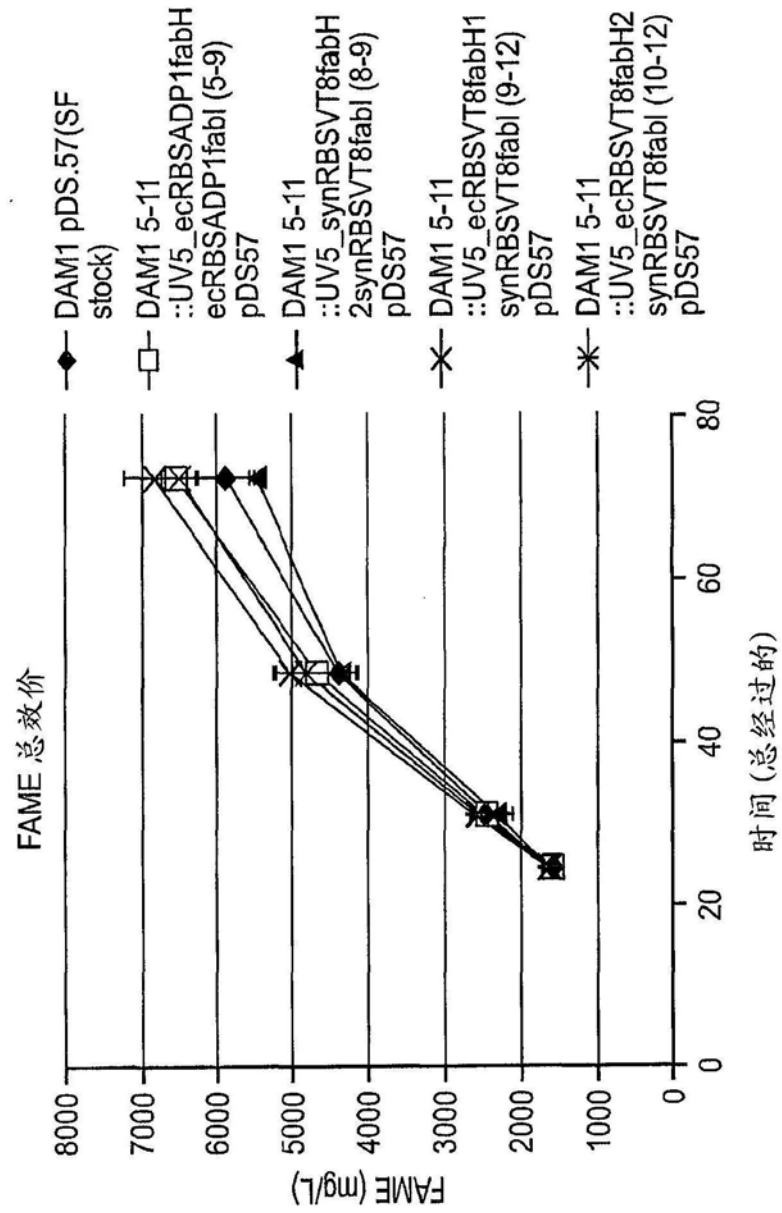


图13

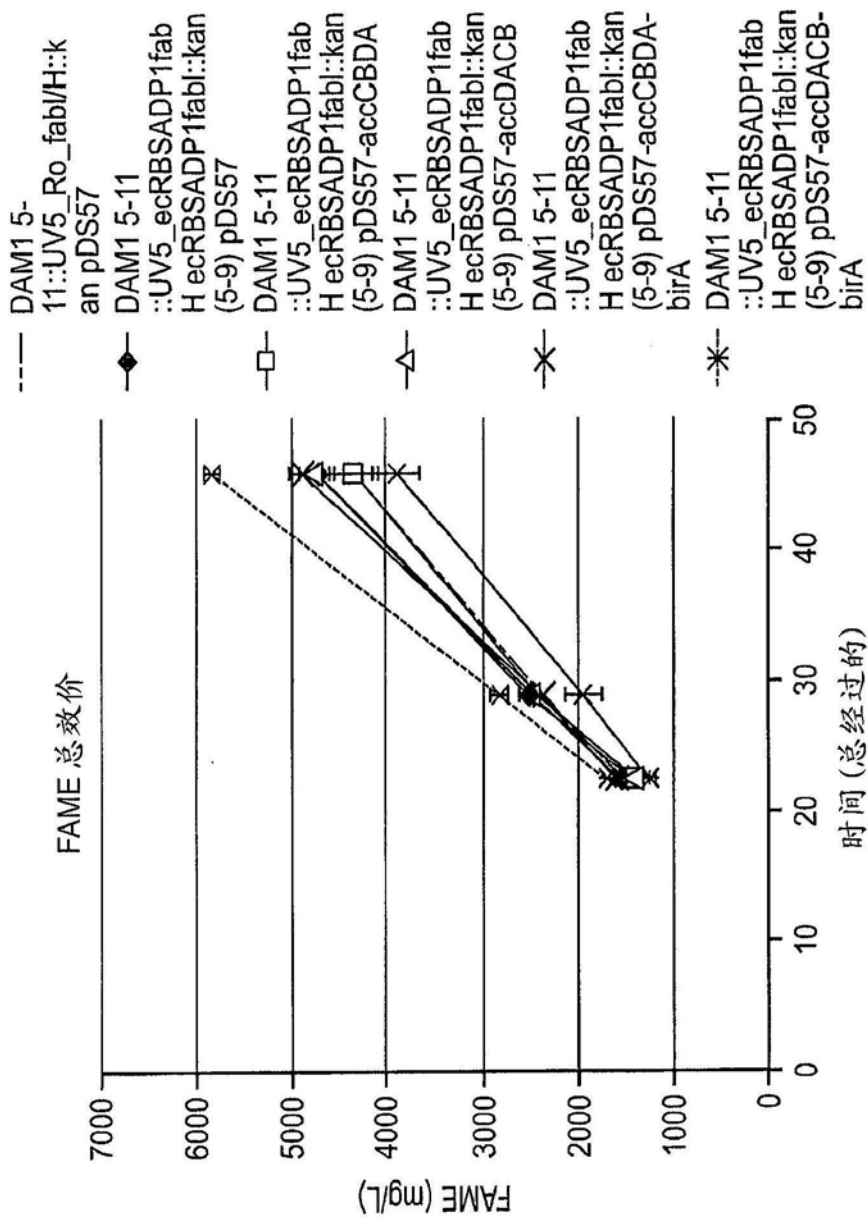


图14

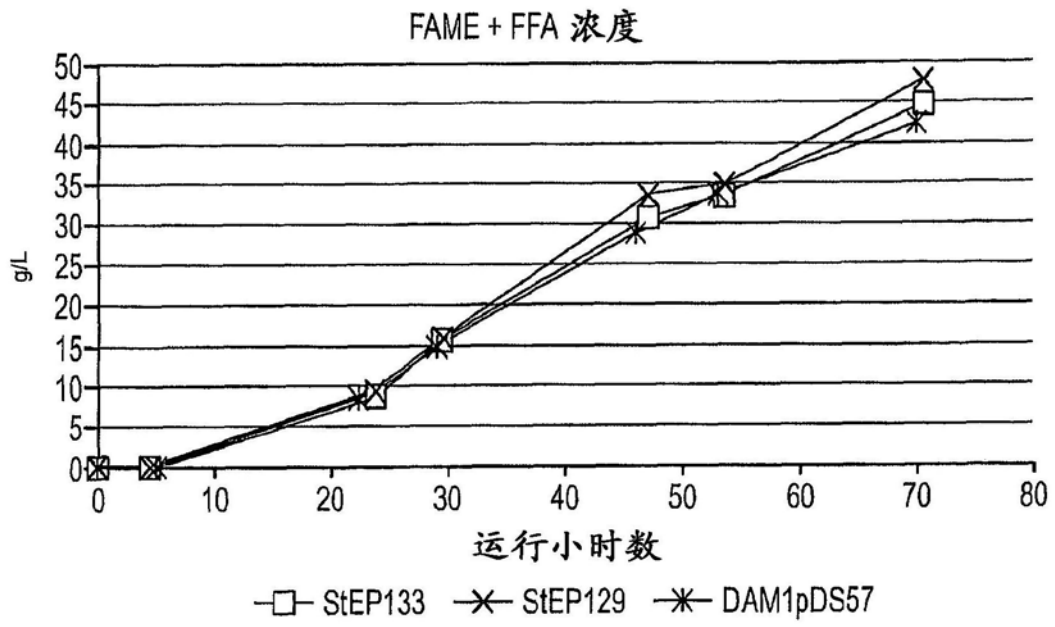


图15

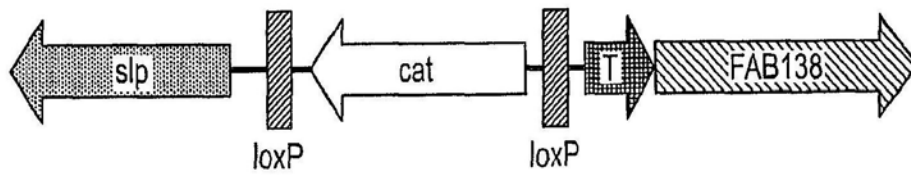


图16A

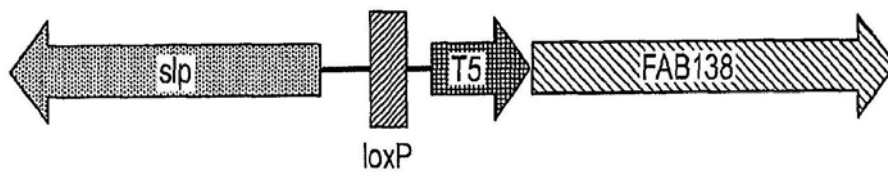


图16B

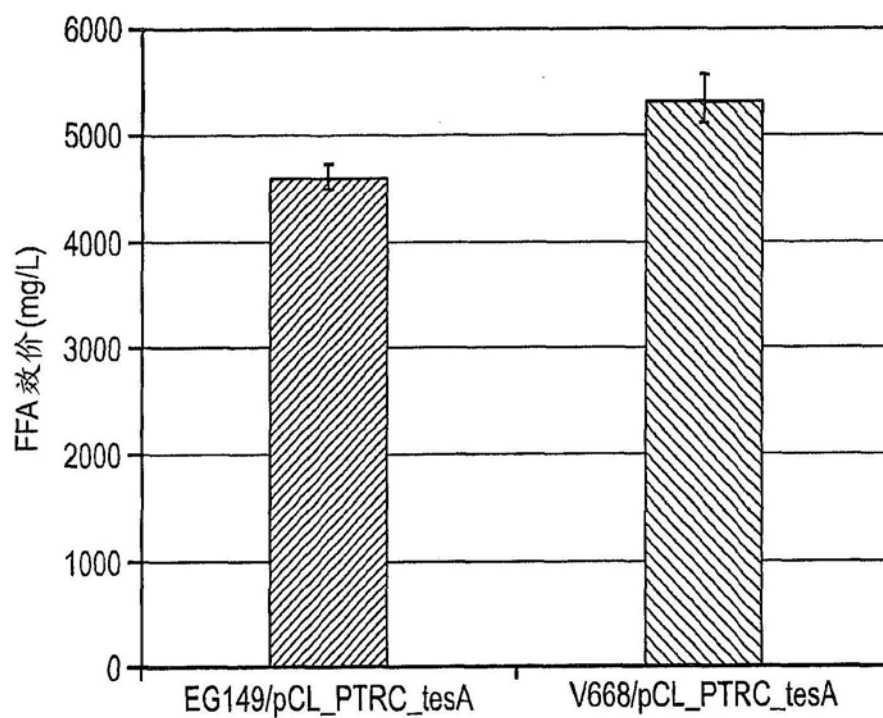


图17

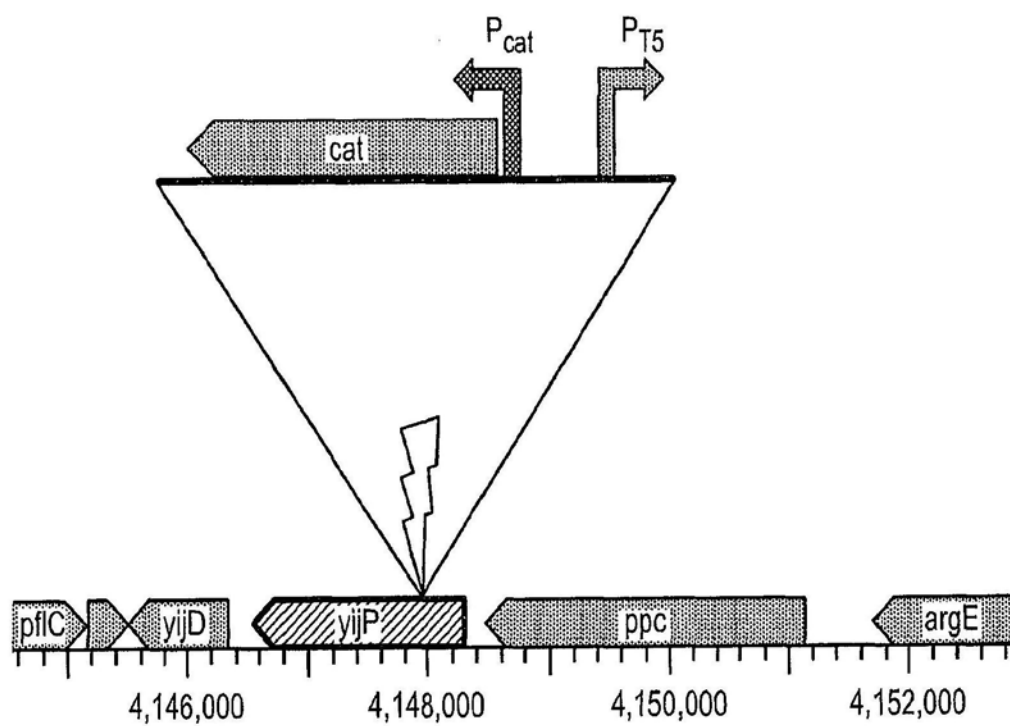


图18

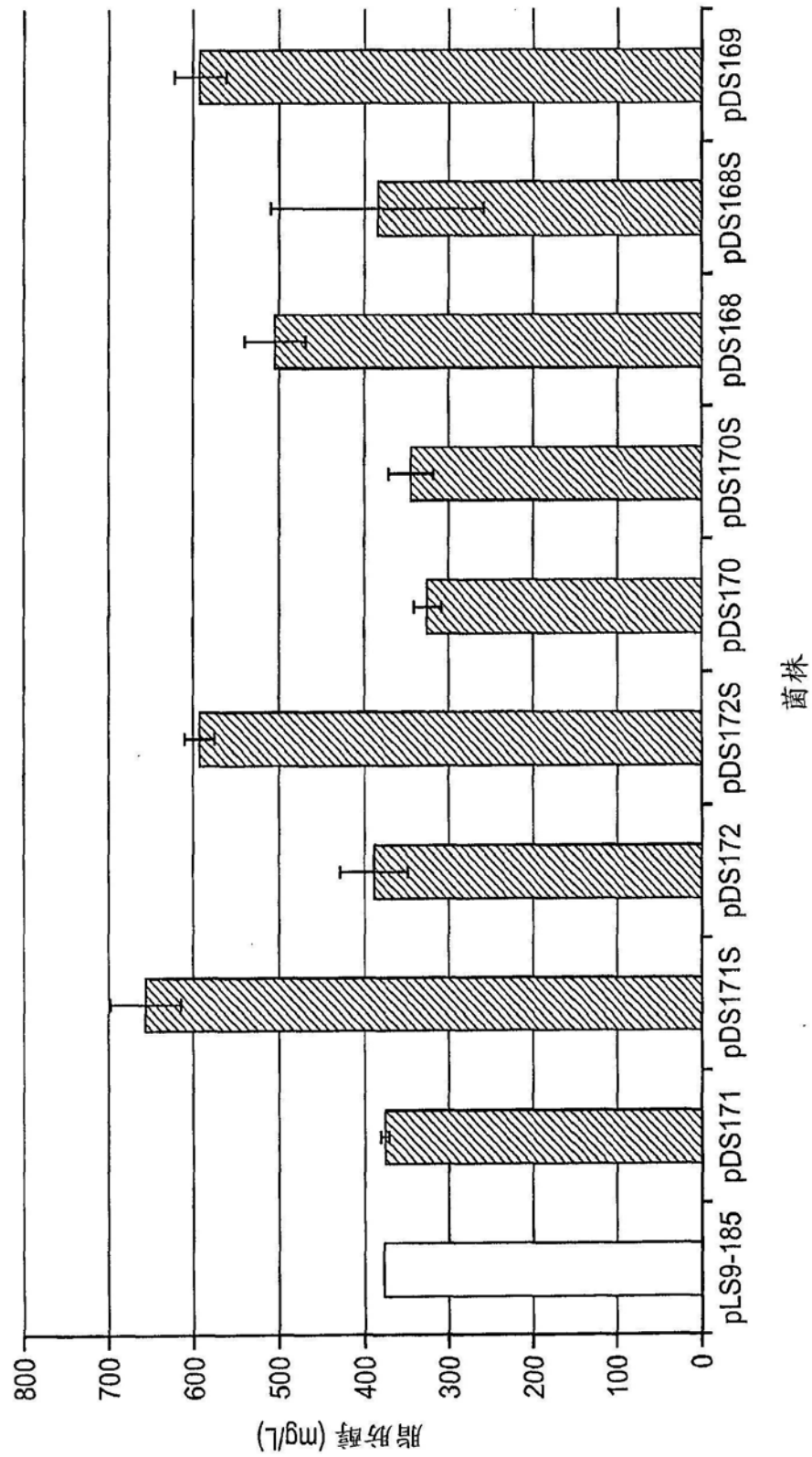


图19

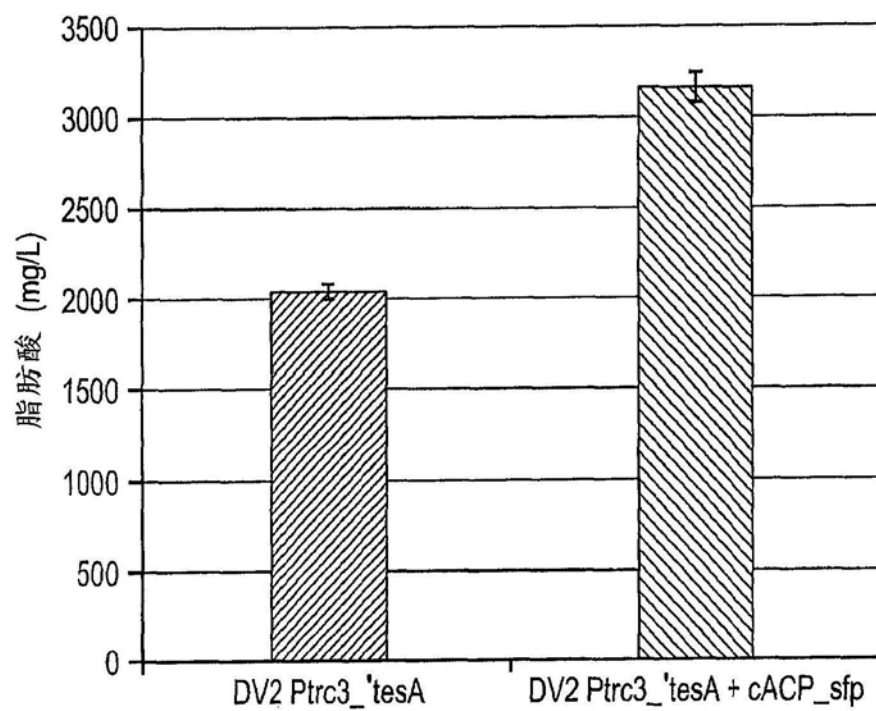


图20