(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 105378486 B
(45)授权公告日 2024.05.14

(21)申请号 201480033702.8
(22)申请日 2014.06.16
(65)同一申请的已公布的文献号
申请公布号 CN 105378486 A
(43)申请公布日 2016.03.02
(30)优先权数据
61/835,464 2013.06.14 US
(85)PCT国际申请进入国家阶段日
2015.12.14
(86)PCT国际申请的申请数据
PCT/US2014/042594 2014.06.16
(87)PCT国际申请的公布数据
WO2014/201474 EN 2014.12.18
(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 A·W·席尔默 H·王
S·B·德尔卡黛莉 胡志浩
L·G·霍姆 B·朱 C·常
E·E·波波瓦
(74)专利代理机构 中国贸促会专利商标事务所有限公司 11038
专利代理师 袁泉
(51)Int.Cl.
C12N 1/21(2006.01)
C12P 7/6409(2022.01)
C12P 13/00(2006.01)
C12P 7/42(2006.01)
C12P 7/18(2006.01)
C12P 7/6436(2022.01)
C12R 1/19(2006.01)
(56)对比文件
US 2008220419 A1,2008.11.11
Sumire Honda et al..Bacterial CYP153A monooxygenases for the synthesis of omega-hydroxylated fatty acids.《ChemComm》.2012,第48卷摘要部分.
审查员 于淼

权利要求书2页 说明书71页
序列表70页 附图46页

(54)发明名称

生产OMEGA-羟基化的脂肪酸衍生物的方法

(57)摘要

本发明涉及omega-羟基化的脂肪酸衍生物及其生产方法。在本发明中,本发明公开涵盖了以高纯度和产率提供omega-羟基化的脂肪酸衍生物的新型且环境友好的生产方法。此外,本发明公开进一步涵盖通过选择性发酵而生产omega-羟基化的脂肪酸衍生物的重组微生物有机体。

CN 105378486 B

1.经改造而表达编码经修饰的ω-羟化酶的核酸序列的重组大肠杆菌,其中所述经修饰的ω-羟化酶与SEQ ID NO:6的差异在于:

(i)相应于V141I、V141T、V141Q、V141G、R27L、R178N、N309R、N407A、V415R、P666A或A796V的突变;

(ii)突变V141I和A231T;

(iii)突变R27L、R82D、V141M、R178N和N407A;

(iv)突变V141I、A231T和A796V;

(v)突变R27L、R82D、V141M、R178N、N407A和A796V;

(vi)突变V141T、A231T和A796V;

(vii)突变V141M和E124Q;

(viii)突变R82D和E271F;

(ix)突变V141I、A231T、T516G、P666M和A796V;

(x)突变V141I、A231T、T516G、P666H和A796V;

(xi)突变V141I、A231T、T516V、P666D和A796V;或

(xii)突变V141I、A231T、P666M和V696T,

并且其中所述经修饰的ω-羟化酶具有经修饰的细胞色素P450单加氧酶(P450)酶活性。

2.权利要求1所述的重组大肠杆菌,其中所述重组大肠杆菌经改造而进一步表达编码硫酯酶或酯合酶的核酸序列。

3.权利要求1或2所述的重组大肠杆菌,其中所述重组大肠杆菌产生ω-羟基脂肪酸衍生物,该ω-羟基脂肪酸衍生物包括ω-羟基脂肪酸或ω-羟基脂肪酸甲基酯。

4.权利要求1所述的重组大肠杆菌,其中所述重组大肠杆菌经改造而进一步表达编码醇脱氢酶或者醇氧化酶的核酸序列。

5.权利要求4所述的重组大肠杆菌,其中所述重组大肠杆菌产生ω-羟基脂肪酸衍生物,该ω-羟基脂肪酸衍生物包括ω-氧代脂肪酸或ω-氧代脂肪酸甲基酯。

6.权利要求4所述的重组大肠杆菌,其中所述重组大肠杆菌经改造而进一步表达编码醛脱氢酶或者醛氧化酶的核酸序列。

7.权利要求6所述的重组大肠杆菌,其中所述重组大肠杆菌产生ω-羟基脂肪酸衍生物,该ω-羟基脂肪酸衍生物包括α,ω-二酸或ω-羧基脂肪酸甲基酯。

8.权利要求6所述的重组大肠杆菌,其中所述重组大肠杆菌经改造而进一步表达编码酰基-CoA连接酶或者酰基-CoA转移酶的核酸序列。

9.权利要求8所述的重组大肠杆菌,其中所述重组大肠杆菌产生ω-羟基脂肪酸衍生物,该ω-羟基脂肪酸衍生物包括α,ω-二酯。

10.权利要求4所述的重组大肠杆菌,其中所述重组大肠杆菌经改造而进一步表达编码氨基转移酶或者胺脱氢酶的核酸序列。

11.权利要求10所述的重组大肠杆菌,其中所述重组大肠杆菌产生ω-羟基脂肪酸衍生物,该ω-羟基脂肪酸衍生物包括ω-氨基脂肪酸或ω-氨基脂肪酸甲基酯。

12.权利要求1所述的重组大肠杆菌,其中所述重组大肠杆菌经改造而进一步表达编码醇脱氢酶以及羧酸还原酶的核酸序列。

13.权利要求12所述的重组大肠杆菌,其中所述重组大肠杆菌产生ω-羟基脂肪酸衍生物,该ω-羟基脂肪酸衍生物包括α,ω-二醇。

14.权利要求1所述的重组大肠杆菌,其中所述重组大肠杆菌经改造而进一步表达编码酰基-ACP还原酶和醇脱氢酶的核酸序列,并且其中所述重组大肠杆菌产生α,ω-二醇。

15.包含根据权利要求1-14中任一项所述的大肠杆菌的细胞培养物。

16.权利要求15所述的细胞培养物,其中所述细胞培养物产生ω-羟基脂肪酸衍生物,该ω-羟基脂肪酸衍生物包括ω-羟基游离脂肪酸、ω-羟基脂肪酸甲基酯、ω-氧代脂肪酸、ω-氧代脂肪酸甲基酯、α,ω-二酸、α,ω-二醇、ω-氨基脂肪酸或ω-氨基脂肪酸甲基酯。

17.产生ω-羟基脂肪酸衍生物的方法,其包括:

(a)提供根据权利要求1-14中任一项所述的重组大肠杆菌;

(b)将包含碳源的可再生的供料加入发酵液中;以及

(c)由所述的发酵液分离ω-羟基脂肪酸衍生物。

18.权利要求17所述的重组大肠杆菌,其中所述重组大肠杆菌产生产生ω-羟基脂肪酸衍生物,该ω-羟基脂肪酸衍生物包括ω-羟基游离脂肪酸、ω-羟基脂肪酸甲基酯、ω-氧代脂肪酸、ω-氧代脂肪酸甲基酯、α,ω-二酸、α,ω-二醇、ω-氨基脂肪酸或ω-氨基脂肪酸甲基酯。

生产OMEGA-羟基化的脂肪酸衍生物的方法

[0001] 相关申请的交叉引用

[0002] 本申请要求2013年6月14日提交的美国临时申请No.61/835,464的 权益,该申请的全部公开内容以引用方式并入本文。

[0003] 序列表

[0004] 本申请包括序列表,该序列表以ASCII格式以电子方式提交,并以引 用方式全文并入本文。在2014年6月16日创建的所述的ASCII拷贝命名 为LS00048PCT_SL.txt,并且大小为342,103字节。

技术领域

[0005] 本发明公开涉及omega-羟基化的脂肪酸衍生物以及生产它们的方 法。在本发明中,本发明公开涵盖了以高纯度和高产率提供了omega-羟基 化的脂肪酸衍生物的新的且环境友好的生产方法。此外,本发明公开涵盖了通过选择性发酵而生产omega-羟基化的脂肪酸衍生物的重组微生物有 机体。

背景技术

[0006] Omega-羟基化的(ω-羟基)脂肪酸衍生物作为工业试剂的成分具有许多 商业用途。工业界已经确认了多种类型的ω-羟基脂肪酸衍生物,包括ω- 羟基脂肪酸;ω-羟基脂肪酸甲基酯;ω-氧代脂肪酸;ω-氨基脂肪酸;ω-氨 基脂肪酸甲基酯;alpha-,omega-二酸(α,ω-二酸);omega-羧基脂肪酸甲基酯 (ω-羧基脂肪酸甲基酯);alpha-,omega-二酯(α,ω-二酯);alpha-,omega-二醇 (α,ω-二醇)等。这些分子作为多种其他化合物的前体也是重要的。例如α,ω- 二羧酸和其他的α,ω-双官能分子在聚合物树脂、金属工作液、粘合剂、腐 蚀抑制剂、电容器电解质、二酯合成润滑剂、纤维、粉末涂敷药品、增塑 剂、聚酯涂料、环氧树脂、聚酰胺树脂、香料(flavor)、香精(fragrance)、 表面活性剂、洗涤剂、添加剂等的工业应用中是重要的化学品。目前,ω- 羟基脂肪酸衍生物仍主要由石油基材料或者通过石蜡和脂肪酸的生物转 化而制得。用于生产这些化合物的化学方法需要使用危险试剂,并且是能量密集型的且环境代价大的。相反,新兴的发酵途径尽管被认为是绿色工 艺,但是仍是非常昂贵并局限于可以制备的几种类型的产物。因此,由可 再生的供料来直接生产多种类型和功能性的ω-羟基脂肪酸衍生物的工艺 不仅对环境更安全,而且还显著地更具有成本效益。本发明公开解决了这 种需要。

[0007] 发明概述

[0008] 本发明公开的一个方面提供了在由可再生的供料得到的碳源存在下, 在发酵液中生长时,用于在体内生产omega-羟基化的(ω-羟基)脂肪酸衍生 物的重组微生物有机体。所述微生物有机体包含经改造而表达编码如下多 肽的至少2种核酸序列的途径,所述多肽包括EC 3.1.2.-、3.1.1.5或3.1.2.14 的硫酯酶;或者EC 2.3.1.75或EC 2.3.1.20的酯合酶;以及EC 1.14.15.3的 经修饰的ω-羟化酶。所述经修饰的ω-羟化酶具有经修饰的细胞色素P450 单加氧酶(P450)酶活性,并且在体内有效地催化碳氢化合物链的ω-位置。

在一个实施方案中,经修饰的ω-羟化酶为EC 1.14.15.3的ω-羟化酶杂交融 合蛋白质变体。在另一个实施方案中,经修饰的ω-羟化酶为CYP153A-还 原酶杂交融合多肽。在另一个实施方案中,所述CYP153A-还原酶杂交融 合多肽是自给自足的CYP153A-RedRhF杂交融合蛋白质。在另一个实施方案中,为ω-羟化酶杂交融合蛋白质变体的经修饰的ω-羟化酶与SEQ ID NO: 6具有至少90%的序列一致性,并且具有一个或多个突变,包括V141I、V141T、V141Q、V141G、V141M、V141L、R27L、R82D、R178N、A231T、 N309R、N407A、V415R、T516V、P666A、P666D和/或A796V。在本发 明中,所述ω-羟化酶杂交融合蛋白质变体具有在氨基酸位置27、82、141、 178、231、309、407、415、516、666和/或796的至少一个突变。所述重 组微生物有机体产生ω-羟基脂肪酸衍生物,包括但不限于ω-羟基脂肪酸和 ω-羟基脂肪酸甲基酯。

[0009] 本发明公开的另一个方面提供了在由可再生的供料得到的碳源存在 下,在发酵液中生长时,用于在体内生产ω-羟基脂肪酸衍生物的重组微生 物有机体。所述微生物有机体包含经改造而表达编码如下多肽的至少2种 核酸序列的途径,所述多肽包括EC 3.1.2.-、3.1.1.5或3.1.2.14的硫酯酶; 或者EC 2.3.1.75或EC 2.3.1.20的酯合酶;以及EC 1.14.15.3的经修饰的ω- 羟化酶。所述修饰的ω-羟化酶具有经修饰的细胞色素P450单加氧酶(P450) 酶活性,并且在体内有效地催化碳氢化合物链的ω-位置。在一个方面中,所述重组微生物有机体经改造而进一步表达编码如下多肽的核酸序列,所 述多肽包括EC 1.1.1.1/2的醇脱氢酶,或者EC 1.1.3.13或EC 1.1.3.20的醇 氧化酶。在一个实施方案中,经修饰的ω-羟化酶为EC 1.14.15.3的ω-羟化 酶杂交融合蛋白质变体。在另一个实施方案中,经修饰的ω-羟化酶为 CYP153A-还原酶杂交融合多肽。在另一个实施方案中,CYP153A-还原酶杂交融合多肽是自给自足的CYP153A-RedRhF杂交融合蛋白质。在另一个实施方案中,为ω-羟化酶杂交融合蛋白质变体的经修饰的ω-羟化酶与SEQ ID NO:6具有至少90%的序列一致性,并且具有一个或多个突变,包括 V141I、V141T、V141Q、V141G、V141M、V141L、R27L、R82D、R178N、 A231T、N309R、N407A、V415R、T516V、P666A、P666D和/或A796V。 所述重组微生物有机体产生ω-羟基脂肪酸衍生物,包括但不限于ω-氧代脂肪酸和ω-氧代脂肪酸甲基酯。

[0010] 本发明公开的另一个方面提供了在由可再生的供料得到的碳源存在 下,在发酵液中生长时,用于在体内生产ω-羟基脂肪酸衍生物的重组微生 物有机体。所述微生物有机体包含经改造而表达编码如下多肽的至少2种 核酸序列的途径,所述多肽包括EC 3.1.2.-、3.1.1.5或3.1.2.14的硫酯酶; 或者EC 2.3.1.75或EC 2.3.1.20的酯合酶;以及EC 1.14.15.3的经修饰的ω- 羟化酶(细胞色素P450单加氧酶)。在一个方面中,所述重组微生物有机体 经改造而进一步表达编码如下多肽的核酸序列,所述多肽包括EC 1.1.1.1/2 的醇脱氢酶,或者EC 1.1.3.13或EC 1.1.3.20的醇氧化酶。在另一个方面中,经修饰的微生物有机体经改造而进一步表达编码如下多肽的核酸序 列,所述多肽包括EC 1.2.1.3/4/5的醛脱氢酶或EC 1.2.3.1的醛氧化酶。所 述修饰的ω-羟化酶具有经修饰的细胞色素P450单加氧酶(P450)酶活性, 并且在体内有效地催化碳氢化合物链的ω-位置。在一个实施方案中,经修 饰的ω-羟化酶为EC 1.14.15.3的ω-羟化酶杂交融合蛋白质变体。在另一个 实施方案中,经修饰的ω-羟化酶为CYP153A-还原酶杂交融合多肽。在另一个实施方案中,CYP153A-还原酶杂交融合多肽是自给自足的 CYP153A-RedRhF杂交融合蛋白质。在另一

个实施方案中,为ω-羟化酶杂 交融合蛋白质变体的经修饰的ω-羟化酶与SEQ ID NO:6具有至少90％的 序列一致性,并且具有一个或多个突变,包括V141I、V141T、V141Q、 V141G、V141M、V141L、R27L、R82D、R178N、A231T、N309R、N407A、 V415R、T516V、P666A、P666D和/或A796V。所述重组微生物有机体产 生ω-羟基脂肪酸衍生物,包括但不限于ω-羟基脂肪酸衍生物,其为α,ω- 二酸或ω-羧基脂肪酸甲基酯。

[0011] 而且,本发明的另一个方面提供了在由可再生的供料得到的碳源存在 下,在发酵液中生长时,用于在体内生产ω-羟基脂肪酸衍生物的重组微生 物有机体。所述微生物有机体包含经改造而表达编码如下多肽的至少2种 核酸序列的途径,所述多肽包括EC 3.1.2.-,3.1.1.5或3.1.2.14的硫酯酶; 或者EC 2.3.1.75或EC 2.3.1.20的酯合酶;以及EC 1.14.15.3的经修饰的ω- 羟化酶。在一个方面中,所述重组微生物有机体经改造而进一步表达编码 如下多肽的核酸序列,所述多肽包括EC 1.1.1.1/2的醇脱氢酶,或者EC 1.1.3.13或EC 1.1.3.20的醇氧化酶。在另一个方面中,所述重组微生物有 机体经改造而进一步表达编码如下多肽的核酸序列,所述多肽包括EC 1.2.1.3/4/5的醛脱氢酶或EC 1.2.3.1的醛氧化酶。在另一个方面中,所述重 组微生物有机体经改造而进一步表达编码如下多肽的核酸序列,所述多肽 包括EC 6.2.1.3的酰基-CoA连接酶或EC 2.8.3.6的酰基-CoA转移酶。所 述修饰的ω-羟化酶具有经修饰的细胞色素P450单加氧酶(P450)酶活性,并且在体内有效地催化碳氢化合物链的ω-位置。在一个实施方案中,经修 饰的ω-羟化酶为EC 1.14.15.3的ω-羟化酶杂交融合蛋白质变体。在另一个实施方案中,经修饰的ω-羟化酶为CYP153A-还原酶杂交融合多肽。在另 一个实施方案中,CYP153A-还原酶杂交融合多肽是自给自足的CYP153A-RedRhF杂交融合蛋白质。在另一个实施方案中,为ω-羟化酶杂交融合蛋白质变体的经修饰的ω-羟化酶与SEQ ID NO:6具有至少90％的 序列一致性,并且具有一个或多个突变,包括V141I、V141T、V141Q、 V141G、V141M、V141L、R27L、R82D、R178N、A231T、N309R、N407A、 V415R、T516V、P666A、P666D和/或A796V。所述重组微生物有机体产 生ω-羟基脂肪酸衍生物,包括但不限于ω-羟基脂肪酸衍生物,其为α,ω- 二酯。

[0012] 本发明公开进一步涵盖了在由可再生的供料得到的碳源存在下,在发 酵液中生长时,用于在体内生产ω-羟基脂肪酸衍生物的重组微生物有机 体。所述微生物有机体包含经改造而表达编码如下多肽的至少2种核酸序 列的途径,所述多肽包括EC 3.1.2.-,3.1.1.5或3.1.2.14的硫酯酶;或者EC 2.3.1.75或EC 2.3.1.20的酯合酶;以及EC 1.14.15.3的经修饰的ω-羟化酶。 在另一个方面中,所述重组微生物有机体经改造而进一步表达编码如下多 肽的核酸序列,所述多肽包括EC 1.1.1.1/2的醇脱氢酶或者EC 1.1.3.13或 EC 1.1.3.20的醇氧化酶。在另一个方面中,所述重组微生物有机体经改造 而进一步表达编码如下多肽的核酸序列,所述多肽包括EC 2.6.1的氨基转 移酶或者EC 1.4.9、EC 1.4.98或EC 1.4.99的胺脱氢酶。所述ω-羟化酶具 有经修饰的细胞色素P450单加氧酶(P450)酶活性,并且在体内有效地催化 碳氢化合物链的ω-位置。在一个实施方案中,经修饰的ω-羟化酶为EC 1.14.15.3的ω-羟化酶杂交融合蛋白质变体。在另一个实施方案中,经修饰 的ω-羟化酶为CYP153A-还原酶杂交融合多肽。在另一个实施方案中,CYP153A-还原酶杂交融合多肽是自给自足的CYP153A-RedRhF杂交融合 蛋白质。在另一个实施方案中,为ω-羟化酶杂交融合蛋白质变体的经修饰 的ω-羟化酶与SEQ ID NO:6具有至少90％的序列一致性,并且具有一个 或多个突变,包括V141I、V141T、V141Q、V141G、

V141M、V141L、R27L、R82D、R178N、A231T、N309R、N407A、V415R、T516V、P666A、P666D和/或A796V。所述重组微生物有机体产生ω-羟基脂肪酸衍生物，其包括但不限于ω-氨基脂肪酸和ω-氨基脂肪酸甲基酯。

[0013] 本发明公开的另一个方面提供了在由可再生的供料得到的碳源存在 下,在发酵液中生长时,用于在体内生产ω-羟基脂肪酸衍生物的重组微生 物有机体。所述微生物有机体包含经改造而表达编码如下多肽的至少2种 核酸序列的途径,所述多肽包括EC 3.1.2.-、3.1.1.5或3.1.2.14的硫酯酶；或者EC 2.3.1.75或EC 2.3.1.20的酯合酶;以及EC 1.14.15.3的经修饰的ω- 羟化酶。在一个方面中,所述重组微生物有机体经改造而进一步表达编码 如下多肽的核酸序列,所述多肽包括EC 1.1.-.-的醇脱氢酶或者1.2.99的羧酸还原酶。所述ω-羟化酶具有经修饰的细胞色素P450单加氧酶(P450)酶 活性,并且在体内有效地催化碳氢化合物链的ω-位置。在一个实施方案中， 经修饰的ω-羟化酶为EC 1.14.15.3的ω-羟化酶杂交融合蛋白质变体。在另 一个实施方案中,经修饰的ω-羟化酶为CYP153A-还原酶杂交融合多肽。 在另一个实施方案中,CYP153A-还原酶杂交融合多肽是自给自足的 CYP153A-RedRhF杂交融合蛋白质。在另一个实施方案中,为ω-羟化酶杂交融合蛋白质变体的经修饰的ω-羟化酶与SEQ ID NO:6具有至少90％的 序列一致性,并且具有一个或多个突变,包括V141I、V141T、V141Q、 V141G、V141M、V141L、R27L、R82D、R178N、A231T、N309R、N407A、 V415R、T516V、P666A、P666D和/或A796V。所述重组微生物有机体产生ω-羟基脂肪酸衍生物,其包括但不限于ω-羟基脂肪酸衍生物,其为α,ω- 二醇。

[0014] 本发明公开进一步考虑了包含微生物有机体(如上所述)的细胞培养 物,其中所述的细胞培养物产生ω-羟基脂肪酸衍生物,包括但不限于ω- 羟基脂肪酸,包括ω-羟基游离脂肪酸;ω-羟基脂肪酸甲基酯;ω-氧代脂肪 酸;ω-氧代脂肪酸甲基酯;α,ω-二酸;α,ω-二醇;α,ω-二酯;ω-羧基脂肪酸 甲基酯;ω-氨基脂肪酸;和ω-氨基脂肪酸甲基酯。

[0015] 本发明公开的另一个方面提供了生产ω-羟基脂肪酸衍生物的方法,其 包括在发酵液中提供重组微生物有机体,所述微生物有机体具有经改造而 表达编码如下多肽的至少2种核酸序列的途径,所述多肽包括EC 3.1.2.-、 3.1.1.5或3.1.2.14的硫酯酶;或者EC 2.3.1.75或EC 2.3.1.20的酯合酶;以 及EC 1.14.15.3的经修饰的ω-羟化酶。所述方法进一步包括将包含碳源的 可再生的供料加入发酵液中,以及由发酵液分离ω-羟基脂肪酸衍生物。在一个实施方案中,经修饰的ω-羟化酶为EC 1.14.15.3的ω-羟化酶杂交融合 蛋白质变体。在另一实施方案中,经修饰的ω-羟化酶为CYP153A-还原酶 杂交融合多肽。在另一个实施方案中,CYP153A-还原酶杂交融合多肽是 自给自足的CYP153A-RedRhF杂交融合蛋白质。在另一个实施方案中,为 ω-羟化酶杂交融合蛋白质变体的经修饰的ω-羟化酶与SEQ ID NO:6具有 至少90％的序列一致性,并且具有一个或多个突变,包括V141I、V141T、V141Q、V141G、V141M、V141L、R27L、R82D、R178N、A231T、N309R、 N407A、V415R、T516V、P666A、P666D和/或A796V。在一个方面中,所产生的ω-羟基脂肪酸衍生物为ω-羟基游离脂肪酸或ω-羟基脂肪酸甲基 酯。

[0016] 本发明公开的另一个方面提供了生产ω-羟基脂肪酸衍生物的方法,其 包括在发酵液中提供重组微生物有机体,所述微生物有机体具有经改造而 表达编码如下多肽的至少2种核酸序列的途径,所述多肽包括EC 3.1.2.-、 3.1.1.5或3.1.2.14的硫酯酶;或者EC 2.3.1.75或EC 2.3.1.20的酯合酶;以 及EC 1.14.15.3的ω-羟化酶(细胞色素P450单加氧

酶)。在一个具体的方面 中,所述重组微生物有机体经改造而进一步表达编码如下多肽的核酸序 列,所述多肽包括EC 1.1.1.1/2的醇脱氢酶或者EC 1.1.3.13或EC 1.1.3.20 的醇氧化酶。所述方法进一步包括将包含碳源的可再生的供料加入发酵液 中,以及由发酵液分离ω-羟基脂肪酸衍生物。在一个方面中,所产生的ω- 羟基脂肪酸衍生物为ω-氧代脂肪酸或ω-氧代脂肪酸甲基酯。

[0017] 本发明公开的另一个方面提供了生产ω-羟基脂肪酸衍生物的方法,其 包括在发酵液中提供重组微生物有机体,所述微生物有机体具有经改造而 表达编码如下多肽的至少2种核酸序列的途径,所述多肽包括EC 3.1.2.-、3.1.1.5或3.1.2.14的硫酯酶;或者EC 2.3.1.75或EC 2.3.1.20的酯合酶;以 及EC 1.14.15.3的经修饰的ω-羟化酶。在一个具体的方面中,所述重组微 生物有机体经改造而进一步表达编码如下多肽的核酸序列,所述多肽包括 EC 1.1.1.1/2的醇脱氢酶或者EC 1.1.3.13或EC 1.1.3.20的醇氧化酶。在另 一个特定的方面中,所述重组微生物有机体经改造而进一步表达编码如下 多肽的核酸序列,所述多肽包括EC 1.2.1.3/4/5的醛脱氢酶或EC 1.2.3.1的醛氧化酶。所述方法进一步包括将包含碳源的可再生的供料加入发酵液 中,以及由发酵液分离ω-羟基脂肪酸衍生物。在一个方面中,所产生的ω- 羟基脂肪酸衍生物为α,ω-二酸或α,ω-脂肪酸二甲基酯。在一个实施方案中, 经修饰的ω-羟化酶为EC 1.14.15.3的ω-羟化酶杂交融合蛋白质变体。在另一个实施方案中,经修饰的ω-羟化酶为CYP153A-还原酶杂交融合多肽。 在另一个实施方案中,CYP153A-还原酶杂交融合多肽是自给自足的 CYP153A-RedRhF杂交融合蛋白质。在另一个实施方案中,为ω-羟化酶杂 交融合蛋白质变体的经修饰的ω-羟化酶与SEQ ID NO:6具有至少90％的 序列一致性,并且具有一个或多个突变,包括V141I、V141T、V141Q、V141G、V141M、V141L、R27L、R82D、R178N、A231T、N309R、N407A、 V415R、T516V、P666A、P666D和/或A796V。

[0018] 本发明公开的另一个方面提供了生产ω-羟基脂肪酸衍生物的方法,其 包括在发酵液中提供重组微生物有机体,所述微生物有机体具有经改造而 表达编码如下多肽的至少2种核酸序列的途径,所述多肽包括EC 3.1.2.-、3.1.1.5或3.1.2.14的硫酯酶;或者EC 2.3.1.75或EC 2.3.1.20的酯合酶;以 及EC 1.14.15.3的经修饰的ω-羟化酶。在一个具体的方面中,所述重组微 生物有机体经改造而进一步表达编码如下多肽的核酸序列,所述多肽包括 EC 1.1.1.1/2的醇脱氢酶或者EC 1.1.3.13或EC 1.1.3.20的醇氧化酶。在另 一个特定的方面中,所述重组微生物有机体经改造而进一步表达编码如下 多肽的核酸序列,所述多肽包括EC 2.6.1的氨基转移酶或者EC 1.4.9、EC 1.4.98或EC 1.4.99的胺脱氢酶。所述方法进一步包括将包含碳源的可再生的供料加入发酵液中,以及由发酵液分离ω-羟基脂肪酸衍生物。在一个方 面中,所产生的ω-羟基脂肪酸衍生物为ω-氨基脂肪酸或ω-氨基脂肪酸甲 基酯。在一个实施方案中,经修饰的ω-羟化酶为EC 1.14.15.3的ω-羟化酶杂交融合蛋白质变体。在另一个实施方案中,经修饰的ω-羟化酶为 CYP153A-还原酶杂交融合多肽。在另一个实施方案中,CYP153A-还原酶杂交融合多肽是自给自足的CYP153A-RedRhF杂交融合蛋白质。在另一个 实施方案中,为ω-羟化酶杂交融合蛋白质变体的经修饰的ω-羟化酶与SEQ ID NO:6具有至少90％的序列一致性,并且具有一个或多个突变,包括 V141I、V141T、V141Q、V141G、V141M、V141L、R27L、R82D、R178N、 A231T、N309R、N407A、V415R、T516V、P666A、P666D和/或A796V。

[0019] 本发明公开进一步考虑了生产ω-羟基脂肪酸衍生物的方法,其包括在 发酵液中提供重组微生物有机体,所述微生物有机体具有经改造而表达编 码如下多肽的至少2种核酸序列的途径,所述多肽包括EC 3.1.2.-、3.1.1.5 或3.1.2.14的硫酯酶;或者EC 2.3.1.75或EC 2.3.1.20的酯合酶;以及EC 1.14.15.3的经修饰的ω-羟化酶。在一个具体的方面中,所述重组微生物有 机体经改造而进一步表达编码如下多肽的核酸,所述多肽包括EC 1.2.99.6 的羧酸还原酶和/或EC 1.1.-.-的醇脱氢酶。所述方法进一步包括将包含碳源 的可再生的供料加入发酵液中,以及由发酵液分离ω-羟基脂肪酸衍生物。 在一个方面中,所产生的ω-羟基脂肪酸衍生物为α,ω-二醇。在一个实施方 案中,经修饰的ω-羟化酶为EC 1.14.15.3的ω-羟化酶杂交融合蛋白质变体。 在另一个实施方案中,经修饰的ω-羟化酶为CYP153A-还原酶杂交融合多 肽。在另一个实施方案中,CYP153A-还原酶杂交融合多肽是自给自足的 CYP153A-RedRhF杂交融合蛋白质。在另一个实施方案中,为ω-羟化酶杂交融合蛋白质变体的经修饰的ω-羟化酶与SEQ ID NO:6具有至少90％的 序列一致性,并且具有一个或多个突变,包括V141I、V141T、V141Q、 V141G、V141M、V141L、R27L、R82D、R178N、A231T、N309R、N407A、 V415R、T516V、P666A、P666D和/或A796V。

[0020] 本发明公开的另一个方面提供了方法(如上所述),其中可再生的供 料是碳基的,其包括但不限于玉米、甘蔗、高粱、甜菜、柳枝稷、新鲜秣 草、稻草、木材、纸浆、污水、垃圾、纤维素城市废物、废气、合成气和 二氧化碳。在一个方面中,碳源选自葡萄糖、果糖、甘露糖、半乳糖、木 糖、阿拉伯糖、低聚果糖、低聚半乳糖、淀粉、纤维素、胶质、木聚糖、 蔗糖、麦芽糖、纤维二糖、松二糖、半纤维素、甲基纤维素、羧甲基纤维素钠、琥珀酸酯、乳酸酯、乙酸酯、乙醇、甲醇、甘油和它们的混合物。

[0021] 本发明公开进一步考虑了通过本发明所公开的方法(如上所述)而生 产的聚合物组合物,其中所述聚合物组合物包含但不限于聚亚安酯、聚酯 多元醇、聚酯树脂、烷基涂料树脂、玻璃纤维树脂、凝胶涂料树脂和热塑 性聚酯(polyester thermoplastic)。

[0022] 本发明公开的另一个方面提供了在由可再生的供料得到的碳源存在 下,在发酵液中生长时,用于在体内生产ω-羟基脂肪酸衍生物的重组微生 物有机体,所述微生物有机体包含经改造而表达编码如下多肽的至少3种 核酸序列的途径,所述多肽包括EC 1.2.1.42的酰基-ACP还原酶;EC 1.1.-.- 的醇脱氢酶;以及EC 1.14.15.3的经修饰的ω-羟化酶。所述ω-羟化酶具有 经修饰的细胞色素P450单加氧酶(P450)酶活性,并且在体内有效地催化碳 氢化合物链的ω-位置。在一个实施方案中,经修饰的ω-羟化酶为EC 1.14.15.3的ω-羟化酶杂交融合蛋白质变体。在另一个实施方案中,经修饰 的ω-羟化酶为CYP153A-还原酶杂交融合多肽。在另一个实施方案中, CYP153A-还原酶杂交融合多肽是自给自足的CYP153A-RedRhF杂交融合 蛋白质。在另一个实施方案中,为ω-羟化酶杂交融合蛋白质变体的经修饰 的ω-羟化酶与SEQ ID NO:6具有至少90％的序列一致性,并且具有一个 或多个突变,包括V141I、V141T、V141Q、V141G、V141M、V141L、 R27L、R82D、R178N、A231T、N309R、N407A、V415R、T516V、P666A、 P666D和/或A796V。所产生的ω-羟基脂肪酸衍生物为α,ω-二醇。在一个 方面中,还提供了细胞培养物,其包含本发明所公开的微生物有机体(如 上所述)。

[0023] 然而,本发明公开的另一个方面提供了生产ω-羟基脂肪酸衍生物的方 法,其包括在发酵液中提供重组微生物有机体,所述微生物有机体包含经 改造而表达编码如下多

肽的至少3种核酸序列的途径,所述多肽包括EC 1.2.1.42的酰基-ACP还原酶;EC 1.1.-.-的醇脱氢酶;以及EC 1.14.15.3的 经修饰的ω-羟化酶。所述方法进一步包括将包含碳源的可再生的供料加入发酵液中,以及由发酵液分离ω-羟基脂肪酸衍生物。在一个方面中,经修 饰的ω-羟化酶为EC 1.14.15.3的ω-羟化酶杂交融合蛋白质变体。在另一个 实施方案中,经修饰的ω-羟化酶为CYP153A-还原酶杂交融合多肽。在另一个实施方案中,CYP153A-还原酶杂交融合多肽是自给自足的 CYP153A-RedRhF杂交融合蛋白质。在另一个实施方案中,为ω-羟化酶杂 交融合蛋白质变体的经修饰的ω-羟化酶与SEQ ID NO:6具有至少90%的 序列一致性,并且具有一个或多个突变,包括V141I、V141T、V141Q、 V141G、V141M、V141L、R27L、R82D、R178N、A231T、N309R、N407A、 V415R、T516V、P666A、P666D和/或A796V。在本发明中,ω-羟基脂肪酸衍生物为α,ω-二醇。在一个方面中,可再生的供料是碳基的,并且包括 玉米、甘蔗、高粱、甜菜、柳枝稷、新鲜秣草、稻草、木材、纸浆、污水、垃圾、纤维素城市废物、废气、合成气和二氧化碳。在另一个方面,碳源选自葡萄糖、果糖、甘露糖、半乳糖、木糖、阿拉伯糖、低聚果糖、低聚半乳糖、淀粉、纤维素、胶质、木聚糖、蔗糖、麦芽糖、纤维二糖、松二糖、半纤维素、甲基纤维素、羧甲基纤维素钠、琥珀酸酯、乳酸酯、乙酸酯、乙醇、甲醇、甘油和它们的混合物。通过所述方法生产的聚合物组合物也进一步包含在本发明中,其中所述聚合物组合物包括但不限于聚亚安酯、聚酯多元醇、聚酯树脂、烷基涂料树脂、玻璃纤维树脂、凝胶涂料树脂和热塑性聚酯。

[0024] 本发明公开进一步考虑了通过本文所述方法(如上所述)而生产的香 精化学组合物,其中所述的香精化学组合物为C9至C16的饱和的或不饱 和的大环内酯。所述香精化学组合物包括选自麝香内脂(ambrettolide)、 二氢麝香内脂、15-羟基十五酸的大环内酯、15-羟基十五碳烯酸的大环内 酯和/或其他中的化学实体。

[0025] 附图简述

[0026] 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。

[0027] 图1描绘了用于制备ω-羟基-羧酸、ω-氧代-羧酸、ω-氨基-羧酸和α,ω- 二酸的途径。将具有12个碳原子的脂肪酸衍生物作为实例来描述。

[0028] 图2A描绘了通过甲基酯中间体来制备ω-羟基-羧酸、ω-氨基-羧酸和 α,ω-二酸的途径。将具有12个碳原子的脂肪酸衍生物作为实例来描述。

[0029] 图2B描绘了通过二甲基酯中间体来制备α,ω-二酸的途径。将具有12 个碳原子的脂肪酸衍生物作为实例来描述。

[0030] 图3描绘了使用硫酯酶和羧酸还原酶来制备α,ω-二醇的途径。将具有 12个碳原子的脂肪酸衍生物作为实例来描述。

[0031] 图4描绘了使用酰基-ACP还原酶来制备α,ω-二醇的途径。将具有12 个碳原子的脂肪酸衍生物作为实例来描述。

[0032] 图5示出了将ω-氨基羧酸转化成内酰胺的途径。将具有12个碳原子的脂肪酸衍生物作为实例来描述。

[0033] 图6示出了由表达cyp153A操纵子的重组大肠杆菌菌株得到的提取物 的GC/MS色谱,所述菌株在供入十二烷酸时会形成12-羟基十二烷酸。此 外,还示出了由对照菌株

(MG1655ΔfadD)得到的提取物的GC/MS色谱。

[0034] 图7A至7C示出了衍生的12-羟基十二烷酸(在12.297分钟时达到峰 值)(图7A)、衍生的12-羟基十二烷酸真实标准品(图7B)、和非衍生的12- 羟基十二烷酸(在11.393分钟时达到峰值)(图7C)的质谱。衍生化试剂为 BSTFA+1%TMCS。图7A和7C中的质谱得自大肠杆菌菌株sAS.321的提 取物。

[0035] 图8A至8B示出了衍生的12-羟基十二烷酸(12-三甲基甲硅氧基十二 烷酸三甲基甲硅烷基酯)(图8A)和非衍生的12-羟基十二烷酸(图8B)的离子断裂方式。

[0036] 图9示出了由表达CYP153A操纵子的重组大肠杆菌菌株得到的提取 物的GC/MS色谱,所述菌株在供入十二烷酸时会形成12-羟基十二烷酸。 菌株sAS.335表达了改良的CYP153A操纵子。

[0037] 图10A至10C示出了由表达CYP153A-RedRhF融合蛋白质的大肠杆 菌菌株得到的提取物的GC/MS色谱,所述菌株由十二烷酸形成12-羟基十 二烷酸(图10A),由十二烷酸甲基酯形成12-羟基十二烷酸甲基酯(图10B), 以及由十二烷醇形成1,12-十二烷二醇(图10C)。此外,还示出了对照菌株 MG1655ΔfadD的提取物。

[0038] 图11A至11B描绘了由菌株sAS.336的提取物得到的衍生的1,12-十 二烷二醇(在11.948分钟时达到峰值)(图11A),以及真正的衍生的1,12- 十二烷二醇标准品(图11B)的质谱。衍生化试剂为BSTFA+1%TMCS。

[0039] 图12A至12B描绘了由菌株sAS.336的提取物得到的12-羟基十二烷 酸甲基酯的质谱。其中示出了非衍生(在11.662分钟时达到峰值)(图12A) 的和衍生(在11.107分钟时达到峰值)(图12B)的样品。衍生化试剂为BSTFA+1%TMCS。

[0040] 图13A至13B示出了无衍生化的1,12-十二烷二醇(图13A)和12-羟基 十二烷酸甲基酯(图13B)的离子断裂方式。

[0041] 图14A至14C示出了由表达CYP153A-RedRhF融合蛋白质的重组大 肠杆菌菌株得到的提取物的GC/MS色谱,所述菌株由葡萄糖产生ω-OH脂 肪酸(图14A)、ω-OH脂肪酸甲基酯(图14B)或α,ω-二醇(图14C)。所有 样品是使用BSTFA+1%TMCS进行衍生化的。

[0042] 图15证明从生产细胞有效地分泌ω-羟基脂肪酸。

[0043] 图16显示2种大肠杆菌菌株在葡萄糖上生长时产生ω-羟基脂肪酸的 组合物。

[0044] 图17示出了由重组大肠杆菌菌株得到的提取物的GC/MS色谱,所述 菌株表达CYP153A-RedRhF融合蛋白质、醇氧化酶和醛脱氢酶,从而产生 α,ω-二酸。所有样品是使用BSTFA+1%TMCS进行衍生化的。

[0045] 图18A至18B描绘了由菌株sEP.690的提取物得到的衍生的1,14-十四 烷二酸(在13.367分钟时达到峰值)(图18A),以及真正的衍生的1,14-十 四烷二酸标准品(图18B)的质谱。衍生化试剂为BSTFA+1%TMCS。

[0046] 图19A示出了衍生的1,14-十四烷二酸的离子断裂方式。衍生化试剂 为BSTFA+1%TMCS。

[0047] 图20示出了由表达cyp102A1(F87A)的重组大肠杆菌菌株得到的提取 物的GC/MS色谱,所述菌株产生少量的近末端的(例如ω-1、ω-2和/或ω-3) 羟基化的脂肪酸。此外,还示出了对照菌株AlcV334的提取物的GC/MS 色谱。所述的样品是使用BSTFA+1%TMCS进行衍生化的。

[0048] 图21A至21E描绘了RT 7.195至7.510(得自图20)的峰的质谱,这些 峰识别为11-羟基十二烷酸(图21A)、10-羟基十二烷酸(图21B)、9-羟基 十二烷酸(图21C)、8-羟基十二烷酸(图21D)、7-羟基十二烷酸(图21E)。针对各羟基化位置来描绘诊断离子断裂。样品是使用BSTFA+1%TMCS 进行衍生化的。

[0049] 图22A至22B示出了由表达cyp102A7-蛋白质的重组大肠杆菌菌株的 提取物的GC/MS色谱,所述菌株由葡萄糖产生ω-1、ω-2和ω-3羟基脂肪 酸(图22A)、或者ω-1、ω-2和ω-3羟基脂肪醇(图22B)。所有样品是使用 BSTFA+1%TMCS进行衍生化的。

[0050] 图23A至23C示出了衍生的9-羟基十二烷醇(在9.653分钟时达到峰 值)(图23A)、衍生的10-羟基十二烷醇(在9.808分钟时达到峰值)(图23B)、和衍生的11-羟基十二烷醇(在9.905分钟时达到峰值)(图23C)的质谱和离子断裂方式。衍生化试剂为BSTFA+1%TMCS。样品得自大肠杆菌菌株 XL963的提取物。

[0051] 图24A至24B示出了表达cyp102A7的重组大肠杆菌菌株由葡萄糖生 产的近末端(例如ω-1、ω-2和/或ω-3)羟基化的脂肪酸(图24A)和近末端 羟基化的脂肪醇(图24B)的量。

[0052] 图25示出了由大肠杆菌菌株stEP.798生产的ω-羟基化的脂肪酸。

[0053] 图26示出了由大肠杆菌菌株L1017生产的α,ω-二酸。

[0054] 图27示出了培养表达CYP153A-还原酶杂交融合蛋白质变体的微生物 有机体的结果。如所示,具有V141T(SEQ ID:46)的变体产生了最高的ω- 羟基十六碳烯酸效价,以及由十六碳烯酸的最高的转化。

[0055] 发明详述

[0056] 综述

[0057] 研发用于生产ω-羟基脂肪酸衍生物的新的且环境友好的方法意味着 对工业具有显着的改良。所述方法可以使这些化合物从由可再生的供料衍 生的简单碳源有效地生产。具体而言,所述方法提供了从可再生的材料(例 如由玉米、藤条或木质纤维素生物质得到的碳水化合物),或废品(例如 甘油、废气、合成气),或改进的有机材料(例如生物质、天然气或二氧 化碳)生产ω-羟基脂肪酸衍生物。

[0058] 更具体而言,本公开提供了新型重组微生物有机体,该有机体经改造 而将可再生的供料(例如碳水化合物)转化成特定的ω-羟基脂肪酸衍生物, 其包括ω-羟基脂肪酸;ω-羟基-脂肪酸甲基酯;ω-羧基-脂肪酸甲基酯;ω- 氧代脂肪酸,ω-氨基脂肪酸,ω-氨基脂肪酸甲基酯,α,ω-二酸;α,ω-二酯;α,ω- 二醇等。这样,双功能分子包括但不限于ω-羟基脂肪酸;α,ω-羟基脂肪醇;α,ω-羟基-脂肪酸甲基酯;α,ω-羟基胺;α,ω-二酸;α,ω-二脂肪酸甲基酯,α,ω- 二醇等。所述重组微生物有机体可以进行成本有效的发酵工艺用于生产这 些化合物。本发明公开涵盖微生物脂肪酸代谢以及将其中间体转化成特定 的化学物。

[0059] 本发明公开的优点有很多。本发明公开提供了更简单的生产方法,即,使用简单的发酵过程,而非多个化学的和/或生物催化的工艺,由于所述的 发酵过程产生了更少的废品,所以其更快速、成本更低且更环境友好。使 用可再生的供料(可持续的原材料)和/或工业废品(例如甘油)作为来源材料增加了另一个成本益处,并保护了环境。本发明公开提供了特定目标 产物(即,包含具有选择性链长和化学作用的化学实体的组合物)的选择 性制

造。获得多样的化学功能性可以用于新的目标市场的应用中。

[0060] 定义

[0061] 如本文所用,术语“omega-羟基化的脂肪酸衍生物”、“ω-羟基化的脂肪 酸衍生物”和“ω-羟基脂肪酸衍生物”和“ω-羟基脂肪酸衍生物”和“ω-OH脂 肪酸衍生物”在本发明中可以交换使用,并且是指来源于脂肪酸代谢,并且 在omega位置处具有至少一个OH基团或者由在omega位置处具有至少一 个OH基团的中间体衍生的化学实体。在本发明中,“omega位置”是指就 其伯官能团而言,在其相反的端处脂肪酸衍生物的末端碳原子。此类ω-羟基脂肪酸衍生物包括但不限于ω-羟基脂肪酸;ω-羟基-脂肪酸甲基酯;ω- 羧基-脂肪酸甲基酯;ω-氧代脂肪酸;ω-氨基脂肪酸;ω-氨基脂肪酸甲基酯; 以及α,ω-二酸;α,ω-二酯;和α,ω-二醇。术语“ω-羟基脂肪酸衍生物”包 括“α,ω-双官能的脂肪酸衍生物”。

[0062] “ω-羟基脂肪酸衍生物组合物”在本发明中是指通过重组宿主细胞生产 的,并且通常包含具有不同链长和/或饱和度和/或支化特征的特定类型的ω-羟基脂肪酸衍生物的混合物(例如不同链长和/或饱和度和/或支化特征 的α,ω-二酸;或不同链长和/或饱和度和/或支化特征的α,ω-二酯;或不同 链长和/或饱和度和/或支化特征的α,ω-二醇等)。在一些情况下,ω-羟基脂 肪酸衍生物组合物包含大部分一种类型的ω-羟基脂肪酸衍生物,例如1,12- 十二烯二醇,或1,14-十四烷二醇,或16-羟基十六烷酸甲基酯,或16-羟 基-十六碳烯酸,或15-羟基-十五酸,或15-羟基十五碳烯酸,或18-羟基十 八烯酸,或者这些脂肪酸衍生物的任意一种的甲基酯等。在其他情况下, ω-羟基脂肪酸衍生物组合物包含多于一种类型的ω-羟基脂肪酸衍生物的 混合物,以便提供具体设计的组合物(例如在相同的组合物中具有大约20% 1,12-十二烯二醇和大约80%1,16-十六烷二醇可以作为此类实例)。

[0063] 术语“近末端的”羟基化的脂肪酸衍生物是指在omega-1位置,和/或 omega-2位置,和/或omega-3位置,和/或omega-4位置等(例如ω-1,ω-2 和/或ω-3等)处具有至少一个OH基团(或者由具有至少一个OH基团的 中间体衍生得到)的化学实体。示例性的种类为ω-1,ω-2和/或ω-3-羟基脂 肪酸;或ω-1-羟基脂肪酸甲基酯;或者ω-1,ω-2,ω-3,ω-4,和/或ω-5-羟基 十二烷酸等。

[0064] 术语“酶分类(EC)号”是指表示特定酶活性的编号。EC编号在酶的 命名系统下,根据它们所催化的反应将酶分类。EC编号指定了酶催化的 反应。例如如果由不同的有机体得到的不同的酶催化了相同的反应,则它 们具有相同的EC编号。此外,不同的蛋白质折叠物可以催化相同的反应,由此将分配相同的EC编号(例如非同源的双功能酶或NISE)。EC编号是 由生物国际化学和分子生物学联合会的命名委员会(IUBMB)建立的,其描 述可以在互联网的IUBMB酶命名系统的网站上得到。例如细胞色素P450 单加氧酶(P450)酶活性(包括ω-羟化酶或ω-氧化酶酶活性)分类在EC 1.14.15.3(也称为长链酰基-[酰基-载体-蛋白质]还原酶)之下。属于P450 酶家族的酶的功能性在一个物种至下一物种的大部分的原核生物中是保守的。因此,不同的微生物物种可以具有分类为EC 1.14.15.3之下的相同 的酶活性。表征为EC 1.14.15.3的酶活性的实例为本发明所讨论的 CYP153A-还原酶杂交融合多肽或其变体的酶活性(如上所述)。

[0065] 术语“EC 1.14.15.3的经修饰的ω-羟化酶”和“修饰的ω-羟化酶”在本发 明中可

以交换使用,并且是指在体内(例如在微生物有机体中)有效地催 化碳氢化合物链的ω位置的细胞色素P450单加氧酶酶活性。

[0066] 术语“EC 1.14.15.3的ω-羟化酶杂交融合蛋白质变体”和“ω-羟化酶杂交 融合蛋白质变体”在本发明中可以交换使用,并且是指在其氨基酸序列中具 有至少一个突变的经修饰的ω-羟化酶杂交融合多肽,从而在重组宿主细胞 中表达ω-羟化酶杂交融合蛋白质变体,由此得到与在相应的宿主细胞中天然P450融合蛋白质的表达相比,效价、产率和/或生产率得到改良的ω-OH 脂肪酸和/或ω-OH脂肪酸衍生物组合物。例如当使用ω-羟化酶杂交融合蛋 白质变体来转化细胞时,该细胞可以表达ω-羟化酶杂交融合蛋白质变体(例如重组细胞)。在一个实施方案中,由表达ω-羟化酶杂交融合蛋白质 变体的细胞所产生的ω-OH脂肪酸的效价和/或产率是表达天然P450融合 蛋白质的相应细胞的至少2倍。在另一个实施方案中,由表达ω-羟化酶杂 交融合蛋白质变体的细胞所产生的ω-OH脂肪酸或其衍生物的效价和/或产 率是表达天然P450融合蛋白质的相应细胞的效价和/或产率的至少大约1 倍、至少大约2倍、至少大约3倍、至少大约4倍、至少大约5倍、至少 大约6倍、至少大约7倍、至少大约8倍、至少大约9倍或至少大约10 倍。在一个实施方案中,由表达ω-羟化酶杂交融合蛋白质变体的细胞所产 生的ω-OH脂肪酸或其衍生物的效价和/或产率高于表达天然P450融合蛋 白质的相应细胞的效价和/或产率的至少大约1%、至少大约2%、至少大约 3%、至少大约4%、至少大约5%、至少大约6%、至少大约7%、至少大 约8%、至少大约9%或至少大约10%。在另一个实施方案中,由于ω-羟 化酶杂交融合蛋白质变体的表达而使得重组细胞中产生的ω-OH脂肪酸或 其衍生物的效价和/或产率高于表达天然P450融合蛋白质的相应细胞的效 价和/或产率的至少大约20%至至少大约80%。在一些实施方案中,由所述 的细胞产生的ω-OH脂肪酸的效价和/或产率高于表达天然P450融合蛋白质的相应细胞的效价和/或产率的至少大约至少大约20%、至少大约25%、 至少大约30%、至少大约35%、至少大约40%、至少大约45%至少大约 50%、至少大约55%、至少大约60%、至少大约65%、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%、至少大约97%、至少大约98%或至少大约100%。

[0067] 术语“CYP153A-还原酶杂交融合多肽”是指与SEQ ID NO:6具有至少 70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、 82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、 94%、95%、96%、97%、98%、99%或100%序列一致性的多肽序列。CYP153A- 还原酶杂交融合多肽是自给自足的,并且具有催化脂肪酸反应形成ω-OH 脂肪酸的ω-羟化酶的酶活性。CYP153A-还原酶杂交融合多肽的实例为杂 交的cyp153A-RedRhF型融合多肽。

[0068] 术语“登录号”、“NCBI登录号”或“GenBank登录号”是指表示具体的核 酸序列的编号。在本说明书中所讨论的序列登录号得自于NCBI(National Center for Biotechnology Information)提供的数据库,由the National Institutes of Health,U.S.A.维护,以及得自于由the Swiss Institute of Bioinformatics提供的the UniProt Knowledgebase(UniProtKB)和Swiss-Prot 数据库(也称为UniProtKB登录号)。

[0069] 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)。

[0070] 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经修饰的多核苷酸(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。

[0071] 如本文所用,术语"多肽"和"蛋白质"可以交换使用,其是指氨基酸残 基的聚合物。术语"重组多肽"是指通过重组技术生产的多肽,其中通常, 将编码所表达蛋白质的DNA或RNA插入到合适的表达载体中,该载体进而转化宿主细胞,从而生产多肽。

[0072] 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。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、所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的函数,其考虑了空位的数量和各空位的长度, 必须引入空位的数量和各空位的长度以达到两个序列的最佳比对的目的。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.(1990)J.Mol.Biol.215(3):403-410)。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,(1970)J.Mol.Biol.48:444-453)。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(2005) BMC Bioinformatics 6:278);Altschul et al.(2005)FEBS J.272(20):5101-5109)。

[0073] “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。

[0074] “外源的”多肽是指并非由亲代微生物细胞的基因组编码的多肽。变体 (即,突变体)多肽为外源多肽的实例。

[0075] 术语“异源的”通常是指由不同的物种衍生得到或由不同的有机体衍生 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经实验加入到通常不表达 该蛋白质的细胞中。因此,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对于植物细胞内源的多核苷酸序列引入到细菌宿主细胞中,然后该植物多核苷酸在重组细菌宿主细胞中为异源多核苷酸。

[0076] 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。

[0077] 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。

[0078] 如本文所用,术语“基因”是指编码RNA产物或蛋白质产物的核酸序 列;以及影响RNA或蛋白质的表达的可操作地连接的核酸序列(例如此类序列包括但不限于启动子或增强子序列)或编码影响RNA或蛋白质的 表达的序列的、可操作地连接的核酸序列(例如此类序列包括但不限于核 糖体结合位点或翻译控制序列)。

[0079] 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.(1987) Science 236:1237-1245)。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。

[0080] 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。

[0081] 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。能够使与其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。在原核生物(例如 大肠杆菌)中表达编码多肽的基因最通常的是使用包括构成型或诱导型启 动子的载体来实施,其中所述的启动子指导融合或非融合多肽的表达。融 合载体将一定数量的氨基酸加入到其中编码的多肽中,通常加入到重组多 肽的氨基末端或羧基末端。这种融合载体通常起到以下3种目的中的一种 或多种:(1)增加重组多肽的表达;(2)增加重组多肽的溶解性;以及(3)通 过在亲和纯化中起到配体的作用而有助于重组多肽的纯化。通常,在融合表达的载体中,在融合部分与重组多肽的连接处引入蛋白水解的切割位 点。这使得在纯化融合多肽后,重组多肽与融合部分分离。在某些实施方 案中,本发明公开的多核苷酸序列与由噬菌体T5衍生的启动子可操作地 连接。在某些实施方案中,宿主细胞为酵母细胞,表达载体为酵母表达载 体。用于在酿酒酵母(S.cerevisiae)中表达的载体的实例包括pYepSecl(Baldari et al.(1987)EMBO J.6:229-234)、pMFa(Kurjan et al.(1982)Cell30:933-943)、pJRY88(Schultz et al.(1987)Gene 54:113-123)、pYES2(Invitrogen Corp.,San Diego,CA)和picZ(Invitrogen Corp.,San Diego, CA)。在其他的实施方案中,宿主细胞为昆虫细胞,表达载体为杆状病毒 表达载体。可用于在培养的昆虫细胞(例如Sf9细胞)中表达蛋白质的杆 状病毒载体包括例如pAc系列(Smith et al.(1983)Mol.Cell Biol. 3:2156-2165)和pVL系列(Lucklow et al.(1989)Virology 170:31-39)。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)。

[0082] 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’- 磷酸泛酰亚巯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。

[0083] 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。

[0084] 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 (包括ω-羟基化的脂肪酸衍生物)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(例如ω-羟基化的脂肪酸衍生物)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。

[0085] 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。

[0086] 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”是指使重组宿主细胞群 体在液体或固体培养基中在合适的条件下生长。在具体的实施方案中,培养是指底物以发酵方式生物转化为终产物。培养用的培养基是公知的,并且此类培养物培养基的单个成分可得自商业来源,例如Difco™和BBL™商标。在一个非限定性的实例中,水性营养培养基为包括氮、盐和碳的复 合来源的“富集培养基”,例如YP培养基,其包括此类培养基的10g/L蛋 白胨以及10g/L酵母提取物。根据美国专利5,000,000;5,028,539;5,424,202; 5,482,846;5,602,030;WO 2010127318中所述方法,培养物的宿主细胞可以另外改造,从而有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。

[0087] 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ω-羟基脂肪酸衍生物的任何条件。合适的条件 包括例如发酵条件。

[0088] 术语“重组微生物有机体”是指宿主细胞经基因改造,使得宿主细胞中 的某些酶活性相对于亲代细胞或天然宿主细胞被改变、加入和/或删除。遗 传修饰的宿主细胞为重组微生物有机体的实例。如此,在重组宿主细胞中 的“修饰的或改变的蛋白质活性水平”(例如酶)是指相对于亲代或天然宿 主细胞(其中缺乏相同的修饰)而确定的活性中一种或多种特征的差异。 通常,活性的差异是在重组宿主细胞(具有经修饰的活性)与相应的野生型宿主细胞(不具有经修饰的活性)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,多肽 为本发明所述的任何多肽的突变体或变体。在某些实施方案中,用于本发 明所述多肽的编码序列为针对具体的宿主细胞的表达而优化的密码

子。例 如为了在大肠杆菌中表达,可以优化一种或多种密码子(例如在Grosjean et al.(1982)Gene 18:199-209中所述)。

[0089] 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)。

[0090] 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、代谢物或产物(例如ω-羟基脂肪酸 衍生物)在相应的野生型细胞中的浓度相比,多核苷酸、多肽、代谢物或 产物在改造的宿主细胞中以不同的浓度存在。

[0091] 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ω- 脂肪酸衍生物的定量。效价可以指由给定的重组宿主细胞培养物所生产的 具体的ω-脂肪酸衍生物或ω-脂肪酸衍生物的组合。

[0092] 如本文所用,“由宿主细胞生产的ω-羟基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(例如ω-羟基脂肪酸,α,ω-二酸等) 的效率。产率可以指由给定的重组宿主细胞培养物所生产的具体的ω-脂肪 酸衍生物或ω-脂肪酸衍生物的组合。ω-羟基脂肪酸衍生物包括但不限于ω- 羟基脂肪酸,包括ω-羟基游离脂肪酸;ω-羟基脂肪酸甲基酯;ω-氧代脂肪 酸;ω-氧代脂肪酸甲基酯;双功能化合物,例如α,ω-二酸;α,ω-二醇;α,ω- 二酯;ω-羧基脂肪酸甲基酯;ω-氨基脂肪酸;和ω-氨基脂肪酸甲基酯。

[0093] 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ω-羟基脂肪酸衍生物或衍生物的定量。生产率可以指由给定的 重组宿主细胞培养物所生产的具体的ω-脂肪酸衍生物或脂肪酸衍生物的 组合。

[0094] 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)。

[0095] 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CO_2)。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;琥珀 酸酯、乳酸酯和醋酸酯;醇,例如乙醇、甲醇和甘油;或它们的混合物。此外,碳源还可以为光合成产物,例如葡萄糖。在某些实施方案中,碳源 为生物质。在其他的实施方案中,碳源为葡萄糖。在其他的实施方案中, 碳源为蔗糖。

[0096] 当术语“得自可再生的供料的碳源”单独使用或者参照供料来源使用 时,其是指包括除了油脂化学品(即,由植物和动物得到的精制的油,例 如脂肪酸、脂肪酸酯、TAG、羟基脂肪酸等)和石油化学品(即,衍生自 石油的化学品,例如烷、烯等)以外的、可以衍生出碳的任何生物材料(包 括可再生的供料和/或生物质和/或废品)。因此,如本文所用,术语“衍生自可再生的供料的碳源”排除了衍生自油脂化学品和石油化学品的碳。在一 些实施

方案中,碳源包括糖或碳水化合物(例如单糖、二糖或多糖)。在 一些实施方案中,碳源为葡萄糖和/或蔗糖。在其他的实施方案中,碳源衍 生自可再生的供料,例如得自玉米、甘蔗或木质纤维素生物质的碳水化合 物;废品,例如甘油、废气、合成气;改进的有机材料,例如生物质或天 然气;或者碳源为通过光合成而固定的二氧化碳。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,而可以直接用作碳源。此类生物质的示例性来源为植 物物质或蔬菜,例如柳枝稷。另一种示例性的碳源包括代谢废物,例如动 物物质(例如牛粪便)。其他的示例性碳源包括海藻和其他海生植物。另一种碳源(包括生物质)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、发酵生物质、甘油/丙三醇、新鲜秣草、稻草、木材、 污水、垃圾、独立固体废物、纤维素城市废物和剩饭剩菜。

[0097] 如本文所用,对产物(例如ω-羟基脂肪酸衍生物)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。通过本发明所述方法生产的ω-羟基脂肪酸衍生物和ω-羟基脂肪酸衍生物组合物在发酵液及细胞质中可以是相对不可混合的。因此,所述的 ω-羟基脂肪酸衍生物及和ω-羟基脂肪酸衍生物组合物可以在细胞内或细 胞外收集于有机相中。在一个优选的方法中,ω-羟基脂肪酸衍生物及和ω- 羟基脂肪酸衍生物组合物在细胞外收集。

[0098] 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(separation)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%不含(例如至少大约70%不含、 至少大约75%不含、至少大约85%不含、至少大约90%不含、至少大约 95%不含、至少大约97%不含、至少大约99%不含)它们所关联的其他成 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ω-羟基脂肪酸衍生物时,可以通过去除宿主细胞蛋白质来纯化ω-羟基脂肪 酸衍生物。纯化后,样品中ω-羟基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ω-羟基脂肪酸衍生 物时,纯化的ω-羟基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)基本隔离的ω-羟基脂肪酸衍生 物。

[0099] 说明书和权利要求书中使用的术语“在体内产生ω-羟基脂肪酸衍生 物”是指在活的和/或基因改造的宿主细胞中由诸如碳水化合物之类的可再 生的供料来产生ω-羟基脂肪酸衍生物,其中所述的可再生的供料作为碳源 加入发酵液中,这样宿主细胞可以在发酵过程中吸收并代谢碳源。这与在体外产生ω-羟基脂肪酸衍生物的方法不同,在该方法中,使用纯化的酶或 细胞裂解物,并且用于酶转化的直接底物(例如脂肪酸或脂肪酸衍生物) 加入纯化的酶中或加入细胞裂解物溶液中。此外,本发明的方法还不同于 其中在生物转化中产生ω-羟基脂肪酸衍生物的方法,在该方法中,使用休 眠细胞,并且用于酶转化的直接底物(例如脂肪酸或脂肪酸衍生物)以外 源方式加入休眠细胞中。

[0100] 途径的改造和酶的活性

[0101] 脂肪酸合成是细菌生物合成机器的最保守的系统之一。脂肪酸合酶 (FAS)多酶复合物存在于所有细菌和真核细胞中。大部分的FAS相关基 因对于细胞生长和生存而言是必不可少的。真核和细菌FAS本质上驱动了 相同类型的生物化学转化。在真核细胞中,FAS称

为FAS I,并且其大部 分的催化结构域由一个多肽链(不可分离的)编码。在诸如细菌之类的原 核生物中,FAS称为FAS II,其单一的酶以及载体蛋白质由编码离散的(分 离的)蛋白质的分开基因所编码。如此,FAS II为具有明显改变和不同特 性的复合物系统。

[0102] 酰基载体蛋白质(ACP)与FAS途径中的酶一起控制了在天然有机体 中生产的脂肪酸的长度、饱和度和支化。该途径中的步骤是由脂肪酸生物 合成(FAB)和乙酰基-CoA羧化酶(ACC)基因家族的酶所催化的。例如可 以包含在FAS途径中的酶包括AccABCD、FabD、FabH、FabG、FabA、 FabZ、FabI、FabK、FabL、FabM、FabB和FabF。根据所需的产物,可以 减弱或过表达这些基因中的一种或多种。这样,原核生物经改造从而增加了由可再生的供料(例如葡萄糖或其他碳源)形成脂肪酸衍生物的生产。 在本发明中,主要的目标是增加关键性控制酶(其调节脂肪酸衍生物的生产)的活性,从而将细菌菌株转化成用于脂肪酸衍生物(包括脂肪酸甲酯 (FAME)、脂肪酸乙酯(FAEE)和脂肪醇(FALC))生产的微生物工厂(例如参见美国专利No.8,283,143,该文献以引用方式并入本文)。

[0103] 本发明公开鉴定了编码酶功能的多肽的多核苷酸,从而修饰了用于生 产所需的化合物(例如ω-羟基脂肪酸衍生物)的酶途径。在本发明中通过 酶登录号(EC编号)识别的这些多肽可以用于改造脂肪酸途径,这可以产生双功能分子,例如ω-羟基脂肪酸衍生物。图1-4描绘了经改造而生产 这些化合物的途径。图5描绘了形成内酰胺和聚合物(例如尼龙)的途径。

[0104] 在一个实施方案中,图1至4描绘了使用诸如葡萄糖之类的可再生的 供料来生产ω-羟基脂肪酸衍生物的途径。葡萄糖通过天然有机体而转化成 酰基-ACP(参见图1至4中的步骤1)。编码具有脂肪酸降解酶活性的多 肽的多核苷酸可以根据所需的产物而可任选地减弱(参见下文的实施例)。 此类多肽的非限定性实例为酰基-CoA合酶(FadD)和酰基-CoA脱氢酶(FadE)。表1提供了代谢途径中的酶活性(见下文)的大量列举物,包括 可以根据本领域已知的方法可任选地减弱的多种脂肪酸降解酶(例如参见 美国专利No.8,283,143,见下文)。

[0105] 例如FadR(参见表1,见下文)为与脂肪酸降解和脂肪酸生物合成途 径有关的重要调节因子(Cronan et al.,Mol.Microbiol.,29(4):937-943 (1998))。大肠杆菌酶FadD(参见表1,见下文)和脂肪酸运送蛋白质FadL 为脂肪酸摄入系统的组分。FadL介导了脂肪酸向细菌细胞的运送,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 菌摄入并转化成酰基-CoA酯,其可以与转录因子FadR结合,并抑制fad 基因的表达,所述fad基因编码负责脂肪酸运送(FadL)、活化(FadD)和β- 氧化(FadA,FadB和FadE)的蛋白质。当可以利用备选碳源时,细菌合成脂 肪酸作为酰基-ACP,其用于磷脂合成,但不能作为β-氧化的底物。因此,酰基-CoA和酰基-ACP都是脂肪酸的独立的来源,其可以得到不同的终产 物(Caviglia et al.,J.Biol.Chem.,279(12):1163-1169(2004))。

[0106] 表1:酶的活性

[0107]

基因名称	来源有机体	酶的名称	登录号#	EC 编号	示例性用途
脂肪酸生产的增加					
accA	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
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 *W3110*	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 *K12*	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 *K12*	[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 *K12*	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 *K12*	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 *K12*	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 *K12*	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 *K12*	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（*Vibrio cholerae*）	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 *K12*	(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 *K13*	酰基-CoA 脱氢酶	AAC73325	1.3.99.3, 1.3.99.-	减少脂肪酸的降解
fadD	大肠杆菌 *K12*	酰基-CoA 合酶	NP_416319	6.2.1.3	减少脂肪酸的降解

[0108]

基因名称	来源有机体	酶的名称	登录号#	EC 编号	示例性用途
fadA	大肠杆菌 *K12*	3-酮酰基-CoA 硫解酶	YP_02627	2.3.1.16	减少脂肪酸的降解
fadB	大肠杆菌 *K12*	烯酰基-CoA 水合酶, 3-OH 酰基-CoA 差向异构体/脱氢酶	NP_418288	4.2.1.17, 5.1.2.3, 1.1.1.35	减少脂肪酸的降解
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
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	C18:1 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（*Umbellularia californica*）	硫酯酶	Q41635	3.1.2.14	C12:0 链长
fatB2	萼距花（*Cuphea hookeriana*）	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（*Cinnamomum camphora*）	硫酯酶	Q39473	3.1.2.14	C14:0 链长
fatB	拟南芥（*Arabidopsis thaliana*）	硫酯酶	CAA85388	3.1.2.14	C16:1 链长
fatB1	加州月桂	硫酯酶	Q41635	3.1.2.14	C12:0 链长
fatA1	向日葵（*Helianthus annuus*）	硫酯酶	AAL79361	3.1.2.14	C18:1 链长
fatA	拟南芥	硫酯酶	NP_189147, NP_193041	3.1.2.14	C18:1 链长
fatA	芥菜（*Brassica juncea*）	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	*Photbacterium profundum*	硫酯酶	YP_130990	3.1.2.14	链长
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

[0109]

基因名称	来源有机体	酶的名称	登录号#	EC 编号	示例性用途
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
饱和度水平的控制					
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
酯的生产					
AT3G51970	拟南芥	长链醇 O-脂肪-酰基转移酶	NP_190765	2.3.1.26	酯的生产
ELO1	安格斯毕赤酵母（*Pichia angusta*）	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	酯的生产
DAGAT/DGAT	拟南芥	二酰基甘油酰基转移酶	AAF19262	2.3.1.20	酯的生产
hWS	智人（*Homo sapiens*）	酰基-CoA 蜡醇酰基转移酶	AAX48018	2.3.1.20	酯的生产
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	酯的生产
ES9	除烃海杆菌（*Marinobacter hydrocarbonoclasticus*）	蜡酯合酶	ABO21021	2.3.1.20	酯的生产
mWS	霍霍巴（*Simmondsia chinensis*）	蜡酯合酶	AAD38041	2.3.1.-	酯的生产
脂肪醇的输出					
		硫酯酶(如上所述)			增加脂肪酸/脂肪醇的生产
BmFAR	家蚕	FAR(形成脂肪醇的酰	BAC79425	1.1.1.-	将酰基-CoA 转化为脂肪醇

[0110]

基因名称	来源有机体	酶的名称	登录号#	EC 编号	示例性用途
	(*Bombyxmori*)	基-CoA 还原酶)			
acr1	不动杆菌菌种 *ADP1*	酰基-CoA 还原酶	YP_047869	1.2.1.42	将脂肪酰基-CoA还原成脂肪醛
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 *ADP1*	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(*Geobacillusthermodenitrificans*)*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(*Synechococcus elongatus*)	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(*Mycobacterium smegmatis*)	羧酸还原酶蛋白质	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 合酶	NP_416319	6.2.1.3	将脂肪酸活化为脂肪酰基-CoA
atoB	欧文氏菌(*Erwiniacarotovora*)	乙酰基-CoA 乙酰基转移酶	YP_049388	2.3.1.9	丁醇的生产
hbd	溶纤维丁酸弧菌(*Butyrivibriofibrisolvens*)	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(*Clostridium perfringens*)	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(*Clostridium beijerinckii*)	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 *CFT073*	醛-醇脱氢酶	AAN80172	1.1.1.1	丁醇的生产

[0111]

基因名称	来源有机体	酶的名称	登录号#	EC 编号	示例性用途
				1.2.1.10	
脂肪醇乙酰基酯的输出					
		硫酯酶(如上所述)			改变输出
acr1	不动杆菌菌种 *ADP1*	酰基-CoA 还原酶	YP_047869	1.2.1.42	改变输出
yqhD	大肠杆菌 *K12*	醇脱氢酶	AP_003562	1.1.-.-	改变输出
AAT	草莓（*Fragaria x ananassa*）	醇 O-乙酰基转移酶	AAG13130	2.3.1.84	改变输出
终末石蜡的输出					
OleT	威海鲜球菌的菌种	脂肪酸脱羧酶	HQ709266	1.11.2.4	脱羧化的脂肪酸
产物的出口（export）					
AtMRP5	拟南芥	与拟南芥多药耐药相关的	NP_171908	无	改变产物出口的量
AmiS2	红球菌菌种	ABC 运送因子 AmiS2	JC5491	无	改变产物出口的量
AtPGP1	拟南芥	拟南芥 p 糖蛋白 1	NP_181228	无	改变产物出口的量
AcrA	*CandidatusProtochlamydiaamoebophila UWE25*	推断的多药外排运送蛋白质 acrA	CAF23274	无	改变产物出口的量
AcrB	*CandidatusProtochlamydiaamoebophila UWE25*	可能的多药外排运送蛋白质 acrB	CAF23275	无	改变产物出口的量
TolC	土拉弗朗西丝菌 novicida 亚种（*Francisellatularensis subsp. novicida*）	外膜蛋白质（细胞包膜的生物发生）	ABD59001	无	改变产物出口的量
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	改变产物出口的量
AcrF	大肠杆菌	吖啶黄抗性蛋白质 F	P24181	无	改变产物出口的量
tll1619	细长嗜热聚球蓝细菌（*Thermosynechococcus elongatus*）*[BP-1]*	多药外排运送蛋白质	NP_682409.1	无	改变产物出口的量
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	改变产物出口的量

[0112]

基因名称	来源有机体	酶的名称	登录号#	EC 编号	示例性用途
发酵					
复制的检查关口基因					增加输出效率
umuD	宋内志贺菌(*Shigellasonnei*)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(*Shigellaflexneri*)	NADH:NADPH 转氢酶(α 和 β 亚基)	P07001, P0AB70	1.6.1.2	增加输出效率
其他					
fabK	肺炎链球菌（*Streptococcus pneumoniae*）	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（*Bacillus licheniformis*)*DSM 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

[0113] 表2A：ω-羟化酶/ω-氧化酶(EC 1.14.15.3)的实例

[0114]

基因名称	来源有机体	登录号	氧化还原系统	羟基化的位置
cyp153A (aciA)	不动杆菌属的菌种 OC4	BAE78452	具有铁氧还蛋白和铁氧还蛋白还原酶的操纵子	ω-羟化酶
cyp153A16	海分枝杆菌（*Mycobacterium marinum*）M	YP_001851443	具有铁氧还蛋白和铁氧还蛋白还原酶的操纵子	ω-羟化酶
cyp153A6	分枝杆菌属的菌种 HXN-1500	AJ833989	具有铁氧还蛋白和铁氧还蛋白还原酶的操纵子	ω-羟化酶
cyp153A	水油海杆菌（Marinobacter aquaeolei） VT8	YP_957888	具有铁氧还蛋白和铁氧还蛋白还原酶的操纵子	ω-羟化酶
alkB	恶臭假单胞菌（*Pseudomonas putida*）GPo1	CAB54050	需要红素氧还蛋白和红素氧化蛋白还原酶	ω-羟化酶
alkB	荧光假单胞菌	CAB51045	需要红素氧还蛋白和红素氧化	ω-羟化酶

[0115]

基因名称	来源有机体	登录号	氧化还原系统	羟基化的位置
	(*Pseudomonas fluorescens*) CHA0		蛋白还原酶	
alkM	*Acinetobacter baylyi*	YP_046098	需要红素氧还蛋白和红素氧化蛋白还原酶	ω-羟化酶
alkB	大头茶属的物种SoGc	ADT82701	需要红素氧还蛋白和红素氧化蛋白还原酶	ω-羟化酶
alkW1	*Dietzia* sp. DQ12-45-1b	HQ850582	c-末端红素氧化蛋白融合物,需要红素氧还蛋白还原酶	ω-羟化酶
alkB	恶臭假单胞菌 GPo1	CAB54050	需要红素氧还蛋白和红素氧化蛋白还原酶	ω-羟化酶
alkB	荧光假单胞菌 CHA0	CAB51045	需要红素氧还蛋白和红素氧化蛋白还原酶	ω-羟化酶

[0116] 表2B:用于ω-羟化酶/ω-氧化酶(EC 1.14.15.3)氧化还原作用的伴侣的实例

[0117]

名称/命名	有机体	登录号#
铁氧还蛋白,铁氧还蛋白还原酶	不动杆菌属的菌种OC4	BAE78451,BAE78453
铁氧还蛋白,铁氧还蛋白还原酶	海分枝杆菌M	YP_001851444,YP_001851442
铁氧还蛋白,铁氧还蛋白还原酶	水油海杆菌VT8	YP_957887,YP_957889
alkG,alkT	恶臭假单胞菌GPo1	CAB54052,CAB54063
rubA,rubB	Acinetobacter baylyi ADP1	CAA86925,CAA86926

[0118] 表2C:自给自足的ω-1,ω-2,ω-3-羟化酶/氧化酶(EC 1.14.14.1)融合蛋白质的实例

[0119]

基因名称	来源有机体	酶的名称	氧化还原系统	羟基化的位置
P450-BM3 (cyp102A1)	巨大芽孢杆菌(*Bacillus megaterium*)	AAA87602	具有还原酶结构域的融合蛋白质	ω-1,-2,-3 羟化
yrhJ (cyp102A3)	枯草芽胞杆菌(*Bacillus subtilis*)	NP_390594	具有还原酶结构域的融合蛋白质	ω-1,-2,-3 羟化
yrhJ (cyp102A7)	地衣芽孢杆菌(*Bacillus licheniformis*)	AAU41718	具有还原酶结构域的融合蛋白质	ω-1,-2,-3 羟化

[0120] 表2D:自给自足的I类P450融合的PFOR融合蛋白质的实例

[0121]

名称/命名	有机体	登录号#
P450RhF	红球菌属的菌种 *NCIMB 9784*	AAM67416
REQ_44300	马红球菌(*Rhodococcus equi*) *103S*	YP_004009071
HMPREF0018_01193	抗辐射不动杆菌(*Acinetobacter radioresistens*) *SH164*	ZP_06072406
BMAA1669	鼻疽伯克霍尔德氏菌(*Burkholderia mallei*) *ATCC 23344*	YP_106239
Rmet_4932	耐金属贪铜菌(*Cupriavidus metallidurans*) *CH34*	YP_587063
H16_B1279	真氧产碱杆菌(*Ralstonia eutropha*) *H16*	YP_840799

[0122] 表3A:醇脱氢酶(EC 1.1.1.1/2)或醇氧化酶(EC 1.1.3.13,EC 1.1.3.20)的实例

[0123]

名称/命名	有机体	登录号#
alkJ	恶臭假单胞菌GPo1	CAB54054
alkJ	泊库岛食烷菌(Alcanivorax borkumensis)AP1	CAC38030
cddC	赤红球菌(Rhodococcus ruber)SC1	AAL14237

[0124] 表3B:醛脱氢酶(EC 1.2.1.3/4/5/)或醛氧化酶(EC 1.2.3.1)的实例

[0125]

名称/命名	有机体	登录号#
alkH	恶臭假单胞菌GPo1	CAB51050
alkH	泊库岛食烷菌AP1	CAC38029
cddD	赤红球菌SC1	AAL14238

[0126] 表4:氨基转移酶/转氨酶(EC 2.6.1)和胺脱氢酶(EC 1.4.9,EC 1.4.98,EC 1.4.99)的实例

[0127]

名称/命名	**功能**	**有机体**	**登录号#**
β-丙氨酸-丙酮酸转氨酶	β-丙氨酸-丙酮酸转氨酶	绿脓杆菌（*Pseudomonas aeruginosa*）PA7	YP_001345604
ygjG	丁二胺氨基转移酶	大肠杆菌 MG1655	NP_417544
gabT	5-氨基戊酸转氨酶	绿脓杆菌 PA01	AAG03655
Lat	L-赖氨酸 6-转氨酶	结核分枝杆菌（*Mycobacterium tuberculosis*）H37Rv	NP_217807
GABA-T	4-氨基丁酸转氨酶	野猪	NP_999428
Ald	丙氨酸脱氢酶	枯草芽孢杆菌纳豆亚种（*Bacillus subtilis subsp.*	BAI86717

[0128]

名称/命名	功能	有机体	登录号#
		natto）BEST195	
gdhA	谷氨酸脱氢酶(NADPH)	大肠杆菌 MG1655	NP_416275
Gdh	谷氨酸脱氢酶(NADH)	不解糖嗜胨菌（*Peptoniphilus asaccharolyticus*）	AAA25611
L-赖氨酸 6-脱氢酶	L-赖氨酸 6-脱氢酶	反硝化无色杆菌（*Achromobacter denitrificans*）	AAZ94428
mauRFBEDACJGMN	甲基胺脱氢酶	脱氮副球菌（*Paracoccus denitrificans*）	P52685.1 P29897.2 P29894.1 P29896.2 P29895.2 P22619.2 P22364.1 P22566.2 ABL72797.1 ABL72798.1 AAA86469.1

[0129] 表5:酯酶(EC 3.1.1.1)和脂肪酶(EC 3.1.1.3)的实例

[0130]

名称/命名	有机体	登录号#
lipA	荧光假单胞菌B52	AAF80996
phaZ	荧光假单胞菌GK13	AAA64538

[0131] 表6:水解酶(EC 3.5.2.12)的实例

[0132]

名称/命名	有机体	登录号#
ω-十二内酰胺水解酶	食酸菌属的菌种T31	BAH09870
ω-十二内酰胺水解酶	贪铜菌属的菌种U124	BAH09871

[0133] 表7:酰基-CoA合酶/酰基-CoA连接酶(EC 6.2.1.3)/转移酶(EC 2.8.3.6) 的实例

[0134]

名称/命名	有机体	登录号#
fadD	大肠杆菌 MG1655	NP_416319
fadK	大肠杆菌 MG1655	NP_416216
pimA	沼泽红假单胞菌（*Rhodopseudomonas palustris*）	NP_949053

[0135]

dcaI/dcaJ	Acinetobacter baylyi ADP1	YP_046360/ YP_046361

[0136] 表8:酰胺合酶的实例

[0137]

名称/命名	有机体	登录号#
棕榈酰腐胺合酶(PPS)	非培养的细菌	AAV33349
N-(4-氨基-2-羟基丁基)肉都可酰胺合酶(AhtS)	非培养的细菌	ACX33975

[0138] 表9:酯合酶(EC 2.3.1.75或EC 2.3.1.20)的实例

[0139]

名称/命名	有机体	登录号#
AtfA	不动杆菌属的菌种ADP1	Q8GGG1
AtfA1	泊库岛食烷菌SK2	YP_694462
AtfA2	泊库岛食烷菌SK2	YP_693524
WS/DGAT	Marinobacter alginolyticus	WP_007153340
WS/DGAT	苯酚降解菌属的菌种MED105	WP_008251579
ES9	除烃海杆菌	ABO21021

[0140] 图1和2示出了其中酰基-ACP可以通过2个相似的途径而转化成α,ω- 二酸的途径,其中所述的2个途径分别使用C12游离脂肪酸(FFA)作为前 体中间体(参见图1),或者C12脂肪酸甲基酯(FAME)作为中间体(参见 图2A和2B)。

[0141] 在一个实施方案中,图1示出了多种化学化合物的生产,包括ω-羟基 脂肪酸,ω-氧代脂肪酸,α,ω-二酸和ω-氨基脂肪酸。在图1的步骤2中,使 用硫酯酶将酰基-ACP转化成FFA。在某些实施方案中,编码硫酯酶的基因为tesA,‘tesA、tesB、fatB1、fatB2、fatB3、fatA1或fatA。(同样参见表1, 其示出了具有硫酯酶的酶活性的多肽,其中所述的硫酯酶可以用于催化本 步骤,如上所述)。在步骤3中,还称为ω-氧化酶的ω-羟化酶用于生成ω-羟基脂肪酸。如图1中可见,脂肪酸的omega位置为羟基。

[0142] 在重组宿主细胞中表达的CYP153A-还原酶杂交融合多肽

[0143] 合适的ω-羟化酶/ω-氧化酶(EC 1.14.15.3)和它们的氧化还原伴侣的实 例列于表2A和2B中(如上所述)。这些是某些非血红素亚铁氧化酶(例 如得自恶臭假单胞菌GPo1的alkB),或者某些血红素型P450氧化酶(例 如得自水油海杆菌(Marinobacter aquaeolei)的cyp153A),也称为细胞色 素P450。细胞色素P450为广泛分布的酶,其具有高度的复杂性,并且展 示出广泛领域的活性。它们是由转化多种底物并催化多种化学反应的基因 超家族所编码的蛋白质。Cyp153A为可溶性细菌细胞色素P450的超家族, 其以对ω-位置具有高度的选择性二羟基化碳氢化合物链(van Beilen et al.(2006)Appl.Environ.Microbiol.72:59-65)。cyp153A家族的成员在体外显示 出选择性地羟基化烷、脂肪酸或脂肪醇的ω-位置,所述的成员例如为得自 分枝杆菌属菌种HXN-1500的cyp153A6(Funhoff et al.(2006)J.Bacteriol.188:5220-5227),得自海分枝杆菌的cyp153A16和得自极地单胞菌属菌种 JS666的cyp153A(Scheps et al.(2011)Org.Biomol.Chem.9:6727-6733), 以及得自水油海杆菌的cyp153A(Honda-Malca et al.(2012)Chem.Commun.48:5115-5117)。

[0144] 如同所有的细胞色素P450,Cyp153Aω-羟化酶需要电子用于它们的催 化活性,这些电子是通过特定的氧化还原蛋白质来提供的,例如铁氧还蛋 白和铁氧还蛋白还原酶。这些是与cyp153A相互作用的离散的蛋白质。之 前,已经通过将得自泊库岛食烷菌SK2的cyp153A(Kubota et al.(2005)Biosci.Biotechnol.Biochem.69:2421-2430;Fujita et al.(2009)Biosci. Biotechnol.Biochem.73:1825-1830)与得自P450RhF的还原酶结构域(其 包含黄素单核苷酸(FMN)和NADPH结合位点,以及[2FeS]铁氧还蛋白中 心(Hunter et al.(2005)FEBS Lett.579:2215-2220))融合而创建了自给自足 的杂交(嵌合)cyp153A氧化酶(即,其活性不需要离散的铁氧还蛋白和 铁氧还蛋白还原酶蛋白质的氧化酶)。P450RhF属于I类P450融合的PFOR (DeMot and Parret(2003)Trends Microbiol.10:502)。

这种杂交 cyp153A-RedRhF融合蛋白质在体外生物转化中显示在ω-位置羟基化辛烷,并且还羟基化其他的化合物,例如环己烷或丁基苯。天然P450-还原 酶融合蛋白质的实例示于表2C和2D(如上所述)。

[0145] 考虑到cyp153A家族的氧化酶对碳氢化合物链的ω-位置具有高度的 选择性,其显示为可用于由可再生的碳源来生产α,ω-双功能脂肪酸衍生物 的合适候选物的良好实例。这允许研发商业可行的方法来生产这些有价值 的化合物。但是,如同其他的细胞色素P450,cyp153A家族的蛋白质到目 前为止主要用于体外试验中,这些试验使用了纯化的酶或细胞裂解物粗 品,或者处于休眠细胞的生物转化(其中以外源方式加入了脂肪酸衍生物或碳氢化合物)中(Kubota et al.,Fujita et al.,Honda-Malca et al.,如上所述)。但是,使用杂交融合的体外过程或休眠细胞生物转化不能大规模实 施并成本有效地生产ω-羟基脂肪酸衍生物。本领域中广泛接受的知识是许 多细胞色素P450以及alkB型ω-羟化酶在重组微生物有机体中不容易进行 功能性地表达,这是因为这些酶通常是无活性的,并且它们的化学作用难以阐明。事实上,目前尝试使用不同于脂肪酸衍生物的可再生碳源的唯一 的体内工作使用了alkBω-羟化酶,并且在高细胞密度的发酵中仅取得了低 效价的ω-羟基脂肪酸衍生物(WO2013/024114A2)。

[0146] 本申请人创建了CYP153A-还原酶杂交融合蛋白质及其变体,它们在体内能够有效地由可再生的碳源来生产ω-羟基脂肪酸衍生物。更具体而言,编码CYP153A(G307A)P450催化结构域的杂交融合蛋白质(其中在位置307处甘氨酸(G)被取代为丙氨酸(A))的来自水油海杆菌的基因与来自红球菌属的菌种NCIMB9784的编码P450RhF的含c-末端FMN-和Fe/S还原酶结构域的基因融合(参见实施例6,参见下文)。所得的多肽为CYP153A-RedRhF杂交融合多肽(SEQ ID NO:6),具有相应的核酸序列(SEQ ID NO:5)。当这种CYP153A-还原酶杂交融合蛋白质在大肠杆菌细胞(使用简单的碳源,例如葡萄糖)中表达时,脂肪酸衍生物可以有效地转化成ω-羟基脂肪酸衍生物(参见实施例,参见下文)。合适的ω-羟化酶(EC 1.14.15.3)和它们的氧化还原伴侣(可以用于生成相似的CYP153A-还原酶杂交融合多肽)的其他实例列于表2A和2B中(如上所述)。

[0147] 本发明公开提供了在体内可以有效地且选择性地生产ω-羟基脂肪酸 衍生物(包括α,ω-双功能脂肪酸衍生物)的微生物有机体。在一个实施方 案中,CYP153A-RedRhF杂交融合蛋白质经改造而在微生物有机体中表达, 使得所述微生物有机体在体内能够由碳源(例如葡萄糖)有效地将诸如桂 酸或十二烷酸甲基酯之类的化合物转化成12-羟基十二烷酸或12-羟基十二 烷酸甲基酯。任何可再生的供料都可以代替葡萄糖而用作碳源。因此,第一次显示当具有P450酶活性的改造的杂交融合蛋白质(即,通过 CYP153A-RedRhF杂交融合蛋白质说明)与硫酯酶在宿主细胞(例如大肠 杆菌)中共表达并由可再生的供料供入碳源时,所述的杂交融合蛋白质在体内可以有效地将脂肪酸转化成特定的所需的ω-羟基化的化合物(例如ω- 羟基脂肪酸;ω-羟基脂肪酸甲基酯;α,ω-羟基二酸;α,ω-羟基二酯;和α,ω- 二醇)(参见实施例6以及图25和26)。根据本发明公开,可以通过将P450 基因(例如编码cyp153A蛋白质的基因)与还原酶基因(例如编码I类P450 融合的PFOR蛋白质的c-末端还原酶结构域或其他的结构域)连接来改造 其他的杂交融合蛋白质。表2A和2D(如上所述)分别给出了cyp153A和 I类P450融合的PFOR蛋白质的实例。根据这些指导,可以由其他类型的 ω-羟化酶/ω-氧化酶来创建类似的融合蛋白质。

[0148] 如图2A中所示、通过ω-羟基脂肪酸甲基酯形成α,ω-二酸或ω-氨基脂 肪酸的途径,或者如图2B中所示、通过α,ω-脂肪酸二甲基酯形成α,ω-二 酸的途径是有利的,这是因为甲基酯是不带电荷的,这提供了大规模生产 和回收的优点。此外,通过ω-羟基脂肪酸甲基酯的途径允许通过某些水解 酶的催化而进行直接的酶的内酰胺化。在图1或图2A和2B的步骤3中, 醇脱氢酶或氧化酶可以进一步将ω-羟基脂肪酸转化成ω-氧代脂肪酸(例如表3A显示具有醇脱氢酶或氧化酶的酶活性的多肽,其中所述的醇脱氢酶 或氧化酶可以用于催化该步骤,如上所述)。例如可以将12-羟基十二烷酸 或12-羟基十二烷酸甲基酯氧化成12-氧代十二烷酸或12-氧代十二烷酸甲 基酯的合适的酶为的醇氧化酶(黄素蛋白,例如得自恶臭假单胞菌的alkJ)(EC 1.1.3.13,EC 1.1.3.20)(参见SEQ ID NO:67)或NAD(P)依赖的醇脱 氢酶(例如得自赤红球菌的cddC)(EC 1.1.1.1)(参见表3A,如上所述)。 就此而言,所述的途径由此可以根据2种不同的备选途径,即,通过醛脱 氢酶或氧化酶将ω-氧代脂肪酸转化成α,ω-二酸(例如表3B显示具有醛脱 氢酶或氧化酶的酶活性的多肽,其中所述的醛脱氢酶或氧化酶可以用于催 化该步骤,如上所述),或者通过氨基转移酶、转氨酶或氨基酸脱氢酶将ω- 氧代脂肪酸转化成ω-氨基脂肪酸(例如表4显示具有氨基转移酶、转氨酶或氨基酸脱氢酶的酶活性的多肽,其中所述的氨基转移酶、转氨酶或氨基 酸脱氢酶可以用于催化该步骤,如上所述)。例如用于将12-氧代十二烷酸 或12-氧代十二烷酸甲基酯转化成12-氨基十二烷酸或12-氨基甲基酯的合 适的转氨酶示于表4中(如上所述)。这些酶将胺基由合适的供体转移至ω- 含氧酸或酯,从而生成相应的终末胺(terminal amine)。可以通过多种氨 基酸脱氢酶(例如丙氨酸脱氢酶、谷氨酸盐脱氢酶,参见表4,如上所述)的共表达而在体内改良胺供体(例如丙氨酸或谷氨酸盐)的可利用性。此 外,谷氨酸盐脱氢酶(可以利用NADH代替NADPH)的同源物或其他变 体的共表达可以提供用于生产的额外的改良。此外,适用于将12-氧代十 二烷酸或12-氧代十二烷酸甲基酯转化成12-氨基十二烷酸或12-氨基甲基 酯的其他种类的酶(例如甲基胺脱氢酶或赖氨酸6-脱氢酶)的实例也列于表4中。

[0149] 在另一个实施方案中,图2A和2B示出了通过甲基酯作为中间体而生 产多种化学化合物,包括α,ω-二酸和ω-氨基脂肪酸。图2B的步骤6示出 了α,ω-二酯的生产(使用酰基-CoA连接酶/转移酶)。在图2A和2B的步 骤2中,使用酯合酶来将酰基-ACP转化成脂肪酸甲基酯(FAME)。在某些 实施方案中,编码酯合酶的基因为编码酶分类为EC 2.3.1.75或EC 2.3.1.20 的酶的基因,编码wax/dgat(得自霍霍巴、不动杆菌属的菌种ADP1、泊 库岛食烷菌、绿脓杆菌、Fundibacter jadensis、拟南芥或Alkaligenes eutrophus 的双功能酯合酶/酰基-CoA:二酰基甘油酰基转移酶)的基因,或编码AtfA1、 AtfA2、ES9、ES8或它们的变体的基因(此外,参见表9,其示出了具有 酯合酶的酶活性的多肽,其中所述的酯合酶可以用于催化该步骤,如上所 述)。在步骤3中,使用ω-羟化酶(或ω-氧化酶)来生成ω-羟基脂肪酸甲基酯(例如表2A示出了具有酶活性的多肽,该酶活性可以用于催化该步骤, 如上所述)。如图2A中可见,脂肪酸甲基酯的omega位置为羟基。在图 2A的步骤3中,醇脱氢酶或氧化酶可以用于将ω-羟基脂肪酸甲基酯转化 成ω-氧代脂肪酸甲基酯(例如表3A示出了具有醇脱氢酶或氧化酶的酶活 性的多肽,其中所述的醇脱氢酶或氧化酶可以用于催化该步骤,如上所述)。类似地,所述的途径由此可以根据2种不同的备选途径,即,一种 途径通过醛脱氢酶或氧化酶(参见表3B,如上所述)以及最终的酯酶或脂 肪酶(参见表5,如上所述)、

或者通过化学转化而将ω-氧代脂肪酸甲基 酯转化成α,ω-二酸;而另一种途径通过氨基转移酶或转氨酶或氨基酸脱氢 酶(参见表4,如上所述)、以及最终的酯酶或脂肪酶(参见表5,如上所 述)而将ω-氧代脂肪酸甲基酯转化成ω-氨基脂肪酸。最终的步骤(即,其 中酯酶或脂肪酶催化将α,ω-脂肪酸甲基酯或ω-氨基脂肪酸甲基酯转化成 所示的最终产物)可以为酶转化步骤以及化学转化步骤的结果。例如可以催化该步骤的合适的酶列于表5中。所选的脂肪酶(EC 3.1.1.3)或酯酶(EC 3.1.1.1)位于周质,或者分泌至上清液中。可以根据本领域已知的方法来 改造合适的分泌信号或锚定序列。在另一个实施方案中,图2B示出了通 过α,ω-脂肪酸二甲基酯来生产α,ω-二酸。该途径类似于图2A中的左侧, 但是其使用了其他的酰基-CoA合酶/酰基-CoA连接酶或转移酶(对于合适 的候选物,参见表7,如上所述)。此外,催化第一步骤的酯合酶/酰基转 移酶还可以催化倒数第二步骤。

[0150] 在另一个实施方案中,图3和4描绘了使用由可再生的供料得到的碳 源(例如葡萄糖)由细胞中生产的脂肪醇(FALC)来生产α,ω-二醇的途 径。在一个实施方案中,图3示出了其中通过使用硫酯酶或羧酸还原酶活 性来制备α,ω-二醇的途径(对于可以催化这些步骤的多肽,参见表1,如 上所述)。在本发明中,在步骤2中,硫酯酶可以催化酰基-ACP转化成游 离脂肪酸(FFA)。在步骤3中,羧酸还原酶可以将所得的FFA转化成脂 肪醛。在步骤4中,醇脱氢酶(参见表1,如上所述)可以将脂肪醛转化 成脂肪醇(FALC)。最后,在步骤5中,ω-羟化酶或ω-氧化酶(参见表2A, 如上所述)可以将FALC转化成α,ω-二醇。备选地,酰基-ACP可以通过 酰基-ACP还原酶而直接转化成脂肪醛,如图4所示。(表1示出了具有硫酯酶、羧酸还原酶、酰基-ACP还原酶或醇脱氢酶的酶活性的多肽,从而 催化这些步骤,如上所述)。

[0151] 在另一个实施方案中,示出了生产内酰胺的途径,所述的内酰胺可以 以化学方式转化成聚合物。这以通过ω-氨基脂肪酸(例如12-氨基十二烷 酸)的实例而示于图5中,其中所述的ω-氨基脂肪酸通过酰基-CoA-转移 酶(参见表7,如上所述)和酰胺合酶(参见表8,如上所述)而转化成 内酰胺。在这种具体的实例中,所得的内酰胺为12-氨基十二烷内酰胺或 十二内酰胺。由其衍生的聚合物的实例为尼龙12,如图5所示。用于将 12-氨基十二烷酸甲基酯直接内酰胺化成相应的内酰胺的合适的酶为某些水解酶(参见表6,如上所述),其在相同的生物催化剂中可以与其他酶组 合使用,从而指导十二内酰胺的生产;或者可以用于12-氨基十二烷酸甲 基酯的分开的生物转化中。脂肪酶或酯酶(参见表5,如上所述)还可以将12-氨基十二烷酸甲基酯直接内酰胺化成相应的内酰胺。

[0152] 本发明公开识别了编码具有酶活性的多肽的多核苷酸,其中所述多肽 可以用于重组宿主细胞中和生产方法中。具有酶活性的多肽有助于生产包 含所述的化合物的组合物。普遍公认的是无需与此类多核苷酸的绝对的序 列一致性。例如可以使特定的多核苷酸序列(例如编码具有酶功能的多肽的多核苷酸)发生变化,并针对活性来筛选所编码的多肽。此类变化通常 包含保守的突变和沉默突变(例如密码子优化)。可以使用本领域已知的方法,针对所需的功能来筛选基因改造的或修饰的多核苷酸及所编码的变 体多肽,其中所述的功能包括但不限于增加的增加的催化活性、增加到稳 定性或降低的抑制(例如反馈降低的抑制)。

[0153] 此外,本发明公开根据酶分类(EC)编号,鉴定了与经改造的途径(涉 及本文所述的ω-羟基化的脂肪酸衍生物的生产,如上所述)的多个步骤 (即,反应)有关的酶活性,并

提供了通过此类EC编号而分类的示例性 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(本发明中,其是通过登录号和/或序列标识号(SEQ ID NO)来识 别的)用于改造亲代宿主细胞中的脂肪酸途径,该途径可以生产ω-羟基脂 肪酸衍生物,包括其他双功能分子,例如α,ω-二酸,从而得到本发明所述 的基因修饰的宿主细胞。本发明所述多肽和多核苷酸是示例性的且非限定 性的。本领域的那些技术人员可以通过多个数据库(例如the National Center for Biotechnology Information(NCBI)提供的Entrez数据库,the Swiss Institute of Bioinformatics提供的ExPasy数据库,the Technical University of Braunschweig提供的BRENDA数据库,以及the Bioinformatics Center of Kyoto University and University of Tokyo提供的KEGG数据库,所有这些均可得自互联网)得到本发明所述的示例性多肽的同源物的序列。

[0154] 在一个实施方案中,本发明所述重组微生物有机体(如上所述)会生 产具有偶数碳链的ω-羟基脂肪酸衍生物。在另一个实施方案中,所述重组 微生物有机体可以被改造,从而生产具有奇数碳链的ω-羟基脂肪酸衍生 物。例如ω-羟化酶途径可以在过度生产奇数链的脂肪酸衍生物的重组细胞 (例如大肠杆菌)中表达(通过某些细菌菌株而过度生产的奇数链脂肪酸衍生物在美国专利申请公开US2012/0070868中有所描述,该文献以引用方式并入本文)。这允许生产奇数链的ω-羟基化的脂肪酸衍生物。在一个 实施方案中,过度生产奇数链脂肪酸的重组菌株当与表达ω-羟化酶(例如 CYP153A-RedRhF杂交融合蛋白质)和脂肪酸脱羧酶(例如oleT)(例如 参见Rude et al.(2011) Appl.Environ.Microbiol.77:1718)相结合时,生产具有末端双键的偶数链的脂肪醇,例如9-癸烯-1-醇、11-十二烯-1-醇、13- 十四烯-1-醇和15-十六烯-1-醇。

[0155] 在另一个实施方案中,提供了近末端羟基化的脂肪酸衍生物。这些化 合物为在omega-1位置,和/或omega-2位置,和/或omega-3位置,和/或 omega-4位置等(例如ω-1,ω-2和/或ω-3等)具有至少一个OH基团的化 合物。实施例8和9示出了通过使用得自cyp102A家族的细胞色素P450 氧化酶而在基因修饰的宿主细胞中生产这些化合物(Whitehouse et al. (2012)Chem.Soc.Rev.41:1218)。cyp102A家族的细胞色素P450氧化酶适 用于生产近末端羟基化的脂肪酸衍生物。但是,这些酶不适用于在重组宿 主中有效地生产ω-羟基化的脂肪酸衍生物,除非它们的底物特异性被改 变,使得它们主要羟基化ω-位置的脂肪酸衍生物,到目前为止,这仅是部分成功的(Lentz et al.(2006) ChemBioChem.7:345)。

[0156] 在重组宿主细胞中表达的CYP153A-还原酶杂交融合多肽变体

[0157] 本发明公开鉴定了CYP153A-还原酶杂交融合多肽变体,当其在重组 宿主细胞中表达时,得到高的效价、产率和/或生产率的ω-羟基化的脂肪酸衍生物组合物。在本发明公开的非限定性实例中, CYP153A(G307A)-RedRhF杂交融合多肽(参见实施例6,参见下文)被用 作有效地改造CYP153A-还原酶杂交融合多肽变体的模板(参见实施例 14-10,参见下文),从而生产增加量的ω-OH脂肪酸和ω-OH脂肪酸衍生 物。例如当CYP153A-还原酶杂交融合多肽变体在宿主细胞中表达时,其 在体内可以有效地由简单的碳源(例如葡萄糖)将诸如十二烷酸之类的化 合物转化成12-羟基十二烷酸。任何简单的碳源(例如由可再生的供料衍 生得到)都是合适的。已经显示,当改造的CYP153A-还原酶杂交融合多 肽变体(即,通过改造的CYP153A-RedRhF杂交融合多肽变体来说明)与 硫酯酶在宿主细胞(例如大肠杆

菌)中共表达时,其在体内可以通过利用由可再生的供料得到的碳源(例如葡萄糖)而有效地将脂肪酸转化成特定 的所需的化合物(包括ω-OH脂肪酸)(参见实施例,参见下文)。根据本 发明公开,可以通过将突变的基因(例如编码CYP153A蛋白质的基因) 与编码c-末端还原酶结构域的突变的基因相连接来改造其他的杂交融合多 肽变体(参见表2A至2D,如上所述)。本发明涵盖了多种改变,例如2 个基因的突变(P5450和还原酶结构域)或1个基因的突变(P450或还原酶结构域)。根据这些指导,可以由其他类型的ω-羟化酶来创建类似的融合蛋白质变体,并使它们在重组宿主细胞中表达,从而生产ω-OH脂肪酸 衍生物。

[0158] 因此,本发明公开涉及表达CYP153A-还原酶杂交融合多肽变体的重 组宿主细胞,当使用碳源(例如葡萄糖、蔗糖或者由可再生的供料衍生的 任何其他的碳源)培养所述的细胞时,其产生高的效价、产率和/或生产率 的ω-羟基化的脂肪酸衍生物组合物。CYP153A-还原酶杂交融合多肽变体 在CYP153A结构域或还原结构域或两者中具有一个或多个突变。在一个 实施方案中,本发明公开提供了CYP153A-还原酶杂交融合多肽变体,其与SEQ ID NO:6具有至少大约90%、91%、92%、93%、94%、95%、96%、 97%、98%或99%的序列一致性,并且在氨基酸位置处具有一个或多个突 变,其中所述的氨基酸位置包括位置27、82、141、178、231、309、407、 415、516、666和/或796,其中所述的CYP153A-还原酶杂交融合多肽变体在体内催化脂肪酸转化成ω-OH脂肪酸。更具体而言,CYP153A-还原酶 杂交融合多肽变体具有以下突变的一个或多个,包括:R27L,其中精氨酸 (R)取代赖氨酸(L);位置R82D,其中精氨酸(R)取代天冬氨酸(D);位置 V141I,其中缬氨酸取代异亮氨酸(I);位置V141Q,其中缬氨酸(V)取代谷氨酰胺(Q);位置V141G,其中缬氨酸(V)取代甘氨酸(G);位置V141M, 其中缬氨酸(V)取代蛋氨酸(M);位置V141L,其中缬氨酸(V)取代亮氨酸(L);位置V141T,其中缬氨酸(V)取代苏氨酸(T);位置R178N,其中精氨 酸(R)取代天冬酰氨(N);位置A231T,其中丙氨酸(A)取代苏氨酸(T);位置 N309R,其中天冬酰氨(N)取代精氨酸(R);位置N407A,其中天冬酰氨(N) 取代丙氨酸(A);位置V415R,其中缬氨酸(V)取代精氨酸(R);位置T516V,其中苏氨酸(T)取代缬氨酸(V);位置P666A,其中脯氨酸(P)取代丙氨酸(A); 位置P666D,其中脯氨酸(P)取代天冬氨酸(D);以及位置A796V,其中丙 氨酸(A)取代缬氨酸(V)。CYP153A-还原酶杂交融合多肽变体的实例包括 SEQ ID NO:8、SEQ ID NO:10、SEQ ID NO:12、SEQ ID NO:14、SEQ ID NO:16、SEQ ID NO:18、SEQ ID NO:20、SEQ ID NO:22、SEQ ID NO:24、SEQ ID NO:26、SEQ ID NO:28或SEQ ID NO:30、SEQ ID NO:32、SEQ ID NO:34、SEQ ID NO:36、SEQ ID NO:38、SEQ ID NO:40、SEQ ID NO: 42、SEQ ID NO:44和SEQ ID NO:46。在一个实施方案中,CYP153A-还 原酶杂交融合多肽变体为杂交cyp153A-RedRhF-型融合蛋白质变体。在另 一个实施方案中,重组宿主细胞中的CYP153A-还原酶杂交融合多肽变体得到比通过在相应的宿主细胞中表达CYP153A-还原酶杂交融合多肽而生产的ω-OH脂肪酸组合物的效价更高效价的ω-OH脂肪酸衍生物组合物。在另一个实施方案中,重组宿主细胞中的CYP153A-还原酶杂交融合多肽 变体得到更高效价的ω-OH脂肪酸衍生物组合物,包括但不限于ω-OH C_{12}、 ω-OH C_{14}、ω-OH C_{16}、ω-OH C_{18}、ω-OH $C_{12:1}$、ω-OH $C_{14:1}$、ω-OH $C_{16:1}$和ω-OH $C_{18:1}$脂肪酸衍生物组合物。

[0159] 在另一个实施方案中,CYP153A-还原酶杂交融合多肽变体在氨基酸 位置141处具有突变,包括V141I和/或V141T。在本发明中,具有突变V141I或V141T的CYP153A-还原酶杂交融合多肽变体在重组宿主细胞中 的表达得到分别比通过表达CYP153A-还原酶杂交融合

多肽而生产的 ω-OH C_{12}或C_{16}脂肪酸组合物的效价更高效价的ω-OH C_{12}或C_{16}脂肪酸组 合物。在一个实施方案中,CYP153A-还原酶杂交融合多肽变体具有突变 V141I和A231T(SEQ ID NO:32),并且当其在具有硫酯酶的酶功能的宿主 细胞中表达时,生产增加量的ω-OH C_{12}脂肪酸。在另一个实施方案中, CYP153A-还原酶杂交融合多肽变体具有突变R27L、R82D、V141M、R178N 和N407A(SEQ ID NO:34),并且当其在具有硫酯酶的酶功能的宿主细胞中 表达时,生产增加量的ω-OH C_{12}脂肪酸。在另一个实施方案中,CYP153A- 还原酶杂交融合多肽变体具有突变P666A(SEQ ID NO:36),并且当其在具有硫酯酶的酶功能的宿主细胞中表达时,生产增加量的ω-OH C_{12}脂肪酸。 在另一个实施方案中,CYP153A-还原酶杂交融合多肽变体具有突变 A796V(SEQ ID NO:38),并且当其在具有硫酯酶的酶功能的宿主细胞中表达时,生产增加量的ω-OH C_{12}脂肪酸。在另一个实施方案中,CYP153A- 还原酶杂交融合多肽变体具有突变A796V、P666D和T516V(SEQ ID NO: 40),并且当其在具有硫酯酶的酶功能的宿主细胞中表达时,生产增加量的ω-OH C_{12}脂肪酸。在另一个实施方案中,CYP153A-还原酶杂交融合多肽 变体具有突变V141I、A231T和A796V(SEQ ID NO:42),并且当其在具有 硫酯酶的酶功能的宿主细胞中表达时,生产增加量的ω-OH C_{12}和C_{16}脂肪 酸。在另一个实施方案中,CYP153A-还原酶杂交融合多肽变体具有突变R27L、R82D、V141M、R178N、N407A和A796V(SEQ ID NO:44),并且 当其在具有硫酯酶的酶功能的宿主细胞中表达时,生产增加量的ω-OH C_{12}脂肪酸。在另一个实施方案中,CYP153A-还原酶杂交融合多肽变体具有 突变V141T、A231T和A796V(SEQ ID NO:46),并且当其在具有硫酯酶的 酶功能的宿主细胞中表达时,生产增加量的ω-OH C_{16}脂肪酸。

[0160] 本发明公开鉴定了CYP153A-还原酶杂交融合相关的多核苷酸和多肽 变体。CYP153A-还原酶杂交融合多肽变体包括SEQ ID NO:8、10、12、 14、16、18、20、22、24、26、28、30、32、34、36、38、40、42、44和 46。CYP153A-还原酶杂交融合核酸变体(DNA序列)包括SEQ ID NO:7、 9、11、13、15、17、19、21、23、25、27、29、31、33、35、37、39、41、 43、45和47。但是,公认的是无需与CYP153A-还原酶杂交融合多核苷酸变体的绝对的序列一致性。例如可以使特定的多核苷酸序列发生变化,并 针对活性来筛选编码的多肽。此类变化通常包括保守突变和沉默突变(例 如通过密码子优化)。可以使用本领域已知的方法,针对所需的功能来筛选改造的或修饰的(即,突变的)多核苷酸及所编码的变体多肽,其中所 述的功能例如为比野生型或模板多肽改良的功能,包括但不限于增加的催 化活性、增加的稳定性或降低的抑制(例如降低的反馈抑制)。本发明公开根据酶分类(EC)编号,鉴定了与本文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,并提供了通过此类EC编号而分类的示例性多肽(例如其发挥特定的酶的作用并展示特定的酶的活性)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(本发明中,其 是通过序列标识号(SEQ ID NO;如上所述)来识别的)用于改造宿主细胞 中的脂肪酸途径,例如图1中所示的途径。但是,应该理解的是本发明所 述多肽和多核苷酸是示例性的,因此为非限定性的。本领域的那些技术人 员可以使用多个数据库(例如the National Center for Biotechnology Information(NCBI)提供的Entrez数据库,the Swiss Institute of Bioinformatics提供的ExPasy数据库,the Technical University of Braunschweig提供的BRENDA数据库,以及the Bioinformatics Center of Kyoto University and University of Tokyo提供的KEGG数据库,所有这些 均可得自互联网)得到本发明所述的示例性多肽的同源物的序列。

[0161] 在一个实施方案中,用于实施本发明公开的CYP153A-还原酶杂交融 合多肽变体与SEQ ID NO:8、SEQ ID NO:10、SEQ ID NO:12、SEQ ID NO: 14、SEQ ID NO:16、SEQ ID NO: 18、SEQ ID NO:20、SEQ ID NO:22、SEQ ID NO:24、SEQ ID NO:26、SEQ ID NO:28或SEQ ID NO:30、SEQ ID NO:32、SEQ ID NO:34、SEQ ID NO:36、SEQ ID NO:38、SEQ ID NO: 40、SEQ ID NO:42、SEQ ID NO:44和SEQ ID NO:46具有至少大约90％、 91％、92％、93％、94％、95％、96％、97％、98％、99％或100％的序列一 致性。在一些实施方案中,CYP153A-还原酶杂交融合多肽变体衍生自水 油海杆菌的CYP153A(G307A)多肽,其中甘氨酸(G)被取代为丙氨酸(A), 并且与得自红球菌属的菌种NCIMB9784的P450RhF还原酶结构域融合。细胞色素P450RhF是自给自足的,展现出高度的底物混杂性,并且催化广 泛的官能团。在其他的实施方案中,用于实施本发明公开的CYP153A-还 原酶杂交融合多肽变体与SEQ ID NO:6、SEQ ID NO:8、SEQ ID NO:10、 SEQ ID NO:12、SEQ ID NO:14、SEQ ID NO:16、SEQ ID NO:18、SEQ ID NO:20、SEQ ID NO:22、SEQ ID NO:24、SEQ ID NO:26、SEQ ID NO:28、 SEQ ID NO:30、SEQ ID NO:32、SEQ ID NO:34、SEQ ID NO:36、SEQ ID NO:38、SEQ ID NO:40、SEQ ID NO:42、SEQ ID NO:44或SEQ ID NO:46 具有至少大约75％、76％、77％、78％、79％、80％、81％、82％、83％、 84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、 96％、97％、98％或者至少99％的序列一致性,并且还可以包含一个或多个取代,这可以得到本发明所述的有用的特征和/或性质。在其他实施方案 中,用于实施本发明公开的CYP153A-还原酶杂交融合多肽变体与SEQ ID NO:8、SEQ ID NO:10、SEQ ID NO:12、SEQ ID NO:14、SEQ ID NO:16、 SEQ ID NO:18、SEQ ID NO:20、SEQ ID NO:22、SEQ ID NO:24、SEQ ID NO:26、SEQ ID NO:28、SEQ ID NO:30、SEQ ID NO:32、SEQ ID NO:34、 SEQ ID NO:36、SEQ ID NO:38、SEQ ID NO:40、SEQ ID NO:42、SEQ ID NO:44或SEQ ID NO:46具有100％、99％、98％、97％、96％、95％、94％、 93％、92％、91％或90％的序列一致性。在其他的实施方案中,CYP153A- 还原酶杂交融合多肽变体的P450催化结构域衍生自除了水油海杆菌以外的物种。此类其他的物种包括但不限于不动杆菌属的菌种、海分枝杆菌、Polaromonas sp.、泊库岛食烷菌、真菌伯克霍尔德菌(Burkholderia fungorum)、新月柄杆菌(Caulobacter crescentus)、Hyphomonas neptunium、 沼泽红假单胞菌(Rhodopseudomonas palustris)、鞘氨醇单胞菌属的菌种、 分支杆菌属的菌种。在其他的实施方案中,CYP153A-还原酶杂交融合多 肽变体的还原酶结构域衍生自除了红球菌属的菌种以外的物种。此类其他的物种包括但不限于马红球菌、抗辐射不动杆菌、鼻疽伯克霍尔德氏菌、 鼻疽伯克霍尔德氏菌、真氧产碱杆菌、耐金属贪铜菌。

[0162] 在相关的实施方案中,本发明公开包含CYP153A-还原酶杂交融合多 核苷酸变体,其与SEQ ID NO:5、SEQ ID NO:7、SEQ ID NO:9、SEQ ID NO:11、SEQ ID NO:13、SEQ ID NO:15、SEQ ID NO:17、SEQ ID NO:19、 SEQ ID NO:21、SEQ ID NO:23、SEQ ID NO:25、SEQ ID NO:27、SEQ ID NO:29、SEQ ID NO:31、SEQ ID NO:33、SEQ ID NO:35、SEQ ID NO:37、 SEQ ID NO:39、SEQ ID NO:41、SEQ ID NO:43、SEQ ID NO:45或SEQ ID NO:47具有至少75％、76％、77％、78％、79％、80％、81％、82％、83％、 84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、 96％、97％、98％或者至少99％的序列一致性。在一些实施方案中,核酸序列编码具有一个或多个取代的CYP153A-还原酶杂交融合多肽变体,这 会得到本发明所述的改良的特征和/或性质。在另一个相关的实施方案中, 用于实施本发明公开

的CYP153A-还原酶杂交融合多肽变体是由如下核苷 酸序列编码的,该核苷酸序列与SEQ ID NO:7、SEQ ID NO:9、SEQ ID NO: 11、SEQ ID NO:13、SEQ ID NO:15、SEQ ID NO:17、SEQ ID NO:19、SEQ ID NO:21、SEQ ID NO:23、SEQ ID NO:25、SEQ ID NO:27、SEQ ID NO:29、SEQ ID NO:31、SEQ ID NO:33、SEQ ID NO:35、SEQ ID NO:37、 SEQ ID NO:39、SEQ ID NO:41、SEQ ID NO:43、SEQ ID NO:45或SEQ ID NO:47具有100％、99％、98％、97％、96％、95％、94％、93％、92％、 91％或90％的序列一致性。在另一个方面中,本发明公开涉及涵盖了由如 下核酸序列编码的氨基酸序列的CYP153A-还原酶杂交融合多肽变体,所 述核酸序列在严谨条件下在与SEQ ID NO:7、SEQ ID NO:9、SEQ ID NO:11、SEQ ID NO:13、SEQ ID NO:15、SEQ ID NO:17、SEQ ID NO:19、 SEQ ID NO:21、SEQ ID NO:23、SEQ ID NO:25、SEQ ID NO:27、SEQ ID NO:29、SEQ ID NO:31、SEQ ID NO:33、SEQ ID NO:35、SEQ ID NO:37、 SEQ ID NO:39、SEQ ID NO:41、SEQ ID NO:43、SEQ ID NO:45或SEQ ID NO:47相应的基本全长的核酸序列上杂交。在一些实施方案中, CYP153A-还原酶杂交融合多肽变体衍生自水油海杆菌菌种。在其他的实 施方案中,P450杂交融合多肽衍生自不动杆菌属的菌种、海分枝杆菌、Polaromonas sp.、泊库岛食烷菌、真菌伯克霍尔德菌(Burkholderia fungorum)、新月柄杆菌(Caulobacter crescentus)、Hyphomonas neptunium、 沼泽红假单胞菌(Rhodopseudomonas palustris)、鞘氨醇单胞菌属的菌种、分支杆菌属的菌种。

[0163] 改变和突变

[0164] 本发明使用的变体多肽是指这样的多肽,其具有与野生型或模板多肽 至少一个不同的氨基酸的氨基酸序列。例如变体(例如突变体)可以具有 以下保守氨基酸取代的一个或多个,包括但不限于脂肪族氨基酸(例如丙 氨酸,缬氨酸,亮氨酸和异亮氨酸)被另一个脂肪族氨基酸代替;丝氨酸 被苏氨酸代替;苏氨酸被丝氨酸代替;酸性残基(例如天冬氨酸和谷氨酸) 被另一个酸性残基代替;带有酰胺基团的残基(例如天冬酰胺和谷氨酰胺)被另一个带有酰胺基团的残基代替;碱性残基(例如赖氨酸和精氨酸)与 另一个碱性残基交换;以及芳香族残基(例如苯丙氨酸和酪氨酸)被另一 个芳香族残基代替。在一些实施方案中,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。发挥变体或突变体功能的多肽的一些 优选的片段保留了相应的野生型多肽的一些或所有生物功能(例如酶活 性)。在一些实施方案中,所述的片段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,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。在其他的实施方案中,一些片段展现出比相应的野生型多肽 增高的生物学功能。测定哪个氨基酸残基可以被取代、插入或删除而不影 响生物学活性的指南可以使用本领域公知的计算机程序来发现,例如 LASERGENE软件(DNASTAR,Inc.,Madison,WI)。在一些实施方案中, 片段展现出比相应的野生型多肽或模板多肽增高的生物学功能。例如,片段可以展现出比相应的野生型多肽或模板多肽改良至少10％、至少25％、 至少50％、至少75％或者至少90％的酶活性。在其他的实施方案中,所述 的片段展现出比相应的野生型多肽或模板多肽改良至少100％、至少200％ 或者至少500％的酶活性。

[0165] 应该理解的是本发明所述多肽可以具有不实质上影响多肽功能的额 外保守的或非必需的氨基酸取代。如本领域所知(参见Bowie et al.(1990)Science,247:1306-1310),可以测定特定的取代是否被允许(即,不会不利 地影响所需的生物学功能,例如ω-羟化酶

的酶活性)。保守的氨基酸取代 为其中氨基酸残基被具有相似侧链的氨基酸残基所代替的取代。本领域已 经定义了具有相似侧链的氨基酸残基的家族。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),具有酸性侧链的氨基酸(例如 天冬氨酸,谷氨酸),具有不带电荷的极性侧链的氨基酸(例如甘氨酸,天 冬酰氨,谷氨酰胺,丝氨酸,苏氨酸,酪氨酸,半胱氨酸),具有非极性侧链 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、 蛋氨酸,色氨酸),具有β-支化的侧链的氨基酸(例如苏氨酸,缬氨酸,异 亮氨酸),以及具有芳香族侧链的氨基酸(例如酪氨酸,苯丙氨酸、色氨酸, 组氨酸)。

[0166] 变体可以天然形成或者在体外创建。具体而言,可以使用基因工程技 术来创建此类变体,例如定点诱变、随机化学诱变、核酸外切酶III的删 除过程或者标准的克隆技术。备选地,可以使用化学合成或修饰过程来创 建此类变体、突变体、片段、类似物或衍生物。制备变体的方法是本领域 公知的。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 是本领域公知的(例如参见Arnold(1993)Curr.Opin.Biotech.4:450-455)。可以使用易错PCR来取得随机诱变(例如参见Leung et al.(1989)Technique1:11-15;and Caldwell et al.(1992)PCR Methods Applic.2:28-33)。在易错 PCR中,实际的PCR是在以下条件下实施的,其中DNA聚合酶的复制保 真度较低,使得在沿着PCR产物的全长上获得高比率的点突变。简言之, 在此类过程中,将待诱变的核酸(例如编码P450蛋白质或P450杂交融合 多肽的多核苷酸序列)与PCR引物、反应缓冲剂、$MgCl_2$、$MnCl_2$、Taq 聚合酶以及合适浓度的dNTP混合,用于在沿着PCR产物的全长上获得高 比率的点突变。例如可以使用20fmoles待诱变的核酸,30pmole各种PCR 引物,反应缓冲剂(包含50mMKCl,10mM Tris HCl(pH 8.3),0.01%凝胶,7 mM $MgCl_2$,0.5mM $MnCl_2$,5单位的Taq聚合酶,0.2mM dGTP,0.2mM dATP,1mM dCTP和1mM dTTP)来实施反应。PCR可以实施30个循环: 94℃,1min;45℃,1min;以及72℃,1min。但是,本领域的那些技术 人员应该理解的是这些参数可以适当地改变。然后,将诱变的核酸克隆到合适的载体中,并评价由诱变的核酸所编码的多肽的活性。可以使用寡核 苷酸定点诱变来实施定点诱变,从而在所关注的任何克隆的DNA中生成 点位特异性的突变。寡核苷酸诱变在本领域中有所描述(例如参见Reidhaar-Olson et al.(1988)Science 241:53-57)。简言之,在此类过程中, 带有一个或多个突变的待引入至克隆的DNA中的多个双链寡核苷酸被合 成,并插入至待诱变的克隆的DNA中(例如编码P450多肽或P450杂交 融合多肽的多核苷酸序列)。回收包含诱变的DNA的克隆,并评估它们所 编码的多肽的活性。

[0167] 用于生成变体的另一种方法是组装PCR。组装PCR涉及由小的DNA 片段的混合物组装成PCR产物。大量的不同的PCR反应在小瓶中平行进 行,其中一个反应的产物引发了另一反应的产物(参见美国专利 5,965,408)。生成变体的另一种方法为有性PCR诱变。在有性PCR诱变中, DNA分子的随机片段基于序列同源性的结果是在不同的但是高度相关的 体外DNA序列的DNA分子之间发生被迫的同源重组。然后,在PCR反 应中,通过引物的延伸而使得交叠(crossover)被固定。有性PCR诱变在 本领域已知的出版物中有所描述(例如参见Stemmer(1994)Proc.Natl. Acad.Sci.U.S.A.91:10747-10751)。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)中扩增所述的 序列而在核酸序列中生成随机突变,所述菌株在一个或多个DNA修复途 径中携带突变。此类增变基因菌株比野生型菌株具有更高的随机突变率。在这些菌株中的一种菌株中扩增DNA

序列(例如编码P450杂交融合多肽 的多核苷酸序列)将在DNA中最终生成随机突变。适用于体内诱变的增 变基因菌株在本领域的出版物中有所描述(例如参见国际专利申请公开No.WO1991/016427)。此外,还可以使用盒式诱变来生成变体。在盒式诱变中, 双链DNA分子的一段小的区域被不同于天然序列的合成寡核苷酸盒代替。 寡核苷酸通常包含全部的和/或部分的随机化的天然序列。此外,还是可以 使用递归整体诱变(Recursive ensemble mutagenesis)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,研发该算法用于生 产表现型相关的突变体的不同群体,这些群体的成员的氨基酸序列不同。该方法使用反馈机制来控制连续几轮的组合的盒式诱变(例如参见Arkin et al.(1992)Proc.Natl.Acad.Sci.,U.S.A.89:7811-7815)。在一些实施方案中,变体是使用指数整体诱变(exponential ensemble mutagenesis)创建的。指 数整体诱变是使用高百分率的独特并功能性的变体来生成组合文库的方 法,其中几个小组的残基平行地随机化,从而识别各改变位点的、形成功 能蛋白质的氨基酸(例如参见Delegrave et al.(1993)Biotech.Res.11:1548-1552)。在一些实施方案中,使用改组过程来创建变体,其中在大 量编码不同的多肽的核酸中,其中的一部分核酸被融合在一起,从而创建 编码嵌合多肽的嵌合核酸序列(例如在美国专利No.5,965,408和5,939,250 中所述)。

[0168] 表达载体

[0169] 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,重组载体包括选自以下的至少一个序列:与多核苷酸序列可操作地缀合的表达控制序列;与多核苷酸序列可操作地缀合的选择标记; 与多核苷酸序列可操作地缀合的标记序列;与多核苷酸序列可操作地缀合 的纯化部分;与多核苷酸序列可操作地缀合的分泌序列;以及与多核苷酸 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适用于在宿主 细胞中表达多核苷酸序列的形式的多核苷酸序列。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,从而生产由上文所述的多核苷酸序列(如上所述)编码的多肽, 包括融合多肽。在原核生物(例如大肠杆菌)中,表达编码多肽的基因大 部分是使用包括构成型或诱导型启动子的载体来实施的,其中所述的启动 子指导融合或非融合多肽的表达。融合载体将一定数量的氨基酸加入到其 中编码的多肽中,通常加入到重组多肽的氨基末端或羧基末端。这种融合 载体通常起到以下3种目的中的一种或多种:增加重组多肽的表达;增加重组多肽的溶解性;以及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载体(Pharmacia Biotech,Inc.,Piscataway,NJ;Smith et al.(1988)Gene 67:31-40)、pMAL载体(New England Biolabs,Beverly,MA)、和pRITS载体(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。

[0170] 诱导型非融合大肠杆菌表达载体的实例包括pTrc载体(Amann et al. (1988)

Gene 69:301-315)和pET11d载体(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(例如BL21(DE3)和HMS174(DE3)),由容留T7gn1基因的居民λ噬菌体在 lacUV5启动子的转录控制之下来供应的。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.(1989)Molecular Cloning:A Laboratory Manual,second edition, Cold Spring Harbor Laboratory)。诱导 型非融合大肠杆菌表达载体的实例包括pTrc载体(Amann et al.(1988)Gene69:301-315)和PET11d载体(Studier et al.(1990)Gene Expression Technology:Methods in Enzymology 185,Academic Press,San Diego,CA, pp.60-89)。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.(如上所述)中找到。就稳定地转化细菌细胞而言,已知(取 决于所使用的表达载体和转化技术)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,编码可选择标记(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。可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 可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。

[0171] 可任选的途径改造

[0172] 本发明公开的宿主细胞或微生物有机体包括宿主菌株或宿主细胞,这 些宿主菌株或宿主细胞经基因改造或修饰而包含一些改变,从而检验特定 突变对酶活性(即,重组细胞或微生物有机体)的功效。根据原始宿主细胞中存在的天然酶途径的模式,可以由一种宿主至另一种宿主中交替地使 用多种可任选的基因操作和改变。在一个实施方案中,可以使用宿主菌株 来检验与其他生物合成多肽(例如酶)组合的CYP153A-还原酶杂交融合 多肽或其变体的表达。宿主菌株可以涵盖大量的基因改变,从而检验特定 的变量,其包括但不限于培养条件,包括发酵成分、碳源(例如供料)、 温度、压力、减低的培养物污染条件和氧的水平。

[0173] 在一个实施方案中,宿主菌株涵盖可任选的fadE和fhuA的删除。酰 基-CoA脱氢酶(FadE)为对于代谢脂肪酸而言重要的酶。其催化脂肪酸利用 (β-氧化)中的第二步骤,其为将长链脂肪酸(酰基-CoA)断裂成乙酰基 -CoA分子的方法。更具体而言,细菌中脂肪酸降解的β-氧化循环的第二 步骤是将酰基-CoA氧化成2-烯醇-CoA,这是由FadE催化的。当大肠杆菌 缺乏FadE时,其不能在脂肪酸作为碳源上生长,但是可以在乙酸酯上生 长。不能利用任何链长的脂肪酸符合对fadE菌株的所报告的表现型,即, 其中FadE功能被破坏的fadE突变体菌株。fadE基因可以可任选地敲除或 减弱,以确保可以作为脂肪酸衍生物途径的中间体的酰基-CoA可以在细胞 中累积,这样所有的酰基-CoA可以有效地转化成脂肪酸衍生物。但是,由于在使用糖作为碳源的条件下,FadE的表达可能被阻抑,因此FadE可以 仅以少量存

在并且不能有效地与用于酰基-CoA底物的酯合酶或其他酶竞 争,所以当使用糖作为碳源时,fadE减弱是可任选地。FadE由于分解代 谢物阻抑而被阻抑。大肠杆菌和许多其他的微生物优选消耗糖超过脂肪酸,这样当2种来源均可得到时,通过阻抑fad调节子而首先消耗糖(参 见D.Clark,J Bacteriol.(1981)148(2):521-6)。此外,缺乏糖而存在脂肪酸 会诱导FadE的表达。由于fad调节子(包括FadE)表达的蛋白质受到上 调,并会有效地夺取酰基-CoA,所以酰基-CoA中间体可以丧失于β氧化 途径。因此,有利的是使fadE基因被敲除或减弱。因为大多数碳源主要是 糖基,可选的是减弱FadE。基因fhuA编码TonA蛋白质,其为大肠杆菌 的外膜中的能量偶联的运送因子和受体(V.Braun(2009)J Bacteriol. 191(11):3431-3436)。其删除是可任选的。fhuA删除使得细胞对噬菌体的 攻击更具有抗性,这在某些发酵条件下是有利的。因此,理想的是删除宿主细胞中的fhuA,其中所述的细胞在发酵运行中可能经历了潜在的污染。

[0174] 在另一个实施方案中,宿主菌株(如上所述)还涵盖了以下一种或多 种基因的可任选的过表达:fadR、fabA、fabD、fabG、fabH、fabV,和/或fabF。此类基因的实例为得自大肠杆菌的fadR、得自鼠伤寒沙门氏杆菌 (Salmonella typhimurium)(NP_460041)的fabA、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、得自霍乱弧菌(YP_001217283)的 fabV、以及得自丙酮丁醇梭菌(Clostridium acetobutylicum)(NP_350156) 的fabF。这些基因(其编码脂肪酸生物合成的酶和调节因子)的一种或多 种的过表达可以在多种培养条件下增加脂肪酸衍生物化合物的效价,包括 ω-OH脂肪酸及其衍生物。

[0175] 在另一个实施方案中,使用大肠杆菌菌株作为用于生产ω-OH脂肪酸 及其衍生物的宿主细胞。类似地,这些宿主细胞提供了一种或多种生物合 成基因(即,编码脂肪酸生物合成的酶和调节因子的基因)的可任选的过 表达,这样可以在多种培养条件下进一步增加或增强脂肪酸衍生物化合物 例如脂肪酸衍生物(例如ω-OH脂肪酸和α,ω-二酸等)的效价,其中所述 的基因包括但不限于fadR、fabA、fabD、fabG、fabH、fabV和/或fabF。基因改变的实例包括得自大肠杆菌的fadR、得自鼠伤寒沙门氏杆菌 (NP_460041)的fabA,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,得自霍乱弧菌(YP_001217283)的fabV,以及得自丙酮丁醇梭菌 (NP_350156)的fabF。在一些实施方案中,携带这些生物合成基因的合成 操纵子可以被改造并在细胞中表达,从而得以在多种培养条件下检验P450 的表达和/或进一步增强ω-OH脂肪酸和α,ω-二酸的生产。此类合成操纵子 包含一种或多种生物合成基因。例如ifab138操纵子为改造的操纵子,其 包含可任选的脂肪酸生物合成基因,包括得自霍乱弧菌的fabV,得自鼠伤 寒沙门氏杆菌的fabH,得自鼠伤寒沙门氏杆菌的fabD,得自鼠伤寒沙门 氏杆菌的fabG,得自鼠伤寒沙门氏杆菌的fabA,和/或得自丙酮丁醇梭菌 的fabF,这些基因可以用于促进脂肪酸衍生物的过表达,以便检验特定的培养条件。此类合成操纵子的一个益处在于ω-OH脂肪酸衍生物生产的速 率可以被进一步增加或增强。

[0176] 在一些实施方案中,用于表达ACP和生物合成酶(例如ω-羟化酶,硫 酯酶等)的宿主细胞或微生物有机体进一步表达涵盖某些酶活性的基因, 其中所述的酶活性可以增加一种或多种特定脂肪酸衍生物的生产,例如 ω-OH脂肪酸、ω-OH脂肪酸衍生物、α,ω-二酸等。在一个实施方案中,所 述的宿主细胞具有用于生产脂肪酸的硫酯酶活性

(E.C.3.1.2.*、E.C.3.1. 2.14或E.C.3.1.1.5),其中所述的硫酯酶活性可以通过过表达所述的基因 而增加。在另一个实施方案中,所述的宿主细胞具有用于生产脂肪酯的酯合酶活性(E.C.2.3.1.75)。在另一个实施方案中,所述的宿主细胞具有用 于生产脂肪醇的酰基-ACP还原酶(AAR)(E.C.1.2.1.80)活性和/或醇脱氢酶 活性(E.C.1.1.1.1.)和/或脂肪醇酰基-CoA还原酶(FAR)(E.C.1.1.1.*)活性和/ 或羧酸还原酶(CAR)(EC 1.2.99.6)活性。在另一个实施方案中,所述的宿主 细胞具有用于生产脂肪醛的酰基-ACP还原酶(AAR)(E.C.1.2.1.80)活性。在另一个实施方案中,所述的宿主细胞具有用于生产烷和烯的酰基-ACP 还原酶(AAR)(E.C.1.2.1.80)活性和脱羧酶(ADC)活性。在另一个实施方案 中,所述的宿主细胞具有用于生产脂肪醇的酰基-CoA还原酶(E.C.1.2.1.50) 活性,酰基-CoA合酶(FadD)(E.C.2.3.1.86)活性和硫酯酶(E.C.3.1.2.*、E.C. 3.1.2.14或E.C.3.1.1.5)活性。在另一个实施方案中,所述的宿主细胞具有 用于生产脂肪酯的酯合酶活性(E.C.2.3.1.75),酰基-CoA合酶(FadD)(E.C. 2.3.1.86)活性和硫酯酶(E.C.3.1.2.*、E.C.3.1.2.14或E.C.3.1.1.5)活性。在另一个实施方案中,所述的宿主细胞具有用于生产内烯烃的OleBCD活性。在另一个实施方案中,所述的宿主细胞具有用于生产脂肪醇的酰基-ACP 还原酶(AAR)(E.C.1.2.1.80)活性和醇脱氢酶活性(E.C.1.1.1.1.)。在另一个 实施方案中,所述的宿主细胞具有用于制备末端烯烃的硫酯酶(E.C. 3.1.2.*、E.C.3.1.2.14或E.C.3.1.1.5)活性和脱羧酶活性。酶活性在微生物 有机体和微生物细胞中的表达由美国专利号8,097,439;8,110,093; 8,110,670;8,183,028;8,268,599;8,283,143;8,232,924;8,372,610和8,530,221 所教导,这些文献以引用方式并入本文。在其他的实施方案中,用于表达ACP和其他生物合成酶的宿主细胞或微生物有机体包含某些天然的酶活 性,这些酶受到上调或被过表达、从而生产一种或多种特定的脂肪酸衍生 物(多个),例如ω-OH脂肪酸、ω-OH脂肪酸衍生物和α,ω-二酸。在一个 实施方案中,所述的宿主细胞具有用于生产脂肪酸的天然硫酯酶(E.C. 3.1.2.*、E.C.3.1.2.14或E.C.3.1.1.5)活性,其中所述的硫酯酶可以通过过表达所述的硫酯酶基因而增加。

[0177] 本发明公开包括宿主菌株或微生物有机体,其表达编码CYP153A-还 原酶杂交融合多肽及其变体、以及其他的生物合成酶(如上所述)的基因。重组宿主细胞生产脂肪酸衍生物,包括ω-羟基脂肪酸衍生物(例如ω-羟基 脂肪酸,包括ω-羟基游离脂肪酸;ω-羟基脂肪酸甲基酯;ω-氧代脂肪酸; ω-氧代脂肪酸甲基酯;α,ω-二酸;α,ω-二醇;α,ω-二酯;ω-羧基脂肪酸甲基 酯;ω-氨基脂肪酸和ω-氨基脂肪酸甲基酯),以及它们的组合物和共混物。脂肪酸衍生物通常由培养物培养基中回收,和/或与宿主细胞分离。在一个 实施方案中,脂肪酸衍生物由培养物培养基(细胞外)中回收。在另一个实施方案中,脂肪酸衍生物由宿主细胞(细胞内)回收。在另一个实施方 案中,脂肪酸衍生物由培养物培养基中回收并与宿主细胞分离。可以使用 本领域已知的方法(例如GC-FID)来分析由宿主细胞生产的脂肪酸衍生 物组合物,从而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(例如ω-羟基脂肪酸,包括ω-羟基游离脂肪酸;ω-羟基脂肪酸甲基 酯;ω-氧代脂肪酸;ω-氧代脂肪酸甲基酯;双功能化合物,例如α,ω-二酸; α,ω-二醇;α,ω-二酯;ω-羧基脂肪酸甲基酯;ω-氨基脂肪酸;ω-氨基脂肪 酸甲基酯等)的链长和饱和度。

[0178] 在其他的实施方案中,用于表达ACP和其他的生物合成酶的宿主细胞 或微生物有

机体将包含天然的酶活性,这些酶受到上调或被过表达,从而 生产一种或多种特定的脂肪酸衍生物,包括ω-羟基脂肪酸衍生物(例如ω- 羟基脂肪酸,包括ω-羟基游离脂肪酸;ω-羟基脂肪酸甲基酯;ω-氧代脂肪 酸;ω-氧代脂肪酸甲基酯;双功能化合物,例如α,ω-二酸;α,ω-二醇;α,ω- 二酯;ω-羧基脂肪酸甲基酯;ω-氨基脂肪酸和ω-氨基脂肪酸甲基酯)。在 一个实施方案中,所述的宿主细胞具有用于生产脂肪酸的天然硫酯酶(E.C.3.1.2.*、E.C.3.1.2.14或E.C.3.1.1.5)活性,其中所述的硫酯酶可以通过过 表达硫酯酶基因而增加。

[0179] 重组宿主细胞和发酵

[0180] 为了生产ω-羟基脂肪酸衍生物,进行了大量的修饰,从而产生宿主细 胞(如上所述)。因此,本发明公开提供了重组宿主细胞,其经改造而提供了相对于未改造的或天然宿主细胞(例如发挥对照细胞作用的野生型宿 主细胞)而言的ω-脂肪酸生物合成途径,这是通过例如特定菌株的改良而 完成的。本发明公开的生产宿主有机体包括植物、动物或微生物细胞。诸 如细菌、蓝细菌、酵母、海藻或丝状真菌之类的微生物有机体或微生物细 胞可以用作生产宿主。可以用作生产宿主的微生物有机体的非限定性实例 包括大肠杆菌、酿酒酵母等(如上所述)。微生物菌株将来自可再生的供 料的葡萄糖或其他碳源有效转化成脂肪酸或脂肪酸酯,例如脂肪酸甲基酯 (FAME)、脂肪酸乙酯(FAEE)和脂肪醇(FALC)。为了达到上述目的,将所 述的菌株仔细地改造从而表达关键的酶,包括用于生产脂肪酸的硫酯酶(例如得自大肠杆菌的TesA)或者用于生产FAME的酯合酶(例如得自 除烃海杆菌的ES9)。用于生产多种化合物的高密度发酵的方案和过程已经建立(参见美国专利No.8,372,610;8,323,924;8,313,934;8,283,143; 8,268,599;8,183,028;8,110,670;8,110,093和8,097,439,这些文献以引用 方式并入本文)。

[0181] 可以使用基因工程技术将本发明所讨论的酶步骤(其发挥生物催化剂 的作用)加入这些微生物有机体中,从而创建能够生产双功能脂肪酸衍生 物(例如图1和2中所示的α,ω-二酸,以及图3和4中所示的α,ω-二醇) 的新型微生物有机体。预计所述的产物被分泌至细胞的外部,从而允许通 过离心而容易地收获。某些重组酶步骤可以与一种微生物有机体结合,用于直接生产特定的化合物,例如图1和2所示的12-氨基十二烷酸或12- 氨基十二烷酸甲基酯。备选地,使用由可再生的来源生产的中间体的生物 转化可以用于本发明公开的发酵方法中。例如由表达酶(例如硫酯酶或酯 合酶)的重组微生物有机体得到的十二烷酸甲基酯可以被供入表达酶(例 如醇脱氢酶或氧化酶;氨基转移酶或转氨酶;和/或酯酶或脂肪酶)的重组 微生物有机体中。这样,由表达催化某些步骤的某些酶的一种重组微生物有机体得到的化学实体可以被供入表达催化其他步骤的酶的另一种重组 微生物有机体中。因此,所述的宿主细胞提供了生产系统,其中中间体是 可交换的,并且发酵过程可以用于所有的宿主细胞中,从而生成所需的ω- 羟基脂肪酸衍生物。值得注意的是,由葡萄糖或其他的可再生的供料(除 了外源脂肪酸或石蜡以外)直接且有效地生产ω-羟基脂肪酸衍生物(例如α,ω-二酸;α,ω-二醇或ω-氨基脂肪酸)的方法到目前为止并不存在。但是, 这些双功能脂肪酸衍生物对于聚合物的合成而言是重要的前体。如本发明 所提出的用于生产ω-羟基脂肪酸衍生物的基于发酵的方法提供了对于本领域使用的化学方法的快速且环境友好的备选。

[0182] 本发明公开提供了由可再生的供料(例如得自玉米、藤条或木质纤维 素生物质的

碳水化合物;或者诸如甘油、废气、合成气的废品;或者改进 的有机材料,例如生物质)衍生的碳源直接生产ω-羟基脂肪酸衍生物。在 一个实施方案中,所述方法依赖于用于生产这些重要化学品的成本有效且 可再生的备选来源。所述方法包括通过以下过程来生产ω-羟基脂肪酸衍生 物:在发酵液中提供重组微生物有机体(例如宿主细胞);将可再生的供 料加入发酵液中;以及由发酵液分离ω-羟基脂肪酸衍生物。特定的微生物有机体的宿主细胞包含经改造而表达编码如下多肽的至少2种核酸序列的 途径,所述多肽包括硫酯酶或酯合酶、以及ω-羟化酶。

[0183] 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,重组 宿主细胞将由可再生的供料衍生的碳源转化为ω-OH脂肪酸衍生物及其组 合物。允许生产的条件是指允许宿主细胞生产所需产物(例如ω-OH脂肪 酸衍生物)的任何条件。类似地,其中载体的多核苷酸被表达的一种或多种条件是指允许宿主细胞合成多肽的任何条件。合适的条件包括例如发酵条件。发酵条件可以包括例如许多参数,包括但不限于温度范围、通风水 平、供料速率和培养基组成。这些条件(单独及组合)的每一种均允许宿 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。

[0184] 就小规模的生产而言,经改造的宿主细胞可以以例如大约100μL、200 μL、300μL、400μL、500μL、1mL、5mL、10mL、15mL、25mL、50mL、 75mL、100mL、500mL、1L、2L、5L或10L的批次生长;发酵;并 被诱导而表达所需的多核苷酸序列,例如编码单独的CYP153A-还原酶杂 交融合多肽或者与其他酶功能性组合的多核苷酸序列。就大规模的生产而 言,经改造的宿主细胞可以以例如大约10L、100L、1000L、10,000L、 100,000L、1,000,000L或更大的批次生长;发酵;并被诱导而表达所需的 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 ω-OH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。ω-OH脂肪酸或其衍生物可 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 离其ω-OH脂肪酸衍生物。

[0185] 在一些实施方案中,宿主细胞在包含初始浓度的碳源(例如大约2g/L 至大约100g/L的可再生的供料)的培养物培养基中培养。在其他的实施 方案中,培养物培养基包含初始浓度为大约2g/L至大约10g/L的碳源,大 约10g/L至大约20g/L的碳源,大约20g/L至大约30g/L的碳源,大约30 g/L至大约40g/L的碳源或者大约40g/L至大约50g/L的碳源。在一些实 施方案中,针对培养物培养基中碳源的水平来监测发酵情况。在一些实施 方案中,所述方法进一步包括当培养基中的碳源低于大约0.5g/L时,向该培养物培养基中加入补充碳源。在一些实施方案中,当培养基中碳源的水 平低于大约0.4g/L,低于大约0.3g/L,低于大约0.2g/L或者低于大约0.1 g/L时,将补充的碳源加入培养物培养基中。在一些实施方案中,加入补 充的碳源,从而保持碳源的水平为大约1g/L至大约25g/L。在一些实施 方案中,加入补充的碳源,从而保持碳源的水平为大约2g/L或更高(例 如大约2g/L或更高,大约3g/L或更高,大约4g/L或更高)。在某些实施 方案中,加入补充的碳源,从而保持碳源的水平为大约5g/L或更低(例如大约5g/L或更低,大约4g/L或更低,大约3g/L或更低)。在一些实施

方案中,加入补充的碳源,从而保持碳源的水平为大约2g/L至大约5g/L, 大约5g/L至大约10g/L或者大约10g/L至大约25g/L。在一些实施方案 中,碳源为葡萄糖或另一种类型的可再生的供料,例如甘油。

[0186] 在一些实施方案中,生产浓度为大约1g/L至大约200g/L的ω-羟基 脂肪酸衍生物。在一些实施方案中,生产浓度为大约1g/L或更高(例如 大约1g/L或更高,大约10g/L或更高,大约20g/L或更高,大约50g/L或 更高,大约100g/L或更高)的ω-羟基脂肪酸衍生物。在一些实施方案中, 生产浓度为大约1g/L至大约170g/L、大约1g/L至大约10g/L、大约40g/L 至大约170g/L、大约100g/L至大约170g/L、大约10g/L至大约100g/L、大约1g/L至大约40g/L、大约40g/L至大约100g/L或者大约1g/L至大 约100g/L的ω-羟基脂肪酸衍生物。

[0187] 在一些实施方案中,生产效价为大约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525 mg/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750 mg/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1425 mg/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1850 mg/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2个值所 限定的范围的ω-羟基脂肪酸衍生物。在其他的实施方案中,生产效价为高 于100g/L,高于200g/L,高于300g/L或者更高(例如500g/L,700g/L,1000g/L,1200g/L,1500g/L或2000g/L)的ω-脂肪酸衍生物。根据本发明公开 的方法由重组宿主细胞生产的ω-羟基脂肪酸衍生物的优选的效价为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ω-羟基脂肪酸 衍生物的效价为大约1g/L至大约250g/L,更具体为90g/L至大约120g/L。 效价是指由给定的重组宿主细胞培养物生产的具体的ω-脂肪酸衍生物或ω-脂肪酸衍生物的组合。

[0188] 在其他的实施方案中,根据本发明公开的方法经改造而生产ω-羟基脂 肪酸衍生物的宿主细胞具有至少1%、至少2%、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%、至少40%或者由上述任意2个值所限定的范围的产率。在其他的实施方案中,生产产率为高于30%、40%、50%、60%、70%、80%、90%或更高的ω-羟基脂肪酸衍生物或衍生物。备选地或此外,产率为大约30%

或更低,大约27％或更低,大约25％或更低或者大约22％或更低。因此,所述的产率可以由上述任意2个端点所限定。例如根据本发明公开的方法由重组宿主细胞生产的ω-羟基脂肪酸衍生物或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％或者20％至 30％。在具体的实施方案中,由重组宿主细胞生产的ω-羟基脂肪酸衍生物或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ω-羟基脂肪酸衍生物或衍生物的产率为大约25％至大约30％。产率可以指由给定的重组宿主细胞培养物生产的具体的ω-羟基脂肪酸衍生物或ω-羟基脂肪酸衍生物的组合。此外,产率还取决于所使用的供料。

[0189] 在一些实施方案中,由重组宿主细胞生产的ω-羟基脂肪酸衍生物或衍 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1000 mg/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2300 mg/L/小时、至少2400mg/L/小时或至少2500mg/L/小时。例如根据本发明 公开的方法由重组宿主细胞生产的ω-羟基脂肪酸衍生物或多种衍生物的生产率可以为500mg/L/小时至2500mg/L/小时,或者700mg/L/小时至 2000mg/L/小时。在一个具体的实施方案中,生产率为大约0.7g/L/h至大 约3g/L/h。生产率可以指由给定的重组宿主细胞培养物生产的具体的ω- 羟基脂肪酸衍生物或ω-羟基脂肪酸衍生物的组合。

[0190] 用于增加重组宿主细胞生产ω-OH脂肪酸组合物的策略包括通过在生产型宿主中过表达CYP153A-还原酶杂交融合基因和硫酯酶基因而增加的通过脂肪酸生物合成途径的流量。如本文所用,术语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/或经修饰的基因元件。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(Myceliophtora)、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 (Stenotrophamonas)、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。在一些实 施方案中,宿主细胞为大肠杆菌B细胞、大肠杆菌C细胞、大肠杆菌K 细胞或者大肠杆菌W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 lichenoformis)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、 或解淀粉牙胞杆菌(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 fumigates)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 lanuginose)细胞、不透明红球菌细胞、米黑根毛霉(Rhizomucormiehei)细胞或米黑毛霉(Mucormichei)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。在其他的实施方案中,宿 主细胞为放线菌属(Actinomycetes)细胞。在一些实施方案中,宿主细胞 为酿酒酵母(Saccharomyces cerevisiae)细胞。

[0191] 在其他的实施方案中,宿主细胞为真核植物细胞、海藻细胞、蓝细菌 细胞、绿硫细菌细胞、绿色非硫细菌细胞、紫硫细菌细胞、紫色非硫细菌 细胞、嗜极生物细胞、酵母细胞、真菌细胞、本发明所述的任何物种的经 改造的细胞或合成有机体。在一些实施方案中,宿主细胞为光依赖的或固 定碳。在一些实施方案中,宿主细胞具有自养活性。在一些实施方案中,例如在光存在下,宿主细胞具有光合自养活性。在一些实施方案中,在光 缺乏下,宿主细胞是异养的或兼养的。在某些实施方案中,宿主细胞为得 自拟南芥(Arabidopsis thaliana)、柳枝稷(Panicum virgatum)、巨芒 (Miscanthus giganteus)、玉米(Zea mays)、葡萄藻(Botryococcuse braunii)、 莱茵衣藻(Chlamydomonas reinhardtii)、杜氏盐藻(Dunaliela 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 elongates) BP-1、绿硫细菌(Chlorobium tepidum)、Chlorojlexus auranticus、Chromatiumm vinosum、深红红螺菌(Rhodospirillum rubrum)、荚膜红细菌(Rhodobacter capsulatus)、沼泽红假单胞菌(Rhodopseudomonas palusris)、杨氏梭菌 (Clostridium ljungdahlii)、热纤梭菌(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)、萤光假单孢菌(Pseudomonas fluorescens)或运动发酵单胞菌(Zymomonas mobilis)的细胞。在一个实 施方案中,微生物细胞得自蓝细菌,其包括但不限于原绿球藻 (Prochlorococcus),聚球藻(Synechococcus),集胞藻(Synechocystis), 蓝杆藻(Cyanothece)和Nostoc punctiforme。在另一个实施方案中,微生物细胞得自特定的蓝细菌物种,其包括但不限于细长聚球藻 (Synechococcus elongatus)PCC7942,集胞藻的物种PCC6803和聚球藻的物种PCC7001。

[0192] 由重组宿主细胞衍生的产物

[0193] 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HOxI和HOxII)定义的相同的含义。基本定义是指 0.95乘以$^{14}C/^{12}C$同位素比HOxI(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。 生物产物(例如脂肪酸衍生物,包括根据本发明公开生产的ω-OH脂肪酸 和衍生物)包括生物学方法生产的有机化合物。具体而言,使用本发明的 脂肪酸生物合成途径生产的脂肪酸衍生物(例如ω-OH脂肪酸及其衍生物) 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No.7,169,588)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($^{13}C/^{12}C$),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$^{13}C/^{12}C$比是大气二氧化碳中该二氧化碳被固定时的$^{13}C/^{12}C$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$^{13}C/^{12}C$ 以及相应的$\delta^{13}C$值中全部表现出显著差异。进一步地,由于代谢途径的原 因,C3和C4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 分的材料不同。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$^{13}C/^{12}C$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。δ13C值被表达为千分比(每密耳), 缩写为‰,并且按照下式进行计算:

[0194] $\delta^{13}C(‰)=[(^{13}C/^{12}C)样品-(^{13}C/^{12}C)标准品]/(^{13}C/^{12}C)标准品x1000$

[0195] 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$\delta^{13}C$。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,包括例如脂肪酸衍生物产 物。

具体而言,所述的生物产物的$\delta^{13}C$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$\delta^{13}C$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$\delta^{13}C$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$1.2x10^{-12}$的稳态宇宙发生(大气)基线 同位素率($^{14}C/^{12}C$),并且大致的驰豫"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$^{14}C/^{12}C$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。

[0196] 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。

[0197] ω-羟基脂肪酸衍生物及配制物

[0198] 诸如ω-羟基脂肪酸衍生物之类的化合物为在许多任务业应用中具有 价值的且理想的分子。本发明公开通过重组微生物有机体在体内生产此类 化合物,由此生成多种有用的产物。此类产物包括ω-羟基脂肪酸衍生物。ω-羟基脂肪酸衍生物包括但不限于ω-羟基脂肪酸;ω-羟基-脂肪酸甲基酯; ω-氨基脂肪酸;ω-氧代脂肪酸,ω-氨基脂肪酸甲基酯;ω-氧代脂肪酸甲基酯; α,ω-二酸;α,ω-二酯;α,ω-二醇以及它们的组合。尽管本发明描述了大部的 偶数链的ω-羟基-脂肪酸衍生物,但是还包括奇数链的ω-羟基-脂肪酸衍生 物,例如具有7、9、11、13、15、19等个碳的那些。

[0199] ω-羟基脂肪酸的实例包括但不限于8-羟基辛酸、10-羟基癸烯酸、10- 羟基癸酸、12-羟基十二碳烯酸、12-羟基十二烷酸、14-羟基十四碳烯酸、14-羟基十四烷酸、16-羟基十六碳烯酸、16-羟基十六烷酸、18-羟基十八碳 烯酸和18-羟基十八碳烯酸。

[0200] ω-羟基-脂肪酸甲基酯的实例包括但不限于8-羟基辛酸甲基酯、10-羟 基癸烯酸甲基酯、10-羟基癸酸甲基酯、12-羟基十二碳烯酸甲基酯、12-羟 基十二烷酸甲基酯、14-羟基十四碳烯酸甲基酯、14-羟基十四烷酸甲基酯、 16-羟基十六碳烯酸甲基酯、16-羟基十六烷酸甲基酯、18-羟基十八碳烯酸 甲基酯和18-羟基十八碳烷酸甲基酯。

[0201] ω-氨基脂肪酸的实例包括但不限于8-氨基辛酸、10-氨基癸烯酸、10- 氨基癸酸、12-氨基十二碳烯酸、12-氨基十二烷酸、14-氨基十四碳烯酸、 14-氨基十四烷酸、16-氨基十六碳烯酸、16-氨基十六烷酸、18-氨基十八碳 烯酸和18-氨基十八碳烯酸。

[0202] ω-氨基脂肪酸甲基酯的实例包括但不限于8-氨基辛酸甲基酯、10-氨基 癸烯酸甲基酯、10-氨基癸酸甲基酯、12-氨基十二碳烯酸甲基酯、12-氨基 十二烷酸甲基酯、14-氨基十四碳烯酸甲基酯、14-氨基十四烷酸甲基酯、 16-氨基十六碳烯酸甲基酯、16-氨基十六烷酸甲基酯、18-氨基十八碳烯酸 甲基酯和18-氨基十八烷酸甲基酯。

[0203] ω-氧代脂肪酸的实例包括但不限于8-氧代辛酸、10-氧代癸烯酸、10- 氧代癸酸、12-氧代十二碳烯酸、12-氧代十二烷酸、14-氧代十四碳烯酸、 14-氧代十四烷酸、16-氧代十六碳烯酸、16-氧代十六烷酸、18-氧代十八碳 烯酸和18-氧代十八碳烷酸。

[0204] ω-氧代脂肪酸甲基酯的实例包括但不限于8-氧代辛酸甲基酯、10-氧代 癸烯酸甲基酯、10-氧代癸酸甲基酯、12-氧代十二碳烯酸甲基酯、12-氧代 十二碳烷酸甲基酯、14-氧代十四碳烯酸甲基酯、14-氧代十四烷酸甲基酯、 16-氧代十六碳烯酸甲基酯、16-氧代十六烷酸甲基酯、18-氧代十八碳烯酸 甲基酯和18-氧代十八碳烯酸甲基酯。

[0205] α,ω-二酸的实例包括但不限于1,8-辛酸、1,10-癸烯二酸(decenedioic acid)、1,10-癸二酸、1,12-十二碳烯二酸、1,12-十二烷二酸、1,14-十四碳 烯二酸、1,14-十四烷二酸、1,16-十六碳烯二酸、1,16-十六烷二酸、1,18- 十八碳烯二酸、1,18-十八烷二酸。

[0206] α,ω-二酸的实例包括但不限于1,8-辛酸、1,10-癸烯二酸(decenedioic acid)、

1,10-癸二酸、1,12-十二碳烯二酸、1,12-十二烷二酸、1,14-十四碳 烯二酸、1,14-十四烷二酸、1,16-十六碳烯二酸、1,16-十六烷二酸、1,18- 十八碳烯二酸、1,18-十八烷二酸。

[0207] α,ω-二酯的实例包括但不限于1,8-辛酸二甲基酯、1,10-癸烯酸二甲基酯、1,10-癸酸二甲基酯、1,12-十二碳烯酸二甲基酯、1,12-十二烷酸二甲基 酯、1,14-十四碳烯酸二甲基酯、1,14-十四烷酸二甲基酯、1,16-十六碳烯酸 二甲基酯、1,16-十六烷酸二甲基酯、1,18-十八碳烯酸二甲基酯、1,18-十八 烷酸二甲基酯。

[0208] α,ω-二醇的实例包括但不限于1,8-辛二醇、1,10-癸烯二醇、1,10-癸二 醇、1,12-十二烯二醇、1,12-十二烷二醇、1,14-十四烯二醇、1,14-十四烷二 醇、1,16-十六烯二醇、1,16-十六烷二醇、1,18-十八烯二醇和1,18-十八烷 二醇。

[0209] 尽管本发明描述了偶数链的ω-羟基-脂肪酸衍生物,但是还包括奇数 链的ω-羟基-脂肪酸衍生物,例如具有7、9、11、13、15、19等个碳的那 些。奇数链ω-羟基脂肪酸的实例包括但不限于例如11-羟基十一碳烯酸、 11-羟基十一酸、13-羟基十三碳烯酸、13-羟基十三酸、15-羟基十五碳烯酸、 15-羟基十五酸、17-羟基十七碳烯酸和17-羟基十七酸。奇数链α,ω-二酸的 实例包括但不限于例如1,11-十一烯二酸、1,11-十二烷二酸、1,13-十三烯二 酸、13-十三酸、1,15-十五烯二酸、1,15-十五酸、1,17-十七烯二酸和1,17-十七酸。奇数链α,ω-二醇的实例包括但不限于例如1,11-十一烯二醇、1,11- 十一烷二醇、1,13-十三烯二醇、13-十三烷二醇、1,15-十五烯二醇、1,15- 十五烷二醇、1,17-十七烯二醇和1,17-十七烷二醇。

实施例

[0210] 以下实施例进一步说明了本发明公开,但是不应该解释为以任何方式 限定了本发明的范围。

[0211] 实施例1:培养用于生产ω-羟基脂肪酸衍生物的重组大肠杆菌菌株

[0212] 分析柱:DB-1HT,15m×250μm×0.1μm,得自Agilent,编号为#J&W 122-1111E

[0213] 炉温:初始为50℃,保持5分钟,以25℃/min升高至300℃,并保持5.24 分钟,总体运行时间为24分钟

[0214] 柱流速:1.2mL/分钟

[0215] 入口温度:300℃

[0216] 样品量:1μL

[0217] 分流比:20:1

[0218] 软件:ChemStation E.02.01.1177

[0219] MS参数

[0220] 传输管路的温度:300℃

[0221] MS源:230℃

[0222] MS Quad:150℃

[0223] 自动取样器

[0224] 由LEAP Technologies分配的Combi PAL(CTC analytics)

[0225] GC/FID参数:

[0226] 装配有FID或等同物的Gas Chromatograph Agilent 7890

[0227] 数据系统ChemStation软件B.04.03,或等同物

[0228] 分析柱:DB-1(10m×0.18mm×0.2μM),或等同物

[0229] 自动取样器:Combi PAL,CTC analytics(Leap Technologies)

[0230] 初始温度:60℃,保持0.5分钟,以25℃/min升高至300℃,并保持0.9分 钟,总体运行时间为11分钟

[0231] 注射器温度:320℃

[0232] 检测器:火焰电离检测器(FID)

[0233] 检测器温度:350℃

[0234] 氢气流速:40mL/分钟

[0235] 空气流速:450mL/分钟

[0236] 组成成分流速:45mL/分钟(N_2)

[0237] 分流比:50:1

[0238] 柱流速:0.8mL/分钟

[0239] 样品量:1μL

[0240] 实施例3:由表达2种cyp153A P450氧化酶操纵子的大肠杆菌菌株将 十二烷酸转化成12(ω)-羟基十二烷酸

[0241] 本实施例显示由表达cyp153A P450氧化酶操纵子的重组大肠杆菌菌 株将外源加入的脂肪酸转化成ω-羟基脂肪酸。本试验的目的在于研究cyp153A P450氧化酶操纵子的效率。到目前为止,cyp153A P450氧化酶仅 在体外使用(例如参见Honda-Malca et al.(2012)Chem.Commun.48:5115-5117),并且本试验的目的是检测当在体内提供外源脂肪酸时,这 些操纵子是否能够生产ω-羟基脂肪酸衍生物。

[0242] 得自2种细菌(Marinobacter aquaeoli(编号YP_957888;SEQ ID NO: 3)和海分枝杆菌(编号YP_001851443;SEQ ID NO:59))的cyp153A操纵 子由这些有机体的基因组DNA中进行PCR扩增。所述的操纵子由编码铁 氧还蛋白(fd)、cyp153A P450氧化酶和铁氧还蛋白还原酶(fdR)的基因组成 (参见表2A和2B,如上所述)。这些基因和基因间区域的自然顺序被保 留,但是得自海分枝杆菌的cyp153A16中的GTG起始密码子通过交叠PCR 被ATG起始密码子代替。将PCR扩增引物(amplimer)克隆至pCL1920- 衍生物载体(SC101复制子,奇霉素抗性标记)中,使得操纵子的转录受到IPTG诱导型Ptrc启动子的控制。所得的质粒pAS.017(参见表11,参见 下文)和pAS.018被转化至大肠杆菌MG1655中,其中fadE基因(编码酰基 -CoA脱氢酶)或fadD基因(编码酰基-CoA合酶)被删除。这些菌株不能降解 脂肪酸,因此可以通过删除这些基因或编码其他脂肪酸降解酶的基因(例 如fadA或fadB)来实现用于增加转化成产物的脂肪酸的利用性。但是, 这是可任选的,并且可以在游离脂肪酸是外源供入或者为产物途径的中间 体时来实施的。(表1(如上所述)提供了在所述的代谢途径内的酶活性的 广泛列举物,包括可以被衰减从而增加脂肪酸在宿主菌株中的利用性的多种脂肪酸降解酶。)4种所得的菌株概括于表12(参见下文)。针对如实施 例1和2中所述的十二烷酸向12-羟基十二烷酸的转化来分析所述的菌株。

[0243] 与对照菌株相比MG1655△fadD(其不表达cyp153A操纵子),在表达 cyp153A操纵子的所有4种菌株的GC-MS色谱中,检测在RT 12.303分钟 的新的峰(在BSTFA衍生化后)(参见图6,其中仅示出菌株sAS.320和 sAS.321)。在RT 12.303分钟的峰的质谱示于图7A中。断

裂方式表明该峰 为12-三甲基甲硅氧基十二烷酸三甲基甲硅烷基酯,其是12-羟基十二烷酸 的衍生化形式。12-三甲基甲硅氧基脂肪酸三甲基甲硅烷基酯的特征离子碎 片示于图8A中。在m/z＝129,147,204和217(图中未示出,但是存在) 的离子为这些化合物的有用的诊断标记。在m/z 255下的离子用于确定在 羧基一侧失去CH_3后的链长(m/z＝345)以及该化合物的羟基一侧的 $(CH_3)_3SiOH$(参见图8A)。该峰的正确鉴定通过与图7B所示的12-三甲基 甲硅氧基十二烷酸三甲基甲硅烷基酯真实标准品的停留时间和质谱比较 而进一步被证实。此外,还记录了该化合物在衍生化之前的质谱,并示于 图7C中。未衍生化的化合物的特征离子碎片示于图8B中。在m/z＝98和 84(显示的)下的离子为用于ω-羟基脂肪酸的有用的诊断标记。在m/z 186 下的离子为12-羟基十二烷酸在羟基一侧失去CH_2O的特征碎片。通过在 m/z 186下将CH_2O加入碎片离子中可以测定所述的化合物的分子离子为 216。

[0244] 因此,表达cyp153A操纵子(得自Marinobacter aquaeoli和海分枝杆 菌)的大肠杆菌在宿主细胞中在体内将外源十二烷酸转化成12-羟基十二 烷酸。这样,可以确认所述的酶实际上可以生产ω-羟基脂肪酸衍生物。但 是该酶的转化效率相当低。因此,预计cyp153A操纵子的酶活性本身对于 生产ω-羟基脂肪酸衍生物而言是不理想的,并且确定需要进一步改造来设 计具有更高的转化效率的酶。

[0245] 表11:用于生产ω-羟基化的脂肪酸衍生物的表达质粒

[0246]

治疗	描述
pAS.017	pCL-fd-cyp153A-fdR_Maqu
pAS.018	pCL-fd-cyp153A16-fdR_Mmar
pAS.022	pCL-fd_synIGR_cyp153A(G307A)-fdR_synIGR_Maqu
pAS.023	pCL-cyp153A(G307A) _Maqu-P450-RedRhF_Rhod
pEP.125	pACYC-cyp153A(G307A) _Maqu-P450-RedRhF_Rhod
pEP.126	pACYC-alkJH_Pput

[0247]

治疗	描述
pSN.09	pACYC-cyp102A1(F87A)
pSN.12	pCL- cyp102A1(F87A)

[0248] 表12:用于将脂肪酸衍生物转化成ω-羟基化的脂肪酸衍生物的、表 达cyp153A的重组大肠杆菌菌株

[0249]

菌株	菌株背景	质粒
sAS.314	MG1655 fadE fhuA	pAS.017
sAS.315	MG1655 fadE fhuA	pAS.018
sAS.320	MG1655 fadD	pAS.017
sAS.321	MG1655 fadD	pAS.018
sAS.335	MG1655 fadD	pAS.022
sAS.336	MG1655 fadD	pAS.023

[0250] 实施例4:由表达经修饰的cyp153A P450氧化酶操纵子的大肠杆菌菌 株将十二烷酸转化成12(ω)-羟基十二烷酸

[0251] 本实施例显示由表达经修饰的cyp153A P450氧化酶操纵子的重组大 肠杆菌菌株将外源加入的脂肪酸转化成ω-羟基脂肪酸。类似地,本试验的 目的是研究这种经修饰的cyp153A P450氧化酶操纵子在体内生产ω-羟基 脂肪酸衍生物的能力和效率。

[0252] 之前已经描述了由M.aquaeoli得到的cyp153A的突变体(其中甘氨 酸307被丙氨酸代替)(参见Honda Malca et al.(2012)Chem.Commun. 48:5115)。按照以下方式来修饰质粒pAS.017(参见实施例3,如上所述): 定义了cyp153_Maqu的Gly307的密码子由GGC变为GCC,其定义Ala, 此外,使用合成的IGR(TAAGGAGGAAAACAAA)(SEQ ID NO:65)代替 cyp156操纵子的2个天然的基因间区域(IGR)。所得的质粒命名为pAS.022 (参见表11,如上所述),并转化至大肠杆菌MG1655ΔfadD中,从而得 到菌株sAS.335(参见表12,如上所述)。针对如实施例1和2中所述的十 二烷酸转化成12-羟基十二烷酸来分析所述的菌株(如上所述)。如图9可 见,与sAS.320相比,sAS.335将更多的十二烷酸转化成12-羟基十二烷酸, 并且12-羟基十二烷酸的量确定为37.4±0.3 mg/L。

[0253] 因此,表达经修饰的cyp153A操纵子(得自Marinobacter aquaeoli) 的大肠杆菌显示当在体内供入外源脂肪酸时,外源十二烷酸向12-羟基十 二烷酸的转化得到一定的改良(与实施例3相比)。但是,这种酶的转化 效率仍是较低的。因此,预计经修饰的cyp153A操纵子(得自Marinobacter aquaeoli)的酶活性本身对于生产ω-羟基脂肪酸衍生物而言是不理想的, 并且确定需要进一步的改造来设计具有较高的转化效率的酶。

[0254] 实施例5:由表达杂交cyp153A-Red450RhF融合蛋白质的大肠杆菌 菌株将脂肪酸衍生物转化成ω-羟基化的脂肪酸衍生物

[0255] 本实施例显示由表达杂交蛋白质的重组大肠杆菌菌株将外源加入的 脂肪酸、脂肪酸甲基酯或脂肪醇分别转化成ω-羟基脂肪酸、ω-羟基脂肪酸 甲基酯或α,ω-二醇,其中cyp153A P450氧化酶与还原酶结构域融合。本试 验的目的是创建杂交融合蛋白质,其中cyp153A P450氧化酶与用于显着改 良ω-羟基脂肪酸衍生物生产的还原酶结构域融合。

[0256] 自给自足的细胞色素P450氧化酶为其中还原酶伴侣与细胞色素P450 催化蛋白质融合的酶。得自红球菌属的菌种NCIMB 9784的P450RhF代表 了一类自给自足的细菌细胞色素P450氧化酶(Roberts et al.(2003)J.Biol. Chem.278:48914;Hunter et al.(2005)FEBS Lett.579:2215),并且成为“I 类P450融合的PFOR”(DeMot and Parret(2002)Trends Microbiol.10:502)。

[0257] 在本试验中,按照如下方式创建编码杂交融合蛋白质(由Marinobacter aquaeoli得到cyp153A(G307A)P450催化蛋白质以及由红球菌属的菌种 NCIMB9784得到的P450RhF的含c-末端FMN-和Fe/S还原酶结构域组成) 的基因:由pAS.022扩增cyp165A(G307A)_Maqu基因,并通过交叠PCR 与密码子优化的合成P450RhF还原酶结构域融合。将所得的融合基因(SEQ ID NO:5)克隆至pCL1920-衍生物(即,SC101复制子,奇霉素抗性标记) 中,使得其转录受到IPTG诱导型Ptrc启动子的控制。所述的质粒命名为 pAS.023(参见表11,如上所述),并转化至大肠杆菌MG1655ΔfadD中, 从而得到菌株sAS.336(参见表12,参见上文)。

[0258] 然后,针对如实施例1和2中所述的(i)十二烷酸向12-羟基十二烷酸、(ii)十二烷酸甲基酯向12-羟基十二烷酸以及(iii)十二烷醇向1,12-十二烷二醇的转化来分析sAS.336菌株。与对照菌株MG1655△fadD(如 上所述)相比,在GC/MS色谱中,在sAS.336中鉴定出所有3种加入的 化合物的新的峰(在BSTFA衍生化后,参见图10A至10C)。如实施例3 所

述,鉴定出12-羟基十二烷酸(如上所述)。在RT 11.946和RT 11.668 分钟(在BSTFA衍生化后)的新的峰分别鉴定为1,12-十二烷二醇和12- 羟基十二烷酸甲基酯。

[0259] 图11A示出了在RT 11.948分钟时的峰的质谱。离子断裂方式表明该 峰为1,12-双(三甲基硅氧基)十二烷,其衍生自1,12-十二烷二醇。图13A 示出1,12-双(三甲基硅氧基)十二烷的离子断裂方式。在m/z=147的离子为 所有二醇相关的化合物的特征碎片,而在m/z=55,69,83,97,111和125 的离子为失去$HOSi(CH_3)_3$部分后特征性烯丙基的断裂碎片。在m/z=241 的离子为用于确定α,ω-二醇化合物的链长的特征性诊断标记,如图13A所 示。该峰的正确鉴定通过与1,12-十二烷二醇(使用BSTFA+1%TMCS衍 生化,图11B)真实标准品的停留时间和质谱比较而进一步被证实。类似 地,可以测定其他α,ω-二醇的链长。例如,如果峰具有在m/z 269的碎片 和在m/z 147,149,111,97,83,69和55的其他特异性离子,则该峰为1,14 双(三甲基硅氧基)十四烷。为了鉴定不饱和的α,ω-二醇,可以使用类似的 规则。由于不饱和的α,ω-二醇比它们的饱和的相对物少2个质量单位,则 1,12-双(三甲基硅氧基)十二烯的特征性诊断碎片为m/z=239。

[0260] 图12A示出为了在RT 11.668分钟时峰的质谱。根据12-三甲基硅氧 基十二烷酸甲基酯的离子断裂方式,所述的峰鉴定为12-三甲基硅氧基十 二烷酸甲基酯。如图13B中所证明的那样,所述的分子下断裂成2种主要 的离子,在m/z=287(M-CH_3)和255(M-CH_3-$HOCH_3$)。此外,质谱显示在 m/z=271(M-OCH_3)的特征性离子,表明该分子具有甲基酯部分。在m/z=103下的离子表明三甲基硅氧基处于末端的位置。此外,在m/z=287和255 下的离子为确定该化合物的链长的特征性离子(参见图13B)。因此,如果 峰具有在m/z=315和283的断裂离子和在m/z 55和103的其他离子,则 所述的峰可以鉴定为14-三甲基硅氧基十四烷酸甲基酯。使用该规则可以 鉴定由菌株stEP677生产的不同链长的ω-羟基脂肪酸甲基酯(参见实施例 6,参见下文)。

[0261] 表13(参见下文)示出了菌株sAS.336在32℃下在18h内转化的ω- 羟基化的脂肪酸衍生物的量。如表13中可见,与所检验的其他酶相比, 由pAS.023表达的杂交cyp153A-RedRhF融合蛋白质有效地将外源脂肪酸 衍生物转化成ω-羟基化的脂肪酸衍生物(如上所述)。因此,选择这种改 造的酶通过可再生的供料,通过改造的生产宿主在体内生产ω-羟基脂肪酸 衍生物(参见实施例6,参见下文)。

[0262] 表13:由sAS.336形成的ω-羟基化的脂肪酸衍生物

[0263]

加入的脂肪酸衍生物(1 g/L)	所形成的ω-羟基化的脂肪酸衍生物(mg/L)*	
十二烷酸	12-十二烷酸	814±13
十二烷酸甲基酯	12-羟基十二烷酸甲基酯	114±10
十二烷醇	1,12-十二烷二醇	148±2

[0264] *一式三份

[0265] 实施例6:由表达CYP153A-Red450RhF杂交融合蛋白质的重组大肠 杆菌菌株由葡萄糖生产ω-羟基化的脂肪酸衍生物

[0266] 本实施例显示通过表达嵌合杂交蛋白质(其中CYP153A P450氧化酶 与还原酶结构域融合)的重组大肠杆菌菌株,由诸如葡萄糖之类的可再生 碳水化合物供料来生产ω-羟基脂肪酸、ω-羟基脂肪酸甲基酯和α,ω-二醇。

[0267] 由pAS.023扩增编码CYP153A(G307A)-RedRhF杂交融合基因的基 因,并将其克隆

至pACYC衍生物载体(p15a复制子,卡那霉素抗性标记) 中,使得融合基因的转录受到IPTG诱导型Ptrc启动子的控制。所得的质 粒命名为pEP.125(参见表11,如上所述)。

[0268] 使用质粒pAS.023或pEP.125分别转化单独地经改造而由碳水化合物 底物(例如葡萄糖)过量生产脂肪酸或脂肪酸衍生物的6种重组大肠杆菌 菌株(参见表14,参见下文),从而创建生产ω-羟基化的脂肪酸衍生物的 菌株(参见表15,参见下文)。此外,过量生产脂肪酸-或脂肪酸衍生物的 菌株还作为对照菌株,从而容易地鉴定在生产ω-羟基化的脂肪酸衍生物的 菌株中的新的化合物。这些过量生产脂肪酸-或脂肪酸衍生物的菌株在本发明中简单地描述,但是不应该解释为是进行限定。按照以下方式创建菌株 ALcV334基因组:删除fadE(酰基-CoA脱氢酶)基因,并过表达硫酯酶tesA 基因的变体。除了在ALcV334中进行基因扩增以外,菌株XL897的基因 组包含以下操作:过表达磷酸泛酰巯基乙胺基转移酶基因和合成的脂肪酸 生物合成操纵子(包括表1中所述的多个基因);并整合多个操纵子,包括羧酸还原酶(carB)变体;硫酯酶(tesA)变体;醇脱氢酶(AlrA);3-酮-酰基 -ACP合酶(fabB)变体;以及转录调节因子(fadR)。按照以下方式操作菌株 DAM1基因组:删除酰基-CoA脱氢酶(fadE)基因,并过表达大肠杆菌硫酯 酶(tesA)和酰基-CoA合酶(fadD)基因。

[0269] 按照以下方式操作菌株stNH1293,KASH286和stNT29的基因组:删 除酰基-CoA脱氢酶(fadE)基因,并过表达转录调节因子(fadR)和合成的脂 肪酸生物合成操纵子。此外,菌株stNH1293包含表达植物硫酯酶、酰基 载体蛋白质(acp)基因和乙酰基-CoA羧化酶(acc)基因复合物的质粒。菌株 KASH286包含表达酯合酶变体的质粒。菌株stNT29包含表达酰基ACP 还原酶(AAR)变体、醇脱氢酶(AlrA)、酰基载体蛋白质(acp)和乙酰基-CoA 羧化酶(acc)基因复合物的质粒。

[0270] 表14:过量生产脂肪酸或脂肪酸衍生物的重组大肠杆菌菌株

[0271]

菌株	表现型
AlcV334	脂肪酸生产者
stNH1293	
LC972	
DAM1	脂肪酸和脂肪酸甲基酯生产者
KASH286	
XL897	脂肪醇生产者
stNT29	
XL959	

[0272] 然后,如实施例1和2所述,针对由诸如葡萄糖之类的可再生的供料 来生产ω-羟基化的脂肪酸衍生物的能力来分析经改造而生产ω-羟基化的 脂肪酸衍生物(如上所述)的菌株。如实施例3和5中所述来识别ω-羟基 化的脂肪酸,脂肪酸甲基酯和脂肪醇。

[0273] 表15:用于由可再生的碳水化合物供料来生产ω-羟基化的脂肪酸衍 生物的、表达cyp153A(G307A)-RedRhF融合蛋白质的重组大肠杆菌菌株

[0274]

菌株	基因型	表现型
stEP675	AlcV334/pAS.023	ω-羟基化的脂肪酸生产者

[0275]

stEP682	stNH1293/pEP.125	
stEP677	DAM1/pAS.023	ω-羟基化的脂肪酸和脂肪酸甲基酯生产者
stEP684	KASH286/pEP.125	
stEP676	XL897/pAS.023	α,ω-二醇醇生产者
stEP685	stNT29/pEP.125	
StEP690	ALKV334/pAS.023 + pEP.127	α,ω-二酸生产者

[0276] 表16:重组大肠杆菌菌株由葡萄糖生产的ω-羟基化的脂肪酸衍生物

[0277]

菌株	由葡萄糖形成的ω-羟基化的脂肪酸衍生物(mg/L)*	
stEP675	ω-羟基化的脂肪酸 (C12,C14,C16, C18)	465 ± 19
stEP682	ω-羟基化的脂肪酸 (C12,C14,C16)	1177 ± 19
stEP677	ω-羟基化的脂肪酸甲基酯 (C12,C14,C16) ω-羟基化的脂肪酸 (C12,C14,C16)	91 ± 4 56 ± 2
stEP684	ω-羟基化的脂肪酸甲基酯 (C14,C16) ω-羟基化的脂肪酸 (C14,C16)	70 ± 8 14 ± 1
stEP676	ω-羟基化的脂肪醇 (C12,C14,C16)	140 ± 2
stEP685	ω-羟基化的脂肪醇 (C14,C16)	201 ± 32
stEP690	α,ω-二酸 (C12, C14, C16, C18)	552 ± 28

[0278] *一式三份

[0279] 与对照菌株(参见表14,如上所述)相比,鉴定表达cyp153A-RedRhF 融合蛋白质(参见表15,如上所述)(对应于ω-羟基脂肪酸、ω-羟基脂肪酸甲基酯和α,ω-二醇)的所有菌株中的新的峰(还参见图14A-14C)。表 16显示这些重组菌株在20h内由葡萄糖生产的ω-羟基脂肪酸衍生物的量。如表16可见,由重组宿主细胞制备的ω-羟基脂肪酸衍生物的量是显著的。具体而言,发现ω-羟基脂肪酸衍生物的生产是非常高效的。值得注意的是 有菌株stEP675生产的ω-羟基脂肪酸几乎唯独在上清液中发现(参见图 15),这表明细胞将它们的产物释放至上清液中。图16显示由菌株stEP675 和stEP682生产的ω-羟基脂肪酸的组成。菌株stEP682生产的最丰富的ω- 羟基脂肪酸为12-羟基十二烷酸(79%)。14-羟基十四烷酸(42%)为stEP675 生产的最丰富的ω-羟基脂肪酸。

[0280] 总之,菌株stEP675生产以下ω-羟基脂肪酸:12-羟基十二碳烯酸、12- 羟基十二烷酸、14-羟基十四碳烯酸、14-羟基十四烷酸、16-羟基十六碳烯酸、16-羟基十六烷酸和18-羟基十八碳烯酸(参见图14A)。菌株stEP682 生产以下ω-羟基脂肪酸:12-羟基十二碳烯酸、12-羟基十二烷酸、14-羟基十四碳烯酸、14-羟基十四烷酸、16-羟基十六碳烯酸和16-羟基十六烷酸。此外,菌株stEP677除了生产ω-羟基脂肪酸以外,还生产以下ω-羟基脂肪酸甲基酯:12-羟基十二烷酸甲基酯、14-羟基十四烷酸甲基酯和16-羟基十六烷酸甲基酯(参见图14B)。此外,菌株stEP684除了生产ω-羟基脂肪酸以外,还生产以下ω-羟基脂肪酸甲基酯:14-羟基十四烷酸甲基酯、16-羟基十六碳烯酸甲基酯和16-羟基十六烷酸甲基酯。菌株stEP676生产以下α,ω-二醇:1,12-十二烯二醇、1,12-十二烷二醇、1,14-十四烷二醇、1,16-十六碳烯二醇和1,16-十六烷二醇(图14C)。菌株stEP685生产以下α,ω-二醇: 1,14-十四烷二醇、1,16-十六碳烯二醇和1,16-十六烷二醇。

[0281] 值得注意的是,本实施例显示经改造而当与CYP153A-RedRhF杂交融 合蛋白质的表达组合时过量生产脂肪酸衍生物的大肠杆菌菌株会有效地 由葡萄糖(作为唯一的碳源)生产ω-羟基化的脂肪酸衍生物。此外,ω- 羟基化的脂肪酸被有效地分泌至发酵液(即,生产细胞或宿主细胞将产物 分泌至发酵液中)中,这是本发明的方法的理想特征。

[0282] 实施例7:重组大肠杆菌菌株由葡萄糖生产α,ω-二酸

[0283] 本实施例证明表达嵌合杂交蛋白质(其中cyp153A P450氧化酶与还原 酶结构域和醇氧化酶和醛脱氢酶融合)的重组大肠杆菌菌株由诸如葡萄糖 之类的可再生的碳水化合物供料生产α,ω-二酸。

[0284] 由恶臭假单胞菌ATCC 29347的基因组DNA来扩增编码醇氧化酶alkJ (登录号CAB54054)(SEQ ID NO:66)和醛脱氢酶alkH(登录号CAB51050) (SEQ ID NO:68)的基因,并将其克隆至pACYC衍生物载体(p15a复制子, 卡那霉素抗性标记)中,使得2个基因形成操纵子,并且该操纵子的转录 受到IPTG诱导型Ptrc启动子的控制。所得的质粒命名为pEP.126(参见表 11,如上所述)。将质粒pEP126与质粒pAS.023(参见表11,如上所述) 共转染至菌株AlcV334(参见表14,如上所述),从而得到菌株sEP690。如实施例1和2中所述,针对由葡萄糖生产α,ω-二酸的能力来分析菌株。与对照菌株AlcV334相比,如图17可见,在sEP690中鉴定多个新的峰。这些峰鉴定为α,ω-二酸和α,ω-双(三甲基甲硅烷基)脂肪酸酯的衍生化形式。α,ω-双(三甲基甲硅烷基)脂肪酸酯的离子断裂方式与α,ω-双(三甲基硅氧基)脂肪醇(其为α,ω-二酸的衍生化的形式)极为相似(参见实施例3),不同 之处在于α,ω-双(三甲基甲硅烷基)脂肪酸酯不生成在m/z=103的离子。因 此,该离子为区分α,ω-双(三甲基硅氧基)脂肪醇与α,ω-双(三甲基甲硅烷基) 脂肪酸酯(以及ω-1,ω-2和ω-3三甲基硅氧基脂肪醇,如下文所述,参见下文)的重要信息。

[0285] 由菌株stEP690生产的主要的化合物为1,12-十二烷二酸、1,14-碳烯酸、 1,14-十四烷二酸、1,16-十六碳烯酸、1,16-十六烷二酸、1,18-十八碳烯二酸 (参见图17)。由该菌株生产的最丰富的化合物为在RT 13.367分钟时的 1,14-十四烷二酸。图18A示出了该化合物在BSTFA衍生化后的质谱。该 化合物的分子离子为m/z=387($M-CH_3$),并且其断裂方式示于图19中。通过与在BSTFA衍生化后的真实标准品1,14-十四烷二酸的质谱和停留时 间比较,而证明1,14-十四烷二酸双(三甲基甲硅烷基)酯的鉴定(图18B)。

[0286] 表16显示菌株sEP690由葡萄糖在20h内生产的α,ω-二酸的量,其为 ~550g/L。该实施例显示经改造用于过量生产脂肪酸衍生物的大肠杆菌菌 株(当与CYP153A-RedRhF杂交融合蛋白质的表达结合时,并且与醇氧化酶(SEQ ID NO:67)和醛脱氢酶(SEQ ID NO:69)的表达结合时)由葡萄糖 (作为唯一的碳源)有效地生产α,ω-二酸。

[0287] 实施例8:由表达cyp102A1(得自巨大芽孢杆菌)的大肠杆菌菌株生 产近末端羟基化的脂肪酸

[0288] 本试验的目的是研究通过使用表达得自巨大芽孢杆菌的细胞色素 P450cyp102A1(P450-BM3)的F87A变体的基因修饰的宿主菌株是否在体 内生产ω-羟基化的脂肪酸或α,ω-二酸。发现,通过使用表达这种cyp102A1 变体的重组大肠杆菌菌株可以生产少量的近末端羟基化的(ω-1,ω-2,ω-3, ω-4,ω-5)脂肪酸。但是,通过使用相同的大肠杆菌菌株不能生产ω-羟基 脂肪酸或α,ω-二酸。

[0289] 通过巨大芽孢杆菌基因组DNA(登录号AAA87602;SEQ ID NO:61) 的交叠PCR来创

建编码得自巨大芽孢杆菌(P450-BM3)的cyp102A1基因变 体(其中位置87处的苯丙氨酸被丙氨酸代替(F87A))的基因。将扩增的 DNA克隆至pCL衍生物(SC101复制子,奇霉素抗性标记)和pACYC衍 生物(p15a复制子,卡那霉素抗性标记)载体中,使得基因的转录受到IPTG诱导型Ptrc启动子的控制。将所得的质粒pSN.012和pSN.009(参见表17, 参见下文)转化至过量生产脂肪酸的菌株AlcV334(参见表14和实施例 16,如上所述)中,从而得到菌株sSN.012和sSN.013(参见表18,参见 下文)。然后,如实施例1和2所述,针对这些菌株由葡萄糖生产ω-羟基 化的脂肪酸的能力来分析这些菌株。与对照菌株AlcV334相比,在2种菌 株中鉴定小的新的峰(图20中sSN.012所示,在BSTFA衍生化之后)。RT 7.195至7.510与8.122至8.414之间的新峰的质谱的断裂方式与ω-羟基 脂肪酸不匹配(参见实施例3,参见上文),但是,它们与近末端ω-1,ω-2, ω-3,ω-4和ω-5位置处的脂肪酸羟基化的预计断裂方式匹配。尽管近末端 羟基化的脂肪酸显示出与ω-羟基脂肪酸相似的离子断裂方式(参见图8A,其为衍生化的12-羟基十二烷酸),这些化合物显示其他的离子碎片,这取 决于衍生化后m/z=117(ω-1),131(ω-2),145(ω-3),159(ω-4)或173(ω-5)的羟基化位点(参见图21A-21E)。关于链长,在m/z=117和345的离子的 组合为衍生化后11ω-1)-羟基十二烷酸的特异性离子。类似地,在m/z=131 和331,m/z=145和317,m/z=159和303,m/z=173和289的组合为ω-2, ω-3,ω-4和ω-5羟基十二烷酸的特异性离子。因此,由RT 7.195至7.510 的峰鉴定为7(ω-5)-,8(ω-4)-,9(ω-3)-,10(ω-2)-和11(ω-1)-羟基十二烷酸,而RT 8.122至8.414的峰鉴定为9(ω-5)-,10(ω-4)-,11(ω-3)-,12(ω-2)-和13(ω-1)-羟基十四酸。此外,针对α,ω-二酸的存在还仔细地检测菌株sSN.012和 sSN.013的GC/MS色谱,但是未检测到这些双功能分子。

[0290] 综上,本实施例显示经改造用于过量生产脂肪酸的大肠杆菌菌株当与 由巨大芽孢杆菌得到的cyp102A1的F87A变体的表达结合时,不生产ω- 羟基化的脂肪酸,但是生产痕量的ω-1,ω-2,ω-3,ω-4和ω-5羟基脂肪酸。已经报告在由巨大芽孢杆菌得到的cyp102A1中,F87A点突变在体外改变 了酶的底物特异性,使得所述的酶几乎专门羟基化十二烷酸或十四烷酸的 ω-位置(参见Oliver et al.(1997)Biochem.36:1567)。此外还报告,由枯草 芽孢杆菌得到的P450-BM3(cyp102A3)和P450-BM3(F87A)(cyp153(F87A))生产14-羟基十四烷酸和1,14,十四烷二酸(例如参见WO 2012/071439)。但是,这尚未证实,并且似乎与本发明相矛盾。此外,关于改变枯草芽孢 杆菌的cyp102A3(cyp102A1的密切的同源物)的底物特异性的近期报告支持本发明提出的数据(例如Lentz et al.(2004)J.Biotechnol.108:41and Lentz et al.(2006)ChemBioChem.7:345)。这样,cyp102A3中等价的苯丙氨酸残基突变成缬氨酸(F88V)主要羟基化脂肪酸的ω-1,ω-2,ω-3和ω-4位置,并且在体外,使用十二烷酸或十六烷酸作为底物未观察到ω-羟基化的的产 物。因此,可以断定上文所述的由巨大芽孢杆菌得到的cyp102A1的变体 以及由枯草芽孢杆菌得到的cyp102A3不适用于生产ω-羟基脂肪酸或α,ω- 二酸(其需要ω-羟基脂肪酸作为中间体)。

[0291] 表17:用于生产近末端羟基化的脂肪酸衍生物而构建的表达质粒

[0292]

质粒	描述
pSN.009	pACYC-cyp102A1(F87A)_巨大芽孢杆菌
pSN.012	pCL-cyp102A1(F87A)_巨大芽孢杆菌
pHM105	pCL-cyp102A7_地衣芽孢杆菌

pSL170.02	pCL-cyp102A7_地衣芽孢杆菌-tesA*-alrA-fabB*-fadR

[0293] 表18:表达cyp102A蛋白质的重组大肠杆菌菌株由可再生的碳水化 合物供料生产近末端羟基化的脂肪酸衍生物

[0294]

菌株	基因型	表现型
sSN.012	AlcV334/pSN.012	生产痕量的近末端羟基化的脂肪酸
sSN.013	AlcV334/pSN.009	
XL960	LC972/pHM105	生产大量的近末端羟基化的脂肪酸
XL961	LC972/pSL170.02	
XL962	XL959/pHM105	生产大量的近末端羟基化的脂肪醇
XL963	XL959/ pSL170.02	

[0295] 实施例9:表达cyp102A7(得自地衣芽孢杆菌)的大肠杆菌菌株由葡 萄糖有效地生产ω-1,ω-2和ω-3-羟基化的脂肪酸

[0296] 本实施例的目的是检验表达细胞色素P450-BM3-型氧化酶cyp102A7 (得自地衣芽孢杆菌)的大肠菌种生产近末端羟基化的脂肪酸。数据证明 表达细胞色素P450-BM3-型氧化酶cyp102A7的大肠杆菌菌株有效地生产ω-1,ω-2和ω-3-羟基脂肪酸和脂肪醇。这是惊人的且意外的。

[0297] 由地衣芽孢杆菌ATCC14580基因组DNA(登录号AAU41718;SEQ ID NO:63)扩增编码由地衣芽孢杆菌得到的P450-BM3-型氧化酶cyp102A7的 yrhJ基因(Dietrich et al.(2008)Appl.Microbiol.Biotechnol.79:931)。将该 基因克隆至pCL1920-衍生物(SC101复制子,奇霉素抗性标记)中,使得 诱导型Ptrc启动子控制其转录。此外,将所述的基因克隆至pCL1920-衍 生物中,使得诱导型Ptrc启动子操纵子,其由(依次)yrhJ,硫酯酶变体(tesA),醇脱氢酶(AlrA),3-酮-酰基-ACP合酶变体(fabB)和转录调节因子 fadR组成。所得的质粒分别命名为pHM105和pSL170.02,并转化至大肠 杆菌菌株LC972和XL959中(参见表14,如上所述)。本发明简单地描述 了这些菌株(同样参见表14)。按照以下方式改造菌株LC972的基因组: 删除酰基-CoA脱氢酶(fadE)基因。磷酸泛酰巯基乙胺基转移酶,2个拷贝的硫酯酶变体(tesA)以及合成的脂肪酸生物合成操纵子(由表1中所述的多 个基因组成)过表达。菌株XL959为LC972,其具有表达羧酸还原酶carB 变体的pACYC衍生物质粒。如实施例1和2所述,针对由葡萄糖生产羟 基化的脂肪酸衍生物的能力来分析4种新的重组大肠杆菌菌株 XL960-XL963(参见表18,如上所述),不同之处在使用35℃的温育温度。 所述的菌株均不能生产任何的ω-羟基化的脂肪酸或脂肪酸衍生物。但是, 如实施例8所述,所有的菌株均生产近末端羟基化的脂肪酸衍生物。图22A 示出了由菌株XL961得到的BFTSA衍生化后的色谱。如实施例8所述, 所述的峰鉴定为以下的ω-1,ω-2和ω-3-羟基脂肪酸:11-羟基十二烷酸(RT=10.563),10-羟基十二烷酸(RT=10.512),9-羟基十二烷酸(RT=10.412), 13-羟基十四烷酸(RT=11.279),12-羟基十四烷酸(RT=11.233),11-羟基十四烷酸(RT=11.133),15-羟基十六烷酸(RT=11.679),14-羟基十六烷酸 (RT=11.623)和13-羟基十六烷酸(RT=11.439)。此外,还检测了最小的峰, 其最可能相当于不饱和的近末端羟基化的脂肪酸。

[0298] 图22B显示由菌株XL963得到的BFTSA衍生化后的色谱。所述的峰 鉴定为以下ω-1,ω-2和ω-3-羟基脂肪醇:在RT 8.303,8.480和8.572分钟 的峰分别为7,8和9-羟基癸

醇,在RT 9.264,9.653和9.905的峰分别为9,10 和11-羟基十二烷醇,在RT 10.889,11.044和11.124的峰分别为11,12和 13-十四烷醇。用于鉴定9,10和11-羟基十二烷醇的详细的质谱和离子断裂 方式示于图23A,23B和23C中。考虑到实施例8中的类似的菌株仅生产 痕量的近末端羟基化的脂肪酸,则令人惊奇且意外的是所有4种菌株均生 产大量的这些产物(高达~1.6g/L)。图24A显示由菌株XL960和XL961 生产的ω-1,ω-2和ω-3-羟基脂肪酸的量,图24B显示由菌株XL962和 XL963生产的ω-1,ω-2和ω-3-羟基脂肪醇的量。因此,本实施例显示经改 造用于当与表达得自地衣芽孢杆菌的cyp102A7结合时过量生产脂肪酸衍生物的大肠杆菌,由葡萄糖(作为唯一的碳源)有效地生产ω-1,ω-2和ω-3- 羟基脂肪酸衍生物。

[0299] 实施例10:重组大肠杆菌菌株由葡萄糖生产α,ω-二酯

[0300] 将酰基-CoA合酶/连接酶或转移酶基因克隆至pCL衍生物载体中编码 cyp153A-RedRhF融合蛋白质的基因的下游,使得这2种基因形成操纵子, 并且该操纵子的转录受到IPTG诱导型Ptrc启动子的控制。合适的cyp153A 和RedRhF融合伴侣的实例在表2A和2D中给出。合适的酰基-CoA合酶/ 连接酶或转移酶的实例在表7中给出。将所得的质粒转化至生产脂肪酸甲 基酯的大肠杆菌菌株中,例如KASH286(参见实施例6和表14,如上所述)。如实施例1和2所述,针对由葡萄糖生产ω-羟基化的脂肪酸衍生物 的能力来分析所述的菌株。预计该菌株会生产α,ω-二酯。

[0301] 实施例11:重组大肠杆菌菌株由葡萄糖生产ω-氨基脂肪酸衍生物

[0302] 将编码cyp153A-RedRhF融合蛋白质的基因、编码醇氧化酶或脱氢酶 的基因、以及编码氨基转移酶或转氨酶的基因克隆至pCL衍生物载体中, 使得这3种基因形成操纵子,并且该操纵子的转录受到IPTG诱导型Ptrc 启动子的控制。合适的cyp153A和RedRhF融合伴侣的实例在表2A和2D 中给出。合适的醇氧化酶或脱氢酶的实例在表3A中给出,而合适的氨基转移酶或转氨酶的实例在表4中给出。将所得的质粒转化至生产脂肪酸的大肠杆菌菌株中,例如stNH1293(参见实施例6和表14,如上所述)。如 实施例1和2所述,针对由葡萄糖生产ω-羟基化的脂肪酸衍生物的能力来 分析所述的菌株。预计该菌株会生产ω-氨基脂肪酸。备选地,将所得的质 粒转化至生产脂肪酸甲基酯的大肠杆菌菌株中,例如KASH286(参见实施 例6和表14,如上所述)。如实施例1和2所述,针对由葡萄糖生产ω-羟基化的脂肪酸衍生物的能力来分析所述的菌株。预计该菌株会生产ω-氨基 脂肪酸甲基酯。

[0303] 实施例12:由多种供料来生产ω-羟基化的脂肪酸衍生物

[0304] 由于自然中脂肪酸的生物合成是普遍的,所以可以在其他有机体中进 一步实施本发明所提供的实施例,其中所述的有机体天然地使用除了碳水 化合物以外的供料。例如这些途径可以在光合作用的微生物有机体中表 达,从而允许由CO_2生产ω-羟基化的脂肪酸衍生物。具体而言,所述的途 径可以在蓝细菌的细胞质中表达,其中当所述的蓝细菌在合适的条件(例 如光生物反应器或开放的池塘中)下生长时,它们生产能够与培养物分离的ω-羟基化的脂肪酸衍生物。备选地,本领域的任一技术人员清楚的是这 些途径可以在使用一氧化碳的有机体中表达,例如得自梭菌属的那些。例 如当这些经改造的微生物有机体在合适的条件(例如,在供入由钢厂废气 或合成气得到的CO的反应器中,其中所述的合成气由有机材料的改进而 衍生得到,例如天然气或生物质)下生长时,它们会生产可以由培养物中回收的ω-羟基化的脂肪酸衍生物。

[0305] 实施例13:表达多种CYP153A-还原酶杂交融合蛋白质的重组大肠杆 菌菌株由葡萄糖生产ω-羟基化的脂肪酸衍生物

[0306] 本实施例显示表达另外的嵌合杂交蛋白质(其中多种CYP153A P450 氧化酶结构域与多种还原酶结构域融合)的重组大肠杆菌菌株由诸如葡萄 糖之类的可再生的碳水化合物生产ω-羟基脂肪酸。

[0307] 在本实验中,由基因组DNA扩增编码由3种微生物得到的CYP153A P450催化蛋白质的基因以及编码由4种微生物得到的还原酶结构域蛋白质 的基因,或者将这些基因合成为密码子优化的DNA(参见下表19)。还原 酶结构域蛋白质包含RedRhF-型以及BM3-型蛋白质。如实施例5所述, 将编码嵌合杂交蛋白质的基因组装并克隆至pCL1920-衍生物载体中(如上 所述)。将所得的质粒转化至菌株AlcV334(参见表14)中。然后,如实 施例1和2所述,针对由诸如葡萄糖之类的可再生的供料生产ω-羟基化的 脂肪酸衍生物的能力来分析经改造而生产ω-羟基化的脂肪酸衍生物的这 些另外的6种菌株。下表20示出了与StEP675相比,所述的这些菌株在 32℃下在18h内生产的ω-羟基化的脂肪酸的量。StEP675为表达嵌合蛋白质的菌株AlcV 334,其中所述的嵌合蛋白质由得自水油海杆菌(M. aquaeolei)的CYP153A和得自红球菌属的菌种NCIMB 9784的还原酶结构域组成(参见实施例6)。如下表20可见,表达CYP153A-还原酶杂交 融合蛋白质的大部分菌株由葡萄糖(作为唯一的碳源)有效地生产ω-羟基化的脂肪酸衍生物。综上,当嵌合CYP153A-还原酶杂交融合蛋白质在大肠杆菌中表达时,由葡萄糖作为唯一的碳源有效地生产ω-羟基化的脂肪酸 衍生物,其中所述的嵌合CYP153A-还原酶杂交融合蛋白质由得自 CYP153A家族的不同成员以及RedRhF-型或BM3-型还原酶家族的不同成 员的蛋白质组成。

[0308] 表19:用于由可再生的碳水化合物供料生产ω-羟基化的脂肪酸衍生 物的、表达杂交cyp153A-还原酶融合蛋白质的另外的重组大肠杆菌菌株

[0309]

	P450 cyp153A 结构域		还原酶结构域		
菌株	来源微生物	登录号#	来源微生物	登录号#	SEQ ID
sIB.012	水油海杆菌	YP_957888	耐金属贪铜菌	YP_587063	48
sIB.013	水油海杆菌	YP_957888	抗辐射不动杆菌	ZP_06072406	50
sIB.027	*Sphingomonas macrogoltabida*	CAH61448	红球菌属的菌种 *NCIMB9784*	AAM67416	52
sIB.028	水油海杆菌	YP_957888	巨大芽胞杆菌	AAA87602	54
sIB.029	水油海杆菌	YP_957888	地衣芽孢杆菌	AAU41718	56
sIB.030	泊库岛食烷菌	WP_011587498	红球菌属的菌种 *NCIMB9784*	AAM67416	58

[0310] 表20:另外的重组大肠杆菌菌株由葡萄糖生产的ω-羟基化的脂肪酸衍生 物

[0311]

菌株	由葡萄糖形成的ω-羟基化的脂肪酸衍生物(mg/L)*	
stEP675	ω-羟基化的脂肪酸 (C12,C14,C16)	626±75
sIB.012	ω-羟基化的脂肪酸 (C12,C14,C16)	751±19
sIB.013	ω-羟基化的脂肪酸 (C12,C14,C16)	451±31
sIB.027	ω-羟基化的脂肪酸 (C12,C14,C16)	38±9

[0312]

sIB.028	ω-羟基化的脂肪酸(C12,C14,C16)	802±155
sIB.029	ω-羟基化的脂肪酸(C12,C14,C16)	825±11
sIB.030	ω-羟基化的脂肪酸(C12,C14,C16)	462±83

[0313] 以下方案和方法适合实施例14至19。

[0314] 方案和方法

[0315] 筛选文库

[0316] 本发明所述的所有方案均依赖于用于使培养物生长的96孔板-深孔板 -2mL系统(Greiner Bio-One,Monroe,NC or Corning,Amsterdam,The Netherlands)以及用于由培养液中提取脂肪酸物质的平板(Costar,Inc.)。下 文提供的方案为发酵条件的实例。可以使用备选方案来评价脂肪酸物质的生产。

[0317] 32℃Plim培养的方案:

[0318] 使用30μL LB培养物(由在96孔板中生长的LB培养物得到)接种 290μL Plim培养基(下表21),然后,将该培养基在32℃摇动温育大约16 小时。使用40μL过夜的种子来接种360μL Plim培养基。在32℃生长2小 时后,使用IPTG(最终浓度为1mM)诱导培养物(下表21)。然后,如 果未另外说明,则将培养物在32℃摇动温育20小时,此后,根据下文所述的标准的提取方案来提取这些培养物。

[0319] 35℃Nlim培养方案:

[0320] 使用40μL LB培养物(由在96孔板中生长的LB培养物得到)接种 360μL LB培养基(下表21),然后,将该培养基在32℃摇动温育大约4 小时。使用40μL LB种子来接种360μL Nlim培养基。在32℃生长2小时 后,使用IPTG(最终浓度为1mM)诱导培养物(下表21)。然后,如果 未另外说明,则将培养物在35℃摇动=下温育20小时,此后,根据下文所述的标准的提取方案来提取这些培养物。

[0321] 表21:培养基的名称和配方

[0322]

培养基名称	配方		
Plim	1	x	5x Plim 盐溶液，包含$(NH4)_2SO_4$
	1	x	1000x 痕量维生素
	1	mg/L	10mg/mL 硫胺

[0323]

培养基名称	配方		
Plim	1	x	5x Plim 盐溶液，包含$(NH4)_2SO_4$
	1	x	1000x 痕量维生素
	1	mg/L	10mg/mL 硫胺

[0324]

培养基名称	配方		
	1	mM	1M MgSO4
	0.1	mM	1M CaCl2
	40	g/L	500g/L 葡萄糖
	1	x	1000x 痕量矿物质
	10	mg/L	10g/L 柠檬酸 Fe
	100	μg/mL	100 mg/ml 奇霉素
	100	mM	2M BisTris (pH7.0)
	0.5	mM	氨基乙酰丙酸
Nlim	1	x	5x 盐溶液，包含 NH4Cl
	1	x	1000x 痕量维生素
	1	mg/L	10mg/mL 硫胺
	1	mM	1M MgSO4
	0.1	mM	1M CaCl2
	40	g/L	500g/L 葡萄糖
	1	x	1000x 痕量矿物质
	10	mg/L	10g/L 柠檬酸 Fe
	100	μg/mL	100 mg/ml 奇霉素
	100	mM	2M BisTris (pH7.0)
	0.5	mM	氨基乙酰丙酸

[0325] 脂肪酸物质的标准提取方案：

[0326] 向各待提取的孔中,加入80μL 1M HCl,然后加入400μL乙酸丁酯(使 用500mg/L十五烷醇作为内部标准品)。然后,使用平板热封仪(ALPS-300 加热器;Abgene,ThermoScientific,Rockford,IL)将96孔板热封,并使用 MIXMATE混合器(Eppendorf,Hamburg,Germany)在2000rpm下摇动15分 钟。在摇动后,将平板以4500rpm在室温(Allegra X-15R,rotor SX4750A, Beckman Coulter,Brea,CA)离心10分钟,从而分离水层和有机层。将100μL 有机层转移至96孔板(聚丙烯,Corning,Amsterdam,The Netherlands)中,并使用100uL BSTFA进行衍生化。随后将平板热封,并储存在-20℃下, 直至按照以下方式使用w-OH FFA方法通过GC-FID进行评价:将1μL样 品注射至Agilent 7890A GC Ultra装置(Agilent,Santa Clara,CA)的分析柱 (DB-1,10m×180μm×0.2μM膜厚,得自JW 121-101A)上,其中所述的 Agilent 7890A GC Ultra装置具有1-20分流的火焰离子化检测仪(FID)。建 立所述的仪器用于检测和定量C_{10}至C_{18}脂肪酸和ω-羟基脂肪酸。上文详 细描述的方案代表了标准的条件,其可以根据需要修改以优化分析结果。

[0327] 建立易错文库

[0328] 使用本领域那些技术人员已知的标准技术来制备易错文库。在一个实 施例中,在有利于引入错配的核苷酸的条件下,由DNA模板通过PCR扩 增在DNA插入物中进行多样性创建时,使用载体中的限制性核酸内切酶 制备载体的骨架。在一种方法中,使用INFUSION

Cloning System(Clontech Laboratories,Inc.,Mountain View,CA),根据制造商的方案来进行载体骨架 和具有多样性的DNA插入物的克隆。

[0329] 建立饱和文库

[0330] 使用本领域那些技术人员已知的标准技术来制备饱和文库。在一个实 施例中,在使用降解引物在DNA插入物中进行多样性创建时,使用载体 中的限制性核酸内切酶制备载体骨架。在一种方法中,使用INFUSION Cloning System(Clontech Laboratories,Inc.,Mountain View,CA),根据制造商的方案来进行载体骨架和具有多样性的DNA插入物的克隆。

[0331] 建立组合文库

[0332] 将识别为有利的突变组合,从而提供ω-OH脂肪酸衍生物物质的生产 得到进一步改良的CYP153-还原酶杂交融合多肽变体(例如 CYP153A-RedRhF杂交蛋白质变体)。使用本领域那些技术人员已知的标准技术来制备组合文库。在一个实施例中,在使用引入了所需突变的引物 在DNA插入物中进行多样性创建时,使用载体中的限制性核酸内切酶制 备载体骨架。如上文所述,在一种方法中,使用INFUSION Cloning System(Clontech Laboratories,Inc.,Mountain View,CA),根据制造商的方案来进行 载体骨架和具有多样性的DNA插入物的克隆。可以使用转移PCR(tPCR) 方案(Erijman et al.(2011) J.Structural Bio.175:171-177)生成组合文库。

[0333] 文库筛选

[0334] 一旦在易错文库、饱和文库或组合文库中生成文库多样性后,便使用 上文所述的一种方法来筛选这些文库。识别2种类型的符合情况(hit):(1) ω-羟基脂肪酸的量(ωOH FFA效价)增加;和/或(2)脂肪酸向ω-羟基脂 肪酸的转化增加。使用本领域那些技术人员常用的标准技术,通过测序来 识别各种符合情况中的杂交cyp153A-RedRhF蛋白质变体的突变。下表23、24和25列出了饱和文库中识别为有利的突变(符合情况)。

[0335] 实施例14:用于文库筛选的菌株和质粒构建

[0336] 本实施例描述了用于饱和或组合诱变文库筛选而构建的菌株和质粒。

[0337] 按照以下方式创建用于编码杂交融合蛋白质(由得自Marinobacter aquaeoli的CYP153A(G307A)P450催化蛋白质和得自红球菌属的菌种 NCIMB9784的P450RhF的含c-末端FMN-和Fe/S的还原酶结构域制备得 到)的基因:由基因组DNA扩增cyp165A(G307A)_Maqu基因,并通过交 叠PCR与密码子优化的合成P450RhF还原酶结构域融合。将所述的融合 基因(SEQ ID NO:5)克隆至pACYC-衍生物(即,p15A复制子,卡那霉素 抗性标记)中,使得其转录受到IPTG诱导型Ptrc启动子的控制。所述的 质粒命名为pEP125(参见表22,参见下文)。此外,由pEP125扩增编码 CYP153A(G307A)-Red450RhF杂交融合蛋白质的基因,并将其克隆至pCL1920-衍生物载体(SC101复制子,奇霉素抗性标记)中,使得其转录 受到IPTG诱导型Ptrc启动子的控制,并且其形成了操纵子,该操纵子具 有编码植物硫酯酶(fatB1)、3-酮-酰基-ACP合酶变体(fabB)和转录调节因子 (fadR)的基因。所述的质粒命名为pLC81(参见表22,参见下文)。

[0338] 按照以下方式创建另外的质粒:将编码植物硫酯酶(fatB1)(得自加州 月桂)的基因合成为密码子优化的DNA,并将其克隆至pCL1920-衍生物 载体(SC101复制子,奇霉素抗性标记)中,使得其转录受到IPTG诱导 型Ptrc启动子的控制,并且其形成了操纵子,该操纵

子具有编码乙酰基 -CoA羧化酶(accDACB)、生物素连接酶(birA)和酰基载体蛋白质的基因。所述的质粒命名为pNH305(参见表22,参见下文)。通过使用密码子优化 的合成的植物硫酯酶(fatA3)(得自拟南芥)代替pNH305中的fatB1,来创 建质粒pAS033(参见表22,参见下文)。通过使用密码子优化的合成的植 物硫酯酶(fatA3)(得自拟南芥)代替pLC81中fatB1,来创建质粒pEP146 (参见表22,参见下文)。此外,pEP146还携带质粒编码的repA蛋白质 中的突变。

[0339] 用于质粒转化的基础菌株为GLP077和BZ128。简言之,按照以下方 式操作基础菌株GLPH077的基因组:删除酰基-CoA脱氢酶(fadE)基因, 并且转录调节因子(fadR)和合成的脂肪酸生物合成操纵子被过表达。简言 之,按照以下方式操作基础菌株BZ128的基因组:删除fadE(酰基-CoA脱 氢酶)基因,并且合成的脂肪酸生物合成操纵子、β-羟基脂肪酰基-ACP脱 水酶(fabZ)和硫酯酶变体(tesA)被过表达。此外,所述的菌株事先经过转座(transposon)以及N-甲基-N′-硝基-N-亚硝基胍(NTG)诱变和筛选。

[0340] 表22:用于文库筛选的质粒

[0341]

质粒	描述
pAS033	pCL-fatA3_Atal-accDCBAbirA_Cglu-acp_Ecol
pEP125	pACYC-cyp153A(G307A)_Maqu-RedRhF_Rhod
pNH305	pCL-fatB1_Ucal-accDCBAbirA_Cglu-acp_Ecol
pLC81	pCL-cyp153A(G307A)_Maqu-RedRhF_Rhod-fatB1_Ucal-fadB_Ecol-fadR_Ecol
pEP146	pCL*-cyp153A(G307A)_Maqu-RedRhF_Rhod-fatA3-Atal-fadB_Ecol-fadR_Ecol

[0342] 检验杂交cyp153A(G307A)-Red450RhF融合蛋白质,以查看宿主细胞 中的表达是否可以生产ω-OH脂肪酸衍生物。表达SEQ ID NO:5的微生物有机体能够由葡萄糖生产超过1g/L的ω-OH脂肪酸衍生物。因此,选择 这种改造的酶以用于进一步的演变研究。

[0343] 实施例15:cyp153A(G307A)-Red450RhF融合蛋白质的P450催化结 构域的饱和文库

[0344] 建立CYP153A-Red450RhF融合蛋白质的P450催化结构域的完整饱和 文库,并针对变体进行筛选,其中所述的变体显示超过 CYP153A(G307A)-Red450RhF(即,模板多肽)的改良。G307A(甘氨酸 残基代替位置307处的丙氨酸)为改良CYP153A的ω-羟化酶活性的有利突变(参见Honda Malca et al.(2012)Chem.Commun.48:5115)。符合情况 的选择标准为(1)ω-羟基脂肪酸的量(ωOH FFA效价)增加;和/或(2) 脂肪酸向ω-羟基脂肪酸的的转化增加。

[0345] 使用本领域那些技术人员已知的标准技术来制备饱和文库。使用质粒 pEP125和pLC81(参见表22,如上所述)来制备完整的饱和文库。筛选3 种饱和文库:对于第一文库,pEP125与pNH305一起转化至菌株GLPH077中;对于第二文库,pLC81转化至BZ128中;而对于第三文库,pEP125 与pAS.033一起转化至GLPH077菌株中。在具体针对ω-羟基十二烷酸的形成得到改良的变体中筛选第一和第二文库,并且在具体针对ω-羟基十六 碳烯酸的形成得到改良的变体中筛选第三文库。使用上文所述的一种标准 的方案来筛选所述的文库。改良的变体示于下表23至25中(参见下文)。 具体而言,多次鉴定位置141处的变体,并发现其为形成ω-羟基十二烷酸 和ω-羟基十六碳烯酸的显著改良的酶。

[0346] 表23:由CYP153A(G307A)-Red450RhF的催化结构域的第一位点饱 和文库得到的

改良变体的概述

[0347]

ω-OH FFA	总 FAS	% ω-OH FFA	FIOC	C12:0 中的% C12:0 ω-OH	FIOC	氨基酸	密码子
1346.3	2236.6	60.2	1.33	83.1	1.08	V141Q	GTG/CAG
1201.1	2149.3	55.9	1.23	84.1	1.10	D134G	GAC/GGG
1106.2	2006.9	55.1	1.22	82	1.07	R40H	AGG/CAC
1007.9	1839.7	54.8	1.21	86.1	1.12	V141I	GTG/ATC
962.5	1791.2	53.7	1.19	81.1	1.06	K41V	AAG/GTG
1228.6	2298.6	53.4	1.18	80.2	1.05	M419V	ATG/GTC
1046.8	1958.5	53.4	1.18	80.1	1.05	V154A	GTG/GCC
990.7	1865.4	53.1	1.17	84.9	1.11	D134G	GAC/GGT
1203.1	2313.1	52	1.15	81.6	1.07	D134G	GAC/GGG
908.7	1773.2	51.2	1.13	80.3	1.05	I11C	ATT/TGC
1020.1	2057	49.6	1.09	81.4	1.06	R205L	CGC/TTG
1256	2688.4	46.7	1.03	72.6	0.95	L304W	CTC/TGG
883.2	**1960.8**	**45.3**	**1.00**	**76.6**	**1.00**		

[0348] FIOC:超过对照的改良的倍数;对照为粗体。

[0349] 表24:由CYP153A(G307A)-Red450RhF的催化结构域的第二位点饱 和文库得到的改良变体的概述

[0350]

突变 1	突变 2	总 ω-OH FFA	总 FAS	% ω-OH FFA	FIOC	C12:0 FAS 中的% C12:0 ω-OH	FIOC
V415R	0	928.10	2880.10	32.23	1.85	33.29	1.96
V415R	0	941.13	2980.97	31.58	1.81	32.98	1.94
V154A	0	694.63	2959.63	23.47	1.35	23.06	1.36
V154A	0	716.00	2963.77	24.16	1.39	23.88	1.40
V154A	0	686.93	2926.97	23.47	1.35	23.40	1.38
V141M	E142Q	717.80	2873.73	24.98	1.44	28.51	1.68
V141I	0	749.07	2971.23	25.21	1.45	31.96	1.88
V141I	0	778.87	2886.77	26.98	1.55	34.27	2.02

[0351]

V141I	0	754.67	2918.90	25.85	1.49	32.85	1.93
V141I	R258Y	672.13	2909.13	23.10	1.33	29.24	1.72
V141I	0	810.23	2912.67	27.83	1.60	35.86	2.11
S233R	0	720.13	2838.00	25.37	1.46	30.82	1.81
S233R	0	746.20	2912.97	25.61	1.47	31.15	1.83
S233N	0	735.57	2905.40	25.33	1.46	25.77	1.52
S233N	0	698.80	2915.17	23.97	1.38	24.40	1.44
S233N	0	732.47	2949.93	24.83	1.43	25.29	1.49
S233N	0	725.97	3018.60	24.05	1.38	24.76	1.46
R82D	E271F	629.03	2914.83	21.58	1.24	20.90	1.23
R6F	R178N	792.33	2845.17	27.85	1.60	28.56	1.68
R6F	V141I	833.13	2871.87	29.01	1.67	36.28	2.13
R27L	0	742.57	2857.53	25.99	1.49	26.10	1.54
R178N	0	701.17	2983.60	23.50	1.35	24.98	1.47
Q129R	0	675.07	2847.37	23.71	1.36	27.97	1.65
Q129R	0	812.23	3044.30	26.68	1.53	31.29	1.84
Q129R	0	660.53	2967.23	22.26	1.28	26.24	1.54
P149R	S157V	684.03	3011.80	22.71	1.31	23.04	1.36
P149R	0	771.40	2959.70	26.06	1.50	26.12	1.54
P149R	0	731.10	2966.13	24.65	1.42	24.75	1.46
P149R	0	757.97	3014.93	25.14	1.45	25.49	1.50
P149R	0	765.90	2963.50	25.84	1.49	26.16	1.54
P149R	0	734.30	2923.70	25.12	1.44	25.50	1.50
P149R	0	745.00	2993.83	24.88	1.43	25.47	1.50
P136T	0	724.53	2980.20	24.31	1.40	24.97	1.47
P136T	0	729.37	3017.67	24.17	1.39	24.90	1.46
P136T	0	678.33	2850.87	23.79	1.37	24.39	1.43
P136C	0	702.27	2947.23	23.83	1.37	25.36	1.49
P136C	0	689.77	3069.63	22.47	1.29	24.01	1.41
N407A	0	731.50	3042.77	24.04	1.38	24.56	1.44
N407A	0	704.47	3015.93	23.36	1.34	23.75	1.40
M228R	0	344.60	2992.27	11.52	0.66	18.33	1.08
L168V	0	793.20	2938.23	27.00	1.55	27.84	1.64
G161P	0	718.33	2938.47	24.45	1.41	24.28	1.43
G161A	0	639.93	2943.40	21.74	1.25	21.65	1.27
G138F	N407A	667.93	2825.43	23.64	1.36	26.09	1.53
F116R	V415R	678.77	2854.97	23.78	1.37	24.14	1.42
E142R	0	663.67	2925.83	22.68	1.30	22.86	1.34
E142R	0	628.03	2930.57	21.43	1.23	21.62	1.27
E142R	0	639.23	2972.03	21.51	1.24	21.86	1.29
D153G	0	787.87	3018.90	26.13	1.50	26.94	1.58
D153G	0	746.20	3039.10	24.55	1.41	25.31	1.49
0	**0**	**543.65**	**3117.75**	**17.44**	**1.00**	**17.04**	**1.00**

[0352] FIOC:超过对照的改良的倍数;对照为粗体。

[0353] 表25:由CYP153A(G307A)-Red450RhF的催化结构域的第三位点饱 和文库得到的改良变体的概述

[0354]

ω-OH		% ω-OH		C16:0 中的% C16:0		C16:1 中的% C16:1			
FFA	总 FAS	FFA	FIOC	ω-OH	FIOC	ω-OH	FIOC	氨基酸	密码子
1298.5	2342.5	55.43	1.53	64.61	1.33	49.02	2.00	N309R	AAC/CGG
1095.9	2374.3	46.16	1.28	58.36	1.20	34.41	1.40	V141G	GTG/GGG
1564	3448.1	45.36	1.25	62.78	1.29	32.88	1.34	L132T	CTC/ACT
1092.9	2391.4	45.70	1.26	60.82	1.25	32.96	1.34	F144R	TTC/AGG
1170.5	2529.6	46.27	1.28	62.41	1.28	31.91	1.30	I131L	ATT/TTG
1232.9	2685.8	45.90	1.27	55.17	1.13	37.63	1.53	G308W	GGC/TGG
931.1	**2570.1**	**36.2**	**1.00**	**48.70**	**1.00**	**24.53**	**1.00**		

[0355] FIOC:超过对照的改良的倍数;对照为粗体。

[0356] 实施例16:CYP153A(G307A)-Red450RhF融合蛋白质的还原酶结构 域的部分位点饱和文库

[0357] 建立杂交CYP153A-Red450RhF融合蛋白质的还原酶结构域的部分饱 和文库(每第10个氨基酸被突变),并针对如下去变体进行筛选,其中所 述变体显示超过CYP153A(V141I,A231T,G307A)-Red450RhF(SEQ ID NO:32)(其为在催化P450CYP153A结构域的位点饱和诱变文库中鉴定的变 体的改良。用于符合情况的选择标准为(1)ω-羟基十二烷酸的量(ωOH FFA 效价)增加;和/或(2)十二烷酸向ω-羟基十二烷酸的转化增加。使用本 领域那些技术人员已知的标准技术来制备饱和文库。对于所述的文库,将容留CYP153A(V141I,A231T,G307A)-Red450RhF的pLC81转化至BZ128 中。使用上文所述的一种标准方案来筛选文库。改良的变体示于下表26 中。具体而言,变体A796V(SEQ ID:42)和P666A为限制改良的酶。

[0358] 表26:由CYP153A(V141I A231T G307A)-Red450RhF的还原酶结构 域的部分饱和文库得到的改良变体的概述

[0359]

RhF 突变	ω-OH FFA	FAS	% ω-OH FFA	FIOC	C12:0 FAS 中的 C12:0 ω-OH	FIOC
P666K	1012.1	2945.5	34.36	1.09	44.08	1.07
P666A	1575.9	2918.7	53.99	1.71	68.35	1.66
T516E	1150.4	2966.2	38.78	1.23	49.01	1.19
V696K	983.4	2955.4	33.27	1.05	43.02	1.05
0	**950.3**	**3004.6**	**31.63**	**1.00**	**41.13**	**1.00**

[0360]

A796V	2458.0	3884.7	63.27	1.81	76.58	1.70
0	**1363.7**	**3905.2**	**34.92**	**1.00**	**44.96**	**1.00**

[0361] FIOC:超过对照的改良的倍数;对照为粗体。

[0362] 实施例17:CYP153A(G307A)-Red450RhF融合蛋白质的还原酶结构 域的组合文库

[0363] 在还原酶结构域的部分饱和文库中所鉴定的有利突变(实施例17)是 进一步改良CYP153A(G307A)-Red450RhF融合蛋白质的组合文库的基础。选择标准为(1)ω-羟基十二烷酸的量(ωOH FFA效价)增加;和/或(2) 十二烷酸向ω-羟基十二烷酸的转化增加。

[0364] 在容留CYP153A(V141I,A231T,G307A)-Red450RhF(SEQ ID:32)的 pLC81中构建组合文库,并将其转化至BZ128中。使用本领域那些技术人 员已知的标准技术来制备组合文库。使用上文所述的一种标准方案来筛选 文库。改良的变体示于下表27中。

[0365] 表27:由CYP153A(V141I,A231T,G307A)-Red450RhF的还原酶结 构域的组合文库

得到的改良变体的概述

[0366]

P450 突变	RhF 突变	ω-OH FFA	FAS	% ω-OH FFA	FIOC	C12:0 FAS 中的% C12:0 ω-OH	FIOC
141I, 231T	T516G, P666M, A796V	851	983	86.8	1.29	88.3	1.23
141I, 231T	T516G, P666H, A796V	1557	2214	69.2	1.03	73.1	1.02
141I, 231T	T516V, P666D, A796V	1491	1999	74.5	1.11	76.9	1.07
141I, 231T	P666M, V696T	916.88	1125	81.4	1.21	82.9	1.15
141I, 231T	A796V	**1528.33**	**2280**	**67.1**	**1.00**	**71.8**	**1.00**

[0367] FIOC:超过对照的改良的倍数;对照为粗体。

[0368] 实施例18:CYP153A(G307A)-Red450RhF融合蛋白质的催化和还原 酶结构域的组合文库

[0369] 在饱和文库中所识别的有利突变(实施例16和17)是进一步改良 CYP153A(G307A)-Red450RhF融合蛋白质的组合文库的基础。选择标准为 (1)ω-羟基十二烷酸的量(ωOH FFA效价)增加;和/或(2)十二烷酸向ω-羟基十二烷酸的转化增加。在pLC81中构建组合文库,并将其转化 至BZ128中。使用本领域那些技术人员已知的标准技术来制备组合文库。使用上文所述的一种标准方案来筛选文库。最好的2种改良变体示于表28 中。

[0370] 表28:由CYP153A(G307A)-Red450RhF的组合文库得到的最佳改良变体

[0371]

突变	SEQ ID	ω-OH FFA*	FAS*	ω-OH FFA	C12:0 FAS 中的 C12:0 ω-OH
R27l, R82D, V141M, R178N, N407A	34	2290.3	3665.1	62.4%	74.1%
R27l, R82D, V141M, R178N, N407A, A796V	44	3499.5	4154.9	84.5%	93.1%

[0372] *48h后的效价(mg/L)

[0373] 实施例19:CYP153A(G307A,A796V)-Red450RhF的位置141和309 的位点饱和诱变

[0374] 注意到位置141处的变化影响底物特异性。因此,在CYP153A(G307A, A796V)-Red450RhF中在这2个位置处实施位点饱和诱变。符合情况的选 择标准为(1)ω-羟基十六碳烯酸的量增加;和/或(2)十六碳烯酸向ω- 羟基十六碳烯酸的转化增加。

[0375] 对于所述的文库,将容留CYP153A(G307A A796V)-Red450RhF(SEQ ID:38)的pEP146转化至BZ128中。使用本领域那些技术人员已知的标准技术来制备位点饱和文库。使用上文所述的一种标准方法来筛选文库。具 体而言,具有V141T的变体(SEQ ID:46)显示出最高的ω-羟基十六碳烯酸 效价以及由十六碳烯酸的最高的转化(图27)。

[0376] 实施例20:表达改良的杂交CYP153A-Red450RhF融合蛋白质的重 组大肠杆菌菌株由葡萄糖高效价地生产ω-羟基化的脂肪酸

[0377] 本实施例显示表达改良的杂交CYP153A-Red450RhF融合蛋白质的重 组大肠杆菌菌株由诸如葡萄糖之类的可再生碳水化合物供料高产率地生 产ω-羟基脂肪酸。

[0378] 将编码变体杂交CYP153A-Red450RhF融合蛋白质的基因(SEQ ID No: 46)克隆至pCL1920-衍生物载体(修饰的SC101复制子,奇霉素抗性标记) 中,使得其转录受到IPTG诱导型Ptrc启动子的控制,并且其形成了具有植物硫酯酶(fatA3)、3-酮-酰基-ACP合酶变体(fabB)和转录调节因子(fadR) 的操纵子。将所述的质粒转化至菌株L439中,从而得到菌株stEP.798。简 言之,基础菌株L439的基因组包含以下操作:删除fadE(酰基-CoA脱氢 酶)基因,并且合成的脂肪酸生物合成操纵子和硫酯酶变体(tesA)被过表达。此外,所述的菌株事先经过转座以及N-甲基-N′-硝基-N-亚硝基胍(NTG)诱 变和筛选。

[0379] 按照以下方式在生物反应器中运行所述的菌株:将所述的菌株的细胞 库小瓶在32℃在包含奇霉素(115mg/L)的LB摇瓶中培养,直至培养物的 OD读数>1。将5%v/v的这种培养物转移至FA种子培养基(2g/L氯化铵、 0.5g/L氯化钠、0.3g/L磷酸二氢钾、1mM硫酸镁、0.1mM氯化钙、20g/L 葡萄糖、1mL/L痕量元素溶液、10mg/L一水柠檬酸铁、100mM bis-tris缓冲液以及115mg奇霉素)中并在32℃过夜培养。然后,使用该种子培 养物接种用于生产的制备型生物反应器。用于该工艺的初始生物反应器培养基包含:0.5g/L氯化铵、1g/L氯化钠、4g/L磷酸二氢钾、2.2g/L七水 硫酸镁、140mg/L二水氯化钙、10mL/L痕量元素溶液、80mg/L一水柠 檬酸铁、0.6mL/L痕量维生素溶液以及5g/L玉米糖浆粉。生物反应器的 灭菌后加入物包含:0.2mM氨基乙酰丙酸、30g/L葡萄糖和115mg/L奇 霉素。

[0380] 在接种之前,稳定生物反应器的参数并且将控制回路打开—溶解氧设 定值:30%;温度设定值:29℃;通风设定值:0.5vvm;pH设定值:6.9。 当培养物的密度高于OD 30时,使用5%v/v种子培养物接种生物反应器, 并使用1mM IPTG诱导。使用DO触发器(当培养基中的葡萄糖被耗尽的 时,所述的DO触发器对控制器作出指示)将复合葡萄糖供料溶液(586g/L 葡萄糖、2.2g/L七水硫酸镁、0.4g/L磷酸二氢钾、80mg/L一水柠檬酸铁 以及10mL/L痕量元素溶液)以最大速率10g/L葡萄糖(根据公称培养物 体积)供入培养物中。在运行的整个过程中对生物反应器取样,并在培养 72小时后收获。

[0381] 图25示出了在30.5℃在72h的整个过程中由菌株stEP.798生产的ω- 羟基化的脂肪酸的量,其在72h后达到高的效价16.0g/L。所生产的ω-羟 基化的脂肪酸由63.1%ω-羟基十六碳烯酸(C16:1)、26.4%ω-羟基十六烷酸 (C16:0)、7.6%ω-羟基十四烷酸(C14:0)、1.9%ω-羟基十四碳烯酸(C14:1) 以及少量的ω-羟基十二碳烯酸(C12:0)和ω-羟基十二碳烯酸(C12:1)(C12低于1%)组成。此外,stEP.798在72h时生产3.0g/L脂肪酸。综上,改良的 杂交CYP153A-Red450RhF融合蛋白质在生产脂肪酸的大肠杆菌菌株中的 表达能够由可再生的碳水化合物供料高效价地生产ω-羟基化的脂肪酸。

[0382] 实施例21:表达改良的杂交CYP153A-Red450RhF融合蛋白质的重 组大肠杆菌菌株由葡萄糖高效价地生产α,ω-二酸

[0383] 本实施例显示表达改良的杂交CYP153A-Red450RhF融合蛋白质以及 异源醇氧化酶(alkJ)和醛脱氢酶(alkH)的重组大肠杆菌菌株由诸如葡萄糖之类的可再生碳水化合物供料高产率地生产α,ω-二酸。

[0384] 将编码变体杂交CYP153A-Red450RhF融合蛋白质的基因(SEQ ID No: 42)克隆至pCL1920-衍生物载体(修饰的SC101复制子,奇霉素抗性标记)中,使得其转录受到IPTG诱导型Ptrc启动子的控制,并且其形成了具有 植物硫酯酶(fatB1)、醇氧化酶(alkJ)、醛脱氢酶(alkH)、3-酮-酰基-ACP合 酶变体(fabB)和转录调节因子(fadR)的操纵子。将所述的质粒转化至菌株 L1012中,从而得到菌株L1017。简言之,基础菌株L1012的基因组包含 以下操作:删除fadE(酰基-CoA脱氢酶)和adhE(醇脱氢酶)基因,并且合 成的脂肪酸生物合成操纵子、β-羟基脂肪酰基-ACP脱水酶(fabZ)和硫酯酶变体(tesA)被过表达。此外,所述的菌株事先经过转座以及N-甲基-N′-硝基 -N-亚硝基胍(NTG)诱变和筛选。

[0385] 按照以下方式在生物反应器中运行所述菌株:将所述菌株的细胞库小 瓶在32℃下在包含奇霉素(115mg/L)的LB摇瓶中培养,直至培养物的OD 读数>1。将2%v/v的这种培养物转移至FA种子培养基(2g/L氯化铵、 0.5g/L氯化钠、0.3g/L磷酸二氢钾、1mM硫酸镁、

0.1mM氯化钙、20g/L 葡萄糖、1mL/L痕量元素溶液、10mg/L一水柠檬酸铁、100mM bis-tris缓冲液以及115mg奇霉素)中并在32℃过夜培养。然后,使用该种子培 养物接种用于生产的制备型生物反应器。用于该工艺的初始生物反应器培养基包含:0.5g/L氯化铵、1g/L氯化钠、4g/L磷酸二氢钾、2.2g/L七水 硫酸镁、140mg/L二水氯化钙、10mL/L痕量元素溶液、80mg/L一水柠 檬酸铁、0.6mL/L痕量维生素溶液以及5g/L玉米糖浆粉。生物反应器的 灭菌后加入物包含:0.2mM氨基乙酰丙酸、30g/L葡萄糖和115mg/L奇 霉素。

[0386] 霉素。

[0387] 在接种之前,稳定生物反应器的参数并且将控制回路打开—溶解氧设 定值:30%;温度设定值:31℃;通风设定值:0.5vvm;pH设定值:6.9。当培养物的密度高于OD 30时,使用5%v/v种子培养物接种生物反应器, 并使用1mM IPTG诱导。使用pH触发器(当培养基中的葡萄糖被耗尽的 时,所述的pH触发器对控制器作出指示)将复合葡萄糖供料溶液(586g/L 葡萄糖、2.2g/L七水硫酸镁、0.4g/L磷酸二氢钾、80mg/L一水柠檬酸铁 以及10mL/L痕量元素溶液)以10g/L团料(根据公称培养物体积)供入 培养物中。在运行的整个过程中对生物反应器取样,并在培养48小时后 收获。

[0388] 图26示出了在30.5℃在48h的整个过程中,由菌株L1017生产的α,ω- 二酸的量。所述的菌株在31h时生产21.2g/Lα,ω-二酸,并且在48h后达 到高的效价23.9g/L。所生产的α,ω-二酸由85.9%α,ω-十二烷酸(C12:0)、4.7%α,ω-十二碳烯酸(C12:1)、5.7%α,ω-十四烷酸(C14:0)、2.9%α,ω-十四 碳烯酸(C14:1)以及少量α,ω-十六碳烯酸(C16:1)(C16低于1%)组成。此外, L1017在48h时生产9.3g/L脂肪酸。仅检测到痕量的ω-羟基脂肪酸。综 上,改良的杂交CYP153A-Red450RhF融合蛋白质在过量生产脂肪酸的大 肠杆菌菌株中的表达以及与醇氧化酶(alkJ)和醛脱氢酶(alkH)的共表达使得能够由可再生的碳水化合物供料高效价地生产α,ω-二酸。

序列表

<110> LS9, INC.
<120> 生产OMEGA-羟基化的脂肪酸衍生物的方法
<130> LS00048PCT
<140>
<141>
<150> 61/835,464
<151> 2013-06-14
<160> 69
<170> PatentIn version 3.5

<210> 1
<211> 1413
<212> DNA
<213> 水油海杆菌
<400> 1

[0001]

```
atgccaacac tgcccagaac atttgacgac attcagtccc gactgattaa cgccacctcc      60
agggtggtgc cgatgcagag gcaaattcag ggactgaaat tcttaatgag cgccaagagg     120
aagaccttcg gcccacgccg accgatgccc gaattcgttg aaacacccat cccggacgtt     180
aacacgctgg cccttgagga catcgatgtc agcaatccgt ttttataccg gcagggtcag     240
tggcgcgcct atttcaaacg gttgcgtgat gaggcgccgg tccattacca gaagaacagc     300
cctttcggcc ccttctggtc ggtaactcgg tttgaagaca tcctgttcgt ggataagagt     360
cacgacctgt tttccgccga gccgcaaatc attctcggtg accctccgga ggggctgtcg     420
gtggaaatgt tcatagcgat ggatccgccg aaacacgatg tgcagcgcag ctcggtgcag     480
ggagtagtgg caccgaaaaa cctgaaggag atggaggggc tgatccgatc acgcaccggc     540
gatgtgcttg acagcctgcc tacagacaaa ccctttaact gggtacctgc tgtttccaag     600
gaactcacag gccgcatgct ggcgacgctt ctggattttc cttacgagga acgccacaag     660
ctggttgagt ggtcggacag aatggcaggt gcagcatcgg ccaccggcgg ggagtttgcc     720
gatgaaaatg ccatgtttga cgacgcggca gacatggccc ggtctttctc caggctttgg     780
cgggacaagg aggcgcgccg cgcagcaggc gaggagcccg gtttcgattt gatcagcctg     840
ttgcagagca acaaagaaac gaaagacctg atcaatcggc cgatggagtt tatcggtaat     900
ttgacgctgc tcatagtcgg cggcaacgat acgacgcgca actcgatgag tggtggcctg     960
gtggccatga acgaattccc cagggaattt gaaaaattga aggcaaaacc ggagttgatt    1020
ccgaacatgg tgtcggaaat catccgctgg caaacgccgc tggcctatat gcgccgaatc    1080
gccaagcagg atgtcgaact gggcggccag accatcaaga agggtgatcg agttgtcatg    1140
tggtacgcgt cgggtaaccg ggacgagcgc aaatttgaca accccgatca gttcatcatt    1200
gatcgcaagg acgcacgaaa ccacatgtcg ttcggctatg gggttcaccg ttgcatgggc    1260
aaccgtctgg ctgaactgca actgcgcatc ctctgggaag aaatactcaa gcgttttgac    1320
aacatcgaag tcgtcgaaga gcccgagcgg gtgcagtcca acttcgtgcg gggctattcc    1380
aggttgatgg tcaaactgac accgaacagt taa                                 1413
```

<210> 2
<211> 470
<212> PRT
<213> 水油海杆菌
<400> 2

```
Met Pro Thr Leu Pro Arg Thr Phe Asp Asp Ile Gln Ser Arg Leu Ile
1               5                  10                  15
Asn Ala Thr Ser Arg Val Val Pro Met Gln Arg Gln Ile Gln Gly Leu
            20                  25                  30
Lys Phe Leu Met Ser Ala Lys Arg Lys Thr Phe Gly Pro Arg Arg Pro
        35                  40                  45
Met Pro Glu Phe Val Glu Thr Pro Ile Pro Asp Val Asn Thr Leu Ala
    50                  55                  60
Leu Glu Asp Ile Asp Val Ser Asn Pro Phe Leu Tyr Arg Gln Gly Gln
65                  70                  75                  80
Trp Arg Ala Tyr Phe Lys Arg Leu Arg Asp Glu Ala Pro Val His Tyr
                85                  90                  95
Gln Lys Asn Ser Pro Phe Gly Pro Phe Trp Ser Val Thr Arg Phe Glu
            100                 105                 110
Asp Ile Leu Phe Val Asp Lys Ser His Asp Leu Phe Ser Ala Glu Pro
        115                 120                 125
Gln Ile Ile Leu Gly Asp Pro Pro Glu Gly Leu Ser Val Glu Met Phe
    130                 135                 140
Ile Ala Met Asp Pro Pro Lys His Asp Val Gln Arg Ser Ser Val Gln
145                 150                 155                 160
Gly Val Val Ala Pro Lys Asn Leu Lys Glu Met Glu Gly Leu Ile Arg
                165                 170                 175
Ser Arg Thr Gly Asp Val Leu Asp Ser Leu Pro Thr Asp Lys Pro Phe
            180                 185                 190
Asn Trp Val Pro Ala Val Ser Lys Glu Leu Thr Gly Arg Met Leu Ala
        195                 200                 205
Thr Leu Leu Asp Phe Pro Tyr Glu Glu Arg His Lys Leu Val Glu Trp
    210                 215                 220
Ser Asp Arg Met Ala Gly Ala Ala Ser Ala Thr Gly Gly Glu Phe Ala
```

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

<210> 3
<211> 1413
<212> DNA
<213> 水油海杆菌
<400> 3

[0002]

```
atgccaacac tgcccagaac atttgacgac attcagtccc gactgattaa cgccacctcc      60
agggtggtgc cgatgcagag gcaaattcag ggactgaaat tcttaatgag cgccaagagg     120
aagaccttcg gcccacgccg accgatgccc gaattcgttg aaacacccat cccggacgtt     180
aacacgctgg cccttgagga catcgatgtc agcaatccgt ttttataccg gcagggtcag     240
tggcgcgcct atttcaaacg gttgcgtgat gaggcgccgg tccattacca gaagaacagc     300
cctttcggcc ccttctggtc ggtaactcgg tttgaagaca tcctgttcgt ggataagagt     360
cacgacctgt tttccgccga gccgcaaatc attctcggtg accctccgga ggggctgtcg     420
gtggaaatgt tcatagcgat ggatccgccg aaacacgatg tgcagcgcag ctcggtgcag     480
ggagtagtgg caccgaaaaa cctgaaggag atggaggggc tgatccgatc acgcaccggc     540
gatgtgcttg acagcctgcc tacagacaaa ccctttaact gggtacctgc tgtttccaag     600
gaactcacag gccgcatgct ggcgacgctt ctggattttc cttacgagga acgccacaag     660
ctggttgagt ggtcggacag aatggcaggt gcagcatcgg ccaccggcgg ggagtttgcc     720
gatgaaaatg ccatgtttga cgacgcggca gacatggccc ggtctttctc caggctttgg     780
cgggacaagg aggcgcgccg cgcagcaggc gaggagcccg gtttcgattt gatcagcctg     840
ttgcagagca acaaagaaac gaaagacctg atcaatcggc cgatggagtt tatcggtaat     900
ttgacgctgc tcatagtcgc cggcaacgat acgacgcgca actcgatgag tggtggcctg     960
gtggccatga acgaattccc cagggaattt gaaaaattga aggcaaaacc ggagttgatt    1020
ccgaacatgg tgtcggaaat catccgctgg caaacgccgc tggcctatat gcgccgaatc    1080
gccaagcagg atgtcgaact gggcggccag accatcaaga agggtgatcg agttgtcatg    1140
tggtacgcgt cgggtaaccg ggacgagcgc aaatttgaca accccgatca gttcatcatt    1200
gatcgcaagg acgcacgaaa ccacatgtcg ttcggctatg gggttcaccg ttgcatgggc    1260
aaccgtctgg ctgaactgca actgcgcatc ctctgggaag aaatactcaa gcgttttgac    1320
aacatcgaag tcgtcgaaga gcccgagcgg gtgcagtcca acttcgtgcg gggctattcc    1380
aggttgatgg tcaaactgac accgaacagt taa                                 1413
```

<210> 4
<211> 470
<212> PRT
<213> 水油海杆菌
<400> 4

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

```
65                  70                  75                  80
Trp Arg Ala Tyr Phe Lys Arg Leu Arg Asp Glu Ala Pro Val His Tyr
                85                  90                  95
Gln Lys Asn Ser Pro Phe Gly Pro Phe Trp Ser Val Thr Arg Phe Glu
            100                 105                 110
Asp Ile Leu Phe Val Asp Lys Ser His Asp Leu Phe Ser Ala Glu Pro
        115                 120                 125
Gln Ile Ile Leu Gly Asp Pro Pro Glu Gly Leu Ser Val Glu Met Phe
    130                 135                 140
Ile Ala Met Asp Pro Pro Lys His Asp Val Gln Arg Ser Ser Val Gln
145                 150                 155                 160
Gly Val Val Ala Pro Lys Asn Leu Lys Glu Met Glu Gly Leu Ile Arg
                165                 170                 175
Ser Arg Thr Gly Asp Val Leu Asp Ser Leu Pro Thr Asp Lys Pro Phe
            180                 185                 190
Asn Trp Val Pro Ala Val Ser Lys Glu Leu Thr Gly Arg Met Leu Ala
        195                 200                 205
Thr Leu Leu Asp Phe Pro Tyr Glu Glu Arg His Lys Leu Val Glu Trp
    210                 215                 220
Ser Asp Arg Met Ala Gly Ala Ala Ser Ala Thr Gly Gly Glu Phe Ala
225                 230                 235                 240
Asp Glu Asn Ala Met Phe Asp Asp Ala Ala Asp Met Ala Arg Ser Phe
                245                 250                 255
Ser Arg Leu Trp Arg Asp Lys Glu Ala Arg Arg Ala Ala Gly Glu Glu
            260                 265                 270
Pro Gly Phe Asp Leu Ile Ser Leu Leu Gln Ser Asn Lys Glu Thr Lys
        275                 280                 285
Asp Leu Ile Asn Arg Pro Met Glu Phe Ile Gly Asn Leu Thr Leu Leu
    290                 295                 300
Ile Val Ala Gly Asn Asp Thr Thr Arg Asn Ser Met Ser Gly Gly Leu
305                 310                 315                 320
Val Ala Met Asn Glu Phe Pro Arg Glu Phe Glu Lys Leu Lys Ala Lys
                325                 330                 335
Pro Glu Leu Ile Pro Asn Met Val Ser Glu Ile Ile Arg Trp Gln Thr
            340                 345                 350
Pro Leu Ala Tyr Met Arg Arg Ile Ala Lys Gln Asp Val Glu Leu Gly
        355                 360                 365
Gly Gln Thr Ile Lys Lys Gly Asp Arg Val Val Met Trp Tyr Ala Ser
    370                 375                 380
Gly Asn Arg Asp Glu Arg Lys Phe Asp Asn Pro Asp Gln Phe Ile Ile
385                 390                 395                 400
Asp Arg Lys Asp Ala Arg Asn His Met Ser Phe Gly Tyr Gly Val His
                405                 410                 415
Arg Cys Met Gly Asn Arg Leu Ala Glu Leu Gln Leu Arg Ile Leu Trp
            420                 425                 430
Glu Glu Ile Leu Lys Arg Phe Asp Asn Ile Glu Val Val Glu Glu Pro
        435                 440                 445
Glu Arg Val Gln Ser Asn Phe Val Arg Gly Tyr Ser Arg Leu Met Val
    450                 455                 460
Lys Leu Thr Pro Asn Ser
465                 470
```

[0003]

<210> 5
<211> 2400
<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> 5

```
atgccaacac tgcccagaac atttgacgac attcagtccc gactgattaa cgccacctcc       60
agggtggtgc cgatgcagag gcaaattcag ggactgaaat tcttaatgag cgccaagagg      120
aagaccttcg gcccacgccg accgatgccc gaattcgttg aaacacccat cccggacgtt      180
aacacgctgg cccttgagga catcgatgtc agcaatccgt ttttataccg gcagggtcag      240
tggcgcgcct atttcaaacg gttgcgtgat gaggcgccgg tccattacca gaagaacagc      300
cctttcggcc ccttctggtc ggtaactcgg tttgaagaca tcctgttcgt ggataagagt      360
cacgacctgt tttccgccga gccgcaaatc attctcggtg accctccgga ggggctgtcg      420
gtggaaatgt tcatagcgat ggatccgccg aaacacgatg tgcagcgcag ctcggtgcag      480
ggagtagtgg caccgaaaaa cctgaaggag atggaggggc tgatccgatc acgcaccggc      540
gatgtgcttg acagcctgcc tacagacaaa ccctttaact gggtacctgc tgtttccaag      600
gaactcacag gccgcatgct ggcgacgctt ctggattttc cttacgagga acgccacaag      660
ctggttgagt ggtcggacag aatggcaggt gcagcatcgg ccaccggcgg ggagtttgcc      720
gatgaaaatg ccatgtttga cgacgcggca gacatggccc ggtctttctc caggctttgg      780
cgggacaagg aggcgcgccg cgcagcaggc gaggagcccg gtttcgattt gatcagcctg      840
ttgcagagca acaaagaaac gaaagacctg atcaatcggc cgatggagtt tatcggtaat      900
ttgacgctgc tcatagtcgc cggcaacgat acgacgcgca actcgatgag tggtggcctg      960
```

gtggccatga acgaattccc cagggaattt gaaaaattga aggcaaaacc ggagttgatt 1020
ccgaacatgg tgtcggaaat catccgctgg caaacgccgc tggcctatat gcgccgaatc 1080
gccaagcagg atgtcgaact gggcggccag accatcaaga agggtgatcg agttgtcatg 1140
tggtacgcgt cgggtaaccg ggacgagcgc aaatttgaca accccgatca gttcatcatt 1200
gatcgcaagg acgcacgaaa ccacatgtcg ttcggctatg gggttcaccg ttgcatgggc 1260
aaccgtctgg ctgaactgca actgcgcatc ctctgggaag aaatactcaa gcgttttgac 1320
aacatcgaag tcgtcgaaga gcccgagcgg gtgcagtcca acttcgtgcg gggctattcc 1380
aggttgatgg tcaaactgac accgaacagt gtactccatc gtcatcaacc tgtcaccatc 1440
ggcgagccgg ccgctcgtgc tgtgagccgc acggtgaccg ttgagcgtct tgatcgcatt 1500
gccgacgatg tccttcgcct ggtccttcgc gatgctggag gtaaaaccct cccgacgtgg 1560
acgcctggcg ctcacatcga cctggatctg ggtgctctga gccgtcagta ttcgctctgc 1620
ggcgctccgg atgctccgtc gtacgaaatc gccgtgcact tagatccgga aagccgtggt 1680
ggaagccgct atattcatga acagctggaa gttggaagtc cgctgcgtat gcgtggccca 1740
cgcaaccatt tcgccctgga tccgggtgcg gaacattacg tgtttgttgc cgggggtatc 1800
ggcatcacgc cggtgctggc aatggcggat catgcccgtg cgcgtggttg gtcgtacgaa 1860
ctgcattatt gtggtcgtaa tcgtagcggt atggcttacc tggaacgcgt cgcgggacat 1920
ggtgaccgcg ctgccttgca cgtatctgaa gaaggcaccc gcattgatct ggcggcatta 1980
cttgctgaac cggcgccggg cgtgcaaatc tacgcctgcg gtccgggccg tttattagcg 2040
ggtcttgaag acgcgtctcg taattggccg gatggcgcgc ttcatgtgga gcatttcact 2100
tcgagtttag ccgctttgga tccggatgtc gaacatgcct ttgatttgga gctgcgtgac 2160
tctggcctta ccgttcgcgt cgagccaact cagaccgttt tagacgcttt gcgtgcgaac 2220
aatatcgacg tcccgtcgga ttgcgaagag gggctgtgtg gttcttgcga agtagccgtt 2280
ctggatggcg aggttgatca ccgtgatacc gttctgacta aggccgagcg cgccgcgaat 2340
cgtcagatga tgacttgctg cagtcgtgca tgcggtgatc gtctggcgct gcgcctctaa 2400

<210> 6
<211> 799
<212> PRT
<213> 人工序列
<220>
<221> 来源
<223> /注="人工序列的描述：合成的多肽"
<400> 6

[0004]

Met Pro Thr Leu Pro Arg Thr Phe Asp Asp Ile Gln Ser Arg Leu Ile
1 5 10 15
Asn Ala Thr Ser Arg Val Val Pro Met Gln Arg Gln Ile Gln Gly Leu
20 25 30
Lys Phe Leu Met Ser Ala Lys Arg Lys Thr Phe Gly Pro Arg Arg Pro
35 40 45
Met Pro Glu Phe Val Glu Thr Pro Ile Pro Asp Val Asn Thr Leu Ala
50 55 60
Leu Glu Asp Ile Asp Val Ser Asn Pro Phe Leu Tyr Arg Gln Gly Gln
65 70 75 80
Trp Arg Ala Tyr Phe Lys Arg Leu Arg Asp Glu Ala Pro Val His Tyr
85 90 95
Gln Lys Asn Ser Pro Phe Gly Pro Phe Trp Ser Val Thr Arg Phe Glu
100 105 110
Asp Ile Leu Phe Val Asp Lys Ser His Asp Leu Phe Ser Ala Glu Pro
115 120 125
Gln Ile Ile Leu Gly Asp Pro Pro Glu Gly Leu Ser Val Glu Met Phe
130 135 140
Ile Ala Met Asp Pro Pro Lys His Asp Val Gln Arg Ser Ser Val Gln
145 150 155 160
Gly Val Val Ala Pro Lys Asn Leu Lys Glu Met Glu Gly Leu Ile Arg
165 170 175
Ser Arg Thr Gly Asp Val Leu Asp Ser Leu Pro Thr Asp Lys Pro Phe
180 185 190

Asn Trp Val Pro Ala Val Ser Lys Glu Leu Thr Gly Arg Met Leu Ala
195 200 205
Thr Leu Leu Asp Phe Pro Tyr Glu Glu Arg His Lys Leu Val Glu Trp
210 215 220
Ser Asp Arg Met Ala Gly Ala Ala Ser Ala Thr Gly Gly Glu Phe Ala
225 230 235 240
Asp Glu Asn Ala Met Phe Asp Asp Ala Ala Asp Met Ala Arg Ser Phe
245 250 255
Ser Arg Leu Trp Arg Asp Lys Glu Ala Arg Arg Ala Ala Gly Glu Glu
260 265 270
Pro Gly Phe Asp Leu Ile Ser Leu Leu Gln Ser Asn Lys Glu Thr Lys
275 280 285
Asp Leu Ile Asn Arg Pro Met Glu Phe Ile Gly Asn Leu Thr Leu Leu
290 295 300
Ile Val Ala Gly Asn Asp Thr Thr Arg Asn Ser Met Ser Gly Gly Leu
305 310 315 320
Val Ala Met Asn Glu Phe Pro Arg Glu Phe Glu Lys Leu Lys Ala Lys
325 330 335

```
Pro Glu Leu Ile Pro Asn Met Val Ser Glu Ile Ile Arg Trp Gln Thr
            340                 345                 350
Pro Leu Ala Tyr Met Arg Arg Ile Ala Lys Gln Asp Val Glu Leu Gly
        355                 360                 365
Gly Gln Thr Ile Lys Lys Gly Asp Arg Val Val Met Trp Tyr Ala Ser
    370                 375                 380
Gly Asn Arg Asp Glu Arg Lys Phe Asp Asn Pro Asp Gln Phe Ile Ile
385                 390                 395                 400
Asp Arg Lys Asp Ala Arg Asn His Met Ser Phe Gly Tyr Gly Val His
                405                 410                 415
Arg Cys Met Gly Asn Arg Leu Ala Glu Leu Gln Leu Arg Ile Leu Trp
            420                 425                 430
Glu Glu Ile Leu Lys Arg Phe Asp Asn Ile Glu Val Val Glu Glu Pro
        435                 440                 445
Glu Arg Val Gln Ser Asn Phe Val Arg Gly Tyr Ser Arg Leu Met Val
    450                 455                 460
Lys Leu Thr Pro Asn Ser Val Leu His Arg His Gln Pro Val Thr Ile
465                 470                 475                 480
Gly Glu Pro Ala Ala Arg Ala Val Ser Arg Thr Val Thr Val Glu Arg
                485                 490                 495
Leu Asp Arg Ile Ala Asp Asp Val Leu Arg Leu Val Leu Arg Asp Ala
            500                 505                 510
Gly Gly Lys Thr Leu Pro Thr Trp Thr Pro Gly Ala His Ile Asp Leu
        515                 520                 525
Asp Leu Gly Ala Leu Ser Arg Gln Tyr Ser Leu Cys Gly Ala Pro Asp
    530                 535                 540
Ala Pro Ser Tyr Glu Ile Ala Val His Leu Asp Pro Glu Ser Arg Gly
545                 550                 555                 560
Gly Ser Arg Tyr Ile His Glu Gln Leu Glu Val Gly Ser Pro Leu Arg
                565                 570                 575
Met Arg Gly Pro Arg Asn His Phe Ala Leu Asp Pro Gly Ala Glu His
            580                 585                 590
Tyr Val Phe Val Ala Gly Gly Ile Gly Ile Thr Pro Val Leu Ala Met
        595                 600                 605
Ala Asp His Ala Arg Ala Arg Gly Trp Ser Tyr Glu Leu His Tyr Cys
    610                 615                 620
Gly Arg Asn Arg Ser Gly Met Ala Tyr Leu Glu Arg Val Ala Gly His
625                 630                 635                 640
Gly Asp Arg Ala Ala Leu His Val Ser Glu Glu Gly Thr Arg Ile Asp
                645                 650                 655
Leu Ala Ala Leu Leu Ala Glu Pro Ala Pro Gly Val Gln Ile Tyr Ala
            660                 665                 670
Cys Gly Pro Gly Arg Leu Leu Ala Gly Leu Glu Asp Ala Ser Arg Asn
        675                 680                 685
Trp Pro Asp Gly Ala Leu His Val Glu His Phe Thr Ser Ser Leu Ala
    690                 695                 700

Ala Leu Asp Pro Asp Val Glu His Ala Phe Asp Leu Glu Leu Arg Asp
705                 710                 715                 720
Ser Gly Leu Thr Val Arg Val Glu Pro Thr Gln Thr Val Leu Asp Ala
                725                 730                 735
Leu Arg Ala Asn Asn Ile Asp Val Pro Ser Asp Cys Glu Glu Gly Leu
            740                 745                 750
Cys Gly Ser Cys Glu Val Ala Val Leu Asp Gly Glu Val Asp His Arg
        755                 760                 765
Asp Thr Val Leu Thr Lys Ala Glu Arg Ala Ala Asn Arg Gln Met Met
    770                 775                 780
Thr Cys Cys Ser Arg Ala Cys Gly Asp Arg Leu Ala Leu Arg Leu
785                 790                 795
```

[0005]

<210> 7
<211> 2400
<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> 7

```
atgccaacac tgcccagaac atttgacgac attcagtccc gactgattaa cgccacctcc      60
agggtggtgc cgatgcagct gcaaattcag ggactgaaat tcttaatgag cgccaagagg     120
aagaccttcg gcccacgccg accgatgccc gaattcgttg aaacacccat cccggacgtt     180
aacacgctgg cccttgagga catcgatgtc agcaatccgt ttttataccg gcagggtcag     240
tggcgcgcct atttcaaacg gttgcgtgat gaggcgccgg tccattacca gaagaacagc     300
cctttcggcc ccttctggtc ggtaactcgg tttgaagaca tcctgttcgt ggataagagt     360
cacgacctgt tttccgccga gccgcaaatc attctcggtg accctccgga ggggctgtcg     420
gtggaaatgt tcatagcgat ggatccgccg aaacacgatg tgcagcgcag ctcggtgcag     480
```

```
ggagtagtgg caccgaaaaa cctgaaggag atggaggggc tgatccgatc acgcaccggc      540
gatgtgcttg acagcctgcc tacagacaaa ccctttaact gggtacctgc tgtttccaag      600
gaactcacag gccgcatgct ggcgacgctt ctggattttc cttacgagga acgccacaag      660
ctggttgagt ggtcggacag aatggcaggt gcagcatcgg ccaccggcgg ggagtttgcc      720
gatgaaaatg ccatgtttga cgacgcggca gacatggccc ggtctttctc caggctttgg      780
cgggacaagg aggcgcgccg cgcagcaggc gaggagcccg gtttcgattt gatcagcctg      840
ttgcagagca acaaagaaac gaaagacctg atcaatcggc cgatggagtt tatcggtaat      900
ttgacgctgc tcatagtcgc cggcaacgat acgacgcgca actcgatgag tggtggcctg      960
gtggccatga acgaattccc cagggaattt gaaaaattga aggcaaaacc ggagttgatt     1020
ccgaacatgg tgtcggaaat catccgctgg caaacgccgc tggcctatat gcgccgaatc     1080
gccaagcagg atgtcgaact gggcggccag accatcaaga agggtgatcg agttgtcatg     1140
tggtacgcgt cgggtaaccg ggacgagcgc aaatttgaca accccgatca gttcatcatt     1200
gatcgcaagg acgcacgaaa ccacatgtcg ttcggctatg gggttcaccg ttgcatgggc     1260
aaccgtctgg ctgaactgca actgcgcatc ctctgggaag aaatactcaa gcgttttgac     1320
aacatcgaag tcgtcgaaga gcccgagcgg gtgcagtcca acttcgtgcg gggctattcc     1380
aggttgatgg tcaaactgac accgaacagt gtactccatc gtcatcaacc tgtcaccatc     1440
ggcgagccgg ccgctcgtgc tgtgagccgc acggtgaccg ttgagcgtct tgatcgcatt     1500
gccgacgatg tccttcgcct ggtccttcgc gatgctggag gtaaaaccct cccgacgtgg     1560
acgcctggcg ctcacatcga cctggatctg ggtgctctga gccgtcagta ttcgctctgc     1620
ggcgctccgg atgctccgtc gtacgaaatc gccgtgcact tagatccgga aagccgtggt     1680
ggaagccgct atattcatga acagctggaa gttggaagtc cgctgcgtat gcgtggccca     1740
cgcaaccatt tcgccctgga tccgggtgcg gaacattacg tgtttgttgc cggggtatc      1800
ggcatcacgc cggtgctggc aatggcggat catgcccgtg cgcgtggttg gtcgtacgaa     1860
ctgcattatt gtggtcgtaa tcgtagcggt atggcttacc tggaacgcgt cgcgggacat     1920
ggtgaccgcg ctgccttgca cgtatctgaa gaaggcaccc gcattgatct ggcggcatta     1980
cttgctgaac cggcgccggg cgtgcaaatc tacgcctgcg gtccgggccg tttattagcg     2040
ggtcttgaag acgcgtctcg taattggccg gatggcgcgc ttcatgtgga gcatttcact     2100
tcgagtttag ccgctttgga tccggatgtc gaacatgcct ttgatttgga gctgcgtgac     2160
tctggcctta ccgttcgcgt cgagccaact cagaccgttt tagacgcttt gcgtgcgaac     2220
aatatcgacg tcccgtcgga ttgcgaagag gggctgtgtg gttcttgcga agtagccgtt     2280
ctggatggcg aggttgatca ccgtgatacc gttctgacta aggccgagcg cgccgcgaat     2340
cgtcagatga tgacttgctg cagtcgtgca tgcggtgatc gtctggcgct gcgcctctaa     2400
```

<210> 8
<211> 799
<212> PRT
<213> 人工序列
<220>

[0006]

<221> 来源
<223> /注="人工序列的描述: 合成的多肽"
<400> 8

```
Met Pro Thr Leu Pro Arg Thr Phe Asp Asp Ile Gln Ser Arg Leu Ile
1               5                  10                  15
Asn Ala Thr Ser Arg Val Val Pro Met Gln Leu Gln Ile Gln Gly Leu
            20                  25                  30
Lys Phe Leu Met Ser Ala Lys Arg Lys Thr Phe Gly Pro Arg Arg Pro
        35                  40                  45
Met Pro Glu Phe Val Glu Thr Pro Ile Pro Asp Val Asn Thr Leu Ala
    50                  55                  60
Leu Glu Asp Ile Asp Val Ser Asn Pro Phe Leu Tyr Arg Gln Gly Gln
65                  70                  75                  80
Trp Arg Ala Tyr Phe Lys Arg Leu Arg Asp Glu Ala Pro Val His Tyr
                85                  90                  95
Gln Lys Asn Ser Pro Phe Gly Pro Phe Trp Ser Val Thr Arg Phe Glu
            100                 105                 110
Asp Ile Leu Phe Val Asp Lys Ser His Asp Leu Phe Ser Ala Glu Pro
        115                 120                 125
Gln Ile Ile Leu Gly Asp Pro Pro Glu Gly Leu Ser Val Glu Met Phe
    130                 135                 140
Ile Ala Met Asp Pro Pro Lys His Asp Val Gln Arg Ser Ser Val Gln
145                 150                 155                 160
Gly Val Val Ala Pro Lys Asn Leu Lys Glu Met Glu Gly Leu Ile Arg
                165                 170                 175
Ser Arg Thr Gly Asp Val Leu Asp Ser Leu Pro Thr Asp Lys Pro Phe
            180                 185                 190
Asn Trp Val Pro Ala Val Ser Lys Glu Leu Thr Gly Arg Met Leu Ala
        195                 200                 205
Thr Leu Leu Asp Phe Pro Tyr Glu Glu Arg His Lys Leu Val Glu Trp
    210                 215                 220
Ser Asp Arg Met Ala Gly Ala Ala Ser Ala Thr Gly Gly Glu Phe Ala
225                 230                 235                 240
Asp Glu Asn Ala Met Phe Asp Asp Ala Ala Asp Met Ala Arg Ser Phe
                245                 250                 255
Ser Arg Leu Trp Arg Asp Lys Glu Ala Arg Arg Ala Ala Gly Glu Glu
            260                 265                 270
Pro Gly Phe Asp Leu Ile Ser Leu Leu Gln Ser Asn Lys Glu Thr Lys
```

```
        275                 280                 285
Asp Leu Ile Asn Arg Pro Met Glu Phe Ile Gly Asn Leu Thr Leu Leu
    290                 295                 300
Ile Val Ala Gly Asn Asp Thr Thr Arg Asn Ser Met Ser Gly Gly Leu
305                 310                 315                 320
Val Ala Met Asn Glu Phe Pro Arg Glu Phe Glu Lys Leu Lys Ala Lys
                325                 330                 335
Pro Glu Leu Ile Pro Asn Met Val Ser Glu Ile Ile Arg Trp Gln Thr
            340                 345                 350
Pro Leu Ala Tyr Met Arg Arg Ile Ala Lys Gln Asp Val Glu Leu Gly
        355                 360                 365
Gly Gln Thr Ile Lys Lys Gly Asp Arg Val Val Met Trp Tyr Ala Ser
    370                 375                 380
Gly Asn Arg Asp Glu Arg Lys Phe Asp Asn Pro Asp Gln Phe Ile Ile
385                 390                 395                 400
Asp Arg Lys Asp Ala Arg Asn His Met Ser Phe Gly Tyr Gly Val His
                405                 410                 415
Arg Cys Met Gly Asn Arg Leu Ala Glu Leu Gln Leu Arg Ile Leu Trp
            420                 425                 430
Glu Glu Ile Leu Lys Arg Phe Asp Asn Ile Glu Val Val Glu Glu Pro
        435                 440                 445
Glu Arg Val Gln Ser Asn Phe Val Arg Gly Tyr Ser Arg Leu Met Val
    450                 455                 460
Lys Leu Thr Pro Asn Ser Val Leu His Arg His Gln Pro Val Thr Ile
465                 470                 475                 480
Gly Glu Pro Ala Ala Arg Ala Val Ser Arg Thr Val Thr Val Glu Arg
                485                 490                 495
Leu Asp Arg Ile Ala Asp Asp Val Leu Arg Leu Val Leu Arg Asp Ala
            500                 505                 510
Gly Gly Lys Thr Leu Pro Thr Trp Thr Pro Gly Ala His Ile Asp Leu
        515                 520                 525
Asp Leu Gly Ala Leu Ser Arg Gln Tyr Ser Leu Cys Gly Ala Pro Asp
    530                 535                 540
Ala Pro Ser Tyr Glu Ile Ala Val His Leu Asp Pro Glu Ser Arg Gly
545                 550                 555                 560
Gly Ser Arg Tyr Ile His Glu Gln Leu Glu Val Gly Ser Pro Leu Arg
                565                 570                 575
Met Arg Gly Pro Arg Asn His Phe Ala Leu Asp Pro Gly Ala Glu His
            580                 585                 590
Tyr Val Phe Val Ala Gly Gly Ile Gly Ile Thr Pro Val Leu Ala Met
        595                 600                 605
Ala Asp His Ala Arg Ala Arg Gly Trp Ser Tyr Glu Leu His Tyr Cys
    610                 615                 620
Gly Arg Asn Arg Ser Gly Met Ala Tyr Leu Glu Arg Val Ala Gly His
625                 630                 635                 640
Gly Asp Arg Ala Ala Leu His Val Ser Glu Glu Gly Thr Arg Ile Asp
                645                 650                 655
Leu Ala Ala Leu Leu Ala Glu Pro Ala Pro Gly Val Gln Ile Tyr Ala
            660                 665                 670
Cys Gly Pro Gly Arg Leu Leu Ala Gly Leu Glu Asp Ala Ser Arg Asn
        675                 680                 685
Trp Pro Asp Gly Ala Leu His Val Glu His Phe Thr Ser Ser Leu Ala
    690                 695                 700
Ala Leu Asp Pro Asp Val Glu His Ala Phe Asp Leu Glu Leu Arg Asp
705                 710                 715                 720
Ser Gly Leu Thr Val Arg Val Glu Pro Thr Gln Thr Val Leu Asp Ala
                725                 730                 735
Leu Arg Ala Asn Asn Ile Asp Val Pro Ser Asp Cys Glu Glu Gly Leu
            740                 745                 750
Cys Gly Ser Cys Glu Val Ala Val Leu Asp Gly Glu Val Asp His Arg
        755                 760                 765
Asp Thr Val Leu Thr Lys Ala Glu Arg Ala Ala Asn Arg Gln Met Met
    770                 775                 780
Thr Cys Cys Ser Arg Ala Cys Gly Asp Arg Leu Ala Leu Arg Leu
785                 790                 795
```

[0007]

<210> 9
<211> 2400
<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> 9

```
atgccaacac tgcccagaac atttgacgac attcagtccc gactgattaa cgccacctcc      60
agggtggtgc cgatgcagag gcaaattcag ggactgaaat tcttaatgag cgccaagagg      120
```

```
aagaccttcg gcccacgccg accgatgccc gaattcgttg aaacacccat cccggacgtt      180
aacacgctgg cccttgagga catcgatgtc agcaatccgt ttttataccg gcagggtcag      240
tgggacgcct atttcaaacg gttgcgtgat gaggcgccgg tccattacca gaagaacagc      300
cctttcggcc ccttctggtc ggtaactcgg tttgaagaca tcctgttcgt ggataagagt      360
cacgacctgt tttccgccga gccgcaaatc attctcggtg accctccgga ggggctgtcg      420
gtggaaatgt tcatagcgat ggatccgccg aaacacgatg tgcagcgcag ctcggtgcag      480
ggagtagtgg caccgaaaaa cctgaaggag atggaggggc tgatccgatc acgcaccggc      540
gatgtgcttg acagcctgcc tacagacaaa ccctttaact gggtacctgc tgtttccaag      600
gaactcacag gccgcatgct ggcgacgctt ctggattttc cttacgagga acgccacaag      660
ctggttgagt ggtcggacag aatggcaggt gcagcatcgg ccaccggcgg ggagtttgcc      720
gatgaaaatg ccatgtttga cgacgcggca gacatggccc ggtctttctc caggctttgg      780
cgggacaagg aggcgcgccg cgcagcaggc gaggagcccg gtttcgattt gatcagcctg      840
ttgcagagca acaaagaaac gaaagacctg atcaatcggc cgatggagtt tatcggtaat      900
ttgacgctgc tcatagtcgc cggcaacgat acgacgcgca actcgatgag tggtggcctg      960
gtggccatga acgaattccc cagggaattt gaaaaattga aggcaaaacc ggagttgatt     1020
ccgaacatgg tgtcggaaat catccgctgg caaacgccgc tggcctatat gcgccgaatc     1080
gccaagcagg atgtcgaact gggcggccag accatcaaga agggtgatcg agttgtcatg     1140
tggtacgcgt cgggtaaccg ggacgagcgc aaatttgaca accccgatca gttcatcatt     1200
gatcgcaagg acgcacgaaa ccacatgtcg ttcggctatg gggttcaccg ttgcatgggc     1260
aaccgtctgg ctgaactgca actgcgcatc ctctgggaag aaatactcaa gcgttttgac     1320
aacatcgaag tcgtcgaaga gcccgagcgg gtgcagtcca acttcgtgcg gggctattcc     1380
aggttgatgg tcaaactgac accgaacagt gtactccatc gtcatcaacc tgtcaccatc     1440
ggcgagccgg ccgctcgtgc tgtgagccgc acggtgaccg ttgagcgtct tgatcgcatt     1500
gccgacgatg tccttcgcct ggtccttcgc gatgctggag gtaaaaccct cccgacgtgg     1560
acgcctggcg ctcacatcga cctggatctg ggtgctctga gccgtcagta ttcgctctgc     1620
ggcgctccgg atgctccgtc gtacgaaatc gccgtgcact tagatccgga aagccgtggt     1680
ggaagccgct atattcatga acagctggaa gttggaagtc cgctgcgtat gcgtggccca     1740
cgcaaccatt tcgccctgga tccgggtgcg gaacattacg tgtttgttgc cgggggtatc     1800
ggcatcacgc cggtgctggc aatggcggat catgcccgtg cgcgtggttg gtcgtacgaa     1860
ctgcattatt gtggtcgtaa tcgtagcggt atggcttacc tggaacgcgt cgcgggacat     1920
ggtgaccgcg ctgccttgca cgtatctgaa gaaggcaccc gcattgatct ggcggcatta     1980
cttgctgaac cggcgccggg cgtgcaaatc tacgcctgcg gtccgggccg tttattagcg     2040
ggtcttgaag acgcgtctcg taattggccg gatggcgcgc ttcatgtgga gcatttcact     2100
tcgagtttag ccgctttgga tccggatgtc gaacatgcct ttgatttgga gctgcgtgac     2160
tctggcctta ccgttcgcgt cgagccaact cagaccgttt tagacgcttt gcgtgcgaac     2220
aatatcgacg tcccgtcgga ttgcgaagag gggctgtgtg gttcttgcga agtagccgtt     2280
ctggatggcg aggttgatca ccgtgatacc gttctgacta aggccgagcg cgccgcgaat     2340
cgtcagatga tgacttgctg cagtcgtgca tgcggtgatc gtctggcgct gcgcctctaa     2400
```

[0008]

<210> 10
<211> 799
<212> PRT
<213> 人工序列
<220>
<221> 来源
<223> /注="人工序列的描述：合成的多肽"
<400> 10

```
Met Pro Thr Leu Pro Arg Thr Phe Asp Asp Ile Gln Ser Arg Leu Ile
1               5                  10                  15
Asn Ala Thr Ser Arg Val Val Pro Met Gln Arg Gln Ile Gln Gly Leu
            20                  25                  30
Lys Phe Leu Met Ser Ala Lys Arg Lys Thr Phe Gly Pro Arg Arg Pro
        35                  40                  45
Met Pro Glu Phe Val Glu Thr Pro Ile Pro Asp Val Asn Thr Leu Ala
    50                  55                  60
Leu Glu Asp Ile Asp Val Ser Asn Pro Phe Leu Tyr Arg Gln Gly Gln
65                  70                  75                  80
Trp Asp Ala Tyr Phe Lys Arg Leu Arg Asp Glu Ala Pro Val His Tyr
                85                  90                  95
Gln Lys Asn Ser Pro Phe Gly Pro Phe Trp Ser Val Thr Arg Phe Glu
            100                 105                 110
Asp Ile Leu Phe Val Asp Lys Ser His Asp Leu Phe Ser Ala Glu Pro
        115                 120                 125
Gln Ile Ile Leu Gly Asp Pro Pro Glu Gly Leu Ser Val Glu Met Phe
    130                 135                 140
Ile Ala Met Asp Pro Pro Lys His Asp Val Gln Arg Ser Ser Val Gln
145                 150                 155                 160
Gly Val Val Ala Pro Lys Asn Leu Lys Glu Met Glu Gly Leu Ile Arg
                165                 170                 175
Ser Arg Thr Gly Asp Val Leu Asp Ser Leu Pro Thr Asp Lys Pro Phe
            180                 185                 190
Asn Trp Val Pro Ala Val Ser Lys Glu Leu Thr Gly Arg Met Leu Ala
        195                 200                 205
Thr Leu Leu Asp Phe Pro Tyr Glu Glu Arg His Lys Leu Val Glu Trp
    210                 215                 220
Ser Asp Arg Met Ala Gly Ala Ala Ser Ala Thr Gly Gly Glu Phe Ala
```

225 230 235 240
Asp Glu Asn Ala Met Phe Asp Asp Ala Ala Asp Met Ala Arg Ser Phe
245 250 255
Ser Arg Leu Trp Arg Asp Lys Glu Ala Arg Arg Ala Ala Gly Glu Glu
260 265 270
Pro Gly Phe Asp Leu Ile Ser Leu Leu Gln Ser Asn Lys Glu Thr Lys
275 280 285
Asp Leu Ile Asn Arg Pro Met Glu Phe Ile Gly Asn Leu Thr Leu Leu
290 295 300
Ile Val Ala Gly Asn Asp Thr Thr Arg Asn Ser Met Ser Gly Gly Leu
305 310 315 320
Val Ala Met Asn Glu Phe Pro Arg Glu Phe Glu Lys Leu Lys Ala Lys
325 330 335
Pro Glu Leu Ile Pro Asn Met Val Ser Glu Ile Ile Arg Trp Gln Thr
340 345 350
Pro Leu Ala Tyr Met Arg Arg Ile Ala Lys Gln Asp Val Glu Leu Gly
355 360 365
Gly Gln Thr Ile Lys Lys Gly Asp Arg Val Val Met Trp Tyr Ala Ser
370 375 380
Gly Asn Arg Asp Glu Arg Lys Phe Asp Asn Pro Asp Gln Phe Ile Ile
385 390 395 400
Asp Arg Lys Asp Ala Arg Asn His Met Ser Phe Gly Tyr Gly Val His
405 410 415
Arg Cys Met Gly Asn Arg Leu Ala Glu Leu Gln Leu Arg Ile Leu Trp
420 425 430
Glu Glu Ile Leu Lys Arg Phe Asp Asn Ile Glu Val Val Glu Glu Pro
435 440 445
Glu Arg Val Gln Ser Asn Phe Val Arg Gly Tyr Ser Arg Leu Met Val
450 455 460
Lys Leu Thr Pro Asn Ser Val Leu His Arg His Gln Pro Val Thr Ile
465 470 475 480
Gly Glu Pro Ala Ala Arg Ala Val Ser Arg Thr Val Thr Val Glu Arg
485 490 495
Leu Asp Arg Ile Ala Asp Asp Val Leu Arg Leu Val Leu Arg Asp Ala
500 505 510
Gly Gly Lys Thr Leu Pro Thr Trp Thr Pro Gly Ala His Ile Asp Leu
515 520 525
[0009] Asp Leu Gly Ala Leu Ser Arg Gln Tyr Ser Leu Cys Gly Ala Pro Asp
530 535 540
Ala Pro Ser Tyr Glu Ile Ala Val His Leu Asp Pro Glu Ser Arg Gly
545 550 555 560
Gly Ser Arg Tyr Ile His Glu Gln Leu Glu Val Gly Ser Pro Leu Arg
565 570 575
Met Arg Gly Pro Arg Asn His Phe Ala Leu Asp Pro Gly Ala Glu His
580 585 590
Tyr Val Phe Val Ala Gly Gly Ile Gly Ile Thr Pro Val Leu Ala Met
595 600 605
Ala Asp His Ala Arg Ala Arg Gly Trp Ser Tyr Glu Leu His Tyr Cys
610 615 620
Gly Arg Asn Arg Ser Gly Met Ala Tyr Leu Glu Arg Val Ala Gly His
625 630 635 640
Gly Asp Arg Ala Ala Leu His Val Ser Glu Glu Gly Thr Arg Ile Asp
645 650 655
Leu Ala Ala Leu Leu Ala Glu Pro Ala Pro Gly Val Gln Ile Tyr Ala
660 665 670
Cys Gly Pro Gly Arg Leu Leu Ala Gly Leu Glu Asp Ala Ser Arg Asn
675 680 685
Trp Pro Asp Gly Ala Leu His Val Glu His Phe Thr Ser Ser Leu Ala
690 695 700
Ala Leu Asp Pro Asp Val Glu His Ala Phe Asp Leu Glu Leu Arg Asp
705 710 715 720
Ser Gly Leu Thr Val Arg Val Glu Pro Thr Gln Thr Val Leu Asp Ala
725 730 735
Leu Arg Ala Asn Asn Ile Asp Val Pro Ser Asp Cys Glu Glu Gly Leu
740 745 750
Cys Gly Ser Cys Glu Val Ala Val Leu Asp Gly Glu Val Asp His Arg
755 760 765
Asp Thr Val Leu Thr Lys Ala Glu Arg Ala Ala Asn Arg Gln Met Met
770 775 780
Thr Cys Cys Ser Arg Ala Cys Gly Asp Arg Leu Ala Leu Arg Leu
785 790 795

<210> 11
<211> 2400
<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> 11
atgccaacac tgcccagaac atttgacgac attcagtccc gactgattaa cgccacctcc 60
agggtggtgc cgatgcagag gcaaattcag ggactgaaat tcttaatgag cgccaagagg 120
aagaccttcg gcccacgccg accgatgccc gaattcgttg aaacacccat cccggacgtt 180
aacacgctgg cccttgagga catcgatgtc agcaatccgt ttttataccg gcagggtcag 240
tggcgcgcct atttcaaacg gttgcgtgat gaggcgccgg tccattacca gaagaacagc 300
cctttcggcc ccttctggtc ggtaactcgg tttgaagaca tcctgttcgt ggataagagt 360
cacgacctgt tttccgccga gccgcaaatc attctcggtg accctccgga ggggctgtcg 420
atcgaaatgt tcatagcgat ggatccgccg aaacacgatg tgcagcgcag ctcggtgcag 480
ggagtagtgg caccgaaaaa cctgaaggag atggaggggc tgatccgatc acgcaccggc 540
gatgtgcttg acagcctgcc tacagacaaa ccctttaact gggtacctgc tgtttccaag 600
gaactcacag gccgcatgct ggcgacgctt ctggattttc cttacgagga acgccacaag 660
ctggttgagt ggtcggacag aatggcaggt gcagcatcgg ccaccggcgg ggagtttgcc 720
gatgaaaatg ccatgtttga cgacgcggca gacatggccc ggtctttctc caggctttgg 780
cgggacaagg aggcgcgccg cgcagcaggc gaggagcccg gtttcgattt gatcagcctg 840
ttgcagagca acaaagaaac gaaagacctg atcaatcggc cgatggagtt tatcggtaat 900
ttgacgctgc tcatagtcgc cggcaacgat acgacgcgca actcgatgag tggtggcctg 960
gtggccatga acgaattccc cagggaattt gaaaaattga aggcaaaacc ggagttgatt 1020
ccgaacatgg tgtcggaaat catccgctgg caaacgccgc tggcctatat gcgccgaatc 1080
gccaagcagg atgtcgaact gggcggccag accatcaaga agggtgatcg agttgtcatg 1140
tggtacgcgt cgggtaaccg ggacgagcgc aaatttgaca accccgatca gttcatcatt 1200
gatcgcaagg acgcacgaaa ccacatgtcg ttcggctatg gggttcaccg ttgcatgggc 1260
aaccgtctgg ctgaactgca actgcgcatc ctctgggaag aaatactcaa gcgttttgac 1320
aacatcgaag tcgtcgaaga gcccgagcgg gtgcagtcca acttcgtgcg gggctattcc 1380
aggttgatgg tcaaactgac accgaacagt gtactccatc gtcatcaacc tgtcaccatc 1440
ggcgagccgg ccgctcgtgc tgtgagccgc acggtgaccg ttgagcgtct tgatcgcatt 1500
gccgacgatg tccttcgcct ggtccttcgc gatgctggag gtaaaaccct cccgacgtgg 1560
acgcctggcg ctcacatcga cctggatctg ggtgctctga gccgtcagta ttcgctctgc 1620
ggcgctccgg atgctccgtc gtacgaaatc gccgtgcact tagatccgga aagccgtggt 1680
ggaagccgct atattcatga acagctggaa gttggaagtc cgctgcgtat gcgtggccca 1740
cgcaaccatt tcgccctgga tccgggtgcg gaacattacg tgtttgttgc cgggggtatc 1800
ggcatcacgc cggtgctggc aatggcggat catgcccgtg cgcgtggttg gtcgtacgaa 1860
ctgcattatt gtggtcgtaa tcgtagcggt atggcttacc tggaacgcgt cgcgggacat 1920
[0010] ggtgaccgcg ctgccttgca cgtatctgaa gaaggcaccc gcattgatct ggcggcatta 1980
cttgctgaac cggcgccggg cgtgcaaatc tacgcctgcg gtccgggccg tttattagcg 2040
ggtcttgaag acgcgtctcg taattggccg gatggcgcgc ttcatgtgga gcatttcact 2100
tcgagtttag ccgctttgga tccggatgtc gaacatgcct ttgatttgga gctgcgtgac 2160
tctggcctta ccgttcgcgt cgagccaact cagaccgttt tagacgcttt gcgtgcgaac 2220
aatatcgacg tcccgtcgga ttgcgaagag gggctgtgtg gttcttgcga agtagccgtt 2280
ctggatggcg aggttgatca ccgtgatacc gttctgacta aggccgagcg cgccgcgaat 2340
cgtcagatga tgacttgctg cagtcgtgca tgcggtgatc gtctggcgct gcgcctctaa 2400

<210> 12
<211> 799
<212> PRT
<213> 人工序列
<220>
<221> 来源
<223> /注="人工序列的描述：合成的多肽"
<400> 12
Met Pro Thr Leu Pro Arg Thr Phe Asp Asp Ile Gln Ser Arg Leu Ile
1 5 10 15
Asn Ala Thr Ser Arg Val Val Pro Met Gln Arg Gln Ile Gln Gly Leu
20 25 30
Lys Phe Leu Met Ser Ala Lys Arg Lys Thr Phe Gly Pro Arg Arg Pro
35 40 45
Met Pro Glu Phe Val Glu Thr Pro Ile Pro Asp Val Asn Thr Leu Ala
50 55 60
Leu Glu Asp Ile Asp Val Ser Asn Pro Phe Leu Tyr Arg Gln Gly Gln
65 70 75 80
Trp Arg Ala Tyr Phe Lys Arg Leu Arg Asp Glu Ala Pro Val His Tyr
85 90 95
Gln Lys Asn Ser Pro Phe Gly Pro Phe Trp Ser Val Thr Arg Phe Glu
100 105 110
Asp Ile Leu Phe Val Asp Lys Ser His Asp Leu Phe Ser Ala Glu Pro
115 120 125
Gln Ile Ile Leu Gly Asp Pro Pro Glu Gly Leu Ser Ile Glu Met Phe
130 135 140
Ile Ala Met Asp Pro Pro Lys His Asp Val Gln Arg Ser Ser Val Gln
145 150 155 160
Gly Val Val Ala Pro Lys Asn Leu Lys Glu Met Glu Gly Leu Ile Arg
165 170 175

[0011]

```
Ser Arg Thr Gly Asp Val Leu Asp Ser Leu Pro Thr Asp Lys Pro Phe
            180                 185                 190
Asn Trp Val Pro Ala Val Ser Lys Glu Leu Thr Gly Arg Met Leu Ala
        195                 200                 205
Thr Leu Leu Asp Phe Pro Tyr Glu Glu Arg His Lys Leu Val Glu Trp
    210                 215                 220
Ser Asp Arg Met Ala Gly Ala Ala Ser Ala Thr Gly Gly Glu Phe Ala
225                 230                 235                 240
Asp Glu Asn Ala Met Phe Asp Asp Ala Ala Asp Met Ala Arg Ser Phe
                245                 250                 255
Ser Arg Leu Trp Arg Asp Lys Glu Ala Arg Arg Ala Ala Gly Glu Glu
            260                 265                 270
Pro Gly Phe Asp Leu Ile Ser Leu Leu Gln Ser Asn Lys Glu Thr Lys
        275                 280                 285
Asp Leu Ile Asn Arg Pro Met Glu Phe Ile Gly Asn Leu Thr Leu Leu
    290                 295                 300
Ile Val Ala Gly Asn Asp Thr Thr Arg Asn Ser Met Ser Gly Gly Leu
305                 310                 315                 320
Val Ala Met Asn Glu Phe Pro Arg Glu Phe Glu Lys Leu Lys Ala Lys
                325                 330                 335
Pro Glu Leu Ile Pro Asn Met Val Ser Glu Ile Ile Arg Trp Gln Thr
            340                 345                 350
Pro Leu Ala Tyr Met Arg Arg Ile Ala Lys Gln Asp Val Glu Leu Gly
        355                 360                 365
Gly Gln Thr Ile Lys Lys Gly Asp Arg Val Val Met Trp Tyr Ala Ser
    370                 375                 380
Gly Asn Arg Asp Glu Arg Lys Phe Asp Asn Pro Asp Gln Phe Ile Ile
385                 390                 395                 400
Asp Arg Lys Asp Ala Arg Asn His Met Ser Phe Gly Tyr Gly Val His
                405                 410                 415
Arg Cys Met Gly Asn Arg Leu Ala Glu Leu Gln Leu Arg Ile Leu Trp
            420                 425                 430
Glu Glu Ile Leu Lys Arg Phe Asp Asn Ile Glu Val Val Glu Glu Pro
        435                 440                 445
Glu Arg Val Gln Ser Asn Phe Val Arg Gly Tyr Ser Arg Leu Met Val
    450                 455                 460
Lys Leu Thr Pro Asn Ser Val Leu His Arg His Gln Pro Val Thr Ile
465                 470                 475                 480
Gly Glu Pro Ala Ala Arg Ala Val Ser Arg Thr Val Thr Val Glu Arg
                485                 490                 495
Leu Asp Arg Ile Ala Asp Asp Val Leu Arg Leu Val Leu Arg Asp Ala
            500                 505                 510
Gly Gly Lys Thr Leu Pro Thr Trp Thr Pro Gly Ala His Ile Asp Leu
        515                 520                 525
Asp Leu Gly Ala Leu Ser Arg Gln Tyr Ser Leu Cys Gly Ala Pro Asp
    530                 535                 540
Ala Pro Ser Tyr Glu Ile Ala Val His Leu Asp Pro Glu Ser Arg Gly
545                 550                 555                 560
Gly Ser Arg Tyr Ile His Glu Gln Leu Glu Val Gly Ser Pro Leu Arg
                565                 570                 575
Met Arg Gly Pro Arg Asn His Phe Ala Leu Asp Pro Gly Ala Glu His
            580                 585                 590
Tyr Val Phe Val Ala Gly Gly Ile Gly Ile Thr Pro Val Leu Ala Met
        595                 600                 605
Ala Asp His Ala Arg Ala Arg Gly Trp Ser Tyr Glu Leu His Tyr Cys
    610                 615                 620
Gly Arg Asn Arg Ser Gly Met Ala Tyr Leu Glu Arg Val Ala Gly His
625                 630                 635                 640
Gly Asp Arg Ala Ala Leu His Val Ser Glu Glu Gly Thr Arg Ile Asp
                645                 650                 655
Leu Ala Ala Leu Leu Ala Glu Pro Ala Pro Gly Val Gln Ile Tyr Ala
            660                 665                 670
Cys Gly Pro Gly Arg Leu Leu Ala Gly Leu Glu Asp Ala Ser Arg Asn
        675                 680                 685
Trp Pro Asp Gly Ala Leu His Val Glu His Phe Thr Ser Ser Leu Ala
    690                 695                 700
Ala Leu Asp Pro Asp Val Glu His Ala Phe Asp Leu Glu Leu Arg Asp
705                 710                 715                 720
Ser Gly Leu Thr Val Arg Val Glu Pro Thr Gln Thr Val Leu Asp Ala
                725                 730                 735
Leu Arg Ala Asn Asn Ile Asp Val Pro Ser Asp Cys Glu Glu Gly Leu
            740                 745                 750
Cys Gly Ser Cys Glu Val Ala Val Leu Asp Gly Glu Val Asp His Arg
        755                 760                 765
Asp Thr Val Leu Thr Lys Ala Glu Arg Ala Ala Asn Arg Gln Met Met
    770                 775                 780
```

Thr Cys Cys Ser Arg Ala Cys Gly Asp Arg Leu Ala Leu Arg Leu
785 790 795

<210> 13
<211> 2400
<212> DNA
<213> 人工序列
<220>
<221> 来源
<223> /注="人工序列的描述：合成的多核苷酸"
<400> 13

[0012]

atgccaacac tgcccagaac atttgacgac attcagtccc gactgattaa cgccacctcc 60
agggtggtgc cgatgcagag gcaaattcag ggactgaaat tcttaatgag cgccaagagg 120
aagaccttcg gcccacgccg accgatgccc gaattcgttg aaacacccat cccggacgtt 180
aacacgctgg cccttgagga catcgatgtc agcaatccgt ttttataccg gcagggtcag 240
tggcgcgcct atttcaaacg gttgcgtgat gaggcgccgg tccattacca gaagaacagc 300
cctttcggcc ccttctggtc ggtaactcgg tttgaagaca tcctgttcgt ggataagagt 360
cacgacctgt tttccgccga gccgcaaatc attctcggtg accctccgga ggggctgtcg 420
caggaaatgt tcatagcgat ggatccgccg aaacacgatg tgcagcgcag ctcggtgcag 480
ggagtagtgg caccgaaaaa cctgaaggag atggaggggc tgatccgatc acgcaccggc 540
gatgtgcttg acagcctgcc tacagacaaa ccctttaact gggtacctgc tgtttccaag 600
gaactcacag gccgcatgct ggcgacgctt ctggattttc cttacgagga acgccacaag 660
ctggttgagt ggtcggacag aatggcaggt gcagcatcgg ccaccggcgg ggagtttgcc 720
gatgaaaatg ccatgtttga cgacgcggca gacatggccc ggtctttctc caggctttgg 780
cgggacaagg aggcgcgccg cgcagcaggc gaggagcccg gtttcgattt gatcagcctg 840
ttgcagagca acaaagaaac gaaagacctg atcaatcggc cgatggagtt tatcggtaat 900
ttgacgctgc tcatagtcgc cggcaacgat acgacgcgca actcgatgag tggtggcctg 960
gtggccatga acgaattccc cagggaattt gaaaaattga aggcaaaacc ggagttgatt 1020
ccgaacatgg tgtcggaaat catccgctgg caaacgccgc tggcctatat gcgccgaatc 1080
gccaagcagg atgtcgaact gggcggccag accatcaaga agggtgatcg agttgtcatg 1140
tggtacgcgt cgggtaaccg ggacgagcgc aaatttgaca accccgatca gttcatcatt 1200
gatcgcaagg acgcacgaaa ccacatgtcg ttcggctatg gggttcaccg ttgcatgggc 1260
aaccgtctgg ctgaactgca actgcgcatc ctctgggaag aaatactcaa gcgttttgac 1320
aacatcgaag tcgtcgaaga gcccgagcgg gtgcagtcca acttcgtgcg gggctattcc 1380
aggttgatgg tcaaactgac accgaacagt gtactccatc gtcatcaacc tgtcaccatc 1440
ggcgagccgg ccgctcgtgc tgtgagccgc acggtgaccg ttgagcgtct tgatcgcatt 1500
gccgacgatg tccttcgcct ggtccttcgc gatgctggag gtaaaaccct cccgacgtgg 1560
acgcctggcg ctcacatcga cctggatctg ggtgctctga gccgtcagta ttcgctctgc 1620
ggcgctccgg atgctccgtc gtacgaaatc gccgtgcact tagatccgga aagccgtggt 1680
ggaagccgct atattcatga acagctggaa gttggaagtc cgctgcgtat gcgtggccca 1740
cgcaaccatt tcgccctgga tccgggtgcg gaacattacg tgtttgttgc cgggggtatc 1800
ggcatcacgc cggtgctggc aatggcggat catgcccgtg cgcgtggttg gtcgtacgaa 1860
ctgcattatt gtggtcgtaa tcgtagcggt atggcttacc tggaacgcgt cgcgggacat 1920
ggtgaccgcg ctgccttgca cgtatctgaa gaaggcaccc gcattgatct ggcggcatta 1980
cttgctgaac cggcgccggg cgtgcaaatc tacgcctgcg gtccgggccg tttattagcg 2040
ggtcttgaag acgcgtctcg taattggccg gatggcgcgc ttcatgtgga gcatttcact 2100
tcgagtttag ccgctttgga tccggatgtc gaacatgcct ttgatttgga gctgcgtgac 2160
tctggcctta ccgttcgcgt cgagccaact cagaccgttt tagacgcttt gcgtgcgaac 2220
aatatcgacg tcccgtcgga ttgcgaagag gggctgtgtg gttcttgcga agtagccgtt 2280
ctggatggcg aggttgatca ccgtgatacc gttctgacta aggccgagcg cgccgcgaat 2340
cgtcagatga tgacttgctg cagtcgtgca tgcggtgatc gtctggcgct gcgcctctaa 2400

<210> 14
<211> 799
<212> PRT
<213> 人工序列
<220>
<221> 来源
<223> /注="人工序列的描述：合成的多肽"
<400> 14

Met Pro Thr Leu Pro Arg Thr Phe Asp Asp Ile Gln Ser Arg Leu Ile
1 5 10 15
Asn Ala Thr Ser Arg Val Val Pro Met Gln Arg Gln Ile Gln Gly Leu
20 25 30
Lys Phe Leu Met Ser Ala Lys Arg Lys Thr Phe Gly Pro Arg Arg Pro
35 40 45
Met Pro Glu Phe Val Glu Thr Pro Ile Pro Asp Val Asn Thr Leu Ala
50 55 60
Leu Glu Asp Ile Asp Val Ser Asn Pro Phe Leu Tyr Arg Gln Gly Gln
65 70 75 80
Trp Arg Ala Tyr Phe Lys Arg Leu Arg Asp Glu Ala Pro Val His Tyr
85 90 95
Gln Lys Asn Ser Pro Phe Gly Pro Phe Trp Ser Val Thr Arg Phe Glu
100 105 110
Asp Ile Leu Phe Val Asp Lys Ser His Asp Leu Phe Ser Ala Glu Pro
115 120 125

[0013]

```
Gln Ile Ile Leu Gly Asp Pro Pro Glu Gly Leu Ser Gln Glu Met Phe
    130                 135                 140
Ile Ala Met Asp Pro Pro Lys His Asp Val Gln Arg Ser Ser Val Gln
145                 150                 155                 160
Gly Val Val Ala Pro Lys Asn Leu Lys Glu Met Glu Gly Leu Ile Arg
                165                 170                 175
Ser Arg Thr Gly Asp Val Leu Asp Ser Leu Pro Thr Asp Lys Pro Phe
            180                 185                 190
Asn Trp Val Pro Ala Val Ser Lys Glu Leu Thr Gly Arg Met Leu Ala
        195                 200                 205
Thr Leu Leu Asp Phe Pro Tyr Glu Glu Arg His Lys Leu Val Glu Trp
    210                 215                 220
Ser Asp Arg Met Ala Gly Ala Ala Ser Ala Thr Gly Gly Glu Phe Ala
225                 230                 235                 240
Asp Glu Asn Ala Met Phe Asp Asp Ala Ala Asp Met Ala Arg Ser Phe
                245                 250                 255
Ser Arg Leu Trp Arg Asp Lys Glu Ala Arg Arg Ala Ala Gly Glu Glu
            260                 265                 270
Pro Gly Phe Asp Leu Ile Ser Leu Leu Gln Ser Asn Lys Glu Thr Lys
        275                 280                 285
Asp Leu Ile Asn Arg Pro Met Glu Phe Ile Gly Asn Leu Thr Leu Leu
    290                 295                 300
Ile Val Ala Gly Asn Asp Thr Thr Arg Asn Ser Met Ser Gly Gly Leu
305                 310                 315                 320
Val Ala Met Asn Glu Phe Pro Arg Glu Phe Glu Lys Leu Lys Ala Lys
                325                 330                 335
Pro Glu Leu Ile Pro Asn Met Val Ser Glu Ile Ile Arg Trp Gln Thr
            340                 345                 350
Pro Leu Ala Tyr Met Arg Arg Ile Ala Lys Gln Asp Val Glu Leu Gly
        355                 360                 365
Gly Gln Thr Ile Lys Lys Gly Asp Arg Val Val Met Trp Tyr Ala Ser
    370                 375                 380
Gly Asn Arg Asp Glu Arg Lys Phe Asp Asn Pro Asp Gln Phe Ile Ile
385                 390                 395                 400
Asp Arg Lys Asp Ala Arg Asn His Met Ser Phe Gly Tyr Gly Val His
                405                 410                 415
Arg Cys Met Gly Asn Arg Leu Ala Glu Leu Gln Leu Arg Ile Leu Trp
            420                 425                 430
Glu Glu Ile Leu Lys Arg Phe Asp Asn Ile Glu Val Val Glu Glu Pro
        435                 440                 445
Glu Arg Val Gln Ser Asn Phe Val Arg Gly Tyr Ser Arg Leu Met Val
    450                 455                 460
Lys Leu Thr Pro Asn Ser Val Leu His Arg His Gln Pro Val Thr Ile
465                 470                 475                 480
Gly Glu Pro Ala Ala Arg Ala Val Ser Arg Thr Val Thr Val Glu Arg
                485                 490                 495
Leu Asp Arg Ile Ala Asp Asp Val Leu Arg Leu Val Leu Arg Asp Ala
            500                 505                 510
Gly Gly Lys Thr Leu Pro Thr Trp Thr Pro Gly Ala His Ile Asp Leu
        515                 520                 525
Asp Leu Gly Ala Leu Ser Arg Gln Tyr Ser Leu Cys Gly Ala Pro Asp
    530                 535                 540
Ala Pro Ser Tyr Glu Ile Ala Val His Leu Asp Pro Glu Ser Arg Gly
545                 550                 555                 560
Gly Ser Arg Tyr Ile His Glu Gln Leu Glu Val Gly Ser Pro Leu Arg
                565                 570                 575
Met Arg Gly Pro Arg Asn His Phe Ala Leu Asp Pro Gly Ala Glu His
            580                 585                 590
Tyr Val Phe Val Ala Gly Gly Ile Gly Ile Thr Pro Val Leu Ala Met
        595                 600                 605
Ala Asp His Ala Arg Ala Arg Gly Trp Ser Tyr Glu Leu His Tyr Cys
    610                 615                 620
Gly Arg Asn Arg Ser Gly Met Ala Tyr Leu Glu Arg Val Ala Gly His
625                 630                 635                 640
Gly Asp Arg Ala Ala Leu His Val Ser Glu Glu Gly Thr Arg Ile Asp
                645                 650                 655
Leu Ala Ala Leu Leu Ala Glu Pro Ala Pro Gly Val Gln Ile Tyr Ala
            660                 665                 670
Cys Gly Pro Gly Arg Leu Leu Ala Gly Leu Glu Asp Ala Ser Arg Asn
        675                 680                 685
Trp Pro Asp Gly Ala Leu His Val Glu His Phe Thr Ser Ser Leu Ala
    690                 695                 700
Ala Leu Asp Pro Asp Val Glu His Ala Phe Asp Leu Glu Leu Arg Asp
705                 710                 715                 720
Ser Gly Leu Thr Val Arg Val Glu Pro Thr Gln Thr Val Leu Asp Ala
                725                 730                 735
```

```
Leu Arg Ala Asn Asn Ile Asp Val Pro Ser Asp Cys Glu Glu Gly Leu
            740                 745                 750
Cys Gly Ser Cys Glu Val Ala Val Leu Asp Gly Glu Val Asp His Arg
        755                 760                 765
Asp Thr Val Leu Thr Lys Ala Glu Arg Ala Ala Asn Arg Gln Met Met
    770                 775                 780
Thr Cys Cys Ser Arg Ala Cys Gly Asp Arg Leu Ala Leu Arg Leu
785                 790                 795

<210> 15
<211> 2400
<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> 15
atgccaacac tgcccagaac atttgacgac attcagtccc gactgattaa cgccacctcc      60
agggtggtgc cgatgcagag gcaaattcag ggactgaaat tcttaatgag cgccaagagg     120
aagaccttcg gcccacgccg accgatgccc gaattcgttg aaacacccat cccggacgtt     180
aacacgctgg cccttgagga catcgatgtc agcaatccgt ttttataccg gcagggtcag     240
tggcgcgcct atttcaaacg gttgcgtgat gaggcgccgg tccattacca gaagaacagc     300
cctttcggcc ccttctggtc ggtaactcgg tttgaagaca tcctgttcgt ggataagagt     360
cacgacctgt tttccgccga gccgcaaatc attctcggtg accctccgga ggggctgtcg     420
ggggaaatgt tcatagcgat ggatccgccg aaacacgatg tgcagcgcag ctcggtgcag     480
ggagtagtgg caccgaaaaa cctgaaggag atggaggggc tgatccgatc acgcaccggc     540
gatgtgcttg acagcctgcc tacagacaaa ccctttaact gggtacctgc tgtttccaag     600
gaactcacag gccgcatgct ggcgacgctt ctggattttc cttacgagga acgccacaag     660
ctggttgagt ggtcggacag aatggcaggt gcagcatcgg ccaccggcgg ggagtttgcc     720
gatgaaaatg ccatgtttga cgacgcggca gacatggccc ggtctttctc caggctttgg     780
cgggacaagg aggcgcgccg cgcagcaggc gaggagcccg gtttcgattt gatcagcctg     840
ttgcagagca acaaagaaac gaaagacctg atcaatcggc cgatggagtt tatcggtaat     900
ttgacgctgc tcatagtcgc cggcaacgat acgacgcgca actcgatgag tggtggcctg     960
gtggccatga acgaattccc cagggaattt gaaaaattga aggcaaaacc ggagttgatt    1020
ccgaacatgg tgtcggaaat catccgctgg caaacgccgc tggcctatat gcgccgaatc    1080
gccaagcagg atgtcgaact gggcggccag accatcaaga agggtgatcg agttgtcatg    1140
tggtacgcgt cgggtaaccg ggacgagcgc aaatttgaca accccgatca gttcatcatt    1200
gatcgcaagg acgcacgaaa ccacatgtcg ttcggctatg gggttcaccg ttgcatgggc    1260
aaccgtctgg ctgaactgca actgcgcatc ctctgggaag aaatactcaa gcgttttgac    1320
aacatcgaag tcgtcgaaga gcccgagcgg gtgcagtcca acttcgtgcg gggctattcc    1380
aggttgatgg tcaaactgac accgaacagt gtactccatc gtcatcaacc tgtcaccatc    1440
ggcgagccgg ccgctcgtgc tgtgagccgc acggtgaccg ttgagcgtct tgatcgcatt    1500
gccgacgatg tccttcgcct ggtccttcgc gatgctggag gtaaaaccct cccgacgtgg    1560
acgcctggcg ctcacatcga cctggatctg ggtgctctga gccgtcagta ttcgctctgc    1620
ggcgctccgg atgctccgtc gtacgaaatc gccgtgcact tagatccgga aagccgtggt    1680
ggaagccgct atattcatga acagctggaa gttggaagtc cgctgcgtat gcgtggccca    1740
cgcaaccatt tcgccctgga tccgggtgcg gaacattacg tgtttgttgc cgggggtatc    1800
ggcatcacgc cggtgctggc aatggcggat catgcccgtg cgcgtggttg gtcgtacgaa    1860
ctgcattatt gtggtcgtaa tcgtagcggt atggcttacc tggaacgcgt cgcgggacat    1920
ggtgaccgcg ctgccttgca cgtatctgaa gaaggcaccc gcattgatct ggcggcatta    1980
cttgctgaac cggcgccggg cgtgcaaatc tacgcctgcg gtccgggccg tttattagcg    2040
ggtcttgaag acgcgtctcg taattggccg gatggcgcgc ttcatgtgga gcatttcact    2100
tcgagtttag ccgctttgga tccggatgtc gaacatgcct ttgatttgga gctgcgtgac    2160
tctggcctta ccgttcgcgt cgagccaact cagaccgttt tagacgcttt gcgtgcgaac    2220
aatatcgacg tcccgtcgga ttgcgaagag gggctgtgtg gttcttgcga agtagccgtt    2280
ctggatggcg aggttgatca ccgtgatacc gttctgacta aggccgagcg cgccgcgaat    2340
cgtcagatga tgacttgctg cagtcgtgca tgcggtgatc gtctggcgct gcgcctctaa    2400

<210> 16
<211> 799
<212> PRT
<213> 人工序列
<220>
<221> 来源
<223> /注="人工序列的描述：合成的多肽"
<400> 16
Met Pro Thr Leu Pro Arg Thr Phe Asp Asp Ile Gln Ser Arg Leu Ile
1               5                  10                  15
Asn Ala Thr Ser Arg Val Val Pro Met Gln Arg Gln Ile Gln Gly Leu
            20                  25                  30
Lys Phe Leu Met Ser Ala Lys Arg Lys Thr Phe Gly Pro Arg Arg Pro
        35                  40                  45
Met Pro Glu Phe Val Glu Thr Pro Ile Pro Asp Val Asn Thr Leu Ala
    50                  55                  60
Leu Glu Asp Ile Asp Val Ser Asn Pro Phe Leu Tyr Arg Gln Gly Gln
65                  70                  75                  80
```

[0014]

[0015]

```
Trp Arg Ala Tyr Phe Lys Arg Leu Arg Asp Glu Ala Pro Val His Tyr
                85                  90                  95
Gln Lys Asn Ser Pro Phe Gly Pro Phe Trp Ser Val Thr Arg Phe Glu
            100                 105                 110
Asp Ile Leu Phe Val Asp Lys Ser His Asp Leu Phe Ser Ala Glu Pro
        115                 120                 125
Gln Ile Ile Leu Gly Asp Pro Pro Glu Gly Leu Ser Gly Glu Met Phe
    130                 135                 140
Ile Ala Met Asp Pro Pro Lys His Asp Val Gln Arg Ser Ser Val Gln
145                 150                 155                 160
Gly Val Val Ala Pro Lys Asn Leu Lys Glu Met Glu Gly Leu Ile Arg
                165                 170                 175
Ser Arg Thr Gly Asp Val Leu Asp Ser Leu Pro Thr Asp Lys Pro Phe
            180                 185                 190
Asn Trp Val Pro Ala Val Ser Lys Glu Leu Thr Gly Arg Met Leu Ala
        195                 200                 205
Thr Leu Leu Asp Phe Pro Tyr Glu Glu Arg His Lys Leu Val Glu Trp
    210                 215                 220
Ser Asp Arg Met Ala Gly Ala Ala Ser Ala Thr Gly Gly Glu Phe Ala
225                 230                 235                 240
Asp Glu Asn Ala Met Phe Asp Asp Ala Ala Asp Met Ala Arg Ser Phe
                245                 250                 255
Ser Arg Leu Trp Arg Asp Lys Glu Ala Arg Arg Ala Ala Gly Glu Glu
            260                 265                 270
Pro Gly Phe Asp Leu Ile Ser Leu Leu Gln Ser Asn Lys Glu Thr Lys
        275                 280                 285
Asp Leu Ile Asn Arg Pro Met Glu Phe Ile Gly Asn Leu Thr Leu Leu
    290                 295                 300
Ile Val Ala Gly Asn Asp Thr Thr Arg Asn Ser Met Ser Gly Gly Leu
305                 310                 315                 320
Val Ala Met Asn Glu Phe Pro Arg Glu Phe Glu Lys Leu Lys Ala Lys
                325                 330                 335
Pro Glu Leu Ile Pro Asn Met Val Ser Glu Ile Ile Arg Trp Gln Thr
            340                 345                 350
Pro Leu Ala Tyr Met Arg Arg Ile Ala Lys Gln Asp Val Glu Leu Gly
        355                 360                 365
Gly Gln Thr Ile Lys Lys Gly Asp Arg Val Val Met Trp Tyr Ala Ser
    370                 375                 380
Gly Asn Arg Asp Glu Arg Lys Phe Asp Asn Pro Asp Gln Phe Ile Ile
385                 390                 395                 400
Asp Arg Lys Asp Ala Arg Asn His Met Ser Phe Gly Tyr Gly Val His
                405                 410                 415
Arg Cys Met Gly Asn Arg Leu Ala Glu Leu Gln Leu Arg Ile Leu Trp
            420                 425                 430
Glu Glu Ile Leu Lys Arg Phe Asp Asn Ile Glu Val Val Glu Glu Pro
        435                 440                 445
Glu Arg Val Gln Ser Asn Phe Val Arg Gly Tyr Ser Arg Leu Met Val
    450                 455                 460
Lys Leu Thr Pro Asn Ser Val Leu His Arg His Gln Pro Val Thr Ile
465                 470                 475                 480
Gly Glu Pro Ala Ala Arg Ala Val Ser Arg Thr Val Thr Val Glu Arg
                485                 490                 495
Leu Asp Arg Ile Ala Asp Asp Val Leu Arg Leu Val Leu Arg Asp Ala
            500                 505                 510
Gly Gly Lys Thr Leu Pro Thr Trp Thr Pro Gly Ala His Ile Asp Leu
        515                 520                 525
Asp Leu Gly Ala Leu Ser Arg Gln Tyr Ser Leu Cys Gly Ala Pro Asp
    530                 535                 540
Ala Pro Ser Tyr Glu Ile Ala Val His Leu Asp Pro Glu Ser Arg Gly
545                 550                 555                 560
Gly Ser Arg Tyr Ile His Glu Gln Leu Glu Val Gly Ser Pro Leu Arg
                565                 570                 575
Met Arg Gly Pro Arg Asn His Phe Ala Leu Asp Pro Gly Ala Glu His
            580                 585                 590
Tyr Val Phe Val Ala Gly Gly Ile Gly Ile Thr Pro Val Leu Ala Met
        595                 600                 605
Ala Asp His Ala Arg Ala Arg Gly Trp Ser Tyr Glu Leu His Tyr Cys
    610                 615                 620
Gly Arg Asn Arg Ser Gly Met Ala Tyr Leu Glu Arg Val Ala Gly His
625                 630                 635                 640
Gly Asp Arg Ala Ala Leu His Val Ser Glu Glu Gly Thr Arg Ile Asp
                645                 650                 655
Leu Ala Ala Leu Leu Ala Glu Pro Ala Pro Gly Val Gln Ile Tyr Ala
            660                 665                 670
Cys Gly Pro Gly Arg Leu Leu Ala Gly Leu Glu Asp Ala Ser Arg Asn
        675                 680                 685
```

```
Trp Pro Asp Gly Ala Leu His Val Glu His Phe Thr Ser Ser Leu Ala
    690                 695                 700
Ala Leu Asp Pro Asp Val Glu His Ala Phe Asp Leu Glu Leu Arg Asp
705                 710                 715                 720
Ser Gly Leu Thr Val Arg Val Glu Pro Thr Gln Thr Val Leu Asp Ala
                725                 730                 735
Leu Arg Ala Asn Asn Ile Asp Val Pro Ser Asp Cys Glu Glu Gly Leu
            740                 745                 750
Cys Gly Ser Cys Glu Val Ala Val Leu Asp Gly Glu Val Asp His Arg
        755                 760                 765
Asp Thr Val Leu Thr Lys Ala Glu Arg Ala Ala Asn Arg Gln Met Met
    770                 775                 780
Thr Cys Cys Ser Arg Ala Cys Gly Asp Arg Leu Ala Leu Arg Leu
785                 790                 795
```

<210> 17
<211> 2400
<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> 17

[0016]

```
atgccaacac tgcccagaac atttgacgac attcagtccc gactgattaa cgccacctcc      60
agggtggtgc cgatgcagag gcaaattcag ggactgaaat tcttaatgag cgccaagagg     120
aagaccttcg gcccacgccg accgatgccc gaattcgttg aaacacccat cccggacgtt     180
aacacgctgg cccttgagga catcgatgtc agcaatccgt ttttataccg gcagggtcag     240
tggcgcgcct atttcaaacg gttgcgtgat gaggcgccgg tccattacca gaagaacagc     300
cctttcggcc ccttctggtc ggtaactcgg tttgaagaca tcctgttcgt ggataagagt     360
cacgacctgt tttccgccga gccgcaaatc attctcggtg accctccgga ggggctgtcg     420
atggaaatgt tcatagcgat ggatccgccg aaacacgatg tgcagcgcag ctcggtgcag     480
ggagtagtgg caccgaaaaa cctgaaggag atggaggggc tgatccgatc acgcaccggc     540
gatgtgcttg acagcctgcc tacagacaaa ccctttaact gggtacctgc tgtttccaag     600
gaactcacag gccgcatgct ggcgacgctt ctggattttc cttacgagga acgccacaag     660
ctggttgagt ggtcggacag aatggcaggt gcagcatcgg ccaccggcgg ggagtttgcc     720

gatgaaaatg ccatgtttga cgacgcggca gacatggccc ggtctttctc caggctttgg     780
cgggacaagg aggcgcgccg cgcagcaggc gaggagcccg gtttcgattt gatcagcctg     840
ttgcagagca acaaagaaac gaaagacctg atcaatcggc cgatggagtt tatcggtaat     900
ttgacgctgc tcatagtcgc cggcaacgat acgacgcgca actcgatgag tggtggcctg     960
gtggccatga acgaattccc cagggaattt gaaaaattga aggcaaaacc ggagttgatt    1020
ccgaacatgg tgtcggaaat catccgctgg caaacgccgc tggcctatat gcgccgaatc    1080
gccaagcagg atgtcgaact gggcggccag accatcaaga agggtgatcg agttgtcatg    1140
tggtacgcgt cgggtaaccg ggacgagcgc aaatttgaca accccgatca gttcatcatt    1200
gatcgcaagg acgcacgaaa ccacatgtcg ttcggctatg gggttcaccg ttgcatgggc    1260
aaccgtctgg ctgaactgca actgcgcatc ctctgggaag aaatactcaa gcgttttgac    1320
aacatcgaag tcgtcgaaga gcccgagcgg gtgcagtcca acttcgtgcg gggctattcc    1380
aggttgatgg tcaaactgac accgaacagt gtactccatc gtcatcaacc tgtcaccatc    1440
ggcgagccgg ccgctcgtgc tgtgagccgc acggtgaccg ttgagcgtct tgatcgcatt    1500
gccgacgatg tccttcgcct ggtccttcgc gatgctggag gtaaaaccct cccgacgtgg    1560
acgcctggcg ctcacatcga cctggatctg ggtgctctga gccgtcagta ttcgctctgc    1620
ggcgctccgg atgctccgtc gtacgaaatc gccgtgcact tagatccgga aagccgtggt    1680
ggaagccgct atattcatga acagctggaa gttggaagtc cgctgcgtat gcgtggccca    1740
cgcaaccatt tcgccctgga tccgggtgcg gaacattacg tgtttgttgc cgggggtatc    1800
ggcatcacgc cggtgctggc aatggcggat catgcccgtg cgcgtggttg gtcgtacgaa    1860
ctgcattatt gtggtcgtaa tcgtagcggt atggcttacc tggaacgcgt cgcgggacat    1920
ggtgaccgcg ctgccttgca cgtatctgaa gaaggcaccc gcattgatct ggcggcatta    1980
cttgctgaac cggcgccggg cgtgcaaatc tacgcctgcg gtccgggccg tttattagcg    2040
ggtcttgaag acgcgtctcg taattggccg gatggcgcgc ttcatgtgga gcatttcact    2100
tcgagtttag ccgctttgga tccggatgtc gaacatgcct ttgatttgga gctgcgtgac    2160
tctggcctta ccgttcgcgt cgagccaact cagaccgttt tagacgcttt gcgtgcgaac    2220
aatatcgacg tcccgtcgga ttgcgaagag gggctgtgtg gttcttgcga agtagccgtt    2280
ctggatggcg aggttgatca ccgtgatacc gttctgacta aggccgagcg cgccgcgaat    2340
cgtcagatga tgacttgctg cagtcgtgca tgcggtgatc gtctggcgct gcgcctctaa    2400
```

<210> 18
<211> 799
<212> PRT
<213> 人工序列
<220>
<221> 来源
<223> /注="人工序列的描述：合成的多肽"
<400> 18

```
Met Pro Thr Leu Pro Arg Thr Phe Asp Asp Ile Gln Ser Arg Leu Ile
1               5                  10                  15
Asn Ala Thr Ser Arg Val Val Pro Met Gln Arg Gln Ile Gln Gly Leu
```

[0017]

```
                20                  25                  30
Lys Phe Leu Met Ser Ala Lys Arg Lys Thr Phe Gly Pro Arg Arg Pro
        35                  40                  45
Met Pro Glu Phe Val Glu Thr Pro Ile Pro Asp Val Asn Thr Leu Ala
    50                  55                  60
Leu Glu Asp Ile Asp Val Ser Asn Pro Phe Leu Tyr Arg Gln Gly Gln
65                  70                  75                  80
Trp Arg Ala Tyr Phe Lys Arg Leu Arg Asp Glu Ala Pro Val His Tyr
                85                  90                  95
Gln Lys Asn Ser Pro Phe Gly Pro Phe Trp Ser Val Thr Arg Phe Glu
            100                 105                 110
Asp Ile Leu Phe Val Asp Lys Ser His Asp Leu Phe Ser Ala Glu Pro
        115                 120                 125
Gln Ile Ile Leu Gly Asp Pro Pro Glu Gly Leu Ser Met Glu Met Phe
    130                 135                 140
Ile Ala Met Asp Pro Pro Lys His Asp Val Gln Arg Ser Ser Val Gln
145                 150                 155                 160
Gly Val Val Ala Pro Lys Asn Leu Lys Glu Met Glu Gly Leu Ile Arg
                165                 170                 175
Ser Arg Thr Gly Asp Val Leu Asp Ser Leu Pro Thr Asp Lys Pro Phe
            180                 185                 190
Asn Trp Val Pro Ala Val Ser Lys Glu Leu Thr Gly Arg Met Leu Ala
        195                 200                 205
Thr Leu Leu Asp Phe Pro Tyr Glu Glu Arg His Lys Leu Val Glu Trp
    210                 215                 220
Ser Asp Arg Met Ala Gly Ala Ala Ser Ala Thr Gly Gly Glu Phe Ala
225                 230                 235                 240
Asp Glu Asn Ala Met Phe Asp Asp Ala Ala Asp Met Ala Arg Ser Phe
                245                 250                 255
Ser Arg Leu Trp Arg Asp Lys Glu Ala Arg Arg Ala Ala Gly Glu Glu
            260                 265                 270
Pro Gly Phe Asp Leu Ile Ser Leu Leu Gln Ser Asn Lys Glu Thr Lys
        275                 280                 285
Asp Leu Ile Asn Arg Pro Met Glu Phe Ile Gly Asn Leu Thr Leu Leu
    290                 295                 300
Ile Val Ala Gly Asn Asp Thr Thr Arg Asn Ser Met Ser Gly Gly Leu
305                 310                 315                 320
Val Ala Met Asn Glu Phe Pro Arg Glu Phe Glu Lys Leu Lys Ala Lys
                325                 330                 335
Pro Glu Leu Ile Pro Asn Met Val Ser Glu Ile Ile Arg Trp Gln Thr
            340                 345                 350
Pro Leu Ala Tyr Met Arg Arg Ile Ala Lys Gln Asp Val Glu Leu Gly
        355                 360                 365
Gly Gln Thr Ile Lys Lys Gly Asp Arg Val Val Met Trp Tyr Ala Ser
    370                 375                 380
Gly Asn Arg Asp Glu Arg Lys Phe Asp Asn Pro Asp Gln Phe Ile Ile
385                 390                 395                 400
Asp Arg Lys Asp Ala Arg Asn His Met Ser Phe Gly Tyr Gly Val His
                405                 410                 415
Arg Cys Met Gly Asn Arg Leu Ala Glu Leu Gln Leu Arg Ile Leu Trp
            420                 425                 430
Glu Glu Ile Leu Lys Arg Phe Asp Asn Ile Glu Val Val Glu Glu Pro
        435                 440                 445
Glu Arg Val Gln Ser Asn Phe Val Arg Gly Tyr Ser Arg Leu Met Val
    450                 455                 460
Lys Leu Thr Pro Asn Ser Val Leu His Arg His Gln Pro Val Thr Ile
465                 470                 475                 480
Gly Glu Pro Ala Ala Arg Ala Val Ser Arg Thr Val Thr Val Glu Arg
                485                 490                 495
Leu Asp Arg Ile Ala Asp Asp Val Leu Arg Leu Val Leu Arg Asp Ala
            500                 505                 510
Gly Gly Lys Thr Leu Pro Thr Trp Thr Pro Gly Ala His Ile Asp Leu
        515                 520                 525
Asp Leu Gly Ala Leu Ser Arg Gln Tyr Ser Leu Cys Gly Ala Pro Asp
    530                 535                 540
Ala Pro Ser Tyr Glu Ile Ala Val His Leu Asp Pro Glu Ser Arg Gly
545                 550                 555                 560
Gly Ser Arg Tyr Ile His Glu Gln Leu Glu Val Gly Ser Pro Leu Arg
                565                 570                 575
Met Arg Gly Pro Arg Asn His Phe Ala Leu Asp Pro Gly Ala Glu His
            580                 585                 590
Tyr Val Phe Val Ala Gly Gly Ile Gly Ile Thr Pro Val Leu Ala Met
        595                 600                 605
Ala Asp His Ala Arg Ala Arg Gly Trp Ser Tyr Glu Leu His Tyr Cys
    610                 615                 620
Gly Arg Asn Arg Ser Gly Met Ala Tyr Leu Glu Arg Val Ala Gly His
```

625 630 635 640
Gly Asp Arg Ala Ala Leu His Val Ser Glu Glu Gly Thr Arg Ile Asp
645 650 655

Leu Ala Ala Leu Leu Ala Glu Pro Ala Pro Gly Val Gln Ile Tyr Ala
660 665 670
Cys Gly Pro Gly Arg Leu Leu Ala Gly Leu Glu Asp Ala Ser Arg Asn
675 680 685
Trp Pro Asp Gly Ala Leu His Val Glu His Phe Thr Ser Ser Leu Ala
690 695 700
Ala Leu Asp Pro Asp Val Glu His Ala Phe Asp Leu Glu Leu Arg Asp
705 710 715 720
Ser Gly Leu Thr Val Arg Val Glu Pro Thr Gln Thr Val Leu Asp Ala
725 730 735
Leu Arg Ala Asn Asn Ile Asp Val Pro Ser Asp Cys Glu Glu Gly Leu
740 745 750
Cys Gly Ser Cys Glu Val Ala Val Leu Asp Gly Glu Val Asp His Arg
755 760 765
Asp Thr Val Leu Thr Lys Ala Glu Arg Ala Ala Asn Arg Gln Met Met
770 775 780
Thr Cys Cys Ser Arg Ala Cys Gly Asp Arg Leu Ala Leu Arg Leu
785 790 795

<210> 19
<211> 2400
<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

[0018]

atgccaacac tgcccagaac atttgacgac attcagtccc gactgattaa cgccacctcc 60
agggtggtgc cgatgcagag gcaaattcag ggactgaaat tcttaatgag cgccaagagg 120
aagaccttcg gcccacgccg accgatgccc gaattcgttg aaacacccat cccggacgtt 180
aacacgctgg cccttgagga catcgatgtc agcaatccgt ttttataccg gcagggtcag 240
tggcgcgcct atttcaaacg gttgcgtgat gaggcgccgg tccattacca gaagaacagc 300
cctttcggcc ccttctggtc ggtaactcgg tttgaagaca tcctgttcgt ggataagagt 360
cacgacctgt tttccgccga gccgcaaatc attctcggtg accctccgga ggggctgtcg 420
ctggaaatgt tcatagcgat ggatccgccg aaacacgatg tgcagcgcag ctcggtgcag 480
ggagtagtgg caccgaaaaa cctgaaggag atggaggggc tgatccgatc acgcaccggc 540
gatgtgcttg acagcctgcc tacagacaaa ccctttaact gggtacctgc tgtttccaag 600
gaactcacag gccgcatgct ggcgacgctt ctggattttc cttacgagga acgccacaag 660
ctggttgagt ggtcggacag aatggcaggt gcagcatcgg ccaccggcgg ggagtttgcc 720
gatgaaaatg ccatgtttga cgacgcggca gacatggccc ggtctttctc caggctttgg 780
cgggacaagg aggcgcgccg cgcagcaggc gaggagcccg gtttcgattt gatcagcctg 840
ttgcagagca acaaagaaac gaaagacctg atcaatcggc cgatggagtt tatcggtaat 900
ttgacgctgc tcatagtcgc cggcaacgat acgacgcgca actcgatgag tggtggcctg 960
gtggccatga acgaattccc cagggaattt gaaaaattga aggcaaaacc ggagttgatt 1020
ccgaacatgg tgtcggaaat catccgctgg caaacgccgc tggcctatat gcgccgaatc 1080
gccaagcagg atgtcgaact gggcggccag accatcaaga agggtgatcg agttgtcatg 1140
tggtacgcgt cgggtaaccg ggacgagcgc aaatttgaca accccgatca gttcatcatt 1200
gatcgcaagg acgcacgaaa ccacatgtcg ttcggctatg gggttcaccg ttgcatgggc 1260
aaccgtctgg ctgaactgca actgcgcatc ctctgggaag aaatactcaa gcgttttgac 1320
aacatcgaag tcgtcgaaga gcccgagcgg gtgcagtcca acttcgtgcg gggctattcc 1380
aggttgatgg tcaaactgac accgaacagt gtactccatc gtcatcaacc tgtcaccatc 1440
ggcgagccgg ccgctcgtgc tgtgagccgc acggtgaccg ttgagcgtct tgatcgcatt 1500
gccgacgatg tccttcgcct ggtccttcgc gatgctggag gtaaaaccct cccgacgtgg 1560
acgcctggcg ctcacatcga cctggatctg ggtgctctga gccgtcagta ttcgctctgc 1620
ggcgctccgg atgctccgtc gtacgaaatc gccgtgcact tagatccgga aagccgtggt 1680
ggaagccgct atattcatga acagctggaa gttggaagtc cgctgcgtat gcgtggccca 1740
cgcaaccatt tcgccctgga tccgggtgcg gaacattacg tgtttgttgc cgggggtatc 1800
ggcatcacgc cggtgctggc aatggcggat catgcccgtg cgcgtggttg gtcgtacgaa 1860
ctgcattatt gtggtcgtaa tcgtagcggt atggcttacc tggaacgcgt cgcgggacat 1920
ggtgaccgcg ctgccttgca cgtatctgaa gaaggcaccc gcattgatct ggcggcatta 1980
cttgctgaac cggcgccggg cgtgcaaatc tacgcctgcg gtccgggccg tttattagcg 2040
ggtcttgaag acgcgtctcg taattggccg gatggcgcgc ttcatgtgga gcatttcact 2100
tcgagtttag ccgctttgga tccggatgtc gaacatgcct ttgatttgga gctgcgtgac 2160
tctggcctta ccgttcgcgt cgagccaact cagaccgttt tagacgcttt gcgtgcgaac 2220
aatatcgacg tcccgtcgga ttgcgaagag gggctgtgtg gttcttgcga agtagccgtt 2280
ctggatggcg aggttgatca ccgtgatacc gttctgacta aggccgagcg cgccgcgaat 2340
cgtcagatga tgacttgctg cagtcgtgca tgcggtgatc gtctggcgct gcgcctctaa 2400

<210> 20
<211> 799
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注="人工序列的描述：合成的多肽"
<400> 20

```
Met Pro Thr Leu Pro Arg Thr Phe Asp Asp Ile Gln Ser Arg Leu Ile
1               5                  10                  15
Asn Ala Thr Ser Arg Val Val Pro Met Gln Arg Gln Ile Gln Gly Leu
            20                  25                  30
Lys Phe Leu Met Ser Ala Lys Arg Lys Thr Phe Gly Pro Arg Arg Pro
        35                  40                  45
Met Pro Glu Phe Val Glu Thr Pro Ile Pro Asp Val Asn Thr Leu Ala
    50                  55                  60
Leu Glu Asp Ile Asp Val Ser Asn Pro Phe Leu Tyr Arg Gln Gly Gln
65                  70                  75                  80
Trp Arg Ala Tyr Phe Lys Arg Leu Arg Asp Glu Ala Pro Val His Tyr
                85                  90                  95
Gln Lys Asn Ser Pro Phe Gly Pro Phe Trp Ser Val Thr Arg Phe Glu
            100                 105                 110
Asp Ile Leu Phe Val Asp Lys Ser His Asp Leu Phe Ser Ala Glu Pro
        115                 120                 125
Gln Ile Ile Leu Gly Asp Pro Pro Glu Gly Leu Ser Leu Glu Met Phe
    130                 135                 140
Ile Ala Met Asp Pro Pro Lys His Asp Val Gln Arg Ser Ser Val Gln
145                 150                 155                 160
Gly Val Val Ala Pro Lys Asn Leu Lys Glu Met Glu Gly Leu Ile Arg
                165                 170                 175
Ser Arg Thr Gly Asp Val Leu Asp Ser Leu Pro Thr Asp Lys Pro Phe
            180                 185                 190
Asn Trp Val Pro Ala Val Ser Lys Glu Leu Thr Gly Arg Met Leu Ala
        195                 200                 205
Thr Leu Leu Asp Phe Pro Tyr Glu Glu Arg His Lys Leu Val Glu Trp
    210                 215                 220
Ser Asp Arg Met Ala Gly Ala Ala Ser Ala Thr Gly Gly Glu Phe Ala
225                 230                 235                 240
Asp Glu Asn Ala Met Phe Asp Asp Ala Ala Asp Met Ala Arg Ser Phe
                245                 250                 255
Ser Arg Leu Trp Arg Asp Lys Glu Ala Arg Arg Ala Ala Gly Glu Glu
            260                 265                 270
Pro Gly Phe Asp Leu Ile Ser Leu Leu Gln Ser Asn Lys Glu Thr Lys
        275                 280                 285
Asp Leu Ile Asn Arg Pro Met Glu Phe Ile Gly Asn Leu Thr Leu Leu
    290                 295                 300
Ile Val Ala Gly Asn Asp Thr Thr Arg Asn Ser Met Ser Gly Gly Leu
305                 310                 315                 320
Val Ala Met Asn Glu Phe Pro Arg Glu Phe Glu Lys Leu Lys Ala Lys
                325                 330                 335
Pro Glu Leu Ile Pro Asn Met Val Ser Glu Ile Ile Arg Trp Gln Thr
            340                 345                 350
Pro Leu Ala Tyr Met Arg Arg Ile Ala Lys Gln Asp Val Glu Leu Gly
        355                 360                 365
Gly Gln Thr Ile Lys Lys Gly Asp Arg Val Val Met Trp Tyr Ala Ser
    370                 375                 380
Gly Asn Arg Asp Glu Arg Lys Phe Asp Asn Pro Asp Gln Phe Ile Ile
385                 390                 395                 400
Asp Arg Lys Asp Ala Arg Asn His Met Ser Phe Gly Tyr Gly Val His
                405                 410                 415
Arg Cys Met Gly Asn Arg Leu Ala Glu Leu Gln Leu Arg Ile Leu Trp
            420                 425                 430
Glu Glu Ile Leu Lys Arg Phe Asp Asn Ile Glu Val Val Glu Glu Pro
        435                 440                 445
Glu Arg Val Gln Ser Asn Phe Val Arg Gly Tyr Ser Arg Leu Met Val
    450                 455                 460
Lys Leu Thr Pro Asn Ser Val Leu His Arg His Gln Pro Val Thr Ile
465                 470                 475                 480
Gly Glu Pro Ala Ala Arg Ala Val Ser Arg Thr Val Thr Val Glu Arg
                485                 490                 495
Leu Asp Arg Ile Ala Asp Asp Val Leu Arg Leu Val Leu Arg Asp Ala
            500                 505                 510
Gly Gly Lys Thr Leu Pro Thr Trp Thr Pro Gly Ala His Ile Asp Leu
        515                 520                 525
Asp Leu Gly Ala Leu Ser Arg Gln Tyr Ser Leu Cys Gly Ala Pro Asp
    530                 535                 540
Ala Pro Ser Tyr Glu Ile Ala Val His Leu Asp Pro Glu Ser Arg Gly
545                 550                 555                 560
Gly Ser Arg Tyr Ile His Glu Gln Leu Glu Val Gly Ser Pro Leu Arg
                565                 570                 575
```

[0019]

```
Met Arg Gly Pro Arg Asn His Phe Ala Leu Asp Pro Gly Ala Glu His
            580                 585                 590
Tyr Val Phe Val Ala Gly Gly Ile Gly Ile Thr Pro Val Leu Ala Met
        595                 600                 605
Ala Asp His Ala Arg Ala Arg Gly Trp Ser Tyr Glu Leu His Tyr Cys
    610                 615                 620
Gly Arg Asn Arg Ser Gly Met Ala Tyr Leu Glu Arg Val Ala Gly His
625                 630                 635                 640
Gly Asp Arg Ala Ala Leu His Val Ser Glu Glu Gly Thr Arg Ile Asp
                645                 650                 655
Leu Ala Ala Leu Leu Ala Glu Pro Ala Pro Gly Val Gln Ile Tyr Ala
            660                 665                 670
Cys Gly Pro Gly Arg Leu Leu Ala Gly Leu Glu Asp Ala Ser Arg Asn
        675                 680                 685
Trp Pro Asp Gly Ala Leu His Val Glu His Phe Thr Ser Ser Leu Ala
    690                 695                 700
Ala Leu Asp Pro Asp Val Glu His Ala Phe Asp Leu Glu Leu Arg Asp
705                 710                 715                 720
Ser Gly Leu Thr Val Arg Val Glu Pro Thr Gln Thr Val Leu Asp Ala
                725                 730                 735
Leu Arg Ala Asn Asn Ile Asp Val Pro Ser Asp Cys Glu Glu Gly Leu
            740                 745                 750
Cys Gly Ser Cys Glu Val Ala Val Leu Asp Gly Glu Val Asp His Arg
        755                 760                 765
Asp Thr Val Leu Thr Lys Ala Glu Arg Ala Ala Asn Arg Gln Met Met
    770                 775                 780
Thr Cys Cys Ser Arg Ala Cys Gly Asp Arg Leu Ala Leu Arg Leu
785                 790                 795
```

<210> 21
<211> 2400
<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> 21

[0020]

```
atgccaacac tgcccagaac atttgacgac attcagtccc gactgattaa cgccacctcc      60
agggtggtgc cgatgcagag gcaaattcag ggactgaaat tcttaatgag cgccaagagg     120
aagaccttcg gcccacgccg accgatgccc gaattcgttg aaacacccat cccggacgtt     180
aacacgctgg cccttgagga catcgatgtc agcaatccgt ttttataccg gcagggtcag     240
tggcgcgcct atttcaaacg gttgcgtgat gaggcgccgg tccattacca gaagaacagc     300
cctttcggcc ccttctggtc ggtaactcgg tttgaagaca tcctgttcgt ggataagagt     360
cacgacctgt tttccgccga gccgcaaatc attctcggtg accctccgga ggggctgtcg     420
acggaaatgt tcatagcgat ggatccgccg aaacacgatg tgcagcgcag ctcggtgcag     480
ggagtagtgg caccgaaaaa cctgaaggag atggaggggc tgatccgatc acgcaccggc     540
gatgtgcttg acagcctgcc tacagacaaa ccctttaact gggtacctgc tgtttccaag     600
gaactcacag gccgcatgct ggcgacgctt ctggattttc cttacgagga acgccacaag     660
ctggttgagt ggtcggacag aatggcaggt gcagcatcgg ccaccggcgg ggagtttgcc     720
gatgaaaatg ccatgtttga cgacgcggca gacatggccc ggtctttctc caggctttgg     780
cgggacaagg aggcgcgccg cgcagcaggc gaggagcccg gtttcgattt gatcagcctg     840
ttgcagagca acaaagaaac gaaagacctg atcaatcggc cgatggagtt tatcggtaat     900
ttgacgctgc tcatagtcgc cggcaacgat acgacgcgca actcgatgag tggtggcctg     960
gtggccatga acgaattccc cagggaattt gaaaaattga aggcaaaacc ggagttgatt    1020
ccgaacatgg tgtcggaaat catccgctgg caaacgccgc tggcctatat gcgccgaatc    1080
gccaagcagg atgtcgaact gggcggccag accatcaaga agggtgatcg agttgtcatg    1140
tggtacgcgt cgggtaaccg ggacgagcgc aaatttgaca accccgatca gttcatcatt    1200
gatcgcaagg acgcacgaaa ccacatgtcg ttcggctatg gggttcaccg ttgcatgggc    1260
aaccgtctgg ctgaactgca actgcgcatc ctctgggaag aaatactcaa gcgttttgac    1320
aacatcgaag tcgtcgaaga gcccgagcgg gtgcagtcca acttcgtgcg gggctattcc    1380
aggttgatgg tcaaactgac accgaacagt gtactccatc gtcatcaacc tgtcaccatc    1440
ggcgagccgg ccgctcgtgc tgtgagccgc acggtgaccg ttgagcgtct tgatcgcatt    1500
gccgacgatg tccttcgcct ggtccttcgc gatgctggag gtaaaaccct cccgacgtgg    1560
acgcctggcg ctcacatcga cctggatctg ggtgctctga gccgtcagta ttcgctctgc    1620
ggcgctccgg atgctccgtc gtacgaaatc gccgtgcact tagatccgga aagccgtggt    1680
ggaagccgct atattcatga acagctggaa gttggaagtc cgctgcgtat gcgtggccca    1740
cgcaaccatt tcgccctgga tccgggtgcg gaacattacg tgtttgttgc cgggggtatc    1800
ggcatcacgc cggtgctggc aatggcggat catgcccgtg cgcgtggttg gtcgtacgaa    1860
ctgcattatt gtggtcgtaa tcgtagcggt atggcttacc tggaacgcgt cgcgggacat    1920
ggtgaccgcg ctgccttgca cgtatctgaa gaaggcaccc gcattgatct ggcggcatta    1980
cttgctgaac cggcgccggg cgtgcaaatc tacgcctgcg gtccgggccg tttattagcg    2040
ggtcttgaag acgcgtctcg taattggccg gatggcgcgc ttcatgtgga gcatttcact    2100
tcgagtttag ccgctttgga tccggatgtc gaacatgcct ttgatttgga gctgcgtgac    2160
tctggcctta ccgttcgcgt cgagccaact cagaccgttt tagacgcttt gcgtgcgaac    2220
aatatcgacg tcccgtcgga ttgcgaagag gggctgtgtg gttcttgcga agtagccgtt    2280
ctggatggcg aggttgatca ccgtgatacc gttctgacta aggccgagcg cgccgcgaat    2340
```

cgtcagatga tgacttgctg cagtcgtgca tgcggtgatc gtctggcgct gcgcctctaa 2400

<210> 22
<211> 799
<212> PRT
<213> 人工序列
<220>
<221> 来源
<223> /注="人工序列的描述：合成的多肽"
<400> 22

[0021]

```
Met Pro Thr Leu Pro Arg Thr Phe Asp Asp Ile Gln Ser Arg Leu Ile
1               5                  10                  15
Asn Ala Thr Ser Arg Val Val Pro Met Gln Arg Gln Ile Gln Gly Leu
            20                  25                  30
Lys Phe Leu Met Ser Ala Lys Arg Lys Thr Phe Gly Pro Arg Arg Pro
        35                  40                  45
Met Pro Glu Phe Val Glu Thr Pro Ile Pro Asp Val Asn Thr Leu Ala
    50                  55                  60
Leu Glu Asp Ile Asp Val Ser Asn Pro Phe Leu Tyr Arg Gln Gly Gln
65                  70                  75                  80
Trp Arg Ala Tyr Phe Lys Arg Leu Arg Asp Glu Ala Pro Val His Tyr
                85                  90                  95
Gln Lys Asn Ser Pro Phe Gly Pro Phe Trp Ser Val Thr Arg Phe Glu
            100                 105                 110
Asp Ile Leu Phe Val Asp Lys Ser His Asp Leu Phe Ser Ala Glu Pro
        115                 120                 125
Gln Ile Ile Leu Gly Asp Pro Pro Glu Gly Leu Ser Thr Glu Met Phe
    130                 135                 140
Ile Ala Met Asp Pro Pro Lys His Asp Val Gln Arg Ser Ser Val Gln
145                 150                 155                 160
Gly Val Val Ala Pro Lys Asn Leu Lys Glu Met Glu Gly Leu Ile Arg
                165                 170                 175
Ser Arg Thr Gly Asp Val Leu Asp Ser Leu Pro Thr Asp Lys Pro Phe
            180                 185                 190
Asn Trp Val Pro Ala Val Ser Lys Glu Leu Thr Gly Arg Met Leu Ala
        195                 200                 205
Thr Leu Leu Asp Phe Pro Tyr Glu Glu Arg His Lys Leu Val Glu Trp
    210                 215                 220
Ser Asp Arg Met Ala Gly Ala Ala Ser Ala Thr Gly Gly Glu Phe Ala
225                 230                 235                 240
Asp Glu Asn Ala Met Phe Asp Asp Ala Ala Asp Met Ala Arg Ser Phe
                245                 250                 255
Ser Arg Leu Trp Arg Asp Lys Glu Ala Arg Arg Ala Ala Gly Glu Glu
            260                 265                 270
Pro Gly Phe Asp Leu Ile Ser Leu Leu Gln Ser Asn Lys Glu Thr Lys
        275                 280                 285
Asp Leu Ile Asn Arg Pro Met Glu Phe Ile Gly Asn Leu Thr Leu Leu
    290                 295                 300
Ile Val Ala Gly Asn Asp Thr Thr Arg Asn Ser Met Ser Gly Gly Leu
305                 310                 315                 320
Val Ala Met Asn Glu Phe Pro Arg Glu Phe Glu Lys Leu Lys Ala Lys
                325                 330                 335
Pro Glu Leu Ile Pro Asn Met Val Ser Glu Ile Ile Arg Trp Gln Thr
            340                 345                 350
Pro Leu Ala Tyr Met Arg Arg Ile Ala Lys Gln Asp Val Glu Leu Gly
        355                 360                 365
Gly Gln Thr Ile Lys Lys Gly Asp Arg Val Val Met Trp Tyr Ala Ser
    370                 375                 380
Gly Asn Arg Asp Glu Arg Lys Phe Asp Asn Pro Asp Gln Phe Ile Ile
385                 390                 395                 400
Asp Arg Lys Asp Ala Arg Asn His Met Ser Phe Gly Tyr Gly Val His
                405                 410                 415
Arg Cys Met Gly Asn Arg Leu Ala Glu Leu Gln Leu Arg Ile Leu Trp
            420                 425                 430
Glu Glu Ile Leu Lys Arg Phe Asp Asn Ile Glu Val Val Glu Glu Pro
        435                 440                 445
Glu Arg Val Gln Ser Asn Phe Val Arg Gly Tyr Ser Arg Leu Met Val
    450                 455                 460
Lys Leu Thr Pro Asn Ser Val Leu His Arg His Gln Pro Val Thr Ile
465                 470                 475                 480
Gly Glu Pro Ala Ala Arg Ala Val Ser Arg Thr Val Thr Val Glu Arg
                485                 490                 495
Leu Asp Arg Ile Ala Asp Asp Val Leu Arg Leu Val Leu Arg Asp Ala
            500                 505                 510
Gly Gly Lys Thr Leu Pro Thr Trp Thr Pro Gly Ala His Ile Asp Leu
        515                 520                 525
```

```
Asp Leu Gly Ala Leu Ser Arg Gln Tyr Ser Leu Cys Gly Ala Pro Asp
    530                 535                 540
Ala Pro Ser Tyr Glu Ile Ala Val His Leu Asp Pro Glu Ser Arg Gly
545                 550                 555                 560
Gly Ser Arg Tyr Ile His Glu Gln Leu Glu Val Gly Ser Pro Leu Arg
                565                 570                 575
Met Arg Gly Pro Arg Asn His Phe Ala Leu Asp Pro Gly Ala Glu His
            580                 585                 590
Tyr Val Phe Val Ala Gly Gly Ile Gly Ile Thr Pro Val Leu Ala Met
        595                 600                 605
Ala Asp His Ala Arg Ala Arg Gly Trp Ser Tyr Glu Leu His Tyr Cys
    610                 615                 620
Gly Arg Asn Arg Ser Gly Met Ala Tyr Leu Glu Arg Val Ala Gly His
625                 630                 635                 640
Gly Asp Arg Ala Ala Leu His Val Ser Glu Glu Gly Thr Arg Ile Asp
                645                 650                 655
Leu Ala Ala Leu Leu Ala Glu Pro Ala Pro Gly Val Gln Ile Tyr Ala
            660                 665                 670
Cys Gly Pro Gly Arg Leu Leu Ala Gly Leu Glu Asp Ala Ser Arg Asn
        675                 680                 685
Trp Pro Asp Gly Ala Leu His Val Glu His Phe Thr Ser Ser Leu Ala
    690                 695                 700
Ala Leu Asp Pro Asp Val Glu His Ala Phe Asp Leu Glu Leu Arg Asp
705                 710                 715                 720
Ser Gly Leu Thr Val Arg Val Glu Pro Thr Gln Thr Val Leu Asp Ala
                725                 730                 735
Leu Arg Ala Asn Asn Ile Asp Val Pro Ser Asp Cys Glu Glu Gly Leu
            740                 745                 750
Cys Gly Ser Cys Glu Val Ala Val Leu Asp Gly Glu Val Asp His Arg
        755                 760                 765
Asp Thr Val Leu Thr Lys Ala Glu Arg Ala Ala Asn Arg Gln Met Met
    770                 775                 780
Thr Cys Cys Ser Arg Ala Cys Gly Asp Arg Leu Ala Leu Arg Leu
785                 790                 795
```

[0022]

<210> 23
<211> 2400
<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> 23

```
atgccaacac tgcccagaac atttgacgac attcagtccc gactgattaa cgccacctcc      60
agggtggtgc cgatgcagag gcaaattcag ggactgaaat tcttaatgag cgccaagagg     120
aagaccttcg gcccacgccg accgatgccc gaattcgttg aaacacccat cccggacgtt     180
aacacgctgg cccttgagga catcgatgtc agcaatccgt ttttataccg gcagggtcag     240
tggcgcgcct atttcaaacg gttgcgtgat gaggcgccgg tccattacca gaagaacagc     300
cctttcggcc ccttctggtc ggtaactcgg tttgaagaca tcctgttcgt ggataagagt     360
cacgacctgt tttccgccga gccgcaaatc attctcggtg accctccgga ggggctgtcg     420
gtggaaatgt tcatagcgat ggatccgccg aaacacgatg tgcagcgcag ctcggtgcag     480
ggagtagtgg caccgaaaaa cctgaaggag atggaggggc tgatccgatc aaacaccggc     540
gatgtgcttg acagcctgcc tacagacaaa ccctttaact gggtacctgc tgtttccaag     600
gaactcacag gccgcatgct ggcgacgctt ctggattttc cttacgagga acgccacaag     660
ctggttgagt ggtcggacag aatggcaggt gcagcatcgg ccaccggcgg ggagtttgcc     720
gatgaaaatg ccatgtttga cgacgcggca gacatggccc ggtctttctc caggctttgg     780
cgggacaagg aggcgcgccg cgcagcaggc gaggagcccg gtttcgattt gatcagcctg     840
ttgcagagca acaaagaaac gaaagacctg atcaatcggc cgatggagtt tatcggtaat     900
ttgacgctgc tcatagtcgc cggcaacgat acgacgcgca actcgatgag tggtggcctg     960
gtggccatga acgaattccc cagggaattt gaaaaattga aggcaaaacc ggagttgatt    1020
ccgaacatgg tgtcggaaat catccgctgg caaacgccgc tggcctatat gcgccgaatc    1080
gccaagcagg atgtcgaact gggcggccag accatcaaga agggtgatcg agttgtcatg    1140
tggtacgcgt cgggtaaccg ggacgagcgc aaatttgaca accccgatca gttcatcatt    1200
gatcgcaagg acgcacgaaa ccacatgtcg ttcggctatg gggttcaccg ttgcatgggc    1260
aaccgtctgg ctgaactgca actgcgcatc ctctgggaag aaatactcaa gcgttttgac    1320
aacatcgaag tcgtcgaaga gcccgagcgg gtgcagtcca acttcgtgcg gggctattcc    1380
aggttgatgg tcaaactgac accgaacagt gtactccatc gtcatcaacc tgtcaccatc    1440
ggcgagccgg ccgctcgtgc tgtgagccgc acggtgaccg ttgagcgtct tgatcgcatt    1500
gccgacgatg tccttcgcct ggtccttcgc gatgctggag gtaaaaccct cccgacgtgg    1560
acgcctggcg ctcacatcga cctggatctg ggtgctctga gccgtcagta ttcgctctgc    1620
ggcgctccgg atgctccgtc gtacgaaatc gccgtgcact tagatccgga aagccgtggt    1680
ggaagccgct atattcatga acagctggaa gttggaagtc cgctgcgtat gcgtggccca    1740
cgcaaccatt tcgccctgga tccgggtgcg gaacattacg tgtttgttgc cgggggtatc    1800
ggcatcacgc cggtgctggc aatggcggat catgcccgtg cgcgtggttg gtcgtacgaa    1860
ctgcattatt gtggtcgtaa tcgtagcggt atggcttacc tggaacgcgt cgcgggacat    1920
ggtgaccgcg ctgccttgca cgtatctgaa gaaggcaccc gcattgatct ggcggcatta    1980
```

```
cttgctgaac cggcgccggg cgtgcaaatc tacgcctgcg gtccgggccg tttattagcg     2040
ggtcttgaag acgcgtctcg taattggccg gatggcgcgc ttcatgtgga gcatttcact     2100
tcgagtttag ccgctttgga tccggatgtc gaacatgcct ttgatttgga gctgcgtgac     2160
tctggcctta ccgttcgcgt cgagccaact cagaccgttt tagacgcttt gcgtgcgaac     2220
aatatcgacg tcccgtcgga ttgcgaagag gggctgtgtg gttcttgcga agtagccgtt     2280
ctggatggcg aggttgatca ccgtgatacc gttctgacta aggccgagcg cgccgcgaat     2340
cgtcagatga tgacttgctg cagtcgtgca tgcggtgatc gtctggcgct gcgcctctaa     2400
```

<210> 24
<211> 799
<212> PRT
<213> 人工序列
<220>
<221> 来源
<223> /注="人工序列的描述：合成的多肽"
<400> 24

```
Met Pro Thr Leu Pro Arg Thr Phe Asp Asp Ile Gln Ser Arg Leu Ile
1               5                  10                  15
Asn Ala Thr Ser Arg Val Val Pro Met Gln Arg Gln Ile Gln Gly Leu
            20                  25                  30
Lys Phe Leu Met Ser Ala Lys Arg Lys Thr Phe Gly Pro Arg Arg Pro
        35                  40                  45
Met Pro Glu Phe Val Glu Thr Pro Ile Pro Asp Val Asn Thr Leu Ala
    50                  55                  60
Leu Glu Asp Ile Asp Val Ser Asn Pro Phe Leu Tyr Arg Gln Gly Gln
65                  70                  75                  80
Trp Arg Ala Tyr Phe Lys Arg Leu Arg Asp Glu Ala Pro Val His Tyr
                85                  90                  95
Gln Lys Asn Ser Pro Phe Gly Pro Phe Trp Ser Val Thr Arg Phe Glu
            100                 105                 110
Asp Ile Leu Phe Val Asp Lys Ser His Asp Leu Phe Ser Ala Glu Pro
        115                 120                 125
Gln Ile Ile Leu Gly Asp Pro Pro Glu Gly Leu Ser Val Glu Met Phe
    130                 135                 140
Ile Ala Met Asp Pro Pro Lys His Asp Val Gln Arg Ser Ser Val Gln
145                 150                 155                 160
Gly Val Val Ala Pro Lys Asn Leu Lys Glu Met Glu Gly Leu Ile Arg
                165                 170                 175
Ser Asn Thr Gly Asp Val Leu Asp Ser Leu Pro Thr Asp Lys Pro Phe
            180                 185                 190
Asn Trp Val Pro Ala Val Ser Lys Glu Leu Thr Gly Arg Met Leu Ala
        195                 200                 205
Thr Leu Leu Asp Phe Pro Tyr Glu Glu Arg His Lys Leu Val Glu Trp
    210                 215                 220
Ser Asp Arg Met Ala Gly Ala Ala Ser Ala Thr Gly Gly Glu Phe Ala
225                 230                 235                 240
Asp Glu Asn Ala Met Phe Asp Asp Ala Ala Asp Met Ala Arg Ser Phe
                245                 250                 255
Ser Arg Leu Trp Arg Asp Lys Glu Ala Arg Arg Ala Ala Gly Glu Glu
            260                 265                 270
Pro Gly Phe Asp Leu Ile Ser Leu Leu Gln Ser Asn Lys Glu Thr Lys
        275                 280                 285
Asp Leu Ile Asn Arg Pro Met Glu Phe Ile Gly Asn Leu Thr Leu Leu
    290                 295                 300
Ile Val Ala Gly Asn Asp Thr Thr Arg Asn Ser Met Ser Gly Gly Leu
305                 310                 315                 320
Val Ala Met Asn Glu Phe Pro Arg Glu Phe Glu Lys Leu Lys Ala Lys
                325                 330                 335
Pro Glu Leu Ile Pro Asn Met Val Ser Glu Ile Ile Arg Trp Gln Thr
            340                 345                 350
Pro Leu Ala Tyr Met Arg Arg Ile Ala Lys Gln Asp Val Glu Leu Gly
        355                 360                 365
Gly Gln Thr Ile Lys Lys Gly Asp Arg Val Val Met Trp Tyr Ala Ser
    370                 375                 380
Gly Asn Arg Asp Glu Arg Lys Phe Asp Asn Pro Asp Gln Phe Ile Ile
385                 390                 395                 400
Asp Arg Lys Asp Ala Arg Asn His Met Ser Phe Gly Tyr Gly Val His
                405                 410                 415
Arg Cys Met Gly Asn Arg Leu Ala Glu Leu Gln Leu Arg Ile Leu Trp
            420                 425                 430
Glu Glu Ile Leu Lys Arg Phe Asp Asn Ile Glu Val Val Glu Glu Pro
        435                 440                 445
Glu Arg Val Gln Ser Asn Phe Val Arg Gly Tyr Ser Arg Leu Met Val
    450                 455                 460
Lys Leu Thr Pro Asn Ser Val Leu His Arg His Gln Pro Val Thr Ile
465                 470                 475                 480
```

[0023]

```
Gly Glu Pro Ala Ala Arg Ala Val Ser Arg Thr Val Thr Val Glu Arg
                485                 490                 495
Leu Asp Arg Ile Ala Asp Asp Val Leu Arg Leu Val Leu Arg Asp Ala
            500                 505                 510
Gly Gly Lys Thr Leu Pro Thr Trp Thr Pro Gly Ala His Ile Asp Leu
        515                 520                 525
Asp Leu Gly Ala Leu Ser Arg Gln Tyr Ser Leu Cys Gly Ala Pro Asp
    530                 535                 540
Ala Pro Ser Tyr Glu Ile Ala Val His Leu Asp Pro Glu Ser Arg Gly
545                 550                 555                 560
Gly Ser Arg Tyr Ile His Glu Gln Leu Glu Val Gly Ser Pro Leu Arg
                565                 570                 575
Met Arg Gly Pro Arg Asn His Phe Ala Leu Asp Pro Gly Ala Glu His
            580                 585                 590
Tyr Val Phe Val Ala Gly Gly Ile Gly Ile Thr Pro Val Leu Ala Met
        595                 600                 605
Ala Asp His Ala Arg Ala Arg Gly Trp Ser Tyr Glu Leu His Tyr Cys
    610                 615                 620
Gly Arg Asn Arg Ser Gly Met Ala Tyr Leu Glu Arg Val Ala Gly His
625                 630                 635                 640
Gly Asp Arg Ala Ala Leu His Val Ser Glu Glu Gly Thr Arg Ile Asp
                645                 650                 655
Leu Ala Ala Leu Leu Ala Glu Pro Ala Pro Gly Val Gln Ile Tyr Ala
            660                 665                 670
Cys Gly Pro Gly Arg Leu Leu Ala Gly Leu Glu Asp Ala Ser Arg Asn
        675                 680                 685
Trp Pro Asp Gly Ala Leu His Val Glu His Phe Thr Ser Ser Leu Ala
    690                 695                 700
Ala Leu Asp Pro Asp Val Glu His Ala Phe Asp Leu Glu Leu Arg Asp
705                 710                 715                 720
Ser Gly Leu Thr Val Arg Val Glu Pro Thr Gln Thr Val Leu Asp Ala
                725                 730                 735
Leu Arg Ala Asn Asn Ile Asp Val Pro Ser Asp Cys Glu Glu Gly Leu
            740                 745                 750
Cys Gly Ser Cys Glu Val Ala Val Leu Asp Gly Glu Val Asp His Arg
        755                 760                 765
Asp Thr Val Leu Thr Lys Ala Glu Arg Ala Ala Asn Arg Gln Met Met
    770                 775                 780
Thr Cys Cys Ser Arg Ala Cys Gly Asp Arg Leu Ala Leu Arg Leu
785                 790                 795
```

[0024]

<210> 25
<211> 2400
<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> 25

```
atgccaacac tgcccagaac atttgacgac attcagtccc gactgattaa cgccacctcc       60
agggtggtgc cgatgcagag gcaaattcag ggactgaaat tcttaatgag cgccaagagg      120
aagaccttcg gcccacgccg accgatgccc gaattcgttg aaacacccat cccggacgtt      180
aacacgctgg cccttgagga catcgatgtc agcaatccgt ttttataccg gcagggtcag      240
tggcgcgcct atttcaaacg gttgcgtgat gaggcgccgg tccattacca gaagaacagc      300
cctttcggcc ccttctggtc ggtaactcgg tttgaagaca tcctgttcgt ggataagagt      360
cacgacctgt tttccgccga gccgcaaatc attctcggtg accctccgga ggggctgtcg      420
gtggaaatgt tcatagcgat ggatccgccg aaacacgatg tgcagcgcag ctcggtgcag      480
ggagtagtgg caccgaaaaa cctgaaggag atggaggggc tgatccgatc acgcaccggc      540
gatgtgcttg acagcctgcc tacagacaaa ccctttaact gggtacctgc tgtttccaag      600
gaactcacag gccgcatgct ggcgacgctt ctggattttc cttacgagga acgccacaag      660
ctggttgagt ggtcggacag aatggcaggt gcagcatcgg ccaccggcgg ggagtttgcc      720
gatgaaaatg ccatgtttga cgacgcggca gacatggccc ggtctttctc caggctttgg      780
cgggacaagg aggcgcgccg cgcagcaggc gaggagcccg gtttcgattt gatcagcctg      840
ttgcagagca acaaagaaac gaaagacctg atcaatcggc cgatggagtt tatcggtaat      900
ttgacgctgc tcatagtcgc cggccgggat acgacgcgca actcgatgag tggtggcctg      960
gtggccatga acgaattccc cagggaattt gaaaaattga aggcaaaacc ggagttgatt     1020
ccgaacatgg tgtcggaaat catccgctgg caaacgccgc tggcctatat gcgccgaatc     1080
gccaagcagg atgtcgaact gggcggccag accatcaaga agggtgatcg agttgtcatg     1140
tggtacgcgt cgggtaaccg ggacgagcgc aaatttgaca accccgatca gttcatcatt     1200
gatcgcaagg acgcacgaaa ccacatgtcg ttcggctatg gggttcaccg ttgcatgggc     1260
aaccgtctgg ctgaactgca actgcgcatc ctctgggaag aaatactcaa gcgttttgac     1320
aacatcgaag tcgtcgaaga gcccgagcgg gtgcagtcca acttcgtgcg gggctattcc     1380
aggttgatgg tcaaactgac accgaacagt gtactccatc gtcatcaacc tgtcaccatc     1440
ggcgagccgg ccgctcgtgc tgtgagccgc acgggtgaccg ttgagcgtct tgatcgcatt     1500
gccgacgatg tccttcgcct ggtccttcgc gatgctggag gtaaaaccct cccgacgtgg     1560
acgcctggcg ctcacatcga cctggatctg ggtgctctga gccgtcagta ttcgctctgc     1620
```

```
ggcgctccgg atgctccgtc gtacgaaatc gccgtgcact tagatccgga aagccgtggt     1680
ggaagccgct atattcatga acagctggaa gttggaagtc cgctgcgtat gcgtggccca     1740
cgcaaccatt tcgccctgga tccgggtgcg gaacattacg tgtttgttgc cgggggtatc     1800
ggcatcacgc cggtgctggc aatggcggat catgcccgtg cgcgtggttg gtcgtacgaa     1860
ctgcattatt gtggtcgtaa tcgtagcggt atggcttacc tggaacgcgt cgcgggacat     1920
ggtgaccgcg ctgccttgca cgtatctgaa gaaggcaccc gcattgatct ggcggcatta     1980
cttgctgaac cggcgccggg cgtgcaaatc tacgcctgcg gtccgggccg tttattagcg     2040
ggtcttgaag acgcgtctcg taattggccg gatggcgcgc ttcatgtgga gcatttcact     2100
tcgagtttag ccgctttgga tccggatgtc gaacatgcct ttgatttgga gctgcgtgac     2160
tctggcctta ccgttcgcgt cgagccaact cagaccgttt tagacgcttt gcgtgcgaac     2220
aatatcgacg tcccgtcgga ttgcgaagag gggctgtgtg gttcttgcga agtagccgtt     2280
ctggatggcg aggttgatca ccgtgatacc gttctgacta aggccgagcg cgccgcgaat     2340
cgtcagatga tgacttgctg cagtcgtgca tgcggtgatc gtctggcgct gcgcctctaa     2400
```

<210> 26
<211> 799
<212> PRT
<213> 人工序列
<220>
<221> 来源
<223> /注="人工序列的描述：合成的多肽"
<400> 26

[0025]

```
Met Pro Thr Leu Pro Arg Thr Phe Asp Asp Ile Gln Ser Arg Leu Ile
1               5                  10                  15
Asn Ala Thr Ser Arg Val Val Pro Met Gln Arg Gln Ile Gln Gly Leu
            20                  25                  30
Lys Phe Leu Met Ser Ala Lys Arg Lys Thr Phe Gly Pro Arg Arg Pro
        35                  40                  45
Met Pro Glu Phe Val Glu Thr Pro Ile Pro Asp Val Asn Thr Leu Ala
    50                  55                  60
Leu Glu Asp Ile Asp Val Ser Asn Pro Phe Leu Tyr Arg Gln Gly Gln
65                  70                  75                  80
Trp Arg Ala Tyr Phe Lys Arg Leu Arg Asp Glu Ala Pro Val His Tyr
                85                  90                  95
Gln Lys Asn Ser Pro Phe Gly Pro Phe Trp Ser Val Thr Arg Phe Glu
            100                 105                 110
Asp Ile Leu Phe Val Asp Lys Ser His Asp Leu Phe Ser Ala Glu Pro
        115                 120                 125
Gln Ile Ile Leu Gly Asp Pro Pro Glu Gly Leu Ser Val Glu Met Phe
    130                 135                 140
Ile Ala Met Asp Pro Pro Lys His Asp Val Gln Arg Ser Ser Val Gln
145                 150                 155                 160
Gly Val Val Ala Pro Lys Asn Leu Lys Glu Met Glu Gly Leu Ile Arg
                165                 170                 175
Ser Arg Thr Gly Asp Val Leu Asp Ser Leu Pro Thr Asp Lys Pro Phe
            180                 185                 190
Asn Trp Val Pro Ala Val Ser Lys Glu Leu Thr Gly Arg Met Leu Ala
        195                 200                 205
Thr Leu Leu Asp Phe Pro Tyr Glu Glu Arg His Lys Leu Val Glu Trp
    210                 215                 220
Ser Asp Arg Met Ala Gly Ala Ala Ser Ala Thr Gly Gly Glu Phe Ala
225                 230                 235                 240
Asp Glu Asn Ala Met Phe Asp Asp Ala Ala Asp Met Ala Arg Ser Phe
                245                 250                 255
Ser Arg Leu Trp Arg Asp Lys Glu Ala Arg Arg Ala Ala Gly Glu Glu
            260                 265                 270
Pro Gly Phe Asp Leu Ile Ser Leu Leu Gln Ser Asn Lys Glu Thr Lys
        275                 280                 285
Asp Leu Ile Asn Arg Pro Met Glu Phe Ile Gly Asn Leu Thr Leu Leu
    290                 295                 300
Ile Val Ala Gly Arg Asp Thr Thr Arg Asn Ser Met Ser Gly Gly Leu
305                 310                 315                 320
Val Ala Met Asn Glu Phe Pro Arg Glu Phe Glu Lys Leu Lys Ala Lys
                325                 330                 335
Pro Glu Leu Ile Pro Asn Met Val Ser Glu Ile Ile Arg Trp Gln Thr
            340                 345                 350
Pro Leu Ala Tyr Met Arg Arg Ile Ala Lys Gln Asp Val Glu Leu Gly
        355                 360                 365
Gly Gln Thr Ile Lys Lys Gly Asp Arg Val Val Met Trp Tyr Ala Ser
    370                 375                 380
Gly Asn Arg Asp Glu Arg Lys Phe Asp Asn Pro Asp Gln Phe Ile Ile
385                 390                 395                 400
Asp Arg Lys Asp Ala Arg Asn His Met Ser Phe Gly Tyr Gly Val His
                405                 410                 415
Arg Cys Met Gly Asn Arg Leu Ala Glu Leu Gln Leu Arg Ile Leu Trp
            420                 425                 430
```

```
Glu Glu Ile Leu Lys Arg Phe Asp Asn Ile Glu Val Val Glu Glu Pro
        435                 440                 445
Glu Arg Val Gln Ser Asn Phe Val Arg Gly Tyr Ser Arg Leu Met Val
    450                 455                 460
Lys Leu Thr Pro Asn Ser Val Leu His Arg His Gln Pro Val Thr Ile
465                 470                 475                 480
Gly Glu Pro Ala Ala Arg Ala Val Ser Arg Thr Val Thr Val Glu Arg
                485                 490                 495
Leu Asp Arg Ile Ala Asp Asp Val Leu Arg Leu Val Leu Arg Asp Ala
            500                 505                 510
Gly Gly Lys Thr Leu Pro Thr Trp Thr Pro Gly Ala His Ile Asp Leu
        515                 520                 525
Asp Leu Gly Ala Leu Ser Arg Gln Tyr Ser Leu Cys Gly Ala Pro Asp
    530                 535                 540
Ala Pro Ser Tyr Glu Ile Ala Val His Leu Asp Pro Glu Ser Arg Gly
545                 550                 555                 560
Gly Ser Arg Tyr Ile His Glu Gln Leu Glu Val Gly Ser Pro Leu Arg
                565                 570                 575
Met Arg Gly Pro Arg Asn His Phe Ala Leu Asp Pro Gly Ala Glu His
            580                 585                 590
Tyr Val Phe Val Ala Gly Gly Ile Gly Ile Thr Pro Val Leu Ala Met
        595                 600                 605
Ala Asp His Ala Arg Ala Arg Gly Trp Ser Tyr Glu Leu His Tyr Cys
    610                 615                 620
Gly Arg Asn Arg Ser Gly Met Ala Tyr Leu Glu Arg Val Ala Gly His
625                 630                 635                 640
Gly Asp Arg Ala Ala Leu His Val Ser Glu Glu Gly Thr Arg Ile Asp
                645                 650                 655
Leu Ala Ala Leu Leu Ala Glu Pro Ala Pro Gly Val Gln Ile Tyr Ala
            660                 665                 670
Cys Gly Pro Gly Arg Leu Leu Ala Gly Leu Glu Asp Ala Ser Arg Asn
        675                 680                 685
Trp Pro Asp Gly Ala Leu His Val Glu His Phe Thr Ser Ser Leu Ala
    690                 695                 700
Ala Leu Asp Pro Asp Val Glu His Ala Phe Asp Leu Glu Leu Arg Asp
705                 710                 715                 720
Ser Gly Leu Thr Val Arg Val Glu Pro Thr Gln Thr Val Leu Asp Ala
                725                 730                 735
Leu Arg Ala Asn Asn Ile Asp Val Pro Ser Asp Cys Glu Glu Gly Leu
            740                 745                 750
Cys Gly Ser Cys Glu Val Ala Val Leu Asp Gly Glu Val Asp His Arg
        755                 760                 765
Asp Thr Val Leu Thr Lys Ala Glu Arg Ala Ala Asn Arg Gln Met Met
    770                 775                 780
Thr Cys Cys Ser Arg Ala Cys Gly Asp Arg Leu Ala Leu Arg Leu
785                 790                 795
```

[0026]

<210> 27
<211> 2400
<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> 27

```
atgccaacac tgcccagaac atttgacgac attcagtccc gactgattaa cgccacctcc      60
agggtggtgc cgatgcagag gcaaattcag ggactgaaat tcttaatgag cgccaagagg     120
aagaccttcg gcccacgccg accgatgccc gaattcgttg aaacacccat cccggacgtt     180
aacacgctgg cccttgagga catcgatgtc agcaatccgt ttttataccg gcagggtcag     240
tggcgcgcct atttcaaacg gttgcgtgat gaggcgccgg tccattacca gaagaacagc     300
cctttcggcc ccttctggtc ggtaactcgg tttgaagaca tcctgttcgt ggataagagt     360
cacgacctgt tttccgccga gccgcaaatc attctcggtg accctccgga ggggctgtcg     420
gtggaaatgt tcatagcgat ggatccgccg aaacacgatg tgcagcgcag ctcggtgcag     480
ggagtagtgg caccgaaaaa cctgaaggag atggaggggc tgatccgatc acgcaccggc     540
gatgtgcttg acagcctgcc tacagacaaa ccctttaact gggtacctgc tgtttccaag     600
gaactcacag gccgcatgct ggcgacgctt ctggattttc cttacgagga acgccacaag     660
ctggttgagt ggtcggacag aatggcaggt gcggcatcgg ccaccggcgg ggagtttgcc     720
gatgaaaatg ccatgtttga cgacgcggca gacatggccc ggtctttctc caggctttgg     780
cgggacaagg aggcgcgccg cgcagcaggc gaggagcccg gtttcgattt gatcagcctg     840
ttgcagagca acaaagaaac gaaagacctg atcaatcggc cgatggagtt tatcggtaat     900
ttgacgctgc tcatagtcgc cggcaacgat acgacgcgca actcgatgag tggtggcctg     960
gtggccatga acgaattccc cagggaattt gaaaaattga aggcaaaacc ggagttgatt    1020
ccgaacatgg tgtcggaaat catccgctgg caaacgccgc tggcctatat gcgccgaatc    1080
gccaagcagg atgtcgaact gggcggccag accatcaaga agggtgatcg agttgtcatg    1140
tggtacgcgt cgggtaaccg ggacgagcgc aaatttgaca accccgatca gttcatcatt    1200
gatcgcaagg acgcacgaaa ccacatgtcg ttcggctatg gggttcaccg ttgcatgggc    1260
```

```
aaccgtctgg ctgaactgca actgcgcatc ctctgggaag aaatactcaa gcgttttgac     1320
aacatcgaag tcgtcgaaga gcccgagcgg gtgcagtcca acttcgtgcg gggctattcc     1380
aggttgatgg tcaaactgac accgaacagt gtactccatc gtcatcaacc tgtcaccatc     1440
ggcgagccgg ccgctcgtgc tgtgagccgc acggtgaccg ttgagcgtct tgatcgcatt     1500
gccgacgatg tccttcgcct ggtccttcgc gatgctggag gtaaaaccct cccgacgtgg     1560
acgcctggcg ctcacatcga cctggatctg ggtgctctga gccgtcagta ttcgctctgc     1620
ggcgctccgg atgctccgtc gtacgaaatc gccgtgcact tagatccgga aagccgtggt     1680
ggaagccgct atattcatga acagctggaa gttggaagtc cgctgcgtat gcgtggccca     1740
cgcaaccatt tcgccctgga tccgggtgcg gaacattacg tgtttgttgc cgggggtatc     1800
ggcatcacgc cggtgctggc aatggcggat catgcccgtg cgcgtggttg gtcgtacgaa     1860
ctgcattatt gtggtcgtaa tcgtagcggt atggcttacc tggaacgcgt cgcgggacat     1920
ggtgaccgcg ctgccttgca cgtatctgaa gaaggcaccc gcattgatct ggcggcatta     1980
cttgctgaac cggcgccggg cgtgcaaatc tacgcctgcg gtccgggccg tttattagcg     2040
ggtcttgaag acgcgtctcg taattggccg gatggcgcgc ttcatgtgga gcatttcact     2100
tcgagtttag ccgctttgga tccggatgtc gaacatgcct ttgatttgga gctgcgtgac     2160
tctggcctta ccgttcgcgt cgagccaact cagaccgttt tagacgcttt gcgtgcgaac     2220
aatatcgacg tcccgtcgga ttgcgaagag gggctgtgtg gttcttgcga agtagccgtt     2280
ctggatggcg aggttgatca ccgtgatacc gttctgacta aggccgagcg cgccgcgaat     2340
cgtcagatga tgacttgctg cagtcgtgca tgcggtgatc gtctggcgct gcgcctctaa     2400
```

<210> 28
<211> 799
<212> PRT
<213> 人工序列
<220>
<221> 来源
<223> /注="人工序列的描述：合成的多肽"
<400> 28

[0027]

```
Met Pro Thr Leu Pro Arg Thr Phe Asp Asp Ile Gln Ser Arg Leu Ile
1               5                  10                  15
Asn Ala Thr Ser Arg Val Val Pro Met Gln Arg Gln Ile Gln Gly Leu
            20                  25                  30
Lys Phe Leu Met Ser Ala Lys Arg Lys Thr Phe Gly Pro Arg Arg Pro
        35                  40                  45
Met Pro Glu Phe Val Glu Thr Pro Ile Pro Asp Val Asn Thr Leu Ala
    50                  55                  60
Leu Glu Asp Ile Asp Val Ser Asn Pro Phe Leu Tyr Arg Gln Gly Gln
65                  70                  75                  80
Trp Arg Ala Tyr Phe Lys Arg Leu Arg Asp Glu Ala Pro Val His Tyr
                85                  90                  95
Gln Lys Asn Ser Pro Phe Gly Pro Phe Trp Ser Val Thr Arg Phe Glu
            100                 105                 110
Asp Ile Leu Phe Val Asp Lys Ser His Asp Leu Phe Ser Ala Glu Pro
        115                 120                 125
Gln Ile Ile Leu Gly Asp Pro Pro Glu Gly Leu Ser Val Glu Met Phe
    130                 135                 140
Ile Ala Met Asp Pro Pro Lys His Asp Val Gln Arg Ser Ser Val Gln
145                 150                 155                 160
Gly Val Val Ala Pro Lys Asn Leu Lys Glu Met Glu Gly Leu Ile Arg
                165                 170                 175
Ser Arg Thr Gly Asp Val Leu Asp Ser Leu Pro Thr Asp Lys Pro Phe
            180                 185                 190
Asn Trp Val Pro Ala Val Ser Lys Glu Leu Thr Gly Arg Met Leu Ala
        195                 200                 205
Thr Leu Leu Asp Phe Pro Tyr Glu Glu Arg His Lys Leu Val Glu Trp
    210                 215                 220
Ser Asp Arg Met Ala Gly Ala Ala Ser Ala Thr Gly Gly Glu Phe Ala
225                 230                 235                 240
Asp Glu Asn Ala Met Phe Asp Asp Ala Ala Asp Met Ala Arg Ser Phe
                245                 250                 255
Ser Arg Leu Trp Arg Asp Lys Glu Ala Arg Arg Ala Ala Gly Glu Glu
            260                 265                 270
Pro Gly Phe Asp Leu Ile Ser Leu Leu Gln Ser Asn Lys Glu Thr Lys
        275                 280                 285
Asp Leu Ile Asn Arg Pro Met Glu Phe Ile Gly Asn Leu Thr Leu Leu
    290                 295                 300
Ile Val Ala Gly Asn Asp Thr Thr Arg Asn Ser Met Ser Gly Gly Leu
305                 310                 315                 320
Val Ala Met Asn Glu Phe Pro Arg Glu Phe Glu Lys Leu Lys Ala Lys
                325                 330                 335
Pro Glu Leu Ile Pro Asn Met Val Ser Glu Ile Ile Arg Trp Gln Thr
            340                 345                 350
Pro Leu Ala Tyr Met Arg Arg Ile Ala Lys Gln Asp Val Glu Leu Gly
        355                 360                 365
Gly Gln Thr Ile Lys Lys Gly Asp Arg Val Val Met Trp Tyr Ala Ser
    370                 375                 380
```

```
Gly Asn Arg Asp Glu Arg Lys Phe Asp Asn Pro Asp Gln Phe Ile Ile
385                 390                 395                 400
Asp Arg Lys Asp Ala Arg Asn His Met Ser Phe Gly Tyr Gly Val His
                405                 410                 415
Arg Cys Met Gly Asn Arg Leu Ala Glu Leu Gln Leu Arg Ile Leu Trp
            420                 425                 430
Glu Glu Ile Leu Lys Arg Phe Asp Asn Ile Glu Val Val Glu Glu Pro
        435                 440                 445
Glu Arg Val Gln Ser Asn Phe Val Arg Gly Tyr Ser Arg Leu Met Val
    450                 455                 460
Lys Leu Thr Pro Asn Ser Val Leu His Arg His Gln Pro Val Thr Ile
465                 470                 475                 480
Gly Glu Pro Ala Ala Arg Ala Val Ser Arg Thr Val Thr Val Glu Arg
                485                 490                 495
Leu Asp Arg Ile Ala Asp Asp Val Leu Arg Leu Val Leu Arg Asp Ala
            500                 505                 510
Gly Gly Lys Thr Leu Pro Thr Trp Thr Pro Gly Ala His Ile Asp Leu
        515                 520                 525
Asp Leu Gly Ala Leu Ser Arg Gln Tyr Ser Leu Cys Gly Ala Pro Asp
    530                 535                 540
Ala Pro Ser Tyr Glu Ile Ala Val His Leu Asp Pro Glu Ser Arg Gly
545                 550                 555                 560
Gly Ser Arg Tyr Ile His Glu Gln Leu Glu Val Gly Ser Pro Leu Arg
                565                 570                 575
Met Arg Gly Pro Arg Asn His Phe Ala Leu Asp Pro Gly Ala Glu His
            580                 585                 590
Tyr Val Phe Val Ala Gly Gly Ile Gly Ile Thr Pro Val Leu Ala Met
        595                 600                 605
Ala Asp His Ala Arg Ala Arg Gly Trp Ser Tyr Glu Leu His Tyr Cys
    610                 615                 620
Gly Arg Asn Arg Ser Gly Met Ala Tyr Leu Glu Arg Val Ala Gly His
625                 630                 635                 640
Gly Asp Arg Ala Ala Leu His Val Ser Glu Glu Gly Thr Arg Ile Asp
                645                 650                 655
Leu Ala Ala Leu Leu Ala Glu Pro Ala Pro Gly Val Gln Ile Tyr Ala
            660                 665                 670
Cys Gly Pro Gly Arg Leu Leu Ala Gly Leu Glu Asp Ala Ser Arg Asn
        675                 680                 685
Trp Pro Asp Gly Ala Leu His Val Glu His Phe Thr Ser Ser Leu Ala
    690                 695                 700
Ala Leu Asp Pro Asp Val Glu His Ala Phe Asp Leu Glu Leu Arg Asp
705                 710                 715                 720
Ser Gly Leu Thr Val Arg Val Glu Pro Thr Gln Thr Val Leu Asp Ala
                725                 730                 735
Leu Arg Ala Asn Asn Ile Asp Val Pro Ser Asp Cys Glu Glu Gly Leu
            740                 745                 750
Cys Gly Ser Cys Glu Val Ala Val Leu Asp Gly Glu Val Asp His Arg
        755                 760                 765
Asp Thr Val Leu Thr Lys Ala Glu Arg Ala Ala Asn Arg Gln Met Met
    770                 775                 780
Thr Cys Cys Ser Arg Ala Cys Gly Asp Arg Leu Ala Leu Arg Leu
785                 790                 795
```

[0028]

<210> 29
<211> 2400
<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> 29

```
atgccaacac tgcccagaac atttgacgac attcagtccc gactgattaa cgccacctcc      60
agggtggtgc cgatgcagag gcaaattcag ggactgaaat tcttaatgag cgccaagagg     120
aagaccttcg gcccacgccg accgatgccc gaattcgttg aaacacccat cccggacgtt     180
aacacgctgg cccttgagga catcgatgtc agcaatccgt ttttataccg gcagggtcag     240
tggcgcgcct atttcaaacg gttgcgtgat gaggcgccgg tccattacca gaagaacagc     300
cctttcggcc ccttctggtc ggtaactcgg tttgaagaca tcctgttcgt ggataagagt     360
cacgacctgt tttccgccga gccgcaaatc attctcggtg accctccgga ggggctgtcg     420
gtggaaatgt tcatagcgat ggatccgccg aaacacgatg tgcagcgcag ctcggtgcag     480
ggagtagtgg caccgaaaaa cctgaaggag atggaggggc tgatccgatc acgcaccggc     540
gatgtgcttg acagcctgcc tacagacaaa ccctttaact gggtacctgc tgtttccaag     600
gaactcacag gccgcatgct ggcgacgctt ctggattttc cttacgagga acgccacaag     660
ctggttgagt ggtcggacag aatggcaggt gcagcatcgg ccaccggcgg ggagtttgcc     720
gatgaaaatg ccatgtttga cgacgcggca gacatggccc ggtctttctc caggctttgg     780
cgggacaagg aggcgcgccg cgcagcaggc gaggagcccg gtttcgattt gatcagcctg     840
ttgcagagca acaaagaaac gaaagacctg atcaatcggc cgatggagtt tatcggtaat     900
```

```
ttgacgctgc tcatagtcgc cggcaacgat acgacgcgca actcgatgag tggtggcctg      960
gtggccatga acgaattccc cagggaattt gaaaaattga aggcaaaacc ggagttgatt     1020
ccgaacatgg tgtcggaaat catccgctgg caaacgccgc tggcctatat gcgccgaatc     1080
gccaagcagg atgtcgaact gggcggccag accatcaaga agggtgatcg agttgtcatg     1140
tggtacgcgt cgggtaaccg ggacgagcgc aaatttgaca accccgatca gttcatcatt     1200
gatcgcaagg acgcacgaaa ccacatgtcg ttcggctatg ggcgtcaccg ttgcatgggc     1260
aaccgtctgg ctgaactgca actgcgcatc ctctgggaag aaatactcaa gcgttttgac     1320
aacatcgaag tcgtcgaaga gcccgagcgg gtgcagtcca acttcgtgcg gggctattcc     1380
aggttgatgg tcaaactgac accgaacagt gtactccatc gtcatcaacc tgtcaccatc     1440
ggcgagccgg ccgctcgtgc tgtgagccgc acggtgaccg ttgagcgtct tgatcgcatt     1500
gccgacgatg tccttcgcct ggtccttcgc gatgctggag gtaaaaccct cccgacgtgg     1560
acgcctggcg ctcacatcga cctggatctg ggtgctctga gccgtcagta ttcgctctgc     1620
ggcgctccgg atgctccgtc gtacgaaatc gccgtgcact tagatccgga aagccgtggt     1680
ggaagccgct atattcatga acagctggaa gttggaagtc cgctgcgtat gcgtggccca     1740
cgcaaccatt tcgccctgga tccgggtgcg gaacattacg tgtttgttgc cgggggtatc     1800
ggcatcacgc cggtgctggc aatggcggat catgcccgtg cgcgtggttg gtcgtacgaa     1860
ctgcattatt gtggtcgtaa tcgtagcggt atggcttacc tggaacgcgt cgcgggacat     1920
ggtgaccgcg ctgccttgca cgtatctgaa gaaggcaccc gcattgatct ggcggcatta     1980
cttgctgaac cggcgccggg cgtgcaaatc tacgcctgcg gtccgggccg tttattagcg     2040
ggtcttgaag acgcgtctcg taattggccg gatggcgcgc ttcatgtgga gcatttcact     2100
tcgagtttag ccgctttgga tccggatgtc gaacatgcct ttgatttgga gctgcgtgac     2160
tctggcctta ccgttcgcgt cgagccaact cagaccgttt tagacgcttt gcgtgcgaac     2220
aatatcgacg tcccgtcgga ttgcgaagag gggctgtgtg gttcttgcga agtagccgtt     2280
ctggatggcg aggttgatca ccgtgatacc gttctgacta aggccgagcg cgccgcgaat     2340
cgtcagatga tgacttgctg cagtcgtgca tgcggtgatc gtctggcgct gcgcctctaa     2400
```

<210> 30
<211> 799
<212> PRT
<213> 人工序列
<220>
<221> 来源
<223> /注="人工序列的描述：合成的多肽"
<400> 30

[0029]

```
Met Pro Thr Leu Pro Arg Thr Phe Asp Asp Ile Gln Ser Arg Leu Ile
1               5                  10                  15
Asn Ala Thr Ser Arg Val Val Pro Met Gln Arg Gln Ile Gln Gly Leu
            20                  25                  30
Lys Phe Leu Met Ser Ala Lys Arg Lys Thr Phe Gly Pro Arg Arg Pro
        35                  40                  45
Met Pro Glu Phe Val Glu Thr Pro Ile Pro Asp Val Asn Thr Leu Ala
    50                  55                  60
Leu Glu Asp Ile Asp Val Ser Asn Pro Phe Leu Tyr Arg Gln Gly Gln
65                  70                  75                  80
Trp Arg Ala Tyr Phe Lys Arg Leu Arg Asp Glu Ala Pro Val His Tyr
                85                  90                  95
Gln Lys Asn Ser Pro Phe Gly Pro Phe Trp Ser Val Thr Arg Phe Glu
            100                 105                 110
Asp Ile Leu Phe Val Asp Lys Ser His Asp Leu Phe Ser Ala Glu Pro
        115                 120                 125
Gln Ile Ile Leu Gly Asp Pro Pro Glu Gly Leu Ser Val Glu Met Phe
    130                 135                 140
Ile Ala Met Asp Pro Pro Lys His Asp Val Gln Arg Ser Ser Val Gln
145                 150                 155                 160
Gly Val Val Ala Pro Lys Asn Leu Lys Glu Met Glu Gly Leu Ile Arg
                165                 170                 175
Ser Arg Thr Gly Asp Val Leu Asp Ser Leu Pro Thr Asp Lys Pro Phe
            180                 185                 190
Asn Trp Val Pro Ala Val Ser Lys Glu Leu Thr Gly Arg Met Leu Ala
        195                 200                 205
Thr Leu Leu Asp Phe Pro Tyr Glu Glu Arg His Lys Leu Val Glu Trp
    210                 215                 220
Ser Asp Arg Met Ala Gly Ala Ala Ser Ala Thr Gly Gly Glu Phe Ala
225                 230                 235                 240
Asp Glu Asn Ala Met Phe Asp Asp Ala Ala Asp Met Ala Arg Ser Phe
                245                 250                 255
Ser Arg Leu Trp Arg Asp Lys Glu Ala Arg Arg Ala Ala Gly Glu Glu
            260                 265                 270
Pro Gly Phe Asp Leu Ile Ser Leu Leu Gln Ser Asn Lys Glu Thr Lys
        275                 280                 285
Asp Leu Ile Asn Arg Pro Met Glu Phe Ile Gly Asn Leu Thr Leu Leu
    290                 295                 300
Ile Val Ala Gly Asn Asp Thr Thr Arg Asn Ser Met Ser Gly Gly Leu
305                 310                 315                 320
Val Ala Met Asn Glu Phe Pro Arg Glu Phe Glu Lys Leu Lys Ala Lys
                325                 330                 335
```

```
Pro Glu Leu Ile Pro Asn Met Val Ser Glu Ile Ile Arg Trp Gln Thr
            340                 345                 350
Pro Leu Ala Tyr Met Arg Arg Ile Ala Lys Gln Asp Val Glu Leu Gly
        355                 360                 365
Gly Gln Thr Ile Lys Lys Gly Asp Arg Val Val Met Trp Tyr Ala Ser
    370                 375                 380
Gly Asn Arg Asp Glu Arg Lys Phe Asp Asn Pro Asp Gln Phe Ile Ile
385                 390                 395                 400
Asp Arg Lys Asp Ala Arg Asn His Met Ser Phe Gly Tyr Gly Arg His
                405                 410                 415
Arg Cys Met Gly Asn Arg Leu Ala Glu Leu Gln Leu Arg Ile Leu Trp
            420                 425                 430
Glu Glu Ile Leu Lys Arg Phe Asp Asn Ile Glu Val Val Glu Glu Pro
        435                 440                 445
Glu Arg Val Gln Ser Asn Phe Val Arg Gly Tyr Ser Arg Leu Met Val
    450                 455                 460
Lys Leu Thr Pro Asn Ser Val Leu His Arg His Gln Pro Val Thr Ile
465                 470                 475                 480
Gly Glu Pro Ala Ala Arg Ala Val Ser Arg Thr Val Thr Val Glu Arg
                485                 490                 495
Leu Asp Arg Ile Ala Asp Asp Val Leu Arg Leu Val Leu Arg Asp Ala
            500                 505                 510
Gly Gly Lys Thr Leu Pro Thr Trp Thr Pro Gly Ala His Ile Asp Leu
        515                 520                 525
Asp Leu Gly Ala Leu Ser Arg Gln Tyr Ser Leu Cys Gly Ala Pro Asp
    530                 535                 540
Ala Pro Ser Tyr Glu Ile Ala Val His Leu Asp Pro Glu Ser Arg Gly
545                 550                 555                 560
Gly Ser Arg Tyr Ile His Glu Gln Leu Glu Val Gly Ser Pro Leu Arg
                565                 570                 575
Met Arg Gly Pro Arg Asn His Phe Ala Leu Asp Pro Gly Ala Glu His
            580                 585                 590
Tyr Val Phe Val Ala Gly Gly Ile Gly Ile Thr Pro Val Leu Ala Met
        595                 600                 605
Ala Asp His Ala Arg Ala Arg Gly Trp Ser Tyr Glu Leu His Tyr Cys
    610                 615                 620
Gly Arg Asn Arg Ser Gly Met Ala Tyr Leu Glu Arg Val Ala Gly His
625                 630                 635                 640
Gly Asp Arg Ala Ala Leu His Val Ser Glu Glu Gly Thr Arg Ile Asp
                645                 650                 655
Leu Ala Ala Leu Leu Ala Glu Pro Ala Pro Gly Val Gln Ile Tyr Ala
            660                 665                 670
Cys Gly Pro Gly Arg Leu Leu Ala Gly Leu Glu Asp Ala Ser Arg Asn
        675                 680                 685
Trp Pro Asp Gly Ala Leu His Val Glu His Phe Thr Ser Ser Leu Ala
    690                 695                 700
Ala Leu Asp Pro Asp Val Glu His Ala Phe Asp Leu Glu Leu Arg Asp
705                 710                 715                 720
Ser Gly Leu Thr Val Arg Val Glu Pro Thr Gln Thr Val Leu Asp Ala
                725                 730                 735
Leu Arg Ala Asn Asn Ile Asp Val Pro Ser Asp Cys Glu Glu Gly Leu
            740                 745                 750
Cys Gly Ser Cys Glu Val Ala Val Leu Asp Gly Glu Val Asp His Arg
        755                 760                 765
Asp Thr Val Leu Thr Lys Ala Glu Arg Ala Ala Asn Arg Gln Met Met
    770                 775                 780
Thr Cys Cys Ser Arg Ala Cys Gly Asp Arg Leu Ala Leu Arg Leu
785                 790                 795
```

[0030]

<210> 31
<211> 2400
<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> 31

```
atgccaacac tgcccagaac atttgacgac attcagtccc gactgattaa cgccacctcc      60
agggtggtgc cgatgcagag gcaaattcag ggactgaaat tcttaatgag cgccaagagg     120
aagaccttcg gcccacgccg accgatgccc gaattcgttg aaacacccat cccggacgtt     180
aacacgctgg cccttgagga catcgatgtc agcaatccgt ttttataccg gcagggtcag     240
tggcgcgcct atttcaaacg gttgcgtgat gaggcgccgg tccattacca gaagaacagc     300
cctttcggcc ccttctggtc ggtaactcgg tttgaagaca tcctgttcgt ggataagagt     360
cacgacctgt tttccgccga gccgcaaatc attctcggtg accctccgga ggggctgtcg     420
atcgaaatgt tcatagcgat ggatccgccg aaacacgatg tgcagcgcag ctcggtgcag     480
ggagtagtgg caccgaaaaa cctgaaggag atggaggggc tgatccgatc acgcaccggc     540
```

```
gatgtgcttg acagcctgcc tacagacaaa ccctttaact gggtacctgc tgtttccaag      600
gaactcacag gccgcatgct ggcgacgctt ctggattttc cttacgagga acgccacaag      660
ctggttgagt ggtcggacag aatggcaggt acagcatcgg ccaccggcgg ggagtttgcc      720
gatgaaaatg ccatgtttga cgacgcggca gacatggccc ggtctttctc caggctttgg      780
cgggacaagg aggcgcgccg cgcagcaggc gaggagcccg gtttcgattt gatcagcctg      840
ttgcagagca acaaagaaac gaaagacctg atcaatcggc cgatggagtt tatcggtaat      900
ttgacgctgc tcatagtcgc cggcaacgat acgacgcgca actcgatgag tggtggcctg      960
gtggccatga acgaattccc cagggaattt gaaaaattga aggcaaaacc ggagttgatt     1020
ccgaacatgg tgtcggaaat catccgctgg caaacgccgc tggcctatat gcgccgaatc     1080
gccaagcagg atgtcgaact gggcggccag accatcaaga agggtgatcg agttgtcatg     1140
tggtacgcgt cgggtaaccg ggacgagcgc aaatttgaca accccgatca gttcatcatt     1200
gatcgcaagg acgcacgaaa ccacatgtcg ttcggctatg gggttcaccg ttgcatgggc     1260
aaccgtctgg ctgaactgca actgcgcatc ctctgggaag aaatactcaa gcgttttgac     1320
aacatcgaag tcgtcgaaga gcccgagcgg gtgcagtcca acttcgtgcg gggctattcc     1380
aggttgatgg tcaaactgac accgaacagt gtactccatc gtcatcaacc tgtcaccatc     1440
ggcgagccgg ccgctcgtgc tgtgagccgc acggtgaccg ttgagcgtct tgatcgcatt     1500
gccgacgatg tccttcgcct ggtccttcgc gatgctggag gtaaaaccct cccgacgtgg     1560
acgcctggcg ctcacatcga cctggatctg ggtgctctga gccgtcagta ttcgctctgc     1620
ggcgctccgg atgctccgtc gtacgaaatc gccgtgcact tagatccgga aagccgtggt     1680
ggaagccgct atattcatga acagctggaa gttggaagtc cgctgcgtat gcgtggccca     1740
cgcaaccatt tcgccctgga tccgggtgcg gaacattacg tgtttgttgc cgggggtatc     1800
ggcatcacgc cggtgctggc aatggcggat catgcccgtg cgcgtggttg gtcgtacgaa     1860
ctgcattatt gtggtcgtaa tcgtagcggt atggcttacc tggaacgcgt cgcgggacat     1920
ggtgaccgcg ctgccttgca cgtatctgaa gaaggcaccc gcattgatct ggcggcatta     1980
cttgctgaac cggcgccggg cgtgcaaatc tacgcctgcg gtccgggccg tttattagcg     2040
ggtcttgaag acgcgtctcg taattggccg gatggcgcgc ttcatgtgga gcatttcact     2100
tcgagtttag ccgctttgga tccggatgtc gaacatgcct ttgatttgga gctgcgtgac     2160
tctggcctta ccgttcgcgt cgagccaact cagaccgttt tagacgcttt gcgtgcgaac     2220
aatatcgacg tcccgtcgga ttgcgaagag gggctgtgtg gttcttgcga agtagccgtt     2280
ctggatggcg aggttgatca ccgtgatacc gttctgacta aggccgagcg cgccgcgaat     2340
cgtcagatga tgacttgctg cagtcgtgca tgcggtgatc gtctggcgct gcgcctctaa     2400
```

<210> 32
<211> 799
<212> PRT
<213> 人工序列
<220>
[0031] <221> 来源
<223> /注="人工序列的描述：合成的多肽"
<400> 32

```
Met Pro Thr Leu Pro Arg Thr Phe Asp Asp Ile Gln Ser Arg Leu Ile
1               5                  10                  15
Asn Ala Thr Ser Arg Val Val Pro Met Gln Arg Gln Ile Gln Gly Leu
            20                  25                  30
Lys Phe Leu Met Ser Ala Lys Arg Lys Thr Phe Gly Pro Arg Arg Pro
        35                  40                  45
Met Pro Glu Phe Val Glu Thr Pro Ile Pro Asp Val Asn Thr Leu Ala
    50                  55                  60
Leu Glu Asp Ile Asp Val Ser Asn Pro Phe Leu Tyr Arg Gln Gly Gln
65                  70                  75                  80
Trp Arg Ala Tyr Phe Lys Arg Leu Arg Asp Glu Ala Pro Val His Tyr
                85                  90                  95
Gln Lys Asn Ser Pro Phe Gly Pro Phe Trp Ser Val Thr Arg Phe Glu
            100                 105                 110
Asp Ile Leu Phe Val Asp Lys Ser His Asp Leu Phe Ser Ala Glu Pro
        115                 120                 125
Gln Ile Ile Leu Gly Asp Pro Pro Glu Gly Leu Ser Ile Glu Met Phe
    130                 135                 140
Ile Ala Met Asp Pro Pro Lys His Asp Val Gln Arg Ser Ser Val Gln
145                 150                 155                 160
Gly Val Val Ala Pro Lys Asn Leu Lys Glu Met Glu Gly Leu Ile Arg
                165                 170                 175
Ser Arg Thr Gly Asp Val Leu Asp Ser Leu Pro Thr Asp Lys Pro Phe
            180                 185                 190
Asn Trp Val Pro Ala Val Ser Lys Glu Leu Thr Gly Arg Met Leu Ala
        195                 200                 205
Thr Leu Leu Asp Phe Pro Tyr Glu Glu Arg His Lys Leu Val Glu Trp
    210                 215                 220
Ser Asp Arg Met Ala Gly Thr Ala Ser Ala Thr Gly Gly Glu Phe Ala
225                 230                 235                 240
Asp Glu Asn Ala Met Phe Asp Asp Ala Ala Asp Met Ala Arg Ser Phe
                245                 250                 255
Ser Arg Leu Trp Arg Asp Lys Glu Ala Arg Arg Ala Ala Gly Glu Glu
            260                 265                 270
Pro Gly Phe Asp Leu Ile Ser Leu Leu Gln Ser Asn Lys Glu Thr Lys
        275                 280                 285
```

```
Asp Leu Ile Asn Arg Pro Met Glu Phe Ile Gly Asn Leu Thr Leu Leu
    290                 295                 300
Ile Val Ala Gly Asn Asp Thr Thr Arg Asn Ser Met Ser Gly Gly Leu
305                 310                 315                 320
Val Ala Met Asn Glu Phe Pro Arg Glu Phe Glu Lys Leu Lys Ala Lys
                325                 330                 335
Pro Glu Leu Ile Pro Asn Met Val Ser Glu Ile Ile Arg Trp Gln Thr
            340                 345                 350
Pro Leu Ala Tyr Met Arg Arg Ile Ala Lys Gln Asp Val Glu Leu Gly
        355                 360                 365
Gly Gln Thr Ile Lys Lys Gly Asp Arg Val Val Met Trp Tyr Ala Ser
    370                 375                 380
Gly Asn Arg Asp Glu Arg Lys Phe Asp Asn Pro Asp Gln Phe Ile Ile
385                 390                 395                 400
Asp Arg Lys Asp Ala Arg Asn His Met Ser Phe Gly Tyr Gly Val His
                405                 410                 415
Arg Cys Met Gly Asn Arg Leu Ala Glu Leu Gln Leu Arg Ile Leu Trp
            420                 425                 430
Glu Glu Ile Leu Lys Arg Phe Asp Asn Ile Glu Val Val Glu Glu Pro
        435                 440                 445
Glu Arg Val Gln Ser Asn Phe Val Arg Gly Tyr Ser Arg Leu Met Val
    450                 455                 460
Lys Leu Thr Pro Asn Ser Val Leu His Arg His Gln Pro Val Thr Ile
465                 470                 475                 480
Gly Glu Pro Ala Ala Arg Ala Val Ser Arg Thr Val Thr Val Glu Arg
                485                 490                 495
Leu Asp Arg Ile Ala Asp Asp Val Leu Arg Leu Val Leu Arg Asp Ala
            500                 505                 510
Gly Gly Lys Thr Leu Pro Thr Trp Thr Pro Gly Ala His Ile Asp Leu
        515                 520                 525
Asp Leu Gly Ala Leu Ser Arg Gln Tyr Ser Leu Cys Gly Ala Pro Asp
    530                 535                 540
Ala Pro Ser Tyr Glu Ile Ala Val His Leu Asp Pro Glu Ser Arg Gly
545                 550                 555                 560
Gly Ser Arg Tyr Ile His Glu Gln Leu Glu Val Gly Ser Pro Leu Arg
                565                 570                 575
Met Arg Gly Pro Arg Asn His Phe Ala Leu Asp Pro Gly Ala Glu His
            580                 585                 590
Tyr Val Phe Val Ala Gly Gly Ile Gly Ile Thr Pro Val Leu Ala Met
        595                 600                 605
Ala Asp His Ala Arg Ala Arg Gly Trp Ser Tyr Glu Leu His Tyr Cys
    610                 615                 620
Gly Arg Asn Arg Ser Gly Met Ala Tyr Leu Glu Arg Val Ala Gly His
625                 630                 635                 640
Gly Asp Arg Ala Ala Leu His Val Ser Glu Glu Gly Thr Arg Ile Asp
                645                 650                 655
Leu Ala Ala Leu Leu Ala Glu Pro Ala Pro Gly Val Gln Ile Tyr Ala
            660                 665                 670
Cys Gly Pro Gly Arg Leu Leu Ala Gly Leu Glu Asp Ala Ser Arg Asn
        675                 680                 685
Trp Pro Asp Gly Ala Leu His Val Glu His Phe Thr Ser Ser Leu Ala
    690                 695                 700
Ala Leu Asp Pro Asp Val Glu His Ala Phe Asp Leu Glu Leu Arg Asp
705                 710                 715                 720
Ser Gly Leu Thr Val Arg Val Glu Pro Thr Gln Thr Val Leu Asp Ala
                725                 730                 735
Leu Arg Ala Asn Asn Ile Asp Val Pro Ser Asp Cys Glu Glu Gly Leu
            740                 745                 750
Cys Gly Ser Cys Glu Val Ala Val Leu Asp Gly Glu Val Asp His Arg
        755                 760                 765
Asp Thr Val Leu Thr Lys Ala Glu Arg Ala Ala Asn Arg Gln Met Met
    770                 775                 780
Thr Cys Cys Ser Arg Ala Cys Gly Asp Arg Leu Ala Leu Arg Leu
785                 790                 795
```

[0032]

<210> 33
<211> 2400
<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> 33

```
atgccaacac tgcccagaac atttgacgac attcagtccc gactgattaa cgccacctcc      60
agggtggtgc cgatgcagct gcaaattcag ggactgaaat tcttaatgag cgccaagagg     120
aagaccttcg gcccacgccg accgatgccc gaattcgttg aaacacccat cccggacgtt     180
```

```
aacacgctgg cccttgagga catcgatgtc agcaatccgt ttttataccg gcagggtcag      240
tgggacgcct atttcaaacg gttgcgtgat gaggcgccgg tccattacca gaagaacagc      300
cctttcggcc ccttctggtc ggtaactcgg tttgaagaca tcctgttcgt ggataagagt      360
cacgacctgt tttccgccga gccgcaaatc attctcggtg accctccgga ggggctgtcg      420
atggaaatgt tcatagcgat ggatccgccg aaacacgatg tgcagcgcag ctcggtgcag      480
ggagtagtgg caccgaaaaa cctgaaggag atggaggggc tgatccgatc aaacaccggc      540
gatgtgcttg acagcctgcc tacagacaaa ccctttaact gggtacctgc tgtttccaag      600
gaactcacag gccgcatgct ggcgacgctt ctggattttc cttacgagga acgccacaag      660
ctggttgagt ggtcggacag aatggcaggt gcagcatcgg ccaccggcgg ggagtttgcc      720
gatgaaaatg ccatgtttga cgacgcggca gacatggccc ggtctttctc caggctttgg      780
cgggacaagg aggcgcgccg cgcagcaggc gaggagcccg gtttcgattt gatcagcctg      840
ttgcagagca acaaagaaac gaaagacctg atcaatcggc cgatggagtt tatcggtaat      900
ttgacgctgc tcatagtcgc cggcgcggat acgacgcgca actcgatgag tggtggcctg      960
gtggccatga acgaattccc cagggaattt gaaaaattga aggcaaaacc ggagttgatt     1020
ccgaacatgg tgtcggaaat catccgctgg caaacgccgc tggcctatat gcgccgaatc     1080
gccaagcagg atgtcgaact gggcggccag accatcaaga agggtgatcg agttgtcatg     1140
tggtacgcgt cgggtaaccg ggacgagcgc aaatttgaca accccgatca gttcatcatt     1200
gatcgcaagg acgcacgaaa ccacatgtcg ttcggctatg gggttcaccg ttgcatgggc     1260
aaccgtctgg ctgaactgca actgcgcatc ctctgggaag aaatactcaa gcgttttgac     1320
aacatcgaag tcgtcgaaga gcccgagcgg gtgcagtcca acttcgtgcg gggctattcc     1380
aggttgatgg tcaaactgac accgaacagt gtactccatc gtcatcaacc tgtcaccatc     1440
ggcgagccgg ccgctcgtgc tgtgagccgc acggtgaccg ttgagcgtct tgatcgcatt     1500
gccgacgatg tccttcgcct ggtccttcgc gatgctggag gtaaaaccct cccgacgtgg     1560
acgcctggcg ctcacatcga cctggatctg ggtgctctga gccgtcagta ttcgctctgc     1620
ggcgctccgg atgctccgtc gtacgaaatc gccgtgcact tagatccgga aagccgtggt     1680
ggaagccgct atattcatga acagctggaa gttggaagtc cgctgcgtat gcgtggccca     1740
cgcaaccatt tcgccctgga tccgggtgcg gaacattacg tgtttgttgc cgggggtatc     1800
ggcatcacgc cggtgctggc aatggcggat catgcccgtg cgcgtggttg gtcgtacgaa     1860
ctgcattatt gtggtcgtaa tcgtagcggt atggcttacc tggaacgcgt cgcgggacat     1920
ggtgaccgcg ctgccttgca cgtatctgaa gaaggcaccc gcattgatct ggcggcatta     1980
cttgctgaac cggcgccggg cgtgcaaatc tacgcctgcg gtccgggccg tttattagcg     2040
ggtcttgaag acgcgtctcg taattggccg gatggcgcgc ttcatgtgga gcatttcact     2100
tcgagtttag ccgctttgga tccggatgtc gaacatgcct ttgatttgga gctgcgtgac     2160
tctggcctta ccgttcgcgt cgagccaact cagaccgttt tagacgcttt gcgtgcgaac     2220
aatatcgacg tcccgtcgga ttgcgaagag gggctgtgtg gttcttgcga agtagccgtt     2280
ctggatggcg aggttgatca ccgtgatacc gttctgacta aggccgagcg cgccgcgaat     2340
cgtcagatga tgacttgctg cagtcgtgca tgcggtgatc gtctggcgct gcgcctctaa     2400
```

[0033]

<210> 34
<211> 799
<212> PRT
<213> 人工序列
<220>
<221> 来源
<223> /注="人工序列的描述：合成的多肽"
<400> 34

```
Met Pro Thr Leu Pro Arg Thr Phe Asp Asp Ile Gln Ser Arg Leu Ile
1               5                  10                  15
Asn Ala Thr Ser Arg Val Val Pro Met Gln Leu Gln Ile Gln Gly Leu
            20                  25                  30
Lys Phe Leu Met Ser Ala Lys Arg Lys Thr Phe Gly Pro Arg Arg Pro
        35                  40                  45
Met Pro Glu Phe Val Glu Thr Pro Ile Pro Asp Val Asn Thr Leu Ala
    50                  55                  60
Leu Glu Asp Ile Asp Val Ser Asn Pro Phe Leu Tyr Arg Gln Gly Gln
65                  70                  75                  80
Trp Asp Ala Tyr Phe Lys Arg Leu Arg Asp Glu Ala Pro Val His Tyr
                85                  90                  95
Gln Lys Asn Ser Pro Phe Gly Pro Phe Trp Ser Val Thr Arg Phe Glu
            100                 105                 110
Asp Ile Leu Phe Val Asp Lys Ser His Asp Leu Phe Ser Ala Glu Pro
        115                 120                 125
Gln Ile Ile Leu Gly Asp Pro Pro Glu Gly Leu Ser Met Glu Met Phe
    130                 135                 140
Ile Ala Met Asp Pro Pro Lys His Asp Val Gln Arg Ser Ser Val Gln
145                 150                 155                 160
Gly Val Val Ala Pro Lys Asn Leu Lys Glu Met Glu Gly Leu Ile Arg
                165                 170                 175
Ser Asn Thr Gly Asp Val Leu Asp Ser Leu Pro Thr Asp Lys Pro Phe
            180                 185                 190
Asn Trp Val Pro Ala Val Ser Lys Glu Leu Thr Gly Arg Met Leu Ala
        195                 200                 205
Thr Leu Leu Asp Phe Pro Tyr Glu Glu Arg His Lys Leu Val Glu Trp
    210                 215                 220
Ser Asp Arg Met Ala Gly Ala Ala Ser Ala Thr Gly Gly Glu Phe Ala
225                 230                 235                 240
```

Asp Glu Asn Ala Met Phe Asp Asp Ala Ala Asp Met Ala Arg Ser Phe
245 250 255
Ser Arg Leu Trp Arg Asp Lys Glu Ala Arg Arg Ala Ala Gly Glu Glu
260 265 270
Pro Gly Phe Asp Leu Ile Ser Leu Leu Gln Ser Asn Lys Glu Thr Lys
275 280 285
Asp Leu Ile Asn Arg Pro Met Glu Phe Ile Gly Asn Leu Thr Leu Leu
290 295 300
Ile Val Ala Gly Ala Asp Thr Thr Arg Asn Ser Met Ser Gly Gly Leu
305 310 315 320
Val Ala Met Asn Glu Phe Pro Arg Glu Phe Glu Lys Leu Lys Ala Lys
325 330 335
Pro Glu Leu Ile Pro Asn Met Val Ser Glu Ile Ile Arg Trp Gln Thr
340 345 350
Pro Leu Ala Tyr Met Arg Arg Ile Ala Lys Gln Asp Val Glu Leu Gly
355 360 365
Gly Gln Thr Ile Lys Lys Gly Asp Arg Val Val Met Trp Tyr Ala Ser
370 375 380
Gly Asn Arg Asp Glu Arg Lys Phe Asp Asn Pro Asp Gln Phe Ile Ile
385 390 395 400
Asp Arg Lys Asp Ala Arg Asn His Met Ser Phe Gly Tyr Gly Val His
405 410 415
Arg Cys Met Gly Asn Arg Leu Ala Glu Leu Gln Leu Arg Ile Leu Trp
420 425 430
Glu Glu Ile Leu Lys Arg Phe Asp Asn Ile Glu Val Val Glu Glu Pro
435 440 445
Glu Arg Val Gln Ser Asn Phe Val Arg Gly Tyr Ser Arg Leu Met Val
450 455 460
Lys Leu Thr Pro Asn Ser Val Leu His Arg His Gln Pro Val Thr Ile
465 470 475 480
Gly Glu Pro Ala Ala Arg Ala Val Ser Arg Thr Val Thr Val Glu Arg
485 490 495
Leu Asp Arg Ile Ala Asp Asp Val Leu Arg Leu Val Leu Arg Asp Ala
500 505 510
Gly Gly Lys Thr Leu Pro Thr Trp Thr Pro Gly Ala His Ile Asp Leu
515 520 525
Asp Leu Gly Ala Leu Ser Arg Gln Tyr Ser Leu Cys Gly Ala Pro Asp
530 535 540
[0034] Ala Pro Ser Tyr Glu Ile Ala Val His Leu Asp Pro Glu Ser Arg Gly
545 550 555 560
Gly Ser Arg Tyr Ile His Glu Gln Leu Glu Val Gly Ser Pro Leu Arg
565 570 575
Met Arg Gly Pro Arg Asn His Phe Ala Leu Asp Pro Gly Ala Glu His
580 585 590
Tyr Val Phe Val Ala Gly Gly Ile Gly Ile Thr Pro Val Leu Ala Met
595 600 605
Ala Asp His Ala Arg Ala Arg Gly Trp Ser Tyr Glu Leu His Tyr Cys
610 615 620
Gly Arg Asn Arg Ser Gly Met Ala Tyr Leu Glu Arg Val Ala Gly His
625 630 635 640
Gly Asp Arg Ala Ala Leu His Val Ser Glu Glu Gly Thr Arg Ile Asp
645 650 655
Leu Ala Ala Leu Leu Ala Glu Pro Ala Pro Gly Val Gln Ile Tyr Ala
660 665 670
Cys Gly Pro Gly Arg Leu Leu Ala Gly Leu Glu Asp Ala Ser Arg Asn
675 680 685
Trp Pro Asp Gly Ala Leu His Val Glu His Phe Thr Ser Ser Leu Ala
690 695 700
Ala Leu Asp Pro Asp Val Glu His Ala Phe Asp Leu Glu Leu Arg Asp
705 710 715 720
Ser Gly Leu Thr Val Arg Val Glu Pro Thr Gln Thr Val Leu Asp Ala
725 730 735
Leu Arg Ala Asn Asn Ile Asp Val Pro Ser Asp Cys Glu Glu Gly Leu
740 745 750
Cys Gly Ser Cys Glu Val Ala Val Leu Asp Gly Glu Val Asp His Arg
755 760 765
Asp Thr Val Leu Thr Lys Ala Glu Arg Ala Ala Asn Arg Gln Met Met
770 775 780
Thr Cys Cys Ser Arg Ala Cys Gly Asp Arg Leu Ala Leu Arg Leu
785 790 795

<210> 35
<211> 2400
<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> 35

[0035]

```
atgccaacac tgcccagaac atttgacgac attcagtccc gactgattaa cgccacctcc      60
agggtggtgc cgatgcagag gcaaattcag ggactgaaat tcttaatgag cgccaagagg     120
aagaccttcg gcccacgccg accgatgccc gaattcgttg aaacacccat cccggacgtt     180
aacacgctgg cccttgagga catcgatgtc agcaatccgt ttttataccg gcagggtcag     240
tggcgcgcct atttcaaacg gttgcgtgat gaggcgccgg tccattacca gaagaacagc     300
cctttcggcc ccttctggtc ggtaactcgg tttgaagaca tcctgttcgt ggataagagt     360
cacgacctgt tttccgccga gccgcaaatc attctcggtg accctccgga ggggctgtcg     420
gtggaaatgt tcatagcgat ggatccgccg aaacacgatg tgcagcgcag ctcggtgcag     480
ggagtagtgg caccgaaaaa cctgaaggag atggaggggc tgatccgatc acgcaccggc     540
gatgtgcttg acagcctgcc tacagacaaa ccctttaact gggtacctgc tgtttccaag     600
gaactcacag gccgcatgct ggcgacgctt ctggattttc cttacgagga acgccacaag     660
ctggttgagt ggtcggacag aatggcaggt gcagcatcgg ccaccggcgg ggagtttgcc     720
gatgaaaatg ccatgtttga cgacgcggca gacatggccc ggtctttctc caggctttgg     780
cgggacaagg aggcgcgccg cgcagcaggc gaggagcccg gtttcgattt gatcagcctg     840
ttgcagagca acaaagaaac gaaagacctg atcaatcggc cgatggagtt tatcggtaat     900
ttgacgctgc tcatagtcgc cggcaacgat acgacgcgca actcgatgag tggtggcctg     960
gtggccatga acgaattccc cagggaattt gaaaaattga aggcaaaacc ggagttgatt    1020
ccgaacatgg tgtcggaaat catccgctgg caaacgccgc tggcctatat gcgccgaatc    1080
gccaagcagg atgtcgaact gggcggccag accatcaaga agggtgatcg agttgtcatg    1140
tggtacgcgt cgggtaaccg ggacgagcgc aaatttgaca accccgatca gttcatcatt    1200
gatcgcaagg acgcacgaaa ccacatgtcg ttcggctatg gggttcaccg ttgcatgggc    1260
aaccgtctgg ctgaactgca actgcgcatc ctctgggaag aaatactcaa gcgttttgac    1320
aacatcgaag tcgtcgaaga gcccgagcgg gtgcagtcca acttcgtgcg gggctattcc    1380
aggttgatgg tcaaactgac accgaacagt gtactccatc gtcatcaacc tgtcaccatc    1440
ggcgagccgg ccgctcgtgc tgtgagccgc acggtgaccg ttgagcgtct tgatcgcatt    1500
gccgacgatg tccttcgcct ggtccttcgc gatgctggag gtaaaaccct cccgacgtgg    1560
acgcctggcg ctcacatcga cctggatctg ggtgctctga gccgtcagta ttcgctctgc    1620
ggcgctccgg atgctccgtc gtacgaaatc gccgtgcact tagatccgga aagccgtggt    1680
ggaagccgct atattcatga acagctggaa gttggaagtc cgctgcgtat gcgtggccca    1740
cgcaaccatt tcgccctgga tccgggtgcg gaacattacg tgtttgttgc cgggggtatc    1800
ggcatcacgc cggtgctggc aatggcggat catgcccgtg cgcgtggttg gtcgtacgaa    1860
ctgcattatt gtggtcgtaa tcgtagcggt atggcttacc tggaacgcgt cgcgggacat    1920
ggtgaccgcg ctgccttgca cgtatctgaa gaaggcaccc gcattgatct ggcggcatta    1980
cttgctgaac cggcggcggg cgtgcaaatc tacgcctgcg gtccgggccg tttattagcg    2040
ggtcttgaag acgcgtctcg taattggccg gatggcgcgc ttcatgtgga gcatttcact    2100
tcgagtttag ccgctttgga tccggatgtc gaacatgcct ttgatttgga gctgcgtgac    2160
tctggcctta ccgttcgcgt cgagccaact cagaccgttt tagacgcttt gcgtgcgaac    2220
aatatcgacg tcccgtcgga ttgcgaagag gggctgtgtg gttcttgcga agtagccgtt    2280
ctggatggcg aggttgatca ccgtgatacc gttctgacta aggccgagcg cgccgcgaat    2340
cgtcagatga tgacttgctg cagtcgtgca tgcggtgatc gtctggcgct gcgcctctaa    2400
```

<210> 36
<211> 799
<212> PRT
<213> 人工序列
<220>
<221> 来源
<223> /注="人工序列的描述：合成的多肽"
<400> 36

```
Met Pro Thr Leu Pro Arg Thr Phe Asp Asp Ile Gln Ser Arg Leu Ile
1               5                  10                  15
Asn Ala Thr Ser Arg Val Val Pro Met Gln Arg Gln Ile Gln Gly Leu
            20                  25                  30
Lys Phe Leu Met Ser Ala Lys Arg Lys Thr Phe Gly Pro Arg Arg Pro
        35                  40                  45
Met Pro Glu Phe Val Glu Thr Pro Ile Pro Asp Val Asn Thr Leu Ala
    50                  55                  60
Leu Glu Asp Ile Asp Val Ser Asn Pro Phe Leu Tyr Arg Gln Gly Gln
65                  70                  75                  80
Trp Arg Ala Tyr Phe Lys Arg Leu Arg Asp Glu Ala Pro Val His Tyr
                85                  90                  95
Gln Lys Asn Ser Pro Phe Gly Pro Phe Trp Ser Val Thr Arg Phe Glu
            100                 105                 110
Asp Ile Leu Phe Val Asp Lys Ser His Asp Leu Phe Ser Ala Glu Pro
        115                 120                 125
Gln Ile Ile Leu Gly Asp Pro Pro Glu Gly Leu Ser Val Glu Met Phe
    130                 135                 140
Ile Ala Met Asp Pro Pro Lys His Asp Val Gln Arg Ser Ser Val Gln
145                 150                 155                 160
Gly Val Val Ala Pro Lys Asn Leu Lys Glu Met Glu Gly Leu Ile Arg
                165                 170                 175
Ser Arg Thr Gly Asp Val Leu Asp Ser Leu Pro Thr Asp Lys Pro Phe
            180                 185                 190
```

```
Asn Trp Val Pro Ala Val Ser Lys Glu Leu Thr Gly Arg Met Leu Ala
        195                 200                 205
Thr Leu Leu Asp Phe Pro Tyr Glu Glu Arg His Lys Leu Val Glu Trp
    210                 215                 220
Ser Asp Arg Met Ala Gly Ala Ala Ser Ala Thr Gly Gly Glu Phe Ala
225                 230                 235                 240
Asp Glu Asn Ala Met Phe Asp Asp Ala Ala Asp Met Ala Arg Ser Phe
                245                 250                 255
Ser Arg Leu Trp Arg Asp Lys Glu Ala Arg Arg Ala Ala Gly Glu Glu
            260                 265                 270
Pro Gly Phe Asp Leu Ile Ser Leu Leu Gln Ser Asn Lys Glu Thr Lys
        275                 280                 285
Asp Leu Ile Asn Arg Pro Met Glu Phe Ile Gly Asn Leu Thr Leu Leu
    290                 295                 300
Ile Val Ala Gly Asn Asp Thr Thr Arg Asn Ser Met Ser Gly Gly Leu
305                 310                 315                 320
Val Ala Met Asn Glu Phe Pro Arg Glu Phe Glu Lys Leu Lys Ala Lys
                325                 330                 335
Pro Glu Leu Ile Pro Asn Met Val Ser Glu Ile Ile Arg Trp Gln Thr
            340                 345                 350
Pro Leu Ala Tyr Met Arg Arg Ile Ala Lys Gln Asp Val Glu Leu Gly
        355                 360                 365
Gly Gln Thr Ile Lys Lys Gly Asp Arg Val Val Met Trp Tyr Ala Ser
    370                 375                 380
Gly Asn Arg Asp Glu Arg Lys Phe Asp Asn Pro Asp Gln Phe Ile Ile
385                 390                 395                 400
Asp Arg Lys Asp Ala Arg Asn His Met Ser Phe Gly Tyr Gly Val His
                405                 410                 415
Arg Cys Met Gly Asn Arg Leu Ala Glu Leu Gln Leu Arg Ile Leu Trp
            420                 425                 430
Glu Glu Ile Leu Lys Arg Phe Asp Asn Ile Glu Val Val Glu Glu Pro
        435                 440                 445
Glu Arg Val Gln Ser Asn Phe Val Arg Gly Tyr Ser Arg Leu Met Val
    450                 455                 460
Lys Leu Thr Pro Asn Ser Val Leu His Arg His Gln Pro Val Thr Ile
465                 470                 475                 480
Gly Glu Pro Ala Ala Arg Ala Val Ser Arg Thr Val Thr Val Glu Arg
                485                 490                 495
Leu Asp Arg Ile Ala Asp Asp Val Leu Arg Leu Val Leu Arg Asp Ala
            500                 505                 510
Gly Gly Lys Thr Leu Pro Thr Trp Thr Pro Gly Ala His Ile Asp Leu
        515                 520                 525
Asp Leu Gly Ala Leu Ser Arg Gln Tyr Ser Leu Cys Gly Ala Pro Asp
    530                 535                 540
Ala Pro Ser Tyr Glu Ile Ala Val His Leu Asp Pro Glu Ser Arg Gly
545                 550                 555                 560
Gly Ser Arg Tyr Ile His Glu Gln Leu Glu Val Gly Ser Pro Leu Arg
                565                 570                 575
Met Arg Gly Pro Arg Asn His Phe Ala Leu Asp Pro Gly Ala Glu His
            580                 585                 590
Tyr Val Phe Val Ala Gly Gly Ile Gly Ile Thr Pro Val Leu Ala Met
        595                 600                 605
Ala Asp His Ala Arg Ala Arg Gly Trp Ser Tyr Glu Leu His Tyr Cys
    610                 615                 620
Gly Arg Asn Arg Ser Gly Met Ala Tyr Leu Glu Arg Val Ala Gly His
625                 630                 635                 640
Gly Asp Arg Ala Ala Leu His Val Ser Glu Glu Gly Thr Arg Ile Asp
                645                 650                 655
Leu Ala Ala Leu Leu Ala Glu Pro Ala Ala Gly Val Gln Ile Tyr Ala
            660                 665                 670
Cys Gly Pro Gly Arg Leu Leu Ala Gly Leu Glu Asp Ala Ser Arg Asn
        675                 680                 685
Trp Pro Asp Gly Ala Leu His Val Glu His Phe Thr Ser Ser Leu Ala
    690                 695                 700
Ala Leu Asp Pro Asp Val Glu His Ala Phe Asp Leu Glu Leu Arg Asp
705                 710                 715                 720
Ser Gly Leu Thr Val Arg Val Glu Pro Thr Gln Thr Val Leu Asp Ala
                725                 730                 735
Leu Arg Ala Asn Asn Ile Asp Val Pro Ser Asp Cys Glu Glu Gly Leu
            740                 745                 750
Cys Gly Ser Cys Glu Val Ala Val Leu Asp Gly Glu Val Asp His Arg
        755                 760                 765
Asp Thr Val Leu Thr Lys Ala Glu Arg Ala Ala Asn Arg Gln Met Met
    770                 775                 780
Thr Cys Cys Ser Arg Ala Cys Gly Asp Arg Leu Ala Leu Arg Leu
785                 790                 795
```

[0036]

<210> 37
<211> 2400
<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> 37

[0037]

atgccaacac tgcccagaac atttgacgac attcagtccc gactgattaa cgccacctcc 60
agggtggtgc cgatgcagag gcaaattcag ggactgaaat tcttaatgag cgccaagagg 120
aagaccttcg gcccacgccg accgatgccc gaattcgttg aaacacccat cccggacgtt 180
aacacgctgg cccttgagga catcgatgtc agcaatccgt ttttataccg gcagggtcag 240
tggcgcgcct atttcaaacg gttgcgtgat gaggcgccgg tccattacca gaagaacagc 300
cctttcggcc ccttctggtc ggtaactcgg tttgaagaca tcctgttcgt ggataagagt 360
cacgacctgt tttccgccga gccgcaaatc attctcggtg accctccgga ggggctgtcg 420
gtggaaatgt tcatagcgat ggatccgccg aaacacgatg tgcagcgcag ctcggtgcag 480
ggagtagtgg caccgaaaaa cctgaaggag atggaggggc tgatccgatc acgcaccggc 540
gatgtgcttg acagcctgcc tacagacaaa ccctttaact gggtacctgc tgtttccaag 600
gaactcacag gccgcatgct ggcgacgctt ctggattttc cttacgagga acgccacaag 660
ctggttgagt ggtcggacag aatggcaggt gcagcatcgg ccaccggcgg ggagtttgcc 720
gatgaaaatg ccatgtttga cgacgcggca gacatggccc ggtctttctc caggctttgg 780
cgggacaagg aggcgcgccg cgcagcaggc gaggagcccg gtttcgattt gatcagcctg 840
ttgcagagca acaaagaaac gaaagacctg atcaatcggc cgatggagtt tatcggtaat 900
ttgacgctgc tcatagtcgc cggcaacgat acgacgcgca actcgatgag tggtggcctg 960
gtggccatga acgaattccc cagggaattt gaaaaattga aggcaaaacc ggagttgatt 1020
ccgaacatgg tgtcggaaat catccgctgg caaacgccgc tggcctatat gcgccgaatc 1080
gccaagcagg atgtcgaact gggcggccag accatcaaga agggtgatcg agttgtcatg 1140
tggtacgcgt cgggtaaccg ggacgagcgc aaatttgaca accccgatca gttcatcatt 1200
gatcgcaagg acgcacgaaa ccacatgtcg ttcggctatg gggttcaccg ttgcatgggc 1260
aaccgtctgg ctgaactgca actgcgcatc ctctgggaag aaatactcaa gcgttttgac 1320
aacatcgaag tcgtcgaaga gcccgagcgg gtgcagtcca acttcgtgcg gggctattcc 1380
aggttgatgg tcaaactgac accgaacagt gtactccatc gtcatcaacc tgtcaccatc 1440
ggcgagccgg ccgctcgtgc tgtgagccgc acggtgaccg ttgagcgtct tgatcgcatt 1500
gccgacgatg tccttcgcct ggtccttcgc gatgctggag gtaaaaccct cccgacgtgg 1560
acgcctggcg ctcacatcga cctggatctg ggtgctctga gccgtcagta ttcgctctgc 1620
ggcgctccgg atgctccgtc gtacgaaatc gccgtgcact tagatccgga aagccgtggt 1680
ggaagccgct atattcatga acagctggaa gttggaagtc cgctgcgtat gcgtggccca 1740
cgcaaccatt tcgccctgga tccgggtgcg gaacattacg tgtttgttgc cgggggtatc 1800
ggcatcacgc cggtgctggc aatggcggat catgcccgtg cgcgtggttg gtcgtacgaa 1860
ctgcattatt gtggtcgtaa tcgtagcggt atggcttacc tggaacgcgt cgcgggacat 1920
ggtgaccgcg ctgccttgca cgtatctgaa gaaggcaccc gcattgatct ggcggcatta 1980
cttgctgaac cggcgccggg cgtgcaaatc tacgcctgcg gtccgggccg tttattagcg 2040
ggtcttgaag acgcgtctcg taattggccg gatggcgcgc ttcatgtgga gcatttcact 2100
tcgagtttag ccgctttgga tccggatgtc gaacatgcct ttgatttgga gctgcgtgac 2160
tctggcctta ccgttcgcgt cgagccaact cagaccgttt tagacgcttt gcgtgcgaac 2220
aatatcgacg tcccgtcgga ttgcgaagag gggctgtgtg gttcttgcga agtagccgtt 2280
ctggatggcg aggttgatca ccgtgatacc gttctgacta aggccgagcg cgccgcgaat 2340
cgtcagatga tgacttgctg cagtcgtgca tgcggtgatc gtctggtcct gcgcctctaa 2400

<210> 38
<211> 799
<212> PRT
<213> 人工序列
<220>
<221> 来源
<223> /注="人工序列的描述：合成的多肽"
<400> 38

Met Pro Thr Leu Pro Arg Thr Phe Asp Asp Ile Gln Ser Arg Leu Ile
1 5 10 15
Asn Ala Thr Ser Arg Val Val Pro Met Gln Arg Gln Ile Gln Gly Leu
20 25 30
Lys Phe Leu Met Ser Ala Lys Arg Lys Thr Phe Gly Pro Arg Arg Pro
35 40 45
Met Pro Glu Phe Val Glu Thr Pro Ile Pro Asp Val Asn Thr Leu Ala
50 55 60
Leu Glu Asp Ile Asp Val Ser Asn Pro Phe Leu Tyr Arg Gln Gly Gln
65 70 75 80
Trp Arg Ala Tyr Phe Lys Arg Leu Arg Asp Glu Ala Pro Val His Tyr
85 90 95
Gln Lys Asn Ser Pro Phe Gly Pro Phe Trp Ser Val Thr Arg Phe Glu
100 105 110
Asp Ile Leu Phe Val Asp Lys Ser His Asp Leu Phe Ser Ala Glu Pro
115 120 125
Gln Ile Ile Leu Gly Asp Pro Pro Glu Gly Leu Ser Val Glu Met Phe
130 135 140

[0038]

```
Ile Ala Met Asp Pro Pro Lys His Asp Val Gln Arg Ser Ser Val Gln
145                 150                 155                 160
Gly Val Val Ala Pro Lys Asn Leu Lys Glu Met Glu Gly Leu Ile Arg
                165                 170                 175
Ser Arg Thr Gly Asp Val Leu Asp Ser Leu Pro Thr Asp Lys Pro Phe
            180                 185                 190
Asn Trp Val Pro Ala Val Ser Lys Glu Leu Thr Gly Arg Met Leu Ala
        195                 200                 205
Thr Leu Leu Asp Phe Pro Tyr Glu Glu Arg His Lys Leu Val Glu Trp
    210                 215                 220
Ser Asp Arg Met Ala Gly Ala Ala Ser Ala Thr Gly Gly Glu Phe Ala
225                 230                 235                 240
Asp Glu Asn Ala Met Phe Asp Asp Ala Ala Asp Met Ala Arg Ser Phe
                245                 250                 255
Ser Arg Leu Trp Arg Asp Lys Glu Ala Arg Arg Ala Ala Gly Glu Glu
            260                 265                 270
Pro Gly Phe Asp Leu Ile Ser Leu Leu Gln Ser Asn Lys Glu Thr Lys
        275                 280                 285
Asp Leu Ile Asn Arg Pro Met Glu Phe Ile Gly Asn Leu Thr Leu Leu
    290                 295                 300
Ile Val Ala Gly Asn Asp Thr Thr Arg Asn Ser Met Ser Gly Gly Leu
305                 310                 315                 320
Val Ala Met Asn Glu Phe Pro Arg Glu Phe Glu Lys Leu Lys Ala Lys
                325                 330                 335
Pro Glu Leu Ile Pro Asn Met Val Ser Glu Ile Ile Arg Trp Gln Thr
            340                 345                 350
Pro Leu Ala Tyr Met Arg Arg Ile Ala Lys Gln Asp Val Glu Leu Gly
        355                 360                 365
Gly Gln Thr Ile Lys Lys Gly Asp Arg Val Val Met Trp Tyr Ala Ser
    370                 375                 380
Gly Asn Arg Asp Glu Arg Lys Phe Asp Asn Pro Asp Gln Phe Ile Ile
385                 390                 395                 400
Asp Arg Lys Asp Ala Arg Asn His Met Ser Phe Gly Tyr Gly Val His
                405                 410                 415
Arg Cys Met Gly Asn Arg Leu Ala Glu Leu Gln Leu Arg Ile Leu Trp
            420                 425                 430
Glu Glu Ile Leu Lys Arg Phe Asp Asn Ile Glu Val Val Glu Glu Pro
        435                 440                 445
Glu Arg Val Gln Ser Asn Phe Val Arg Gly Tyr Ser Arg Leu Met Val
    450                 455                 460
Lys Leu Thr Pro Asn Ser Val Leu His Arg His Gln Pro Val Thr Ile
465                 470                 475                 480
Gly Glu Pro Ala Ala Arg Ala Val Ser Arg Thr Val Thr Val Glu Arg
                485                 490                 495
Leu Asp Arg Ile Ala Asp Asp Val Leu Arg Leu Val Leu Arg Asp Ala
            500                 505                 510
Gly Gly Lys Thr Leu Pro Thr Trp Thr Pro Gly Ala His Ile Asp Leu
        515                 520                 525
Asp Leu Gly Ala Leu Ser Arg Gln Tyr Ser Leu Cys Gly Ala Pro Asp
    530                 535                 540
Ala Pro Ser Tyr Glu Ile Ala Val His Leu Asp Pro Glu Ser Arg Gly
545                 550                 555                 560
Gly Ser Arg Tyr Ile His Glu Gln Leu Glu Val Gly Ser Pro Leu Arg
                565                 570                 575
Met Arg Gly Pro Arg Asn His Phe Ala Leu Asp Pro Gly Ala Glu His
            580                 585                 590
Tyr Val Phe Val Ala Gly Gly Ile Gly Ile Thr Pro Val Leu Ala Met
        595                 600                 605
Ala Asp His Ala Arg Ala Arg Gly Trp Ser Tyr Glu Leu His Tyr Cys
    610                 615                 620
Gly Arg Asn Arg Ser Gly Met Ala Tyr Leu Glu Arg Val Ala Gly His
625                 630                 635                 640
Gly Asp Arg Ala Ala Leu His Val Ser Glu Glu Gly Thr Arg Ile Asp
                645                 650                 655
Leu Ala Ala Leu Leu Ala Glu Pro Ala Pro Gly Val Gln Ile Tyr Ala
            660                 665                 670
Cys Gly Pro Gly Arg Leu Leu Ala Gly Leu Glu Asp Ala Ser Arg Asn
        675                 680                 685
Trp Pro Asp Gly Ala Leu His Val Glu His Phe Thr Ser Ser Leu Ala
    690                 695                 700
Ala Leu Asp Pro Asp Val Glu His Ala Phe Asp Leu Glu Leu Arg Asp
705                 710                 715                 720
Ser Gly Leu Thr Val Arg Val Glu Pro Thr Gln Thr Val Leu Asp Ala
                725                 730                 735
Leu Arg Ala Asn Asn Ile Asp Val Pro Ser Asp Cys Glu Glu Gly Leu
            740                 745                 750
```

```
Cys Gly Ser Cys Glu Val Ala Val Leu Asp Gly Glu Val Asp His Arg
        755                 760                 765
Asp Thr Val Leu Thr Lys Ala Glu Arg Ala Ala Asn Arg Gln Met Met
    770                 775                 780
Thr Cys Cys Ser Arg Ala Cys Gly Asp Arg Leu Val Leu Arg Leu
785                 790                 795
```

<210> 39
<211> 2400
<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> 39

[0039]

```
atgccaacac tgcccagaac atttgacgac attcagtccc gactgattaa cgccacctcc       60
agggtggtgc cgatgcagag gcaaattcag ggactgaaat tcttaatgag cgccaagagg      120
aagaccttcg gcccacgccg accgatgccc gaattcgttg aaacacccat cccggacgtt      180
aacacgctgg cccttgagga catcgatgtc agcaatccgt ttttataccg gcagggtcag      240
tggcgcgcct atttcaaacg gttgcgtgat gaggcgccgg tccattacca gaagaacagc      300
cctttcggcc ccttctggtc ggtaactcgg tttgaagaca tcctgttcgt ggataagagt      360
cacgacctgt tttccgccga gccgcaaatc attctcggtg accctccgga ggggctgtcg      420
gtggaaatgt tcatagcgat ggatccgccg aaacacgatg tgcagcgcag ctcggtgcag      480
ggagtagtgg caccgaaaaa cctgaaggag atggaggggc tgatccgatc acgcaccggc      540
gatgtgcttg acagcctgcc tacagacaaa ccctttaact gggtacctgc tgtttccaag      600
gaactcacag gccgcatgct ggcgacgctt ctggattttc cttacgagga acgccacaag      660
ctggttgagt ggtcggacag aatggcaggt gcagcatcgg ccaccggcgg ggagtttgcc      720
gatgaaaatg ccatgtttga cgacgcggca gacatggccc ggtctttctc caggctttgg      780
cgggacaagg aggcgcgccg cgcagcaggc gaggagcccg gtttcgattt gatcagcctg      840
ttgcagagca acaaagaaac gaaagacctg atcaatcggc cgatggagtt tatcggtaat      900
ttgacgctgc tcatagtcgc cggcaacgat acgacgcgca actcgatgag tggtggcctg      960
gtggccatga acgaattccc cagggaattt gaaaaattga aggcaaaacc ggagttgatt     1020
ccgaacatgg tgtcggaaat catccgctgg caaacgccgc tggcctatat gcgccgaatc     1080
gccaagcagg atgtcgaact gggcggccag accatcaaga agggtgatcg agttgtcatg     1140
tggtacgcgt cgggtaaccg ggacgagcgc aaatttgaca accccgatca gttcatcatt     1200
gatcgcaagg acgcacgaaa ccacatgtcg ttcggctatg gggttcaccg ttgcatgggc     1260
aaccgtctgg ctgaactgca actgcgcatc ctctgggaag aaatactcaa gcgttttgac     1320
aacatcgaag tcgtcgaaga gcccgagcgg gtgcagtcca acttcgtgcg gggctattcc     1380
aggttgatgg tcaaactgac accgaacagt gtactccatc gtcatcaacc tgtcaccatc     1440
ggcgagccgg ccgctcgtgc tgtgagccgc acggtgaccg ttgagcgtct tgatcgcatt     1500
gccgacgatg tccttcgcct ggtccttcgc gatgctggag gtaaagtcct cccgacgtgg     1560
acgcctggcg ctcacatcga cctggatctg ggtgctctga gccgtcagta ttcgctctgc     1620
ggcgctccgg atgctccgtc gtacgaaatc gccgtgcact tagatccgga aagccgtggt     1680
ggaagccgct atattcatga acagctggaa gttggaagtc cgctgcgtat gcgtggccca     1740
cgcaaccatt tcgccctgga tccgggtgcg gaacattacg tgtttgttgc cgggggtatc     1800
ggcatcacgc cggtgctggc aatggcggat catgcccgtg cgcgtggttg gtcgtacgaa     1860
ctgcattatt gtggtcgtaa tcgtagcggt atggcttacc tggaacgcgt cgcgggacat     1920
ggtgaccgcg ctgccttgca cgtatctgaa gaaggcaccc gcattgatct ggcggcatta     1980
cttgctgaac cggcggacgg cgtgcaaatc tacgcctgcg gtccgggccg tttattagcg     2040
ggtcttgaag acgcgtctcg taattggccg gatggcgcgc ttcatgtgga gcatttcact     2100
tcgagtttag ccgctttgga tccggatgtc gaacatgcct ttgatttgga gctgcgtgac     2160
tctggcctta ccgttcgcgt cgagccaact cagaccgttt tagacgcttt gcgtgcgaac     2220
aatatcgacg tcccgtcgga ttgcgaagag gggctgtgtg gttcttgcga agtagccgtt     2280
ctggatggcg aggttgatca ccgtgatacc gttctgacta aggccgagcg cgccgcgaat     2340
cgtcagatga tgacttgctg cagtcgtgca tgcggtgatc gtctggtcct gcgcctctaa     2400
```

<210> 40
<211> 799
<212> PRT
<213> 人工序列
<220>
<221> 来源
<223> /注="人工序列的描述：合成的多肽"
<400> 40

```
Met Pro Thr Leu Pro Arg Thr Phe Asp Asp Ile Gln Ser Arg Leu Ile
1               5                  10                  15
Asn Ala Thr Ser Arg Val Val Pro Met Gln Arg Gln Ile Gln Gly Leu
            20                  25                  30
Lys Phe Leu Met Ser Ala Lys Arg Lys Thr Phe Gly Pro Arg Arg Pro
        35                  40                  45
Met Pro Glu Phe Val Glu Thr Pro Ile Pro Asp Val Asn Thr Leu Ala
    50                  55                  60
Leu Glu Asp Ile Asp Val Ser Asn Pro Phe Leu Tyr Arg Gln Gly Gln
65                  70                  75                  80
Trp Arg Ala Tyr Phe Lys Arg Leu Arg Asp Glu Ala Pro Val His Tyr
                85                  90                  95
```

[0040]

```
Gln Lys Asn Ser Pro Phe Gly Pro Phe Trp Ser Val Thr Arg Phe Glu
                100                 105                 110
Asp Ile Leu Phe Val Asp Lys Ser His Asp Leu Phe Ser Ala Glu Pro
            115                 120                 125
Gln Ile Ile Leu Gly Asp Pro Pro Glu Gly Leu Ser Val Glu Met Phe
        130                 135                 140
Ile Ala Met Asp Pro Pro Lys His Asp Val Gln Arg Ser Ser Val Gln
145                 150                 155                 160
Gly Val Val Ala Pro Lys Asn Leu Lys Glu Met Glu Gly Leu Ile Arg
                165                 170                 175
Ser Arg Thr Gly Asp Val Leu Asp Ser Leu Pro Thr Asp Lys Pro Phe
            180                 185                 190
Asn Trp Val Pro Ala Val Ser Lys Glu Leu Thr Gly Arg Met Leu Ala
        195                 200                 205
Thr Leu Leu Asp Phe Pro Tyr Glu Glu Arg His Lys Leu Val Glu Trp
    210                 215                 220
Ser Asp Arg Met Ala Gly Ala Ala Ser Ala Thr Gly Gly Glu Phe Ala
225                 230                 235                 240
Asp Glu Asn Ala Met Phe Asp Asp Ala Ala Asp Met Ala Arg Ser Phe
                245                 250                 255
Ser Arg Leu Trp Arg Asp Lys Glu Ala Arg Arg Ala Ala Gly Glu Glu
            260                 265                 270
Pro Gly Phe Asp Leu Ile Ser Leu Leu Gln Ser Asn Lys Glu Thr Lys
        275                 280                 285
Asp Leu Ile Asn Arg Pro Met Glu Phe Ile Gly Asn Leu Thr Leu Leu
    290                 295                 300
Ile Val Ala Gly Asn Asp Thr Thr Arg Asn Ser Met Ser Gly Gly Leu
305                 310                 315                 320
Val Ala Met Asn Glu Phe Pro Arg Glu Phe Glu Lys Leu Lys Ala Lys
                325                 330                 335
Pro Glu Leu Ile Pro Asn Met Val Ser Glu Ile Ile Arg Trp Gln Thr
            340                 345                 350
Pro Leu Ala Tyr Met Arg Arg Ile Ala Lys Gln Asp Val Glu Leu Gly
        355                 360                 365
Gly Gln Thr Ile Lys Lys Gly Asp Arg Val Val Met Trp Tyr Ala Ser
    370                 375                 380
Gly Asn Arg Asp Glu Arg Lys Phe Asp Asn Pro Asp Gln Phe Ile Ile
385                 390                 395                 400
Asp Arg Lys Asp Ala Arg Asn His Met Ser Phe Gly Tyr Gly Val His
                405                 410                 415
Arg Cys Met Gly Asn Arg Leu Ala Glu Leu Gln Leu Arg Ile Leu Trp
            420                 425                 430
Glu Glu Ile Leu Lys Arg Phe Asp Asn Ile Glu Val Val Glu Glu Pro
        435                 440                 445
Glu Arg Val Gln Ser Asn Phe Val Arg Gly Tyr Ser Arg Leu Met Val
    450                 455                 460
Lys Leu Thr Pro Asn Ser Val Leu His Arg His Gln Pro Val Thr Ile
465                 470                 475                 480
Gly Glu Pro Ala Ala Arg Ala Val Ser Arg Thr Val Thr Val Glu Arg
                485                 490                 495
Leu Asp Arg Ile Ala Asp Asp Val Leu Arg Leu Val Leu Arg Asp Ala
            500                 505                 510
Gly Gly Lys Val Leu Pro Thr Trp Thr Pro Gly Ala His Ile Asp Leu
        515                 520                 525
Asp Leu Gly Ala Leu Ser Arg Gln Tyr Ser Leu Cys Gly Ala Pro Asp
    530                 535                 540
Ala Pro Ser Tyr Glu Ile Ala Val His Leu Asp Pro Glu Ser Arg Gly
545                 550                 555                 560
Gly Ser Arg Tyr Ile His Glu Gln Leu Glu Val Gly Ser Pro Leu Arg
                565                 570                 575
Met Arg Gly Pro Arg Asn His Phe Ala Leu Asp Pro Gly Ala Glu His
            580                 585                 590
Tyr Val Phe Val Ala Gly Gly Ile Gly Ile Thr Pro Val Leu Ala Met
        595                 600                 605
Ala Asp His Ala Arg Ala Arg Gly Trp Ser Tyr Glu Leu His Tyr Cys
    610                 615                 620
Gly Arg Asn Arg Ser Gly Met Ala Tyr Leu Glu Arg Val Ala Gly His
625                 630                 635                 640
Gly Asp Arg Ala Ala Leu His Val Ser Glu Glu Gly Thr Arg Ile Asp
                645                 650                 655
Leu Ala Ala Leu Leu Ala Glu Pro Ala Asp Gly Val Gln Ile Tyr Ala
            660                 665                 670
Cys Gly Pro Gly Arg Leu Leu Ala Gly Leu Glu Asp Ala Ser Arg Asn
        675                 680                 685
Trp Pro Asp Gly Ala Leu His Val Glu His Phe Thr Ser Ser Leu Ala
    690                 695                 700
```

```
Ala Leu Asp Pro Asp Val Glu His Ala Phe Asp Leu Glu Leu Arg Asp
705                 710                 715                 720
Ser Gly Leu Thr Val Arg Val Glu Pro Thr Gln Thr Val Leu Asp Ala
                725                 730                 735
Leu Arg Ala Asn Asn Ile Asp Val Pro Ser Asp Cys Glu Glu Gly Leu
            740                 745                 750
Cys Gly Ser Cys Glu Val Ala Val Leu Asp Gly Glu Val Asp His Arg
        755                 760                 765
Asp Thr Val Leu Thr Lys Ala Glu Arg Ala Ala Asn Arg Gln Met Met
    770                 775                 780
Thr Cys Cys Ser Arg Ala Cys Gly Asp Arg Leu Val Leu Arg Leu
785                 790                 795
```

<210> 41
<211> 2400
<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> 41

[0041]

```
atgccaacac tgcccagaac atttgacgac attcagtccc gactgattaa cgccacctcc       60
agggtggtgc cgatgcagag gcaaattcag ggactgaaat tcttaatgag cgccaagagg      120
aagaccttcg gcccacgccg accgatgccc gaattcgttg aaacacccat cccggacgtt      180
aacacgctgg cccttgagga catcgatgtc agcaatccgt ttttataccg gcagggtcag      240
tggcgcgcct atttcaaacg gttgcgtgat gaggcgccgg tccattacca gaagaacagc      300
cctttcggcc ccttctggtc ggtaactcgg tttgaagaca tcctgttcgt ggataagagt      360
cacgacctgt tttccgccga gccgcaaatc attctcggtg accctccgga ggggctgtcg      420
atcgaaatgt tcatagcgat ggatccgccg aaacacgatg tgcagcgcag ctcggtgcag      480
ggagtagtgg caccgaaaaa cctgaaggag atggaggggc tgatccgatc acgcaccggc      540
gatgtgcttg acagcctgcc tacagacaaa ccctttaact gggtacctgc tgtttccaag      600
gaactcacag gccgcatgct ggcgacgctt ctggattttc cttacgagga acgccacaag      660
ctggttgagt ggtcggacag aatggcaggt acagcatcgg ccaccggcgg ggagtttgcc      720
gatgaaaatg ccatgtttga cgacgcggca gacatggccc ggtctttctc caggctttgg      780
cgggacaagg aggcgcgccg cgcagcaggc gaggagcccg gtttcgattt gatcagcctg      840
ttgcagagca acaaagaaac gaaagacctg atcaatcggc cgatggagtt tatcggtaat      900
ttgacgctgc tcatagtcgc cggcaacgat acgacgcgca actcgatgag tggtggcctg      960
gtggccatga acgaattccc cagggaattt gaaaaattga aggcaaaacc ggagttgatt     1020
ccgaacatgg tgtcggaaat catccgctgg caaacgccgc tggcctatat gcgccgaatc     1080
gccaagcagg atgtcgaact gggcggccag accatcaaga agggtgatcg agttgtcatg     1140
tggtacgcgt cgggtaaccg ggacgagcgc aaatttgaca accccgatca gttcatcatt     1200
gatcgcaagg acgcacgaaa ccacatgtcg ttcggctatg gggttcaccg ttgcatgggc     1260
aaccgtctgg ctgaactgca actgcgcatc ctctgggaag aaatactcaa gcgttttgac     1320
aacatcgaag tcgtcgaaga gcccgagcgg gtgcagtcca acttcgtgcg gggctattcc     1380
aggttgatgg tcaaactgac accgaacagt gtactccatc gtcatcaacc tgtcaccatc     1440
ggcgagccgg ccgctcgtgc tgtgagccgc acggtgaccg ttgagcgtct tgatcgcatt     1500
gccgacgatg tccttcgcct ggtccttcgc gatgctggag gtaaaaccct cccgacgtgg     1560
acgcctggcg ctcacatcga cctggatctg ggtgctctga gccgtcagta ttcgctctgc     1620
ggcgctccgg atgctccgtc gtacgaaatc gccgtgcact tagatccgga aagccgtggt     1680
ggaagccgct atattcatga acagctggaa gttggaagtc cgctgcgtat gcgtggccca     1740
cgcaaccatt tcgccctgga tccgggtgcg gaacattacg tgtttgttgc cgggggtatc     1800
ggcatcacgc cggtgctggc aatggcggat catgcccgtg cgcgtggttg gtcgtacgaa     1860
ctgcattatt gtggtcgtaa tcgtagcggt atggcttacc tggaacgcgt cgcgggacat     1920
ggtgaccgcg ctgccttgca cgtatctgaa gaaggcaccc gcattgatct ggcggcatta     1980
cttgctgaac cggcgccggg cgtgcaaatc tacgcctgcg gtccgggccg tttattagcg     2040
ggtcttgaag acgcgtctcg taattggccg gatggcgcgc ttcatgtgga gcatttcact     2100
tcgagtttag ccgctttgga tccggatgtc gaacatgcct ttgatttgga gctgcgtgac     2160
tctggcctta ccgttcgcgt cgagccaact cagaccgttt tagacgcttt gcgtgcgaac     2220
aatatcgacg tcccgtcgga ttgcgaagag gggctgtgtg gttcttgcga agtagccgtt     2280
ctggatggcg aggttgatca ccgtgatacc gttctgacta aggccgagcg cgccgcgaat     2340
cgtcagatga tgacttgctg cagtcgtgca tgcggtgatc gtctggtcct gcgcctctaa     2400
```

<210> 42
<211> 799
<212> PRT
<213> 人工序列
<220>
<221> 来源
<223> /注="人工序列的描述：合成的多肽"
<400> 42

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

```
Met Pro Glu Phe Val Glu Thr Pro Ile Pro Asp Val Asn Thr Leu Ala
    50                  55                  60
Leu Glu Asp Ile Asp Val Ser Asn Pro Phe Leu Tyr Arg Gln Gly Gln
65                  70                  75                  80
Trp Arg Ala Tyr Phe Lys Arg Leu Arg Asp Glu Ala Pro Val His Tyr
                85                  90                  95
Gln Lys Asn Ser Pro Phe Gly Pro Phe Trp Ser Val Thr Arg Phe Glu
            100                 105                 110
Asp Ile Leu Phe Val Asp Lys Ser His Asp Leu Phe Ser Ala Glu Pro
        115                 120                 125
Gln Ile Ile Leu Gly Asp Pro Pro Glu Gly Leu Ser Ile Glu Met Phe
    130                 135                 140
Ile Ala Met Asp Pro Pro Lys His Asp Val Gln Arg Ser Ser Val Gln
145                 150                 155                 160
Gly Val Val Ala Pro Lys Asn Leu Lys Glu Met Glu Gly Leu Ile Arg
                165                 170                 175
Ser Arg Thr Gly Asp Val Leu Asp Ser Leu Pro Thr Asp Lys Pro Phe
            180                 185                 190
Asn Trp Val Pro Ala Val Ser Lys Glu Leu Thr Gly Arg Met Leu Ala
        195                 200                 205
Thr Leu Leu Asp Phe Pro Tyr Glu Glu Arg His Lys Leu Val Glu Trp
    210                 215                 220
Ser Asp Arg Met Ala Gly Thr Ala Ser Ala Thr Gly Gly Glu Phe Ala
225                 230                 235                 240
Asp Glu Asn Ala Met Phe Asp Asp Ala Ala Asp Met Ala Arg Ser Phe
                245                 250                 255
Ser Arg Leu Trp Arg Asp Lys Glu Ala Arg Arg Ala Ala Gly Glu Glu
            260                 265                 270
Pro Gly Phe Asp Leu Ile Ser Leu Leu Gln Ser Asn Lys Glu Thr Lys
        275                 280                 285
Asp Leu Ile Asn Arg Pro Met Glu Phe Ile Gly Asn Leu Thr Leu Leu
    290                 295                 300
Ile Val Ala Gly Asn Asp Thr Thr Arg Asn Ser Met Ser Gly Gly Leu
305                 310                 315                 320
Val Ala Met Asn Glu Phe Pro Arg Glu Phe Glu Lys Leu Lys Ala Lys
                325                 330                 335
Pro Glu Leu Ile Pro Asn Met Val Ser Glu Ile Ile Arg Trp Gln Thr
            340                 345                 350
Pro Leu Ala Tyr Met Arg Arg Ile Ala Lys Gln Asp Val Glu Leu Gly
        355                 360                 365
Gly Gln Thr Ile Lys Lys Gly Asp Arg Val Val Met Trp Tyr Ala Ser
    370                 375                 380
Gly Asn Arg Asp Glu Arg Lys Phe Asp Asn Pro Asp Gln Phe Ile Ile
385                 390                 395                 400
Asp Arg Lys Asp Ala Arg Asn His Met Ser Phe Gly Tyr Gly Val His
                405                 410                 415
Arg Cys Met Gly Asn Arg Leu Ala Glu Leu Gln Leu Arg Ile Leu Trp
            420                 425                 430
Glu Glu Ile Leu Lys Arg Phe Asp Asn Ile Glu Val Val Glu Glu Pro
        435                 440                 445
Glu Arg Val Gln Ser Asn Phe Val Arg Gly Tyr Ser Arg Leu Met Val
    450                 455                 460
Lys Leu Thr Pro Asn Ser Val Leu His Arg His Gln Pro Val Thr Ile
465                 470                 475                 480
Gly Glu Pro Ala Ala Arg Ala Val Ser Arg Thr Val Thr Val Glu Arg
                485                 490                 495
Leu Asp Arg Ile Ala Asp Asp Val Leu Arg Leu Val Leu Arg Asp Ala
            500                 505                 510
Gly Gly Lys Thr Leu Pro Thr Trp Thr Pro Gly Ala His Ile Asp Leu
        515                 520                 525
Asp Leu Gly Ala Leu Ser Arg Gln Tyr Ser Leu Cys Gly Ala Pro Asp
    530                 535                 540
Ala Pro Ser Tyr Glu Ile Ala Val His Leu Asp Pro Glu Ser Arg Gly
545                 550                 555                 560
Gly Ser Arg Tyr Ile His Glu Gln Leu Glu Val Gly Ser Pro Leu Arg
                565                 570                 575
Met Arg Gly Pro Arg Asn His Phe Ala Leu Asp Pro Gly Ala Glu His
            580                 585                 590
Tyr Val Phe Val Ala Gly Gly Ile Gly Ile Thr Pro Val Leu Ala Met
        595                 600                 605
Ala Asp His Ala Arg Ala Arg Gly Trp Ser Tyr Glu Leu His Tyr Cys
    610                 615                 620
Gly Arg Asn Arg Ser Gly Met Ala Tyr Leu Glu Arg Val Ala Gly His
625                 630                 635                 640
Gly Asp Arg Ala Ala Leu His Val Ser Glu Glu Gly Thr Arg Ile Asp
                645                 650                 655
```

[0042]

```
Leu Ala Ala Leu Leu Ala Glu Pro Ala Pro Gly Val Gln Ile Tyr Ala
            660                 665                 670
Cys Gly Pro Gly Arg Leu Leu Ala Gly Leu Glu Asp Ala Ser Arg Asn
        675                 680                 685
Trp Pro Asp Gly Ala Leu His Val Glu His Phe Thr Ser Ser Leu Ala
    690                 695                 700
Ala Leu Asp Pro Asp Val Glu His Ala Phe Asp Leu Glu Leu Arg Asp
705                 710                 715                 720
Ser Gly Leu Thr Val Arg Val Glu Pro Thr Gln Thr Val Leu Asp Ala
                725                 730                 735
Leu Arg Ala Asn Asn Ile Asp Val Pro Ser Asp Cys Glu Glu Gly Leu
            740                 745                 750
Cys Gly Ser Cys Glu Val Ala Val Leu Asp Gly Glu Val Asp His Arg
        755                 760                 765
Asp Thr Val Leu Thr Lys Ala Glu Arg Ala Ala Asn Arg Gln Met Met
    770                 775                 780
Thr Cys Cys Ser Arg Ala Cys Gly Asp Arg Leu Val Leu Arg Leu
785                 790                 795
```

<210> 43
<211> 2400
<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> 43

[0043]

```
atgccaacac tgcccagaac atttgacgac attcagtccc gactgattaa cgccacctcc      60
agggtggtgc cgatgcagct gcaaattcag ggactgaaat tcttaatgag cgccaagagg     120
aagaccttcg gcccacgccg accgatgccc gaattcgttg aaacacccat cccggacgtt     180
aacacgctgg cccttgagga catcgatgtc agcaatccgt ttttataccg gcagggtcag     240
tgggacgcct atttcaaacg gttgcgtgat gaggcgccgg tccattacca gaagaacagc     300
cctttcggcc ccttctggtc ggtaactcgg tttgaagaca tcctgttcgt ggataagagt     360
cacgacctgt tttccgccga gccgcaaatc attctcggtg accctccgga ggggctgtcg     420
atggaaatgt tcatagcgat ggatccgccg aaacacgatg tgcagcgcag ctcggtgcag     480
ggagtagtgg caccgaaaaa cctgaaggag atggaggggc tgatccgatc aaacaccggc     540
gatgtgcttg acagcctgcc tacagacaaa ccctttaact gggtacctgc tgtttccaag     600
gaactcacag gccgcatgct ggcgacgctt ctggattttc cttacgagga acgccacaag     660
ctggttgagt ggtcggacag aatggcaggt gcagcatcgg ccaccggcgg ggagtttgcc     720
gatgaaaatg ccatgtttga cgacgcggca gacatggccc ggtctttctc caggctttgg     780
cgggacaagg aggcgcgccg cgcagcaggc gaggagcccg gtttcgattt gatcagcctg     840
ttgcagagca acaaagaaac gaaagacctg atcaatcggc cgatggagtt tatcggtaat     900
ttgacgctgc tcatagtcgc cggcgcggat acgacgcgca actcgatgag tggtggcctg     960
gtggccatga acgaattccc cagggaattt gaaaaattga aggcaaaacc ggagttgatt    1020
ccgaacatgg tgtcggaaat catccgctgg caaacgccgc tggcctatat gcgccgaatc    1080
gccaagcagg atgtcgaact gggcggccag accatcaaga agggtgatcg agttgtcatg    1140
tggtacgcgt cgggtaaccg ggacgagcgc aaatttgaca accccgatca gttcatcatt    1200
gatcgcaagg acgcacgaaa ccacatgtcg ttcggctatg gggttcaccg ttgcatgggc    1260
aaccgtctgg ctgaactgca actgcgcatc ctctgggaag aaatactcaa gcgttttgac    1320
aacatcgaag tcgtcgaaga gcccgagcgg gtgcagtcca acttcgtgcg gggctattcc    1380
aggttgatgg tcaaactgac accgaacagt gtactccatc gtcatcaacc tgtcaccatc    1440
ggcgagccgg ccgctcgtgc tgtgagccgc acggtgaccg ttgagcgtct tgatcgcatt    1500
gccgacgatg tccttcgcct ggtccttcgc gatgctggag gtaaaaccct cccgacgtgg    1560
acgcctggcg ctcacatcga cctggatctg ggtgctctga gccgtcagta ttcgctctgc    1620
ggcgctccgg atgctccgtc gtacgaaatc gccgtgcact tagatccgga aagccgtggt    1680
ggaagccgct atattcatga acagctggaa gttggaagtc cgctgcgtat gcgtggccca    1740
cgcaaccatt tcgccctgga tccgggtgcg gaacattacg tgtttgttgc cgggggtatc    1800
ggcatcacgc cggtgctggc aatggcggat catgcccgtg cgcgtggttg gtcgtacgaa    1860
ctgcattatt gtggtcgtaa tcgtagcggt atggcttacc tggaacgcgt cgcgggacat    1920
ggtgaccgcg ctgccttgca cgtatctgaa gaaggcaccc gcattgatct ggcggcatta    1980
cttgctgaac cggcgccggg cgtgcaaatc tacgcctgcg gtccgggccg tttattagcg    2040
ggtcttgaag acgcgtctcg taattggccg gatggcgcgc ttcatgtgga gcatttcact    2100
tcgagtttag ccgctttgga tccggatgtc gaacatgcct ttgatttgga gctgcgtgac    2160
tctggcctta ccgttcgcgt cgagccaact cagaccgttt tagacgcttt gcgtgcgaac    2220
aatatcgacg tcccgtcgga ttgcgaagag gggctgtgtg gttcttgcga agtagccgtt    2280
ctggatggcg aggttgatca ccgtgatacc gttctgacta aggccgagcg cgccgcgaat    2340
cgtcagatga tgacttgctg cagtcgtgca tgcggtgatc gtctggtcct gcgcctctaa    2400
```

<210> 44
<211> 799
<212> PRT
<213> 人工序列
<220>
<221> 来源
<223> /注="人工序列的描述：合成的多肽"
<400> 44

[0044]

```
Met Pro Thr Leu Pro Arg Thr Phe Asp Asp Ile Gln Ser Arg Leu Ile
1               5                  10                  15
Asn Ala Thr Ser Arg Val Val Pro Met Gln Leu Gln Ile Gln Gly Leu
            20                  25                  30
Lys Phe Leu Met Ser Ala Lys Arg Lys Thr Phe Gly Pro Arg Arg Pro
        35                  40                  45
Met Pro Glu Phe Val Glu Thr Pro Ile Pro Asp Val Asn Thr Leu Ala
    50                  55                  60
Leu Glu Asp Ile Asp Val Ser Asn Pro Phe Leu Tyr Arg Gln Gly Gln
65                  70                  75                  80
Trp Asp Ala Tyr Phe Lys Arg Leu Arg Asp Glu Ala Pro Val His Tyr
                85                  90                  95
Gln Lys Asn Ser Pro Phe Gly Pro Phe Trp Ser Val Thr Arg Phe Glu
            100                 105                 110
Asp Ile Leu Phe Val Asp Lys Ser His Asp Leu Phe Ser Ala Glu Pro
        115                 120                 125
Gln Ile Ile Leu Gly Asp Pro Pro Glu Gly Leu Ser Met Glu Met Phe
    130                 135                 140
Ile Ala Met Asp Pro Pro Lys His Asp Val Gln Arg Ser Ser Val Gln
145                 150                 155                 160
Gly Val Val Ala Pro Lys Asn Leu Lys Glu Met Glu Gly Leu Ile Arg
                165                 170                 175
Ser Asn Thr Gly Asp Val Leu Asp Ser Leu Pro Thr Asp Lys Pro Phe
            180                 185                 190
Asn Trp Val Pro Ala Val Ser Lys Glu Leu Thr Gly Arg Met Leu Ala
        195                 200                 205
Thr Leu Leu Asp Phe Pro Tyr Glu Glu Arg His Lys Leu Val Glu Trp
    210                 215                 220
Ser Asp Arg Met Ala Gly Ala Ala Ser Ala Thr Gly Gly Glu Phe Ala
225                 230                 235                 240
Asp Glu Asn Ala Met Phe Asp Asp Ala Ala Asp Met Ala Arg Ser Phe
                245                 250                 255
Ser Arg Leu Trp Arg Asp Lys Glu Ala Arg Arg Ala Ala Gly Glu Glu
            260                 265                 270
Pro Gly Phe Asp Leu Ile Ser Leu Leu Gln Ser Asn Lys Glu Thr Lys
        275                 280                 285
Asp Leu Ile Asn Arg Pro Met Glu Phe Ile Gly Asn Leu Thr Leu Leu
    290                 295                 300
Ile Val Ala Gly Ala Asp Thr Thr Arg Asn Ser Met Ser Gly Gly Leu
305                 310                 315                 320
Val Ala Met Asn Glu Phe Pro Arg Glu Phe Glu Lys Leu Lys Ala Lys
                325                 330                 335
Pro Glu Leu Ile Pro Asn Met Val Ser Glu Ile Ile Arg Trp Gln Thr
            340                 345                 350
Pro Leu Ala Tyr Met Arg Arg Ile Ala Lys Gln Asp Val Glu Leu Gly
        355                 360                 365
Gly Gln Thr Ile Lys Lys Gly Asp Arg Val Val Met Trp Tyr Ala Ser
    370                 375                 380
Gly Asn Arg Asp Glu Arg Lys Phe Asp Asn Pro Asp Gln Phe Ile Ile
385                 390                 395                 400
Asp Arg Lys Asp Ala Arg Asn His Met Ser Phe Gly Tyr Gly Val His
                405                 410                 415
Arg Cys Met Gly Asn Arg Leu Ala Glu Leu Gln Leu Arg Ile Leu Trp
            420                 425                 430
Glu Glu Ile Leu Lys Arg Phe Asp Asn Ile Glu Val Val Glu Glu Pro
        435                 440                 445
Glu Arg Val Gln Ser Asn Phe Val Arg Gly Tyr Ser Arg Leu Met Val
    450                 455                 460
Lys Leu Thr Pro Asn Ser Val Leu His Arg His Gln Pro Val Thr Ile
465                 470                 475                 480
Gly Glu Pro Ala Ala Arg Ala Val Ser Arg Thr Val Thr Val Glu Arg
                485                 490                 495
Leu Asp Arg Ile Ala Asp Asp Val Leu Arg Leu Val Leu Arg Asp Ala
            500                 505                 510
Gly Gly Lys Thr Leu Pro Thr Trp Thr Pro Gly Ala His Ile Asp Leu
        515                 520                 525
Asp Leu Gly Ala Leu Ser Arg Gln Tyr Ser Leu Cys Gly Ala Pro Asp
    530                 535                 540
Ala Pro Ser Tyr Glu Ile Ala Val His Leu Asp Pro Glu Ser Arg Gly
545                 550                 555                 560
Gly Ser Arg Tyr Ile His Glu Gln Leu Glu Val Gly Ser Pro Leu Arg
                565                 570                 575
Met Arg Gly Pro Arg Asn His Phe Ala Leu Asp Pro Gly Ala Glu His
            580                 585                 590
Tyr Val Phe Val Ala Gly Gly Ile Gly Ile Thr Pro Val Leu Ala Met
        595                 600                 605
```

```
Ala Asp His Ala Arg Ala Arg Gly Trp Ser Tyr Glu Leu His Tyr Cys
    610                 615                 620
Gly Arg Asn Arg Ser Gly Met Ala Tyr Leu Glu Arg Val Ala Gly His
625                 630                 635                 640
Gly Asp Arg Ala Ala Leu His Val Ser Glu Glu Gly Thr Arg Ile Asp
                645                 650                 655
Leu Ala Ala Leu Leu Ala Glu Pro Ala Pro Gly Val Gln Ile Tyr Ala
            660                 665                 670
Cys Gly Pro Gly Arg Leu Leu Ala Gly Leu Glu Asp Ala Ser Arg Asn
        675                 680                 685
Trp Pro Asp Gly Ala Leu His Val Glu His Phe Thr Ser Ser Leu Ala
    690                 695                 700
Ala Leu Asp Pro Asp Val Glu His Ala Phe Asp Leu Glu Leu Arg Asp
705                 710                 715                 720
Ser Gly Leu Thr Val Arg Val Glu Pro Thr Gln Thr Val Leu Asp Ala
                725                 730                 735
Leu Arg Ala Asn Asn Ile Asp Val Pro Ser Asp Cys Glu Glu Gly Leu
            740                 745                 750
Cys Gly Ser Cys Glu Val Ala Val Leu Asp Gly Glu Val Asp His Arg
        755                 760                 765
Asp Thr Val Leu Thr Lys Ala Glu Arg Ala Ala Asn Arg Gln Met Met
    770                 775                 780
Thr Cys Cys Ser Arg Ala Cys Gly Asp Arg Leu Val Leu Arg Leu
785                 790                 795
```

<210> 45
<211> 2400
<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> 45

[0045]

```
atgccaacac tgcccagaac atttgacgac attcagtccc gactgattaa cgccacctcc      60
agggtggtgc cgatgcagag gcaaattcag ggactgaaat tcttaatgag cgccaagagg     120
aagaccttcg gcccacgccg accgatgccc gaattcgttg aaacacccat cccggacgtt     180
aacacgctgg cccttgagga catcgatgtc agcaatccgt ttttataccg gcagggtcag     240
tggcgcgcct atttcaaacg gttgcgtgat gaggcgccgg tccattacca gaagaacagc     300
cctttcggcc ccttctggtc ggtaactcgg tttgaagaca tcctgttcgt ggataagagt     360
cacgacctgt tttccgccga gccgcaaatc attctcggtg accctccgga ggggctgtcg     420
acggaaatgt tcatagcgat ggatccgccg aaacacgatg tgcagcgcag ctcggtgcag     480
ggagtagtgg caccgaaaaa cctgaaggag atggaggggc tgatccgatc acgcaccggc     540
gatgtgcttg acagcctgcc tacagacaaa ccctttaact gggtacctgc tgtttccaag     600
gaactcacag gccgcatgct ggcgacgctt ctggattttc cttacgagga acgccacaag     660
ctggttgagt ggtcggacag aatggcaggt acagcatcgg ccaccggcgg ggagtttgcc     720
gatgaaaatg ccatgtttga cgacgcggca gacatggccc ggtctttctc caggctttgg     780
cgggacaagg aggcgcgccg cgcagcaggc gaggagcccg gtttcgattt gatcagcctg     840
ttgcagagca acaaagaaac gaaagacctg atcaatcggc cgatggagtt tatcggtaat     900
ttgacgctgc tcatagtcgc cggcaacgat acgacgcgca actcgatgag tggtggcctg     960
gtggccatga acgaattccc cagggaattt gaaaaattga aggcaaaacc ggagttgatt    1020
ccgaacatgg tgtcggaaat catccgctgg caaacgccgc tggcctatat gcgccgaatc    1080
gccaagcagg atgtcgaact gggcggccag accatcaaga agggtgatcg agttgtcatg    1140
tggtacgcgt cgggtaaccg ggacgagcgc aaatttgaca accccgatca gttcatcatt    1200
gatcgcaagg acgcacgaaa ccacatgtcg ttcggctatg gggttcaccg ttgcatgggc    1260
aaccgtctgg ctgaactgca actgcgcatc ctctgggaag aaatactcaa gcgttttgac    1320
aacatcgaag tcgtcgaaga gcccgagcgg gtgcagtcca acttcgtgcg gggctattcc    1380
aggttgatgg tcaaactgac accgaacagt gtactccatc gtcatcaacc tgtcaccatc    1440
ggcgagccgg ccgctcgtgc tgtgagccgc acggtgaccg ttgagcgtct tgatcgcatt    1500
gccgacgatg tccttcgcct ggtccttcgc gatgctggag gtaaaaccct cccgacgtgg    1560
acgcctggcg ctcacatcga cctggatctg ggtgctctga gccgtcagta ttcgctctgc    1620
ggcgctccgg atgctccgtc gtacgaaatc gccgtgcact tagatccgga aagccgtggt    1680
ggaagccgct atattcatga acagctggaa gttggaagtc cgctgcgtat gcgtggccca    1740
cgcaaccatt tcgccctgga tccgggtgcg gaacattacg tgtttgttgc cgggggtatc    1800
ggcatcacgc cggtgctggc aatggcggat catgcccgtg cgcgtggttg gtcgtacgaa    1860
ctgcattatt gtggtcgtaa tcgtagcggt atggcttacc tggaacgcgt cgcgggacat    1920
ggtgaccgcg ctgccttgca cgtatctgaa gaaggcaccc gcattgatct ggcggcatta    1980
cttgctgaac cggcgccggg cgtgcaaatc tacgcctgcg gtccgggccg tttattagcg    2040
ggtcttgaag acgcgtctcg taattggccg gatggcgcgc ttcatgtgga gcatttcact    2100
tcgagtttag ccgctttgga tccggatgtc gaacatgcct ttgatttgga gctgcgtgac    2160
tctggcctta ccgttcgcgt cgagccaact cagaccgttt tagacgcttt gcgtgcgaac    2220
aatatcgacg tcccgtcgga ttgcgaagag gggctgtgtg gttcttgcga agtagccgtt    2280
ctggatggcg aggttgatca ccgtgatacc gttctgacta aggccgagcg cgccgcgaat    2340
cgtcagatga tgacttgctg cagtcgtgca tgcggtgatc gtctggtcct gcgcctctaa    2400
```

<210> 46
<211> 799

<212> PRT
<213> 人工序列
<220>
<221> 来源
<223> /注="人工序列的描述：合成的多肽"
<400> 46

```
Met Pro Thr Leu Pro Arg Thr Phe Asp Asp Ile Gln Ser Arg Leu Ile
1               5                  10                  15
Asn Ala Thr Ser Arg Val Val Pro Met Gln Arg Gln Ile Gln Gly Leu
            20                  25                  30
Lys Phe Leu Met Ser Ala Lys Arg Lys Thr Phe Gly Pro Arg Arg Pro
        35                  40                  45
Met Pro Glu Phe Val Glu Thr Pro Ile Pro Asp Val Asn Thr Leu Ala
    50                  55                  60
Leu Glu Asp Ile Asp Val Ser Asn Pro Phe Leu Tyr Arg Gln Gly Gln
65                  70                  75                  80
Trp Arg Ala Tyr Phe Lys Arg Leu Arg Asp Glu Ala Pro Val His Tyr
                85                  90                  95
Gln Lys Asn Ser Pro Phe Gly Pro Phe Trp Ser Val Thr Arg Phe Glu
            100                 105                 110
Asp Ile Leu Phe Val Asp Lys Ser His Asp Leu Phe Ser Ala Glu Pro
        115                 120                 125
Gln Ile Ile Leu Gly Asp Pro Pro Glu Gly Leu Ser Thr Glu Met Phe
    130                 135                 140
Ile Ala Met Asp Pro Pro Lys His Asp Val Gln Arg Ser Ser Val Gln
145                 150                 155                 160
Gly Val Val Ala Pro Lys Asn Leu Lys Glu Met Glu Gly Leu Ile Arg
                165                 170                 175
Ser Arg Thr Gly Asp Val Leu Asp Ser Leu Pro Thr Asp Lys Pro Phe
            180                 185                 190
Asn Trp Val Pro Ala Val Ser Lys Glu Leu Thr Gly Arg Met Leu Ala
        195                 200                 205
Thr Leu Leu Asp Phe Pro Tyr Glu Glu Arg His Lys Leu Val Glu Trp
    210                 215                 220
Ser Asp Arg Met Ala Gly Thr Ala Ser Ala Thr Gly Gly Glu Phe Ala
225                 230                 235                 240
Asp Glu Asn Ala Met Phe Asp Asp Ala Ala Asp Met Ala Arg Ser Phe
                245                 250                 255
Ser Arg Leu Trp Arg Asp Lys Glu Ala Arg Arg Ala Ala Gly Glu Glu
            260                 265                 270
Pro Gly Phe Asp Leu Ile Ser Leu Leu Gln Ser Asn Lys Glu Thr Lys
        275                 280                 285
Asp Leu Ile Asn Arg Pro Met Glu Phe Ile Gly Asn Leu Thr Leu Leu
    290                 295                 300
Ile Val Ala Gly Asn Asp Thr Thr Arg Asn Ser Met Ser Gly Gly Leu
305                 310                 315                 320
Val Ala Met Asn Glu Phe Pro Arg Glu Phe Glu Lys Leu Lys Ala Lys
                325                 330                 335
Pro Glu Leu Ile Pro Asn Met Val Ser Glu Ile Ile Arg Trp Gln Thr
            340                 345                 350
Pro Leu Ala Tyr Met Arg Arg Ile Ala Lys Gln Asp Val Glu Leu Gly
        355                 360                 365
Gly Gln Thr Ile Lys Lys Gly Asp Arg Val Val Met Trp Tyr Ala Ser
    370                 375                 380
Gly Asn Arg Asp Glu Arg Lys Phe Asp Asn Pro Asp Gln Phe Ile Ile
385                 390                 395                 400
Asp Arg Lys Asp Ala Arg Asn His Met Ser Phe Gly Tyr Gly Val His
                405                 410                 415
Arg Cys Met Gly Asn Arg Leu Ala Glu Leu Gln Leu Arg Ile Leu Trp
            420                 425                 430
Glu Glu Ile Leu Lys Arg Phe Asp Asn Ile Glu Val Val Glu Glu Pro
        435                 440                 445
Glu Arg Val Gln Ser Asn Phe Val Arg Gly Tyr Ser Arg Leu Met Val
    450                 455                 460
Lys Leu Thr Pro Asn Ser Val Leu His Arg His Gln Pro Val Thr Ile
465                 470                 475                 480
Gly Glu Pro Ala Ala Arg Ala Val Ser Arg Thr Val Thr Val Glu Arg
                485                 490                 495
Leu Asp Arg Ile Ala Asp Asp Val Leu Arg Leu Val Leu Arg Asp Ala
            500                 505                 510
Gly Gly Lys Thr Leu Pro Thr Trp Thr Pro Gly Ala His Ile Asp Leu
        515                 520                 525
Asp Leu Gly Ala Leu Ser Arg Gln Tyr Ser Leu Cys Gly Ala Pro Asp
    530                 535                 540
Ala Pro Ser Tyr Glu Ile Ala Val His Leu Asp Pro Glu Ser Arg Gly
545                 550                 555                 560
```

[0046]

```
Gly Ser Arg Tyr Ile His Glu Gln Leu Glu Val Gly Ser Pro Leu Arg
                565                 570                 575
Met Arg Gly Pro Arg Asn His Phe Ala Leu Asp Pro Gly Ala Glu His
            580                 585                 590
Tyr Val Phe Val Ala Gly Gly Ile Gly Ile Thr Pro Val Leu Ala Met
        595                 600                 605
Ala Asp His Ala Arg Ala Arg Gly Trp Ser Tyr Glu Leu His Tyr Cys
    610                 615                 620
Gly Arg Asn Arg Ser Gly Met Ala Tyr Leu Glu Arg Val Ala Gly His
625                 630                 635                 640
Gly Asp Arg Ala Ala Leu His Val Ser Glu Glu Gly Thr Arg Ile Asp
                645                 650                 655
Leu Ala Ala Leu Leu Ala Glu Pro Ala Pro Gly Val Gln Ile Tyr Ala
            660                 665                 670
Cys Gly Pro Gly Arg Leu Leu Ala Gly Leu Glu Asp Ala Ser Arg Asn
        675                 680                 685
Trp Pro Asp Gly Ala Leu His Val Glu His Phe Thr Ser Ser Leu Ala
    690                 695                 700
Ala Leu Asp Pro Asp Val Glu His Ala Phe Asp Leu Glu Leu Arg Asp
705                 710                 715                 720
Ser Gly Leu Thr Val Arg Val Glu Pro Thr Gln Thr Val Leu Asp Ala
                725                 730                 735
Leu Arg Ala Asn Asn Ile Asp Val Pro Ser Asp Cys Glu Glu Gly Leu
            740                 745                 750
Cys Gly Ser Cys Glu Val Ala Val Leu Asp Gly Glu Val Asp His Arg
        755                 760                 765
Asp Thr Val Leu Thr Lys Ala Glu Arg Ala Ala Asn Arg Gln Met Met
    770                 775                 780
Thr Cys Cys Ser Arg Ala Cys Gly Asp Arg Leu Val Leu Arg Leu
785                 790                 795
```

<210> 47
<211> 2415
<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"

[0047]

<400> 47

```
atgccaacac tgcccagaac atttgacgac attcagtccc gactgattaa cgccacctcc       60
agggtggtgc cgatgcagag gcaaattcag ggactgaaat tcttaatgag cgccaagagg      120
aagaccttcg gcccacgccg accgatgccc gaattcgttg aaacacccat cccggacgtt      180
aacacgctgg cccttgagga catcgatgtc agcaatccgt ttttataccg gcagggtcag      240
tggcgcgcct atttcaaacg gttgcgtgat gaggcgccgg tccattacca gaagaacagc      300
cctttcggcc ccttctggtc ggtaactcgg tttgaagaca tcctgttcgt ggataagagt      360
cacgacctgt tttccgccga gccgcaaatc attctcggtg accctccgga ggggctgtcg      420
gtggaaatgt tcatagcgat ggatccgccg aaacacgatg tgcagcgcag ctcggtgcag      480
ggagtagtgg caccgaaaaa cctgaaggag atggaggggc tgatccgatc acgcaccggc      540
gatgtgcttg acagcctgcc tacagacaaa ccctttaact gggtacctgc tgtttccaag      600
gaactcacag gccgcatgct ggcgacgctt ctggattttc cttacgagga acgccacaag      660
ctggttgagt ggtcggacag aatggcaggt gcagcatcgg ccaccggcgg ggagtttgcc      720
gatgaaaatg ccatgtttga cgacgcggca gacatggccc ggtctttctc caggctttgg      780
cgggacaagg aggcgcgccg cgcagcaggc gaggagcccg gtttcgattt gatcagcctg      840
ttgcagagca acaaagaaac gaaagacctg atcaatcggc cgatggagtt tatcggtaat      900
ttgacgctgc tcatagtcgc cggcaacgat acgacgcgca actcgatgag tggtggcctg      960
gtggccatga acgaattccc cagggaattt gaaaaattga aggcaaaacc ggagttgatt     1020
ccgaacatgg tgtcggaaat catccgctgg caaacgccgc tggcctatat gcgccgaatc     1080
gccaagcagg atgtcgaact gggcggccag accatcaaga agggtgatcg agttgtcatg     1140
tggtacgcgt cgggtaaccg ggacgagcgc aaatttgaca accccgatca gttcatcatt     1200
gatcgcaagg acgcacgaaa ccacatgtcg ttcggctatg gggttcaccg ttgcatgggc     1260
aaccgtctgg ctgaactgca actgcgcatc ctctgggaag aaatactcaa gcgttttgac     1320
aacatcgaag tcgtcgaaga gcccgagcgg gtgcagtcca acttcgtgcg gggctattcc     1380
aggttgatgg tcaaactgac accgaacagt gtgctggcgc cacgtgacgc ggtccgcatc     1440
ggcgaaccga ctggcggcac caccggtcgc acgctcattg tcgagcgcgt cgagacggcc     1500
gcgcagggtg tgtcgcggat ccgcctggtt tcgcccgatg gtcgcgcact accgcgttgg     1560
tcgccgggct cgcatatcga catcgaatgc ggccacaccg gcatctcgcg ccagtattcg     1620
ctgtgcggcg accccgccga taccagcgcc ttcgagatcg cggtgctgcg cgagcccgaa     1680
agccgtggtg gatcggcgtg gattcatgcc agtctgcgcg caggcgacaa gctcaaggtt     1740
cgcggcccgc gcaatcactt ccggctcgac gagacctgcc gtcgcgcgat cttcatcgcc     1800
ggtggcattg gcgtgactcc ggtcagcgcc atggcaaggc gtgcgaaaga actgggcgtc     1860
gactacacct tccactattg cggccgcagc cgtgcctcca tggcgatgat cgatgaactg     1920
cgcgccctgc atggtgatcg cgtgcggatt catgccgcgg atgaaggcca gcgcgccgat     1980
ctcgcgcaag tgctcggcgc acccgatgcg aacacgcaga tctacgcttg tggaccggcc     2040
cggatgatcg aggcgctgga agccctgtgc gcgacatggc ccgaggattc gctgcgcgtc     2100
gaacacttca gttcgaaact tggaacgctc gatccctcca gggaacagcc gtttgcggtc     2160
gaactgaagg actcggggct gacgcttgaa gttcctccgg accagacgct gctcgccacc     2220
```

```
ctgcgcgccg cgaacatcga cgtgcaaagc gattgcgagg aaggcctgtg tggatcgtgt     2280
gaagtgcgcg tgctggccgg cgagatcgac caccgcgacg tcgtgctgac gcgcggcgag     2340
cgtgatgcga acaaccggat gatggcctgc tgctcgcgag cggcgaaggg cggaaagatc     2400
gtgctggggc tgtaa                                                     2415
```

<210> 48
<211> 804
<212> PRT
<213> 人工序列
<220>
<221> 来源
<223> /注="人工序列的描述：合成的多肽"
<400> 48

```
Met Pro Thr Leu Pro Arg Thr Phe Asp Asp Ile Gln Ser Arg Leu Ile
1               5                  10                  15
Asn Ala Thr Ser Arg Val Val Pro Met Gln Arg Gln Ile Gln Gly Leu
            20                  25                  30
Lys Phe Leu Met Ser Ala Lys Arg Lys Thr Phe Gly Pro Arg Arg Pro
        35                  40                  45
Met Pro Glu Phe Val Glu Thr Pro Ile Pro Asp Val Asn Thr Leu Ala
    50                  55                  60
Leu Glu Asp Ile Asp Val Ser Asn Pro Phe Leu Tyr Arg Gln Gly Gln
65                  70                  75                  80
Trp Arg Ala Tyr Phe Lys Arg Leu Arg Asp Glu Ala Pro Val His Tyr
                85                  90                  95
Gln Lys Asn Ser Pro Phe Gly Pro Phe Trp Ser Val Thr Arg Phe Glu
            100                 105                 110
Asp Ile Leu Phe Val Asp Lys Ser His Asp Leu Phe Ser Ala Glu Pro
        115                 120                 125
Gln Ile Ile Leu Gly Asp Pro Pro Glu Gly Leu Ser Val Glu Met Phe
    130                 135                 140
Ile Ala Met Asp Pro Pro Lys His Asp Val Gln Arg Ser Ser Val Gln
145                 150                 155                 160
Gly Val Val Ala Pro Lys Asn Leu Lys Glu Met Glu Gly Leu Ile Arg
                165                 170                 175
Ser Arg Thr Gly Asp Val Leu Asp Ser Leu Pro Thr Asp Lys Pro Phe
            180                 185                 190
Asn Trp Val Pro Ala Val Ser Lys Glu Leu Thr Gly Arg Met Leu Ala
        195                 200                 205
Thr Leu Leu Asp Phe Pro Tyr Glu Glu Arg His Lys Leu Val Glu Trp
    210                 215                 220
Ser Asp Arg Met Ala Gly Ala Ala Ser Ala Thr Gly Gly Glu Phe Ala
225                 230                 235                 240
Asp Glu Asn Ala Met Phe Asp Asp Ala Ala Asp Met Ala Arg Ser Phe
                245                 250                 255
Ser Arg Leu Trp Arg Asp Lys Glu Ala Arg Arg Ala Ala Gly Glu Glu
            260                 265                 270
Pro Gly Phe Asp Leu Ile Ser Leu Leu Gln Ser Asn Lys Glu Thr Lys
        275                 280                 285
Asp Leu Ile Asn Arg Pro Met Glu Phe Ile Gly Asn Leu Thr Leu Leu
    290                 295                 300
Ile Val Ala Gly Asn Asp Thr Thr Arg Asn Ser Met Ser Gly Gly Leu
305                 310                 315                 320
Val Ala Met Asn Glu Phe Pro Arg Glu Phe Glu Lys Leu Lys Ala Lys
                325                 330                 335
Pro Glu Leu Ile Pro Asn Met Val Ser Glu Ile Ile Arg Trp Gln Thr
            340                 345                 350
Pro Leu Ala Tyr Met Arg Arg Ile Ala Lys Gln Asp Val Glu Leu Gly
        355                 360                 365
Gly Gln Thr Ile Lys Lys Gly Asp Arg Val Val Met Trp Tyr Ala Ser
    370                 375                 380
Gly Asn Arg Asp Glu Arg Lys Phe Asp Asn Pro Asp Gln Phe Ile Ile
385                 390                 395                 400
Asp Arg Lys Asp Ala Arg Asn His Met Ser Phe Gly Tyr Gly Val His
                405                 410                 415
Arg Cys Met Gly Asn Arg Leu Ala Glu Leu Gln Leu Arg Ile Leu Trp
            420                 425                 430
Glu Glu Ile Leu Lys Arg Phe Asp Asn Ile Glu Val Val Glu Glu Pro
        435                 440                 445
Glu Arg Val Gln Ser Asn Phe Val Arg Gly Tyr Ser Arg Leu Met Val
    450                 455                 460
Lys Leu Thr Pro Asn Ser Val Leu Ala Pro Arg Asp Ala Val Arg Ile
465                 470                 475                 480
Gly Glu Pro Thr Gly Gly Thr Thr Gly Arg Thr Leu Ile Val Glu Arg
                485                 490                 495
Val Glu Thr Ala Ala Gln Gly Val Ser Arg Ile Arg Leu Val Ser Pro
```

[0048]

```
            500                 505                 510
Asp Gly Arg Ala Leu Pro Arg Trp Ser Pro Gly Ser His Ile Asp Ile
        515                 520                 525
Glu Cys Gly His Thr Gly Ile Ser Arg Gln Tyr Ser Leu Cys Gly Asp
    530                 535                 540
Pro Ala Asp Thr Ser Ala Phe Glu Ile Ala Val Leu Arg Glu Pro Glu
545                 550                 555                 560
Ser Arg Gly Gly Ser Ala Trp Ile His Ala Ser Leu Arg Ala Gly Asp
                565                 570                 575
Lys Leu Lys Val Arg Gly Pro Arg Asn His Phe Arg Leu Asp Glu Thr
            580                 585                 590
Cys Arg Arg Ala Ile Phe Ile Ala Gly Gly Ile Gly Val Thr Pro Val
        595                 600                 605
Ser Ala Met Ala Arg Arg Ala Lys Glu Leu Gly Val Asp Tyr Thr Phe
    610                 615                 620
His Tyr Cys Gly Arg Ser Arg Ala Ser Met Ala Met Ile Asp Glu Leu
625                 630                 635                 640
Arg Ala Leu His Gly Asp Arg Val Arg Ile His Ala Ala Asp Glu Gly
                645                 650                 655
Gln Arg Ala Asp Leu Ala Gln Val Leu Gly Ala Pro Asp Ala Asn Thr
            660                 665                 670
Gln Ile Tyr Ala Cys Gly Pro Ala Arg Met Ile Glu Ala Leu Glu Ala
        675                 680                 685
Leu Cys Ala Thr Trp Pro Glu Asp Ser Leu Arg Val Glu His Phe Ser
    690                 695                 700
Ser Lys Leu Gly Thr Leu Asp Pro Ser Arg Glu Gln Pro Phe Ala Val
705                 710                 715                 720
Glu Leu Lys Asp Ser Gly Leu Thr Leu Glu Val Pro Pro Asp Gln Thr
                725                 730                 735
Leu Leu Ala Thr Leu Arg Ala Ala Asn Ile Asp Val Gln Ser Asp Cys
            740                 745                 750
Glu Glu Gly Leu Cys Gly Ser Cys Glu Val Arg Val Leu Ala Gly Glu
        755                 760                 765
Ile Asp His Arg Asp Val Val Leu Thr Arg Gly Glu Arg Asp Ala Asn
    770                 775                 780
Asn Arg Met Met Ala Cys Cys Ser Arg Ala Ala Lys Gly Gly Lys Ile
785                 790                 795                 800
Val Leu Gly Leu
```

[0049]

<210> 49
<211> 2400
<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> 49

```
atgccaacac tgcccagaac atttgacgac attcagtccc gactgattaa cgccacctcc      60
agggtggtgc cgatgcagag gcaaattcag ggactgaaat tcttaatgag cgccaagagg     120
aagaccttcg gcccacgccg accgatgccc gaattcgttg aaacacccat cccggacgtt     180
aacacgctgg cccttgagga catcgatgtc agcaatccgt ttttataccg gcagggtcag     240
tggcgcgcct atttcaaacg gttgcgtgat gaggcgccgg tccattacca gaagaacagc     300
cctttcggcc ccttctggtc ggtaactcgg tttgaagaca tcctgttcgt ggataagagt     360
cacgacctgt tttccgccga gccgcaaatc attctcggtg accctccgga ggggctgtcg     420
gtggaaatgt tcatagcgat ggatccgccg aaacacgatg tgcagcgcag ctcggtgcag     480
ggagtagtgg caccgaaaaa cctgaaggag atggaggggc tgatccgatc acgcaccggc     540
gatgtgcttg acagcctgcc tacagacaaa ccctttaact gggtacctgc tgtttccaag     600
gaactcacag gccgcatgct ggcgacgctt ctggattttc cttacgagga acgccacaag     660
ctggttgagt ggtcggacag aatggcaggt gcagcatcgg ccaccggcgg ggagtttgcc     720
gatgaaaatg ccatgtttga cgacgcggca gacatggccc ggtctttctc caggctttgg     780
cgggacaagg aggcgcgccg cgcagcaggc gaggagcccg gtttcgattt gatcagcctg     840
ttgcagagca acaaagaaac gaaagacctg atcaatcggc cgatggagtt tatcggtaat     900
ttgacgctgc tcatagtcgc cggcaacgat acgacgcgca actcgatgag tggtggcctg     960
gtggccatga acgaattccc cagggaattt gaaaaattga aggcaaaacc ggagttgatt    1020
ccgaacatgg tgtcggaaat catccgctgg caaacgccgc tggcctatat gcgccgaatc    1080
gccaagcagg atgtcgaact gggcggccag accatcaaga agggtgatcg agttgtcatg    1140
tggtacgcgt cgggtaaccg ggacgagcgc aaatttgaca accccgatca gttcatcatt    1200
gatcgcaagg acgcacgaaa ccacatgtcg ttcggctatg gggttcaccg ttgcatgggc    1260
aaccgtctgg ctgaactgca actgcgcatc ctctgggaag aaatactcaa gcgttttgac    1320
aacatcgaag tcgtcgaaga gcccgagcgg gtgcagtcca acttcgtgcg gggctattcc    1380
aggttgatgg tcaaactgac accgaacagt actgccagac aaattaccag ttttcctatt    1440
ggagcgcctg atctcaaggc catgacccgc ccggtacgtg tccaggcagt ttatcctgaa    1500
gcagaagata ttatacgtat tgagctggct gctatacacg gtgaagaatt accacgctgg    1560
agtgccggtt cacatattga actggtattg cctaatggcc tgagccgtaa gtactctcta    1620
tgcggtttag cgactgatca attttatacg attgctgtaa agagagagca ggaaagccgg    1680
```

```
ggtggttcac aatggattca tcagtattta aaagcaggag aacagatcta cattaaggga     1740
cctaaaaact tttttaaatt aaatttacag gcgagccagt atgtgctgat cgcaggtggg     1800
ataggtatta ctcctattct gagcatggcc agcagcttac gtgaacaggg gcgcccttac     1860
cggctaattt atctttcacg ccagcgggct agtatggcat tacttaagga agttgctgcg     1920
catggtagtg ctgccgaact ctatatttct tctgaaggta aacggataga cctgcaacag     1980
ctcttgtcag cgctgcctgc cggtacacag gtttgtgcct gtggtccgga agcattactt     2040
gataccttga ccaactatac cgaagacttg tcacaggttc agctcacggt cgagcatttt     2100
ggttcaggga agaacctctt tttatatgaa aatgataccg actttgaagt tgaactgctg     2160
gatagcggtt taacgctgac agtagcccgg gatcagacct tattggactg tttgctggac     2220
aaaggaattg atgtcagctt tgactgtacc gagggtttat gtggaagctg tcagcttccg     2280
gttgaagaag gtgaaattga tcatcgcgac aaggtgttga cccgagccga gcgtgacgga     2340
atgaaatcgg taatcagttg ctgttcacga ggaaaaggaa aactcaaact caagctttaa     2400
```

<210> 50
<211> 799
<212> PRT
<213> 人工序列
<220>
<221> 来源
<223> /注="人工序列的描述：合成的多肽"
<400> 50

[0050]

```
Met Pro Thr Leu Pro Arg Thr Phe Asp Asp Ile Gln Ser Arg Leu Ile
1               5                  10                  15
Asn Ala Thr Ser Arg Val Val Pro Met Gln Arg Gln Ile Gln Gly Leu
            20                  25                  30
Lys Phe Leu Met Ser Ala Lys Arg Lys Thr Phe Gly Pro Arg Arg Pro
        35                  40                  45
Met Pro Glu Phe Val Glu Thr Pro Ile Pro Asp Val Asn Thr Leu Ala
    50                  55                  60
Leu Glu Asp Ile Asp Val Ser Asn Pro Phe Leu Tyr Arg Gln Gly Gln
65                  70                  75                  80
Trp Arg Ala Tyr Phe Lys Arg Leu Arg Asp Glu Ala Pro Val His Tyr
                85                  90                  95
Gln Lys Asn Ser Pro Phe Gly Pro Phe Trp Ser Val Thr Arg Phe Glu
            100                 105                 110
Asp Ile Leu Phe Val Asp Lys Ser His Asp Leu Phe Ser Ala Glu Pro
        115                 120                 125
Gln Ile Ile Leu Gly Asp Pro Pro Glu Gly Leu Ser Val Glu Met Phe
    130                 135                 140
Ile Ala Met Asp Pro Pro Lys His Asp Val Gln Arg Ser Ser Val Gln
145                 150                 155                 160
Gly Val Val Ala Pro Lys Asn Leu Lys Glu Met Glu Gly Leu Ile Arg
                165                 170                 175
Ser Arg Thr Gly Asp Val Leu Asp Ser Leu Pro Thr Asp Lys Pro Phe
            180                 185                 190
Asn Trp Val Pro Ala Val Ser Lys Glu Leu Thr Gly Arg Met Leu Ala
        195                 200                 205
Thr Leu Leu Asp Phe Pro Tyr Glu Glu Arg His Lys Leu Val Glu Trp
    210                 215                 220
Ser Asp Arg Met Ala Gly Ala Ala Ser Ala Thr Gly Gly Glu Phe Ala
225                 230                 235                 240
Asp Glu Asn Ala Met Phe Asp Asp Ala Ala Asp Met Ala Arg Ser Phe
                245                 250                 255
Ser Arg Leu Trp Arg Asp Lys Glu Ala Arg Arg Ala Ala Gly Glu Glu
            260                 265                 270
Pro Gly Phe Asp Leu Ile Ser Leu Leu Gln Ser Asn Lys Glu Thr Lys
        275                 280                 285
Asp Leu Ile Asn Arg Pro Met Glu Phe Ile Gly Asn Leu Thr Leu Leu
    290                 295                 300
Ile Val Ala Gly Asn Asp Thr Thr Arg Asn Ser Met Ser Gly Gly Leu
305                 310                 315                 320
Val Ala Met Asn Glu Phe Pro Arg Glu Phe Glu Lys Leu Lys Ala Lys
                325                 330                 335
Pro Glu Leu Ile Pro Asn Met Val Ser Glu Ile Ile Arg Trp Gln Thr
            340                 345                 350
Pro Leu Ala Tyr Met Arg Arg Ile Ala Lys Gln Asp Val Glu Leu Gly
        355                 360                 365
Gly Gln Thr Ile Lys Lys Gly Asp Arg Val Val Met Trp Tyr Ala Ser
    370                 375                 380
Gly Asn Arg Asp Glu Arg Lys Phe Asp Asn Pro Asp Gln Phe Ile Ile
385                 390                 395                 400
Asp Arg Lys Asp Ala Arg Asn His Met Ser Phe Gly Tyr Gly Val His
                405                 410                 415
Arg Cys Met Gly Asn Arg Leu Ala Glu Leu Gln Leu Arg Ile Leu Trp
            420                 425                 430
Glu Glu Ile Leu Lys Arg Phe Asp Asn Ile Glu Val Val Glu Glu Pro
```

```
        435                 440                 445
Glu Arg Val Gln Ser Asn Phe Val Arg Gly Tyr Ser Arg Leu Met Val
    450                 455                 460
Lys Leu Thr Pro Asn Ser Thr Ala Arg Gln Ile Thr Ser Phe Pro Ile
465                 470                 475                 480
Gly Ala Pro Asp Leu Lys Ala Met Thr Arg Pro Val Arg Val Gln Ala
                485                 490                 495
Val Tyr Pro Glu Ala Glu Asp Ile Ile Arg Ile Glu Leu Ala Ala Ile
            500                 505                 510
His Gly Glu Glu Leu Pro Arg Trp Ser Ala Gly Ser His Ile Glu Leu
        515                 520                 525
Val Leu Pro Asn Gly Leu Ser Arg Lys Tyr Ser Leu Cys Gly Leu Ala
    530                 535                 540
Thr Asp Gln Phe Tyr Thr Ile Ala Val Lys Arg Glu Gln Glu Ser Arg
545                 550                 555                 560
Gly Gly Ser Gln Trp Ile His Gln Tyr Leu Lys Ala Gly Glu Gln Ile
                565                 570                 575
Tyr Ile Lys Gly Pro Lys Asn Phe Phe Lys Leu Asn Leu Gln Ala Ser
            580                 585                 590
Gln Tyr Val Leu Ile Ala Gly Gly Ile Gly Ile Thr Pro Ile Leu Ser
        595                 600                 605
Met Ala Ser Ser Leu Arg Glu Gln Gly Arg Pro Tyr Arg Leu Ile Tyr
    610                 615                 620
Leu Ser Arg Gln Arg Ala Ser Met Ala Leu Leu Lys Glu Val Ala Ala
625                 630                 635                 640
His Gly Ser Ala Ala Glu Leu Tyr Ile Ser Ser Glu Gly Lys Arg Ile
                645                 650                 655
Asp Leu Gln Gln Leu Leu Ser Ala Leu Pro Ala Gly Thr Gln Val Cys
            660                 665                 670
Ala Cys Gly Pro Glu Ala Leu Leu Asp Thr Leu Thr Asn Tyr Thr Glu
        675                 680                 685
Asp Leu Ser Gln Val Gln Leu Thr Val Glu His Phe Gly Ser Gly Lys
    690                 695                 700
Asn Leu Phe Leu Tyr Glu Asn Asp Thr Asp Phe Glu Val Glu Leu Leu
705                 710                 715                 720
Asp Ser Gly Leu Thr Leu Thr Val Ala Arg Asp Gln Thr Leu Leu Asp
                725                 730                 735
Cys Leu Leu Asp Lys Gly Ile Asp Val Ser Phe Asp Cys Thr Glu Gly
            740                 745                 750
Leu Cys Gly Ser Cys Gln Leu Pro Val Glu Glu Gly Glu Ile Asp His
        755                 760                 765
Arg Asp Lys Val Leu Thr Arg Ala Glu Arg Asp Gly Met Lys Ser Val
    770                 775                 780
Ile Ser Cys Cys Ser Arg Gly Lys Gly Lys Leu Lys Leu Lys Leu
785                 790                 795
```

[0051]

<210> 51
<211> 2238
<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> 51

```
atggaacata caggacaaag cgcggcggcg acgatgccgc tcgattcgat cgacgtgagt       60
atcccagaac tcttctacaa cgacagtgtt ggcgaatatt tcaagcggct gcgcaaggat      120
gatccagtcc actattgcgc cgacagcgcg tttgggccct actggtccat cacgaaatat      180
aatgacataa tgcacgtcga cacaaatcat gacatcttct cgtccgatgc aggatatggc      240
ggcatcataa tcgatgacgg cattcaaaaa ggtggcgatg gcggactgga tcttcccaat      300
ttcatcgcga tggatcggcc gcgacacgac gaacaaagaa aagctgtaag cccgattgta      360
gcccccgcca atttggccgc gcttgaaggc accattcgcg aacgagttag caaaacgctc      420
gatggtcttc cggtgggtga ggagttcgat tgggtagatc gcgtgtcgat cgaaatcacc      480

actcaaatgc tcgccaccct gttcgacttt ccgtttgaag agcgccgcaa gcttacccgc      540
tggtcggatg tgacaaccgc agcacccggc ggcggagttg tcgaaagctg ggatcagaga      600
aaaaccgaat tgttggaatg tgccgcttat ttccaggtgc tttggaatga gcgtgtcaac      660
aaggaccccg gcaacgatct catttcgatg ctggcacatt cgccagccac gcggaacatg      720
acgcccgaag agtatctggg caatgtactt ctcctgatcg ttgccgggaa cgataccaca      780
cgcaattcga tgaccggtgg cgttttagct ctccacaaga atccggacca gtttgccaag      840
ctaaaagcca accctgcgct ggttgaaacg atggtccctg aaatcattcg ctggcaaaca      900
ccgcttgctc atatgcgccg cacggccatt gcagattccg aactgggtgg gaagaccatc      960
cgcaagggcg acaaggtcgt catgtggtat tattcgggta atcgcgatga cgaagtgatt     1020
gaccgtcccg aagaatttat catcgaccgc ccccggcccc gccagcattt gtcattcggc     1080
tttggcatcc accgttgcgt tggcaatcgg ctagccgaaa tgcagctccg gatattgtgg     1140
gaagaaattc tcacgcgttt cagtcgtatc gaagtgatgg ccgaaccgga acgggtccgt     1200
tcaaattttg tgcgcggtta cgccaagatg atggttcgcg tccacgcggt actccatcgt     1260
```

```
catcaacctg tcaccatcgg cgagccggcc gctcgtgctg tgagccgcac ggtgaccgtt     1320
gagcgtcttg atcgcattgc cgacgatgtc cttcgcctgg tccttcgcga tgctggaggt     1380
aaaaccctcc cgacgtggac gcctggcgct cacatcgacc tggatctggg tgctctgagc     1440
cgtcagtatt cgctctgcgg cgctccggat gctccgtcgt acgaaatcgc cgtgcactta     1500
gatccggaaa gccgtggtgg aagccgctat attcatgaac agctggaagt tggaagtccg     1560
ctgcgtatgc gtggcccacg caaccatttc gccctggatc cgggtgcgga acattacgtg     1620
tttgttgccg ggggtatcgg catcacgccg gtgctggcaa tggcggatca tgcccgtgcg     1680
cgtggttggt cgtacgaact gcattattgt ggtcgtaatc gtagcggtat ggcttacctg     1740
gaacgcgtcg cgggacatgg tgaccgcgct gccttgcacg tatctgaaga aggcacccgc     1800
attgatctgg cggcattact tgctgaaccg gcgccgggcg tgcaaatcta cgcctgcggt     1860
ccgggccgtt tattagcggg tcttgaagac gcgtctcgta attggccgga tggcgcgctt     1920
catgtggagc atttcacttc gagtttagcc gctttggatc cggatgtcga acatgccttt     1980
gatttggagc tgcgtgactc tggccttacc gttcgcgtcg agccaactca gaccgtttta     2040
gacgctttgc gtgcgaacaa tatcgacgtc ccgtcggatt gcgaagaggg gctgtgtggt     2100
tcttgcgaag tagccgttct ggatggcgag gttgatcacc gtgataccgt tctgactaag     2160
gccgagcgcg ccgcgaatcg tcagatgatg acttgctgca gtcgtgcatg cggtgatcgt     2220
ctggcgctgc gcctctaa                                                  2238
```

<210> 52
<211> 745
<212> PRT
<213> 人工序列
<220>
<221> 来源
<223> /注="人工序列的描述：合成的多肽"
<400> 52

[0052]

```
Met Glu His Thr Gly Gln Ser Ala Ala Ala Thr Met Pro Leu Asp Ser
1               5                  10                  15
Ile Asp Val Ser Ile Pro Glu Leu Phe Tyr Asn Asp Ser Val Gly Glu
            20                  25                  30
Tyr Phe Lys Arg Leu Arg Lys Asp Asp Pro Val His Tyr Cys Ala Asp
        35                  40                  45
Ser Ala Phe Gly Pro Tyr Trp Ser Ile Thr Lys Tyr Asn Asp Ile Met
    50                  55                  60
His Val Asp Thr Asn His Asp Ile Phe Ser Ser Asp Ala Gly Tyr Gly
65                  70                  75                  80
Gly Ile Ile Ile Asp Asp Gly Ile Gln Lys Gly Gly Asp Gly Gly Leu
                85                  90                  95
Asp Leu Pro Asn Phe Ile Ala Met Asp Arg Pro Arg His Asp Glu Gln
            100                 105                 110
Arg Lys Ala Val Ser Pro Ile Val Ala Pro Ala Asn Leu Ala Ala Leu
        115                 120                 125
Glu Gly Thr Ile Arg Glu Arg Val Ser Lys Thr Leu Asp Gly Leu Pro
    130                 135                 140
Val Gly Glu Glu Phe Asp Trp Val Asp Arg Val Ser Ile Glu Ile Thr
145                 150                 155                 160
Thr Gln Met Leu Ala Thr Leu Phe Asp Phe Pro Phe Glu Glu Arg Arg
                165                 170                 175
Lys Leu Thr Arg Trp Ser Asp Val Thr Thr Ala Ala Pro Gly Gly Gly
            180                 185                 190
Val Val Glu Ser Trp Asp Gln Arg Lys Thr Glu Leu Leu Glu Cys Ala
        195                 200                 205
Ala Tyr Phe Gln Val Leu Trp Asn Glu Arg Val Asn Lys Asp Pro Gly
    210                 215                 220
Asn Asp Leu Ile Ser Met Leu Ala His Ser Pro Ala Thr Arg Asn Met
225                 230                 235                 240
Thr Pro Glu Glu Tyr Leu Gly Asn Val Leu Leu Leu Ile Val Ala Gly
                245                 250                 255
Asn Asp Thr Thr Arg Asn Ser Met Thr Gly Gly Val Leu Ala Leu His
            260                 265                 270
Lys Asn Pro Asp Gln Phe Ala Lys Leu Lys Ala Asn Pro Ala Leu Val
        275                 280                 285
Glu Thr Met Val Pro Glu Ile Ile Arg Trp Gln Thr Pro Leu Ala His
    290                 295                 300
Met Arg Arg Thr Ala Ile Ala Asp Ser Glu Leu Gly Gly Lys Thr Ile
305                 310                 315                 320
Arg Lys Gly Asp Lys Val Val Met Trp Tyr Tyr Ser Gly Asn Arg Asp
                325                 330                 335
Asp Glu Val Ile Asp Arg Pro Glu Glu Phe Ile Ile Asp Arg Pro Arg
            340                 345                 350
Pro Arg Gln His Leu Ser Phe Gly Phe Gly Ile His Arg Cys Val Gly
        355                 360                 365
Asn Arg Leu Ala Glu Met Gln Leu Arg Ile Leu Trp Glu Glu Ile Leu
    370                 375                 380
Thr Arg Phe Ser Arg Ile Glu Val Met Ala Glu Pro Glu Arg Val Arg
385                 390                 395                 400
```

```
Ser Asn Phe Val Arg Gly Tyr Ala Lys Met Met Val Arg Val His Ala
                405                 410                 415
Val Leu His Arg His Gln Pro Val Thr Ile Gly Glu Pro Ala Ala Arg
            420                 425                 430
Ala Val Ser Arg Thr Val Thr Val Glu Arg Leu Asp Arg Ile Ala Asp
        435                 440                 445
Asp Val Leu Arg Leu Val Leu Arg Asp Ala Gly Gly Lys Thr Leu Pro
    450                 455                 460
Thr Trp Thr Pro Gly Ala His Ile Asp Leu Asp Leu Gly Ala Leu Ser
465                 470                 475                 480
Arg Gln Tyr Ser Leu Cys Gly Ala Pro Asp Ala Pro Ser Tyr Glu Ile
                485                 490                 495
Ala Val His Leu Asp Pro Glu Ser Arg Gly Gly Ser Arg Tyr Ile His
            500                 505                 510
Glu Gln Leu Glu Val Gly Ser Pro Leu Arg Met Arg Gly Pro Arg Asn
        515                 520                 525
His Phe Ala Leu Asp Pro Gly Ala Glu His Tyr Val Phe Val Ala Gly
    530                 535                 540
Gly Ile Gly Ile Thr Pro Val Leu Ala Met Ala Asp His Ala Arg Ala
545                 550                 555                 560
Arg Gly Trp Ser Tyr Glu Leu His Tyr Cys Gly Arg Asn Arg Ser Gly
                565                 570                 575
Met Ala Tyr Leu Glu Arg Val Ala Gly His Gly Asp Arg Ala Ala Leu
            580                 585                 590
His Val Ser Glu Glu Gly Thr Arg Ile Asp Leu Ala Ala Leu Leu Ala
        595                 600                 605
Glu Pro Ala Pro Gly Val Gln Ile Tyr Ala Cys Gly Pro Gly Arg Leu
    610                 615                 620
Leu Ala Gly Leu Glu Asp Ala Ser Arg Asn Trp Pro Asp Gly Ala Leu
625                 630                 635                 640
His Val Glu His Phe Thr Ser Ser Leu Ala Ala Leu Asp Pro Asp Val
                645                 650                 655
Glu His Ala Phe Asp Leu Glu Leu Arg Asp Ser Gly Leu Thr Val Arg
            660                 665                 670
Val Glu Pro Thr Gln Thr Val Leu Asp Ala Leu Arg Ala Asn Asn Ile
        675                 680                 685
Asp Val Pro Ser Asp Cys Glu Glu Gly Leu Cys Gly Ser Cys Glu Val
    690                 695                 700
Ala Val Leu Asp Gly Glu Val Asp His Arg Asp Thr Val Leu Thr Lys
705                 710                 715                 720
Ala Glu Arg Ala Ala Asn Arg Gln Met Met Thr Cys Cys Ser Arg Ala
                725                 730                 735
Cys Gly Asp Arg Leu Ala Leu Arg Leu
            740                 745
```

[0053]

<210> 53
<211> 3180
<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> 53

```
atgccaacac tgcccagaac atttgacgac attcagtccc gactgattaa cgccacctcc       60
agggtggtgc cgatgcagag gcaaattcag ggactgaaat tcttaatgag cgccaagagg      120
aagaccttcg gcccacgccg accgatgccc gaattcgttg aaacacccat cccggacgtt      180
aacacgctgg cccttgagga catcgatgtc agcaatccgt ttttataccg gcagggtcag      240
tggcgcgcct atttcaaacg gttgcgtgat gaggcgccgg tccattacca gaagaacagc      300
cctttcggcc ccttctggtc ggtaactcgg tttgaagaca tcctgttcgt ggataagagt      360
cacgacctgt tttccgccga gccgcaaatc attctcggtg accctccgga ggggctgtcg      420
gtggaaatgt tcatagcgat ggatccgccg aaacacgatg tgcagcgcag ctcggtgcag      480
ggagtagtgg caccgaaaaa cctgaaggag atggaggggc tgatccgatc acgcaccggc      540
gatgtgcttg acagcctgcc tacagacaaa ccctttaact gggtacctgc tgtttccaag      600
gaactcacag gccgcatgct ggcgacgctt ctggattttc cttacgagga acgccacaag      660
ctggttgagt ggtcggacag aatggcaggt gcagcatcgg ccaccggcgg ggagtttgcc      720
gatgaaaatg ccatgtttga cgacgcggca gacatggccc ggtctttctc caggctttgg      780
cgggacaagg aggcgcgccg cgcagcaggc gaggagcccg gtttcgattt gatcagcctg      840
ttgcagagca acaaagaaac gaaagacctg atcaatcggc cgatggagtt tatcggtaat      900
ttgacgctgc tcatagtcgc cggcaacgat acgacgcgca actcgatgag tggtggcctg      960
gtggccatga acgaattccc cagggaattt gaaaaattga aggcaaaacc ggagttgatt     1020
ccgaacatgg tgtcggaaat catccgctgg caaacgccgc tggcctatat gcgccgaatc     1080
gccaagcagg atgtcgaact gggcggccag accatcaaga agggtgatcg agttgtcatg     1140
tggtacgcgt cgggtaaccg ggacgagcgc aaatttgaca accccgatca gttcatcatt     1200
gatcgcaagg acgcacgaaa ccacatgtcg ttcggctatg gggttcaccg ttgcatgggc     1260
aaccgtctgg ctgaactgca actgcgcatc ctctgggaag aaatactcaa gcgttttgac     1320
aacatcgaag tcgtcgaaga gcccgagcgg gtgcagtcca acttcgtgcg gggctattcc     1380
```

```
aggttgatgg tcaaactgac accgaacagt tcacctagca ctgaacagtc tgctaaaaaa     1440
gtacgcaaaa aggcagaaaa cgctcataat acgccgctgc ttgtgctata cggttcaaat     1500
atgggaacag ctgaaggaac ggcgcgtgat ttagcagata ttgcaatgag caaaggattt     1560
gcaccgcagg tcgcaacgct tgattcacac gccggaaatc ttccgcgcga aggagctgta     1620
ttaattgtaa cggcgtctta taacggtcat ccgcctgata acgcaaagca atttgtcgac     1680
tggttagacc aagcgtctgc tgatgaagta aaaggcgttc gctactccgt atttggatgc     1740
ggcgataaaa actgggctac tacgtatcaa aaagtgcctg cttttatcga tgaaacgctt     1800
gccgctaaag gggcagaaaa catcgctgac cgcggtgaag cagatgcaag cgacgacttt     1860
gaaggcacat atgaagaatg gcgtgaacat atgtggagtg acgtagcagc ctactttaac     1920
ctcgacattg aaaacagtga agataataaa tctactcttt cacttcaatt tgtcgacagc     1980
gccgcggata tgccgcttgc gaaaatgcac ggtgcgtttt caacgaacgt cgtagcaagc     2040
aaagaacttc aacagccagg cagtgcacga agcacgcgac atcttgaaat tgaacttcca     2100
aaagaagctt cttatcaaga aggagatcat ttaggtgtta ttcctcgcaa ctatgaagga     2160
atagtaaacc gtgtaacagc aaggttcggc ctagatgcat cacagcaaat ccgtctggaa     2220
gcagaagaag aaaaattagc tcatttgcca ctcgctaaaa cagtatccgt agaagagctt     2280
ctgcaatacg tggagcttca agatcctgtt acgcgcacgc agcttcgcgc aatggctgct     2340
aaaacggtct gcccgccgca taaagtagag cttgaagcct tgcttgaaaa gcaagcctac     2400
aaagaacaag tgctggcaaa acgtttaaca atgcttgaac tgcttgaaaa atacccggcg     2460
tgtgaaatga aattcagcga atttatcgcc cttctgccaa gcatacgccc gcgctattac     2520
tcgatttctt catcacctcg tgtcgatgaa aaacaagcaa gcatcacggt cagcgttgtc     2580
tcaggagaag cgtggagcgg atatggagaa tataaaggaa ttgcgtcgaa ctatcttgcc     2640
gagctgcaag aaggagatac gattacgtgc tttatttcca caccgcagtc agaatttacg     2700
ctgccaaaag accctgaaac gccgcttatc atggtcggac cgggaacagg cgtcgcgccg     2760
tttagaggct ttgtgcaggc gcgcaaacag ctaaaagaac aaggacagtc acttggagaa     2820
gcacatttat acttcggctg ccgttcacct catgaagact atctgtatca agaagagctt     2880
gaaaacgccc aaagcgaagg catcattacg cttcataccg ctttttctcg catgccaaat     2940
cagccgaaaa catacgttca gcacgtaatg gaacaagacg gcaagaaatt gattgaactt     3000
cttgatcaag gagcgcactt ctatatttgc ggagacggaa gccaaatggc acctgccgtt     3060
gaagcaacgc ttatgaaaag ctatgctgac gttcaccaag tgagtgaagc agacgctcgc     3120
ttatggctgc agcagctaga agaaaaaggc cgatacgcaa aagacgtgtg ggctgggtaa     3180
```

<210> 54
<211> 1059
<212> PRT
<213> 人工序列
<220>
<221> 来源
<223> /注="人工序列的描述：合成的多肽"

[0054]

<400> 54

```
Met Pro Thr Leu Pro Arg Thr Phe Asp Asp Ile Gln Ser Arg Leu Ile
1               5                  10                  15
Asn Ala Thr Ser Arg Val Val Pro Met Gln Arg Gln Ile Gln Gly Leu
            20                  25                  30
Lys Phe Leu Met Ser Ala Lys Arg Lys Thr Phe Gly Pro Arg Arg Pro
        35                  40                  45
Met Pro Glu Phe Val Glu Thr Pro Ile Pro Asp Val Asn Thr Leu Ala
    50                  55                  60
Leu Glu Asp Ile Asp Val Ser Asn Pro Phe Leu Tyr Arg Gln Gly Gln
65                  70                  75                  80
Trp Arg Ala Tyr Phe Lys Arg Leu Arg Asp Glu Ala Pro Val His Tyr
                85                  90                  95
Gln Lys Asn Ser Pro Phe Gly Pro Phe Trp Ser Val Thr Arg Phe Glu
            100                 105                 110
Asp Ile Leu Phe Val Asp Lys Ser His Asp Leu Phe Ser Ala Glu Pro
        115                 120                 125
Gln Ile Ile Leu Gly Asp Pro Pro Glu Gly Leu Ser Val Glu Met Phe
    130                 135                 140
Ile Ala Met Asp Pro Pro Lys His Asp Val Gln Arg Ser Ser Val Gln
145                 150                 155                 160
Gly Val Val Ala Pro Lys Asn Leu Lys Glu Met Glu Gly Leu Ile Arg
                165                 170                 175
Ser Arg Thr Gly Asp Val Leu Asp Ser Leu Pro Thr Asp Lys Pro Phe
            180                 185                 190
Asn Trp Val Pro Ala Val Ser Lys Glu Leu Thr Gly Arg Met Leu Ala
        195                 200                 205
Thr Leu Leu Asp Phe Pro Tyr Glu Glu Arg His Lys Leu Val Glu Trp
    210                 215                 220
Ser Asp Arg Met Ala Gly Ala Ala Ser Ala Thr Gly Gly Glu Phe Ala
225                 230                 235                 240
Asp Glu Asn Ala Met Phe Asp Asp Ala Ala Asp Met Ala Arg Ser Phe
                245                 250                 255
Ser Arg Leu Trp Arg Asp Lys Glu Ala Arg Arg Ala Ala Gly Glu Glu
            260                 265                 270
Pro Gly Phe Asp Leu Ile Ser Leu Leu Gln Ser Asn Lys Glu Thr Lys
        275                 280                 285
Asp Leu Ile Asn Arg Pro Met Glu Phe Ile Gly Asn Leu Thr Leu Leu
```

290 295 300
Ile Val Ala Gly Asn Asp Thr Thr Arg Asn Ser Met Ser Gly Gly Leu
305 310 315 320
Val Ala Met Asn Glu Phe Pro Arg Glu Phe Glu Lys Leu Lys Ala Lys
325 330 335
Pro Glu Leu Ile Pro Asn Met Val Ser Glu Ile Ile Arg Trp Gln Thr
340 345 350
Pro Leu Ala Tyr Met Arg Arg Ile Ala Lys Gln Asp Val Glu Leu Gly
355 360 365
Gly Gln Thr Ile Lys Lys Gly Asp Arg Val Val Met Trp Tyr Ala Ser
370 375 380
Gly Asn Arg Asp Glu Arg Lys Phe Asp Asn Pro Asp Gln Phe Ile Ile
385 390 395 400
Asp Arg Lys Asp Ala Arg Asn His Met Ser Phe Gly Tyr Gly Val His
405 410 415
Arg Cys Met Gly Asn Arg Leu Ala Glu Leu Gln Leu Arg Ile Leu Trp
420 425 430
Glu Glu Ile Leu Lys Arg Phe Asp Asn Ile Glu Val Val Glu Glu Pro
435 440 445
Glu Arg Val Gln Ser Asn Phe Val Arg Gly Tyr Ser Arg Leu Met Val
450 455 460
Lys Leu Thr Pro Asn Ser Ser Pro Ser Thr Glu Gln Ser Ala Lys Lys
465 470 475 480
Val Arg Lys Lys Ala Glu Asn Ala His Asn Thr Pro Leu Leu Val Leu
485 490 495
Tyr Gly Ser Asn Met Gly Thr Ala Glu Gly Thr Ala Arg Asp Leu Ala
500 505 510
Asp Ile Ala Met Ser Lys Gly Phe Ala Pro Gln Val Ala Thr Leu Asp
515 520 525
Ser His Ala Gly Asn Leu Pro Arg Glu Gly Ala Val Leu Ile Val Thr
530 535 540
Ala Ser Tyr Asn Gly His Pro Pro Asp Asn Ala Lys Gln Phe Val Asp
545 550 555 560
Trp Leu Asp Gln Ala Ser Ala Asp Glu Val Lys Gly Val Arg Tyr Ser
565 570 575
Val Phe Gly Cys Gly Asp Lys Asn Trp Ala Thr Thr Tyr Gln Lys Val
580 585 590
[0055] Pro Ala Phe Ile Asp Glu Thr Leu Ala Ala Lys Gly Ala Glu Asn Ile
595 600 605
Ala Asp Arg Gly Glu Ala Asp Ala Ser Asp Asp Phe Glu Gly Thr Tyr
610 615 620
Glu Glu Trp Arg Glu His Met Trp Ser Asp Val Ala Ala Tyr Phe Asn
625 630 635 640
Leu Asp Ile Glu Asn Ser Glu Asp Asn Lys Ser Thr Leu Ser Leu Gln
645 650 655
Phe Val Asp Ser Ala Ala Asp Met Pro Leu Ala Lys Met His Gly Ala
660 665 670
Phe Ser Thr Asn Val Val Ala Ser Lys Glu Leu Gln Gln Pro Gly Ser
675 680 685
Ala Arg Ser Thr Arg His Leu Glu Ile Glu Leu Pro Lys Glu Ala Ser
690 695 700
Tyr Gln Glu Gly Asp His Leu Gly Val Ile Pro Arg Asn Tyr Glu Gly
705 710 715 720
Ile Val Asn Arg Val Thr Ala Arg Phe Gly Leu Asp Ala Ser Gln Gln
725 730 735
Ile Arg Leu Glu Ala Glu Glu Glu Lys Leu Ala His Leu Pro Leu Ala
740 745 750
Lys Thr Val Ser Val Glu Glu Leu Leu Gln Tyr Val Glu Leu Gln Asp
755 760 765
Pro Val Thr Arg Thr Gln Leu Arg Ala Met Ala Ala Lys Thr Val Cys
770 775 780
Pro Pro His Lys Val Glu Leu Glu Ala Leu Leu Glu Lys Gln Ala Tyr
785 790 795 800
Lys Glu Gln Val Leu Ala Lys Arg Leu Thr Met Leu Glu Leu Leu Glu
805 810 815
Lys Tyr Pro Ala Cys Glu Met Lys Phe Ser Glu Phe Ile Ala Leu Leu
820 825 830
Pro Ser Ile Arg Pro Arg Tyr Tyr Ser Ile Ser Ser Ser Pro Arg Val
835 840 845
Asp Glu Lys Gln Ala Ser Ile Thr Val Ser Val Val Ser Gly Glu Ala
850 855 860
Trp Ser Gly Tyr Gly Glu Tyr Lys Gly Ile Ala Ser Asn Tyr Leu Ala
865 870 875 880
Glu Leu Gln Glu Gly Asp Thr Ile Thr Cys Phe Ile Ser Thr Pro Gln
885 890 895
Ser Glu Phe Thr Leu Pro Lys Asp Pro Glu Thr Pro Leu Ile Met Val

```
            900                 905                 910
Gly Pro Gly Thr Gly Val Ala Pro Phe Arg Gly Phe Val Gln Ala Arg
        915                 920                 925
Lys Gln Leu Lys Glu Gln Gly Gln Ser Leu Gly Glu Ala His Leu Tyr
    930                 935                 940
Phe Gly Cys Arg Ser Pro His Glu Asp Tyr Leu Tyr Gln Glu Glu Leu
945                 950                 955                 960
Glu Asn Ala Gln Ser Glu Gly Ile Ile Thr Leu His Thr Ala Phe Ser
                965                 970                 975
Arg Met Pro Asn Gln Pro Lys Thr Tyr Val Gln His Val Met Glu Gln
            980                 985                 990
Asp Gly Lys Lys Leu Ile Glu Leu  Leu Asp Gln Gly Ala  His Phe Tyr
        995                1000                1005
Ile Cys  Gly Asp Gly Ser Gln  Met Ala Pro Ala Val  Glu Ala Thr
    1010                1015                1020
Leu Met  Lys Ser Tyr Ala Asp  Val His Gln Val Ser  Glu Ala Asp
    1025                1030                1035
Ala Arg  Leu Trp Leu Gln Gln  Leu Glu Glu Lys Gly  Arg Tyr Ala
    1040                1045                1050
Lys Asp  Val Trp Ala Gly
    1055
```

<210> 55
<211> 3246
<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> 55

[0056]

```
atgccaacac tgcccagaac atttgacgac attcagtccc gactgattaa cgccacctcc      60
agggtggtgc cgatgcagag gcaaattcag ggactgaaat tcttaatgag cgccaagagg     120
aagaccttcg gcccacgccg accgatgccc gaattcgttg aaacacccat cccggacgtt     180
aacacgctgg cccttgagga catcgatgtc agcaatccgt ttttataccg gcagggtcag     240
tggcgcgcct atttcaaacg gttgcgtgat gaggcgccgg tccattacca gaagaacagc     300
cctttcggcc ccttctggtc ggtaactcgg tttgaagaca tcctgttcgt ggataagagt     360
cacgacctgt tttccgccga gccgcaaatc attctcggtg accctccgga ggggctgtcg     420
gtggaaatgt tcatagcgat ggatccgccg aaacacgatg tgcagcgcag ctcggtgcag     480
ggagtagtgg caccgaaaaa cctgaaggag atggaggggc tgatccgatc acgcaccggc     540
gatgtgcttg acagcctgcc tacagacaaa ccctttaact gggtacctgc tgtttccaag     600
gaactcacag gccgcatgct ggcgacgctt ctggattttc cttacgagga acgccacaag     660
ctggttgagt ggtcggacag aatggcaggt gcagcatcgg ccaccggcgg ggagtttgcc     720
gatgaaaatg ccatgtttga cgacgcggca gacatggccc ggtctttctc caggctttgg     780
cgggacaagg aggcgcgccg cgcagcaggc gaggagcccg gtttcgattt gatcagcctg     840
ttgcagagca acaaagaaac gaaagacctg atcaatcggc cgatggagtt tatcggtaat     900
ttgacgctgc tcatagtcgc cggcaacgat acgacgcgca actcgatgag tggtggcctg     960
gtggccatga acgaattccc cagggaattt gaaaaattga aggcaaaacc ggagttgatt    1020
ccgaacatgg tgtcggaaat catccgctgg caaacgccgc tggcctatat gcgccgaatc    1080
gccaagcagg atgtcgaact gggcggccag accatcaaga agggtgatcg agttgtcatg    1140
tggtacgcgt cgggtaaccg ggacgagcgc aaatttgaca accccgatca gttcatcatt    1200
gatcgcaagg acgcacgaaa ccacatgtcg ttcggctatg gggttcaccg ttgcatgggc    1260
aaccgtctgg ctgaactgca actgcgcatc ctctgggaag aaatactcaa gcgttttgac    1320
aacatcgaag tcgtcgaaga gcccgagcgg gtgcagtcca acttcgtgcg gggctattcc    1380
aggttgatgg tcaaactgac accgaacagt ggcgctcagc ctgaagagaa cggacggcag    1440
gaagaacggc cttccgcacc ggcggcggaa aatacgcacg gaaccctct tcttgtgctc    1500
tacggttcaa atctcggcac agccgaagag attgcgaagg agcttgctga agaagcgcgt    1560
gagcaagggt ttcacagccg gacggcggag cttgatcaat acgcaggcgc catcccggca    1620
gaaggggctg ttatcattgt gacggcttcc tataacggaa acccgcccga ttgcgcaaag    1680
gaatttgtca attggcttga gcatgatcag acagacgatt tgcgtggtgt caaatatgcg    1740
gtattcggct gcggtaaccg cagctgggcc agcacctacc agcggattcc gcgcctgatt    1800
gacagcgtat tggaaaaaaa aggcgcccaa aggctgcaca agcttggaga aggggatgca    1860
ggcgatgatt ttgaaggaca gtttgagtca tggaaatatg atctgtggcc gcttttaaga    1920
accgaatttt cattggccga acccgagccg aatcaaacag aaacagacag gcaagcctta    1980
tctgtcgagt tcgtaaacgc acctgcggct tcgccgctgg ctaaagctta tcaggtgttc    2040
acagcgaaga tatcggcaaa ccgagaactg cagtgtgaaa agagcgggag aagcacaagg    2100
catattgaaa tatcgcttcc tgaaggcgcc gcatatcagg agggagacca tctcggtgtg    2160
ctaccgcaaa acagcgaagt gctgattggc cgcgtttttc agcggtttgg gctgaacgga    2220
aatgaacaaa ttctgattag cggccggaat caagcatcac atttgccttt ggagaggccc    2280
gttcatgtca aagacctttt tcaacattgc gtcgagctcc aggaaccggc cacaagggcc    2340
cagatacgcg agctggcggc tcatactgtt tgtccgcctc atcagcgcga gcttgaagac    2400
ctgctgaaag atgacgtcta taaggatcaa gtgttgaata agcggctgac aatgcttgac    2460
ctgcttgagc aatacccggc ctgtgaactg ccgttcgccc gttttctggc gcttcttcct    2520
ccgctaaaac cgaggtacta ttcgatttcc agttcgccgc agcttaaccc gcggcaaaca    2580
agcatcaccg tctctgtcgt aagtggcccg gcgttgagcg gccgcgggca ttataaggga    2640
gttgcatcga actatctcgc cggccttgag ccgggagacg cgatttcgtg tttcatcaga    2700
gagcctcagt caggcttccg gcttcccgaa gatcctgaaa caccggtgat catggtcggg    2760
```

```
ccgggcaccg gaatcgcccc ttaccgcgga tttcttcagg cgcgccgcat ccagcgcgat     2820
gccggtgtga agctcggtga agcgcatttg tacttcggct gccgccgtcc gaacgaagat     2880
tttctgtatc gagacgagtt ggagcaagcg gaaaaggacg gaatcgtcca tctgcataca     2940
gcgttttccc ggcttgaggg ccggccgaaa acatatgtgc aagatttgct cagagaggat     3000
gcagccttgc tgattcactt gttgaacgaa ggcggccgcc tgtatgtgtg cggagacgga     3060
agccgcatgg ctccagctgt tgaacaagct ttgtgcgagg cgtatcgcat agtacagggt     3120
gcgagtcggg aagagtcgca aagctggctg tccgcacttt tagaagaagg gcgctatgca     3180
aaggatgtat gggacggcgg cgtttcccaa cataatgtga aggcggactg cattgcaaga     3240
acgtaa                                                               3246
```

<210> 56
<211> 1081
<212> PRT
<213> 人工序列
<220>
<221> 来源
<223> /注="人工序列的描述：合成的多肽"
<400> 56

[0057]

```
Met Pro Thr Leu Pro Arg Thr Phe Asp Asp Ile Gln Ser Arg Leu Ile
1               5                  10                  15
Asn Ala Thr Ser Arg Val Val Pro Met Gln Arg Gln Ile Gln Gly Leu
            20                  25                  30
Lys Phe Leu Met Ser Ala Lys Arg Lys Thr Phe Gly Pro Arg Arg Pro
        35                  40                  45
Met Pro Glu Phe Val Glu Thr Pro Ile Pro Asp Val Asn Thr Leu Ala
    50                  55                  60
Leu Glu Asp Ile Asp Val Ser Asn Pro Phe Leu Tyr Arg Gln Gly Gln
65                  70                  75                  80
Trp Arg Ala Tyr Phe Lys Arg Leu Arg Asp Glu Ala Pro Val His Tyr
                85                  90                  95
Gln Lys Asn Ser Pro Phe Gly Pro Phe Trp Ser Val Thr Arg Phe Glu
            100                 105                 110
Asp Ile Leu Phe Val Asp Lys Ser His Asp Leu Phe Ser Ala Glu Pro
        115                 120                 125
Gln Ile Ile Leu Gly Asp Pro Pro Glu Gly Leu Ser Val Glu Met Phe
    130                 135                 140
Ile Ala Met Asp Pro Pro Lys His Asp Val Gln Arg Ser Ser Val Gln
145                 150                 155                 160
Gly Val Val Ala Pro Lys Asn Leu Lys Glu Met Glu Gly Leu Ile Arg
                165                 170                 175
Ser Arg Thr Gly Asp Val Leu Asp Ser Leu Pro Thr Asp Lys Pro Phe
            180                 185                 190
Asn Trp Val Pro Ala Val Ser Lys Glu Leu Thr Gly Arg Met Leu Ala
        195                 200                 205
Thr Leu Leu Asp Phe Pro Tyr Glu Glu Arg His Lys Leu Val Glu Trp
    210                 215                 220
Ser Asp Arg Met Ala Gly Ala Ala Ser Ala Thr Gly Gly Glu Phe Ala
225                 230                 235                 240
Asp Glu Asn Ala Met Phe Asp Asp Ala Ala Asp Met Ala Arg Ser Phe
                245                 250                 255
Ser Arg Leu Trp Arg Asp Lys Glu Ala Arg Arg Ala Ala Gly Glu Glu
            260                 265                 270
Pro Gly Phe Asp Leu Ile Ser Leu Leu Gln Ser Asn Lys Glu Thr Lys
        275                 280                 285
Asp Leu Ile Asn Arg Pro Met Glu Phe Ile Gly Asn Leu Thr Leu Leu
    290                 295                 300
Ile Val Ala Gly Asn Asp Thr Thr Arg Asn Ser Met Ser Gly Gly Leu
305                 310                 315                 320
Val Ala Met Asn Glu Phe Pro Arg Glu Phe Glu Lys Leu Lys Ala Lys
                325                 330                 335
Pro Glu Leu Ile Pro Asn Met Val Ser Glu Ile Ile Arg Trp Gln Thr
            340                 345                 350
Pro Leu Ala Tyr Met Arg Arg Ile Ala Lys Gln Asp Val Glu Leu Gly
        355                 360                 365
Gly Gln Thr Ile Lys Lys Gly Asp Arg Val Val Met Trp Tyr Ala Ser
    370                 375                 380
Gly Asn Arg Asp Glu Arg Lys Phe Asp Asn Pro Asp Gln Phe Ile Ile
385                 390                 395                 400
Asp Arg Lys Asp Ala Arg Asn His Met Ser Phe Gly Tyr Gly Val His
                405                 410                 415
Arg Cys Met Gly Asn Arg Leu Ala Glu Leu Gln Leu Arg Ile Leu Trp
            420                 425                 430
Glu Glu Ile Leu Lys Arg Phe Asp Asn Ile Glu Val Val Glu Glu Pro
        435                 440                 445
Glu Arg Val Gln Ser Asn Phe Val Arg Gly Tyr Ser Arg Leu Met Val
    450                 455                 460
```

```
Lys Leu Thr Pro Asn Ser Gly Ala Gln Pro Glu Glu Asn Gly Arg Gln
465                 470                 475                 480
Glu Glu Arg Pro Ser Ala Pro Ala Ala Glu Asn Thr His Gly Thr Pro
                485                 490                 495
Leu Leu Val Leu Tyr Gly Ser Asn Leu Gly Thr Ala Glu Glu Ile Ala
            500                 505                 510
Lys Glu Leu Ala Glu Glu Ala Arg Glu Gln Gly Phe His Ser Arg Thr
        515                 520                 525
Ala Glu Leu Asp Gln Tyr Ala Gly Ala Ile Pro Ala Glu Gly Ala Val
    530                 535                 540
Ile Ile Val Thr Ala Ser Tyr Asn Gly Asn Pro Pro Asp Cys Ala Lys
545                 550                 555                 560
Glu Phe Val Asn Trp Leu Glu His Asp Gln Thr Asp Asp Leu Arg Gly
                565                 570                 575
Val Lys Tyr Ala Val Phe Gly Cys Gly Asn Arg Ser Trp Ala Ser Thr
            580                 585                 590
Tyr Gln Arg Ile Pro Arg Leu Ile Asp Ser Val Leu Glu Lys Lys Gly
        595                 600                 605
Ala Gln Arg Leu His Lys Leu Gly Glu Gly Asp Ala Gly Asp Asp Phe
    610                 615                 620
Glu Gly Gln Phe Glu Ser Trp Lys Tyr Asp Leu Trp Pro Leu Leu Arg
625                 630                 635                 640
Thr Glu Phe Ser Leu Ala Glu Pro Glu Pro Asn Gln Thr Glu Thr Asp
                645                 650                 655
Arg Gln Ala Leu Ser Val Glu Phe Val Asn Ala Pro Ala Ala Ser Pro
            660                 665                 670
Leu Ala Lys Ala Tyr Gln Val Phe Thr Ala Lys Ile Ser Ala Asn Arg
        675                 680                 685
Glu Leu Gln Cys Glu Lys Ser Gly Arg Ser Thr Arg His Ile Glu Ile
    690                 695                 700
Ser Leu Pro Glu Gly Ala Ala Tyr Gln Glu Gly Asp His Leu Gly Val
705                 710                 715                 720
Leu Pro Gln Asn Ser Glu Val Leu Ile Gly Arg Val Phe Gln Arg Phe
                725                 730                 735
Gly Leu Asn Gly Asn Glu Gln Ile Leu Ile Ser Gly Arg Asn Gln Ala
            740                 745                 750
Ser His Leu Pro Leu Glu Arg Pro Val His Val Lys Asp Leu Phe Gln
        755                 760                 765
```

[0058]

```
His Cys Val Glu Leu Gln Glu Pro Ala Thr Arg Ala Gln Ile Arg Glu
    770                 775                 780
Leu Ala Ala His Thr Val Cys Pro Pro His Gln Arg Glu Leu Glu Asp
785                 790                 795                 800
Leu Leu Lys Asp Asp Val Tyr Lys Asp Gln Val Leu Asn Lys Arg Leu
                805                 810                 815
Thr Met Leu Asp Leu Leu Glu Gln Tyr Pro Ala Cys Glu Leu Pro Phe
            820                 825                 830
Ala Arg Phe Leu Ala Leu Leu Pro Pro Leu Lys Pro Arg Tyr Tyr Ser
        835                 840                 845
Ile Ser Ser Ser Pro Gln Leu Asn Pro Arg Gln Thr Ser Ile Thr Val
    850                 855                 860
Ser Val Val Ser Gly Pro Ala Leu Ser Gly Arg Gly His Tyr Lys Gly
865                 870                 875                 880
Val Ala Ser Asn Tyr Leu Ala Gly Leu Glu Pro Gly Asp Ala Ile Ser
                885                 890                 895
Cys Phe Ile Arg Glu Pro Gln Ser Gly Phe Arg Leu Pro Glu Asp Pro
            900                 905                 910
Glu Thr Pro Val Ile Met Val Gly Pro Gly Thr Gly Ile Ala Pro Tyr
        915                 920                 925
Arg Gly Phe Leu Gln Ala Arg Arg Ile Gln Arg Asp Ala Gly Val Lys
    930                 935                 940
Leu Gly Glu Ala His Leu Tyr Phe Gly Cys Arg Arg Pro Asn Glu Asp
945                 950                 955                 960
Phe Leu Tyr Arg Asp Glu Leu Glu Gln Ala Glu Lys Asp Gly Ile Val
                965                 970                 975
His Leu His Thr Ala Phe Ser Arg Leu Glu Gly Arg Pro Lys Thr Tyr
            980                 985                 990
Val Gln Asp Leu Leu Arg Glu Asp  Ala Ala Leu Leu Ile  His Leu Leu
        995                 1000                1005
Asn Glu  Gly Gly Arg Leu Tyr  Val Cys Gly Asp Gly  Ser Arg Met
    1010                1015                1020
Ala Pro  Ala Val Glu Gln Ala  Leu Cys Glu Ala Tyr  Arg Ile Val
    1025                1030                1035
Gln Gly  Ala Ser Arg Glu Glu  Ser Gln Ser Trp Leu  Ser Ala Leu
    1040                1045                1050
Leu Glu  Glu Gly Arg Tyr Ala  Lys Asp Val Trp Asp  Gly Gly Val
    1055                1060                1065
```

```
Ser Gln  His Asn Val Lys Ala  Asp Cys Ile Ala Arg  Thr
    1070                1075                1080

<210> 57
<211> 2400
<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> 57
atgtcaacga gttcaagtac aagtaatgac atccaggcaa aaataattaa cgccacatcc      60
aaagtcgtgc caatgcatct acagatcaag gcactaaaaa acttgatgaa ggtgaagcgg     120
aagaccattg gcacttcccg ccctcaggtg cactttgttg aaaccgattt gcctgacgtc     180
aatgatttgg cgatagaaga tatcgatacg agtaaccctt ttttataccg acaaggtaag     240
gcgaatgcgt actttaagcg gttgcgtgat gaagcgccgg tgcactatca gaagaacagt     300
gctttcgggc cgttctggtc ggtaacacgc tacgaagata tcgtcttcgt ggacaagagc     360
catgatttgt tttccgccga accccaaatt atcttgggtg atcctccgga aggcctgtcg     420
gttgaaatgt tcatcgctat ggatcctccc aagcacgacg tacagcgtcg ggcagtccag     480
ggtgttgttg cgcccaagaa cctgaaagaa atggaaggac tgatccgcaa gcgcaccggg     540
gacgtactgg atagcctgcc gttggacact ccgttcaact gggtgccggt ggtgtcgaaa     600
gagctgaccg ggcgcatgct agcctcactg ttagatttcc cgtatgacga acgcgaaaaa     660
ctggttggct ggtcggatcg attgtccggc gcgtcctcgg caaccggcgg cgagtttacg     720
aatgaagatg tgtttttga tgacgcggca gatatggcgt gggctttctc caagctttgg      780
cgtgataaag aagcccgtca aaaagcaggt gaagagccgg gtttcgattt gatcagcatg     840
cttcagtcca atgaagacac aaaagatctg atcaatcgtc ctttggaatt cattggtaat     900
ctcgcgttgt tgattgttgg cggtaatgac accacgcgta actcaatgag cgggggggtg     960
ctggctttaa atcagttccc agagcaattc gagaagctaa aggcgaaccc aaagcttatc    1020
cccaatatgg tctctgaaat cattcgctgg caaacgccgc ttgcgtatat gcgccgggtt    1080
gccaagcagg atgtggagct gaacggacag accatcaaga agggtgatcg cgtgctgatg    1140
tggtatgcgt cgggcaacca ggatgagaga aaatttgaga atcctgagca attcatcatc    1200
gaccgcaaag atacgcgtaa ccatgtgtcg tttggttatg gggttcaccg ttgtatgggc    1260
aaccgccttg ccgaactgca gctgcgtatt ctgtgggaag agcttctccc tcgctttgaa    1320
aacatcgaag tgatcggtga gccggagcgc gtgcaatcga actttgtgcg gggctattcc    1380
aagatgatgg ttaagttgac ggctaaaaaa gtactccatc gtcatcaacc tgtcaccatc    1440
ggcgagccgg ccgctcgtgc tgtgagccgc acggtgaccg ttgagcgtct tgatcgcatt    1500
gccgacgatg tccttcgcct ggtccttcgc gatgctggag gtaaaaccct cccgacgtgg    1560
acgcctggcg ctcacatcga cctggatctg ggtgctctga gccgtcagta ttcgctctgc    1620
ggcgctccgg atgctccgtc gtacgaaatc gccgtgcact tagatccgga aagccgtggt    1680
ggaagccgct atattcatga acagctggaa gttggaagtc cgctgcgtat gcgtggccca    1740
cgcaaccatt tcgccctgga tccgggtgcg gaacattacg tgtttgttgc cgggggtatc    1800
ggcatcacgc cggtgctggc aatggcggat catgcccgtg cgcgtggttg gtcgtacgaa    1860
ctgcattatt gtggtcgtaa tcgtagcggt atggcttacc tggaacgcgt cgcgggacat    1920
ggtgaccgcg ctgccttgca cgtatctgaa gaaggcaccc gcattgatct ggcggcatta    1980
cttgctgaac cggcgccggg cgtgcaaatc tacgcctgcg gtccgggccg tttattagcg    2040
ggtcttgaag acgcgtctcg taattggccg gatggcgcgc ttcatgtgga gcatttcact    2100
tcgagtttag ccgctttgga tccggatgtc gaacatgcct ttgatttgga gctgcgtgac    2160
tctggcctta ccgttcgcgt cgagccaact cagaccgttt tagacgcttt gcgtgcgaac    2220
aatatcgacg tcccgtcgga ttgcgaagag gggctgtgtg gttcttgcga agtagccgtt    2280
ctggatggcg aggttgatca ccgtgatacc gttctgacta aggccgagcg cgccgcgaat    2340
cgtcagatga tgacttgctg cagtcgtgca tgcggtgatc gtctggcgct gcgcctctaa    2400

<210> 58
<211> 799
<212> PRT
<213> 人工序列
<220>
<221> 来源
<223> /注="人工序列的描述：合成的多肽"
<400> 58
Met Ser Thr Ser Ser Ser Thr Ser Asn Asp Ile Gln Ala Lys Ile Ile
1               5                  10                  15
Asn Ala Thr Ser Lys Val Val Pro Met His Leu Gln Ile Lys Ala Leu
            20                  25                  30
Lys Asn Leu Met Lys Val Lys Arg Lys Thr Ile Gly Thr Ser Arg Pro
        35                  40                  45
Gln Val His Phe Val Glu Thr Asp Leu Pro Asp Val Asn Asp Leu Ala
    50                  55                  60
Ile Glu Asp Ile Asp Thr Ser Asn Pro Phe Leu Tyr Arg Gln Gly Lys
65                  70                  75                  80
Ala Asn Ala Tyr Phe Lys Arg Leu Arg Asp Glu Ala Pro Val His Tyr
                85                  90                  95
Gln Lys Asn Ser Ala Phe Gly Pro Phe Trp Ser Val Thr Arg Tyr Glu
            100                 105                 110
Asp Ile Val Phe Val Asp Lys Ser His Asp Leu Phe Ser Ala Glu Pro
        115                 120                 125
```

[0059]

```
Gln Ile Ile Leu Gly Asp Pro Pro Glu Gly Leu Ser Val Glu Met Phe
    130                 135                 140
Ile Ala Met Asp Pro Pro Lys His Asp Val Gln Arg Arg Ala Val Gln
145                 150                 155                 160
Gly Val Val Ala Pro Lys Asn Leu Lys Glu Met Glu Gly Leu Ile Arg
                165                 170                 175
Lys Arg Thr Gly Asp Val Leu Asp Ser Leu Pro Leu Asp Thr Pro Phe
            180                 185                 190
Asn Trp Val Pro Val Val Ser Lys Glu Leu Thr Gly Arg Met Leu Ala
        195                 200                 205
Ser Leu Leu Asp Phe Pro Tyr Asp Glu Arg Glu Lys Leu Val Gly Trp
    210                 215                 220
Ser Asp Arg Leu Ser Gly Ala Ser Ser Ala Thr Gly Gly Glu Phe Thr
225                 230                 235                 240
Asn Glu Asp Val Phe Phe Asp Asp Ala Ala Asp Met Ala Trp Ala Phe
                245                 250                 255
Ser Lys Leu Trp Arg Asp Lys Glu Ala Arg Gln Lys Ala Gly Glu Glu
            260                 265                 270
Pro Gly Phe Asp Leu Ile Ser Met Leu Gln Ser Asn Glu Asp Thr Lys
        275                 280                 285
Asp Leu Ile Asn Arg Pro Leu Glu Phe Ile Gly Asn Leu Ala Leu Leu
    290                 295                 300
Ile Val Gly Gly Asn Asp Thr Thr Arg Asn Ser Met Ser Gly Gly Val
305                 310                 315                 320
Leu Ala Leu Asn Gln Phe Pro Glu Gln Phe Glu Lys Leu Lys Ala Asn
                325                 330                 335
Pro Lys Leu Ile Pro Asn Met Val Ser Glu Ile Ile Arg Trp Gln Thr
            340                 345                 350
Pro Leu Ala Tyr Met Arg Arg Val Ala Lys Gln Asp Val Glu Leu Asn
        355                 360                 365
Gly Gln Thr Ile Lys Lys Gly Asp Arg Val Leu Met Trp Tyr Ala Ser
    370                 375                 380
Gly Asn Gln Asp Glu Arg Lys Phe Glu Asn Pro Glu Gln Phe Ile Ile
385                 390                 395                 400
Asp Arg Lys Asp Thr Arg Asn His Val Ser Phe Gly Tyr Gly Val His
                405                 410                 415
Arg Cys Met Gly Asn Arg Leu Ala Glu Leu Gln Leu Arg Ile Leu Trp
            420                 425                 430
Glu Glu Leu Leu Pro Arg Phe Glu Asn Ile Glu Val Ile Gly Glu Pro
        435                 440                 445
Glu Arg Val Gln Ser Asn Phe Val Arg Gly Tyr Ser Lys Met Met Val
    450                 455                 460
Lys Leu Thr Ala Lys Lys Val Leu His Arg His Gln Pro Val Thr Ile
465                 470                 475                 480
Gly Glu Pro Ala Ala Arg Ala Val Ser Arg Thr Val Thr Val Glu Arg
                485                 490                 495
Leu Asp Arg Ile Ala Asp Asp Val Leu Arg Leu Val Leu Arg Asp Ala
            500                 505                 510
Gly Gly Lys Thr Leu Pro Thr Trp Thr Pro Gly Ala His Ile Asp Leu
        515                 520                 525
Asp Leu Gly Ala Leu Ser Arg Gln Tyr Ser Leu Cys Gly Ala Pro Asp
    530                 535                 540
Ala Pro Ser Tyr Glu Ile Ala Val His Leu Asp Pro Glu Ser Arg Gly
545                 550                 555                 560
Gly Ser Arg Tyr Ile His Glu Gln Leu Glu Val Gly Ser Pro Leu Arg
                565                 570                 575
Met Arg Gly Pro Arg Asn His Phe Ala Leu Asp Pro Gly Ala Glu His
            580                 585                 590
Tyr Val Phe Val Ala Gly Gly Ile Gly Ile Thr Pro Val Leu Ala Met
        595                 600                 605
Ala Asp His Ala Arg Ala Arg Gly Trp Ser Tyr Glu Leu His Tyr Cys
    610                 615                 620
Gly Arg Asn Arg Ser Gly Met Ala Tyr Leu Glu Arg Val Ala Gly His
625                 630                 635                 640
Gly Asp Arg Ala Ala Leu His Val Ser Glu Glu Gly Thr Arg Ile Asp
                645                 650                 655
Leu Ala Ala Leu Leu Ala Glu Pro Ala Pro Gly Val Gln Ile Tyr Ala
            660                 665                 670
Cys Gly Pro Gly Arg Leu Leu Ala Gly Leu Glu Asp Ala Ser Arg Asn
        675                 680                 685
Trp Pro Asp Gly Ala Leu His Val Glu His Phe Thr Ser Ser Leu Ala
    690                 695                 700
Ala Leu Asp Pro Asp Val Glu His Ala Phe Asp Leu Glu Leu Arg Asp
705                 710                 715                 720
Ser Gly Leu Thr Val Arg Val Glu Pro Thr Gln Thr Val Leu Asp Ala
                725                 730                 735
```

[0060]

```
Leu Arg Ala Asn Asn Ile Asp Val Pro Ser Asp Cys Glu Glu Gly Leu
            740                 745                 750
Cys Gly Ser Cys Glu Val Ala Val Leu Asp Gly Glu Val Asp His Arg
        755                 760                 765
Asp Thr Val Leu Thr Lys Ala Glu Arg Ala Ala Asn Arg Gln Met Met
    770                 775                 780
Thr Cys Cys Ser Arg Ala Cys Gly Asp Arg Leu Ala Leu Arg Leu
785                 790                 795
```

<210> 59
<211> 1392
<212> DNA
<213> 海分枝杆菌
<400> 59

```
atgagcaata ttcgcgaggc agtcactgcc aaggctcagg caacaattcc gatggaccga       60
ataatccagg gcgcccacct ctacgacaga acgcggcgct gggtcaccgg caccaacggt      120
gaaaaaatct tcatcgagcg accgatcccg ccggctgacg aggttgaact gaccgacatc      180
gaccttagca atcctttcct ctatcgtcag ggtcgctgga agtcctatta cgagcgccta      240
cgcaacgagg ctcccgtgca ctatcaggcc cacagcgcgt tcggcccgtt ctggtcggtg      300
acgcggcatg ccgacatcgt ggccgtcgac aagaaccacg aggtcttctc ctccgagccg      360
ttcatcgtca tcgggagccc gccgcgcttc ctcgacattg cgatgttcat cgcgatggac      420
cccccaaaac acgaccggca acggcaggct gtccagggtg tggtcgcacc gaagaacctg      480
cgtgagatgg agggcctcat ccgcgagcgg gtggtagacg tgctcgacgc tctgccgctt      540
ggcgaaccgt tcaactgggt gcagcacgtc tcgatcgagc taaccgcgcg catgctggcc      600
acgctgctgg acttcccgtt cgagcagcgg cgcaagctcg tccaatggtc cgatctcgcc      660
acctccatgg agcaagccaa cggtgggccc tcggacaacg acgagatatt tcgcggcatg      720
gtcgatatgg ctagaggtct cagcgctcac tggcgggaca aggcagcccg gacagctgcc      780
ggagagctgc ccggcttcga tctgatcacc atgttgcaga gcgacgagag caccaaggac      840
ctgatcgatc gcccgatgga gttcttgggc aacttggtat tgctgatcgt gggtggcaac      900
gatacgaccc gcaattccat gagcggtggt gttctggcgc tgaacgagtt ccctgaccag      960
tttgagaagc tgaaggcgaa ccccgagctg atccccaaca tggtctcgga gatcatccgg     1020
tggcaaaccc cgctcgcgca tatgcgccgg atcgccaagg ccgacactgt gctcaacggg     1080
cagttcatcc gcaagggcga caaggtcctg atgtggtacg cctcgggcaa ccgcgacgag     1140
cgcgtgttcg atcggcccga tgacctgatt atcgatcggg ccaacgcccg taaccacatc     1200
tccttcggtt tcggcgtgca ccgctgtatg ggtaaccggc tggccgagat gcagttgcgg     1260
atcctgtggg aggagctgct tccgcggttc gagaacatcg aggtcgtcgg tgagcccgag     1320
tacgtgcagt ccaacttcgt gaggggggatc agtaagctga tggtccgcct caccccgaaa     1380
ggtggcgcat ga                                                         1392
```

[0061]

<210> 60
<211> 463
<212> PRT
<213> 海分枝杆菌
<400> 60

```
Met Ser Asn Ile Arg Glu Ala Val Thr Ala Lys Ala Gln Ala Thr Ile
1               5                  10                  15
Pro Met Asp Arg Ile Ile Gln Gly Ala His Leu Tyr Asp Arg Thr Arg
            20                  25                  30
Arg Trp Val Thr Gly Thr Asn Gly Glu Lys Ile Phe Ile Glu Arg Pro
        35                  40                  45
Ile Pro Pro Ala Asp Glu Val Glu Leu Thr Asp Ile Asp Leu Ser Asn
    50                  55                  60
Pro Phe Leu Tyr Arg Gln Gly Arg Trp Lys Ser Tyr Tyr Glu Arg Leu
65                  70                  75                  80
Arg Asn Glu Ala Pro Val His Tyr Gln Ala His Ser Ala Phe Gly Pro
                85                  90                  95
Phe Trp Ser Val Thr Arg His Ala Asp Ile Val Ala Val Asp Lys Asn
            100                 105                 110
His Glu Val Phe Ser Ser Glu Pro Phe Ile Val Ile Gly Ser Pro Pro
        115                 120                 125
Arg Phe Leu Asp Ile Ala Met Phe Ile Ala Met Asp Pro Pro Lys His
    130                 135                 140
Asp Arg Gln Arg Gln Ala Val Gln Gly Val Val Ala Pro Lys Asn Leu
145                 150                 155                 160
Arg Glu Met Glu Gly Leu Ile Arg Glu Arg Val Val Asp Val Leu Asp
                165                 170                 175
Ala Leu Pro Leu Gly Glu Pro Phe Asn Trp Val Gln His Val Ser Ile
            180                 185                 190
Glu Leu Thr Ala Arg Met Leu Ala Thr Leu Leu Asp Phe Pro Phe Glu
        195                 200                 205
Gln Arg Arg Lys Leu Val Gln Trp Ser Asp Leu Ala Thr Ser Met Glu
    210                 215                 220
Gln Ala Asn Gly Gly Pro Ser Asp Asn Asp Glu Ile Phe Arg Gly Met
225                 230                 235                 240
Val Asp Met Ala Arg Gly Leu Ser Ala His Trp Arg Asp Lys Ala Ala
                245                 250                 255
```

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

<210> 61
<211> 3150
<212> DNA
<213> 巨大芽孢杆菌
<400> 61

[0062]

```
atgacaatta aagaaatgcc tcagccaaaa acgtttggag agcttaaaaa tttaccgtta      60
ttaaacacag ataaaccggt tcaagctttg atgaaaattg cggatgaatt aggagaaatc     120
tttaaattcg aggcgcctgg tcgtgtaacg cgctacttat caagtcagcg tctaattaaa     180
gaagcatgcg atgaatcacg ctttgataaa aacttaagtc aagcgcttaa atttgtacgt     240
gattttgcag gagacgggtt agctacaagc tggacgcatg aaaaaaattg gaaaaaagcg     300
cataatatct tacttccaag cttcagtcag caggcaatga aaggctatca tgcgatgatg     360
gtcgatatcg ccgtgcagct tgttcaaaag tgggagcgtc taaatgcaga tgagcatatt     420
gaagtaccgg aagacatgac acgtttaacg cttgatacaa ttggtctttg cggctttaac     480
tatcgcttta acagctttta ccgagatcag cctcatccat ttattacaag tatggtccgt     540
gcactggatg aagcaatgaa caagctgcag cgagcaaatc cagacgaccc agcttatgat     600
gaaaacaagc gccagtttca agaagatatc aaggtgatga acgacctagt agataaaatt     660
attgcagatc gcaaagcaag cggtgaacaa agcgatgatt tattaacgca tatgctaaac     720
ggaaaagatc cagaaacggg tgagccgctt gatgacgaga acattcgcta tcaaattatt     780
acattcttaa ttgcgggaca cgaaacaaca agtggtcttt tatcatttgc gctgtatttc     840
ttagtgaaaa atccacatgt attacaaaaa gcagcagaag aagcagcacg agttctagta     900
gatcctgttc caagctacaa acaagtcaaa cagcttaaat atgtcggcat ggtcttaaac     960
gaagcgctgc gcttatggcc aactgctcct gcgttttccc tatatgcaaa agaagatacg    1020
gtgcttggag gagaatatcc tttagaaaaa ggcgacgaac taatggttct gattcctcag    1080
cttcaccgtg ataaaacaat ttggggagac gatgtggaag agttccgtcc agagcgtttt    1140
gaaaatccaa gtgcgattcc gcagcatgcg tttaaaccgt ttggaaacgg tcagcgtgcg    1200
tgtatcggtc agcagttcgc tcttcatgaa gcaacgctgg tacttggtat gatgctaaaa    1260
cactttgact ttgaagatca tacaaactac gagctggata ttaaagaaac tttaacgtta    1320
aaacctgaag gctttgtggt aaaagcaaaa tcgaaaaaaa ttccgcttgg cggtattcct    1380
tcacctagca ctgaacagtc tgctaaaaaa gtacgcaaaa aggcagaaaa cgctcataat    1440
acgccgctgc ttgtgctata cggttcaaat atgggaacag ctgaaggaac ggcgcgtgat    1500
ttagcagata ttgcaatgag caaaggattt gcaccgcagg tcgcaacgct tgattcacac    1560
gccggaaatc ttccgcgcga aggagctgta ttaattgtaa cggcgtctta taacggtcat    1620
ccgcctgata acgcaaagca atttgtcgac tggttagacc aagcgtctgc tgatgaagta    1680
aaaggcgttc gctactccgt atttggatgc ggcgataaaa actgggctac tacgtatcaa    1740
aaagtgcctg cttttatcga tgaaacgctt gccgctaaag gggcagaaaa catcgctgac    1800
cgcggtgaag cagatgcaag cgacgacttt gaaggcacat atgaagaatg gcgtgaacat    1860
atgtggagtg acgtagcagc ctactttaac ctcgacattg aaaacagtga agataataaa    1920
tctactcttt cacttcaatt tgtcgacagc gccgcggata tgccgcttgc gaaaatgcac    1980
ggtgcgtttt caacgaacgt cgtagcaagc aaagaacttc aacagccagg cagtgcacga    2040
agcacgcgac atcttgaaat tgaacttcca aaagaagctt cttatcaaga aggagatcat    2100
ttaggtgtta ttcctcgcaa ctatgaagga atagtaaacc gtgtaacagc aaggttcggc    2160
ctagatgcat cacagcaaat ccgtctggaa gcagaagaag aaaaattagc tcatttgcca    2220
ctcgctaaaa cagtatccgt agaagagctt ctgcaatacg tggagcttca agatcctgtt    2280
acgcgcacgc agcttcgcgc aatggctgct aaaacggtct gcccgccgca taaagtagag    2340
cttgaagcct tgcttgaaaa gcaagcctac aaagaacaag tgctggcaaa acgtttaaca    2400
atgcttgaac tgcttgaaaa atacccggcg tgtgaaatga aattcagcga atttatcgcc    2460
cttctgccaa gcatacgccc gcgctattac tcgatttctt catcacctcg tgtcgatgaa    2520
aaacaagcaa gcatcacggt cagcgttgtc tcaggagaag cgtggagcgg atatggagaa    2580
tataaaggaa ttgcgtcgaa ctatcttgcc gagctgcaag aaggagatac gattacgtgc    2640
```

```
tttatttcca caccgcagtc agaatttacg ctgccaaaag accctgaaac gccgcttatc     2700
atggtcggac cgggaacagg cgtcgcgccg tttagaggct ttgtgcaggc gcgcaaacag     2760
ctaaaagaac aaggacagtc acttggagaa gcacatttat acttcggctg ccgttcacct     2820
catgaagact atctgtatca agaagagctt gaaaacgccc aaagcgaagg catcattacg     2880
cttcataccg ctttttctcg catgccaaat cagccgaaaa catacgttca gcacgtaatg     2940
gaacaagacg gcaagaaatt gattgaactt cttgatcaag gagcgcactt ctatatttgc     3000
ggagacggaa gccaaatggc acctgccgtt gaagcaacgc ttatgaaaag ctatgctgac     3060
gttcaccaag tgagtgaagc agacgctcgc ttatggctgc agcagctaga agaaaaaggc     3120
cgatacgcaa aagacgtgtg ggctgggtaa                                      3150
```

<210> 62
<211> 1049
<212> PRT
<213> 巨大芽孢杆菌
<400> 62

```
Met Thr Ile Lys Glu Met Pro Gln Pro Lys Thr Phe Gly Glu Leu Lys
1               5                  10                  15
Asn Leu Pro Leu Leu Asn Thr Asp Lys Pro Val Gln Ala Leu Met Lys
            20                  25                  30
Ile Ala Asp Glu Leu Gly Glu Ile Phe Lys Phe Glu Ala Pro Gly Arg
        35                  40                  45
Val Thr Arg Tyr Leu Ser Ser Gln Arg Leu Ile Lys Glu Ala Cys Asp
    50                  55                  60
Glu Ser Arg Phe Asp Lys Asn Leu Ser Gln Ala Leu Lys Phe Val Arg
65                  70                  75                  80
Asp Phe Ala Gly Asp Gly Leu Ala Thr Ser Trp Thr His Glu Lys Asn
                85                  90                  95
Trp Lys Lys Ala His Asn Ile Leu Leu Pro Ser Phe Ser Gln Gln Ala
            100                 105                 110
Met Lys Gly Tyr His Ala Met Met Val Asp Ile Ala Val Gln Leu Val
        115                 120                 125
Gln Lys Trp Glu Arg Leu Asn Ala Asp Glu His Ile Glu Val Pro Glu
    130                 135                 140
Asp Met Thr Arg Leu Thr Leu Asp Thr Ile Gly Leu Cys Gly Phe Asn
145                 150                 155                 160
Tyr Arg Phe Asn Ser Phe Tyr Arg Asp Gln Pro His Pro Phe Ile Thr
                165                 170                 175
Ser Met Val Arg Ala Leu Asp Glu Ala Met Asn Lys Leu Gln Arg Ala
            180                 185                 190
Asn Pro Asp Asp Pro Ala Tyr Asp Glu Asn Lys Arg Gln Phe Gln Glu
        195                 200                 205
Asp Ile Lys Val Met Asn Asp Leu Val Asp Lys Ile Ile Ala Asp Arg
    210                 215                 220
Lys Ala Ser Gly Glu Gln Ser Asp Asp Leu Leu Thr His Met Leu Asn
225                 230                 235                 240
Gly Lys Asp Pro Glu Thr Gly Glu Pro Leu Asp Asp Glu Asn Ile Arg
                245                 250                 255
Tyr Gln Ile Ile Thr Phe Leu Ile Ala Gly His Glu Thr Thr Ser Gly
            260                 265                 270
Leu Leu Ser Phe Ala Leu Tyr Phe Leu Val Lys Asn Pro His Val Leu
        275                 280                 285
Gln Lys Ala Ala Glu Glu Ala Ala Arg Val Leu Val Asp Pro Val Pro
    290                 295                 300
Ser Tyr Lys Gln Val Lys Gln Leu Lys Tyr Val Gly Met Val Leu Asn
305                 310                 315                 320
Glu Ala Leu Arg Leu Trp Pro Thr Ala Pro Ala Phe Ser Leu Tyr Ala
                325                 330                 335
Lys Glu Asp Thr Val Leu Gly Gly Glu Tyr Pro Leu Glu Lys Gly Asp
            340                 345                 350
Glu Leu Met Val Leu Ile Pro Gln Leu His Arg Asp Lys Thr Ile Trp
        355                 360                 365
Gly Asp Asp Val Glu Glu Phe Arg Pro Glu Arg Phe Glu Asn Pro Ser
    370                 375                 380
Ala Ile Pro Gln His Ala Phe Lys Pro Phe Gly Asn Gly Gln Arg Ala
385                 390                 395                 400
Cys Ile Gly Gln Gln Phe Ala Leu His Glu Ala Thr Leu Val Leu Gly
                405                 410                 415
Met Met Leu Lys His Phe Asp Phe Glu Asp His Thr Asn Tyr Glu Leu
            420                 425                 430
Asp Ile Lys Glu Thr Leu Thr Leu Lys Pro Glu Gly Phe Val Val Lys
        435                 440                 445
Ala Lys Ser Lys Lys Ile Pro Leu Gly Gly Ile Pro Ser Pro Ser Thr
    450                 455                 460
Glu Gln Ser Ala Lys Lys Val Arg Lys Lys Ala Glu Asn Ala His Asn
465                 470                 475                 480
Thr Pro Leu Leu Val Leu Tyr Gly Ser Asn Met Gly Thr Ala Glu Gly
```

[0063]

```
                485                 490                 495
Thr Ala Arg Asp Leu Ala Asp Ile Ala Met Ser Lys Gly Phe Ala Pro
            500                 505                 510
Gln Val Ala Thr Leu Asp Ser His Ala Gly Asn Leu Pro Arg Glu Gly
        515                 520                 525
Ala Val Leu Ile Val Thr Ala Ser Tyr Asn Gly His Pro Pro Asp Asn
    530                 535                 540
Ala Lys Gln Phe Val Asp Trp Leu Asp Gln Ala Ser Ala Asp Glu Val
545                 550                 555                 560
Lys Gly Val Arg Tyr Ser Val Phe Gly Cys Gly Asp Lys Asn Trp Ala
                565                 570                 575
Thr Thr Tyr Gln Lys Val Pro Ala Phe Ile Asp Glu Thr Leu Ala Ala
            580                 585                 590
Lys Gly Ala Glu Asn Ile Ala Asp Arg Gly Glu Ala Asp Ala Ser Asp
        595                 600                 605
Asp Phe Glu Gly Thr Tyr Glu Glu Trp Arg Glu His Met Trp Ser Asp
    610                 615                 620
Val Ala Ala Tyr Phe Asn Leu Asp Ile Glu Asn Ser Glu Asp Asn Lys
625                 630                 635                 640
Ser Thr Leu Ser Leu Gln Phe Val Asp Ser Ala Ala Asp Met Pro Leu
                645                 650                 655
Ala Lys Met His Gly Ala Phe Ser Thr Asn Val Val Ala Ser Lys Glu
            660                 665                 670
Leu Gln Gln Pro Gly Ser Ala Arg Ser Thr Arg His Leu Glu Ile Glu
        675                 680                 685
Leu Pro Lys Glu Ala Ser Tyr Gln Glu Gly Asp His Leu Gly Val Ile
    690                 695                 700
Pro Arg Asn Tyr Glu Gly Ile Val Asn Arg Val Thr Ala Arg Phe Gly
705                 710                 715                 720
Leu Asp Ala Ser Gln Gln Ile Arg Leu Glu Ala Glu Glu Glu Lys Leu
                725                 730                 735
Ala His Leu Pro Leu Ala Lys Thr Val Ser Val Glu Glu Leu Leu Gln
            740                 745                 750
Tyr Val Glu Leu Gln Asp Pro Val Thr Arg Thr Gln Leu Arg Ala Met
        755                 760                 765
Ala Ala Lys Thr Val Cys Pro Pro His Lys Val Glu Leu Glu Ala Leu
    770                 775                 780
Leu Glu Lys Gln Ala Tyr Lys Glu Gln Val Leu Ala Lys Arg Leu Thr
785                 790                 795                 800
Met Leu Glu Leu Leu Glu Lys Tyr Pro Ala Cys Glu Met Lys Phe Ser
                805                 810                 815
Glu Phe Ile Ala Leu Leu Pro Ser Ile Arg Pro Arg Tyr Tyr Ser Ile
            820                 825                 830
Ser Ser Ser Pro Arg Val Asp Glu Lys Gln Ala Ser Ile Thr Val Ser
        835                 840                 845
Val Val Ser Gly Glu Ala Trp Ser Gly Tyr Gly Glu Tyr Lys Gly Ile
    850                 855                 860
Ala Ser Asn Tyr Leu Ala Glu Leu Gln Glu Gly Asp Thr Ile Thr Cys
865                 870                 875                 880
Phe Ile Ser Thr Pro Gln Ser Glu Phe Thr Leu Pro Lys Asp Pro Glu
                885                 890                 895
Thr Pro Leu Ile Met Val Gly Pro Gly Thr Gly Val Ala Pro Phe Arg
            900                 905                 910
Gly Phe Val Gln Ala Arg Lys Gln Leu Lys Glu Gln Gly Gln Ser Leu
        915                 920                 925
Gly Glu Ala His Leu Tyr Phe Gly Cys Arg Ser Pro His Glu Asp Tyr
    930                 935                 940
Leu Tyr Gln Glu Glu Leu Glu Asn Ala Gln Ser Glu Gly Ile Ile Thr
945                 950                 955                 960
Leu His Thr Ala Phe Ser Arg Met Pro Asn Gln Pro Lys Thr Tyr Val
                965                 970                 975
Gln His Val Met Glu Gln Asp Gly Lys Lys Leu Ile Glu Leu Leu Asp
            980                 985                 990
Gln Gly Ala His Phe Tyr Ile Cys  Gly Asp Gly Ser Gln  Met Ala Pro
        995                 1000                1005
Ala Val  Glu Ala Thr Leu Met  Lys Ser Tyr Ala Asp  Val His Gln
    1010                1015                1020
Val Ser  Glu Ala Asp Ala Arg  Leu Trp Leu Gln Gln  Leu Glu Glu
    1025                1030                1035
Lys Gly  Arg Tyr Ala Lys Asp  Val Trp Ala Gly
    1040                1045
```

[0064]

<210> 63
<211> 3225
<212> DNA
<213> 地衣芽孢杆菌

<400> 63
atgaacaagt tagatggaat tccaatccct aaaacttacg ggccgctcgg caacctgcct 60
ttgcttgaca aaaacagggt ctcccagtca ctttggaaaa tcgcggatga gatggggcct 120
atctttcaat ttaagtttgc ggatgcgatt ggggtttttg tgtccagcca tgaactggtt 180
aaagaagtct ctgaagaatc ccgttttgac aaaaacatgg ggaaggggct attgaaagtt 240
cgcgagttca gcggagacgg gctctttaca agctggacgg aagaacccaa ttggcggaaa 300
gcccacaaca tccttctgcc gagcttcagc cagaaagcga tgaagggata ccatcccatg 360
atgcaggata tcgccgtcca gctcattcaa aagtggtccc gtctcaatca ggatgaaagc 420
attgatgtgc cggacgatat gacgcggctg acgctggaca cgatcggctt atgcgggttt 480
aactaccgct ttaacagctt ctaccgtgaa gggcagcatc cgtttattga gagcatggtc 540
cggggtttga gcgaagcgat gagacagacg aagcgcttcc cgctgcagga taagctgatg 600
attcaaacga agcgccggtt taacagcgat gtcgagtcga tgttttctct tgttgaccgg 660
atcatcgctg accggaagca ggccgagagt gaaagcggaa atgacctctt gtcgcttatg 720
cttcatgcga aagaccctga gaccggcgaa aaactggatg atgagaatat ccgctatcaa 780
attattacat ttttgattgc cggacacgag acgacgagcg gtttattatc gtttgcaatc 840
tatctgctcc tgaagcatcc ggataagctt aagaaagcgt atgaagaagc agaccgcgtg 900
ctgaccgatc ccgtcccatc ctacaaacag gttcagcagc tgaaatacat ccgaatgatt 960
ttgaatgaat cgataaggct ttggccgacg gcaccggctt tctctcttta tgcaaaagaa 1020
gaaacggtta tcgggggaaa atatttgatt ccaaaaggac agagcgttac agtgctcatc 1080
ccaaaactgc acagagatca aagcgtctgg ggagaagatg ccgaggcatt ccggcctgaa 1140
cggttcgagc agatggacag cattccggcg cacgcataca aaccgtttgg caacggccaa 1200
agggcatgca tcggcatgca gttcgccctt catgaagcga cgcttgtgct cggcatgatt 1260
cttcagtact ttgatcttga agatcatgca aactaccaat tgaagatcaa agaatcgctg 1320
acattaaaac cggatggttt cacaatccgg gtgaggccga ggaaaaaaga agcaatgacg 1380
gcgatgccgg gcgctcagcc tgaagagaac ggacggcagg aagaacggcc ttccgcaccg 1440
gcggcggaaa atacgcacgg aacccctctt cttgtgctct acggttcaaa tctcggcaca 1500
gccgaagaga ttgcgaagga gcttgctgaa gaagcgcgtg agcaagggtt tcacagccgg 1560
acggcggagc ttgatcaata cgcaggcgcc atcccggcag aaggggctgt tatcattgtg 1620
acggcttcct ataacggaaa cccgcccgat tgcgcaaagg aatttgtcaa ttggcttgag 1680
catgatcaga cagacgattt gcgtggtgtc aaatatgcgg tattcggctg cggtaaccgc 1740
agctgggcca gcacctacca gcggattccg cgcctgattg acagcgtatt ggaaaaaaaa 1800
ggcgcccaaa ggctgcacaa gcttggagaa ggggatgcag gcgatgattt tgaaggacag 1860
tttgagtcat ggaaatatga tctgtggccg cttttaagaa ccgaattttc attggccgaa 1920
cccgagccga atcaaacaga aacagacagg caagccttat ctgtcgagtt cgtaaacgca 1980
cctgcggctt cgccgctggc taaagcttat caggtgttca cagcgaagat atcggcaaac 2040
cgagaactgc agtgtgaaaa gagcgggaga agcacaaggc atattgaaat atcgcttcct 2100
gaaggcgccg catatcagga gggagaccat ctcggtgtgc taccgcaaaa cagcgaagtg 2160
ctgattggcc gcgtttttca gcggtttggg ctgaacggaa atgaacaaat tctgattagc 2220
[0065] ggccggaatc aagcatcaca tttgcctttg gagaggcccg ttcatgtcaa agacctttt 2280
caacattgcg tcgagctcca ggaaccggcc acaagggccc agatacgcga gctggcggct 2340
catactgttt gtccgcctca tcagcgcgag cttgaagacc tgctgaaaga tgacgtctat 2400
aaggatcaag tgttgaataa gcggctgaca atgcttgacc tgcttgagca atacccggcc 2460
tgtgaactgc cgttcgcccg ttttctggcg cttcttcctc cgctaaaacc gaggtactat 2520
tcgatttcca gttcgccgca gcttaacccg cggcaaacaa gcatcaccgt ctctgtcgta 2580
agtggcccgg cgttgagcgg ccgcgggcat tataagggag ttgcatcgaa ctatctcgcc 2640
ggccttgagc cgggagacgc gatttcgtgt ttcatcagag agcctcagtc aggcttccgg 2700
cttcccgaag atcctgaaac accggtgatc atggtcgggc cgggcaccgg aatcgccct 2760
taccgcggat ttcttcaggc gcgccgcatc cagcgcgatg ccggtgtgaa gctcggtgaa 2820
gcgcatttgt acttcggctg ccgccgtccg aacgaagatt ttctgtatcg agacgagttg 2880
gagcaagcgg aaaaggacgg aatcgtccat ctgcatacag cgttttcccg gcttgagggc 2940
cggccgaaaa catatgtgca agatttgctc agagaggatg cagccttgct gattcacttg 3000
ttgaacgaag gcggccgcct gtatgtgtgc ggagacggaa gccgcatggc tccagctgtt 3060
gaacaagctt tgtgcgaggc gtatcgcata gtacagggtg cgagtcggga agagtcgcaa 3120
agctggctgt ccgcactttt agaagaaggg cgctatgcaa aggatgtatg ggacggcggc 3180
gtttcccaac ataatgtgaa ggcggactgc attgcaagaa cgtaa 3225

<210> 64
<211> 1074
<212> PRT
<213> 地衣芽孢杆菌
<400> 64
Met Asn Lys Leu Asp Gly Ile Pro Ile Pro Lys Thr Tyr Gly Pro Leu
1 5 10 15
Gly Asn Leu Pro Leu Leu Asp Lys Asn Arg Val Ser Gln Ser Leu Trp
20 25 30
Lys Ile Ala Asp Glu Met Gly Pro Ile Phe Gln Phe Lys Phe Ala Asp
35 40 45
Ala Ile Gly Val Phe Val Ser Ser His Glu Leu Val Lys Glu Val Ser
50 55 60
Glu Glu Ser Arg Phe Asp Lys Asn Met Gly Lys Gly Leu Leu Lys Val
65 70 75 80
Arg Glu Phe Ser Gly Asp Gly Leu Phe Thr Ser Trp Thr Glu Glu Pro
85 90 95
Asn Trp Arg Lys Ala His Asn Ile Leu Leu Pro Ser Phe Ser Gln Lys
100 105 110
Ala Met Lys Gly Tyr His Pro Met Met Gln Asp Ile Ala Val Gln Leu

[0066]

```
          115                 120                 125
Ile Gln Lys Trp Ser Arg Leu Asn Gln Asp Glu Ser Ile Asp Val Pro
      130                 135                 140
Asp Asp Met Thr Arg Leu Thr Leu Asp Thr Ile Gly Leu Cys Gly Phe
145                 150                 155                 160
Asn Tyr Arg Phe Asn Ser Phe Tyr Arg Glu Gly Gln His Pro Phe Ile
                165                 170                 175
Glu Ser Met Val Arg Gly Leu Ser Glu Ala Met Arg Gln Thr Lys Arg
            180                 185                 190
Phe Pro Leu Gln Asp Lys Leu Met Ile Gln Thr Lys Arg Arg Phe Asn
          195                 200                 205
Ser Asp Val Glu Ser Met Phe Ser Leu Val Asp Arg Ile Ile Ala Asp
      210                 215                 220
Arg Lys Gln Ala Glu Ser Glu Ser Gly Asn Asp Leu Leu Ser Leu Met
225                 230                 235                 240
Leu His Ala Lys Asp Pro Glu Thr Gly Glu Lys Leu Asp Asp Glu Asn
                245                 250                 255
Ile Arg Tyr Gln Ile Ile Thr Phe Leu Ile Ala Gly His Glu Thr Thr
            260                 265                 270
Ser Gly Leu Leu Ser Phe Ala Ile Tyr Leu Leu Leu Lys His Pro Asp
          275                 280                 285
Lys Leu Lys Lys Ala Tyr Glu Glu Ala Asp Arg Val Leu Thr Asp Pro
      290                 295                 300
Val Pro Ser Tyr Lys Gln Val Gln Gln Leu Lys Tyr Ile Arg Met Ile
305                 310                 315                 320
Leu Asn Glu Ser Ile Arg Leu Trp Pro Thr Ala Pro Ala Phe Ser Leu
                325                 330                 335
Tyr Ala Lys Glu Glu Thr Val Ile Gly Gly Lys Tyr Leu Ile Pro Lys
            340                 345                 350
Gly Gln Ser Val Thr Val Leu Ile Pro Lys Leu His Arg Asp Gln Ser
          355                 360                 365
Val Trp Gly Glu Asp Ala Glu Ala Phe Arg Pro Glu Arg Phe Glu Gln
      370                 375                 380
Met Asp Ser Ile Pro Ala His Ala Tyr Lys Pro Phe Gly Asn Gly Gln
385                 390                 395                 400
Arg Ala Cys Ile Gly Met Gln Phe Ala Leu His Glu Ala Thr Leu Val
                405                 410                 415
Leu Gly Met Ile Leu Gln Tyr Phe Asp Leu Glu Asp His Ala Asn Tyr
            420                 425                 430
Gln Leu Lys Ile Lys Glu Ser Leu Thr Leu Lys Pro Asp Gly Phe Thr
          435                 440                 445
Ile Arg Val Arg Pro Arg Lys Lys Glu Ala Met Thr Ala Met Pro Gly
      450                 455                 460
Ala Gln Pro Glu Glu Asn Gly Arg Gln Glu Glu Arg Pro Ser Ala Pro
465                 470                 475                 480
Ala Ala Glu Asn Thr His Gly Thr Pro Leu Leu Val Leu Tyr Gly Ser
                485                 490                 495
Asn Leu Gly Thr Ala Glu Glu Ile Ala Lys Glu Leu Ala Glu Glu Ala
            500                 505                 510
Arg Glu Gln Gly Phe His Ser Arg Thr Ala Glu Leu Asp Gln Tyr Ala
          515                 520                 525
Gly Ala Ile Pro Ala Glu Gly Ala Val Ile Ile Val Thr Ala Ser Tyr
      530                 535                 540
Asn Gly Asn Pro Pro Asp Cys Ala Lys Glu Phe Val Asn Trp Leu Glu
545                 550                 555                 560
His Asp Gln Thr Asp Asp Leu Arg Gly Val Lys Tyr Ala Val Phe Gly
                565                 570                 575
Cys Gly Asn Arg Ser Trp Ala Ser Thr Tyr Gln Arg Ile Pro Arg Leu
            580                 585                 590
Ile Asp Ser Val Leu Glu Lys Lys Gly Ala Gln Arg Leu His Lys Leu
          595                 600                 605
Gly Glu Gly Asp Ala Gly Asp Asp Phe Glu Gly Gln Phe Glu Ser Trp
      610                 615                 620
Lys Tyr Asp Leu Trp Pro Leu Leu Arg Thr Glu Phe Ser Leu Ala Glu
625                 630                 635                 640
Pro Glu Pro Asn Gln Thr Glu Thr Asp Arg Gln Ala Leu Ser Val Glu
                645                 650                 655
Phe Val Asn Ala Pro Ala Ala Ser Pro Leu Ala Lys Ala Tyr Gln Val
            660                 665                 670
Phe Thr Ala Lys Ile Ser Ala Asn Arg Glu Leu Gln Cys Glu Lys Ser
          675                 680                 685
Gly Arg Ser Thr Arg His Ile Glu Ile Ser Leu Pro Glu Gly Ala Ala
      690                 695                 700
Tyr Gln Glu Gly Asp His Leu Gly Val Leu Pro Gln Asn Ser Glu Val
705                 710                 715                 720
Leu Ile Gly Arg Val Phe Gln Arg Phe Gly Leu Asn Gly Asn Glu Gln
```

[0067]

```
                725                 730                 735
Ile Leu Ile Ser Gly Arg Asn Gln Ala Ser His Leu Pro Leu Glu Arg
            740                 745                 750
Pro Val His Val Lys Asp Leu Phe Gln His Cys Val Glu Leu Gln Glu
        755                 760                 765
Pro Ala Thr Arg Ala Gln Ile Arg Glu Leu Ala Ala His Thr Val Cys
    770                 775                 780
Pro Pro His Gln Arg Glu Leu Glu Asp Leu Leu Lys Asp Asp Val Tyr
785                 790                 795                 800
Lys Asp Gln Val Leu Asn Lys Arg Leu Thr Met Leu Asp Leu Leu Glu
                805                 810                 815
Gln Tyr Pro Ala Cys Glu Leu Pro Phe Ala Arg Phe Leu Ala Leu Leu
            820                 825                 830
Pro Pro Leu Lys Pro Arg Tyr Tyr Ser Ile Ser Ser Ser Pro Gln Leu
        835                 840                 845
Asn Pro Arg Gln Thr Ser Ile Thr Val Ser Val Val Ser Gly Pro Ala
    850                 855                 860
Leu Ser Gly Arg Gly His Tyr Lys Gly Val Ala Ser Asn Tyr Leu Ala
865                 870                 875                 880
Gly Leu Glu Pro Gly Asp Ala Ile Ser Cys Phe Ile Arg Glu Pro Gln
                885                 890                 895
Ser Gly Phe Arg Leu Pro Glu Asp Pro Glu Thr Pro Val Ile Met Val
            900                 905                 910
Gly Pro Gly Thr Gly Ile Ala Pro Tyr Arg Gly Phe Leu Gln Ala Arg
        915                 920                 925
Arg Ile Gln Arg Asp Ala Gly Val Lys Leu Gly Glu Ala His Leu Tyr
    930                 935                 940
Phe Gly Cys Arg Arg Pro Asn Glu Asp Phe Leu Tyr Arg Asp Glu Leu
945                 950                 955                 960
Glu Gln Ala Glu Lys Asp Gly Ile Val His Leu His Thr Ala Phe Ser
                965                 970                 975
Arg Leu Glu Gly Arg Pro Lys Thr Tyr Val Gln Asp Leu Leu Arg Glu
            980                 985                 990
Asp Ala Ala Leu Leu Ile His Leu  Leu Asn Glu Gly Gly  Arg Leu Tyr
        995                1000                1005
Val Cys  Gly Asp Gly Ser Arg  Met Ala Pro Ala Val  Glu Gln Ala
    1010                1015                1020
Leu Cys  Glu Ala Tyr Arg Ile  Val Gln Gly Ala Ser  Arg Glu Glu
    1025                1030                1035
Ser Gln  Ser Trp Leu Ser Ala  Leu Leu Glu Glu Gly  Arg Tyr Ala
    1040                1045                1050
Lys Asp  Val Trp Asp Gly Gly  Val Ser Gln His Asn  Val Lys Ala
    1055                1060                1065
Asp Cys  Ile Ala Arg Thr
    1070
```

<210> 65
<211> 16
<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> 65
taaggaggaa aacaaa 16

<210> 66
<211> 1677
<212> DNA
<213> 恶臭假单胞菌
<400> 66

```
atgtacgact atataatcgt tggtgctgga tctgcaggat gtgtgcttgc taatcgtctt   60
tcggccgacc cctctaaaag agtttgttta cttgaagctg ggccgcgaga tacgaatccg  120
ctaattcata tgccgttagg tattgctttg ctttcaaata gtaaaaagtt gaattgggct  180
tttcaaactg cgccacagca aaatctcaac ggccggagcc ttttctggcc acgaggaaaa  240
acgttaggtg gttcaagctc aatcaacgca atggtctata tccgagggca tgaagacgat  300
taccacgcat gggagcaggc ggccggccgc tactggggtt ggtaccgggc tcttgagttg  360
ttcaaaaggc ttgaatgcaa ccagcgattc gataagtccg agcaccatgg ggttgacgga  420
gaattagctg ttagtgattt aaaatatatc aatccgctta gcaaagcatt cgtgcaagcc  480
ggcatggagg ccaatattaa tttcaacgga gatttcaacg gcgagtacca ggacggcgta  540
gggttctatc aagtaaccca aaaaaatgga caacgctgga gctcggcgcg tgcattcttg  600
cacggtgtac tttccagacc aaatctagac atcattactg atgcgcatgc atcaaaaatt  660
ctttttgaag accgtaaggc ggttggtgtt tcttatataa agaaaaatat gcaccatcaa  720
gtcaagacaa cgagtggtgg tgaagtactt cttagtcttg gcgcagtcgg cacgcctcac  780
cttctaatgc tttctggtgt tggggctgca gccgagctta aggaacatgg tgtttctcta  840
gtccatgatc ttcctgaggt ggggaaaaat cttcaagatc atttggacat cacattgatg  900
```

```
tgcgcagcaa attcgagaga gccgataggt gttgctcttt ctttcatccc tcgtggtgtc      960
tcgggtttgt tttcatatgt gtttaagcgc gaggggtttc tcactagtaa cgtggcagag     1020
tcgggtggtt ttgtaaaaag ttctcctgat cgtgatcggc ccaatttgca gtttcatttc     1080
cttccaactt atcttaaaga tcacggtcga aaaatagcgg gtggttatgg ttatacgcta     1140
catatatgtg atcttttgcc taagagccga ggcagaattg gcctaaaaag cgccaatcca     1200
ttacagccgc ctttaattga cccgaactat cttagcgatc atgaagatat taaaaccatg     1260
attgcgggta ttaagatagg gcgcgctatt ttgcaggccc catcgatggc gaagcatttt     1320
aagcatgaag tagtaccggg ccaggctgtt aaaactgatg atgaaataat cgaagatatt     1380
cgtaggcgag ctgagactat ataccatccg gtaggtactt gtaggatggg taaagatcca     1440
gcgtcagttg ttgatccgtg cctgaagatc cgtgggttgg caaatattag agtcgttgat     1500
gcgtcaatta tgccgcactt ggtcgcgggt aacacaaacg ctccaactat tatgattgca     1560
gaaaatgcgg cagaaataat tatgcggaat cttgatgtgg aagcattaga ggctagcgct     1620
gagtttgctc gcgagggtgc agagctagag ttggccatga tagctgtctg catgtaa        1677

<210> 67
<211> 558
<212> PRT
<213> 恶臭假单胞菌
<400> 67
Met Tyr Asp Tyr Ile Ile Val Gly Ala Gly Ser Ala Gly Cys Val Leu
1               5                  10                  15
Ala Asn Arg Leu Ser Ala Asp Pro Ser Lys Arg Val Cys Leu Leu Glu
            20                  25                  30
Ala Gly Pro Arg Asp Thr Asn Pro Leu Ile His Met Pro Leu Gly Ile
        35                  40                  45
Ala Leu Leu Ser Asn Ser Lys Lys Leu Asn Trp Ala Phe Gln Thr Ala
    50                  55                  60
Pro Gln Gln Asn Leu Asn Gly Arg Ser Leu Phe Trp Pro Arg Gly Lys
65                  70                  75                  80
Thr Leu Gly Gly Ser Ser Ser Ile Asn Ala Met Val Tyr Ile Arg Gly
                85                  90                  95
His Glu Asp Asp Tyr His Ala Trp Glu Gln Ala Ala Gly Arg Tyr Trp
            100                 105                 110
Gly Trp Tyr Arg Ala Leu Glu Leu Phe Lys Arg Leu Glu Cys Asn Gln
        115                 120                 125
Arg Phe Asp Lys Ser Glu His His Gly Val Asp Gly Glu Leu Ala Val
    130                 135                 140
Ser Asp Leu Lys Tyr Ile Asn Pro Leu Ser Lys Ala Phe Val Gln Ala
145                 150                 155                 160
Gly Met Glu Ala Asn Ile Asn Phe Asn Gly Asp Phe Asn Gly Glu Tyr
                165                 170                 175
Gln Asp Gly Val Gly Phe Tyr Gln Val Thr Gln Lys Asn Gly Gln Arg
            180                 185                 190
Trp Ser Ser Ala Arg Ala Phe Leu His Gly Val Leu Ser Arg Pro Asn
        195                 200                 205
Leu Asp Ile Ile Thr Asp Ala His Ala Ser Lys Ile Leu Phe Glu Asp
    210                 215                 220
Arg Lys Ala Val Gly Val Ser Tyr Ile Lys Lys Asn Met His His Gln
225                 230                 235                 240
Val Lys Thr Thr Ser Gly Gly Glu Val Leu Leu Ser Leu Gly Ala Val
                245                 250                 255
Gly Thr Pro His Leu Leu Met Leu Ser Gly Val Gly Ala Ala Ala Glu
            260                 265                 270
Leu Lys Glu His Gly Val Ser Leu Val His Asp Leu Pro Glu Val Gly
        275                 280                 285
Lys Asn Leu Gln Asp His Leu Asp Ile Thr Leu Met Cys Ala Ala Asn
    290                 295                 300
Ser Arg Glu Pro Ile Gly Val Ala Leu Ser Phe Ile Pro Arg Gly Val
305                 310                 315                 320
Ser Gly Leu Phe Ser Tyr Val Phe Lys Arg Glu Gly Phe Leu Thr Ser
                325                 330                 335
Asn Val Ala Glu Ser Gly Gly Phe Val Lys Ser Ser Pro Asp Arg Asp
            340                 345                 350
Arg Pro Asn Leu Gln Phe His Phe Leu Pro Thr Tyr Leu Lys Asp His
        355                 360                 365
Gly Arg Lys Ile Ala Gly Gly Tyr Gly Tyr Thr Leu His Ile Cys Asp
    370                 375                 380
Leu Leu Pro Lys Ser Arg Gly Arg Ile Gly Leu Lys Ser Ala Asn Pro
385                 390                 395                 400
Leu Gln Pro Pro Leu Ile Asp Pro Asn Tyr Leu Ser Asp His Glu Asp
                405                 410                 415
Ile Lys Thr Met Ile Ala Gly Ile Lys Ile Gly Arg Ala Ile Leu Gln
            420                 425                 430
Ala Pro Ser Met Ala Lys His Phe Lys His Glu Val Val Pro Gly Gln
        435                 440                 445
Ala Val Lys Thr Asp Asp Glu Ile Ile Glu Asp Ile Arg Arg Arg Ala
```

[0068]

```
    450                 455                 460
Glu Thr Ile Tyr His Pro Val Gly Thr Cys Arg Met Gly Lys Asp Pro
465                 470                 475                 480
Ala Ser Val Val Asp Pro Cys Leu Lys Ile Arg Gly Leu Ala Asn Ile
                485                 490                 495
Arg Val Val Asp Ala Ser Ile Met Pro His Leu Val Ala Gly Asn Thr
            500                 505                 510
Asn Ala Pro Thr Ile Met Ile Ala Glu Asn Ala Ala Glu Ile Ile Met
        515                 520                 525
Arg Asn Leu Asp Val Glu Ala Leu Glu Ala Ser Ala Glu Phe Ala Arg
    530                 535                 540
Glu Gly Ala Glu Leu Glu Leu Ala Met Ile Ala Val Cys Met
545                 550                 555
```

<210> 68
<211> 1452
<212> DNA
<213> 恶臭假单胞菌
<400> 68

[0069]

```
atgaccatac caattagcct agccaagtta aactctagtg ccgataccca ttcagcgctt      60
gaagtattta atttgcagaa agttgcaagt agtgcgcgtc gtggtaaatt tggcatagca     120
gagcgcatcg ctgctcttaa tttacttaag gaaactattc agcgtcgtga gcctgaaatt     180
attgctgcac ttgcagcgga ctttcgcaag ccggcaagcg aggtgaagct aacagaaatc     240
tttccggtat tgcaagaaat taatcatgcc aaacggaacc ttaaagattg gatgaagcca     300
cggcgagtga gggcggcact tagtgtagcg ggcacgcggg caggacttcg ttacgagcct     360
aagggtgtct gtttgataat tgcgccgtgg aactatccat tcaaccttag tttcggtcct     420
cttgtatctg cgttagcggc aggaaatagc gttgttataa agccgtctga attgacacca     480
cacactgcaa cactgatcgg atctatagtc agggaggcat tctctgtcga cctagtcgct     540
gtggtggagg gtgatgccgc agtttcccag gagctgttgg ctctgccatt tgaccatatt     600
ttttttactg gtagtcctag ggtcggcaag ttagtgatgg aagcggcgtc aaaaacactc     660
gcttcggtta ctttggagtt aggcggaaaa tctccaacca ttattggacc aacagcaaat     720
ttgccgaaag ctgcgcgcaa catagtgtgg ggaaagtttt caaacaacgg ccagacgtgc     780
atagcgcctg atcacgtatt tgttcatcgg tgtatagccc agaaattcaa tgaaattctt     840
gtgaaagaga ttgtgcgagt ttatgggaag gattttgctg cgcagcgtag atcggcagac     900
tattgcagga tcgtcaatga tcaacatttc aatcgaatta ataaactcct gactgacgcg     960
aaagctaaag gtgcaaaaat tctgcaaggg ggtcaagttg acgcgactga gaggcttgtg    1020
gtgccaacgg ttttatctaa cgtcactgct gctatggata ttaaccatga ggaaatattc    1080
gggccgctac ttcctataat tgaatacgat gatatagatt ctgtaattaa gcgtgtgaat    1140
gacggtgaca agcccctggc gctgtatgtc ttttctgaag ataaacaatt tgtaaataac    1200
atcgtggctc gtacaagctc tggttcggtc ggagttaatc tgagtgtcgt gcacttttg     1260
caccctaatc tcccatttgg cggtgtcaat aatagtggta tcggcagtgc tcatggagtt    1320
tacgggttca gggcgttttc tcacgaaaaa ccagttctta tagataagtt ctcaatcacg    1380
cattggttgt ttccgcctta taccaagaag gtgaagcagt tgattggtat cacagttaag    1440
tatttgagct ga                                                       1452
```

<210> 69
<211> 483
<212> PRT
<213> 恶臭假单胞菌
<400> 69

```
Met Thr Ile Pro Ile Ser Leu Ala Lys Leu Asn Ser Ser Ala Asp Thr
1               5                   10                  15
His Ser Ala Leu Glu Val Phe Asn Leu Gln Lys Val Ala Ser Ser Ala
            20                  25                  30
Arg Arg Gly Lys Phe Gly Ile Ala Glu Arg Ile Ala Ala Leu Asn Leu
        35                  40                  45
Leu Lys Glu Thr Ile Gln Arg Arg Glu Pro Glu Ile Ile Ala Ala Leu
    50                  55                  60
Ala Ala Asp Phe Arg Lys Pro Ala Ser Glu Val Lys Leu Thr Glu Ile
65                  70                  75                  80
Phe Pro Val Leu Gln Glu Ile Asn His Ala Lys Arg Asn Leu Lys Asp
                85                  90                  95
Trp Met Lys Pro Arg Arg Val Arg Ala Ala Leu Ser Val Ala Gly Thr
            100                 105                 110
Arg Ala Gly Leu Arg Tyr Glu Pro Lys Gly Val Cys Leu Ile Ile Ala
        115                 120                 125
Pro Trp Asn Tyr Pro Phe Asn Leu Ser Phe Gly Pro Leu Val Ser Ala
    130                 135                 140
Leu Ala Ala Gly Asn Ser Val Val Ile Lys Pro Ser Glu Leu Thr Pro
145                 150                 155                 160
His Thr Ala Thr Leu Ile Gly Ser Ile Val Arg Glu Ala Phe Ser Val
                165                 170                 175
Asp Leu Val Ala Val Val Glu Gly Asp Ala Ala Val Ser Gln Glu Leu
            180                 185                 190
Leu Ala Leu Pro Phe Asp His Ile Phe Phe Thr Gly Ser Pro Arg Val
        195                 200                 205
```

[0070]

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

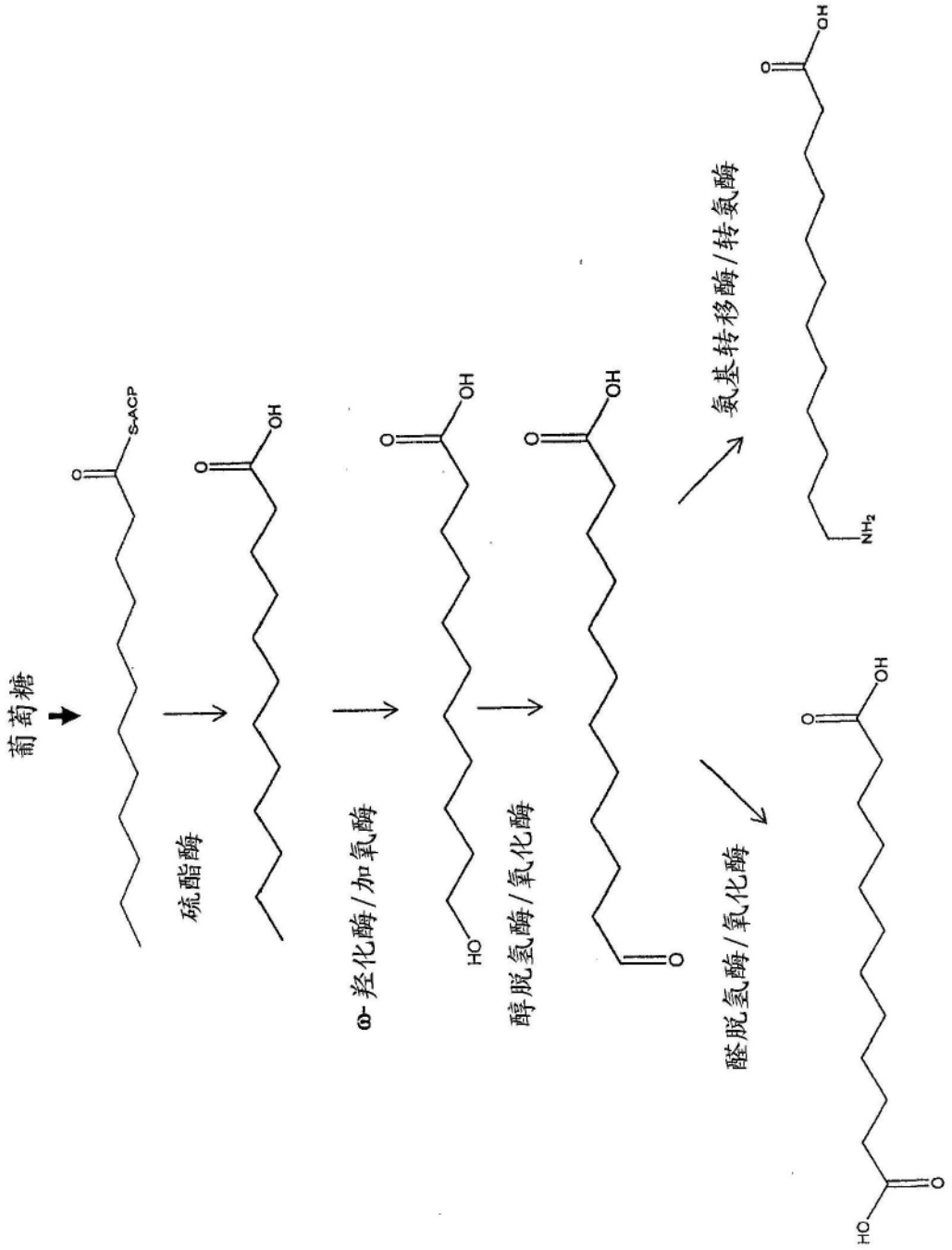
葡萄糖
S-ACP
硫酯酶
OH
ω-羟化酶/加氧酶
HO
OH
醇脱氢酶/氧化酶
OH
O
氨基转移酶/转氨酶
OH
NH2
醛脱氢酶/氧化酶
OH
HO
O

图1

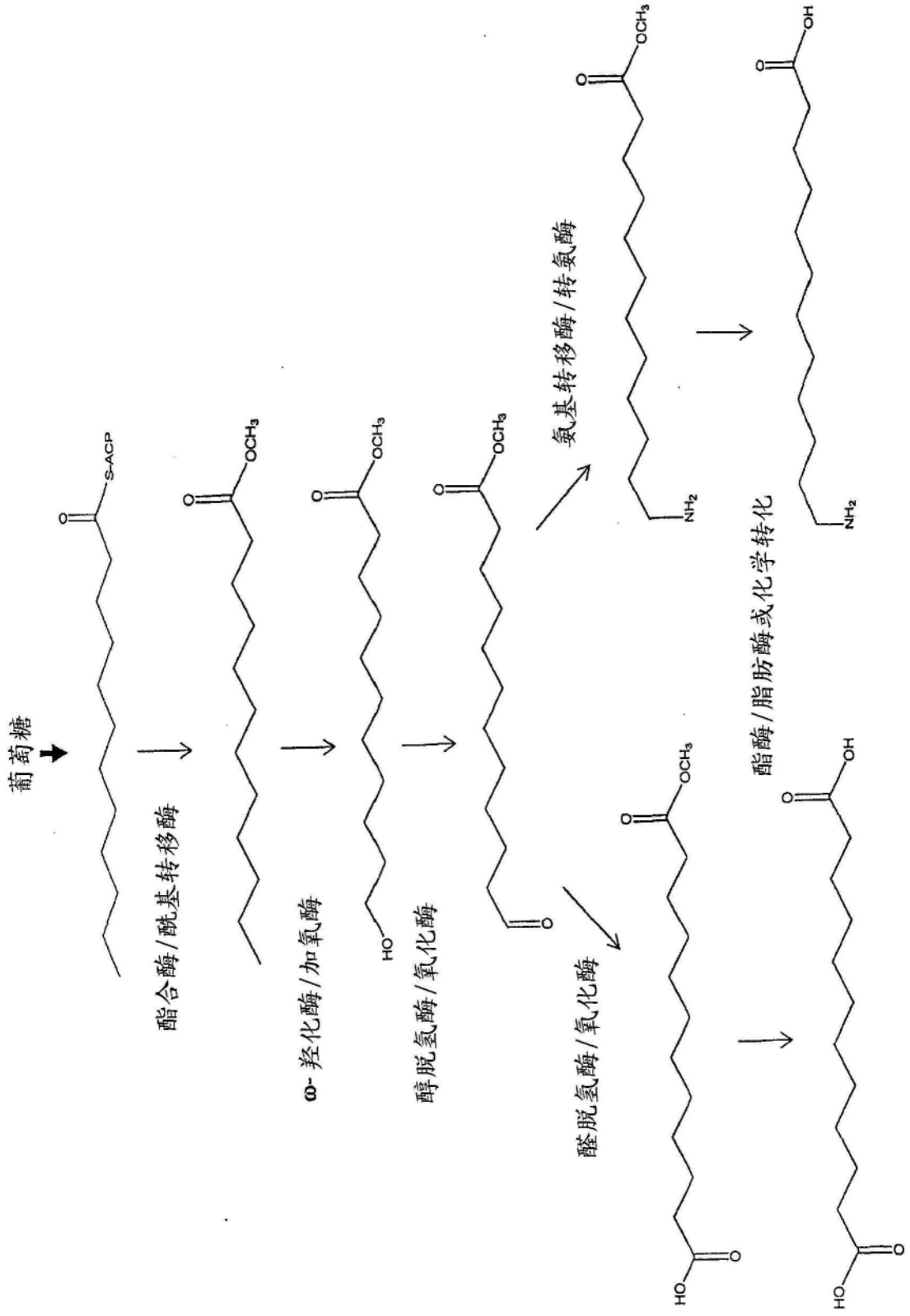
葡萄糖
S-ACP
酯合酶/酰基转移酶
OCH3
ω-羟化酶/加氧酶
HO
醇脱氢酶/氧化酶
氨基转移酶/转氨酶
NH2
醛脱氢酶/氧化酶
酯酶/脂肪酶或化学转化
OH

图2A

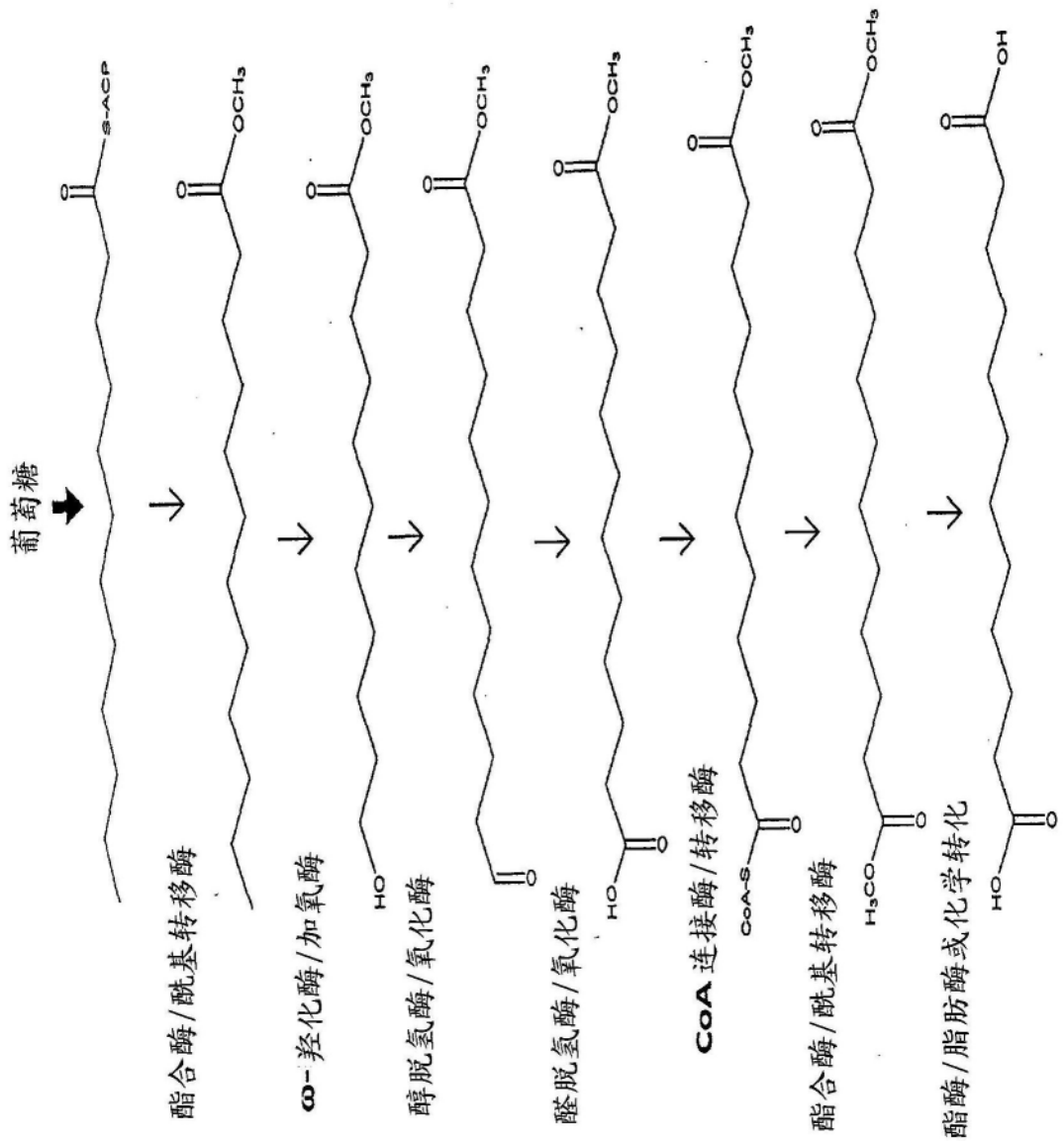
葡萄糖
S-ACP
OCH3
酯合酶/酰基转移酶
ω-羟化酶/加氧酶
HO
醇脱氢酶/氧化酶
O
醛脱氢酶/氧化酶
CoA 连接酶/转移酶
CoA-S
酯合酶/酰基转移酶
H3CO
酯酶/脂肪酶或化学转化
OH

图2B

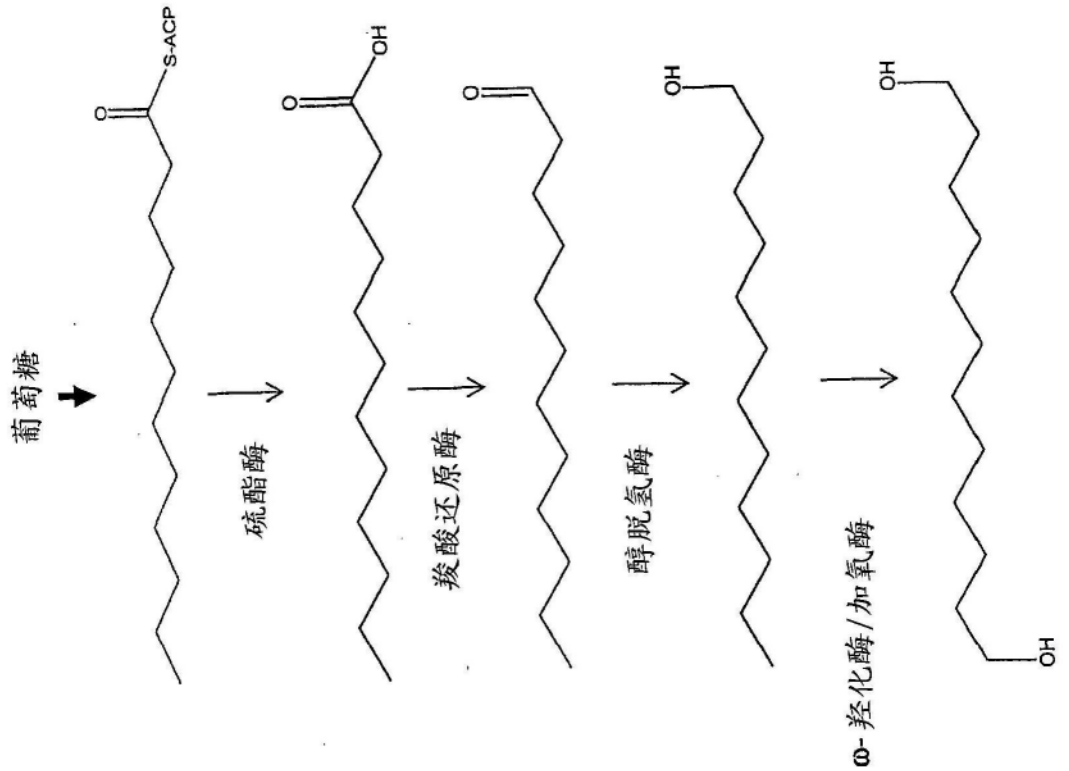
葡萄糖
S-ACP
硫酯酶
OH
羧酸还原酶
醇脱氢酶
ω- 羟化酶/加氧酶
OH

图3

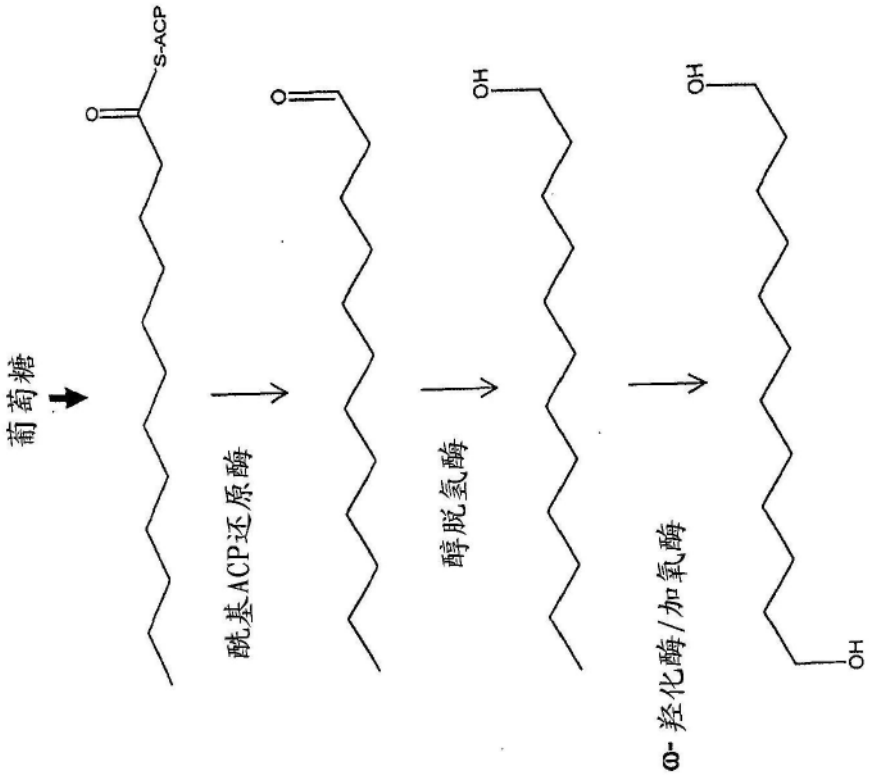
葡萄糖
S-ACP
酰基ACP还原酶
醇脱氢酶
OH
ω- 羟化酶/加氧酶
OH

图4

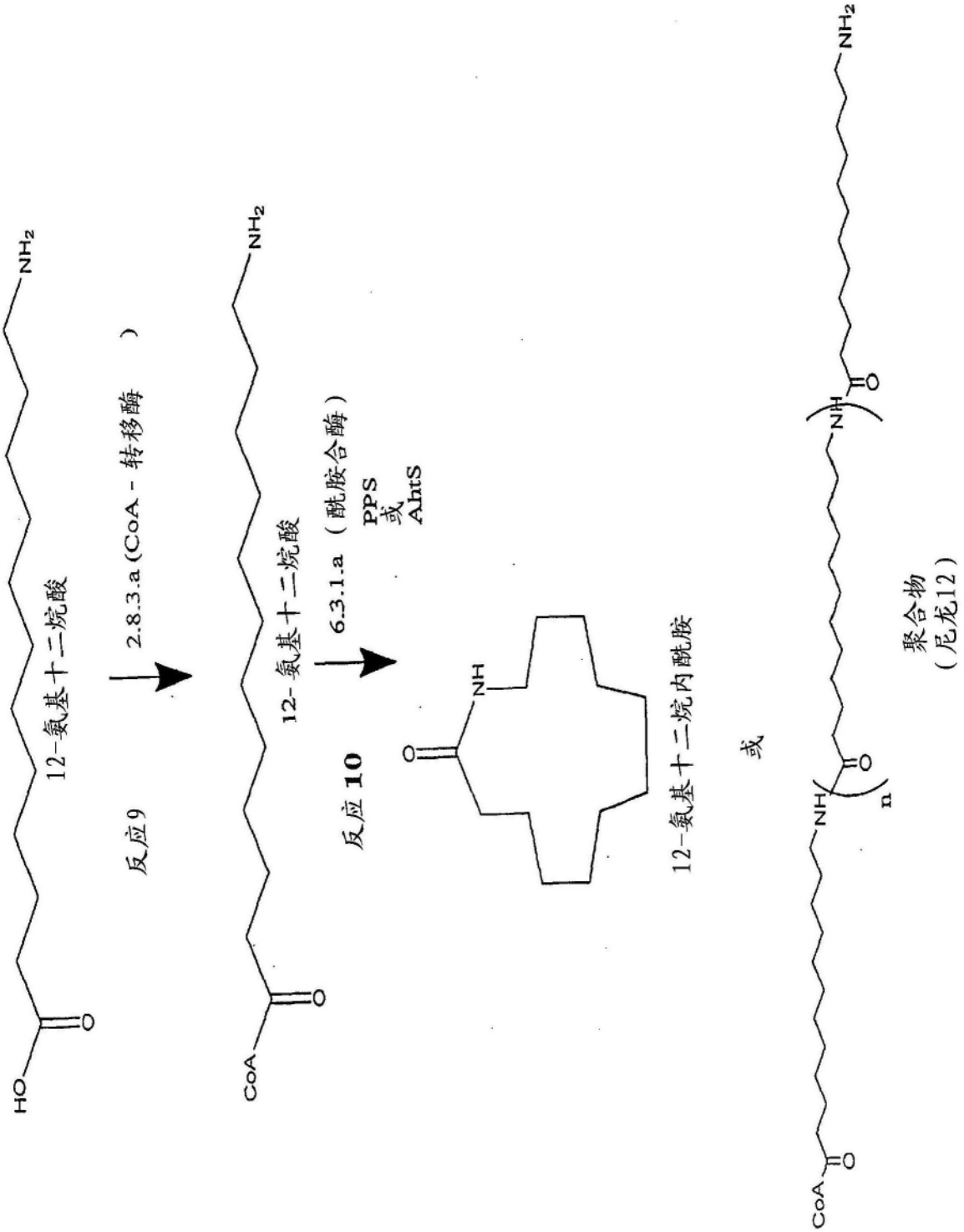
12-氨基十二烷酸
反应9
2.8.3.a (CoA - 转移酶)
12-氨基十二烷酸
反应 10
6.3.1.a （酰胺合酶）
PPS
或
AntS
12-氨基十二烷内酰胺
或
聚合物
（尼龙12）

图5

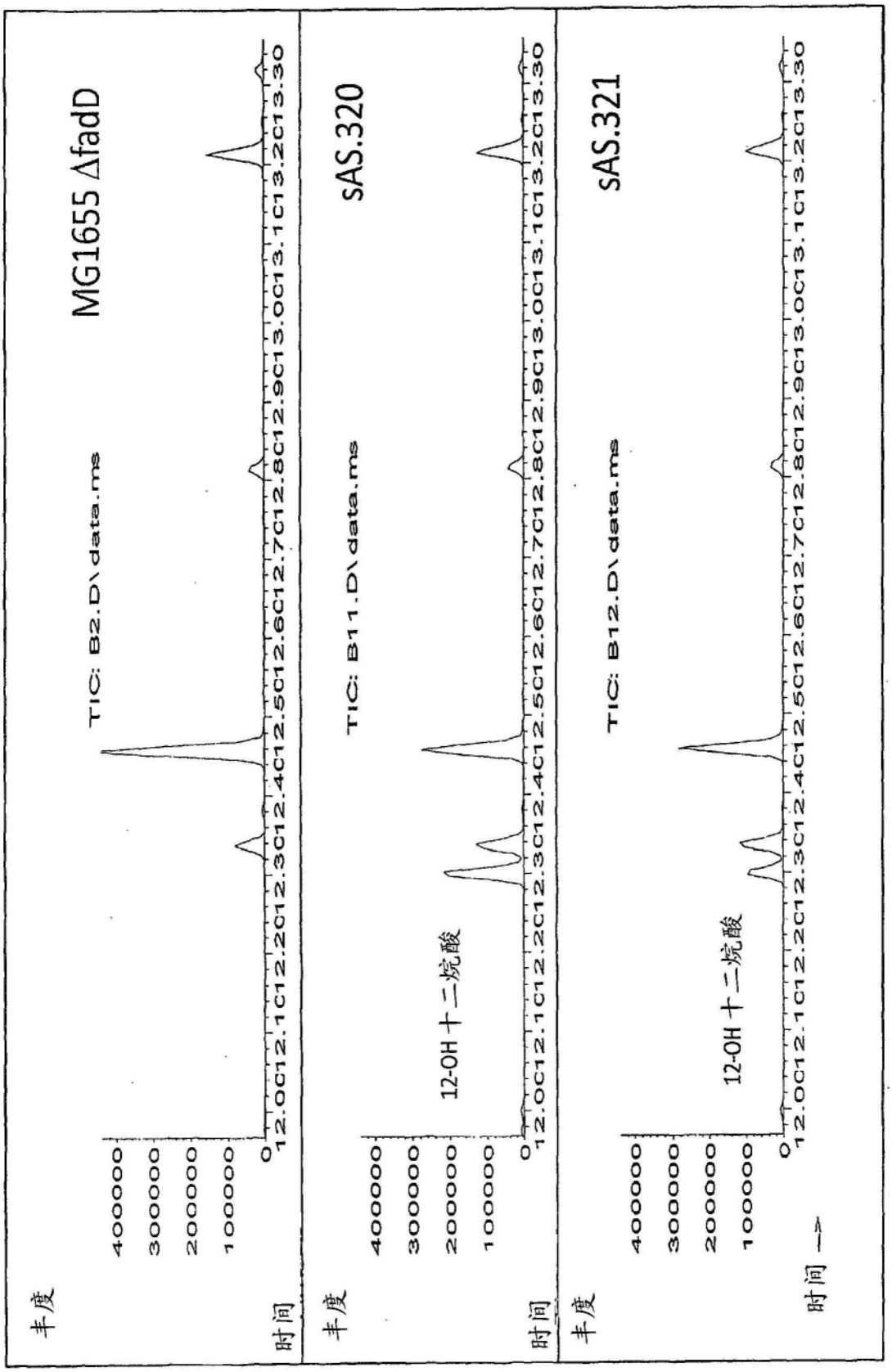
MG1655 ΔfadD
TIC: B2.D\data.ms
sAS.320
TIC: B11.D\data.ms
12-OH 十二烷酸
sAS.321
TIC: B12.D\data.ms
12-OH 十二烷酸
丰度
时间
400000
300000
200000
100000
0
12.0C12.1C12.2C12.3C12.4C12.5C12.6C12.7C12.8C12.9C13.0C13.1C13.2C13.30

图6

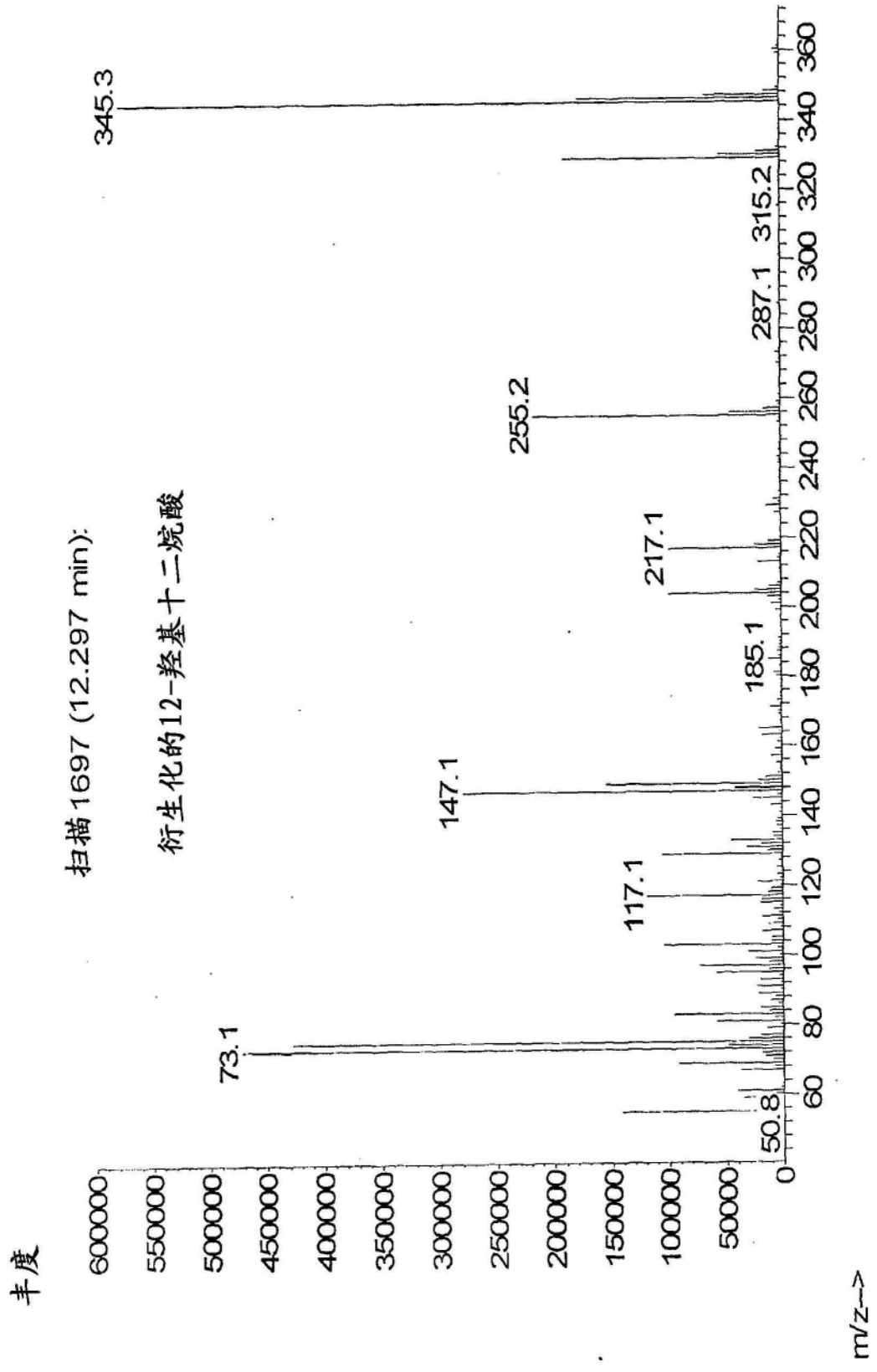
扫描1697 (12.297 min):
衍生化的12-羟基十二烷酸
丰度
m/z-->
345.3
315.2
287.1
255.2
217.1
185.1
147.1
117.1
73.1
50.8

图7A

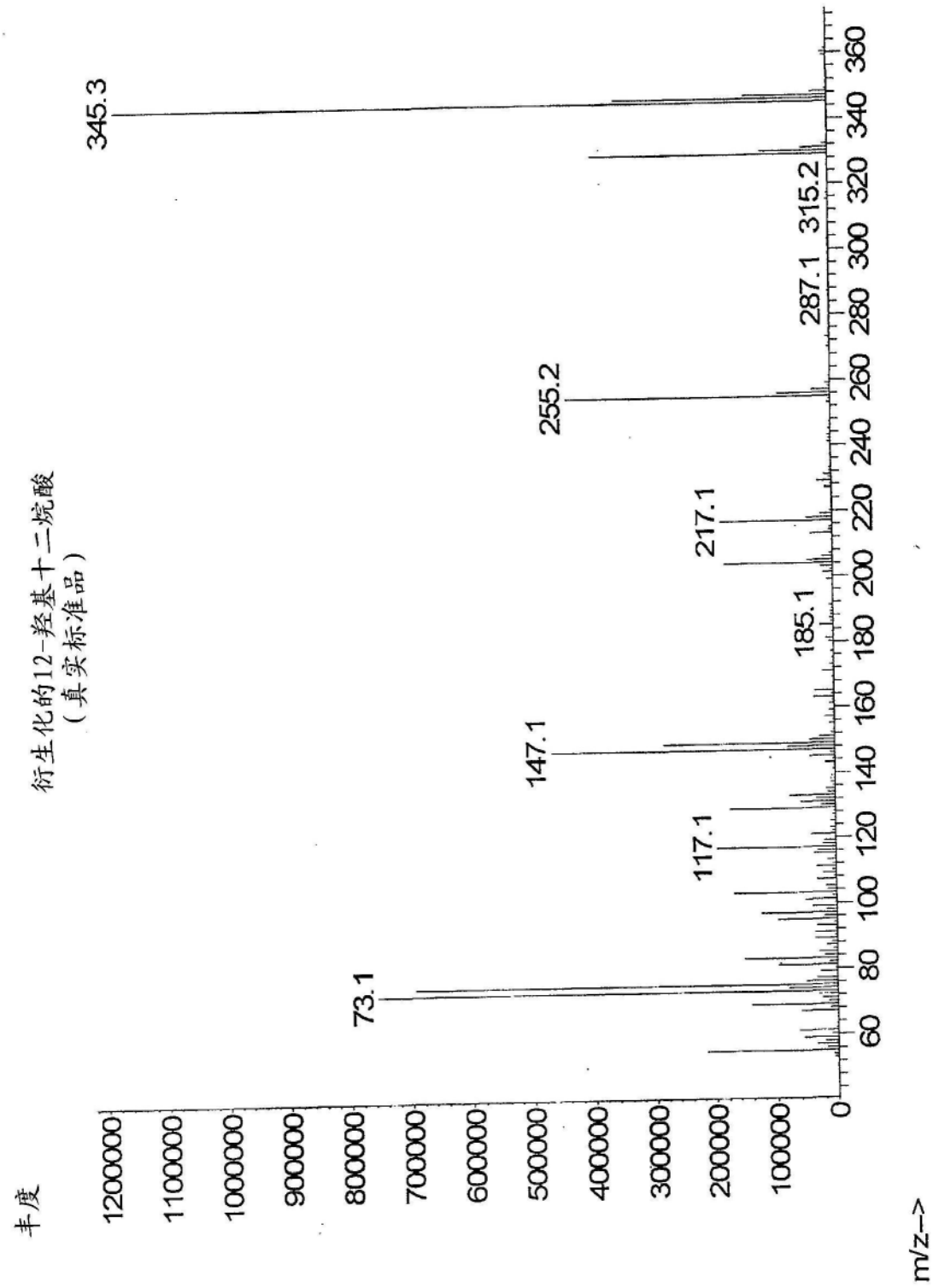
衍生化的12-羟基十二烷酸
（真实标准品）
丰度
m/z-->
345.3
315.2
287.1
255.2
217.1
185.1
147.1
117.1
73.1

图7B

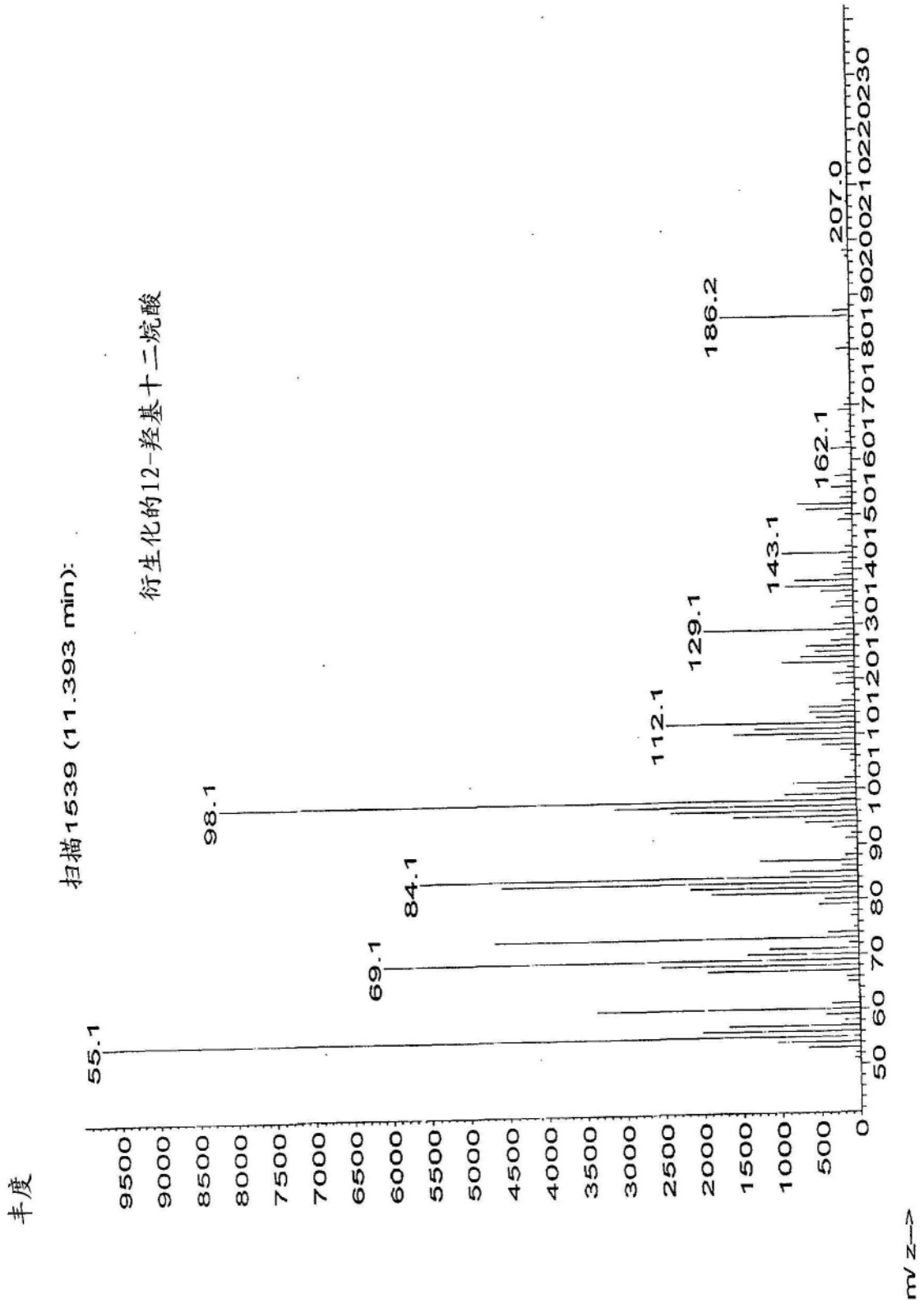
丰度
扫描1539 (11.393 min):
衍生化的12-羟基十二烷酸
55.1
98.1
69.1
84.1
112.1
129.1
143.1
162.1
186.2
207.0
m/z-->

图7C

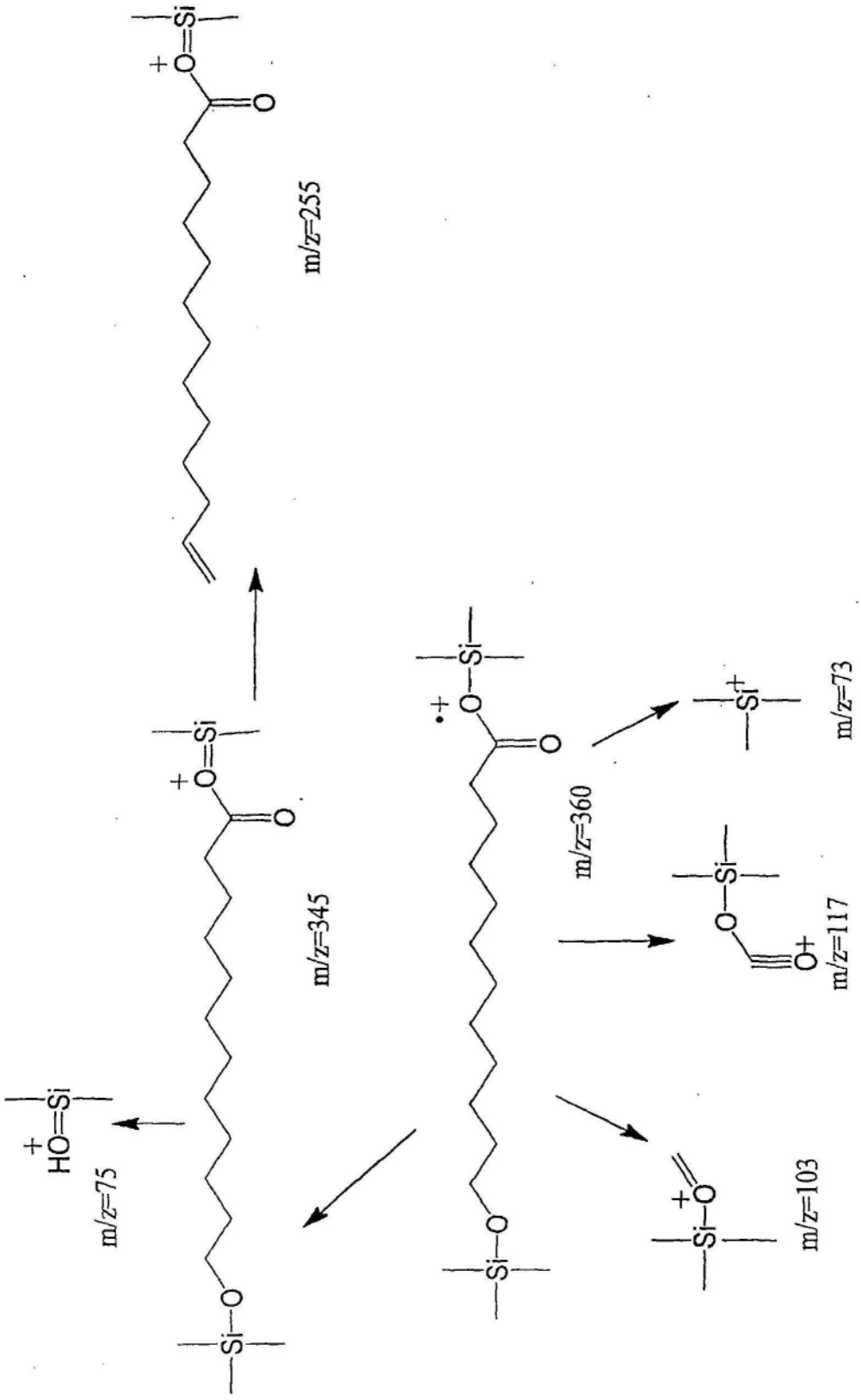

图8A

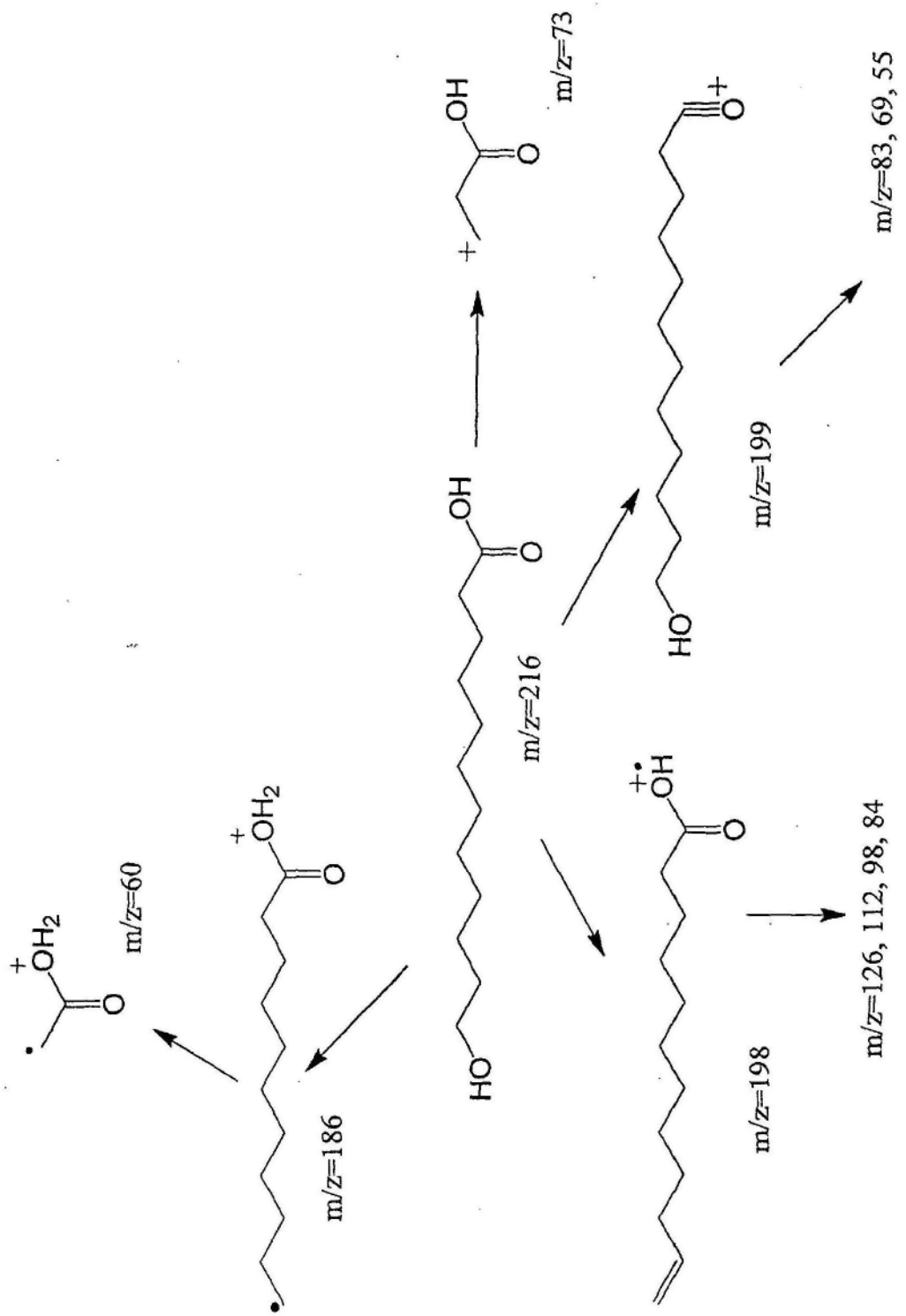

图8B

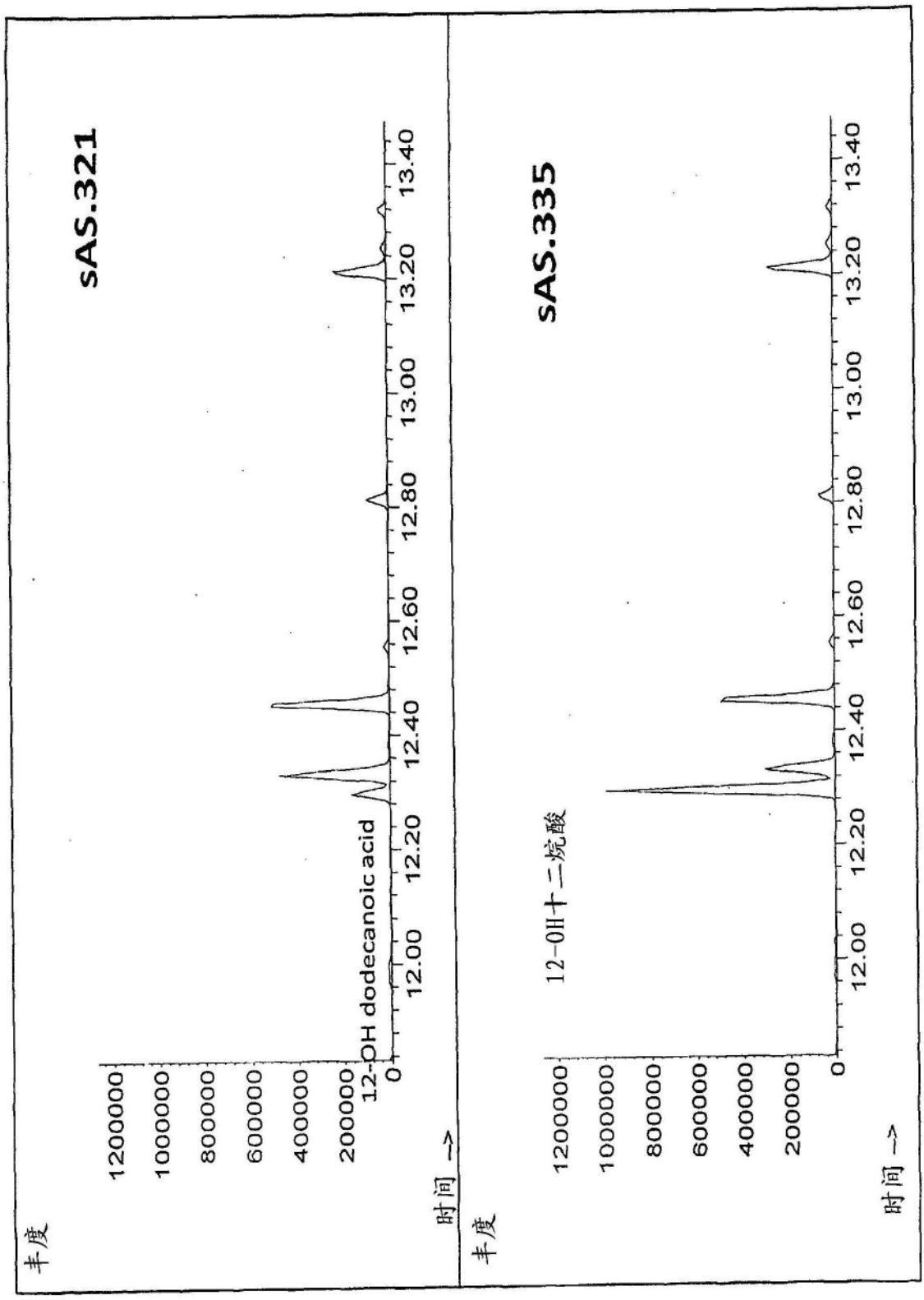
sAS.321
丰度
1200000
1000000
800000
600000
400000
200000
0
12-OH dodecanoic acid
12.00
12.20
12.40
12.60
12.80
13.00
13.20
13.40
时间 -->
sAS.335
丰度
12-OH十二烷酸
1200000
1000000
800000
600000
400000
200000
0
12.00
12.20
12.40
12.60
12.80
13.00
13.20
13.40
时间 -->

图9

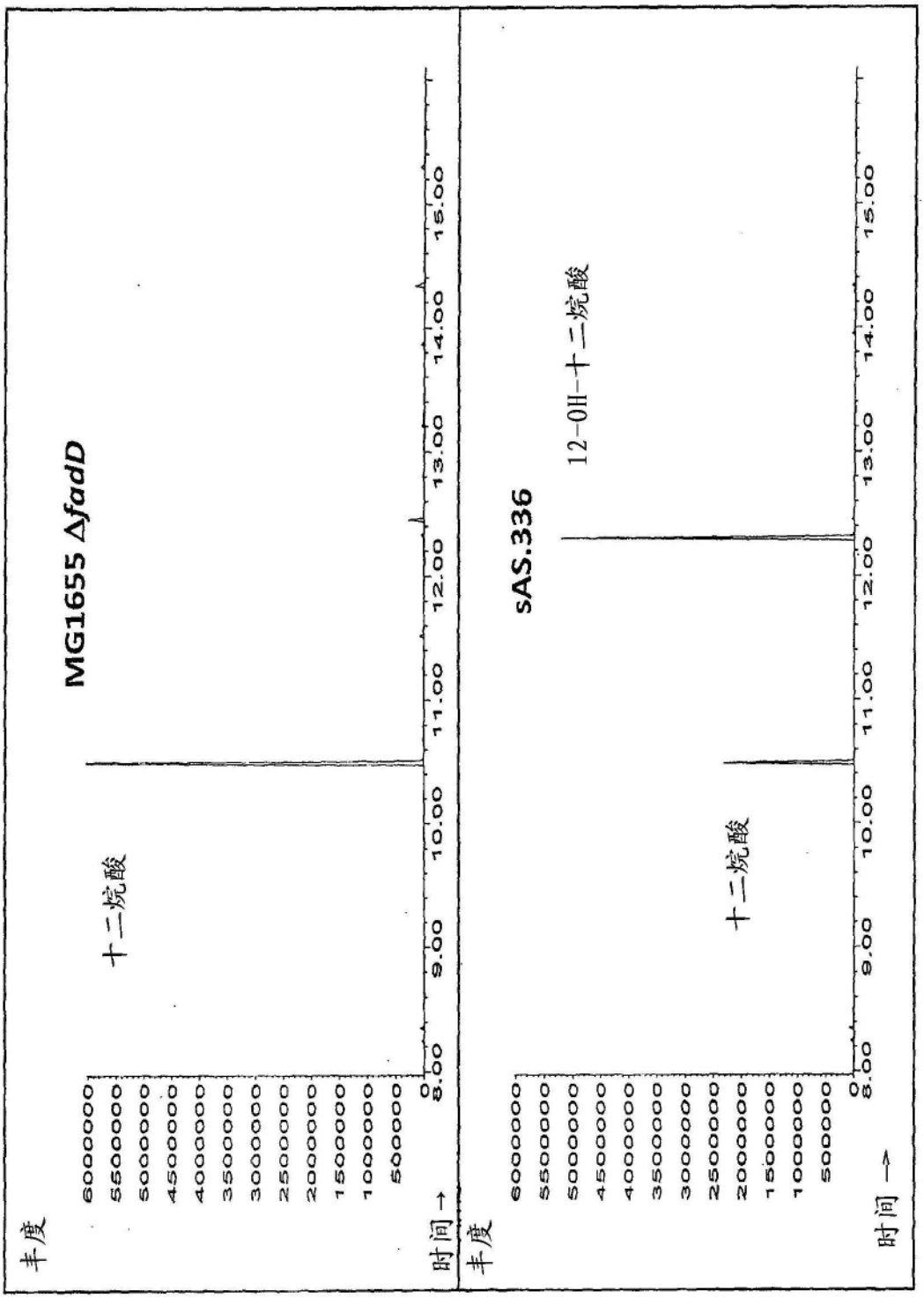

图10A

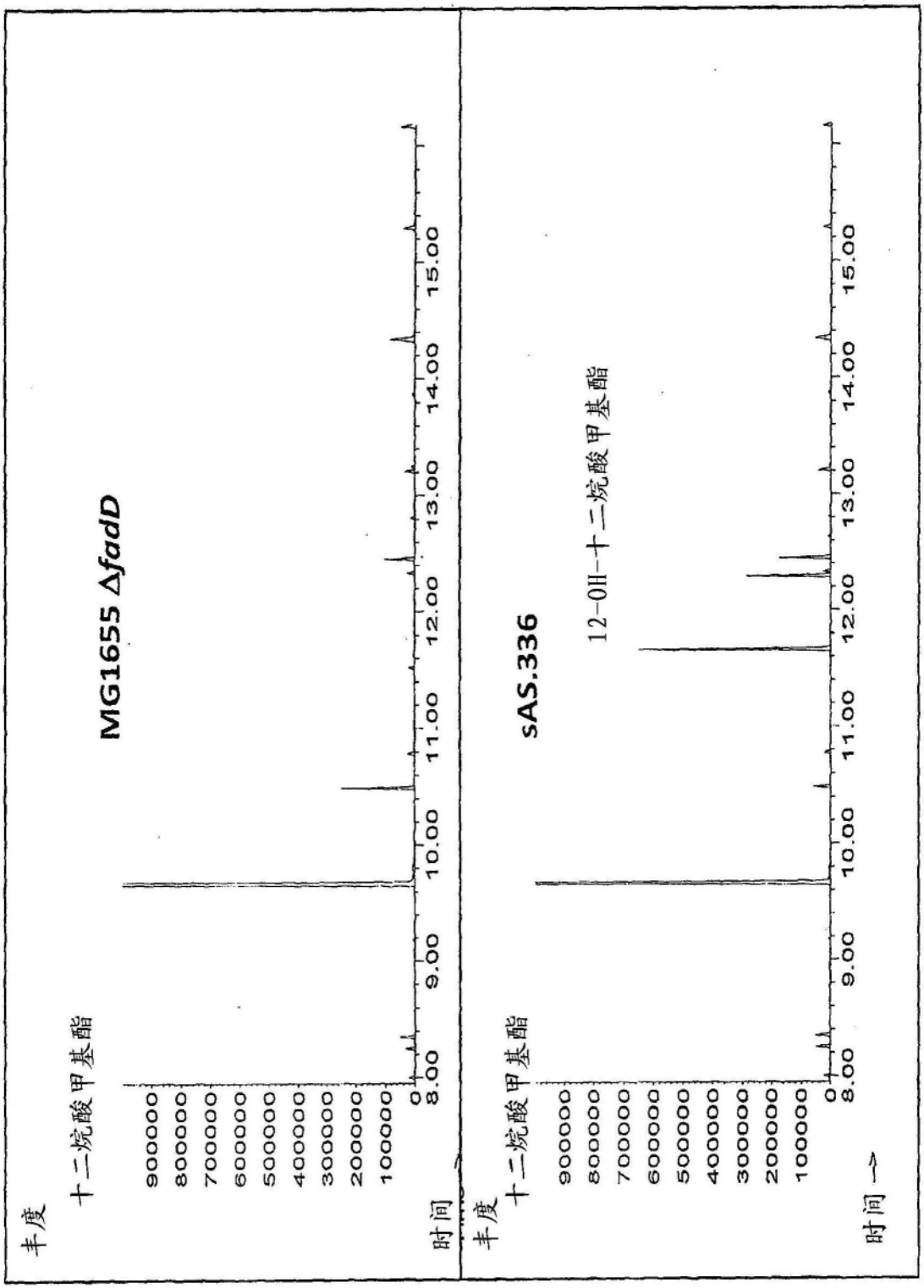
MG1655 ΔfadD
丰度
十二烷酸甲基酯
900000
800000
700000
600000
500000
400000
300000
200000
100000
0
8.00
9.00
10.00
11.00
12.00
13.00
14.00
15.00
时间 →
sAS.336
丰度
十二烷酸甲基酯
12-OH-十二烷酸甲基酯
900000
800000
700000
600000
500000
400000
300000
200000
100000
0
8.00
9.00
10.00
11.00
12.00
13.00
14.00
15.00
时间 →

图10B

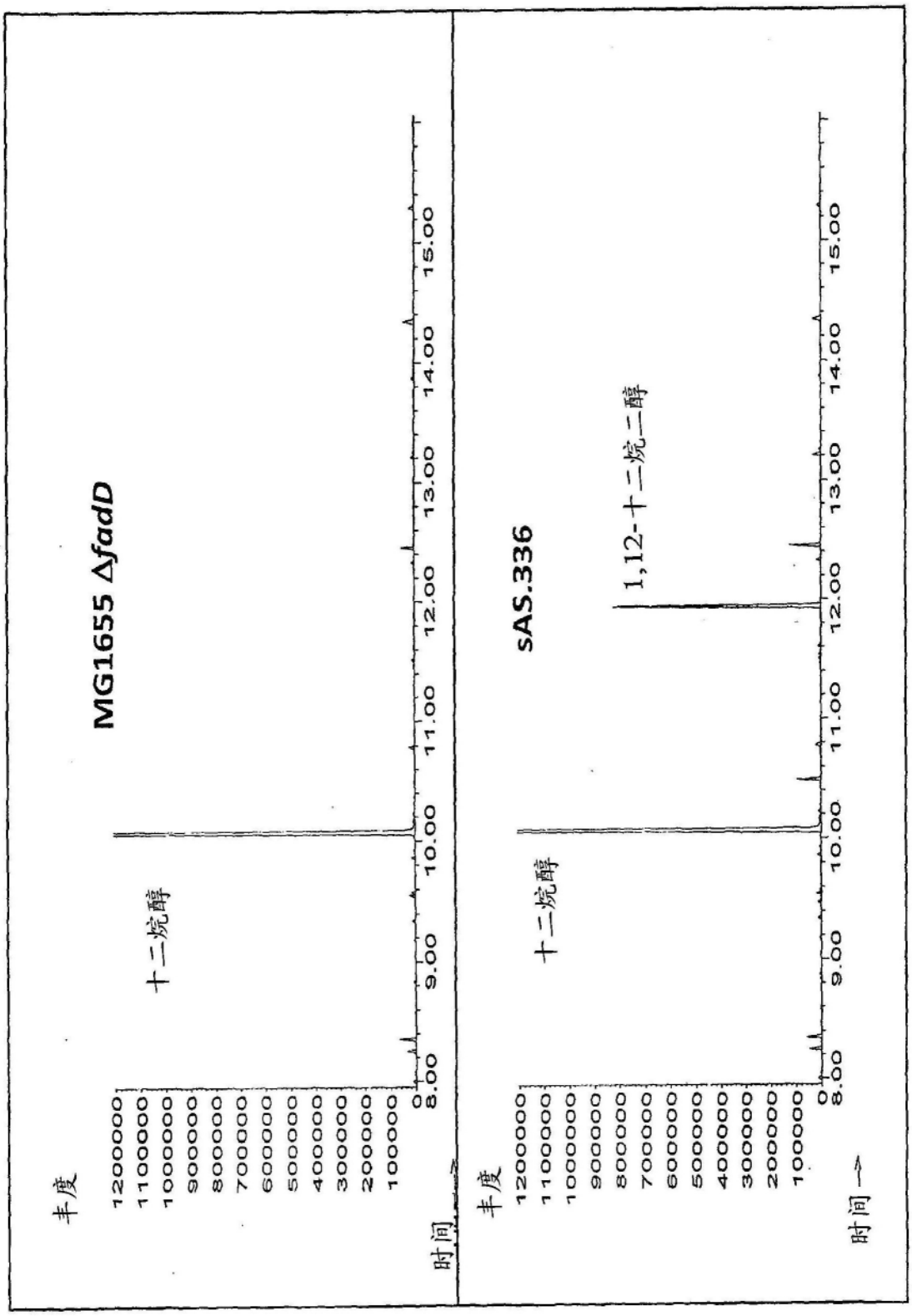

图10C

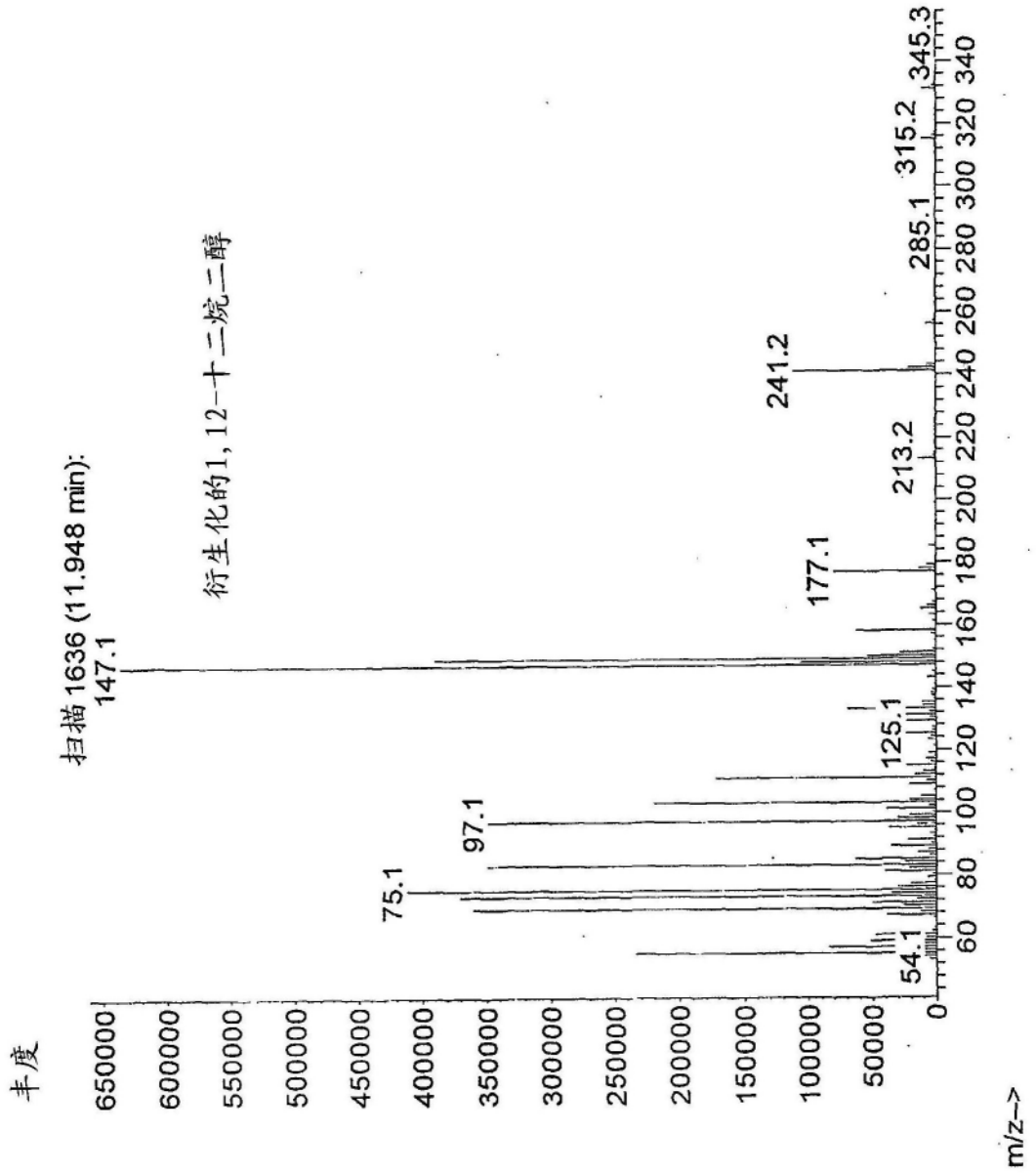
扫描1636 (11.948 min):
衍生化的1,12-十二烷二醇
丰度
m/z-->
147.1
75.1
97.1
125.1
177.1
213.2
241.2
285.1
315.2
345.3
54.1

图11A

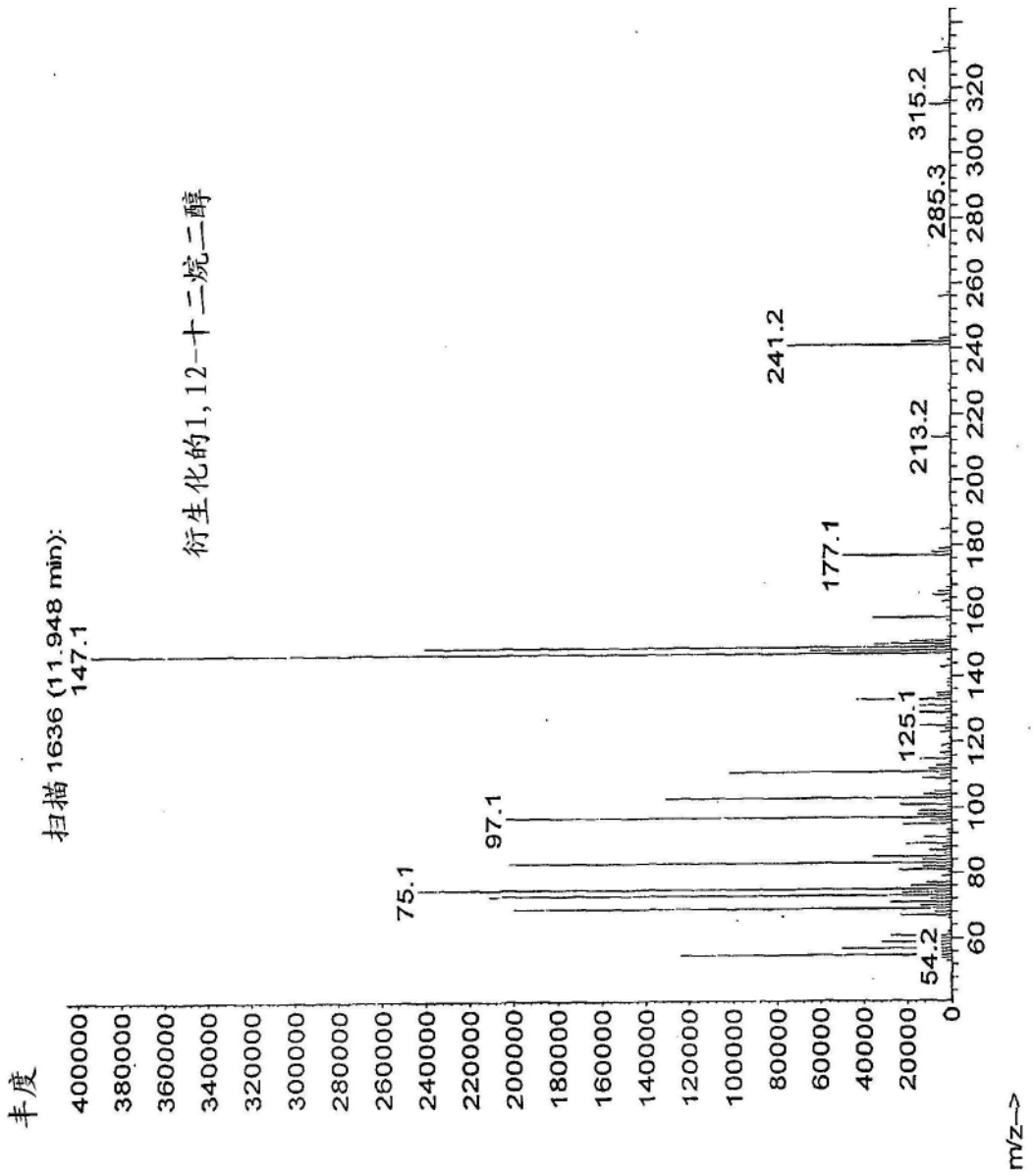
丰度
扫描 1636 (11.948 min):
147.1
衍生化的1,12-十二烷二醇
75.1
97.1
125.1
54.2
177.1
213.2
241.2
285.3
315.2
m/z-->
400000
380000
360000
340000
320000
300000
280000
260000
240000
220000
200000
180000
160000
140000
120000
100000
80000
60000
40000
20000
0
60
80
100
120
140
160
180
200
220
240
260
280
300
320

图11B

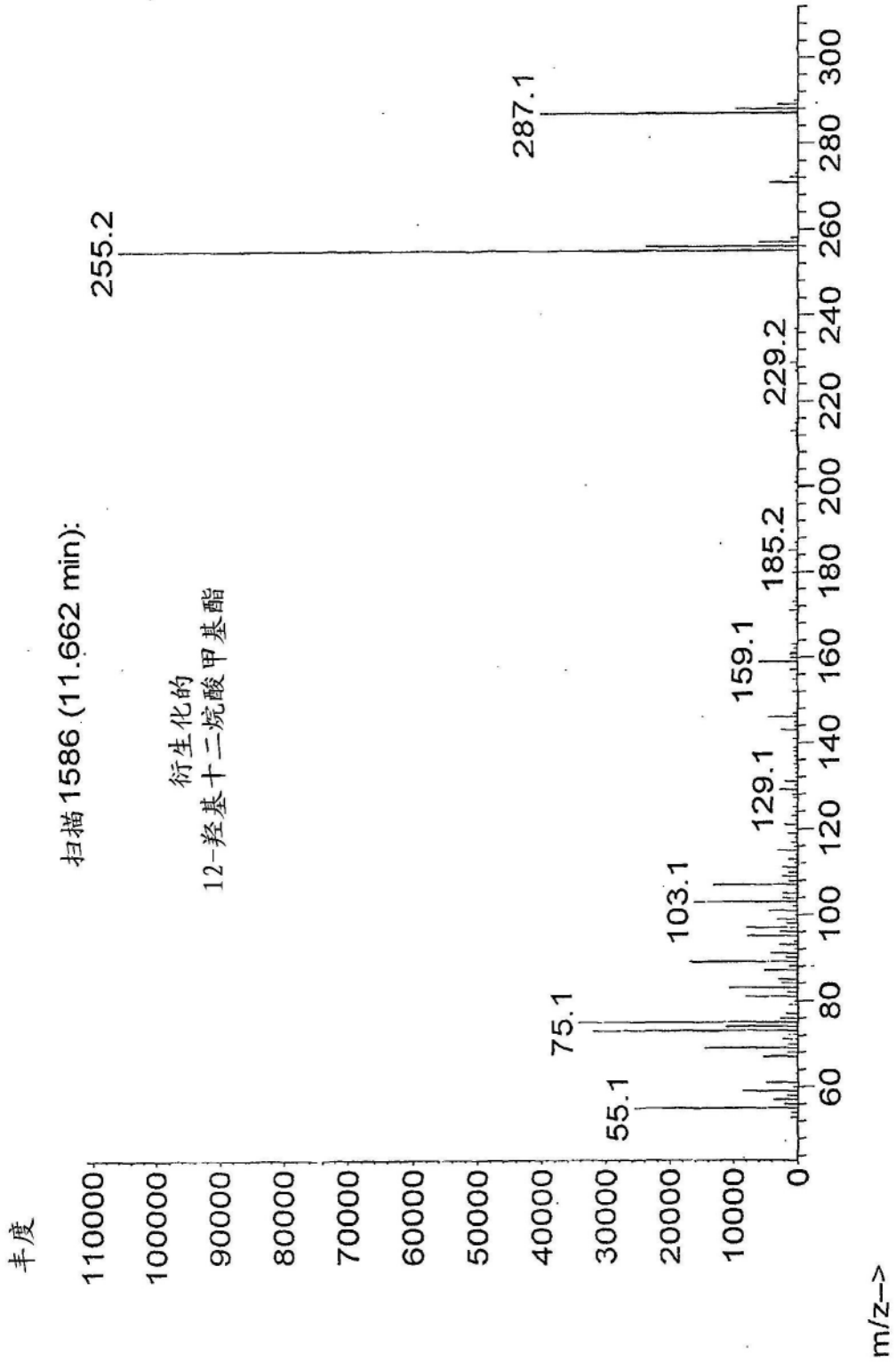
扫描1586 (11.662 min):
衍生化的
12-羟基十二烷酸甲基酯
丰度
m/z-->
255.2
287.1
229.2
185.2
159.1
129.1
103.1
75.1
55.1

图12A

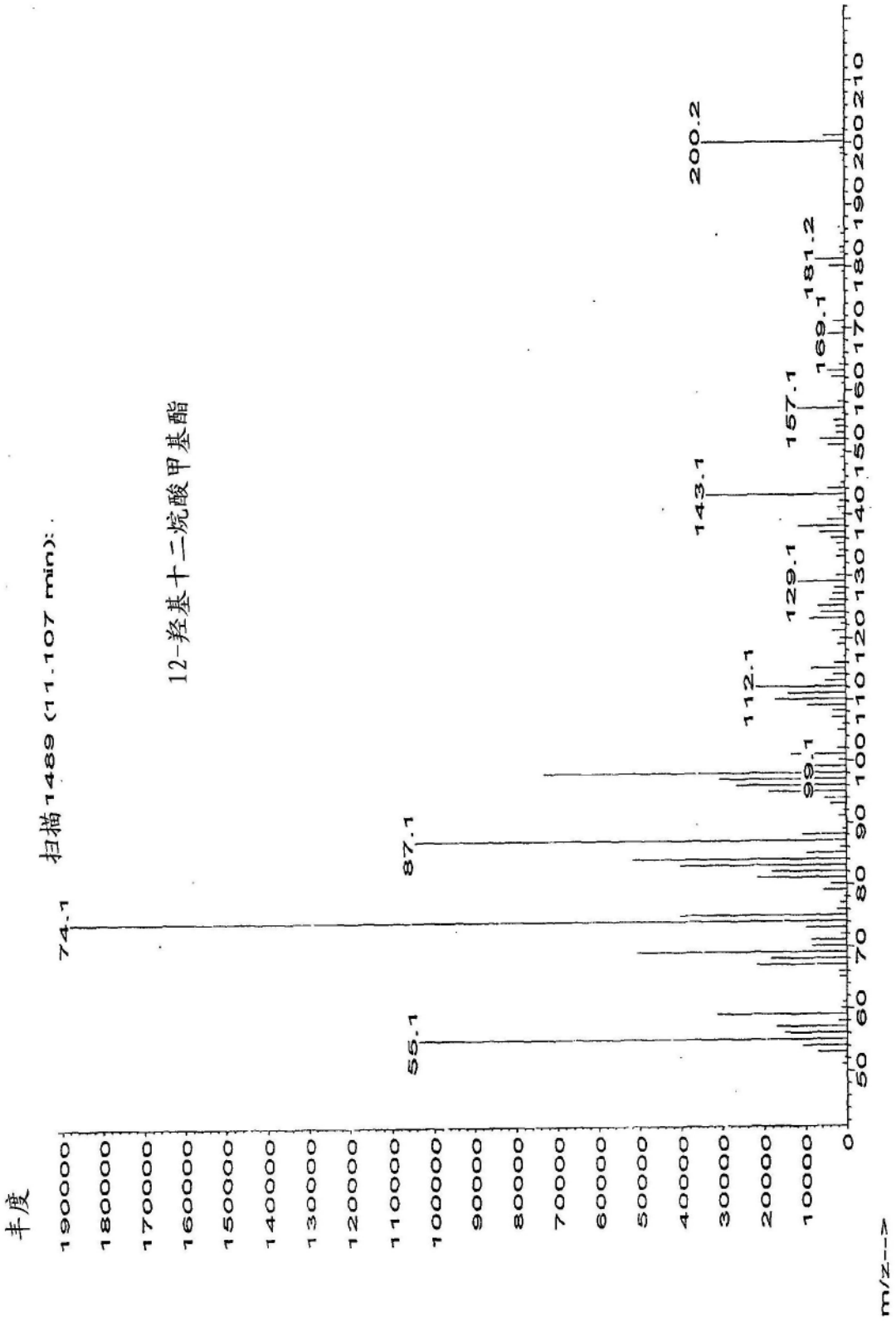

图12B

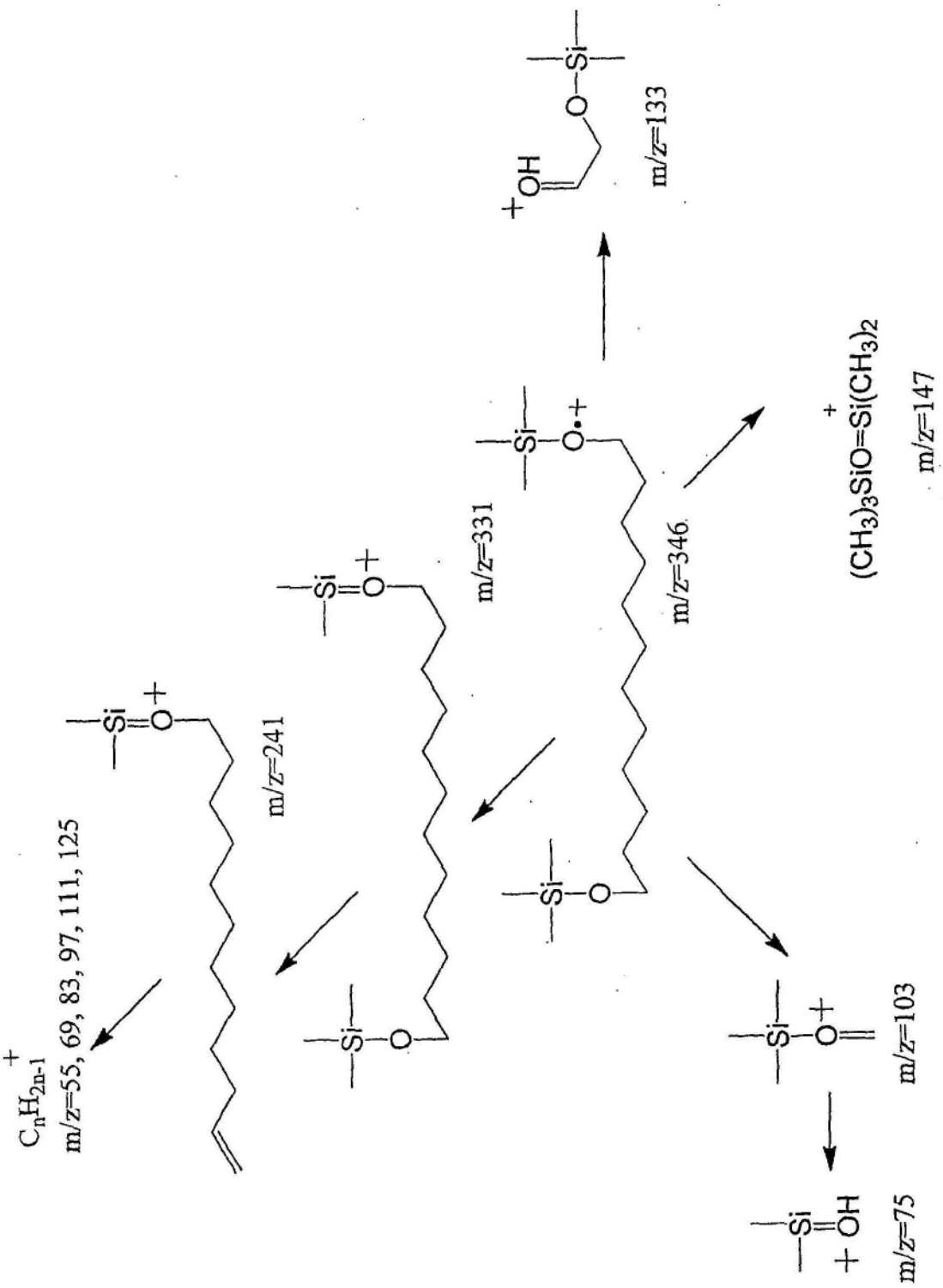
$C_nH_{2n-1}^+$
m/z=55, 69, 83, 97, 111, 125
m/z=241
m/z=331
m/z=346
m/z=133
$(CH_3)_3SiO=Si(CH_3)_2^+$
m/z=147
m/z=103
m/z=75

图13A

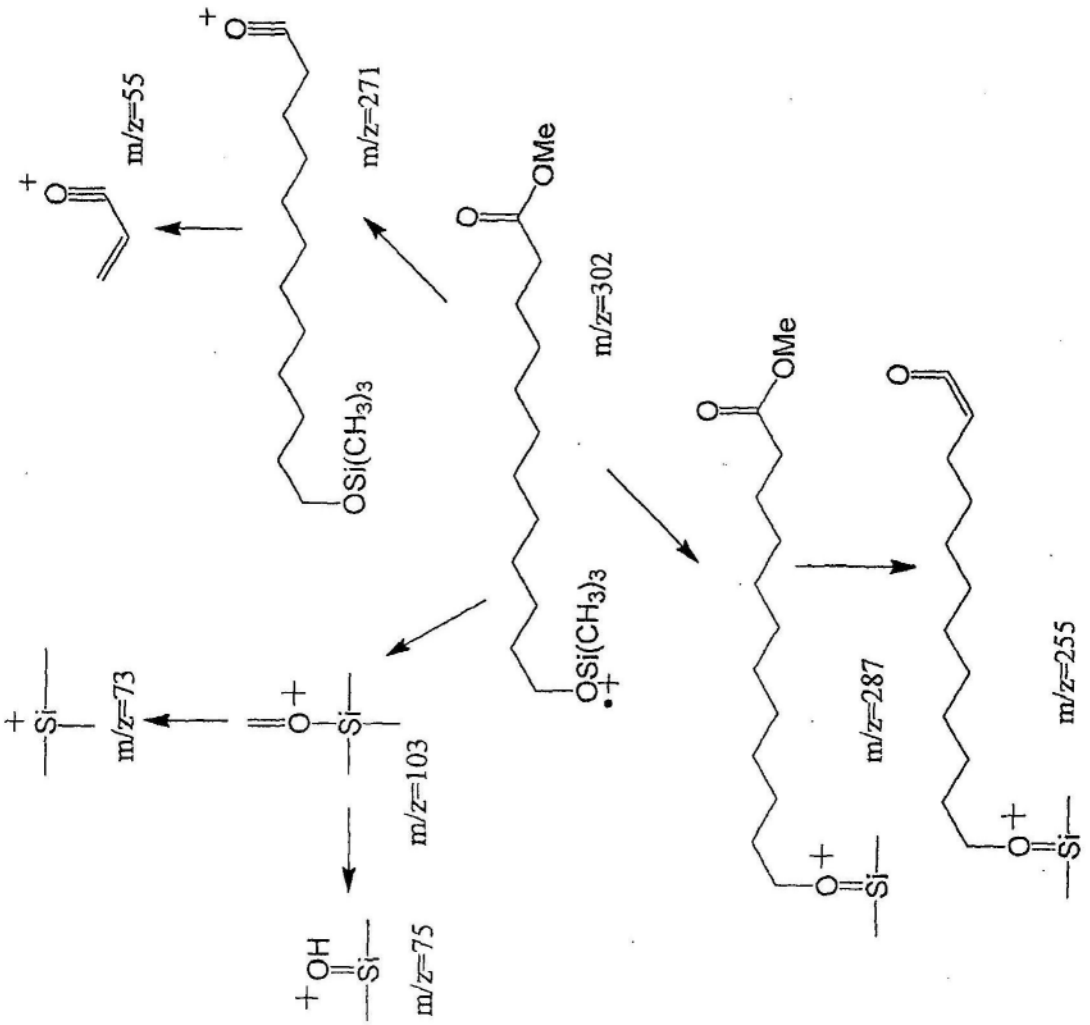
m/z=55
m/z=271
m/z=302
m/z=73
m/z=103
m/z=75
m/z=287
m/z=255
OMe
OSi(CH3)3
Si
OH

图13B

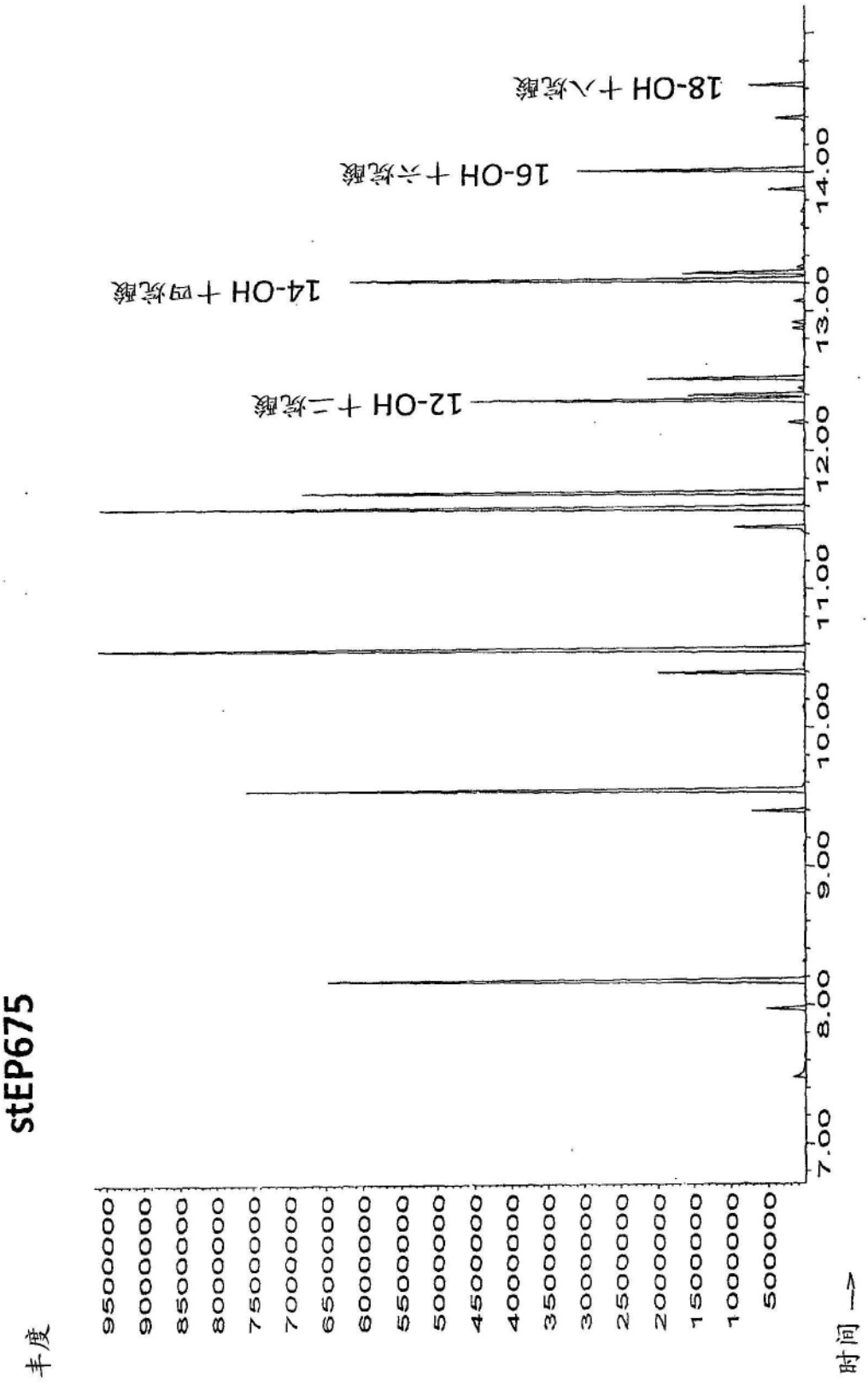

图14A

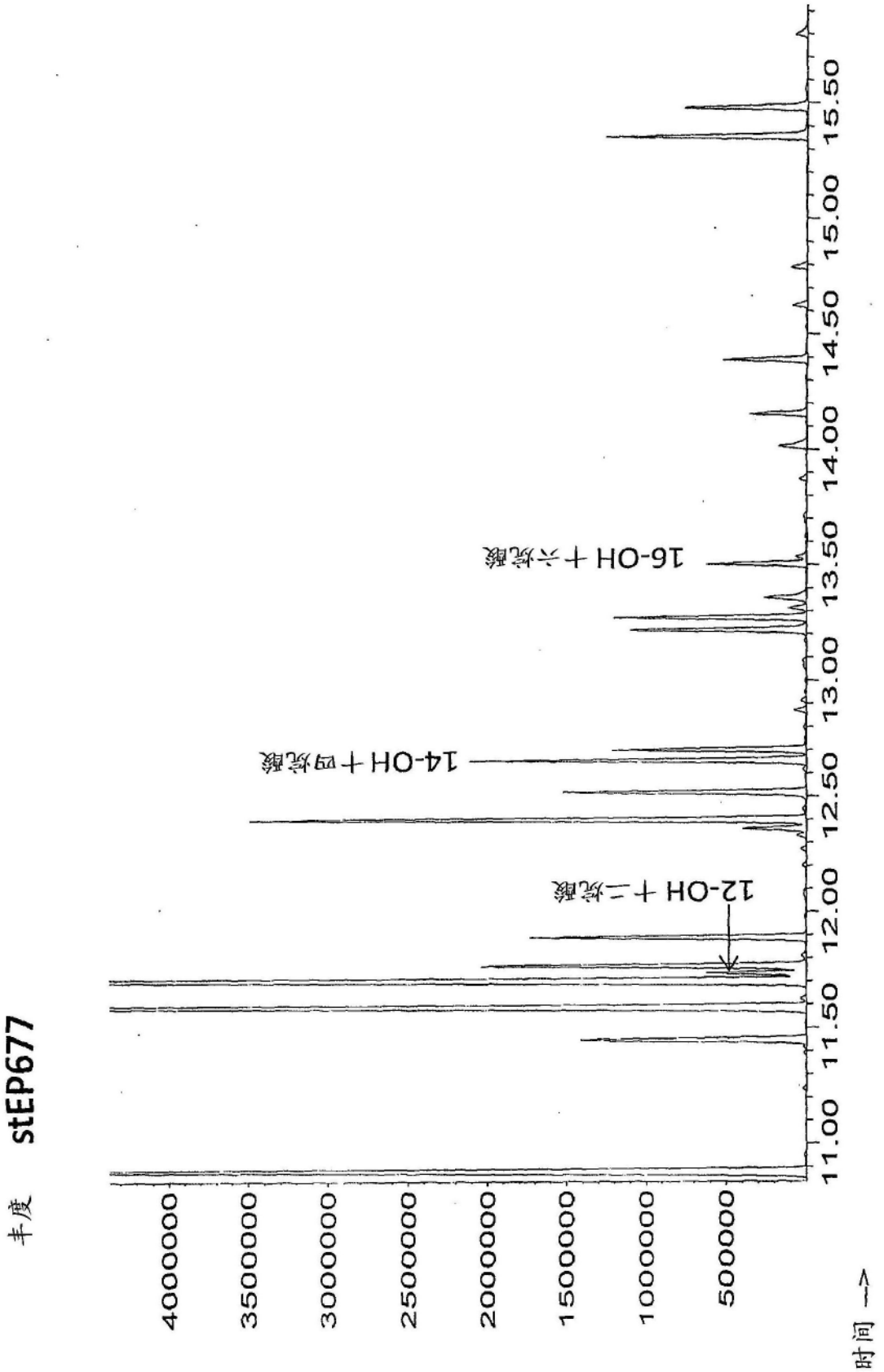
stEP677
丰度
时间 —>
4000000
3500000
3000000
2500000
2000000
1500000
1000000
500000
11.00
11.50
12.00
12.50
13.00
13.50
14.00
14.50
15.00
15.50
12-OH十二烷酸
14-OH十四烷酸
16-OH十六烷酸

图14B

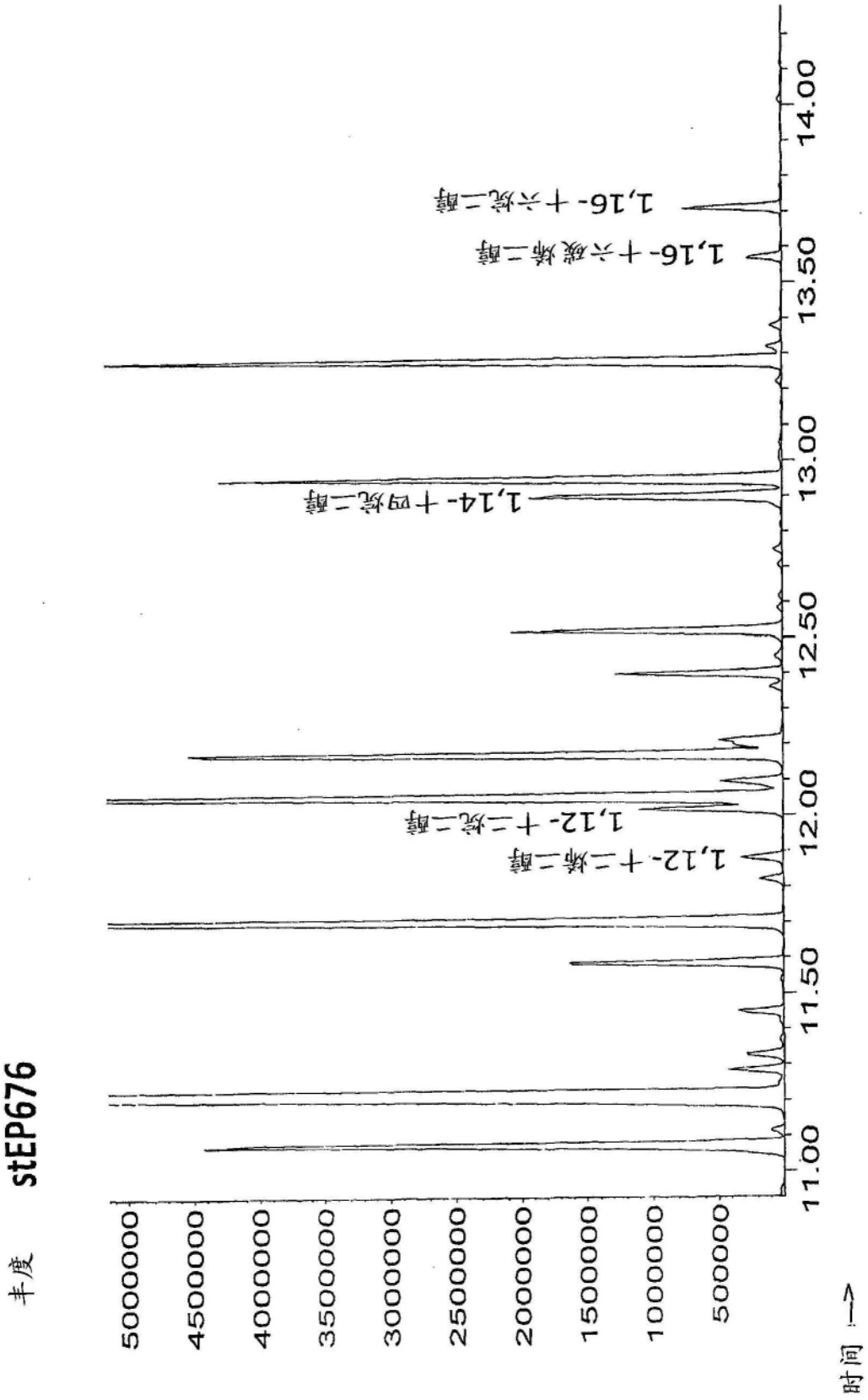
丰度
stEP676
5000000
4500000
4000000
3500000
3000000
2500000
2000000
1500000
1000000
500000
11.00
11.50
12.00
12.50
13.00
13.50
14.00
时间
1,12-十二烯二醇
1,12-十二烷二醇
1,14-十四烷二醇
1,16-十六碳烯二醇
1,16-十六烷二醇

图14C

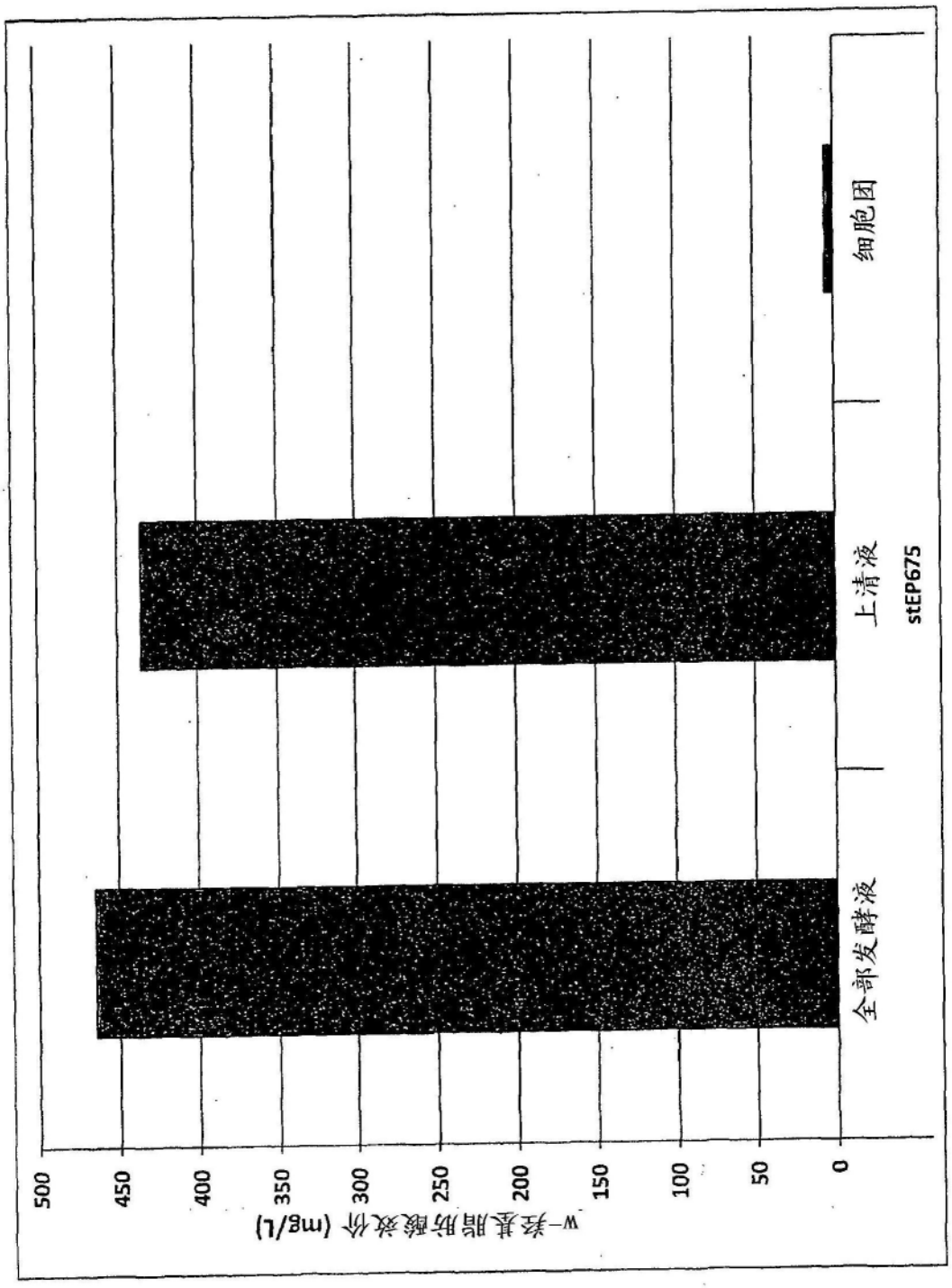
500
450
400
350
300
250
200
150
100
50
0
ω-羟基脂肪酸效价 (mg/L)
全部发酵液
上清液
细胞团
stEP675

图15

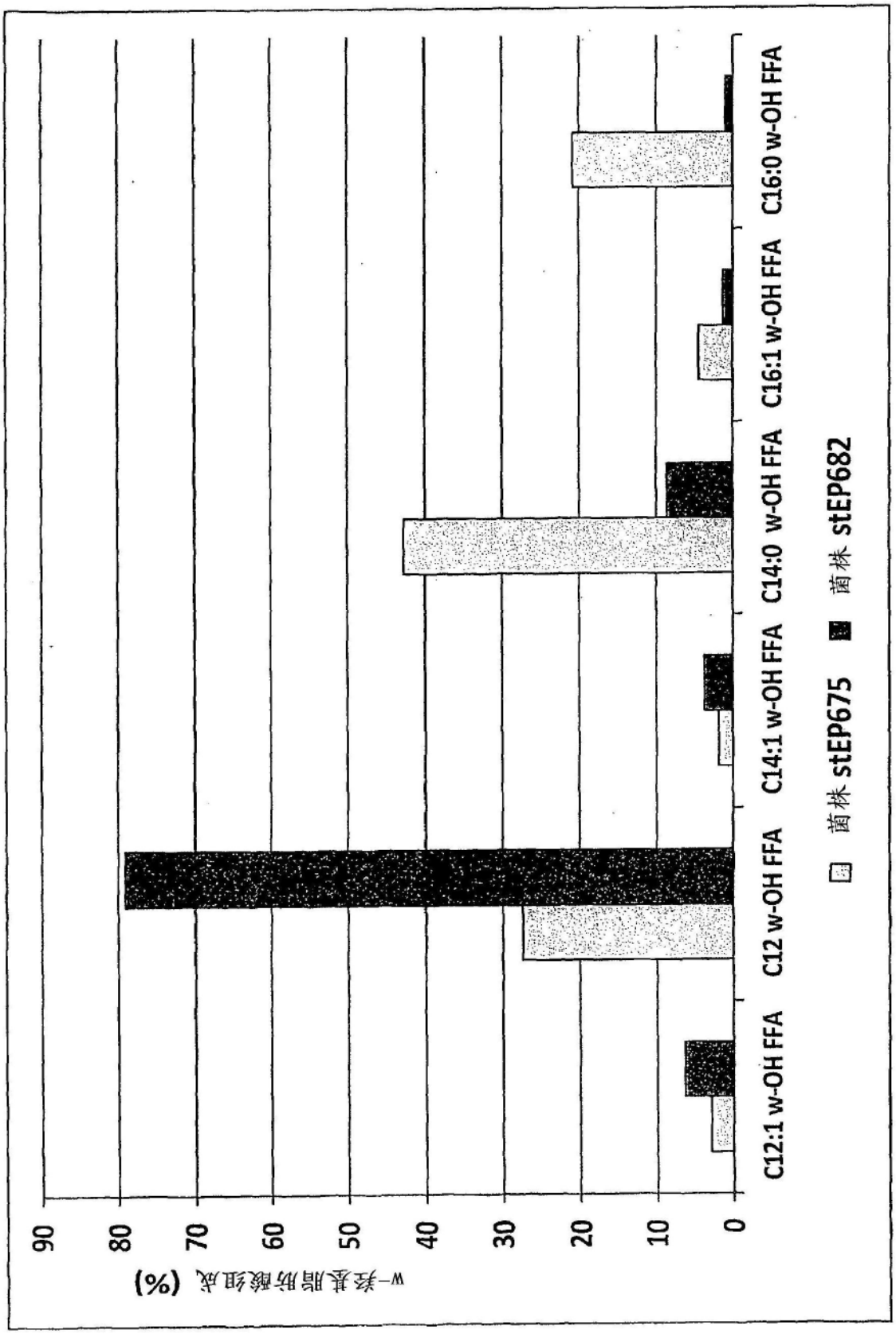
ω-羟基脂肪酸组成 (%)
0
10
20
30
40
50
60
70
80
90
C12:1 ω-OH FFA
C12 ω-OH FFA
C14:1 ω-OH FFA
C14:0 ω-OH FFA
C16:1 ω-OH FFA
C16:0 ω-OH FFA
菌株 stEP675
菌株 stEP682

图16

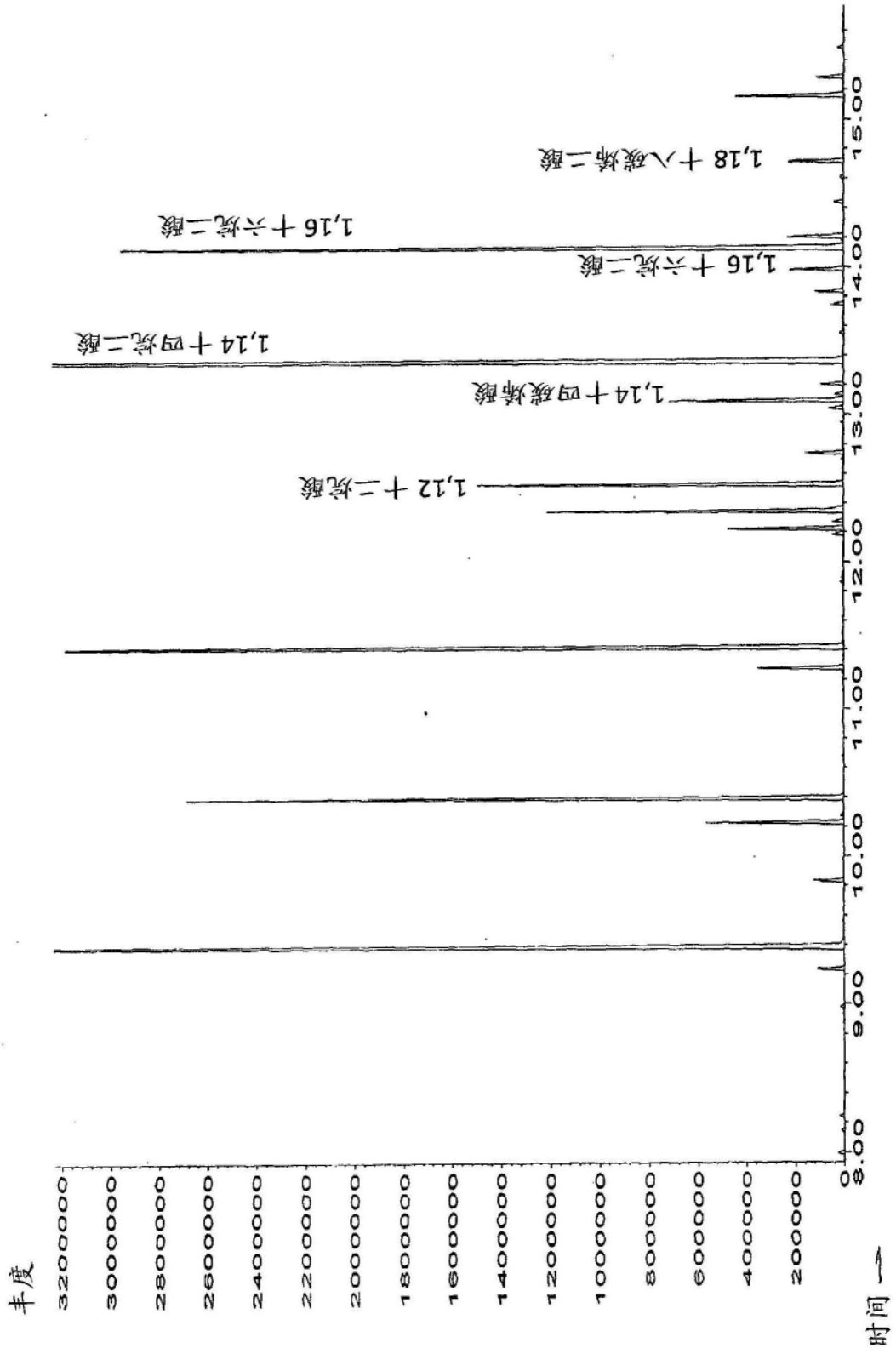
丰度
32000000
30000000
28000000
26000000
24000000
22000000
20000000
18000000
16000000
14000000
12000000
10000000
8000000
6000000
4000000
2000000
0
8.00
9.00
10.00
11.00
12.00
13.00
14.00
15.00
时间
1,12＋二烷酸
1,14＋四碳烯酸
1,14＋四烷二酸
1,16＋六烷二酸
1,16＋六烷二酸
1,18＋八碳烯二酸

图17

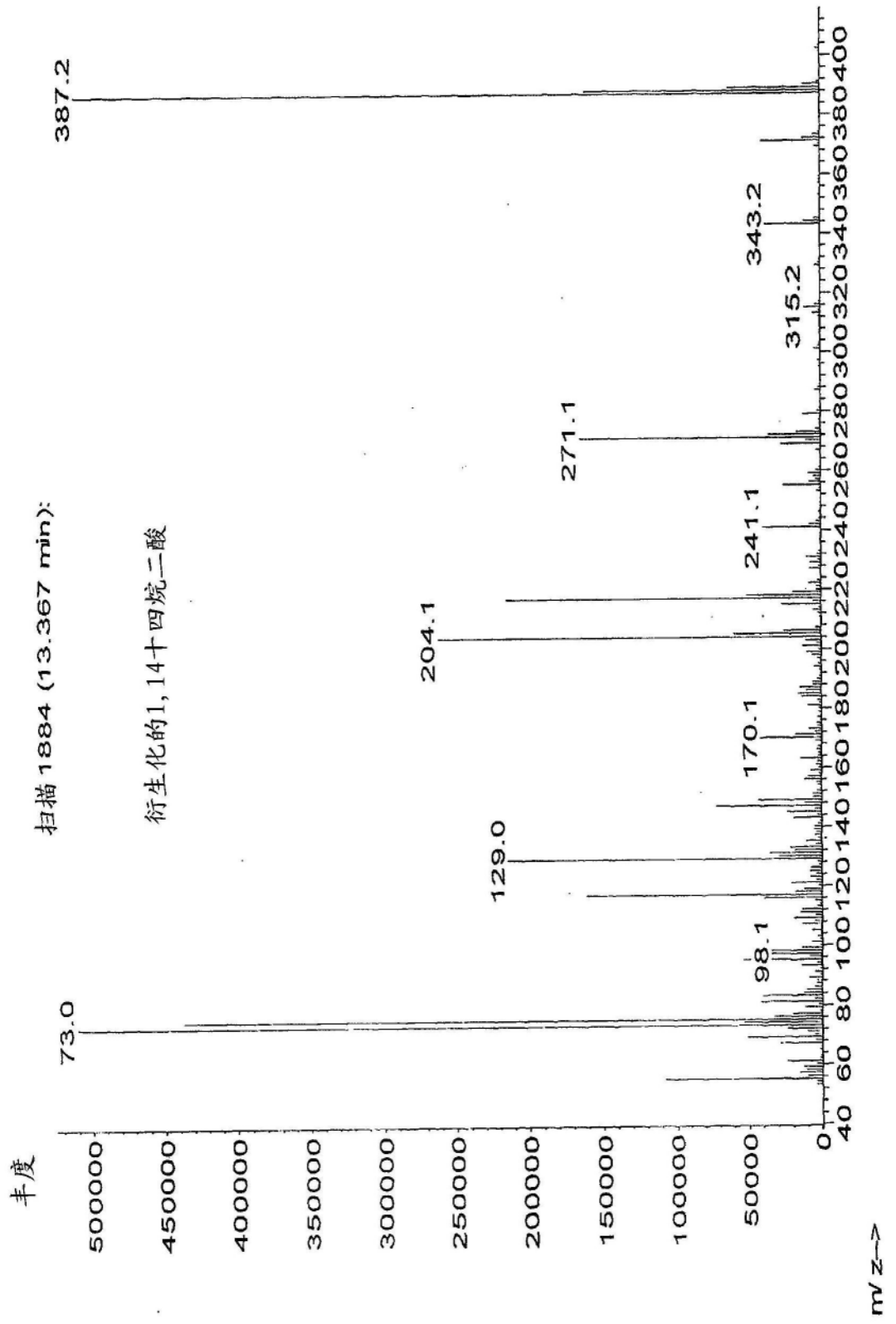
扫描 1884 (13.367 min):
衍生化的1,14十四烷二酸
丰度
m/z-->
73.0
98.1
129.0
170.1
204.1
241.1
271.1
315.2
343.2
387.2

图18A

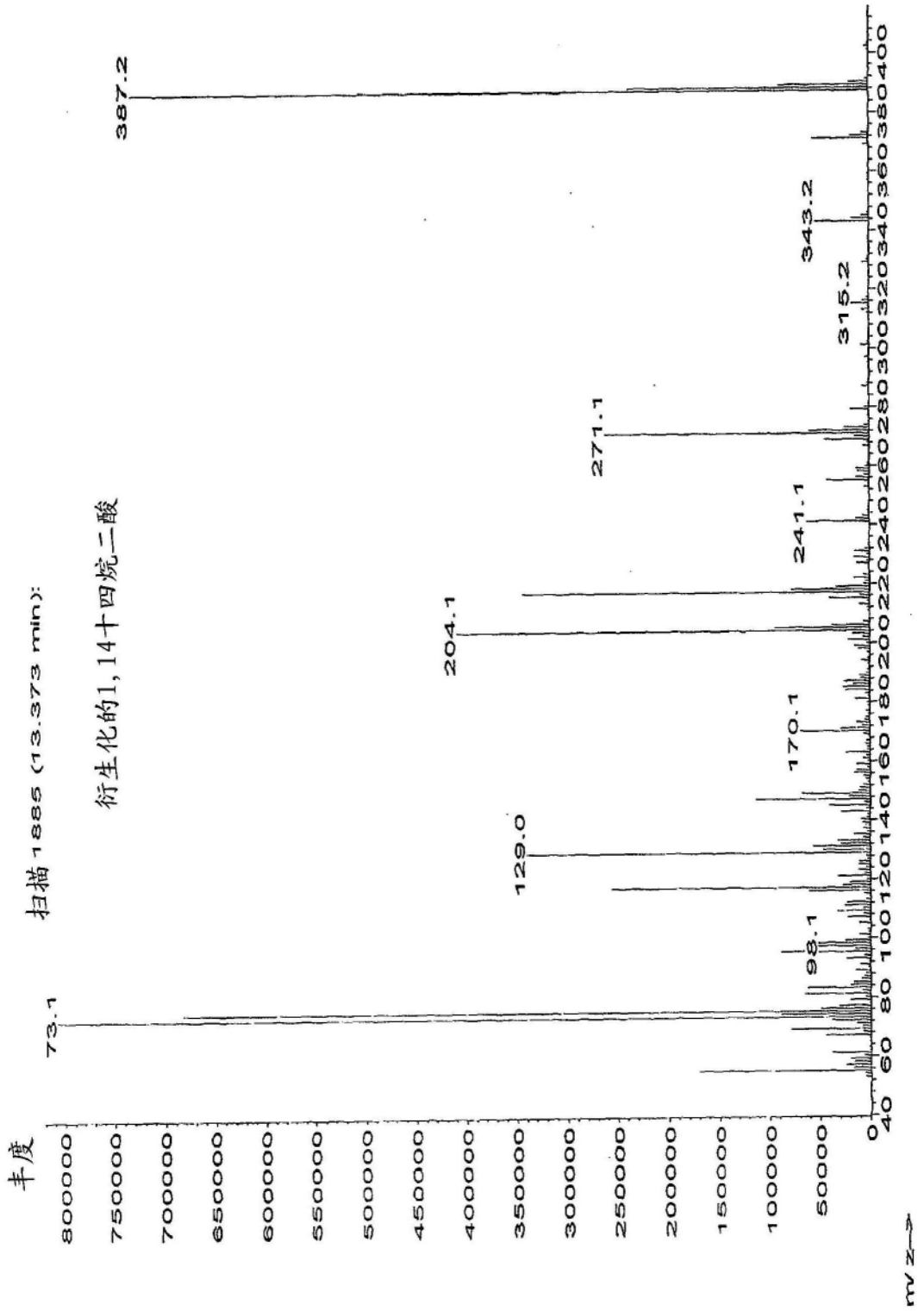
衍生化的1,14十四烷二酸
扫描 1885 (13.373 min):
丰度
m/z
73.1
98.1
129.0
170.1
204.1
241.1
271.1
315.2
343.2
387.2

图18B

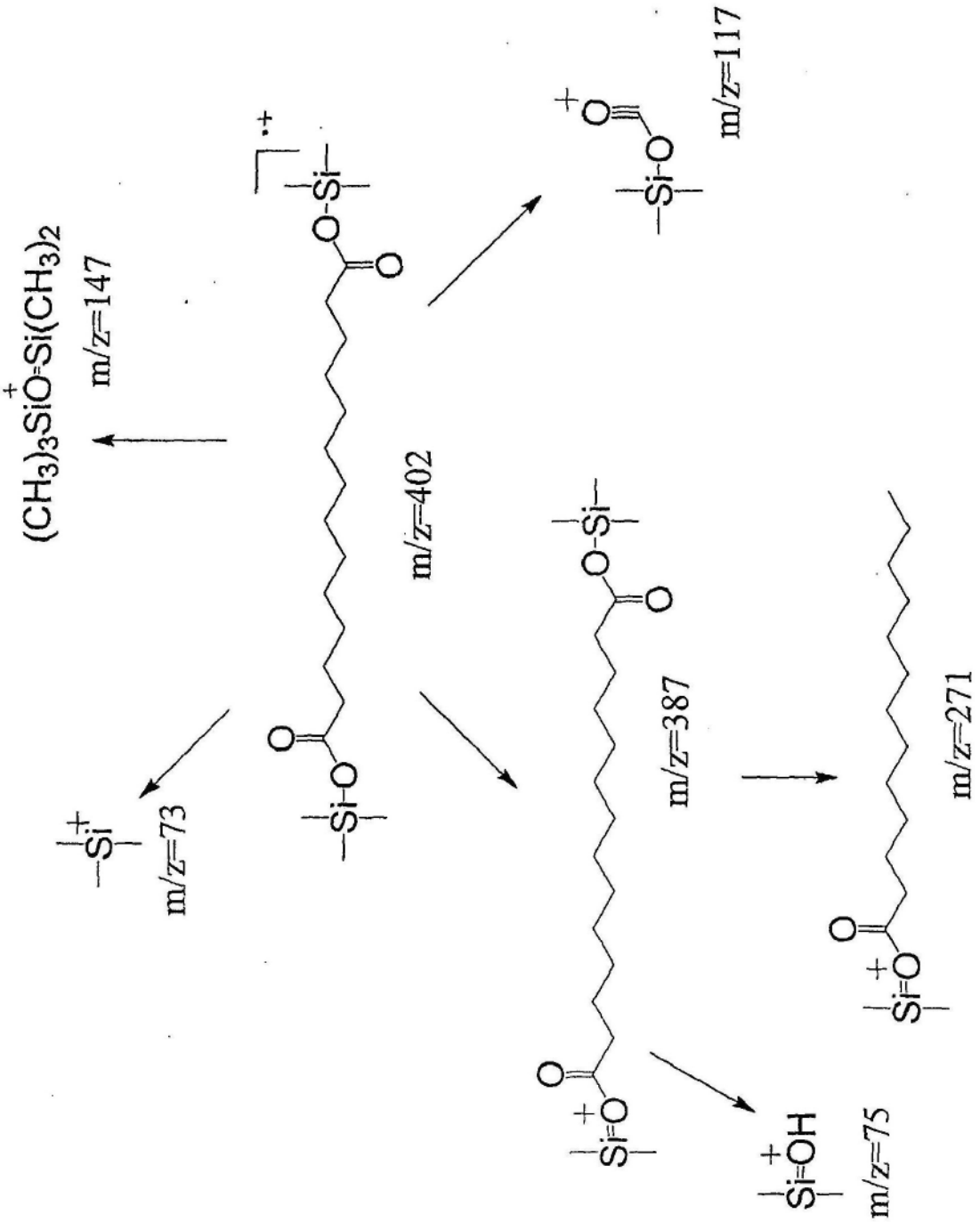

图19

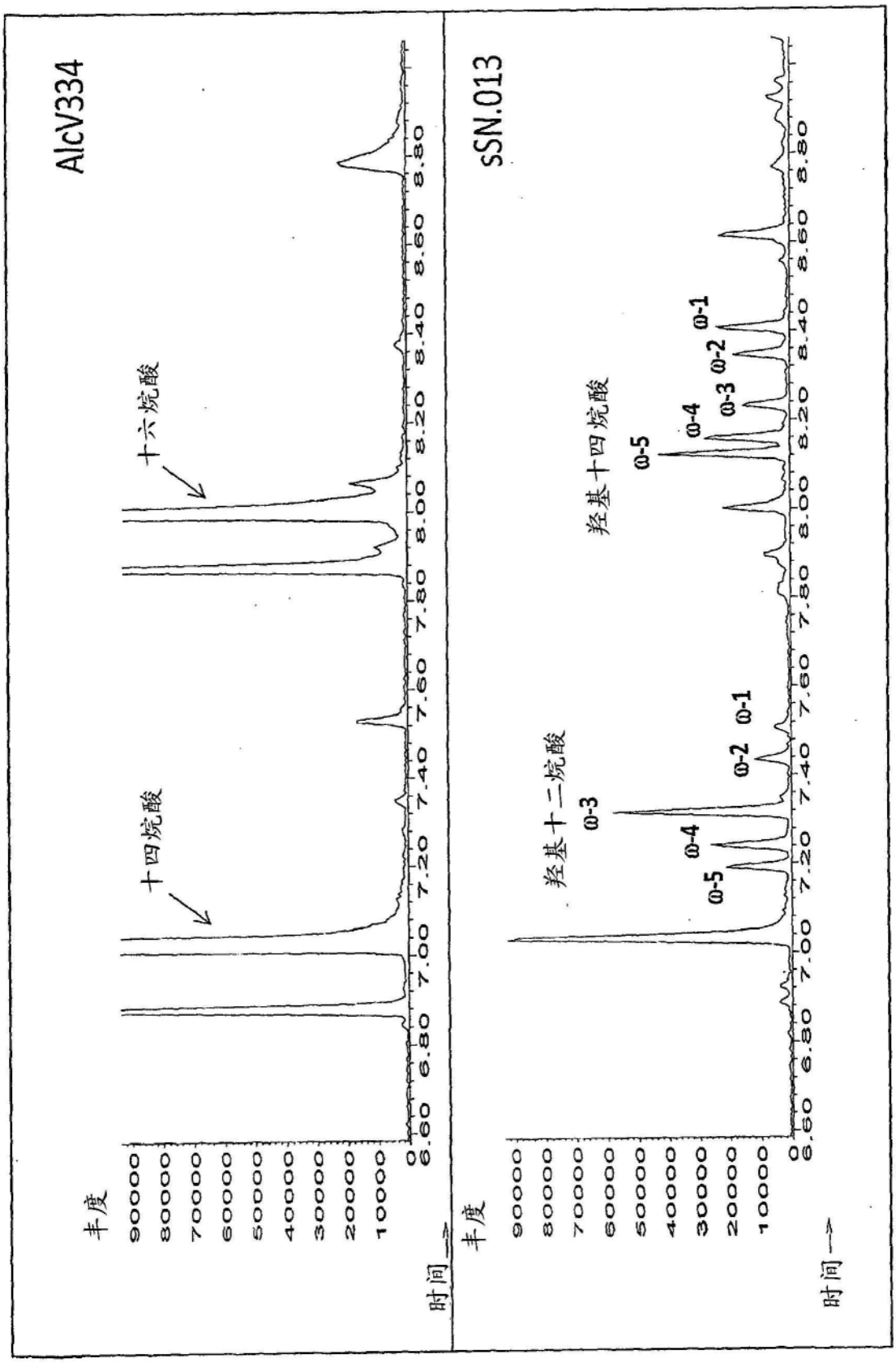

图20

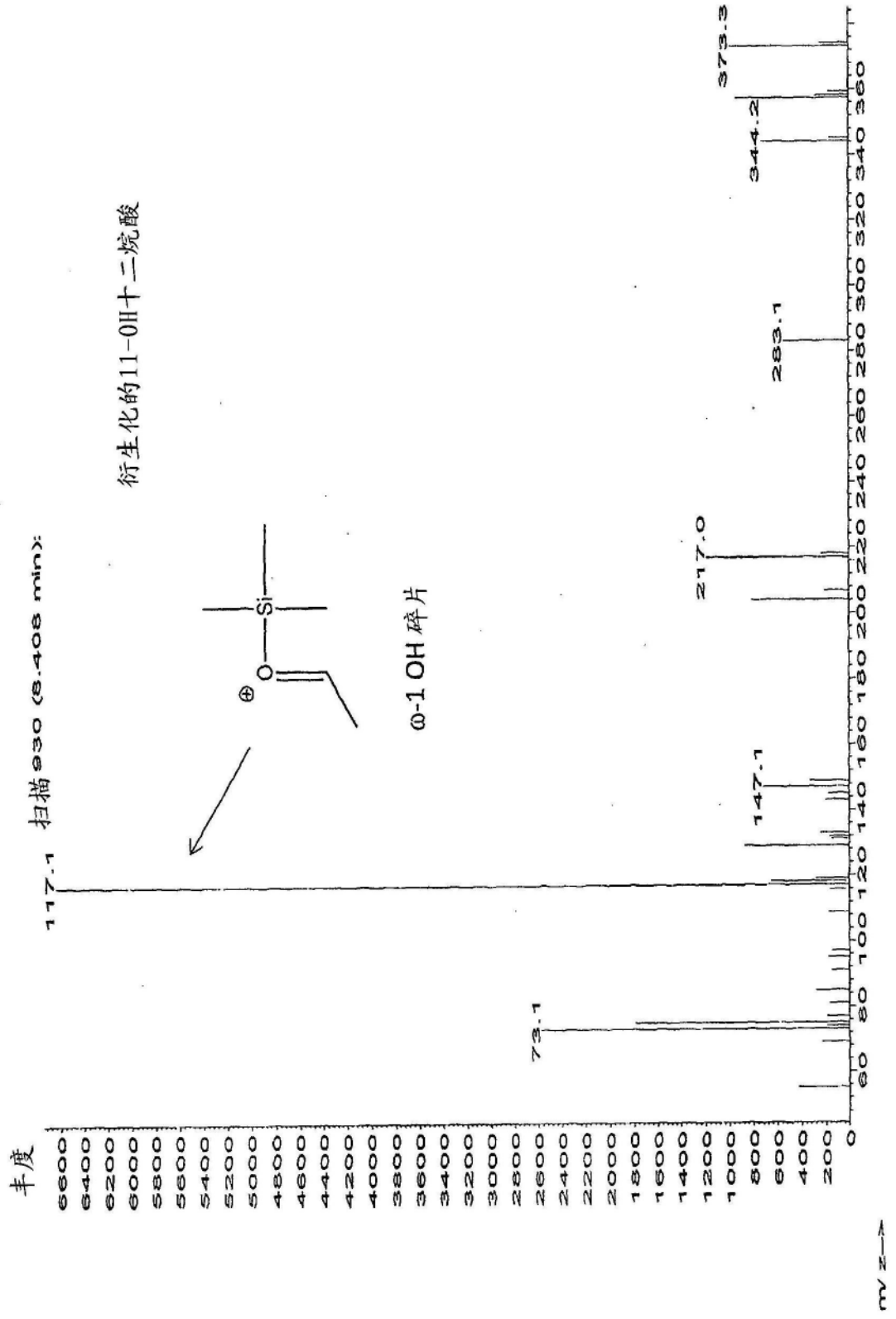

图21A

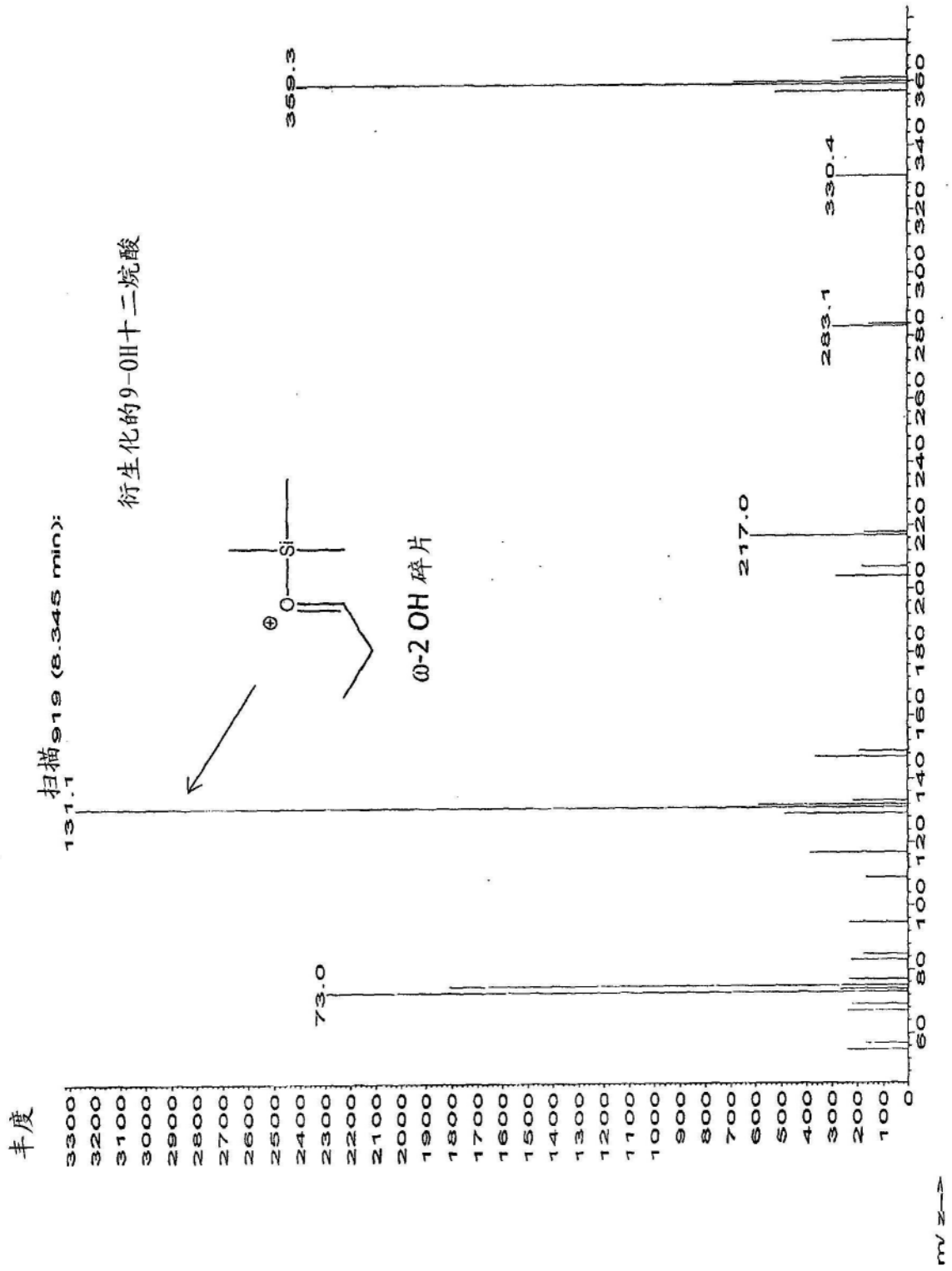
衍生化的9-OH十二烷酸
扫描919 (8.345 min):
131.1
73.0
217.0
283.1
330.4
359.3
ω-2 OH 碎片
丰度
m/z

图21B

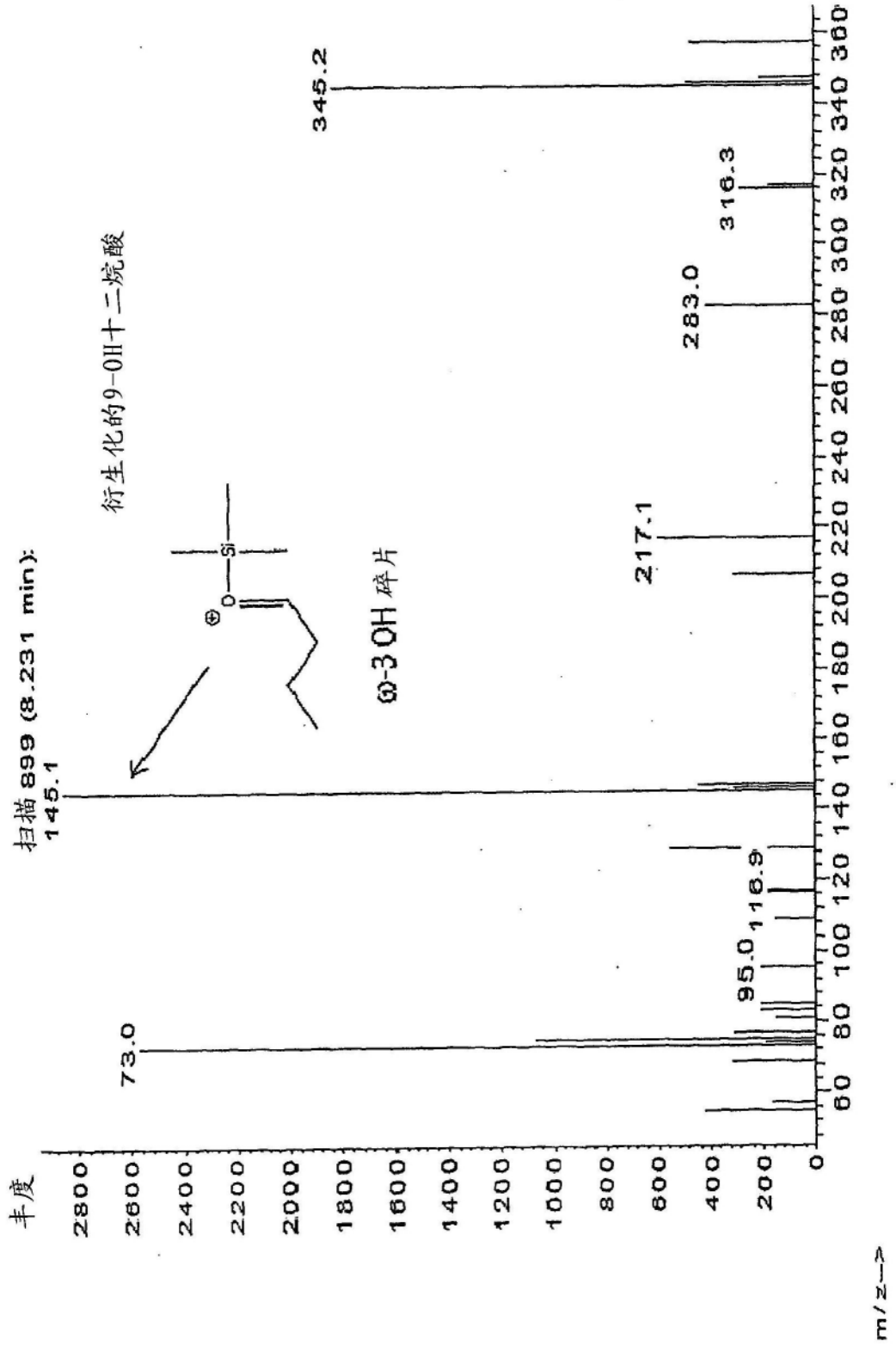
扫描 899 (8.231 min):
衍生化的9-OH十二烷酸
145.1
73.0
95.0
116.9
217.1
283.0
316.3
345.2
ω-3 OH 碎片
丰度
m/z-->

图21C

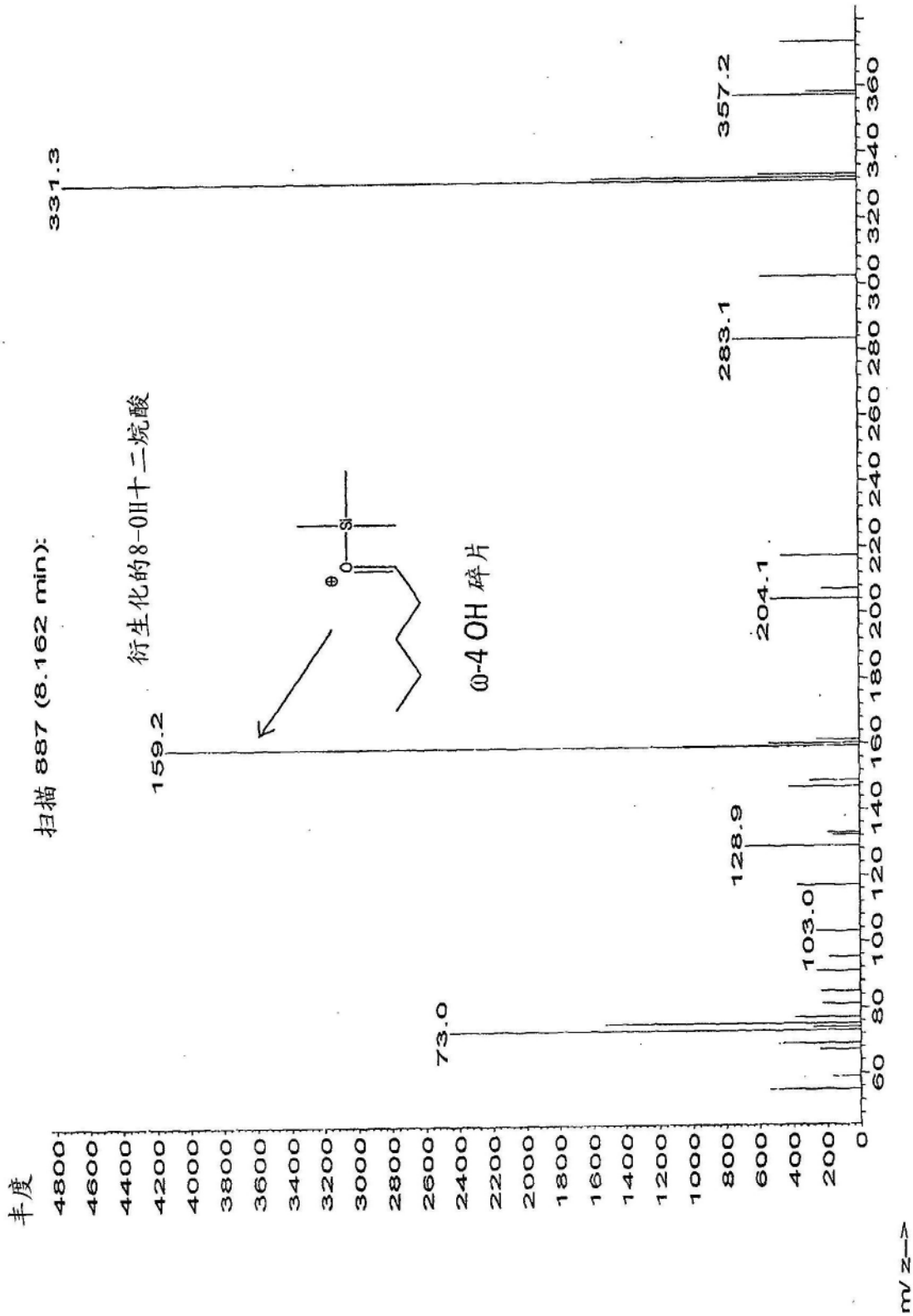
扫描 887 (8.162 min):
衍生化的8-OH十二烷酸
ω-4 OH 碎片
丰度
m/z-->
331.3
357.2
283.1
204.1
159.2
128.9
103.0
73.0

图21D

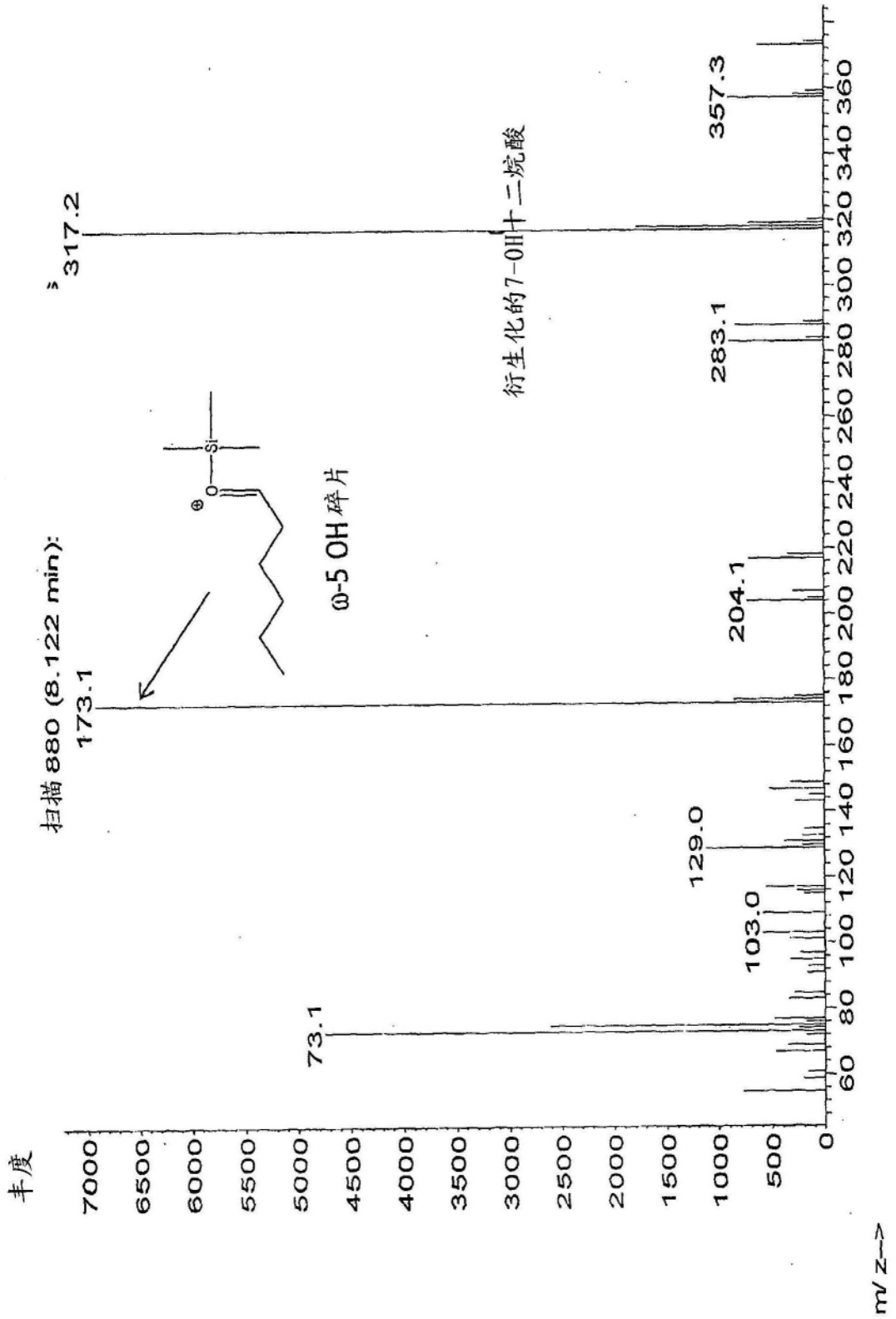
扫描 880 (8.122 min):
173.1
317.2
357.3
283.1
204.1
129.0
103.0
73.1
ω-5 OH 碎片
衍生化的7-OH十二烷酸
丰度
m/z-->

图21E

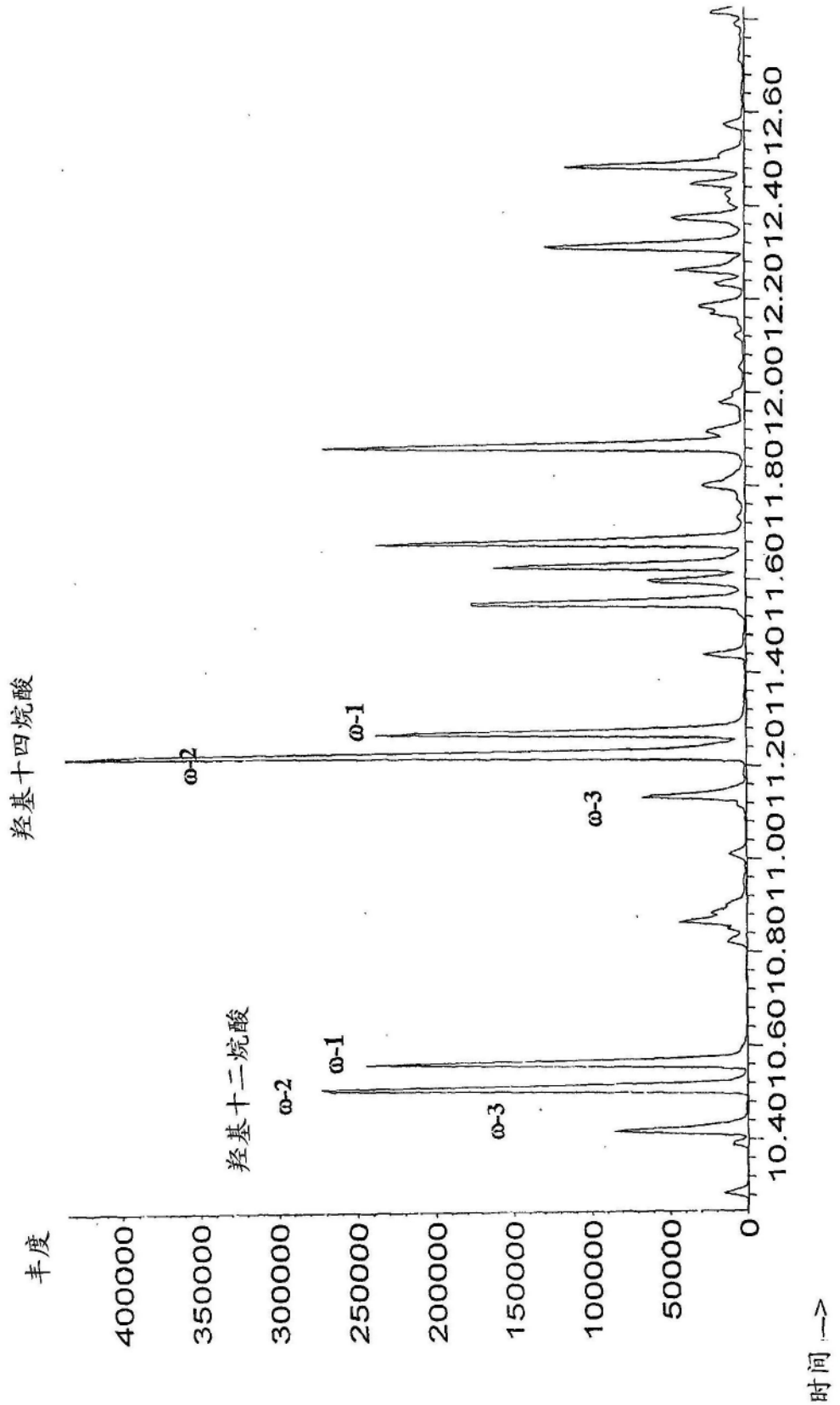
羟基十四烷酸
羟基十二烷酸
ω-1
ω-2
ω-3
丰度
时间 —>
400000
350000
300000
250000
200000
150000
100000
50000
0
10.40
10.60
10.80
11.00
11.20
11.40
11.60
11.80
12.00
12.20
12.40
12.60

图22A

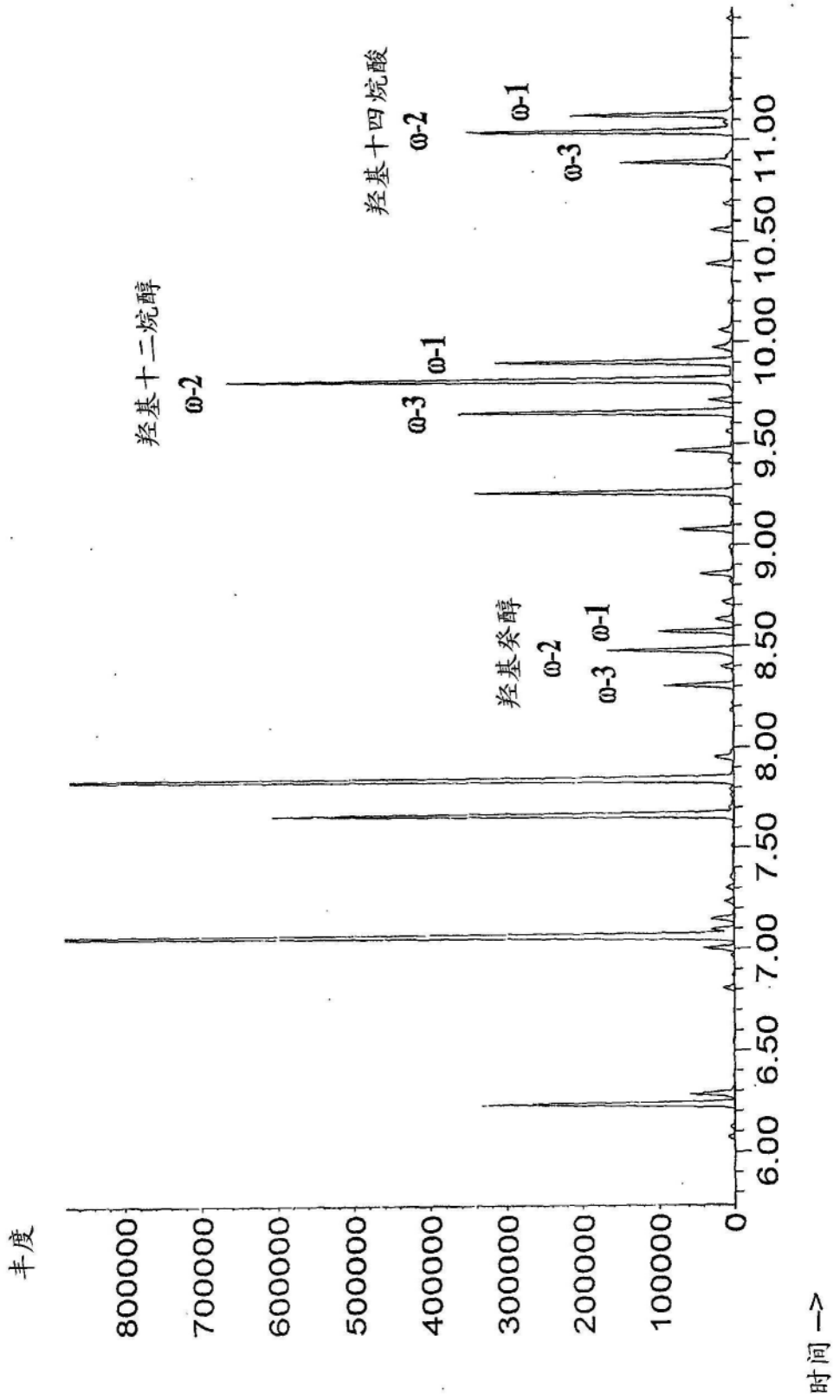
丰度
800000
700000
600000
500000
400000
300000
200000
100000
0
6.00
6.50
7.00
7.50
8.00
8.50
9.00
9.50
10.00
10.50
11.00
时间 —>
羟基十二烷醇
ω-2
羟基十四烷酸
ω-2
ω-3
ω-1
ω-1
ω-3
羟基癸醇
ω-2
ω-3
ω-1

图22B

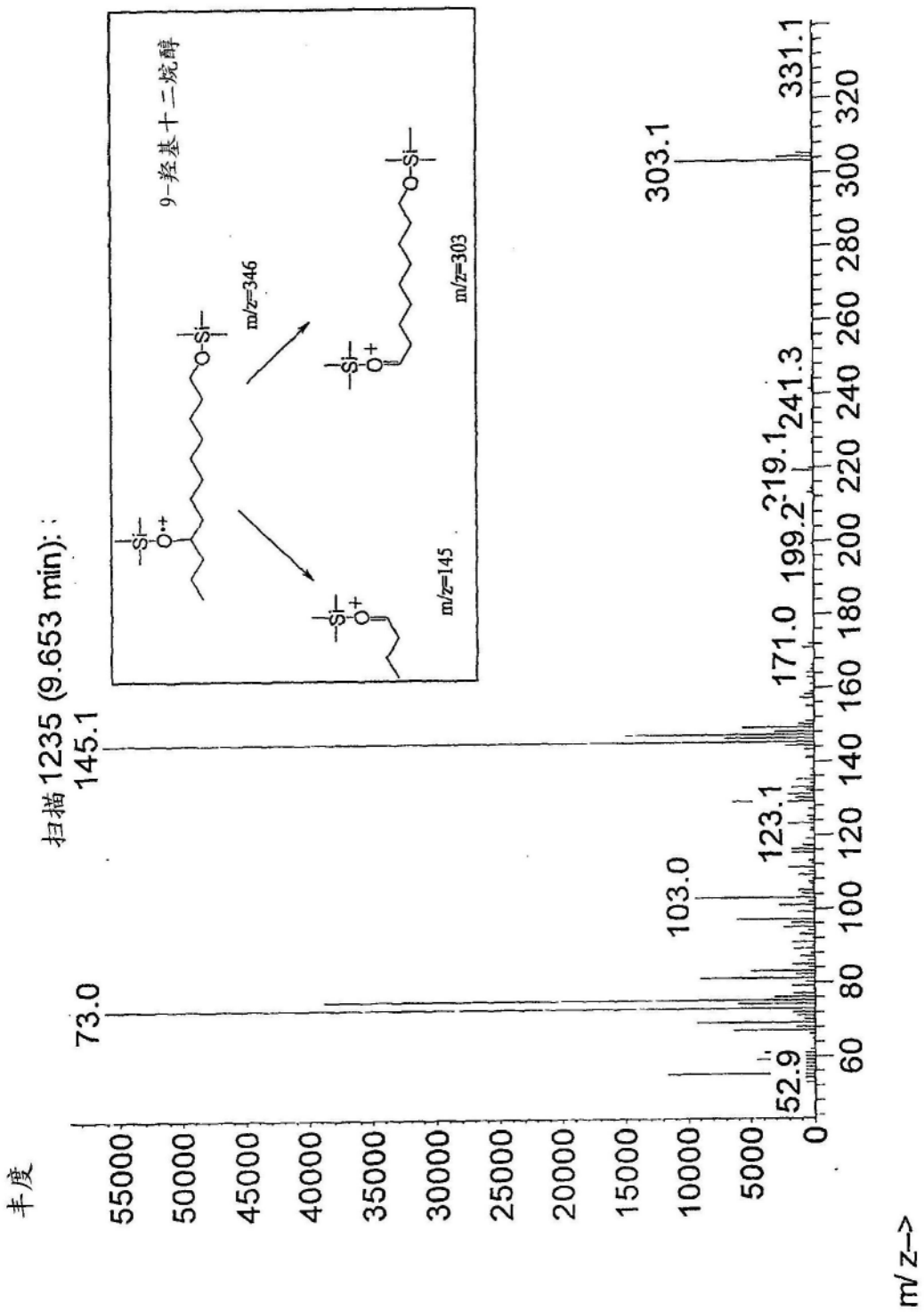
扫描1235 (9.653 min):
丰度
m/z-->
9-羟基十二烷醇
m/z=346
m/z=303
m/z=145
73.0
145.1
103.0
123.1
171.0
199.2
219.1
241.3
303.1
331.1
52.9

图23A

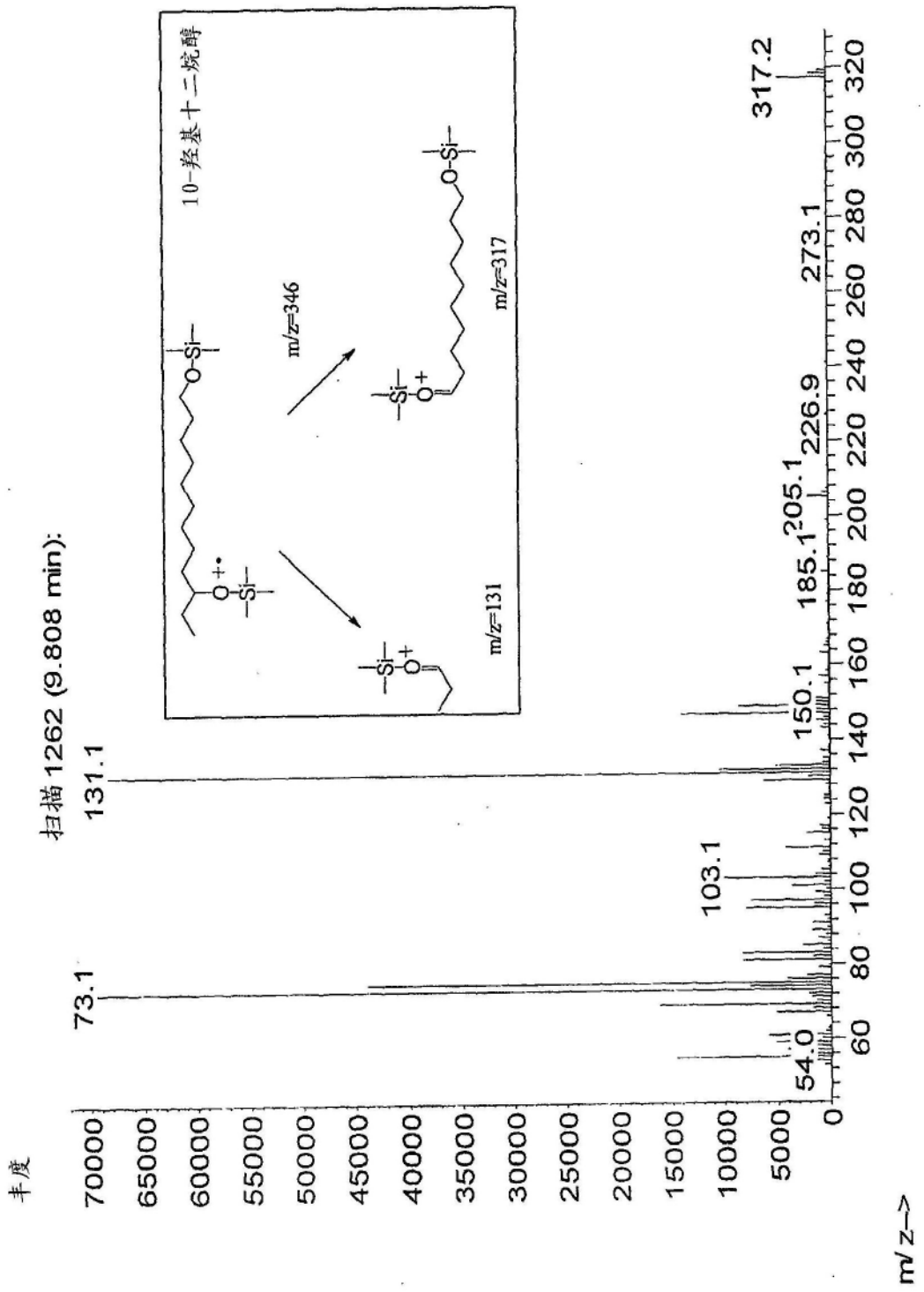
扫描 1262 (9.808 min):
丰度
m/z-->
10-羟基十二烷醇
m/z=346
m/z=131
m/z=317
73.1
131.1
103.1
150.1
185.1
205.1
226.9
273.1
317.2
54.0

图23B

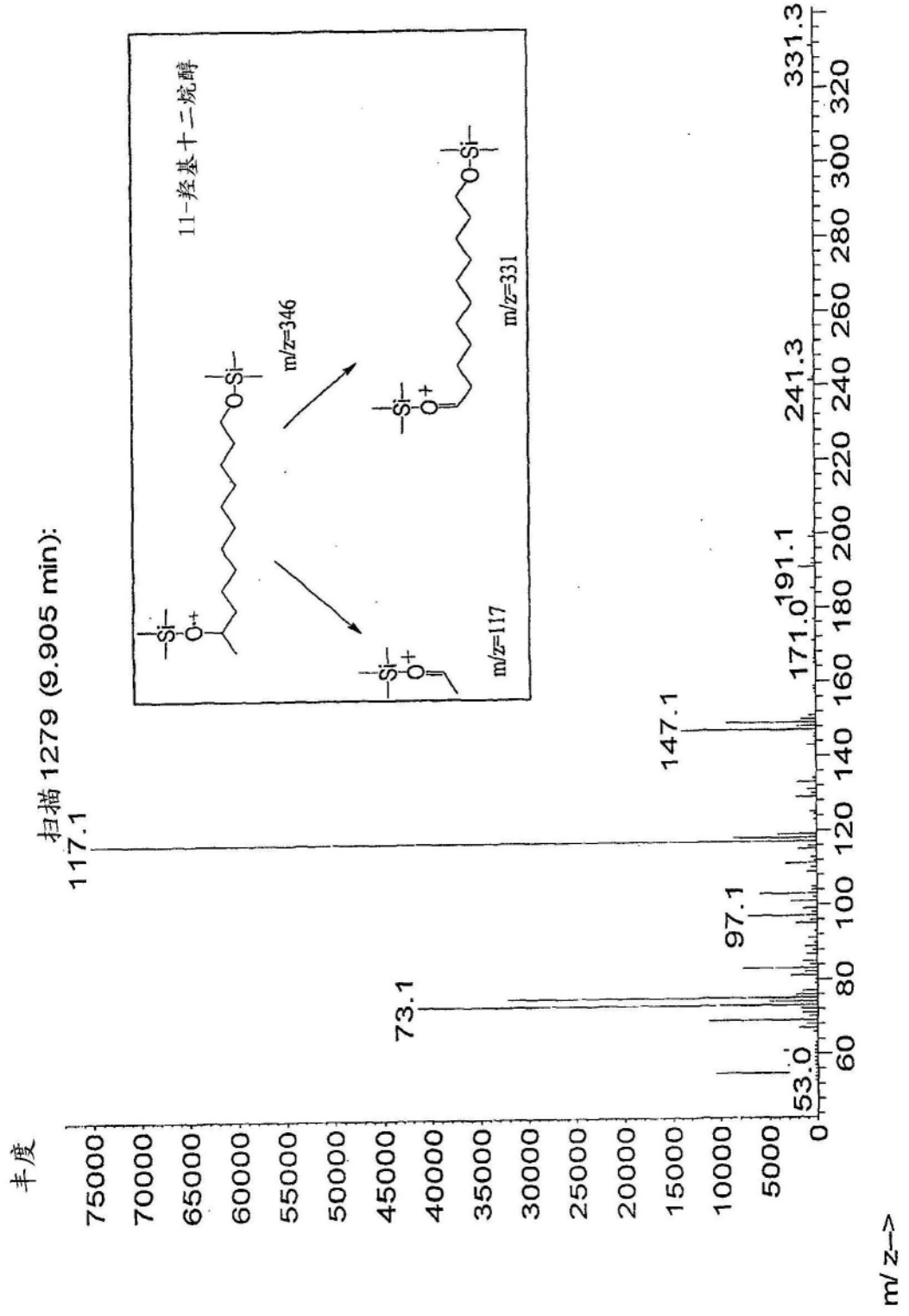
扫描 1279 (9.905 min):
丰度
m/z-->
11-羟基十二烷醇
m/z=346
m/z=331
m/z=117
117.1
147.1
73.1
97.1
53.0
171.0
191.1
241.3
331.3

图23C

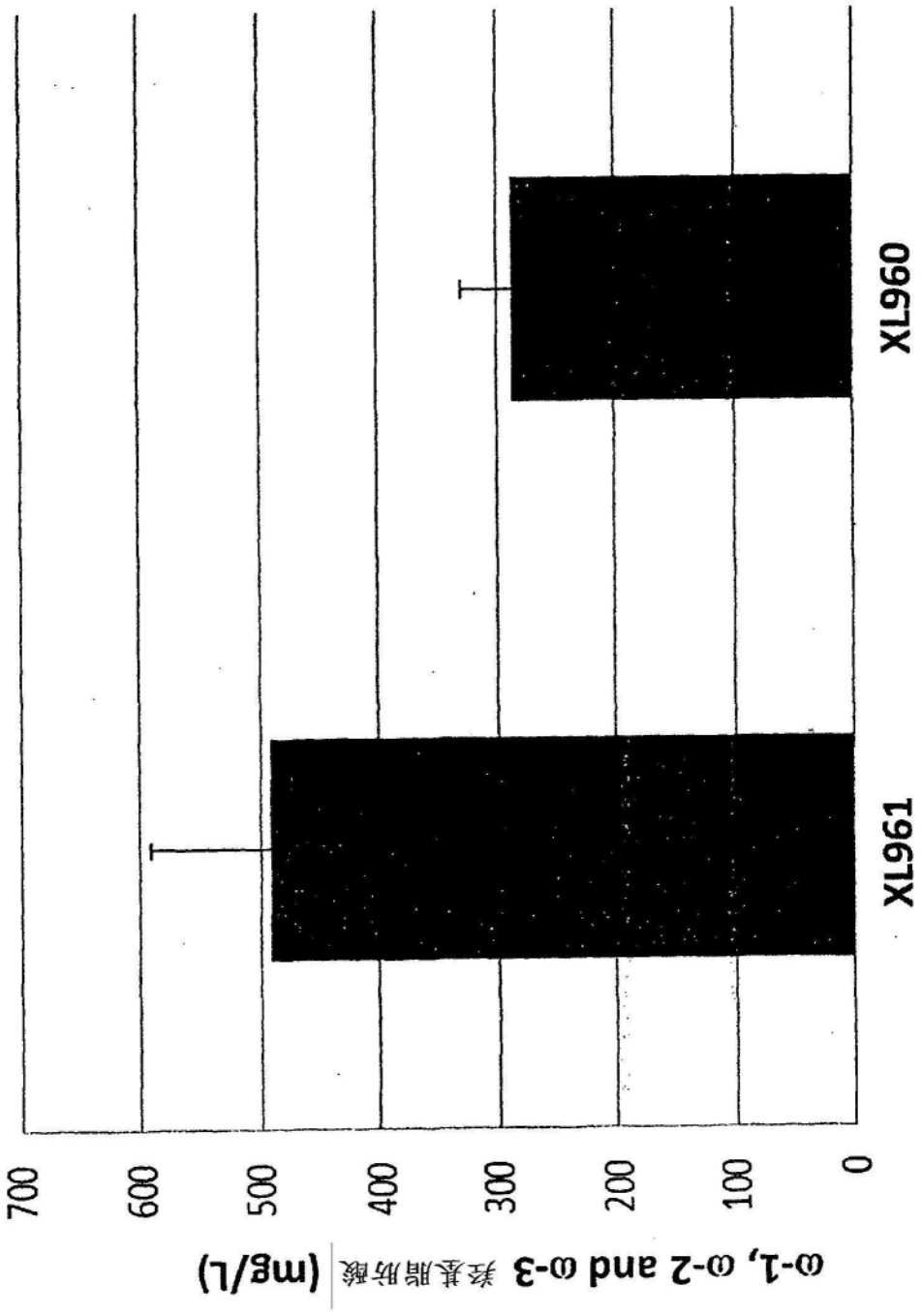

图24A

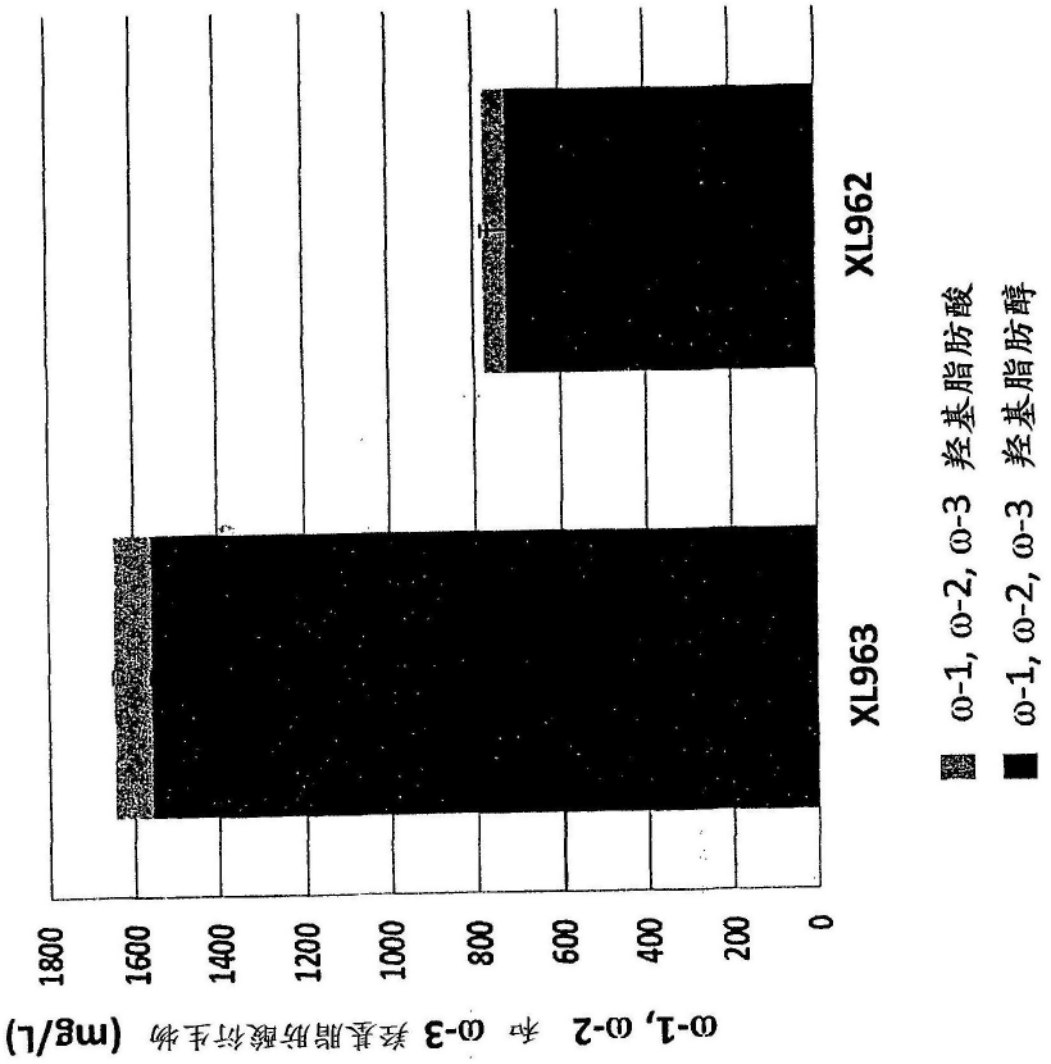
ω-1, ω-2 和 ω-3 羟基脂肪酸衍生物 (mg/L)
1800
1600
1400
1200
1000
800
600
400
200
0
XL963
XL962
ω-1, ω-2, ω-3 羟基脂肪酸
ω-1, ω-2, ω-3 羟基脂肪醇

图24B

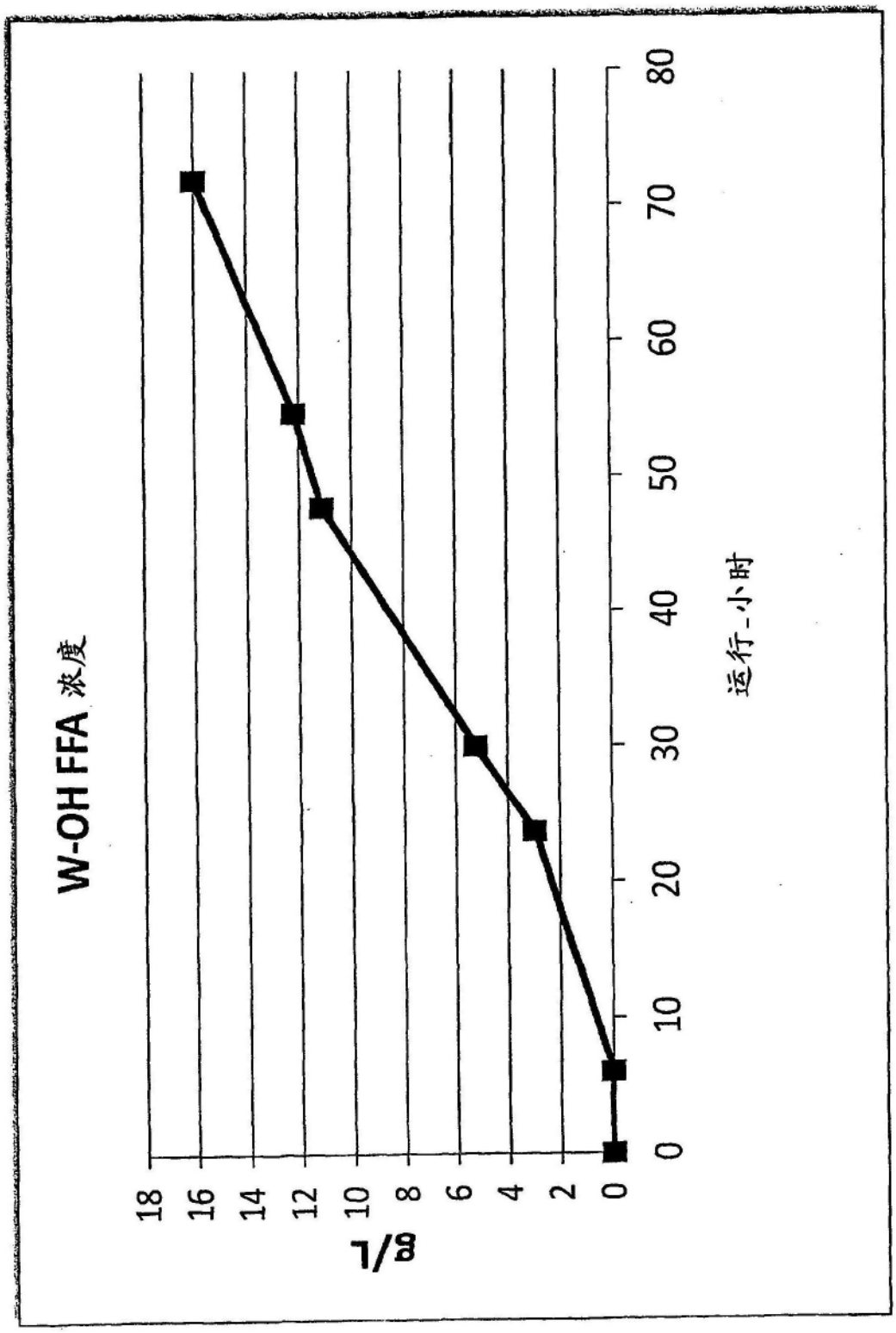

图25

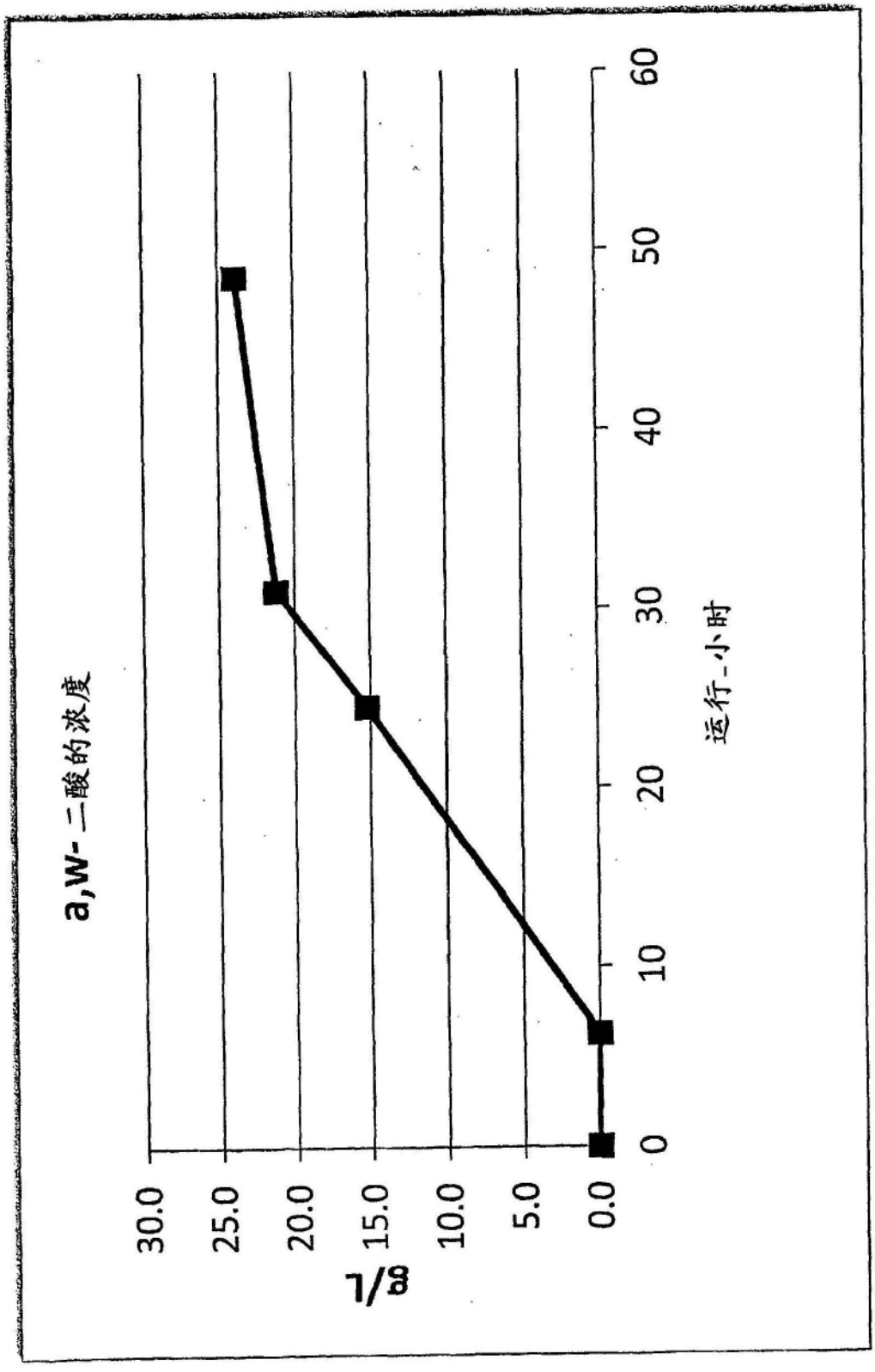

图26

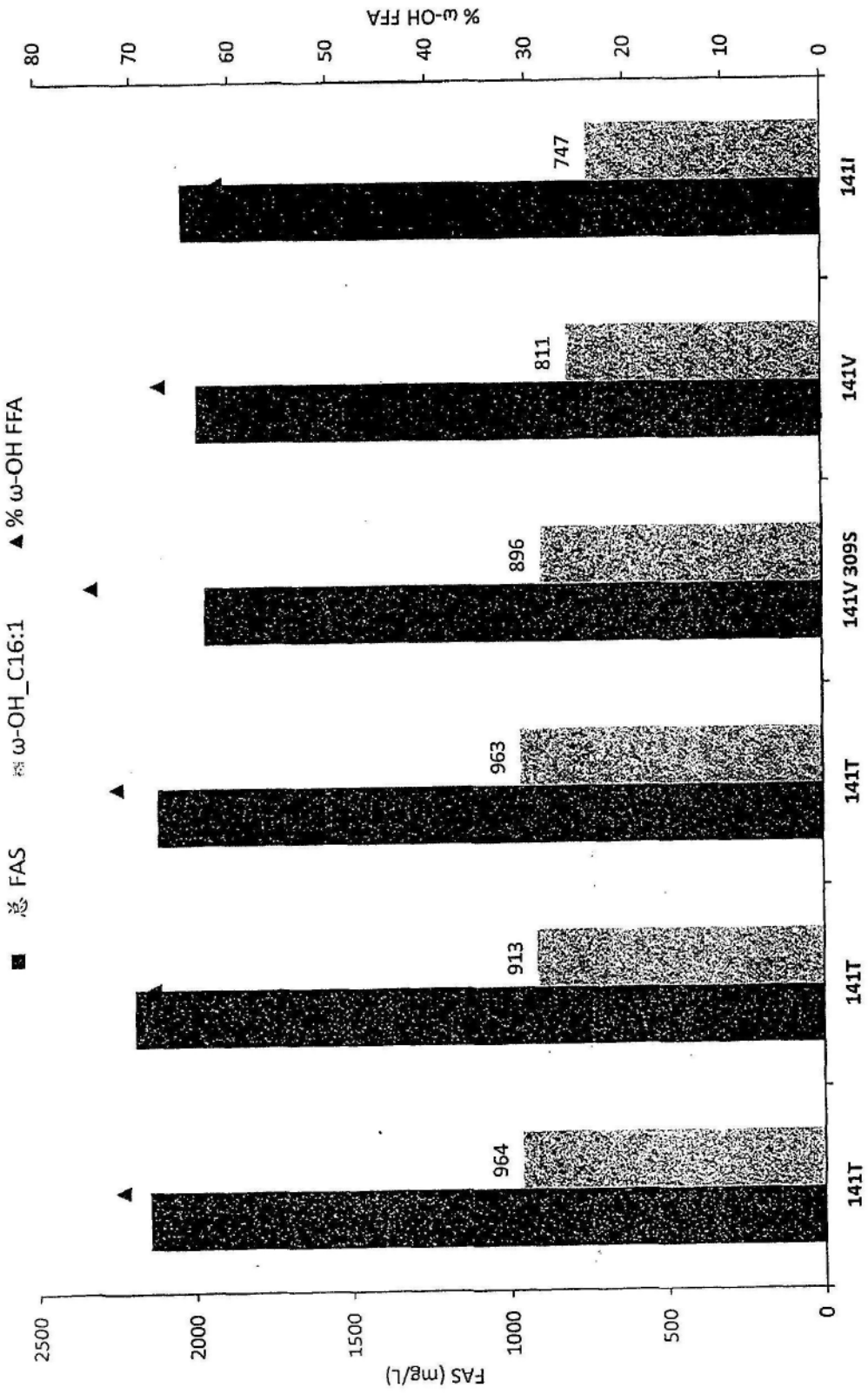
FAS (mg/L)
2500
2000
1500
1000
500
0
% ω-OH FFA
80
70
60
50
40
30
20
10
0
总 FAS
ω-OH_C16:1
% ω-OH FFA
964
913
963
896
811
747
141T
141T
141T
141V 309S
141V
141I

图27