

[19] 中华人民共和国国家知识产权局



# [12] 发明专利申请公布说明书

[21] 申请号 200780050429. X

[51] Int. Cl.

C12N 9/18 (2006.01)

C12N 15/75 (2006.01)

C12N 1/21 (2006.01)

[43] 公开日 2010年2月17日

[11] 公开号 CN 101652474A

[22] 申请日 2007.1.25

[21] 申请号 200780050429. X

[86] 国际申请 PCT/IB2007/000558 2007.1.25

[87] 国际公布 WO2008/090395 英 2008.7.31

[85] 进入国家阶段日期 2009.7.24

[71] 申请人 丹尼斯科有限公司

地址 丹麦哥本哈根

[72] 发明人 马克·科尔克曼

约恩·达尔高·米克尔森

里克·赫格·洛伦森

[74] 专利代理机构 隆天国际知识产权代理有限公司

代理人 吴小瑛 吕俊清

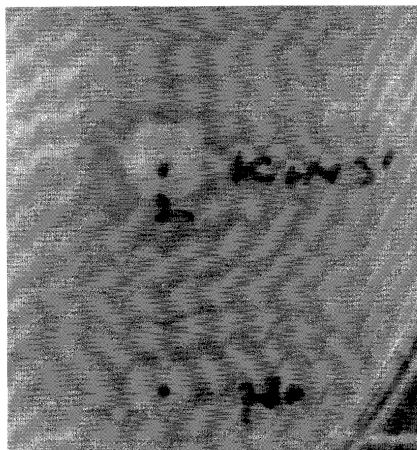
权利要求书4页 说明书61页 附图41页

## [54] 发明名称

由地衣芽孢杆菌转化细胞制备脂酰基转移酶

## [57] 摘要

本发明涉及脂酰基转移酶的制备方法，步骤包括：(i) 提供地衣芽孢杆菌细胞；(ii) 用编码脂酰基转移酶的异源核苷酸序列转化地衣芽孢杆菌细胞；和(iii) 在启动子序列的控制下在细胞中表达脂酰基转移酶。此外，本发明还涉及地衣芽孢杆菌在表达脂酰基转移酶中的应用，包含异源脂酰基转移酶的地衣芽孢杆菌宿主细胞，和包含编码脂酰基转移酶的核苷酸序列的载体，所述核苷酸序列可操作地连接与地衣芽孢杆菌同源的启动子序列。



- 1、用于制备脂酰基转移酶的方法，包括以下步骤：
  - (i) 提供地衣芽孢杆菌细胞；
  - (ii) 用编码脂酰基转移酶的异源核苷酸序列转化所述地衣芽孢杆菌细胞；和
  - (iii) 在启动子序列的控制下，在所述细胞中表达脂酰基转移酶。
- 2、如权利要求1所述的方法，其中，所述启动子序列与编码脂酰基转移酶的核苷酸序列不是天然连接。
- 3、如权利要求2所述的方法，其中，编码信号肽的核苷酸序列可操作地连接所述编码脂酰基转移酶的异源核苷酸序列。
- 4、如前述任一项权利要求所述的方法，其中，所述方法还包括分离/回收所述脂酰基转移酶的步骤。
- 5、如前述任一项权利要求所述的方法，其中，所述启动子序列与宿主细胞是同源的。
- 6、如前述任一项权利要求所述的方法，其中，所述启动子序列可以选自以下组成的组： $\alpha$ -淀粉酶启动子序列、蛋白酶启动子序列、枯草杆菌蛋白酶启动子序列、谷氨酸特异性蛋白酶启动子序列和果聚糖蔗糖酶启动子序列。
- 7、如前述任一项权利要求所述的方法，其中，所述启动子序列为  $\alpha$ -淀粉酶启动子序列。
- 8、如前述任一项权利要求所述的方法，其中，所述编码脂酰基转移酶的核苷酸序列编码包含 GDSx 基序和/或 GANDY 基序的脂酰基转移酶。
- 9、如前述任一项权利要求所述的方法，其中所述编码脂酰基转移酶的核苷酸序列编码如下所述的脂酰基转移酶，该脂酰基转移酶的特征在于其是具有酰基转移酶活性并包含氨基酸序列基序 GDSX 的酶，其中 X 是以下氨基酸残基中的一种或多种：L、A、V、I、F、Y、H、Q、T、N、M 或 S。
- 10、如前述任一项权利要求所述的方法，其中，所述编码脂酰基转移酶的核苷酸序列编码如下所述的脂酰基转移酶，该脂酰基转移酶可以从一种或多种以下属的生物体中获得：气单孢菌属、链霉菌属、酵母菌属、乳球菌属、分支杆菌属、链球菌属、乳杆菌属、脱亚硫酸菌属、芽孢杆菌、

弯曲菌属、弧菌科、木杆菌属、硫化叶菌属、曲霉属、裂殖酵母属、李斯特菌属、奈瑟菌属、中慢生根瘤菌属、雷尔氏菌属、黄单胞菌属和念珠菌属。

11、如前述任一项权利要求所述的方法，其中，所述编码脂酰基转移酶的核苷酸序列编码可以从来自气单孢菌属的生物体获得的脂酰基转移酶。

12、如前述任一项权利要求所述的方法，其中，所述编码脂酰基转移酶的核苷酸序列编码如下所述的脂酰基转移酶，该脂酰基转移酶在对应于 SEQ ID No. 35 所示的嗜水气单孢菌脂酰基转移酶氨基酸序列的 N-80 的位置上包含天冬氨酸残基。

13、如前述任一项权利要求所述的方法，其中，所述编码脂酰基转移酶的核苷酸序列编码如下所述的脂酰基转移酶，该脂酰基转移酶包含 SEQ ID No. 16 所示的氨基酸序列或与 SEQ ID No. 16 具有至少 75%同源性的氨基酸序列。

14、如前述任一项权利要求所述的方法，其中，所述编码脂酰基转移酶的核苷酸序列编码包含 SEQ ID No. 16 所示的氨基酸序列的脂酰基转移酶。

15、包含编码异源脂酰基转移酶的核苷酸序列的地衣芽孢杆菌宿主细胞。

16、如权利要求 14 所述的宿主细胞，其中，所述编码脂酰基转移酶的核苷酸序列编码包含 GDSx 基序和/或 GANDY 基序的脂酰基转移酶。

17、如权利要求 15 或 16 所述的宿主细胞，其中，所述编码脂酰基转移酶的核苷酸序列编码如下所述的脂酰基转移酶，该脂酰基转移酶的特征在于其是具有酰基转移酶活性并包含氨基酸序列基序 GDSX 的酶，其中 X 是以下氨基酸残基中的一种或多种：L、A、V、I、F、Y、H、Q、T、N、M 或 S。

18、如权利要求 15-17 中任一项所述的宿主细胞，其中，所述脂酰基转移酶可以从一种或多种以下属的生物体中获得：气单孢菌属、链霉菌属、酵母菌属、乳球菌属、分支杆菌属、链球菌属、乳杆菌属、脱亚硫酸菌属、芽孢杆菌、弯曲菌属、弧菌科、木杆菌属、硫化叶菌属、曲霉属、裂殖酵

母属、李斯特菌属、奈瑟菌属、中慢生根瘤菌属、雷尔氏菌属、黄单胞菌属和念珠菌属。

19、如权利要求 15-17 中任一项所述的宿主细胞，其中所述编码脂酰基转移酶的核苷酸序列编码可以从来自气单胞菌属的生物体中获得的脂酰基转移酶。

20、如权利要求 15-19 中任一项所述的宿主细胞，其中，所述编码脂酰基转移酶的核苷酸序列编码如下所述的脂酰基转移酶，该脂酰基转移酶在对应于 SEQ ID No. 35 所示的嗜水气单胞菌的脂酰基转移酶氨基酸序列的 N-80 的位置上包含天冬氨酸残基。

21、如权利要求 15-20 中任一项所述的宿主细胞，其中，所述编码脂酰基转移酶的核苷酸序列编码如下所述的脂酰基转移酶，该脂酰基转移酶包含 SEQ ID No. 16 所示的氨基酸序列或与 SEQ ID No. 16 具有 75%或更高同源性的氨基酸序列。

22、如权利要求 15-21 中任一项所述的宿主细胞，其中，所述编码脂酰基转移酶的核苷酸序列编码包含 SEQ ID No. 16 所示的氨基酸序列的脂酰基转移酶。

23、地衣芽孢杆菌宿主细胞在制备异源脂酰基转移酶中的应用。

24、如权利要求 23 所述的应用，其中，与枯草芽孢杆菌宿主细胞相比，在地衣芽孢杆菌宿主细胞中的表达得到提高。

25、包含编码脂酰基转移酶的核苷酸序列的表达载体，其中，所述核苷酸序列可操作地连接与地衣芽孢杆菌同源的启动子序列。

26、如权利要求 25 所述的表达载体，其中，所述启动子不与编码脂酰基转移酶的核苷酸序列天然连接。

27、如权利要求 25 或 26 所述的表达载体，其中，编码信号肽的核苷酸序列可操作地连接所述编码脂酰基转移酶的异源核苷酸序列。

28、如前述任一项权利要求所述的表达载体，其中，所述启动子序列可以选自以下组成的组： $\alpha$ -淀粉酶启动子序列、蛋白酶启动子序列、枯草杆菌蛋白酶启动子序列、谷氨酸特异性蛋白酶启动子序列和果聚糖蔗糖酶启动子序列。

29、如前述任一项权利要求所述的表达载体，其中，所述启动子序列

为  $\alpha$ -淀粉酶启动子序列。

30、用于制备脂酰基转移酶的方法，包括以下步骤：

(i) 提供芽孢杆菌细胞，其中所述芽孢杆菌细胞为除枯草芽孢杆菌之外的一种；

(ii) 用编码脂酰基转移酶的异源核苷酸序列转化芽孢杆菌细胞，其中所述芽孢杆菌细胞为除枯草芽孢杆菌之外的一种；和

(iii) 在启动子序列的控制下，在所述细胞中表达脂酰基转移酶。

31、芽孢杆菌宿主细胞，其中，所述芽孢杆菌细胞为除枯草芽孢杆菌之外的一种，其包含编码异源脂酰基转移酶的核苷酸序列。

32、芽孢杆菌宿主细胞在制备异源脂酰基转移酶中的应用，其中，所述芽孢杆菌宿主细胞为除枯草芽孢杆菌之外的一种。

33、参照实施例和附图基本如前所述的应用。

34、参照实施例和附图基本如前所述的表达载体。

35、参照实施例和附图基本如前所述的地衣芽孢杆菌宿主细胞。

## 由地衣芽孢杆菌转化细胞制备脂酰基转移酶

### 相关申请的引用

引用以下相关申请：US 2002-0009518、US 2004-0091574、WO2004/064537、WO2004/064987、WO2005/066347、WO2005/066351、2006年2月2日提交的美国申请第60/764,430号和WO2006/008508。所述每个申请以及各申请中所引用的每篇文献（“申请引用的文献”），和申请引用的文献中所参考或引用的每篇文献（无论是出现在这些申请的正文中或是在审查过程中）以及在所述审查中支持先进专利性的辩论，均通过引用并入本文。本文还引用了多份文献（“本文引用的文献”）。因此，每篇被本文引用的文献以及本文引用文献中所引用或参考的每篇文献均通过引用并入本文。

### 技术领域

本发明涉及脂酰基转移酶(lipid acyltransferase)的制备。具体而言，本发明涉及通过在芽孢杆菌(*Bacillus*)宿主细胞(优选地衣芽孢杆菌(*Bacillus licheniformis*)宿主细胞)中表达脂酰基转移酶来制备脂酰基转移酶的方法。此外，本发明还涉及芽孢杆菌(优选为地衣芽孢杆菌)在表达脂酰基转移酶上的应用，还涉及在其基因组中包含编码脂酰基转移酶的基因的芽孢杆菌宿主细胞(优选地衣芽孢杆菌宿主细胞)。

### 背景技术

已知脂酰基转移酶在食品应用中具有优势。已经发现脂酰基转移酶在食物中具有显著的酰基转移酶活性。这种活性在食物的制备方法中有惊人的有益应用。

例如，WO 2004/064537 公开了通过使用脂酰基转移酶来原位制备乳化剂的方法和与其相关的优点。

因此，需要脂酰基转移酶的商业制备方法。

然而，基因通常可能难以在异源宿主中表达，且在宿主细胞中表达脂酰基转移酶可能会有问题。

WO 2004/064537 公开了在枯草芽孢杆菌 (*Bacillus subtilis*) 和大肠杆菌 (*Escherichia coli*) 的表达两种气单胞菌 (*Aeromonas*) 脂酰基转移酶。然而，其在枯草芽孢杆菌中的表达很低，同时大肠杆菌不是 GRAS 生物体并因此不适合作为食品工业用酶的宿主。

US 6,255,076 公开了在芽孢杆菌 (*Bacillus*) 宿主细胞中制备多肽的方法。然而，此种方法需要使用串联的启动子，其中每个启动子序列可操作地连接编码所述多肽序列的单拷贝核酸序列。因此，本领域中需要一种制备脂酰基转移酶的改进方法。

## 发明内容

本发明的各个方面体现在权利要求书和以下详述中。

本发明的一个方面涉及脂酰基转移酶的制备方法，所述方法包括以下步骤：

(i) 提供宿主细胞，优选为芽孢杆菌宿主细胞，其中所述芽孢杆菌宿主细胞为除枯草芽孢杆菌之外的一种，优选为地衣芽孢杆菌细胞；

(ii) 用编码脂酰基转移酶的异源核苷酸序列转化宿主细胞，所述宿主细胞优选为芽孢杆菌宿主细胞，其中所述芽孢杆菌宿主细胞为除枯草芽孢杆菌之外的一种，优选为地衣芽孢杆菌细胞；和

(iii) 在启动子序列的控制下在细胞中表达脂酰基转移酶。

另一方面，本发明涉及包含异源脂酰基转移酶的芽孢杆菌宿主细胞，其中所述芽孢杆菌宿主细胞为除枯草芽孢杆菌之外的一种，优选为地衣芽孢杆菌宿主细胞。

另一方面，本发明涉及芽孢杆菌宿主细胞在制备异源脂酰基转移酶中的应用，其中所述芽孢杆菌宿主细胞为除枯草芽孢杆菌之外的一种，优选为地衣芽孢杆菌宿主细胞。

适合地，与在枯草芽孢杆菌中的表达相比，在芽孢杆菌宿主中的表达可以引起表达的增加，其中所述芽孢杆菌宿主为除枯草芽孢杆菌之外的一种，且优选芽孢杆菌宿主为地衣芽孢杆菌。

再一方面，本发明涉及包含编码脂酰基转移酶的核苷酸序列的表达载体，所述核苷酸序列可操作地连接一个或多个调控序列，使得所述调控序列能够在适合的宿主或宿主细胞中表达编码脂酰基转移酶的核苷酸序列，优选在芽孢杆菌宿主（或细胞），其中所述芽孢杆菌宿主（或细胞）为除枯草芽孢杆菌之外的一种，优选在地衣芽孢杆菌或地衣芽孢杆菌细胞。

适合地，所述脂酰基转移酶可以是重组的脂酰基转移酶。

### 具体实施方式

根据本发明的第一方面，本发明提供了脂酰基转移酶的制备方法，所述方法包括以下步骤：

(i) 提供宿主细胞，优选为芽孢杆菌宿主细胞，其中所述芽孢杆菌宿主细胞为除枯草芽孢杆菌之外的一种，优选为地衣芽孢杆菌细胞；

(ii) 用编码脂酰基转移酶的异源核苷酸序列转化宿主细胞，优选为芽孢杆菌宿主细胞，其中所述芽孢杆菌宿主细胞为除枯草芽孢杆菌之外的一种，优选为地衣芽孢杆菌细胞；和

(iii) 在启动子序列的控制下在细胞中表达脂酰基转移酶。

此外，可以使编码信号肽的核苷酸序列可操作地连接编码脂酰基转移酶的所述异源核苷酸序列。

适合地，本发明的方法还包括分离/回收脂酰基转移酶的附加步骤。

另一方面，本发明涉及包含异源脂酰基转移酶的地衣芽孢杆菌宿主细胞。

适合地，所述脂酰基转移酶可以是重组的脂酰基转移酶。

适合地，在本发明的宿主细胞、载体、方法和/或用途中所用的启动子序列可以与宿主细胞同源。“与宿主细胞同源”是指来源于宿主生物体；即，所述启动子序列是在宿主生物体中自然发现的。适合地，所述启动子序列可以选自由编码以下启动子的核苷酸序列组成的组： $\alpha$ -淀粉酶启动子、蛋白酶启动子、枯草杆菌蛋白酶启动子、谷氨酸特异性蛋白酶启动子和果聚糖蔗糖酶启动子。适合地，所述启动子序列可以是编码以下启动子的核苷酸序列：LAT（如来自地衣芽孢杆菌的 $\alpha$ -淀粉酶启动子，也称作 AmyL）、AprL（如枯草杆菌蛋白酶 Carlsberg 启动子）、EndoGluC（如来自地衣芽孢



杆菌的谷氨酸特异性启动子)、AmyQ (如来自解淀粉芽孢杆菌 (*Bacillus amyloliquefaciens*)  $\alpha$ -淀粉酶启动子的  $\alpha$ -淀粉酶启动子) 和 SacB (如枯草芽孢杆菌的果聚糖蔗糖酶启动子)。

在本发明的一个实施方案中,所述启动子序列为  $\alpha$ -淀粉酶启动子的-35至-10 序列,优选为地衣芽孢杆菌  $\alpha$ -淀粉酶启动子的-35至-10 序列。所述“-35至-10 序列”描述了相对于转录起点的位置。“-35”和“-10”均为框(即多个核苷酸),每个框包含 6 个核苷酸,且这些框被 17 个核苷酸分开。这 17 个核苷酸通常被称为“间隔子”。图 55 对此进行了说明,其中下划线处为-35 框和-10 框。为了避免混淆,本文所用的“-35至-10 序列”是指从-35 框的起始至-10 框的终止的序列,即包括-35 框、长度为 17 个核苷酸的间隔子和-10 框。

在一些方面,用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列可以包含 GDSx 基序和/或 GANDY 基序。

优选地,将所述脂酰基转移酶表征为具有酰基转移酶活性并包含氨基酸序列基序 GDSX 的酶,其中 X 是以下氨基酸残基中的一种或多种: L、A、V、I、F、Y、H、Q、T、N、M 或 S。

适合地,优选用于本发明载体、宿主细胞、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列可以从以下一个或多个属的生物体中获得: 气单孢菌属、链霉菌属 (*Streptomyces*)、酵母菌属 (*Saccharomyces*)、乳球菌属 (*Lactococcus*)、分支杆菌属 (*Mycobacterium*)、链球菌属 (*Streptococcus*)、乳杆菌属 (*Lactobacillus*)、脱亚硫酸菌属 (*Desulfitobacterium*)、芽孢杆菌属、弯曲菌属 (*Campylobacter*)、弧菌科 (*Vibrionaceae*)、木杆菌属 (*Xylella*)、硫化叶菌属 (*Sulfolobus*)、曲霉属 (*Aspergillus*)、裂殖酵母属 (*Schizosaccharomyces*)、李斯特菌属 (*Listeria*)、奈瑟菌属 (*Neisseria*)、中慢生根瘤菌属 (*Mesorhizobium*)、雷尔氏菌属 (*Ralstonia*)、黄单胞菌属 (*Xanthomonas*) 和念珠菌属 (*Candida*)。优选地,所述脂酰基转移酶是可以获得的,且优选从气单孢菌属的生物体中获得。

在本发明的一些方面,用于本发明载体、宿主细胞、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列所编码的脂酰基转移酶在对应于

SEQ ID No. 35 所示的嗜水气单胞菌 (*Aeromonas hydrophila*) 脂酰基转移酶的氨基酸序列中的 N-80 位置上包含天冬氨酸残基。

此外或备选地, 用于本发明载体、宿主细胞、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列编码的脂酰基转移酶包含 SEQ ID No. 16 所示的氨基酸序列, 或与其具有 75% 或更高同源性的氨基酸序列。适合地, 编码脂酰基转移酶的核苷酸序列所编码的脂酰基转移酶可以包含 SEQ ID No. 16 所示的氨基酸序列。

本文所用的术语“异源的”是指源自分离的遗传资源或物种的序列。异源序列是非宿主序列、修饰的序列、来自不同宿主细胞株的序列、或来自宿主细胞的不同染色体位置的同源序列。

“同源”序列是指在同一遗传资源或物种中发现的序列, 即其天然存在于宿主细胞的相关物种中。

本文所用的术语“重组的脂酰基转移酶”是指已经通过遗传重组的方式制备的脂酰基转移酶。例如, 将编码脂酰基转移酶的核苷酸序列插入克隆载体, 得到以存在异源脂酰基转移酶为特征的地衣芽孢杆菌细胞。

#### 宿主细胞

在本发明的一个实施方案中, 用于本发明方法和/或应用中的宿主细胞是地衣芽孢杆菌宿主细胞。

已经发现与其它生物体如枯草芽孢杆菌相比, 使用地衣芽孢杆菌宿主细胞可使脂酰基转移酶的表达增加。

已经将来自于杀鲑气单胞菌 (*Aeromonas salmonicida*) 的脂酰基转移酶插入到多个常规表达载体中, 所述表达载体经设计可分别在枯草芽孢杆菌、多型汉逊酵母 (*Hansenula polymorpha*)、粟酒裂殖酵母 (*Schizosaccharomyces pombe*) 和塔宾曲霉 (*Aspergillus tubigenensis*) 中有最佳的表达。然而, 在多型汉逊酵母、粟酒裂殖酵母和塔宾曲霉中仅检测到极低的水平。这些表达水平低于 1  $\mu\text{g/ml}$ , 因此不可能选择出产生足量蛋白以起始商业生产的细胞 (结果未显示)。相反, 地衣芽孢杆菌能够产生对商业生产的可行性具有吸引力的蛋白水平。

具体而言, 已经发现在 *aprE* 启动子的控制下时, 地衣芽孢杆菌中的表

达比在枯草芽孢杆菌中的表达高约 100 倍,或在 A4 启动子的控制下并与纤维素融合时,比在变铅青链霉菌 (*S. lividans*) 中的表达高约 100 倍(结果未显示在本文中)。

在另一实施方案中,宿主细胞可以为除枯草芽孢杆菌之外的任何芽孢杆菌细胞。优选地,所述芽孢杆菌宿主细胞可以是以下物种中的一种:地衣芽孢杆菌、嗜碱芽孢杆菌 (*B. alkalophilus*)、解淀粉芽孢杆菌、环状芽孢杆菌 (*B. circulans*)、克劳氏芽孢杆菌 (*B. clausii*)、凝结芽孢杆菌 (*B. coagulans*)、坚强芽孢杆菌 (*B. firmus*)、*B. lautus*、缓慢芽孢杆菌 (*B. lentus*)、巨大芽孢杆菌 (*B. megaterium*)、短小芽孢杆菌 (*B. pumilus*) 或嗜热脂肪芽孢杆菌 (*S. stearothermophilus*)。

本发明涉及的术语“宿主细胞”所包括具有以下特征的任何细胞:包含本文所定义的编码脂酰基转移酶的核苷酸序列或如上所述的表达载体,且用于具有如本文所定义的特异性质的脂酰基转移酶的重组制备。

因此,本发明的其它实施方案提供了宿主细胞,其包含(如经转化或转染)本发明的核苷酸序列,或包含表达具有如本文所定义的特异性质的多肽的核苷酸序列。

适合地,在一些实施方案中,宿主细胞可以是蛋白酶缺陷型或无蛋白酶型细胞株和/或  $\alpha$ -淀粉酶缺陷型或无  $\alpha$ -淀粉酶型细胞株。

### 调控序列

在一些应用中,用于本发明宿主细胞、载体、方法和/或应用中任一项的脂酰基转移酶序列可操作地连接能够通过如所选宿主细胞(如地衣芽孢杆菌细胞)表达核苷酸序列的调控序列。

例如,本发明涵盖包含可操作地连接此调控序列的本发明核苷酸序列的载体,即所述载体是表达载体。

术语“可操作地连接”是指并列放置(juxtaposition),其中所述组件的关系允许它们能以其希望的方式发挥功能。“可操作地连接”编码序列的调控序列以这样的方式连接,即能在与控制序列相容的条件下实现编码序列的表达。

术语“调控序列”包括启动子和增强子以及其它表达调控信号。

所用的术语“启动子”为本领域的一般含义，如 RNA 聚合酶结合位点。

编码具有如本文所定义的特性的酶的核苷酸序列的增强表达也可以通过选择调控区（如启动子、分泌引导子和终止子区）来实现，所述终止子区本质上不是编码所述酶的核苷酸序列的调控区。

适合地，本发明的核苷酸序列至少可以可操作地连接启动子。

适合地，编码脂酰基转移酶的核苷酸序列可以可操作地连接编码终止子序列的核苷酸序列。用于本发明载体、宿主细胞、方法和/或应用中任一项的适合的终止子序列的实例包括： $\alpha$ -淀粉酶终止子序列（如，CGGGACTTACCGAAAGAAACCATCAATGATGGTTTCTTTTTTGTTCATAAA-SEQ ID No. 64）、碱性蛋白酶终止子序列（如，CAAGACTAAAGACCGTTCGCCCGTTTTTGCAATAAGCGGGCGAATCTTACATAAAAATA-SEQ ID No. 65）、谷氨酸特异性终止子序列（如，ACGGCCGTTAGATGTGACAGCCCGTTCCAAAAGGAAGCGGGCTGTCTTCGTGTATTATTGT-SEQ ID No. 66）、果聚糖酶终止子序列（如，TCTTTTAAAGGAAAGGCTGGAATGCCCGGCATTCCAGCCACATGATCATCGTTT-SEQ ID No. 67）和枯草杆菌蛋白酶 E 终止子序列（如，GCTGACAAATAAAAAGAAGCAGGTATGGAGGAACCTGCTTCTTTTACTATTATTG）。适合地，编码脂酰基转移酶的核苷酸序列可以可操作地连接  $\alpha$ -淀粉酶终止子，如地衣芽孢杆菌  $\alpha$ -淀粉酶终止子。

### 启动子

根据本发明所用的启动子序列与编码脂酰基转移酶的序列可以是异源的或同源的。

所述启动子序列可以是能够指导脂酰基转移酶在所选宿主细胞中表达的任何启动子序列。

适合地，所述启动子序列可以与芽孢杆菌（如地衣芽孢杆菌）同源。优选地，所述启动子序列与所选宿主细胞是同源的。

用于本发明的适合的启动子序列包括：地衣芽孢杆菌  $\alpha$ -淀粉酶基因的启动子、地衣芽孢杆菌枯草杆菌蛋白酶基因的启动子、枯草芽孢杆菌枯草杆菌蛋白酶基因的启动子、地衣芽孢杆菌碱性蛋白酶基因（枯草杆菌蛋白

酶 Carlsberg 基因) 的启动子、地衣芽孢杆菌谷氨酸特异性蛋白酶基因的启动子、解淀粉芽孢杆菌  $\alpha$ -淀粉酶基因的启动子、枯草芽孢杆菌果聚糖蔗糖酶的启动子和具有  $\alpha$ -淀粉酶基因的“-35”区序列 TTGACA 和“-10”区序列 TATAAT 的“共有”启动子(即-35至-10启动子)。

适合在本发明方法中指导核酸序列转录的启动子的其它实例包括: 缓慢芽孢杆菌碱性蛋白酶基因(aprH)的启动子、枯草芽孢杆菌  $\alpha$ -淀粉酶基因(amyE)的启动子、嗜热脂肪芽孢杆菌麦芽糖淀粉酶基因(amyM)的启动子、地衣芽孢杆菌青霉素酶基因(penP)的启动子、枯草芽孢杆菌 xylA 和 xylB 基因的启动子、和/或苏云金芽孢杆菌拟步行甲亚种(*Bacillus thuringiensis* subsp. *tenebrionis*) CryIII A 基因的启动子。

在优选实施方案中, 所述启动子序列为  $\alpha$ -淀粉酶启动子(如地衣芽孢杆菌  $\alpha$ -淀粉酶启动子)。优选地, 所述启动子序列包含地衣芽孢杆菌  $\alpha$ -淀粉酶启动子的-35至-10序列(参见图 53 和图 55)。

### 信号肽

根据所用的序列和/或载体, 宿主细胞通过表达编码脂酰基转移酶的核苷酸序列所产生的脂酰基转移酶可以被分泌出或包含在细胞内。

信号序列可以用于指导编码序列分泌通过特定的细胞膜。所述信号序列对脂酰基转移酶编码序列来说可以是天然的或外源的。例如, 所述信号肽编码序列是从芽孢杆菌物种, 优选是从地衣芽孢杆菌获得的淀粉酶或蛋白酶基因。

可以由一种或多种以下基因获得适合的信号肽编码序列: 麦芽糖  $\alpha$ -淀粉酶基因、枯草杆菌蛋白酶基因、 $\beta$ -内酰胺酶基因、中性蛋白酶基因、prsA 基因和/或酰基转移酶基因。

优选地, 所述信号肽为地衣芽孢杆菌  $\alpha$ -淀粉酶的信号肽、气单孢菌酰基转移酶的信号肽(如, mkkwfvcllgialtvqa-SEQ ID No. 21)、枯草芽孢杆菌枯草杆菌蛋白酶的信号肽(如, mrskklwisllfaltliftmafnsmsaqa-SEQ ID No. 22) 或地衣芽孢杆菌枯草杆菌蛋白酶的信号肽(如, mmrkksfwgmltafmlvftmefdsasa-SEQ ID No. 23)。适合地, 所述信号肽可以为地衣芽孢杆菌  $\alpha$ -淀粉酶的信号肽。

然而，可以使用能够指导所表达的脂酰基转移酶进入所选芽孢杆菌宿主细胞（优选为地衣芽孢杆菌宿主细胞）的分泌途径的任何信号肽编码序列。

在本发明的一些实施方案中，编码信号肽的核苷酸序列可以可操作地连接编码所选脂酰基转移酶的核苷酸序列。

所选脂酰基转移酶可以在本文所定义的宿主细胞中作为融合蛋白表达。

### 表达载体

术语“表达载体”是指能够在体内或体外表达的构建体。

优选地，所述表达载体被引入到生物体如地衣芽孢杆菌宿主的基因组中。术语“引入”优选涵盖稳定的引入到基因组中。

如本文所定义的编码脂酰基转移酶的核苷酸序列可以存在于载体中，在该载体中所述核苷酸序列可操作地连接调控序列，使得所述调控序列能够通过适合的宿主生物体（如地衣芽孢杆菌）表达所述核苷酸序列，即所述载体是表达载体。

本发明的载体可以被转化到上述的适合的宿主细胞中，以表达具有如本文所定义的脂酰基转移酶活性的多肽。

通常根据其将被引入的宿主细胞来选择载体（如质粒、粘粒、病毒或噬菌体载体、基因组插入物）。本发明可以涵盖具有等同功能且本领域已知或可知的其它形式的表达载体。

一旦转化到所选择的宿主细胞中，载体可以不依赖宿主细胞的基因组而复制并发挥功能，或可以自行整合进入基因组。

所述载体可以包含一种或多种选择性标记基因，如提供抗生素抗性的基因，如氨苄青霉素抗性、卡那霉素抗性、氯霉素抗性 or 四环素抗性的基因。或者，可以通过共转化完成选择（如 WO91/17243 中所述）。

载体可以在体外使用，例如，用于制备 RNA 或用于转染或转化宿主细胞。

因此，在另一实施方案中，本发明提供用于本发明载体、宿主细胞、其它方法和/或应用中任一项的本发明的核苷酸序列或编码具有如本文所

定义的特定性质的多肽的核苷酸序列的制备方法，所述方法是通过将核苷酸序列引入可复制载体，将所述载体引入相容的宿主细胞，并使宿主细胞在可以进行载体复制的条件下生长。

所述载体可以进一步包含能使该载体在所述宿主细胞中复制的核苷酸序列。所述序列的实例为质粒 pUC19、pACYC177、pUB110、pE194、pAMB1 和 pIJ702 的复制起点。

### 脂酰基转移酶

用于本发明方法、载体和/或应用中任一项的编码脂酰基转移酶的核苷酸序列可以编码天然的脂酰基转移酶或脂酰基转移酶变体。

例如，用于本发明的编码脂酰基转移酶的核苷酸序列可以是 WO2004/064537、WO2004/064987、WO2005/066347 或 WO2006/008508 中所述的一种。这些文献通过引用并入本文。

本文所用的术语“脂酰基转移酶”优选地是指具有脂酰基转移酶活性的酶（通常被归类为 E.C. 2.3.1.x，如 2.3.1.43），其中所述酶能够将酰基基团从脂质转移到一个或多个受体（acceptor）物质，如以下的一种或多种：固醇（sterol）、甾烷醇（stanol）、碳水化合物、蛋白、蛋白亚基、糖醇（如抗坏血酸和/或甘油），优选甘油和/或固醇，如胆固醇。

优选地，用于本发明载体、宿主细胞、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列编码能够将酰基基团从磷脂（如本文所定义）转移到糖醇（如抗坏血酸和/或甘油，最优选为甘油）的脂酰基转移酶。

对于一些方面，本发明的“酰基受体”可以是包含羟基基团（-OH）的任何化合物，例如多价的醇，包括甘油、固醇、甾烷醇、碳水化合物；羟酸，包括果酸、柠檬酸、酒石酸、乳酸和抗坏血酸；蛋白或其亚基，如氨基酸、蛋白水解产物和肽（部分水解的蛋白）；和其混合物和衍生物。优选地，根据本发明的“酰基受体”不是水。优选地，根据本发明的“酰基受体”是糖醇如多元醇，最优选为甘油。出于本发明的目的，抗坏血酸也被认为是糖醇。

优选地，所述酰基受体不为甘油单酯。

优选地，所述酰基受体不为甘油二酯。

一方面，用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列所编码的脂酰基转移酶可以且能够将酰基基团从脂质转移到甘油，此外能够将酰基基团从脂质转移到以下的一种或多种物质：碳水化合物、蛋白、蛋白亚基、固醇和/或甾烷醇，优选其能够转移到糖醇（如抗坏血酸和/或甘油，最优选为固醇（如胆固醇））和/或植物固醇/甾烷醇。

优选地，脂酰基作用的脂质底物是以下脂质中的一种或多种：磷脂，如卵磷脂，如磷脂酰胆碱。

所述脂质底物在本文中可以被称为“脂酰基供体”。本文所用的术语卵磷脂涵盖磷脂酰胆碱、磷脂酰乙醇胺、磷脂酰肌醇、磷脂酰丝氨酸和磷脂酰甘油。

对于一些方面，用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列优选编码不能或基本上不能作用于甘油三酯和/或 1-甘油单酯和/或 2-甘油单酯的脂酰基转移酶。

对于一些方面，用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列编码不表现出三酰基甘油（triacylglycerol）脂肪酶活性（E.C. 3.1.1.3）或不表现出显著的三酰基甘油脂肪酶活性（E.C. 3.1.1.3）的脂酰基转移酶。

水解甘油三酯的能力（E.C. 3.1.1.3 活性）可以由根据 Food Chemical Codex (3rd Ed., 1981, pp 492-493)（修改为葵花油、pH 5.5 以替代橄榄油、pH 6.5）测定的脂肪酶活性来测定。脂肪酶活性以 LUS（葵花的脂肪酶单位）来测定，其中将 1 LUS 定义为在以上试验条件下每分钟可以从葵花油中释放 1  $\mu\text{mol}$  脂肪酸的酶量。或者，也可以使用如 WO9845453 中定义的 LUT 试验。此文献通过引用并入本文。

用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列可以编码基本上不能作用于甘油三酯的脂酰基转移酶，所述脂酰基转移酶的 LUS/mg 可以小于 1000，如小于 500、如小于 300、优选为小于 200、更优选为小于 100、更优选为小于 50、更优选为小于 20、更优选为小于 10，如小于 5、小于 2、更优选为小于 1 LUS/mg。优选地，LUT/mg 活性小于 500，如小于 300、优选为小于 200、更优选为小于 100、



更优选为小于 50、更优选为小于 20、更优选为小于 10，如小于 5、小于 2、更优选为小于 1 LUT/mg。

用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列可以编码基本上不能作用于甘油单酯的脂酰基转移酶，所述脂酰基转移酶可以在 LUS 试验中使用单油酸酯（M7765 1-油酰-*rac*-甘油 99%）以替代葵花油来测定。将 1 MGHU 定义为在所述试验条件下每分钟可以从甘油单酯中释放 1  $\mu\text{mol}$  脂肪酸的酶量。

用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列可以编码基本上不能作用于甘油三酯的脂酰基转移酶，所述脂酰基转移酶具有的 MGHU/mg 可以小于 5000，如小于 1000、如小于 500、如小于 300、优选为小于 200、更优选为小于 100、更优选为小于 50、更优选为小于 20、更优选为小于 10（如小于 5、小于 2）、更优选为小于 1 MGHU/mg。

适合地，用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列编码可以表现出一种或多种以下磷脂酶活性的脂酰基转移酶：磷脂酶 A2 活性（E.C. 3.1.1.4）和/或磷脂酶 A1 活性（E.C. 3.1.1.32）。所述脂酰基转移酶也可以具有磷脂酶 B 活性（E.C. 3.1.1.5）。

适合地，对于一些方面，所述脂酰基转移酶可以能够将酰基基团从磷脂转移到糖醇（优选为甘油和/或抗坏血酸）。

对于一些方面，优选地用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列编码能够将酰基基团从磷脂转移到固醇和/或甾烷醇以至少形成固醇酯和/或甾烷醇酯的脂酰基转移酶。

所述脂酰基转移酶可以能够将酰基基团从脂质转移到多元醇（如甘油）和/或固醇（如胆固醇或植物固醇）和/或甾烷醇。因此，在一个实施方案中，本发明的“酰基受体”可以为甘油和/或胆固醇或植物固醇和/或甾烷醇。

优选地，所述脂酰基转移酶可以使用以下标准进行表征：

所述酶具有可以被定义为酯转移活性的酰基转移酶活性，通过此活性将脂酰基供体的原初酯键的酰基部分转移到酰基受体（优选为甘油或胆固醇）以形成新的酯；和

所述酶包含氨基酸序列基序 GDSX，其中 X 是以下氨基酸残基中的一

种或多种：L、A、V、I、F、Y、H、Q、T、N、M 或 S。

优选地，GDSX 基序的 X 为 L 或 Y。更优选地，GDSX 基序的 X 为 L。因此，优选本发明的酶包含氨基酸序列基序 GDSL。

GDSX 基序由四个保守的氨基酸构成。优选地，该基序中的丝氨酸为所述脂酰基转移酶的催化性丝氨酸。适合地，GDSX 基序的丝氨酸的位置对应于 Brumlik 和 Buckley (Journal of Bacteriology Apr. 1996, Vol. 178, No. 7, p 2060-2064)中教导的嗜水气单胞菌脂酰基转移酶的 Ser-16。

为了测定蛋白是否具有本发明的 GDSX 基序，优选使用 WO2004/064537 或 WO2004/064987 (均通过引用并入本文)中教导的方法，将序列与 pfam 数据库的隐马尔可夫模型序型(hidden markov model profile) (HMM 序型) 进行比较。

优选地，所述脂酰基转移酶可以使用 Pfam00657 共有序列进行比对(完整的解释可参见 WO2004/064537 或 WO2004/064987)。

优选地，与 pfam00657 结构域家族的隐马尔可夫模型序型 (HMM 序型) 的阳性匹配表示存在本发明的 GDSL 或 GDSX 结构域。

优选地，当与 Pfam00657 共有序列进行比对时，用于本发明方法和/或应用的脂酰基转移酶可以具有至少一个、优选为大于一个、更优选为大于两个的下述部分(block): GDSx、GANDY、HPT。适合地，所述脂酰基转移酶可以具有 GDSx 部分和 GANDY 部分。或者，所述酶可以具有 GDSx 部分和 HPT 部分。优选地，所述酶至少包含 GDSx 部分。更多细节请参见 WO2004/064537 或 WO2004/064987。

优选地，GANDY 基序的残基选自 GANDY、GGNDA、GGNDL，最优选为 GANDY。

优选地，当与 Pfam00657 共有序列进行比对时，与嗜水气单胞菌多肽参考序列 (SEQ ID No. 1) 相比，用于本发明方法和/或应用的酶具有以下氨基酸残基中的至少一个、优选大于两个、优选大于三个、优选大于四个、优选大于五个、优选大于六个、优选大于七个、优选大于八个、优选大于九个、优选大于十个、优选大于十一个、优选大于十二个、优选大于十三个、优选大于十四个：28hid、29hid、30hid、31hid、32gly、33Asp、34Ser、35hid、130hid、131Gly、132Hid、133Asn、134Asp、135hid、309His。

Pfam00657 GDSX 结构域是区分有此结构域蛋白和其它酶的独特标志。

表 3 中显示了 pfam00657 共有序列 (SEQ ID No. 2)。其源自于对第 6 版数据库 pfam 家族 00657 (在本文中也可以称为 pfam00657.6) 的鉴定。

所述共有序列可以通过使用较新版 pfam 数据库来升级 (如参见 WO2004/064537 或 WO2004/064987)。

在一个实施方案中,用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列编码可以使用以下标准进行表征的脂酰基转移酶:

(i) 所述酶具有可以被称为酯转移活性的酰基转移酶活性,通过此活性将脂酰基供体的原初酯键的酰基部分转移到酰基受体 (优选为甘油或胆固醇) 以形成新的酯,优选分别为甘油单酯或胆固醇酯;

(ii) 所述酶包含氨基酸序列基序 GDSX,其中 X 是以下氨基酸残基中的一种或多种: L、A、V、I、F、Y、H、Q、T、N、M 或 S;

(iii) 所述酶包含 His-309 或在图 2 和 4 中 (SEQ ID No. 1 或 SEQ ID No. 3) 所示的嗜水气单胞菌脂酰基转移酶中 His-309 的对应位置上包含组氨酸残基。

优选地,所述 GDSX 基序的氨基酸残基为 L。

在 SEQ ID No. 3 或 SEQ ID No. 1 中,起始的 18 个氨基酸残基形成信号序列。全长序列 (即包含信号序列的蛋白) 中的 His-309 等同于蛋白的成熟部分 (即不含信号序列的序列) 中的 His-291。

在一个实施方案中,用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列所编码的脂酰基转移酶包含以下催化三联体: Ser-34、Asp-306 和 His-309,或在图 4 (SEQ ID No. 3) 或图 2 (SEQ ID No. 1) 中所示的嗜水气单胞菌脂酰基转移酶中 Ser-34、Asp-306 和 His-309 的对应位置上分别包含丝氨酸残基、天冬氨酸残基和组氨酸残基。如上所述,在 SEQ ID No. 3 或 SEQ ID No. 1 所示的序列中,起始的 18 个氨基酸残基形成信号序列。全长序列 (即包含信号序列的蛋白) 中的 Ser-34、Asp-306 和 His-309 等同于蛋白的成熟部分 (即不含信号序列的序列) 中的 Ser-16、Asp-288 和 His-291。在图 3 (SEQ ID No. 2) 所示的 pfam00657

共有序列中，活性位点残基为 Ser-7、Asp-345 和 His-348。

在一个实施方案中，用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列编码可以使用以下标准进行表征的脂酰基转移酶：

所述酶具有可以被称为酯转移活性的酰基转移酶活性，通过此活性将第一脂酰基供体的原初酯键的酰基部分转移到酰基受体以形成新的酯；和

所述酶至少包含 Gly-32、Asp-33、Ser-34、Asp-134 和 His-309，或在 SEQ ID No. 3 或 SEQ ID No. 1 所示的嗜水气单胞菌脂酰基转移酶中 Gly-32、Asp-33、Ser-34、Asp-306 和 His-309 的对应位置上分别包含甘氨酸残基、天冬氨酸残基、丝氨酸残基、天冬氨酸残基和组氨酸残基。

适合地，用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列可以是以下核苷酸序列中的一种：

- (a) SEQ ID No. 36 所示的核苷酸序列（参见图 29）、
- (b) SEQ ID No. 38 所示的核苷酸序列（参见图 31）、
- (c) SEQ ID No. 39 所示的核苷酸序列（参见图 32）、
- (d) SEQ ID No. 42 所示的核苷酸序列（参见图 35）、
- (e) SEQ ID No. 44 所示的核苷酸序列（参见图 37）、
- (f) SEQ ID No. 46 所示的核苷酸序列（参见图 39）、
- (g) SEQ ID No. 48 所示的核苷酸序列（参见图 41）、
- (h) SEQ ID No. 49 所示的核苷酸序列（参见图 57）、
- (i) SEQ ID No. 50 所示的核苷酸序列（参见图 58）、
- (j) SEQ ID No. 51 所示的核苷酸序列（参见图 59）、
- (k) SEQ ID No. 52 所示的核苷酸序列（参见图 60）、
- (l) SEQ ID No. 53 所示的核苷酸序列（参见图 61）、
- (m) SEQ ID No. 54 所示的核苷酸序列（参见图 62）、
- (n) SEQ ID No. 55 所示的核苷酸序列（参见图 63）、
- (o) SEQ ID No. 56 所示的核苷酸序列（参见图 64）、
- (p) SEQ ID No. 57 所示的核苷酸序列（参见图 65）、
- (q) SEQ ID No. 58 所示的核苷酸序列（参见图 66）、
- (r) SEQ ID No. 59 所示的核苷酸序列（参见图 67）、

(s) SEQ ID No. 60 所示的核苷酸序列 (参见图 68)、  
(t) SEQ ID No. 61 所示的核苷酸序列 (参见图 69)、  
(u) SEQ ID No. 62 所示的核苷酸序列 (参见图 70)、  
(v) SEQ ID No. 63 所示的核苷酸序列 (参见图 71) 或  
(w) 与 SEQ ID No. 36、SEQ ID No. 38、SEQ ID No. 39、SEQ ID No. 42、SEQ ID No. 44、SEQ ID No. 46、SEQ ID No. 48、SEQ ID No. 49、SEQ ID No. 50、SEQ ID No. 51、SEQ ID No. 52、SEQ ID No. 53、SEQ ID No. 54、SEQ ID No. 55、SEQ ID No. 56、SEQ ID No. 57、SEQ ID No. 58、SEQ ID No. 59、SEQ ID No. 60、SEQ ID No. 61、SEQ ID No. 62 或 SEQ ID No. 63 中所示的任一序列具有 70%或更高, 优选为 75%或更高同一性的核苷酸序列。

适合地, 所述核苷酸序列可以与 SEQ ID No. 36、SEQ ID No. 38、SEQ ID No. 39、SEQ ID No. 42、SEQ ID No. 44、SEQ ID No. 46、SEQ ID No. 48、SEQ ID No. 49、SEQ ID No. 50、SEQ ID No. 51、SEQ ID No. 52、SEQ ID No. 53、SEQ ID No. 54、SEQ ID No. 55、SEQ ID No. 56、SEQ ID No. 57、SEQ ID No. 58、SEQ ID No. 59、SEQ ID No. 60、SEQ ID No. 61、SEQ ID No. 62 或 SEQ ID No. 63 中所示的任一序列具有 80%或更高, 优选为 85%或更高, 更优选为 90%或更高, 且更优选为 95%或更高同一性。

在一个实施方案中, 用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列为与 SEQ ID No. 49、SEQ ID No. 50、SEQ ID No. 51、SEQ ID No. 62 和 SEQ ID No. 63 中所示的任一序列具有 70%或更高, 优选为 75%或更高同一性的核苷酸序列。适合地, 所述核苷酸序列可以与 SEQ ID No. 49、SEQ ID No. 50、SEQ ID No. 51、SEQ ID No. 62 和 SEQ ID No. 63 中所示的任一序列具有 80%或更高, 优选为 85%或更高, 更优选为 90%或更高, 且更优选为 95%或更高同一性。

在一个实施方案中, 用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列为与 SEQ ID No. 49 中所示的序列具有 70%或更高, 75%或更高, 80%或更高, 优选为 85%或更高, 更优选为 90%或更高, 且更优选为 95%或更高同一性的核苷酸序列。

适合地, 用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列可以编码包含一种或多种以下氨基酸序列的脂

酰基转移酶：

- (i) SEQ ID No. 3 所示的氨基酸序列、
- (ii) SEQ ID No. 4 所示的氨基酸序列、
- (iii) SEQ ID No. 5 所示的氨基酸序列、
- (iv) SEQ ID No. 6 所示的氨基酸序列、
- (v) SEQ ID No. 7 所示的氨基酸序列、
- (vi) SEQ ID No. 8 所示的氨基酸序列、
- (vii) SEQ ID No. 9 所示的氨基酸序列、
- (viii) SEQ ID No. 10 所示的氨基酸序列、
- (ix) SEQ ID No. 11 所示的氨基酸序列、
- (x) SEQ ID No. 12 所示的氨基酸序列、
- (xi) SEQ ID No. 13 所示的氨基酸序列、
- (xii) SEQ ID No. 14 所示的氨基酸序列、
- (xiii) SEQ ID No. 1 所示的氨基酸序列、
- (xiv) SEQ ID No. 15 所示的氨基酸序列或

与 SEQ ID No. 1、SEQ ID No. 3、SEQ ID No. 4、SEQ ID No. 5、SEQ ID No. 6、SEQ ID No. 7、SEQ ID No. 8、SEQ ID No. 9、SEQ ID No. 10、SEQ ID No. 11、SEQ ID No. 12、SEQ ID No. 13、SEQ ID No. 14 或 SEQ ID No. 15 中所示的任一序列具有 75%、80%、85%、90%、95%、98%或更高同一性的氨基酸序列。

适合地，用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列所编码的脂酰基转移酶可以包含如 SEQ ID No. 3 或 SEQ ID No. 4 或 SEQ ID No. 1 或 SEQ ID No. 15 所示的氨基酸序列，或包含与 SEQ ID No. 3 所示的氨基酸序列或 SEQ ID No. 4 所示的氨基酸序列或 SEQ ID No. 1 所示的氨基酸序列或 SEQ ID No. 15 所示的氨基酸序列具有 75%或更高、优选为 80%或更高、优选为 85%或更高、优选为 90%或更高、优选为 95%或更高的同一性的氨基酸序列。

适合地，用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列所编码的脂酰基转移酶可以包含与 SEQ ID No. 3、SEQ ID No. 4、SEQ ID No. 5、SEQ ID No. 6、SEQ ID No. 7、SEQ ID No.

8、SEQ ID No. 9、SEQ ID No. 10、SEQ ID No. 11、SEQ ID No. 12、SEQ ID No. 13、SEQ ID No. 14、SEQ ID No. 1 或 SEQ ID No. 15 中所示的任一序列具有 80%或更高、优选为 85%或更高、更优选为 90%或更高、且更优选为 95%或更高的同一性的氨基酸序列。

适合地，用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列可以编码包含一种或多种以下氨基酸序列的脂酰基转移酶：

(a) SEQ ID No. 3 或 SEQ ID No. 1 的氨基酸残基 1-100 所示的氨基酸序列；

(b) SEQ ID No. 3 或 SEQ ID No. 1 的氨基酸残基 101-200 所示的氨基酸序列；

(c) SEQ ID No. 3 或 SEQ ID No. 1 的氨基酸残基 201-300 所示的氨基酸序列；或

(d) 与以上 (a) 至 (c) 中所定义的氨基酸序列中任一序列具有 75%或更高、优选为 85%或更高、更优选为 90%或更高、且更优选为 95%或更高的同一性的氨基酸序列。

适合地，用于本发明方法和/或应用中的脂酰基转移酶可以包含一种或多种以下氨基酸序列：

(a) SEQ ID No. 3 或 SEQ ID No. 1 的氨基酸残基 28-39 所示的氨基酸序列；

(b) SEQ ID No. 3 或 SEQ ID No. 1 的氨基酸残基 77-88 所示的氨基酸序列；

(c) SEQ ID No. 3 或 SEQ ID No. 1 的氨基酸残基 126-136 所示的氨基酸序列；

(d) SEQ ID No. 3 或 SEQ ID No. 1 的氨基酸残基 163-175 所示的氨基酸序列；

(e) SEQ ID No. 3 或 SEQ ID No. 1 的氨基酸残基 304-311 所示的氨基酸序列；或

(f) 与以上 (a) 至 (e) 中所定义的氨基酸序列中任一序列具有 75%或更高、优选为 85%或更高、更优选为 90%或更高、且更优选为 95%或更

高的同一性的氨基酸序列。

一方面，用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列所编码的脂酰基转移酶可以是如 EP 1 275 711 中教导的来自于近平滑假丝酵母 (*Candida parapsilosis*) 的脂酰基转移酶。因此，一方面，用于本发明方法和应用中的脂酰基转移酶可以为包含 SEQ ID No. 17 或 SEQ ID No. 18 所教导的一种氨基酸序列的脂酰基转移酶。

优选地，用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列所编码的脂酰基转移酶可以是包含 SEQ ID No. 16 所示的氨基酸序列的脂酰基转移酶，或者是包含与 SEQ ID No. 16 具有 75%或更高、优选为 85%或更高、更优选为 90%或更高、更优选为 95%或更高、更优选为 98%或更高、或更优选为 99%或更高的同一性的氨基酸序列的脂酰基转移酶。所述酶可以被认为酶的变体。

一方面，用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列所编码的脂酰基转移酶可以为卵磷脂:胆固醇酰基转移酶 (LCAT) 或其变体 (如由分子进化产生的变体)。

适合的 LCAT 为本领域已知的，且可以获得来自一种或多种如下的生物体，如：哺乳类、大鼠、小鼠、鸡、黑腹果蝇 (*Drosophila melanogaster*)、植物 (包括拟南芥 (*Arabidopsis*) 和水稻 (*Oryza sativa*))、线虫、真菌和酵母。

在一个实施方案中，用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列所编码的脂酰基转移酶可以是可获得的，优选从带有 pPet12aAhydro 和 pPet12aASalmo 的大肠杆菌株 TOP 10 获得的脂酰基转移酶，所述大肠杆菌株 TOP 10 由 Danisco A/S of Langebrogade 1, DK-1001 Copenhagen K, Denmark 根据国际承认用于专利程序的微生物保藏布达佩斯条约保藏在 National Collection of Industrial, Marine and Food Bacteria (NCIMB) 23 St. Machar Street, Aberdeen Scotland, GB, 保藏日为 2003 年 12 月 22 日，保藏号分别为 NCIMB 41204 和 NCIMB 41205。

用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列可以编码磷脂甘油酰基转移酶。磷脂甘油酰基转移酶包



括分离自气单胞菌，优选为嗜水气单胞菌或杀鲑气单胞菌，最优选为杀鲑气单胞菌或其变体的那些磷脂甘油酰基转移酶。用于本发明的最优选的脂酰基转移酶可以由 SEQ ID No. 1、3、4、15 和 16 所编码。本领域技术人员可以理解，优选酰基转移酶的信号肽已经在转移酶的表达过程中被切割。SEQ ID No. 1、3、4、15 和 16 的信号肽为氨基酸 1-18。因此，最优选的区域为 SEQ ID No. 1 和 SEQ ID No. 3（嗜水气单胞菌）中的氨基酸 19-335，和 SEQ ID No. 4、SEQ ID No. 15 和 SEQ ID No. 16（杀鲑气单胞菌）中的氨基酸 19-336。当用于测定氨基酸序列的同一性或同源性时，优选使用成熟序列进行本文所述的比对。

因此，用于测定同源性（同一性）的最优选的区域为 SEQ ID No. 1 和 3（嗜水气单胞菌）中的氨基酸 19-335，和 SEQ ID No. 4、15 和 16（杀鲑气单胞菌）中的氨基酸 19-336。SEQ ID 34 和 35 分别是来自嗜水气单胞菌和杀鲑气单胞菌的脂酰基转移酶的成熟蛋白序列。

用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列所编码的脂酰基转移酶也可以分离自喜热裂孢菌（*Thermobifida*），优选为褐色喜热裂孢菌（*T. fusca*），最优选其由 SEQ ID No. 28 编码。

用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列所编码的核苷酸序列也可以分离自链霉菌属，优选为阿维链霉菌（*S. avermitis*），最优选其由 SEQ ID No. 32 编码。用于本发明的来自链霉菌的其它可能的酶包括由 SEQ ID No. 5、6、9、10、11、12、13、14、31 和 33 所编码的那些酶。

用于本发明的酶也可以分离自棒状杆菌属（*Corynebacterium*），优选为 *C. efficiens*，最优选其由 SEQ ID No. 29 编码。

适合地，用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列所编码的脂酰基转移酶可以包含 SEQ ID No. 37、38、40、41、43、45 或 47 所示的任一氨基酸序列，或与其具有至少 70%、75%、80%、85%、90%、95%、96%、97% 或 98% 同一性的氨基酸序列，或由 SEQ ID No. 36、39、42、44、46 或 48 所示的任一核苷酸序列所编码，或由与其具有至少 70%、75%、80%、85%、90%、95%、96%、97%

或 98%同一性的核苷酸序列所编码。

在一个实施方案中，用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列选自由以下组成的组：

a) 包含 SEQ ID No. 36 所示的核苷酸序列的核酸；

b) 由于遗传密码简并性而与 SEQ ID No. 36 的核苷酸序列相关的核酸；  
和

c) 包含与 SEQ ID No. 36 所示的核苷酸序列具有至少 70%同一性的核苷酸序列的核酸。

在一个实施方案中，用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列所编码的脂酰基转移酶包含 SEQ ID No. 37 所示的氨基酸序列或与其具有至少 60%同一性的氨基酸序列。

在另一实施方案中，用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列所编码的脂酰基转移酶可以包含 SEQ ID No. 37、38、40、41、43、45 或 47 所示的任一氨基酸序列，或与其具有至少 70%、75%、80%、85%、90%、95%、96%、97%或 98%同一性的氨基酸序列，或由 SEQ ID No. 39、42、44、46 或 48 所示的任一核苷酸序列所编码，或由与其具有至少 70%、75%、80%、85%、90%、95%、96%、97%或 98%同一性的核苷酸序列所编码。

在另一实施方案中，用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列所编码的脂酰基转移酶可以包含 SEQ ID No. 38、40、41、45 或 47 所示的任一氨基酸序列，或与其具有至少 70%、75%、80%、85%、90%、95%、96%、97%或 98%同一性的氨基酸序列，以用于本发明。

在另一实施方案中，用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列所编码的脂酰基转移酶可以包含 SEQ ID No. 38、40 或 47 所示的任一氨基酸序列，或与其具有至少 70%、75%、80%、85%、90%、95%、96%、97%或 98%同一性的氨基酸序列，以用于本发明。

更优选地，在一个实施方案中，用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列所编码的脂酰基转移酶

可以包含 SEQ ID No. 47 所示的氨基酸序列，或与其具有至少 70%、75%、80%、85%、90%、95%、96%、97%或 98%同一性的氨基酸序列。

在另一实施方案中，用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列所编码的脂酰基转移酶可以包含 SEQ ID No. 43 或 44 所示的氨基酸序列，或与其具有至少 80%、85%、90%、95%、96%、97%或 98%同一性的氨基酸序列。

在另一实施方案中，用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列所编码的脂酰基转移酶可以包含 SEQ ID No. 41 所示的氨基酸序列，或与其具有至少 70%、75%、80%、85%、90%、95%、96%、97%或 98%同一性的氨基酸序列。

在一个实施方案中，用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列选自由以下组成的组：

- a) 包含 SEQ ID No. 36 所示的核苷酸序列的核酸；
- b) 由于遗传密码简并性而与 SEQ ID No. 36 的核苷酸序列相关的核酸；

和

- c) 包含与 SEQ ID No. 36 所示的核苷酸序列具有至少 70%同一性的核苷酸序列的核酸。

在一个实施方案中，根据本发明的脂酰基转移酶可以是可获得的，优选从链霉菌属株 L130 或 L131 获得的脂酰基转移酶，所述链霉菌属株 L130 或 L131 由 Danisco A/S of Langebrogade 1, DK-1001 Copenhagen K, Denmark 根据国际承认用于专利程序的微生物保藏布达佩斯条约保藏在 National Collection of Industrial, Marine and Food Bacteria (NCIMB) 23 St. Machar Street, Aberdeen Scotland, GB，保藏日为 2004 年 6 月 25 日，保藏号分别为 NCIMB 41226 和 NCIMB 41227。

用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的适合的核苷酸序列可以编码脂酰基转移酶 (SEQ ID No. 16)，或可以编码脂酰基转移酶的氨基酸序列 (SEQ ID No. 17)。

用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的适合核苷酸序列可编码如下所述的氨基酸序列，该氨基酸序列可使用 Align X (使用默认设置的 VectorNTI 的 Clustal W 成对比对算法) 通过

与 L131 (SEQ ID No. 37) 序列的比对进行鉴定。

L131 与来自灰色链霉菌 (*S. avermitilis*) 和褐色喜热裂孢菌的同源物的比对可以显示 GDSx 基序 (L131 以及灰色链霉菌和褐色喜热裂孢菌中为 GDSY)、GANDY 框 (其为 GGND A 或 GGND L) 和 HPT 部分 (被认为是保守的催化性组氨酸) 的保守性。这三个保守部分突出标记在图 42 中。

与 pfam Pfam00657 共有序列 (如 WO04/064987 中所述) 和/或本文公开的 L131 序列 (SEQ ID No. 37) 比对时, 可以鉴定出三个保守区, GDSx 部分、GANDY 部分和 HTP 部分 (详述参见 WO04/064987)。

与 pfam Pfam00657 共有序列 (如 WO04/064987 中所述) 和/或本文公开的 L131 序列 (SEQ ID No. 37) 比对时,

i) 用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列可以编码具有 GDSx 基序, 更优选为选自 GDSL 或 GDSY 基序的 GDSx 基序的脂酰基转移酶; 和/或

ii) 用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列可以编码具有 GANDY 部分, 更优选为包含氨基 GGNDx 的 GANDY 部分, 更优选为 GGND A 或 GGND L 的脂酰基转移酶; 和/或

iii) 用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列可以编码优选具有 HTP 部分的脂酰基转移酶; 和优选地,

iv) 用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列可以编码优选具有 GDSx 或 GDSY 基序, 和包含氨基 GGNDx (优选为 GGND A 或 GGND L) 的 GANDY 部分, 和 HTP 部分 (保守的组氨酸) 的脂酰基转移酶。

### **脂酰基转移酶变体**

在优选的实施方案中, 用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列所编码的脂酰基转移酶为脂酰基转移酶变体。

可以使用对磷脂的活性增加的变体, 如对磷脂的水解活性增加和/或转

移酶活性增加的变体，优选为对磷脂的转移酶活性增加的变体。

优选地，通过如上定义的脂酰基转移酶的一个或多个氨基酸修饰来制备所述脂酰基转移酶变体。

适合地，当用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列所编码的脂酰基转移酶为脂酰基转移酶变体时，在此情况中，所述酶的特征为所述酶包含氨基酸序列基序 GDSX，其中 X 是下列氨基酸残基中的一种或多种：L、A、V、I、F、Y、H、Q、T、N、M 或 S，且其中所述酶变体相对于亲本序列在第 2 组、第 4 组、第 6 组或第 7 组限定的一个或多个氨基酸残基处包含一个或多个氨基酸修饰(如 WO2005/066347 和下文所定义)。

例如，脂酰基转移酶变体可以被表征为所述酶包含氨基酸序列基序 GDSX，其中 X 是下列氨基酸残基中的一种或多种：L、A、V、I、F、Y、H、Q、T、N、M 或 S，且其中酶变体相对于亲本序列在第 2 组、第 4 组、第 6 组或第 7 组限定的一个或多个氨基酸残基处包含一个或多个氨基酸修饰(如 WO2005/066347 和下文所定义)，所述组是通过所述亲本序列与本文定义的 P10480 结构模型在结构上进行比对而鉴定出来的，其优选是通过 P10480 晶体结构等同物(crystal structure coordinates)与如 WO2005/066347 和下文所定义的 1IVN.PDB 和/或 1DEO.PDB 的结构比对而获得的。

在其它实施方案中，用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列可以编码脂酰基转移酶变体，其可以被表征为所述酶包含氨基酸序列基序 GDSX，其中 X 是下列氨基酸残基中的一种或多种：L、A、V、I、F、Y、H、Q、T、N、M 或 S，且其中酶变体相对于亲本序列在第 2 组所教导的一个或多个氨基酸残基处包含一个或多个氨基酸修饰，所述第 2 组是在将所述亲本序列与 pfam 共有序列(SEQ ID No. 2-图 3) 比对时被鉴定出来的，并根据 P10480 的结构模型进行修饰以确保如 WO2005/066347 和下文所定义的最佳符合交叠(best fit overlap)。

适合地，当用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列可以编码可包含以下氨基酸序列的脂酰基转移酶变体，其中所述氨基酸序列如 SEQ ID No. 34、SEQ ID No. 3、SEQ ID No. 4、SEQ ID No. 5、SEQ ID No. 6、SEQ ID No. 7、SEQ ID No. 8、SEQ ID

No. 19、SEQ ID No. 10、SEQ ID No. 11、SEQ ID No. 12、SEQ ID No. 13、SEQ ID No. 14、SEQ ID No. 1、SEQ ID No. 15、SEQ ID No. 25、SEQ ID No. 26、SEQ ID No. 27、SEQ ID No. 28、SEQ ID No. 29、SEQ ID No. 30、SEQ ID No. 32 或 SEQ ID No. 33 所示，并在第 2 组、第 4 组、第 6 组或第 7 组限定的一个或多个氨基酸残基处有一个或多个氨基酸修饰（如 WO2005/066347 和下文所定义），所述组是通过与 SEQ ID No. 34 的序列比对而鉴定出来的。

或者，编码脂酰基转移酶的核苷酸序列可以编码包含以下氨基酸序列的脂酰基转移酶变体，其中所述氨基酸序列如 SEQ ID No. 34、SEQ ID No. 3、SEQ ID No. 4、SEQ ID No. 5、SEQ ID No. 6、SEQ ID No. 7、SEQ ID No. 8、SEQ ID No. 19、SEQ ID No. 10、SEQ ID No. 11、SEQ ID No. 12、SEQ ID No. 13、SEQ ID No. 14、SEQ ID No. 1、SEQ ID No. 15、SEQ ID No. 25、SEQ ID No. 26、SEQ ID No. 27、SEQ ID No. 28、SEQ ID No. 29、SEQ ID No. 30、SEQ ID No. 32 或 SEQ ID No. 33 所示，并在第 2 组、第 4 组、第 6 组或第 7 组限定的一个或多个氨基酸残基处有一个或多个氨基酸修饰（如 WO2005/066347 和下文所定义），所述组是通过所述亲本序列与本文定义的 P10480 的结构模型在结构上进行比对而鉴定出来的，其优选是通过 P10480 晶体结构等同物与如 WO2005/066347 和下文所教导的 1IVN.PDB 和/或 1IDEO.PDB 的结构比对而获得的。

或者，编码脂酰基转移酶的核苷酸序列可以编码包含以下氨基酸序列的脂酰基转移酶变体，其中所述氨基酸序列如 SEQ ID No. 34、SEQ ID No. 3、SEQ ID No. 4、SEQ ID No. 5、SEQ ID No. 6、SEQ ID No. 7、SEQ ID No. 8、SEQ ID No. 19、SEQ ID No. 10、SEQ ID No. 11、SEQ ID No. 12、SEQ ID No. 13、SEQ ID No. 14、SEQ ID No. 1、SEQ ID No. 15、SEQ ID No. 25、SEQ ID No. 26、SEQ ID No. 27、SEQ ID No. 28、SEQ ID No. 29、SEQ ID No. 30、SEQ ID No. 32 或 SEQ ID No. 33 所示，并在第 2 组所教导的一个或多个氨基酸残基处有一个或多个氨基酸修饰，所述第 2 组是在将所述亲本序列与 pfam 共有序列（SEQ ID No. 2）比对时被鉴定出来的，并根据 P10480 的结构模型进行修饰以确保如 WO2005/066347 和下文所教导的最佳符合交叠。

优选地，所述亲本酶是包含 SEQ ID No. 34 和/或 SEQ ID No. 15 和/或 SEQ ID No. 35 所示氨基酸序列或与它们同源的酶。

优选地，编码脂酰基转移酶的核苷酸序列可以编码包含以下氨基酸序列的脂酰基转移酶变体，其中所述氨基酸序列如 SEQ ID No. 34 或 SEQ ID No. 35 所示，并在如 WO2005/066347 和下文所定义的第 2 组、第 4 组、第 6 组或第 7 组限定的一个或多个氨基酸残基处有一个或多个氨基酸修饰。

### 组的定义

#### 第 1 组氨基酸：

第 1 组氨基酸（注意这些是图 53 和图 54 的 1IVN 中的氨基酸）

Gly8、Asp9、Ser10、Leu11、Ser12、Tyr15、Gly44、Asp45、Thr46、Glu69、Leu70、Gly71、Gly72、Asn73、Asp74、Gly75、Leu76、Gln106、Ile107、Arg108、Leu109、Pro110、Tyr113、Phe121、Phe139、Phe140、Met141、Tyr145、Met151、Asp154、His157、Gly155、Ile156、Pro158

从第 1 组中不包含高度保守的基序，如 GDSx 和催化残基（标有下划线的残基）。为了避免争议，第 1 组定义了 1IVN 模型活性位点中在甘油的中心碳原子 10 Å 以内的氨基酸残基。

#### 第 2 组氨基酸：

第 2 组氨基酸（注意氨基酸的编号是指 P10480 成熟序列中的氨基酸）

Leu17、Lys22、Met23、Gly40、Asn80、Pro81、Lys82、Asn87、Asn88、Trp111、Val112、Ala114、Tyr117、Leu118、Pro156、Gly159、Gln160、Asn161、Pro162、Ser163、Ala164、Arg165、Ser166、Gln167、Lys168、Val169、Val170、Glu171、Ala172、Tyr179、His180、Asn181、Met209、Leu210、Arg211、Asn215、Lys284、Met285、Gln289 和 Val290。

第 1 组和第 2 组中所选残基的比较表

IVN	IVN 模型 嗜水气单胞菌同源物		P10480 成熟序列中的残基编号
	PFAM	结构	
Gly8	Gly32		
Asp9	Asp33		
Ser10	Ser34		
Leu11	Leu35		Leu17
Ser12	Ser36		Ser18
			Lys22
			Met23
Tyr15	Gly58		Gly40
Gly44	Asn98		Asn80
Asp45	Pro99		Pro81
Thr46	Lys100		Lys82
			Asn87
			Asn88
Glu69	Trp129		Trp111
Leu70	Val130		Val112
Gly71	Gly131		
Gly72	Ala132		Ala114
Asn73	Asn133		
Asp74	Asp134		
Gly75	Tyr135		Tyr117
Leu76	Leu136		Leu118
Gln106		Pro174	Pro156
Ile107		Gly177	Gly159
Arg108		Gln178	Gln160
Leu109		Asn179	Asn161
Pro110		180-190	Pro162
Tyr113			Ser163
			Ala164
			Arg165
			Ser166
			Gln167
			Lys168
			Val169
			Val170
			Glu171
			Ala172
Phe121	His198	Tyr197	Tyr179
		His198	His180
		Asn199	Asn181
Phe139	Met227		Met209



Phe140	Leu228		Leu210
Met141	Arg229		Arg211
Tyr145	Asn233		Asn215
			Lys284
Met151	Met303		Met285
Asp154	Asp306		
Gly155	Gln307		Gln289
Ile156	Val308		Val290
His157	His309		
Pro158	Pro310		

### 第 3 组氨基酸：

第三组氨基酸与第 2 组相同，但是是指杀鲑气单胞菌（SEQ ID No.4）的编码序列，即第 3 组中的氨基酸残基编号要大 18，这反映了成熟蛋白（SEQ ID No. 34）与包含信号序列的蛋白（SEQ ID No. 25）中氨基酸编号之间的差异。

杀鲑气单胞菌 GDSX（SEQ ID No.4）和嗜水气单胞菌 GDSX（SEQ ID No.34）的成熟蛋白有五个氨基酸不同。它们是 Thr3Ser、Gln182Lys、Glu309Ala、Ser310Asn 和 Gly318-，其中杀鲑气单胞菌的残基列在前面，而嗜水气单胞菌的残基列在后面。嗜水气单胞菌蛋白的长度只有 317 个氨基酸，且在第 318 位上缺少残基。与嗜水气单胞菌蛋白相比，杀鲑气单胞菌 GDSX 对极性脂质（如半乳糖脂底物）具有相当高的活性。对全部五个氨基酸位置进行了位点扫描。

### 第 4 组氨基酸：

第 4 组氨基酸是 S3、Q182、E309、S310 和-318。

### 第 5 组氨基酸：

F13S、D15N、S18G、S18V、Y30F、D116N、D116E、D157 N、Y226F、D228N、Y230F。

### 第 6 组氨基酸：

第 6 组氨基酸是 Ser3、Leu17、Lys22、Met23、Gly40、Asn80、Pro81、

Lys82、Asn87、Asn88、Trp111、Val112、Ala114、Tyr117、Leu118、Pro156、Gly159、Gln160、Asn161、Pro162、Ser163、Ala164、Arg165、Ser166、Gln167、Lys168、Val169、Val170、Glu171、Ala172、Tyr179、His180、Asn181、Gln182、Met209、Leu210、Arg211、Asn215、Lys284、Met285、Gln289、Val290、Glu309、Ser310、-318。

第6组中的氨基酸的编号是指 P10480 (SEQ ID No. 25) 中的氨基酸残基, 其它序列主链上的相应氨基酸可以通过与 P10480 和/或 1IVN 的同源比对和/或结构比对来确定。

#### 第7组氨基酸:

第7组氨基酸是 Ser3、Leu17、Lys22、Met23、Gly40、Asn80、Pro81、Lys82、Asn87、Asn88、Trp111、Val112、Ala114、Tyr117、Leu118、Pro156、Gly159、Gln160、Asn161、Pro162、Ser163、Ala164、Arg165、Ser166、Gln167、Lys168、Val169、Val170、Glu171、Ala172、Tyr179、His180、Asn181、Gln182、Met209、Leu210、Arg211、Asn215、Lys284、Met285、Gln289、Val290、Glu309、Ser310、-318、Y30X (其中 X 选自 A、C、D、E、G、H、I、K、L、M、N、P、Q、R、S、T、V 或 W)、Y226X (其中 X 选自 A、C、D、E、G、H、I、K、L、M、N、P、Q、R、S、T、V 或 W)、Y230X (其中 X 选自 A、C、D、E、G、H、I、K、L、M、N、P、Q、R、S、T、V 或 W)、S18X (其中 X 选自 A、C、D、E、F、H、I、K、L、M、N、P、Q、R、T、W 或 Y)、D157X (其中 X 选自 A、C、E、F、G、H、I、K、L、M、P、Q、R、S、T、V、W 或 Y)。

第7组中氨基酸的编号是指 P10480 (SEQ ID No. 25) 中的氨基酸残基, 其它序列主链上的相应氨基酸可以通过与 P10480 和/或 1IVN 的同源比对和/或结构比对来确定。

适合地, 与亲本酶相比, 所述酶变体包含一种或多种以下氨基酸修饰:

S3E、A、G、K、M、Y、R、P、N、T 或 G

E309Q、R 或 A, 优选为 Q 或 R

-318Y、H、S 或 Y, 优选为 Y。

优选地, GDSX 基序的 X 是 L。因此, 优选所述亲本酶包含氨基酸基

序 GDSL。

适合地，所述第一亲本脂酰基转移酶可以包含任何一种以下氨基酸序列：SEQ ID No. 34、SEQ ID No. 3、SEQ ID No. 4、SEQ ID No. 5、SEQ ID No. 6、SEQ ID No. 7、SEQ ID No. 8、SEQ ID No. 19、SEQ ID No. 10、SEQ ID No. 11、SEQ ID No. 12、SEQ ID No. 13、SEQ ID No. 14、SEQ ID No. 1、SEQ ID No. 15、SEQ ID No. 25、SEQ ID No. 26、SEQ ID No. 27、SEQ ID No. 28、SEQ ID No. 29、SEQ ID No. 30、SEQ ID No. 32 或 SEQ ID No. 33。

适合地，所述第二相关脂酰基转移酶可以包含任何一种以下氨基酸序列：SEQ ID No. 3、SEQ ID No. 34、SEQ ID No. 4、SEQ ID No. 5、SEQ ID No. 6、SEQ ID No. 7、SEQ ID No. 8、SEQ ID No. 19、SEQ ID No. 10、SEQ ID No. 11、SEQ ID No. 12、SEQ ID No. 13、SEQ ID No. 14、SEQ ID No. 1、SEQ ID No. 15、SEQ ID No. 25、SEQ ID No. 26、SEQ ID No. 27、SEQ ID No. 28、SEQ ID No. 29、SEQ ID No. 30、SEQ ID No. 32 或 SEQ ID No. 33。

与所述亲本酶相比，所述酶变体必须包含至少一个氨基酸修饰。在一些实施方案中，与亲本酶相比，所述酶变体可以包含至少 2 个、优选为至少 3 个、优选为至少 4 个、优选为至少 5 个、优选为至少 6 个、优选为至少 7 个、优选为至少 8 个、优选为至少 9 个、优选为至少 10 个氨基酸修饰。

当在本文中提及具体的氨基酸残基时，可以从序列变体与 SEQ ID No. 34 或 SEQ ID No. 35 所示的参考序列的比对获得编号。

一方面，优选酶变体包含一种或多种以下氨基酸取代：

S3A、C、D、E、F、G、H、I、K、L、M、N、P、Q、R、T、V、W 或 Y；和/或

L17A、C、D、E、F、G、H、I、K、M、N、P、Q、R、S、T、V、W 或 Y；和/或

S18A、C、D、E、F、H、I、K、L、M、N、P、Q、R、T、W 或 Y；和/或

K22A、C、D、E、F、G、H、I、L、M、N、P、Q、R、S、T、V、W 或 Y；和/或

M23A、C、D、E、F、G、H、I、K、L、N、P、Q、R、S、T、V、W 或 Y；和/或

Y30A、C、D、E、G、H、I、K、L、M、N、P、Q、R、S、T、V 或 W；和/或

G40A、C、D、E、F、H、I、K、L、M、N、P、Q、R、S、T、V、W 或 Y；和/或

N80A、C、D、E、F、G、H、I、K、L、M、P、Q、R、S、T、V、W 或 Y；和/或

P81A、C、D、E、F、G、H、I、K、L、M、N、Q、R、S、T、V、W 或 Y；和/或

K82A、C、D、E、F、G、H、I、L、M、N、P、Q、R、S、T、V、W 或 Y；和/或

N87A、C、D、E、F、G、H、I、K、L、M、P、Q、R、S、T、V、W 或 Y；和/或

N88A、C、D、E、F、G、H、I、K、L、M、P、Q、R、S、T、V、W 或 Y；和/或

W111A、C、D、E、F、G、H、I、K、L、M、N、P、Q、R、S、T、V、W 或 Y；和/或

V112A、C、D、E、F、G、H、I、K、L、M、N、P、Q、R、S、T、W 或 Y；和/或

A114C、D、E、F、G、H、I、K、L、M、N、P、Q、R、S、T、V、W 或 Y；和/或

Y117A、C、D、E、F、G、H、I、K、L、M、N、P、Q、R、S、T、V 或 W；和/或

L118A、C、D、E、F、G、H、I、K、M、N、P、Q、R、S、T、V、W 或 Y；和/或

P156A、C、D、E、F、G、H、I、K、L、M、N、Q、R、S、T、V、W 或 Y；和/或

D157A、C、E、F、G、H、I、K、L、M、P、Q、R、S、T、V、W 或 Y；和/或

G159A、C、D、E、F、H、I、K、L、M、N、P、Q、R、S、T、V、W 或 Y；和/或

Q160A、C、D、E、F、G、H、I、K、L、M、N、P、R、S、T、V、  
W 或 Y；和/或

N161A、C、D、E、F、G、H、I、K、L、M、P、Q、R、S、T、V、  
W 或 Y；和/或

P162A、C、D、E、F、G、H、I、K、L、M、N、Q、R、S、T、V、  
W 或 Y；和/或

S163A、C、D、E、F、G、H、I、K、L、M、N、P、Q、R、T、V、  
W 或 Y；和/或

A164C、D、E、F、G、H、I、K、L、M、N、P、Q、R、S、T、V、  
W 或 Y；和/或

R165A、C、D、E、F、G、H、I、K、L、M、N、P、Q、S、T、V、  
W 或 Y；和/或

S166A、C、D、E、F、G、H、I、K、L、M、N、P、Q、R、T、V、  
W 或 Y；和/或

Q167A、C、D、E、F、G、H、I、K、L、M、N、P、R、S、T、V、  
W 或 Y；和/或

K168A、C、D、E、F、G、H、I、L、M、N、P、Q、R、S、T、V、  
W 或 Y；和/或

V169A、C、D、E、F、G、H、I、K、L、M、N、P、Q、R、S、T、  
W 或 Y；和/或

V170A、C、D、E、F、G、H、I、K、L、M、N、P、Q、R、S、T、  
W 或 Y；和/或

E171A、C、D、F、G、H、I、K、L、M、N、P、Q、R、S、T、V、  
W 或 Y；和/或

A172C、D、E、F、G、H、I、K、L、M、N、P、Q、R、S、T、V、  
W 或 Y；和/或

Y179A、C、D、E、F、G、H、I、K、L、M、N、P、Q、R、S、T、  
V 或 W；和/或

H180A、C、D、E、F、G、I、K、L、M、P、Q、R、S、T、V、W 或  
Y；和/或

N181A、C、D、E、F、G、H、I、K、L、M、P、Q、R、S、T、V、  
W 或 Y；和/或

Q182A、C、D、E、F、G、H、I、K、L、M、N、P、R、S、T、V、  
W 或 Y，优选为 K；和/或

M209A、C、D、E、F、G、H、I、K、L、N、P、Q、R、S、T、V、  
W 或 Y；和/或

L210A、C、D、E、F、G、H、I、K、M、N、P、Q、R、S、T、V、  
W 或 Y；和/或

R211A、C、D、E、F、G、H、I、K、L、M、N、P、Q、R、S、T、  
V、W 或 Y；和/或

N215A、C、D、E、F、G、H、I、K、L、M、N、P、Q、R、S、T、  
V、W 或 Y；和/或

Y226A、C、D、E、G、H、I、K、L、M、N、P、Q、R、S、T、V  
或 W；和/或

Y230A、C、D、E、G、H、I、K、L、M、N、P、Q、R、S、T、V  
或 W；和/或

K284A、C、D、E、F、G、H、I、L、M、N、P、Q、R、S、T、V、  
W 或 Y；和/或

M285A、C、D、E、F、G、H、I、K、L、N、P、Q、R、S、T、V、  
W 或 Y；和/或

Q289A、C、D、E、F、G、H、I、K、L、M、N、P、R、S、T、V、  
W 或 Y；和/或

V290A、C、D、E、F、G、H、I、K、L、M、N、P、Q、R、S、T、  
W 或 Y；和/或

E309A、C、D、F、G、H、I、K、L、M、N、P、Q、R、S、T、V、  
W 或 Y；和/或

S310A、C、D、E、F、G、H、I、K、L、M、N、P、Q、R、T、V、  
W 或 Y。

此外或可选地，可以有一个或多个 C 末端延长。优选地，所述附加的 C 末端延长由一个或多个脂肪族氨基酸构成，优选为非极性氨基酸，更优

选为 I、L、V 或 G。因此，本发明进一步提供了包含以下 C 末端延长中的一个或多个的酶变体：318I、318L、318V、318G。

优选的酶变体可以降低对磷脂如磷脂酰胆碱（PC）的水解活性，并可以升高对磷脂的转移酶活性。

优选的酶变体可以升高对磷脂如磷脂酰胆碱（PC）的转移酶活性，也可以升高对磷脂的水解活性。

一个或多个以下残基的修饰可以产生对磷脂具有升高的绝对转移酶活性的酶变体：S3、D157、S310、E309、Y179、N215、K22、Q289、M23、H180、M209、L210、R211、P81、V112、N80、L82、N88、N87。

可以提供对磷脂具有改进的转移酶活性的酶变体的优选具体修饰可以选自以下一种或多种：

S3A、C、D、E、F、G、H、I、K、L、M、N、P、Q、R、T、V、W 或 Y，优选为 N、E、K、R、A、P 或 M，最优选为 S3A；

D157A、C、E、F、G、H、I、K、L、M、N、P、Q、R、S、T、V、W 或 Y，优选为 D157S、R、E、N、G、T、V、Q、K 或 C；

S310A、C、D、E、F、G、H、I、K、L、M、N、P、Q、R、T、V、W 或 Y，优选为 S310T、-318E；

E309A、C、D、E、F、G、H、I、K、L、M、N、P、Q、R、T、V、W 或 Y，优选为 E309 R、E、L、R 或 A；

Y179A、C、D、E、F、G、H、I、K、L、M、N、P、Q、R、S、T、V 或 W，优选为 Y179D、T、E、R、N、V、K、Q 或 S，更优选为 E、R、N、V、K 或 Q；

N215A、C、D、E、F、G、H、I、K、L、M、P、Q、R、S、T、V、W 或 Y，优选为 N215 S、L、R 或 Y；

K22A、C、D、E、F、G、H、I、L、M、N、P、Q、R、S、T、V、W 或 Y，优选为 K22 E、R、C 或 A；

Q289A、C、D、E、F、G、H、I、K、L、M、N、P、R、S、T、V、W 或 Y，优选为 Q289R、E、G、P 或 N；

M23A、C、D、E、F、G、H、I、K、L、N、P、Q、R、S、T、V、W 或 Y，优选为 M23K、Q、L、G、T 或 S；

H180A、C、D、E、F、G、I、K、L、M、P、Q、R、S、T、V、W 或 Y，优选为 H180Q、R 或 K；

M209A、C、D、E、F、G、H、I、K、L、N、P、Q、R、S、T、V、W 或 Y，优选为 M209Q、S、R、A、N、Y、E、V 或 L；

L210A、C、D、E、F、G、H、I、K、M、N、P、Q、R、S、T、V、W 或 Y，优选为 L210R、A、V、S、T、I、W 或 M；

R211A、C、D、E、F、G、H、I、K、L、M、N、P、Q、S、T、V、W 或 Y，优选为 R211T；

P81A、C、D、E、F、G、H、I、K、L、M、N、Q、R、S、T、V、W 或 Y，优选为 P81G；

V112A、C、D、E、F、G、H、I、K、L、M、N、P、Q、R、S、T、W 或 Y，优选为 V112C；

N80A、C、D、E、F、G、H、I、K、L、M、P、Q、R、S、T、V、W 或 Y；优选为 N80R、G、N、D、P、T、E、V、A 或 G

L82A、C、D、E、F、G、H、I、M、N、P、Q、R、S、T、V、W 或 Y，优选为 L82N、S 或 E；

N88A、C、D、E、F、G、H、I、K、L、M、P、Q、R、S、T、V、W 或 Y，优选为 N88C；

N87A、C、D、E、F、G、H、I、K、L、M、P、Q、R、S、T、V、W 或 Y，优选为 N87M 或 G；

一个或多个以下残基的优选修饰产生对磷脂具有升高的绝对转移酶活性的酶变体：

S3 N、R、A、G

M23 K、Q、L、G、T、S

H180 R

L82 G

Y179 E、R、N、V、K 或 Q

E309 R、S、L 或 A

一个优选的修饰是 N80D。尤其当使用参考序列 SEQ ID No. 35 作为骨



架时更是如此。因此，参考序列可以为 SEQ ID No. 16。所述修饰可以与一个或多个其它修饰组合。因此，在本发明的优选实施方案中，用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列所编码的脂酰基转移酶可以包含 SEQ ID No. 35，或与 SEQ ID No. 35 具有 75%或更高、优选为 85%或更高、更优选为 90%或更高、更优选为 95%或更高、更优选为 98%或更高、或更优选为 99%或更高同一性的氨基酸序列。

如上指出，当在本文中提及具体氨基酸残基时，编号是通过序列变体与 SEQ ID No. 34 或 SEQ ID No. 35 所示的参考序列的比对获得的。

优选地，用于本发明宿主细胞、载体、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列可以编码包含 SEQ ID No. 16 所示的氨基酸序列的脂，或与 SEQ ID No. 16 具有 75%或更高、优选为 85%或更高、更优选为 90%或更高、更优选为 95%或更高、更优选为 98%或更高、或更优选为 99%或更高同一性的氨基酸序列的脂。所述酶可以被认为是酶变体。

出于本发明的目的，同一性的程度是基于相同序列元件（sequence elements）的数目。根据本发明，可以通过本领域已知的计算机程序的方法（如 Vector NTI 10（Invitrogen Corp.））来适当地确定氨基酸序列的同一性程度。对于成对比对，所用的评分优选为空位开放罚分为 10.0、空位延伸罚分为 0.1 的 BLOSUM62。

适合地，关于氨基酸序列的同一性程度，在至少 20 个连续的氨基酸内、优选为至少 30 个连续的氨基酸内、优选为至少 40 个连续的氨基酸内、优选为至少 50 个连续的氨基酸内、优选为至少 60 个连续的氨基酸内进行测定。

适合地，可以在全序列中测定氨基酸序列的同一性程度。

适合地，编码本发明脂酰基转移酶的核苷酸序列是可以获得的，优选从一种或多种以下属的生物体中获得：气单孢菌属、链霉菌属、酵母菌属、乳球菌属、分支杆菌属、链球菌属、乳杆菌属、脱亚硫酸菌属、芽孢杆菌、弯曲菌属、弧菌科、木杆菌属、硫化叶菌属、曲霉属、裂殖酵母属、李斯特菌属、奈瑟菌属、中慢生根瘤菌属、雷尔氏菌属、黄单胞菌属、念珠菌属、喜热裂孢菌属和棒状杆菌属。

适合地，编码本发明脂酰基转移酶的核苷酸序列是可以获得的，优选从一种或多种以下生物体中获得：嗜水气单胞菌、杀鲑气单胞菌、天蓝色链霉菌 (*Streptomyces coelicolor*)、龟裂链霉菌 (*Streptomyces rimosus*)、分支杆菌属、酿脓链球菌 (*Streptococcus pyogenes*)、乳酸乳球菌 (*Lactococcus lactis*)、酿脓链球菌、嗜热链球菌 (*Streptococcus thermophilus*)、*Streptomyces thermosacchari*、灰色链霉菌、瑞士乳杆菌 (*Lactobacillus helveticus*)、脱卤脱亚硫酸菌 (*Desulfitobacterium dehalogenans*)、芽孢杆菌、空肠弯曲杆菌 (*Campylobacter jejuni*)、弧菌科 (*Vibrionaceae*)、苛养木杆菌 (*Xylella fastidiosa*)、硫磺矿硫化叶菌 (*Sulfolobus solfataricus*)、酿酒酵母 (*Saccharomyces cerevisiae*)、土霉菌 (*Aspergillus terreus*)、粟酒裂殖酵母 (*Schizosaccharomyces pombe*)、无害李斯特菌 (*Listeria innocua*)、单核细胞增生李斯特菌 (*Listeria monocytogenes*)、脑膜炎奈瑟氏球菌 (*Neisseria meningitidis*)、百脉根中慢生根瘤菌 (*Mesorhizobium loti*)、茄科雷尔氏菌 (*Ralstonia solanacearum*)、野油菜黄单胞菌 (*Xanthomonas campestris*)、地毯草黄单胞菌 (*Xanthomonas axonopodis*)、近平滑假丝酵母、褐色喜热裂孢菌和 *Corynebacterium efficiens*。

一方面，优选用于本发明宿主细胞、载体、方法和/或应用中任一项的编码本发明脂酰基转移酶的核苷酸序列所编码的脂酰基转移酶是可以获得的，优选获得自或来自气单胞菌属、嗜水气单胞菌或杀鲑气单胞菌的一种或多种。

可以使用 WO2004/064537 的实施例 12 中教导的试验对本发明具有脂酰基转移酶功能的酶进行常规鉴定。所述试验使用很高的水含量(约 95%)，本发明的脂酰基转移酶是具有至少 2%的酰基转移酶活性(相对转移酶活性)，优选为至少 5%的相对转移酶活性，优选为至少 10%的相对转移酶活性，优选为至少 15%、20%、25%、26%、28%、30%、40%、50%、60%或 75%的相对转移酶活性。

磷脂酶在低水环境中可以起酰基转移酶的作用。因此，当在低水环境中进行脂肪食用油的修饰过程时，可以考虑使用磷脂酶替代磷脂酰基转移酶，或除磷脂酰基转移酶之外，可以同时使用磷脂酶。

本文所用的术语“高水”是指水含量大于 3%、优选为大于 4%、大于

5%、大于 6%、大于 7%、大于 8%、大于 9%、大于 10%、大于 20%、大于 30%、大于 40%、大于 50%、大于 60%、大于 70%、大于 80%或大于 90%的任何底物或食物。

本文所用的术语“低水”是指水含量水含量小于 3%、优选为小于 2%、小于 1%或小于 0.5%、小于 0.3%、小于 0.2、小于 0.1、小于 0.05 或小于 0.01%的任何底物或食物。

为了避免争议，牛奶为高水环境，而乳脂（butterfat）为低水环境。

用于本发明的适合的磷脂酶包括磷脂酶 A1、磷脂酶 A2 或磷脂酶 B。也可以将磷脂酶 A1、磷脂酶 A2 或磷脂酶 B 与脂酰基转移酶活性协同使用。也可以将磷脂酶 C 和/或 D 与脂酰基转移酶活性/磷脂酶 A1、A2 和/或 B 活性协同使用，这类似于 WO2005/089562。优选的磷脂酶可以包括磷脂酶 A2，如 Lecitase<sup>TM</sup> 或 WO2004/97012（Novozymes/Chr. Hansen）公开的镰孢霉（*Fusarium venenatum*）的磷脂酶和 *Tuber albidum* 的磷脂酶。Novozymes 销售的镰孢霉磷脂酶名为 MAX YIELD<sup>TM</sup>。

#### 分离的

一方面，本发明的方法包括回收/分离脂酰基转移酶的额外步骤。因此，所制备的脂酰基转移酶可以是分离的形式。

另一方面，用于本发明的编码脂酰基转移酶的核苷酸序列可以是分离的形式。

术语“分离的（或分离）”是指序列或蛋白至少基本不包含其它至少一种成分，所述成分在自然中与该序列或蛋白天然连接，并以这种连接状态发现于自然中。

#### 纯化的

一方面，本发明的方法包括对所述脂酰基转移酶进行纯化的额外步骤。

另一方面，用于本发明的编码脂酰基转移酶的核苷酸序列可以是纯化的形式。

术语“纯化的”是指序列处于相对纯的状态，如至少约 51%纯的、或至少约 75%纯的、或至少约 80%纯的、或至少约 90%纯的、或至少约 95%纯的、或至少约 98%纯的。

克隆编码本发明多肽的核苷酸序列

编码具有如本文所定义的具体性质的多肽或适合修饰的多肽的核苷酸序列可以从产生所述多肽的任何细胞或生物体中分离得到。用于核苷酸序列分离的各种方法都是本领域熟知的。

例如，可以使用产生所述多肽的生物体的染色体 DNA 或信使 RNA 来构建基因组 DNA 和/或 cDNA 文库。如果所述多肽的氨基酸序列是已知的，可以合成经标记的寡核苷酸探针，并将其用于从由该生物体制备的基因组文库中鉴定编码多肽的克隆。或者，也可以使用包含与另一已知多肽基因同源的序列的经标记寡核苷酸探针来鉴定编码多肽的克隆。在后一种情况下，使用严谨性较低的杂交和清洗条件。

或者，可以通过如下方式鉴定编码多肽的克隆：将基因组 DNA 的片段插入到表达载体（如质粒）中，使用所得的基因组 DNA 文库转化酶为阴性的细菌，并随后将转化的细菌涂布在含由能够被所述多肽抑制的酶的琼脂上，由此可以鉴定出表达多肽的克隆。

再者，也可以通过已建立的标准方法通过合成来制备编码所述多肽的核苷酸序列，如 Beaucage S.L.等(1981) Tetrahedron Letters 22, p 1859-1869 描述的亚磷酰胺法，或 Matthes 等(1984) EMBO J. 3, p 801-805 描述的方法。在亚磷酰胺法中，在如自动 DNA 合成仪上合成寡核苷酸，然后将其纯化、退火、连接并克隆到适当的载体中。

所述核苷酸序列可以是源自基因组片段和合成片段的连接物、源自合成片段和 cDNA 片段的连接物、或源自基因组片段和 cDNA 片段的连接物，其根据标准技术通过连接源自合成的、基因组的或 cDNA 的片段（根据需要）而制得。每个连接的片段对应于整个核苷酸序列的不同部分。所述 DNA 序列也可以使用特定引物通过聚合酶链式反应（PCR）来制备，如 US 4,683,202 或 Saiki R K 等(Science (1988) 239, pp 487-491)中所述。

核苷酸序列

本发明还涵盖编码具有如本文所定义的具体性质的多肽的核苷酸序列。本文所用的术语“核苷酸序列”是指寡核苷酸序列或多核苷酸序列和

其变体、同源物、片段和衍生物（如其部分）。所述核苷酸序列可以是源自基因组或合成物或重组物的，其可以是双链的或单链的（无论其代表正义链还是反义链）。

本发明的术语“核苷酸序列”包括基因组 DNA、cDNA、合成的 DNA 和 RNA。优选地，其指 DNA，更优选为编码序列的 cDNA。

在优选的实施方案中，编码具有如本文所定义的特定性质的多肽的核苷酸序列不涵盖在天然环境中存在的并与其天然相连序列（同处于其天然环境中）连接的天然核苷酸序列。为了便于参考，我们将此优选实施方案称为“非天然核苷酸序列”。由此，术语“天然核苷酸序列”是指与其天然相连的完整启动子（同处于其天然环境中）可操作地连接的完整核苷酸序列（同处于其天然环境中）。因此，本发明的多肽可以利用核苷酸序列在其天然生物体中表达，但是其中所述核苷酸序列不受所述生物体中与其天然结合的启动子的控制。

优选地，所述多肽不是天然多肽。由此，术语“天然多肽”是指处于其天然环境中并已经被其天然核苷酸序列表达的整个多肽。

通常，使用重组 DNA 技术制备编码具有如本文所定义的特定性质的多肽的核苷酸序列（即重组的 DNA）。然而，在本发明的另一个具体实施方案中，可以使用本领域已知的化学方法来合成全部或部分核苷酸序列（参见 Caruthers MH 等(1980) *Nuc Acids Res Symp Ser* 215-23 和 Horn T 等(1980) *Nuc Acids Res Symp Ser* 225-232）。

### 分子进化

分离出编码酶的核苷酸序列或鉴定出编码推定酶的核苷酸序列后，就可能需要对所选择的核苷酸序列进行修饰，例如需要对所述序列进行突变以制备本发明的酶。

可以使用合成的寡核苷酸来引入突变。这些寡核苷酸包含目标突变位点侧翼的核苷酸序列。

适合的方法公开在 Morinaga 等(*Biotechnology* (1984) 2, p646-649)。另一种向编码酶的核苷酸序列中引入突变的方法描述在 Nelson 和 Long (*Analytical Biochemistry* (1989), 180, p 147-151)。

除了如上文所述的定点诱变，可以随机引入突变，如使用商品试剂盒，如来自 Stratagene 的 GeneMorph PCR 诱变试剂盒或来自 Clontech 的 Diversify PCR 随机诱变试剂盒。EP 0 583 265 提到基于 PCR 的诱变的优化方法，其也可以结合使用 DNA 突变类似物，如 EP 0 866 796 中所述的那些。易错 PCR 技术也适用于制备具有优选的性质的脂酰基转移酶变体。WO0206457 提到了脂肪酶的分子进化。

获得新型序列的第三种方法是使用任何数量的限制性酶或如 Dnase I 的酶来将不同的核苷酸序列片段化，并重新组装成编码功能性蛋白的全长核苷酸序列。或者，可以使用一个或多个不同的核苷酸序列，并在重新组装全长核苷酸序列时引入突变。DNA 改组 (shuffling) 和家族改组技术适用于制备具有优选性质的脂酰基转移酶变体。适合进行“改组”的方法可以参见 EP0 752 008、EP1 138 763、EP1 103 606。改组也可以与如 US 6,180,406 和 WO 01/34835 中所述的其它形式的 DNA 诱变相结合。

因此，有可能在体内或在体外使核苷酸序列中产生大量定点突变或随机突变，并随后通过多种手段筛选功能性得到改进的编码多肽。使用例如 *in silico* 和 *exo* 介导的重组方法 (参见 WO 00/58517、US 6,344,328、US 6,361,974)，可以进行分子进化，其中所产生的变体保留了与已知酶或蛋白的很低同源性。由此获得的所述变体可与已知的转移酶具有显著的结构类似性，但却具有很低的氨基酸序列同源性。

此外，如在非限定性实施例中，多核苷酸序列的突变或天然变体也可以与野生型或其它突变或天然变体重组以生产新的变体。也可以筛选所述新的变体以获得功能性得到改进的编码多肽。

应用以上以及类似的分子进化方法能够在没有关于蛋白结构或功能的任何现有知识的情况下鉴定和选择具有优选特性的本发明的酶变体，并能够产生不可预测的但有益的突变或变体。在本领域中应用分子进化来优化或改变酶活性的例子有很多，所述例子包括但不限于以下的一种或多种：优化在宿主细胞或体外的表达和/或活性、增加酶活性、改变底物和/或产物特异性、增加或降低酶稳定性或结构稳定性、改变在优选环境条件 (如温度、pH、底物) 中酶的活性/特异性。

使用分子进化工具可以对酶进行改变以提高该酶的功能性，这对于本

领域的技术人员来说是显而易见的。

适合地，用于本发明的编码脂酰基转移酶的核苷酸序列可以编码脂酰基转移酶变体，即当与亲本酶比较时，所述脂酰基转移酶可以包含至少一个氨基酸的取代、缺失或加入。变体酶与亲本酶保持至少 1%、2%、3%、5%、10%、15%、20%、30%、40%、50%、60%、70%、80%、90%、95%、97%、99%的同源性。适合的亲本酶可以包括具有酯酶或脂肪酶活性的任何酶。优选地，亲本酶对应于 pfam00657 共有序列。

在优选的实施方案中，脂酰基转移酶变体保留或掺入了在 GDSx、GANDY 和 HPT 部分中发现的至少一个或多个 pfam00657 共有序列氨基酸残基。

可以使用分子进化工具使酶（如在含水环境中没有或具有低脂酰基转移酶活性的脂肪酶）突变，以引入或增强转移酶活性，有此生产适合用于本发明的组合物和方法的具有显著转移酶活性的脂酰基转移酶。

适合地，用于本发明载体、宿主细胞、方法和/或应用中任一项的编码脂酰基转移酶的核苷酸序列所编码的脂酰基转移酶可以是变体，与亲本酶相比，该变体对极性脂质（优选为磷脂和/或糖脂）具有增强的酶活性。优选地，所述变体还可以对可溶性极性脂质（lyso polar lipids）具有低活性或无活性。对极性脂质、磷脂和/或糖脂的增强的活性可能是由于水解和/或转移酶活性或二者组合的结果。

与亲本酶相比，所述脂酰基转移酶变体对甘油三酯和/或甘油单酯和/或甘油二酯的活性可以降低。

适合地，所述酶变体对甘油三酯、和/或甘油单酯和/或甘油二酯可以没有活性。

或者，所述酶变体对甘油三酯的活性可以增加，和/或对以下的一种或多种的活性也可以增加：极性脂质、磷脂、卵磷脂、磷脂酰胆碱、糖脂、二半乳糖基甘油单酯、单半乳糖基甘油单酯。

脂酰基转移酶变体是已知的，且一种或多种所述变体可以适用于本发明的方法和应用，和/或本发明的酶组合物。仅举例来说，根据本发明可以使用以下文献中阐述的脂酰基转移酶变体：Hilton & Buckley *J Biol. Chem.* 1991 Jan 15; 266 (2): 997-1000; Robertson 等 *J. Biol. Chem.* 1994 Jan 21;

269(3):2146-50; Brumlik 等 J. Bacteriol 1996 Apr; 178 (7): 2060-4; Peelman 等 Protein Sci. 1998 Mar; 7(3):587-99。

### 氨基酸序列

本发明还涵盖由用于本发明载体、宿主细胞、方法和/或应用中任一项目的编码脂酰基转移酶的核苷酸序列所编码的氨基酸序列。

本文所使用的术语“氨基酸序列”与术语“多肽”和/或术语“蛋白”同义。在一些实例中，术语“氨基酸序列”与术语“肽”同义。

所述氨基酸序列可以从适合的来源制备/分离，或其可以通过合成制备、或其可以使用重组 DNA 技术制备。

适合地，所述氨基酸序列可以通过标准技术从本文教导的分离多肽中获得。

一种适合测定分离多肽的氨基酸序列的方法如下：

可以将纯化的多肽冻干，并将 100  $\mu\text{g}$  的冻干原料溶解于 8 M 尿素和 0.4 M 碳酸氢铵的 50  $\mu\text{l}$  混合物 (pH 8.4) 中。在覆盖氮气并加入 5  $\mu\text{l}$  45 mM 二硫苏糖醇后，可以将溶解的蛋白在 50°C 变性并还原 15 分钟。冷却至室温后，可以加入 5  $\mu\text{l}$  100 mM 碘乙酰胺，从而使半胱氨酸残基在室温、避光和氮气下衍生 15 分钟。

可以向以上反应混合物中加入 135  $\mu\text{l}$  水和含 5  $\mu\text{g}$  内切蛋白酶 Lys-C 的 5  $\mu\text{l}$  水溶液，并在氮气下在 37°C 进行 24 小时消化。

可以使用溶剂 A (0.1%TFA 的水溶液) 和溶剂 B (0.1%TFA 的乙腈溶液) 在 VYDAC C18 柱 (0.46x15cm; 10 $\mu\text{m}$ ; The Separation Group, California, USA) 上通过反向 HPLC 分离所得到的肽。在 N 末端测序前，可以使用相同的溶剂体系在 Develosil C18 柱上对所选择的肽进行重新层析。可以使用 Applied Biosystems 476A 测序仪，根据生产商的说明书 (Applied Biosystems, California, USA) 使用脉冲液态快速循环来完成测序。

### 序列同一性或序列同源性

在此，术语“同源物”是指与主题氨基酸序列和主题核苷酸序列具有一定同源性的实体。在此，术语“同源性”可以等同于“同一性”。



所述同源氨基酸序列和/或核苷酸序列可以提供和/或编码保留所述酶的功能活性和/或增强所述酶的活性的多肽。

在本文中，认为同源序列包括可以与主题序列有至少 75、85 或 90% 同一性，优选为至少 95 或 98% 同一性的氨基酸序列。通常，同源物将包含与主题氨基酸序列相同的活性位点等。虽然同源性也可以被看作为相似性（即氨基酸残基具有相似的化学性质/功能），在本发明的内容中，在序列同一性方面优选同源性来表达。

在本文中，认为同源序列包括可以与编码本发明多肽的核苷酸序列（主题序列）有至少 75、85 或 90% 同一性的核苷酸序列，优选为至少 95 或 98% 同一性的核苷酸序列。通常，同源物将包括与主题序列相同的活性位点编码序列等。虽然同源性也可以被看作为相似性（即氨基酸残基具有相似的化学性质/功能），在本发明的内容中，在序列同一性方面优选同源性来表达。

可以通过目测进行同源性比较，或者更通常是借助易于获得的序列比较程序进行同源性比较。这些商业性计算机程序可以计算两个或多个序列间的同源性%。

可以在连续的序列中计算同源性%，即一个序列与其它序列比对，且将一个序列中的每个氨基酸与其它序列中的相应氨基酸直接比较，每次一个残基。这种称为“无缺口”比对。通常所述无缺口比对仅在数量相对少的残基中进行。

虽然这是非常简单且可靠的方法，但是其未能考虑到如在其它方面相同的成对序列中，一个插入或缺失将引起随后的氨基酸残基无法比对，因此可能在整体比对时导致同源性%的大大降低。因此，大部分序列比对方法被设计成在考虑可能的插入和缺失的情况下产生最佳的比对，而不过度地处罚整体同源性的得分。这可以通过在序列中插入“缺口”以设法将局部同源性最大化而实现的。

然而，这些更加复杂的方法给比对中出现的每个缺口赋予“缺口罚分”，使得对于相同数目的相同氨基酸，具有尽量少缺口的序列比对（反映了两个比较的序列间更高的相关性）将比具有更多缺口的序列比对具有更高的得分。通常使用“仿射缺口代价”（Affine gap costs），即对缺口的存在处以

较高代价，而对缺口中的每个后续残基处以较小罚分。这是最常用的缺口评分系统。高缺口罚分将无疑将产生具有更少缺口的优化比对。大多数比对程序允许修正缺口罚分。然而，当使用所述软件进行序列比较时，优选使用默认值。

因此，最大同源性%的计算首先需要在考虑缺口罚分下产生最佳的比对。适合用于进行所述比对的计算机程序为 Vector NTI (Invitrogen Corp.)。可以进行序列比较的其它软件的实例包括但不限于 BLAST 软件包 (参见 Ausubel 等 1999 Short Protocols in Molecular Biology, 4<sup>th</sup> Ed, Chapter 18) 和 FASTA (Altschul 等 1990 J. Mol. Biol. 403-410)。BLAST 和 FASTA 均可进行离线和在线搜索 (参见 Ausubel 等 1999, 7-58 页至 7-60 页)。然而，对于一些应用，优选使用 Vector NTI 程序。也可以将称为 BLAST 2 Sequences 的新型工具用于比较蛋白和核苷酸序列 (参见 FEMS Microbiol Lett 1999 174(2): 247-50; FEMS Microbiol Lett 1999 177(1): 187-8 和 [tatiana@ncbi.nlm.nih.gov](mailto:tatiana@ncbi.nlm.nih.gov))。

虽然根据同一性来测定最终的同源性%，但是比对方法本身通常不是基于全是或全非成对比较。而是，通常使用大规模相似性评分矩阵 (scaled similarity score matrix)，基于化学相似性或进化距离对每对比较评分。通常所用的此矩阵的实例为 BLOSUM62 矩阵 (BLAST 程序套的默认矩阵)。Vector NTI 程序通常使用公开的默认值，或者也可能提供的自定义的符号比较表 (详见用户手册)。对于一些应用，优选使用 Vector NTI 软件包的默认值。

或者，也可以使用基于与 CLUSTAL (Higgins DG & Sharp PM (1988), Gene 73(1), 237-244) 类似的算法的 Vector NTI (Invitrogen Corp.) 中的多重比对特征来计算同源性%。

一旦软件生成了最佳比对，就可以计算同源性%，优选为序列同一性%。软件通常将其作为序列比较的一部分来执行，并生成数值结果。

如果在测定序列同一性时使用缺口罚分，随后优选使用以下参数进行成对比对：

BLAST	
缺口开口	0
缺口延伸	0

CLUSTAL	DNA	蛋白	
字长 (WORD SIZE)	2	1	K 三联体
缺口罚分	15	10	
缺口延伸	6.66	0.1	

在一个实施方案中，优选使用具有以上定义的缺口罚分和缺口延伸值的 CLUSTAL 来测定核苷酸序列的序列同一性。

适合地，在至少 20 个连续的核苷酸中、优选为在至少 30 个连续的核苷酸中、优选为在至少 40 个连续的核苷酸中、优选为在至少 50 个连续的核苷酸中、优选为在至少 60 个连续的核苷酸中、优选为在至少 100 连续的核苷酸中测定核苷酸序列的同一性程度。

适合地，在全序列中测定核苷酸序列中的同一性程度。

在一个实施方案中，本发明的氨基酸序列的同一性程度可以通过本领域已知的计算机程序的方法，如 Vector NTI 10 (Invitrogen Corp.) 来适当地测定。对于成对比对，所用的矩阵优选为缺口开口罚分为 10.0 且缺口延伸罚分为 0.1 的 BLOSUM62。

适合地，在至少 20 个连续的氨基酸中、优选为在至少 30 个连续的氨基酸中、优选为在至少 40 个连续的氨基酸中、优选为在至少 50 个连续的氨基酸中、优选为在至少 60 个连续的氨基酸中测定氨基酸序列的同一性程度。

适合地，在全序列中测定氨基酸序列的同一性程度。

所述序列也可以具有产生沉默改变和产生功能等同物的氨基酸残基的缺失、插入或取代。可以根据残基在极性、电荷、溶解性、疏水性、亲水性和/或两亲性质上的相似性来进行谨慎的氨基酸取代，只要保持该物质的次要结合活性即可。例如，带负电荷的氨基酸包括天冬氨酸和谷氨酸；带正电荷的氨基酸包括赖氨酸和精氨酸；具有相似亲水性值的不带电荷的极性头基团 (polar head groups) 的氨基酸包括亮氨酸、异亮氨酸、缬氨酸、

甘氨酸、丙氨酸、天冬酰胺、谷酰胺、丝氨酸、苏氨酸、苯丙氨酸和酪氨酸。

例如可以根据下表进行保守取代。在第二列中相同栏中的氨基酸，优选为在第三列中相同行中的氨基酸可以相互取代：

脂肪族	非极性	G、A、P
		I、L、V
	极性-不带电荷	C、S、T、M
		N、Q
	极性-带电荷	D、E
		K、R
芳香族		H、F、W、Y

本发明还涵盖可以发生的同源取代（本文所用的取代和替换均指现存氨基酸残基与备选残基的互换），即相似取代，如碱性氨基酸取代碱性氨基酸、酸性氨基酸取代酸性氨基酸、极性氨基酸取代极性氨基酸等。也可以发生非同源取代，即从一类残基变成另一类残基，或涉及非天然氨基酸，如鸟氨酸（下文称为 Z）、二氨基丁酸鸟氨酸（下文称为 B）、正亮氨酸鸟氨酸（下文称为 O）、吡啶丙氨酸、噻吩丙氨酸、萘丙氨酸和苯甘氨酸。

也可以使用非天然氨基酸进行替换。

氨基酸变体序列可以包含可以在序列的任何两个氨基酸残基间插入的适合的间隔子基团，除氨基酸间隔子如甘氨酸或  $\beta$ -丙氨酸残基外，还包括烷基基团如甲基、乙基或丙基基团。本领域的技术人员将理解其它形式的变异，其涉及以类肽（peptoid）形式存在的一个或多个氨基酸残基。为了避免争议，“类肽形式”用来指  $\alpha$ -碳取代基基团在残基的氮原子上而非在  $\alpha$ -碳上的氨基酸残基变体。制备类肽形式的肽的方法是本领域内已知的，如 Simon RJ 等，PNAS (1992) 89(20), 9367-9371 和 Horwell DC, Trends Biotechnol. (1995) 13(4), 132-134。

本发明所使用的或编码具有本文定义的特定性质的多肽的核苷酸序列可以在其内部包含合成的或修饰的核苷酸。本领域中已知对寡核苷酸的许多不同类型的修饰。这些修饰包括甲基磷酸酯和硫代磷酸酯主链和/或在分子的 3'末端和/或 5'末端添加吡啶或多赖氨酸链。出于本发明的目的，应该

理解可以用本领域中的任何方法来修饰本文所述的核苷酸序列。可以进行所述修饰以增强核苷酸序列的体内活性或寿命。

本发明还涵盖与本文所讨论的序列或其任何衍生物、片段或衍生物互补的核苷酸序列的应用。如果序列与其片段互补，此序列可以被用作鉴定其它生物体中相似编码序列等的探针。

可以以多种方式获得与本发明的序列并非 100%同源但却落入本发明范围中的多核苷酸。可以通过例如探查由一系列个体（如来自不同种群的个体）制备的 DNA 文库来获得本文所述序列的其它变体。此外，可以获得其它病毒/细菌或细胞的同源物，特别是发现于哺乳类细胞（如大鼠、小鼠、牛和灵长类细胞）中的细胞同源物，且所述同源物或其片段将通常能够选择性地与本文序列表中所示的序列杂交。可以通过探查从其它动物物种制备的 cDNA 文库或基因组 DNA 文库，且在中度至高度严谨条件使用包含所附序列表中任何一个序列的全部或部分的探针来探查所述文库来获得所述序列。类似的考虑用于获得本发明多肽或核苷酸序列的物种同源物和等位基因变体。

也可以使用简并 PCR 来获得变体和株系/物种同源物，所述简并 PCR 将使用设计为靶向变体和同源物中编码本发明序列中的保守氨基酸序列的序列。例如可以通过比对来自多个变体/同源物的氨基酸序列来预测保守序列。可以使用本领域已知的计算机软件进行序列比对。例如广泛使用 GCG Wisconsin PileUp 程序。

简并 PCR 中所用的引物将包含一个或多个简并位点，且其使用条件的严谨性将低于通过针对已知序列的单一序列引物克隆序列的那些条件。

或者，也可以通过已表征序列的定点诱变来获得所述多核苷酸。例如当需要在表达多核苷酸序列的特定宿主细胞中改变沉默密码子序列以优化密码子偏爱性时，这可能是有用的。可能需要其它序列改变，以引入限制性多肽识别位点，或改变多核苷酸编码的多肽的性质或功能。

可以使用本发明的多核苷酸（核苷酸序列）来制备引物，如 PCR 引物，用于可选扩增反应的引物；探针，如使用放射性或非放射性标记物通过常规方法用显现标记物标记的探针；或可以将多核苷酸克隆到载体中。所述引物、探针和其它片段在长度上将为至少 15 个、优选为至少 20 个、如至

少 25 个、30 个或 40 个核苷酸，且其也被如本文所用的本发明的术语多核苷酸所涵盖。

可以重组、合成或通过本领域技术人员可利用的任何方法来制备根据本发明的多核苷酸（如 DNA 多核苷酸）和探针。它们也可以通过标准技术进行克隆。

通常，将通过合成方法来制备引物，该方法包括以每次一个核苷酸的方式逐步制备所需要的核酸序列。使用自动化技术来实现这一点的技术是本领域中易于获得的。

通常使用重组方法，如使用 PCR（聚合酶链式反应）克隆技术来制备更长的多核苷酸。这包括制备所需要克隆的脂质靶向序列区域侧翼的一对引物（如约 15 至 30 个核苷酸），使引物接触从动物或人细胞获得的 mRNA 或 cDNA，在可以使所需区域扩增的条件下进行聚合酶链式反应，分离扩增的片段（如通过在琼脂糖凝胶上纯化反应混合物），并回收扩增的 DNA。可以设计引入使其包含适合的限制性酶识别位点，以便可以将扩增的 DNA 克隆到适合的克隆载体中。

## 杂交

本发明还涵盖与本发明的序列互补的序列，或能够与本发明的序列杂交或与其互补序列杂交的序列。

本文所用的术语“杂交”包括“一条核酸链通过碱基配对与互补链结合的过程”，以及在聚合酶链式反应（PCR）技术中进行扩增的过程。

本发明还涵盖所述核苷酸序列的应用，所述核苷酸序列能够和与本文所讨论的主题序列、或其任何衍生物、片段或衍生物互补的序列杂交。

本发明还涵盖能够与本文所讨论的核苷酸序列杂交的序列的互补序列。

杂交条件基于核苷酸结合复合物的熔解温度( $T_m$ )，如 Berger 和 Kimmel (1987, Guide to Molecular Cloning Techniques, Methods in Enzymology, Vol. 152, Academic Press, San Diego CA)中所教导，且赋予了下文所解释的定义的“严谨性”。

最高严谨性通常出现在约  $T_m-5^\circ\text{C}$ （比探针的  $T_m$  低  $5^\circ\text{C}$ ）；高严谨性出

现在比  $T_m$  低约  $5^{\circ}\text{C}$  至  $10^{\circ}\text{C}$ ；中等严谨性出现在比  $T_m$  低约  $10^{\circ}\text{C}$  至  $20^{\circ}\text{C}$ ；且低严谨性出现在比  $T_m$  低约  $20^{\circ}\text{C}$  至  $25^{\circ}\text{C}$ 。正如本领域技术人员所理解的是，最高严谨性杂交可以用于鉴定或检测相同的核苷酸序列，而中等（或低）严谨性杂交可以用于鉴定或检测相似或相关的多核苷酸序列。

优选地，本发明涵盖能够在高严谨性条件或中等严谨性条件下与编码具有如本文定义的特定性质的多肽的核苷酸序列杂交的序列所互补的序列。

更优选地，本发明涵盖能够在高严谨性条件（如  $65^{\circ}\text{C}$  和  $0.1\times\text{SSC}$  { $1\times\text{SSC} = 0.15\text{ M NaCl}$ 、 $0.015\text{ M}$  柠檬酸钠， $\text{pH } 7.0$ }）下与编码具有如本文定义的特定性质的多肽的核苷酸序列发生杂交的序列的互补序列。

本发明还涉及能够与本文讨论的核苷酸序列（包括本文讨论的那些序列的互补序列）杂交的核苷酸序列。

本发明还涉及能够与本文讨论的核苷酸序列（包括本文讨论的那些序列的互补序列）杂交的序列的互补核苷酸序列。

本发明的范围还包括在中等至最高严谨性条件下能够与本文所讨论的核苷酸序列杂交的多核苷酸序列。

在优选的方面，本发明涵盖在严谨性条件（如  $50^{\circ}\text{C}$  和  $0.2\times\text{SSC}$ ）下能够与本文所讨论的核苷酸序列或其互补序列杂交的核苷酸序列。

在更优选的方面，本发明涵盖在高严谨性条件（如  $65^{\circ}\text{C}$  和  $0.1\times\text{SSC}$ ）下能够与本文所讨论的核苷酸序列或其互补序列杂交的核苷酸序列。

### 多肽的表达

可以将用于本发明的或用于编码具有如本文所定义的特定性质的多肽的核苷酸序列引入重组的复制型载体中。载体可以用于在相容的宿主细胞中和/或从相容的宿主细胞复制并以多肽形式表达核苷酸序列。可以使用包含启动子/增强子和其它表达调控信号在内的控制序列来控制表达。可以使用原核生物的启动子和在真核细胞中有功能的启动子。可以使用组织特异的或刺激特异性启动子。也可以使用包含来自上述两个或更多不同启动子的序列元件的嵌合启动子。

根据所用的序列和/或载体，由宿主重组细胞通过表达核苷酸序列所产

生的多肽可以被分泌，或被包含在细胞内。编码序列经过设计可以具有信号序列，该信号序列可指导物质的编码序列分泌通过特定的原核生物或真核细胞膜。

### 构建体

术语“构建体”，与术语如“结合物(或连接物)”、“盒(cassette)”和“杂合体(hybrid)”同义，其包含依照本发明使用的编码具有本文所定义的特定性质的多肽的核苷酸序列，且其直接或间接地与启动子连接。间接连接的实例是在启动子和本发明的核苷酸序列之间提供适合的间隔基如内含子序列，如 Sh1 内含子或 ADH 内含子。术语“融合”在本文中也是如此，其包括直接或间接连接。在一些情况中，这些术语不涵盖编码所述蛋白的核苷酸序列与其通常连接的野生型基因启动子的天然组合，并且此时这两者均处于其天然环境中。

所述构建体甚至可以包含或表达允许选择基因构建体的标记物。

对于一些应用，优选构建体至少包含本发明的核苷酸序列，或编码具有如本文定义的特定性质的多肽且可操作地与启动子连接的核苷酸序列。

### 生物体

与本发明有关的术语“生物体”包括任何包含本发明的核苷酸序列或编码具有如本文定义的特定性质的多肽的核苷酸序列和/或由其获得的产物的生物体。

与本发明有关的术语“转基因生物体”包括任何包含编码具有如本文定义的特定性质的多肽的核苷酸序列和/或由其获得的产物的生物体，和/或其中启动子能够允许编码具有如本文定义的特定性质的多肽的核苷酸序列在所述生物体内表达。优选核苷酸序列被引入到生物体的基因组中。

术语“转基因生物体”不涵盖在处于自身天然环境中并同时受到其同处于自身天然环境中的天然启动子控制的天然核苷酸编码序列。

因此，本发明的转基因生物体包括包含以下任何一种或组合的生物体：编码具有如本文定义的特定性质的多肽的核苷酸序列、本文定义的构建体、本文定义的载体、本文定义的质粒、本文定的细胞或其产物。例如，所述



转基因生物体还可以包含编码具有如本文定义的特定性质的多肽且受到启动子控制的核苷酸序列，其中所述启动子与脂酰基转移酶编码基因不是天然相连。

#### 宿主细胞/生物体的转化

原核生物体宿主的转化为本领域所公知，例如参见 Sambrook 等 (Molecular Cloning: A Laboratory Manual, 2nd edition, 1989, Cold Spring Harbor Laboratory Press)。如果使用原核生物体宿主，则在转化前需要通过如去除内含子来适当地修饰核苷酸序列。

已知芽孢杆菌物种转化的各种方法。

#### 分泌

通常，期望多肽被从表达宿主分泌到培养基中，由此可以更容易地回收酶。根据本发明，可以基于所需的表达宿主来选择分泌前导序列。本发明中也可以使用杂交信号序列。

不与天然的编码脂酰基转移酶的核苷酸序列连接的分泌前导序列的典型实例为那些源于真菌淀粉葡萄糖苷酶 (AG) 基因 (glaA-18 个氨基酸和 24 个氨基酸两种形式，如来自曲霉菌属)、 $\alpha$ -因子基因 (酵母，如酵母属 (saccharomyces)、克鲁维酵母 (*Kluyveromyces*) 和汉逊酵母 (*Hansenula*)) 或  $\alpha$ -淀粉酶基因 (芽孢杆菌) 的序列。

#### 检测

本领域已知多种用于检测和测定氨基酸序列表达的方法。实例包括酶联免疫吸附分析 (ELISA)、放射免疫分析 (RIA) 和荧光活化的细胞分选 (FACS)。

本领域技术人员已知多种标记物和偶联技术，且它们可以被用在各种核酸和氨基酸分析中。

许多公司如 Pharmacia Biotech (Piscataway, NJ)、Promega (Madison, WI) 和 US Biochemical Corp (Cleveland, OH) 为这些步骤提供商业试剂盒和方法。

适合的报告分子或标记物包括那些放射性核素、酶、荧光剂、化学发

光剂或显色剂，以及底物、辅助因子、抑制剂、磁性粒子等。教导这些标记物应用的专利包括 US-A-3,817,837、US-A-3,850,752、US-A-3,939,350、US-A-3,996,345、US-A-4,277,437、US-A-4,275,149 和 US-A-4,366,241。

此外，可以如 US-A-4,816,567 中所示制备重组免疫球蛋白。

### 融合蛋白

在本发明的方法中，脂酰基转移酶可以作为融合蛋白来制备，如以便有助于其提取和纯化。融合蛋白配偶体（partners）的实例包括谷胱甘肽-S-转移酶（GST）、6xHis、GAL4（DNA 结合和/或转录激活结构域）和  $\beta$ -半乳糖苷酶。在融合蛋白配偶体和目的蛋白序列间也可以包含蛋白水解切割位点以便于去除融合蛋白序列。优选地，融合蛋白将不会妨碍蛋白序列的活性。

Curr. Opin. Biotechnol. (1995) 6(5):501-6 已经对大肠杆菌中的基因融合表达系统进行了综述。

在本发明的另一实施方案中，具有如本文定义的特性的多肽的氨基酸序列可以与非天然序列连接以编码融合蛋白。例如，对于为获得能够影响实质活性的试剂而进行的肽库筛选，编码这样的嵌合物可能是有用的，即它表达可被商购抗体所识别的非天然表位。

以下将参照附图和实施例仅以举例的方式描述本发明。

### 附图说明

图 1 显示了杀鲑气单胞菌成熟脂酰基转移酶突变体（GCAT）的氨基酸序列，该突变体具有 Asn80Asp 突变（注意，第 80 位氨基酸指在成熟序列中）（SEQ ID 16）；

图 2 显示了来自嗜水气单胞菌（ATCC #7965）的脂酰基转移酶的氨基酸序列（SEQ ID No. 1）；

图 3 显示了来自第 6 版数据库的 pfam00657 共有序列（SEQ ID No. 2）；

图 4 显示了从生物体嗜水气单胞菌获得的氨基酸序列（SEQ ID No. 3）（P10480；GI:121051）；

图 5 显示了从生物体杀鲑气单胞菌获得的氨基酸序列（SEQ ID No. 4）

(AAG098404; GI:9964017);

图 6 显示了从生物体天蓝色链霉菌 A3(2)获得的氨基酸序列 (SEQ ID No. 5) (Genbank 登录号 NP\_631558);

图 7 显示了从生物体天蓝色链霉菌 A3(2)获得的氨基酸序列 (SEQ ID No. 6) (Genbank 登录号 CAC42140);

图 8 显示了从生物体酿酒酵母获得的氨基酸序列 (SEQ ID No. 7) (Genbank 登录号 P41734);

图 9 显示了从生物体雷尔氏菌属获得的氨基酸序列 (SEQ ID No. 8) (Genbank 登录号 AL646052);

图 10 显示了 SEQ ID No. 9。Scoe1 为 NCBI 蛋白登录号 CAB39707.1 GI:4539178 的保守的假定蛋白[天蓝色链霉菌 A3(2)];

图 11 显示了如 SEQ ID No. 10 所示的氨基酸。Scoe2 为 NCBI 蛋白登录号 CAC01477.1 GI:9716139 的保守的假定蛋白[天蓝色链霉菌 A3(2)];

图 12 显示了氨基酸序列 (SEQ ID No. 11)。Scoe3 为 NCBI 蛋白登录号 CAB88833.1 GI:7635996 的公知分泌性蛋白[天蓝色链霉菌 A3(2)];

图 13 显示了氨基酸序列 (SEQ ID No. 12)。Scoe4 为 NCBI 蛋白登录号 CAB89450.1 GI:7672261 的公知分泌性蛋白[天蓝色链霉菌 A3(2)];

图 14 显示了氨基酸序列 (SEQ ID No. 13)。Scoe5 为 NCBI 蛋白登录号 CAB62724.1 GI:6562793 的公知脂蛋白[天蓝色链霉菌 A3(2)];

图 15 显示了氨基酸序列 (SEQ ID No. 14)。Srim1 为 NCBI 蛋白登录号 AAK84028.1 GI:15082088 的 GDSSL-脂肪酶[龟裂链霉菌];

图 16 显示了来自杀鲑气单胞菌杀鲑亚种(*Aeromonas salmonicida subsp. Salmonicida*) (ATCC#14174) 的脂酰基转移酶的氨基酸序列 (SEQ ID No. 15);

图 17 显示了 SEQ ID No. 19。Scoe1 为 NCBI 蛋白登录号 CAB39707.1 GI:4539178 的保守的假定蛋白[天蓝色链霉菌 A3(2)];

图 18 显示了用于嗜水气单胞菌脂酰基转移酶基因诱变的融合构建体的氨基酸序列 (SEQ ID No. 25)。下划线的氨基酸为木聚糖酶信号肽;

图 19 显示了来自链霉菌属的脂酰基转移酶的多肽序列 (SEQ ID No. 26);

图 20 显示了来自喜热裂孢菌属的脂酰基转移酶的多肽序列 (SEQ ID No. 27);

图 21 显示了来自喜热裂孢菌属的脂酰基转移酶的多肽序列 (SEQ ID No. 28);

图 22 显示了来自 *Corynebacterium efficiens* GDSx 300 氨基酸的脂酰基转移酶的多肽序列 (SEQ ID No. 29);

图 23 显示了来自 *Novosphingobium aromaticivorans* GDSx 284 氨基酸的脂酰基转移酶的多肽 (SEQ ID No. 30);

图 24 显示了来自天蓝色链霉菌 GDSx 269 氨基酸的脂酰基转移酶的多肽 (SEQ ID No. 31);

图 25 显示了来自灰色链霉菌/GDSx 269 氨基酸的脂酰基转移酶的多肽 (SEQ ID No. 32);

图 26 显示了来自链霉菌属的脂酰基转移酶的多肽 (SEQ ID No. 33);

图 27 显示了从生物体嗜水气单胞菌获得的氨基酸序列 (SEQ ID No. 34) (P10480; GI:121051) (注意, 其为成熟序列);

图 28 显示了杀鲑气单胞菌成熟脂酰基转移酶突变体 (GCAT) 的氨基酸序列 (SEQ ID 35) (注意, 其为成熟序列);

图 29 显示了来自 *Streptomyces thermosacchari* 的核苷酸序列 (SEQ ID No. 36);

图 30 显示了来自 *Streptomyces thermosacchari* 的核苷酸序列 (SEQ ID No. 37);

图 31 显示了来自褐色喜热裂孢菌/GDSx 548 氨基酸的氨基酸序列 (SEQ ID No. 38);

图 32 显示了来自褐色喜热裂孢菌的核苷酸序列 (SEQ ID No. 39);

图 33 显示了来自褐色喜热裂孢菌/GDSx 的氨基酸序列 (SEQ ID No. 40);

图 34 显示了来自 *Corynebacterium efficiens*/GDSx 300 氨基酸的氨基酸序列 (SEQ ID No. 41);

图 35 显示了来自 *Corynebacterium efficiens* 的核苷酸序列 (SEQ ID No. 42);

图 36 显示了来自天蓝色链霉菌/GDSx 268 氨基酸的氨基酸序列 (SEQ ID No. 43);

图 37 显示了来自天蓝色链霉菌的核苷酸序列 (SEQ ID No. 44);

图 38 显示了来自灰色链霉菌的氨基酸序列 (SEQ ID No. 45);

图 39 显示了来自灰色链霉菌的核苷酸序列 (SEQ ID No. 46);

图 40 显示了来自褐色喜热裂孢菌/GDSx 的氨基酸序列 (SEQ ID No. 47);

图 41 显示了来自褐色喜热裂孢菌/GDSx 的核苷酸序列 (SEQ ID No. 48);

图 42 显示了 L131 和来自灰色链霉菌和褐色喜热裂孢菌的同源物的比对,说明了 GDSx 基序 (在 L131 以及灰色链霉菌和褐色喜热裂孢菌中均为 GDSY)、GANDY 框 (其为 GGND A 或 GGND L) 和 HPT 部分 (被认为是保守的催化组氨酸) 的保守性。这三个保守部分有突出标记;

图 43 显示了来自近平滑假丝酵母脂酰基转移酶的氨基酸序列 (SEQ ID No. 17);

图 44 显示了来自近平滑假丝酵母脂酰基转移酶的氨基酸序列 (SEQ ID No. 18);

图 45 显示了活性位点中具有甘油的 1IVN.PDB 晶体结构的带状显示图。此图使用 Deep View Swiss-PDB viewer 制作。

图 46 显示了活性位点中具有甘油的 1IVN.PDB 晶体结构的侧视图 (使用 Deep View Swiss-PDB viewer), 黑色部分为在活性位点甘油的 10 Å 内的残基。

图 47 显示了活性位点中具有甘油的 1IVN.PDB 晶体结构的俯视图 (使用 Deep View Swiss-PDB viewer), 黑色部分为在活性位点甘油的 10 Å 内的残基。

图 48 显示了比对 1;

图 49 显示了比对 2;

图 50 和 51 显示了 1IVN 与 P10480 的比对 (P10480 为嗜水气单胞菌酶数据库序列), 此比对从 PFAM 数据库获得并用在模型构建过程中; 和

图 52 显示了比对, 其中 P10480 为嗜水气单胞菌数据库序列。此序列

用于模型构建和位点选择。注意绘制出了完整的蛋白 (SEQ ID No. 25), 成熟蛋白 (等价于 SEQ ID No. 34) 起始于第 19 位残基。A. sal 为杀鲑气单胞菌的 GDSX 脂肪酶 (SEQ ID No. 4), A. hyd 为嗜水气单胞菌的 GDSX 脂肪酶 (SEQ ID No. 34)。所列序列间的差异位置在共有序列中用\*表示。

图 53 显示了用在实施例 1 中的基因构建体;

图 54 显示了用在实施例 1 中的密码子优化的基因构建体(no. 052907);  
和

图 55 显示了包含 LAT-KLM3'前体基因的 XhoI 插入物的序列, 下划线处为-35 和-10 框;

图 56 显示了在三丁酸甘油酯琼脂上于 37°C 生长 48 小时后的 BML780-KLM3'CAP50 (包含 SEQ ID No. 16-上面的菌落) 和 BML780 (空宿主菌株-下面的菌落)。

图 57 显示了来自杀鲑气单胞菌的包含信号序列 (前 LAT-从第 1 位至第 87 位) 的核苷酸序列 (SEQ ID No. 49);

图 58 显示了从生物体嗜水气单胞菌获得的本发明的编码脂酰基转移酶的核苷酸序列 (SEQ ID No. 50);

图 59 显示了从生物体杀鲑气单胞菌获得的本发明的编码脂酰基转移酶的核苷酸序列 (SEQ ID No. 50);

图 60 显示了从生物体天蓝色链霉菌 A3(2)获得的本发明的编码脂酰基转移酶的核苷酸序列 (SEQ ID No. 52) (Genbank 登录号 NC\_003888.1:8327480..8328367);

图 61 显示了从生物体天蓝色链霉菌 A3(2)获得的本发明的编码脂酰基转移酶的核苷酸序列 (SEQ ID No. 53) (Genbank 登录号 AL939131.1:265480..266367);

图 62 显示了从生物体灰色链霉菌获得的本发明的编码脂酰基转移酶的核苷酸序列 (SEQ ID No. 54) (Genbank 登录号 Z75034);

图 63 显示了从生物体雷尔氏菌属获得的本发明的编码脂酰基转移酶的核苷酸序列 (SEQ ID No. 55);

图 64 显示了如 SEQ ID No. 56 所示的核苷酸序列, 其编码 NCBI 蛋白登录号 CAB39707.1 GI:4539178 的保守的假定蛋白[天蓝色链霉菌 A3(2)];

图 65 显示了如 SEQ ID No. 57 所示的核苷酸序列,其编码 Scoe2,NCBI 蛋白登录号 CAC01477.1 GI:9716139 的保守的假定蛋白[天蓝色链霉菌 A3(2)];

图 66 显示了如 SEQ ID No. 58 所示的核苷酸序列,其编码 Scoe3,NCBI 蛋白登录号 CAB88833.1 GI:7635996 的公知分泌性蛋白[天蓝色链霉菌 A3(2)];

图 67 显示了如 SEQ ID No. 59 所示的核苷酸序列,其编码 Scoe4,NCBI 蛋白登录号 CAB89450.1 GI:7672261 的公知分泌性蛋白[天蓝色链霉菌 A3(2)];

图 68 显示了如 SEQ ID No. 60 所示的核苷酸序列,其编码 Scoe5,NCBI 蛋白登录号 CAB62724.1 GI:6562793 的公知脂蛋白[天蓝色链霉菌 A3(2)];

图 69 显示了如 SEQ ID No. 61 所示的核苷酸序列,其编码 Srim1,NCBI 蛋白登录号 AAK84028.1 GI:15082088 GDSL-脂肪酶[龟裂链霉菌];

图 70 显示了来自嗜水气单胞菌 (ATCC #7965) 的编码脂酰基转移酶的核苷酸序列 (SEQ ID No. 62);

图 71 显示了来自杀鲑气单胞菌杀鲑亚种 (ATCC#14174) 的编码脂酰基转移酶的核苷酸序列 (SEQ ID No. 63); 和

图 72 显示了来自嗜水气单胞菌的编码包含木聚糖酶信号肽的酶的核苷酸序列 (SEQ ID No. 24)。

## 实施例 1

### 地衣芽孢杆菌中 KLM3' 的表达

编码脂酰基转移酶 (SEQ. ID No. 16, 下文称为 KLM3') 的核苷酸序列 (SEQ ID No. 49) 在地衣芽孢杆菌中作为融合蛋白被表达, 所述融合蛋白具有地衣芽孢杆菌  $\alpha$ -淀粉酶 (LAT) 的信号肽 (参见图 53 和 54)。为了优化在芽孢杆菌中的表达, 从 Genart (Genart AG, Regensburg, Germany) 订购了密码子优化的基因构建体 (no. 052907)。

构建体 no. 052907 包含位于 LAT-KLM3' 前体基因之前的不完全 LAT 启动子 (仅-10 序列) 和 LAT-KLM3' 前体基因下游的 LAT 转录物 (Tlat) (参见图 53 和 55)。为了构建包含 5' 末端侧翼具有完全 LAT 启动子、3' 末端侧

翼具有 LAT 终止子的 LAT-KLM3'前体基因的 *XhoI* 片段, Plat5*XhoI*\_FW 和 EBS2*XhoI*\_RV 为引物, 基因构建体 052907 作为模板进行 PCR (聚合酶链式反应) 扩增。

Plat5*XhoI*\_FW:

```
ccccgctcgaggcttttctttggaagaaaatataggaaaatggtacttgtaaaaattc
ggaatattatacaatatcatatgtttcacattgaaagggg
```

EBS2*XhoI*\_RV: tggaatcctcgaggttttatcctttacctgtctcc

在热循环仪上, 根据生产商的说明书 (退火温度 55°C) 使用 Phusion 高保真度 DNA 聚合酶 (Finnzymes OY, Espoo, Finland) 来进行 PCR。

根据生产商的说明书 (Invitrogen, Carlsbad, Calif. USA), 所得 PCR 片段用限制性酶 *XhoI* 消化, 并使用 T4 DNA 连接酶将其连接到 *XhoI* 消化过的 pICatH 中。

如美国专利申请 US20020182734 (国际公布 WO 02/14490) 所述, 将连接混合物转化到枯草芽孢杆菌菌株 SC6.1 中。通过 DNA 测序确定包含 LAT-KLM3'前体基因的 *XhoI* 插入物的序列 (BaseClear, Leiden, The Netherlands), 并将其中的一个正确质粒克隆命名为 pICatH-KLM3'(ori1) (图 53)。在许可的温度 (37°C) 下将 pICatH-KLM3'(ori1) 转化地衣芽孢杆菌菌株 BML780 (BRA7 和 BML612 的衍生物, 参见 WO2005111203) 中。

选择一个新霉素抗性 (neoR) 和氯霉素抗性 (CmR) 的转化株并将其命名为 BML780(plCatH-KLM3'(ori1))。通过非许可温度 (50°C) 下、于含有 5 µg/ml 氯霉素的培养基中培养菌株, 使 BML780(plCatH-KLM3'(ori1)) 中的质粒整合到地衣芽孢杆菌基因组的 *catH* 区中。选择一个 CmR 抗性克隆并命名为 BML780-plCatH-KLM3'(ori1)。再次在许可温度、没有抗生素的情况下培养 BML780-plCatH-KLM3'(ori1) 数代, 以使载体序列环出 (loop-out), 并随后选择一个新霉素敏感 (neoS) 的 CmR 克隆。在此克隆中, 染色体上的 plCatH 载体序列被切除 (包括新霉素抗性基因), 并只剩下 *catH*-LATKLM3'盒。随后, 通过使菌株在氯霉素浓度增加的培养基中/上生长来扩增染色体上的 *catH*-LATKLM3'盒。在扩增数轮后, 选择一个克



隆（抗 50  $\mu\text{g/ml}$  氯霉素），并将其命名为 BML780-KLM3'CAP50。为了确定 KLM3' 表达，使 BML780-KLM3'CAP50 和 BML780（空宿主菌株）在添加有 1% 三丁酸甘油酯的 Heart Infusion (Bacto) 琼脂培养板上于 37°C 下培养 48 小时。清晰的区域（指示脂酰基转移酶活性）在 BML780-KLM3'CAP50 菌落周围而非宿主菌株 BML780 周围清晰可见（参见图 56）。此结果显示 KLM3' 在地衣芽孢杆菌菌株 BML780-KLM3'CAP50 中被大量的表达，且这些 KLM3' 分子是有功能的。

### 比较例 1

#### 载体构建体

所述质粒构建体为 pCS32new N80D，其为携带有第 80 位氨基酸由 Asn 取代成 Asp 的成熟形式的天然杀鲑气单胞菌甘油磷酸酯-胆固醇酰基转移酶的编码序列的 pCCmini 衍生物（KLM3'），其在 p32 启动子的控制下并具有 CGTase 信号序列。

用于表达的宿主菌株为枯草芽孢杆菌 OS21 $\Delta$ aprE 菌株。

以用酯化胆固醇的百分数表示的转移酶活性来计算表达水平，在 PC ( $T_{PC}$ ) 为供体和胆固醇为受体分子的反应中，根据参比样品中游离胆固醇和酶样品中游离胆固醇的差值来计算。

#### 培养条件

将单一菌落接种在添加有 50 mg/l 卡那霉素的 5 ml 的 LB 肉汤培养基（酪蛋白酶促消化物，10 g/l；低钠酵母提取物，5 g/l；氯化钠，5 g/l；惰性片化助剂（Inert tableting aids），2 g/l）中，并将 LB 肉汤培养基中于 205 rpm 下、30°C 孵育 6 小时。将 0.7 ml 此培养物接种到添加有 50 mg/l 卡那霉素和高麦芽糖淀粉水解物溶液（60 g/l）的 50 ml SAS 培养基（K<sub>2</sub>HPO<sub>4</sub>，10 g/l；MOPS（3-吗啡啉丙烷磺酸），40 g/l；氯化钠，5 g/l；消泡剂（Sin 260），5 滴/l；脱脂豆粉，20 g/l；Biospringer 106（100% dw YE），20 g/l）中。在 30°C 和 180 rpm 下继续孵育 40 小时，随后以 19000 rpm 离心 30 分钟来分离培养物上清液。将上清液转移到清洁试管并直接用于转移酶活性测定。

#### 底物的制备和酶促反应

以 9: 1 的比例称量 PC（Avanti Polar Lipids #441601）和胆固醇（Sigma

C8503), 将其溶解在氯仿中, 并蒸发至干燥。

通过将 3% PC 和胆固醇 (9: 1) 分散到 50 mM Hepes 缓冲液 (pH 7) 中来制备底物。

将 0.250 ml 底物溶液转移到带有螺纹盖的 3 ml 玻璃试管中。加入 0.025 ml 培养物上清液并将混合物在 40°C 下孵育 2 小时。用水代替酶来制备参比样品。将反应混合物在沸水浴中加热 10 分钟来终止酶反应。进行胆固醇试验分析前向反应混合物中加入 2 ml 99% 的乙醇。

#### 胆固醇试验

将 100  $\mu$ l 底物在 37°C 下孵育 5 分钟, 随后加入 5  $\mu$ l 酶反应样品并混合, 其中所述底物为含有 1.4 U/ml 胆固醇氧化酶 (SERVA Electrophoresis GmbH cat. No 17109)、0.4 mg/ml ABTS (Sigma A-1888)、6 U/ml 过氧化物酶 (Sigma 6782) 的 0.1 M Tris-HCl (pH 6.6) 和 0.5% Triton X-100 (Sigma X-100) 溶液。将反应混合物继续孵育 5 分钟并测量 OD<sub>405</sub>。根据对胆固醇标准溶液的分析计算胆固醇的含量, 所述胆固醇标准溶液为含有 0.4 mg/ml、0.3 mg/ml、0.20 mg/ml、0.1 mg/ml、0.05 mg/ml 和 0 mg/ml 胆固醇的 99% EtOH 溶液。

#### 结果

此表显示了 8 个单独表达培养物的平均值

菌株	T <sub>PC</sub> <sup>a</sup>
OS21 $\Delta$ AprE[pCS32new]	74.2 $\pm$ 10.1 <sup>b</sup>

<sup>a</sup>T<sub>PC</sub> 为转移酶活性, 用酯化胆固醇的百分数表示, 在 PC 为供体分子和胆固醇为受体分子的反应中, 根据参比样品中游离胆固醇和酶样品中游离胆固醇的差值来计算。

<sup>b</sup> 8 个单独表达培养物的平均值

以上说明书提到的全部出版物均通过引用并入本文。本发明所述方法和体系的各种修改和改变对本领域的技术人员来说都是显而易见的, 并不脱离本发明的范围和精神。虽然本发明是根据具体优选的实施方案进行描述的, 但是应该理解所主张的发明不仅限于所述具体实施方案。实际上, 用于实施本发明的所述模式的各种修改对于生物化学和生物工程或相关领域的技术人员来说都是显而易见的, 均应落入本发明权利要求书的范围。

图 1 (SEQ ID No. 16)

```
1  ADTRPAFSRI VMFGDSLSDT GKMYSKMRGY LPSSPPYYEG RFSNGPVWLE QLTQFPGLT
61  IANEAEAGGAT AVAYNKISWD PKYQVINNLD YEVTFQLQKD SFKPDDLVLW WVGANDYLA
121 GWNTEQDAKR VRDAISDAAN RMVLNGAKQI LLFNLPDLGQ NPSARSQKVV EAVSHVSAYH
181 NKLLLNLARQ LAPTMVKLF EIDKQFAEML RDPQNFGLSD VENPCYDGGY VWKPFATRSV
241 STDRQLSAFS PQLRLAIAGN PLLAQAVASP MARRSASPLN CEGKMFWDQV HPTTVVHAAL
301 SERAATFIET QYEFLAHG
```

图 2 (SEQ ID No. 1)

```
1  MKKWFVCLLG LVALTVQAAD SRPAFSRIVM FGDSLSDTGK MYSKMRGYLP
51  SSPPYYEGRF SNGPVWLEQL TKQFPGLTIA NEAEGGATAV AYNKISWNP
101 YQVINNLDYE VTQFLQKDSF KPDDLVLWV GANDYLA
151 DAISDAANRM VLNGAKQILL FNLPDLGQNP SARSQKVVEA VSHVSAYHNQ
201 LLLNLARQLA PTGMVKLFEI DKQFAEMLRD PQNFGLSDVE NPCYDGGYVW
251 KPFATRSVST DRQLSAFSPQ ERLAIAGNPL LAQAVASPMA RRSASPLNCE
301 GKMFWDQVHP TTVVHAALSE RAATFIANQY EFLAH*
```

图 3 (SEQ ID No. 2)

```
1 ivafGDSlTd geayygsdsd ggwgagladr Ltallrlrar prgvdfnrg isGrtsdGrl
61 ivDalvallF laqslglpnL pPYLsgdflr GANFAsagAt Ilptsgpfli QvqFkdfksq
121 vlelrqalgl lqellrllpv ldakspdlvt imiGtNDlit saffgpkste sdrnsvvpef
181 kdnlrqlikr Lrsnngarii vlitlvilnl gplGClPlkl alalassknv dasgclerln
241 eavadfneal relaiskled qlrkdglp dv kgadvpyvDl ysifqldgi qnpsayvyGF
301 ettkACCyG gryNynrvCG naglcnvtak aCnpssylls flfwDgfHps ekGykavAea
361 l
```

图 4 (SEQ ID No. 3)

```
1 mkkwfvcllg lvaltvqaad srpafsrivm fgdsldtgk myskmrqylp sspyyegrif
61 sngpvwleql tnefpgltia neaeggptav aynkiswnpk yqvinnldye vtqflqkdsf
121 kpddlvilwv gandylaygw nteqdakrvr daisdaanrm vlngakeill fnlpldgqnp
181 sarsqkvvea ashvsayhnq lllnlarqla ptgmvlkfei dkqfaemlrd pqnfglsdqr
241 nacyggsyvw kpfasrsast dsqlsafnpq erlaiagnpl laqavaspmarsastlnce
301 gkmfwdqvhp ttvvaalse paatfiesqy eflah
```

图 5 (SEQ ID No. 4)

```
1 mkkwfvcllg lialtvqaad trpafsrivm fgdsldtgk myskmrgylp ssppyyegrif
61 sngpvwleql tkqfpgltia neaeggatav aynkiswnpk yqvynnldye vtqflqkdsf
121 kpddlvilwv gandylaygw ntegdakrvr daisdaanrm vlngakqill fnlpldlgqnp
181 sarsqkvvea vshvsayhnk lllnlarqla ptgmvklfei dkqfaemlrd pqnfglsdve
241 npcydgggyvw kpfatrsvst drqlsafspq erlaiagnpl laqavaspma rrsasplnce
301 gkmfwdqvhp ttvvhaalse raatfietqy eflahg
```

图 6 (SEQ ID No. 5)

```
1 mpkpalrrvm tatvaavgtl algltdatah aapaqatptl dyvalgdsys agsgvlpvdp
61 anllclrsta nyphviadtt garltdvtcg aaqtadftra qypgvapqld algtgtdlvt
121 ltiggndnst finaitacgt agvlsggkgs pckdrhgtsf ddeieantyp alkeallgvr
181 arapharvaa lgywitpat adpscflklp laagdvpylr aiqahlndav rraaetgat
241 yvdfsgvsdg hdaceapgtr wiepllfghs lvpvhpnalg errmaehtmd vlgld
```

图 7 (SEQ ID No. 6)

```
1 mpkpalrrvm tatvaavgtl algltdatah aapaqatptl dyvalgdsys agsgvlpvdp
61 anllclrsta nyphviadtt garltdvtcg aaqtadftra qypgvapqld algtgtdlvt
121 ltiggndnst finaitacgt agvlsggkgs pckdrhgtsf ddeieantyp alkeallgvr
181 arapharvaa lgywitpat adpscflklp laagdvpylr aiqahlndav rraaetgat
241 yvdfsgvsdg hdaceapgtr wiepllfghs lvpvhpnalg errmaehtmd vlgld
```

图 8 (SEQ ID No. 7)

```
1 mdyekfllfg dsitefafnt rpiedgkdqy algaalvney trkmdilqrg fkgytsrwal
61 kilpeilkhe snivmatifl gandacsagp qsvplpefid nirqmvslmk syhirpiiig
121 pglvdrekwe kekseeialg yftrtnenfai ysdalaklan eekvpfvaln kafqqeggda
181 wqqlldglh fsgkgykifh dellkvietf ypyqhpknmq yklkdwrdbl ddgsnims
```

图 9 (SEQ ID No. 8)

10	20	30	40	50	60
MNLRQWMGAA	TAALALGLAA	CGGGGTDQSG	NPNVAKVQRM	VV <b>FGDSL</b> SDI	GTYTPVAQAV
70	80	90	100	110	120
GGGKFTTNP	PIWAETVAAQ	LGVTLTPAVM	GYATSVQNCP	KAGCFDYAQQ	GSRVTDPNGI
130	140	150	160	170	180
GHNGGAGALT	YPVQQQLANF	YAASNNTFNG	NNDVVFVLAG	SNDIFFWTTA	AATSGSGVTP
190	200	210	220	230	240
AIATAQVQQA	ATDLVGYVKD	MIAKGATQVY	VFNLPDSSLT	PDGVASGTTG	QALLHALVGT
250	260	270	280	290	300
FNTTLQSGLA	GTSARIIDFN	AQLTAAIQNG	ASFGFANTSA	RACDATKINA	LVPSAGGSSL
310	320	330	340		
FCSANTLVAS	GADQSYLFAD	GVHPTTAGHR	LIASNVLARL	LADNVAH	

图 10 (SEQ ID No. 9)

```

1 migsvavgd sftegvdpd pdgafvgwad rlavlladdr pegdftytnl avrgrlldqi
61 vaevprvvg lapdlvsfaa ggndiirpqt dpdevaerfe lavaaltaaa gtlvlttgfd
121 trgvplkhl rgkiatyngv vraidrygc pvldlwlslr vqdrrawdada rhlhspeght
181 rvalragqal glrvpadpdq pwpplpprgt ldvrrddvhw areylvpwig rrlrgessgd
241 hvtagtllsp daiktriaav a

```

图 11 (SEQ ID No. 10)

```

1 mqtncpytsl vavgdsfteg msdllpdgsy rgwadllatr maarspgfry anlavrgkli
61 gqivdeqvvd aaamgadvit lvggldntlr pkcdmarvrd lltqaverla phceqlvlmr
121 spgrqgvkvl rfrprmealv aiddlagrh gavvvdlyga qsladprmwv vdrhlhtaeg
181 hrrvaeavwq slghepedpe whapipatpp pgwvtrrtad vrfarqhlip wigrlltgrs
241 sgdglpakrp dllpyedpar

```

图 12 (SEQ ID No. 11)

```

1 mtrgrdggag apptkhrall aaivtlivai saaiyagasa ddgsrdhalq aggrlprgda
61 apastgawvg awatapaaae pgtettglag rsvrnvhts vggtagarilt snlygqspit
121 vthasialaa gpdtaaaiaad tmrrltfggs arviipagqg vmsdtariai pyganvlvtt
181 yspipsgpvt yhpqarqtsy ladgdrtadv tavayttptp ywryltaldv lsheadgtvv
241 afgdsitdga rsqsdanhrw tdvlaarlhe aagdgrdtp rsvvnegisg nrlttrpgr
301 padnpsglsr fgrdvlertn vkavvvvlgv ndvlinspela drdailtglr tlvdraharg
361 lrvvgatitp fgggygytea retmrqevne eirsgrvfdt vvdfdkalrd pydprmrds
421 ydsghlhpq dkgyarmgav idlaalkgaa pvka

```

图 13 (SEQ ID No. 12)

```

1 mtsmsrarva rriaagaayg gggiglagaa avglvvaevq larrrvvgvt ptrvpnaqgl
61 yggtlptagd pplrlmmlgd staaggvhr agqtpgalla sglaavaerp vrlgsvaqpg
121 acsddldrvq alvlaepdrv pdicvimvga ndvthmpat rsvrhlssav rrlrtagaev
181 vvgtcpdlgt iervrqplrw larrasrqla aaqtigaveq ggtrvslgdl lgpefaqnpr
241 elfgpdnyhp saegyataam avlpsvcaal glwpadeehp dalrregflp varaaaaaas
301 eagtevaam ptgprgpwal lkrrrrrrvs eaepsspsgv

```

图 14 (SEQ ID No. 13)

```

1 mgrgtdqrtr ygrrrarval aaltaavlgv gvagcdsvgg dspapsgsps krtrtapawd
61 tspasvaavg dsitrgfdac avlsdcpevs watgssakvd slavrligka daaehswnya
121 vtgarmadlt aqvtraaqre pelvavmaga ndacrsttsa mtpvadfrac feeamatlrk
181 klpkaqvyvs sipdlkrlws qgrtnplgkq vwklglcpsm lgdadslds atlrntvrd
241 rvadynevlr evcakdrrcr sddgavhefr fgtdqlshwd wfhpsvdgqa rlaeiayrav
301 taknp

```

图 15 (SEQ ID No. 14)

```

1 mrlsrraata sallltpala lfgasaavsa priqatdyva lgdsyssgvg agsydsssgs
61 ckrstksypa lwaashtgtr fnftacsgar tgdvlakqlt pvnsqtdlvs itiggndagf
121 adtmttcnlq gesaclaria karayiqqt1 paqldqvyda idsrapaaqv vvlgyprfyk
181 lggscavgl5 eksraainaa addinavtak raadhgafag dvnttfaghe lcsgapwlhs
241 vtlpvensyh ptangqskgy lplnsat

```

图 16 (SEQ ID No. 15)

```

1 MKKWFVCLLG LIALTVQAAD TRPAFSRIVM FGDSLSDTGK MYSKMRGYLP
51 SSPPYYEGRF SNGPVWLEQL TKQFPGLTIA NEAEGGATAV AYNKISWNEK
101 YQVINNL DYE VTQFLQKDSF KPDDLVLWV GANDYLAYGW NTEQDAKRVR
151 DAISDAANRM VLNGAKQILL FNLPDLGQNP SARSQKVVEA VSHVSAYHMK
201 LLLNLARQLA PTGMVKLFEI DKQFAEMLRD PQNFGLSDVE NPCYDGGYVW
251 KPFATRSVST DRQLSAFSPQ ERLAIAGNPL LAQAVASPMA RRSASPLNCE
301 GKMFWDQVHP TTVVHAALSE RAATFIETQY EFLAHG*

```

图 17 (SEQ ID No. 19)

```

1 migsyvavgd sftegvgdpg pdgafvgwad rlavlladdr pegdftytnl avrgrlldqi
61 vaeqvprvvg lapdlvsfaa ggndiirpgt dpdevaerfe lavaalaaa gtlvttgfd
121 trgvplkhl rgkiatyngv vraiadrygc pvldlwslrs vqdrrawdada rhlhspegt
181 rvalragqal glrvpadpdq pwpplpprgt ldvrrddvhw areylvpwig rrlrgessgd
241 hvtakgtlsp daiktriaav a

```



图 18 (SEQ ID No. 25)

1 MFKEKKNFLV GLSAALMSIS LFSATASAAS ADSRPAFSRI VMFGDSLSDT  
51 GKMYSKMRGY LPSSPPYYEG RFSNGPVWLE QLTKQFPGLT IANEAEGGAT  
101 AVAYNKISWN PKYQVINNLD YEVTQFLQKD SFKPDDLIVL WVGANDYLAY  
151 GWNTEQDAKR VRDAISDAAN RMVLNGAKQI LLFNLPDLGQ NPSARSQKVV  
201 EAVSHVSAYH NQLLLNLRQ LAPTGMVKLF EIDKQFAEML RDPQNFGLSD  
251 VENPCYDGGY VWKPFATRSV STDRQLSAFS PQRERLAIAGN PLLAQAVASP  
301 MARRSASPLN CEGKMFWDQV HPTTVVHAAL SERAATFIAN QYEFLAH\*\*

图 19 (SEQ ID NO. 26)

MRLTRLSAASVIVFALLLALLGISPAQAAGPAYVALGDSYSSNGAGSYIDSSGDCHRSN  
NAYPARWAAANAPSSFTFAACSGAVTTDVINNQLGALNASTGLVSITIGGNDAGFADAMTT  
CVTSSDSTCLNRLATATNYINTTLLARLDAVYSQIKARAPNARVVVLGYPRMYLASNPWYC  
LGLSNTKRAAINTTADTLNSVISSRATAHGFRFGDVRPTFNNHELFFGNDWLHSLTLPVWE  
SYHPTSTGHQSGYLPVLNANSST

图 20 (SEQ ID No. 27)

ZP 00058717

```
1 mlphpagerg evgaffallv gtpqdrirl echetrplrg rcgcgerrvp pltlpgdgvl
61 cttssstrdae tvwrkhlqpr pdggfrphlg vgcllagqgs pglwlcgreg crfevcrrdt
121 pglstrngd ssppfragws lppkcgeisq sarktpavpr ysllrtdrpd gprgrfvgs
181 praatrrrlf lgipalvlvt altlavlavpt gretlwrwc eatqdwclgv pvdsrgqpae
241 dgeflllspv qaatwgnyya lgdsyssgdg ardyypgtav kggcwrsana ypelvaeayd
301 faghlsflac sgqrqyamld aidevgsqld wnsphstslvt igiggndlgf stvlktcmvr
361 vplldskact dqedairkm akfettfeel isevrtrapd arilvgypr ifpeeptgay
421 ytltasnqrw lnetiqefnq qlaeavavhd eeiaasggvg svefvdvyha ldgheigsde
481 pwvngvqlrd latgvtvdrs tfhpnaaghr avgervieqi etgpgrplya tfavvagatv
541 dtlagevg
```

图 21 (SEQ ID No. 28)

```
1 msggptraatr rrlflgipal vlvtaltlvt avptgretlw rmwceatqdw clgvpvdsrg
61 qpaedgefll lspvqaatwg nyyalgdsys sgdgadyyp gtavkkgcwr sanaypelva
121 eaydfaghls flacsgqrqy amldaidevg sqldwnspht slvtigiggn dlgstvlkt
181 cmvrpllds kactdqedai rkrmakfett feelisevrt rapdarilvv gyprifpeep
241 tgayytltas nqrwlnetiq efnqqlaeav avhdeeiaas ggvgsvefd vyhaldghei
301 gsdepwvngv qlrdlatgvt vdrstfhpna aghravger v ieqietgpgr plyatfavva
361 gatvdtlage vg
```

图 22 (SEQ ID No. 29)

```
1 mrttviaasa llllagcadg areetagapp gessggiree gaeastsitd vyialgdsya
61 amggrdqplr gepfclrssg nypellhaev tdltcqgavt gdlleprtl g ertlpaqvda
121 ltedttlvtl siggndlgfg evagcireri agenaddcvd llgetigeql dqlppqldrv
181 heairdragd aqvvtgylp lvsagdcpel gdvseadrrw aveltgqine tvreaaerhd
241 alfvlpddad ehtscappq rwadiqqqt dayplhptsa gheamaavr dalglepvqp
```

图 23 (SEQ ID No. 30)

ZP 00094165

```
1 mgqvklfarr capvllalag lapaatvare aplaegaryv algssfaagp gvgnapgpsp
61 ercgrgtlly phllaealkl dlvdatsga tthhvlgpwn evppqidsvn gdtrlvlti
121 ggndvsfvgn ifaaacekma spdprcgkwr eiteewqad eermrsivrq iharaplary
181 vvdvityvlp psgtcaamai spdrlagsrs aakrlarita rvareegasl lkfshisrrh
241 hpcsakpwsn glsapaddgi pvhpnrlgha eaaaalvklv klmk //
```

图 24 (SEQ ID No. 31)

NP\_625998.

```

1 mrrfrlvgfl sslvlaagaa ltgaataqaa qpaaadgyva lgdsyssvgv agsyisssgd
61 ckrstkahpy lwaaahpst fdftacsgar tgdvlsqqlg plssgtglvs isiggndagf
121 adtmttcvlq sessclsria taeaydstl pgkldgvysa isdkapnahv vvigyprfyk
181 lgttciglse tkrtainkas dhlntvlaqr aaahgftfgd vrttftghel csgspwlhsv
241 nwl nigesyh ptaagqsggy lpvlnгаа

```

//

图 25 (SEQ ID No. 32)

NP\_827753.

```

1 mrrsritayv tslllavgca ltgaataqas paaaatgyva lgdsyssvgv agsylvssgd
61 ckrsskaypy lwqaahspss fsfmacsgar tgdvlnqlg tlnsstglvs ltiggndagf
121 sdvmttcvlq sdsaclsrin takaydstl pgqldsvyta istkapsahv avlgyprfyk
181 lggscлагls etkrsainda adylnsaiak raadhgftfg dvkstftghe icssstwlhs
241 ldllnigqsy hptaagqsgg ylpvmnsva

```

//

图 26 (SEQ ID No. 33)

```

MRLTRLSAASVIVFALLLALLGISPAQAAGPAYVALGDSYSSNGAGSYIDSSGDCHRSN
NAYPARWAAANAPSSFTFAACSGAVTTDVINNQLGALNASTGLVSIITIGNDAGFADAMTT
CVTSSDSTCLNRLATATNYINTTLLARLDAVYSQIKARAPNARVVVLGYPRMYLASNPWYC
LGLSNTKRAAINTTADTLNSVISSRATAHGFRFGDVRPTFNNHELFFGNDWLHSLTLPVWE
SYHPTSTGHQSGYLPVLNANSST

```

图 27 (SEQ ID No. 34)

ADSRPAFSRIVMFGDSLSDTGKMYSKMRGYLPSSPPYYEGRFSNGPVWLEQLTNEFPGLTIANEAEGGPTAVAYNKIS  
WNPKYQVINNLDYEVTFQKDSFKPDDLVLWVGANDYLAYGWNTAQDAKVRDAISDAANRMVLNGAKEILLFNLP  
DLGQNPSARSQKVVEAASHVSAYHNQLLLNLARQLAPTGMVKLFEIDKQFAEMLRDPQNFGLSDQRNACYGGSYVWKP  
FASRSASTDSQLSAFNPQERLAIAGNPLLAQAVASPMARSASTLNCE  
GKMFWDQVHPTTVVHAALSEPAATFIESQYEFLLAH

图 28 (SEQ ID No. 35)

1 ADTRPAFSRI VMFGDSLSDT GKMYSKMRGY LPSSPPYYEG RFSNGPVWLE QLTQFPGLT  
61 IANEAEGGAT AVAYNKISWN PKYQVINNLD YEVTQFLQKD SFKPDDLVL WVGANDYLAY  
121 GWNTAQDAKR VRDAISDAAN RMVLNGAKQI LFNLPDLGQ NPSARSQKV EAVSHVSAYH  
181 NKLLNLRARQ LAPTMVKLF EIDKQFAEML RDPQNFGLSD VENPCYDGGY VWKPFATRSV  
241 STDRQLSAFS PQERLAIAGN PLLAQAVASP MARRASPLN CEGKMFWDQV HPTTVVHAAL  
301 SERAATFIET QYEFLLAHG

**图 29 (SEQ ID No. 36)**

ACAGGCCGATGCACGGAACCGTACCTTTCGCGAGTGAAGCGCTCTCCCCCATCGTTTCGC  
 CGGGACTTCATCCGCGATTTTGGCATGAACACTTCCTTCAACGCGCGTAGCTTGCTACAA  
 GTGCGGCAGCAGACCCGCTCGTTGGAGGCTCAGTGAGATTGACCCGATCCCTGTCGGCCG  
 CATCCGTCATCGTCTTCGCCCTGCTGCTCGCGCTGCTGGGCATCAGCCCGGCCAGGCAG  
 CCGGCCCGGCTATGTGCCCTGGGGGATTCTATTCTCGGGCAACGGCGCCGGAAGTT  
 ACATCGATTGAGCGGTGACTGTCACCGCAGCAACAACCGGTACCCCGCCCGCTGGGCGG  
 CGGCCAACGCACCGTCCCTTACCTTCGCGGCCTGCTCGGGAGCGGTGACCACGGATG  
 TGATCAACAATCAGCTGGGCGCCCTCAACGCGTCCACCGGCCTGGTGAGCATCACCATCG  
 GCGGCAATGACGCGGGCTTCGCGGACGCGATGACCACCTGCGTCACCAGCTCGGACAGCA  
 CCTGCCTCAACCGGCTGGCCACCGCCACCAACTACATCAACACCACCCTGCTCGCCCGG  
 TCGACGCGGTCTACAGCCAGATCAAGGCCCGTGCCCCAACGCCCGCGTGGTTCCTCG  
 GCTACCCGCGCATGTACCTGGCCTCGAACCCTGGTACTGCCTGGGCCTGAGCAACACCA  
 AGCGCGCGGCCATCAACACCACCGCCGACACCCTCAACTCGGTGATCTCCTCCCGGCCA  
 CCGCCCACGGATTCCGATTTCGGCGATGTCCGCCCGACCTTCAACAACCACGAACTGTTCT  
 TCGGCAACGACTGGCTGCACTCACTCACCTGCCGGTGTGGGAGTCGTACCACCCACCA  
 GCACGGGCCATCAGAGCGGCTATCTGCCGGTCTCAACGCCAACAGCTCGACCTGATCAA  
 CGCACGGCCGTGCCGCCCGCGCGTACGCTCGGCGGGCGCCGAGCGGTTGATCA  
 GCCCACAGTGCCGGTGACGGTCCCACCGTACGCGTTCGAGGGTGTACGTCACGGTGGCGCC  
 GCTCCAGAAGTGGAACGTCAGCAGGACCGTGGAGCCGTCCCTGACCTCGTGAAGAATC  
 CGGGGTACGCGTGATCACCCCTCCCCGTAGCCGGGGGGAAGGCGGCGCGAACTCCTT  
 GTAGGACGTCAGTTCGTCGGCCCGCGTTCGCCACCGTCCGCGTAGACCGCTTCCATGGT  
 CGCCAGCCGGTCCCCGCGGAACTCGGTGGGGATGTCCGTGCCCAAGGTGGTCCCGGTGGT  
 GTCCGAGAGCACCGGGGCTCGTACCGGATGATGTGCAGATCCAAAGAATT

**图 30 (SEQ ID NO. 37):**

MRLTRLSAASVIVFALLLALLGISPAQAAGPAYVALGDSYSSGNGAGSYIDSSGDCHRSN  
 NAYPARWAAANAPSSFTFAACSGAVTTDVINNQLGALNASTGLVSIITIGGNDAGFADAMTT  
 CVTSSDSTCLNRLATATNYINTTLLARLDAVYSQIKARAPNARVVVLGYPRMYLASNPWYC  
 LGLSNTKRAAINTTADTLNSVISSRATAHGFRFGDVRPTFNNHELFFGNDWLHSLTLPVWE  
 SYHPTSTGHQSGYLPVLNANSST

图 31 (SEQ ID No. 38)

```
1 mlphpagerg evgaffallv gtpqdrirl echetrplrg rccgerrvp pltlpgdgv1
61 cttssstrdae tvwrkhlqpr pdggfrphlg vgccllaggs pgvlwcgreg crfevcrrdt
121 pglrstrngd ssppfragws lppkcgeisq sarktpavpr ysllrtdrpd gprgrfvgsg
181 praatrrrlf lgipalvlvt altlavlavpt gretlwmwc eatqdwclgv pvdsrgqpae
241 dgeflllspv qaatwgnyya lgdsyssgdg ardyypgtav kggcwrsana ypelvaeayd
301 faghlsflac sgqrgyamld aidevgsqld wnsphtslvt igiggndlgf stvlkctmvr
361 vplldskact dgedairkrm akfettfeel isevrtrapd arilvvgypr ifpeeptgay
421 ytltasqrw lnetiqefnq qlaeavavhd eeiaasggvg svefvdvyha ldgheigsde
481 pwvngvqlrd latgvtvdrs tfhpnaaghr avgervieqi etgpgrplya tfavvagatv
541 dtlagevg
```

图 32 (SEQ ID No. 39)

```

1 ggtggtgaac cagaacaccc ggtcgtcggc gtggggcgtcc aggtgcaggt gcaggttctt
61 caactgctcc agcaggatgc cgccgtggcc gtgcacgatg gccttgggca ggcctgtggt
121 ccccgacgag tacagcacc atagcggatg gtcgaacggc agcggggtga actccagttc
181 cgcgccttcg cccgcggctt cgaactccgc ccaggacagg gtgtcggcga cagggccgca
241 gcccaggtac ggcaggacga cgggtgtctg caggctgggc atgccgtcgc gcagggttt
301 gagcacgtca cggcggtcga agtccttacc gccgtagcgg tagccgtcca cggccagcag
361 cactttcggg tcgatctcgc cgaaccggtc gaggacgctg cgcaccccca agtcggggga
421 acaggacgac caggtcgcac cgatcgcggc gcaggcgagg aatgcggccg tcgcctcggc
481 gatgttcggc aggttaggcca cgaccggctc gccggggccc accccgaggc tcgggagggc
541 cgcagcgatc gcggcgggtc gggtcgcgag ttctcccag gtccactcgg tcaacggccg
601 gagttcggac gcgtgcccga tcgccacggc tgatgggtca cggtcgcgga agatgtgctc
661 ggcgtagtgt aggggtggcg cggggaacca gacggcgccg ggcagtgccg cggaggcgag
721 cactgtggtg tacgggggtg cggcgcgcac ccggtagtac tcccagatcg cggaccagaa
781 tccttcgagg tcggttaccg accagcgcca cagtgcctcg tagtcgggtg cgtccacacc
841 gcggtgctcc cgcaccagc ggggtgaacgc ggtgaggttg gcgcgttctt tgcgctcctc
901 gtcgggactc cacaggatcg gcggctgcgg cttgagtgtc atgaaacgcg accccttcgt
961 ggacggtgcg gatgcgggtg gcgtcgggtg cctcccctaa cgctcccggg tgacggagtg
1021 ttgtgcacca catctagcac gcgggacgcg gaaaccgtat ggagaaaaca cctacaacc
1081 cggccggacg gtgggttctc gccacactta ggggtcgggt gcctgcttgc cggcagggc
1141 agtcccgggg tgctgtggtg cgggcgggag ggctgtcgtc tcgaggtgtg ccggcgggac
1201 actccgggcc tcagccgtac ccgcaacggg gacagttctc ctccctccg ggctggatgg
1261 tcccttcccc cgaatgcgg cgagatctcc cagtcagccc ggaaaacacc cgctgtgcc
1321 aggtactctt tgcttcgaac agacaggccg gacggtcac gggggaggtt tgtgggcagc
1381 ggaccacgtg cggcgaccag acgacggttg ttctcggta tccccgctct tgtacttgtg
1441 acagcgtca cgtggtctt ggctgtccc acggggcgcg agacgctgtg gcgctgtgg
1501 tgtgaggcca cccaggactg gtgcctgggg gtgccggtcg actcccggg acagcctgcg
1561 gaggacggcg agtttctgct gctttctccg gtccaggcag cgacctgggg gaactattac
1621 gcgctcgggg attcgtactc ttcgggggac ggggcccgcg actactatcc cggcaccgcg
1681 gtgaaggcg gttgctgcg gtccgctaac gcctatccg agctggtcgc cgaagcctac
1741 gacttcgccc gacactgtc gttcctggcc tgcagcggcc agcgcggcta cgcatgctt
1801 gacgctatcg acgaggtcgg ctgcagctg gactggaact cccctcacac gtcgctggtg
1861 acgacgga tcggcgcaa cgatctggg ttctccagc ttttgaagac ctgcatggtg
1921 cgggtgccgc tgctggacag caaggcgtgc acggaccagg aggacgctat ccgcaagcgg
1981 atggcgaat tcgagacgac gtttgaagag ctcatcagc aagtgcgcac ccgcgcgcc
2041 gacgcccgga tccttgtcgt gggctacccc cggatthttc cggaggaacc gaccggcgcc
2101 tactacacgc tgaccgcgag caaccagcgg tggctcaacg aaaccattca ggagttcaac
2161 cagcagctcg ccgaggtgt cgcggtccac gacgaggaga ttgccgcgtc gggcggggtg
2221 ggcagcgtgg agttcgtgga cgtctaccac gcgttgacg gccacgagat cggctcggac
2281 gagccgtggg tgaacggggt gcagttgcgg gacctcgcca ccggggtgac tgtggaccgc
2341 agtaccttcc accccaacgc cgctgggcac cgggcggtcg gtgagcgggt catcgagcag
2401 atcgaaaccg gcccgggccg tccgctctat gccactttcg cgggtggtggc gggggcgacc
2461 gtggacactc tcgcgggcga ggtggggtga cccggcttac cgctccggcc gcaggtctgc
2521 gagcactcgc gcgatctggt ccaactgcca gtgcagttcg tcttcggtga tgaccagcgg
2581 cggggagagc cggatcgttg agcgtgctg gtctttgacg agcacacccc gctgcaggag
2641 ccgttcgcac agttctcttc cgggtggccag agtcgggtcg acgtcgatcc cagcccacag
2701 gccgatgctg cgggccgcga ccacgcggt gccgaccagt tggtcgaggc gggcgcgcag
2761 cacgggggcg agggcgcgga catggtccag gtaagggccg tcgcgacga ggctcaccac
2821 ggcagtgcg accgcgcagg cgagggcgtt gccgccgaag gtgctgccgt gctggccggg
2881 gcggatcacg tcgaagactt ccgcgtcgc taccgccgc gccacgggca ggatgccgcc
2941 gccagcgtct ttgccgaaca ggtagatata ggcgtcgaact ccgctgtggt cgcaggcccc

```



**图 33** (SEQ ID No. 40)

```
1  vsgspraatr  rrlflgipal  vlvtaltlvl  avptgretlw  rmwceatqdw  clgvpvdsrg
61  qpaedgefl1  lspvqaatwg  nyyalgdsys  sgdgardyyp  gtavkggcwr  sanaypelva
121 eaydfaghls  flacsgqrgy  amldaidevg  sqldwnspht  slvtigiggn  dlgfstvlkt
181 cmvrvpllds  kactdqedai  rkrmakfett  feelisevrt  rapdarilvv  gyprifpeep
241 tgayytltas  nqrwlnetiq  efnqqlaeav  avhdeeiaas  ggvgsvefvd  vyhaldghei
301 gsdepwvngv  qlrdlatgvt  vdrstfhpna  aghravgerv  ieqietgpgr  plyatfavva
361 gatvdtlage  vg
```

**图 34** (SEQ ID No. 41)

```
1  mrttviaasa  llllagcadg  areetagapp  gessggiree  gaeastsitd  vyialgdsya
61  amggrdqplr  gepfclrsg  nypellhaev  tdltcggavt  gdllleprtl  g ertlpaqvda
121 ltedttlvtl  siggndlgfg  evagcireri  agenaddcvd  llgetigeql  dqlppqldrv
181 heairdragd  aqvvtgylp  lvsagdcpel  gdvseadrrw  aveltgqine  tvreaaerhd
241 alfvlpddad  ehtscappqg  rwadiqqgqt  dayplhptsa  gheamaaavr  dalglepvqp
```

图 35 (SEQ ID No. 42)

```

1  ttctgggggtg  ttatgggggtt  gttatcggct  cgtcctgggt  ggatcccgcc  aggtggggta
61  ttcacggggg  acttttgtgt  ccaacagccg  agaatgagtg  ccctgagcgg  tgggaatgag
121  gtggggcggg  ctgtgtcggc  atgagggggc  ggcgggctct  gtggtgcccc  ggcacccccg
181  gccccggtga  gcggtgaatg  aaatccggct  gtaatcagca  tcccgtgcc  acccgtcgg
241  ggaggtcagc  gcccggagtg  tctacgcagt  cggatcctct  cggactcggc  catgctgtcg
301  gcagcatcgc  gctcccgggt  cttggcgtcc  ctcggctgtt  ctgcctgctg  tccctggaag
361  gcgaaatgat  caccggggag  tgatacaccg  gtggtctcat  cccggatgcc  cacttcggcg
421  ccatccggca  attcgggcag  ctccgggtgg  aagtagtggt  catccgatgc  gtcggtgacg
481  ccatagtggg  cgaagatctc  atcctgctcg  aggtggtcga  ggccactctc  cggatcgata
541  tcggggggcgt  ccttgatggc  gtccttgctg  aaaccgaggt  gcagcttggt  ggcttccaat
601  ttcgcaccac  ggagcgggac  gaggctgga  tgacggccga  agagcccgtg  gtggacctca
661  acgaaggtgg  gtagtcccgt  gtcattctg  aggaacacgc  cctccaccgc  acccagcttg
721  tggccggagt  tgtcgtaggc  gctggcatcc  agaagggaaa  cgatctcata  tttgtcggtg
781  tgctcagaca  tgatcttcct  ttgctgtcgg  tgtctgttac  taccacggta  gggctgaatg
841  caactgttat  ttttctgtta  ttttaggaat  tggctccatat  cccacaggct  ggctgtggtc
901  aaatcgtcat  caagtaatcc  ctgtcacaca  aaatgggtgg  tgggagccct  ggtcgcgggt
961  ccgtgggagg  cgccgtgcc  cgcaggatcg  tcggcatcgg  cggatctggc  cggtaaccgg
1021  cggtgaataa  aatcattctg  taacctcat  cacggttggt  tttaggtatc  cgcccccttc
1081  gtcctgaccc  cgtccccggc  gcgcgggagc  ccgcggttg  cggtagacag  gggagacgtg
1141  gacaccatga  ggacaacggg  catcgcagca  agcgcattac  tccttctcgc  cggatgcggc
1201  gatggggccc  gggaggagac  cgcgggtgca  ccgcccgggt  agtcctccgg  gggcatccgg
1261  gaggaggggg  cggaggcgtc  gacaagcatc  accgacgtct  acatgccct  cggggattcc
1321  tatgcggcga  tgggcggggc  ggatcagccg  ttacgggggt  agccgttctg  cctgcgctcg
1381  tccggtaatt  acccgaact  cctccagca  gaggtcaccg  atctcacctg  ccaggggcg
1441  gtgaccgggg  atctgctcga  acccaggacg  ctgggggagc  gcacgctgcc  ggcgagggtg
1501  gatgcgctga  cggaggacac  caccctggtc  accctctcca  tcgggggcaa  tgacctcgga
1561  ttccggggagg  tggcgggatg  catccgggaa  cggatcggcg  gggagaacgc  tgatgattgc
1621  gtggacctgc  tgggggaaac  catcggggag  cagctcgatc  agcttcccc  gcagctggac
1681  cgcgtgcacg  aggctatccg  ggaccgcgcc  ggggacgcgc  aggttgtggt  caccggttac
1741  ctgccgctcg  tgtctgcgg  ggactgcccc  gaactggggg  atgtctccga  ggcggatcgt
1801  cgttggcgcg  ttgagctgac  cgggcagatc  aacgagaccg  tgcgcgagcg  ggcggaacga
1861  cacgatgcc  tctttgtcct  gcccagcat  gccgatgagc  acaccagttg  tgacccccca
1921  cagcagcgt  gggcggatat  ccagggccaa  cagaccgatg  cctatccgct  gcaccggacc
1981  tccgccggcc  atgaggcgat  ggccgccgcc  gtccgggacg  cgtgggct  ggaaccggtc
2041  cagccgtagc  gccgggcggc  cgcttgtcga  cgaccaacc  atgccaggct  gcagtcacat
2101  ccgcacatag  cgcgcgcggg  cgatggagta  cgcaccatag  aggatgagcc  cgatgccagc
2161  gatgatgagc  agcacactgc  cgaagggttg  ttccccgagg  gtgcgcagag  ccgagtcag
2221  acctgcggcc  tgctccggat  catgggcca  accggcgatg  acgatcaaca  cccccaggat
2281  ccgaaggcg  ataccacggg  cgacataacc  ggtgttccg  gtgatgatga  tcggggtccc
2341  gacctgccct  gaccccgcac  ccgcctccag  atcctcccg  aaatcccggg  tggccccctt
2401  ccagaggttg  tagacaccgg  cccccagtac  caccagccc  gcgaccacaa  ccagcaccac
2461  accccagggt  tgggatagga  cgggtggcgg  gacatcggtg  gcggtctccc  catcggagggt
2521  gctgccgccc  cgggcgaagg  tggaggtggt  caccgccagg  gagaagtaga  ccatggccat
2581  gaccgcccc  ttggccctt  ccttgaggtc  ctgcccggc  agcagctggc  tcaattgcca
2641  gagtcccagg  gccgccaggg  cgatgacggc  aaccacagg  aggaactgcc  caccgggagc
2701  ctccgcgatg  gtggccaggg  cacctgaatt  cgaggcctca  tcaccgaac  cgccggatcc
2761  agtggcgatg  cgcaccgcga  tccaccgat  gaggatgtgc  agtatgcca  ggacaatgaa
2821  accacctctg  gccagggtgg  tcagcgcggg  gtggtcctcg  gctgtgctgg  cagcccgttc
2881  gatcgtccgt  ttcgcggtac  tgggtgtcgg  cttatccata  gctcccattg  aaccgcttg
2941  aggggtgggc  ggcactgtc  agggcggatt  gtgatctgaa  ctgtgatgtt  ccatcaacc

```

图 36 (SEQ ID No. 43)

```

1 mrrfrlvglf sslvlaagaa ltgaataqaa qpaaadyva lgdsyssvgv agsyisssgd
61 ckrstkahpy lwaaahspst fdftacs gar tgdvlsqqlg plssgtglvs isiggn dagf
121 adtmttcvlq sessclsria taeayvdstl pgkldgvysa isdkapnavh vvigyprfyk
181 lgttciglse tkrtainkas dhlntvlaqr aaahgftfgd vrtrftghel csgspwlhsv
241 nwl nigesyh ptaagqsggy lpvln gaa

```

图 37 (SEQ ID No. 44)

```

1 cccggcgcc cgtgcaggag cagcagccgg cccgcgatgt cctcggcgt cgtcttcac
61 agggcgtcca tgcgctcggc gaccggcgcc gtgtagttgg cccggacctc gtcccagggtg
121 cccggcgga tctggcgggt ggtgcggtgc gggccgcgcc gaggggagac gtaccagaag
181 cccatcgtca cgttctccgg ctgcggttcg ggctcgtccg ccgctccgtc cgtcgcctcg
241 ccgagcact tctcggcgag gtcggcgctg gtcgcccgtca ccgtgacgtc ggcgccccgg
301 ctccagcgg agatcagcag cgtccagccg tcgccctccg ccagcgtcgc gctgcggctg
361 tcgttcggg cgatccgcag cagcgcggcg cggggcgga gcagcgtggc gccggaccgt
421 acgcggtcga tgttcgccgc gtgcgagtac ggctgctcac ccgtggcga acggccgagg
481 aacagcgcgt cgacgacgtc ggacggggag tcgctgtcgt ccacgttgag ccggatcggc
541 agggcttctg ggggttcac ggacatgtcg ccatgatcgg gcaccggcc gccgcgtgca
601 cccgcttcc cgggcacgca cgacagggc tttctcggc tcttccgtcc gaacttgaac
661 gagtgtcagc catttcttg catggacct tccagtcaac ggcgtagct gctaccagg
721 ttgtggcagc aatcctgcta agggagttc catgagact tccgacttg tcggcttct
781 gagttcgctc gtcctcggc cggcgccgc cctcaccgg gcagcgacc cccaggcggc
841 ccaaccggc gccgcccag gctatgtggc cctcggcgc tctactcct cgggggtcgg
901 agcgggcagc tacatcagct cgagcggcga ctgcaagcgc agcacgaagg cccatcccta
961 cctgtgggcg gccgcccact cgcctccac gttcgaact accgcctgt cggcgcccg
1021 tacgggtgat gttctctccg gacagctcgg cccgctcagc tcggcaccg gcctcgtctc
1081 gatcagcatc ggcggaacg acgcccgttt cgcgacacc atgacgacct gtgtgctcca
1141 gtcgagagc tctgcctgt cgcgcatcg caccgcccag gcgtacgtc actcagcgt
1201 gcccggcaag ctgcagggc tctactcggc aatcagcgc aaggcggcga acgcccagc
1261 cgtcgtcacc ggctaccgc gcttctacaa gctcggcacc acctgcatc gcctgtccga
1321 gaccaagcgg acggcgatca acaaggcctc cgaccacctc aacaccgtcc tcgcccagcg
1381 cgcgcccgc cacggttca ccttcggcga cgtacgcacc accttcaccg gccacgagct
1441 gtgctccggc agcccctggc tgcacagcgt caactggctg aacatcggcg agtgcgtacca
1501 ccccaccgg gccggccagt ccggtggcta cctgcccgtc ctcaacggcg ccgctgacc
1561 tcaggcggaa ggagaagaag aaggagcgg gggagacgag gagtgggag ccccggcga
1621 cggggtccc gteccgctc ccgtctccg cccgggtccc caagtcaccg agaacgccac
1681 cgcgtcggac gtggcccga ccggactccg cacctccag cgcacggcac tctcgaacgc
1741 gccgggtgct tcgtgcgtc teaccaccac gccgtcctg cgcgagcgt cgcgcccga
1801 cgggaaggac agcgtccgcc accccggatc ggagaccgac ccgtccggc tcaccaccg
1861 gtagccgacc tccgcccga gcccccgc cgtgaacgtc gccgtgaac cgggtgcccg
1921 gtcgtcggc ggcgacagg cccccagta gtgggtgcg gagcccacca cggtcacctc
1981 caccgactgc gctgcggggc

```

图 38 (SEQ ID No. 45)

```

1 mrrsritayv tslllavgca ltgaataqas paaaatgyva lgdsyssvgv agsylsssgd
61 ckrsskaypy lwqaahspss fsfmacsgar tgdvlanqlg tlnsstglvs ltiggndagf
121 sdvmttcvlq sdsaclsrin takayvdstl pgqldsvyta istkapsahv avlgyprfyk
181 lggscлагls etkrasinda adylnsaiak raadhgftfg dvkstftghe icssstwlhs
241 ldllnigqsy hptaagqsgg ylpvmnsva

```

图 39 (SEQ ID No. 46)

```

1 ccaccgccgg gtcggcggcg agtctcctgg cctcggtcgc ggagaggttg gccgtgtagc
61 cgttcagcgc ggcgccgaac gtcttcttca ccgtgccgcc gtactcgttg atcaggccct
121 tgcccttgct cgacgcggcc ttgaagccgg tgcccttctt gagcgtgacg atgtagctgc
181 ccttgatcgc ggtgggggag ccggcggcga gcaccgtgcc ctcgccggg gtggcctggg
241 cgggcagtgc ggtgaatccg cccacgaggg cgccggtcgc cacggcgggt atcgccgga
301 tccgatctt cttgctacgc agctgtgcca tacgagggag tcctcctctg gccagcggcg
361 cgcctgggtg gggcgcacgg ctgtgggggg tgcgcgcgtc atcacgcaca cggccctgga
421 gcgctcgtgt cgcgccctgg ttgagtaaag cctcggccat ctacgggggt ggctcaaggg
481 agttgagacc ctgtcatgag tctgacatga gcacgcaatc aacggggccg tgagcaccct
541 ggggcgaccc cggaaaagtgc cgagaagtct tggcatggac acttcctgtc aacacgcgta
601 gctggtacga cggttacggc agagatcctg ctaaagggag gttccatgag acgttcccga
661 attacggcat acgtgacctc actcctcctc gccgtcggct gcgccctcac cggggcagcg
721 acggcgcagg cgtccccagc cgccgcggcc acgggctatg tggccctcgg cgactcgtac
781 tcgtccggtg tcggcggccg cagctacctc agctccagcg gcgactgcaa gcgcagttcg
841 aaggcctatc cgtacctctg gcaggcccg cattcacctt cgtcgttcag tttcatggct
901 tgctcggggc ctgtagcggg tgatgtcctg gccaatcagc tcggcaccct gaactcgtcc
961 accggcctgg tctccctcac catcgaggc aacgacgcgg gcttctccga cgtcatgacg
1021 acctgtgtgc tccagtcgga cagcgcctgc ctctcccgca tcaacacggc gaaggcgtac
1081 gtcgactcca ccctgcccg ccaactcgac agcgtgtaca cggcgatcag cacgaaggcc
1141 ccgtcggccc atgtggccgt gctgggctac ccccgttctt acaaactggg cggctcctgc
1201 ctgcggggcc tctcggagac caagcggctc gccatcaacg acgcccggca ctatctgaa
1261 agcggccatc ccaagcgcgc cgccgaccac ggcttcacct tcggcgacgt caagagcacc
1321 ttcaccggcc atgagatctg ctccagcagc acctggctgc acagtctcga cctgctgaac
1381 atcgccagt cctaccacc caccgcggcc ggccagtccg gcggctatct gccggctatg
1441 aacagcgtgg cctgagctcc cacggcctga atttttaagg cctgaatttt taaggcgaag
1501 gtgaaccgga agcggaggcc ccgtccgtcg gggctctcct cgcacaggtc accgagaacg
1561 gcacggagtt ggacgtcgtg cgcaccgggt cgcgcacctc gacggcgatc tcgttcgaga
1621 tcgttccgct cgtgtcgtac gtggtgacga acacctgctt ctgctgggtc tttccgccgc
1681 tcgccgggaa ggacagcgtc ttccagccc gatccgggac ctgcacctc ttggtcacc
1741 agcggtaact cacctcgacc ggcaccggc ccaccgtgaa ggtcggcgtg aacgtggcg
1801 cctggggcgt gggcggcggg caggcaccgg agtagtcggt gtgcacggc gtgaccgtca
1861 ccttcacgga ctgggcccgc ggggtcgtcg taccgccgcc gccaccgcc cctcccggag
1921 tggagcccga gctgtggtcg ccccgcctg cggcgttctc gtctcggg gttttcgaa

```

图 40 (SEQ ID No. 47)

```

1  mgsqpraatr rrlflgipal vlvtaltlvtl avptgretlw rmwceatqdw clgvpvdsrg
61  qpaedgefll lspvqaatwg nyyalgdsys sgdgardyyp gtavkggcwr sanaypelva
121 eaydfaghls flacsgqrgy amldaidevg sqldwnspht slvtigiggn dlqfstvlkt
181 cmvrvpllds kactdqedai rkrmakfett feelisevrt rapdarilvv gyprifpeep
241 tgayytltas nqrwlnetiq efnqqlaeav avhdeeiias ggvgsvfvd vyhaldghei
301 gsdepwvngv qlrdlatgvt vdrstfhpna aghravgerv ieqietgpgr plyatfavva
361 gatvdtlage vg

```

图 41 (SEQ ID No. 48)

```

1      ctgcagacac ccgccccgcc ttctcccga tcgtcatgtt cggcgactcc ctcagcgaca
61     ccggcaagat gtactccaag atgcgcggct acctgccgtc ctccccgcgc tactacgagg
121    gccgcttctc gaacggcccc gtctggctgg agcagctgac gaagcagttc cccggcctga
181    cgatcgccaa cgaggccgag gggggcgoga ccgcagtcgc ctacaacaag atctcctgga
241    acccgaagta ccaggtcatt aacaacctcg actacgaggt caccagttc ttgcagaagg
301    actcgttcaa gcccgacgac ctggteatcc tgtgggtggg cgccaacgac tacctggcct
361    acggttgaa cacggagcag gacgccaagc ggggtgcgga cgccatctcg gacgcggcaa
421    accgcatggt cctgaacggc gcgaagcaga tcctgctgtt caacctgccc gacctgggcc
481    agaaccgctc cgcccgtccc cagaaggctc tcgaggccgt ctgcacgtg tcgcctacc
541    acaacaagct gtcctcaac ctgccccgcc agctcgcccc gacgggcatg gtcaagctgt
601    tcgagatcga caagcagttc gcggagatgc tgcgcgacc ccagaacttc ggcctgagcg
661    acgtggagaa cccgtgctac gacggcggct acgtgtggaa gccgttcgcc acccggtccc
721    tctcgaccga ccggcagctg tcggcettct cgccccagga gcgcctggcg atcgctggca
781    accgctcct ggcacaggcg gtagcttcgc cgatggccc ccgctcggcc tcgcccctca
841    actgcgaggg caagatgttc tgggaccagg tccaccccac caccgtggtc cacgccgcc
901    tctcgagcgc cgccgccacc ttcacgaga cccagtacga gttcctcgcc cactagtcta
961    gaggatcc

```

图 42

1. L131
2. 灰色链霉菌 (*S. avermitilis*)
3. 褐色喜热裂孢菌 (*T. fusca*)
4. 共有序列

```

1          1                               50
1 (1) -----MRLTRSLSAASVIVFALLLALLGISPAQAAG-----
2 (1) -----MRRSRITAYVTSLLLAVGCALTGAATAQASPA-----
3 (1) VSGSPRAATRRLFLGIPALVLTALTLVLAVPTGRET LWRMWCEATQDW
4 (1)          MRRSRFLA  ALILLTLA  AL  GAA  ARAAP

          51                               100
1 (32) -----P-AYVALGDSYSSGNGAGSYID
2 (33) -----AAATGYVALGDSYSSGVGAGSYLS
3 (51) CLGVPVDSRGQPAEDGEFLLLSPVQAATWGNYYALGDSYSSGDGARDYYP
4 (51)          A A  YVALGDSYSSG  GAGSY

          101                              150
1 (53) SSGD---CHRNNAYPARWAAAANAP---SSFTFAACSGAVTTDVIN----
2 (57) SSGD---CKRSSKAYPYLWQAAHSP---SSFSEMACSGARTGDVLA----
3 (101) GTAVKGGCWRSANAYPELVAEAYDFA--GHLSFLACSGQRGYAMLDAIDE
4 (101) SSGD  C  RSTKAYPALWAAAHA      SSFSF  ACSGARTYDVLA

          151                              200
1 (93) --NQLGALNAST--GLVSIGGNDAGFADAMTTCVTS-----SDSTCL
2 (97) --NQLGTLNSST--GLVSLGGNDAGFSDVMTTCVLQ-----SDSACL
3 (149) VGSQLDWNSPHT--SLVTIGGNDLGFSTVLKTCMVR-----VPLLDS
4 (151)  QL  LNS  T   LVSITGGNDAGFAD  MTTCVL          SDSACL

          201                              250
1 (133) NRLATATNYINTLLA-----RLDAVYSQIKARAPNARVVVLGYPRMY
2 (137) SRINTAKAYVDSTLPG-----QLDSVYTAISTKAPSAHVAVLGYPRFY
3 (191) KACTDQEDAIRKRMAKF----ETTFEELISEVRTRAPDARILVVGYPRI
4 (201)  RIA  AK  YI  TLPA          RLDSVYSAI  TRAP  ARVVVLGYPRIY

          251                              300
1 (176) LASNPWYCLGLSNTKRAAINTTADTLNSVISSRATAH-----GF
2 (180) KLGG-SCLAGLSETKRSAINDAADYLNSAIKRAADH-----GF
3 (237) PEEPTGAYYTLTASNQRWLNETIQEFNQQLAEAVAVHDEEIAASGGVGSV
4 (251)  SG    LGLS  TKRAAINDAAD  LNSVIAKRAADH          GF

          301                              350
1 (215) RFGDVRPTFNNHELFFGNDWLHSLTLP-----VWESYH
2 (218) TFGDVKSTFTGHEICSSSTWLHSLDLLN-----IGQSYH
3 (287) EFVDVYHALDGHEIGSDEPWVNGVQLRDLATG-----VTVDRSTFH
4 (301) TFGDV  TF  GHELCSA  PWLHSLTLP          V  SYH

          351                              395
1 (248) PTSTGHQSGYLPVLNANSST-----
2 (252) PTAAGQSGGYLPVMNSVA-----
3 (328) PNAAGHRAVGERVIEQIETGPGRPLYATFAVVAGATVDTLAGEVG
4 (351) PTA  GHAAGYLPVLNSI  T

```

**图 43**

来自近平滑假丝酵母脂质酰基转移酶的氨基酸序列 SEQ ID No. 17;

```
MRYFAIAFLL INTISAFVLA PKKPSQDDFY TPPQGYEAQP LGSILKTRNV PNPLTNVFTP VKVQNAWQLL
VRSEDTFGNP NAIVTTHIQP FNAKKDKLVS YQTFEDSGKL DCAPSYAIQY GSDISTLTQ GEMYYISALL
DQGYVVTVD YEGPKSTFTV GLQSGRATLN SLRATLKSGN LTGVSSDAET LLWGYSGGSL ASGWAAAIQK
EYAPELSKNL LGAALGGFVT NITATAEAVD SGPFAGIISN ALAGIGNEYF DFKNYLLKKV SPLLSITYRL
GNTHCLLDGG IAYFGKSFFS
RIIRYFPDGW DLVNQEPIKT ILQDNGLVYQ PKDLTPQIPL FIYHGTLDAI VPIVNSRKTF QQWCDWGLKS
GEYNEDLTNG HITESIVGAP AALTWIINRF NGQPPVDGCQ HNVRSANLEY PGTPQSIKNY FEALHALIG
FDLGPDKRD KVTLGGLLKL ERFAP
```

**图 44**

来自近平滑假丝酵母脂质酰基转移酶的氨基酸序列 SEQ ID No. 18;

```
MRYFAIAFLL INTISAFVLA PKKPSQDDFY TPPQGYEAQP LGSILKTRNV PNPLTNVFTP VKVQNAWQLL
VRSEDTFGNP NAIVTTHIQP FNAKKDKLVS YQTFEDSGKL DCAPSYAIQY GSDISTLTQ GEMYYISALL
DQGYVVTVD YEGPKSTFTV GLQSGRATLN SLRATLKSGN LTGVSSDAET LLWGYSGGSL ASGWAAAIQK
EYAPELSKNL LGAALGGFVT NITATAEAVD SGPFAGIISN ALAGIGNEYF DFKNYLLKKV SPLLSITYRL
GNTHCLLDGG IAYFGKSFFS RIIRYFPDGW DLVNQEPIKT ILQDNGLVYQ PKDLTPQIPL FIYHGTLDAI
VPIVNSRKTF QQWCDWGLKS GEYNEDLTNG HITESIVGAP AALTWIINRF NGQPPVDGCQ HNVRSANLEY
PGTPQSIKNY FEALHALIG FDLGPDKRD KVTLGGLLKL ERFAPHHHHH H
```

图 45





图 46

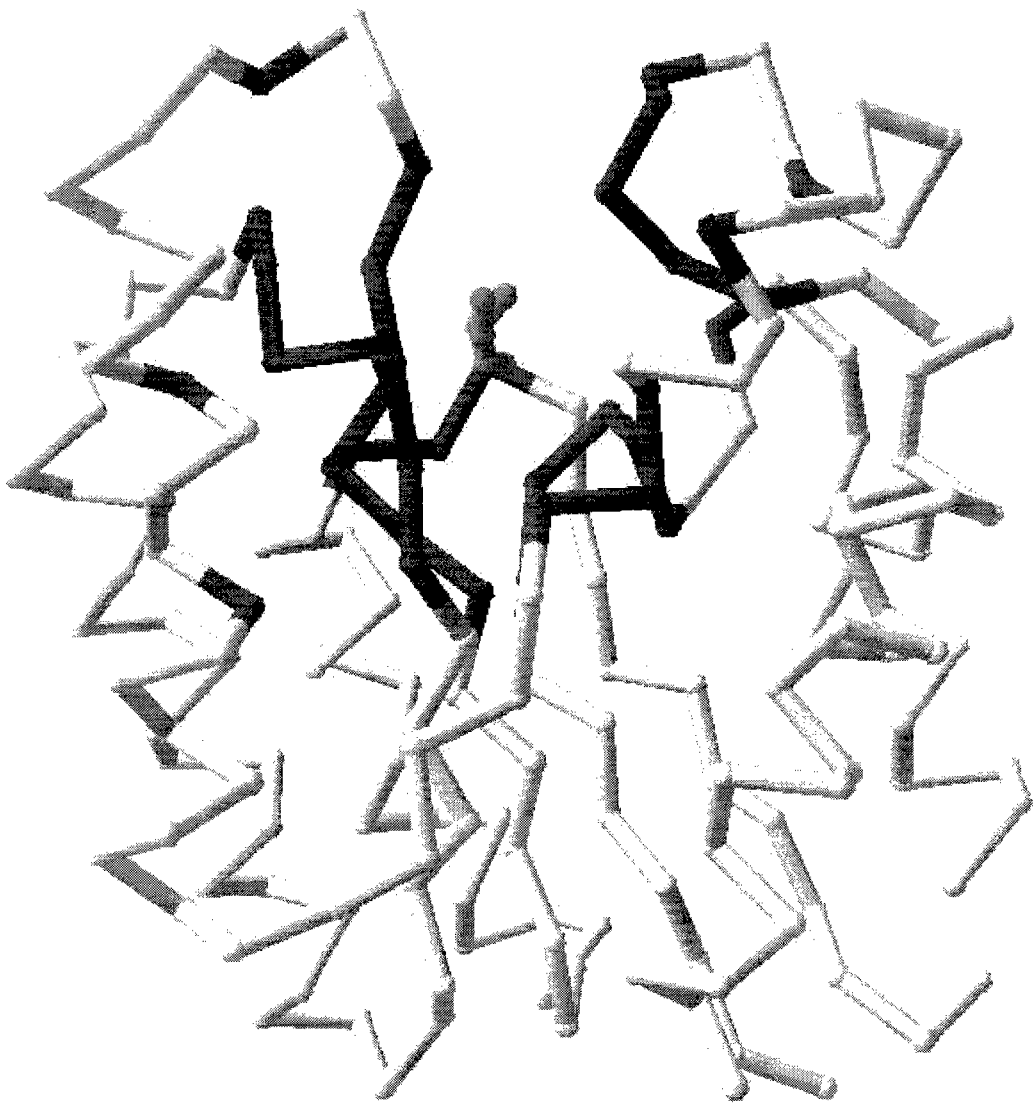


图 47

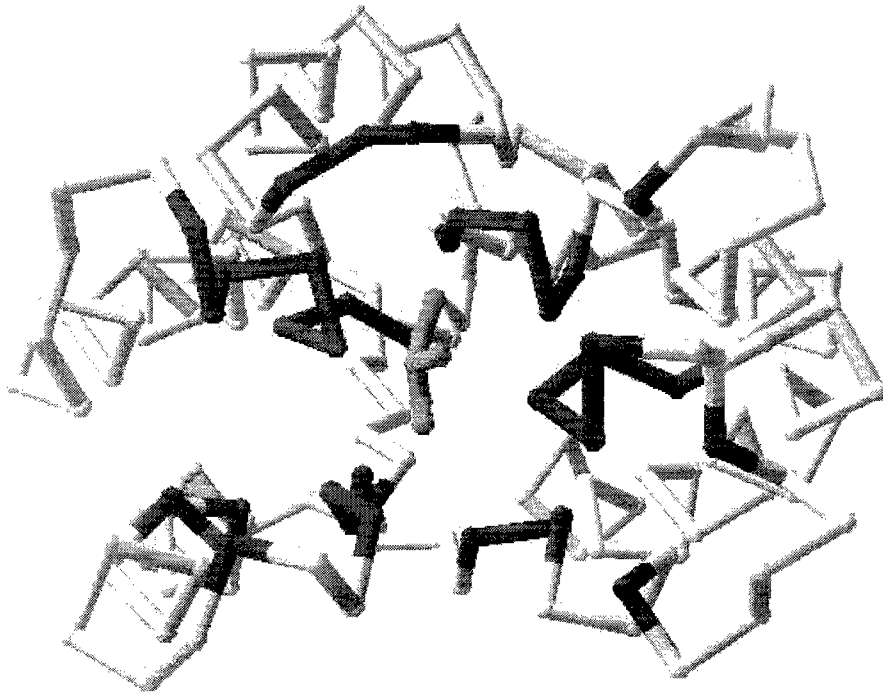










图 51

```

          10      20      30      40      50      60
      .....*.....|.....*.....|.....*.....|.....*.....|.....*.....|.....*.....|
1IVN_A  4  LLILGDSLSAG-----YRMSASAAWPALLNDKWqsk-----
34
P10480 28  IVMFGDSLSDTgkmyskmrgylpssppyyeGRFSNGPVWLEQLTNEFPGLTianaeggp
87

          70      80      90      100     110     120
      .....*.....|.....*.....|.....*.....|.....*.....|.....*.....|.....*.....|
1IVN_A 35  -tsvVNASISGDT-----SQOGLARLPALLKQHQP RW
65
P10480 88  tavaYNKISWNPkyq-----vINNLDYEVTQFLQKDSFKPDDL
125

          130     140     150     160     170     180
      .....*.....|.....*.....|.....*.....|.....*.....|.....*.....|.....*.....|
1IVN_A 66  VLVELGGNDG-----LRGFQPQOTEQT
87
P10480 126 VILWVGANDY-----LA--YGWNTAQDAKRVRDA
152

          190     200     210     220     230     240
      .....*.....|.....*.....|.....*.....|.....*.....|.....*.....|.....*.....|
1IVN_A 88  LRQILQDVkaANAEPllmqiRLPANYGR-----
115
P10480 153 ISDAANRMV-LNGAK-----EILLFNLPdlg-----qnP
180

          250     260     270     280     290     300
      .....*.....|.....*.....|.....*.....|.....*.....|.....*.....|.....*.....|
1IVN_A 116 -----RYNEAFSAIYPKLake-----fdVPLLPFFME
142
P10480 181 SARSQKVVEAASHVSAyHNQLLLNLArqlaptg-----mvklfeiDKQFAEMLRD
230

          310     320     330     340     350     360
      .....*.....|.....*.....|.....*.....|.....*.....|.....*.....|.....*.....|
1IVN_A 143 EVYLKPQW-----
150
P10480 231 PQNFGLSDQRNacyggsyvwkpfasrsastdsqf safnpqerlaiagnp llaqavaspma
290

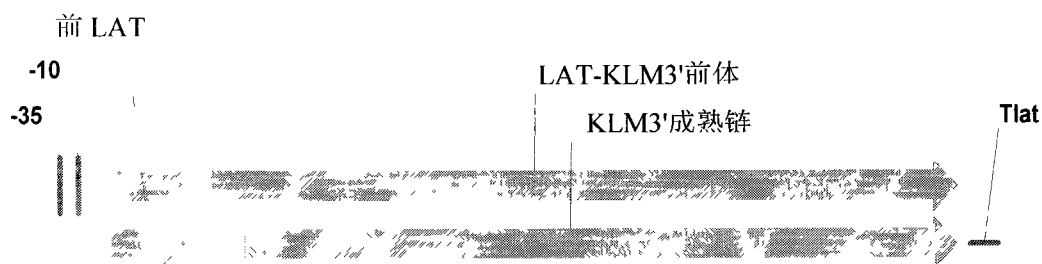
          370     380     390     400
      .....*.....|.....*.....|.....*.....|.....*.....|
1IVN_A 151 -----MQDDGI-----HPNRDAQPFADWM 170
P10480 291 arsastlncegkMFWDQV-----HPTTVVHAALSEPA 322

```

图 52

		1		50
P10480	(1)	MKKWFVCLLGLVALTVQAADSRPAFSRIVMFGDSLSDTGKMYSKMRGYLP		
杀鲟气单胞菌	(1)	-----ADTRPAFSRIVMFGDSLSDTGKMYSKMRGYLP		
嗜水气单胞菌	(1)	-----ADSRPAFSRIVMFGDSLSDTGKMYSKMRGYLP		
共有序列	(1)	AD*RPAFSRIVMFGDSLSDTGKMYSKMRGYLP		
		51		100
P10480	(51)	SSPPYYEGRFSNGPVWLEQLTNEFPGLTIANEAEGGPTAVAYNKISWNPK		
杀鲟气单胞菌	(33)	SSPPYYEGRFSNGPVWLEQLTKQFPGLTIANEAEGGATAVAYNKISWNPK		
嗜水气单胞菌	(33)	SSPPYYEGRFSNGPVWLEQLTKQFPGLTIANEAEGGATAVAYNKISWNPK		
共有序列	(51)	SSPPYYEGRFSNGPVWLEQLT*FPGLTIANEAEGG*TAAYNKISWNPK		
		101		150
P10480	(101)	YQVINNLDYEVTQFLQKDSFKPDDLVLWVGANDYLAYGWNTQDAKRVR		
杀鲟气单胞菌	(83)	YQVINNLDYEVTQFLQKDSFKPDDLVLWVGANDYLAYGWNTQDAKRVR		
嗜水气单胞菌	(83)	YQVINNLDYEVTQFLQKDSFKPDDLVLWVGANDYLAYGWNTQDAKRVR		
共有序列	(101)	YQVINNLDYEVTQFLQKDSFKPDDLVLWVGANDYLAYGWNTQDAKRVR		
		151		200
P10480	(151)	DAISDAANRMVLNGAKEILLFNLPDLGQNPSARSQKVVEAASHVSAYHNQ		
杀鲟气单胞菌	(133)	DAISDAANRMVLNGAKQILLFNLPDLGQNPSARSQKVVEAVSHVSAYHNK		
嗜水气单胞菌	(133)	DAISDAANRMVLNGAKQILLFNLPDLGQNPSARSQKVVEAVSHVSAYHNQ		
共有序列	(151)	DAISDAANRMVLNGAK*IILLFNLPDLGQNPSARSQKVVEA*SHVSAYHN*		
		201		250
P10480	(201)	LLLNLARQLAPTGMVKLFEIDKQFAEMLRDPQNFGLSDQRNACYGGSYVW		
杀鲟气单胞菌	(183)	LLLNLARQLAPTGMVKLFEIDKQFAEMLRDPQNFGLSDVENPCYDGGYVW		
嗜水气单胞菌	(183)	LLLNLARQLAPTGMVKLFEIDKQFAEMLRDPQNFGLSDVENPCYDGGYVW		
共有序列	(201)	LLLNLARQLAPTGMVKLFEIDKQFAEMLRDPQNFGLSD**N*CY*G*YVW		
		251		300
P10480	(251)	KPFASRSASTDSQLSAFNPQERLAIAGNPLLAQAVASPMAARSASTLNCE		
杀鲟气单胞菌	(233)	KPFATRSVSTDRQLSAFSPQERLAIAGNPLLAQAVASPMARRSASPLNCE		
嗜水气单胞菌	(233)	KPFATRSVSTDRQLSAFSPQERLAIAGNPLLAQAVASPMARRSASPLNCE		
共有序列	(251)	KPFA*RS*STD*QLSAF*PQERLAIAGNPLLAQAVASPMA*RSAS*LNCE		
		301		336
P10480	(301)	GKMFWDQVHPTTVVHAALSEPAATFIESQYEFLLAH-		
杀鲟气单胞菌	(283)	GKMFWDQVHPTTVVHAALSERAATFIETQYEFLLAHG		
嗜水气单胞菌	(283)	GKMFWDQVHPTTVVHAALSERAATFIANQYEFLLAH-		
共有序列	(301)	GKMFWDQVHPTTVVHAALSE*AATFI**QYEFLLAH*		

图 53



用以表达 KLM3'的基因构建体



图 54

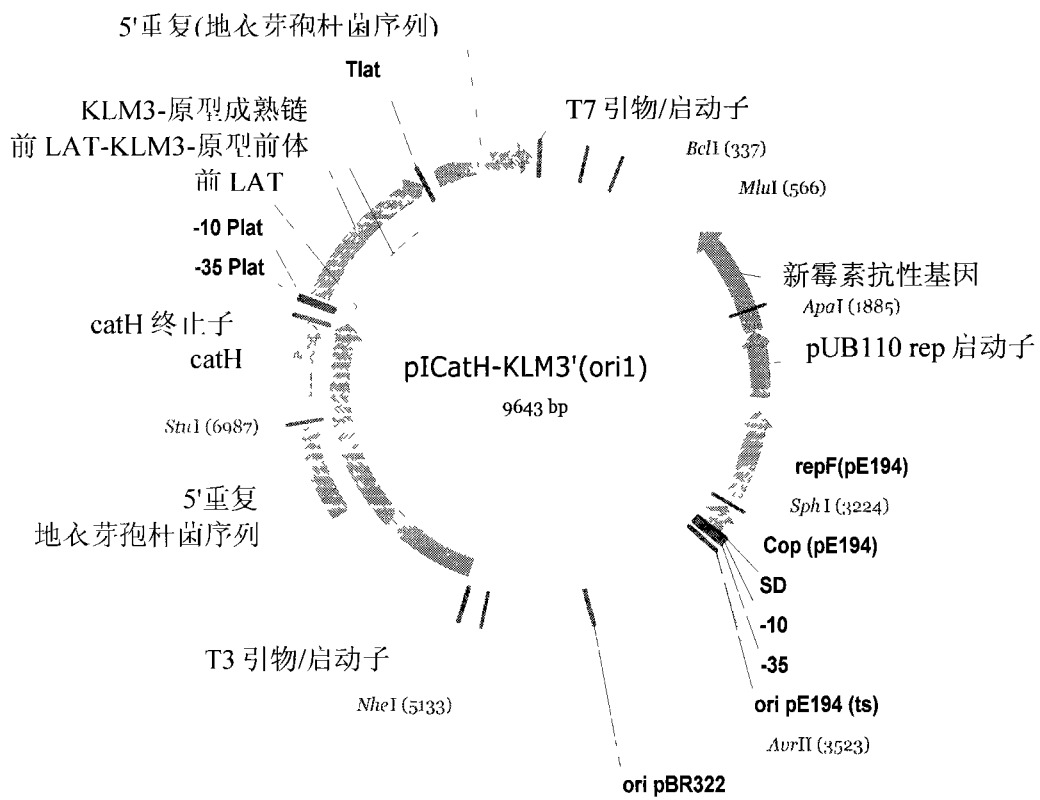


图 55

-35

```

1  GCTTTTCTTT TGGAAGAAAA TATAGGGAAA ATGGTACTTG TTA AAAAATTC GGAATATTTA
   CGAAAAGAAA ACCTTCTTTT ATATCCCTTT TACCATGAAC AATTTTAAAG CCTTATAAAT
   -10                                     M K Q Q K R L
51  TACAATATCA TATGTTTCAC ATTGAAAGGG GAGGAGAATC ATGAAACAAC AAAAACGGCT
   ATGTTATAGT ATACAAAGTG TAACTTTCCC CTCCTCTTAG TACTTTGTTG TTTTGGCCGA
   · Y A R L L T L L F A L I F L L P H S A A
21  TTACGCCCGA TTGCTGACGC TGTATTTGCG GCTCATCTTC TTGCTGCCTC ATTCTGCAGC
   AATGCGGGCT AACGACTGCG ACAATAAACG CGAGTAGAAG AACGACGGAG TAAGACGTCG
   · S A A D T R P A F S R I V M F G D S L S
33  TTCAGCAGCA GATACAAGAC CGGCGTTTAG CCGGATCGTC ATGTTTGGAG ATAGCCTGAG
   AAGTCGTCGT CTATGTTCTG GCGCAAAATC GGCCTAGCAG TACAAACCTC TATCGGACTC
   · D T G K M Y S K M R G Y L P S S P P Y Y
44  CGATACGGGG AAAATGTATA GCAAAATGAG AGGCTATCTT CCGTCAAGCC CGCCGTATTA
   GCTATGCCCC TTTTACATAT CGTTTACTC TCCGATAGAA GGCAGTTCGG GCGGCATAAT
   · E G R F S N G P V W L E Q L T K Q F P G
54  TGAAGGCCCG TTTAGCAATG GACCGGTCTG GCTGGAACAA CTGACGAAA AATTTCCGGG
   ACTTCCGGCG AAATCGTTAC CTGGCCAGAC CGACCTTGTG GACTGCTTTG TTAAGGCC
   · L T I A N E A E G G A T A V A Y N K I S
66  ACTGACGATC GCTAATGAAG CAGAAGGAGG AGCAACAGCG GTCGCCTATA ACAAATCAG
   TGACTGCTAG CGATTACTTC GTCTTCCCTC TCGTTGTGCG CAGCGGATAT TGTTTTAGTC
   · W D P K Y Q V I N N L D Y E V T Q F L Q
77  CTGGGACCCG AAATATCAGG TCATCAACAA CCTGGACTAT GAAGTCACAC AGTTTCTTCA
   GACCCTGGGC TTTATAGTCC AGTAGTTGTT GGACCTGATA CTTCAGTGTG TCAAAGAAGT
   · K D S F K P D D L V I L W V G A N D Y L
88  GAAAGACAGC TTTAAACCGG ATGATCTGGT CATCCTTTGG GTCGGCGCCA ATGATTATCT
   CTTTCTGTGCG AAATTTGGCC TACTAGACCA GTAGGAAACC CAGCCGCGGT TACTAATAGA
   · A Y G W N T E Q D A K R V R D A I S D A
99  GGCGTATGGC TGGAACACAG AACAAAGATG CAAAAGAGTC AGAGATGCCA TCAGCGATGC
   CCGCATACCG ACCTTGTGTC TTGTTCTACG GTTTTCTCAG TCTCTACGGT AGTCGCTACG
   · A N R M V L N G A K Q I L L F N L P D L
111  CGCTAATAGA ATGGTCTGA ACGGCGCCAA ACAAATCCTG CTGTTTAAAC TGCCGGATCT
   GCGATTATCT TACCAGGACT TGCCGCGGTT TGTTTAGGAC GACAAATGG ACGGCCTAGA
   · G Q N P S A R S Q K V V E A V S H V S A
122  GGGACAAAAT CCGAGCGCCA GAAGCCAAA AGTCGTGCGA GCAGTCAGCC ATGTCAGCGC
   CCCTGTTTTA GGCTCGGGT CTTCGGTTTT TCAGCAGCTT CGTCAGTCGG TACAGTCGG
   · Y H N K L L L N L A R Q L A P T G M V K
133  GTATCATAAC AAATGCTGC TGAACCTGGC AAGACAATG GCACCGACGG GAATGGTTAA
   CATAGTATTG TTTGACGAGC ACTTGGACCG TTCTGTTAAC CGTGGCTGCC CTTACCAATT
   · L F E I D K Q F A E M L R D P Q N F G L
144  ATGTTTGAAT ATTGACAAAC AGTTTGCCGA AATGCTGAGA GATCCGCAA ATTTTGGCCT
   TAACAACTT TAATGTTTG TCAAACGGCT TTACGACTCT CTAGGCGTTT TAAAACCGGA
   · S D V E N P C Y D G G Y V W K P F A T R
155  GAGCGATGTC GAAAACCCGT GCTATGATGG CCGATATGTC TGGAAACCGT TTGCCACAAG
   CTGCTACAG CTTTGGGCA CGATACTACC GCCTATACAG ACCTTTGGCA AACGGTGTTC
   · S V S T D R Q L S A F S P Q E R L A I A
166  AAGCGTCAGC ACGGATAGAC AACTGTGACG GTTTAGCCCG CAAGAAAGAC TGGAATCGC
   TTCGACGTCG TGCCTATCTG TTGACAGTCG CAAATCGGGC GTTCTTTCTG ACCGTTAGCG
   · G N P L L A Q A V A S P M A R R S A S P
177  CGGAAATCCG CTTTGGCAC AAGCAGTTGC TTCACCGATG GCAAGAAGAT CAGCAAGCCC
   GCCTTTAGGC GAAAACCGTG TTCGTCAACG AAGTGGTAC CGTCTTCTA GTCGTTCCGG
   · L N C E G K M F W D Q V H P T T V V H A
188  GCTGAATTGC GAAGGCAAAA TGTTTTGGGA TCAGGTCCAT CCGACAACAG TTGTCCATGC
   CGACTTAACG CTTCCGTTTT ACAAACCCCT AGTCCAGGTA GGCTGTTGTC AACAGGTACG
   · A L S E R A A T F I E T Q Y E F L A H G
199  TGCCCTTTCA GAAAGAGCGG CGACGTTTAT CGAAACACAG TATGAATTC TGGCCCATGG
   ACGGAAAGT CTTTCTCGCC GCTGCAATA GCTTGTGTG ATACTTAAAG ACCGGGTACC
   · stop
211  CTGAGTTAAC AGAGGACGGA TTTCTGAAG GAAATCCGTT TTTTATTTT AAGCTTGGAG
   GACTCAATTG TCTCTGCCT AAAGGACTTC CTTTAGGCAA AAAAATAAAA TTCGACCTC
222  ACAAGGTAAG GGATAAAACC TCGAG
   TGTTCCATTT CCTATTTTGG AGCTC

```

图 56



图 57 (SEQ ID No 49)

```

1   ATGAAACAAC AAAAACGGCT TTACGCCCGA TTGCTGACGC TGTTATTTGC
    TACTTTGTTG TTTTGGCCGA AATGCGGGCT AACGACTGCG ACAATAAACG

51  GCTCATCTTC TTGCTGCCTC ATTCTGCAGC TTCAGCAGCA GATACAAGAC
    CGAGTAGAAG AACGACGGAG TAAGACGTCG AAGTCGTCGT CTATGTTCTG

101 CGGCGTTTAG CCGGATCGTC ATGTTTGAG ATAGCCTGAG CGATACGGGC
    GCCGAAATC GGCTAGCAG TACAAACCTC TATCGGACTC GCTATGCCCG

151 AAAATGTATA GCAAATGAG AGGCTATCTT CCGTCAAGCC CGCCGTATTA
    TTTTACATAT CGTTTACTC TCCGATAGAA GGCAGTTCGG GCGGCATAAT

201 TGAAGGCCGC TTTAGCAATG GACCGGTCTG GCTGGAACAA CTGACGAAAC
    ACTTCCGGCG AAATCGTTAC CTGGCCAGAC CGACCTTGTT GACTGCTTTG

251 AATTTCCGGG ACTGACGATC GCTAATGAAG CAGAAGGAG AGCAACAGCG
    TTAAAGGCC TGACTGCTAG CGATTACTTC GTCTTCCTCC TCGTTGTGCG

301 GTCGCCTATA ACAAATCAG CTGGGACCCG AAATATCAGG TCATCAACAA
    CAGCGGATAT TGTTTTAGTC GACCCGCGGC TTTATAGTCC AGTAGTTGTT

351 CCTGGACTAT GAAGTCACAC AGTTTCTTCA GAAAGACAGC TTTAAACCGG
    GGACCTGATA CTTCAAGTGT TCAAAGAAGT CTTTCTGTCG AAATTTGGCC

401 ATGATCTGGT CATCCTTTGG GTCGGCGCCA ATGATTATCT GCGGTATGGC
    TACTAGACCA GTAGGAAACC CAGCCGCGGT TACTAATAGA CCGCATACCG

451 TGGAACACAG AACAAAGTGC CAAAAGAGTC AGAGATGCCA TCAGCGATGC
    ACCTTGTGTC TTGTCTACG GTTTCTCAG TCTCTACGGT AGTCGCTACG

501 CGCTAATAGA ATGGTCCTGA ACGGCGCCAA ACAAATCCTG CTGTTTAAAC
    GCGATTATCT TACCAGGACT TGCCGCGGTT TGTTTAGGAC GACAAATTGG

551 TGCCGGATCT GGGACAAAT CCGAGCGCCA GAAGCCAAA AGTCGTGAA
    ACGGCTAGA CCCTGTTTTA GGCTCGCGGT CTTGCGTTT TCAGCAGCTT

601 GCAGTCAGCC ATGTCAGCGC CTATCATAAC AACTGCTGC TGAACCTGGC
    CGTCAGTCGG TACAGTCGCG GATAGTATG TTTGACGACG ACTTGGACCG

651 AAGACAATTG GCACCGACGG GAATGGTTAA ATTGTTTGAA ATTGACAAAC
    TTCTGTTAAC CGTGGCTGCC CTTACCAATT TAACAAACTT TAACTGTTG

701 AGTTTGCCGA AATGCTGAGA GATCCGCAA ATTTTGGCCT GAGCGATGC
    TCAAACGGCT TTACGACTCT CTAGGCGTTT TAAAACCGGA CTCGCTACAG

751 GAAAACCCGT GCTATGATGG CGGATATGTC TGGAAACCGT TTGCCACAAG
    CTTTGGGCA CGATACTACC GCCTATACAG ACCTTTGGCA AACGGTGTTT

801 AAGCGTCAGC ACGGATAGAC AACTGTCAGC GTTTAGCCCG CAAGAAAGAC
    TTCGAGTCG TGCCATCTG TTGACAGTCG CAAATCGGGC GTTCTTTCTG

851 TGGCAATCGC CGGAAATCCG CTTTGGCAC AAGCAGTTC TTCACCGATG
    ACCGTTAGCG GCCTTAGGC GAAAACCGTG TCGTCAACG AAGTGGCTAC

901 GCAAGAAGAT CAGCAAGCCC GCTGAATTGC GAAGGCAAAA TGTTTGGGA
    CGTTCTTCTA GTCGTTCCGG CGACTTAACG CTTCCGTTTT ACAAACCTT

951 TCAGGTCCAT CCGACAACAG TTGTCCATGC TGCCCTTTC GAAAGAGCGG
    AGTCCAGGTA GGCTGTTGTC AACAGGTACG ACGGGAAAGT CTTTCTCGCC

1001 CGACGTTTAT CGAAACACAG TATGAATTC TGGCCATGG CTGA
    GCTGCAAATA GCTTTGTGTC ATACTTAAAG ACCGGGTACC GACT

```

图 58 (SEQ ID No. 50)

```

1  ATGAAAAAAT GGTTTGTGTG TTTATTGGGA TTGGTCGCGC TGACAGTTCA GGCAGCCGAC
61  AGCCGTCCCG CCTTCTCCCG GATCGTGATG TTTGGCGACA GCCTCTCCGA TACCGGCAAG
121 ATGTACAGCA AGATGCGCGG TTACCTCCCC TCCAGCCCCC CCTACTATGA GGGCCGCTTC
181 TCCAACGGGC CCGTCTGGCT GGAGCAGCTG ACCAACGAGT TCCCGGGCCT GACCATAGCC
241 AACGAGCGG AAGGCGGACC GACCGCCGTG GCTTACAACA AGATCTCCTG GAATCCCAAG
301 TATCAGGTCA TCAACAACCT GGACTACGAG GTCACCCAGT TCCTGCAAAA AGACAGCTTC
361 AAGCCGGACG ATCTGGTGAT CCTCTGGGTC GGCGCCAACG ACTATCTGGC CTATGGCTGG
421 AACACAGAGC AGGATGCCAA GCGGTGCGC GACGCCATCA GCGATGCGGC CAACCGCATG
481 GTGCTGAACG GCGCCAAGGA GATACTGCTG TTCAACCTGC CGGATCTGGG CCAGAACCCC
541 TCGCCCCGCA GCCAGAAGGT GGTGAGGCG GCCAGCCATG TCTCCGCTA CCACACCAG
601 CTGCTGCTGA ACCTGGCAGC CCAGCTGGCT CCCACCGGCA TGGTGAAGCT GTTCGAGATC
661 GACAAGCAGT TTGCCGAGAT GCTGCGTGAT CCGCAGAACT TCGGCCTGAG CGACCAGAGG
721 AACCCCTGCT ACGGTGGCAG CTATGTATGG AAGCCGTTTG CCTCCCGCAG CGCCAGCACC
781 GACAGCCAGC TCTCCGCTT CAACCCGAG GAGCGCTCG CCATCGCCGG CAACCCGCTG
841 CTGGCCAGG CCGTCGCCAG CCCCATGGCT GCCCGCAGCG CCAGCACCTT CAACTGTGAG
901 GGCAAGATGT TCTGGGATCA GGTCCACCCC ACCACTGTG TGCACGCCGC CCTGAGCGAG
961 CCCGCCGCA CCTTCATCGA GAGCCAGTAC GAGTTCCTCG CCCAC

```

图 59 (SEQ ID No. 51)

```

1  ATGAAAAAAT GGTTTGTTTG TTTATTGGGG TTGATCGCGC TGACAGTTCA GGCAGCCGAC
61  ACTCGCCCCG CCTTCTCCCG GATCGTGATG TTCGGCGACA GCCTCTCCGA TACCGGCAAA
121 ATGTACAGCA AGATGCGCGG TTACCTCCCC TCCAGCCCGC CCTACTATGA GGGCCGTTTC
181 TCCAACGGAC CCGTCTGGCT GGAGCAGCTG ACCAAGCAGT TCCCGGTCT GACCATCGCC
241 AACGAAGCGG AAGGCGGTGC CACTGCCGTG GCTTACAACA AGATCTCCTG GAATCCCAAG
301 TATCAGGTCT ACAACAACCT GGACTACGAG GTCACCCAGT TCTTGCAAAA AGACAGCTTC
361 AAGCCGGACG ATCTGGTGAT CCTCTGGGTC GGTGCCAATG ACTATCTGGC ATATGGCTGG
421 AATACGGAGC AGGATGCCAA GCGAGTTCGC GATGCCATCA GCGATGCGGC CAACCGCATG
481 GFACTGAACG GTGCCAAGCA GATACTGCTG TTCAACCTGC CGGATCTGGG CCAGAACCCG
541 TCAGCCCCGA GTCAGAAGGT GGTGAGGCG GTCAGCCATG TCTCCGCTA TCACAACAAG
601 CTGCTGCTGA ACCTGGCAGC CCAGCTGGCC CCCACCGGCA TGGTAAAGCT GTTCGAGATC
661 GACAAGCAAT TTGCCGAGAT GCTGCGTGAT CCGCAGAACT TCGGCCTGAG CGACGTCGAG
721 AACCCCTGCT ACGACGGCGG CTATGTATGG AAGCCGTTTG CCACCCGAG CGTCAGCACC
781 GACCGCCAGC TCTCCGCTT CAGTCCGAG GAACGCCTCG CCATCGCCGG CAACCCGCTG
841 CTGGCACAGG CCGTTGCCAG TCTATGGCC GCGCGCAGCG CCAGCCCTT CAACTGTGAG
901 GGCAAGATGT TCTGGGATCA GGTACACCCG ACCACTGTG TGCACGCAGC CCTGAGCGAG
961 CGCGCCGCA CCTTCATCGA GACCCAGTAC GAGTTCCTCG CCCACGGATG A

```

图 60 (SEQ ID No. 52)

```

1  ATGCCGAAGC CTGCCCTTCG CCGTGTCATG ACCGCGACAG TCGCCGCCGT CGGCACGCTC
61  GCCCTCGGCC TCACCGACGC CACCGCCAC GCCGCGCCG CCCAGGCCAC TCCGACCCTG
121 GACTACGTCG CCCTCGGCGA CAGCTACAGC GCCGGCTCCG GCGTCTGCG CGTCGACCCC
181 GCCAACCTGC TCTGTCTGGC CTCGACGGCC AACTACCCCG ACGTCATCGC GGACACGACG
241 GCGGCCCGCC TCACGGACGT CACCTGCGGC GCCGCGCAGA CCGCCACTT CACGCGGGCC
301 CAGTACCCGG CCGTCGCACC CCAGTTGGAC GCGTCCGCA CCGGCACGGA CCTGGTCACG
361 CTCACCATCG GCGGCAACGA CAACAGCACC TTCATCAACG CCATCACGGC CTGCGGCACG
421 GCGGTGTGCC TCAGCGGCGG CAAGGGCAGC CCCTGCAAGG ACAGGCACGG CACCTCCTTC
481 GACGACGAGA TCGAGGCCAA CAGTACCCC GCGCTCAAGG AGGCGTGTCT CGGCGTCCGC
541 GCCAGGGCTC CCCACGCCAG GGTGGCGGCT CTCGGCTACC CGTGGATCAC CCCGGCCACC
601 GCCGACCCGT CCTGCTTCCT GAAGCTCCCC CTCGCGCCG GTGACGTGCC CTACCTGCGG
661 GCCATCCAGG CACACCTCAA CGACGCGGTC CCGCGGGCCG CCAGGAGAC CGGAGCCACC
721 TACGTGGACT TCTCCGGGTT GTCCGACGGC CACGACGCTT GCGAGGCCCC CGGCACCCGC
781 TGGATCGAAC CCGTGTCTTT CGGCACAGC CTCGTTCCCG TCCACCCCAA CGCCTTGGGC
841 GAGCGCGCA TGGCCGAGCA CACGATGGAC GTCCTCGGCC TGGACTGA

```

图 61 (SEQ ID No. 53)

```

1   TCAGTCCAGG CCGAGGACGT CCATCGTGTG CTCGGCCATG CGCCGCTCGC CCAGGGCGTT
61  GGGGTGGACG GGAACGAGGC TGTGCCCGAA GAGCAGCGGT TCGATCCAGC GGGTGCCGGG
121 GGCTTCGCAG GCCTCGTGGC CGTCGGACAC CCCGGAGAAG TCCACGTAGG TGGCTCCGGT
181 CTCTCGGCGG GCCCGCCGGA CCGCGTCGTT GAGGTGTGCC TGGATGGCCC GCAGGTAGGG
241 CACGTACCCG GCGGCGAGGG GGAGCTTCAG GAAGCAGGAC GGGTCGGCGG TGGCCGGGGT
301 GATCCACGGG TAGCCGAGAG CCGCCACCCT GGCGTGGGGA GCCCTGGCGC GGACGCCGAG
361 CAGCGCCTCC TTGAGCGCGG GGTACGTGTT GGCTCGATC TCGTCTCGA AGGAGGTGCC
421 GTGCCTGTCC TTGCAGGGGC TGCCTTGCC GCCGCTGAGG ACACCCGCCG TGCCCGAGGC
481 CGTGATGGCG TTGATGAAGG TGCTGTTGTC GTTGCCCGCG ATGGTGAGCG TGACCAGGTC
541 CGTGCCGGTG CCGAGCGCGT CCAACTGGGG TCGCAGCCCC GGGTACTGGG CCCGCGTGAA
601 GTCGGCGGTC TCGCGGGCGC CGCAGGTGAC GTCCGTGAGG CGGGCGCCCC TCGTGTCCGC
661 GATGACGTGG GGTAGTTGG CCGTCCGAGC CAGACAGAGC AGGTTGGCGG GGTCCGCGGG
721 CAGGACGCCG GAGCCGGCGC TGTAGCTGTC GCCGAGGGCG ACGTAGTCCA GGGTCGGAGT
781 GGCCTGGGCG GGGCGGGCGT GGGCGGTGGC GTCCGTGAGG CCGAGGGCGA GCGTCCCGAC
841 GGCGGCGACT GTCGCGGTCA TGACACGGCG AAGGGCAGGC TTCGGCAT

```

图 62 (SEQ ID No. 54)

```

1   ATGGATTACG AGAAGTTTCT GTTATTTGGG GATTCCATTA CTGAATTGC TTTTAATACT
61  AGGCCATTG  AAGATGGCAA AGATCAGTAT GCTCTTGGAG CCGCATTAGT CAACGAATAT
121 ACAGAAAAAA TGGATATTCT TCAAAGAGGG TTCAAAGGGT ACACTTCTAG ATGGGCGTTG
181 AAAATACTTC CTGAGATTTT AAAGCATGAA TCCAATATTG TCATGGCCAC AATATTTTTG
241 GGTGCCAACG ATGCATGCTC AGCAGGTCCC CAAAGTGTCC CCTCCCGCA ATTTATCGAT
301 AATATTCGTC AAATGGTATC TTTGATGAAG TCTTACCATA TCCGTCTAT TATAATAGGA
361 CCGGGGCTAG TAGATAGAGA GAAGTGGGAA AAAGAAAAAT CTGAAGAAAT AGCTCTCGGA
421 TACTTCCGTA CCAACGAGAA CTTTGCCATT TATTCCGATG CCTTAGCAA ACTAGCCAAT
481 GAGGAAAAAG TTCCCTTCGT GCCTTTGAAT AAGGCCTTTC AACAGGAAGG TGGTGATGCT
541 TGGAACAAC  TGCTAACAGA TGGACTGCAC TTTTCCGAA AAGGGTACAA AATTTTTTCAT
601 GACGAATTAT TGAAGTTCAT TGAGACATTC TACCCCAAT ATCATCCAA AAACATGCAG
661 TACAAACTGA AAGATTGGAG AGATGTGCTA GATGATGGAT CTAACATAAT GTCTTGA

```

图 63 (SEQ ID No. 55)

```

atgaacctgc gtcaatggat gggcgccgcc acggetgccc ttgccttggg cttggccgcg          60
tgcgggggcy gtdggaccga ccagagcggc aatcccaatg tcgccaaggt gcagcgcag          120
gtgtgtttcg gcgacagcct gacgatatc ggcacctaca cccccgtcgc gcagcgggtg          180
ggcgcgggca agttcaccac caaccggggc cgatctggg ccgagaccgt ggccgcgcaa          240
ctgggctgta cgctcacgcc ggcggtgatg ggctacgcca cctccgtgca gaattgcccc          300
aaggccggct gcttcgacta tgcgcagggc ggctcgcgcg tgaccgatcc gaacggcatc          360
ggccacaacg gcgcgcgggg ggcgctgacc taccgggttc agcagcagct cgccaacttc          420
tacgcgcca gcaacaacac attcaacggc aataacgatg tcgtcttcgt gctggccggc          480
agcaacgaca ttttcttctg gaccactgcy gcgccacca gcggtccgg cgtgacgccc          540
gccattgcca cggcccaggt gcagcaggcc gcgacggacc tggctggcta tgtcaaggac          600
atgatcgcca aggtgcgac gcaggtctac gtgttcaacc tgcccgaag cagcctgacg          660
ccggacggcg tggcaagcgg cagcaccggc caggcgctgc tgcaagcgt ggtgggcacg          720
ttcaacacga cgctgcaaag cgggctggcc ggcacctcgg cgcgatcat cgacttcaac          780
gcacaactga ccgcgcgcat ccagaatggc gctcgttcg gcttcgcaa caccagcgcc          840
cgggctgcy acgccaacaa gatcaatgcc ctggtgccga gcgcccggcg cagctcgtg          900
ttctgctcgg ccaaacgct ggtggttcc ggtgcgacc agagctacct gttcgccgac          960
ggcgtgacc cgaccagggc cggccatcgc ctgatcgcca gcaacgtgct ggcgcgctg          1020
ctggcggata acgtcgcgca ctga          1044

```

图 64 (SEQ ID No. 56)

```

1  gtgatcgggt  cgtacgtggc  ggtgggggac  agcttcaccg  agggcgtcgg  cgaccccggc
61  cccgacgggg  cgttcgtcgg  ctgggcccgc  cggctcggcg  tactgctcgc  ggaccggcgc
121  cccgagggcg  acttcaagta  cacgaacctc  gccgtgcggc  gcaggctcct  cgaccagatc
181  gtggcggaac  aggtcccgcg  ggtcgtcggc  ctccgcccgc  acctcgtctc  gttcgcggcg
241  ggcggcaacg  acatcatccg  gcccgccacc  gatcccgcag  aggtcggcga  gcggttcggg
301  ctggcgggtg  ccgcgctgac  cgcgcggccc  ggaacctcct  tggtgaccac  cgggttcggc
361  acccgggggg  tgcccgtcct  caagcacctg  cgcggcaaga  tcgccaccga  caacgggcac
421  gtccgcgcca  tcgcccaccg  ctacggctgc  ccgggtgctc  acctgtggtc  gctgcggagc
481  gtccaggacc  gcagggcgtg  ggacgcccgc  cggctgcacc  tgtcgcggga  ggggcacacc
541  cgggtggcgc  tcgcccggcg  gcaggccctg  ggctgcggcg  tcccggccga  ccctgaccag
601  ccttggccgc  cctgcggccc  gcgcccgcag  ctgcagctcc  ggcgcgacga  cgtgactggg
661  gcgcccagat  acctggtgcc  gtgatcggg  cgcggctgc  ggggcgagtc  gtcggcggac
721  cactgacgag  ccaaggggac  gctgtcggcg  gacgccatca  agacgcggat  cgccgcgggt
781  gcctga

```

图 65 (SEQ ID No. 57)

```

1  atgcagacga  accccgcgta  caccagtctc  gtcgcccgtc  gcgactcctt  caccgagggc
61  atgtcggacc  tgctgcccga  cggctcctac  cgtggctggg  ccgacctcct  cgccaccggg
121  atggcggccc  gctcccccg  ctccgggtac  gccaacctgg  cgggtgcggc  gaagctgatc
181  ggacagatcg  tcgacgagca  ggtggacgtg  gccgcccga  tgggagccga  cgtgatcacg
241  ctggtcggcg  ggtcaacga  cagctgcgg  cccaagtgcg  acatggcccg  ggtgcccggc
301  ctgctgacc  aggcgtgga  acggctcgc  ccgcaactgc  agcagctggt  gctgatgcgc
361  agtcccggtc  gccagggtcc  ggtgctggag  cgcttcggc  cccgcatgga  ggcctgttc
421  gccgtgatcg  acgacctg  cggcggccac  ggcccgctg  tcgtcgcact  gtaccggggc
481  cagtccgtg  ccgacctc  gatgtgggac  gtggaccggc  tgcacctgac  cgccgagggc
541  caccgcccgg  tcgcccggcg  ggtgtggcag  tcgctcggcc  acgagcccga  ggaccggag
601  tggcacgccc  cgatcccgg  gacgcccgg  ccgggggtgg  tgacgcccag  gaccgcccag
661  gtccggttcg  cccggcagca  cctgctgccc  tggataggcc  gcaggctgac  cgggcccgtc
721  tccggggcag  gctgcccgg  caagcggccc  gacctgctgc  cctacgagga  ccccgcacgg
781  tga

```

图 66 (SEQ ID No. 58)

```

1  atgaccggg  gtcgtgacg  ggtgcccgg  gcgccccca  ccaagcacg  tgcctgctc
61  gcggcgatc  tcaccctgat  agtggcgatc  tccgcccga  tatacggcg  agcgtcccg
121  gacgacggca  gcagggacca  cgcgctgac  gccggaggcc  gtctcccag  aggagacgcc
181  gcccccgcgt  ccaccggtgc  ctgggtggcg  gcctgggcca  ccgaccggc  cgcggccgag
241  ccgggcaccg  agacgaccgg  cctggcgggc  cgctccgtgc  gcaacgtcgt  gcacacctc
301  gtcggcggca  ccggcggcgg  gatcaccctc  tcgaaacctg  acgggcagtc  gccgctgacc
361  gtcacacacg  cctcgatcgc  cctggcccgc  gggcccgaca  ccgcccggc  gatcggccg
421  accatgccc  ggtcacctt  cggcggcagc  gccgggtg  tcatcccgg  gggcggccag
481  gtgatgagcg  acaccgccc  cctcggcacc  ccctacgggg  cgaacgtcct  ggtcaccag
541  tactccccca  tcccgtccgg  gccggtgacc  taccatccgc  aggcccggca  gaccagctac
601  ctggccgacg  gcgaccgcac  ggcggacgtc  accgcccgtc  cgtacaccac  ccccacggcc
661  tactgcccgt  acctgaccgc  cctcagcgtg  ctgagccacg  aggcggcgg  caggtcgtg
721  gcgttcggcg  actccatcac  cgacggcggc  cgctcgcaga  gcgacgcca  ccaccgctgg
781  accgacgtcc  tcgcccgcag  cctgcaacgag  gcggcgggg  acggcccgg  cagcccggc
841  tacagcgtcg  tcaacgaggg  catcagcggc  aaccggctcc  tgaccagcag  gccggggcgg
901  ccggcccaca  acccgagcgg  actgagcccg  ttccagcgg  acgtgctgga  acgaccaaac
961  gteaaggccg  tcgtcgtcgt  cctcggcgtc  aacgacgtcc  tgaacagccc  ggaactcggc
1021  gaccgcgacg  ccattcctgac  cggcctgccc  accctcgtcg  accggggcga  cgcccgggga
1081  ctgcccggctg  tcggcggccc  gatcacgccc  ttggcgggct  acggcgggta  caccgagggc
1141  cgcgagacga  acccgacgga  ggtcaacgag  gagatccgct  ccggcccgg  cttcgacacg
1201  gtcgctgact  tcgacaaggc  cctcggcggc  ccgtacgacc  cgcgcccgg  gcgctcccac
1261  tacgacagcg  gcgaccacct  gcaccccgg  gacaaggggt  acgcccggat  gggcggcggt
1321  atgcacctgg  ccgcccgtgaa  gggcggcggc  ccggctcaagg  cgtag

```

图 67 (SEQ ID No. 59)

```

1 atgacgagca tgtcgagggc gagggtggcg cggcggatcg cggccggcgc ggcgtacggc
61 ggcggcggca tggcctggc gggagcggcg gcggtcggtc tgggtggggc cgaggtgcag
121 ctggccagac gcaggggtggg ggtggggcac ccgacccggg tgccgaacgc gcagggactg
181 tacggcggca ccctgccac ggcggcgac ccgcccgtgc ggctgatgat gctggggcgc
241 tccacggccg cggggcaggg cgtgcaccgg gccgggcaga cggcggcgc gctgctggcg
301 tccgggctcg cggcgggtggc ggagcggccg gtgcggctgg ggtcggtcgc ccagccgggg
361 gcgtgctcgg acgacctgga ccggcaggtg gcgctggtgc tcgcccagcc ggaccgggtg
421 cccgacatct gcgtgatcat ggtcggcgc aacgacgtca ccaccggat gccggcgacc
481 cgctcgggtc gccacctgtc ctccgggta cggcggctgc gcacggcgg tggggagggtg
541 gtggtcggca cctgtccgga cctggggcac atcgagcggg tcgggcagcc gctgcgctgg
601 ctggcccggc gggcctcacg gcagctcgc gcggcacaga ccatcggcgc cgtcgagcag
661 ggcggcgca cgtgtcgcct gggcgacctg ctgggtccgg agttcgcgca gaaccggcg
721 gagctcttcg gccccgacaa ctaccaccc tcggccgagg ggtacgccac ggcggcgatg
781 gcggtactgc cctcgggtgt cggcgcgctc ggctgtggc cggccgacga ggagcaccg
841 gacgcgctgc gccgcgaggg ctctctcgc gtggcgcgcg cggcggcgga ggcggcgctc
901 gagcgggta cggaggtcgc cggcccatg cctacggggc ctggggggc ctggggcgtg
961 ctgaagcgc ggagacggcg tcgggtgctg gaggcggaac cgtccagccc gtcggcgctt
1021 tga

```

图 68 (SEQ ID No. 60)

```

1 atgggtcgag ggacggacca gcggacggc taccggcgtc gccggggcg tgtcgcgctc
61 gccgccctga ccgcccggc cctggggcgt ggctggcgg gctgcgactc cgtggggcgc
121 gactcaccgc ctctctccg cagcccgtcg aagcggacga ggacggcgc ccgctgggac
181 accagccggc cgtccgctgc cgcggtggc gactccatca cgcggcgtt cgacgcctgt
241 gcggtgctgt cggactgccc ggaggtgctg tggggcagc gcagcagcgc gaaggtcgac
301 tcgctggccg tacggctgct ggggaaggcg gacgcggccg agcacagctg gaactacgcg
361 gtcaccgggg cccggatggc ggacctgacc gctcaggtga cgcggggcgc gcagcgcgag
421 ccggagctgg tggcgggtgat ggcggggcg aacgacgct gccgggtccac gaacctggcg
481 atgacgcggc tggcggactt ccggggcag ttcgaggagg cgatggccac cctgcgcaag
541 aagctcccca agcgcaggt gtacgtgctg agcatcccgc acctcaagcg gctctggtcc
601 cagggccgca ccaaccgctc gggcaagcag gtgtggaagc tcggcctgtg cccgctgatg
661 ctggggcagc cggactccct ggactcggcg gcgaccctgc ggcgcaacac ggtgcgcgac
721 cgggtggcgg actacaacga ggtgctgcgg gaggtctgcg cgaaggaccg gcggtgcccg
781 agcgaacgag gcggtgca cgaattccgg ttcggcacgg accagttgag ccactgggac
841 tggttccacc cgagtgtgga cggccaggcc cggctggcgg agatgccta ccgcccggc
901 acccgaaga atccctga

```

图 69 (SEQ ID No. 61)

```

1 ttcatacaaa cgatgtcaca acaccggcca tccgggtcat ccctgatcgt gggaatgggt
61 gacaagcctt cccgtgacga aagggtcctg ctacatcaga aatgacagaa atcctgctca
121 gggaggttcc atgagactgt cccgacgcgc ggccacggcg tccgctcc tcctcaccoc
181 ggcgctcgcg ctcttcggcg cgagcgcgc cgtgtccgc ccgcaatcc aggccaccga
241 ctacgtggcc ctccggcact cctactctc ggggtcggc gcgggagct acgacagcag
301 cagcttccgc tgtaagcga gcaccaagtc ctaccggcc cgtggggcg cctcgcaac
361 cggtaacggc tcaacttca ccgctgttc gggcggccc acaggagacg tctgggcaa
421 gcagctgacc ccggtcaact ccggcaccga cctgggtcagc attaccatcg gcggcaacga
481 cgcgggcttc gccgacacca tgaccacctg caacctccag ggcgagagcg cgtgcctggc
541 cgccttcgcc aagcgcgcgc cctacatcca gcagacgctg ccgcccagc tggaccagg
601 ctacgacgcc atcgacagcc gggcccccgc agcccagtc gtcgtcctgg gctaccggc
661 cttctacaag ctggggcgca gctgcgccgt cgtctctcg gagaagtccc gcgcccac
721 caacgcgcgc gccgacgaca tcaacgcctg caccgccaag cgcgcccgc accacggctt
781 cgccttcggg gacgtcaaca cgaccttcgc cgggcacgag ctgtgctccg gcgcccctg
841 gctgcacagc gtcaccttc ccgtggagaa ctctaccac cccacggcca acggacagtc
901 caagggttac ctgcccgtcc tgaactccgc cacctgatct cgcggctact ccgcccctga
961 cgaagtccc cccccggcg gggcttcgccc gtagggtcgc gtaccggcct cggcccgtcg
1021 gccgggtggc ccgcccgtacg tgcccgcgc cccgacgcg gtcgggtc

```



图 70 (SEQ ID No. 62)

```

1  ATGAAAAAAT GGTTTGTGTG TTTATTGGGA TTGGTCGCGC TGACAGTTCA
   TACTTTTTTA CCAAACACAC AAATAACCCT AACCAGCGCG ACTGTCAAGT

51  GGCAGCCGAC AGTCGCCCCG CCTTTTCCCG GATCGTGATG TTCGGCGACA
   CCGTCGGCTG TCAGCGGGGC GGAAAAGGGC CTAGCACTAC AAGCCGCTGT

101  GCCTCTCCGA TACCGGCAAA ATGTACAGCA AGATGCGCGG TTACCTCCCC
   CGGAGAGGCT ATGGCCGTTT TACATGTCGT TCTACGCGCC AATGGAGGGG

151  TCCAGCCCGC CCTACTATGA GGGCCGTTC TCCAACGGAC CCGTCTGGCT
   AGGTCGGGCG GGATGATACT CCCGGCAAAG AGGTTGCCTG GGCAGACCGA

201  GGAGCAGCTG ACCAAACAGT TCCCGGTCT GACCATCGCC AACGAAGCGG
   CCTCGTCGAC TGTTTTGTCA AGGGCCAGA CTGGTAGCGG TTGCTTCGCG

251  AAGCGGCTGC CACTGCCGTG GCTTACAACA AGATCTCCTG GAATCCCAAG
   TTCCGCCACG GTGACGGCAC CGAATGTTGT TCTAGAGGAC CTTAGGGTTC

301  TATCAGGTCA TCAACAACCT GACTACGAG GTCACCCAGT TCTTGAGAA
   ATAGTCCAGT AGTTGTTGGA CCTGATGCTC CAGTGGGTCA AGAACGTCTT

351  AGACAGCTTC AAGCCGGACG ATCTGGTGAT CCTCTGGGTC GGTGCCAATG
   TCTGTGGAAG TTCGGCCTGC TAGACCACTA GGAGACCAG CCACGGTTAC

401  ACTATCTGGC CTATGGCTGG AACACGGAGC AGGATGCCAA GCGGGTTCGC
   TGATAGACCG GATACCGACC TTGTGCCCTG TCCTACGGTT CGCCCAAGCG

451  GATGCCATCA GCGATGCGGC CAACCGCATG GTACTGAACG GTGCCAAGCA
   CTACGGTAGT CGTACGCCG GTTGGCGTAC CATGACTTGC CACGGTTCGT

501  GATACTGCTG TTCAACCTGC CGGATCTGGG CCAGAACCCG TCAGCTCGCA
   CTATGACGAC AAGTTGGACG GCCTAGACCC GGTCTTGGGC AGTCGAGCGT

551  GTCAGAAGGT GGTGAGGCGG GTCAGCCATG TCTCCGCCTA TCACAACCGA
   CAGTCTTCCA CCAGCTCCGC CAGTCGGTAC AGAGGCGGAT AGTGTGGTTC

601  CTGCTGCTGA ACCTGGCAGC CCAGCTGGCC CCCACCGGCA TGGTAAAGCT
   GACGACGACT TGGACCGTGC GGTGACCGG GGTGCGCGT ACCATTTGCA

651  GTTCGAGATC GACAAGCAAT TTGCCGAGAT GCTGCGTGAT CCGCAGAACT
   CAAGCTCTAG CTGTTGTTA AACGGCTCTA CGACGACTA GCGCTCTGA

701  TCGGCCTGAG CGACGTCGAG AACCCCTGCT ACGACGGCGG CTATGTGTGG
   AGCCGGACTC GCTGCAGCTC TTGGGGACGA TGCTGCCGCC GATACACACC

751  AAGCCGTTTG CCACCCGCAG CGTCAGCACC GACCGCCAGC TCTCCGCCTT
   TTCGGCAAAC GGTGGGCGTC GCAGTCGTGG CTGGCGGTG AGAGGCGGAA

801  CAGTCCGCAG GAACGCTCG CCATCGCCGG CAACCCGCTG CTGGCACAGG
   GTCAGGCGTC CTTGCGGAGC GGTAGCGGCC GTTGGGCGAC GACCGTGTCC

851  CCGTTGCCAG TCCTATGGCC CGCCGAGCG CCAGCCCCCT CAACTGTGAG
   GGCAACGGTC AGGATACCGG GCGGCGTCCG GGTGCGGGGA GTTGACACTC

901  GGCAAGATGT TCTGGGATCA GGTACACCCG ACCACTGTG TGCACGCAGC
   CCGTCTTACA AGACCCTAGT CCATGTGGGC TGGTGACAGC ACGTGCCTCG

951  CCTGAGCGAG CGCGCCGCCA CCTTCATCGC GAACCACTG GAGTTCCTCG
   GGACTCGCTC GCGCGCGGT GGAAGTAGCG CTTGGTCATG CTAAGGAGC

1001  CCCAC TGA
      GGGTG ACT

```

图 71 (SEQ ID No. 63)

```

1  ATGAAAAAAT GGTGTGTTG TTTATTGGGG TTGATCGCGC TGACAGTTCA
   TACTTTTTTA CCAAACAAAC AAATAACCCC AACTAGCGCG ACTGTCAAGT
51  GGCAGCCGAC ACTCGCCCGG CCTTCTCCCG GATCGTGATG TTCGGCGACA
   CCGTCGGCTG TGAGCGGGGC GGAAGAGGGC CTAGCACTAC AAGCCGCTGT
101 GCCTCTCCGA TACCGGCAAA ATGTACAGCA AGATGCGCGG TTACCTCCCC
   CGGAGAGGCT ATGGCCGTTT TACATGTCGT TCTACGCGCC AATGGAGGGG
151 TCCAGCCCGC CCTACTATGA GGGCCGTTTC TCCAACGGAC CCGTCTGGCT
   AGGTCGGGCG GGATGATACT CCCGGCAAAG AGGTTGCCTG GGCAGACCGA
201 GGAGCAGCTG ACCAAGCAGT TCCCGGGTCT GACCATCGCC AACGAAGCGG
   CCTCGTCGAC TGGTTCGTCA AGGGCCAGG CTGGTAGCGG TTGCTTCGCC
251 AAGCGCGTGC CACTGCCGTG GCTTACAACA AGATCTCCTG GAATCCCAAG
   TTCCGCCACG GTGACGGCAC CGAATGTTGT TCTAGAGGAC CTAGGGTTC
301 TATCAGGTCA TCAACAACCT GGACTACGAG GTCACCCAGT TCTTGCAGAA
   ATAGTCCAGT AGTTGTTGGA CCTGATGCTC CAGTGGGTCA AGAACGTCTT
351 AGACAGCTTC AAGCCGGACG ATCTGGTGAT CCTCTGGGTC GGTGCCAATG
   TCTGTCGAAG TTCGGCCTGC TAGACCACTA GGAGACCCAG CCACGGTTAC
401 ACTATCTGGC ATATGGCTGG AATACGGAGC AGGATGCCAA GCGAGTTCGC
   TGATAGACCG TATACCGACC TTATGCCTCG TCCTACGGTT CGCTCAAGCG
451 GATGCCATCA GCGATGCGGC CAACCGCATG GTECTGAACG GTGCCAAGCA
   CTACGGTAGT CGGTACGCCG GTTGGCGTAC CATGACTTGC CACGGTTCGT
501 GATACTGCTG TTCAACCTGC CGGATCTGGG CCAGAACCCG TCAGCCCGCA
   CTATGACGAC AAGTTGGACG GCCTAGACCC GGTCTTGGGC AGTCGGGCGT
551 GTCAGAAGGT GGTGAGGGCG GTCAGCCATG TCTCCGCTA TCACAACAAG
   CAGTCTTCCA CCAGCTCCGC CAGTCGGTAC AGAGGCGGAT AGTGTGTTTC
601 CTGTGCTGA ACCTGGCAGC CCAGCTGGCC CCCACCGGCA TGGTAAAGCT
   GACGACGACT TGGACCGTGC GGTGACCCGG GGTGGCCGT ACCATTTCGA
651 GTTCGAGATC GACAAGCAAT TTGCCGAGAT GCTGCGTGAT CCGCAGAACT
   CAAGCTCTAG CTGTTCGTTA AACGGCTCTA CGACGCACTA GCGCTCTTGA
701 TCGGCCTGAG CGACGTCGAG AACCCCTGCT ACGACGGCGG CTATGTGTGG
   AGCCGGACTC GCTGCAGCTC TTGGGGACGA TGCTGCCGCC GATACACACC
751 AAGCCGTTTG CCACCCGACG CGTCAGCACC GACCGCCAGC TCTCCGCCTT
   TTCGGCAAAC GGTGGGCGTC GCAGTCGTGG CTGGCGGTCG AGAGGCGGAA
801 CAGTCCGCAG GAACGCCTCG CCATCGCCGG CAACCCGCTG CTGGCACAGG
   GTCAGGCGTC CTTGCGGAGC GGTAGCGGCC GTTGGGCGAC GACCGTGTCC
851 CCGTTGCCAG TCCTATGGCC CGCCGACGCG CCAGCCCCCT CAACTGTGAG
   GGCAACGGTC AGGATACCGG GCGGCGTCCG GGTGCGGGGA GTTGACACTC
901 GGCAAGATGT TCTGGGATCA GGTACACCCG ACCACTGTGCG TGACCGCAGC
   CCGTTCTACA AGACCCTAGT CCATGTGGGC TGGTGACAGC ACGTGCCTCG
951 CCTGAGCGAG CGCGCCGCCA CCTTCATCGA GACCCAGTAC GAGTTCCTCG
   GGACTCGCTC GCGCGGCGGT GGAAGTAGCT CTGGGTCATG CTAAGGAGC
1001 CCCACGGATG A
     GGGTGCCTAC T

```

图 72 (SEQ ID No. 24)

```

1  ATGTTTAAGT TTAAAAAGAA TTTCTTAGTT GGATTATCGG CAGCTTTAAT
   TACAAATTC AATTTTCTT AAAGAATCAA CCTAATAGCC GTCGAAATTA

51  GAGTATTAGC TTGTTTTTCG CAACCGCCTC TGCAGCTAGC GCCGACAGCC
   CTCATAATCG AACAAAAGCC GTTGGCGGAG ACGTCGATCG CGGCTGTCGG

101  GTCCCGCCTT TTCCCGGATC GTGATGTTTC GCGACAGCCT CTCCGATACC
   CAGGGCCGAA AAGGGCCTAG CACTACAAGC CGCTGTCGGA GAGGCTATGG

151  GGCAAAATGT ACAGCAAGAT GCGCGGTTAC CTCCCCTCCA GCCCGCCCTA
   CCGTTTTACA TGTGTTCTA CCGGCCAATG GAGGGGAGGT CGGGCGGGAT

201  CTATGAGGGC CGTTTCTCCA ACGGACCCGT CTGGCTGGAG CAGCTGACCA
   GATACTCCCG GCAAAGAGGT TGCTGGGCA GACCGACCTC GTCGACTGGT

251  AACAGTTCCC GGGTCTGACC ATCGCCAACG AAGCGGAAG CGGTGCCACT
   TTGTCAAGGG CCCAGACTGG TAGCGGTTGC TTCGCCTTCC GCCACGGTGA

301  GCCGTGGCTT ACAACAAGAT CTCCTGGAAT CCAAGTATC AGGTCATCAA
   CGGCACCGAA TGTGTTCTA GAGGACCTA GGTTCATAG TCCAGTAGTT

351  CAACCTGGAC TACGAGGTCA CCCAGTTCTT GCAGAAAGAC AGCTTCAAGC
   GTTGACCTG ATGCTCCAGT GGTCAAGAA CGTCTTCTG TCGAAGTTTCG

401  CGGACGATCT GGTGATCCTC TGGTTCGGTG CCAATGACTA TCTGGCCTAT
   GCCTGCTAGA CCACTAGGAG ACCCAGCCAC GGTACTGAT AGACCGGATA

451  GGCTGGAACA CGGAGCAGGA TGCCAAGCGG GTTCGCGATG CCATCAGCGA
   CCGACCTTGT GCCTCGTCTT ACGGTTCCGC CAAGCGCTAC GGTAGTCGCT

501  TGCGGCCAAC CGCATGGTAC TGAACGGTGC CAAGCAGATA CTGCTGTTC A
   ACGCCGGTTG GCTTACCATG ACTTGCCACG GTTCGCTAT GACGACAAGT

551  ACCTGCCGGA TCTGGGCCAG AACCCGTCAG CTCGCAGTCA GAAGGTGGTC
   TGGACGGCCT AGACCCGGTC TTGGGCAGTC GAGCGTCAGT CTCCACCAG

601  GAGGCGGTCA GCCATGTCTC CGCCTATCAC AACCAGCTGC TGCTGAACCT
   CTCCGCCAGT CCGTACAGAG GCGGATAGTG TTGGTCGACG ACGACTTGGA

651  GGCACGCCAG CTGGCCCCA CCGCATGGT AAAGCTGTTC GAGATCGACA
   CCGTGCCTG GACCGGGGT GGCCGTACCA TTTCGACAAG CTCTAGCTGT

701  AGCAATTTGC CGAGATGCTG CGTGATCCGC AGAACTTCGG CCTGAGCGAC
   TCGTTAAACG GCTCTACGAC GCACTAGGCG TCTGAAGCC GGACTCGCTG

751  GTCGAGAACC CCTGCTACGA CGGCGGCTAT GTGTGGAAGC CGTTTGCCAC
   CAGCTCTTGG GGACGATGCT GCCGCCGATA CACACCTTCG GCAAACGGTG

801  CCGCAGCGTC AGCACCAGC GCCAGCTCTC CGCCTTCAGT CCGCAGGAAC
   GCGCTCGCAG TCGTGGCTGG CGGTCGAGAG GCGGAAGTCA GGCGTCTTG

851  GCCTCGCCAT CGCCGGCAAC CCGCTGCTGG CACAGGCCGT TGCCAGTCTT
   CGGAGCGGTA GCGGCCGTTG GCGCAGCACC GTGTCCGCA ACGGTACGGA

901  ATGGCCCGCC GCAGCGCCAG CCCCTCAAC TGTGAGGGCA AGATGTTCTG
   TACCGGGCGG CGTCGCGGTC GGGGGAGTTG AACTCCCGT TCTACAAGAC

951  GGATCAGGTA CACCCGACCA CTGTCGTGCA CGCAGCCCTG AGCGAGCGCG
   CCTAGTCCAT GTGGGCTGGT GACAGCACGT GCGTCGGGAC TCGCTCGCCG

1001  CCGCCACCTT CATCGCAAC CAGTACGAGT TCCTCGCCCA CTGATGA
   GCGGGTGGAA GTAGCGCTTG GTCATGCTCA AGGAGCGGGT GACTACT

```