

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
C12N 9/28 (2006.01)



[12] 发明专利说明书

专利号 ZL 99815544.6

[45] 授权公告日 2008 年 3 月 12 日

[11] 授权公告号 CN 100374557C

[22] 申请日 1999.11.16 [21] 申请号 99815544.6

[30] 优先权

[32] 1998.11.16 [33] DK [31] PA199801495

[86] 国际申请 PCT/DK1999/000628 1999.11.16

[87] 国际公布 WO2000/029560 英 2000.5.25

[85] 进入国家阶段日期 2001.7.10

[73] 专利权人 诺维信公司

地址 丹麦鲍斯韦

[72] 发明人 阿伦·斯文德森 索伦·谢鲁尔夫

亨里克·比斯加德弗兰岑

卡斯滕·安德森

[56] 参考文献

WO9510603A1 1995.4.20

WO9623873A1 1996.8.8

WO9743424 1997.11.20

WO9526397A1 1995.10.5

审查员 徐 莉

[74] 专利代理机构 北京市柳沈律师事务所

代理人 巫肖南

权利要求书 2 页 说明书 58 页 附图 4 页

[54] 发明名称

α - 淀粉酶变体

[57] 摘要

本发明涉及亲代 Termamyl - 样 α - 淀粉酶的变体, 该变体在 2, 3, 4, 5, 或 6 个区域/位置中含有突变。所述变体在高温下(相对于其亲代)具有增加的稳定性。本发明还涉及含有编码本发明 α - 淀粉酶变体之 DNA 序列的 DNA 构建体, 携带本发明 DNA 构建体的重组表达载体, 用本发明的 DNA 构建体转化的细胞, 本发明的 α - 淀粉酶变体在洗涤和/或洗碟、织物退浆、淀粉液化方面的用途, 含有本发明的 α - 淀粉酶变体的去污剂添加剂, 含有本发明的 α - 淀粉酶变体的手洗或机洗洗碟剂组合物, 产生亲代 Termamyl - 样 α - 淀粉酶的变体的方法, 所述变体显示增加。

1. SEQ ID NO:4 所示地衣芽孢杆菌 α -淀粉酶和 SEQ ID NO:5 所示解淀粉芽孢杆菌 α -淀粉酶的亲代杂合体 α -淀粉酶的变体, 所述杂合体 α -淀粉酶与 SEQ ID NO:4 所示的地衣芽孢杆菌 α -淀粉酶等同, 不同之处在于其成熟蛋白质 N-末端的 35 个氨基酸残基被 SEQ ID NO:5 所示的解淀粉芽孢杆菌 α -淀粉酶成熟蛋白质 N-末端的 33 个氨基酸残基所取代, 并且进一步具有在下列依据 SEQ ID NO:4 中使用的编号的位置上的突变: H156Y+A181T+N190F+A209V+Q264S。

2. 权利要求 1 的变体, 其中亲代 α -淀粉酶依据在 SEQ ID NO:4 中使用的编号还在下列一个或多个位置具有突变: K176, I201 和 H205。

3. 权利要求 2 的变体, 其中亲代 α -淀粉酶依据在 SEQ ID NO:4 中使用的编号含有一个或多个下列取代: K176R, I201F 或 H205N。

4. 权利要求 3 的变体, 其中所述亲代 α -淀粉酶依据在 SEQ ID NO:4 中使用的编号具有下述取代 K176R+ I201F+H205N。

5. 权利要求 1-4 中任一项的变体, 其中所述变体依据在 SEQ ID NO:4 中使用的编号具有一个或多个下列取代: E376, S417, A420, S356 或 Y358。

6. 权利要求 5 的变体, 所述变体具有一个或多个下列取代: S356A 或 Y358F。

7. 权利要求 5 的变体, 所述变体具有一个或多个下列取代: E376K 或 A420R。

8. 权利要求 5 的变体, 所述变体具有一个或多个下列取代: S417T 或 A420Q。

9. DNA 构建体, 其含有编码权利要求 1 至 8 中任一项的 α -淀粉酶变体的 DNA 序列。

10. 重组表达载体, 其携有权利要求 9 的 DNA 构建体。

11. 用权利要求 9 的 DNA 构建体或权利要求 10 的载体转化的细胞。

12. 权利要求 11 的细胞, 其为微生物。

13. 权利要求 12 的细胞, 其是革兰氏阳性细菌。

14. 权利要求 13 的细胞, 其选自枯草芽孢杆菌, 地衣芽孢杆菌, 迟缓芽孢杆菌, 短芽孢杆菌, 嗜热脂肪芽孢杆菌, 嗜碱芽孢杆菌, 解淀粉芽孢杆菌,

凝结芽孢杆菌，环状芽孢杆菌，灿烂芽孢杆菌或苏云金芽孢杆菌。

15.洗涤剂添加剂，其含有权利要求1至8中任一项的 α -淀粉酶变体。

16. 权利要求15的洗涤剂添加剂，所述洗涤剂添加剂是无粉尘的颗粒，稳定化的液体或被保护的酶的形式。

17.权利要求15的洗涤剂添加剂，其含有0.02-200 mg 酶蛋白质/g 添加剂。

18.权利要求15-17中任一项的洗涤剂添加剂，其中还含有选自下述的另外的酶：蛋白酶，脂肪酶，过氧化物酶，另一种淀粉分解酶，和/或纤维素酶。

19.洗涤剂组合物，其含有权利要求1至8中任一项的 α -淀粉酶变体。

20.权利要求19的洗涤剂组合物，其中还含有选自下述的另外的酶：蛋白酶，脂肪酶，过氧化物酶，另一种淀粉分解酶和/或纤维素酶。

21. 权利要求19的洗涤剂组合物，其中所述洗涤剂组合物为手用或机用洗碟洗涤剂组合物。

22.权利要求21的洗涤剂组合物，其中还含有选自下述的另外的酶：蛋白酶，脂肪酶，过氧化物酶，另一种淀粉分解酶，和/或纤维素酶。

23. 权利要求19的洗涤剂组合物，其中所述洗涤剂组合物为手用或机用洗衣组合物。

24.权利要求23的洗涤剂组合物，其中还含有选自下述的另外的酶：蛋白酶，脂肪酶，过氧化物酶，淀粉分解酶和/或纤维素酶。

25.权利要求1至8中任一项的 α -淀粉酶变体用于洗涤或用于织物脱浆或用于淀粉液化的用途。

26. 权利要求25的用途，其中的洗涤为洗碟。

α -淀粉酶变体

发明领域

本发明涉及亲代 Termamyl-样 α -淀粉酶的新变体，相对于亲代 α -淀粉酶而言，所述变体具有经改变的特性。所述特性包括例如在酸性 pH，低钙浓度和/或高温下增加的稳定性。所述变体适于多种用途，特别是工业淀粉处理(如淀粉液化或糖化)。

发明背景

α -淀粉酶(α -1, 4-葡聚糖-4-葡聚糖水解酶, EC 3.2.1.1)构成一组能催化淀粉和其它线性和分支 1, 4-葡糖苷寡-和多糖水解的酶。

大量专利和科学文献中涉及这一类在工业上非常重要的酶。可从例如 WO 90/11352, WO 95/10603, WO 95/26397, WO 96/23873 和 WO 96/23874 中了解多种 α -淀粉酶(如 Termamyl-样 α -淀粉酶)的变体。

WO 96/23874 提供了 Termamyl-样 α -淀粉酶的三维，X-射线晶体结构数据，该 α -淀粉酶由解淀粉芽孢杆菌 α -淀粉酶的 300 个 N-末端氨基酸残基，和含有氨基酸序列的地衣芽孢杆菌 α -淀粉酶 C-末端的氨基酸 301-483 组成(后者有商品 Termamyl™(商标名))，因此，与工业上重要的芽孢杆菌 α -淀粉酶(在本发明的上下文中，包含在术语“Termamyl-样 α -淀粉酶”的定义中，它特别地包括地衣芽孢杆菌，解淀粉芽孢杆菌和嗜热脂肪芽孢杆菌的 α -淀粉酶)密切相关。WO 96/23874 进一步描述了在分析亲代 Termamyl-样 α -淀粉酶结构的基础上设计亲代 Termamyl-样 α -淀粉酶变体的方法学，相对于亲代而言，所述变体表现出有所改变的特性。

发明简述

本发明涉及 Termamyl-样 α -淀粉酶的新的 α -淀粉分解变体(突变体)，尤其是在高温和酸性 pH 下(相对于亲代而言)表现出增加的稳定性的变体，如美国专利 3,912,590 和欧洲专利公开号 252,730 和 63,909 所述，所述变体有利于对淀粉进行工业处理(淀粉液化，糖化等)。

淀粉转化

“传统”的淀粉转化方法将淀粉降解为较低分子量的碳水化合物成分，

如糖或脂肪替代物，该方法包括脱支步骤。

“淀粉至糖”的转化

当将淀粉转化为糖时，淀粉被解聚。所述解聚过程由预处理步骤和两个或三个连续处理步骤(即液化处理，糖化处理，根据所需终产物，还任选包括异构化处理)组成。

预-处理天然淀粉

天然淀粉由室温下不溶于水的微粒组成。加热水淀粉浆液时，微粒膨胀，最后胀破，将淀粉分子分散至溶液中。在“凝胶化”处理的过程中，粘度显著增加。由于一般工业过程中的固体水平为30-40%，因此不得不稀释或“液化”淀粉以使其便于处理。目前，主要通过酶促降解使粘度降低。

液化

在液化步骤中，通过 α -淀粉酶(如本文的 Termamyl™ SEQ ID NO:4)将长链淀粉分解为分支的和线性的较短单位(麦芽糖糊精)。液化过程在105-110℃进行5至10分钟，接着在95℃进行1-2小时。pH为5.5至6.2之间。为了确保这些条件下最佳的酶稳定性，需加入1mM钙(40ppm游离的钙离子)。如此处理之后，液化淀粉的“葡萄糖当量”(DE)为10-15。

糖化

液化过程之后，通过加入葡糖淀粉酶(如 AMG™)和脱支酶，如异淀粉酶(美国专利 4,335,208)或支链淀粉酶(如 Promozyme™)(美国专利 4,560,651)，将麦芽糖糊精转变为葡萄糖。在此步骤之前，将pH值降低为4.5以下，维持高温(95℃以上)以灭活液化 α -淀粉酶，减少不能被脱支酶正确水解的短寡糖(被称为“潘糖前体”)的形成。

将温度降低为60℃，加入葡糖淀粉酶和脱支酶。将糖化过程进行24-72小时。

通常，当在液化步骤之后变性 α -淀粉酶时，约0.2-0.5%的糖化产物是分支的三糖 6^2 - α -葡糖基麦芽糖(潘糖)，该糖不能被支链淀粉酶降解。如果糖化过程中存在得自液化步骤的活性淀粉酶(即未变性)，该水平可以高至1-2%，这非常不合乎需要，因为它显著降低了糖化产量。

异构化

当所需的最终糖产物是例如高果糖浆时，葡萄糖浆可被转变为果糖。糖化过程之后，将pH值增加为6-8，优选为pH7.5，通过离子交换除去钙。

然后使用例如固定化的葡萄糖异构酶(如 Sweetzyme™)将葡萄糖浆转变为高果糖浆。

在本发明的上下文中，术语“酸性 pH”指的是 pH 低于 7.0，尤其是低于上述传统工业淀粉液化过程中所用的 pH 范围，即为 pH5.5-6.2。

在本发明的上下文中，术语“低钙浓度”指的是浓度低于传统的工业淀粉液化过程中所用的正常水平，如浓度为 0-40 ppm，优选为 10-30 ppm，如 15-25 ppm 钙。正常浓度根据谷物中游离 Ca^{2+} 的浓度而变化。通常加入相当于 1mM(40ppm)的剂量，该剂量与谷物中的钙水平一起给出 40-60ppm 游离的 Ca^{2+} 。

在本发明的上下文中，术语“高温”指的是温度为 95 至 160 °C，尤其是正常进行工业淀粉液化处理所用的温度范围，所述范围是 95 至 105 °C。

本发明进一步涉及编码本发明变体的 DNA 构建体，制备本发明变体的方法，和本发明变体单独或与其它 α -淀粉分解酶联合用于多种工业过程，尤其是淀粉液化的用途。

命名法

在本说明书和权利要求书中，使用了常规的氨基酸残基一字母和三字母密码。

为了易于参照，利用下述命名法描述本发明的 α -淀粉酶变体：原来的氨基酸：位置：取代的氨基酸。

根据该命名法，用天冬酰胺取代第 30 位的丙氨酸被表示为：Ala30Asn 或 A30N，在相同位置处缺失丙氨酸被表示为：Ala30*或 A30*，而插入另一个氨基酸残基，如赖氨酸被表示为：Ala30AlaLys 或 A30AK。缺失连续的一段氨基酸残基，如氨基酸残基 30-33 被表示为(30-33)*或 Δ (A30-N33)。

当特定的 α -淀粉酶相对于其它 α -淀粉酶而言含有“缺失”，并在该位置插入某个氨基酸，例如在第 36 位插入天冬氨酸时，被表示为*36Asp 或*36D。

多个突变由加号隔开，即：Ala30Asp+Glu34Ser 或 A30N+E34S 分别表示在第 30 和 34 位，丙氨酸和谷氨酸分别用天冬酰胺和丝氨酸取代。多个突变也可以由下列与加号意义相同的符号隔开：Ala30Asp /Glu34Ser 或 A30N/E34S。

当在给定的位置插入一个或多个其它氨基酸残基时，被表示为 A30N,E

或 A30N 或 A30E。

另外，当本文鉴定出适于修饰的位置，而没有暗示任何具体的修饰时，应理解为可用任何氨基酸残基取代该位置存在的氨基酸残基。因此，例如，当提到修饰第 30 位的丙氨酸，但未具体说明时，应理解为丙氨酸可被缺失，或用任何其它氨基酸取代，即下列任何一个氨基酸所取代：R, N, D, A, C, Q, E, G, H, I, L, K, M, F, P, S, T, W, Y, V。

附图简述

图 1 是本发明上下文中 6 个亲代 Termamyl-样 α -淀粉酶的氨基酸序列对比。最左边的数字表示以下各个氨基酸序列：

1 : SEQ ID NO:2 ,

2 : 淀粉酶

3 : SEQ ID NO:1 ,

4 : SEQ ID NO:5 ,

5 : SEQ ID NO:4 ,

6 : SEQ ID NO:3 。

图 2 表示实施例 1 中所用的 PCR 策略。

发明详述

Termamyl-样 α -淀粉酶

众所周知，通过芽孢杆菌种产生的多种 α -淀粉酶在氨基酸水平上高度同源。例如，已发现含有 SEQ ID NO:4 所示氨基酸序列的地衣芽孢杆菌 α -淀粉酶(商品为 Termamyl™)与含有 SEQ ID NO:5 所示氨基酸序列的解淀粉芽孢杆菌 α -淀粉酶约 89%同源，与含有 SEQ ID NO:3 所示氨基酸序列的嗜热脂肪芽孢杆菌 α -淀粉酶约 79%同源。其它同源的 α -淀粉酶包括得自芽孢杆菌菌种 NCIB 12289，NCIB 12512，NCIB 12513 或 DSM 9375 的 α -淀粉酶(皆详细描述于 WO 95/26397)，和描述于 Tsukamoto 等，生物化学和生物物理研究通讯, 151(1988), p25-31 中的 α -淀粉酶。

其它同源的 α -淀粉酶包括由 EP 0252666 所述的地衣芽孢杆菌菌株(ATCC 27811)所产生的 α -淀粉酶，和 WO 91/00353 和 WO 94/18314 中鉴定的 α -淀粉酶。其它可商购的 Termamyl-样地衣芽孢杆菌 α -淀粉酶是 Optitherm™和 Takatherm™(可得自 Solvay)，Maxamyl™(可得自 Gist-brocades/Genencor)，Spezym AA™和 Spezyme Delta AA™(可得自 Genencor)

和 Keistase™(可得自 Daiwa)。

由于这些 α -淀粉酶基本上同源，因此可以认为它们属于同一类 α -淀粉酶，即“Termamyl-样 α -淀粉酶”。

因此，在本发明的上下文中，术语“Termamyl-样 α -淀粉酶”欲指在氨基酸水平上与 Termamyl™(即具有本文 SEQ ID NO:4 所示氨基酸序列的地衣芽孢杆菌 α -淀粉酶)基本上同源的 α -淀粉酶。换句话说，Termamyl-样 α -淀粉酶是具有本文 SEQ ID NO:1, 2, 3, 4, 5, 6, 7 或 8 所示氨基酸序列，和 WO 95/26397 的 SEQ ID NO:1(与本文 SEQ ID NO:7 所示氨基酸序列相同)或 WO 95/26397 的 SEQ ID NO:2(与本文 SEQ ID NO:8 所示氨基酸序列相同)或 Tsukamoto 等, 1988(其氨基酸序列示于本文 SEQ ID NO:6)所示氨基酸序列的 α -淀粉酶，或者，Termamyl-样 α -淀粉酶是：i)与 SEQ ID NO:1 或 2 或 3 或 4 或 5 或 6 或 7 或 8 所示的至少一个所述氨基酸序列的同源性(同一性)至少为 60%，优选至少为 70%，更优选至少为 75%，甚至更优选至少为 80%，尤其是至少为 85%，尤其优选至少为 90%，尤其是至少为 95%，甚至尤其更优选至少为 97%，尤其是至少为 99%的 α -淀粉酶，和/或 ii)能与针对一种或多种所述 α -淀粉酶产生的抗体发生免疫交叉反应的 α -淀粉酶，和/或 iii)由能在低至很高严紧条件下(所述条件在下文中描述)与编码上述-特定 α -淀粉酶的 DNA 序列杂交的 DNA 序列编码的 α -淀粉酶，所述特定 α -淀粉酶的 DNA 序列分别示于本申请的 SEQ ID NO:9, 10, 11, 12 和 32(分别编码本文 SEQ ID NO:1, 2, 3, 4 和 5 所示的氨基酸序列)，WO 95/26397 的 SEQ ID NO:4(该 DNA 序列与终止密码子 TAA 一起示于本文 SEQ ID NO:13，它编码本文 SEQ ID NO:8 所示的氨基酸序列)和 WO 95/26397 的 SEQ ID NO:5(示于本文 SEQ ID NO:14)。

关于特性 i)，可利用任何常规的算法测定“同源性”(同一性)，优选使用 GCG 程序包第 8 版(1994 年 8 月)的缺口程序，该程序使用缺口得分的缺省值，即缺口产生得分为 3.0，缺口延伸得分为 0.1 (Genetic Computer Group(1991) GCG 程序包程序手册，第 8 版，575 Science Drive, Madison, Wisconsin, USA 53711)。

在一个实施方案中，亲代 Termamyl-样 α -淀粉酶骨架具有的氨基酸序列与 SEQ ID NO:4 的同一性程度(按上述方法测定)至少为 65%，优选至少为 70%，优选至少为 75%，更优选至少为 80%，更优选至少为 85%，甚至更优

选至少约为 90%，甚至更优选至少为 95%，甚至更优选至少为 97%，甚至更优选至少为 99%。

可使用 Termamyl®(SEQ ID NO:4)和 Termamyl-样 α -淀粉酶之间的结构对比鉴定其它 Termamyl-样 α -淀粉酶中等同的/相应的位置。得到所述结构对比的一个方法是使用 GCG 程序包中的 Pile Up 程序，该程序使用缺省的缺口得分值，即缺口产生得分为 3.0，缺口延伸得分为 0.1。其它结构对比方法包括疏水簇分析(Gaboriaud 等, (1987), FEBS LETTERS 224, p.149-155)和反向穿梭(threading)法(Huber, T; Torda, AE, 蛋白质科学, 卷 7, 1 期, 142-149(1998))。

例如，在已提及的多种 Termamyl-样 α -淀粉酶的氨基酸序列中，地衣芽孢杆菌 α -淀粉酶 C-结构域中靶残基的相应位置如下：

Termamyl-样 α -淀粉酶

地衣芽孢杆菌(SEQ ID NO:4)	S356 Y358 E376 S417 A420
解淀粉芽孢杆菌(SEQ ID NO:5)	S356 Y358 E376 S417 A420
嗜热脂肪芽孢杆菌(SEQ ID NO:3) Y361
Bac.WO95/26397(SEQ ID NO:2) Y363 S419
Bac.WO95/26397(SEQ ID NO:1) Y363

这些保守氨基酸残基的突变对于增加高温下酸性 pH 和/或低钙浓度时的稳定性非常重要，这一点将在下文中进一步描述。

使用针对相关 Termamyl-样 α -淀粉酶的至少一个表位的抗体或与所述表位反应的抗体，可以检测 α -淀粉酶的特性 ii)(见上文)，即免疫交叉反应性。通过本领域已知的方法，如 Hudson 等，实用免疫学，第 3 版(1989)，Blackwell Scientific Publications 可以产生抗体，所述抗体可以是单克隆或多克隆抗体。使用本领域已知的试验，例如 Western 印迹或放射状免疫扩散试验(如 Hudson 等, 1989 所述)可以测定免疫交叉-反应性。在此方面，已发现了分别具有氨基酸序列 SEQ ID NO:1, 2, 3, 4, 5, 6, 7 或 8 的 α -淀粉酶之间的免疫交叉-反应性。

可以在所述 α -淀粉酶的完整或部分核苷酸或氨基酸序列的基础上适当制备寡核苷酸探针，然后使用所述探针，根据上述特性 iii)鉴定 Termamyl-样 α -淀粉酶。

检测杂交的适当条件包括：在 5×SSC 中预浸泡，于~40 °C，在 20%甲酰

胺，5×Denhardt's 溶液，50mM 磷酸钠，pH6.8 和 50mg 变性的经超声处理的小牛胸腺 DNA 中预杂交 1 小时，接着于~40 °C，在添加有 100mM ATP 的相同溶液中杂交 18 小时，接着于 40 °C，用 2×SSC，0.2% SDS 将滤膜洗涤 3 次，每次 30 分钟(低严谨性)，优选在 50 °C(中度严谨性)，更优选在 65 °C(高严谨性)，甚至更优选在~75 °C(很高的严谨性)进行洗涤。有关杂交方法的更多细节描述于 Sambrook 等，分子克隆：实验室手册，第 2 版，冷泉港，1989。

在本发明的上下文中，“得自”不仅仅表示由所述生物体产生或可由所述生物体产生的 α -淀粉酶，还指由分离自所述菌株的 DNA 序列编码的 α -淀粉酶，和由所述 DNA 序列转化的宿主生物体产生的 α -淀粉酶。最后，该术语还表示由合成的和/或 cDNA 来源的 DNA 序列编码的，并具有所述 α -淀粉酶的鉴定特征的 α -淀粉酶。该术语还表示亲代 α -淀粉酶可以是天然 α -淀粉酶的变体，即修饰(插入，取代，缺失)天然 α -淀粉酶的一个或多个氨基酸残基而得到的变体。

亲代杂合 α -淀粉酶

亲代 α -淀粉酶(骨架)可以是杂合 α -淀粉酶，即含有得自至少两种 α -淀粉酶的部分氨基酸序列组合的 α -淀粉酶。

亲代杂合 α -淀粉酶可以是在氨基酸同源性和/或免疫交叉-反应性和/或 DNA 杂交(如上所述)的基础上被确定为属于 Termamyl-样 α -淀粉酶家族的杂合 α -淀粉酶。此时，杂合 α -淀粉酶一般由 Termamyl-样 α -淀粉酶的至少一个部分和一种或多种其它 α -淀粉酶的部分组成，所述其它 α -淀粉酶选自微生物(细菌或真菌)和/或哺乳动物来源的 Termamyl-样 α -淀粉酶或非-Termamyl-样 α -淀粉酶。

因此，亲代杂合 α -淀粉酶可含有得自至少两种 Termamyl-样 α -淀粉酶，或得自至少一种 Termamyl-样和至少一种非-Termamyl-样细菌 α -淀粉酶，或得自至少一种 Termamyl-样和至少一种真菌 α -淀粉酶的部分氨基酸序列组合。衍生得到部分氨基酸序列的 Termamyl-样 α -淀粉酶可以是例如本文所述的任何特定的 Termamyl-样 α -淀粉酶。

例如，亲代 α -淀粉酶可含有得自地衣芽孢杆菌菌株的 α -淀粉酶的 C-末端部分，和得自解淀粉芽孢杆菌菌株或嗜热脂肪芽孢杆菌菌株的 α -淀粉酶的 N-末端部分。例如，亲代 α -淀粉酶可含有地衣芽孢杆菌 α -淀粉酶 C-末端部分的至少 430 个氨基酸残基。所述杂合 Termamyl-样 α -淀粉酶可以与 SEQ ID NO:4

所示的地衣芽孢杆菌 α -淀粉酶相同，不同之处仅在于：(成熟蛋白质)N-末端的35个氨基酸残基被SEQ ID NO:5所示的解淀粉芽孢杆菌 α -淀粉酶(BAN)成熟蛋白质N-末端的33个氨基酸残基所取代。该杂合体也可由对应于嗜热脂肪芽孢杆菌 α -淀粉酶(具有SEQ ID NO:3所示氨基酸序列)的68个N-末端氨基酸残基的氨基酸区段，和对应于地衣芽孢杆菌 α -淀粉酶(具有SEQ ID NO:4所示氨基酸序列)的415个C-末端氨基酸残基的氨基酸区段组成。

非-Termamyl-样 α -淀粉酶可以是例如真菌 α -淀粉酶，哺乳动物或植物 α -淀粉酶或细菌 α -淀粉酶(不同于Termamyl-样 α -淀粉酶)。这种 α -淀粉酶的具体例子包括米曲霉 TAKA α -淀粉酶，黑曲霉酸性 α -淀粉酶，枯草芽孢杆菌 α -淀粉酶，猪胰腺 α -淀粉酶和大麦 α -淀粉酶。所有这些 α -淀粉酶都具有已阐明的，与本文所述典型Termamyl-样 α -淀粉酶的结构显著不同的结构。

上述真菌 α -淀粉酶，即得自黑曲霉和米曲霉的 α -淀粉酶在氨基酸水平上高度同源，一般认为它们属于相同的 α -淀粉酶家族。得自米曲霉的真菌 α -淀粉酶可以商购，其商品名为Fungamyl™。

另外，当提及Termamyl-样 α -淀粉酶的特定变体(本发明的变体)时，提到该变体是通过常规方法，修饰(例如缺失或取代)特定Termamyl-样 α -淀粉酶氨基酸序列中的特定氨基酸残基而获得的，应理解在另一个Termamyl-样 α -淀粉酶的等同位置(由各个氨基酸序列之间最有可能的氨基酸序列对比测定)修饰得到的变体也包括在本发明的范围内。

本发明变体的优选实施方案是得自地衣芽孢杆菌 α -淀粉酶(作为亲代Termamyl-样 α -淀粉酶)的变体，例如，上述 α -淀粉酶中的一个，如具有SEQ ID NO:4所示氨基酸序列的地衣芽孢杆菌 α -淀粉酶的变体。

本发明变体经改变的特性

下文将讨论本发明变体中可能存在的改变/突变与由此产生的合乎需要的特性改变(相对于亲代Termamyl-样 α -淀粉酶而言)之间的关系。

在高温，酸性pH和/或低钙浓度下稳定性的增加

本发明涉及亲代Termamyl-样 α -淀粉酶的变体，与亲代 α -淀粉酶相比，所述变体 α -淀粉酶表面的一个或多个暴露于溶剂的氨基酸残基发生变化，从而使 α -淀粉酶的总体亲水性增加和/或表面上所述暴露于溶剂之氨基酸残基的侧链中的甲基总数增加。

在优选的实施方案中，向内弯曲的凹面上的一个或多个暴露于溶剂的氨

基酸残基变为更具疏水性的氨基酸残基。

在另一个优选的实施方案中，凸面上的一个或多个暴露于溶剂的氨基酸残基经改变后增加了侧链中的甲基数目。

本发明涉及亲代 Termamyl-样 α -淀粉酶的 α -淀粉酶变体，该变体在选自下列的一个或多个位置处具有变化：E376, S417, A420, S356, Y358；

其中：(a)所述变化各为：

(i)在占用上述位置的氨基酸的下游插入一个氨基酸，

(ii)缺失占用上述位置的氨基酸，或

(iii)用不同的氨基酸取代占用上述位置的氨基酸，

(b)所述变体具有 α -淀粉酶活性，和(c)每个位置对应于具有 SEQ ID NO:4 氨基酸序列的亲代 Termamyl-样 α -淀粉酶的氨基酸序列位置。

在一个实施方案中，变化是下列取代之—：

E376A, R, D, C, Q, G, H, I, K, L, M, N, F, P, S, T, W, Y, V；

在优选的实施方案中，取代是：E376K。

在一个实施方案中，变化是下列取代之—：

S417A, R, D, C, E, Q, G, H, I, K, L, M, N, F, P, T, W, Y, V；

在优选的实施方案中，取代是：S417T。

在一个实施方案中，变化是下列取代之—：

A420R, D, C, E, Q, G, H, I, K, L, M, N, F, P, S, T, W, Y, V；

在优选的实施方案中，取代是：A420Q, R。

在一个实施方案中，变化是下列取代之—：

S356A, R, D, C, E, Q, G, H, I, K, L, M, N, F, P, T, W, Y, V。

在一个实施方案中，变化是下列取代之—：

Y358A, R, D, C, E, Q, G, H, I, K, L, M, N, F, P, S, T, W, V。

在优选的实施方案中，取代是：Y358F。

在本发明的一个实施方案中，变体含有一个或多个下列取代：E376K, S417T, A420Q, R, S356A, Y358F。

使用下文阐明本发明的实施例 2 中所述的方法，测定在高温，酸性 pH 和/或低钙浓度下稳定性的增加。

用作制备本发明变体所用骨架的亲代 Termamyl-样 α -淀粉酶可以是上文所述的任何 Termamyl-样 α -淀粉酶。

特别希望它们是选自下列的亲代 Termamyl-样 α -淀粉酶：得自地衣芽孢杆菌，如地衣芽孢杆菌菌株 ATCC 27811，解淀粉芽孢杆菌，嗜热脂肪芽孢杆菌，芽孢杆菌菌种 NCIB 12289, NCIB12512, NCIB 12513 或 DSM 9375 的亲代 Termamyl-样 α -淀粉酶，和 SEQ ID NO:1, 2, 3, 4, 5, 6, 7 和 8 所示的亲代 Termamyl-样 α -淀粉酶。

在本发明的一个实施方案中，亲代 Termamyl-样 α -淀粉酶是与 SEQ ID NO:4 所示的地衣芽孢杆菌 α -淀粉酶(Termamyl)相同的杂合 α -淀粉酶，不同之处在于(成熟蛋白质)N-末端的 35 个氨基酸残基被 SEQ ID NO:5 所示解淀粉芽孢杆菌 α -淀粉酶(BAN)成熟蛋白质 N-末端的 33 个氨基酸残基所取代。亲代 Termamyl-样杂合 α -淀粉酶可以是上述的杂合 Termamyl-样 α -淀粉酶，其进一步具有下列突变：H156Y+181T+190F+209V+264S(使用 SEQ ID NO:4 中的编号)。所述骨架在下文中被称为“LE174”。

亲代 α -淀粉酶进一步在下列一个或多个位置具有突变较为有利：K176, I201 和 H205(使用 SEQ ID NO:4 中的编号)，尤其有利的是具有一个或多个下列取代：K176R, I201F 和 H205N(使用 SEQ ID NO:4 中的编号)，例如特别是下列取代：K176R+I201F+H205N(使用 SEQ ID NO:4 中的编号)。

本发明人发现：相对于亲代 Termamyl-样 α -淀粉酶而言，上述变体在 95-160 °C 的温度(即高温)，低于 7.0 的 pH(即酸性 pH)和/或低于 1mM(40ppm)的钙浓度(即低钙浓度)下具有增加的稳定性。

改变(如通过取代来改变)一个或多个暴露于溶剂的氨基酸残基可以 1)增加酶的总体亲水性，或 2)增加暴露于溶剂的氨基酸残基侧链中的甲基数目，从而改善温度稳定性。优选将向内弯曲的凹面上的氨基酸残基改变(如通过取代来改变)为更具疏水性的残基。优选改变(如通过取代来改变)凸面上的氨基酸残基以增加侧链中的甲基数目。

使用国际互联网站点 <http://sunrise.cbs.umn.edu/cast/> 版本 1.0 中的 CAST 程序(1998 年 2 月发布)(参考文献: Jie Liang, Herbert Edelsbrunner 和 Clare Woodward, 1998, 分析蛋白质袋和沟槽: 测定结合位点几何学并涉及配体设计, 蛋白质科学, 7, pp.1884-1897), 可以鉴定出靠近表面的凹面区域。由于直径为 1.4 埃的探针可进可出, 因此在该程序中确定了通向表面的路径。在 CAST 程序中使用缺省参数, 使用 Brookhaven 数据库(1BPL)中的地衣芽孢杆菌钙-耗尽的 α -淀粉酶结构可以发现凹面沟槽:

三种类型的相互作用可被合理化：

- A. 残基侧链和蛋白质之间的相互作用，
- B. 残基侧链和周围环境中的水之间的相互作用，
- C. 水和蛋白质之间的相互作用。

将 SEQ ID NO:4 所示的亲代 Termamyl-样 α -淀粉酶用作骨架，认为下列位置暴露于溶剂，并可对其进行适当改变：

E376, S417, A420, S356, Y358。

使用 W. kabsch 和 C. Sander, 生物聚合物 22(1983) pp.2577-2637 的 dssp 程序，可鉴定其它 Termamyl-样 α -淀粉酶表面相应的和与其它暴露于溶剂的位置。使用 WHATIF 程序包中的 AACAVI 程序部分(G. Vriend, Whatif 和药物设计程序, J. Mol. Graph. 8, pp.52-56 (1990) 版本 19980317), 可以鉴定凸起的表面。

在本发明的一个实施方案中，变体含有一个或多个下列取代：E376K, S417T, A420Q, R, S356A, Y358F。

本发明人发现：通过将上述位置，即 E376, S417, A420, S356, Y358(使用 SEQ ID NO:4 的编号)的突变与下列一个或多个位置，即 K176, I201 和 H205 的突变联合，可以使高温，酸性 pH 和/或低钙浓度下的稳定性有更大的提高。

下列其它取代是优选的：

K176A, R, D, C, E, Q, G, H, I, L, M, N, F, P, S, T, W, Y, V；

I201A, R, D, C, E, Q, G, H, L, K, M, N, F, P, S, T, W, Y, V；

H205A, R, D, C, E, Q, G, I, L, K, M, N, F, P, S, T, W, Y, V；

如阐明本发明的实施例 2 所示，使用被称为 LE174 的杂合 α -淀粉酶作为亲代 Termamyl-样 α -淀粉酶，联合下列突变可以使稳定性增加：

K176+I201F+H205N+E376K+A420R 或

K176+I201F+H205N+S417T+A420Q 或

K176+I201F+H205N+S356A+Y358F。

本发明变体中的一般突变

优选本发明的变体除了含有上述变化外，还含有一个或多个修饰。因此，经修饰的 α -淀粉酶变体中一个或多个脯氨酸残基被非脯氨酸残基取代较为有利，所述非脯氨酸残基可以是任何可能的天然非脯氨酸残基，优选为丙

氨酸，甘氨酸，丝氨酸，苏氨酸，缬氨酸或亮氨酸。

类似地，优选亲代 α -淀粉酶被修饰的氨基酸残基中存在的一个或多个半胱氨酸残基被非半胱氨酸残基取代，所述非半胱氨酸残基可以是例如丝氨酸，丙氨酸，苏氨酸，甘氨酸，缬氨酸或亮氨酸。

另外，本发明的变体可以(作为唯一的修饰或与上述任何修饰联合)被修饰，以使对应于 SEQ ID NO:4 中 185-209 的氨基酸片段中存在的一个或多个 Asp 和/或 Glu 分别被 Asn 和/或 Gln 取代。另外，用 Arg 取代 Termamyl-样 α -淀粉酶中对应于 SEQ ID NO:4 中 185-209 的氨基酸片段中存在的一个或多个 Lys 残基也很有意义。

应理解本发明包含掺入了两个或多个上述修饰的变体。

另外，在本文所述的任何变体中导入点突变较为有利。

克隆编码本发明 α -淀粉酶的 DNA 序列

可使用本领域众所周知的多种方法，从产生所述 α -淀粉酶的任何细胞或微生物中分离出编码亲代 α -淀粉酶的 DNA 序列。首先，应使用源自产生待研究之 α -淀粉酶的生物体的染色体 DNA 或信使 RNA 构建基因组 DNA 和/或 cDNA 文库。然后，如果 α -淀粉酶的氨基酸序列是已知的，可合成同源的经标记的寡核苷酸探针，使用该探针从制备自所述生物体的基因组文库中鉴定编码 α -淀粉酶的克隆。或者，将含有与已知 α -淀粉酶基因同源之序列的经标记的寡核苷酸探针用作探针，使用较低严谨度的杂交和洗涤条件鉴定编码 α -淀粉酶的克隆。

另一种鉴定编码 α -淀粉酶的克隆的方法包括：将基因组 DNA 片段插入表达载体，如质粒，用所得基因组 DNA 文库转化 α -淀粉酶-阴性细菌，然后将经转化的细菌铺于含有 α -淀粉酶底物的琼脂上，从而鉴定出表达 α -淀粉酶的克隆。

或者，可通过已建立的标准方法合成制备编码酶的 DNA 序列，所述方法例如 S. L. Beaucage 和 M. H. Caruthers(1981)所述的磷酰胺法或 Matthes 等(1984)所述的方法。在磷酰胺法中，可在例如自动化的 DNA 合成仪中合成寡核苷酸，对其进行纯化，退火，再连接并克隆至适当的载体。

最终，DNA 序列可以是根据标准技术，通过连接合成的，基因组的或 cDNA 来源的片段(适当时是对应于完整 DNA 序列的多个部分的片段)而制备的混合的基因组和合成来源，混合的合成和 cDNA 来源或混合的基因组和

cDNA 来源的 DNA 序列。也可以使用特定的引物，如 US 4,683,202 或 R. K. Saiki 等(1988)所述的引物经聚合酶链反应(PCR)制备 DNA 序列。

定点诱变

一旦分离出编码 α -淀粉酶的 DNA 序列，并鉴定出合乎需要的突变位点，可使用合成的寡核苷酸导入突变。这些寡核苷酸含有侧翼于所需突变位点的核苷酸序列；在合成寡核苷酸的过程中插入突变的核苷酸。在特定的方法中，在携有 α -淀粉酶基因的载体中产生桥连 α -淀粉酶-编码序列的单链 DNA 缺口。然后，使携有所需突变的合成核苷酸与该单链 DNA 的同源部分退火。然后用 DNA 聚合酶 I (Klenow 片段)补平其余缺口，使用 T4 连接酶连接构建体。该方法的特例描述于 Morinaga 等(1984)。US 4,760,025 公开了通过对盒进行微小的改变而导入编码多个突变的寡核苷酸。然而，由于 Morinaga 法可以导入多种长度的多个寡核苷酸，因此可在任何一次导入甚至更多个突变。

Nelson 和 Long(1989)描述了另一种将突变导入编码 α -淀粉酶的 DNA 序列的方法。所述方法包括以 3 个步骤产生含有所需突变的 PCR 片段，所述突变是通过使用化学合成的 DNA 链作为 PCR 反应的一个引物而导入的。通过用限制性内切核酸酶裂解，可从 PCR-产生的片段中分离出携有突变的 DNA 片段，并将该片段重新插入表达质粒。

随机诱变

可在翻译成本文所示氨基酸序列的基因的至少 3 个部分中，或在整个基因内适当地进行随机诱变，所述诱变可以是定域的或区域-特异性的随机诱变。

利用本领域已知的任何方法，可以方便地对编码亲代 α -淀粉酶的 DNA 序列进行随机诱变。

关于上文，本发明的另一方面涉及产生亲代 α -淀粉酶的变体的方法，例如，其中变体相对于亲代而言，表现出经改变的或增加的热稳定性，所述方法包括：

- (a)对编码亲代 α -淀粉酶的 DNA 序列进行随机诱变，
- (b)在宿主细胞中表达步骤(a)中得到的突变 DNA 序列，和
- (c)筛选表达 α -淀粉酶变体的宿主细胞，所述变体相对于亲代 α -淀粉酶而言具有经改变的特性(即热稳定性)。

优选使用添加引物进行本发明上述方法中的步骤(a)。

例如，可使用适当的物理或化学诱变剂，使用适当的寡核苷酸，或通过对 DNA 序列进行 PCR 以产生诱变来进行随机诱变。另外，可使用这些诱变剂的任何组合进行随机诱变。诱变剂可以是例如诱导碱基转换，颠换，倒位，倒频，缺失和/或插入的试剂。

适用于本发明目的的物理或化学诱变剂的例子包括紫外线(UV)照射，羟胺，N-甲基-N'-硝基-N-亚硝基胍(MNNG)，邻甲基羟胺，亚硝酸，乙基甲磺酸(EMS)，亚硫酸氢钠，甲酸和核苷酸类似物。当使用这些诱变剂时，一般通过在适于发生诱变的条件下，在选定诱变剂的存在下保温待诱变的编码亲代酶的 DNA 序列来进行诱变，然后筛选具有所需特性的突变 DNA。

当利用寡核苷酸进行诱变时，在合成待改变位置处的寡核苷酸的过程中，可同时添加寡核苷酸和 3 个非-亲代核苷酸。添加的目的是避免不必要氨基酸的密码子。可通过任何公开的技术，例如使用 PCR，LCR 或适当时使用任何 DNA 聚合酶和连接酶，将添加的寡核苷酸掺入编码 α -淀粉酶的 DNA。

优选使用“恒随机的添加”进行添加，其中各个位置的野生型和突变的百分比是预先确定好的。另外，添加可以导致优先导入某些核苷酸，从而优先导入一个或多个特定的氨基酸残基。例如，进行添加后可在每个位置导入 90%野生型和 10%突变。选择添加方案的其它考虑基于遗传学以及蛋白质-结构的限制。可使用 DOPE 程序制订添加方案，所述程序可特别地确保不导入终止密码子。

当使用 PCR-产生的诱变时，可在增加核苷酸错误掺入的条件下，对经化学处理或未经处理的编码亲代 α -淀粉酶的基因进行 PCR(Deshler 1992; Leung 等, 技术, Vol.1, 1989, pp.11-15)。

可使用大肠杆菌(Fowler 等, Molec. Gen. Genet., 133, 1974, pp.179-191), 酿酒酵母或任何其它微生物的增变株对编码 α -淀粉酶的 DNA 进行随机诱变，诱变方法包括例如：将含有亲代糖基酶的质粒转化至增变株，培养含有质粒的增变株，从增变株中分离突变的质粒。随后，将突变的质粒转化至表达生物体。

待诱变的 DNA 序列可以方便地存在于基因组或 cDNA 文库中，所述文库制备自表达亲代 α -淀粉酶的生物体。或者，DNA 序列可以存在于适当的

载体，如质粒或噬菌体上，再与诱变剂一起保温或要不然暴露于诱变剂中。待诱变的 DNA 也可存在于宿主细胞中，或者整合于所述细胞的基因组中，或者存在于细胞所携带的载体上。最后，待诱变的 DNA 也可以是分离的形式。应理解接受随机诱变的 DNA 序列优选为 cDNA 或基因组 DNA 序列。

在某些情况下，在进行表达步骤 b) 或筛选步骤 c) 前，可以方便地扩增突变的 DNA 序列。可根据本领域已知的方法进行扩增，本发明优选的方法是：使用根据亲代酶的 DNA 或氨基酸序列制备的寡核苷酸引物进行 PCR 扩增。

与诱变剂保温或暴露于诱变剂之后，通过在允许发生表达的条件下培养携有突变 DNA 序列的适当宿主细胞来表达突变的 DNA。用于此目的的宿主细胞可以是被突变的 DNA 序列(任选其存在于载体上)转化的宿主细胞，或者是在诱变处理的过程中携有编码亲代酶的 DNA 序列的宿主细胞。适当宿主细胞的例子如下：革兰氏阳性细菌，如枯草芽孢杆菌，地衣芽孢杆菌，迟缓芽孢杆菌，短芽孢杆菌，嗜热脂肪芽孢杆菌，嗜碱芽孢杆菌，解淀粉芽孢杆菌，凝结芽孢杆菌，环状芽孢杆菌，灿烂芽孢杆菌，巨大芽孢杆菌，苏云金芽孢杆菌，浅青紫链霉菌或鼠灰链霉菌；和革兰氏阴性细菌，如大肠杆菌。

突变的 DNA 序列可进一步含有能使突变的 DNA 序列表达的 DNA 序列。

定域随机诱变

随机诱变可有利地定域于所述亲代 α -淀粉酶的一部分。例如，当酶的某些区域被鉴定为对酶的给定特性特别重要时，以及当预期修饰能导致具有改良特性的变体时，定域随机诱变较为有利。当亲代酶的三级结构已被阐明，且所述结构与酶的功能相关时，一般可鉴定出所述区域。

通过使用上述的 PCR 产生的诱变技术或本领域已知的任何其它适当的技术，可以方便地进行定域的或区域-特异性的随机诱变。或者，通过例如插入适当的载体来分离编码待修饰的 DNA 序列部分的 DNA 序列，随后可使用上述任何诱变方法对所述部分进行诱变。

提供 α -淀粉酶变体的其它方法

提供 α -淀粉酶变体的其它方法包括本领域已知的基因改组方法，所述方法包括例如 WO 95/22625(Affymax Technologies N. V.)和 WO 96/00343(Novo Nordisk A/S)所述的方法。

表达本发明的 α -淀粉酶变体

根据本发明，可使用表达载体，以酶的形式表达通过上述方法，或通过本领域已知的任何其它方法产生的编码变体的 DNA 序列，所述表达载体一般包括编码启动子，操纵子，核糖体结合位点，翻译起始信号的控制序列，并任选包括阻抑基因或多种激活基因。

携有编码本发明 α -淀粉酶变体的 DNA 序列的重组表达载体可以是能对其方便地进行重组 DNA 操作的任何载体，载体的选择经常取决于导入该载体的宿主细胞。因此，载体可以是自我复制的载体，即作为染色体外实体存在的载体，所述载体的复制独立于染色体的复制，所述载体包括例如质粒，噬菌体或染色体外元件，微型染色体或人工染色体。或者，载体可以是当导入宿主细胞时可以整合至宿主细胞基因组，并与整合了该载体的染色体一起复制的载体。

在载体中，DNA 序列应该与适当的启动子序列可操作相连。启动子可以是在选定宿主细胞中显示出转录活性的任何 DNA 序列，启动子可以得自编码与宿主细胞同源或异源之蛋白质的基因。介导编码本发明 α -淀粉酶变体之 DNA 序列转录，尤其是在细菌宿主中的转录的适当启动子的例子是：大肠杆菌 *lac* 操纵子的启动子，天蓝色链霉菌琼脂酶基因 *dagA* 启动子，地衣芽孢杆菌 α -淀粉酶基因(*amyL*)启动子，嗜热脂肪芽孢杆菌产麦芽糖淀粉酶基因(*amyM*)启动子，解淀粉芽孢杆菌 α -淀粉酶(*amyQ*)启动子，枯草芽孢杆菌 *xylA* 和 *xylB* 基因的启动子等。为了在真菌宿主中转录，有用的启动子的例子是得自编码下列酶的基因的启动子：米曲霉 TAKA 淀粉酶，米赫根毛霉 (*Rhizomucor miehei*)天冬氨酸蛋白酶，黑曲霉中性 α -淀粉酶，黑曲霉酸稳定的 α -淀粉酶，黑曲霉葡萄糖淀粉酶，米赫根毛霉脂肪酶，米曲霉碱性蛋白酶，米曲霉丙糖磷酸异构酶或构巢曲霉乙酰胺酶。

本发明的表达载体也含有适当的转录终止子，在真核生物中，还含有与编码本发明 α -淀粉酶变体的 DNA 序列可操作相连的聚-腺苷酸化序列。终止和聚腺苷酸化序列可适当地得自与启动子相同的来源。

载体可进一步含有能使载体在所述宿主细胞中复制的 DNA 序列。所述序列的例子是质粒 pUC19, pACYC177, pUB110, pE194, pAMB1 和 pIJ702 的复制起点。

载体也可含有选择标记，例如其产物可以补偿宿主细胞缺陷的基因，如枯草芽孢杆菌或地衣芽孢杆菌的 *dal* 基因，或能赋予抗生素抗性的基因，所

述抗生素抗性如氨基青霉素，卡那霉素，氯霉素或四环素抗性。另外，载体可含有曲霉属选择标记，如 *amdS*, *argB*, *niaD* 和 *sC*，产生潮霉素抗性的标记，或者可通过 WO 91/17243 中所述的共-转化来完成选择。

尽管细胞内表达在某些方面(例如当使用某些细菌作为宿主细胞时)有利，但一般优选在细胞外进行表达。通常，本文所述的芽孢杆菌 α -淀粉酶含有允许表达的蛋白酶分泌至培养基的前区。必要时，该前区可被不同的前区或信号序列取代，通过取代编码各个前区的 DNA 序列可以方便地实现此目的。

分别连接编码 α -淀粉酶变体的本发明 DNA 构建体，启动子，终止子和其它元件，并将它们插入含有复制所需信息的适当载体的方法是本领域技术人员众所周知的(参见例如 Sambrook 等，分子克隆：实验室手册，第 2 版，冷泉港，1989)。

含有上文所定义的本发明 DNA 构建体或表达载体的本发明细胞可以有利地用作宿主细胞，以重组产生本发明的 α -淀粉酶变体。可以方便地通过将编码变体的本发明 DNA 构建体(一个或多个拷贝)整合至宿主染色体，用所述 DNA 构建体转化细胞。一般认为整合是有利的，因为整合后 DNA 序列可以在细胞中更加稳定地维持。根据常规方法，例如通过同源或异源重组可以将 DNA 构建体整合至宿主染色体。或者，可用与不同类型的宿主细胞有关的上述表达载体转化细胞。

本发明的细胞可以是高等生物，如哺乳动物或昆虫的细胞，但优选其为微生物细胞，例如细菌或真菌(包括酵母)细胞。

适当细菌的例子是：革兰氏阳性细菌，如枯草芽孢杆菌，地衣芽孢杆菌，迟缓芽孢杆菌，短芽孢杆菌，嗜热脂肪芽孢杆菌，嗜碱芽孢杆菌，解淀粉芽孢杆菌，凝结芽孢杆菌，环状芽孢杆菌，灿烂芽孢杆菌，巨大芽孢杆菌，苏云金芽孢杆菌，或浅青紫链霉菌或鼠灰链霉菌；或革兰氏阴性细菌，如大肠杆菌。通过例如原生质体转化，或通过以本身已知的方式使用感受态细胞即可实现细菌的转化。

酵母生物体可有利地选自糖酵母属或裂殖酵母属的种，例如酿酒酵母。丝状真菌可有利地属于曲霉属的种，例如米曲霉或黑曲霉。通过以下方法可转化真菌细胞，所述方法包括按本身已知的方式形成原生质体，转化原生质体；接着再生细胞壁。转化曲霉属宿主细胞的适当方法描述于 EP 238 023。

另一方面，本发明涉及产生本发明 α -淀粉酶变体的方法，所述方法包括在有利于产生变体的条件下培养上述宿主细胞，并从细胞和/或培养基中回收变体。

用于培养细胞的培养基可以是适于培养所述宿主细胞，并表达本发明 α -淀粉酶变体的任何常规培养基。适当的培养基可以商购，或者可根据公开的配方(例如美国典型培养物保藏中心目录中所述的配方)制备。

通过众所周知的方法，包括通过离心或过滤将培养基与细胞分开，利用诸如硫酸铵的盐沉淀培养基中的蛋白质成分，接着利用层析法，如离子交换层析，亲和层析等，可以从培养基中方便地回收宿主细胞分泌的 α -淀粉酶变体。

工业应用

本发明的 α -淀粉酶变体具有有价值的特性，所述特性有多种工业用途。本发明的酶变体可用作洗涤，洗碟和硬-表面清洗去污剂组合物中的组分。多种变体对由淀粉产生增甜剂和乙醇，和/或织物脱浆特别有用。常规淀粉-转变过程，包括淀粉液化和/或糖化过程的条件描述于例如 US 3,912,590 和 EP 专利公开号 252,730 和 63,909。

由淀粉产生增甜剂：

“传统的”由淀粉转变为果糖浆的方法一般由三个连续的酶促过程组成，即液化过程，接着是糖化过程和异构化过程。在液化过程中，在温度为 95-160 °C，pH 值为 5.5-6.2 时，用 α -淀粉酶(如 Termamyl™)将淀粉降解约 2 小时，使其成为糊精。为了确保这些条件下的最适酶稳定性，需加入 1mM 钙(40ppm 游离的钙离子)。

液化过程之后，通过加入葡糖淀粉酶(如 AMG™)和脱支酶，如异淀粉酶或支链淀粉酶(如 Promozyme™)，将糊精转变为葡萄糖。在此步骤之前，将 pH 值降低为 4.5 以下，维持高温(95 °C 以上)，使液化 α -淀粉酶活性变性。将温度降低为 60 °C，加入葡糖淀粉酶和脱支酶。将糖化过程进行 24-72 小时。

糖化过程之后，将 pH 增加至 pH 值范围为 6-8，优选为 pH7.5，通过离子交换除去钙。然后使用固定化的葡萄糖异构酶(如 Sweetzyme™)将葡萄糖浆转变为高果糖浆。

此方法的至少 1 个酶促改良是可以想象到的。即降低了液化 α -淀粉酶的钙依赖性。添加游离的钙是确保 α -淀粉酶足够高的稳定性所必需的，但游离

的钙强烈地抑制了葡萄糖异构酶的活性，必需通过操作昂贵装置以除去游离的钙，使其水平降至 3-5ppm 以下。如果这种操作可以避免，液化过程在不加入游离钙离子时也能进行的话，就可以节省开支。

为了达到此目的，需要一种对钙的依赖性较低的 Termamyl-样 α -淀粉酶，该 α -淀粉酶在游离钙浓度低(<40ppm)时是稳定的，且具有高活性。这种 Termamyl-样 α -淀粉酶应该在 pH 范围为 4.5-6.5 时，优选在 pH 范围为 4.5-5.5 时具有最适 pH。

去污剂组合物

如上所述，可将本发明的变体适当掺入去污剂组合物中。关于去污剂组合物(如洗衣或洗碟去污剂)的相关成分，将变体配制在所述去污剂组合物中的适当方法，和相关类型的去污剂组合物的例子的其它细节可参见例如 WO 96/23874 和 WO 97/07202。

含有本发明变体的去污剂组合物还可含有一种或多种其它的酶，例如脂肪酶，角质酶，蛋白酶，纤维素酶，过氧化物酶或漆酶和/或另一种 α -淀粉酶。

以常规使用的浓度将本发明的 α -淀粉酶变体掺入去污剂。本发明希望使用常规的去污剂剂量水平，以相当于每升洗涤/洗碟液 0.00001-1mg(以纯的活性酶蛋白质计) α -淀粉酶的量掺入本发明的变体。

材料和方法

酶：

LE174 杂合 α -淀粉酶变体：LE174 是与 Termamyl 序列，即 SEQ ID NO:4 所示的地衣芽孢杆菌 α -淀粉酶相同的杂合 Termamyl-样 α -淀粉酶，不同之处在于：(成熟蛋白质)N-末端的 35 个氨基酸残基被 SEQ ID NO:5 所示的解淀粉芽孢杆菌 α -淀粉酶(BAN)成熟蛋白质 N-末端的 33 个氨基酸残基所取代，其进一步具有下列突变：H156Y+A181T+N190F+A209V+Q264S(使用 SEQ ID NO:4 的编号)。

构建 pSNK101

该大肠杆菌/芽孢杆菌穿梭载体可用于导入突变，而不用在大肠杆菌中表达 α -淀粉酶，然后以 α -淀粉酶在芽孢杆菌中具有活性的方式修饰该载体。按下述构建载体：通过将含有来源于大肠杆菌之片段的 1.2kb 片段插入 α -淀粉酶基因 5'编码区的 PstI 位点，来灭活 pX 载体(在 amyL 中具有下列改变的 pDN1528：BAN(1-33), H156Y, A181T, N190F, A209V, Q264S，实施例 1 中

将进一步描述质粒 pDN1528)中的 α -淀粉酶基因。使用正向引物 1 : 5'-gacctgcagtcaggc aacta-3' (SEQ ID NO:28) 和反向引物 1 : 5'-tagagtcgacctgcaggcat-3'(SEQ ID NO:29), 由 pUC19(GenBank 登记号: X02514) 扩增此片段。于 37 °C, 用 PstI 消化 PCR 扩增子和含有 α -淀粉酶基因的 pX 质粒达 2 小时。室温下连接 pX 载体片段和来源于大肠杆菌的扩增子达 1 小时, 通过电转化转化至大肠杆菌。所得载体被称为 pSnK101。

该大肠杆菌/芽孢杆菌穿梭载体可用于导入突变, 而不用在大肠杆菌中表达 α -淀粉酶, 然后以 α -淀粉酶在芽孢杆菌中具有活性的方式修饰该载体。按下述构建载体: 通过将含有来源于大肠杆菌之片段的 1.2kb 片段插入 α -淀粉酶基因 5'编码区的 PstI 位点, 来灭活 pX 载体(在 amyL 中具有下列改变的 pDN1528 : BAN(1-33), H156Y+A181T+N190F+A209V+Q264S, 实施例 1 中将进一步描述质粒 pDN1528)中的 α -淀粉酶基因。使用正向引物 2 : 5'-gacctgcagtcaggcaacta-3' (SEQ ID NO:30) 和反向引物 2 : 5'-tagagtcgacctgcaggcat-3'(SEQ ID NO:31), 由 pUC19(GenBank 登记号: X02514) 扩增此片段。于 37 °C, 用 PstI 消化 PCR 扩增子和含有 α -淀粉酶基因的 pX 质粒达 2 小时。室温下连接 pX 载体片段和来源于大肠杆菌的扩增子达 1 小时, 通过电转化转化至大肠杆菌。所得载体被称为 pSnK101。

低 pH 滤膜试验

于 37 °C, 将芽孢杆菌文库铺于含 10 μ g/ml 氯霉素的 TY 琼脂平板上的醋酸纤维素(OE67, Schleicher & Schuell, Dassel, 德国)和硝酸纤维素滤膜(Protran-Ba 85, Schleicher & Schuell, Dassel, 德国)夹层, 至少放置 21 小时。醋酸纤维素层位于 TY 琼脂平板上。

在铺平板之后, 但在保温之前用针头特异性地标记每个滤膜夹层, 以使阳性变体定位于滤膜上, 将结合有变体的硝酸纤维素滤膜转移至含柠檬酸盐缓冲液, pH4.5 的容器, 90 °C 保温 15 分钟。室温下, 将含菌落的醋酸纤维素滤膜储存于 TY-平板上待用。保温之后, 在含有 1% 琼脂糖, 0.2% 淀粉(皆溶于柠檬酸盐缓冲液, pH6.0)的试验平板上检测残留的活性。按与滤膜夹层相同的方式标记含硝酸纤维素滤膜的试验平板, 50 °C 保温 2 小时。取出滤膜之后, 用 10% Lugol 溶液染色试验平板。降解淀粉的变体被测定为深蓝色背景上的白点, 然后在储存平板上鉴定该变体。在与第一次筛选相同的条件下再将阳性变体筛选 2 次。

二级筛选

重新筛选之后，从储存平板上挑选阳性转化子，在二级平板试验中进行检测。于 37 °C，在 5ml LB+氯霉素中将阳性转化子培养 22 小时。于 90 °C，在柠檬酸盐缓冲液，pH4.5 中保温每个阳性转化子和对照 LE174 变体的芽孢杆菌培养物，在 0, 10, 20, 30, 40, 60 和 80 分钟时取样。将 3 微升样品点在试验平板上。用 10% Lugol 溶液染色试验平板。将试验板上残留活性高于对照的晕圈定为改良变体。通过核苷酸测序测定这些改良的变体。

发酵和纯化 α -淀粉酶变体

在含有 15 μ g/ml 氯霉素的 LB-琼脂平板上，用 - 80 °C 原种划线携有相关表达质粒的枯草芽孢杆菌菌株，37 °C 培养过夜。将菌落转移至 500ml 摇瓶内的 100ml 添加有 15 μ g/ml 氯霉素的 BPX 培养基中。

BPX 培养基的组成：

马铃薯淀粉	100g/l
大麦粉	50g/l
BAN 5000 SKB	0.1g/l
酪蛋白酸钠	10g/l
大豆粉	20g/l
Na ₂ HP ₄ , 12H ₂ O	9g/l
Pluronic™	0.1g/l

于 37 °C，以 270rpm 振荡培养 5 天。

通过以 4500rpm 离心 20-25 分钟，从发酵肉汤中除去细胞和细胞碎片。然后，过滤上清液，得到十分清亮的溶液。在 UF-滤器(10000 截留膜)上浓缩和冲洗滤液，将缓冲液变为 20mM 醋酸盐，pH5.5。将 UF-滤液上样于 S-Sepharose F. F.，用含 0.2M NaCl 的相同缓冲液逐步洗脱。洗脱液对 10mM Tris, pH9.0 透析，然后上样于 Q-sepharose F. F.，用超过 6 倍柱体积的 0-0.3M NaCl 线性梯度进行洗脱。集中含有活性(通过 Phadebas 试验测定)的级分，将 pH 调节至 pH7.5，用 0.5% W/vol 的活性炭处理 5 分钟除去残留的颜色。

稳定性的测定

使用相同的设置进行所有的稳定性试验。方法是：

在相关条件(1-4)下保温酶。在 0, 5, 10, 15 和 30 分钟时取样，用试验缓冲液(0.1M 50mM Britton 缓冲液 pH7.3)稀释 25 倍(对所有取出的样品都作相

同的稀释), 在标准条件下(pH7.3, 37 °C)使用 Phadebas 试验(Pharmacia)测定活性。

将保温前(0 分钟)测定的活性用作参照(100%)。百分比的下降被计算为保温时间的函数。表中显示了保温 30 分钟之后残留的活性。

活性测定 - (KNU)

在根据下列条件测定 α -淀粉酶的 Novo Nordisk's 标准方法中, 1000 个 α -淀粉酶单位(1KNU)是每小时降解 5.26g 淀粉(Merck, Amylum Solubile, Erg. B6, 批号 9947275)所需的酶量:

底物	可溶性淀粉
溶剂中的钙含量	0.0043M
反应时间	7-20 分钟
温度	37 °C
pH	5.6

Novo Nordisk's 分析法(AF 9)的详细描述即要即得。

比活的测定

检测 α -淀粉酶的活性

通过利用 Phadebas[®]片为底物的方法测定 α -淀粉酶的活性。Phadebas 片(Phadebas[®] Amylase Test, 由 Pharmacia Diagnostic 提供)含有交联的不可溶的蓝色淀粉聚合物, 所述聚合物已与牛血清白蛋白和缓冲物质混合, 并被制成片状。

对每一个单次测定而言, 将 1 片 Phadebas[®]悬浮于含有 5ml 50mM Britton-Robinson 缓冲液(50mM 醋酸, 50mM 磷酸, 50mM 硼酸, 0.1mM CaCl₂, 用 NaOH 将 pH 调节至所需的值)的试管中。在所需温度的水浴中进行试验。用 x ml 50mM Britton-Robinson 缓冲液稀释待检测的 α -淀粉酶。在 5ml 50mM Britton-Robinson 缓冲液中加入 1ml 这种 α -淀粉酶溶液。通过 α -淀粉酶水解淀粉, 给出可溶性的蓝色片段。用分光光度计测定所得蓝色溶液在 620nm 处的光吸收值, 该值是 α -淀粉酶活性的函数。

重要的是, 保温 10 或 15 分钟之后(检测时间)测定的 620nm 光吸收值范围是 0.2 至 2.0 个光吸收单位。在此光吸收值范围内, 活性和光吸收值之间呈现线性关系(Lambert-Beer 定律)。因此, 必须调节酶的稀释度以符合此标准。在特定的一套条件(温度, pH, 反应时间, 缓冲液浓度)下, 1mg 给定

的 α -淀粉酶可水解一定量的底物，并产生蓝色。测定 620nm 处的颜色强度。在给定的一套条件下，测定的光吸收值与所述 α -淀粉酶的比活(活性/mg 纯 α -淀粉酶蛋白质)直接成比例。

实施例

实施例 1

通过随机诱变构建 Termamyl-样 LE174 α -淀粉酶变体，与亲代酶相比，所述变体在低 pH 下具有改善的稳定性，其稳定性对钙离子的依赖性降低

随机诱变

为了改善亲代 LE174 α -淀粉酶变体在低 pH 和低钙浓度下的稳定性，在预先选定的区域进行随机诱变：

所述区域是：

区域	残基
SERI	A425-Y438
SERII	W411-L424
SERIII	G397-G410
SERV	T369-H382
SERVII	G310-F323
SERIX	L346-P359

对 6 个区域中的每一个而言，在上述区域中的每一个核苷酸位置使用相同的突变率(主要成分 97%，三种其余核苷酸中的每一种各 1%共给出 3%突变)，例如对 A425 密码子的第 1 位而言：97%C, 1%A, 1%T, 1%G 来合成随机寡核苷酸。6 个随机的寡核苷酸和必要时使用的互补 SOE 辅助引物示于具有下列 4 种核苷酸分布的表 1-6：

表 1

RSERI: 5'-GC GTT TTG CCG GCC GAC ATA 312 234 322 243 333 133
444 233 423 242 212 211 243 343 CAA ACC TGA ATT-3' (SEQ ID NO:
15)

表 2

RSERII: 5'-GC GTT TTG CCG GCC GAC ATA CAT TCG CTT TGC CCC ACC
GGG TCC GTC TGT TAT TAA TGC CGC 311 133 241 122 243 113 341 432
423 433 223 332 242 331 GCC GAC AAT GTC ATG GTG-3' (SEQ ID NO:
16)

表 3

RSERIII: 5'-GTC GCC TTC CCT TGT CCA 433 413 112 423 124 424 423
411 121 123 124 324 243 233 GTA CGC ATA CTG TTT TCT-3' (SEQ ID
NO: 17)

辅助引物 FSERIII : 5'-TGG ACA AGG GAA GGC GAC AG-3' (SEQ ID
NO:18)

表 4

RSERV: 5-TAA GAT CGG TTC AAT TTT 424 222 311 443 144 112 223
434 324 441 423 233 222 342 CCC GTA CAT ATC CCC GTA GAA-3 (SEQ
ID NO: 19)

辅助引物 FSERV : 5'-AAA ATT GAA CCG ATC TTA-3' (SEQ ID
NO:20)

表 5

FSERVII: 5'-TT CCA TGC TGC ATC GAC ACA GGG AGG CGG CTA TGA TAT
GAG GAA ATT GCT GAA 344 213 442 342 223 311 431 233 422 411 123
442 213 122 TGT CGA TAA CCA-3' (SEQ ID NO: 21)

辅助引物 RSERVII : 5'-TGT CGA TGC AGC ATG GAA-3' (SEQ ID
NO:22)

表 6

FSERIX: 5'-GT CCA AAC ATG GTT TAA GCC 432 243 221 343 222 212
232 313 114 441 123 244 121 333 TCA GGT TTT CTA CGG GGA-3' (SEQ
ID NO: 23)

辅助引物 RSERIX : 5'-GGC TTA AAC CAT GTT TGG AC-3' (SEQ ID
NO:24)

每个突变的核苷酸位置的核苷酸分布:

1 : 97%A, 1%T, 1%C, 1%G

2 : 97%T, 1%A, 1%C, 1%G

3 : 97%C, 1%A, 1%T, 1%G

4 : 97%G, 1%A, 1%T, 1%C

构建质粒文库

使用引物 1B : 5'-CGA TTG CTG ACG CTG TTA TTT GCG-3'和表 1 所示的随机寡核苷酸, 或表 2 所示的随机寡核苷酸, 经 PCR 扩增两个约 1.4kb 的片段。用 EcoRV 和 EagI 消化 pSnK101 和所述 PCR 片段达 2 小时。纯化约 3.6kb 的载体片段和约 1.3kb 的 PCR 片段, 连接过夜, 转化至大肠杆菌, 然后进一步按下文所述转化至芽孢杆菌宿主菌株。使用表 3-6 所示的针对各个区域的随机寡核苷酸(在图 2 中通称为 aSER 和 bSER), 和覆盖了 LE174 序列中的 EcoRV 和 EagI 位点的特定的地衣芽孢杆菌引物 1B(SEQ ID NO:26) 和#63: 5'-CTA TCT TTG AAC ATA AAT TGA AAC C-3'(SEQ ID NO:27), 经重叠延伸法(Horton 等, 基因, 77(1989), pp61-68)产生 PCR-文库-片段。图 2 显示了 PCR 策略。将 PCR 片段克隆至大肠杆菌/芽孢杆菌穿梭载体 pSNK101(见材料和方法), 从而在大肠杆菌中诱变, 并立即在枯草芽孢杆菌中表达, 以防止 α -淀粉酶在大肠杆菌中的致死积累。在大肠杆菌中建立了克隆的 PCR 片段之后, 消化质粒, 得到经修饰的 pUC19 片段, 物理连接启动子和突变的 Termamyl 基因, 在芽孢杆菌宿主中进行表达。

筛选

在上文“材料和方法”一节中所述的低 pH 滤膜试验中筛选 6 个文库。

按实施例 1 所述制备下文实施例 2 表中所列的所有变体。

实施例 2

测定稳定性

通常, 为了改善 95 °C-105 °C 时的稳定性, 在 pH6.0-6.2, 加入约 40ppm 游离钙时进行工业液化过程。制备本发明的变体, 使之在

1.pH 低于 pH6.2 和/或

2.游离钙水平低于 40ppm 游离钙时, 稳定性改善。

使用在酸性 pH(pH5.0), 5ppm 游离钙的存在下测定稳定性的试验测定稳定性的增加。

在下列条件下保温 10 μ g 变体: 0.1M 醋酸盐溶液, pH 调节至 pH5.0, 含有 5ppm 钙和 5%w/w 普通的玉米淀粉(不含钙)。在 95 °C 水浴中保温 30 分钟。

结果:

在 pH5.0, 5ppm 钙, 95 °C 保温时增加的稳定性

保温时间	LE174, 具有 K176R+ I201F+ H205N	LE174, 具有 K176R+ I201F+ H205N+ E376K+ A420R	LE174, 具有 K176R+ I201F+ H205N+ S417T+ A420Q	LE174, 具有 K176R+ I201F+ H205N+ S356A+ Y358F
0	100	100	100	100
5	65	61	66	66
10	58	53	60	59
15	51	48	55	56
30	36	39	45	49

比活测定

使用 Phadebas 试验(Pharmacia)(上述)将比活测定为活性/mg 酶。使用本文材料和方法一节中所述的 α -淀粉酶试验测定活性。

具有下列取代的 LE174 :

K176R+I201F+H205N

所测比活为: 13400NU/mg

具有下列取代的 LE174 :

K176R+I201F+H205N+E376K+A420R:

所测比活为: 14770NU/mg

具有下列取代的 LE174 :

K176R+I201F+H205N+S417T+A420Q:

所测比活为: 16670NU/mg

具有下列取代的 LE174 :

K176R+I201F+H205N+S356A+Y358F:

所测比活为: 15300NU/mg

参考文献

- Klein, C 等, 生物化学 1992, 31, 8740-8746 ,
- Mizuno, H 等, 分子生物学杂志, (1993), 234, 1282-1283 ,
- Chang, C 等, 分子生物学杂志, (1993), 229, 235-238 ,
- Larson, S. B., 分子生物学杂志, (1994), 235, 1560-1584 ,
- Lawson, C. L., 分子生物学杂志, (1994), 236, 590-600 ,
- Qian, M 等, 分子生物学杂志, (1993), 231, 785-799 ,
- Brady, R. L 等, Acta Crystallogr. sect. B. 47, 527-535 ,
- Swift, H. J 等, Acta Crystallogr. sect. B. 47, 535-544 ,
- A. Kadziola 博士论文: “通过 X-射线晶体学研究的大麦 α -淀粉酶及其与底物类似物抑制剂的复合物”, 哥本哈根大学化学系, 1993 ,
- MacGregor, E. A., 食品水胶体, 1987, Vol.1, No.5-6, p.
- B. Diderichsen 和 L. Christiansen, 嗜热脂肪芽孢杆菌中克隆产麦芽糖的 α -淀粉酶, FEMS Microbiol letters: 56: pp.53-60(1988),
- Hudson 等, 实用免疫学, 第3版(1989), Blackwell Scientific Publications ,
- Sambrook 等, 分子克隆: 实验室手册, 第2版, 冷泉港, 1989 ,
- S. L. Beaucage 和 M. H. Caruthers, 四面体通讯, 22, 1981, pp.1859-1869 ,
- Matthes 等, The EMBO J. 3, 1984, pp.801-805 ,
- R. K. Saiki 等, 科学, 239, 1988, pp.487-491 ,
- Morinaga 等, (1984, 生物技术 2:646-639),
- Nelson 和 Long, 分析生物化学, 180, 1989, pp. 147-151 ,
- Hunkapiller 等, 1984, 自然, 310:105-111 ,
- R. Higuchi, B. Krummel 和 R. K. Saiki(1988), 体外制备和特异性诱变 DNA 片段的一般方法: 研究蛋白质和 DNA 的相互作用, 核酸研究, 16:7351-7367 ,
- Dubnau 等, 1971, 分子生物学杂志, 56, pp.209-221 ,
- Gryczan 等, 1978, 细菌学杂志, 134, pp.318-329 ,
- S. D. Erlich, 1977, Proc. Natl. Acad. Sci. 74, pp.1680-1682 ,
- Boel 等, 1990, 生物化学, 29, pp.6244-6249 .

序列表

(1) 序列资料:

(i) 申请人:

(A) 姓名: NOVO NORDISK A/S

(B) 街道: Novo Alle

(C) 城市: DK-2880 Bagsvaerd

(E) 国家: 丹麦

(F) 邮编: DK-2880

(G) 电话: +45 4444 8888

(H) 电传: +45 4449 3256

(ii) 发明标题: α -淀粉酶变体

(iii) 序列数: 32

(iv) 计算机可读形式:

(A) 介质类型: 软盘

(B) 计算机: IBM PC 兼容机

(C) 操作系统: PC-DOS/MS-DOS

(2) SEQ ID NO: 1 的资料:

(i) 序列特征:

(A) 长度: 485 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 单链

(D) 拓扑构型: 线性

(ii) 分子类型: 蛋白质

(iii) 生物: 芽孢杆菌属的种

(xi) 序列描述: SEQ ID NO: 1

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

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

Gly Asn Ser Ser His Pro Asn Ser Gly Leu Ala Thr Ile Met Ser Asp
 420 425 430

Gly Pro Gly Gly Asn Lys Trp Met Tyr Val Gly Lys Asn Lys Ala Gly
 435 440 445

Gln Val Trp Arg Asp Ile Thr Gly Asn Arg Thr Gly Thr Val Thr Ile
 450 455 460

Asn Ala Asp Gly Trp Gly Asn Phe Ser Val Asn Gly Gly Ser Val Ser
 465 470 475 480

Val Trp Val Lys Gln
 485

(2) SEQ ID NO: 2 的资料:

(i) 序列特征:

- (A) 长度: 485 个氨基酸
- (B) 类型: 氨基酸
- (C) 链型: 单链
- (D) 拓扑构型: 线性

(ii) 分子类型: 蛋白质

(iii) 生物: 芽孢杆菌属的种

(xi) 序列描述: SEQ ID NO: 2

His His Asn Gly Thr Asn Gly Thr Met Met Gln Tyr Phe Glu Trp His
 1 5 10 15

Leu Pro Asn Asp Gly Asn His Trp Asn Arg Leu Arg Asp Asp Ala Ser
 20 25 30

Asn Leu Arg Asn Arg Gly Ile Thr Ala Ile Trp Ile Pro Pro Ala Trp
 35 40 45

Lys Gly Thr Ser Gln Asn Asp Val Gly Tyr Gly Ala Tyr Asp Leu Tyr
 50 55 60

Asp Leu Gly Glu Phe Asn Gln Lys Gly Thr Val Arg Thr Lys Tyr Gly
 65 70 75 80

Thr Arg Ser Gln Leu Glu Ser Ala Ile His Ala Leu Lys Asn Asn Gly
 85 90 95

Val Gln Val Tyr Gly Asp Val Val Met Asn His Lys Gly Gly Ala Asp
 100 105 110

Ala Thr Glu Asn Val Leu Ala Val Glu Val Asn Pro Asn Asn Arg Asn
 115 120 125

Gln Glu Ile Ser Gly Asp Tyr Thr Ile Glu Ala Trp Thr Lys Phe Asp
 130 135 140

Phe Pro Gly Arg Gly Asn Thr Tyr Ser Asp Phe Lys Trp Arg Trp Tyr
 145 150 155 160

His Phe Asp Gly Val Asp Trp Asp Gln Ser Arg Gln Phe Gln Asn Arg
 165 170 175
 Ile Tyr Lys Phe Arg Gly Asp Gly Lys Ala Trp Asp Trp Glu Val Asp
 180 185 190
 Ser Glu Asn Gly Asn Tyr Asp Tyr Leu Met Tyr Ala Asp Val Asp Met
 195 200 205
 Asp His Pro Glu Val Val Asn Glu Leu Arg Arg Trp Gly Glu Trp Tyr
 210 215 220
 Thr Asn Thr Leu Asn Leu Asp Gly Phe Arg Ile Asp Ala Val Lys His
 225 230 235 240
 Ile Lys Tyr Ser Phe Thr Arg Asp Trp Leu Thr His Val Arg Asn Ala
 245 250 255
 Thr Gly Lys Glu Met Phe Ala Val Ala Glu Phe Trp Lys Asn Asp Leu
 260 265 270
 Gly Ala Leu Glu Asn Tyr Leu Asn Lys Thr Asn Trp Asn His Ser Val
 275 280 285
 Phe Asp Val Pro Leu His Tyr Asn Leu Tyr Asn Ala Ser Asn Ser Gly
 290 295 300
 Gly Asn Tyr Asp Met Ala Lys Leu Leu Asn Gly Thr Val Val Gln Lys
 305 310 315 320
 His Pro Met His Ala Val Thr Phe Val Asp Asn His Asp Ser Gln Pro
 325 330 335
 Gly Glu Ser Leu Glu Ser Phe Val Gln Glu Trp Phe Lys Pro Leu Ala
 340 345 350
 Tyr Ala Leu Ile Leu Thr Arg Glu Gln Gly Tyr Pro Ser Val Phe Tyr
 355 360 365
 Gly Asp Tyr Tyr Gly Ile Pro Thr His Ser Val Pro Ala Met Lys Ala
 370 375 380
 Lys Ile Asp Pro Ile Leu Glu Ala Arg Gln Asn Phe Ala Tyr Gly Thr
 385 390 395 400
 Gln His Asp Tyr Phe Asp His His Asn Ile Ile Gly Trp Thr Arg Glu
 405 410 415
 Gly Asn Thr Thr His Pro Asn Ser Gly Leu Ala Thr Ile Met Ser Asp
 420 425 430
 Gly Pro Gly Gly Glu Lys Trp Met Tyr Val Gly Gln Asn Lys Ala Gly
 435 440 445
 Gln Val Trp His Asp Ile Thr Gly Asn Lys Pro Gly Thr Val Thr Ile
 450 455 460
 Asn Ala Asp Gly Trp Ala Asn Phe Ser Val Asn Gly Gly Ser Val Ser
 465 470 475 480
 Ile Trp Val Lys Arg
 485

(2) SEQ ID NO: 3 的资料:

(i) 序列特征:

(A) 长度: 514 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 单链

(D) 拓扑构型: 线性

(ii) 分子类型: 蛋白质

(iii) 生物: 嗜热脂肪芽孢杆菌

(xi) 序列描述: SEQ ID NO: 3

```

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

```

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

(2) SEQ ID NO: 4 的资料:

(i) 序列特征:

- (A) 长度: 483 个氨基酸
- (B) 类型: 氨基酸
- (C) 链型: 单链
- (D) 拓扑构型: 线性

(ii) 分子类型: 蛋白质

(iii) 生物: 地衣芽孢杆菌

(xi) 序列描述: SEQ ID NO: 4

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

Val Thr Phe Val Asp Asn His Asp Thr Gln Pro Gly Gln Ser Leu Glu
 325 330 335
 Ser Thr Val Gln Thr Trp Phe Lys Pro Leu Ala Tyr Ala Phe Ile Leu
 340 345 350
 Thr Arg Glu Ser Gly Tyr Pro Gln Val Phe Tyr Gly Asp Met Tyr Gly
 355 360 365
 Thr Lys Gly Asp Ser Gln Arg Glu Ile Pro Ala Leu Lys His Lys Ile
 370 375 380
 Glu Pro Ile Leu Lys Ala Arg Lys Gln Tyr Ala Tyr Gly Ala Gln His
 385 390 395 400
 Asp Tyr Phe Asp His His Asp Ile Val Gly Trp Thr Arg Glu Gly Asp
 405 410 415
 Ser Ser Val Ala Asn Ser Gly Leu Ala Ala Leu Ile Thr Asp Gly Pro
 420 425 430
 Gly Gly Ala Lys Arg Met Tyr Val Gly Arg Gln Asn Ala Gly Glu Thr
 435 440 445
 Trp His Asp Ile Thr Gly Asn Arg Ser Glu Pro Val Val Ile Asn Ser
 450 455 460
 Glu Gly Trp Gly Glu Phe His Val Asn Gly Gly Ser Val Ser Ile Tyr
 465 470 475 480
 Val Gln Arg

(2) SEQ ID NO: 5 的资料:

(i) 序列特征:

- (A) 长度: 480 个氨基酸
- (B) 类型: 氨基酸
- (C) 链型: 单链
- (D) 拓扑构型: 线性

(ii) 分子类型: 蛋白质

(iii) 生物: 解淀粉芽孢杆菌

(xi) 序列描述: SEQ ID NO: 5

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

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

Gly Gly Ser Lys Arg Met Tyr Ala Gly Leu Lys Asn Ala Gly Glu Thr
 435 440 445
 Trp Tyr Asp Ile Thr Gly Asn Arg Ser Asp Thr Val Lys Ile Gly Ser
 450 455 460
 Asp Gly Trp Gly Glu Phe His Val Asn Asp Gly Ser Val Ser Ile Tyr
 465 470 475 480

(2) SEQ ID NO: 6 的资料:

(i) 序列特征:

(A) 长度: 485 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 单链

(D) 拓扑构型: 线性

(ii) 分子类型: 蛋白质

(iii) 生物: 芽孢杆菌属的种

(xi) 序列描述: SEQ ID NO: 6

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

Ile Tyr Lys Phe Arg Gly His Gly Lys Ala Trp Asp Trp Glu Val Asp
 180 185 190
 Thr Glu Asn Gly Asn Tyr Asp Tyr Leu Met Tyr Ala Asp Ile Asp Met
 195 200 205
 Asp His Pro Glu Val Val Asn Glu Leu Arg Asn Trp Gly Val Trp Tyr
 210 215 220
 Thr Asn Thr Leu Gly Leu Asp Gly Phe Arg Ile Asp Ala Val Lys His
 225 230 235 240
 Ile Lys Tyr Ser Phe Thr Arg Asp Trp Ile Asn His Val Arg Ser Ala
 245 250 255
 Thr Gly Lys Asn Met Phe Ala Val Ala Glu Phe Trp Lys Asn Asp Leu
 260 265 270
 Gly Ala Ile Glu Asn Tyr Leu Gln Lys Thr Asn Trp Asn His Ser Val
 275 280 285
 Phe Asp Val Pro Leu His Tyr Asn Leu Tyr Asn Ala Ser Lys Ser Gly
 290 295 300
 Gly Asn Tyr Asp Met Arg Asn Ile Phe Asn Gly Thr Val Val Gln Arg
 305 310 315 320
 His Pro Ser His Ala Val Thr Phe Val Asp Asn His Asp Ser Gln Pro
 325 330 335
 Glu Glu Ala Leu Glu Ser Phe Val Glu Glu Trp Phe Lys Pro Leu Ala
 340 345 350
 Tyr Ala Leu Thr Leu Thr Arg Glu Gln Gly Tyr Pro Ser Val Phe Tyr
 355 360 365
 Gly Asp Tyr Tyr Gly Ile Pro Thr His Gly Val Pro Ala Met Arg Ser
 370 375 380
 Lys Ile Asp Pro Ile Leu Glu Ala Arg Gln Lys Tyr Ala Tyr Gly Lys
 385 390 395 400
 Gln Asn Asp Tyr Leu Asp His His Asn Ile Ile Gly Trp Thr Arg Glu
 405 410 415
 Gly Asn Thr Ala His Pro Asn Ser Gly Leu Ala Thr Ile Met Ser Asp
 420 425 430
 Gly Ala Gly Gly Ser Lys Trp Met Phe Val Gly Arg Asn Lys Ala Gly
 435 440 445
 Gln Val Trp Ser Asp Ile Thr Gly Asn Arg Thr Gly Thr Val Thr Ile
 450 455 460
 Asn Ala Asp Gly Trp Gly Asn Phe Ser Val Asn Gly Gly Ser Val Ser
 465 470 475 480
 Ile Trp Val Asn Lys
 485

(2) SEQ ID NO: 7 的资料:

(i) 序列特征:

- (A) 长度: 485 个氨基酸
 (B) 类型: 氨基酸
 (C) 链型: 单链
 (D) 拓扑构型: 线性
 (ii) 分子类型: 蛋白质
 (iii) 生物: 芽孢杆菌属的种
 (xi) 序列描述: SEQ ID NO: 7

```

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

```

Gly Ala Ile Glu Asn Tyr Leu Asn Lys Thr Ser Trp Asn His Ser Val
 275 280 285
 Phe Asp Val Pro Leu His Tyr Asn Leu Tyr Asn Ala Ser Asn Ser Gly
 290 295 300
 Gly Tyr Tyr Asp Met Arg Asn Ile Leu Asn Gly Ser Val Val Gln Lys
 305 310 315 320
 His Pro Thr His Ala Val Thr Phe Val Asp Asn His Asp Ser Gln Pro
 325 330 335
 Gly Glu Ala Leu Glu Ser Phe Val Gln Gln Trp Phe Lys Pro Leu Ala
 340 345 350
 Tyr Ala Leu Val Leu Thr Arg Glu Gln Gly Tyr Pro Ser Val Phe Tyr
 355 360 365
 Gly Asp Tyr Tyr Gly Ile Pro Thr His Gly Val Pro Ala Met Lys Ser
 370 375 380
 Lys Ile Asp Pro Leu Leu Gln Ala Arg Gln Thr Phe Ala Tyr Gly Thr
 385 390 395 400
 Gln His Asp Tyr Phe Asp His His Asp Ile Ile Gly Trp Thr Arg Glu
 405 410 415
 Gly Asn Ser Ser His Pro Asn Ser Gly Leu Ala Thr Ile Met Ser Asp
 420 425 430
 Gly Pro Gly Gly Asn Lys Trp Met Tyr Val Gly Lys Asn Lys Ala Gly
 435 440 445
 Gln Val Trp Arg Asp Ile Thr Gly Asn Arg Thr Gly Thr Val Thr Ile
 450 455 460
 Asn Ala Asp Gly Trp Gly Asn Phe Ser Val Asn Gly Gly Ser Val Ser
 465 470 475 480
 Val Trp Val Lys Gln
 485

(2) SEQ ID NO: 8 的资料:

(i) 序列特征:

(A) 长度: 485 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 单链

(D) 拓扑构型: 线性

(ii) 分子类型: 蛋白质

(iii) 生物: 芽孢杆菌属的种

(xi) 序列描述: SEQ ID NO: 8

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

Gly Glu Ser Leu Glu Ser Phe Val Gln Glu Trp Phe Lys Pro Leu Ala
 340 345 350
 Tyr Ala Leu Ile Leu Thr Arg Glu Gln Gly Tyr Pro Ser Val Phe Tyr
 355 360 365
 Gly Asp Tyr Tyr Gly Ile Pro Thr His Ser Val Pro Ala Met Lys Ala
 370 375 380
 Lys Ile Asp Pro Ile Leu Glu Ala Arg Gln Asn Phe Ala Tyr Gly Thr
 385 390 395 400
 Gln His Asp Tyr Phe Asp His His Asn Ile Ile Gly Trp Thr Arg Glu
 405 410 415
 Gly Asn Thr Thr His Pro Asn Ser Gly Leu Ala Thr Ile Met Ser Asp
 420 425 430
 Gly Pro Gly Gly Glu Lys Trp Met Tyr Val Gly Gln Asn Lys Ala Gly
 435 440 445
 Gln Val Trp His Asp Ile Thr Gly Asn Lys Pro Gly Thr Val Thr Ile
 450 455 460
 Asn Ala Asp Gly Trp Ala Asn Phe Ser Val Asn Gly Gly Ser Val Ser
 465 470 475 480
 Ile Trp Val Lys Arg
 485

(2) SEQ ID NO: 9 的资料:

(i) 序列特征:

- (A) 长度: 1455 个碱基对
- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑构型: 线性

(ii) 分子类型: DNA(基因组)

(iii) 生物: 芽孢杆菌属的种

(xi) 序列描述: SEQ ID NO: 9

CATCATAATG GAACAAATGG TACTATGATG CAATATTTTCG AATGGTATTT GCCAAATGAC 60
 GGGAATCATT GGAACAGGTT GAGGGATGAC GCAGCTAACT TAAAGAGTAA AGGGATAACA 120
 GCTGTATGGA TCCCACCTGC ATGGAAGGGG ACTTCCCAGA ATGATGTAGG TTATGGAGCC 180
 TATGATTTAT ATGATCTTGG AGAGTTTAAC CAGAAGGGGA CGGTTTCGTAC AAAATATGGA 240
 ACACGCAACC AGCTACAGGC TGCGGTGACC TCTTTAAAAA ATAACGGCAT TCAGGTATAT 300
 GGTGATGTCC TCATGAATCA TAAAGGTGGA GCAGATGGTA CGGAAATTGT AAATGCGGTA 360
 GAAGTGAATC GGAGCAACCG AAACCAGGAA ACCTCAGGAG AGTATGCAAT AGAAGCGTGG 420
 ACAAAGTTTG ATTTTCCTGG AAGAGGAAAT AACCATTCCA GCTTTAAGTG GCGCTGGTAT 480
 CATTTTGATG GGACAGATTG GGATCAGTCA CGCCAGCTTC AAAACAAAAT ATATAAATTC 540

AGGGGAACAG GCAAGGCCTG GGA	CTGACTAT	600
CTTATGTATG CAGACGTGGA TATGGATCAC CCAGAAGTAA TACATGAACT TAGAAACTGG		660
GGAGTGTGGT ATACGAATAC ACTGAACCTT GATGGATTTA GAATAGATGC AGTGAAACAT		720
ATAAAATATA GCTTTACGAG AGATTGGCTT ACACATGTGC GTAACACCAC AGGTAAACCA		780
ATGTTTGCGAG TGGCTGAGTT TTGGAAAAAT GACCTTGGTG CAATTGAAAA CTATTTGAAT		840
AAAACAAGTT GGAATCACTC GGTGTTTGAT GTTCCTCTCC ACTATAATTT GTACAATGCA		900
TCTAATAGCG GTGGTTATTA TGATATGAGA AATATTTTAA ATGGTTCTGT GGTGCAAAAA		960
CATCCAACAC ATGCCGTTAC TTTTGTGGAT AACCATGATT CTCAGCCCGG GGAAGCATTG		1020
GAATCCTTTG TTCAACAATG GTTTAAACCA CTTGCATATG CATTGGTTCT GACAAGGGAA		1080
CAAGGTTATC CTTCCGTATT TTATGGGGAT TACTACGGTA TCCCAACCCA TGGTGTTCGG		1140
GCTATGAAAT CTAATAAGTA CCCTCTTCTG CAGGCACGTC AAACCTTTTGC CTATGGTACG		1200
CAGCATGATT ACTTTGATCA TCATGATATT ATCGGTTGGA CAAGAGAGGG AAATAGCTCC		1260
CATCCAAATT CAGGCCTTGC CACCATTATG TCAGATGGTC CAGGTGGTAA CAAATGGATG		1320
TATGTGGGGA AAAATAAAGC GGGACAAGTT TGGAGAGATA TTACCGGAAA TAGGACAGGC		1380
ACCGTCACAA TTAATGCAGA CGGATGGGGT AATTTCTCTG TTAATGGAGG GTCCGTTTCG		1440
GTTTGGGTGA AGCAA		1455

(2) SEQ ID NO: 10 的资料:

(i) 序列特征:

(A) 长度: 1455 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑构型: 线性

(ii) 分子类型: DNA(基因组)

(iii) 生物: 芽孢杆菌属的种

(xi) 序列描述: SEQ ID NO: 10

CATCATAATG GGACAAATGG GACGATGATG CAATACTTTG AATGGCACTT GCCTAATGAT		60
GGGAATCACT GGAATAGATT AAGAGATGAT GCTAGTAATC TAAGAAATAG AGGTATAACC		120
GCTATTTGGA TTCCGCCTGC CTGGAAAGGG ACTTCGCAA ATGATGTGGG GTATGGAGCC		180
TATGATCTTT ATGATTTAGG GGAATTTAAT CAAAAGGGGA CGGTTTCGTAC TAAGTATGGG		240
ACACGTAGTC AATTGGAGTC TGCCATCCAT GCTTTAAAGA ATAATGGCGT TCAAGTTTAT		300
GGGGATGTAG TGATGAACCA TAAAGGAGGA GCTGATGCTA CAGAAAACGT TCTTGCTGTC		360

GAGGTGAATC CAAATAACCG GAATCAAGAA ATATCTGGGG ACTACACAAT TGAGGCTTGG	420
ACTAAGTTTG ATTTTCCAGG GAGGGTAAT ACATACTCAG ACTTTAAATG GCGTTGGTAT	480
CATTTTCGATG GTGTAGATTG GGATCAATCA CGACAATTCC AAAATCGTAT CTACAAATTC	540
CGAGGTGATG GTAAGGCATG GGATTGGGAA GTAGATTCCG AAAATGGAAA TTATGATTAT	600
TTAATGTATG CAGATGTAGA TATGGATCAT CCGGAGGTAG TAAATGAGCT TAGAAGATGG	660
GGAGAATGGT ATACAAATAC ATTAATCTT GATGGATTTA GGATCGATGC GGTGAAGCAT	720
ATTAAATATA GCTTTACACG TGATTGGTTG ACCCATGTAA GAAACGCAAC GGGAAAAGAA	780
ATGTTTGCTG TTGCTGAATT TTGGAAAAAT GATTTAGGTG CCTTGAGAGAA CTATTTAAAT	840
AAAACAACT GGAATCATTG TGTCTTTGAT GTCCCCCTTC ATTATAATCT TTATAACCGG	900
TCAAATAGTG GAGGCAACTA TGACATGGCA AAACCTCTTA ATGGAACGGT TGTTCAAAAAG	960
CATCCAATGC ATGCCGTAAC TTTTGTGGAT AATCACGATT CTCAACCTGG GGAATCATTG	1020
GAATCATTG TACAAGAATG GTTTAAGCCA CTTGCTTATG CGCTTATTTT AACAAGAGAA	1080
CAAGGCTATC CCTCTGTCTT CTATGGTGAC TACTATGGAA TTCCAACACA TAGTGTCCCA	1140
GCAATGAAAG CCAAGATTGA TCCAATCTTA GAGGCGCGTC AAAATTTTGC ATATGGAACA	1200
CAACATGATT ATTTTGACCA TCATAATATA ATCGGATGGA CACGTGAAGG AAATACCACG	1260
CATCCCAATT CAGGACTTGC GACTATCATG TCGGATGGGC CAGGGGGAGA GAAATGGATG	1320
TACGTAGGGC AAAATAAAGC AGGTCAAGTT TGGCATGACA TAACTGGAAA TAAACCAGGA	1380
ACAGTTACGA TCAATGCAGA TGGATGGGCT AATTTTTTCAG TAAATGGAGG ATCTGTTTCC	1440
ATTTGGGTGA AACGA	1455

(2) SEQ ID NO: 11 的资料:

(i) 序列特征:

(A) 长度: 1548 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑构型: 线性

(ii) 分子类型: DNA(基因组)

(iii) 生物: 嗜热脂肪芽孢杆菌

(xi) 序列描述: SEQ ID NO: 11

GCCGCACCGT TTAACGGCAC CATGATGCAG TATTTTGAAT GGTACTTGCC GGATGATGGC	60
ACGTTATGGA CCAAAGTGGC CAATGAAGCC AACAACTTAT CCAGCCTTGG CATCACCGCT	120
CTTTGGCTGC CGCCCCTTA CAAAGGAACA AGCCGCAGCG ACGTAGGGTA CGGAGTATAC	180
GACTTGATG ACCTCGGCGA ATTCAATCAA AAAGGGACCG TCCGCACAAA ATACGGAACA	240

AAAGCTCAAT ATCTTCAAGC CATTCAAGCC GCCCACGCCG CTGGAATGCA AGTGTACGCC	300
GATGTCGTGT TCGACCATAA AGGCGGCGCT GACGGCACGG AATGGGTGGA CGCCGTCGAA	360
GTCAATCCGT CCGACCGCAA CCAAGAAATC TCGGGCACCT ATCAAATCCA AGCATGGACG	420
AAATTTGATT TTCCCGGGCG GGGCAACACC TACTCCAGCT TTAAGTGGCG CTGGTACCAT	480
TTTGACGGCG TTGATTGGGA CGAAAGCCGA AAATTGAGCC GCATTTACAA ATTCCGCGGC	540
ATCGGCAAAG CGTGGGATTG GGAAGTAGAC ACGGAAAACG GAAACTATGA CTACTIONAATG	600
TATGCCGACC TTGATATGGA TCATCCCGAA GTCGTGACCG AGCTGAAAAA CTGGGGGAAA	660
TGGTATGTCA ACACAACGAA CATTGATGGG TTCCGGCTTG ATGCCGTCAA GCATATTAAG	720
TTCAGTTTTT TTCCTGATTG GTTGTCTGAT GTGCGTTCTC AGACTGGCAA GCCGCTATTT	780
ACCGTCGGGG AATATTGGAG CTATGACATC AACAAGTTGC ACAATTACAT TACGAAAACA	840
GACGGAACGA TGTCTTTGTT TGATGCCCGG TTACACAACA AATTTTATAC CGCTTCCAAA	900
TCAGGGGGCG CATTGATAT GCGCACGTTA ATGACCAATA CTCTCATGAA AGATCAACCG	960
ACATTGGCCG TCACCTTCGT TGATAATCAT GACACCGAAC CCGCCAAGC GCTGCAGTCA	1020
TGGGTCGACC CATGGTTCAA ACCGTTGGCT TACGCCTTTA TTCTAACTCG GCAGGAAGGA	1080
TACCCGTGCG TCTTTTATGG TGACTATTAT GGCATTCCAC AATATAACAT TCCTTCGCTG	1140
AAAAGCAAAA TCGATCCGCT CCTCATCGCG CGCAGGGATT ATGCTTACGG AACGCAACAT	1200
GATTATCTTG ATCACTCCGA CATCATCGGG TGGACAAGGG AAGGGGGCAC TGAAAAACCA	1260
GGATCCGGAC TGGCCGCACT GATCACCGAT GGGCCGGGAG GAAGCAAATG GATGTACGTT	1320
GGCAAACAAC ACGCTGGAAA AGTGTCTAT GACCTTACCG GCAACCGGAG TGACACCGTC	1380
ACCATCAACA GTGATGGATG GGGGGAATTC AAAGTCAATG GCGGTTCCGT TTCGGTTTGG	1440
GTTCTAGAAA AAACGACCGT TTCTACCATC GCTCGGCCGA TCACAACCCG ACCGTGGACT	1500
GGTGAATTCG TCCGTTGGAC CGAACCACGG TTGGTGGCAT GGCCTTGA	1548

(2) SEQ ID NO: 12 的资料:

(i) 序列特征:

(A) 长度: 1920 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑构型: 线性

(ii) 分子类型: DNA(基因组)

(iii) 生物: 地衣芽孢杆菌

(xi) 序列描述: SEQ ID NO: 12

CGGAAGATTG GAAGTACAAA AATAAGCAAA AGATTGTCAA TCATGTCATG AGCCATGCGG	60
GAGACGGAAA AATCGTCTTA ATGCACGATA TTTATGCAAC GTTCGCAGAT GCTGCTGAAG	120
AGATTATTAA AAAGCTGAAA GCAAAAGGCT ATCAATTGGT AACTGTATCT CAGCTTGAAG	180
AAGTGAAGAA GCAGAGAGGC TATTGAATAA ATGAGTAGAA GCGCCATATC GCGCCTTTTC	240
TTTTGGAAGA AAATATAGGG AAAATGGTAC TTGTTAAAAA TTCGGAATAT TTATAACAACA	300
TCATATGTTT CACATTGAAA GGGGAGGAGA ATCATGAAAC AACAAAAACG GCTTTACGCC	360
CGATTGCTGA CGCTGTTATT TCGCTCATC TTCTTGCTGC CTCATTCTGC AGCAGCGGCG	420
GCA AAT CTT AAT GGG ACG CTG ATG CAG TAT TTT GAA TGG TAC ATG CCC	468
AAT GAC GGC CAA CAT TGG AGG CGT TTG CAA AAC GAC TCG GCA TAT TTG	516
GCT GAA CAC GGT ATT ACT GCC GTC TGG ATT CCC CCG GCA TAT AAG GGA	564
ACG AGC CAA GCG GAT GTG GGC TAC GGT GCT TAC GAC CTT TAT GAT TTA	612
GGG GAG TTT CAT CAA AAA GGG ACG GTT CGG ACA AAG TAC GGC ACA AAA	660
GGA GAG CTG CAA TCT GCG ATC AAA AGT CTT CAT TCC CGC GAC ATT AAC	708
GTT TAC GGG GAT GTG GTC ATC AAC CAC AAA GGC GGC GCT GAT GCG ACC	756
GAA GAT GTA ACC GCG GTT GAA GTC GAT CCC GCT GAC CGC AAC CGC GTA	804
ATT TCA GGA GAA CAC CTA ATT AAA GCC TGG ACA CAT TTT CAT TTT CCG	852
GGG CGC GGC AGC ACA TAC AGC GAT TTT AAA TGG CAT TGG TAC CAT TTT	900
GAC GGA ACC GAT TGG GAC GAG TCC CGA AAG CTG AAC CGC ATC TAT AAG	948
TTT CAA GGA AAG GCT TGG GAT TGG GAA GTT TCC AAT GAA AAC GGC AAC	996
TAT GAT TAT TTG ATG TAT GCC GAC ATC GAT TAT GAC CAT CCT GAT GTC	1044
GCA GCA GAA ATT AAG AGA TGG GGC ACT TGG TAT GCC AAT GAA CTG CAA	1092
TTG GAC GGT TTC CGT CTT GAT GCT GTC AAA CAC ATT AAA TTT TCT TTT	1140
TTG CGG GAT TGG GTT AAT CAT GTC AGG GAA AAA ACG GGG AAG GAA ATG	1188
TTT ACG GTA GCT GAA TAT TGG CAG AAT GAC TTG GGC GCG CTG GAA AAC	1236
TAT TTG AAC AAA ACA AAT TTT AAT CAT TCA GTG TTT GAC GTG CCG CTT	1284
CAT TAT CAG TTC CAT GCT GCA TCG ACA CAG GGA GGC GGC TAT GAT ATG	1332
AGG AAA TTG CTG AAC GGT ACG GTC GTT TCC AAG CAT CCG TTG AAA TCG	1380
GTT ACA TTT GTC GAT AAC CAT GAT ACA CAG CCG GGG CAA TCG CTT GAG	1428
TCG ACT GTC CAA ACA TGG TTT AAG CCG CTT GCT TAC GCT TTT ATT CTC	1476
ACA AGG GAA TCT GGA TAC CCT CAG GTT TTC TAC GGG GAT ATG TAC GGG	1524
ACG AAA GGA GAC TCC CAG CGC GAA ATT CCT GCC TTG AAA CAC AAA ATT	1572

GAA CCG ATC TTA AAA GCG AGA AAA CAG TAT GCG TAC GGA GCA CAG CAT	1620
GAT TAT TTC GAC CAC CAT GAC ATT GTC GGC TGG ACA AGG GAA GGC GAC	1668
AGC TCG GTT GCA AAT TCA GGT TTG GCG GCA TTA ATA ACA GAC GGA CCC	1716
GGT GGG GCA AAG CGA ATG TAT GTC GGC CGG CAA AAC GCC GGT GAG ACA	1764
TGG CAT GAC ATT ACC GGA AAC CGT TCG GAG CCG GTT GTC ATC AAT TCG	1812
GAA GGC TGG GGA GAG TTT CAC GTA AAC GGC GGG TCG GTT TCA ATT TAT	1860
GTT CAA AGA TAG AAGAGCAGAG AGGACGGATT TCCTGAAGGA AATCCGTTTT	1912
TTTATTTTT	1920

(2) SEQ ID NO: 13 的资料:

(i) 序列特征:

(A) 长度: 1455 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑构型: 线性

(ii) 分子类型: DNA(基因组)

(iii) 生物: 芽孢杆菌属的种

(xi) 序列描述: SEQ ID NO: 13

CATCATAATG GAACAAATGG TACTATGATG CAATATTTTCG AATGGTATTT GCCAAATGAC	60
GGGAATCATT GGAACAGGTT GAGGGATGAC GCAGCTAACT TAAAGAGTAA AGGGATAACA	120
GCTGTATGGA TCCCACCTGC ATGGAAGGGG ACTTCCCAGA ATGATGTAGG TTATGGAGCC	180
TATGATTTAT ATGATCTTGG AGAGTTTAAC CAGAAGGGGA CGGTTTCGTAC AAAATATGGA	240
ACACGCAACC AGCTACAGGC TGCGGTGACC TCTTTAAAAA ATAACGGCAT TCAGGTATAT	300
GGTGATGTCTG TCATGAATCA TAAAGGTGGA GCAGATGGTA CGGAAATTGT AAATGCGGTA	360
GAAGTGAATC GGAGCAACCG AAACCAGGAA ACCTCAGGAG AGTATGCAAT AGAAGCGTGG	420
ACAAAAGTTG ATTTTCCTGG AAGAGGAAAT AACCATTTCCA GCTTTAAGTG GCGCTGGTAT	480
CATTTTGATG GGACAGATTG GGATCAGTCA CGCCAGCTTC AAAACAAAAT ATATAAATTC	540
AGGGGAACAG GCAAGGCCTG GGAAGTGGAA GTCGATACAG AGAATGGCAA CTATGACTAT	600
CTTATGTATG CAGACGTGGA TATGGATCAC CCAGAAGTAA TACATGAACT TAGAACTGG	660
GGAGTGTGGT ATACGAATAC ACTGAACCTT GATGGATTTA GAATAGATGC AGTGAAACAT	720
ATAAAATATA GCTTTACGAG AGATTGGCTT ACACATGTGC GTAACACCAC AGGTAAACCA	780
ATGTTTGCAG TGGCTGAGTT TTGGAAAAAT GACCTTGCTG CAATTGAAAA CTATTTGAAT	840
AAAACAAGTT GGAATCACTC GGTGTTTGGT GTTCCTCTCC ACTATAATTT GTACAATGCA	900
TCTAATAGCG GTGGTTATTA TGATATGAGA AATATTTTAA ATGGTTCTGT GGTGCAAAAA	960

CATCCAACAC ATGCCGTTAC TTTTGTGAT AACCATGATT CTCAGCCCGG GGAAGCATTG	1020
GAATCCTTTG TTCAACAATG GTTTAAACCA CTTGCATATG CATTGGTTCT GACAAGGGAA	1080
CAAGGTTATC CTTCCGTATT TTATGGGGAT TACTACGGTA TCCCAACCCA TGGTGTTCGG	1140
GCTATGAAAT CTAAAATAGA CCCTCTTCTG CAGGCACGTC AAACCTTTGC CTATGGTACG	1200
CAGCATGATT ACTTTGATCA TCATGATATT ATCGGTTGGA CAAGAGAGGG AAATAGCTCC	1260
CATCCAAATT CAGGCCTTGC CACCATTATG TCAGATGGTC CAGGTGGTAA CAAATGGATG	1320
TATGTGGGGA AAAATAAAGC GGGACAAGTT TGGAGAGATA TTACCGGAAA TAGGACAGGC	1380
ACCGTCACAA TTAATGCAGA CGGATGGGGT AATTTCTCTG TTAATGGAGG GTCCGTTTCG	1440
GTTTGGGTGA AGCAA	1455

(2) SEQ ID NO: 14 的资料:

(i) 序列特征:

(A) 长度: 1455 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑构型: 线性

(ii) 分子类型: DNA(基因组)

(iii) 生物: 芽孢杆菌属的种

(xi) 序列描述: SEQ ID NO: 14

CATCATAATG GGACAAAATGG GACGATGATG CAATACTTTG AATGGCACTT GCCTAATGAT	60
GGGAATCACT GGAATAGATT AAGAGATGAT GCTAGTAATC TAAGAAATAG AGGTATAACC	120
GCTATTTGGA TTCCGCCCTGC CTGGAAAGGG ACTTCGCAA ATGATGTGGG GTATGGAGCC	180
TATGATCTTT ATGATTTAGG GGAATTTAAT CAAAAGGGGA CGGTTTCGTAC TAAGTATGGG	240
ACACGTAGTC AATTGGAGTC TGCCATCCAT GCTTTAAAGA ATAATGGCGT TCAAGTTTAT	300
GGGGATGTAG TGATGAACCA TAAAGGAGGA GCTGATGCTA CAGAAAACGT TCTTGCTGTC	360
GAGGTGAATC CAAATAACCG GAATCAAGAA ATATCTGGGG ACTACACAAT TGAGGCTTGG	420
ACTAAGTTTG ATTTTCCAGG GAGGGGTAAT ACATACTCAG ACTTTAAATG GCGTTGGTAT	480
CATTCGATG GTGTAGATTG GGATCAATCA CGACAATTCC AAAATCGTAT CTACAAATTC	540
CGAGGTGATG GTAAGGCATG GGATTGGGAA GTAGATTCGG AAAATGGAAA TTATGATTAT	600
TTAATGTATG CAGATGTAGA TATGGATCAT CCGGAGGTAG TAAATGAGCT TAGAAGATGG	660
GGAGAATGGT ATACAAATAC ATTAATCTT GATGGATTTA GGATCGATGC GGTGAAGCAT	720
ATTAATATA GCTTTACACG TGATTGGTTG ACCCATGTAA GAAACGCAAC GGGAAAAGAA	780
ATGTTTGCTG TTGCTGAATT TTGGAAAAAT GATTTAGGTG CCTTGGAGAA CTATTTAAAT	840

```

AAAACAACT GGAATCATTG TGTCTTTGAT GTCCCCCTTC ATTATAATCT TTATAACGCG      900
TCAAATAGTG GAGGCAACTA TGACATGGCA AACTTCTTA ATGGAACGGT TGTTCAAAG      960
CATCCAATGC ATGCCGTAAC TTTTGTGGAT AATCACGATT CTCAACCTGG GGAATCATT      1020
GAATCATTTG TACAAGAATG GTTTAAGCCA CTTGCTTATG CGCTTATTTT AACAAGAGAA      1080
CAAGGCTATC CCTCTGTCTT CTATGGTGAC TACTATGGAA TTCCAACACA TAGTGTCCCA      1140
GCAATGAAAG CCAAGATTGA TCCAATCTTA GAGGCGCGTC AAAATTTTGC ATATGGAACA      1200
CAACATGATT ATTTTGACCA TCATAATATA ATCGGATGGA CACGTGAAGG AAATACCACG      1260
CATCCCAATT CAGGACTTGC GACTATCATG TCGGATGGGC CAGGGGGAGA GAAATGGATG      1320
TACGTAGGGC AAAATAAAGC AGGTCAAGTT TGGCATGACA TAACTGGAAA TAAACCAGGA      1380
ACAGTTACGA TCAATGCAGA TGGATGGGCT AATTTTTTCAG TAAATGGAGG ATCTGTTTCC      1440
ATTTGGGTGA AACGA      1455

```

(2) SEQ ID NO: 15 的资料:

(i) 序列特征:

- (A) 长度: 74 个碱基对
- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑构型: 线性

(ii) 分子类型: 其它核酸

(ix) 特征:

- (A) 名称/关键词: misc-特征
- (B) 其它信息: /desc = "RSERI"

(ix) 特征:

- (A) 名称/关键词: misc-特征
- (B) 位置: 21-62
- (D) 其它信息: /Note = 1: 97% A, 1% T, 1% C, 1% G
2: 97% T, 1% A, 1% C, 1% G
3: 97% C, 1% A, 1% T, 1% G
4: 97% G, 1% A, 1% T, 1% C

(xi) 序列描述: SEQ ID NO: 15

```

GCGTTTTGCC GGCCGACATA 3122343222 4333313344
4233423242 2122112433 43CAAACCTG AATT

```

74

(2) SEQ ID NO: 16 的资料:

(i) 序列特征:

- (A) 长度: 122 个碱基对
- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑构型: 线性

(ii) 分子类型: 其它核酸

(ix) 特征:

- (A) 名称/关键词: misc-特征
- (B) 其它信息: /desc = "RSERII"

(ix) 特征:

- (A) 名称/关键词: misc-特征
- (B) 位置: 63-104
- (D) 其它信息: /Note = 1: 97% A, 1% T, 1% C, 1% G
2: 97% T, 1% A, 1% C, 1% G
3: 97% C, 1% A, 1% T, 1% G
4: 97% G, 1% A, 1% T, 1% C

(xi) 序列描述: SEQ ID NO: 16

```
GCGTTTTGCC GGCCGACATA CATTGCTTT GCCCCACCGG GTCCGTCTGT
TATTAATGCC GC31113324 1122243113 3414324234 3322333224
2331GCCGAC AATGTCATGG TG
```

122

(2) SEQ ID NO: 17 的资料:

(i) 序列特征:

- (A) 长度: 78 个碱基对
- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑构型: 线性

(ii) 分子类型: 其它核酸

(ix) 特征:

- (A) 名称/关键词: misc-特征

(B) 其它信息: /desc = "RSERIII"

(ix) 特征:

(A) 名称/关键词: misc-特征

(B) 位置: 19-60

(D) 其它信息: /Note = 1: 97% A, 1% T, 1% C, 1% G

2: 97% T, 1% A, 1% C, 1% G

3: 97% C, 1% A, 1% T, 1% G

4: 97% G, 1% A, 1% T, 1% C

(xi) 序列描述: SEQ ID NO: 17

GTCGCCTTCC CTTGTCCA43 3413112423 1244244234 1112112312
4324243233 GTACGCATAC TGTTCCT

78

(2) SEQ ID NO: 18 的资料:

(i) 序列特征:

(A) 长度: 20 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑构型: 线性

(ii) 分子类型: 其它核酸

(ix) 特征:

(A) 名称/关键词: misc-特征

(B) 其它信息: /desc = "FSERIII"

(xi) 序列描述: SEQ ID NO: 18

TGGACAAGGG AAGGCGACAG

20

(2) SEQ ID NO: 19 的资料:

(i) 序列特征:

(A) 长度: 81 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑构型: 线性

(ii) 分子类型: 其它核酸

(ix) 特征:

(A) 名称/关键词: misc-特征

(B) 其它信息: /desc = "RSERV"

(ix) 特征:

(A) 名称/关键词: misc-特征

(B) 位置: 19-60

(D) 其它信息: /Note = 1: 97% A, 1% T, 1% C, 1% G

2: 97% T, 1% A, 1% C, 1% G

3: 97% C, 1% A, 1% T, 1% G

4: 97% G, 1% A, 1% T, 1% C

(xi) 序列描述: SEQ ID NO: 19

TAAGATCGGT TCAATTTT42 4222311443 1441122234 3432444142
3233222342 CCCGTACATA TCCCCGTAGA A

(2) SEQ ID NO: 20 的资料:

(i) 序列特征:

(A) 长度: 18 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑构型: 线性

(ii) 分子类型: 其它核酸

(ix) 特征:

(A) 名称/关键词: misc-特征

(B) 其它信息: /desc = "FSERV"

(xi) 序列描述: SEQ ID NO: 20

AAAATTGAAC CGATCTTA

18

(2) SEQ ID NO: 21 的资料:

(i) 序列特征:

(A) 长度: 107 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑构型: 线性

(ii) 分子类型: 其它核酸

(ix) 特征:

(A) 名称/关键词: misc-特征

(B) 其它信息: /desc = "FSERVII"

(ix) 特征:

(A) 名称/关键词: misc-特征

(B) 位置: 54-95

(D) 其它信息: /Note = 1: 97% A, 1% T, 1% C, 1% G

2: 97% T, 1% A, 1% C, 1% G

3: 97% C, 1% A, 1% T, 1% G

4: 97% G, 1% A, 1% T, 1% C

(xi) 序列描述: SEQ ID NO: 21

TTCCATGCTG CATCGACACA GGGAGGCGGC TATGATATGA GGAAATTGCT
GAA3442134 4234222331 1431233422 4111234422 13122TGTCG
ATAACCA

108

(2) SEQ ID NO: 22 的资料:

(i) 序列特征:

(A) 长度: 18 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑构型: 线性

(ii) 分子类型: 其它核酸

(ix) 特征:

(A) 名称/关键词: misc-特征

(B) 其它信息: /desc = "RSERVII"

(xi) 序列描述: SEQ ID NO: 22

TGTCGATGCA GCATGGAA

18

(2) SEQ ID NO: 23 的资料:

(i) 序列特征:

(A) 长度: 80 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑构型: 线性

(ii) 分子类型: 其它核酸

(ix) 特征:

(A) 名称/关键词: misc-特征

(B) 其它信息: /desc = "FSERIX"

(ix) 特征:

(A) 名称/关键词: misc-特征

(B) 位置: 21-62

(D) 其它信息: /Note = 1: 97% A, 1% T, 1% C, 1% G

2: 97% T, 1% A, 1% C, 1% G

3: 97% C, 1% A, 1% T, 1% G

4: 97% G, 1% A, 1% T, 1% C

(xi) 序列描述: SEQ ID NO: 23

```

GTCCAAACAT GGTTTAAGCC 4322432213 4322221223 2313114441
1232441213 33TCAGGTTT TCTACGGGGA 80

```

(2) SEQ ID NO: 24 的资料:

(i) 序列特征:

(A) 长度: 20 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑构型: 线性

(ii) 分子类型: 其它核酸

(ix) 特征:

(A) 名称/关键词: misc-特征

(B) 其它信息: /desc = "RSERIX"

(xi) 序列描述: SEQ ID NO: 25

GGCTTAAACC ATGTTTGGAC

20

(2) SEQ ID NO: 26 的资料:

(i) 序列特征:

(A) 长度: 24 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑构型: 线性

(ii) 分子类型: 其它核酸

(ix) 特征:

(A) 名称/关键词: misc-特征

(B) 其它信息: /desc = “引物 1B”

(xi) 序列描述: SEQ ID NO: 26

CGATTGCTGA CGCTGTTATT TGCG

24

(2) SEQ ID NO: 27 的资料:

(i) 序列特征:

(A) 长度: 25 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑构型: 线性

(ii) 分子类型: 其它核酸

(ix) 特征:

(A) 名称/关键词: misc-特征

(B) 其它信息: /desc = “引物 # 63”

(xi) 序列描述: SEQ ID NO: 27

CTATCTTTGA ACATAAATTG AAACC

25

(2) SEQ ID NO: 28 的资料:

(i) 序列特征:

(A) 长度: 20 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑构型: 线性

(ii) 分子类型: 其它核酸

(ix) 特征:

(A) 名称/关键词: misc-特征

(B) 其它信息: /desc = “正向引物 1”

(xi) 序列描述: SEQ ID NO: 28

gacctgcagt caggcaacta

20

(2) SEQ ID NO: 29 的资料:

(i) 序列特征:

(A) 长度: 20 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑构型: 线性

(ii) 分子类型: 其它核酸

(ix) 特征:

(A) 名称/关键词: misc-特征

(B) 其它信息: /desc = “反向引物 1”

(xi) 序列描述: SEQ ID NO: 29

tagagtcgac ctgcagggcat

20

(2) SEQ ID NO: 30 的资料:

(i) 序列特征:

(A) 长度: 20 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑构型: 线性

(ii) 分子类型: 其它核酸

(ix) 特征:

(A) 名称/关键词: misc-特征

(B) 其它信息: /desc = “正向引物 2”

(xi) 序列描述: SEQ ID NO: 30

gacctgcagt caggcaacta

20

(2) SEQ ID NO: 31 的资料:

(i) 序列特征:

(A) 长度: 25 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑构型: 线性

(ii) 分子类型: 其它核酸

(ix) 特征:

(A) 名称/关键词: misc-特征

(B) 其它信息: /desc = “反向引物 2”

(xi) 序列描述: SEQ ID NO: 31

tagagtcgac ctgcaggcat

20

(2) SEQ ID NO: 32 的资料:

(i) 序列特征:

(A) 长度: 2084 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑构型: 线性

(ii) 分子类型: DNA(基因组)

(iii) 生物: 解淀粉芽孢杆菌

(ix) 特征:

(A) 名称/关键词: CDS

(B) 位置: 343...1794

(xi) 序列描述: SEQ ID NO: 32

GCCCCGCACA TACGAAAAGA CTGGCTGAAA ACATTGAGCC TTTGATGACT GATGATTTGG	60
CTGAAGAAGT GGATCGATTG TTTGAGAAAA GAAGAAGACC ATAAAAATAC CTTGTCTGTG	120
ATCAGACAGG GTATTTTTTA TGCTGTCCAG ACTGTCCGCT GTGTAAAAAT AAGGAATAAA	180
GGGGGGTTGT TATTATTTTA CTGATATGTA AAATATAATT TGTATAAGAA AATGAGAGGG	240
AGAGGAAACA TGATTCAAAA ACGAAAGCGG ACAGTTTCGT TCAGACTTGT GCTTATGTGC	300
ACGCTGTTAT TTGTCAGTTT GCCGATTACA AAAACATCAG CC GTA AAT GGC ACG	354

CTG ATG CAG TAT TTT GAA TGG TAT ACG CCG AAC GAC GGC CAG CAT TGG	402
AAA CGA TTG CAG AAT GAT GCG GAA CAT TTA TCG GAT ATC GGA ATC ACT	450
GCC GTC TGG ATT CCT CCC GCA TAC AAA GGA TTG AGC CAA TCC GAT AAC	498
GGA TAC GGA CCT TAT GAT TTG TAT GAT TTA GGA GAA TTC CAG CAA AAA	546
GGG ACG GTC AGA ACG AAA TAC GGC ACA AAA TCA GAG CTT CAA GAT GCG	594
ATC GGC TCA CTG CAT TCC CGG AAC GTC CAA GTA TAC GGA GAT GTG GTT	642
TTG AAT CAT AAG GCT GGT GCT GAT GCA ACA GAA GAT GTA ACT GCC GTC	690
GAA GTC AAT CCG GCC AAT AGA AAT CAG GAA ACT TCG GAG GAA TAT CAA	738
ATC AAA GCG TGG ACG GAT TTT CGT TTT CCG GGC CGT GGA AAC ACG TAC	786
AGT GAT TTT AAA TGG CAT TGG TAT CAT TTC GAC GGA GCG GAC TGG GAT	834
GAA TCC CGG AAG ATC AGC CGC ATC TTT AAG TTT CGT GGG GAA GGA AAA	882
GCG TGG GAT TGG GAA GTA TCA AGT GAA AAC GGC AAC TAT GAC TAT TTA	930
ATG TAT GCT GAT GTT GAC TAC GAC CAC CCT GAT GTC GTG GCA GAG ACA	978
AAA AAA TGG GGT ATC TGG TAT GCG AAT GAA CTG TCA TTA GAC GGC TTC	1026
CGT ATT GAT GCC GCC AAA CAT ATT AAA TTT TCA TTT CTG CGT GAT TGG	1074
GTT CAG GCG GTC AGA CAG GCG ACG GGA AAA GAA ATG TTT ACG GTT GCG	1122
GAG TAT TGG CAG AAT AAT GCC GGG AAA CTC GAA AAC TAC TTG AAT AAA	1170
ACA AGC TTT AAT CAA TCC GTG TTT GAT GTT CCG CTT CAT TTC AAT TTA	1218
CAG GCG GCT TCC TCA CAA GGA GGC GGA TAT GAT ATG AGG CGT TTG CTG	1266
GAC GGT ACC GTT GTG TCC AGG CAT CCG GAA AAG GCG GTT ACA TTT GTT	1314
GAA AAT CAT GAC ACA CAG CCG GGA CAG TCA TTG GAA TCG ACA GTC CAA	1362
ACT TGG TTT AAA CCG CTT GCA TAC GCC TTT ATT TTG ACA AGA GAA TCC	1410
GGT TAT CCT CAG GTG TTC TAT GGG GAT ATG TAC GGG ACA AAA GGG ACA	1458
TCG CCA AAG GAA ATT CCC TCA CTG AAA GAT AAT ATA GAG CCG ATT TTA	1506
AAA GCG CGT AAG GAG TAC GCA TAC GGG CCC CAG CAC GAT TAT ATT GAC	1554
CAC CCG GAT GTG ATC GGA TGG ACG AGG GAA GGT GAC AGC TCC GCC GCC	1602
AAA TCA GGT TTG GCC GCT TTA ATC ACG GAC GGA CCC GGC GGA TCA AAG	1650
CGG ATG TAT GCC GGC CTG AAA AAT GCC GGC GAG ACA TGG TAT GAC ATA	1698
ACG GGC AAC CGT TCA GAT ACT GTA AAA ATC GGA TCT GAC GGC TGG GGA	1746
GAG TTT CAT GTA AAC GAT GGG TCC GTC TCC ATT TAT GTT CAG AAA TAA	1794
GGTAATAAAA AAACACCTCC AAGCTGAGTG CGGGTATCAG CTTGGAGGTG CGTTTATTTT	1854
TTCAGCCGTA TGACAAGGTC GGCATCAGGT GTGACAAATA CGGTATGCTG GCTGTCATAG	1914
GTGACAAATC CGGGTTTTGC GCCGTTTGGC TTTTTCACAT GTCTGATTTT TGTATAATCA	1974
ACAGGCACGG AGCCGGAATC TTTCCGCTTG GAAAAATAAG CGGCGATCGT AGCTGCTTCC	2034
AATATGGATT GTTCATCGGG ATCGCTGCTT TTAATCACAA CGTGGGATCC	2084

		1				50
1		HHNGTNGTMM	QYFEWHL PND	GNHWNRLRDD	ASNLRNRGIT	AIWIPPAWKG
5	2	. .NGTNGTMM	QYFEWYL PND	GNHWNRLRSD	ASNLKDKGIS	AVWIPPAWKG
	3	HHNGTNGTMM	QYFEWYL PND	GNHWNRLRDD	AANLKSKGIT	AVWIPPAWKG
	4VNGTLM	QYFEWYTPND	GQHWKRLQND	AEHLSDIGIT	AVWIPPAYKG
	5	. .ANLNGTLM	QYFEWYMPND	GQHWRRLQND	SAYLAEHGIT	AVWIPPAYKG
	6	.AAPFNGTMM	QYFEWYLPDD	GTLWTKVANE	ANNLSSLGIT	ALWLPPAYKG
10						
		51				100
	1	TSQNDVGYGA	YDLYDLGEFN	QKGTVRTKYG	TRSQLESaih	ALKNNGVQVY
	2	ASQNDVGYGA	YDLYDLGEFN	QKGTIRT KYG	TRNQLQAAVN	ALKSNGIQVY
	3	TSQNDVGYGA	YDLYDLGEFN	QKGTVRTKYG	TRNQLQAAVT	SLKNNGIQVY
15	4	LSQSDNGYGP	YDLYDLGEFQ	QKGTVRTKYG	TKSELQDAIG	SLHSRNVQVY
	5	TSQADVGYGA	YDLYDLGEFH	QKGTVRTKYG	TKGELQSAIK	SLHSRDIN VY
	6	TSRSDVGYGV	YDLYDLGEFN	QKGTVRTKYG	TKAQYLQAIQ	AAHAAGMQVY
		101				150
20	1	GDVVMNHKGG	ADATENVLAV	EVNPNNRNQE	ISGDYTI EAW	TKFDFPGRGN
	2	GDVVMNHKGG	ADATEMVR AV	EVNPNNRNQE	VSGEYTI EAW	TKFDFPGRGN
	3	GDVVMNHKGG	ADGTEI VNAV	EVNRSNRNQE	TSGEYAIEAW	TKFDFPGRGN
	4	GDVVLN HKAG	ADATEDVTAV	EVNPNANRNQE	TSEYQIKAW	TDFRFPGRGN
	5	GDVVIN HKGG	ADATEDVTAV	EVDPADRN RV	ISGEHLI KAW	THFHFPGRGS
25	6	ADV VFDHKGG	ADGTEWVDAV	EVNPSDRNQE	ISGTYQIQAW	TKFDFPGRGN
		151				200
	1	TYSDFKWRWY	HFDGVDWDQS	RQFQNRIYKF	RGDGKAWDWE	VDSSENGNYDY
	2	THSNFKWRWY	HFDGVDWDQS	RKLNNRIYKF	RGDGKAWDWE	VDTENGN YDY
30	3	NHSSF KWRWY	HFDGTDWDQS	RQLQNKIYKF	RGTGKAWDWE	VDTENGN YDY
	4	TYSDFKWHWY	HFDGADWDES	RKI .SRIYKF	RGEGKAWDWE	VSSSENGNYDY
	5	TYSDFKWHWY	HFDGTDWDES	RKL .NRIYKF	. .QGKAWDWE	VSNSENGNYDY
	6	TYSSF KWRWY	HFDGVDWDES	RKL .SRIYKF	RGIGKAWDWE	VDTENGN YDY

图 1

5	201					250
1	LMYADVDMDH	PEVVNELRRW	GEWYTNLNL	DGFRIDAVKH	IKYSFTRDWL	
2	LMYADIDMDH	PEVVNELRNW	GVWYTNLGL	DGFRIDAVKH	IKYSFTRDWS	
3	LMYADVDMDH	PEVIHELNRW	GVWYTNLNL	DGFRIDAVKH	IKYSFTRDWL	
4	LMYADVVDYDH	PDVVAETKKW	GIWYANESL	DGFRIDAANK	IKFSFLRDWV	
10	5	LMYADIDYDH	PDVAAEIKRW	GTWYANELQL	DGFRLDVAVKH	IKFSFLRDWV
6	LMYADLDMDH	PEVVTELKNW	GKWYVNTTNI	DGFRLDVAVKH	IKFSFFPDWL	
	251					300
1	THVRNATGKE	MFAVAEFWKN	DLGALENYLN	KTNWNHVSVD	VPLHYNLYNA	
15	2	IHVRSATGKN	MFAVAEFWKN	DLGAIENYLN	KTNWNHVSVD	VPLHYNLYNA
3	THVRNTTGKP	MFAVAEFWKN	DLGAIENYLN	KTSWNHSAFD	VPLHYNLYNA	
4	QAVRQATGKE	MFTVAEYWQN	NAGKLENYLN	KTSFNQSVFD	VPLHFNLQAA	
5	NHVREKTGKE	MFTVAEYWQN	DLGALENYLN	KTNFNHSVFD	VPLHYQFHAA	
6	SYVRSQTGKP	LFTVGEYWSY	DINKLHNYIT	KTDGTMSLFD	APLHNKFYTA	
20	301					350
1	SNSGGNYDMA	KLLNGTVVQK	HPMHAVTFVD	NHDSQPGEAL	ESFVQEWFKP	
2	SKSGGNYDMR	QIFNGTVVQR	HPMHAVTFVD	NHDSQPGEAL	ESFVQEWFKP	
3	SNSGGYDMR	NILNGSVVQK	HPTHAVTFVD	NHDSQPGEAL	ESFVQWFKP	
25	4	SSQGGGYDMR	RLLDGTVVS	HPEKAVTFVE	NHDTQPGQSL	ESTVQWFKP
5	STQGGGYDMR	KLLNGTVVSK	HPLKSVTFVD	NHDTQPGQSL	ESTVQWFKP	
6	SKSGGAFDMR	TLMTNTLMKD	QPTLAVTFVD	NHDTQPGQAL	QSWVDPWFKP	
	351					400
30	1	LAYALILTRE	QGYPSVFGD	YYGIPTHS..	.VPAMKAKID	PILEARQNFA
2	LAYALTLTRE	QGYPSVFGD	YYGIPTHG..	.VPAMKSKID	PILEARQKYA	
3	LAYALVLTRE	QGYPSVFGD	YYGIPTHG..	.VPAMKSKID	PLLQARQTF	
4	LAYAFILTRE	SGYPQVFGD	MYGTKGTSK	EIPSLKDNIE	PILKARKEYA	
5	LAYAFILTRE	SGYPQVFGD	MYGTKGDSQR	EIPALKHKIE	PILKARKQYA	
35	6	LAYAFILTRQ	EGYPCVFGD	YYGIPQYN..	.IPSLKSKID	PLLIARRDYA
	401					450
1	YGTQHDYFDH	HNIIGWTREG	NTHPNSGLA	TIMSDGPGGE	KWMYVGQNK	
2	YGRQN.....	
40	3	YGTQHDYFDH	HDIIGWTREG	NSSHNSGLA	TIMSDGPGGN	KWMYVGKNA
4	YGPQHDYIDH	PDVIGWTREG	DSSAAKSGLA	ALITDGPGG	KRMYAGLNA	
5	YGAQHDYFDH	HDIVGTREG	DSSVANSGLA	ALITDGPGG	KRMYVGRQNA	
6	YGTQHDYLDH	SDIIGWTREG	GTEKPGSGLA	ALITDGPGG	KWMYVGKQHA	

图 1 (续)

		451							500
5	1	GQVWHDITGN	KPGTVTINAD	GWANFSVNGG	SVSIWVKR..			
	2			
	3	GQVWRDITGN	RTGTVTINAD	GWGNFSVNGG	SVSVWVKQ..			
	4	GETWYDITGN	RSDTVKIGSD	GWGEFHVNDG	SVSIYVQ...			
	5	GETWHDITGN	RSEPVVINSE	GWGEFHVNGG	SVSIYVQR..			
10	6	GKVFYDLTGN	RSDTV TINSD	GWGEFKVNGG	SVSVWVPRKT	TVSTIARPI			
		501		519					
	1						
	2						
15	3						
	4						
	5						
	6	TRPWTGEFVR	WTEPRLVAW						

图 1 (续)

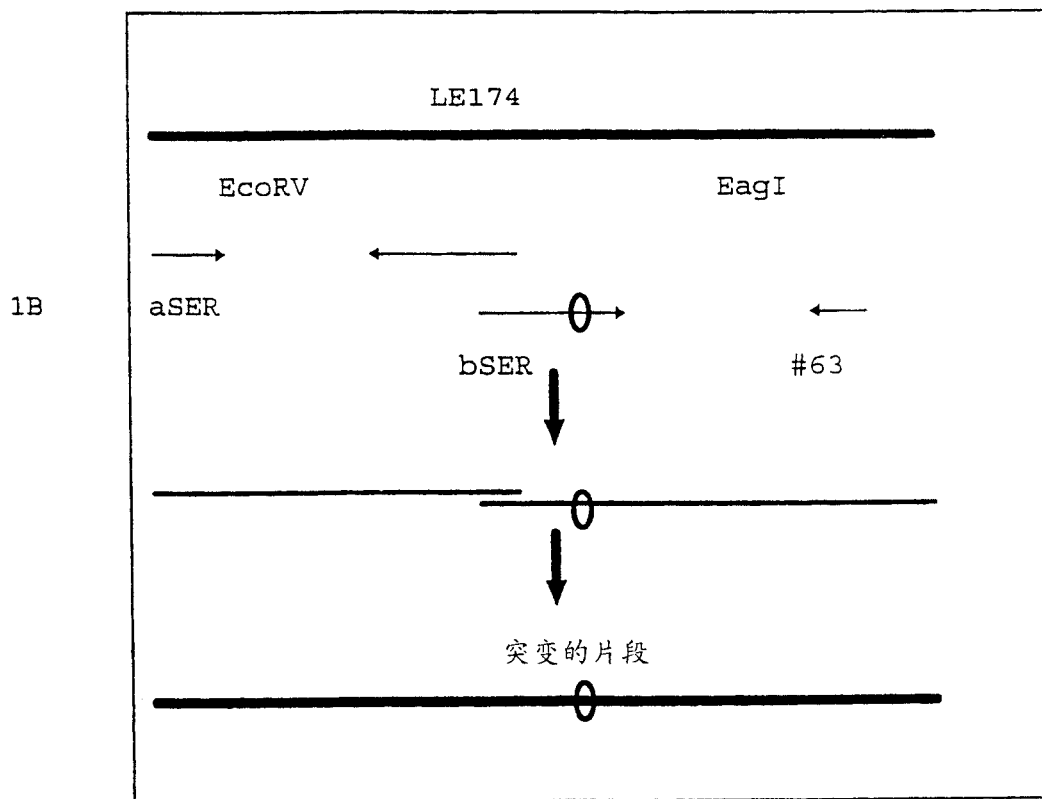


图 2