[19] 中华人民共和国国家知识产权局

[12] 发明专利说明书

专利号 ZL 03824541.8

[51] Int. Cl.
C12N 15/56 (2006.01)
C12R 1/01 (2006.01)
C12R 1/645 (2006.01)
C12N 9/34 (2006.01)
C12N 9/44 (2006.01)
C12P 7/20 (2006.01)

[45] 授权公告日 2009 年 8 月 26 日

[11] 授权公告号 CN 100532557C

[51] Int. Cl. (续)
C07H 21/04 (2006.01)
[22] 申请日 2003.8.25 [21] 申请号 03824541.8
[30] 优先权
[32] 2002. 8.23 [33] US [31] 60/405,896
[86] 国际申请 PCT/US2003/026760 2003.8.25
[87] 国际公布 WO2004/018645 英 2004.3.4
[85] 进入国家阶段日期 2005.4.22
[73] 专利权人 纳幕尔杜邦公司
地址 美国特拉华州
[72] 发明人 P·G·凯米
[56] 参考文献
EP1227152A 2002.7.31
我国淀粉糖工业现状及其发展趋势. 粮食与油脂. 粮食与油脂，第 3 期. 1997

审查员 滕 蕾
[74] 专利代理机构 中国专利代理(香港)有限公司
代理人 李 波 王景朝

权利要求书 3 页 说明书 103 页 附图 4 页

[54] 发明名称

淀粉制品在利用发酵的生物生产中的应用

[57] 摘要

本发明涉及在发酵工艺中使用低纯化的淀粉的方法。 一个实例是，具有外源性胞外异淀粉酶活性的重组大肠杆菌能在发酵中利用含有(1，6)键的小寡聚体(包括但不限于异麦芽糖和潘糖)来生产有用的产物。 通过降低底物碳水化合物的成本，本发明可以用于大规模的工业生物发酵。

1. 一种分离的核酸分子，其编码能水解α(1,6)-连接的葡萄糖寡糖的多肽，该核酸分子选自：

(a)分离的核酸分子，其编码SEQ ID NO:6的氨基酸序列;和

(b)核酸分子，其与(a)完全互补。

2. 根据权利要求1的分离的核酸分子，其由SEQ ID NO:5组成。

3. 权利要求1的分离的核酸分子编码的多肽。

4. 根据权利要求3的多肽，其由SEQ ID NO:6组成。

5. 一种分离的核酸分子，其编码能水解α(1,6)-连接的葡萄糖寡糖的多肽，该核酸分子选自：

(a)分离的核酸分子，其编码由信号肽和SEQ ID NO:6所示的能水解α(1,6)-连接的葡萄糖寡糖的多肽组成的嵌合蛋白，其中信号肽可操作地连接到SEQ ID NO:6;和

(b)核酸分子，其与(a)完全互补。

6. 根据权利要求5的分离的核酸分子,其中所述信号肽是SEQ ID NO:24或SEQ ID NO: 25。

7. 根据权利要求5的编码能水解α(1,6)-连接的葡萄糖寡糖的多肽的分离的核酸分子，所述分离的核酸分子由SEQ ID NO: 5所述序列组成。

8. 嵌合基因，其含有可操作地连接到合适的调控序列的权利要求1或5的分离的核酸分子。

9. 权利要求8的嵌合基因，其中所述合适的调控序列选自*CYC1, HIS3, GAL1, GAL10, ADH1, PGK, PHO5, GAPDH, ADC1, TRP1, URA3, LEU2, ENO, TPI, AOX1, lac, ara, tet, trp,* IP_L, IP_R, *T7, tac, trc, apr, npr, nos*和*GI*。

10. 载体，其含有权利要求8的嵌合基因。

11. 转化的宿主细胞，其含有权利要求8的嵌合基因。

12. 根据权利要求11的转化的宿主细胞，其中所述嵌合基因整合在染色体中或是质粒-携带的。

13. 根据权利要求11的转化的宿主细胞，其中所述宿主细胞选自细菌、酵母和丝状真菌。

14. 根据权利要求13的转化的宿主细胞，其中所述转化的宿主细

胞选自：曲霉属、木霉菌属、糖酵母属、毕赤酵母属、假丝酵母属、汉逊酵母属、红球菌属、不动杆菌属、节杆菌属、短杆菌属、食酸菌属、芽孢杆菌属、链霉菌属、埃希氏菌属、沙门氏菌属、假单胞菌属和棒杆菌属。

15. 根据权利要求 13 的转化的宿主细胞，其中所述转化的宿主细胞是大肠杆菌。

16. 一种在重组宿主细胞中生产靶分子的方法，其包括：

(a)在存在极限糊精、在合适的条件下接触转化的宿主细胞，所述宿主细胞包含：

(i)分离的核酸分子，其编码 SEQ ID NO:6 的氨基酸序列；或

(ii)核酸分子，其与(i)完全互补；和

(iii)至少一种选自下组的嵌合基因：甘油脱水酶基因、二醇脱水酶基因、1,3-丙二醇氧化还原酶基因、NAD^+-连接的甘油脱氢酶基因、$NADP^+$-连接的甘油脱氢酶基因、二羟基丙酮激酶基因、果糖-1,6-二磷酸途径基因、Embden-Meyerhof Parnas 途径基因、卡尔文循环基因、一磷酸核酮糖途径基因、丝氨酸途径基因、一磷酸木酮糖途径基因、甲胺途径基因、丙氨酸途径基因、油酸途径基因、甘油-3-磷酸脱氢酶基因和甘油-3-磷酸酶基因，

由此生成靶分子；和

(b)任选地回收(a)中生成的靶分子。

17. 一种在重组宿主细胞中生产 1,3-丙二醇的方法，其包括：

(a)在存在极限糊精、在合适的条件下接触转化的宿主细胞，所述宿主细胞包含：

(i)分离的核酸分子，其编码 SEQ ID NO:6 的氨基酸序列；或

(ii)核酸分子，其与(i)完全互补；和

(iii)至少一种选自下组的嵌合基因：甘油脱水酶基因、二醇脱水酶基因和 1,3-丙二醇氧化还原酶基因，

由此生成 1,3-丙二醇；和

(b)任选地回收(a)中生成的 1,3-丙二醇。

18. 一种在重组宿主细胞中生产甘油的方法，其包括：

(a)在存在极限糊精、在合适的条件下接触转化的宿主细胞，所述

宿主细胞包含:

(i)分离的核酸分子，其编码 SEQ ID NO:6 所示的氨基酸序列;或

(ii)核酸分子，其与(i)完全互补; 和

(iii)至少一种选自下组的嵌合基因: 果糖-1,6-二磷酸途径基因、Embden-Meyerhof Parnas 途径基因、卡尔文循环基因、一磷酸核酮糖途径基因、丝氨酸途径基因、一磷酸木酮糖途径基因、甲胺途径基因、丙氨酸途径基因、油酸途径基因、甘油-3-磷酸脱氢酶基因和甘油-3-磷酸酶基因，

由此生成甘油; 和

(b)任选地回收(a)中生成的甘油。

19. 一种在重组宿主细胞中生产细胞物质的方法，其包括:

(a)在存在极限糊精、在合适的条件下接触转化的宿主细胞，所述宿主细胞包含:

(i)分离的核酸分子，其编码 SEQ ID NO:6 的氨基酸序列;或

(ii)核酸分子，其与(i)完全互补;

由此生成细胞物质; 和

(b)任选地回收(a)中生成的细胞物质。

20. 一种降解极限糊精的方法，其包括:

(a)使转化的宿主细胞在合适的生长条件下接触有效量的极限糊精底物，所述宿主细胞包含:

(i)核酸分子，其编码 SEQ ID NO:6 的酶;或

(ii)核酸分子，其与(i)完全互补;和

(b)任选地回收步骤(a)的产物。

淀粉制品在利用发酵的生物生产中的应用

本申请要求享有2002年8月23日提交的美国临时申请号60/405,896的利益。

技术领域

本发明涉及分子生物学领域。更具体地，它描述了含有能表达酶的基因的微生物宿主，所述酶能有效地将淀粉制品转化成发酵产物。

背景技术

发酵是将可再生的原料生物催化地转化成理想产物的重要技术。碳水化合物是发酵工业的传统原料。经常见到的情况是，用作底物的碳水化合物在生产成本中的比重比任何其它的单一组分都大。根据具体的工艺，25～70%的总发酵成本可以归因于碳水化合物源(Crueger 和 Crueger, Biotechnology: A Textbook of Industrial Microbiology, Sinauer Associates: Sunderland, MA. , 第 124-174 页(1990); Atkinson 和 Mavituna, Biochemical Engineering and Biotechnology Handbook, 第2版; Stockton Press: New York, 第243-364页(1991))。出于这样的经济原因，很少用高纯化的葡萄糖或蔗糖作为底物。

淀粉（碳水化合物）是两种不同多糖的混合物，每一种多糖由连接的重复单糖（葡萄糖）单元的链组成。该混合物由两种分离的多糖组成，即直链淀粉和支链淀粉。直链淀粉是线性多糖，含有专有地通过α(1,4)糖苷键连接的葡萄糖单元。支链淀粉中的葡萄糖单元也是通过α(1,4)糖苷键连接，另外还通过α(1,6)糖苷键连接，每30个葡萄糖残基约有一个。淀粉中的支链淀粉与直链淀粉的比例根据植物物种的不同而变化，但是其范围通常是3-4至1 (Kainuma, Starch 第125-150 页; Whistler, Bemiller 和 Pashcall 编, Academic Press, Orlando, FL (1984))。

商业淀粉主要是通过湿磨法生产的。但是，湿磨的终产物包括非常少的未加工的淀粉。到目前为止，生产的大多数产品是完全加工的淀

粉（单糖，包括葡萄糖）或源自淀粉的更小的降解产物的形式。一般地，用淀粉酶将淀粉分解成更小的链(Blanchard, Technology of Corn wet Milling (1992), Elseiver, Amsterdam, 荷兰, 第 174-215 页)。存在各种商业化的 α-淀粉酶源，但是，无论酶源如何，反应产物关于大小和键合类型通常是相同的。已知淀粉的淀粉酶消化产物是极限糊精，其包含含有 2-10 个葡萄糖单元(寡糖)的小淀粉链。因为淀粉酶不能水解支链淀粉中的 α(1,6)糖苷键，所以极限糊精含有 α(1,4)-和 α(1,6)-连接的葡萄糖寡糖。或者，可以通过非酶法（例如，通过酸水解）处理粗淀粉，生产基本上类似于极限糊精的淀粉制品。

在湿磨工业中，极限糊精被进一步加工成葡萄糖，其用作发酵的碳源，以生产各种化学制剂、商业酶或抗生素。当流行的生物催化剂大肠杆菌（Escherichia coli）用在发酵工艺中时，相对较纯的葡萄糖是优选的碳水化合物源。这是因为，大肠杆菌不能利用极限糊精组分(即，潘糖、异麦芽糖和链长超过约 10 个 α(1,4)-连接的葡萄糖单元的高分子量的寡糖)，它们通常包含在备用的廉价发酵培养基中(Lin,Escherichia coli and Salmonella typhimuium, 第 245-265 页, Neidhardt, 编; American Society for Microbiology, Washington, D. C. (1987))。含有 α(1,6)-键的葡萄糖寡聚体不能运输到细胞中，当胞外地供给该材料时，大肠杆菌不能生成能降解它的酶(Palmer 等, Eur. J. Biochem. 39: 601-612 (1973))。

能从适合于大肠杆菌使用的淀粉生产出相对较纯的葡萄糖，这需要许多工艺步骤和其它的酶，它们显著地增加了产品的生产成本。

因而，要解决的问题是，缺少一种能在大规模发酵生产工艺中利用廉价的淀粉制品的方法。如果能更彻底地发酵廉价的、部分地降解的淀粉，将会降低通过发酵生产的产品的生产成本。

发明概述

申请人已经提供了一种分离的编码 α(1,6)-连接的葡萄糖寡糖水解酶的选自下述的核酸分子：(a)分离的核酸分子，其编码选自 SEQ ID NO: 2、4 和 6 的氨基酸序列；(b)核酸分子，其能在下述杂交条件下与(a)杂交：0. 1X SSC, 0.1% SDS, 65℃，以及用 2X SSC, 0.1% SDS 洗涤，然后用 0. 1X SSC, 0.1% SDS 洗涤；和(c)核酸分子，其与(a)

或(b)互补。

申请人已经提供了核酸组合物，其包含信号肽和α(1,6)-连接的葡萄糖寡糖水解酶的编码区，所以能表达指导细胞质外的（胞外的）水解活性的嵌合蛋白。分离的核酸分子可以编码如SEQ ID NO: 24或SEQ ID NO: 25所述信号肽。信号序列的核酸序列是SEQ ID NO: 26或SEQ ID NO: 27。分离的核酸分子可以编码如SEQ ID NO:2、4、6、17或31所述能水解α(1,6)-连接的葡萄糖寡糖的多肽。

申请人已经提供了重组生物，其包含α(1,6)-连接的葡萄糖寡糖水解酶，它能利用外源地添加的α(1,6)-连接的葡萄糖寡糖(例如，异麦芽糖和潘糖)，以发酵生产有用产物。能水解α(1,6)-连接的葡萄糖寡糖的多肽可以选自SEQ ID NO: 2、SEQ ID NO: 6、SEQ ID NO: 17或SEQ ID NO: 31。本发明还包括SEQ ID NO: 1、3、5、16或30所述核酸分子编码的能水解α(1,6)-连接的葡萄糖寡糖的多肽。本发明还包括选自SEQ ID NO: 3、SEQ ID NO: 28、SEQ ID NO: 32、SEQ ID NO: 34、SEQ ID NO: 36、SEQ ID NO: 38、SEQ ID NO: 40或SEQ ID NO: 42的分离的核酸分子。本发明还包括多肽SEQ ID NO: 4、SEQ ID NO: 29、SEQ ID NO: 33、SEQ ID NO: 35、SEQ ID NO: 37、SEQ ID NO: 39、SEQ ID NO: 41和SEQ ID NO: 43。

本发明还包括嵌合基因，其包含可操作地连接到合适的调控序列的本文所述分离的核酸分子。合适的调控序列选自 *CYC1, HIS3, GAL1, GAL10, ADH1, PGK, PHO5, GAPDH, ADC1, TRP1, URA3, LEU2, ENO, TPI, AOX1, lac, ara, tet, trp,* IP_L, IP_R, *T7, tac, trc, apr, npr, nos*和 *CI*。本发明包括转化的宿主细胞，其中嵌合基因整合在染色体中或是质粒-携带的。

申请人还已经提供了一种降解极限糊精的方法，其包括:

(a)使转化的宿主细胞在合适的生长条件下接触有效量的极限糊精底物，所述宿主细胞包含:

(i)核酸分子，其编码选自SEQ ID NO: 2、6、17和31的酶;

(ii)核酸分子，其能在下述杂交条件下与(i)杂交:0.1X SSC, 0.1% SDS, 65℃，用2X SSC, 0.1% SDS洗涤，然后用0.1X SSC, 0.1% SDS洗涤; 或

(iii)核酸分子，其与(i)或(ii)互补;和

(b)任选地回收步骤(a)的产物。

本发明还包括一种在重组宿主细胞中生产靶分子的方法，其包括在存在极限糊精、在合适的条件下使转化的宿主细胞与用于将单糖转化成靶分子的嵌合基因接触，所述宿主细胞包含：(i)分离的核酸分子，其编码由连接到能水解 α(1,6)-连接的葡萄糖寡糖的多肽的信号肽组成的嵌合蛋白；(ii)核酸分子，其能在下述杂交条件下与(i)杂交：0.1X SSC，0.1% SDS，65℃，用 2X SSC， 0.1% SDS 洗涤，然后用 0.1X SSC， 0.1% SDS 洗涤；或(iii)核酸分子，其与(i)或(ii)互补，由此生成靶分子；和任选地回收生成的靶分子。信号肽可以选自 SEQ ID NO: 24 或 SEQ ID NO: 25。能水解 α(1,6)-连接的葡萄糖寡糖的多肽可以选自 SEQ ID NO: 2、SEQ ID NO: 6、SEQ ID NO: 17 或 SEQ ID NO: 31。转化的宿主细胞可以选自细菌、酵母或丝状真菌。 本发明包括从极限糊精生产 1,3 丙二醇、甘油和细胞因（Cell Mass）

本发明还包括多肽，当在 BLASTP 比对方法的基础上与具有 SEQ ID NO: 17 所述序列的多肽（具有 α(1,6)-连接的葡萄糖寡糖水解活性的多肽）相比较时，其氨基酸序列具有至少 69%的相同性。

附图简述，生物保藏和序列说明

图 1a 至 1d 显示了大肠杆菌菌株 DH5α 的结果，该菌株含有质粒 pUC18 (图 1a)(阴性对照)和含有来自克隆 j20 (图 1b)、k1 (图 1c)或 h12 (图 1d)的成熟编码序列的 pUC18 。从超声处理的细胞分离出总蛋白提取物，与潘糖(250 μg/ml)在 37 ℃温育 2 小时。显示了消化后的产物的高效阴离子交换层析谱。

根据布达佩斯条约中关于用于专利程序的微生物保藏的国际认可的条款，申请人在美国典型培养物保藏中心（American Type Culture Collection）(ATCC)10801 University Boulevard, Manassas, VA 20110-2209 进行了下述生物保藏：

保藏人鉴别参考	国际保藏编码	保藏日期
大肠杆菌 RJ8n	ATCC PTA-4216	2002 年 4 月 9 日

列出的保藏物将在指定的国际保藏单位保藏至少 30 年，并且在公

开它的专利授权后可以由公众获得。该保藏物的可获得性并不构成对实施主题发明的许可，所述实施会损害通过政府行为授予的专利权。

申请人提供了含有 43 个序列的序列表。这些序列能满足 37 C. F. R. 1.821-1.825（“对含有核苷酸序列和/或氨基酸序列公开内容的专利申请的要求——序列规则”），并且与世界知识产权组织 (WIPO) 标准 ST. 25 (1998)和 EPO 和 PCT (细则 5.2 和 49.5 (a-bis)和管理规程 208 部分和附件 C)以及相应的美国专利商标局细则 37 C. F. R. §1.822 所述的序列表要求相一致。

ORF 名称	基因名称	SEQ ID 碱基	SEQ ID 肽	来源菌株
mbc1g.pk007.h12	*algB*	1	2	短双歧杆菌
mbc2g.pk018.j20	*algA*	3	4	短双歧杆菌
mbc1g.pk026.k1	*algA*	5	6	短双歧杆菌
dexB	*dexB*	16	17	变异链球菌

SEQ ID NO: 1-6 是从短双岐杆菌（Bifidobacterium breve）ATCC 15700 得到的 3 种基因/基因产物的核酸和氨基酸序列。

SEQ ID NO: 7-15 和 18-23 是用于 PCR 的引物。

SEQ ID NO: 16-17 是公开数据文库中公开的变异链球菌（Streptococcus mutans）(ATCC 25175D)的核酸和氨基酸序列。

SEQ ID NO: 24 是来自短双岐杆菌基因 mbc2g. pk018. j20 (也包含在 SEQ ID NO: 3 中)的天然信号肽的氨基酸序列。

SEQ ID NO: 25 是用于靶向由短双岐杆菌 mbc1g. pk026. k1 和变异链球菌 dexB 基因编码的酶的非天然的信号肽的氨基酸序列。

SEQ ID NO: 26 是短双岐杆菌基因 mbc2g. pk018. J20 信号肽的核酸序列。

SEQ ID NO: 27 是枯草芽孢芽孢杆菌（Bacillus subtilis）中性蛋白酶基因信号肽的核酸序列。

SEQ ID NO: 28 是短双岐杆菌基因 mbc2g. pk018. j20 信号肽的核酸序列，其连接到短双岐杆菌 mbc2g. pk018. h12 基因的编码序列。

SEQ ID NO: 29 是短双岐杆菌基因 mbc2g. pk018. j20 信号肽的氨基酸序列，其连接到短双岐杆菌 mbc2g. pk018. h12 基因的氨基酸序列。

SEQ ID NO: 30 是没有其天然信号肽序列的短双岐杆菌基因 mbc2g.

pk018. j20 的核酸序列。

SEQ ID NO: 31 是没有其天然信号肽序列的短双岐杆菌基因 mbc2g. pk018. j20 的氨基酸序列。

SEQ ID NO: 32 是短双岐杆菌基因 mbc2g. pk018. j20 信号肽的核酸序列，其连接到短双岐杆菌 mbc2g. pk018. k1 基因的编码序列。

SEQ ID NO: 33 是短双岐杆菌基因 mbc2g. pk018. j20 信号肽的氨基酸序列，其连接到短双岐杆菌 mbc2g. pk018.k1 基因的氨基酸序列。

SEQ ID NO: 34 是短双岐杆菌基因 mbc2g. pk018. j20 信号肽的核酸序列，其连接到变异链球菌 dexB 基因的编码序列。

SEQ ID NO: 35 是短双岐杆菌基因 mbc2g. pk018. j20 信号肽的氨基酸序列，其连接到变异链球菌 dexB 基因的氨基酸序列。

SEQ ID NO: 36 是枯草芽孢杆菌中性蛋白酶基因信号肽的核酸序列，其连接到短双岐杆菌 mbc2g. pk018. h12 基因的编码序列。

SEQ ID NO: 37 是枯草芽孢杆菌中性蛋白酶基因信号肽的氨基酸序列，其连接到短双岐杆菌 mbc2g. pk018. h12 基因的氨基酸序列。

SEQ ID NO: 38 是枯草芽孢杆菌中性蛋白酶基因信号肽的核酸序列，其连接到短双岐杆菌 mbc2g. pk018. j20 基因的编码序列。

SEQ ID NO: 39 是枯草芽孢杆菌中性蛋白酶基因信号肽的氨基酸序列，其连接到短双岐杆菌 mbc2g. pk018. j20 基因的氨基酸序列。

SEQ ID NO: 40 是枯草芽孢杆菌中性蛋白酶基因信号肽的核酸序列，其连接到短双岐杆菌 mbc2g. pk018. k1 基因的编码序列。

SEQ ID NO: 41 是枯草芽孢杆菌中性蛋白酶基因信号肽的氨基酸序列 ，其连接到短双岐杆菌 mbc2g. pk018. k1 基因的氨基酸序列。

SEQ ID NO: 42 是枯草芽孢杆菌中性蛋白酶基因信号肽的核酸序列，其连接到变异链球菌 dexB 基因的编码序列。

SEQ ID NO: 43 是枯草芽孢杆菌中性蛋白酶基因信号肽的氨基酸序列，其连接到变异链球菌 dexB 基因的氨基酸序列。

发明详述

申请人已经解决了所述问题。本发明提供了几种酶，当在生产宿主中表达时，它们能使宿主利用 α(1,6)-连接的葡萄糖寡糖，该寡糖是

廉价的淀粉制品组分。本发明还提供了信号序列，它能使 α(1,6)-连接的葡萄糖寡糖水解酶靶向胞外。

可以得到廉价的淀粉制品，例如通过市场上可买到的淀粉酶对粗淀粉和其它含有 α(1,6)-连接的葡萄糖寡糖的原料的作用，该作用会生成极限糊精。廉价淀粉制品的有效利用需要对宿主生物（例如大肠杆菌）进行基因工程处理，从而使该重组生物生产能降解 α(1,6)-连接的葡萄糖寡糖的酶。能降解 α(1,6)-连接的葡萄糖寡糖的酶是已知的(Vihinen 和 Mantsala Crit. Rev. in Biochem. Mol. Biol. 4: 329-427 (1989))。而且，已知能降解这些键的酶以其天然状态细胞内地（在细胞质内）和胞外地（在细胞质外）存在。

当宿主生物缺少运输系统时，通过添加天然的或非天然的信号肽，可以对细胞内酶进行工程处理以使其可接近极限糊精（或其它含有 α(1,6)-连接的葡萄糖寡糖的原料)。信号肽使能降解 α(1,6)-连接的葡萄糖寡糖的蛋白被导向到胞外位置(细胞质外)，并接近没有摄入到细胞中的底物(Nagarajan 等，Gene 114: 121-126 (1992))。能将蛋白穿过细胞膜移位的信号肽的实例包括但不限于 SEQ ID NO: 24 和 25。含有信号肽的蛋白指向分泌途径，然后穿过细胞膜移位。蛋白分泌的普通机制在所有革兰氏阴性的和革兰氏阳性的细菌都是保守的(Simonen 和 Palva (1993)Microbiol. Rev. 57: 109-137; Fekkes 和 Driessen (1999)Microbiol. Rev. 63: 161-173)。所有细菌的信号肽都包含 13～20 个疏水氨基酸串(Bae 和 Schneewind, J. Bacteriol., 185: 2910-2919 (2003))。

天然的大肠杆菌不能水解 α(1,6)-糖苷键，因而在发酵中不能利用含有(1,6)-键的化合物。(1,6)-键能被“异淀粉酶”和“葡糖苷酶”所水解(异麦芽糖和潘糖是(1,6)-连接的寡糖的模型化合物)。含有非天然的胞外“异淀粉酶”或“葡糖苷酶”的重组大肠杆菌能在发酵中利用含有 (1,6)-键的化合物(例如，异麦芽糖和潘糖)来生产有用的产物。而且，含有非天然的胞外“异淀粉酶”或“葡糖苷酶”的任何重组生物都能更有效地利用含有(1,6)-键的化合物。提高的使用效率将会贯穿重组的“异淀粉酶”或“葡糖苷酶”基因的组成型表达或改变的定时。重组基因的表达也会提高活性水平，使其超过可能存在的任何内源“异淀粉酶”或“葡糖苷酶”基因的水平，从而提高对(1,6)-

连接的底物的利用。

本发明可以用于生产各种生物发酵产物，包括但不限于有机酸、抗生素、氨基酸、酶、维生素、醇（例如生物醇（bioethanol））和细胞物质（cell mass）。通过使用信号肽，用极限糊精作为宿主微生物的碳源来生物生产甘油、1,3-丙二醇和细胞物质，这是本发明的示例。

多元醇 1,3-丙二醇是对生产聚酯纤维和生产聚氨基甲酸酯和环状化合物有用的单体。通过单一生物从碳底物（例如葡萄糖或其它糖）生物生产 1,3-丙二醇的方法已经描述在美国专利号 5,686,276 中，它在这里引作参考。

淀粉是葡萄糖的同多糖。它在高等植物中合成为含有两种组分的颗粒，即直链淀粉和支链淀粉 (Vihinen 和 Mantsala, Crit. Rev. Biochem. Mol. Biol., 24: 329-418 (1989))。直链淀粉是通过α(1,4)-连接的葡萄糖残基形成的基本上线性的多糖，占颗粒的 15-25%（含量根据植物物种不同而异）。相反地，支链淀粉是高度分支的，约 4－5%的糖苷键是α(1,6)-连接的葡萄糖残基。已经深入地研究了能降解淀粉的淀粉分解酶。高等植物物种中的淀粉代谢是通过首先将聚合物降解成单个的葡萄糖残基，这需要几种淀粉分解酶的相互作用。

淀粉分解酶在单独作用时，通常仅能将淀粉部分地降解为更小的直链或支链。可以用淀粉分解酶的组合或者与酸处理一起的酶组合来提高淀粉的解聚。

酶和酶组合可以将淀粉部分地降解成更小的直链或支链，或完全降解成葡萄糖。α-葡糖苷酶能水解寡糖中的(1,4)-和(1,6)-键，所述寡糖是通过其它淀粉分解酶（例如 α-淀粉酶、ß-淀粉酶、葡糖淀粉酶、异淀粉酶和支链淀粉酶）或通过酸和热处理的作用形成的。

α-葡糖苷酶 (α-D-葡糖苷葡糖水解酶; 例如, EC 3.2.1.20)广泛地分布在微生物中。它们能水解(1,4)-和(1,6)-键，并从非还原末端释放出 α-D-葡萄糖单元。已经在下述物种中发现了这些具有不同的(和广泛的）底物特异性的酶的不同类型：细菌例如芽孢杆菌属（Bacillus），链球菌属(Streptococcus)，埃希氏菌属(Escherichia)，假单胞菌属(Pseudomonas)，超嗜热古细菌例如火球菌属（Pyrococcus）、热球菌属（Thermococcus）和栖热孢菌病（Thermotoga），和真菌物种例如青霉属（Penicillium）、四膜虫属

(Tetrahymena)、糖酵母属(Saccharomyces)和曲霉属(Aspergillus)。

已经广泛地研究了来自黑曲霉(Aspergillus niger)的酶多年,其具有广泛的底物特异性。其水解这样的底物,例如麦芽糖、曲二糖、黑曲糖、异麦芽糖、苯基-α-葡糖苷、苯基-α-麦芽糖苷、寡糖、麦芽糖糊精和可溶性的淀粉。来自米曲霉(A. oryzae)、枯草芽孢杆菌和蜡状芽孢杆菌(B.cereus)和超嗜热古细菌的α-葡糖苷酶表现出类似的性质。

寡-(1,6)-葡糖苷酶或异麦芽糖酶(糊精6-α-D-葡聚糖水解酶,EC 3.2.1.10; dexB基因编码)是类似于α-葡糖苷酶的酶(Krasikov等,Biochemistry (Moscow). 66: 332-348 (2001))。它能催化由α-淀粉酶水解淀粉和糖原生成的异麦芽糖和糊精中的(1,6)-α-D-葡糖苷键(Enzyme Nomenclature, C. Webb, 编 (1984)Academic Press, San Diego, CA.)。该酶比α-葡糖苷酶更少地分布,但是在生物中有所发现,例如芽孢杆菌属物种,包括热葡糖苷芽孢杆菌(B. thermoglucosidius) KP1006、蜡状芽孢杆菌ATCC 7064和可能的解淀粉芽孢杆菌(B. amyloliquefaciens) ATCC 23844 (Vihinen和Mantsala, Critical Reviews in Biochemistry. 24: 329-418 (1989)),以及凝结芽孢杆菌(Bacillus coagulans) (Suzuki和Tomura, Eur. J. Biochem., 158: 77-83 (1986))。芽孢杆菌酶的大小一般是60-63 kDa。还已经从热厌氧菌属(Thermoanaerobium) Tok6-B1分离出寡-(1,6)-α-葡糖苷酶 (EC 3.2.1.10),报道其分子量为30-33 kDa。

来自变异链球菌的dexB酶的活性模式类似于葡聚糖酶 (EC 3.2.1.11),后者能催化葡聚糖中的(1,6)-α-D-葡糖苷键的内水解(endohydolysis)。dexB酶和芽孢杆菌属物种寡-(1,6)-葡糖苷酶之间有较高程度的相似性(Whiting等, J. Gen. Microbiol., 139: 2019-2026 (1993))。DexB大小约是62 kDa (Aduse-Opoku等, J. Gen Microbiol., 137: 757-764 (1991))。

具有α(1,6)水解酶活性的酶属于超过81种鉴别出的葡糖基水解酶家族的非常宽的种类(Henrissat, Biochem. J., 280: 309-316 (1991); Henrissat和Bairoch, Biochem. J., 293: 781-788

(1993))。通过使用少至69%氨基酸序列的相同性的酶证实本发明的用途，进一步强调了广分类的能利用α(1,6)连接的葡萄糖单元作为发酵底物的酶。已知能解聚含有α(1,6)-连接的葡萄糖残基的寡糖的酶，包括：葡糖淀粉酶(EC 3.2.1.3，也称作淀粉葡糖苷酶)，它在序列中的下一键是α(1,4)-α-D-葡糖苷键时能快速地水解(1,6)-α-D-葡糖苷键或连接；α-糊精内-(1,6)-α-葡聚糖苷酶（glucanosidase）(EC 3.2.1.41,也称作支链淀粉酶)，它能降解支链淀粉、支链淀粉、糖原和支链淀粉和糖原的α-和β-淀粉酶极限糊精中的(1,6)-α-D-葡糖苷键；蔗糖酶(EC 3.2.1.48)，它分离自肠粘膜，且具有抗异麦芽糖的活性;异淀粉酶 (EC 3.2.1.68)，它能水解糖原、支链淀粉和它们的β-极限糊精中的(1,6)-α-D-葡糖苷键；和葡聚糖(1,6)-α-葡糖苷酶 (EC 3.2.1.70)，它能水解来自(1,6)-α-D-葡聚糖和衍生的寡糖的连续葡萄糖残基。

在本公开内容的上下文中，使用了许多术语。

术语“淀粉”是指由通过糖苷键连接的 D-葡萄糖单元组成的同多糖，其形成植物中的营养储存文库。淀粉存在 2 种形式，即直链淀粉和支链淀粉。在直链淀粉中，D-葡萄糖单元专有地通过α(1,4)糖苷键连接。认为由多个α(1,4)糖苷键组成的链是线性的或未分支的。在支链淀粉中,尽管主要的连接是通过α(1,4)糖苷键,但偶尔存在的α(1,6)糖苷键形成了另外的线性部分的分支点。支链淀粉在每 30 个 α(1,4)键中含有约 1 个α(1,6)键。

术语“单糖”是指经验式$(CH_2O)_n$的化合物，其中 $n \geq 3$，碳骨架是未分支的，除了 1 个碳原子外的每个碳原子都含有羟基，剩余的碳原子是在碳原子 2 上的醛或酮。术语“单糖”还指细胞内的环状半缩醛或半酮缩醇形式。最熟悉的单糖是 D-葡萄糖。环状形式的 D-葡萄糖涉及碳原子 5 的羟基与碳原子 1 的醛形成半缩醛的反应，其中羰基碳被称为异头碳。

术语“糖苷键”和“糖苷连接”是指通过异头碳与醇羟基的反应形成的缩醛。1 个 D-葡萄糖分子的异头碳与第 2 个 D-葡萄糖分子的碳原子 4 上的羟基的反应产生α(1,4)糖苷键或连接。类似地，1 个 D-葡萄糖分子的异头碳与第 2 个 D-葡萄糖分子的碳原子 6 上的羟基的反应产生α(1,6)糖苷键或连接。本领域的技术人员能够认识到，糖苷键可以

是α或ß构型。糖苷键构型通过例如α(1,4)和α(1,6)命名。

术语“α”是指键在环平面之上的构象。相反地，“ß”是指键在环平面之下的构象。

术语“寡糖”是指含有 2～10 个通过糖苷键连接的单糖单元的化合物。术语“多糖”是指含有超过 10 个通过糖苷键连接的单糖单元的化合物，且通常是指大分子量种类的混合物。由单一单体单元组成的多糖由术语“同多糖”表示。

术语“异麦芽糖（isomaltosaccharide）”是指具有至少一个α(1,6)-键的寡糖。

术语“(1,4)键”是指两个糖之间的关系，其中 1 个糖单元的 C1 结合到第 2 个糖单元的 C4。

术语“(1,6)键”是指两个糖之间的关系，其中 1 个糖单元的 C1 结合到第 2 个糖单元的 C6。

术语“淀粉酶”和“α-淀粉酶”是指能催化 α(1,4)糖苷键的水解的酶。α(1,4)糖苷键的水解活性是通过术语“淀粉酶活性”或“淀粉分解活性”表示的。淀粉酶包括但不限于 IUBMB 分类 EC 3.2.1.1 (淀粉酶)、EC 3.2.1.60 ((1,4)-α-maltotetraohydrolase) 和 EC 3.2.1.98 ((1,4)-α-(maltohexaosidase))。

术语“异淀粉酶”和“α-异淀粉酶”是指能催化 α(1,6)糖苷键的水解的酶。α(1,6)糖苷键的水解活性是通过术语“异淀粉酶活性”或“异淀粉分解活性”表示。异淀粉酶包括但不限于 IUBMB 分类 EC 3.2.1.10 (寡-(1,6)-葡糖苷酶)、EC 3.2.1.11 (葡聚糖酶)、EC 3.2.1.41 (支链淀粉酶)和 EC 3.2.1.68 (异淀粉酶)。

术语“葡糖苷酶”和“α-葡糖苷酶”是指能催化 α(1,4)糖苷键和α(1,6)糖苷键的水解并从寡糖的非还原末端释放出 α-D-葡萄糖单元的酶。葡糖苷酶具有淀粉分解活性和异淀粉分解活性。葡糖苷酶包括但不限于 IUBMB 分类 EC 3 2.1. 3 (淀粉葡糖苷酶)和 EC 3.2.1.20 (α-葡糖苷酶)。

术语“α(1,6)-连接的葡萄糖寡糖水解酶”是指具有催化 α(1,6)糖苷键水解的功能活性的酶。具有这样的功能活性的酶的具体实例包括异淀粉酶、α-异淀粉酶、葡糖苷酶和α-葡糖苷酶。

术语“异麦芽糖酶”或“寡-(1,6)-葡糖苷酶”或“糊精 6-α-D-葡聚

糖水解酶”是指酶 (EC 3.2.1.10)，它仅能水解寡糖的非还原末端的α(1,6)-键。

术语“DexB”是指由变异链球菌的 dexB 基因(GenBank 登记号 M77351)编码的(1,6)-α-葡糖苷酶，它从α(1,6)-连接的异麦芽糖和葡聚糖的非还原末端释放出葡萄糖。

术语“极限糊精”是指含有单糖和寡糖的淀粉的淀粉分解降解产物。淀粉酶对支链淀粉的作用会生成单糖(D-葡萄糖)、二糖(α(1,4)连接的麦芽糖和 α(1,6)连接的异麦芽糖)和更高的寡糖的混合物。更高的寡糖可以是线性的(专有地含有 α(1,4)键)或分支的(主要含有α(1,4)键和α(1,6)键)。

术语“聚合度”或“DP”是指在糖混合物的各组分中的单体单元的数目；例如，单糖例如 D-葡萄糖）的 DP 为 1，二糖（例如麦芽糖）的 DP 为 2，三糖（例如潘糖）的 DP 为 3，等。当应用到多糖混合物或寡糖混合物时，DP 是指每个分子中的单体的平均数目。

术语“葡萄糖当量”("DE")是指通过 Lane-Eynon 滴定测定的“表示为干物质的葡萄糖百分比的还原糖含量”(Handbook of Starch Hydrolysis Products and their Derivatives, M. W. Kearsely 和 S. Z. Dziedzic, 编, Blackie Academic & Professional, 第 86 页）。DE 大小表示了淀粉的水解度，淀粉具有 0 DE 标称值而最终水解产物具有 100 DE 值。

淀粉酶和异淀粉酶活性可以是胞内的或胞外的。为了本发明的目的，术语“胞内活性”是指当提供底物时能从破碎的细胞或细胞提取物观察到、但是当胞外地提供底物时不能从完整细胞观察到的酶活性。术语“胞外活性”是指当胞外地提供底物时能从完整细胞（包括生长的细胞）观察到的活性。酶底物不能被动扩散或主动运输进细胞，由术语“胞内活性”和“胞外活性”暗示。

“靶分子”是指生物催化地生成的产物。这可以是通过生物催化剂自然地生成的化合物，或者非天然的基因可以进行基因工程处理进入微生物中，以用于它们在生物发酵中的功能表达。在本上下文中的“靶分子”还指生物发酵的任何副产物，它们可以理想地从生物发酵系统中选择性地除去，以消除反馈抑制和/或使生物催化剂活性最大化。

“体积生产率”是指单位时间内在给定体积的生物发酵罐中生成的

靶分子质量，其单位为克/（升·小时）(缩写为 g/（L·hr））。该度量是通过生物催化剂的特异性活性和生物催化剂的浓度测定的。它从效价、运行时间和生物发酵罐的工作体积计算出。

“效价”是指靶分子的浓度，其单位为克/升(缩写为 g/L)。

术语“多核苷酸”或“多核苷酸序列”、“寡核苷酸”、“核酸序列”和“核酸片段”或“分离的核酸片段”在本文中互换地使用。这些术语包括核苷酸序列等。多核苷酸可以是单链的或双链的 RNA 或 DNA 的聚合物，所述 RNA 或 DNA 任选地含有合成的、非自然的或改变的核苷酸碱基。DNA 聚合物形式的多核苷酸可以由 cDNA、基因组 DNA、合成的 DNA 的一个或多个片段或它们的混合物组成。

术语“分离的”是指诸如核酸分子和/或蛋白的材料基本上不含有通常在自然环境中伴随该材料或与其相互作用的组分，或者已经从其中分离出。分离的多核苷酸可以是从它们自然存在的宿主细胞中纯化出的。可以使用本领域技术人员已知的常规核酸纯化方法来获取分离的多核苷酸。该术语还包括重组多核苷酸和化学合成的多核苷酸。

如本文所使用的，“分离的核酸分子”或“分离的核酸片段” 是单链的或双链的 RNA 或 DNA 的聚合物，所述 RNA 或 DNA 任选地含有合成的、非自然的或改变的核苷酸碱基。DNA 聚合物形式的分离的核酸片段可以由 cDNA、基因组 DNA 或合成的 DNA 的一个或多个片段组成。

术语“互补”用于描述能与另一个核苷酸碱基杂交的核苷酸碱基之间的关系。例如，关于 DNA，腺嘌呤与胸腺嘧啶互补，胞嘧啶与鸟嘌呤互补。因此，本发明还包括能与所附序列表报道的完整序列以及基本上类似的核酸序列互补的分离的核酸片段。

如本文所使用的，“基本上类似的”是指核酸片段，其中的一个或多个核苷酸碱基的变化造成了一个或多个氨基酸的替代，但是不影响该核苷酸序列编码的多肽功能性质。因此可以明白，本发明不仅包括具体示例的核苷酸或氨基酸序列，还包括它们的功能等同物。术语“基本上类似的”和“基本上对应的”在本文中互换地使用。

而且，本领域熟知能导致在指定的位点生成化学上等同的氨基酸、但是不影响编码的多肽的功能性质的核酸片段中的改变。因此，氨基酸丙氨酸（一种疏水氨基酸）的密码子可以替代为编码另一个更少疏水性残基（例如甘氨酸）的密码子或编码另一个更多疏水性残基（例

如缬氨酸、亮氨酸或异亮氨酸）的密码子。类似地，可以预期能导致一个带负电荷的残基替代另一个（例如天冬氨酸替代谷氨酸）或一个带正电荷的残基替代另一个（例如赖氨酸替代精氨酸）的变化也能生成功能上的等价产物。还可以预期，能导致多肽分子的 N-端和 C-端部分的改变的核苷酸变化不会改变多肽的活性。每一种推定的修饰都是本领域的常规技术，保持编码产物的生物学活性的测定也是如此。

而且，基本上类似的核酸片段还可以通过它们的杂交能力来表征。这种同源性是通过严格条件下的 DNA-DNA 或 DNA-RNA 杂交来估计的，如本领域的技术人员所熟知的 (Hames 和 Higgins, 编 (1985) Nucleic Acid Hybridisation, IRL Press, Oxford, 英国)。可选择严格条件以用于筛选中等相似的片段，例如来自关系较远的生物的同源序列，至高度相似的片段，例如能加倍来自密切相关的生物的功能酶的基因。杂交后的洗涤决定了严格条件。一组优选的条件使用一系列的洗涤：开始是 6X SSC, 0.5% SDS，在室温进行 15 分钟，然后重复 2X SSC, 0.5% SDS，在 45 ℃ 进行 30 分钟，然后重复 2 次 0.2X SSC, 0.5% SDS，在 50 ℃进行 30 分钟。一组更优选的严格条件使用更高的温度，其中的洗涤与上面相同，只是最后 2 次在 0.2X SSC, 0.5% SDS 中的 30 分钟洗涤的温度提高到了 60 ℃。另一个优选的高度严格的条件使用两次在 0.1X SSC, 0.1% SDS 中、在 65 ℃的最终洗涤。

本发明的基本上类似的核酸片段还可以通过它们编码的氨基酸序列与本文所公开的氨基酸序列的百分比相同性来表征，如通过本领域的技术人员通常采用的算法所测定的。合适的核酸片段(本发明的分离的多核苷酸)编码的多肽与本文所报道的氨基酸序列至少 70%相同、优选地至少 80%相同。优选的核酸片段编码的氨基酸序列与本文所报道的氨基酸序列至少 85%相同。更优选的核酸片段编码的氨基酸序列与本文所报道的氨基酸序列至少 90%相同。最优选的核酸片段编码的氨基酸序列与本文所报道的氨基酸序列至少 95%相同。合适的核酸片段不仅具有上述的相同性，一般还编码具有至少 50 氨基酸、优选地至少 100 氨基酸、更优选地至少 150 氨基酸、再优选地至少 200 氨基酸和最优选地至少 250 氨基酸的多肽。

本领域的技术人员熟知，序列相同性的许多水平可以用于鉴别相关的多肽序列。百分比相同性的有用实例是 50%、55%、60%、65%、70%、

75%、80%、85%、90%或95%或55%至100%之间的任何整数百分比。如本领域所知的，术语"%相同性"是指两个或多个多肽序列或两个或多个多核苷酸序列之间的关系，如通过序列比较所测定的。在本领域"相同性"还指多肽或多核苷酸序列之间的序列关联性程度，根据具体情况而定，如通过这些序列串之间的匹配所测定的。"相同性"和"相似性"可以容易地通过已知方法计算出，包括但不限于下面所述：Computational Molecular Bioloay (Lesk, A. M. ,编) Oxford University Press, New York (1988); Biocomputing: Informatics and Genome Projects (Smith, D. W. ,编)Academic Press, New York (1993); Computer Analysis of Sequence Data, Part I (Griffin, A. M. 和 Griffin, H. G. ,编) Humana Press, New Jersey (1994); Sequence Analysis in Molecular Biology (von Heinje, G. , 编)Academic Press (1987); 和 Sequence Analysis Primer (Gribskov, M. 和 Devereux, J., 编)Stockton Press, New York (1991)。测定相同性的优选方法能能在测试的序列之间产生最佳匹配。测定相同性和相似性的方法已经编程成公众可获得的计算机程序。可以使用LASERGENE生物信息计算组件(DNASTAR Inc., Madison, WI)的Megalign程序进行序列比对和百分比相同性计算。使用Clustal比对方法 (Higgins 和 Sharp (1989) CABIOS. 5: 151-153)、默认参数(GAP PENALTY=10, GAP LENGTH PENALTY=10)进行了多次序列比对。使用 Clustal 方法进行逐对比对的默认参数是 KTUPLE 1,GAP PENALTY=3, WINDOW=5 和 DIAGONALS SAVED=5。

合适的核酸片段 (本发明的分离的多核苷酸)编码的多肽与本文报道的氨基酸序列有至少约60%的相同、优选地至少约80%的相同。优选的核酸片段编码的氨基酸序列与本文报道的氨基酸序列有至少约 85%的相同。更优选的核酸片段编码的氨基酸序列与本文报道的氨基酸序列有至少约90%的相同。最优选的核酸片段编码的氨基酸序列与本文报道的氨基酸序列有至少约95%的相同。

氨基酸或核苷酸序列的"基本部分"包含足以提供氨基酸或核苷酸序列含有的蛋白或基因的推定鉴别的氨基酸或核苷酸序列。通过本领域的技术人员手工地或通过使用基于计算机的序列对比和鉴别工具，可以评价氨基酸或核苷酸序列，所述工具采用算法例如BLAST (Basic

Local Alignment Search Tool；Altschul 等 (1993) J. Mol. Biol. 215: 403-410；还参见国家健康研究所国家医药图书馆的国家生物技术信息中心（National Center for Biotechnology Information at the National Library of Medicine of the National Institutes of Health）的环球网站点上的对 BLAST 算法的解释)。通常，为了推定地鉴别多肽或核酸序列与已知蛋白或基因的相同性，10 个或更多的连续氨基酸或 30 个或更多的连续核苷酸的序列是必须的。而且，关于核苷酸序列，含有 30 个或更多的连续核苷酸的基因-特异性的寡核苷酸探针可以用于依赖于序列的基因鉴别（例如，DNA 杂交）和分离（例如，细菌菌落或噬菌体噬斑的原位杂交）方法。另外，含有 12 个或更多的核苷酸的短寡核苷酸可以用作 PCR 的扩增引物，以得到含有该引物的特定核酸片段。因此，核苷酸序列的“基本部分”包含能提供含有该序列的核酸片段的特异性的鉴别和/或分离的核苷酸序列。本说明书教导了编码含有一个或多个特定植物蛋白的多肽的氨基酸和核苷酸序列。利用本文报道的序列，本领域的技术人员现在可以使用公开的序列的全部或基本部分用于本领域的技术人员已知的目的。因此，本发明包含所附序列表报道的完整序列以及上面定义的那些序列的基本部分。

“密码子简并性”是指能允许核苷酸序列变化而不影响编码的多肽的氨基酸序列的遗传密码的趋异。因此，本发明涉及含有编码本文所述氨基酸序列的全部或基本部分的核苷酸序列的任何核酸片段。本领域的技术人员熟知特定的宿主细胞在使用核苷酸密码子以指定特定的氨基酸时表现出来的“密码子偏倚”。因此，当合成用于提高宿主细胞中的表达的核酸片段时，需要设计核酸片段，使其密码子的使用频率接近宿主细胞的优选密码子的使用频率。

从使用本领域的技术人员已知的方法化学合成的寡核苷酸结构单元可以装配出“合成的核酸片段”或“合成的基因”。连接和退火这些结构单元以形成更大的核酸片段，然后可以将它们酶切地装配，构建完整的理想的核酸片段。与核酸片段相关的“化学合成的”是指组分核苷酸是体外装配的。使用充分确立的方法可以手工地化学合成核酸片段，或者使用许多市场上可买到的装置之一来自动地化学合成。因此，可以将核酸片段裁剪以进行最佳的基因表达，其基于核苷酸序列

的最优化以反映宿主细胞的密码子偏倚。如果密码子使用偏倚于宿主偏好的密码子，本领域的技术人员能够明白成功的基因表达的可能性。优选密码子的确定可以以调查源自宿主细胞的基因为基础，该基因的序列信息是可获得的。

术语"序列分析软件"是指能用于核苷酸或氨基酸序列分析的任何计算机算法或软件程序。可以从市场上买到或独立地开发出"序列分析软件"。一般的序列分析软件包括但不限于 GCG 程序套(Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, WI), BLASTP, BLASTN, BLASTX (Altschul 等, J. Mol. Biol. 215: 403-410 (1990)和 DNASTAR (DNASTAR, Inc. 1228 S. Park St. Madison, WI 53715 美国)和采用 Smith-Waterman 算法的 FASTA 程序(W. R. Pearson, Comput. Methods Genome Res., [Proc. Int. Symp.] (1994), Meeting Date 1992, 111-20. Editor (s): Suhai, Sandor. Publisher: Plenum, New York, NY)。在本申请的上下文中应当明白，在用序列分析软件进行分析时，除非另有说明，分析的结果是基于使用的程序的"默认值"的。如本文所使用的，"默认值"是指当初次启动时在软件中原始调用的数值或参数的任何设置。

"基因"是指能表达特定蛋白的核酸片段，其包括在编码序列之前(5′非编码序列)和之后(3′非编码序列)的调控序列。"天然基因"是指具有其自己的调控序列的自然存在的基因。"嵌合基因"是指非天然基因的任何基因，其含有非自然地一起存在的调控序列和编码序列。因此，嵌合基因可以含有不同来源或相同来源、但是以不同于自然存在的方式排列的调控序列和编码序列。"嵌合蛋白"是由嵌合基因编码的蛋白。"内源基因"是指在生物的基因组中的自然位置的天然基因。"外源基因"是指在宿主生物中通常不存在的基因，但是它通常经基因转移导入了宿主生物。外源基因可以包含插入到非天然的生物中的天然基因、重组 DNA 构建体或嵌合基因。"转基因"是已经通过转化方法导入基因组中的基因。

"编码序列"是指编码特定的氨基酸序列的核苷酸序列。

"调控序列"和"合适的调控序列"是指位于编码序列的上游(5′非编码序列)、之中或下游(3′非编码序列)的核苷酸序列，其影响相关的编码序列的转录、RNA 加工或稳定性或翻译。调控序列可以包含启动

子、翻译前导序列、内含子和多腺苷酸化识别序列。

“启动子”是指能够控制编码序列或功能性RNA的表达的核苷酸序列。通常，编码序列是位于启动子序列的3’。启动子可以整体地源自天然基因，或者可以由源自不同的自然存在的启动子的不同元件组成，或者可以甚至包含合成的核苷酸片段。本领域的技术人员能够明白，不同的启动子可以指导不同组织或细胞类型、或不同发育阶段，或者作为对不同环境条件的反应的基因表达。能使核酸片段在大多数细胞类型中在大多时间表达的启动子通常称作“组成型启动子”。还认识到，由于在大多数情况下尚未完整地定义调控序列的精确边界，不同长度的核酸片段可以具有相同的启动子活性。

有许多能用于驱动本发明的基因在理想宿主细胞中的表达的启动子，且该启动子是本领域的技术人员所熟悉的。实际上，能驱动这些基因的任何启动子都是适用于本发明的，包括但不限于：CYC1,HIS3,GAL1,GAL10, ADH1,PGK, PHO5, GAPDH, ADC1,TRP1,URA3, LEU2, ENO, TPI（用于糖酵母属中的表达）；AOX1（用于毕赤酵母属（Pichia）中的表达）;和lac, ara, tet, trp, IP_L, IP_R, T7, tac和trc（用于大肠杆菌中的表达），Streptomyces lividins GI，以及amy, apr和npr启动子和用于在芽孢杆菌属中表达的不同噬菌体启动子。

“翻译前导序列”是指位于基因的启动子序列和编码序列之间的核苷酸序列。翻译前导序列位于翻译起始序列的完全加工的mRNA的上游。翻译前导序列会影响mRNA初级转录物的加工、mRNA稳定性或翻译效率。翻译前导序列的实例已经描述在(Turner 和 Foster (1995)Mol. Biotechnol. 3: 225-236)。

“3′非编码序列”是指位于编码序列下游的DNA序列，其包括多腺苷酸化识别序列和编码能影响mRNA加工或基因表达的调节信号的其它序列。多腺苷酸化信号的特征通常是影响多腺苷酸序列段向mRNA前体的3′端的添加。不同的3′非编码序列的使用由Ingelbrecht等((1989) Plant Cell 1: 671-680) 示例。

“RNA转录物”是指RNA聚合酶-催化的DNA序列的转录产物。当RNA转录物是DNA序列的完全互补拷贝时，它称作初级转录物，或者它可以是源自初级转录物的转录后加工的RNA序列，此时它称作成熟的RNA。“信使RNA (mRNA)”是指不含有内含子且可以被细胞翻译成多肽

的 RNA。“cDNA”是指与 mRNA 模板互补且来自它的 DNA。cDNA 可以是单链形式，或者使用例如 DNA 聚合酶 I 的 Klenow 片段转化成双链形式。“有义-RNA”是指包括 mRNA 且因而可以由细胞翻译成多肽的 RNA 转录物。“反义 RNA” 是指与全部或部分目标初级转录物或 mRNA 互补的 RNA 转录物，其能阻止目标基因的表达(见，美国专利号 5,107,065，本文中引作参考)。反义 RNA 的互补性可以是特定核苷酸序列的任何部分的，即在 5′非编码序列、3′非编码序列、内含子或编码序列。“功能性的 RNA”是指有义 RNA、反义 RNA、核酶 RNA 或其它的不能被翻译但是仍然对细胞过程有影响的 RNA。

术语“可操作地连接”是指位于一个多核苷酸上且彼此相连的 2 个或更多的核酸片段，所以一个的功能会影响其它的功能。例如，当它能影响编码序列的表达时，该启动子是可操作地连接到该编码序列(即，该编码序列是在该启动子的转录控制下)。编码序列可以以有义的或反义的方向可操作地连接到调控序列。

如本文所使用的，术语“表达”是指源自本发明的核酸片段的有义(mRNA)或反义 RNA 的转录和稳定积累。表达还可以指 mRNA 向多肽的翻译。“反义抑制”是指能抑制目标蛋白的表达的反义 RNA 转录物的产生。“超表达”是指在转基因生物中的基因产物的产生超过了在正常的或非转化的生物中的产生水平。“共抑制”是指能抑制相同的或基本上类似的外源或内源基因的表达的有义 RNA 转录物的产生(美国专利号 5,231,020，本文中引作参考)。

“蛋白”或“多肽”是以编码多肽的多核苷酸中的编码序列决定的特定顺序排列的氨基酸链。每一种蛋白或多肽具有独特的功能。

“信号序列”是指编码信号肽的核苷酸序列。

“转化”是指核酸片段向宿主生物或宿主生物的基因组中的转移，其导致遗传上稳定的遗传。含有转化的核酸片段的宿主生物称作“重组的”、“转基因的”或“转化的”生物。因此，本发明的分离的多核苷酸可以整合进能导入宿主细胞并在其中复制的重组构建体中，一般是 DNA 构建体。这样的构建体可以是包括复制系统和能在给定的宿主细胞中转录和翻译编码多肽的序列的序列的载体。一般地，表达载体包括例如一个或多个在 5′和 3′调控序列和选择性标记转录控制下的克隆基因。这样的载体还可以含有启动子调节区(例如，控制诱导型的或

组成型的、环境地或发育地调节的或位置特异性的表达的调节区)、转录起始位点、核糖体结合位点、转录终止位点和/或多腺苷酸化信号。

术语“宿主细胞”或“宿主生物”是指能接受外源基因或异源基因或内源基因的多个拷贝并表达这些基因以生成有活性的基因产物的微生物。

术语“DNA 构建体”或“构建体”是指人工构建的 DNA 片段。这样的构建体可以单独使用或者与载体结合使用。

术语“质粒”、“载体”和“盒”是指通常携带基因的染色体外的元件，所述基因不是细胞中心代谢的部分，其通常是环状双链 DNA 分子的形式。这样的元件可以是自主复制序列、整合入基因组的序列、噬菌体或任何来源的单链的或双链的 DNA 或 RNA 的线性的或环状的核苷酸序列，其中许多核苷酸序列已经插入或重组进独特的构建体中，该构建体能将选定的基因产物的启动子片段和 DNA 序列与合适的 3′非翻译序列一起导入细胞中。“转化盒”是指含有外源基因的特定载体，除了外源基因以外，其还含有能促进特定宿主细胞的转化的元件。“表达盒” 是指含有外源基因的特定载体，除了外源基因以外，其还含有能增强该基因在外源宿主中的表达的元件。

术语“编码”是指基因通过转录和翻译机制生产氨基酸序列的过程。编码特定氨基酸序列的过程包括这样的 DNA 序列，其可以含有不造成编码的氨基酸变化的碱基变化，或者其含有可以改变一个或多个氨基酸、但是不影响该 DNA 序列编码的蛋白的功能性质的碱基变化。因此应当明白，本发明不只包括具体的示例性序列。

“PCR”或“聚合酶链反应”是本领域的技术人员熟知的用于扩增特定 DNA 片段的技术(美国专利号 4,683, 195 和 4,800, 159)。

“ORF”或“可读框”是编码肽或蛋白的 DNA 分子中的核苷酸序列。在已经确定了 DNA 片段的序列、但是不知道编码的蛋白的功能时，经常使用该术语。

术语“可发酵的碳底物”是指能被本发明的宿主生物代谢的碳源，更具体地是选自单糖、寡糖、多糖和一碳底物或其混合物的那些碳源。

同系物的分离

本发明的核酸片段可以用于从相同的或其它微生物物种中分离编

码同源蛋白的基因。使用序列依赖性的方法分离同源基因是本领域熟知的。序列依赖性的方法的例子包括但不限于核酸杂交方法、DNA 和 RNA 扩增方法，如通过核酸扩增技术的各种应用所示例的(例如，聚合酶链反应(PCR)，Mullis 等，美国专利 4,683, 202)，连接酶链反应(LCR)，Tabor 等，Proc. Acad. Sci. USA 82, 1074, (1985))或链置换扩增(SDA, Walker 等，Proc. Natl. Acad. Sci. U. S. A., 89, 392, (1992))。

一般地，在 PCR-类扩增技术中，引物具有不同的序列，且彼此不互补。根据所需的实验条件，引物序列应该设计成能有效地和正确地复制目标核酸。PCR 引物设计的方法是常见的和本领域熟知的 (Thein 和 Wallace,"The use of oligonucleotide as specific hybridization probes in the Diagnosis of Genetic Disorders", in Human Genetic Diseases: A Practical Approach, K. E. Davis 编，(1986) 第 33-50 页 IRL PRESS, Herndon, Virginia); Rychlik, W. (1993) In White, B. A. (ed.), Methods in Molecular Biology, Vol. 15, 第 31-39 页，PCR Protocols: Current Methods and Applications. Humania Press, Inc., Totowa, NJ.)。

杂交方法被充分地详细说明了。一般地，必须将探针和样品在能允许核酸杂交的条件下混合。这包括在有无机盐或有机盐存在的情况下、在合适的浓度和温度条件下使探针与样品接触。探针和样品核酸必须接触足够长的时间，使探针和样品核酸之间能够发生任何可能的杂交。混合物中的探针或目标的浓度会决定发生杂交所需的时间。探针或目标的浓度越高，所需的杂交温育时间越短。

可以采用各种杂交溶液。一般地，它们包括约 20 ~60%体积、优选地 30%的极性有机溶剂。常用的杂交溶液含有约 30-50%v/v 甲酰胺、约 0.15 ~ 1M 氯化钠、约 0.05 ~ 0.1 M 缓冲液例如柠檬酸钠、Tris-HCl、PIPES 或 HEPES (pH 范围约 6-9)、约 0.05~0.2%去污剂例如十二烷基硫酸钠或 0.5-20 mM 之间的 EDTA、FICOLL (Pharmacia Inc.)(约 300-500 千道尔顿)聚乙烯吡咯烷酮(约 250-500 千道尔顿)和血清清蛋白。一般的杂交溶液还包括约 0.1 ~ 5 mg/mL 未标记的载体核酸，片段化的核 DNA(例如小牛胸腺或鲑鱼精子 DNA)或酵母 RNA 和任选地约 0.5 ~ 2%重量/体积的甘氨酸。还可以包括其它的添加剂，例如体积排

阻剂，其包括多种极性的水溶性的或可膨胀的试剂例如聚乙二醇，阴离子聚合物例如聚丙烯酸酯或聚丙烯酸甲酯，和阴离子糖类聚合物例如葡聚糖硫酸酯。

重组表达—微生物

本发明的序列的基因和基因产物可以导入到微生物宿主细胞中。用于表达本发明的基因和核酸分子的优选宿主细胞是微生物宿主，其可以广泛发现于真菌或细菌家族中，且可以在广范围的温度、pH 值和溶剂耐受性生长。在光合的或化能自养的宿主的情况下，大规模的微生物生长和功能基因的表达可以使用广范围的简单的或复杂的碳水化合物、有机酸和醇、饱和烃（例如甲烷）或二氧化碳。但是，功能基因可以由特定的生长条件所调节、抑制或降低，所述条件可以包括氮、亚磷、硫、氧、碳或任何痕量微量营养物（包括小无机离子）的形式和量。另外，功能基因的调节可以通过存在或不存在特定的调节分子来实现，所述调节分子是加入到培养物中但是一般不认为是营养物或能源。生长速率也可以是基因表达的重要调节因子。合适的宿主菌株的示例包括但不限于真菌或酵母物种例如曲霉属、木霉菌属（Trichoderma）、糖酵母属、毕赤酵母属、假丝酵母属（Candida）、汉逊酵母属（Hansenula），或细菌物种例如多变细菌（Proteotacteria）和放线菌以及特定的红球菌属（Rhodococcus）、不动杆菌属（Acinetobacter）、节杆菌属（Arthrobacter）、短杆菌属（Brevitacterium）、食酸菌属(Acidovorax)、芽孢杆菌属、链霉菌属(Streptomyces)、埃希氏菌属、沙门氏菌属(Salmonella)、假单胞菌属和棒杆菌属(Corynebacterium)的成员。

大肠杆菌特别适用于用作本发明的发酵工艺中的宿主微生物。大肠杆菌不能代谢含有 α(1,6)键的寡糖，也难以代谢 DP > 7 的任何寡糖。

本领域的技术人员熟知含有能指导外源蛋白的高水平表达的调控序列的微生物表达系统和表达载体。它们中的任意一个都可以用于构建嵌合基因，以生成本发明的任何基因产物。然后可以通过转化技术将这些嵌合基因导入合适的微生物中，以高水平地表达酶。

用于转化合适的宿主细胞的载体或盒是本领域熟知的。一般地，载体或盒含有能指导相关基因、选择性标记和使得能进行自主复制或染色体整合的序列转录和翻译的序列。合适的载体包含含有转录起始控

制的基因的 5′区域和控制转录终止的 DNA 片段的 3′区域。当两个控制区都源自与转化的宿主细胞同源的基因时是最优选的，尽管能够明白，这样的控制区不需要源自对于选择的作为生产宿主的特定物种而言是天然的基因。

起始控制区或启动子用于驱动基因产物的表达。终止控制区也可以源自对于优选的宿主而言是天然的基因。任选地，终止位点可以是非必须的，但是，如果包括则是最优选的。

对于一些应用，可以有益地指导本发明的蛋白进入不同的细胞隔室。因而可以想象，通过改变编码序列使其编码具有合适的胞内导向序列（例如转运序列）的酶，可以进一步补充上述的嵌合基因。

具有提高的活性的酶

可以预期，本发明的序列可以用于生产具有提高的或改变的活性的基因产物。已知用于突变天然基因序列、生产具有改变的或提高的活性的基因产物的多种方法，包括但不限于易错的 PCR (Melnikov 等, Nucleic Acids Research, (Feb. 15,1999) Vol. 27, No. 4, 第 1056-1062 页)；定点诱变(Coombs 等, Proteins (1998), 259-311,1 plate. Editor (s): Angeletti, Ruth Hogue. Publisher: Academic, San Diego, CA)和"基因改组（gene shuffling）" (US 5,605, 793; US 5,811,238; US 5,830, 721;和 US 5,837, 458, 本文中引作参考)。

途径调控

关于本发明的基因序列的知识将用于操作具有这样的途径的任何生物中的糖代谢途径。操纵遗传途径的方法是常见的，也是本领域熟知的。特定途径中的选定的基因可以通过多种方法上调或下调。另外，竞争途径的生物可以通过基因破坏和类似技术消除或升华。

一旦已经鉴别出了关键的遗传途径并进行了测序，就可以上调具体基因以增加该途径的输出。例如，可以将目标基因的其它拷贝在多拷贝质粒例如 pBR322 上导入到宿主细胞中。或者可以修饰目标基因，使其在非天然的启动子的控制下。当需要使途径在细胞周期的特定点或在发酵过程中运行时，可以用受调节的或诱导型启动子替换目标基因的天然启动子。类似地，在一些情况下，可以修饰天然的或内源的启动子来提高基因的表达。例如，可以通过突变、缺失和/或替代来体内地改变内源启动子(见，Kmiec，美国专利 5,565, 350; Zarling 等,

PCT/US93/03868)。

在本发明的上下文中，可以通过许多熟知方法（例如， 反义、辐射-或化学-诱导的突变、基因改组等）中的任一种来有益地调控糖代谢途径的表达。例如，本发明提供了许多基因，它们编码糖代谢途径中的关键酶、从而生产简单的糖。分离的基因包括 α-葡糖苷酶和异麦芽糖酶基因。当例如需要积累葡萄糖或麦芽糖时，上述方法中的任一种都可以用于超表达本发明的 α-葡糖苷酶和异麦芽糖酶基因。类似地，可以通过上述的任一种方法破坏下游基因（例如糖酵解途径中的基因）来影响葡萄糖或麦芽糖的生物合成基因的积累。

生物发酵

本发明适用于多种生物发酵方法，特别是适用于大规模工业方法的那些。本发明可以使用分批、补料-分批或连续工艺来实现，但是优选地以补料-分批模式来实现。这些生物发酵方法是常见的，且是本领域熟知的(Brock, T. D.; Biotechnology: A Textbook of Industrial Microbiology, 第 2 版; Sinauer Associates: Sunderland, MA (1989); 或 Deshpande, Appl. Biochem. Biotechnol. 36: 227 (1992))。

“生物发酵系统”或“生物发酵”是指使用生物催化剂来催化底物和产物之间的反应的系统。

生物催化剂

生物催化剂能启动或改变底物和产物之间的化学反应的速率。生物催化剂可以是整个的生物或者是分离的酶催化剂的形式。整个的微生物细胞可以不经任何预处理（例如透化）地用作生物催化剂。或者，可以通过本领域的技术人员熟悉的方法使整个细胞透化(例如，用甲苯、去污剂或冻-融处理)，以提高材料扩散进出细胞的速率。

用于本发明的微生物可以包括但不限于：细菌(例如肠细菌埃希氏菌属和沙门氏菌属，例如，以及芽孢杆菌属、不动杆菌属、链霉菌属、甲基细菌属（Methylobacter）、红球菌属和假单胞菌属)；蓝细菌(例如红细菌属（Rhodobacter）和集胞蓝细菌属（synechocystis)；酵母(例如糖酵母属、接合酵母属（Zygosaccharomyces）、克鲁维酵母属（Klayveromyces）、假丝酵母属、汉逊酵母属、德巴利酵母属（Debaryomyces）、毛霉菌属（Mucor）、毕赤酵母属和球拟酵母属

(Torulopsis)); 丝状真菌(例如曲霉属和节丛孢属(Arthrobotrys)); 和藻类。为了本申请的目的,“微生物”还包括来自昆虫、动物或植物的细胞。

培养条件

用于微生物培养物的维持和生长的材料和方法是微生物学或生物发酵科学领域技术人员熟知的(Bailey 和 Ollis, Biochemical Engineering Fundamentals, 第 2 版; McGraw-Hill: NY (1986))。根据微生物进行理想的功能基因的表达的具体需要,必须考虑合适的培养基、pH、温度和对需氧、微好氧或厌氧条件的要求。

培养基和碳底物

用于本发明的生物发酵培养基(液体培养基或溶液)必须含有合适的碳底物,其根据生物催化剂的需要进行选择。合适的底物可以包括但不限于: 单糖(例如葡萄糖和果糖), 二糖(例如乳糖或蔗糖), 寡糖和多糖(例如淀粉或纤维素或其混合物)或来自可再生的原料的未纯化的混合物 (例如干酪乳清渗透物、玉米浆、甜菜糖蜜和大麦芽)。碳底物还可以是一碳底物(例如二氧化碳、甲醇或甲烷)。

除了合适的碳源外,生物发酵培养基必须含有本领域的技术人员已知的合适的矿物质、盐、微生物、辅因子、缓冲剂和其它组分(Bailey 和 Ollis, Biochemical Engineering Fundamentals, 第 2 版; 第 383-384 和 620-622 页; McGraw-Hill: New York (1986))。这些补充物必须适合于生物催化剂的生成,并促进生成生物发酵目标产物所必须的酶途径。

最后,能表达工业上有用的产物的功能基因可以由特定的生长条件(例如氮、亚磷、硫、氧、碳或任何痕量微量营养物(包括小无机离子)的形式和量)所调节、抑制或降低。功能基因的调节可以通过存在或不存在特定的调节分子(例如安慰诱导物)来实现,所述调节分子是加入到培养物中但是一般地不认为是营养物或能源。生长速率也可以是基因表达的重要调节因子。

实施例

在下面的实施例中进一步定义了本发明。应当明白,尽管这些实施例显示了本发明的优选实施方案,但它们仅仅用于解释目的。从上面

的讨论和这些实施例，本领域的技术人员可以明白本发明的基本特征，并且可以在不背离其精神和范围的情况下作出本发明的各种变化和修饰，使其适用于各种用途和条件。

缩写的含义如下："h"代表小时，"min"代表分钟，"sec"代表秒，"d"代表天，"mL"代表毫升，"L"代表升，"mM"代表毫摩尔，"nm"代表纳米，"g"代表克，且"kg"代表千克，"HPLC"代表高效液相层析，"RI"代表折射率。

一般方法

适用于细菌培养物的维持和生长的材料和方法是本领域熟知的。适用于下面的实施例的技术可以参见：Manual of Methods for General Bacteriology；Phillipp Gerhardt, R. G. E. Murray, Ralph N. Costilow, Eugene W. Nester, Willis A. Wood, Noel R. Krieg 和 G. Briggs Philips, 编, American Society for Microbiology: Washington, D. C. (1994) 或 Biotechnology: A Textbook of Industrial Microbiology；Brock, T. D., 第 2 版; Sinauer Associates: Sunderland, MA (1989)。

通过 HPLC 监控甘油向 1,3-丙二醇的转变。使用层析领域的技术人员可获得的标准技术和物质进行分析。一种合适的方法使用 Waters Maxima 820 HPLC 系统，其采用 UV (210 nm) 和 RI 检测。将样品注射到 Shodex SH-1011 柱 (8 mm x 300 mm, 购自 Waters, Milford, MA)，其配备有 Shodex SH-1011P 前置柱 (6 mm x 50 mm)，温度控制在 50 ℃，使用 0.01 N H_2SO_4 作为流动相，流速 0.5 mL/分钟。当需要定量分析时，使用已知量的三甲基乙酸作为外标物制备样品。一般地，葡萄糖 (RI 检测)、甘油、1,3-丙二醇 (RI 检测) 和三甲基乙酸 (UV 和 RI 检测) 的保留时间分别是 15.27 分钟、20.67 分钟、26.08 分钟和 35.03 分钟。

实施例 1 短双岐杆菌 ATCC 15700 的基因组测序

短双岐杆菌 (ATCC 15700) 购自美国典型培养物保藏中心 (American Type Culture Collection), P. O. Box 1549, Manassas, VA 20108, 美国。得到细胞沉淀并悬浮在含有 10 mM Na-EDTA 和 50 mM Tris-HCl、pH 7.5 的溶液中。根据标准方法从短双岐杆菌 (ATCC 15700)

中分离出基因组 DNA。根据公开的方法制备基因组 DNA 和文库构建(Fraser 等, Science 270 (5235): 397-403 (1995))。

基因组 DNA 的制备: 悬浮后，在 0.2%肌氨酸、20 mM β-巯基乙醇和 150 单位/mL 溶细胞酶中温和地裂解细胞，并在 37 ℃温育 30 分钟。用 Tris-平衡的苯酚提取 DNA 两次，再用氯仿提取两次。使 DNA 在 70%乙醇中沉淀，再悬浮到含有 1 mM Na-EDTA 和 10 mM Tris-HCl、pH 7.5 的溶液中。用核糖核酸酶的混合物处理 DNA 溶液，然后用 Tris-平衡的苯酚提取两次，再用氯仿提取两次。然后在乙醇中沉淀，并悬浮到 1 mM Na-EDTA 和 10 mM Tris-HCl、pH 7.5 中。

文库构建: 将 50~100 μg 染色体 DNA 悬浮到含有 30%甘油、300 mM 醋酸钠、1 mM Na-EDTA,和 10 mM Tris-HCl、pH 7.5 的溶液中，在 Aeromist Downdraft Nebulizer 室(IBI Medical products, Chicago, IL)中在 12 psi 剪切 60 秒。沉淀、悬浮 DNA，并用 BAL-31 核酸酶处理。在低熔点的琼脂糖凝胶上进行大小分级后，将一个级分(2.0 kb 或 5.0 kb)切离、清洁，并使用 T4 DNA 连接酶(New England Biolabs, Inc., Beverly, MA)连接到 pUC18 (Amersham Biosciences)的磷酸酶处理过的 SmaI 位点。将连接混合物上凝胶，切离代表了载体+一个插入的连接产物的 DNA 带，用 T4 DNA 聚合酶 (New England Biolabs)处理，然后重新连接。用该两步连接方法生产高效价的文库，其具有超过 99%的单一插入物。

测序: 采用鸟枪测序方法进行整个微生物基因组的测序(Fleischmann, R. 等, Science 269 (5223): 496-512 (1995))。组合使用载体和插入物-特异性的引物，使用染料终止子技术(U. S. 5,366,860; EP 272, 007)，在 ABI 自动测序仪(Applied Biosystems, Foster City, CA)上进行测序。在 DNAStar (DNA Star Inc., Madison, WI)或 Wisconsin GCG 程序(Wisconsin Package 9.0 版, Genetics Computer Group (GCG), Madison, WI)和 CONSED 包(7.0 版)中进行序列编辑。所有序列代表在两个方向上至少 2 倍的覆盖度。使用 Phred/Phrap 软件包(版本为 0.961028. m/ 0.990319)进行序列装配。

实施例 2 碳水化合物降解基因的鉴别

通过进行 BLAST (Basic Local Alignment Search Tool ; Altschul 等, J. Mol. Biol. 215: 403-410 (1993); 还参见, www.ncbi.nlm.nih.gov/ BLAST/)搜索与包含在 BLAST "nr"数据文库中的序列(含有全部非多余的 GenBank CDS 翻译, 源自 3-维结构 Brookhaven Protein Data Bank, SWISS-PROT 蛋白序列数据文库、EMBL 和 DDBJ 数据文库的序列)的相似性, 鉴别出了编码异淀粉酶活性的基因。使用国家生物技术信息中心(NCBI)提供的 BLASTN 算法, 分析得到的序列与所有可公开获得的包含在"nr"数据文库中的 DNA 序列的相似性。将 DNA 序列翻译成所有可读框, 并使用 NCBI 提供的 BLASTP 算法(Gish 和 States, Nature Genetics 3: 266-272 (1993)), 对比与所有可公开获得的包含在"nr"数据文库中的蛋白序列的相似性。

使用 BLASTN 或 BLASTP 算法进行所有对比。BLAST 对比的结果如表 1 所示, 其总结了它们具有最大相似性的序列。表 1 显示了基于 BLASTP 算法的数据, 其含有以期望值表示的值。期望值(E-值)是不同比对的数目, 所述比对的分数等同于或高于在随机数据文库搜索中预期得到的特定分数 S。E-值越小, 分数的显著性越大。

表1

克隆名称	鉴别出的相似性	SEQ ID	SEQ ID	% 相同性[a]	% 相似性[b]	E-值[c]	引用
mbc1g.pk007.h12 短双岐杆菌	(AF411186) α-葡糖苷酶[青春双岐杆菌 Bifido bacterium adolescentis]	1	2	61	70	0.0	Van den Broek, L.A.M. et al. "Cloning and characterization of two alpha-glucosidases from *Bifidobacterium adolescentis*" NCBI database
Mbc2g.pk018.j20 短双岐杆菌	(AF358444) α-葡糖苷酶[青春双岐杆菌]	3	4	73	84	0.0	Van den Broek, L.A.M. et al. "Cloning and characterization of two alpha-glucosidases from *Bifidobacterium adolescentis*" NCBI database
mbc1g.pk026.k1 短双岐杆菌	(AF358444) α-葡糖苷酶[青春双岐杆菌]	5	6	69	82	0.0	Van den Broek, L.A.M. et al. "Cloning and characterization of two alpha-glucosidases from *Bifidobacterium adolescentis*" NCBI database
DexB 变异链球菌	(M7351) 葡聚糖葡糖苷酶[变异链球菌]	16	17	100	100	0.0	Russell, R.R. and Ferretti, J.J. "Nucleotide sequence of the dextran glucosidase (*dexB*) gene of *Streptococcus mutans*" *J. Gen. Microbiol.* 136 (Pt 5), 803-810 (1990)

[a]% 相同性：定义为两种蛋白之间相同的氨基酸的百分比。

[b]% 相似性：定义为两种蛋白之间相同的或保守的氨基酸的百分比。

[c] 期望值。该期望值估计了与给定分数的匹配的统计显著性，说明了匹配的数目，该分数是在绝对随机地搜索该大小的数据文库时所预期的。

实施例 3 含有变异链球菌 dexB 基因的大肠杆菌中的胞内异淀粉酶的活性

为了克隆dexB基因,使用 Jagusztyn 等 (J. Gen. Microbiol. 128: 1135- 1145 (1982))所述方法从变异链球菌(ATCC 25175D)分离出基因组 DNA。

在变异链球菌(dexB)DNA 序列 (Ferretti 等, Infection and Immunity 56: 1585-1588 (1988))的基础上设计了寡核苷酸引物(SEQ ID NO: 7 和 SEQ ID NO: 8),且包括 BamHI 和 SalI 限制位点。使用包含在 HotStartTaq™ 试剂盒(Qiagen, Valencia, CA)中的标准 PCR 方法扩增 dexB 基因。反应物含有 1 ng 基因组 DNA 和 1μm 每种引物。用 BamHI 和 SalI 酶消化得到的 1.6 kb DNA 片段。将消化的片段直接克隆进质粒 pTRC99a(amp^R) (Amersham-Pharmacia, Amersham, 英国),得到与 LacZ 基因的翻译融合体。该命名为 pTRC99-dexB 的质粒还含有 LacZ 基因的前 10 个氨基酸的编码序列,其在表达后融合到天然 DexB 蛋白的 N-末端。使用生产商的方法(Invitrogen, Carlsbad, CA),将 pTRC99-dexB 质粒转化进大肠杆菌 DH5α细胞中,并涂布到含有 100μg/mL 氨苄青霉素的 Luria Broth(LB)培养基上。

在大肠杆菌中表达后,从粗蛋白提取物估计异淀粉酶活性。在 LB 培养基中过夜培养大肠杆菌 DH5α/pTRC99-dexB 的单一菌落,然后在新鲜的 LB 培养基(3.0 mL)中以 1: 100 稀释,在 37℃再培养 2 小时。该培养后,通过加入异丙基β-D-1-硫代吡喃半乳糖苷(IPTG)至 1 mM 的终浓度诱导 DexB 基因。另外培养 2 小时后,从诱导的细胞提取粗蛋白。为了分离粗蛋白提取物,通过离心(1 x 8000 g)收集细胞,然后悬浮到 0.5 mL 磷酸盐缓冲液(10 mM, pH 6.8)中。超声处理悬浮液,以释放出总细胞蛋白,离心(1 x 14,000 g)除去细胞碎片。通过与异麦芽糖、或分开地与潘糖在 10 mM 磷酸盐缓冲液(pH 6.8)中在 37 ℃温育 2 小时,检测上清液中的总蛋白的异淀粉酶活性。通过高效阴离子交换层析(HPAEC)表征反应产物。

为了进行 HPAEC,以下述方式制备和分析样品。与异麦芽糖或潘糖温育 2 小时后,通过 0.22μM Spin-X (R)离心管过滤器(Costar, Corning, NY)过滤总蛋白提取物,并用无菌过滤水稀释。使用 PA10 柱以 100 mM 氢氧化钠作为洗脱剂和 0-150 mM 醋酸钠线性梯度,通过

HPAEC (Dionex, Sunnyvale, CA)分析样品。使用 pTRC99-dexB 细胞提取物降解异麦芽糖的结果如表 2 所示。通过与 pTRC99-dexB 细胞提取物温育降解潘糖和形成的产物如表 3 所示。

表 2 DexB 粗蛋白提取物与异麦芽糖的活性(250μg/mL)

细胞系	异麦芽糖(μg/mL)
DH5α/pTRC99a (阴性对照)	256
DH5α/pTRC99-dexB	ND

ND = 未检出

表 3 DexB 粗蛋白提取物与潘糖的活性(150μg/mL)

细胞系	潘糖 (μg/mL)	麦芽糖 (μg/mL)	异麦芽糖 (μg/mL)	葡萄糖 (μg/mL)
DH5α/pTRC99a (阴性对照)	122	ND	ND	ND
DH5α/pTRC99-dexB	ND	74	8	82

ND = 未检出

实施例 4 短双岐杆菌异淀粉分解基因在大肠杆菌中的表达

将来自短双岐杆菌(ATCC 15700)文库的几个可读框鉴别为推定的备选基因，其具有针对 α(1,6)-连接的葡萄糖寡糖的活性(实施例 2)。使用含有 α(1,6)-连接的葡萄糖的寡糖,选择了 3 个推定的克隆 mbc1g.pk007.h12 (h12)、mbc1g.pk026.k1 (k1)和 mbc2g.pk018.j20 (j20)来详细表征异淀粉分解活性。

将含有克隆的在来自实施例 1 的 pUC18 中的推定的异淀粉分解双岐杆菌属基因的全长编码序列的大肠杆菌 DH5α 菌株接种到 LB 培养基中，在 37℃培养。20 小时后，将培养物在新鲜 LB 培养基中稀释(1:100)，并在 37 ℃另外培养 3-4 小时。如实施例 3 所述从细胞中制备总蛋白提取物。通过与异麦芽糖、或分开地与潘糖在 10 mM 磷酸盐缓冲液(pH 6.8)中在 37 ℃温育 2 小时，检测上清液中的总蛋白的异淀粉分解活性。如实施例 3 所述，制备样品并通过高效阴离子交换层析

(HPAEC)表征反应产物。结果表明，从克隆 h12、k1 和 j20 生成的酶将异麦芽糖降解成了葡萄糖(表 4)。

表 4 短双歧杆菌粗提取物与异麦芽糖的活性(150μg/mL)

细胞系	异麦芽糖	葡萄糖
DH5α/pUC18(阴性对照)	107	37
DH5α-h12	6	187
DH5α-k1	5	165
DH5α-j20	8	154

ND = 未测出

将总蛋白提取物与潘糖(250μg/mL)温育 2 小时，然后通过 0.22μM Spin-X (R)离心管过滤器(Costar, Corning, NY)过滤。如实施例 3 所述通过 HPAEC 分析样品。温育后的潘糖的消失表明，从克隆 h12、k1 和 j20 生成的酶能降解潘糖。图 1 表明，克隆 h12 能将潘糖彻底降解为葡萄糖（还显示了阴性对照，大肠杆菌 DH5α中的质粒 pUC18）。图 1 还表明，来自 k1 和 j20 克隆的酶能将潘糖降解成葡萄糖和麦芽糖。

实施例 5 天然短双歧杆菌 j20 异淀粉酶基因在大肠杆菌中的表达

天然短双歧杆菌基因 j20 (从实施例 1 得到)似乎在成熟编码序列的 NH-端具有信号肽(通过 pSort 预测软件检测; Nakai 和 Kanehisa, Expert, PROTEINS: Structure, Function and Genetics 11: 95-110 (1991))。编码α(1,6)-连接的葡萄糖寡糖水解活性的短双歧杆菌 j20 基因的核酸和氨基酸序列分别是 SEQ ID NO: 30 和 SEQ ID NO: 31。

因此，使用完整的全细胞测试了异麦芽糖的代谢。这是通过将能表达 j20基因的大肠杆菌DH5α细胞的单一菌落在含有异麦芽糖(500μg/mL)的 LB 培养基中、在 37 ℃培养 24 小时来实现的。培养后，从培养基中除去细胞，并通过实施例 3 所述 HPAEC 方法制备和分析培养基。通过与阴性对照(只含有原始 pUC18 质粒的大肠杆菌 DH5α细胞)相比异麦芽糖降低的水平，证实了在能表达短双歧杆菌 j20 基因的细胞中存在胞外的异淀粉酶活性。表 5 中的结果表明，能表达天然 j20 基因的大肠杆菌细胞降解了胞外供给的异麦芽糖。

表 5 天然的 j20 基因代谢的异麦芽糖

细胞系	异麦芽糖	葡萄糖
DH5α/pUC18(阴性对照)	508	26
DH5α-j20	180	22

实施例 6 变异链球菌 dexB 和短双歧杆菌异淀粉酶的胞外导向

因为短双歧杆菌 k1 和变异链球菌 dexB 基因似乎不含有天然的信号肽(pSort 预测软件; Nakai 和 Kanehisa, Expert, PROTEINS: Structure, Function and Genetics 11: 95-110 (1991)), 通过 PCR 方法将成熟的编码序列在翻译融合体中连接到信号肽上,以进行胞外表达。

在从质粒 pBE505 (Borchert 和 Nagarajan, J. Bacteriol. 173: 276-282 (1991)) 和 pBE311 (Nagarajan 和 Borchert, Res. Microbiol. 142: 787-792 (1991))开始的一系列步骤中, 构建了含有枯草芽孢芽孢杆菌碱性和中性蛋白酶基因的模块表达载体。用限制酶 KpnI 和 NruI 消化质粒。分离得到的来自 pBE505 的 969 个碱基对的 KpnI-NruI 片段, 并连接到来自 pBE311 的 7.2kb 的大 KpnI-NruI 片段中, 生成 pBE559。

然后用限制酶 KpnI 和 EcoRV 消化质粒 pBE559 和 pBE597 (Chen 和 Nagarajan, J. Bacteriol. 175: 5697-5700 (1993))。将来自 pBE559 的 941 个碱基对的 KpnI-EcoRV 片段连接到来自 pBE597 的 8.9 kb 的 KpnI-EcoRV 片段, 生成质粒 pBE592。

将质粒 pBE26 (Ribbe 和 Nagarajan, Mol. Gen. Genet. 235: 333-339 (1992))用作扩增解淀粉芽孢杆菌碱性蛋白酶(apr)启动子区域的模板, 其中使用实施例 3 所述 PCR 方法。设计并合成了寡核苷酸引物 SEQ ID NO: 9, 以在碱性蛋白酶信号断裂位点导入 NheI 限制位点, 并紧邻该断裂位点在下游导入 EcoRV 限制位点。寡核苷酸引物 SEQ ID NO: 10 用于退火到 pBE26 中的 apr 启动子区域的上游的 5′多接头区。使用所述引物和质粒 pBE26 模板 DNA 进行 PCR 反应。用 KpnI 和 EcoRV 消化得到的 1.2 kb 的 PCR 产物, 并连接到来自 pBE592 的大 KpnI-EcoRV 片段, 生成质粒 pBE92。

将质粒 pBE80 (Nagarajan 等,Gene 114: 121-126 (1992))用作扩增解淀粉芽孢杆菌中性蛋白酶 (npr)启动子区域的模板，其中使用实施例 3 所述 PCR 方法。设计并合成了下游引物 SEQ ID NO: 11，以在中性蛋白酶信号断裂位点导入 NheI 限制位点，并紧邻该断裂切位点在下游导入 EcoRV 限制位点。引物 SEQ ID NO: 12 用于退火到 pBE80 中的 Npr 启动子的 5′区。使用所述引物和 DNA 模板进行 PCR 反应。用 KpnI 和 EcoRV 酶促地消化得到的 350 个碱基对的 PCR 产物，并连接到来自 pBE592 的大 KpnI-EcoRV 片段，生成质粒 pBE93。

使用表 6 所述寡核苷酸引物,将 k1 和 dexB 基因翻译地融合到在载体 pBE92 和 pBE93 中的枯草芽孢杆菌碱性和中性蛋白酶基因的信号肽。通过实施例 3 所述方法，使用分别来自短双岐杆菌(ATCC 15700)或 pTRC99-dexB 质粒的基因组 DNA 作为模板，进行 PCR 扩增。

用工程地加入了 NheI 和 BamHI 位点的寡核苷酸引物 SEQ ID NO: 14 和 SEQ ID NO: 15 扩增 1.8 kb 的 k1 基因 DNA 片段。含有 NheI 和 SalI 限制酶位点的寡核苷酸引物 SEQ ID NO: 13 和 SEQ ID NO: 8 扩增了 1.6 kb 的 dexB 基因 DNA 片段。用合适的酶消化该片段，并克隆进模块载体 pBE92 和 pBE93 中。

得到的质粒(分别命名为 pBE92-dexB、pBE93-dexB、pBE92-k1 和 pBE93-k1)含有翻译地融合到信号肽的天然酶，所以该信号肽断裂位点(Ala Ser Ala)是保守的。连接到短双岐杆菌 k1 基因的枯草芽孢杆菌中性蛋白酶信号肽的核酸和氨基酸序列分别是 SEQ ID NO: 40 和 41。连接到变异链球菌 dexB 基因的枯草芽孢杆菌中性蛋白酶信号肽的核酸和氨基酸序列分别是 SEQ ID NO: 42 和 43。使用生产商的方法(Invitrogen, Carlsbad, CA)，将质粒转化进大肠杆菌 DH5α 细胞中，并涂布到含有氨苄青霉素(100μg/mL)的 Luria Broth(LB)培养基上。

通过接种 3.0 mL 含有氨苄青霉素(100μg/mL)和异麦芽糖(0.250 mg/ mL)的 LB 培养基，研究了含有 pBE93 (阴性对照)、pBE93-dexB 或 pBE93-k1 质粒的大肠杆菌 DH5α 细胞的活性特征。该细胞在 37℃生长 20 小时。培养后，从培养基中除去细胞，并通过实施例 3 所述方法制备和分析。通过与阴性对照(只含有原始 pBE92 质粒的大肠杆菌 DH5α 细胞)相比异麦芽糖降低的水平，证实了在含有 pBE93、pBE93-dexB 或 pBE93-k1 质粒的细胞中存在胞外的异淀粉酶活性。表 6 中的结果表

明，Npr-基因融合蛋白降解了胞外供给的异麦芽糖。

表 6 大肠杆菌 DH5α 细胞中的 DexB 和 K1 胞外融合蛋白的活性

细胞系	异麦芽糖（μg/mL）
pBE93（阴性对照）	215
pBE93-dexB（分离物 4）	117
pBE93-dexB（分离物 8）	89
pBE93-k1（分离物 7）	76
pBE93-k1（分离物 8）	74
pBE93-k1（分离物 9）	62

含有 pBE93-dexB 或 pBE93-k1 质粒的大肠杆菌 DH5α 细胞降解了异麦芽糖；但是，在含有异麦芽糖作为唯一碳源的基本培养基中的细胞生长是对异淀粉酶活性的更严格的测试。因此，将 pBE93-dexB 和 pBE93-k1 质粒转化进了大肠杆菌菌株 FM5。与 DH5α 不同，FM5 菌株具有在基本培养基中生长的能力，该培养基除了碳源以外仅含有盐和痕量金属(Maniatis 等(1982) Molecular Cloning ; Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y; Neidhardt (1987) Escherichia coli and Salmonella typhimurium, ASM Press, Washington, DC) 。与 DH5α 菌株类似，天然的 FM5 细胞不能利用异麦芽糖作为碳源。为了证实这一点，将用质粒 pBE93 转化的 FM5 细胞接种到含有葡萄糖(1 mg/mL)或异麦芽糖(1 mg/mL)的 M9 培养基(Maniatis 等，同上；Neidhardt，同上)中，在 37 ℃培养至少 20 小时。20 小时后，在含有葡萄糖的烧瓶中观察到细胞生长，但是即使培养 60 小时后也未在含有异麦芽糖的烧瓶中观察到细胞生长。

与阴性对照不同，含有 Npr-DexB 和 Npr-k1 融合蛋白(分别是 pBE93- dexB 和 pBE93-k1)的 FM5 细胞在含有异麦芽糖的 M9 培养基中培养 20 小时后生长良好。为此实验，将 FM5/pBE93、FM5/pBE93-dexB 和 FM5/pBE93-k1 菌株接种到 2.0 mL M9 培养基中，其中补充了葡萄糖或异麦芽糖(1 mg/mL)作为唯一碳源。表 7 所示的结果表明，当翻译融合地连接到 Npr 信号肽上的 dexB 或 k1 基因在 FM5 细胞中表达时，异麦芽糖被代谢，且支持细胞的生长。

表 7 大肠杆菌 FM5 细胞中的 DexB 和 K1 胞外融合蛋白的活性

细胞系	异麦芽糖（μg/mL）
pBE93（阴性对照）	1091
pBE93-dexB（分离物 2）	319
pBE93-dexB（分离物 15）	197
pBE93-dexB（分离物 3）	183
pBE93-k1（分离物 5）	34
pBE93-k1（分离物 4）	20
pBE93-k1（分离物 3）	17

实施例 7 Npr-dexB 和 Npr-k1 融合基因在大肠杆菌中的表达增强了各种发酵产物的合成

当提供糖类混合物作为碳源时，生产宿主的代谢含有 α(1,6)-连接的葡萄糖残基的寡糖的能力会提高发酵产物的产量。通过首先将质粒 pBE93-dexB 和 pBE93-k1 转化进工程化的生产甘油的细胞系中，测试了 Npr-dexB 和 Npr-k1 融合蛋白的降解 α(1,6)-键的能力。

用 1 微克质粒 DNA 转化大肠杆菌菌株 RJ8n（ATCC PTA-4216），其还含有质粒 pSYC0101(spec^R)（描述于美国专利申请 10/420,587，在本文中引作参考），其编码来自酿酒糖酵母（Saccharomyces cerevisiase）的 DAR1 和 GPP2 基因和来自肺炎克雷伯氏菌（Klebsiella pneumoniae）的 dhaB 和 orf 操纵子。当加入维生素 B12 时，转化的大肠杆菌菌株能从葡萄糖生产甘油和 1,3-丙二醇。从葡萄糖生产甘油和 1,3-丙二醇的方法详细地描述在美国专利号 6,358,716 和美国专利号 6,013,494 中，它们在本文中引作参考。在这两篇美国专利及上文引用的美国专利号 5,686,276 中公开的用于生产甘油和/或 1,3-丙二醇的特异性基因和途径包括：甘油脱水酶基因、二醇脱水酶基因、1,3-丙二醇氧化还原酶基因、NAD^+-连接的甘油脱氢酶基因、$NADP^+$-连接的甘油脱氢酶基因、二羟基丙酮激酶基因、果糖-1,6-二磷酸途径基因、Embden-Meyerhof Parnas 途径基因、卡尔文循环基因、一磷酸核酮糖途径基因、丝氨酸途径基因、一磷酸木酮糖途径基因、甲胺途径基因、丙氨酸途径基因、油酸途径基因、甘油-3-磷酸脱氢酶基因和甘油-3-磷酸酶基因。将转化的 RJ8n 细胞涂布到含有 50μg/mL 壮观霉素和 100μg/mL 氨苄青霉素的 LB 培养基上。将单一菌落接种到 2.0 mL TM2 培养基（磷酸钾 7.5 g/L；柠檬酸 2.0 g/L；硫酸铵 3.0 g/L；硫酸镁 2.0 g/L；氯化钙 0.2 g/L；柠檬酸铁铵 0.33 g/L；酵母提取物(Difco-BD, Sparks, MD) 5.0 g/L；痕量元素(硫酸锌、硫酸铜、氯化钴、硫酸锰、硫酸铁、氯化钠)；氢氧化铵调 pH 至 6.5；还含有葡萄糖或异麦芽糖(1 mg/mL)。培养物在 37℃生长 24 小时。通过实施例 3 所述方法制备和分析细胞。

在用葡萄糖作为碳源的TM2培养基中在37℃培养24小时时，在仅含有质粒pSYCO101的大肠杆菌RJ8n细胞中表现出甘油积累(表8)。但是，当用异麦芽糖替代培养基中的葡萄糖时，该阴性对照系生产了可忽略水平的甘油，这表明α(1,6)-连接的葡萄糖不支持发酵产物的积累。相反地，当提供异麦芽糖或葡萄糖作为唯一碳源时，在含有质粒pSYCO101和pBE93-dexB或pBE93-k1的大肠杆菌RJ8n中生产了甘油(表8)。当异麦芽糖用作碳源时，与阴性对照系、只含有pSYCO101的RJ8n相比，含有pBE93-dexB和pSYCO101质粒的大肠杆菌RJ8n生产的甘油高8~9倍。与阴性对照相比，利用异麦芽糖的含有pSYCO101和pBE93-k1的系的甘油积累是6~10倍。表8中的数据表明，Npr-dexB或Npr-k1基因的表达使供给了异麦芽糖的培养物生产甘油。该数据还表明，无论碳源是葡萄糖还是异麦芽糖，积累的产物的水平都与含有融合蛋白的培养物相当。

表8 Npr-DexB或Npr-K1的表达造成的甘油积累

细胞系	甘油积累（μg/mL）	
	供给葡萄糖的培养物	供给异麦芽糖的培养物
RJ8n/pSYCO101	430	39
RJ8n/pSYCO101/pBE93-dexB (分离物4)	381	362
RJ8n/pSYCO101/pBE93-dexB (分离物8)	353	354
RJ8n/pSYCO101/pBE93-k1 (分离物2)	383	401
RJ8n/pSYCO101/pBE93-k1 (分离物6)	412	226

通过在含有潘糖(1 mg/mL)的TM2培养基中进行培养并与使用葡萄糖作为底物(1 mg/mL)的相同系的结果进行对比，进一步表征了含有质粒pSYCO101和pBE93-k1的大肠杆菌系RJ8n使用α(1,6)-连接的葡萄糖作为底物生产发酵产物的能力。

表 9 中的数据还表明，当提供潘糖作为 TM2 培养基中的唯一碳水化合物源时，仅含有质粒 pSYCO101 的大肠杆菌菌株 RJ8n(阴性对照)不能合成甘油。但是，当该相同菌株中存在质粒 pBE93-k1 且在含有潘糖的 TM2 培养基中培养时，能生产甘油。当提供葡萄糖或潘糖作为碳水化合物源时，含有质粒 pSYCO101 和 pBE93-k1 的大肠杆菌 RJ8n 中的甘油积累是相当的。

表 9 Npr-K1 的表达造成的甘油积累

细胞系	甘油积累（μg/mL）	
	供给葡萄糖的培养物	供给异麦芽糖的培养物
RJ8n/pSYCO101	417	25
RJ8n/pSYCO101/pBE93-k1(9)	396	363
RJ8n/pSYCO101/pBE93-k1(7)	376	347

上述数据表明，当异麦芽糖或潘糖代表了培养基中的唯一碳水化合物源时，Npr-dexB 或 Npr-k1 融合蛋白在大肠杆菌中的表达会增强甘油的生产。通过使用相同的融合蛋白表达系统合成另一种发酵产物(1,3-丙二醇)证实了，该结果不只是限于甘油的生产。

将用质粒 pSYCO101 和 pBE93-dexB 或 pBE93-k1 转化的 RJ8n 细胞接种到 2.0 mL TM2 培养基（如上所述）中，其还含有葡萄糖(1 mg/mL)或异麦芽糖(1 mg/mL)和维生素 B12 (100 ng/L)。培养物在 37 ℃生长 20 小时。通过实施例 3 所述方法制备和分析细胞。

表 10 中的数据表明，当在仅含有异麦芽糖作为碳水化合物源的培养基中生长时，阴性对照系(RJ8n/pSYCO101)不能合成 1,3-丙二醇。但是，当 Npr-dexB 或 Npr-k1 融合蛋白在 RJ8n 细胞中表达时，表现出异麦芽糖的代谢。这导致了发酵产物 1,3-丙二醇的积累。该数据还表明，当提供葡萄糖或异麦芽糖作为唯一的碳水化合物时，能表达 Npr-dexB 或 Npr-k1 融合蛋白的 RJ8n 细胞合成的 1,3-丙二醇水平是相当的。

表 10 Npr-dexB 或 Npr-k1 的表达造成的 1,3-丙二醇积累

细胞系	1,3-丙二醇（mg/mL）		
	供给葡萄糖的培养物	供给异麦芽糖的培养物	异麦芽糖（μg/mL）
RJ8n/pSYC0101	2.8	ND	1225
RJ8n/pSYC0101/pBE93-k1(9)	1.7	2.8	12
RJ8n/pSYC0101/pBE93-k1(7)	3.0	2.9	14
RJ8n/pSYC0101/pBE93-dexB	3.0	3.1	27

ND = 未测出

实施例 8 短双歧杆菌 k1 基因在使用可变启动子的大肠杆菌中的表达

用可变启动子来指导优选基因的表达通常是非常理想的。可变启动子可以用于改变基因表达的水平或定时，从而提高优选底物的利用。

通过将质粒 pBE93-k1（实施例 6）中的中性蛋白酶启动子替换为葡萄糖异构酶(GI)启动子和 GI-启动子的变体，证实了使用可变启动子的短双歧杆菌 k1 异淀粉酶基因的有效表达。Streptomyces lividins GI-启动子的分离和变体启动子的产生描述在美国专利申请 10/420,587 中。在替换 NPR-启动子之前，对中性蛋白酶信号肽的非编码核苷酸序列和 k1 基因进行了修饰。该序列修饰产生了限制酶位点，它们将用于随后的克隆步骤。

通过使用引物 SEQ ID NO. 18 和 SEQ ID NO. 19 的 PCR，将限制酶位点 SacI 和 PacI 分别加入到中性蛋白酶信号肽和 k1 基因序列的 5′和 3′-端。通过实施例 3 所述方法进行 PCR 扩增。分离 1919 个碱基对的 PCR 产物，并连接到如美国专利申请 10/420,587 所公开的 pSYC0109mcs 野生型 GI yqhD 质粒，其也用酶 SacI 和 PacI 消化过。得到的质粒含有野生型 GI 启动子和翻译融合地连接到 k1 基因的 NPR-信号序列。该构建体命名为 WTGI-ss-K1。还用变体 GI 启动子指导 NPR-信号肽/K1 融合体的表达。使用 SacI 和 PacI 限制酶位点，将从使用引物 SEQ ID NO: 18 和 SEQ ID NO: 19 的反应得到的 1919 个碱基对的 PCR 产物置入 pSYC0109mcs-短 1.6 GI yqhD 质粒。得到的质粒命名为 LowGI-ss-K1。当可操作地连接到 yqhD 基因时，与野生型 GI 启动

子-yqhD构建体相比，该变体启动子以前表现出指导低水平的基因表达(美国专利申请10/420,587)。

通过活性实验，证实了使用野生型和变体GI启动子的k1基因的有效表达。用质粒WTGI-ss-K1和LowGI-ss-K1转化了大肠杆菌细胞(菌株DH5α, Invitrogen, Carlsbad, CA)，在LB培养基中生长过夜。通过离心回收细胞沉淀，悬浮到1/10体积的磷酸钠缓冲液(10 mM, pH 7.0)中。用弗氏压碎器裂解悬浮液中的细胞，通过离心除去细胞碎片。通过Bradford实验(Bio-Rad, Hercules, CA)检测总蛋白浓度。使用4-硝基苯基-α-D-吡喃葡糖苷(PNPG, Sigma, ST. Louis, MO)检测总蛋白分离物中的k1基因产物的活性。将来自含有质粒WTGI-ss-K1、LowGI-ss-K1、NPR-ss-K1（阳性对照)和pSYC0109（阴性对照)的细胞的总蛋白提取物在10 mM磷酸钠缓冲液（含有10 mM PNPG）中在30℃温育多达30分钟。从PNPG释放出的葡萄糖残基导致PNP积累，其吸收400 nm光。通过在400 nm波长的吸光度，监控了k1酶活性直接导致的PNP积累随时间的变化。下面的表11表明，除了中性蛋白酶启动子之外的启动子可以用于指导有活性的K1基因的表达。该结果还表明，可变启动子可以用于改变K1表达水平，且K1活性与相关的启动子强度水平相对应。

表11 K1酶活性导致的PNP生产速率

质粒	活性（mM PNP/mg蛋白 分钟$^{-1}$）
WTGI-ss-K1（高表达者）	0.0144
NPR-ss-K1（阳性对照）	0.0104
LowGI-ss-K1（低表达者）	0.0028
pSYC0109（阴性对照）	0.0002

实施例9 短双歧杆菌k1基因向大肠杆菌基因组的整合

所需DNA向细胞基因组中的整合可以增强基因表达随时间过去和在各种发酵条件下的稳定性。但是，整合位置会影响基因表达水平，并最终影响所需酶活性的有效性。

通过首先克隆进质粒pKD3 (Datsenko和Wanner, Proc. Natl. Acad. Sci. 97: 6640-6645 (2000))中，完成了k1表达盒（NPR启动子-信号肽-k1基因)向大肠杆菌(菌株FM5)基因组的整合，并证实了

使用 α(1,6)-连接的葡萄糖底物的利用。选择了宿主的 aldA (醛脱氢酶 A)和 aldB (醛脱氢酶 B)基因组位点用于整合。将 PCR 引物设计成与质粒 pKD3、aldA 或 aldB 和 k1 基因序列(SEQ ID NO: 20 至 23)同源。

通过实施例 3 所述方法进行 PCR 扩增。分离从引物 SEQ ID NO. 21-23 的反应得到的 PCR 产物和含有 k1 表达盒的质粒 pKD3，连接并转化进大肠杆菌(FM5)。通过在含有氯霉素的 LB 培养基上生长，选择含有整合的 k1 表达盒的细胞。使用引物 SEQ ID NO: 7 和 SEQ ID NO: 8，通过 PCR 反应检测氯霉素阳性菌落中 k1 基因的存在。

通过在含有异麦芽糖作为唯一碳水化合物源的培养基中的生长分析，进一步检测了含有整合的 k1 表达盒的 FM5 系的活性。将氯霉素和 PCR 阳性的菌落接种到含有 0.5%异麦芽糖(w/v)的 TM2 培养基(见实施例 7)中，在 35 ℃生长。在各个时间点取样，通过光密度(A600nm)和异麦芽糖消耗(通过 HPLC，参考一般方法)表征细胞物质积累。

下面的表 12 表明，FM5 细胞本身不能代谢作为唯一碳水化合物源提供的异麦芽糖。当在含有 0.5%异麦芽糖的 TM2 培养基中生长时，低水平的细胞物质积累证实了这一点。观察到了阴性系 FM5 的低水平生长，但仅仅是由于小量的可发酵的麦芽糖污染了异麦芽糖源材料(Sigma, St. Louis, MO)。含有整合的 K1 表达盒的细胞于高得多的速率生长，并在培养 25 小时后达到更高的最终光密度。将含有整合在 aldA 位点的 k1 表达盒的 PCR-阳性的菌落命名为 A2-3。将含有整合在 aldB 位点的 k1 表达盒的 PCR 阳性的菌落命名为 B1-1 和 B1-2。

表 12 细胞物质积累(A600nm)

时间(小时)	FM5	FM5-A2-3	FM5-B1-1	FM5-B1-2
0	0.02	0.02	0.02	0.02
3	0.66	0.72	0.76	0.75
6	2.75	3.17	6.60	6.01
8	3.34	4.50	10.40	9.92
11	3.72	8.34	10.41	10.10
25	3.66	10.16	11.10	10.78

还通过 HPLC 分析，将含有整合的 k1 表达盒的细胞对异麦芽糖的消耗与 FM5 阴性对照系进行了对比。表 13 中的数据表明，与不能利用该碳水化合物的阴性对照相比，整合后的 k1 表达盒是有活性的，并且能

使细胞彻底利用可获得的含有 α(1,6)-连接的葡萄糖的糖。该数据还表明，在基因已经整合进 aldA 位点的系中，异麦芽糖的消耗速率与整合进 aldB 位点的不同。

表 13 异麦芽糖消耗(g/L)

时间(小时)	FM5	FM5-A2-3	FM5-B1-1	FM5-B1-2
0	5.56	5.46	5.36	5.31
3	5.52	5.35	5.31	5.30
6	5.60	4.73	1.81	1.78
8	5.48	3.64	0	0
11	5.77	1.34	0	0
25	5.55	0	0	0

序列表

<110> E.I. DuPont de Nemours & Co., Inc.

<120> 淀粉制品在利用发酵的生物生产中的应用

<130> CL2123 PCT

<150> 60/405896
<151> 08-23-2002

<160> 43

<170> PatentIn version 3.2

<210> 1
<211> 1815
<212> DNA
<213> 短双岐杆菌 ATCC#15700

<400> 1
atgaccgcca acaacctcaa tgacgactgg tggaagcagg ccgtcgttta ccagatttac 60

ccgcgcagct tcaaggacgt taacggcgac ggcatcggcg acatcgccgg cgttaccgag 120

aaaatggact acctgaaaaa cctcggcgtg gacgccatct ggctctcccc gttctacccc 180

tccgatctgg cggacggcgg ctacgacgtg atcgactacc gcaacgtcga cccgcgactg 240

ggcaccatgg acgacttcga cgccatggcc aaagccgcgc atgaggccgg catcaaggtg 300

atcgtggaca tcgtgcccaa tcacaccgcc gacaagcacg tgttcttcaa ggaagccctc 360

gccgccgagc ccggctcccc ggcgcgcgac cgctacatct tccgcgacgg ccgcggcgag 420

cacggcgaac tgccgcccaa cgactggcag tccttcttcg gcggcccggc ctgggctcgc 480

gtggccgacg gccagtggta tctgcacctg ttcgacaagg cgcaaccgga cgtcaactgg 540

aagaacccgg acatccacga ggaattcaag aaaaccctgc gcttctggtc cgaccacggc 600

accgacggct tccgcatcga cgtggcgcac ggtctggcca aagaccttga atccaagccg 660

ctggaggagc tcggccgcga atacagcgtg gtcggcgtgc tgaatcacga cttcagccat 720

ccgctgttcg accgccgcga agtgcacgac atctaccgcg aatggcgcaa ggtgttcaac 780

gagtacgacc cgccgcgctt tgccgtggcc gaggcgtggg tggtacccga gcaccagcac 840

ctgtatgcct cgatggatga gctggggcag tccttcaact tcgactttgc gcaggccagc 900

tggtatgccg atgagttccg cgcagccatc gccgcgggtc tcaaggccgc tgccgaaacc 960

ggcggttcca ccaccacgtg ggtgatgaac aaccatgacg tgccgcgcag ccgctccgcc 1020

tatggtctac cgcaggtcaa gggcgcgcct taccaccagc tgccgcacga ctggctgctg 1080

cgcaacggca ccacgtatcc cgaggatcgc gagcttggca ccgccgcgcc ccgcgccgcc 1140

gctttgatgg agctcggcct gcccggcgcc gcctatatct atcagggcga ggagctgggc 1200

ctgtttgaag tggccgatat tccgtgggat cgactggaag atccgaccgc tttccacacc 1260

gctcaggcca cgatggacaa gggccgagac ggctgccgcg tgccgattcc gtggaccgct 1320

gcaaacgaac cgaccttggc tgatttcagc cgcccgatcc cggccgatga cggcaccggc 1380

gagaaccacg tgccgctgtg cgccgccggc cagttcggca cgggcgcttc cttcggcttc 1440

tcgccggcta cgcgcgctga gggcgtgacg ccggccgccg acccgcacct gccgcagccg 1500

ttgtggttca aggattacgc ggtggacgtg gagcaggccg acccggattc aatgctcgcg 1560

ctgtatcatg cggcgttggc gattcgtcag gagtcgctga ccgccacgcg tgacaccacc 1620

gctgagcagg tggatatggg gccggacgtg gtggcctaca cccgcgcggc ggttggtggc 1680

cgcacgttca cctcgatcac caacttcggc accgagccgg tggagctgcc tggaggctcc 1740

gtcgtgctga cgtccggccc gctgaccccc gacggccagc tccccaccga cacttctgcg 1800

tgggttatca agtag 1815

<210> 2
<211> 604
<212> PRT
<213> 短双岐杆菌 ATCC#15700

<400> 2

Met Thr Ala Asn Asn Leu Asn Asp Asp Trp Trp Lys Gln Ala Val Val
1 5 10 15

Tyr Gln Ile Tyr Pro Arg Ser Phe Lys Asp Val Asn Gly Asp Gly Ile

20 25 30

Gly Asp Ile Ala Gly Val Thr Glu Lys Met Asp Tyr Leu Lys Asn Leu
35 40 45

Gly Val Asp Ala Ile Trp Leu Ser Pro Phe Tyr Pro Ser Asp Leu Ala
50 55 60

Asp Gly Gly Tyr Asp Val Ile Asp Tyr Arg Asn Val Asp Pro Arg Leu
65 70 75 80

Gly Thr Met Asp Asp Phe Asp Ala Met Ala Lys Ala Ala His Glu Ala
85 90 95

Gly Ile Lys Val Ile Val Asp Ile Val Pro Asn His Thr Ala Asp Lys
100 105 110

His Val Phe Phe Lys Glu Ala Leu Ala Ala Glu Pro Gly Ser Pro Ala
115 120 125

Arg Asp Arg Tyr Ile Phe Arg Asp Gly Arg Gly Glu His Gly Glu Leu
130 135 140

Pro Pro Asn Asp Trp Gln Ser Phe Phe Gly Gly Pro Ala Trp Ala Arg
145 150 155 160

Val Ala Asp Gly Gln Trp Tyr Leu His Leu Phe Asp Lys Ala Gln Pro
165 170 175

Asp Val Asn Trp Lys Asn Pro Asp Ile His Glu Glu Phe Lys Lys Thr
180 185 190

Leu Arg Phe Trp Ser Asp His Gly Thr Asp Gly Phe Arg Ile Asp Val
195 200 205

Ala His Gly Leu Ala Lys Asp Leu Glu Ser Lys Pro Leu Glu Glu Leu

210 215 220

Gly Arg Glu Tyr Ser Val Val Gly Val Leu Asn His Asp Phe Ser His
225 230 235 240

Pro Leu Phe Asp Arg Arg Glu Val His Asp Ile Tyr Arg Glu Trp Arg
245 250 255

Lys Val Phe Asn Glu Tyr Asp Pro Pro Arg Phe Ala Val Ala Glu Ala
260 265 270

Trp Val Val Pro Glu His Gln His Leu Tyr Ala Ser Met Asp Glu Leu
275 280 285

Gly Gln Ser Phe Asn Phe Asp Phe Ala Gln Ala Ser Trp Tyr Ala Asp
290 295 300

Glu Phe Arg Ala Ala Ile Ala Ala Gly Leu Lys Ala Ala Ala Glu Thr
305 310 315 320

Gly Gly Ser Thr Thr Thr Trp Val Met Asn Asn His Asp Val Pro Arg
325 330 335

Ser Pro Ser Arg Tyr Gly Leu Pro Gln Val Lys Gly Ala Pro Tyr His
340 345 350

Gln Leu Pro His Asp Trp Leu Leu Arg Asn Gly Thr Thr Tyr Pro Glu
355 360 365

Asp Arg Glu Leu Gly Thr Arg Arg Ala Arg Ala Ala Ala Leu Met Glu
370 375 380

Leu Gly Leu Pro Gly Ala Ala Tyr Ile Tyr Gln Gly Glu Glu Leu Gly
385 390 395 400

Leu Phe Glu Val Ala Asp Ile Pro Trp Asp Arg Leu Glu Asp Pro Thr

405 410 415

Ala Phe His Thr Ala Gln Ala Thr Met Asp Lys Gly Arg Asp Gly Cys
420 425 430

Arg Val Pro Ile Pro Trp Thr Ala Ala Asn Glu Pro Thr Leu Ala Asp
435 440 445

Phe Ser Arg Pro Ile Pro Ala Asp Asp Gly Thr Gly Glu Asn His Val
450 455 460

Pro Leu Cys Ala Ala Gly Gln Phe Gly Thr Gly Ala Ser Phe Gly Phe
465 470 475 480

Ser Pro Ala Thr Arg Ala Glu Gly Val Thr Pro Ala Ala Asp Pro His
485 490 495

Leu Pro Gln Pro Leu Trp Phe Lys Asp Tyr Ala Val Asp Val Glu Gln
500 505 510

Ala Asp Pro Asp Ser Met Leu Ala Leu Tyr His Ala Ala Leu Ala Ile
515 520 525

Arg Gln Glu Ser Leu Thr Ala Thr Arg Asp Thr Thr Ala Glu Gln Val
530 535 540

Asp Met Gly Pro Asp Val Val Ala Tyr Thr Arg Ala Ala Val Gly Gly
545 550 555 560

Arg Thr Phe Thr Ser Ile Thr Asn Phe Gly Thr Glu Pro Val Glu Leu
565 570 575

Pro Gly Gly Ser Val Val Leu Thr Ser Gly Pro Leu Thr Pro Asp Gly
580 585 590

Gln Leu Pro Thr Asp Thr Ser Ala Trp Val Ile Lys

595 600

<210> 3
<211> 1845
<212> DNA
<213> 短双岐杆菌 ATCC#15700

<400> 3
atgaataagg agccaacgat gactactttc aacegcacaa taattcccga cgccattcgc 60

accaacggag ccacgcccaa cccgtggtgg tcgaacgccg tcgtctacca gatttaccca 120

cgttccttcc aggacacgaa cggcgatggt ctcggcgacc tgaagggcat cacctcccgc 180

ctcgactatc tcgccgacct cggcgtggat gtgctgtggc tctctccggt ctacaggtcc 240

ccgcaagacg acaacggcta cgacatctcc gactaccggg acatcgaccc gctgttcggc 300

acgcttgacg acatggacga gctgctcgcc gaagcgcaca agcgcggcct caagatcgtg 360

atggacctgg tggtcaatca cacctctgac gagcacgcgt ggttcgaggc gtcgaaggac 420

aaggacgacc cgcacgccga ctggtactgg tggcgtcccg cccgcccggc ccacgagccg 480

ggcacgcccg gcgccgagcc gaatcagtgg ggctcctact tcggcggttc cgcatgggag 540

tacagccccgg agcgcggcga gtactacctg caccagttct cgaagaagca gcctgatctc 600

aactgggaga accccggccgt gcgccgtgca gtgtacgaca tgatgaactg gtggctcgat 660

cgcggcatcg acggcttccg tatggacgtc atcaccctta tctccaagcg caccgacccc 720

aacggcaggc tccccggcga ggccggttcc gagctgcagg acctgccggt gggggaggag 780

ggctactccg accccgaatcc gttctgtgcg gacggccccc gtcaggatga attcctggct 840

gaaatgcgcc gtgaggtatt cgaagggcgt gacggcttcc tgactgtagg cgaggcgcca 900

ggcgtcacag cccagcgcaa cgaatacatc accgatccgg ccaatggcga gctggatatg 960

ctcttcctat ttgagcatgt tgattttgat tgcgaaggta ccaagtggaa gccgttgccg 1020

ctcgatctgc cgaagcttaa gagcatcatg gccggctatc aggccgctgt gcagaacgca 1080

ggatgggcca gcctattcac cggcaaccac gatcagccgc gcgtggtttc ccgctggggt 1140

gacgattcct cggaagaggc tcgggtccgc tcggccaagg cccttggcct gatgctgcac 1200

ctgcaccgtg gtaccccgta catctatcag ggtgaagaat tgggcatgac cgacgcccac 1260

ttcactcgtc tcgaccagta ccgcgacctt gaatccctga acgcctaccg tcaaagggtc 1320

gaagaggcca aggtgcagtc gccccgaatcc atgatggccg gtatcgccgc ccgcggtcgt 1380

gacaactcac gcacaccgat gcagtgggat ggctccgtct acgccggttt caccgcacct 1440

gacgcagccg ccgagccatg gatctccgtg aatccgaatc atgccgagat caacgccgcc 1500

ggcgaattcg atgatccgga ttcggtgtac tccttctaca agcggctcat cgcgctgcgc 1560

cacgacatgc ctgtcgtgga ggccggcgac tggcatctgc tcgacgcgga cgatgcgcat 1620

gtgtatgcct tcactcgtac cctcggtgac gagaagttgt tggtcgtggt caatatgtcc 1680

gggcgaactg ttgatttgcc tcgcgaatcc gccgaactgt tggcagtggc cgatggcctt 1740

gccgagtcga acgtggtgat cagcacgtat gatgccccgc acgctgtgac cgctcttgcc 1800

ggccgtgagc ttgcaccatg ggagggcgtc gtcgtcagcc tataa 1845

<210> 4
<211> 613
<212> PRT
<213> 短双岐杆菌 ATCC#15700

<400> 4

Met Asn Lys Glu Pro Thr Met Thr Thr Phe Asn Arg Thr Ile Ile Pro
1 5 10 15

Asp Ala Ile Arg Thr Asn Gly Ala Thr Pro Asn Pro Trp Trp Ser Asn
20 25 30

Ala Val Val Tyr Gln Ile Tyr Pro Arg Ser Phe Gln Asp Thr Asn Gly
35 40 45

Asp Gly Leu Gly Asp Leu Lys Gly Ile Thr Ser Arg Leu Asp Tyr Leu
50 55 60

Ala Asp Leu Gly Val Asp Val Leu Trp Leu Ser Pro Val Tyr Arg Ser

65 70 75 80

Pro Gln Asp Asp Asn Gly Tyr Asp Ile Ser Asp Tyr Arg Asp Ile Asp
85 90 95

Pro Leu Phe Gly Thr Leu Asp Asp Met Asp Glu Leu Leu Ala Glu Ala
100 105 110

His Lys Arg Gly Leu Lys Ile Val Met Asp Leu Val Val Asn His Thr
115 120 125

Ser Asp Glu His Ala Trp Phe Glu Ala Ser Lys Asp Lys Asp Asp Pro
130 135 140

His Ala Asp Trp Tyr Trp Trp Arg Pro Ala Arg Pro Gly His Glu Pro
145 150 155 160

Gly Thr Pro Gly Ala Glu Pro Asn Gln Trp Gly Ser Tyr Phe Gly Gly
165 170 175

Ser Ala Trp Glu Tyr Ser Pro Glu Arg Gly Glu Tyr Tyr Leu His Gln
180 185 190

Phe Ser Lys Lys Gln Pro Asp Leu Asn Trp Glu Asn Pro Ala Val Arg
195 200 205

Arg Ala Val Tyr Asp Met Met Asn Trp Trp Leu Asp Arg Gly Ile Asp
210 215 220

Gly Phe Arg Met Asp Val Ile Thr Leu Ile Ser Lys Arg Thr Asp Pro
225 230 235 240

Asn Gly Arg Leu Pro Gly Glu Ala Gly Ser Glu Leu Gln Asp Leu Pro
245 250 255

Val Gly Glu Glu Gly Tyr Ser Asp Pro Asn Pro Phe Cys Ala Asp Gly

260 265 270

Pro Arg Gln Asp Glu Phe Leu Ala Glu Met Arg Arg Glu Val Phe Glu
275 280 285

Gly Arg Asp Gly Phe Leu Thr Val Gly Glu Ala Pro Gly Val Thr Ala
290 295 300

Gln Arg Asn Glu Tyr Ile Thr Asp Pro Ala Asn Gly Glu Leu Asp Met
305 310 315 320

Leu Phe Leu Phe Glu His Val Asp Phe Asp Cys Glu Gly Thr Lys Trp
325 330 335

Lys Pro Leu Pro Leu Asp Leu Pro Lys Leu Lys Ser Ile Met Ala Gly
340 345 350

Tyr Gln Ala Ala Val Gln Asn Ala Gly Trp Ala Ser Leu Phe Thr Gly
355 360 365

Asn His Asp Gln Pro Arg Val Val Ser Arg Trp Gly Asp Asp Ser Ser
370 375 380

Glu Glu Ala Arg Val Arg Ser Ala Lys Ala Leu Gly Leu Met Leu His
385 390 395 400

Leu His Arg Gly Thr Pro Tyr Ile Tyr Gln Gly Glu Glu Leu Gly Met
405 410 415

Thr Asp Ala His Phe Thr Arg Leu Asp Gln Tyr Arg Asp Leu Glu Ser
420 425 430

Leu Asn Ala Tyr Arg Gln Arg Val Glu Glu Ala Lys Val Gln Ser Pro
435 440 445

Glu Ser Met Met Ala Gly Ile Ala Ala Arg Gly Arg Asp Asn Ser Arg

450 455 460

Thr Pro Met Gln Trp Asp Gly Ser Val Tyr Ala Gly Phe Thr Ala Pro
465 470 475 480

Asp Ala Ala Ala Glu Pro Trp Ile Ser Val Asn Pro Asn His Ala Glu
485 490 495

Ile Asn Ala Ala Gly Glu Phe Asp Asp Pro Asp Ser Val Tyr Ser Phe
500 505 510

Tyr Lys Arg Leu Ile Ala Leu Arg His Asp Met Pro Val Val Glu Ala
515 520 525

Gly Asp Trp His Leu Leu Asp Ala Asp Asp Ala His Val Tyr Ala Phe
530 535 540

Thr Arg Thr Leu Gly Asp Glu Lys Leu Leu Val Val Val Asn Met Ser
545 550 555 560

Gly Arg Thr Val Asp Leu Pro Arg Glu Ser Ala Glu Leu Leu Ala Val
565 570 575

Ala Asp Gly Leu Ala Glu Ser Asn Val Val Ile Ser Thr Tyr Asp Ala
580 585 590

Pro His Ala Val Thr Ala Leu Ala Gly Arg Glu Leu Ala Pro Trp Glu
595 600 605

Gly Val Val Val Ser
610

<210> 5
<211> 1818
<212> DNA
<213> 短双岐杆菌 ATCC#15700

<400> 5

atgatgacct ctttcaaccg tgaacccctg cccgacgccg tccgcacgaa tggcgcctcc 60

ccgaacccgt ggtggtcgaa cgccgtcgtc taccagattt acccacgttc cttccaggac 120

acgaacggcg atggtttcgg agatcttaag ggcattactt cccgcctcga ctatcttgcc 180

gacctcggcg tggatgtgct gtggctctcc ccggtctaca ggtcccccgca agacgacaac 240

ggctacgaca tctccgacta ccgggacatc gacccgctgt tcggcacgct cgacgacatg 300

gacgagctgc tcgccgaagc gcacaagcgc ggcctcaaga tcgtgatgga cctggtcgtc 360

aaccacacct ccgacgagca cgcgtggttc gaggcgtcga aggacaagga cgacccgcac 420

gccgactggt actggtggcg tcccgccccgc cccggccacg agcccggcac gcccggcgcc 480

gagccgaacc agtggggctc ctacttcggc ggctccgcat gggaatattg ccccgagcgt 540

ggtgagtact atctccacca gttctcgaag aagcagccgg acctcaactg ggagaacccg 600

gccgtacgcc gagccgtgta cgacatgatg aactggtggc tcgatcgcgg catcgacggc 660

ttccgcatgg acgtcatcac cctgatctcc aagcgcacgg atgcaaacgg caggctgccc 720

ggcgagaccg gttccgagct gcaggacctg ccggtggggg aggagggcta ctccaacccg 780

aacccgttct gcgccgacgg tccgcgtcag gacgagttcc tcgccgagat gcgccgcgag 840

gtgttcgacg ggcgtgacgg cttcctcacc gtcggcgagg ccccccggcat caccgccgaa 900

cgcaacgagc acatcaccga ccccgccaac ggcgagctgg atatgctctt cctgttcgaa 960

cacatgggcg tcgaccaaac ccccgaatcg aaatgggacg acaaaccatg gacgccggcc 1020

gacctcgaaa ccaagctcgc cgaacaacag gacgccatcg cccgacacgg ctgggccagc 1080

ctgttcctcg acaaccacga ccagccgcgt gtcgtctccc gttggggcga cgacaccagc 1140

aagaccggcc gcatccgctc cgccaaggcg ctcgcgctgc tgctgcacat gcaccgcggc 1200

actccgtatg tctaccaggg cgaggagctc ggcatgacca atgcgcactt cacctcgctc 1260

gaccagtacc gcgacctcga atccatcaac gcctaccatc aacgcgtcga ggaaaccggg 1320

atacggacat cggagaccat gatgcgatcc ctcgcccgat acggcaggga caacgcgcgc 1380

accccgatgc aatgggacga ctccacctac gccggcttca ccatgcccga cgccccggtc 1440

gaaccctgga tcgccgtcaa cccgaaccac acggagatca acgccgccga cgagatcgac 1500

gaccccgact ccgtgtactc gttccacaaa cggctcatcg ccctgcgtca caccgacccc 1560

gtggtcgccg ccggcgacta ccgacgcgtg gaaaccggaa acgaccggat catcgccttc 1620

accagaaccc tcgacgagcg aaccatcctc accgtcatca acctctcgcc cacacaggcc 1680

gcaccggccg gagaactgga aacgatgccc gacggcacga tcctcatcgc caacacggac 1740

gatcccgtag gaaacctgaa aaccacggga acactcggac catgggaggc gttcgccatg 1800

gaaaccgatc cggaataa 1818

<210> 6
<211> 605
<212> PRT
<213> 短双岐杆菌 ATCC#15700

<400> 6

Met Met Thr Ser Phe Asn Arg Glu Pro Leu Pro Asp Ala Val Arg Thr
1 5 10 15

Asn Gly Ala Ser Pro Asn Pro Trp Trp Ser Asn Ala Val Val Tyr Gln
20 25 30

Ile Tyr Pro Arg Ser Phe Gln Asp Thr Asn Gly Asp Gly Phe Gly Asp
35 40 45

Leu Lys Gly Ile Thr Ser Arg Leu Asp Tyr Leu Ala Asp Leu Gly Val
50 55 60

Asp Val Leu Trp Leu Ser Pro Val Tyr Arg Ser Pro Gln Asp Asp Asn
65 70 75 80

Gly Tyr Asp Ile Ser Asp Tyr Arg Asp Ile Asp Pro Leu Phe Gly Thr
85 90 95

Leu Asp Asp Met Asp Glu Leu Leu Ala Glu Ala His Lys Arg Gly Leu

100 105 110

Lys Ile Val Met Asp Leu Val Val Asn His Thr Ser Asp Glu His Ala
115 120 125

Trp Phe Glu Ala Ser Lys Asp Lys Asp Asp Pro His Ala Asp Trp Tyr
130 135 140

Trp Trp Arg Pro Ala Arg Pro Gly His Glu Pro Gly Thr Pro Gly Ala
145 150 155 160

Glu Pro Asn Gln Trp Gly Ser Tyr Phe Gly Gly Ser Ala Trp Glu Tyr
165 170 175

Cys Pro Glu Arg Gly Glu Tyr Tyr Leu His Gln Phe Ser Lys Lys Gln
180 185 190

Pro Asp Leu Asn Trp Glu Asn Pro Ala Val Arg Arg Ala Val Tyr Asp
195 200 205

Met Met Asn Trp Trp Leu Asp Arg Gly Ile Asp Gly Phe Arg Met Asp
210 215 220

Val Ile Thr Leu Ile Ser Lys Arg Thr Asp Ala Asn Gly Arg Leu Pro
225 230 235 240

Gly Glu Thr Gly Ser Glu Leu Gln Asp Leu Pro Val Gly Glu Glu Gly
245 250 255

Tyr Ser Asn Pro Asn Pro Phe Cys Ala Asp Gly Pro Arg Gln Asp Glu
260 265 270

Phe Leu Ala Glu Met Arg Arg Glu Val Phe Asp Gly Arg Asp Gly Phe
275 280 285

Leu Thr Val Gly Glu Ala Pro Gly Ile Thr Ala Glu Arg Asn Glu His

290 295 300

Ile Thr Asp Pro Ala Asn Gly Glu Leu Asp Met Leu Phe Leu Phe Glu
305 310 315 320

His Met Gly Val Asp Gln Thr Pro Glu Ser Lys Trp Asp Asp Lys Pro
325 330 335

Trp Thr Pro Ala Asp Leu Glu Thr Lys Leu Ala Glu Gln Gln Asp Ala
340 345 350

Ile Ala Arg His Gly Trp Ala Ser Leu Phe Leu Asp Asn His Asp Gln
355 360 365

Pro Arg Val Val Ser Arg Trp Gly Asp Asp Thr Ser Lys Thr Gly Arg
370 375 380

Ile Arg Ser Ala Lys Ala Leu Ala Leu Leu Leu His Met His Arg Gly
385 390 395 400

Thr Pro Tyr Val Tyr Gln Gly Glu Glu Leu Gly Met Thr Asn Ala His
405 410 415

Phe Thr Ser Leu Asp Gln Tyr Arg Asp Leu Glu Ser Ile Asn Ala Tyr
420 425 430

His Gln Arg Val Glu Glu Thr Gly Ile Arg Thr Ser Glu Thr Met Met
435 440 445

Arg Ser Leu Ala Arg Tyr Gly Arg Asp Asn Ala Arg Thr Pro Met Gln
450 455 460

Trp Asp Asp Ser Thr Tyr Ala Gly Phe Thr Met Pro Asp Ala Pro Val
465 470 475 480

Glu Pro Trp Ile Ala Val Asn Pro Asn His Thr Glu Ile Asn Ala Ala

485 490 495

Asp Glu Ile Asp Asp Pro Asp Ser Val Tyr Ser Phe His Lys Arg Leu
500 505 510

Ile Ala Leu Arg His Thr Asp Pro Val Val Ala Ala Gly Asp Tyr Arg
515 520 525

Arg Val Glu Thr Gly Asn Asp Arg Ile Ile Ala Phe Thr Arg Thr Leu
530 535 540

Asp Glu Arg Thr Ile Leu Thr Val Ile Asn Leu Ser Pro Thr Gln Ala
545 550 555 560

Ala Pro Ala Gly Glu Leu Glu Thr Met Pro Asp Gly Thr Ile Leu Ile
565 570 575

Ala Asn Thr Asp Asp Pro Val Gly Asn Leu Lys Thr Thr Gly Thr Leu
580 585 590

Gly Pro Trp Glu Ala Phe Ala Met Glu Thr Asp Pro Glu
595 600 605

<210> 7
<211> 39
<212> DNA
<213> 人工序列

<220>
<223> 引物

<400> 7
gcatgcggat ccatgcaaaa acattggtgg cacaaggca 39

<210> 8
<211> 44
<212> DNA
<213> 人工序列

<220>
<223> 引物

<400> 8
ggtaccgtcg acttagttta tcttaataca aaaagcatcc caag 44

<210> 9
<211> 39
<212> DNA
<213> 人工序列

<220>
<223> 引物

<400> 9
gacgtatatg atatccgcgc tagcagagga tgtgctgcc 39

<210> 10
<211> 17
<212> DNA
<213> 人工序列

<220>
<223> 引物

<400> 10
gaattcgagc tcggtac 17

<210> 11
<211> 39
<212> DNA
<213> 人工序列

<220>
<223> 引物

<400> 11
gacgtatatg atatccgcgc tagcacccgg cagactgat 39

<210> 12
<211> 29
<212> DNA
<213> 人工序列

<220>
<223> 引物

<400> 12
atgcatggta ccgatctaac attttcccc 29

<210> 13
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 引物

<400> 13
gagtctgcta gcgcgatgca aaaacattgg tggcac 36

<210> 14
<211> 42
<212> DNA
<213> 人工序列

<220>
<223> 引物

<400> 14
cgcggatccg ctagcgcgat gatgacctct ttcaaccgtg aa 42

<210> 15
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 引物

<400> 15
gagtctaagc ttttattccg gatcggtttc catggc 36

<210> 16
<211> 1611
<212> DNA
<213> 变异链球菌 ATCC#25175D

<400> 16

atgcaaaaac attggtggca caaggcaact gtttatcaaa tttatccaaa atcttttatg 60

gatacaaatg gtgatggaat tggtgatctc aaaggtatta cgagtaaatt ggattatttg 120

caaaagttag gggttatggc tatttggcta tctccagttt atgatagccc catggatgac 180

aatggctatg acattgcgaa ctatgaagca attgcggata tttttggcaa tatggctgat 240

atggataatt tgctgacgca ggcaaaaatg cgcgacataa aaatcattat ggatctagtg 300

gttaatcata cctcagatga acatacttgg tttattgaag cacgtgagca tccagacagt 360

tctgaacgcg attattatat ttggtgtgac cagccaaatg atttggaatc tattttcggt 420

ggttctgctt ggcagtatga tgataagtcc gatcaatatt atttgcattt ttttagtaag 480

aagcagccag atctaaactg ggaaaacgca aacttacgtc agaagattta tgatatgatg 540

aatttctgga ttgataaagg tattggcggc tttcggatgg acgtcattga tatgattggg 600

aaaattcctg ctcagcatat tgtcagtaac ggaccaaaat tgcatgctta tcttaaggag 660

atgaatgccg ctagttttgg tcaacatgat ctgctgactg tgggggaaac ttggggagca 720

acgcctgaga ttgcgaagca atattcaaat ccagtcaatc acgaactctc tatgattttt 780

caatttgaac atattggtct tcagcataaa ccagaagctc ctaaatggga ttatgtgaag 840

gaacttaatg ttcctgcttt aaaaacaatc tttaataaat ggcagactga gttggaatta 900

ggacaggggt ggaattcgtt attctggaat aaccatgacc tgcctcgtgt tttatcaatc 960

tggggaaata cgggcaaata tcgtgagaag tctgctaaag cactggctat tcttcttcac 1020

cttatgcgtg ggacacctta tatttatcaa ggtgaagaga ttgggatgac caattatcct 1080

tttaaagatt taaatgaact tgatgatatt gaatcactta attatgctaa ggaagctttt 1140

acaaatggta agtctatgga aactatcatg gacagtattc gtatgattgg ccgtgataat 1200

gccagaacac ctatgcaatg ggatgcttct caaaatgccg gattttcaac agcggataaa 1260

acatggctgc cagttaatcc aaactataaa gacatcaatg ttcaagcagc tctgaaaaat 1320

tccaattcta tcttttacac ctatcaacaa ctcattcagc ttcgaaaaga aaatgattgg 1380

ctagtagatg ccgattttga attgctccct acagcggaca aagtatttgc ctatttacga 1440

aaggtaagag aagaaaggta tcttatagtg gtcaatgttt cagatcagga agaagttcta 1500

gagattgatg ttgacaaaca agaaactctc attagcaata caaatgaaag cgctgctctt 1560

gccaatcaca aactccagcc ttgggatgct ttttgtatta agataaacta a 1611

<210> 17
<211> 536
<212> PRT
<213> 变异链球菌 ATCC#25175D

<400> 17

Met Gln Lys His Trp Trp His Lys Ala Thr Val Tyr Gln Ile Tyr Pro
1 5 10 15

Lys Ser Phe Met Asp Thr Asn Gly Asp Gly Ile Gly Asp Leu Lys Gly
20 25 30

Ile Thr Ser Lys Leu Asp Tyr Leu Gln Lys Leu Gly Val Met Ala Ile
35 40 45

Trp Leu Ser Pro Val Tyr Asp Ser Pro Met Asp Asp Asn Gly Tyr Asp
50 55 60

Ile Ala Asn Tyr Glu Ala Ile Ala Asp Ile Phe Gly Asn Met Ala Asp
65 70 75 80

Met Asp Asn Leu Leu Thr Gln Ala Lys Met Arg Asp Ile Lys Ile Ile
85 90 95

Met Asp Leu Val Val Asn His Thr Ser Asp Glu His Thr Trp Phe Ile
100 105 110

Glu Ala Arg Glu His Pro Asp Ser Ser Glu Arg Asp Tyr Tyr Ile Trp
115 120 125

Cys Asp Gln Pro Asn Asp Leu Glu Ser Ile Phe Gly Gly Ser Ala Trp

130 135 140

Gln Tyr Asp Asp Lys Ser Asp Gln Tyr Tyr Leu His Phe Phe Ser Lys
145 150 155 160

Lys Gln Pro Asp Leu Asn Trp Glu Asn Ala Asn Leu Arg Gln Lys Ile
165 170 175

Tyr Asp Met Met Asn Phe Trp Ile Asp Lys Gly Ile Gly Gly Phe Arg
180 185 190

Met Asp Val Ile Asp Met Ile Gly Lys Ile Pro Ala Gln His Ile Val
195 200 205

Ser Asn Gly Pro Lys Leu His Ala Tyr Leu Lys Glu Met Asn Ala Ala
210 215 220

Ser Phe Gly Gln His Asp Leu Leu Thr Val Gly Glu Thr Trp Gly Ala
225 230 235 240

Thr Pro Glu Ile Ala Lys Gln Tyr Ser Asn Pro Val Asn His Glu Leu
245 250 255

Ser Met Ile Phe Gln Phe Glu His Ile Gly Leu Gln His Lys Pro Glu
260 265 270

Ala Pro Lys Trp Asp Tyr Val Lys Glu Leu Asn Val Pro Ala Leu Lys
275 280 285

Thr Ile Phe Asn Lys Trp Gln Thr Glu Leu Glu Leu Gly Gln Gly Trp
290 295 300

Asn Ser Leu Phe Trp Asn Asn His Asp Leu Pro Arg Val Leu Ser Ile
305 310 315 320

Trp Gly Asn Thr Gly Lys Tyr Arg Glu Lys Ser Ala Lys Ala Leu Ala

325 330 335

Ile Leu Leu His Leu Met Arg Gly Thr Pro Tyr Ile Tyr Gln Gly Glu
340 345 350

Glu Ile Gly Met Thr Asn Tyr Pro Phe Lys Asp Leu Asn Glu Leu Asp
355 360 365

Asp Ile Glu Ser Leu Asn Tyr Ala Lys Glu Ala Phe Thr Asn Gly Lys
370 375 380

Ser Met Glu Thr Ile Met Asp Ser Ile Arg Met Ile Gly Arg Asp Asn
385 390 395 400

Ala Arg Thr Pro Met Gln Trp Asp Ala Ser Gln Asn Ala Gly Phe Ser
405 410 415

Thr Ala Asp Lys Thr Trp Leu Pro Val Asn Pro Asn Tyr Lys Asp Ile
420 425 430

Asn Val Gln Ala Ala Leu Lys Asn Ser Asn Ser Ile Phe Tyr Thr Tyr
435 440 445

Gln Gln Leu Ile Gln Leu Arg Lys Glu Asn Asp Trp Leu Val Asp Ala
450 455 460

Asp Phe Glu Leu Leu Pro Thr Ala Asp Lys Val Phe Ala Tyr Leu Arg
465 470 475 480

Lys Val Arg Glu Glu Arg Tyr Leu Ile Val Val Asn Val Ser Asp Gln
485 490 495

Glu Glu Val Leu Glu Ile Asp Val Asp Lys Gln Glu Thr Leu Ile Ser
500 505 510

Asn Thr Asn Glu Ser Ala Ala Leu Ala Asn His Lys Leu Gln Pro Trp

515 520 525

Asp Ala Phe Cys Ile Lys Ile Asn
530 535

<210> 18
<211> 35
<212> DNA
<213> 人工序列

<220>
<223> 引物

<400> 18
tccgagctca tgggtttagg taagaaattg tctgt 35

<210> 19
<211> 37
<212> DNA
<213> 人工序列

<220>
<223> 引物

<400> 19
accttaatta aggttattcc ggatcggttt ccatggc 37

<210> 20
<211> 89
<212> DNA
<213> 人工序列

<220>
<223> 引物

<400> 20
gtgatagctg tcgtaaagct gttaccgact ggcgaagatt tcgccagtca cgtctaccct 60

tgttataccg tgtaggctgg agctgcttc 89

<210> 21
<211> 93
<212> DNA

<213> 人工序列

<220>
<223> 引物

<400> 21
tcagaacagc cccaacggtt tatccgagta gctcaccagc aggcacttgg tttgctggta 60

atgctccagc ttattccgga tcggtttcca tgg 93

<210> 22
<211> 89
<212> DNA
<213> 人工序列

<220>
<223> 引物

<400> 22
atgtcagtac ccgttcaaca tcctatgtat atcgatggac agtttgttac ctggcgtgga 60

gacgcatggg tgtaggctgg agctgcttc 89

<210> 23
<211> 93
<212> DNA
<213> 人工序列

<220>
<223> 引物

<400> 23
ttaagactgt aaataaacca cctgggtctg cagatattca tgcaagccat gtttaccatc 60

tgcgccgcca ttattccgga tcggtttcca tgg 93

<210> 24
<211> 19
<212> PRT
<213> 人工序列

<220>
<223> NPR信号

<400> 24

Met Asn Lys Glu Pro Thr Met Thr Thr Phe Asn Arg Thr Ile Pro Asp
1 5 10 15

Ala Ile Arg

<210> 25
<211> 14
<212> PRT
<213> 人工序列

<220>
<223> 短双歧杆菌信号

<400> 25

Met Ser Leu Thr Ile Ser Leu Pro Gly Val Gln Ala Ser Ala
1 5 10

<210> 26
<211> 62
<212> DNA
<213> 人工序列

<220>
<223> 短双歧杆菌信号

<400> 26
atgaataagg agccaacgat gactactttc aaccgcacaa taattcccga cgccattcgc 60

ac 62

<210> 27
<211> 42
<212> DNA
<213> 人工序列

<220>
<223> NPR信号

<400> 27
atgagtttaa ccatcagtct gccgggtgtt caggctagcg cg 42

<210> 28
<211> 1877
<212> DNA
<213> 短双歧杆菌

<400> 28
atgaataagg agccaacgat gactactttc aaccgcacaa taattcccga cgccattcgc 60

acatgaccgc caacaacctc aatgacgact ggtggaagca ggccgtcgtt taccagattt 120

accccgccag cttcaaggac gttaacggcg acggcatcgg cgacatcgcc ggcgttaccg 180

agaaaatgga ctacctgaaa aacctcggcg tggacgccat ctggctctcc ccgttctacc 240

cctccgatct ggcggacggc ggctacgacg tgatcgacta ccgcaacgtc gaccccgcgac 300

tgggcaccat ggacgacttc gacgccatgg ccaaagccgc gcatgaggcc ggcatcaagg 360

tgatcgtgga catcgtgccc aatcacaccg ccgacaagca cgtgttcttc aaggaagccc 420

tcgccgccga gccccggctcc ccggcgcgcg accgctacat cttccgcgac ggccgcggcg 480

agcacggcga actgccgccc aacgactggc agtccttctt cggcggcccg gcctgggctc 540

gcgtggccga cggccagtgg tatctgcacc tgttcgacaa ggcgcaaccg gacgtcaact 600

ggaagaaccc ggacatccac gaggaattca agaaaaccct gcgcttctgg tccgaccacg 660

gcaccgacgg cttccgcatc gacgtggcgc acggtctggc caaagacctt gaatccaagc 720

cgctggagga gctcggccgc gaatacagcg tggtcggcgt gctgaatcac gacttcagcc 780

atccgctgtt cgaccgccgc gaagtgcacg acatctaccg cgaatggcgc aaggtgttca 840

acgagtacga cccgccgcgc tttgccgtgg ccgaggcgtg ggtggtaccc gagcaccagc 900

acctgtatgc ctcgatggat gagctggggc agtccttcaa cttcgacttt gcgcaggcca 960

gctggtatgc cgatgagttc cgcgcagcca tcgccgcggg tctcaaggcc gctgccgaaa 1020

ccggcggttc caccaccacg tgggtgatga acaaccatga cgtgccgcgc agccccctcc 1080

gctatggtct accgcaggtc aagggcgcgc cttaccacca gctgccgcac gactggctgc 1140

tgcgcaacgg caccacgtat cccgaggatc gcgagcttgg cacccgccgc gcccgcgccg 1200

ccgctttgat ggagctcggc ctgcccggcg ccgcctatat ctatcagggc gaggagctgg 1260

gcctgtttga agtggccgat attccgtggg atcgactgga agatccgacc gctttccaca 1320

ccgctcaggc cacgatggac aagggccgag acggctgccg cgtgccgatt ccgtggaccg 1380

ctgcaaacga accgaccttg gctgatttca gccgcccgat cccggccgat gacggcaccg 1440

gcgagaacca cgtgccgctg tgcgccgccg gccagttcgg cacgggcgct tccttcggct 1500

tctcgccggc tacgcgcgct gagggcgtga cgccggccgc cgacccgcac ctgccgcagc 1560

cgttgtggtt caaggattac gcggtggacg tggagcaggc cgacccggat tcaatgctcg 1620

cgctgtatca tgcggcgttg gcgattcgtc aggagtcgct gaccgccacg cgtgacacca 1680

ccgctgagca ggtggatatg gggccggacg tggtggccta cacccgcgcg gcggttggtg 1740

gccgcacgtt cacctcgatc accaacttcg gcaccgagcc ggtggagctg cctggaggct 1800

ccgtcgtgct gacgtccggc ccgctgaccc ccgacggcca gctccccacc gacacttctg 1860

cgtgggttat caagtag 1877

<210> 29
<211> 624
<212> PRT
<213> 短双歧杆菌

<400> 29

Met Asn Lys Glu Pro Thr Met Thr Thr Phe Asn Arg Thr Ile Ile Pro
1 5 10 15

Asp Ala Ile Arg Met Thr Ala Asn Asn Leu Asn Asp Asp Trp Trp Lys
20 25 30

Gln Ala Val Val Tyr Gln Ile Tyr Pro Arg Ser Phe Lys Asp Val Asn
35 40 45

Gly Asp Gly Ile Gly Asp Ile Ala Gly Val Thr Glu Lys Met Asp Tyr
50 55 60

Leu Lys Asn Leu Gly Val Asp Ala Ile Trp Leu Ser Pro Phe Tyr Pro

65 70 75 80

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

Asp Pro Arg Leu Gly Thr Met Asp Asp Phe Asp Ala Met Ala Lys Ala
100 105 110

Ala His Glu Ala Gly Ile Lys Val Ile Val Asp Ile Val Pro Asn His
115 120 125

Thr Ala Asp Lys His Val Phe Phe Lys Glu Ala Leu Ala Ala Glu Pro
130 135 140

Gly Ser Pro Ala Arg Asp Arg Tyr Ile Phe Arg Asp Gly Arg Gly Glu
145 150 155 160

His Gly Glu Leu Pro Pro Asn Asp Trp Gln Ser Phe Phe Gly Gly Pro
165 170 175

Ala Trp Ala Arg Val Ala Asp Gly Gln Trp Tyr Leu His Leu Phe Asp
180 185 190

Lys Ala Gln Pro Asp Val Asn Trp Lys Asn Pro Asp Ile His Glu Glu
195 200 205

Phe Lys Lys Thr Leu Arg Phe Trp Ser Asp His Gly Thr Asp Gly Phe
210 215 220

Arg Ile Asp Val Ala His Gly Leu Ala Lys Asp Leu Glu Ser Lys Pro
225 230 235 240

Leu Glu Glu Leu Gly Arg Glu Tyr Ser Val Val Gly Val Leu Asn His
245 250 255

Asp Phe Ser His Pro Leu Phe Asp Arg Arg Glu Val His Asp Ile Tyr

260 265 270

Arg Glu Trp Arg Lys Val Phe Asn Glu Tyr Asp Pro Pro Arg Phe Ala
275 280 285

Val Ala Glu Ala Trp Val Val Pro Glu His Gln His Leu Tyr Ala Ser
290 295 300

Met Asp Glu Leu Gly Gln Ser Phe Asn Phe Asp Phe Ala Gln Ala Ser
305 310 315 320

Trp Tyr Ala Asp Glu Phe Arg Ala Ala Ile Ala Ala Gly Leu Lys Ala
325 330 335

Ala Ala Glu Thr Gly Gly Ser Thr Thr Thr Trp Val Met Asn Asn His
340 345 350

Asp Val Pro Arg Ser Pro Ser Arg Tyr Gly Leu Pro Gln Val Lys Gly
355 360 365

Ala Pro Tyr His Gln Leu Pro His Asp Trp Leu Leu Arg Asn Gly Thr
370 375 380

Thr Tyr Pro Glu Asp Arg Glu Leu Gly Thr Arg Arg Ala Arg Ala Ala
385 390 395 400

Ala Leu Met Glu Leu Gly Leu Pro Gly Ala Ala Tyr Ile Tyr Gln Gly
405 410 415

Glu Glu Leu Gly Leu Phe Glu Val Ala Asp Ile Pro Trp Asp Arg Leu
420 425 430

Glu Asp Pro Thr Ala Phe His Thr Ala Gln Ala Thr Met Asp Lys Gly
435 440 445

Arg Asp Gly Cys Arg Val Pro Ile Pro Trp Thr Ala Ala Asn Glu Pro

450 455 460

Thr Leu Ala Asp Phe Ser Arg Pro Ile Pro Ala Asp Asp Gly Thr Gly
465 470 475 480

Glu Asn His Val Pro Leu Cys Ala Ala Gly Gln Phe Gly Thr Gly Ala
485 490 495

Ser Phe Gly Phe Ser Pro Ala Thr Arg Ala Glu Gly Val Thr Pro Ala
500 505 510

Ala Asp Pro His Leu Pro Gln Pro Leu Trp Phe Lys Asp Tyr Ala Val
515 520 525

Asp Val Glu Gln Ala Asp Pro Asp Ser Met Leu Ala Leu Tyr His Ala
530 535 540

Ala Leu Ala Ile Arg Gln Glu Ser Leu Thr Ala Thr Arg Asp Thr Thr
545 550 555 560

Ala Glu Gln Val Asp Met Gly Pro Asp Val Val Ala Tyr Thr Arg Ala
565 570 575

Ala Val Gly Gly Arg Thr Phe Thr Ser Ile Thr Asn Phe Gly Thr Glu
580 585 590

Pro Val Glu Leu Pro Gly Gly Ser Val Val Leu Thr Ser Gly Pro Leu
595 600 605

Thr Pro Asp Gly Gln Leu Pro Thr Asp Thr Ser Ala Trp Val Ile Lys
610 615 620

<210> 30
<211> 1611
<212> DNA
<213> 短双歧杆菌

<400> 30

atgcaaaaac attggtggca caaggcaact gtttatcaaa tttatccaaa atcttttatg 60

gatacaaatg gtgatggaat tggtgatctc aaaggtatta cgagtaaatt ggattatttg 120

caaaagttag gggttatggc tatttggcta tctccagttt atgatagccc catggatgac 180

aatggctatg acattgcgaa ctatgaagca attgcggata tttttggcaa tatggctgat 240

atggataatt tgctgacgca ggcaaaaatg cgcgacataa aaatcattat ggatctagtg 300

gttaatcata cctcagatga acatacttgg tttattgaag cacgtgagca tccagacagt 360

tctgaacgcg attattatat ttggtgtgac cagccaaatg atttggaatc tattttcggt 420

ggttctgctt ggcagtatga tgataagtcc gatcaatatt atttgcattt ttttagtaag 480

aagcagccag atctaaactg ggaaaacgca aacttacgtc agaagattta tgatatgatg 540

aatttctgga ttgataaagg tattggcggc tttcggatgg acgtcattga tatgattggg 600

aaaattcctg ctcagcatat tgtcagtaac ggaccaaaat tgcatgctta tcttaaggag 660

atgaatgccg ctagttttgg tcaacatgat ctgctgactg tgggggaaac ttggggagca 720

acgcctgaga ttgcgaagca atattcaaat ccagtcaatc acgaactctc tatgattttt 780

caatttgaac atattggtct tcagcataaa ccagaagctc ctaaatggga ttatgtgaag 840

gaacttaatg ttcctgcttt aaaaacaatc tttaataaat ggcagactga gttggaatta 900

ggacaggggt ggaattcgtt attctggaat aaccatgacc tgcctcgtgt tttatcaatc 960

tggggaaata cgggcaaata tcgtgagaag tctgctaaag cactggctat tcttcttcac 1020

cttatgcgtg ggacacctta tatttatcaa ggtgaagaga ttgggatgac caattatcct 1080

tttaaagatt taaatgaact tgatgatatt gaatcactta attatgctaa ggaagctttt 1140

acaaatggta agtctatgga aactatcatg gacagtattc gtatgattgg ccgtgataat 1200

gccagaacac ctatgcaatg ggatgcttct caaaatgccg gattttcaac agcggataaa 1260

acatggctgc cagttaatcc aaactataaa gacatcaatg ttcaagcagc tctgaaaaat 1320

tccaattcta tcttttacac ctatcaacaa ctcattcagc ttcgaaaaga aaatgattgg 1380

ctagtagatg ccgattttga attgctccct acagcggaca aagtatttgc ctatttacga 1440

aaggtaagag aagaaaggta tcttatagtg gtcaatgttt cagatcagga agaagttcta 1500

gagattgatg ttgacaaaca agaaactctc attagcaata caaatgaaag cgctgctctt 1560

gccaatcaca aactccagcc ttgggatgct ttttgtatta agataaacta a 1611

<210> 31
<211> 536
<212> PRT
<213> 短双歧杆菌

<400> 31

Met Gln Lys His Trp Trp His Lys Ala Thr Val Tyr Gln Ile Tyr Pro
1 5 10 15

Lys Ser Phe Met Asp Thr Asn Gly Asp Gly Ile Gly Asp Leu Lys Gly
20 25 30

Ile Thr Ser Lys Leu Asp Tyr Leu Gln Lys Leu Gly Val Met Ala Ile
35 40 45

Trp Leu Ser Pro Val Tyr Asp Ser Pro Met Asp Asp Asn Gly Tyr Asp
50 55 60

Ile Ala Asn Tyr Glu Ala Ile Ala Asp Ile Phe Gly Asn Met Ala Asp
65 70 75 80

Met Asp Asn Leu Leu Thr Gln Ala Lys Met Arg Asp Ile Lys Ile Ile
85 90 95

Met Asp Leu Val Val Asn His Thr Ser Asp Glu His Thr Trp Phe Ile
100 105 110

Glu Ala Arg Glu His Pro Asp Ser Ser Glu Arg Asp Tyr Tyr Ile Trp
115 120 125

Cys Asp Gln Pro Asn Asp Leu Glu Ser Ile Phe Gly Gly Ser Ala Trp

130 135 140

Gln Tyr Asp Asp Lys Ser Asp Gln Tyr Tyr Leu His Phe Phe Ser Lys
145 150 155 160

Lys Gln Pro Asp Leu Asn Trp Glu Asn Ala Asn Leu Arg Gln Lys Ile
165 170 175

Tyr Asp Met Met Asn Phe Trp Ile Asp Lys Gly Ile Gly Gly Phe Arg
180 185 190

Met Asp Val Ile Asp Met Ile Gly Lys Ile Pro Ala Gln His Ile Val
195 200 205

Ser Asn Gly Pro Lys Leu His Ala Tyr Leu Lys Glu Met Asn Ala Ala
210 215 220

Ser Phe Gly Gln His Asp Leu Leu Thr Val Gly Glu Thr Trp Gly Ala
225 230 235 240

Thr Pro Glu Ile Ala Lys Gln Tyr Ser Asn Pro Val Asn His Glu Leu
245 250 255

Ser Met Ile Phe Gln Phe Glu His Ile Gly Leu Gln His Lys Pro Glu
260 265 270

Ala Pro Lys Trp Asp Tyr Val Lys Glu Leu Asn Val Pro Ala Leu Lys
275 280 285

Thr Ile Phe Asn Lys Trp Gln Thr Glu Leu Glu Leu Gly Gln Gly Trp
290 295 300

Asn Ser Leu Phe Trp Asn Asn His Asp Leu Pro Arg Val Leu Ser Ile
305 310 315 320

Trp Gly Asn Thr Gly Lys Tyr Arg Glu Lys Ser Ala Lys Ala Leu Ala

325 330 335

Ile Leu Leu His Leu Met Arg Gly Thr Pro Tyr Ile Tyr Gln Gly Glu
340 345 350

Glu Ile Gly Met Thr Asn Tyr Pro Phe Lys Asp Leu Asn Glu Leu Asp
355 360 365

Asp Ile Glu Ser Leu Asn Tyr Ala Lys Glu Ala Phe Thr Asn Gly Lys
370 375 380

Ser Met Glu Thr Ile Met Asp Ser Ile Arg Met Ile Gly Arg Asp Asn
385 390 395 400

Ala Arg Thr Pro Met Gln Trp Asp Ala Ser Gln Asn Ala Gly Phe Ser
405 410 415

Thr Ala Asp Lys Thr Trp Leu Pro Val Asn Pro Asn Tyr Lys Asp Ile
420 425 430

Asn Val Gln Ala Ala Leu Lys Asn Ser Asn Ser Ile Phe Tyr Thr Tyr
435 440 445

Gln Gln Leu Ile Gln Leu Arg Lys Glu Asn Asp Trp Leu Val Asp Ala
450 455 460

Asp Phe Glu Leu Leu Pro Thr Ala Asp Lys Val Phe Ala Tyr Leu Arg
465 470 475 480

Lys Val Arg Glu Glu Arg Tyr Leu Ile Val Val Asn Val Ser Asp Gln
485 490 495

Glu Glu Val Leu Glu Ile Asp Val Asp Lys Gln Glu Thr Leu Ile Ser
500 505 510

Asn Thr Asn Glu Ser Ala Ala Leu Ala Asn His Lys Leu Gln Pro Trp

515 520 525

Asp Ala Phe Cys Ile Lys Ile Asn
530 535

<210> 32
<211> 1880
<212> DNA
<213> 短双歧杆菌

<400> 32
atgaataagg agccaacgat gactactttc aaccgcacaa taattcccga cgccattcgc 60

acatgatgac ctctttcaac cgtgaacccc tgcccgacgc cgtccgcacg aatggcgcct 120

ccccgaaccc gtggtggtcg aacgccgtcg tctaccagat ttacccacgt tccttccagg 180

acacgaacgg cgatggtttc ggagatctta agggcattac ttcccgcctc gactatcttg 240

ccgacctcgg cgtggatgtg ctgtggctct ccccggtcta caggtccccg caagacgaca 300

acggctacga catctccgac taccgggaca tcgacccgct gttcggcacg ctcgacgaca 360

tggacgagct gctcgccgaa gcgcacaagc gcggcctcaa gatcgtgatg gacctggtcg 420

tcaaccacac ctccgacgag cacgcgtggt tcgaggcgtc gaaggacaag gacgacccgc 480

acgccgactg gtactggtgg cgtcccgccc gccccggcca cgagcccggc acgcccggcg 540

ccgagccgaa ccagtggggc tcctacttcg gcggctccgc atgggaatat tgccccgagc 600

gtggtgagta ctatctccac cagttctcga agaagcagcc ggacctcaac tgggagaacc 660

cggccgtacg ccgagccgtg tacgacatga tgaactggtg gctcgatcgc ggcatcgacg 720

gcttccgcat ggacgtcatc accctgatct ccaagcgcac ggatgcaaac ggcaggctgc 780

ccggcgagac cggttccgag ctgcaggacc tgccggtggg ggaggagggc tactccaacc 840

cgaacccgtt ctgcgccgac ggtccgcgtc aggacgagtt cctcgccgag atgcgccgcg 900

aggtgttcga cgggcgtgac ggcttcctca ccgtcggcga ggccccggc atcaccgccg 960

aacgcaacga gcacatcacc gaccccgcca acggcgagct ggatatgctc ttcctgttcg 1020

aacacatggg cgtcgaccaa acccccgaat cgaaatggga cgacaaacca tggacgccgg 1080

ccgacctcga aaccaagctc gccgaacaac aggacgccat cgcccgacac ggctgggcca 1140

gcctgttcct cgacaaccac gaccagccgc gtgtcgtctc ccgttggggc gacgacacca 1200

gcaagaccgg ccgcatccgc tccgccaagg cgctcgcgct gctgctgcac atgcaccgcg 1260

gcactccgta tgtctaccag ggcgaggagc tcggcatgac caatgcgcac ttcacctcgc 1320

tcgaccagta ccgcgacctc gaatccatca acgcctacca tcaacgcgtc gaggaaaccg 1380

ggatacggac atcggagacc atgatgcgat ccctcgcccg atacggcagg gacaacgcgc 1440

gcaccccgat gcaatgggac gactccacct acgccggctt caccatgccc gacgcccccgg 1500

tcgaaccctg gatcgccgtc aacccgaacc acacggagat caacgccgcc gacgagatcg 1560

acgacccccga ctccgtgtac tcgttccaca aacggctcat cgccctgcgt cacaccgacc 1620

ccgtggtcgc cgccggcgac taccgacgcg tggaaaccgg aaacgaccgg atcatcgcct 1680

tcaccagaac cctcgacgag cgaaccatcc tcaccgtcat caacctctcg cccacacagg 1740

ccgcaccggc cggagaactg gaaacgatgc ccgacggcac gatcctcatc gccaacacgg 1800

acgatcccgt aggaaacctg aaaaccacgg gaacactcgg accatgggag gcgttcgcca 1860

tggaaaccga tccggaataa 1880

<210> 33
<211> 625
<212> PRT
<213> 短双歧杆菌

<400> 33

Met Asn Lys Glu Pro Thr Met Thr Thr Phe Asn Arg Thr Ile Ile Pro
1 5 10 15

Asp Ala Ile Arg Met Met Thr Ser Phe Asn Arg Glu Pro Leu Pro Asp
20 25 30

Ala Val Arg Thr Asn Gly Ala Ser Pro Asn Pro Trp Trp Ser Asn Ala
35 40 45

Val Val Tyr Gln Ile Tyr Pro Arg Ser Phe Gln Asp Thr Asn Gly Asp
50 55 60

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

Asp Leu Gly Val Asp Val Leu Trp Leu Ser Pro Val Tyr Arg Ser Pro
85 90 95

Gln Asp Asp Asn Gly Tyr Asp Ile Ser Asp Tyr Arg Asp Ile Asp Pro
100 105 110

Leu Phe Gly Thr Leu Asp Asp Met Asp Glu Leu Leu Ala Glu Ala His
115 120 125

Lys Arg Gly Leu Lys Ile Val Met Asp Leu Val Val Asn His Thr Ser
130 135 140

Asp Glu His Ala Trp Phe Glu Ala Ser Lys Asp Lys Asp Asp Pro His
145 150 155 160

Ala Asp Trp Tyr Trp Trp Arg Pro Ala Arg Pro Gly His Glu Pro Gly
165 170 175

Thr Pro Gly Ala Glu Pro Asn Gln Trp Gly Ser Tyr Phe Gly Gly Ser
180 185 190

Ala Trp Glu Tyr Cys Pro Glu Arg Gly Glu Tyr Tyr Leu His Gln Phe
195 200 205

Ser Lys Lys Gln Pro Asp Leu Asn Trp Glu Asn Pro Ala Val Arg Arg
210 215 220

Ala Val Tyr Asp Met Met Asn Trp Trp Leu Asp Arg Gly Ile Asp Gly
225 230 235 240

Phe Arg Met Asp Val Ile Thr Leu Ile Ser Lys Arg Thr Asp Ala Asn
245 250 255

Gly Arg Leu Pro Gly Glu Thr Gly Ser Glu Leu Gln Asp Leu Pro Val
260 265 270

Gly Glu Glu Gly Tyr Ser Asn Pro Asn Pro Phe Cys Ala Asp Gly Pro
275 280 285

Arg Gln Asp Glu Phe Leu Ala Glu Met Arg Arg Glu Val Phe Asp Gly
290 295 300

Arg Asp Gly Phe Leu Thr Val Gly Glu Ala Pro Gly Ile Thr Ala Glu
305 310 315 320

Arg Asn Glu His Ile Thr Asp Pro Ala Asn Gly Glu Leu Asp Met Leu
325 330 335

Phe Leu Phe Glu His Met Gly Val Asp Gln Thr Pro Glu Ser Lys Trp
340 345 350

Asp Asp Lys Pro Trp Thr Pro Ala Asp Leu Glu Thr Lys Leu Ala Glu
355 360 365

Gln Gln Asp Ala Ile Ala Arg His Gly Trp Ala Ser Leu Phe Leu Asp
370 375 380

Asn His Asp Gln Pro Arg Val Val Ser Arg Trp Gly Asp Asp Thr Ser
385 390 395 400

Lys Thr Gly Arg Ile Arg Ser Ala Lys Ala Leu Ala Leu Leu Leu His
405 410 415

Met His Arg Gly Thr Pro Tyr Val Tyr Gln Gly Glu Glu Leu Gly Met
420 425 430

Thr Asn Ala His Phe Thr Ser Leu Asp Gln Tyr Arg Asp Leu Glu Ser
435 440 445

Ile Asn Ala Tyr His Gln Arg Val Glu Glu Thr Gly Ile Arg Thr Ser
450 455 460

Glu Thr Met Met Arg Ser Leu Ala Arg Tyr Gly Arg Asp Asn Ala Arg
465 470 475 480

Thr Pro Met Gln Trp Asp Asp Ser Thr Tyr Ala Gly Phe Thr Met Pro
485 490 495

Asp Ala Pro Val Glu Pro Trp Ile Ala Val Asn Pro Asn His Thr Glu
500 505 510

Ile Asn Ala Ala Asp Glu Ile Asp Asp Pro Asp Ser Val Tyr Ser Phe
515 520 525

His Lys Arg Leu Ile Ala Leu Arg His Thr Asp Pro Val Val Ala Ala
530 535 540

Gly Asp Tyr Arg Arg Val Glu Thr Gly Asn Asp Arg Ile Ile Ala Phe
545 550 555 560

Thr Arg Thr Leu Asp Glu Arg Thr Ile Leu Thr Val Ile Asn Leu Ser
565 570 575

Pro Thr Gln Ala Ala Pro Ala Gly Glu Leu Glu Thr Met Pro Asp Gly
580 585 590

Thr Ile Leu Ile Ala Asn Thr Asp Asp Pro Val Gly Asn Leu Lys Thr
595 600 605

Thr Gly Thr Leu Gly Pro Trp Glu Ala Phe Ala Met Glu Thr Asp Pro
610 615 620

Glu
625

<210> 34
<211> 1673
<212> DNA
<213> 短双歧杆菌 + 变异链球菌

<400> 34
atgaataagg agccaacgat gactactttc aaccgcacaa taattcccga cgccattcgc 60

acatgcaaaa acattggtgg cacaaggcaa ctgtttatca aatttatcca aaatctttta 120

tggatacaaa tggtgatgga attggtgatc tcaaaggtat tacgagtaaa ttggattatt 180

tgcaaaagtt aggggttatg gctatttggc tatctccagt ttatgatagc cccatggatg 240

acaatggcta tgacattgcg aactatgaag caattgcgga tatttttggc aatatggctg 300

atatggataa tttgctgacg caggcaaaaa tgcgcgacat aaaaatcatt atggatctag 360

tggttaatca tacctcagat gaacatactt ggtttattga agcacgtgag catccagaca 420

gttctgaacg cgattattat atttggtgtg accagccaaa tgatttggaa tctatttttcg 480

gtggttctgc ttggcagtat gatgataagt ccgatcaata ttatttgcat ttttttagta 540

agaagcagcc agatctaaac tgggaaaacg caaacttacg tcagaagatt tatgatatga 600

tgaatttctg gattgataaa ggtattggcg gctttcggat ggacgtcatt gatatgattg 660

ggaaaattcc tgctcagcat attgtcagta acggaccaaa attgcatgct tatcttaagg 720

agatgaatgc cgctagtttt ggtcaacatg atctgctgac tgtgggggaa acttggggag 780

caacgcctga gattgcgaag caatattcaa atccagtcaa tcacgaactc tctatgattt 840

ttcaatttga acatattggt cttcagcata aaccagaagc tcctaaatgg gattatgtga 900

aggaacttaa tgttcctgct ttaaaaacaa tctttaataa atggcagact gagttggaat 960

taggacaggg gtggaattcg ttattctgga ataaccatga cctgcctcgt gttttatcaa 1020

tctggggaaa tacgggcaaa tatcgtgaga agtctgctaa agcactggct attcttcttc 1080

accttatgcg tgggacacct tatatttatc aaggtgaaga gattgggatg accaattatc 1140

cttttaaaga tttaaatgaa cttgatgata ttgaatcact taattatgct aaggaagctt 1200

ttacaaatgg taagtctatg gaaactatca tggacagtat tcgtatgatt ggccgtgata 1260

atgccagaac acctatgcaa tgggatgctt ctcaaaatgc cggattttca acagcggata 1320

aaacatggct gccagttaat ccaaactata aagacatcaa tgttcaagca gctctgaaaa 1380

attccaattc tatcttttac acctatcaac aactcattca gcttcgaaaa gaaaatgatt 1440

ggctagtaga tgccgatttt gaattgctcc ctacagcgga caaagtattt gcctatttac 1500

gaaaggtaag agaagaaagg tatcttatag tggtcaatgt ttcagatcag gaagaagttc 1560

tagagattga tgttgacaaa caagaaactc tcattagcaa tacaaatgaa agcgctgctc 1620

ttgccaatca caaactccag ccttgggatg ctttttgtat taagataaac taa 1673

<210> 35
<211> 556
<212> PRT
<213> 短双歧杆菌 + 变异链球菌

<400> 35

Met Asn Lys Glu Pro Thr Met Thr Thr Phe Asn Arg Thr Ile Ile Pro
1 5 10 15

Asp Ala Ile Arg Met Gln Lys His Trp Trp His Lys Ala Thr Val Tyr
20 25 30

Gln Ile Tyr Pro Lys Ser Phe Met Asp Thr Asn Gly Asp Gly Ile Gly
35 40 45

Asp Leu Lys Gly Ile Thr Ser Lys Leu Asp Tyr Leu Gln Lys Leu Gly
50 55 60

Val Met Ala Ile Trp Leu Ser Pro Val Tyr Asp Ser Pro Met Asp Asp
65 70 75 80

Asn Gly Tyr Asp Ile Ala Asn Tyr Glu Ala Ile Ala Asp Ile Phe Gly

85 90 95

Asn Met Ala Asp Met Asp Asn Leu Leu Thr Gln Ala Lys Met Arg Asp
100 105 110

Ile Lys Ile Ile Met Asp Leu Val Val Asn His Thr Ser Asp Glu His
115 120 125

Thr Trp Phe Ile Glu Ala Arg Glu His Pro Asp Ser Ser Glu Arg Asp
130 135 140

Tyr Tyr Ile Trp Cys Asp Gln Pro Asn Asp Leu Glu Ser Ile Phe Gly
145 150 155 160

Gly Ser Ala Trp Gln Tyr Asp Asp Lys Ser Asp Gln Tyr Tyr Leu His
165 170 175

Phe Phe Ser Lys Lys Gln Pro Asp Leu Asn Trp Glu Asn Ala Asn Leu
180 185 190

Arg Gln Lys Ile Tyr Asp Met Met Asn Phe Trp Ile Asp Lys Gly Ile
195 200 205

Gly Gly Phe Arg Met Asp Val Ile Asp Met Ile Gly Lys Ile Pro Ala
210 215 220

Gln His Ile Val Ser Asn Gly Pro Lys Leu His Ala Tyr Leu Lys Glu
225 230 235 240

Met Asn Ala Ala Ser Phe Gly Gln His Asp Leu Leu Thr Val Gly Glu
245 250 255

Thr Trp Gly Ala Thr Pro Glu Ile Ala Lys Gln Tyr Ser Asn Pro Val
260 265 270

Asn His Glu Leu Ser Met Ile Phe Gln Phe Glu His Ile Gly Leu Gln

275 280 285

His Lys Pro Glu Ala Pro Lys Trp Asp Tyr Val Lys Glu Leu Asn Val
290 295 300

Pro Ala Leu Lys Thr Ile Phe Asn Lys Trp Gln Thr Glu Leu Glu Leu
305 310 315 320

Gly Gln Gly Trp Asn Ser Leu Phe Trp Asn Asn His Asp Leu Pro Arg
325 330 335

Val Leu Ser Ile Trp Gly Asn Thr Gly Lys Tyr Arg Glu Lys Ser Ala
340 345 350

Lys Ala Leu Ala Ile Leu Leu His Leu Met Arg Gly Thr Pro Tyr Ile
355 360 365

Tyr Gln Gly Glu Glu Ile Gly Met Thr Asn Tyr Pro Phe Lys Asp Leu
370 375 380

Asn Glu Leu Asp Asp Ile Glu Ser Leu Asn Tyr Ala Lys Glu Ala Phe
385 390 395 400

Thr Asn Gly Lys Ser Met Glu Thr Ile Met Asp Ser Ile Arg Met Ile
405 410 415

Gly Arg Asp Asn Ala Arg Thr Pro Met Gln Trp Asp Ala Ser Gln Asn
420 425 430

Ala Gly Phe Ser Thr Ala Asp Lys Thr Trp Leu Pro Val Asn Pro Asn
435 440 445

Tyr Lys Asp Ile Asn Val Gln Ala Ala Leu Lys Asn Ser Asn Ser Ile
450 455 460

Phe Tyr Thr Tyr Gln Gln Leu Ile Gln Leu Arg Lys Glu Asn Asp Trp

465 470 475 480

Leu Val Asp Ala Asp Phe Glu Leu Leu Pro Thr Ala Asp Lys Val Phe
485 490 495

Ala Tyr Leu Arg Lys Val Arg Glu Glu Arg Tyr Leu Ile Val Val Asn
500 505 510

Val Ser Asp Gln Glu Glu Val Leu Glu Ile Asp Val Asp Lys Gln Glu
515 520 525

Thr Leu Ile Ser Asn Thr Asn Glu Ser Ala Ala Leu Ala Asn His Lys
530 535 540

Leu Gln Pro Trp Asp Ala Phe Cys Ile Lys Ile Asn
545 550 555

<210> 36
<211> 1857
<212> DNA
<213> 芽孢杆菌属和短双歧杆菌

<400> 36
atgagtttaa ccatcagtct gccgggtgtt caggctagcg cgatgaccgc caacaacctc 60

aatgacgact ggtggaagca ggccgtcgtt taccagattt accccgccag cttcaaggac 120

gttaacggcg acggcatcgg cgacatcgcc ggcgttaccg agaaaatgga ctacctgaaa 180

aacctcggcg tggacgccat ctggctctcc ccgttctacc cctccgatct ggcggacggc 240

ggctacgacg tgatcgacta ccgcaacgtc gacccgcgac tgggcaccat ggacgacttc 300

gacgccatgg ccaaagccgc gcatgaggcc ggcatcaagg tgatcgtgga catcgtgccc 360

aatcacaccg ccgacaagca cgtgttcttc aaggaagccc tcgccgccga gccccggctcc 420

ccggcgcgcg accgctacat cttccgcgac ggccgcggcg agcacggcga actgccgccc 480

aacgactggc agtccttctt cggcggcccg gcctgggctc gcgtggccga cggccagtgg 540

tatctgcacc tgttcgacaa ggcgcaaccg gacgtcaact ggaagaaccc ggacatccac 600

gaggaattca agaaaaccct gcgcttctgg tccgaccacg gcaccgacgg cttccgcatc 660

gacgtggcgc acggtctggc caaagacctt gaatccaagc cgctggagga gctcggccgc 720

gaatacagcg tggtcggcgt gctgaatcac gacttcagcc atccgctgtt cgaccgccgc 780

gaagtgcacg acatctaccg cgaatggcgc aaggtgttca acgagtacga cccgccgcgc 840

tttgccgtgg ccgaggcgtg ggtggtaccc gagcaccagc acctgtatgc ctcgatggat 900

gagctggggc agtccttcaa cttcgacttt gcgcaggcca gctggtatgc cgatgagttc 960

cgcgcagcca tcgccgcggg tctcaaggcc gctgccgaaa ccggcggttc caccaccacg 1020

tgggtgatga acaaccatga cgtgccgcgc agcccctccc gctatggtct accgcaggtc 1080

aagggcgcgc cttaccacca gctgccgcac gactggctgc tgcgcaacgg caccacgtat 1140

cccgaggatc gcgagcttgg cacccgccgc gcccgcgccg ccgctttgat ggagctcggc 1200

ctgcccggcg ccgcctatat ctatcagggc gaggagctgg gcctgtttga agtggccgat 1260

attccgtggg atcgactgga agatccgacc gctttccaca ccgctcaggc cacgatggac 1320

aagggccgag acggctgccg cgtgccgatt ccgtggaccg ctgcaaacga accgaccttg 1380

gctgatttca gccgcccgat cccggccgat gacggcaccg gcgagaacca cgtgccgctg 1440

tgcgcccgcg gccagttcgg cacgggcgct tccttcggct tctcgccggc tacgcgcgct 1500

gagggcgtga cgccggccgc cgacccgcac ctgccgcagc cgttgtggtt caaggattac 1560

gcggtggacg tggagcaggc cgacccggat tcaatgctcg cgctgtatca tgcggcgttg 1620

gcgattcgtc aggagtcgct gaccgccacg cgtgacacca ccgctgagca ggtggatatg 1680

gggccggacg tggtggccta cacccgcgcg gcggttggtg gccgcacgtt cacctcgatc 1740

accaacttcg gcaccgagcc ggtggagctg cctggaggct ccgtcgtgct gacgtccggc 1800

ccgctgaccc ccgacggcca gctccccacc gacacttctg cgtgggttat caagtag 1857

<210> 37
<211> 618
<212> PRT
<213> 芽孢杆菌属和短双歧杆菌

<400> 37

Met Ser Leu Thr Ile Ser Leu Pro Gly Val Gln Ala Ser Ala Met Thr
1 5 10 15

Ala Asn Asn Leu Asn Asp Asp Trp Trp Lys Gln Ala Val Val Tyr Gln
20 25 30

Ile Tyr Pro Arg Ser Phe Lys Asp Val Asn Gly Asp Gly Ile Gly Asp
35 40 45

Ile Ala Gly Val Thr Glu Lys Met Asp Tyr Leu Lys Asn Leu Gly Val
50 55 60

Asp Ala Ile Trp Leu Ser Pro Phe Tyr Pro Ser Asp Leu Ala Asp Gly
65 70 75 80

Gly Tyr Asp Val Ile Asp Tyr Arg Asn Val Asp Pro Arg Leu Gly Thr
85 90 95

Met Asp Asp Phe Asp Ala Met Ala Lys Ala Ala His Glu Ala Gly Ile
100 105 110

Lys Val Ile Val Asp Ile Val Pro Asn His Thr Ala Asp Lys His Val
115 120 125

Phe Phe Lys Glu Ala Leu Ala Ala Glu Pro Gly Ser Pro Ala Arg Asp
130 135 140

Arg Tyr Ile Phe Arg Asp Gly Arg Gly Glu His Gly Glu Leu Pro Pro
145 150 155 160

Asn Asp Trp Gln Ser Phe Phe Gly Gly Pro Ala Trp Ala Arg Val Ala
165 170 175

Asp Gly Gln Trp Tyr Leu His Leu Phe Asp Lys Ala Gln Pro Asp Val

180 185 190

Asn Trp Lys Asn Pro Asp Ile His Glu Glu Phe Lys Lys Thr Leu Arg
195 200 205

Phe Trp Ser Asp His Gly Thr Asp Gly Phe Arg Ile Asp Val Ala His
210 215 220

Gly Leu Ala Lys Asp Leu Glu Ser Lys Pro Leu Glu Glu Leu Gly Arg
225 230 235 240

Glu Tyr Ser Val Val Gly Val Leu Asn His Asp Phe Ser His Pro Leu
245 250 255

Phe Asp Arg Arg Glu Val His Asp Ile Tyr Arg Glu Trp Arg Lys Val
260 265 270

Phe Asn Glu Tyr Asp Pro Pro Arg Phe Ala Val Ala Glu Ala Trp Val
275 280 285

Val Pro Glu His Gln His Leu Tyr Ala Ser Met Asp Glu Leu Gly Gln
290 295 300

Ser Phe Asn Phe Asp Phe Ala Gln Ala Ser Trp Tyr Ala Asp Glu Phe
305 310 315 320

Arg Ala Ala Ile Ala Ala Gly Leu Lys Ala Ala Ala Glu Thr Gly Gly
325 330 335

Ser Thr Thr Thr Trp Val Met Asn Asn His Asp Val Pro Arg Ser Pro
340 345 350

Ser Arg Tyr Gly Leu Pro Gln Val Lys Gly Ala Pro Tyr His Gln Leu
355 360 365

Pro His Asp Trp Leu Leu Arg Asn Gly Thr Thr Tyr Pro Glu Asp Arg

370 375 380

Glu Leu Gly Thr Arg Arg Ala Arg Ala Ala Ala Leu Met Glu Leu Gly
385 390 395 400

Leu Pro Gly Ala Ala Tyr Ile Tyr Gln Gly Glu Glu Leu Gly Leu Phe
405 410 415

Glu Val Ala Asp Ile Pro Trp Asp Arg Leu Glu Asp Pro Thr Ala Phe
420 425 430

His Thr Ala Gln Ala Thr Met Asp Lys Gly Arg Asp Gly Cys Arg Val
435 440 445

Pro Ile Pro Trp Thr Ala Ala Asn Glu Pro Thr Leu Ala Asp Phe Ser
450 455 460

Arg Pro Ile Pro Ala Asp Asp Gly Thr Gly Glu Asn His Val Pro Leu
465 470 475 480

Cys Ala Ala Gly Gln Phe Gly Thr Gly Ala Ser Phe Gly Phe Ser Pro
485 490 495

Ala Thr Arg Ala Glu Gly Val Thr Pro Ala Ala Asp Pro His Leu Pro
500 505 510

Gln Pro Leu Trp Phe Lys Asp Tyr Ala Val Asp Val Glu Gln Ala Asp
515 520 525

Pro Asp Ser Met Leu Ala Leu Tyr His Ala Ala Leu Ala Ile Arg Gln
530 535 540

Glu Ser Leu Thr Ala Thr Arg Asp Thr Thr Ala Glu Gln Val Asp Met
545 550 555 560

Gly Pro Asp Val Val Ala Tyr Thr Arg Ala Ala Val Gly Gly Arg Thr

565 570 575

Phe Thr Ser Ile Thr Asn Phe Gly Thr Glu Pro Val Glu Leu Pro Gly
580 585 590

Gly Ser Val Val Leu Thr Ser Gly Pro Leu Thr Pro Asp Gly Gln Leu
595 600 605

Pro Thr Asp Thr Ser Ala Trp Val Ile Lys
610 615

<210> 38
<211> 1653
<212> DNA
<213> 芽孢杆菌属和短双歧杆菌

<400> 38
atgagtttaa ccatcagtct gccgggtgtt caggctagcg cgatgcaaaa acattggtgg 60

cacaaggcaa ctgtttatca aatttatcca aaatctttta tggatacaaa tggtgatgga 120

attggtgatc tcaaaggtat tacgagtaaa ttggattatt tgcaaaagtt aggggttatg 180

gctatttggc tatctccagt ttatgatagc cccatggatg acaatggcta tgacattgcg 240

aactatgaag caattgcgga tatttttggc aatatggctg atatggataa tttgctgacg 300

caggcaaaaa tgcgcgacat aaaaatcatt atggatctag tggttaatca tacctcagat 360

gaacatactt ggtttattga agcacgtgag catccagaca gttctgaacg cgattattat 420

atttggtgtg accagccaaa tgatttggaa tctattttcg gtggttctgc ttggcagtat 480

gatgataagt ccgatcaata ttatttgcat ttttttagta agaagcagcc agatctaaac 540

tgggaaaacg caaacttacg tcagaagatt tatgatatga tgaatttctg gattgataaa 600

ggtattggcg gctttcggat ggacgtcatt gatatgattg ggaaaattcc tgctcagcat 660

attgtcagta acggaccaaa attgcatgct tatcttaagg agatgaatgc cgctagtttt 720

ggtcaacatg atctgctgac tgtgggggaa acttggggag caacgcctga gattgcgaag 780

caatattcaa atccagtcaa tcacgaactc tctatgattt ttcaatttga acatattggt 840

cttcagcata aaccagaagc tcctaaatgg gattatgtga aggaacttaa tgttcctgct 900

ttaaaaacaa tctttaataa atggcagact gagttggaat taggacaggg gtggaattcg 960

ttattctgga ataaccatga cctgcctcgt gttttatcaa tctggggaaa tacgggcaaa 1020

tatcgtgaga agtctgctaa agcactggct attcttcttc accttatgcg tgggacacct 1080

tatatttatc aaggtgaaga gattgggatg accaattatc cttttaaaga tttaaatgaa 1140

cttgatgata ttgaatcact taattatgct aaggaagctt ttacaaatgg taagtctatg 1200

gaaactatca tggacagtat tcgtatgatt ggccgtgata atgccagaac acctatgcaa 1260

tgggatgctt ctcaaaatgc cggattttca acagcggata aaacatggct gccagttaat 1320

ccaaactata aagacatcaa tgttcaagca gctctgaaaa attccaattc tatcttttac 1380

acctatcaac aactcattca gcttcgaaaa gaaaatgatt ggctagtaga tgccgatttt 1440

gaattgctcc ctacagcgga caaagtattt gcctatttac gaaaggtaag agaagaaagg 1500

tatcttatag tggtcaatgt ttcagatcag gaagaagttc tagagattga tgttgacaaa 1560

caagaaactc tcattagcaa tacaaatgaa agcgctgctc ttgccaatca caaactccag 1620

ccttgggatg ctttttgtat taagataaac taa 1653

<210> 39
<211> 550
<212> PRT
<213> 芽孢杆菌属和短双歧杆菌

<400> 39

Met Ser Leu Thr Ile Ser Leu Pro Gly Val Gln Ala Ser Ala Met Gln
1 5 10 15

Lys His Trp Trp His Lys Ala Thr Val Tyr Gln Ile Tyr Pro Lys Ser
20 25 30

Phe Met Asp Thr Asn Gly Asp Gly Ile Gly Asp Leu Lys Gly Ile Thr
35 40 45

Ser Lys Leu Asp Tyr Leu Gln Lys Leu Gly Val Met Ala Ile Trp Leu
50 55 60

Ser Pro Val Tyr Asp Ser Pro Met Asp Asp Asn Gly Tyr Asp Ile Ala
65 70 75 80

Asn Tyr Glu Ala Ile Ala Asp Ile Phe Gly Asn Met Ala Asp Met Asp
85 90 95

Asn Leu Leu Thr Gln Ala Lys Met Arg Asp Ile Lys Ile Ile Met Asp
100 105 110

Leu Val Val Asn His Thr Ser Asp Glu His Thr Trp Phe Ile Glu Ala
115 120 125

Arg Glu His Pro Asp Ser Ser Glu Arg Asp Tyr Tyr Ile Trp Cys Asp
130 135 140

Gln Pro Asn Asp Leu Glu Ser Ile Phe Gly Gly Ser Ala Trp Gln Tyr
145 150 155 160

Asp Asp Lys Ser Asp Gln Tyr Tyr Leu His Phe Phe Ser Lys Lys Gln
165 170 175

Pro Asp Leu Asn Trp Glu Asn Ala Asn Leu Arg Gln Lys Ile Tyr Asp
180 185 190

Met Met Asn Phe Trp Ile Asp Lys Gly Ile Gly Gly Phe Arg Met Asp
195 200 205

Val Ile Asp Met Ile Gly Lys Ile Pro Ala Gln His Ile Val Ser Asn
210 215 220

Gly Pro Lys Leu His Ala Tyr Leu Lys Glu Met Asn Ala Ala Ser Phe
225 230 235 240

Gly Gln His Asp Leu Leu Thr Val Gly Glu Thr Trp Gly Ala Thr Pro
245 250 255

Glu Ile Ala Lys Gln Tyr Ser Asn Pro Val Asn His Glu Leu Ser Met
260 265 270

Ile Phe Gln Phe Glu His Ile Gly Leu Gln His Lys Pro Glu Ala Pro
275 280 285

Lys Trp Asp Tyr Val Lys Glu Leu Asn Val Pro Ala Leu Lys Thr Ile
290 295 300

Phe Asn Lys Trp Gln Thr Glu Leu Glu Leu Gly Gln Gly Trp Asn Ser
305 310 315 320

Leu Phe Trp Asn Asn His Asp Leu Pro Arg Val Leu Ser Ile Trp Gly
325 330 335

Asn Thr Gly Lys Tyr Arg Glu Lys Ser Ala Lys Ala Leu Ala Ile Leu
340 345 350

Leu His Leu Met Arg Gly Thr Pro Tyr Ile Tyr Gln Gly Glu Glu Ile
355 360 365

Gly Met Thr Asn Tyr Pro Phe Lys Asp Leu Asn Glu Leu Asp Asp Ile
370 375 380

Glu Ser Leu Asn Tyr Ala Lys Glu Ala Phe Thr Asn Gly Lys Ser Met
385 390 395 400

Glu Thr Ile Met Asp Ser Ile Arg Met Ile Gly Arg Asp Asn Ala Arg
405 410 415

Thr Pro Met Gln Trp Asp Ala Ser Gln Asn Ala Gly Phe Ser Thr Ala
420 425 430

Asp Lys Thr Trp Leu Pro Val Asn Pro Asn Tyr Lys Asp Ile Asn Val
435 440 445

Gln Ala Ala Leu Lys Asn Ser Asn Ser Ile Phe Tyr Thr Tyr Gln Gln
450 455 460

Leu Ile Gln Leu Arg Lys Glu Asn Asp Trp Leu Val Asp Ala Asp Phe
465 470 475 480

Glu Leu Leu Pro Thr Ala Asp Lys Val Phe Ala Tyr Leu Arg Lys Val
485 490 495

Arg Glu Glu Arg Tyr Leu Ile Val Val Asn Val Ser Asp Gln Glu Glu
500 505 510

Val Leu Glu Ile Asp Val Asp Lys Gln Glu Thr Leu Ile Ser Asn Thr
515 520 525

Asn Glu Ser Ala Ala Leu Ala Asn His Lys Leu Gln Pro Trp Asp Ala
530 535 540

Phe Cys Ile Lys Ile Asn
545 550

<210> 40
<211> 1860
<212> DNA
<213> 芽孢杆菌属和短双歧杆菌

<400> 40
atgagtttaa ccatcagtct gccgggtgtt caggctagcg cgatgatgac ctctttcaac 60

cgtgaacccc tgcccgacgc cgtccgcacg aatggcgcct ccccgaaccc gtggtggtcg 120

aacgccgtcg tctaccagat ttacccacgt tccttccagg acacgaacgg cgatggtttc 180

ggagatctta agggcattac ttcccgcctc gactatcttg ccgacctcgg cgtggatgtg 240

ctgtggctct ccccggtcta caggtccccg caagacgaca acggctacga catctccgac 300

tacgggaca tcgacccgct gttcggcacg ctcgacgaca tggacgagct gctgccgaa 360

gcgcacaagc gcggcctcaa gatcgtgatg gacctggtcg tcaaccacac ctccgacgag 420

cacgcgtggt tcgaggcgtc gaaggacaag gacgacccgc acgccgactg gtactggtgg 480

cgtcccgccc gccccggcca cgagcccggc acgcccggcg ccgagccgaa ccagtggggc 540

tcctacttcg gcggctccgc atgggaatat tgccccgagc gtggtgagta ctatctccac 600

cagttctcga agaagcagcc ggacctcaac tgggagaacc cggccgtacg ccgagccgtg 660

tacgacatga tgaactggtg gctcgatcgc ggcatcgacg gcttccgcat ggacgtcatc 720

accctgatct ccaagcgcac ggatgcaaac ggcaggctgc ccggcgagac cggttccgag 780

ctgcaggacc tgccggtggg ggaggagggc tactccaacc cgaacccgtt ctgcgccgac 840

ggtccgcgtc aggacgagtt cctcgccgag atgcgccgcg aggtgttcga cgggcgtgac 900

ggcttcctca ccgtcggcga ggccccggc atcaccgccg aacgcaacga gcacatcacc 960

gaccccgcca acggcgagct ggatatgctc ttcctgttcg aacacatggg cgtcgaccaa 1020

accccgaat cgaaatggga cgacaaacca tggacgccgg ccgacctcga aaccaagctc 1080

gccgaacaac aggacgccat cgccccgacac ggctgggcca gcctgttcct cgacaaccac 1140

gaccagccgc gtgtcgtctc ccgttggggc gacgacacca gcaagaccgg ccgcatccgc 1200

tccgccaagg cgctcgcgct gctgctgcac atgcaccgcg gcactccgta tgtctaccag 1260

ggcgaggagc tcggcatgac caatgcgcac ttcacctcgc tcgaccagta ccgcgacctc 1320

gaatccatca acgcctacca tcaacgcgtc gaggaaaccg ggatacggac atcggagacc 1380

atgatgcgat ccctcgcccg atacggcagg gacaacgcgc gcaccccgat gcaatgggac 1440

gactccacct acgccggctt caccatgccc gacgcccccgg tcgaaccctg gatcgccgtc 1500

aaccccgaacc acacggagat caacgccgcc gacgagatcg acgaccccga ctccgtgtac 1560

tcgttccaca aacggctcat cgccctgcgt cacaccgacc ccgtggtcgc cgccggcgac 1620

taccgacgcg tggaaaccgg aaacgaccgg atcatcgcct tcaccagaac cctcgacgag 1680

cgaaccatcc tcaccgtcat caacctctcg cccacacagg ccgcaccggc cggagaactg 1740

gaaacgatgc ccgacggcac gatcctcatc gccaacacgg acgatcccgt aggaaacctg 1800

aaaaccacgg gaacactcgg accatgggag gcgttcgcca tggaaaccga tccggaataa 1860

<210> 41
<211> 619
<212> PRT
<213> 芽孢杆菌属和短双歧杆菌

<400> 41

Met Ser Leu Thr Ile Ser Leu Pro Gly Val Gln Ala Ser Ala Met Met
1 5 10 15

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

Ala Ser Pro Asn Pro Trp Trp Ser Asn Ala Val Val Tyr Gln Ile Tyr
35 40 45

Pro Arg Ser Phe Gln Asp Thr Asn Gly Asp Gly Phe Gly Asp Leu Lys
50 55 60

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

Leu Trp Leu Ser Pro Val Tyr Arg Ser Pro Gln Asp Asp Asn Gly Tyr
85 90 95

Asp Ile Ser Asp Tyr Arg Asp Ile Asp Pro Leu Phe Gly Thr Leu Asp
100 105 110

Asp Met Asp Glu Leu Leu Ala Glu Ala His Lys Arg Gly Leu Lys Ile
115 120 125

Val Met Asp Leu Val Val Asn His Thr Ser Asp Glu His Ala Trp Phe
130 135 140

Glu Ala Ser Lys Asp Lys Asp Asp Pro His Ala Asp Trp Tyr Trp Trp
145 150 155 160

Arg Pro Ala Arg Pro Gly His Glu Pro Gly Thr Pro Gly Ala Glu Pro
165 170 175

Asn Gln Trp Gly Ser Tyr Phe Gly Gly Ser Ala Trp Glu Tyr Cys Pro
180 185 190

Glu Arg Gly Glu Tyr Tyr Leu His Gln Phe Ser Lys Lys Gln Pro Asp
195 200 205

Leu Asn Trp Glu Asn Pro Ala Val Arg Arg Ala Val Tyr Asp Met Met
210 215 220

Asn Trp Trp Leu Asp Arg Gly Ile Asp Gly Phe Arg Met Asp Val Ile
225 230 235 240

Thr Leu Ile Ser Lys Arg Thr Asp Ala Asn Gly Arg Leu Pro Gly Glu
245 250 255

Thr Gly Ser Glu Leu Gln Asp Leu Pro Val Gly Glu Glu Gly Tyr Ser
260 265 270

Asn Pro Asn Pro Phe Cys Ala Asp Gly Pro Arg Gln Asp Glu Phe Leu
275 280 285

Ala Glu Met Arg Arg Glu Val Phe Asp Gly Arg Asp Gly Phe Leu Thr
290 295 300

Val Gly Glu Ala Pro Gly Ile Thr Ala Glu Arg Asn Glu His Ile Thr
305 310 315 320

Asp Pro Ala Asn Gly Glu Leu Asp Met Leu Phe Leu Phe Glu His Met
325 330 335

Gly Val Asp Gln Thr Pro Glu Ser Lys Trp Asp Asp Lys Pro Trp Thr
340 345 350

Pro Ala Asp Leu Glu Thr Lys Leu Ala Glu Gln Gln Asp Ala Ile Ala
355 360 365

Arg His Gly Trp Ala Ser Leu Phe Leu Asp Asn His Asp Gln Pro Arg
370 375 380

Val Val Ser Arg Trp Gly Asp Asp Thr Ser Lys Thr Gly Arg Ile Arg
385 390 395 400

Ser Ala Lys Ala Leu Ala Leu Leu Leu His Met His Arg Gly Thr Pro
405 410 415

Tyr Val Tyr Gln Gly Glu Glu Leu Gly Met Thr Asn Ala His Phe Thr
420 425 430

Ser Leu Asp Gln Tyr Arg Asp Leu Glu Ser Ile Asn Ala Tyr His Gln
435 440 445

Arg Val Glu Glu Thr Gly Ile Arg Thr Ser Glu Thr Met Met Arg Ser
450 455 460

Leu Ala Arg Tyr Gly Arg Asp Asn Ala Arg Thr Pro Met Gln Trp Asp
465 470 475 480

Asp Ser Thr Tyr Ala Gly Phe Thr Met Pro Asp Ala Pro Val Glu Pro
485 490 495

Trp Ile Ala Val Asn Pro Asn His Thr Glu Ile Asn Ala Ala Asp Glu
500 505 510

Ile Asp Asp Pro Asp Ser Val Tyr Ser Phe His Lys Arg Leu Ile Ala
515 520 525

Leu Arg His Thr Asp Pro Val Val Ala Ala Gly Asp Tyr Arg Arg Val
530 535 540

Glu Thr Gly Asn Asp Arg Ile Ile Ala Phe Thr Arg Thr Leu Asp Glu
545 550 555 560

Arg Thr Ile Leu Thr Val Ile Asn Leu Ser Pro Thr Gln Ala Ala Pro
565 570 575

Ala Gly Glu Leu Glu Thr Met Pro Asp Gly Thr Ile Leu Ile Ala Asn
580 585 590

Thr Asp Asp Pro Val Gly Asn Leu Lys Thr Thr Gly Thr Leu Gly Pro
595 600 605

Trp Glu Ala Phe Ala Met Glu Thr Asp Pro Glu
610 615

<210> 42
<211> 1653
<212> DNA
<213> 芽孢杆菌属和变异链球菌

<400> 42
atgagtttaa ccatcagtct gccgggtgtt caggctagcg cgatgcaaaa acattggtgg 60

cacaaggcaa ctgtttatca aatttatcca aaatctttta tggatacaaa tggtgatgga 120

attggtgatc tcaaaggtat tacgagtaaa ttggattatt tgcaaaagtt aggggttatg 180

gctatttggc tatctccagt ttatgatagc cccatggatg acaatggcta tgacattgcg 240

aactatgaag caattgcgga tatttttggc aatatggctg atatggataa tttgctgacg 300

caggcaaaaa tgcgcgacat aaaaatcatt atggatctag tggttaatca tacctcagat 360

gaacatactt ggtttattga agcacgtgag catccagaca gttctgaacg cgattattat 420

atttggtgtg accagccaaa tgatttggaa tctatttttcg gtggttctgc ttggcagtat 480

gatgataagt ccgatcaata ttatttgcat ttttttagta agaagcagcc agatctaaac 540

tgggaaaacg caaacttacg tcagaagatt tatgatatga tgaatttctg gattgataaa 600

ggtattggcg gctttcggat ggacgtcatt gatatgattg ggaaaattcc tgctcagcat 660

attgtcagta acggaccaaa attgcatgct tatcttaagg agatgaatgc cgctagtttt 720

ggtcaacatg atctgctgac tgtgggggaa acttggggag caacgcctga gattgcgaag 780

caatattcaa atccagtcaa tcacgaactc tctatgattt ttcaatttga acatattggt 840

cttcagcata aaccagaagc tcctaaatgg gattatgtga aggaacttaa tgttcctgct 900

ttaaaaacaa tctttaataa atggcagact gagttggaat taggacaggg gtggaattcg 960

ttattctgga ataaccatga cctgcctcgt gttttatcaa tctggggaaa tacgggcaaa 1020

tatcgtgaga agtctgctaa agcactggct attcttcttc accttatgcg tgggacacct 1080

tatatttatc aaggtgaaga gattgggatg accaattatc cttttaaaga tttaaatgaa 1140

cttgatgata ttgaatcact taattatgct aaggaagctt ttacaaatgg taagtctatg 1200

gaaactatca tggacagtat tcgtatgatt ggccgtgata atgccagaac acctatgcaa 1260

tgggatgctt ctcaaaatgc cggattttca acagcggata aaacatggct gccagttaat 1320

ccaaactata aagacatcaa tgttcaagca gctctgaaaa attccaattc tatcttttac 1380

acctatcaac aactcattca gcttcgaaaa gaaaatgatt ggctagtaga tgccgatttt 1440

gaattgctcc ctacagcgga caaagtattt gcctatttac gaaaggtaag agaagaaagg 1500

tatcttatag tggtcaatgt ttcagatcag gaagaagttc tagagattga tgttgacaaa 1560

caagaaactc tcattagcaa tacaaatgaa agcgctgctc ttgccaatca caaactccag 1620

ccttgggatg ctttttgtat taagataaac taa 1653

<210> 43
<211> 550
<212> PRT
<213> 芽孢杆菌属和变异链球菌

<400> 43

Met Ser Leu Thr Ile Ser Leu Pro Gly Val Gln Ala Ser Ala Met Gln

1 5 10 15

Lys His Trp Trp His Lys Ala Thr Val Tyr Gln Ile Tyr Pro Lys Ser
20 25 30

Phe Met Asp Thr Asn Gly Asp Gly Ile Gly Asp Leu Lys Gly Ile Thr
35 40 45

Ser Lys Leu Asp Tyr Leu Gln Lys Leu Gly Val Met Ala Ile Trp Leu
50 55 60

Ser Pro Val Tyr Asp Ser Pro Met Asp Asp Asn Gly Tyr Asp Ile Ala
65 70 75 80

Asn Tyr Glu Ala Ile Ala Asp Ile Phe Gly Asn Met Ala Asp Met Asp
85 90 95

Asn Leu Leu Thr Gln Ala Lys Met Arg Asp Ile Lys Ile Ile Met Asp
100 105 110

Leu Val Val Asn His Thr Ser Asp Glu His Thr Trp Phe Ile Glu Ala
115 120 125

Arg Glu His Pro Asp Ser Ser Glu Arg Asp Tyr Tyr Ile Trp Cys Asp
130 135 140

Gln Pro Asn Asp Leu Glu Ser Ile Phe Gly Gly Ser Ala Trp Gln Tyr
145 150 155 160

Asp Asp Lys Ser Asp Gln Tyr Tyr Leu His Phe Phe Ser Lys Lys Gln
165 170 175

Pro Asp Leu Asn Trp Glu Asn Ala Asn Leu Arg Gln Lys Ile Tyr Asp
180 185 190

Met Met Asn Phe Trp Ile Asp Lys Gly Ile Gly Gly Phe Arg Met Asp

195 200 205

Val Ile Asp Met Ile Gly Lys Ile Pro Ala Gln His Ile Val Ser Asn
210 215 220

Gly Pro Lys Leu His Ala Tyr Leu Lys Glu Met Asn Ala Ala Ser Phe
225 230 235 240

Gly Gln His Asp Leu Leu Thr Val Gly Glu Thr Trp Gly Ala Thr Pro
245 250 255

Glu Ile Ala Lys Gln Tyr Ser Asn Pro Val Asn His Glu Leu Ser Met
260 265 270

Ile Phe Gln Phe Glu His Ile Gly Leu Gln His Lys Pro Glu Ala Pro
275 280 285

Lys Trp Asp Tyr Val Lys Glu Leu Asn Val Pro Ala Leu Lys Thr Ile
290 295 300

Phe Asn Lys Trp Gln Thr Glu Leu Glu Leu Gly Gln Gly Trp Asn Ser
305 310 315 320

Leu Phe Trp Asn Asn His Asp Leu Pro Arg Val Leu Ser Ile Trp Gly
325 330 335

Asn Thr Gly Lys Tyr Arg Glu Lys Ser Ala Lys Ala Leu Ala Ile Leu
340 345 350

Leu His Leu Met Arg Gly Thr Pro Tyr Ile Tyr Gln Gly Glu Glu Ile
355 360 365

Gly Met Thr Asn Tyr Pro Phe Lys Asp Leu Asn Glu Leu Asp Asp Ile
370 375 380

Glu Ser Leu Asn Tyr Ala Lys Glu Ala Phe Thr Asn Gly Lys Ser Met

385 390 395 400

Glu Thr Ile Met Asp Ser Ile Arg Met Ile Gly Arg Asp Asn Ala Arg
405 410 415

Thr Pro Met Gln Trp Asp Ala Ser Gln Asn Ala Gly Phe Ser Thr Ala
420 425 430

Asp Lys Thr Trp Leu Pro Val Asn Pro Asn Tyr Lys Asp Ile Asn Val
435 440 445

Gln Ala Ala Leu Lys Asn Ser Asn Ser Ile Phe Tyr Thr Tyr Gln Gln
450 455 460

Leu Ile Gln Leu Arg Lys Glu Asn Asp Trp Leu Val Asp Ala Asp Phe
465 470 475 480

Glu Leu Leu Pro Thr Ala Asp Lys Val Phe Ala Tyr Leu Arg Lys Val
485 490 495

Arg Glu Glu Arg Tyr Leu Ile Val Val Asn Val Ser Asp Gln Glu Glu
500 505 510

Val Leu Glu Ile Asp Val Asp Lys Gln Glu Thr Leu Ile Ser Asn Thr
515 520 525

Asn Glu Ser Ala Ala Leu Ala Asn His Lys Leu Gln Pro Trp Asp Ala
530 535 540

Phe Cys Ile Lys Ile Asn
545 550

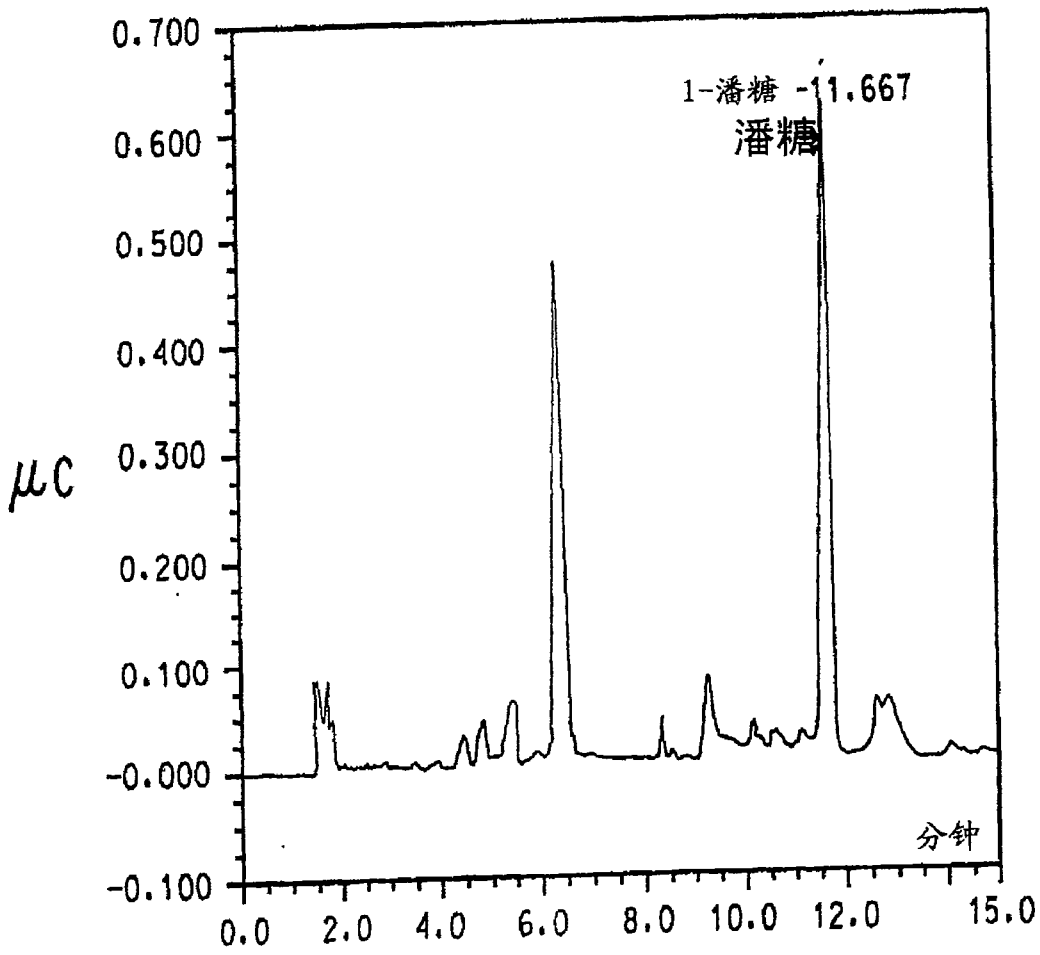

0.700
0.600
0.500
0.400
0.300
0.200
0.100
-0.000
-0.100
μC
1-潘糖 -11.667
潘糖
分钟
0.0
2.0
4.0
6.0
8.0
10.0
12.0
15.0

图 1A

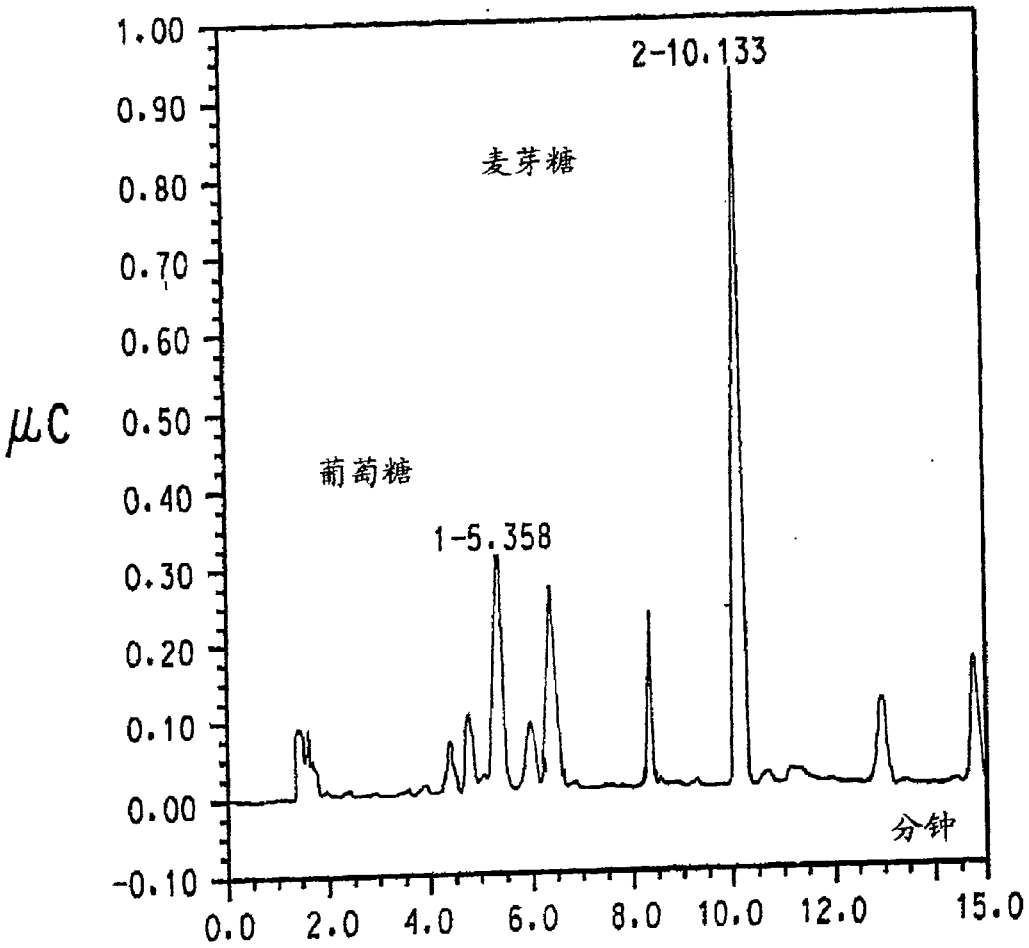

1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
-0.10
μC
2-10.133
麦芽糖
葡萄糖
1-5.358
分钟
0.0
2.0
4.0
6.0
8.0
10.0
12.0
15.0

图 1B

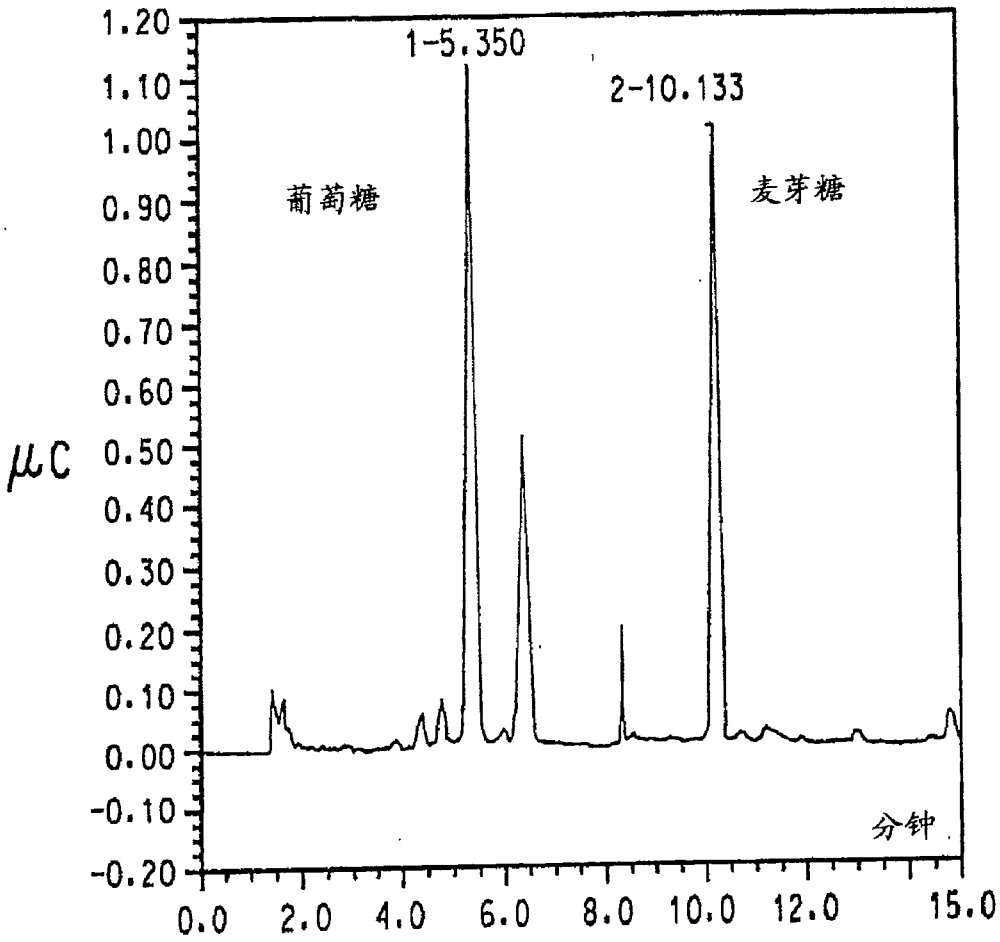

1-5.350
2-10.133
葡萄糖
麦芽糖
μC
分钟
1.20
1.10
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
-0.10
-0.20
0.0
2.0
4.0
6.0
8.0
10.0
12.0
15.0

图 1C

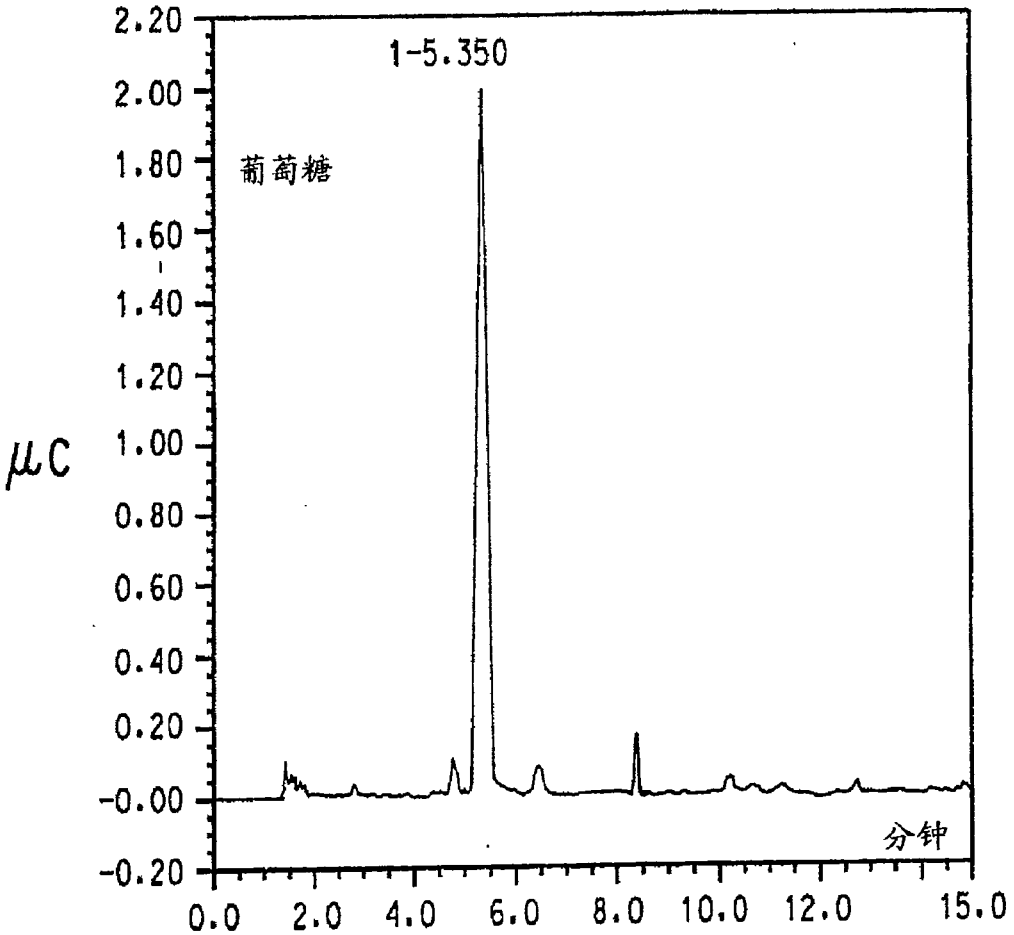

2.20
2.00
1.80
1.60
1.40
1.20
1.00
0.80
0.60
0.40
0.20
-0.00
-0.20
μC
1-5.350
葡萄糖
分钟
0.0
2.0
4.0
6.0
8.0
10.0
12.0
15.0

图 1D