



(12) 发明专利

(10) 授权公告号 CN 1468311 B

(45) 授权公告日 2012. 12. 12

(21) 申请号 01816886. 8

C13K 13/00 (2006. 01)

(22) 申请日 2001. 08. 29

A01H 1/00 (2006. 01)

(30) 优先权数据

PQ9768 2000. 08. 29 AU

(56) 对比文件

US 5985622 A, 1999. 11. 16, 说明书第 9 栏第 38 — 41 行, 序列表 SEQ ID NO: 10 和 9; 和说明书第 3 — 8 栏.

(85) PCT 申请进入国家阶段日

2003. 04. 04

作者: Yong K. Park, et al. 标

(86) PCT 申请的申请数据

PCT/AU2001/001084 2001. 08. 29

题: BIOTECHNOLOGY LETTERS 卷号: 14 期号: 7. 1992, 14(7), 547-551.

(87) PCT 申请的公布数据

W002/18603 EN 2002. 03. 07

审查员 曾繁辉

(73) 专利权人 昆士兰大学

地址 澳大利亚昆士兰州

(72) 发明人 R·G·博奇 武禄光

(74) 专利代理机构 中国国际贸易促进委员会专

利商标事务所 11038

代理人 李瑛

(51) Int. Cl.

C12N 15/61 (2006. 01)

C12N 15/31 (2006. 01)

C12N 9/90 (2006. 01)

权利要求书 2 页 说明书 53 页

序列表 34 页 附图 7 页

(54) 发明名称

6-果糖- α -葡糖苷合酶

(57) 摘要

6-果糖- α -葡糖苷合酶(蔗糖, EC5. 4. 99. 11) 将蔗糖转化为 6-果糖- α -葡糖苷(6-O- α -D-吡喃葡糖基-D-呋喃果糖)。该酶已被从大黄欧文氏菌和另一个细菌分离物 68J 中分离出来。还公开了用于将蔗糖转化为 6-果糖- α -葡糖苷的方法, 转化植物和细菌以使它们能够转化蔗糖, 检测源自样品的 6-果糖- α -葡糖苷合酶的方法和鉴定含有源自环境样品的 6-果糖- α -葡糖苷合酶的细菌的方法。

1. 一种分离的具有蔗糖异构酶活性的多肽,其含有:
 - (a) SEQ ID NO :8 或 10 所示的氨基酸序列 ;或
 - (b) 由 SEQ ID NO :7 或 9 编码的氨基酸序列。
2. 一种编码权利要求 1 所述多肽的分离的多核苷酸。
3. 一种分离的多核苷酸,其编码具有蔗糖异构酶活性的氨基酸序列,并含有:
 - (a) SEQ ID NO :7 或 9 所示的核苷酸序列 ;或
 - (b) 编码 SEQ ID NO :8 或 10 所示的氨基酸序列的核苷酸序列。
4. 一种含有与调节多核苷酸可操作性连接的权利要求 2-3 中任一项所述多核苷酸的表达载体。
5. 一种含有权利要求 4 所述表达载体的宿主细胞。
6. 权利要求 5 的宿主细胞,其是一种细菌或其它原核细胞。
7. 权利要求 5 的宿主细胞,其是一种植物细胞或其它真核细胞。
8. 一种含有权利要求 4 所述表达载体的植物细胞,其中所述植物是能够合成和 / 或积累蔗糖的物种。
9. 权利要求 8 的植物细胞,其中所述植物选自甘蔗或甜菜。
10. 权利要求 8 的植物细胞,其中所述植物是甘蔗。
11. 一种生产重组多肽的方法,所述重组多肽含有权利要求 1 所定义的多肽,所述方法包括:
 - 培养含有权利要求 4 所述的表达载体的宿主细胞以使所述重组多肽由所述多核苷酸表达 ;和
 - 分离所述重组多肽。
12. 一种生产多肽的生物活性片段的方法,所述多肽含有:
 - (a) SEQ ID NO :8 或 10 所示的氨基酸序列 ;或
 - (b) 由 SEQ ID NO :7 或 9 编码的氨基酸序列 ;所述方法包括:
 - 生产所述多肽的片段,和
 - 检测与所述片段相关的蔗糖异构酶活性,结果表明它是生物活性片段。
13. 权利要求 12 的生产多肽的生物活性片段的方法,所述方法包括:
 - 将一种多核苷酸引入到细胞中,由该多核苷酸可生产所述多肽的片段 ;
 - 从所述多核苷酸生产所述片段 ;和
 - 检测与所述片段的产生有关的蔗糖异构酶活性,结果表明它是生物活性片段。
14. 一种检测样品中的蔗糖异构酶的方法,包括:
 - 使所述样品与抗原结合分子接触,所述抗原结合分子与权利要求 1 的蔗糖异构酶具有免疫相互作用而与葡萄糖苷酶没有免疫相互作用,其中所述抗原结合分子选自免疫球蛋白、免疫球蛋白片段和源自非免疫球蛋白的呈现抗原结合活性的蛋白质构架 ;和
 - 检测所述被接触样品中存在的含有所述抗原结合分子和多肽的复合物,其中所述多肽如权利要求 1 中所定义。
15. 一种用于检测样品中的蔗糖异构酶的方法,包括 :

- 检测编码权利要求 1 中所定义的多肽的多核苷酸在所述样品中的表达。

16. 一种含有权利要求 4 所述的表达载体的转化植物细胞。

17. 权利要求 16 的植物细胞,其中所述植物是能够合成和 / 或积累蔗糖的物种。

18. 权利要求 16 的植物细胞,其中所述植物选自甘蔗或甜菜。

19. 权利要求 18 的植物细胞,其中所述植物是甘蔗。

20. 一种生产 6- 果糖 - α - 葡糖苷的方法,包括:

- 培养具有含有权利要求 4 所述表达载体的植物细胞的分化植物 ;和

- 从所述被培养的植物中收获 6- 果糖 - α - 葡糖苷。

21. 一种分离的生产 6- 果糖 - α - 葡糖苷的细菌,其生产权利要求 1 的多肽。

22. 权利要求 21 的细菌,其中所述多肽是由权利要求 3 的多核苷酸编码的。

23. 一种由蔗糖生产 6- 果糖 - α - 葡糖苷的方法,所述方法包括使蔗糖或含蔗糖的底物与权利要求 1 所述的多肽或与权利要求 5-7 任一项所述的宿主细胞或与权利要求 21-22 任一项的细菌在足以产生 6- 果糖 - α - 葡糖苷的条件下接触一段时间。

24. 一种由蔗糖生产 6- 果糖 - α - 葡糖苷的方法,所述方法包括使蔗糖或含蔗糖的培养基或底物与被编码权利要求 1 的多肽的 DNA 序列转化的生物的胞外产物或细胞提取物在蔗糖被所述多肽至少部分转化为 6- 果糖 - α - 葡糖苷的条件下进行接触。

6- 果糖 - α - 葡糖苷合酶

技术领域

[0001] 本发明总体涉及将蔗糖转化为 6- 果糖 - α - 葡糖苷的酶。更具体地,本发明涉及新型蔗糖异构酶,涉及编码这些蔗糖异构酶的多核苷酸,涉及分离这类多核苷酸的方法以及涉及表达这些多核苷酸的核酸构建体。本发明还涉及细胞,特别是被转化的细菌或植物细胞,以及涉及含有细胞的分化植物,所述细胞含有所述的核酸构建体。本发明进一步涉及本发明的多肽、多核苷酸、细胞和植物用于生产 6- 果糖 - α - 葡糖苷的用途。

背景技术

[0002] acarigenic 糖替代品,6- 果糖 - α - 葡糖苷(巴糖),是一种由 α -1,6- 糖苷键连接的葡萄糖和果糖组成的杂二糖。利用细菌蔗糖异构酶进行蔗糖酶促重排可以大规模地生产 6- 果糖 - α - 葡糖苷。

[0003] 最初,利用被固定化的自然产生蔗糖异构酶的细菌(例如红色精肱杆菌,大黄欧文氏菌和普城沙雷氏菌)细胞有利于 6- 果糖 - α - 葡糖苷的大规模生产。最近,利用重组技术已经获得了高产率的 6- 果糖 - α - 葡糖苷。在这方面,Mattes 等人(US5786140)公开了从红色精肱杆菌(CBS 54777),大黄欧文氏菌(NCPPB 1578),微生物 SZ 62(肠杆菌属某菌种)和微生物 MX-45(FERM 11808 或 FERM BP 3619)中分离到的编码部分或全长蔗糖异构酶的多核苷酸。

[0004] Mattes 等人还公开了保守氨基酸序列,按照该序列可以将简并寡核苷酸设计成适合于通过聚合酶链反应(PCR)克隆编码蔗糖异构酶的多核苷酸。

[0005] 发明内容

[0006] 在先于本发明的研究工作中,基于由 Mattes 等人公开的保守氨基酸序列,使用简并寡核苷酸通过 PCR 从大黄欧文氏菌(登记号 WAC2928)和 30 个独立的蔗糖异构酶阴性细菌分离物中扩增了编码蔗糖异构酶的多核苷酸。该 PCR 扩增反应生成了多种来源于大多数测试细菌的 DNA 产物。然而,却发现这些产物不能编码蔗糖异构酶。对 12 种单独的 PCR 产物,其中包括 6 种由大黄欧文氏菌扩增的产物,进行的核酸序列分析表明没有一种 DNA 产物显示出与蔗糖异构酶基因具有明显的序列同源性。相反,这些产物中的大多数表明与已知葡糖苷酶基因具有很高的序列同源性。因此,所得的结论是由 Mattes 等人公开的保守序列不是蔗糖异构酶所特有的,而是包括葡糖苷酶在内的其它种类的酶所共有的。

[0007] 尽管如此,本发明人仍然研究出一种新的功能性筛选测定法,该方法用于对编码用于生产 6- 果糖 - α - 葡糖苷的蔗糖异构酶的新型多核苷酸进行分离和定性。利用该测定法已经克隆出若干个这种新型多核苷酸,而且发现这些新型多核苷酸中的一些多核苷酸所编码的多肽的蔗糖异构酶活性高于由 Mattes 等人公开的那些多肽。推测出的多肽序列与已知蔗糖异构酶或葡糖苷酶的多肽序列的比较结果表明一些保守基元是蔗糖异构酶所特有的,因而能被专用于设计蔗糖异构酶特异性寡核苷酸。这类寡核苷酸的优势在于它们首次提供了利用核酸扩增技术容易地分离编码蔗糖异构酶的多核苷酸的方法。

[0008] 本发明人已经将上述发现付诸如下文所述的用于生产 6- 果糖 - α - 葡糖苷的新的

分离的分子、重组细胞和植物的实践中。

[0009] 因此,一方面,本发明提供了一种用于分离编码生产 6-果糖- α -葡糖苷的蔗糖异构酶的新型多核苷酸的方法,该方法包括:

[0010] (a) 从特定区域获得一种环境样品,在该特定区域中,能够将蔗糖转化为 6-果糖- α -葡糖苷的生物具有选择优势;

[0011] (b) 筛选能由蔗糖生产 6-果糖- α -葡糖苷的生物;和

[0012] (c) 从产 6-果糖- α -葡糖苷的生物中分离编码生产 6-果糖- α -葡糖苷的蔗糖异构酶的多核苷酸。

[0013] 优选地,所述方法进一步包括选样或富集蔗糖-和 6-果糖- α -葡糖苷-双重代谢生物的步骤,该生物既能利用蔗糖又能利用 6-果糖- α -葡糖苷作为生长的碳源。

[0014] 合适地,所述筛选利用了一种定量测定由生物产生的 6-果糖- α -葡糖苷产量的测定法。

[0015] 另一方面,本发明提供了一种分离的多肽,或其生物活性片段,或其变异体或其衍生物,所述多肽含有选自 SEQ ID NO:2、4、8 和 10 的氨基酸序列,前提条件是 SEQ ID NO:2 或 4 的生物活性片段含有一段被包含在 SEQ ID NO:26 或其变异体中的氨基酸的连续序列。

[0016] 合适地,所述变异体与 SEQ ID NO:2、4、8、10 和 26 所示氨基酸序列中的任意一个序列具有至少 75%,优选地是至少 80%,更优选地是至少 85%,更优选地是至少 90%,以及更加优选地是至少 95% 的序列相同性。

[0017] 合适地,所述生物活性片段是至少 6 个氨基酸的长度。

[0018] 优选地,所述变异体含有 SEQ ID NO:19、20、21、22、23 和 24 中的任意一个或多个序列所示的共有序列或其变异体。

[0019] 合适地,所述共有序列变异体与 SEQ ID NO:19、20、21、22、23 和 24 所示氨基酸序列中的任意一个序列具有至少 80%,更优选地是至少 90%,以及更加优选地是至少 95% 的序列相同性。

[0020] 另一方面,本发明提供了一种编码上述多肽、片段、变异体或衍生物的分离的多核苷酸。优选地,所述多核苷酸含有 SEQ ID NO:1、3、7 和 9 中的任意一个所示序列,或其生物活性片段,或其多核苷酸变异体,其前提条件是 SEQ ID NO:1 或 3 的生物活性片段含有一段被包含在 SEQ ID NO:25 中的核苷酸的连续序列或其多核苷酸变异体。

[0021] 在一个实施方案中,所述多核苷酸变异体与 SEQ ID NO:1、3、7、和 9 所示多核苷酸中的任意一个序列具有至少 60%,更优选地是至少 70%,以及更加优选地是至少 90% 的序列相同性。

[0022] 在另一个实施方案中,所述多核苷酸变异体在至少低严格条件,优选地在至少中等严格条件,以及更优选地在高严格条件下,能够与 SEQ ID NO:1、3、7、和 9 所示多核苷酸中的任意一个序列杂交。

[0023] 合适地,所述生物活性片段是至少 18 个核苷酸长度。

[0024] 优选地,所述多核苷酸变异体含有编码 SEQ ID NO:19、20、21、22、23 和 24 中的任意一个或多个所示的共有序列的核苷酸序列或其变异体。

[0025] 合适地,所述共有序列是由 SEQ ID NO:27、28、29、30、31、32、33、34、35 和 36 中的任意一个所示核苷酸序列或其核苷酸序列变异体编码的。

[0026] 在一个实施方案中,所述核苷酸序列变异体与 SEQ ID NO :27、28、29、30、31、32、33、34、35 和 36 所示序列中的任意一个具有至少 60%,优选地至少 70%,更优选地至少 80%,以及更加优选地至少 90%的序列相同性。

[0027] 在另一个实施方案中,所述核苷酸序列变异体在至少低严格条件,优选地在至少中等严格条件,以及更优选地在高严格条件下,能够与 SEQ ID NO :27、28、29、30、31、32、33、34、35 和 36 所示序列中的任意一个杂交。

[0028] 另一方面,本发明描述了一种含有上述多核苷酸的表达载体,其中所述多核苷酸可操作性地连接于调节多核苷酸上。

[0029] 另一方面,本发明提供了一种含有所述表达载体的宿主细胞。

[0030] 合适地,所述宿主细胞是一种细菌或其它原核生物,或者是一种植物细胞或其它真核生物。

[0031] 优选地,所述植物是甘蔗(甘蔗属某种)或另一种能够合成和/或积累蔗糖的物种(例如甜菜)。

[0032] 本发明还描述了一种生产上述重组多肽、片段、变异体或衍生物的方法,包括:

[0033] - 培养含有上述表达载体的宿主细胞以便由所述多核苷酸表达所述的重组多肽、片段、变异体或衍生物;和

[0034] - 分离所述的重组多肽、片段、变异体或衍生物。

[0035] 另一方面,本发明提供了一种生产上述多肽的生物活性片段的方法,包括:

[0036] - 检测与 SEQ ID NO :2、4、8 和 10 中的任意一个所示多肽的片段相关的蔗糖异构酶活性,活性检测结果表明所述片段就是所述的生物活性片段。

[0037] 另一方面,本发明提供了一种生产上述生物活性片段的方法,包括:

[0038] - 将一种多核苷酸引入细胞中,由该多核苷酸可以制备 SEQ ID NO :2、4、8 和 10 中的任意一个所示多肽的片段;和

[0039] - 检测蔗糖异构酶活性,检测结果表明所述片段就是所述的生物活性片段。

[0040] 另一方面,本发明提供了一种生产母体多肽的多肽变异体的方法,所述母体多肽含有 SEQ ID NO :2、4、8 和 10 中的任意一个所示的序列,或其生物活性片段所述方法包括:

[0041] - 通过至少一个氨基酸的取代、缺失或插入来生产一种改构多肽,该多肽的序列不同于母体多肽;和

[0042] - 检测与改构多肽相关的蔗糖异构酶活性,检测结果表明所述改构多肽就是所述的多肽变异体。

[0043] 另一方面,本发明设计了一种生产母体多肽的多肽变异体的方法,所述母体多肽含有 SEQ ID NO :2、4、8 和 10 中的任意一个所示的序列,或其生物活性片段所述方法包括:

[0044] - 制备一种由其能够产生上述改构多肽的多核苷酸;

[0045] - 将所述多核苷酸引入细胞;和

[0046] - 检测蔗糖异构酶活性,检测结果表明所述改构多肽就是所述的多肽变异体。

[0047] 根据本发明的另一方面,提供了一种由蔗糖生产 6-果糖- α -葡萄糖苷的方法,所述方法包括在足以产生 6-果糖- α -葡萄糖苷的条件下使蔗糖或一种含蔗糖底物与上述的多肽、片段、变异体或衍生物,或者与上述宿主细胞接触一段时间。

[0048] 另一方面,本发明在于提供了一种抗原结合分子,该抗原结合分子能够与本发明

所述的多肽、片段、变异体或衍生物具有免疫相互作用。

[0049] 优选地,所述抗原结合分子与 SEQ ID NO:19、20、21、22、23 和 24 中所示的任意一个氨基酸序列或其变异体具有免疫相互作用。

[0050] 本发明的另一方面提供了一种用于检测特异性多肽或多核苷酸的方法,该方法包括检测以下序列:

[0051] (a) SEQ ID NO:2、4、8 和 10,或其至少 6 个氨基酸长度的生物活性片段,或这些变异体,条件是 SEQ ID NO:2 或 4 的生物活性片段含有一段被包含在 SEQ ID NO:26 中的氨基酸的连续序列或其变异体;或

[0052] (b) 编码 (a) 的多核苷酸。

[0053] 在一个优选的实施方案中,(b) 的序列选自 SEQ ID NO:1、3、7 和 9,或其至少 18 个核苷酸长度的生物活性片段,或其多核苷酸变体,条件是 SEQ ID NO:1 或 3 的生物活性片段含有一段被包含在 SEQ ID NO:25 中的核苷酸的连续序列或其多核苷酸变异体。

[0054] 根据本发明的另一个方面,提供了一种检测样品中的蔗糖异构酶的方法,该方法包括:

[0055] - 使所述样品与上述抗原结合分子接触;和

[0056] - 检测所述被接触样品中存在的含有所述抗原结合分子和所述多肽、片段、变异体或衍生物的复合物。

[0057] 另一方面,本发明提供了一种检测上述多肽、片段、变异体或衍生物的方法,该方法包括:

[0058] - 检测编码所述多肽、片段、变异体或衍生物的多核苷酸在细胞中的表达。

[0059] 本发明的另一方面提供了一种含有核苷酸序列的探针,该核苷酸序列能够在至少低严格条件下与编码 SEQ ID NO:2、4、8 和 10 的至少一部分核苷酸序列杂交。

[0060] 在一个优选的实施方案中,所述探针含有能够在至少低严格条件下与 SEQ ID NO:1、3、7 和 9 的至少一部分杂交的核苷酸序列。

[0061] 根据本发明的另一个方面,提供了一种含有一种上述表达载体的转化植物细胞。

[0062] 在一个优选的实施方案中,所述植物细胞是甘蔗(甘蔗属某种)。

[0063] 另一方面,本发明提供了一种分化植物,该植物具有含有一种上述表达载体的植物细胞。

[0064] 另一方面,本发明提供了从上述分化植物中收获的 6-果糖- α -葡萄糖苷。

附图说明

[0065] 图 1. 在分离的细菌中蔗糖向 6-果糖- α -葡萄糖苷的转化。峰:1-蔗糖,2-6-果糖- α -葡萄糖苷,3-果糖,4-葡萄糖。“点状”电泳图谱是蔗糖和 6-果糖- α -葡萄糖苷标准品。

[0066] 图 2. 在表达蔗糖异构酶基因的大肠杆菌中蔗糖向 6-果糖- α -葡萄糖苷的转化,该基因在 SuperCosTM 载体中被克隆。峰:1-蔗糖,2-6-果糖- α -葡萄糖苷,3-果糖,4-葡萄糖。“点状”电泳图谱是蔗糖和 6-果糖- α -葡萄糖苷标准品。

[0067] 图 3. 从大黄欧文氏菌中克隆的蔗糖异构酶的核苷酸序列。

[0068] 图 4. 从 68J 中克隆的蔗糖异构酶的核苷酸序列。

[0069] 图 5. 从大黄欧文氏菌中克隆的蔗糖异构酶的推测的氨基酸序列。

[0070] 图 6. 从 68J 中克隆的蔗糖异构酶的推测的氨基酸序列。

[0071] 图 7. 通过表达克隆的蔗糖异构酶基因的大肠杆菌使蔗糖向 6-果糖- α -葡萄糖苷转化的效率。结果是得自 3 次重复试验的平均值 $\pm$ 标准误差。

[0072] 图 8. 在被稳定转化的甘蔗愈伤组织中蔗糖向 6-果糖- α -葡萄糖苷的转化, 该甘蔗愈伤组织表达被克隆的蔗糖异构酶基因。峰: 1-蔗糖, 2-6-果糖- α -葡萄糖苷, 3-果糖, 4-葡萄糖。“点状”电泳图谱是蔗糖和 6-果糖- α -葡萄糖苷标准品。迹线: a-pUbi Er+2.5mM 6-果糖- α -葡萄糖苷, b-pUbi Er, c-pUbi 14S, d-2.5mM 蔗糖和 6-果糖- α -葡萄糖苷标准品, e-pUbi 68J, f-pUbi 68J+2.5mM 6-果糖- α -葡萄糖苷。

[0073] 序列的简要描述: 一览表

[0074] 表 A

[0075]

序列 ID 序号	序列	长度
SEQ ID NO: 1	来自大黄欧文氏菌(登记号 WAC2928)的全长蔗糖异构酶编码序列	1899 个碱基
SEQ ID NO: 2	来自大黄欧文氏菌(登记号 WAC2928)的全长蔗糖异构酶多肽序列	632 个残基
SEQ ID NO: 3	来自大黄欧文氏菌(登记号 WAC2928)的编码成熟蔗糖异构酶的多核苷酸序列	1791 个碱基
SEQ ID NO: 4	来自大黄欧文氏菌(登记号 WAC2928)的成熟蔗糖异构酶多肽序列	596 个残基
SEQ ID NO: 5	来自大黄欧文氏菌(登记号 WAC2928)的与蔗糖异构酶相关的信号肽编码序列	108 个碱基
SEQ ID NO: 6	来自大黄欧文氏菌(登记号 WAC2928)的与蔗糖异构酶相关的信号肽	36 个残基
SEQ ID NO: 7	来自细菌分离物 68J 的全长蔗糖异构酶编码序列	1797 个碱基
SEQ ID NO: 8	来自细菌分离物 68J 的全长蔗糖异构酶多肽序列	598 个残基
SEQ ID NO: 9	来自细菌分离物 68J 的编码成熟蔗糖异构酶的多核苷酸序列	1698 个碱基
SEQ ID NO: 10	来自细菌分离物 68J 的成熟蔗糖异构酶多肽序列	565 个残基
SEQ ID NO: 11	来自细菌分离物 68J 的与蔗糖异构酶相关的信号肽编码序列	99 个碱基

[0076]

SEQ ID NO: 12	来自细菌分离物 68J 的与蔗糖异构酶相关的信号肽	33 个残基
SEQ ID NO: 13	用于扩增 68J 分离物的 5' 寡核苷酸引物	34 个碱基
SEQ ID NO: 14	用于扩增 68J 分离物的 3' 寡核苷酸引物	30 个碱基
SEQ ID NO: 15	用于扩增大黄欧文氏菌(登记号 WAC2928)的 5' 寡核苷酸引物	35 个碱基
SEQ ID NO: 16	用于扩增大黄欧文氏菌(登记号 WAC2928)的 3' 寡核苷酸引物	28 个碱基
SEQ ID NO: 17	用于扩增 14S 分离物的 5' 寡核苷酸引物	35 个碱基
SEQ ID NO: 18	用于扩增 14S 分离物的 3' 寡核苷酸引物	30 个碱基
SEQ ID NO: 19	蔗糖异构酶共有序列	7 个残基
SEQ ID NO: 20	蔗糖异构酶共有序列	10 个残基
SEQ ID NO: 21	蔗糖异构酶共有序列	6 个残基
SEQ ID NO: 22	蔗糖异构酶共有序列	6 个残基
SEQ ID NO: 23	蔗糖异构酶共有序列	13 个残基
SEQ ID NO: 24	蔗糖异构酶共有序列	16 个残基
SEQ ID NO: 25	编码源自大黄欧文氏菌(登记号 WAC2928)的蔗糖异构酶的新型羧基端部分的多核苷酸序列	594 个碱基
SEQ ID NO: 26	源自大黄欧文氏菌(登记号 WAC2928)的蔗糖异构酶的新型羧基端部分的多肽序列	197 个残基
SEQ ID NO: 27	编码 SEQ ID NO: 19 所示的共有序列的 SEQ ID NO: 1 的亚序列	21 个碱基
SEQ ID NO: 28	编码 SEQ ID NO: 20 所示的共有序列的 SEQ ID NO: 1 的亚序列	30 个碱基

[0077]

SEQ ID NO: 29	编码 SEQ ID NO: 21 所示的共有序列的 SEQ ID NO: 1 的亚序列	18 个碱基
SEQ ID NO: 30	编码 SEQ ID NO: 23 所示的共有序列的 SEQ ID NO: 1 的亚序列	39 个碱基
SEQ ID NO: 31	编码 SEQ ID NO: 24 所示的共有序列的 SEQ ID NO: 1 的亚序列	48 个碱基
SEQ ID NO: 32	编码 SEQ ID NO: 19 所示的共有序列的 SEQ ID NO: 7 的亚序列	21 个碱基
SEQ ID NO: 33	编码 SEQ ID NO: 20 所示的共有序列的 SEQ ID NO: 7 的亚序列	30 个碱基
SEQ ID NO: 34	编码 SEQ ID NO: 21 所示的共有序列的 SEQ ID NO: 7 的亚序列	18 个碱基
SEQ ID NO: 35	编码 SEQ ID NO: 23 所示的共有序列的 SEQ ID NO: 7 的亚序列	39 个碱基
SEQ ID NO: 36	编码 SEQ ID NO: 24 所示的共有序列的 SEQ ID NO: 7 的亚序列	48 个碱基
SEQ ID NO: 37	Geysen 文库肽	8 个残基
SEQ ID NO: 38	基于 Mattes 的正向引物	17 个碱基
SEQ ID NO: 39	基于 Mattes 的反向引物	19 个碱基

[0078] 本发明的详述

[0079] 1. 定义

[0080] 除非另有定义, 本文使用的所有科学技术术语具有本发明所属领域普通技术人员通常理解的含义。尽管与本文所述的方法和材料相似或相同的任何方法和材料都可被用于实施或检验本发明, 但还是对优选的方法和材料进行了描述。为了实现本发明的目的, 对下列术语进行了定义。

[0081] 冠词“a”和“an”在本文中被用来指一个或多于一个(即指至少一个)的冠词的语法对象。举例说明, “一个元件”意谓一个元件或多于一个的元件。

[0082] 术语“大约”在本文中被用来指相对于参比数量变化多达 30%，优选地多达 20% 以及更优选地多达 10% 的序列、参考水平、参考值、基准尺寸、长度、位置、大小或量。

[0083] “扩增产物”是指通过核酸扩增技术产生的核酸产物。

[0084] “抗原结合分子”是指与靶抗原具有结合亲合力的分子。应当理解为该术语泛指免疫球蛋白、免疫球蛋白片段和源自非免疫球蛋白的呈现抗原结合活性的蛋白质构架。

[0085] 本文中使用的术语“特异性结合”等是指结合本发明的多肽或多肽片段但不明显结合现有技术中的同源多肽的抗原结合分子。

[0086] “生物活性片段”是指全长母体多肽的片段，该片段保留了母体多肽的活性。因而生物活性片段将具有将蔗糖转化为 6-果糖- α -葡萄糖苷的蔗糖异构酶活性。本文中使用的术语“生物活性片段”包括缺失突变体和小肽，例如至少 8 个，优选地至少 10 个，更优选地至少 20 个以及更加优选地至少 30 个连续氨基酸的缺失突变体和小肽，所述缺失突变体和小肽仍具有上述活性。这类肽可以通过采用标准的核酸重组技术来获得或通过利用常规的液相或固相合成技术来进行合成。例如，可以参考由 Atherton 和 Shephard 所著的题为“肽合成”的文章中的第 9 章所记载的溶液合成方法或固相合成方法，该文章被收集在由 Nicholson 编辑的和由 Blackwell 科学出版社出版的题为“合成疫苗”的出版物中。或者，肽可通过采用蛋白酶如内切 Lys-C、内切 Arg-C、内切 Glu-C 和葡萄球菌 V8-蛋白酶消化本发明的多肽来生产。被消化的片段可以通过例如高效液相色谱 (HPLC) 技术来纯化。

[0087] 在整个说明书中，除非上下文另有说明，单词“包含”将被理解成是指包括所述步骤或元件或者包括一组步骤或元件但不排除任何其它步骤或元件或任何其它一组步骤或元件。

[0088] “相当于”或“相应于”是指 (a) 具有与参比多核苷酸序列的全长或部分序列基本上相同或互补的核苷酸序列的多核苷酸或 (b) 编码与肽或蛋白质中的氨基酸序列相同的氨基酸序列的多核苷酸。该词组在其范围内还包括具有与参比肽或蛋白质中的氨基酸序列基本上相同的氨基酸序列的肽或多肽。

[0089] “衍生物”是指通过修饰，例如通过与其它化学部分缀合或复合或通过本领域中熟悉的翻译后修饰技术由基础序列所衍生的多肽。术语“衍生物”在其范围内还包括对母体序列所做的包括插入或缺失在内的改变，这种改变可提供功能上等同的分子。因此，术语衍生物包括具有蔗糖异构酶活性的分子。

[0090] “同源性”是指相同氨基酸或构成下文表 B 中定义的保守取代的氨基酸的百分数。同源性可以利用序列对比程序如 GAP (Deveraux 等, 1984, 核酸研究 12, 387-395) 来确定。通过这种方式，长度与本文所列举的序列相似或基本上不同的序列可以通过将待测定的缺口插入到顺序中来进行比较，例如通过 GAP 采用的比较运算法。

[0091] “杂交”被本文用来表示互补核苷酸序列的配对从而产生 DNA-DNA 杂合子或 DNA-RNA 杂合子。互补碱基序列是那些按照碱基配对规则相关联的序列。在 DNA 中，A 与 T 配对以及 C 与 G 配对。在 RNA 中，U 与 A 配对，以及 C 与 G 配对。在这方面，如本文使用的术语“配对”和“错配”是指配对核苷酸在核酸互补链中的杂交潜力。配对的核苷酸能有效地进行杂交，如上述的经典 A-T 和 G-C 碱基对。错配是不能进行有效杂交的核苷酸的其它组合。

[0092] 本文提到的“免疫相互作用”包括分子间的以及尤其是其中所述分子中的一个

或类似于免疫系统的一种成分的分子间的任何相互作用、反应或其它形式的联系。

[0093] “免疫相互作用片段”是指 SEQ ID NO :2、4、8 和 10 中的任意一个所示多肽的片段，该片段能激发免疫应答，该免疫应答包括产生与所述多肽或其变异体或其衍生物特异结合的元件。如本文所使用的术语“免疫相互作用片段”包括缺失突变体和小肽，例如至少 6 个，优选地至少 8 个以及更优选地至少 20 个连续氨基酸的缺失突变体和小肽，所述缺失突变体和小肽含有抗原决定簇或表位。若干个这类片段可以被连接在一起。

[0094] “分离的”是指基本上或本质上不含有通常在其天然状态下与其共存的成分的物质。例如，本文中使用的“分离的多核苷酸”是指从在天然存在状态下与其侧接的序列中纯化出来的多核苷酸，例如从通常与所述片段相邻的序列中取出的 DNA 片段。

[0095] “标记基因”是指赋予了能够表达标记基因的细胞以不同的表型从而使这类转化细胞与不具有该标记的细胞区分开来的基因。选择性标记基因带来的特性是可以基于对选择剂（例如除草剂、抗生素、辐射、加热或其它能够损伤未转化细胞的处理方法）的抗性进行‘选择’。筛选性标记基因（或报道基因）带来的特性是可以通过观察或检测，即通过‘筛选’（例如 β -葡萄糖醛酸酶、荧光素酶或不存在于未转化细胞中的其它酶活性）来鉴别。

[0096] “获自于”是指样品如核酸提取物或多肽提取物分离自，或得自特定的来源。例如，所述提取物可以直接分离自任何代谢蔗糖的生物，优选地分离自代谢蔗糖的微生物，更优选地分离自土壤杆菌属、肠杆菌属、欧文氏菌属、克雷伯氏菌属、明串珠粒菌属、精氨酸菌属、假单胞菌属和沙雷氏菌属的微生物或分离自获自于特定区域的微生物，在该特定区域中能将蔗糖转化成 6-果糖- α -葡糖苷的生物具有例如如本文所述的选择优势。

[0097] 本文使用的术语“寡核苷酸”是指由多个借助磷酸二酯键连接在一起的核苷酸单元（脱氧核糖核苷酸或核糖核苷酸，或其相关结构变异体或其合成类似物）组成的聚合物。因此，尽管术语“寡核苷酸”通常是指其中核苷酸以及核苷酸间的键合是天然存在的核苷酸聚合物，但是应当理解该术语在其范围内还包括各种类似物，这些类似物包括，但不限于肽核酸（PNA）、氨基磷酸酯、硫代磷酸酯、甲基膦酸酯、2-O-甲基核糖核酸等。分子的准确大小可随着特定用途的变化而变化。寡核苷酸的长度通常很短，一般约 10 至 30 个核苷酸，尽管术语“多核苷酸”或“核酸”通常被用于大的寡核苷酸，但术语“寡核苷酸”可以指任意长度的分子。

[0098] “可操作性地连接”是指转录和翻译调节核酸以这样一种方式相对于编码多肽的多核苷酸被定位以致于所述多核苷酸被转录而且任选地翻译所述多肽。

[0099] 本文使用的“植物”和“分化植物”是指含有分化的植物细胞类型、组织和 / 或器官系统的完整的植物或植物部分。小植物和植物种子也被包括在上述术语的含义内。本发明所包括的植物是任何适用于转化技术的植物，所述植物包括被子植物、裸子植物、单子叶植物和双子叶植物。

[0100] 本文使用的术语“植物细胞”是指得自植物的原生质体或其它细胞、产生配子的细胞和再生完整植物的细胞。植物细胞包括植物中的细胞以及原生质体或培养物中的其它细胞。

[0101] “植物组织”是指分化组织和未分化组织，所述组织得自根、枝条花粉、种子、根瘤组织如冠瘿，和培养物中的植物细胞的各种形式的聚集物如胚芽和愈伤组织。

[0102] “组成型启动子”是指指导可操作性连接的可转录序列的在植物的多种或所有组

织中的表达的启动子。

[0103] “茎-特异性启动子”是指与在植物的叶、根或其它组织中的表达相比,优先指导可操作性连接的可转录序列的在植物的秆或茎组织中的表达的启动子。

[0104] 本文使用的术语“多核苷酸”或“核酸”是指 mRNA、RNA、cRNA、cDNA 或 DNA。该术语通常是指长度大于 30 个核苷酸的寡核苷酸。

[0105] 术语“多核苷酸变异体”和“变异体”是指其序列与参比多核苷酸序列基本相同的多核苷酸或在下文定义的严格条件下与参比序列杂交的多核苷酸。这些术语还包括其中一个或多个核苷酸被插入或缺失,或被不同核苷酸取代的多核苷酸。在这方面,在本技术领域应当充分认识到可以针对参比多核苷酸进行包括突变、插入、缺失和取代在内的某些改变,而且被改变的多核苷酸保留了参比多核苷酸的生物功能或活性。术语“多核苷酸变异体”和“变异体”还包括天然存在的等位变异体。

[0106] “多肽”、“肽”和“蛋白质”被本文交替用来指氨基酸残基的聚合物以及指该聚合物的变异体和合成类似物。因此,这些术语适用于其中一个或多个氨基酸残基是合成的、非天然存在的氨基酸的氨基酸聚合物,如相应天然存在的氨基酸的化学类似物,这些术语也指天然存在的氨基酸聚合物。

[0107] 术语“多肽变异体”是指其中一个或多个氨基酸已经被不同氨基酸取代的多肽。在本技术领域中很容易理解如下文所述某些氨基酸可被换成具有大致相似特性而不会改变所述多肽的活性特征的其它氨基酸(保守取代)。这些术语还包括其中一个或多个氨基酸被插入或缺失,或被不同氨基酸取代的多肽。因此,本文使用的多肽变异体包括具有蔗糖异构酶活性的多肽。

[0108] “引物”是指当与 DNA 链配对时,能够在有适当聚合剂存在的条件下启动引物延伸产物的合成的寡核苷酸。所述引物优选地是适于以最大效率进行扩增的单链但是也可以是双链的。引物必须足够长以便在有聚合剂存在的条件下引发延伸产物的合成。引物的长度取决于多种因素,该因素包括应用情况、要采用的温度、模板反应条件、其它试剂和引物的来源。例如取决于靶序列的复杂性,所述寡核苷酸引物通常含有 15 至 35 个或更多个核苷酸,但也可以含有较少的核苷酸。引物可以是大的多核苷酸,如从约 200 个核苷酸至几千个碱基或更多。可以选择与模板上的序列“基本互补”的引物,该引物被设计成能够与所述杂交并能作为合成的起始位点。“基本互补”是指引物充分互补从而能与靶核苷酸序列进行杂交。优选地,引物不含有与其被设计成能与之杂交的模板的错配,但这并不是必需的。例如,非互补的核苷酸可以被连接到引物的 5' 端,而引物序列的其余部分与模板互补。或者,非互补性的核苷酸或一段非互补核苷酸可以散布在引物中,条件是所述引物序列与所要进行杂交的模板序列具有充分互补性从而形成用于合成引物延伸产物的模板。

[0109] “探针”是指与特定序列或亚序列或另一分子的其它部分进行结合的分子。除了另有说明,术语“探针”通常是指,通过互补碱基配对,与常被称作“靶核酸”的另一核酸进行结合的多核苷酸探针。取决于杂交条件的严格性,探针可以结合缺乏与该探针完全序列互补性的靶核酸。探针可被直接或间接标记。

[0110] 本文使用的术语“重组多核苷酸”是指通过将核酸处理成一种通常不能在自然界发现的形式从而在体外形成的多核苷酸。例如,重组多核苷酸可以是表达载体的形式。一般,这类表达载体包括转录和翻译的调节核酸,该核酸与核苷酸序列进行可操作性的连接。

[0111] “重组多肽”是指利用重组技术制备的多肽,即通过重组多核苷酸的表达制备的多肽。

[0112] 本文使用的与植物材料有关的术语“再生”是指由一种植物细胞、一组植物细胞、植物部分(包括种子),或植物块(例如,由原生质体、愈伤组织或组织部分)培养出完整的、被分化的植物。

[0113] 本说明书中使用的“报道分子”是指一种通过其所具有的化学特性能够提供分析上可鉴别信号的分子,该信号使得含有抗原结合分子及其靶抗原的复合物得以检测。术语“报道分子”还延伸到细胞凝集的用途或凝集的抑制作用如乳胶珠粒上的红细胞等。

[0114] 用来描述两个或多个多核苷酸或多肽间的序列关系的术语包括“参比序列”、“比较窗口”、“序列相同性”、“序列相同性的百分率”和“基本相同”。“参比序列”是至少 12 个但常常是 15 至 18 个以及常常是至少 25 个单体单位长度,所述单体单位包括核苷酸和氨基酸残基。由于两个多核苷酸可以各自含有 (1) 在两个多核苷酸之间相似的一段序列(即,只是完整多核苷酸序列的一部分),和 (2) 在两个多核苷酸之间不同的一段序列,两个(或多个)多核苷酸之间的序列比较通常按照以下方式进行:利用“比较窗口”比较两个多核苷酸的序列从而鉴别和比较局部区域的序列相似性。“比较窗口”是指至少 6 个连续位点,通常约 50 至约 100 个,更经常约 100 至约 150 个连续位点的概念性片段,在该片段中将两个序列进行最佳排列后,将序列与连续位点数目相同的参比序列进行比较。与用于两个序列的最佳排列的参比序列(其不含插入或缺失)相比,比较窗口可含有约 20%或低于 20%的插入或缺失(即,缺口)。用于排列比较窗口的序列的最佳排列可以通过以下方式进行:通过计算机进行的运算(Wisconsin 遗传学软件包 7.0 版,遗传学计算机组,575 ScienceDrive Madison, WI, 美国,中的 GAP, BESTFIT, 和 TFASTA)或通过利用所选多种方法中的任意一种进行的检查和最佳排列(即,通过比较窗口产生最高百分率的同源性)。也可以参考例如在 Altschul 等,1997,核酸研究(Nucl. Acids Res). 25 ;3389 中公开的 BLAST 家族的程序。序列分析的详细讨论可以参见 Ausubel 等人,“精编分子生物学实验室指南(Current Protocols in Molecular Biology)”, John Wiley & Sons Inc, 1994-1998, 第 15 章的 19.3 单元。

[0115] 本文使用的术语“序列相同性”是指通过比较窗口进行逐个核苷酸或逐个氨基酸比较后得出的序列相同的程度。因此,“序列相同性的百分率”的计算方法是利用比较窗口比较两个最佳排列的序列,确定其中相同核酸碱基(例如, A, T, C, G, I)或相同氨基酸残基(例如, Ala, Pro, Ser, Thr, Gly, Val, Leu, Ile, Phe, Tyr, Trp, Lys, Arg, His, Asp, Glu, Asn, Gln, Cys 和 Met)出现在两个序列中的位点数目从而得到配对位点的数目,将配对位点的数目除以比较窗口中位点的总数目(即窗口大小),以及将结果乘以 100 从而得到序列相同性的百分率。为了本发明的目的,“序列相同性”可被理解为“配对百分率”,该配对百分率通过 DNASIS 计算机程序(windows 2.5 版;可获自于 Hitachi 软件工程有限公司,南旧金山,加利福尼亚,美国)计算得到,该计算机程序利用了随软件所附的参考手册中使用的标准缺省值。

[0116] 本文使用的“严格性”是指在杂交和洗涤过程中,温度和离子强度条件,以及有或没有某些有机溶剂的存在。严格性愈高,固定化靶核苷酸序列与被标记探针多核苷酸序列之间的互补程度愈高,所述被标记探针多核苷酸序列在洗涤后仍保持与靶核苷酸序列杂

交。

[0117] “严格条件”是指在其下只有具有高频率互补性碱基的核苷酸序列才能杂交的温度和离子强度条件。所需严格性是由核苷酸序列决定的,取决于杂交和随后洗涤过程中存在的多种成分,以及这些过程所允许的时间。通常,为了使杂交率最大化,选择非严格杂交条件;比热熔点 (T_m) 低约 20 至 25°C。 T_m 是在具有确定的离子强度和 pH 的溶液中有 50% 的特异性靶序列与完全互补性探针杂交的温度。通常,为了获得杂交序列的至少约 85% 的核苷酸互补性,选择比 T_m 低约 5 至 15°C 的高严格洗涤条件。为了获得杂交序列的至少约 70% 的核苷酸互补性,选择比 T_m 低约 15 至 30°C 的中等严格洗涤条件。高度随意的(低严格性)洗涤条件可低至比 T_m 低 50°C,使得杂交序列间发生高水平的错配。本技术领域的技术人员能够认识到在杂交和洗涤阶段中的其它物理和化学参数也可以被改变从而影响可检测杂交信号的结果,该信号来自靶序列和探针序列间同源性的特定水平。严格条件的其它实例在 3.3 节中进行了描述。

[0118] 术语“转化”是指通过引入外源性或内源性核酸,生物表型发生的改变,所述生物例如是细菌或植物。

[0119] “转基因子”是指转化过程中的中间产物。

[0120] “载体”是指核酸分子,优选地是源自例如质粒、噬菌体或植物病毒的 DNA 分子,在所述载体中可以插入或克隆核酸序列。载体优选地含有一个或多个独特的限制位点并且可能能够在包括靶细胞或靶组织或始祖细胞或其组织在内的确定的宿主细胞内自主复制,或者能与确定的宿主细胞的基因组整合在一起以使克隆序列可复制。因此,载体可以是自主复制的载体,即以染色体外实体存在的、其复制不依赖于染色体的复制的载体,所述染色体外实体例如是线性的或闭环的质粒、染色体外因子、小染色体或合成染色体。载体可含有保证自我复制的任何元件。或者,载体可以是一种当被引入细胞中时,其被整合到受体细胞的基因组中而且与其被整合到其中的染色体一起进行复制。载体系统可以包括单个载体或质粒,两个或多个载体或质粒,其共同含有要被引入宿主基因组中的总 DNA,或转座子。载体的选择通常取决于载体与载体要引入其中的细胞间的相容性。载体还可以包括一个选择标记如可用于选择合适转化体的抗生素抗性基因。这类抗性基因的实例对于本技术领域的技术人员来说是熟知的。

[0121] 2. 分离的多肽、生物活性片段、多肽变异体和衍生物

[0122] 2.1 本发明的多肽

[0123] 本发明部分基于对源自大黄欧文氏菌(登记号 WAC2928)的蔗糖异构酶的全长序列和源自被称作 68J 的细菌分离物的新型蔗糖异构酶的全长序列的确定。

[0124] 大黄欧文氏菌蔗糖异构酶的全长氨基酸序列长度为 632 个残基而且与由 Mattes 等人(见上文)公开的序列相比包括羧基端序列的 197 个附加残基(SEQ ID NO:26 中所示)。大黄欧文氏菌多肽包括一个前导肽或信号肽,如 SEQ ID NO:6 中所示,所述前导肽或信号肽的长度为从 SEQ ID NO:2 的第 1 个残基至约第 36 个残基。所述信号肽只是成熟多肽在特定细胞区室(例如,在外膜内、在内膜内或在外膜和内膜之间的壁膜间隙内)进行正定位所必需的。SEQ ID NO:4 所示的成熟多肽的长度为从约第 37 个残基至第 632 个残基。因此,在一个实施方案中,本发明提供了一种 SEQ ID NO:2 所示的分离的前体多肽,该多肽含有一个 SEQ ID NO:6 所示的前导肽,该前导肽在读框内与 SEQ ID NO:4 所示多肽融

合在一起。在另一个实施方案中,本发明提供了一种含有 SEQ ID NO:4 所示序列的分离的成熟多肽。

[0125] 68J 蔗糖异构酶的全长氨基酸序列的长度为 SEQ ID NO:8 中所示的 598 个残基,并且含有一个 SEQ ID NO:12 所示的信号肽,该信号肽的长度为 SEQ ID NO:8 的第 1 个至约第 33 个残基。SEQ IDNO:10 所示的成熟多肽的长度为 SEQ ID NO:8 的约第 34 个残基至第 598 个残基。因此,在一个实施方案中,本发明提供了一种 SEQ ID NO:8 的分离前体多肽,该多肽含有一个 SEQ ID NO:12 的前导肽,该前导肽在读框内与 SEQ ID NO:10 的多肽融合在一起。在另一个实施方案中,本发明提供了一种含有 SEQ ID NO:10 所示序列的分离成熟多肽。

[0126] 2.2 生物活性片段

[0127] 生物活性片段可以按照本领域已知的任何合适的方法来生产。例如,合适的方法可包括首先生成所述多肽的片段和然后测试该片段的适当生物活性。在一个实施方案中,可以测试所述片段的蔗糖异构酶活性。本发明提供了检测或优选地测定蔗糖异构酶活性的任何测定方法。优选地,蔗糖异构酶活性通过本文记载的苯胺/二苯胺测定法和毛细管电泳来测定。

[0128] 在另一个实施方案中,所述片段的生物活性通过以下步骤来测定:将一种能够翻译所述多肽片段的多核苷酸引入细胞,检测蔗糖异构酶活性,该检测表明所述片段是所述生物活性片段。

[0129] 本发明还设计了上述多肽的至少 6 个以及优选地至少 8 个氨基酸长度的生物片段,该生物片段能够激发动物体内的免疫应答以便产生与本发明的蔗糖异构酶具有免疫相互作用的抗体。例如,范例性的能够激发免疫应答的 8 个氨基酸长度的多肽片段包括但不限于 SEQ IDNO:2 的残基 1-8,9-16,17-24,25-32,33-40,41-48,49-56,57-64,65-72,73-80,81-88,89-96,97-104,105-112,113-120,121-128,129-136,137-144,145-152,153-160,161-168,169-176,177-184,185-192,193-200,201-208,209-216,217-224,225-232,223-240,241-248,249-256,257-264,265-272,273-280,281-288,289-296,297-304,305-312,313-320,321-328,329-336,337-344,345-352,353-360,361-368,369-376,377-384,385-392,393-400,401-408,409-416,417-424,425-432,423-440,441-448,449-456,457-464,465-472,473-480,481-488,489-496,497-504,505-512,513-520,521-528,529-536,537-544,545-552,553-560,561-568,569-576,577-584,585-592 和 589-596,或 SEQ ID NO:4 的残基 1-8,9-16,17-24,25-32,33-40,41-48,49-56,57-64,65-72,73-80,81-88,89-96,97-104,105-112,113-120,121-128,129-136,137-144,145-152,153-160,161-168,169-176,177-184,185-192,193-200,201-208,209-216,217-224,225-232,223-240,241-248,249-256,257-264,265-272,273-280,281-288,289-296,297-304,305-312,313-320,321-328,329-336,337-344,345-352,353-360,361-368,369-376,377-384,385-392,393-400,401-408,409-416,417-424,425-432,423-440,441-448,449-456,457-464,465-472,473-480,481-488,489-496,497-504,505-512,513-520,521-528,529-536,537-544,545-552,553-560 和 559-566。在一个优选的这类实施方案中,生物活性片段含有至少一种选自 SEQ ID NO:19,20,21,22,23 或 24 的蔗糖异构酶共有序列。

[0130] 2.3 多肽变异体

[0131] 本发明还设计了本发明多肽的多肽变异体,其中所述变异体具有蔗糖异构酶活性。制备多肽变异体的合适方法包括,例如,通过取代、缺失和/或插入至少一个氨基酸来制备其序列不同于母体多肽的改构多肽,其中所述母体多肽含有 SEQ ID NO :2,4,8 和 10 任一个所示的序列。然后检测改构多肽的蔗糖异构酶活性,如果具有所述活性,则表明所述改构多肽就是所述的变异体。

[0132] 在另一个实施方案中,多肽变异体通过以下步骤来制备:将一种能够翻译改构多肽的多核苷酸引入细胞,检测与细胞有关的蔗糖异构酶活性,该检测表明所述改构多肽是所述多肽变异体。

[0133] 一般来说,变异体与例如 SEQ ID NO :2,4,8 和 10 所示的多肽或其生物活性片段具有至少 60%,更合适地至少 70%,优选地至少 80%,和更优选地至少 90%的同源性。优选地是所述变异体显示出与例如 SEQ ID NO :2,4,8 和 10 所示的多肽或其生物活性片段至少 60%,更合适地至少 70%,优选地至少 75%,更优选地至少 80%,更优选地至少 85%,更优选地至少 90%以及更加优选地至少 95%的序列相同性。在这方面,比较窗口优选地跨越大约所述多肽或所述生物活性片段的全长。

[0134] 合适的变异体可获自于任何合适的蔗糖代谢生物。优选地,所述变异体可获自于蔗糖代谢细菌如下文 3.3 节所记载的细菌。

[0135] 2.4 生产多肽变异体的方法

[0136] 2.4.1 诱变

[0137] 本发明的多肽变异体既可以合理地,也可以借助已经建立的诱变方法(例如参见 Watson, J. D. 等, " 基因的分子生物学 (MOLECULAR BIOLOGY OF THE GENE) ", 第四版, Benjamin/Cummings, Menlo Park, 加利福尼亚, 1987) 来确定。显然,随机诱变方法不需要关于待突变的基因序列的推理的信息。该方法的优势在于根据其功能来评价特定突变体的需要性,因此不需要了解所获得的突变蛋白质是怎样或为什么采取了特定的构象。实际上,靶基因序列的随机突变一直是一种被用来获取具有理想特性的突变蛋白质的方法 (Leatherbarrow, R. 1986, 蛋白质工程杂志 (J. Prot. Eng) 1 :7-16 ; Knowles, J. R. , 1987, 科学 (Science) 236 :1252-1258 ; Shaw, W. V. , 1987, 生物化学杂志 (Biochem. J) 246 :1-17 ; Gerit, J. A. 1987, 化学评论 (Chem. Rev) 87 :1079-1105)。或者,如果需要特定的序列改变,可以采用定点诱变的方法。因此,这类方法可以被用来选择性地仅仅改变那些被认为重要的蛋白质的氨基酸 (Craik, C. S. , 1985, 科学 (Science) 228 :291-297 ; Cronin, 等, 1988, 生物化学 (Biochem) 27 :4572-4579 ; Wilks, 等, 1988, 科学 (Science) 242 :1541-1544)。

[0138] 通过推理的或已经建立的诱变方法或通过下文所述的组合化学得到的变异肽或多肽可包括保守的氨基酸取代。本发明的多肽或多肽片段中的范例性保守取代可以按照下表进行:

[0139] 表 B

[0140]

原始残基范例性取代	
Ala	Ser
Arg	Lys
Asn	Gln, His

Asp	Glu
Cys	Ser
Gln	Asn
Glu	Asp
Gly	Pro
His	Asn, Gln
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile,
Phe	Met, Leu, Tyr
Ser	Thr
Thr	Ser
Trp	Tyr
Tyr	Trp, Phe
Val	Ile, Leu

[0141] 通过选择保守性低于表 B 所列取代的取代可以使功能发生巨大的变化。其它取代将是非保守性取代而且其中只有相对较少的取代能被允许。一般,可能引起多肽的特性发生最大变化的取代是那些其中发生了下列取代的取代:(a) 亲水性残基(例如 Ser 或 Thr)被疏水性残基(例如,Ala,Leu,Ile,Phe 或 Val)所取代;(b) 半胱氨酸或脯氨酸被其它任意残基所取代;(c) 具有带正电的侧链的残基(例如,Arg, His 或 Lys)被带负电的残基(例如, Glu 或 Asp)所取代或(d) 具有庞大侧链的残基(例如, Phe 或 Trp)被具有较小侧链的残基(例如,Ala, Ser)或不具有侧链的残基(例如, Gly)所取代。

[0142] 可以通过常规技术来确定怎样构建合适的变异体。例如,编码 SEQ ID NO :2,4,8 和 10 的多肽的核酸可以利用随机诱变例如利用转座子诱变,或者利用例如下文 3.3 节所记载的定点诱变来进行突变。

[0143] 2.4.2 利用组合化学制备的肽文库

[0144] 许多灵活适用的组合技术可被用来合成浩大多样性的分子文库。在本案中,多肽变异体,或者优选地是本发明的多肽片段可以利用这类技术来进行合成。然后可以利用 2.3 节所记载的方法来筛选变异体。

[0145] 优选地,制备可溶性的合成肽组合文库(SPCL)从而提供了在溶液中处理游离肽的优势,因而允许对肽浓度进行调节以便适应特定的测定系统。SPCL 被适当地制备成六聚体。在这方面,已知大多数的结合位点涉及到四至六个残基。半胱氨酸被优选地从混合位点中排除以便避免形成二硫键和较难确定的聚合物。制备 SPCL 的范例性方法被 Houghten 等人(1991,自然(Nature)354:84-86;1992,生物技术(BioTechniques)13:412-421), Appel 等人(1992,免疫方法(Immunomethods)1:1723),和 Pinilla 等人(1992,生物技术(BioTechniques)13:901-905;1993,基因(Gene)128:71-76)所公开。

[0146] 组合合成肽文库的制备可以采用叔丁氧羰基(t-Boc)或 9-芴基甲氧羰基(Fmoc)化学法(参见 Coligan 等人,上文的 9.1 章;Stewart 和 Young,1984,固相肽合成(Solid Phase Peptide Synthesis),第二版,Pierce Chemical Co.,Rockford,IL;和 Atherton 和 Sheppard,1989,固相肽合成(Solid Phase Peptide Synthesis):实用方法(A Practical Approach)IRL 出版社,牛津大学),优选地但不是唯一地,利用两种不同方法中的一种方法。这些方法中的第一种方法,被适当地称作“断裂-加工-重组”或“裂解合成”方法,

首次由 Furka 等人 (1988, 第 14 届国际生物化学大会 (14th Int. Congr. Biochem), 布拉格, Czechoslovakia 5:47; 1991, 肽、蛋白质研究国际杂志 (Int. J. Pept. Protein Res) 37: 487-493) 和 Lam 等人 (1991, 自然 (Nature) 354:82-84) 进行了记载, 以及后来由 Eichler 等人 (1995, 医药研究综述 (Medicinal Research Reviews) 15(6):481-496) 和 Balkenhohl 等人 (1996, Angew. Chem. Int. Ed. Engl. 35:2288-2337) 进行了评述。概括地说, 所述裂解合成方法包括将多个固体载体如聚合物珠粒分成 n 等分, 该等分代表用于每步合成的有效氨基酸数目 (例如, 20 个 L-氨基酸), 分别将单个氨基酸偶联到相应等分的每个聚合物珠粒上, 然后充分混合所有等分的聚合物珠粒。将此过程总共重复 x 个循环以便制备出高达 N^x 种不同化合物的随机集合物。对如此制备的肽文库可进行蔗糖异构酶活性的筛选。通过检测, 选择一些阳性珠粒用于测序从而确定活性肽。随后可以将这种肽从珠粒上裂解下来, 并按照上述方法进行测定。

[0147] 第二种方法, 化学比率方法, 通过利用根据经验确定的特定比率的氨基酸使得在每个偶联步骤等摩尔掺入每种氨基酸来制备混合型的肽树脂。每个树脂珠粒含有肽的混合物。接近等摩尔的表示法可以通过氨基酸分析 (Dooley 和 Houghten, 1993, 美国国家科学院院报 (Proc. Natl. Acad. Sci. U. S. A). 90:10811-10815; Eichler 和 Houghten, 1993, 生物化学 (Biochemistry) 32:11035-11041) 来得到证实。优选地, 例如按照 Geysen 等人公开的方法 (1986, 分子免疫学 (Mol. Immunol). 23:709-715) 在作为固体载体的聚乙烯小棒状物或针形物上制备所述合成肽文库。一个范例性的这类肽文库可由八肽组成, 该八肽中的第三个和第四个位点代表选自天然和非天然氨基酸的确定氨基酸, 剩下的六个位点代表氨基酸的随机混合物。该肽文库可以由式 $\text{Ac-XXO}_1\text{O}_2\text{XXXX-Ss}$ [SEQ ID NO:37] 来表示, 其中 Ss 代表固体载体。肽混合物仍然保留在针形物上用于测定目的。例如, 可以首次筛选肽文库将蔗糖转化为 6-果糖- α -葡萄糖苷的能力。然后选择最大活性肽用于另一循环的测试, 该测试包括将附加的残基连接到起始肽上 (或通过内部修饰原起始肽的组分), 然后检测这组候选肽的蔗糖异构酶活性。重复该过程直到确定出具有理想蔗糖异构酶活性的肽。一旦得到确定, 与固相载体连接的肽的特性可通过肽测序来测定。

[0148] 2.4.3 丙氨酸扫描诱变

[0149] 在一个实施方案中, 本发明通过对本发明的多肽或多肽片段进行系统分析来确定所述多肽或片段中的残基, 该残基参与蔗糖向 6-果糖- α -葡萄糖苷转化的催化作用。利用 DNA 重组技术可以方便地进行这类分析。一般, 对编码多肽或其片段的 DNA 序列进行克隆和操作从而可在适宜的宿主中对其进行表达。编码所述多肽或其片段的 DNA 可获自基因组文库、获自得自于表达所述多肽或其片段的细胞中的 mRNA 的 cDNA、或者通过合成构建所述的 DNA 序列 (Sambrook 等, 上文; Ausubel 等, 上文) 获得。

[0150] 然后如本文所述将编码所述多肽或其片段的野生型 DNA 插入到合适质粒或载体中。尤其是, 原核生物被优选用于克隆和表达 DNA 序列从而生产出所述多肽或其片段的变异体。例如, 可以使用大肠杆菌 K12 菌株 294 (ATCC No. 31446), 也可以使用大肠杆菌 B, 大肠杆菌 X1776 (ATCC No. 31537), 和大肠杆菌 c600 和 c600hf1, 和大肠杆菌 W3110 (F^- , γ^- , 原养型的, ATCC No. 27325), 杆菌如枯草杆菌, 和其它肠杆菌科如鼠伤寒杆菌或退色沙雷氏菌, 和多种假单胞菌属的菌种。优选的原核生物是大肠杆菌 W3110 (ATCC 27325)。

[0151] 一旦克隆了该多肽或其片段, 就可以对被克隆的 DNA 进行例如由 Carter 等人

(1986, 核酸研究 (Nucl. Acids. Res.), 13 :4331) 或 Zoller 等人 (1987, 核酸研究 (Nucl. Acids. Res.), 10 :6487) 所记载的定点诱变、例如由 Wells 等人 (1985, 基因 (Gene), 34 : 315) 所记载的盒式诱变、例如由 Wells 等人 (1986, Philos. Trans. R. Soc. London SerA, 317 :415) 所记载的限制选择诱变或其它已知技术的操作从而生产出变异体 DNA, 该变异体 DNA 编码通过将被取代残基限定的氨基酸序列的改变。当被可操作性地连接到合适表达载体中的调节多核苷酸上时, 就获得了变异体多肽。在某些情况下, 通过利用一个与编码所述变异体的 DNA 序列可操作性连接的合适信号序列来表达和分泌这类分子可以方便地回收所述变异体。这类方法对于所述技术领域技术人员是熟知的。当然, 也可以采用其它方法如体外化学合成所需的多肽变异体 (Barany 等, 肽 (The Peptides), 由 E. Gross 和 J. Meienhofer 编辑 (纽约科学出版社 (Academic Press :N. Y). 1979), Vol. 2, 3-254 页) 来生产这类多肽或其片段。

[0152] 一旦不同的变异体被生产出来, 就使它们与蔗糖或含蔗糖的底物进行接触并且如果发生转化, 则针对每个变异体测定蔗糖向 6-果糖- α -葡萄糖苷的转化。将这些蔗糖异构酶活性与母体多肽或其片段的活性进行对比从而确定活性位点中的哪些氨基酸残基与蔗糖异构化有关。

[0153] 任何常规测定方法例如本文记载的测定方法都可以被用来分别测定母体和变异体的蔗糖异构酶活性。尽管可以采用任意个数的分析测量值来比较活性时, 但是用于酶活性的一个方便的分析测量值是与母体多肽或其片段的 K_m 相比变异体的 Michaelis 常数 K_m 。通常, 每当类似残基通过取代反应被取代时, K_m 增大或减小两倍表明被取代的残基在母体多肽或其片段与底物相互作用的过程中是具有活性的。

[0154] 当对未知或已知的活性氨基酸残基进行扫描氨基酸分析时, 与其直接相邻的氨基酸残基也应当被扫描到。在这种分析中采用的扫描氨基酸可以是任何不同于被取代的氨基酸的氨基酸, 即其它 19 种天然存在的氨基酸中的任何氨基酸。可以制备出三个残基被取代的多肽。其中一种在 N 位上含有一个扫描氨基酸, 优选的是丙氨酸, 其中所述 N 位上是未知的或已知的活性氨基酸。其它两种在 N+1 位点和 N-1 位点上含有扫描氨基酸。如果每个被取代的多肽或其片段对相对底物的 K_m 的影响大于约两倍, 那么扫描氨基酸就在 N+2 位点和 N-2 位点上被取代。对此进行重复直到每个方向上的至少一个, 优选地是四个残基得到鉴定或者直到到达母体多肽或其片段的任意一端, 所述残基对 K_m 的影响小于约两倍。以这种方式, 沿着连续的氨基酸序列可以鉴定出一个或多个与蔗糖向 6-果糖- α -葡萄糖苷转化的催化过程有关的氨基酸。

[0155] 通过氨基酸扫描鉴定的活性氨基酸残基通常是与蔗糖直接接触的氨基酸残基。然而, 活性氨基酸也可以通过与其它残基或小分子如 H_2O 或离子如 Na^+ , Ca^{+2} , Mg^{+2} , 或 Zn^{+2} 形成的盐桥与蔗糖间接接触。

[0156] 在某些情况下, 一个或多个氨基酸残基上的扫描氨基酸的取代导致残基被取代的多肽的产生, 该多肽不能以使得分离量足以进行其蔗糖异构酶活性分析的水平来进行表达。在这类情况下, 可以使用一种不同的扫描氨基酸, 优选地是一种等配氨基酸。

[0157] 优选的扫描氨基酸是较小的、中性的氨基酸。这类氨基酸包括丙氨酸、甘氨酸、丝氨酸和半胱氨酸。由于丙氨酸除去了 β -碳原子以外的侧链并且似乎不太可能改变变异体的主链构象, 所以它是这组氨基酸中的优选的扫描氨基酸。由于丙氨酸是最常用的氨基

酸,所以它也是最优选的。而且,丙氨酸经常出现在掩埋的和暴露的位点 (Creighton, 蛋白质 (The Proteins), W. H. Freeman & Co., N. Y. ;Chothia, 1976, 分子生物学杂志 (J. Mol. Biol), 150 :1)。如果丙氨酸的取代不能产生足够量的变异体,则可以使用等配氨基酸。或者,可以使用下列优先顺序递减的氨基酸 :Ser, Asn 和 Leu。

[0158] 一旦确定了活性氨基酸残基,就可以取代等配氨基酸。这种等配取代不需要发生在所有情况下并且可以在任何活性氨基酸被确定之前进行。进行这种等配氨基酸取代以便使可能对构象产生的破坏作用降到最小,所述破坏作用由一些取代所引起。等配氨基酸如下所示

[0159] 表 C

[0160]

多肽氨基酸	等配扫描氨基酸
Ala (A)	Ser, Gly
Glu (E)	Gln, Asp
Gln (Q)	Asn, Glu
Asp (D)	Asn, Glu
Asn (N)	Ala, Asp
Leu (L)	Met, Ile
Gly (G)	Pro, Ala
Lys (K)	Met, Arg
Ser (S)	Thr, Ala
Val (V)	Ile, Thr
Arg (R)	Lys, Met, Asn
Thr (T)	Ser, Val
Pro (P)	Gly
Ile (I)	Met, Leu, Val
Met (M)	Ile, Leu
Phe (F)	Tyr
Tyr (Y)	Phe
Cys (C)	Ser, Ala
Trp (W)	Phe
His (H)	Asn, Gln

[0161] 本文的方法可被用来检测本发明多肽或其片段的不同区域中的活性氨基酸残基。一旦进行了此鉴定,就可以对母体多肽或其片段进行各种修饰从而改变了母体多肽或其片段与其底物之间的相互作用。

[0162] 2.4.4 由噬菌体展示制备的多肽或肽文库

[0163] 通过利用噬菌体 (或噬菌粒) 展示蛋白配体筛选系统有助于变异体的鉴定,所述系统例如是由 Lowman 等人 (1991, 生物化学 (Biochem). 30 :10832-10838)、Markland 等人 (1991, 基因 (Gene) 109 :13-19)、Roberts 等人 (1992, 美国国家科学院院报 (Proc. Natl. Acad. Sci. (U. S. A.)) 89 :2429-2433)、Smith, G. P. (1985, 科学 (Science) 228 :1315-1317)、Smith, 等人 (1990, 科学 (Science) 248 :1126-1128) 和 Lardner 等人 (美国专利 5, 223, 409) 所记载的系统。通常,该方法包括融合蛋白质的表达,在该融合蛋白中所需蛋白质配体被融合到病毒外壳蛋白 (如 M13 基因 III 外壳蛋白,或 λ 外壳蛋白) 的 N- 端上。

[0164] 在一个实施方案中,对噬菌体文库进行改造以便展示噬菌体外壳蛋白序列中的新型肽。通过利用易错 PCR 对编码本发明多肽或其生物活性片段的基因片段进行的随机诱变或通过 E. Coli 突变株细胞的体内突变来产生新型肽序列。使在噬菌体表面展示的所述新型肽处于与蔗糖或含有蔗糖的底物接触的状态。然后选择展示外壳蛋白的噬菌体,该外壳蛋白含有

能够将蔗糖异构化为 6-果糖- α -葡萄糖苷的肽。可以对被选出的噬菌体进行扩增,可以对编码其外壳蛋白的 DNA 进行测序。以这种方式,可以推测出内含的肽或多肽的氨基酸序列。

[0165] 更详细地,所述方法包括 (a) 构建可复制的表达载体,该载体含有第一个编码本发明多肽或其片段的基因,编码天然或野生型噬菌体外壳蛋白中的至少一部分序列的第二个基因,其中第一个基因与第二个基因是异源的,以及一个转录调节元件,该元件可操作性地与第一个基因和第二个基因进行连接,因而形成编码融合蛋白的融合基因;(b) 在第一个基因内的一个或多个选择位点对所述载体进行突变从而形成一族相关质粒;(c) 用该质粒转化合适的宿主细胞;(d) 用含有编码所述噬菌体外壳蛋白的辅助噬菌体感染被转化的宿主细胞;(e) 在适于形成重组噬菌粒颗粒的条件下培养被感染的转化宿主细胞,所述重组噬菌粒含有所述质粒的至少一部分并且能够转化宿主细胞,调节所述条件使得只有少量的噬菌粒颗粒在颗粒表面展示多于一个拷贝的融合蛋白;(f) 使所述噬菌粒颗粒与蔗糖或含有蔗糖的底物接触;和 (g) 将能使蔗糖异构化为 6-果糖- α -葡萄糖苷的噬菌粒颗粒与不能使蔗糖异构化为 6-果糖- α -葡萄糖苷的噬菌粒颗粒分离。优选地,该方法进一步包括用能将蔗糖异构化为 6-果糖- α -葡萄糖苷的重组噬菌粒颗粒转化合适的宿主细胞并且将步骤 (d) 至 (g) 重复一次或多次。

[0166] 优选地,在该方法中,所述质粒是在转录调节元件的严格控制之下,以及将所述培养条件调节到能使在噬菌粒颗粒表面展示多于一个拷贝的融合蛋白的噬菌粒颗粒的量或数目低于约 20%。更优选地,展示多于一个拷贝的融合蛋白的噬菌粒颗粒的数目比展示单个拷贝的融合蛋白的噬菌粒颗粒的数目低 10%。更优选地,所述数目低于 1%。

[0167] 通常在该方法中,表达载体将进一步含有一个被融合到编码所述多肽的每个亚基的 DNA 上的分泌信号序列并且所述转录调节元件将是一个启动子系统。优选的启动子系统选自 lac Z, λ_{pL} , tac, T7 聚合酶,色氨酸和碱性磷酸酯酶启动子及其组合。通常,该方法还可以使用一个选自 M13K07, M13R408, M13-VCS 和 Phi X 174 的辅助噬菌体。优选的辅助噬菌体是 M13K07,以及优选的外壳蛋白是 M13 噬菌体基因 III 外壳蛋白。所述优选的宿主是 E. Coli 和 E. Coli 的蛋白酶缺陷菌株。

[0168] 通过由多个选择循环过程选择的多个氨基酸变化的噬菌粒选择,反复循环的变异体选择过程被用来选择愈来愈高的结合亲和力。进行了第一轮噬菌粒选择(包括在配体多肽的第一个区域或氨基酸中选择)后,在配体多肽的其它区域或氨基酸中进行另外几轮噬菌粒选择。重复噬菌粒选择的循环过程直到所述多肽的亲合特性符合要求为止。

[0169] 应当认识到形成多肽或其片段的活性位点的氨基酸残基不可能是顺序连接的而且可能定位于所述多肽或其片段的不同亚基上。也就是说,结合区域与活性位点上的特定二级结构有关而与一级结构无关。因此,突变通常被引入编码氨基酸的密码子中,所述氨基酸位于特定二级结构中的定向偏离多肽内部的位点上因而它们将具有与蔗糖或含蔗糖底物相互作用的潜力。

[0170] 本文的噬菌粒展示方法包括将编码多肽或其片段的多核苷酸(多核苷酸 1)融合到第二个多核苷酸(多核苷酸 2)上以便在转录过程中产生融合蛋白。多核苷酸 2 通常是一种噬菌体的外壳蛋白基因,并且优选地它是噬菌体 M13 基因 III 外壳蛋白基因,或其片段。多核苷酸 1 和 2 的融合可以通过将多核苷酸 2 插入到含有多核苷酸 1 的质粒上的特定位点中,或者通过将多核苷酸 1 插入到含有多核苷酸 2 的质粒上的特定位点中来实现。

[0171] 在多核苷酸 1 和多核苷酸 2 之间,可以插入编码终止密码子的 DNA,这类终止密码子是 UAG(琥珀),UAA(赭石),和 UGA(蛋白石)(参见例如,Davis 等,微生物学(Microbiology)(Harper 和 Row:纽约,1980),237,245-247,和 274 页)。在野生型宿主细胞中表达的终止密码子导致没有附连的多核苷酸 2 蛋白的多核苷酸 1 蛋白产物的合成。然而,在抑制性宿主细胞中的生长导致可检测量的融合蛋白的合成。这类抑制性宿主细胞含有一种被修饰的 tRNA,从而在 mRNA 的终止密码子位点插入一个氨基酸,因而导致可检测量的融合蛋白的产生。这种类型的抑制性宿主细胞是熟知的并且是有记载的,如大肠杆菌抑制性菌株,如 JM101 或 XL1-Blue(Bullock 等,1987,生物技术(BioTechniques),5:376-379)。任何可接受的方法都可以被用来将这类终止密码子置于编码该融合多肽的 mRNA 中。

[0172] 可以将可抑制密码子插入到编码多肽或其片段的多核苷酸与编码噬菌体外壳蛋白的至少一部分序列的第二个多核苷酸之间。或者,通过取代所述多肽/片段中的最后一个氨基酸三联体或噬菌体外壳蛋白中的第一个氨基酸可以将所述可抑制终止密码子插入到与融合位点相邻的位点上。当含有所述可抑制密码子的噬菌粒在抑制性宿主细胞中培育时,可以导致可检测的含有所述多肽或其片段和外壳蛋白的融合蛋白的产生。当所述噬菌粒在非抑制性宿主细胞中培育时,由于终止于被插入的编码三联体的可抑制 UAG,UAA 或 UGA,所以大量地合成了不与噬菌体外壳蛋白融合的多肽或其片段。在非抑制性细胞中,由于不存在被融合的噬菌体外壳蛋白,所以多肽被合成后,从宿主细胞中分泌出来,其中所述被融合的噬菌体外壳蛋白以另外的方式将多肽锚定到宿主细胞中。

[0173] 通过一个或多个被选择的密码子的改变可以改变多肽或其片段。所述改变被定义为在编码多肽或其片段的基因中一个或多个密码子的取代、缺失或插入,结果导致氨基酸序列与所述多肽或其片段的未改变的或天然的序列相比发生了变化。优选地,所述改变是在所述分子的一个或多个区域中以任何其它的氨基酸取代至少一个氨基酸。所述改变可以通过所述技术领域多种已知方法来产生,例如 2.3 和 2.4.1 节中所记载的方法。这些方法包括但不限于本文实施例中记载的寡核苷酸介导的诱变和盒式诱变。

[0174] 噬菌粒颗粒文库然后在合适的条件下与蔗糖或含蔗糖的底物进行接触。通常,包括 pH、离子强度、温度等在内的所述条件将模拟生理条件。然后从那些具有低活性的噬菌粒颗粒中选出蔗糖异构酶活性高的噬菌粒颗粒。

[0175] 采用所述被选出的噬菌粒颗粒和辅助噬菌体感染合适的宿主细胞,在适合于所述噬菌粒颗粒扩增的条件下培养所述宿主细胞。然后收集所述噬菌粒颗粒并且将所述选择过程重复一次或多次直到选出对靶分子具有所需亲合力的结合体为止。

[0176] 2.4.5 合理的药物设计

[0177] 本发明的分离多肽的变异体或其生物活性片段还可以利用常规的或合理的药物设计原则来获得,所述的常规的或合理的药物设计原则例如是由 Andrews 等人(见:"PROCEEDINGS OF THE ALFRED BENZON SYMPOSIUM",28 卷,145-165 页,Munksgaard,Copenhagen,1990),McPherson, A. (1990,欧洲生物化学杂志(Eur. J. Biochem). 189:1-24),Hold 等(见:"MOLECULAR RECOGNITION:CHEMICAL AND BIOCHEMICAL PROBLEMS",Roberts, S. M. (编辑.);英国化学协会(Royal Society of Chemistry);84-93 页,1989),Hol, W. G. J. (1989,Arzneim-Forsch. 39:1016-1018),Hol, W. G. J. (1986, Agnew Chem. Int. Ed. Engl. 25:767-778) 所记载的。

[0178] 根据常规药物设计的方法,通过随机检测其结构与本发明的母体多肽或其生物活性片段的结构具有共同属性的分子而获得所需要的变异分子。通过测定母体多肽或多肽片段与候选多肽变异体之间的竞争或协同性可以确定由特定一组结合分子中的变化产生的定量贡献。

[0179] 在合理药物设计的一个实施方案中,所述多肽变异体被设计成具有本发明的多肽或多肽片段的最稳定三维构象的属性。因此,该变异体可以被设计成拥有以一种方式定向的化学基团,所述方式足以引起离子的、疏水的或范德瓦耳斯的相互作用,该相互作用类似于由本发明多肽或多肽片段所显示的相互作用。在合理设计的第二种方法中,特定多肽或多肽片段进行构象“呼吸”的能力得到了利用。这种“呼吸”-不同分子构象的瞬变的和可逆转的假说-是一种很好理解的现象,并且是由温度、热力学因素和分子的催化活性引起的。多肽或多肽片段的三维结构方面的知识有助于这种评估。对多肽或多肽片段的天然构象的变化进行评估有助于识别潜在的铰合位点、由于分子的呼吸等而使氢键、离子键或范德瓦耳斯键在其上可能形成或可能被消除的潜在位点。这类识别允许对多肽或多肽片段可能呈现的附加构象进行鉴定,并且可以实现具有这类构象的模拟多肽变异体的合理设计和生产。

[0180] 用于进行合理模拟设计的优选方法采用了能够呈现多肽或多肽片段(诸如那些利用 RIBBON(Priestle, J., 1988, 分子图形显示杂志(J. Mol. Graphics) 21: 572), QUANTA(Polygen), InSite(Biosyn), 或 Nanovision(美国化学协会(American Chemical Society))获得的)的三维结构的计算机系统。这类分析由 Ho1 等人(见:“分子识别:化学和生物化学的难题(MOLECULAR RECOGNITION:CHEMICAL AND BIOCHEMICAL PROBLEMS)”, 上文, Ho1, W. G. J. (1989, 上文)和 Ho1, W. G. J., (1986, 上文)进行了举例说明。

[0181] 当对候选多肽变异体进行这类直接比较评估时,筛选测定方法可以被用来鉴定这类分子。这类测定方法优选地利用了变异体催化蔗糖转化为6-果糖- α -葡萄糖苷的能力。

[0182] 2.5 多肽衍生物

[0183] 关于本发明的合适衍生物,这类衍生物包括对本发明的多肽、片段或变异体的氨基酸缺失和/或插入,其中所述衍生物催化蔗糖向6-果糖- α -葡萄糖苷的转化。氨基酸的“插入”可以包括本发明多肽、片段和多肽变异体与其它多肽或蛋白质的融合。例如,应当理解为所述多肽、片段或变异体可被掺加到较大的多肽中,而且预计这类较大的多肽也能如上所述催化蔗糖向6-果糖- α -葡萄糖苷的转化。

[0184] 本发明的多肽、片段或变异体可以被融合到其它的蛋白中,例如所述蛋白不是源自原始宿主。所述其它蛋白可能有助于融合蛋白的纯化。例如,多组氨酸标记或麦芽糖结合蛋白可以在这方面被采用,详细内容见以下描述。其它可能的融合蛋白是那些能够产生免疫调节反应的融合蛋白。这类蛋白的特定例子包括蛋白A或谷胱甘肽S-转移酶(GST)。

[0185] 本发明设计的其它衍生物包括但不限于对侧链的修饰,在肽、多肽或蛋白质的合成过程中,非天然氨基酸和/或其衍生物的掺入以及交联剂和其它方法的利用,所述交联剂和其它方法对本发明的多肽、片段和变异体的构象施加限制。

[0186] 由本发明设计的侧链修饰的实例包括氨基的修饰,该氨基的修饰例如通过以下方式进行:用乙酸酐进行酰化;用琥珀酸酐和四氢化邻苯二甲酸酐使氨基酰化;用甲基乙

酰亚氨酸酯进行脒基化 ; 用氰酸盐使氨基甲氨酰化 ; 用 5- 磷酸吡哆醛使赖氨酸吡哆醛化, 接着用 NaBH_4 进行还原 ; 通过与醛反应, 接着用 NaBH_4 进行还原进行还原性烷基化 ; 和用 2, 4, 6- 三硝基苯磺酸 (TNBS) 使氨基三硝基苄基化。

[0187] 羧基的修饰可以通过碳化二亚胺的激活来进行, 先形成 O- 酰基异脒, 接着衍生化成例如相应的酰胺。

[0188] 精氨酸残基的胍基可以通过与试剂如 2, 3- 丁二酮、苯甲酰甲醛和异二醛形成杂环缩合产物来进行修饰。

[0189] 巯基可以通过以下方法如过甲酸氧化成磺基丙氨酸 ; 使用 4- 氯 汞苯磺酸、4- 氯汞苯甲酸酯、2- 氯汞 -4- 硝基苯酚, 氯化苯汞和其它汞制剂形成汞的衍生物 ; 与其它硫醇化合物形成混合的二硫化物 ; 与马来酰亚胺、马来酸酐或其它被取代的马来酰亚胺反应 ; 用碘乙酸或碘乙酰胺进行羧甲基化 ; 和在碱性 pH 下用氰酸盐进行甲氨酰化来进行修饰。

[0190] 色氨酸残基可以例如通过吲哚环与 2- 羟基 -5- 硝基苄基溴或磺酰卤的烷基化作用或通过与 N- 溴代琥珀酰亚胺的氧化作用来进行修饰。

[0191] 酪氨酸残基可以通过与四硝基甲烷发生硝化作用从而形成 3- 硝基酪氨酸衍生物来进行修饰。

[0192] 组氨酸残基的咪唑环可以通过与焦碳酸二乙酯的 N- 乙酯基化作用或通过与碘乙酸衍生物的烷基化作用来进行修饰。

[0193] 在肽合成过程中, 掺入非天然氨基酸和衍生物的实例包括但不限于 4- 氨基丁酸、6- 氨基己酸、4- 氨基 -3- 羟基 -5- 苯基戊酸、4- 氨基 -3- 羟基 -6- 甲基庚酸、叔 - 丁基甘氨酸、正亮氨酸、正缬氨酸、苯基甘氨酸、鸟氨酸、肌氨酸、2- 噻吩基丙氨酸和 / 或氨基酸的 D 型异构体的使用。表 D 列出了本发明设计的一系列非天然氨基酸。

[0194] 表 D

[0195]

非常规氨基酸	非常规氨基酸
α - 氨基丁酸	L-N- 甲基丙氨酸
α - 氨基 - α - 甲基丁酸酯	L-N- 甲基精氨酸
氨基环丙烷 - 羧化物	L-N- 甲基天冬酰胺
氨基异丁酸	L-N- 甲基天冬氨酸
氨基降冰片基 - 羧化物	L-N- 甲基半胱氨酸
环己基丙氨酸	L-N- 甲基谷氨酰胺
环戊基丙氨酸	L-N- 甲基谷氨酸
L-N- 甲基异亮氨酸	L-N- 甲基组氨酸

D- 丙氨酸 D- 精氨酸 D- 天冬氨酸 D- 半胱氨酸 D- 谷氨酸盐 D- 谷氨酸 D- 组氨酸 D- 异亮氨酸 D- 亮氨酸 D- 赖氨酸 D- 蛋氨酸 D- 鸟氨酸 D- 苯丙氨酸 D- 脯氨酸 D- 丝氨酸 D- 苏氨酸 D- 色氨酸 D- 酪氨酸 D- 缬氨酸 D- α - 甲基丙氨酸 D- α - 甲基精氨酸 D- α - 甲基天冬酰胺 D- α - 甲基天冬氨酸盐 D- α - 甲基半胱氨酸 D- α - 甲基谷氨酰胺 D- α - 甲基组氨酸 D- α - 甲基异亮氨酸 D- α - 甲基亮氨酸 D- α - 甲基赖氨酸	L-N- 甲基亮氨酸 L-N- 甲基赖氨酸 L-N- 甲基蛋氨酸 L-N- 甲基正亮氨酸 L-N- 甲基正缬氨酸 L-N- 甲基鸟氨酸 L-N- 甲基苯丙氨酸 L-N- 甲基脯氨酸 L-N- 甲基丝氨酸 L-N- 甲基苏氨酸 L-N- 甲基色氨酸 L-N- 甲基酪氨酸 L-N- 甲基缬氨酸 L-N- 甲基乙基甘氨酸 L-N- 甲基- 叔- 丁基甘氨酸 L- 正亮氨酸 L- 正缬氨酸 α - 甲基- 氨基异丁酸酯 α - 甲基- γ - 氨基丁酸酯 α - 甲基环己基丙氨酸 α - 甲基环戊基丙氨酸 α - 甲基- α - 萘基丙氨酸 α - 甲基青霉胺 N-(4- 氨基丁基) 甘氨酸 N-(2- 氨基乙基) 甘氨酸 N-(3- 氨基丙基) 甘氨酸 N- 氨基- α - 甲基丁酸酯 α - 萘基丙氨酸 N- 苄基甘氨酸
D- α - 甲基蛋氨酸 D- α - 甲基鸟氨酸 D- α - 甲基苯丙氨酸 D- α - 甲基脯氨酸 D- α - 甲基丝氨酸 D- α - 甲基苏氨酸 D- α - 甲基色氨酸 D- α - 甲基酪氨酸 L- α - 甲基亮氨酸 L- α - 甲基蛋氨酸 L- α - 甲基正缬氨酸 L- α - 甲基苯丙氨酸 L- α - 甲基丝氨酸 L- α - 甲基色氨酸 L- α - 甲基缬氨酸 N-(N-2,2- 二苯乙基氨基甲酰甲基) 甘氨酸 1- 羧基-1-(2,2- 二苯基- 乙基氨基) 环丙烷	N-(2- 氨基甲酰乙基) 甘氨酸 N-(2- 氨基甲酰甲基) 甘氨酸 N-(2- 羧乙基) 甘氨酸 N-(2- 羧甲基) 甘氨酸 N- 环丁基甘氨酸 N- 环庚基甘氨酸 N- 环己基甘氨酸 N- 环癸基甘氨酸 L- α - 甲基赖氨酸 L- α - 甲基正亮氨酸 L- α - 甲基鸟氨酸 L- α - 甲基脯氨酸 L- α - 甲基苏氨酸 L- α - 甲基酪氨酸 L-N- 甲基高苯丙氨酸 N-(N-3,3- 二苯丙基氨基甲酰甲基) 甘氨酸

[0196] 还考虑了交联剂的使用,例如利用同-双功能交联剂如双功能亚氨基酯、戊二醛、N- 羟基琥珀酰亚胺酯和异-双功能试剂来稳定本发明的多肽、片段或变异体的 3D 构象,所述亚氨基酯具有其中 $n = 1$ 至 $n = 6$ 的 $(CH_2)_n$ 间隔基团,所述异-双功能试剂通常含有一个氨基活性部分如 N- 羟基琥珀酰亚胺和另外一个基团特异性活性部分如马来酰亚氨基或二硫基部分或碳化二亚胺。另外,可以对肽进行构象上的限制,例如,通过在氨基酸的 C_α 和 C_β 原子之间引入双键、通过掺入 C_α 和 N_α - 甲基氨基酸以及通过引入共价键如在两条侧链之间 N 和 C 端或在一条侧链与肽或其类似物的 N 或 C 端之间形成酰胺键从而形成环状 肽

或其类似物来进行构象上的限制。例如,可以参考:由 Marlowe (1993,生物有机化学与医药化学通讯 (Biorganic & Medicinal Chemistry Letters) 3:437-44) 记载的利用三甲基甲硅烷基酯 (TMSE) 作为正交保护基团在 TFA 树脂上进行肽的环化;由 Pallin 和 Tam (1995,化学协会化学通讯杂志 (J. Chem. Soc. Chem. Comm) 2021-2022) 记载的未被保护的肽通过在水溶液中形成胍而进行的环化;由 Algin 等人 (1994,四面体通讯 (Tetrahedron Letters) 35:9633-9636) 公开的通过赖氨酸侧链的锚定而进行的头尾环化肽的固相合成;由 Kates 等人 (1993,四面体通讯 (Tetrahedron Letters) 34:1549-1552) 所记载的通过三维固相方案而进行的头尾环状肽的生产;由 Tumelty 等人 (1994,化学协会化学通讯杂志 (J. Chem. Soc. Chem. Comm) 1067-1068) 所记载的由固定化的激活中间体进行的环状肽的合成,其中对固定化肽进行激活的同时保持 N-保护基团的完整无缺,然后去除 N-端保护基团从而导致环化;由 McMurray 等人 (1994,肽研究 (Peptide Research) 7:195-206) 所公开的借助天冬氨酸和谷氨酸的侧链与不溶性载体连接的肽的头尾环化;由 Hruby 等人 (1994,反应性聚合物 (Reactive Polymers) 22:231-241) 所教导的用于通过固体载体进行肽环化的替代方法;以及由 Schmid t 和 Langer (1997,肽研究杂志 (J. Peptide Res) 49:67-73) 所公开的用于合成环四肽和环五肽的方法。上述方法可以被用来生产催化蔗糖向 6-果糖- α -葡萄糖苷转化的构象受到限制的多肽。

[0197] 本发明还设计了已经利用普通分子生物学技术进行了修饰的本发明的多肽、片段或变异体以便提高它们对蛋白降解作用的抗性或者优化溶解特性或者使得它们更适合作为致免疫剂。

[0198] 2.6 制备本发明肽的方法

[0199] 本发明的多肽可以通过本技术领域中的技术人员已知的任何适当方法来制备。例如,所述多肽可以通过包括以下步骤的方法来制备:

[0200] (a) 制备含有编码所述多肽的核苷酸序列的重组多核苷酸,所述多肽具有 SEQ ID NO:2,4,8 和 10 中的任意一个所示的序列,或其变异体 或其衍生物,所述核苷酸序列可操作性地连接于转录和翻译调节核酸上;(b) 将所述重组多肽引入适当的宿主细胞;(c) 培养所述宿主细胞以便由所述重组多核苷酸表达重组多肽;和 (d) 分离重组多肽。合适地,所述核苷酸序列含有 SEQ ID NO:1,3,7 和 9 中的任意一个所示的序列。

[0201] 重组多肽优选地是一种表达载体的形式,该表达载体可以是一种自我复制的染色体外载体如质粒,或一种整合到宿主基因组中的载体的形式。

[0202] 转录和翻译调节核酸通常需要适合于用于表达的宿主细胞。已知本技术领域中具有多种类型的适用于多种宿主细胞的表达载体和适宜的调节序列。

[0203] 通常,转录和翻译调节核酸可以包括但不限于启动子序列、前导序列或信号序列、核糖体结合位点、转录起始和终止序列、翻译起始和终止序列和增强子或激活子序列。

[0204] 本发明考虑了本技术领域中的已知的组成型或诱导型启动子。所述启动子既可以是天然存在的启动子,也可以是组合了一个以上启动子元件的杂合启动子。

[0205] 在一个优选的实施方案中,表达载体含有一个能够用来选择转化宿主细胞的选择性标记基因。选择性标记基因在本技术领域是已知的而且还可以随着所用宿主细胞的不同而不同。

[0206] 表达载体还可以含有融合配偶体(通常由表达载体提供)以便使本发明的重组多

肽被表达为与所述融合配偶体形成的融合多肽形式。融合配偶体的主要优势在于其有助于所述融合多肽的鉴定和 / 或纯化。

[0207] 为了表达所述融合多肽,有必要将本发明的多核苷酸连接到表达载体中以使融合配偶体与多核苷酸的翻译读框一致。

[0208] 熟知的融合配偶体的实例包括但不限于谷胱甘肽-S-转移酶 (GST)、人 IgG 的 Fc 片段、麦芽糖结合蛋白 (MBP) 和六组氨酸 (HIS₆), 这些融合配偶体特别适用于通过亲和和色谱分离融合多肽。为了通过亲和和色谱纯化融合多肽的目的,用于亲和和色谱的相关基质包括 但不限于谷胱甘肽-、直链淀粉-和镍-或钴-结合的树脂。许多这类基质是以“试剂盒”的形式提供的,如与 (HIS₆) 融合配偶体共同使用的 QIAexpress™ 系统 (Qiagen) 和 Pharmacia GST 纯化系统。在一个优选的实施方案中,重组多核苷酸在如下文中更加详细描述的商业载体 pFLAG 中进行表达。

[0209] 本技术领域中的另一个熟知的融合配偶体是绿色荧光蛋白 (GFP)。该融合配偶体作为荧光“标记”使得本发明的融合多肽能够通过荧光显微镜或通过流式细胞仪进行鉴定。所述 GFP 标记适用于评估本发明融合多肽的亚细胞定位,或者适用于分离表达本发明融合多肽的细胞。流式细胞计数方法如被荧光激活的细胞分选 (FACS) 特别适用于以后这方面的应用。

[0210] 优选地,所述融合配偶体还具有蛋白酶切割位点,诸如 Xa 因子或凝血酶的切割位点,这些蛋白酶切割位点使得相关蛋白酶能够部分消化本发明的融合多肽因而从中释放出本发明的重组多肽。然后能够通过后续的色谱分离将被释放出的多肽与融合配偶体分离。

[0211] 本发明的融合配偶体在它们的范围内还包括“表位标记”,该表位标记通常是能够获得其特异性抗体的短肽序列。能够容易获得其特异性单克隆抗体的表位标记的已知实例包括 c-Myc、流感病毒、血细胞凝集素和 FLAG 标记。

[0212] 将重组多核苷酸引入宿主细胞的步骤可以通过任何合适的包括转染和转化在内的方法来进行,这些方法的选择取决于被采用的宿主细胞。这类方法为所述技术领域中的技术人员所熟知。

[0213] 本发明的重组多肽可以通过培养用一个表达载体转化的宿主细胞来生产,所述表达载体含有编码本发明的多肽、其生物活性片段、变异体或衍生物的核酸。适于蛋白质表达的条件随着所选表达载体和宿主细胞的变化而变化。这一点可以由本技术领域中的技术人员通过常规实验来确定。

[0214] 用于表达的合适宿主细胞可以是原核细胞或真核细胞。用于表达本发明多肽的一种优选宿主细胞是细菌。所述细菌可以是大肠杆菌。或者,所述宿主细胞可以是昆虫细胞,例如是可以与杆状病毒表达系统一起使用的 SF9 细胞。

[0215] 重组蛋白可以由所述技术领域的技术人员通过利用标准技术方案来制备,所述标准技术方案例如参见记载在 Sambrook 等,《分子克隆,实验室手册》(MOLECULAR CLONING, A LABORATORY MANUAL) (冷泉港出版社 (Cold Spring Harbor Press), 1989), 具体见第 16 节和第 17 节; Ausubel 等,《精编分子生物学 (CURRENT PROTOCOLS IN MOLECULAR BIOLOGY)》(John Wiley & Sons, Inc. 1994-1998), 具体见第 10 和 16 章中的技术方案; 和 Coligan 等,《精编蛋白质科学 (CURRENT PROTOCOLS IN PROTEIN SCIENCE)》(John Wiley & Sons, Inc. 1995-1997), 具体见第 1, 5 和 6 章中的技术方案。

[0216] 或者,本发明的多肽、其片段、变异体或衍生物可以利用溶液合成方法或固相合成方法来进行合成,例如参见 Atherton 和 Shephard(上文)的第9章和 Roberge 等(1995,科学(Science)269:202)。

[0217] 3. 本发明的多核苷酸

[0218] 3.1 分离编码生产 6-果糖- α -葡糖苷的蔗糖异构酶的多核苷酸的方法

[0219] 本发明提供了一种分离编码生产 6-果糖- α -葡糖苷的蔗糖异构酶的新型多核苷酸的方法。该方法包括从特定区域获得环境样品,该环境样品中含有的能够将蔗糖转化为 6-果糖- α -葡糖苷的生物具有选择性优势。所述环境样品可以包括,例如,土壤或包括植物表皮或组织在内的植物材料(例如花)。所述环境样品优选地获自于可以从中能够定期或持久地获得大量蔗糖浓缩物的特定区域,该特定区域包括但不限于涉及含糖植物或植物部分的加工或贮存的工厂和含有被收割的含糖植物的剩余物的田地。优选地,但不是唯一地,所述含糖植物是甜菜或甘蔗。

[0220] 所述方法优选地进一步包括选择或富集双重代谢蔗糖-和 6-果糖- α -葡糖苷的生物,该生物既能利用蔗糖也能利用 6-果糖- α -葡糖苷作为生长的碳源。例如,可以利用含 6-果糖- α -葡糖苷的培养基在足以对代谢 6-果糖- α -葡糖苷的生物进行选择或富集的条件下将所述生物培养一段时间。随后可以利用含蔗糖的培养基在足以对双重代谢 6-果糖- α -葡糖苷和蔗糖的生物进行选择或富集的条件下将如此选择或富集的生物培养一段时间。如果需要,可以颠倒在上述培养基中培养所述生物的顺序。

[0221] 利用至少一种对 6-果糖- α -葡糖苷的产生进行定量分析的测定方法将能够由蔗糖产生 6-果糖- α -葡糖苷的生物筛选出来。优选地,但不是唯一地,所述测定方法是一种苯胺/二苯胺测定方法,例如下文实施例 3 和 4 中所公开方法。或者,或另外,优选地采用一种定量分析蔗糖向 6-果糖- α -葡糖苷转化的测定方法。一种合适的这类测定方法可定量测定 6-果糖- α -葡糖苷产物相对蔗糖和/或相关代谢产物的量。例如下文实施例 5 和 6 中记载的毛细管电泳测定方法可以被用于这方面。

[0222] 然后从 6-果糖- α -葡糖苷产生生物中分离编码蔗糖异构酶的多核苷酸。该分离优选地包括筛选得自于 6-果糖- α -葡糖苷产生生物的核酸文库以及任选地筛选该文库的亚克隆,该亚克隆是编码产生 6-果糖- α -葡糖苷的蔗糖异构酶的多核苷酸的亚克隆。利用对编码蔗糖异构酶的多核苷酸特异的引物或探针,例如本文中所公开的引物或探针,能够适当加快筛选过程。核酸文库优选地是一种表达文库,该文库适当由基因组核酸或 cDNA 来制备。可以利用如上所述的对 6-果糖- α -葡糖苷的产生进行定量的测定方法来检测所需的多核苷酸。实施例 7 至 12 记载了功能性筛选多核苷酸的范例性方案。

[0223] 然后对 6-果糖- α -葡糖苷产生被测试为阳性的克隆进行核酸序列分析从而鉴定出相对于已知蔗糖异构酶是新的基因和/或基因产物。然后将所需产物的酶活性,产率和纯度与已知的参比酶在适当的条件下进行比较,从而鉴定出编码具有较高蔗糖异构酶活性的多肽的分离多核苷酸。

[0224] 3.2 编码本发明的多肽的多核苷酸

[0225] 本发明进一步提供了一种编码如上所定义的多肽、其片段、变异体或衍生物的多核苷酸。在一个实施方案中,所述多核苷酸含有 SEQID NO:1 所示的核苷酸的整个序列。SEQID NO:1 对应于全长的大黄欧文氏菌蔗糖异构酶编码序列。该序列限定

了：(1) 编码信号肽的第一个区域，该区域从第 1 个核苷酸至约第 108 个核苷酸；和 (2) 编码成熟蔗糖异构酶的第二个区域，该区域从约第 109 个核苷酸至第 1899 个核苷酸。合适地，所述多核苷酸含有 SEQ ID NO :3 所示的序列，该序列限定了不含信号序列的编码成熟蔗糖异构酶多肽的区域。本发明的编码序列在相对于 Mattes 等（上文）的编码大黄欧文氏菌蔗糖异构酶的多核苷酸的 3' 端含有一个附加的 594 个碱基对的序列。

[0226] 在另一个实施方案中，所述多核苷酸含有 SEQ ID NO :8 所示核苷酸的完整序列。SEQ ID NO :8 对应于细菌分离物 68J 的 1791 个碱基对的全长蔗糖异构酶编码序列。SEQ ID NO :12 限定了 (1) 编码信号肽的第一个区域，该区域从第 1 个核苷酸至约第 99 个核苷酸；和 (2) 编码成熟蔗糖异构酶的第二个区域，该区域从约第 100 个核苷酸至第 1791 个核苷酸。合适地，所述多核苷酸含有 SEQ ID NO :10 所示序列，该序列限定了不含信号序列的编码成熟蔗糖异构酶多肽的区域。

[0227] 3.3 多核苷酸变异体

[0228] 通常，本发明的多核苷酸变异体含有与相同大小的参比多核苷酸序列（“比较窗口”）相比或当与一个其中的排列是由本领域中已知的计算机同源性程序进行的排列序列相比时具有至少 60%，更合适地至少 70%，优选地至少 80% 以及更优选地至少 90% 序列相同性的区域。通过常规技术可以确定构建合适变异体的因素。例如，SEQ IDNO :1, 3, 7 和 9 中的任意一个所示的多核苷酸可以通过对早期制备的变异体或非变异体形式的本发明的分离的天然启动子进行随机诱变（例如转座子诱变）、寡核苷酸介导的（或定点的）诱变、PCR 诱变和盒式诱变来进行突变。

[0229] 寡核苷酸介导的诱变是一种用于制备本发明多核苷酸的核苷酸取代变异体的优选方法。该技术为本技术领域所熟知，例如采用 Adelman 等人（1983, DNA 2 :183）所记载的方法。扼要地说，通过将编码所需突变的寡核苷酸与模板 DNA 杂交来改变 SEQ ID NO :1, 3, 7 或 9 中的任何一个所示的多核苷酸，其中所述模板是单链形式的含有未改变的或母体 DNA 序列的质粒或噬菌体。杂交后，使用 DNA 聚合酶来合成所述模板的第二条完整互补链，该互补链将被掺入寡核苷酸引物中并且编码所述母体 DNA 序列中所选择的改变。

[0230] 通常，使用长度至少为 25 个核苷酸的寡核苷酸。最适的寡核苷酸将具有 12 至 15 个与编码突变的核苷酸的任意一侧上的模板完全互补的核苷酸。这确保了寡核苷酸能够与单链 DNA 模板分子适当地杂交。

[0231] DNA 模板可以由噬菌体 M13 载体衍生的那些载体或由 Viera 等人（1987, 酶学方法 (Methods Enzymol) 153 :3) 记载的含有单链噬菌体复制起点的那些载体来产生。因此，可以将待突变的 DNA 插入到一个载体中来生产单链模板。例如 Sambrook 等人（1989, 上文）的 4.21-4.41 节记载了单链模板的生产方法。

[0232] 或者，可以通过利用标准技术对双链质粒（或其它 DNA）进行变性来生成单链模板。

[0233] 为了改变天然 DNA 序列，在适当的杂交条件下使寡核苷酸与单链模板杂交。然后加入一种 DNA 聚合酶，通常是 DNA 聚合酶 I 的 Klenow 片段，利用寡核苷酸作为合成的引物来合成模板的互补链。由此形成了异源双链分子以致于 DNA 的一条链编码被测试的多肽或其片段的突变形式，而另一条链（原始模板）编码被测试的多肽或其片段的天然的未改变的序列。然后将该异源双链分子转化到合适的宿主细胞中，通常是一种原核生物如大肠杆

菌中。对所述细胞进行培养后,将其涂布到琼脂平板上并利用含有可检测标记的寡核苷酸引物进行筛选从而鉴定出含有突变 DNA 的细菌菌落。然后利用常规技术将产生的突变 DNA 片段克隆到合适的表达宿主如大肠杆菌中并且检测保留了所需蔗糖异构酶活性的克隆。一旦利用随机诱变技术得到了克隆,就必须对阳性克隆进行测序以便检测突变。

[0234] 或者, DNA 的接头扫描诱变可以被用来在已被克隆到质粒载体中的整个目的序列中引入点突变簇。例如,可以参考 Ausubel 等人的文章,上文,(具体地是 8.4 章),其中记载了利用互补寡核苷酸和需要一个与待诱变区域相邻的独特的限制位点的第一个方案。在所述区域中首先生成一系列连锁式缺失突变。合成一对互补的寡核苷酸来填补缺失终点上的接头与附近限制位点间的目的序列中的缺口。该接头序列实际上提供了所需的点突变簇,因为按照它在缺失突变系列的被改变的终点上的位置而在所述区域进行移动或“扫描”。Ausubel 等人,上文还记载了另一个方案,该方案利用定点诱变方法将小的点突变簇引入整个靶区域中。简要地说,通过将一条含有一个或多个错配的合成寡核苷酸退火到被克隆到单链 M13 载体中的目的序列上而将突变引入该序列中。在大肠杆菌 *dut⁻ung⁻* 菌株中培养该模板,从而使得尿嘧啶被掺入到该模板链中。将寡核苷酸退火到经纯化的模板上并且利用 T4DNA 聚合物进行延伸从而产生双链异源双链体。最后,将该异源双链体引入野生型大肠杆菌菌株中,由于模板链中尿嘧啶的存在可防止模板链的复制,因而产生仅含突变 DNA 的噬菌斑。

[0235] 也可以通过利用 PCR 的区域特异性诱变和定向诱变来构建本发明的多核苷酸变异体。在这方面,可以参见,例如 Ausubel 等的文章,上文,具体地是 8.2A 和 8.5 章。

[0236] 或者,可以按照以下步骤来制备本发明的合适的多核苷酸序列变异体:(i) 制备任选地是简并的引物,其中每个引物含有编码本发明的参比多肽或其片段的参比多核苷酸的一部分序列,优选地含有编码 SEQ ID NO:1,3,7 或 9 中的任何一个所示序列的一部分序列;(ii) 从代谢蔗糖的生物体内获得核酸提取物,所述生物优选地是细菌,更优选地是从获自于特定区域的物种中获得核酸提取物,其中所述特定区域中的能够将蔗糖转化为 6-果糖- α -葡萄糖苷的生物能够获得本文所记载的选择优势;(iii) 通过核酸扩增技术,利用所述引物从所述核酸提取物中扩增至少一种扩增产物,其中所述扩增产物相应于多核苷酸变异体。

[0237] 合适的核酸扩增技术是技术人员所熟知的,并且包括例如由 Ausubel 等人(上文)所记载的聚合酶链反应(PCR);例如在美国专利 US 5,422,252 中记载的链取代扩增(SDA);例如由 Liu 等人(1996,美国化学协会杂志(J. Am. Chem. Soc)118:1587-1594 和国际专利申请 WO 92/01813)和 Lizardi 等人(国际专利申请 WO 97/19193)记载的滚环复制(RCR);例如由 Sooknanan 等人(1994,生物技术(Biotechniques)17:1077-1080)所记载的基于核酸序列的扩增(NASBA);和例如由 Tyagi 等人(1996,美国国家科学院院报(Proc. Natl. Acad. Sci)USA 93:5395-5400)记载的 Q- β 复制酶扩增。

[0238] 通常,与参比多核苷酸基本互补的多核苷酸变异体是通过印迹技术进行鉴定的,该技术包括将核酸固定在一基质(优选地是一种合成膜如硝酸纤维素)上的步骤,接着是杂交步骤,以及检测步骤。DNA 印迹法被用来鉴定互补 DNA 序列;RNA 印迹法被用来鉴定互补 RNA 序列。点印迹法和狭线印迹法可以被用来鉴定互补 DNA/DNA, DNA/RNA 或 RNA/RNA 多核苷酸序列。这类技术对于所述技术领域的技术人员来说是熟知的,并且已经被记载在

Ausubel 等人撰写的文章(1994-1998, 上文)第 2.9.1 至 2.9.20 页中。

[0239] 按照这类方法, DNA 印迹法包括通过凝胶电泳将 DNA 分子按照大小进行分离, 将按大小分离的 DNA 转移到合成膜上, 使与膜结合的 DNA 与放射性标记的、酶标记的或荧光标记的互补核苷酸序列杂交。在点印迹法和狭线印迹法中, DNA 样品在进行上述杂交之前被直接点到合成膜上。

[0240] 当鉴定 cDNA 或基因组 DNA 文库中的互补多核苷酸时可以使用另一种印迹步骤, 如通过噬菌斑或菌落杂交的方法。该方法的一个典型的实例被记载在 Sambrook 等人撰写的文章《分子克隆, 实验室手册》(Molecular Cloning, A Laboratory Manual), (冷泉港出版社 (Cold Spring Harbour Press), 1989) 的第 8-12 章中。

[0241] 通常, 采用下列常规方法来确定杂交条件。按照上述方法将多核苷酸印迹 / 转移到合成膜上。按照上述方法标记参比多核苷酸如本发明的多核苷酸, 并分析该被标记的多核苷酸与被固定的多核苷酸进行杂交的能力。

[0242] 一名技术人员应当认识到许多因素都影响杂交。放射性标记的多核苷酸序列的比活性通常应该大于或等于约 10^8 dpm/mg 从而提供了一种可检测的信号。比活性为 10^8 至 10^9 dpm/mg 的放射性标记的核苷酸序列能够检测大约 0.5 pg 的 DNA。在本技术领域已知为了能够进行测定, 必须将足量的 DNA 固定在膜上。理想的是使用过量的固定化 DNA, 通常为 10 μ g。在杂交过程中添加惰性聚合物如 10% (w/v) 的葡聚糖硫酸酯 (分子量 (MW) 为 500,000) 或聚乙二醇 6000 也能够增加杂交的灵敏度 (参见 Ausubel 上文 2.10.10)。

[0243] 为了从被固定在膜上的多核苷酸与被标记的多核苷酸之间的杂交获得有意义的结果, 在洗涤步骤后, 必须使足量的被标记的多核苷酸与被固定的多核苷酸进行杂交。洗涤步骤确保被标记的多核苷酸只与具有所需与被标记的多核苷酸的互补度的被固定的多核苷酸杂交。

[0244] 应当理解本发明的多核苷酸变异体能够在低严格条件下与参比多核苷酸杂交。本文所指的低严格条件包括用于在 42°C 下杂交的至少约 1% v/v 到至少约 15% v/v 的甲酰胺和至少约 1M 到至少约 2M 的盐, 以及用于在 42°C 下洗涤的至少约 1M 到至少约 2M 的盐。低严格条件还可以包括用于在 65°C 下杂交的 1% 的牛血清白蛋白 (BSA), 1mM EDTA, 0.5M NaHPO₄ (pH 7.2), 7% SDS, 以及用于在室温下洗涤的 (i) 2xSSC, 0.1% SDS; 或 (ii) 0.5% BSA, 1mM EDTA, 40mMNaHPO₄ (pH 7.2), 5% SDS。

[0245] 合适地, 多核苷酸变异体在至少中等严格条件下与参比多核苷酸进行杂交。中等严格条件包括用于在 42°C 下杂交的至少约 16% v/v 至少约 30% v/v 的甲酰胺和至少约 0.5M 到至少约 0.9M 的盐, 以及用于在 55°C 下洗涤的至少约 0.1M 到至少约 0.2M 的盐。中等严格条件还可以包括用于在 65°C 下杂交的 1% 的牛血清白蛋白 (BSA), 1mMEDTA, 0.5M NaHPO₄ (pH 7.2), 7% SDS, 和用于在 60-65°C 下洗涤的 (i) 2xSSC, 0.1% SDS; 或 (ii) 0.5% BSA, 1mM EDTA, 40mMNaHPO₄ (pH 7.2), 5% SDS。

[0246] 优选地, 多核苷酸变异体在高严格条件下与参比多核苷酸进行杂交。高严格条件包括用于在 42°C 下杂交的至少约 31% v/v 至少约 50% v/v 的甲酰胺和至少约 0.01M 到至少约 0.15M 的盐, 以及用于在 55°C 下洗涤的至少约 0.01M 到至少约 0.02M 的盐。高严格条件还可以包括用于在 65°C 下杂交的 1% 的牛血清白蛋白 (BSA), 1mM EDTA, 0.5M NaHPO₄ (pH 7.2), 7% SDS, 和用于在高于 65°C 的温度下洗涤的 (i) 0.2xSSC, 0.1% SDS; 或 (ii) 0.5%

BSA, 1mM EDTA, 40mM NaHPO₄ (pH 7.2), 1% SDS。

[0247] 其它严格条件为所述技术领域所熟知。一名技术人员应当认识到可以利用多种因素来优化杂交的特异性。对最后洗涤步骤的严格性进行优化能够起到确保高度杂交的作用。更详细的实例参见 Ausubel 等撰写的文章, 上文, 第 2.10.1 至 2.10.16 页和 Sambrook 等撰写的文章 (1989, 上文), 1.101 至 1.104 节。

[0248] 尽管通常在约 42°C 下至 68°C 的温度下进行严格洗涤, 但是本技术领域的技术人员应当认识到其它温度对于严格条件也可能是合适的。最大杂交率通常出现在比形成 DNA-DNA 杂合体的 T_m 低约 20°C 至 25°C 的温度下。在本技术领域中已知 T_m 是解链温度, 或者是两条互补多核苷酸序列解离时的温度。评估 T_m 的方法是所述技术领域中熟知的 (参见上文, Ausubel 等, 第 2.10.8 页)。

[0249] 通常, 完全配对的 DNA 双链体的 T_m 的近似值可以通过以下公式来预测:

$$[0250] \quad T_m = 81.5 + 16.6 (\log_{10} M) + 0.41 (\% G+C) - 0.63 (\% \text{甲酰胺}) - (600 / \text{长度})$$

[0251] 其中: M 是 Na⁺ 的浓度, 优选地范围在 0.01 摩尔至 0.4 摩尔; % G+C 是鸟苷和胞嘧啶碱基的总数相对碱基总数的百分率, 范围在 30% 至 75% G+C; % 甲酰胺是甲酰胺的体积百分比浓度; 长度是 DNA 双链体中的碱基对的数目。

[0252] 伴随着随机错配的碱基对数目每增加 1%, 双链体 DNA 的 T_m 就降低约 1°C。洗涤对于高严格条件通常在 $T_m - 15^\circ\text{C}$ 下, 或对于中等严格条件通常在 $T_m - 30^\circ\text{C}$ 下进行。

[0253] 在一个优选的杂交方法中, 一种含有固定化 DNA 的膜 (例如硝酸纤维素膜或尼龙膜) 于 42°C 下在含有被标记探针的杂交缓冲液 (50% 去离子化的甲酰胺, 5xSSC, 5x Denhardt 溶液 (0.1% Ficoll, 0.1% 聚乙烯吡咯烷酮和 0.1% 牛血清白蛋白), 0.1% SDS 和 200mg/Ml 变性鲑精 DNA) 中杂交过夜。对所述膜顺序进行两次中等严格条件的洗涤 (即, 在 45°C 下 2xSSC, 0.1% SDS 洗涤 15 分钟, 接着 50°C 下 2xSSC, 0.1% SDS 洗涤 15 分钟), 然后顺序进行两次较高严格条件的洗涤 (即, 在 55°C 下 0.2xSSC, 0.1% SDS 洗涤 12 分钟, 接着在 65-68°C 下 0.2xSSC 和 0.1% SDS 溶液洗涤 12 分钟)。

[0254] 用于检测被标记的多核苷酸与被固定的多核苷酸杂交的方法对于所述技术领域的专业人员来说是熟知的。这类方法包括放射自显影术、磷光成像和化学发光法、荧光和比色检测法。

[0255] 4. 抗原结合分子

[0256] 本发明还设计与上述多肽、片段、变异体和衍生物特异性结合的抗原结合分子。优选地, 本发明的抗原结合分子与 SEQ ID NO: 2, 4, 8, 10, 19, 20, 21, 22, 23 和 24 所示的任意一个或多个氨基酸序列或其变异体具有免疫相互作用。

[0257] 例如, 所述抗原结合分子可以含有完整的多克隆抗体。这类抗体可以, 例如通过将本发明的多肽、片段、变异体或衍生物注射到生产物种中从而获得多克隆抗血清来进行制备, 其中所述生产物种包括小鼠或兔子。生产多克隆抗体的方法是本技术领域的技术人员所熟知的。可以使用的范例性方案例如被记载在 Coligan 等人的精编免疫学 (CURRENT PROTOCOLS IN IMMUNOLOGY), (John Wiley & Sons, Inc, 1991) 和 Ausubel 等人的文章 (1994-1998, 上文), 具体是第 11 章的第 III 节中。

[0258] 代替在生产物种中获得的多克隆抗血清, 可以利用例如由 Köhler 和 Milstein (1975, 自然 (Nature) 256, 495-497) 所记载的标准方法, 或者通过例如由 Coligan

等人(1991, 上文)所记载的上述方法的更新改进方法来生产单克隆抗体, 所述更新改进方法是通过使源自生产物种的脾或其它产生抗体的细胞无限增殖, 其中所述生产物种中已经被注入了本发明的多肽、片段、变异体或衍生物中的一种或多种。

[0259] 本发明还设计了抗原结合分子的 Fv, Fab, Fab' 和 F(ab')₂ 免疫球蛋白片段。

[0260] 或者, 所述抗原结合分子可以包括一种合成的稳定化 Fv 片段。这类片段的范例性片段包括单链 Fv 片段 (sFv, 经常被称作 scFv), 在该片段中肽接头 (肽连接序列) 被用来分别桥接 V_H 区的 N 端或 C 端与 V_L 区的 C 端或 N 端。ScFv 缺少整个抗体的所有恒定区域并且不能激活补体。用于连接 V_H 区和 V_L 区的合适肽接头是那些能够使 V_H 区和 V_L 区折叠成具有抗原结合位点的多肽单链的接头, 所述抗原结合位点的三维结构与从中能够获得 Fv 片段的整个抗体的抗原结合位点的三维结构相类似。具有所需特性的接头可以通过美国专利 US4, 946, 778 中所公开的方法来获得。然而, 在某些情况下不存在接头。ScFvs 可以按照例如 Kreber 等人 (Kreber 等 1997, 免疫学方法杂志 (J. Immunol. Methods) ;201(1) : 35-55) 中所概括的方法来制备。或者, 它们可以通过利用美国专利 US 5, 091, 513, 欧洲专利 EP 239, 400 或由 Winter 和 Milstein(1991, 自然 (Nature)349 :293) 和 Plickthun 等人 (1996, 在抗体工程中 (In Antibodyengineering) : 实用性方法 (A practical approach) 203-252) 所撰写的文章中记载的方法来进行制备。

[0261] 或者, 所述合成稳定化 Fv 片段含有一个通过二硫键稳定的 Fv(dsFv), 其中所述通过二硫键稳定的 Fv 中的半胱氨酸残基被引入 V_H 区和 V_L 区致使在完全折叠的 Fv 分子中的两个残基之间能够形成二硫键。生产 dsFv 的合适方法被记载在例如 (Glockscutter 等, 生物化学 (Biocheml) 29 :1363-1367 ;Reiterd 等, 1994, 生物化学杂志 (J. Biol. Chem) 269 : 18327-18331 ;Reiter 等, 1994, 生物化学 (Biochem) 33 :5451-5459 ;Reiter 等, 1994, 癌症研究 (CancerRes) 54 :2714-2718 ;Webber 等, 1995, 分子免疫学 (Mol. Immunol) 32 : 249-258) 中。

[0262] 所设计的抗原结合分子是例如由 Ward 等人 (1989, Nature 341 :544546) ; Hamers-Casterman 等人 (1993, Nature. 363 :446-448) ;Davies & Riechmann, (1994, FEBS Lett. 339 :285-290) 所公开的单可变区 (被称作 dAbs)。

[0263] 或者, 所述抗原结合分子可以含有一种“小体 (minibody)”。在这点上, 小体是小形式的完整抗体, 其以单链的形式编码完整抗体的必需元件。合适地, 所述小体由被融合到例如由美国专利 US5, 837, 821 公开的免疫球蛋白分子的铰链区和 CH3 区上的天然抗体的 V_H 区和 V_L 区组成。

[0264] 在另一个实施方案中, 所述抗原结合分子可以含有非免疫球蛋白来源的蛋白构架。例如, 可以参考 Ku & Schultz, (1995, 美国国家科学院院报 (Proc. Natl. Acad. Sci. USA), 92 :652-6556) 的文章, 其中公开了一种四螺旋束蛋白细胞色素 b562, 该细胞色素具有两个被随机化而形成互补性决定区 (CDRs) 的环, 所述互补性决定区被选择用于结合抗原。

[0265] 抗原结合分子可以是多价的 (即, 具有一个以上的抗原结合位点)。这类多价分子对于一个或多个抗原可以是特异性的。这类多价分子可以通过例如由 Adams 等人 (1993, 癌症研究 (Cancer Res) 53 :4026-4034) 和 Cumber 等人 (1992, 免疫学杂志 (J. Immunol) 149 : 120-126) 所公开的含半胱氨酰的肽使两个抗体片段二聚化来进行制备。或者, 可以通过将

抗体片段融合到自然发生二聚化的两亲性螺旋体上来促进二聚化 (Pack P. Plünckthun, 1992, 生物化学 (Biochem) 31 :1579-1584), 或者通过利用优先进行异二聚化的区域 (Kostelny 等, 1992, 免疫学杂志 (J. Immunol) 148 :1547-1553) 来促进二聚化。在另一个实施方案中, 所述多价分子可以含有一个多价单链抗体 (多价 -scFv), 该多价单链抗体含有至少两个通过肽接头连接在一起的 scFvs。在这方面, 可以使用非共价或共价连接的被称作“二体 (diabodies)”的 scFv 二聚体。根据所采用的具有不同抗原结合特异性的 scFv 的数目, 多价 -scFv 可以是双特异性的或较大的。多价 -scFv 可以例如通过美国专利 US 5, 892, 020 中所公开的方法来进行制备。

[0266] 本发明的抗原结合分子可以被用于亲和色谱来分离本发明的天然或重组的多肽或生物活性片段。例如可以参考由 Coligan 等人撰写的文章 (1995-1997, 上文) 的第 9.5 章所记载的免疫亲和色谱方法。

[0267] 所述抗原结合分子可以被用来筛选本文所记载的本发明的变异多肽的表达文库。它们还可以被用来检测和 / 或分离本发明的多肽、片段、变异体和衍生物。因此, 本发明还设计了抗原结合分子用于分离蔗糖异构酶的用途, 利用例如任何合适的基于免疫亲和性的方法, 所述基于免疫亲和性的方法包括但不限于免疫色谱法和免疫沉淀法。一种优选的方法利用了固相吸附, 在该方法中将抗蔗糖异构酶的抗原结合分子附连到合适的树脂上, 使所述树脂与被怀疑含有蔗糖异构酶的样品接触, 如果存在蔗糖异构酶的话, 随后将蔗糖异构酶从树脂中洗脱出来。优选的树脂包括: **Sepharose®** (Pharmacia), **Poros®** 树脂 (Roche Molecular Biochemicals, Indianapolis), ActigelSuperflow™ 树脂 (Sterogene Bioseparations Inc., Carlsbad Calif.) 和 Dynabeads™ (Dynal Inc., Lake Success, N. Y.)。

[0268] 5. 检测方法

[0269] 5.1 本发明多肽的检测

[0270] 本发明还涉及到一种检测样品中的如上文概述的多肽、片段、变异体或衍生物的方法, 该方法包括使样品与第 4 节所记载的抗原结合分子接触, 然后检测所述被接触样品中的含有所述抗原结合分子和所述多肽、片段、变异体或衍生物的复合物。

[0271] 任何用于测定复合物形成的合适技术都可以被采用。例如, 本发明的其中结合了报道分子的抗原结合分子可以被用于免疫测定。这类免疫测定包括但不限于本技术领域的技术人员所熟知的放射免疫测定 (RIA), 酶联免疫吸附测定 (ELISA) 和免疫色谱技术 (ICT), 蛋白质印迹分析。例如, 可以参考“精编免疫学 (CURRENT PROTOCOLS IN IMMUNOLOGY)” (1994, 上文), 其中记载了可被用于本发明的各种免疫测定。免疫测定可以包括所述技术领域所理解的或例如下文中所记载的竞争性测定。应当理解本发明包括定性和定量的免疫测定。

[0272] 合适的免疫测定技术被记载在例如美国专利 US 4, 016, 043、US4, 424, 279 和 US 4, 018, 653 中。这些免疫测定技术既包括非竞争类型的单位点测定也包括非竞争类型的双位点测定, 以及传统的竞争性结合测定。这些测定方法还包括被标记的抗原结合分子与靶抗原的直接结合。

[0273] 两种位点测定方法尤其适用于本发明。存在着这些测定方法的多种变化形式, 所有变化形式的测定方法都被包括在本发明中。简言之, 在一种典型的正向测定法中, 将未被

标记的抗原结合分子如未被标记的抗体固定在一固体基质上并使待测样品与被结合的分子接触。进行一段合适时间的温育后,保持足够的一段时间从而形成一种抗体-抗原复合物,然后加入被能够产生可检测信号的报告分子标记的另一种抗原结合分子,合适地是特异于抗原的第二种抗体,并且温育足够长的时间以便形成另一种抗体-抗原-被标记抗体的复合物。将任何未反应的物质洗掉后通过观察由报告分子产生的信号来确定抗原的存在。通过简单观察可视信号可以得到定性结果,或者通过与含有已知量抗原的对照样品进行比较可以得到定量结果。正向测定方法的变化形式包括同步测定方法,在该方法中样品和被标记抗体被同时加到被结合的抗体中。这些技术是本技术领域熟知的技术,包括显而易见的较小的改动。根据本发明,所述样品是可能含有蔗糖异构酶如来源于蔗糖-代谢生物的蔗糖异构酶。优选地,所述蔗糖-代谢生物是细菌,该细菌合适地获自于其中能够将蔗糖转化为 6-果糖- α -葡萄糖苷的生物具有选择性优势的特定区域。

[0274] 在一个典型的正向测定方法中,对抗原或其抗原部分具有特异性 的第一抗体被共价或被动地结合到一种固体表面。所述固体表面通常是玻璃或聚合物,最常用的聚合物是纤维素、聚丙烯酰胺、尼龙、聚苯乙烯、聚氯乙烯或聚丙烯。所述固体载体可以是管状形式的、珠粒形式的、盘状小平板形式的,或适于进行免疫测定的任何其它表面形式的。所述结合方法是本技术领域熟知的并且通常由交联、共价结合或物理吸附组成。在制备过程中洗涤用于测试样品的聚合物-抗体复合物。然后将待测的等分样品加到固相复合物中并且在合适的条件下温育足够长的时间以便使存在的所有抗原都结合到所述抗体上。在温育期以后,抗原-抗体复合物被洗涤和干燥并与特异于该抗原的一部分的第二抗体一起进行温育。第二抗体通常含有与其连接的报告分子,该报告分子被用来显示第二抗体与抗原的结合。通过被连接的报告分子确定的被标记抗体的结合量与结合到被固定的第一抗体上的抗原量成比例。

[0275] 另一种方法包括将抗原固定在生物学样品中,然后使被固定的抗原与可能被报告分子标记的或可能没被报告分子标记的特异性抗体接触。根据靶的数量和报告分子信号的强度,通过采用抗体直接进行标记可以对被结合的抗原进行检测。或者,使特异于第一抗体的被标记的第二抗体与靶-第一抗体复合物接触从而形成靶-第一抗体-第二抗体三元复合物。该复合物通过由报告分子发出的信号进行检测。

[0276] 从上述内容应当理解到与抗原结合分子连接的报告分子可包括以下几种:

[0277] (a) 报告分子与抗原结合分子直接相连;

[0278] (b) 报告分子与抗原结合分子间接相连;即报告分子与另一种分析试剂相连后再与抗原结合分子结合;和

[0279] (c) 与抗原结合分子的后续反应产物相连。

[0280] 所述报告分子可以选自:色原、催化剂、酶、荧光染料、化学发光分子、镧系元素离子如铕(Eu^{3+})、放射性同位素和直视标记物。

[0281] 对于直视标记物来说,可以使用胶体金属或非金属颗粒、染料颗粒、酶或底物、有机聚合物、乳胶颗粒、脂质体或含有产生信号的物质的其它囊泡等。

[0282] 适于用作报告分子的多种酶被公开在美国专利 US 4,366,241, US 4,843,000 和 US,4,849,338 的说明书中。用于本发明的合适酶包括碱性磷酸酶、辣根过氧化物酶、荧光素酶、 β -半乳糖苷酶、葡萄糖氧化酶、溶菌酶、苹果酸脱氢酶等。这些酶可被单独使用或与溶

液中的第二种酶联合使用。

[0283] 合适的荧光染料包括但不限于异硫氰酸荧光素 (FITC)、异硫氰酸四甲基碱性蕊香红 (TRITC)、R-藻红素 (RPE) 和德克萨斯红 (TexasRed)。其它范例性的荧光染料包括那些由 Dower 等人 (国际专利 W093/06121) 论述的荧光染料。可以参考美国专利 US 5,573,909 (Singer 等), US 5,326,692 (Brinkley 等) 中记载的荧光染料。或者,可以参考美国专利 US 5,227,487, US 5,274,113, US5,405,975, US 5,433,896, US 5,442,045, US 5,451,663, US5,453,517, US 5,459,276, US 5,516,864, US 5,648,270 和 US5,723,218 中记载的荧光染料。

[0284] 在酶免疫测定的情况中,通常借助戊二醛或过碘酸盐将酶缀合到第二抗体上。然而,很容易理解,存在着对技术人员来说容易得到的多种不同缀合技术。与特异性酶一起使用的底物通常被选择用于通过相应酶的水解作用,产生一种可检测的颜色变化。合适的酶的实例包括上文记载的那些酶。也可以采用产生荧光产物的荧光底物,而不采用上述的显色底物。在所有情况下,将酶-标记的抗体加到第一抗体-抗原复合物中。然后使其结合,并将过量的试剂洗掉。然后将含有合适底物的溶液加到抗体-抗原-抗体的复合物中。所述底物与被连接到第二抗体上的酶发生反应,结果产生定性的可视信号,该信号通常可以利用分光光度测定法来进一步进行定量,从而显示出存在于样品中的抗原量。

[0285] 荧光化合物如荧光素、碱性蕊香红和镧系元素-铕 (EU) 可以交替地被化学偶联到抗体上而不改变它们的结合能力。当采用特定波长的光进行照射而被激活时,荧光染料标记的抗体可吸收光能,从而在分子中诱导一种可激发性状态,接着发出可利用光学显微镜进行目视检测的具有特征性颜色的光。使荧光标记的抗体与第一抗体-抗原复合物结合。将未结合的试剂洗掉后,使剩下的三元复合物暴露在波长合适的光下。观察到的荧光表明所需抗体的存在。本技术领域已经建立起了免疫荧光测定法 (IFMA)。然而,也可以采用其它报道分子,诸如放射性同位素、化学发光或生物发光分子。

[0286] 5.2 本发明的多核苷酸的检测

[0287] 在另一个实施方案中,测定方法包括检测编码所述多肽、片段、变异体或衍生物的多核苷酸在细胞中的表达。所述多核苷酸的表达可以利用任何合适的技术来检测。例如,编码所述成员的被标记的多核酸可以在获自于肌细胞的 RNA 提取物的 RNA 印迹中被用作探针。优选地,在核酸扩增反应如 RT-PCR 中,源自于动物的核酸提取物与对应于编码所述成员的多核苷酸的有义和反义序列或其侧翼序列的寡核苷酸引物一起使用。许多自动固相检测技术也是适用的。例如规模非常大的固定化引物阵列 (VLSIPS™) 被用于核苷酸的检测,例如由 Fodor 等人 (1991, 科学 (Science) 251:767-777) 和 Kazal 等人 (1996, 自然医学 (Nature Medicine) 2:753-759) 记载的核苷酸的检测。上述常规技术对于本技术领域的技术人员来说是熟知的。

[0288] 6. 嵌合核酸构建体

[0289] 6.1 原核表达

[0290] 本发明进一步涉及被设计用于遗传转化原核细胞的嵌合核酸构建体,该构建体含有与启动子序列可操作性连接的本发明的多核苷酸、片段或变异体。优选地,所述嵌合构建体在革兰氏阴性原核细胞中是可操作性的。可被用作构建嵌合核酸构建体基础的多种原核表达载体可以被用来表达本发明的多核苷酸、片段或变异体。这些载体包括但不限于

染色体载体（例如噬菌体如 λ 噬菌体），染色体外载体（例如质粒或粘粒表达载体）。所述表达载体还通常含有一个能够使载体进行自主复制的复制起点和一个或多个能够使被转化细胞进行表型选择的基因。含有组成型和诱导型启动子序列的许多合适启动子序列中的任何一种都可被用在表达载体（例如参见 Bitter 等, 1987, 酶学方法 (Methods in Enzymology) 153 :516-544) 中。例如, 可以使用诱导型启动子如 γ 噬菌体的 pL、plac、ptrp、ptac ptrp-lac 杂合启动子等。然后可以用嵌合核酸构建体转化所需的原核宿主细胞从而制备重组的原核宿主细胞, 该重组的原核宿主细胞能够产生如上所述的重组多肽或者能够产生下文所述的 6-果糖- α -葡萄糖苷。

[0291] 6.2 真核表达

[0292] 本发明还设计了用于在真核宿主细胞中表达本发明的多核苷酸、片段或变异体的嵌合核苷酸构成物。许多真核宿主表达载体系统都可以被用于这方面。这些表达载体系统包括但不限于被重组酵母表达载体转化的酵母; 被重组病毒表达载体（例如杆状病毒）感染的昆虫细胞系统; 或被重组病毒表达载体（例如反录病毒、腺病毒、痘苗病毒）感染的动物细胞系统, 或被改造用于稳定表达的被转化的动物细胞系统。优选地, 所述嵌合核酸构建体被设计用于下文所述植物的遗传转化。

[0293] 6.3 植物表达

[0294] 在一个优选的实施方案中, 将本发明的多核苷酸、片段或变异体融合到启动子序列和 3' 非翻译序列上从而制备成一种被设计用于植物的遗传转化的嵌合 DNA 构建体。

[0295] 6.3.1 植物启动子

[0296] 本发明设计的启动子序列可以是待被转化宿主植物本身固有的或可以得自于其它来源, 其中所述区域在所述宿主植物中是功能性的。其它来源包括土壤杆菌 T-DNA 基因, 诸如用于胭脂氨酸、章鱼氨酸、甘露氨酸或其它冠瘿氨基酸生物合成的启动子; 源自植物的启动子, 诸如遍在蛋白质启动子; 组织特异性启动子（例如参见授予 Conkling 等人的美国专利 US 5, 459, 252; Advanced Technologies 的 W091/13992）; 源自病毒（包括宿主特异性病毒）的启动子, 或者部分或全部合成的启动子。在单子叶或双子叶植物中具有功能的众多启动子在本技术领域是熟知的（例如参见 Greve, 1983, 分子应用遗传学杂志 (J. Mol. Appl. Genet). 1 :499-511; Salomon 等, 1984, EMBO J. 3 :141-146; Garfinkel 等, 1983, 细胞 (Cell) 27 :143-153; Barker 等, 1983, 植物分子生物学 (Plant Mol. Biol). 2 :235-350）; 包括从植物中分离出来的各种启动子（诸如源自玉米 ubi-1 基因的 Ubi 启动子, Christensen 和 Quail, 1996）（例如参见美国专利 US 4, 962, 028）和从病毒中分离出来的各种启动子（诸如花椰菜花叶病毒启动子, CaMV 35S）。

[0297] 启动子序列可以包括调节转录的区域, 其中所述调节包括例如化学或物理抑制或诱导（例如基于代谢物、光或其它物理化学因素的调节; 例如参见其中公开了线虫效应启动子的 WO 93/06710）或基于细胞分化的调节（诸如与植物中的叶、根、种子等有关; 例如参见其中公开了根特异性启动子的美国专利 US 5, 459, 252）。因此, 启动子区域或这个区域的调节部分获自于被如此调节的合适基因。例如, 核酮糖-1,5-二磷酸羧化酶基因是光诱导型的并且可以被用于转录的起始。其它由应力、温度、创伤、病原体作用等诱导的基因是已知的。

[0298] 用于在培养的细胞中表达的优选启动子是一种强组成型启动子, 或者是一种

对特异性诱导物产生应答的启动子 (Gatz 和 Lenk, 1998, 植物科学趋势 (Trends Plant Science) 3 :352-8)。用于在完整植物中表达的优选启动子是一种在蔗糖贮存组织中被表达的启动子 (诸如成熟的甘蔗茎和甜菜的块茎), 或者是一种在收获前的后期使蔗糖转化为 6-果糖- α -葡萄糖苷同时对植物的其它生长和发育过程破坏作用最小的诱导型启动子。

[0299] 6.3.23' 非翻译区域

[0300] 本发明的嵌合基因构建体可能含有 3' 非翻译序列。3' 非翻译序列是指包括含有多腺苷酸化信号和其它任何能够实现 mRNA 加工或基因表达的调节信号的 DNA 片段的基因的一部分。所述多腺苷酸化信号的特征在于实现多腺苷酸序列向 mRNA 前体的 3' 端的插入。尽管变异不是罕见的, 但是多腺苷酸化信号通常因与规范形式 5' AATAAA-3' 具有同源性而被识别。

[0301] 3' 非翻译调节 DNA 序列优选地包括约 50 至 1000 个核苷酸碱基对而且除了含有多腺苷酸化信号和其它任何能够实现 mRNA 加工和基因表达的调节信号外还可以含有植物转录和翻译终止序列。合适的 3' 非翻译序列的实例是 3' 转录非翻译区, 该 3' 转录非翻译区含有一个源自根癌土壤杆菌的胭脂氨酸合酶 (nos) 基因的多腺苷酸化信号 (Bevan 等, 1983, 核酸研究 (Nucl. Acid Res), 11 :369) 和源自根癌土壤杆菌的章鱼氨酸合酶基因的 T7 转录物的终止子。或者, 尽管也可以采用本技术领域技术人员已知的其它 3' 元件, 但合适的 3' 非翻译序列可以得自于植物基因如源自马铃薯或番茄的蛋白酶抑制剂 I 或 II 基因的 3' 端, 大豆贮存蛋白基因的 3' 端和核酮糖-1,5-二磷酸羧化酶 (ssRUBISCO) 基因的豌豆 E9 小亚基的 3' 端。或者, 3' 非翻译调节序列可以按照例如由 An 记载的方法 (1987, 酶学方法 (Methods in Enzymology), 153 :292) 来从头获得, 其内容被本文引入作为参考。

[0302] 6.3.3 任选序列

[0303] 根据需要, 本发明的嵌合 DNA 构建体还可包括增强子, 即翻译或转录增强子。这些增强子区域对于本技术领域技术人员来说是已知的, 并且可以含有 ATG 起始密码子和相邻序列。所述起始密码子必须与涉及外源或内源 DNA 序列的编码序列的读框同相从而确保整个序列的翻译。翻译控制信号和起始密码子可以是多种来源的, 既可以是天然的也可以是合成的。翻译起始区可由转录起始区的来源来提供, 或者由外源或内源 DNA 序列来提供。所述序列还可以得自于被选用来驱动转录的启动子的来源, 并且这些序列可以被特异性修饰以便增强 mRNA 的翻译。

[0304] 转录增强子的实例包括但不限于源于例如由 Last 等人记载的 CaMV 35S 启动子和章鱼氨酸合酶基因 (美国专利 US 5, 290, 924, 其被引入本文作为参考) 的元件。提出的建议是当在植物转化的前后应用时, 使用增强子元件如 ocs 元件, 以及尤其是多拷贝元件, 将产生使由相邻启动子转录的水平提高的效果。或者, 可以使用源自烟草花叶病毒 (Gallie 等人, 1987) 的外壳蛋白基因的 ω 序列来增强由本发明的多核苷酸转录的 mRNA 的翻译。

[0305] 当插入在转录起始位点和编码序列的起始端, 即未被翻译的前导序列之间的 DNA 序列能够影响基因表达时, 还可以采用一个特定的前导序列。优选的前导序列包括那些含有被选用于指导外源或内源 DNA 序列进行最佳表达的序列的前导序列。例如, 这类前导序列包括一个优选的能够增加或保持 mRNA 稳定性和防止不恰当的翻译起始的共有序列, 例如由 Joshi (1987, 核酸研究 (Nucl. Acid Res), 15 :6643) 记载的共有序列, 其内容被引入本文作为参考。然而, 其它前导序列, 例如 RTBV 的前导序列具有高度的二级结构, 该结构预计

能够降低 mRNA 的稳定性和 / 或降低 mRNA 的翻译水平。因此, (i) 不具有高度的二级结构, (ii) 具有高度的二级结构, 其中该二级结构不抑制 mRNA 的稳定性和 / 或降低翻译水平, 或 (iii) 得自于在植物中被高表达的基因的前导序列将是最优选的。

[0306] 如果需要的话, 也可以包括调节元件如蔗糖合酶内含子, 例如由 Vasil 等人 (1989, 植物生理学 (Plant Physiol), 91 :5175) 记载的蔗糖合酶内含子, Adh 内含子 I, 例如由 Callis 等人 (1987, 基因研究 (Genes Develop), II) 记载的 Adh 内含子 I, 或 TMV ω 元件, 例如由 Gallie 等人 (1989, 植物细胞 (The Plant Cell), 1 :301) 记载的 TMV ω 元件。被用于实施本发明的其它这类调节元件对于本技术领域的技术人员来说是已知的。

[0307] 另外, 靶向序列可以被用来将外源或内源 DNA 序列的蛋白产物引导到植物细胞的胞内区室或胞外环境中。例如, 编码转运肽或信号肽序列的 DNA 序列可以被可操作性地连接到编码所需蛋白的序列上, 以致于当被翻译时, 所述转运肽或信号肽能够将所述蛋白转运到特定的胞内或胞外目的地, 然后所述转运肽或信号肽通过翻译后的加工被除去。转运肽或信号肽通过促进蛋白质转运过胞内膜而起作用, 所述胞内膜例如是内质网膜、液泡膜、泡囊膜、质体膜、线粒体膜和质膜。例如, 所述靶向序列可以将所需蛋白定向到特定的细胞器如液泡或质体 (例如叶绿体) 中, 而不是定向到胞质中。因此, 所述嵌合 DNA 构建体可进一步含有编码质体转运肽的 DNA 序列, 该 DNA 序列可操作性地连接在启动子区或本发明的启动子变异体和外源或内源 DNA 序列之间。例如, 可以参考 Heijne 等人 (1989, 欧洲生物化学杂志 (Eur. J. Biochem), 180 :535) 和 Keegstra 等人 (1989, 植物生理学, 植物分子生物学年鉴 (Ann. Rev. Plant Physiol. Plant Mol. Biol), 40 :471) 的文章, 其内容被引入本文作为参考。

[0308] 嵌合 DNA 构建体也可以被引入载体, 如质粒中。质粒载体包括有利于表达盒在真核和真核细胞中进行筛选、扩增和转化的附加 DNA 序列, 例如 pUC 来源的载体, pSK 来源的载体, pGEM 来源的载体, pSP 来源的载体, 或 pBS 来源的载体。附加 DNA 序列包括供载体的自主复制用的复制起点, 选择性标记基因, 优选地是编码抗生素或除草剂抗性的选择性标记基因, 单一多克隆位点和增强原核和真核细胞转化的序列, 其中所述单一多克隆位点为在嵌合 DNA 构建体中插入被编码的 DNA 序列或基因提供了多个位点。

[0309] 载体优选地含有一个能使载体稳定整合到宿主细胞基因组中或者能使载体在细胞中独立于该细胞基因组进行自主复制的元件。当被引入宿主细胞中时, 所述载体可以被整合到宿主细胞的基因组中。所述载体可以依靠其中所存在的外源或内源 DNA 序列或通过同源重组将载体稳定整合到基因组的该载体的任何其它元件来进行整合。或者, 载体可以含有用于指导通过同源重组而整合到宿主细胞基因组中的附加核酸序列。该附加核酸序列使载体能以在染色体中的精确位置整合到宿主细胞基因组中。为了提高在精确位置上进行整合的可能性, 整合元件应当优选地含有足够数目的核酸, 诸如 100 至 1500 个碱基对, 优选地 400 至 1500 个碱基对, 以及最优选地 800 至 1500 个碱基对, 所述核酸与相应的靶序列高度同源从而增强同源重组的可能性。整合元件可以是任何与宿主细胞基因组中的靶序列同源的序列。而且, 所述整合元件可以是非编码核酸序列或编码核酸序列。

[0310] 为了进行克隆和亚克隆, 所述载体可以进一步含有能够使载体在宿主细胞如细菌细胞中进行自主复制的复制起点。细菌的复制起点的实例是能够在大肠杆菌中进行复制的质粒 pBR322, pUC19, pACYC177 和 pACYC184 的复制起点和能够在芽孢杆菌中进行复制的

pUB110, pE194, pTA1060 和 pAM β 1 的复制起点。复制起点可以是具有一个突变的复制起点从而使其在芽孢杆菌细胞中的功能是温度敏感性的(例如参见 Ehrlich, 1978, 美国国家科学院院报 (Proc. Natl. Acad. Sci. USA) 75 :1433)。

[0311] 6.3.4 标记基因

[0312] 为了有助于转化体的鉴定, 嵌合 DNA 构建体理想地、或者除了本发明的多核苷酸序列外, 含有一个可选择的或可筛选的选标记基因。标记的实际选择并不关键, 只要该标记与被选植物细胞结合具有功能(即选择性的)就行。由于在植物转化过程中例如美国专利 US4, 399, 216 中所记载的未被连接的基因的共转化也是一种有效的方法, 所以标记基因和所需的外源或内源 DNA 序列不是必须被连接的。

[0313] 术语可选择的或可筛选的标记基因包括编码“可分泌的标记”的基因, 该可分泌的标记的分泌可以作为鉴定或选择转化细胞的工具被检测。实例包括编码可分泌的抗原或可分泌的酶的标记, 所述可分泌的抗原可通过抗体的相互作用来检测, 所述可分泌的酶可通过它们的催化活性来检测。可分泌的蛋白包括但不限于被插入或被截留在细胞壁中的蛋白(例如包括一个前导序列如在伸展蛋白或烟草 PR-S 的表达单元中发现的前导序列的蛋白); 可通过例如 ELISA 检测的可扩散的小分子蛋白; 和可在胞外溶液中检测的小分子活性酶(例如 α -淀粉酶, β -内酰胺酶, 膦丝菌素乙酰转移酶)。

[0314] 6.3.5 选择性标记

[0315] 细菌的选择性标记的实例是源自枯草芽孢杆菌或地衣形芽孢杆菌的 *dal* 基因, 或赋予了抗生素抗性如氨苄青霉素、卡那霉素、红霉素、氯霉素或四环素抗性的标记。用于选择植物转化体的选择性标记包括但不限于编码潮霉素 B 抗性的 *hyg* 基因; 例如由 Potrykus 等人 (1985, 分子基因遗传学 (Mol. Gen. Genet). 199 :183) 记载的赋予对卡那霉素、巴龙霉素、G418 等的抗性的新霉素磷酸转移酶 (*neo*) 基因; 例如被记载在 EP-A 256 223 中的源自大鼠肝脏的赋予对由谷胱甘肽衍生的除草剂的抗性的谷胱甘肽-S-转移酶基因; 例如被记载在 WO 87/05327 中的通过超表达赋予对谷氨酰胺合成酶抑制剂如膦丝菌素的抗性的谷氨酰胺合成酶基因; 例如被记载在 EP-A 275 957 中的源自产绿色链霉素的赋予对选择剂膦丝菌素的抗性的乙酰转移酶基因; 例如由 Hinchee 等人 (1988, 生物技术 (Biotech), 6 :915) 记载的赋予对 N-膦酰基甲基甘氨酸的耐受性的编码 5-烯醇莽草酸 (*enolshikimate*)-3-磷酸合酶 (EPSPS) 的基因; 例如被记载在 WO91/02071 中的赋予抗 bialaphos 抗性的 *bar* 基因; 源自臭鼻克雷伯氏菌的赋予对溴草腈的抗性的腈酶基因; 如 *bxn* (Stalker 等, 1988, 科学 (Science), 242 :419); 赋予氨甲蝶呤抗性的二氢叶酸还原酶 (DHFR) 基因 (Thillet 等, 1988, 生物化学杂志 (J. Biol. Chem), 263 :12500); 突变的乙酰乳酸合酶基因 (ALS), 该基因能够赋予对咪唑啉酮的抗性、对磺酰脲的抗性或对其它 ALS-化学抑制剂的抗性 (EP-A-154204); 赋予对 5-甲基色氨酸的抗性的突变的邻氨基苯甲酸合酶基因; 或赋予对除草剂的抗性的茅草枯脱卤酶基因。

[0316] 6.3.6 可筛选的标记

[0317] 优选的可筛选的标记包括但不限于编码其多种显色底物是已知的 β -葡萄糖醛酸糖苷酶 (GUS) 的 *uidA* 基因; 编码其显色底物是已知的酶的 β -半乳糖苷酶; 水母发光蛋白基因 (Prasher 等, 1985, 生物化学, 生物物理学研究通讯 (Biochem. Biophys. Res. Comm), 126 :1259), 该基因可被用于钙敏感性生物发光检测中; 绿色荧光蛋白基因 (Niedz 等, 1995

植物细胞报道 (Plant Cell Reports), 14 :403) ; 荧光素酶 (luc) 基因 (Ow 等, 1986, 科学 (Science), 234 :856), 该基因可被用于生物发光检测 ; β -内酰胺酶基因 (Sutcliffe, 1978, 美国国家科学院院报 (Proc. Natl. Acad. Sci. USA) 75 :3737), 该基因编码其多种显色底物是已知的酶 (例如 PADAC, 产生头孢菌素) ; R- 基因座基因, 该基因编码一种对花色素苷色素在植物组织中的产生起调节作用的产物 (Dellaporta 等, 1988, 见染色体结构和功能 (inChromosome Structure and Function), 263-282 页) ; α -淀粉酶基因 (Ikuta 等, 1990, 生物技术 (Biotech), 8 :241) ; 酪氨酸酶基因 (Katz 等, 1983, 遗传微生物学杂志 (J. Gen. Microbiol), 129 :2703), 该基因编码一种能够将酪氨酸氧化成多巴和多巴醌的酶, 所述多巴醌继而缩合成可容易检测的化合物-黑色素 ; 或 xylE 基因 (Zukowsky 等, 1983, 美国国家科学院院报 (Proc. Natl. Acad. Sci. USA) 80 :1101), 该基因编码一种能够转化显色儿茶酚的儿茶酚双加氧酶。

[0318] 7. 嵌合构建体向植物细胞中的引入

[0319] 许多技术可以被用来将 DNA 引入植物宿主细胞中。已有许多植物转化技术是本技术领域技术人员所熟知的, 而且新技术也在不断为人们所知晓。转化技术的具体选择由该技术转化某些植物物种的效率以及采用选择的具体方法实施本发明的人员的经验和偏爱来决定。对于技术人员显而易见的是将嵌合 DNA 构建体引入植物细胞中的转化系统的具体选择对于本发明不是必需的或限制性的, 前提条件是该系统可达到核酸转移的可接受的水平。Birch (1997, 植物生理学, 植物分子生物学年鉴 (Annu. Rev. Plant Physiol. Plant Molec. Biol). 48 :297-326) 提供了用于植物改良的转化系统的实际应用指南。

[0320] 原则上, 能够进行转化的双子叶植物和单子叶植物都可以通过将本发明的嵌合 DNA 构建体引入受体细胞中并且对含有和表达本发明多核苷酸的新物进行培养来进行改良。

[0321] 利用根癌土壤杆菌的根瘤诱导 (Ti) 质粒的 T-DNA 在双子叶 (宽叶) 植物如烟草、马铃薯和苜蓿中引入和表达外源或嵌合 DNA 序列已经被证明是可能的 (参见例如, Umbeck, 美国专利 US 5, 004, 863, 和国际专利申请 PCT/US93/02480)。可以利用含有 Ti 质粒的根癌土壤杆菌将本发明的构建体引入到植物细胞中。在利用根癌土壤杆菌培养物作为转化载体中, 最有利的是利用土壤杆菌属的非致癌菌株作为载体以使转化组织进行正常的非致癌分化成为可能。优选地是所述土壤杆菌 含有一个双 Ti 质粒系统。这种双系统包含 (1) 含有毒性区的第一 Ti 质粒, 所述区域对于转移 DNA (T-DNA) 向植物中的引入是必要的, 和 (2) 一个嵌合质粒。所述嵌合质粒含有野生型 Ti 质粒的 T-DNA 区的至少一个边缘区, 所述边缘区位于待被转移核酸的侧翼。双 Ti 质粒系统已经被证明对于转化植物细胞是有效的, 例如参见由 De Framond (1983, 生物技术 (Biotechnology), 1 :262) 和 Hoekema 等人 (1983, 自然 (Nature), 303 :179) 所记载的内容。由于这种双系统不需要整合到土壤杆菌的 Ti 质粒中所以是特别优选的。

[0322] 涉及利用土壤杆菌的方法包括但不限于 : (a) 将土壤杆菌与培养分离的原生质体进行共培养 ; (b) 用土壤杆菌转化植物细胞或组织 ; 或 (c) 用土壤杆菌转化种子、顶端或分生组织。

[0323] 最近, 属于单子叶植物的稻米和玉米已证明也易被土壤杆菌所转化。然而, 利用土壤杆菌介导的转化方法还没有成功地转化包括燕麦、高粱、黍和黑麦在内的许多其它重要

的单子叶农作物。然而,将来 Ti 质粒可能被利用来作为这些其它单子叶植物的载体。另外,利用 Ti 质粒作为模型系统可能能够人工构建用于这些植物的转化载体。也可能通过人工的方法如微注射,或使单子叶植物的原生质体与含有 T- 区域的细菌的原生质球融合将 Ti 质粒引入单子叶植物中,然后将其整合到植物的核 DNA 中。

[0324] 另外,如 Bechtold 等人 (1993, C. R. Acad. Sci. Paris, 316 :1194) 所记载的,可以通过利用土壤杆菌的原位转化来实现基因的转移。该方法基于土壤杆菌细胞悬浮液的真空渗透。

[0325] 或者,可利用土壤杆菌的根诱导 (Ri) 质粒作为载体来引入嵌合构建体。

[0326] 花椰菜花叶病毒 (CaMV) 也可以被用作将外源核酸引入植物细胞的载体 (美国专利 US 4, 407, 956)。CaMV DNA 基因组被插入到亲本细菌质粒中从而生成一种能在细菌中增殖的重组 DNA 分子。克隆后,所述重组质粒可以再次被克隆并且通过引入所需的核酸序列而得到进一步的修饰。然后将重组质粒的被修饰病毒部分从亲本细菌质粒上切下来,而且被用于接种植物细胞或植物。

[0327] 例如由 Fromm 等人 (1985, 美国国家科学院院报 (Proc. Natl. Acad. Sci., U. S. A), 82 :5824) 和 Shimamoto 等人 (1989, 自然 (Nature) 338 :274-276) 所记载的,也可以通过电穿孔将嵌合核酸构建体引入到植物细胞中。在该项技术中,在含有相关核酸序列的载体或核酸存在的条件下对植物原生质体进行电穿孔。高场强度的电脉冲可逆性地穿透膜从而引入核酸。被电穿孔的植物原生质体重新形成细胞壁,然后进行分裂并且形成植物愈伤组织。

[0328] 用于将嵌合核酸构建体引入植物细胞的另一种方法是通过利用在小珠粒或颗粒基质内部或者其表面含有待引入核酸的小粒子进行的高速冲击穿透法 (也被称作粒子轰击或微粒轰击),例如参见由 Klein 等人 (1987, 自然 (Nature) 327 :70) 记载的内容。尽管通常只需要新核酸序列的单个引入方法,但是该方法特别提供了多个引入方法。

[0329] 或者,通过利用机械或化学手段与植物细胞接触可以将嵌合核酸构建体引入植物细胞中。例如,通过利用微量吸管进行微注射可以直接将核酸机械地转移到植物细胞中。或者,通过利用聚乙二醇可以将核酸转移到植物细胞中,所述聚乙二醇与被细胞接纳的遗传物质形成沉淀复合物。

[0330] 目前已有多种已知用于转化单子叶植物的方法。目前,用于转化单子叶植物的优选方法是外植体或悬浮细胞的微粒轰击,和 DNA 的直接摄取或电穿孔,例如参见由 Shimamoto 等人 (1989, 上文) 所记载的内容。通过微粒轰击将吸水链霉素 bar 基因引入玉米悬浮培养物的胚发生细胞已经获得了转基因玉米植物 (Gordon-Kamm, 1990, 植物细胞 (Plant Cell), 2 :603-618)。已经有关于将遗传物质引入其它单子叶农作物如小麦和大麦的糊粉原生质体中的报道 (Lee, 1989, 植物分子生物学 (Plant Mol. Biol) 13 :21-30)。通过只选择老化的致密和小结状的胚发生愈伤组织来建立胚发生悬浮培养物,已经由胚发生悬浮培养物再生出小麦植物 (Vasil, 1990, 生物技术 (Bio/Technol) 8 :429-434)。与这些农作物的转化系统联合使用能够将本发明应用到单子叶植物中。这些方法还可以被应用于双子叶植物的转化和再生。由例如 Bower 等人 (1996, 分子育种 (Molecular Breeding) 2 :239-249) 所记载的从胚发生愈伤组织已经再生出转基因甘蔗植物。

[0331] 或者,可以联合采用不同的技术来提高转化过程的效率,例如通过采用被土壤杆

菌包衣的微粒进行轰击 (EP-A-486234) 或通过微粒轰击来导致创伤,接着与土壤杆菌进行共培养 (EP-A-486233)。

[0332] 8. 分化的转基因植物的生产和定性

[0333] 8.1 再生

[0334] 被用来使被转化细胞再生成分化植物的方法对于本发明并不关键,任何适用于靶植物的方法都可以被采用。通常,在转化过程后,使植物细胞再生从而获得完整的植物。

[0335] 由原生质体进行的再生根据植物物种的不同而不同,但是,通常首先制备原生质体的悬浮液。对于某些物种,诱导胚芽形成的过程是由原生质体悬浮液到成熟期和萌发出天然的胚芽。培养基通常含有各种生长和再生所必需的氨基酸和激素。使用的激素的实例包括苗长素和细胞激动素。向培养基中添加谷氨酸和脯氨酸有时是有利的,尤其是针对玉米和苜蓿这类物种来说是有利的。有效的再生将取决于培养基、基因型和培养过程。如果这些变量得到了控制,那么再生是可重复的。还可以由植物的愈伤组织、外植体、器官或部分进行再生。转化可以在器官或植物部分再生的前后进行,例如参见在酶学方法 (Methods in Enzymology), Vol. 118 和 Klee 等人 (1987, 植物生理学年鉴 (Annual Review of Plant Physiology), 38 :467) 中记载的内容,该内容被引入本文作为参考。利用 Horsch 等人的叶片-转化-再生方法 (1985, 科学 (Science), 227 :1229, 引入本文作为参考), 在选择培养基上培养叶片, 大约 2-4 周内形成茎干。从愈伤组织上切下发育的茎干并将其移植到合适的根诱导选择培养基中。一旦出现了根, 尽可能快地将带根的小植株转种到土壤中。根据需要可以将小植株移栽到花盆中, 直到发育成熟。

[0336] 对于营养繁殖的农作物, 通过取插条或通过组织培养技术来对成熟的转基因植物进行繁殖从而生产出多株相同的植物。选择所需的转基因子从而获得新品种并对新品种进行营养繁殖以备商用。

[0337] 对于种子繁殖的农作物, 成熟的转基因植物可以自交产生纯合同系繁殖的植物。该同系繁殖的植物产生含有新引入的外源基因的种子。可以对这些种子进行培养从而长出能够产生被选表型, 例如开花早的植物来。

[0338] 由再生植物得到的部分如花、种子、叶、枝、果实等被包括在本发明中, 前提条件是这些部分含有所述被转化的细胞。再生植物的子代和变异体和突变体也被包括在本发明的范围内, 前提条件是这些部分含有被引入的核酸序列。

[0339] 应当理解所述文献记载了多种用于再生特定植物类型的技术而且更多的技术不断为人所知。本技术领域的普通技术人员可以参考该文献以便得到更加详细的内容而且可以在不需要进行过多实验的情况下选择合适的技术。

[0340] 8.2 定性

[0341] 为了证实再生植物中含有本发明的多核苷酸, 可以进行多种测定。这类测定包括, 例如本技术领域的技术人员熟知的“分子生物学”测定, 诸如 DNA 和 RNA 印迹法和 PCR; 可以按照例如本文记载的方法来测定由本发明的多核苷酸表达的蛋白质的蔗糖异构酶活性。

[0342] 9.6- 果糖 - α - 葡萄糖苷的生产

[0343] 本发明进一步涉及利用本文记载的多核苷酸或多肽序列或利用其变异体或片段生产 6- 果糖 - α - 葡萄糖苷的方法。该方法包括使蔗糖或含蔗糖的培养基或底物与至少一种选自 (a) 一种被编码具有蔗糖异构酶活性的 DNA 序列转化的生物, 例如遗传修饰的细菌或

植物；(b) 这种细胞或生物的胞外产物或细胞提取物；和 (c) 具有蔗糖异构酶活性的分离形式的蛋白质的成员在蔗糖被蔗糖异构酶至少部分转化为 6-果糖- α -葡萄糖苷的条件下进行接触。接着，从所述培养基或生物中得到 6-果糖- α -葡萄糖苷并按照本技术领域中的已知方法进行纯化。用于工业生产 6-果糖- α -葡萄糖苷的方法，例如利用与含蔗糖的培养基接触的固定化细胞或蔗糖异构酶生产 6-果糖- α -葡萄糖苷的方法是已知的 (Cheetham 等, 1985, 生物技术, 生物工程 (Biotech. Bioeng) 27 :471-481 ;Takazoe, 1989, 巴糖-蔗糖的异构替代品。见甜味剂进展 (Progress in Sweeteners) (Grenby, T. H., 编辑) Barking :Elsevier, 143-167 页；以及各自其中的参考文献)。本发明通过提供新型的蔗糖异构酶而对这些方法进行改进，改进方法具有包括能够高效生产 6-果糖- α -葡萄糖苷在内的有益特性。

[0344] 而且，本发明首次揭示了在植物中直接生产 6-果糖- α -葡萄糖苷的能力。在植物中直接生产 6-果糖- α -葡萄糖苷是非常有利的，因为这样可以节省从植物中提取蔗糖并以该此作为底物，由其它生物、提取物或分离酶通过工业发酵转化为 6-果糖- α -葡萄糖苷所需要的费用。相反，由本文所记载的经遗传修饰的植物中的光合作用而产生的蔗糖通过植物组织中的蔗糖异构酶活性被转化为 6-果糖- α -葡萄糖苷。然后利用现有的用于从植物中收获其它糖，尤其是蔗糖的方法收获所产生的 6-果糖- α -葡萄糖苷。首先收获贮存有 6-果糖- α -葡萄糖苷的植物材料，然后挤压出含有 6-果糖- α -葡萄糖苷的汁液和 / 或流经扩散装置以便从不溶性植物材料中提取可溶性 6-果糖- α -葡萄糖苷。然后通过去除杂质的处理方法对 6-果糖- α -葡萄糖苷进行纯化并通过本技术领域技术人员熟知的蒸发和结晶步骤进行浓缩 (Cooke 和 Scott, 1993, 甜菜作物 (The Sugar Beet Crop) :科学实践 (science intopractice), London :Chapman & Hall ;Meade, 1977, 蔗糖手册 (Cane Sugar Handbook), 纽约 :Wiley, 以及各自其中的参考文献)。

具体实施方式

[0345] 为了能容易地理解本发明并付诸实践，通过下列非限制性实施例的方式来描述特别优选的实施方案。

[0346] 实施例

[0347] 实施例 1

[0348] 基于由 Mattes 等人指定的区域利用寡核苷酸引物对编码蔗糖异构酶的多核苷酸进行分离

[0349] 采用该设计方案对表达蔗糖异构酶的已知细菌 (登记号为 WAC2928 的大黄欧文氏菌)，和其它 30 个独立的细菌分离物进行了试验。简并 PCR 引物是基于由 Mattes 等人 (上文) 指定为保守区的区域进行设计的，所述保守区是 Mattes 等人对他们已知的蔗糖异构酶基因进行分析后被指定的。

[0350] 正向引物由从 SEQ ID NO :1 的第 139 位核苷酸延伸到第 155 位核苷酸的序列组成 :5' -tgg tgg aa(a, g) ga(g, a) gct gt-3' [SEQ ID NO :38]。

[0351] 反向引物由从 SEQ ID NO :1 的第 625 位核苷酸延伸到第 644 位核苷酸的序列组成 :5' -tcc cag tta g(g, a)t ccg gct g-3' [SEQ ID NO :39]。

[0352] 细菌基因组 DNA 被用作 PCR 的模板。按照 Ausubel 等人的方法 (1989, 上文) 提取基因组 DNA。以含 100ng DNA, 5 μ L 的 10 X PCR 缓冲液 (Promega), 2 μ L dNTPs (每种

NTP5mM), 正向引物和反向引物各 250ng, Taq 聚合酶 1 μ L(Promega) 的 50 μ L 终体积进行 PCR 反应。通过利用三种不同的退火温度: 46 $^{\circ}$ C, 50 $^{\circ}$ C 或 53 $^{\circ}$ C 平行进行了三次 PCR。首先在 94 $^{\circ}$ C 下反应 1 分钟, 然后进行 35 个循环, 每个循环包括在 94 $^{\circ}$ C 下反应 1 分钟, 在退火温度下反应 1 分钟并在 72 $^{\circ}$ C 下反应 1 分钟。

[0353] PCR 产物在 1% 的琼脂糖凝胶上跑电泳后, 回收大小范围在 0.3 至 1.0kb 的带并利用 TOPOTMTA Cloning[®] 试剂盒 (Invitrogen) 和按照试剂盒所附的说明书将所收集的带克隆到 pCR[®] 2.1 载体中。在澳大利亚基因组研究实验室 (Australian Genomic Research Facility), 利用 ABI PRISM 大染料终止子循环测序快速反应试剂盒 (ABI PRISM Big Dye Terminator Cycle Sequencing Ready Reaction Kit), 采用可在载体上得到的通用 M13 反向引物或 M13 正向引物对质粒插入序列进行测序。以被测序的 DNA 作为查询对象, 利用通过 ANGIS 的 FASTA 程序对 GenBank 数据库进行检索。

[0354] 利用源自由 Mattes 等人 (上文) 指定的 ‘保守区’ 的引物, 由大黄欧文氏菌而且也由后来发现对蔗糖异构酶活性呈阴性的细菌扩增了 PCR 产物。由琼脂糖凝胶电泳显示的 PCR 产物的图形包括: 2 个分离物没有带, 来自 3 个分离物的一条带, 和来自包括大黄欧文氏菌在内的其它所有细菌的多条带。对包括由大黄欧文氏菌扩增的 6 条带在内的 12 条带中的 DNA 进行了克隆和测序。被测序的带中没有一条带显示与蔗糖异构酶具有明显的同源性, 包括由 Mattes 等人 (上文) 记载的源自大黄欧文氏菌的基因区域。大多数被测序的带显示出与已知的葡萄糖苷酶基因的相似性。

[0355] 因此, 结论是由 Mattes 等人指定的保守序列不是蔗糖异构酶所特有的, 而是包括葡萄糖苷酶在内的其它类酶所共有的。因此, 这些保守序列不能在不采用 PCR 条件进行繁重的实验的情况下直接用于蔗糖异构酶的克隆和借助其它识别异构酶克隆的手段进行筛选。

[0356] 实施例 2

[0357] 对使蔗糖转化为 6-果糖- α -葡糖苷的细菌进行的功能性筛选

[0358] 细菌的收集和分离

[0359] 从一定范围的被选环境地点收集细菌样品, 所述被选环境地点具有产生新型的代谢蔗糖的细菌的潜力。特别是, 选择能够定期获得蔗糖的地点, 这样的地点可能有利于能使蔗糖转化为贮存异构体如 6-果糖- α -葡糖苷的生物。来自昆士兰东南部地点的约 100 个样品被收集到 MIM 液体培养基中。MIM 是 0.2% 6-果糖- α -葡糖苷 (6-0- α -D-吡喃葡萄糖基-D-呋喃果糖) 加上 MM (含有 0.5% Na_2PO_4 , 0.45% KH_2PO_4 , 0.1% NH_4Cl , 0.05% $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 0.005% 柠檬酸铁铵和 0.0005% CaCl_2 的基本培养基)。于室温下以 200rpm 的转速在定轨摇床上培养 2 小时后, 将 100 μ L 的样品划线到 MSM (MM 加上 4% 蔗糖) 琼脂平板上并在 28 $^{\circ}$ C 下培养过夜。经过这两步富集后, 将形态不同的菌落分离到单独的 LB 或 MSM 新平板上进行进一步培养 (总共 578 个菌落)。在划线培养以确保单菌落分离物的纯度后, 将这些分离物一式两份转移到复制斑点平板上和含有 5mL SLB (含有 4% 蔗糖的 LB) 的 30mL 普通试管中以便在分析测定中进一步进行功能性筛选, 所述分析测定优先显示具有较高 6-果糖- α -葡糖苷生产能力的生物。

[0360] 实施例 3

[0361] 用于苯胺 / 二苯胺测定的样品的制备

[0362] 在室温下以 $10000 \times g$ 的速度对在 5mL SLB 中培养过夜的培养物进行离心。小心倒掉上清液并加入 2mL 50% 的蔗糖在柠檬酸盐 / 磷酸盐缓冲液 (pH6) 中的溶液。轻轻地将细胞再悬浮起来并于 28°C 下在摇床中培养 48 小时。培养后, 将 1.5mL 的培养物转移到新 Eppendorf 管中, 在 100°C 下煮沸 15 分钟并在室温下以 $16000 \times g$ 的速率离心 20 分钟。不要碰离心沉淀物, 将上清液保存在新管中以备苯胺 / 二苯胺测定和毛细管电泳用。

[0363] 实施例 4

[0364] 苯胺 / 二苯胺测定

[0365] 将样品均匀地点在 Whatman#1 滤纸的外周边缘, 同时将阳性对照 (源自大黄欧文氏菌) 和阴性对照 (源自大肠杆菌) 置于滤纸的中心点。样品被点样到滤纸上后, 在进行 15 分钟干燥的同时制备显色试剂。

[0366] 按照以下方式制备所述试剂:

[0367] a. 用 A. R. 丙酮将 4mL 的苯胺配成 100mL;

[0368] b. 用 A. R. 丙酮将 4g 二苯胺配成 100mL;

[0369] c. 20mL 85% 的正磷酸。

[0370] 在通风橱中分别制备组分 (a) 和 (b) 以确保苯胺 / 二苯胺被充分混合 / 溶解在丙酮中, 然后在玻璃烧杯中将组分 (a) 和 (b) 合并, 接着加入所述的酸。刚加入酸后, 形成混浊的白色沉淀, 通过剧烈涡动搅拌后白色沉淀溶解从而产生一种澄清的褐色溶液。

[0371] 使制备好的滤纸透过“显色剂”, 确保每张滤纸接受到均匀和等同的显色剂。然后在通风橱中使滤纸在纸巾上干燥 15 分钟, 然后在 80°C 的烘箱中加热 10 分钟。记录结果 (斑点的颜色) 或用数码相机拍下结果。

[0372] 如果存在 6- 果糖 - α - 葡萄糖苷, 由于存在 1,6 键连接的葡萄糖, 所述反应则产生黄色至黄褐色的斑点; 而由于存在 1,2 连接键, 葡萄糖则产生深灰色的斑点, 果糖产生银灰色的斑点, 以及蔗糖产生紫色 - 褐色斑点。颜色的深浅取决于含糖浓度。通过苯胺 / 二苯胺测定试验的结果显示从 578 个菌落中选出 12 个候选菌落。然后, 为了分辨和鉴定相关的代谢物, 通过利用毛细管电泳的定量分析证实了被选分离物中含有 6- 果糖 - α - 葡萄糖苷产物。

[0373] 实施例 5

[0374] 用于毛细管电泳的样品的制备

[0375] 为了进行进一步的分析, 在向毛细管中加样前需要除去用于苯胺 / 二苯胺测定的上清液中的离子型物质。通过流经购自 Varian 的强阳离子交换柱 (Bond Elut-SCX, 1210-2013) 和强阴离子交换柱 (BondElut-SAX, 1210-2017) 来去除所述的离子型物质。按照以下方式对所述柱子进行预处理: 用一个体积的甲醇冲洗, 接着用一个体积的水冲洗, 其中借助注射器迫使清洗液通过柱子。

[0376] 在首先通过 SCX 柱然后通过 SAX 柱进行处理之前, 用无菌 Milli-Q (SMQ) 水将细菌上清液稀释 150 倍。将 1mL 的被稀释的上清液加载到 SCX 柱上。借助 50mL 的注射器迫使样品通过柱子。将洗脱液直接收集进 SAX 柱子中。采用收集在 1.5mL 的 Eppendorf 管中的最终洗脱液以类似的方式迫使样品通过柱子。

[0377] 实施例 6

[0378] 毛细管电泳

[0379] 利用 Beckman P/ACE 5000 系列 C. E. 系统进行高效毛细管电泳 (HPCE) 的分离, 其

中所述系统采用了由所配置的氙灯发出的 190 至 380nm 的光源和用于样品检测的 Beckman P/ACE 紫外吸收检测器 (254nm[±10nm] 滤光轮)。

[0380] 毛细管是裸露的、熔凝硅石毛细管, I. D. 50 μ m, O. D. 363 μ m (Supelco 目录号 70550-U)。毛细管总长度是 77cm, 进样口至检测器窗口的长度是 69cm。毛细管的检测器窗口通过用火柴点燃烧掉毛细管上的涂层, 并用甲醇擦拭来制备。

[0381] 为了使迁移时间达到最大程度的再现, 每天早晚按照下述冲洗程序对毛细管进行预处理: 用 SMQ 冲洗 2 分钟, 用 0.1M HCl 冲洗 10 分钟, 用 SMQ 冲洗 2 分钟, 用 0.1M NaOH 冲洗 10 分钟, 用 SMQ 冲洗 2 分钟, 用 0.5M 氨水冲洗 15 分钟以及用 SMQ 冲洗 2 分钟。所有溶液都用 SMQ 来溶解 / 稀释并通过 0.45 μ m 微孔过滤器进行过滤。

[0382] 能够基于紫外吸收进行直接检测的碱性硫酸铜电解液除了被用来分辨和检测包括预计存在于细胞提取物中的葡萄糖和果糖在内的其它糖外, 还被用来分辨和检测低浓度的蔗糖及其异构体 6-果糖- α -葡萄糖苷。利用由 6mM 硫酸铜 (II) 和 500mM 氨水组成的, pH 为 11.6 的电解液, 在中性糖与铜离子 (II) 于碱性条件下发生螯合反应的基础上实现中性糖的分离和直接 UV 检测。

[0383] 电解质缓冲液 (EB) 在每天的开始新配制并在使用前脱气 15 分钟。达到要求后, 用 EB 将毛细管冲洗 15 分钟。在样品分离之间也采用 EB 将毛细管冲洗 10 分钟。表 1 中列出了分批操作的程序参数。每个样品中都包含了上述的阳性对照和阴性对照。另外, 在第一个样品之前面和最后一个样品之后点上标准品 (由蔗糖和 6-果糖- α -葡萄糖苷组成), 以便使由诸如 EB 的耗尽、毛细管加热等因素导致的迁移时间的差异能够得到测量和校正。

[0384] 表 I 用于毛细管电泳分批操作的参数

[0385]

功能	持续时间	进样小瓶	取样小瓶	备注
EB 冲洗	5 分钟	11	10	正向, 20psi
加压注射	5 秒钟	样品小瓶	10	正向, 20psi
分离	30 分钟	12、	1	25KV, 254nm
EB 冲洗	5 分钟	13	10	正向, 20psi

[0386] 被命名为 349J, 14s 和 68J 的三种分离物被证实具有将蔗糖转化为 6-果糖- α -葡萄糖苷的能力。分别掺入了 5Mm 蔗糖, 0.5mM 6-果糖- α -葡萄糖苷, 0.5mM 果糖或 0.5mM 葡萄糖后, 对由所述三种阳性分离物得到的稀释上清液进行再测试以便基于与已知糖的共迁移来检验样品中的峰的特性。

[0387] 实施例 7

[0388] 细菌基因组文库的构建

[0389] 粘粒载体 SuperCos 1 (Stratagene) 被用于构建大黄欧文氏菌 (登记号 WAC2928) 的澳大利亚分离物和细菌分离物 14S, 68J 和 349J 的基因组文库。所述载体适于容纳 30 至 45kb 的基因组 DNA 片段。

[0390] 实施例 8

[0391] 基因组 DNA 插入片段的制备

[0392] 由于需要大片段用于在 SuperCos 1 载体中进行克隆, 所以基本采用 Priefer 等人 (1984, 用粘粒进行克隆 (Cloning with cosmids), 见现代分子遗传学 (Advanced Molecular Genetics) (Pühler, A. 和 Timmis, K. N., 编辑) 柏林: Springer-Verlag, 190-201 页) 的方

法来提取基因组 DNA 从而在消化前获得高分子量 (~ 150kb) 的 DNA。在 TE 缓冲液中将钩状的 DNA 于 65℃ 下溶解 3 小时或在 4℃ 下溶解 2 天, 该过程中不进行振荡。利用 0.4% 的琼脂糖凝胶检查来估计分子的大小。为了克隆到 SuperCos 1 载体的 BamH I 位点中, 利用限制内切核酸酶 Sau 3A 对染色体 DNA 进行部分消化。进行一系列的部分消化试验以便确定用于获得所需插入大小范围的理想条件。利用 1X Sau 3A 缓冲液将 10 μg 的基因组 DNA 配制成 135 μL 体积的反应物后在 37℃ 下预平衡 5 分钟。然后加入 0.5 个单位的 Sau 3A, 并且在 0, 5, 10, 15, 20, 25, 30, 40 分钟后, 取出等分样品 (15 μL) 并立即在 68℃ 保持 20 分钟来终止反应。将所述等分样品加到 0.5% 的琼脂糖凝胶上进行电泳。确定了片段平均大小为 50kb 的最佳消化时间。将反应规模扩大到在 675 μL 的总体积中的 50 μg 的基因组 DNA。消化后, 向样品中加入 13 μL 的 0.5M EDTA, pH 8.0。进行苯酚 / 氯仿萃取后, 按照 Sambrook 等人 (1989) 的方法通过加入 1/10 体积的醋酸钠 (3M, pH 5.2) 和 2.5 体积的乙醇使 DNA 沉淀。将沉淀物重新悬浮在 450 μL 的 1X CIAP 缓冲液中并在 37℃ 下对 DNA 进行 60 分钟的 CIAP 处理。对经 CIAP 处理过的 DNA 再次进行苯酚 / 氯仿萃取。最后将 DNA 溶解在 30 μL 的 TE 缓冲液中以备连接反应应用。

[0393] 实施例 9

[0394] 载体 DNA 的制备

[0395] 在 37℃ 下利用 Xba I 将 20 μg 的 SuperCos 1 载体消化 3 小时后, 向反应物中按每 μg DNA 加入一个单位的 CIAP 并在 37℃ 下再温育一个小时。利用上述方法对被处理的 DNA 进行苯酚 / 氯仿萃取和乙醇沉淀。经 Xba I/CIAP 处理的 SuperCos 1 DNA 被重悬在 TE 缓冲液中并利用 0.8% 的琼脂糖凝胶进行检查从而观察到大小为 7.6kb 的单一的线性带。载体 DNA 采用 BamH I 进一步进行消化, 采用苯酚 / 氯仿进行萃取, 并进行乙醇沉淀, 然后以 1 μg/μL 的比例被重悬在 TE 缓冲液中以备连接反应应用。

[0396] 实施例 10

[0397] DNA 的连接和包装

[0398] 于 70℃ 下, 在 15 μL 的体积中, 将 2.5 μg 的 Sau 3A 部分消化的细菌基因组 DNA 和经 Xba I/CIAP/BamH I 处理的 1.0 μg SuperCos 1 载体 DNA 加热 5 分钟。然后加入 2 μL 10mM ATP, 2 μL 10X 连接缓冲液和 1 μL T4 DNA 连接酶 (Invitrogen) 从而使达到 20 μL 的总体积。在室温下温育 4 小时后, 使连接反应在 4℃ 下维持过夜。通过 2 μL 反应液相对载体和插 DNA 的未连接混合物在 0.8% 的琼脂糖凝胶上跑电泳来观测连接效率。

[0399] 按照厂商提供的说明书 (Gigapack III Gold Packaging Extract, Stratagene) 对四分之一的连接物进行体外包装。

[0400] 在 37℃ 下, 在含有 0.2% 麦芽糖和 10mM MgSO₄ 的 LB 培养基中对大肠杆菌 NM554 (Stratagene) 的宿主细胞进行振荡培养使元从单个菌落生长到 1.0 的 OD₆₀₀ 值。在 4℃ 下以 2000xg 的速度离心 10 分钟来收获细胞, 然后将培养细胞轻轻地再悬浮在 10mM MgSO₄ 的溶液中以达到 0.5 的 OD₆₀₀ 值。将 10 μL 被包装的粘粒文库与 50 μL 的 NM554 细胞在 1.5mL 的试管中混合后, 在室温下将其培养 30 分钟, 然后向试管中加 400 μL 的 LB。为了使抗生素抗性得到表达, 在 37℃ 下将细胞再培养一个小时, 同时每 15 分钟轻轻摇动一次。将细胞离心 30 秒钟后轻轻地重悬在 100 μL 的新鲜 LB 肉汤中。将 50 μL 涂布到含有 50 μg/mL 氨苄青霉素的 LB 平板上。

[0401] 实施例 11

[0402] 粘粒文库的功能性筛选

[0403] 对四个粘粒文库的每个文库中的 600 个菌落进行功能性筛选,如上所述进行苯胺 / 二苯胺测定和 CE 后,结果表明源自大黄欧文氏菌的 4 个克隆,源自 14S 的 4 个克隆,源自 349J 的 3 个克隆和源自 68J 的 3 个克隆具有将蔗糖转化为 6- 果糖 - α - 葡糖苷的能力。

[0404] 实施例 12

[0405] 亚克隆和测序

[0406] 按照 Sambrook 等人 (1989) 的方法制备源自阳性菌落的粘粒 DNA。为了找到含有蔗糖异构酶的最小功能性片段,通过分别利用 EcoR I, BamH I 或 Hind III 进行部分消化来制备粘粒 DNA 的亚克隆插入片段。将刚被 EcoR I, BamH I 或 Hind III 消化的 pZerOTM-2 载体 (Invitrogen) 用于与插入片段的连接反应中。全部克隆过程如连接和向 Top 10 大肠杆菌菌株种的转化都按照由 Invitrogen 提供的说明书进行。将每种连接物的 200 个转化体挑选出来,进行涂布和培养以便按照上述的苯胺 / 二苯胺测定方法进行功能性筛选。通过 CE 分析进一步证实功能为阳性的亚克隆。从经 CE 证实为阳性的亚克隆中分离质粒 DNA 以便检验 EcoR I, BamH I 或 Hind III 的消化模式。源自粘粒插入片段的消化片段被进一步亚克隆到 pZerOTM-2 载体中,按照上述方法进行测定并进行大小排列以便获得具有最小插入片段的功能性克隆用于测序。

[0407] 在澳大利亚基因组研究实验室 (Australian Genomic Research Facility), 利用 ABI PRISM 大染料终止子循环测序快速反应试剂盒 (ABI PRISM Big Dye Terminator Cycle Sequencing Ready Reaction Kit) 对质粒插入片段进行了测序。为了进行第一轮的测序,使用了起始于可在 pZerOTM-2 载体上获得的位点的通用引物 (Sp6, T7, M13 反向或 M13 正向), 然后特制引物被用于序列的延伸。对两条 DNA 链都进行了序列测定和确认。

[0408] 实施例 13

[0409] 三种蔗糖异构酶基因在大肠杆菌中的表达

[0410] 基于通过上述功能性筛选方法克隆的基因序列,设计了三对用于将三种蔗糖异构酶基因亚克隆到表达载体 pET 24b 中的引物。通过 PCR, 删除了非编码区和前导序列并掺入了一个人造起始密码子。每个正向引物: 1) 包括一个起始密码子, 2) 创建一个用于翻译起始的植物样邻近序列 (a plant-like context), 3) 插入一个便于克隆基因和与该基因可读框相配的 BamH I 限制位点。每个反向引物掺入了一个 Kpn I 限制位点而且包括一个终止密码子。所述引物的碱基对如下所示:

[0411] 大黄欧文氏菌正向: 5' -gga tcc aac aat ggc aac cgt tcagca atc aaa tg-3' [SEQ ID NO:15]

[0412] 14S 正向: 5' -gga tcc aac aat ggc aac cgt tca caa gga aagtg-3' [SEQ ID NO:17]

[0413] 68J 正向: 5' -gga tcc aac aat ggc aac gaa tat aca aaa gtcc-3' [SEQ ID NO:13]

[0414] 大黄欧文氏菌反向: 5' -ata ggt acc tta ctt aaa cgc gtg gatg-3' [SEQ ID NO:16]

[0415] 14S 反向: 5' -ata ggt acc tta ccg cag ctt ata cac acc-3' [SEQ ID NO:18]

[0416] 68J 反向:5' -ata ggt acc tca gtt cag ctt ata gat ccc-3' [SEQ ID NO:14]

[0417] 高保真 DNA 聚合酶 pfu (Stratagene) 被用于 PCR。利用 TOPOTMTA Cloning[®] 试剂盒 (Invitrogen) 并按照试剂盒中的说明书将 PCR 产物直接克隆到 pCR[®] 2.1 载体中。

[0418] pCR[®] 2.1 载体中的三个蔗糖异构酶基因被切割下来并被克隆到 pGEM[®] -3Zf(+) 中, 然后被克隆到用于在大肠杆菌 BL21 (DE3) 菌株中进行表达的 pET 24b 载体 (Novagen) 中。含有 50 μ g/mL 卡那霉素的 5mL LB 培养基被用于 BL21 (DE3) 细胞的培养。最初由每个构建体得到了 15 个培养物。在 37°C 下 225rpm 的转速振荡培养细胞。从每个构建体中选出 6 至 10 个 OD₆₀₀ 为 1.000 $\pm$ 0.005 的培养物用于进一步的诱导。从每个培养物中取出 0.5mL 样品后, 向所述的培养物中加入 IPTG 使终浓度达到 1.0mM。将培养物再继续培养 3 小时。进一步选出只具有 OD₆₀₀ 1.750 $\pm$ 0.005 的诱导培养物用于蔗糖转化分析和蛋白质测定, 每个构建体只允许分析三个复制培养物。从每个被选出的 IPTG 诱导的培养物中取出 1.5mL 的样品用于蛋白质定量测定, 0.5mL 的样品用于蛋白质的 SDS-PAGE, 1.0mL 的样品用于蔗糖向 6-果糖- α -葡萄糖苷转化效率的定量测定。

[0419] 实施例 14

[0420] 蛋白质测定

[0421] 通过离心 (3,000xg, 4°C, 分钟) 收获细胞。将细胞沉淀物重新悬浮在 50 μ L 的 50mM Tris-HCl pH 8.0 和 2mM EDTA 中, 然后再离心。立即将细胞沉淀物冷冻在液氮中并贮存在 -70°C 下。将细胞悬浮在 0.5mL 萃取缓冲液 (20mM Tris-HCl, pH 7.4, 200mM NaCl, 1mM EDTA, 1mM 叠氮化物, 10mM β -巯基乙醇) 中, 然后通过超声波处理 (在由 Branson 超声波降解器 450 微探头发出的 50 瓦特下脉冲 9x15 秒) 进行溶胞, 再离心 (10,000xg, 4°C, 10 分钟)。上清液通过 Acrodisch[®] 32 Super[®] 0.45 μ m 膜过滤装置 (Gelman Science) 进行过滤。

[0422] 用牛血清白蛋白作为标准品, 按照 Bradford (1976, 分析生物化学 (Anal. Biochem) 72:248-254) 所记载的方法进行蛋白质测定。将 10 μ L 的上述蛋白质提取物与 90 μ L 的 0.15M NaCl 和 1mL 的考马斯亮蓝溶液 (溶于 50mL 的 95% 乙醇 + 100mL 的 85% 磷酸 + 850mL SMQ 中的 100mg 考马斯亮蓝 G-250) 混合。测定 A₅₉₅ 并由标准曲线计算蛋白质含量。

[0423] 实施例 15

[0424] SDS-PAGE

[0425] SDS 聚丙烯酰胺凝胶被聚合后按照 Laemmli (1970, 自然 (Nature) 227:680-685) 所记载的方法跑电泳。在 100°C 下将蛋白样品在 1xSDS PAGE 样品缓冲液 (25mM Tris-HCl pH 6.8, 1% (w/v) SDS, 5% (v/v) β -巯基乙醇, 10% (v/v) 甘油, 0.005% (v/v) 溴酚兰) 中加热 5 分钟, 在 12,000xg 下离心 1 分钟并将上清液加载到凝胶中。每个样品被加载到两条相邻的泳道上。跑胶后, 在 0.025% (w/v) 考马斯蓝 R-250 中对凝胶的一条泳道进行染色, 然后在 30% (v/v) 的甲醇、10% (v/v) 的乙酸中进行脱色, 从对应于染色凝胶泳道上的相对迁移位置的未染色泳道上切下被表达的蔗糖异构酶。通过将凝胶切片浸没在提取缓冲液中于 4°C 下轻轻振荡过夜而使蔗糖异构酶蛋白被从所述凝胶切片中洗脱出来。利用上述的蛋白质定量方法对洗脱出来的蔗糖异构酶进行定量测定。

[0426] 实施例 16

[0427] 通过在大肠杆菌中表达的蔗糖异构酶将蔗糖转化为 6-果糖- α -葡萄糖苷的转化率

[0428] 对 1.0mL 培养物进行离心,然后重悬在被柠檬酸盐 / 磷酸盐 (pH6.0) 缓冲的 50% 蔗糖溶液中并通过上述的 CE 分析方法测定 6-果糖- α -葡萄糖苷的转化率。利用 Beckman P/ACE 5000 系列 C.E. 系统的软件,通过利用已知浓度的标准品对蔗糖峰面积和 6-果糖- α -葡萄糖苷峰面积进行标准化来计算转化率。

[0429] 实施例 17

[0430] 构建体 DNA 的制备

[0431] 将 pET 24b 载体中的蔗糖异构酶 (SI) 插入片段进一步克隆到玉米 μ bi-1 基因的 Ubi 启动子 (Christensen 和 Quail,1996,转基因研究 (Transgen. Res)5 :215-218) 和土壤杆菌 nos 终止子 (Bevan 等,1983,自然 (Nature)304 :183-187) 之间以便驱动在甘蔗细胞中的表达。

[0432] 通过碱性提取方法 (Sambrook 等,1989,上文) 将具有蔗糖异构酶基因的质粒 (pU3ZErw,pU3Z14s 或 pU3Z68J) 和 aph A 构建体质粒 pEmuKN (作为选择性标记) 分离出来并溶解在 TE 缓冲液中。通过凝胶电泳检查质粒的完整性和基因组 DNA 或 RNA 的缺乏,并且通过分光光度法测定浓度。基本上按照 Bower 等人 (1996,分子育种 (Molec. Breed)2 :239-249) 所记载的方法将蔗糖异构酶 (UbiSI) 基因构建体和选择性标记构建体共沉淀到钨微粒上后引入甘蔗愈伤组织中,接着选择被转化的愈伤组织,并且使转基因植物再生。

[0433] 实施例 18

[0434] 粒子轰击

[0435] 通过在 4℃ 下将下列物质顺序加到 1.5mL 的微量离心管中来进行沉淀反应:5 μ L pEmuKN 质粒 DNA (1mg/mL),5 μ L UbiSI 质粒 DNA (1 μ g/ μ L),50 μ L 钨 (Bio-Rad M10, 100 μ g/ μ L),50 μ LCaCl₂ (2.5M),20 μ L 亚精胺 (100mM 游离碱基)。加入各种试剂后迅速混合制备物,使加入 CaCl₂ 和加入亚精胺之间的间隔时间最小。然后使钨在冰上静置 5 分钟,接着取出 100 μ L 的上清液并且通过摇动试管架使试管底部转动而使钨重悬。悬浮液以 4 μ L/轰击的负荷在 15 分钟内进行使用,使粒子重悬后立即取出各等分样品。假定在反应过程中,全部 DNA 被沉淀下来,这相当于 1.3 μ g DNA/轰击,或 667 μ g 钨/轰击。

[0436] 源自甘蔗栽培品种 Q117 的胚发生愈伤组织被用于轰击。按照 Bower 等人 (1996,上文) 所记载的方法在真空室中通过压缩注射滤器柄使粒子直接夹带在氦气脉冲中而加速进入靶愈伤组织中。轰击前后对组织进行 4 个小时的渗透调节。在不含抗生素的固体培养基上恢复 48 小时后,将被轰击的愈伤组织转移到含有 45mg/L 遗传霉素培养基中用于选择、愈伤组织发育和植物再生。

[0437] 实施例 19

[0438] 转化体在蔗糖转化过程中的功能

[0439] 从独立的转基因愈伤组织中收集样品并在液氮中进行研磨。而且,以未转化的 Q117 愈伤组织和被 Ubi-luc 转化的愈伤组织作为阴性对照。在 4℃ 下以 16,000xg 对被研磨的组织进行离心从而使细胞碎片沉淀。用 SMQ 将上清液稀释 10 倍,然后煮沸 20 分钟。再次进行离心以便除去变性蛋白,使上清液流过 Bond Elut™SCX 和 SAX。按照上述方法进行 CE 分析。

[0440] 有关实施例的结果和讨论

[0441] 三株具有蔗糖异构酶活性的细菌菌株被分离出来

[0442] 将一个大黄欧文氏菌(登记号:WAC 2928)的澳大利亚分离物用作 6-果糖- α -葡萄糖苷产生的阳性对照,这是由于该菌种以前已被证明能够产生将蔗糖转化为 6-果糖- α -葡萄糖苷的蔗糖异构酶(Cheetham,1985,上文)。在通过富集过程分离的总共 578 株细菌中,在苯胺/二苯胺测定中三株菌株发生了表明产生 6-果糖- α -葡萄糖苷的黄色反应,以及在 CE 测定中产生一个与 6-果糖- α -葡萄糖苷标准品和大黄欧文氏菌的 6-果糖- α -葡萄糖苷相对应的新型峰(图 1)。这三株被命名为 14S,68J 和 349J 的菌株都是能以蔗糖或 6-果糖- α -葡萄糖苷作为唯一碳源的革兰氏阴性菌。所有三株菌株在 22-30℃ 下生长良好,而且 68J 在 4℃ 下也能缓慢生长。

[0443] 三种蔗糖异构酶基因被功能性克隆并且进行测序

[0444] 在大肠杆菌中对大黄欧文氏菌、14S,349J 和 68J 的基因组粘粒文库进行功能性筛选得到了能够将蔗糖转化为 6-果糖- α -葡萄糖苷的克隆(图 2)。经过几轮亚克隆到 pZer0™-2 载体中和进行功能性筛选, pZer0™-2 载体中的最小功能性插入片段的大小范围为 3 至 5kb。

[0445] 源自大黄欧文氏菌的序列(图 3)具有一个编码 632 个氨基酸的 1899bp ORF(图 5)。对 349J 亚克隆中第一条链的测序表明存在一种与这个大黄欧文氏菌 ORF 具有 99% 相同性的基因,因此停止了 349J 的测序。14S 的序列表明具有一个编码 598 个氨基酸的 1797bpORF。利用 FASTA 进行的数据库检索表明源自大黄欧文氏菌的 1305bp 的 SI 基因和源自 14S 的全长 SI 基因已经被 Mattes 等人(上文)公开了。源自 68J 的序列(图 4)表明存在一种含有一个 1797bp 的 ORF 的新型 SI 基因。在核苷酸的水平上,其与已知蔗糖异构酶的相同性小于 70%,含或不含前导片段(表 2)。在氨基酸的水平上,与其它蔗糖异构酶的相同性在 63.4%至 70.6%之间且含有前导片段,或者 64.6%至 73.7%且不含前导片段。推定 68J 的 SI 基因产物是一种具有 598 个氨基酸的蛋白质(图 6),由 78 个碱性氨基酸残基和 69 个酸性氨基酸残基所产生的 Mr 为 69291 和等电点为 7.5。对氨基酸序列进行的种系发生分析显示了 68J SI 基因与已知基因之间的亲缘性。所有蔗糖异构酶和葡萄糖苷酶都具有糖结合区域的保守产物。结果保守序列和由 Mattes 等人(上文)记载的相应的引物对蔗糖异构酶不是特异性的而且将会产生许多源自不同生物的非 SI 基因。就象与嗜中酸假单胞菌的已知 SI 基因一样,68J 的 SI 基因与各种葡萄糖苷酶具有几乎相同水平的核苷酸相同性。

[0446]

表 2: 68J, 其它蔗糖异构酶、蔗糖异构酶片段和葡萄糖苷酶之间的特性比较

序列登 记号	菌种	注释	ORE 长度		含有前导序列的肽		不含前导序列的肽		核苷酸相同性	
			核苷酸 (bps)	肽 (氨基酸)	相似性 (%)	相同性 (%)	相似性 (%)	相同性 (%)	含有前导序 列	不含前导序 列
UQ68J		全长 UQ 分离物	1797	598						
a45846 Sudz#1	红色精朊杆菌	全长	1890	629	81.1	70.3	83.0	73.3	68.2	69.4
a45854 Sudz#9	红色精朊杆菌 (变种)	全长	1803	600	81.5	70.6	83.4	73.7	68.2	69.4
a45856 Sudz#11	肠杆菌属菌种	全长	1794	597	80.7	68.5	82.6	71.4	67.3	68.5
UQ281	大黄欧文氏菌	全长 UQ 分离物	1899	632	79.6	68.8	82.1	72.0	66.4	67.7
a45858 Sudz#13	嗜中酸假单胞菌	全长	1782	593	75.5	63.4	76.3	64.6	60.9	62.4
a45860 Sudz#16	嗜中酸假单胞菌	全长	1704	567	不含前导序列		70.4	52.4	-	56.4
a45850 Sudz#3	肠杆菌属菌种	Sudz#11 的 PCR 片段 (非功能性的)	471	157	-	-	89.1*	81.4*	-	75.4*
a45848 Sudz#2	大黄欧文氏菌	UQ281 的 N 端区域 同源性 (非功能性的)	1305	435	84.2*	74.9*	87.0*	78.5*	67.2*	72.4*
Bco16g1	蜡状芽孢杆菌	葡萄糖苷酶	1677	599	不含前导序列		68.2	48.8	-	53.2

*68J 与源自非全长蔗糖异构酶基因的非功能片段之间的比较。
Sudz#序列被公开在授予 Sudzucker 的专利 (Mattes 等) 中。

[0447] 源自 68J 的蔗糖异构酶的转化效率在所测试的异构酶中最高

[0448] 当利用相同载体 (相同启动子、起始密码子和终止序列) 分别对源自大黄欧文

氏菌、14S 和 68J 的 SI 基因进行表达时,其总蛋白含量或蔗糖异构酶的表达水平之间没有显著差异,总蛋白含量都约为 10% (表 3)。然而,被克隆的 68J 基因产物使蔗糖向 6-果糖- α -葡萄糖苷转化的转化效率为大黄欧文氏菌的 10 倍,并且是 14S 基因产物的 18 倍(图 7)。另外,与 14S 和大黄欧文氏菌的蔗糖异构酶相比,68J 的蔗糖异构酶产生葡萄糖和果糖的比例较小。基因表达过程中的所有其它因素和酶活性定量测定是相同的:用于基因构建体的相同 ATG 起始密码子邻近序列;相同的载体 pET 24b,相同的宿主细胞菌株 BL21 (DE3),相同的培养条件,IPTG 诱导前后相同的细胞密度,用于蔗糖转化的相同的细胞量,被加载到 SDS-PAGE 上的相同的总蛋白量和被加载到 CE 上的总蛋白含量相同的相同体积的上清液。所述实验进行三次且具有相同的结果。

[0449] 所述实验的结果表明源自 68J 的蔗糖异构酶在工业化生产 6-果糖- α -葡萄糖苷的应用中具有很高的潜力。

[0450] 表 3. 在含有大黄欧文氏菌、14S 或 68J[#] 的 SI 基因的大肠杆菌中的总蛋白含量和推测的蔗糖异构酶蛋白含量

[0451]

蔗糖异构酶	总蛋白含量 (%干重)				蔗糖异构酶含量 (%总蛋白*)			
大黄欧文氏菌 14S 68J 对照	15.97±1.63	15.75±1.38	16.12±1.79	14.36±2.04	12.2±1.5	11.8±0.5	12.4±1.2	1.9±0.6

[0452] [#] 结果为来源于 3 次重复实验的平均值 $\pm$ 标准误差

[0453] * 包括与蔗糖异构酶共同迁移的约 2%蛋白的本底。

[0454] 含有 68J 蔗糖异构酶的甘蔗转基因愈伤组织在所测试的蔗糖异构酶基因构建体中也具有最高的转化率

[0455] 在表达蔗糖异构酶基因的转基因甘蔗愈伤组织的细胞提取物中也发现含有 6-果糖- α -葡萄糖苷。三个被测试的 68J 转基因系在 CE 电描记图上的 6-果糖- α -葡萄糖苷峰都比蔗糖峰高(图 8A)。相反,七个被测试的 14S 转基因系中有三个转基因系的 6-果糖- α -葡萄糖苷峰比蔗糖峰低(图 8B)。在其它四个被测试的 14S 转基因系的愈伤组织中没有检测到 6-果糖- α -葡萄糖苷。含有大黄欧文氏菌基因的转基因愈伤组织中的 6-果糖- α -葡萄糖苷水平甚至低于 14S 系(图 8C)。

[0456] 这些结果首次显示了通过在植物中表达蔗糖异构酶来生产 6-果糖- α -葡萄糖苷的可行性以及蔗糖异构酶 68J 用于此目的的高潜力。

[0457] 本文引用的所有专利、专利申请和出版物的公开内容被全文引入本文作为参考。

[0458] 本文引用的任何参考文献都不应当被解释为本申请的内容,而是能够作为本申请的“现有技术”加以利用。

[0459] 整个说明书的目的始终在于描述本发明的优选实施方案而并非将本发明限定到任何一个实施方案或特定的特征组合。因此本技术领域技术人员能够理解,在本申请公开内容的启示下,对范例性的特定实施方案可以进行各种修改或改动而不偏离本发明的保护范围。所有这些修改和改动都将被包括在所附权利要求的保护范围内。

序列表

<110> 昆士兰大学 (除美国之外的所有指定国)

Birch, Robert (只有美国)

<120> 新型多肽和多核苷酸及其用途

<130> 蔗糖异构酶

<140> 还未指定

<141>2001-08-29

<150>AU PQ9788/00

<151>2000-08-29

<160>39

<170>PatentIn 版本 3.1

<210>1

<211>1899

<212>DNA

<213> 大黄欧文氏菌

<220>

<221>CDS

<222>(1).. (1896)

<223>

<220>

<221> 信号肽

<222>(1).. (108)

<223>

<220>

<221> 成熟肽

<222>(109).. (1899)

<223>

<220>

<221>misc_feature

<222>(707).. (707)

<223>n = 未知核苷酸

<220>

<221>misc_feature

<222>(1347).. (1347)

<223>n = 未知核苷酸

<400>1

atg tcc tct caa gaa ttg aaa gcg gct gtc gct att ttt ctt gca acc 48

Met Ser Ser Gln Glu Leu Lys Ala Ala Val Ala Ile Phe Leu Ala Thr

-35	-30	-25	
act ttt tct gcc aca tcc tat cag gcc tgc agt gcc ggg cca gat acc	96		
Thr Phe Ser Ala Thr Ser Tyr Gln Ala Cys Ser Ala Gly Pro Asp Thr			
-20	-15	-10	-5
gcc ccc tca ctc acc gtt cag caa tca aat gcc ctg ccc aca tgg tgg	144		
Ala Pro Ser Leu Thr Val Gln Gln Ser Asn Ala Leu Pro Thr Trp Trp			
-1 1	5	10	
aag cag gct gtt ttt tat cag gta tat cca cgc tca ttt aaa gat acg	192		
Lys Gln Ala Val Phe Tyr Gln Val Tyr Pro Arg Ser Phe Lys Asp Thr			
15	20	25	
aat ggg gat ggc att ggg gat tta aac ggt att att gag aat tta gac	240		
Asn Gly Asp Gly Ile Gly Asp Leu Asn Gly Ile Ile Glu Asn Leu Asp			
30	35	40	
tat ctg aag aaa ctg ggt att gat gcg att tgg atc aat cca cat tac	288		
Tyr Leu Lys Lys Leu Gly Ile Asp Ala Ile Trp Ile Asn Pro His Tyr			
45	50	55	60
gat tcg ccg aat acg gat aat ggt tat gac atc cgg gat tac cgt aag	336		
Asp Ser Pro Asn Thr Asp Asn Gly Tyr Asp Ile Arg Asp Tyr Arg Lys			
65	70	75	
ata atg aaa gaa tac ggt acg atg gaa gac ttt gac cgt ctt att tca	384		
Ile Met Lys Glu Tyr Gly Thr Met Glu Asp Phe Asp Arg Leu Ile Ser			
80	85	90	
gaa atg aag aaa cgc aat atg cgt ttg atg att gat att gtt atc aac	432		
Glu Met Lys Lys Arg Asn Met Arg Leu Met Ile Asp Ile Val Ile Asn			
95	100	105	
cac acc agc gat cag cat gcg tgg ttt gtt cag agc aaa tcg ggt aag	480		
His Thr Ser Asp Gln His Ala Trp Phe Val Gln Ser Lys Ser Gly Lys			
110	115	120	
aac aac ccc tac agg gac tat tac ttc tgg cgt gac ggt aag gat ggc	528		
Asn Asn Pro Tyr Arg Asp Tyr Tyr Phe Trp Arg Asp Gly Lys Asp Gly			
125	130	135	140
cat gcc ccc aat aac tat ccc tcc ttc ttc ggt ggc tca gcc tgg gaa	576		
His Ala Pro Asn Asn Tyr Pro Ser Phe Phe Gly Gly Ser Ala Trp Glu			
145	150	155	
aaa gac gat aaa tca ggc cag tat tac ctc cat tac ttt gcc aaa cag	624		
Lys Asp Asp Lys Ser Gly Gln Tyr Tyr Leu His Tyr Phe Ala Lys Gln			
160	165	170	
caa ccc gac ctc aac tgg gac aat ccc aaa gtc cgt caa gac ctg tat	672		
Gln Pro Asp Leu Asn Trp Asp Asn Pro Lys Val Arg Gln Asp Leu Tyr			

175	180	185	
gac atg ctc cgc ttc tgg tta gat aaa ggc gtt tnt ggt tta cgc ttt			720
Asp Met Leu Arg Phe Trp Leu Asp Lys Gly Val Xaa Gly Leu Arg Phe			
190	195	200	
gat acc gtt gcc acc tat tca aaa atc ccg aac ttc cct gac ctt agc			768
Asp Thr Val Ala Thr Tyr Ser Lys Ile Pro Asn Phe Pro Asp Leu Ser			
205	210	215	220
caa cag cag tta aaa aat ttc gcc gag gaa tat act aaa ggt cct aaa			816
Gln Gln Gln Leu Lys Asn Phe Ala Glu Glu Tyr Thr Lys Gly Pro Lys			
225	230	235	
att cac gac tac gtg aat gaa atg aac aga gaa gta tta tcc cac tat			864
Ile His Asp Tyr Val Asn Glu Met Asn Arg Glu Val Leu Ser His Tyr			
240	245	250	
gat atc gcc act gcg ggg gaa ata ttt ggg gtt cct ctg gat aaa tcg			912
Asp Ile Ala Thr Ala Gly Glu Ile Phe Gly Val Pro Leu Asp Lys Ser			
255	260	265	
att aag ttt ttc gat cgc cgt aga aat gaa tta aat ata gcg ttt acg			960
Ile Lys Phe Phe Asp Arg Arg Arg Asn Glu Leu Asn Ile Ala Phe Thr			
270	275	280	
ttt gat ctg atc aga ctc gat cgt gat gct gat gaa aga tgg cgg cga			1008
Phe Asp Leu Ile Arg Leu Asp Arg Asp Ala Asp Glu Arg Trp Arg Arg			
285	290	295	300
aaa gac tgg acc ctt tcg cag ttc cga aaa att gtc gat aag gtt gac			1056
Lys Asp Trp Thr Leu Ser Gln Phe Arg Lys Ile Val Asp Lys Val Asp			
305	310	315	
caa acg gca gga gag tat ggg tgg aat gcc ttt ttc tta gac aat cac			1104
Gln Thr Ala Gly Glu Tyr Gly Trp Asn Ala Phe Phe Leu Asp Asn His			
320	325	330	
gac aat ccc cgc gcg gtt tct cac ttt ggt gat gat cga cca caa tgg			1152
Asp Asn Pro Arg Ala Val Ser His Phe Gly Asp Asp Arg Pro Gln Trp			
335	340	345	
cgc gag cat gcg gcg aaa gca ctg gca aca ttg acg ctg acc cag cgt			1200
Arg Glu His Ala Ala Lys Ala Leu Ala Thr Leu Thr Leu Thr Gln Arg			
350	355	360	
gca acg ccg ttt atc tat cag ggt tca gaa ctc ggt atg acc aat tat			1248
Ala Thr Pro Phe Ile Tyr Gln Gly Ser Glu Leu Gly Met Thr Asn Tyr			
365	370	375	380
ccc ttt aaa aaa atc gat gat ttc gat gat gta gag gtg aaa ggt ttt			1296
Pro Phe Lys Lys Ile Asp Asp Phe Asp Asp Val Glu Val Lys Gly Phe			

385	390	395	
tgg caa gac tac gtt gaa aca ggc aaa gtg aaa gct gag gaa ttc ctt			1344
Trp Gln Asp Tyr Val Glu Thr Gly Lys Val Lys Ala Glu Glu Phe Leu			
400	405	410	
can aac gta cgc caa acc agc cgt gat aac agc aga acc ccc ttc cag			1392
Thr Asn Val Arg Gln Thr Ser Arg Asp Asn Ser Arg Thr Pro Phe Gln			
415	420	425	
tgg gat gca agc aaa aat gcg ggc ttt acc agc gga acc cct tgg tta			1440
Trp Asp Ala Ser Lys Asn Ala Gly Phe Thr Ser Gly Thr Pro Trp Leu			
430	435	440	
aaa atc aat ccc aat tat aaa gaa atc aac agc gca gat cag att aac			1488
Lys Ile Asn Pro Asn Tyr Lys Glu Ile Asn Ser Ala Asp Gln Ile Asn			
445	450	455	460
aat cca aat tcc gta ttt aac tat tat aga aag ctc att aac att cgc			1536
Asn Pro Asn Ser Val Phe Asn Tyr Tyr Arg Lys Leu Ile Asn Ile Arg			
465	470	475	
cac gac atc cct gcc tta acc tac ggc agt tat att gat tta gct cct			1584
His Asp Ile Pro Ala Leu Thr Tyr Gly Ser Tyr Ile Asp Leu Ala Pro			
480	485	490	
gac aac aat tca gtc tat gct tac act cga acg ttt ggc gct gaa aaa			1632
Asp Asn Asn Ser Val Tyr Ala Tyr Thr Arg Thr Phe Gly Ala Glu Lys			
495	500	505	
tat ctt gtg gtc att aat ttt aaa gaa gaa gtg atg cac tac acc ctg			1680
Tyr Leu Val Val Ile Asn Phe Lys Glu Glu Val Met His Tyr Thr Leu			
510	515	520	
cct ggg gat tta tcc atc aat aag gtg att act gaa aac aac agt cac			1728
Pro Gly Asp Leu Ser Ile Asn Lys Val Ile Thr Glu Asn Asn Ser His			
525	530	535	540
act att gtg aat aaa aat gac gta gaa gat cct cgt ggg gct aca agc			1776
Thr Ile Val Asn Lys Asn Asp Val Glu Asp Pro Arg Gly Ala Thr Ser			
545	550	555	
gtt tgt agc ccc ttc cag gct caa aaa agg cct ggc gac ccg ggt tac			1824
Val Cys Ser Pro Phe Gln Ala Gln Lys Arg Pro Gly Asp Pro Gly Tyr			
560	565	570	
tct gct gcc cat teg att cgg ttc ttg ccc cgg ttt ttc gct tca tac			1872
Ser Ala Ala His Ser Ile Arg Phe Leu Pro Arg Phe Phe Ala Ser Tyr			
575	580	585	
agg ggc gac atc cac gcg ttt aag taa			1899
Arg Gly Asp Ile His Ala Phe Lys			

590 595
 <210>2
 <211>632
 <212>PRT
 <213> 大黄欧文氏菌
 <220>
 <221>misc_feature
 <222>(200).. (200)
 <223>200 位上的 ' Xaa' 代表 Tyr、Cyr、Ser 或 Phe
 <220>
 <221>misc_feature
 <222>(707).. (707)
 <223>n = 未知核苷酸
 <220>
 <221>misc_feature
 <222>(1347).. (1347)
 <223>n = 未知核苷酸
 <400>2
 Met Ser Ser Gln Glu Leu Lys Ala Ala Val Ala Ile Phe Leu Ala Thr
 -35 -30 -25
 Thr Phe Ser Ala Thr Ser Tyr Gln Ala Cys Ser Ala Gly Pro Asp Thr
 -20 -15 -10 -5
 Ala Pro Ser Leu Thr Val Gln Gln Ser Asn Ala Leu Pro Thr Trp Trp
 -1 1 5 10
 Lys Gln Ala Val Phe Tyr Gln Val Tyr Pro Arg Ser Phe Lys Asp Thr
 15 20 25
 Asn Gly Asp Gly Ile Gly Asp Leu Asn Gly Ile Ile Glu Asn Leu Asp
 30 35 40
 Tyr Leu Lys Lys Leu Gly Ile Asp Ala Ile Trp Ile Asn Pro His Tyr
 45 50 55 60
 Asp Ser Pro Asn Thr Asp Asn Gly Tyr Asp Ile Arg Asp Tyr Arg Lys
 65 70 75
 Ile Met Lys Glu Tyr Gly Thr Met Glu Asp Phe Asp Arg Leu Ile Ser
 80 85 90
 Glu Met Lys Lys Arg Asn Met Arg Leu Met Ile Asp Ile Val Ile Asn
 95 100 105
 His Thr Ser Asp Gln His Ala Trp Phe Val Gln Ser Lys Ser Gly Lys
 110 115 120
 Asn Asn Pro Tyr Arg Asp Tyr Tyr Phe Trp Arg Asp Gly Lys Asp Gly

125	130	135	140
His Ala Pro Asn Asn Tyr Pro Ser Phe Phe Gly Gly Ser Ala Trp Glu			
145	150	155	
Lys Asp Asp Lys Ser Gly Gln Tyr Tyr Leu His Tyr Phe Ala Lys Gln			
160	165	170	
Gln Pro Asp Leu Asn Trp Asp Asn Pro Lys Val Arg Gln Asp Leu Tyr			
175	180	185	
Asp Met Leu Arg Phe Trp Leu Asp Lys Gly Val Xaa Gly Leu Arg Phe			
190	195	200	
Asp Thr Val Ala Thr Tyr Ser Lys Ile Pro Asn Phe Pro Asp Leu Ser			
205	210	215	220
Gln Gln Gln Leu Lys Asn Phe Ala Glu Glu Tyr Thr Lys Gly Pro Lys			
225	230	235	
Ile His Asp Tyr Val Asn Glu Met Asn Arg Glu Val Leu Ser His Tyr			
240	245	250	
Asp Ile Ala Thr Ala Gly Glu Ile Phe Gly Val Pro Leu Asp Lys Ser			
255	260	265	
Ile Lys Phe Phe Asp Arg Arg Arg Asn Glu Leu Asn Ile Ala Phe Thr			
270	275	280	
Phe Asp Leu Ile Arg Leu Asp Arg Asp Ala Asp Glu Arg Trp Arg Arg			
285	290	295	300
Lys Asp Trp Thr Leu Ser Gln Phe Arg Lys Ile Val Asp Lys Val Asp			
305	310	315	
Gln Thr Ala Gly Glu Tyr Gly Trp Asn Ala Phe Phe Leu Asp Asn His			
320	325	330	
Asp Asn Pro Arg Ala Val Ser His Phe Gly Asp Asp Arg Pro Gln Trp			
335	340	345	
Arg Glu His Ala Ala Lys Ala Leu Ala Thr Leu Thr Leu Thr Gln Arg			
350	355	360	
Ala Thr Pro Phe Ile Tyr Gln Gly Ser Glu Leu Gly Met Thr Asn Tyr			
365	370	375	380
Pro Phe Lys Lys Ile Asp Asp Phe Asp Asp Val Glu Val Lys Gly Phe			
385	390	395	
Trp Gln Asp Tyr Val Glu Thr Gly Lys Val Lys Ala Glu Glu Phe Leu			
400	405	410	
Thr Asn Val Arg Gln Thr Ser Arg Asp Asn Ser Arg Thr Pro Phe Gln			
415	420	425	
Trp Asp Ala Ser Lys Asn Ala Gly Phe Thr Ser Gly Thr Pro Trp Leu			
430	435	440	

Lys Ile Asn Pro Asn Tyr Lys Glu Ile Asn Ser Ala Asp Gln Ile Asn
445 450 455 460
Asn Pro Asn Ser Val Phe Asn Tyr Tyr Arg Lys Leu Ile Asn Ile Arg
465 470 475
His Asp Ile Pro Ala Leu Thr Tyr Gly Ser Tyr Ile Asp Leu Ala Pro
480 485 490
Asp Asn Asn Ser Val Tyr Ala Tyr Thr Arg Thr Phe Gly Ala Glu Lys
495 500 505
Tyr Leu Val Val Ile Asn Phe Lys Glu Glu Val Met His Tyr Thr Leu
510 515 520
Pro Gly Asp Leu Ser Ile Asn Lys Val Ile Thr Glu Asn Asn Ser His
525 530 535 540
Thr Ile Val Asn Lys Asn Asp Val Glu Asp Pro Arg Gly Ala Thr Ser
545 550 555
Val Cys Ser Pro Phe Gln Ala Gln Lys Arg Pro Gly Asp Pro Gly Tyr
560 565 570
Ser Ala Ala His Ser Ile Arg Phe Leu Pro Arg Phe Phe Ala Ser Tyr
575 580 585
Arg Gly Asp Ile His Ala Phe Lys
590 595

<210>3

<211>1791

<212>DNA

<213> 大黄欧文氏菌

<220>

<221>CDS

<222>(1).. (1788)

<223>

<220>

<221> 成熟肽

<222>(1).. (1791)

<223>

<220>

<221>misc_feature

<222>(599).. (599)

<223>n = 未知核苷酸

<220>

<221>misc_feature

<222>(1239).. (1239)

<223>n = 未知核苷酸

<400>3

acc gtt cag caa tca aat gcc ctg ccc aca tgg tgg aag cag gct gtt	48
Thr Val Gln Gln Ser Asn Ala Leu Pro Thr Trp Trp Lys Gln Ala Val	
1 5 10 15	
ttt tat cag gta tat cca cgc tca ttt aaa gat acg aat ggg gat ggc	96
Phe Tyr Gln Val Tyr Pro Arg Ser Phe Lys Asp Thr Asn Gly Asp Gly	
20 25 30	
att ggg gat tta aac ggt att att gag aat tta gac tat ctg aag aaa	144
Ile Gly Asp Leu Asn Gly Ile Ile Glu Asn Leu Asp Tyr Leu Lys Lys	
35 40 45	
ctg ggt att gat gcg att tgg atc aat cca cat tac gat tcg ccg aat	192
Leu Gly Ile Asp Ala Ile Trp Ile Asn Pro His Tyr Asp Ser Pro Asn	
50 55 60	
acg gat aat ggt tat gac atc cgg gat tac cgt aag ata atg aaa gaa	240
Thr Asp Asn Gly Tyr Asp Ile Arg Asp Tyr Arg Lys Ile Met Lys Glu	
65 70 75 80	
tac ggt acg atg gaa gac ttt gac cgt ctt att tca gaa atg aag aaa	288
Tyr Gly Thr Met Glu Asp Phe Asp Arg Leu Ile Ser Glu Met Lys Lys	
85 90 95	
cgc aat atg cgt ttg atg att gat att gtt atc aac cac acc agc gat	336
Arg Asn Met Arg Leu Met Ile Asp Ile Val Ile Asn His Thr Ser Asp	
100 105 110	
cag cat gcg tgg ttt gtt cag agc aaa tcg ggt aag aac aac ccc tac	384
Gln His Ala Trp Phe Val Gln Ser Lys Ser Gly Lys Asn Asn Pro Tyr	
115 120 125	
agg gac tat tac ttc tgg cgt gac ggt aag gat ggc cat gcc ccc aat	432
Arg Asp Tyr Tyr Phe Trp Arg Asp Gly Lys Asp Gly His Ala Pro Asn	
130 135 140	
aac tat ccc tcc ttc ttc ggt ggc tca gcc tgg gaa aaa gac gat aaa	480
Asn Tyr Pro Ser Phe Phe Gly Gly Ser Ala Trp Glu Lys Asp Asp Lys	
145 150 155 160	
tca ggc cag tat tac ctc cat tac ttt gcc aaa cag caa ccc gac ctc	528
Ser Gly Gln Tyr Tyr Leu His Tyr Phe Ala Lys Gln Gln Pro Asp Leu	
165 170 175	
aac tgg gac aat ccc aaa gtc cgt caa gac ctg tat gac atg ctc cgc	576
Asn Trp Asp Asn Pro Lys Val Arg Gln Asp Leu Tyr Asp Met Leu Arg	
180 185 190	
ttc tgg tta gat aaa ggc gtt tnt ggt tta cgc ttt gat acc gtt gcc	624

Phe	Trp	Leu	Asp	Lys	Gly	Val	Xaa	Gly	Leu	Arg	Phe	Asp	Thr	Val	Ala	
195				200				205								
acc	tat	tca	aaa	atc	ccg	aac	ttc	cct	gac	ctt	agc	caa	cag	cag	tta	672
Thr	Tyr	Ser	Lys	Ile	Pro	Asn	Phe	Pro	Asp	Leu	Ser	Gln	Gln	Gln	Leu	
210				215				220								
aaa	aat	ttc	gcc	gag	gaa	tat	act	aaa	ggt	cct	aaa	att	cac	gac	tac	720
Lys	Asn	Phe	Ala	Glu	Glu	Tyr	Thr	Lys	Gly	Pro	Lys	Ile	His	Asp	Tyr	
225				230				235				240				
gtg	aat	gaa	atg	aac	aga	gaa	gta	tta	tcc	cac	tat	gat	atc	gcc	act	768
Val	Asn	Glu	Met	Asn	Arg	Glu	Val	Leu	Ser	His	Tyr	Asp	Ile	Ala	Thr	
245				250				255								
gcg	ggg	gaa	ata	ttt	ggg	gtt	cct	ctg	gat	aaa	tcg	att	aag	ttt	ttc	816
Ala	Gly	Glu	Ile	Phe	Gly	Val	Pro	Leu	Asp	Lys	Ser	Ile	Lys	Phe	Phe	
260				265				270								
gat	cgc	cgt	aga	aat	gaa	tta	aat	ata	gcg	ttt	acg	ttt	gat	ctg	atc	864
Asp	Arg	Arg	Arg	Asn	Glu	Leu	Asn	Ile	Ala	Phe	Thr	Phe	Asp	Leu	Ile	
275				280				285								
aga	ctc	gat	cgt	gat	gct	gat	gaa	aga	tgg	cgg	cga	aaa	gac	tgg	acc	912
Arg	Leu	Asp	Arg	Asp	Ala	Asp	Glu	Arg	Trp	Arg	Arg	Lys	Asp	Trp	Thr	
290				295				300								
ctt	tcg	cag	ttc	cga	aaa	att	gtc	gat	aag	gtt	gac	caa	acg	gca	gga	960
Leu	Ser	Gln	Phe	Arg	Lys	Ile	Val	Asp	Lys	Val	Asp	Gln	Thr	Ala	Gly	
305				310				315				320				
gag	tat	ggg	tgg	aat	gcc	ttt	ttc	tta	gac	aat	cac	gac	aat	ccc	cgc	1008
Glu	Tyr	Gly	Trp	Asn	Ala	Phe	Phe	Leu	Asp	Asn	His	Asp	Asn	Pro	Arg	
325				330				335								
gcg	gtt	tct	cac	ttt	ggt	gat	gat	cga	cca	caa	tgg	cgc	gag	cat	gcg	1056
Ala	Val	Ser	His	Phe	Gly	Asp	Asp	Arg	Pro	Gln	Trp	Arg	Glu	His	Ala	
340				345				350								
gcg	aaa	gca	ctg	gca	aca	ttg	acg	ctg	acc	cag	cgt	gca	acg	ccg	ttt	1104
Ala	Lys	Ala	Leu	Ala	Thr	Leu	Thr	Leu	Thr	Gln	Arg	Ala	Thr	Pro	Phe	
355				360				365								
atc	tat	cag	ggt	tca	gaa	ctc	ggt	atg	acc	aat	tat	ccc	ttt	aaa	aaa	1152
Ile	Tyr	Gln	Gly	Ser	Glu	Leu	Gly	Met	Thr	Asn	Tyr	Pro	Phe	Lys	Lys	
370				375				380								
atc	gat	gat	ttc	gat	gat	gta	gag	gtg	aaa	ggt	ttt	tgg	caa	gac	tac	1200
Ile	Asp	Asp	Phe	Asp	Asp	Val	Glu	Val	Lys	Gly	Phe	Trp	Gln	Asp	Tyr	
385				390				395				400				
gtt	gaa	aca	ggc	aaa	gtg	aaa	gct	gag	gaa	ttc	ctt	can	aac	gta	cgc	1248

Val Glu Thr Gly Lys Val Lys Ala Glu Glu Phe Leu Thr Asn Val Arg	
405 410 415	
caa acc agc cgt gat aac agc aga acc ccc ttc cag tgg gat gca agc	1296
Gln Thr Ser Arg Asp Asn Ser Arg Thr Pro Phe Gln Trp Asp Ala Ser	
420 425 430	
aaa aat gcg ggc ttt acc agc gga acc cct tgg tta aaa atc aat ccc	1344
Lys Asn Ala Gly Phe Thr Ser Gly Thr Pro Trp Leu Lys Ile Asn Pro	
435 440 445	
aat tat aaa gaa atc aac agc gca gat cag att aac aat cca aat tcc	1392
Asn Tyr Lys Glu Ile Asn Ser Ala Asp Gln Ile Asn Asn Pro Asn Ser	
450 455 460	
gta ttt aac tat tat aga aag ctc att aac att cgc cac gac atc cct	1440
Val Phe Asn Tyr Tyr Arg Lys Leu Ile Asn Ile Arg His Asp Ile Pro	
465 470 475 480	
gcc tta acc tac ggc agt tat att gat tta gct cct gac aac aat tca	1488
Ala Leu Thr Tyr Gly Ser Tyr Ile Asp Leu Ala Pro Asp Asn Asn Ser	
485 490 495	
gtc tat gct tac act cga acg ttt ggc gct gaa aaa tat ctt gtg gtc	1536
Val Tyr Ala Tyr Thr Arg Thr Phe Gly Ala Glu Lys Tyr Leu Val Val	
500 505 510	
att aat ttt aaa gaa gaa gtg atg cac tac acc ctg cct ggg gat tta	1584
Ile Asn Phe Lys Glu Glu Val Met His Tyr Thr Leu Pro Gly Asp Leu	
515 520 525	
tcc atc aat aag gtg att act gaa aac aac agt cac act att gtg aat	1632
Ser Ile Asn Lys Val Ile Thr Glu Asn Asn Ser His Thr Ile Val Asn	
530 535 540	
aaa aat gac gta gaa gat cct cgt ggg gct aca agc gtt tgt agc ccc	1680
Lys Asn Asp Val Glu Asp Pro Arg Gly Ala Thr Ser Val Cys Ser Pro	
545 550 555 560	
ttc cag gct caa aaa agg cct ggc gac ccg ggt tac tct gct gcc cat	1728
Phe Gln Ala Gln Lys Arg Pro Gly Asp Pro Gly Tyr Ser Ala Ala His	
565 570 575	
tcg att cgg ttc ttg ccc cgg ttt ttc gct tca tac agg ggc gac atc	1776
Ser Ile Arg Phe Leu Pro Arg Phe Phe Ala Ser Tyr Arg Gly Asp Ile	
580 585 590	
cac gcg ttt aag taa	1791
His Ala Phe Lys	
595	
<210>4	

<211>596

<212>PRT

<213> 大黄欧文氏菌

<220>

<221>misc_feature

<222>(200)..(200)

<223>200 位上的 ' Xaa' 代表 Tyr、Cys、Ser 或 Phe.

<220>

<221>misc_feature

<222>(599)..(599)

<223>n = 未知核苷酸

<220>

<221>misc_feature

<222>(1239)..(1239)

<223>n = 未知核苷酸

<400>4

Thr Val Gln Gln Ser Asn Ala Leu Pro Thr Trp Trp Lys Gln Ala Val

1 5 10 15

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

20 25 30

Ile Gly Asp Leu Asn Gly Ile Ile Glu Asn Leu Asp Tyr Leu Lys Lys

35 40 45

Leu Gly Ile Asp Ala Ile Trp Ile Asn Pro His Tyr Asp Ser Pro Asn

50 55 60

Thr Asp Asn Gly Tyr Asp Ile Arg Asp Tyr Arg Lys Ile Met Lys Glu

65 70 75 80

Tyr Gly Thr Met Glu Asp Phe Asp Arg Leu Ile Ser Glu Met Lys Lys

85 90 95

Arg Asn Met Arg Leu Met Ile Asp Ile Val Ile Asn His Thr Ser Asp

100 105 110

Gln His Ala Trp Phe Val Gln Ser Lys Ser Gly Lys Asn Asn Pro Tyr

115 120 125

Arg Asp Tyr Tyr Phe Trp Arg Asp Gly Lys Asp Gly His Ala Pro Asn

130 135 140

Asn Tyr Pro Ser Phe Phe Gly Gly Ser Ala Trp Glu Lys Asp Asp Lys

145 150 155 160

Ser Gly Gln Tyr Tyr Leu His Tyr Phe Ala Lys Gln Gln Pro Asp Leu

165 170 175

Asn Trp Asp Asn Pro Lys Val Arg Gln Asp Leu Tyr Asp Met Leu Arg

180					185					190						
Phe	Trp	Leu	Asp	Lys	Gly	Val	Xaa	Gly	Leu	Arg	Phe	Asp	Thr	Val	Ala	
195					200					205						
Thr	Tyr	Ser	Lys	Ile	Pro	Asn	Phe	Pro	Asp	Leu	Ser	Gln	Gln	Gln	Leu	
210					215					220						
Lys	Asn	Phe	Ala	Glu	Glu	Tyr	Thr	Lys	Gly	Pro	Lys	Ile	His	Asp	Tyr	
225					230					235					240	
Val	Asn	Glu	Met	Asn	Arg	Glu	Val	Leu	Ser	His	Tyr	Asp	Ile	Ala	Thr	
245					250					255						
Ala	Gly	Glu	Ile	Phe	Gly	Val	Pro	Leu	Asp	Lys	Ser	Ile	Lys	Phe	Phe	
260					265					270						
Asp	Arg	Arg	Arg	Asn	Glu	Leu	Asn	Ile	Ala	Phe	Thr	Phe	Asp	Leu	Ile	
275					280					285						
Arg	Leu	Asp	Arg	Asp	Ala	Asp	Glu	Arg	Trp	Arg	Arg	Lys	Asp	Trp	Thr	
290					295					300						
Leu	Ser	Gln	Phe	Arg	Lys	Ile	Val	Asp	Lys	Val	Asp	Gln	Thr	Ala	Gly	
305					310					315					320	
Glu	Tyr	Gly	Trp	Asn	Ala	Phe	Phe	Leu	Asp	Asn	His	Asp	Asn	Pro	Arg	
325					330					335						
Ala	Val	Ser	His	Phe	Gly	Asp	Asp	Arg	Pro	Gln	Trp	Arg	Glu	His	Ala	
340					345					350						
Ala	Lys	Ala	Leu	Ala	Thr	Leu	Thr	Leu	Thr	Gln	Arg	Ala	Thr	Pro	Phe	
355					360					365						
Ile	Tyr	Gln	Gly	Ser	Glu	Leu	Gly	Met	Thr	Asn	Tyr	Pro	Phe	Lys	Lys	
370					375					380						
Ile	Asp	Asp	Phe	Asp	Asp	Val	Glu	Val	Lys	Gly	Phe	Trp	Gln	Asp	Tyr	
385					390					395					400	
Val	Glu	Thr	Gly	Lys	Val	Lys	Ala	Glu	Glu	Phe	Leu	Thr	Asn	Val	Arg	
405					410					415						
Gln	Thr	Ser	Arg	Asp	Asn	Ser	Arg	Thr	Pro	Phe	Gln	Trp	Asp	Ala	Ser	
420					425					430						
Lys	Asn	Ala	Gly	Phe	Thr	Ser	Gly	Thr	Pro	Trp	Leu	Lys	Ile	Asn	Pro	
435					440					445						
Asn	Tyr	Lys	Glu	Ile	Asn	Ser	Ala	Asp	Gln	Ile	Asn	Asn	Pro	Asn	Ser	
450					455					460						
Val	Phe	Asn	Tyr	Tyr	Arg	Lys	Leu	Ile	Asn	Ile	Arg	His	Asp	Ile	Pro	
465					470					475					480	
Ala	Leu	Thr	Tyr	Gly	Ser	Tyr	Ile	Asp	Leu	Ala	Pro	Asp	Asn	Asn	Ser	
485					490					495						

Val Tyr Ala Tyr Thr Arg Thr Phe Gly Ala Glu Lys Tyr Leu Val Val
500 505 510
Ile Asn Phe Lys Glu Glu Val Met His Tyr Thr Leu Pro Gly Asp Leu
515 520 525
Ser Ile Asn Lys Val Ile Thr Glu Asn Asn Ser His Thr Ile Val Asn
530 535 540
Lys Asn Asp Val Glu Asp Pro Arg Gly Ala Thr Ser Val Cys Ser Pro
545 550 555 560
Phe Gln Ala Gln Lys Arg Pro Gly Asp Pro Gly Tyr Ser Ala Ala His
565 570 575
Ser Ile Arg Phe Leu Pro Arg Phe Phe Ala Ser Tyr Arg Gly Asp Ile
580 585 590
His Ala Phe Lys
595

<210>5

<211>108

<212>DNA

<213> 大黄欧文氏菌

<220>

<221>CDS

<222>(1).. (108)

<223>

<220>

<221> 信号肽

<222>(1).. (108)

<223>

<400>5

atg tcc tct caa gaa ttg aaa gcg gct gtc gct att ttt ctt gca acc 48
Met Ser Ser Gln Glu Leu Lys Ala Ala Val Ala Ile Phe Leu Ala Thr
1 5 10 15
act ttt tct gcc aca tcc tat cag gcc tgc agt gcc ggg cca gat acc 96
Thr Phe Ser Ala Thr Ser Tyr Gln Ala Cys Ser Ala Gly Pro Asp Thr
20 25 30
gcc ccc tca ctc 108
Ala Pro Ser Leu
35

<210>6

<211>36

<212>PRT

<213> 大黄欧文氏菌

<400>6

```
Met Ser Ser Gln Glu Leu Lys Ala Ala Val Ala Ile Phe Leu Ala Thr
1           5           10           15
Thr Phe Ser Ala Thr Ser Tyr Gln Ala Cys Ser Ala Gly Pro Asp Thr
           20           25           30
Ala Pro Ser Leu
           35
```

<210>7

<211>1797

<212>DNA

<213> 细菌分离物 68J

<220>

<221>CDS

<222>(1).. (1794)

<223>

<220>

<221> 信号肽

<222>(1).. (99)

<223>

<220>

<221> 成熟肽

<222>(100).. (1797)

<223>

<220>

<221>misc_feature

<222>(1478).. (1478)

<223>n = 未知核苷酸

__n

<400>7

```
atg ttt ctt aat gga ttt aag aca gtt att gct ctg act atg gca agc      48
Met Phe Leu Asn Gly Phe Lys Thr Val Ile Ala Leu Thr Met Ala Ser
           -30           -25           -20
tcg ttt tat ctt gcc gcc agc ccg tta act aag cca tcg acc cct att      96
Ser Phe Tyr Leu Ala Ala Ser Pro Leu Thr Lys Pro Ser Thr Pro Ile
           -15           -10           -5
gcc gca acg aat ata caa aag tcc gct gat ttt ccc att tgg tgg aaa     144
Ala Ala Thr Asn Ile Gln Lys Ser Ala Asp Phe Pro Ile Trp Trp Lys
-1  1           5           10           15
```

cag gca gta ttt tac cag att tat ccc cgc tca ttt aaa gat agc aat	192
Gln Ala Val Phe Tyr Gln Ile Tyr Pro Arg Ser Phe Lys Asp Ser Asn	
20 25 30	
ggt gat ggt atc ggc gat att ccc ggt atc att gag aaa ctg gac tat	240
Gly Asp Gly Ile Gly Asp Ile Pro Gly Ile Ile Glu Lys Leu Asp Tyr	
35 40 45	
tta aaa atg ctg gga gtt gat gct atc tgg ata aac ccg cac tat gag	288
Leu Lys Met Leu Gly Val Asp Ala Ile Trp Ile Asn Pro His Tyr Glu	
50 55 60	
tct cct aac acc gac aat ggt tac gat att agt gat tat cgt aaa atc	336
Ser Pro Asn Thr Asp Asn Gly Tyr Asp Ile Ser Asp Tyr Arg Lys Ile	
65 70 75	
atg aag gag tac ggc agc atg gct gac ttt gac cgt ctg gtt gcc gaa	384
Met Lys Glu Tyr Gly Ser Met Ala Asp Phe Asp Arg Leu Val Ala Glu	
80 85 90 95	
atg aat aaa cgt ggt atg cgc ctg atg att gat att gtt atc aat cat	432
Met Asn Lys Arg Gly Met Arg Leu Met Ile Asp Ile Val Ile Asn His	
100 105 110	
acc agc gat cgt cac cgc tgg ttt gtg cag agc cgt tca ggt aaa gat	480
Thr Ser Asp Arg His Arg Trp Phe Val Gln Ser Arg Ser Gly Lys Asp	
115 120 125	
aat cct tac cgc gac tat tat ttc tgg cgt gat ggt aaa cag gga cag	528
Asn Pro Tyr Arg Asp Tyr Tyr Phe Trp Arg Asp Gly Lys Gln Gly Gln	
130 135 140	
gct ccc aat aac tat ccc tct ttc ttt ggc ggt tca gcc tgg caa ctg	576
Ala Pro Asn Asn Tyr Pro Ser Phe Phe Gly Gly Ser Ala Trp Gln Leu	
145 150 155	
gat aaa cag act gac cag tat tat ctg cac tat ttt gca cca cag cag	624
Asp Lys Gln Thr Asp Gln Tyr Tyr Leu His Tyr Phe Ala Pro Gln Gln	
160 165 170 175	
ccg gat ctg aac tgg gat aac cca aaa gtt cgg gct gaa ctc tac gat	672
Pro Asp Leu Asn Trp Asp Asn Pro Lys Val Arg Ala Glu Leu Tyr Asp	
180 185 190	
att ctg cgt ttc tgg ctg gat aaa ggc gta tcc gga cta cgt ttt gat	720
Ile Leu Arg Phe Trp Leu Asp Lys Gly Val Ser Gly Leu Arg Phe Asp	
195 200 205	
acc gtg gct act ttc tcc aaa att cct ggc ttc ccg gac ctg tca aaa	768
Thr Val Ala Thr Phe Ser Lys Ile Pro Gly Phe Pro Asp Leu Ser Lys	
210 215 220	

gcg cag ctg aag aat ttt gcc gaa gct tat act gag ggg ccg aat att	816
Ala Gln Leu Lys Asn Phe Ala Glu Ala Tyr Thr Glu Gly Pro Asn Ile	
225 230 235	
cat aaa tat atc cat gaa atg aac cgc cag gta ctg tct aaa tat aat	864
His Lys Tyr Ile His Glu Met Asn Arg Gln Val Leu Ser Lys Tyr Asn	
240 245 250 255	
gtt gcc acc gct ggt gaa atc ttc ggt gtg cca gtg agt gct atg ccg	912
Val Ala Thr Ala Gly Glu Ile Phe Gly Val Pro Val Ser Ala Met Pro	
260 265 270	
gat tat ttt gac cgg cgg cgt gaa gaa ctc aat att gct ttc acc ttt	960
Asp Tyr Phe Asp Arg Arg Arg Glu Glu Leu Asn Ile Ala Phe Thr Phe	
275 280 285	
gat ttg atc agg ctc gat cgt tat ccc gat cag cgc tgg cgt cgt aaa	1008
Asp Leu Ile Arg Leu Asp Arg Tyr Pro Asp Gln Arg Trp Arg Arg Lys	
290 295 300	
cca tgg aca tta agc cag ttt cgt caa gtt atc tct cag act gac cgt	1056
Pro Trp Thr Leu Ser Gln Phe Arg Gln Val Ile Ser Gln Thr Asp Arg	
305 310 315	
gcc gcc ggt gaa ttt ggc tgg aac gcc ttt ttc ctt gat aac cat gat	1104
Ala Ala Gly Glu Phe Gly Trp Asn Ala Phe Phe Leu Asp Asn His Asp	
320 325 330 335	
aac ccg cgc cag gtc tca cac ttt ggt gac gac agc cca caa tgg cgc	1152
Asn Pro Arg Gln Val Ser His Phe Gly Asp Asp Ser Pro Gln Trp Arg	
340 345 350	
gaa cgc tcg gca aaa gca ctg gca acg ctg ctg ctg acg cag cgt gcc	1200
Glu Arg Ser Ala Lys Ala Leu Ala Thr Leu Leu Leu Thr Gln Arg Ala	
355 360 365	
acg ccg ttt atc ttt cag ggg gcg gag ttg gga atg act aat tac ccc	1248
Thr Pro Phe Ile Phe Gln Gly Ala Glu Leu Gly Met Thr Asn Tyr Pro	
370 375 380	
ttt aaa aat ata gag gaa ttt gat gat att gag gtt aaa ggc ttc tgg	1296
Phe Lys Asn Ile Glu Glu Phe Asp Asp Ile Glu Val Lys Gly Phe Trp	
385 390 395	
aac gac tat gta gcc agc gga aaa gta aac gct gct gaa ttt tta cag	1344
Asn Asp Tyr Val Ala Ser Gly Lys Val Asn Ala Ala Glu Phe Leu Gln	
400 405 410 415	
gag gtt cgc atg acc agc cgc gat aac agc cga aca cca atg cag tgg	1392
Glu Val Arg Met Thr Ser Arg Asp Asn Ser Arg Thr Pro Met Gln Trp	
420 425 430	

aac gac tct gtt aat gcc gga ttc acc cag ggc aaa ccc tgg ttt cac	1440
Asn Asp Ser Val Asn Ala Gly Phe Thr Gln Gly Lys Pro Trp Phe His	
435 440 445	
ctc aat ccc aac tat aag caa atc aat gcc gcc agg gng gtg aat aaa	1488
Leu Asn Pro Asn Tyr Lys Gln Ile Asn Ala Ala Arg Xaa Val Asn Lys	
450 455 460	
ccc gac tcg gta ttc agt tac tac cgt caa ctg atc aac ctg cgt cac	1536
Pro Asp Ser Val Phe Ser Tyr Tyr Arg Gln Leu Ile Asn Leu Arg His	
465 470 475	
cag atc ccg gca ctg acc agt ggt gaa tac cgt gat ctc gat ccg cag	1584
Gln Ile Pro Ala Leu Thr Ser Gly Glu Tyr Arg Asp Leu Asp Pro Gln	
480 485 490 495	
aat aac cag gtc tat gcc tat acc cgt ata ctg gat aat gaa aaa tat	1632
Asn Asn Gln Val Tyr Ala Tyr Thr Arg Ile Leu Asp Asn Glu Lys Tyr	
500 505 510	
ctg gtg gta gtt aat ttt aaa cct gag cag ctg cat tac gct ctg cca	1680
Leu Val Val Val Asn Phe Lys Pro Glu Gln Leu His Tyr Ala Leu Pro	
515 520 525	
gat aat ctg act att gcc agc agt ctg ctg gaa aat gtc cac caa cca	1728
Asp Asn Leu Thr Ile Ala Ser Ser Leu Leu Glu Asn Val His Gln Pro	
530 535 540	
tca ctg caa gaa aat gcc tcc acg ctg act ctt gct ccg tgg caa gcc	1776
Ser Leu Gln Glu Asn Ala Ser Thr Leu Thr Leu Ala Pro Trp Gln Ala	
545 550 555	
ggg atc tat aag ctg aac tga	1797
Gly Ile Tyr Lys Leu Asn	
560 565	
<210>8	
<211>598	
<212>PRT	
<213> 细菌分离物 68J	
<220>	
<221>misc_feature	
<222>(460).. (460)	
<223>460 位上的 ' Xaa' 代表 Glu、Gly、Ala 或 Val	
<220>	
<221>misc_feature	
<222>(1478).. (1478)	
<223>n = 未知核苷酸	

<400>8

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

Asp Tyr Phe Asp Arg Arg Arg Glu Glu Leu Asn Ile Ala Phe Thr Phe			
275	280	285	
Asp Leu Ile Arg Leu Asp Arg Tyr Pro Asp Gln Arg Trp Arg Arg Lys			
290	295	300	
Pro Trp Thr Leu Ser Gln Phe Arg Gln Val Ile Ser Gln Thr Asp Arg			
305	310	315	
Ala Ala Gly Glu Phe Gly Trp Asn Ala Phe Phe Leu Asp Asn His Asp			
320	325	330	335
Asn Pro Arg Gln Val Ser His Phe Gly Asp Asp Ser Pro Gln Trp Arg			
340	345	350	
Glu Arg Ser Ala Lys Ala Leu Ala Thr Leu Leu Leu Thr Gln Arg Ala			
355	360	365	
Thr Pro Phe Ile Phe Gln Gly Ala Glu Leu Gly Met Thr Asn Tyr Pro			
370	375	380	
Phe Lys Asn Ile Glu Glu Phe Asp Asp Ile Glu Val Lys Gly Phe Trp			
385	390	395	
Asn Asp Tyr Val Ala Ser Gly Lys Val Asn Ala Ala Glu Phe Leu Gln			
400	405	410	415
Glu Val Arg Met Thr Ser Arg Asp Asn Ser Arg Thr Pro Met Gln Trp			
420	425	430	
Asn Asp Ser Val Asn Ala Gly Phe Thr Gln Gly Lys Pro Trp Phe His			
435	440	445	
Leu Asn Pro Asn Tyr Lys Gln Ile Asn Ala Ala Arg Xaa Val Asn Lys			
450	455	460	
Pro Asp Ser Val Phe Ser Tyr Tyr Arg Gln Leu Ile Asn Leu Arg His			
465	470	475	
Gln Ile Pro Ala Leu Thr Ser Gly Glu Tyr Arg Asp Leu Asp Pro Gln			
480	485	490	495
Asn Asn Gln Val Tyr Ala Tyr Thr Arg Ile Leu Asp Asn Glu Lys Tyr			
500	505	510	
Leu Val Val Val Asn Phe Lys Pro Glu Gln Leu His Tyr Ala Leu Pro			
515	520	525	
Asp Asn Leu Thr Ile Ala Ser Ser Leu Leu Glu Asn Val His Gln Pro			
530	535	540	
Ser Leu Gln Glu Asn Ala Ser Thr Leu Thr Leu Ala Pro Trp Gln Ala			
545	550	555	
Gly Ile Tyr Lys Leu Asn			
560	565		
<210>9			

<211>1698

<212>DNA

<213> 细菌分离物 68J

<220>

<221>CDS

<222>(1)..(1695)

<223>

<220>

<221>misc_feature

<222>(1379)..(1379)

<223>n = 未知核苷酸

<220>

<221> 成熟肽

<222>(1)..(1698)

<223>

<400>9

gca acg aat ata caa aag tcc gct gat ttt ccc att tgg tgg aaa cag 48

Ala Thr Asn Ile Gln Lys Ser Ala Asp Phe Pro Ile Trp Trp Lys Gln

1 5 10 15

gca gta ttt tac cag att tat ccc cgc tca ttt aaa gat agc aat ggt 96

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

20 25 30

gat ggt atc ggc gat att ccc ggt atc att gag aaa ctg gac tat tta 144

Asp Gly Ile Gly Asp Ile Pro Gly Ile Ile Glu Lys Leu Asp Tyr Leu

35 40 45

aaa atg ctg gga gtt gat gct atc tgg ata aac ccg cac tat gag tct 192

Lys Met Leu Gly Val Asp Ala Ile Trp Ile Asn Pro His Tyr Glu Ser

50 55 60

cct aac acc gac aat ggt tac gat att agt gat tat cgt aaa atc atg 240

Pro Asn Thr Asp Asn Gly Tyr Asp Ile Ser Asp Tyr Arg Lys Ile Met

65 70 75 80

aag gag tac ggc agc atg gct gac ttt gac cgt ctg gtt gcc gaa atg 288

Lys Glu Tyr Gly Ser Met Ala Asp Phe Asp Arg Leu Val Ala Glu Met

85 90 95

aat aaa cgt ggt atg cgc ctg atg att gat att gtt atc aat cat acc 336

Asn Lys Arg Gly Met Arg Leu Met Ile Asp Ile Val Ile Asn His Thr

100 105 110

agc gat cgt cac cgc tgg ttt gtg cag agc cgt tca ggt aaa gat aat 384

Ser Asp Arg His Arg Trp Phe Val Gln Ser Arg Ser Gly Lys Asp Asn

115	120	125	
cct tac cgc gac tat tat ttc tgg cgt gat ggt aaa cag gga cag gct			432
Pro Tyr Arg Asp Tyr Tyr Phe Trp Arg Asp Gly Lys Gln Gly Gln Ala			
130	135	140	
ccc aat aac tat ccc tct ttc ttt ggc ggt tca gcc tgg caa ctg gat			480
Pro Asn Asn Tyr Pro Ser Phe Phe Gly Gly Ser Ala Trp Gln Leu Asp			
145	150	155	160
aaa cag act gac cag tat tat ctg cac tat ttt gca cca cag cag ccg			528
Lys Gln Thr Asp Gln Tyr Tyr Leu His Tyr Phe Ala Pro Gln Gln Pro			
165	170	175	
gat ctg aac tgg gat aac cca aaa gtt cgg gct gaa ctc tac gat att			576
Asp Leu Asn Trp Asp Asn Pro Lys Val Arg Ala Glu Leu Tyr Asp Ile			
180	185	190	
ctg cgt ttc tgg ctg gat aaa ggc gta tcc gga cta cgt ttt gat acc			624
Leu Arg Phe Trp Leu Asp Lys Gly Val Ser Gly Leu Arg Phe Asp Thr			
195	200	205	
gtg gct act ttc tcc aaa att cct ggc ttc ccg gac ctg tca aaa gcg			672
Val Ala Thr Phe Ser Lys Ile Pro Gly Phe Pro Asp Leu Ser Lys Ala			
210	215	220	
cag ctg aag aat ttt gcc gaa gct tat act gag ggg ccg aat att cat			720
Gln Leu Lys Asn Phe Ala Glu Ala Tyr Thr Glu Gly Pro Asn Ile His			
225	230	235	240
aaa tat atc cat gaa atg aac cgc cag gta ctg tct aaa tat aat gtt			768
Lys Tyr Ile His Glu Met Asn Arg Gln Val Leu Ser Lys Tyr Asn Val			
245	250	255	
gcc acc gct ggt gaa atc ttc ggt gtg cca gtg agt gct atg ccg gat			816
Ala Thr Ala Gly Glu Ile Phe Gly Val Pro Val Ser Ala Met Pro Asp			
260	265	270	
tat ttt gac cgg cgg cgt gaa gaa ctc aat att gct ttc acc ttt gat			864
Tyr Phe Asp Arg Arg Arg Glu Glu Leu Asn Ile Ala Phe Thr Phe Asp			
275	280	285	
ttg atc agg ctc gat cgt tat ccc gat cag cgc tgg cgt cgt aaa cca			912
Leu Ile Arg Leu Asp Arg Tyr Pro Asp Gln Arg Trp Arg Arg Lys Pro			
290	295	300	
tgg aca tta agc cag ttt cgt caa gtt atc tct cag act gac cgt gcc			960
Trp Thr Leu Ser Gln Phe Arg Gln Val Ile Ser Gln Thr Asp Arg Ala			
305	310	315	320
gcc ggt gaa ttt ggc tgg aac gcc ttt ttc ctt gat aac cat gat aac			1008
Ala Gly Glu Phe Gly Trp Asn Ala Phe Phe Leu Asp Asn His Asp Asn			

325	330	335	
ccg cgc cag gtc tca cac ttt ggt gac gac agc cca caa tgg cgc gaa			1056
Pro Arg Gln Val Ser His Phe Gly Asp Asp Ser Pro Gln Trp Arg Glu			
340	345	350	
cgc tcg gca aaa gca ctg gca acg ctg ctg ctg acg cag cgt gcc acg			1104
Arg Ser Ala Lys Ala Leu Ala Thr Leu Leu Leu Thr Gln Arg Ala Thr			
355	360	365	
ccg ttt atc ttt cag ggg gcg gag ttg gga atg act aat tac ccc ttt			1152
Pro Phe Ile Phe Gln Gly Ala Glu Leu Gly Met Thr Asn Tyr Pro Phe			
370	375	380	
aaa aat ata gag gaa ttt gat gat att gag gtt aaa ggc ttc tgg aac			1200
Lys Asn Ile Glu Glu Phe Asp Asp Ile Glu Val Lys Gly Phe Trp Asn			
385	390	395	400
gac tat gta gcc agc gga aaa gta aac gct gct gaa ttt tta cag gag			1248
Asp Tyr Val Ala Ser Gly Lys Val Asn Ala Ala Glu Phe Leu Gln Glu			
405	410	415	
gtt cgc atg acc agc cgc gat aac agc cga aca cca atg cag tgg aac			1296
Val Arg Met Thr Ser Arg Asp Asn Ser Arg Thr Pro Met Gln Trp Asn			
420	425	430	
gac tct gtt aat gcc gga ttc acc cag ggc aaa ccc tgg ttt cac ctc			1344
Asp Ser Val Asn Ala Gly Phe Thr Gln Gly Lys Pro Trp Phe His Leu			
435	440	445	
aat ccc aac tat aag caa atc aat gcc gcc agg gng gtg aat aaa ccc			1392
Asn Pro Asn Tyr Lys Gln Ile Asn Ala Ala Arg Xaa Val Asn Lys Pro			
450	455	460	
gac tcg gta ttc agt tac tac cgt caa ctg atc aac ctg cgt cac cag			1440
Asp Ser Val Phe Ser Tyr Tyr Arg Gln Leu Ile Asn Leu Arg His Gln			
465	470	475	480
atc ccg gca ctg acc agt ggt gaa tac cgt gat ctc gat ccg cag aat			1488
Ile Pro Ala Leu Thr Ser Gly Glu Tyr Arg Asp Leu Asp Pro Gln Asn			
485	490	495	
aac cag gtc tat gcc tat acc cgt ata ctg gat aat gaa aaa tat ctg			1536
Asn Gln Val Tyr Ala Tyr Thr Arg Ile Leu Asp Asn Glu Lys Tyr Leu			
500	505	510	
gtg gta gtt aat ttt aaa cct gag cag ctg cat tac gct ctg cca gat			1584
Val Val Val Asn Phe Lys Pro Glu Gln Leu His Tyr Ala Leu Pro Asp			
515	520	525	
aat ctg act att gcc agc agt ctg ctg gaa aat gtc cac caa cca tca			1632
Asn Leu Thr Ile Ala Ser Ser Leu Leu Glu Asn Val His Gln Pro Ser			

530	535	540	
ctg caa gaa aat gcc tcc acg ctg act ctt gct ccg tgg caa gcc ggg			1680
Leu Gln Glu Asn Ala Ser Thr Leu Thr Leu Ala Pro Trp Gln Ala Gly			
545	550	555	560
atc tat aag ctg aac tga			1698
Ile Tyr Lys Leu Asn			
565			
<210>10			
<211>565			
<212>PRT			
<213> 细菌分离物 68J			
<220>			
<221>misc_feature			
<222>(460).. (460)			
<223>460 位上的' Xaa' 代表 Glu、Gly、Ala 或 Val			
<220>			
<221>misc_feature			
<222>(1379).. (1379)			
<223>n = 未知核苷酸			
<400>10			
Ala Thr Asn Ile Gln Lys Ser Ala Asp Phe Pro Ile Trp Trp Lys Gln			
1 5 10 15			
Ala Val Phe Tyr Gln Ile Tyr Pro Arg Ser Phe Lys Asp Ser Asn Gly			
20 25 30			
Asp Gly Ile Gly Asp Ile Pro Gly Ile Ile Glu Lys Leu Asp Tyr Leu			
35 40 45			
Lys Met Leu Gly Val Asp Ala Ile Trp Ile Asn Pro His Tyr Glu Ser			
50 55 60			
Pro Asn Thr Asp Asn Gly Tyr Asp Ile Ser Asp Tyr Arg Lys Ile Met			
65 70 75 80			
Lys Glu Tyr Gly Ser Met Ala Asp Phe Asp Arg Leu Val Ala Glu Met			
85 90 95			
Asn Lys Arg Gly Met Arg Leu Met Ile Asp Ile Val Ile Asn His Thr			
100 105 110			
Ser Asp Arg His Arg Trp Phe Val Gln Ser Arg Ser Gly Lys Asp Asn			
115 120 125			
Pro Tyr Arg Asp Tyr Tyr Phe Trp Arg Asp Gly Lys Gln Gly Gln Ala			
130 135 140			
Pro Asn Asn Tyr Pro Ser Phe Phe Gly Gly Ser Ala Trp Gln Leu Asp			

145	150	155	160
Lys Gln Thr Asp Gln Tyr Tyr Leu His Tyr Phe Ala Pro Gln Gln Pro			
165	170	175	
Asp Leu Asn Trp Asp Asn Pro Lys Val Arg Ala Glu Leu Tyr Asp Ile			
180	185	190	
Leu Arg Phe Trp Leu Asp Lys Gly Val Ser Gly Leu Arg Phe Asp Thr			
195	200	205	
Val Ala Thr Phe Ser Lys Ile Pro Gly Phe Pro Asp Leu Ser Lys Ala			
210	215	220	
Gln Leu Lys Asn Phe Ala Glu Ala Tyr Thr Glu Gly Pro Asn Ile His			
225	230	235	240
Lys Tyr Ile His Glu Met Asn Arg Gln Val Leu Ser Lys Tyr Asn Val			
245	250	255	
Ala Thr Ala Gly Glu Ile Phe Gly Val Pro Val Ser Ala Met Pro Asp			
260	265	270	
Tyr Phe Asp Arg Arg Arg Glu Glu Leu Asn Ile Ala Phe Thr Phe Asp			
275	280	285	
Leu Ile Arg Leu Asp Arg Tyr Pro Asp Gln Arg Trp Arg Arg Lys Pro			
290	295	300	
Trp Thr Leu Ser Gln Phe Arg Gln Val Ile Ser Gln Thr Asp Arg Ala			
305	310	315	320
Ala Gly Glu Phe Gly Trp Asn Ala Phe Phe Leu Asp Asn His Asp Asn			
325	330	335	
Pro Arg Gln Val Ser His Phe Gly Asp Asp Ser Pro Gln Trp Arg Glu			
340	345	350	
Arg Ser Ala Lys Ala Leu Ala Thr Leu Leu Leu Thr Gln Arg Ala Thr			
355	360	365	
Pro Phe Ile Phe Gln Gly Ala Glu Leu Gly Met Thr Asn Tyr Pro Phe			
370	375	380	
Lys Asn Ile Glu Glu Phe Asp Asp Ile Glu Val Lys Gly Phe Trp Asn			
385	390	395	400
Asp Tyr Val Ala Ser Gly Lys Val Asn Ala Ala Glu Phe Leu Gln Glu			
405	410	415	
Val Arg Met Thr Ser Arg Asp Asn Ser Arg Thr Pro Met Gln Trp Asn			
420	425	430	
Asp Ser Val Asn Ala Gly Phe Thr Gln Gly Lys Pro Trp Phe His Leu			
435	440	445	
Asn Pro Asn Tyr Lys Gln Ile Asn Ala Ala Arg Xaa Val Asn Lys Pro			
450	455	460	

〈213〉 细菌分离物 68.J

<400>12

Met Phe Leu Asn Gly Phe Lys Thr Val Ile Ala Leu Thr Met Ala Ser

1

5

10

15

Ser Phe Tyr Leu Ala Ala Ser Pro Leu Thr Lys Pro Ser Thr Pro Ile

20

25

30

Ala

<210>13

<211>34

<212>DNA

<213> 合成序列

<220>

<223> 合成序列的描述 :用于 68J 分离物扩增的 5' 寡核苷酸引物

<400>13

ggatccaaca atggcaacga atatacaaaa gtcc

34

<210>14

<211>30

<212>DNA

<213> 合成序列

<220>

<223> 合成序列的描述 :用于 68J 分离物扩增的 3' 寡核苷酸引物

<400>14

ataggtacct cagttcagct tatagatccc

30

<210>15

<211>35

<212>DNA

<213> 合成序列

<220>

<223> 合成序列的描述 :用于大黄欧文氏菌 (登记号 WAC2928) 扩增的 5' 寡核苷酸引

物

<400>15

ggatccaaca atggcaaccg ttcagcaatc aaatg

35

<210>16

<211>28

<212>DNA

<213> 合成序列

<220>

<223> 合成序列的描述 :用于大黄欧文氏菌 (登记号 WAC2928) 扩增的 3' 寡核苷酸引

物

<400>16

ataggtacct tacttaaacg cgtggatg	28
<210>17	
<211>35	
<212>DNA	
<213> 合成序列	
<220>	
<223> 合成序列的描述 :用于 14S 分离物扩增的 5' 寡核苷酸引物	
<400>17	
ggatccaaca atggcaaccg ttcacaagga aagtg	35
<210>18	
<211>30<212>DNA	
<213> 合成序列	
<220>	
<223> 合成序列的描述 :用于 14S 分离物扩增的 3' 寡核苷酸引物	
<400>18	
ataggtacct taccgcagct tatacacacc	30
<210>19	
<211>7	
<212>PRT	
<213> 合成序列	
<220>	
<223> 合成序列的描述 :蔗糖异构酶共有序列	
<400>19	
Asp Leu Ile Arg Leu Asp Arg	
1 5	
<210>20	
<211>10	
<212>PRT	
<213> 合成序列	
<220>	
<223> 合成序列的描述 :蔗糖异构酶共有序列	
<220>	
<221>MISC_FEATURE	
<222>(7).. (7)	
<223>X =任何氨基酸	
<400>20	
Glu Val Lys Gly Phe Trp Xaa Asp Tyr Val	
1 5 10	
<210>21	

<211>6

<212>PRT

<213> 合成序列

<220>

<223> 合成序列的描述 :蔗糖异构酶共有序列

<400>21

Arg Pro Gln Trp Arg Glu

1 5

<210>22

<211>6

<212>PRT

<213> 合成序列

<220>

<223> 合成序列的描述 :蔗糖异构酶共有序列

<400>22

Ser Pro Gln Trp Arg Glu

1 5

<210>23

<211>13

<212>PRT

<213> 合成序列

<220>

<223> 合成序列的描述 :蔗糖异构酶共有序列

<400>23

Pro Asn Asn Tyr Pro Ser Phe Phe Gly Gly Ser Ala Trp

1 5 10

<210>24

<211>16

<212>PRT

<213> 合成序列

<220>

<223> 合成序列的描述 :蔗糖异构酶共有序列

<220>

<221>MISC_FEATURE

<222>(8)..(8)

<223>X =任何氨基酸

<220>

<221>MISC_FEATURE

<222>(9)..(9)

<223>X =任何氨基酸

<400>24

Gln Tyr Tyr Leu His Tyr Phe Xaa Xaa Gln Gln Pro Asp Leu Asn Trp
1 5 10 15

<210>25

<211>594

<212>DNA

<213> 大黄欧文氏菌

<220>

<221>CDS

<222>(1)..(594)

<223>

<220>

<221>misc_feature

<222>(42)..(42)

<223>n =未知核苷酸

<400>25

tac gtt gaa aca ggc aaa gtg aaa gct gag gaa ttc ctt can aac gta 48

Tyr Val Glu Thr Gly Lys Val Lys Ala Glu Glu Phe Leu Thr Asn Val
1 5 10 15

cgc caa acc agc cgt gat aac agc aga acc ccc ttc cag tgg gat gca 96

Arg Gln Thr Ser Arg Asp Asn Ser Arg Thr Pro Phe Gln Trp Asp Ala
20 25 30

agc aaa aat gcg ggc ttt acc agc gga acc cct tgg tta aaa atc aat 144

Ser Lys Asn Ala Gly Phe Thr Ser Gly Thr Pro Trp Leu Lys Ile Asn
35 40 45

ccc aat tat aaa gaa atc aac agc gca gat cag att aac aat cca aat 192

Pro Asn Tyr Lys Glu Ile Asn Ser Ala Asp Gln Ile Asn Asn Pro Asn
50 55 60

tcc gta ttt aac tat tat aga aag ctc att aac att cgc cac gac atc 240

Ser Val Phe Asn Tyr Tyr Arg Lys Leu Ile Asn Ile Arg His Asp Ile
65 70 75 80

cct gcc tta acc tac ggc agt tat att gat tta gct cct gac aac aat 288

Pro Ala Leu Thr Tyr Gly Ser Tyr Ile Asp Leu Ala Pro Asp Asn Asn
85 90 95

tca gtc tat gct tac act cga acg ttt ggc gct gaa aaa tat ctt gtg 336

Ser Val Tyr Ala Tyr Thr Arg Thr Phe Gly Ala Glu Lys Tyr Leu Val
100 105 110

gtc att aat ttt aaa gaa gaa gtg atg cac tac acc ctg cct ggg gat 384

Val Ile Asn Phe Lys Glu Glu Val Met His Tyr Thr Leu Pro Gly Asp	
115 120 125	
tta tcc atc aat aag gtg att act gaa aac aac agt cac act att gtg	432
Leu Ser Ile Asn Lys Val Ile Thr Glu Asn Asn Ser His Thr Ile Val	
130 135 140	
aat aaa aat gac gta gaa gat cct cgt ggg gct aca agc gtt tgt agc	480
Asn Lys Asn Asp Val Glu Asp Pro Arg Gly Ala Thr Ser Val Cys Ser	
145 150 155 160	
ccc ttc cag gct caa aaa agg cct ggc gac ccg ggt tac tct gct gcc	528
Pro Phe Gln Ala Gln Lys Arg Pro Gly Asp Pro Gly Tyr Ser Ala Ala	
165 170 175	
cat tcg att cgg ttc ttg ccc cgg ttt ttc gct tca tac agg ggc gac	576
His Ser Ile Arg Phe Leu Pro Arg Phe Phe Ala Ser Tyr Arg Gly Asp	
180 185 190	
atc cac gcg ttt aag taa	594
Ile His Ala Phe Lys	

195

<210>26

<211>197

<212>PRT

<213> 大黄欧文氏菌

<220>

<221>misc_feature

<222>(42).. (42)

<223>n = 未知核苷酸

<400>26

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

100	105	110
Val Ile Asn Phe Lys Glu Glu Val Met His Tyr Thr Leu Pro Gly Asp		
115	120	125
Leu Ser Ile Asn Lys Val Ile Thr Glu Asn Asn Ser His Thr Ile Val		
130	135	140
Asn Lys Asn Asp Val Glu Asp Pro Arg Gly Ala Thr Ser Val Cys Ser		
145	150	155
Pro Phe Gln Ala Gln Lys Arg Pro Gly Asp Pro Gly Tyr Ser Ala Ala		
165	170	175
His Ser Ile Arg Phe Leu Pro Arg Phe Phe Ala Ser Tyr Arg Gly Asp		
180	185	190
Ile His Ala Phe Lys		
195		

<210>27

<211>21

<212>DNA

<213> 大黄欧文氏菌

<400>27

gatctgatca gactcgatcgt

21

<210>28

<211>30

<212>DNA

<213> 大黄欧文氏菌

<400>28

gaggtgaaag gtttttggca agactacgtt

30

<210>29

<211>18

<212>DNA

<213> 大黄欧文氏菌

<400>29

cgaccacaat ggcgcgag

18

<210>30

<211>39

<212>DNA

<213> 大黄欧文氏菌

<400>30

cccaataact atccctcctt ctteggtggc tcagcctgg

39

<210>31

<211>48

<212>DNA	
<213> 大黄欧文氏菌	
<400>31	
cagtattacc tccattactt tgccaaacag caacccgacc tcaactgg	48
<210>32	
<211>21	
<212>DNA	
<213> 细菌分离物 68J	
<400>32	
gatttgatca ggctcgatcg t	21
<210>33	
<211>30	
<212>DNA	
<213> 细菌分离物 68J	
<400>33	
gaggttaaag gcttctggaa cgactatgta	30
<210>34	
<211>18	
<212>DNA	
<213> 细菌分离物 68J	
<400>34	
agcccacaat ggcgcgaa	18
<210>35	
<211>39	
<212>DNA	
<213> 细菌分离物 68J	
<400>35	
cccaataact atccctcttt ctttggcggt tcagcctgg	39
<210>36	
<211>48	
<212>DNA	
<213> 细菌分离物 68J	
<400>36	
cagtattatc tgcactatit tgcaccacag cagccggatc tgaactgg	48
<210>37	
<211>8	
<212>PRT	
<213> 合成的	
<220>	

<221>MOD_RES
<222>(1)..(1)
<223>乙酰化
<220>
<221>MISC_FEATURE
<222>(1)..(1)
<223>X =任何氨基酸
<220>
<221>MISC_FEATURE
<222>(2)..(2)
<223>X =任何氨基酸
<220>
<221>MISC_FEATURE
<222>(3)..(3)
<223>X =限定的氨基酸
<220>
<221>MISC_FEATURE
<222>(4)..(4)
<223>X =限定的氨基酸
<220>
<221>MISC_FEATURE
<222>(5)..(5)
<223>X =任何氨基酸
<220>
<221>MISC_FEATURE
<222>(6)..(6)
<223>X =任何氨基酸
<220>
<221>MISC_FEATURE
<222>(7)..(7)
<223>X =任何氨基酸
<220>
<221>MISC_FEATURE
<222>(8)..(8)
<223>X =任何氨基酸
<400>37
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
1 5
<210>38

<211>17	
<212>DNA	
<213> 合成的	
<400>38	
tggtggaarg argctgt	17
<210>39	
<211>19	
<212>DNA	
<213> 合成的	
<400>39	
tcccagtttag rtccggctg	19

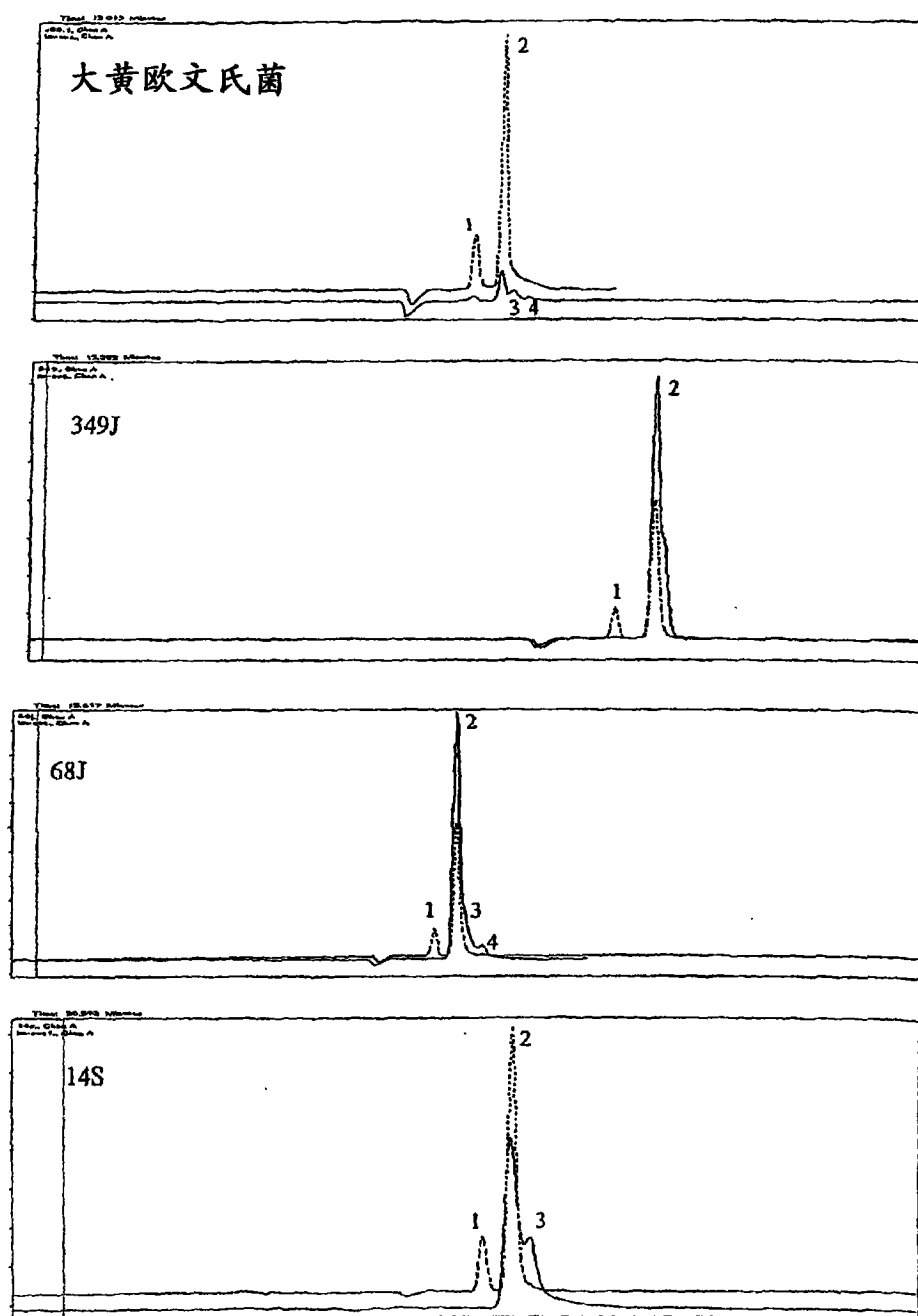


图 1

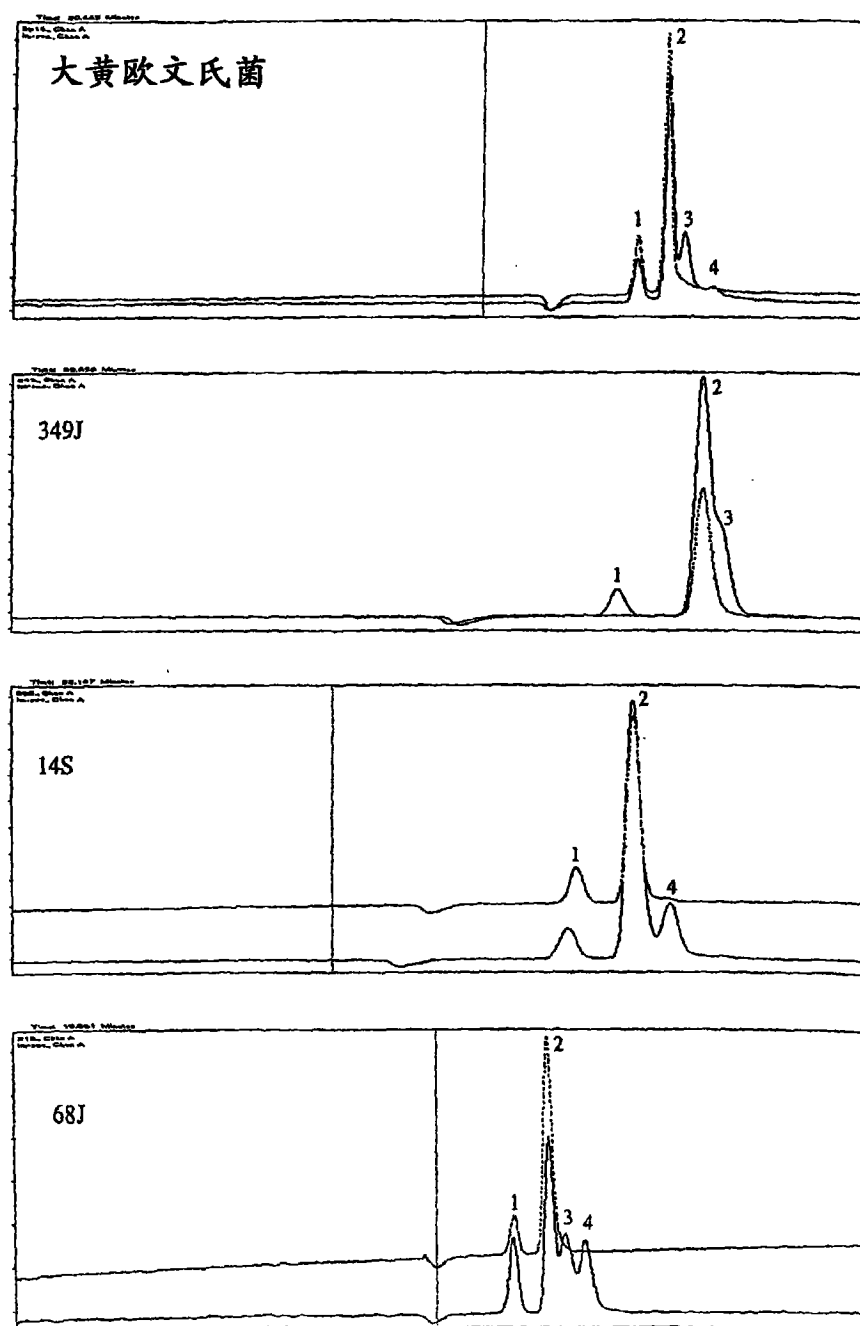


图 2

```

1 ATGTCCTCTC AAGAATTGAA AGCGGCTGTC GCTATTTTTC TTGCAACCAC 50
51 TTTTCTGCC ACATCCTATC AGGCCTGCAG TGCCGGGCCA GATACCGCCC 100
101 CCTCACTCAC CGTTCAGCAA TCAAATGCCC TGCCACATG GTGGAAGCAG 150
151 GCTGTTT TTT ATCAGGTATA TCCACGCTCA TTTAAAGATA CGAATGGGGA 200
201 TGGCATTGGG GATTTAAACG GTATTATTGA GAATTTAGAC TATCTGAAGA 250
251 AACTGGGTAT TGATGCGATT TGGATCAATC CACATTACGA TTCGCCGAAT 300
301 ACGGATAATG GTTATGACAT CCGGATTAC CGTAAGATAA TGAAAGAATA 350
351 CGGTACGATG GAAGACTTTG ACCGTCTTAT TTCAGAAATG AAGAAACGCA 400
401 ATATGCGTTT GATGATTGAT ATTGTATCA ACCACACCAG CGATCAGCAT 450
451 GCGTGGTTTG TTCAGAGCAA ATCGGGTAAG AACAAACCCCT ACAGGGACTA 500
501 TTACTTCTGG CGTGACGGTA AGGATGGCCA TGCCCCCAAT AACTATCCCT 550
551 CCTTCTTCGG TGGCTCAGCC TGGGAAAAAG ACGATAAATC AGGCCAGTAT 600
601 TACCTCCATT ACTTTGCCAA ACAGCAACCC GACCTCAACT GGGACAATCC 650
651 CAAAGTCCGT CAAGACCTGT ATGACATGCT CCGCTTCTGG TTAGATAAAG 700
701 GCGTTTNTGG TTTACGCTTT GATACCGTTG CCACCTATTC AAAAATCCCG 750
751 AACTTCCCTG ACCTTAGCCA ACAGCAGTTA AAAAATTTCTG CCGAGGAATA 800
801 TACTAAAGGT CCTAAATTC ACGACTACGT GAATGAAATG AACAGAGAAG 850
851 TATTATCCCA CTATGATATC GCCACTGCGG GGGAAATATT TGGGGTTCCT 900
901 CTGGATAAAT CGATTAAAGTT TTTGATCGC CGTAGAAATG AATTAAATAT 950
951 AGCGTTTACG TTTGATCTGA TCAGACTCGA TCGTGATGCT GATGAAAGAT 1000
1001 GCGGCGGAAA AGACTGGACC CTTTCGCACT TCCGAAAAAT TGTGATAAG 1050
1051 GTTGACCAA CCGCAGGAGA GTATGGGTGG AATGCCTTTT TCTTAGACAA 1100
1101 TCACGACAAT CCCCAGCGCG TTTCTCACTT TGGTGATGAT CGACCACAAT 1150
1151 GCGCGGAGCA TGCGGCGAAA GCACTGGCAA CATTGACGCT GACCCAGCGT 1200
1201 GCAACGCCGT TTATCTATCA GGGTTCAGAA CTCGGTATGA CCAATTATCC 1250
1251 CTTTAAAAAA ATCGATGATT TCGATGATGT AGAGGTGAAA GGTTTTTGCG 1300
1301 AAGACTACGT TGAAACAGGC AAAGTGAAAG CTGAGGAATT CCTTCANAAC 1350
1351 GTACGCCAAA CCAGCCGTGA TAACAGCAGA ACCCCCTTCC AGTGGGATGC 1400
1401 AAGCAAAAAT GCGGGCTTTA CCAGCGGAAC CCCTTGTTA AAAATCAATC 1450
1451 CCAATTATAA AGAAATCAAC AGCGCAGATC AGATTAAACA TCCAAATTCC 1500
1501 GTATTTAACT ATTATAGAAA GCTCATTAC ATTGCCCACG ACATCCCTGC 1550
1551 CTTAACCTAC GGCAGTTATA TTGATTTAGC TCCTGACAAC AATTCAGTCT 1600
1601 ATGCTTACAC TCGAACGTTT GCGCTGAAA AATATCTTGT GGTCATTAA 1650
1651 TTTAAGAAG AAGTGATGCA CTACACCCTG CCTGGGGATT TATCCATCAA 1700
1701 TAAGGTGATT ACTGAAAACA ACAGTCACAC TATTGTGAAT AAAAATGACG 1750
1751 TAGAAGATCC TCGTGGGGCT ACAAGCGTTT GTAGCCCCCT CCAGGCTCAA 1800
1801 AAAAGGCCTG GCGACCCGGG TTAATCTGCT GCCCATTCGA TTCGGTTCTT 1850
1851 GCCCCGGTTT TCGCTTCAT ACAGGGGCGA CATCCACGCG TTTAAGTAA 1899

```

图 3

```

1  ATGTTTCTTA ATGGATTAA GACAGTTATT GCTCTGACTA TGGCAAGCTC 50
51  GTTTTATCTT GCCGCCAGCC CGTTAACTAA GCCATCGACC CCTATTGCCG 100
101 CAACGAATAT AAAAAAGTCC GCTGATTTTC CCATTGGTG GAAACAGGCA 150
151 GTATTTTACC AGATTTATCC CCGCTCATT AAAGATAGCA ATGGTGATGG 200
201 TATCGGCGAT ATTCCCGGTA TCATTGAGAA ACTGGACTAT TTAATAATGC 250
251 TGGGAGTTGA TGCTATCTGG ATAAACCCGC ACTATGAGTC TCCTAACACC 300
301 GACAATGGTT ACGATATTAG TGATTATCGT AAAATCATGA AGGAGTACGG 350
351 CAGCATGGCT GACTTTGACC GTCTGGTTGC CGAAATGAAT AAACGTGGTA 400
401 TCGCCTGAT GATTGATATT GTTATCAATC ATACCAGCGA TCGTCACCGC 450
451 TGGTTTGTGC AGAGCCGTT AGGTAAAGAT AATCCTTACC GCGACTATTA 500
501 TTTCTGGCGT GATGGTAAAC AGGGACAGGC TCCCAATAAC TATCCCTCTT 550
551 TCTTTGGCGG TTCAGCCTGG CAACTGGATA AACAGACTGA CCAGTATTAT 600
601 CTGCATATT TTGCACCACA GCAGCCGGAT CTGAACGGG ATAACCCAAA 650
651 AGTTCCGGCT GAACTCTACG ATATTCTGCG TTTCTGGCTG GATAAAGGCG 700
701 TATCCGGACT ACGTTTGTG ACCGTGGCTA CTTTCTCCAA AATTCTGGC 750
751 TTCCCGGACC TGTCAAAAGC GCAGCTGAAG AATTTTGCCG AAGCTTATAC 800
801 TGAGGGGCCG AATATTCTA AATATATCCA TGAAATGAAC CGCCAGGTAC 850
851 TGTCTAAATA TAATGTTGCC ACCGCTGGTG AAATCTTCGG TGTGCCAGTG 900
901 AGTGCTATGC CGGATTATTT TGACCGGCGG CGTGAAGAAC TCAATATTGC 950
951 TTTACCTTT GATTGATCA GGCTCGATCG TTATCCCGAT CAGCGCTGGC 1000
1001 GTCGTAAACC ATGGACATTA AGCCAGTTTC GTCAAGTTAT CTCTCAGACT 1050
1051 GACCGTGCCG CCGGTGAATT TGGCTGGAAC GCCTTTTTC TTGATAACCA 1100
1101 TGATAACCCG CGCCAGGTCT CACACTTTGG TGACGACAGC CCACAATGGC 1150
1151 GCGAACGCTC GGCAAAAGCA CTGGCAACGC TGCTGCTGAC GCAGCGTGCC 1200
1201 ACGCCGTTTA TCTTTCAGGG GGCGGAGTTG GGAATGACTA ATTACCCCTT 1250
1251 TAAAAATATA GAGGAATTTG ATGATATTGA GGTAAAGGC TTCTGGAACG 1300
1301 ACTATGTAGC CAGCGGAAAA GTAAACGCTG CTGAATTTTT ACAGGAGGTT 1350
1351 CGCATGACCA GCCGCGATAA CAGCCGAACA CCAATGCAGT GGAACGACTC 1400
1401 TGTTAATGCC GGATTACCC AGGGCAAACC CTGGTTTCAC CTCAATCCCA 1450
1451 ACTATAAGCA AATCAATGCC GCCAGGNGG TGAATAAACC CGACTCGGTA 1500
1501 TTCAGTACT ACCGTCAACT GATCAACCTG CGTCACCAGA TCCCGGCACT 1550
1551 GACCAGTGGT GAATACCGTG ATCTCGATCC GCAGAATAAC CAGGTCTATG 1600
1601 CCTATACCCG TATACTGGAT AATGAAAAAT ATCTGGTGGT AGTTAATTTT 1650
1651 AAACCTGAGC AGCTGCATTA CGCTCTGCCA GATAATCTGA CTATTGCCAG 1700
1701 CAGTCTGCTG GAAAATGTCC ACCAACCATC ACTGCAAGAA AATGCCTCCA 1750
1751 CGCTGACTCT TGCTCCGTGG CAAGCCGGA TCTATAAGCT GAACTGA 1797

```

图 4

```

1 MSSQELKAAV AIFLATTFSA TSYQACSAGP DTAPSLTVQQ SNALPTWWKQ 50
51 AVFYQVYPRS FKDTNGDGIG DLNGIENLD YLKKLGIDAI WINPHYDSPN 100
101 TDNGYDIRDY RKIMKEYGTM EDFDRLISEM KKRNMRLMID TVINHTSDQH 150
151 AWFVQSKSGK NNPYRDYYFW RDGKDGHAPN NYPSFFGGSA WEKDDKSGQY 200
201 YLHYFAKQQP DLNWDNPKVR QDLYDMLRFW LDKGVXGLRF DTVATYSKIP 250
251 NFPDLSQQQL KNFAEYTKG PKIHDEVNEM NREVLSHYDI ATAGEIFGVP 300
301 LDKSIKFFDR RRNELNIAFT FDLIRLDRDA DERWRRKDWT LSQFRKIVDK 350
351 VDQTAGEYGW NAFFLDNHDN PRAVSHFGDD RPQWREHAAK ALATLTLTQR 400
401 ATPFIYQGSE LGMTNYPFKK IDDFDDVEVK GFWQDYVETG KVKAEFLXN 450
451 VRQTSRDNRS TPFQWDASKN AGFTSGTPWL KINPNYKEIN SADQINNPN 500
501 VFNYRKLIN IRHDIPALTY GSYIDLAPDN NSVYAYTRTF GAEKYLVVIN 550
551 FKEEVMHYTL PGDLSINKVI TENNSHTIVN KNDVEDPRGA TSVCSPPQAA 600
601 KRPGDPGYSA AHSIRFLPRF FASYRGDIHA FK* 632

```

图 5

```

1 MFLNGFKTVI ALTMASFFYL AASPLTKPST PIAATNIQKS ADFPIWWKQA 50
51 VFYQIYPRSF KOSNGDGIGD IPGIIEKLDY LKMLGVDAIW INPHYESPNT 100
101 DNGYDISDYR KIMKEYGSMA DFDRLVAEMN KRGMRMLIDI VINHTSDRHR 150
151 WFWQSRSGKD NPYRDYYFWR DGKQGQAPNN YPSFFGGSAW QLDKQTDQYY 200
201 LHYFAPQQPD LNWDNPKVRA ELYDILRFLW DKGVSGLRFD TVATFSKIPG 250
251 FPDLSKAQLK NFAYEYTEGP NIHKYIHEMN RQVLSKYNVA TAGEIFGVPV 300
301 SAMPDYFDRR REELNIAFTF DLIRLDRYPD QRWRRKPWTL SQFRQVISQT 350
351 DRAAGEFGWN AFFLDNHDNP RQVSHFGDDS PQWRERSAKA LATLLLTQRA 400
401 TPFIFQGAEL GMTNYPFKNI EEFDDIEVKG FWN DYVASGK VNAAEFLQEV 450
451 RMTSRDNST PMQWNSVNA GFTQGKPFWH LNPYKQINA ARXVKNPDSV 500
501 FSYYRQLINL RHQIPALTSG EYRDLDPQNN QVYAYTRILD NEKYLVVVNF 550
551 KPEQLHYALP DNLTIASSLL ENVHQPQLQE NASTLTLPW QAGIYKLN* 598

```

图 6

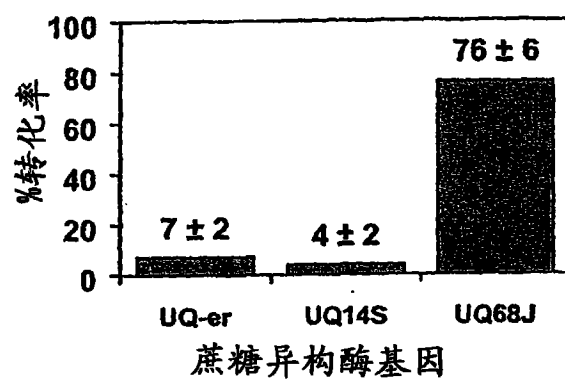


图 7

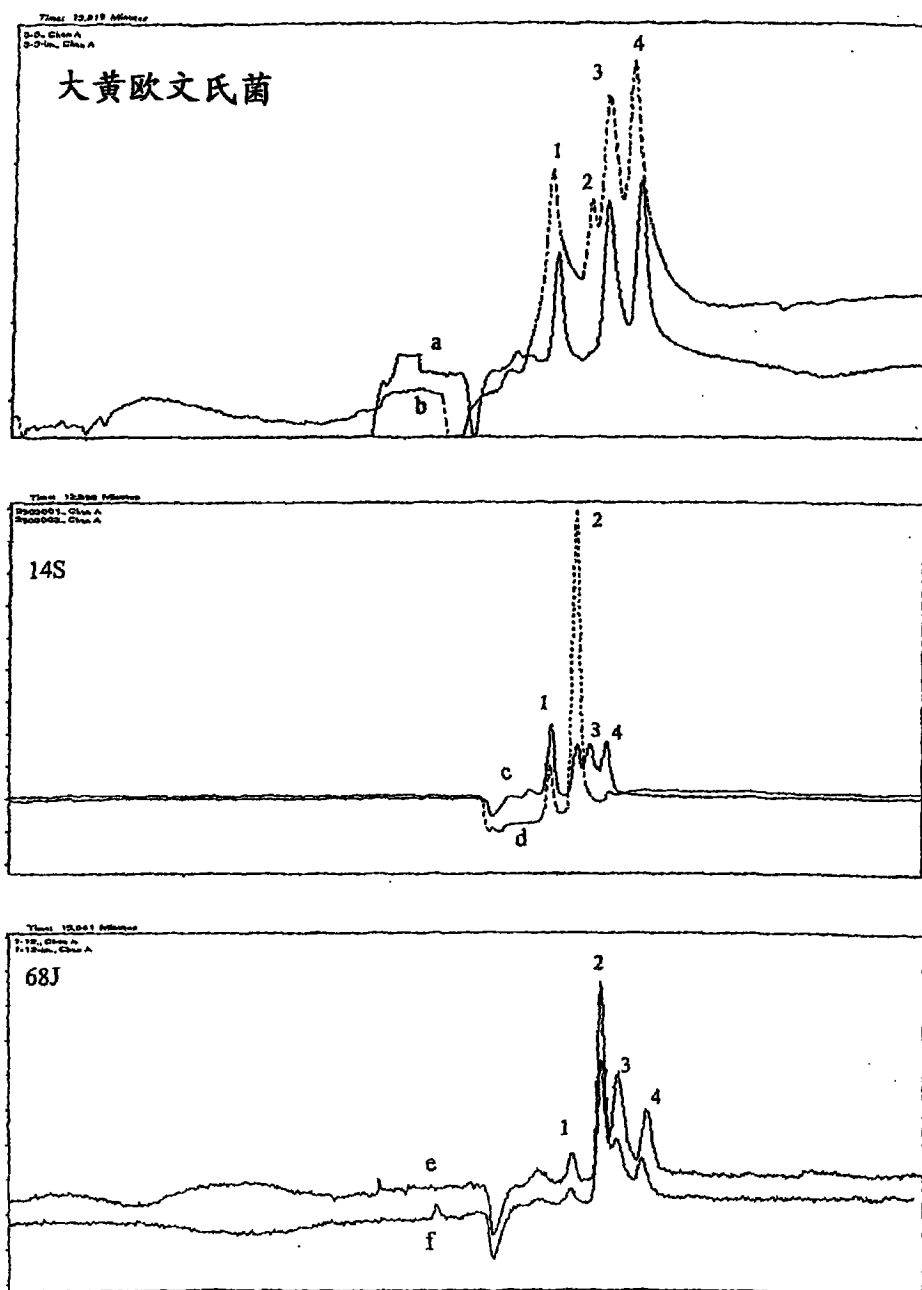


图 8

专利名称(译)	6-果糖- α -葡糖苷合酶		
公开(公告)号	CN1468311B	公开(公告)日	2012-12-12
申请号	CN01816886.8	申请日	2001-08-29
[标]申请(专利权)人(译)	昆士兰大学		
申请(专利权)人(译)	昆士兰大学		
当前申请(专利权)人(译)	昆士兰大学		
[标]发明人	RG博奇 武禄光		
发明人	R·G·博奇 武禄光		
IPC分类号	C12N15/61 C12N15/31 C12N9/90 C13K13/00 A01H1/00 A01H5/00 C07H3/04 C07K16/40 C12N1/21 C12N5/10 C12N9/44 C12N15/09 C12N15/82 C12P19/24 C12Q1/68 G01N31/00 G01N33/53 G01N33/566 G01N33/573		
CPC分类号	C12N15/8245 C12N9/90 C12P19/24 C13K13/00		
代理人(译)	李瑛		
审查员(译)	曾繁辉		
优先权	2000PQ9768 2000-08-29 AU		
其他公开文献	CN1468311A		
外部链接	Espacenet SIPO		

摘要(译)

6-果糖- α -葡糖苷合酶(蔗糖, EC5.4.99.11)将蔗糖转化为6-果糖- α -葡糖苷(6-O- α -D-吡喃葡萄糖基-D-呋喃果糖)。该酶已被从大黄欧文氏菌和另一个细菌分离物68J中分离出来。还公开了用于将蔗糖转化为6-果糖- α -葡糖苷的方法, 转化植物和细菌以使它们能够转化蔗糖, 检测源自样品的6-果糖- α -葡糖苷合酶的方法和鉴定含有源自环境样品的6-果糖- α -葡糖苷合酶的细菌的方法。

SEQ ID NO: 29	编码 SEQ ID NO: 21 所示的共有序列的 SEQ ID NO: 1 的亚序列	18 个碱基
SEQ ID NO: 30	编码 SEQ ID NO: 23 所示的共有序列的 SEQ ID NO: 1 的亚序列	39 个碱基
SEQ ID NO: 31	编码 SEQ ID NO: 24 所示的共有序列的 SEQ ID NO: 1 的亚序列	48 个碱基
SEQ ID NO: 32	编码 SEQ ID NO: 19 所示的共有序列的 SEQ ID NO: 7 的亚序列	21 个碱基
SEQ ID NO: 33	编码 SEQ ID NO: 20 所示的共有序列的 SEQ ID NO: 7 的亚序列	30 个碱基
SEQ ID NO: 34	编码 SEQ ID NO: 21 所示的共有序列的 SEQ ID NO: 7 的亚序列	18 个碱基
SEQ ID NO: 35	编码 SEQ ID NO: 23 所示的共有序列的 SEQ ID NO: 7 的亚序列	39 个碱基
SEQ ID NO: 36	编码 SEQ ID NO: 24 所示的共有序列的 SEQ ID NO: 7 的亚序列	48 个碱基
SEQ ID NO: 37	Geysen 文库肽	8 个残基
SEQ ID NO: 38	基于 Mattes 的正向引物	17 个碱基
SEQ ID NO: 39	基于 Mattes 的反向引物	19 个碱基