

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.

C12P 19/28 (2006.01)

C12N 1/21 (2006.01)



[12] 发明专利申请公布说明书

[21] 申请号 200580017663.3

[43] 公开日 2009年5月6日

[11] 公开号 CN 101426925A

[22] 申请日 2005.3.31

[21] 申请号 200580017663.3

[30] 优先权

[32] 2004.3.31 [33] US [31] 60/558,507

[86] 国际申请 PCT/US2005/010939 2005.3.31

[87] 国际公布 WO2005/098016 英 2005.10.20

[85] 进入国家阶段日期 2006.11.30

[71] 申请人 诺维信生物聚合物公司

地址 丹麦鲍斯韦

[72] 发明人 威廉·威德纳 艾伦·斯洛马

迈克尔·托马斯 玛丽亚·唐

[74] 专利代理机构 北京市柳沈律师事务所

代理人 张文辉 巫肖南

权利要求书 7 页 说明书 43 页 序列表 58 页
附图 27 页

[54] 发明名称

在芽孢杆菌细胞中生产透明质酸的方法

[57] 摘要

本发明涉及生产透明质酸的方法，包括：(a) 在有助于生产透明质酸的培养基中培养芽孢杆菌宿主细胞，其中芽孢杆菌细胞包含含有三联启动子的核酸构建体，三联启动子包含具有对应于 SEQ ID NO: 1 的位置 590 的突变的变体 amyL 启动子，“-35”区域具有 TTGACA 序列和“-10”区域具有 TATAAT 序列的共有启动子和 cryIII A 启动子，其中三联启动子的每个启动子序列可操作地与一个或多个涉及透明质酸生物合成的编码序列相连；和 (b) 从培养基中分离透明质酸。本发明也涉及包含核酸构建体的芽孢杆菌细胞，其中核酸构建体包含 (i) 含有具有对应于 SEQ ID NO: 1 的位置 590 的突变的变体 amyL 启动子，“-35”区域具有 TTGACA 序列和“-10”区域具有 TATAAT 序列的共有启动子和 cryIII A 启动子的三联启动子，其中三联启

动子的每个启动子序列可操作地与该个或多个涉及透明质酸生物合成的编码序列相连。

1. 生产透明质酸的方法，包括：

(a) 在有助于生产透明质酸的培养基中培养芽孢杆菌细胞，其中芽孢杆菌细胞包含含有三联启动子的核酸构建体，三联启动子包含具有对应于 SEQ ID NO: 1 的位置 590 的突变的变体 *amyL* 启动子，"-35"区域具有 TTGACA 序列和"-10"区域具有 TATAAT 序列的共有启动子和 *cryIIIA* 启动子，其中三联启动子的每个启动子序列可操作地与一个或多个涉及透明质酸生物合成的编码序列相连；和

(b) 从培养基中分离透明质酸。

2. 权利要求1的方法，其中变体 $amyL$ 启动子是SEQ ID NO: 1。

3. 权利要求1的方法，其中共有启动子是从任何细菌启动子获得的。

4. 权利要求1的方法，其中共有启动子是从芽孢杆菌启动子获得的。

5. 权利要求1的方法，其中共有启动子是从由大肠杆菌 lac 操纵子、天蓝色链霉菌琼脂糖酶基因 (*dagA*)、迟缓芽孢杆菌碱性蛋白酶基因 (*aprH*)、地衣芽孢杆菌碱性蛋白酶基因 (枯草杆菌蛋白酶Carlsberg基因)、枯草芽孢杆菌果聚糖蔗糖酶基因 (*sacB*)、枯草芽孢杆菌 α -淀粉酶基因 (*amyE*)、地衣芽孢杆菌 α -淀粉酶基因 (*amyL*)、嗜热脂肪芽孢杆菌生麦芽糖淀粉酶基因 (*amyM*)、解淀粉芽孢杆菌 α -淀粉酶基因 (*amyQ*)、地衣芽孢杆菌青霉素酶基因 (*penP*)、枯草芽孢杆菌 $xylA$ 和 $xylB$ 基因、苏云金芽孢杆菌亚种 $tenebrionis$ *cryIIIA*基因或其部分，或者原核生物的 β -内酰胺酶基因获得的启动子中得到的。

6. 权利要求1的方法，其中共有启动子是从解淀粉芽孢杆菌 α -淀粉酶基因 (*amyQ*) 中获得的。

7. 权利要求6的方法，其中共有 $amyQ$ 启动子具有SEQ ID NO: 42或SEQ ID NO: 43的核苷酸序列。

8. 权利要求7的方法，其中共有 $amyQ$ 启动子具有SEQ ID NO: 42或SEQ ID NO: 43的核苷酸86-185的核苷酸序列。

9. 权利要求1的方法，其中 $cryIIIA$ 启动子是从苏云金芽孢杆菌亚种 $tenebrionis$ 获得的。

10. 权利要求1的方法，其中核酸构建体更进一步地包含位于三联启动子下游和一个或多个涉及透明质酸生物合成的编码序列上游的mRNA加

工/稳定序列。

11. 权利要求10的方法，其中mRNA加工/稳定序列是cryIIIA mRNA加工/稳定序列。

12. 权利要求10的方法，其中mRNA加工/稳定序列是SP82 mRNA加工/稳定序列。

13. 权利要求1的方法，其中芽孢杆菌细胞包含一个或多个拷贝的核酸构建体。

14. 权利要求1的方法，其中芽孢杆菌细胞包含一个拷贝的核酸构建体。

15. 权利要求1的方法，其中核酸构建体更进一步地包含可选择的标记基因。

16. 权利要求1的方法，其中芽孢杆菌细胞不包含外来的可选择标记基因。

17. 权利要求1的方法，其中涉及透明质酸生物合成的一个或多个编码序列选自由类透明质酸合酶、UDP-葡萄糖6-脱氢酶、UDP-葡萄糖焦磷酸化酶、UDP-N-乙酰葡萄糖胺焦磷酸化酶、6-磷酸葡萄糖异构酶、己糖激酶、葡萄糖磷酸变位酶、酰胺转移酶、变位酶和乙酰转移酶基因组成的群组。

18. 权利要求1的方法，其中核酸构建体包含在芽孢杆菌细胞的染色体中。

19. 权利要求1的方法，其中核酸构建体包含在染色体外元件中。

20. 权利要求1的方法，其中芽孢杆菌属宿主细胞是嗜碱芽孢杆菌、解淀粉芽孢杆菌、短芽孢杆菌、环状芽孢杆菌、克劳氏芽孢杆菌细胞、凝结芽孢杆菌、坚强芽孢杆菌、灿烂芽孢杆菌、迟缓芽孢杆菌、地衣芽孢杆菌、巨大芽孢杆菌、短小芽孢杆菌、嗜热脂肪芽孢杆菌、枯草芽孢杆菌或苏云金杆菌细胞。

21. 权利要求1的方法，其中芽孢杆菌细胞是枯草芽孢杆菌细胞。

22. 权利要求1的方法，其中芽孢杆菌细胞是地衣芽孢杆菌细胞。

23. 包含核酸构建体的芽孢杆菌细胞，其中核酸构建体含有(a)包含具有对应于SEQ ID NO: 1的位置590的突变的变体*amyL*启动子，"-35"区域具有TTGACA序列和"-10"区域具有TATAAT序列的共有启动子和*cryIIIA*启动子的三联启动子，其中三联启动子的每个启动子序列可操作地与一个或多个涉及透明质酸生物合成的编码序列相连，和任选地(b)位于三联启

动子下游和一个或多个涉及透明质酸生物合成的编码序列的上游的mRNA加工/稳定序列。

24. 权利要求23的芽孢杆菌细胞,其中变体amyL启动子是SEQ ID NO: 1。

25. 权利要求23的芽孢杆菌细胞,其中共有启动子是从任何细菌启动子获得的。

26. 权利要求23的芽孢杆菌细胞,其中共有启动子是从芽孢杆菌属启动子获得的。

27. 权利要求23的芽孢杆菌细胞,其中共有启动子是从由大肠杆菌lac操纵子、天蓝色链霉菌琼脂糖酶基因(dagA)、迟缓芽孢杆菌碱性蛋白酶基因(aprH)、地衣芽孢杆菌碱性蛋白酶基因(枯草杆菌蛋白酶Carlsberg基因)、枯草芽孢杆菌果聚糖蔗糖酶基因(sacB)、枯草芽孢杆菌 α -淀粉酶基因(amyE)、地衣芽孢杆菌 α -淀粉酶基因(amyL)、嗜热脂肪芽孢杆菌生麦芽糖淀粉酶基因(amyM)、解淀粉芽孢杆菌 α -淀粉酶基因(amyQ)、地衣芽孢杆菌青霉素酶基因(penP)、枯草芽孢杆菌xylA和xylB基因、苏云金杆菌亚种tenebrionis cryIIIA基因或其部分,或者原核生物的 β -内酰胺酶基因获得的启动子中得到的。

28. 权利要求23的芽孢杆菌细胞,其中共有启动子是从解淀粉芽孢杆菌 α -淀粉酶基因(amyQ)中获得的。

29. 权利要求28的芽孢杆菌细胞,其中共有amyQ启动子具有SEQ ID NO: 42或SEQ ID NO: 43的核苷酸序列。

30. 权利要求29的芽孢杆菌细胞,其中共有amyQ启动子具有SEQ ID NO: 42或SEQ ID NO: 43的核苷酸86-185的核苷酸序列。

31. 权利要求23的方法,其中cryIIIA启动子是从苏云金杆菌亚种tenebrionis获得的。

32. 权利要求23的芽孢杆菌细胞,其中核酸构建体更进一步地包含位于三联启动子下游和一个或多个涉及透明质酸生物合成的编码序列上游的mRNA加工/稳定序列。

33. 权利要求32的芽孢杆菌细胞,其中mRNA加工/稳定序列是cryIIIA mRNA加工/稳定序列。

34. 权利要求32的芽孢杆菌细胞,其中mRNA加工/稳定序列是SP82 mRNA加工/稳定序列。

35. 权利要求23的芽孢杆菌细胞,其中芽孢杆菌细胞包含一个或多个拷贝的核酸构建体。

36. 权利要求23的芽孢杆菌细胞,其中芽孢杆菌细胞包含一个拷贝的核酸构建体。

37. 权利要求23的芽孢杆菌细胞,其中核酸构建体更进一步地包含可选择的标记基因。

38. 权利要求23的芽孢杆菌细胞,其中芽孢杆菌细胞不包含外来的可选择标记基因。

39. 权利要求23的芽孢杆菌细胞,其中涉及透明质酸生物合成的一个或多个编码序列选自由类透明质酸合酶、UDP-葡萄糖6-脱氢酶、UDP-葡萄糖焦磷酸化酶、UDP-N-乙酰葡萄糖胺焦磷酸化酶、6-磷酸葡萄糖异构酶、己糖激酶、葡萄糖磷酸变位酶、酰胺转移酶、变位酶和乙酰转移酶基因组成的群组。

40. 权利要求23的芽孢杆菌细胞,其中核酸构建体包含在芽孢杆菌细胞的染色体中。

41. 权利要求23的芽孢杆菌细胞,其中核酸构建体包含在染色体外元件中。

42. 权利要求23的芽孢杆菌细胞,其中芽孢杆菌属宿主细胞是嗜碱芽孢杆菌、解淀粉芽孢杆菌、短芽孢杆菌、环状芽孢杆菌、克劳氏芽孢杆菌细胞、凝结芽孢杆菌、坚强芽孢杆菌、灿烂芽孢杆菌、迟缓芽孢杆菌、地衣芽孢杆菌、巨大芽孢杆菌、短小芽孢杆菌、嗜热脂肪芽孢杆菌、枯草芽孢杆菌或苏云金杆菌细胞。

43. 权利要求23的芽孢杆菌细胞,其中芽孢杆菌细胞是枯草芽孢杆菌细胞。

44. 权利要求23的芽孢杆菌细胞,其中芽孢杆菌细胞是地衣芽孢杆菌细胞。

45. 产生没有可选择标记的芽孢杆菌细胞突变体的方法,包括删除芽孢杆菌细胞的可选择标记基因,其中芽孢杆菌细胞包含含有三联启动子的核酸构建体,三联启动子包含具有对应于SEQ ID NO:1的位置590的突变的变体*amyL*启动子,"-35"区域具有TTGACA序列和"-10"区域具有TATAAT序列的共有启动子和*cryIIIa*启动子,其中三联启动子的每个启动子序列可操作地与一个或多个涉及透明质酸生物合成的编码序列相连。

46. 权利要求45的方法，其中变体amyL启动子是SEQ ID NO: 1。
47. 权利要求45的方法，其中共有启动子是从任何细菌启动子获得的。
48. 权利要求45的方法，其中共有启动子是从芽孢杆菌启动子获得的。
49. 权利要求45的方法，其中共有启动子是从由大肠杆菌*lac*操纵子、天蓝色链霉菌琼脂糖酶基因(*dagA*)、迟缓芽孢杆菌碱性蛋白酶基因(*aprH*)、地衣芽孢杆菌碱性蛋白酶基因(枯草杆菌蛋白酶Carlsberg基因)、枯草芽孢杆菌果聚糖蔗糖酶基因(*sacB*)、枯草芽孢杆菌 α -淀粉酶基因(*amyE*)、地衣芽孢杆菌 α -淀粉酶基因(*amyL*)、嗜热脂肪芽孢杆菌生麦芽糖淀粉酶基因(*amyM*)、解淀粉芽孢杆菌 α -淀粉酶基因(*amyQ*)、地衣芽孢杆菌青霉素酶基因(*penP*)、枯草芽孢杆菌*xylA*和*xyIB*基因、苏云金杆菌亚种*tenebrionis cryIIIA*基因或其部分，或者原核生物的 β -内酰胺酶基因获得的启动子中得到的。
50. 权利要求45的方法，其中共有启动子是从解淀粉芽孢杆菌 α -淀粉酶基因(*amyQ*)中获得的。
51. 权利要求50的方法，其中共有*amyQ*启动子具有SEQ ID NO: 42或SEQ ID NO: 43的核苷酸序列。
52. 权利要求51的方法，其中共有*amyQ*启动子具有SEQ ID NO: 42或SEQ ID NO: 43的核苷酸86-185的核苷酸序列。
53. 权利要求45的方法，其中*cryIIIA*启动子从苏云金杆菌亚种*tenebrionis*获得的。
54. 权利要求45的方法，其中核酸构建体更进一步地包含位于三联启动子下游和一个或多个涉及透明质酸生物合成的编码序列上游的mRNA加工/稳定序列。
55. 权利要求54的方法，其中mRNA加工/稳定序列是*cryIIIA* mRNA加工/稳定序列。
56. 权利要求54的方法，其中mRNA加工/稳定序列是SP82 mRNA加工/稳定序列。
57. 权利要求45的方法，其中芽孢杆菌细胞包含一个或多个拷贝的核酸构建体。
58. 权利要求45的方法，其中芽孢杆菌细胞包含一个拷贝的核酸构建

体。

59. 权利要求45的方法，其中核酸构建体更进一步地包含可选择的标记基因。

60. 权利要求45的方法，其中芽孢杆菌细胞不包含外来的可选择标记基因。

61. 权利要求45的方法，其中涉及透明质酸生物合成的一个或多个编码序列是从由类透明质酸合酶、UDP-葡萄糖6-脱氢酶、UDP-葡萄糖焦磷酸化酶、UDP-N-乙酰葡萄糖胺焦磷酸化酶、6-磷酸葡萄糖异构酶、己糖激酶、葡萄糖磷酸变位酶、酰胺转移酶、变位酶和乙酰转移酶基因组成的群组中选择出来的。

62. 权利要求45的方法，其中核酸构建体包含在芽孢杆菌细胞的染色体中。

63. 权利要求45的方法，其中核酸构建体包含在染色体外元件中。

64. 权利要求45的方法，其中芽孢杆菌属宿主细胞是嗜碱芽孢杆菌、解淀粉芽孢杆菌、短芽孢杆菌、环状芽孢杆菌、克劳氏芽孢杆菌细胞、凝结芽孢杆菌、坚强芽孢杆菌、灿烂芽孢杆菌、迟缓芽孢杆菌、地衣芽孢杆菌、巨大芽孢杆菌、短小芽孢杆菌、嗜热脂肪芽孢杆菌、枯草芽孢杆菌或苏云金杆菌细胞。

65. 权利要求45的方法，其中芽孢杆菌细胞是枯草芽孢杆菌细胞。

66. 权利要求45的方法，其中芽孢杆菌细胞是地衣芽孢杆菌细胞。

67. 通过权利要求45的方法获取的没有可选择标记的芽孢杆菌细胞突变体。

68. 获取芽孢杆菌属宿主细胞的方法，包括将包含三联启动子的核酸构建体引入芽孢杆菌细胞，其中三联启动子包含具有对应于SEQ ID NO: 1的位置590的突变的变体*amyL*启动子，"-35"区域具有TTGACA序列和"-10"区域具有TATAAT序列的共有启动子和*cryIII A*启动子，其中三联启动子的每个启动子序列可操作地与一个或多个涉及透明质酸生物合成的编码序列相连。

69. 权利要求68的方法，其中核酸构建体更进一步地包含位于三联启动子下游和一个或多个涉及透明质酸生物合成的编码序列上游的mRNA加工/稳定序列。

70. 权利要求69的方法，其中mRNA加工/稳定序列是*cryIII A* mRNA加

工/稳定序列。

71. 权利要求69的方法,其中mRNA加工/稳定序列是SP82 mRNA加工/稳定序列。

72. 包含三联启动子的核酸构建体,其中三联启动子包含具有对应于SEQ ID NO:1的位置590的突变的变体*amyL*启动子,“-35”区域具有TTGACA序列和“-10”区域具有TATAAT序列的共有启动子和*cryIIIA*启动子,其中三联启动子的每个启动子序列可操作地与一个或多个涉及透明质酸生物合成的编码序列相连。

73. 权利要求72的核酸构建体,其中核酸构建体更进一步地包含位于三联启动子下游和一个或多个涉及透明质酸生物合成的编码序列上游的mRNA加工/稳定序列。

74. 权利要求73的核酸构建体,其中mRNA加工/稳定序列是*cryIIIA* mRNA加工/稳定序列。

75. 权利要求73的核酸构建体,其中mRNA加工/稳定序列是SP82 mRNA加工/稳定序列。

在芽孢杆菌细胞中生产透明质酸的方法

发明背景

发明领域

本发明涉及在细菌细胞中生产透明质酸的方法。

相关技术的描述

透明质酸是由通过交替的 β -1,4-和 β -1,3-糖苷键连接在一起的N-乙酰葡萄糖胺(GlcNAc)和葡萄糖醛酸(GlcUA)的重复二糖单位组成的未硫酸化的糖胺聚糖。已经鉴定透明质酸在体内的许多作用(参见, Laurent T. C. and Fraser J. R. E., 1992, FASEB J. 6: 2397-2404 和 Toole B.P., 1991, "Proteoglycans and hyaluronan in morphogenesis and differentiation." In: Cell Biology of the Extracellular Matrix, pp. 305-341, Hay E D., ed., Plenum, New York)。透明质酸存在于透明软骨、滑膜关节液体和皮肤组织中,在真皮和表皮中都存在。怀疑透明质酸在许多生理机能,例如粘附、发育、细胞运动性、癌症、血管发生和伤口愈合中起作用。由于透明质酸独特的物理和生物学特性,已经在眼睛和关节手术中使用它。也已经研究出供整形外科、风湿病学和皮肤病学使用的透明质酸产物。

公鸡冠已经是透明质酸的常规来源。但是,利用含有生物合成透明质酸的基因的重组微生物正作为一种替换方式出现。

已经很好地确定芽孢杆菌(Bacilli)是生产天然和重组蛋白质的宿主细胞系统。美国专利申请 No.2002/0160489 公开了包含酿脓链球菌(*Streptococcus pyogenes*)类透明质酸合酶和 UDP-葡萄糖脱氢酶基因中的一个或两个基因的三个枯草芽孢杆菌菌株的构建。美国专利申请 No.2003/0092118 描述了利用包含启动子控制下的来自类马链球菌(*Streptococcus equisimilis*)、化脓性链球菌(*Streptococcus pyogenes*)、乳房链球菌(*Streptococcus uberis*)、出血败血性巴斯德(氏)菌(*Pasteurella multocida*)、*Sulfolobus solfataricus*、炭疽芽孢杆菌 pXO1 (*Bacillus anthracis* pXO1)、小球藻病毒(*Chlorella virus*)或长囊水云病毒(*Ectocarpus siliculosus virus*)的生产透明质酸的类透明质酸

合酶基因的重组芽孢杆菌宿主细胞的用途。WO 03/054163 公开了通过培养包含含有类透明质酸合酶编码序列的核酸构建体的芽孢杆菌宿主细胞生产透明质酸的方法，其中类透明质酸（hyaluronan）合酶编码序列可操作地与类透明质酸合酶编码序列的外来启动子序列连接。

美国专利 Nos.6,255,076 和 5,955,310 描述了供在芽孢杆菌细胞中表达酶所用的串联启动子，构建体和方法。其中也描述了利用 cryIIIA 稳定序列改进芽孢杆菌中生产的情况。WO 03/095658 公开了由 amyL4199、短共有 amyQ 和 cryIIIA 启动子序列组成的三联启动子。

提供在芽孢杆菌菌株中生产透明质酸的改进方法是本发明的目标。

发明概述

本发明涉及生产透明质酸的方法，包括：（a）在有助于生产透明质酸的培养基中培养芽孢杆菌宿主细胞，其中芽孢杆菌细胞包含含有三联启动子的核酸构建体，三联启动子包含具有对应于 SEQ ID NO:1 的位置 590 的突变的变体 amyL 启动子，"-35"区域具有 TTGACA 序列和"-10"区域具有 TATAAT 序列的共有启动子和 cryIIIA 启动子，其中三联启动子的每个启动子序列可操作地与一个或多个涉及透明质酸生物合成的编码序列相连；和（b）从培养基中分离透明质酸。在优选方式中，核酸构建体更进一步地包含位于三联启动子下游，一个或多个涉及透明质酸生物合成的编码序列上游的 mRNA 加工/稳定序列。

本发明也涉及包含含有三联启动子的核酸构建体的芽孢杆菌细胞，其中三联启动子包含具有对应于 SEQ ID NO:1 的位置 590 的突变的变体 amyL 启动子，"-35"区域具有 TTGACA 序列和"-10"区域具有 TATAAT 序列的共有启动子和 cryIIIA 启动子，其中三联启动子的每个启动子序列可操作地与一个或多个涉及透明质酸生物合成的编码序列相连。在优选方式中，核酸构建体更进一步地包含位于三联启动子下游，一个或多个涉及透明质酸生物合成的编码序列上游的 mRNA 加工/稳定序列。

本发明也涉及产生没有可选择标记的芽孢杆菌细胞突变体的方法，包括删除芽孢杆菌细胞的可选择标记，其中芽孢杆菌细胞包含含有三联启动子的核酸构建体，三联启动子包含具有对应于 SEQ ID NO:1 的位置 590 的突变的变体 amyL 启动子，"-35"区域具有 TTGACA 序列和"-10"区域具有 TATAAT 序列的共有启动子和 cryIIIA 启动子，其中三联启动子的每个启动

子序列可操作地与一个或多个涉及透明质酸生物合成的编码序列相连。

本发明也涉及获取芽孢杆菌宿主细胞的方法，包括将包含三联启动子的核酸构建体引入芽孢杆菌细胞，其中三联启动子包含具有对应于 SEQ ID NO: 1 的位置 590 的突变的变体 amyL 启动子，"-35"区域具有 TTGACA 序列和"-10"区域具有 TATAAT 序列的共有启动子和 cryIIIA 启动子，其中三联启动子的每个启动子序列可操作地与一个或多个涉及透明质酸生物合成的编码序列相连。

本发明更进一步地涉及包含三联启动子的核酸构建体，其中三联启动子包含具有对应于 SEQ ID NO: 1 的位置 590 的突变的变体 amyL 启动子，"-35"区域具有 TTGACA 序列和"-10"区域具有 TATAAT 序列的共有启动子和 cryIIIA 启动子，其中三联启动子的每个启动子序列可操作地与一个或多个涉及透明质酸生物合成的编码序列相连。

附图简述

- 图 1 显示 pNBT28 的限制性酶切图谱。
- 图 2 显示 pMRT038 的限制性酶切图谱。
- 图 3 显示 pNBT29 的限制性酶切图谱。
- 图 4 显示 pWWi001.1 的限制性酶切图谱。
- 图 5 显示 pWWi005 的限制性酶切图谱。
- 图 6 显示 pNBT30 的限制性酶切图谱。
- 图 7 显示 pNBT31 的限制性酶切图谱。
- 图 8 显示 pNBT33 的限制性酶切图谱。
- 图 9 显示 pMDT006 的限制性酶切图谱。
- 图 10 显示 pMDT007 的限制性酶切图谱。
- 图 11 显示 pNBT37 的限制性酶切图谱。
- 图 12 显示 pNBT38 的限制性酶切图谱。
- 图 13 显示 pNBT39 的限制性酶切图谱。
- 图 14 显示 pMRT040 的限制性酶切图谱。
- 图 15 显示 pMRT044 的限制性酶切图谱。
- 图 16 显示 pMRT070 的限制性酶切图谱。
- 图 17 显示 pMRT075 的限制性酶切图谱。
- 图 18 显示 pNBT40 的限制性酶切图谱。

- 图 19 显示 pMRT077 的限制性酶切图谱。
图 20 显示 pTH012 的限制性酶切图谱。
图 21 显示 pMB1024-1 的限制性酶切图谱。
图 22 显示 pMB1242 的限制性酶切图谱。
图 23 显示 pTH029 的限制性酶切图谱。
图 24 显示 pTH026 的限制性酶切图谱。
图 25 显示 pTH013 的限制性酶切图谱。
图 26 显示 pTH020 的限制性酶切图谱。
图 27 显示通过地衣芽孢杆菌 TH15 生产透明质酸。

发明详述

本发明涉及生产透明质酸的方法，包括：(a) 在有助于生产透明质酸的培养基中培养芽孢杆菌细胞，其中芽孢杆菌细胞包含含有三联启动子的核酸构建体，三联启动子包含具有对应于 SEQ ID NO: 1 的位置 590 的突变的变体 amyL 启动子，“-35”区域具有 TTGACA 序列和“-10”区域具有 TATAAT 序列的共有启动子和 cryIIIA 启动子，其中三联启动子的每个启动子序列可操作地与一个或多个涉及透明质酸生物合成的编码序列相连；和 (b) 从培养基中分离透明质酸。在本发明的方法中，核酸构建体优选更进一步地包含位于三联启动子下游和一个或多个涉及透明质酸生物合成的编码序列上游的 mRNA 加工/稳定序列。

透明质酸

“透明质酸”在这里被定义为由通过交替的 β -1,4-和 β -1,3-糖苷键连接在一起的 N-乙酰葡糖胺 (GlcNAc) 和葡糖醛酸 (GlcUA) 的重复二糖单位组成的未硫酸化的糖胺聚糖。透明质酸又名类透明质酸 (hyaluronan)、透明质酸盐或 HA，他们在这里可交换地使用。

在优选方式中，通过本发明的方法获取的透明质酸具有大约 10,000-大约 10,000,000 Da 的分子量。在更优选的方式中，通过本发明的方法获取的透明质酸具有大约 25,000-大约 5,000,000 Da 的分子量。在最优选的方式中，通过本发明的方法获取的透明质酸具有大约 50,000-大约 3,000,000 Da 的分子量。

可以依照改良的卞啉方法 (Bitter and Muir, 1962, Anal Biochem. 4:

330-334) 确定通过本发明的芽孢杆菌宿主细胞产生的透明质酸的水平。而且, 利用文献中的标准方法, 例如 Ueno et al., 1988, Chem. Pharm. Bull. 36: 4971-4975; Wyatt, 1993, Anal. Chim. Acta 272: 1-40 和 Wyatt Technologies, 1999, "Light Scattering University DAWN Course Manual" and "DAWN EOS Manual", Wyatt Technology Corporation, Santa Barbara, California 中描述的方法可以确定透明质酸的平均分子量。

可以用本领域已知的修饰透明质酸的各种技术, 例如美国专利 Nos.5,616,568、5,652,347 和 5,874,417 中描述的对透明质酸进行交联修饰通过本发明的方法获取的透明质酸。而且, 可以利用本领域已知的技术改变透明质酸的分子量。

宿主细胞

本发明也涉及包含三联启动子的芽孢杆菌细胞, 其中三联启动子包含具有对应于 SEQ ID NO: 1 的位置 590 的突变的变体 amyL 启动子, "-35"区域具有 TTGACA 序列和"-10"区域具有 TATAAT 序列的共有启动子和 cryIIIA 启动子, 其中三联启动子的每个启动子序列可操作地与一个或多个涉及透明质酸生物合成的编码序列相连。在优选方式中, 核酸构建体更进一步地包含位于三联启动子下游和一个或多个涉及透明质酸生物合成的编码序列上游的 mRNA 加工/稳定序列。在另一个优选方式中, 芽孢杆菌细胞没有外来或异源的可选择标志基因。

本发明也涉及获取芽孢杆菌宿主细胞的方法, 包括将含有三联启动子的核酸构建体引入芽孢杆菌细胞, 其中三联启动子包含具有对应于 SEQ ID NO: 1 的位置 590 的突变的变体 amyL 启动子, "-35"区域具有 TTGACA 序列和"-10"区域具有 TATAAT 序列的共有启动子和 cryIIIA 启动子, 其中三联启动子的每个启动子序列可操作地与一个或多个涉及透明质酸生物合成的编码序列相连。在优选方式中, 核酸构建体更进一步地包含位于三联启动子下游和一个或多个涉及透明质酸生物合成的编码序列上游的 mRNA 加工/稳定序列。

在本发明的方法中, 芽孢杆菌宿主细胞可以是适合于重组生产透明质酸的任何芽孢杆菌细胞。芽孢杆菌宿主细胞可以是野生型芽孢杆菌细胞或其突变体。对实施本发明有用的芽孢杆菌细胞包括, 但不限于 *Bacillus agaraderhens*、嗜碱芽孢杆菌(*Bacillus alkalophilus*)、解淀粉芽孢杆菌(*Bacillus*

amyloliquefaciens)、短芽孢杆菌 (*Bacillus brevis*)、环状芽孢杆菌 (*Bacillus circulans*)、克劳氏芽孢杆菌细胞 (*Bacillus clausii*)、凝结芽孢杆菌 (*Bacillus coagulans*)、坚强芽孢杆菌 (*Bacillus firmus*)、灿烂芽孢杆菌 (*Bacillus lautus*)、迟缓芽孢杆菌 (*Bacillus lentus*)、地衣芽孢杆菌 (*Bacillus licheniformis*)、巨大芽孢杆菌 (*Bacillus magaterium*)、短小芽孢杆菌 (*Bacillus pumilus*)、嗜热脂肪芽孢杆菌 (*Bacillus stearothermophilus*)、枯草芽孢杆菌 (*Bacillus subtilis*) 或苏云金芽孢杆菌 (*Bacillus thuringiensis*) 细胞。在 WO 98/22598 中描述了特别适合于重组表达的枯草芽孢杆菌细胞突变体。非包囊 (encapsulating) 芽孢杆菌细胞对本发明特别有用。

在优选方式中, 芽孢杆菌宿主细胞是解淀粉芽孢杆菌 (*Bacillus amyloliquefaciens*)、克劳氏芽孢杆菌 (*Bacillus clausii*)、迟缓芽孢杆菌 (*Bacillus lentus*)、地衣芽孢杆菌、嗜热脂肪芽孢杆菌 (*Bacillus stearothermophilus*) 或枯草杆菌 (*Bacillus subtilis*) 细胞。在更优选的方式中, 芽孢杆菌细胞是解淀粉芽孢杆菌 (*Bacillus amyloliquefaciens*) 细胞。在另一个更优选的方式中, 芽孢杆菌细胞是克劳氏芽孢杆菌细胞 *Bacillus clausii*)。在另一个更优选的方式中, 芽孢杆菌细胞是迟缓芽孢杆菌细胞 (*Bacillus lentus*)。在另一个更优选的方式中, 芽孢杆菌细胞是地衣芽孢杆菌细胞 (*Bacillus licheniformis*)。在另一个更优选的方式中, 芽孢杆菌细胞是枯草芽孢杆菌细胞 (*Bacillus subtilis*)。在最优选方式中, 芽孢杆菌宿主细胞是枯草杆菌 A164 Δ 5 或枯草杆菌 168 Δ 4 (参见美国专利 No.5,891,701)。在另一个最优选方式中, 芽孢杆菌宿主细胞是地衣芽孢杆菌 SJ1904 (参见美国专利 No.5,733,753)。

例如, 可以通过原生质体转化 (例如, 参见 Chang and Cohen, 1979, *Molecular General Genetics* 168: 111-115)、使用感受态细胞 (例如, 参见 Young and Spizizen, 1961, *Journal of Bacteriology* 81: 823-829, or Dubnau and Davidoff-Abelson, 1971, *Journal of Molecular Biology* 56: 209-221)、通过电穿孔 (例如, 参见 Shigekawa and Dower, 1988, *Biotechniques* 6: 742-751)、或通过结合作用 (例如, 参见 Koehler and Thorne, 1987, *Journal of Bacteriology* 169: 5271-5278), 用本发明的核酸构建体转化芽孢杆菌宿主细胞。

核酸构建体

通过利用本领域众所周知的方法修饰一个或多个基因, 将三联启动子

或者三联启动子和 mRNA 加工/稳定序列可操作地连接到涉及透明质酸生物合成的一个或多个基因上，将构建体插入载体中，并通过同源重组将载体导入芽孢杆菌细胞的染色体或者将载体作为染色体外的自主复制元件，例如质粒导入芽孢杆菌细胞来完成对包含可操作地与这一个或多个基因相连的三联启动子，或三联启动子和 mRNA 加工/稳定序列的芽孢杆菌细胞的构建。但是，应理解也可以利用本领域众所周知的方法在芽孢杆菌细胞内操作这一个或多个基因。

"核酸构建体"在这里限定为分离自天然发生基因或已经经过修饰的包含核酸片段的单链或双链的核酸分子，其中核酸片段以不存在于自然界的方式组合和连接。当核酸构建体包含表达编码序列需要的全部控制序列时，术语核酸构建体可能与术语表达盒同义。

"启动子"在这里限定为涉及结合 RNA 聚合酶启动基因转录的核苷酸序列。

"三联启动子"在这里限定为串联的三个启动子序列，其中每个启动子可操作地与一个编码序列或多个编码序列相连，并介导编码序列转录成 mRNA。

"操作性连接"在这里限定为其中控制序列，例如三联启动子恰当地位于编码序列的相对位置使控制序列指导编码序列编码的透明质酸的生产的构型。

"编码序列"在这里被限定为在适当的控制序列的支配下，被转录成 mRNA，并被翻译成涉及透明质酸生物合成的酶（或其他蛋白质）的核苷酸序列。通常通过正好位于 mRNA 5'端的开放阅读框上游的核糖体结合部位和正好位于 mRNA 3'端的开放阅读框下游的转录终止序列确定编码序列的范围。编码序列可以包括，但是不限于基因组 DNA、cDNA、半合成的、合成的核酸和重组的核酸。

用于分离或克隆编码多肽，例如编码酶的基因的技术为本领域熟知，例如包括从基因组 DNA 分离、从 cDNA 制备或他们的组合。例如，通过利用表达库的抗体筛选法检测克隆的具有共用结构特点的 DNA 片段或众所周知的聚合酶链反应（PCR）可以完成从这种基因组 DNA 克隆基因。例如：参见 Innis et al., 1990, PCR Protocols: A Guide to Methods and Application, Academic Press, New York。可以使用其他的核酸扩增过程，例如连接酶链反应、连接活化转录和基于核酸序列的扩增。克隆过程涉及切割和分离包含

多肽编码基因的目的核酸片段，将片段插入到载体分子中，以及将重组载体整合到核苷酸序列的克隆可以在其中复制的芽孢杆菌细胞中。基因可以是基因组、cDNA、RNA、半合成来源，合成来源的基因或他们的任何组合。

可以用多种方式操作分离的编码涉及透明质酸生物合成的酶（或其他蛋白质）的基因，为该酶（或其他蛋白质）的表达作准备。将基因序列插入到构建体或载体之前对基因序列进行操作可能是合乎需要的或需要的，这取决于表达载体或芽孢杆菌宿主细胞。利用克隆方法修饰核苷酸序列的技术在本领域为大家所熟知。应理解也可以利用本领域众所周知的方法在宿主细胞内操作基因序列。

许多酶与透明质酸的生物合成有关。在本发明的方法中，一种或多种涉及透明质酸生物合成的基因包括，但不限于编码类透明质酸合酶、UDP-葡萄糖 6-脱氢酶、UDP-葡萄糖焦磷酸化酶、UDP-N-乙酰葡萄糖胺焦磷酸化酶、6-磷酸葡萄糖异构酶、己糖激酶、葡萄糖磷酸变位酶、酰胺转移酶、变位酶和乙酰转移酶的基因。类透明质酸合酶是生产透明质酸的关键酶。

"类透明质酸合酶"在这里被限定为通过添加 GlcUA 和 GlcNAc 糖前体催化类透明质酸链延伸的合酶。链球菌的类透明质酸合酶，脊椎动物的类透明质酸合酶和病毒的类透明质酸合酶的氨基酸序列不同于巴斯德菌属 (*Pasteurella*) 的类透明质酸合酶，已经建议将他们分类为 I 组和 II 组类透明质酸合酶，I 组类透明质酸合酶包括链球菌的类透明质酸合酶 (DeAngelis, 1999, *Cell. Mol. Life Sci.* 56: 670-682)。为了在芽孢杆菌宿主细胞中生产类透明质酸，可以使用真核来源的类透明质酸合酶，例如哺乳动物类透明质酸合酶，但是这种类透明质酸合酶优选程度低。

类透明质酸合酶基因可以是任何能够在芽孢杆菌宿主细胞中表达的任何类透明质酸合酶基因。基因可以是任何来源的基因。优选的类透明质酸合酶基因包括 I 组或 II 组的任何基因，例如 I 组的来自类马链球菌、化脓性链球菌、乳房链球菌、马链球菌兽疫亚种 (*Streptococcus equi* subsp. *Zooepidemicus*) 的类透明质酸合酶基因，或 II 组的 *Pasteurella multocida* 的类透明质酸合酶基因。

可以用这里公开的核苷酸序列或其子序列，以及其氨基酸序列或其片段来设计核酸探针，以依照本领域众所周知的方法从不同属或种的菌株中鉴定和克隆编码涉及透明质酸生物合成的酶的 DNA。特别地，在进行标准的 Southern 印迹过程后，这种探针可用于与有价值的属或种的基因组或

cDNA 杂交，以鉴定和分离其中对应的基因。这种探针可以大大短于全长序列，但是长度应该是至少 14 个、优选至少 25 个、更优选至少 35 个，最优选至少 70 个核苷酸。也可以使用长些的探针。DNA 和 RNA 探针都可以使用。为了检测对应的基因，典型地可以标记探针（例如：用 ^{32}P 、 ^3H 、 ^{35}S 、生物素或亲和素）。

因此，可以筛选从这种其他生物体制备的基因组 DNA 或 cDNA 库，以选择与如上所述的探针杂交和编码透明质酸生物合成途径中的酶的 DNA。通过琼脂糖凝胶电泳或聚丙烯酰胺凝胶电泳，或其他分离方法可以从这种其他生物体分离基因组 DNA 或其他 DNA。来自库的 DNA 或分离的 DNA 可以被转移和固定到硝酸纤维素或其他适当的载体上。为了识别与这里公开的核苷酸序列或其子序列同源的克隆或 DNA，在 Southern 印迹中使用了载体。对本发明来说，杂交指核酸序列在极低到很高的严谨条件下，与相当于这里公开的核苷酸序列的标记的核酸探针其互补链或子序列杂交。可以利用 X 线胶片检测在这些条件下与核酸探针杂交的分子。

对长度为至少 100 个核苷酸的长探针来说，极低到很高的严谨条件被定义为遵循标准 Southern 印迹程序， 42°C 在 5XSSPE，0.3% SDS，200ug/ml 的剪切变性的鲑鱼精 DNA，以及 25%，35% 或 50% 的甲酰胺中预杂交和杂交最佳 12-24 小时，其中 25%，35% 或 50% 甲酰胺分别代表极低和低严谨条件，中等和中等-高严谨条件，以及高和很高的严谨条件。

对长度为至少 100 个核苷酸的长探针来说，最后优选在至少 45°C （极低严谨条件），更优选至少 50°C （低严谨条件），更优选至少 55°C （中等严谨条件），更优选至少 60°C （中等-高严谨条件），甚至更优选至少 65°C （高严谨条件），最优选至少 70°C （很高的严谨条件）下，用 2XSSC，0.2% SDS 洗涤载体三次，每次 15 分钟。

对长度为大约 15-70 个核苷酸的短探针来说，严谨条件定义为遵循标准 Southern 印迹程序，在低于计算的 T_m 大约 $5-10^\circ\text{C}$ 的条件下，在 0.9M NaCl，0.09M Tris-HCl，pH 7.6，6mM EDTA，0.5% NP-40，1X Denhardt's 溶液，1mM 焦磷酸钠，1mM 磷酸二氢钠，0.1mM ATP 和每毫升 0.2mg 酵母 RNA 中预杂交，杂交，洗涤最佳 12-24 小时，其中 T_m 是参照 Bolton 和 McCarthy（1962, Proceedings of the National Academy of Sciences USA 48:1390）计算的。

对长度为大约 15-70 个核苷酸的短探针来说，在 6XSSC，0.1% SDS 中

洗涤载体一次，冲洗 15 分钟，再在低于计算的 $T_m-10^\circ\text{C}$ 的条件下利用 6XSSC 洗涤两次，每次 15 分钟。

在优选方案中，类透明质酸合酶基因是 I 组的类透明质酸合酶基因。

在更优选的方案中，I 组类透明质酸合酶基因选自由 (a) 编码的类透明质酸合酶的氨基酸序列与 SEQ ID NO: 3, SEQ ID NO: 5 或 SEQ ID NO: 7 具有最少 70%，优选至少 75%，更优选至少 80%，更优选至少 85%，更加优选至少 90%，最优选至少 95%，或者甚至最优选至少 97% 的同一性的基因；b) 在低，中等，中等-高，或高严谨条件下与 SEQ ID NO: 2, SEQ ID NO: 4 或 SEQ ID NO: 6 杂交的基因；和 (c) (a) 或 (b) 的互补链组成的群组。对本发明来说，通过 Clustal 方法 (Higgins, 1989, CABIOS 5: 151-153)，利用具有同一性表和下列多重比对参数：缺口计罚分 (gap penalty) 10，间隙长度计罚分 (gap length penalty) 10 的 LASERGENE™ MEGALIGN™ 软件 (DNASTAR, Inc., Madison, WI) 确定两个氨基酸序列之间的同一性程度。成对地比对参数是 Ktuple=1，缺口罚分=3，窗口=5 和对角线=5。

在最优选的方案中，I 组类透明质酸合酶基因编码具有 SEQ ID NO: 3, SEQ ID NO: 5 或 SEQ ID NO: 7 的氨基酸序列的类透明质酸合酶或其具有类透明质酸合酶活性的片段。

在另一个优选方案中，类透明质酸合酶基因是 II 组的类透明质酸合酶基因。

在更优选的方案中，II 组类透明质酸合酶基因选自由 (a) 编码的类透明质酸合酶的氨基酸序列与 SEQ ID NO: 9 具有最少 70%，优选至少 75%，更优选至少 80%，更优选至少 85%，更加优选至少 90%，最优选至少 95%，或者甚至最优选至少 97% 的同一性的基因；b) 在低，中等，中等-高，或高严谨条件下与 SEQ ID NO: 8 杂交的基因；和 (c) (a) 或 (b) 的互补链组成的群组。

在最优选的方案中，II 组类透明质酸合酶基因编码具有 SEQ ID NO: 9 的氨基酸序列的类透明质酸合酶或其具有类透明质酸合酶活性的片段。

可用于本发明的其他类透明质酸合酶基因是来自炭疽芽孢杆菌、硫磺矿硫化叶菌 (Sulfolobus solfataricus)、长囊水云病毒 (Ectocarpus siliculosus virus) 和小球藻病毒 1 (paramecium bursaria Chlorella virus) (PBCV - 1) 的类透明质酸合酶基因。

在另一个更优选的方案中，类透明质酸合酶基因选自由 (a) 编码的类

透明质酸合酶的氨基酸序列与 SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 15 或 SEQ ID NO: 17 具有最少 70%, 优选至少 75%, 更优选至少 80%, 更优选至少 85%, 更加优选至少 90%, 最优选至少 95%, 或者甚至最优选至少 97% 的同一性的基因; b) 在低, 中等, 中等-高, 或高严谨条件下与 SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 14 或 SEQ ID NO: 16 杂交的基因; 和 (c) (a) 或 (b) 的互补链组成的群组。

在更优选的方案中, 类透明质酸合酶基因编码具有 SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 15 或 SEQ ID NO: 17 的氨基酸序列的类透明质酸合酶或其具有类透明质酸合酶活性的片段。

本发明的方法也包括核酸(nucleic)构建体, 在构建体中通过由存在于构建体中的内源基因, 非内源基因, 或者内源和非内源基因的组合编码类透明质酸前体糖而将前体糖提供给宿主细胞。前体糖可以是 D-葡萄糖醛酸或 N-乙酰氨基葡萄糖。

在本发明的方法中, 核酸构建体可以更进一步地包含编码涉及类透明质酸前体糖生物合成的酶的一个或多个基因。换句话说, 芽孢杆菌宿主细胞可以更进一步地包含含有编码涉及前体糖生物合成的酶的一个或多个基因的一个或多个第二核酸构建体。通过利用具有指导类透明质酸前体糖生物合成路径中的步骤的一个或多个基因的构建体可以改进类透明质酸的生产。短语"指导类透明质酸前体糖生物合成路径中的步骤"在这里指在 N-乙酰氨基葡萄糖或 D-葡萄糖醛酸, 或者 N-乙酰氨基葡萄糖或 D-葡萄糖醛酸前体糖的形成中, 该基因表达的酶是有活性的。

在供给前体糖的优选方法中, 通过培养具有重组构建体的宿主细胞来提供改进天然包含类透明质酸合酶基因的宿主细胞中类透明质酸的生产的重组构建体, 其中重组构建体具有可操作地与编码类透明质酸前体糖的生物合成路径中的酶的一个或多个基因相连的三联启动子。在优选方法中, 宿主细胞也包含具有可操作地与类透明质酸合酶相连的三联启动子的重组构建体。因此, 本发明也涉及通过利用具有指导类透明质酸前体糖生物合成路径中的步骤的一个或多个基因的构建体改进类透明质酸生产的构建体。核酸构建体中的这种基因可操作地与这里描述的三联启动子相连。

涉及生产透明质酸的前体糖的生物合成的基因包括, 但不限于 UDP-葡萄糖 6-脱氢酶基因、UDP-葡萄糖焦磷酸化酶基因、UDP-N-乙酰葡萄糖胺焦磷酸化酶基因、6-磷酸葡萄糖异构酶基因、己糖激酶基因、葡萄糖磷酸变位酶基

因、酰胺转移酶基因、变位酶基因和乙酰转移酶基因。

在包含类透明质酸合酶基因的细胞中，可以表达 hasB、hasC、hasD 或其同系物中的任何一种，任何两种或更多种的组合以增加类透明质酸合酶可用的前体糖库。在 Kunst, et al., Nature 390, 249-256, "The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*" (20 November 1997) 中描述了枯草芽孢杆菌基因组。在有些情况下，例如宿主细胞没有天然类透明质酸合酶活性的情况下，构建体更进一步地包括 hasA 基因。

编码生物合成酶的基因可以是宿主细胞与生俱来的，但在其它情况下，可以使用异源基因，或者天然和异源基因的组合。如果一个或多个基因包括在构建体中，他们可能是天然操纵子中彼此有关的基因，例如包含 hasA、hasB、hasC 和 hasD 的类马链球菌的 HAS 操纵子的基因。在其他情况下，利用没有包括操纵子所有组件的前体基因的某些组合也可能是想要的。在其它情况下，利用一些宿主细胞天然基因和其他外源基因也可能是优选的。选择取决于给定宿主细胞中可利用的糖库，在不干扰宿主细胞其他功能的情况下细胞适应生产过剩的能力，以及细胞对天然基因和外源基因表达的调节是否不同。

在一个实施例中这取决于细胞的代谢需求和生长条件，可利用的前体糖库，因此通过表达编码 UDP-N-乙酰葡萄糖胺焦磷酸化酶的基因，例如 hasD 基因，芽孢杆菌 *gcaD* 基因或其同系物增加 N-乙酰氨基葡萄糖的生产是合乎需要的。换句话说，前体糖可以是 D 葡萄糖醛酸。在这样的一个方案中，基因编码 UDP-葡萄糖 6-脱氢酶。这种基因包括芽孢杆菌 *tuaD* 基因，链球菌的 hasB 基因或他们的同系物。另一个基因可以编码 UDP 葡萄糖焦磷酸化酶，例如芽孢杆菌 *gtaB* 基因，链球菌的 hasC 基因或他们的同系物。

在本发明的方法中，UDP-葡萄糖 6-脱氢酶基因可能是 hasB 基因或 *tuaD* 基因，或他们的同系物。

在优选方案中，hasB 基因选自由 (a) 编码氨基酸序列与 SEQ ID NO: 19, SEQ ID NO: 21 或 SEQ ID NO: 23 有最少 70%，优选至少 75%，更优选至少 80%，更优选至少 85%，更加优选至少 90%，最优选至少 95%，或者甚至最优选至少 97% 同一性的 UDP-葡萄糖 6-脱氢酶的基因；(b) 在低，中等，中等-高，或高严谨条件下与 SEQ ID NO: 18, SEQ ID NO: 20 或 SEQ ID NO: 22 杂交的基因；和 (c) (a) 或 (b) 的互补链组成的群组。

在更优选的方案中，hasB 基因编码具有 SEQ ID NO: 19, SEQ ID NO: 21

或 SEQ ID NO: 23 的氨基酸序列的 UDP-葡萄糖 6-脱氢酶或其具有 UDP-葡萄糖 6-脱氢酶活性的片段。

在另一个优选的方案中, tuaD 基因选自由 (a) 编码氨基酸序列与 SEQ ID NO: 25 有最少 70%, 优选至少 75%, 更优选至少 80%, 更优选至少 85%, 更加优选至少 90%, 最优选至少 95%, 或者甚至最优选至少 97% 同一性的多肽的核苷酸序列; b) 在低, 中等, 中等-高, 或高严谨条件下与 SEQ ID NO: 24 杂交的核苷酸序列; 和 (c) (a) 或 (b) 的互补链组成的群组。

在另一个更优选的方案中, tuaD 基因编码具有 SEQ ID NO: 25 的氨基酸序列的 UDP-葡萄糖 6-脱氢酶或其具有 UDP-葡萄糖 6-脱氢酶活性的片段。

在本发明的方法中, UDP-葡萄糖焦磷酸化酶基因可能是 hasC 基因或 gtaB 基因, 或他们的同系物。

在优选方案中, hasC 基因选自由 (a) 编码氨基酸序列与 SEQ ID NO: 27, SEQ ID NO: 29 或 SEQ ID NO: 31 有最少 70%, 优选至少 75%, 更优选至少 80%, 更优选至少 85%, 更加优选至少 90%, 最优选至少 95%, 或者甚至最优选至少 97% 同一性的 UDP-葡萄糖焦磷酸化酶的基因; (b) 在低, 中等, 中等-高, 或高严谨条件下与 SEQ ID NO: 26, SEQ ID NO: 28 或 SEQ ID NO: 30 杂交的基因; 和 (c) (a) 或 (b) 的互补链组成的群组。

在另一个更优选的方案中, hasC 基因编码具有 SEQ ID NO: 27, SEQ ID NO: 29 或 SEQ ID NO: 31 的氨基酸序列的 UDP-葡萄糖焦磷酸化酶或其具有 UDP-葡萄糖焦磷酸化酶活性的片段。

在另一个优选的方案中, gtaB 基因选自由 (a) 编码氨基酸序列与 SEQ ID NO: 33 有最少 70%, 优选至少 75%, 更优选至少 80%, 更优选至少 85%, 更加优选至少 90%, 最优选至少 95%, 或者甚至最优选至少 97% 同一性的 UDP-葡萄糖焦磷酸化酶的基因; b) 在低, 中等, 中等-高, 或高严谨条件下与 SEQ ID NO: 32 杂交的基因; 和 (c) (a) 或 (b) 的互补链组成的群组。

在另一个更优选的方案中, gtaB 基因编码具有 SEQ ID NO: 33 的氨基酸序列的 UDP-葡萄糖焦磷酸化酶或其具有 UDP-葡萄糖焦磷酸化酶活性的片段。

在本发明的方法中, UDP-N-乙酰葡萄糖胺焦磷酸化酶基因可能是 hasD 或 gcaD 基因, 或他们的同系物。

在优选方案中, hasD 基因选自由 (a) 编码氨基酸序列与 SEQ ID NO: 35 有最少 70%, 优选至少 75%, 更优选至少 80%, 更优选至少 85%, 更加优

选至少 90%，最优选至少 95%，或者甚至最优选至少 97%同一性的 UDP-N-乙酰葡萄糖胺焦磷酸化酶的基因；(b) 在低，中等，中等-高，或高严谨条件下与 SEQ ID NO: 34 杂交的基因；和 (c) (a) 或 (b) 的互补链组成的群组。

在另一个更优选的方案中，hasD 基因编码具有 SEQ ID NO: 35 的氨基酸序列的 UDP-N-乙酰葡萄糖胺焦磷酸化酶或其具有 UDP-N-乙酰葡萄糖胺焦磷酸化酶活性的片段。

在优选的方案中，gcaD 基因选自由 (a) 编码氨基酸序列与 SEQ ID NO: 37 有最少 70%，优选至少 75%，更优选至少 80%，更优选至少 85%，更加优选至少 90%，最优选至少 95%，或者甚至最优选至少 97%同一性的 UDP-N-乙酰葡萄糖胺焦磷酸化酶的基因；(b) 在低，中等，中等-高，或高严谨条件下与 SEQ ID NO: 36 杂交的基因；和 (c) (a) 或 (b) 的互补链组成的群组。

在另一个更优选的方案中，gcaD 基因编码具有 SEQ ID NO: 37 的氨基酸序列的 UDP-N-乙酰葡萄糖胺焦磷酸化酶或其具有 UDP-N-乙酰葡萄糖胺焦磷酸化酶活性的片段。

在本发明的方法中，6-磷酸葡萄糖异构酶基因可能是 hasE 或其同系物。

在优选方案中，hasE 基因选自由 (a) 编码氨基酸序列与 SEQ ID NO: 39 有最少 70%，优选至少 75%，更优选至少 80%，更优选至少 85%，更加优选至少 90%，最优选至少 95%，或者甚至最优选至少 97%同一性的 6-磷酸葡萄糖的基因；(b) 在低，中等，中等-高，或高严谨条件下与 SEQ ID NO: 38 杂交的基因；和 (c) (a) 或 (b) 的互补链组成的群组。

在另一个更优选的方案中，hasE 基因编码具有 SEQ ID NO: 39 的氨基酸序列的 6-磷酸葡萄糖或其具有 6-磷酸葡萄糖异构酶活性的片段。

本发明也涉及包含编码类透明质酸合酶操纵子的分离的多核苷酸的核酸构建体，其中操纵子包含类透明质酸合酶基因和 UDP-葡萄糖 6-脱氢酶基因，以及任选地由 UDP-葡萄糖焦磷酸化酶基因，UDP-N-乙酰葡萄糖胺焦磷酸化酶基因和 6-磷酸葡萄糖异构酶基因组成的群组选择出来的一个或多个基因。

可以模仿类马链球菌的操纵子 (WO 03/054163) 或化脓性链球菌的操纵子 (Crater and van de Rijn, 1995, J. Biol. Chem. 270: 18452-18458) 构建"人工操纵子"。这种人工操纵子包含 hasA、hasB、hasC 和 hasD，或他们的同系物，可选地，包括的成分可以少于存在于类马链球菌操纵子中的全部补足物 (full complement)。人工操纵子也可以同时包含 6-磷酸葡萄糖异构

酶基因 (hasE), 以及从由己糖激酶基因、葡糖磷酸变位酶基因、酰胺转移酶基因、变位酶基因和乙酰转移酶基因组成的群组选择出来的一个或多个基因。在 SEQ ID NO: 40 中发现了编码类马链球菌类透明质酸合酶操纵子大部分的多核苷酸。这一序列包括分别是枯草芽孢杆菌 tuaD 基因 (SEQ ID NO: 24) 和 gtaB 基因 (SEQ ID NO: 32) 同系物的 hasB (SEQ ID NO: 18) 和 hasC (SEQ ID NO: 26), 这和化脓性链球菌(*Streptococcus equisimilis*)的 gcaD 基因的同系物 (SEQ ID NO: 36) 已经被命名为 hasD (SEQ ID NO: 34) 的情况一样。枯草芽孢杆菌 gcaD 基因编码参与 N 乙酰氨基葡萄糖合成的 UDP-N-乙酰葡萄糖胺焦磷酸化酶, N-乙酰氨基葡萄糖是类透明质酸的双组分中的一种。类马链球菌的 gcaD、hasD 同系物由类马链球菌排列在类透明质酸合酶操纵子上。多核苷酸也包含 hasA 基因的一部分 (SEQ ID NO: 2 的最后 1156bp)。

在优选方案中, 核酸构建体包含从由 hasA、hasB、tuaD、hasC、gtaB、hasD、gcaD 和 hasE 组成的群组中选择出来的一个或多个基因。

在另一个优选方案中, 核酸构建体包含 hasA。

在另一个优选方案中, 核酸构建体包含 hasA 和 hasB 或 tuaD。在另一个优选方案中, 核酸构建体包含 hasA 和 hasC 或 gtaB。在另一个优选方案中, 核酸构建体包含 hasA 和 hasB 或 gcaD。在另一个优选方案中, 核酸构建体包含 hasA 和 hasE。在另一个优选方案中, 每一个如上所述的核酸构建体都不包含 hasA。

在另一个优选方案中, 核酸构建体包含 hasA、hasB 或 tuaD, 以及 hasC 或 gtaB。在另一个优选方案中, 核酸构建体包含 hasA、hasB 或 tuaD, 以及 hasD 或 gcaD。在另一个优选方案中, 核酸构建体包含 hasA、hasB 或 tuaD, 以及 hasE。在另一个优选方案中, 核酸构建体包含 hasA、hasC 或 gtaB, 以及 hasD 或 gcaD。在另一个优选方案中, 核酸构建体包含 hasA、hasC 或 gtaB, 以及 hasE。在另一个优选方案中, 每一个如上所述的核酸构建体都不包含 hasA。

在另一个优选方案中, 核酸构建体包含 hasA、hasB 或 tuaD, hasC 或 gcaD, 以及 hasD。在另一个优选方案中, 核酸构建体包含 hasA、hasB、hasD 或 gcaD, 以及 hasE。在另一个优选方案中, 核酸构建体包含 hasA、hasC 或 gtaD, hasD 或 gcaD, 以及 hasE。在另一个优选方案中, 核酸构建体包含 hasA、hasB 或 tuaD, hasC 或 gtaD, 以及 hasE。在另一个优选方案中,

每一个如上所述的核酸构建体都不包含 hasA。

基于以上所述的优选方案，可以用记录的基因的其他同系物替换记录的基因。

在本发明的方法中，核酸构建体包含可操作地与三联启动子连接的涉及类透明质酸生物合成的一个或多个基因，其中三联启动子包含具有对应于 SEQ ID NO : 1 的位置 590 的突变的变体 amyL 启动子，"-35"区域具有 TTGACA 序列和"-10"区域具有 TATAAT 序列的共有启动子和 cryIIIA 启动子，三联启动子的每个启动子序列可操作地与一个或多个涉及透明质酸生物合成的编码序列相连。三联启动子的启动子顺序可以是任何次序。

在本发明的方法中，可以从任何细菌来源中获得三联启动子成分。在优选方案中，从革兰氏阳性细菌，例如芽孢杆菌菌株，例如嗜碱芽孢杆菌 (*Bacillus alkalophilus*)、解淀粉芽孢杆菌 (*Bacillus amyloliquefaciens*)、短芽孢杆菌 (*Bacillus brevis*)、环状芽孢杆菌 (*Bacillus circulans*)、克劳氏芽孢杆菌细胞 (*Bacillus clausii*)、凝结芽孢杆菌 (*Bacillus coagulans*)、坚强芽孢杆菌 (*Bacillus firmus*)、灿烂芽孢杆菌 (*Bacillus lautus*)、迟缓芽孢杆菌 (*Bacillus lentus*)、地衣芽孢杆菌 (*Bacillus licheniformis*)、巨大芽孢杆菌 (*Bacillus magaterium*)、短小芽孢杆菌 (*Bacillus pumilus*)、嗜热脂肪芽孢杆菌 (*Bacillus stearothermophilus*)、枯草芽孢杆菌 (*Bacillus subtilis*) 或苏云金芽孢杆菌 (*Bacillus thuringiensis*)；或链霉菌菌株，例如变铅青链霉菌 (*Streptomyces lividans*) 或 *Streptomyces murinus*；或革兰氏阴性细菌，例如大肠杆菌或假单胞菌中获得启动子序列。

供本发明的方法使用的适当的 amyL 启动子的实例是地衣芽孢杆菌 α -淀粉酶基因的启动子 (amyL)。供本发明的方法使用的适当的 cryIIIA 启动子的实例是苏芸金芽孢杆菌亚种 tenebrionis cryIIIA 基因的启动子。

在本发明的方法中可以依照美国专利 Nos. 5,698,415 和 6,100,063 获取具有对应于 SEQ ID NO : 1 的位置 590 的突变的变体 amyL 启动子，其中突变是将位置 590 的 T 转换为 A 以产生 SEQ ID NO : 1。美国专利 No. 5,698,415 要求保护来源于地衣芽孢杆菌 amyL 启动子的启动子变异体。参考上述专利的权利要求 1 可知，这种启动子变异体是该权利要求中给出的序列的片段，其中 N2-N9 具有序列 ATGTATCA。通过将想要的突变掺入到覆盖 amyL 启动子区域的长 PCR 引物，28902 (美国专利 No. 6,100,063) 中来构建这种启动子变异体。另一个 PCR 引物，LWN3216 (美国专利 No.6,100,063) 从跨

越 AmyL 信号肽编码区域的 PstI 位点向上游阅读。这些引物允许 PCR 扩增来源于亲代 amyL 启动子的变体 amyL 启动子片段。

在本发明中，亲代 amyL 启动子是 (a) 具有与 SEQ ID NO: 1 有最少 70% 同一性的核苷酸序列的多核苷酸；或 (b) 具有至少在低严谨条件下与 SEQ ID NO: 1 或其互补链杂交的核苷酸序列的多核苷酸。

在第一个方案中，亲代 amyL 启动子包含与 SEQ ID NO: 1 具有最少 70%，优选至少 75%，更优选至少 80%，更优选至少 85%，更加优选至少 90%，最优选至少 95%，或者甚至最优选至少 97% 同一性程度的核苷酸序列（在下文中指“同源 amyL 启动子”）。

优选地，亲代 amyL 启动子包含 SEQ ID NO: 1 的核苷酸序列或其具有启动子活性的片段。在优选实施例中，亲代 amyL 启动子包含 SEQ ID NO: 1 的核苷酸序列。在另一个优选实施例中，亲代 amyL 启动子由 SEQ ID NO: 1 的核苷酸序列组成。

在第二个方案中，亲代 amyL 启动子是在低严谨条件下，优选中等严谨条件下，更优选中等-高严谨条件下，更加优选高严谨条件下和最优选很高的严谨条件下与 SEQ ID NO: 1 或它的互补链杂交的核苷酸序列（J. Sambrook, E.F. Fritsch, and T. Maniatus, 1989, Molecular Cloning, A Laboratory Manual, 2d edition, Cold Spring Harbor, New York）。在这里限定了这种严谨条件。

依照本领域众所周知的方法，可以用 SEQ ID NO: 1 的核苷酸序列或其片段来鉴定和克隆来自不同的属或种的菌株的同源 amyL 启动子。

本发明中，分离的变体 amyL 启动子包含与 SEQ ID NO: 1 具有最少 70%，优选至少 75%，更优选至少 80%，更优选至少 85%，更加优选至少 90%，最优选至少 95%，或者甚至最优选至少 97% 同一性程度的核苷酸序列。

对本发明来说，通过 Wilbur-Lipman 方法（Wilbur and Lipman, 1983, Proceedings of the National Academy of Science USA 80: 726-730），利用具有同一性表和下列多重比对参数：缺口计罚分 10，间隙长度计罚分 10 的 LASERGENE™ MEGALIGN™ 软件（DNASTAR, Inc., Madison, WI）确定两个核苷酸序列之间的同一性程度。成对地比对参数是 Ktuple=3，缺口罚分=3 和窗口=20。

通过定位诱变可以完成“共有”启动子的构建，产生更完美地符合确定的

枯草芽孢杆菌营养(vegetative)" σ A型"启动子的"-10"和"-35"区的共有序列的启动子。"-35"区域的共有序列是 TTGACA, "-10"区域的共有序列是 TATAAT。可以从能够在芽孢杆菌宿主细胞中起作用的任何启动子中获得共有序列。

在优选方案中, "共有"启动子是从由大肠杆菌 lac 操纵子、天蓝色链霉菌琼脂糖酶基因(dagA)、迟缓芽孢杆菌碱性蛋白酶基因(aprH)、地衣芽孢杆菌碱性蛋白酶基因(枯草芽孢杆菌蛋白酶 Carlsberg 基因)、枯草芽孢杆菌果聚糖蔗糖酶基因(sacB)、枯草芽孢杆菌 α -淀粉酶基因(amyE)、地衣芽孢杆菌 α -淀粉酶基因(amyL)、嗜热脂肪芽孢杆菌生麦芽糖淀粉酶基因(amyM)、解淀粉芽孢杆菌 α -淀粉酶基因(amyQ)、地衣芽孢杆菌青霉素酶基因(penP)、枯草芽孢杆菌 xylA 和 xylB 基因、苏云金芽孢杆菌亚种 tenebrionis (*Bacillus thuringiensis* subsp.tenebrionis) cryIIIA 基因 (SEQ ID NO: 41) 或其部分, 或者原核生物的 β -内酰胺酶基因获得的启动子中得到的。还可以从 spo1 细菌噬菌体启动子中获得"共有"启动子。

在更优选的方案中, "共有"启动子是从解淀粉芽孢杆菌 α -淀粉酶基因(amyQ)中获得的。在最优选的方案中, 共有启动子是包含在 SEQ ID NO: 42 或 SEQ ID NO: 43 的核苷酸 1-185 中的"共有"amyQ 启动子。在另一个最优选的方案中, 共有启动子是包含在 SEQ ID NO: 42 或 SEQ ID NO: 43 的核苷酸 86-185 中的短"共有"amyQ 启动子。SEQ ID NO: 42 的"共有"amyQ 启动子包含含有野生型 amyQ 启动子 (SEQ ID NO 44) 核苷酸序列的下列突变: 在-35 区域(相对于转录起始位点)中的位置 135 和 136 分别出现的 T \rightarrow A 和 T \rightarrow C, 和 SEQ ID NO: 44 的-10 区域的位置 156 的 A \rightarrow T 变化。SEQ ID NO: 43 的"共有"amyQ 启动子更进一步地在位置 116, 即-35 区域上游大约 20 个碱基处, 包含 T \rightarrow A 变化 (SEQ ID NO: 43), 其中变化对启动子功能显然没有不利影响, 因为能很好地从关键的-10 和-35 区域除去它。

在优选方案中, 三联启动子包含具有对应于 SEQ ID NO: 1 的位置 590 的突变的变体 amyL 启动子。在另一个优选方案中, 三联启动子包含 SEQ ID NO: 1 的变体 amyL 启动子。在另一个优选方案中, 三联启动子包含"-35"区域具有序列 TTGACA 和"-10"区域具有 TATAAT 的共有 amyQ 启动子。在另一个优选方案中, 三联启动子包含"-35"区域具有序列 TTGACA 和"-10"区域具有 TATAAT 的短共有 amyQ 启动子。在另一个优选方案中, 三联启动子包含 cryIIIA 启动子或其部分 (Agaisse and Lereclus, 1994, Molecular

Microbiology 13: 97-107)。

在更优选的方案中，三联启动子包含具有对应于 SEQ ID NO: 1 的位置 590 的突变的变体 amyL 启动子，“-35”区域具有序列 TTGACA 和“-10”区域具有 TATAAT 的短共有 amyQ 启动子和 cryIIIA 启动子。

在另一个更优选的方案中，三联启动子包含以上述 5'到 3'次序排列的具有对应于 SEQ ID NO: 1 的位置 590 的突变的变体 amyL 启动子，“-35”区域具有序列 TTGACA 和“-10”区域具有 TATAAT 的短共有 amyQ 启动子和 cryIIIA 启动子。

在最优选的方案中，三联启动子包含 SEQ ID NO: 1 的变体 amyL 启动子，“-35”区域具有序列 TTGACA 和“-10”区域具有 TATAAT 的短共有 amyQ 启动子和 SEQ ID NO: 41 的 cryIIIA 启动子。

在另一个最优选的方案中，三联启动子包含以上述 5'到 3'次序排列的 SEQ ID NO: 1 的变体 amyL 启动子，“-35”区域具有序列 TTGACA 和“-10”区域具有 TATAAT 的短共有 amyQ 启动子和 SEQ ID NO: 41 的 cryIIIA 启动子。

"mRNA 加工/稳定序列"在这里限定为位于三联启动子的一个或多个启动子序列下游，位于一个或多个编码序列上游的序列，其中三联启动子序列中的每一个都可操作地连接到一个或多个编码序列上，因此可以处理从一个或多个启动子序列合成的全部 mRNAs 以产生在转录产物 5'端有稳定序列的 mRNA 转录产物。在 mRNA 转录产物 5'端存在这种稳定序列会增加 mRNA 转录产物的半衰期(Agaisse and Lereclus, 1994, supra, Hue et al., 1995, Journal of Bacteriology 177: 3465-3471)。mRNA 加工/稳定序列与细菌 16S 核糖体 RNA 的 3'端互补。在优选方案中，mRNA 加工/稳定序列产生在转录产物 5 端具有稳定序列的大小基本均一的转录产物。在另一个优选方案中，mRNA 加工/稳定序列位于整个三联启动子下游和一个或多个编码序列上游。在另一个优选方案中，mRNA 加工/稳定序列位于三联启动子的 cryIIIA 启动子下游和一个或多个编码序列上游。

mRNA 加工/稳定序列优选位于三联启动子下游和一个或多个涉及透明质酸生物合成的编码序列的上游。但是，mRNA 加工/稳定序列可以位于三联启动子中任何启动子序列的下游和一个或多个涉及透明质酸生物合成的编码序列的上游。而且，mRNA 加工/稳定序列可以位于三联启动子的每一个启动子序列的下游和一个或多个涉及透明质酸生物合成的编码序列的上

游。mRNA 处理稳定序列或序列可能与三联启动子的一个或多个启动子序列异源和/或彼此异源。

在优选方案中，mRNA 加工/稳定序列是 WO 94/25612 和上文的 Agaisse 和 Lereclus, 1994 中公开的苏云金芽孢杆菌 cryIIIA mRNA 加工/稳定序列或其保留 mRNA 处理/稳定功能的部分。在另一个更优选的方案中，mRNA 加工/稳定序列是上文 Hue 等, 1995 中公开的枯草芽孢杆菌 SP82 的 mRNA 加工/稳定序列或其保留 mRNA 处理/稳定功能的部分。

当 cryIIIA 启动子和其 mRNA 加工/稳定序列被用于本发明的方法时，可以使用包含 WO 94/25612 和上文 Agaisse 和 Lereclus, 1994 中公开的序列，通过 SEQ ID NO: 41 的核苷酸-635 到-22 描绘的 DNA 片段或者其保留启动子和 mRNA 处理/稳定功能的部分。当 cryIIIA mRNA 加工/稳定序列被包含在核苷酸-551 到-22 中时，用核苷酸-635 到-552 来描绘 cryIIIA 启动子。在优选方案中，cryIIIA mRNA 加工/稳定序列包含在含有核苷酸-568 到-22 的片段中。在另一个优选方案中，cryIIIA mRNA 加工/稳定序列包含在含有核苷酸-367 到-21 的片段中。而且，可以利用本领域众所周知的方法制备仅包含 cryIIIA 启动子和/或仅包含 cryIIIA mRNA 加工/稳定序列的 DNA 片段，以构建各种三联启动子和 mRNA 加工/稳定序列组合。

在优选方式中，cryIIIA 启动子和它的 mRNA 加工/稳定序列优选放置在组成三联启动子的其他启动子序列的下游和一个或多个编码序列的上游。

在更优选方案中，三联启动子包含具有对应于 SEQ ID NO: 1 的位置 590 的突变的变体 amyL 启动子，其中启动子序列是任何次序的"-35"区域具有序列 TTGACA 和"-10"区域具有 TATAAT 的短共有 amyQ 启动子和 cryIIIA 启动子，以及 cryIIIA mRNA 加工/稳定序列。

在另一个更优选方案中，三联启动子包含 SEQ ID NO: 1 的变体 amyL 启动子，其中启动子序列是任何次序的"-35"区域具有序列 TTGACA 和"-10"区域具有 TATAAT 的短共有 amyQ 启动子和 SEQ ID NO: 41 的 cryIIIA 启动子，以及 cryIIIA mRNA 加工/稳定序列。

在最优选的方案中，三联启动子包含以上述 5'到 3'次序排列的具有对应于 SEQ ID NO: 1 的位置 590 的突变的变体 amyL 启动子，"-35"区域具有序列 TTGACA 和"-10"区域具有 TATAAT 的短共有 amyQ 启动子，cryIIIA 启动子和 cryIIIA mRNA 加工/稳定序列。

在另一个最优选的方案中，三联启动子包含以上述 5'到 3'次序排列的

SEQ ID NO : 1 的变体 amyL 启动子, "-35"区域具有序列 TTGACA 和"-10"区域具有 TATAAT 的短共有 amyQ 启动子, SEQ ID NO : 41 的 cryIIIA 启动子和 cryIIIA mRNA 加工/稳定序列。

通过将一个或多个编码序列可操作地连接到一个或多个附加的控制序列上,可以更进一步地操作涉及透明质酸生物合成的一个或多个编码序列,其中控制序列指导编码序列在与控制序列相适合的情形下在芽孢杆菌细胞中表达。应理解表达包括涉及多肽生产的任何步骤,包括但不限于转录、转录后修饰、翻译、翻译后修饰和移位。利用克隆方法修饰核苷酸序列的技术在本领域为大家所熟知。

术语"控制序列"在这里限定为包括表达编码序列必需的或对编码序列表达有利的所有组件。每个控制序列可以是一个或多个编码序列的天然控制序列或者与一个或多个编码序列异源。除较早描述的三联启动子之外,这种控制序列包括,但不局限于引导序列,信号序列和转录终止子。可以用连接子来提供控制序列,这样可以插入特异的限制性内切位点,便于将控制序列与编码区连接。

控制序列也可以是适当的被芽孢杆菌细胞识别后能终止转录的转录终止序列。终止序列可操作地与编码多肽的多核苷酸的 3'末端相连。在选择的芽孢杆菌细胞中有功能的任何终止子都可以用于本发明。

控制序列也可以是适当的引导序列,引导序列是对芽孢杆菌细胞进行的翻译很重要的 mRNA 非翻译区。引导序列可操作地与编码多肽的多核苷酸的 5'末端相连。在选择的芽孢杆菌细胞中有功能的任何引导序列都可以用于本发明。

控制序列也可以是编码与转移表达的多肽,例如将表达的多肽转移到细胞膜的多肽的氨基末端连接的氨基酸序列的信号肽编码区。信号肽编码区可以是多肽与生俱来的或者是从外来来源中获得的。编码序列的 5'端可以固有地包含在翻译阅读框中天然地与编码移位多肽的编码区的片段连接的信号肽编码区。或者,编码序列的 5'端可以包含与编码移位多肽的编码序列的部分异源的信号肽编码区。当编码序列通常不包含信号肽编码区时,可能需要外来的信号肽编码区。或者,为了使多肽的移位相对于通常与编码序列有关的天然信号肽编码区有所增强,可以用外来的信号肽编码区简单地替换天然的信号肽编码区。能够指导表达的多肽移位的任何信号肽编码区都可以用于本发明。

芽孢杆菌细胞可以包含至少两个不同核酸构建体的一个或多个拷贝，其中按照上文的描述构建每个构建体。

核酸构建体可以更进一步地包含一个或多个允许容易选择转化细胞的可选择标记。可选择标记是其产物提供生物杀伤剂抗性，重金属抗性和营养缺陷原养型等等的基因。细菌的可选择标记的实例有来自枯草芽孢杆菌或地衣芽孢杆菌的 *dal* 基因，或者是赋予抗生素抗性，例如氨基青霉素、卡那霉素、红霉素、氯霉素或四环素抗性的标记。而且，可以通过如 WO 91/09129 中描述的共转化完成选择，在共转化中可选择标记位于分离的载体上。

本发明的特殊优点是可以产生不含外来可选择标记的芽孢杆菌细胞，也就是说，在将核酸构建体导入芽孢杆菌细胞之后，可以从芽孢杆菌细胞上删除可选择标记基因以产生没有标记的细胞。因为调节和环境方面的原因，除去可选择标记基因产生没有这种标记的芽孢杆菌细胞可能是优选的。

可以用基因缺失或取代技术完全去除可选择标记基因。例如：利用构建的包含毗连的（contiguous）侧接可选择标记基因的 5'和 3'区的质粒，通过同源重组可以完成可选择标记基因的缺失。可以在允许质粒定居在细胞中的许可温度条件下，将与第二可选择标记相联系的温度敏感质粒，例如 pE194 上连续的 5'和 3'区引入芽孢杆菌细胞。然后将细胞转入非许可温度以选择质粒在一个同源侧接区被整合到染色体中的细胞。质粒整合的选择受第二可选择标记的影响。在整合以后，通过将细胞改变到许可温度培养几代，不进行选择可以促进第二同源侧接区的重组事件。将细胞接种到平皿获取单个集落，检查丧失两个可选择标记的集落（例如，参见 Perego, 1993, In A.L. Sonneshein, J.A. Hoch, and R. Losick, editors, *Bacillus subtilis and Other Gram-Positive Bacteria*, Chapter 42, American Society of Microbiology, Washington, D.C., 1993）。也可以使用本领域众所周知的其他方法。

表达载体

在本发明的方法中，可以利用包含涉及透明质酸生物合成的一个或多个基因，三联启动子，或者也同时包含 mRNA 加工/稳定序列，转录和翻译终止信号的重组表达载体来进行重组生产。可以将如上所述的各种核酸和控制序列连接在一起产生包含一个或多个方便的限制性内切位点的重组表达载体，其中限制性内切位点允许在该位点插入或替换多核苷酸。或者，

通过将一个或多个涉及透明质酸生物合成的基因或包含一个或多个基因的核酸构建体插入到适当的表达载体中可以表达一个或多个涉及透明质酸生物合成的基因。在构建表达载体的过程中，一个或多个涉及透明质酸生物合成的基因位于载体中，这样一个编码序列或多个编码序列可操作地与三联启动子，或者三联启动子和 mRNA 加工/稳定序列，以及其他任何适当的表达控制序列和可能的移位控制序列连接。

重组体表达载体可以是能方便地经受重组 DNA 过程和能使涉及透明质酸生物合成的一个或多个基因在导入芽孢杆菌细胞中时表达的任何载体。典型地，载体的选择取决于载体与载体被插入其中的芽孢杆菌细胞的相容性。载体可以是线性质粒或闭合环状质粒。载体可以是自主复制载体，即作为染色体外实体存在的载体，例如质粒，染色体外元件，微型染色体或人工染色体，这种载体的复制不依赖染色体复制。载体可以包含任何保证自我复制的工具。换句话说，载体可以是被引入芽孢杆菌细胞时，能被整合到染色体中并和已经整合它的染色体一起复制的载体。载体系统可以是单个载体或质粒或两个或多个载体或质粒，或转座子。

"导入"指将包含涉及透明质酸生物合成的一个或多个基因的载体引入到芽孢杆菌细胞中，以便以染色体整合体或自我复制的染色体外载体来维持该载体。因为一个或多个编码序列或基因更有可能稳定地维持在细胞中，因此通常认为整合是一个优点。通过同源重组，非同源重组或转位可以将载体整合到染色体中。

将表达载体导入芽孢杆菌细胞可以受原生质体转化，使用感受态细胞，电穿孔或这里描述的结合作用的影响。

为了进行整合作用，载体可以依靠载体的任何成分以通过同源重组将载体稳定整合到基因组中。载体可以包含指导通过同源重组整合到芽孢杆菌细胞基因组的附加多核苷酸序列。附加多核苷酸序列使载体能够在染色体的精确位置整合到芽孢杆菌细胞基因组中。为了增加在精确位置整合的可能性，整合成分应优选包含足够量的核酸，例如 100-10,000 个碱基对，优选 400-10,000 个碱基对，最优选 800-10,000 个碱基对，其中核酸与对应的目标序列高度同源以增强同源重组的概率。整合成分可以是与芽孢杆菌细胞基因组中的目标序列同源的任何多核苷酸序列。而且，整合成分可以是非编码或编码序列。

为了自主复制，载体可以更进一步地包含使载体能够在正讨论的芽孢

杆菌细胞中自主复制的复制起点。复制的细菌起点的实例有允许在大肠杆菌中复制的质粒 pBR322、pUC19、pACYC177 和 pACYC184 的复制起点和允许在芽孢杆菌属中复制的 pUB110、pE194、pTA1060 和 pAMB1。复制起点可以是具有使它在芽孢杆菌细胞中的功能变得对温度敏感的突变的复制起点（例如，参见 Ehrlich, 1978, Proceedings of the National Academy of Sciences USA 75:1433）。

用来连接如上所述的组件以构建重组表达载体的过程为本领域熟练技术人员所熟知（例如，参见 Sambrook et al., 1989, 上文）。

生产

在本发明的方法中，使用本领域已知的方法在适合于生产透明质酸的营养培养基中培养芽孢杆菌属宿主细胞。例如：可以在适当的培养基，以及允许表达涉及透明质酸合成的酶和分离透明质酸的条件，在实验室或工业发酵罐中通过摇瓶培养，小规模或大规模发酵（连续不断的发酵、分批发酵、分级-分批发酵或固态发酵）培养细胞。使用本领域已知的方法，在包含碳源，氮源和无机盐的合适的营养培养基中进行培养。可以直接从培养基中回收分泌的透明质酸。

可以利用本领域众所周知的方法分离产生的透明质酸。例如，通过包括但不限于离心、过滤、抽提、喷雾干燥、蒸发或沉淀的常规程序可以从营养培养基中分离透明质酸。然后通过本领域已知的多种方法，包括但不限于色谱法（例如离子交换、亲和、疏水性、色谱聚焦和排阻色谱（size exclusion））、电泳方法（例如，准备的等电聚焦）、有差别的可溶性（differential solubility）（例如，硫酸铵沉淀）或抽提（例如，参见 Protein Purification, J.-C. Janson and Lars Ryden, editors, VCH Publishers, New York, 1989）可以更进一步地纯化分离的透明质酸。

因为重组芽孢杆菌细胞的类透明质酸直接表达到培养基中，因此利用简单的处理就可以从培养基中分离出类透明质酸。首先，采用物理方法从培养基中除去芽孢杆菌细胞和细胞碎片。可以先稀释培养基，如果需要，以降低培养基的粘性。从培养基中除去细胞的许多方法，例如离心或微量过滤，本领域熟练技术人员已经已知。如果愿意，接下来可以过滤剩余的上清液，例如通过超滤以浓缩类透明质酸和从类透明质酸中除去小分子污染物。在除去细胞和细胞碎片后，通过已知的机制可以完成从培养基简单

沉淀类透明质酸。可以用盐、乙醇或者盐和乙醇的组合从滤液中沉淀类透明质酸。一旦还原成沉淀，通过物理手段可以很容易地从溶液中分离类透明质酸。换句话说，通过使用本领域已知的蒸发技术，例如喷雾干燥可以从滤出溶液干燥或浓缩类透明质酸。依照这里描述的修饰的咋唑方法可以确定由本发明的芽孢杆菌属宿主细胞产生的透明质酸的水平。

在本发明的方法中，当在相同的生产条件下培养时，芽孢杆菌细胞相对于仅包含可操作地与涉及透明质酸生物合成的一个或多个编码序列相连的三联启动子的一个启动子序列的芽孢杆菌细胞优选产生至少 25%以上、更优选产生至少 50%以上、更优选产生至少 75%以上、更优选产生至少 100%以上、甚至更优选产生至少 200%以上、最优选产生至少 300%以上、甚至最优选产生至少 400%以上的透明质酸。

缺失/破坏

可以用基因缺失或取代技术完全去除外来的或异源的可选择标记基因或其他不受欢迎的基因。在这种方法中，利用已经构建的包含连续的侧接可选择标记基因的 5'和 3'区的质粒，通过同源重组可以完成可选择标记基因的删除。例如，可以在允许质粒定居在细胞中的许可温度条件下，将与第二可选择标记相联系的温度敏感质粒，例如 pE194 上连续的 5'和 3'区引入芽孢杆菌细胞。然后将细胞转入非许可温度以选择质粒在一个同源侧接区被整合到染色体中那些细胞。整合质粒的选择受第二可选择标记的影响。在整合以后，通过将细胞改变到许可温度培养几代，不进行选择可以促进第二同源侧接区的重组事件。将细胞接种到平皿以获取单个集落，检查丧失两个可选择标记的集落（例如，参见 Perego, 1993, In A.L. Sonneshein, J.A. Hoch, and R. Losick, editors, *Bacillus subtilis and Other Gram-Positive Bacteria*, Chapter 42, American Society of Microbiology, Washington, D.C., 1993）。

将包含缺陷基因 5'和 3'区，但缺乏可选择标记基因的核酸片段引入突变细胞进行同源重组，随后在反选择培养基上进行选择也可以除去可选择标记基因。通过同源重组，包含可选择标记基因的缺陷基因被缺乏可选择标记的核酸片段替换。也可以使用本领域已知的其他方法。

如上所述的方法还可以用于删除或破坏不想要的基因。美国专利 No.5,891,701 公开了删除包含 spoIIAC、aprE、nprE 和 amyE 的几个基因的技术。

通过上述方法也可以除去其他不想要的生物化合物，例如 *cypX*（登记号 BG12580）和/或 *yvmC*（登记号 BG14121）合成的红颜料。

在优选方式中，芽孢杆菌宿主细胞没有标记外来或异源的可选择标记。在另一个优选方案中，芽孢杆菌属宿主细胞不产生 *cypX* 和 *yvmC* 合成的任何红颜料。

下列实施例更进一步地描述了本发明，不应认为实施例是限制本发明的范围。

实施例

引物和寡核苷酸

依照厂家的说明在 Applied Biosystems Model 394 Synthesizer（Applied Biosystems, Inc., Foster City, CA）上合成全部引物和寡核苷酸。

实施例 1：构建在 *amyL* 位点包含 P11 启动子/*amyL* 表达盒的地衣芽孢杆菌 MaTa2

通过标准基因替换方法，在地衣芽孢杆菌 SJ1904（美国专利 No.5,733,753）*amyL* 基因（基本上和登记号 M13256 一致，Gray et al., 1986, J. Bacteriol. 166: 635，除在位置 474 用 C 代替 T，在位置 523 用 G 代替 C，在位置 524 用 C 代替 G，以及在位置 768-770 用 AAA 代替 TTT 之外）的上游插入 P11 启动子（Pr_{short} "共有"amyQ/Pr_{cryIII A stab}，美国专利 No.6,255,076）。因为使质粒负载位于功能启动子下游的 *amyL* 编码区的片段存在困难，因此按照如下所述的两个步骤完成构建。首先，用缺乏 RBS 的复合（composite）启动子替换 *amyL* 启动子和 *amyL* 核糖体结合部位（RBS）（防止 *amyL* 表达）。然后进行第二次基因替换再导入 RBS，从而产生有功能的表达盒。

通过电穿孔将基于 pE194 的温度敏感质粒 pMRT064.1（WO 03/054163）引入地衣芽孢杆菌菌株 SJ1904（Xue et al., 1999, Journal of Microbiological Methods 34: 183-191）。在 28°C 孵育 24-48 小时后，将细胞接种到每毫升补充有 1ug 红霉素和 25ug 林可霉素的胰蛋白胍血琼脂培养基（TBAB）-琼脂平板上。TBAB 琼脂平板由每升 33g 胰蛋白胍血琼脂培养基组成。分离红霉素抗性转化体，并在存在红霉素（5ug/ml）的条件下在 50°C 的非许可温

度培养转化体。在这个温度，pE194 复制起点是非活性的。细胞仅通过将质粒整合到细菌染色体的 amyL 位点才能在存在红霉素的条件下生长。为了促进导致用 P11 启动子替换内源 amyL 启动子和随后丧失 RBS 的质粒丧失或 "looping out"，在 30°C 的许可温度，在 Luria-Bertani (LB) 培养基生长整合体几代，不进行选择。LB 培养基由每升 10g 胰蛋白胍，5g 酵母抽提物和 5g NaCl 组成。在这个温度，pE194 复制起点是有活性的，能促进从基因组切除质粒 (Molecular Biological Methods for Bacillus, edited by C.R. Harwood and S.M. Cutting, 1990, John Wiley and Sons Ltd.)。然后将细胞接种到非选择性 LB 琼脂平板上 (每升 LB 培养基加 15g Bacto 琼脂)，通过下列标准鉴定包含想要的启动子替换和丧失基于 pE194 的复制子的集落：(1) 在包含 0.5% 淀粉-azure (Sigma Chemical Co., St. Louis, MO)，加 TBAB 盖层的 TBAB 平皿上不能形成晕圈表明 P11 启动子存在，amyL RBS 丧失；和 (2) 红霉素敏感性表明基于 pE194 的复制子丧失。选择一个满足这三个标准的转化体，将其命名为地衣芽孢杆菌 SJ1904::pMRT064.1。

按照如下步骤恢复上述菌株中的 amyL RBS。用 PacI 消化质粒 pNBT23 (pDG268MCS Δ neo-Pr_{short} "共有"amyQ/Pr_{cryIII Δ} /cryIII Δ stab/SAV, 美国专利 No. 6,255,076)，用 T4 DNA 聚合酶 I (Roche Applied Science, Indianapolis, IN) 使末端变钝，然后用 SalI 进行消化。用 Ecl136II 和 SalI 消化质粒 pUC19 (Yanisch-Perron et al., 1985, Gene 33: 103-119)。利用 44 mM Tris Base, 44 mM 硼酸, 0.5 mM EDTA (0.5X TBE) 缓冲液，在 0.8% 琼脂糖凝胶上分辨消化物，依照厂家的说明利用 QIAquick DNA 抽提试剂盒 (QIAGEN Inc., Valencia, CA) 从凝胶中纯化来自 pUC19 的大些的载体片段 (大约 2661bp) 和来自 pNBT23 的较小的 cryIII Δ stab/aprH 5' 片段 (大约 1069bp)。依照厂家的说明，利用 T4 DNA 连接酶 (Roche Applied Science; Indianapolis, IN) 将两个纯化的片段连接在一起，连接的混合物被转化到大肠杆菌 SURE® 感受态细胞中 (Stratagene, Inc., La Jolla, CA)。在每毫升补充有 100ug 氨苄青霉素的 2X 酵母-胰蛋白胍 (2XYT) 琼脂平板上选择转化体。2XYT 平板由每升 16g 胰蛋白胍，10g 酵母抽提物，5g NaCl 和 15g Bacto 琼脂组成。依照厂家的说明，利用 Bio Robot 9600 (QIAGEN Inc., Valencia, CA) 从几个转化体中纯化质粒 DNA，用 EcoRI 加 SalI 消化质粒 DNA，再利用 0.5XTBE 缓冲液，在 0.8% 琼脂糖凝胶上进行分析。通过存在大约 1071 bp 的 EcoRI/SalI cryIII Δ stab/aprH 5' 片段鉴定出正确质粒，并将其命名为 pNBT28 (图 1)。

依照 Horton et al., 1989, Gene 77: 61-8 的方法, 通过重叠延伸拼接法 (SOE) 构建质粒 pMRT038。分别利用如下所示的引物对 733-45-1 和 733-45-2, 以及 733-68-1 和 733-70-1 PCR 扩增来自质粒 pDN1981 (美国专利 No.5,698,415) 的 amyL 启动子区和 5'编码序列。在由 pDN1981 DNA 1ng, 每种引物 0.4uM, dATP、dCTP、dGTP 和 dTTP 各 200uM, 具有 2.5mM MgCl₂ 的 1XPCR Buffer II (Applied Biosystems, Inc., Foster City, CA) 和 2.5 单位 AmpliTaq Gold™ DNA 聚合酶 (Applied Biosystems, Inc., Foster City, CA) 组成的 50ul 反应中进行 PCR 扩增。在 RoboCycler 40 热循环仪 (Stratagene, Inc., La Jolla, CA) 中进行反应, 反应程序为 95°C 10 分钟, 1 个循环; 95°C 1 分钟, 50°C 1 分钟, 25 循环; 72°C 1 分钟, 25 个循环; 和 72°C 7 分钟, 1 个循环。用 0.5XTBE 缓冲液, 在 0.8% 琼脂糖凝胶上目测检验 PCR 产物。amyL 启动子区和 5'编码序列的预期片段分别是大约 600 和 500bp。利用引物 733-45-1 和 733-70-1 扩增最后的 SOE 片段。利用 TA-TOPO 克隆试剂盒 (Stratagene, Inc., La Jolla, CA) 将最后的 SOE 片段克隆到 pCR2.1 中。在每毫升补充有 100ug 氨苄青霉素的 2XYT 琼脂平板上选择转化体, 并在 37°C 孵育 16 小时。利用 M13 正反向引物 (Invitrogen, Inc, Carlsbad, CA), 通过 DNA 测序验证携带正确质粒的转化体。质粒被命名为 pMRT038 (图 2)。

引物 733-45-1:

5'-GTCCTTCTTGGTACCTGGAAGCAGAGC-3' (SEQ ID NO: 45)

引物

733-45-2: 5'-GTATAAATATTCGGCCCTTAAGGCCAGTACCATTTTCCC-3'
(SEQ ID NO: 46)

引物 733-68-1:

5'-TGGTACTGGCCTTAAGGGCCGAATATTTATACAATATCATGAGCT
CCACATTGAA-3' (SEQ ID NO: 47)

引物

733-70-1: 5'-GGTGTTCTCTAGAGCGGCCGCGGTTGCGGTCAGC-3' (SEQ
ID NO: 48)

用 BglIII 消化质粒 pNBT28, 用 Klenow 片段使末端变钝, 然后用 SacI 进行消化。用 EcoRI 消化质粒 pMRT038, 用 Klenow 片段使末端变钝, 然后用 SacI 进行消化。利用 0.5XTBE 缓冲液, 在 0.8% 琼脂糖凝胶上分辨消化物, 依照厂家的说明利用 QIAquick DNA 抽提试剂盒从凝胶中纯化来自

pNBT28 的大些的载体片段(大约 3253bp)和来自 pMRT038 的较小片段(大约 593bp)。利用 T4 DNA 连接酶将两个纯化的片段连接在一起,连接的混合物被转化到大肠杆菌 SURE®感受态细胞中。在每毫升补充有 100ug 氨苄青霉素的 2XYT 琼脂平板上选择转化体。依照厂家的说明,利用 QIAGEN Bio Robot 9600 从几个转化体中纯化质粒 DNA,用 EcoRI 加 HindIII 消化质粒 DNA 后,再利用 0.5XTBE 缓冲液,在 0.8%琼脂糖凝胶上进行分析。通过存在大约 1168bp 的 EcoRI/HindIII 片段鉴定出正确质粒,并将其命名为 pNBT29 (图 3)。

用 EcoRI 和 HindIII 消化质粒 pNBT29 和 pCJ791 (WO 03/054163)。利用 0.5XTBE 缓冲液,在 0.8%琼脂糖凝胶上分辨消化物,依照厂家的说明利用 QIAquick DNA 抽提试剂盒从凝胶中纯化来自 pCJ791 的大些的载体片段(大约 4340bp)和来自 pNBT29 的较小片段(大约 1168bp)。利用 T4 DNA 连接酶将两个纯化的片段连接在一起,连接的混合物被转化到枯草芽孢杆菌 168Δ4 (WO 03/054163) 中。依照厂家的说明,利用 tip-20 柱子(QIAGEN Inc., Valencia, CA)从几个转化体中纯化质粒 DNA,用 EcoRI 加 HindIII 消化质粒 DNA 后,再利用 0.5XTBE 缓冲液,在 0.8%琼脂糖凝胶上进行分析。通过存在大约 1168bp 的 EcoRI/HindIII 片段鉴定出正确质粒,并将其命名为 pWWi001.1 (图 4)。

为了恢复地衣芽孢杆菌菌株 SJ1904::pMRT064.1 中的 amyL RBS,通过电穿孔将质粒 pWWi001.1 引入该菌株中,然后按照实施例 1 的描述将其合并从染色体中切除。通过 amyL 表达的恢复鉴定出想要的克隆,其中通过在包含 0.5%淀粉-azure 加 TBAB 盖层的 TBAB 平板上生长菌株来测定 amyL 的表达。产生淀粉酶的菌株在集落或碎片周围形成清晰的晕圈。

实施例 2:构建在天然 amyL 位点包含 P12 启动子/amyL 表达盒的地衣芽孢杆菌 MaTa3

用填充了 Klenow 片段的 HindIII 消化质粒 pMRT064.1,然后用 SfiI 进行消化。用 SfiI 和 Ecl136II 消化质粒 pNBT23 (美国专利 No.6,255,076)。利用 0.5XTBE 缓冲液,在 0.8%琼脂糖凝胶上分辨消化物,依照厂家的说明利用 QIAquick DNA 抽提试剂盒从凝胶中纯化来自 pMRT064.1 的大些的载体片段(大约 4892bp)和来自 pNBT23 的较小片段(大约 728bp)。利用 T4 DNA 连接酶将两个纯化的片段连接在一起,连接的混合物被转化到枯草芽孢杆

菌 168Δ4 中。依照厂家的说明, 利用 QIAGEN tip-20 柱子从几个转化体中纯化质粒 DNA, 用 EcoRI 加 HindIII 消化质粒 DNA 后, 再利用 0.5XTBE 缓冲液, 在 0.8%琼脂糖凝胶上进行分析。通过限制性内切酶和/或 PCR 分析验证正确的质粒, 并将其命名为 pWWi005 (图 5)。

通过电穿孔将质粒 pWWi005 引入地衣芽孢杆菌 MaTa2 中, 然后按照实施例 1 的描述将其整合并从染色体中切除, 其中用 P12 启动子 (Pr_{short} "共有 amyQ/Pr_{cryIII Λ} /cryIII Λ stab, 美国专利 No.6,255,076) 替换 P11 启动子产生地衣芽孢杆菌菌株 MaTa3。利用下文显示的引物 961197 和 94-935, 通过 PCR 分析鉴定想要的克隆。在由 200ng 地衣芽孢杆菌 MaTa3 染色体 DNA (按照 Pitcher et al., 1989, Letters in Applied Microbiology 8: 151-156 的描述分离), 每种引物 0.4uM, dATP、dCTP、dGTP 和 dTTP 各 200uM, 具有 2.5mM MgCl₂ 的 1X PCR Buffer II 和 2.5 单位 AmpliTaq Gold™ DNA 聚合酶组成的 50ul 反应中进行 PCR 扩增。在 RoboCycler 40 热循环仪中进行反应, 反应程序为 95°C 10 分钟, 1 个循环; 95°C 1 分钟, 50°C 1 分钟, 25 循环; 72°C 1 分钟, 25 个循环; 和 72°C 7 分钟, 1 个循环。利用 0.5XTBE 缓冲液, 在 2%琼脂糖凝胶上目测检验 PCR 产物。预期片段是大约 230bp。

引物 961197:

5'-GGCCTTAAGGGCCTGCTGTCCAGACTGTCCGCT-3' (SEQ ID NO: 49)

引物 94-935:

5'-GGCGTTACAATTCAAAGA-3' (SEQ ID NO: 50)

实施例 3: 构建在天然 amyL 位点包含 P17 启动子/amyL 表达盒的地衣芽孢杆菌 MDT217

利用如下所示的引物, 从地衣芽孢杆菌 SJ1904(美国专利 No.5,733,753) 的染色体 DNA PCR 扩增 amyL4199 启动子 (Pr_{amyL4199}, 美国专利 No.6,100,063), 此后被命名为 P6 启动子), 其中引物分别整合了 SfiI 位点和 SacI 位点。

引物 950872:

5'-CCAGGCCTTAAGGGCCGCATGCGTCCTTCTTTGTGCT-3' (SEQ ID NO: 51)

引物 991151:

5'-GAGCTCCTTTCAATGTGATACATATGA-3' (SEQ ID NO : 52)

在由 50ng 地衣芽孢杆菌 SJ1904 染色体 DNA (按照上文 Pitcher et al., 1989 的描述获取), 950872 和 991151 引物各 0.4 uM, dATP、dCTP、dGTP 和 dTTP 各 200 uM, 具有 2.5 mM MgCl₂ 的 1X PCR Buffer II 和 2.5 单位 AmpliTaq Gold™ DNA 聚合酶组成的 50ul 反应中进行一式三份的 PCR 扩增。在 RoboCycler 40 热循环仪中进行反应, 反应程序为 95°C 9 分钟, 1 个循环; 95°C 1 分钟, 55°C 1 分钟, 30 循环; 72°C 1 分钟, 30 个循环; 和 72°C 5 分钟, 1 个循环。用 0.5XTBE 缓冲液, 在 0.8% 琼脂糖凝胶上目测检验 PCR 产物。预期片段是大约 600bp。

利用 TA-TOPO 克隆试剂盒将 600bp 的 PCR 片段克隆到 pCR2.1 中, 并依照厂家的说明将其转化到大肠杆菌 OneShot™ 感受态细胞中 (Stratagene, Inc., La Jolla, CA)。37°C 在每毫升补充有 100ug 氨苄青霉素的 2XYT 琼脂平板上生长 16 小时后, 选择出转化体。依照厂家的说明, 利用 QIAGEN Bio Robot 9600 从这些转化体中纯化质粒 DNA, 用 EcoRI 消化质粒 DNA 后, 再利用 0.5XTBE 缓冲液, 在 0.8% 琼脂糖凝胶上进行分析。发现质粒具有正确的片段 (大约 3913bp 和 640bp)。利用 M13 (-20) 正向引物和 M13 反向引物 (Invitrogen, Inc, Carlsbad, CA), 通过 DNA 测序确认插入 DNA 的 DNA 序列。具有正确克隆序列的质粒被命名为 pNBT30 (图 6)。

用 SfiI 和 SacI 消化质粒 pNBT11 (pDG268MCSΔneo-Pr_{cryIII_A}/SAV, 美国专利 No.6,255,076) 和 pNBT30。利用 0.5XTBE 缓冲液, 在 0.8% 琼脂糖凝胶上分辨消化物, 依照厂家的说明利用 QIAquick DNA 抽提试剂盒从凝胶中纯化来自 pNBT11 的大些的载体片段 (大约 7931bp) 和来自 pNBT30 的较小片段 (大约 611bp)。利用 T4 DNA 连接酶将两个纯化的片段连接在一起, 连接的混合物被转化到大肠杆菌 SURE® 感受态细胞中。在每毫升补充有 100ug 氨苄青霉素的 2XYT 琼脂平板上选择转化体。依照厂家的说明, 利用 QIAGEN Bio Robot 9600 从这些转化体中纯化质粒 DNA, 用 NcoI 消化质粒 DNA 后, 再利用 0.5XTBE 缓冲液, 在 0.8% 琼脂糖凝胶上进行分析。发现质粒具有正确的片段 (大约 6802 bp 和 1741 bp), 并将其命名为 pNBT31 (图 7)。

Pr_{amyL4199}/Pr_{short}"共有"amyQ/cryIII_Astab 复合启动子 (P16) 的构建如下。用 DraIII 和 Ecl136II 消化质粒 pNBT31。利用 SfiI 消化质粒 pNBT24 (pDG268MCSΔneo-Pr_{short}"共有"amyQ/长 cryIII_Astab/SAV, 美国专利 No.

6,255,076), 利用 T4 DNA 聚合酶 I 使末端变钝, 然后用 DraIII 进行消化。利用 0.5XTBE 缓冲液, 在 0.8%琼脂糖凝胶上分辨消化物, 依照厂家的说明利用 QIAquick DNA 抽提试剂盒从凝胶中纯化来自 pNBT31 的大些的载体片段和来自 pNBT24 的较小启动子片段(大约 1100bp)。利用 T4 DNA 连接酶将两个纯化的片段连接在一起, 依照厂家的说明将连接的混合物转化到大肠杆菌 SURE®感受态细胞中。在每毫升补充有 100ug 氨苄青霉素的 2XYT 琼脂平板上选择转化体。依照厂家的说明, 利用 QIAGEN Bio Robot 9600 从几个转化体中纯化质粒 DNA, 用 SfiI 加 SacI 消化质粒 DNA 后, 再利用 0.5XTBE 缓冲液, 在 0.8%琼脂糖凝胶上进行分析。通过存在大约 1400bp 的 SfiI/SacI 片段鉴定出正确质粒, 并将其命名为 pNBT33 (图 8)。

用 AvaII 和 Ecl136II 消化质粒 pNBT33。用 NotI 消化质粒 pMRT074(WO 03/054163), 利用 T4 DNA 聚合酶使末端变钝, 然后用 AvaII 进行消化。利用 0.5XTBE 缓冲液, 在 0.8%琼脂糖凝胶上分辨消化物, 依照厂家的说明利用 QIAquick DNA 抽提试剂盒从凝胶中纯化来自 pMRT074 的大些的载体片段(大约 4433bp)和来自 pNBT30 的较小片段(大约 1157bp)。利用 T4 DNA 连接酶将两个纯化的片段连接在一起, 将连接的混合物转化到枯草芽孢杆菌 168Δ4 中。在 28℃孵育 24-48 小时后, 在每毫升补充有 1ug 红霉素和 25ug 林可霉素的 TBAB-琼脂平板上选择转化体。依照厂家的说明, 利用 QIAGEN tip-20 柱子从几个转化体中分离质粒 DNA, 通过用 BamHI 加 HindIII 消化质粒进行验证。产生的质粒被命名为 pMDT006 (图 9)。

通过电穿孔将质粒 pMDT006 引入地衣芽孢杆菌 MaTa3。分离红霉素抗性转化体, 按照实施例 1 的描述整合质粒并从染色体中切除质粒, 其中导致用 P16 启动子替换 P12 串联启动子产生地衣芽孢杆菌菌株 MDT216。地衣芽孢杆菌 MDT216 基本上是地衣芽孢杆菌 MDT206 的利福平敏感型式(参见实施例 6)。利用引物 94-935 和 94-919 (上述的), 通过 PCR 分析鉴定想要的克隆。

用 SfiI 和 EcoRI 消化质粒 pWWi005 (实施例 2) 以除去位于 amyL 基因上游的 DNA 序列。用 T4 DNA 聚合酶使末端钝化, 在 0.8%琼脂糖凝胶上利用 0.5XTBE 缓冲液分辨片段。依照厂家的说明利用 QIAquick DNA 抽提试剂盒从凝胶中纯化大些的载体片段(大约 5069bp)。利用 T4 DNA 连接酶将纯化的片段连接在一起, 将连接的混合物转化到枯草芽孢杆菌 168Δ4 中。在 28℃孵育 24-48 小时后, 在每毫升补充有 1ug 红霉素和 25ug 林可霉素的

TBAB-琼脂平板上选择转化体。依照厂家的说明,利用 QIAGEN tip-20 柱子从几个转化体中分离质粒 DNA,通过用 BamHI 加 HindIII 消化质粒进行验证。产生的质粒被命名为 pMDT007 (图 10)。

通过电穿孔将质粒 pMDT007 引入地衣芽孢杆菌 MDT216。分离红霉素抗药性转化体,按照实施例 1 的描述整合质粒并从染色体中切除质粒,导致 P16 启动子转换成 Pr_{amyL4199}/Pr_{short}共有"amyQ/Pr_{cryIII A}/cryIII A_{stab} 三联的串联启动子 (P17) 产生地衣芽孢杆菌菌株 MDT217。利用如上所述的引物 94-935 和 94-919,通过 PCR 分析鉴定想要的克隆。

实施例 4:地衣芽孢杆菌 MDT217 的 C-组件蛋白酶被删除的衍生物--地衣芽孢杆菌 MDT220 的构建

通过按照如下所述的方法删除地衣芽孢杆菌 MDT217 (实施例 3) 的 C-组件 (component) 蛋白酶基因构建地衣芽孢杆菌 MDT220,地衣芽孢杆菌 MDT220 是包含在 P17 三联串联启动子控制下的 amyL 基因 (登记号 M13256) 的 C-组件阴性菌株。

利用 SOE (Horton et al., 1989, 上文), 通过 PCR 构建编码地衣芽孢杆菌 C-组件蛋白酶的基因被删除的型式 (美国专利 No.5,459,064, 登记 no.D10060, Kakudo et al., 1992, J. Biol. Chem. 267: 23782)。利用显示如下的扩增 5'C-组件片段的引物 991173 (插入 5'EcoRI 限制性内切位点) 和引物 991174, 以及扩增 3'C-组件片段的引物 991175 和 991176 (插入 3'HindIII 限制性内切位点), 通过 PCR 扩增 C-组件基因的 5'和 3'区。

引物 991173:

5'-GAATTCGACGGCTTCCCGTGCGCC-3' (SEQ ID NO: 53)

引物 991174:

5'-GCAAGCGAGCACGGATTGTAAGTACAAGTTAGATA-3' (SEQ ID NO: 54)

引物 991175:

5'-AACTTGTA CTTACAATCCGTGCTCGCTTGCCGTAC-3' (SEQ ID NO: 55)

引物 991176:

5'-AAGCTTCCATTCAAACCTGGTGAGGAAG-3' (SEQ ID NO: 56)

在由 50ng 地衣芽孢杆菌 SJ1904 染色体 DNA (按照上文 Pitcher et al.,

1989 的描述获取), 扩增 5'C-组件片段的 991173 和 991174 引物对或扩增 3'C-组件片段的引物对 991175 和 991176 中的每种引物各 0.4 uM, dATP、dCTP、dGTP 和 dTTP 各 200 uM, 具有 2.5 mM MgCl₂ 的 1X PCR Buffer II 和 2.5 单位 AmpliTaq Gold™ DNA 聚合酶组成的 30ul 反应中进行一式三份的 PCR 扩增。在 RoboCycler 40 热循环仪中进行反应, 反应程序为 95°C 9 分钟, 1 个循环; 95°C 1 分钟, 52°C 1 分钟, 72°C 1 分钟, 3 个循环; 95°C 1 分钟, 55°C 1 分钟, 72°C 1 分钟, 27 个循环; 和 72°C 5 分钟, 1 个循环。用 0.5XTBE 缓冲液, 在 0.8% 琼脂糖凝胶上目测检验 PCR 产物。预期片段的大小是大约 290 bp。依照 Horton et al., 1989 的描述, 利用引物对 991173 和 991176 产生最后的 SOE 片段, 并利用 TA-TOPO 克隆试剂盒将片段克隆到 pCR2.1 中。在 37°C 孵育 16 小时之后, 在每毫升补充有 100ug 氨苄青霉素的 2XYT 琼脂平板上选择转化体。依照厂家的说明, 利用 QIAGEN tip-20 柱子从几个转化体中分离质粒 DNA, 利用 M13 (-20) 正向引物和 M13 反向引物通过 DNA 测序进行验证。包含 SOE 片段的质粒被命名为 (图 11)。

用 EcoRI 和 HindIII 消化质粒 pCJ791 (WO 03/054163) 和 pNBT37。利用 0.5XTBE 缓冲液, 在 0.8% 琼脂糖凝胶上分辨消化物, 依照厂家的说明利用 QIAquick DNA 抽提试剂盒从凝胶中纯化来自 pCJ791 的大些的载体片段 (大约 4340bp) 和来自 pNBT37 的较小的 C-组件删除片段 (大约 580bp)。利用 T4 DNA 连接酶将两个纯化的片段连接在一起, 将连接的混合物转化到枯草芽孢杆菌 168Δ4 中。在 28°C 孵育 24-48 小时后, 在每毫升补充有 1ug 红霉素和 25ug 林可霉素的 TBAB-琼脂平板上选择转化体。依照厂家的说明, 利用 QIAGEN tip-20 柱子从几个转化体中分离质粒 DNA, 通过用 EcoRI 加 HindIII 消化和琼脂糖凝胶电泳进行验证。鉴定出产生大小为大约 4340bp 和大约 580bp 的预期片段的质粒, 并将其命名为 pNBT38 (图 12)。

通过电穿孔将质粒 pNBT38 引入地衣芽孢杆菌 MDT217。分离红霉素抗性转化体, 按照实施例 1 的描述整合质粒并从染色体中切除质粒。依照上文 Pitcher et al., 1989 的描述从几个转化体中分离染色体 DNA, 利用如上所述的引物 991173 和 991176 通过 PCR 分析染色体 DNA 以鉴定 C-组件删除菌株。鉴定出产生大约 580bp 的预期大小的 PCR 片段的几个转化体, 大约 580bp 的 PCR 片段证实 C-组件基因部分缺失。选择一个这样的转化体, 并将其命名为 MDT220。

实施例 5：构建质粒 pTH012

利用下列引物,通过 PCR 从芽孢杆菌 JP170 的染色体 DNA 扩增编码芽孢杆菌 JP170 碱性蛋白酶的基因 (WO 98/56927 和美国专利 No. 5891701, 登记号 AR069954)。

引物 992843：

5'-CGAGCTCGATGTGTTATAAATTGAGAGGAG-3' (SEQ ID NO : 57)

引物 961252：

5'-GCGGCCGCGTCATAAACGTTGCAATCGTGCTC-3' (SEQ ID NO : 58)

在由 50ng 芽孢杆菌 JP170 染色体 DNA(按照上文 Pitcher et al., 1989 的描述获取), 引物 992843 和 961252 各 0.4 uM, dATP、dCTP、dGTP 和 dTTP 各 200 uM, 具有 2.5 mM MgCl₂ 的 1X PCR Buffer II 和 2.5 单位 AmpliTaq Gold™ DNA 聚合酶组成的 50ul 反应中进行一式三份的 PCR 扩增。在 RoboCycler 40 热循环仪中进行反应, 反应程序为 95°C 9 分钟, 1 个循环; 95°C 1 分钟, 55°C 1 分钟, 30 循环; 72°C 1 分钟, 30 个循环; 和 72°C 5 分钟, 1 个循环。用 0.5XTBE 缓冲液, 在 0.8%琼脂糖凝胶上目测检验 PCR 产物。预期片段是大约 2163bp。

利用 TA-TOPO 克隆试剂盒将 2163bp 的 PCR 片段克隆到 pCR2.1 中, 并依照厂家的说明将其转化到大肠杆菌 OneShot™ 感受态细胞中。37°C 在每毫升补充有 100ug 氨苄青霉素的 2XYT 琼脂平板上生长 16 小时之后进行转化体选择。依照厂家的说明, 利用 QIAGEN Bio Robot 9600 从几个转化体中纯化质粒 DNA, 用 SfiI 加 NotI 消化质粒 DNA 后, 再利用 0.5XTBE 缓冲液, 在 0.8%琼脂糖凝胶上进行分析。鉴定具有正确大小插入物的质粒, 利用 M13 (-20) 正向引物和 M13 反向引物及下面的内在引物通过 DNA 测序确认插入物的 DNA 序列。鉴定具有正确序列的质粒, 并将其命名为 pNBT39 (图 13)。

引物 992843：

5'-CGAGCTCGATGTGTTATAAATTGAGAGGAG-3' (SEQ ID NO : 59)

引物 961021：

5'-GTCGAATATGATGGGGATG-3' (SEQ ID NO : 60)

引物 960898：

5'-GGACAAGGACAGATTGTAGCAGTTGCTGATACTGG-3' (SEQ ID

NO : 61)

引物 961048 :

5'-GCGATTACAGTTGGGGCAACCG-3' (SEQ ID NO : 62)

引物 961222 :

5'-GGTAGCACGACGGCATCACTAAC-3' (SEQ ID NO : 63)

按照如下步骤构建质粒 pMRT077 (图 19)。分别利用显示如下的引物对 733-45-1 和 733-45-7, 以及引物对 757-19-1 和 733-45-6 通过 SOE 将 amyL 基因的上游区和 3'区融合。

引物 733-45-1 :

5'-GTCCTTCTTGGTACCTGGAAGCAGAGC-3' (SEQ ID NO : 64)

引物 733-45-7 :

5'-CATGCTGGGCCCTTAAGGCCAGTACCATTTTCCC-3' (SEQ ID NO : 65)

引物 757-19-1 :

5'-CAGTAGGCCTTAAGGGCCCAGCATGATTGAGCTCACCAC
CATGGGATCCGCGG CCGCACAAGGGAAGGC-3' (SEQ ID NO : 66)

引物 733-45-6 :

5'-CAATTCATCCTCTAGAGTCTCAGG-3' (SEQ ID NO : 67)

在由 1ng pDN1981DNA (美国专利 No. 5,698,415), 每个引物各 0.4 uM, dATP、dCTP、dGTP 和 dTTP 各 200 uM, 具有 2.5 mM MgCl₂ 的 1X PCR Buffer II 和 2.5 单位 AmpliTaq Gold™ DNA 聚合酶组成的 50ul 反应中进行 PCR 扩增。在 RoboCycler 40 热循环仪中进行反应, 反应程序为 95°C 10 分钟, 1 个循环; 95°C 1 分钟, 50°C 1 分钟, 25 循环; 72°C 1 分钟, 25 个循环; 和 72°C 7 分钟, 1 个循环。利用 0.5XTBE 缓冲液, 通过在 0.8% 琼脂糖凝胶上的电泳目测检验 PCR 产物。预期片段分别是大约 600 和 500bp。利用引物 733-45-1 和 733-45-6 扩增最后的片段。利用 TA-TOPO 克隆试剂盒将最后的片段克隆到 pCR2.1 载体中。在每毫升补充有 100ug 氨苄青霉素的 2XYT 琼脂平板上选择转化体, 并在 37°C 孵育 16 小时。利用 M13 正反向引物通过 DNA 测序验证携带正确质粒的转化体。该质粒被命名为 pMRT040 (图 14)。

用 KpnI/XbaI 消化质粒 pMRT040, 并用 Klenow 片段补平(filled in), 然后依照厂家的说明利用 QIAquick DNA 抽提试剂盒从 0.8% 琼脂糖-0.5XTBE 凝胶中分离大约 1000bp 的片段。将该片段克隆到用 EcoRV 消化的 pShV3

(WO 03/054163) 中, 并依照厂家的说明将其转化到大肠杆菌 XL1Blue 细胞中 (Stratagene, Inc., La Jolla, CA)。在每毫升补充有 100ug 氨苄青霉素的 2XYT 琼脂平板上选择转化体, 并在 37°C 孵育 16 小时。依照厂家的说明, 利用 QIAGEN tip-20 柱子从几个转化体中分离质粒 DNA, 通过用 SacI/SphI 的限制性分析在利用 0.5XTBE 缓冲液的 0.8% 的琼脂糖凝胶上验证质粒 DNA。产生的质粒被命名为 pMRT044 (图 15)。

用 SacI 加 HindIII 消化质粒 pMRT044 和 pNBT3 (pDG268MCS Δ neo-Pr_{cryIIIA}stab/SAV, 美国专利 No. 6,255,076)。利用 0.5XTBE 缓冲液, 在 0.8% 琼脂糖凝胶上分辨消化物, 依照厂家的说明利用 QIAquick DNA 抽提试剂盒从凝胶中纯化来自 pMRT044 的大些的载体片段 (大约 7500bp) 和来自 pNBT3 的较小的 cryIIIA 的稳定片段 (大约 540bp)。利用 T4 DNA 聚合酶将两个纯化的片段连接转化一起, 依照厂家的说明将连接的混合物用于转化大肠杆菌 XL1Blue 细胞。在每毫升补充有 100ug 氨苄青霉素的 2XYT 琼脂平板上选择转化体, 并在 37°C 孵育 16 小时。依照厂家的说明, 利用 QIAGEN tip-20 柱子从几个转化体中分离质粒 DNA, 通过用 SacI/HindIII 的限制性分析在利用 0.5XTBE 缓冲液的 0.8% 的琼脂糖凝胶上验证质粒 DNA。产生的质粒被命名为 pMRT070 (图 16)。

用 EcoRI/HindIII 消化质粒 pMRT074 (WO 03/054163) 和 pMRT070。依照厂家的说明, 利用 0.5XTBE 缓冲液和 QIAquick DNA 纯化试剂盒从 0.8% 琼脂糖凝胶中分离来自 pMRT074 的大约 3500bp 的片段和来自 pMRT070 的大约 1100bp 的片段, 连接分离的片段, 并将他们转化到枯草芽孢杆菌 168 Δ 4 感受态细胞中。在每毫升补充有 1ug 红霉素和 25ug 林可霉素的 TBAB-琼脂平板上选择转化体, 并在 30°C 孵育 24-48 小时。依照厂家的说明, 利用 QIAGEN tip-20 柱子从几个转化体中分离质粒 DNA, 通过用 EcoRI/HindIII 的限制性分析在 0.8% 的琼脂糖凝胶上利用 0.5XTBE 缓冲液验证质粒 DNA。产生的质粒被命名为 pMRT075 (图 17)。

用 SacI 加 NotI 消化质粒 pMRT075 和 pNBT40 (图 18)。质粒 pNBT40 基本上是包含编码来自解淀粉芽孢杆菌的中性蛋白酶的基因 (npr [BamP]) 的 pCR2.1-TOPO 载体 (Vasanth et al., 1984, J. Bacteriol. 159: 811, accession no. K02497)。依照厂家的说明, 利用 0.5XTBE 缓冲液和 QIAquick DNA 纯化试剂盒从 0.8% 琼脂糖凝胶中分离来自 pMRT075 的大约 5500bp 的片段和来自 pNBT40 的大约 1600bp 的片段, 连接分离的片段, 并将他们转化到枯

草芽孢杆菌 168Δ4 感受态细胞中。在每毫升补充有 1%的脱脂乳，1μg 红霉素和 1μg 林可霉素的 TBAB-琼脂平板上选择转化体，并在 30℃ 孵育 24-48 小时。获取在 TBAB-琼脂脱脂乳平皿上产生清晰区带的转化体，产生的质粒被命名为 pMRT077 (图 19)。

用 SacI 加 NotI 消化质粒 pNBT39 和 pMRT077。利用 0.5XTBE 缓冲液，在 0.8%琼脂糖凝胶上分辨消化物，依照厂家的说明利用 QIAquick DNA 抽提试剂盒从凝胶中纯化来自 pMRT077 的大些的载体片段 (大约 5400bp) 和来自 pNBT39 的较小的 JP170 蛋白酶片段 (大约 2163bp)。利用 T4 DNA 连接酶将两个纯化的片段连接在一起，将连接的混合物转化到枯草芽孢杆菌 168Δ4 感受态细胞中。依照厂家的说明，利用 QIAGEN tip-20 柱子从几个转化体中纯化质粒 DNA，用 SacI 加 NotI 消化质粒 DNA 后，再利用 0.5XTBE 缓冲液，在 0.8%琼脂糖凝胶上进行分析。通过存在大约 2163 的 SfiI/NotI JP170 片段鉴定出正确质粒，并将其命名为 pTH012 (图 20)。

实施例 6：在 amyL 位点包含一个拷贝的 P17 启动子 /sc hasA/tuaD/gtaB 表达盒的地衣芽孢杆菌菌株 TH15 的构建

质粒 pMB748 包含融合到地衣芽孢杆菌 amyL 基因的 amyL 信号序列上的来自芽孢杆菌属 I633 (WO 99/64619) 的甘露聚糖酶基因的成熟部分。利用显示如下的引物对 80501D1B11 和 172965，以质粒 pMB748 作为模板 DNA，通过删除甘露聚糖酶基因的 3'部分构建该基因的截短型式。在由 50ng pMB748 DNA，引物 80501D1B11 和 172965 各 0.4 μM，dATP、dCTP、dGTP 和 dTTP 各 200 μM，具有 2.5 mM MgCl₂ 的 50X PCR Buffer II 和 2.5 单位 AmpliTaq Gold™ DNA 聚合酶组成的 50ul 反应中进行 PCR 扩增。在 RoboCycler 40 热循环仪中完成反应，反应程序为 95℃9 分钟，1 个循环；95℃1 分钟，55℃1 分钟，30 循环；72℃1 分钟，30 个循环；和 72℃5 分钟，1 个循环。用 0.5XTBE 缓冲液，在 0.8%琼脂糖凝胶上目测检验大约 900bp 的 PCR 产物。

引物 80501d1b11：

5'-CATTCTGCAGCCGCGGCAAATTCCGGATTTTATGTAAGCGG-3'

(SEQ ID NO : 68)

引物 172965：

5'-CATCATATGCGGCCGCTTATCATTGAAAAACGGTGCTTAATCTCG

AAG-3' (SEQ ID NO : 69)

PCR片段被克隆到用 SacII 和 NotI 消化的质粒 pMOL944(WO 99/64619) 中。依照厂家的说明, 利用 QIAquick DNA 纯化试剂盒从 0.8%琼脂糖-0.5XTBE 凝胶中分离质粒载体片段和 amyL/甘露聚糖酶 PCR 片段, 连接分离的片段, 并将他们转化到枯草芽孢杆菌菌株 PL2306 (美国专利 No.6,677,147) 中, 在每毫升补充有 10ug 卡那霉素的 TBAB 平皿上选择卡那霉素抗性菌株。依照厂家的说明, 利用 QIAprep Spin Miniprep 试剂盒 (QIAGEN Inc., Valencia, CA) 从几个转化体纯化质粒。通过限制性内切酶消化分析质粒。鉴定包含正确插入物的质粒, 并将其命名为 pMB1024-1 (图 21)。

枯草芽孢杆菌 MDT206 的构建如下。通过电穿孔将质粒 pMDT006 (实施例 3) 引入地衣芽孢杆菌 MaTa4。分离红霉素抗性转化体, 按照实施例 1 的描述整合质粒并从染色体中切除质粒, 导致用 P16 启动子替换 P11 启动子产生地衣芽孢杆菌菌株 MDT206。利用引物 94-935 (上述的) 和下列引物, 通过 PCR 分析鉴定想要的克隆。

引物 94-919 :

5-GGAAGTACAAAAATAAGC-3' (SEQ ID NO : 70)

按照如下步骤从地衣芽孢杆菌菌株 MDT206 PCR 扩增和克隆 cryIIIA mRNA 处理稳定序列和 amyL 信号序列。在由 50ng 地衣芽孢杆菌 MDT206 染色体 DNA (按照上文 Pitcher et al., 1989 的描述获取), 226370 和 219916 引物各 0.4 uM, dATP、dCTP、dGTP 和 dTTP 各 200 uM, 具有 2.5 mM MgCl₂ 的 1X PCR Buffer II 和 2.5 单位 AmpliTaq Gold™ DNA 聚合酶组成的 50ul 反应中进行 PCR 扩增。在 RoboCycler 40 热循环仪中进行反应, 反应程序为 95°C 9 分钟, 1 个循环; 95°C 1 分钟, 55°C 1 分钟, 30 循环; 72°C 1 分钟, 30 个循环; 和 72°C 5 分钟, 1 个循环。用 0.5XTBE 缓冲液, 在 0.8%琼脂糖凝胶上目测检验大约 650bp 的 PCR 产物。

引物 226370 :

5'-CATCCCCCGGGAGCTTAATTAAGATAATATCTTTGAATTG-3'

(SEQ ID NO : 71)

引物 219916 :

5'-TGCCGCGGCTGCAGAATGAGGCAG-3' (SEQ ID NO : 72)

用 XmaI 和 PstI 消化质粒 pMB1024-1 和 cryIIIA/amyL 信号序列的 PCR

片段。依照厂家的说明，利用 QIAquick DNA 纯化试剂盒从 0.8%琼脂糖-0.5XTBE 凝胶中分离来自 pMB1024-1 的大些的载体片段（大约 5500bp）和 amyL 信号序列的 PCR 片段（大约 650bp），连接分离的片段，并将他们转化到枯草芽孢杆菌菌株 MOL2023 中，在每毫升补充有 10ug 卡那霉素的 TBAB 平皿上选择卡那霉素抗性。该菌株基本上是具有来自质粒 pE194（Horinouchi and Weisblum, 1982, J. Bacteriol. 150 : 804 - 814）的红霉素抗性基因的枯草芽孢杆菌 A164Δ10，其中质粒 pE194 被插入 ydhT 基因（编码甘露糖内-1,4-β-甘露糖苷酶）中。枯草芽孢杆菌 A164Δ10 源自于枯草芽孢杆菌 A164Δ5（美国专利 No.5,891,701），而且在下列基因：spoIIAC、aprE、nprE、amyE、srfAC、wprA、bpr、vpr、mpr 和 epr 中有缺失。依照厂家的说明，利用 QIAprep Spin Miniprep 试剂盒从几个转化体纯化质粒，并通过限制性内切酶消化进行分析。鉴定出包含正确插入物的质粒，并将其命名为 pMB1242（图 22）。

利用显示如下的引物 993634 和引物 733-68-1（实施例 1），通过 PCR 从 pMB1242 扩增 RBS 和甘露聚糖酶基因。

引物 993634：

5'-GTTAACTTGAAAAACGGTGCTTAATC-3'（SEQ ID NO : 73）

在由 50ng pMB1242，引物 993634 和 733-68-1 各 0.4 uM，dATP、dCTP、dGTP 和 dTTP 各 200 uM，具有 2.5 mM MgCl₂ 的 50X PCR Buffer II 和 2.5 单位 AmpliTaq Gold™ DNA 聚合酶组成的 50ul 反应中进行一式三份的 PCR 扩增。在 RoboCycler 40 热循环仪中 A 进行反应，反应程序为 95℃9 分钟，1 个循环；95℃1 分钟，55℃1 分钟，30 循环；72℃1 分钟，30 个循环；和 72℃5 分钟，1 个循环。利用 0.5XTBE 缓冲液，在 0.8%琼脂糖凝胶上目测检验 PCR 产物。预期片段是大约 1052 bp。

利用 TA-TOPO 克隆试剂盒将 1052bp 的 PCR 片段克隆到 pCR2.1 中，并依照厂家的说明将其转化到大肠杆菌 OneShot™ 感受态细胞中。37℃在每毫升补充有 100ug 氨苄青霉素的 2XYT 琼脂平板上生长 16 小时之后进行转化体选择。依照厂家的说明，利用 QIAGEN Bio Robot 9600 从这些转化体中纯化质粒 DNA，用 SfiI 加 NotI 消化质粒 DNA。鉴定质粒，利用 M13（-20）正向和 M13 反向引物通过 DNA 测序确认插入物的 DNA 序列。鉴定具有正确序列的质粒，并将其命名为 pTH029（图 23）。

接下来,按照下列步骤将甘露聚糖酶基因作为 *SacI/HpaI* 片段克隆到 pNBT18 (pDG268MCS Δ neo-长 cryIII Δ stab/SAV, 美国专利 No.6,255,076) 中替换 aprH 编码区。用 *SacI* 加 *HpaI* 消化质粒 pTH029。用 *SacI* 加 *HpaI* 消化质粒 pNBT18。利用 0.5XTBE 缓冲液,在 0.8%琼脂糖凝胶上分辨消化物,依照厂家的说明利用 QIAquick DNA 抽提试剂盒从凝胶中纯化来自 pNBT18 的大些的载体片段(大约 7330bp)和来自 pTH029 的较小的甘露聚糖酶基因片段(大约 1010bp)。利用 T4 DNA 连接酶将两个纯化的片段连接在一起,依照厂家的说明将连接的混合物转化到大肠杆菌 OneShot™ 感受态细胞中。37°C 在每毫升补充有 100ug 氨苄青霉素的 2XYT 琼脂平板上生长 16 小时之后进行转化体选择。依照厂家的说明,利用 QIAGEN 自动机从这些转化体纯化质粒 DNA。通过存在大约 1095bp 的 *SfiI/NotI* 甘露聚糖酶基因片段鉴定出正确的质粒,并将其命名为 pTH026 (图 24)。

然后,按照如下步骤将甘露聚糖酶基因和来自 pTH026 的 aprH 基因终止子作为 *SacI/NotI* 片段插入 pTH012 中替换 JP170 蛋白酶基因。用 *SacI* 和 *NotI* 消化质粒 pTH012 和 pTH026。利用 0.5XTBE 缓冲液,在 0.8%琼脂糖凝胶上分辨消化物,依照厂家的说明利用 QIAquick DNA 抽提试剂盒从凝胶中纯化来自 pTH012 的大些的载体片段(大约 5359bp)和来自 pTH026 的较小的甘露聚糖酶基因片段(大约 1095bp)。利用 T4 DNA 连接酶将两个纯化的片段连接在一起,将连接的混合物转化到枯草芽孢杆菌 168 Δ 4 中。依照厂家的说明,利用 QIAGEN tip-20 柱子从几个转化体中纯化质粒 DNA,用 *SacI* 加 *NotI* 消化质粒 DNA 后,再利用 0.5XTBE 缓冲液,在 0.8%琼脂糖凝胶上进行分析。通过存在大约 1095 的 *SfiI/NotI* 甘露聚糖酶基因片段(该片段携带有 amyL RBS, amyL 信号序列-甘露聚糖酶编码序列融合物和 aprH 终止子,在下文中被称为"甘露聚糖酶基因")鉴定出正确的质粒。产生的质粒被命名为 pTH013 (图 25)。

构建质粒 pTH020,将包含类马链球菌类透明质酸合酶基因 (hasA) 的人工透明质酸 (HA) 操纵子引入在 P17 三联启动子控制下的地衣芽孢杆菌 MDT220 的染色体中。按照如下步骤将来自 pHA3 (WO 03/054163) 的 cryIII Δ stab/hasA/tuaD/gtaB 人工操纵子作为 *SacI/NotI* 片段插入质粒 pTH013 中替换甘露聚糖酶基因。用 *SacI* 和 *NotI* 消化质粒 pTH013 和 pHA3。利用 0.5XTBE 缓冲液,在 0.8%琼脂糖凝胶上分辨消化物,依照厂家的说明利用

QIAquick DNA 抽提试剂盒从凝胶中纯化来自 pTH013 的大些的载体片段 (大约 5400bp) 和来自 pHA3 的较小的 cryIIIAstab/hasA/tuaD/gtaB 人工操纵子片段 (大约 3800bp)。利用 T4 DNA 连接酶将两个纯化的片段连接在一起, 将连接的混合物转化到枯草芽孢杆菌 168 Δ 4 中。依照厂家的说明, 利用 QIAGEN tip-20 柱子从几个转化体中纯化质粒 DNA, 用 SacI 加 NotI 消化质粒 DNA 后, 再利用 0.5XTBE 缓冲液, 在 0.8% 琼脂糖凝胶上进行分析。通过限制性内切酶消化和/或 PCR 分析鉴定正确的质粒。产生的质粒被命名为 pTH020 (图 26)。

通过电穿孔将质粒 pTH020 引入地衣芽孢杆菌 MDT220。分离红霉素抗性转化体, 按照实施例 1 的描述整合质粒并从染色体中切除质粒。从几个转化体分离染色体 DNA (上文, Pitcher et al., 1989), 并通过 PCR 进行分析。利用引物 992463 和 992468 (hasA 基因)、引物 992464 和 992472 (tuaD 基因), 以及引物 992471 和 992477 (gtaB 基因) 完成 PCR 核对。在 HA 操纵子被整合到染色体之后, 鉴定最后的菌株是蛋白酶阴性和红霉素敏感菌株, 利用相同的引物再次确认存在的操纵子。选择一个这样的转化体, 并将其命名为地衣芽孢杆菌 TH15。

实施例 7: 地衣芽孢杆菌 TH15 的发酵

在两升的发酵罐 (Applikon, Inc., Holland) 中评估地衣芽孢杆菌 TH15 生产透明质酸 (HA) 的情况, pH 控制在 7.0 ± 0.2 , 培养基由每升 6.5g KH_2PO_4 、4.5g Na_2HPO_4 、3.0g $(\text{NH}_4)_2\text{SO}_4$ 、2.0g 柠檬酸钠、3.0g $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ 、15g 蔗糖、0.5g $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$ 、6.0ml 微量金属溶液 (组成见下文) 和 3.0ml 消泡剂组成。补料 (feed) 由 20% (w/w) 蔗糖组成。微量金属溶液由每升 100g 柠檬酸、20g $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ 、5g $\text{MnSO}_4 \cdot \text{H}_2\text{O}$ 、2g $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ 和 2g ZnCl_2 组成。在 48 小时发酵期间, 温度控制在 37°C 。最大气流和搅动速度分别是 1.5 vvm 和 1300 rpm。

显示在图 27 中的结果表明了 p17 三联启动子在地衣芽孢杆菌 TH15 中驱动 hasA/tuaD/gtaB 操纵子表达产生 HA 的能力。

本发明在这里描述和要求保护的内容并不限于这里公开的特定方案的范围, 因为这些方案只是用于解释说明本发明的几个方面。规定任何等效

方案都在本发明的范围内。实际上，在上述内容的基础上，这里显示和描述的方案之外的本发明的各种修饰对本领域熟练技术人员来说是很明显的。这种修饰也在附加的权利要求的范围内。在发生矛盾时，本发明中包含定义的公开内容将起调控作用。

这里引用的各种参考文献的公开内容全部引入作为参考。

| | | |
|-------|--------------------------------------------------------------------|-----|
| <110> | 诺维信生物聚合物公司 (Novozymes Biopolymer A/S) | |
| <120> | 在芽孢杆菌细胞中生产透明质酸的方法 | |
| <130> | 10596.204-W0 | |
| <150> | 60/558,507 | |
| <151> | 2004-03-31 | |
| <160> | 73 | |
| <170> | PatentIn version 3.2 | |
| <210> | 1 | |
| <211> | 614 | |
| <212> | DNA | |
| <213> | 地衣芽孢杆菌 (Bacillus licheniformis) | |
| <400> | 1 | |
| | gcatgctcc ttccttgtgc ttggaagcag agcccaatat tatcccgaaa cgataaaacg | 60 |
| | galgctgaag gaaggaaacg aagtcggcaa ccattcctgg gaccatccgt tallgacaag | 120 |
| | gctglcaaac gaaaaagcgt atcaggagat taacgacacg caagaaatga tcgaaaaaat | 180 |
| | cagcggacac ctgcctgtac acttgctgcc tccatacggc gggatcaatg attccgtccg | 240 |
| | ctcgccttcc aatctgaagg ttccattgig ggatgllgat ccggaagatt ggaagtacaa | 300 |
| | aaalaagcaa aagattgtca atcatgtcat gagccatgcg ggagacggaa aaatcgtcct | 360 |
| | aatgcacgat atttatgcaa cgttcgcaga tgctgctgaa gagattatta aaaagctgaa | 420 |
| | agcaaaaggc tatcaattgg taactgtatc tcagcttgaa gaaglgaa gaagcagagagg | 480 |
| | ctattgaata aatgagtaga agcgccatat cggcgctttt cttttggaag aaaatatagg | 540 |
| | gaaaatggta ctggttaaaa attcggaaata ttatacaat atcatatgta tcacattgaa | 600 |
| | aggggaggag aatc | 614 |
| <210> | 2 | |
| <211> | 1251 | |
| <212> | DNA | |
| <213> | 类马链球菌 (Streptococcus equisimilis) | |
| <400> | 2 | |
| | atgagaacat taaaaaacct cataactggt gtggccctta gtattttttg ggtactgttg | 60 |
| | atllacgtca atgtttatct ctttgggtgct aaaggaagct tgtcaattta tggctttttg | 120 |
| | ctgalagctt acctattagt caaaatgtcc ttatcctttt tttaacaagcc atttaaggga | 180 |
| | agggctgggc aatataaggt tgcagccatt attccctctt ataacgaaga tgctgagtca | 240 |
| | ttgctagaga ccttaaaaag tgltcagcag caaacctatc ccctagcaga aatttatggt | 300 |

gttgacgatg gaagtgctga tgagacaggt attaagcgca ttgaagacta tgtgcgtgac 360
 actggtgacc tatcaagcaa tgtcattggt caccggctcag aaaaaaatca aggaaagcgt 420
 catgcacagg cctgggcctt tgaagatca gacgctgatg tctttttgac cgttgactca 480
 galacttata tctacctga tgccttagag gagttgttaa aaacctttaa tgaccaact 540
 gtllltgctg cgacgggtca ccttaatgtc agaaatagac aaaccaatct cttaacacgc 600
 ttgacagata ttgcclatga taatgctttt ggcgttgaac gagctgcca atccgtlaca 660
 gglaataatc tegtltgctc aggcccgctt agcgtttaca gacgcgaggt ggttgttctt 720
 aacatagata gatacatcaa ccagaccttc cigggtattc ctgtaagtat cggatgatgac 780
 aggtgcltga ccaactatgc aactgattta ggaaagactg tttatcaatc cactgctaaa 840
 tgtattacag atgttcttga caagatgtct acttacttga agcagcaaaa ccgctggaac 900
 aagtccttct ttagagagtc cattatttct gtaagaaaa tcatgaacaa tctttttgta 960
 gccclatgga ccatacttga ggtgtctatg tttatgatgc ttgtttattc tgtggtggat 1020
 ttctttglag acaatgtcag agaatttgat tgctcaggg ttttggcctt tctggtgatt 1080
 atcttcatlg ttgctctttg tctaatatt cactatatgc ttaagcacc gctgtccttc 1140
 ttgttatctc cgtttlatgg ggtactgcat ttgtttgtcc tacagccctt gaaattglat 1200
 tctcttttla ctattagaaa tctgactgg ggaacacgta aaaaattatt a 1251

<210> 3

<211> 417

<212> PRT

<213> 类马链球菌 (Streptococcus equisimilis)

<400> 3

Met Arg Thr Leu Lys Asn Leu Ile Thr Val Val Ala Phe Ser Ile Phe
1 5 10 15

Trp Val Leu Leu Ile Tyr Val Asn Val Tyr Leu Phe Gly Ala Lys Gly
20 25 30

Ser Leu Ser Ile Tyr Gly Phe Leu Leu Ile Ala Tyr Leu Leu Val Lys
35 40 45

Met Ser Leu Ser Phe Phe Tyr Lys Pro Phe Lys Gly Arg Ala Gly Gln
50 55 60

Tyr Lys Val Ala Ala Ile Ile Pro Ser Tyr Asn Glu Asp Ala Glu Ser
65 70 75 80

Leu Leu Glu Thr Leu Lys Ser Val Gln Gln Gln Thr Tyr Pro Leu Ala
85 90 95

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

Leu

<210> 4
 <211> 1257
 <212> DNA
 <213> 化脓性链球菌 (*Streptococcus pyogenes*)

<400> 4
 gfgcctatftt ttaaaaaaac tttaatigtg ttatccftta tttttttgat atctatcttg 60
 atttatctaa atatgtatct atttggaaca tcaactgtag gaatttatgg agtaatatta 120
 ataacctatc tagttattaa actlggattia tctttccftt atgagccatt taaaggaaag 180
 ccacatgact ataaagtftg tgcftgtaatt cfttcttata atgaagatgc cgagtcatta 240
 ttagaaacfc ttaaaagtgt gttagcacag acctatccgt taltcagaaat ttatatigtg 300
 gatgatggga gttcaaacac agatgcaata caattaattg aagagtatgt aaatagagaa 360
 gftgataftt gfcgaaacgt tategttcaac cgttcccttg tcaataaagg aaaacgcat 420
 gctcaagcgt gggcafttga aagatctgac gctgacgttt ttttaaccgt agattcagat 480
 acttataftc atccaaalgc cttagaagaa ctcctaaaaa gcttcaatga tgagacagtt 540
 tatgctgcaa caggacattt gaatgctaga aacagacaaa ctaatctatt aacgcgactt 600
 acagataftc gttacgataa tgccfttggg gftggagcgtg ctgctcaatc attaacaggt 660
 aatafttftag ttgctcagg accattgagt atttatcgac gftgaagtgat tattcctaac 720
 tttagagcgt ataaaaatca aacattccia ggtttacctg tttagcatgg ggatgatcga 780
 tglftlaacaa attatgctat tgatfttagga cgcactgtct accaatcaac agctagatgt 840
 galactgatg taccfttcca attaaaaagt tattftaaagc aacaaaaatg atggaataaa 900
 tctftftftta aagaatctat tattftctgtt aaaaaaatc tttcftaatcc catcgttggc 960
 ttatggacia tfttcgaagt cgttatgttt atgatgttga ttgctcgaat tgggaatctt 1020
 ttgftlaatc aagctattca attagacctt attaaacttt ttgctftttt atccatcadc 1080
 tftatcgttg cfttatgtcg taatgttcat tatatgatca aacatccftg tagttftttg 1140
 ttatcftccfc tglatggaat attacactftg tftgtctftac agccctaaa actfttattct 1200
 ttatgcacca ttaaaaaaac ggaatgggga acacgtaaaa aggtcactat tftftaaa 1257

<210> 5
 <211> 419
 <212> PRT
 <213> 化脓性链球菌 (*Streptococcus pyogenes*)

<400> 5

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

Glu Ser Ile Ile Ser Val Lys Lys Ile Leu Ser Asn Pro Ile Val Ala
305 310 315 320

Leu Trp Thr Ile Phe Glu Val Val Met Phe Met Met Leu Ile Val Ala
325 330 335

Ile Gly Asn Leu Leu Phe Asn Gln Ala Ile Gln Leu Asp Leu Ile Lys
340 345 350

Leu Phe Ala Phe Leu Ser Ile Ile Phe Ile Val Ala Leu Cys Arg Asn
355 360 365

Val His Tyr Met Ile Lys His Pro Ala Ser Phe Leu Leu Ser Pro Leu
370 375 380

Tyr Gly Ile Leu His Leu Phe Val Leu Gln Pro Leu Lys Leu Tyr Ser
385 390 395 400

Leu Cys Thr Ile Lys Asn Thr Glu Trp Gly Thr Arg Lys Lys Val Thr
405 410 415

Ile Phe Lys

<210> 6

<211> 1251

<212> DNA

<213> 乳房链球菌 (*Streptococcus uberis*)

<400> 6

```

atggaaaaac taaaaaatct cattacattt atgacttita ttttcctgtg gctcataatt      60
attgggcitta atgtttttgt atttggaact aaaggaagtc taacagtgtg tgggattatt      120
ctattaacct atttgtgat aaaaatggga ttatcttttt tttatcgtcc ctataaagga      180
agtgtaggtc aatataaggt agcagctatt atcccatctt ataatgagga tgggtcgggt      240
ttaclagaaa ctctaaagag tgttcaaaaa caaacatata caattgcaga aattttcgta      300
atlgacgatg ggcagtaga taaaacaggt ataaaattgg tcgaagacta tgtgaagtta      360
aatggctttg gagaccaagt tatcgttcat cagatgcctg aaaatgttgg taaaagacat      420
gctcaggctt gggcatttga aaggtctgat gctgatgitt tcttaacagt ggattcagat      480
acctacatct atcctgatgc tcttgaagaa ttattaaaga catttaatga tccagaggtc      540
tacgctgcaa ctggtcattt aaatgcaaga aatagacaaa ctaatctctt aactagactg      600
actgalattc gttacgataa tgcatttggg glagaacgtg ctgctcagtc tgttacggga      660
aatattttgg ttgttccgg acctttaagt atttatagac gttecgctcg tattccaat      720
cttgaacgct ataccacaca aacatttctt ggtgtccctg taagcatagg ggatgaccgt      780
tgtttgacaa attatgcaac tgatttggga aaaacggttt atcagtcaac tgcaagatgt      840

```

gatactgacg ticcagataa gtttaagggt ttcacaaac aacaaaatcg ttggaataag 900
 tcatttttta gggagtctat tatctctggt aagaagtat tagccacacc aagtgttgct 960
 gttiggacta tiacagaagt tccatgttc atcatgctag tttattctat ctttagctta 1020
 ttgalaggag aggcicaaga atttaatcic ataaaactgg ttgctttttt agttattatt 1080
 ttcalagtag ctcctttagtag aaatgttcat tacatgggta agcatccatt tgctttttta 1140
 ttgicaccgt ttatggatt galacatcia ttcgttttgc aacctcttaa gatatattcg 1200
 ttatitacta taagaaatgc tacatgggga actcgtaaaa agacaagtaa a 1251

<210> 7

<211> 416

<212> PRT

<213> 乳房链球菌 (Streptococcus uberis)

<400> 7

Met Glu Lys Leu Lys Asn Leu Ile Thr Phe Met Thr Phe Ile Phe Leu
1 5 10 15

Trp Leu Ile Ile Ile Gly Leu Asn Val Phe Val Phe Gly Thr Lys Gly
20 25 30

Ser Leu Thr Val Tyr Gly Ile Ile Leu Leu Thr Tyr Leu Ser Ile Lys
35 40 45

Met Gly Leu Ser Phe Phe Tyr Arg Pro Tyr Lys Gly Ser Val Gly Gln
50 55 60

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

Leu Leu Glu Thr Leu Lys Ser Val Gln Lys Gln Thr Tyr Pro Ile Ala
85 90 95

Glu Ile Phe Val Ile Asp Asp Gly Ser Val Asp Lys Thr Gly Ile Lys
100 105 110

Leu Val Glu Asp Tyr Val Lys Leu Asn Gly Phe Gly Asp Gln Val Ile
115 120 125

Val His Gln Met Pro Glu Asn Val Gly Lys Arg His Ala Gln Ala Trp
130 135 140

Ala Phe Glu Arg Ser Asp Ala Asp Val Phe Leu Thr Val Asp Ser Asp
145 150 155 160

Thr Tyr Ile Tyr Pro Asp Ala Leu Glu Glu Leu Leu Lys Thr Phe Asn
165 170 175

Asp Pro Glu Val Tyr Ala Ala Thr Gly His Leu Asn Ala Arg Asn Arg
180 185 190

Gln Thr Asn Leu Leu Thr Arg Leu Thr Asp Ile Arg Tyr Asp Asn Ala
195 200 205

Phe Gly Val Glu Arg Ala Ala Gln Ser Val Thr Gly Asn Ile Leu Val
210 215 220

Cys Ser Gly Pro Leu Ser Ile Tyr Arg Arg Ser Val Gly Ile Pro Asn
225 230 235 240

Leu Glu Arg Tyr Thr Ser Gln Thr Phe Leu Gly Val Pro Val Ser Ile
245 250 255

Gly Asp Asp Arg Cys Leu Thr Asn Tyr Ala Thr Asp Leu Gly Lys Thr
260 265 270

Val Tyr Gln Ser Thr Ala Arg Cys Asp Thr Asp Val Pro Asp Lys Phe
275 280 285

Lys Val Phe Ile Lys Gln Gln Asn Arg Trp Asn Lys Ser Phe Phe Arg
290 295 300

Glu Ser Ile Ile Ser Val Lys Lys Leu Leu Ala Thr Pro Ser Val Ala
305 310 315 320

Val Trp Thr Ile Thr Glu Val Ser Met Phe Ile Met Leu Val Tyr Ser
325 330 335

Ile Phe Ser Leu Leu Ile Gly Glu Ala Gln Glu Phe Asn Leu Ile Lys
340 345 350

Leu Val Ala Phe Leu Val Ile Ile Phe Ile Val Ala Leu Cys Arg Asn
355 360 365

Val His Tyr Met Val Lys His Pro Phe Ala Phe Leu Leu Ser Pro Phe
370 375 380

Tyr Gly Leu Ile His Leu Phe Val Leu Gln Pro Leu Lys Ile Tyr Ser
385 390 395 400

Leu Phe Thr Ile Arg Asn Ala Thr Trp Gly Thr Arg Lys Lys Thr Ser
405 410 415

<210> 8

<211> 2916

<212> DNA

<213> 出血败血性巴斯德(氏)菌(Pasteurella multocida)

<400> 8

atgaatcat taccacaagc aataaaagca tataacagca atgactatca attagcactc 60

aaattatttg aaaagtcggc ggaatctat ggacggaaaa ttgttgaatt tcaaattacc 120

aaatgccaag aaaaactctc agcacatcct tcigttaatt cagcacatct ttctgtaaat 180

aaagaagaaa aagcaatgt ttgcgatagt ccgtagata ttgcaacaca actgttactt 240

tccaacglaa aaaaattagt acitcttgac tcggaaaaaa acacgttaaa aaataaatgg 300

| | |
|---------------------------------------------------------------------|------|
| aaatigctca ctgagaagaa atctgaaaat gcggaggtaa gagcggtcgc ccttgtagca | 360 |
| aaagattttc ccaaagatct ggtttttagcg cctttacctg atcatgttaa tgattttaca | 420 |
| tggtacaaaa agcgaagaa aagacttggc ataaaacctg aacatcaaca tgttggcttt | 480 |
| tctatlatcg ttacaacatt caatcgacca gcaatlliat cgattacatt agcctgttta | 540 |
| gtaaaccaaa aaacacatta cccgtttgaa gttatcgiga cagatgatgg tagtcaggaa | 600 |
| gatclatcac cgatcattcg ccaatatgaa aataaatigg atattcgcta cgtcagacaa | 660 |
| aaagataacg gtttcaagc cagtgccgct cggaatatgg gattacgctt agcaaaaat | 720 |
| gacitlattg gcttactcga ctgtgatatg gcgccaaatc cattatgggt tcattcttat | 780 |
| gttgagagc tattagaaga tgatgattia acaatcattg gtccaagaaa atacatcgat | 840 |
| acacaacata ttgacccaaa agacttctta aataacgcga gtttgcttga atcattacca | 900 |
| gaagtgaaaa ccaataatag tgttgccgca aaaggggaag gaacagtttc tctggattgg | 960 |
| cgcttagaac aatcgaaaa aacagaaaat ctccgcttat ccgattcgcc ttcccgtttt | 1020 |
| tttgcggcgg gtaatgttgc ttctgctaaa aaatggctaa ataaatccgg ttctttgat | 1080 |
| gaggaaattt atcactgggg tggagaagat gtggaatttg gatatcgctt attccgttac | 1140 |
| ggtagtttct taaaaactat tgatggcatt atggccctacc atcaagagcc accaggtaaa | 1200 |
| gaaaatgaaa ccgatcgtga agcgggaaaa aatattacgc tcalattat gagagaaaag | 1260 |
| gtcccttata tctatagaaa actttttacca atagaagatt cgcaatcaa tagagtacct | 1320 |
| ttagtttcaa ttatataccc agcttataac tgtgcaaac1 atattcaacg ttgcgtagat | 1380 |
| agtgacatga atcagactgt tgttgatctc gaggtttgta ttgttaacga tggttcaaca | 1440 |
| gataatacct tagaaglgat caataagctt tatggtaata atcctaggg1 acgcatcatg | 1500 |
| tctaaaccaa atggcggaat agcctcagca tcaaatgcag ccgtttcttt tgctaaaggt | 1560 |
| tattacattg ggagttaga ttcagatgat tatcttgagc ctgatgcagt tgaactgtgt | 1620 |
| ttaaaagaat ttttaaaaga taaaacgcta gcttgtgttt ataccactaa tagaaacgtc | 1680 |
| aatccggatg gtagcttaat cgctaattgg tacaattggc cagaattttc acgagaaaa | 1740 |
| ctcacaacgg ctatgattgc tcaccacttt agaatgttca cgattagagc ttggcattta | 1800 |
| actgatggat tcaatgaaaa aattgaaaaat gccgtagact atgacatgtt cctcaaacctc | 1860 |
| agtgaagtig gaaaatttaa acatcttaat aaaatctgct ataaccgtgt attacatggt | 1920 |
| gataacacat caattaagaa acttggcatt caaaagaaaa accatlttgt tglagtcaat | 1980 |
| cagtcaltta atagacaagg cataacttat tataattaig acgaatttga tgatttagat | 2040 |
| gaaagtagaa agtatatltt caataaaacc gctgaatatc aagaagagat tgatatctta | 2100 |

aaagatatta aaatcatcca gaataaagat gccaaaatcg cagtcagtat ttttatccc 2160
 aatacattaa acggccttagt gaaaaaacta aacaatatta ttgaatataa taaaaatata 2220
 ttcggtattg ttctacatgt tgataagaat catcttacac cagatatcaa aaaagaaata 2280
 ctagccctct atcataaaca tcaagtgaat attttactaa ataatgatat ctcatattac 2340
 acgaglaata gaitaataaa aactgaggcg catttaagta atattaataa attaagtcag 2400
 ttaaatctaa atigtgaata catcattttt gataatcatg acagcctatt cgttaaaaat 2460
 gacagctatg cttataigaa aaaatatgat gtcggcatga atttctcagc attaacacat 2520
 gattggatcg agaaaatcaa tgcgcatcca ccatttaaaa agctcattaa aacttatttt 2580
 aatgacaatg acttaaaaag tatgaatgtg aaaggggcat cacaaggat gttiatgacg 2640
 tatgcgctag cgcatgagct tctgacgatt attaaagaag tcatcacatc ttgccagtca 2700
 attgatagtg tgcagaata taacactgag gatatttggc tccaatttgc acttttaatc 2760
 ttgaaaaaga aaaccggcca tglattlaat aaaacatcga ccctgactta tatgccttgg 2820
 gaacgaaaat tacaatggac aaatgaacaa attgaaagtg caaaaagagg agaaaatata 2880
 cctglttaaca agltcattat taatagtata actcta 2916

<210> 9

<211> 972

<212> PRT

<213> 出血败血性巴斯德(氏)菌(*Pasteurella multocida*)

<400> 9

Met Asn Thr Leu Ser Gln Ala Ile Lys Ala Tyr Asn Ser Asn Asp Tyr
1 5 10 15

Gln Leu Ala Leu Lys Leu Phe Glu Lys Ser Ala Glu Ile Tyr Gly Arg
20 25 30

Lys Ile Val Glu Phe Gln Ile Thr Lys Cys Gln Glu Lys Leu Ser Ala
35 40 45

His Pro Ser Val Asn Ser Ala His Leu Ser Val Asn Lys Glu Glu Lys
50 55 60

Val Asn Val Cys Asp Ser Pro Leu Asp Ile Ala Thr Gln Leu Leu Leu
65 70 75 80

Ser Asn Val Lys Lys Leu Val Leu Ser Asp Ser Glu Lys Asn Thr Leu
85 90 95

Lys Asn Lys Trp Lys Leu Leu Thr Glu Lys Lys Ser Glu Asn Ala Glu
100 105 110

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

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

Asn Lys Asn Ile Phe Val Ile Val Leu His Val Asp Lys Asn His Leu
 740 745 750
 Thr Pro Asp Ile Lys Lys Glu Ile Leu Ala Phe Tyr His Lys His Gln
 755 760 765
 Val Asn Ile Leu Leu Asn Asn Asp Ile Ser Tyr Tyr Thr Ser Asn Arg
 770 775 780
 Leu Ile Lys Thr Glu Ala His Leu Ser Asn Ile Asn Lys Leu Ser Gln
 785 790 795 800
 Leu Asn Leu Asn Cys Glu Tyr Ile Ile Phe Asp Asn His Asp Ser Leu
 805 810 815
 Phe Val Lys Asn Asp Ser Tyr Ala Tyr Met Lys Lys Tyr Asp Val Gly
 820 825 830
 Met Asn Phe Ser Ala Leu Thr His Asp Trp Ile Glu Lys Ile Asn Ala
 835 840 845
 His Pro Pro Phe Lys Lys Leu Ile Lys Thr Tyr Phe Asn Asp Asn Asp
 850 855 860
 Leu Lys Ser Met Asn Val Lys Gly Ala Ser Gln Gly Met Phe Met Thr
 865 870 875 880
 Tyr Ala Leu Ala His Glu Leu Leu Thr Ile Ile Lys Glu Val Ile Thr
 885 890 895
 Ser Cys Gln Ser Ile Asp Ser Val Pro Glu Tyr Asn Thr Glu Asp Ile
 900 905 910
 Trp Phe Gln Phe Ala Leu Leu Ile Leu Glu Lys Lys Thr Gly His Val
 915 920 925
 Phe Asn Lys Thr Ser Thr Leu Thr Tyr Met Pro Trp Glu Arg Lys Leu
 930 935 940
 Gln Trp Thr Asn Glu Gln Ile Glu Ser Ala Lys Arg Gly Glu Asn Ile
 945 950 955 960
 Pro Val Asn Lys Phe Ile Ile Asn Ser Ile Thr Leu
 965 970

<210> 10

<211> 1200

<212> DNA

<213> 炭疽杆菌 (Bacillus anthracis)

<400> 10

taaggttcca aatatgaaag tggcagtagt ggtccccca tataatgaaa gtgcaagtgc 60

tattgtaat acaattiaata gcgttttagc tcaagattat ccaattcatg aaatttictt 120

tgttgatgat ggiagtaagg ataaatcggc ttatgaagta gcacttaaaa tgagggagga 180

acttcltaga acicaacgag aaattgctgc tacaactaag aatatttggt ctgaaatatt 240

aggtaticct gacttaatcg tacatcgttt acctaagaat tgcgggaaaa gacatgctca 300
 attatgggct tftaaacgga caacagcaga tgctatigt accattgatt cagatggatga 360
 ttigtlccca aatgctgtta gagagttatt gaaacccttt aatgatgaaa aagtaatggc 420
 cacaactggc cacglgaaca ttcgtaatag aaatgataat ttattaaaca aactaattga 480
 talgcgllat gacaatgcgt tccgtgtgga gcgtgcagca cagtcgtaa caggaaatgt 540
 tcttgtttgt agtgggcccgt taagtgttga tctagagaa gtaataactg aaaatttaga 600
 acattatgga agtcagatgt tccctggatga ggagggtcag ttggagatg atagatgctt 660
 aactaattat gctattttga aagggaaac agtttatcaa tccactgctc gatgtattac 720
 tgalgctcca actacattaa aacaatttct taaacagcaa ctacgttga acaagtcatt 780
 tttlagagaa agtttaattt cacttggcat tggatgaaa aaaccaaatg ttctgtttg 840
 gacaattttc gaaatatcgt tatggatttt atttggcctt tccctacttc taagtattat 900
 tctcaaggca agtcatgtag ggtaatttt ggctgtttat tatttgggtt atatttcatt 960
 agctgtatat gctagaaatg tattttatct attaaaacat ccccttactt tcttactggc 1020
 gccattatat ggaattctcc atgtatttagc actattacct atacgctttt atgctttact 1080
 aactattaaa tctaattggtt gggaacacg ttaattacag taattttatg tattttttta 1140
 ggaggatatt attaatgtaa gattagaaaa gcgattatcc cagcagcggg attaggcaca 1200

<210> 11
 <211> 366
 <212> PRT
 <213> 炭疽杆菌 (Bacillus anthracis)

<400> 11

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

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

<210> 12

<211> 1380

<212> DNA

<213> 硫磺矿硫化叶菌 (Sulfolobus solfataricus)

<400> 12

tgccctcatg cggttaagct tgaatggcc ttttttagaa agagaaagga gttatctagt

60

taaccaatct tgccaccaga aagttatcat agcatatagg gaagcaatat attgcactgc 120
 tagtgcgata gaaccgtata ttagagtttt caatttatct tctggaatta aataaatcat 180
 gglcattaca aagggtatta tagctataaa tcctccgtaa tgcaagaata aataaaagat 240
 aaagaagtcg gaatggtaac gggttggcag atacaalagt aataagagct taglattaat 300
 tacatlaatt gaggaataga tcittaagaat tctagtgaaa ctcalataaa ggaacaagag 360
 cgtaaaataat gglaacagat tggigtatat cataatlaaaa acgtataatg aaccctttt 420
 acttatacia ccgicagcia tciccccttat aaaattaaga taatttgctc ttgtccatct 480
 agtlacttgt ttcgtaaaca tttttaigtc tctagggggg tttgtatatg ccactgcatc 540
 aaagactttt acagccctat accctttttt tataacaaaa tgggttaaat ctctatcatc 600
 ggaaatllta attggctctc caaacatttt cggctctaaa aactctllag ataatatata 660
 tggllllacg agtccggicc tatatattac acattgicca ctlaalatta tagcacttcc 720
 aaaatagitt accgcccgtt ttactatctc acttattctc tcaaagaatt caccataata 780
 atalgcatat ttatttttct cgtcalacat aattctaata ttggcccta ctccacctac 840
 tgaiccatca aaaacactta acatctllag tatagagctt ttataaataa tclatcact 900
 atclagaaac atcactagag gagatcttac atacttaact ccctcggcta acgcgtatct 960
 tttcccctta tgltcacgca talaataaaa ttaccacca tatcttccg taatlgattt 1020
 gtaiggttct agaacactat cccctacaac aataaatctt aacctgtgt catataaagt 1080
 ccttalcact tttcaaaaa tatclatttc ctcttataa actggtaalca caactgtaag 1140
 atcagagaga ttataaaaac ttgagtggtg agtttttcta ttattactta ttactgcaaa 1200
 aaatgaaltc aaaaagaaat aaagaatagt talaattgtg aatgaaagag aataaatgaa 1260
 atalgagact ccglgaaata agtgaaacat aatcaccact ataatgctcg atatcgaaat 1320
 atataacgat ttttcciaat tcaccattcg aattctccgt tcaaaaaggg gttagttaac 1380

<210> 13

<211> 415

<212> PRT

<213> 硫磺矿硫化叶菌 (Sulfolobus solfataricus)

<400> 13

Met Val Ile Met Phe His Leu Phe His Gly Val Ser Tyr Phe Ile Tyr
 1 5 10 15

Ser Leu Ser Phe Thr Ile Ile Thr Ile Leu Tyr Phe Phe Leu Asn Ser
 20 25 30

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

Lys Leu Leu Leu Leu Tyr Leu Pro Thr Arg Tyr His Ser Asp Phe
340 345 350

Phe Ile Phe Tyr Leu Phe Leu His Tyr Gly Gly Phe Ile Ala Ile Ile
355 360 365

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

Thr Leu Ile Tyr Gly Ser Ile Ala Leu Ala Val Gln Tyr Ile Ala Ser
385 390 395 400

Leu Tyr Ala Met Ile Thr Phe Trp Trp Gln Asp Trp Leu Thr Arg
405 410 415

<210> 14

<211> 1680

<212> DNA

<213> *Ectocarpus siliculosus*

<400> 14

gccatttaca tcttggcgcg atcgttgcaa agtaaaaata atgttttctg tgacggaaac 60

aggggcatga catgcttctc ggccatgaaga aaaacctctt gttcattgcg tctctgttct 120

tcaatgccct tttaacgctt ttgttgcttg gattcgaltt cggctacatc atcglatcga 180

tattcgttgt cggcggacac ttcagagaig tgatcaacgt agcataccaa ttacttcaca 240

tgcacaggat acicaggcgt tglgcccaca tcccggaaaga cgatgccaaag atagttatit 300

gttgcttggg tccgggtgtac aacgagaagc cttcgatgtt aaagaagaat ctigatgctc 360

tgacgacaca gaagctatcg gaaaacacca agttgggtgg gatgcttctc ttigacggac 420

tgaacaacca caacgcagat ctcttcaalg ccgtcgtcga tgccattggc cttgacaccg 480

gatgcggaga agagcaatgg tticcgaatt ggaagagcaa gctgctgaag aagttgggtg 540

acaaaaatcg catatacaat gacacgtcgg tcatcctgtc gtacaaggag acaaatcgg 600

gaaaaaaaga ctctctcatc atcggggaaa acttcatcgt gctcggcatc ccgaggatcg 660

aatctctgga cgtacgacaa gtggatttca tctatcacac ggacggcgac accatttccg 720

acgaaaactg ttgaacgag atggatgaagt ctctcgttga tgatccagac ctcgacggcg 780

tctctggcct cctgagaaca taccitcaagg acgacgcgac ttgctcggaa agtgcgttct 840

tagcgatgca agactttcag tacttcttct ccattgttgt ccgtaggatg acggagagca 900

taatgaattc aactacctgc ctcccggggt gctccaacat gatcaggata agcggaaaaga 960

ctcacgctgc gattgaaaaa tacggaaacc ttccgggtcaa gaagagcggg ctgggtcaga 1020

cagtcacgcg gatgcaagga accgaccgac gatacaccac gctcttctct agacagggtt 1080

ccaagctaca gatgaattgg cgtgcgtttg ttacacagga gccaccgctg aacgcgacgg 1140

cgtttgtgaa tcaacgcaga cgatggcttt caaactcctt cttcaattcc atgatcacgc 1200
 tgtactccaa caacatcccg atgtacatca agctatcgaa ccttgtcgac atcgcccagag 1260
 tcttcaccac gatccttcgc gtgatatcgt acttgtgctt ttgggtttac gtcaagaatt 1320
 tttcccttgt caacategtt tttttctcta tatttatcgc ccttccttac ctctatgcct 1380
 tcgcciggat attctgtatc gtccagagt ggaagcagat gatagccggg ttttttttga 1440
 acaaaaatcct cagccctttt ttatctgtga tcgcggtcac aaagatgttc ttcacttcaa 1500
 ccgalltcgc ttggggcagt acgcggttga caccaccgga tgcagcgtct taaactcaac 1560
 caacaatcgt ctctattaat aaacctgatt cgtgattttg ggttcgattc tectgtttcc 1620
 ttttttccct ttttttccct tttttccttt ttttccctatt ttcctttttg attttgtttt 1680

<210> 15

<211> 493

<212> PRT

<213> Ectocarpus siliculosus 病毒

<400> 15

Met Leu Leu Gly Leu Lys Lys Asn Leu Leu Phe Ile Ala Tyr Leu Phe
1 5 10 15

Phe Asn Ala Leu Leu Thr Leu Leu Leu Leu Gly Phe Asp Tyr Gly Tyr
20 25 30

Ile Ile Val Ser Ile Phe Val Val Gly Gly His Phe Arg Asp Val Ile
35 40 45

Asn Val Ala Tyr Gln Leu Leu His Met His Arg Ile Leu Arg Arg Cys
50 55 60

Ala Asp Ile Pro Glu Asp Asp Ala Lys Ile Val Ile Cys Cys Leu Val
65 70 75 80

Pro Val Tyr Asn Glu Lys Pro Ser Met Leu Lys Lys Asn Leu Asp Ala
85 90 95

Leu Thr Thr Gln Lys Leu Ser Glu Asn Thr Lys Leu Val Val Met Leu
100 105 110

Leu Phe Asp Gly Leu Asn Asn His Asn Ala Asp Leu Phe Asn Ala Val
115 120 125

Val Asp Ala Ile Gly Leu Asp Thr Gly Cys Gly Glu Glu Gln Trp Phe
130 135 140

Pro Asn Trp Lys Ser Lys Leu Leu Lys Lys Leu Val Tyr Lys Ile Gly
145 150 155 160

Ile Tyr Asn Asp Thr Ser Val Ile Leu Ser Tyr Lys Glu Asn Asn Ser
165 170 175

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

Trp Gly Ser Thr Arg Leu Thr Pro Pro Asp Ala Ala Ser
 485 490

<210> 16
 <211> 1740
 <212> DNA
 <213> Paramecium bursaria Chlorella virus 1

<400> 16
 aagacttctt gaaagttaca atgggtaaaa atataatcat aatggtttccg tggtagacca 60
 tcataacitc aaatcclaac gcggttggag gagcctcict aatcttggct ccggcaatta 120
 ctgggatatgt tctacattgg aatattgctc tctcgacaat ctggggagta tcagcttatg 180
 gtattttcgt ttttgggttt ttccctggcac aagttttatt ttcagaactg aacaggaaac 240
 gtcttcgcaa gttgatttct ctacagaccia agggttggaa tgatgttcgt ttggctgtga 300
 tcatitgcigg atatcgcgag gatccttata tgttccagaa gtgccctcag tctgtacgtg 360
 actctgatta tggcaacgtt gcccgctctga ttgtgtgat tgacgggatg gaggacgatg 420
 atatgaggat ggctgccgtt tacaaggcga tctacaatga taatatcaag aagcccagat 480
 ttgttctgtg tgagtcagac gacaaggaag gtgaacgcat cgactctgat ttctctcgcg 540
 acatttgggt ccaccagcct catcgtggaa aacgggagtg tcttataact gggtttcaac 600
 ttgcaaagat ggaccccagt gtcaatgcig tegtctgat tgacagcgat accgttctcg 660
 agaaggatgc tattctggaa gttgtatacc cacttgcatt cgatcccag atccaagccg 720
 ttgcaggiga gttgaagatt tggaaacacag acactctttt gagtcttctc gtcgcttggc 780
 ggtactatct tgcgttttgt gtggagagga gtgccccagtc ttttttcagg actgttccagt 840
 gcgttggggg gccactgggt gcctacaaga ttgatatacat taaggagatt aaggaccctt 900
 ggatttccca gcgctttctt ggtcagaagt gtacttacgg tgacgaccgc cggctaacca 960
 acgagatctt gatgcgtggg aaaaagggtt gtttcactcc atttgcgtt ggttggctctg 1020
 acagtcggac caatgtgttt cggtagatcg ttacagcagac ccgctggagt aagtcgtgg 1080
 gccgcgaaat ttggtagacc ctcttcgccg cgtggaagca cggtttctct ggaatttggc 1140
 tggccttga atgtttgat caaattacat acttcttctt cgtgatttac ctctttctc 1200
 gcctagccgt tgaggccgac ctctgcgccc agacagccac ggtgattgtg agcaccacgg 1260
 ttgcatlgat taagtggtgg tatttttcat tccgagccaa ggatattcgg gcgttttact 1320
 ttgtgcttta tacatttgtt tactttttct gtatgatcc gccaggatt actgcaatga 1380
 tgacgccttg ggacattggc tggatactc gcggtggaaa cgagaagcct tccgttggca 1440
 cccgggtcgc tctgtgggca aagcaatata tcatitgcata tatgtgggtg gccgcggtt 1500

ttggcgctgg agttacagc atcgccata actggatggt cgattggaat tctctttctt 1560
 atcgttttgc ttiggttggg atttgttctt acattgtttt tattgttatt gtgctgggtg 1620
 ttatltcac cggcaaaatt acgacttggg atttcacgaa gcttcagaag gagctaactg 1680
 aggatcgcgt tcigtacgat gcaactacca atgctcagtc tgtgtgattt ttcttgaag 1740

<210> 17

<211> 567

<212> PRT

<213> 小球藻病毒(Paramecium bursaria Chlorella virus)

<400> 17

Met Gly Lys Asn Ile Ile Ile Met Val Ser Trp Tyr Thr Ile Ile Thr
 1 5 10 15

Ser Asn Leu Ile Ala Val Gly Gly Ala Ser Leu Ile Leu Ala Pro Ala
 20 25 30

Ile Thr Gly Tyr Val Leu His Trp Asn Ile Ala Leu Ser Thr Ile Trp
 35 40 45

Gly Val Ser Ala Tyr Gly Ile Phe Val Phe Gly Phe Phe Leu Ala Gln
 50 55 60

Val Leu Phe Ser Glu Leu Asn Arg Lys Arg Leu Arg Lys Trp Ile Ser
 65 70 75 80

Leu Arg Pro Lys Gly Trp Asn Asp Val Arg Leu Ala Val Ile Ile Ala
 85 90 95

Gly Tyr Arg Glu Asp Pro Tyr Met Phe Gln Lys Cys Leu Glu Ser Val
 100 105 110

Arg Asp Ser Asp Tyr Gly Asn Val Ala Arg Leu Ile Cys Val Ile Asp
 115 120 125

Gly Asp Glu Asp Asp Asp Met Arg Met Ala Ala Val Tyr Lys Ala Ile
 130 135 140

Tyr Asn Asp Asn Ile Lys Lys Pro Glu Phe Val Leu Cys Glu Ser Asp
 145 150 155 160

Asp Lys Glu Gly Glu Arg Ile Asp Ser Asp Phe Ser Arg Asp Ile Cys
 165 170 175

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

Gln Leu Ala Lys Met Asp Pro Ser Val Asn Ala Val Val Leu Ile Asp
 195 200 205

Ser Asp Thr Val Leu Glu Lys Asp Ala Ile Leu Glu Val Val Tyr Pro
 210 215 220

Leu Ala Cys Asp Pro Glu Ile Gln Ala Val Ala Gly Glu Cys Lys Ile
 225 230 235 240
 Trp Asn Thr Asp Thr Leu Leu Ser Leu Leu Val Ala Trp Arg Tyr Tyr
 245 250 255
 Ser Ala Phe Cys Val Glu Arg Ser Ala Gln Ser Phe Phe Arg Thr Val
 260 265 270
 Gln Cys Val Gly Gly Pro Leu Gly Ala Tyr Lys Asp Ile Ile Lys Glu
 275 280 285
 Ile Lys Asp Pro Trp Ile Ser Gln Arg Phe Leu Gly Gln Lys Cys Thr
 290 295 300
 Tyr Gly Asp Asp Arg Arg Leu Thr Asn Glu Ile Leu Met Arg Gly Lys
 305 310 315 320
 Lys Val Val Phe Thr Pro Phe Ala Val Gly Trp Ser Asp Ser Pro Thr
 325 330 335
 Asn Val Phe Arg Tyr Ile Val Gln Gln Thr Arg Trp Ser Lys Ser Trp
 340 345 350
 Cys Arg Glu Ile Trp Tyr Thr Leu Phe Ala Ala Trp Lys His Gly Leu
 355 360 365
 Ser Gly Ile Trp Leu Ala Phe Glu Cys Leu Tyr Gln Ile Thr Tyr Phe
 370 375 380
 Phe Leu Val Ile Tyr Leu Phe Ser Arg Leu Ala Val Glu Ala Asp Pro
 385 390 395 400
 Arg Ala Gln Thr Ala Thr Val Ile Val Ser Thr Thr Val Ala Leu Ile
 405 410 415
 Lys Cys Gly Tyr Phe Ser Phe Arg Ala Lys Asp Ile Arg Ala Phe Tyr
 420 425 430
 Phe Val Leu Tyr Thr Phe Val Tyr Phe Phe Cys Met Ile Pro Ala Arg
 435 440 445
 Ile Thr Ala Met Met Thr Leu Trp Asp Ile Gly Trp Asp Thr Arg Gly
 450 455 460
 Gly Asn Glu Lys Pro Ser Val Gly Thr Arg Val Ala Leu Trp Ala Lys
 465 470 475 480
 Gln Tyr Leu Ile Ala Tyr Met Trp Trp Ala Ala Val Val Gly Ala Gly
 485 490 495
 Val Tyr Ser Ile Val His Asn Trp Met Phe Asp Trp Asn Ser Leu Ser
 500 505 510
 Tyr Arg Phe Ala Leu Val Gly Ile Cys Ser Tyr Ile Val Phe Ile Val
 515 520 525
 Ile Val Leu Val Val Tyr Phe Thr Gly Lys Ile Thr Thr Trp Asn Phe
 530 535 540

<212> PRT

<213> 链球菌属

<400> 19

Val Lys Ile Ser Val Ala Gly Ser Gly Tyr Val Gly Leu Ser Leu Ser
1 5 10 15

Ile Leu Leu Ala Gln His Asn Asp Val Thr Val Val Asp Ile Ile Asp
20 25 30

Glu Lys Val Arg Leu Ile Asn Gln Gly Ile Ser Pro Ile Lys Asp Ala
35 40 45

Asp Ile Glu Glu Tyr Leu Lys Asn Ala Pro Leu Asn Leu Thr Ala Thr
50 55 60

Leu Asp Gly Ala Ser Ala Tyr Ser Asn Ala Asp Leu Ile Ile Ile Ala
65 70 75 80

Thr Pro Thr Asn Tyr Asp Ser Glu Arg Asn Tyr Phe Asp Thr Arg His
85 90 95

Val Glu Glu Val Ile Glu Gln Val Leu Asp Leu Asn Ala Ser Ala Thr
100 105 110

Ile Ile Ile Lys Ser Thr Ile Pro Leu Gly Phe Ile Lys His Val Arg
115 120 125

Glu Lys Tyr Gln Thr Asp Arg Ile Ile Phe Ser Pro Glu Phe Leu Arg
130 135 140

Glu Ser Lys Ala Leu Tyr Asp Asn Leu Tyr Pro Ser Arg Ile Ile Val
145 150 155 160

Ser Tyr Glu Lys Asp Asp Ser Pro Arg Val Ile Gln Ala Ala Lys Ala
165 170 175

Phe Ala Gly Leu Leu Lys Glu Gly Ala Lys Ser Lys Asp Thr Pro Val
180 185 190

Leu Phe Met Gly Ser Gln Glu Ala Glu Ala Val Lys Leu Phe Ala Asn
195 200 205

Thr Phe Leu Ala Met Arg Val Ser Tyr Phe Asn Glu Leu Asp Thr Tyr
210 215 220

Ser Glu Ser Lys Gly Leu Asp Ala Gln Arg Val Ile Glu Gly Val Cys
225 230 235 240

His Asp Gln Arg Ile Gly Asn His Tyr Asn Asn Pro Ser Phe Gly Tyr
245 250 255

Gly Gly Tyr Cys Leu Pro Lys Asp Ser Lys Gln Leu Leu Ala Asn Tyr
260 265 270

Arg Gly Ile Pro Gln Ser Leu Met Ser Ala Ile Val Glu Ser Asn Lys
275 280 285

Ile Arg Lys Ser Tyr Leu Ala Glu Gln Ile Leu Asp Arg Ala Ser Ser
 290 295 300

Gln Lys Gln Ala Gly Val Pro Leu Thr Ile Gly Phe Tyr Arg Leu Ile
 305 310 315 320

Met Lys Ser Asn Ser Asp Asn Phe Arg Glu Ser Ala Ile Lys Asp Ile
 325 330 335

Ile Asp Ile Ile Asn Asp Tyr Gly Val Asn Ile Val Ile Tyr Glu Pro
 340 345 350

Met Leu Gly Glu Asp Ile Gly Tyr Arg Val Val Lys Asp Leu Glu Gln
 355 360 365

Phe Lys Asn Glu Ser Thr Ile Ile Val Ser Asn Arg Phe Glu Asp Asp
 370 375 380

Leu Gly Asp Val Ile Asp Lys Val Tyr Thr Arg Asp Val Phe Gly Arg
 385 390 395 400

Asp

<210> 20

<211> 1206

<212> DNA

<213> 链球菌属

<400> 20

atgaaaatag cagttgctgg atcaggatat gttggattat cactaggagt tctttatca 60
 ctcaaaaacg aagtcactat tgttgatatt ctccctcta aagttgataa gattaataat 120
 ggcttateac caattcaaga tgaatatatt gaatattact taaaagtaa gcaattatct 180
 attaaagcaa cttlagatag caaagcagct tataagaag cggaactggc cattattgcc 240
 acacctacaa attacaacag tagaattaat tatttgata cacagcatgt tgaacagtt 300
 atcaaagagg tactaagcgt taatagccat gcaactccta tcatcaaatc acaattcca 360
 ataggtttca ttactgaaat gagacagaaa ticcaaactg atcgtattat cttcagccct 420
 gaatttttaa gagaatctaa agctttatat gacaacttat atccaagccg aattattggt 480
 tcttggaag aaaacgattc tccaaaagta aaggcagacg cagaaaaatt tgcactttta 540
 ttaaagctcg cagctaaaaa aaataatgta ccagtactta ttatgggagc ttcagaagct 600
 gaagcagtaa aactatttgc caatacttat ttagcgtaa gggtagctta ttttaatgag 660
 ttagacactt acgcagaatc gagaaaatta aatagtcaca tgattattca aggaatttct 720
 tatgatgac gaalaggaat gcattataat aacctatcat ttggttatgg aggttattgt 780
 ctacclaaag atacgaagca attattggca aattacaata atattcctca aacgctaatt 840

gaagctatcg tticatcaaa taatgtgcgc aagtcctata ttgctaagca aattatcaac 900
gtcctagaag agcgggagtc cccagtaaaa gtagtcgggg tttaccgitt aattatgaaa 960
agtaactcag ataattttag agaaagtgc atcaaagatg ttattgacat tcttaaaagt 1020
aaagacatta agataattat ttatgagcca atgttaaaca aacttgaatc tgaagatcaa 1080
tcctgiacttg taaatgattt agagaatttc aagaaacaag caaatattat cgtaactaat 1140
cgctatgata atgaattaca agatgttaaa aataaagitt acagtagaga tatttttaat 1200
agagac 1206

<210> 21
<211> 402
<212> PRT
<213> 链球菌属

<400> 21

Met Lys Ile Ala Val Ala Gly Ser Gly Tyr Val Gly Leu Ser Leu Gly
1 5 10 15

Val Leu Leu Ser Leu Gln Asn Glu Val Thr Ile Val Asp Ile Leu Pro
20 25 30

Ser Lys Val Asp Lys Ile Asn Asn Gly Leu Ser Pro Ile Gln Asp Glu
35 40 45

Tyr Ile Glu Tyr Tyr Leu Lys Ser Lys Gln Leu Ser Ile Lys Ala Thr
50 55 60

Leu Asp Ser Lys Ala Ala Tyr Lys Glu Ala Glu Leu Val Ile Ile Ala
65 70 75 80

Thr Pro Thr Asn Tyr Asn Ser Arg Ile Asn Tyr Phe Asp Thr Gln His
85 90 95

Val Glu Thr Val Ile Lys Glu Val Leu Ser Val Asn Ser His Ala Thr
100 105 110

Leu Ile Ile Lys Ser Thr Ile Pro Ile Gly Phe Ile Thr Glu Met Arg
115 120 125

Gln Lys Phe Gln Thr Asp Arg Ile Ile Phe Ser Pro Glu Phe Leu Arg
130 135 140

Glu Ser Lys Ala Leu Tyr Asp Asn Leu Tyr Pro Ser Arg Ile Ile Val
145 150 155 160

Ser Cys Glu Glu Asn Asp Ser Pro Lys Val Lys Ala Asp Ala Glu Lys
165 170 175

Phe Ala Leu Leu Leu Lys Ser Ala Ala Lys Lys Asn Asn Val Pro Val
180 185 190

Leu Ile Met Gly Ala Ser Glu Ala Glu Ala Val Lys Leu Phe Ala Asn
 195 200 205
 Thr Tyr Leu Ala Leu Arg Val Ala Tyr Phe Asn Glu Leu Asp Thr Tyr
 210 215 220
 Ala Glu Ser Arg Lys Leu Asn Ser His Met Ile Ile Gln Gly Ile Ser
 225 230 235 240
 Tyr Asp Asp Arg Ile Gly Met His Tyr Asn Asn Pro Ser Phe Gly Tyr
 245 250 255
 Gly Gly Tyr Cys Leu Pro Lys Asp Thr Lys Gln Leu Leu Ala Asn Tyr
 260 265 270
 Asn Asn Ile Pro Gln Thr Leu Ile Glu Ala Ile Val Ser Ser Asn Asn
 275 280 285
 Val Arg Lys Ser Tyr Ile Ala Lys Gln Ile Ile Asn Val Leu Glu Glu
 290 295 300
 Arg Glu Ser Pro Val Lys Val Val Gly Val Tyr Arg Leu Ile Met Lys
 305 310 315 320
 Ser Asn Ser Asp Asn Phe Arg Glu Ser Ala Ile Lys Asp Val Ile Asp
 325 330 335
 Ile Leu Lys Ser Lys Asp Ile Lys Ile Ile Ile Tyr Glu Pro Met Leu
 340 345 350
 Asn Lys Leu Glu Ser Glu Asp Gln Ser Val Leu Val Asn Asp Leu Glu
 355 360 365
 Asn Phe Lys Lys Gln Ala Asn Ile Ile Val Thr Asn Arg Tyr Asp Asn
 370 375 380
 Glu Leu Gln Asp Val Lys Asn Lys Val Tyr Ser Arg Asp Ile Phe Asn
 385 390 395 400
 Arg Asp

<210> 22

<211> 1203

<212> DNA

<213> 链球菌属

<400> 22

gtgaaaattg cagttgcagg ttctggctat gttggcctat cattaagtgt attattagca 60

cagaaaaatc ctgltacagt thtagatatt attgagaaga aagtaaatct cataaatcaa 120

aaacaatcac caatccagga tgttgatatt gaaaactatt taaaagaaaa aaagttaaa 180

ttaagagcta ctctagacgc cgalcaagca tttagggatg cagatatact aattattgct 240

acaccaacca attatgatgt ggagaagaat ttttttgata ctagtcatgt tgagactgta 300

atlgagaaag cttlagcttt aaatagtcag gctttggttag ttattaaatc aacgatacca 360
 cttggtttta ttaaaaagat gcgtcaaaaa tadcagacag accgtattat ttttagtccc 420
 gaattlcitta gagagtciaa agctttaaaa gataatcttt atcctagtcg aataattggt 480
 tcctilgaag atgaitgattc taiggaagta atagaagcag caaagacttt tgctcaattg 540
 ttaaaagaig giiclligga taaagaigt cctgtacttt ttatgggttc agcagaggct 600
 gaagcagtaa aattalltgc caatacctat ttagctatgc gigtctccta ttttaatgag 660
 ttagalacat atgctgaaaa gaalggttta cgtgiggata atattatga gggcgtttgc 720
 catgalcgcac gcataggaat tcattataat aaccctctt ttggctatgg aggatactgc 780
 ttacclaaag ataccaaaaca gtigctagca ggctatgatg gtattcctca atcgcttata 840
 aaagcaattg ttgallctaa taaaattcgt aaagagtata tcgcatcaca aattttacaa 900
 caattgagtg atattaatgt agalcctaaa gatgcaacga ttggtattta ccgccttacc 960
 atgaaaagta acictgataa tttcagagag agtgcaataa aagatattat tgalcalatt 1020
 aagagctatc aaattaatat agtcttgtat gagccaatga tgaatgaaga ttttgattta 1080
 ccaatcattg atgatttacc tgacttcaaa gccatgtcac atattatcgt ttcaaataga 1140
 tatgatttag ccttagaaga tgttaaagaa aaagtttaca ccagagatat ttacgggtg 1200
 gat 1203

<210> 23
 <211> 401
 <212> PRT
 <213> 链球菌属

<400> 23

Val Lys Ile Ala Val Ala Gly Ser Gly Tyr Val Gly Leu Ser Leu Ser
1 5 10 15

Val Leu Leu Ala Gln Lys Asn Pro Val Thr Val Val Asp Ile Ile Glu
20 25 30

Lys Lys Val Asn Leu Ile Asn Gln Lys Gln Ser Pro Ile Gln Asp Val
35 40 45

Asp Ile Glu Asn Tyr Leu Lys Glu Lys Lys Leu Gln Leu Arg Ala Thr
50 55 60

Leu Asp Ala Asp Gln Ala Phe Arg Asp Ala Asp Ile Leu Ile Ile Ala
65 70 75 80

Thr Pro Thr Asn Tyr Asp Val Glu Lys Asn Phe Phe Asp Thr Ser His
85 90 95

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

<210> 24
 <211> 1383
 <212> DNA
 <213> 枯草芽孢杆菌(bacillus subtilis)

<400> 24
 gtgaaaaaaa tagctgtcat tggaacaggt tatgtaggac tcgtatcagg cacttgcttt 60
 gcggagatcg gcaataaagt tgtttgctgt gatatcgatg aatcaaaaat cagaagcctg 120
 aaaaaatgggg taatcccaat ctatgaacca gggcttgtag acttagtga aaaaaatgtg 180
 ctggatcagc gcctgacctt tacgaacgat atcccgtctg ccattcgggc ctcagatatt 240
 atllatattg cagtcggaac gcciatgtcc aaaacaggtg aagctgattt aacgtacgtc 300
 aaagcggcgg cgaaaacaat cgggtgagcat cttaacggct acaaagtgat cgtaaataaa 360
 agcacagtcg cggllggaac agggaaacig glgcaatcta tcgttcaaaa agcccaaaag 420
 gggagatact calllgatgl tglatctaac cctgaattcc ttcgggaagg gtcagcgatt 480
 calgacacga lgaatatgga gcgtgccgtg attggttcaa caagtcataa agccgctgcc 540
 atcallgagg aactlcatca gccattccat gctcctgtca ttaaaacaaa cctagaaagt 600
 gcagaaatga tlaaatagc cgcgaaatgca tttctggcga caaagatttc cttatcaac 660
 gatatcgcaa acatllgtga gcgagtcggc gcagacgitt caaaagtigc tgatgggttt 720
 ggtcltgaca gccgtatcgg cagaaagttc cttaaagctg gtatlggatt cggcggttca 780
 tgtlltccaa aggatacaac cgcgctgctt caaatcgcaa aatcggcagg ctatccattc 840
 aagclcatcg aagcigtcat tgaacgaac gaaaagcagc gtgttcatat ttagataaaa 900
 ctlllgactg tlatgggaag cgtcaaaggg agaaccattt cagtcctggg attagccttc 960
 aaaccgaata cgaacgatgt gagatccgct ccagcgttg atattatccc aatgctgcag 1020
 cagctgggcg cccatgtaaa agcatacgat ccgattgcta ttcctgaagc ttcagcgatc 1080
 ctllgggaac aggtcgagta ttaacacagat gigtatgctg cgatggaaga cactgatgca 1140
 tgcctgattt taacggattg gccggaagtg aaagaaatgg agcttgtaaa agtgaaaacc 1200
 ctcttaaaac agccagtcat cattgacggc agaaatttat tttcacttga agagatgcag 1260
 gcagccgat acatttatca ctctatcggc cgtcccgtg ttcggggaac ggaaccctct 1320
 gacaaglatt ttccgggctt gccgcttgaa gaattggcta aagacttggg aagcgtcaat 1380
 tta 1383

<210> 25
 <211> 461

<212> PRT

<213> 枯草芽孢杆菌(bacillus subtilis)

<400> 25

Val Lys Lys Ile Ala Val Ile Gly Thr Gly Tyr Val Gly Leu Val Ser
1 5 10 15

Gly Thr Cys Phe Ala Glu Ile Gly Asn Lys Val Val Cys Cys Asp Ile
20 25 30

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

Glu Pro Gly Leu Ala Asp Leu Val Glu Lys Asn Val Leu Asp Gln Arg
50 55 60

Leu Thr Phe Thr Asn Asp Ile Pro Ser Ala Ile Arg Ala Ser Asp Ile
65 70 75 80

Ile Tyr Ile Ala Val Gly Thr Pro Met Ser Lys Thr Gly Glu Ala Asp
85 90 95

Leu Thr Tyr Val Lys Ala Ala Ala Lys Thr Ile Gly Glu His Leu Asn
100 105 110

Gly Tyr Lys Val Ile Val Asn Lys Ser Thr Val Pro Val Gly Thr Gly
115 120 125

Lys Leu Val Gln Ser Ile Val Gln Lys Ala Ser Lys Gly Arg Tyr Ser
130 135 140

Phe Asp Val Val Ser Asn Pro Glu Phe Leu Arg Glu Gly Ser Ala Ile
145 150 155 160

His Asp Thr Met Asn Met Glu Arg Ala Val Ile Gly Ser Thr Ser His
165 170 175

Lys Ala Ala Ala Ile Ile Glu Glu Leu His Gln Pro Phe His Ala Pro
180 185 190

Val Ile Lys Thr Asn Leu Glu Ser Ala Glu Met Ile Lys Tyr Ala Ala
195 200 205

Asn Ala Phe Leu Ala Thr Lys Ile Ser Phe Ile Asn Asp Ile Ala Asn
210 215 220

Ile Cys Glu Arg Val Gly Ala Asp Val Ser Lys Val Ala Asp Gly Val
225 230 235 240

Gly Leu Asp Ser Arg Ile Gly Arg Lys Phe Leu Lys Ala Gly Ile Gly
245 250 255

Phe Gly Gly Ser Cys Phe Pro Lys Asp Thr Thr Ala Leu Leu Gln Ile
260 265 270

Ala Lys Ser Ala Gly Tyr Pro Phe Lys Leu Ile Glu Ala Val Ile Glu
275 280 285

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

<210> 26
 <211> 900
 <212> DNA
 <213> 链球菌属

<400> 26
 atgacaaagg tcagaaaagc cattatccca gccgccggcc taggcaactcg ctctctaccc 60
 gccaccaagg cactggccaa ggaaatgctc ccaatcgicg ataagccaac cattcaattc 120
 atcgtcgagg aagctctaaa ggccggtatc gaggagattc ttgtcgtcac cgcaaggcc 180
 aaacgtctta ttgaagacca ctttgactcc aacttegagc tcgaatacaa tctccaagcc 240
 aagggcaaaa ccgagctgct caagctcggt gatgagacca ctgccaatcaa cctgcacttc 300
 attcgtcaga gccaccctag aggactaggg gacgctgtcc tccaggccaa ggccittgtg 360
 ggcaatgagc cctttgtggt catgctgggg gatgaccica tggatattac caatcctagt 420
 gccaaagcct tggccaagca gctcattgag gattatgatt gcacacacgc ctcaacgatt 480
 gcagtgatga gggcgccgca tgaggagggt tccaattatg gcgtgattgc accgcaaggg 540
 aaggctgtta agggcttgta tagtgiggag accttltgtg agaagccaag tccagatgag 600

gcaccgagtg acttagcgat tattggtcga tatttgttga cgcctgagat ttttgccata 660
 ttggagaatc aggcgcctgg ggctggcaat gaggtacagc tagccgatgc gattgacaag 720
 ctcaacaaga ctcagcgggt ttttgcgagg gagttaaagg gagagcggta tgaigtggg 780
 gacaagtttg gcattatgaa gaccicacti gactatgctc tcaagcacc tcaggccaag 840
 gacgacctca ctgactacat tataaagctc agtaagcaac tgaacaagga cgtaaaaaa 900

<210> 27
 <211> 300
 <212> PRT
 <213> 链球菌属

<400> 27

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

Gly Arg Tyr Leu Leu Thr Pro Glu Ile Phe Ala Ile Leu Glu Asn Gln
 210 215 220

Ala Pro Gly Ala Gly Asn Glu Val Gln Leu Ala Asp Ala Ile Asp Lys
 225 230 235 240

Leu Asn Lys Thr Gln Arg Val Phe Ala Arg Glu Phe Lys Gly Glu Arg
 245 250 255

Tyr Asp Val Gly Asp Lys Phe Gly Phe Met Lys Thr Ser Leu Asp Tyr
 260 265 270

Ala Leu Lys His Pro Gln Val Lys Asp Asp Leu Thr Asp Tyr Ile Ile
 275 280 285

Lys Leu Ser Lys Gln Leu Asn Lys Asp Val Lys Lys
 290 295 300

<210> 28

<211> 912

<212> DNA

<213> 链球菌属

<400> 28

atgaccaaag tcagaaaagc cattattcct gctgcaggtc taggaacacg ttttttacct 60

gctaccaaaag cctttgccaa agagatgttg cccatcgttg ataaaccaac catccagttt 120

atcgtcgaag aagcgcataa atctggcatc gaggaatcc ttgtggtgac cggaaaagct 180

aaacgctcta tcgaggacca ttttgattca aactttgaat tagaatacaa cctccaagct 240

aaggggaaaa algaacigtg gaaattagtg gatgaaacca ctgccattaa ccttcatttt 300

atccgtcaaa gccaccaag agggctggga gatgctgtct tacaagccaa agcctttgtg 360

ggcaatgaac cctttgtggt catgcttggg gatgacttaa tggacattac aaatgcatcc 420

gctaaaccic tcacaaaca actcatggag gactatgaca agacgcatgc atccactatc 480

gctlgatga aagttcctca tgaagatgig tctagctatg gggttatcgc tcccaaggc 540

aaggctgtca agggccttta cagtgtagac acctttgttg aaaaaccaca accagaagat 600

gcgcctagtg atttggctat tattggctgt tacctcctaa cccctgaaat ttttggattt 660

ttggaaagac agaccctgg agcaggtaac gaagtgcaac tcacagatgc tatcgatacc 720

ctcaataaaa ctacgcgtgt ctttgcacga gaatttaaag gcaatcgta cgatgttggg 780

gataaatitg galcatgaa aacatctatc gactatgcct tagaacacce acaggtcaaa 840

gaggacttga aaaattacat tatcaacta ggaaaagctt tggaaaaaag taaagtacca 900

acacattcaa ag 912

<210> 29

<211> 304
 <212> PRT
 <213> 链球菌属

<400> 29

Met Thr Lys Val Arg Lys Ala Ile Ile Pro Ala Ala Gly Leu Gly Thr
 1 5 10 15

Arg Phe Leu Pro Ala Thr Lys Ala Leu Ala Lys Glu Met Leu Pro Ile
 20 25 30

Val Asp Lys Pro Thr Ile Gln Phe Ile Val Glu Glu Ala Leu Lys Ser
 35 40 45

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

Glu Asp His Phe Asp Ser Asn Phe Glu Leu Glu Tyr Asn Leu Gln Ala
 65 70 75 80

Lys Gly Lys Asn Glu Leu Leu Lys Leu Val Asp Glu Thr Thr Ala Ile
 85 90 95

Asn Leu His Phe Ile Arg Gln Ser His Pro Arg Gly Leu Gly Asp Ala
 100 105 110

Val Leu Gln Ala Lys Ala Phe Val Gly Asn Glu Pro Phe Val Val Met
 115 120 125

Leu Gly Asp Asp Leu Met Asp Ile Thr Asn Ala Ser Ala Lys Pro Leu
 130 135 140

Thr Lys Gln Leu Met Glu Asp Tyr Asp Lys Thr His Ala Ser Thr Ile
 145 150 155 160

Ala Val Met Lys Val Pro His Glu Asp Val Ser Ser Tyr Gly Val Ile
 165 170 175

Ala Pro Gln Gly Lys Ala Val Lys Gly Leu Tyr Ser Val Asp Thr Phe
 180 185 190

Val Glu Lys Pro Gln Pro Glu Asp Ala Pro Ser Asp Leu Ala Ile Ile
 195 200 205

Gly Arg Tyr Leu Leu Thr Pro Glu Ile Phe Gly Ile Leu Glu Arg Gln
 210 215 220

Thr Pro Gly Ala Gly Asn Glu Val Gln Leu Thr Asp Ala Ile Asp Thr
 225 230 235 240

Leu Asn Lys Thr Gln Arg Val Phe Ala Arg Glu Phe Lys Gly Asn Arg
 245 250 255

Tyr Asp Val Gly Asp Lys Phe Gly Phe Met Lys Thr Ser Ile Asp Tyr
 260 265 270

Ala Leu Glu His Pro Gln Val Lys Glu Asp Leu Lys Asn Tyr Ile Ile
 275 280 285

Lys Leu Gly Lys Ala Leu Glu Lys Ser Lys Val Pro Thr His Ser Lys
 290 295 300

<210> 30
 <211> 912
 <212> DNA
 <213> 链球菌属

<400> 30
 atgactaaag taagaaaagc cattattcca gctgccggac ttggcacacg tttttacca 60
 gcaacaaaag ctctcgctaa ggaaatgttg cccatcgttg acaaaccaac cattcaattc 120
 atcglggaag aagctttgcg ttctggcatt gaagaaatct tggtcglaac aggaaaatca 180
 aaacgctcca ttgaagacca ttttgattcc aactttgaac tcgaatataa tttgcaagaa 240
 aaagggaaaa ctgaactctt aaaattagtt gatgaaacca cttctataaa cttgcatttc 300
 attcglcaaa gicatcccaa aggcttaggg gatgctgttt tacaagcaaa agcttttgta 360
 ggaaatgaac ccttcattgt tatgcttggg gacgatttga tggacattac aaatacaaaa 420
 gctglcccat taaccaaaca attaatggac gattatgaaa caacacatgc tctacaata 480
 gccglaatga aagttectca cgatgacgla tctcttatg gtgtcattgc tccaaacggc 540
 aaagccttga atggcttata tagcgtggat acctttgttg aaaaacaaa accigaggac 600
 gcaccaagtg accttgctat cattggacgt tateccttaa caccigaaat tttgacatt 660
 ctlgaaaatc aagcaccagg tgccggaaac gaagtccaat taactgatgc tatcgatacc 720
 ctcaacaaaa cacaacgtgt ttttgctcgt gagtttactg gcaaacgcta cgatgttggg 780
 gacaaglttg gcttcatgaa aacatctatc gattatgccc taaaacacca tcaagtcaaa 840
 gatgacctaa aagcttatai tatcaagtta ggtaaagaat tagaaaaagc acaagattcc 900
 aaagaaagca aa 912

<210> 31
 <211> 304
 <212> PRT
 <213> 链球菌属

<400> 31

Met Thr Lys Val Arg Lys Ala Ile Ile Pro Ala Ala Gly Leu Gly Thr
 1 5 10 15

Arg Phe Leu Pro Ala Thr Lys Ala Leu Ala Lys Glu Met Leu Pro Ile
 20 25 30

Val Asp Lys Pro Thr Ile Gln Phe Ile Val Glu Glu Ala Leu Arg Ser
35 40 45

Gly Ile Glu Glu Ile Leu Val Val Thr Gly Lys Ser Lys Arg Ser Ile
50 55 60

Glu Asp His Phe Asp Ser Asn Phe Glu Leu Glu Tyr Asn Leu Gln Glu
65 70 75 80

Lys Gly Lys Thr Glu Leu Leu Lys Leu Val Asp Glu Thr Thr Ser Ile
85 90 95

Asn Leu His Phe Ile Arg Gln Ser His Pro Lys Gly Leu Gly Asp Ala
100 105 110

Val Leu Gln Ala Lys Ala Phe Val Gly Asn Glu Pro Phe Ile Val Met
115 120 125

Leu Gly Asp Asp Leu Met Asp Ile Thr Asn Thr Lys Ala Val Pro Leu
130 135 140

Thr Lys Gln Leu Met Asp Asp Tyr Glu Thr Thr His Ala Ser Thr Ile
145 150 155 160

Ala Val Met Lys Val Pro His Asp Asp Val Ser Ser Tyr Gly Val Ile
165 170 175

Ala Pro Asn Gly Lys Ala Leu Asn Gly Leu Tyr Ser Val Asp Thr Phe
180 185 190

Val Glu Lys Pro Lys Pro Glu Asp Ala Pro Ser Asp Leu Ala Ile Ile
195 200 205

Gly Arg Tyr Leu Leu Thr Pro Glu Ile Phe Asp Ile Leu Glu Asn Gln
210 215 220

Ala Pro Gly Ala Gly Asn Glu Val Gln Leu Thr Asp Ala Ile Asp Thr
225 230 235 240

Leu Asn Lys Thr Gln Arg Val Phe Ala Arg Glu Phe Thr Gly Lys Arg
245 250 255

Tyr Asp Val Gly Asp Lys Phe Gly Phe Met Lys Thr Ser Ile Asp Tyr
260 265 270

Ala Leu Lys His His Gln Val Lys Asp Asp Leu Lys Ala Tyr Ile Ile
275 280 285

Lys Leu Gly Lys Glu Leu Glu Lys Ala Gln Asp Ser Lys Glu Ser Lys
290 295 300

<210> 32

<211> 876

<212> DNA

<213> 枯草芽孢杆菌(bacillus subtilis)

<400> 32

atgaaaaaag tacgtaaagc cataattcca gcagcaggct taggaacacg ttttcttccg

60

gctacgaaag caatgccgaa agaaatgctt cctatcgttg ataaacctac cattcaatac 120
 ataattgaag aagctgttga agccgggtatt gaagatatta ttatcgtaac aggaaaaagc 180
 aagcgtgcga ttgaggatca ttttgattac tctcctgagc ttgaaagaaa cctagaagaa 240
 aaaggaaaaa ctgagctgct tgaaaaagtg aaaaaggctt ctaacctggc tgacattcac 300
 tataiccgcc aaaaagaacc taaaggcttc ggacatgctg tctggtgcgc acgcaacttt 360
 atcggcgatg agccgtttgc ggtactgctt ggtgacgata ttgttcaggc tgaaactcca 420
 gggttgcgcc aatlaatgga tgaatatgaa aaaacacttt cttctattat cgggtttcag 480
 cagglgcccc aagaagaaac acaccgctac ggcatlaltg acccgctgac aagtgaaggc 540
 cgccgllatc aggtgaaaaa cttcgttgaa aaaccgccta aaggcacagc accttctaata 600
 ctlgccatct taggcegtta cgtattcacg cctgagatct tcatgtatit agaagagcag 660
 caggltggcg ccggcggaga aattcagctc acagacgcca ttcaaaagct gaatgaaatt 720
 caaagagigt ttgcctacga ttttgaaggc aagcgttatg atgttggatg aaagctcggc 780
 tttatcacia caactcttga atttgcgatg caggataaag agcttcgcca tcagctcgtt 840
 ccatlaltgg aagglttact aaacaaagaa gaaatc 876

<210> 33

<211> 292

<212> PRT

<213> 枯草芽孢杆菌(bacillus subtilis)

<400> 33

Met Lys Lys Val Arg Lys Ala Ile Ile Pro Ala Ala Gly Leu Gly Thr
1 5 10 15

Arg Phe Leu Pro Ala Thr Lys Ala Met Pro Lys Glu Met Leu Pro Ile
20 25 30

Val Asp Lys Pro Thr Ile Gln Tyr Ile Ile Glu Glu Ala Val Glu Ala
35 40 45

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

Glu Asp His Phe Asp Tyr Ser Pro Glu Leu Glu Arg Asn Leu Glu Glu
65 70 75 80

Lys Gly Lys Thr Glu Leu Leu Glu Lys Val Lys Lys Ala Ser Asn Leu
85 90 95

Ala Asp Ile His Tyr Ile Arg Gln Lys Glu Pro Lys Gly Leu Gly His
100 105 110

Ala Val Trp Cys Ala Arg Asn Phe Ile Gly Asp Glu Pro Phe Ala Val
 115 120 125

Leu Leu Gly Asp Asp Ile Val Gln Ala Glu Thr Pro Gly Leu Arg Gln
 130 135 140

Leu Met Asp Glu Tyr Glu Lys Thr Leu Ser Ser Ile Ile Gly Val Gln
 145 150 155 160

Gln Val Pro Glu Glu Glu Thr His Arg Tyr Gly Ile Ile Asp Pro Leu
 165 170 175

Thr Ser Glu Gly Arg Arg Tyr Gln Val Lys Asn Phe Val Glu Lys Pro
 180 185 190

Pro Lys Gly Thr Ala Pro Ser Asn Leu Ala Ile Leu Gly Arg Tyr Val
 195 200 205

Phe Thr Pro Glu Ile Phe Met Tyr Leu Glu Glu Gln Gln Val Gly Ala
 210 215 220

Gly Gly Glu Ile Gln Leu Thr Asp Ala Ile Gln Lys Leu Asn Glu Ile
 225 230 235 240

Gln Arg Val Phe Ala Tyr Asp Phe Glu Gly Lys Arg Tyr Asp Val Gly
 245 250 255

Glu Lys Leu Gly Phe Ile Thr Thr Thr Leu Glu Phe Ala Met Gln Asp
 260 265 270

Lys Glu Leu Arg Asp Gln Leu Val Pro Phe Met Glu Gly Leu Leu Asn
 275 280 285

Lys Glu Glu Ile
 290

<210> 34

<211> 1380

<212> DNA

<213> 链球菌属

<400> 34

atgaaaaact acgccattat cctagcagct ggaaagggaa cccgcatgaa ttcaggcctt 60

tccaagggtgc tgcacaaggt atcaggccta agcatgctgg agcatgtcct caagagcgtc 120

tcagccctag ctctcaaaa gcaactcaca gtgatcggtc atcaggcaga gcaagtacgt 180

gccgtcctag gtgatcaatt actgacagtg gtgcaagagg agcagctagg aacaggccat 240

gcagtcatga tggcagaaga ggagctatct ggcttagaag ggcagaccct agtgattgca 300

ggtgacaccc ccttgatcag aggagaaagc ctcaaggctc tgctagaacta tcatacaga 360

gaaaagaatg tggcaacat tctcacagcc aatgccaaagg atcccittgg ctacggccga 420

atcattcgca atgcagcagg agaggtggtc aacatcgttg aacaaaagga cgctaattgag 480

gcagagcaag aggtcaagga gatcaacaca gggacctata tctttgacaa taagcgctc 540
 tttagggctc taaagcatct cactgactgat aatgcccaag gggaatatta cctaaccgat 600
 gtgatcaglia ttttcaaggc cagccaagaa aaggttggag cttacctgct gaaggatfff 660
 gatgaaagcc taggggttaa tgatcgccca gctctagccc aggctgaggt gatcatgcag 720
 gagcggatca acaagcagca catgcittaat ggggtgaccc tgcaaaaccc tgcagctacc 780
 tatalcgaaa gcagtgtaga gattgcgccc gacgtcttga ttgaagctaa tgtgacccta 840
 aagggacaga clagaattgg cagcagaagt gttataacca atgggagcta tacccttgat 900
 tcaaggcttg gtgagggcgt agtgggtgagc cagtcagtga ttgagggctc agtcctagca 960
 gatgggttga cagtagggcc ctatgcacac attcgcccgg actctcagct cgatgagtgt 1020
 gttcatattg ggaactttgt agaggttaag gggctctcacc taggggcca taccaaggca 1080
 gggcatttga cttatctggg gaatgccgag attggctcag aggttaatat tggtcagga 1140
 agcattacgg tlaattaatga tggcaacgg aaataccaga cagtgatgg cgatcacgct 1200
 tttattggga gtcattcgac ttgatagct ccggtagagg ttggggagaa tgcittaca 1260
 gcagcagggt clacgatagc ccagtcgggt ccagcagaca gtgtggctat agggcgtagc 1320
 cgtcagggtg tgaaggaagg ctatgcccaag aggcaccac atcacccgga tcagccccag 1380

<210> 35
 <211> 460
 <212> PRT
 <213> 链球菌属

<400> 35

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

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

Asp Ser Val Ala Ile Gly Arg Ser Arg Gln Val Val Lys Glu Gly Tyr
 435 440 445

Ala Lys Arg Leu Pro His His Pro Asp Gln Pro Gln
 450 455 460

<210> 36

<211> 1368

<212> DNA

<213> 枯草芽孢杆菌(bacillus subtilis)

<400> 36

```

atggataagc gglllgcagt lgttttagcg gctggacaag gaacgagaat gaaatcgaag      60
cittalaaag tccllcatcc agtttgcggt aagcclatgg tagagcacgt cgtggacgaa     120
gccllaaaat talctllatc aaagcttgtc acgattgtcg gacalggltgc ggaagaagtg     180
aaaaagcagc ttgglgataa aagcgagtac gcgcttcaag caaaacagct tggcactgct     240
calgcigtta aacaggcaca gccatltctt gctgacgaaa aaggcgtcac aattgtcatt     300
tgcggagata cgccgctttt gacagcagag acgatggaac agatgctgaa agaacataca     360
caaagagaag cgaaagctac gatlltaact gcggttgtag aagatccaac tggatacggc     420
cgcatlaltc gcagcgaaaa cggagcggll caaaaaatag ttgagcataa ggacgcctct     480
gaagaagaac gtlclgtaac tgagatcaac accggtacgt attgttttga caatgaagcg     540
ctalllcggg clattgatca ggltgtclaat gataatgcac aaggcgagta tlatttgccg     600
gatgicatag agatlcitaa aaatgaaggc gaaactgttg ccgcttacca gactggtaat     660
tlccaagaaa cgctcggagt taatgataga gllgtclctt ctcaggcaga acaatttatg     720
aaagagcgca ttaataaacg gcatatgcaa aatggcggtga cgttgattga cccgatgaat     780
acgtataltt clcctgacgc lgtlatcgga agcgatactg tgatttacc c tggaaactgtg     840
attaaagggt aggtgcaaat cggagaagat acgattattg gccctcatac ggagattatg     900
aataglgcca llggcagccg tacggttatt aaacaatcgg tagtcaatca cagtaaagtg     960
gggaatgatg taacatagg accttttgct cacatcagac ctgattctgt catcgggaat    1020
gaagtgaaag tcgggaatlt tglagaaatt aaaaagactc aattcggaga ccgaagcaag    1080
gcalcicatc taagclatgt cggcgatgct gaggtaggca ctgatgtaaa cctgggctgc    1140
ggllcaatta ctgcaatta tgatggaaag aataagtatt tgacaaaaat tgaagatggc    1200
gcgtllatcg gctgcaatc caacttggtt gccctgtca cagtcggaga aggccttat    1260
gtggcggcag gllcaactgt tacggaagat gtacctggaa aagcacttgc tattgccaga    1320
gcgagacaag taaataaaga cgatlatgtg aaaaatattc ataaaaaa    1368

```

<210> 37
 <211> 456
 <212> PRT
 <213> 枯草芽孢杆菌(bacillus subtilis)

<400> 37

Met Asp Lys Arg Phe Ala Val Val Leu Ala Ala Gly Gln Gly Thr Arg
 1 5 10 15

Met Lys Ser Lys Leu Tyr Lys Val Leu His Pro Val Cys Gly Lys Pro
 20 25 30

Met Val Glu His Val Val Asp Glu Ala Leu Lys Leu Ser Leu Ser Lys
 35 40 45

Leu Val Thr Ile Val Gly His Gly Ala Glu Glu Val Lys Lys Gln Leu
 50 55 60

Gly Asp Lys Ser Glu Tyr Ala Leu Gln Ala Lys Gln Leu Gly Thr Ala
 65 70 75 80

His Ala Val Lys Gln Ala Gln Pro Phe Leu Ala Asp Glu Lys Gly Val
 85 90 95

Thr Ile Val Ile Cys Gly Asp Thr Pro Leu Leu Thr Ala Glu Thr Met
 100 105 110

Glu Gln Met Leu Lys Glu His Thr Gln Arg Glu Ala Lys Ala Thr Ile
 115 120 125

Leu Thr Ala Val Ala Glu Asp Pro Thr Gly Tyr Gly Arg Ile Ile Arg
 130 135 140

Ser Glu Asn Gly Ala Val Gln Lys Ile Val Glu His Lys Asp Ala Ser
 145 150 155 160

Glu Glu Glu Arg Leu Val Thr Glu Ile Asn Thr Gly Thr Tyr Cys Phe
 165 170 175

Asp Asn Glu Ala Leu Phe Arg Ala Ile Asp Gln Val Ser Asn Asp Asn
 180 185 190

Ala Gln Gly Glu Tyr Tyr Leu Pro Asp Val Ile Glu Ile Leu Lys Asn
 195 200 205

Glu Gly Glu Thr Val Ala Ala Tyr Gln Thr Gly Asn Phe Gln Glu Thr
 210 215 220

Leu Gly Val Asn Asp Arg Val Ala Leu Ser Gln Ala Glu Gln Phe Met
 225 230 235 240

Lys Glu Arg Ile Asn Lys Arg His Met Gln Asn Gly Val Thr Leu Ile
 245 250 255

Asp Pro Met Asn Thr Tyr Ile Ser Pro Asp Ala Val Ile Gly Ser Asp
 260 265 270

Thr Val Ile Tyr Pro Gly Thr Val Ile Lys Gly Glu Val Gln Ile Gly
275 280 285

Glu Asp Thr Ile Ile Gly Pro His Thr Glu Ile Met Asn Ser Ala Ile
290 295 300

Gly Ser Arg Thr Val Ile Lys Gln Ser Val Val Asn His Ser Lys Val
305 310 315 320

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

Val Ile Gly Asn Glu Val Lys Ile Gly Asn Phe Val Glu Ile Lys Lys
340 345 350

Thr Gln Phe Gly Asp Arg Ser Lys Ala Ser His Leu Ser Tyr Val Gly
355 360 365

Asp Ala Glu Val Gly Thr Asp Val Asn Leu Gly Cys Gly Ser Ile Thr
370 375 380

Val Asn Tyr Asp Gly Lys Asn Lys Tyr Leu Thr Lys Ile Glu Asp Gly
385 390 395 400

Ala Phe Ile Gly Cys Asn Ser Asn Leu Val Ala Pro Val Thr Val Gly
405 410 415

Glu Gly Ala Tyr Val Ala Ala Gly Ser Thr Val Thr Glu Asp Val Pro
420 425 430

Gly Lys Ala Leu Ala Ile Ala Arg Ala Arg Gln Val Asn Lys Asp Asp
435 440 445

Tyr Val Lys Asn Ile His Lys Lys
450 455

<210> 38

<211> 1347

<212> DNA

<213> 链球菌属

<400> 38

atgtcacata ttacatttga ttattcaaag gtctttgagc aatttgccgg acagcatgaa 60

attgactttt tacaaggcca ggtaacagag gctgatcagg cactacgtca gggcactgga 120

cctggatcag atttcttggg ctggcttgag ttacctgaaa actatgacaa agaagaattt 180

gctcgtatcc ttaaagcagc tgagaagatt aaggctgaca gtgacgttct tgttgtgatt 240

ggatattggg gctcttacct tgggtgctaag gctgcaattg actttttgaa cagccatttt 300

gccaacctac aaacagcaaa agagcgcgaaa gcaccacaaa ttctttatgc tggttaactcc 360

atctcatcaa gctatcttgc tgatctttgt gactatgttc aagataaaga tttctctgtt 420

aacgtgatit ctaagtcagg tacaacaaca gagcctgcaa tgccttttcg tgtctttaaa 480

gaattacttg ttaaaaagta cggtaagaa gaggccaaca agcgtatcta tgcaacgact 540
 gataaggta aggggtgctgt taaggttgag gctgatgcaa atcattggga aacctttgtt 600
 glgccagata atgttggtgg ccgtttctca gtgctgacag ctgtgggctt gctaccaatt 660
 gcagcatcag gggctgatat taccgctg aggaaggag caaatgcagc tcgtaaggac 720
 ctgtcatcag ataaaaatctc agaaaacatc gcttaccaat atgctgtggt ccgcaatata 780
 ctctatcgca aaggclatgt aactgaaatt ttggcaact atgagccatc attgcagtat 840
 tttagcgaat ggiggaagca actggctggt gagtctgaag gaaaggacca aaagggtatt 900
 tacccaactt cagctaattt ctgacagac ctgcattctc ttggtcaatt tatccaagaa 960
 ggctaccgia accctittga gacagtgatt cgtgtggaca agccacgta aatgtgatt 1020
 atcccagaaa tggctgagga ccttgatggt cttggctacc tacaaggaaa agacgttgac 1080
 ttigtcaaca aaaaagcaac agatgggtgc cttcttgccc atacagatgg tgggtgcca 1140
 aataiglita tcacgcttcc agagcaagac gaatttacac taggctatac gatctacttc 1200
 ttigagcttg ctattgcctt ttcaggctac ctcaacgggg tcaatccatt tgatcagcca 1260
 ggcgltgagg cttaacaagaa aaacatgttt gcccttcttg gtaagccagg ctttgaagag 1320
 ctaggagcag cgtcaacgc acgcttg 1347

<210> 39
 <211> 449
 <212> PRT
 <213> 链球菌属

<400> 39

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

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

Phe Asp Gln Pro Gly Val Glu Ala Tyr Lys Lys Asn Met Phe Ala Leu
 420 425 430

Leu Gly Lys Pro Gly Phe Glu Glu Leu Gly Ala Ala Leu Asn Ala Arg
 435 440 445

Leu

<210> 40
 <211> 5158
 <212> DNA
 <213> 类马链球菌 (*Streptococcus equisimilis*)

<400> 40
 tcaatttatg gctttttgct gatagcttac ctattagtca aaatgicctt atcctttttt 60
 tacaagccat ttaagggaag ggctgggcaa tataagggtg cagccattat tccctcttat 120
 aacgaagatg ctgagtcatt gctagagacc ttaaaaagtg ttcagcagca aacctatccc 180
 ctagcagaaa tttatgttgt tgacgatgga agtgctgatg agacaggat taagcgcatt 240
 gaagaclatg tgcgtgacac tggtagccia tcaagcaatg tcatigtica tcggtcagag 300
 aaaaatcaag gaaagcgica tgcacaggcc tgggcctttg aaagatcaga cgcgatgctc 360
 tttttgaccg ttgactcaga tacttataic taccctgatg ctttagagga gtigttaaaa 420
 acctttaatg acccaactgt ttttgctgcg acgggtcacc ttaatgicag aaatagacaa 480
 accaatctct taacacgctt gacagalatt cgctatgata atgcitttgg cgttgaacga 540
 gctgccaat ccgttacagg taalatecct gtttgcctag gtccgcttag cgtttacaga 600
 cgcgaggigg ttgttccctaa calagataga tacatcaacc agaccttctt gggtattcct 660
 glaaglatig gtagtgacag gtgcttgacc aactatgcaa ctgatttagg aaagactggt 720
 talcaateca ctgctaaatg tattacagat gttcctgaca agatgcttac ttacttgaag 780
 cagcaaaaacc gciggaacaa gtccttcttt agagagtcca ttatttctgt taagaaaatc 840
 atgaacaatc cttttgtagc cctatggacc atacttgagg tgictatggt tatgatgctt 900
 gtttattctg tggiggallt cttttaggce aaigtacagag aatttgattg gctcagggtt 960
 ttagccittc tggigallat cticattggt gccctgtgtc ggaacattca ttacatgctt 1020
 aagcaccgcg tgccttctt gttatctccg ttttatgggg tgctgcattt gtttgcctc 1080
 cagccctiga aallatattc tctttttact attagaaatg ctgactgggg aacacgtaaa 1140
 aaattiatat aaaccaacta gacctagggt ctgacaaggg agctaagcta gggataaaca 1200
 aagagltttg atccgactcg agcagctcat aaacgaaagc tatcccactt gtaattgaag 1260
 ciaagagctt ttagcttgca gctctataaa gacgaaccag aggctgagtg tcagctttgg 1320

| | |
|---------------------------------------------------------------------|------|
| tgtgagggct aggtcattat gatccttcag gtgtggcacc tgagctccgg cagtagctaa | 1380 |
| ctgtactaag giatcaaagg aaaaaatgaa gtgaaaattt ctgtagcagg ctcaggatat | 1440 |
| gtcggcctat ccttgaglat tttactggca caacataatg acgtcactgt tgttgacatt | 1500 |
| atlgatgaaa aggtgagatt gatcaatcaa ggcatatcgc caatcaagga tgctgatatt | 1560 |
| gaggaglatl taaaaaatgc gccgctaaat ctacacagca cgcttgalgg cgcaagcgct | 1620 |
| tatagcaatg cagaccittat taitcattgct actccgacaa attatgacag cgaacgcaac | 1680 |
| tactttgaca caaggcatgt tgaagaggic atcgagcagg tcctagacct aaatgcgtca | 1740 |
| gcaaccatta ttalcaaatc aaccatacca ctaggcttta tcaagcatgt tagggaaaaa | 1800 |
| taccagacag atcglattat ttttagccca gaatttttaa gagaatcaaa agccttatac | 1860 |
| galaaccittt acccaagtgc gatcattggt tcttatgaaa aggacgactc accaagggtt | 1920 |
| atlcaggctg claaagcctt tgctggctct ttaaaggaag gagccaaaag caaggatact | 1980 |
| ccggctttat ttatgggctc acaggaggct gaggcggctca agctatttgc gaataccttt | 2040 |
| ttggctatgc ggggtcttta ctttaatgaa ttagacacct attccgaaag caagggtcta | 2100 |
| gatgctcagc gcgtgattga aggagtcctgt catgatcagc gcattggtaa ccattacaat | 2160 |
| aacccttccct ttggatatgg cggctattgc ctgccaaagg acagcaagca gctgttggca | 2220 |
| aallatagag gcatlcccca glccttgalg tcagcgattg ttgaatccaa caagatacga | 2280 |
| aaatcttatt tggctgaaca aatattagac agagcctcta gtcaaaaagca ggctgggtga | 2340 |
| ccatlaacga ttggctttta ccgcttgatt atgaaaagca actctgataa tttccgagaa | 2400 |
| agcgccatta aagatattat tgatatcacc aacgactatg gggttaatat tgtcatttac | 2460 |
| gaaccatgc ttggcgagga taltggctac agggttgca aggacttaga gcagttcaaa | 2520 |
| aacgagtcta caalcattgt gtcaaatcgc tttgaggacg acctaggaga tgtcattgat | 2580 |
| aaggtttata cgagagatgt ctttgggaaga gactagtcag aaaacgaatg gcactcataa | 2640 |
| ggaaccacaa atcaaggagg aactcatgac aaaggtcaga aaagccatta tcccagccgc | 2700 |
| cgccclaggc actcgcctcc tgcccgccac caaggcactg gccaaaggaaa tgcctccaat | 2760 |
| cgctegalaag ccaaccatcc aaltcatcgt cgaggaagcc ctaaaggcag gtatcgagga | 2820 |
| gattcttgc gtcaccggca aggccaaacg ctctatcgag gaccactttg actccaactt | 2880 |
| cgagctcgaa tacaatctcc aagccaaggg caaaaaccgag ctactcaagc tcgttgatga | 2940 |
| gaccactgcc alcaacctgc acitcattcg tcagagccac cctagaggac taggggacgc | 3000 |
| tgtccctcaa gccaaagcct ttgttggcaa tgagcccttt gtggctcatgc tgggggatga | 3060 |

| | |
|--------------------------------------------------------------------|------|
| ccicatggat attaccaatc ctagtgccaa gcccttgacc aagcagctta ttgaggatta | 3120 |
| tgatlgcaca cacgcctcaa cgattgcagt gatgagggtg ccgcatgagg aggtttccaa | 3180 |
| ttalgggtgtg attgaccgc aaggggaaggc tgtaagggc ttgtatagtg tggagacctt | 3240 |
| tgltgagaag ccaagtccag atgaggcacc gagtgacitg gcgattattg gtcgatattt | 3300 |
| gttgacgcct gagatllttg ccatattgga gaagcaggcg cctggagctg gcaatgaggt | 3360 |
| acagctgacc galgcgattg acaagctcaa taagacacag cgggtttttg cgagggagtt | 3420 |
| taaggagag cggtatgag ttggggacaa gtttggttt atgaagacct cacttgacta | 3480 |
| tgctctcaag cacccicagg tcaaggacga cctcactgac tacattataa agctcagtaa | 3540 |
| gcaactgaac aaggacgtca agaaataggc gtttattgat cagctattgc agagctattt | 3600 |
| aaaagcattt agagctllaa ggtgggatac tagaggattg gtatcicact ttllaggctg | 3660 |
| acttglatta alaccaaaag ccaaaactag gcagataagc ataaggaatt agatlaaaaa | 3720 |
| taaggaacca aaacatgaaa aactacgcca ttatcctagc agctggaaag ggaacgcgca | 3780 |
| tgaagtcagc gcttcccaag glgctgcaca aggtatcagg cctaagcatg ctggagcatg | 3840 |
| tcctcaagag tglctcagcc ctageccctc aaaagcagct cacagtgatc ggtcatcagg | 3900 |
| cagagcaggt gcgtgctgtc ctaggagagc aatcgctaac agtggtgcaa gaggagcagc | 3960 |
| tagggacagg ccatgcagtc atgatggcag aagaggagct atctggctta gaggggcaaa | 4020 |
| ccctagtgat tgcaggtgac acccccttga tcagaggaga aagcctcaag gctctgctag | 4080 |
| actatcatat cagagaaaag aatgtggcaa ccattctcac agccaatgcc aaggatccct | 4140 |
| tiggctatgg acgaatcatt cgcaatgcag caggagaggt ggtcaacatc gttgagcaaa | 4200 |
| aggatgctaa tgaggcagag caagaggica aggagatcaa cacagggact tatactttg | 4260 |
| acaalaagcg cctllttgag gctctaaagc atctcacgac tgataatgcc caaggggagt | 4320 |
| actacctaac cgatlgatc agtattttca aggctggcca agaaagggtt ggcgcttacc | 4380 |
| tgctgaagga clllgatgag agcctagggg ttaatgatcg cttagctcta gcccaggccg | 4440 |
| agglgattat gcaagagcgg atcaacaggc agcacaatgct taatggggtg accctgcaaa | 4500 |
| accggcagc tacctatatt gaaagcagtg tagagattgc accagacgtc ttgattgaag | 4560 |
| ccaatlgac ctlaaaggga cagactagaa tiggcagcag aagtgtcata agcaatggga | 4620 |
| gctatalcct tgattcgagg cttgggtgagg gtgtagtgt tagccagtcg gtgattgagg | 4680 |
| cttcagctt agcagaigga gtgacagtag ggccatatgc acacattcgc ccggactccc | 4740 |
| agctcgatga gigtgticat attgggaact ttgttagaggt taaggggtct catctagggg | 4800 |
| ccaataccaa ggcagggcat ttgacttacc tggggaatgc cgagattggc tcagaggtta | 4860 |

acattggtgc aggaagcatt acggttaatt atgatggica acggaaatac cagacagtga 4920
 ttggcgalca cgcttttatt gggagtcatt cgactttgat agctccggta gaggttgggg 4980
 agaatgcitt aacagcagca gggctctacga lagcccagtc agtgccggca gacagtgtgg 5040
 ctatagggcg cagccgtcag gtggtgaagg aaggctatgc caagaggctg ccgcaccacc 5100
 caaatcaagc claatcgctc aacccaaaaga ggcaggigag aaaacctagg ccattaaa 5158

<210> 41

<211> 3050

<212> DNA

<213> 苏云金芽孢杆菌 (*Bacillus thuringiensis*)

<400> 41

tcgaaacgta agatgaaacc ttagataaaa gtgctttttt tgttgcaatt gaagaattat 60
 taatgllaag cilaattaaa gataatatct ttgaattgta acgcccctca aaagtaagaa 120
 ctacaaaaaa agaatacgtt atatagaaat atgttlgaac cttcttcaga ttacaaatat 180
 attcggacgg actctaccctc aaatgcttat ctaactatag aatgacatac aagcacaacc 240
 ttgaaaaatt gaaaatataa ctaccaatga acttggtcat gtgaattatc gctgtattta 300
 attttcicaa ticaatatal aalatgccaa tacattgita caagtagaaa ttaagacacc 360
 ctlgatagcc ttaactatacc taacatgatg tagtattaaa tgaatatgta aatataatla 420
 tgataagaag cgacttattt ataatcattt catatttttc tattggaatg attaagattc 480
 caatagaata gigtataaat talttatctt gaaaggaggg atgcctaaaa acgaagaaca 540
 ttaaaaaaat atatttgcac cgtctaatgg atttatgaaa aatcatttta tcagittgaa 600
 aattatgtat taigalaaga aaggaggaa gaaaaatgaa tccgaacaat cgaagtgaac 660
 atgatacaat aaaaactact gaaaataatg aggtgccaac taacctgtt caalatecctt 720
 tagcggaaac tccaaatcca acactagaag atttaaatta taaagattt ttaagaatga 780
 ctgcagataa taatcggaa gcactagata gctctacaac aaaagatgic attcaaaaag 840
 gcatttccgt agtaggtgat ctccctaggcg tagtaggttt cccgtttggt ggagcgttg 900
 ttctglttta tacaacitlt ttaaactact ttiggccaag tgaagaccog tggaaggctt 960
 ttatggaaca aglagaagca ttgatggatc agaaaatagc tgattatgca aaaaataaag 1020
 ctcttgcaga gttacagggc ctccaataa atgtcgaaga ttatgtgagt gcattgagtt 1080
 catggcaaaa aaatccctgtg agttcacgaa atccacatag ccaggggagg ataagagagc 1140
 tgltttcica agcagaaagt cattttcgta attcaatgcc ttctgttgca atttctggat 1200
 acgaggltct atttctaaca acatatgcac aagctgccaa cacacattta tttttactaa 1260

| | |
|---------------------------------------------------------------------|------|
| aagacgctca aatltatgga gaagaatggg gatacgaaaa agaagatatt gctgaatltt | 1320 |
| ataaaagaca actaaaactt acgcaagaat atactgacca ttgtgtcaaa tggataaatg | 1380 |
| ttggattaga taaattaaga ggttcatctt atgaatcttg ggtaaacttt aaccgttatt | 1440 |
| gcagagagat gacattaaaca gtattagatt taattgcact atttccattg tatgatgttc | 1500 |
| ggctataccc aaaagaagt aaaccgaat taacaagaga cgttttaaca gatccaattg | 1560 |
| tcggagtcaa caaccttagg ggctatggaa caaccttctc taatatagaa aattatattc | 1620 |
| gaaaaccaca tclatttgac tctctgcala gaattcaatt tcacacgcgg ttccaaccag | 1680 |
| gatattatgg aatgactct tccaattatt ggtccggtaa ttatgtttca actagaccaa | 1740 |
| gcataggatc aatgatata atcacatctc cattctatgg aaataaatcc agtgaacctg | 1800 |
| tacaaaatlt agaatlttaal ggagaaaaag tctatagagc cgtagcaaat acaaatcttg | 1860 |
| cggctctggcc glccgcctga tcttcagggt ttacaaaagt ggaatttagc caatataatg | 1920 |
| atcaaacaga tgaagcaagt acacaaactc acgactcaaa aagaaatgtt ggcgcgggtca | 1980 |
| gctgggattc tctcgaatcaa ttgcctccag aaacaacaga tgaacctcta gaaaagggat | 2040 |
| atagccaatca actcaattat gtaatgtgct ttttaattgca gggtagtaga ggaacaatcc | 2100 |
| cagtgtaaac ttggacacat aaaagtgtag acttttttaa catgattgat tcgaaaaaaaa | 2160 |
| ttacacaact tccgttagta aaggcataia agttacaatc ttgtgtcttc gttgtcgcag | 2220 |
| gtccatagggt tacaggagga gatattcattc aatgcacaga aatgggaagt ggcgcaacta | 2280 |
| tttacgttac accggatgtg tctactctc aaaaatatcg agctagaatt cattatgctt | 2340 |
| ctacatctca galaacattt acactcagtt tagacggggc accatttaat caatactatt | 2400 |
| tcgataaaac galaataaaa ggagacacat taacgtataa ttcatttaat ttagcaagtt | 2460 |
| tcagcacacc attcgaatta tcagggaata acttacaat aggcgtcaca ggattaagtg | 2520 |
| ciggagataa agtttatata gacaaaattg aatttattcc agtgaattaa attactaga | 2580 |
| aagtaaagaa gtagtgacca tctatgatag taagcaaagg ataaaaaaaa gagttcataa | 2640 |
| aatgaataac atagtgctt tcaactttcg ctttttgaag gtagatgaag aacctattt | 2700 |
| ttattttcaa aatgaaggaa gttttaaata tgaatcatt taaagggaac aatgaaagta | 2760 |
| ggaaaalaagt cattatctat aacaaaataa catttttata tagccagaaa tgaattataa | 2820 |
| tattaatctt ttclaaattg acgtttttct aaacgttcta tagcttcaag acgcttagaa | 2880 |
| tcatcaatat ttgtatacag agctgttggt tccatcgagt tatgtccat ttgattcgct | 2940 |
| aatagaacaa gatctttatt ttcgttataa tgattgggtg cataagiatg gcgtaattta | 3000 |

tgaggccttt tcttttcatac aaaagccctc gtgtatttct ctgtaagctt 3050

<210> 42
 <211> 185
 <212> DNA
 <213> 地衣芽孢杆菌 (*Bacillus licheniformis*)

<400> 42
 ggccctaagg gcctgcaatc gattgtttga gaaaagaaga agaccataaa aataccttgt 60
 ctgtcatcag acagggattt ttttatgctg tccagactgt ccgctgtgta aaaaaaagga 120
 ataaaggggg gttgacatta ttttactgat atgtataata taatttgtat aagaaaatgg 180
 agctc 185

<210> 43
 <211> 185
 <212> DNA
 <213> 地衣芽孢杆菌 (*Bacillus licheniformis*)

<400> 43
 ggccctaagg gcctgcaatc gattgtttga gaaaagaaga agaccataaa aataccttgt 60
 ctgtcatcag acagggattt ttttatgctg tccagactgt ccgctgtgta aaaaatagga 120
 ataaaggggg gttgacatta ttttactgat atgtataata taatttgtat aagaaaatgg 180
 agctc 185

<210> 44
 <211> 185
 <212> DNA
 <213> 地衣芽孢杆菌 (*Bacillus licheniformis*)

<400> 44
 ggccctaagg gcctgcaatc gattgtttga gaaaagaaga agaccataaa aataccttgt 60
 ctgtcatcag acagggattt ttttatgctg tccagactgt ccgctgtgta aaaaatagga 120
 ataaaggggg gttgttatta ttttactgat atgtaaaata taatttgtat aagaaaatgg 180
 agctc 185

<210> 45
 <211> 27
 <212> DNA
 <213> 芽孢杆菌属 (*Bacillus*)

<400> 45
 gtccctcttg gtaccctggaa gcagagc 27

<210> 46
 <211> 39

| | | |
|------------------------------------------------------------------|--|----|
| <212> DNA | | |
| <213> 芽孢杆菌属 | | |
| <400> 46 | | |
| gtataaatat tcggccctta aggccagtac cattttccc | | 39 |
| <210> 47 | | |
| <211> 59 | | |
| <212> DNA | | |
| <213> 芽孢杆菌属 | | |
| <400> 47 | | |
| tggtactggc cllaagggcc gaatatttat acaatatcat gagctccaca ttgaaaggg | | 59 |
| <210> 48 | | |
| <211> 34 | | |
| <212> DNA | | |
| <213> 芽孢杆菌属 | | |
| <400> 48 | | |
| ggtgtctct agagcggccg cggttgcggt cagc | | 34 |
| <210> 49 | | |
| <211> 33 | | |
| <212> DNA | | |
| <213> 芽孢杆菌属 | | |
| <400> 49 | | |
| ggcctlaagg gcctgctgic cagactgtcc gct | | 33 |
| <210> 50 | | |
| <211> 18 | | |
| <212> DNA | | |
| <213> 芽孢杆菌属 | | |
| <400> 50 | | |
| ggcgtlaca ttcaaaga | | 18 |
| <210> 51 | | |
| <211> 37 | | |
| <212> DNA | | |
| <213> 地衣芽孢杆菌(<i>Bacillus licheniformis</i>) | | |
| <400> 51 | | |
| ccaggcctta agggccgcat gcgtccttct ttgtgct | | 37 |
| <210> 52 | | |
| <211> 27 | | |
| <212> DNA | | |
| <213> 地衣芽孢杆菌(<i>Bacillus licheniformis</i>) | | |
| <400> 52 | | |
| gagctccttt caatgtgata catatga | | 27 |

| | |
|------------------------------------------------|----|
| <210> 53 | |
| <211> 24 | |
| <212> DNA | |
| <213> 地衣芽孢杆菌 (<i>Bacillus licheniformis</i>) | |
| <400> 53 | |
| gaattcgacg gcttcccgtg cgcc | 24 |
| <210> 54 | |
| <211> 35 | |
| <212> DNA | |
| <213> 地衣芽孢杆菌 (<i>Bacillus licheniformis</i>) | |
| <400> 54 | |
| gcaagcgagc acggattgta agtacaagtt agata | 35 |
| <210> 55 | |
| <211> 35 | |
| <212> DNA | |
| <213> 地衣芽孢杆菌 (<i>Bacillus licheniformis</i>) | |
| <400> 55 | |
| aacttgtaact tacaatccgt gctcgttgc cgtac | 35 |
| <210> 56 | |
| <211> 28 | |
| <212> DNA | |
| <213> 地衣芽孢杆菌 (<i>Bacillus licheniformis</i>) | |
| <400> 56 | |
| aagcttccat tcaaacctgg tgaggaag | 28 |
| <210> 57 | |
| <211> 30 | |
| <212> DNA | |
| <213> 芽孢杆菌属 | |
| <400> 57 | |
| cgagctcgat gtgtataaa ttgagaggag | 30 |
| <210> 58 | |
| <211> 32 | |
| <212> DNA | |
| <213> 芽孢杆菌属 | |
| <400> 58 | |
| gcggcccgct cataaacggt gcaatcgtgc tc | 32 |
| <210> 59 | |
| <211> 30 | |
| <212> DNA | |

| | |
|----------------------------------------|----|
| <213> 芽孢杆菌属 | |
| <400> 59 | |
| cgagctcgat gtgtataaa ttgagaggag | 30 |
| <210> 60 | |
| <211> 19 | |
| <212> DNA | |
| <213> 芽孢杆菌属 | |
| <400> 60 | |
| gtcgaatatg atggggatg | 19 |
| <210> 61 | |
| <211> 35 | |
| <212> DNA | |
| <213> 芽孢杆菌属 | |
| <400> 61 | |
| ggacaaggac agattgtagc agttgctgat actgg | 35 |
| <210> 62 | |
| <211> 21 | |
| <212> DNA | |
| <213> 芽孢杆菌属 | |
| <400> 62 | |
| gcgattacag ttggggcaac c | 21 |
| <210> 63 | |
| <211> 23 | |
| <212> DNA | |
| <213> 芽孢杆菌属 | |
| <400> 63 | |
| ggtagcacga cggcatcact aac | 23 |
| <210> 64 | |
| <211> 27 | |
| <212> DNA | |
| <213> 芽孢杆菌属 | |
| <400> 64 | |
| gtccttcttg gtacctggaa gcagagc | 27 |
| <210> 65 | |
| <211> 34 | |
| <212> DNA | |
| <213> 芽孢杆菌属 | |
| <400> 65 | |
| catgctgggc ccttaaggcc agtaccattt tccc | 34 |

| | |
|-------------------------------------------------------------------|----|
| <210> 66 | |
| <211> 69 | |
| <212> DNA | |
| <213> 芽孢杆菌属 | |
| <400> 66 | |
| cagtaggcct taagggccca gcatgattga gctcaccacc atgggatccg cggccgcaca | 60 |
| agggaaggc | 69 |
| <210> 67 | |
| <211> 24 | |
| <212> DNA | |
| <213> 芽孢杆菌属 | |
| <400> 67 | |
| caaticatcc tctagagtct cagg | 24 |
| <210> 68 | |
| <211> 41 | |
| <212> DNA | |
| <213> 芽孢杆菌属 | |
| <400> 68 | |
| cattctgcag ccgcgcaaa ttccggattt tatgtaagcg g | 41 |
| <210> 69 | |
| <211> 48 | |
| <212> DNA | |
| <213> 芽孢杆菌属 | |
| <400> 69 | |
| catcataigc ggccgcittat cattgaaaaa cggtgcttaa tctcgaag | 48 |
| <210> 70 | |
| <211> 18 | |
| <212> DNA | |
| <213> 芽孢杆菌属 | |
| <400> 70 | |
| ggaagtacaa aaataagc | 18 |
| <210> 71 | |
| <211> 42 | |
| <212> DNA | |
| <213> 芽孢杆菌属 | |
| <400> 71 | |
| catcccccg ggagcittaat taaagataat atctttgaat tg | 42 |
| <210> 72 | |
| <211> 24 | |

<212> DNA
<213> 芽孢杆菌属

<400> 72
tgccgcggct gcagaatgag gcag 24

<210> 73
<211> 26
<212> DNA
<213> 芽孢杆菌属

<400> 73
glttaacttga aaaacgggtgc ttaatc 26

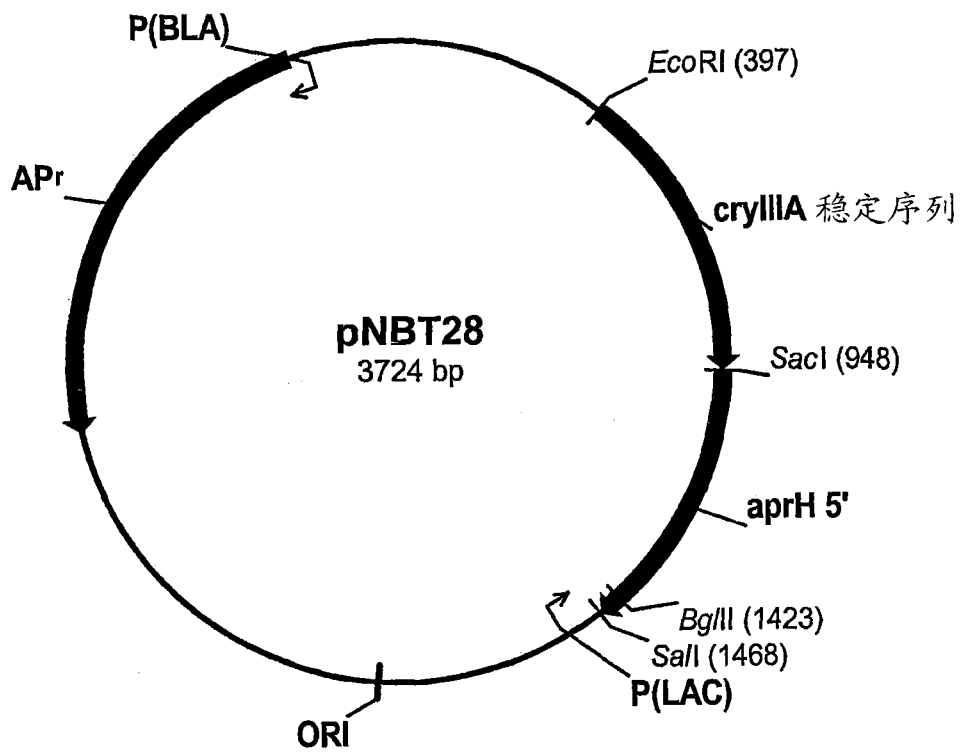


图 1

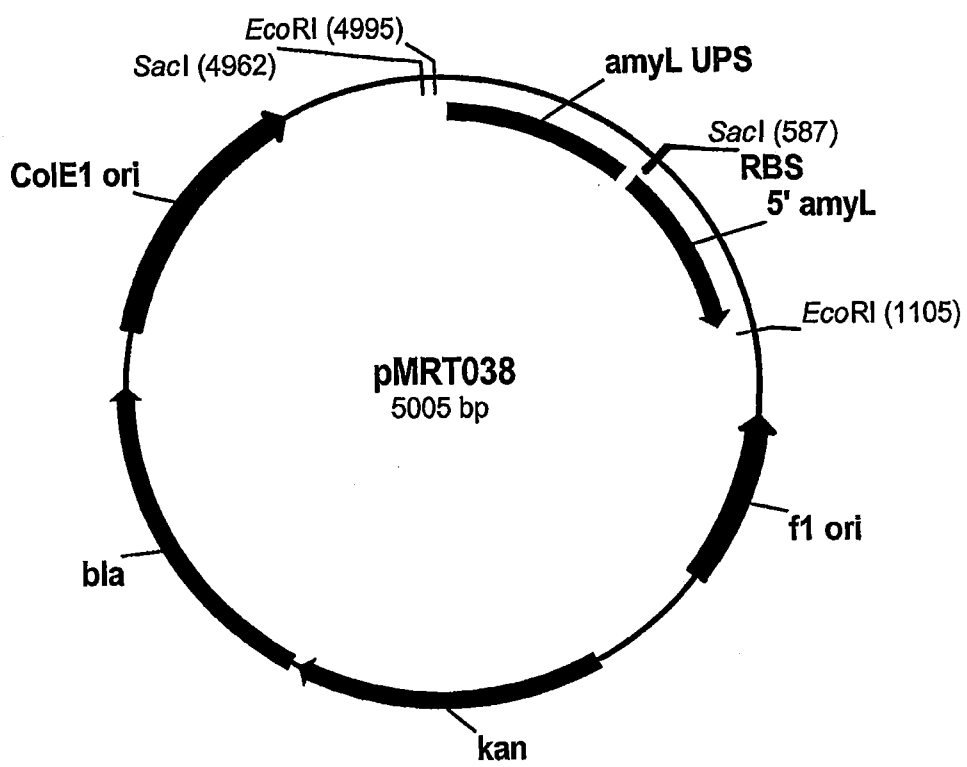


图 2

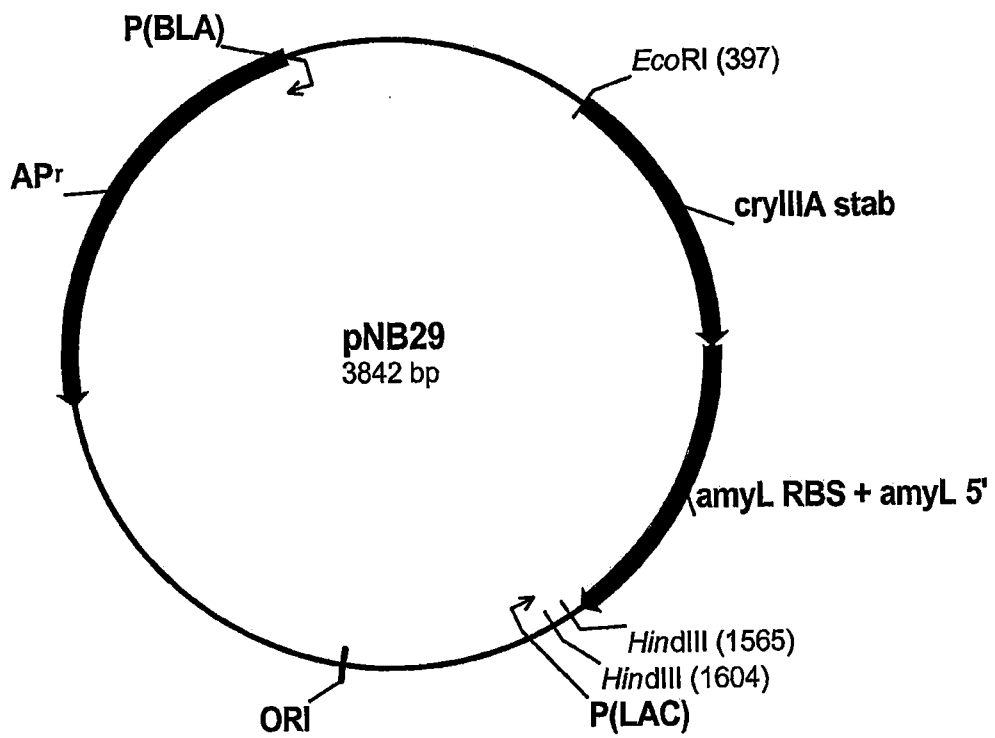


图 3

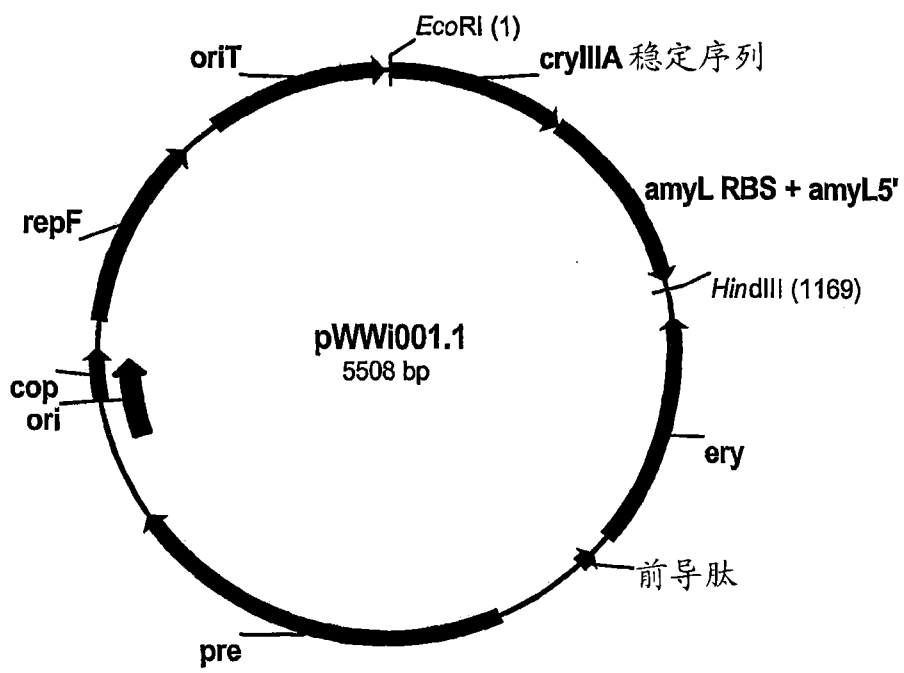


图 4

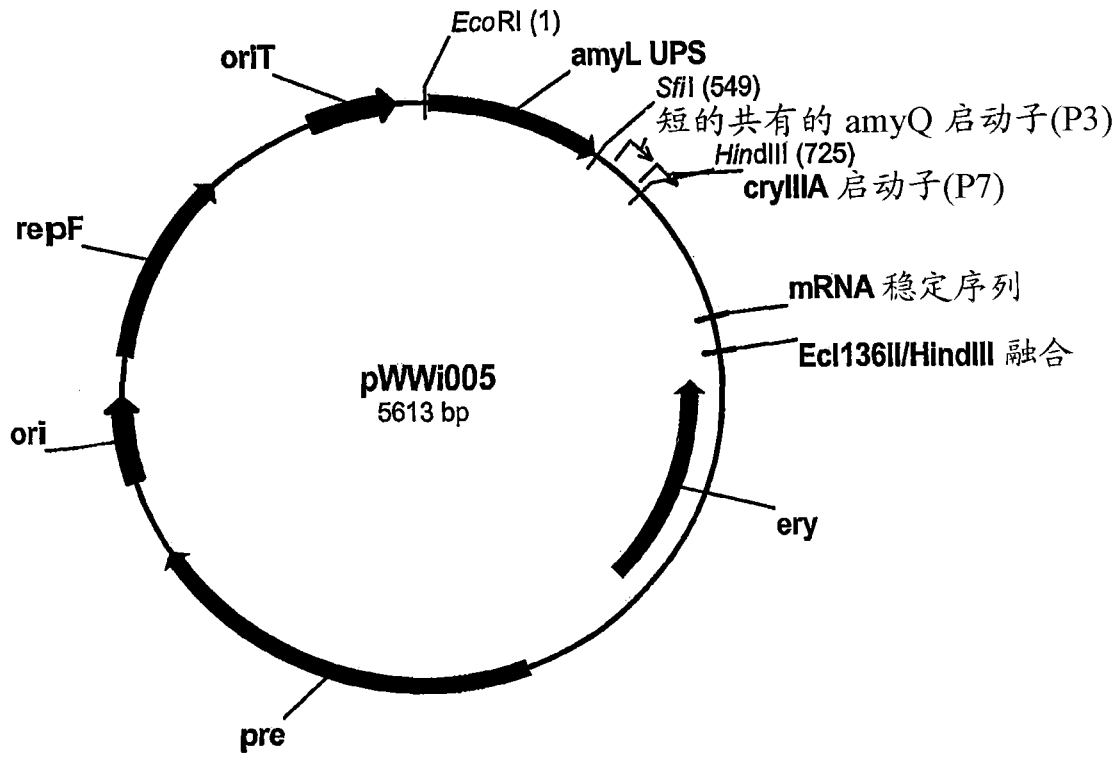


图 5

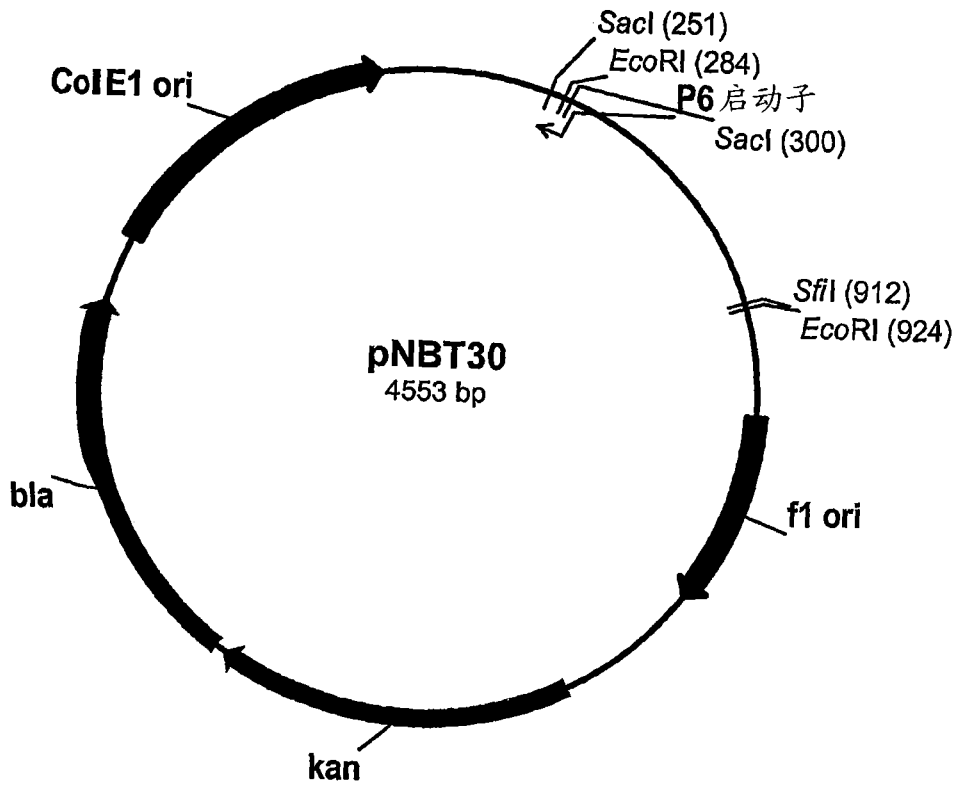


图 6

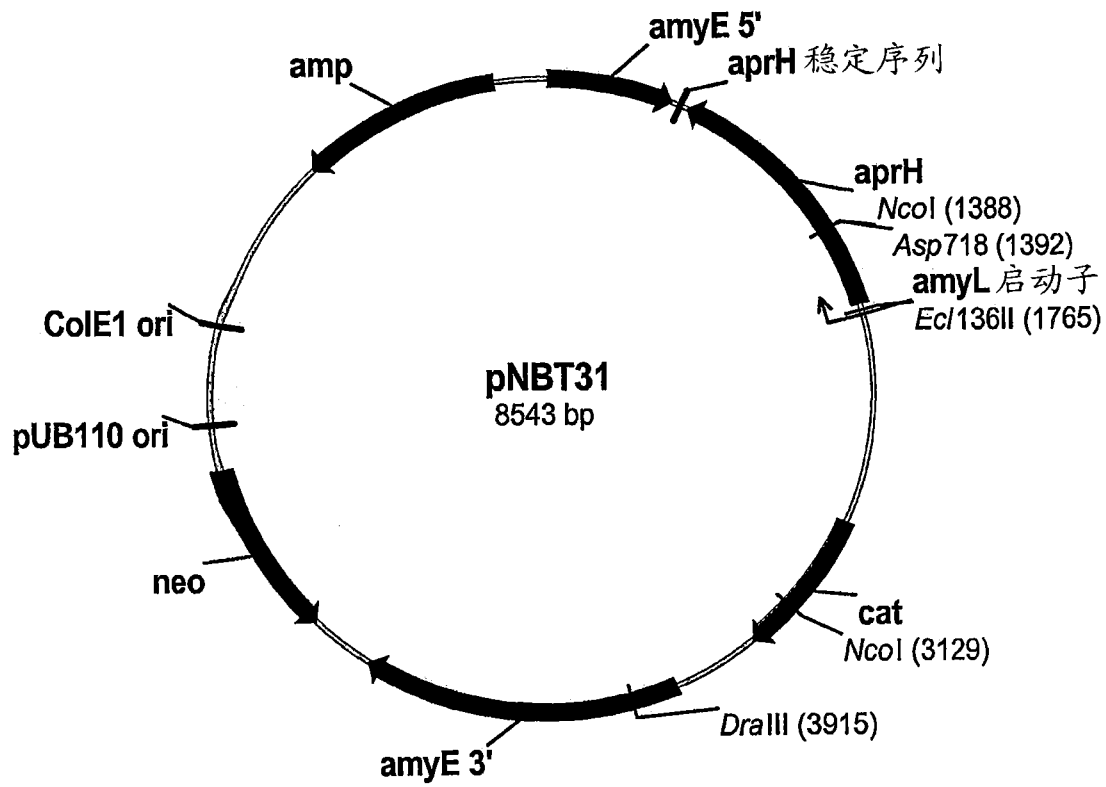


图 7

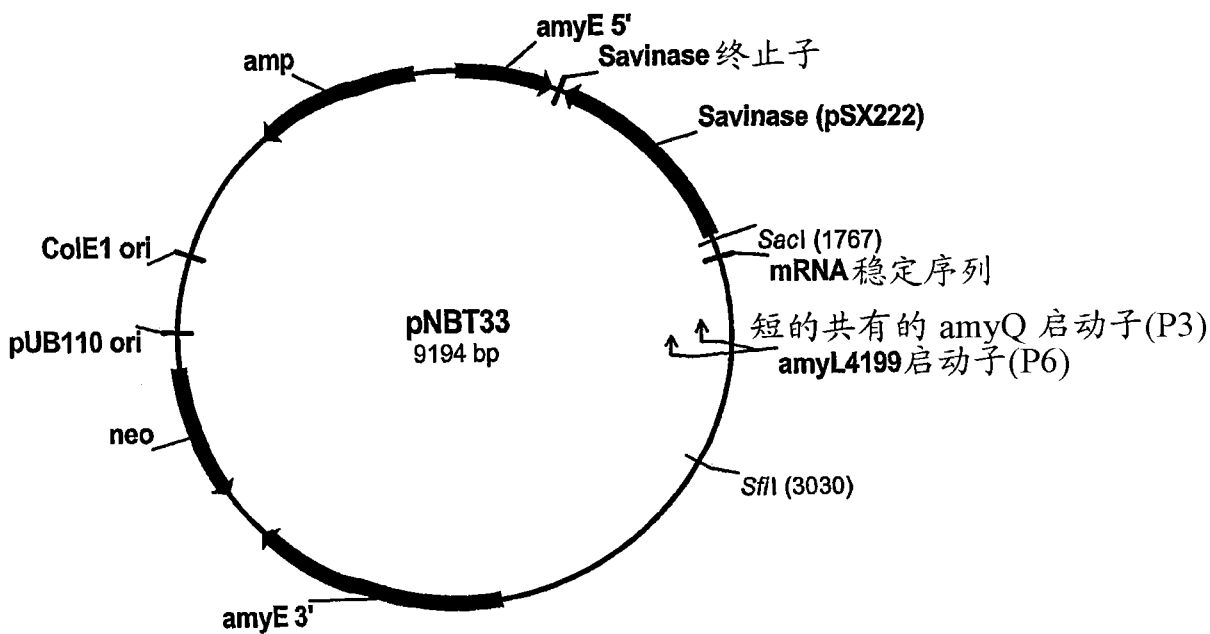


图 8

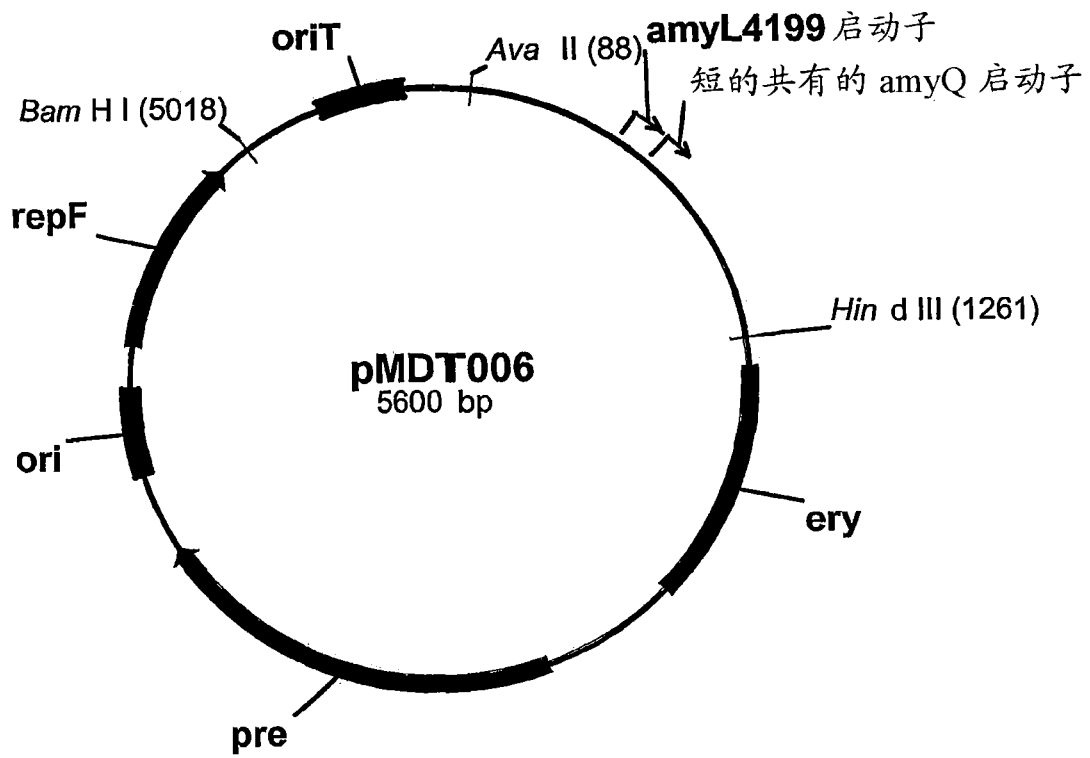


图 9

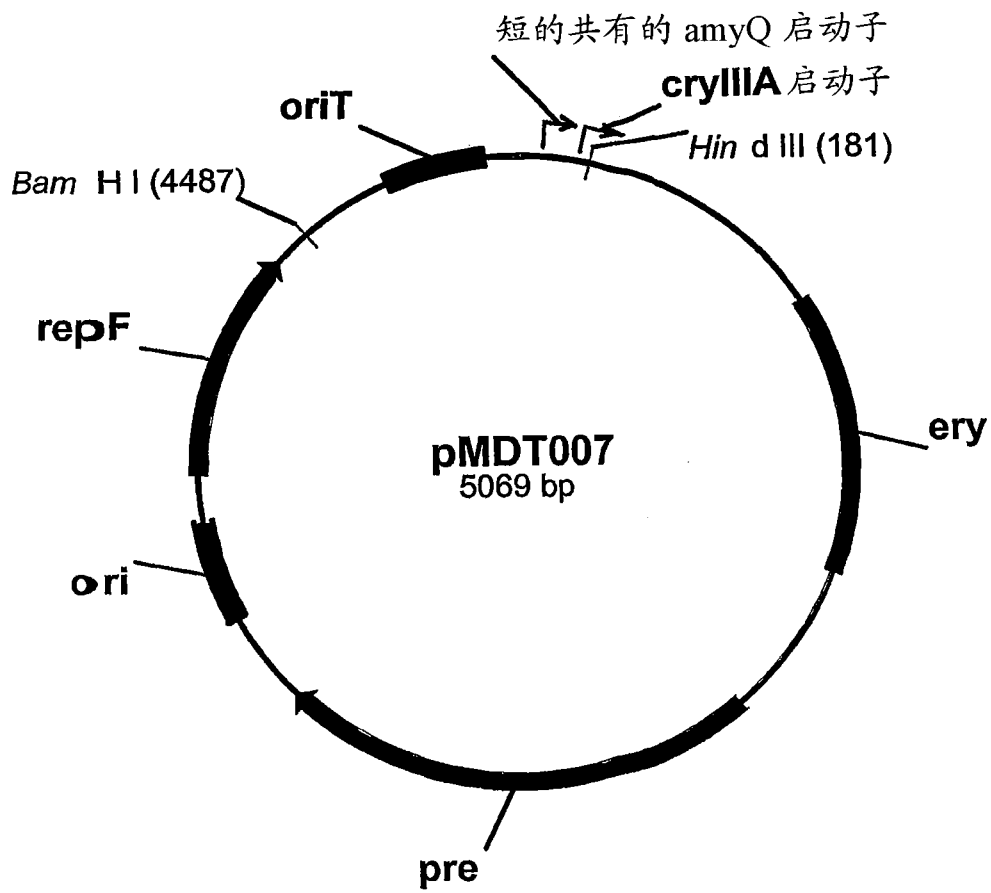


图 10

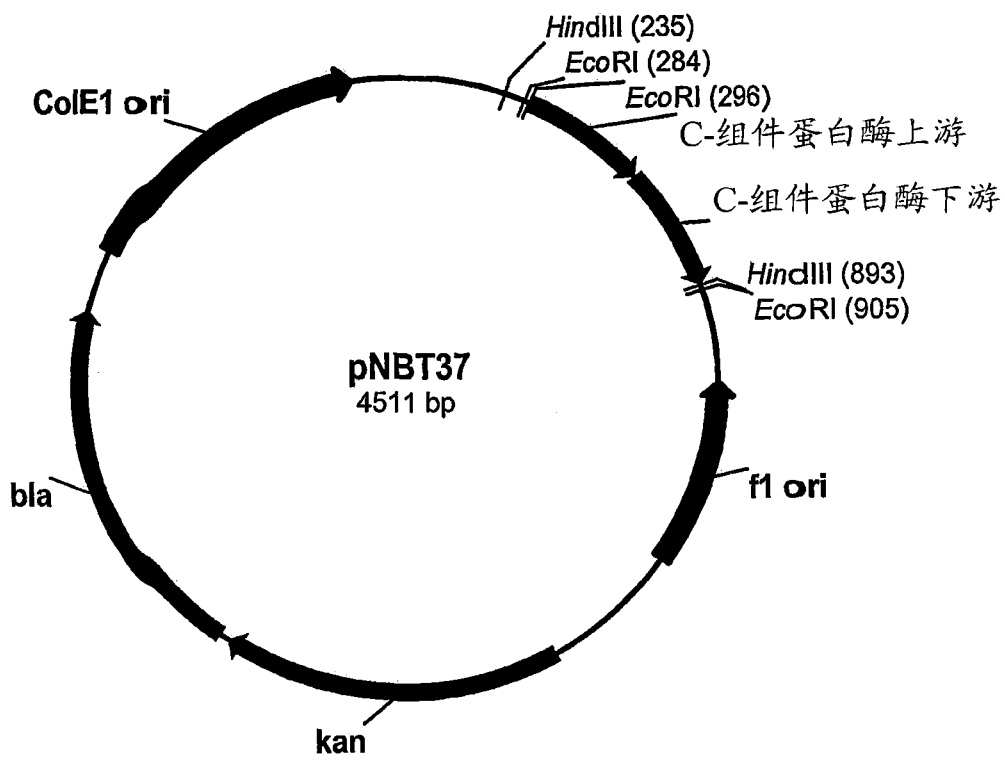


图 11

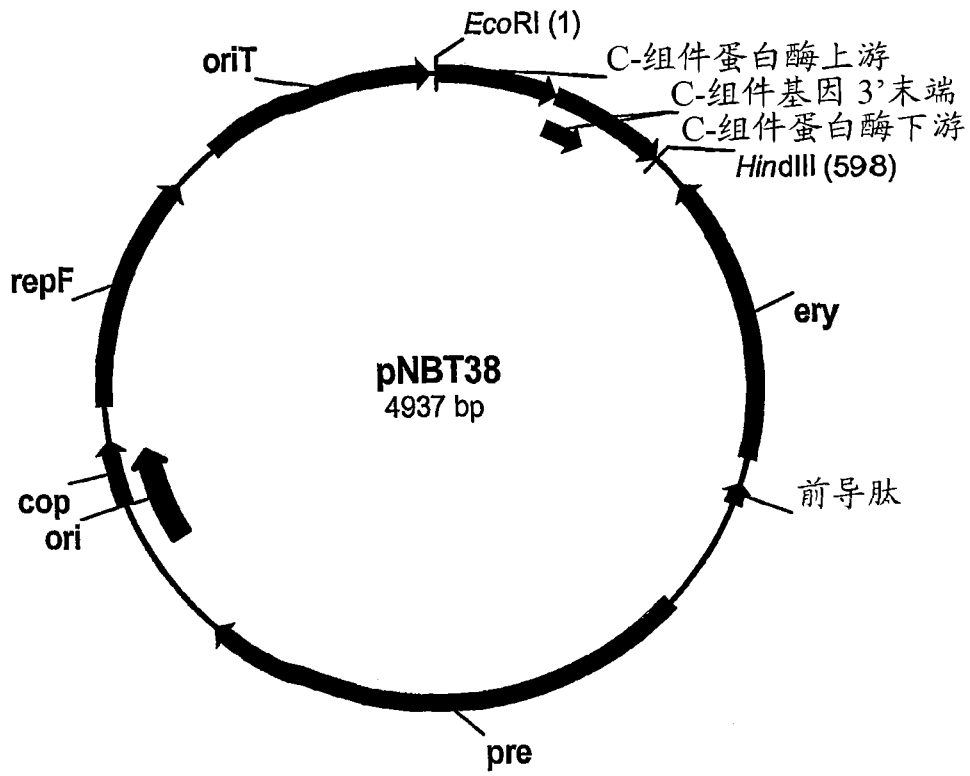


图 12

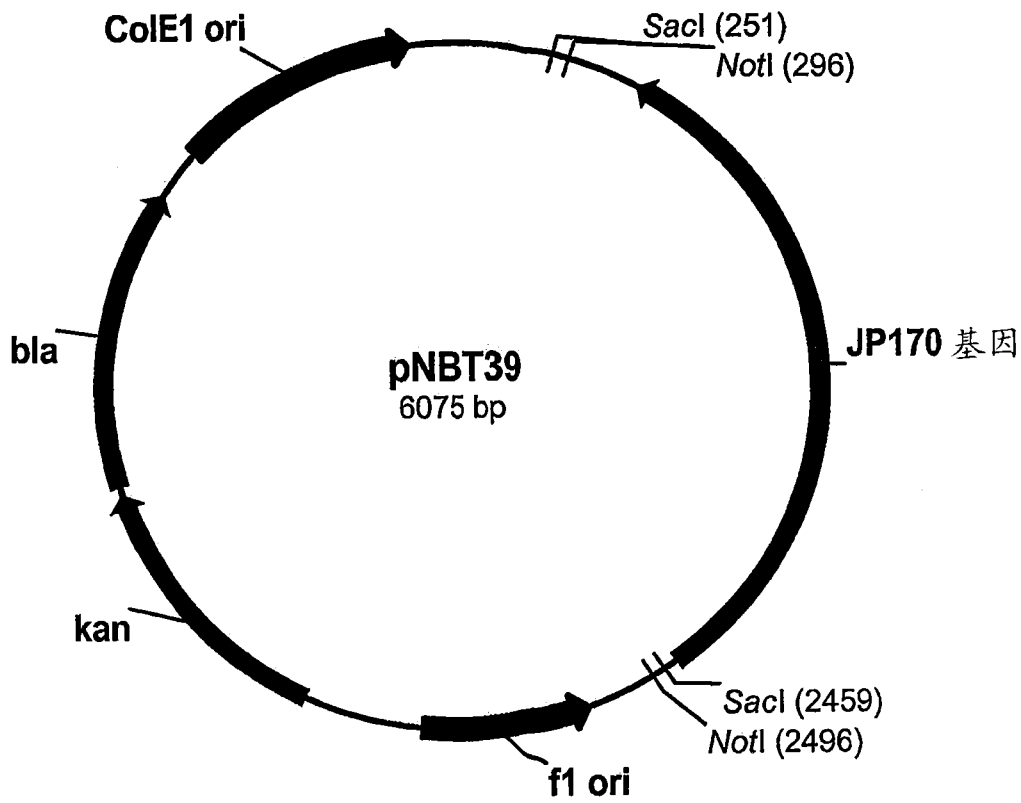


图 13

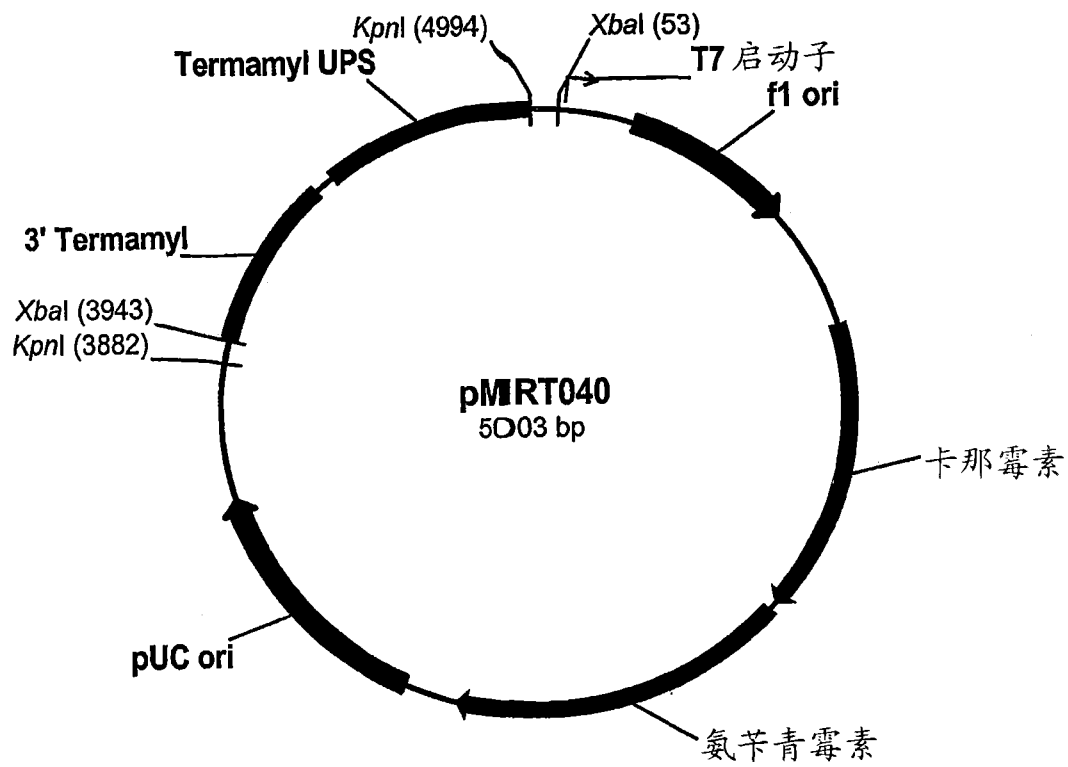


图 14

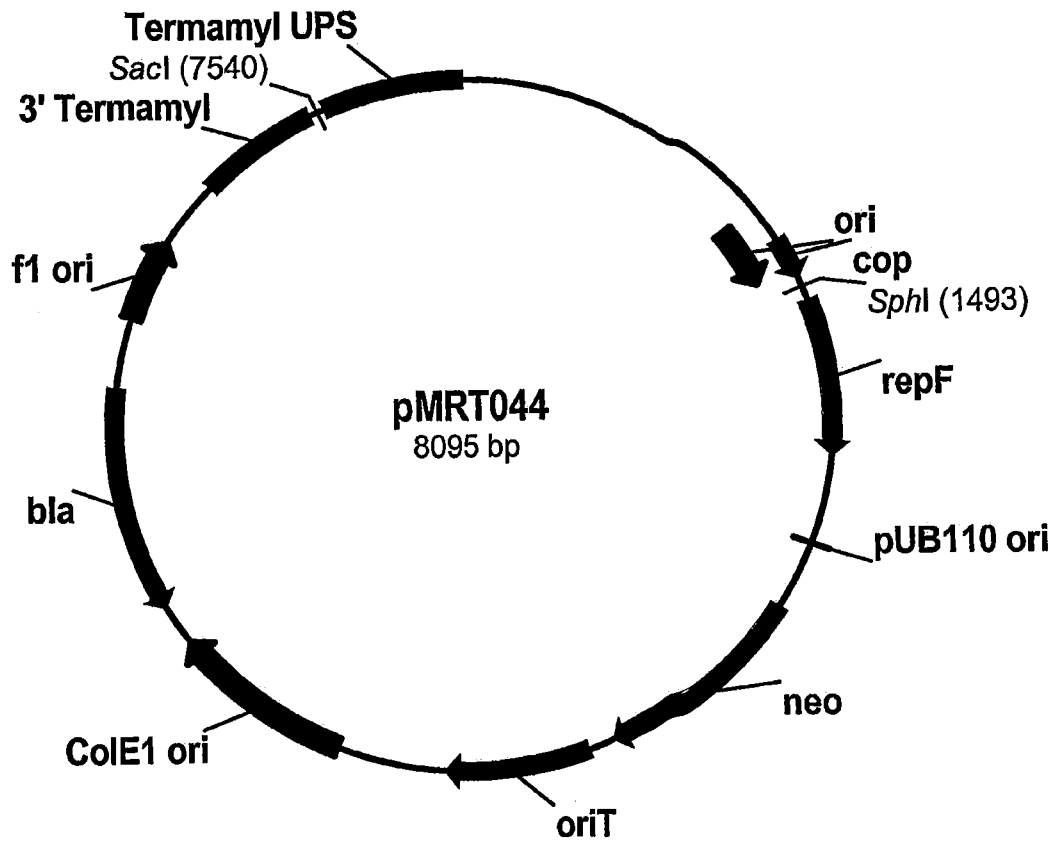


图 15

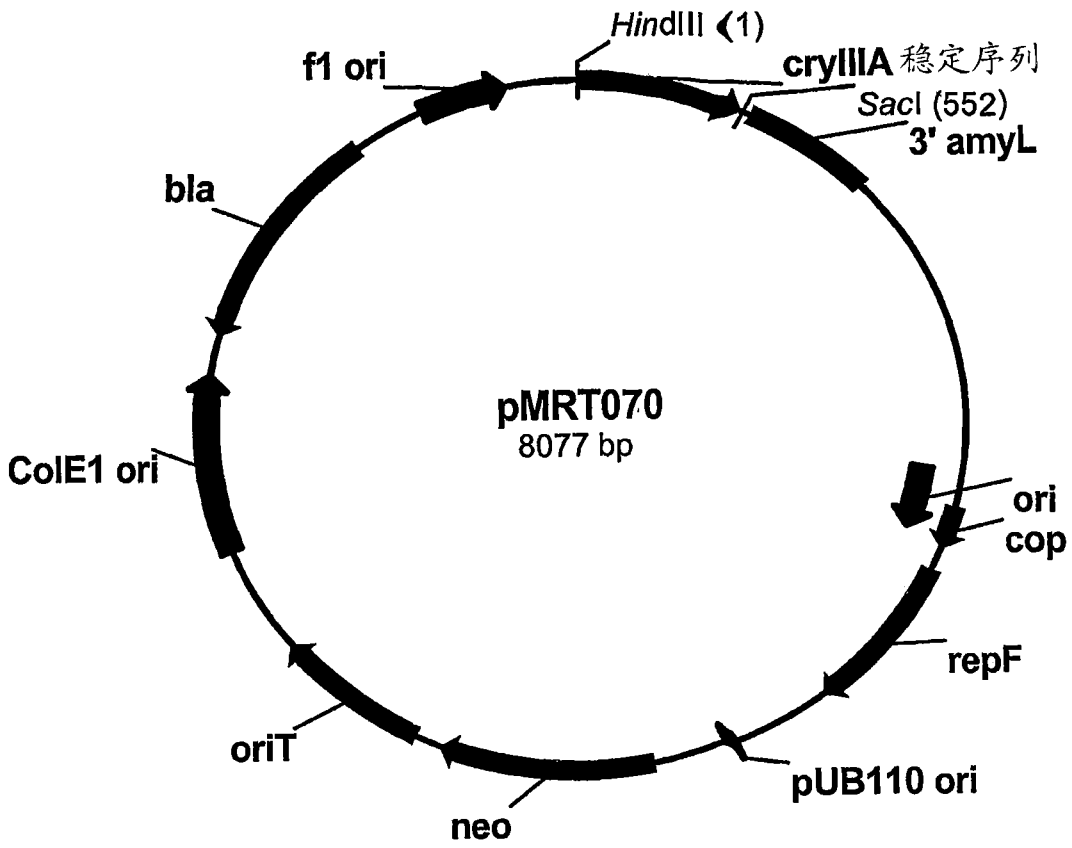


图 16

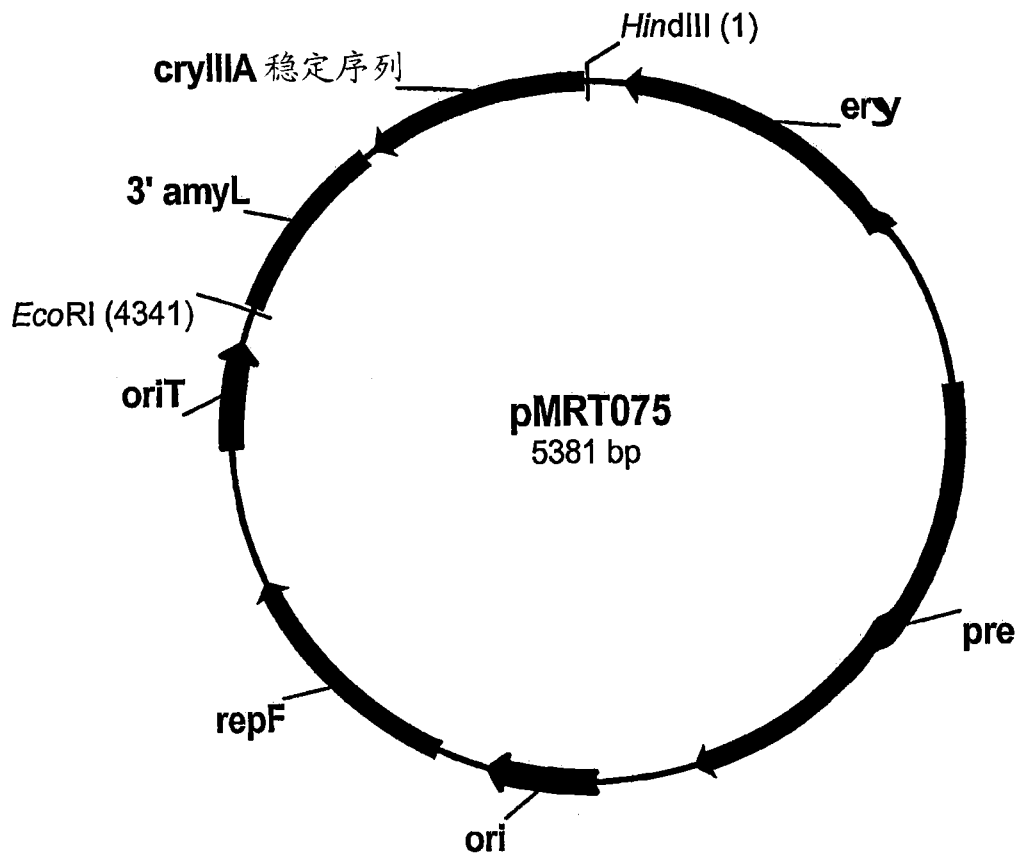


图 17

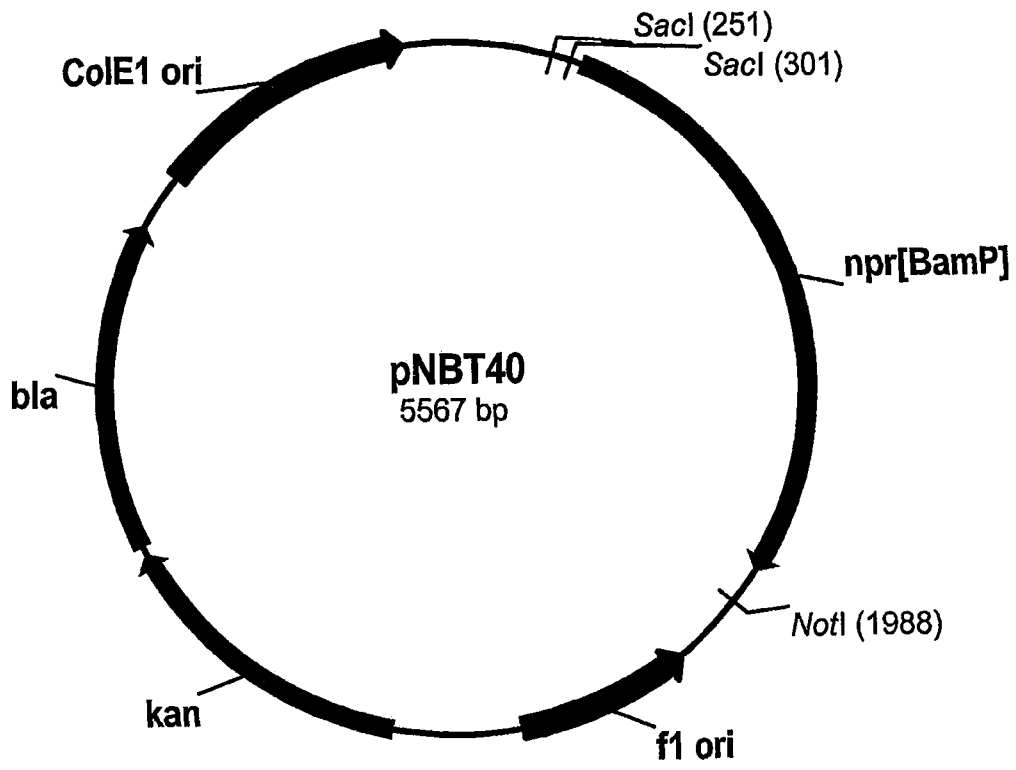


图 18

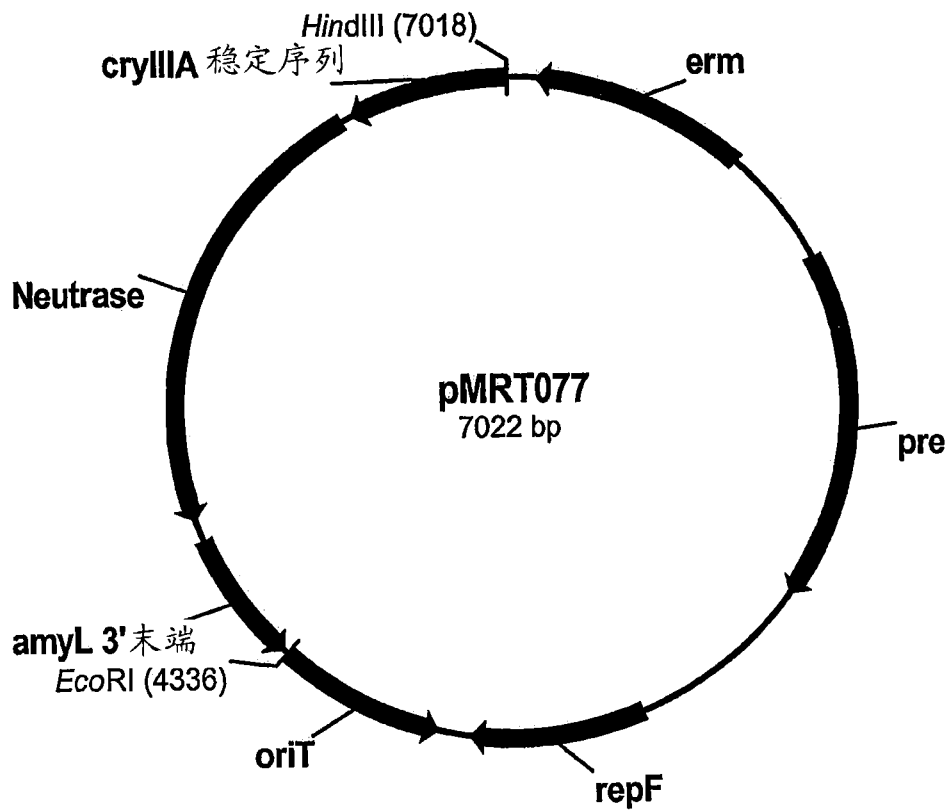


图 19

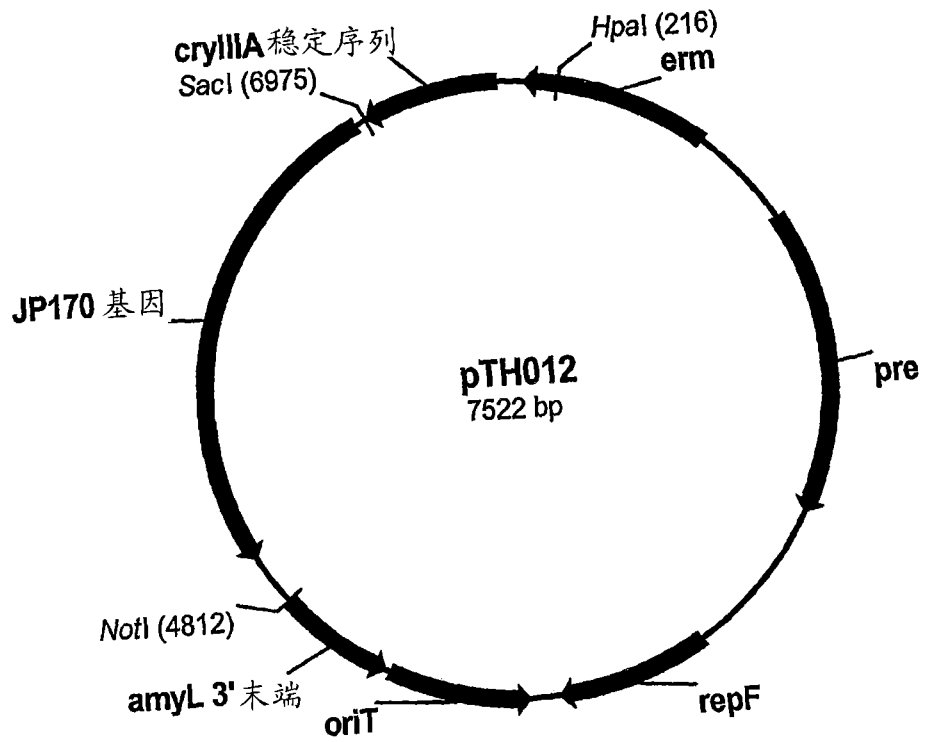


图 20

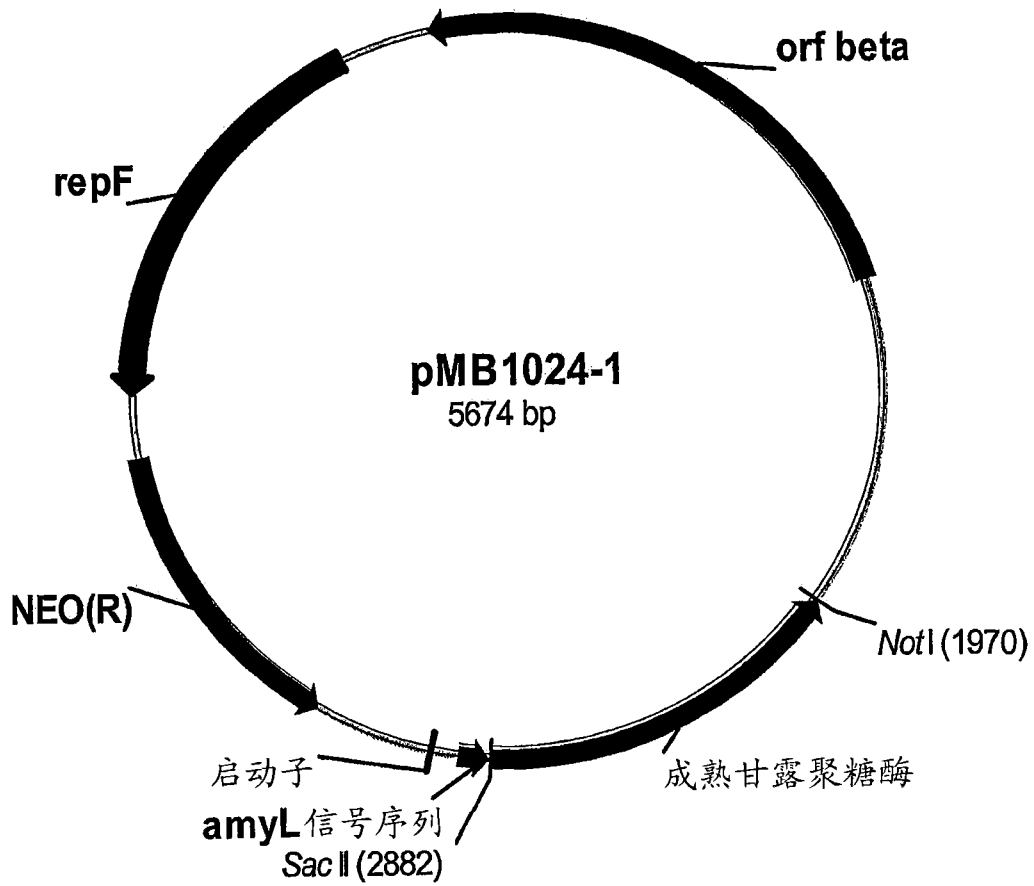


图 21

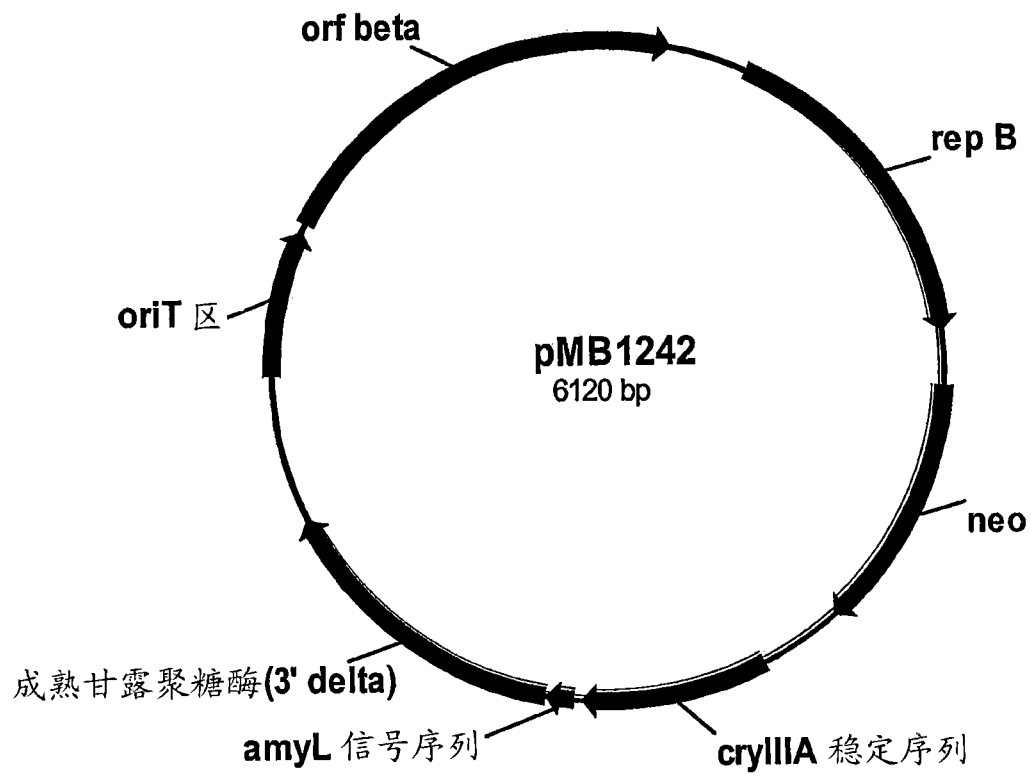


图 22

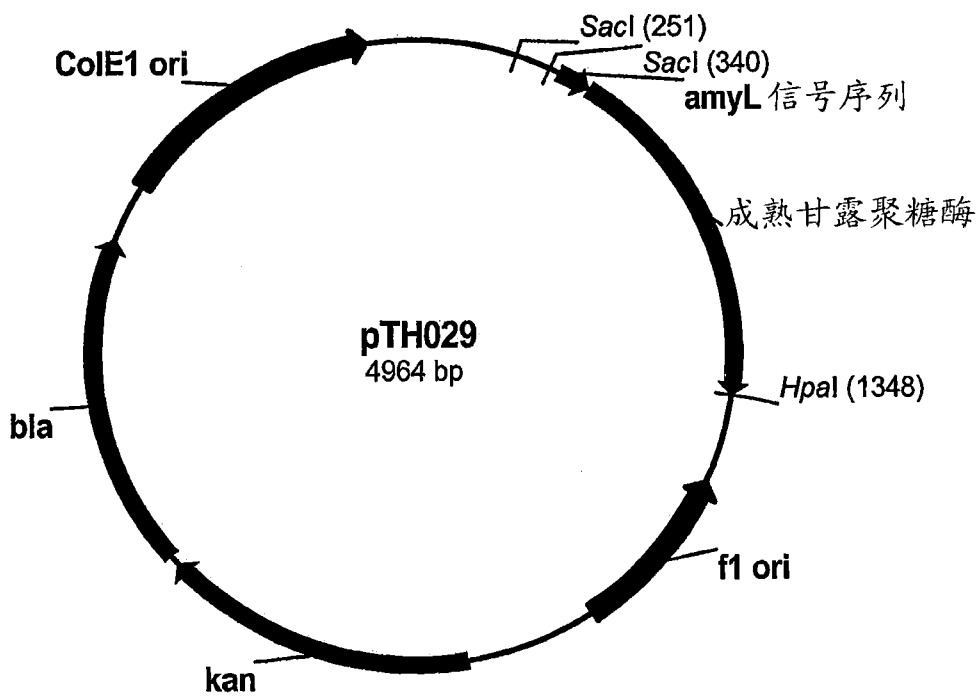


图 23

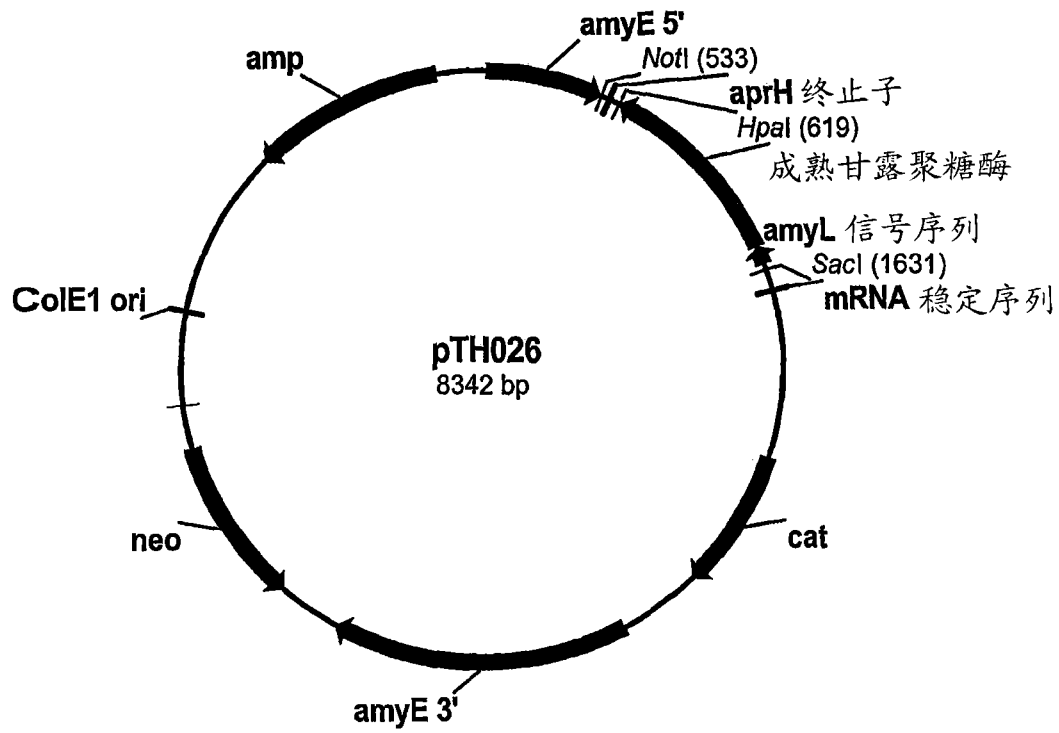


图 24

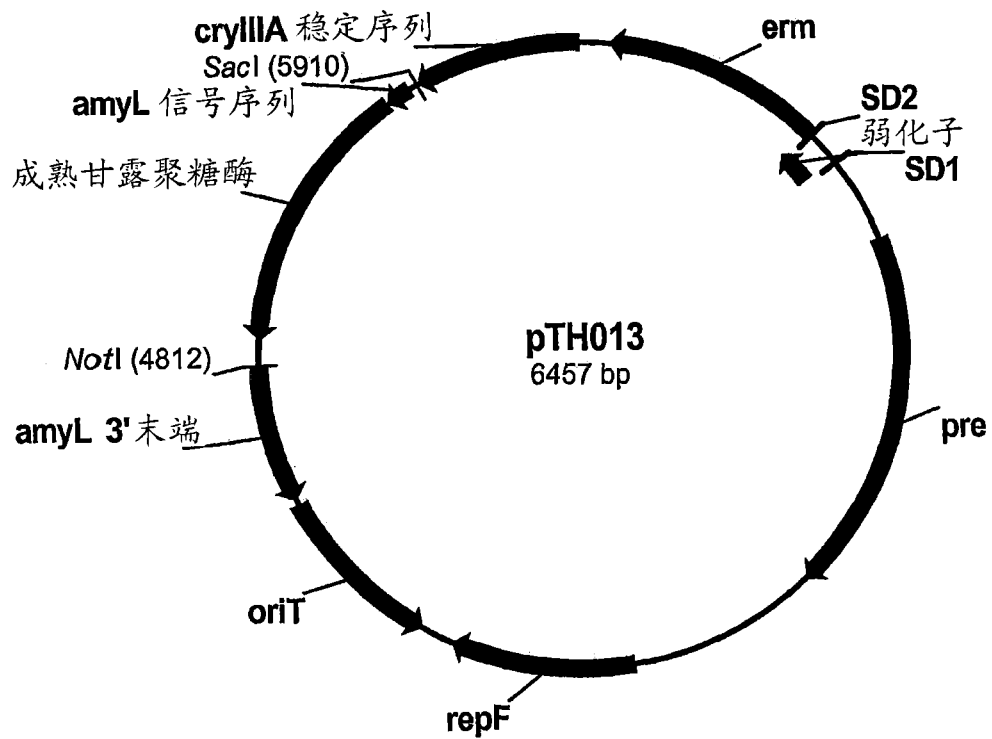


图 25

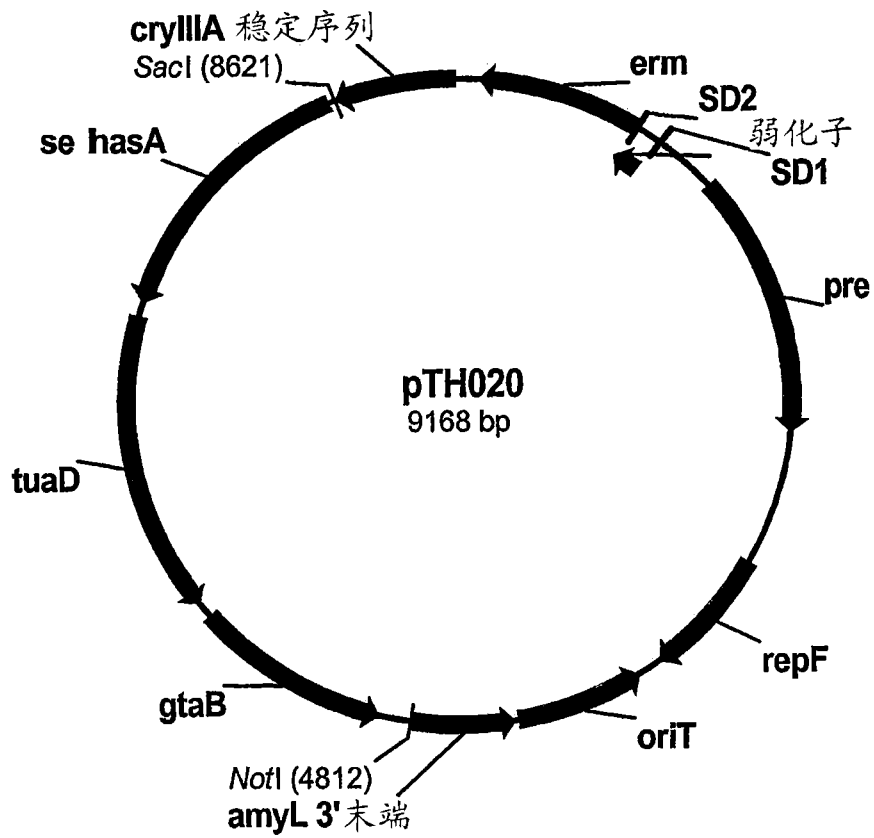


图 26

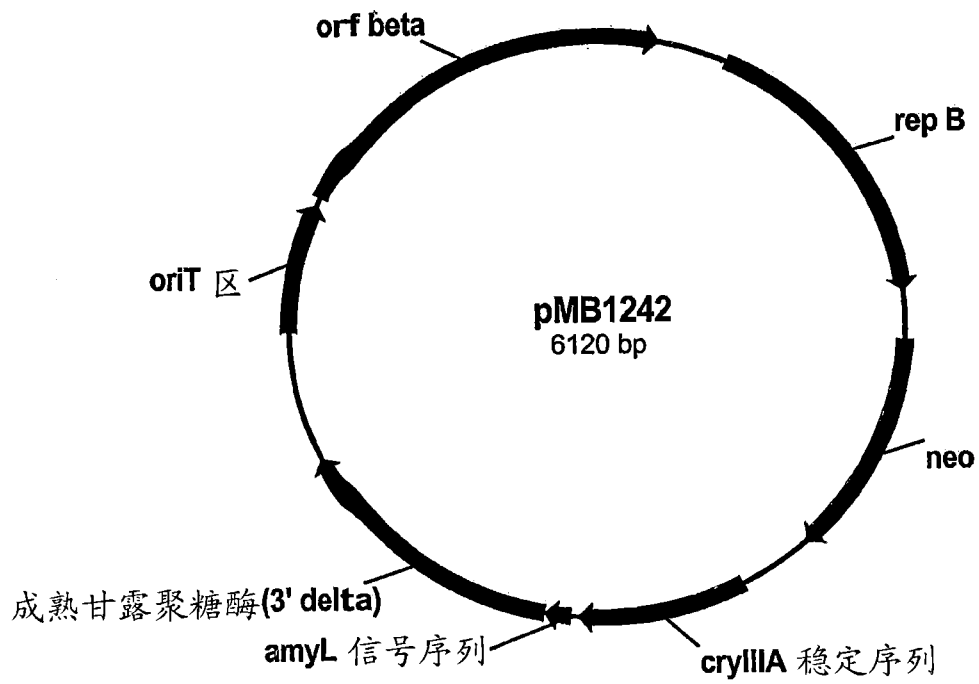


图 27