

[19] 中华人民共和国国家知识产权局



[12] 发明专利说明书

专利号 ZL 98806054. X

[51] Int. Cl.

C12N 15/57 (2006.01)

C12N 15/75 (2006.01)

C12N 9/54 (2006.01)

C07K 14/00 (2006.01)

[45] 授权公告日 2006年7月12日

[11] 授权公告号 CN 1263859C

[22] 申请日 1998.6.9 [21] 申请号 98806054. X

[30] 优先权

[32] 1997. 6. 12 [33] US [31] 08/873,479

[86] 国际申请 PCT/US1998/012005 1998.6.9

[87] 国际公布 WO98/056927 英 1998.12.17

[85] 进入国家阶段日期 1999.12.10

[71] 专利权人 诺沃奇梅兹有限公司

地址 美国加利福尼亚州

[72] 发明人 A·斯洛玛 L·克里斯坦森

审查员 陈中伟

[74] 专利代理机构 中国国际贸易促进委员会专利
商标事务所

代理人 黄革生

权利要求书2页 说明书71页 附图12页

[54] 发明名称

具有蛋白酶活性的多肽的编码核酸

[57] 摘要

本发明涉及得自芽孢杆菌、编码有蛋白酶活性的多肽的分离核酸序列。本发明还涉及包含所述核酸序列的核酸构建体、载体和宿主细胞以及所述多肽的制备方法。

1. 编码有蛋白酶活性的多肽的分离核酸序列，其中所述多肽包含 SEQ ID NO: 42 所示氨基酸序列或者 SEQ ID NO: 42 中第 209 - 641 位氨基酸所示氨基酸序列。

2. 权利要求 1 的核酸序列，其编码包含 SEQ ID NO: 42 中第 209 - 641 位氨基酸的多肽。

3. 权利要求 2 的核酸序列，其编码由 SEQ ID NO: 42 中第 209 - 641 位氨基酸所组成的多肽。

4. 权利要求 1 的核酸序列，其编码包含 SEQ ID NO: 42 氨基酸序列的多肽。

5. 权利要求 4 的核酸序列，其编码由 SEQ ID NO: 42 氨基酸序列组成的多肽。

6. 权利要求 1 的核酸序列，其中该核酸序列得自芽孢杆菌属 (*Bacillus*) 菌株。

7. 权利要求 1 的核酸序列，其中该核酸序列得自嗜碱芽孢杆菌。

8. 权利要求 1 的核酸序列，其中该核酸序列得自嗜碱芽孢杆菌 NCIB 12513。

9. 权利要求 1 的核酸序列，其中包含图 5 所示的第 1471-2769 位核苷酸。

10. 权利要求 1 的核酸序列，其中含有枯草芽孢杆菌 NRRL B-21680 中所含质粒 p170BAN 上的蛋白酶编码核酸序列。

11. 含有权利要求 1 的核酸序列与一或多个控制序列可操作连接在一起的核酸构建体，该控制序列指导权利要求 1 所述核酸序列所编码的多肽在表达宿主中的产生。

12. 含有权利要求 11 的核酸构建体的重组表达载体。

13. 含有权利要求 11 的核酸构建体的一或多个拷贝的重组宿主细胞。

14. 权利要求 13 的重组宿主细胞，它是细菌细胞。

15. 权利要求 14 的重组宿主细胞，其中所述细菌细胞是芽孢杆菌 (*Bacillus*)、链霉菌 (*Streptomyces*) 或假单胞菌 (*Pseudomonas*) 细胞。

16. 权利要求 15 的重组宿主细胞，其中所述芽孢杆菌细胞为嗜碱芽孢杆菌 (*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*)，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 megaterium*)，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*) 菌株。

17. 权利要求 1 所述核酸序列所编码的具有蛋白酶活性的多肽的制备方法，包括(a)培养权利要求 13 的重组宿主细胞以生产多肽；和 (b) 回收多肽。

18. 产生亲本细胞的突变体的方法，包括破坏或删除权利要求 1 的核酸序列，得到生产权利要求 1 所述核酸序列所编码的多肽比亲本细胞少的突变细胞。

19. 由权利要求 18 的方法获得的突变细胞。

20. 权利要求 19 的突变细胞，其中进一步含有编码异源蛋白的核酸序列的一或多个拷贝。

具有蛋白酶活性的多肽的编码核酸

相关申请的相互参考

本申请是 1997 年 6 月 12 日提交的待审美国申请系列号 08/873, 479 的继续申请, 该申请全文引入本文作为参考。

技术领域

本发明涉及编码有蛋白酶活性的多肽的分离核酸序列。本发明还涉及包含所述核酸序列的核酸构建体、载体和宿主细胞, 以及所述多肽的重组制备方法。

背景技术

已知与蛋白水解酶一起配制的洗涤剂具有除去污渍的改善性能。例如, SAVINASE™(Novo Nordisk A/S, Bagsvaerd, Denmark), 得自迟缓芽孢杆菌的一种微生物蛋白酶, 现已用于许多商业品牌的洗涤剂中。

WO88/01293 中公开了得自一种嗜碱性芽孢杆菌、对过氧化物型漂白剂有更强稳定性的蛋白酶。

JP1497182 中公开了编码一种芽孢杆菌碱性蛋白酶 Y 的 DNA 序列, 所述蛋白酶据称有较好的碱性和表面活性剂耐受性, 并且去污力有所改善。

许多洗涤剂在溶液中是碱性的(如 pH10 左右)。需要有在高 pH 下有高活性、对漂白剂比较稳定的新型蛋白水解酶。公开于 WO88/01293 中的蛋白酶类型具有这些特性, 故用于洗涤剂组合物中非常理想。然而, 迄今尚无重组生产这些酶的方法。

本发明的一个目的是重组生产这些有价值的酶。

发明内容

本发明涉及编码有蛋白酶活性的多肽的分离核酸序列, 它们选

自:

(a)编码氨基酸序列与 SEQ ID NO: 43 氨基酸序列至少 95%相同的多肽的核酸序列;

(b)编码氨基酸序列与 SEQ ID NO: 42 氨基酸序列至少 85%相同的多肽的核酸序列;

(c)与 SEQ ID NO: 41 核酸序列的成熟多肽编码区至少 95%同源的核酸序列;

(d)(a)、(b)或(c)的等位变体; 和

(e)(a)、(b)、(c)或(d)的亚序列, 其中所述亚序列编码有蛋白酶活性的多肽片段。

本发明还涉及包含该核酸序列的核酸构建体、载体和宿主细胞以及制备所述多肽的重组方法。

附图说明

图 1 显示 pShv2 的限制图谱。

图 2 显示 pSJ1678 的限制图谱。

图 3 显示 pSJ2882-MCS 的限制图谱。

图 4 显示 pPL1759 的限制图谱。

图 5A 和 5B 显示芽孢杆菌 JP170(NCIB12513)蛋白酶基因的核酸序列和推导出的氨基酸序列。

图 6A 和 6B 显示芽孢杆菌 JP170(NCIB12513)蛋白酶基因推导出的氨基酸序列与其它蛋白酶推导出的氨基酸序列之间的比较。

图 7 显示 pPL2419 的限制图谱。

图 8 显示 pCAsub2 的限制图谱。

图 9 显示芽孢杆菌 JP170 的蛋白酶与 SAVINASE™ 的模型洗涤剂在除去棉上草类杂质的洗涤效果比较。

图 10 显示芽孢杆菌 JP170 的蛋白酶与 SAVINASE™ 的 Koso Top 洗涤剂在除去棉上草类杂质的洗涤效果比较。

具体实施方式

编码有蛋白酶活性的多肽的分离核酸序列

本文所用的术语“分离的核酸序列”表示基本上没有其他核酸序列的核酸序列，例如通过琼脂糖电泳确定为至少大约 20% 纯，优选至少大约 40% 纯，更优选大约 60% 纯，还要优选的是至少大约 80% 纯，最优选至少大约 90% 纯。例如，通过基因工程中用来将核酸序列从其天然位置移置到其它位置(在此它将复制)的标准克隆操作，能够获得分离的核酸序列。克隆操作可能包括切割并分离出含有多肽编码核酸序列的所需核酸片段，将该片段插入载体分子中，再将重组载体导入宿主细胞内，在其中将复制出该核酸序列的多个拷贝或克隆。核酸序列可以是基因组的、cDNA、RNA、半合成的、合成的序列，或者是其混合形式。

在第二个实施方案中，本发明涉及编码有蛋白酶活性的多肽的分离核酸序列，所述多肽的氨基酸序列与 SEQ ID NO: 43 氨基酸序列之间的相同程度至少大约 95%、优选至少大约 97% (该多肽以下称为“同源多肽”)。

在第三个实施方案中，本发明涉及编码有蛋白酶活性(优选在翻译后加工之后)的多肽的分离核酸序列，所述多肽含有的氨基酸序列与 SEQ ID NO: 42 氨基酸序列之间的相同程度至少大约 85%、优选至少大约 90%、更优选至少大约 95%、最优选至少大约 97% (该多肽以下称为“同源多肽”)。

在一个优选实施方案中，同源多肽具有的氨基酸序列与 SEQ ID NO: 43 氨基酸序列有 5 个氨基酸不同，优选 4 个氨基酸不同，更优选 3 个氨基酸不同，更优选 2 个氨基酸不同，最优选 1 个氨基酸不同。为了本发明的目的，两个氨基酸序列之间的相同程度是通过 Clustal 方法 (Higgins, 1989, CABIOS5: 151-153)，用相同性表，缺口罚分和缺口长度罚分均为 10 确定的。

优选地，本发明核酸序列编码含有 SEQ ID NO: 42 或 SEQ ID NO: 43 氨基酸序列的多肽或者其等位变体。在一个更优选的实施方案中，本发明核酸序列编码含有 SEQ ID NO: 42 或 SEQ ID NO: 43 氨基酸序列的多肽。在另一个优选实施方案中，本发明核酸序列编

码含有 SEQ ID NO: 42 或 SEQ ID NO: 43 氨基酸序列的多肽或者其有蛋白酶活性的片段。在最优选的实施方案中, 该核酸序列编码含有 SEQ ID NO: 42 或 SEQ ID NO: 43 氨基酸序列的多肽。本发明还包括编码具有 SEQ ID NO: 42 或 SEQ ID NO: 43 氨基酸序列的多肽的核酸序列, 所述核酸序列因遗传密码的简并性而与 SEQ ID NO: 41 不同。本发明还涉及 SEQ ID NO: 41 的亚序列, 其编码有蛋白酶活性的 SEQ ID NO: 42 或 SEQ ID NO: 43 片段。

SEQ ID NO: 41 的亚序列是包括在 SEQ ID NO: 41 中的一段核酸序列, 不同之处在于从 5' 和/或 3' 末端删除了一个或多个核苷酸。优选亚序列含有至少 1029 个核苷酸, 更优选含有至少 1119 个核苷酸, 最优选含有至少 1209 个核苷酸。SEQ ID NO: 42 或 SEQ ID NO: 43 的片段是从该氨基酸序列的氨基和/或羧基末端删除了一个或多个氨基酸的多肽。优选地, 片段含有至少 343 个氨基酸残基, 更优选至少 373 个氨基酸残基, 最优选有至少 403 个氨基酸残基。

等位变体表示一个基因占据相同染色体位点的两种或多种可替换形式中任意一种。等位变体是通过突变自然产生的, 并可能导致种群内的表型多态性。基因突变可以是沉默的(编码的多肽没有变化)或者可以编码氨基酸序列改变的多肽。术语多肽的等位变体表示基因的等位变体所编码的多肽。

同源多肽的氨基酸序列与 SEQ ID NO: 42 或 SEQ ID NO: 43 的氨基酸序列不同之处可能在于插入或缺失了 1 或多个氨基酸残基和/或用不同的氨基酸残基取代了 1 或多个氨基酸残基。优选地, 氨基酸改变是小的变化, 即不会显著影响蛋白质的折叠和/或活性的保守性氨基酸取代; 小片段缺失, 通常是 1 到大约 30 个氨基酸的缺失; 小的氨基或羧基末端延伸, 如氨基末端甲硫氨酸残基; 有多达大约 20-25 个残基的小连接肽; 或者能通过改变净电荷而有助于纯化或者有其它功能的小延伸, 如多聚组氨酸串、抗原性表位或结合结构域。

保守性取代的例子是在碱性氨基酸(如精氨酸、赖氨酸和组氨酸)、酸性氨基酸(如谷氨酸和天冬氨酸)、极性氨基酸(如谷氨

酰胺和天冬酰胺)、疏水性氨基酸(如亮氨酸、异亮氨酸和缬氨酸)、芳香族氨基酸(如苯丙氨酸、色氨酸和酪氨酸)和小氨基酸(如丙氨酸、甘氨酸、丝氨酸、苏氨酸和甲硫氨酸)各组内进行的取代。通常不会改变特异性作用的氨基酸取代是本领域已知的,例如 H.Neurath 和 R.L.Hill(1979)等(《蛋白质》, Academic Press, New York)描述过。最常见的替换是 Ala/Ser, Val/Ile, Asp/Glu, Thr/Ser, Ala/Gly, Ala/Thr, Ser/Asn, Ala/Val, Ser/Gly, Tyr/Phe, Ala/Pro, Lys/Arg, Asp/Asn, Leu/Ile, Leu/Val, Ala/Glu 和 Asp/Gly 以及反向进行的替换。

在第三个实施方案中,本发明涉及编码有蛋白酶活性的多肽、与 SEQ ID NO: 41 的成熟多肽编码序列之间的同源性程度至少大约 95%、优选至少大约 97% 的分离核酸序列;或者 SEQ ID No:41 的等位变体或其编码有蛋白酶活性的多肽片段的亚序列。为了本发明的目的,通过 Clustal 方法(Higgins,1989, CABIOS5: 151-153)用相同性表、缺口罚分和缺口长度罚分均为 10 来确定两个核酸序列之间的同源性程度。

在第四个实施方案中,本发明涉及编码有蛋白酶活性的多肽的分离核酸序列,所述核酸序列在低度严谨条件下,更优选在中度严谨条件下,最优选在高度严谨条件下,与在相同的条件下和 SEQ ID NO: 41 核酸序列或其互补链能杂交上的寡核苷酸探针能够杂交上(J.Sambrook, E.F.Fritsch,和 T.Maniatus,分子克隆:实验指南,第2版, Cold Spring Harbor,New York);或者 SEQ ID NO: 41 的等位变体,或其编码有蛋白酶活性的多肽片段的亚序列。

可以利用 SEQ ID NO: 41 的核酸序列或其亚序列, SEQ ID NO: 42 或 SEQ ID NO: 43 的氨基酸序列或其部分序列来设计寡核苷酸探针,以根据本领域熟知的方法从不同属或种的菌株中鉴定和克隆到编码有蛋白酶活性的多肽的 DNA。具体来说,可以按照标准 Southern 印迹操作,用这些探针与所研究的属或种的基因组或 cDNA 杂交,以便鉴定和分离其中的相应基因。这些探针可以比完整序列短得

多,但应至少有 15 个、优选至少 25 个、更优选至少 40 个核苷酸长。也可以使用更长的探针。DNA 和 RNA 探针都可使用。为了检测相应基因,通常要将探针进行标记(例如,用 ^{32}P 、 ^3H 、 ^{35}S 、生物素或抗生物素蛋白标记)。

因此,可以从得自其他生物的基因组、cDNA 或组合化学文库筛选能与上述探针杂交、并编码有蛋白酶活性的多肽的 DNA。可以通过琼脂糖或聚丙烯酰胺凝胶电泳,或者其他分离技术分离来自其他生物的基因组 DNA 或其他 DNA。可以将来自文库的 DNA 或分离的 DNA 转移并固定到硝酸纤维素膜或其他合适的载体材料上。为了鉴定出与 SEQ ID NO: 41 同源的克隆或 DNA,在 Southern 印迹中使用了载体材料。杂交是指按照标准 Southern 印迹操作进行实验,核酸序列与对应于 SEQ ID NO: 41 所示核酸序列中多肽编码部分的寡核苷酸探针在低度到高度严谨条件下[即,在 $5\times\text{SSPE}$ 、 $0.3\%\text{SDS}$ 、 $200\mu\text{g/ml}$ 剪切变性的鲑精 DNA、甲酰胺(对低度、中度和高度严谨条件分别为 25、35 或 50%)中于 42°C 预杂交和杂交]能发生杂交。最终用 $2\times\text{SSC}$ 、 $0.2\%\text{SDS}$ 将该载体材料洗 3 次,每次 30 分钟,洗涤温度优选至少 50°C (极低严谨度),更优选至少 55°C (低严谨度),更优选至少 60°C (中等严谨度),更优选至少 65°C (中高严谨度),还要优选的是至少 70°C (高严谨度),最优选的是至少 75°C (极高严谨度)。用 X-光胶片检测出在这些条件下与寡核苷酸探针杂交的分子。

本发明的核酸序列可以得自任何属的微生物。就本发明的目的而言,文中与特定来源连在一起使用的术语“得自”意味着所述核酸序列编码的多肽由该来源产生或由插入了来自该来源的核酸序列的细胞产生。

所述核酸序列可以来源于细菌。例如,可以从革兰氏阳性细菌,如芽孢杆菌或链霉菌中得到这些核酸序列,例如浅青紫链霉菌或鼠灰链霉菌;或者可以从革兰氏阴性细菌,例如大肠杆菌或假单胞菌中得到。

在一个优选实施方案中,本发明核酸序列得自芽孢杆菌属的菌

株，该属的定义参见 Fergus G. Priest 在 Abraham L. Sonenshein, James A. Hoch, 和 Richard Losick 所编“枯草芽孢杆菌和其它革兰氏阳性细菌”，美国微生物学会，华盛顿特区，1993，第 3-16 页的定义。

在一个优选实施方案中，该核酸序列得自嗜碱芽孢杆菌、解淀粉芽孢杆菌、短芽孢杆菌、环状芽孢杆菌、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芽孢杆菌菌株。

在一个最优选的实施方案中，该核酸序列得自芽孢杆菌 NCIB12513，如 SEQ ID NO:41 所示的核酸序列。在另一个最优选的实施方案中，该核酸序列是枯草芽孢杆菌 LC20 NRRL B-21680 中所含质粒 p170BAN 中包含的序列。

公众可很容易从许多培养物保藏中心（如美国典型培养物保藏中心 ATCC, 德意志微生物保藏中心(DSM), 真菌菌种保藏中心(CBS) 和北方地区研究中心农业研究机构保藏中心(NRRL)）获得这些种的菌株。

另外，用前面提到的探针可以从其他来源，包括从自然界（如土壤、沉积物、水体等）分离到的微生物中鉴定和获得这些核酸序列。用于从其自然生存环境分离微生物的技术是本领域已知的。类似地，通过筛选另一种微生物的基因组或 cDNA 文库可以得到所述核酸序列。一旦用探针检测到了编码多肽的核酸序列，就可以采用本领域技术人员已知的技术（参见，例如 Sambrook 等，1989，出处同前）对序列进行分离或克隆。

用于分离或克隆编码多肽的核酸序列的技术是本领域已知的，包括从基因组 DNA 中分离、由 cDNA 制备或将它们组合运用。为了从基因组 DNA 克隆本发明所述核酸序列，可以使用已知的聚合酶链式反应（PCR）或用抗体筛选表达文库来检测有共同结构特征的克隆 DNA 片段。参见例如，Innis 等，1990，PCR: 方法和应用指南，

Academic Press, New York. 还可以使用其他核酸扩增操作, 如连接酶链反应 (LCR)、连接活化转录 (LAT) 和基于核酸序列的扩增方法 (NASBA)。可以从芽孢杆菌属的一个或另一个菌株或者相关微生物中克隆得到所述核酸序列, 因此, 该核酸序列例如可以是所述核酸序列的多肽编码区之等位基因或种变体。

合成与所述多肽基本相似的多肽时, 可能有必要对本发明的核酸序列进行修饰。术语与所述多肽“基本相似”是指非天然形式的多肽。这些多肽可能与从天然来源分离到的多肽在某些设计方式上不同。例如, 可能希望用例如定点诱变来合成多肽的变体, 这些变体有不同的特异活性、热稳定性或最佳 pH 等。可以在 SEQ ID NO: 41 中多肽编码部分所示核酸序列的基础上构建出类似序列, 例如其亚序列; 并/或通过引入这样的核苷酸取代而构建, 这些取代不会使产生的氨基酸序列与原核酸序列编码的多肽不同, 但它们符合制备酶所用宿主生物对密码子的使用习惯; 或通过引入会产生不同氨基酸序列的核苷酸取代而构建。关于核苷酸取代的综述, 参见例如 Ford 等, 蛋白质表达和纯化 (Protein Expression and Purification) 2: 95-107。

对本领域技术人员显而易见的是, 可以在分子功能关键区以外作这些取代, 而仍可以得到活性多肽。可以依照本领域的已知操作, 例如定点诱变或丙氨酸扫描诱变 (参见例如, Cunningham 和 Wells, 1989, 科学 (Science) 244: 1081-1085), 鉴定出那些为本发明所述分离核酸序列编码的多肽的活性所必需、因此最好不进行取代的氨基酸残基。后一种技术中, 向分子中每个带正电荷的残基导入突变, 检测所得突变分子的蛋白酶活性从而鉴定出对分子活性比较关键的氨基酸残基。通过分析例如用核磁共振、晶体学或光亲和标记确定的立体结构, 也可以确定底物-酶相互作用的位点 (参见例如, de Vos 等, 1992, 科学 255: 306-312; Smith 等, 1992, 分子生物学杂志 (Journal of Molecular Biology) 224: 899-904; Wlodaver 等, 1992, FEBS 快报 (FEBS Letters) 309: 59-64)。

本发明核酸序列还可能编码融合多肽或可裂解的融合多肽，这些融合多肽是在所述多肽或其片段的N-或C-末端融合上了另一种多肽。融合多肽是通过将编码另一多肽的核酸序列（或者其一部分）与本发明的核酸序列（或者其一部分）融合在一起制备的。用于制备融合多肽的技术是本领域已知的，包括读框一致地连接诸多肽编码序列，并且使融合多肽的表达受到相同的启动子和终止子的调控。

核酸构建体

本发明还涉及包含本发明所述核酸序列的核酸构建体，在这些核酸构建体中，核酸序列可操纵地连接了1或多个调控序列，该调控序列在与其相容的条件下能指导编码序列在合适的宿主细胞中进行表达。表达应理解为包括生产有蛋白酶活性的多肽中所涉及的任何步骤，包括，但不限于转录、转录后修饰、翻译、翻译后修饰和分泌。

“核酸构建体”在文中定义为单链或双链核酸分子，它们分离自天然基因，或者已被修饰而含有以非天然方式组合和排列的核酸片段的基因。当核酸构建体包含本发明所述编码序列表达必需的所有调控序列时，术语核酸构建体与表达盒同义。术语“编码序列”在文中定义为能转录为mRNA并翻译成多肽的序列。编码序列的范围通常是由位于mRNA 5'端开放读码框上游的核糖体结合位点和位于mRNA 3'端开放读码框下游的转录终止序列确定的。编码序列可以包括，但不限于DNA、cDNA和重组核酸序列。

可以以多种方式操作编码多肽的分离核酸序列，使其表达有蛋白酶活性的多肽。在将核酸序列插入载体中之前对其进行加工可能比较理想，或者是必需的，这取决于具体的表达载体。利用克隆方法修饰核酸序列的技术为本领域所熟知。

本文中术语“控制序列”定义为包括多肽的表达所必需或有利的的所有组分。每个调控序列对于编码多肽的核酸序列可以是天然含有的或外来的。这些调控序列包括，但不限于，前导序列、前肽序

列、启动子、信号序列和转录终止子。最低限度，调控序列要包括启动子以及转录和翻译终止信号。为了导入特定的限制位点以便于将调控序列与编码多肽的核酸序列的编码区连接起来，可以提供带有接头的调控序列。术语“可操纵地连接”在文中定义为这样一种结构，其中调控序列相对于核酸序列的编码序列适当地位于能指导产生多肽的位置上。

调控序列可以是合适的启动子序列，即用来表达核酸序列的宿主细胞能识别的核酸序列。启动子序列含有介导多肽表达的转录调控序列。启动子可以是在所选用的宿主细胞中有转录活性的任何核酸序列，包括突变的、截短的和杂合的启动子，可以得自编码对于宿主细胞是内源或外源、胞外或胞内多肽的基因。

指导本发明所述核酸构建体转录（特别是在细菌宿主细胞中）的合适启动子的例子，是得自以下来源的启动子：大肠杆菌 lac 操纵子、天蓝色链霉菌琼脂糖酶基因（dagA）、枯草芽孢杆菌果聚糖酶基因(sacB)、地衣芽孢杆菌 α -淀粉酶基因(amyL)、嗜热脂肪芽孢杆菌麦芽糖源淀粉酶基因(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，原核细胞的 β -内酰胺酶基因（Villa-Kamaroff 等,1978,美国科学院学报（Proceedings of the National Academy of Sciences USA）75: 3727- 3731），以及 tac 启动子（DeBoer 等，1983，美国科学院学报，80:21-25）。“来自重组细菌的有用蛋白质”（科学美国人,1980，242: 74-94）一文以及 Sambrook 等（1989，出处同前）描述过其他启动子。

调控序列还可以是合适的转录终止序列，即能被宿主细胞识别从而终止转录的一段序列。终止序列可操纵地连接在编码多肽的核酸序列的 3' 末端。在所选宿主细胞中有功能的任何终止子都可以用于本发明。

调控序列还可以是合适的前导序列，即对宿主细胞翻译非常重要的 mRNA 非翻译区。前导序列可操纵地连接在编码多肽的核酸序

列的 5'末端。在所选宿主细胞中有功能的任何前导序列都可以用于本发明。

调控序列还可以是信号肽编码区，该区编码连在多肽氨基端、能引导编码多肽进入细胞分泌途径的一段氨基酸序列。核酸序列编码区的 5'端可能天然含有信号肽编码区，该编码区与分泌多肽的编码区片段翻译读框一致地连接在一起。可选择地，核酸序列编码区的 5'端可含有对编码序列是外来的信号肽编码区。在正常情况下不含有信号肽编码区的编码序列，可能需要添加外来信号肽编码区。可选择地，为了增强多肽分泌，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。信号肽编码区可以得自芽孢杆菌的淀粉酶或蛋白酶基因、或者小牛前凝乳酶原基因。但是，能引导表达后的多肽进入所用宿主细胞的分泌途径的任何信号肽编码区都可以用于本发明。

用于细菌宿主细胞的有效信号肽编码区是得自芽孢杆菌 NCIB11837 的麦芽糖源淀粉酶基因、嗜热脂肪芽孢杆菌的 α -淀粉酶基因、地衣芽孢杆菌的枯草蛋白酶基因、地衣芽孢杆菌的 β -内酰胺酶基因、嗜热脂肪芽孢杆菌的中性蛋白酶基因 (nprT、nprS、nprM) 或枯草芽孢杆菌 prsA 基因的信号肽编码区。Simonen 和 Palva(1993, 微生物学综述 (Microbiological Reviews) 57: 109-137) 描述过其他的信号肽。

调控序列还可以是肽原编码区，该区编码位于多肽氨基末端的一段氨基酸序列。这样产生的多肽已知是酶原或多肽原（或者在某些情况中是酶原）形式。多肽原通常没有活性，可以由多肽原催化性地或自身催化性地切割肽原而转化为成熟的活性多肽。肽原编码区可以得自枯草芽孢杆菌碱性蛋白酶基因 (aprE)、枯草芽孢杆菌中性蛋白酶基因 (nprT)。

在多肽的氨基末端有信号肽和肽原区的情况中，肽原区与多肽的氨基末端相邻，而信号肽区与肽原区的氨基末端相邻。

本发明所述核酸构建体还可以包含 1 或多个核酸序列，这些序

列编码 1 或多个有助于指导多肽表达的因子,例如转录激活因子(例如,反式作用因子)、伴侣蛋白和加工蛋白酶。任何可用于所选宿主细胞的因子都可以用于本发明。编码 1 或多个这类因子的核酸不必与编码多肽的核酸序列串联在一起。

转录激活因子是一种能激活编码多肽的核酸序列进行转录的蛋白质(Kudla等,1990,EMBO Journal 9:1355-1364;Jarai和Buxton,1994,现代遗传学(Current Genetics) 26:2238-2244;Verdier,1990,酵母 6:271-291)。编码转录激活因子的核酸序列可以得自编码嗜热脂肪芽孢杆菌 NprA 的基因(nprA)。

伴侣蛋白是能协助其他多肽进行正确折叠的蛋白质(Hartl等,1994,TIBS19: 20-25;Bergeron等,1994,TIBS19: 124-128;Demolder等,1994,生物技术杂志(Journal of Biotechnology) 32: 179-189;Craig,1993,科学 260: 1902-1903;Gething和Sambrook,1992,自然(Nature) 355: 33-45;Puig和Gilbert,1994,生物学化学杂志 269: 7764-7771;Wang和Tsou,1993,FASEB杂志 7: 1515-11157;Robinson等,1994,生物/技术(Bio/Technology) 1: 381-384;Jacobs等,1993,分子微生物学(Molecular Microbiology) 8: 957-966)。编码伴侣蛋白的核酸序列可以得自编码枯草芽孢杆菌 GroE 蛋白和枯草芽孢杆菌 PrsA 的基因。另外的例子可参见 Gething 和 Sambrook,1992,出处同前和 Hartl 等,1994,出处同前。

加工蛋白酶是能切割肽原从而产生成熟的有生物化学活性的多肽的蛋白酶(Enderlin 和 Ogrydziak,1994,酵母 10: 67-79;Fuller等,1989,美国科学院学报 86: 1434-1438;Julius等,1984,细胞(Cell)37:1075-1089;Julius等,1983,细胞 32: 839-852;美国专利 5702934)。编码加工蛋白酶的核酸序列可以得自编码下述蛋白质的基因:酿酒酵母二肽基氨肽酶、酿酒酵母 Kex2、*Yarrowia lipolytica* 双碱性加工内蛋白酶(xpr6)和尖镰孢金属蛋白酶(p45基因)。

添加能根据宿主细胞的生长情况来调节多肽表达的调控序列可能也是需要的。调控系统的例子是那些能对化学或物理刺激物(包

括在有调控化合物的情况下)作出反应,从而开放或关闭基因表达的系统。原核细胞体系中的调控系统包括 lac、tac 和 trp 操纵子系统。调控序列的其他例子是那些能使基因发生扩增的调控序列。在真核细胞体系中,这类调控序列包括二氢叶酸还原酶基因(该基因在有氨甲蝶呤的情况下发生扩增)和金属硫蛋白基因(重金属可使其扩增)。在这些情况中,应将编码多肽的核酸序列与调控序列可操纵地连接在一起。

表达载体

本发明还涉及包含本发明核酸序列、启动子和转录及翻译终止信号的重组表达载体。可以将上述各种核酸和调控序列连接在一起来制备重组表达载体,该载体可以包括 1 或多个方便的限制位点,以便在这些位点插入或取代编码多肽的核酸序列。可选择地,通过将核酸序列或包含该序列的核酸构建体插入用于表达的合适载体中,可以表达本发明所述核酸序列。为了进行表达并可能分泌,在制备表达载体时,将编码序列放在载体中的一定位置以便编码序列与适当的调控序列可操纵地连接在一起。

重组表达载体可以是便于进行重组 DNA 操作并使核酸序列表达的任何载体(例如质粒或病毒)。选择载体通常取决于载体与它将要导入的宿主细胞的相容性。载体可以是线性或闭环质粒。载体可以是自主复制型载体(即以染色体外个体存在的载体,其复制独立于染色体的复制),例如质粒、染色体外元件、微小染色体或人工染色体。载体可以含有任何保证自我复制的元件。或者,载体可以是一个当导入宿主细胞时,将整合到基因组中并与所整合到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,或者是两个或多个载体或质粒,它们一同含有要导入到宿主细胞基因组中的全部 DNA,或者是转座子。

优选本发明所述载体含有 1 或多个使得容易挑选出转化细胞的选择标记。选择标记是这样一个基因,其产物赋予对杀生物剂或病毒的抗性、对重金属的抗性、相对营养缺陷型的原养型等。细菌选

择标记的例子是来自枯草芽孢杆菌或地衣芽孢杆菌的 *dal* 基因，或者能赋予抗生素抗性（如氨基青霉素、卡那霉素、氯霉素或四环素抗性）的标记。

优选本发明所述载体包含这样的元件，它（们）允许载体稳定地整合到宿主细胞基因组中，或者保证载体在细胞中独立于细胞的基因组进行自主复制。

就整合到宿主细胞基因组的情况而言，载体可以依赖于编码多肽的核酸序列或载体的其他元件，以便载体通过同源或非同源重组稳定地整合到基因组中。可选择地，载体可含有用于引导通过同源重组整合到宿主细胞基因组中的附加核酸序列。附加的核酸序列能使载体整合到宿主细胞基因组中染色体上的精确位点。为了增加在精确位点整合的可能性，优选整合元件应含有足够数量的核酸，如 100 到 1500 个碱基对，优选 400 到 1500 个碱基对，最优选 800 到 1500 个碱基对，它们与相应的目标序列高度同源，从而增加了同源重组的可能性。整合元件可以是与宿主细胞基因组中的目标序列同源的任何序列。另外，整合元件可以是非编码或编码核酸序列。另一方面，载体可以通过非同源重组整合到宿主细胞的基因组中。

就进行自主复制的情况而言，载体还可以包含复制起点，使载体能在目标宿主细胞中自主地复制。细菌复制起点的例子是以下质粒的复制起点：允许在大肠杆菌中复制的 pBR322、pUC19、pACYC177 和 pACYC184 的复制起点，以及允许在芽孢杆菌中复制的 pUB110、pE194、pTA1060 和 pAM β 1 的复制起点。复制起点可以带有使其在宿主细胞中成为温度敏感型的突变（参见例如，Ehrlich, 1978, 美国科学院学报 75: 1433）。

可以向宿主细胞中插入 1 个以上拷贝的本发明核酸序列来提高基因产物的产量。可以通过将至少 1 个附加拷贝的序列整合到宿主细胞基因组中或者与该核酸序列一起插入一个可扩增的选择标记，来使核酸序列拷贝数增加，后一情况中，可以通过在有合适选择试剂存在下培养细胞，挑选出含有扩增拷贝的选择标记基因、从而含

有附加拷贝核酸序列的细胞。

用于连接前面描述的各元件来构建本发明所述重组表达载体的操作是本领域技术人员熟知的（参见例如，Sambrook等，1989，出处同前）。

宿主细胞

本发明还涉及包含本发明核酸序列的重组宿主细胞，这些细胞适合用来重组地制备多肽。术语“宿主细胞”涵盖由于复制期间发生的突变而与亲本细胞不同的任何后代。

将包含本发明所述核酸序列的载体导入宿主细胞，使载体保持染色体整合的或染色体外自我复制的形态。通常认为整合有优势，因为这样核酸序列更可能稳定地保持在细胞内。载体可以通过如上描述的同源或非同源重组整合到宿主的染色体中。

选择宿主细胞很大程度上取决于编码多肽的基因及其来源。宿主细胞可以是单细胞微生物（例如原核细胞）或非单细胞微生物（例如真核细胞）。有用的单细胞是细菌细胞，如革兰氏阳性细菌，包括，但不限于芽孢杆菌属的细胞，例如：嗜碱芽孢杆菌、解淀粉芽孢杆菌、短芽孢杆菌、环状芽孢杆菌、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、苏云金芽孢杆菌；或者链霉菌属的细胞，例如浅青紫链霉菌或鼠灰链霉菌；或者是革兰氏阴性细菌，例如大肠杆菌或假单胞菌。在一个优选实施方案中，细菌宿主细胞是迟缓芽孢杆菌、地衣芽孢杆菌、嗜热脂肪芽孢杆菌或枯草芽孢杆菌细胞。

可以通过例如原生质体转化（参见，例如，Chang和Cohen,1979,分子普通遗传学（Molecular General Genetics）168:111-115）、使用感受态细胞（参见，例如，Young和Spizizin,1961,细菌学杂志（Journal of Bacteriology）81:823-829,或Dubnau和Davidoff-Abelson,1971,分子生物学杂志（Journal of Molecular Biology）

56:209-221)、电穿孔(参见,例如, Shigekawa 和 Dower,1988,生物技术 6:742-751)或接合(参见,例如, Koehler 和 Thorne,1987,细菌学杂志 169: 5771-5278)将载体导入细菌宿主细胞。

制备方法

本发明还涉及多肽的制备方法,该方法包括(a)在适合所述多肽产生的条件下培养宿主细胞;和(b)回收该多肽。

在本发明所述制备方法中,用本领域已知方法在适合多肽产生的营养基质中培养细胞。例如,可以在合适的培养基中,在允许多肽表达和/或分离的条件下,通过摇瓶培养、在实验室或工业发酵罐中小规模或大规模发酵(包括连续、分批、分批加料或固态发酵)来培养细胞。在包含碳和氮源以及无机盐的合适的培养基中,采用本领域已知的步骤进行培养(参见,例如 M.V.Arbige 等在 Abraham L.Sonenshein, James A.Hoch 和 Richard Losick 所编《枯草芽孢杆菌及其它革兰氏阳性细菌》,美国微生物学会, Washington, D.C., 1993 中所述)。合适的培养基有供应商提供或者可以参照公开的成分(例如,美国典型培养物保藏中心的目录)来制备。如果多肽被分泌到培养基中,则可以直接从培养基中提纯多肽。如果多肽不分泌,则可以从细胞裂解物中纯化。

可以用本领域已知的所述多肽的特异方法检测多肽。这些检测方法包括使用特异抗体、形成酶产物或酶底物的消失。例如,可以用酶检测法确定多肽的活性。用于确定蛋白酶活性的步骤是本领域已知的,包括,例如测量因异硫氰酸荧光素标记的酪蛋白水解所发出的荧光。

可以用本领域已知方法回收所产生的多肽。例如,可以通过常规操作(包括,但不限于离心、过滤、萃取、喷雾干燥、蒸发或沉淀)从培养基中回收多肽。

可以通过各种本领域已知的操作来纯化本发明所述多肽,这些操作包括,但不限于层析(例如,离子交换层析、亲和层析、疏水

层析、层析聚焦、和大小排阻层析)、电泳操作(例如,制备性等电点聚焦)、分级溶解(例如硫酸铵沉淀)、SDS-PAGE或萃取(参见例如,蛋白质纯化, J.C.Janson 和 Lars Ryden 编, VCH Publishers, New York, 1989)。

去除或降低蛋白酶活性

本发明还涉及制备亲本细胞的突变细胞的方法,包括破坏或删除本发明核酸序列或者它的调控序列,从而得到由该核酸序列编码的多肽的产量比亲本减少的突变细胞。

可以将该有蛋白酶活性的多肽在细胞中进行表达所必需的本发明核酸序列修饰或使它失活来方便地构建出蛋白酶活性减少的菌株。要进行修饰或失活的核酸序列可以是,例如编码所述多肽或其表现蛋白酶活性所必需的部分的核酸序列,或者是由核酸序列的编码序列表达为多肽时所需要的有调控功能的核酸序列。这种调控或调节序列的例子可以是启动子序列或它的功能性部分(即足够影响多肽表达的那部分)。其他可能被修饰的调控序列如上所述。

可以对细胞进行诱变处理并挑选出其蛋白酶产生能力减少的细胞,从而使核酸序列修饰或失活。可以通过例如使用合适的物理或化学诱变剂、使用合适的寡核苷酸、或将DNA序列做PCR进行(特异性或随机)诱变。另外,可以将这些诱变剂组合起来进行诱变。

适用于本发明目的的物理或化学诱变剂的例子包括紫外线(UV)照射、羟胺、N-甲基-N'-硝基-N-亚硝基胍(MNNG)、O-甲基羟胺、亚硝酸、甲磺酸乙酯(EMS)、亚硫酸氢钠、甲酸和核苷酸类似物。

使用这类试剂时,通常于合适的条件下,在有所选诱变剂存在的情况下培养细胞使其突变,并挑选出蛋白酶活性或产量呈现减少的细胞。

可以通过在核酸序列或者其转录或翻译所必需的调控元件中导入、取代或删除1或多个核苷酸来完成本发明核酸序列所编码的多肽的修饰或失活。例如,可以插入或删除核苷酸来导入终止密码、

去掉起始密码子或改变开放读码框。可以依照本领域已知的方法通过定点诱变或 PCR 诱变来达到修饰或失活的目的。尽管一般来说，修饰可以在体内进行，即直接在待修饰核酸序列的表达细胞上进行，但优选象下面的例子一样在体外进行修饰。

使所用宿主细胞表达失活或减少的方便方法的例子是基于基因置换或基因破坏技术。例如，在基因破坏方法中，在体外将对应于目标内源基因或基因片段的核酸序列进行诱变，从而产生缺陷的核酸序列，然后将该序列转化到宿主细胞中，以产生缺陷型基因。经过同源重组，缺陷型核酸序列将取代内源基因或基因片段。可能希望缺陷基因或基因片段还编码标记物，以使用该标记物挑选出编码多肽的基因已经被修饰或破坏的转化子。

可选择地，可以利用成熟的反义技术，用与多肽编码序列互补的核苷酸序列使本发明核酸序列修饰或失活。更具体地说，可以导入与编码多肽的核酸序列互补的核苷酸序列，该序列在细胞中可以进行转录并能与细胞中产生的多肽 mRNA 杂交，从而减少或消除细胞中该多肽的产量。在能使得互补的反义核苷酸序列与多肽 mRNA 杂交的条件下，翻译出的多肽产量就会减少或消除。

优选地，欲按照本发明所述方法进行修饰的细胞是微生物来源的，例如是适合用来产生（对于细胞是同源或异源的）目标蛋白质产物的芽孢杆菌菌株。

本发明还涉及亲本细胞的突变细胞，其中所含编码多肽的核酸序列或其调控序列被破坏或缺失，从而导致突变细胞中的多肽产量少于亲本细胞。

这样得到的多肽缺陷型突变细胞特别适合作为表达同源和/或异源多肽的宿主细胞。因此，本发明还涉及同源或异源多肽的生产方法，包括（a）在适合多肽产生的条件下培养突变细胞；和（b）回收多肽。在本文中，术语“异源多肽”此处定义为宿主细胞的非天然多肽、其中有修饰而使天然序列改变的天然蛋白质，或者因采用重组 DNA 技术操作宿主细胞而使其表达量有改变的天然蛋白

质。

另一方面，本发明涉及通过发酵能产生本发明核酸序列编码的多肽和目标蛋白质产物的细胞来生产基本没有蛋白酶活性的蛋白质产物的方法。该方法包括在发酵期间或发酵完成后，向发酵液中添加足够量的能抑制蛋白酶活性的试剂，从发酵液中回收目标产物，以及任选将回收的产物进一步纯化。在下面的实施例中将对该方法作进一步阐述。

本发明另一可选择的方面涉及生产基本上没有蛋白酶活性的蛋白质产物的方法，其中的目标蛋白质产物由细胞中的 DNA 序列编码，该细胞中也含有编码有蛋白酶活性的多肽的本发明核酸序列。该方法包括在允许产物表达的条件下培养细胞，将得到的培养液进行 pH 和温度的组合处理以便较大地减少蛋白酶活性，以及从培养液中回收产物。可选择地，可以针对从培养液中回收的酶制剂进行 pH 和温度的组合处理。任选地，可以将 pH 和温度的组合处理与用蛋白酶抑制剂进行的处理结合使用。

根据本发明的这一方面，可以将蛋白酶活性消除至少 60%、优选至少 75%、更优选至少 85%、还要优选的是至少 95%，最优选至少 99%。预计通过使用这些方法可以完全消除蛋白酶活性。

优选在 pH6.5-7、25-70℃ 的范围内，使 pH 和温度的组合处理进行充分的时间以达到预期的效果，通常约 30-60 分钟即可。

可以采用本领域已知的方法培养和纯化目标产物。

本发明所述用于生产基本没有蛋白酶活性的产物的方法对于生产原核多肽（尤其是芽孢杆菌蛋白质，比如酶）特别有意义。所述的酶可以选自，例如，淀粉分解酶、脂类分解酶、蛋白水解酶、纤维素分解酶、氧化还原酶或植物细胞壁降解酶。这些酶的例子包括氨肽酶、淀粉酶、淀粉糖苷酶、糖酶、羧肽酶、过氧化氢酶、纤维素酶、壳多糖酶、角质酶、环糊精糖基转移酶(cyclodextrin glycosyltransferase)、脱氧核糖核酸酶、酯酶、半乳糖苷酶、β-半乳糖苷酶、葡糖淀粉酶、葡萄糖氧化酶、葡糖苷酶、卤化物过氧化物

酶 (haloperoxidase)、半纤维素酶、转化酶、异构酶、漆酶、连接酶、脂酶、裂解酶、甘露糖苷酶、氧化酶、果胶水解酶、过氧化物酶、植酸酶、酚氧化酶、多酚氧化酶、蛋白水解酶、核糖核酸酶、转移酶、转谷氨酰胺酶或木聚糖酶。可以用蛋白酶缺陷的细胞表达有药用价值的异源蛋白质。

术语“原核细胞多肽”应理解为不仅包括天然多肽，也包括通过氨基酸取代、缺失或添加或者其他修饰使它们的活性、热稳定性、pH 耐受能力等增强的其它一些多肽 (例如酶)。

本发明另一个方面涉及通过本发明方法制备的基本上没有蛋白酶活性的蛋白质产物。

用途

本发明核酸序列编码的重组多肽可以用于蛋白水解酶的常规应用中，在高 pH 下尤其有用，例如用于衣物和碗碟洗涤剂中，公共机构和工业清洗中，以及皮革加工中。由于重组多肽对碱性条件下的氧化作用 (例如过氧化物型漂白剂) 的稳定性有所提高，因此它们尤其可用于洗涤剂中。

重组多肽也可用于多种其它应用中，包括增强或降低蛋白水解物的水解程度，通过蛋白质的水解调味，降解不需要的肽以及酶促合成肽。本领域中早已确认了蛋白酶在这些及其它应用中的用途。

下述实施例进一步描述了本发明，这些实施例不应理解为是对本发明范围的限制。

实施例

所有引物和寡聚物均在应用生物系统 394 型合成仪 (Applied Biosystems Inc., Foster City, C.A.)，根据生产厂家的说明所合成。

实施例 1: 枯草芽孢杆菌供体菌株 BW154 的构建

这里所述枯草芽孢杆菌宿主菌株 A164(ATCC6051A) 和 1630(NCFB 736) 中均有数个基因 (*spoIIAC*, *aprE*, *nprE*, *amyE*, 和

srfC)被缺失。为了完成这个任务,利用 pLS20介导的接合系统(Koehler和Thorn,1987,见前),向这些菌株中导入了含有这些基因的缺失形式的质粒。简要地说,这个系统包括一个含有大质粒 pLS20的枯草芽孢杆菌“供体”菌株。pLS20编码将 pLS20 迁移至枯草芽孢杆菌受体菌株所必需的功能。此外,已知质粒诸如 pUB110和 pBC16也可由这个接合系统(在 pLS20存在下)进行迁移。这些质粒含有顺式作用区(*oriT*)和编码作用于 *oriT* 位点并推动这些质粒向受体菌株转移的反式作用功能的基因(*orf-beta*)。如果供体菌株含有 pLS20 以及 pUB110和 pBC16二者中的任一种,则仅含有 *oriT* 的质粒也能被迁移(这时, *orf-beta* 功能由反式提供)。

有效的供体菌株必须含有质粒 pLS20 或其衍生质粒如 pXO503(Koehler和 Thorne,1987,见前)。此外,也希望在接合完成之后有反筛选供体菌株的方法。已经发展了一种非常“干净”(无背景)且易于实施的反筛选方案。这涉及在供体菌株的 *dal* 基因(编码细胞壁合成所需的 D-丙氨酸消旋酶)中导入缺失以及将接合实验后所得细胞混合物涂板于 D-丙氨酸缺失的固体培养基上而对供体菌株进行筛选(枯草芽孢杆菌 *dal*-菌株只能在添加了外源丙氨酸的培养基中才能生长)。

为了缺失上面所提到的基因,必须将含有这些基因的缺失形式的 pE194复制子(红霉素抗性)(Gryczan 等, 1982, 细菌学杂志 152: 722-735)和 *oriT* 序列转移到枯草芽孢杆菌 A164和 A1630菌株中。合适的供体菌株应有以下特征: 1) *dal* 基因已缺失(为了进行反筛选), 和 2) 它还必须含有质粒 pLS20(此时 pXO503 不适用, 因为 pE194 复制子必须通过红霉素筛选保持, 而 pXO503 已赋予对红霉素的抗性), 以及 pUB110或 pBC16以反式提供 *orf-beta* 功能。有关枯草芽孢杆菌 BW154 如何构建成为供体菌株的叙述如下。

(A)在枯草芽孢杆菌中导入 *dal* 缺失突变, 得到枯草芽孢杆菌 BW96。

首先, 选择 *bac-1* 基因内有突变的枯草芽孢杆菌菌株(该基因突变使得菌株失去了合成二肽抗生素杆菌溶素的能力), 因为在接合过程

中野生型枯草芽孢杆菌细胞会杀死其他种芽孢杆菌，而*bac-1*-细胞中这种潜在的杀伤力则大大降低。因此，所有供体菌株均是在*bac-1*背景下构建的。

构建合适供体菌株的第一步为缺失背景为*bac-1*的枯草芽孢杆菌菌株1A758 (Bacillus Stock Center, 哥伦比亚, 俄亥俄州) 中*dal*基因的部分。体外构建了能够替换细菌染色体上野生型*dal*基因的*dal*基因缺失形式。利用PCR扩增得到*dal*基因的5'和3'部分，引物1和2用来扩增该基因的5'部分(核苷酸19-419, ATG密码子的A是+1位)，引物3和4用来扩增该基因的3'部分(核苷酸618-1037)。

引物1: 5'-GAGCTCACAGAGATACGTGGGC-3' (SEQ ID NO:1)

引物2: 5'-GGATCCACACCAAGTCTGTTCAT-3' (SEQ ID NO:2)(划线部分为BamHI位点)

引物3: 5'-GGATCCGCTGGACTCCGGCTG-3' (SEQ ID NO:3)(划线部分为BamHI位点)

引物4: 5'-AAGCTTATCTCATCCATGGAAA-3' (SEQ ID NO:4)
(划线部分为HindIII位点)

扩增反应(100 μ l)含有以下组分: 200ng枯草芽孢杆菌168染色体DNA, 每一引物浓度均为0.5 μ M, dATP、dCTP、dGTP、dTTP浓度各为200 μ M, 1 \times Taq聚合酶缓冲液, 以及1U Taq DNA聚合酶。依照Pitcher等, 1989, 应用微生物学通讯 8: 151-156所述方法得到枯草芽孢杆菌168的染色体DNA。反应条件如下: 95 $^{\circ}$ C 3分钟, 然后再进行30个循环, 每个循环程序为95 $^{\circ}$ C 1分钟、50 $^{\circ}$ C 1分钟和72 $^{\circ}$ C 1分钟, 最后为72 $^{\circ}$ C 5分钟。反应产物经琼脂糖凝胶电泳检测。根据生产商的操作指示, 利用TA克隆试剂盒(Invitrogen, 圣地亚哥, 加州) 将5'和3' PCR产物克隆至pCRII载体中。鉴定出含有*dal*基因5'端一半、且由PCR引物导入的BamHI位点与pCRII载体多接头的BamHI位点相邻的pCRII克隆(另一方向的克隆中BamHI位点间相距要远得多)。然后用BamHI和HindIII消化含有*dal*基因3'端一半的pCRII克隆, 再将该*dal*基因片段克隆至前述含有*dal*基因5'端一半的pCRII

克隆的BamHI-HindIII位点, 这样所产生的pCRII载体含有中部缺失了约200bp的*dal*基因, 在该基因的5'末端有一NotI位点(pCRII多接头的部分), 3'末端为HindIII位点。

为了将该*dal*缺失导入细菌染色体, 将该缺失基因克隆至枯草芽孢杆菌温度敏感型复制子pE194(Gryczan 等, 1982, 见前)中。然后经两步将该缺失*dal*基因导入染色体中: 首先通过同源重组将此质粒整合到染色体*dal*位点, 接着再除去质粒(还是通过同源重组), 使得*dal*基因的缺失形式仍保留在细菌染色体上。这可通过下列步骤完成: 将该缺失*dal*基因片段(如上述)克隆至温度敏感型质粒pSK⁺/pE194的NotI-HindIII位点(实质上是用*dal*Δ片段替代pSK⁺载体序列)。质粒pSK⁺/pE194的构建方法如下: 用XbaI酶切消化BluescriptSK⁺(Stratagene, La Jolla, CA)和pE194, 然后用小牛肠碱性磷酸酶处理pSK⁺载体, 再将两质粒连接起来。用连接混合物转化大肠杆菌菌株DH5α, 转化体在含氨苄青霉素(100 μg/ml)和X-gal(5-溴-4-氯-3-吡啶-β-D-半乳糖苷)的LB培养基上进行筛选。从数个“白色”菌落中纯化出质粒, 利用限制酶切消化并随后凝胶电泳鉴定出含有pE194和pSK⁺的嵌合体。用HindIII和NotI消化该质粒。然后凝胶纯化含有pE194复制子的片段, 并将其与同样经凝胶纯化的*dal*Δ基因片段(HindIII-NotI)连接。连接产物被用来转化枯草芽孢杆菌*bac*-1菌株1A758(Bacillus Stock Center, 哥伦布, OH), 并在含红霉素(5mg/ml)的胰蛋白胨血琼脂碱(TBAB)平板中选择转化体, 在许可温度34℃下进行培养。从5个红霉素抗性转化体中纯化出质粒DNA, 并经限制酶切消化/凝胶电泳分析。鉴定出对应于含有*dal*缺失片段的pE194的质粒。带有该质粒的菌株随后被用来通过同源重组向染色体中导入*dal*缺失。

为了获得首次交换(*dal*缺失质粒整合到染色体上的*dal*基因内), 将转化菌株在含有D-丙氨酸(0.1mg/ml)和红霉素(5 μg/ml)的TBAB平板上划线, 在非许可温度45℃下生长过夜。取一大菌落再次于相同条件下进行划线, 得到染色体上的*dal*基因内整合有该温度敏感型质粒

的细胞的均一群体。在非许可温度下，由于质粒不能复制，因此只有在染色体上含有该质粒的细胞才能在红霉素平板上生长。为了获得第二次交换事件(导致质粒从染色体上切出，只留下*dal*基因的缺失形式)，将一环细胞转移至20ml含D-丙氨酸(0.1mg/ml)的Luria培养液中，在许可温度34℃下进行无选择生长至对数后期，以使得复制起点保持功能，并发生第二次交换事件。细胞再转移四次(每次转移稀释100倍)，以使得质粒从染色体上切除并从群体中分离出去。最后，细胞在34℃下涂板于添加有D-丙氨酸(0.1mg/ml)的TBAB平板上以生长出单菌落，并将菌落印迹转移至无D-丙氨酸(0.1mg/ml)的TBAB平板上和含有D-丙氨酸(0.1mg/ml)和红霉素(5 μg/ml)的TBAB平板上，以便对*dal*-和*erm^S*菌落计数。50个菌落中有2个具有这种表型。将所得菌株命名为枯草芽孢杆菌BW96，这是一种*bac-1*, *dal*-菌株。

(B) 将pLS20和pBC16导入枯草芽孢杆菌*bac-1*, *dal*缺失菌株，得到接合有效供体菌株枯草芽孢杆菌BW154。

选择一种供体菌株，用于向枯草芽孢杆菌BW96中导入pLS20和pBC16质粒，该供体菌株应具有以下特征：本身为红霉素敏感型(为了提供对抗供体菌株的反筛选)含有pLS20和pBC16的枯草芽孢杆菌菌株。一种含有pLS20和pBC16的*dal*缺失枯草芽孢杆菌菌株被选为合适的供体菌株，其构建如下：用pHV1248(Petit等, 1990, 细菌学杂志 172: 6736-6740)转化枯草芽孢杆菌DN1686(美国专利号 4,920,048),使得细胞获得红霉素抗性。通过与枯草芽孢杆菌(*natto*) 3335UM8(Koehler和Thorne,1987,见前)的接合转移,接合因子pLS20和质粒pBC16一起被转移至枯草芽孢杆菌DN1686(pHV1248)菌株中。筛选得到的转接合子为含有*dal*缺失突变的四环素和红霉素抗性菌落。根据带有pLS20的菌落通过接合转移pBC16至其它枯草芽孢杆菌菌株的能力对它们进行评分。最后，通过提高培养温度至50℃，使得接合菌株丢掉pHV1248，从而得到供体菌株：含有pLS20和pBC16的枯草芽孢杆菌DN1686。

为了将这些质粒导入枯草芽孢杆菌BW96中，必须执行恰当的

反筛选, 因此, 用赋予红霉素抗性的温度敏感型质粒pSK⁺/pE194转化枯草芽孢杆菌BW96, 其随后可通过生长于非允许温度下而除去。根据以下程序, 将pLS20和pBC16从含有pLS20和pBC16的枯草芽孢杆菌DN1686中迁移至枯草芽孢杆菌BW96(带有pSK⁺/pE194)中。每种类型细胞各取一接种环的量, 混合涂在添加了D-丙氨酸(50 μg/ml)的TBAB平板上, 33℃下温育5小时。从平板上刮取细胞, 转移至1ml LB培养基中。将多种稀释度的细胞铺在添加了四环素(10 μg/ml), 红霉素(5 μg/ml)和D-丙氨酸(50 μg/ml)的TBAB平板上, 34℃生长, 以筛选获得了pBC16以及许多情况下也获得了pLS20的受体细胞。为了检测pLS20是否也存在于任一转接合子中, 测试了十个菌落转移pBC16至枯草芽孢杆菌PL1801中的能力。枯草芽孢杆菌PL1801为缺失了apr和npr基因的枯草芽孢杆菌168(芽孢杆菌储藏中心, 哥伦布, OH)。然而, 也可用枯草芽孢杆菌168。能够迁移pBC16的供体一定也含有pLS20。一旦鉴定出能有效接合的菌株(含有pLS20及pBC16和pSK⁺/pE194的*bac-1*, *dal*-枯草芽孢杆菌), 即通过在添加了四环素(5 μg/ml)和D-丙氨酸(50 μg/ml)的LB培养基中于45℃过夜增殖细胞、在33℃铺板于添加了D-丙氨酸(50 μg/ml)的TBAB平板上以获得单菌落, 并鉴定出红霉素敏感型菌落, 以便从该菌株中除去pSK⁺/pE194质粒。该程序产生了枯草芽孢杆菌BW154, 即含有pLS20和pBC16的*bac-1*, *dal*-枯草芽孢杆菌。

表1简要总结了各芽孢杆菌菌株和质粒。

表1: 芽孢杆菌菌株和质粒

枯草芽孢杆菌菌株:

枯草芽孢杆菌(natto)	pLS20
DN1686	<i>dal</i> -
DN1280	<i>dal</i> -
MT101	DN1280(pXO503)
1A758	168 <i>bac-1</i> (芽孢杆菌储藏中)

心, 哥伦布, Ohio)

BW96	1A758 <i>dal</i> Δ
BW97	1A758 <i>dal</i> Δ :: <i>cat</i> (pXO503)
BW99	1A758 <i>dal</i> Δ (pPL2541-tet)
BW100	1A758 <i>dal</i> Δ (pXO503),(pXL2541-tet)
PL1801	<i>apr</i> Δ, <i>npr</i> Δ

质粒:

pBC16	Mob ⁺ , Tc ^r
pE194	温度敏感型
pLS20	Tra ⁺
pXO503	Tra ⁺ , MLS ^r (=pLS20::Tn917)
pPL2541-tet	Mob ⁺ , Tc ^r (pE194 <i>ts ori</i>)
pCAsub2	Mob ⁺ , Cm ^r , Apr ^r , (pE194 <i>ts ori</i>)
pSK ⁺ /pE194	Em ^r , Apr ^r , 温度敏感型
pShv2	Tra ⁺ , Em ^r , Cm ^r , 温度敏感型
pHV1248	Em ^r , 温度敏感型

Tra⁺表示该质粒赋与携有它的任一枯草芽孢杆菌菌株进行接合的能力, 也就是说, 该质粒编码将一接合型质粒从供体细胞转移至受体细胞所需的全部功能。

Mob⁺表示质粒可被含有Tra⁺质粒(pLS20或pXO503)的菌株通过接合进行转移。该质粒必须含有顺式作用序列和编码反式作用蛋白的基因(例如在pBC16中分别为oriT和orf-beta), 或仅有oriT序列(例如pPL254-tet, 这时细胞中还必须存在反式提供orf-beta功能的质粒, 如pBC16)。

实施例2: 枯草芽孢杆菌A164(ATCC 6051A)中spoIIAC基因的缺失
利用重叠延伸剪切(SOE)技术(Horton等, 1989, 基因77: 61-68), 制备spoIIAC基因的缺失形式, 该基因编码sigmaF, 使得细胞通过孢

子形成II期。利用Pitcher等(1989, 出处同前)的方法获得枯草芽孢杆菌A164(ATCC6051A)的染色体DNA。合成下面列出的引物5和6, 以用于PCR扩增枯草芽孢杆菌A164的染色体DNA上从spoIIAC基因ATG起始密码子上游205个核苷酸至ATG起点下游209个核苷酸的区域。上游引物的下划线核苷酸为添加的HindIII位点。下游引物的下划线核苷酸与ATG翻译起始密码子下游507至524位碱基互补。合成引物7和8, 以用于PCR扩增ATG翻译起始密码子下游从507位延伸至884位核苷酸的序列区。引物7的下划线区与用于扩增上游片段的引物6的3'端部分完全互补。

引物5: 5'-AAGCTTAGGCATTACAGATC-3' (SEQ ID NO:5)

引物6: 5'-CGGATCTCCGTCATTTTCCAGCCCGATGCAGCC-3' (SEQ ID NO:6)

引物7: 5'-GGCTGCATCGGGCTGGAAAATGACGGAGATCCG-3' (SEQ ID NO:7)

引物8: 5'-GATCACATCTTTCGGTGG-3' (SEQ ID NO:8)

在各自的PCR扩增中使用这两套引物来分别扩增spoIIAC基因的上游和下游片段。扩增反应体系(25 μ l)含有以下组分: 200ng枯草芽孢杆菌A164染色体DNA, 每条引物0.5 μ M, dATP, dCTP, dGTP, 和dTTP各为200 μ M, 1 \times Taq DNA聚合酶缓冲液, 以及0.625U的Taq DNA聚合酶。按照Pitcher等(1989, 出处同前)的操作获得枯草芽孢杆菌A164染色体DNA。反应条件如下: 96 $^{\circ}$ C 3分钟, 然后依次按96 $^{\circ}$ C 1分钟, 50 $^{\circ}$ C 1分钟和72 $^{\circ}$ C 1分钟为一个循环进行30个循环, 最后再72 $^{\circ}$ C 3分钟以确保扩增片段末端加上腺苷酸残基(Invitrogene, 圣地亚哥, 加州)。目标产物的扩增经1.5%琼脂糖凝胶电泳加以确证。

然后在含上述每种扩增反应产物各2.5 μ l, 而除了只加有引物5和8外其它反应条件同上的新PCR反应体系中完成扩增, 得到长为1089个核苷酸的“剪切”片段, 代表内部缺少298个核苷酸的spoIIAC基因。利用Invitrogen TA 克隆试剂盒, 按照生产商的操作指示, 将该片段克隆至pCRII载体中, 作为HindIII-EcoRI片段切出, 再克隆至用HindIII/EcoRI消化过的pShv2中。pShv2(图1)是一个穿梭载体, 通过将含有pUB110 oriT的pBCSK⁺ (Stratagene, La Jolla, 加

州)XbaI酶切片段与pE194 XbaI酶切片段连接,再连入来自pUB110的经PCR扩增出的含有SstI相容性末端的oriT片段而构建成。通过pLS20介导的接合(Battisti等, 1985, 细菌学杂志 162:543-550), oriT片段使得该质粒迁移至枯草芽孢杆菌A164中。pShv2 - Δ spoIIAC被转入供体菌株枯草芽孢杆菌BW154(实施例1)。枯草芽孢杆菌BW154(pShv2 - Δ spoIIAC)被用作供体菌株, 向枯草芽孢杆菌A164中导入该带有缺失基因的穿梭载体。

通过转有pShv2 - Δ spoIIAC的枯草芽孢杆菌BW154与枯草芽孢杆菌A164的接合, 筛选红霉素抗性转接合子, 并于45℃培养, 可实现缺失基因与完整染色体基因间的交换。在45℃下, pE194复制子没有活性, 细胞只能通过该含有缺失基因的质粒在spoIIAC位点的Campbell整合才能保持红霉素抗性。通过在34℃, 即允许pE194复制子行使功能的温度下, 在无抗生素选择的LB培养基中培养菌株两轮, 完成第二次重组, 导致载体DNA脱落(loopout), 并且缺失基因替代了完整spoIIAC基因。根据以下标准选择出已经发生基因替代的菌落: 1)穿梭载体pShv2编码的红霉素(erm)抗性缺失, 2)在孢子形成培养基上菌落不透明度降低, 表明不能产生孢子, 和3)用引物5和8经PCR扩增可获得一条791个核苷酸的片段而不是该基因未缺失形式的1089个核苷酸。

实施例3 枯草芽孢杆菌A164 Δ spoIIAC中nprE基因的缺失

根据实施例2所述方法制备出枯草芽孢杆菌A164 Δ spoIIAC染色体DNA作为模板, 利用下述引物9和10, PCR扩增出中性蛋白酶(nprE)基因的上游部分序列(GTG起始密码子下游第40-610位核苷酸)。利用下列引物11和12, 经PCR扩增出nprE基因的下游部分(核苷酸1040-1560)。引物10和11设计成在两片段中有15个碱基对重叠(下划线部分)。扩增反应体系(25 μ l)含有实施例2所述的相同组分, 并在相同条件下完成。

引物9: 5'-CGTTTATGAGTTTATCAATC-3' (SEQ ID NO:9)

引物10: 5'-AGACTTCCCAGTTTGCAGGT-3' (SEQ ID NO:10)

引物11: 5'-CAAACTGGGAAGTCTCGACGGTTCATTCTTCTCTC-3' (SEQ ID NO:11)

引物12: 5'-TCCAACAGCATTCCAGGCTG-3' (SEQ ID NO:12)

参照生产商的操作指示, 对所扩增的上游和下游片段用 QiaexII 试剂盒(Qiagen, Chatsworth, 加州)进行凝胶纯化。在含每种纯化片段各约20ng的新的PCR混合物(100 μl)中进行反应。以下条件用来进行SOE反应: 按实施例2的条件, 先不加引物进行第1-3个循环, 产生“剪接”片段, 再在加入引物9和12的条件下进行第4-30个循环。将所扩增的SOE片段克隆进pCRII载体, 并进行限制性酶切分析确证。该片段随后作为BamHI-XhoI片段克隆进pShv2。这个质粒, pShv2-Δ nprE被转化至枯草芽孢杆菌BW154中, 得到可用于接合的合适供体菌株。该质粒随后被迁移至枯草芽孢杆菌A164Δ spoIIAC中。利用实施例2所述的温度变换方法, 将Δ nprE基因引入枯草芽孢杆菌A164Δ spoIIAC的染色体上。将erm^S菌落涂在加有1%脱脂奶粉的TBAB琼脂平板上, 37℃生长过夜, 从而对nprE-表型评分。(nprE-菌株的清亮带显著减弱。)利用引物9和12进行染色体DNA的PCR分析, 证实有430个碱基对被删除。

实施例4 : 枯草芽孢杆菌A164Δ spoIIACΔ nprE中aprE基因的缺失

利用SOE方法产生出枯草芽孢杆菌中编码碱性枯草溶素杆菌蛋白酶(subtilisin)的aprE基因的缺失形式。根据实施例2所述方法, 利用下列引物13和14, 从枯草芽孢杆菌A164染色体DNA上PCR扩增出aprE基因的上游部分序列, 得到从翻译起始密码子上游189个核苷酸至起始点下游328个核苷酸的一个片段。引物13中带下划线的核苷酸为添加的EcoRI位点。引物14中带下划线的核苷酸处添加有SalI位点, 并提供与下游PCR片段的互补性。选用引物15和16, 经PCR扩增出aprE基因的下游部分序列, 得到从aprE翻译起始密码子下游789位核苷酸至1306位核苷酸的一个片段。引物14和引物15中带下划线的区域为上游片段与下游片段间提供互补。引物16的下划线核苷酸添加了HindIII位点。扩增反应体系(25 μl)含有如实施例2所述的相同组分, 并在相同条件下执行。

引物13: 5'-GCGAATTCTACCTAAATAGAGATAAAATC-3' (SEQ ID NO:13)

引物14: 5'-GTTTACCGCACCTACGTCGACCCTGTGTAGCCTTGA-3' (SEQ ID NO:14)

引物15: 5'-TCAAGGCTACACAGGGTTCGACGTAGGTGCGGTAAAC-3' (SEQ ID NO:15)

引物16: 5'-GCAAGCTTGACAGAGAACAGAGAAGCCAG-3' (SEQ ID NO:16)

参照生产商的操作指示, 利用 Qiaquick PCR 纯化试剂盒 (Qiagen, Chatsworth, 加州) 纯化出扩增得到的上游和下游片段。然后利用引物13和16将两种纯化片段剪接在一起。除了染色体DNA用上游和下游PCR产物各2 μ l 替代外, 扩增反应体系(50 μ l)含有与前述相同的组分。反应先在96 $^{\circ}$ C 3分钟温育1个循环(无dNTP和Taq聚合酶), 然后进行30个循环, 每个循环程序为96 $^{\circ}$ C 1分钟再72 $^{\circ}$ C 1分钟, 这样产生了aprE中缺失编码区460个核苷酸的缺失形式。反应产物经琼脂糖电泳分离, 克隆至pCRII中, 以EcoRI-HindIII片段切出, 再克隆至经EcoRI/HindIII消化的pShv2载体中, 得到pShv2- Δ aprE。将该质粒导入以上所述的供体菌株中, 用于接合转移至枯草芽孢杆菌A164 Δ spoIIAC Δ nprE中。

根据以上所述spoIIAC和nprE的缺失方法, 用缺失基因替换aprE基因。利用红霉素敏感型以及在加有1%脱脂奶粉上层的TBAB琼脂平板上清亮带的减弱, 选择出aprE基因已被缺失的菌落。aprE的缺失经PCR确证。

枯草芽孢杆菌A164 Δ spoIIAC Δ nprE Δ aprE在这里被称为枯草芽孢杆菌A164 Δ 3。

实施例5: 枯草芽孢杆菌A164 Δ spoIIAC Δ nprE Δ aprE中amyE基因的缺失

利用SOE方法产生编码枯草芽孢杆菌 α -淀粉酶的amyE基因的缺失形式。利用下列引物17和18从枯草芽孢杆菌A164染色体DNA中扩增出amyE的上游部分序列, 得到从amyE翻译起始密码子上游421位核苷酸延伸至amyE编码区第77位核苷酸的DNA片段, 并在上游末端加有SalI位点, 在下游末端加有SfiI和NotI位点。利用如下所示的引物19和20, 经PCR扩增出amyE基因的下游部分序列, 得到从

amyE编码区第445位核苷酸至953位核苷酸的片段，并在上游末端加有SfiI和NotI位点，在下游末端加有HindIII位点。限制性酶切位点用下划线标出。扩增反应体系(25 μl)含有与实施例2所述的相同组分，并在相同条件下执行。

利用引物17和20通过PCR将两条片段剪接起来。除了其中染色体DNA被换成上游和下游PCR产物各2 μl外，扩增反应体系(25 μl)含有与前述相同的组分。反应先在96℃ 3分钟温育1个循环(无dNTP和Taq聚合酶)，然后进行30个循环，每个循环程序为96℃ 1分钟再72℃ 1分钟。该反应通过两片段在互补区的重叠(SfiI和NotI位点)将这两片段融合，得到缺少编码区367个核苷酸、并在amyE的两部分序列之间导入了SfiI和NotI位点的amyE片段。根据标准方法进行1%琼脂糖凝胶电泳，分离出反应产物。参照生产商的操作指示，将该片段克隆至pCRII中，产生pCRII-Δ amyE。

引物17: 5'-CGTCGACGCCTTTGCGGTAGTGGTGCTT-3' (SEQ ID NO:17)

(划线处为SalI位点)

引物18: 5'-CGCGGCCCGCAGGCCCTTAAGGCCAGAACCAAATGAA-3' (SEQ ID NO:18)

(划线处为NotI和SfiI位点)

引物19: 5'-TGGCCTTAAGGGCCTGCGGCCGCGATTTCCAATG-3' (SEQ ID NO:19)

(划线处为SfiI和NotI位点)

引物20: 5'-GAAGCTTCTTCATCATCATTGGCATACG-3' (SEQ ID NO:20)

(划线处为HindIII位点)

用NotI消化pShv2，用Klenow片段和dNTP补平粘性末端，并重新连接该质粒，得到pShv2.1。该步骤破坏了pShv2的NotI位点。从pCRII-Δ amyE上以SalI-HindIII片段形式切下缺失型amyE片段，克隆至用SalI和HindIII双消化的pShv2.1中，得到pShv2.1-Δ amyE。将该质粒导入枯草芽孢杆菌BW154，以便接合转移至枯草芽孢杆菌A164Δ spoIIACΔ nprEΔ aprE中。

根据以上就spoIIAC，nprE和aprE所述的方法，用缺失基因替代amyE基因。利用红霉素敏感性以及失去在加有天青色淀粉顶层的平

板上产生清亮带的能力，选择出已经发生基因替代的菌落。通过利用引物17和20对染色体DNA进行该缺失基因的PCR扩增，证实amyE基因的缺失。

实施例6: 缺失枯草芽孢杆菌A164 Δ spoIIAC Δ nprE Δ aprE Δ amyE中的srfC基因，得到枯草芽孢杆菌A164 Δ spoIIAC Δ nprE Δ aprE Δ amyE Δ srfC

合成下列引物21-24，用于缺失枯草菌脂肽操纵子的srfC基因。引物21与srfC基因中已存在的HindIII位点(下划线)有重叠，与引物22相配合，可以PCR扩增出从srfC翻译起始点下游410位核苷酸至848位核苷酸的DNA片段。引物22的下划线部分与ATG起始密码子下游第1709-1725位核苷酸互补。用引物23和24经PCR扩增出srfC翻译起始点下游1709至2212位核苷酸的区域。引物23的下划线部分与ATG密码子下游835-848位核苷酸互补。扩增反应体系(25 μ l)含有与实施例2所述相同的组分，并在相同条件下执行。

引物21: 5'-AAGCTTTGAATGGGTGTGG-3' (SEQ ID NO:21)

引物22: 5'-CCGCTTGTTCCTTTCATCCCCTGAAACAACTGTACCG-3' (SEQ ID NO:22)

引物23: 5'-CAGTTGTTTCAGGGGATGAAAGAACAAGCGGCTG-3' (SEQ ID NO:23)

引物24: 5'-CTGACATGAGGCACTGAC-3' (SEQ ID NO:24)

利用Qiagen PCR离心柱(Qiagen, Chatsworth, 加州)，从PCR产物中除去引物和其它污染物。经PCR产生的两个片段间的互补性使得可通过SOE完成剪接。除前3个循环是在加入引物之前进行的以延伸重叠区外，其它PCR条件与以上所述相同，加入“外端引物”——引物21和24，将PCR产物(每种2 μ l或大约50ng)剪接到一起。经SOE反应得到缺失srfC基因内部859个核苷酸的长为955个核苷酸的片段。srfC基因被缺失的部分是负责向枯草菌脂肽分子添加第七个氨基酸(即亮氨酸)的区域，它的缺失进一步引起基因的移码突变，导致肽在类似硫酯酶活性位点的区域之前终止，而推测该区域参与SrfC蛋白的枯草菌脂肽释放(Cosmina等，1993，见前)。

根据以上就spoIIAC, nprE, aprE和amyE述及的缺失方法，用

缺失基因替代srfC基因。利用红霉素敏感性特征以及失去在血琼脂平板上产生清亮带的能力(Grossman等, 1993, 细菌学杂志 175: 6203-6211), 以及在加有50ml含有10%蔗糖、4%大豆粉、0.42%无水Na₂HPO₄和0.5% CaCO₃并添加5 μg/ml氯霉素的PS-1培养基的250ml摇瓶中于37℃ 250rpm培养4天后缺少泡沫的产生, 来选择已经发生基因替代的菌落。利用引物21和24, 从染色体DNA中PCR扩增缺失基因, 确证srfC基因已发生缺失。

枯草芽孢杆菌A164Δ spoIIACΔ nprEΔ aprEΔ amyEΔ srfC在这里被称为枯草芽孢杆菌A164Δ 5。

实施例7: 枯草芽孢杆菌A1630Δ spoIIACΔ nprEΔ aprEΔ amyEΔ srfC的构建

根据实施例1-6就枯草芽孢杆菌A164Δ spoIIACΔ nprEΔ aprEΔ amyEΔ srfC(枯草芽孢杆菌A164Δ 5)所述的相同程序, 利用为枯草芽孢杆菌A164缺失所构建的诸缺失质粒, 从枯草芽孢杆菌A1630(NCFB736, 旧称NCDO736)构建出枯草芽孢杆菌A1630Δ spoIIACΔ nprEΔ aprEΔ amyEΔ srfC。

枯草芽孢杆菌A1630Δ spoIIACΔ nprΔ aprΔ amyEΔ srfC在这里被称为枯草芽孢杆菌A1630Δ 5。

实施例8: 芽孢杆菌JP170染色体DNA的制备

芽孢杆菌JP170(NCIB12513)在含0.1M NaHCO₃, pH 8的50ml LB培养液中, 于37℃过夜生长。基因组DNA的制备见Pitcher等(1989, 见上)的所述方法。

实施例9: 芽孢杆菌JP170蛋白酶基因的探针制备

根据以下所示的芽孢杆菌JP170蛋白酶(JP4197182)的N-末端和内部氨基酸序列, 合成用于克隆芽孢杆菌JP170蛋白酶基因的引物:

N-末端: NDVARGIVKADVAQNNFGLYGQGQIVADTGLDTGRNDS (SEQ ID NO:25)

内部肽: GAADVGLGFPNGNQGWGRVTLDK (SEQ ID NO:26)

如下所示命名为170-291, 1701和1702B(其中I为肌苷)的引

物被用于以下所述的扩增反应。

170-291: 5'-CCCCAICCITGITTICITTIGGGIAAICC-3' (SEQ ID NO:27)

1701: 5'-GGIATIGTIAAIGCIGAIGTIGCICAAIAAAITTTGG-3' (SEQ ID NO:28)

1702B: 5'-TAIGGICAIGGICAIATIGTIGCIGTIGCIGAIACIGG-3' (SEQ ID NO:29)

制备含有下述组分的扩增反应物: 50 pmol 引物 1701 和 170-291 或者引物 1702B 和 170-291, 7 μ g 芽孢杆菌 JP170 染色体 DNA 作为模板, 1 \times PCR 缓冲液 (Perkin-Elmer 公司, Foster City, CA), dATP, dCTP, dGTP, dTTP 每种各 100 μ M, 和 0.5U AmpliTaq Gold (Perkin-Elmer 公司, Foster City, CA)。将反应物温育于 Stratagene Robocycler 40 (Stratagene 公司, La Jolla, CA) 热循环仪上, 扩增程序设置为先 96 $^{\circ}$ C 3 分钟一个循环, 再依 40 $^{\circ}$ C 1 分钟, 40 $^{\circ}$ C 1 分钟, 和 72 $^{\circ}$ C 1 分钟为一个循环进行 30 个循环。

由引物 170-291 和 1701 扩增出被命名为 1/291 的 905 bp 产物, 由引物 1702B 和 170-291 扩增出被命名为 2B/291 的 863 bp 产物。所有 PCR 产物分别依照生产商用法说明克隆至 Invitrogen TA 克隆试剂盒载体 pCR2.1 (Invitrogen 公司, San Diego, CA) 中。利用 Applied Biosystems Model 377 Sequencer (Applied Biosystems 公司, Foster Foster City, CA) 测序, 结果显示, 基于 GeneAssist 1.1b4 数据库 (Applied Biosystems 公司, Foster Foster City, CA) 中推导的氨基酸序列的比对, 这些 PCR 产物与 JP4197182 所公开的 Ya 蛋白酶氨基酸序列有 90% 的同一性。这些 PCR 产物的氨基酸序列与芽孢杆菌丝氨酸蛋白酶枯草杆菌蛋白酶的氨基酸序列有 35% 的同一性。

然后利用 Genius System PCR DIG 探针合成试剂盒 (Boehringer Mannheim Corporation, Indianapolis, IN), 依照生产商用法说明, 在如上所述同样的 PCR 条件下, 用引物 170-291、1701 和 1702B PCR 扩增出 1/291 和 2B/291 的 DIG 标记探针。

实施例 10: 染色体文库的筛选

利用实施例 9 中所述探针 2B/291 筛选芽孢杆菌 JP170 的染色体

文库。将 **Sau3A** 部分消化 (4-8Kb) 的芽孢杆菌 JP170 染色体 DNA 连至载体 pSJ1678 (图 2) 的 **BamHI** 位点, 构建出染色体文库。用该染色体文库转化大肠杆菌 DH5 α (Gibico BRL 公司, Gaithersburg, MD), 并依照 Genius System 的操作说明, 利用 DIG 标记探针 2B/291 进行菌落浸提物筛选。大约筛选 4600 个菌落之后, 找到一个菌落可与探针杂交, 被命名为克隆 1。利用 QIAprep 8 质粒提取试剂盒 (Qiagen 公司, Chatsworth, CA) 从克隆 1 中提取出质粒 DNA。质粒 DNA 的限制性酶切消化显示克隆 1 含有大约 13Kb 的插入物。

利用 2B/291 作探针, 进行克隆 1 DNA 和芽孢杆菌 JP170 染色体 DNA 的 Southern 杂交分析。特别地, 7 μ g 芽孢杆菌 JP170 染色体 DNA 和 16 ng 克隆 1 质粒 DNA 用 **EcoRI** 和 **HindIII** 消化, 消化产物经 1% 琼脂糖凝胶电泳。DNA 依生产商用法说明经毛细管法转移至 Nytran Plus 膜 (Schleicher & Schuell 公司, Keene, NH) 上。再按照 Genius System 用法说明对膜进行探查。

Southern 杂交结果显示, 2B/291 探针可与 **EcoRI** 消化染色体 DNA 产生的 1800bp 和 1400bp 两条带以及 **EcoRI** 消化克隆 1 DNA 产生的约 2000bp 和 1800bp 两条带杂交。2B/291 探针也可与 **HindIII** 消化染色体 DNA 产生的约 2000bp 和 1800bp 两条带以及 **HindIII** 消化克隆 1 DNA 产生的大约 2000bp 的一条带杂交。这些结果表明克隆 1 不含有完整的基因, 因为只有 2000bp 条带与 2B/291 探针杂交。经克隆 1 的 **HindIII** 片段序列分析并基于与 JP4197182 所公开蛋白酶的同源性比较, 表明它含有包括蛋白酶基因 5' 末端 1200bp 的部分开放阅读框。

由于 Southern 杂交结果显示 3' 末端位于 1800bp 的 **HindIII** 片段中, 所以构建了一个新文库。芽孢杆菌 JP170 染色体 DNA 用 **HindIII** 消化, 消化产物经 1% 琼脂糖凝胶电泳。从胶上切下 1500bp - 2200bp 范围大小的片段并经 QIAquick 凝胶抽提试剂盒 (Qiagen 公司, Chatsworth, CA) 纯化。然后将这些片段连至 pUC118 的 **HindIII** 位点上。连接产物依照生产商用法说明转化大肠杆菌 DH5 α (Gibico

BRL 公司, Gaithersburg, MD), 并用以上所述 2B/291 探针筛选转化子。筛选了 3200 个转化子, 确认出 5 个阳性克隆。利用 QIAprep 8 质粒提取试剂盒 (Qiagen 公司, Chatsworth, CA) 依照生产商用法说明分别从 5 个转化子中提取出质粒 DNA, 用 HindIII 消化。将所得限制性酶切片段经凝胶电泳与克隆 1 质粒 DNA 限制性酶切片段进行比较。5 个克隆皆含有与前述所克隆的芽孢杆菌 JP170 蛋白酶基因 5' 末端大小一致的片段。

实施例 11: 经反向 PCR 分离芽孢杆菌 JP170 蛋白酶基因 3' 末端片段

通过扩增实施例 10 中经文库筛选所分离的染色体克隆 (克隆 1) 的下游区, 利用反向 PCR 分离芽孢杆菌 JP170 蛋白酶基因 3' 末端片段。染色体 DNA 的 Southern 杂交结果显示基因的 3' 末端应含在 1800bp 的 EcoRI 片段 (实施例 10) 中。通过用 EcoRI 消化芽孢杆菌 JP170 染色体 DNA 并经 1% 琼脂糖凝胶电泳, 制得经大小选择的染色体 DNA 片段。经 QIAquick 凝胶抽提试剂盒分离大约 1600bp-2000bp 范围大小的片段, 并洗脱于 30 μ l TE 中。使 EcoRI 片段在 10 μ l 含以下组分的连接反应体系中自连: 1 μ l 所选大小的 DNA, 1 \times 连接缓冲液 (Boehringer Mannheim Corporation, Indianapolis, IN), 和 1 μ l T4 DNA 连接酶 (Boehringer Mannheim Indianapolis, IN)。连接反应物温育于 14 $^{\circ}$ C。然后在 20 μ l 反应体系中用 HindIII 酶切 3 μ l 连接混合物, 使得自连的 EcoRI 片段在 PCR 引物的结合位点之间线性化。将该线性化 DNA 用作 PCR 反应的模板, 两个别异的引物 17011 和 17012 如下所示, 其序列是基于克隆 1 所含蛋白酶基因的序列。

17011: 5'-GTAGGTTTTCGGTTGCCCAACTGTAATCGC-3' (SEQ ID NO:30)

17012: 5'-GGTCCTACTAGAGATGGACGTATTAAGCCGG-3' (SEQ ID NO:31)

扩增反应使用 GeneAmp 试剂盒 (Perkin-Elmer 公司, Foster City, CA), 依照生产商的指示完成。

扩增结果产生 1700bp PCR 产物。将该 1700bp PCR 产物克隆至 TA 克隆试剂盒的 pCR2.1 载体上并用前述方法测序。将所推断的氨基酸序列与 JP4197182 所公开的已知氨基酸序列相比较, 表明克隆的反向 PCR 产物含有芽孢杆菌 JP170 蛋白酶基因的 3' 末端。

实施例 12: 芽孢杆菌 JP170 蛋白酶基因的重构

将芽孢杆菌 JP170 蛋白酶基因 5' 和 3' 末端克隆至芽孢杆菌载体 pSJ2882-MCS (图 3) 上, 以重构出芽孢杆菌 JP170 蛋白酶基因。pSJ2882-MCS 来自 pHP13 (Haima 等, 1987, 分子普通遗传学 (Molecular General Genetics) 209 : 335-342), 但含有 SfiI-NotI-侧翼的多克隆位点, 并且还含有包括来自 pUB110 的 oriT 区的 0.5 kb SstI 片段。后面的片段允许通过由 pLS20 介导的接合使质粒向枯草芽孢杆菌 A164 发生迁移 (Battisti 等, 1985, 细菌学杂志, 162; 543-550)。

利用加有新限制性酶切位点的引物从芽孢杆菌 JP170 染色体 DNA 中进行 PCR 扩增, 可分别将 5' 和 3' 片段克隆至质粒上。以下引物被用来添加 5'SmaI 位点至芽孢杆菌 JP170 蛋白酶基因片段 5' 端;

170Sma: 5'-CTCCCCCGGGGATGTGTTATAAAATTGAGAGGAG-3' (SEQ ID NO:32)

17030R: 5'-CCTCGTGAAGAGAATTGAGCAACATGG-3' (SEQ ID NO:33)

用以下引物将 3'NotI 位点添加至芽孢杆菌 JP170 蛋白酶基因片段 3' 端;

17027F: 5'-GCGATTACAGTTGGGGCAACC-3' (SEQ ID NO:34)

17035NOT: 5'-GCGGCCGCGTACTCTCATCAATTTCCCAAGC-3' (SEQ ID NO:35)

17036NOT: 5'-GCGGCCGCGTCATAAACGTTGCAATCGTGCTC-3' (SEQ ID NO:36)

扩增反应在与实施例 9 所述同样的条件下完成。

5' 末端 PCR 产物包括 ATG (包括核糖体结合位点) 上游 35bp 的新 SmaI 位点, 并延伸至其内部原有的 HindIII 位点区。将这个片段作为 SmaI- HindIII 片段克隆至 pSJ2882-MCS 的 SmaI- HindIII 位

点。扩增由 HindIII 位点开始一直至其终止密码子下游 192bp 的 3' 末端，添加一个 NotI 位点，并作为 HindIII-NotI 片段克隆至 5' 末端下游。

根据实施例 9 所述扩增条件，使用以下所列引物 37 和 38 PCR 扩增出 amyQ 启动子（被称之为 BANTM（Novo Nordisk A/S 公司，Bagsverd，丹麦）的地衣芽孢杆菌淀粉酶编码基因的启动子）：

引物 37: 5'-TTTGGCCTTAAGGGCCTGCAATCGATTGTTTGAGAAAAGAAG-3'

（划线部分分别为 SfiI 和 ClaI 位点）（SEQ ID NO: 37）

引物 38: 5'-TTTGAGCTCCATTTTCTTATACAAATTATATTTTACATATCAG-3'

（划线部分为 SstI 位点）（SEQ ID NO: 38）

根据实施例 9 所述扩增条件，从 pPL1759（图 4，即衍生于 pUB110 并含有 amyL 启动子的质粒）经 PCR 扩增出 amyL 启动子（被称之为 TERMAMYLTM（Novo Nordisk A/S 公司，Bagsverd，丹麦）的解淀粉芽孢杆菌淀粉酶编码基因的启动子）。使用引物 term1SFi 在扩增中在 5' 端添加 SfiI 位点，使用引物 2iSfi 在 3' 端添加 SacI 位点：

引物 termSFi: 5'-CCAGGCCTTAAGGGCCGCATGCGTCCTTCTTTG-3' (SEQ ID NO:39)

引物 2iSfi: 5'-CCAGAGCTCCTTTCAATGTAACATATGA-3' (SEQ ID NO:40)

然后将 amyQ 启动子（BANTM 启动子）和 amyL 启动子（TERMAMYLTM 启动子）作为 SfiI-Ecl136II（平端）片段插入重构基因上游区的 SfiI-SmaI 位点，分别产生出 p170BAN 和 p170TERM。

实施例 13: 芽孢杆菌 JP170 蛋白酶基因的序列分析

依照生产商用法说明使用 Applied Biosystems Model 377 测序仪对重构芽孢杆菌蛋白酶基因进行序列测定。

重构蛋白酶基因的 DNA 序列分析结果显示有如图 5（SEQ ID NO: 41）所示的 1923bp 的开放阅读框。如图 5（SEQ ID NO: 42）所示的推导氨基酸序列由包括 33 个氨基酸的信号肽序列和 175 个氨基酸的前原区的 641 个氨基酸组成。根据 GeneAssist 软件（PE

Applied Biosystems 公司, Foster Foster City, CA) 和 LaserGene 软件 (DNASTAR 公司, Madison, WI) 的计算, 完整的蛋白 (包括信号肽和前原区) 与 JP4197182 所公开的蛋白酶有 77% 的同一性, 并且推导的成熟蛋白与同样的蛋白酶 (图 6, SEQ ID NO: 43) 有 89% 的同一性。特别地, 它也含有 JP4197182 所公开蛋白酶中可见的 C-末端延伸区。在蛋白质数据库中与之具最大同源性的枯草杆菌蛋白酶前体, 根据 GeneAssist 两者只有 35% 的同一性 (图 6, SEQ ID NO: 44)。

实施例 14: p170BAN 和 p170TERM 转化枯草芽孢杆菌

根据 Petit et al. (1990, 见上) 的方法将质粒 p170BAN 和 p170TERM 转入枯草芽孢杆菌菌株 A164 Δ 5 感受态细胞中, 并用氯霉素抗性筛选。

将转化子涂板于含有 5 μ g/ml 氯霉素和 1% 奶粉的 TBAB 平板上, 37 $^{\circ}$ C 过夜生长以测试蛋白酶的产生。比较而言, 含有 p170BAN 或 p170TERM 的菌株产生淡晕圈, 而只含有空载体的菌株不产生晕圈。

将质粒 p170BAN 也如上述方法转化至枯草芽孢杆菌菌株 168 aprE- nprE- amyE- spoIIE::Tn917 的感受态细胞中。被命名为枯草芽孢杆菌 LC20 的转化子在 1% 奶粉-TBAB 平板上产生晕圈环带。

实施例 15: 将 pLC20 和 pLC21 整合至枯草芽孢杆菌中

为了构建整合载体 pCAsub2, 经 BclI 和 BglII 消化 pPL2419 (图 7) 切下新霉素抗性基因, 并用来自 pMI1101 (Youngman et al., 1984, 质粒 12:1-9) 的含氯霉素乙酰转移酶基因 (cat) 的 BamHI 片段替换, 得到质粒 pPL2419-cat。 (BamHI 粘性末端与 BclI 和 BglII 粘性末端相容。) 然后, 用通过下述两寡核苷酸 (SEQ ID NO: 45 和 SEQ ID NO: 46) 一起退火所产生的含有 SfiI 和 NotI 位点的新 MCS 替换 pPL2419-cat 的多克隆位点 (MCS):

5'-AGCTTGGCCTTAAGGGCCCGATATCGGATCCGCGCCGCTGCAGGTAC-3'

(下划线所示为 HindIII 和 KpnI 相容位点, 双划线部分为 SfiI 和 NotI) (SEQ ID NO: 45)

5'-CTGCAGCGGCCGCGGATCCGATATCGGGCCCTTAAGGCCA-3' (SEQ ID NO:46)

将退火后的寡核苷酸连至经 HindIII 和 KpnI 切开的 pPL2419-cat 上, 产生 p2419MCS5-cat. 然后, 取枯草芽孢杆菌菌株 A164 Δ 5 染色体 DNA 为模板, 经含有 NotI 和 KpnI (Asp718) 接头的引物 (SEQ ID NO: 47 和 SEQ ID NO: 48) PCR 扩增出 amyE (GenBank Locus BSAMYL, 接受号 V00101, J01547) 的第 942 至 1751 位核苷酸, 并插至经 NotI 和 Asp718 消化的 p2419MCS5 上, 产生整合载体 pCAsub2 (图 8), CAsub 是指用于枯草芽孢杆菌宿主 (*subtilis*) 的氯霉素 (*chloramphenicol*) 抗性的淀粉酶 (*amylase*) 同源物。

5'-GCGGCCGCGATTTCCAATGAG-3' (划线部分为添加上以形成 NotI 位点的核苷酸) (SEQ ID NO: 47)

5'-GGTACCTGCATTTGCCAGCAC-3' (划线部分为添加上以形成 Asp718 位点的核苷酸) (SEQ ID NO: 48)

将这个载体单独整合至枯草芽孢杆菌 168 中并涂板于铺有蓝色淀粉上层的平板上, 显示淀粉酶活性完全消失。

由基于 pSJ2882-MCS 的质粒 p170BAN 和 p170TERM 中分离 amyQ 启动子和 amyL 启动子芽孢杆菌 JP170 蛋白酶基因盒, 并克隆进芽孢杆菌整合载体 pCAsub2 的 SfiI-NotI 位点, 分别得到 pLC20 和 pLC21. pSJ2882-MCS 不能在芽孢杆菌内独立复制, 所以要稳定保存下来就必须整合至染色体上。它带有被部分删除的 amyE 基因以作为同源区, 通过单交换发生的整合导致完整质粒插入 amyE 位点。

依照 Petit 等 (1990, 见上) 的方法, 将 pLC20 (amyQ 启动子) 和 pLC21 (amyL 启动子) 转入枯草芽孢杆菌菌株 A164 Δ 5 和 A1630 Δ 5 感受态细胞中。整合子被命名为枯草芽孢杆菌 A164 Δ 5-B-JP170, A164 Δ 5-T-JP170, A1630 Δ 5-B-JP170, A1630 Δ 5-T-JP170,

其中 B 代表 BANTM 启动子，T 代表 TERMAMYLTM 启动子，而 JP170 是指蛋白酶基因。每种氯霉素抗性转化子皆在 1% 牛奶-TBAB 培养基上检测蛋白酶产生情况。

所有检测的转化子所产生的晕圈皆比基于 pSJ2882MCS 的多拷贝转化子更大，更明显。用如实施例 16 所述的 PCR 方法证实芽孢杆菌 JP170 蛋白酶的存在和在 amyE 位点的整合。

实施例 16: 整合筛选

利用 PCR 筛选实施例 15 中所述推断的整合子，以证实蛋白酶基因的存在并正确整合于 amyE 位点。推断整合子的基因组 DNA 由以下过程制备：将单菌落悬浮于 100 μ l H₂O 中，放置于干冰上冻 5 分钟后，再置于沸水浴中煮沸 5 分钟，然后重复以上循环 3 次。将悬浮液离心 10 分钟。取 5 μ l 上清液为模板，用下述蛋白酶引物，依实施例 9 所述方法进行 PCR 反应：

17020: 5'-GCTGCACTATTGTCTTCTG-3' (SEQ ID NO:49)

17025: 5'-CAGCAACTGCTACAATCTG-3' (SEQ ID NO:50)

利用下列引物进行整合筛选：

17037: 5'-GTGCAGGCTTACAATGTACCAG-3' (SEQ ID NO:51)

LCamyREV: 5'-GCATTTACCTGGCTCCAATGATTC-3' (SEQ ID NO:52)

如果菌株中存在蛋白酶，那么用蛋白酶引物进行扩增就能得到 665bp 的条带。如果蛋白酶基因整合在 amyE 位点，那么用整合引物扩增可得 1555bp 的条带。

PCR 产物经琼脂糖凝胶电泳显示一条 1555bp 的条带，这证实芽孢杆菌 JP170 蛋白酶基因已整合至染色体上。

实施例 17: 芽孢杆菌 JP170 蛋白酶基因表达盒的扩增

在整合的枯草芽孢杆菌菌株 A164 Δ 5-B-JP170，枯草芽孢杆菌菌株 A164 Δ 5-T-JP170，枯草芽孢杆菌菌株 A1630 Δ 5-B-JP170 和枯草芽孢杆菌菌株 A1630 Δ 5-T-JP170 中扩增 amyQ (BANTM 启动子) 启动子和 amyL 启动子 (TERMAMYLTM 启动子) 芽孢杆菌 JP170

蛋白酶基因盒。这是通过将它们各自涂板于含递增氯霉素浓度（分别为 15, 30, 45, 60, 和 80 μ g/ml）的 TBAB 平板上完成的。

通过点斑于含 1%牛奶的各种氯霉素浓度 TBAB 平板上以证实扩增后蛋白酶整合的稳定性。晕圈的产生显示为 100%的稳定性。数小时以后，扩增菌株产生的晕圈在大小上即与过夜生长的未扩增菌株产生的晕圈相当。

实施例 18: 确定拷贝数

进行 Southern 杂交，以估计扩增和未扩增形式的枯草芽孢杆菌菌株 A164 Δ 5-T-JP170 和枯草芽孢杆菌菌株 A1630 Δ 5-B-JP170 中芽孢杆菌 JP170 蛋白酶基因表达盒的拷贝数。依照生产商用法说明，利用 Qiagen Genomic DNA 手册（Qiagen 公司，Chatsworth, CA）所述的细菌 DNA 分离操作方法从菌株中制备出基因组 DNA，经 HindIII 酶切，再 0.8%琼脂糖凝胶电泳后，使用 PosiBlot Pressure Blotter 和 Pressure Control Station（Stratagene 公司，La Jolla, CA）系统转膜，并根据生产商用法说明用探针 1/291（实施例 9）以及 DIG 系统杂交和检测试剂盒（Boehringer Mannheim 公司，Indianapolis, IN）杂交检测。依照生产商用法说明，利用 Storm Imaging System Model 860 (Molecular Dynamics 公司，Sunnyvale, CA)系统估计出在每种菌株里基因表达盒数至少扩增了 4 倍。

已扩增枯草芽孢杆菌菌株 A164 Δ 5-T-JP170 的 Southern 杂交显示 amyL 启动子（TERMAMYLTTM 启动子）芽孢杆菌 JP170 蛋白酶基因盒中有 300bp 的删除。然而，依照生产商用法说明，利用 Novex 14% Tris-甘氨酸预制备 1.0mm \times 15 孔胶和 Novex DryEase 微型凝胶电泳干燥系统(Novel Experimental Technology 公司, San Diego, CA)进行的 SDS-PAGE 分析显示，枯草芽孢杆菌 JP170 基因的表达不受该删除的影响。

利用一系列 PCR 反应，确定删除发生在芽孢杆菌 JP170 蛋白酶基因的 5'端并包括 amyL 启动子。PCR 反应是利用上述数条引物和下列引物完成的：

17021: 5'-CCAATAGTAGAAGGACTG-3' (SEQ ID NO:53)

RB1701: 5'-CTTCAGATTGGAAAGCGAGCGGACGGAATCATTGATC-3' (SEQ ID NO:54)

RB1702: 5'-CTCAGCTTGAAGAAGTGA-3' (SEQ ID NO:55)

RB1703: 5'-GAAGCAGAGAGGCTATTG-3' (SEQ ID NO:56)

RB1704: 5'-GAAAATATAGGGAAAATGT-3' (SEQ ID NO:57)

PCR 反应是利用下列引物对完成的: 17037/17036Not, Term1Sfi/RB1701, RB1702/17021, RB1703/17021, RB1704/17021, 17036Not/Term1Sfi, 17020/17025, 170Sma/17021, M13-48Rev./17021, 并且反应体系中含有 5 μ g 浓度为 40 μ g/ml 的模板 DNA, 含有 15mM MgCl₂ 的 2.5 μ l 10 \times PCR 缓冲液 (Perkin-Elmer 公司, Foster City, CA), 1 μ l 10mM MgCl₂, 5 μ l 1mM dNTP 混合物, 2.5 μ l 浓度为 5 pmol/ μ l 的每种引物对, 0.125 μ l 浓度为 5U/ μ l 的 AmpliTaq Gold 聚合酶 (Perkin-Elmer 公司, Foster City, CA), 和 6.375 μ l 去离子水。反应物温育于 Stratagene Robocycler 40 热循环仪上, 程序设置为先 96 $^{\circ}$ C 10 分钟为一个循环, 再依 96 $^{\circ}$ C 1 分钟, 55 $^{\circ}$ C 1 分钟, 和 72 $^{\circ}$ C 1 分钟为一个循环进行 30 个循环, 最后再 72 $^{\circ}$ C 5 分钟一个循环。

因为在已扩增枯草芽孢杆菌菌株 A164 Δ 5-T-JP170 中不存在 amyL 启动子, 因此存在于 amyL 启动子上游的 pUC19 序列 (lacZ 启动子) 可能用作芽孢杆菌 JP170 基因的驱动启动子。

通过如实施例 17 所述将枯草芽孢杆菌菌株 A164 Δ 5-T-JP170 涂板于具递增氯霉素浓度的平板上实现重扩增, 以获得启动子未删除型蛋白酶盒。枯草芽孢杆菌 A164 Δ 5-T-JP170 基因组 DNA 由以下过程制备: 将单菌落悬浮于 100 μ l 去离子水中, 置于沸水浴中煮沸 5 分钟, 再冻 5 分钟, 然后重复以上循环 3 次。悬浮液离心 10 分钟。取 5 μ l 上清液为模板, 用引物对 Term1Sfi/17021 依以上所述方法进行 PCR 反应。在氯霉素浓度为 20 μ g/ml 时, 显示新扩增的基因

形式中存在删除。

为了获得启动子未删除型蛋白酶盒，完成了用 pLC21 重转化枯草芽孢杆菌菌株 A164 Δ 5。用引物对 M13-48Reverse/17021 依以上所述方法进行 PCR 反应，显示这个未扩增菌株为未删除型。通过如实施例 17 所述分别涂板于氯霉素浓度增加的平板上实现这个菌株的扩增。利用引物对 M13-48Reverse/17021 进行 PCR 反应，显示扩增形式（氯霉素浓度增至 40 μ g/ml）为未删除型。然而，未删除型扩增形式很难生长，并且与含 amyL 删除的扩增菌株相比较，它在 1% 牛奶-TBAB 平板上只产生非常小的晕圈。

利用如枯草芽孢杆菌 A164 Δ 5-T-JP170 同样的方法完成枯草芽孢杆菌 A1630 Δ 5-B-JP170 的 Southern 杂交，显示在启动子/蛋白酶盒中并不存在任何删除。

实施例 19: 芽孢杆菌 JP170 蛋白酶在摇瓶中的表达

枯草芽孢杆菌菌株 A164 Δ 5-B-JP170，枯草芽孢杆菌菌株 A164 Δ 5-T-JP170，枯草芽孢杆菌菌株 A1630 Δ 5-B-JP170 和枯草芽孢杆菌菌株 A1630 Δ 5-T-JP170 分别接种在含有 50 ml 由 10%蔗糖、4%大豆粉、0.42%无水磷酸氢二钠和 0.5%碳酸钙组成并且每毫升添加 5 μ g 氯霉素的 PS-1 培养液的摇瓶中，于 37 $^{\circ}$ C，转速 250 rpm 培养 5 天。此外，含有整合载体的枯草芽孢杆菌 A164 Δ 5::pCAsub2 用作负对照。

通过在每次分析的开始和末尾如实施例 18 所述用酪蛋白涂板法确证蛋白酶整合的稳定性。每次实例中，可见过夜后产生大的晕圈（晕圈可在数小时内看到），这表明整合是 100%稳定的。

如实施例 18 所述，利用 Novex 预制备凝胶进行 SDS-PAGE 分析，以确定所有分析中的表达水平。这四种菌株相互比较时，观察到与枯草芽孢杆菌菌株 A164 Δ 5-B-JP170 相比，枯草芽孢杆菌菌株 A164 Δ 5-T-JP170 具更高的表达水平。枯草芽孢杆菌菌株 A1630 Δ 5 则正好相反，与枯草芽孢杆菌菌株 A1630 Δ 5-T-JP170 相比，枯草芽孢杆菌菌株 A1630 Δ 5-B-JP170 具更高的表达水平。负对照不产生可检测

水平的蛋白酶。

实施例 20: 芽孢杆菌 JP170 蛋白酶与 SAVINASE™ 的比较

通过洗涤实验完成对芽孢杆菌 JP170 蛋白酶 (SP444) 与 SAVINASE™ 的比较。如 WO88/01293 所述方法获得芽孢杆菌 JP170 蛋白酶。SAVINASE™ 来自 Novo Nordisk A/S 公司, Bagsverd, 丹麦。

洗涤实验条件列举于表 2。

表 2

	模型蛋白酶	Koso Top
	洗涤剂	洗涤剂
洗涤剂剂量	3g/l	0.7g/l
pH	9.5	10.5
洗涤时间	15 分钟	10 分钟
温度	15°C	20°C
水硬度	5.6 dH	2.8 dH
	~ 1 mM Ca ²⁺ /Mg ²⁺	~ 0.5 mM Ca ²⁺ /Mg ²⁺
酶浓度	0, 3, 6, 9, 12, 15, 30, 60, 90nM	
测试方法	微洗(Miniwash)	
样品/体积	5 个样品 (2.5cm) /50ml	
测试材料	棉中的草(水中漂洗)	

Koso Top (Lion Corp., 东京, 日本)是一种商业洗涤剂, 因此在洗涤实验前对洗涤剂中的蛋白酶进行失活。在微波炉中将含 10g 洗涤剂的 100ml 去离子水加热至 85°C 使蛋白酶失活。

模型洗涤剂由 25%STP (Na₅P₃O₁₀), 25%Na₂SO₄, 10%Na₂CO₃, 20%LAS (Nansa 80S), 5% NI (Dobanol 25-7), 0.5% Na₂Si₂O₅, 0.5%羧甲基纤维素 (CMC), 和 9.5%水组成。将 pH 调至 9.5。

利用 Elrepho 2000 光度计 (无紫外) 在 460nm 进行实验材料的 R 值 (Remission, 缓解值) 测量。测量结果与表达的关系用下式表

示:

$$\Delta R = \{[(a)(\Delta R_{\max})(c)]/[\Delta R_{\max} + (a)(c)]\} + b$$

改进因子 (IF) 可用初始斜率计算: $IF = a/a_{\text{ref}}$ 。 ΔR 为缓解单位表示的酶洗效果; a 是实验点曲线的初始斜率 ($c \rightarrow 0$); a_{ref} 是参考酶类的初始斜率; b 是实验点曲线与 y 轴的交点值; c 是单位为纳摩尔活性酶类每升的酶浓度, ΔR_{\max} 是缓解单位表示的酶洗效果的理论最大值 ($C \rightarrow \infty$)。

如表 3 所示, 洗涤实验结果表明, 在模型洗涤剂中 JP170 蛋白酶的 IF 为 6.2, 相对应地 SAVINASE™ 值是 1.0。在 Koso Top 洗涤剂中 JP170 蛋白酶的 IF 也有 4.6, 相对应地 SAVINASE™ 值是 1.0。

表 3

蛋白酶	浓度	改进因子	
		模型洗涤剂	Koso Top 洗涤剂
SAVINASE™	$8.1 \times 10^{-4}M$	1.0	1.0
JP170 (SP144)	$3.77 \times 10^{-5}M$	6.2	4.6

如图 9 所示经模型洗涤剂洗涤的结果显示, 应用于从棉中除去草类杂质时的效果, JP170 蛋白酶 (SP144) 显著优于 SAVOINASE™。

如图 10 所示经 Koso Top 洗涤剂洗涤的结果显示, 应用于从棉中除去草类杂质时的效果, JP170 蛋白酶 (SP144) 显著优于 SAVINASE™。

生物材料保藏

已根据布达佩斯协议将以下生物材料保藏于农业研究机构保藏中心, 北方地区研究中心, 1815 University Street, Peoria, Illinois,

61604, 保藏号如下:

保藏物

保藏号

保藏日

枯草芽孢杆菌 LC20(p170BAN)

NRRL B-21680

1997年4月4日

序列表

(1) 一般资料:

(i) 申请人: Sloma, Alan

Lynne, Christianson

(ii) 发明名称: 蛋白酶活性多肽的编码核酸

(iii) 序列数: 57

(iv) 通讯地址:

(A) 收件人: Novo Nordisk of North America

(B) 街道: 405 Lexington Avenue

(C) 城市: New York

(D) 州: NY

(E) 国家: 美国

(F) ZIP: 10174

(v) 计算机可读形式:

(A) 介质类型: 软盘

(B) 计算机: IBM 可兼容机

(C) 操作系统: DOS

(D) 软件: FastSEQ, Windows Version 2.0

(vi) 目前的申请资料:

(A) 申请号: 待定

(B) 申请日: 1998年6月12日

(C) 分类号:

(viii) 代理人/代理资料:

(A) 姓名: Starnes, Robert L.

(B) 登记号: 41,324

(C) 资料/文档号: 5251,200-US

(ix) 通讯资料:

(A) 电话: 212-867-0123

(B) 传真: 212-878-9655

(C) 电传:

(2) SEQ ID NO: 1 的资料:

(i) 序列特征:

(A) 长度: 22 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(xi) 序列描述: SEQ ID NO: 1:

GAGCTCACAG AGATACGTGG GC

22

(2) SEQ ID NO: 2 的资料:

(i) 序列特征:

(A) 长度: 23 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(xi) 序列描述: SEQ ID NO: 2:

GGATCCACAC CAAGTCTGTT CAT

23

(2) SEQ ID NO: 3 的资料:

(i) 序列特征:

(A) 长度: 21 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(xi) 序列描述: SEQ ID NO: 3:

GGATCCGCTG GACTCCGGCT G

21

(2) SEQ ID NO: 4 的资料:**(i) 序列特征:****(A) 长度: 22 个碱基对****(B) 类型: 核酸****(C) 链型: 单链****(D) 拓扑结构: 线性****(xi) 序列描述: SEQ ID NO: 4:**

AAGCTTATCT CATCCATGGA AA

22

(2) SEQ ID NO: 5 的资料:**(i) 序列特征:****(A) 长度: 20 个碱基对****(B) 类型: 核酸****(C) 链型: 单链****(D) 拓扑结构: 线性****(xi) 序列描述: SEQ ID NO: 5:**

AAGCTTAGGC ATTACAGATC

20

(2) SEQ ID NO: 6 的资料:**(i) 序列特征:****(A) 长度: 33 个碱基对****(B) 类型: 核酸****(C) 链型: 单链****(D) 拓扑结构: 线性****(xi) 序列描述: SEQ ID NO: 6:**

CGGATCTCCG TCATTTTCCA GCCCGATGCA GCC

33

(2) SEQ ID NO: 7 的资料:**(i) 序列特征:**

- (A) 长度: 33 个碱基对
 - (B) 类型: 核酸
 - (C) 链型: 单链
 - (D) 拓扑结构: 线性
- (xi) 序列描述: SEQ ID NO: 7:

GGCTGCATCG GGCTGGAAAA TGACGGAGAT CCG

33

(2) SEQ ID NO: 8 的资料:

- (i) 序列特征:
- (A) 长度: 18 个碱基对
 - (B) 类型: 核酸
 - (C) 链型: 单链
 - (D) 拓扑结构: 线性
- (xi) 序列描述: SEQ ID NO: 8:

GATCACATCT TTCGGTGG

18

(2) SEQ ID NO: 9 的资料:

- (i) 序列特征:
- (A) 长度: 20 个碱基对
 - (B) 类型: 核酸
 - (C) 链型: 单链
 - (D) 拓扑结构: 线性
- (xi) 序列描述: SEQ ID NO: 9:

CGTTTATGAG TTTATCAATC

20

(2) SEQ ID NO: 10 的资料:

- (i) 序列特征:
- (A) 长度: 20 个碱基对
 - (B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(xi) 序列描述: SEQ ID NO: 10:

AGACTTCCCA GTTTGCAGGT

20

(2) SEQ ID NO: 11 的资料:

(i) 序列特征:

(A) 长度: 35 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(xi) 序列描述: SEQ ID NO: 11:

CAAACCTGGGA AGTCTCGACG GTTCATTCTT CTCTC

35

(2) SEQ ID NO: 12 的资料:

(i) 序列特征:

(A) 长度: 20 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(xi) 序列描述: SEQ ID NO: 12:

TCCAACAGCA TTCCAGGCTG

20

(2) SEQ ID NO: 13 的资料:

(i) 序列特征:

(A) 长度: 29 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(xi) 序列描述: SEQ ID NO: 13:

GCGAATTCTA CCTAAATAGA GATAAAATC

29

(2) SEQ ID NO: 14 的资料:**(i) 序列特征:****(A) 长度: 36 个碱基对****(B) 类型: 核酸****(C) 链型: 单链****(D) 拓扑结构: 线性****(xi) 序列描述: SEQ ID NO: 14:**

GTTTACCGCA CCTACGTCGA CCCTGTGTAG CCTTGA

36

(2) SEQ ID NO: 15 的资料:**(i) 序列特征:****(A) 长度: 36 个碱基对****(B) 类型: 核酸****(C) 链型: 单链****(D) 拓扑结构: 线性****(xi) 序列描述: SEQ ID NO: 15:**

TCAAGGCTAC ACAGGGTCGA CGTAGGTGCG GTAAAC

36

(2) SEQ ID NO: 16 的资料:**(i) 序列特征:****(A) 长度: 29 个碱基对****(B) 类型: 核酸****(C) 链型: 单链****(D) 拓扑结构: 线性****(xi) 序列描述: SEQ ID NO: 16:**

GCAAGCTTGA CAGAGAACAG AGAAGCCAG

29

(2) SEQ ID NO: 17 的资料:**(i) 序列特征:****(A) 长度: 28 个碱基对****(B) 类型: 核酸****(C) 链型: 单链****(D) 拓扑结构: 线性****(xi) 序列描述: SEQ ID NO: 17:**

CGTCGACGCC TTTGCGGTAG TGGTGCTT

28

(2) SEQ ID NO: 18 的资料:**(i) 序列特征:****(A) 长度: 36 个碱基对****(B) 类型: 核酸****(C) 链型: 单链****(D) 拓扑结构: 线性****(xi) 序列描述: SEQ ID NO: 18:**

CGCGGCCGCEA GGCCCTTAAG GCCAGAACCA AATGAA

36

(2) SEQ ID NO: 19 的资料:**(i) 序列特征:****(A) 长度: 34 个碱基对****(B) 类型: 核酸****(C) 链型: 单链****(D) 拓扑结构: 线性****(xi) 序列描述: SEQ ID NO: 19:**

TGGCCTTAAG GGCCTGCGGC CGCGATTTCC AATG

34

(2) SEQ ID NO: 20 的资料:

(i) 序列特征:**(A) 长度: 28 个碱基对****(B) 类型: 核酸****(C) 链型: 单链****(D) 拓扑结构: 线性****(xi) 序列描述: SEQ ID NO: 20:**

GAAGCTTCTT CATCATCATT GGCATACG

28

(2) SEQ ID NO: 21 的资料:**(i) 序列特征:****(A) 长度: 19 个碱基对****(B) 类型: 核酸****(C) 链型: 单链****(D) 拓扑结构: 线性****(xi) 序列描述: SEQ ID NO: 21:**

AAGCTTTGAA TGGGTGTGG

19

(2) SEQ ID NO: 22 的资料:**(i) 序列特征:****(A) 长度: 36 个碱基对****(B) 类型: 核酸****(C) 链型: 单链****(D) 拓扑结构: 线性****(xi) 序列描述: SEQ ID NO: 22:**

CCGCTTGTTT TTTTCATCCCC TGAAACAACCT GTACCG

36

(2) SEQ ID NO: 23 的资料:**(i) 序列特征:****(A) 长度: 34 个碱基对**

- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑结构: 线性

(xi) 序列描述: SEQ ID NO: 23:

CAGTTGTTTC AGGGGATGAA AGAACAAGCG GCTG

34

(2) SEQ ID NO: 24 的资料:

(i) 序列特征:

- (A) 长度: 18 个碱基对
- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑结构: 线性

(xi) 序列描述: SEQ ID NO: 24:

CTGACATGAG GCACTGAC

18

(2) SEQ ID NO:25 的资料:

(i) 序列特征:

- (A) 长度: 38 个氨基酸
- (B) 类型: 氨基酸
- (C) 链型: 单链
- (D) 拓扑结构: 线性

(xi) 序列描述: SEQ ID NO:25:

```

Asn Asp Val Ala Arg Gly Ile Val Lys Ala Asp Val Ala Gln Asn Asn
 1           5           10           15
Phe Gly Leu Tyr Gly Gln Gly Gln Ile Val Ala Asp Thr Gly Leu Asp
 20           25           30
Thr Gly Arg Asn Asp Ser
 35

```

(2) SEQ ID NO:26 的资料:

(i) 序列特征:

- (A) 长度: 23 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(xi) 序列描述: SEQ ID NO:26:

Gly	Ala	Ala	Asp	Val	Gly	Leu	Gly	Phe	Pro	Asn	Gly	Asn	Gln	Gly	Trp
1				5					10					15	
Gly	Arg	Val	Thr	Leu	Asp	Lys									
			20												

(2) SEQ ID NO:27 的资料:

(i) 序列特征:

(A) 长度: 29 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(xi) 序列描述: SEQ ID NO:27:

CCCCANCCNT GNTTNCNTT NGGNAANCC

29

(2) SEQ ID NO:28 的资料:

(i) 序列特征:

(A) 长度: 38 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(xi) 序列描述: SEQ ID NO:28:

GGNATNGTNA ANGCNGANGT NGCNCANAAN AANTTNGG

38

(2) SEQ ID NO:29 的资料:

(i) 序列特征:

(A) 长度: 38 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(xi) 序列描述: SEQ ID NO:29:

TANGGNCANG GNCANATNGT NGCNGTNGCN GANACNGG

38

(2) SEQ ID NO:30 的资料:

(i) 序列特征:

(A) 长度: 31 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(xi) 序列描述: SEQ ID NO:30:

GTAGGTTTTTC GGTTGCCCCA ACTGTAATCG C

31

(2) SEQ ID NO:31 的资料:

(i) 序列特征:

(A) 长度: 31 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(xi) 序列描述: SEQ ID NO:31:

GGTCCTACTA GAGATGGACG TATTAAGCCG G

32

(2) SEQ ID NO:32 的资料:

(i) 序列特征:

(A) 长度: 33 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(xi) 序列描述: SEQ ID NO:32:

CTCCCCGGG GATGTGTTAT AAATTGAGAG GAG

33

(2) SEQ ID NO:33 的资料:**(i) 序列特征:****(A) 长度: 27 个碱基对****(B) 类型: 核酸****(C) 链型: 单链****(D) 拓扑结构: 线性****(xi) 序列描述: SEQ ID NO:33:**

CCTCGTGAAG AGAATTGAGC AACATGG

27

(2) SEQ ID NO:34 的资料:**(i) 序列特征:****(A) 长度: 21 个碱基对****(B) 类型: 核酸****(C) 链型: 单链****(D) 拓扑结构: 线性****(xi) 序列描述: SEQ ID NO:34:**

GCGATTACAG TTGGGGCAAC C

21

(2) SEQ ID NO:35 的资料:**(i) 序列特征:****(A) 长度: 31 个碱基对****(B) 类型: 核酸****(C) 链型: 单链****(D) 拓扑结构: 线性****(xi) 序列描述: SEQ ID NO:35:**

GCGGCCGCGT ACTCTCATCA ATTTCCAAG C

31

(2) SEQ ID NO:36 的资料:**(i) 序列特征:****(A) 长度: 32 个碱基对****(B) 类型: 核酸****(C) 链型: 单链****(D) 拓扑结构: 线性****(xi) 序列描述: SEQ ID NO:36:**

GCGGCCGCGT CATAAACGTT GCAATCGTGC TC

32

(2) SEQ ID NO:37 的资料:**(i) 序列特征:****(A) 长度: 42 个碱基对****(B) 类型: 核酸****(C) 链型: 单链****(D) 拓扑结构: 线性****(xi) 序列描述: SEQ ID NO:37:**

TTTGGCCTTA AGGGCCTGCA ATCGATTGTT TGAGAAAAGA AG

42

(2) SEQ ID NO:38 的资料:**(i) 序列特征:****(A) 长度: 43 个碱基对****(B) 类型: 核酸****(C) 链型: 单链****(D) 拓扑结构: 线性****(xi) 序列描述: SEQ ID NO:38:**

TTTGAGCTCC ATTTTCTTAT ACAAATTATA TTTTACATAT CAG

43

(2) SEQ ID NO:39 的资料:

(i) 序列特征:

- (A) 长度: 33 个碱基对
- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑结构: 线性

(xi) 序列描述: SEQ ID NO:39:

CCAGGCCTTA AGGGCCGCAT GCGTCCTTCT TTG

33

(2) SEQ ID NO:40 的资料:**(i) 序列特征:**

- (A) 长度: 28 个碱基对
- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑结构: 线性

(xi) 序列描述: SEQ ID NO:40:

CCAGAGCTCC TTTCAATGTA ACATATGA

28

(2) SEQ ID NO: 41 的资料:**(i) 序列特征:**

- (A) 长度: 3003 个碱基对
- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑结构: 线性

(ii) 分子类型: 基因组**(xi) 序列描述: SEQ ID NO: 41:**

cttaggcaag ctttactcta tacagagatt acatcctcaa gccattgaag aattcgaaaa 60
 aagttattat ttaaaagagg ataggggggtt agacagtaaa ttaaattcga tttattgtct 120
 tttgatggaa tacgataaca tggaagattc tactcaatgt agaaaatggt tagaaattgg 180
 gaaatctttg ctaactagtc cagacgaatt ggtagaatat cattattatt tcaccatttt 240

tgactatgtc	ctagcagaca	atatggatga	gcttgatgtc	tatttccaag	aagtcgtttt	300
accttttttt	caacaacaag	atftaaaaga	accaattatt	aaatatgcag	agaggctcgc	360
catctatfff	gaatcttggt	ataaatataa	aaaagcaagc	tactactatt	cgttatgcta	420
ccaagaaatt	aaagaacaaa	cttttttata	ctaaggggag	ggtaatatga	aaaaaaaaact	480
gttgcttgta	gttttagttg	gaattctfff	tttagtaggt	actttggaaa	aatctattca	540
agagcctcaa	gtaattgcac	atggcgaggt	tactgcttta	aaagatgaac	atcctgagcc	600
gcttccaaat	ggttaaaaac	aataaagaac	ttctctact	ggagaggggt	ctttttttct	660
ttcatfff	tagaaaatat	tgaatggctg	ctgtagtctg	gcttgacagt	aatfffccat	720
tgggaaagta	tgagcccaaa	aagcgaatta	tgaagctatt	ttaatctgaa	ttttcccaat	780
ataaagtttt	tgtttcctgt	gataaattaa	tgatgtgta	taaattgaga	ggagttgagc	840
tatagaatga	gaaagaaagg	atcgaagagg	gtttttttat	ccgttttatc	agttgctgca	900
ctattgtctt	ctgttgcttt	aagcagtcct	tctactattg	gggcgaacaa	ttttgaattg	960
gactttaagg	ggatagagac	acttacgcta	gagaaggctg	ccaccaagca	aggaaaaacg	1020
ggaaaggcat	cttttcttgt	aaactctgaa	aatgtgaaaa	tcccaagag	tattcaaaag	1080
aaactagaag	tagttccagc	ggataacaag	ctatatatcg	ttcaatttga	cggacctatt	1140
ttagaggaaa	cgcaacttca	actagagaag	acgggagcga	aaattctcga	ttacatacca	1200
gattacgctt	atattgtcga	atatgatggg	gatgtaaagg	ccgtaactaa	cgcaattgcg	1260
catttggaat	cggttgaacc	atatttacct	ttatataaaa	tagaccgcga	attatfffcc	1320
agaggagctt	ctgaattagt	agaaacagta	gctttagata	aaaagcaaag	aagtaaagaa	1380
gtacgtttaa	gaggattgga	aaaattgcc	caatacgcga	caaataatga	tgtattatac	1440
gtaaccccaa	agcctgaata	cgaagttttg	aatgacgtgg	cccgtggcat	tgtgaaagca	1500
gacgtcgcac	aaaataactt	tggcttatat	ggacaaggac	agattgtagc	agttgctgat	1560
actgggcttg	atacaggaag	aatgacagt	tcgatgcatg	aagcattccg	cggtaaagatt	1620
accgcactat	atgcactggg	cagaacgaat	aacgccaatg	atccaaatgg	acatggaacc	1680
catgttgctg	gatctgtggt	aggaaatgct	aaaataaag	ggatggcacc	gcaagccaat	1740
ctagtctttc	aatctattat	ggatagtggg	ggagggctgg	gaggactacc	tgctaattcta	1800
caaacattat	tcagtcaagc	atatagtgct	ggagcgagaa	ttcatacgaa	ttcatggggg	1860
gctccagtaa	acggtgccct	tacgacagac	tctcgaaatg	ttgatgatta	tgtgagaaaa	1920
aatgatatga	cgattctfff	tgcggccgga	aatgagggac	caggtagcgg	tacaatcagt	1980
gcaccaggaa	cagcaaaaaa	tgcgattaca	gttggggcaa	ccgaaaacct	acgtccaagc	2040

ttcggatcctt atgcggataa tattaacccat gttgctcaat tctcttcacg aggtcctact	2100
agagatggac gtattaagcc ggacgctcgt gcaccaggta cgtatattct ctctgctaga	2160
tcatecattag ctccagattc ctccattctgg gcaaaccatg atagtaaata tgcctacatg	2220
ggtggtactt ctatggctac tccaattgta gcaggtaatg ttgcacaatt aagggagcat	2280
tttgtgaaaa atagaggggt aactcctaag ccttcctttt taaaagctgc tttaattgca	2340
ggtgctgcgg atgttgact tggctttcca aatggtaacc aaggatgggg aagagtaacg	2400
ttagataaat cctaataatgt cgcattttgtg aatgaaacga gccctttatc aacaagtcaa	2460
aaagcaacat attcgtttac ggctcaagct ggtaaaccct taaaaatata acttgtttgg	2520
tcagatgcac caggtagcac gacggcatca ctaactttag tgaatgattt agacttagta	2580
atcactgcac caaatggaac taaatacgtc ggaaatgact ttacagcacc gtatgataac	2640
aattgggatg gcagaaacaa cgtggaaaat gtgtttatca atgctcctca aagcggaacg	2700
tatacagtcg aagtgcaggc ttacaatgta ccagtaagtc cgcaaaccct ttcttttagcg	2760
attgtacatt aaaatattgg aaggaagagt tgttgatgaa tatatcagca gctctttttt	2820
tgattaagct cttttcgtaa aggttgttgc tttaagtcgg taaaaagtcg gtatttggac	2880
tttttaccag tcattttgct tgggaaattg atgagagtac tttcattact gatggaaaag	2940
agcaacgattg caacgtttat gacgggggtga tttctattta cgaaaagcaa caaagtatgc	3000
gaaa	3004

(2) SEQ ID NO:42 的资料:

(i) 序列特征:

(A) 长度: 641 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(ii) 分子类型: 蛋白质

(v) 片段类型: 内部的

(xi) 序列描述: SEQ ID NO:42:

Met	Arg	Lys	Lys	Gly	Ser	Lys	Arg	Val	Phe	Leu	Ser	Val	Leu	Ser	Val
1				5					10					15	
Ala	Ala	Leu	Leu	Ser	Ser	Val	Ala	Leu	Ser	Ser	Pro	Ser	Thr	Ile	Gly
			20					25					30		
Ala	Asn	Asn	Phe	Glu	Leu	Asp	Phe	Lys	Gly	Ile	Glu	Thr	Leu	Thr	Leu
		35					40					45			
Glu	Lys	Ala	Ala	Thr	Lys	Gln	Gly	Lys	Thr	Gly	Lys	Ala	Ser	Phe	Leu
	50					55					60				
Val	Asn	Ser	Glu	Asn	Val	Lys	Ile	Pro	Lys	Ser	Ile	Gln	Lys	Lys	Leu
65					70					75					80
Glu	Val	Val	Pro	Ala	Asp	Asn	Lys	Leu	Tyr	Ile	Val	Gln	Phe	Asp	Gly
				85					90					95	
Pro	Ile	Leu	Glu	Glu	Thr	Gln	Leu	Gln	Leu	Glu	Lys	Thr	Gly	Ala	Lys
			100					105						110	
Ile	Leu	Asp	Tyr	Ile	Pro	Asp	Tyr	Ala	Tyr	Ile	Val	Glu	Tyr	Asp	Gly
		115					120					125			
Asp	Val	Lys	Ala	Val	Thr	Asn	Ala	Ile	Ala	His	Leu	Glu	Ser	Val	Glu
	130					135					140				
Pro	Tyr	Leu	Pro	Leu	Tyr	Lys	Ile	Asp	Pro	Gln	Leu	Phe	Ser	Arg	Gly
145					150					155					160
Ala	Ser	Glu	Leu	Val	Glu	Thr	Val	Ala	Leu	Asp	Lys	Lys	Gln	Arg	Ser
				165					170					175	
Lys	Glu	Val	Arg	Leu	Arg	Gly	Leu	Glu	Gln	Ile	Ala	Gln	Tyr	Ala	Thr
			180					185					190		
Asn	Asn	Asp	Val	Leu	Tyr	Val	Thr	Pro	Lys	Pro	Glu	Tyr	Glu	Val	Leu
		195					200					205			
Asn	Asp	Val	Ala	Arg	Gly	Ile	Val	Lys	Ala	Asp	Val	Ala	Gln	Asn	Asn
		210				215					220				
Phe	Gly	Leu	Tyr	Gly	Gln	Gly	Gln	Ile	Val	Ala	Val	Ala	Asp	Thr	Gly
225					230					235					240
Leu	Asp	Thr	Gly	Arg	Asn	Asp	Ser	Ser	Met	His	Glu	Ala	Phe	Arg	Gly
				245					250					255	
Lys	Ile	Thr	Ala	Leu	Tyr	Ala	Leu	Gly	Arg	Thr	Asn	Asn	Ala	Asn	Asp
			260					265						270	
Pro	Asn	Gly	His	Gly	Thr	His	Val	Ala	Gly	Ser	Val	Leu	Gly	Asn	Ala
		275					280						285		
Thr	Asn	Lys	Gly	Met	Ala	Pro	Gln	Ala	Asn	Leu	Val	Phe	Gln	Ser	Ile
	290					295					300				
Met	Asp	Ser	Gly	Gly	Gly	Leu	Gly	Gly	Leu	Pro	Ala	Asn	Leu	Gln	Thr
305					310					315					320
Leu	Phe	Ser	Gln	Ala	Tyr	Ser	Ala	Gly	Ala	Arg	Ile	His	Thr	Asn	Ser
				325					330					335	
Trp	Gly	Ala	Pro	Val	Asn	Gly	Ala	Tyr	Thr	Thr	Asp	Ser	Arg	Asn	Val
			340					345					350		
Asp	Asp	Tyr	Val	Arg	Lys	Asn	Asp	Met	Thr	Ile	Leu	Phe	Ala	Ala	Gly
		355					360					365			
Asn	Glu	Gly	Pro	Gly	Ser	Gly	Thr	Ile	Ser	Ala	Pro	Gly	Thr	Ala	Lys
	370					375					380				
Asn	Ala	Ile	Thr	Val	Gly	Ala	Thr	Glu	Asn	Leu	Arg	Pro	Ser	Phe	Gly
385					390					395					400
Ser	Tyr	Ala	Asp	Asn	Ile	Asn	His	Val	Ala	Gln	Phe	Ser	Ser	Arg	Gly
				405					410					415	
Pro	Thr	Arg	Asp	Gly	Arg	Ile	Lys	Pro	Asp	Val	Met	Ala	Pro	Gly	Thr
			420					425					430		
Tyr	Ile	Leu	Ser	Ala	Arg	Ser	Ser	Leu	Ala	Pro	Asp	Ser	Ser	Phe	Trp
		435					440					445			
Ala	Asn	His	Asp	Ser	Lys	Tyr	Ala	Tyr	Met	Gly	Gly	Thr	Ser	Met	Ala
	450					455					460				
Thr	Pro	Ile	Val	Ala	Gly	Asn	Val	Ala	Gln	Leu	Arg	Glu	His	Phe	Val
465					470					475					480
Lys	Asn	Arg	Gly	Val	Thr	Pro	Lys	Pro	Ser	Leu	Leu	Lys	Ala	Ala	Leu
				485					490					495	
Ile	Ala	Gly	Ala	Ala	Asp	Val	Gly	Leu	Gly	Phe	Pro	Asn	Gly	Asn	Gln
			500				505						510		
Gly	Trp	Gly	Arg	Val	Thr	Leu	Asp	Lys	Ser	Leu	Asn	Val	Ala	Phe	Val
		515					520						525		

```

Asn Glu Thr Ser Pro Leu Ser Thr Ser Gln Lys Ala Thr Tyr Ser Phe
   530                               535                               540
Thr Ala Gln Ala Gly Lys Pro Leu Lys Ile Ser Leu Val Trp Ser Asp
545                               550                               555                               560
Ala Pro Gly Ser Thr Thr Ala Ser Leu Thr Leu Val Asn Asp Leu Asp
                               565                               570                               575
Leu Val Ile Thr Ala Pro Asn Gly Thr Lys Tyr Val Gly Asn Asp Phe
                               580                               585                               590
Thr Ala Pro Tyr Asp Asn Asn Trp Asp Gly Arg Asn Asn Val Glu Asn
595                               600                               605
Val Phe Ile Asn Ala Pro Gln Ser Gly Thr Tyr Thr Val Glu Val Gln
610                               615                               620
Ala Tyr Asn Val Pro Val Ser Pro Gln Thr Phe Ser Leu Ala Ile Val
625                               630                               635                               640
His

```

(2) SEQ ID NO:43 的资料:

(i) 序列特征:

(A) 长度: 635 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(xi) 序列描述: SEQ ID NO:42:

```

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

```

```

Pro Gln Ala Asn Leu Val Phe Gln Ser Ile Met Asp Ser Ser Gly Gly
 290                295                300
Leu Gly Gly Leu Pro Ser Asn Leu Asn Thr Leu Phe Ser Gln Ala Trp
305                310                315                320
Asn Ala Gly Ala Arg Ile His Thr Asn Ser Trp Gly Ala Pro Val Asn
                325                330                335
Gly Ala Tyr Thr Ala Asn Ser Arg Gln Val Asp Glu Tyr Val Arg Asn
                340                345                350
Asn Asp Met Thr Val Leu Phe Ala Ala Gly Asn Glu Gly Pro Asn Ser
355                360                365
Gly Thr Ile Ser Ala Pro Gly Thr Ala Lys Asn Ala Ile Thr Val Gly
370                375                380
Ala Thr Glu Asn Tyr Arg Pro Ser Phe Gly Ser Ile Ala Asp Asn Pro
385                390                395                400
Asn His Ile Ala Gln Phe Ser Ser Arg Gly Ala Thr Arg Asp Gly Arg
                405                410                415
Ile Lys Pro Asp Val Thr Ala Pro Gly Thr Phe Ile Leu Ser Ala Arg
                420                425                430
Ser Ser Leu Ala Pro Asp Ser Ser Phe Trp Ala Asn Tyr Asn Ser Lys
435                440                445
Tyr Ala Tyr Met Gly Gly Thr Ser Met Ala Thr Pro Ile Val Ala Gly
450                455                460
Asn Val Ala Gln Leu Arg Glu His Phe Ile Lys Asn Arg Gly Ile Thr
465                470                475                480
Pro Lys Pro Ser Leu Ile Lys Ala Ala Leu Ile Ala Gly Ala Thr Asp
                485                490                495
Val Gly Leu Gly Tyr Pro Ser Gly Asp Gln Gly Trp Gly Arg Val Thr
                500                505                510
Leu Asp Lys Ser Leu Asn Val Ala Tyr Val Asn Glu Ala Thr Ala Leu
515                520                525
Ala Thr Gly Gln Lys Ala Thr Tyr Ser Phe Gln Ala Gln Ala Gly Lys
530                535                540
Pro Leu Lys Ile Ser Leu Val Trp Thr Asp Ala Pro Gly Ser Thr Thr
545                550                555                560
Ala Ser Tyr Thr Leu Val Asn Asp Leu Asp Leu Val Ile Thr Ala Pro
                565                570                575
Asn Gly Gln Lys Tyr Val Gly Asn Asp Phe Ser Tyr Pro Tyr Asp Asn
580                585                590
Asn Trp Asp Gly Arg Asn Asn Val Glu Asn Val Phe Ile Asn Ala Pro
595                600                605
Gln Ser Gly Thr Tyr Ile Ile Glu Val Gln Ala Tyr Asn Val Pro Ser
610                615                620
Gly Pro Gln Arg Phe Ser Leu Ala Ile Val His
625                630                635

```

(2) SEQ ID NO:44 的资料:

(i) 序列特征:

(A) 长度: 418 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(xi) 序列描述: SEQ ID NO:44:

```

Met Lys Arg Ser Gly Lys Ile Phe Thr Thr Ala Met Leu Ala Val Thr
 1                5                10                15
Leu Met Met Pro Ala Ile Gly Val Ser Ala Asn Arg Gly Asn Ala Ala
                20                25                30

```

```

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

Ile Gln Asn Gly Thr Tyr Arg Val Ala Asp Phe Ser Ser Arg Gly His
  305                               310                               315                               320
Lys Thr Ala Gly Asp Tyr Val Ile Gln Lys Gly Asp Val Glu Ile Ser
  325                               330                               335
Ala Pro Gly Ala Ala Val Tyr Ser Thr Trp Phe Asp Gly Gly Tyr Ala
  340                               345                               350
Thr Ile Ser Gly Thr Ser Met Ala Ser Pro His Ala Ala Gly Leu Ala
  355                               360                               365
Ala Lys Ile Trp Ala Gln Ser Pro Ala Ala Ser Asn Val Asp Val Arg
  370                               375                               380
Gly Glu Leu Gln Thr Arg Ala Ser Val Asn Asp Ile Leu Ser Gly Asn
  385                               390                               395                               400
Ser Ala Gly Ser Gly Asp Asp Ile Ala Ser Gly Phe Gly Phe Ala Lys
  405                               410                               415
Val Gln

```

(2) SEQ ID NO:45 的资料:

(i) 序列特征:

- (A) 长度: 48 个碱基对
- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑结构: 线性

(xi) 序列描述: SEQ ID NO:45:

AGCTTGGCCT TAAGGGCCCC ATATCGGATC CGCGGCCGCT GCAGGTAC

48

(2) SEQ ID NO:46 的资料:**(i) 序列特征:****(A) 长度: 40 个碱基对****(B) 类型: 核酸****(C) 链型: 单链****(D) 拓扑结构: 线性****(xi) 序列描述: SEQ ID NO:46:**

CTGCAGCGGC CGCGGATCCG ATATCGGGCC CTTAAGGCCA

40

(2) SEQ ID NO:47 的资料:**(i) 序列特征:****(A) 长度: 21 个碱基对****(B) 类型: 核酸****(C) 链型: 单链****(D) 拓扑结构: 线性****(xi) 序列描述: SEQ ID NO:47:**

GCGGCCGCGA TTTCCAATGA G

21

(2) SEQ ID NO:48 的资料:**(i) 序列特征:****(A) 长度: 21 个碱基对****(B) 类型: 核酸****(C) 链型: 单链****(D) 拓扑结构: 线性****(xi) 序列描述: SEQ ID NO:48:**

GGTACCTGCA TTTGCCAGCA C

21

(2) SEQ ID NO:49 的资料:**(i) 序列特征:**

- (A) 长度: 19 个碱基对
- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑结构: 线性

(xi) 序列描述: SEQ ID NO:49:

GCTGCACTAT TGTCTTCTG

19

(2) SEQ ID NO:50 的资料:**(i) 序列特征:**

- (A) 长度: 19 个碱基对
- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑结构: 线性

(xi) 序列描述: SEQ ID NO:50:

CAGCAACTGC TACAATCTG

19

(2) SEQ ID NO:51 的资料:**(i) 序列特征:**

- (A) 长度: 22 个碱基对
- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑结构: 线性

(xi) 序列描述: SEQ ID NO:51:

GTGCAGGCTT ACAATGTACC AG

22

(2) SEQ ID NO:52 的资料:

(i) 序列特征:**(A) 长度: 24 个碱基对****(B) 类型: 核酸****(C) 链型: 单链****(D) 拓扑结构: 线性****(xi) 序列描述: SEQ ID NO:52:**

GCATTTACCT GGCTCCAATG ATTC

24

(2) SEQ ID NO:53 的资料:**(i) 序列特征:****(A) 长度: 18 个碱基对****(B) 类型: 核酸****(C) 链型: 单链****(D) 拓扑结构: 线性****(xi) 序列描述: SEQ ID NO:53:**

CCAATAGTAG AAGGACTG

18

(2) SEQ ID NO:54 的资料:**(i) 序列特征:****(A) 长度: 37 个碱基对****(B) 类型: 核酸****(C) 链型: 单链****(D) 拓扑结构: 线性****(xi) 序列描述: SEQ ID NO:54:**

CTTCAGATTG GAAAGCGAGC GGACGGAATC ATTGATC

37

(2) SEQ ID NO:55 的资料:**(i) 序列特征:****(A) 长度: 18 个碱基对**

- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑结构: 线性
- (xi) 序列描述: SEQ ID NO:55:

CTCAGCTTGA AGAAGTGA

18

(2) SEQ ID NO:56 的资料:

- (i) 序列特征:
 - (A) 长度: 18 个碱基对
 - (B) 类型: 核酸
 - (C) 链型: 单链
 - (D) 拓扑结构: 线性
- (xi) 序列描述: SEQ ID NO:56:

GAAGCAGAGA GGCTATTG

18

(2) SEQ ID NO:57 的资料:

- (i) 序列特征:
 - (A) 长度: 19 个碱基对
 - (B) 类型: 核酸
 - (C) 链型: 单链
 - (D) 拓扑结构: 线性
- (xi) 序列描述: SEQ ID NO:57:

GAAAAATATAG GGAAAAATGT

19

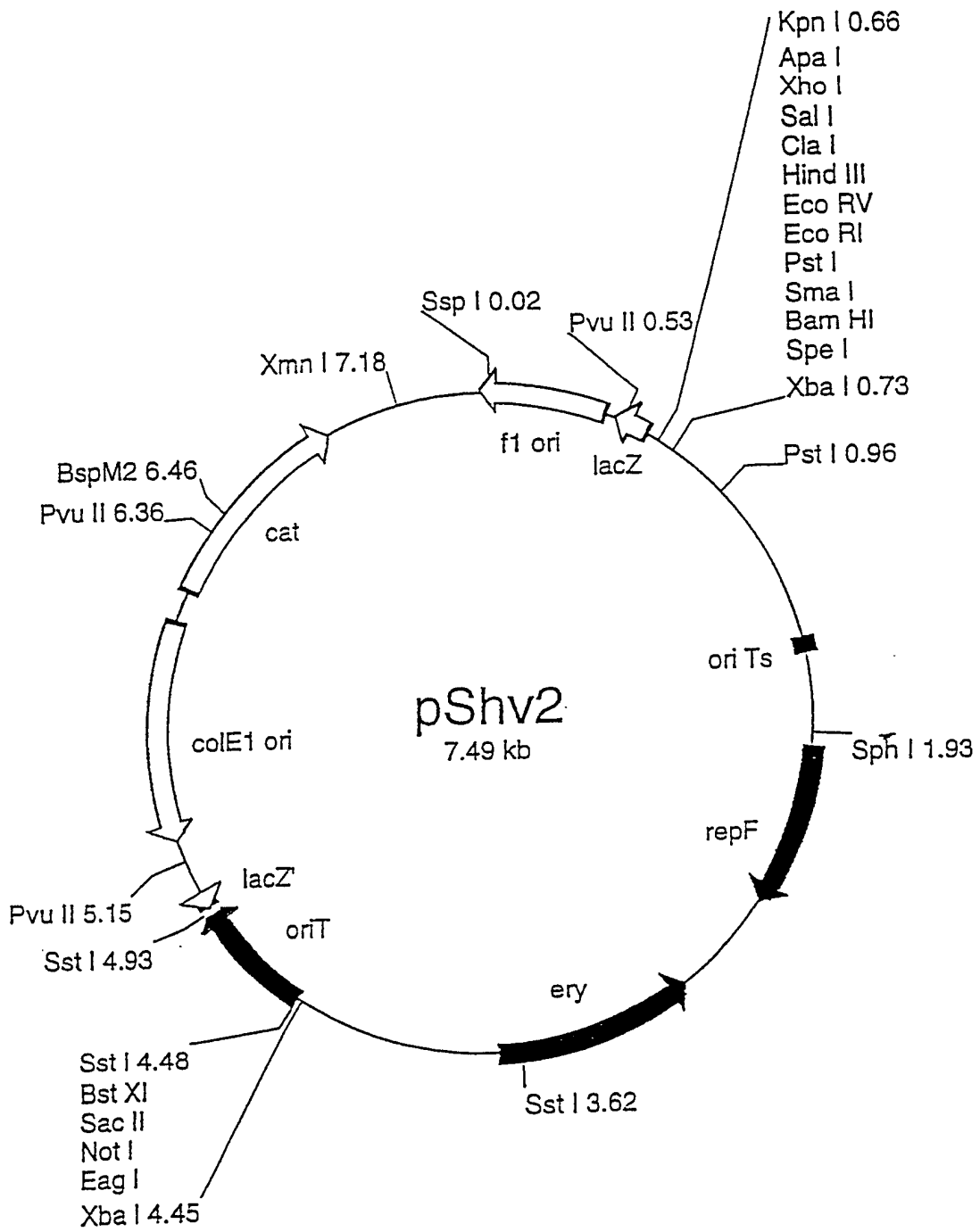


图 1

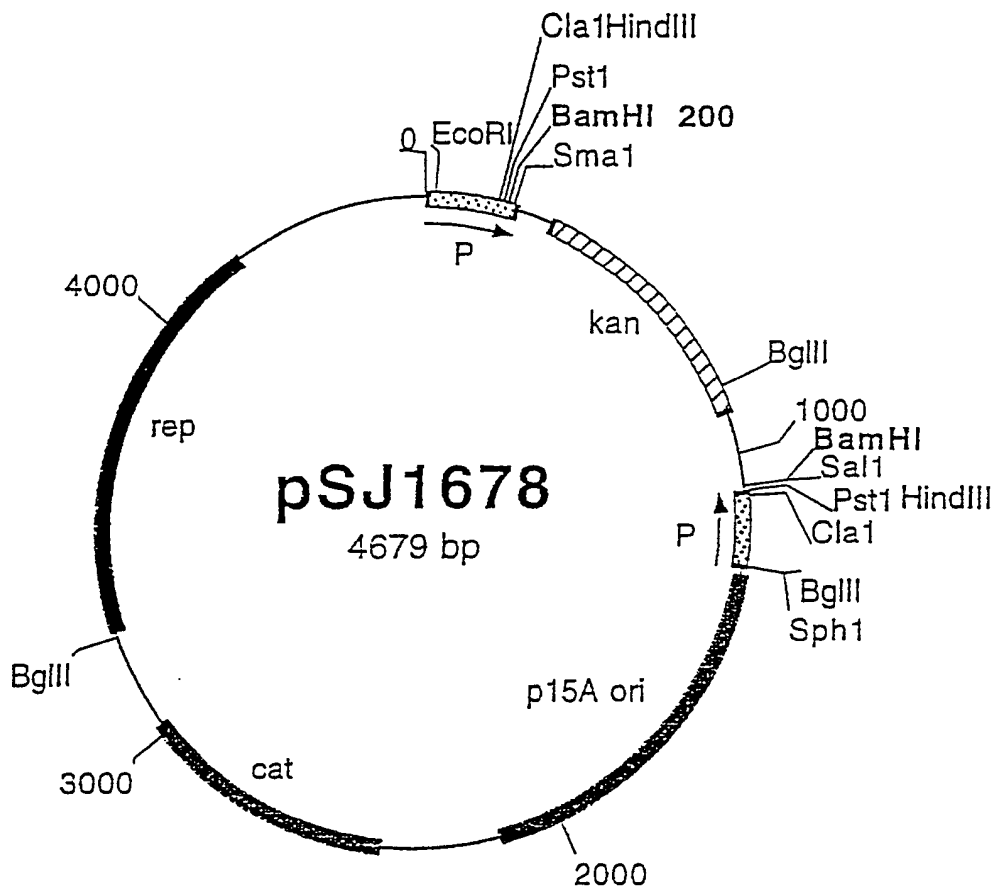


图 2

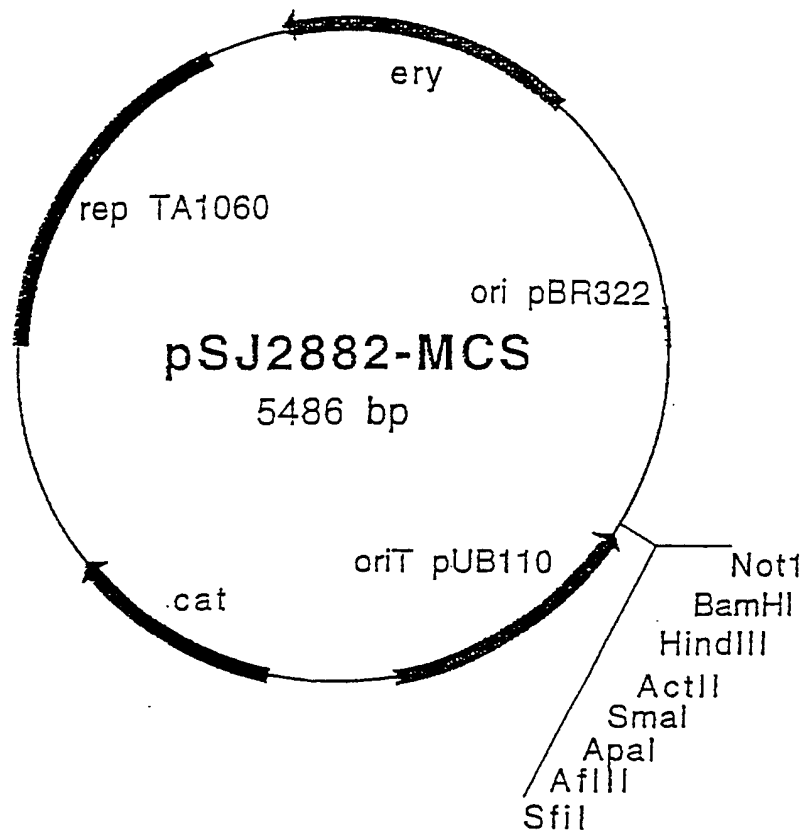


图 3

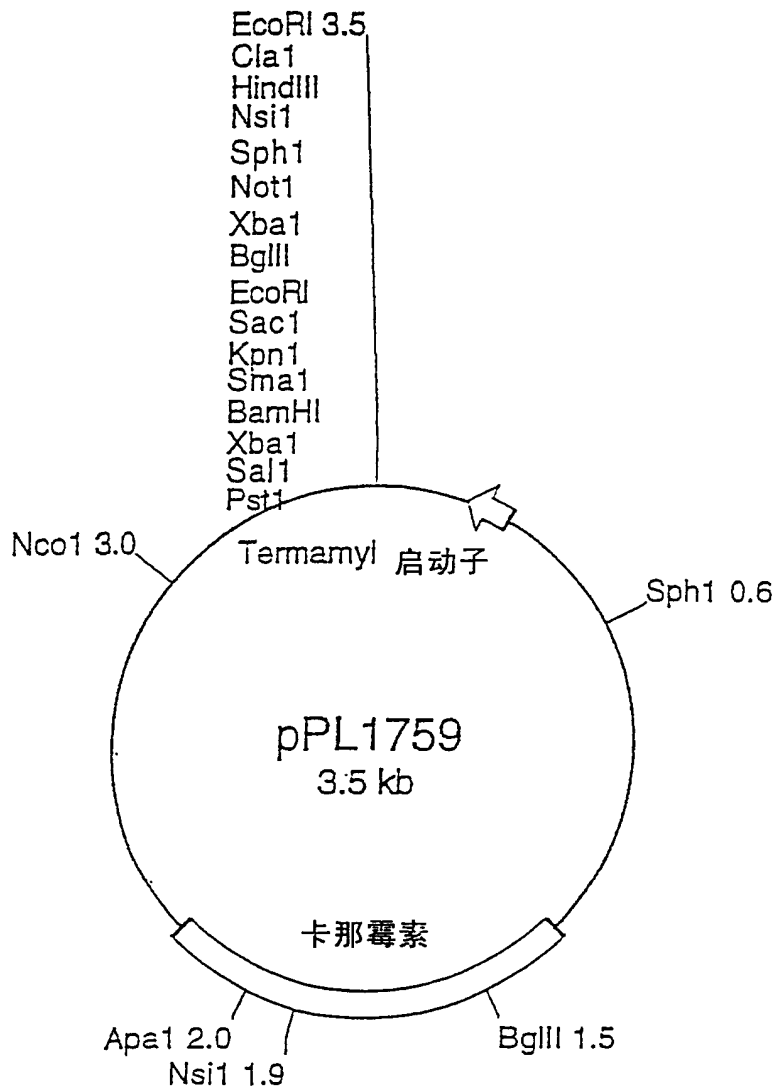


图 4

CTTAGGCAAGCTTTACTCTATACAGAGATTACATCCTCAAGCCATTGAAGAATTCGAAAAAGTTATTATTTAAA 75
 AGAGGATAGGGGGTTAGACAGTAAATTTAAATTCGATTTATTGTCTTTTGATGGAATACGATAACATGGAAGATTC 180
 TACTCAATGTAGAAAATGGTTAGAAAATTTGGGAAATCTTTGCTAACTAGTCCAGACGSAATTGGTAGAATATCATT 225
 TTATTTCCACATTTTTGACTATGTCCTAGCAGACAATATGGATGAGCTTGATGCTATTTCCAAGAAGTCGTTTT 300
 ACETTTTTTTCAACAACAAGATTTAAAAGAACCAATTATTAATATGCAGAGAGGGCTCGCCATCTATTTGAATC 375
 TTGTTATAAATACAAAAAAGCAAGCTACTACTATTCGTTATGCTACCAAGAAATTAAGAACAACACTTTTTTATA 450
 CTAAGGGGAGGGTAATATGAAAAAAAAGCTGTTGCTTGTAGTTTTAGTGGAAATCTTTTTTTAGTAGGTACTTT 525
 GGAATAATCTATCAAGAGCCTCAAGTAATTGCACATGCGGAGGTTACTGCTTAAAAGATGAACATCCTGAGCC 600
 GCTTCCAAATGGTTAAAACAATAAAGAACTTTCTCTACTGGAGAGGGTTCTTTTTTTCTTCATTTTTTTAGAA 675
 AATATTGAATGGTCGCTGCTGCTGGCTTGACAGTAATTTCCATTGGGAAAGTATGAGCCCAAAAAGCGAATTA 750
 TGAAGCTATTTAATCTGAATTTCCCAATATAAAGTTTTGTTTCTGTGATAAATTAATGATGTTATAAAT 825
 TGAGAGGAGTTGAGCTATAGAATGAGAAAGAAAGGATCGAAGAGGGTTTTTTTTATCCGTTTTATCAGTTGCTGCA 900
 M R K K G S K R V F L S V L S V A A
 CTATTGTCTTCTGTTGCTTTAAGCAGTCCCTTCTACTATTGGGGCGAACAATTTTGAATTGGACTTTAAGGGGATA 975
 L L S S V A L S S P S T I G . A N N F E L D F K G I
 GAGACACTTACGCTAGAGAAGGCTGCCACCAAGCAAGGAAAAACGGGAAAGGCATCTTTCTTAAACTCTGAA 1050
 E T L T L E K A A T K Q G K T G K A S F L V N S E
 AATGTGAAAATCCCAAAGAGTATTCAAAGAAACTAGAAGTAGTTCCAGCGGATAACAAGCTATATATCGTTCAA 1125
 N V K I P K S I Q K K L E V V P A D N K L Y I V Q
 TTTGACGGACCTATTTAGAGGAAACGCAACTTCAACTAGAGAAGACGGGAGCGGAAAAATCTCGATTACATACCA 1200
 F D G P I L E E T Q L Q L E K T G A K I L D Y I P
 GATTACGCTTATATTGCGAATATGATGGGGATGTAAGGCCGTAACCTAACGCAATTGCGCATTTGGAATCGGTT 1275
 D Y A Y I V E Y D G D V K A V T N A I A H L E S V
 GAACCATATTTACCTTTATATAAAAATAGACCCGCAATTTTCCAGAGGAGCTTCTGAATTAGTAGAAACAGTA 1350
 E P Y L P L Y K I D P Q L F S R G A S E L V E T V
 GCTTTAGATAAAAAGCAAAGAAGTAAAGAAGTACGTTAAGAGGATTGGAACAAATGCCCCAATACGCGACAAT 1425
 A L D K K Q R S K E V R L R G L E Q I A Q Y A T N
 AATGATGATTATACGTAACCCCAAGCCTGAATACGAAGTTTTGAATGACGTGGCCCTGGCATTGTGAAAGCA 1500
 N D V L Y V T P K P E Y E V L N D V A R G I V K A
 GACGTCGCACAAAATAACTTTGGCTTATATGGACAAGGACAGATTGTAGCAGTTGCTGATCTGGGCTTGATACA 1575
 D V A Q N N F G L Y G Q G Q I V A V A D T G L D T
 GGAAGAAATGACAGTTCGATGCATGAAGCATTCCGCGGTAAGATTACCGCACTATATGCACTGGGCAGAACGAAT 1650
 G R N D S S M H E A F R G K I T A L Y A L G R T N
 AACGCCAATGATCCAAATGGACATGGAACCCATGTTGCTGGATCTGTGTTAGGAAATGCTACAATAAAGGGATG 1725
 N A N D P N G H G T H Y A G S V L G N A T N K G M
 GCACCGCAAGCCAATCTAGTCTTTCAATCTATTATGGATAGTGGTGGAGGGCTGGGAGGACTACCTGCTAATCTA 1800
 A P Q A N L V F Q S I M D S G G G L G G L P A N L
 CAAACATTATTCAGTCAAGCATATAGTGTGGAGCGAGAATTCATACGAATTCATGGGGGGCTCCAGTAAACGGT 1875
 Q T L F S Q A Y S A G A R I H T N S W G A P V N G
 GCCTATACGACAGACTCTCGAAATGTTGATGATTATGTGAGAAAAATGATATGACGATTCTTTTTGCGGCCGGA 1950
 A Y T T D S R N V D D Y V R K N D M T I L F A A G
 AATGAGGGACCAGGTAGCGGTACAATCAGTGCACCAGGAACAGCAAAAAATGCGATTACAGTTGGGGCAACCGAA 2025
 N E G P G S G T I S A P G T A K N A I T V G A T E
 AACCTACGTCGAAGCTTCGGATCTTATGCGGATAATATTAACCATGTTGCTCAATTCCTTCACGAGSTCCTACT 2100
 N L R P S F G S Y A D N I N H Y A Q F S S R G P T
 AGAGATGGACGTATTAAGCCGGACGTCATGGCACCAGGTACGTATATTCTCTCTGCTAGATCATCATTAGCTCCA 2175
 R D G R I K P D V M A P G T Y I L S A R S S L A P
 GATTCCTCATTCTGGGCAAAACCATGATAGTAAATATGCCTACATGGGTGGTACTTCTATGGCTACTCCAATTGTA 2250
 D S S F W A N H D S K Y A Y M G G T S M A T P I V

图 5A

```

GCAGGTAATGTTGCACAATTAAGGGAGCATTGTTGTGAAAAATAGAGGGTAACCTCCTAAGCCTTCCCTTTTAAAA 2325
A G N V A Q L R E H F Y K N R G V T P K P S L L K
GCTGCTTTAATTGCAGGTGCTGCCGATGTTGGACTTGGCTTTCCAAATGGTAACCAAGGATGGGGAAGAGTAACG 2400
A A L I A G A A D Y G L G F P N G N G G W G R V T
TTAGATAAATCCCTAAATGTCGCATTTGTGAATGAAACGAGCCCTTATCAACAAGTCAAAAAGCAACATATTCG 2475
L D K S L N V A F V N E T S P L S T S Q K A T Y S
TTTACGGCTCAAGCTGGTAAACCCCTTAAAAATATCACTTGTGTTGGTCAGATGCACCAGGTAGCACCAGCCATCA 2550
F T A Q A G K P L K I S L V W S D A P G S T T A S
CTAACTTTAGTGAATGATTTAGACTTAGTAATCACTGCACCAAATGGAACATAAATACGTCGGAAATGACTTTACA 2625
L T L V N D L D L V I T A P N G T K Y V G N D F T
GCACCGTATGATAACAATTGGGATGGCAGAAAACAACGTGGAAAAATGTGTTTATCAATGCTCCTCAAAGCGGAACG 2700
A P Y D N N W D G R N N V E N V F I N A P G S G T
TATACAGTCGAAGTGCAGGCTTACAATGTACCAGTAAGTCCGCAAACCTTTTCTTTAGCGATTGTACATTAATAA 2775
Y T V E V Q A Y N V F V S P Q T F S L A I V H
ATTGGAAGGAAGAGTTGTTGATGAATATATCAGCAGCTCTTTTTTTGATTAAGCTCTTTTCGTAAAGGTTGTTGC 2850
TTTAAGTCGGTAAAAAGTCGGTATTTGGACTTTTACCAGTCATTTTGCTTGGGAAATGATGAGAGTACTTTCA 2925
TTACTGATGGAAAAGAGCACGATTGCAACGTTTATGACGGGGTGATTTCTATTTACGAAAAGCAACAAAGTATGC 3000
GAAA 3004

```

图 5B

图 6B

JP170 与枯草杆菌蛋白酶之间的比较

1 MRKKGSKRVFLSVLSVAALLSSVALSSPSTIGANNFELDFKGIETL/LEKAATKQG

57 KTGKASFLVNSENVKIPKSIQKKLEVVPADNKLYIVQFDGPILEETQLQLEKYGAR

1 MKRSGKIFTTAMLAVTLM

113 ILDYIPDYAYIVEYDGDVKAVTNAIAHLESVEPYLPLYKIDPQLF SRGASELVEVTV

20 MPAIGVSANRGNADGNEKFRVLVDSANQNNLKNVKEQYGVWDFAGEGFTTMNE

169 ALDKKQRSKEVRLRGLAQIAQYATNNDVLYVTPKPEYEVLNDVARGIVKADVAQNN

76 KQFNALQNNKNLTVKVPPELEIATATNKPPEALYNAMAASQSTPWGKAIYNNSNLT

225 FGLYGQQQIVAVADTGLDTPRNDSSMHEAFRGKITALYALGRTNNANDPNPEGTEV
 | - | - + + | | - | | | + + | + + - | - | - + + + - + - | - - + + - + - - - | + + + | - + | | | | | |

132 STSGGAGINIAVLDTGVNTNHPDLS . NNVEQCKDFVGTNFTDNSCTDRQGGEGTEV

281 AGSVLGN . AT . N . . KGMPQANL . VFQSIMDSGGGLG . GLPANLQTLFSQAYSAGA
 | | | + | + | - + | - + - - - | + | | + | + | - + + + - + - | + | + | - + - + - - - + - - - + | - - - + +

187 AGSALANGGTGSGVYGVAP EADLWAYKVLGDDGSGYADDIAEAI RHAGDQATALNT

331 RIETN . SWGAPVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGPGSGTISAPGTAKN
 + + - - | - | - | + - - + + + - - | + + - - - | - | + - - + - + - | + - | | | + | | - - | + | + - | + - - |

243 KVVINMSLGSSESSLITNA . . V . DYAYDKGVLI IAAAGNSGPKPGSIGYPGALVN

386 AITVGATENLRPSFGSYADNINEVAQFSSRGPTRDGRIPDVMA PGTYILSARSSL
 | + + | + | - | | - - - + - | + | - - - - + | | + | | | | | - - | - - - - - - - | + - - | - - - + | | - - - +

296 AVAVAALENTIQN . GTY R V A D F S S R G H K R . . T A G D Y V I Q K G D V E I S A . P G A

442 APDSSFWANEDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGVTPKPSLLKAAALI
 | - - - | - | - - - | + - | | - + + | | | | | + | - + | | - + | + +

343 AV . YSTW . . F D G G Y A T I S G T S M A S P E A A G L A A K I W A Q S P A A S N V D V R G E L Q T P R A S V

498 AGAADVGLGFPNGNQGWGRVTLDDKSLNVAFVNETSPLSTSQKATYSFTAQAGKPLK

396 NDILSGNSAGSGDDIASGF GFAKVQ

554 ISLVWSDAPGSTTASLTLVNDLDELVITAPNGTKYVGNDFTPYDNNWGDGRNNVENV

610 FINAPQSGTYTVEVQAYNVFVSPQTFSLAIVH

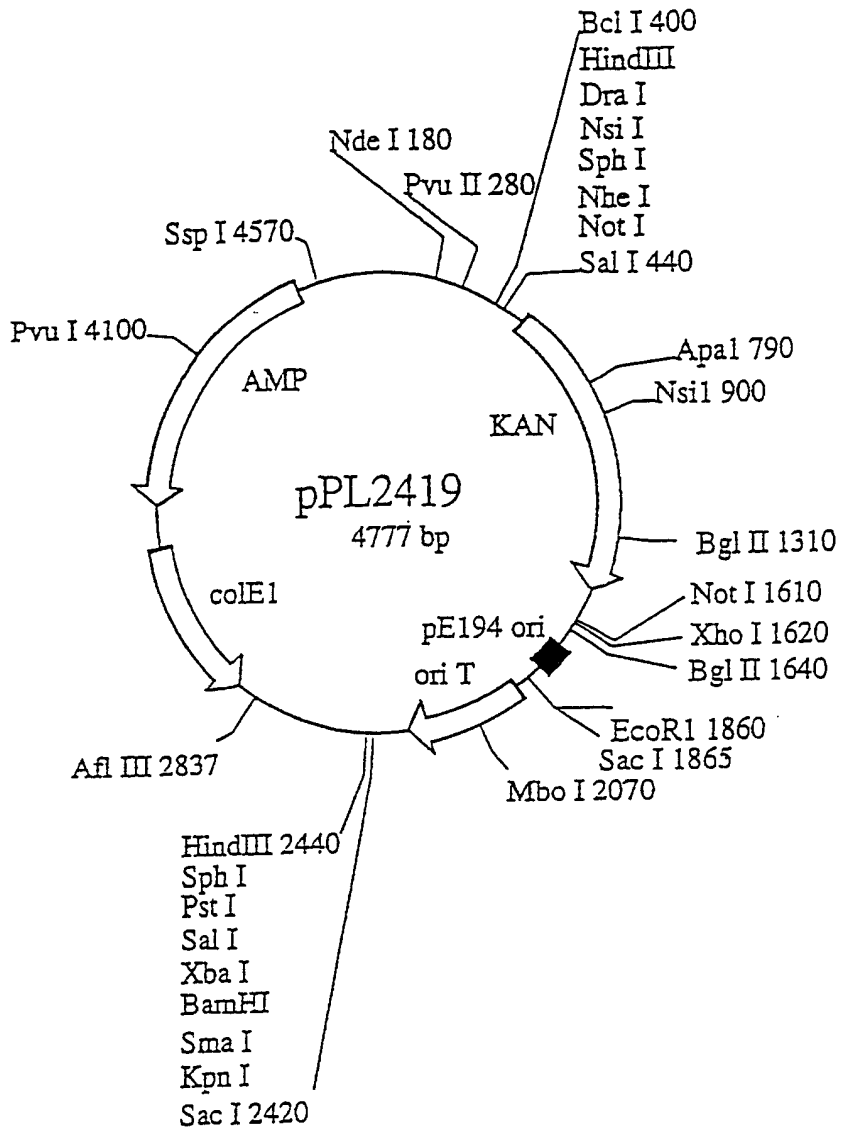


图 7

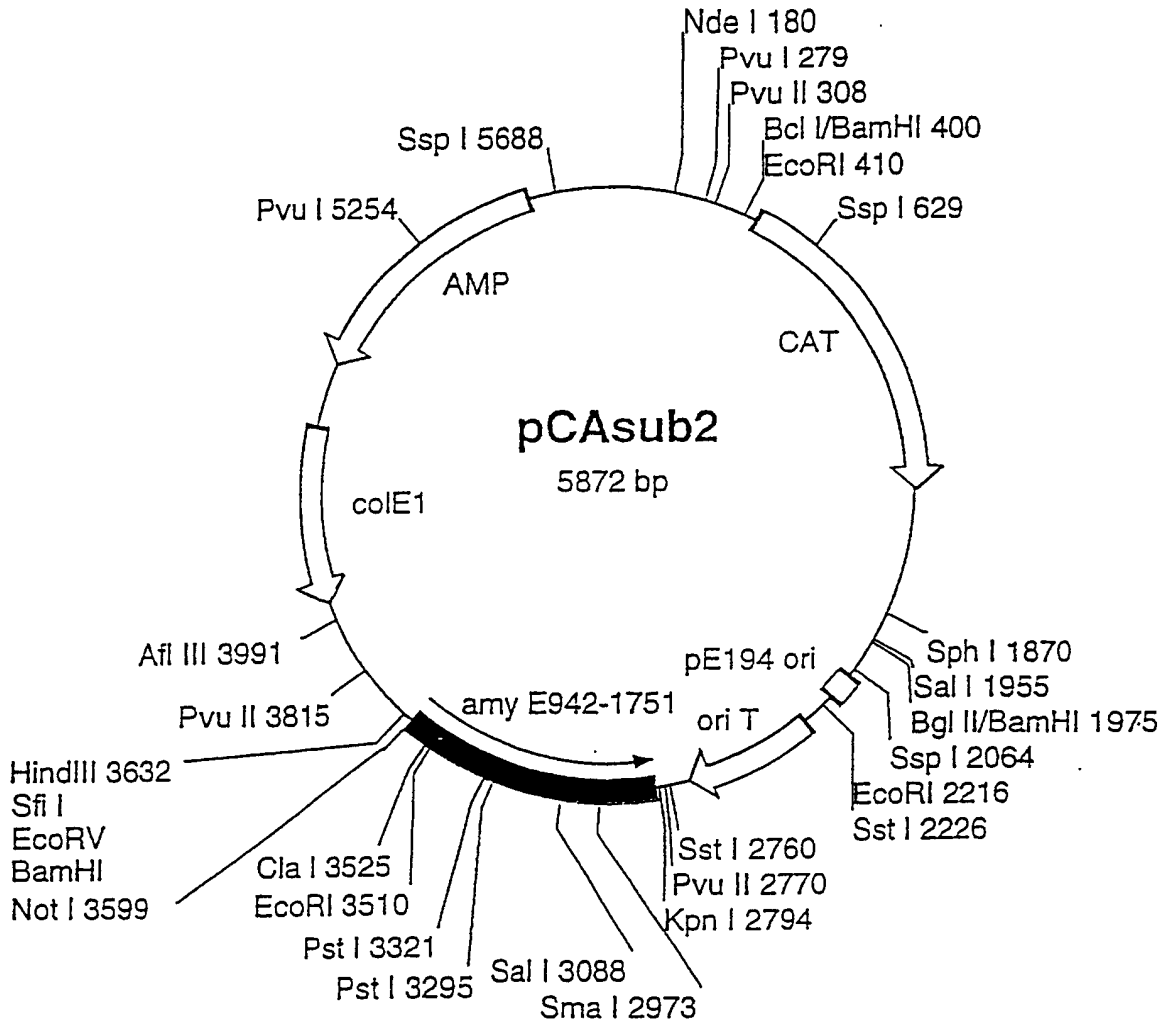


图 8

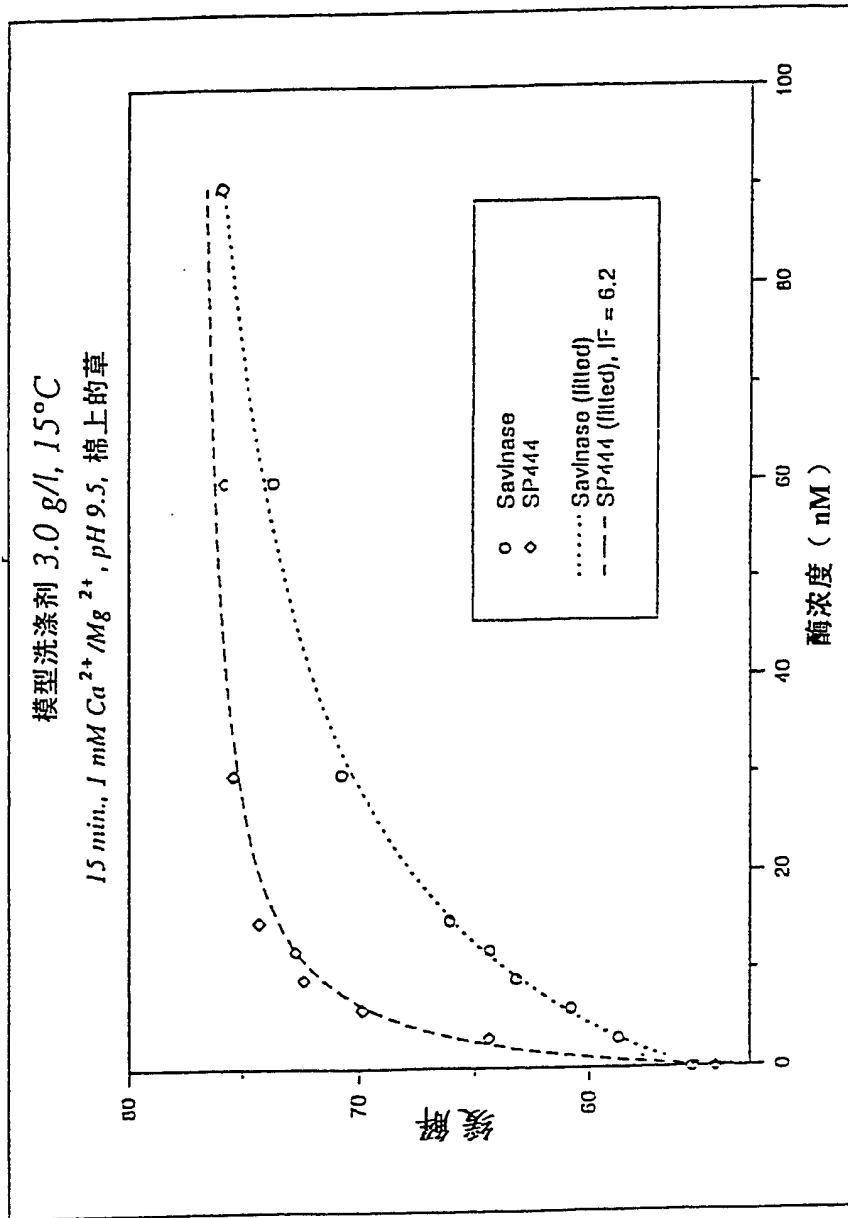


图 9

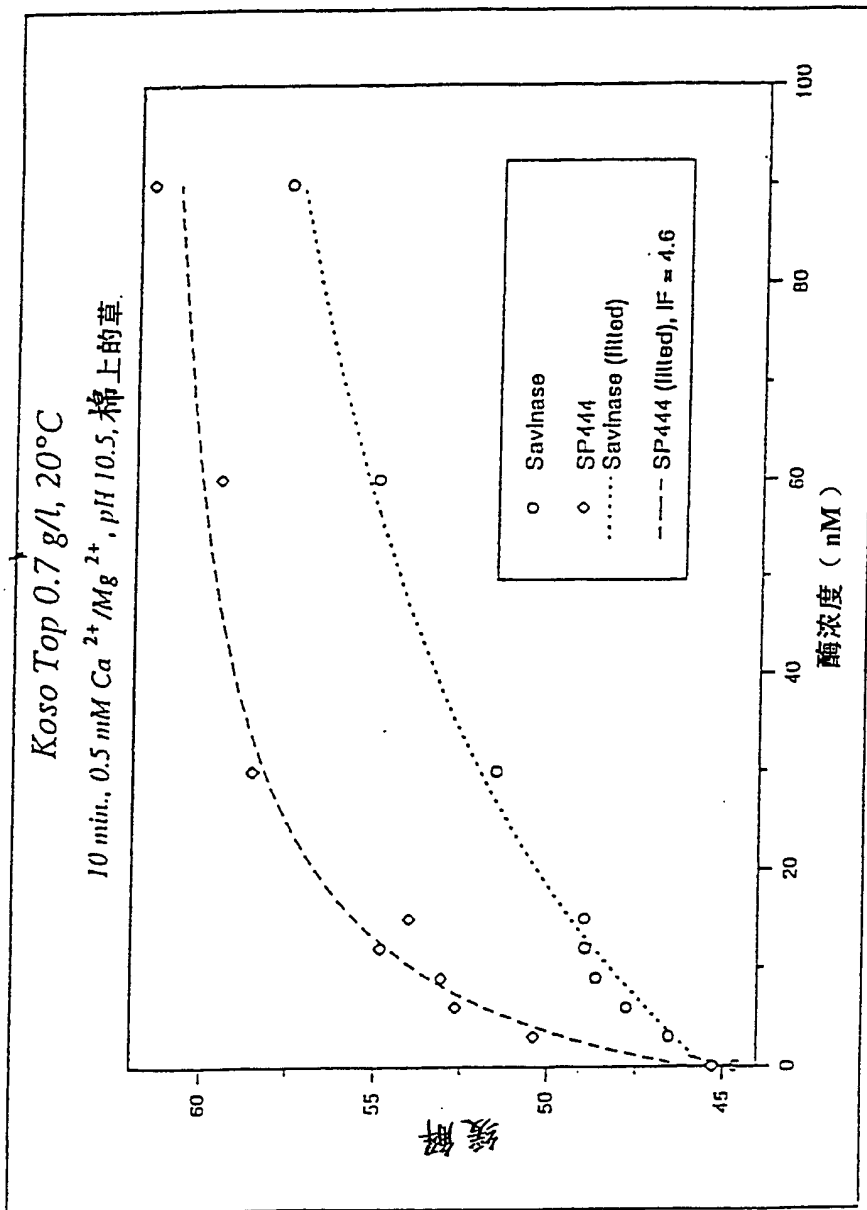


图 10