

[19]中华人民共和国国家知识产权局

[51]Int. Cl⁷

[12] 发明专利申请公开说明书

[21] 申请号 97181395.7

C12N 15/32
C12N 15/62 C07K 14/325
A01N 63/00 C12Q 1/68
A01H 5/00 C07K 16/12
G01N 33/08

[43]公开日 2000年9月27日

[11]公开号 CN 1268180A

[22]申请日 1997.11.20 [21]申请号 97181395.7

[30]优先权

[32]1996.11.20US [33]US [31]08/754,490

[32]1997.9.3 [33]US [31]08/922,505

[86]国际申请 PCT/US97/21587 1997.11.20

[87]国际公布 WO98/22595 英 1998.5.28

[85]进入国家阶段日期 1999.7.20

[71]申请人 艾可根公司

地址 美国宾夕法尼亚州

[72]发明人 T·马尔瓦

A·J·吉尔默

[74]专利代理机构 中国专利代理(香港)有限公司

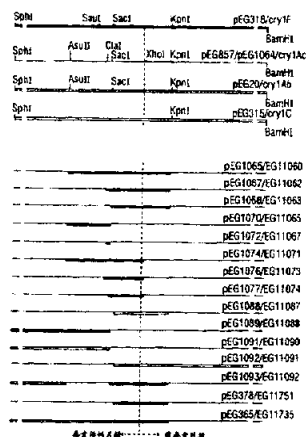
代理人 卢新华 齐曾度

权利要求书 5 页 说明书 185 页 附图页数 4 页

[54]发明名称 广谱 δ -内毒素

[57]摘要

公开的为新的人工修饰的苏云金芽孢杆菌嵌合晶体蛋白,其对鞘翅目、双翅目和鳞翅目昆虫具有增强的杀虫活性。也公开了编码这些新肽的核酸片段。制备和使用这些基因和蛋白的方法以及用于重组表达和合适宿主细胞的转化方法也被公开。表达该修饰内毒素的转化的宿主细胞和转基因植物也是本发明的内容。



ISSN 1008-4274

权 利 要 求 书

1. 分离的核酸片段, 包括编码具有 SEQ ID NO: 10、SEQ ID NO: 12、SEQ ID NO: 14、SEQ ID NO: 26、SEQ ID NO: 28、SEQ ID NO: 30 或 SEQ ID NO: 34 氨基酸序列多肽的基因。

5 2. 权利要求 1 的核酸片段, 其中所述的基因编码具有抗草地贪夜蛾、甜菜夜蛾、烟草夜蛾、玉米穗夜蛾或欧洲玉米螟杀虫活性的多肽。

3. 前面任一权利要求的核酸片段, 其中所述的核酸片段可从苏云金芽孢杆菌 NRRL B - 21579、NRRL B - 21580、NRRL B - 21581、NRRL B - 21635、NRRL B - 21636、NRRL B - 21780
10 或 NRRL B - 21781 中分离。

4. 前面任一权利要求的核酸片段, 其中所述的核酸片段特异性地与具有 SEQ ID NO: 9、SEQ ID NO: 11、SEQ ID NO: 13、SEQ ID NO: 25、SEQ ID NO: 27、SEQ ID NO: 29、或 SEQ ID NO: 33 序列或其互补序列的核酸片段杂交。

15 5. 前面任一权利要求的核酸片段, 其中所述的核酸片段包括 SEQ ID NO: 9、SEQ ID NO: 11、SEQ ID NO: 13、SEQ ID NO: 25、SEQ ID NO: 27、SEQ ID NO: 29 或 SEQ ID NO: 33 的核酸序列或其互补的核酸序列。

6. 前面任一权利要求的进一步定义为 DNA 片段的核酸片段。

20 7. 前面任一权利要求的核酸片段, 其中所述的基因可操作地连接到在宿主细胞中表达所述基因的启动子上。

8. 前面任一权利要求的核酸片段, 其被包括在重组载体中。

9. 前面任一权利要求的核酸片段, 其被包括进质粒、粘粒、噬菌体、噬菌粒、病毒、杆状病毒、细菌人工染色体或酵母人工染色体重组载体
25 中。

10. 前面任一权利要求的核酸片段, 其被包括进质粒载体 pEG 1068、pEG 1077、pEG 11092、pEG 1093、pEG 365、pEG 378 或 pEG 381 中。

30 11. 根据前面任一权利要求的核酸片段用于重组表达方法中以制备重组多肽。

12. 根据前面任一权利要求的核酸片段用于制备昆虫抗性转基因植

物。

13. 使用根据前面任一权利要求核酸片段的方法，包括在宿主细胞中表达所述的基因并且收集表达的多肽。

14. 根据权利要求 1 到 12 中任一权利要求的核酸片段在制备重组多肽组合物中的应用。

15. 根据权利要求 1 到 12 中任一权利要求的核酸片段在构建用于制备昆虫抗性转基因植物中的应用。

16. 根据权利要求 1 到 12 中任一权利要求的核酸片段在构建昆虫抗性转基因植物中的应用。

10 17. 宿主细胞，包括根据权利要求 1 到 12 中任一权利要求的核酸片段。

18. 权利要求 17 的宿主细胞，其中所述的宿主细胞为细菌细胞。

19. 权利要求 17 或 18 的宿主细胞，其中所述的细胞为大肠杆菌、苏云金芽孢杆菌、枯草芽孢杆菌、巨大芽孢杆菌或假单胞菌的细胞。

15 20. 权利要求 17 到 19 中任一权利要求的宿主细胞，其中所述的细胞为苏云金芽孢杆菌 EG 11063、EG 11074、EG 11090、EG 11092、EG 11735、EG 11751、EG 11768、NRRL B - 21579、NRRL B - 21580、NRRL B - 21581、NRRL B - 21635、NRRL B - 21636、NRRL B - 21780 或 NRRL B - 21781。

20 21. 权利要求 17 的宿主细胞，其中所述的细胞为真核细胞。

22. 权利要求 21 的宿主细胞，其中所述的宿主细胞为植物细胞。

23. 权利要求 21 或 22 的宿主细胞，其中所述的细胞为谷物、树木、蔬菜、水果、浆果、坚果、草、仙人掌、肉质或观赏植物细胞。

25 24. 权利要求 21 到 23 的任一权利要求的宿主细胞，其中所述的细胞为玉米、水稻、烟草、土豆、番茄、亚麻、canola、向日葵、棉花、小麦、燕麦、大麦或黑麦细胞。

25. 权利要求 21 到 24 中任一权利要求的宿主细胞，其中所述的细胞被包括进转基因植物中。

30 26. 权利要求 21 到 25 中任一权利要求的宿主细胞，其中所述的细胞产生具有抗草地贪夜蛾、甜菜叶蛾、烟芽夜蛾、玉米穗夜蛾或欧洲玉米螟杀虫活性的多肽。

27. 根据权利要求 17 到 26 中任一权利要求的宿主细胞用于重组多肽的表达。

28. 根据权利要求 17 到 26 中任一权利要求的宿主细胞用于制备转基因植物。

5 29. 根据权利要求 17 到 26 中任一权利要求的宿主细胞在制备多能植物细胞中的应用。

30. 根据权利要求 17 到 26 中任一权利要求的宿主细胞在制备杀虫多肽制品中的应用。

10 31. 组合物, 包括含有 SEQ ID NO: 10、SEQ ID NO: 12、SEQ ID NO: 14、SEQ ID NO: 26、SEQ ID NO: 28、SEQ ID NO: 30 或 SEQ ID NO: 34 氨基酸序列的分离多肽。

32. 权利要求 31 的组合物, 其中所述的多肽对草地贪夜蛾、甜菜叶蛾、烟芽夜蛾、玉米穗夜蛾或欧洲玉米螟具有杀虫活性。

15 33. 权利要求 31 或 32 的组合物, 其中所述的多肽可从苏云金芽孢杆菌 EG 11060、EG 11062、EG 11063、EG 11071、EG 11073、EG 11074、EG 11090、EG 11091、EG 11092、EG 11735、EG 11751、EG 11768、NRRL B - 21579、NRRL B - 21580、NRRL B - 21581、NRRL B - 21635、NRRL B - 21636、NRRL B - 21780 或 NRRL B - 21781 中分离。

20 34. 权利要求 31 到 33 中任一权利要求的组合物, 其中所述的多肽包括约 0.5 % 到约 99 % 重量的所述组合物。

35. 权利要求 31 到 34 中任何一个权利要求的组合物, 其中所述的多肽包括约 50 % 到约 99 % 重量的所述组合物。

25 36. 组合物, 包括可通过下列方法制备的多肽, 该方法包括以下步骤:

(a) 在有效产生包括苏云金芽孢杆菌多肽组合物的条件下培养苏云金芽孢杆菌 NRRL B - 21579、NRRL B - 21580、NRRL B - 21581、NRRL B - 21635、NRRL B - 21636、NRRL B - 21780、NRRL B - 21781、EG 11768、EG 11090、EG 11063、
30 EG 11074、EG 11092、EG 11735、或 EG 11751 细胞; 和

(b) 从所述细胞中获取所述的组合物。

37. 根据权利要求 31 到 36 中任一权利要求的组合物用于杀死昆虫细胞。

38. 根据权利要求 31 到 37 中任一权利要求的组合物在制备杀虫剂中的应用。

5 39. 根据权利要求 31 到 37 中任一权利要求的组合物在制备植物保护性喷洒制剂中的应用。

40. 制备苏云金芽孢杆菌晶体蛋白的方法, 包括:

10 (a) 在有效产生苏云金芽孢杆菌晶体蛋白的条件下培养苏云金芽孢杆菌 NRRL B - 21579、NRRL B - 21580、NRRL B - 21581、NRRL B - 21635、NRRL B - 21636、NRRL B - 21780、NRRL B - 21781、EG 11768、EG 11090、EG 11063、EG 11074、EG 11092、EG 11735 或 EG 11751 细胞; 和

(b) 从所述的细胞中获取所述的苏云金芽孢杆菌晶体蛋白。

15 41. 杀死昆虫细胞的方法, 包括提供给昆虫细胞杀虫有效量的根据权利要求 31 到 37 中任一权利要求的组合物。

42. 权利要求 41 的方法, 其中所述的昆虫细胞在昆虫体内。

43. 权利要求 42 的方法, 其中所述的昆虫通过摄食覆盖有所组合物的植物而摄食所述的组合物。

20 44. 权利要求 42 或 43 的方法, 其中所述的昆虫通过摄食表达所述组合物的转基因植物而摄食所述的组合物。

45. 特异性地与包括 SEQ ID NO: 10、SEQ ID NO: 12、SEQ ID NO: 14、SEQ ID NO: 26、SEQ ID NO: 28、SEQ ID NO: 30 或 SEQ ID NO: 34 氨基酸序列的多肽结合的纯化抗体。

46. 权利要求 45 的抗体, 其被可操作地结合到可检测的标记上。

25 47. 免疫检测试剂盒, 包括位于适当容器中的根据权利要求 45 抗体和免疫检测试剂。

48. 用于检测生物样品中杀虫多肽的方法, 包括在有效形成免疫复合物的条件下将怀疑含有所述杀虫多肽的生物样品与根据权利要求 45 的抗体接触并且检测这样形成的免疫复合物。

30 49. 在其基因组中具有掺入的转基因的转基因植物, 其中的转基因编码包括 SEQ ID NO: 10、SEQ ID NO: 12、SEQ ID NO: 14、

SEQ ID NO: 26、SEQ ID NO: 28、SEQ ID NO: 30 或 SEQ ID NO: 34 氨基酸序列的多肽。

50. 权利要求 49 的转基因植物，其中所述的转基因包括 SEQ ID NO: 9、SEQ ID NO: 11、SEQ ID NO: 13、SEQ ID NO: 25、
5 SEQ ID NO: 27、SEQ ID NO: 29 或 SEQ ID NO: 33 的核酸序列。

51. 权利要求 49 或 50 植物的后代。

52. 来自权利要求 49 到 51 中任一权利要求的植物或后代的种子。

53. 来自权利要求 52 种子的植物。

说明书

广谱 δ - 内毒素

1.0 发明背景

5 本发明是 1997 年 9 月 3 日提交的美国专利申请系列 No 08/922, 505 的部分延续, 后者是 1996 年 11 月 20 日提交的美国专利申请系列 No 08/754, 490 的部分延续, 特异性在此插入其全部内容供参考。

1.1 发明领域

10 本发明提供新的编码对鞘翅目、双翅目和鳞翅目昆虫具有毒性的苏云金芽孢杆菌 (*Bacillus thuringiensis*) 晶体蛋白的基因, 同时提供的是包括具有增强杀虫活性和增加杀虫特异性嵌合晶体蛋白的蛋白组合物。

1.2 相关领域描述

1.2.1 苏云金芽孢杆菌晶体蛋白

众所周知格兰氏阳性土壤细菌苏云金芽孢杆菌产生对多种鳞翅目、鞘翅目和双翅目幼虫具有毒性的蛋白伴胞晶体或 δ - 内毒素。苏云金芽孢杆菌在芽胞形成过程中产生对某些种昆虫具有特异性毒性的晶体蛋白。已显示苏云金芽孢杆菌的不同菌株产生杀虫晶体蛋白, 并且由于其对特定目标昆虫的毒性而对植物和其它非目标生物没有毒性, 包括产生具有杀虫活性蛋白的苏云金芽孢杆菌菌株的组合物已在商业上用作环境可接受的杀虫剂。

25 自然产生的苏云金芽孢杆菌分离物的商业制剂已长期用于农业害虫的生物控制。在商业制备中, 浓缩和制备从发酵过程中获得的芽胞和晶体用于根据常规的农业实践进行施用。

1.2.2 晶体蛋白的命名

30 Höfte 等 (1989) 的综述描述了有关鉴定出的对抵抗鳞翅目昆虫, 即毛毛虫类昆虫具有活性的大多数杀虫性苏云金芽孢杆菌菌株领域的一般状态。该论文也描述了对双翅目昆虫 (即苍蝇和蚊子) 和鞘翅目昆虫

(即, 甲虫) 具有杀虫活性的苏云金芽孢杆菌菌株。已从数种苏云金芽孢杆菌菌株中克隆了许多编码晶体蛋白的基因。Höfte 等 (1989) 讨论了在 1990 年之前在苏云金芽孢杆菌中鉴定出的基因和蛋白, 并且提出了已常规应用于苏云金芽孢杆菌基因和蛋白的命名和分类方案。cry 1 基因编码鳞翅目 - 毒性的 Cry 1 蛋白。cry 2 基因编码同时对鳞翅目和双翅目具有毒性的 Cry 2 蛋白。cry 3 基因编码鞘翅目 - 毒性的 Cry 3 蛋白, 而 cry 4 基因编码双翅目 - 毒性的 Cry 4 蛋白等。

最近提出一种新的命名方法, 其根据氨基酸序列同源性而不是昆虫目标的特异性系统地对 Cry 蛋白进行分类。该分类方案总结于表 1 中。

10

表 1
修订的苏云金芽孢杆菌 δ - 内毒素命名^A

| 新 | 旧 | GenBank 许可号 |
|--------|----------|-------------|
| CryIAa | CryIA(a) | M11250 |
| CryIAb | CryIA(b) | M13898 |
| CryIAc | CryIA(c) | M11068 |
| CryIAd | CryIA(d) | M73250 |
| CryIAe | CryIA(e) | M65252 |
| CryIBa | CryIB | X06711 |
| CryIBb | ET5 | L32020 |
| CryIBc | PEG5 | Z46442 |
| CryIBd | CryE1 | U70726 |
| CryICa | CryIC | X07518 |
| CryICb | CryIC(b) | M97880 |

99.07.14

| 新 | 旧 | GenBank 许可号 |
|--------|----------|-------------|
| Cry1Da | CryID | X54160 |
| Cry1Db | PrtB | Z22511 |
| Cry1Ea | CryIE | X53985 |
| Cry1Eb | CryIE(b) | M73253 |
| Cry1Fa | CryIF | M63897 |
| Cry1Fb | PrtD | Z22512 |
| Cry1Ga | PrtA | Z22510 |
| Cry1Gb | CryH2 | U70725 |
| Cry1Ha | PrtC | Z22513 |
| Cry1Hb | | U35780 |
| Cry1Ia | CryV | X62821 |
| Cry1Ib | CryV | U07642 |
| Cry1Ja | ET4 | L32019 |
| Cry1Jb | ET1 | U31527 |
| Cry1K | | U28801 |
| Cry2Aa | CryIIA | M31738 |
| Cry2Ab | CryIIB | M23724 |
| Cry2Ac | CryIIC | X57252 |
| Cry3A | CryIIIA | M22472 |
| Cry3Ba | CryIIIB | X17123 |
| Cry3Bb | CryIIIB2 | M89794 |
| Cry3C | CryIIID | X59797 |
| Cry4A | CryIVA | Y00423 |
| Cry4B | CryIVB | X07423 |
| Cry5Aa | CryVA(a) | L07025 |
| Cry5Ab | CryVA(b) | L07026 |
| Cry5B | | U19725 |
| Cry6A | CryVIA | L07022 |
| Cry6B | CryVIB | L07024 |
| Cry7Aa | CryIIIC | M64478 |

| | 新 | 旧 | GenBank 许可号 |
|----|--------|----------|-------------|
| | Cry7Ab | CryIIICb | U04367 |
| | Cry8A | CryIIIE | U04364 |
| 5 | Cry8B | CryIIIG | U04365 |
| | Cry8C | CryIIIF | U04366 |
| | Cry9A | CryIG | X58120 |
| | Cry9B | CryIX | X75019 |
| | Cry9C | CryIH | Z37527 |
| | Cry10A | CryIVC | M12662 |
| 10 | Cry11A | CryIVD | M31737 |
| | Cry11B | Jeg80 | X86902 |
| | Cry12A | CryVB | L07027 |
| | Cry13A | CryVC | L07023 |
| | Cry14A | CryVD | U13955 |
| 15 | Cry15A | 34kDa | M76442 |
| | Cry16A | cbm71 | X94146 |
| | Cry17A | cbm71 | X99478 |
| | Cry18A | CryBP1 | X99049 |
| | Cry19A | Jeg65 | Y08920 |
| | Cyt1Aa | CytA | X03182 |
| 20 | Cyt1Ab | CytM | X98793 |
| | Cyt1B | | U37196 |
| | Cyt2A | CytB | Z14147 |
| | Cyt2B | CytB | U52043 |

^a 节选自: http://epunix.biols.susx.ac.uk/Home/Neil_Crickmore/Bt/index.html

25

1.2.3 晶体蛋白毒性的模式

所有的 δ -内毒素晶体通过摄食而对昆虫幼虫具有毒性。该晶体在昆虫中肠中的溶解释放前毒素形式的 δ -内毒素, 在大多数情况下其随后被中肠蛋白酶加工成活性的毒素。活化的毒素经受体蛋白识别并结合到昆虫中肠上皮的刷状缘上。从一些昆虫幼虫中已分离出数种推断性的晶体蛋白受体 (Knight 等, 1995; Gill 等, 1995; Masson 等,

30

1995)。活性毒素结合后插入和聚集毒素分子以在中肠上皮内形成穿孔。该过程导致形成中肠上皮衬里的细胞渗透失衡、膨胀、裂解，并且最终幼虫死亡。

5 1.2.4 δ -内毒素的分子生物学

随着分子遗传技术的到来，已分离各种 δ -内毒素基因并且测定其 DNA 序列。这些基因已用于构建一些已获准商业使用的遗传工程苏云金芽孢杆菌产品。最近的进展又见到新开发出的 δ -内毒素转运系统，包括含有并表达遗传工程操作的 δ -内毒素基因的植物。

10 Höfte 和 Whiteley 在 1989 年已描述和总结过对多种苏云金芽孢杆菌菌株的许多 δ -内毒素基因的克隆和测序。设计用于在大肠杆菌或苏云金芽孢杆菌中克隆与表达 δ -内毒素基因的质粒穿梭载体由 Gawron - Burke 和 Baum (1991) 描述过。美国专利 No. 5, 441, 884 公开了用于构建含有无苏云金芽孢杆菌外源 DNA 的 δ -内毒素基因重组苏云金芽孢杆菌菌株的定点重组系统。

15 主要对抵抗鳞翅目害虫具有活性的 Cry 1 家族晶体蛋白是研究最为彻底的一类 δ -内毒素。前毒素形式的 Cry 1 δ -内毒素由 2 个大致相同长度的片段所组成。羧基端或前-毒素片段没有毒性并且认为对于晶体形成是重要的 (Arvidson 等, 1989)。前毒素的氨基端包括 Cry 1 分子活性的毒素片段并且通过最近所述的 Cry 1 Aa δ -内毒素活性毒素片段的晶体学结构测定可进一步分为三个结构域 (Grochulski 等, 1995)。结构域 1 占据该活性毒素的头三分之一并且对于通道的形成是必需的 (Thompson 等, 1995)。结构域 2 和结构域 3 分别占据该活性毒性毒素的中部和后三分之一。依据被检测的昆虫和 δ -内毒素，结构域 2 和 3 均涉及受体结合和昆虫的特异性 (Thompson 等, 1995)。

1.2.5 嵌合晶体蛋白

30 近年来，研究者集中致力于构建杂合 δ -内毒素，从而希望制备具有增强活性或改进特性的蛋白。过去十年中分子遗传学领域的进展促进了加工具有改进特性蛋白的逻辑和有条不紊的方法。定点和随机诱变方法、聚合酶链式反应 (PCR™) 方法学的出现和开发用于产生基因融合

和构建嵌合蛋白的重组方法促进对以下方法的分类，包括改变蛋白的氨基酸序列、将两个或更多个蛋白成分一起融合到一个重组蛋白中和改变编码商业目的蛋白的遗传序列。

不幸的是，对于结晶蛋白，这些技术仅仅以有限的方式得到利用。

- 5 从众多已鉴定出天然蛋白部分中随机产生具有增强特性嵌合蛋白的可能性很少得到复杂性质的蛋白结构、折叠、寡聚化、活化和正确加工嵌合的前毒素成为活性的部分。仅仅通过对每种蛋白内部特性靶区域的细心筛选和随后的蛋白加工才可合成具有改进杀虫活性的毒素。

然而，文献中已报导过在该领域的一些成功。例如，在下面有关领域
10 域中报导了一些杂合 δ -内毒素的构建：国际专利申请公开 No. WO 95/30753 公开了用于在荧光假单胞菌 (*Pseudomonas fluorescens*) 中制备的杂合苏云金假单胞菌 δ -内毒素的构建，其中的 Cry 1 F 的非毒性前毒素片段已由美国专利 5,128,130 中公开的 Cry 1 Ac/Cry 1 Ab 的非毒性前毒素片段所代替。

- 15 美国专利 5,128,130 公开了用于在荧光假单胞菌中制备的杂合苏云金假单胞菌 δ -内毒素的构建，其中的 Cry 1 Ac 的非毒性前毒素片段的部分已由 Cry 1 Ab 的相应非毒性前毒素片段所代替。美国专利 5,055,294 公开了在 Cry 1 Ac (氨基酸残基 1 - 466) 和 Cry 1 Ab (氨基酸残基 466 - 1155) 之间构建用于在荧光假单胞菌中制备的特异性杂合 δ
20 -内毒素。尽管前述的专利公开了在活性毒素片段内杂合毒素的构建，但没有列出有关该杂合毒素杀虫活性的特异性。国际专利申请公开 No. WO 95/30752 公开了用于在荧光假单胞菌中制备的杂合苏云金假单胞菌 δ -内毒素的构建，其中的 Cry 1 C 的非毒性前毒素片段已由 Cry 1 Ab 的非毒性前毒素片段所替代。前述的申请进一步公开了该杂合 δ -内毒素
25 素抗甜菜夜蛾的活性高于母本活性毒素，Cry 1 C 的活性。

国际专利申请公开 No. WO 95/06730 公开了由偶联到结构域 3 上的结构域 1 和 2 和 Cry 1 C 非毒性前毒素片段所组成的杂合苏云金芽孢杆菌 δ -内毒素的构建。对烟草天蛾 (*Manduca sexta*) (对 Cry 1 C 和 Cry 1 E 敏感)、甜菜夜蛾 (*Spodoptera exigua*) (对 Cry 1 C 敏感)
30 和甘蓝夜蛾 (*Mamestra brassicae*) (对 Cry 1 C 敏感) 进行的昆虫生物测定显示该杂合 Cry 1 E / Cry 1 C 杂合毒素对抵抗烟草天蛾、甜

菜夜蛾和烟草夜蛾是有活性的。该生物分析结果表示为 EC_{50} 值（得到 50% 生长降低的毒素浓度）而不是 LC_{50} 值（得到 50% 死亡率的毒素浓度）。尽管用于生物分析的 δ -内毒素在苏云金芽孢杆菌中制备，但是仅仅使用人工产生的 δ -内毒素的活性片段，而不使用由存在于商业苏云金芽孢杆菌制剂中的苏云金芽孢杆菌一般产生的天然产生的晶体。生物分析结果表明该杂合 Cry 1 E / Cry 1 C 晶体抗草地贪夜蛾（*S. frugiperda*）的 LC_{50} 值较天然 Cry 1 C 低 1.5 到 1.7 倍（活性更高）。该文章也公开了在 Cry 1 Ab（结构域 1 和 2）和 Cry 1 C（结构域 3 和非毒性前毒素片段）之间构建杂合的苏云金芽孢杆菌 δ -内毒素，尽管没有得到有关该杂合毒素活性或使用的数据。

Lee 等（1995）报导了在活性的毒素片段内部建 Cry 1 Ac 和 Cry 1 Aa 之间的杂合苏云金芽孢杆菌 δ -内毒素。将人工产生的该杂合毒素的活性片段用于在易感昆虫刷状缘膜囊泡（BBMV）中检测蛋白的相互作用。没有报导该杂合毒素的生物活性。

Honee 等（1991）报导了 Cry 1 C（结构域 1）和 Cry 1 Ab（结构域 2 和 3）之间杂合 δ -内毒素的构建和 Cry 1 Ab（结构域 1）和 Cry 1 C（结构域 2 和 3）之间交互杂合体的构建。这些杂合体在抵抗易感昆虫活性中没有显示出明显的增加。此外，发现 Cry 1 C（结构域 1）/Cry 1 Ab（结构域 2 和 3）杂合毒素对蛋白酶降解是超敏的。Schnepf 等（1991）的报导公开了 Cry 1 Ac 杂合毒素的构建，其中的结构域 2 的小部分由 Cry 1 Aa 的相应区域所代替，虽然没有观察到抵抗易感昆虫幼虫活性明显的增加。

1.3 现有技术的缺陷

在制备具有增强活性嵌合晶体蛋白中有限的成功通过挫败制备用于商业开发的重组加工晶体蛋白和扩展已知内毒素的毒性特征和宿主特异性的努力而对本领域产生负面影响。因此，现有技术中所缺少的是可靠的方法和包括重组加工晶体蛋白的组合物，其中的晶体蛋白具有增强的杀虫活性，广谱的宿主特异性（broad-host-range）并且适于在苏云金芽孢杆菌中商业制备。

2.0 发明概述

本发明通过提供具有增强的杀虫特性和广谱特异性的新的嵌合 δ - 内毒素而克服了现有技术中的这些或其它的局限性。

公开的为构建包括天然 Cry 1 Ac 和 Cry 1 F 晶体蛋白氨基酸序列的苏云金芽孢杆菌杂合 δ - 内毒素的构建。这些杂合蛋白（其中的全部或部分的 Cry 1 Ac 结构域 2、全部或部分的 Cry 1 Ac 结构域 3 以及全部或部分的 Cry 1 Ac 前毒素片段由 Cry 1 F 的相应部分所代替）不仅具有母本 δ - 内毒素的杀虫特征，而且具有没有预期到和增强的广谱特异性的显著特性，该特性在从中加工嵌合蛋白的 2 个天然 δ - 内毒素中没有任何一个充分表现出来。

具体地，本发明公开并且要求保护包括融合到 Cry 1 F 晶体蛋白部分上的 Cry 1 Ac 晶体蛋白部分的遗传加工的杂合 δ - 内毒素。这些嵌合的内毒素对在此所述的害虫具有广谱特异性。

在进一步的实施方案中，本发明也公开并要求保护包括 Cry 1 Ab、Cry 1 F 和 Cry 1 Ac 部分的重组苏云金芽孢杆菌杂合内毒素，其中的全部或部分的 Cry 1 Ab 结构域 2 或全部或部分的 Cry 1 Ab 结构域 3 由 Cry 1 F 的相应部分所代替并且全部或部分的 Cry 1 Ab 前毒素片段由 Cry 1 Ac 的相应部分所代替。Cry 1 Ab 和 Cry 1 F 之间杂合 δ - 内毒素的实例鉴定于 SEQ ID NO: 13 和 SEQ ID NO: 14 中。

本发明的一个方面显示了没有预料到的如下结果，即一些来源于 Cry 1 Ac 和 Cry 1 F 蛋白的杂合 δ - 内毒素不仅显示出母本 δ - 内毒素的杀虫特征，而且具有 2 个母本 δ - 内毒素中任何一个均没有充分显示出来的杀虫活性。

本发明的另一方面进一步显示了没有预料到的如下结果，即一些嵌合的 Cry 1 Ab / Cry 1 F 蛋白不仅维持了母本 δ - 内毒素的杀虫特征，而且显示出天然 Cry 1 Ab 或 Cry 1 F 内毒素中任何一个均没有显示出的杀虫活性。

本发明也包括维持对在苏云金芽孢杆菌中商业制备所必需特征的 Cry 1 Ac / Cry 1 F 和 Cry 1 Ab / Cry 1 F 杂合 δ - 内毒素。具体地，在 SEQ ID NO: 10、SEQ ID NO: 12、SEQ ID NO: 14、SEQ ID NO: 26、SEQ ID NO: 28、SEQ ID NO: 30 和 SEQ ID NO: 34

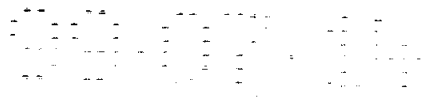
中鉴定出的杂合 δ - 内毒素可在苏云金芽孢杆菌中有效地形成蛋白伴胞包涵物并且具有母本 δ - 内毒素溶解性、蛋白酶敏感性和杀虫活性的有利特征。

5 在进一步的实施方案中，本发明也公开并且要求保护包括 Cry 1 Ac 和 Cry 1 C 部分的重组苏云金芽孢杆菌杂合 δ - 内毒素，其中的全部或部分的 Cry 1 Ac 结构域 3 由 Cry 1 C 的相应部分所代替并且全部或部分的 Cry 1 Ac 前毒素片段由 Cry 1 C 的相应部分所代替。Cry 1 Ac 和 Cry 1 C 之间杂合 δ - 内毒素的实例鉴定于 SEQ ID NO: 29 和 SEQ ID NO: 30 中。

10 本发明的一个方面显示了没有预料到的如下结果，即虽然 Cry 1 Ac 和 Cry 1 C 均不具有草地贪夜蛾活性，但是由 SEQ ID NO: 29 和 SEQ ID NO: 30 所鉴定出的 Cry 1 Ac / Cry 1 C 杂合 δ - 内毒素具有显著的抗草地贪夜蛾的活性。此外，由 SEQ ID NO: 29 和 SEQ ID NO: 30 所鉴定出的 Cry 1 Ac / Cry 1 C 杂合 δ - 内毒素较 Cry 1 C 母本 δ - 内毒素具有显著更好的抗甜菜夜蛾的活性。

15 构建这些和其它杂合 δ - 内毒素的动力是产生具有增强杀虫活性、增加的宿主范围特异性和改进的制备特征的新的毒素。列于表 8 中的 DNA 序列定义了与本发明有关的杂合 δ - 内毒素的交换位点并且作为寡核苷酸引物，可在适当到高度严格条件下通过 Southern 或菌落杂交而用于鉴定同样或类似的杂合 δ - 内毒素。本领域熟练的研究人员将认识到在二种或更多种 δ - 内毒素之间选定的交换位点的重要性可用众多的体内或体外分子遗传技术加以完成。在二种或更多种 δ - 内毒素之间的交换区域中的微小变异可能产生类似的结果或者，如 EG 11062 和 EG 11063 所显示，对所需特征产生负面影响。类似地，如 EG 11063 和 EG 25 11074 所显示，在二种或更多种 δ - 内毒素之间的交换区域中的大的变异可能对所需特征没有影响，或者如 EG 11060 和 EG 11063 所显示，可能对所需特征产生负面影响。

关于增强的杀虫活性、增加的宿主范围和改进的制备特征的有利特征可通过其它这样的杂合 δ - 内毒素加以完成，包括但不限于 cry 1、30 cry 2、cry 3、cry 4、cry 5、cry 6、cry 7、cry 8、cry 9、cry 10、cry 11、cry 12、cry 13、cry 14、cry 15 类 δ - 内毒素基因和苏云



金芽孢杆菌细胞裂解性 cyt 1 和 cyt 2 基因。这类苏云金芽孢杆菌杀虫剂蛋白的成员包括但不限于 Cry 1 Aa、Cry 1 Ab、Cry 1 Ac、Cry 1 Ad、Cry 1 Ae、Cry 1 Ba、Cry 1 Bb、Cry 1 Ca、Cry 1 Cb、Cry 1 Da、Cry 1 Db、Cry 1 Ea、Cry 1 Eb、Cry 1 Fa、Cry 1 Fb、Cry 1 Ga、Cry 1 Ha、Cry 2a、Cry 2b、Cry 1 Ja、Cry 1 Ka、Cry 11 Aa、Cry 11 Ab、Cry 12 Aa、Cry 3 Ba、Cry 3 Bb、Cry 3 C、Cry 4a、Cry 4 Ba、Cry 5a、Cry 5 Ab、Cry 6 Aa、Cry 6 Ba、Cry 7 Aa、Cry 7 Ab、Cry 8 Aa、Cry 8 Ba、Cry 8 Ca、Cry 9 Aa、Cry 9 Ba、Cry 9 Ca、Cry 10 Aa、Cry 11 Aa、Cry 12 Aa、Cry 13 Aa、Cry 14 Aa、Cry 15 Aa、Cyt 1 Aa 和 Cyt 2 Aa。有关的杂合 δ -内毒素将由前述 δ -内毒素中的其中一种毒素的氨基部分所组成，该部分包括融合到另一种前述 δ -内毒素全部或部分结构域 3 上的全部或部分的 5 结构域 1 或结构域 2。该杂合 δ -内毒素的无活性前毒素片段可由任何一种前述的可稳定该杂合 δ -内毒素的 δ -内毒素前毒素片段所组成，如 EG 11087 和 EG 11091 所显示（参见，如表 5）。与本发明中所述的 δ -内毒素具有类似特征的杂合 δ -内毒素可通过在杂合 δ -内毒素内部保守或“类似”的氨基酸替代而加以构建。这种替代将模仿该蛋白中任何位置天然氨基酸的生化 and 生物物理特性。认为类似的氨基酸包括，例如，但不限于：

- 20 Ala、Ser 和 Thr；
- Asp 和 Glu；
- Asn 和 Gln；
- Lys 和 Arg；
- Ile、Leu、Met 和 Val；和
- 25 Phe、Tyr 和 Trp。

本领域的技术人员将认识到赋予杂合 δ -内毒素的增强的杀虫活性、增加的宿主范围和改进的制备特征可通过改变将杂合 δ -内毒素中一个或更多个氨基酸位置的遗传密码从而使该位置或这些位置由任何其它的氨基酸所代替而得到进一步改进。这可通过任何种类确定的诱变技术，包括与本发明有关的那些方法针对该蛋白的某一或某些区域进行诱变而完成。这些技术包括定点诱变（Kunkel, 1985；Kunkel 等，

1987)、DNA 改组 (Shuffling) (Stemmer, 1994) 和 PCR™ 重叠延伸 (Horton 等, 1989)。由于位于或靠近蛋白表面的氨基酸很可能负责其与其它蛋白或非蛋白成分之间的相互作用, 故它们可用作诱变的“目标”区域。该表面裸露的区域可能由, 但不限于该蛋白活性毒素片段内的表面裸露的氨基酸残基所组成并且包括在结构域 1 内部分隔 α -螺旋的 δ -内毒素的中间- α -螺旋或中间- β -链“环”-区域和结构域 2 和结构域 3 内部的 β -链。该步骤可有利地改变此蛋白的生化和生物物理特征或其作用方式, 如第 1 部分中所总结的。这些特征包括但不限于: 1) 增强的晶体形成, 2) 增强的蛋白稳定性或降低的蛋白酶降解, 3) 增强的昆虫膜受体识别和结合, 4) 在昆虫中肠上皮中增强的寡聚化或通道形成, 和 5) 由于上述任何一种或所有原因所致的增强的杀虫活性或杀虫特异性。

本发明提供了编码包括 SEQ ID NO: 10、SEQ ID NO: 12、SEQ ID NO: 14、SEQ ID NO: 26、SEQ ID NO: 28、SEQ ID NO: 30 或 SEQ ID NO: 34 氨基酸序列的多肽的分离的核酸片段。优选该核酸片段编码具有抗草地贪夜蛾、甜菜夜蛾、烟芽夜蛾 (*Heliothis virescens*)、玉米穗夜蛾 (*Helicoverpa zea*) 或欧洲玉米螟 (*Ostrinia nubilalis*) 杀虫活性的多肽。该核酸片段可分别从苏云金芽孢杆菌 NRRL B-21579、NRRL B-21580、NRRL B-21581、NRRL B-21635、NRRL B-21636、NRRL B-21780 或 NRRL B-21781 细胞中得到分离。

在优选的实施方案中, 这些核酸片段与具有 SEQ ID NO: 9、SEQ ID NO: 11、SEQ ID NO: 13、SEQ ID NO: 25、SEQ ID NO: 27、SEQ ID NO: 29 或 SEQ ID NO: 33 序列或其互补序列的核酸片段特异性杂交, 并且在高度优选的实施方案中, 这些核酸片段包括分别在 SEQ ID NO: 9、SEQ ID NO: 11、SEQ ID NO: 13、SEQ ID NO: 25、SEQ ID NO: 27、SEQ ID NO: 29 和 SEQ ID NO: 33 中公开的核酸序列。

可将这样的核酸片段可操作地连接到启动子上以在宿主细胞中表达该核酸片段。在该实施方案中, 可将此核酸片段包括进重组载体如质粒、粘粒、噬菌体、噬菌粒、病毒、杆状病毒、细菌人工染色体或酵母人工

染色体中。质粒载体的实例包括载体 pEG 1068、pEG 1077、pEG 11092、pEG 11093、pEG 365、pEG 378 和 pEG 381。

本发明的进一步方面涉及这里所述的核酸片段在制备重组多肽组合物、产生用于制备转化宿主细胞的载体和产生昆虫抗性的转基因植物中的使用。

宿主细胞也代表本发明重要的方面。该宿主细胞一般包括上述一种或更多种的核酸片段。优选的宿主细胞包括细菌细胞如大肠杆菌、苏云金芽孢杆菌、枯草芽孢杆菌、巨大芽孢杆菌 (*B. megaterium*) 和假单胞菌细胞, 其中的苏云金芽孢杆菌 EG 11060、EG 11062、EG 11063、EG 11071、EG 11073、EG 11074、EG 11090、EG 11091、EG 11092、EG 11735、EG 11751、EG 11768、NRRL B - 21579、NRRL B - 21580、NRRL B - 21581、NRRL B - 21635、NRRL B - 21636、NRRL B - 21780 和 NRRL B - 21781 细胞是特别优选的。优选的宿主细胞也可包括真核细胞, 如植物和动物细胞。优选的植物细胞包括谷物、树木、蔬菜、水果、浆果、板栗、草、仙人掌、肉质和观赏植物的细胞, 其中商业作物如玉米、水稻、烟草、土豆、番茄、亚麻、canola、向日葵、棉花、小麦、燕麦、大麦和黑麦是特别优选的。

在一个实施方案中, 可将此宿主细胞包括进转基因植物或可用于制备转基因植物或产生多能植物细胞。或者, 该宿主细胞可用于晶体蛋白的重组表达或包括一种或更多种这里公开的毒素的杀虫多肽制剂的制备。该组合物优选包括一种或更多种具有 SEQ ID NO: 10、SEQ ID NO: 12、SEQ ID NO: 14、SEQ ID NO: 26、SEQ ID NO: 28、SEQ ID NO: 30 或 SEQ ID NO: 34 氨基酸序列的分离的多肽。这样的组合物在杀死昆虫细胞、制备杀虫剂和制备植保喷剂中具有特别的使用。

本发明也提供了用于制备苏云金芽孢杆菌晶体蛋白的方法。该方法一般包括在有效产生苏云金芽孢杆菌晶体蛋白的条件下培养苏云金芽孢杆菌 NRRL B - 21579、NRRL B - 21580、NRRL B - 21581、NRRL B - 21635、NRRL B - 21636、NRRL B - 21780、NRRL B - 21781、EG 11768、EG 11090、EG 11063、EG 11074、EG 11092、EG 11735 或 EG 11751 细胞且然后从该细胞中获取所述的苏云

金芽孢杆菌晶体蛋白。这些方法在重组制备这里公开的晶体蛋白中相当有用。

5 本发明也公开和要求保护杀死昆虫细胞的方法。该方法一般包括提供给昆虫细胞杀虫有效量的一种或更多种这里公开的杀虫组合物。该细胞可以为分离的细胞，或者作为选择，可包括进昆虫本身体内。一般地，该组合物将通过下面任一种方式提供给昆虫，包括直接喷洒昆虫，或者通过昆虫直接摄入该组合物，或者通过摄入已被此组合物覆盖的植物，或者通过摄入表达一种或更多种杀虫组合物的转基因植物的一部分。

10 本发明的进一步方面是与包括 SEQ ID NO: 10、SEQ ID NO: 12、SEQ ID NO: 14、SEQ ID NO: 26、SEQ ID NO: 28、SEQ ID NO: 30 或 SEQ ID NO: 34 氨基酸序列的多肽特异性结合的纯化抗体。可将该抗体可操作地附着到可检测的标记上，在免疫检测试剂盒中加以提供或者用于通过以下方法检测生物样品中杀虫多肽的方法中，其中的方法包括将怀疑含有所述杀虫多肽的生物样品在有效形成免疫复合物的条件下与该抗体接触并且然后检测形成的免疫复合物。

15 本发明另一个重要的方面涉及其基因组中掺入有转基因的转基因植物，其中的转基因编码包括 SEQ ID NO: 10、SEQ ID NO: 12、SEQ ID NO: 14、SEQ ID NO: 26、SEQ ID NO: 28、SEQ ID NO: 30 或 SEQ ID NO: 34 氨基酸序列的多肽。优选地，该转基因植物将包括一种或更多种在 SEQ ID NO: 9、SEQ ID NO: 11、SEQ ID NO: 13、SEQ ID NO: 25、SEQ ID NO: 27、SEQ ID NO: 29 和 SEQ ID NO: 33 中公开的核酸序列。该转基因植物的子代和种子及其后代或子代也是本发明的重要方面，这将在本文别处详细描述。

25 2.1 晶体蛋白转基因和转基因植物

在另一方面，本发明提供了制备表达编码本发明新的嵌合晶体蛋白核酸片段的转基因植物的方法。制备转基因植物的方法在本领域中是众所周知的。一般地，该方法包括用下面的 DNA 片段转化合适的宿主细胞，该 DNA 片段含有可操作地连接到编码苏云金芽孢杆菌 Cry 1 Ac - 1 F 或 Cry 1 Ab - 1 F、Cry 1 Ac - 1 C 或 Cry 1 Ab - 1 Ac - 1 F 嵌合晶体蛋白 30 编码区域上的启动子。该编码区一般可操作地连接到转录终止区域，而

该启动子能够驱动编码区在细胞中的转录，并且因此提供细胞以体内产生该重组蛋白的能力。作为选择，当需要控制、调节或减少在特定的转基因细胞中表达的特定重组晶体蛋白的量时，本发明也提供用于晶体蛋白表达的反义 mRNA。反义 mRNA 在细胞中作为控制或降低给定目的蛋白量的方法的使用在本领域中是众所周知的。

本发明的另一方面包括表达编码一种或更多种这里所公开新多肽组合物基因或基因片段的转基因植物。如这里所使用的，术语“转基因植物”意指具有掺入 DNA 序列的植物，该序列包括但不限于也许正常不存在的基因。正常不转录成 RNA 或翻译成蛋白（“表达”）的 DNA 序列、或者人们希望导入非转化植物中的任何其它基因或 DNA 序列，如可能正常存在于非转化植物中但人们希望遗传加工或具有改变表达的基因。植物中合成苏云金芽孢杆菌基因的构建与表达已详细描述于美国专利 5,500,365 和 5,380,831（每个特异性地在此掺入供参考）中。

应该注意到有时本发明转基因植物的基因组经过稳定导入一种或更多种天然、人工修饰或进一步突变的 cry 1 Ac - 1F、cry 1 Ab - 1F、cry 1 Ac - 1C、或 cry 1 Ab - 1 Ac - 1 F 转基因后将会增大。在有些情况，多于一种转基因将被掺入到转化宿主植物细胞的基因组中。当将多于一种晶体蛋白编码 DNA 片段掺入到这样植物的基因组中时的情况便是实例。在有些情况中，在转化的转基因植物中掺入和稳定表达一种、二种、三种或甚至更多种苏云金芽孢杆菌晶体蛋白（或者是天然的或者是重组加工的）可能是必要的。

优选的基因，如可被导入的在 SEQ ID NO： 9、SEQ ID NO： 11、SEQ ID NO： 13、SEQ ID NO： 25、SEQ ID NO： 27、SEQ ID NO： 29 和 SEQ ID NO： 33 中公开的那些基因包括，例如，细菌来源的晶体蛋白编码 DNA 序列和特别是从芽孢杆菌获得的一种或更多种这种公开的那些基因。高度优选的核酸序列为那些从苏云金芽孢杆菌中获得的序列或被遗传加工以降低或增加在这样的转化宿主细胞中该晶体蛋白杀虫活性的那些序列中的任何序列。

用于转化植物细胞并且制备转基因细胞系的方法在本领域中是众所周知的，并且在此进行讨论。用于转化该细胞的载体、质粒、粘粒、酵母人工染色体（YACs）和核酸片段当然一般包括天然或合成产生的本

发明操纵子、基因或基因的衍生序列并且特别是那些编码已公开晶体蛋白的基因。这些 DNA 构建体可进一步包括这些结构，如启动子、增强子、多接头或者甚至是按需要正负调节特定目的基因活性的基因序列。此 DNA 片段或基因可能编码将要在得到的重组细胞中表达并且/或者将赋予再生植物改进表型的天然或修饰的晶体蛋白。被优化用于在植物中表达的核酸序列已公开于国际专利申请公开 No. WO 93/07278（特异性在此掺入供参考）。

通过向该植物中掺入编码具有广泛昆虫特异性的 Cry 1 Ac - 1 F 和/或 Cry 1 Ac - 1 C、和/或 Cry 1 Ab - 1 F 和/或 Cry 1 Ab - 1 Ac - 1 F 晶体蛋白的转基因 DNA 片段，该转基因植物可能对增加单子叶或双子叶植物的杀虫抗性是理想的。特别优选的植物如谷物，包括但不限于玉米（corn）、小麦、黑麦、水稻、玉米（maize）和大麦；棉花；大豆及其它豆科植物；树木，包括但不限于观赏植物、灌木、果树、板栗；蔬菜、草皮和牧草、浆果、柑橘和其它商业目的的作物；如园艺作物和/或居室植物（houseplants）、肉汁植物、仙人掌及开花的种类。

在有关方面，本发明也包括由转化植物产生的种子、由该种子来源的后代和根据上述方法产生的最初转基因植物的后代产生的种子。此后代和种子将稳定地具有稳定掺入其基因组中的晶体蛋白转基因，并且该后代植物将以孟德尔方式遗传通过稳定转基因的导入而赋予的特征。其基因组中具有掺入的编码一种或更多种嵌合晶体蛋白或多肽转基因 DNA 片段的所有这些转基因植物均是本发明的方面。

2.2 晶体蛋白的筛选和免疫检测试剂盒

本发明涉及到用于筛选怀疑含有晶体蛋白多肽或晶体蛋白有关多肽的样品或产生这样多肽的细胞的方法和试剂盒。这些蛋白的实例包括在 SEQ ID NO: 10、SEQ ID NO: 12、SEQ ID NO: 14、SEQ ID NO: 26、SEQ ID NO: 28、SEQ ID NO: 30 和 SEQ ID NO: 34 中公开的那些蛋白。所述的试剂盒可能含有本发明的核酸片段或抗体。该试剂盒可能含有用于检测样品和本发明核酸或抗体之间相互作用的试剂。该提供的试剂可能被放射、荧光或酶标记。该试剂盒可能含有能够与本发明核酸或抗体结合或相互作用的已知的放射性标记试剂。

该试剂盒的试剂可用附着到固体支持物上的液体溶液或干粉形式提供。优选地，当该试剂以液体溶液形式提供时，该液体溶液为水溶液。优选地，当该提供的试剂附着到固体支持物上时，该固体支持物可以为层析介质、多孔测试平板或显微镜载玻片。当提供的试剂为干粉时，可
5 通过加入可能已提供的适当溶剂而重新调制此粉末。

在更进一步的实施方案中，本发明涉及免疫检测方法和有关的试剂盒。已经提出本发明的晶体蛋白或肽可以用于检测与其具有反应性的抗体，或者，根据本发明制备的抗体可用于检测晶体蛋白或含有晶体蛋白
10 相关表位的肽。一般地，这些方法将包括首先获取怀疑含有该蛋白、肽或抗体的样品，在有效形成免疫复合物的条件下，可根据实情将该样品与本发明的抗体或肽接触，并且然后检测该免疫复合物的存在。

一般地，免疫复合物形成的检测在本领域中是十分熟知的并且可用众多的方法完成。例如，本发明涉及到应用 ELISA、RIA、免疫印迹（如，点印迹）、间接免疫荧光技术等。一般地，免疫复合物的形成将通过使
15 用标记，如放射标记或酶标签（如碱性磷酸酶、辣根过氧化物酶等）加以检测。当然，通过使用二级结合配体如本领域已知的二级抗体或生物素/亲合素配体结合组合，人们可能会发现其它的优势。

为了分析的目的，已提出实际上可以利用根据实际情况待测的怀疑包括晶体蛋白或肽或晶体蛋白有关的肽或抗体的任何样品。应该注意到
20 这样的实施方案可能在筛选杂交瘤等的抗原或抗体样品滴定中具有应用。在有关的实施方案中，本发明涉及到可用于检测样品中晶体蛋白或有关的肽和/或抗体的存在。样品可以包括怀疑含有晶体蛋白或肽的细胞、细胞上清液、细胞悬液、细胞提取物、酶的馏份、蛋白提取物或其它无细胞组合物。一般来说，根据本发明的试剂盒将包括合适的晶体蛋
25 白、肽或抗该蛋白或肽的抗体以及免疫测定试剂和用于容纳该抗体或抗原和试剂的方式。免疫测定试剂将一般包括与该抗体或抗原有关或与二级结合配体有关的标记。配体的实例可以包括抗一级抗体的二级抗体或抗原或者生物素或具有有关标记的亲合素（或链霉亲和素）。当然，如上所述，许多标记的实例在本领域中是已知的并且这些标记可与本发明
30 结合使用。

容器将一般包括其中可放置抗体、抗原或检测试剂并且优选适当分

装的小瓶。本发明试剂盒一般也将包括为了商业销售而在密闭条件下容纳该抗体、抗原和试剂容器的方式。这些容器可以包括其中放置所需小瓶的注射或吹制塑料容器。

5 2.3 ELISA 和免疫沉淀

ELISA 可与本发明联合使用。在 ELISA 分析中，将掺入晶体蛋白抗原序列的蛋白或肽固定于选定的表面，优选是显示出蛋白亲和力的表面如聚苯乙烯微滴定板的小孔。在清洗以去除不完全吸附的物质后，以非特异性蛋白结合或包被分析平板小孔是必要，该非特异性蛋白已知在抗原上
10 原上与测试抗血清如牛血清蛋白（BSA）、酪蛋白或奶粉溶液是无关的（neutral）。这样便封闭了固定表面上的非特异性吸附位点且因此减少由非特异性结合抗血清至该表面上所造成的背景。

结合抗原物质至小孔上以后，以非反应性物质包被以降低背景并且清洗以去除未结合的物质，以有利于免疫复合物（抗原/抗体）形成的方式将固定的表面与待测的抗血清或临床或生物学提取物接触。该条件优选包括以稀释剂如 BSA、牛 γ 球蛋白（BGG）和磷酸缓冲盐（PBS）/ Tween[®] 稀释该抗血清。这些加入的试剂也趋向于有助于降低非特异性背景。然后使分层的抗血清在优选约 25 °C 到约 27 °C 水平的温度孵育约 2
15 到 4 小时。孵育后，清洗该接触抗血清的表面从而去除非免疫复合的物质。优选的清洗方法包括以溶液如 PBS/ Tween[®] 或硼酸盐缓冲液清洗。

在测试样品与结合抗原之间形成特异性免疫复合物并且随后清洗后，可通过将此样品与对第一级抗体具有特异性的二级抗体接触而测定免疫复合物形成的频率且甚至是形成的量。为了提供检测手段，该二级抗体将优选与酶结合，当与适当的生色底物孵育时该酶将产生一定的颜色。因而，例如，人们希望在有利于免疫复合物形成的条件（如在含有
25 PBS 的溶液如 PBS - Tween[®] 中于室温孵育 2 小时）下接触并孵育结合抗血清的表面与脲酶或过氧化物酶 - 交联的抗 - 人 IgG 一段时间。

与二级酶标抗体孵育并且随后清洗以去除未结合的物质后，通过在作为酶标的过氧化物酶存在下与生色底物如尿素和溴甲酚紫或 2, 2' - 连
30 氮 - 双 - (3 - 乙基 - 苯噻唑啉) - 6 - 磺酸 (ABTS) 和 H₂O₂ 孵育加以定量。然后通过，如，用可见光分光光度计测定生色的程度而完成定

量。

本发明的抗-晶体蛋白抗体对于通过免疫沉淀分离其它的晶体蛋白抗原特别有用。免疫沉淀包括从复合物的混合物中分离靶抗原成分，并且用于区分或分离微量的蛋白。对于膜蛋白的分离，必须将细胞溶解成去污剂微团。非离子盐是优选的，因为其它试剂如胆盐在酸性 pH 或在二价阳离子存在下沉淀。

在候选的实施方案中，本发明的抗体对紧密并联二种抗原是有用的。这对于增加特定位置抗原的浓度，如酶底物对是特别有用的。

10 2.4 Western 印迹

本发明的组合物在免疫印迹或 western 印迹分析中将会发现有巨大的应用价值。该抗-肽的抗体可用作鉴定固定到固相支持基质，如硝酸纤维素、尼龙或其组合上蛋白的高亲和一级试剂。与免疫沉淀连用，然后凝胶电泳，这些可用作用于检测抗原中的单一步骤的试剂，其中的用于检测抗原的二级试剂导致相反的背景。当研究过的抗原为免疫球蛋白（排除使用结合免疫球蛋白的细菌细胞壁成分），这些研究过的抗原与检测试剂起交叉反应或者它们作为交叉-反应信号以相同的相对分子量进行迁移时这将特别有用。

用于与 Western 印迹交联的基于免疫学的检测方法包括认为抗毒素成分的酶、放射标记或荧光标记的二级抗体在此方面有特别的使用。

2.5 表位核心序列

本发明也涉及无完整细胞和其它肽的蛋白或肽组合物，其包括掺入在免疫学上与一种或更多种抗-晶体蛋白抗体交叉反应的表位的纯化蛋白或肽。特别地，本发明涉及来自 Cry 蛋白或肽的表位核心序列。

如这里所使用的，术语“掺入在免疫学上与一种或更多种抗-晶体蛋白抗体交叉反应的表位”意指包括与位于晶体蛋白或多肽内部的表位类似的一级、二级或三级结构的肽或蛋白抗原。类似的水平将一般达到这样的水平，即抗该晶体蛋白或多肽的单克隆或多克隆抗体将同时与交叉反应肽或蛋白抗原结合，起反应或对其进行识别。各种免疫分析方法，如 Western 印迹、ELISA、RIA 等可与这些抗体联合使用，这些方法

对于本领域中的技术人员来说是已知的。

适用于疫苗中的 Cry 免疫显性表位和/或其功能等价物的鉴定是相对简单的事情。例如，人们可以利用美国专利 4, 554, 101（在此引入供参考）中讲述的 Hopp 方法，该方法讲述了根据氨基酸序列的亲水性鉴定和制备表位。在几篇其它论文中所述的方法和基于其的软件程序也可用于鉴定表位核心序列（参见，例如，Jameson 和 Wolf，1988；Wolf 等，1988；美国专利。4, 554, 101）。然后可经过用肽合成或重组技术容易地将这些“表位核心序列”的氨基酸序列掺入到肽中。

根据本发明优选使用的肽将一般在约 8 到 20 个氨基酸长度，并且更为优选约 8 到 15 个氨基酸长度。已提出较短的抗原晶体蛋白 - 衍生的肽将在有些情况，例如，在准备免疫学检测分析中具有优势。优势的实例包括制备和纯化的简单易行、相对较低的成本和提高了的制备重复性以及优越的生物分布。

已提出本发明的特定优势可通过制备合成肽加以实现，该合成肽包括修饰和/或延伸的表位/免疫原性核心序列，其导致对晶体蛋白且特别是 Cry 和 Cry - 有关的序列“广泛”的表位肽。在此特定方面鉴定出这些表位核心序列为特定多肽抗原的亲水区域。已提出这些区域代表最有可能促进 T - 细胞或 B - 细胞刺激，并且，因此引发特异性抗体产生的那些区域。

这里使用的表位核心序列是相对短的氨基酸片段，其与这里公开的针对晶体蛋白的抗体上的抗原结合位点“互补”并且因此将与之结合。此外，或者作为选择，表位核心序列为将要引发与针对本发明肽组合物抗体起交叉反应的抗体的序列。应该明白在本文的上下文中，术语“互补”指相互间显示出吸引力的氨基酸或肽。因此，可根据与相应的抗 - 蛋白抗血清之间竞争结合所需蛋白抗原或者也许是代替它们之间结合的能力而可操性地定义本发明的一些表位核心序列。

一般地，只要其至少足够大以带有鉴定出的核心序列或多个序列，那么该多肽抗原的大小不认为是特别重要的。本发明预期的最小有用的核心序列一般将为大约 8 个氨基酸的长度，其中 10 到 20 个氨基酸长度的序列是优选的。因此，该大小将一般相应于根据本发明制备的最小的肽抗原。然而，只要其含有基本的表位核心序列，那么该抗原的大小在

需要时可以更大些。

表位核心序列的鉴定对于本领域的技术人员是已知的，例如，如美国专利 4, 554, 101 中所述，在此掺入供参考，其中讲述了根据氨基酸的亲水性鉴定和制备表位。然而，众多的计算机程序可用于预言蛋白的抗原部分（参见，如，Jameson 和 Wolf，1988；Wolf 等，1988）。计算机肽序列分析程序（如，DNAS[®]Star 软件，DNAS[®]Star 公司，Madison, WI）也可用于设计根据本发明的合成肽。

在其序列内部包括抗原表位的表位序列或肽的合成用常规的合成技术如固相方法可容易地完成（如，通过使用商业可得到的肽合成仪如应用生物系统 430A 型肽合成仪）。然后将从该方式合成的肽抗原以预定的量分装并且以常规方式，如水溶液或，甚至更为优选以粉末或冻干状态贮存备用。

一般地，由于肽的相对稳定性，如果需要，它们可实际上容易地贮存于任何水溶液中相当长时间，如达 6 个月或更长时间而没有可觉察到的降解或抗原活性的丧失。然而，当需要大量贮存时，其一般有必要将包括含有缓冲液如 Tris 或磷酸缓冲液的试剂从维持约 7.0 到 7.5 的 pH。此外，包括将会抑制微生物生长的试剂，如叠氮化钠或硫柳汞可能是必要的。为了在水中大量贮存，于大约 4 °C，或更为优选冰冻下贮存溶液将是必要的。当然，当这些肽以冻干或粉末状态贮存时，它们实际上可，如以固定的等份无限制地贮存，这些等份在使用前可以预定量的水（优选蒸馏过的）或缓冲液重新水化。

2.6 编码晶体蛋白嵌合体的核酸片段

本发明也涉及可以合成或从实际上任何来源分离的天然、合成和诱变的 DNA 片段，其不含总基因组 DNA 并且编码这里公开的新的嵌合肽。编码这些肽的 DNA 片段可能证明是编码晶体蛋白有关的蛋白、多肽、亚单位、功能域等或其它无关的基因产物。此外，这些 DNA 片段可用本领域技术人员熟知的方法在体外完整地合成。

如这里所使用的，术语“DNA 片段”指分离的无特定物种总基因组 DNA 的 DNA 分子。因此，编码晶体蛋白或肽的 DNA 片段指含有晶体蛋白编码序列然而是从获得该 DNA 片段的物种总基因组 DNA 中分离或纯

化的 DNA 片段，其中的基因组直接的是格兰氏阳性细菌，芽孢杆菌属的基因组，并且特别是称为苏云金芽孢杆菌的那些芽孢杆菌的基因组。术语“DNA 片段”所包括的有 DNA 片段和该片段更小的片段，并且也包括重组载体，包括，例如，质粒、粘粒、噬菌粒、噬菌体、病毒等。

5 类似地，包括分离或纯化的编码晶体蛋白基因的 DNA 片段指除了可包括肽编码序列外，还包括一些其它元件，如，实际上从其它天然基因或蛋白编码序列分离的调节序列的 DNA 片段。在该方面，术语“基因”为了简洁而使用从而指有功能蛋白-、多肽-或肽-编码单位。本领域人员将会明白，该功能术语同时包括表达或可以被加工表达蛋白、多肽或肽的基因组序列、操纵子序列和更小加工的基因片段。

10 “基本上与其它编码序列相分离”意为此时的目的基因，编码细菌晶体蛋白的基因形成了该 DNA 片段编码区的重要部分并且该 DNA 片段不含有大片的天然编码 DNA，如大的染色体片段或其它有功能的基因或操纵子编码区。当然，这是指最初分离的 DNA 片段并且不排除后来人工添加到该片段上的基因、重组基因、合成接头或编码区。

 在特定的实施方案中，本发明涉及分离的 DNA 片段和重组载体，其中掺入有编码多种 Cry 肽的 DNA 序列，这些 Cry 肽的氨基酸序列内包括基本上由 SEQ ID NO: 10、SEQ ID NO: 12、SEQ ID NO: 14、SEQ ID NO: 26、SEQ ID NO: 28、SEQ ID NO: 30、SEQ ID NO: 34 中列出的氨基酸序列。

20 术语“基本上由 SEQ ID NO: 10、SEQ ID NO: 12、SEQ ID NO: 14、SEQ ID NO: 26、SEQ ID NO: 28、SEQ ID NO: 30 或 SEQ ID NO: 34 中列出的序列”意为该序列基本上相应于 SEQ ID NO: 10、SEQ ID NO: 12、SEQ ID NO: 14、SEQ ID NO: 26、SEQ ID NO: 28、SEQ ID NO: 30 或 SEQ ID NO: 34 序列的一部分并且与这些序列中的任何一个序列的氨基酸具有很少不相同、生物学功能相当的氨基酸。术语“生物学功能相当”在本领域中众所周知并在此进一步详细描述(如，参见说明性实施方案)。因此，与 SEQ ID NO: 10、SEQ ID NO: 12、SEQ ID NO: 14、SEQ ID NO: 26、SEQ ID NO: 28、SEQ ID NO: 30 或 SEQ ID NO: 34 的氨基酸具有约 70%到 80%，或更为优选具有 81%到 90%，或甚至更为优选具有约 91%

到 99 % 氨基酸序列一致性或功能等价的序列将是“基本上由 SEQ ID NO: 10、SEQ ID NO: 12、SEQ ID NO: 14、SEQ ID NO: 26、SEQ ID NO: 28、SEQ ID NO: 30 或 SEQ ID NO: 34 所列出的”序列。

5 也将明白该氨基酸和核酸序列可以包括额外的残基，如额外的 N - 或 C - 末端氨基酸或 5' 或 3' 序列，并且只要该序列达到上述标准，包括当蛋白表达时维持蛋白的生物学活性，那么该序列仍然为基本上由这里公开的其中一个序列所列出。添加末端序列特别应用于这样的核酸序列，即，例如其可包括在该编码区 5' 或 3' 部分的各种侧翼非编码序列
10 或可包括各种内部序列，即已知在基因内部出现的内含子。

 本发明的核酸片段，无论其本身编码序列有多长，都可与其它 DNA 序列，如启动子、多腺苷信号、额外的限制酶位点、多克隆位点、其它的编码片段等连接从而它们的整体长度可能有相当大变化。因此，应注意
15 到可以使用几乎任何长度的核酸片段，其总长度优选由在目的重组 DNA 方法中制备和使用的简便性而加以确定。例如，可制备这样的核酸片段，即其包括编码在 SEQ ID NO: 10、SEQ ID NO: 12、SEQ ID NO: 14、SEQ ID NO: 26、SEQ ID NO: 28、SEQ ID NO: 30 或 SEQ ID NO: 34 中公开的任一肽序列的短连续片段，或与其与编
20 码 SEQ ID NO: 10、SEQ ID NO: 12、SEQ ID NO: 14、SEQ ID NO: 26、SEQ ID NO: 28、SEQ ID NO: 30 或 SEQ ID NO: 34 中公开的任何一种肽的 DNA 序列相同或互补，且特别是与 SEQ ID NO: 9、SEQ ID NO: 11、SEQ ID NO: 13、SEQ ID NO: 25、SEQ ID NO: 27、SEQ ID NO: 29 或 SEQ ID NO: 33 中公开的那些 DNA
25 片段相同或互补。例如，约 14 个核苷酸，并且达约 10,000、约 5,000、约 3,000、约 2,000、约 1,000、约 500、约 200、约 100、约 50 和约 14 个碱基对长度（包括所有居间长度）的 DNA 序列也被认为是有用的。

 将容易理解上下文中的“居间长度”意为引用范围之间的任何长度，如 14、15、16、17、18、19、20 等； 21、22、23 等； 30、31、
30 32 等； 50、51、52、53 等； 100、101、102、103 等； 150、151、152、153 等； 包括 200 - 500 的所有整数； 500 - 1, 000； 1, 000 - 2,000； 2, 000 - 3, 000； 3, 000 - 5, 000； 并且达到并包括约 10, 000

个核苷酸的序列等。

也应认为本发明不限于编码本发明肽的特定核酸序列或编码 SEQ ID NO: 10、SEQ ID NO: 12、SEQ ID NO: 14、SEQ ID NO: 26、SEQ ID NO: 28、SEQ ID NO: 30 或 SEQ ID NO: 34 氨基酸序列的特定核酸序列，包括特别是在 SEQ ID NO: 9、SEQ ID NO: 11、SEQ ID NO: 13、SEQ ID NO: 25、SEQ ID NO: 27、SEQ ID NO: 29 或 SEQ ID NO: 33 中公开的那些 DNA 序列。重组载体和分离的 DNA 片段可能因此多样性地包括肽编码区本身、在碱基编码区域中具有选定改变或修饰的编码区，或者它们可编码只不过包括这些肽编码区的较大多肽或者可编码具有变异氨基酸序列的生物学上有功能的等价蛋白或肽。

本发明的 DNA 片段包括生物学上有功能的等价肽。这些序列可能是由已知天然出现于核酸序列内和其编码的蛋白内的密码子多余和功能等价的后果所致。作为选择，可通过应用重组 DNA 技术产生功能上相当的蛋白或肽，其中可根据被交换氨基酸的特性加工该蛋白结构中的改变。人工设计的改变可经应用定点诱变技术而导入从而，如，导入对该蛋白抗原性的改进或为了在分子水平检测活性而测试突变子。

如果必要，人们也可制备融合蛋白和肽，如，有时将该肽编码区与其它具有所需功能，如用于纯化或免疫检测目的（如，蛋白可分别通过亲和层析和酶标编码区加以纯化）的其它蛋白或肽位于相同的表达单位中。

重组载体形成本发明进一步的方面。特别有用的载体认为是其中的 DNA 片段的编码区，无论编码全长蛋白或更小的肽，位于启动子控制下的那些载体。该启动子可以为与编码本发明肽基因天然连接的启动子形式，其可，例如，用重组克隆和/或 PCR™ 技术联合这里公开的组合物通过分离位于该编码片段或外显子上游的 5' 非-编码序列而获得。

2.7 重组载体和蛋白表达

在其它实施方案中，已注意到将编码 DNA 片段置于重组、或异源启动子控制下将会获得一些优势。正如这里所使用的，重组或异源启动子意指通常在自然环境中不与编码晶体蛋白或肽的 DNA 片段连接。这些启

5 动子可以包括通常与其它基因相连接的启动子和/或从任何细菌、病毒、真核或植物细胞中分离的启动子。一般地，利用在选定用于表达的细胞类型、有机体或甚至是动物中有效指导 DNA 片段表达的启动子将是重要的。启动子和用于蛋白表达的细胞类型组合的使用对于分子生物学领域

10 中的技术人员一般是已知的，例如，参见 Sambrook 等，1989。利用的启动子可以是组成性的或可诱导的，并且可在适当的条件下使用以指导导入 DNA 片段的高水平表达，这在大规模制备重组蛋白或肽中是有利的。打算用于高水平表达的适当启动子系统包括但不限于毕赤酵母表达载体系统（发玛西亚 LKB 生物技术公司）。

15 在与表达实例连接以制备重组蛋白和肽的过程中，应该考虑到较长的 DNA 片段将是最常使用的，其中编码完整肽序列的 DNA 片段最为优选。然而，应该明白使用较短的 DNA 片段以指导晶体肽或表位核心区域的表达从而可用于产生抗晶体蛋白抗体，这也在本发明范围之内。认为编码约 8 到 50 个氨基酸长度、或更为优选地，从约 8 到 30 个氨基酸长度，或甚至更为优选地，从约 8 到约 20 个氨基酸长度肽抗原的 DNA 片

20 段特别有用。这样的肽表位可以是包括来自 SEQ ID NO: 10、SEQ ID NO: 12、SEQ ID NO: 14、SEQ ID NO: 26、SEQ ID NO: 28、SEQ ID NO: 30 或 SEQ ID NO: 34 的连续氨基酸序列的氨基酸序列或是由 SEQ ID NO: 9、SEQ ID NO: 11、SEQ ID NO: 13、SEQ ID NO: 25、SEQ ID NO: 27、SEQ ID NO: 29、SEQ ID NO: 33 的核酸序列所编码的任何肽表位。

25 用于重组表达晶体蛋白的方法和用于表达编码晶体蛋白的 DNA 构建体的载体描述于国际专利申请公开 No. WO 95/02058，在此特异性掺入供参考。

2.8 重组宿主细胞

30 已根据布达佩斯条约的条款保藏的本发明重组宿主细胞列于表 2 中。

表 2
具有布达佩斯条约 NRRL 的保藏的菌株

| 菌株 | 质粒 | 许可号 | 保藏日期 |
|----------|----------|----------------|------------------|
| EG 11063 | pEG 1068 | NRRL B - 21579 | 1996 年 6 月 26 日 |
| EG 11074 | pEG 1077 | NRRL B - 21580 | 1996 年 6 月 26 日 |
| EG 11091 | pEG 1092 | NRRL B - 21780 | 1997 年 5 月 21 日 |
| EG 11092 | pEG 1093 | NRRL B - 21635 | 1996 年 11 月 14 日 |
| EG 11735 | pEG 365 | NRRL B - 21581 | 1996 年 6 月 26 日 |
| EG 11751 | pEG 378 | NRRL B - 21636 | 1996 年 11 月 14 日 |
| EG 11768 | pEG 381 | NRRL B - 21781 | 1997 年 5 月 21 日 |

2.9 作为杂交探针和引物的 DNA 片段

除了其在指导本发明晶体蛋白或肽表达中的使用外，在此涉及的核酸序列也具有多种其它的用途。例如，它们也可用作核酸杂交实施方案中的探针或引物。尽管如此，但应注意到包括由至少 14 个核苷酸长的连续序列所组成序列区域的核酸片段将会有特别的用途，其中的连续序列与 SEQ ID NO: 9、SEQ ID NO: 11、SEQ ID NO: 13、SEQ ID NO: 25、SEQ ID NO: 27、SEQ ID NO: 29 或 SEQ ID NO: 33 中的 14 个核苷酸长的连续 DNA 片段具有相同的序列或与其互补。同样，编码来自 SEQ ID NO: 10、SEQ ID NO: 12、SEQ ID NO: 14、SEQ ID NO: 26、SEQ ID NO: 28、SEQ ID NO: 30 或 SEQ ID NO: 34 的至少 6 个氨基酸连续序列的核酸片段也是优选的。更长连续且相同或互补的序列，如，约 20、30、40、50、100、200、500、1000、2000、5000、10000 等（包括所有居间长度并且达到和包括全长序列的那些序列也将在有些实施方案中是有用的。

这样的核酸探针与晶体蛋白 - 编码序列特异性杂交的能力将使其能够用于检测给定样品中互补序列的存在。然而，又发现其它的用途，包括用该序列信息制备突变种类的引物或用于制备其它遗传构建体的引物。

具有由与 SEQ ID NO: 9、SEQ ID NO: 11、SEQ ID NO: 13、SEQ ID NO: 25、SEQ ID NO: 27、SEQ ID NO: 29 或 SEQ ID

NO: 33 DNA 序列相同或互补的大约 10 - 14、15 - 20、30、50、或甚至 100 - 200 个核苷酸的连续核苷酸片段所组成序列区域的核酸分子被特别认为用作，如 Southern 和 Northern 印迹中的杂交探针。较小的片段一般将在杂交实施方案中具有用途，其中的连续互补区域的长度可以改变，如在约 10 - 14 到约 100 或 200 个核苷酸之间，但是根据人们希望检测的互补序列的长度可以使用较大的连续互补片段。

当然，也可通过其它技术，如，机械剪切或限制酶消化而获得这些片段。小的核酸片段或节段可容易地通过，例如，化学方法直接合成该片段而加以制备，这可用自动寡核苷酸合成仪常规地完成。同样，可通过应用核酸复制技术，如美国专利 4,683,195 和 4,683,202（在此特异性地掺入每例供参考）的 PCR™ 技术、将选定的序列导入用于重组制备的重组载体中和分子生物学领域中技术人员通常已知的其它重组 DNA 技术获得这些片段。

因此，可以使用本发明核苷酸序列与互补 DNA 片段选择性形成双链分子的能力。根据观察到的应用，人们希望利用不同的杂交条件来完成不同程度选择性检测靶序列。对于要求高选择性的应用，人们一般将希望利用相对严格的条件以形成杂交体，如，人们将选择相对低盐和/或高温条件，如约 0.02M 到 0.15M NaCl，约 50 °C 到 70 °C 的温度条件。该选择性条件耐受，如果有的话是极少的探针与模板或靶链之间的错配，并且将特别适用于分离编码晶体蛋白的 DNA 片段。通过杂交检测 DNA 片段对本领域的技术人员是众所周知的，并且美国专利 4,965,188 和 5,176,995（在此特异性地掺入每例供参考）的讲义是这些杂交分析方法的实例。特别有关的讲义，如在 Maloy 等，1994；Segal 1976；Prokop, 1991；和 Kuby, 1994 的文章中出现的内容。

当然，对于有些应用，例如，当人们希望用杂交到潜在(underlying)模板上的突变引物链制备突变体或寻求从有关种类、功能等价物等中分离编码晶体蛋白的序列时，为了形成异源双链一般将需要较少严格的杂交条件。在这些情况中，人们可能希望利用这样的条件，如约 0.15M 到 0.9M 的盐，约 20 °C 到 55 °C 的温度范围。因此，相对于对照杂交，交叉杂交的种类可容易地鉴定为阳性杂交信号。在任何情况，一般应该明白这些条件可通过加入逐渐增加量的甲醛而变得更为严格，其中的甲醛用

于随着温度的增加以相同的方式使杂交双链变得不稳定。因此，杂交条件可以容易地进行操作，并因此一般将根据所需的结果成为最佳方法。

在有些实施方案中，联合使用本发明的核酸序列与适当的方法，如用于检测杂交的探针将是优越的。多种适当的指示方式在本领域中是已知的，包括荧光、放射性、酶或其它配体，如亲和素/生物素，它们能够得到可检测的信号。在优选的实施方案中，人们将很可能希望利用荧光标记或酶标记，如脲酶、碱性磷酸酶或过氧化物酶代替放射性或其它对环境不利的试剂。在酶标记的案例中，比色指示底物是已知的。它们可用于提供对人眼可见的手段或分光光度的手段从而鉴定出与含有互补核酸样品之间的特异性杂交。

一般地，已观察到这里所述的杂交探针将可同时用作溶液杂交中的试剂以及利用固相的实施方案中。在包括固相的实施方案中，将测试 DNA（或 RNA）吸附或附着到选定的基质或表面上。然后将该固定的单链核酸在所需的条件下与选定的探针进行特异性杂交。根据所需的特定标准（例如，依据 G + C 含量、靶核酸的类型、核酸的来源、杂交探针的大小等），此选定的条件将依赖于特定的情况。清洗杂交表面以去除非特异性结合的探针分子后，检测特异性杂交，或者甚至通过标记方式加以定量。

2.10 生物学功能等价物

在本发明肽结构或编码它们的 DNA 片段中可造成修饰和改变并且仍然获得编码具有所需特征蛋白或肽的有功能的分子。下面是基于改变蛋白的氨基酸以产生等价物，或者甚至是改进的、二代分子的讨论。在本发明特定的实施方案中，认为突变的晶体蛋白对于增强该蛋白的杀虫活性并且因此增强该重组转基因在植物细胞中的杀虫活性和/或表达是有用的。氨基酸改变可根据表 3 中列出的密码子改变该 DNA 序列的密码子来完成。

表 3

| 氨基酸 | | 密码子 | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|--|
| 丙氨酸 | Ala | A | GCA | GCC | GCG | GCU | | | |
| 半胱氨酸 | Cys | C | UGC | UGU | | | | | |
| 天冬氨酸 | Asp | D | GAC | GAU | | | | | |
| 谷氨酸 | Glu | E | GAA | GAG | | | | | |
| 苯丙氨酸 | Phe | F | UUC | UUU | | | | | |
| 甘氨酸 | Gly | G | GGA | GGC | GGG | GGU | | | |
| 组氨酸 | His | H | CAC | CAU | | | | | |
| 异亮氨酸 | Ile | I | AUA | AUC | AUU | | | | |
| 赖氨酸 | Lys | K | AAA | AAG | | | | | |
| 亮氨酸 | Leu | L | UUA | UUG | CUA | CUC | CUG | CUU | |
| 甲硫氨酸 | Met | M | AUG | | | | | | |
| 天冬酰胺 | Asn | N | AAC | AAU | | | | | |
| 脯氨酸 | Pro | P | CCA | CCC | CCG | CCU | | | |
| 谷氨酰胺 | Gln | Q | CAA | CAG | | | | | |
| 精氨酸 | Arg | R | AGA | AGG | CGA | CGC | CGG | CGU | |
| 丝氨酸 | Ser | S | AGC | AGU | UCA | UCC | UCG | UCU | |
| 苏氨酸 | Thr | T | ACA | ACC | ACG | ACU | | | |
| 缬氨酸 | Val | V | GUA | GUC | GUG | GUU | | | |
| 色氨酸 | Trp | W | UGG | | | | | | |
| 酪氨酸 | Tyr | Y | UAC | UAU | | | | | |

例如，在蛋白结构中有些氨基酸可以用其它氨基酸代替而没有可觉察到的与一些结构，如抗体的抗原结合区或底物分子上的结合位点之间相互结合能力的丧失。由正是蛋白的这种相互间作用能力和性质决定了该蛋白的生物学功能活性，故可以在蛋白序列，和，当然其潜在的 DNA 编码序列中造成一些氨基酸序列替代但是可得到具有同样特性的蛋白。因此，本发明者考虑到了在此公开的组合物肽序列或编码所述肽的相应 DNA 序列中可造成各种变化而其生物学用途或活性没有可觉察到的表

失。

在进行这样的变化中，可以考虑氨基酸的亲水性指数。亲水性氨基酸指数在赋予蛋白相互间生物功能中的重要性在本领域中已得到一致认同（Kyte 和 Doolittle, 1982, 在此掺入供参考）。现已接受氨基酸的相对亲水特征赋予所得到蛋白的二级结构，其反过来又决定了该蛋白与其它分子，例如，酶、底物、受体、DNA、抗体、抗原等之间的相互作用。

根据其疏水性和电荷特征，每种氨基酸均被赋予一亲水指数（Kyte 和 Doolittle, 1982），它们是：异亮氨酸（+ 4.5）；缬氨酸（+ 4.2）；亮氨酸（+ 3.8）；苯丙氨酸（+ 2.8）；半胱氨酸/胱氨酸（+ 2.5）；甲硫氨酸（+ 1.9）；丙氨酸（+ 1.8）；甘氨酸（- 0.4）；苏氨酸（- 0.7）；丝氨酸（- 0.8）；色氨酸（- 0.9）；酪氨酸（- 1.3）；脯氨酸（- 1.6）；组氨酸（- 3.2）；谷氨酸（- 3.5）；谷氨酰胺（- 3.5）；天冬氨酸（- 3.5）；天冬酰胺（- 3.5）；赖氨酸（- 3.9）；和精氨酸（- 4.5）。

在本领域中已知有些氨基酸可由具有类似亲水指数或得分的其它氨基酸所代替并且仍可得到具有类似生物学活性的蛋白，即，仍然获得生物功能相当的蛋白。在制备这样的改变中，替代其亲水指数在 ± 2 以内的氨基酸是优选的，替代其亲水指数在 ± 1 以内的那些氨基酸是特别优选的，并且替代那些亲水指数在 ± 0.5 以内的那些氨基酸甚至是更为特别优选的。

也应明白在本领域中替代类似的氨基酸可根据疏水性而有效地完成。美国专利 4, 554, 101（在此掺入供参考）叙述了由其邻近氨基酸亲水性所支配的蛋白最大局部平均亲水性与该蛋白的生物学特性相关。

如在美国专利 4, 554, 101 中所详述的，下面的疏水值被赋予每个氨基酸：精氨酸（+ 3.0）；赖氨酸（+ 3.0）；天冬氨酸（+ 3.0 \pm 1）；谷氨酸（+ 3.0 \pm 1）；丝氨酸（+ 0.3）；天冬酰胺（+ 0.2）；谷氨酰胺（+ 0.2）；甘氨酸（0）；苏氨酸（- 0.4）；脯氨酸（- 0.5 \pm 1）；丙氨酸（- 0.5）；组氨酸（- 0.5）；半胱氨酸（- 1.0）；甲硫氨酸（- 1.3）；缬氨酸（- 1.5）；亮氨酸（- 1.8）；异亮氨酸（- 1.8）；酪氨酸（- 2.3）；苯丙氨酸（- 2.5）；色氨酸（- 3.4）。

应该明白氨基酸可由具有类似疏水值的其它氨基酸所代替并且仍然获得生物学上相当的，且特别是免疫学上相当的蛋白。在这些改变中，替代其疏水值在 ± 2 以内的氨基酸是优选的，替代其疏水值在 ± 1 以内的氨基酸是特别优选的，并且替代其疏水值在 $+ 0.5$ 以内的氨基酸是甚至更为特别优选的。

如上面所总结的，因此氨基酸替代一般是根据氨基酸侧-链替代物的相对类似性，例如疏水性、亲水性、电荷、大小等。考虑到前述各种特征的替代的实例对本领域的技术人员是众所周知的并且包括：精氨酸与赖氨酸；谷氨酸与天冬氨酸；丝氨酸与苏氨酸；谷氨酰胺与天冬酰胺；以及缬氨酸、亮氨酸与异亮氨酸。

2.11 定点诱变

定点诱变是经过特异性诱变潜在 DNA 而用于制备单个肽、或生物学功能相当的蛋白或肽的技术。该技术进一步提供通过导入一种或更多种核苷酸序列改变至该 DNA 中而制备和测试，例如，掺入一种或更多种前述因素的序列突变体的即用能力。定点诱变使能够经过使用特异性的寡核苷酸序列而制备突变体，其中的寡核苷酸序列编码所需突变的 DNA 序列以及足够数目的相邻核苷酸以提供足够大小的引物序列和序列复杂度从而在跨越缺失接头处的二侧形成稳定的双链。一般地，约 17 到 25 个核苷酸长度的引物是优选的，其中待改变序列接头的二侧具有约 5 到 10 个残基。

一般来说，定点诱变技术在本领域中是众所周知的，如各种出版物中所列举的那样。应当明白，该技术一般利用既表现出单链又表现出双链形式的噬菌体载体。用于定点诱变的典型载体包括载体如 M 13 噬菌体。该噬菌体在商业上容易得到并且其使用对于本领域的技术人员是众所周知的。双链质粒也经常用于定点诱变从而免除了将目的基因从质粒转移至噬菌体的步骤。

一般地，根据这里的定点诱变通过首先获得单链载体或使其序列中包括编码所需肽的 DNA 序列的双链载体二条链解链而完成。制备，一般是合成制备具有所需突变序列的寡核苷酸引物。然后将该引物与单链载体退火，并且用 DNA 聚合酶如大肠杆菌聚合酶 I Klenow 片段以完

成具有突变链的合成。这样，形成了异源双链，其中的一条链编码原初的非突变序列并且第二条链具有所需的突变。然后将此异源双链载体用于转化合适的细胞，如大肠杆菌细胞，并且筛选出包括有突变序列组合的重组载体的克隆。

5 用定点诱变制备编码选定肽的 DNA 片段的序列体将提供为制备十分有用物种的手段并且并非意在只限于此因为还有其它方法，在这些方法中可获得肽的序列突变体和编码它们的 DNA 序列。例如，可以用诱变剂，如羟胺处理编码所需肽序列的重组载体以获得序列变异体。

10 2.12 作为杀虫剂的晶体蛋白组合物和使用方法

本发明人考虑到在此公开的嵌合晶体蛋白组合物作为局部和/或系统施用至田间作物、草、果树和蔬菜以及观赏植物的杀虫剂将会有特别的用途。在优选的实施方案中，该生物杀虫剂组合物包括表达这里公开的新晶体蛋白细菌细胞的油性可流动的悬液。优选地，这些细胞为苏云
15 金芽孢杆菌细胞，然而，认为表达这里公开的新的核酸片段并且产生晶体蛋白的任何这样的细菌宿主细胞均是有用的，如巨大芽孢杆菌、枯草芽孢杆菌、大肠杆菌或假单胞菌。

在另一个重要的实施方案中，该生物杀虫剂组合物包括可在水中分散的微粒。该微粒包括表达这里公开的新的晶体蛋白的细菌细胞。优选
20 的细菌细胞为苏云金芽孢杆菌细胞，然而，以这里公开的 DNA 片段转化并且表达晶体蛋白的细菌如巨大芽孢杆菌、枯草芽孢杆菌、大肠杆菌或假单胞菌细胞也认为是有用的。

在第三个重要的实施方案中，该生物杀虫剂组合物包括可增湿的粉末、粉末（dust）、颗粒（pellet）或胶态浓缩物。该粉末包括表达这
25 里公开的新的晶体蛋白的细菌细胞。优选的细菌细胞为苏云金芽孢杆菌细胞，然而，以这里公开的 DNA 片段转化并且表达该晶体蛋白的细菌如巨大芽孢杆菌、枯草芽孢杆菌、大肠杆菌或假单胞菌也认为是有用的。可制备这种干燥形式的杀虫组合物从而在增湿时立即溶解，或者作为选择，以控释、缓释或其它时间依赖性的方式溶解。

30 在第四个重要的实施方案中，该生物杀虫剂组合物包括细菌细胞如上述的表达该晶体蛋白的那些细菌细胞的水悬液。此水悬液可提供为浓

缩的贮备液，在施用前稀释、或者作为选择，提供为即用的稀释液。

对于包括施用细菌细胞的这些方法，含有该晶体蛋白基因的细胞宿主可培养于任何方便的营养培养基中，其中的 DNA 构建体提供了选择优势，提供选择培养基从而基本上全部或所有的细胞均带有苏云金芽孢杆菌基因。然后根据常规方式收获这些细胞。作为选择，可在收获之前处理这些细胞。

当该杀虫剂组合物包括表达目的蛋白的完整苏云金芽孢杆菌细胞时，可以多种方法制备这种细菌。通过与各种惰性材料，如无机矿物质（页硅酸盐、碳酸盐、硫酸盐、磷酸盐等）或植物材料（粉末状玉米穗轴、稻壳、板栗壳等）混合，它们可用作可增湿的粉末、微粒或粉末。这些制品可包括分散-粘着佐剂、稳定剂、其它杀虫添加剂或表面活性剂。液体制剂可以为基于水或无水的形式并且可用作泡沫、悬液、可乳化的浓缩物等。这些成分可包括流变剂、表面活性剂、乳化剂、分散剂或聚合物。

通过重组细菌表达系统在体外可制备新的嵌合 Cry 蛋白并且分离用于随后的田间施用。该蛋白可以位于粗细胞裂解液、悬液、胶体等中或在配制成活性的生物杀虫剂制剂之前可被纯化、提炼、缓冲和/或进一步加工。同样，在有些情况下，从表达该晶体蛋白的细菌培养物中分离晶体和/或孢子并且施用作为活性生物杀虫剂组合物的该晶体和/或孢子的溶液、悬液或胶体制品可能是必要的。

无论施用方法如何，以杀虫有效的量施用该活性组合物，这个量将依赖于这些因素，如待控制的具体的鞘翅目昆虫、待处理的具体植物或作物、环境条件和施用该杀虫-活性组合物的方法、速度和量。

所述的杀虫剂组合物可通过配制该细菌细胞、晶体和/或孢子悬液、或分离的蛋白组分与所需的农业上可接受的载体而完成。该组合物可在施用之前以适当的方式配制，如冷冻干燥、冰冻干燥、脱水或位于水性载体、介质或合适的稀释剂，如盐或其它缓冲液。该配制的组合物可以为粉末或微粒物质的形式、或油（蔬菜或矿物油）中的悬液的形式、或者水或油/水乳剂的形式、或者为可增湿的粉末形式或者与任何其它适于农业施用的载体材料联合使用的形式。合适的农业载体可以为固体或液体并且在本领域中是众所周知的。术语“农业上可接受的载体”包括所

有的佐剂，如，通常用于杀虫剂制备技术中的惰性成分、分散剂、表面活性剂、粘剂（tackifiers）、粘合剂等；这些对于杀虫剂制备的技术人员来说是众所周知的。该制剂可与一种或更多种固体或液体佐剂混合并且用常规的制备技术通过各种方法，如均匀混合、混和（blend）和/或研磨该杀虫组合物与合适的佐剂加以制备。

本发明的杀虫剂组合物施用至目标鞘翅目昆虫的环境中，一般通过常规方法，优选通过喷洒施用至该植物或要保护的作物的叶片上。杀虫剂施用的强度和um时间将依据待处理的特定害虫、作物的具体状况和特定的环境条件来确定。活性成分对载体的比例将一般依赖于该杀虫剂组合物的化学性质和稳定性以及所涉及的特定配方。

在某些情况下，如导致根部或茎部感染的昆虫或施用至纤弱的植物或观赏植物，其它施用技术，如，喷粉、喷灌、浸泡、土壤注射、种子涂层（coating）、幼苗涂层、喷洒、通气（aerating）、喷雾（misting）、雾化（atomizing）等也是可行的并且可能是必要的。这些施用方法对于本领域中的技术人员也是众所周知的。

本发明的杀虫剂组合物可以仅仅用于本发明方法中或者与其它化合物，包括但不限于其它杀虫剂联合使用。本发明方法也可与其它处理方案，如表面活性剂、去污剂、聚合物或时间-释放制剂联合使用。本发明的杀虫剂组合物可以配制成系统或局部使用。

用于环境、系统、或叶片施用的杀虫剂组合物的浓度将依据特定制剂、施用方式、环境条件和生物杀虫活性的程度而有较大变化。一般地，该生物杀虫剂组合物将以至少约 0.5 % 重量的浓度存在于施用的制剂中并且可达到并包括约 99 % 的重量。该组合物的干制剂可以是约 0.5 % 到 99 % 或更多该组合物的重量，而液体制剂可一般包括约 0.5 % 到 99 % 或更多该活性成分的重量。包括完整细菌细胞的制剂将一般含有约 10^4 到 10^{12} 个细胞/mg。

该杀虫剂制剂根据需要可以一次或更多次施用至特定的植物或靶区域，每公顷的典型田间施用量在约 50 g 至约 500 g 活性成分的范围，或者约 500 g 到约 1000 g、或者约 1000 g 到约 5000 g 或更多活性成分的范围。

2.13 抗体组合物和制备方法

在特定的实施方案中，本发明者考虑到了与这里公开的晶体蛋白结合的单克隆或多克隆抗体的使用。用于制备这些抗体并对其进行特征分析的方法在本领域中是已知的（参见，如，Harlow 和 Lane，1988；
5 在此掺入供参考）。用于制备单克隆抗体（mAbs）的方法一般与用于制备多克隆抗体的方法以相同的路线开始。简言之，通过以根据本发明的免疫原性组合物免疫动物并且从该免疫的动物中收集抗血清而制备多克隆抗体。多种动物种类可用于抗血清的制备。一般用于制备抗血清的动物为兔子、小鼠、大鼠、仓鼠、豚鼠或山羊。由于兔子的血量较大，
10 故兔子是用于制备多克隆抗体的优选的选择。

正如在本领域中所众所周知的，一种给定的组合物的免疫原性可以改变。因此，增强宿主免疫系统常是必需的，这可通过将肽或多肽免疫原偶联至载体上而完成。优选载体的实例有匙孔血蓝蛋白（KLH）和牛血清蛋白（BSA）。其它清蛋白如卵清蛋白、小鼠血清蛋白或兔血清蛋白也可用作载体。用于将多肽交联至载体蛋白上的方法在本领域中是众所周知的并且包括戊二醛、间马来酰亚胺苯甲酰-N-羧琥珀酰亚胺酯、碳化二亚胺和双二氯化联苯胺。

正如同样在本领域中所众所周知的，可通过使用非特异性的名为佐剂的免疫应答刺激剂而增强特定免疫原组合物的免疫原性。优选佐剂的实例包括完全弗氏佐剂（含有灭活结核分枝杆菌（*Mycobacterium tuberculosis*）的免疫应答非特异性的刺激剂）、不完全弗氏佐剂和氢氧化铝佐剂。

用于制备多克隆抗体中的免疫原组合物的量依据该免疫原的性质以及用于免疫的动物而改变。可用多种途径施用免疫原（皮下、肌肉内、
25 皮内、静脉内和腹膜内）。多克隆抗体的产生可通过在免疫后不同的时间点采集免疫动物的血样加以监测。也可能要给以第二次增强注射。重复增强和滴定的过程直到达到合适的滴度。当获得所需水平的免疫原性时，将免疫的动物放血并且分离和贮存血清，且/或将该动物用于产生 mAbs。

30 通过使用众所周知的技术，如在美国专利 4,196,265（在此特异性掺入供参考）中例举的技术可容易地制备 mAbs。一般地，该技术包括

以选定的免疫原组合物，如纯化或部分纯化的晶体蛋白免疫合适的动物。以有效刺激产生抗体的细胞的方式施用该免疫组合物。啮齿动物如小鼠和大鼠是优选的动物，然而，使用兔子、绵羊、蛙细胞也是可能的。使用大鼠可能提供某些优势（Goding, 1986, 60 - 61 页），但小鼠是优选的，其中 BALB/c 小鼠是最为优选的因为其最常使用并且一般达到更高百分比的稳定融合体。

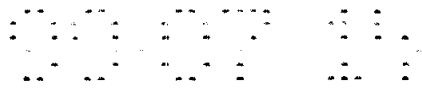
免疫后，筛选出具有产生抗体能力的体细胞，特别是 B 淋巴细胞（B 细胞）用于 mAb 的制备方法中。这些细胞可获自活检的脾脏、扁桃体或淋巴结，或者获自外周血样品。脾脏细胞和外周血细胞是优选的，前者是因为它们为处于分裂的浆母细胞期产生抗体细胞的丰富来源，并且后者是因为外周血容易得到。通常，将免疫一组动物并且取出具有最高抗体滴度的动物脾脏且然后通过用注射器搅匀脾脏而获取脾脏淋巴细胞。一般地，来自免疫小鼠的脾脏含有大约 5×10^7 到 2×10^8 个淋巴细胞。

然后将来自免疫动物的该产生抗体的 B 淋巴细胞与永生的骨髓瘤细胞，一般是与该免疫动物相同种动物的一种细胞融合。适用于产生杂交瘤融合方法中的杂交瘤细胞系优选为不产生抗体、具有高的融合效率和酶的缺失、此缺失然后使其不能够培养于一些支持仅仅所需融合细胞（杂交瘤）生长的选择培养基中。

可以使用众多骨髓瘤细胞中的任何一种，这些对于本领域的技术人员来说是已知的（Goding, pp. 65 - 66, 1986; Campbell, pp. 75 - 83, 1984）。例如，其中的免疫动物为小鼠时，人们可以使用 P 3 - X 63/Ag 8、X 63 - Ag 8.653、NS 1/1. Ag 4 1、Sp 210 - Ag 14、FO、NSO/U、MPC - 11、MPC 11 - X 45 - GTG 1.7 和 S 194/5 XX0 Bul；对于大鼠，人们可以使用 R 210. RCY 3、Y 3 - Ag 1. 2. 3、IR 983 F 和 4B 210；并且 U - 266、GM 1500 - GRG 2、LICR - LON - HMy 2 和 UC 729 - 6 均可与人类细胞融合联用。

一种优选的鼠骨髓瘤细胞为 NS - 1 骨髓瘤细胞系（也称为 P 3 - NS - 1 - Ag 4 - 1），其通过查寻细胞系保藏号 GM 3573 可容易地从 NIGMS 人类遗传突变细胞贮藏库中得到。另一种可用的小鼠骨髓瘤细胞系为不产生抗 - 8 - 氮鸟嘌呤小鼠骨髓瘤 SP 2/0 的细胞系。

用于制备产生抗体的脾脏或淋巴结细胞与骨髓瘤细胞的杂交体的方



法通常包括在一种或多种促进细胞膜融合（化学或电）试剂存在下以 2 : 1 的比例混合体细胞与骨髓瘤细胞，但该比例可分别从约 20 : 1 变化至约 1 : 1。已描述过用仙台病毒的融合方法（Kohler 和 Milstein, 1975; 1976）和用聚乙二醇（PEG），如 37 %（体积/体积）PEG（Geffer 等, 1977）的方法。使用电诱导融合方法也是适当的（Goding, 1986, pp. 71 - 74）。

融合方法通常以低频率，约 1×10^{-6} 到 1×10^{-8} 产生存活的杂交体。然而，这将不会带来麻烦，因为通过在选择培养基中培养就可从母本未融合的细胞（特别是将正常继续无限分裂的未融合骨髓瘤细胞）中分离出活的融合杂交体。选择培养基一般为含有阻止在组织培养基中核苷酸从头合成试剂的培养基。优选试剂的实例为氨基喋呤、氨基甲喋呤和重氮丝氨酸。氨基喋呤和氨基甲喋呤同时阻止嘌呤和嘧啶的从头合成，而重氮丝氨酸仅阻止嘌呤的合成。当使用氨基喋呤或氨基甲喋呤时，向该培养基中补充次黄嘌呤和胸苷作为核苷酸的来源（HAT 培养基）。当使用重氮丝氨酸时，该培养基补充以次黄嘌呤。

优选的选择培养基为 HAT。仅仅能够运行核苷酸补救途径的细胞才能够在 HAT 培养基中存活。骨髓瘤细胞缺少补救途径的关键酶，如，次黄嘌呤磷酸核糖基转移酶（HPRT）并且它们不能够存活。B - 细胞可运行该途径，但它们在培养中具有有限的生活周期并且一般在大约 2 周内死亡。因此，仅仅是可在选择培养基中存活的细胞才是由骨髓瘤和 B - 细胞所形成的那些杂交体。

该培养提供了杂交瘤种群，从中可筛选出特异性的杂交瘤。一般地，通过在微滴定板中的单克隆稀释培养细胞，然后测试该单克隆上清液所需反应性（大约 2 到 3 周后）来完成杂交瘤的筛选。此分析应该是敏感、简单而快速的，如放免分析、酶免分析、细胞毒性分析、噬菌斑分析、点印迹结合分析等。

然后将系列稀释这些筛选出的杂交瘤并且克隆成为产生单一抗体的细胞系，然后可无限扩增这些克隆以提供 mAbs。该细胞系可以二种基本的方式利用于 mAb 的制备。可将该杂交瘤样品注射（经常是进入腹腔）到组织相容性的动物类型中，该种动物是用于提供原初融合的体细胞和骨髓瘤细胞。注射后的动物形成分泌由该融合细胞杂交体产生的特

异性单克隆抗体的肿瘤。然后可排出该动物的体液，如血清或腹水以提供高浓度的 mAbs。也可在体外培养该单个细胞系，其中的 mAbs 自然分泌至培养基中，从中可容易地获取高浓度的 mAbs。如果必要，可用过滤、离心和各种层析方法如 HPLC 或亲和层析进一步纯化由这两种方法中任何一种所产生的 mAbs。

3.0 附图简述

下面的附图形成了本说明书的一部分并且被包括进是进一步说明的有些方面。通过参考一个或更多个这些附图结合这里列出的具体实施方案的详述可更好地理解本发明。

图. 1. 图 1A 中描绘了野生型 δ -内毒素和用于构建与本发明有关的杂合 δ -内毒素的有关限制性位点。显示了仅仅为编码含在所示质粒（由“pEG”前缀鉴别）中 δ -内毒素的 DNA。含有所示质粒的苏云金芽孢杆菌菌株由“EG”前缀所鉴别。本发明中所述的杂合 δ -内毒素描绘于图 1B 中并且与图 1A 中的野生型 δ -内毒素作比较。

图. 2. 通过 SDS - PAGE 分析相同量的每种清洗过的形成芽孢的苏云金芽孢杆菌培养物。a 道：产生 Cry 1 Ac 的对照苏云金芽孢杆菌菌株 EG 11070、b：EG 11060、c：EG 11062、d：EG 11063、e：EG 11065、f：EG 11067、g：EG 11071、h：EG 11073、i：EG 11074、j：EG 11088、k：EG 11090、和 l：EG 11091。

图. 3. 将溶解的杂合 δ -内毒素暴露于胰蛋白酶 0、15、30、60 和 120 分钟。通过 SDS - PAGE 分析得到的物质。通过用分子动力学 300A 型光密度计进行光密度扫描定量残余活性 δ -内毒素片段的量。残余活性毒素的百分比对时间作图。野生型 Cry 1Ac δ -内毒素（空方框）用作对照。

图. 4. 图示用于构建由 pEG 381 所编码的并且在 EG 11768 中表达的杂合 δ -内毒素的野生型毒素和有关的限制位点。仅仅显示编码含于所示质粒（由“pEG”前缀所识别）中 δ -内毒素的 DNA。

4.0 序列标识简述

SEQ ID NO: 1 为寡核苷酸引物 A。



SEQ ID NO: 2 为寡核苷酸引物 B。

SEQ ID NO: 3 为寡核苷酸引物 C。

SEQ ID NO: 4 为寡核苷酸引物 D。

SEQ ID NO: 5 为寡核苷酸引物 E。

5 SEQ ID NO: 6 为寡核苷酸引物 F。

SEQ ID NO: 7 为寡核苷酸引物 G。

SEQ ID NO: 8 为寡核苷酸引物 H。

SEQ ID NO: 9 为 EG 11063 杂合 δ -内毒素的核苷酸和推导的氨基酸序列。

10 SEQ ID NO: 10 以三-字母的简写形式表示在 SEQ ID NO: 9 中定义的杂合 δ -内毒素的氨基酸序列。

SEQ ID NO: 11 为 EG 11074 杂合 δ -内毒素的核苷酸和推导的氨基酸序列。

15 SEQ ID NO: 12 以三-字母的简写形式表示在 SEQ ID NO: 11 中定义的杂合 δ -内毒素的氨基酸序列。

SEQ ID NO: 13 为 EG 11735 杂合 δ -内毒素的核苷酸和推导的氨基酸序列。

SEQ ID NO: 14 以三-字母简写形式表示在 SEQ ID NO: 13 中定义的杂合 δ -内毒素的氨基酸序列。

20 SEQ ID NO: 15 为 pEG 1065、pEG 1070 和 pEG 1074 的 5' 交换位点。

SEQ ID NO: 16 为 pEG 1067、pEG 1072 和 pEG 1076 的 5' 交换位点。

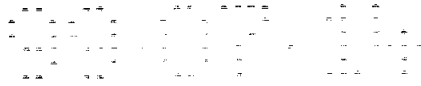
25 SEQ ID NO: 17 为 pEG 1068、pEG 1077 和 pEG 365 的 5' 交换位点。

SEQ ID NO: 18 为 pEG 1088 和 pEG 1092 的 5' 交换位点。

SEQ ID NO: 19 为 pEG 1089 的 5' 交换位点和 pEG 1070 和 pEG 1072 的 3' 交换位点。

SEQ ID NO: 20 为 pEG 1091 的 5' 交换位点。

30 SEQ ID NO: 21 为 pEG 1065、pEG 1067、pEG 1068、pEG 1093、pEG 378 和 pEG 365 的 3' 交换位点。



SEQ ID NO: 22 为 pEG 1088 的 3' 交换位点。

SEQ ID NO: 23 为寡核苷酸引物 I。

SEQ ID NO: 24 为寡核苷酸引物 J。

5 SEQ ID NO: 25 为 EG 11092 的编码杂合晶体蛋白基因的核酸序列和推导的氨基酸序列。

SEQ ID NO: 26 为由 SEQ ID NO: 25 所编码的菌株 EG 11092 所产生的杂合晶体蛋白的三-字母简写形式的氨基酸序列。

SEQ ID NO: 27 为 EG 11751 的编码杂合晶体蛋白基因的核酸序列和推导的氨基酸序列。

10 SEQ ID NO: 28 为由 SEQ ID NO: 27 所编码的菌株 EG 11751 所产生的杂合晶体蛋白的三-字母简写形式的氨基酸序列。

SEQ ID NO: 29 为 EG 11091 的编码杂合晶体蛋白基因的核酸序列和推导的氨基酸序列。

15 SEQ ID NO: 30 为由 SEQ ID NO: 29 所编码的菌株 EG 11091 所产生的杂合晶体蛋白的三-字母简写形式的氨基酸序列。

SEQ ID NO: 31 为寡核苷酸引物 K。

SEQ ID NO: 32 为 pEG 378 和 pEG 381 的 5' 交换位点。

SEQ ID NO: 33 为 EG 11768 的编码杂合晶体蛋白基因的核酸序列和推导的氨基酸序列。

20 SEQ ID NO: 34 以三-字母的简写形式表示由 SEQ ID NO: 33 所编码的菌株 EG 11768 产生的杂合晶体蛋白的氨基酸序列。

SEQ ID NO: 35 为 pEG 1074、pEG 1076、pEG 1077 和 pEG 381 的 3' 交换位点。

25 5.0 说明性实施方案的描述

5.1 用于培养苏云金芽孢杆菌以制备 Cry 蛋白的方法

可用已知标准的培养基和发酵技术培养这里所述的苏云金芽孢杆菌菌株。发酵循环完成后, 可通过本领域中众所周知的方法首先从发酵培养基中分离苏云金芽孢杆菌孢子和晶体而收获细菌。回收的苏云金芽孢
30 和晶体可通过添加表面活性剂、分散剂、惰性载体和其它成分而配制成可增湿的粉末、液体浓缩物、微粒或其它制剂以利于处理和施用至特定

的靶害虫。此配制和施用方法在本领域中是众所周知的并且与抗鳞翅目，如毛虫的活性苏云金芽孢杆菌（HD - 1）商业菌株一起使用。

5.2 用于CRY基因表达的重组宿主细胞

5 本发明的核苷酸序列可导入多种微生物宿主中。毒素基因的表达直接或间接导致了细胞内产生和具有杀虫剂。对于适宜的宿主，如假单胞菌，可将该微生物施用至鳞翅目昆虫的位点，这些昆虫将在此繁殖并且摄食这些微生物。结果是控制了不需要的昆虫。作为选择，可在延长细胞中产生毒素活性的条件下处理具有毒素基因的微生物。然后将处理的
10 细胞施用至靶害虫的环境中。得到的产物维持有苏云金芽孢杆菌毒素的毒性。

合适的宿主细胞（其中将处理含有杀虫剂的细胞以延长细胞中毒素的活性，然后将处理的细胞施用至靶害虫的环境中）可以包括原核或真核生物，一般限于不产生对高等生物，如哺乳动物具有毒性物质的那些
15 细胞。然而，当该毒素不稳定或施用水平足够低从而避免对哺乳动物宿主体任何可能性或毒性时，也可使用产生对高等生物具有毒性物质的生物。作为宿主，特别感兴趣的将是原核生物和低等的真核生物，如真菌。说明性的格兰氏 - 阴性和格兰氏 - 阳性原核生物包括肠杆菌科，如埃希氏菌属、欧文氏菌属、志贺氏菌属、沙门氏菌属、和变形菌属；芽孢杆菌科；根瘤菌科，如根瘤菌属；螺菌科，如发光杆菌属、发酵单胞菌属、沙雷氏菌属、气单胞菌属、弧菌属、脱硫弧菌属、螺菌属；乳杆菌科，假单胞菌科，如假单胞菌属和醋杆菌属；固氮菌科，放线菌目和硝化杆菌科。真核生物中有真菌，如藻状菌纲和子囊菌纲，其包括酵母如酵母属和裂殖酵母属；和担子菌纲酵母，如红酵母属、短柄霉属、掷孢酵母
20 属等。
25

在筛选用于制备宿主细胞中的特定目的特征包括导入苏云金芽孢杆菌基因至宿主中的简单易行性、表达系统的可获得性、表达的效率、宿主中杀虫剂的稳定性和辅助性遗传能力的存在。用作杀虫剂微胶囊（microcapsule）目的的特征包括对杀虫剂的保护性特征，如厚的细胞
30 壁、色素形成、和细胞内包装或形成包涵体；叶片亲和性；缺乏哺乳动物毒性；对害虫摄食的趣向性；杀死和固定而不损害毒素的简单易行性

等。其它因素包括配制和处理的简单易行性、经济性、贮存的稳定性等。

特定目的的宿主生物包括酵母，如红酵母属、短柄霉属、酵母属和
 掷孢酵母属等；叶面（*phylloplane*）生物如假单胞菌属、欧文氏菌属和
 黄杆菌属的种类；或者一些其它的生物如埃希氏菌属、乳杆菌属、芽孢
 5 杆菌属、链霉菌属等。具体的生物包括铜绿假单胞菌、荧光假单胞菌、
 啤酒糖酵母、苏云金芽孢杆菌、枯草芽孢杆菌、大肠杆菌、浅青紫链霉
 菌等。

只要其既不有害地影响毒素的特性，又不降低细胞保护该毒素的能
 力，那么可用化学或物理方法，或同时用化学和/或物理方法处理微生物
 10 细胞，如含有苏云金芽孢杆菌毒素基因的微生物。化学试剂的实例为卤
 化试剂，特别是原子序数在 17 - 80 的卤素。更为特别地，碘可在温和
 条件下和达到所需结果的足够时间下可以使用。其它合适的技术包括以
 下面的试剂处理，包括醛，如甲醛和戊二醛；抗菌剂，如苯扎氯铵和氯
 化十六烷基吡啶；醇，如异丙醇和乙醇；各种组织学固定剂，如卢戈尔
 15 碘溶液、布安固定剂和 Helly's 固定剂（参见，如，Humason, 1967）；
 或者物（热）和化学试剂的组合，当将细胞施用至适宜的宿主时它们保
 持并延长细胞中产生的毒素的活性。物理方法的实例有短波长辐射如 γ
 - 辐射和 X - 辐射、冷冻、UV 照射、冷冻干燥等。尽管在有些情况可
 以利用孢子，但利用的细胞将一般是完整的并且当被处理时基本上为增
 20 殖的形式而不是孢子的形式。

当通过合适的载体将苏云金芽孢杆菌毒素基因导入到微生物宿主
 中，并且将所述活性形式的宿主施用至环境中时，使用某些宿主微生物
 是必要的。筛选出已知占据一种或更多种目的作物“植物圈”（叶面、
 叶际（*phyllosphere*）、根际、和/或根表）的微生物宿主。筛选出这些
 25 微生物是为了在特定的环境中（作物和其它昆虫栖息地）能够与野生型
 微生物成功地竞争，提供表达该多肽杀虫剂基因的稳定的维持与表达，
 并且，最理想地是将对该杀虫剂更好的保护从而免受环境降解和失活。

已知多种微生物栖息于许多种重要作物的叶面（植物叶子表面）和/
 或根际（围绕植物根的土壤）。这些微生物包括细菌、藻类和真菌。具
 30 有特别意义的微生物有，如细菌，例如，芽孢杆菌属、假单胞菌属、欧
 文氏菌属、沙雷氏菌属、克雷伯氏菌属、*Zanthomonas*、链霉菌属、根

瘤菌属、红假单胞菌属、嗜甲基菌属、农杆菌属、醋杆菌属、乳杆菌属、
节杆菌属、固氮菌属、明串珠菌属和产碱菌属；真菌，特别是酵母、如糖
酵母属、隐球酵母属、克鲁维酵母属、掷孢酵母属、红酵母属和短柄霉
属。特别有意义的植物圈细菌种类如丁香假单胞菌、荧光假单胞菌、粘
5 质沙雷氏菌、木醋杆菌、根癌农杆菌、类球红细菌、野油菜黄单胞菌、
苜蓿根瘤菌、真氧产碱菌和棕色固氮菌；和植物圈酵母种类如深红酵母
(*Rhodotorula rubra*)、胶粘红酵母(*R. glutinis*)、海滨红酵母(*R.*
marina)、橙黄红酵母(*R. aurantiaca*)、浅白色隐球酵母(*Cryptococcus*
albidus)、流散隐球酵母(*C. diffluens*)、罗仑氏隐球酵母(*C.*
10 *laurentii*)、罗茜糖酵母(*Saccharomyces rosei*)、善地糖酵母(*S.*
pretoriensis)、啤酒糖酵母、红色掷孢酵母(*Sporobolomyces roseus*)、
香气掷孢酵母(*S. odorus*)、佛地克鲁维氏酵母(*Kluyveromyces*
veronae)和出芽短柄霉(*Aureobasidium pollulans*)。

15 5.3 定义

下面的单词和词语具有以下意义。

广 - 谱：指广泛的昆虫种类。

广 - 谱杀虫活性：对各种昆虫种类的毒性。

20 表达：细胞内过程的组合，包括由编码的 DNA 分子如结构基因的转
录和翻译以产生多肽。

杀虫活性：对昆虫的毒性。

杀虫特异性：由晶体蛋白对多种昆虫种类所显示的毒性。

目内特异性 (*Intraorder Specificity*)：特定的晶体蛋白对一个目
内 (如，鳞翅目) 昆虫种类的毒性。

25 目间特异性：特定的晶体蛋白对不同目 (如鳞翅目和双翅目) 昆虫
种类的毒性。

LC₅₀：导致处理的昆虫 50 % 死亡率的晶体蛋白的致死浓度。

LC₉₅：导致处理的昆虫 95 % 死亡率的晶体蛋白的致死浓度。

30 启动子：一个或一组 DNA 序上提供结构基因表达控制元件的识别位
点并且 RNA 聚合酶特异性地结合到该基因的此位置并起始 RNA 合成
(转录)。

再生：从植物细胞（如，植物原生质体或外植体）中长出植物的过程。

结构基因：被表达以产生多肽的基因。

5 转化：将外源 DNA 序列（如，载体、重组 DNA 分子）导入细胞或原生质体中的过程，该外源 DNA 序列掺入到细胞或原生质体的染色体中或能够自主复制。

转化的细胞：通过导入外源 DNA 分子到该细胞中而变了其 DNA 的细胞。

10 转基因：当经过如转化、电穿孔、粒子轰击等方法而导入到宿主细胞的基因组中时由该宿主细胞表达并且整合入这些细胞的基因组中从而由该转基因的表达所产生的一个或多个特征遗传至该转化细胞的后代中的外源基因。

15 转基因细胞：从转化的细胞衍生或再生的或从转基因细胞衍生的任何细胞。转基因细胞的实例包括源自转化植物细胞的植物愈伤组织和特定的细胞如叶、根、茎，如获自转基因植物的体细胞或生殖（胚芽）细胞。

20 转基因植物：源自转化植物细胞或原生质体的植物或其后代，其中的植物 DNA 含有最初不存在于天然、非-转基因相同品系的植物中的外源 DNA 分子。术语“转基因植物”和“转化的植物”有时在本领域中用作同义语以表示其 DNA 含有外源 DNA 分子的植物。然而，认为称获自转化植物细胞或原生质体的再生植物或愈伤组织为转基因植物是更为科学而准确的，并且其用途将在后面描述。

载体：能够在宿主细胞中复制并且/或可与另一种 DNA 片段可操作地连接从而导致附着片段的复制的 DNA 分子。质粒为载体的实例。

25

5.4 探针和引物

在另一方面，由本发明提供的 DNA 序列信息允许制备具有与这里公开的筛选出的多核苷酸基因序列特异性杂交能力的相对较短的 DNA（或 RNA）序列。在这些方面，根据筛选出晶体蛋白基因序列，如在 SEQ ID NO： 9、SEQ ID NO： 11、SEQ ID NO： 13、SEQ ID NO： 25、30 SEQ ID NO： 27、SEQ ID NO： 29 或 SEQ ID NO： 33 中显示出的

序列制备适当长度的核酸探针。该核酸探针与编码晶体蛋白基因序列特异性杂交的能力赋予它们在多种实施方案中特定的用途。更为重要的是，这些探针可用于多种分析中以用于测定互补序列在给定样品中的存在。

5 在一些实施方案中，使用寡核苷酸引物是优越的。用本发明的多核苷酸设计这些引物序列从而用 PCR™ 技术测定、扩增或突变苏云金芽孢杆菌晶体蛋白一定的片段。来自其它种类有关晶体蛋白基因片段也可用这些引物通过 PCR™ 加以扩增。

10 为了提供根据本发明的一些优势，用于杂交研究或分析的优选的核酸序列包括与编码晶体蛋白序列的至少 14 到 30 个左右核苷酸长度的片段互补的序列，编码晶体蛋白的序列片段如 SEQ ID NO: 9、SEQ ID NO: 11、SEQ ID NO: 13、SEQ ID NO: 25、SEQ ID NO: 27、SEQ ID NO: 29 或 SEQ ID NO: 33 中所显示，至少 14 个核苷酸长度的大小有助于保证该片段将形成稳定而选择性的双链分子的足够长度。

15 为了增加杂交体的稳定性和选择性并且因此提高获得的特异性杂交分子的质量和程度，具有与跨越大于 14 个碱基长度片段互补序列的分子一般是优选的。当需要时，人们将一般喜欢设计具有 14 到 20 个核苷酸基因互补片段的核酸分子或更长的分子。例如，通过化学方法直接合成该片段，通过用核酸复制技术，如美国专利 4,683,195 和 4,683,202 的 PCR™
20 技术（在此特异性掺入每个仅供参考），或通过从含有适当插入子和适当限制位点的重组质粒中切下选定的 DNA 片段可容易地制备这些片段。

5.5 表达载体

25 本发明涉及包括本发明多核苷酸的表达载体。因此，在一个实施方案中，表达载体为包括可操作地连接到编码本发明多肽编码区上的启动子的分离纯化的 DNA 分子，其编码区可操作地连接到转录终止区域，继而该启动子驱动此编码区域的转录。

30 如这里所使用的，术语“可操作地连接”意为以这样的方式将启动子连接到编码区上从而该编码区的转录受该启动子的控制和调节。用于可操作地连接启动子到编码区上的方法在本领域中是众所周知的。

在细菌中起作用的启动子在本领域中是众所周知的。用于芽孢杆菌晶体蛋白的优选启动子实例包括 sig A、sig E 和 sig K 基因启动子。作为选择，也可使用编码天然、诱变或重组晶体蛋白基因的启动子本身。

5 当本发明的表达载体要用于转化植物时，选择具有在植物中驱动表达能力的启动子。在植物中起作用的启动子在本领域中是众所周知的。在植物中表达该多肽中有用的启动子有可诱导的、病毒的、合成的、组成性的启动子，如已描述过的那样（Poszkowski 等，1989；Odell 等，1985）和时间调控、空间调控和时-空调控的启动子（Chau 等，1989）。

10 也选择这样的启动子，其具有介导转化植物细胞或转基因植物编码区转录活性的能力。结构基因可由植物组织中多种启动子所驱动。启动子可以是近-组成性的，如影响双子叶或单子叶的 CaMV 35 S 启动子、或组织特异性或发育特异性启动子。

15 当该启动子为近-组成性启动子如 CaMV - 35 S 时，发现在多种转化的植物组织中有多肽表达的增加（如，愈伤组织、叶片、种子和根部）。作为选择，通过用含有组织特异性启动子的植物整合载体可将转化的效应归结于特定的植物组织。

20 组织特异性启动子的实例为对种子组织特异性的凝集素启动子。大豆中的该凝集素蛋白由单基因（Lel）所编码，其仅在种子成熟过程中表达并且占种子总 mRNA 的约 2% 到 5%。此凝集素基因和种子-特异性启动子已进行过彻底的特征分析并用于在转基因烟草植物中指导种子特异性的表达（Vodkin 等，1983；Lindstrom 等，1990）。

25 将含有编码目的多肽编码区的表达载体进行加工从而使其在凝集素启动子的控制下并且用，例如，原生质体转化方法（Dhir 等，1991）将该载体导入植物中。该多肽的表达具体与该转基因植物的种子有关。

从以组织特异性启动子转化的植物细胞制备的本发明转基因植物可与从以不同组织特异性启动子转化的植物细胞形成的第二种转基因植物杂交从而制备显示出不止一种特异性组织中转化效应的杂交转基因植物。

30 组织特异性启动子的实例有玉米蔗糖合成酶 1（Yang 等，1990）、玉米乙醇脱氢酶 1（Vogel 等，1989）、玉米集光复合体（Simpson，

1986)、玉米热休克蛋白(Odell等, 1985)、豌豆小亚单位RuBP羧化酶(Poulsen等, 1986; Cashmore等, 1983)、Ti质粒甘露氨酸合酶(Langridge等, 1989)、Ti质粒胭脂氨酸合酶(Langridge等, 1989)、矮牵牛查耳酮异构酶(Van Tunen等, 1988)、蚕豆富含甘氨酸蛋白1(Keller等, 1989)、CaMV 35 s转录本(Odell等, 1985)和土豆patatin(Wenzler等, 1989)的启动子。优选的启动子为花椰菜花叶病毒(CaMV 35 S)启动子和S-E9小亚单位RuBP羧化酶启动子。

表达载体和最终多肽编码区与哪一个启动子可操作地连接的选择直接依赖于所需的功能特性, 如蛋白表达的位置和时间选择, 和待转化的宿主细胞。这些是在构建重组DNA分子领域中众所周知的固有限制。然而, 用于完成本发明的载体能够指导与载体可操作地连接的多肽编码区域的表达。

用于高等植物中基因表达的典型载体在本领域中是众所周知的并且包括来源于已述根癌农杆菌诱导-肿瘤(Ti)质粒的载体(Rogers等, 1987)。然而, 已知数种其它植物整合载体系统在植物中起作用, 包括已述的pCaMVCN转移控制载体(Fromm等, 1985)。pCaMVCN(可从Pharmacia, Piscataway, NJ获得)包括花椰菜花叶病毒CaMV 35 S启动子。

在优选的实施方案中, 用于表达该多肽的载体包括在植物细胞中有效的选择标记, 优选为药物抗性选择标记。一种优选的药物抗性标记为其表达导致卡那霉素抗性的基因, 即已述的含有胭脂氨酸合酶启动子、Tn 5新霉素磷酸转移酶II(npt II)和胭脂氨酸合酶3'非翻译区的嵌合基因(Rogers等, 1988)。

RNA聚合酶经过聚腺苷酸化出现的位点转录编码DNA序列。一般地, 位于聚腺苷酸化下游数百个碱基对处的DNA序列用于终止转录。这些DNA序列在此称为转录终止区。这些区域对于转录信使RNA(mRNA)的有效聚腺苷酸化是必需的。

用于制备表达载体的方法在本领域中是众所周知的。用于转化植物(转化载体)的表达和制备这些载体的方法描述于美国专利4,971,908、4,940,835、4,769,091和4,757,011(在此特异性地掺入每例供参考)

中。可修饰这些载体以包括根据本发明的编码序列。

已开发出多种通过互补粘性末端或钝端可操作地将 DNA 连接到载体上的方法。例如，将互补的同聚物特征添加到待插入的 DNA 片段上和载体 DNA 上。然后通过互补同聚物尾部之间的氢键连接载体与 DNA 片段以形成重组 DNA 分子。

编码具有赋予细胞杀虫活性多肽的编码区优选为嵌合的编码苏云金芽孢杆菌晶体蛋白的基因。在优选的实施方案中，该多肽具有 SEQ ID NO: 10、SEQ ID NO: 12、SEQ ID NO: 14、SEQ ID NO: 26、SEQ ID NO: 28、SEQ ID NO: 30 或 SEQ ID NO: 34 的氨基酸残基序列；或这些序列中一个或更多个功能等价物的序列。根据此实施方案，包括 SEQ ID NO: 9、SEQ ID NO: 11、SEQ ID NO: 13、SEQ ID NO: 25、SEQ ID NO: 27、SEQ ID NO: 29 或 SEQ ID NO: 33 DNA 序列的编码区域也是优选的。

5.6 转化或转基因植物细胞

以本发明表达载体转化的细菌、酵母细胞、或植物细胞或植物也被涉及到。来自该转化或转基因细胞的转基因细菌、酵母细胞、植物细胞或植物也被涉及到。用于转化细菌和酵母细胞的方法在本领域中是众所周知的。一般地，转化的方法与众所周知的用于转化其它细菌或酵母如大肠杆菌或啤酒糖酵母的方法类似。

用于植物细胞 DNA 转化的方法包括农杆菌 - 介导的植物转化、原生质体转化、基因转移入花粉、注射至生殖器官中、注射至未成熟胚胎和粒子轰击。这些方法中的每一个具有独特的优点和缺点。因此，一种将基因导入特定植物品系中的特定方法可能不一定对于另一种植物品系是最为有效的，但是哪一种方法对于特定植物品系是有用的是众所周知的。

有许多种将转化 DNA 片段导入细胞中的方法，但并不是所有方法均适用于将 DNA 运载至植物细胞中。认为适宜的方法包括实际上可将 DNA 导入细胞中的任何方法，如根癌农杆菌及有关农杆菌的感染、DNA 的直接转运，如，通过 PEG - 介导的原生质体转化 (Omirulleh 等, 1993)、干燥/抑制 - 介导的 DNA 摄入、电穿孔、碳化硅纤维的搅拌、DNA 包被

粒子的加速等。在有些实施方案中，加速方法是优选的并且包括，例如，微粒轰击等。

用于导入 DNA 至细胞中的技术对于本领域中的技术人员是众所周知的。已描述过四种将基因导入细胞中的常用方法：（1）化学方法（Graham 和 van der Eb，1973）；（2）物理方法如显微注射（Capecchi，1980）、电穿孔（Wong 和 Neumann，1982；Fromm 等，1985）和基因枪（Johnston 和 Tang，1994；Fynan 等，1993）；（3）病毒载体（Clapp，1993；Lu 等，1993；Eglitis 和 Anderson，1988 a；1988 b）；和（4）受体-介导的机制（Curiel 等，1991；1992；Wagner 等，1992）。

5.6.1 电穿孔

应用简单、高压电脉冲至多种动植物细胞导致在质膜上形成纳米大小的孔。然后或者经过这些孔或者由于伴随孔的闭合而膜组分的重组而直接将 DNA 摄入细胞质中。电穿孔可能是极其有效的并且可同时用于克隆基因的短暂表达和带有目的基因整合拷贝的细胞系的建立。电穿孔，与磷酸钙-介导的转染和原生质体融合相比，频繁得到带有一个，或至多数个外源基因整合拷贝的细胞系。

通过电穿孔方法导入 DNA 对于本领域的技术人员是众所周知的。在该方法中，利用一些降解细胞壁的酶，如降解果胶的酶使靶受体细胞较未处理的细胞对于电穿孔转化更为敏感。作为选择，通过机械研磨使受体细胞对转化更为敏感。为了完成电穿孔转化，人们可以利用易碎的组织如细胞悬液，或胚胎发育的愈伤组织，或者作为选择，人们可以直接转化未成熟胚胎或其它结构的组织。人们将通过将其暴露于降解果胶的酶（溶果胶酶）或以控制的方式机械研磨而部分地降解选定细胞的细胞壁。然后这些细胞将对于电穿孔的 DNA 转移是感受态的（recipient），可在此时完成该过程，且然后根据新掺入 DNA 的性质通过合适的筛选或选择步骤鉴定转化的细胞。

30 5.6.2 微粒轰击

用于转运转化的 DNA 片段至植物细胞的进一步优越的方法是粒子

轰击。在该方法中，粒子可以核酸包被并且通过推进力转入细胞中。粒子的实例包括钨、金、铂等。

除了其是可重复而稳定转化单子叶植物的有效方法外，粒子轰击的优势在于其既不需要分离原生质体（Cristou 等，1988）又不需对农杆菌感染的易感性。用于通过加速将 DNA 运载至玉米细胞中的方法的说明性实施方案为 **Biolistics Particle Delivery System**，其可于经筛子、如不锈钢或 Nytex 筛子将包被 DNA 或细胞的粒子推进至覆盖有培养悬液中玉蜀黍（corn）细胞的滤膜表面。此筛子分散这些颗粒从而它们不以大团块的形式运载至受体细胞中。据认为在发射器与待轰击细胞之间的筛子减少投射物的大小并且可通过减少太大的投射物对受体细胞造成的损害而促进更高频率的转化。

对于轰击，悬液中的细胞优选于滤膜上或固体培养基上进行浓缩。作为选择，未成熟胚胎或其它靶细胞可安排于固体培养基上。将待轰击的细胞放置于大粒子（macroprojectile）终止板下适当的距离处。如果必要，在加速器和待轰击细胞间也可放置一个或更多个筛子。通过使用前述的技术，人们可以获得达 1000 或更多灶（foci）短暂表达标记基因的细胞。细胞灶中轰击 48 小时后表达外源基因产物的细胞数目常为 1 到 10 并平均 1 到 3 个。

在轰击转化中，人们可优化轰击前的培养条件和轰击参数从而产生最大数目的稳定转化子。轰击的物理和生物学参数在该技术中是重要的。物理因素为包括操作 DNA/微粒沉淀物的那些因素或影响大或小微粒飞行和速度的那些因素。生物学因素包括涉及轰击前和轰击后即刻操作细胞的所有步骤、为了帮助减轻与轰击有关的损伤而对靶细胞的渗透调节和转化 DNA 的性质，如线性 DNA 或完整超螺旋质粒。据认为轰击前的操作对于成功转化未成熟胚胎是特别重要的。

因此，应该考虑到人们在小规模研究中可能希望调整各种轰击参数以彻底地优化条件。人们可能特别希望调整物理参数如间隙距离（gap distance）、飞行距离、组织间距、和氨压。通过改变影响受体细胞生理状态的条件并且因此可影响转化和整合效率，人们也可最大限度地减少损伤减少因素（TRFs）。例如，可从调整渗透状态、组织水化和传代培养时期或受体细胞的细胞周期以用于最佳转化。根据本发明内容指导

下，进行其它常规的调整对于本领域中的技术人员将是众所周知的。

粒子-介导的转化方法对于本领域中的技术人员是众所周知的。美国专利 5, 015, 580 (特异性在此掺入供参考) 描述了用该技术对大豆的转化。

5

5.6.3 农杆菌-介导的转移

农杆菌-介导的转移是可广泛用于将基因导入植物细胞中的系统，因为该 DNA 可导入完整的植物组织中，继而绕过了从原生质体再生完整植株的必要。用农杆菌介导的植物整合载体将 DNA 导入植物细胞中在本领域 10 中是众所周知的。参见，例如，已述的方法 (Fraley 等, 1985 ; Rogers 等, 1987)。用农杆菌-介导的转移对棉花植物的遗传工程描述于美国专利 5, 004, 863 (在此特异性地掺入供参考) 中，而莴苣植物的转化描述于美国专利 5, 349, 124 (在此特异性地掺入供参考) 中。此外，Ti - DNA 的整合是很少导致重排的相对精确的方法。待转移的 DNA 区域 15 由边界序列所限定，并且通常按已述将间插 DNA 插入到植物基因组中 (Spielmann 等, 1986 ; Jorgensen 等, 1987)。

当今的农杆菌转化载体能够在大肠杆菌以及农杆菌中复制，能够按所述进行方便的操作 (Klee 等, 1985)。此外，最近在农杆菌-介导基因转移载体中的技术进展改进了载体中基因和限制位点组织从而利于 20 构建能够表达各种多肽编码基因的载体。所述的载体 (Roger 等, 1987) 具有方便的多接头区域并且适用于本发明的目的，其中多接头区域的二侧为用于插入多肽编码基因直接表达的启动子和聚腺苷酸化位点。此外，同时含有有臂 (armed) 和无臂 (disarmed) Ti 基因的农杆菌可用于转化。在农杆菌-介导的转化有效的那些植物品系中，由于 25 基因转移简单易行且特定的性质而成为最佳的方法。

农杆菌介导的对叶片和其它组织如子叶和下胚轴的转化似乎限于农杆菌自然感染的植物。农杆菌介导的转化在双子叶植物中最为有效。尽管已按所述用农杆菌在芦笋中制备了转基因植物 (Bytebier 等, 1987)，但很少单子叶植物似乎是农杆菌的天然宿主。因此，商业重要的禾谷如水稻、玉蜀黍和小麦通常必须用其它方法加以转化。然而，如 30 上述，也可用农杆菌完成对芦笋的转化 (参见，如 Bytebier 等, 1987)。

用农杆菌转化方法形成的转基因植物一般在染色体上含有一个基因。由于该添加的基因，此转基因植物可称为杂合的。然而，由于单词“杂合”的使用通常指在一对染色体的第二条染色体相同位点存在互补基因，并且在此处含有一个添加基因的植物中没有这样的基因，认为这样植物更为准确的名字为独立的分离子，因为添加的外源基因在有丝分裂和减数分裂期间独立地分离。

更为优选的是转基因植物对于添加的结构基因是纯合的；即含有 2 个添加基因的转基因植物，即在一对染色体的每条染色体相同位点各具一个基因。纯合的转基因植物可通过以下方法而获得，包括有性交配（自交）含有一个添加基因的独立分离的转基因植物、萌发一些产生的种子并且分析得到植物较对照（天然、非转基因）或独立分离转基因植物增强的羧化酶活性。

应当明白也可交配两种不同的转基因植物以制备含有二种独立分离的添加外源基因的后代。适当后代的自交可以产生对编码目的多肽的 2 个添加外源基因均为纯合的植物。也可考虑与母本植物的回交和与非转基因植物的异型杂交。

植物原生质体的转化可根据下面的方法完成，包括磷酸钙沉淀、聚乙二醇沉淀、电穿孔、和这些处理的组合（参见，如，Potrykus 等，1985；Lorz 等，1985；Fromm 等，1985；Uchimiya 等，1986；Callis 等，1987；Macrott 等，1988）。

不同植物品系中这些系统的施用依赖于从原生质体中再生此特定植物的能力。已描述过从原生质体中再生禾谷类的说明性方法（参见，如 Fujimura 等，1985；Toriyama 等，1986；Yamada 等，1986；Abdullah 等，1986）。

为了转化不能从原生质体中成功再生的植物品系，可以利用其它将 DNA 导入完整细胞或组织中的方法。例如，按所述可从未成熟胚胎或外植体中再生出禾谷类（Vasil，1988）。此外，可以利用“粒子枪”或高速微粒技术（Vasil，1992）。

用此后者技术，按所述将 DNA 经细胞壁带入小金属颗粒上的原生质体中（Klein 等，1987；Klein 等，1988；McCabe 等，1988）。该金属粒子穿透数层细胞且因而允许对组织外植体内部细胞的转化。

5.6.4 植物中基因的表达

与单细胞生物相比，植物编码子的使用更与人类和其它高等生物类似的事实，未修饰的细菌基因常在转基因植物中表达较差。在密码子第三位对 GC 含量的明显的总的偏向性已由 Murray 等（1990）进行过详述。该书中描述的 207 种植物基因允许对植物中氨基酸密码子的倾向性进行编辑（compilation）。这些作者描述了在单子叶和双子叶植物中密码子使用的差异以及叶绿体编码基因与核编码基因之间密码子使用的差异。利用供给的密码子频率表，本领域中的技术人员通过修饰该 DNA 序列可加工此细菌序列用于在植物中表达从而提供在第三位置密码子对 G 或 C 的偏向性。

Diehn 等（1996）的工作详述了修饰原核生物来源的基因序列以让其在植物中表达。Iannacone 等（1997）描述了以编码 cry 3 类内毒素的遗传工程苏云金芽孢杆菌基因对番茄（egg plant）的转化。利用避免聚腺苷酸化序列，ATTA 序列和剪切位点，构建使编码的毒素在植物体内表达的合成基因。Rouwental 等（1997）已描述水母绿色荧光蛋白在转基因烟草中的表达。用合成的基因，建立第三位置的密码子对 C + G 的偏向性从而使外源基因在植物中表达。

Fütterer 和 Hohn（1996）描述了 mRNA 序列、前导序列、多顺反子信息和内部核糖体结合位点 motifs 对植物中表达的影响。通过构建合成基因对这些序列的修饰使病毒 mRNA 能够在转基因植物细胞中表达。

尽管近年在制备表达细菌蛋白如苏云金芽孢杆菌晶体蛋白的植物中取得巨大的进展，但在植物中表达天然细菌基因的结果又常是令人失望的。与微生物遗传学不同，有关影响外源基因在植物中异源表达因素的早期植物遗传学知之甚少。然而，近年来，数种重要的因素作为主导者参与改变特定编码序列的蛋白表达水平的程度。例如，科学家现已知道在细胞中维持显著水平的特定 mRNA 的确是关键因素。不幸的是，编码外源蛋白 mRNA 的低稳定状态水平原因有很多。首先，全长 RNA 合成可能不以高频率发现。这可能，例如，由于转录期间 RNA 的提早终止或者由于转录期间没有预期到的 mRNA 加工所造成的。其次，植物细胞中可能产生全长 RNA，但然后在核中以产生无功能 mRNA 的方式进行加工（剪切、polyA 添加）。如果该 RNA 没有被适当合成、终止和聚腺苷

酸化，其不能够移动到细胞质中用于翻译。类似地，在细胞质中，如果 mRNA 具有降低的半寿期（由其一级或二级结构序列所测定），那么将产生不充足的蛋白产物。此外，还不能确定翻译效率对 mRNA 半寿期影响的幅度。此外，每种 RNA 分子折叠成由其序列所决定的特定的结构或结构家族。任何 RNA 的特定结构可能导致细胞质中更大或更少的稳定性。结构本身也可能是核内 mRNA 加工的决定因素。不幸的是，不可能预言并且几乎不可能测定体外或体内任何 RNA（除 tRNA）的结构。然而，很可能剧烈地改变 RNA 的序列将对其折叠的结构具有较大的影响。很可能结构本身或特定的结构特征在决定 RNA 稳定性中也具有作用。

10 为了克服在外源基因表达中的这些局限，研究人员已鉴定出 RNA 中对其稳定性具有特定影响的特定序列和信号。在本发明的一些实施方案中，因此，需要优化公开的核酸序列片段在植物体内的表达。这样做的一个特定方法是通过改变细菌基因以去除降低在转化的植物细胞中表达的序列或基序。加工编码序列从而在植物体中最佳表达的过程经常称为
15 “植物化” DNA 序列。

特别麻烦的序列是富含 A + T 的那些序列。不幸的是，由于苏云金芽孢杆菌具有富含 A + T 的基因组，故必须经常修饰天然晶体蛋白基因序列从而在植物中最佳表达。基序 ATTTA（或当其出现于 RNA 中的 AUUUA）已用作对哺乳动物细胞 mRNA 进行去稳定化的序列（Shaw
20 和 Kamen，1986）。许多短寿命的 mRNA 具有富含 A + T 的 3' 非翻译区，并且这些区域经常具有 ATTTA 序列，有时以多拷贝或多聚体（如，ATTTATTTA ...）存在。Shaw 和 Kamen 显示将不稳定 mRNA 的 3' 末端转移至稳定的 RNA（珠蛋白或 VA 1）大大降低了稳定 RNA 的半寿期。他们进一步显示五聚体 ATTTA 对稳定的信使具有复杂的脱稳定效应，并且此信号无论位于 3' 末端或编码序列内部均可发挥其效应。然而，ATTTA 序列的数目和/或其所处位置左右的序列对于决定是否它们作为去稳定序列而起作用中也是重要的。Shaw 和 Kamen 显示 ATTTA
25 三聚体较五聚体对 mRNA 稳定性的影响要少得多并且二聚体或单聚体对其稳定性没有影响（Shaw 和 Kamen，1987）。注意到 ATTTA 多聚体如五聚体自动产生富含 A + T 的区域。这被显示为细胞质效应，而非核
30 效应。在其它不稳定的 mRNA 中，此 ATTTA 序列可能仅以单拷贝存在，

但其经常含于富含 A + T 的区域中。根据目前收集到的动物细胞数据，似乎至少在一些序列中的 ATTTA 在稳定性中是重要的，但仍不可能预言 ATTTA 的哪一种出现几率为去稳定因素或者是否这些效应中的任何一个效应有可能在植物中见到。

5 对动物细胞中 mRNA 降解的一些研究表明 RNA 降解有时可能在富含 A + T 区域中开始受到核苷酸降解攻击。还不清楚是否这些降解在 ATTTA 序列处出现。也有这样的 mRNA 实例，即依据其表达的细胞类型或其表达所处的细胞周期时期而具有不同的稳定性。例如，组蛋白 mRNA 在 DNA 合成过程中是稳定的但如果 DNA 合成被打断则不稳定。
10 有些组蛋白 mRNA 的 3' 末端似乎负责此效应 (Pandey 和 Merzluff, 1987)。其似乎不是由 ATTTA 所介导，也不清楚是什么控制了该 mRNA 的不同稳定性。另一个实例是在 B 细胞成熟过程中 B 淋巴细胞中 IgG mRNA 的不同稳定性 (Genovese 和 Milcarek, 1988)。最后一个实例是突变的 β - thallemic 珠蛋白 mRNA 的不稳定性。在其中的该基因
15 正常表达的骨髓细胞中，突变的 mRNA 是不稳定的，而野生型 mRNA 是稳定的。当突变的基因在体外的 Hela 或 L 细胞中表达时，此突变的 mRNA 没有显示出稳定性 (Lim 等, 1988)。这些实例均提供了这样的证据，即 mRNA 稳定性可由细胞类型或细胞特异性因素所介导。此外，此种不稳定性还与特异的序列无关。有了这些不确定性，预言哪一
20 种 RNA 可能在给定的细胞中是不稳定的是不可能的。此外，根据其中存在 RNA 的细胞的性质，甚至 ATTTA 基序可起不同作用。Shaw 和 Kamen (1987) 报导了蛋白激酶 C 的活化可阻止 ATTTA 介导的降解。

添加聚腺苷酸链 3' 末端对于大多数真核 mRNA，包括植物和动物 mRNA 是常见的事。当今已接受的 polyA 添加的观点是新生的转录本延
25 伸越过成熟的 3' 末端。此转录本中含有的聚腺苷酸化和适当的 3' 末端形成的信号。此 3' 末端处的加工包括 mRNA 的切割和添加 polyA 至成熟的 3' 末端。通过寻找动物和植物 mRNA 中靠近 polyA 的序列 (tract) 处的共有序列，鉴定明显参与 polyA 添加和 3' 末端切割的共有序列已是可能的。相同的共有序列似乎对这两种过程均是重要的。这
30 些信号通常是在序列 AATATT 上的变异。在动物细胞中，已鉴定出此序

列有功能的一些变异体；在植物细胞中，似乎存存在广泛范围的功能序列（Wickens 和 Stephenson，1984；Dean 等，1986）。因为这些共有序列在 AATAAA 上具有变异，故它们是富含 A + T 的序列。发此序列一般为成熟 mRNA 中 polyA 序列前的 15 到 20 bp。动物细胞中的研究
 5 表明此序列同时参与 polyA 添加和 3' 成熟。此序列中的定点突变可破坏这些功能（Conway 和 Wickens，1988；Wickens 等，1987）。然而，已观察到推断的 polyA 信号 3' 达 50 到 100 bp 的序列也是必需的，即，具有正常 AATAAA 但在下游已被代替或破坏的基因没有得到适当的聚腺苷酸化（Gil 和 Proudfoot，1984；Sadofsky 和 Alwine，1984；
 10 McDevitt 等，1984）。即，polyA 信号本身对于完全而适当的加工是不够的。仍不知道除了 polyA 信号外，什么特定的下游序列是必需的，或者是否有具有该功能的特异性序列。因此，序列分析仅可鉴定潜在的 polyA 信号。

在正常聚腺苷酸化的天然 mRNA 中，已观察到通过改变该 mRNA
 15 中的 polyA 信号或其它序列，可获得功能 mRNA 水平的复杂效应。这已在数种天然 mRNA 中观察到，目前已有特异基因的结果。

已显示在天然 mRNA 中，适当的聚腺苷酸化在 mRNA 积累中是重要的，并且破坏该过程可显著影响 mRNA 水平。然而，预言在正常基因中变化的效应方面的知识存在还不充足。在杂合基因中，预言此后果甚至更为困难。然而，鉴定出的推断位点功能失常是可能的。即，这些位点
 20 可能不是作为适当的 polyA 位点而起作用，相反，是作为异常位点而起作用从而得到不稳定的 mRNA。

在动物细胞系统中，AATAAA 是在 polyA 上游 mRNA 中鉴定出最为常见的信号，但至少已发现 4 种变异体（Wickens 和 Stephenson，
 25 1984）。在植物中，还没有做出这么多的分析，但很明显可以使用类似于 A..TAAA 的多个序列。表 4 中称为主要或次要植物位点仅仅指 Dean 等（1986）的研究，其仅仅分析了三种植物基因。聚腺苷酸化位点命名为主要或次要仅仅指其在已分析过的天然基因中作为有功能位点而出现的频率。在植物中这方面数据很有限。以任何确定性预言存在于杂合基
 30 因如编码本发明晶体蛋白那些基因中命名为主要或次要位点更有可能或更少有可能部分或完全起作用困难的。

表 4
植物基因中聚腺苷酸化位点

| | | | |
|----|------|--------|--------|
| | PA | AATAAA | 主要共有位点 |
| 5 | P1A | AATAAT | 主要植物位点 |
| | P2A | AACCAA | 次要植物位点 |
| | P3A | ATATAA | " |
| | P4A | AATCAA | " |
| | P5A | ATACTA | " |
| | P6A | ATAAAA | " |
| 10 | P7A | ATGAAA | " |
| | P8A | AAGCAT | " |
| | P9A | ATTAAT | " |
| | P10A | ATACAT | " |
| | P11A | AAAATA | " |
| | P12A | ATTAAT | 次要动物位点 |
| 15 | P13A | AATTAA | " |
| | P14A | AATACA | " |
| | P15A | CATAAA | " |

20 本发明提供用于制备合成植物基因的方法，该基因较以前通常用于植物转化的野生型基因以显著更高的水平表达其蛋白产物。在另一方面，本发明也提供了编码非-植物蛋白的新的合成植物基因。

如上述，天然苏云金芽孢杆菌基因在植物中的表达经常是困难的。苏云金芽孢杆菌基因编码序列的性质使其与植物中的基因以及许多其它在植物中表达的杂合基因有所不同。特别地，苏云金芽孢杆菌基因富含
25 (~ 62 %) 腺嘌呤 (A) 和胸腺嘧啶 (T) 而植物基因和大多数其它在植物中表达的细菌基因在 45 ~ 55 % A + T 的数量级。

由于遗传密码子的简并性和任何氨基酸密码子选择数目的限制，有些芽孢杆菌的结构编码序列的大多数“过多” A + T 位于密码子的第三位。即，一些芽孢杆菌基因在许多密码子中具有 A 或 T 作为第三个核苷酸。因此 A + T 含量可部分决定密码子使用偏向性。此外，很明显在其
30 进化的生物中，基因向最大的功能进化。这意味着位于一种生物基因中

的特定核苷酸序列（在该生物中此基因除了编码特定片段的氨基酸外可能没有作用）有可能在另一种生物中被识别为基因调控元件（如转录启动子或终止子、polyA 添加位点、内含子剪切位点、或特异性 mRNA 降解信号）。此错读的信号不是异源基因表达的更为常见的特征，但这可
5 部分地由许多生物相对均一 A + T 含量（~ 50 %）加以解释。此 A + T 含量加上遗传密码的性质对任何特定寡核苷酸序列出现的可能性赋予清晰的限制。因此，与苏云金芽孢杆菌基因相比，具有 50 % A + T 含量的大肠杆菌基因很少有可能含有任何特定的富含 A + T 的片段。

通常，为了获得 δ -内毒素基因在植物中高水平的表达，通过对包
10 括结构基因 DNA 的定点诱变而去除 ATTTA 序列和推断的聚腺苷酸化信号而修饰编码该 δ -内毒素的已存在的结构编码序列（“结构基因”）。虽然已观察到仅部分去除上面鉴定出的任一序列时有增加的表达水平，但是基本上去除所有聚腺苷酸化信号和 ATTTA 序列是最为优选的。作为选择，如果制备编码用于目的蛋白表达的合成基因，选择的密码子应
15 避开 ATTTA 序列和推断的聚腺苷酸化信号。对于本发明的目的，推断的聚腺苷酸化信号包括但不一定限于 AATAAA、AATAAT、AACCAA、ATATAA、AATCAA、ATACTA、ATAAAA、ATGAAA、AAGCAT、ATTAAT、ATACAT、AAAATA、ATTTAA、AATTAA、AATACA 和 CATAAA。在替代 ATTTA 序列
20 和聚腺苷酸化信号中，优选利用避开很少见于植物基因组中密码子的密码子。

扫描选定的 DNA 序列以鉴定出具有大于 4 个连续腺嘌呤（A）或胸腺嘧啶（T）核苷酸的区域。扫描出 A + T 区域潜在的植物聚腺苷酸化信号。虽然缺乏 5 个或更多个连续 A 或 T 核苷酸排除了大多数植物的
25 聚腺苷酸化信号，但如果在相互间的 10 个核苷酸内鉴定出多于一个次要聚腺苷酸化信号时，那么优选改变此区域的核苷酸序列以去除这些信号而同时维持原先编码的氨基酸序列。

第二步是考虑在步骤一中鉴定出的富含 A + T 区域周围的大约 15
30 到 30 个左右的核苷酸残基。如果周围区域的 A + T 含量不到 80 %，那么应该检查该区域的聚腺苷酸化信号。依据聚腺苷酸化信号对该区域的改变依赖于（1）存在的聚腺苷酸化信号的数目和（2）主要植物聚腺

苷酸化信号的存在。

检查延伸区域植物聚腺苷酸化信号的存在。通过对该 DNA 序列的定点诱变而去除聚腺苷酸化信号。也检查延伸区域中的多拷贝 ATTTA 序列，其也通过诱变去除。

- 5 破坏包括许多连续 A + T 碱基或 G + C 碱基的区域也是优选的因为预期这些区域由于自身互补性而具有形成发夹结构更高的可能性。因此，插入异源碱基对将减小自身 - 互补二级结构形成的可能性，已知该二级结构在有些生物中抑制转录和/或翻译。在大多数情况，通过使用不含有多于 5 个连续 A + T 或 G + C 序列可最大限度地减少副作用。

10

5.7 昆虫抗性转基因植物的制备

因此，通过用粒子轰击方法转化这些植物可增该植物如玉蜀黍中编码目的多肽基因的量（即，具有抗一种或更多种昆虫杀虫活性的细菌晶体蛋白或多肽）。通过实例的方式，用粒子枪运载包被于微粒上的 DNA 将含有苏云金芽孢杆菌晶体蛋白编码区和适当可选择标记的表达载体转化入玉米（玉蜀黍）胚胎细胞悬液中。从表达该公开杀虫剂晶体蛋白的转化胚胎愈伤组织中再生转基因植物。粒子轰击已用于成功转化小麦（Vasil 等，1992）。

也可按所述通过直接将 DNA 转入花粉中而将 DNA 导入植物中（Zhou 等，1983；Hess，1987；Luo 等，1988）。也可按所述将 DNA 注射入植物的生殖器官而获得编码多肽基因的表达（Pena 等，1987）。也可按所述将 DNA 直接注射入未成熟胚胎的细胞中并且对于干燥的胚胎进行重新水合（Neuhaus 等，1987；Benbrook 等，1986）。

从单个植物原生质体或各种外植体中或再生植株在本领域中是众所周知的（Weissbach 和 Weissbach，1988）。此再生和生长方法一般包括筛选转化的细胞、培养这些单个化的细胞经过胚胎发育的正常期经过生根的小植株时期。类似地再生转基因胚胎和种子。之后将得到的转基因生根的幼枝种植于适当的植物培养基中如土壤中。

可按所述的本领域中众所周知的方法从叶片外植体中发育或再生含有通过农杆菌导入的编码目的多肽外源基因的植株（Horsch 等，1985）。在该方法中，按所述在选择试剂存在下和诱导被转化植物品系

中幼枝再生的培养基中培养转化子 (Fraley 等, 1983)。特别地, 美国专利 5, 349, 124 (在此掺入说明书供参考) 详述了遗传转化莴苣细胞的产生和从其中得到表达赋予该植物对鳞翅目幼虫杀虫活性杂合晶体蛋白的植物。

5 此方法一般在 2 到 4 个月内产生幼枝并且然后将这些幼枝转移至适当的含有筛选剂和抗生素的诱导根的培养基中从而防止细菌的生长。然后将选择剂存在下生根以形成小植株的幼苗移植至土壤或其它培养基中以便能够生根。这些方法随着所用的特定植物品系而变化, 此变化在本领域中是众所周知的。

10 优选地, 按前述将再生的植物自花传粉以提供纯合的转基因植物。或者, 将获自此再生植物的花粉与农业上重要的种子培育出的植物、优选为近交系进行杂交。反过来, 用这些重要品系植物的花粉给再生植物授粉。用对本领域中技术人员众所周知的方法培养含有所需多肽的本发明转基因植物。

15 本发明的转基因植物因此具有编码一种或更多种这里公开的嵌合 Cry 多肽的增加量的编码区 (如, cry 基因)。优选的转基因植物为独立的分离分子并且可将该基因和其活性传递至其子代中。更为优选的转基因植物为该基因的纯合体并且将该基因在有性交配中传递至其所有子代中。转基因植物的种子可种植于田野或温室中, 并将得到的有性成熟转基因植物自花传粉以产生真正的生育植物。通过实例的方式, 这些植物的子代变成真正的生育系, 评估其在一定的环境条件下, 优选在田间对鞘翅目昆虫具有增加的杀虫能力。本发明者考虑到了本发明在产生下列转基因植物中将会有特别的应用, 包括玉蜀黍 (corn)、大豆、棉花、烟草、番茄、土豆、亚麻、黑麦、大麦、canola、小麦、燕麦、水稻、
20 其它谷物、蔬菜、水果、果树、浆果、草坪草、观赏植物、藻木和乔木。
25

5.8 核酶

核酶是切割特定 mRNA 种类的酶促 RNA 分子。在有些实施方案中, 本发明者考虑到选择和利用能够切割本发明 RNA 片段的核酶及其用于
30 减少靶 mRNA 在特定细胞类型或组织中的活性。

目前已知天然酶促 RNA 有 6 种基本的变异体。在生理条件下每个可

催化反式 RNA 磷酸二酯键的水解（并且因此可切割其它 RNA 分子）。一般地，酶促核酸通过首先结合到靶 RNA 上而起作用。此结合经过靶向结合酶促核酸部分而发生，该部分紧紧与起切割靶 RNA 作用分子的酶促部分相邻。因此，该酶促核酸首先经过互补碱基 - 配对识别且然后结合靶 RNA，且一旦结合到正确的位点，起酶促作用以切割靶 RNA。策略性切割此靶 RNA 将破坏其指导编码蛋白合成的能力。酶促核酸结合并切割其 RNA 靶子后，其从该 RNA 中释放出来以寻找另一个靶子并可反复结合和切割新的靶子。

核酶的酶促性质优越于许多技术，如反义技术（其中核酸分子简单地结合到核酸靶子上以阻止其翻译）因为完成治疗所必需的核酶浓度低于反义寡核苷酸的浓度。此优势反映了核酶酶促作用的能力。因此，单个核酶分子能够切割许多分子的靶 RNA。此外，核酶为高度特异性的抑制剂，其抑制特异性不仅依赖于结合到靶 RNA 上的碱基配对机制，而且依赖于靶 RNA 的切割机制。靠近切割位点的单个错配或碱基替代可完全消除核酶的催化活性。在反义分子中类似的错配不阻止其作用（Woolf 等，1992）。因此，核酶作用的特异性大于结合相同 RNA 位点的反义寡核苷酸的特异性。

酶促核酸分子可形成于锤头、发夹、肝炎 δ 病毒、组 I 内含子或 RNaseP RNA（与 RNA 前导序列有关）或 Neurospora VS RNA 基序中。锤头基序的实例由 Rossi 等（1992）描述过；发夹基序的实例由 Hampel 等描述过（欧洲专利，EP 0360257），Hampel 和 Tritz（1989），Hampel 等（1990）和 Cech 等（美国专利，5,631,359）；肝炎 δ 病毒基序的实例由 Perrotta 和 Been（1992）描述过；RNaseP 基序的实例由 Gruerrier - Takada 等（1983）描述过，Neurospora VS RNA 核酶基序由 Collins 描述过（Saville 和 Collins，1990；Saville 和 Collins，1991；Collins 和 Olive，1993）；并且 I 组内含子的实例由 Cech 等描述过（美国专利 4,987,071）。本发明酶促核酸分子中所有重要的因素是其具有与一种或更多种靶基因 RNA 区域互补的特异性底物结合位点并且其在底物结合位点内部或周围具有赋予该分子 RNA 切割活性的核苷酸序列。因此该核酶构建体不必要限制于这里所述的特定基序。

本发明提供了用于制备一类对所需靶 RNA 显示出高度特异性酶促

切割试剂的方法。该酶促核酸分子优选靶向于靶 RNA 高度保守的序列区域从而可提供一种或数种酶促核酸给特定疾病或病情的治疗中。此酶促核酸分子可外源性地运载至所需的特异性细胞。作为选择，此核酶可由运载至特定细胞的 DNA 或 RNA 载体所表达。

5 小的酶促核酸基序（如，锤头或发夹结构）可用于外源运输。这些分子的简单结构增加了此酶促核酸侵袭 mRNA 结构靶区域的能力。作为选择，催化 RNA 分子可在细胞中由真核生物启动子控制下表达（如，Scanlon 等，1991；Kashani - Sabet 等，1992；Dropulic 等，1992；Weerasinghe 等，1991；Ojwang 等，1992；Chen 等，1992；Sarver
10 等，1990）。本领域中的技术人员认识到任何核酶均可在真核细胞的适当 DNA 载体中表达。通过第二种核酶使其从初级转录本中释放可增加此核酶的活性（Draper 等，国际专利申请公开 No. WO 93/23569，和 Sullivan 等，国际专利申请公开 No. WO 94/02595，在此掺入此二者供参考其完整性；Ohkawa 等，1992；Taira 等，1991；Ventura 等，
15 1993）。

核酶可直接加入、或可与阳离子脂复合、脂类复合物、包装于脂质体中或其它方式运载至靶细胞中。此 RNA 或 RNA 复合物可经注射、烟雾剂吸入、灌注泵或 stent，掺入或不掺入生物聚合物而局部施用至体内外的有关组织。

20 按 Draper 等所述(国际专利申请公开 No. WO 93/23569)或 Sullivan 等所述（国际专利申请公开 No. WO 94/02595）设计核酶并合成从而按所述进行体内外测试。也可优化此核酶以用于运输。虽然提供了具体的实例，当必要时本领域中的人员将认识到可以利用其它种类中的相当的 RNA 靶子。

25 可通过计算机折叠（Jaeger 等，1989）单个地分析锤头或发夹核酶以评价是否此核酶序列折叠成适当的二级结构。对具有不利的分子内结合臂与催化核心之间相互作用的那些核酶不予考虑。可选择改变结合臂长度以优化活性。一般地，每条臂上的至少 5 个碱基能够与靶 RNA 结合或与其相互作用。

30 可设计锤头或发夹基序的核酶从而与 mRNA 信使中的各种位点退火，并可对其化学合成。所用的合成方法按照 Usman 等（1987）和

Scaringe 等 (1990) 所述的通常 RNA 合成方法并且使用常规的核酸保护和偶联基团, 如 5'-末端的二甲氧三苯甲基和 3'-末端的亚磷酰胺。平均每步的偶联产量一般 > 98 %。可以两部分合成发夹核酶并且退火以重建活性的核酶 (Chowrira 和 Burke, 1992)。通过修饰核酸酶抗性基因, 例如, 2'-氨基、2'-C-烯丙基、2'-氟基、2'-O-甲基、2'-H 可广泛地修饰核酶以增强稳定性 (综述参见 Usman 和 Cedergren, 1992)。可通过用一般方法的凝胶电泳或高压液相色谱来纯化核酶并重悬于水中。

核酶活性可通过以下方法加以优化, 包括改变核酶结合臂的长度或化学合成具有防止其被血清核糖核酸酶降解修饰的核酶 (参见, 如, 国际专利申请公开 No. WO 92/07065; Perrault 等, 1990; Pieken 等, 1991; Usman 和 Cedergren, 1992; 国际专利申请公开 No. WO 93/15187; 国际专利申请公开 No. WO 91/03162; 美国专利 5, 334, 711; 和国际专利申请公开 No. WO 94/13688, 其中描述了可制备对酶促 RNA 分子的糖部分进行的各种化学修饰), 该核酶具有增强其在细胞中效率的修饰并去除了茎段 II 碱基以缩短 RNA 合成时间和降低化学需求。

Sullivan 等 (国际专利申请公开 No. WO 94/02595) 描述了用于运载酶促 RNA 分子的一般方法。可通过各种已知的对本领域熟练人员熟悉的方法, 包括但不限于脂质体包被、离子电渗疗法或插入其它载体, 如水凝胶、环化糊精、生物可降的微胶囊和生物粘性微球体将核酶施用至细胞。对于有些情况, 可在体外直接将核酶运载至具有或不具有前述载体的细胞或组织。作为选择, 可通过直接吸入、直接注射或用导管、灌注泵或 stent 局部运载 RNA/载体组合。其它运载途径包括但不限于血管内、肌肉内、皮下或联合注射、烟雾剂吸入、口腔 (片剂或药物形式)、局部、全身、眼、腹膜内和/或子囊内 (intrathecal) 的运载。核酶运载和施用更为详细的描述提供于 Sullivan 等 (国际专利申请公开 No. WO 94/02595) 和 Draper 等 (国际专利申请公开 No. WO 93/23569) 文献中, 在此引用供参考。

另一种在细胞中积累高浓度核酶的方法是掺入编码核酶的序列到 DNA 表达载体中。核酶序列的转录由真核生物 RNA 聚合酶 I (pol I)、RNA 聚合酶 II (pol II) 或 RNA 聚合酶 III (pol III) 的启动子所驱

动。 pol II 或 pol III 启动子的转录本将以高水平在所有细胞中表达；在给定细胞类型中给定 pol II 启动子的水平将依赖于附近存在的基因调控序列（增强子、沉寂子等）的性质。如果原核生物 RNA 聚合酶在适当的细胞中表达，则也可使用原核生物 RNA 聚合酶启动子（Elroy - Stein 和 Moss, 1990； Gao 和 Huang, 1993； Lieber 等, 1993； Zhou 等, 1990）。由这些启动子表达的核酶可在哺乳动物细胞中起作用（如, Kashani - Saber 等, 1992； Ojwang 等, 1992； Chen 等, 1992； Yu 等, 1993； L'Huillier 等, 1992； Lisziewicz 等, 1993）。此转录单位可掺入多种用于导入哺乳动物中的载体，包括但不限于质粒 DNA 载体、病毒 DNA 载体（如腺病毒或腺相关载体）、或病毒 RNA 载体（如逆转录病毒、semliki 森林病毒、新培斯病毒载体）。

本发明的核酶可作诊断手段以检测细胞系或细胞类型中的遗传漂变和突变。它们也可用于分析靶 RNA 分子的水平。核酶活性与靶 RNA 结构间的密切关系使能够检测该分子中任何区域改变靶 RNA 碱基配对和三维结构的突变。通过用本发明所述的多种核酶，人们可在体外以及细胞和组织中找出对 RNA 结构重要的核苷酸改变。用核酶切割靶 RNA 可用于抑制特定细胞或细胞类型中的基因表达并且（基本上）限定特定基因产物的作用。

20 5.9 同源基因和基因片段的分离

根据本发明的基因和 δ - 内毒素不仅包括这里公开的全长序列，而且包括具有这里特异性例举序列特征性杀虫活性的这些序列的片段或融合蛋白。

可经过数种方法鉴定和获得杀虫剂 δ - 内毒素，这一点对于本领域的技术人员来说应是显而易见的。此特异性基因或其部分可获自培养库或通过例如，用基因仪合成构建。用制备点突变的标准技术可容易地构建这些基因的变异形式。同样，也可根据标准的方法用商业上可得到的外切核酸酶或内切核酸酶制备这些基因的片段。例如，可用酶如 Bal 31 或定点诱变从这些基因的末端系统地切下核苷酸。同样，用多种其它限制酶可获得编码活性片段的基因。蛋白酶可用于直接获得这些 δ - 内毒素的活性片段。

也可用这里提供的讲义从芽孢杆菌菌株和/或 DNA 文库中分离等价

5 δ - 内毒素和/或编码这些等价 δ - 内毒素的基因。例如，这里公开并要求保护的抗 δ - 内毒素抗体可用于从蛋白混合物中鉴定和分离其它 δ - 内毒素。具体地，可产生针对该 δ - 内毒素中最为常见而又与其它苏云金芽孢杆菌 δ - 内毒素最为不同的部分的抗体。这些抗体可用于鉴定等价的 δ - 内毒素，该内毒素通过免疫沉淀、酶联免疫分析 (ELISA) 或 Western 印迹等分析而具有特征性的杀虫活性。

10 用于鉴定本发明 δ - 内毒素和基因的进一步方法是通过使用寡核苷酸探针。这些探针为具有可检测标记的核苷酸序列。如本领域中所众所周知的，如果探针分子和核酸样品通过在二个分子间形成强键而杂交，那么可有理由认为该探针与样品是基本上一致的。探针的可检测标记提供了从已知方式检测无论什么杂交的方法。此探针分析提供了用于鉴定本发明灭蚊 δ - 内毒素基因的快速方法。

15 用作本发明探针的核苷酸片段可通过用标准的方法用 DNA 合成仪加以合成。因此核苷酸片段作为探针时，用本领域技术人员已知的任何适当的标记，包括放射和非 - 放射性标记标记特定的探针。典型的放射标记包括 ^{32}P 、 ^{125}I 、 ^{35}S 等。可用 DNA 酶和 DNA 聚合酶通过常规的缺口翻译反应从与 DNA 样品互补的核苷酸序列中构建放射性同位素标记的探针。然后可将该探针和样品混合于杂交缓冲液中并维持于适当的温度中直至退火出现。之后，洗去膜上的外源物质，一般通过放射自显影和/或液闪计数检测和定量样品和结合的探针分子。

20 非放射性标记包括，例如，配体如生物素或甲状腺素，以及酶如水解酶或过氧化物酶或各种化学发光剂如萤光素、或萤光化合物样荧光素及其衍生物。为了分离的方便，例如，也可通过用在上述一个末端以同位素标记并且在另一端用生物素标记而以不同种标记在 2 个末端同时标记探针。

25 双链形成和稳定性依赖于 2 条杂交链间实质的互补性，并且如上所述，可耐受一定程度的错配。因此，本发明的探针包括突变（单个及多个）、缺失、插入所述的序列及其组合，其中所述的突变、插入和缺失允许与目的靶核苷酸形成稳定的杂交体。可通过对普通技术人员目前已知的的方法并且或许通过将来可能弄清的其它方法以多种方式在给定的多核苷酸序列中制备突变、插入和缺失。

列出探针中潜在的变异部分是由于遗传密码的丰余性。由于遗传密码的丰余性，即不止一种编码的核苷酸三联体（密码子）可用于制备蛋白的大多数氨基酸。因此，不同的核苷酸序列可编码特定的某一氨基酸。因此，可通过编码该蛋白或肽的相同氨基酸序列的等价核苷酸序列制备苏云金芽孢杆菌 δ -内毒素和肽的氨基酸序列。因此，本发明包括这种等价的核苷酸序列。同样，反向或互补序列也是本发明的一方面并可由本领域的技术人员容易地使用。此外，已显示如果此改变不影响蛋白的二级结构可通过改变氨基酸序列而构建已鉴定过结构和功能的蛋白（Kaiser 和 Kezdy, 1984）。因此，本发明包括这里所述的不改变蛋白二级结构的氨基酸序列突变体或，如果结构被改变，但生物学活性基本上未变的氨基酸序列突变体。此外，本发明也包括具有完整或部分本发明编码 δ -内毒素基因生物的突变体。此突变体可通过对本领域技术人员众所周知的技术加以制备。例如，可用 UV 照射制备宿主生物的突变体。同样，此突变体可包括也可通过本领域众所周知的方法而制备的不产芽孢的宿主细胞。

6.0 实施例

包括进下面的实例以说明本发明优选的实施方案。本发明的技术人员应该明白在后面的实施例中公开的技术代表着由本发明者发现的技术从而在完成本发明中很好地发挥作用，且因此可认为构成该实践优选的方式。然而，在本发明的指导下，本领域中的技术人员应该明白在不偏离本发明精神和范围情况下，在已公开的特定实施方案中可制备许多改变并且仍可获得同样或类似的结果。

6.1 实施例 1-- 杂合苏云金芽孢杆菌 δ -内毒素的构建

由于本发明的苏云金芽孢杆菌穿梭载体 pEG 853、pEG 854 和 pEG 857 已描述过（Baum 等, 1990）。pEG 857 含有作为 SphI - BamHI DNA 片段而克隆入 pEG 853 中的 Cry 1 Ac 基因。以这种方式构建 pEG 1064，保存 cry 1 Ac 基因内部的 KpnI 位点并且去掉 pEG 857 多克隆位点（MCS）中的 KpnI 位点。这通过以下一系列的方法完成，包括对 pEG 857 DNA 进行有限的 KpnI 消化从而切掉唯一的 KpnI 位点。通过 DNA

聚合酶 I Klenow 片段补平 KpnI 5' 突出端从而产生平的 DNA 末端，并且通过 T4 DNA 连接酶连接此 DNA 的平末端。 pEG 318 含有作 XhoI - Sall DNA 片段而克隆入 pEG 854 XhoI 位点中的 cry 1F 基因(Chambers 等, 1991)。 pEG 315 含有作为 Sall - BamHI DNA 片段而克隆入 pEG 854 XhoI-BamHI 位点中的菌株 EG 6346 的 cry 1C 基因(Chambers 等,1991)。

图 1A 图解显示分别编码 pEG 854/pEG 1064、 pEG 20、 pEG 315 和 pEG 318 中含有的完整 cry 1Ac、 cry 1Ab、 cry 1C 和 cry 1F 基因的 DNA。同时显示了用于构建一些杂合基因的独特限制性位点。图 1B 图解显示与本发明有关的杂合基因。在有些情况，用以诱变寡核苷酸引物的标准 PCRTM 扩增将适当的限制位点掺入到用于杂合基因构建的 DNA 片段中。一些杂合基因的构建不能够通过限制性片段亚克隆来完成。在这些情况，用 PCRTM 重叠延伸 (POE) 构建所需的杂合基因 (Horton 等, 1989)。使用了下面的寡核苷酸引物 (购自整合 DNA 技术公司, Coralville, IA) :

- 15 引物 A: 5'-GGATAGCACTCATCAAAGGTACC-3' (SEQ ID NO:1)
- 引物 B: 5'-GAAGATATCCAATTTCGAACAGTTTCCC-3' (SEQ ID NO:2)
- 引物 C: 5'-CATATTCTGCCTCGAGTGTGCAGTAAC-3' (SEQ ID NO:3)
- 引物 D: 5'-CCCGATCGGCCGCATGC-3' (SEQ ID NO:4)
- 引物 E: 5'-CATTGGAGCTCTCCATG-3' (SEQ ID NO:5)
- 引物 F: 5'-GCACTACGATGTATCC-3' (SEQ ID NO:6)
- 20 引物 G: 5'-CATCGTAGTGCAACTCTTAC-3' (SEQ ID NO:7)
- 引物 H: 5'-CCAAGAAAATACTAGAGCTCTGTAAAAAAGGTGTTC-3' (SEQ ID NO:8)
- 引物 I: 5'-ATTTGAGTAATACTATCC-3' (SEQ ID NO:23)
- 引物 J: 5'-ATTACTCAAATACCATTGG-3' (SEQ ID NO:24)
- 引物 K: 5'-TCGTTGCTCTGTTCCCG-3' (SEQ ID NO:31)

25 图 1B 中所述的含有与本发明有关的杂合 δ -内毒素基因的质粒描述如下。通过限制性消化质粒 DNA、PCRTM 扩增或 POE 而产生的 DNA 片段的分离或纯化依次应用琼脂糖 - TAE 凝胶电泳并根据制造商建议使用 Geneclean 试剂盒 (Bio 101)。通过用引物对 A 和 B 以及 pEG 318 作为模板对 cry 1F DNA 片段进行 PCRTM 扩增而构建 pEG 1065。分离得到的 PCRTM 产物、以 AsuII 和 KpnI 切割并用于代替 pEG 857 中相应的 AsuII - KpnI DNA 片段。用分别从 pEG 318 和 pEG 857 中分离的

POE 和 DNA 片段, cry 1F 的 *SauI* - *KpnI* 和 cry 1Ac 的 *AsuII* - *ClaI* 构建质粒 pEG 1067。将得到的 POE 产物用引物对 A 和 B 进行 PCR™ 扩增、以 *AsuII* 和 *KpnI* 切割并且用于代替 pEG 857 中相应的 *AsuII* - *KpnI* 片段。

5 通过以从 cry 1F (pEG 318) 中分离的相应的 *SacI* - *KpnI* DNA 片段代替从 pEG 857 中分离的 cry 1Ac 的 *SacI* - *KpnI* DNA 片段构建 pEG 1068。通过以从 cry 1Ac (pEG 857) 中分离的相应的 *SacI* - *KpnI* DNA 片段代替从 pEG 1065 中分离的 *SacI* - *KpnI* DNA 片段而构建 pEG 1070。通过以从 cry 1Ac (pEG 857) 中分离的相应 *SacI* - *KpnI* DNA 片段代替从 pEG 1067 中分离的 *SacI* - *KpnI* DNA 片段而构建 pEG 1072。用引物对 C 和 D 分别以从 pEG 1065、pEG 1067、pEG 1068 中 PCR™ 扩增出的 *SphI* - *XhoI* DNA 片段代替 pEG 1064 中的 *SphI* - *XhoI* DNA 片段而构建 pEG 1074、pEG 1076 和 pEG 1077。以分离的并且以 *SphI* 和 *SacI* 切割的用引物对 D 和 E 并以模板 pEG 318 而产生的 cry 1F PCR™ 产物代替 pEG 1064 的 *SphI* - *SacI* DNA 片段而构建 pEG 1089。

通过以分离的并从 *SphI* 和 *SacI* 切割的用引物对 D 和 H 以及模板 pEG 315 而产生的 cry 1C 的 PCR™ 产物代替 pEG 1064 的 *SphI* - *SacI* DNA 片段而构建 pEG 1091。

20 用引物对 B 和 F 而产生的 cry 1Ac DNA 片段和用引物 A 和 G 而产生的 cry 1C DNA 片段通过 POE 而构建 pEG 1088。从得到的 POE 产物中分离 *SacI* - *KpnI* 片段并用于代替 pEG 1064 中的相应 *SacI* - *KpnI* 片段。

通过首先以从 pEG 20 中分离的相应 cry 1Ab DNA 片段代替 pEG 1065 的 *SphI* - *KpnI* DNA 片段以得到 pEG 364 而构建 pEG 365。

通过以从 pEG 315 中分离的相应的 DNA 片段代替 pEG 1088 中的 *KpnI* - *BamHI* DNA 片段而构建 pEG 1092。pEG 1092 不同于国际专利申请公开 No. WO 95/06730 中公开的 cry 1Ab/cry 1C 杂合 δ - 内毒素基因。

30 通过以从 pEG 20 中分离的相应的 *SphI* - *AsuII* DNA 片段代替 pEG 1068 中的 *SphI* - *AsuII* DNA 片段而构建 pEG 1093。

用引物对 B 和 I 及 pEG 857 作为模板而产生的 cry 1Ac DNA 片段和用引物对 A 和 J 及 pEG 318 作为模板而产生的 cry 1F DNA 片段通过 POE 构建 pEG 378。将得到的 POE 产物以 *AsuII* 和 *KpnI* 切割并将得到的分离的 DNA 片段用于代替 pEG 1064 中相应的 *AsuII* - *KpnI* DNA 片段。

通过从以用引物对 C 和 K 对 pEG 378 的 PCR™ 扩增中分离的相应的 *AsuII* - *XhoI* DNA 片段代替 pEG 1064 中的 *AsuII* - *XhoI* DNA 片段构建 pEG 381。

6.2 实施例 2 -- 在苏云金芽孢杆菌中杂合毒素的制备

按所述将实施例 1 中所述的编码杂合毒素的质粒转化入苏云金芽孢杆菌中 (Mettus 和 Macaluso, 1990)。将得到的苏云金芽孢杆菌菌株培养于 50ml C-2 培养基中直至该培养物完全形成芽孢并被裂解(大约 48 小时)。由于晶体的形成是在苏云金芽孢杆菌中对 δ -内毒素进行有效商业制备的前提, 故用显微镜分析来鉴定形成孢子培养物中的晶体(表 5)。

表 5
杂合 δ - 内毒素的晶体形成

| 菌株 | 质粒 | 母本 δ -内毒素 | 晶体形成 | |
|----|---------|------------------|-------------------------|---|
| 5 | EG11060 | pEG1065 | Cry1Ac + Cry1F | + |
| | EG11062 | pEG1067 | Cry1Ac + Cry1F | + |
| | EG11063 | pEG1068 | Cry1Ac + Cry1F | + |
| | EG11065 | pEG1070 | Cry1Ac + Cry1F | - |
| | EG11067 | pEG1072 | Cry1Ac + Cry1F | - |
| 10 | EG11071 | pEG1074 | Cry1Ac + Cry1F | + |
| | EG11073 | pEG1076 | Cry1Ac + Cry1F | + |
| | EG11074 | pEG1077 | Cry1Ac + Cry1F | + |
| | EG11087 | pEG1088 | Cry1Ac + Cry1C | - |
| | EG11088 | pEG1089 | Cry1F + Cry1Ac | - |
| | EG11090 | pEG1091 | Cry1C + Cry1Ac | - |
| 15 | EG11091 | pEG1092 | Cry1Ac + Cry1C | + |
| | EG11092 | pEG1093 | Cry1Ab + Cry1Ac + Cry1F | + |
| | EG11735 | pEG365 | Cry1Ab + Cry1F + Cry1Ac | + |
| | EG11751 | pEG378 | Cry1Ac + Cry1F | + |
| | EG11768 | pEG381 | Cry1Ac + Cry1F | + |

20

按 Baum 等, 1990 所述通过十二烷基磺酸钠聚丙烯酰胺凝胶电泳 (SDS - PAGE) 检测表 5 中列出的一些苏云金芽孢杆菌菌株 δ - 内毒素的产生。在 C - 2 培养基中培养每种苏云金芽孢杆菌菌株等体积的培养物直到完全形成芽孢并被裂解。离心培养物并以等体积的去离子蒸馏水 25 水洗芽胞/晶体沉淀物两次。将最终的沉淀物悬浮于一半体积的 0.005 % Triton X-100[®] 中。通过 SDS - PAGE 分析等体积的每种洗过的培养物, 如图 2 中所示。

包括 Cry 1 Ac 和 Cry 1 F 的大多数杂交体在苏云金芽孢杆菌中形成稳定的晶体。一个明显的例外是 EG 11088, 其中的活性毒素片段将是 30 EG 11063 的互换形式。包括 Cry 1 Ac 与 Cry 1 C、EG 11087 和 EG 11090 的三个杂交体中的两个未能在苏云金芽孢杆菌中产生的晶体即使

这些交互杂交体模仿活化的编码晶体的 EG 11063 和 EG 11074 的毒素片段。

由 SDS - PAGE 检测的每种菌株均产生一定水平的 δ - 内毒素。然而，如所预期的，那些鉴定为晶体阴性的培养物产生很少的蛋白（如，
5 e 道： EG 11065， f 道： EG 11067， j 道： EG 11088 和 k 道： EG 11090）。例如，显示了形成 δ - 内毒素晶体 Cry 1 Ac（ a 道）的一般产量。数种杂合 δ - 内毒素产生相近水平的蛋白，包括 EG 11060（ b 道）， EG 11062（ c 道）， EG 11063（ d 道； SEQ ID NO： 10）和 EG 11074（ i 道； SEQ ID NO： 12）。此数据清晰地显示在苏云金芽
10 孢杆菌中有效的杂合 δ - 内毒素的产量是不可预期的并且随着用于构建该杂合体的母本 δ - 内毒素而变化。

6.3 实施例 3 -- 杂合 δ - 内毒素的蛋白水解加工

一旦 δ - 内毒素晶体溶解于幼虫中肠时前毒素形式的 δ - 内毒素
15 蛋白降解成稳定活性的毒素。对 δ - 内毒素潜在活性的一个测定是活性的 δ - 内毒素在蛋白水解环境中的稳定性。为了测试此杂合 δ - 内毒素的蛋白酶水解敏感性，将溶解的毒素用胰蛋白酶消化。从形成芽胞的苏云金芽孢杆菌培养物中纯化 δ - 内毒素并按所述定量（ Chambers 等， 1991）。精确地将 250 μ g 的每种杂合 δ - 内毒素晶体溶解于 30 mM
20 NaHCO₃、 10 mM DTT（总体积 0.5 ml）中。以 1:10 的比例向溶解的毒素中加入胰蛋白酶。在适当的时间点取出 50 μ l 等份至 50 μ l Laemmli 缓冲液中，加热至 100 $^{\circ}$ C 3 分钟，并且冷冻于干冷的乙醇浴中用于随后的分析。通过 SDS - PAGE 分析溶解毒素的胰蛋白酶消化产物并且用光密度定量每个时间点活性 δ - 内毒素的量。这些研究结果图示于图 3 中。
25

快速将野生型 Cry 1 Ac 加工成在整个研究期间均稳定的活性 δ - 内毒素片段。也加工 EG 11063 和 EG 11074 的杂合 δ - 内毒素成为整个研究期间均稳定的活性 δ - 内毒素。以更慢的速度和此活性 δ - 内毒素片段以更高百分比维持在每个时间点对 EG 11063 δ - 内毒素进行加工。
30 尽管此 EG 11060 与 EG 11062 的杂合 δ - 内毒素被加工成活性的 δ - 内毒素片段，但是这些片段对进一步切割更为敏感并在研究期间以不同

的速度降解。EG 11062 和 EG 11063 δ -内毒素 cry 1 Ac 与 cry 1 F 之间的 5' 交换位点得到仅有 21 个氨基酸残基不同的毒素 (见图 1)。然而, 由于 EG 11062 δ -内毒素更为快速的降解, 在这些位置维持 Cry 1 Ac 序列的重要性是明显的。这些数据显示用相同的母本 δ -内毒素构建的不同杂合 δ -内毒素在生物化学特征如蛋白水解稳定性中可有显著不同。

6.4 实施例 4-- 杂合 δ -内毒素的生物活性

将表达所需 δ -内毒素的苏云金芽孢杆菌培养物培养至完全形成芽胞并被裂解且按实施例 2 中所述进行清洗。按所述 (Baum 等, 1990) 通过 SDS - PAGE 定量每种培养物 δ -内毒素的水平。在生物测定筛选时, 将单个适当浓度的每种洗过的 δ -内毒素培养物局部放置于每孔含有 1.0 ml 人工营养物的 30 个小孔中 (表面积 175 mm²)。将新生的单个幼虫置于每个处理过的小孔中并将栽培木箱覆盖上一层有孔的干净聚酯薄膜。摄食 7 天后计数幼虫的致死率并将百分致死率表示为死亡幼虫数目对处理的总的幼虫数目的比例, 为 32。

在 LC₅₀ 测定 (得到 50 % 致死率的 δ -内毒素浓度) 中, 从苏云金芽孢杆菌培养物中纯化 δ -内毒素并按 Chambers 等 (1991) 所述进行定量。通过在 0.005 % 的 Triton X - 100[®] 中连续稀释而制备 8 种浓度的 δ -内毒素并将每种浓度局部施用至含有 1.0 ml 人工营养物的小孔中。摄食 7 天后计数幼虫致死率 (每种 δ -内毒素浓度 32 只幼虫)。在所有情况, 将稀释剂用作对照。

通过生物测定筛选对 Cry 1 A / cry 1 F 杂合毒素的比较示于表 6 中。菌株 EG 11063 和 EG 11074 的杂合 δ -内毒素维持有母本 Cry 1 A 和 Cry 1 F δ -内毒素的活性。此外, EG 11735 杂合 δ -内毒素维持有其母本 Cry 1 Ab 和 Cry 1 F δ -内毒素的活性。由菌株 EG 11061、EG 11062、EG 11071 和 EG 11073 所产生的 δ -内毒素对测试的昆虫幼虫没有杀虫活性尽管其 1) 至少包括一种对抵抗所示幼虫有活性的母本 δ -内毒素和 2) 在苏云金芽孢杆菌中形成稳定、十分确切的晶体。这些结果显示杂合毒素构建体不可预期的性质。

数据列于表 6 中。测试所有菌株洗过的形成芽胞的培养物。对于每

种测试的昆虫，使用相同量的 δ - 内毒素并且杀虫活性是基于表现出最高百分致死率的菌株 (++++)。

表 6

5

杂合 CRY1A/CRY1F δ - 内毒素的生物测定筛选

| 菌株 | 草地贪夜蛾 | 甜菜夜蛾 | 烟草夜蛾 | 玉米穗夜蛾 | 欧洲玉米螟 |
|---------|-------------------|------|------|-------|-------|
| Cry1Ac | - | - | ++++ | ++++ | +++ |
| Cry1F | ++++ | ++ | ++ | ++ | ++ |
| Cry1Ab | ++ | + | +++ | ++ | +++ |
| EG11060 | - | - | - | - | - |
| EG11062 | - | - | - | - | - |
| EG11063 | ++++ | ++++ | +++ | +++ | ++++ |
| EG11071 | - | - | - | - | - |
| EG11073 | - | - | - | - | - |
| EG11074 | ++++ | ++++ | +++ | +++ | ++++ |
| EG11090 | - | +++ | - | - | - |
| EG11091 | ++++ | ++++ | - | - | N.D. |
| EG11092 | ++++ | ++++ | +++ | +++ | N.D. |
| EG11735 | ++++ | ++++ | +++ | +++ | N.D. |
| EG11751 | N.D. ^a | ++++ | N.D. | ++++ | N.D. |

^aN. D. = 未测定。

测试图 1 中所述的并在生物检测筛选中显示出杀虫活性的 δ - 内毒素纯化晶体以测定其 LC_{50} (见表 7)。与任何一种测试过的野生型 δ - 内毒素相比，从菌株 EG 11063、EG 11074、EG 11091 和 EG 11735 中纯化的 δ - 内毒素均显示出增加的 (抗) 粘虫 (草地贪夜蛾和甜菜夜蛾) 活性。EG 11063 和 EG 11074 δ - 内毒素将产生相同活性的毒素片段 (图 1 B)，这由其对检测过的昆虫的类似 LC_{50} 值所证明。这些数据所证明的没有预料到的结果是杂合 δ - 内毒素如 EG 11063、EG 11092、EG 11074、EG 11735 或 EG 11751 可维持其各自母本 δ - 内

5 毒素活性并且抗一些昆虫如甜菜夜蛾，并且可能具有较任何一种母本 δ - 内毒素大得多的活性。接近或低于母本 δ - 内毒素时该广泛的杀虫活性以及野生型水平的毒素产生（实施例 2）使这些蛋白特别适用于在苏云金芽孢杆菌中产生。虽然 EG 11091 来源的 δ - 内毒素较其母本 δ - 内毒素具有更好的抗草地贪夜蛾和甜菜夜蛾活性，但其丧失母本 Cry 1Ac 的烟草夜蛾和玉米穗夜蛾活性。EG 11091 δ - 内毒素（实施例 2）有限的宿主范围以及观察到的降低的毒素产量使其较少适于在苏云金芽孢杆菌中产生。

10

表 7

纯化杂合 δ - 内毒素的 LC_{50} 值^A

| 毒素 | 草地贪夜蛾 | 甜菜夜蛾 | 烟草夜蛾 | 玉米穗夜蛾 | 欧洲玉米螟 |
|------------------------|--------|--------|--------|--------|-------------------|
| Cry1Ac | >10000 | >10000 | 9 | 100 | 23 |
| Cry1Ab | 1435 | 4740 | 118 | 400 | 17 |
| Cry1C | >10000 | 490 | >10000 | >10000 | >10000 |
| Cry1F | 1027 | 3233 | 54 | 800 | 51 |
| EG11063 (Cry1Ac/1F) | 550 | 114 | 33 | 80 | 7 |
| EG11074 (Cry1Ac/1F) | 468 | 77 | 25 | 76 | 9 |
| EG11091 (Cry1Ac/1C) | 21 | 21 | 219 | >10000 | N.D. ^a |

^aN. D. = 未测定。

15 在表 7 中， LC_{50} 值以每孔（ 175 mm^2 ）中纯化的 δ - 内毒素的纳克数表示并且是 2 到 6 个同样样品的复合（composite）值。

表 8

CRY1 杂合 δ - 内毒素的 DNA 交换位点

| 质粒 | SEQ ID NO: | 5' 交换位点 | SEQ ID NO: | 3' 交换位点 |
|---------|------------|------------------------|------------|----------------------|
| pEG1065 | 15 | TATCCAATTGGAACGTCATC | 21 | ACTACCAGGTACCTTTGATG |
| pEG1067 | 16 | TTTAGTCATCGATTAAATCA | 21 | ACTACCAGGTACCTTTGATG |
| pEG1068 | 17 | ATAATAAGAGCTCCAATGTT | 21 | ACTACCAGGTACCTTTGATG |
| pEG1070 | 15 | TATCCAATTGGAACGTCATC | 19 | TCAJGGAGAGCTCCTATGTT |
| pEG1072 | 16 | TTTAGTCATCGATTAAATCA | 19 | TCAJGGAGAGCTCCTATGTT |
| pEG1074 | 15 | TATCCAATTGGAACGTCATC | 35 | TGCAACACTCGAGGCTGAAT |
| pEG1076 | 16 | TTTAGTCATCGATTAAATCA | 35 | TGCAACACTCGAGGCTGAAT |
| pEG1077 | 17 | ATAATAAGAGCTCCAATGTT | 35 | TGCAACACTCGAGGCTGAAT |
| pEG1088 | 18 | TACATCGTAGTGCAACTCTT | 22 | ACTACCGGGTACCTTTGATA |
| pEG1089 | 19 | TCATGGAGAGCTCCTATGTT | - | NA |
| pEG1091 | 20 | TTAACAAAGAGCTCCTATGTT | - | NA |
| pEG1092 | 18 | TACATCGTAGTGCAACTCTT | - | NA |
| pEG1093 | - | ND ^b | 21 | ACTACCAGGTACCTTTGATG |
| pEG365 | 17 | ATAATAAGAGCTCCAATGTT | 21 | ACTACCAGGTACCTTTGATG |
| pEG378 | 32 | TCAAAATACCCATTGGTAAAAG | 21 | ACTACCAGGTACCTTTGATG |
| pEG381 | 32 | TCAAAATACCCATTGGTAAAAG | 35 | TGCAACACTCGAGGCTGAAT |

^aNA = 不可施用。这些杂合毒素仅含有一个交换位点，如图 1 中所示。

^bND = 不可区分。这些杂合蛋白的交换位点由与任何一种母本毒素不可区分的 DNA 序列所鉴定。

表 8 描述了与本发明有关的杂合 δ - 内毒素 5' 和 3' 交换位点周围的 DNA。由该 SEQ ID NO 所证实, 有些杂合 δ - 内毒素具有相同的交换位点。

为了检测其它小的改变在选定用于杂合内毒素构建的交换位点中的作用, 比较 EG 11751 和 EG 11063 对甜菜夜蛾和玉米穗夜蛾的活性。这些数据清楚地显示通过改变 2 个母本 δ - 内毒素之间的交换位点可得到对杂合 δ - 内毒素的改进。在该实施例中, 与 EG 11063 δ - 内毒素相比, EG 11751 δ - 内毒素中的交换位点去除了 3' 的 75 个碱基对并且得到增强的杀虫活性。虽然在 EG 11063 和 EG 11751 之间没有观察到对甜菜夜蛾活性的显著改进, 但观察到 EG 11751 对玉米穗夜蛾活性有几乎 4 倍的明显提高。注意到通过改变交换位点而提高杂合 δ - 内毒素生物活性是不可预期的, 这点很重要。在 EG 11062 情况时, 去除 EG 11063 交换位点 5' 的 63 个碱基对交换位点破坏了其杀虫活性, 如表 8 中所示。

15

表 9

EG 11063 和 EG 11751 的生物活性

| 苏云金芽孢杆菌菌株 | 清洗过形成芽胞培养物的 LC ₅₀ 值 | |
|-----------|--------------------------------|-------|
| | 甜菜夜蛾 | 玉米穗夜蛾 |
| EG11063 | 106 | 38 |
| EG 11751 | 90 | 10 |

为了进一步检测杂合 δ - 内毒素交换位点中改变的效应, 比较由 pEG 381 编码的杂合 δ - 内毒素与由 pEG 378 和 pEG 1068 所编码的杂合 δ - 内毒素。在此实施例中, 与 pEG 378 杂合 δ - 内毒素相比, 编码 pEG 381 的杂合 δ - 内毒素的 3' 交换位点被去除 5' 的 340 个碱基对。表 9 中的数据显示与编码 pEG 378 和 pEG 1066 的 δ - 内毒素相比, 此改变得到对草地贪夜蛾活性的增加而同时维持所观察到的编码 pEG 378 δ - 内毒素较编码 pEG 1068 的 δ - 内毒素增加的活性 (参见表 8)。这些结果是没有预料到的因为来自编码 pEG 378 和 pEG 381 δ - 内毒素蛋白水解的活化的毒素应该是相同的。此实施例进一步证明 δ - 内毒素的前毒素片段内的交换位点对杀虫活性可具有复杂的效应。

25



表 10

由 pEG 378、 pEG 381 和 pEG 1068 编码的毒素的生物活性

| 质粒 | 纯化晶体的 LC ₅₀ 值 | | | |
|----------|--------------------------|-------|-------|---------------|
| | 草地贪夜蛾 | T. ni | 玉米穗夜蛾 | P. xylostella |
| pEG 378 | 464 | 57.7 | 37.5 | 3.02 |
| pEG 381 | 274 | 56.0 | 36.6 | 2.03 |
| pEG 1068 | 476 | 66.7 | 72.7 | 3.83 |

6.5 实施例 5 -- 杂合毒素对其它害虫的活性

- 5 也对本发明的毒素抗其它害虫的活性进行生物测定，包括西南玉米蛀虫和两种大豆害虫。溶解毒素蛋白、加入营养物中并生物分析其抗靶害虫的活性。该杂合毒素对所有三种害虫表现出很有效的控制。

表 11

杂合毒素对西南玉米蛀虫的 LC₅₀ 和 EC₅₀ 范围^{1,2}

| | EG 11063 | EG 11074 | EG 11091 | EG 11751 |
|------------------|----------|----------|----------|----------|
| LC ₅₀ | 20 | 10 - 20 | 10 - 20 | 10 - 20 |
| EC ₅₀ | 0.2 - 2 | 0.2 - 2 | 0.2 - 2 | 0.2 - 2 |

¹ 所有值以 µg/ml 营养物表示。

² SWCB 数据范围分别代表 LC₅₀ 和 EC₅₀ 范围 (由 > 第 1 龄期虫 % 所测定)。

5

表 12

嵌合晶体蛋白对大豆害虫的 LC₅₀ 值¹

| 害虫 | EG 11063 | EG 11074 | EG 11091 | EG 11751 | EG 11768 |
|-------|----------|----------|----------|----------|----------|
| 绒毛豆毛虫 | 0.9 | 0.6 | 0.3 | 0.1 | 0.06 |
| 大豆尺蠖 | 0.9 | 0.8 | 0.6 | 0.7 | 0.2 |

¹ 所有值以 µg/ml 营养物表示。

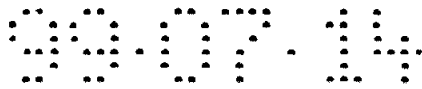
² 绒毛豆毛虫 (*Anticarsia gemmatalis*) 和大豆尺蠖 (*Pseudoplusi includens*) 均为夜蛾科成员。

10

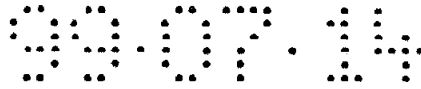
6.6 实施例 6 -- 新的晶体蛋白的氨基酸序列

6.6.1 EG 11063 晶体蛋白的氨基酸序列 (SEQ ID NO: 10)

MetAspAsnAsnProAsnIleAsnGluCysIleProTyrAsnCysLeuSerAsnProGluValGluValLeu
 GlyGlyGluArgIleGluThrGlyTyrThrProIleAspIleSerLeuSerLeuThrGlnPheLeuLeuSer
 GluPheValProGlyAlaGlyPheValLeuGlyLeuValAspIleIleTrpGlyIlePheGlyProSerGln
 TrpAspAlaPheLeuValGlnIleGluGlnLeuIleAsnGlnArgIleGluGluPheAlaArgAsnGlnAla
 IleSerArgLeuGluGlyLeuSerAsnLeuTyrGlnIleTyrAlaGluSerPheArgGluTrpGluAlaAsp
 ProThrAsnProAlaLeuArgGluGluMetArgIleGlnPheAsnAspMetAsnSerAlaLeuThrThrAla
 IleProLeuPheAlaValGlnAsnTyrGlnValProLeuLeuSerValTyrValGlnAlaAlaAsnLeuHis
 LeuSerValLeuArgAspValSerValPheGlyGlnArgTrpGlyPheAspAlaAlaThrIleAsnSerArg
 TyrAsnAspLeuThrArgLeuIleGlyAsnTyrThrAspTyrAlaValArgTrpTyrAsnThrGlyLeuGlu
 ArgValTrpGlyProAspSerArgAspTrpValArgTyrAsnGlnPheArgArgGluLeuThrLeuThrVal
 LeuAspIleValAlaLeuPheProAsnTyrAspSerArgArgTyrProIleArgThrValSerGlnLeuThr
 ArgGluIleTyrThrAsnProValLeuGluAsnPheAspGlySerPheArgGlySerAlaGlnGlyIleGlu
 ArgSerIleArgSerProHisLeuMetAspIleLeuAsnSerIleThrIleTyrThrAspAlaHisArgGly

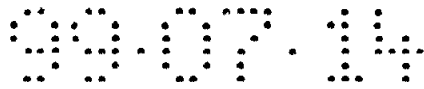


TyrTyrTyrTrpSerGlyHisGlnIleMetAlaSerProValGlyPheSerGlyProGluPheThrPhePro
LeuTyrGlyThrMetGlyAsnAlaAlaProGlnGlnArgIleValAlaGlnLeuGlyGlnGlyValTyrArg
ThrLeuSerSerThrLeuTyrArgArgProPheAsnIleGlyIleAsnAsnGlnGlnLeuSerValLeuAsp
GlyThrGluPheAlaTyrGlyThrSerSerAsnLeuProSerAlaValTyrArgLysSerGlyThrValAsp
5 SerLeuAspGluIleProProGlnAsnAsnAsnValProProArgGlnGlyPheSerHisArgLeuSerHis
ValSerMetPheArgSerGlyPheSerAsnSerSerValSerIleIleArgAlaProMetPheSerTrpThr
HisArgSerAlaThrProThrAsnThrIleAspProGluArgIleThrGlnIleProLeuValLysAlaHis
ThrLeuGlnSerGlyThrThrValValArgGlyProGlyPheThrGlyGlyAspIleLeuArgArgThrSer
GlyGlyProPheAlaTyrThrIleValAsnIleAsnGlyGlnLeuProGlnArgTyrArgAlaArgIleArg
10 TyrAlaSerThrThrAsnLeuArgIleTyrValThrValAlaGlyGluArgIlePheAlaGlyGlnPheAsn
LysThrMetAspThrGlyAspProLeuThrPheGlnSerPheSerTyrAlaThrIleAsnThrAlaPheThr
PheProMetSerGlnSerSerPheThrValGlyAlaAspThrPheSerSerGlyAsnGluValTyrIleAsp
ArgPheGluLeuIleProValThrAlaThrPheGluAlaGluTyrAspLeuGluArgAlaGlnLysAlaVal
AsnAlaLeuPheThrSerIleAsnGlnIleGlyIleLysThrAspValThrAspTyrHisIleAspGlnVal
15 SerAsnLeuValAspCysLeuSerAspGluPheCysLeuAspGluLysArgGluLeuSerGluLysValLys
HisAlaLysArgLeuSerAspGluArgAsnLeuLeuGlnAspProAsnPheLysGlyIleAsnArgGlnLeu
AspArgGlyTrpArgGlySerThrAspIleThrIleGlnArgGlyAspAspValPheLysGluAsnTyrVal
ThrLeuProGlyThrPheAspGluCysTyrProThrTyrLeuTyrGlnLysIleAspGluSerLysLeuLys
AlaPheThrArgTyrGlnLeuArgGlyTyrIleGluAspSerGlnAspLeuGluIleTyrLeuIleArgTyr
20 AsnAlaLysHisGluThrValAsnValProGlyThrGlySerLeuTrpProLeuSerAlaGlnSerProIle
GlyLysCysGlyGluProAsnArgCysAlaProHisLeuGluTrpAsnProAspLeuAspCysSerCysArg
AspGlyGluLysCysAlaHisHisSerHisHisPheSerLeuAspIleAspValGlyCysThrAspLeuAsn
GluAspLeuGlyValTrpValIlePheLysIleLysThrGlnAspGlyHisAlaArgLeuGlyAsnLeuGlu
PheLeuGluGluLysProLeuValGlyGluAlaLeuAlaArgValLysArgAlaGluLysLysTrpArgAsp
25 LysArgGluLysLeuGluTrpGluThrAsnIleValTyrLysGluAlaLysGluSerValAspAlaLeuPhe
ValAsnSerGlnTyrAspGlnLeuGlnAlaAspThrAsnIleAlaMetIleHisAlaAlaAspLysArgVal
HisSerIleArgGluAlaTyrLeuProGluLeuSerValIleProGlyValAsnAlaAlaIlePheGluGlu
LeuGluGlyArgIlePheThrAlaPheSerLeuTyrAspAlaArgAsnValIleLysAsnGlyAspPheAsn
AsnGlyLeuSerCysTrpAsnValLysGlyHisValAspValGluGluGlnAsnAsnGlnArgSerValLeu
30 ValValProGluTrpGluAlaGluValSerGlnGluValArgValCysProGlyArgGlyTyrIleLeuArg
ValThrAlaTyrLysGluGlyTyrGlyGluGlyCysValThrIleHisGluIleGluAsnAsnThrAspGlu
LeuLysPheSerAsnCysValGluGluGluIleTyrProAsnAsnThrValThrCysAsnAspTyrThrVal
AsnGlnGluGluTyrGlyGlyAlaTyrThrSerArgAsnArgGlyTyrAsnGluAlaProSerValProAla
AspTyrAlaSerValTyrGluGluLysSerTyrThrAspGlyArgArgGluAsnProCysGluPheAsnArg
35 GlyTyrArgAspTyrThrProLeuProValGlyTyrValThrLysGluLeuGluTyrPheProGluThrAsp
LysValTrpIleGluIleGlyGluThrGluGlyThrPheIleValAspSerValGluLeuLeuLeuMetGlu
Glu



6.6.2 EG 11074 晶体蛋白的氨基酸序列 (SEQ ID NO: 12)

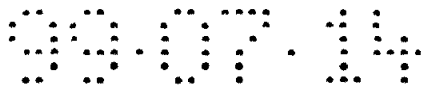
MetAspAsnAsnProAsnIleAsnGluCysIleProTyrAsnCysLeuSerAsnProGluValGluValLeu
GlyGlyGluArgIleGluThrGlyTyrThrProIleAspIleSerLeuSerLeuThrGlnPheLeuLeuSer
GluPheValProGlyAlaGlyPheValLeuGlyLeuValAspIleIleTrpGlyIlePheGlyProSerGln
5 TrpAspAlaPheLeuValGlnIleGluGlnLeuIleAsnGlnArgIleGluGluPheAlaArgAsnGlnAla
IleSerArgLeuGluGlyLeuSerAsnLeuTyrGlnIleTyrAlaGluSerPheArgGluTrpGluAlaAsp
ProThrAsnProAlaLeuArgGluGluMetArgIleGlnPheAsnAspMetAsnSerAlaLeuThrThrAla
IleProLeuPheAlaValGlnAsnTyrGlnValProLeuLeuSerValTyrValGlnAlaAlaAsnLeuHis
LeuSerValLeuArgAspValSerValPheGlyGlnArgTrpGlyPheAspAlaAlaThrIleAsnSerArg
10 TyrAsnAspLeuThrArgLeuIleGlyAsnTyrThrAspTyrAlaValArgTrpTyrAsnThrGlyLeuGlu
ArgValTrpGlyProAspSerArgAspTrpValArgTyrAsnGlnPheArgArgGluLeuThrLeuThrVal
LeuAspIleValAlaLeuPheProAsnTyrAspSerArgArgTyrProIleArgThrValSerGlnLeuThr
ArgGluIleTyrThrAsnProValLeuGluAsnPheAspGlySerPheArgGlySerAlaGlnGlyIleGlu
ArgSerIleArgSerProHisLeuMetAspIleLeuAsnSerIleThrIleTyrThrAspAlaHisArgGly
15 TyrTyrTyrTrpSerGlyHisGlnIleMetAlaSerProValGlyPheSerGlyProGluPheThrPhePro
LeuTyrGlyThrMetGlyAsnAlaAlaProGlnGlnArgIleValAlaGlnLeuGlyGlnGlyValTyrArg
ThrLeuSerSerThrLeuTyrArgArgProPheAsnIleGlyIleAsnAsnGlnGlnLeuSerValLeuAsp
GlyThrGluPheAlaTyrGlyThrSerSerAsnLeuProSerAlaValTyrArgLysSerGlyThrValAsp
SerLeuAspGluIleProProGlnAsnAsnAsnValProProArgGlnGlyPheSerHisArgLeuSerHis
20 ValSerMetPheArgSerGlyPheSerAsnSerSerValSerIleIleArgAlaProMetPheSerTrpThr
HisArgSerAlaThrProThrAsnThrIleAspProGluArgIleThrGlnIleProLeuValLysAlaHis
ThrLeuGlnSerGlyThrThrValValArgGlyProGlyPheThrGlyGlyAspIleLeuArgArgThrSer
GlyGlyProPheAlaTyrThrIleValAsnIleAsnGlyGlnLeuProGlnArgTyrArgAlaArgIleArg
TyrAlaSerThrThrAsnLeuArgIleTyrValThrValAlaGlyGluArgIlePheAlaGlyGlnPheAsn
25 LysThrMetAspThrGlyAspProLeuThrPheGlnSerPheSerTyrAlaThrIleAsnThrAlaPheThr
PheProMetSerGlnSerSerPheThrValGlyAlaAspThrPheSerSerGlyAsnGluValTyrIleAsp
ArgPheGluLeuIleProValThrAlaThrLeuGluAlaGluTyrAsnLeuGluArgAlaGlnLysAlaVal
AsnAlaLeuPheThrSerThrAsnGlnLeuGlyLeuLysThrAsnValThrAspTyrHisIleAspGlnVal
SerAsnLeuValThrTyrLeuSerAspGluPheCysLeuAspGluLysArgGluLeuSerGluLysValLys
30 HisAlaLysArgLeuSerAspGluArgAsnLeuLeuGlnAspSerAsnPheLysAspIleAsnArgGlnPro
GluArgGlyTrpGlyGlySerThrGlyIleThrIleGlnGlyGlyAspAspValPheLysGluAsnTyrVal
ThrLeuSerGlyThrPheAspGluCysTyrProThrTyrLeuTyrGlnLysIleAspGluSerLysLeuLys
AlaPheThrArgTyrGlnLeuArgGlyTyrIleGluAspSerGlnAspLeuGluIleTyrLeuIleArgTyr
AsnAlaLysHisGluThrValAsnValProGlyThrGlySerLeuTrpProLeuSerAlaGlnSerProIle
35 GlyLysCysGlyGluProAsnArgCysAlaProHisLeuGluTrpAsnProAspLeuAspCysSerCysArg
AspGlyGluLysCysAlaHisHisSerHisHisPheSerLeuAspIleAspValGlyCysThrAspLeuAsn
GluAspLeuGlyValTrpValIlePheLysIleLysThrGlnAspGlyHisAlaArgLeuGlyAsnLeuGlu
PheLeuGluGluLysProLeuValGlyGluAlaLeuAlaArgValLysArgAlaGluLysLysTrpArgAsp



LysArgGluLysLeuGluTrpGluThrAsnIleValTyrLysGluAlaLysGluSerValAspAlaLeuPhe
ValAsnSerGlnTyrAspGlnLeuGlnAlaAspThrAsnIleAlaMetIleHisAlaAlaAspLysArgVal
HisSerIleArgGluAlaTyrLeuProGluLeuSerValIleProGlyValAsnAlaAlaIlePheGluGlu
LeuGluGlyArgIlePheThrAlaPheSerLeuTyrAspAlaArgAsnValIleLysAsnGlyAspPheAsn
5 AsnGlyLeuSerCysTrpAsnValLysGlyHisValAspValGluGluGlnAsnAsnGlnArgSerValLeu
ValValProGluTrpGluAlaGluValSerGlnGluValArgValCysProGlyArgGlyTyrIleLeuArg
ValThrAlaTyrLysGluGlyTyrGlyGluGlyCysValThrIleHisGluIleGluAsnAsnThrAspGlu
LeuLysPheSerAsnCysValGluGluGluIleTyrProAsnAsnThrValThrCysAsnAspTyrThrVal
AsnGlnGluGluTyrGlyGlyAlaTyrThrSerArgAsnArgGlyTyrAsnGluAlaProSerValProAla
10 AspTyrAlaSerValTyrGluGluLysSerTyrThrAspGlyArgArgGluAsnProCysGluPheAsnArg
GlyTyrArgAspTyrThrProLeuProValGlyTyrValThrLysGluLeuGluTyrPheProGluThrAsp
LysValTrpIleGluIleGlyGluThrGluGlyThrPheIleValAspSerValGluLeuLeuLeuMetGlu
Glu

15 **6.6.3 EG 11735 晶体蛋白的氨基酸序列 (SEQ ID NO: 14)**

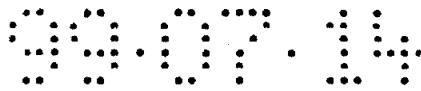
MetAspAsnAsnProAsnIleAsnGluCysIleProTyrAsnCysLeuSerAsnProGluValGluValLeu
GlyGlyGluArgIleGluThrGlyTyrThrProIleAspIleSerLeuSerLeuThrGlnPheLeuLeuSer
GluPheValProGlyAlaGlyPheValLeuGlyLeuValAspIleIleTrpGlyIlePheGlyProSerGln
TrpAspAlaPheLeuValGlnIleGluGlnLeuIleAsnGlnArgIleGluGluPheAlaArgAsnGlnAla
20 IleSerArgLeuGluGlyLeuSerAsnLeuTyrGlnIleTyrAlaGluSerPheArgGluTrpGluAlaAsp
ProThrAsnProAlaLeuArgGluGluMetArgIleGlnPheAsnAspMetAsnSerAlaLeuThrThrAla
IleProLeuPheAlaValGlnAsnTyrGlnValProLeuLeuSerValTyrValGlnAlaAlaAsnLeuHis
LeuSerValLeuArgAspValSerValPheGlyGlnArgTrpGlyPheAspAlaAlaThrIleAsnSerArg
TyrAsnAspLeuThrArgLeuIleGlyAsnTyrThrAspHisAlaValArgTrpTyrAsnThrGlyLeuGlu
25 ArgValTrpGlyProAspSerArgAspTrpIleArgTyrAsnGlnPheArgArgGluLeuThrLeuThrVal
LeuAspIleValSerLeuPheProAsnTyrAspSerArgThrTyrProIleArgThrValSerGlnLeuThr
ArgGluIleTyrThrAsnProValLeuGluAsnPheAspGlySerPheArgGlySerAlaGlnGlyIleGlu
GlySerIleArgSerProHisLeuMetAspIleLeuAsnSerIleThrIleTyrThrAspAlaHisArgGly
GluTyrTyrTrpSerGlyHisGlnIleMetAlaSerProValGlyPheSerGlyProGluPheThrPhePro
30 LeuTyrGlyThrMetGlyAsnAlaAlaProGlnGlnArgIleValAlaGlnLeuGlyGlnGlyValTyrArg
ThrLeuSerSerThrLeuTyrArgArgProPheAsnIleGlyIleAsnAsnGlnGlnLeuSerValLeuAsp
GlyThrGluPheAlaTyrGlyThrSerSerAsnLeuProSerAlaValTyrArgLysSerGlyThrValAsp
SerLeuAspGluIleProProGlnAsnAsnAsnValProProArgGlnGlyPheSerHisArgLeuSerHis
ValSerMetPheArgSerGlyPheSerAsnSerSerValSerIleIleArgAlaProMetPheSerTrpThr
35 HisArgSerAlaThrProThrAsnThrIleAspProGluArgIleThrGlnIleProLeuValLysAlaHis
ThrLeuGlnSerGlyThrThrValValArgGlyProGlyPheThrGlyGlyAspIleLeuArgArgThrSer
GlyGlyProPheAlaTyrThrIleValAsnIleAsnGlyGlnLeuProGlnArgTyrArgAlaArgIleArg



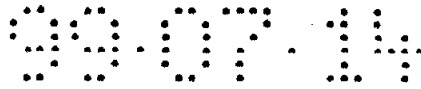
TyrAlaSerThrThrAsnLeuArgIleTyrValThrValAlaGlyGluArgIlePheAlaGlyGlnPheAsn
LysThrMetAspThrGlyAspProLeuThrPheGlnSerPheSerTyrAlaThrIleAsnThrAlaPheThr
PheProMetSerGlnSerSerPheThrValGlyAlaAspThrPheSerSerGlyAsnGluValTyrIleAsp
ArgPheGluLeuIleProValThrAlaThrPheGluAlaGluTyrAspLeuGluArgAlaGlnLysAlaVal
5 AsnAlaLeuPheThrSerIleAsnGlnIleGlyIleLysTnrAspValThrAspTyrHisIleAspGlnVal
SerAsnLeuValAspCysLeuSerAspGluPheCysLeuAspGluLysArgGluLeuSerGluLysValLys
HisAlaLysArgLeuSerAspGluArgAsnLeuLeuGlnAspProAsnPheLysGlyIleAsnArgGlnLeu
AspArgGlyTrpArgGlySerThrAspIleThrIleGlnArgGlyAspAspValPheLysGluAsnTyrVal
ThrLeuProGlyThrPheAspGluCysTyrProThrTyrLeuTyrGlnLysIleAspGluSerLysLeuLys
10 AlaPheThrArgTyrGlnLeuArgGlyTyrIleGluAspSerGlnAspLeuGluIleTyrLeuIleArgTyr
AsnAlaLysHisGluThrValAsnValProGlyThrGlySerLeuTrpProLeuSerAlaGlnSerProIle
GlyLysCysGlyGluProAsnArgCysAlaProHisLeuGluTrpAsnProAspLeuAspCysSerCysArg
AspGlyGluLysCysAlaHisHisSerHisHisPheSerLeuAspIleAspValGlyCysThrAspLeuAsn
GluAspLeuGlyValTrpValIlePheLysIleLysThrGlnAspGlyHisAlaArgLeuGlyAsnLeuGlu
15 PheLeuGluGluLysProLeuValGlyGluAlaLeuAlaArgValLysArgAlaGluLysLysTrpArgAsp
LysArgGluLysLeuGluTrpGluThrAsnIleValTyrLysGluAlaLysGluSerValAspAlaLeuPhe
ValAsnSerGlnTyrAspGlnLeuGlnAlaAspThrAsnIleAlaMetIleHisAlaAlaAspLysArgVal
HisSerIleArgGluAlaTyrLeuProGluLeuSerValIleProGlyValAsnAlaAlaIlePheGluGlu
LeuGluGlyArgIlePheThrAlaPheSerLeuTyrAspAlaArgAsnValIleLysAsnGlyAspPheAsn
20 AsnGlyLeuSerCysTrpAsnValLysGlyHisValAspValGluGluGlnAsnAsnGlnArgSerValLeu
ValValProGluTrpGluAlaGluValSerGlnGluValArgValCysProGlyArgGlyTyrIleLeuArg
ValThrAlaTyrLysGluGlyTyrGlyGluGlyCysValThrIleHisGluIleGluAsnAsnThrAspGlu
LeuLysPheSerAsnCysValGluGluGluIleTyrProAsnAsnThrValThrCysAsnAspTyrThrVal
AsnGlnGluGluTyrGlyGlyAlaTyrThrSerArgAsnArgGlyTyrAsnGluAlaProSerValProAla
25 AspTyrAlaSerValTyrGluGluLysSerTyrThrAspGlyArgArgGluAsnProCysGluPheAsnArg
GlyTyrArgAspTyrThrProLeuProValGlyTyrValThrLysGluLeuGluTyrPheProGluThrAsp
LysValTrpIleGluIleGlyGluThrGluGlyThrPheIleValAspSerValGluLeuLeuLeuMetGlu
Glu

30 6.6.4 EG 11092 晶体蛋白的氨基酸序列 (SEQ ID NO : 26)

MetAspAsnAsnProAsnIleAsnGluCysIleProTyrAsnCysLeuSerAsnProGluValGluValLeu
GlyGlyGluArgIleGluThrGlyTyrThrProIleAspIleSerLeuSerLeuThrGlnPheLeuLeuSer
GluPheValProGlyAlaGlyPheValLeuGlyLeuValAspIleIleTrpGlyIlePheGlyProSerGln
TrpAspAlaPheLeuValGlnIleGluGlnLeuIleAsnGlnArgIleGluGluPheAlaArgAsnGlnAla
35 IleSerArgLeuGluGlyLeuSerAsnLeuTyrGlnIleTyrAlaGluSerPheArgGluTrpGluAlaAsp
ProThrAsnProAlaLeuArgGluGluMetArgIleGlnPheAsnAspMetAsnSerAlaLeuThrThrAla
IleProLeuPheAlaValGlnAsnTyrGlnValProLeuLeuSerValTyrValGlnAlaAlaAsnLeuHis



LeuSerValLeuArgAspValSerValPheGlyGlnArgTrpGlyPheAspAlaAlaThrIleAsnSerArg
TyrAsnAspLeuThrArgLeuIleGlyAsnTyrThrAspHisAlaValArgTrpTyrAsnThrGlyLeuGlu
ArgValTrpGlyProAspSerArgAspTrpIleArgTyrAsnGlnPheArgArgGluLeuThrLeuThrVal
LeuAspIleValSerLeuPheProAsnTyrAspSerArgThrTyrProIleArgThrValSerGlnLeuThr
5 ArgGluIleTyrThrAsnProValLeuGluAsnPheAspGlySerPheArgGlySerAlaGlnGlyIleGlu
ArgSerIleArgSerProHisLeuMetAspIleLeuAsnSerIleThrIleTyrThrAspAlaHisArgGly
TyrTyrTyrTrpSerGlyHisGlnIleMetAlaSerProValGlyPheSerGlyProGluPheThrPhePro
LeuTyrGlyThrMetGlyAsnAlaAlaProGlnGlnArgIleValAlaGlnLeuGlyGlnGlyValTyrArg
ThrLeuSerSerThrLeuTyrArgArgProPheAsnIleGlyIleAsnAsnGlnGlnLeuSerValLeuAsp
10 GlyThrGluPheAlaTyrGlyThrSerSerAsnLeuProSerAlaValTyrArgLysSerGlyThrValAsp
SerLeuAspGluIleProProGlnAsnAsnAsnValProProArgGlnGlyPheSerHisArgLeuSerHis
ValSerMetPheArgSerGlyPheSerAsnSerSerValSerIleIleArgAlaProMetPheSerTrpThr
HisArgSerAlaThrProThrAsnThrIleAspProGluArgIleThrGlnIleProLeuValLysAlaHis
ThrLeuGlnSerGlyThrThrValValArgGlyProGlyPheThrGlyGlyAspIleLeuArgArgThrSer
15 GlyGlyProPheAlaTyrThrIleValAsnIleAsnGlyGlnLeuProGlnArgTyrArgAlaArgIleArg
TyrAlaSerThrThrAsnLeuArgIleTyrValThrValAlaGlyGluArgIlePheAlaGlyGlnPheAsn
LysThrMetAspThrGlyAspProLeuThrPheGlnSerPheSerTyrAlaThrIleAsnThrAlaPheThr
PheProMetSerGlnSerSerPheThrValGlyAlaAspThrPheSerSerGlyAsnGluValTyrIleAsp
ArgPheGluLeuIleProValThrAlaThrPheGluAlaGluTyrAspLeuGluArgAlaGlnLysAlaVal
20 AsnAlaLeuPheThrSerIleAsnGlnIleGlyIleLysThrAspValThrAspTyrHisIleAspGlnVal
SerAsnLeuValAspCysLeuSerAspGluPheCysLeuAspGluLysArgGluLeuSerGluLysValLys
HisAlaLysArgLeuSerAspGluArgAsnLeuLeuGlnAspProAsnPheLysGlyIleAsnArgGlnLeu
AspArgGlyTrpArgGlySerThrAspIleThrIleGlnArgGlyAspAspValPheLysGluAsnTyrVal
ThrLeuProGlyThrPheAspGluCysTyrProThrTyrLeuTyrGlnLysIleAspGluSerLysLeuLys
25 AlaPheThrArgTyrGlnLeuArgGlyTyrIleGluAspSerGlnAspLeuGluIleTyrLeuIleArgTyr
AsnAlaLysHisGluThrValAsnValProGlyThrGlySerLeuTrpProLeuSerAlaGlnSerProIle
GlyLysCysGlyGluProAsnArgCysAlaProHisLeuGluTrpAsnProAspLeuAspCysSerCysArg
AspGlyGluLysCysAlaHisHisSerHisHisPheSerLeuAspIleAspValGlyCysThrAspLeuAsn
GluAspLeuGlyValTrpValIlePheLysIleLysThrGlnAspGlyHisAlaArgLeuGlyAsnLeuGlu
30 PheLeuGluGluLysProLeuValGlyGluAlaLeuAlaArgValLysArgAlaGluLysLysTrpArgAsp
LysArgGluLysLeuGluTrpGluThrAsnIleValTyrLysGluAlaLysGluSerValAspAlaLeuPhe
ValAsnSerGlnTyrAspGlnLeuGlnAlaAspThrAsnIleAlaMetIleHisAlaAlaAspLysArgVal
HisSerIleArgGluAlaTyrLeuProGluLeuSerValIleProGlyValAsnAlaAlaIlePheGluGlu
LeuGluGlyArgIlePheThrAlaPheSerLeuTyrAspAlaArgAsnValIleLysAsnGlyAspPheAsn
35 AsnGlyLeuSerCysTrpAsnValLysGlyHisValAspValGluGluGlnAsnAsnGlnArgSerValLeu
ValValProGluTrpGluAlaGluValSerGlnGluValArgValCysProGlyArgGlyTyrIleLeuArg
ValThrAlaTyrLysGluGlyTyrGlyGluGlyCysValThrIleHisGluIleGluAsnAsnThrAspGlu
LeuLysPheSerAsnCysValGluGluGluIleTyrProAsnAsnThrValThrCysAsnAspTyrThrVal



AsnGlnGluGluTyrGlyGlyAlaTyrThrSerArgAsnArgGlyTyrAsnGluAlaProSerValProAla
AspTyrAlaSerValTyrGluGluLysSerTyrThrAspGlyArgArgGluAsnProCysGluPheAsnArg
GlyTyrArgAspTyrThrProLeuProValGlyTyrValThrLysGluLeuGluTyrPheProGluThrAsp
LysValTrpIleGluIleGlyGluThrGluGlyThrPheIleValAspSerValGluLeuLeuLeuMetGlu
Glu

.5

6.6.5 EG 11751 晶体蛋白的氨基酸序列 (SEQ ID NO: 28)

MetAspAsnAsnProAsnIleAsnGluCysIleProTyrAsnCysLeuSerAsnProGluValGluValLeu
GlyGlyGluArgIleGluThrGlyTyrThrProIleAspIleSerLeuSerLeuThrGlnPheLeuLeuSer
GluPheValProGlyAlaGlyPheValLeuGlyLeuValAspIleIleTrpGlyIlePheGlyProSerGln
TrpAspAlaPheLeuValGlnIleGluGlnLeuIleAsnGlnArgIleGluGluPheAlaArgAsnGlnAla
IleSerArgLeuGluGlyLeuSerAsnLeuTyrGlnIleTyrAlaGluSerPheArgGluTrpGluAlaAsp
ProThrAsnProAlaLeuArgGluGluMetArgIleGlnPheAsnAspMetAsnSerAlaLeuThrThrAla
IleProLeuPheAlaValGlnAsnTyrGlnValProLeuLeuSerValTyrValGlnAlaAlaAsnLeuHis
LeuSerValLeuArgAspValSerValPheGlyGlnArgTrpGlyPheAspAlaAlaThrIleAsnSerArg
TyrAsnAspLeuThrArgLeuIleGlyAsnTyrThrAspTyrAlaValArgTrpTyrAsnThrGlyLeuGlu
ArgValTrpGlyProAspSerArgAspTrpValArgTyrAsnGlnPheArgArgGluLeuThrLeuThrVal
LeuAspIleValAlaLeuPheProAsnTyrAspSerArgArgTyrProIleArgThrValSerGlnLeuThr
ArgGluIleTyrThrAsnProValLeuGluAsnPheAspGlySerPheArgGlySerAlaGlnGlyIleGlu
ArgSerIleArgSerProHisLeuMetAspIleLeuAsnSerIleThrIleTyrThrAspAlaHisArgGly
TyrTyrTyrTrpSerGlyHisGlnIleMetAlaSerProValGlyPheSerGlyProGluPheThrPhePro
LeuTyrGlyThrMetGlyAsnAlaAlaProGlnGlnArgIleValAlaGlnLeuGlyGlnGlyValTyrArg
ThrLeuSerSerThrLeuTyrArgArgProPheAsnIleGlyIleAsnAsnGlnGlnLeuSerValLeuAsp
GlyThrGluPheAlaTyrGlyThrSerSerAsnLeuProSerAlaValTyrArgLysSerGlyThrValAsp
SerLeuAspGluIleProProGlnAsnAsnAsnValProProArgGlnGlyPheSerHisArgLeuSerHis
ValSerMetPheArgSerGlyPheSerAsnSerSerValSerIleIleArgAlaProMetPheSerTrpIle
HisArgSerAlaGluPheAsnAsnIleIleAlaSerAspSerIleThrGlnIleProLeuValLysAlaHis
ThrLeuGlnSerGlyThrThrValValArgGlyProGlyPheThrGlyGlyAspIleLeuArgArgThrSer
GlyGlyProPheAlaTyrThrIleValAsnIleAsnGlyGlnLeuProGlnArgTyrArgAlaArgIleArg
TyrAlaSerThrThrAsnLeuArgIleTyrValThrValAlaGlyGluArgIlePheAlaGlyGlnPheAsn
LysThrMetAspThrGlyAspProLeuThrPheGlnSerPheSerTyrAlaThrIleAsnThrAlaPheThr
PheProMetSerGlnSerSerPheThrValGlyAlaAspThrPheSerSerGlyAsnGluValTyrIleAsp
ArgPheGluLeuIleProValThrAlaThrPheGluAlaGluTyrAspLeuGluArgAlaGlnLysAlaVal
AsnAlaLeuPheThrSerIleAsnGlnIleGlyIleLysThrAspValThrAspTyrHisIleAspGlnVal
SerAsnLeuValAspCysLeuSerAspGluPheCysLeuAspGluLysArgGluLeuSerGluLysValLys
HisAlaLysArgLeuSerAspGluArgAsnLeuLeuGlnAspProAsnPheLysGlyIleAsnArgGlnLeu
AspArgGlyTrpArgGlySerThrAspIleThrIleGlnArgGlyAspAspValPheLysGluAsnTyrVal

10

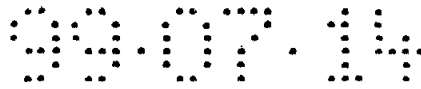
15

20

25

30

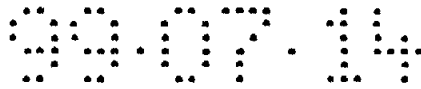
35



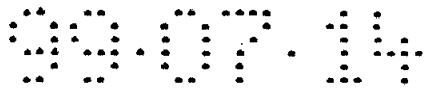
ThrLeuProGlyThrPheAspGluCysTyrProThrTyrLeuTyrGlnLysIleAspGluSerLysLeuLys
AlaPheThrArgTyrGlnLeuArgGlyTyrIleGluAspSerGlnAspLeuGluIleTyrLeuIleArgTyr
AsnAlaLysHisGluThrValAsnValProGlyThrGlySerLeuTrpProLeuSerAlaGlnSerProIle
GlyLysCysGlyGluProAsnArgCysAlaProHisLeuGluTrpAsnProAspLeuAspCysSerCysArg
5 AspGlyCysLysCysAlaHisHisSerHisHisPheSerLeuAspIleAspValGlyCysThrAspLeuAsn
GluAspLeuGlyValTrpValIlePheLysIleLysThrGlnAspGlyHisAlaArgLeuGlyAsnLeuGlu
PheLeuGluGluLysProLeuValGlyGluAlaLeuAlaArgValLysArgAlaGluLysLysTrpArgAsp
LysArgGluLysLeuGluTrpGluThrAsnIleValTyrLysGluAlaLysGluSerValAspAlaLeuPhe
ValAsnSerGlnTyrAspGlnLeuGlnAlaAspThrAsnIleAlaMetIleHisAlaAlaAspLysArgVal
10 HisSerIleArgGluAlaTyrLeuProGluLeuSerValIleProGlyValAsnAlaAlaIlePheGluGlu
LeuGluGlyArgIlePheThrAlaPheSerLeuTyrAspAlaArgAsnValIleLysAsnGlyAspPheAsn
AsnGlyLeuSerCysTrpAsnValLysGlyHisValAspValGluGluGlnAsnAsnGlnArgSerValLeu
ValValProGluTrpGluAlaGluValSerGlnGluValArgValCysProGlyArgGlyTyrIleLeuArg
ValThrAlaTyrLysGluGlyTyrGlyGluGlyCysValThrIleHisGluIleGluAsnAsnThrAspGlu
15 LeuLysPheSerAsnCysValGluGluGluIleTyrProAsnAsnThrValThrCysAsnAspTyrThrVal
AsnGlnGluGluTyrGlyGlyAlaTyrThrSerArgAsnArgGlyTyrAsnGluAlaProSerValProAla
AspTyrAlaSerValTyrGluGluLysSerTyrThrAspGlyArgArgGluAsnProCysGluPheAsnArg
GlyTyrArgAspTyrThrProLeuProValGlyTyrValThrLysGluLeuGluTyrPheProGluThrAsp
LysValTrpIleGluIleGlyGluThrGluGlyThrPheIleValAspSerValGluLeuLeuLeuMetGlu
20 Glu

6.6.6 EG 11091 晶体蛋白的氨基酸序列 (SEQ ID NO: 30)

MetAspAsnAsnProAsnIleAsnGluCysIleProTyrAsnCysLeuSerAsnProGluValGluValLeu
GlyGlyGluArgIleGluThrGlyTyrThrProIleAspIleSerLeuSerLeuThrGlnPheLeuLeuSer
25 GluPheValProGlyAlaGlyPheValLeuGlyLeuValAspIleIleTrpGlyIlePheGlyProSerGln
TrpAspAlaPheLeuValGlnIleGluGlnLeuIleAsnGlnArgIleGluGluPheAlaArgAsnGlnAla
IleSerArgLeuGluGlyLeuSerAsnLeuTyrGlnIleTyrAlaGluSerPheArgGluTrpGluAlaAsp
ProThrAsnProAlaLeuArgGluGluMetArgIleGlnPheAsnAspMetAsnSerAlaLeuThrThrAla
IleProLeuPheAlaValGlnAsnTyrGlnValProLeuLeuSerValTyrValGlnAlaAlaAsnLeuHis
30 LeuSerValLeuArgAspValSerValPheGlyGlnArgTrpGlyPheAspAlaAlaThrIleAsnSerArg
TyrAsnAspLeuThrArgLeuIleGlyAsnTyrThrAspTyrAlaValArgTrpTyrAsnThrGlyLeuGlu
ArgValTrpGlyProAspSerArgAspTrpValArgTyrAsnGlnPheArgArgGluLeuThrLeuThrVal
LeuAspIleValAlaLeuPheProAsnTyrAspSerArgArgTyrProIleArgThrValSerGlnLeuThr
ArgGluIleTyrThrAsnProValLeuGluAsnPheAspGlySerPheArgGlySerAlaGlnGlyIleGlu
35 ArgSerIleArgSerProHisLeuMetAspIleLeuAsnSerIleThrIleTyrThrAspAlaHisArgGly
TyrTyrTyrTrpSerGlyHisGlnIleMetAlaSerProValGlyPheSerGlyProGluPheThrPhePro
LeuTyrGlyThrMetGlyAsnAlaAlaProGlnGlnArgIleValAlaGlnLeuGlyGlnGlyValTyrArg

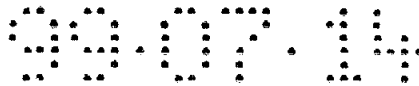


ThrLeuSerSerThrLeuTyrArgArgProPheAsnIleGlyIleAsnAsnGlnGlnLeuSerValLeuAsp
GlyThrGluPheAlaTyrGlyThrSerSerAsnLeuProSerAlaValTyrArgLysSerGlyThrValAsp
SerLeuAspGluIleProProGlnAsnAsnAsnValProProArgGlnGlyPheSerHisArgLeuSerHis
ValSerMetPheArgSerGlyPheSerAsnSerSerValSerIleIleArgAlaProMetPheSerTrpIle
5 HisArgSerAlaThrLeuThrAsnThrIleAspProGluArgIleAsnGlnIleProLeuValLysGlyPhe
ArgValTrpGlyGlyThrSerValIleThrGlyProGlyPheThrGlyGlyAspIleLeuArgArgAsnThr
PheGlyAspPheValSerLeuGlnValAsnIleAsnSerProIleThrGlnArgTyrArgLeuArgPheArg
TyrAlaSerSerArgAspAlaArgValIleValLeuThrGlyAlaAlaSerThrGlyValGlyGlyGlnVal
SerValAsnMetProLeuGlnLysThrMetGluIleGlyGluAsnLeuThrSerArgThrPheArgTyrThr
10 AspPheSerAsnProPheSerPheArgAlaAsnProAspIleIleGlyIleSerGluGlnProLeuPheGly
AlaGlySerIleSerSerGlyGluLeuTyrIleAspLysIleGluIleIleLeuAlaAspAlaThrPheGlu
AlaGluSerAspLeuGluArgAlaGlnLysAlaValAsnAlaLeuPheThrSerSerAsnGlnIleGlyLeu
LysThrAspValThrAspTyrHisIleAspGlnValSerAsnLeuValAspCysLeuSerAspGluPheCys
LeuAspGluLysArgGluLeuSerGluLysValLysHisAlaLysArgLeuSerAspGluArgAsnLeuLeu
15 GlnAspProAsnPheArgGlyIleAsnArgGlnProAspArgGlyTrpArgGlySerThrAspIleThrIle
GlnGlyGlyAspAspValPheLysGluAsnTyrValThrLeuProGlyThrValAspGluCysTyrProThr
TyrLeuTyrGlnLysIleAspGluSerLysLeuLysAlaTyrThrArgTyrGluLeuArgGlyTyrIleGlu
AspSerGlnAspLeuGluIleTyrLeuIleArgTyrAsnAlaLysHisGluIleValAsnValProGlyThr
GlySerLeuTrpProLeuSerAlaGlnSerProIleGlyLysCysGlyGluProAsnArgCysAlaProHis
20 LeuGluTrpAsnProAspLeuAspCysSerCysArgAspGlyGluLysCysAlaHisHisSerHisHisPhe
ThrLeuAspIleAspValGlyCysThrAspLeuAsnGluAspLeuGlyValTrpValIlePheLysIleLys
ThrGlnAspGlyHisAlaArgLeuGlyAsnLeuGluPheLeuGluGluLysProLeuLeuGlyGluAlaLeu
AlaArgValLysArgAlaGluLysLysTrpArgAspLysArgGluLysLeuGlnLeuGluThrAsnIleVal
TyrLysGluAlaLysGluSerValAspAlaLeuPheValAsnSerGlnTyrAspArgLeuGlnValAspThr
25 AsnIleAlaMetIleHisAlaAlaAspLysArgValHisArgIleArgGluAlaTyrLeuProGluLeuSer
ValIleProGlyValAsnAlaAlaIlePheGluGluLeuGluGlyArgIlePheThrAlaTyrSerLeuTyr
AspAlaArgAsnValIleLysAsnGlyAspPheAsnAsnGlyLeuLeuCysTrpAsnValLysGlyHisVal
AspValGluGluGlnAsnAsnHisArgSerValLeuValIleProGluTrpGluAlaGluValSerGlnGlu
ValArgValCysProGlyArgGlyTyrIleLeuArgValThrAlaTyrLysGluGlyTyrGlyGluGlyCys
30 ValThrIleHisGluIleGluAspAsnThrAspGluLeuLysPheSerAsnCysValGluGluGluValTyr
ProAsnAsnThrValThrCysAsnAsnTyrThrGlyThrGlnGluGluTyrGluGlyThrTyrThrSerArg
AsnGlnGlyTyrAspGluAlaTyrGlyAsnAsnProSerValProAlaAspTyrAlaSerValTyrGluGlu
LysSerTyrThrAspGlyArgArgGluAsnProCysGluSerAsnArgGlyTyrGlyAspTyrThrProLeu
ProAlaGlyTyrValThrLysAspLeuGluTyrPheProGluThrAspLysValTrpIleGluIleGlyGlu
35 ThrGluGlyThrPheIleValAspSerValGluLeuLeuLeuMetGluGlu



6.6.7 EG 11768 晶体蛋白的氨基酸序列 (SEQ ID NO: 34)

MetAspAsnAsnProAsnIleAsnGluCysIleProTyrAsnCysLeuSerAsnProGluValGluValLeuGlyGlyGluArgIleGluThrGlyTyrThrProIleAspIleSerLeuSerLeuThrGlnPheLeuLeuSerGluPheValProGlyAlaGlyPheValLeuGlyLeuValAspIleIleTrpGlyIlePheGlyProSerGlnTrpAspAlaPheLeuValGlnIleGluGlnLeuIleAsnGlnArgIleGluGluPheAlaArgAsnGlnAlaIleSerArgLeuGluGlyLeuSerAsnLeuTyrGlnIleTyrAlaGluSerPheArgGluTrpGluAlaAspProThrAsnProAlaLeuArgGluGluMetArgIleGlnPheAsnAspMetAsnSerAlaLeuThrThrAlaIleProLeuPheAlaValGlnAsnTyrGlnValProLeuLeuSerValTyrValGlnAlaAlaAsnLeuHisLeuSerValLeuArgAspValSerValPheGlyGlnArgTrpGlyPheAspAlaAlaThrIleAsnSerArgTyrAsnAspLeuThrArgLeuIleGlyAsnTyrThrAspTyrAlaValArgTrpTyrAsnThrGlyLeuGluArgValTrpGlyProAspSerArgAspTrpValArgTyrAsnGlnPheArgArgGluLeuThrLeuThrValLeuAspIleValAlaLeuPheProAsnTyrAspSerArgArgTyrProIleArgThrValSerGlnLeuThrArgGluIleTyrThrAsnProValLeuGluAsnPheAspGlySerPheArgGlySerAlaGlnGlyIleGluArgSerIleArgSerProHisLeuMetAspIleLeuAsnSerIleThrIleTyrThrAspAlaHisArgGlyTyrTyrTyrTrpSerGlyHisGlnIleMetAlaSerProValGlyPheSerGlyProGluPheThrPheProLeuTyrGlyThrMetGlyAsnAlaAlaProGlnGlnArgIleValAlaGlnLeuGlyGlnGlyValTyrArgThrLeuSerSerThrLeuTyrArgArgPropheAsnIleGlyIleAsnAsnGlnGlnLeuSerValLeuAspGlyThrGluPheAlaTyrGlyThrSerSerAsnLeuProSerAlaValTyrArgLysSerGlyThrValAspSerLeuAspGluIleProProGlnAsnAsnAsnValProProArgGlnGlyPheSerHisArgLeuSerHisValSerMetPheArgSerGlyPheSerAsnSerSerValSerIleIleArgAlaProMetPheSerTrpIleHisArgSerAlaGluPheAsnAsnIleIleAlaSerAspSerIleThrGlnIleProLeuValLysAlaHisThrLeuGlnSerGlyThrThrValValArgGlyProGlyPheThrGlyGlyAspIleLeuArgArgThrSerGlyGlyProPheAlaTyrThrIleValAsnIleAsnGlyGlnLeuProGlnArgTyrArgAlaArgIleArgTyrAlaSerThrThrAsnLeuArgIleTyrValThrValAlaGlyGluArgIlePheAlaGlyGlnPheAsnLysThrMetAspThrGlyAspProLeuThrPheGlnSerPheSerTyrAlaThrIleAsnThrAlaPheThrPheProMetSerGlnSerSerPheThrValGlyAlaAspThrPheSerSerGlyAsnGluValTyrIleAspArgPheGluLeuIleProValThrAlaThrLeuGluAlaGluTyrAsnLeuGluArgAlaGlnLysAlaValAsnAlaLeuPheThrSerThrAsnGlnLeuGlyLeuLysThrAsnValThrAspTyrHisIleAspGlnValSerAsnLeuValThrTyrLeuSerAspGluPheCysLeuAspGluLysArgGluLeuSerGluLysValLysHisAlaLysArgLeuSerAspGluArgAsnLeuLeuGlnAspSerAsnPheLysAspIleAsnArgGlnProGluArgGlyTrpGlyGlySerThrGlyIleThrIleGlnGlyGlyAspAspValPheLysGluAsnTyrValThrLeuSerGlyThrPheAspGluCysTyrProThrTyrLeuTyrGlnLysIleAspGluSerLysLeuLysAlaPheThrArgTyrGlnLeuArgGlyTyrIleGluAspSerGlnAspLeuGluIleTyrLeuIleArgTyrAsnAlaLysHisGluThrValAsnValProGlyThrGlySerLeuTrpProLeuSerAlaGlnSerProIleGlyLysCysGlyGluProAsnArgCysAlaProHisLeuGluTrpAsnProAspLeuAspCysSerCysArgAspGlyGluLysCysAlaHisHisSerHisHisPheSerLeuAspIleAspValGlyCysThrAspLeuAsnGluAspLeuGlyValTrpValIlePheLysIleLysThrGlnAspGlyHisAlaArgLeuGlyAsnLeuGluPheLeuGluGluLysProLeuValGlyGluAlaLeuAlaArgValLysArgAlaGluLysLysTrpArgAspLysArgGluLysLeuGluTrpGluThrAsnIleValTyrLysGluAlaLysGluSerValAspAlaLeuPheValAsnSerGlnTyrAspGlnLeuGlnAlaAspThrAsnIleAlaMetIleHisAlaAlaAspLysArgValHisS

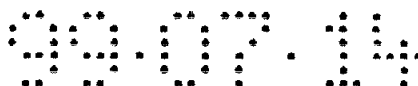


erIleArgGluAlaTyrLeuProGluLeuSerValIleProGlyValAsnAlaAlaIlePheGluGluLeuGluGly
yArgIlePheThrAlaPheSerLeuTyrAspAlaArgAsnValIleLysAsnGlyAspPheAsnAsnGlyLeuSer
CysTrpAsnValLysGlyHisValAspValGluGluGlnAsnAsnGlnArgSerValLeuValValProGluTrpG
luAlaGluValSerGlnGluValArgValCysProGlyArgGlyTyrIleLeuArgValThrAlaTyrLysGluGly
5 yTyrGlyGluGlyCysValThrIleHisGluIleGluAsnAsnThrAspGluLeuLysPheSerAsnCysValGlu
GluGluIleTyrProAsnAsnThrValThrCysAsnAspTyrThrValAsnGlnGluGluTyrGlyGlyAlaTyrT
hrSerArgAsnArgGlyTyrAsnGluAlaProSerValProAlaAspTyrAlaSerValTyrGluGluLysSerTy
rThrAspGlyArgArgGluAsnProCysGluPheAsnArgGlyTyrArgAspTyrThrProLeuProValGlyTyr
ValThrLysGluLeuGluTyrPheProGluThrAspLysValTrpIleGluIleGlyGluThrGluGlyThrPheI
10 leValAspSerValGluLeuLeuLeuMetGluGlu

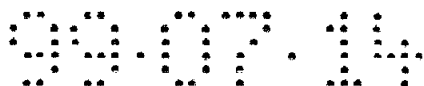
6.7 实施例 7 -- 编码新的晶体蛋白的 DNA 序列

6.7.1 编码 EG 11063 晶体蛋白的 DNA 序列 (SEQ ID NO: 9)

| | | |
|----|---|------|
| 15 | ATG GAT AAC AAT CCG AAC ATC AAT GAA TGC ATT CCT TAT AAT TGT TTA | 48 |
| | AGT AAC CCT GAA GTA GAA GTA TTA GGT GGA GAA AGA ATA GAA ACT GGT | 96 |
| | TAC ACC CCA ATC GAT ATT TCC TTG TCG CTA ACG CAA TTT CTT TTG AGT | 144 |
| | GAA TTT GTT CCC GGT GCT GGA TTT GTG TTA GGA CTA GTT GAT ATA ATA | 192 |
| | TGG GGA ATT TTT GGT CCC TCT CAA TGG GAC GCA TTT CTT GTA CAA ATT | 240 |
| | GAA CAG TTA ATT AAC CAA AGA ATA GAA GAA TTC GCT AGG AAC CAA GCC | 288 |
| 20 | ATT TCT AGA TTA GAA GGA CTA AGC AAT CTT TAT CAA ATT TAC GCA GAA | 336 |
| | TCT TTT AGA GAG TGG GAA GCA GAT CCT ACT AAT CCA GCA TTA AGA GAA | 384 |
| | GAG ATG CGT ATT CAA TTC AAT GAC ATG AAC AGT GCC CTT ACA ACC GCT | 432 |
| | ATT CCT CTT TTT GCA GTT CAA AAT TAT CAA GTT CCT CTT TTA TCA GTA | 480 |
| | TAT GTT CAA GCT GCA AAT TTA CAT TTA TCA GTT TTG AGA GAT GTT TCA | 528 |
| 25 | GTG TTT GGA CAA AGG TGG GGA TTT GAT GCC GCG ACT ATC AAT AGT CGT | 576 |
| | TAT AAT GAT TTA ACT AGG CTT ATT GGC AAC TAT ACA GAT TAT GCT GTA | 624 |
| | CGC TGG TAC AAT ACG GGA TTA GAA CGT GTA TGG GGA CCG GAT TCT AGA | 672 |
| | GAT TGG GTA AGG TAT AAT CAA TTT AGA AGA GAA TTA ACA CTA ACT GTA | 720 |
| | TTA GAT ATC GTT GCT CTG TTC CCG AAT TAT GAT AGT AGA AGA TAT CCA | 768 |
| 30 | ATT CGA ACA GTT TCC CAA TTA ACA AGA GAA ATT TAT ACA AAC CCA GTA | 816 |
| | TTA GAA AAT TTT GAT GGT AGT TTT CGA GGC TCG GCT CAG GGC ATA GAA | 864 |
| | AGA AGT ATT AGG AGT CCA CAT TTG ATG GAT ATA CTT AAC AGT ATA ACC | 912 |
| | ATC TAT ACG GAT GCT CAT AGG GGT TAT TAT TAT TGG TCA GGG CAT CAA | 960 |
| | ATA ATG GCT TCT CCT GTA GGG TTT TCG GGG CCA GAA TTC ACT TTT CCG | 1008 |
| 35 | CTA TAT GGA ACT ATG GGA AAT GCA GCT CCA CAA CAA CGT ATT GTT GCT | 1056 |
| | CAA CTA GGT CAG GGC GTG TAT AGA ACA TTA TCG TCC ACT TTA TAT AGA | 1104 |
| | AGA CCT TTT AAT ATA GGG ATA AAT AAT CAA CAA CTA TCT GTT CTT GAC | 1152 |



| | | | | | | | | | | | | | | | | | |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| | GGG | ACA | GAA | TTT | GCT | TAT | GGA | ACC | TCC | TCA | AAT | TTG | CCA | TCC | GCT | GTA | 1200 |
| | TAC | AGA | AAA | AGC | GGA | ACG | GTA | GAT | TCG | CTG | GAT | GAA | ATA | CCG | CCA | CAG | 1248 |
| | AAT | AAC | AAC | GTG | CCA | CCT | AGG | CAA | GGA | TTT | AGT | CAT | CGA | TTA | AGC | CAT | 1296 |
| | GTT | TCA | ATG | TTT | CGT | TCA | GGC | TTT | AGT | AAT | AGT | AGT | GTA | AGT | ATA | ATA | 1344 |
| 5 | AGA | GCT | CCA | ATG | TTT | TCT | TGG | ACG | CAC | CGT | AGT | GCA | ACC | CCT | ACA | AAT | 1392 |
| | ACA | ATT | GAT | CCG | GAG | AGG | ATT | ACT | CAA | ATA | CCA | TTG | GTA | AAA | GCA | CAT | 1440 |
| | ACA | CTT | CAG | TCA | GGT | ACT | ACT | GTT | GTA | AGA | GGG | CCC | GGG | TTT | ACG | GGA | 1488 |
| | GGA | GAT | ATT | CTT | CGA | CGA | ACA | AGT | GGA | GGA | CCA | TTT | GCT | TAT | ACT | ATT | 1536 |
| | GTT | AAT | ATA | AAT | GGG | CAA | TTA | CCC | CAA | AGG | TAT | CGT | GCA | AGA | ATA | CGC | 1584 |
| 10 | TAT | GCC | TCT | ACT | ACA | AAT | CTA | AGA | ATT | TAC | GTA | ACG | GTT | GCA | GGT | GAA | 1632 |
| | CGG | ATT | TTT | GCT | GGT | CAA | TTT | AAC | AAA | ACA | ATG | GAT | ACC | GGT | GAC | CCA | 1680 |
| | TTA | ACA | TTC | CAA | TCT | TTT | AGT | TAC | GCA | ACT | ATT | AAT | ACA | GCT | TTT | ACA | 1728 |
| | TTC | CCA | ATG | AGC | CAG | AGT | AGT | TTC | ACA | GTA | GGT | GCT | GAT | ACT | TTT | AGT | 1776 |
| | TCA | GGG | AAT | GAA | GTT | TAT | ATA | GAC | AGA | TTT | GAA | TTG | ATT | CCA | GTT | ACT | 1824 |
| 15 | GCA | ACA | TTT | GAA | GCA | GAA | TAT | GAT | TTA | GAA | AGA | GCA | CAA | AAG | GCG | GTG | 1872 |
| | AAT | GCG | CTG | TTT | ACT | TCT | ATA | AAC | CAA | ATA | GGG | ATA | AAA | ACA | GAT | GTG | 1920 |
| | ACG | GAT | TAT | CAT | ATT | GAT | CAA | GTA | TCC | AAT | TTA | GTG | GAT | TGT | TTA | TCA | 1968 |
| | GAT | GAA | TTT | TGT | CTG | GAT | GAA | AAG | CGA | GAA | TTG | TCC | GAG | AAA | GTC | AAA | 2016 |
| | CAT | GCG | AAG | CGA | CTC | AGT | GAT | GAG | CGG | AAT | TTA | CTT | CAA | GAT | CCA | AAC | 2064 |
| 20 | TTC | AAA | GGC | ATC | AAT | AGG | CAA | CTA | GAC | CGT | GGT | TGG | AGA | GGA | AGT | ACG | 2112 |
| | GAT | ATT | ACC | ATC | CAA | AGA | GGA | GAT | GAC | GTA | TTC | AAA | GAA | AAT | TAT | GTC | 2160 |
| | ACA | CTA | CCA | GGT | ACC | TTT | GAT | GAG | TGC | TAT | CCA | ACA | TAT | TTG | TAT | CAA | 2208 |
| | AAA | ATC | GAT | GAA | TCA | AAA | TTA | AAA | GCC | TTT | ACC | CGT | TAT | CAA | TTA | AGA | 2256 |
| | GGG | TAT | ATC | GAA | GAT | AGT | CAA | GAC | TTA | GAA | ATC | TAT | TTA | ATT | CGC | TAC | 2304 |
| 25 | AAT | GCA | AAA | CAT | GAA | ACA | GTA | AAT | GTG | CCA | GGT | ACG | GGT | TCC | TTA | TGG | 2352 |
| | CCG | CTT | TCA | GCC | CAA | AGT | CCA | ATC | GGA | AAG | TGT | GGA | GAG | CCG | AAT | CGA | 2400 |
| | TGC | GCG | CCA | CAC | CTT | GAA | TGG | AAT | CCT | GAC | TTA | GAT | TGT | TCG | TGT | AGG | 2448 |
| | GAT | GGA | GAA | AAG | TGT | GCC | CAT | CAT | TCG | CAT | CAT | TTC | TCC | TTA | GAC | ATT | 2496 |
| | GAT | GTA | GGA | TGT | ACA | GAC | TTA | AAT | GAG | GAC | CTA | GGT | GTA | TGG | GTG | ATC | 2544 |
| 30 | TTT | AAG | ATT | AAG | ACG | CAA | GAT | GGG | CAC | GCA | AGA | CTA | GGG | AAT | CTA | GAG | 2592 |
| | TTT | CTC | GAA | GAG | AAA | CCA | TTA | GTA | GGA | GAA | GCG | CTA | GCT | CGT | GTG | AAA | 2640 |
| | AGA | GCG | GAG | AAA | AAA | TGG | AGA | GAC | AAA | CGT | GAA | AAA | TTG | GAA | TGG | GAA | 2688 |
| | ACA | AAT | ATC | GTT | TAT | AAA | GAG | GCA | AAA | GAA | TCT | GTA | GAT | GCT | TTA | TTT | 2736 |
| | GTA | AAC | TCT | CAA | TAT | GAT | CAA | TTA | CAA | GCG | GAT | ACG | AAT | ATT | GCC | ATG | 2784 |
| 35 | ATT | CAT | GCG | GCA | GAT | AAA | CGT | GTT | CAT | AGC | ATT | CGA | GAA | GCT | TAT | CTG | 2832 |
| | CCT | GAG | CTG | TCT | GTG | ATT | CCG | GGT | GTC | AAT | GCG | GCT | ATT | TTT | GAA | GAA | 2880 |
| | TTA | GAA | GGG | CGT | ATT | TTC | ACT | GCA | TTC | TCC | CTA | TAT | GAT | GCG | AGA | AAT | 2928 |
| | GTC | ATT | AAA | AAT | GGT | GAT | TTT | AAT | AAT | GGC | TTA | TCC | TGC | TGG | AAC | GTG | 2976 |



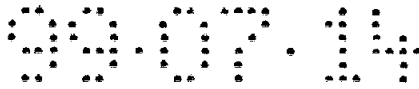
AAA GGG CAT GTA GAT GTA GAA GAA CAA AAC AAC CAA CGT TCG GTC CTT 3024
GTT GTT CCG GAA TGG GAA GCA GAA GTG TCA CAA GAA GTT CGT GTC TGT 3072
CCG GGT CGT GGC TAT ATC CTT CGT GTC ACA GCG TAC AAG GAG GGA TAT 3120
GGA GAA GGT TGC GTA ACC ATT CAT GAG ATC GAG AAC AAT ACA GAC GAA 3168
5 CTG AAG TTT AGC AAC TGC GTA GAA GAG GAA ATC TAT CCA AAT AAC ACG 3216
GTA ACG TGT AAT GAT TAT ACT GTA AAT CAA GAA GAA TAC GGA GGT GCG 3264
TAC ACT TCT CGT AAT CGA GGA TAT AAC GAA GCT CCT TCC GTA CCA GCT 3312
GAT TAT GCG TCA GTC TAT GAA GAA AAA TCG TAT ACA GAT GGA CGA AGA 3360
GAG AAT CCT TGT GAA TTT AAC AGA GGG TAT AGG GAT TAC ACG CCA CTA 3408
10 CCA GTT GGT TAT GTG ACA AAA GAA TTA GAA TAC TTC CCA GAA ACC GAT 3456
AAG GTA TGG ATT GAG ATT GGA GAA ACG GAA GGA ACA TTT ATC GTG GAC 3504
AGC GTG GAA TTA CTC CTT ATG GAG GAA 3531

6.7.2 编码 EG 11074 晶体蛋白的 DNA 序列 (SEQ ID NO: 11)

15 ATG GAT AAC AAT CCG AAC ATC AAT GAA TGC ATT CCT TAT AAT TGT TTA 48
AGT AAC CCT GAA GTA GAA GTA TTA GGT GGA GAA AGA ATA GAA ACT GGT 96
TAC ACC CCA ATC GAT ATT TCC TTG TCG CTA ACG CAA TTT CTT TTG AGT 144
GAA TTT GTT CCC GGT GCT GGA TTT GTG TTA GGA CTA GTT GAT ATA ATA 192
TGG GGA ATT TTT GGT CCC TCT CAA TGG GAC GCA TTT CTT GTA CAA ATT 240
20 GAA CAG TTA ATT AAC CAA AGA ATA GAA GAA TTC GCT AGG AAC CAA GCC 288
ATT TCT AGA TTA GAA GGA CTA AGC AAT CTT TAT CAA ATT TAC GCA GAA 336
TCT TTT AGA GAG TGG GAA GCA GAT CCT ACT AAT CCA GCA TTA AGA GAA 384
GAG ATG CGT ATT CAA TTC AAT GAC ATG AAC AGT GCC CTT ACA ACC GCT 432
ATT CCT CTT TTT GCA GTT CAA AAT TAT CAA GTT CCT CTT TTA TCA GTA 480
25 TAT GTT CAA GCT GCA AAT TTA CAT TTA TCA GTT TTG AGA GAT GTT TCA 528
GTG TTT GGA CAA AGG TGG GGA TTT GAT GCC GCG ACT ATC AAT AGT CGT 576
TAT AAT GAT TTA ACT AGG CTT ATT GGC AAC TAT ACA GAT TAT GCT GTA 624
CGC TGG TAC AAT ACG GGA TTA GAA CGT GTA TGG GGA CCG GAT TCT AGA 672
GAT TGG GTA AGG TAT AAT CAA TTT AGA AGA GAA TTA ACA CTA ACT GTA 720
30 TTA GAT ATC GTT GCT CTG TTC CCG AAT TAT GAT AGT AGA AGA TAT CCA 768
ATT CGA ACA GTT TCC CAA TTA ACA AGA GAA ATT TAT ACA AAC CCA GTA 816
TTA GAA AAT TTT GAT GGT AGT TTT CGA GGC TCG GCT CAG GGC ATA GAA 864
AGA AGT ATT AGG AGT CCA CAT TTG ATG GAT ATA CTT AAC AGT ATA ACC 912
ATC TAT ACG GAT GCT CAT AGG GGT TAT TAT TAT TGG TCA GGG CAT CAA 960
35 ATA ATG GCT TCT CCT GTA GGG TTT TCG GGG CCA GAA TTC ACT TTT CCG 1008
CTA TAT GGA ACT ATG GGA AAT GCA GCT CCA CAA CAA CGT ATT GTT GCT 1056
CAA CTA GGT CAG GGC GTG TAT AGA ACA TTA TCG TCC ACT TTA TAT AGA 1104



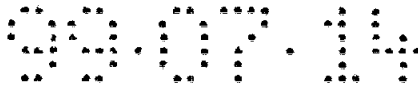
| | | |
|----|---|------|
| | AGA CCT TTT AAT ATA GGG ATA AAT AAT CAA CAA CTA TCT GTT CTT GAC | 1152 |
| | GGG ACA GAA TTT GCT TAT GGA ACC TCC TCA AAT TTG CCA TCC GCT GTA | 1200 |
| | TAC AGA AAA AGC GGA ACG GTA GAT TCG CTG GAT GAA ATA CCG CCA CAG | 1248 |
| | AAT AAC AAC GTG CCA CCT AGG CAA GGA TTT AGT CAT CGA TTA AGC CAT | 1296 |
| 5 | GTT TCA ATG TTT CGT TCA GGC TTT AGT AAT AGT AGT GTA AGT ATA ATA | 1344 |
| | AGA GCT CCA ATG TTT TCT TGG ACG CAC CGT AGT GCA ACC CCT ACA AAT | 1392 |
| | ACA ATT GAT CCG GAG AGG ATT ACT CAA ATA CCA TTG GTA AAA GCA CAT | 1440 |
| | ACA CTT CAG TCA GGT ACT ACT GTT GTA AGA GGG CCC GGG TTT ACG GGA | 1488 |
| | GGG GAT ATT CTT CGA CGA ACA AGT GGA GGA CCA TTT GCT TAT ACT ATT | 1536 |
| 10 | GTT AAT ATA AAT GGG CAA TTA CCC CAA AGG TAT CGT GCA AGA ATA CGC | 1584 |
| | TAT GCC TCT ACT ACA AAT CTA AGA ATT TAC GTA ACG GTT GCA GGT GAA | 1632 |
| | CGG ATT TTT GCT GGT CAA TTT AAC AAA ACA ATG GAT ACC GGT GAC CCA | 1680 |
| | TTA ACA TTC CAA TCT TTT AGT TAC GCA ACT ATT AAT ACA GCT TTT ACA | 1728 |
| | TTC CCA ATG AGC CAG AGT AGT TTC ACA GTA GGT GCT GAT ACT TTT AGT | 1776 |
| 15 | TCA GGG AAT GAA GTT TAT ATA GAC AGA TTT GAA TTG ATT CCA GTT ACT | 1824 |
| | GCA ACA CTC GAG GCT GAA TAT AAT CTG GAA AGA GCG CAG AAG GCG GTG | 1872 |
| | AAT GCG CTG TTT ACG TCT ACA AAC CAA CTA GGG CTA AAA ACA AAT GTA | 1920 |
| | ACG GAT TAT CAT ATT GAT CAA GTG TCC AAT TTA GTT ACG TAT TTA TCG | 1968 |
| | GAT GAA TTT TGT CTG GAT GAA AAG CGA GAA TTG TCC GAG AAA GTC AAA | 2016 |
| 20 | CAT GCG AAG CGA CTC AGT GAT GAA CGC AAT TTA CTC CAA GAT TCA AAT | 2064 |
| | TTC AAA GAC ATT AAT AGG CAA CCA GAA CGT GGG TGG GGC GGA AGT ACA | 2112 |
| | GGG ATT ACC ATC CAA GGA GGG GAT GAC GTA TTT AAA GAA AAT TAC GTC | 2160 |
| | ACA CTA TCA GGT ACC TTT GAT GAG TGC TAT CCA ACA TAT TTG TAT CAA | 2208 |
| | AAA ATC GAT GAA TCA AAA TTA AAA GCC TTT ACC CGT TAT CAA TTA AGA | 2256 |
| 25 | GGG TAT ATC GAA GAT AGT CAA GAC TTA GAA ATC TAT TTA ATT CGC TAC | 2304 |
| | AAT GCA AAA CAT GAA ACA GTA AAT GTG CCA GGT ACG GGT TCC TTA TGG | 2352 |
| | CCG CTT TCA GCC CAA AGT CCA ATC GGA AAG TGT GGA GAG CCG AAT CGA | 2400 |
| | TGC GCG CCA CAC CTT GAA TGG AAT CCT GAC TTA GAT TGT TCG TGT AGG | 2448 |
| | GAT GGA GAA AAG TGT GCC CAT CAT TCG CAT CAT TTC TCC TTA GAC ATT | 2496 |
| 30 | GAT GTA GGA TGT ACA GAC TTA AAT GAG GAC CTA GGT GTA TGG GTG ATC | 2544 |
| | TTT AAG ATT AAG ACG CAA GAT GGG CAC GCA AGA CTA GGG AAT CTA GAG | 2592 |
| | TTT CTC GAA GAG AAA CCA TTA GTA GGA GAA GCG CTA GCT CGT GTG AAA | 2640 |
| | AGA GCG GAG AAA AAA TGG AGA GAC AAA CGT GAA AAA TTG GAA TGG GAA | 2688 |
| | ACA AAT ATC GTT TAT AAA GAG GCA AAA GAA TCT GTA GAT GCT TTA TTT | 2736 |
| 35 | GTA AAC TCT CAA TAT GAT CAA TTA CAA GCG GAT ACG AAT ATT GCC ATG | 2784 |
| | ATT CAT GCG GCA GAT AAA CGT GTT CAT AGC ATT CGA GAA GCT TAT CTG | 2832 |
| | CCT GAG CTG TCT GTG ATT CCG GGT GTC AAT GCG GCT ATT TTT GAA GAA | 2880 |
| | TTA GAA GGG CGT ATT TTC ACT GCA TTC TCC CTA TAT GAT GCG AGA AAT | 2928 |



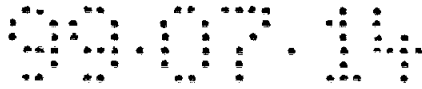
GTC ATT AAA AAT GGT GAT TTT AAT AAT GGC TTA TCC TGC TGG AAC GTG 2976
AAA GGG CAT GTA GAT GTA GAA GAA CAA AAC AAC CAA CGT TCG GTC CTT 3024
GTT GTT CCG GAA TGG GAA GCA GAA GTG TCA CAA GAA GTT CGT GTC TGT 3072
CCG GGT CGT GGC TAT ATC CTT CGT GTC ACA GCG TAC AAG GAG GGA TAT 3120
5 GGA GAA GGT TGC GTA ACC ATT CAT GAG ATC GAG AAC AAT ACA GAC GAA 3168
CTG AAG TTT AGC AAC TGC GTA GAA GAG GAA ATC TAT CCA AAT AAC ACG 3216
GTA ACG TGT AAT GAT TAT ACT GTA AAT CAA GAA GAA TAC GGA GGT GCG 3264
TAC ACT TCT CGT AAT CGA GGA TAT AAC GAA GCT CCT TCC GTA CCA GCT 3312
GAT TAT GCG TCA GTC TAT GAA GAA AAA TCG TAT ACA GAT GGA CGA AGA 3360
10 GAG AAT CCT TGT GAA TTT AAC AGA GGG TAT AGG GAT TAC ACG CCA CTA 3408
CCA GTT GGT TAT GTG ACA AAA GAA TTA GAA TAC TTC CCA GAA ACC GAT 3456
AAG GTA TGG ATT GAG ATT GGA GAA ACG GAA GGA ACA TTT ATC GTG GAC 3504
AGC GTG GAA TTA CTC CTT ATG GAG GAA 3531

15 **6.7.3 编码 EG 11735 晶体蛋白的 DNA 序列 (SEQ ID NO: 13)**

ATG GAT AAC AAT CCG AAC ATC AAT GAA TGC ATT CCT TAT AAT TGT TTA 48
AGT AAC CCT GAA GTA GAA GTA TTA GGT GGA GAA AGA ATA GAA ACT GGT 96
TAC ACC CCA ATC GAT ATT TCC TTG TCG CTA ACG CAA TTT CTT TTG AGT 144
GAA TTT GTT CCC GGT GCT GGA TTT GTG TTA GGA CTA GTT GAT ATA ATA 192
20 TGG GGA ATT TTT GGT CCC TCT CAA TGG GAC GCA TTT CTT GTA CAA ATT 240
GAA CAG TTA ATT AAC CAA AGA ATA GAA GAA TTC GCT AGG AAC CAA GCC 288
ATT TCT AGA TTA GAA GGA CTA AGC AAT CTT TAT CAA ATT TAC GCA GAA 336
TCT TTT AGA GAG TGG GAA GCA GAT CCT ACT AAT CCA GCA TTA AGA GAA 384
GAG ATG CGT ATT CAA TTC AAT GAC ATG AAC AGT GCC CTT ACA ACC GCT 432
25 ATT CCT CTT TTT GCA GTT CAA AAT TAT CAA GTT CCT CTT TTA TCA GTA 480
TAT GTT CAA GCT GCA AAT TTA CAT TTA TCA GTT TTG AGA GAT GTT TCA 528
GTG TTT GGA CAA AGG TGG GGA TTT GAT GCC GCG ACT ATC AAT AGT CGT 576
TAT AAT GAT TTA ACT AGG CTT ATT GGC AAC TAT ACA GAT CAA GCT GTA 624
CGC TGG TAC AAT ACG GGA TTA GAG CGT GTA TGG GGA CCG GAT TCT AGA 672
30 GAT TGG ATA AGA TAT AAT CAA TTT AGA AGA GAA TTA ACA CTA ACT GTA 720
TTA GAT ATC GTT TCT CTA TTT CCG AAC TAT GAT AGT AGA ACG TAT CCA 768
ATT CGA ACA GTT TCC CAA TTA ACA AGA GAA ATT TAT ACA AAC CCA GTA 816
TTA GAA AAT TTT GAT GGT AGT TTT CGA GGC TCG GCT CAG GGC ATA GAA 864
GGA AGT ATT AGG AGT CCA CAT TTG ATG GAT ATA CTT AAC AGT ATA ACC 912
35 ATC TAT ACG GAT GCT CAT AGA GGA GAA TAT TAT TGG TCA GGG CAT CAA 960
ATA ATG GCT TCT CCT GTA GGG TTT TCG GGG CCA GAA TTC ACT TTT CCG 1008
CTA TAT GGA ACT ATG GGA AAT GCA GCT CCA CAA CAA CGT ATT GTT GCT 1056
CAA CTA GGT CAG GGC GTG TAT AGA ACA TTA TCG TCC ACT TTA TAT AGA 1104



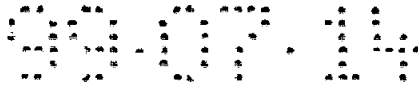
| | | |
|----|---|------|
| | AGA CCT TTT AAT ATA GGG ATA AAT AAT CAA CAA CTA TCT GTT CTT GAC | 1152 |
| | GGG ACA GAA TTT GCT TAT GGA ACC TCC TCA AAT TTG CCA TCC GCT GTA | 1200 |
| | TAC AGA AAA AGC GGA ACG GTA GAT TCG CTG GAT GAA ATA CCG CCA CAG | 1248 |
| | AAT AAC AAC GTG CCA CCT AGG CAA GGA TTT AGT CAT CGA TTA AGC CAT | 1296 |
| 5 | GTT TCA ATG TTT CGT TCA GGC TTT AGT AAT AGT AGT GTA AGT ATA ATA | 1344 |
| | AGA GCT CCA ATG TTT TCT TGG ACG CAC CGT AGT GCA ACC CCT ACA AAT | 1392 |
| | ACA ATT GAT CCG GAG AGG ATT ACT CAA ATA CCA TTG GTA AAA GCA CAT | 1440 |
| | ACA CTT CAG TCA GGT ACT ACT GTT GTA AGA GGG CCC GGG TTT ACG GGA | 1488 |
| | GGA GAT ATT CTT CGA CGA ACA AGT GGA GGA CCA TTT GCT TAT ACT ATT | 1536 |
| 10 | GTT AAT ATA AAT GGG CAA TTA CCC CAA AGG TAT CGT GCA AGA ATA CGC | 1584 |
| | TAT GCC TCT ACT ACA AAT CTA AGA ATT TAC GTA ACG GTT GCA GGT GAA | 1632 |
| | CGG ATT TTT GCT GGT CAA TTT AAC AAA ACA ATG GAT ACC GGT GAC CCA | 1680 |
| | TTA ACA TTC CAA TCT TTT AGT TAC GCA ACT ATT AAT ACA GCT TTT ACA | 1728 |
| | TTC CCA ATG AGC CAG AGT AGT TTC ACA GTA GGT GCT GAT ACT TTT AGT | 1776 |
| 15 | TCA GGG AAT GAA GTT TAT ATA GAC AGA TTT GAA TTG ATT CCA GTT ACT | 1824 |
| | GCA ACA TTT GAA GCA GAA TAT GAT TTA GAA AGA GCA CAA AAG GCG GTG | 1872 |
| | AAT GCG CTG TTT ACT TCT ATA AAC CAA ATA GGG ATA AAA ACA GAT GTG | 1920 |
| | ACG GAT TAT CAT ATT GAT CAA GTA TCC AAT TTA GTG GAT TGT TTA TCA | 1968 |
| | GAT GAA TTT TGT CTG GAT GAA AAG CGA GAA TTG TCC GAG AAA GTC AAA | 2016 |
| 20 | CAT GCG AAG CGA CTC AGT GAT GAG CGG AAT TTA CTT CAA GAT CCA AAC | 2064 |
| | TTC AAA GGC ATC AAT AGG CAA CTA GAC CGT GGT TGG AGA GGA AGT ACG | 2112 |
| | GAT ATT ACC ATC CAA AGA GGA GAT GAC GTA TTC AAA GAA AAT TAT GTC | 2160 |
| | ACA CTA CCA GGT ACC TTT GAT GAG TGC TAT CCA ACA TAT TTG TAT CAA | 2208 |
| | AAA ATC GAT GAA TCA AAA TTA AAA GCC TTT ACC CGT TAT CAA TTA AGA | 2256 |
| 25 | GGG TAT ATC GAA GAT AGT CAA GAC TTA GAA ATC TAT TTA ATT CGC TAC | 2304 |
| | AAT GCA AAA CAT GAA ACA GTA AAT GTG CCA GGT ACG GGT TCC TTA TGG | 2352 |
| | CCG CTT TCA GCC CAA AGT CCA ATC GGA AAG TGT GGA GAG CCG AAT CGA | 2400 |
| | TGC GCG CCA CAC CTT GAA TGG AAT CCT GAC TTA GAT TGT TCG TGT AGG | 2448 |
| | GAT GGA GAA AAG TGT GCC CAT CAT TCG CAT CAT TTC TCC TTA GAC ATT | 2496 |
| 30 | GAT GTA GGA TGT ACA GAC TTA AAT GAG GAC CTA GGT GTA TGG GTG ATC | 2544 |
| | TTT AAG ATT AAG ACG CAA GAT GGG CAC GCA AGA CTA GGG AAT CTA GAG | 2592 |
| | TTT CTC GAA GAG AAA CCA TTA GTA GGA GAA GCG CTA GCT CGT GTG AAA | 2640 |
| | AGA GCG GAG AAA AAA TGG AGA GAC AAA CGT GAA AAA TTG GAA TGG GAA | 2688 |
| | ACA AAT ATC GTT TAT AAA GAG GCA AAA GAA TCT GTA GAT GCT TTA TTT | 2736 |
| 35 | GTA AAC TCT CAA TAT GAT CAA TTA CAA GCG GAT ACG AAT ATT GCC ATG | 2784 |
| | ATT CAT GCG GCA GAT AAA CGT GTT CAT AGC ATT CGA GAA GCT TAT CTG | 2832 |
| | CCT GAG CTG TCT GTG ATT CCG GGT GTC AAT GCG GCT ATT TTT GAA GAA | 2880 |
| | TTA GAA GGG CGT ATT TTC ACT GCA TTC TCC CTA TAT GAT GCG AGA AAT | 2928 |



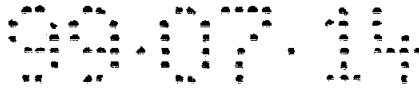
GTC ATT AAA AAT GGT GAT TTT AAT AAT GGC TTA TCC TGC TGG AAC GTG 2976
AAA GGG CAT GTA GAT GTA GAA GAA CAA AAC AAC CAA CGT TCG GTC CTT 3024
GTT GTT CCG GAA TGG GAA GCA GAA GTG TCA CAA GAA GTT CGT GTC TGT 3072
CCG GGT CGT GGC TAT ATC CTT CGT GTC ACA GCG TAC AAG GAG GGA TAT 3120
5 GGA GAA GGT TGC GTA ACC ATT CAT GAG ATC GAG AAC AAT ACA GAC GAA 3168
CTG AAG TTT AGC AAC TGC GTA GAA GAG GAA ATC TAT CCA AAT AAC ACG 3216
GTA ACG TGT AAT GAT TAT ACT GTA AAT CAA GAA GAA TAC GGA GGT GCG 3264
TAC ACT TCT CGT AAT CGA GGA TAT AAC GAA GCT CCT TCC GTA CCA GCT 3312
GAT TAT GCG TCA GTC TAT GAA GAA AAA TCG TAT ACA GAT GGA CGA AGA 3360
10 GAG AAT CCT TGT GAA TTT AAC AGA GGG TAT AGG GAT TAC ACG CCA CTA 3408
CCA GTT GGT TAT GTG ACA AAA GAA TTA GAA TAC TTC CCA GAA ACC GAT 3456
AAG GTA TGG ATT GAG ATT GGA GAA ACG GAA GGA ACA TTT ATC GTG GAC 3504
AGC GTG GAA TTA CTC CTT ATG GAG GAA 3531

15 **6.7.4 编码 EG 11092 晶体蛋白的 DNA 序列 (SEQ ID NO: 25)**

ATG GAT AAC AAT CCG AAC ATC AAT GAA TGC ATT CCT TAT AAT TGT TTA 48
AGT AAC CCT GAA GTA GAA GTA TTA GGT GGA GAA AGA ATA GAA ACT GGT 96
TAC ACC CCA ATC GAT ATT TCC TTG TCG CTA ACG CAA TTT CTT TTG AGT 144
GAA TTT GTT CCC GGT GCT GGA TTT GTG TTA GGA CTA GTT GAT ATA ATA 192
20 TGG GGA ATT TTT GGT CCC TCT CAA TGG GAC GCA TTT CTT GTA CAA ATT 240
GAA CAG TTA ATT AAC CAA AGA ATA GAA GAA TTC GCT AGG AAC CAA GCC 288
ATT TCT AGA TTA GAA GGA CTA AGC AAT CTT TAT CAA ATT TAC GCA GAA 336
TCT TTT AGA GAG TGG GAA GCA GAT CCT ACT AAT CCA GCA TTA AGA GAA 384
GAG ATG CGT ATT CAA TTC AAT GAC ATG AAC AGT GCC CTT ACA ACC GCT 432
25 ATT CCT CTT TTT GCA GTT CAA AAT TAT CAA GTT CCT CTT TTA TCA GTA 480
TAT GTT CAA GCT GCA AAT TTA CAT TTA TCA GTT TTG AGA GAT GTT TCA 528
GTG TTT GGA CAA AGG TGG GGA TTT GAT GCC GCG ACT ATC AAT AGT CGT 576
TAT AAT GAT TTA ACT AGG CTT AT GGC AAC TAT ACA GAT CAT GCT GTA 624
CGC TGG TAC AAT ACG GGA TTA GAG CGT GTA TGG GGA CCG GAT TCT AGA 672
30 GAT TGG ATA AGA TAT AAT CAA TTT AGA AGA GAA TTA ACA CTA ACT GTA 720
TTA GAT ATC GTT TCT CTA TTT CCG AAC TAT GAT AGT AGA ACG TAT CCA 768
ATT CGA ACA GTT TCC CAA TTA ACA AGA GAA ATT TAT ACA AAC CCA GTA 816
TTA GAA AAT TTT GAT GGT AGT TTT CGA GGC TCG GCT CAG GGC ATA GAA 864
AGA AGT ATT AGG AGT CCA CAT TTG ATG GAT ATA CTT AAC AGT ATA ACC 912
35 ATC TAT ACG GAT GCT CAT AGG GGT TAT TAT TAT TGG TCA GGG CAT CAA 960
ATA ATG GCT TCT CCT GTA GGG TTT TCG GGG CCA GAA TTC ACT TTT CCG 1008
CTA TAT GGA ACT ATG GGA AAT GCA GCT CCA CAA CAA CGT ATT GTT GCT 1056



CAA CTA GGT CAG GGC GTG TAT AGA ACA TTA TCG TCC ACT TTA TAT AGA 1104
AGA CCT TTT AAT ATA GGG ATA AAT AAT CAA CAA CTA TCT GTT CTT GAC 1152
GGG ACA GAA TTT GCT TAT GGA ACC TCC TCA AAT TTG CCA TCC GCT GTA 1200
TAC AGA AAA AGC GGA ACG GTA GAT TCG CTG GAT GAA ATA CCG CCA CAG 1248
5 AAT AAC AAC GTG CCA CCT AGG CAA GGA TTT AGT CAT CGA TTA AGC CAT 1296
GTT TCA ATG TTT CGT TCA GGC TTT AGT AAT AGT AGT GTA AGT ATA ATA 1344
AGA GCT CCA ATG TTT TCT TGG ACG CAC CGT AGT GCA ACC CCT ACA AAT 1392
ACA ATT GAT CCG GAG AGG ATT ACT CAA ATA CCA TTG GTA AAA GCA CAT 1440
ACA CTT CAG TCA GGT ACT ACT GTT GTA AGA GGG CCC GGG TTT ACG GGA 1488
10 GGA GAT ATT CTT CGA CGA ACA AGT GGA GGA CCA TTT GCT TAT ACT ATT 1536
GTT AAT ATA AAT GGG CAA TTA CCC CAA AGG TAT CGT GCA AGA ATA CGC 1584
TAT GCC TCT ACT ACA AAT CTA AGA ATT TAC GTA ACG GTT GCA GGT GAA 1632
CGG ATT TTT GCT GGT CAA TTT AAC AAA ACA ATG GAT ACC GGT GAC CCA 1680
TTA ACA TTC CAA TCT TTT AGT TAC GCA ACT ATT AAT ACA GCT TTT ACA 1728
15 TTC CCA ATG AGC CAG AGT AGT TTC ACA GTA GGT GCT GAT ACT TTT AGT 1776
TCA GGG AAT GAA GTT TAT ATA GAC AGA TTT GAA TTG ATT CCA GTT ACT 1824
GCA ACA TTT GAA GCA GAA TAT GAT TTA GAA AGA GCA CAA AAG GCG GTG 1872
AAT GCG CTG TTT ACT TCT ATA AAC CAA ATA GGG ATA AAA ACA GAT GTG 1920
ACG GAT TAT CAT ATT GAT CAA GTA TCC AAT TTA GTG GAT TGT TTA TCA 1968
20 GAT GAA TTT TGT CTG GAT GAA AAG CGA GAA TTG TCC GAG AAA GTC AAA 2016
CAT GCG AAG CGA CTC AGT GAT GAG CGG AAT TTA CTT CAA GAT CCA AAC 2064
TTC AAA GGC ATC AAT AGG CAA CTA GAC CGT GGT TGG AGA GGA AGT ACG 2112
GAT ATT ACC ATC CAA AGA GGA GAT GAC GTA TTC AAA GAA AAT TAT GTC 2160
ACA CTA CCA GGT ACC TTT GAT GAG TGC TAT CCA ACA TAT TTG TAT CAA 2208
25 AAA ATC GAT GAA TCA AAA TTA AAA GCC TTT ACC CGT TAT CAA TTA AGA 2256
GGG TAT ATC GAA GAT AGT CAA GAC TTA GAA ATC TAT TTA ATT CGC TAC 2304
AAT GCA AAA CAT GAA ACA GTA AAT GTG CCA GGT ACG GGT TCC TTA TGG 2352
CCG CTT TCA GCC CAA AGT CCA ATC GGA AAG TGT GGA GAG CCG AAT CGA 2400
TGC GCG CCA CAC CTT GAA TGG AAT CCT GAC TTA GAT TGT TCG TGT AGG 2448
30 GAT GGA GAA AAG TGT GCC CAT CAT TCG CAT CAT TTC TCC TTA GAC ATT 2496
GAT GTA GGA TGT ACA GAC TTA AAT GAG GAC CTA GGT GTA TGG GTG ATC 2544
TTT AAG ATT AAG ACG CAA GAT GGG CAC GCA AGA CTA GGG AAT CTA GAG 2592
TTT CTC GAA GAG AAA CCA TTA GTA GGA GAA GCG CTA GCT CGT GTG AAA 2640
AGA GCG GAG AAA AAA TGG AGA GAC AAA CGT GAA AAA TTG GAA TGG GAA 2688
35 ACA AAT ATC GTT TAT AAA GAG GCA AAA GAA TCT GTA GAT GCT TTA TTT 2736
GTA AAC TCT CAA TAT GAT CAA TTA CAA GCG GAT ACG AAT ATT GCC ATG 2784
ATT CAT GCG GCA GAT AAA CGT GTT CAT AGC ATT CGA GAA GCT TAT CTG 2832
CCT GAG CTG TCT GTG ATT CCG GGT GTC AAT GCG GCT ATT TTT GAA GAA 2880

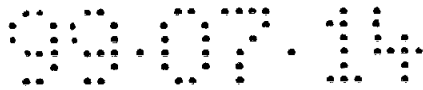


TTA GAA GGG CGT ATT TTC ACT GCA TTC TCC CTA TAT GAT GCG AGA AAT 2928
GTC ATT AAA AAT GGT GAT TTT AAT AAT GGC TTA TCC TGC TGG AAC GTG 2976
AAA GGG CAT GTA GAT GTA GAA GAA CAA AAC AAC CAA CGT TCG GTC CTT 3024
GTT GTT CCG GAA TGG GAA GCA GAA GTG TCA CAA GAA GTT CGT GTC TGT 3072
5 CCG GGT CGT GGC TAT ATC CTT CGT GTC ACA GCG TAC AAG GAG GGA TAT 3120
GGA GAA GGT TGC GTA ACC ATT CAT GAG ATC GAG AAC AAT ACA GAC GAA 3168
CTG AAG TTT AGC AAC TGC GTA GAA GAG GAA ATC TAT CCA AAT AAC ACG 3216
GTA ACG TGT AAT GAT TAT ACT GTA AAT CAA GAA GAA TAC GGA GGT GCG 3264
TAC ACT TCT CGT AAT CGA GGA TAT AAC GAA GCT CCT TCC GTA CCA GCT 3312
10 GAT TAT GCG TCA GTC TAT GAA GAA AAA TCG TAT ACA GAT GGA CGA AGA 3360
GAG AAT CCT TGT GAA TTT AAC AGA GGG TAT AGG GAT TAC ACG CCA CTA 3408
CCA GTT GGT TAT GTG ACA AAA GAA TTA GAA TAC TTC CCA GAA ACC GAT 3456
AAG GTA TGG ATT GAG ATT GGA GAA ACG GAA GGA ACA TTT ATC GTG GAC 3504
AGC GTG GAA TTA CTC CTT ATG GAG GAA TAG 3534

15

6.7.5 编码 EG 11751 晶体蛋白的 DNA 序列 (SEQ ID NO: 27)

ATG GAT AAC AAT CCG AAC ATC AAT GAA TGC ATT CCT TAT AAT TGT TTA 48
AGT AAC CCT GAA GTA GAA GTA TTA GGT GGA GAA AGA ATA GAA ACT GGT 96
TAC ACC CCA ATC GAT ATT TCC TTG TCG CTA ACG CAA TTT CTT TTG AGT 144
20 GAA TTT GTT CCC GGT GCT GGA TTT GTG TTA GGA CTA GTT GAT ATA ATA 192
TGG GGA ATT TTT GGT CCC TCT CAA TGG GAC GCA TTT CTT GTA CAA ATT 240
GAA CAG TTA ATT AAC CAA AGA ATA GAA GAA TTC GCT AGG AAC CAA GCC 288
ATT TCT AGA TTA GAA GGA CTA AGC AAT CTT TAT CAA ATT TAC GCA GAA 336
TCT TTT AGA GAG TGG GAA GCA GAT CCT ACT AAT CCA GCA TTA AGA GAA 384
25 GAG ATG CGT ATT CAA TTC AAT GAC ATG AAC AGT GCC CTT ACA ACC GCT 432
ATT CCT CTT TTT GCA GTT CAA AAT TAT CAA GTT CCT CTT TTA TCA GTA 480
TAT GTT CAA GCT GCA AAT TTA CAT TTA TCA GTT TTG AGA GAT GTT TCA 528
GTG TTT GGA CAA AGG TGG GGA TTT GAT GCC GCG ACT ATC AAT AGT CGT 576
TAT AAT GAT TTA ACT AGG CTT ATT GGC AAC TAT ACA GAT TAT GCT GTA 624
30 CGC TGG TAC AAT ACG GGA TTA GAA CGT GTA TGG GGA CCG GAT TCT AGA 672
GAT TGG GTA AGG TAT AAT CAA TTT AGA AGA GAA TTA ACA CTA ACT GTA 720
TTA GAT ATC GTT GCT CTG TTC CCG AAT TAT GAT AGT AGA AGA TAT CCA 768
ATT CGA ACA GTT TCC CAA TTA ACA AGA GAA ATT TAT ACA AAC CCA GTA 816
TTA GAA AAT TTT GAT GGT AGT TTT CGA GGC TCG GCT CAG GGC ATA GAA 864
35 AGA AGT ATT AGG AGT CCA CAT TTG ATG GAT ATA CTT AAC AGT ATA ACC 912
ATC TAT ACG GAT GCT CAT AGG GGT TAT TAT TAT TGG TCA GGG CAT CAA 960
ATA ATG GCT TCT CCT GTA GGG TTT TCG GGG CCA GAA TTC ACT TTT CCG 1008



CTA TAT GGA ACT ATG GGA AAT GCA GCT CCA CAA CAA CGT ATT GTT GCT 1056
CAA CTA GGT CAG GGC GTG TAT AGA ACA TTA TCG TCC ACT TTA TAT AGA 1104
AGA CCT TTT AAT ATA GGG ATA AAT AAT CAA CAA CTA TCT GTT CTT GAC 1152
GGG ACA GAA TTT GCT TAT GGA ACC TCC TCA AAT TTG CCA TCC GCT GTA 1200
5 TAC AGA AAA AGC GGA ACG GTA GAT TCG CTG GAT GAA ATA CCG CCA CAG 1248
AAT AAC AAC GTG CCA CCT AGG CAA GGA TTT AGT CAT CGA TTA AGC CAT 1296
GTT TCA ATG TTT CGT TCA GGC TTT AGT AAT AGT AGT GTA AGT ATA ATA 1344
AGA GCT CCT ATG TTC TCT TGG ATA CAT CGT AGT GCT GAA TTT AAT AAT 1392
ATA ATT GCA TCG GAT AGT ATT ACT CAA ATA CCA TTG GTA AAA GCA CAT 1440
10 ACA CTT CAG TCA GGT ACT ACT GTT GTA AGA GGG CCC GGG TTT ACG GGA 1488
GGA GAT ATT CTT CGA CGA ACA AGT GGA GGA CCA TTT GCT TAT ACT ATT 1536
GTT AAT ATA AAT GGG CAA TTA CCC CAA AGG TAT CGT GCA AGA ATA CGC 1584
TAT GCC TCT ACT ACA AAT CTA AGA ATT TAC GTA ACG GTT GCA GGT GAA 1632
CGG ATT TTT GCT GGT CAA TTT AAC AAA ACA ATG GAT ACC GGT GAC CCA 1680
15 TTA ACA TTC CAA TCT TTT AGT TAC GCA ACT ATT AAT ACA GCT TTT ACA 1728
TTC CCA ATG AGC CAG AGT AGT TTC ACA GTA GGT GCT GAT ACT TTT AGT 1776
TCA GGG AAT GAA GTT TAT ATA GAC AGA TTT GAA TTG ATT CCA GTT ACT 1824
GCA ACA TTT GAA GCA GAA TAT GAT TTA GAA AGA GCA CAA AAG GCG GTG 1872
AAT GCG CTG TTT ACT TCT ATA AAC CAA ATA GGG ATA AAA ACA GAT GTG 1920
20 ACG GAT TAT CAT ATT GAT CAA GTA TCC AAT TTA GTG GAT TGT TTA TCA 1968
GAT GAA TTT TGT CTG GAT GAA AAG CGA GAA TTG TCC GAG AAA GTC AAA 2016
CAT GCG AAG CGA CTC AGT GAT GAG CGG AAT TTA CTT CAA GAT CCA AAC 2064
TTC AAA GGC ATC AAT AGG CAA CTA GAC CGT GGT TGG AGA GGA AGT ACG 2112
GAT ATT ACC ATC CAA AGA GGA GAT GAC GTA TTC AAA GAA AAT TAT GTC 2160
25 ACA CTA CCA GGT ACC TTT GAT GAG TGC TAT CCA ACA TAT TTG TAT CAA 2208
AAA ATC GAT GAA TCA AAA TTA AAA GCC TTT ACC CGT TAT CAA TTA AGA 2256
GGG TAT ATC GAA GAT AGT CAA GAC TTA GAA ATC TAT TTA ATT CGC TAC 2304
AAT GCA AAA CAT GAA ACA GTA AAT GTG CCA GGT ACG GGT TCC TTA TGG 2352
CCG CTT TCA GCC CAA AGT CCA ATC GGA AAG TGT GGA GAG CCG AAT CGA 2400
30 TGC GCG CCA CAC CTT GAA TGG AAT CCT GAC TTA GAT TGT TCG TGT AGG 2448
GAT GGA GAA AAG TGT GCC CAT CAT TCG CAT CAT TTC TCC TTA GAC ATT 2496
GAT GTA GGA TGT ACA GAC TTA AAT GAG GAC CTA GGT GTA TGG GTG ATC 2544
TTT AAG ATT AAG ACG CAA GAT GGG CAC GCA AGA CTA GGG AAT CTA GAG 2592
TTT CTC GAA GAG AAA CCA TTA GTA GGA GAA GCG CTA GCT CGT GTG AAA 2640
35 AGA GCG GAG AAA AAA TGG AGA GAC AAA CGT GAA AAA TTG GAA TGG GAA 2688
ACA AAT ATC GTT TAT AAA GAG GCA AAA GAA TCT GTA GAT GCT TTA TTT 2736
GTA AAC TCT CAA TAT GAT CAA TTA CAA GCG GAT ACG AAT ATT GCC ATG 2784
ATT CAT GCG GCA GAT AAA CGT GTT CAT AGC ATT CGA GAA GCT TAT CTG 2832

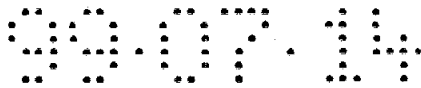
| | | |
|----|---|------|
| | CCT GAG CTG TCT GTG ATT CCG GGT GTC AAT GCG GCT ATT TTT GAA GAA | 2880 |
| | TTA GAA GGG CGT ATT TTC ACT GCA TTC TCC CTA TAT GAT GCG AGA AAT | 2928 |
| | GTC ATT AAA AAT GGT GAT TTT AAT AAT GGC TTA TCC TGC TGG AAC GTG | 2976 |
| | AAA GGG CAT GTA GAT GTA GAA GAA CAA AAC AAC CAA CGT TCG GTC CTT | 3024 |
| 5 | GTT GTT CCG GAA TGG GAA GCA GAA GTG TCA CAA GAA GTT CGT GTC TGT | 3072 |
| | CCG GGT CGT GGC TAT ATC CTT CGT GTC ACA GCG TAC AAG GAG GGA TAT | 3120 |
| | GGA GAA GGT TGC GTA ACC ATT CAT GAG ATC GAG AAC AAT ACA GAC GAA | 3168 |
| | CTG AAG TTT AGC AAC TGC GTA GAA GAG GAA ATC TAT CCA AAT AAC ACG | 3216 |
| | GTA ACG TGT AAT GAT TAT ACT GTA AAT CAA GAA GAA TAC GGA GGT GCG | 3264 |
| 10 | TAC ACT TCT CGT AAT CGA GGA TAT AAC GAA GCT CCT TCC GTA CCA GCT | 3312 |
| | GAT TAT GCG TCA GTC TAT GAA GAA AAA TCG TAT ACA GAT GGA CGA AGA | 3360 |
| | GAG AAT CCT TGT GAA TTT AAC AGA GGG TAT AGG GAT TAC ACG CCA CTA | 3408 |
| | CCA GTT GGT TAT GTG ACA AAA GAA TTA GAA TAC TTC CCA GAA ACC GAT | 3456 |
| | AAG GTA TGG ATT GAG ATT GGA GAA ACG GAA GGA ACA TTT ATC GTG GAC | 3504 |
| 15 | AGC GTG GAA TTA CTC CTT ATG GAG GAA TAG | 3534 |

6.7.6 编码 EG 11091 晶体蛋白的 DNA 序列 (SEQ ID NO: 29)

| | | |
|----|---|-----|
| | ATG GAT AAC AAT CCG AAC ATC AAT GAA TGC ATT CCT TAT AAT TGT TTA | 48 |
| | AGT AAC CCT GAA GTA GAA GTA TTA GGT GGA GAA AGA ATA GAA ACT GGT | 96 |
| 20 | TAC ACC CCA ATC GAT ATT TCC TTG TCG CTA ACG CAA TTT CTT TTG AGT | 144 |
| | GAA TTT GTT CCC GGT GCT GGA TTT GTG TTA GGA CTA GTT GAT ATA ATA | 192 |
| | TGG GGA ATT TTT GGT CCC TCT CAA TGG GAC GCA TTT CTT GTA CAA ATT | 240 |
| | GAA CAG TTA ATT AAC CAA AGA ATA GAA GAA TTC GCT AGG AAC CAA GCC | 288 |
| | ATT TCT AGA TTA GAA GGA CTA AGC AAT CTT TAT CAA ATT TAC GCA GAA | 336 |
| 25 | TCT TTT AGA GAG TGG GAA GCA GAT CCT ACT AAT CCA GCA TTA AGA GAA | 384 |
| | GAG ATG CGT ATT CAA TTC AAT GAC ATG AAC AGT GCC CTT ACA ACC GCT | 432 |
| | ATT CCT CTT TTT GCA GTT CAA AAT TAT CAA GTT CCT CTT TTA TCA GTA | 480 |
| | TAT GTT CAA GCT GCA AAT TTA CAT TTA TCA GTT TTG AGA GAT GTT TCA | 528 |
| | GTG TTT GGA CAA AGG TGG GGA TTT GAT GCC GCG ACT ATC AAT AGT CGT | 576 |
| 30 | TAT AAT GAT TTA ACT AGG CTT ATT GGC AAC TAT ACA GAT TAT GCT GTA | 624 |
| | CGC TGG TAC AAT ACG GGA TTA GAA CGT GTA TGG GGA CCG GAT TCT AGA | 672 |
| | GAT TGG GTA AGG TAT AAT CAA TTT AGA AGA GAA TTA ACA CTA ACT GTA | 720 |
| | TTA GAT ATC GTT GCT CTG TTC CCG AAT TAT GAT AGT AGA AGA TAT CCA | 768 |
| | ATT CGA ACA GTT TCC CAA TTA ACA AGA GAA ATT TAT ACA AAC CCA GTA | 816 |
| 35 | TTA GAA AAT TTT GAT GGT AGT TTT CGA GGC TCG GCT CAG GGC ATA GAA | 864 |
| | AGA AGT ATT AGG AGT CCA CAT TTG ATG GAT ATA CTT AAC AGT ATA ACC | 912 |
| | ATC TAT ACG GAT GCT CAT AGG GGT TAT TAT TAT TGG TCA GGG CAT CAA | 960 |



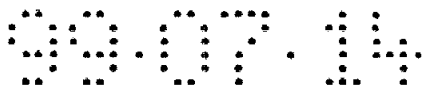
| | | |
|----|---|------|
| | ATA ATG GCT TCT CCT GTA GGG TTT TCG GGG CCA GAA TTC ACT TTT CCG | 1008 |
| | CTA TAT GGA ACT ATG GGA AAT GCA GCT CCA CAA CAA CGT ATT GTT GCT | 1056 |
| | CAA CTA GGT CAG GGC GTG TAT AGA ACA TTA TCG TCC ACT TTA TAT AGA | 1104 |
| | AGA CCT TTT AAT ATA GGG ATA AAT AAT CAA CAA CTA TCT GTT CTT GAC | 1152 |
| 5 | GGG ACA GAA TTT GCT TAT GGA ACC TCC TCA AAT TTG CCA TCC GCT GTA | 1200 |
| | TAC AGA AAA AGC GGA ACG GTA GAT TCG CTG GAT GAA ATA CCG CCA CAG | 1248 |
| | AAT AAC AAC GTG CCA CCT AGG CAA GGA TTT AGT CAT CGA TTA AGC CAT | 1296 |
| | GTT TCA ATG TTT CGT TCA GGC TTT AGT AAT AGT AGT GTA AGT ATA ATA | 1344 |
| | AGA GCT CCT ATG TTC TCT TGG ATA CAT CGT AGT GCA ACT CTT ACA AAT | 1392 |
| 10 | ACA ATT GAT CCA GAG AGA ATT AAT CAA ATA CCT TTA GTG AAA GGA TTT | 1440 |
| | AGA GTT TGG GGG GGC ACC TCT GTC ATT ACA GGA CCA GGA TTT ACA GGA | 1488 |
| | GGG GAT ATC CTT CGA AGA AAT ACC TTT GGT GAT TTT GTA TCT CTA CAA | 1536 |
| | GTC AAT ATT AAT TCA CCA ATT ACC CAA AGA TAC CGT TTA AGA TTT CGT | 1584 |
| | TAC GCT TCC AGT AGG GAT GCA CGA GTT ATA GTA TTA ACA GGA GCG GCA | 1632 |
| 15 | TCC ACA GGA GTG GGA GGC CAA GTT AGT GTA AAT ATG CCT CTT CAG AAA | 1680 |
| | ACT ATG GAA ATA GGG GAG AAC TTA ACA TCT AGA ACA TTT AGA TAT ACC | 1728 |
| | GAT TTT AGT AAT CCT TTT TCA TTT AGA GCT AAT CCA GAT ATA ATT GGG | 1776 |
| | ATA AGT GAA CAA CCT CTA TTT GGT GCA GGT TCT ATT AGT AGC GGT GAA | 1824 |
| | CTT TAT ATA GAT AAA ATT GAA ATT ATT CTA GCA GAT GCA ACA TTT GAA | 1872 |
| 20 | GCA GAA TCT GAT TTA GAA AGA GCA CAA AAG GCG GTG AAT GCC CTG TTT | 1920 |
| | ACT TCT TCC AAT CAA ATC GGG TTA AAA ACC GAT GTG ACG GAT TAT CAT | 1968 |
| | ATT GAT CAA GTA TCC AAT TTA GTG GAT TGT TTA TCA GAT GAA TTT TGT | 2016 |
| | CTG GAT GAA AAG CGA GAA TTG TCC GAG AAA GTC AAA CAT GCG AAG CGA | 2064 |
| | CTC AGT GAT GAG CGG AAT TTA CTT CAA GAT CCA AAC TTC AGA GGG ATC | 2112 |
| 25 | AAT AGA CAA CCA GAC CGT GGC TGG AGA GGA AGT ACA GAT ATT ACC ATC | 2160 |
| | CAA GGA GGA GAT GAC GTA TTC AAA GAG AAT TAC GTC ACA CTA CCG GGT | 2208 |
| | ACC GTT GAT GAG TGC TAT CCA ACG TAT TTA TAT CAG AAA ATA GAT GAG | 2256 |
| | TCG AAA TTA AAA GCT TAT ACC CGT TAT GAA TTA AGA GGG TAT ATC GAA | 2304 |
| | GAT AGT CAA GAC TTA GAA ATC TAT TTG ATC CGT TAC AAT GCA AAA CAC | 2352 |
| 30 | GAA ATA GTA AAT GTG CCA GGC ACG GGT TCC TTA TGG CCG CTT TCA GCC | 2400 |
| | CAA AGT CCA ATC GGA AAG TGT GGA GAA CCG AAT CGA TGC GCG CCA CAC | 2448 |
| | CTT GAA TGG AAT CCT GAT CTA GAT TGT TCC TGC AGA GAC GGG GAA AAA | 2496 |
| | TGT GCA CAT CAT TCC CAT CAT TTC ACC TTG GAT ATT GAT GTT GGA TGT | 2544 |
| | ACA GAC TTA AAT GAG GAC TTA GGT GTA TGG GTG ATA TTC AAG ATT AAG | 2592 |
| 35 | ACG CAA GAT GGC CAT GCA AGA CTA GGG AAT CTA GAG TTT CTC GAA GAG | 2640 |
| | AAA CCA TTA TTA GGG GAA GCA CTA GCT CGT GTG AAA AGA GCG GAG AAG | 2688 |
| | AAG TGG AGA GAC AAA CGA GAG AAA CTG CAG TTG GAA ACA AAT ATT GTT | 2736 |
| | TAT AAA GAG GCA AAA GAA TCT GTA GAT GCT TTA TTT GTA AAC TCT CAA | 2784 |



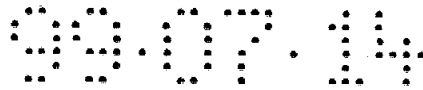
TAT GAT AGA TTA CAA GTG GAT ACG AAC ATC GCA ATG ATT CAT GCG GCA 2832
GAT AAA CGC GTT CAT AGA ATC CGG GAA GCG TAT CTG CCA GAG TTG TCT 2880
GTG ATT CCA GGT GTC AAT GCG GCC ATT TTC GAA GAA TTA GAG GGA CGT 2928
ATT TTT ACA GCG TAT TCC TTA TAT GAT GCG AGA AAT GTC ATT AAA AAT 2976
5 GGC GAT TTC AAT AAT GGC TTA TTA TGC TGG AAC GTG AAA GGT CAT GTA 3024
GAT GTA GAA GAG CAA AAC AAC CAC CGT TCG GTC CTT GTT ATC CCA GAA 3072
TGG GAG GCA GAA GTG TCA CAA GAG GTT CGT GTC TGT CCA GGT CGT GGC 3120
TAT ATC CTT CGT GTC ACA GCA TAT AAA GAG GGA TAT GGA GAG GGC TGC 3168
GTA ACG ATC CAT GAG ATC GAA GAC AAT ACA GAC GAA CTG AAA TTC AGC 3216
10 AAC TGT GTA GAA GAG GAA GTA TAT CCA AAC AAC ACA GTA ACG TGT AAT 3264
AAT TAT ACT GGG ACT CAA GAA GAA TAT GAG GGT ACG TAC ACT TCT CGT 3312
AAT CAA GGA TAT GAC GAA GCC TAT GGT AAT AAC CCT TCC GTA CCA GCT 3360
GAT TAC GCT TCA GTC TAT GAA GAA AAA TCG TAT ACA GAT GGA CGA AGA 3408
GAG AAT CCT TGT GAA TCT AAC AGA GGC TAT GGG GAT TAC ACA CCA CTA 3456
15 CCG GCT GGT TAT GTA ACA AAG GAT TTA GAG TAC TTC CCA GAG ACC GAT 3504
AAG GTA TGG ACT GAG ATC GGA GAA ACA GAA GGA ACA TTC ATC GTG GAT 3552
AGC GTG GAA TTA CTC CTT ATG GAG GAA 3579

6.7.7 编码 EG 11768 晶体蛋白的 DNA 序列 (SEQ ID NO: 33)

20 ATG GAT AAC AAT CCG AAC ATC AAT GAA TGC ATT CCT TAT AAT TGT TTA 48
AGT AAC CCT GAA GTA GAA GTA TTA GGT GGA GAA AGA ATA GAA ACT GGT 96
TAC ACC CCA ATC GAT ATT TCC TTG TCG CTA ACG CAA TTT CTT TTG AGT 144
GAA TTT GTT CCC GGT GCT GGA TTT GTG TTA GGA CTA GTT GAT ATA ATA 192
TGG GGA ATT TTT GGT CCC TCT CAA TGG GAC GCA TTT CTT GTA CAA ATT 240
25 GAA CAG TTA ATT AAC CAA AGA ATA GAA GAA TTC GCT AGG AAC CAA GCC 288
ATT TCT AGA TTA GAA GGA CTA AGC AAT CTT TAT CAA ATT TAC GCA GAA 336
TCT TTT AGA GAG TGG GAA GCA GAT CCT ACT AAT CCA GCA TTA AGA GAA 384
GAG ATG CGT ATT CAA TTC AAT GAC ATG AAC AGT GCC CTT ACA ACC GCT 432
ATT CCT CTT TTT GCA GTT CAA AAT TAT CAA GTT CCT CTT TTA TCA GTA 480
30 TAT GTT CAA GCT GCA AAT TTA CAT TTA TCA GTT TTG AGA GAT GTT TCA 528
GTG TTT GGA CAA AGG TGG GGA TTT GAT GCC GCG ACT ATC AAT AGT CGT 576
TAT AAT GAT TTA ACT AGG CTT ATT GGC AAC TAT ACA GAT TAT GCT GTA 624
CGC TGG TAC AAT ACG GGA TTA GAA CGT GTA TGG GGA CCG GAT TCT AGA 672
GAT TGG GTA AGG TAT AAT CAA TTT AGA AGA GAA TTA ACA CTA ACT GTA 720
35 TTA GAT ATC GTT GCT CTG TTC CCG AAT TAT GAT AGT AGA AGA TAT CCA 768
ATT CGA ACA GTT TCC CAA TTA ACA AGA GAA ATT TAT ACA AAC CCA GTA 816
TTA GAA AAT TTT GAT GGT AGT TTT CGA GGC TCG GCT CAG GGC ATA GAA 864
AGA AGT ATT AGG AGT CCA CAT TTG ATG GAT ATA CTT AAC AGT ATA ACC 912



ATC TAT ACG GAT GCT CAT AGG GGT TAT TAT TAT TGG TCA GGG CAT CAA 960
ATA ATG GCT TCT CCT GTA GGG TTT TCG GGG CCA GAA TTC ACT TTT CCG 1008
CTA TAT GGA ACT ATG GGA AAT GCA GCT CCA CAA CAA CGT ATT GTT GCT 1056
CAA CTA GGT CAG GGC GTG TAT AGA ACA TTA TCG TCC ACT TTA TAT AGA 1104
5 AGA CCT TTT AAT ATA GGG ATA AAT AAT CAA CAA CTA TCT GTT CTT GAC 1152
GGG ACA GAA TTT GCT TAT GGA ACC TCC TCA AAT TTG CCA TCC GCT GTA 1200
TAC AGA AAA AGC GGA ACG GTA GAT TCG CTG GAT GAA ATA CCG CCA CAG 1248
AAT AAC AAC GTG CCA CCT AGG CAA GGA TTT AGT CAT CGA TTA AGC CAT 1296
GTT TCA ATG TTT CGT TCA GGC TTT AGT AAT AGT AGT GTA AGT ATA ATA 1344
10 AGA GCT CCT ATG TTC TCT TGG ATA CAT CGT AGT GCT GAA TTT AAT AAT 1392
ATA ATT GCA TCG GAT AGT ATT ACT CAA ATA CCA TTG GTA AAA GCA CAT 1440
ACA CTT CAG TCA GGT ACT ACT GTT GTA AGA GGG CCC GGG TTT ACG GGA 1488
GGA GAT ATT CTT CGA CGA ACA AGT GGA GGA CCA TTT GCT TAT ACT ATT 1536
GTT AAT ATA AAT GGG CAA TTA CCC CAA AGG TAT CGT GCA AGA ATA CGC 1584
15 TAT GCC TCT ACT ACA AAT CTA AGA ATT TAC GTA ACG GTT GCA GGT GAA 1632
CGG ATT TTT GCT GGT CAA TTT AAC AAA ACA ATG GAT ACC GGT GAC CCA 1680
TTA ACA TTC CAA TCT TTT AGT TAC GCA ACT ATT AAT ACA GCT TTT ACA 1728
TTC CCA ATG AGC CAG AGT AGT TTC ACA GTA GGT GCT GAT ACT TTT AGT 1776
TCA GGG AAT GAA GTT TAT ATA GAC AGA TTT GAA TTG ATT CCA GTT ACT 1824
20 GCA ACA CTC GAG GCT GAA TAT AAT CTG GAA AGA GCG CAG AAG GCG GTG 1872
AAT GCG CTG TTT ACG TCT ACA AAC CAA CTA GGG CTA AAA ACA AAT GTA 1920
ACG GAT TAT CAT ATT GAT CAA GTG TCC AAT TTA GTT ACG TAT TTA TCG 1968
GAT GAA TTT TGT CTG GAT GAA AAG CGA GAA TTG TCC GAG AAA GTC AAA 2016
CAT GCG AAG CGA CTC AGT GAT GAA CGC AAT TTA CTC CAA GAT TCA AAT 2064
25 TTC AAA GAC ATT AAT AGG CAA CCA GAA CGT GGG TGG GGC GGA AGT ACA 2112
GGG ATT ACC ATC CAA GGA GGG GAT GAC GTA TTT AAA GAA AAT TAC GTC 2160
ACA CTA TCA GGT ACC TTT GAT GAG TGC TAT CCA ACA TAT TTG TAT CAA 2208
AAA ATC GAT GAA TCA AAA TTA AAA GCC TTT ACC CGT TAT CAA TTA AGA 2256
GGG TAT ATC GAA GAT AGT CAA GAC TTA GAA ATC TAT TTA ATT CGC TAC 2304
30 AAT GCA AAA CAT GAA ACA GTA AAT GTG CCA GGT ACG GGT TCC TTA TGG 2352
CCG CTT TCA GCC CAA AGT CCA ATC GGA AAG TGT GGA GAG CCG AAT CGA 2400
TGC GCG CCA CAC CTT GAA TGG AAT CCT GAC TTA GAT TGT TCG TGT AGG 2448
GAT GGA GAA AAG TGT GCC CAT CAT TCG CAT CAT TTC TCC TTA GAC ATT 2496
GAT GTA GGA TGT ACA GAC TTA AAT GAG GAC CTA GGT GTA TGG GTG ATC 2544
35 TTT AAG ATT AAG ACG CAA GAT GGG CAC GCA AGA CTA GGG AAT CTA GAG 2592
TTT CTC GAA GAG AAA CCA TTA GTA GGA GAA GCG CTA GCT CGT GTG AAA 2640
AGA GCG GAG AAA AAA TGG AGA GAC AAA CGT GAA AAA TTG GAA TGG GAA 2688
ACA AAT ATC GTT TAT AAA GAG GCA AAA GAA TCT GTA GAT GCT TTA TTT 2736



GTA AAC TCT CAA TAT GAT CAA TTA CAA GCG GAT ACG AAT ATT GCC ATG 2784
ATT CAT GCG GCA GAT AAA CGT GTT CAT AGC ATT CGA GAA GCT TAT CTG 2832
CCT GAG CTG TCT GTG ATT CCG GGT GTC AAT GCG GCT ATT TTT GAA GAA 2880
TTA GAA GGG CGT ATT TTC ACT GCA TTC TCC CTA TAT GAT GCG AGA AAT 2928
GTC ATT AAA AAT GGT GAT TTT AAT AAT GGC TTA TCC TGC TGG AAC GTG 2976
5 AAA GGG CAT GTA GAT GTA GAA GAA CAA AAC AAC CAA CGT TCG GTC CTY 3024
GTT GTT CCG GAA TGG GAA GCA GAA GTG TCA CAA GAA GTT CGT GTC TGT 3072
CCG GGT CGT GGC TAT ATC CTT CGT GTC ACA GCG TAC AAG GAG GGA TAT 3120
GGA GAA GGT TGC GTA ACC ATT CAT GAG ATC GAG AAC AAT ACA GAC GAA 3168
CTG AAG TTT AGC AAC TGC GTA GAA GAG GAA ATC TAT CCA AAT AAC ACG 3216
GTA ACG TGT AAT GAT TAT ACT GTA AAT CAA GAA GAA TAC GGA GGT GCG 3264
TAC ACT TCT CGT AAT CGA GGA TAT AAC GAA GCT CCT TCC GTA CCA GCT 3312
10 GAT TAT GCG TCA GTC TAT GAA GAA AAA TCG TAT ACA GAT GGA CGA AGA 3360
GAG AAT CCT TGT GAA TTT AAC AGA GGG TAT AGG GAT TAC ACG CCA CTA 3408
CCA GTT GGT TAT GTG ACA AAA GAA TTA GAA TAC TTC CCA GAA ACC GAT 3456
AAG GTA TGG ATT GAG ATT GGA GAA ACG GAA GGA ACA TTT ATC GTG GAC 3504
AGC GTG GAA TTA CTC CTT ATG GAG GAA TAG 3534

15 6.8 实施例 8 -- 对 CRY* 变异体抗性的转基因植物的分离

6.8.1 植物基因的构建

以双链 DNA 形式存在的植物基因的表达包括由 RNA 聚合酶从该 DNA 的一条链转录出信使 RNA (mRNA) 以及随后在核内对 mRNA 初级转录本的加工。此加工包括 3' 非翻译区, 其中添加聚腺苷酸核苷酸至该 RNA 的 3' 末端。DNA 到 mRNA 的转录受到常称为“启动子”的 DNA 区域的调控。此启动子区域含有指导 (signal) RNA 聚合酶与 DNA 结合并用该 DNA 的一条链作为模板起始 mRNA 转录以形成相应的 RNA 链的碱基序列。

文献中已描述过许多在植物细胞中有活性的启动子。这些启动子可获自植物或植物病毒, 并且包括但不限于胭脂氨酸合酶 (NOS) 和章鱼氨酸合酶 (OCS) 启动子 (根癌农杆菌的诱导肿瘤的质粒中带有它们)、花椰菜花叶病毒 (CaMV) 19 S 和 35 S 启动子、核酮糖 - 1, 5 - 二磷酸羧化酶 (ssRUBISCO , 一种很丰富的植物多肽) 小亚单位的光诱导启动子、玄参花叶病毒 (FMV) 35 S 启动子。所有这些启动子已用于建立在植物中表达的各种 DNA 构建体 (参见, 如美国专利 No. 5, 463, 175, 在此特异性引入供参考)。

选定的特定启动子应该能够导致酶编码序列充分表达从而产生足够量的蛋白。一套优选的启动子为组成性启动子如在大多数植物中产生高水平表达的 CaMV 35 S 或 FMV 35 S 启动子（美国专利 No. 5, 378, 619, 在此特异性掺入供参考）。另一套优选的启动子为根部增强或特异性的启动子如 CaMV 来源的 4 as - 1 启动子或小麦 POX 1 启动子（美国专利 No. 5, 023, 179, 在此特异性掺入供参考; Hertig 等, 1991）。根部增强的或特异性的启动子对于在转基因玉米植物中控制玉米根虫（rootworm）（*Diabroticus* spp.）。

如果必要, 可修饰用于本发明 DNA 构建体（即嵌合的植物基因）中的启动子以影响其控制特征。例如, 可将在无光时抑制 ssRUBISCO 表达的 CaMV 35 S 启动子连接到 ssRUBISCO 基因的部分以构建在叶片中有活性而在根部无活性的启动子。可按这里所述使用得到的嵌合启动子。为了描述的目的, 术语“CaMV 35 S”启动子因此包括 CaMV 35 S 启动子的变体, 和通过与随或控制诱变等的操纵子区连接而得到的启动子。此外, 可改变该启动子以含有多个“增强子序列”以利于提高基因表达。

由本发明 DNA 构建体而产生的 RNA 也含有 5' 非翻译的前导区序列。此序列可来自被选定表达该基因的启动子并可特异性地加以修饰从而增加 mRNA 的翻译。5' 非翻译区也可获自病毒 RNA、适当的真核基因或合成的基因序列。本发明不限于其中的非翻译区源自伴随该启动子序列 5' 非翻译序列的构建体。

为了优化在单子叶植物如玉米中的表达, 内含子也应包括进该 DNA 表达构建体中。此内含子将一般位于靠近 mRNA 5' 末端的非翻译序列中。该内含子可从但不限于由玉米 hsp 70 内含子（美国专利 No. 5, 424, 412, 在此特异性掺入供参考）或水稻 Act 1 内含子（McElory 等, 1990）所组成的一套内含子中获得。如下面所示, 玉米 hsp 70 内含子在本发明中是有用的。

如上述, 本发明嵌合植物基因的 3' 非翻译区含有在植物中起作用的聚腺苷酸化信号从而导致添加腺苷酸核苷酸至该 RNA 的 3' 末端。优选 3' 区域的实例有（1）含有农杆菌诱导肿瘤（Ti）质粒基因聚腺苷酸化信号的 3' 转录的非翻译区, 如胭脂氨酸合酶（NOS）基因和（2）植物

基因如豌豆 ssRUBISCO E9 基因 (Fischhoff 等, 1987) 。

6.8.2 植物转化与表达

5 可通过任何合适的方法如这里详述的那些方法将本发明含有结构编
码序列的嵌合转基因插入到植物的基因组中。适当的植物转化载体包括
来自根癌农杆菌 Ti 质粒的那些载体以及, 如由 Herrera - Estrella
(1983)、Bevan (1983)、Klee (1985) 和欧洲专利申请公开 No.
EP 0120516 中公开的那些载体。除了来自农杆菌 Ti 或诱导根部 (Ri)
10 质粒的植物转化载体外, 可用其它方法将本发明的 DNA 构建体插入到植
物细胞中。这些方法可包括, 例如, 使用脂质体、电穿孔、通过微粒轰
击而增加自由 DNA 摄入、自由 DNA 转运的化学物质以及用病毒或花粉
的转化 (Fromm 等, 1986 ; Armstrong 等, 1990 ; Fromm 等,
1990) 。

15 6.8.3 用于 CRY*转基因的植物表达载体的构建

为了在此公开的 cry*变异体在转基因植物中的有效表达, 编码该变
异体的基因必须具有适当的序列组成 (Diehn 等, 1996) 。

20 为了将 cry*基因置于适于在单子叶植物中表达的载体中 (即, 在增
强的花椰菜花叶病毒 35 S 启动子的控制下并连接到 hsp 70 内含子上, 后
接胭脂氨酸合酶聚腺苷酸化位点, 如美国专利 No. 5, 424, 412 中所述, 在
此特异性掺入供参考), 用适当的酶如 NcoI 和 EcoRI 消化载体。然后电
泳纯化约 4.6 kb 的较大的载体片段并用 T4 DNA 连接酶连接至含有植物
化 cry*基因的适当限制性片段上。然后将连接混合物转化入大肠杆菌
中, 回收羧苄青霉素抗性的菌落并通过 DNA 小量制备方法回收质粒
25 DNA。然后可将 DNA 以酶如 NcoI 和 EcoRI (共同)、NotI、和 PstI
进行限制性内切核酸酶分析以鉴定出含有在增强的 CaMV 35 S 启动子控
制下融合到 hsp 70 内含子上 cry*基因编码序列的克隆。

30 为了将基因置于适于回收稳定转化和昆虫抗性植物的载体中, 可通
过凝胶电泳和纯化分离来自 pMON 33708 含有在增强的 CaMV 35 S 启动
子控制下融合到 hsp 70 内含子上赖氨酸氧化酶编码序列的限制片段。然
后可将此片段与载体如以 NotI 和牛肠碱性磷酸酶处理过的 pMON 30460

连接 (pMON 30460 含有在 CaMV 35 S 启动子控制下的新霉素磷酸转移酶编码序列)。然后可通过将此连接混合物转化入大肠杆菌中而获得卡那霉素抗性的菌落并且可通过以限制性内切核酸酶消化小量制备的 DNA 而鉴定含有得到质粒的菌落。限制酶如 NotI、EcoRV、HindIII、
5 NcoI、EcoRI 和 BglII 可用于鉴定含有以适当方向插入到 pMON 30460 的相应位点中从而使 2 个基因为串联重复 (即, 将 cry* 基因表达弹夹的 3' 末端连接到 nptIII 表达弹夹的 5' 末端上) 限制性片段的适当的克隆。然后通过将载体电穿孔进入原生质体中然后进行蛋白印迹和 ELISA 分析而确定在植物原生质体中由得到载体对 cry* 蛋白的表达。可基本上按美国专利 No. 5, 424, 412 中所述, 在此特异性掺入供参考, 通过粒子枪轰击
10 将此载体导入植物胚胎如玉米的基因组 DNA 中, 然后通过巴龙霉素筛选以获得表达 cry* 基因的玉米植物。在该实施例中, 通过以赋予潮霉素抗性质粒共轰击 (cobombardment) 而将载体导入玉米未成熟胚胎的盾片中, 然后进行潮霉素筛选和再生。然后通过 ELISA 分析鉴定表达 cry*
15 蛋白的转基因玉米品系。然后测试这些过程的子代种子对易感昆虫摄食的保护。

7.0 参考文献

下面的参考文献提供例举性方案或其它与这里提出内容互补性的细
20 节, 在此特异性地掺入它们供参考之用。

- 1980 年 4 月 1 日授权的美国专利 4, 196, 265。
- 1985 年 11 月 19 日授权的美国专利 4, 554, 101。
- 1987 年 7 月 28 日授权的美国专利 4, 683, 195。
- 25 1987 年 7 月 28 日授权的美国专利 4, 683, 202。
- 1987 年 10 月 27 日授权的美国专利 4, 702, 914。
- 1988 年 7 月 12 日授权的美国专利 4, 757, 011。
- 1988 年 9 月 6 日授权的美国专利 4, 769, 061。
- 1990 年 7 月 10 日授权的美国专利 4, 940, 835。
- 30 1990 年 10 月 23 日授权的美国专利 4, 965, 188。
- 1990 年 11 月 20 日授权的美国专利 4, 971, 908。

- 1991年1月22日授权的美国专利 4,987,071。
1991年4月2日授权的美国专利 5,004,863。
1991年5月14日授权的美国专利 5,015,580。
1991年6月11日授权的美国专利 5,023,179。
5 1991年10月8日授权的美国专利 5,055,294。
1992年7月7日授权的美国专利 5,128,130。
1993年1月5日授权的美国专利 5,176,995。
1994年9月20日授权的美国专利 5,349,124。
1995年1月3日授权的美国专利 5,378,619。
10 1995年1月10日授权的美国专利 5,380,831。
1995年1月24日授权的美国专利 5,384,253。
1995年5月16日授权的美国专利 5,416,102。
1995年6月13日授权的美国专利 5,424,412。
1995年8月15日授权的美国专利 5,441,884。
15 1995年9月12日授权的美国专利 5,449,681。
1995年10月31日授权的美国专利 5,463,175。
1996年3月19日授权的美国专利 5,500,365。
1997年5月20日授权的美国专利 5,631,359。
1997年8月19日授权的美国专利 5,659,123。
20 1984年10月3日授权的欧洲专利 EP 0120516。
1990年3月28日授权的欧洲专利 EP 0360257。
1991年7月25日公开的国际专利申请公开 No. WO 91/10725。
1993年4月15日公开的国际专利申请公开 No. WO 93/07278。
1995年1月19日公开的国际专利申请公开 No. WO 95/02058。
25 1995年3月19日公开的国际专利申请公开 No. WO 95/06730。
1995年11月16日公开的国际专利申请公开 No. WO 95/30752。
1995年11月16日公开的国际专利申请公开 No. WO 95/30753。
1992年4月30日授权的国际专利 WO 92/07065。
1993年8月5日授权的国际专利 WO 93/15187。
30 1993年11月25日授权的国际专利 WO 93/23569。
1994年2月3日授权的国际专利 WO 94/02595。

- 1994 年 6 月 23 日授权的国际专利 WO 94/13688。
1991 年 3 月 21 日授权的国际专利 WO 91/03162。
Abdullah 等, 生物技术, 4: 1087, 1986。
Adami 和 Nevins, 在: RNA 加工, 冷泉港实验室, p. 26, 1988。
5 Adang 等, 在: 用于作物保护的分子策略, Alan R. Liss, 公司, pp. 345 - 357, 1987。
Adelman 等, DNA, 213: 183 - 193, 1983。
Allen 和 Choun, “具有低度摄入网状内皮系统的大的单层脂质体”, FEBS 通讯, 223: 42 - 46, 1987。
10 Altschul 等, “基本的局部对比搜寻工具”, 分子生物学杂志, 215: 403 - 410, 1990。
Arvidson 等, 分子生物学, 3: 1533 - 1534, 1989。
Barton 等, 植物生理, 85: 1103 - 1109, 1987。
Baum 等, 应用环境微生物学, 56: 3420 - 3428, 1990。
15 Benbrook 等, In: Proceedings Bio Expo 1986, Butterworth, Stoneham, MA, pp. 27 - 54, 1986。
Bevan 等, 自然, 304: 184, 1983。
Bolivar 等, 基因, 2: 95, 1977。
Bosch 等, “具有新特性的重组苏云金芽孢杆菌: 用于抗性管理的可能性”, 生物/技术, 12: 915 - 918, 1994。
20 Brady 和 Wold, 在: RNA 加工, 冷泉港实验室, p. 224, 1988。
Brown, 核酸研究, 14 (24): 9549, 1986。
Bytebier 等, 美国自然科学院院报, 84: 5345, 1987。
Callis 等, 基因发育, 1: 1183, 1987。
25 Callis, Fromm, Walbot, 基因与发育, 1: 1183 - 1200, 1987。
Campbell, 在: 单克隆抗体技术、生化与分子生物学实验技术, 13 卷, Burden 与 Von Knippenberg, 编, pp. 75 - 83, Elsevier, Amsterdam, 1984。
Capecchi, “直接显微注射 DNA 至培养的哺乳动物细胞所致的高
30 效转化”, 细胞, 22 (2): 479 - 488, 1980。
Cashmore 等, 植物遗传工程, Plenum 出版社, 纽约, 29 - 38,

1983 .

Chambers 等, 细菌学杂志, 173: 3966 - 3976, 1991 .

Chang 等, 自然, 375: 615, 1978 .

Chau 等, 科学, 244: 174 - 181, 1989 .

5 Chen 等, 核酸研究, 20: 4581 - 9, 1992 .

Clapp. “进入造血细胞内部的体细胞基因治疗。当前的地位和将来的应用”, Clin. Perinatol. 20 (1): 155 - 168, 1993 .

Collins 和 Olive, 生物化学, 32: 2795 - 2799, 1993 .

Conway 和 Wickens, 在: RNA 加工, 冷泉港实验室, p. 40 ,

10 1988 .

Cornellssen 等, EMBO J, 5 (1): 37 - 40, 1986 .

Couvreur 等, “微胶囊 (Nanocapsules), 一种新的溶酶体性 (lysosomotropic) 载体”, FEBS 通讯, 84: 323 - 326, 1977 .

15 Couvreur, “作为胶体药物载体的 Polyalkyleanoacrylates”, Crit. Rev. Ther. Drug Carrier Syst., 5: 1 - 20, 1988 .

Crickmore 等, Abstr. 28 th Annu. Meet. Soc. Invert. Pathol., 康奈尔大学, Ithaca, 纽约, 1995 .

Cristou 等, 植物生理学, 87: 671 - 674, 1988 .

20 Curiel, Agarwal, Wagner, Cotten, “腺病毒增强转铁蛋白 - 聚赖氨酸 - 介导的基因转运”, 美国自然科学院院报, 88 (19): 8850 - 8854, 1991 .

Curiel, Wagner, Cotten, Birnstiel, Agarwal, Li, Loechel, Hu, “由偶联至 DNA - 聚赖氨酸复合物上腺病毒介导的高效基因转移”, 人类基因治疗, 3 (2): 147 - 154, 1992 .

25 Daar 等, 在: RNA 加工, 冷泉港实验室, p. 45, 1988 .

Dean 等, 核酸研究, 14 (5): 2229, 1986 .

Dedrick 等, 生物化学杂志, 262 (19): 90098 - 1106, 1987 .

Dhir 等, 植物细胞报告, 10: 97, 1991 .

30 Diehn, De Rocher, Green, “限制外源基因在植物中表达的困难: 从 B. t. 毒素基因取得的教训”, 在: 遗传工程, 18 卷, J. K. Setlow 编, Plenum 出版社, 纽约, 1996 .

- Donovan 等, 生物化学杂志, 263 (1): 561 - 567, 1988 .
- Doyle 等, 生物化学杂志, 261 (20): 9228 - 9236, 1986 .
- Dropulic 等, 病毒学杂志, 66: 1432 - 41, 1992 .
- Eglitis 与 Anderson, “用于将基因导入哺乳动物细胞中的逆转录病毒载体”, 生物技术, 6 (7): 608 - 614, 1988 .
- Eglitis, Kantoff, Kohn, Karson, Moen, Lothrop, Blaese, Anderson, “逆转录病毒介导的基因进入造血细胞”, 实验医学生物学进展, 241: 19 - 27, 1988 .
- Eichenlaub, 细菌学杂志, 138 (2): 559 - 566, 1979 .
- Elionor 等, 分子遗传遗传学, 218: 78 - 86, 1989 .
- Elroy - Stein 与 Moss, 美国自然科学院院报, 87: 6743 - 7, 1990 .
- Fiers 等, 自然, 273: 113, 1978 .
- Fischhoff 等, 生物/技术, 5: 807, 1987 .
- Fraley 等, 美国自然科学院院报, 80: 4803, 1983 .
- Fraley 等, 生物/技术, 3: 629 - 635, 1985 .
- Fromm, Taylor, Walbot, 自然, 319: 791 - 793, 1986 .
- Fromm, Taylor, Walbot, “由电穿孔转入单子叶和双子叶植物中基因的表达”, 美国自然科学院院报, 82 (17): 5824 - 5828, 1985 .
- Fromm 等, “嵌合基因在转基因玉米植物中的遗传与表达”, 生物技术 (NY), 8 (9): 833 - 839, 1990 .
- Fujimura 等, 植物组织培养通讯, 2: 74, 1985 .
- Fütterer 和 Hohn, “植物中翻译 - 规律与例外”, 植物分子生物学, 32: 159 - 189, 1996 .
- Fynan, Webster, Fuller, Haynes, Santoro, Robinson, “DNA 疫苗: 由肠道外、粘膜和基因枪接种所致的保护性免疫”, 美国自然科学院院报, 90 (24): 11478 - 11482, 1993 .
- Gallego 与 Nadal - Ginard, 在: RNA 加工, 冷泉港实验室, p. 61, 1988 .
- Gao 与 Huang, 核酸研究, 21: 2867 - 72, 1993 .



- Gawron - Burke 与 Baum , 遗传工程, 13: 237 - 263 , 1991 .
Gefter 等, 体细胞遗传学, 3: 231 - 236 , 1977 .
Genovese 与 Mrlcarek , 在: RNA 加工, 冷泉港实验室, p. 62 ,
1988 .
- 5 Gil 与 Proudfoot , 自然. 312: 473 , 1984 .
Gill 等, 生物化学杂志, 270: 27277 - 27282 , 1995 .
Goding , “单克隆抗体: 原理与实践” , pp. 60-74 , 第 2 版, 学术
出版社, Orlando , FL , 1986 .
Goeddel 等, 自然, 281: 544 , 1979 .
- 10 Goeddel 等, 核酸研究, 8: 4057 , 1980 .
Goodall 等, 在: RNA 加工, 冷泉港实验室, p. 63 , 1988 .
Graham 与 Van der Eb , “通过人腺病毒 5 DNA 转化大鼠细胞” ,
病毒学, 54 (2) : 536 - 539 , 1973 .
Green , 核酸研究. 16 (1) : 369 , 1988 .
- 15 Grochulski 等, 分子生物学杂志, 254: 457 - 464 , 1995 .
Gross 等, 在: RNA 加工, 冷泉港实验室, p. 128 , 1988 .
Guerrier - Takada 等, 细胞, 35: 849 , 1983 .
Hampel 与 Tritz , 生物化学, 28: 4929 , 1989 .
Hampel 等, 核酸研究, 18: 299 , 1990 .
- 20 Hampson 与 Rottman , 在: RNA 加工, 冷泉港实验室, p. 68 ,
1988 .
Hanley 与 Schuler , 核酸研究, 16 (14) : 7159 , 1988 .
Harlow 与 Lane , 在: 抗体: 实验指南, 冷泉港实验室, 冷泉港,
纽约, 1988 .
- 25 Helfman 与 Ricci , 在: RNA 加工, 冷泉港实验室, p. 219 , 1988 .
Henry - Michelland 等, “附着抗生素至微粒上; 在体外的制备、药
物释放与抗 - 微生物活性, 国际药理学杂志, 35: 121 - 127 , 1987 .
Herrera - Estrella 等, 自然, 303: 209 , 1983 .
Hess 等, J. Adv. Enzyme Reg., 7: 149 , 1968 .
- 30 Hertig 等, “小麦 (Triticum aestivum L.) 推断性过氧化物酶基因
的序列与组织特异性表达” , 植物分子生物学, 16 (1) : 171 - 174 ,

1991。

Hilber, Bodmer, Smith, Koller, “ Biolistic 转化 *Botryotinia fuckeliana* 的分生孢子”, 当今遗传学, 25 (2): 124 - 127, 1994。

Hitzeman 等, 生物化学杂志, 255: 2073, 1980。

5 Hoekema 等, 分子与细胞生物学, 7: 2914 - 2924, 1987。

Höfte 与 Whiteley, 微生物学综述, 53: 242 - 255, 1989。

Holland 等, 生物化学, 17: 4900, 1978。

Honee 等, 分子微生物学, 5: 2799 - 2806, 1991。

Honee 等, 核酸研究, 16 (13), 1988。

10 Hoover 等 (编), 在: 雷明顿药物科学, 第 15 版, Mack 出版公司, Easton, PA, 1975。

Horsch 等, 科学, 227: 1229 - 1231, 1985。

Horsch 等, 科学, 227: 1229, 1985。

Horton 等, 基因, 77: 61 - 68, 1989。

15 Humason, “动物组织技术”, W. H Freeman & 公司, 纽约, 1967。

Iannacone, Pasquale, Grieco, Francesco, Cellini, “ cry 3B 内毒素基因的特异性序列修饰得到高水平表达和昆虫抗性”, 植物分子生物学, 34: 485 - 496, 1997。

Itakura 等, 科学, 198: 1056, 1977。

20 Jaeger 等, 美国自然科学院院报, 86: 7706 - 7710, 1989。

Jameson 与 Wolf, “抗原指标: 一种预言抗原决定簇的新的算术”, 计算机应用生物科学, 4 (1): 181 - 6, 1988。

Jarret 等, 体外, 17: 825, 1981。

Jarret 等, 植物生理学, 49: 177, 1980。

25 Johnston 与 Tang, “基因枪转染动物细胞与遗传免疫”。细胞生物学方法, 43 (A): 353 - 365, 1994。

Jones, 遗传学, 85: 12, 1977。

Jorgensen 等, 分子遗传学, 207: 471, 1987。

Kaiser 等, “两亲性二级结构: 肽激素的设计”, 科学, 223 (4633): 249 - 255, 1984。

30 Kashani - Saber 等, 反义研究进展, 2: 3 - 15, 1992。

- Kay 等, 科学, 236: 1299 - 1302, 1987。
- Keller 等, EMBO 杂志, 8: 1309 - 14, 1989。
- Kessler 等, 在: RNA 加工, 冷泉港实验室, p. 85, 1988。
- Kingsman 等, 基因, 7: 141, 1979。
- 5 Klee 等, 生物/技术, 3: 637, 1985。
- Klee 等, 生物/技术, 3: 637 - 642, 1985。
- Klein 等, 自然, 327: 70, 1987。
- Klein 等, 美国自然科学院院报, 85: 8502 - 8505, 1988。
- Knight 等, 生物化学杂志, 270: 17765 - 17770, 1995。
- 10 Kohler 与 Milstein, 欧洲免疫学杂志, 6: 511 - 519, 1976。
- Kohler 与 Milstein, 自然, 256: 495 - 497, 1975。
- Kozak, 自然, 308: 241 - 246, 1984。
- Krebbers 等, 植物分子生物学, 11: 745 - 759, 1988。
- Kuby, 免疫学, 第 2 版, W. H. Freeman & 公司, 纽约, 1994。
- 15 Kunkel, “不用表型筛选的快速而有效的定点诱变”, 美国自然科学院院报, 82 (2): 488 - 492, 1985。
- Kunkel 等, “不用表型筛选的快速而有效的定点诱变”, 酶学方法, 154: 367 - 382, 1987。
- Kyte 与 Doolittle, “用于显示蛋白亲水特征的简单方法”, 分子生
- 20 物学杂志, 157 (1): 105 - 132, 1982。
- Langridge 等, 美国自然科学院院报, 86: 3219 - 3223, 1989。
- Lee 等, 生化·生物物理研究通讯, 216: 306 - 312, 1995。
- L'Huillier 等, EMBO 杂志, 11: 4411 - 8, 1992。
- Lieber 等, 酶学方法, 217: 47 - 66, 1993。
- 25 Lisziewicz 等, 美国自然科学院院报, 90: 8000 - 4, 1993。
- Lim 等, “带有无义密码子的人 β - 珠蛋白 mRNA 在鼠红细胞系组织中降解成缺乏外显子 I 或外显子 I 与 II 区域并在 5' 末端具有帽状结构的中间体”, EMBO 杂志, 11 (9): 3271 - 3278, 1992。
- Lindstrom 等, 发育遗传学, 11: 160, 1990。
- 30 Lorz 等, 分子遗传学, 199: 178, 1985。
- Lu, Xiao, Clapp, Li, Broxmeyer, “高效逆转录病毒介导的基

- 因转导入单个人脐带血中分离的未成熟和可移植 (replatable) CD 34 (3 +) 成红干细胞/祖细胞中” , 实验医学杂志, 178 (6): 2089 - 2096 , 1993 .
- Luo 等, 植物分子生物学报告, 6: 165 , 1988 .
- 5 Maddock 等, 第三届国际分子生物学大会, 摘要, No. 372, 1991 .
- Maloy 等, 在: 微生物遗传学, 第 2 版, Jones & Bartlett 出版社, Boston, MA, 1994 .
- Maniatis 等, 在: 分子克隆: 实验指南, 冷泉港实验室, 冷泉港, 纽约, 1982 .
- 10 Marcotte 等, 自然, 335: 454 , 1988 .
- Marrone 等, 经济昆虫学杂志, 78 - 290 - 293 , 1985 .
- Marzluff 与 Pandey , 在: RNA 加工, 冷泉港实验室, p. 244 , 1988 .
- Masson 等, 生物化学杂志, 270: 20309 - 20315 , 1995 .
- 15 McCabe 等, 生物技术, 6: 923 , 1988 .
- McCormick 等, 植物细胞报告, 5: 81 - 84 , 1986 .
- McDevitt 等, 细胞, 37: 993 - 999 , 1984 .
- McElory 等, “水稻 (*Oryza sativa*) 肌动蛋白基因家族的特征分析” , 植物分子生物学, 15 (2): 257 - 268 , 1990 .
- 20 Mettus 与 Macaluso , 应用环境微生物学, 56: 1128 - 1134 , 1990 .
- Murashige 与 Skoog , 植物生理学, 15: 473 , 1962 .
- Murray, Lotzer, Eberle, “植物基因中密码子的使用” , 核酸研究, 17: (2) 477 - 498 , 1989 .
- 25 Neuhaus 等, 理论、应用遗传学, 75: 30 , 1987 .
- Odell 等, 自然, 313: 810 , 1985 .
- Odell 等, 自然, 313 , 810 , 1985 .
- Ojwang 等, 美国自然科学院院报, 89: 10802 - 6 , 1992 .
- Ohkawa 等, Nucl. Acids Symp. Ser., 27: 15 - 6 , 1992 .
- 30 Omirulleh 等, 植物分子生物学, 21: 415 - 428 , 1993 .
- Pandey 与 Marzluff , 在: RNA 加工, 冷泉港实验室, p. 133 ,

1987。

Pieken 等, 科学, 253: 314, 1991。

Pena 等, 自然, 325: 274, 1987。

Perrault 等, 自然, 344: 565, 1990。

5 Poszkowski 等, EMBO 杂志, 3: 2719, 1989。

Potrykus 等, 分子遗传遗传学, 199: 183, 1985。

Poulsen 等, 分子遗传遗传学, 205: 193 - 200, 1986。

Prokop 与 Bajpai, “重组 DNA 技术 I”, 纽约科学院年鉴, 646 卷, 1991。

10 Proudfoot 等, 在: RNA 加工, 冷泉港实验室, p.17, 1987。

Reines 等, 分子生物学杂志, 196: 299 - 312, 1987。

Rogers 等, 在: 植物分子生物学方法, A. Weissbach 与 H. Weissbach, 编, 学术出版公司, 圣地亚哥, CA 1988。

Rogers 等, 酶学方法, 153: 253 - 277, 1987。

15 Rossi 等, Aids 研究。人类逆转录病毒, 8: 183, 1992。

Rouwental, Odette Mendes, Wolbert Douwe de Boer, “通过改变其密码子使用而增强编码绿色荧光蛋白基因在烟草中的表达”, 植物分子生物学, 33: 989 - 999, 1997。

20 Ruud 等, “在配体印迹 (Blats) 中苏云金芽孢杆菌 δ - 内毒素的不同区域可结合昆虫中肠膜蛋白”, 应用环境微生物学, 62 (8): 2753 - 2757, 1996。

Ruud 等, “苏云金芽孢杆菌 δ - 内毒素 Cry IA(b) 区域 III 的替代导致对 *Spodoptera exigua* 更强的毒性和不同的膜蛋白识别”, 应用环境微生物学, 62 (5): 1537 - 1543, 1996。

25 Sadofsky 与 Alwine, 分子与细胞生物学, 4 (8): 1460 - 1468, 1984。

Sambrook 等, 在: 分子克隆: 实验指南, 冷泉港实验室, 冷泉港, 纽约, 1989。

Sanders 等, 核酸研究, 15 (4): 1543, 1987。

30 Saville 与 Collins, 细胞, 61: 685 - 696, 1990。

Saville 与 Collins, 美国自然科学院院报, 88: 8826 - 8830,



1991.

Scanlon 等, 美国自然科学院院报, 88: 10591 - 5, 1991.

Scaringe 等, 核酸研究, 18: 5433 - 5441, 1990.

Schnepf 等, 生物化学杂志, 265: 20923 - 20930, 1990.

5 Schuler 等, 核酸研究, 10 (24): 8225 - 8244, 1982.

Segal, 在: 生化计算, 第 2 版, John Wiley & Sons, 纽约, 1976.

Shaw 与 Kamen, 细胞, 46: 659 - 667, 1986.

Shaw 与 Kamen, 在: RNA 加工, 冷泉港实验室, p. 220, 1987.

Simpson, 科学, 233: 34, 1986.

10 Spielmann 等, 分子遗传遗传学, 205: 34, 1986.

Spoerel, 酶学方法, 152: 588 - 597, 1987.

Stemmer, “通过随机分段与重新装配进行 DNA 重组: 体外重组用于分子进化”, 美国自然科学院院报, 91 (22): 10747 - 10751, 1994.

15 Stinchcomb 等, 自然, 282: 39, 1979.

Taira 等, 核酸研究, 19: 5125 - 30, 1991.

Thompson 等, 遗传工程, 17: 99 - 117, 1995.

Toriyama 等, 理论、应用遗传学, 73: 16, 1986.

Trolinder 与 Goodin, 植物细胞报告, 6: 231 - 234, 1987.

20 Tschemper 等, 基因, 10: 157, 1980.

Tsurushita 与 Korn, 在: RNA 加工, 冷泉港实验室, p. 215, 1987.

Tumer 等, 核酸研究, 14: 8, 3325, 1986.

Uchimiya 等, 分子遗传遗传学, 204: 204, 1986.

25 Usman 与 Cedergren, 化学科学展望, 17: 334, 1992.

Vaeck 等, 自然, 328: 33, 1987.

Van Tunen 等, EMBO 杂志, 7: 1257, 1988.

Vasil 等, “通过微粒轰击可再生的胚胎愈伤组织而获得的杀虫剂抗性可育转基因小麦植物”, 生物技术, 10: 667 - 674, 1992.

30 Vasil, 生物技术, 6: 397, 1988.

Velten 与 Schell, 核酸研究, 13: 6981 - 6998, 1985.

- Velten 等, EMBO 杂志, 3: 2723 - 2730, 1984。
- Ventura 等, 核酸研究, 21: 3249 - 55, 1993。
- Visser 等, 分子遗传遗传学, 212: 219 - 224, 1988。
- Vodkin 等, 细胞, 34: 1023, 1983。
- 5 Vogel 等, 细胞生化杂志 (增刊) 13 D: 312, 1989。
- Wagner, Zatloukal, Cotten, Kirlappos, Mechtler, Curiel, Birnstiel, “腺病毒与转铁蛋白 - 聚赖氨酸/DNA 复合物的偶联大大增强了受体介导的基因转移与转染基因的表达”, 美国自然科学院院报 89 (13): 6099 - 6013, 1992。
- 10 Webb 等, 植物科学通讯, 30: 1, 1983。
- Weerasinghe 等, 病毒学杂志, 65: 5531 - 4, 1991。
- Weissbach 与 Weissbash, 植物分子生物学方法, (编), 学术出版公司, 圣地亚哥, CA, 1988。
- Wenzler 等, 植物分子生物学, 12: 41 - 50, 1989。
- 15 Wickens 与 Stephenson, 科学, 226: 1045, 1984。
- Wickens 等, 在: RNA 加工, 冷泉港实验室, p. 9, 1987。
- Wiebauer 等, 分子与细胞生物学, 8 (5): 2042 - 2051, 1988。
- Woolf 等, 美国自然科学院院报, 89: 7305 - 7309, 1992。
- Wolf 等, “一组完整的氨基酸序列分析程序”, 计算机应用生物科
- 20 学, 4 (1): 187 - 91, 1988。
- Wong 与 Neumann, “电场介导的基因转移”, 生化与生物物理研究通讯, 107 (2): 584 - 587, 1982。
- Yamada 等, 植物细胞复制, 4: 85, 1986。
- Yamamoto 与 Iizuka, 生化与生物物理档案, 227 (1): 233 -
- 25 241, 1983。
- Yang 等, 美国自然科学院院报, 87: 4144 - 48, 1990。
- Zhou 等, 酶学方法, 101: 433, 1983。
- Zhou 等, 分子细胞生物学, 10: 4529 - 37, 1990。

30 在本发明的指导下可制备和执行在此公开并要求保护的所有组合物和方法而无需更多的实验过程。虽然就优选的实施方案而言已描述了本



发明的组合物与方法，但是在不偏离本发明概念、精神和范围情况下多种变化可应用于在此所述的组合物和方法以及步骤或各步骤的序列中，这对于本领域中的技术人员将是显而易见的。更为具体地，很显然，一些同时被化学与物理改变的试剂可被这里所述的试剂所替代而仍然得到相同或类似的结果。认为对本领域中技术人员显而易见的所有此类类似的替代与修饰均在附属权利要求所限定的本发明精神、范围和概念之中。

8.0 序列表

(1) 一般信息:

(i) 申请人:

(A) 名字: Ecogen 公司

5 (B) 街道: 2005 Cabot Boulevard West

(C) 城市: Langhorne

(D) 州: 宾夕法尼亚

(E) 国家: 美国

(F) 邮编 (ZIP): 19047 - 3023

10 (ii) 发明题目: 广谱 δ - 内毒素

(iii) 序列数目: 35

(iv) COMPUTER READABLE FORM:

(A) 计算机可读形式: 软盘

(B) 计算机: IBM PC 兼容机

15 (C) 操作系统: PC - DOS/MS - DOS

(D) 软件: 专利发布 # 1.0, # 1.3 版 (EPO)

(vi) 优先申请数据

(A) 申请号: 美国 08/922, 505

(B) 提交日期: 1997 年 9 月 3 日

20 (vi) 优先申请数据

(A) 申请号: 美国 08/757, 536

(B) 提交日期: 1996 年 11 月 27 日

(vi) 优先申请数据

(A) 申请号: 美国 08/754, 490

25 (B) 提交日期: 1996 年 11 月 20 日

(2) SEQ ID NO: 1 信息:

(i) 序列特征:

(A) 长度: 23 个碱基对

(B) 类型: 核酸

30 (C) 链型: 单链

(D) 拓扑学: 线性



(xi) 序列描述: SEQ ID NO: 1:

GGATAGCACT CATCAAAGGT ACC

23

(2) SEQ ID NO: 2 信息:

5 (i) 序列特征:

(A) 长度: 27 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑学: 线性

10 (xi) 序列描述: SEQ ID NO: 2:

GAAGATATCC AATTCGAACA GTTTCCC

27

(2) SEQ ID NO: 3 信息:

(i) 序列特征:

15 (A) 长度: 28 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑学: 线性

(xi) 序列描述: SEQ ID NO: 3:

20 CATATTCTGC CTCGAGTGTT GCAGTAAC

28

(2) SEQ ID NO: 4 信息:

(i) 序列特征:

(A) 长度: 17 个碱基对

25 (B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑学: 线性

(xi) 序列描述: SEQ ID NO: 4:

30 CCCGATCGGC CGCATGC

17

(2) SEQ ID NO: 5 信息:

(i) 序列特征:

(A) 长度: 17 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

5 (D) 拓扑学: 线性

(xi) 序列描述: SEQ ID NO: 5:

CATTGGAGCT CTCCATG

17

(2) SEQ ID NO: 6 信息:

10 (i) 序列特征:

(A) 长度: 16 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑学: 线性

15 (xi) 序列描述: SEQ ID NO: 6:

GCACTACGAT GTATCC

16

(2) SEQ ID NO: 7 信息:

(i) 序列特征:

20 (A) 长度: 20 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑学: 线性

(xi) 序列描述: SEQ ID NO: 7:

25 CATCGTAGTG CAACTCTTAC

20

(2) SEQ ID NO: 8 信息:

(i) 序列特征:

(A) 长度: 39 个碱基对

30 (B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑学: 线性

(xi) 序列描述: SEQ ID NO: 8:

CCAAGAAAAT ACTAGAGCTC TTGTTAAAAA AGGTGTTCC

39

5 (2) SEQ ID NO: 9 信息:

(i) 序列特征:

(A) 长度: 3531 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

10

(D) 拓扑学: 线性

(ix) 特征:

(A) 名字/关键: CDS

(B) 位置: 1..3531

(xi) 序列描述: SEQ ID NO: 9:

15

ATG GAT AAC AAT CCG AAC ATC AAT GAA TGC ATT CCT TAT AAT TGT TTA 48
Met Asp Asn Asn Pro Asn Ile Asn Glu Cys Ile Pro Tyr Asn Cys Leu
1 5 10 15

AGT AAC CCT GAA GTA GAA GTA TTA GGT GGA GAA AGA ATA GAA ACT GGT 96
Ser Asn Pro Glu Val Glu Val Leu Gly Gly Glu Arg Ile Glu Thr Gly
20 25 30

TAC ACC CCA ATC GAT ATT TCC TTG TCG CTA ACG CAA TTT CTT TTG AGT 144
Tyr Thr Pro Ile Asp Ile Ser Leu Ser Leu Thr Gln Phe Leu Leu Ser
35 40 45

GAA TTT GTT CCC GGT GCT GGA TTT GTG TTA GGA CTA GTT GAT ATA ATA 192
Glu Phe Val Pro Gly Ala Gly Phe Val Leu Gly Leu Val Asp Ile Ile
50 55 60

TGG GGA ATT TTT GGT CCC TCT CAA TGG GAC GCA TTT CTT GTA CAA ATT 240
Trp Gly Ile Phe Gly Pro Ser Gln Trp Asp Ala Phe Leu Val Gln Ile
65 70 75 80

GAA CAG TTA ATT AAC CAA AGA ATA GAA GAA TTC GCT AGG AAC CAA GCC 288
Glu Gln Leu Ile Asn Gln Arg Ile Glu Glu Phe Ala Arg Asn Gln Ala
85 90 95

ATT TCT AGA TTA GAA GGA CTA AGC AAT CTT TAT CAA ATT TAC GCA GAA 336
Ile Ser Arg Leu Glu Gly Leu Ser Asn Leu Tyr Gln Ile Tyr Ala Glu
100 105 110



| | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| TCT | TTT | AGA | GAG | TGG | GAA | GCA | GAT | CCT | ACT | AAT | CCA | GCA | TTA | AGA | GAA | 384 |
| Ser | Phe | Arg | Glu | Trp | Glu | Ala | Asp | Pro | Thr | Asn | Pro | Ala | Leu | Arg | Glu | |
| | | 115 | | | | | 120 | | | | | 125 | | | | |
| GAG | ATG | CGT | ATT | CAA | TTC | AAT | GAC | ATG | AAC | AGT | GCC | CTT | ACA | ACC | GCT | 432 |
| Glu | Met | Arg | Ile | Gln | Phe | Asn | Asp | Met | Asn | Ser | Ala | Leu | Thr | Thr | Ala | |
| | 130 | | | | | 135 | | | | | 140 | | | | | |
| ATT | CCT | CTT | TTT | GCA | GTT | CAA | AAT | TAT | CAA | GTT | CCT | CTT | TTA | TCA | GTA | 480 |
| Ile | Pro | Leu | Phe | Ala | Val | Gln | Asn | Tyr | Gln | Val | Pro | Leu | Leu | Ser | Val | |
| 145 | | | | | 150 | | | | | 155 | | | | | 160 | |
| TAT | GTT | CAA | GCT | GCA | AAT | TTA | CAT | TTA | TCA | GTT | TTG | AGA | GAT | GTT | TCA | 528 |
| Tyr | Val | Gln | Ala | Ala | Asn | Leu | His | Leu | Ser | Val | Leu | Arg | Asp | Val | Ser | |
| | | | | 165 | | | | | | 170 | | | | | 175 | |
| GTG | TTT | GGA | CAA | AGG | TGG | GGA | TTT | GAT | GCC | GCG | ACT | ATC | AAT | AGT | CGT | 576 |
| Val | Phe | Gly | Gln | Arg | Trp | Gly | Phe | Asp | Ala | Ala | Thr | Ile | Asn | Ser | Arg | |
| | | | 180 | | | | | 185 | | | | | | 190 | | |
| TAT | AAT | GAT | TTA | ACT | AGG | CTT | ATT | GGC | AAC | TAT | ACA | GAT | TAT | GCT | GTA | 624 |
| Tyr | Asn | Asp | Leu | Thr | Arg | Leu | Ile | Gly | Asn | Tyr | Thr | Asp | Tyr | Ala | Val | |
| | | 195 | | | | | 200 | | | | | | 205 | | | |
| CGC | TGG | TAC | AAT | ACG | GGA | TTA | GAA | CGT | GTA | TGG | GGA | CCG | GAT | TCT | AGA | 672 |
| Arg | Trp | Tyr | Asn | Thr | Gly | Leu | Glu | Arg | Val | Trp | Gly | Pro | Asp | Ser | Arg | |
| | 210 | | | | | 215 | | | | | | 220 | | | | |
| GAT | TGG | GTA | AGG | TAT | AAT | CAA | TTT | AGA | AGA | GAA | TTA | ACA | CTA | ACT | GTA | 720 |
| Asp | Trp | Val | Arg | Tyr | Asn | Gln | Phe | Arg | Arg | Glu | Leu | Thr | Leu | Thr | Val | |
| 225 | | | | | 230 | | | | | 235 | | | | | 240 | |
| TTA | GAT | ATC | GTT | GCT | CTG | TTC | CCG | AAT | TAT | GAT | AGT | AGA | AGA | TAT | CCA | 768 |
| Leu | Asp | Ile | Val | Ala | Leu | Phe | Pro | Asn | Tyr | Asp | Ser | Arg | Arg | Tyr | Pro | |
| | | | | 245 | | | | | | 250 | | | | | 255 | |
| ATT | CGA | ACA | GTT | TCC | CAA | TTA | ACA | AGA | GAA | ATT | TAT | ACA | AAC | CCA | GTA | 816 |
| Ile | Arg | Thr | Val | Ser | Gln | Leu | Thr | Arg | Glu | Ile | Tyr | Thr | Asn | Pro | Val | |
| | | | 260 | | | | | 265 | | | | | | 270 | | |
| TTA | GAA | AAT | TTT | GAT | GGT | AGT | TTT | CGA | GGC | TCG | GCT | CAG | GGC | ATA | GAA | 864 |
| Leu | Glu | Asn | Phe | Asp | Gly | Ser | Phe | Arg | Gly | Ser | Ala | Gln | Gly | Ile | Glu | |
| | | 275 | | | | | 280 | | | | | | 285 | | | |
| AGA | AGT | ATT | AGG | AGT | CCA | CAT | TTG | ATG | GAT | ATA | CTT | AAC | AGT | ATA | ACC | 912 |
| Arg | Ser | Ile | Arg | Ser | Pro | His | Leu | Met | Asp | Ile | Leu | Asn | Ser | Ile | Thr | |
| | | | 290 | | | | 295 | | | | | | 300 | | | |
| ATC | TAT | ACG | GAT | GCT | CAT | AGG | GGT | TAT | TAT | TAT | TGG | TCA | GGG | CAT | CAA | 960 |
| Ile | Tyr | Thr | Asp | Ala | His | Arg | Gly | Tyr | Tyr | Tyr | Trp | Ser | Gly | His | Gln | |
| 305 | | | | | 310 | | | | | 315 | | | | | 320 | |
| ATA | ATG | GCT | TCT | CCT | GTA | GGG | TTT | TCG | GGG | CCA | GAA | TTC | ACT | TTT | CCG | 1008 |
| Ile | Met | Ala | Ser | Pro | Val | Gly | Phe | Ser | Gly | Pro | Glu | Phe | Thr | Phe | Pro | |
| | | | | | 325 | | | | | 330 | | | | | 335 | |



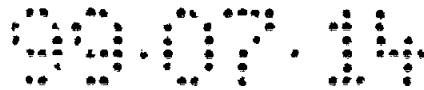
| | |
|---|------|
| CTA TAT GGA ACT ATG GGA AAT GCA GCT CCA CAA CAA CGT ATT GTT GCT Leu Tyr Gly Thr Met Gly Asn Ala Ala Pro Gln Gln Arg Ile Val Ala 340 345 350 | 1056 |
| CAA CTA GGT CAG GGC GTG TAT AGA ACA TTA TCG TCC ACT TTA TAT AGA Gln Leu Gly Gln Gly Val Tyr Arg Thr Leu Ser Ser Thr Leu Tyr Arg 355 360 365 | 1104 |
| AGA CCT TTT AAT ATA GGG ATA AAT AAT CAA CAA CTA TCT GTT CTT GAC Arg Pro Phe Asn Ile Gly Ile Asn Asn Gln Gln Leu Ser Val Leu Asp 370 375 380 | 1152 |
| GGG ACA GAA TTT GCT TAT GGA ACC TCC TCA AAT TTG CCA TCC GCT GTA Gly Thr Glu Phe Ala Tyr Gly Thr Ser Ser Asn Leu Pro Ser Ala Val 385 390 395 400 | 1200 |
| TAC AGA AAA AGC GGA ACG GTA GAT TCG CTG GAT GAA ATA CCG CCA CAG Tyr Arg Lys Ser Gly Thr Val Asp Ser Leu Asp Glu Ile Pro Pro Gln 405 410 415 | 1248 |
| AAT AAC AAC GTG CCA CCT AGG CAA GGA TTT AGT CAT CGA TTA AGC CAT Asn Asn Asn Val Pro Pro Arg Gln Gly Phe Ser His Arg Leu Ser His 420 425 430 | 1296 |
| GTT TCA ATG TTT CGT TCA GGC TTT AGT AAT AGT AGT GTA AGT ATA ATA Val Ser Met Phe Arg Ser Gly Phe Ser Asn Ser Ser Val Ser Ile Ile 435 440 445 | 1344 |
| AGA GCT CCA ATG TTT TCT TGG ACG CAC CGT AGT GCA ACC CCT ACA AAT Arg Ala Pro Met Phe Ser Trp Thr His Arg Ser Ala Thr Pro Thr Asn 450 455 460 | 1392 |
| ACA ATT GAT CCG GAG AGG ATT ACT CAA ATA CCA TTG GTA AAA GCA CAT Thr Ile Asp Pro Glu Arg Ile Thr Gln Ile Pro Leu Val Lys Ala His 465 470 475 480 | 1440 |
| ACA CTT CAG TCA GGT ACT ACT GTT GTA AGA GGG CCC GGG TTT ACG GGA Thr Leu Gln Ser Gly Thr Thr Val Val Arg Gly Pro Gly Phe Thr Gly 485 490 495 | 1488 |
| GGA GAT ATT CTT CGA CGA ACA AGT GGA GGA CCA TTT GCT TAT ACT ATT Gly Asp Ile Leu Arg Arg Thr Ser Gly Gly Pro Phe Ala Tyr Thr Ile 500 505 510 | 1536 |
| GTT AAT ATA AAT GGG CAA TTA CCC CAA AGG TAT CGT GCA AGA ATA CGC Val Asn Ile Asn Gly Gln Leu Pro Gln Arg Tyr Arg Ala Arg Ile Arg 515 520 525 | 1584 |
| TAT GCC TCT ACT ACA AAT CTA AGA ATT TAC GTA ACG GTT GCA GGT GAA Tyr Ala Ser Thr Thr Asn Leu Arg Ile Tyr Val Thr Val Ala Gly Glu 530 535 540 | 1632 |
| CGG ATT TTT GCT GGT CAA TTT AAC AAA ACA ATG GAT ACC GGT GAC CCA Arg Ile Phe Ala Gly Gln Phe Asn Lys Thr Met Asp Thr Gly Asp Pro 545 550 555 560 | 1680 |



| | |
|---|------|
| TTA ACA TTC CAA TCT TTT AGT TAC GCA ACT ATT AAT ACA GCT TTT ACA Leu Thr Phe Gln Ser Phe Ser Tyr Ala Thr Ile Asn Thr Ala Phe Thr 565 570 575 | 1728 |
| TTC CCA ATG AGC CAG AGT AGT TTC ACA GTA GGT GCT GAT ACT TTT AGT Phe Pro Met Ser Gln Ser Ser Phe Thr Val Gly Ala Asp Thr Phe Ser 580 585 590 | 1776 |
| TCA GGG AAT GAA GTT TAT ATA GAC AGA TTT GAA TTG ATT CCA GTT ACT Ser Gly Asn Glu Val Tyr Ile Asp Arg Phe Glu Leu Ile Pro Val Thr 595 600 605 | 1824 |
| GCA ACA TTT GAA GCA GAA TAT GAT TTA GAA AGA GCA CAA AAG GCG GTG Ala Thr Phe Glu Ala Glu Tyr Asp Leu Glu Arg Ala Gln Lys Ala Val 610 615 620 | 1872 |
| AAT GCG CTG TTT ACT TCT ATA AAC CAA ATA GGG ATA AAA ACA GAT GTG Asn Ala Leu Phe Thr Ser Ile Asn Gln Ile Gly Ile Lys Thr Asp Val 625 630 635 640 | 1920 |
| ACG GAT TAT CAT ATT GAT CAA GTA TCC AAT TTA GTG GAT TGT TTA TCA Thr Asp Tyr His Ile Asp Gln Val Ser Asn Leu Val Asp Cys Leu Ser 645 650 655 | 1968 |
| GAT GAA TTT TGT CTG GAT GAA AAG CGA GAA TTG TCC GAG AAA GTC AAA Asp Glu Phe Cys Leu Asp Glu Lys Arg Glu Leu Ser Glu Lys Val Lys 660 665 670 | 2016 |
| CAT GCG AAG CGA CTC AGT GAT GAG CGG AAT TTA CTT CAA GAT CCA AAC His Ala Lys Arg Leu Ser Asp Glu Arg Asn Leu Leu Gln Asp Pro Asn 675 680 685 | 2064 |
| TTC AAA GGC ATC AAT AGG CAA CTA GAC CGT GGT TGG AGA GGA AGT ACG Phe Lys Gly Ile Asn Arg Gln Leu Asp Arg Gly Trp Arg Gly Ser Thr 690 695 700 | 2112 |
| GAT ATT ACC ATC CAA AGA GGA GAT GAC GTA TTC AAA GAA AAT TAT GTC Asp Ile Thr Ile Gln Arg Gly Asp Asp Val Phe Lys Glu Asn Tyr Val 705 710 715 720 | 2160 |
| ACA CTA CCA GGT ACC TTT GAT GAG TGC TAT CCA ACA TAT TTG TAT CAA Thr Leu Pro Gly Thr Phe Asp Glu Cys Tyr Pro Thr Tyr Leu Tyr Gln 725 730 735 | 2208 |
| AAA ATC GAT GAA TCA AAA TTA AAA GCC TTT ACC CGT TAT CAA TTA AGA Lys Ile Asp Glu Ser Lys Leu Lys Ala Phe Thr Arg Tyr Gln Leu Arg 740 745 750 | 2256 |
| GGG TAT ATC GAA GAT AGT CAA GAC TTA GAA ATC TAT TTA ATT CGC TAC Gly Tyr Ile Glu Asp Ser Gln Asp Leu Glu Ile Tyr Leu Ile Arg Tyr 755 760 765 | 2304 |
| AAT GCA AAA CAT GAA ACA GTA AAT GTG CCA GGT ACG GGT TCC TTA TGG Asn Ala Lys His Glu Thr Val Asn Val Pro Gly Thr Gly Ser Leu Trp 770 775 780 | 2352 |

907.14

| | |
|---|------|
| CCG CTT TCA GCC CAA AGT CCA ATC GGA AAG TGT GGA GAG CCG AAT CGA | 2400 |
| Pro Leu Ser Ala Gln Ser Pro Ile Gly Lys Cys Gly Glu Pro Asn Arg | |
| 785 790 795 800 | |
| TGC GCG CCA CAC CTT GAA TGG AAT CCT GAC TTA GAT TGT TCG TGT AGG | 2448 |
| Cys Ala Pro His Leu Glu Trp Asn Pro Asp Leu Asp Cys Ser Cys Arg | |
| 805 810 815 | |
| GAT GGA GAA AAG TGT GCC CAT CAT TCG CAT CAT TTC TCC TTA GAC ATT | 2496 |
| Asp Gly Glu Lys Cys Ala His His Ser His His Phe Ser Leu Asp Ile | |
| 820 825 830 | |
| GAT GTA GGA TGT ACA GAC TTA AAT GAG GAC CTA GGT GTA TGG GTG ATC | 2544 |
| Asp Val Gly Cys Thr Asp Leu Asn Glu Asp Leu Gly Val Trp Val Ile | |
| 835 840 845 | |
| TTT AAG ATT AAG ACG CAA GAT GGG CAC GCA AGA CTA GGG AAT CTA GAG | 2592 |
| Phe Lys Ile Lys Thr Gln Asp Gly His Ala Arg Leu Gly Asn Leu Glu | |
| 850 855 860 | |
| TTT CTC GAA GAG AAA CCA TTA GTA GGA GAA GCG CTA GCT CGT GTG AAA | 2640 |
| Phe Leu Glu Glu Lys Pro Leu Val Gly Glu Ala Leu Ala Arg Val Lys | |
| 865 870 875 880 | |
| AGA GCG GAG AAA AAA TGG AGA GAC AAA CGT GAA AAA TTG GAA TGG GAA | 2688 |
| Arg Ala Glu Lys Lys Trp Arg Asp Lys Arg Glu Lys Leu Glu Trp Glu | |
| 885 890 895 | |
| ACA AAT ATC GTT TAT AAA GAG GCA AAA GAA TCT GTA GAT GCT TTA TTT | 2736 |
| Thr Asn Ile Val Tyr Lys Glu Ala Lys Glu Ser Val Asp Ala Leu Phe | |
| 900 905 910 | |
| GTA AAC TCT CAA TAT GAT CAA TTA CAA GCG GAT ACG AAT ATT GCC ATG | 2784 |
| Val Asn Ser Gln Tyr Asp Gln Leu Gln Ala Asp Thr Asn Ile Ala Met | |
| 915 920 925 | |
| ATT CAT GCG GCA GAT AAA CGT GTT CAT AGC ATT CGA GAA GCT TAT CTG | 2832 |
| Ile His Ala Ala Asp Lys Arg Val His Ser Ile Arg Glu Ala Tyr Leu | |
| 930 935 940 | |
| CCT GAG CTG TCT GTG ATT CCG GGT GTC AAT GCG GCT ATT TTT GAA GAA | 2880 |
| Pro Glu Leu Ser Val Ile Pro Gly Val Asn Ala Ala Ile Phe Glu Glu | |
| 945 950 955 960 | |
| TTA GAA GGG CGT ATT TTC ACT GCA TTC TCC CTA TAT GAT GCG AGA AAT | 2928 |
| Leu Glu Gly Arg Ile Phe Thr Ala Phe Ser Leu Tyr Asp Ala Arg Asn | |
| 965 970 975 | |
| GTC ATT AAA AAT GGT GAT TTT AAT AAT GGC TTA TCC TGC TGG AAC GTG | 2976 |
| Val Ile Lys Asn Gly Asp Phe Asn Asn Gly Leu Ser Cys Trp Asn Val | |
| 980 985 990 | |
| AAA GGG CAT GTA GAT GTA GAA GAA CAA AAC AAC CAA CGT TCG GTC CTT | 3024 |
| Lys Gly His Val Asp Val Glu Glu Gln Asn Asn Gln Arg Ser Val Leu | |
| 995 1000 1005 | |



| | | |
|----|---|------|
| | GTT GTT CCG GAA TGG GAA GCA GAA GTG TCA CAA GAA GTT CGT GTC TGT | 3072 |
| | Val Val Pro Glu Trp Glu Ala Glu Val Ser Gln Glu Val Arg Val Cys | |
| | 1010 1015 1020 | |
| | CCG GGT CGT GGC TAT ATC CTT CGT GTC ACA GCG TAC AAG GAG GGA TAT | 3120 |
| | Pro Gly Arg Gly Tyr Ile Leu Arg Val Thr Ala Tyr Lys Glu Gly Tyr | |
| | 1025 1030 1035 1040 | |
| 5 | GGA GAA GGT TGC GTA ACC ATT CAT GAG ATC GAG AAC AAT ACA GAC GAA | 3168 |
| | Gly Glu Gly Cys Val Thr Ile His Glu Ile Glu Asn Asn Thr Asp Glu | |
| | 1045 1050 1055 | |
| | CTG AAG TTT AGC AAC TGC GTA GAA GAG GAA ATC TAT CCA AAT AAC ACG | 3216 |
| | Leu Lys Phe Ser Asn Cys Val Glu Glu Glu Ile Tyr Pro Asn Asn Thr | |
| | 1060 1065 1070 | |
| 10 | GTA ACG TGT AAT GAT TAT ACT GTA AAT CAA GAA GAA TAC GGA GGT GCG | 3264 |
| | Val Thr Cys Asn Asp Tyr Thr Val Asn Gln Glu Glu Tyr Gly Gly Ala | |
| | 1075 1080 1085 | |
| | TAC ACT TCT CGT AAT CGA GGA TAT AAC GAA GCT CCT TCC GTA CCA GCT | 3312 |
| | Tyr Thr Ser Arg Asn Arg Gly Tyr Asn Glu Ala Pro Ser Val Pro Ala | |
| | 1090 1095 1100 | |
| | GAT TAT GCG TCA GTC TAT GAA GAA AAA TCG TAT ACA GAT GGA CGA AGA | 3360 |
| | Asp Tyr Ala Ser Val Tyr Glu Glu Lys Ser Tyr Thr Asp Gly Arg Arg | |
| | 1105 1110 1115 1120 | |
| 15 | GAG AAT CCT TGT GAA TTT AAC AGA GGG TAT AGG GAT TAC ACG CCA CTA | 3408 |
| | Glu Asn Pro Cys Glu Phe Asn Arg Gly Tyr Arg Asp Tyr Thr Pro Leu | |
| | 1125 1130 1135 | |
| | CCA GTT GGT TAT GTG ACA AAA GAA TTA GAA TAC TTC CCA GAA ACC GAT | 3456 |
| | Pro Val Gly Tyr Val Thr Lys Glu Leu Glu Tyr Phe Pro Glu Thr Asp | |
| | 1140 1145 1150 | |
| 20 | AAG GTA TGG ATT GAG ATT GGA GAA ACG GAA GGA ACA TTT ATC GTG GAC | 3504 |
| | Lys Val Trp Ile Glu Ile Gly Glu Thr Glu Gly Thr Phe Ile Val Asp | |
| | 1155 1160 1165 | |
| | AGC GTG GAA TTA CTC CTT ATG GAG GAA | 3531 |
| | Ser Val Glu Leu Leu Leu Met Glu Glu | |
| | 1170 1175 | |

(2) SEQ ID NO : 10 信息:

25

(i) 序列特征:

(A) 长度: 1177 个氨基酸

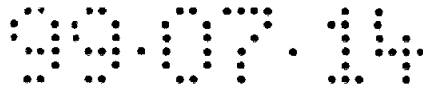
(B) 类型: 氨基酸

(D) 拓扑学: 线性

(ii) 分子类型: 蛋白

30

(xi) 序列描述: SEQ ID NO : 10 :



Met Asp Asn Asn Pro Asn Ile Asn Glu Cys Ile Pro Tyr Asn Cys Leu
1 5 10 15

Ser Asn Pro Glu Val Glu Val Leu Gly Gly Glu Arg Ile Glu Thr Gly
20 25 30

Tyr Thr Pro Ile Asp Ile Ser Leu Ser Leu Thr Gln Phe Leu Leu Ser
35 40 45

Glu Phe Val Pro Gly Ala Gly Phe Val Leu Gly Leu Val Asp Ile Ile
50 55 60

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

Glu Gln Leu Ile Asn Gln Arg Ile Glu Glu Phe Ala Arg Asn Gln Ala
85 90 95

Ile Ser Arg Leu Glu Gly Leu Ser Asn Leu Tyr Gln Ile Tyr Ala Glu
100 105 110

Ser Phe Arg Glu Trp Glu Ala Asp Pro Thr Asn Pro Ala Leu Arg Glu
115 120 125

Glu Met Arg Ile Gln Phe Asn Asp Met Asn Ser Ala Leu Thr Thr Ala
130 135 140

Ile Pro Leu Phe Ala Val Gln Asn Tyr Gln Val Pro Leu Leu Ser Val
145 150 155 160

Tyr Val Gln Ala Ala Asn Leu His Leu Ser Val Leu Arg Asp Val Ser
165 170 175

Val Phe Gly Gln Arg Trp Gly Phe Asp Ala Ala Thr Ile Asn Ser Arg
180 185 190

Tyr Asn Asp Leu Thr Arg Leu Ile Gly Asn Tyr Thr Asp Tyr Ala Val
195 200 205

Arg Trp Tyr Asn Thr Gly Leu Glu Arg Val Trp Gly Pro Asp Ser Arg
210 215 220

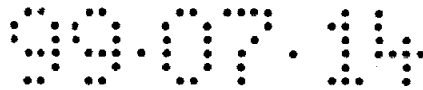
Asp Trp Val Arg Tyr Asn Gln Phe Arg Arg Glu Leu Thr Leu Thr Val
225 230 235 240

Leu Asp Ile Val Ala Leu Phe Pro Asn Tyr Asp Ser Arg Arg Tyr Pro
245 250 255

Ile Arg Thr Val Ser Gln Leu Thr Arg Glu Ile Tyr Thr Asn Pro Val
260 265 270

Leu Glu Asn Phe Asp Gly Ser Phe Arg Gly Ser Ala Gln Gly Ile Glu
275 280 285

Arg Ser Ile Arg Ser Pro His Leu Met Asp Ile Leu Asn Ser Ile Thr
290 295 300



Ile Tyr Thr Asp Ala His Arg Gly Tyr Tyr Tyr Trp Ser Gly His Gln
305 310 315 320

Ile Met Ala Ser Pro Val Gly Phe Ser Gly Pro Glu Phe Thr Phe Pro
325 330 335

Leu Tyr Gly Thr Met Gly Asn Ala Ala Pro Gln Gln Arg Ile Val Ala
340 345 350

Gln Leu Gly Gln Gly Val Tyr Arg Thr Leu Ser Ser Thr Leu Tyr Arg
355 360 365

Arg Pro Phe Asn Ile Gly Ile Asn Asn Gln Gln Leu Ser Val Leu Asp
370 375 380

Gly Thr Glu Phe Ala Tyr Gly Thr Ser Ser Asn Leu Pro Ser Ala Val
385 390 395 400

Tyr Arg Lys Ser Gly Thr Val Asp Ser Leu Asp Glu Ile Pro Pro Gln
405 410 415

Asn Asn Asn Val Pro Pro Arg Gln Gly Phe Ser His Arg Leu Ser His
420 425 430

Val Ser Met Phe Arg Ser Gly Phe Ser Asn Ser Ser Val Ser Ile Ile
435 440 445

Arg Ala Pro Met Phe Ser Trp Thr His Arg Ser Ala Thr Pro Thr Asn
450 455 460

Thr Ile Asp Pro Glu Arg Ile Thr Gln Ile Pro Leu Val Lys Ala His
465 470 475 480

Thr Leu Gln Ser Gly Thr Thr Val Val Arg Gly Pro Gly Phe Thr Gly
485 490 495

Gly Asp Ile Leu Arg Arg Trp Ser Gly Gly Pro Phe Ala Tyr Thr Ile
500 505 510

Val Asn Ile Asn Gly Gln Leu Pro Gln Arg Tyr Arg Ala Arg Ile Arg
515 520 525

Tyr Ala Ser Thr Thr Asn Leu Arg Ile Tyr Val Thr Val Ala Gly Glu
530 535 540

Arg Ile Phe Ala Gly Gln Phe Asn Lys Thr Met Asp Thr Gly Asp Pro
545 550 555 560

Leu Thr Phe Gln Ser Phe Ser Tyr Ala Thr Ile Asn Thr Ala Phe Thr
565 570 575

Phe Pro Met Ser Gln Ser Ser Phe Thr Val Gly Ala Asp Thr Phe Ser
580 585 590

Ser Gly Asn Glu Val Tyr Ile Asp Arg Phe Glu Leu Ile Pro Val Thr
595 600 605



Ala Thr Phe Glu Ala Glu Tyr Asp Leu Glu Arg Ala Gln Lys Ala Val
610 615 620

Asn Ala Leu Phe Thr Ser Ile Asn Gln Ile Gly Ile Lys Thr Asp Val
625 630 635 640

Thr Asp Tyr His Ile Asp Gln Val Ser Asn Leu Val Asp Cys Leu Ser
645 650 655

Asp Glu Phe Cys Leu Asp Glu Lys Arg Glu Leu Ser Glu Lys Val Lys
660 665 670

His Ala Lys Arg Leu Ser Asp Glu Arg Asn Leu Leu Gln Asp Pro Asn
675 680 685

Phe Lys Gly Ile Asn Arg Gln Leu Asp Arg Gly Trp Arg Gly Ser Thr
690 695 700

Asp Ile Thr Ile Gln Arg Gly Asp Asp Val Phe Lys Glu Asn Tyr Val
705 710 715 720

Thr Leu Pro Gly Thr Phe Asp Glu Cys Tyr Pro Thr Tyr Leu Tyr Gln
725 730 735

Lys Ile Asp Glu Ser Lys Leu Lys Ala Phe Thr Arg Tyr Gln Leu Arg
740 745 750

Gly Tyr Ile Glu Asp Ser Gln Asp Leu Glu Ile Tyr Leu Ile Arg Tyr
755 760 765

Asn Ala Lys His Glu Thr Val Asn Val Pro Gly Thr Gly Ser Leu Trp
770 775 780

Pro Leu Ser Ala Gln Ser Pro Ile Gly Lys Cys Gly Glu Pro Asn Arg
785 790 795 800

Cys Ala Pro His Leu Glu Trp Asn Pro Asp Leu Asp Cys Ser Cys Arg
805 810 815

Asp Gly Glu Lys Cys Ala His His Ser His His Phe Ser Leu Asp Ile
820 825 830

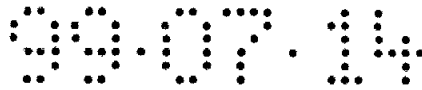
Asp Val Gly Cys Thr Asp Leu Asn Glu Asp Leu Gly Val Trp Val Ile
835 840 845

Phe Lys Ile Lys Thr Gln Asp Gly His Ala Arg Leu Gly Asn Leu Glu
850 855 860

Phe Leu Glu Glu Lys Pro Leu Val Gly Glu Ala Leu Ala Arg Val Lys
865 870 875 880

Arg Ala Glu Lys Lys Trp Arg Asp Lys Arg Glu Lys Leu Glu Trp Glu
885 890 895

Thr Asn Ile Val Tyr Lys Glu Ala Lys Glu Ser Val Asp Ala Leu Phe
900 905 910



Val Asn Ser Gln Tyr Asp Gln Leu Gln Ala Asp Thr Asn Ile Ala Met
915 920 925

Ile His Ala Ala Asp Lys Arg Val His Ser Ile Arg Glu Ala Tyr Leu
930 935 940

Pro Glu Leu Ser Val Ile Pro Gly Val Asn Ala Ala Ile Phe Glu Glu
945 950 955 960

5 Leu Glu Gly Arg Ile Phe Thr Ala Phe Ser Leu Tyr Asp Ala Arg Asn
965 970 975

Val Ile Lys Asn Gly Asp Phe Asn Asn Gly Leu Ser Cys Trp Asn Val
980 985 990

Lys Gly His Val Asp Val Glu Glu Gln Asn Asn Gln Arg Ser Val Leu
995 1000 1005

10 Val Val Pro Glu Trp Glu Ala Glu Val Ser Gln Glu Val Arg Val Cys
1010 1015 1020

Pro Gly Arg Gly Tyr Ile Leu Arg Val Thr Ala Tyr Lys Glu Gly Tyr
1025 1030 1035 1040

Gly Glu Gly Cys Val Thr Ile His Glu Ile Glu Asn Asn Thr Asp Glu
1045 1050 1055

15 Leu Lys Phe Ser Asn Cys Val Glu Glu Glu Ile Tyr Pro Asn Asn Thr
1060 1065 1070

Val Thr Cys Asn Asp Tyr Thr Val Asn Gln Glu Glu Tyr Gly Gly Ala
1075 1080 1085

Tyr Thr Ser Arg Asn Arg Gly Tyr Asn Glu Ala Pro Ser Val Pro Ala
1090 1095 1100

Asp Tyr Ala Ser Val Tyr Glu Glu Lys Ser Tyr Thr Asp Gly Arg Arg
1105 1110 1115 1120

20 Glu Asn Pro Cys Glu Phe Asn Arg Gly Tyr Arg Asp Tyr Thr Pro Leu
1125 1130 1135

Pro Val Gly Tyr Val Thr Lys Glu Leu Glu Tyr Phe Pro Glu Thr Asp
1140 1145 1150

Lys Val Trp Ile Glu Ile Gly Glu Thr Glu Gly Thr Phe Ile Val Asp
1155 1160 1165

25 Ser Val Glu Leu Leu Leu Met Glu Glu
1170 1175

(2) SEQ ID NO : 11 信息:

(i) 序列特征:

(A) 长度: 3531 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

30



(D) 拓扑学: 线性

(ix) 特征:

(A) 名字/关键: CDS

(B) 位置: 1..3531

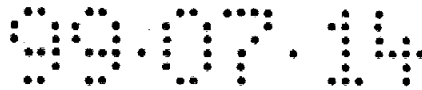
5

(xi) 序列描述: SEQ ID NO: 11:

| | |
|---|-----|
| ATG GAT AAC AAT CCG AAC ATC AAT GAA TGC ATT CCT TAT AAT TGT TTA | 48 |
| Met Asp Asn Asn Pro Asn Ile Asn Glu Cys Ile Pro Tyr Asn Cys Leu | |
| 1 5 10 15 | |
| AGT AAC CCT GAA GTA GAA GTA TTA GGT GGA GAA AGA ATA GAA ACT GGT | 96 |
| Ser Asn Pro Glu Val Glu Val Leu Gly Gly Glu Arg Ile Glu Thr Gly | |
| 20 25 30 | |
| TAC ACC CCA ATC GAT ATT TCC TTG TCG CTA ACG CAA TTT CTT TTG AGT | 144 |
| Tyr Thr Pro Ile Asp Ile Ser Leu Ser Leu Thr Gln Phe Leu Leu Ser | |
| 35 40 45 | |
| GAA TTT GTT CCC GGT GCT GGA TTT GTG TTA GGA CTA GTT GAT ATA ATA | 192 |
| Glu Phe Val Pro Gly Ala Gly Phe Val Leu Gly Leu Val Asp Ile Ile | |
| 50 55 60 | |
| TGG GGA ATT TTT GGT CCC TCT CAA TGG GAC GCA TTT CTT GTA CAA ATT | 240 |
| Trp Gly Ile Phe Gly Pro Ser Gln Trp Asp Ala Phe Leu Val Gln Ile | |
| 65 70 75 80 | |
| GAA CAG TTA ATT AAC CAA AGA ATA GAA GAA TTC GCT AGG AAC CAA GCC | 288 |
| Glu Gln Leu Ile Asn Gln Arg Ile Glu Glu Phe Ala Arg Asn Gln Ala | |
| 85 90 95 | |
| ATT TCT AGA TTA GAA GGA CTA AGC AAT CTT TAT CAA ATT TAC GCA GAA | 336 |
| Ile Ser Arg Leu Glu Gly Leu Ser Asn Leu Tyr Gln Ile Tyr Ala Glu | |
| 100 105 110 | |
| TCT TTT AGA GAG TGG GAA GCA GAT CCT ACT AAT CCA GCA TTA AGA GAA | 384 |
| Ser Phe Arg Glu Trp Glu Ala Asp Pro Thr Asn Pro Ala Leu Arg Glu | |
| 115 120 125 | |
| GAG ATG CGT ATT CAA TTC AAT GAC ATG AAC AGT GCC CTT ACA ACC GCT | 432 |
| Glu Met Arg Ile Gln Phe Asn Asp Met Asn Ser Ala Leu Thr Thr Ala | |
| 130 135 140 | |
| ATT CCT CTT TTT GCA GTT CAA AAT TAT CAA GTT CCT CTT TTA TCA GTA | 480 |
| Ile Pro Leu Phe Ala Val Gln Asn Tyr Gln Val Pro Leu Leu Ser Val | |
| 145 150 155 160 | |
| TAT GTT CAA GCT GCA AAT TTA CAT TTA TCA GTT TTG AGA GAT GTT TCA | 528 |
| Tyr Val Gln Ala Ala Asn Leu His Leu Ser Val Leu Arg Asp Val Ser | |



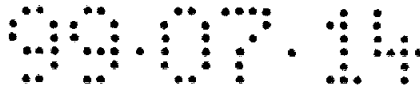
| | | | | | | | |
|---|-----|--|-----|--|-----|--|------|
| | 165 | | 170 | | 175 | | |
| GTG TTT GGA CAA AGG TGG GGA TTT GAT GCC GCG ACT ATC AAT AGT CGT | | | | | | | 576 |
| Val Phe Gly Gln Arg Trp Gly Phe Asp Ala Ala Thr Ile Asn Ser Arg | | | | | | | |
| | 180 | | 185 | | 190 | | |
| TAT AAT GAT TTA ACT AGG CTT ATT GGC AAC TAT ACA GAT TAT GCT GTA | | | | | | | 624 |
| Tyr Asn Asp Leu Thr Arg Leu Ile Gly Asn Tyr Thr Asp Tyr Ala Val | | | | | | | |
| | 195 | | 200 | | 205 | | |
| CGC TGG TAC AAT ACG GGA TTA GAA CGT GTA TGG GGA CCG GAT TCT AGA | | | | | | | 672 |
| Arg Trp Tyr Asn Thr Gly Leu Glu Arg Val Trp Gly Pro Asp Ser Arg | | | | | | | |
| | 210 | | 215 | | 220 | | |
| GAT TGG GTA AGG TAT AAT CAA TTT AGA AGA GAA TTA ACA CTA ACT GTA | | | | | | | 720 |
| Asp Trp Val Arg Tyr Asn Gln Phe Arg Arg Glu Leu Thr Leu Thr Val | | | | | | | |
| | 225 | | 230 | | 235 | | 240 |
| TTA GAT ATC GTT GCT CTG TTC CCG AAT TAT GAT AGT AGA AGA TAT CCA | | | | | | | 768 |
| Leu Asp Ile Val Ala Leu Phe Pro Asn Tyr Asp Ser Arg Arg Tyr Pro | | | | | | | |
| | 245 | | 250 | | 255 | | |
| ATT CGA ACA GTT TCC CAA TTA ACA AGA GAA ATT TAT ACA AAC CCA GTA | | | | | | | 816 |
| Ile Arg Thr Val Ser Gln Leu Thr Arg Glu Ile Tyr Thr Asn Pro Val | | | | | | | |
| | 260 | | 265 | | 270 | | |
| TTA GAA AAT TTT GAT GGT AGT TTT CGA GGC TCG GCT CAG GGC ATA GAA | | | | | | | 864 |
| Leu Glu Asn Phe Asp Gly Ser Phe Arg Gly Ser Ala Gln Gly Ile Glu | | | | | | | |
| | 275 | | 280 | | 285 | | |
| AGA AGT ATT AGG AGT CCA CAT TTG ATG GAT ATA CTT AAC AGT ATA ACC | | | | | | | 912 |
| Arg Ser Ile Arg Ser Pro His Leu Met Asp Ile Leu Asn Ser Ile Thr | | | | | | | |
| | 290 | | 295 | | 300 | | |
| ATC TAT ACG GAT GCT CAT AGG GGT TAT TAT TAT TGG TCA GGG CAT CAA | | | | | | | 960 |
| Ile Tyr Thr Asp Ala His Arg Gly Tyr Tyr Tyr Trp Ser Gly His Gln | | | | | | | |
| | 305 | | 310 | | 315 | | 320 |
| ATA ATG GCT TCT CCT GTA GGG TTT TCG GGG CCA GAA TTC ACT TTT CCG | | | | | | | 1008 |
| Ile Met Ala Ser Pro Val Gly Phe Ser Gly Pro Glu Phe Thr Phe Pro | | | | | | | |
| | 325 | | 330 | | 335 | | |
| CTA TAT GGA ACT ATG GGA AAT GCA GCT CCA CAA CAA CGT ATT GTT GCT | | | | | | | 1056 |
| Leu Tyr Gly Thr Met Gly Asn Ala Ala Pro Gln Gln Arg Ile Val Ala | | | | | | | |
| | 340 | | 345 | | 350 | | |
| CAA CTA GGT CAG GGC GTG TAT AGA ACA TTA TCG TCC ACT TTA TAT AGA | | | | | | | 1104 |
| Gln Leu Gly Gln Gly Val Tyr Arg Thr Leu Ser Ser Thr Leu Tyr Arg | | | | | | | |
| | 355 | | 360 | | 365 | | |
| AGA CCT TTT AAT ATA GGG ATA AAT AAT CAA CAA CTA TCT GTT CTT GAC | | | | | | | 1152 |
| Arg Pro Phe Asn Ile Gly Ile Asn Asn Gln Gln Leu Ser Val Leu Asp | | | | | | | |
| | 370 | | 375 | | 380 | | |
| GGG ACA GAA TTT GCT TAT GGA ACC TCC TCA AAT TTG CCA TCC GCT GTA | | | | | | | 1200 |
| Gly Thr Glu Phe Ala Tyr Gly Thr Ser Ser Asn Leu Pro Ser Ala Val | | | | | | | |
| | 385 | | 390 | | 395 | | 400 |



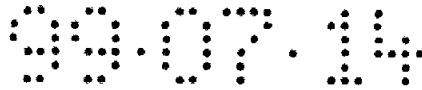
| | |
|---|------|
| TAC AGA AAA AGC GGA ACG GTA GAT TCG CTG GAT GAA ATA CCG CCA CAG Tyr Arg Lys Ser Gly Thr Val Asp Ser Leu Asp Glu Ile Pro Pro Gln 405 410 415 | 1248 |
| AAT AAC AAC GTG CCA CCT AGG CAA GGA TTT AGT CAT CGA TTA AGC CAT Asn Asn Asn Val Pro Pro Arg Gln Gly Phe Ser His Arg Leu Ser His 420 425 430 | 1296 |
| GTT TCA ATG TTT CGT TCA GGC TTT AGT AAT AGT AGT GTA AGT ATA ATA Val Ser Met Phe Arg Ser Gly Phe Ser Asn Ser Ser Val Ser Ile Ile 435 440 445 | 1344 |
| AGA GCT CCA ATG TTT TCT TGG ACG CAC CGT AGT GCA ACC CCT ACA AAT Arg Ala Pro Met Phe Ser Trp Thr His Arg Ser Ala Thr Pro Thr Asn 450 455 460 | 1392 |
| ACA ATT GAT CCG GAG AGG ATT ACT CAA ATA CCA TTG GTA AAA GCA CAT Thr Ile Asp Pro Glu Arg Ile Thr Gln Ile Pro Leu Val Lys Ala His 465 470 475 480 | 1440 |
| ACA CTT CAG TCA GGT ACT ACT GTT GTA AGA GGG CCC GGG TTT ACG GGA Thr Leu Gln Ser Gly Thr Thr Val Val Arg Gly Pro Gly Phe Thr Gly 485 490 495 | 1488 |
| GGA GAT ATT CTT CGA CGA ACA AGT GGA GGA CCA TTT GCT TAT ACT ATT Gly Asp Ile Leu Arg Arg Thr Ser Gly Gly Pro Phe Ala Tyr Thr Ile 500 505 510 | 1536 |
| GTT AAT ATA AAT GGG CAA TTA CCC CAA AGG TAT CGT GCA AGA ATA CGC Val Asn Ile Asn Gly Gln Leu Pro Gln Arg Tyr Arg Ala Arg Ile Arg 515 520 525 | 1584 |
| TAT GCC TCT ACT ACA AAT CTA AGA ATT TAC GTA ACG GTT GCA GGT GAA Tyr Ala Ser Thr Thr Asn Leu Arg Ile Tyr Val Thr Val Ala Gly Glu 530 535 540 | 1632 |
| CGG ATT TTT GCT GGT CAA TTT AAC AAA ACA ATG GAT ACC GGT GAC CCA Arg Ile Phe Ala Gly Gln Phe Asn Lys Thr Met Asp Thr Gly Asp Pro 545 550 555 560 | 1680 |
| TTA ACA TTC CAA TCT TTT AGT TAC GCA ACT ATT AAT ACA GCT TTT ACA Leu Thr Phe Gln Ser Phe Ser Tyr Ala Thr Ile Asn Thr Ala Phe Thr 565 570 575 | 1728 |
| TTC CCA ATG AGC CAG AGT AGT TTC ACA GTA GGT GCT GAT ACT TTT AGT Phe Pro Met Ser Gln Ser Ser Phe Thr Val Gly Ala Asp Thr Phe Ser 580 585 590 | 1776 |
| TCA GGG AAT GAA GTT TAT ATA GAC AGA TTT GAA TIG ATT CCA GTT ACT Ser Gly Asn Glu Val Tyr Ile Asp Arg Phe Glu Leu Ile Pro Val Thr 595 600 605 | 1824 |
| GCA ACA CTC GAG GCT GAA TAT AAT CTG GAA AGA GCG CAG AAG GCG GTG Ala Thr Leu Glu Ala Glu Tyr Asn Leu Glu Arg Ala Gln Lys Ala Val 610 615 620 | 1872 |



| | |
|---|------|
| AAT GCG CTG TTT ACG TCT ACA AAC CAA CTA GGG CTA AAA ACA AAT GTA | 1920 |
| Asn Ala Leu Phe Thr Ser Thr Asn Gln Leu Gly Leu Lys Thr Asn Val | |
| 625 630 635 640 | |
| ACG GAT TAT CAT ATT GAT CAA GTG TCC AAT TTA GTT ACG TAT TTA TCG | 1968 |
| Thr Asp Tyr His Ile Asp Gln Val Ser Asn Leu Val Thr Tyr Leu Ser | |
| 645 650 655 | |
| GAT GAA TTT TGT CTG GAT GAA AAG CGA GAA TTG TCC GAG AAA GTC AAA | 2016 |
| Asp Glu Phe Cys Leu Asp Glu Lys Arg Glu Leu Ser Glu Lys Val Lys | |
| 660 665 670 | |
| CAT GCG AAG CGA CTC AGT GAT GAA CGC AAT TTA CTC CAA GAT TCA AAT | 2064 |
| His Ala Lys Arg Leu Ser Asp Glu Arg Asn Leu Leu Gln Asp Ser Asn | |
| 675 680 685 | |
| TTC AAA GAC ATT AAT AGG CAA CCA GAA CGT GGG TGG GGC GGA AGT ACA | 2112 |
| Phe Lys Asp Ile Asn Arg Gln Pro Glu Arg Gly Trp Gly Gly Ser Thr | |
| 690 695 700 | |
| GGG ATT ACC ATC CAA GGA GGG GAT GAC GTA TTT AAA GAA AAT TAC GTC | 2160 |
| Gly Ile Thr Ile Gln Gly Gly Asp Asp Val Phe Lys Glu Asn Tyr Val | |
| 705 710 715 720 | |
| ACA CTA TCA GGT ACC TTT GAT GAG TGC TAT CCA ACA TAT TTG TAT CAA | 2208 |
| Thr Leu Ser Gly Thr Phe Asp Glu Cys Tyr Pro Thr Tyr Leu Tyr Gln | |
| 725 730 735 | |
| AAA ATC GAT GAA TCA AAA TTA AAA GCC TTT ACC CGT TAT CAA TTA AGA | 2256 |
| Lys Ile Asp Glu Ser Lys Leu Lys Ala Phe Thr Arg Tyr Gln Leu Arg | |
| 740 745 750 | |
| GGG TAT ATC GAA GAT AGT CAA GAC TTA GAA ATC TAT TTA ATT CGC TAC | 2304 |
| Gly Tyr Ile Glu Asp Ser Gln Asp Leu Glu Ile Tyr Leu Ile Arg Tyr | |
| 755 760 765 | |
| AAT GCA AAA CAT GAA ACA GTA AAT GTG CCA GGT ACG GGT TCC TTA TGG | 2352 |
| Asn Ala Lys His Glu Thr Val Asn Val Pro Gly Thr Gly Ser Leu Trp | |
| 770 775 780 | |
| CCG CTT TCA GCC CAA AGT CCA ATC GGA AAG TGT GGA GAG CCG AAT CGA | 2400 |
| Pro Leu Ser Ala Gln Ser Pro Ile Gly Lys Cys Gly Glu Pro Asn Arg | |
| 785 790 795 800 | |
| TGC GCG CCA CAC CTT GAA TGG AAT CCT GAC TTA GAT TGT TCG TGT AGG | 2448 |
| Cys Ala Pro His Leu Glu Trp Asn Pro Asp Leu Asp Cys Ser Cys Arg | |
| 805 810 815 | |
| GAT GGA GAA AAG TGT GCC CAT CAT TCG CAT CAT TTC TCC TTA GAC ATT | 2496 |
| Asp Gly Glu Lys Cys Ala His His Ser His His Phe Ser Leu Asp Ile | |
| 820 825 830 | |
| GAT GTA GGA TGT ACA GAC TTA AAT GAG GAC CTA GGT GTA TGG GTG ATC | 2544 |
| Asp Val Gly Cys Thr Asp Leu Asn Glu Asp Leu Gly Val Trp Val Ile | |
| 835 840 845 | |
| TTT AAG ATT AAG ACG CAA GAT GGG CAC GCA AGA CTA GGG AAT CTA GAG | 2592 |



| | | | | | | | | | | | | | | | | | |
|------|-----|------|------|------|------|------|-----|------|------|------|------|-----|------|------|------|--|------|
| Phe | Lys | Ile | Lys | Thr | Gln | Asp | Gly | His | Ala | Arg | Leu | Gly | Asn | Leu | Glu | | |
| 850 | | | | | | 855 | | | | | 860 | | | | | | |
| TTT | CTC | GAA | GAG | AAA | CCA | TTA | GTA | GGA | GAA | GCG | CTA | GCT | CGT | GTG | AAA | | 2640 |
| Phe | Leu | Glu | Glu | Lys | Pro | Leu | Val | Gly | Glu | Ala | Leu | Ala | Arg | Val | Lys | | |
| 865 | | | | | 870 | | | | | 875 | | | | | 880 | | |
| AGA | GCG | GAG | AAA | AAA | TGG | AGA | GAC | AAA | CGT | GAA | AAA | TTG | GAA | TGG | GAA | | 2688 |
| Arg | Ala | Glu | Lys | Lys | Trp | Arg | Asp | Lys | Arg | Glu | Lys | Leu | Glu | Trp | Glu | | |
| | | | | 885 | | | | | 890 | | | | | 895 | | | |
| ACA | AAT | ATC | GTT | TAT | AAA | GAG | GCA | AAA | GAA | TCT | GTA | GAT | GCT | TTA | TTT | | 2736 |
| Thr | Asn | Ile | Val | Tyr | Lys | Glu | Ala | Lys | Glu | Ser | Val | Asp | Ala | Leu | Phe | | |
| | | | 900 | | | | | 905 | | | | | 910 | | | | |
| GTA | AAC | TCT | CAA | TAT | GAT | CAA | TTA | CAA | GCG | GAT | ACG | AAT | ATT | GCC | ATG | | 2784 |
| Val | Asn | Ser | Gln | Tyr | Asp | Gln | Leu | Gln | Ala | Asp | Thr | Asn | Ile | Ala | Met | | |
| | | 915 | | | | | 920 | | | | | 925 | | | | | |
| ATT | CAT | GCG | GCA | GAT | AAA | CGT | GTT | CAT | AGC | ATT | CGA | GAA | GCT | TAT | CTG | | 2832 |
| Ile | His | Ala | Ala | Asp | Lys | Arg | Val | His | Ser | Ile | Arg | Glu | Ala | Tyr | Leu | | |
| | 930 | | | | | 935 | | | | | 940 | | | | | | |
| CCT | GAG | CTG | TCT | GTG | ATT | CCG | GGT | GTC | AAT | GCG | GCT | ATT | TTT | GAA | GAA | | 2880 |
| Pro | Glu | Leu | Ser | Val | Ile | Pro | Gly | Val | Asn | Ala | Ala | Ile | Phe | Glu | Glu | | |
| 945 | | | | | 950 | | | | | 955 | | | | 960 | | | |
| TTA | GAA | GGG | CGT | ATT | TTC | ACT | GCA | TTC | TCC | CTA | TAT | GAT | GCG | AGA | AAT | | 2928 |
| Leu | Glu | Gly | Arg | Ile | Phe | Thr | Ala | Phe | Ser | Leu | Tyr | Asp | Ala | Arg | Asn | | |
| | | | | 965 | | | | | 970 | | | | 975 | | | | |
| GTC | ATT | AAA | AAT | GGT | GAT | TTT | AAT | AAT | GGC | TTA | TCC | TGC | TGG | AAC | GTG | | 2976 |
| Val | Ile | Lys | Asn | Gly | Asp | Phe | Asn | Asn | Gly | Leu | Ser | Cys | Trp | Asn | Val | | |
| | | | 980 | | | | 985 | | | | | | 990 | | | | |
| AAA | GGG | CAT | GTA | GAT | GTA | GAA | GAA | CAA | AAC | AAC | CAA | CGT | TCG | GTC | CTT | | 3024 |
| Lys | Gly | His | Val | Asp | Val | Glu | Glu | Gln | Asn | Asn | Gln | Arg | Ser | Val | Leu | | |
| | | 995 | | | | 1000 | | | | | 1005 | | | | | | |
| GTT | GTT | CCG | GAA | TGG | GAA | GCA | GAA | GTG | TCA | CAA | GAA | GTT | CGT | GTC | TGT | | 3072 |
| Val | Val | Pro | Glu | Trp | Glu | Ala | Glu | Val | Ser | Gln | Glu | Val | Arg | Val | Cys | | |
| | | 1010 | | | | 1015 | | | | | 1020 | | | | | | |
| CCG | GGT | CGT | GGC | TAT | ATC | CTT | CGT | GTC | ACA | GCG | TAC | AAG | GAG | GGA | TAT | | 3120 |
| Pro | Gly | Arg | Gly | Tyr | Ile | Leu | Arg | Val | Thr | Ala | Tyr | Lys | Glu | Gly | Tyr | | |
| 1025 | | | | | 1030 | | | | | 1035 | | | | | 1040 | | |
| GGA | GAA | GGT | TGC | GTA | ACC | ATT | CAT | GAG | ATC | GAG | AAC | AAT | ACA | GAC | GAA | | 3168 |
| Gly | Glu | Gly | Cys | Val | Thr | Ile | His | Glu | Ile | Glu | Asn | Asn | Thr | Asp | Glu | | |
| | | | | 1045 | | | | | 1050 | | | | | 1055 | | | |
| CTG | AAG | TTT | AGC | AAC | TGC | GTA | GAA | GAG | GAA | ATC | TAT | CCA | AAT | AAC | ACG | | 3216 |
| Leu | Lys | Phe | Ser | Asn | Cys | Val | Glu | Glu | Glu | Ile | Tyr | Pro | Asn | Asn | Thr | | |
| | | | 1060 | | | | | 1065 | | | | | 1070 | | | | |
| GTA | ACG | TGT | AAT | GAT | TAT | ACT | GTA | AAT | CAA | GAA | GAA | TAC | GGA | GGT | GCG | | 3264 |
| Val | Thr | Cys | Asn | Asp | Tyr | Thr | Val | Asn | Gln | Glu | Glu | Tyr | Gly | Gly | Ala | | |



| | 1075 | 1080 | 1085 | |
|----|---|------|------|------|
| | TAC ACT TCT CGT AAT CGA GGA TAT AAC GAA GCT CCT TCC GTA CCA GCT | | | 3312 |
| | Tyr Thr Ser Arg Asn Arg Gly Tyr Asn Glu Ala Pro Ser Val Pro Ala | | | |
| | 1090 | 1095 | 1100 | |
| | GAT TAT GCG TCA GTC TAT GAA GAA AAA TCG TAT ACA GAT GGA CGA AGA | | | 3360 |
| | Asp Tyr Ala Ser Val Tyr Glu Glu Lys Ser Tyr Thr Asp Gly Arg Arg | | | |
| | 1105 | 1110 | 1115 | 1120 |
| 5 | GAG AAT CCT TGT GAA TTT AAC AGA GGG TAT AGG GAT TAC ACG CCA CTA | | | 3408 |
| | Glu Asn Pro Cys Glu Phe Asn Arg Gly Tyr Arg Asp Tyr Thr Pro Leu | | | |
| | 1125 | 1130 | 1135 | |
| | CCA GTT GGT TAT GTG ACA AAA GAA TTA GAA TAC TTC CCA GAA ACC GAT | | | 3456 |
| | Pro Val Gly Tyr Val Thr Lys Glu Leu Glu Tyr Phe Pro Glu Thr Asp | | | |
| | 1140 | 1145 | 1150 | |
| 10 | AAG GTA TGG ATT GAG ATT GGA GAA ACG GAA GGA ACA TTT ATC GTG GAC | | | 3504 |
| | Lys Val Trp Ile Glu Ile Gly Glu Thr Glu Gly Thr Phe Ile Val Asp | | | |
| | 1155 | 1160 | 1165 | |
| | AGC GTG GAA TTA CTC CTT ATG GAG GAA | | | 3531 |
| | Ser Val Glu Leu Leu Leu Met Glu Glu | | | |
| | 1170 | 1175 | | |

(2) SEQ ID NO : 12 信息:

15

(i) 序列特征:

(A) 长度: 1177 个氨基酸

(B) 类型: 氨基酸

(D) 拓扑学: 线性

(ii) 分子类型: 蛋白

20

(xi) 序列描述: SEQ ID NO : 12 :

Met Asp Asn Asn Pro Asn Ile Asn Glu Cys Ile Pro Tyr Asn Cys Leu
 1 5 10 15

Ser Asn Pro Glu Val Glu Val Leu Gly Gly Glu Arg Ile Glu Thr Gly
 20 25 30

Tyr Thr Pro Ile Asp Ile Ser Leu Ser Leu Thr Gln Phe Leu Leu Ser
 35 40 45

Glu Phe Val Pro Gly Ala Gly Phe Val Leu Gly Leu Val Asp Ile Ile
 50 55 60

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

Glu Gln Leu Ile Asn Gln Arg Ile Glu Glu Phe Ala Arg Asn Gln Ala
 85 90 95

Ile Ser Arg Leu Glu Gly Leu Ser Asn Leu Tyr Gln Ile Tyr Ala Glu
 100 105 110



Ser Phe Arg Glu Trp Glu Ala Asp Pro Thr Asn Pro Ala Leu Arg Glu
115 120 125

Glu Met Arg Ile Gln Phe Asn Asp Met Asn Ser Ala Leu Thr Thr Ala
130 135 140

Ile Pro Leu Phe Ala Val Gln Asn Tyr Gln Val Pro Leu Leu Ser Val
145 150 155 160

Tyr Val Gln Ala Ala Asn Leu His Leu Ser Val Leu Arg Asp Val Ser
165 170 175

Val Phe Gly Gln Arg Trp Gly Phe Asp Ala Ala Thr Ile Asn Ser Arg
180 185 190

Tyr Asn Asp Leu Thr Arg Leu Ile Gly Asn Tyr Thr Asp Tyr Ala Val
195 200 205

Arg Trp Tyr Asn Thr Gly Leu Glu Arg Val Trp Gly Pro Asp Ser Arg
210 215 220

Asp Trp Val Arg Tyr Asn Gln Phe Arg Arg Glu Leu Thr Leu Thr Val
225 230 235 240

Leu Asp Ile Val Ala Leu Phe Pro Asn Tyr Asp Ser Arg Arg Tyr Pro
245 250 255

Ile Arg Thr Val Ser Gln Leu Thr Arg Glu Ile Tyr Thr Asn Pro Val
260 265 270

Leu Glu Asn Phe Asp Gly Ser Phe Arg Gly Ser Ala Gln Gly Ile Glu
275 280 285

Arg Ser Ile Arg Ser Pro His Leu Met Asp Ile Leu Asn Ser Ile Thr
290 295 300

Ile Tyr Thr Asp Ala His Arg Gly Tyr Tyr Tyr Trp Ser Gly His Gln
305 310 315 320

Ile Met Ala Ser Pro Val Gly Phe Ser Gly Pro Glu Phe Thr Phe Pro
325 330 335

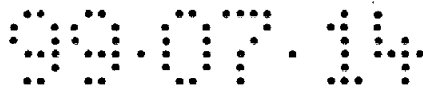
Leu Tyr Gly Thr Met Gly Asn Ala Ala Pro Gln Gln Arg Ile Val Ala
340 345 350

Gln Leu Gly Gln Gly Val Tyr Arg Thr Leu Ser Ser Thr Leu Tyr Arg
355 360 365

Arg Pro Phe Asn Ile Gly Ile Asn Asn Gln Gln Leu Ser Val Leu Asp
370 375 380

Gly Thr Glu Phe Ala Tyr Gly Thr Ser Ser Asn Leu Pro Ser Ala Val
385 390 395 400

Tyr Arg Lys Ser Gly Thr Val Asp Ser Leu Asp Glu Ile Pro Pro Gln
405 410 415



Asn Asn Asn Val Pro Pro Arg Gln Gly Phe Ser His Arg Leu Ser His
420 425 430

Val Ser Met Phe Arg Ser Gly Phe Ser Asn Ser Ser Val Ser Ile Ile
435 440 445

Arg Ala Pro Met Phe Ser Trp Thr His Arg Ser Ala Thr Pro Thr Asn
450 455 460

Thr Ile Asp Pro Glu Arg Ile Thr Gln Ile Pro Leu Val Lys Ala His
465 470 475 480

Thr Leu Gln Ser Gly Thr Thr Val Val Arg Gly Pro Gly Phe Thr Gly
485 490 495

Gly Asp Ile Leu Arg Arg Thr Ser Gly Gly Pro Phe Ala Tyr Thr Ile
500 505 510

Val Asn Ile Asn Gly Gln Leu Pro Gln Arg Tyr Arg Ala Arg Ile Arg
515 520 525

Tyr Ala Ser Thr Thr Asn Leu Arg Ile Tyr Val Thr Val Ala Gly Glu
530 535 540

Arg Ile Phe Ala Gly Gln Phe Asn Lys Thr Met Asp Thr Gly Asp Pro
545 550 555 560

Leu Thr Phe Gln Ser Phe Ser Tyr Ala Thr Ile Asn Thr Ala Phe Thr
565 570 575

Phe Pro Met Ser Gln Ser Ser Phe Thr Val Gly Ala Asp Thr Phe Ser
580 585 590

Ser Gly Asn Glu Val Tyr Ile Asp Arg Phe Glu Leu Ile Pro Val Thr
595 600 605

Ala Thr Leu Glu Ala Glu Tyr Asn Leu Glu Arg Ala Gln Lys Ala Val
610 615 620

Asn Ala Leu Phe Thr Ser Thr Asn Gln Leu Gly Leu Lys Thr Asn Val
625 630 635 640

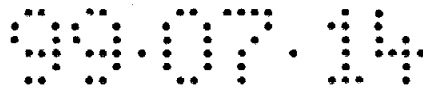
Thr Asp Tyr His Ile Asp Gln Val Ser Asn Leu Val Thr Tyr Leu Ser
645 650 655

Asp Glu Phe Cys Leu Asp Glu Lys Arg Glu Leu Ser Glu Lys Val Lys
660 665 670

His Ala Lys Arg Leu Ser Asp Glu Arg Asn Leu Leu Gln Asp Ser Asn
675 680 685

Phe Lys Asp Ile Asn Arg Gln Pro Glu Arg Gly Trp Gly Gly Ser Thr
690 695 700

Gly Ile Thr Ile Gln Gly Gly Asp Asp Val Phe Lys Glu Asn Tyr Val
705 710 715 720



Thr Leu Ser Gly Thr Phe Asp Glu Cys Tyr Pro Thr Tyr Leu Tyr Gln
725 730 735

Lys Ile Asp Glu Ser Lys Leu Lys Ala Phe Thr Arg Tyr Gln Leu Arg
740 745 750

Gly Tyr Ile Glu Asp Ser Gln Asp Leu Glu Ile Tyr Leu Ile Arg Tyr
755 760 765

Asn Ala Lys His Glu Thr Val Asn Val Pro Gly Thr Gly Ser Leu Trp
770 775 780

Pro Leu Ser Ala Gln Ser Pro Ile Gly Lys Cys Gly Glu Pro Asn Arg
785 790 795 800

Cys Ala Pro His Leu Glu Trp Asn Pro Asp Leu Asp Cys Ser Cys Arg
805 810 815

Asp Gly Glu Lys Cys Ala His His Ser His His Phe Ser Leu Asp Ile
820 825 830

Asp Val Gly Cys Thr Asp Leu Asn Glu Asp Leu Gly Val Trp Val Ile
835 840 845

Phe Lys Ile Lys Thr Gln Asp Gly His Ala Arg Leu Gly Asn Leu Glu
850 855 860

Phe Leu Glu Glu Lys Pro Leu Val Gly Glu Ala Leu Ala Arg Val Lys
865 870 875 880

Arg Ala Glu Lys Lys Trp Arg Asp Lys Arg Glu Lys Leu Glu Trp Glu
885 890 895

Thr Asn Ile Val Tyr Lys Glu Ala Lys Glu Ser Val Asp Ala Leu Phe
900 905 910

Val Asn Ser Gln Tyr Asp Gln Leu Gln Ala Asp Thr Asn Ile Ala Met
915 920 925

Ile His Ala Ala Asp Lys Arg Val His Ser Ile Arg Glu Ala Tyr Leu
930 935 940

Pro Glu Leu Ser Val Ile Pro Gly Val Asn Ala Ala Ile Phe Glu Glu
945 950 955 960

Leu Glu Gly Arg Ile Phe Thr Ala Phe Ser Leu Tyr Asp Ala Arg Asn
965 970 975

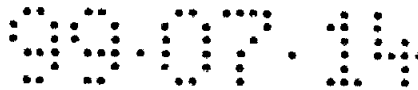
Val Ile Lys Asn Gly Asp Phe Asn Asn Gly Leu Ser Cys Trp Asn Val
980 985 990

Lys Gly His Val Asp Val Glu Glu Gln Asn Asn Gln Arg Ser Val Leu
995 1000 1005

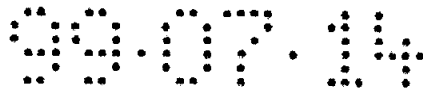
Val Val Pro Glu Trp Glu Ala Glu Val Ser Gln Glu Val Arg Val Cys
1010 1015 1020



| | |
|---|-----|
| AGT AAC CCT GAA GTA GAA GTA TTA GGT GGA GAA AGA ATA GAA ACT GGT | 96 |
| Ser Asn Pro Glu Val Glu Val Leu Gly Gly Glu Arg Ile Glu Thr Gly | |
| 20 25 30 | |
| TAC ACC CCA ATC GAT ATT TCC TTG TCG CTA ACG CAA TTT CTT TTG AGT | 144 |
| Tyr Thr Pro Ile Asp Ile Ser Leu Ser Leu Thr Gln Phe Leu Leu Ser | |
| 35 40 45 | |
| GAA TTT GTT CCC GGT GCT GGA TTT GTG TTA GGA CTA GTT GAT ATA ATA | 192 |
| Glu Phe Val Pro Gly Ala Gly Phe Val Leu Gly Leu Val Asp Ile Ile | |
| 50 55 60 | |
| TGG GGA ATT TTT GGT CCC TCT CAA TGG GAC GCA TTT CTT GTA CAA ATT | 240 |
| Trp Gly Ile Phe Gly Pro Ser Gln Trp Asp Ala Phe Leu Val Gln Ile | |
| 65 70 75 80 | |
| GAA CAG TTA ATT AAC CAA AGA ATA GAA GAA TTC GCT AGG AAC CAA GCC | 288 |
| Glu Gln Leu Ile Asn Gln Arg Ile Glu Glu Phe Ala Arg Asn Gln Ala | |
| 85 90 95 | |
| ATT TCT AGA TTA GAA GGA CTA AGC AAT CTT TAT CAA ATT TAC GCA GAA | 336 |
| Ile Ser Arg Leu Glu Gly Leu Ser Asn Leu Tyr Gln Ile Tyr Ala Glu | |
| 100 105 110 | |
| TCT TTT AGA GAG TGG GAA GCA GAT CCT ACT AAT CCA GCA TTA AGA GAA | 384 |
| Ser Phe Arg Glu Trp Glu Ala Asp Pro Thr Asn Pro Ala Leu Arg Glu | |
| 115 120 125 | |
| GAG ATG CGT ATT CAA TTC AAT GAC ATG AAC AGT GCC CTT ACA ACC GCT | 432 |
| Glu Met Arg Ile Gln Phe Asn Asp Met Asn Ser Ala Leu Thr Thr Ala | |
| 130 135 140 | |
| ATT CCT CTT TTT GCA GTT CAA AAT TAT CAA GTT CCT CTT TTA TCA GTA | 480 |
| Ile Pro Leu Phe Ala Val Gln Asn Tyr Gln Val Pro Leu Leu Ser Val | |
| 145 150 155 160 | |
| TAT GTT CAA GCT GCA AAT TTA CAT TTA TCA GTT TTG AGA GAT GTT TCA | 528 |
| Tyr Val Gln Ala Ala Asn Leu His Leu Ser Val Leu Arg Asp Val Ser | |
| 165 170 175 | |
| GTG TTT GGA CAA AGG TGG GGA TTT GAT GCC GCG ACT ATC AAT AGT CGT | 576 |
| Val Phe Gly Gln Arg Trp Gly Phe Asp Ala Ala Thr Ile Asn Ser Arg | |
| 180 185 190 | |
| TAT AAT GAT TTA ACT AGG CTT ATT GGC AAC TAT ACA GAT CAT GCT GTA | 624 |
| Tyr Asn Asp Leu Thr Arg Leu Ile Gly Asn Tyr Thr Asp His Ala Val | |
| 195 200 205 | |
| CGC TGG TAC AAT ACG GGA TTA GAG CGT GTA TGG GGA CCG GAT TTT AGA | 672 |
| Arg Trp Tyr Asn Thr Gly Leu Glu Arg Val Trp Gly Pro Asp Ser Arg | |
| 210 215 220 | |
| GAT TGG ATA AGA TAT AAT CAA TTT AGA AGA GAA TTA ACA CTA ACT GTA | 720 |
| Asp Trp Ile Arg Tyr Asn Gln Phe Arg Arg Glu Leu Thr Leu Thr Val | |
| 225 230 235 240 | |



| | |
|---|------|
| TTA GAT ATC GTT TCT CTA TTT CCG AAC TAT GAT AGT AGA ACG TAT CCA Leu Asp Ile Val Ser Leu Phe Pro Asn Tyr Asp Ser Arg Thr Tyr Pro 245 250 255 | 768 |
| ATT CGA ACA GTT TCC CAA TTA ACA AGA GAA ATT TAT ACA AAC CCA GTA Ile Arg Thr Val Ser Gln Leu Thr Arg Glu Ile Tyr Thr Asn Pro Val 260 265 270 | 816 |
| TTA GAA AAT TTT GAT GGT AGT TTT CGA GGC TCG GCT CAG GGC ATA GAA Leu Glu Asn Phe Asp Gly Ser Phe Arg Gly Ser Ala Gln Gly Ile Glu 275 280 285 | 864 |
| GGA AGT ATT AGG AGT CCA CAT TTG ATG GAT ATA CTT AAC AGT ATA ACC Gly Ser Ile Arg Ser Pro His Leu Met Asp Ile Leu Asn Ser Ile Thr 290 295 300 | 912 |
| ATC TAT ACG GAT GCT CAT AGA GGA GAA TAT TAT TGG TCA GGG CAT CAA Ile Tyr Thr Asp Ala His Arg Gly Glu Tyr Tyr Trp Ser Gly His Gln 305 310 315 320 | 960 |
| ATA ATG GCT TCT CCT GTA GGG TTT TCG GGG CCA GAA TTC ACT TTT CCG Ile Met Ala Ser Pro Val Gly Phe Ser Gly Pro Glu Phe Thr Phe Pro 325 330 335 | 1008 |
| CTA TAT GGA ACT ATG GGA AAT GCA GCT CCA CAA CAA CGT ATT GTT GCT Leu Tyr Gly Thr Met Gly Asn Ala Ala Pro Gln Gln Arg Ile Val Ala 340 345 350 | 1056 |
| CAA CTA GGT CAG GGC GTG TAT AGA ACA TTA TCG TCC ACT TTA TAT AGA Gln Leu Gly Gln Gly Val Tyr Arg Thr Leu Ser Ser Thr Leu Tyr Arg 355 360 365 | 1104 |
| AGA CCT TTT AAT ATA GGG ATA AAT AAT CAA CAA CTA TCT GTT CTT GAC Arg Pro Phe Asn Ile Gly Ile Asn Asn Gln Gln Leu Ser Val Leu Asp 370 375 380 | 1152 |
| GGG ACA GAA TTT GCT TAT GGA ACC TCC TCA AAT TTG CCA TCC GCT GTA Gly Thr Glu Phe Ala Tyr Gly Thr Ser Ser Asn Leu Pro Ser Ala Val 385 390 395 400 | 1200 |
| TAC AGA AAA AGC GGA ACG GTA GAT TCG CTG GAT GAA ATA CCG CCA CAG Tyr Arg Lys Ser Gly Thr Val Asp Ser Leu Asp Glu Ile Pro Pro Gln 405 410 415 | 1248 |
| AAT AAC AAC GTG CCA CCT AGG CAA GGA TTT AGT CAT CGA TTA AGC CAT Asn Asn Asn Val Pro Pro Arg Gln Gly Phe Ser His Arg Leu Ser His 420 425 430 | 1296 |
| GTT TCA ATG TTT CGT TCA GGC TTT AGT AAT AGT AGT GTA AGT ATA ATA Val Ser Met Phe Arg Ser Gly Phe Ser Asn Ser Ser Val Ser Ile Ile 435 440 445 | 1344 |
| AGA GCT CCA ATG TTT TCT TGG ACG CAC CGT AGT GCA ACC CCT ACA AAT Arg Ala Pro Met Phe Ser Trp Thr His Arg Ser Ala Thr Pro Thr Asn 450 455 460 | 1392 |



ACA ATT GAT CCG GAG AGG ATT ACT CAA ATA CCA TTG GTA AAA GCA CAT 1440
Thr Ile Asp Pro Glu Arg Ile Thr Gln Ile Pro Leu Val Lys Ala His
465 470 475 480

ACA CTT CAG TCA GGT ACT ACT GTT GTA AGA GGG CCC GGG TTT ACG GGA 1488
Thr Leu Gln Ser Gly Thr Thr Val Val Arg Gly Pro Gly Phe Thr Gly
485 490 495

GGA GAT ATT CTT CGA CGA ACA AGT GGA GGA CCA TTT GCT TAT ACT ATT 1536
Gly Asp Ile Leu Arg Arg Thr Ser Gly Gly Pro Phe Ala Tyr Thr Ile
500 505 510

GTT AAT ATA AAT GGG CAA TTA CCC CAA AGG TAT CGT GCA AGA ATA CGC 1584
Val Asn Ile Asn Gly Gln Leu Pro Gln Arg Tyr Arg Ala Arg Ile Arg
515 520 525

TAT GCC TCT ACT ACA AAT CTA AGA ATT TAC GTA ACG GTT GCA GGT GAA 1632
Tyr Ala Ser Thr Thr Asn Leu Arg Ile Tyr Val Thr Val Ala Gly Glu
530 535 540

CGG ATT TTT GCT GGT CAA TTT AAC AAA ACA ATG GAT ACC GGT GAC CCA 1680
Arg Ile Phe Ala Gly Gln Phe Asn Lys Thr Met Asp Thr Gly Asp Pro
545 550 555 560

TTA ACA TTC CAA TCT TTT AGT TAC GCA ACT ATT AAT ACA GCT TTT ACA 1728
Leu Thr Phe Gln Ser Phe Ser Tyr Ala Thr Ile Asn Thr Ala Phe Thr
565 570 575

TTC CCA ATG AGC CAG AGT AGT TTC ACA GTA GGT GCT GAT ACT TTT AGT 1776
Phe Pro Met Ser Gln Ser Ser Phe Thr Val Gly Ala Asp Thr Phe Ser
580 585 590

TCA GGG AAT GAA GTT TAT ATA GAC AGA TTT GAA TTG ATT CCA GTT ACT 1824
Ser Gly Asn Glu Val Tyr Ile Asp Arg Phe Glu Leu Ile Pro Val Thr
595 600 605

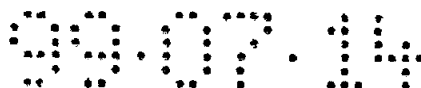
GCA ACA TTT GAA GCA GAA TAT GAT TTA GAA AGA GCA CAA AAG GCG GTG 1872
Ala Thr Phe Glu Ala Glu Tyr Asp Leu Glu Arg Ala Gln Lys Ala Val
610 615 620

AAT GCG CTG TTT ACT TCT ATA AAC CAA ATA GGG ATA AAA ACA GAT GTG 1920
Asn Ala Leu Phe Thr Ser Ile Asn Gln Ile Gly Ile Lys Thr Asp Val
625 630 635 640

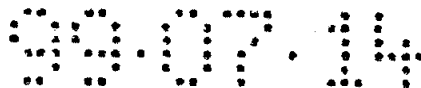
ACG GAT TAT CAT ATT GAT CAA GTA TCC AAT TTA GTG GAT TGT TTA TCA 1968
Thr Asp Tyr His Ile Asp Gln Val Ser Asn Leu Val Asp Cys Leu Ser
645 650 655

GAT GAA TTT TGT CTG GAT GAA AAG CGA GAA TTG TCC GAG AAA GTC AAA 2016
Asp Glu Phe Cys Leu Asp Glu Lys Arg Glu Leu Ser Glu Lys Val Lys
660 665 670

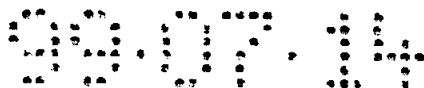
CAT GCG AAG CGA CTC AGT GAT GAG CGG AAT TTA CTT CAA GAT CCA AAC 2064
His Ala Lys Arg Leu Ser Asp Glu Arg Asn Leu Leu Gln Asp Pro Asn
675 680 685



| | |
|---|------|
| TTC AAA GGC ATC AAT AGG CAA CTA GAC CGT GGT TGG AGA GGA AGT ACG Phe Lys Gly Ile Asn Arg Gln Leu Asp Arg Gly Trp Arg Gly Ser Thr 690 695 700 | 2112 |
| GAT ATT ACC ATC CAA AGA GGA GAT GAC GTA TTC AAA GAA AAT TAT GTC Asp Ile Thr Ile Gln Arg Gly Asp Asp Val Phe Lys Glu Asn Tyr Val 705 710 715 720 | 2160 |
| ACA CTA CCA GGT ACC TTT GAT GAG TGC TAT CCA ACA TAT TTG TAT CAA Thr Leu Pro Gly Thr Phe Asp Glu Cys Tyr Pro Thr Tyr Leu Tyr Gln 725 730 735 | 2208 |
| AAA ATC GAT GAA TCA AAA TTA AAA GCC TTT ACC CGT TAT CAA TTA AGA Lys Ile Asp Glu Ser Lys Leu Lys Ala Phe Thr Arg Tyr Gln Leu Arg 740 745 750 | 2256 |
| GGG TAT ATC GAA GAT AGT CAA GAC TTA GAA ATC TAT TTA ATT CGC TAC Gly Tyr Ile Glu Asp Ser Gln Asp Leu Glu Ile Tyr Leu Ile Arg Tyr 755 760 765 | 2304 |
| AAT GCA AAA CAT GAA ACA GTA AAT GTG CCA GGT ACG GGT TCC TTA TGG Asn Ala Lys His Glu Thr Val Asn Val Pro Gly Thr Gly Ser Leu Trp 770 775 780 | 2352 |
| CCG CTT TCA GCC CAA AGT CCA ATC GGA AAG TGT GGA GAG CCG AAT CGA Pro Leu Ser Ala Gln Ser Pro Ile Gly Lys Cys Gly Glu Pro Asn Arg 785 790 795 800 | 2400 |
| TGC GCG CCA CAC CTT GAA TGG AAT CCT GAC TTA GAT TGT TCG TGT AGG Cys Ala Pro His Leu Glu Trp Asn Pro Asp Leu Asp Cys Ser Cys Arg 805 810 815 | 2448 |
| GAT GGA GAA AAG TGT GCC CAT CAT TCG CAT CAT TTC TCC TTA GAC ATT Asp Gly Glu Lys Cys Ala His His Ser His His Phe Ser Leu Asp Ile 820 825 830 | 2496 |
| GAT GTA GGA TGT ACA GAC TTA AAT GAG GAC CTA GGT GTA TGG GTG ATC Asp Val Gly Cys Thr Asp Leu Asn Glu Asp Leu Gly Val Trp Val Ile 835 840 845 | 2544 |
| TTT AAG ATT AAG ACG CAA GAT GGG CAC GCA AGA CTA GGG AAT CTA GAG Phe Lys Ile Lys Thr Gln Asp Gly His Ala Arg Leu Gly Asn Leu Glu 850 855 860 | 2592 |
| TTT CTC GAA GAG AAA CCA TTA GTA GGA GAA GCG CTA GCT CGT GTG AAA Phe Leu Glu Glu Lys Pro Leu Val Gly Glu Ala Leu Ala Arg Val Lys 865 870 875 880 | 2640 |
| AGA GCG GAG AAA AAA TGG AGA GAC AAA CGT GAA AAA TTG GAA TGG GAA Arg Ala Glu Lys Lys Trp Arg Asp Lys Arg Glu Lys Leu Glu Trp Glu 885 890 895 | 2688 |
| ACA AAT ATC GTT TAT AAA GAG GCA AAA GAA TCT GTA GAT GCT TTA TTT Thr Asn Ile Val Tyr Lys Glu Ala Lys Glu Ser Val Asp Ala Leu Phe 900 905 910 | 2736 |



| | |
|---|------|
| GTA AAC TCT CAA TAT GAT CAA TTA CAA GCG GAT ACG AAT ATT GCC ATG Val Asn Ser Gln Tyr Asp Gln Leu Gln Ala Asp Thr Asn Ile Ala Met 915 920 925 | 2784 |
| ATT CAT GCG GCA GAT AAA CGT GTT CAT AGC ATT CGA GAA GCT TAT CTG Ile His Ala Ala Asp Lys Arg Val His Ser Ile Arg Glu Ala Tyr Leu 930 935 940 | 2832 |
| CCT GAG CTG TCT GTG ATT CCG GGT GTC AAT GCG GCT ATT TTT GAA GAA Pro Glu Leu Ser Val Ile Pro Gly Val Asn Ala Ala Ile Phe Glu Glu 945 950 955 960 | 2880 |
| TTA GAA GGG CGT ATT TTC ACT GCA TTC TCC CTA TAT GAT GCG AGA AAT Leu Glu Gly Arg Ile Phe Thr Ala Phe Ser Leu Tyr Asp Ala Arg Asn 965 970 975 | 2928 |
| GTC ATT AAA AAT GGT GAT TTT AAT AAT GGC TTA TCC TGC TGG AAC GTG Val Ile Lys Asn Gly Asp Phe Asn Asn Gly Leu Ser Cys Trp Asn Val 980 985 990 | 2976 |
| AAA GGG CAT GTA GAT GTA GAA GAA CAA AAC AAC CAA CGT TCG GTC CTT Lys Gly His Val Asp Val Glu Glu Gln Asn Asn Gln Arg Ser Val Leu 995 1000 1005 | 3024 |
| GTT GTT CCG GAA TGG GAA GCA GAA GTG TCA CAA GAA GTT CGT GTC TGT Val Val Pro Glu Trp Glu Ala Glu Val Ser Gln Glu Val Arg Val Cys 1010 1015 1020 | 3072 |
| CCG GGT CGT GGC TAT ATC CTT CGT GTC ACA GCG TAC AAG GAG GGA TAT Pro Gly Arg Gly Tyr Ile Leu Arg Val Thr Ala Tyr Lys Glu Gly Tyr 1025 1030 1035 1040 | 3120 |
| GGA GAA GGT TGC GTA ACC ATT CAT GAG ATC GAG AAC AAT ACA GAC GAA Gly Glu Gly Cys Val Thr Ile His Glu Ile Glu Asn Asn Thr Asp Glu 1045 1050 1055 | 3168 |
| CTG AAG TTT AGC AAC TGC GTA GAA GAG GAA ATC TAT CCA AAT AAC ACG Leu Lys Phe Ser Asn Cys Val Glu Glu Glu Ile Tyr Pro Asn Asn Thr 1060 1065 1070 | 3216 |
| GTA ACG TGT AAT GAT TAT ACT GTA AAT CAA GAA GAA TAC GGA GGT GCG Val Thr Cys Asn Asp Tyr Thr Val Asn Gln Glu Glu Tyr Gly Gly Ala 1075 1080 1085 | 3264 |
| TAC ACT TCT CGT AAT CGA GGA TAT AAC GAA GCT CCT TCC GTA CCA GCT Tyr Thr Ser Arg Asn Arg Gly Tyr Asn Glu Ala Pro Ser Val Pro Ala 1090 1095 1100 | 3312 |
| GAT TAT GCG TCA GTC TAT GAA GAA AAA TCG TAT ACA GAT GGA CGA AGA Asp Tyr Ala Ser Val Tyr Glu Glu Lys Ser Tyr Thr Asp Gly Arg Arg 1105 1110 1115 1120 | 3360 |
| GAG AAT CCT TGT GAA TTT AAC AGA GGG TAT AGG GAT TAC ACG CCA CTA Glu Asn Pro Cys Glu Phe Asn Arg Gly Tyr Arg Asp Tyr Thr Pro Leu 1125 1130 1135 | 3408 |



| | |
|---|------|
| CCA GTT GGT TAT GTG ACA AAA GAA TTA GAA TAC TTC CCA GAA ACC GAT | 3456 |
| Pro Val Gly Tyr Val Thr Lys Glu Leu Glu Tyr Phe Pro Glu Thr Asp | |
| 1140 1145 1150 | |
| AAG GTA TGG ATT GAG ATT GGA GAA ACG GAA GGA ACA TTT ATC GTG GAC | 3504 |
| Lys Val Trp Ile Glu Ile Gly Glu Thr Glu Gly Thr Phe Ile Val Asp | |
| 1155 1160 1165 | |
| AGC GTG GAA TTA CTC CTT ATG GAG GAA | 3531 |
| Ser Val Glu Leu Leu Leu Met Glu Glu | |
| 1170 1175 | |

5

(2) SEQ ID NO: 14 信息:

(i) 序列特征:

10

(A) 长度: 1177 个氨基酸

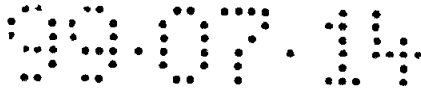
(B) 类型: 氨基酸

(D) 拓扑学: 线性

(ii) 分子类型: 蛋白

(xi) 序列描述: SEQ ID NO: 14:

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



Tyr Val Gln Ala Ala Asn Leu His Leu Ser Val Leu Arg Asp Val Ser
165 170 175

Val Phe Gly Gln Arg Trp Gly Phe Asp Ala Ala Thr Ile Asn Ser Arg
180 185 190

Tyr Asn Asp Leu Thr Arg Leu Ile Gly Asn Tyr Thr Asp His Ala Val
195 200 205

Arg Trp Tyr Asn Thr Gly Leu Glu Arg Val Trp Gly Pro Asp Ser Arg
210 215 220

Asp Trp Ile Arg Tyr Asn Gln Phe Arg Arg Glu Leu Thr Leu Thr Val
225 230 235 240

Leu Asp Ile Val Ser Leu Phe Pro Asn Tyr Asp Ser Arg Thr Tyr Pro
245 250 255

Ile Arg Thr Val Ser Gln Leu Thr Arg Glu Ile Tyr Thr Asn Pro Val
260 265 270

Leu Glu Asn Phe Asp Gly Ser Phe Arg Gly Ser Ala Gln Gly Ile Glu
275 280 285

Gly Ser Ile Arg Ser Pro His Leu Met Asp Ile Leu Asn Ser Ile Thr
290 295 300

Ile Tyr Thr Asp Ala His Arg Gly Glu Tyr Tyr Trp Ser Gly His Gln
305 310 315 320

Ile Met Ala Ser Pro Val Gly Phe Ser Gly Pro Glu Phe Thr Phe Pro
325 330 335

Leu Tyr Gly Thr Met Gly Asn Ala Ala Pro Gln Gln Arg Ile Val Ala
340 345 350

Gln Leu Gly Gln Gly Val Tyr Arg Thr Leu Ser Ser Thr Leu Tyr Arg
355 360 365

Arg Pro Phe Asn Ile Gly Ile Asn Asn Gln Gln Leu Ser Val Leu Asp
370 375 380

Gly Thr Glu Phe Ala Tyr Gly Thr Ser Ser Asn Leu Pro Ser Ala Val
385 390 395 400

Tyr Arg Lys Ser Gly Thr Val Asp Ser Leu Asp Glu Ile Pro Pro Gln
405 410 415

Asn Asn Asn Val Pro Pro Arg Gln Gly Phe Ser His Arg Leu Ser His
420 425 430

Val Ser Met Phe Arg Ser Gly Phe Ser Asn Ser Ser Val Ser Ile Ile
435 440 445

Arg Ala Pro Met Phe Ser Trp Thr His Arg Ser Ala Thr Pro Thr Asn
450 455 460



Thr Ile Asp Pro Glu Arg Ile Thr Gln Ile Pro Leu Val Lys Ala His
465 470 475 480

Thr Leu Gln Ser Gly Thr Thr Val Val Arg Gly Pro Gly Phe Thr Gly
485 490 495

Gly Asp Ile Leu Arg Arg Thr Ser Gly Gly Pro Phe Ala Tyr Thr Ile
500 505 510

Val Asn Ile Asn Gly Gln Leu Pro Gln Arg Tyr Arg Ala Arg Ile Arg
515 520 525

Tyr Ala Ser Thr Thr Asn Leu Arg Ile Tyr Val Thr Val Ala Gly Glu
530 535 540

Arg Ile Phe Ala Gly Gln Phe Asn Lys Thr Met Asp Thr Gly Asp Pro
545 550 555 560

Leu Thr Phe Gln Ser Phe Ser Tyr Ala Thr Ile Asn Thr Ala Phe Thr
565 570 575

Phe Pro Met Ser Gln Ser Ser Phe Thr Val Gly Ala Asp Thr Phe Ser
580 585 590

Ser Gly Asn Glu Val Tyr Ile Asp Arg Phe Glu Leu Ile Pro Val Thr
595 600 605

Ala Thr Phe Glu Ala Glu Tyr Asp Leu Glu Arg Ala Gln Lys Ala Val
610 615 620

Asn Ala Leu Phe Thr Ser Ile Asn Gln Ile Gly Ile Lys Thr Asp Val
625 630 635 640

Thr Asp Tyr His Ile Asp Gln Val Ser Asn Leu Val Asp Cys Leu Ser
645 650 655

Asp Glu Phe Cys Leu Asp Glu Lys Arg Glu Leu Ser Glu Lys Val Lys
660 665 670

His Ala Lys Arg Leu Ser Asp Glu Arg Asn Leu Leu Gln Asp Pro Asn
675 680 685

Phe Lys Gly Ile Asn Arg Gln Leu Asp Arg Gly Trp Arg Gly Ser Thr
690 695 700

Asp Ile Thr Ile Gln Arg Gly Asp Asp Val Phe Lys Glu Asn Tyr Val
705 710 715 720

Thr Leu Pro Gly Thr Phe Asp Glu Cys Tyr Pro Thr Tyr Leu Tyr Gln
725 730 735

Lys Ile Asp Glu Ser Lys Leu Lys Ala Phe Thr Arg Tyr Gln Leu Arg
740 745 750

09.07.14

Gly Tyr Ile Glu Asp Ser Gln Asp Leu Glu Ile Tyr Leu Ile Arg Tyr
755 760 765

Asn Ala Lys His Glu Thr Val Asn Val Pro Gly Thr Gly Ser Leu Trp
770 775 780

Pro Leu Ser Ala Gln Ser Pro Ile Gly Lys Cys Gly Glu Pro Asn Arg
785 790 795 800

Cys Ala Pro His Leu Glu Trp Asn Pro Asp Leu Asp Cys Ser Cys Arg
805 810 815

Asp Gly Glu Lys Cys Ala His His Ser His His Phe Ser Leu Asp Ile
820 825 830

Asp Val Gly Cys Thr Asp Leu Asn Glu Asp Leu Gly Val Trp Val Ile
835 840 845

Phe Lys Ile Lys Thr Gln Asp Gly His Ala Arg Leu Gly Asn Leu Glu
850 855 860

Phe Leu Glu Glu Lys Pro Leu Val Gly Glu Ala Leu Ala Arg Val Lys
865 870 875 880

Arg Ala Glu Lys Lys Trp Arg Asp Lys Arg Glu Lys Leu Glu Trp Glu
885 890 895

Thr Asn Ile Val Tyr Lys Glu Ala Lys Glu Ser Val Asp Ala Leu Phe
900 905 910

Val Asn Ser Gln Tyr Asp Gln Leu Gln Ala Asp Thr Asn Ile Ala Met
915 920 925

Ile His Ala Ala Asp Lys Arg Val His Ser Ile Arg Glu Ala Tyr Leu
930 935 940

Pro Glu Leu Ser Val Ile Pro Gly Val Asn Ala Ala Ile Phe Glu Glu
945 950 955 960

Leu Glu Gly Arg Ile Phe Thr Ala Phe Ser Leu Tyr Asp Ala Arg Asn
965 970 975

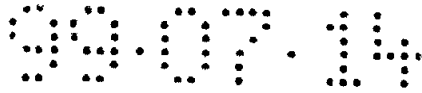
Val Ile Lys Asn Gly Asp Phe Asn Asn Gly Leu Ser Cys Trp Asn Val
980 985 990

Lys Gly His Val Asp Val Glu Glu Gln Asn Asn Gln Arg Ser Val Leu
995 1000 1005

Val Val Pro Glu Trp Glu Ala Glu Val Ser Gln Glu Val Arg Val Cys
1010 1015 1020

Pro Gly Arg Gly Tyr Ile Leu Arg Val Thr Ala Tyr Lys Glu Gly Tyr
1025 1030 1035 1040

Gly Glu Gly Cys Val Thr Ile His Glu Ile Glu Asn Asn Thr Asp Glu
1045 1050 1055



Leu Lys Phe Ser Asn Cys Val Glu Glu Glu Ile Tyr Pro Asn Asn Thr
 1060 1065 1070

Val Thr Cys Asn Asp Tyr Thr Val Asn Gln Glu Glu Tyr Gly Gly Ala
 1075 1080 1085

5 Tyr Thr Ser Arg Asn Arg Gly Tyr Asn Glu Ala Pro Ser Val Pro Ala
 1090 1095 1100

Asp Tyr Ala Ser Val Tyr Glu Glu Lys Ser Tyr Thr Asp Gly Arg Arg
 1105 1110 1115 1120

Glu Asn Pro Cys Glu Phe Asn Arg Gly Tyr Arg Asp Tyr Thr Pro Leu
 1125 1130 1135

10 Pro Val Gly Tyr Val Thr Lys Glu Leu Glu Tyr Phe Pro Glu Thr Asp
 1140 1145 1150

Lys Val Trp Ile Glu Ile Gly Glu Thr Glu Gly Thr Phe Ile Val Asp
 1155 1160 1165

Ser Val Glu Leu Leu Leu Met Glu Glu
 1170 1175

15 (2) SEQ ID NO: 15 信息:

(i) 序列特征:

(A) 长度: 20 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

20 (D) 拓扑学: 线性

(xi) 序列描述: SEQ ID NO: 15:

TATCCAATTC GAACGTCATC

20

(2) SEQ ID NO: 16 信息:

25 (i) 序列特征:

(A) 长度: 20 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑学: 线性

30 (xi) 序列描述: SEQ ID NO: 16:

TTTAGTCATC GATTAAATCA

20

(2) SEQ ID NO : 17 信息:

(i) 序列特征:

(A) 长度: 20 个碱基对

(B) 类型: 核酸

5 (C) 链型: 单链

(D) 拓扑学: 线性

(xi) 序列描述: SEQ ID NO : 17 :

ATAATAAGAG CTCCAATGTT

20

10 (2) SEQ ID NO : 18 信息:

(i) 序列特征:

(A) 长度: 20 个碱基对

(B) 类型: 核酸

15 (C) 链型: 单链

(D) 拓扑学: 线性

(xi) 序列描述: SEQ ID NO : 18 :

TACATCGTAG TGCAACTCTT

20

(2) SEQ ID NO : 19 信息:

20 (i) 序列特征:

(A) 长度: 20 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑学: 线性

25 (xi) 序列描述: SEQ ID NO : 19 :

TCATGGAGAG CTCCTATGTT

20

(2) SEQ ID NO : 20 信息:

(i) 序列特征:

30 (A) 长度: 20 个碱基对

(B) 类型: 核酸



(C) 链型: 单链

(D) 拓扑学: 线性

(xi) 序列描述: SEQ ID NO: 20:

TTAACAAGAG CTCCTATGTT

20

5

(2) SEQ ID NO: 21 信息:

(i) 序列特征:

(A) 长度: 20 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑学: 线性

(xi) 序列描述: SEQ ID NO: 21:

ACTACCAGGT ACCTTTGATG

20

10

15 (2) SEQ ID NO: 22 信息:

(i) 序列特征:

(A) 长度: 20 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑学: 线性

(xi) 序列描述: SEQ ID NO: 22:

ACTACCGGGT ACCTTGATA

20

20

(2) SEQ ID NO: 23 信息:

(i) 序列特征:

(A) 长度: 18 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑学: 线性

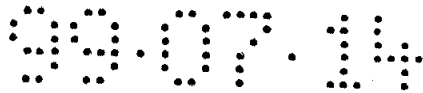
(xi) 序列描述: SEQ ID NO: 23:

ATTGAGTAA TACTATCC

18

25

30



(2) SEQ ID NO : 24 信息:

(i) 序列特征:

(A) 长度: 19 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑学: 线性

(xi) 序列描述: SEQ ID NO : 24 :

ATTACTCAAA TACCATTGG

19

10 (2) SEQ ID NO : 25 信息:

(i) 序列特征:

(A) 长度: 3534 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑学: 线性

(ix) 特征:

(A) 名字/关键: CDS

(B) 位置: 1..3531

(xi) 序列描述: SEQ ID NO : 25:

| | |
|---|----|
| ATG GAT AAC AAT CCG AAC ATC AAT GAA TGC ATT CCT TAT AAT TGT TTA | 48 |
| Met Asp Asn Asn Pro Asn Ile Asn Glu Cys Ile Pro Tyr Asn Cys Leu | |
| 1 5 10 15 | |

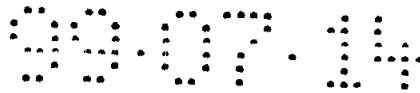
| | |
|---|----|
| AGT AAC CCT GAA GTA GAA GTA TTA GGT GGA GAA AGA ATA GAA ACT GGT | 96 |
| Ser Asn Pro Glu Val Glu Val Leu Gly Gly Glu Arg Ile Glu Thr Gly | |
| 20 25 30 | |

| | |
|---|-----|
| TAC ACC CCA ATC GAT ATT TCC TTG TCG CTA ACG CAA TTT CTT TTG AGT | 144 |
| Tyr Thr Pro Ile Asp Ile Ser Leu Ser Leu Thr Gln Phe Leu Leu Ser | |
| 35 40 45 | |

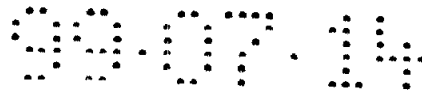
| | |
|---|-----|
| GAA TTT GTT CCC GGT GCT GGA TTT GTG TTA GGA CTA GTT GAT ATA ATA | 192 |
| Glu Phe Val Pro Gly Ala Gly Phe Val Leu Gly Leu Val Asp Ile Ile | |
| 50 55 60 | |

| | |
|---|-----|
| TGG GGA ATT TTT GGT CCC TCT CAA TGG GAC GCA TTT CTT GTA CAA ATT | 240 |
| Trp Gly Ile Phe Gly Pro Ser Gln Trp Asp Ala Phe Leu Val Gln Ile | |
| 65 70 75 80 | |

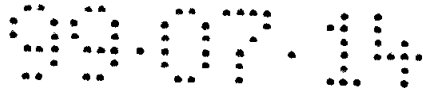
| | |
|---|-----|
| GAA CAG TTA ATT AAC CAA AGA ATA GAA GAA TTC GCT AGG AAC CAA GCC | 288 |
| Glu Gln Leu Ile Asn Gln Arg Ile Glu Glu Phe Ala Arg Asn Gln Ala | |



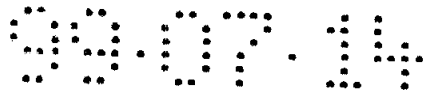
| | | | | | 85 | | | | | | 90 | | | | | | 95 | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----|--|
| ATT | TCT | AGA | TTA | GAA | GGA | CTA | AGC | AAT | CTT | TAT | CAA | ATT | TAC | GCA | GAA | 336 | | |
| Ile | Ser | Arg | Leu | Glu | Gly | Leu | Ser | Asn | Leu | Tyr | Gln | Ile | Tyr | Ala | Glu | | | |
| | | | 100 | | | | | 105 | | | | | 110 | | | | | |
| TCT | TTT | AGA | GAG | TGG | GAA | GCA | GAT | CCT | ACT | AAT | CCA | GCA | TTA | AGA | GAA | 384 | | |
| Ser | Phe | Arg | Glu | Trp | Glu | Ala | Asp | Pro | Thr | Asn | Pro | Ala | Leu | Arg | Glu | | | |
| | | 115 | | | | | 120 | | | | | 125 | | | | | | |
| GAG | ATG | CGT | ATT | CAA | TTC | AAT | GAC | ATG | AAC | AGT | GCC | CTT | ACA | ACC | GCT | 432 | | |
| Glu | Met | Arg | Ile | Gln | Phe | Asn | Asp | Met | Asn | Ser | Ala | Leu | Thr | Thr | Ala | | | |
| | 130 | | | | | 135 | | | | | 140 | | | | | | | |
| ATT | CCT | CTT | TTT | GCA | GTT | CAA | AAT | TAT | CAA | GTT | CCT | CTT | TTA | TCA | GTA | 480 | | |
| Ile | Pro | Leu | Phe | Ala | Val | Gln | Asn | Tyr | Gln | Val | Pro | Leu | Leu | Ser | Val | | | |
| 145 | | | | | 150 | | | | | 155 | | | | | 160 | | | |
| TAT | GTT | CAA | GCT | GCA | AAT | TTA | CAT | TTA | TCA | GTT | TTG | AGA | GAT | GTT | TCA | 528 | | |
| Tyr | Val | Gln | Ala | Ala | Asn | Leu | His | Leu | Ser | Val | Leu | Arg | Asp | Val | Ser | | | |
| | | | 165 | | | | | 170 | | | | | 175 | | | | | |
| GTG | TTT | GGA | CAA | AGG | TGG | GGA | TTT | GAT | GCC | GCG | ACT | ATC | AAT | AGT | CGT | 576 | | |
| Val | Phe | Gly | Gln | Arg | Trp | Gly | Phe | Asp | Ala | Ala | Thr | Ile | Asn | Ser | Arg | | | |
| | | 180 | | | | | | 185 | | | | | 190 | | | | | |
| TAT | AAT | GAT | TTA | ACT | AGG | CTT | ATT | GGC | AAC | TAT | ACA | GAT | CAT | GCT | GTA | 624 | | |
| Tyr | Asn | Asp | Leu | Thr | Arg | Leu | Ile | Gly | Asn | Tyr | Thr | Asp | His | Ala | Val | | | |
| | | 195 | | | | 200 | | | | | | 205 | | | | | | |
| CGC | TGG | TAC | AAT | ACG | GGA | TTA | GAG | CGT | GTA | TGG | GGA | CCG | GAT | TCT | AGA | 672 | | |
| Arg | Trp | Tyr | Asn | Thr | Gly | Leu | Glu | Arg | Val | Trp | Gly | Pro | Asp | Ser | Arg | | | |
| | 210 | | | | | 215 | | | | | 220 | | | | | | | |
| GAT | TGG | ATA | AGA | TAT | AAT | CAA | TTT | AGA | AGA | GAA | TTA | ACA | CTA | ACT | GTA | 720 | | |
| Asp | Trp | Ile | Arg | Tyr | Asn | Gln | Phe | Arg | Arg | Glu | Leu | Thr | Leu | Thr | Val | | | |
| 225 | | | | | 230 | | | | | 235 | | | | | 240 | | | |
| TTA | GAT | ATC | GTT | TCT | CTA | TTT | CCG | AAC | TAT | GAT | AGT | AGA | ACG | TAT | CCA | 768 | | |
| Leu | Asp | Ile | Val | Ser | Leu | Phe | Pro | Asn | Tyr | Asp | Ser | Arg | Thr | Tyr | Pro | | | |
| | | | 245 | | | | | 250 | | | | | | 255 | | | | |
| ATT | CGA | ACA | GTT | TCC | CAA | TTA | ACA | AGA | GAA | ATT | TAT | ACA | AAC | CCA | GTA | 816 | | |
| Ile | Arg | Thr | Val | Ser | Gln | Leu | Thr | Arg | Glu | Ile | Tyr | Thr | Asn | Pro | Val | | | |
| | | | 260 | | | | | 265 | | | | | 270 | | | | | |
| TTA | GAA | AAT | TTT | GAT | GGT | AGT | TTT | CGA | GGC | TCG | GCT | CAG | GGC | ATA | GAA | 864 | | |
| Leu | Glu | Asn | Phe | Asp | Gly | Ser | Phe | Arg | Gly | Ser | Ala | Gln | Gly | Ile | Glu | | | |
| | | 275 | | | | | 280 | | | | | 285 | | | | | | |
| AGA | AGT | ATT | AGG | AGT | CCA | CAT | TTG | ATG | GAT | ATA | CTT | AAC | AGT | ATA | ACC | 912 | | |
| Arg | Ser | Ile | Arg | Ser | Pro | His | Leu | Met | Asp | Ile | Leu | Asn | Ser | Ile | Thr | | | |
| | 290 | | | | | 295 | | | | | 300 | | | | | | | |
| ATC | TAT | ACG | GAT | GCT | CAT | AGG | GGT | TAT | TAT | TAT | TGG | TCA | GGG | CAT | CAA | 960 | | |
| Ile | Tyr | Thr | Asp | Ala | His | Arg | Gly | Tyr | Tyr | Tyr | Trp | Ser | Gly | His | Gln | | | |
| 305 | | | | | 310 | | | | | 315 | | | | | 320 | | | |



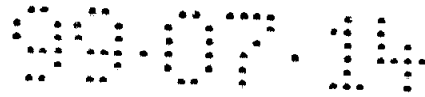
| | |
|---|------|
| ATA ATG GCT TCT CCT GTA GGG TTT TCG GGG CCA GAA TTC ACT TTT CCG | 1008 |
| Ile Met Ala Ser Pro Val Gly Phe Ser Gly Pro Glu Phe Thr Phe Pro | |
| 325 330 335 | |
| CTA TAT GGA ACT ATG GGA AAT GCA GCT CCA CAA CAA CGT ATT GTT GCT | 1056 |
| Leu Tyr Gly Thr Met Gly Asn Ala Ala Pro Gln Gln Arg Ile Val Ala | |
| 340 345 350 | |
| CAA CTA GGT CAG GGC GTG TAT AGA ACA TTA TCG TCC ACT TTA TAT AGA | 1104 |
| Gln Leu Gly Gln Gly Val Tyr Arg Thr Leu Ser Ser Thr Leu Tyr Arg | |
| 355 360 365 | |
| AGA CCT TTT AAT ATA GGG ATA AAT AAT CAA CAA CTA TCT GTT CTT GAC | 1152 |
| Arg Pro Phe Asn Ile Gly Ile Asn Asn Gln Gln Leu Ser Val Leu Asp | |
| 370 375 380 | |
| GGG ACA GAA TTT GCT TAT GGA ACC TCC TCA AAT TTG CCA TCC GCT GTA | 1200 |
| Gly Thr Glu Phe Ala Tyr Gly Thr Ser Ser Asn Leu Pro Ser Ala Val | |
| 385 390 395 400 | |
| TAC AGA AAA AGC GGA ACG GTA GAT TCG CTG GAT GAA ATA CCG CCA CAG | 1248 |
| Tyr Arg Lys Ser Gly Thr Val Asp Ser Leu Asp Glu Ile Pro Pro Gln | |
| 405 410 415 | |
| AAT AAC AAC GTG CCA CCT AGG CAA GGA TTT AGT CAT CGA TTA AGC CAT | 1296 |
| Asn Asn Asn Val Pro Pro Arg Gln Gly Phe Ser His Arg Leu Ser His | |
| 420 425 430 | |
| GTT TCA ATG TTT CGT TCA GGC TTT AGT AAT AGT AGT GTA AGT ATA ATA | 1344 |
| Val Ser Met Phe Arg Ser Gly Phe Ser Asn Ser Ser Val Ser Ile Ile | |
| 435 440 445 | |
| AGA GCT CCA ATG TTT TCT TGG ACG CAC CGT AGT GCA ACC CCT ACA AAT | 1392 |
| Arg Ala Pro Met Phe Ser Trp Thr His Arg Ser Ala Thr Pro Thr Asn | |
| 450 455 460 | |
| ACA ATT GAT CCG GAG AGG ATT ACT CAA ATA CCA TTG GTA AAA GCA CAT | 1440 |
| Thr Ile Asp Pro Glu Arg Ile Thr Gln Ile Pro Leu Val Lys Ala His | |
| 465 470 475 480 | |
| ACA CTT CAG TCA GGT ACT ACT GTT GTA AGA GGG CCC GGG TTT ACG GGA | 1488 |
| Thr Leu Gln Ser Gly Thr Thr Val Val Arg Gly Pro Gly Phe Thr Gly | |
| 485 490 495 | |
| GGA GAT ATT CTT CGA CGA ACA AGT GGA GGA CCA TTT GCT TAT ACT ATT | 1536 |
| Gly Asp Ile Leu Arg Arg Thr Ser Gly Gly Pro Phe Ala Tyr Thr Ile | |
| 500 505 510 | |
| GTT AAT ATA AAT GGG CAA TTA CCC CAA AGG TAT CGT GCA AGA ATA CGC | 1584 |
| Val Asn Ile Asn Gly Gln Leu Pro Gln Arg Tyr Arg Ala Arg Ile Arg | |
| 515 520 525 | |
| TAT GCC TCT ACT ACA AAT CTA AGA ATT TAC GTA ACG GTT GCA GGT GAA | 1632 |
| Tyr Ala Ser Thr Thr Asn Leu Arg Ile Tyr Val Thr Val Ala Gly Glu | |
| 530 535 540 | |



| | |
|---|------|
| CGG ATT TTT GCT GGT CAA TTT AAC AAA ACA ATG GAT ACC GGT GAC CCA Arg Ile Phe Ala Gly Gln Phe Asn Lys Thr Met Asp Thr Gly Asp Pro 545 550 555 560 | 1680 |
| TTA ACA TTC CAA TCT TTT AGT TAC GCA ACT ATT AAT ACA GCT TTT ACA Leu Thr Phe Gln Ser Phe Ser Tyr Ala Thr Ile Asn Thr Ala Phe Thr 565 570 575 | 1728 |
| TTC CCA ATG AGC CAG AGT AGT TTC ACA GTA GGT GCT GAT ACT TTT AGT Phe Pro Met Ser Gln Ser Ser Phe Thr Val Gly Ala Asp Thr Phe Ser 580 585 590 | 1776 |
| TCA GGG AAT GAA GTT TAT ATA GAC AGA TTT GAA TTG ATT CCA GTT ACT Ser Gly Asn Glu Val Tyr Ile Asp Arg Phe Glu Leu Ile Pro Val Thr 595 600 605 | 1824 |
| GCA ACA TTT GAA GCA GAA TAT GAT TTA GAA AGA GCA CAA AAG GCG GTG Ala Thr Phe Glu Ala Glu Tyr Asp Leu Glu Arg Ala Gln Lys Ala Val 610 615 620 | 1872 |
| AAT GCG CTG TTT ACT TCT ATA AAC CAA ATA GGG ATA AAA ACA GAT GTG Asn Ala Leu Phe Thr Ser Ile Asn Gln Ile Gly Ile Lys Thr Asp Val 625 630 635 640 | 1920 |
| ACG GAT TAT CAT ATT GAT CAA GTA TCC AAT TTA GTG GAT TGT TTA TCA Thr Asp Tyr His Ile Asp Gln Val Ser Asn Leu Val Asp Cys Leu Ser 645 650 655 | 1968 |
| GAT GAA TTT TGT CTG GAT GAA AAG CGA GAA TTG TCC GAG AAA GTC AAA Asp Glu Phe Cys Leu Asp Glu Lys Arg Glu Leu Ser Glu Lys Val Lys 660 665 670 | 2016 |
| CAT GCG AAG CGA CTC AGT GAT GAG CGG AAT TTA CTT CAA GAT CCA AAC His Ala Lys Arg Leu Ser Asp Glu Arg Asn Leu Leu Gln Asp Pro Asn 675 680 685 | 2064 |
| TTC AAA GGC ATC AAT AGG CAA CTA GAC CGT GGT TGG AGA GGA AGT ACG Phe Lys Gly Ile Asn Arg Gln Leu Asp Arg Gly Trp Arg Gly Ser Thr 690 695 700 | 2112 |
| GAT ATT ACC ATC CAA AGA GGA GAT GAC GTA TTC AAA GAA AAT TAT GTC Asp Ile Thr Ile Gln Arg Gly Asp Asp Val Phe Lys Glu Asn Tyr Val 705 710 715 720 | 2160 |
| ACA CTA CCA GGT ACC TTT GAT GAG TGC TAT CCA ACA TAT TTG TAT CAA Thr Leu Pro Gly Thr Phe Asp Glu Cys Tyr Pro Thr Tyr Leu Tyr Gln 725 730 735 | 2208 |
| AAA ATC GAT GAA TCA AAA TTA AAA GCC TTT ACC CGT TAT CAA TTA AGA Lys Ile Asp Glu Ser Lys Leu Lys Ala Phe Thr Arg Tyr Gln Leu Arg 740 745 750 | 2256 |
| GGG TAT ATC GAA GAT AGT CAA GAC TTA GAA ATC TAT TTA ATT CGC TAC Gly Tyr Ile Glu Asp Ser Gln Asp Leu Glu Ile Tyr Leu Ile Arg Tyr 755 760 765 | 2304 |
| AAT GCA AAA CAT GAA ACA GTA AAT GTG CCA GGT ACG GGT TCC TTA TGG | 2352 |



| | | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|-----|
| Asn | Ala | Lys | His | Glu | Thr | Val | Asn | Val | Pro | Gly | Thr | Gly | Ser | Leu | Trp | | |
| 770 | | | | | | 775 | | | | | 780 | | | | | | |
| CCG | CTT | TCA | GCC | CAA | AGT | CCA | ATC | GGA | AAG | TGT | GGA | GAG | CCG | AAT | CGA | 2400 | |
| Pro | Leu | Ser | Ala | Gln | Ser | Pro | Ile | Gly | Lys | Cys | Gly | Glu | Pro | Asn | Arg | | 800 |
| 785 | | | | | 790 | | | | 795 | | | | | | | | |
| TGC | GCG | CCA | CAC | CTT | GAA | TGG | AAT | CCT | GAC | TTA | GAT | TGT | TCG | TGT | AGG | 2448 | |
| Cys | Ala | Pro | His | Leu | Glu | Trp | Asn | Pro | Asp | Leu | Asp | Cys | Ser | Cys | Arg | | 815 |
| | | | | 805 | | | | | 810 | | | | | | | | |
| GAT | GGA | GAA | AAG | TGT | GCC | CAT | CAT | TCG | CAT | CAT | TTC | TCC | TTA | GAC | ATT | 2496 | |
| Asp | Gly | Glu | Lys | Cys | Ala | His | His | Ser | His | His | Phe | Ser | Leu | Asp | Ile | | 830 |
| | | | 820 | | | | | 825 | | | | | | | | | |
| GAT | GTA | GGA | TGT | ACA | GAC | TTA | AAT | GAG | GAC | CTA | GGT | GTA | TGG | GTG | ATC | 2544 | |
| Asp | Val | Gly | Cys | Thr | Asp | Leu | Asn | Glu | Asp | Leu | Gly | Val | Trp | Val | Ile | | 845 |
| | | 835 | | | | | 840 | | | | | 845 | | | | | |
| TTT | AAG | ATT | AAG | ACG | CAA | GAT | GGG | CAC | GCA | AGA | CTA | GGG | AAT | CTA | GAG | 2592 | |
| Phe | Lys | Ile | Lys | Thr | Gln | Asp | Gly | His | Ala | Arg | Leu | Gly | Asn | Leu | Glu | | 860 |
| | 850 | | | | | 855 | | | | | 860 | | | | | | |
| TTT | CTC | GAA | GAG | AAA | CCA | TTA | GTA | GGA | GAA | GCG | CTA | GCT | CGT | GTG | AAA | 2640 | |
| Phe | Leu | Glu | Glu | Lys | Pro | Leu | Val | Gly | Glu | Ala | Leu | Ala | Arg | Val | Lys | | 880 |
| | | | | | 870 | | | | | 875 | | | | | | | |
| AGA | GCG | GAG | AAA | AAA | TGG | AGA | GAC | AAA | CGT | GAA | AAA | TTG | GAA | TGG | GAA | 2688 | |
| Arg | Ala | Glu | Lys | Lys | Trp | Arg | Asp | Lys | Arg | Glu | Lys | Leu | Glu | Trp | Glu | | 895 |
| | | | | 885 | | | | | 890 | | | | | | | | |
| ACA | AAT | ATC | GTT | TAT | AAA | GAG | GCA | AAA | GAA | TCT | GTA | GAT | GCT | TTA | TTT | 2736 | |
| Thr | Asn | Ile | Val | Tyr | Lys | Glu | Ala | Lys | Glu | Ser | Val | Asp | Ala | Leu | Phe | | 910 |
| | | | 900 | | | | | 905 | | | | | | 910 | | | |
| GTA | AAC | TCT | CAA | TAT | GAT | CAA | TTA | CAA | GCG | GAT | ACG | AAT | ATT | GCC | ATG | 2784 | |
| Val | Asn | Ser | Gln | Tyr | Asp | Gln | Leu | Gln | Ala | Asp | Thr | Asn | Ile | Ala | Met | | 925 |
| | | 915 | | | | 920 | | | | | | 925 | | | | | |
| ATT | CAT | GCG | GCA | GAT | AAA | CGT | GTT | CAT | AGC | ATT | CGA | GAA | GCT | TAT | CTG | 2832 | |
| Ile | His | Ala | Ala | Asp | Lys | Arg | Val | His | Ser | Ile | Arg | Glu | Ala | Tyr | Leu | | 940 |
| | | 930 | | | | 935 | | | | | 940 | | | | | | |
| CCT | GAG | CTG | TCT | GTG | ATT | CCG | GGT | GTC | AAT | GCG | GCT | ATT | TTT | GAA | GAA | 2880 | |
| Pro | Glu | Leu | Ser | Val | Ile | Pro | Gly | Val | Asn | Ala | Ala | Ile | Phe | Glu | Glu | | 960 |
| | | | | | 950 | | | | | 955 | | | | | | | |
| TTA | GAA | GGG | CGT | ATT | TTC | ACT | GCA | TTC | TCC | CTA | TAT | GAT | GCG | AGA | AAT | 2928 | |
| Leu | Glu | Gly | Arg | Ile | Phe | Thr | Ala | Phe | Ser | Leu | Tyr | Asp | Ala | Arg | Asn | | 975 |
| | | | | 965 | | | | | 970 | | | | | | | | |
| GTC | ATT | AAA | AAT | GGT | GAT | TTT | AAT | AAT | GGC | TTA | TCC | TGC | TGG | AAC | GTG | 2976 | |
| Val | Ile | Lys | Asn | Gly | Asp | Phe | Asn | Asn | Gly | Leu | Ser | Cys | Trp | Asn | Val | | 990 |
| | | | 980 | | | | | 985 | | | | | 990 | | | | |
| AAA | GGG | CAT | GTA | GAT | GTA | GAA | GAA | CAA | AAC | AAC | CAA | CGT | TCG | GTC | CTT | 3024 | |
| Lys | Gly | His | Val | Asp | Val | Glu | Glu | Gln | Asn | Asn | Gln | Arg | Ser | Val | Leu | | |



| | 995 | 1000 | 1005 | |
|----|---|------|------|------|
| | GTT GTT CCG GAA TGG GAA GCA GAA GTG TCA CAA GAA GTT CGT GTC TGT | | | 3072 |
| | Val Val Pro Glu Trp Glu Ala Glu Val Ser Gln Glu Val Arg Val Cys | | | |
| | 1010 | 1015 | 1020 | |
| 5 | CCG GGT CGT GGC TAT ATC CTT CGT GTC ACA GCG TAC AAG GAG GGA TAT | | | 3120 |
| | Pro Gly Arg Gly Tyr Ile Leu Arg Val Thr Ala Tyr Lys Glu Gly Tyr | | | |
| | 1025 | 1030 | 1035 | 1040 |
| | GGA GAA GGT TGC GTA ACC ATT CAT GAG ATC GAG AAC AAT ACA GAC GAA | | | 3168 |
| | Gly Glu Gly Cys Val Thr Ile His Glu Ile Glu Asn Asn Thr Asp Glu | | | |
| | 1045 | 1050 | 1055 | |
| | CTG AAG TTT AGC AAC TGC GTA GAA GAG GAA ATC TAT CCA AAT AAC ACG | | | 3216 |
| | Leu Lys Phe Ser Asn Cys Val Glu Glu Glu Ile Tyr Pro Asn Asn Thr | | | |
| | 1060 | 1065 | 1070 | |
| 10 | GTA ACG TGT AAT GAT TAT ACT GTA AAT CAA GAA GAA TAC GGA GGT GCG | | | 3264 |
| | Val Thr Cys Asn Asp Tyr Thr Val Asn Gln Glu Glu Tyr Gly Gly Ala | | | |
| | 1075 | 1080 | 1085 | |
| | TAC ACT TCT CGT AAT CGA GCA TAT AAC GAA GCT CCT TCC GTA CCA GCT | | | 3312 |
| | Tyr Thr Ser Arg Asn Arg Gly Tyr Asn Glu Ala Pro Ser Val Pro Ala | | | |
| | 1090 | 1095 | 1100 | |
| 15 | GAT TAT GCG TCA GTC TAT GAA GAA AAA TCG TAT ACA GAT GGA CGA AGA | | | 3360 |
| | Asp Tyr Ala Ser Val Tyr Glu Glu Lys Ser Tyr Thr Asp Gly Arg Arg | | | |
| | 1105 | 1110 | 1115 | 1120 |
| | GAG AAT CCT TGT GAA TTT AAC AGA GGG TAT AGG GAT TAC ACG CCA CTA | | | 3408 |
| | Glu Asn Pro Cys Glu Phe Asn Arg Gly Tyr Arg Asp Tyr Thr Pro Leu | | | |
| | 1125 | 1130 | 1135 | |
| | CCA GTT GGT TAT GTG ACA AAA GAA TTA GAA TAC TTC CCA GAA ACC GAT | | | 3456 |
| | Pro Val Gly Tyr Val Thr Lys Glu Leu Glu Tyr Phe Pro Glu Thr Asp | | | |
| | 1140 | 1145 | 1150 | |
| 20 | AAG GTA TGG ATT GAG ATT GGA GAA ACG GAA GGA ACA TTT ATC GTG GAC | | | 3504 |
| | Lys Val Trp Ile Glu Ile Gly Glu Thr Glu Gly Thr Phe Ile Val Asp | | | |
| | 1155 | 1160 | 1165 | |
| | AGC GTG GAA TTA CTC CTT ATG GAG GAA TAG | | | 3534 |
| | Ser Val Glu Leu Leu Leu Met Glu Glu | | | |
| | 1170 | 1175 | | |

25 (2) SEQ ID NO : 26 信息:

(i) 序列特征:

(A) 长度: 1177 个氨基酸

(B) 类型: 氨基酸

(D) 拓扑学: 线性

30 (ii) 分子类型: 蛋白

(xi) 序列描述: SEQ ID NO : 26 :

90714

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

907.14

Ile Tyr Thr Asp Ala His Arg Gly Tyr Tyr Tyr Trp Ser Gly His Gln
 305 310 315 320

Ile Met Ala Ser Pro Val Gly Phe Ser Gly Pro Glu Phe Thr Phe Pro
 325 330 335

Leu Tyr Gly Thr Met Gly Asn Ala Ala Pro Gln Gln Arg Ile Val Ala
 340 345 350

Gln Leu Gly Gln Gly Val Tyr Arg Thr Leu Ser Ser Thr Leu Tyr Arg
 355 360 365

Arg Pro Phe Asn Ile Gly Ile Asn Asn Gln Gln Leu Ser Val Leu Asp
 370 375 380

Gly Thr Glu Phe Ala Tyr Gly Thr Ser Ser Asn Leu Pro Ser Ala Val
 385 390 395 400

Tyr Arg Lys Ser Gly Thr Val Asp Ser Leu Asp Glu Ile Pro Pro Gln
 405 410 415

Asn Asn Asn Val Pro Pro Arg Gln Gly Phe Ser His Arg Leu Ser His
 420 425 430

Val Ser Met Phe Arg Ser Gly Phe Ser Asn Ser Ser Val Ser Ile Ile
 435 440 445

Arg Ala Pro Met Phe Ser Trp Thr His Arg Ser Ala Thr Pro Thr Asn
 450 455 460

Thr Ile Asp Pro Glu Arg Ile Thr Gln Ile Pro Leu Val Lys Ala His
 465 470 475 480

Thr Leu Gln Ser Gly Thr Thr Val Val Arg Gly Pro Gly Phe Thr Gly
 485 490 495

Gly Asp Ile Leu Arg Arg Thr Ser Gly Gly Pro Phe Ala Tyr Thr Ile
 500 505 510

Val Asn Ile Asn Gly Gln Leu Pro Gln Arg Tyr Arg Ala Arg Ile Arg
 515 520 525

Tyr Ala Ser Thr Thr Asn Leu Arg Ile Tyr Val Thr Val Ala Gly Glu
 530 535 540

Arg Ile Phe Ala Gly Gln Phe Asn Lys Thr Met Asp Thr Gly Asp Pro
 545 550 555 560

Leu Thr Phe Gln Ser Phe Ser Tyr Ala Thr Ile Asn Thr Ala Phe Thr
 565 570 575

Phe Pro Met Ser Gln Ser Ser Phe Thr Val Gly Ala Asp Thr Phe Ser
 580 585 590

Ser Gly Asn Glu Val Tyr Ile Asp Arg Phe Glu Leu Ile Pro Val Thr
 595 600 605

907.14

Ala Thr Phe Glu Ala Glu Tyr Asp Leu Glu Arg Ala Gln Lys Ala Val
610 615 620

Asn Ala Leu Phe Thr Ser Ile Asn Gln Ile Gly Ile Lys Thr Asp Val
625 630 635 640

Thr Asp Tyr His Ile Asp Gln Val Ser Asn Leu Val Asp Cys Leu Ser
645 650 655

Asp Glu Phe Cys Leu Asp Glu Lys Arg Glu Leu Ser Glu Lys Val Lys
660 665 670

His Ala Lys Arg Leu Ser Asp Glu Arg Asn Leu Leu Gln Asp Pro Asn
675 680 685

Phe Lys Gly Ile Asn Arg Gln Leu Asp Arg Gly Trp Arg Gly Ser Thr
690 695 700

Asp Ile Thr Ile Gln Arg Gly Asp Asp Val Phe Lys Glu Asn Tyr Val
705 710 715 720

Thr Leu Pro Gly Thr Phe Asp Glu Cys Tyr Pro Thr Tyr Leu Tyr Gln
725 730 735

Lys Ile Asp Glu Ser Lys Leu Lys Ala Phe Thr Arg Tyr Gln Leu Arg
740 745 750

Gly Tyr Ile Glu Asp Ser Gln Asp Leu Glu Ile Tyr Leu Ile Arg Tyr
755 760 765

Asn Ala Lys His Glu Thr Val Asn Val Pro Gly Thr Gly Ser Leu Trp
770 775 780

Pro Leu Ser Ala Gln Ser Pro Ile Gly Lys Cys Gly Glu Pro Asn Arg
785 790 795 800

Cys Ala Pro His Leu Glu Trp Asn Pro Asp Leu Asp Cys Ser Cys Arg
805 810 815

Asp Gly Glu Lys Cys Ala His His Ser His His Phe Ser Leu Asp Ile
820 825 830

Asp Val Gly Cys Thr Asp Leu Asn Glu Asp Leu Gly Val Trp Val Ile
835 840 845

Phe Lys Ile Lys Thr Gln Asp Gly His Ala Arg Leu Gly Asn Leu Glu
850 855 860

Phe Leu Glu Glu Lys Pro Leu Val Gly Glu Ala Leu Ala Arg Val Lys
865 870 875 880

Arg Ala Glu Lys Lys Trp Arg Asp Lys Arg Glu Lys Leu Glu Trp Glu
885 890 895

Thr Asn Ile Val Tyr Lys Glu Ala Lys Glu Ser Val Asp Ala Leu Phe
900 905 910

27

Val Asn Ser Gln Tyr Asp Gln Leu Gln Ala Asp Thr Asn Ile Ala Met
 915 920 925
 Ile His Ala Ala Asp Lys Arg Val His Ser Ile Arg Glu Ala Tyr Leu
 930 935 940
 5 Pro Glu Leu Ser Val Ile Pro Gly Val Asn Ala Ala Ile Phe Glu Glu
 945 950 955 960
 Leu Glu Gly Arg Ile Phe Thr Ala Phe Ser Leu Tyr Asp Ala Arg Asn
 965 970 975
 Val Ile Lys Asn Gly Asp Phe Asn Asn Gly Leu Ser Cys Trp Asn Val
 980 985 990
 10 Lys Gly His Val Asp Val Glu Glu Gln Asn Asn Gln Arg Ser Val Leu
 995 1000 1005
 Val Val Pro Glu Trp Glu Ala Glu Val Ser Gln Glu Val Arg Val Cys
 1010 1015 1020
 Pro Gly Arg Gly Tyr Ile Leu Arg Val Thr Ala Tyr Lys Glu Gly Tyr
 1025 1030 1035 1040
 Gly Glu Gly Cys Val Thr Ile His Glu Ile Glu Asn Asn Thr Asp Glu
 1045 1050 1055
 15 Leu Lys Phe Ser Asn Cys Val Glu Glu Glu Ile Tyr Pro Asn Asn Thr
 1060 1065 1070
 Val Thr Cys Asn Asp Tyr Thr Val Asn Gln Glu Glu Tyr Gly Gly Ala
 1075 1080 1085
 Tyr Thr Ser Arg Asn Arg Gly Tyr Asn Glu Ala Pro Ser Val Pro Ala
 1090 1095 1100
 20 Asp Tyr Ala Ser Val Tyr Glu Glu Lys Ser Tyr Thr Asp Gly Arg Arg
 1105 1110 1115 1120
 Glu Asn Pro Cys Glu Phe Asn Arg Gly Tyr Arg Asp Tyr Thr Pro Leu
 1125 1130 1135
 Pro Val Gly Tyr Val Thr Lys Glu Leu Glu Tyr Phe Pro Glu Thr Asp
 1140 1145 1150
 25 Lys Val Trp Ile Glu Ile Gly Glu Thr Glu Gly Thr Phe Ile Val Asp
 1155 1160 1165
 Ser Val Glu Leu Leu Leu Met Glu Glu
 1170 1175

(2) SEQ ID NO : 27 信息:

(i) 序列特征:

- 30 (A) 长度: 3534 个碱基对
- (B) 类型: 核酸



(C) 键型: 单键

(D) 拓扑学: 线性

(ix) 特征:

(A) 名字/关键: CDS

(B) 位置: 1..3531

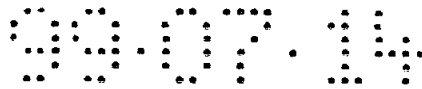
(xi) 序列描述: SEQ ID NO: 27:

5

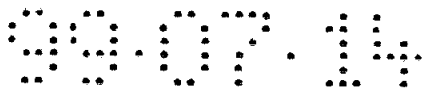
| | |
|---|-----|
| ATG GAT AAC AAT CCG AAC ATC AAT GAA TGC ATT CCT TAT AAT TGT TTA | 48 |
| Met Asp Asn Asn Pro Asn Ile Asn Glu Cys Ile Pro Tyr Asn Cys Leu | |
| 1 5 10 15 | |
| AGT AAC CCT GAA GTA GAA GTA TTA GGT GGA GAA AGA ATA GAA ACT GGT | 96 |
| Ser Asn Pro Glu Val Glu Val Leu Gly Gly Glu Arg Ile Glu Thr Gly | |
| 20 25 30 | |
| TAC ACC CCA ATC GAT ATT TCC TTG TCG CTA ACG CAA TTT CTT TTG AGT | 144 |
| Tyr Thr Pro Ile Asp Ile Ser Leu Ser Leu Thr Gln Phe Leu Leu Ser | |
| 35 40 45 | |
| GAA TTT GTT CCC GGT GCT GGA TTT GTG TTA GGA CTA GTT GAT ATA ATA | 192 |
| Glu Phe Val Pro Gly Ala Gly Phe Val Leu Gly Leu Val Asp Ile Ile | |
| 50 55 60 | |
| TGG GGA ATT TTT GGT CCC TCT CAA TGG GAC GCA TTT CTT GTA CAA ATT | 240 |
| Trp Gly Ile Phe Gly Pro Ser Gln Trp Asp Ala Phe Leu Val Gln Ile | |
| 65 70 75 80 | |
| GAA CAG TTA ATT AAC CAA AGA ATA GAA GAA TTC GCT AGG AAC CAA GCC | 288 |
| Glu Gln Leu Ile Asn Gln Arg Ile Glu Glu Phe Ala Arg Asn Gln Ala | |
| 85 90 95 | |
| ATT TCT AGA TTA GAA GGA CTA AGC AAT CTT TAT CAA ATT TAC GCA GAA | 336 |
| Ile Ser Arg Leu Glu Gly Leu Ser Asn Leu Tyr Gln Ile Tyr Ala Glu | |
| 100 105 110 | |
| TCT TTT AGA GAG TGG GAA GCA GAT CCT ACT AAT CCA GCA TTA AGA GAA | 384 |
| Ser Phe Arg Glu Trp Glu Ala Asp Pro Thr Asn Pro Ala Leu Arg Glu | |
| 115 120 125 | |
| GAG ATG CGT ATT CAA TTC AAT GAC ATG AAC AGT GCC CTT ACA ACC GCT | 432 |
| Glu Met Arg Ile Gln Phe Asn Asp Met Asn Ser Ala Leu Thr Thr Ala | |
| 130 135 140 | |
| ATT CCT CTT TTT GCA GTT CAA AAT TAT CAA GTT CCT CTT TTA TCA GTA | 480 |
| Ile Pro Leu Phe Ala Val Gln Asn Tyr Gln Val Pro Leu Leu Ser Val | |
| 145 150 155 160 | |
| TAT GTT CAA GCT GCA AAT TTA CAT TTA TCA GTT TTG AGA GAT GTT TCA | 528 |
| Tyr Val Gln Ala Ala Asn Leu His Leu Ser Val Leu Arg Asp Val Ser | |
| 165 170 175 | |
| GTG TTT GGA CAA AGG TGG GGA TTT GAT GCC GCG ACT ATC AAT AGT CGT | 576 |
| Val Phe Gly Gln Arg Trp Gly Phe Asp Ala Ala Thr Ile Asn Ser Arg | |
| 180 185 190 | |



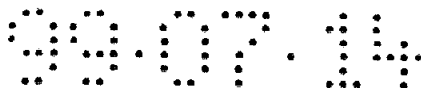
| | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| TAT | AAT | GAT | TTA | ACT | AGG | CTT | ATT | GGC | AAC | TAT | ACA | GAT | TAT | GCT | GTA | 624 |
| Tyr | Asn | Asp | Leu | Thr | Arg | Leu | Ile | Gly | Asn | Tyr | Thr | Asp | Tyr | Ala | Val | |
| | 195 | | | | | | 200 | | | | | 205 | | | | |
| CGC | TGG | TAC | AAT | ACG | GGA | TTA | GAA | CGT | GTA | TGG | GGA | CCG | GAT | TCT | AGA | 672 |
| Arg | Trp | Tyr | Asn | Thr | Gly | Leu | Glu | Arg | Val | Trp | Gly | Pro | Asp | Ser | Arg | |
| | 210 | | | | | | 215 | | | | | 220 | | | | |
| GAT | TGG | GTA | AGG | TAT | AAT | CAA | TTT | AGA | AGA | GAA | TTA | ACA | CTA | ACT | GTA | 720 |
| Asp | Trp | Val | Arg | Tyr | Asn | Gln | Phe | Arg | Arg | Glu | Leu | Thr | Leu | Thr | Val | |
| | 225 | | | | | 230 | | | | 235 | | | | | 240 | |
| TTA | GAT | ATC | GTT | GCT | CTG | TTC | CCG | AAT | TAT | GAT | AGT | AGA | AGA | TAT | CCA | 768 |
| Leu | Asp | Ile | Val | Ala | Leu | Phe | Pro | Asn | Tyr | Asp | Ser | Arg | Arg | Tyr | Pro | |
| | | | | 245 | | | | | | 250 | | | | | 255 | |
| ATT | CGA | ACA | GTT | TCC | CAA | TTA | ACA | AGA | GAA | ATT | TAT | ACA | AAC | CCA | GTA | 816 |
| Ile | Arg | Thr | Val | Ser | Gln | Leu | Thr | Arg | Glu | Ile | Tyr | Thr | Asn | Pro | Val | |
| | | | 260 | | | | | | 265 | | | | | 270 | | |
| TTA | GAA | AAT | TTT | GAT | GGT | AGT | TTT | CGA | GGC | TCG | GCT | CAG | GGC | ATA | GAA | 864 |
| Leu | Glu | Asn | Phe | Asp | Gly | Ser | Phe | Arg | Gly | Ser | Ala | Gln | Gly | Ile | Glu | |
| | | 275 | | | | | | 280 | | | | | | 285 | | |
| AGA | AGT | ATT | AGG | AGT | CCA | CAT | TTG | ATG | GAT | ATA | CTT | AAC | AGT | ATA | ACC | 912 |
| Arg | Ser | Ile | Arg | Ser | Pro | His | Leu | Met | Asp | Ile | Leu | Asn | Ser | Ile | Thr | |
| | | 290 | | | | | 295 | | | | | 300 | | | | |
| ATC | TAT | ACG | GAT | GCT | CAT | AGG | GGT | TAT | TAT | TAT | TGG | TCA | GGG | CAT | CAA | 960 |
| Ile | Tyr | Thr | Asp | Ala | His | Arg | Gly | Tyr | Tyr | Tyr | Trp | Ser | Gly | His | Gln | |
| | 305 | | | | | 310 | | | | | 315 | | | | 320 | |
| ATA | ATG | GCT | TCT | CCT | GTA | GGG | TTT | TCG | GGG | CCA | GAA | TTC | ACT | TTT | CCG | 1008 |
| Ile | Met | Ala | Ser | Pro | Val | Gly | Phe | Ser | Gly | Pro | Glu | Phe | Thr | Phe | Pro | |
| | | | | 325 | | | | | | 330 | | | | | 335 | |
| CTA | TAT | GGA | ACT | ATG | GGA | AAT | GCA | GCT | CCA | CAA | CAA | CGT | ATT | GTT | GCT | 1056 |
| Leu | Tyr | Gly | Thr | Met | Gly | Asn | Ala | Ala | Pro | Gln | Gln | Arg | Ile | Val | Ala | |
| | | | 340 | | | | | | 345 | | | | | | 350 | |
| CAA | CTA | GGT | CAG | GGC | GTG | TAT | AGA | ACA | TTA | TCG | TCC | ACT | TTA | TAT | AGA | 1104 |
| Gln | Leu | Gly | Gln | Gly | Val | Tyr | Arg | Thr | Leu | Ser | Ser | Thr | Leu | Tyr | Arg | |
| | | 355 | | | | | | | 360 | | | | | | 365 | |
| AGA | CCT | TTT | AAT | ATA | GGG | ATA | AAT | AAT | CAA | CAA | CTA | TCT | GTT | CTT | GAC | 1152 |
| Arg | Pro | Phe | Asn | Ile | Gly | Ile | Asn | Asn | Gln | Gln | Leu | Ser | Val | Leu | Asp | |
| | | 370 | | | | | 375 | | | | | 380 | | | | |
| GGG | ACA | GAA | TTT | GCT | TAT | GGA | ACC | TCC | TCA | AAT | TTG | CCA | TCC | GCT | GTA | 1200 |
| Gly | Thr | Glu | Phe | Ala | Tyr | Gly | Thr | Ser | Ser | Asn | Leu | Pro | Ser | Ala | Val | |
| | 385 | | | | | 390 | | | | | 395 | | | | 400 | |
| TAC | AGA | AAA | AGC | GGA | ACG | GTA | GAT | TCG | CTG | GAT | GAA | ATA | CCG | CCA | CAG | 1248 |
| Tyr | Arg | Lys | Ser | Gly | Thr | Val | Asp | Ser | Leu | Asp | Glu | Ile | Pro | Pro | Gln | |
| | | | | 405 | | | | | | 410 | | | | | 415 | |
| AAT | AAC | AAC | GTG | CCA | CCT | AGG | CAA | GGA | TTT | AGT | CAT | CGA | TTA | AGC | CAT | 1296 |



| | | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|--|
| Asn | Asn | Asn | Val | Pro | Pro | Arg | Gln | Gly | Phe | Ser | His | Arg | Leu | Ser | His | | |
| | | | 420 | | | | | 425 | | | | | 430 | | | | |
| GTT | TCA | ATG | TTT | CGT | TCA | GGC | TTT | AGT | AAT | AGT | AGT | GTA | AGT | ATA | ATA | 1344 | |
| Val | Ser | Met | Phe | Arg | Ser | Gly | Phe | Ser | Asn | Ser | Ser | Val | Ser | Ile | Ile | | |
| | | 435 | | | | | 440 | | | | | 445 | | | | | |
| AGA | GCT | CCT | ATG | TTC | TCT | TGG | ATA | CAT | CGT | AGT | GCT | GAA | TTT | AAT | AAT | 1392 | |
| Arg | Ala | Pro | Met | Phe | Ser | Trp | Ile | His | Arg | Ser | Ala | Glu | Phe | Asn | Asn | | |
| | | 450 | | | | 455 | | | | | 460 | | | | | | |
| ATA | ATT | GCA | TCG | GAT | AGT | ATT | ACT | CAA | ATA | CCA | TTG | GTA | AAA | GCA | CAT | 1440 | |
| Ile | Ile | Ala | Ser | Asp | Ser | Ile | Thr | Gln | Ile | Pro | Leu | Val | Lys | Ala | His | | |
| 465 | | | | | 470 | | | | | 475 | | | | | 480 | | |
| ACA | CTT | CAG | TCA | GGT | ACT | ACT | GTT | GTA | AGA | GGG | CCC | GGG | TTT | ACG | GGA | 1488 | |
| Thr | Leu | Gln | Ser | Gly | Thr | Thr | Val | Val | Arg | Gly | Pro | Gly | Phe | Thr | Gly | | |
| | | | | 485 | | | | | 490 | | | | | 495 | | | |
| GGA | GAT | ATT | CTT | CGA | CGA | ACA | AGT | GGA | GGA | CCA | TTT | GCT | TAT | ACT | ATT | 1536 | |
| Gly | Asp | Ile | Leu | Arg | Arg | Thr | Ser | Gly | Gly | Pro | Phe | Ala | Tyr | Thr | Ile | | |
| | | | 500 | | | | | 505 | | | | | 510 | | | | |
| GTT | AAT | ATA | AAT | GGG | CAA | TTA | CCC | CAA | AGG | TAT | CGT | GCA | AGA | ATA | CGC | 1584 | |
| Val | Asn | Ile | Asn | Gly | Gln | Leu | Pro | Gln | Arg | Tyr | Arg | Ala | Arg | Ile | Arg | | |
| | | 515 | | | | 520 | | | | | | 525 | | | | | |
| TAT | GCC | TCT | ACT | ACA | AAT | CTA | AGA | ATT | TAC | GTA | ACG | GTT | GCA | GGT | GAA | 1632 | |
| Tyr | Ala | Ser | Thr | Thr | Asn | Leu | Arg | Ile | Tyr | Val | Thr | Val | Ala | Gly | Glu | | |
| | | 530 | | | | 535 | | | | | 540 | | | | | | |
| CGG | ATT | TTT | GCT | GGT | CAA | TTT | AAC | AAA | ACA | ATG | GAT | ACC | GGT | GAC | CCA | 1680 | |
| Arg | Ile | Phe | Ala | Gly | Gln | Phe | Asn | Lys | Thr | Met | Asp | Thr | Gly | Asp | Pro | | |
| 545 | | | | | 550 | | | | | 555 | | | | | 560 | | |
| TTA | ACA | TTC | CAA | TCT | TTT | AGT | TAC | GCA | ACT | ATT | AAT | ACA | GCT | TTT | ACA | 1728 | |
| Leu | Thr | Phe | Gln | Ser | Phe | Ser | Tyr | Ala | Thr | Ile | Asn | Thr | Ala | Phe | Thr | | |
| | | | | 565 | | | | | 570 | | | | | 575 | | | |
| TTC | CCA | ATG | AGC | CAG | AGT | AGT | TTC | ACA | GTA | GGT | GCT | GAT | ACT | TTT | AGT | 1776 | |
| Phe | Pro | Met | Ser | Gln | Ser | Ser | Phe | Thr | Val | Gly | Ala | Asp | Thr | Phe | Ser | | |
| | | | 580 | | | | | 585 | | | | | 590 | | | | |
| TCA | GGG | AAT | GAA | GTT | TAT | ATA | GAC | AGA | TTT | GAA | TTG | ATT | CCA | GTT | ACT | 1824 | |
| Ser | Gly | Asn | Glu | Val | Tyr | Ile | Asp | Arg | Phe | Glu | Leu | Ile | Pro | Val | Thr | | |
| | | 595 | | | | | 600 | | | | | 605 | | | | | |
| GCA | ACA | TTT | GAA | GCA | GAA | TAT | GAT | TTA | GAA | AGA | GCA | CAA | AAG | GCG | GTG | 1872 | |
| Ala | Thr | Phe | Glu | Ala | Glu | Tyr | Asp | Leu | Glu | Arg | Ala | Gln | Lys | Ala | Val | | |
| | | 610 | | | | 615 | | | | | | 620 | | | | | |
| AAT | GCG | CTG | TTT | ACT | TCT | ATA | AAC | CAA | ATA | GGG | ATA | AAA | ACA | GAT | GTG | 1920 | |
| Asn | Ala | Leu | Phe | Thr | Ser | Ile | Asn | Gln | Ile | Gly | Ile | Lys | Thr | Asp | Val | | |
| 625 | | | | | 630 | | | | | 635 | | | | | 640 | | |
| ACG | GAT | TAT | CAT | ATT | GAT | CAA | GTA | TCC | AAT | TTA | GTG | GAT | TGT | TTA | TCA | 1968 | |
| Thr | Asp | Tyr | His | Ile | Asp | Gln | Val | Ser | Asn | Leu | Val | Asp | Cys | Leu | Ser | | |



| 645 | | | | | 650 | | | | | 655 | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| GAT | GAA | TTT | TGT | CTG | GAT | GAA | AAG | CGA | GAA | TTG | TCC | GAG | AAA | GTC | AAA | 2016 |
| Asp | Glu | Phe | Cys | Leu | Asp | Glu | Lys | Arg | Glu | Leu | Ser | Glu | Lys | Val | Lys | |
| | | | 660 | | | | | 665 | | | | | 670 | | | |
| CAT | GCG | AAG | CGA | CTC | AGT | GAT | GAG | CGG | AAT | TTA | CTT | CAA | GAT | CCA | AAC | 2064 |
| His | Ala | Lys | Arg | Leu | Ser | Asp | Glu | Arg | Asn | Leu | Leu | Gln | Asp | Pro | Asn | |
| | | 675 | | | | | 680 | | | | | 685 | | | | |
| TTC | AAA | GGC | ATC | AAT | AGG | CAA | CTA | GAC | CGT | GGT | TGG | AGA | GGA | AGT | ACG | 2112 |
| Phe | Lys | Gly | Ile | Asn | Arg | Gln | Leu | Asp | Arg | Gly | Trp | Arg | Gly | Ser | Thr | |
| | 690 | | | | | 695 | | | | | 700 | | | | | |
| GAT | ATT | ACC | ATC | CAA | AGA | GGA | GAT | GAC | GTA | TTC | AAA | GAA | AAT | TAT | GTC | 2160 |
| Asp | Ile | Thr | Ile | Gln | Arg | Gly | Asp | Asp | Val | Phe | Lys | Glu | Asn | Tyr | Val | |
| 705 | | | | | 710 | | | | | 715 | | | | | 720 | |
| ACA | CTA | CCA | GGT | ACC | TTT | GAT | GAG | TGC | TAT | CCA | ACA | TAT | TTG | TAT | CAA | 2208 |
| Thr | Leu | Pro | Gly | Thr | Phe | Asp | Glu | Cys | Tyr | Pro | Thr | Tyr | Leu | Tyr | Gln | |
| | | | 725 | | | | | | 730 | | | | | | 735 | |
| AAA | ATC | GAT | GAA | TCA | AAA | TTA | AAA | GCC | TTT | ACC | CGT | TAT | CAA | TTA | AGA | 2256 |
| Lys | Ile | Asp | Glu | Ser | Lys | Leu | Lys | Ala | Phe | Thr | Arg | Tyr | Gln | Leu | Arg | |
| | | | 740 | | | | | 745 | | | | | | | 750 | |
| GGG | TAT | ATC | GAA | GAT | AGT | CAA | GAC | TTA | GAA | ATC | TAT | TTA | ATT | CGC | TAC | 2304 |
| Gly | Tyr | Ile | Glu | Asp | Ser | Gln | Asp | Leu | Glu | Ile | Tyr | Leu | Ile | Arg | Tyr | |
| | | 755 | | | | | 760 | | | | | | | | 765 | |
| AAT | GCA | AAA | CAT | GAA | ACA | GTA | AAT | GTG | CCA | GGT | ACG | GGT | TCC | TTA | TGG | 2352 |
| Asn | Ala | Lys | His | Glu | Thr | Val | Asn | Val | Pro | Gly | Thr | Gly | Ser | Leu | Trp | |
| | | 770 | | | | | 775 | | | | | | | | 780 | |
| CCG | CTT | TCA | GCC | CAA | AGT | CCA | ATC | GGA | AAG | TGT | GGA | GAG | CCG | AAT | CGA | 2400 |
| Pro | Leu | Ser | Ala | Gln | Ser | Pro | Ile | Gly | Lys | Cys | Gly | Glu | Pro | Asn | Arg | |
| | | | 785 | | | | 790 | | | | | | | | 800 | |
| TGC | GCG | CCA | CAC | CTT | GAA | TGG | AAT | CCT | GAC | TTA | GAT | TGT | TCG | TGT | AGG | 2448 |
| Cys | Ala | Pro | His | Leu | Glu | Trp | Asn | Pro | Asp | Leu | Asp | Cys | Ser | Cys | Arg | |
| | | | 805 | | | | | | | | | | | | 815 | |
| GAT | GGA | GAA | AAG | TGT | GCC | CAT | CAT | TCG | CAT | CAT | TTC | TCC | TTA | GAC | ATT | 2496 |
| Asp | Gly | Glu | Lys | Cys | Ala | His | His | Ser | His | His | Phe | Ser | Leu | Asp | Ile | |
| | | | 820 | | | | | | | | | | | | 830 | |
| GAT | GTA | GGA | TGT | ACA | GAC | TTA | AAT | GAG | GAC | CTA | GGT | GTA | TGG | GTG | ATC | 2544 |
| Asp | Val | Gly | Cys | Thr | Asp | Leu | Asn | Glu | Asp | Leu | Gly | Val | Trp | Val | Ile | |
| | | 835 | | | | | | | | | | | | | 845 | |
| TTT | AAG | ATT | AAG | ACG | CAA | GAT | GGG | CAC | GCA | AGA | CTA | GGG | AAT | CTA | GAG | 2592 |
| Phe | Lys | Ile | Lys | Thr | Gln | Asp | Gly | His | Ala | Arg | Leu | Gly | Asn | Leu | Glu | |
| | | 850 | | | | | | | | | | | | | 860 | |
| TTT | CTC | GAA | GAG | AAA | CCA | TTA | GTA | GGA | GAA | GCG | CTA | GCT | CGT | GTG | AAA | 2640 |
| Phe | Leu | Glu | Glu | Lys | Pro | Leu | Val | Gly | Glu | Ala | Leu | Ala | Arg | Val | Lys | |
| | | | 865 | | | | | | | | | | | | 880 | |



| | |
|---|------|
| AGA GCG GAG AAA AAA TGG AGA GAC AAA CGT GAA AAA TTG GAA TGG GAA | 2688 |
| Arg Ala Glu Lys Lys Trp Arg Asp Lys Arg Glu Lys Leu Glu Trp Glu | |
| 885 890 895 | |
| ACA AAT ATC GTT TAT AAA GAG GCA AAA GAA TCT GTA GAT GCT TTA TTT | 2736 |
| Thr Asn Ile Val Tyr Lys Glu Ala Lys Glu Ser Val Asp Ala Leu Phe | |
| 900 905 910 | |
| GTA AAC TCT CAA TAT GAT CAA TTA CAA GCG GAT ACG AAT ATT GCC ATG | 2784 |
| Val Asn Ser Gln Tyr Asp Gln Leu Gln Ala Asp Thr Asn Ile Ala Met | |
| 915 920 925 | |
| ATT CAT GCG GCA GAT AAA CGT GTT CAT AGC ATT CGA GAA GCT TAT CTG | 2832 |
| Ile His Ala Ala Asp Lys Arg Val His Ser Ile Arg Glu Ala Tyr Leu | |
| 930 935 940 | |
| CCT GAG CTG TCT GTG ATT CCG GGT GTC AAT GCG GCT ATT TTT GAA GAA | 2880 |
| Pro Glu Leu Ser Val Ile Pro Gly Val Asn Ala Ala Ile Phe Glu Glu | |
| 945 950 955 960 | |
| TTA GAA GGG CGT ATT TTC ACT GCA TTC TCC CTA TAT GAT GCG AGA AAT | 2928 |
| Leu Glu Gly Arg Ile Phe Thr Ala Phe Ser Leu Tyr Asp Ala Arg Asn | |
| 965 970 975 | |
| GTC ATT AAA AAT GGT GAT TTT AAT AAT GGC TTA TCC TGC TGG AAC GTG | 2976 |
| Val Ile Lys Asn Gly Asp Phe Asn Asn Gly Leu Ser Cys Trp Asn Val | |
| 980 985 990 | |
| AAA GGG CAT GTA GAT GTA GAA GAA CAA AAC AAC CAA CGT TCG GTC CTT | 3024 |
| Lys Gly His Val Asp Val Glu Glu Gln Asn Asn Gln Arg Ser Val Leu | |
| 995 1000 1005 | |
| GTT GTT CCG GAA TGG GAA GCA GAA GTG TCA CAA GAA GTT CGT GTC TGT | 3072 |
| Val Val Pro Glu Trp Glu Ala Glu Val Ser Gln Glu Val Arg Val Cys | |
| 1010 1015 1020 | |
| CCG GGT CGT GGC TAT ATC CTT CGT GTC ACA GCG TAC AAG GAG GGA TAT | 3120 |
| Pro Gly Arg Gly Tyr Ile Leu Arg Val Thr Ala Tyr Lys Glu Gly Tyr | |
| 1025 1030 1035 1040 | |
| GGA GAA GGT TGC GTA ACC ATT CAT GAG ATC GAG AAC AAT ACA GAC GAA | 3168 |
| Gly Glu Gly Cys Val Thr Ile His Glu Ile Glu Asn Asn Thr Asp Glu | |
| 1045 1050 1055 | |
| CTG AAG TTT AGC AAC TGC GTA GAA GAG GAA ATC TAT CCA AAT AAC ACG | 3216 |
| Leu Lys Phe Ser Asn Cys Val Glu Glu Glu Ile Tyr Pro Asn Asn Thr | |
| 1060 1065 1070 | |
| GTA ACG TGT AAT GAT TAT ACT GTA AAT CAA GAA GAA TAC GGA GGT GCG | 3264 |
| Val Thr Cys Asn Asp Tyr Thr Val Asn Gln Glu Glu Tyr Gly Gly Ala | |
| 1075 1080 1085 | |
| TAC ACT TCT CGT AAT CGA GGA TAT AAC GAA GCT CCT TCC GTA CCA GCT | 3312 |
| Tyr Thr Ser Arg Asn Arg Gly Tyr Asn Glu Ala Pro Ser Val Pro Ala | |
| 1090 1095 1100 | |

99.07.14

GAT TAT GCG TCA GTC TAT GAA GAA AAA TCG TAT ACA GAT GGA CGA AGA 3360
 Asp Tyr Ala Ser Val Tyr Glu Glu Lys Ser Tyr Thr Asp Gly Arg Arg
 1105 1110 1115 1120

GAG AAT CCT TGT GAA TTT AAC AGA GGG TAT AGG GAT TAC ACG CCA CTA 3408
 Glu Asn Pro Cys Glu Phe Asn Arg Gly Tyr Arg Asp Tyr Thr Pro Leu
 1125 1130 1135

CCA GTT GGT TAT GTG ACA AAA GAA TTA GAA TAC TTC CCA GAA ACC GAT 3456
 Pro Val Gly Tyr Val Thr Lys Glu Leu Glu Tyr Phe Pro Glu Thr Asp
 1140 1145 1150

AAG GTA TGG ATT GAG ATT GGA GAA ACG GAA GGA ACA TTT ATC GTG GAC 3504
 Lys Val Trp Ile Glu Ile Gly Glu Thr Glu Gly Thr Phe Ile Val Asp
 1155 1160 1165

10 AGC GTG GAA TTA CTC CTT ATG GAG GAA TAG 3534
 Ser Val Glu Leu Leu Leu Met Glu Glu
 1170 1175

(2) SEQ ID NO: 28 信息:

(i) 序列特征:

- 15 (A) 长度: 1177 个氨基酸
 (B) 类型: 氨基酸
 (D) 拓扑学: 线性

(ii) 分子类型: 蛋白

(xi) 序列描述: SEQ ID NO: 28:

Met Asp Asn Asn Pro Asn Ile Asn Glu Cys Ile Pro Tyr Asn Cys Leu
 1 5 10 15

Ser Asn Pro Glu Val Glu Val Leu Gly Gly Glu Arg Ile Glu Thr Gly
 20 25 30

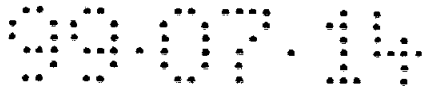
Tyr Thr Pro Ile Asp Ile Ser Leu Ser Leu Thr Gln Phe Leu Leu Ser
 35 40 45

Glu Phe Val Pro Gly Ala Gly Phe Val Leu Gly Leu Val Asp Ile Ile
 50 55 60

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

Glu Gln Leu Ile Asn Gln Arg Ile Glu Glu Phe Ala Arg Asn Gln Ala
 85 90 95

Ile Ser Arg Leu Glu Gly Leu Ser Asn Leu Tyr Gln Ile Tyr Ala Glu
 100 105 110



Ser Phe Arg Glu Trp Glu Ala Asp Pro Thr Asn Pro Ala Leu Arg Glu
115 120 125

Glu Met Arg Ile Gln Phe Asn Asp Met Asn Ser Ala Leu Thr Thr Ala
130 135 140

Ile Pro Leu Phe Ala Val Gln Asn Tyr Gln Val Pro Leu Leu Ser Val
145 150 155 160

Tyr Val Gln Ala Ala Asn Leu His Leu Ser Val Leu Arg Asp Val Ser
165 170 175

Val Phe Gly Gln Arg Trp Gly Phe Asp Ala Ala Thr Ile Asn Ser Arg
180 185 190

Tyr Asn Asp Leu Thr Arg Leu Ile Gly Asn Tyr Thr Asp Tyr Ala Val
195 200 205

Arg Trp Tyr Asn Thr Gly Leu Glu Arg Val Trp Gly Pro Asp Ser Arg
210 215 220

Asp Trp Val Arg Tyr Asn Gln Phe Arg Arg Glu Leu Thr Leu Thr Val
225 230 235 240

Leu Asp Ile Val Ala Leu Phe Pro Asn Tyr Asp Ser Arg Arg Tyr Pro
245 250 255

Ile Arg Thr Val Ser Gln Leu Thr Arg Glu Ile Tyr Thr Asn Pro Val
260 265 270

Leu Glu Asn Phe Asp Gly Ser Phe Arg Gly Ser Ala Gln Gly Ile Glu
275 280 285

Arg Ser Ile Arg Ser Pro His Leu Met Asp Ile Leu Asn Ser Ile Thr
290 295 300

Ile Tyr Thr Asp Ala His Arg Gly Tyr Tyr Tyr Trp Ser Gly His Gln
305 310 315 320

Ile Met Ala Ser Pro Val Gly Phe Ser Gly Pro Glu Phe Thr Phe Pro
325 330 335

Leu Tyr Gly Thr Met Gly Asn Ala Ala Pro Gln Gln Arg Ile Val Ala
340 345 350

Gln Leu Gly Gln Gly Val Tyr Arg Thr Leu Ser Ser Thr Leu Tyr Arg
355 360 365

Arg Pro Phe Asn Ile Gly Ile Asn Asn Gln Gln Leu Ser Val Leu Asp
370 375 380



Gly Thr Glu Phe Ala Tyr Gly Thr Ser Ser Asn Leu Pro Ser Ala Val
385 390 395 400

Tyr Arg Lys Ser Gly Thr Val Asp Ser Leu Asp Glu Ile Pro Pro Gln
405 410 415

Asn Asn Asn Val Pro Pro Arg Gln Gly Phe Ser His Arg Leu Ser His
420 425 430

Val Ser Met Phe Arg Ser Gly Phe Ser Asn Ser Ser Val Ser Ile Ile
435 440 445

Arg Ala Pro Met Phe Ser Trp Ile His Arg Ser Ala Glu Phe Asn Asn
450 455 460

Ile Ile Ala Ser Asp Ser Ile Thr Gln Ile Pro Leu Val Lys Ala His
465 470 475 480

Thr Leu Gln Ser Gly Thr Thr Val Val Arg Gly Pro Gly Phe Thr Gly
485 490 495

Gly Asp Ile Leu Arg Arg Thr Ser Gly Gly Pro Phe Ala Tyr Thr Ile
500 505 510

Val Asn Ile Asn Gly Gln Leu Pro Gln Arg Tyr Arg Ala Arg Ile Arg
515 520 525

Tyr Ala Ser Thr Thr Asn Leu Arg Ile Tyr Val Thr Val Ala Gly Glu
530 535 540

Arg Ile Phe Ala Gly Gln Phe Asn Lys Thr Met Asp Thr Gly Asp Pro
545 550 555 560

Leu Thr Phe Gln Ser Phe Ser Tyr Ala Thr Ile Asn Thr Ala Phe Thr
565 570 575

Phe Pro Met Ser Gln Ser Ser Phe Thr Val Gly Ala Asp Thr Phe Ser
580 585 590

Ser Gly Asn Glu Val Tyr Ile Asp Arg Phe Glu Leu Ile Pro Val Thr
595 600 605

Ala Thr Phe Glu Ala Glu Tyr Asp Leu Glu Arg Ala Gln Lys Ala Val
610 615 620

Asn Ala Leu Phe Thr Ser Ile Asn Gln Ile Gly Ile Lys Thr Asp Val
625 630 635 640

Thr Asp Tyr His Ile Asp Gln Val Ser Asn Leu Val Asp Cys Leu Ser
645 650 655

Asp Glu Phe Cys Leu Asp Glu Lys Arg Glu Leu Ser Glu Lys Val Lys
660 665 670

His Ala Lys Arg Leu Ser Asp Glu Arg Asn Leu Leu Gln Asp Pro Asn
675 680 685

907.14

Phe Lys Gly Ile Asn Arg Gln Leu Asp Arg Gly Trp Arg Gly Ser Thr
 690 695 700

Asp Ile Thr Ile Gln Arg Gly Asp Asp Val Phe Lys Glu Asn Tyr Val
 705 710 715 720

Thr Leu Pro Gly Thr Phe Asp Glu Cys Tyr Pro Thr Tyr Leu Tyr Gln
 725 730 735

Lys Ile Asp Glu Ser Lys Leu Lys Ala Phe Thr Arg Tyr Gln Leu Arg
 740 745 750

Gly Tyr Ile Glu Asp Ser Gln Asp Leu Glu Ile Tyr Leu Ile Arg Tyr
 755 760 765

Asn Ala Lys His Glu Thr Val Asn Val Pro Gly Thr Gly Ser Leu Trp
 770 775 780

Pro Leu Ser Ala Gln Ser Pro Ile Gly Lys Cys Gly Glu Pro Asn Arg
 785 790 795 800

Cys Ala Pro His Leu Glu Trp Asn Pro Asp Leu Asp Cys Ser Cys Arg
 805 810 815

Asp Gly Glu Lys Cys Ala His His Ser His His Phe Ser Leu Asp Ile
 820 825 830

Asp Val Gly Cys Thr Asp Leu Asn Glu Asp Leu Gly Val Trp Val Ile
 835 840 845

Phe Lys Ile Lys Thr Gln Asp Gly His Ala Arg Leu Gly Asn Leu Glu
 850 855 860

Phe Leu Glu Glu Lys Pro Leu Val Gly Glu Ala Leu Ala Arg Val Lys
 865 870 875 880

Arg Ala Glu Lys Lys Trp Arg Asp Lys Arg Glu Lys Leu Glu Trp Glu
 885 890 895

Thr Asn Ile Val Tyr Lys Glu Ala Lys Glu Ser Val Asp Ala Leu Phe
 900 905 910

Val Asn Ser Gln Tyr Asp Gln Leu Gln Ala Asp Thr Asn Ile Ala Met
 915 920 925

Ile His Ala Ala Asp Lys Arg Val His Ser Ile Arg Glu Ala Tyr Leu
 930 935 940

Pro Glu Leu Ser Val Ile Pro Gly Val Asn Ala Ala Ile Phe Glu Glu
 945 950 955 960

Leu Glu Gly Arg Ile Phe Thr Ala Phe Ser Leu Tyr Asp Ala Arg Asn
 965 970 975

Val Ile Lys Asn Gly Asp Phe Asn Asn Gly Leu Ser Cys Trp Asn Val
 980 985 990

Lys Gly His Val Asp Val Glu Glu Gln Asn Asn Gln Arg Ser Val Leu
 995 1000 1005
 Val Val Pro Glu Trp Glu Ala Glu Val Ser Gln Glu Val Arg Val Cys
 1010 1015 1020
 5 Pro Gly Arg Gly Tyr Ile Leu Arg Val Thr Ala Tyr Lys Glu Gly Tyr
 1025 1030 1035 1040
 Gly Glu Gly Cys Val Thr Ile His Glu Ile Glu Asn Asn Thr Asp Glu
 1045 1050 1055
 Leu Lys Phe Ser Asn Cys Val Glu Glu Glu Ile Tyr Pro Asn Asn Thr
 1060 1065 1070
 10 Val Thr Cys Asn Asp Tyr Thr Val Asn Gln Glu Glu Tyr Gly Gly Ala
 1075 1080 1085
 Tyr Thr Ser Arg Asn Arg Gly Tyr Asn Glu Ala Pro Ser Val Pro Ala
 1090 1095 1100
 Asp Tyr Ala Ser Val Tyr Glu Glu Lys Ser Tyr Thr Asp Gly Arg Arg
 1105 1110 1115 1120
 15 Glu Asn Pro Cys Glu Phe Asn Arg Gly Tyr Arg Asp Tyr Thr Pro Leu
 1125 1130 1135
 Pro Val Gly Tyr Val Thr Lys Glu Leu Glu Tyr Phe Pro Glu Thr Asp
 1140 1145 1150
 Lys Val Trp Ile Glu Ile Gly Glu Thr Glu Gly Thr Phe Ile Val Asp
 1155 1160 1165
 Ser Val Glu Leu Leu Leu Met Glu Glu
 1170 1175

20

(2) SEQ ID NO : 29 信息:

(i) 序列特征:

(A) 长度: 3579 个碱基对

(B) 类型: 核酸

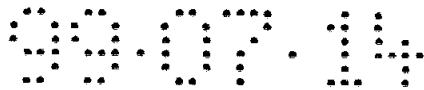
(C) 链型: 单链

(D) 拓扑学: 线性

(xi) 序列描述: SEQ ID NO : 29

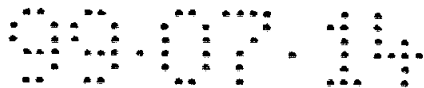
25

ATGGATAACA ATCCGAACAT CAATGAATGC ATTCCTTATA ATTGTTTAAG TAACCCTGAA 60
 GTAGAAGTAT TAGGTGGAGA AAGAATAGAA ACTGGTTACA CCCCAATCGA TATTTCTTGG 120
 TCGCTAACGC AATTTCTTTT GAGTGAATTT GTTCCCGGTG CTGGATTGT GTTAGGACTA 180



GTTGATATAA TATGGGGAAT TTTTGGTCCC TCTCAATGGG ACGCATTCT TGTACAAATT 240
GAACAGTTAA TTAACCAAAG AATAGAAGAA TTCGCTAGGA ACCAAGCCAT TTCTAGATTA 300
GAAGGACTAA GCAATCTTTA TCAAATTTAC GCAGAATCTT TTAGAGAGTG GGAAGCAGAT 360
CCTACTAATC CAGCATTAAAG AGAAGAGATG CGTATTCAAT TCAATGACAT GAACAGTGCC 420
CTTACAACCG CTATTCCTCT TTTTGCAGTT CAAAATTATC AAGTTCCTCT TTTATCAGTA 480
TATGTTCAAG CTGCAAATTT ACATTTATCA GTTTGAGAG ATGTTTCAGT GTTTGGACAA 540
AGGTGGGGAT TTGATGCCGC GACTATCAAT AGTCGTTATA ATGATTTAAC TAGGCTTATT 600
GGCAACTATA CAGATTATGC TGTACGCTGG TACAATACGG GATTAGAACG TGTATGGGGA 660

CCGGATTCTA GAGATTGGGT AAGGTATAAT CAATTTAGAA GAGAATTAAC ACTAACTGTA 720
TTAGATATCG TTGCTCTGTT CCCGAATTAT GATAGTAGAA GATATCCAAT TCGAACAGTT 780
TCCCAATTAA CAAGAGAAAT TTATACAAAC CCAGTATTAG AAAATTTTGA TGGTAGTTTT 840
CGAGGCTCGG CTCAGGGCAT AGAAAGAAGT ATTAGGAGTC CACATTTGAT GGATATACTT 900
AACAGTATAA CCATCTATAC GGATGCTCAT AGGGGTTATT ATTATTGGTC AGGGCATCAA 960
ATAATGGCTT CTCCTGTAGG GTTTTCGGGG CCAGAATTCA CTTTTCCGCT ATATGGAACT 1020
ATGGGAAATG CAGCTCCACA ACAACGTATT GTTGCTCAAC TAGGTCAGGG CGTGTATAGA 1080
ACATTATCGT CCACTTTATA TAGAAGACCT TTTAATATAG GGATAAATAA TCAACAACATA 1140
TCTGTTCTTG ACGGGACAGA ATTTGCTTAT GGAACCTCCT CAAATTTGCC ATCCGCTGTA 1200
TACAGAAAAA GCGGAACGGT AGATTGCTG GATGAAATAC CGCCACAGAA TAACAACGTG 1260
CCACCTAGGC AAGGATTTAG TCATCGATTA AGCCATGTTT CAATGTTTCG TTCAGGCTTT 1320
AGTAATAGTA GTGTAAGTAT AATAAGAGCT CCTATGTTCT CTTGGATACA TCGTAGTGCA 1380
ACTCTTACAA ATACAATTGA TCCAGAGAGA ATTAATCAAA TACCTTTAGT GAAAGGATTT 1440
AGAGTTTGGG GGGGCACCTC TGTCATTACA GGACCAGGAT TTACAGGAGG GGATATCCTT 1500
CGAAGAAATA CCTTTGGTGA TTTTGTATCT CTACAAGTCA ATATTAATTC ACCAATTACC 1560
CAAAGATACC GTTTAAGATT TCGTTACGCT TCCAGTAGGG ATGCACGAGT TATAGTATTA 1620
ACAGGAGCGG CATCCACAGG AGTGGGAGGC CAAGTTAGTG TAAATATGCC TCTTCAGAAA 1680
ACTATGGAAA TAGGGGAGAA CTTAACATCT AGAACATTTA GATATACCGA TTTTAGTAAT 1740
CCTTTTTCAT TTAGAGCTAA TCCAGATATA ATTGGGATAA GTGAACAACC TCTATTTGGT 1800
GCAGGTTCTA TTAGTAGCGG TGAACTTTAT ATAGATAAAA TTGAAATTAT TCTAGCAGAT 1860



GCAACATTTG AAGCAGAATC TGATTAGAA AGAGCACAAA AGGCGGTGAA TGCCCTGTTT 1920
ACTTCTTCCA ATCAAATCGG GTTAAAAACC GATGTGACGG ATTATCATAT TGATCAAGTA 1980
TCCAATTTAG TGGATTGTTT ATCAGATGAA TTTTGTCTGG ATGAAAAGCG AGAATTGTCC 2040
GAGAAAGTCA AACATGCGAA GCGACTCAGT GATGAGCGGA ATTTACTTCA AGATCCAAAC 2100
TTCAGAGGGA TCAATAGACA ACCAGACCGT GGCTGGAGAG GAAGTACAGA TATTACCATC 2160
CAAGGAGGAG ATGACGTATT CAAAGAGAAT TACGTACACAC TACCGGGTAC CGTTGATGAG 2220
TGCTATCCAA CGTATTTATA TCAGAAAATA GATGAGTCGA AATTAAAAGC TTATACCCGT 2280
TATGAATTAA GAGGGTATAT CGAAGATAGT CAAGACTTAG AAATCTATTT GATCCGTTAC 2340

AATGCAAAAC ACGAAATAGT AAATGTGCCA GGCACGGGTT CCTTATGGCC GCTTTCAGCC 2400
CAAAGTCCAA TCGGAAAGTG TGGAGAACCG AATCGATGCG CGCCACACCT TGAATGGAAT 2460
CCTGATCTAG ATTGTTCCCTG CAGAGACGGG GAAAATGTG CACATCATT C CATCATTTTC 2520
ACCTTGGATA TTGATGTTGG ATGTACAGAC TTAAATGAGG ACTTAGGTGT ATGGGTGATA 2580
TTCAAGATTA AGACGCAAGA TGGCCATGCA AGACTAGGGA ATCTAGAGTT TCTCGAAGAG 2640
AAACCATTAT TAGGGGAAGC ACTAGCTCGT GTGAAAAGAG CGGAGAAGAA GTGGAGAGAC 2700
AAACGAGAGA AACTGCAGTT GGAAACAAAT ATTGTTTATA AAGAGGCAAA AGAATCTGTA 2760
GATGCTTTAT TTGTAACTC TCAATATGAT AGATTACAAG TGGATACGAA CATCGCAATG 2820
ATTCATGCGG CAGATAAACG CGTTCATAGA ATCCGGGAAG CGTATCTGCC AGAGTTGTCT 2880
GTGATTCCAG GTGTCAATGC GGCCATTTTC GAAGAATTAG AGGGACGTAT TTTTACAGCG 2940
TATTCCTTAT ATGATGCGAG AAATGTCATT AAAAATGGCG ATTTCAATAA TGGCTTATTA 3000
TGCTGGAACG TGAAAGGTCA TGTAGATGTA GAAGAGCAAA ACAACCACCG TTCGGTCCTT 3060
GTTATCCCAG AATGGGAGGC AGAAGTGTC AAGAGGTTT GTGTCTGTCC AGGTCGTGGC 3120
TATATCCTTC GTGTACAGC ATATAAAGAG GGATATGGAG AGGGCTGCGT AACGATCCAT 3180
GAGATCGAAG ACAATACAGA CGAACTGAAA TTCAGCAACT GTGTAGAAGA GGAAGTATAT 3240
CCAAACAACA CAGTAACGTG TAATAATTAT ACTGGGACTC AAGAAGAATA TGAGGGTACG 3300
TACACTTCTC GTAATCAAGG ATATGACGAA GCCTATGSTA ATAACCCTTC CGTACCAGCT 3360
GATTACGCTT CAGTCTATGA AGAAAAATCG TATACAGATG GACGAAGAGA GAATCCTTGT 3420
GAATCTAACA GAGGCTATGG GGATTACACA CCACTACCGG CTGGTTATGT AACAAAGGAT 3480
TTAGAGTACT TCCCAGAGAC CGATAAGGTA TGGATTGAGA TCGGAGAAAC AGAAGGAACA 3540
TTCATCGTGG ATAGCGTGGA ATTACTCCTT ATGGAGGAA 3579

(2) SEQ ID NO: 30 信息:

(i) 序列特征:

(A) 长度: 1193 个氨基酸

(B) 类型: 氨基酸

(C) 链型:

(D) 拓扑学: 线性

(xi) 序列描述: SEQ ID NO: 30:

```

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

```

907.14

Arg Trp Tyr Asn Thr Gly Leu Glu Arg Val Trp Gly Pro Asp Ser Arg
 210 215 220

Asp Trp Val Arg Tyr Asn Gln Phe Arg Arg Glu Leu Thr Leu Thr Val
 225 230 235 240

Leu Asp Ile Val Ala Leu Phe Pro Asn Tyr Asp Ser Arg Arg Tyr Pro
 245 250 255

Ile Arg Thr Val Ser Gln Leu Thr Arg Glu Ile Tyr Thr Asn Pro Val
 260 265 270

Leu Glu Asn Phe Asp Gly Ser Phe Arg Gly Ser Ala Gln Gly Ile Glu
 275 280 285

Arg Ser Ile Arg Ser Pro His Leu Met Asp Ile Leu Asn Ser Ile Thr
 290 295 300

Ile Tyr Thr Asp Ala His Arg Gly Tyr Tyr Trp Ser Gly His Gln
 305 310 315 320

Ile Met Ala Ser Pro Val Gly Phe Ser Gly Pro Glu Phe Thr Phe Pro
 325 330 335

Leu Tyr Gly Thr Met Gly Asn Ala Ala Pro Gln Gln Arg Ile Val Ala
 340 345 350

Gln Leu Gly Gln Gly Val Tyr Arg Thr Leu Ser Ser Thr Leu Tyr Arg
 355 360 365

Arg Pro Phe Asn Ile Gly Ile Asn Asn Gln Gln Leu Ser Val Leu Asp
 370 375 380

Gly Thr Glu Phe Ala Tyr Gly Thr Ser Ser Asn Leu Pro Ser Ala Val
 385 390 395 400

Tyr Arg Lys Ser Gly Thr Val Asp Ser Leu Asp Glu Ile Pro Pro Gln
 405 410 415

Asn Asn Asn Val Pro Pro Arg Gln Gly Phe Ser His Arg Leu Ser His
 420 425 430

Val Ser Met Phe Arg Ser Gly Phe Ser Asn Ser Ser Val Ser Ile Ile
 435 440 445

Arg Ala Pro Met Phe Ser Trp Ile His Arg Ser Ala Thr Leu Thr Asn
 450 455 460

Thr Ile Asp Pro Glu Arg Ile Asn Gln Ile Pro Leu Val Lys Gly Phe
 465 470 475 480

Arg Val Trp Gly Gly Thr Ser Val Ile Thr Gly Pro Gly Phe Thr Gly
 485 490 495

00714

Gly Asp Ile Leu Arg Arg Asn Thr Phe Gly Asp Phe Val Ser Leu Gln
 500 505 510

Val Asn Ile Asn Ser Pro Ile Thr Gln Arg Tyr Arg Leu Arg Phe Arg
 515 520 525

Tyr Ala Ser Ser Arg Asp Ala Arg Val Ile Val Leu Thr Gly Ala Ala
 530 535 540

Ser Thr Gly Val Gly Gly Gln Val Ser Val Asn Met Pro Leu Gln Lys
 545 550 555 560

Thr Met Glu Ile Gly Glu Asn Leu Thr Ser Arg Thr Phe Arg Tyr Thr
 565 570 575

Asp Phe Ser Asn Pro Phe Ser Phe Arg Ala Asn Pro Asp Ile Ile Gly
 580 585 590

Ile Ser Glu Gln Pro Leu Phe Gly Ala Gly Ser Ile Ser Ser Gly Glu
 595 600 605

Leu Tyr Ile Asp Lys Ile Glu Ile Ile Leu Ala Asp Ala Thr Phe Glu
 610 615 620

Ala Glu Ser Asp Leu Glu Arg Ala Gln Lys Ala Val Asn Ala Leu Phe
 625 630 635 640

Thr Ser Ser Asn Gln Ile Gly Leu Lys Thr Asp Val Thr Asp Tyr His
 645 650 655

Ile Asp Gln Val Ser Asn Leu Val Asp Cys Leu Ser Asp Glu Phe Cys
 660 665 670

Leu Asp Glu Lys Arg Glu Leu Ser Glu Lys Val Lys His Ala Lys Arg
 675 680 685

Leu Ser Asp Glu Arg Asn Leu Leu Gln Asp Pro Asn Phe Arg Gly Ile
 690 695 700

Asn Arg Gln Pro Asp Arg Gly Trp Arg Gly Ser Thr Asp Ile Thr Ile
 705 710 715 720

Gln Gly Gly Asp Asp Val Phe Lys Glu Asn Tyr Val Thr Leu Pro Gly
 725 730 735

Thr Val Asp Glu Cys Tyr Pro Thr Tyr Leu Tyr Gln Lys Ile Asp Glu
 740 745 750

Ser Lys Leu Lys Ala Tyr Thr Arg Tyr Glu Leu Arg Gly Tyr Ile Glu
 755 760 765

Asp Ser Gln Asp Leu Glu Ile Tyr Leu Ile Arg Tyr Asn Ala Lys His
 770 775 780

Glu Ile Val Asn Val Pro Gly Thr Gly Ser Leu Trp Pro Leu Ser Ala
 785 790 795 800

907.14

Gln Ser Pro Ile Gly Lys Cys Gly Glu Pro Asn Arg Cys Ala Pro His
805 810 815

Leu Glu Trp Asn Pro Asp Leu Asp Cys Ser Cys Arg Asp Gly Glu Lys
820 825 830

Cys Ala His His Ser His His Phe Thr Leu Asp Ile Asp Val Gly Cys
835 840 845

Thr Asp Leu Asn Glu Asp Leu Gly Val Trp Val Ile Phe Lys Ile Lys
850 855 860

Thr Gln Asp Gly His Ala Arg Leu Gly Asn Leu Glu Phe Leu Glu Glu
865 870 875 880

Lys Pro Leu Leu Gly Glu Ala Leu Ala Arg Val Lys Arg Ala Glu Lys
885 890 895

Lys Trp Arg Asp Lys Arg Glu Lys Leu Gln Leu Glu Thr Asn Ile Val
900 905 910

Tyr Lys Glu Ala Lys Glu Ser Val Asp Ala Leu Phe Val Asn Ser Gln
915 920 925

Tyr Asp Arg Leu Gln Val Asp Thr Asn Ile Ala Met Ile His Ala Ala
930 935 940

Asp Lys Arg Val His Arg Ile Arg Glu Ala Tyr Leu Pro Glu Leu Ser
945 950 955 960

Val Ile Pro Gly Val Asn Ala Ala Ile Phe Glu Glu Leu Glu Gly Arg
965 970 975

Ile Phe Thr Ala Tyr Ser Leu Tyr Asp Ala Arg Asn Val Ile Lys Asn
980 985 990

Gly Asp Phe Asn Asn Gly Leu Leu Cys Trp Asn Val Lys Gly His Val
995 1000 1005

Asp Val Glu Glu Gln Asn Asn His Arg Ser Val Leu Val Ile Pro Glu
1010 1015 1020

Trp Glu Ala Glu Val Ser Gln Glu Val Arg Val Cys Pro Gly Arg Gly
1025 1030 1035 1040

Tyr Ile Leu Arg Val Thr Ala Tyr Lys Glu Gly Tyr Gly Glu Gly Cys
1045 1050 1055

Val Thr Ile His Glu Ile Glu Asp Asn Thr Asp Glu Leu Lys Phe Ser
1060 1065 1070

Asn Cys Val Glu Glu Glu Val Tyr Pro Asn Asn Thr Val Thr Cys Asn
1075 1080 1085

Asn Tyr Thr Gly Thr Gln Glu Glu Tyr Glu Gly Thr Tyr Thr Ser Arg
1090 1095 1100

Asn Gln Gly Tyr Asp Glu Ala Tyr Gly Asn Asn Pro Ser Val Pro Ala
1105 1110 1115 1120

Asp Tyr Ala Ser Val Tyr Glu Glu Lys Ser Tyr Thr Asp Gly Arg Arg
1125 1130 1135

5

Glu Asn Pro Cys Glu Ser Asn Arg Gly Tyr Gly Asp Tyr Thr Pro Leu
1140 1145 1150

Pro Ala Gly Tyr Val Thr Lys Asp Leu Glu Tyr Phe Pro Glu Thr Asp
1155 1160 1165

Lys Val Trp Ile Glu Ile Gly Glu Thr Glu Gly Thr Phe Ile Val Asp
1170 1175 1180

10

Ser Val Glu Leu Leu Leu Met Glu Glu
1185 1190

(2) SEQ ID NO : 31 信息:

(i) 序列特征:

(A) 长度: 16 个碱基对

15

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑学: 线性

(xi) 序列描述: SEQ ID NO : 31 :

CGTTGCTCTG TTCCCG

16

20

(2) SEQ ID NO : 32 信息:

(i) 序列特征:

(A) 长度: 20 个碱基对

(B) 类型: 核酸

25

(C) 链型: 单链

(D) 拓扑学: 线性

(xi) 序列描述: SEQ ID NO : 32 :

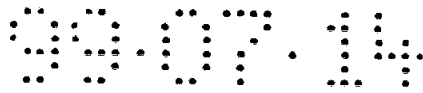
TCAAATACCA TTGGTAAAAG

20

30

(2) SEQ ID NO : 33 信息:

(i) 序列特征:



(A) 长度: 3534 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑学: 线性

5

(xi) 序列描述: SEQ ID NO: 33:

```
ATGGATAACA ATCCGAACAT CAATGAATGC ATTCCTTATA ATTGTTTAAG TAACCCTGAA      60
GTAGAAGTAT TAGGTGGAGA AAGAATAGAA ACTGGTTACA CCCCAATCGA TATTTCTTG      120
TCGCTAACGC AATTTCTTTT GAGTGAATTT GTTCCCGGTG CTGGATTTGT GTTAGGACTA      180
GTTGATATAA TATGGGGAAT TTTTGGTCCC TCTCAATGGG ACGCATTCT TGTACAAATT      240
GAACAGTTAA TTAACCAAAG AATAGAAGAA TTCGCTAGGA ACCAAGCCAT TTCTAGATTA      300
GAAGGACTAA GCAATCTTTA TCAAATTTAC GCAGAATCTT TTAGAGAGTG GGAAGCAGAT      360
CCTACTAATC CAGCATTAAG AGAAGAGATG CGTATTCAAT TCAATGACAT GAACAGTGCC      420
CTTACAACCG CTATTCCTCT TTTTGCAGTT CAAAATTATC AAGTTCCTCT TTTATCAGTA      480
TATGTTCAAG CTGCAAATTT ACATTATCA GTTTTGAGAG ATGTTTCAGT GTTTGGACAA      540
AGGTGGGGAT TTGATGCCGC GACTATCAAT AGTCGTTATA ATGATTTAAC TAGGCTTATT      600
GGCAACTATA CAGATTATGC TGTACGCTGG TACAATACGG GATTAGAACG TGTATGGGGA      660
CCGGATTCTA GAGATTGGGT AAGGTATAAT CAATTTAGAA GAGAATTAAC ACTAACTGTA      720
TTAGATATCG TTGCTCTGTT CCCGAATTAT GATAGTAGAA GATATCCAAT TCGAACAGTT      780
TCCCAATTAA CAAGAGAAAT TTATACAAAC CCAGTATTAG AAAATTTTGA TGGTAGTTTT      840

CGAGGCTCGG CTCAGGGCAT AGAAAGAAGT ATTAGGAGTC CACATTTGAT GGATATACTT      900
AACAGTATAA CCATCTATAC GGATGCTCAT AGGGGTTATT ATTATTGGTC AGGGCATCAA      960
ATAATGGCTT CTCCTGTAGG GTTTTCGGGG CCAGAATTCA CTTTCCGCT ATATGGAACT     1020
ATGGGAAATG CAGCTCCACA ACAACGTATT GTTGCTCAAC TAGGTCAGGG CGTGTATAGA     1080
ACATTATCGT CCACTTTATA TAGAAGACCT TTTAATATAG GGATAAATAA TCAACAACATA     1140
TCTGTTCTTG ACGGGACAGA ATTTGCTTAT GGAACCTCCT CAAATTTGCC ATCCGCTGTA     1200
TACAGAAAAA GCGGAACGGT AGATTGCTG GATGAAATAC CGCCACAGAA TAACAACGTG     1260
CCACCTAGGC AAGGATTTAG TCATCGATTA AGCCATGTTT CAATGTTTCG TTCAGGCTTT     1320
AGTAATAGTA GTGTAAGTAT AATAAGAGCT CCTATGTTCT CTTGGATACA TCGTAGTGCT     1380
```



GAATTTAATA ATATAATTGC ATCGGATAGT ATTACTCAAA TACCATTGGT AAAAGCACAT 1440
ACACTTCAGT CAGGTACTAC TGTTGTAAGA GGGCCCCGGT TTACGGGAGG AGATATTCTT 1500
CGACGAACAA GTGGAGGACC ATTTGCTTAT ACTATTGTTA ATATAAATGG GCAATTACCC 1560
CAAAGGTATC GTGCAAGAAT ACGCTATGCC TCTACTACAA ATCTAAGAAT TTACGTAACG 1620
GTTGCAGGTG AACGGATTTT TGCTGGTCAA TTTAACAAAA CAATGGATAC CGGTGACCCA 1680
TTAACATTCC AATCTTTTAG TTACGCAACT ATTAATACAG CTTTTACATT CCCAATGAGC 1740
CAGAGTAGTT TCACAGTAGG TGCTGATACT TTTAGTTCAG GGAATGAAGT TTATATAGAC 1800
AGATTTGAAT TGATTCCAGT TACTGCAACA CTCGAGGCTG AATATAATCT GGAAAGAGCG 1860
CAGAAGGCGG TGAATGCGCT GTTTACGTCT ACAAACCAAC TAGGGCTAAA AACAAATGTA 1920
ACGGATTATC ATATTGATCA AGTGTCCAAT TTAGTTACGT ATTTATCGGA TGAATTTTGT 1980
CTGGATGAAA AGCGAGAATT GTCCGAGAAA GTCAAACATG CGAAGCGACT CAGTGATGAA 2040
CGCAATTTAC TCCAAGATTC AAATTTCAAA GACATTAATA GGCAACCAGA ACGTGGGTGG 2100
GGCGGAAGTA CAGGGATTAC CATCCAAGGA GGGGATGACG TATTTAAAGA AAATTACGTC 2160
ACACTATCAG GTACCTTTGA TGAGTGCTAT CCAACATATT TGTATCAAAA AATCGATGAA 2220
TCAAAATTAA AAGCCTTTAC CCGTTATCAA TTAAGAGGGT ATATCGAAGA TAGTCAAGAC 2280
TTAGAAATCT ATTTAATTG CTACAATGCA AAACATGAAA CAGTAAATGT GCCAGGTACG 2340
GGTTCCTTAT GGCCGCTTTC AGCCCAAAGT CCAATCGGAA AGTGTGGAGA GCCGAATCGA 2400
TGCGCGCCAC ACCTTGAATG GAATCCTGAC TTAGATTGTT CGTGTAGGGA TGGAGAAAAG 2460
TGTGCCCATC ATTCGCATCA TTTCTCCTTA GACATTGATG TAGGATGTAC AGACTTAAAT 2520

GAGGACCTAG GTGTATGGGT GATCTTTAAG ATTAAGACGC AAGATGGGCA CGCAAGACTA 2580
GGGAATCTAG AGTTTCTCGA AGAGAAACCA TTAGTAGGAG AAGCGCTAGC TCGTGTGAAA 2640
AGAGCGGAGA AAAAATGGAG AGACAAACGT GAAAAATTGG AATGGGAAAC AAATATCGTT 2700
TATAAAGAGG CAAAAGAATC TGTAGATGCT TTATTTGTAA ACTCTCAATA TGATCAATTA 2760
CAAGCGGATA CGAATATTGC CATGATTCAT GCGGCAGATA AACGTGTTCA TAGCATTCGA 2820
GAAGCTTATC TGCCTGAGCT GTCTGTGATT CCGGGTGTCA ATGCGGCTAT TTTTGAAGAA 2880
TTAGAAGGGC GTATTTTCAC TGCATTCTCC CTATATGATG CGAGAAATGT CATTAAAAAT 2940
GGTGATTTTA ATAATGGCTT ATCCTGCTGG AACGTGAAAG GGCATGTAGA TGTAGAAGAA 3000
CAAAACAACC AACGTTCCGGT CCTTGTTGTT CCGGAATGGG AAGCAGAAGT GTCACAAGAA 3060
GTTCTGTCT GTCCGGGTCG TGGCTATATC CTTCGTGTCA CAGCGTACAA GGAGGGATAT 3120

GGAGAAGGTT GCGTAACCAT TCATGAGATC GAGAACAATA CAGACGAACT GAAGTTTAGC 3180
 AACTGCGTAG AAGAGGAAAT CTATCCAAAT AACACGGTAA CGTGTAATGA TTATACTGTA 3240
 AATCAAGAAG AATACGGAGG TCGGTACACT TCTCGTAATC GAGGATATAA CGAAGCTCCT 3300
 5 TCCGTACCAG CTGATTATGC GTCAGTCTAT GAAGAAAAAT CGTATACAGA TGGACGAAGA 3360
 GAGAATCCTT GTGAATTTAA CAGAGGGTAT AGGGATTACA CGCCACTACC AGTTGGTTAT 3420
 GTGACAAAAG AATTAGAATA CTTCCCAGAA ACCGATAAGG TATGGATTGA GATTGGAGAA 3480
 ACGGAAGGAA CATTATATCGT GGACAGCGTG GAATTACTCC TTATGGAGGA ATAG 3534

10 (2) SEQ ID NO: 34 信息:

(i) 序列特征:

(A) 长度: 1177 个氨基酸

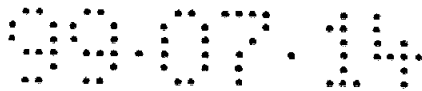
(B) 类型: 核酸

(C) 链型:

15 (D) 拓扑学: 线性

(xi) 序列描述: SEQ ID NO: 34:

Met Asp Asn Asn Pro Asn Ile Asn Glu Cys Ile Pro Tyr Asn Cys Leu
 1 5 10 15
 Ser Asn Pro Glu Val Glu Val Leu Gly Gly Glu Arg Ile Glu Thr Gly
 20 25 30
 Tyr Thr Pro Ile Asp Ile Ser Leu Ser Leu Thr Gln Phe Leu Leu Ser
 35 40 45
 Glu Phe Val Pro Gly Ala Gly Phe Val Leu Gly Leu Val Asp Ile Ile
 50 55 60
 Trp Gly Ile Phe Gly Pro Ser Gln Trp Asp Ala Phe Leu Val Gln Ile
 65 70 75 80
 Glu Gln Leu Ile Asn Gln Arg Ile Glu Glu Phe Ala Arg Asn Gln Ala
 85 90 95
 Ile Ser Arg Leu Glu Gly Leu Ser Asn Leu Tyr Gln Ile Tyr Ala Glu
 100 105 110
 Ser Phe Arg Glu Trp Glu Ala Asp Pro Thr Asn Pro Ala Leu Arg Glu
 115 120 125



Glu Met Arg Ile Gln Phe Asn Asp Met Asn Ser Ala Leu Thr Thr Ala
130 135 140

Ile Pro Leu Phe Ala Val Gln Asn Tyr Gln Val Pro Leu Leu Ser Val
145 150 155 160

Tyr Val Gln Ala Ala Asn Leu His Leu Ser Val Leu Arg Asp Val Ser
165 170 175

Val Phe Gly Gln Arg Trp Gly Phe Asp Ala Ala Thr Ile Asn Ser Arg
180 185 190

Tyr Asn Asp Leu Thr Arg Leu Ile Gly Asn Tyr Thr Asp Tyr Ala Val
195 200 205

Arg Trp Tyr Asn Thr Gly Leu Glu Arg Val Trp Gly Pro Asp Ser Arg
210 215 220

Asp Trp Val Arg Tyr Asn Gln Phe Arg Arg Glu Leu Thr Leu Thr Val
225 230 235 240

Leu Asp Ile Val Ala Leu Phe Pro Asn Tyr Asp Ser Arg Arg Tyr Pro
245 250 255

Ile Arg Thr Val Ser Gln Leu Thr Arg Glu Ile Tyr Thr Asn Pro Val
260 265 270

Leu Glu Asn Phe Asp Gly Ser Phe Arg Gly Ser Ala Gln Gly Ile Glu
275 280 285

Arg Ser Ile Arg Ser Pro His Leu Met Asp Ile Leu Asn Ser Ile Thr
290 295 300

Ile Tyr Thr Asp Ala His Arg Gly Tyr Tyr Tyr Trp Ser Gly His Gln
305 310 315 320

Ile Met Ala Ser Pro Val Gly Phe Ser Gly Pro Glu Phe Thr Phe Pro
325 330 335

Leu Tyr Gly Thr Met Gly Asn Ala Ala Pro Gln Gln Arg Ile Val Ala
340 345 350

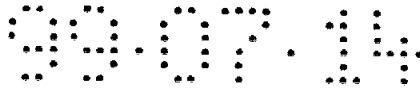
Gln Leu Gly Gln Gly Val Tyr Arg Thr Leu Ser Ser Thr Leu Tyr Arg
355 360 365

Arg Pro Phe Asn Ile Gly Ile Asn Asn Gln Gln Leu Ser Val Leu Asp
370 375 380

Gly Thr Glu Phe Ala Tyr Gly Thr Ser Ser Asn Leu Pro Ser Ala Val
385 390 395 400

Tyr Arg Lys Ser Gly Thr Val Asp Ser Leu Asp Glu Ile Pro Pro Gln
405 410 415

Asn Asn Asn Val Pro Pro Arg Gln Gly Phe Ser His Arg Leu Ser His
420 425 430



Val Ser Met Phe Arg Ser Gly Phe Ser Asn Ser Ser Val Ser Ile Ile
435 440 445

Arg Ala Pro Met Phe Ser Trp Ile His Arg Ser Ala Glu Phe Asn Asn
450 455 460

Ile Ile Ala Ser Asp Ser Ile Thr Gln Ile Pro Leu Val Lys Ala His
465 470 475 480

Thr Leu Gln Ser Gly Thr Thr Val Val Arg Gly Pro Gly Phe Thr Gly
485 490 495

Gly Asp Ile Leu Arg Arg Thr Ser Gly Gly Pro Phe Ala Tyr Thr Ile
500 505 510

Val Asn Ile Asn Gly Gln Leu Pro Gln Arg Tyr Arg Ala Arg Ile Arg
515 520 525

Tyr Ala Ser Thr Thr Asn Leu Arg Ile Tyr Val Thr Val Ala Gly Glu
530 535 540

Arg Ile Phe Ala Gly Gln Phe Asn Lys Thr Met Asp Thr Gly Asp Pro
545 550 555 560

Leu Thr Phe Gln Ser Phe Ser Tyr Ala Thr Ile Asn Thr Ala Phe Thr
565 570 575

Phe Pro Met Ser Gln Ser Ser Phe Thr Val Gly Ala Asp Thr Phe Ser
580 585 590

Ser Gly Asn Glu Val Tyr Ile Asp Arg Phe Glu Leu Ile Pro Val Thr
595 600 605

Ala Thr Leu Glu Ala Glu Tyr Asn Leu Glu Arg Ala Gln Lys Ala Val
610 615 620

Asn Ala Leu Phe Thr Ser Thr Asn Gln Leu Gly Leu Lys Thr Asn Val
625 630 635 640

Thr Asp Tyr His Ile Asp Gln Val Ser Asn Leu Val Thr Tyr Leu Ser
645 650 655

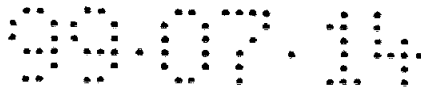
Asp Glu Phe Cys Leu Asp Glu Lys Arg Glu Leu Ser Glu Lys Val Lys
660 665 670

His Ala Lys Arg Leu Ser Asp Glu Arg Asn Leu Leu Gln Asp Ser Asn
675 680 685

Phe Lys Asp Ile Asn Arg Gln Pro Glu Arg Gly Trp Gly Gly Ser Thr
690 695 700

Gly Ile Thr Ile Gln Gly Gly Asp Asp Val Phe Lys Glu Asn Tyr Val
705 710 715 720

Thr Leu Ser Gly Thr Phe Asp Glu Cys Tyr Pro Thr Tyr Leu Tyr Gln
725 730 735



Lys Ile Asp Glu Ser Lys Leu Lys Ala Phe Thr Arg Tyr Gln Leu Arg
740 745 750

Gly Tyr Ile Glu Asp Ser Gln Asp Leu Glu Ile Tyr Leu Ile Arg Tyr
755 760 765

Asn Ala Lys His Glu Thr Val Asn Val Pro Gly Thr Gly Ser Leu Trp
770 775 780

Pro Leu Ser Ala Gln Ser Pro Ile Gly Lys Cys Gly Glu Pro Asn Arg
785 790 795 800

Cys Ala Pro His Leu Glu Trp Asn Pro Asp Leu Asp Cys Ser Cys Arg
805 810 815

Asp Gly Glu Lys Cys Ala His His Ser His His Phe Ser Leu Asp Ile
820 825 830

Asp Val Gly Cys Thr Asp Leu Asn Glu Asp Leu Gly Val Trp Val Ile
835 840 845

Phe Lys Ile Lys Thr Gln Asp Gly His Ala Arg Leu Gly Asn Leu Glu
850 855 860

Phe Leu Glu Glu Lys Pro Leu Val Gly Glu Ala Leu Ala Arg Val Lys
865 870 875 880

Arg Ala Glu Lys Lys Trp Arg Asp Lys Arg Glu Lys Leu Glu Trp Glu
885 890 895

Thr Asn Ile Val Tyr Lys Glu Ala Lys Glu Ser Val Asp Ala Leu Phe
900 905 910

Val Asn Ser Gln Tyr Asp Gln Leu Gln Ala Asp Thr Asn Ile Ala Met
915 920 925

Ile His Ala Ala Asp Lys Arg Val His Ser Ile Arg Glu Ala Tyr Leu
930 935 940

Pro Glu Leu Ser Val Ile Pro Gly Val Asn Ala Ala Ile Phe Glu Glu
945 950 955 960

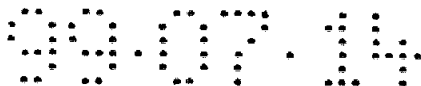
Leu Glu Gly Arg Ile Phe Thr Ala Phe Ser Leu Tyr Asp Ala Arg Asn
965 970 975

Val Ile Lys Asn Gly Asp Phe Asn Asn Gly Leu Ser Cys Trp Asn Val
980 985 990

Lys Gly His Val Asp Val Glu Glu Gln Asn Asn Gln Arg Ser Val Leu
995 1000 1005

Val Val Pro Glu Trp Glu Ala Glu Val Ser Gln Glu Val Arg Val Cys
1010 1015 1020

Pro Gly Arg Gly Tyr Ile Leu Arg Val Thr Ala Tyr Lys Glu Gly Tyr
1025 1030 1035 1040



Gly Glu Gly Cys Val Thr Ile His Glu Ile Glu Asn Asn Thr Asp Glu
 1045 1050 1055

Leu Lys Phe Ser Asn Cys Val Glu Glu Glu Ile Tyr Pro Asn Asn Thr
 1060 1065 1070

Val Thr Cys Asn Asp Tyr Thr Val Asn Gln Glu Glu Tyr Gly Gly Ala
 1075 1080 1085

Tyr Thr Ser Arg Asn Arg Gly Tyr Asn Glu Ala Pro Ser Val Pro Ala
 1090 1095 1100

Asp Tyr Ala Ser Val Tyr Glu Glu Lys Ser Tyr Thr Asp Gly Arg Arg
 1105 1110 1115 1120

Glu Asn Pro Cys Glu Phe Asn Arg Gly Tyr Arg Asp Tyr Thr Pro Leu
 1125 1130 1135

Pro Val Gly Tyr Val Thr Lys Glu Leu Glu Tyr Phe Pro Glu Thr Asp
 1140 1145 1150

Lys Val Trp Ile Glu Ile Gly Glu Thr Glu Gly Thr Phe Ile Val Asp
 1155 1160 1165

Ser Val Glu Leu Leu Leu Met Glu Glu
 1170 1175

5

10

15 (2) SEQ ID NO: 35 信息:

(i) 序列特征:

(A) 长度: 20 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑学: 线性

20

(xi) 序列描述: SEQ ID NO: 35:

TGCAACACTC GAGGCTGAAT

20

说明书附图

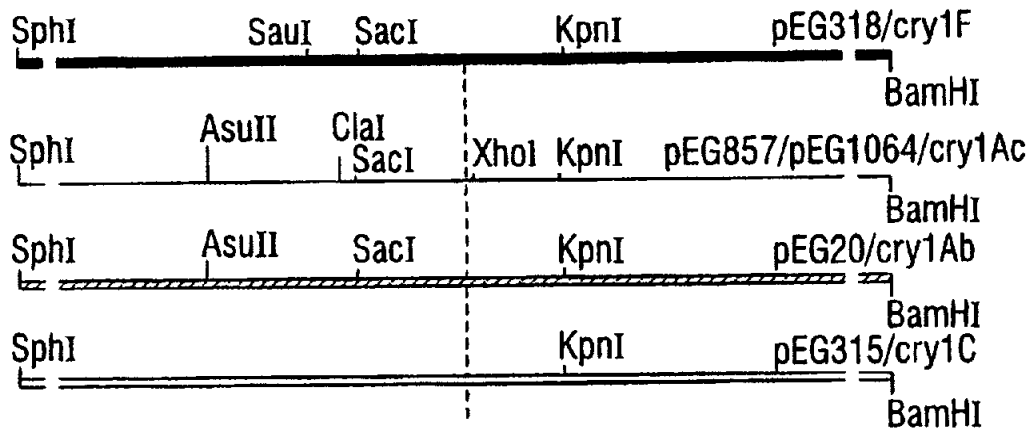


图 1A

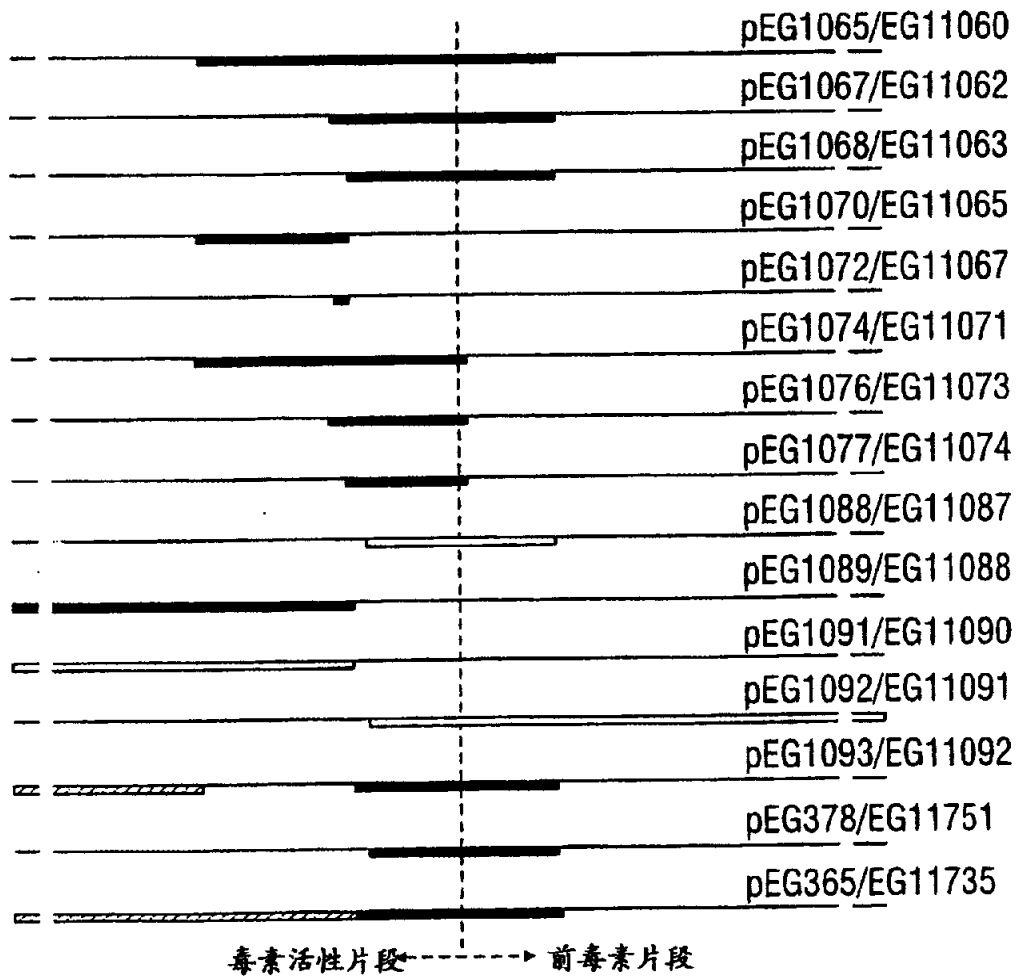


图 1B

99.07.14

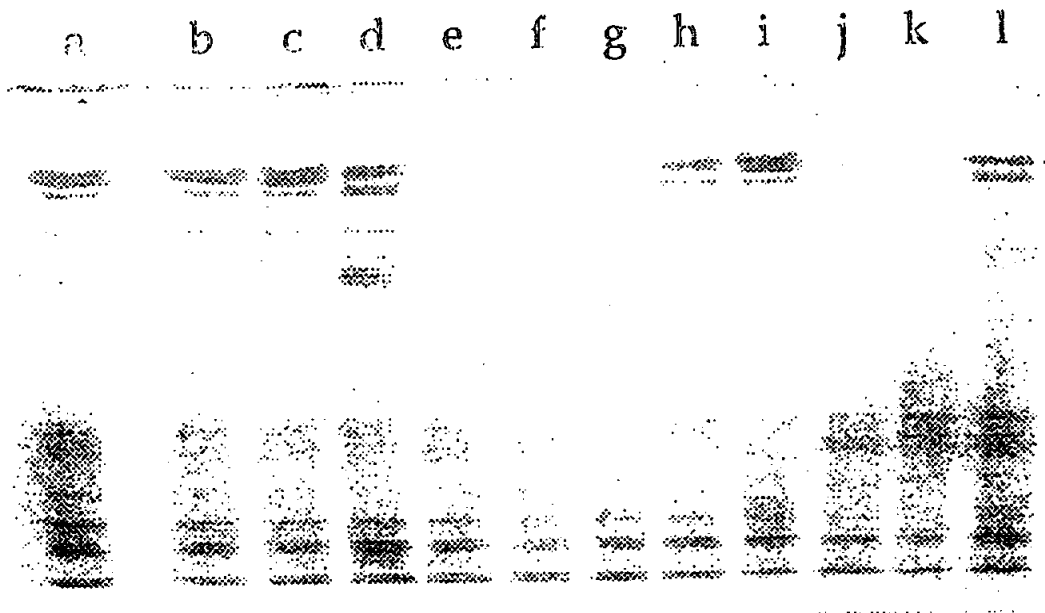


图 2

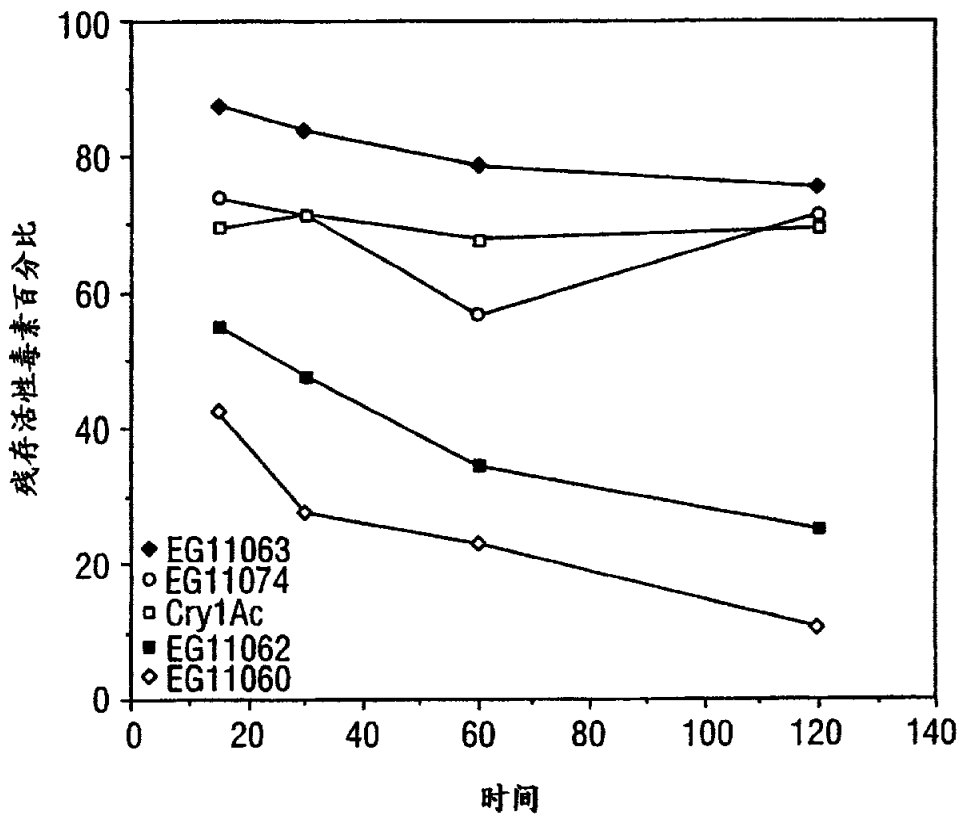


图 3

99.07.14

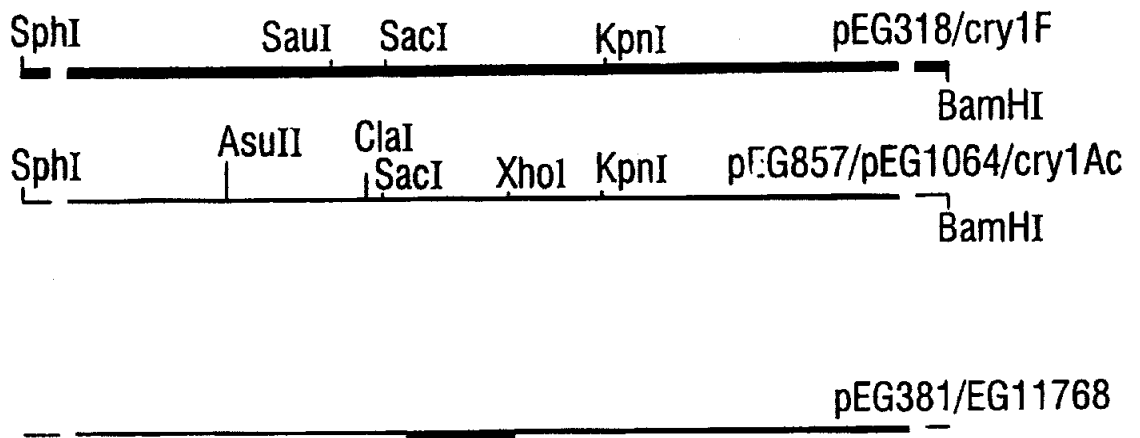


图 4