

[19] 中华人民共和国国家知识产权局



[12] 发明专利说明书

专利号 ZL 01814694.5

[51] Int. Cl.

A01N 63/00 (2006.01)

C12N 15/62 (2006.01)

C12N 15/32 (2006.01)

C12N 15/82 (2006.01)

C12N 5/10 (2006.01)

C07K 14/325 (2006.01)

[45] 授权公告日 2007 年 12 月 12 日

[11] 授权公告号 CN 100353846C

[51] Int. Cl. (续)

C07K 19/00 (2006.01)

[22] 申请日 2001.8.23 [21] 申请号 01814694.5

[30] 优先权

[32] 2000.8.25 [33] US [31] 60/227,956

[86] 国际申请 PCT/EP2001/009751 2001.8.23

[87] 国际公布 WO2002/015701 英 2002.2.28

[85] 进入国家阶段日期 2003.2.25

[73] 专利权人 辛根塔参与股份公司

地址 瑞士巴塞尔

[72] 发明人 N·B·卡洛兹 S·M·拉比

P·J·米勒斯 G·W·瓦伦

P·T·德哈恩

[56] 参考文献

WO9506730A 1995.3.9

Domain III substitution in *Bacillus thuringiensis* delta – endotoxin CryIA (b) results in superior toxicity for *Spodoptera exigua* and altered membrane protein recognition. DE MAAGD RUUD A ET AL. applied and environmental microbiology, Vol. 62 No. 5. 1996

审查员 邹 凯

[74] 专利代理机构 中国国际贸易促进委员会专利
商标事务所

代理人 唐伟杰

权利要求书 2 页 说明书 124 页

[54] 发明名称

新的来自苏云金芽孢杆菌杀虫晶体蛋白的杀
虫毒素

[57] 摘要

为在植物中表达而优化的合成的核苷酸序列编码多种形式的杂合苏云金芽孢杆菌 (*Bacillus thuringiensis*) δ – 内毒素 HO4, 其毒素部分包括 Cry1Ab 的结构域 I 和 II 和 Cry1C 的结构域 III。含所述杀虫毒素的组合物和制剂能控制虫害。本发明进一步涉及制备所述杂合毒素的方法, 以及使用上述核苷酸序列在微生物中控制虫害以及在转基因植物中赋予所述植物昆虫抗性的方法。

1. 合成的核苷酸序列，其编码杂合苏云金芽胞杆菌 (*Bacillus thuringiensis*) 毒素，所述的毒素包含：

来自 Cry1Ab 毒素的结构域 I 和 II，该结构域 I 和 II 按氨基到羧基方向与来自 Cry1C 毒素的结构域 III 连接；和

来自 Cry1Ab 毒素的 C-端尾区域，

其中所述合成的核苷酸序列已经经过修饰而符合植物的特定 GC 含量偏嗜性。

2. 根据权利要求 1 的合成的核苷酸序列，其中，所述的杂合毒素包含 SEQ ID NO: 6, 8, 或 10。

3. 根据权利要求 1 的合成的核苷酸序列，其包含 SEQ ID NO: 5, 7, 或 9。

4. 分离的核酸分子，其包含 SEQ ID NO: 5, 7, 9, 12, 13, 14, 15, 16, 或 17。

5. 权利要求 4 的核酸分子，其包含 SEQ ID NO: 5, 7, 或 9。

6. 权利要求 4 的核酸分子，其包含 SEQ ID NO: 12, 13, 14, 15, 16, 或 17。

7. 嵌合基因，其包含与权利要求 1 至 3 任一项的合成的核苷酸序列可操作地连接的异源启动子序列。

8. 包含权利要求 7 的嵌合基因的重组载体。

9. 包含权利要求 7 的嵌合基因的转基因植物细胞。

10. 控制下述昆虫害虫的方法，所述昆虫害虫选自草地夜蛾 (*fall armyworm*)，红铃麦蛾 (*pink bollworm*)，烟夜蛾 (*tobacco budworm*)，欧洲玉米螟 (*European cornborer*)，和菜蛾 (*diamondback moth*)，该方法包括将该昆虫害虫与包含权利要求 9 的植物细胞的转基因植物接触。

11. 根据权利要求 10 的方法，其中所述转基因植物是玉米，棉花，稻或卷心菜。

12. 产生抗昆虫植物的方法，该方法包括向所述植物中引入根据权利要求 1 至 3 任一项的合成核苷酸序列，其中所述的合成核苷酸序列可在所述植物中以可控制昆虫害虫的有效量表达。

13. 权利要求 12 的方法，其中所述昆虫害虫是鳞翅目昆虫害虫。

14. 权利要求 13 的方法，其中昆虫害虫是草地夜蛾 (*fall armyworm*)，红铃麦蛾 (*pink bollworm*)，烟夜蛾 (*tobacco budworm*)，欧洲玉米螟 (*European cornborer*)，或菜蛾 (*diamondback moth*)。

15. 权利要求 12 的方法，其中所述植物是玉米、棉花、稻或者卷心菜。

新的来自苏云金芽孢杆菌杀虫 晶体蛋白的杀虫毒素

技术领域

本发明涉及新的来自苏云金芽孢杆菌杀虫晶体蛋白的杀虫毒素，表达所述蛋白的核酸序列，以及制备和使用所述毒素和相应核酸序列以控制昆虫的方法。

背景技术

虫害是造成作物减产的主要原因。仅在美国由多种属昆虫侵染造成的损失就达数十亿美国。除大田作物的损失之外，虫害还是蔬菜和水果培育者、观赏花卉生产者的负担，同时它们也是花匠和房主们的祸害。

虫害主要是通过集中施用化学杀虫剂进行控制，这些化学杀虫剂通过抑制昆虫生长，妨碍昆虫进食或繁殖，或者致使昆虫死亡来发挥活性。这样可以很好地控制昆虫，但这些化学物质有时候也影响其它有益的昆虫。广泛应用化学杀虫剂所导致的另一个问题是抗性昆虫变种的出现。虽然这种情况可以通过多种抗性管理策略进行部分地缓解，但对替代性的虫害防治剂的需求仍日益增强。

生物学昆虫防治剂，如表达杀虫毒素的苏云金芽孢杆菌 (*Bacillus thuringiensis*) 菌株的应用已经取得了令人满意的效果，成为化学杀虫剂的一种替代或补充。苏云金芽孢杆菌 (*Bacillus thuringiensis*) 属于革兰氏阳性、需氧、形成内孢子的细菌大组。与其它非常相近的芽孢杆菌种，例如蜡状芽孢杆菌 (*B. cereus*) 或炭疽芽孢杆菌 (*B. anthracis*) 不同，目前已知的苏云金芽孢杆菌 (*Bacillus thuringiensis*) 种大多数在其芽胞形成过程中形成一种伴胞包涵体 (parasporal inclusion body)，鉴于其晶体结构，通常将其称为晶体 (crystalline body)。所述的晶体包含具有杀虫

活性的晶体原毒素蛋白，既所谓的 δ -内毒素。这些蛋白晶体负责苏云金芽胞杆菌 (*Bacillus thuringiensis*) 对昆虫的毒性。所谓的 δ -内毒素直到所述晶体 (crystalline body) 经口摄入并溶解于靶昆虫的肠液后才表现出杀虫活性。大多数情况下，实际的毒性组分是由来自昆虫消化道的蛋白酶对原毒素进行蛋白水解切割而释放的。所述的多种苏云金芽胞杆菌 (*Bacillus thuringiensis*) 菌株的 δ -内毒素具有针对特定的靶昆虫，特别是对于多种鳞翅目，鞘翅目和双翅目幼虫的高度特异性的特征，并且对这些幼虫具有高度的活性。应用苏云金芽胞杆菌 (*Bacillus thuringiensis*) δ -内毒素进一步的优势在于，所述的毒素对人，其它哺乳动物，鸟类和鱼是无害的。

根据序列同源性和杀虫特异性，可将苏云金芽胞杆菌 (*Bacillus thuringiensis*) 晶体蛋白分成不同的类。研究得最多的是 Cry1 类蛋白，该蛋白是作为 140 kDa 的原毒素产生的，对鳞翅目昆虫具有毒性。在一定程度上已阐明了晶体蛋白的作用模式。经口摄入后所述的晶体溶解于幼虫中肠的碱性环境。随后中肠蛋白酶(如胰蛋白酶)将溶解的蛋白加工成约 65kDa 的抗蛋白酶毒性片段，所述的毒性片段与昆虫中肠上皮细胞受体相结合并穿透细胞膜。这最终导致细胞破裂和幼虫死亡。

易感昆虫中肠上皮细胞上存在的受体在很大程度上决定了特定晶体蛋白的活性谱。上述的谱也是由所述晶体蛋白的溶解效率和体内蛋白水解活化作用共同决定的。所述晶体蛋白与中肠上皮受体结合的重要性通过下述事实得到进一步的证明，即当昆虫对一种晶体蛋白产生了抗性时，晶体蛋白与抗性昆虫中肠上皮细胞的结合明显下降。

在过去的几年中，已经分离出了一些编码上述晶体蛋白的基因，其在异源宿主中的表达为控制在经济学上重要的虫害提供了另一种工具。特别地，在转基因植物中杀虫毒素，如苏云金芽胞杆菌 (*Bacillus thuringiensis*) 晶体蛋白的表达对选定的虫害提供了有效的防护，表达这些毒素的转基因植物已商业化，这使农民减少了对化学昆虫防治剂的应用。而且，也可以表达具有经选择的意在增强

针对特定昆虫或昆虫种类杀虫活性，或意在扩大毒素蛋白杀虫谱的组合功能的重组毒素。例如可通过将 δ -内毒素的毒素部分与其它 δ -内毒素原毒素(尾部)部分结合构建具有自然界不存在的新序列的嵌合杀虫蛋白。参见，例如引入此处作为参考的 WO 98/15170。

据认为晶体蛋白毒性片段包含三个不同的结构域。结构域 I，最 N-末端的结构域，由 7 个 α -螺旋组成，可能部分或全部插入到靶细胞膜中。结构域 II 在所谓 Greek key-构象中包含 3 个 β -折叠。大多数的研究者认为结构域 II 与受体相互作用，因此决定毒素的特异性。的确，存在许多的证据暗示结构域 II 残基涉及特异的毒性和高度亲和结合。结构域 III，最 C-末端结构域，由在所谓 jellyroll 构象中的两个 β -折叠组成，也暗示着其与确定特异性相关。例如通过编码区之间的体内重组在毒素之间交换结构域 III 能够导致特异活性的改变。利用这种杂合体进行的结合实验表明结构域 III 涉及与推定的靶昆虫受体的结合，表明结构域 III 可能通过受体识别功能行使其在特异性方面的功能。如果投射在 Cry1 序列上，结构域 I 从约第 28 位到第 260 位残基，结构域 II 由约第 260 位到第 460 位，结构域 III 由约第 460 位到第 600 位。参见 Nakamura 等，*Agric. Biol. Chem.* 54(3): 715-724 (1990); Li 等，*Nature* 353: 815-821 (1991); Ge 等，*J. Biol. Chem.* 266(27): 17954-17958 (1991); 和 Honee 等，*Mol. Microbiol.* 5(11): 2799-2806 (1991); 上述各篇均引入此处作为参考。在此处引入作为参考的 U.S. Pat. No. 5,736,131 描述了在其 C-末端包含第一 Cry 蛋白结构域 III、在其 N-末端包含第二 Cry 蛋白结构域 I 和 II 的苏云金芽胞杆菌 (*Bacillus thuringiensis*) 杂合体毒素片段。这种杂合晶体蛋白具有改变了的杀虫活性。例如，同样在 De Maagd 等，*Appl. Environ. Microbiol.* 62(5): 1537-1543 (1996) 中描述的 H04 杂合毒素在其 N-末端包含 Cry1Ab 的结构域 I 和 II，在其 C-末端包含 Cry1C 的结构域 III。据报道与亲本的 Cry1Ab 毒素相比，H04 对甜菜粘虫 (*Spodoptera exigua*) 具有高毒性，并且比亲本的 Cry1C 也有明显地更高毒性。也

参见 Bosch 等, *FEMS Microbiology Letters* 118: 129-134 (1994); Bosch 等, *Bio/Technology* 12: 915-918 (1994); De Maagd 等, *Appl. Environ. Microbiol.* 62(8): 2753-2757 (1996); 和 De Maagd 等, *Mol. Microbiol.* 31(2): 463-471 (1999); 以上各篇均引入此处作为参考。

尽管通过育种程序和遗传工程掺入昆虫抗性基因获得了成功, 仍然存在长期渴望并仍未实现的对发现新的有效昆虫防治剂的需求。特别需要能靶向经济上重要的虫害, 如欧洲玉米螟 (*European Corn Borer*) 和草地夜蛾 (*Fall Army Worm*), 并能有效控制对现有的昆虫防治剂具有抗性的昆虫种类的控制剂。而且, 需要施用后对环境的影响降到最低的试剂。

发明内容

本发明通过提供编码杂合苏云金芽胞杆菌 (*Bacillus thuringiensis*) 毒素的新的基因序列以满足上述需求, 所述新的基因序列包括为在植物中表达而优化的合成的核苷酸序列。在优选的实施方案中, 所述的新基因序列编码杂合苏云金芽胞杆菌 (*Bacillus thuringiensis*) δ -内毒素 H04 的不同形式, 其毒素部分包含 Cry1Ab 的结构域 I 和 II 和 Cry1C 的结构域 III。由所述的新基因序列编码的杂合苏云金芽胞杆菌毒素对下述经济上重要的虫害具有高度的活性, 如草地夜蛾 (*fall armyworm*), 红铃麦蛾 (*pink bollworm*), 烟夜蛾 (*tobacco budworm*), 欧洲玉米螟 (*European cornborer*) 和菜蛾 (*diamondback moth*)。所述的杂合苏云金芽胞杆菌 (*Bacillus thuringiensis*) 毒素可用于多种昆虫防治策略中, 以获得最高的效率和最低的环境影响。

本发明进一步涉及由本发明的核苷酸序列表达产生的杂合杀虫毒素, 及包含所述杂合杀虫毒素的组合物和制剂, 它们能够抑制害虫存活, 生长或繁殖的能力, 或者限制害虫相关的农作物损害和亏损。本发明进一步涉及制造所述杂合毒素的方法, 以及使用所述核苷酸序列的方法, 例如, 在转基因植物中赋予昆虫抗性, 和使用所述毒素、包

含所述毒素的组合物及制剂的方法，例如，在受昆虫侵染的区域应用所述的毒素、组合物或制剂，或预防性地处理易受昆虫侵染的区域或植物，以赋予其对有害昆虫的防护或抗性。所述的杂合毒素可用于复合昆虫防治策略，以获得最高的效率和最低的环境影响。

一方面本发明提供控制下述昆虫的方法，所述昆虫选自草地夜蛾（*fall armyworm*），红铃麦蛾（*pink bollworm*），烟夜蛾（*tobacco budworm*），欧洲玉米螟（*European cornborer*）和菜蛾（*diamondback moth*），该方法包括向所述昆虫递送有效量的杂合苏云金芽胞杆菌（*Bacillus thuringiensis*）毒素，该毒素包含来自 Cry1Ab 毒素的结构域 I 和 II 以及按由氨基到羧基方向与它们结合的来自 Cry1C 毒素的结构域 III。在一个优选的实施方案中，所述的杂合苏云金芽胞杆菌（*Bacillus thuringiensis*）毒素包含与 SEQ ID NO:2, 4, 6, 8, 或 10 至少 90% 同一的氨基酸序列。在一个更优选的实施方案中，所述的杂合苏云金芽胞杆菌（*Bacillus thuringiensis*）毒素包含 SEQ ID NO:2, 4, 6, 8, 或 10。

在本发明上述方法的又一实施方案中，所述的杂合苏云金芽胞杆菌（*Bacillus thuringiensis*）毒素进一步包含 C-末端尾区域，如 Cry1C 尾区域或 Cry1Ab 尾区域。所述的 C-末端尾区域可以是全长的也可以是截短的，如约 40 个氨基酸长。

在本发明上述方法的一个优选的实施方案中，向昆虫递送有效量的杂合苏云金芽胞杆菌（*Bacillus thuringiensis*）毒素包括用含编码所述杂合苏云金芽胞杆菌（*Bacillus thuringiensis*）毒素的核苷酸序列的重组 DNA 转化的转基因植物组织饲喂或接触昆虫，其中，所述的杂合苏云金芽胞杆菌（*Bacillus thuringiensis*）毒素在所述转基因植物组织中的表达赋予了对所述昆虫的抗性。优选地，所述核苷酸序列基本上与 SEQ ID NO:1, 3, 5, 7, 或 9 同一。

另一方面，本发明提供了包含编码杂合苏云金芽胞杆菌（*Bacillus thuringiensis*）毒素的核苷酸序列的分离的核酸分子，所述毒素包括(a)含来自 Cry1Ab 毒素的结构域 I 和 II 以及按由

氨基到羧基方向与之结合的来自 Cry1C 毒素的结构域 III 的 N-端毒素部分;和(b)来自 Cry1Ab 毒素的 C-端尾区域。优选地,所述的杂合苏云金芽胞杆菌 (*Bacillus thuringiensis*) 毒素包含与 SEQ ID NO:6, 8, 或 10 至少 90% 同一的氨基酸序列。更优选地,所述的杂合苏云金芽胞杆菌 (*Bacillus thuringiensis*) 毒素包含 SEQ ID NO:6, 8, 或 10。更优选地,所述核苷酸序列与 SEQ ID NO:5, 7, 或 9 至少 90% 同一。最优选地,所述核苷酸序列包含 SEQ ID NO: 5, 7, 或 9。

本发明进一步提供包含与异源启动子序列可操作地连接的本发明如上所述核酸分子的嵌合基因;包含所述嵌合基因的重组载体;包含所述嵌合基因的转基因宿主细胞(如,植物细胞);含上述转基因植物细胞的转基因植物(如,玉米,棉花,稻,或甘兰植物);以及上述转基因植物的种子。

又一方面本发明提供使植物抗昆虫的方法,该方法包括在用下述嵌合基因转化的植物中表达杂合苏云金芽胞杆菌 (*Bacillus thuringiensis*) 毒素,所述嵌合基因包含:(a)能在植物中启动相关的编码序列高水平转录的核酸启动子序列,和(b)与所述启动子可操作地连接的本发明的核酸分子,其中,所述的杂合苏云金芽胞杆菌 (*Bacillus thuringiensis*) 毒素在所述植物中的表达能使所述植物抗昆虫。

再一方面,本发明提供生产具有抗昆虫活性的杂合苏云金芽胞杆菌 (*Bacillus thuringiensis*) 毒素的方法,该方法包括(a)获得本发明的转基因宿主细胞;和(b)在所述转基因宿主细胞中表达本发明的核酸分子,从而产生有效地抗昆虫的杂合苏云金芽胞杆菌 (*Bacillus thuringiensis*) 毒素。

再一方面,本发明提供生产抗昆虫植物的方法,该方法包括将本发明的核酸分子引入所述植物,其中所述的核酸分子能在所述植物中以有效地控制昆虫的量表达。

又一方面，本发明提供包含 SEQ ID NO:3, 5, 7, 9, 11, 12, 13, 14, 15, 16 或 17 的分离的核酸分子；含与上述核酸分子可操作地相连的异源启动子的嵌合基因；含上述嵌合基因的重组载体；含上述嵌合基因的转基因宿主细胞(如, 植物细胞)；含上述转基因植物细胞的转基因植物(如, 玉米, 棉花, 稻, 或甘兰植物)；和上述转基因植物的种子。

在又一方面中，本发明提供杂合苏云金芽胞杆菌 (*Bacillus thuringiensis*) 毒素，该毒素包含：(a)含来自 Cry1Ab 毒素的结构域 I 和 II 以及与它们按由氨基端到羧基端方向结合的来自 Cry1C 毒素的结构域 III 的 N-末端毒素部分；和(b)来自 Cry1Ab 毒素的 C-末端尾区域。优选地，所述的杂合苏云金芽胞杆菌 (*Bacillus thuringiensis*) 毒素包含与 SEQ ID NO:6, 8, 或 10 至少 90% 同一的氨基酸序列。更优选地，所述的杂合苏云金芽胞杆菌 (*Bacillus thuringiensis*) 毒素包含 SEQ ID NO: 6, 8, 或 10。

另一方面，本发明提供含控制昆虫有效量的本发明的杂合苏云金芽胞杆菌 (*Bacillus thuringiensis*) 毒素的组合物。

通过本发明的下述描述和非限制性的实施例本领域的技术人员能够清楚地了解本发明的其它方面和有益之处。

序列表中序列的简要描述

SEQ ID NO:1 是 De Maagd 等, *Appl. Environ. Microbiol.* 62(5): 1537-1543 (1996)中描述的编码 H04 杂合毒素的核苷酸序列，其毒素部分包括位于其 N-末端的 Cry1Ab 结构域 I 和 II 和位于其 C-末端的 Cry1C 的结构域 III，加上全长的 Cry1C 尾部分。

SEQ ID NO:2 是 SEQ ID NO:1 中所述的核苷酸序列编码的 H04 杂合毒素的氨基酸序列，其包括 Cry1Ab 毒素结构域 I 和 II 和 Cry1C 的毒素结构域 III，加上全长的 Cry1C 尾部分。

SEQ ID NO:3 是合成的编码不带尾的 H04 毒素部分的核苷酸序列，如同胰蛋白酶位点已被切割。

SEQ ID NO:4 是由 SEQ ID NO:3 描述的合成核苷酸序列编码的 H04 毒素部分的氨基酸序列。

SEQ ID NO:5 是编码 H04 毒素部分加全长 Cry1Ab 尾部分的合成的核苷酸序列。

SEQ ID NO:6 是由 SEQ ID NO:5 中描述的合成核苷酸序列编码的 H04 + Cry1Ab 尾的氨基酸序列。

SEQ ID NO:7 是编码 H04 毒素部分加全长 Cry1Ab 尾部分的另一合成的核苷酸序列。

SEQ ID NO:8 是由 SEQ ID NO:7 中描述的合成核苷酸序列编码的 H04 + Cry1Ab 尾的氨基酸序列。

SEQ ID NO:9 是编码 H04 毒素部分加 Cry1Ab 尾头 40 个氨基酸的合成的核苷酸序列。

SEQ ID NO:10 是由 SEQ ID NO:9 中描述的合成核苷酸序列编码的 H04 + 40 氨基酸的截短的 Cry1Ab 尾的氨基酸序列。

SEQ ID NO:11 是构建体 pNOV1308 的核苷酸序列，其包含与如 SEQ ID NO:3 中所述的编码无尾的 H04 毒素部分的合成核苷酸序列可操作地连接的组成型玉米遍在蛋白启动子。

SEQ ID NO:12 是构建体 pNOV1436 的核苷酸序列，其包含与如 SEQ ID NO:5 中所述的编码 H04 毒素部分加全长 Cry1Ab 尾部分的合成核苷酸序列可操作地连接的根偏好的玉米 MTL 启动子。

SEQ ID NO:13 是构建体 pNOV1441 的核苷酸序列，其包含与如 SEQ ID NO:5 中所述的编码 H04 毒素部分加全长 Cry1Ab 尾部分的合成核苷酸序列可操作地连接的组成型玉米遍在蛋白启动子。

SEQ ID NO:14 是构建体 pNOV1305 的核苷酸序列，其包含与如 SEQ ID NO:7 中所述的编码 H04 毒素部分加全长 Cry1Ab 尾部分的合成核苷酸序列可操作连接的组成型玉米遍在蛋白启动子。

SEQ ID NO:15 是构建体 pNOV1313 的核苷酸序列，其包含与如 SEQ ID NO:7 中所述的编码 H04 毒素部分加全长 Cry1Ab 尾部分的合成核苷酸序列可操作连接的组成型玉米遍在蛋白启动子。

SEQ ID NO:16 是构建体 pNOV1435 的核苷酸序列，其包含与如 SEQ ID NO:9 中所述的编码 H04 毒素部分加 Cry1Ab 尾头 40 个氨基酸的合成核苷酸序列可操作地连接的根偏好的玉米 MTL 启动子。

SEQ ID NO:17 是构建体 pZU578 的核苷酸序列，其包含与如 SEQ ID NO:9 中所述的编码 H04 毒素部分加 Cry1Ab 尾头 40 个氨基酸的合成核苷酸序列可操作地连接的鼠耳芥属 (*Arabidopsis*) 肌动蛋白-2 启动子。

定义

本发明的毒素的“活性”是指毒素作为经口有效的昆虫防治剂而起作用，具有毒性效力，或能破坏或阻止昆虫进食，其可能或不能引起昆虫死亡。当本发明的毒素递送到昆虫，典型的结果是昆虫死亡，或者昆虫不以使所述昆虫获得该毒素的来源为食。

“与.....相关联/可操作地连接”指两核酸序列在物理上或功能上相关联。例如，若两序列可操作地相连，或处于一定的位置，使得调节 DNA 序列影响编码或结构 DNA 序列的表达水平，则称启动子或调节性的 DNA 与编码 RNA 或蛋白的 DNA 序列“相关联”。

“结合位点”是指分子上的某个位点，其中所述位点与毒素间的结合是可逆的，以使得位点与毒素间的 Ka 接近于至少 $10^4 \text{ dm}^3 \text{ mole}^{-1}$ 。

“嵌合基因”是重组的核酸序列，其中启动子或调节核酸序列与编码 mRNA 或表达成蛋白的核酸序列可操作地连接或相关联，从而使调节核酸序列能调节转录或相关核酸分子的转录或表达。嵌合基因中的调节核酸序列通常并不与相关核酸序列如同自然界中所发现的那样可操作地连接或相关联。

“编码序列”是转录成 RNA，如 mRNA，rRNA，tRNA，snRNA，有意 RNA 或反义 RNA 的核酸序列。优选地，所述 RNA 之后在生物体中翻译产生蛋白。

互补的：“互补的”是指两核苷酸序列包含反向平行的核苷酸序列，它们能够通过反向平行的核苷酸序列中的互补碱基残基之间形

成氢键彼此配对。

特定核酸序列的“保守修饰的变异”是指编码同一或基本上同一的氨基酸序列的那些核酸序列,或当所述的核酸序列不编码氨基酸序列时,是指基本上同一的序列。由于遗传密码的简并性,大量功能上同一的核酸编码任一给定的多肽。例如密码子 CGT, CGC, CGA, CGG, AGA, 和 AGG 全部编码精氨酸。因此在密码子编码精氨酸的每个位置,所述的密码子可改变成上述的任一密码子而不改变所编码的蛋白。这种核酸变异是“沉默变异”,其是一种“保守修饰的变异”。除非另有说明,此处所述的编码蛋白的每个核酸序列也描述了每一种可能的沉默变异。本领域的技术人员可知核酸序列内的任一密码子(除通常是甲硫氨酸的唯一密码子的 ATG 外)均可通过常规的技术进行修饰以获得功能同一的分子。因此,编码蛋白的核酸分子的每种“沉默变异”均暗含于所述的序列中。

而且,本领域的技术人员可知当改变导致用化学上类似的氨基酸取代某氨基酸时,在编码序列中改变,添加或缺失单一氨基酸或小百分比的氨基酸(典型地是少于 5%,更典型地是少于 1%)的单独的替代,缺失或添加是“保守修饰的变异”。提供功能类似氨基酸的保守替代是本领域已知的。下述的五组中,每组氨基酸彼此之间均可进行保守替代:脂肪族的:甘氨酸(G),丙氨酸(A),缬氨酸(V),亮氨酸(L),异亮氨酸(I);芳香族的:苯丙氨酸(F),酪氨酸(Y),色氨酸(W);含硫的:甲硫氨酸(M),半胱氨酸(C);碱性的:精氨酸(R),赖氨酸(K),组氨酸(H);酸性的:天冬氨酸(D),谷氨酸(E),天冬酰胺(N),谷氨酰胺(Q)。参见 Creighton (1984) *Proteins*, W.H. Freeman and Company。另外,在编码序列中改变,添加或缺失单一氨基酸或小百分数的氨基酸的单独的替代,缺失或添加也是“保守修饰的变异”。

对昆虫的“控制”通过毒性作用抑制害虫存活、生长、进食和/或繁殖,或限制昆虫相关的作物损害或损失。“控制”昆虫可指也可以不指杀死昆虫,尽管其优选地指杀死昆虫。

相应于:在本发明的上下文中,“相应于”或“相当于”指当不同的

苏云金芽胞杆菌 (*Bacillus thuringiensis*) δ -内毒素的核酸编码序列或氨基酸序列相互对比 (aligned) 时, “相当于” 特定的列举位置的核酸或氨基酸是与这些位置相对比, 但不一定处于与特定 δ -内毒素的各自核酸编码序列或氨基酸序列相应的这些确切数值位置的那些。类似地, 当特定的 δ -内毒素 (例如, Cry1B) 的编码或氨基酸序列与参考的 δ -内毒素 (例如, Cry1Ab) 的编码或氨基酸序列相对比时, 相当于 Cry1Ab 序列中特定的某列举位置的 Cry1B 序列中的核酸或氨基酸是与 Cry1Ab 序列中的这些位置相对比, 但不一定处于 Cry1B 毒素各核酸编码序列或氨基酸序列的这些确切数值位置中的那些。

“递送”毒素是指将所述毒素与昆虫相接触, 导致毒性效果以及对昆虫的控制。所述的毒素可通过多种已知的方法递送, 如通过昆虫摄食经口递送, 或通过转基因植物的表达与昆虫相接触, 配制成蛋白组合物, 喷雾形式的蛋白组合物, 毒饵基质, 或任何本领域已知的毒素递送系统。

此处所用的“表达盒”是指能指导特定的核苷酸序列在适当的宿主中表达的核酸序列, 其包括与有用核苷酸序列可操作连接的启动子, 所述的核苷酸序列与终止信号可操作地相连接。典型地, 它还包括使所述核苷酸序列正确翻译所需的序列。包含有用核苷酸序列的所述表达盒可以是嵌合的, 这就是说, 对于至少一种其它组分而言, 至少一种其中的组分是异源的。所述的表达盒也可以是一种在自然界中存在的, 但为有利于异源表达以重组形式获得的。但典型地, 所述的表达盒相对宿主而言是异源的, 即所述表达盒的特定核酸序列在天然状态下并不存在于宿主细胞中, 必须通过转化事件引入到宿主细胞或其前体中。在表达盒中的核苷酸序列的表达可以在组成型或诱导型启动子的控制之下, 其中诱导型启动子仅当宿主细胞暴露于特定的外部刺激之下时才启动转录。对于多细胞生物, 如植物, 所述的启动子也可以是对特定的组织、器官或发育阶段特异的。

基因: 术语“基因”广泛地用于指任何与生物功能相关的 DNA 片段。因此, 基因包括编码序列和/或其表达所需的调控序列。基因也

包括非表达的 DNA 节段, 例如, 它形成其它蛋白的识别序列。基因可以由多种来源获得, 包括从有用的来源中克隆或由已知的或预测的序列信息合成, 也可以包括被设计为具有所需参数的序列。

“有用基因”是指当其转移到植物中后能赋予植物所需性状的任何基因, 上述所需性状如抗生素抗性, 病毒抗性, 昆虫抗性, 疾病抗性或对其它有害之物的抗性, 除草剂耐受性, 改善了的营养价值, 在工业加工过程中改良了的性能, 或改变了的繁殖能力。所述的“有用基因”也可以是在其转移到植物中后产生有商业价值的酶或代谢物的基因。

此处所用的“H04”是指在 De Maagd 等, *Appl. Environ. Microbiol.* 62(5): 1537-1543 (1996)中所描述的杂合 *Bt* 毒素, 其毒素片段在 N-末端包含 Cry1Ab 的结构域 I 和 II, 在 C-末端包含 Cry1C 的结构域 III。

异源核酸序列: 此处所用的术语“异源核酸[或 DNA]序列”, “外源核酸[或 DNA]节段”或“异源基因”均指来源于相对特定宿主细胞而言是外源的来源, 或者如果是相同来源即是由原始形式经修饰的序列。因此在宿主细胞中的异源基因包括通过例如密码子优化而修饰的特定宿主细胞中的内源基因。该术语也包括那些天然存在序列的非天然多拷贝形式。因此该术语指外来或异源于所述细胞的核酸节段, 或虽与宿主细胞同源, 但处在宿主细胞核酸内该元件通常不存在的某个位置上的核酸节段。外源核酸节段表达产生外源多肽。

“同源”核酸[或 DNA]序列是指与其被引入的宿主细胞天然相关的核酸[或 DNA]序列。

“同源重组”是同源核酸分子间核酸片段的相互交换。

“同质体 (Homoplastidic)”指植物, 植物组织或植物细胞, 其中所有的质体是遗传学上同一的。这是所述质体未经转化, 突变或其它形式的遗传改变前在植物中的正常状态。在不同的组织, 或发育阶段, 所述的质体可以是不同的形式, 如, 叶绿体, 前质体, 黄化质体, 造粉体, 有色体, 等。

在上下文中，术语两个或更多核酸或蛋白序列的“同一”或百分“同一性”是指应用下述的序列比较算法之一或通过视觉检验测量，当比较或对比最大值相应性时，两个或更多的序列或子序列相同，或具有指明百分数的相同氨基酸残基或核苷酸百分数。

“杀虫的”定义为能够控制昆虫的毒性生物活性，优选地是通过杀死昆虫的形式。

当核酸序列编码具有与参考核酸序列编码的多肽相同的氨基酸序列的多肽时所述核酸序列是“同义编码(isocoding with)”参考核酸序列。

“分离的”核酸分子或分离的酶是指通过人工操作使其脱离天然天然环境而存在的核酸分子或酶，因此，其不是天然产物。一种分离的核酸分子或酶可以以纯化的形式存在，或存在于非天然的环境如重组的宿主细胞中。

在杂合毒素中毒素结构域间，即本发明的杂合杀虫毒素结构域 II 和 III 间的“结合点”，是在所述杂合体中的同源交换区域或位点。在所述交换位点左侧的氨基酸来源于一个亲本毒素，而在所述交换位点右侧的氨基酸来源于另一个亲本毒素。

成熟蛋白：正常情况下靶定某种细胞器的蛋白，其中转运肽已被去除。

最小启动子(minimal promoter)：启动子元件，特别是 TATA 元件，其在缺乏上游激活的情况下是非活性的，或具有大大降低的启动子活性。在合适的转录因子存在下，所述最小启动子起作用允许转录。

天然的：指存在于非转化细胞基因组中的基因。

天然地存在：术语“天然地存在”用于描述可在自然界中发现的对象，以区别于人工生产的对象。例如在生物体(包括病毒)中存在的蛋白或核苷酸序列，其可从自然来源中分离，未经过实验室中的有意的人工修饰，是天然存在的。

核酸：术语“核酸”指脱氧核糖核苷酸或核糖核苷酸及其单链或双

链形式的多聚体。除非特别限定，该术语包括与参考核酸具有相似结合特性，含有天然核苷酸的已知类似物，并以与天然存在的核苷酸类似的方式代谢的核酸。除非另有说明，特定的核酸也暗指包括其经保守修饰的变异体（如简并的密码子取代）和互补序列以及明确说明的序列。具体地，简并密码子取代可以通过产生如下序列实现，在所述序列中，一个或多个选定的（或全部）密码子第三位被混合碱基或脱氧肌苷残基所取代（Batzer 等, *nucleic acid Res.* 19: 5081 (1991); Ohtsuka 等, *J. Biol. Chem.* 260: 2605-2608 (1985); Rossolini 等, *Mol. Cell. Probes* 8: 91-98 (1994)）。术语“核酸”或“核酸序列”可与基因，由基因编码的 cDNA, 和 mRNA 相互交换地使用。

“ORF”指开放阅读框。

蛋白的“部分”是指包含在所述蛋白内的肽，其具有所述蛋白至少 80% 的连续序列。

“植物”是在任意发育阶段的植物，特别是种子植物。

“植物细胞”是植物的结构和生理单位，其包括原生质体和细胞壁。所述的植物细胞可以是分离的单细胞或培养细胞形式，或作为高度组织化单位，如植物组织，植物器官，或整个植物的一部分。

“植物细胞培养物”是指下述植物单位的培养物，所述植物单位如，原生质体，细胞培养细胞（cell culture cells），植物组织中的细胞，花粉，花粉管，胚珠，胚囊，受精卵和不同发育阶段的胚。

“植物材料”指叶，茎，根，花或花的组成部分，果实，花粉，卵细胞，受精卵，种子，插条，细胞或组织培养物，或其它部分或植物产物。

“植物器官”是植物明显可见结构化的分化部分，如根，茎，叶，花蕾或胚。

此处所用的“植物组织”是指组织成结构和功能单位的植物细胞群。培养的或在植物中的任何植物组织均包含在内。这一术语包括但不限于，整个植株，植物器官，植物种子或组织培养物和任何组织成结构和功能单位的植物细胞群。单独使用该术语或将该术语与上述列

举的或者包含在这一定义内的任意特定类型的植物组织结合使用，并不意味着排除任何其它类型的植物组织。

“启动子”是位于编码序列上游的非翻译的 DNA 序列，其包含 RNA 聚合酶 II 的结合位点，并起始 DNA 的转录。所述的启动子区域还可以包括作为基因表达调控序列的其它元件。

“原生质体”是没有细胞壁，或只带有细胞壁的一些部分的经分离的植物细胞。

纯化的：术语“纯化的”当应用于核酸或蛋白时，是指所述的核酸或蛋白基本上不带有天然状态下与所述核酸或蛋白相关的其它的细胞组分。优选的是以同质的状态存在，但其也可以是干燥的或水溶液的形式。纯度和同质性典型地是应用分析化学技术，如聚丙烯酰胺凝胶电泳或高效液相色谱来确定。在制剂中占优势种类的蛋白是基本上纯的。术语“纯化的”是指核酸或蛋白在电泳凝胶中基本上给出一条带的情况。具体说来，它是指至少约 50%纯的核酸或蛋白，更优选至少约 85%纯，最优选至少约 99%纯。

当来自两核酸每一个的序列在子代核酸中相联合时，这两核酸“重组”。当两核酸均作为重组底物时，两序列“直接”重组。当两核酸需中间物，如交换寡核苷酸进行重组时，两核酸是“间接重组”。对于间接重组，不多于一个序列作为实际的重组底物，有些情况下两序列均不是用于重组的底物。

“调控元件”是指涉及控制核苷酸序列表达的序列。调控元件包含与有用核苷酸序列可操作连接的启动子和终止信号。典型地，其还包括使所述核苷酸序列正确翻译所需的序列。

基本上同一：在两核酸或蛋白序列上下文中词组“基本上同一”是指应用下述的序列比较算法之一或通过视觉检验测量，当比较或对比最大值相应性时，两个或更多的序列或子序列具有至少 60%，优选 80%，更优选 90%，更优选 95%，最优选至少 99%的核苷酸或氨基酸序列同一性。优选地，在至少约 50 个残基长度，更优选至少约 100 个残基的序列区域中存在基本上同一，最优选该序列在至少约 150 个残基上基

本上同一。在最优选的实施方案中,所述的序列在编码区全长内基本上同一。而且,基本上同一的蛋白或核酸序列执行基本上相同的功能。

对于序列比较,典型地将一个序列作为参考序列,将其它序列与该序列进行比较。当使用序列比较算法时,将待测序列和参考序列输入计算机,在需要时指定子序列坐标,并指定序列算法程序的参数。然后,基于指定的序列参数,所述的序列比较算法计算待测序列相对于参考序列的百分序列同一性。

用于比较的优化序列对比(alignment)可根据下述算法进行, Smith & Waterman, *Adv. Appl. Math.* 2: 482 (1981)的局部同源性算法, Needleman & Wunsch, *J. Mol. Biol.* 48: 443 (1970)的同源对比算法, Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85: 2444 (1988)的相似性搜索方法,这些算法的计算机化执行(GAP, BESTFIT, FASTA, 和 TFASTA, 见 the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI),或通过目测(一般见 Ausubel 等,下文)进行。

一个适合于确定序列百分同一性和序列相似性的算法是 BLAST 算法,其如 Altschul 等, *J. Mol. Biol.* 215: 403-410 (1990)所述。执行 BLAST 分析的软件公众可以从国家生物技术信息中心获得(<http://www.ncbi.nlm.nih.gov/>)。这一算法包括首先通过鉴定在待查询序列中长度为 W 的短字段鉴定高得分序列对(HSPs),当所述的字段与数据库序列中相同长度的字段相对比时该字段匹配或符合某些正值的阈值得分 T。T 是指邻近字段的得分阈值(Altschul 等, 1990)。这些起始邻近字段的命中作为起始搜索以发现包含它们的更长 HSP 的“种晶”。只要累积对比得分能够增长,即可在每个序列中沿两个方向扩展命中的字段。累积得分可通过下述方式计算,对于核苷酸序列,用参数 M (一对匹配残基的回报得分;总是 > 0)和 N (不匹配残基的罚分;总是 < 0)。对于氨基酸序列,可用评分矩阵计算累积得分。当累积对比得分由其矩阵最高获得值衰减量 X, 一个或更多的

负值残基对比的累积使累积分值达到 0 或 0 以下, 或者达到了任一序列的末端, 则在每一方向上命中字段的延伸停止。BLAST 算法参数 W, T, 和 X 决定对比的灵敏度和速度。BLASTN 程序(对于核苷酸序列)字段长度(W)缺省值为 11, 期望值(E)为 10, 截断(cutoff)为 100, M=5, N=-4, 及两条链的比较。对于氨基酸序列, BLASTP 使用的缺省值字段长度(W)为 3, 期望值(E)为 10, BLOSUM62 评分矩阵(参见 Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89: 10915 (1989))。

除了计算序列百分同一性外, BLAST 算法还可进行两序列之间相似性的统计分析(参见, 如, Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* 90: 5873-5787 (1993))。BLAST 算法提供的一种相似性测定是最小总和概率(P(N)), 它提供随机出现两核苷酸或氨基酸序列之间的匹配的可能性的指示。例如, 测试核苷酸序列被认为与参照序列相似, 条件是在测试核苷酸与参照核酸序列比较中的最小总和概率小于约 0.1, 更优选小于约 0.01, 最优选小于约 0.001。

两核酸序列基本上同一的另一种指示是所述的两核酸分子在严谨条件下彼此杂交。词组“特异性杂交”是指一个分子在严谨条件下仅与存在于复杂的(如, 细胞的总)DNA 或 RNA 混合物中的特定核苷酸序列结合, 双螺旋化或杂交。“基本上结合”是指核酸探针与靶核酸之间的互补杂交, 并且包含少量的错配, 这种错配可通过降低杂交介质的严谨度来包容, 以达到所需的靶核酸序列检测。

核酸杂交试验, 如 Southern 和 Northern 杂交中的“严谨杂交条件”和“严谨杂交洗涤条件”是序列依赖性的, 而且在不同的环境参数下也是不同的。较长的序列在较高的温度下特异性杂交。Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with nucleic Acid Probes* part I chapter 2 “Overview of principles of hybridization and the strategy of nucleic acid probe assays” Elsevier, New York 中有对核酸杂交的详细指导。一般地, 高严谨性杂交和洗涤条件选择

为在低于确定的离子强度和 pH 下特定序列的热熔点 (T_m) 约 5°C 的条件。典型地, 在“严谨条件”下探针可与靶序列杂交, 但不与其它序列杂交。

T_m 是指如下温度(在确定的离子强度和 pH 下), 在该温度下 50% 的靶序列与完全匹配的探针杂交。非常严谨的条件是选择为等同于特定探针的 T_m 。一个用于具有 100 个以上互补残基的互补核酸, 在 Southern 或 northern 印迹滤膜上杂交的严谨杂交条件的示例为含 1 mg 肝素的 50% 甲酰胺, 于 42°C 杂交过夜。高度严谨洗涤条件的示例为: 0.1 5M NaCl 于 72°C 下处理约 15 分钟。一个严谨的洗涤条件的例子是于 65°C 下, 0.2x SSC 中洗 15 分钟(参见, Sambrook, 下文, SSC 缓冲液的描述)。通常低严谨的洗涤在高严谨洗涤前应用以去除背景探针信号。对于例如多于 100 个核苷酸的双螺旋的中度严谨洗涤条件的示例为 1x SSC, 45°C 下 15 分钟。对于例如多于 100 个核苷酸的双螺旋的低度严谨洗涤条件的示例为 4-6x SSC, 40°C 下 15 分钟。对于短探针, (如, 约 10 到 50 个核苷酸), 严谨条件典型地包括, 盐浓度低于约 1.0M Na 离子, 典型的是约 0.01 到 1.0 M Na 离子浓度(或其它的盐), pH 7.0 到 8.3, 温度典型的是至少为 30°C。严谨条件也可通过添加去稳定试剂, 如甲酰胺来达到。一般, 在特定杂交试验中, 信噪比比使用非相关探针所达到的结果高 2x (或更高) 表示特异性杂交检测。在严谨条件下不能彼此杂交的核酸在它们所编码的蛋白基本上同一时仍是基本上同一的。这出现在例如, 一个核酸拷贝是利用遗传密码所允许的最大密码子简并性制造的时。

下述的杂交/洗涤条件示例可用于克隆与本发明的参考核苷酸序列基本上同一的同源核苷酸序列: 一种参考核苷酸序列优选地与所述参考核苷酸序列在下述条件下杂交: 于 50°C 下 7% 十二烷基硫酸钠 (SDS), 0.5 M NaPO_4 , 1 mM EDTA 中杂交, 在 50°C 下用 2X SSC, 0.1% SDS 洗涤, 更希望的是于 50°C 下 7% 十二烷基硫酸钠 (SDS), 0.5 M NaPO_4 , 1 mM EDTA 中杂交, 于 50°C 下用 1X SSC, 0.1% SDS 洗涤, 更希望的是于 50°C 下 7% 十二烷基硫酸钠 (SDS), 0.5 M NaPO_4 ,

1 mM EDTA 中杂交, 于 50°C 下在 0.5X SSC, 0.1% SDS 中洗涤, 优选地于 50°C 下 7% 十二烷基硫酸钠 (SDS), 0.5 M NaPO₄, 1 mM EDTA 中杂交, 于 50°C 下在 0.1X SSC, 0.1% SDS 中洗涤, 更优选的是于 50°C 下 7% 十二烷基硫酸钠 (SDS), 0.5 M NaPO₄, 1 mM EDTA 中杂交, 于 65°C 下在 0.1X SSC, 0.1% SDS 中洗涤。

两种核酸序列或蛋白基本上同一的另一指示为由第一核酸编码的蛋白与由第二核酸编码的蛋白产生免疫交叉反应或特异性地相结合。因此, 例如当两蛋白的区别仅在于保守取代时, 第一蛋白与第二蛋白典型地基本上同一。

词组“特异性地(或选择性地)与某种抗体结合”或“特异性(或选择性地)与.....发生免疫反应”, 当指蛋白或肽时, 是指结合反应, 该反应可以确定在异质蛋白群或其它生物制剂中上述蛋白的存在。因此, 在指定的免疫测定条件下, 特定的抗体与特定的蛋白相结合而不以显著的量与存在于样品中的其它蛋白相结合。在这样的条件下与某种抗体的特异性结合需要选择针对特定蛋白具有特异性的抗体。例如, 选择用带有由本发明的任何核酸编码的氨基酸序列的蛋白培育的抗体, 可以获得与上述蛋白发生特异的免疫反应, 且除多态性变异性之外不与其它蛋白发生免疫反应的抗体。多种免疫测定形式均可用于筛选可与特定蛋白进行特异地免疫反应的抗体。例如, 固相 ELISA 免疫测定, Western 印迹, 或免疫组织化学都是常规用于筛选可与某种蛋白发生特异免疫反应的单克隆抗体的方法。用于确定特异性免疫反应的免疫测定形式和条件的描述参见 Harlow 和 Lane (1988) *Antibodies, A Laboratory Manual*, Cold Spring Harbor Publications, New York (“Harlow 和 Lane”)。典型地, 特异性或选择性反应应至少两倍于背景信号或噪音, 更典型地 10 倍以上-100 倍于背景。

“子序列”指核酸或氨基酸序列, 其分别包含较长的核酸或氨基酸(如蛋白)序列的一部分。

“合成的”指包含并不存在于天然序列中的结构特征的核苷酸序

列。例如，非常接近地模仿单子叶和/或双子叶植物（dicot 和/或 monocot）基因 G+C 含量和正常的密码子分布的人工序列即被视为合成的序列。

“转化”是一种将异源核酸引入宿主细胞或生物体的过程。具体说来，“转化”是指将 DNA 分子稳定地整合到目的生物的基因组中。应当理解经转化的细胞，组织，或昆虫不仅包含转化过程的终产物还包括其转基因的子代。

“转化的/转基因的/重组的”是指宿主生物，如细菌或植物，在其中已引入了异源核酸分子。所述的核酸分子可以稳定地整合到宿主生物的基因组中，或者所述核酸分子也可以以染色体外分子的形式存在。这种染色体外分子可以进行自主复制。经转化的细胞，组织，或植物应当理解为不仅包含转化过程的终产物还包含其转基因子代。“非转化的”，“非转基因的”，或“非重组的”宿主指不含异源核酸分子的野生型的生物，如，细菌或植物。

核苷酸通过下述标准缩写表示的碱基表示：腺嘌呤(A)，胞嘧啶(C)，胸腺嘧啶(T)，和鸟嘌呤(G)。氨基酸可由下述的标准缩写表示：丙氨酸(Ala; A)，精氨酸(Arg; R)，天冬酰胺(Asn; N)，天冬氨酸(Asp; D)，半胱氨酸(Cys; C)，谷氨酰胺(Gln; Q)，谷氨酸(Glu; E)，甘氨酸(Gly; G)，组氨酸(His; H)，异亮氨酸(Ile; I)，亮氨酸(Leu; L)，赖氨酸(Lys; K)，甲硫氨酸(Met; M)，苯丙氨酸(Phe; F)，脯氨酸(Pro; P)，丝氨酸(Ser; S)，苏氨酸(Thr; T)，色氨酸(Trp; W)，酪氨酸(Tyr; Y)，和缬氨酸(Val; V)。而且用 (Xaa; X)代表任意氨基酸。

本发明涉及表达为新的毒素的新核酸序列，以及制造和使用所述毒素以控制虫害。具体而言，本发明涉及经优化以便在植物中表达的合成的基因序列，该序列编码多种形式的所述的杂合苏云金芽胞杆菌（*Bacillus thuringiensis*） δ -内毒素 H04，其毒素部分包含 Cry1Ab 的结构域 I 和 II 和 Cry1C 的结构域 III。编码 H04 杂合毒素的杂合基因，如 U.S. Pat. No. 5,736,131 和 De Maagd 等, *Appl.*

Environ. Microbiol. 62(5): 1537-1543 (1996) 所述, 由天然的 cry1Ab 和 Cry1C 基因构建。WO 93/07278 中给出了构建本发明的合成的 H04 基因的优选方法。由所述的新基因序列编码的杂合苏云金芽胞杆菌 (*Bacillus thuringiensis*) 毒素对经济上重要的虫害, 如草地夜蛾 (fall armyworm), 红铃麦蛾 (pink bollworm), 烟夜蛾, 欧洲玉米螟 (European cornborer), 和菜蛾 (diamondback moth) 等具有高度的活性。所述的杂合苏云金芽胞杆菌 (*Bacillus thuringiensis*) 毒素可用于复合昆虫防治策略中, 以获得最高的效率和最低的环境影响。

本发明还包括含有编码本发明的杀虫毒素的核苷酸序列的 DNA 分子。本发明进一步包括含有本发明核酸序列的重组载体。在这样的载体中, 所述的核酸序列优选地包含在含有调控元件的表达盒中, 所述的调控元件使得所述核苷酸序列能在能表达所述核苷酸序列的宿主细胞中表达。这种调控元件通常包含启动子和终止信号, 优选地, 其还包括使本发明的核酸序列有效翻译所需的元件。含有所述核酸序列的载体通常能在特定的宿主细胞中复制, 优选地作为染色体外分子而复制, 由此用于在所述宿主细胞中扩增本发明的核酸序列。在一个实施方案中, 所述载体的宿主细胞是微生物, 如细菌, 特别是苏云金芽胞杆菌 (*Bacillus thuringiensis*) 或大肠杆菌。在另一实施方案中所述重组载体的宿主细胞是内生植物或附生植物。对于所述载体, 一种优选的宿主细胞是真核细胞, 如植物细胞。植物细胞如玉米细胞是最优选的宿主细胞。

在特别优选的实施方案中, 本发明的杀虫毒素在植物中表达。在这种情况下, 表达有效量的毒素的转基因植物可保护自身不受虫害的侵袭。当昆虫开始摄食这种转基因植物后, 它们也摄入了植物表达的毒素。这将制止所述昆虫进一步咬食植物组织甚至危害或杀死所述的昆虫。

本申请中所述的核酸序列可通过重组 DNA 技术整合到植物细胞中。一般来讲, 这包括利用本领域已知的克隆程序, 将本发明的编码

序列插入到一种表达系统中，对所述的表达系统来说上述序列是异源的(即，通常情况下不存在的)。所述的载体包含所插入的蛋白编码序列转录和翻译所必需的元件。本领域已知的大量载体系统均可使用，如质粒，噬菌体病毒和其它经修饰的病毒。合适的载体包括但不限于病毒载体，如 λ 载体系统 λ gt11， λ gt10 和 Charon 4；质粒载体如 pBI121， pBR322， pACYC177， pACYC184， pAR 系列， pKK223-3， pUC8， pUC9， pUC18， pUC19， pLG339， pRK290， pKC37， pKC101， pCDNAII；以及其它类似的系统。转化的细胞可再生完整的植物，以致本发明的核苷酸序列给该转基因植物赋予昆虫抗性。

本发明转化的植物可以是单子叶或双子叶植物 (monocots 或 dicots)，其包括但不限于，玉米，小麦，大麦，黑麦，甜马铃薯，黄豆，豌豆，菊苣，莴苣，卷心菜，嫩茎花椰菜，花椰菜，芜菁，萝卜，菠菜，芦笋，洋葱，大蒜，胡椒，芹菜，南瓜属植物，南瓜，大麻，夏南瓜，苹果，梨，榲桲，瓜(如，西瓜)，李子，樱桃，桃子，油桃，杏，草莓，葡萄，悬钩子，黑莓，凤梨，鳄梨，番木瓜，芒果，香蕉，大豆，番茄，高粱，甘蔗，甜菜，向日葵，菜籽油菜，三叶草，烟草，胡萝卜，棉花，苜蓿，稻，马铃薯，茄子，黄瓜，拟南芥属 (Arabidopsis)，和木本植物，如松柏和落叶树。一旦所需的核苷酸序列转化到特定的植物种中，它可以利用传统的育种技术在该植物种中繁殖或转移到同种植物的其它品种中，特别是包括商业化的品种。

为使本发明的核苷酸序列在转基因植物中表达，它们可能需要被修饰或优化。尽管在很多情况下，来自微生物的基因在不经修饰的情况下也能在植物中高水平表达，但有时由于微生物的核苷酸序列存在植物所不偏嗜的密码子，使得它们在转基因植物中低水平表达。本领域中已知所有的生物对于其所使用的密码子均有特异的偏嗜性，本发明所述的核苷酸序列的密码子可以改变成符合植物偏嗜性并保持其所编码的氨基酸不变。而且，在植物中的高表达由具有至少约 35%，优选高于约 45%，更优选高于约 50%，最优选高于约 60%GC 含量的编码

序列达到。具有低 GC 含量的微生物的核苷酸序列在植物中的表达量低是由于存在 ATTTA 基元，该基元使信息去稳定，并且 AATAAA 基元可能引起不适当的聚腺苷酸化。尽管优选的基因序列可以在单子叶和双子叶植物种中均能足量地表达，也可以对序列进行修饰以满足单子叶植物或双子叶植物各自的特异密码子偏嗜性和 GC 含量偏嗜性，因为已证明两类植物的偏嗜性是不同的(Murray 等 *Nucl. Acids Res.* 17: 477-498 (1989))。另外，还通过筛选所述的核苷酸序列以排除可能造成信息截短的不适当的剪接位点的存在。所有需要在所述核苷酸序列内部进行的改变如上述改变均可利用本领域已知的定点突变，PCR 和合成基因构建等技术，利用已公开的专利申请 EP 0 385 962, EP 0 359 472, 和 WO 93/07278 所描述的方法来进行。

为有效地起始翻译，需要对起始甲硫氨酸临近的序列进行修饰。例如，它们可以通过包括已知的在植物中有效的序列进行修饰。Joshi 建议了一种用于植物的适当的共有序列(NAR 15: 6643-6653 (1987))，Clontech 建议了另外的共有序列翻译起始区(1993/1994 catalog, page 210)。这些共有序列均适合用于本发明的核苷酸序列。将上述序列整合到含所述核苷酸序列的构建体中，直至包括 ATG (留下第二个氨基酸是未经修饰的)，或直至包括 ATG 之后的 GTC (具有修饰转基因中第二个氨基酸的可能性)。

核苷酸序列在转基因植物中的表达被可在植物中行使功能的启动子所驱动。对启动子的选择依表达的时空需要和靶物种的不同而变化。因此，本发明的核苷酸序列表达在叶，穗，花序（如穗状花序，圆锥花序，穗轴，等等），根，和/或幼苗中是优选的。但是，在许多情况下，需要防卫多于一种的虫害，这样就希望在多种组织中表达。尽管有许多来自双子叶植物的启动子在单子叶植物中也可以起作用，反之亦然，但理想的仍然是选择双子叶植物启动子用于在双子叶植物中表达，而单子叶植物启动子用于在单子叶植物中表达。但是对所选择的启动子的起源并没有限制；只要它们能够驱动所述核苷酸序列在所需细胞中表达就足够了。

优选的组成型表达的启动子包括来自编码肌动蛋白和遍在蛋白的基因的启动子和 CaMV 35S 及 19S 启动子。本发明的核苷酸序列也可在可被化学调控的启动子的调控下表达。这使得所述的杀虫毒素只有在所述作物经诱导性化学物质处理后才能合成。优选的化学诱导基因表达技术在已公开的专利申请 EP 0 332 104 和 US 专利 5,614,395 中有描述。优选的化学诱导启动子的例子是烟草 PR-1a 启动子。

优选的启动子的种类是创伤诱导型的。已报道了多种在创伤位点及植物病原体侵染位点表达的启动子。理想地,这种启动子应当仅在侵染位点被局部激活,通过这种方式所述的杀虫毒素仅在需要合成该杀虫毒素以杀死入侵害虫的位点积累。优选的这种类型的启动子包括在下述文献中描述的那些启动子,即 Stanford 等, *Mol. Gen. Genet.* 215: 200-208 (1989), Xu 等, *Plant. Molec. Biol.* 22: 573-588 (1993), Logemann 等, *Plant Cell* 1: 151-158 (1989), Rohrmeier & Lehle, *Plant. Molec. Biol.* 22: 783-792 (1993), Firek 等, *Plant. Molec. Biol.* 22: 129-142 (1993), 和 Warner 等, *Plant J.* 3: 191-201 (1993)。

优选的组织特异性表达方式包括绿色组织特异的,根特异的,茎特异的和花特异的。适合于在绿色组织中表达的启动子包括许多调节涉及光合作用的基因的启动子,它们中许多已从单子叶植物和双子叶植物中被克隆出来。一种优选的启动子是来自磷酸烯醇羧化酶基因的玉米 PEPC 启动子(Hudspeth & Grula, *Plant. Molec. Biol.* 12: 579-589 (1989))。一种优选的根特异性表达启动子是 de Framond (*FEBS* 290: 103-106 (1991); EP 0 452 269 中所描述的玉米金属硫蛋白样(Maize metallothionein-like, MTL)启动子。优选的茎特异性启动子是 US 专利 5,625,136 中所描述的启动子,其驱动玉米 *trpA* 基因的表达。

本发明的特别优选的实施方案是以根优选的或根特异的方式表达至少一种本发明的核苷酸序列的转基因植物。进一步优选的实施方案是以创伤诱导或病原体侵染诱导的方式表达所述核苷酸序列的转基因

植物。

除选择合适的启动子外，用于在植物中表达杀虫毒素的构建体中还需要有合适的转录终止子附加在所述异源核苷酸序列的下游。有几种此种类型的启动子是本领域已知并可以获得的（如 *tm1* 来自 CaMV, E9 来自 *rbcS*）。任何可获得的能在植物中行使功能的已知终止子均可用于本发明。

许多序列均可整合到本发明中所述的表达盒中。它们包括已表明能增强表达的序列，如内含子序列（如来自 *Adh1* 和 *bronze1* 的）和病毒前导序列（如来自 TMV, MCMV 和 AMV 的）。

也可以优选使本发明的核苷酸序列表达靶向在植物中的不同细胞。有些情况下希望在细胞胞液质中表达，另外的情况下可能优选在亚细胞器中表达。利用本领域已知的技术可以进行转基因编码的酶的亚细胞定位。典型地，将编码来源于已知的细胞器靶向性基因产物的靶肽的 DNA 加工并融合到所述核苷酸序列的上游。已知有很多针对叶绿体的这种靶序列，并且已经研究出它们在异源构建体中有功能。本发明的核苷酸序列的表达也可以靶向宿主细胞的内质网或液泡。达到上述目的的技术是本领域已知的。

适合于植物转化的载体已在本说明书的其它部分进行了描述。对于土壤杆菌（*Agrobacterium*）介导的转化，适合使用二元载体或带有至少一种 T-DNA 边缘序列的载体，而对于直接基因转移，任何载体都是合适的，仅含有有用构建体的线性 DNA 可以是优选的。对于直接 DNA 转移的情形，可利用单一的 DNA 种类转化或共转化（Schocher 等 *Biotechnology* 4: 1093-1096 (1986)）。对于直接转移和土壤杆菌介导的转移两者，通常（但不是必须）利用选择标记进行转化，所述的选择标记可提供抗生素（卡那霉素，潮霉素或氨基葡萄糖）或除草剂（basta）抗性。这种选择标记的例子是新霉素磷酸转移酶，潮霉素磷酸转移酶，二氢叶酸还原酶，膦丝菌素乙酰转移酶，2,2-二氯丙酸（2,2-dichloropropionic acid）脱卤素酶，乙酰羟酸合酶，5-烯醇丙酮酰-莽草酸-磷酸合酶，haloarylnitrilase，原卟啉

(protoporphyrinogen) 氧化酶, 乙酰-辅酶 A 羧化酶, 二氢蝶酸 (dihydropteroate) 合酶, 氯霉素乙酰转移酶, 和 β -葡萄糖醛酸糖苷酶。另一种类型的阳性选择标记是甘露糖-6-磷酸异构酶(MPI/PMI)基因, 它提供代谢甘露糖-6-磷酸盐异构酶的能力。但用于植物转化的筛选或选择标记不是本发明的关键。

可利用多种本领域已知的技术将上述的重组 DNA 引入植物细胞。本领域的技术人员知道根据转化靶向的植物类型选择这些方法。转化植物细胞的合适方法包括微注射 (Crossway 等, *BioTechniques* 4:320-334 (1986)), 电穿孔 (Riggs 等, *Proc. Natl. Acad. Sci. USA* 83:5602-5606 (1986)), 土壤杆菌介导的转化 (Hinchey 等, *Biotechnology* 6:915-921 (1988); 也见, Ishida 等, *Nature Biotechnology* 14:745-750 (June 1996) 中的玉米转化), 直接基因转移 (Paszkowski 等, *EMBO J.* 3:2717-2722 (1984); Hayashimoto 等, *Plant. Physiol.* 93:857-863 (1990) (稻)), 和利用 Agracetus, Inc., Madison, Wisconsin 和 Dupont, Inc., Wilmington, Delaware 提供的设备进行的冲击粒子加速作用 (参见, 例如, Sanford 等, U.S. Patent 4,945,050; 和 McCabe 等, *Biotechnology* 6:923-926 (1988))。还可参见, Weissinger 等, *Annual Rev. Genet.* 22:421-477 (1988); Sanford 等, *Particulate Science and Technology* 5:27-37 (1987) (洋葱); Svab 等, *Proc. Natl. Acad. Sci. USA* 87: 8526-8530 (1990) (烟草叶绿体); Christou 等, *Plant Physiol.* 87:671-674 (1988) (大豆); McCabe 等, *Bio/Technology* 6:923-926 (1988) (大豆); Klein 等, *Proc. Natl. Acad. Sci. USA*, 85:4305-4309 (1988) (玉米); Klein 等, *Bio/Technology* 6:559-563 (1988) (玉米); Klein 等, *Plant. Physiol.* 91:440-444 (1988) (玉米); Fromm 等, *Bio/Technology* 8:833-839 (1990); 和 Gordon-Kamm 等, *Plant Cell* 2: 603-618 (1990) (玉米); Koziel 等, *Biotechnology* 11: 194-200 (1993) (玉米); Shimamoto 等, *Nature* 338: 274-277

(1989) (稻); Christou 等, *Biotechnology* 9: 957-962 (1991) (稻); Datta 等, *Bio/Technology* 8:736-740 (1990) (稻); 欧洲专利申请 EP 0 332 581 (鸭茅和其它 *Pooideae*); Vasil 等, *Biotechnology* 11: 1553-1558 (1993) (小麦); Weeks 等, *Plant Physiol.* 102: 1077-1084 (1993) (小麦); Wan 等 *Plant Physiol.* 104: 37-48 (1994) (大麦); Jahne 等, *Theor. Appl. Genet.* 89:525-533 (1994) (大麦); Umbeck 等, *Bio/Technology* 5: 263-266 (1987) (棉花); Casas 等, *Proc. Natl. Acad. Sci. USA* 90:11212-11216 (Dec. 1993) (高粱); Somers 等, *Bio/Technology* 10:1589-1594 (Dec. 1992) (燕麦); Torbert 等, *Plant Cell Reports* 14:635-640 (1995) (燕麦); Weeks 等, *Plant Physiol.* 102:1077-1084 (1993) (小麦); Chang 等, WO 94/13822 (小麦) 和 Nehra 等, *The Plant Journal* 5:285-297 (1994) (小麦)。通过微粒轰击将重组 DNA 分子引入玉米的特别优选的一组实施方案参见 Koziel 等, *Biotechnology* 11: 194-200 (1993), Hill 等, *Euphytica* 85:119-123 (1995) 和 Koziel 等, *Annals of the New York Academy of Sciences* 792:164-171 (1996)。此外的优选实施方案是如 EP 0 292 435 中公开的对于玉米的原生质体转化法。植物转化可用单一的 DNA 种类或多 DNA 种类(即共转化)进行, 这两种技术均适用于本发明的编码序列。

在另一优选的实施方案中, 本发明的核苷酸序列直接转化到质体基因组中。质体转化的一个主要的优势在于质体一般可以表达未经实质修饰的细菌基因, 并且质体还能表达在单一启动子控制下的多个开放阅读框。质体转化技术在 U.S. 专利 Nos. 5,451,513, 5,545,817, 和 5,545,818, 在 PCT 申请 WO 95/16783, 和 McBride 等 (1994) *Proc. Natl. Acad. Sci. USA* 91, 7301-7305 中有详细的描述。叶绿体转化的基本技术包括, 利用生物轰击 (biolistics) 或原生质体转化 (如, 氯化钙或 PEG 介导的转化) 将所克隆的质体 DNA 区域连同有用基因一起引入到合适的靶组织中, 所述的质体 DNA 区域两侧为选择标

记。称作靶向序列的 1 到 1.5 kb 的侧翼区易于与质体基因组进行同源重组，从而对质体基因组 (plastome) 的特定区域进行置换或修饰。起初，利用在叶绿体 16S rRNA 和赋予壮观霉素和/或链霉素抗性的 *rps12* 基因处的点突变作为转化的筛选标记 (Svab, Z., Hajdukiewicz, P., 和 Maliga, P. (1990) *Proc. Natl. Acad. Sci. USA* 87, 8526-8530; Staub, J. M., 和 Maliga, P. (1992) *Plant Cell* 4, 39-45)。这导致以对靶叶轰击约每 100 次中有一次的频率得到稳定的同质体转化体。在这些标记之间存在的克隆位点可用于创建靶定质体的载体，该载体可用于引入外源基因 (Staub, J. M., 和 Maliga, P. (1993) *EMBO J.* 12, 601-606)。转化频率的大幅度提高可通过下述方式实现，即用显性的选择标记，编码壮观霉素解毒酶氨基糖苷-3'-腺嘌呤转移酶 (Svab, Z., 和 Maliga, P. (1993) *Proc. Natl. Acad. Sci. USA* 90, 913-917) 的细菌 *aadA* 基因替换隐性的 rRNA 或 r-蛋白抗生素抗性基因。先前这一标记已成功地用于绿藻 *Chlamydomonas reinhardtii* 质体基因组的高频转化 (Goldschmidt-Clermont, M. (1991) *Nucl. Acids Res.* 19: 4083-4089)。其它本领域已知的对质体转化有用的选择标记也包含在本发明的范围内。典型地，在转化后需要约 15-20 个细胞分裂周期才能达到同质体状态。通过同源重组已使基因插入到存在于每一植物细胞的环状质体基因组的全部几千个拷贝中的质体表达，利用了相对于核表达基因的巨大拷贝数的优势，使得表达水平可容易地超过总的植物可溶蛋白的 10%。在一个优选的实施方案中，本发明的核苷酸序列插入到靶定质体的载体中并转化到所需植物宿主的质体基因组中。由此获得了就含有本发明的核苷酸序列的质体基因组而言同型的植物，所述植物优选地可以高表达所述核苷酸序列。

实施例

本发明参照下述的详细实施例进行进一步详细描述。这些实施例只是处于示例性的目的，除非特别说明，它们不是限制性的。此处所用的标准的重组 DNA 技术和分子克隆技术是本领域已知的，并且在下

述文献中有详细的描述: Ausubel(编), *Current Protocols in Molecular Biology*, John Wiley and Sons, Inc. (1994); T. Maniatis, E. F. Fritsch 和 J. Sambrook, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor laboratory, Cold Spring Harbor, NY (1989); 和 T.J. Silhavy, M.L. Berman, 和 L.W. Enquist, *Experiments with Gene Fusions*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1984)。

实施例 1: H04 毒素片段的表达和纯化

将 H04 杂合毒素基因的截短形式克隆到表达载体, 如 pBluescript SK-, 芽孢杆菌穿梭载体, 或 pET 21b(+) 中以便在大肠杆菌中过表达, 所述的 H04 杂合毒素基因的截短形式, 如 De Maagd 等, *Appl. Environ. Microbiol.* 62(5): 1537-1543 (1996) 中所述, 编码本质上由 Cry1Ab 的结构域 I 和 II 以及 Cry1C 的结构域 III 组成的 Bt 毒素。细胞在含 50 $\mu\text{g/ml}$ 氨苄青霉素的 LB 培养基上于 37°C 下摇床培养 (250 rpm) 24 到 48 小时。7,000 rpm 离心 10 min 收集细胞。所得的颗粒状物用 Bronson 超声波仪超声处理 2 分 30 秒, 脉冲间隔为 2 秒。全部的超声处理物在显微镜下镜检。10,000 rpm 离心 10 min 去除可溶部分。所得的含有晶体蛋白的颗粒状物用含 0.5 M NaCl 的 2% Triton X-100 洗 4-5 次。用 0.5 M NaCl (4-5 次) 连续地洗, 最终的颗粒状物用蒸馏水洗 (2 次)。所得的颗粒状物于含 10 mM 二硫苏糖醇的 50 mM Na_2CO_3 缓冲液在 37°C 下溶解 2 h。12,000 rpm 离心 10 min 将已溶的蛋白与不溶物质分离开。蛋白样品用 50 mM Na_2CO_3 , pH 9.0 的缓冲液透析以备进行生物测定。

实施例 2: 生物测定

用纯化的截短 H04 蛋白测定了草地夜蛾 (*fall armyworm*), 红铃麦蛾 (*pink bollworm*), 烟夜蛾, 和欧洲玉米螟 (*European cornborer*) 的 LC50 值, 所述的 H04 蛋白是例如通过上述实施例 1 制

备的。结果如下：

LC50 草地夜蛾 (<i>fall armyworm</i>)	133 ng/cm ²
LC50 红铃麦蛾 (<i>pink bollworm</i>)	691 ng/cm ²
LC50 烟夜蛾 (<i>tobacco budworm</i>)	299 ng/cm ²
LC50 欧洲玉米螟 (<i>European cornborer</i>)	31 ng/cm ²

实施例 3：合成的 H04 基因构建体

编码 H04 毒素部分的合成的核苷酸序列是利用玉米偏嗜密码子表 (Murray 等, Nucl Acids Res. 17:477-498, 1989, 在此引入作为参考) 通过 University of Wisconsin GCG 程序组提供的 “Backtranslation” 程序, 反向翻译 (backtranslating) De Maagd 等, *Appl. Environ. Microbiol.* 62(5): 1537-1543 (1996) 中描述的 H04 杂合毒素的氨基酸序列 (Cry1Ab 的结构域 I 和 II 和 Cry1C 的结构域 III) 而设计的。优选地, 如 WO 93/07278 中所描述的, 对每个氨基酸均应用使用频率最高的玉米密码子。

编码 H04 毒素部分的合成的核苷酸序列可由几个片段构建。每个片段都是通过代表所述基因两条链的长度为 60-75 个核苷酸的 10 对寡聚体的杂交而构建的。连续的寡核苷酸对之间有约 15 个核苷酸的重叠以便于正确的定位和组装。寡核苷酸可以通过, 例如 Genosys Biotechnologies Inc., TX 来合成。使每一对寡聚体杂交, 并使用供应商提供的条件用多核苷酸激酶磷酸化, 所述的多核苷酸激酶可购自 New England Biolabs, Inc., MA。使激酶作用过的片段对杂交并连接到高拷贝质粒中再转化到例如感受态的 DH5 α 内, 所述的质粒载体包含如, 氨苄青霉素抗性基因。将所述的细胞涂布到含有氨苄青霉素的培养基上于 37°C 下培养过夜。筛选插入了 DNA 的克隆。将所述的 DNA 测序, 筛选出含有正确序列的克隆。利用所述片段间独特的限制性位点进行限制性消化, 连接和转化。

SEQ ID NO:3 是编码 631-氨基酸的 H04 (不带有尾区域) 毒素部分合成的核苷酸序列, SEQ ID NO:4 是由 SEQ ID NO:3 中描述的合成的核苷酸序列编码的 H04 毒素的氨基酸序列。SEQ ID NO:11 是构建体 pNOV1308 的核苷酸序列, 其包含与 SEQ ID NO:3 中所示的合成的 H04 基因序列可操作性连接的组成型的玉米遍在蛋白启动子。

除上述描述的仅编码 H04 杂合体毒素部分(Cry1Ab 的结构域 I 和 II 以及 Cry1C 的结构域 III)的合成基因(SEQ ID NO:3)外, 还利用如 U. S. Patent No. 5,625,136 (引入此处作为参考)所述的与 H04 毒素部分 3' 末端相融合的全部或部分合成的 *cry1Ab* 尾区域构建其它的合成的 H04 基因。这些带有 *cry1Ab* 尾的合成的 H04 基因序列如下:

SEQ ID NO:5 是编码 H04 毒素部分加全长的 Cry1Ab 尾部分的合成的核苷酸序列, SEQ ID NO:6 是由 SEQ ID NO:5 中所述的合成的核苷酸序列编码的 H04 + Cry1Ab 尾的氨基酸序列。SEQ ID NO:12 是构建体 pNOV1436 的核苷酸序列, 其包含与如 SEQ ID NO:5 中所述的合成的 H04 基因序列可操作连接的根优选的玉米 MTL 启动子。SEQ ID NO:13 是构建体 pNOV1441 的核苷酸序列, 其包含与如 SEQ ID NO:5 中所述的合成的 H04 基因序列可操作连接的组成型的玉米遍在蛋白启动子。

SEQ ID NO:7 是编码 H04 毒素部分加全长的 Cry1Ab 尾部分的另一种合成的核苷酸序列, SEQ ID NO:8 由 SEQ ID NO:7 中所述的合成的核苷酸序列编码的 H04 + Cry1Ab 尾的氨基酸序列。SEQ ID NO:14 是构建体 pNOV1305 的核苷酸序列, 其包含与如 SEQ ID NO:7 中所述的合成的 H04 基因序列可操作连接的组成型的玉米遍在蛋白启动子。SEQ ID NO:15 是构建体 pNOV1313 的核苷酸序列, 其包含与如 SEQ ID NO:7 中所述的合成的 H04 基因序列可操作连接的组成型的玉米遍在蛋白启动子。

SEQ ID NO:9 是编码 H04 毒素部分加 Cry1Ab 尾头 40 个氨基酸的合成的核苷酸序列, SEQ ID NO:10 是由 SEQ ID NO:9 中描述的合成的核苷酸序列编码的 H04 + 40 氨基酸的截短的 Cry1Ab 尾的氨基酸序

列, SEQ ID NO:16 是构建体 pNOV1435 的核苷酸序列, 其包含与如 SEQ ID NO:9 中所述的合成 H04 基因序列可操作地连接的根偏好的玉米 MTL 启动子。SEQ ID NO:17 是构建体 pZU578 的核苷酸序列, 其包含与如 SEQ ID NO:9 中所述的合成 H04 基因序列可操作连接的鼠耳芥属 (*Arabidopsis*) 肌动蛋白-2 启动子。

实施例 4: 编码序列及其临近序列的修饰

本申请中所述的核苷酸序列可进行修饰以在转基因植物宿主中表达。表达所述核苷酸序列并在其细胞中产生杀虫毒素的宿主植物具有增强的对昆虫袭击的抗性, 并更好地武装起来以抵抗与所述侵袭相关的作物损失。

微生物来源的基因在植物中进行转基因表达可能需要修饰这些基因以实现并优化其在植物中的表达。具体说来, 编码单独的酶的细菌的 ORFs, 虽然在天然微生物中是由同一个转录本编码的, 但在植物中需要在不同的转录本上才能较好地表达。为此, 每种微生物的 ORF 均需单独分离出来, 并克隆到在所述 ORF 5' 末端提供植物启动子、在所述 ORF 3' 末端提供植物转录终止子的表达盒中。所述的经分离的 ORF 序列优选地包含起始 ATG 和终止密码子, 但也可以包含起始 ATG 和终止密码子以外的序列。除此之外, 所述的 ORF 可以是截短的, 但需保留所需的活性; 对于特别长的 ORFs, 优选地使用其保留活性的截短形式在转基因生物中表达。“植物启动子”和“植物转录终止子 (transcriptional terminator)”是指可在植物细胞中起作用的启动子和转录终止子。这也包括非植物如病毒(一个例子是花椰菜花叶病毒)来源的启动子和转录终止子。

在某些情况下, 不需要对编码序列的 ORF 和邻近序列进行修饰。分离出含有有用的 ORF 的片段并插入到植物启动子的下游就足够。例如, Gaffney 等 (*Science* 261: 754-756 (1993)) 在没有对编码序列进行修饰的前提下, 已在转基因植物中成功地表达了在 CaMV 35S 启动子和 CaMV *tm1* 终止子的控制下的假单胞菌属 (*Pseudomonas*)

nahG 基因, 并且 *nahG* ORF 还附加了 ATG 的上游 x bp 的假单胞菌基因, 和终止密码子下游 y bp。优选地, 尽可能少地在 ATG 的上游和终止密码子的下游遗留微生物的序列。在实践中这种构建要依赖于所能够获得的限制性位点。

在另外一些情况下, 微生物来源的基因的表达可能存在一些问题。这些问题已在现有技术中有详细地描述, 且对某些来源, 如芽孢杆菌的基因有特定的共性。这些问题也存在于本发明的核苷酸序列, 但可以用本领域已知的技术对这些基因进行修饰。可能遇到的问题如下:

1. 密码子的使用

植物中的优选密码子使用与在特定微生物中优选的密码子使用不同。将所克隆的微生物 ORF 中的密码子使用与在植物基因 (特别是来自靶植物的基因) 中的密码子使用进行比较可以鉴定出所述 ORF 中应当优选地被改变的密码子。典型地, 植物进化表现出对于单子叶植物在第三个碱基位置上对核苷酸 C 和 G 的强烈偏嗜性, 而对于双子叶植物在这一位置上常常使用核苷酸 A 或 T。通过对基因进行修饰掺入在特定的靶生物种中优选使用的密码子, 下述的许多有关 GC/AT 含量和不适当剪接的问题都可以得到克服。

2. GC/AT 含量

植物基因典型地具有高于 35% 的 GC 含量。富含 A 和 T 核苷酸的 ORF 序列将在植物中引起许多问题。首先, 据信 ATTTA 基元引起信使 RNA (message) 的去稳定, 且发现于许多短寿的 mRNA 的 3' 末端。第二, 据信聚腺苷酸化信号, 如 AATAAA 在信使 RNA 内不适当位置的出现将引起转录的成熟前截断。另外, 单子叶植物可能将富含-AT 的序列作为剪接位点识别 (见下)。

3. 邻近起始甲硫氨酸的序列

与微生物不同，植物的信使 RNA 没有明确的核糖体结合位点。而是，据信核糖体附加于信使 RNA 的 5' 末端扫描第一个可利用的 ATG 由此起始翻译。虽然如此，其对临近 ATG 的某些序列具有偏嗜，因此可以通过在微生物基因中 ATG 处包含真核细胞共有的翻译起始序列来增强其表达。Clontech (1993/1994 catalog, page 210, 在此引入作为参考) 建议了一种用于在植物中表达大肠杆菌 *uidA* 基因的作为翻译起始子的序列。另外，Joshi (NAR 15: 6643-6653 (1987), 在此引入作为参考) 比较了许多邻近 ATG 的植物序列，提示了另一种共有序列。当在植物中表达微生物 ORF 遇到困难时，在起始 ATG 处掺入上述序列中的一种可能会对翻译有改善。在这些情况下，由于对第二 AA 残基的修饰，共有序列的最后三个核苷酸可能不适于掺入修饰的序列中。在不同植物种中，优选的邻近起始甲硫氨酸的序列可能是不同的。对 14 种记录于 GenBank 数据库中的玉米基因进行调查给出了下述的结果：

在 14 种玉米基因中起始 ATG 前的位置：

	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1
C	3	8	4	6	2	5	6	0	10	7
T	3	0	3	4	3	2	1	1	1	0
A	2	3	1	4	3	2	3	7	2	3
G	6	3	6	0	6	5	4	6	1	5

这一分析可针对有待掺入所述核苷酸序列的目的植物来进行，修饰邻近 ATG 的序列以掺入优选的核苷酸。

4. 去除非正常的剪接位点

由非植物来源克隆的且未经为适于在植物中表达而进行优化的基因也可能包含在植物中被识别为 5'或 3'剪接位点的基元,从而使所述基因被切割产生截短的或缺失的信使 RNA。这些位点可使用本领域熟知的方法去除。

用于修饰编码序列及邻近序列的技术是本领域已知的。在微生物 ORF 起始表达量低并且确定适合于对所述序列进行上述改变的情况下，可以利用本领域已知的技术来构建合成的基因。它们是在已出版的专利说明书 EP 0 385 962, EP 0 359 472 和 WO 93/07278 中描述的，上述文献均引入此处作为参考。在大多数情况下优选在其转移到转基因植物中以前，用瞬时分析方法（其在本领域是已知的）分析基因构建体的表达。

实施例 5：植物表达盒的构建

用于在转基因植物中表达的编码序列首先组装在表达盒中置于能在植物中进行表达的启动子后。所述的表达盒还可包括转基因表达所需或为其选择的其它序列。这些序列包括但不限于，转录终止子，加强表达的外来序列，如内含子，必需序列（vital sequences），和为使基因产物靶定特定的细胞器和细胞区室的序列。这些表达盒易于转移到下述的植物转化载体中。下面是对典型的表达盒的不同元件的描述。

1. 启动子

用于表达盒的启动子的选择将确定转基因在转基因植物中的时空表达方式。所选择的启动子将在特定的细胞类型（如叶表皮细胞，叶肉细胞，根皮层细胞）、或在特定的组织或器官（例如，根，叶或花）中表达转基因，所述的选择将反映有待积累所述基因产物的位置。另外，所选择的启动子可以在多种诱导条件下驱动所述基因的表达。不同启动子其强度（即启动转录的能力）不同。根据所使用的宿主细胞系统，可使用任何一种合适的启动子，它们包括所述基因的天然启动子。下面是可用于表达盒中的启动子的非限制性例子。

a. 组成型表达，遍在蛋白启动子：

遍在蛋白是已知在多种细胞类型中积累的基因产物，已从许多物

种中克隆出了它的启动子以用于转基因植物（如向日葵 - Binet 等 *Plant Science* 79: 87-94 (1991); 玉米 - Christensen 等 *Plant Molec. Biol.* 12: 619-632 (1989); 和拟南芥属 - Norris 等, *Plant Mol. Biol.* 21:895-906 (1993))。所述的玉米遍在蛋白启动子已在转基因单子叶系统中开发出来, 其序列和为进行单子叶植物转化所构建的载体已在专利出版物 EP 0 342 926 中公开, 该文献在此引入作为参考。Taylor 等 (*Plant cell Rep.* 12: 491-495 (1993)) 描述了一种载体 (pAHC25), 其包含玉米遍在蛋白启动子和第一内含子, 当它经微粒轰击引入多种单子叶植物后, 在所述植物细胞悬液中具有高活性。拟南芥属遍在蛋白启动子可理想地与本发明的核苷酸序列一起应用。所述的遍在蛋白启动子适合于在转基因的单子叶植物和双子叶植物中表达。合适的载体是经引入合适的遍在蛋白启动子和/或内含子序列而修饰的 pAHC25 或本说明书中所述的任意一种转化载体的衍生物。

b. 组成型的表达, CaMV 35S 启动子:

质粒 pCGN1761 的构建在已出版的专利申请 EP 0 392 225 (实施例 23) 中描述, 该文献引入此处作为参考。pCGN1761 包含“双” CaMV 35S 启动子和 *tm1* 转录终止子, 并在所述的启动子和终止子之间带有独特的 *EcoRI* 位点, 且具有 pUC-型的骨架。一种构建的 pCGN1761 衍生物具有经修饰的多接头, 除已存在的 *EcoRI* 位点外该接头还包含 *NotI* 和 *XhoI* 位点。将这种衍生物命名为 pCGN1761ENX。pCGN1761ENX 对于在其多接头内克隆 cDNA 序列或编码序列 (包含微生物的 ORF 序列) 是有用的, 这可以使所述基因在转基因植物中在 35S 启动子的控制下表达。这种构建体的完整的 35S 启动子-编码序列-*tm1* 终止子盒可被启动子 5'端的 *HindIII*, *SphI*, *SaII*, 和 *XbaI* 位点和终止子 3'端的 *XbaI*, *BamHI* 和 *BglII* 位点切割下来, 转移到转化载体, 如下述的载体中。而且, 所述的双 35S 启动子片段可用 *HindIII*, *SphI*, *SaII*, *XbaI*, 或 *PstI* 由 5'切割, 或用任意一种多接

头限制位点(*EcoRI*, *NotI* 或 *XhoI*)由 3'切割, 替换为其它的启动子。如果需要还可以通过在克隆位点周围引入增强翻译的序列以进行修饰。如果需要过表达则这种方法特别有用。例如, pCGN1761ENX 可由 U.S. Patent No. 5,639,9493 实施例 37 中描述的对翻译起始位点的优化进行修饰, 上述文献在此引入作为参考。

c. 组成型表达, 肌动蛋白启动子:

已知几种肌动蛋白同种型在大多数细胞类型中表达, 因此, 所述的肌动蛋白启动子是组成型启动子的一个很好的选择。具体说来, 来自稻 *Act1* 基因的启动子已被克隆和描述(McElroy 等 Plant Cell 2: 163-171 (1990))。已发现在该启动子的 1.3kb 片段内包含在稻原生质体中表达所需的所有调控元件。而且, 已构建了多种基于 *Act1* 启动子的表达载体, 特定地用于单子叶植物 (McElroy 等 Mol. Gen. Genet. 231: 150-160 (1991))。它们并入了 *Act1*-内含子 1, *Adh1* 5' 侧翼序列和 *Adh1*-内含子 1 (来自玉米醇脱氢酶基因) 和来自 CaMV 35S 启动子的序列。表现出最高表达的载体是 35S 和 *Act1* 内含子或 *Act1* 5'侧翼序列以及 *Act1* 内含子的融合体。在起始 ATG (属 GUS 报告基因)周围序列的优化也可增强表达。可容易地对 McElroy 等 (Mol. Gen. Genet. 231: 150-160 (1991))所述的启动子表达盒进行修饰以用于基因表达, 特别是在单子叶植物宿主中。例如, 由 McElroy 构建体中分离出含启动子的片段, 用于替换 pCGN1761ENX 中的双 35S 启动子, 使其可用于插入特异的基因序列。可将上述构建的融合基因转移到合适的转化载体中。在单独的报道中还发现, 带有其第一内含子的稻 *Act1* 启动子也可以指导在培养的大麦细胞中的高表达(Chibbar 等 Plant Cell Rep. 12: 506-509 (1993))。

d. 诱导型表达, PR-1 启动子:

pCGN1761ENX 中的双 35S 启动子可用任意其它的启动子替换以期达到合适的高表达水平。通过例如, 一种 U.S. Patent No.

5,614,395 中所描述的化学调节的启动子,如烟草 PR-1a 启动子,可用于替换上述的双 35S 启动子。或者,也可使用 Lebel 等, *Plant J.* 16:223-233 (1998) 中所述的拟南芥属 PR-1 启动子。优选地用限制性酶将所选择的启动子从其来源中切割出来,但也可由带有合适的末端限制位点的引物通过 PCR 扩增。若进行 PCR 扩增,则在扩增启动子克隆到靶载体中后,所述的启动子应当经过再测序,检查是否有扩增错误。所述的可由化学/病原体调节的烟草 PR-1a 启动子可由质粒 pCIB1004 (有关的构建参见 EP 0 332 104 中的实施例 21,该文献引入此处作为参考) 中切割出来,然后转移到质粒 pCGN1761ENX (Uknes 等, *Plant Cell* 4: 645-656 (1992)) 中。pCIB1004 再经 *NcoI* 切割,所得的线性片段的 3' 突出端通过 T4 DNA 多聚酶的处理钝化。再将所述片段用 *HindIII* 切割,所得的含 PR-1a 启动子的片段经凝胶纯化后克隆到已去除了双 35S 启动子的 pCGN1761ENX 中。这是通过下述操作完成的,即用 *XhoI* 切割,再由 T4 聚合酶钝化,接下来用 *HindIII* 切割,然后分离克隆了 pCIB1004 启动子片段的大的含有载体终止子的片段。由此产生了 pCGN1761ENX 衍生物,其带有 PR-1a 启动子和 *tml* 终止子以及所插入的带有独特的 *EcoRI* 和 *NotI* 位点的多接头。可将选择的编码序列插入这一载体中,所得的融合产物(即,启动子-基因-终止子)可转移到任何所选择的转化载体中,包括上文中所述的载体。多种化学调节剂可用于诱导所选择的编码序列在根据本发明转化的植物中的表达,所述的化学调节剂包括 U.S. 专利 Nos. 5,523,311 和 5,614,395 中所述的苯并噻二唑(benzothiadiazole),异烟酸,和水杨酸化合物。

e. 诱导型表达,乙醇可诱导的启动子:

可由特定的醇类或酮类,如乙醇诱导的启动子也可用于本发明编码序列的诱导型表达。这种启动子如来自构巢曲霉(*Aspergillus nidulans*)的 *alcA* 基因启动子(Caddick 等(1998) *Nat. Biotechnol* 16:177-180)。在构巢曲霉中,所述的 *alcA* 基因编码醇

脱氢酶 I, 在存在化学诱导剂的情况下, 其表达可受 AlcR 转录因子的调节。为本发明的目的, 在含与最小 35S 启动子(Caddick 等 (1998) *Nat. Biotechnol* 16:177-180)融合的 *alcA* 基因启动子序列的质粒 p_{alcA}: CAT 中 CAT 编码序列被本发明的编码序列所替代, 形成具有在 *alcA* 基因启动子控制下的编码序列的表达盒。这可由本领域已知的方法进行。

f. 诱导型表达, 糖皮质激素诱导的启动子

本发明还包括用基于类固醇激素的系统诱导本发明的核苷酸序列表达。例如, 使用糖皮质激素介导的诱导系统(Aoyama 和 Chua (1997) *The Plant Journal* 11: 605-612), 基因表达可通过应用糖皮质激素来诱导, 例如合成的糖皮质激素, 优选地塞米松, 优选浓度范围是 0.1mM 到 1mM, 更优选从 10mM 到 100mM。为本发明的目的, 萤光素酶基因序列可由本发明的核酸序列所替代以形成具有本发明核酸序列的表达盒, 所述核酸序列受 6 拷贝的与 35S 最小启动子相融合的 GAL4 上游激活序列调控。这可用本领域中已知的技术进行操作。反式作用因子包含 GAL4 DNA 结合结构域(Keegan 等 (1986) *Science* 231: 699-704), 其与疱疹病毒蛋白 VP16 反式活化域(Triezenberg 等 (1988) *Genes Devel.* 2: 718-729)相融合, 后者与大鼠糖皮质激素受体的激素结合结构域 (Picard 等 (1988) *Cell* 54: 1073-1080)相融合。所述融合蛋白的表达受本领域已知的或此处所述的任何合适的启动子调控。这一表达盒也包含在含有与 6xGAL4/最小启动子相融合的本发明核酸序列的植物中。由此, 实现融合蛋白的组织或器官特异性, 导致所述杀虫毒素的可诱导的组织或器官特异性。

g. 根特异性的表达:

另一种基因表达方式是根表达。合适的根启动子是 de Framond (FEBS 290: 103-106 (1991))及 U. S. Patent No. 5,466,785 中所描述的玉米金属硫蛋白样(MTL)基因的启动子, 上述文献引入此处作

为参考。这种“MTL”启动子转移到合适的载体，如 pCGN1761ENX 中以插入所选择的基因，接下来转移完全的启动子-基因-终止子盒到目的转化载体中。

h. 创伤诱导型启动子：

创伤诱导型启动子也适合应用于基因表达。目前已有多种此种类型的启动子的描述(如 Xu 等 Plant Molec. Biol. 22: 573-588 (1993), Logemann 等 Plant Cell 1: 151-158 (1989), Rohrmeier & Lehle, Plant Molec. Biol. 22: 783-792 (1993), Firek 等 Plant Molec. Biol. 22: 129-142 (1993), Warner 等 Plant J. 3: 191-201 (1993)), 所有这些启动子均适用于本发明。Logemann 等描述了双子叶的马铃薯 *wun1* 基因 5'上游序列。Xu 等描述了来自双子叶植物马铃薯 (*pin2*) 的创伤诱导型启动子在单子叶的稻中具有活性。而 Rohrmeier & Lehle 描述了玉米 *Wip1* cDNA 的克隆, 其也是创伤诱导的且可用于通过标准的技术分离同族的启动子。类似地, Firek 等和 Warner 等描述了来自单子叶的石刁柏 (*Asparagus officinalis*) 的创伤诱导型启动子, 该启动子在局部创伤和病原体侵入位点表达。利用本领域已知的克隆技术, 可将这些启动子转化到合适的载体中, 与本发明的基因相融合, 用于在植物创伤位点表达所述基因。

i. 髓 (pith) 偏好的表达：

引入此处作为参考的专利申请 WO 93/07278 描述了玉米 *trpA* 基因的分离, 该基因偏好在髓细胞中表达。给出了所述的基因序列和由转录起点延伸至-1726 bp 的启动子。利用标准的分子生物学技术, 可将这一启动子, 或其片段转移到载体, 如 pCGN1761 中, 在该载体中其可替换 35S 启动子, 并用于驱动外源基因以髓偏好的方式表达。实际上, 含有髓偏好启动子或其部分的片段可转移到任何载体中, 并可被修饰以用于在转基因植物中表达。

j. 叶特异性的表达:

编码磷酸烯醇羧化酶 (PEPC) 的玉米基因已被 Hudspeth 和 Grula (Plant Molec Biol 12: 579-589 (1989)) 公开。利用标准的分子生物学技术, 该基因的启动子可在转基因植物中以叶特异的方式, 驱动任何基因的表达。

k. 花粉特异的表达:

WO 93/07278 中描述了玉米钙依赖性蛋白激酶 (CDPK) 基因的分离, 该基因在花粉细胞中表达。该基因序列和启动子延伸至从转录起点起多达 1400 bp。利用标准的分子生物学技术, 可将这一启动子, 或其片段转移到载体, 如 pCGN1761 中, 在该载体中其可替换 35S 启动子, 并用于驱动本发明核酸序列以花粉特异性的方式表达。

1. 在化学配体存在下由受体介导的反式激活作用:

引入此处作为参考的 U.S. Patent No. 5,880,333 中描述了一种系统, 其中 II 类激素受体, 如作为异源二聚体行使功能的蜕皮激素受体 (EcR) 和 Ultraspiracle (USP), 在存在合适的化学配体, 如 tebufenozide 时, 在植物细胞中调节靶多肽的表达。

2. 转录终止子

有多种可用于表达盒的转录终止子。它们负责使转基因转录终止并进行正确的聚腺苷酸化。合适的转录终止子是已知可在植物中行使功能的那些, 其包括 CaMV 35S 终止子, *tml* 终止子, 胭脂碱合酶终止子和豌豆 *rbcS* E9 终止子。它们均可用于单子叶植物和双子叶植物。另外, 也可以使用基因的天然转录终止子。

3. 增强或调节表达的序列

已发现有很多序列可用于增强转录单位中的基因表达, 可将这些

序列与本发明的基因相连接以提高所述基因在转基因植物中的表达。

已表明有多种内含子序列可增强基因的表达，特别是在单子叶植物细胞中。例如当将玉米 *AdhI* 基因的内含子引入玉米细胞时，其能显著地增强在其同类启动子控制下的基因的表达。已表明内含子 1 特别有效并增强带有氯霉素乙酰转移酶基因的融合构建体的表达 (Callis 等, *Genes Develop.* 1: 1183-1200 (1987))。在相同的实验系统中，来自玉米 *bronze1* 基因的内含子对增强表达有类似的效果。内含子序列已常规地引入到植物转化载体中，典型地，处在非翻译前导序列中。

已知衍生自病毒的大量非翻译前导序列有增强表达的作用，并且在双子叶植物细胞中特别有效。具体地，来自烟草花叶病毒 (TMV, "W-序列"), 玉米萎黄病斑点病毒 (Maize Chlorotic Mottle Virus, MCMV), 和苜蓿花叶病毒 (AMV) 的前导序列能有效地增强表达 (如 Gallie 等 *Nucl. Acids Res.* 15: 8693-8711 (1987); Skuzeski 等 *Plant Molec. Biol.* 15: 65-79 (1990))。

4. 在细胞内基因产物的靶定 (targeting)

已知在植物中存在多种靶定基因产物的机制，并且所述机制中控制该功能的序列已有所描述。例如，将基因产物靶定到叶绿体是由在多种蛋白氨基末端发现的信号序列控制的，所述的信号序列在向叶绿体运输过程中被切除以获得成熟的蛋白 (如 Comai 等 *J. Biol. Chem.* 263: 15104-15109 (1988))。这些信号序列可与异源基因产物相融合使所述的异源产物运送到叶绿体 (van den Broeck, 等 *Nature* 313: 358-363 (1985))。编码合适的信号序列的 DNA 可由编码 RUBISCO 蛋白、CAB 蛋白, EPSP 合酶, GS2 蛋白以及其它已知定位于叶绿体的蛋白的 cDNA 的 5' 末端分离。参见 U.S. Patent No. 5,639,949 实施例 37 中以“靶定叶绿体的表达”为标题的部分。

其它的基因产物定位于其它的细胞器，如线粒体和过氧化物酶体 (如 Unger 等 *Plant Molec. Biol.* 13: 411-418 (1989))。也可对

编码这些产物的 cDNA 进行进行操作以实现所述异源基因向这些细胞器的靶定。这种序列的例子是核编码的 ATP 酶和对线粒体特异的天冬氨酸氨基转移酶的同种型。Rogers 等描述了导向性细胞蛋白体 (Proc. Natl. Acad. Sci. USA 82: 6512-6516 (1985))。

另外也有对使基因产物靶定其它细胞区室的序列的描述。氨基末端序列负责靶向 ER, 质外体, 及由糊粉细胞向胞外分泌 (Koehler & Ho, Plant Cell 2: 769-783 (1990))。另外, 氨基末端序列与羧基末端序列联合负责基因产物的液泡靶定 (Shinshi 等 Plant Molec. Biol. 14: 357-368 (1990))。

通过将合适的上述靶向序列与有用的转基因序列相融合可能指导转基因产物到任何细胞器或细胞区室。对于叶绿体靶定, 例如, 可将来自 RUBISCO 蛋白、CAB 蛋白, EPSP 合酶, GS2 基因的叶绿体信号序列与转基因的氨基末端 ATG 进行框内融合。所选择的信号序列应包含已知的切割位点, 而且构建融合构建体应考虑到任何处于切割位点之后的对切割所必需的氨基酸。有时候为了达到上述要求可在切割位点和转基因 ATG 之间添加少量的氨基酸, 或者替换转基因序列内的某些氨基酸。用于输入叶绿体而构建的融合体可按照 Bartlett 等在 Edelman 等 (编) Methods in chloroplast Molecular Biology, Elsevier pp 1081-1091 (1982) 和 Wasmann 等在 Mol. Gen. Genet. 205: 446-453 (1986) 中所描述的技术, 通过体外转录的构建体的体外翻译检测叶绿体摄取效果。这些构建技术是本领域周知的, 也可应用于线粒体和过氧化物酶体。

上面所描述的细胞靶定机制不仅可与同源启动子 (Cognate promoters) 结合使用, 也可以与异源启动子一起使用, 在该启动子的转录调节下实现特定的细胞靶定目的, 该启动子的表达方式与靶定信号序列原有的启动子的表达方式是不同的。

实施例 6: 植物转化载体的构建

本领域的技术人员已知存在多种用于植物转化的转化载体, 本发

明的基因可与上述的任何一种载体结合使用。所述载体的选择依赖于优选的转化技术和待转化的靶物种。对于特定的靶物种，不同的抗生素或除草剂选择标记可能是优选的。在转化中常规使用的选择标记包括 *nptII* 基因，其可赋予卡那霉素以及相关的抗生素抗性 (Messing & Vierra. *Gene* 19: 259-268 (1982); Bevan 等, *Nature* 304:184-187 (1983)), *bar* 基因，其可赋予除草剂膦丝菌素抗性 (White 等, *Nucl. Acids Res* 18: 1062 (1990), Spencer 等 *Theor. Appl. Genet* 79: 625-631 (1990)), *hph* 基因，其可赋予抗生素潮霉素抗性 (Blochinger & Diggelmann, *Mol Cell Biol* 4: 2929-2931), 和 *dhfr* 基因，其可赋予 methatrexate 抗性 (Bourouis 等, *EMBO J.* 2(7): 1099-1104 (1983)), EPSPS 基因，其可赋予草甘膦抗性 (U.S. Patent Nos. 4,940,935 and 5,188,642), 和甘露糖-6-磷酸异构酶基因，其可提供代谢甘露糖的能力 (U.S. Patent Nos. 5,767,378 和 5,994,629)。

1. 适合于土壤杆菌转化的载体

许多载体适用于利用根癌土壤杆菌 (*Agrobacterium tumefaciens*) 的转化。它们典型地至少带有一个 T-DNA 边界序列，包括载体例如 pBIN19 (Bevan, *Nucl. Acids Res.* (1984)) 和 pXYZ。下面描述了两种适合用于利用土壤杆菌进行转化的典型载体。

a. pCIB200 和 pCIB2001:

用所述的二元载体 pCIB200 和 pCIB2001 构建与土壤杆菌一起使用的重组载体，其构建方式如下。pTJS75kan 是通过 *NarI* 消化 pTJS75 而创建的 (Schmidhauser & Helinski, *J. Bacteriol.* 164: 446-455 (1985))，其切除了四环素抗性基因，然后插入来自带有 NPTII 的 pUC4K 的 *AccI* 片段 (Vieira & Messing, *Gene* 19: 259-268 (1982); Bevan 等, *Nature* 304: 184-187 (1983); McBride 等, *Plant Molecular Biology* 14: 266-276 (1990))。XhoI 接头与 PCIB7 的

EcoRV 片段相连接, 所述的 *EcoRV* 片段包含左右 T-DNA 边界, 植物选择性的 *nos/nptII* 嵌合基因和 pUC 多接头 (Rothstein 等, Gene 53: 153-161 (1987)), 将 *XhoI* 消化的片段克隆到 *SaII*-消化的 pTJS75kan 中以构建 pCIB200 (参见 EP 0 332 104, 实施例 19)。pCIB200 包含下述独特的多接头限制性位点: *EcoRI*, *SstI*, *KpnI*, *BglII*, *XbaI*, 和 *SaII*。pCIB2001 是 pCIB200 的衍生物, 其是通过将附加的位点插入多接头中构建的。pCIB2001 的多接头中的独特的限制性位点是 *EcoRI*, *SstI*, *KpnI*, *BglII*, *XbaI*, *SaII*, *MluI*, *BclI*, *AvrII*, *ApaI*, *HpaI*, 和 *StuI*。pCIB2001, 除包含上述独特的限制性位点外, 还包含植物和细菌卡那霉素选择标记, 用于土壤杆菌介导的转化的左右 T-DNA 边界, RK2-衍生的 *trfA* 的功能, 用于在大肠杆菌和其它宿主之间的移动, 和也来自 RK2 的 *OriT* 和 *OriV* 功能。所述的 pCIB2001 多接头适合于克隆含有其本身调节信号的植物表达框。

b. pCIB10 及其潮霉素选择衍生物:

二元载体 pCIB10 包含编码卡那霉素抗性的基因以便于在植物中进行筛选, 和 T-DNA 左右侧边界序列, 并掺入了来自宽宿主范围的质粒 pRK252 的序列, 使其能在大肠杆菌和土壤杆菌间复制。这一构建体已由 Rothstein 等 (Gene 53: 153-161 (1987)) 描述。构建了多种 pCIB10 衍生物, 其掺入了 Gritz 等 (Gene 25: 179-188 (1983)) 描述的潮霉素 B 磷酸转移酶基因。这些衍生物使得可以仅通过潮霉素 (pCIB743), 或通过潮霉素和卡那霉素 (pCIB715, pCIB717) 筛选转基因植物细胞。

2. 适合于非土壤杆菌转化的载体

不使用根癌土壤杆菌 (*Agrobacterium tumefaciens*) 就避免了选择载体时对 T-DNA 序列的需要, 因此除例如上述的含 T-DNA 序列的载体外, 还可以使用不含 T-DNA 序列的载体。无需借助于土壤杆菌的转化技术包括通过粒子轰击, 原生质体摄入 (如 PEG 和电穿孔) 以及

微注射的转化。对载体的选择很大程度上依赖于对被转化物种的优选筛选。下面描述了用于非土壤杆菌转化的典型载体的构建。

a. pCIB3064:

pCIB3064 是 pUC-衍生的载体, 适用于与用除草剂 basta (或膦丝菌素) 筛选相结合的直接基因转移。质粒 pCIB246 包含与大肠杆菌 GUS 基因可操作地融合的 CaMV 35S 启动子和 CaMV 35S 转录终止子, 该质粒已在 PCT 公开申请 WO 93/07278 中公开。该载体的 35S 启动子包含两个起始位点 5'端的 ATG 序列。可使用标准的 PCR 技术使上述位点突变, 由此去除 ATG 并产生限制性位点 *SspI* 和 *PvuII*。新的限制性位点分别距独特的 *SalI* 位点 96 和 37bp, 距实际的起始位点 101 和 42bp。将所得的 pCIB246 衍生物命名为 pCIB3025。然后通过 *SalI* 和 *SacI* 消化将 GUS 从 pCIB3025 中切除, 使末端成为平端并再连接以形成质粒 pCIB3060。质粒 pJIT82 可由 John Innes Centre, Norwich 获得, 切割含有来自绿色产色链霉菌 (*Streptomyces viridochromogenes*) 的 *bar* 基因的 400bp *SmaI* 片段, 插入 pCIB3060 的 *HpaI* 位点 (Thompson 等 EMBO J 6: 2519-2523 (1987))。由此产生了 pCIB3064, 其包含在 CaMV 35S 启动子和终止子控制下的 *bar* 基因用于除草剂筛选, 以及氨基青霉素抗性基因 (用于在大肠杆菌中进行筛选) 和带有下述独特位点的多接头, 即 *SphI*, *PstI*, *HindIII*, 和 *BamHI*。这一载体适合于克隆含有自身调节信号的植物表达盒。

b. pSOG19 和 pSOG35:

pSOG35 是一种转化载体, 其利用大肠杆菌基因, 赋予氨基喋呤抗性的二氢叶酸还原酶 (DFR) 作为选择标记。用 PCR 自 pSOG10 扩增 35S 启动子 (-800 bp), 来自玉米 *Adh1* 基因的内含子 6 (-550 bp) 和 18 bp GUS 非翻译前导序列。编码大肠杆菌二氢叶酸还原酶 II 型基因的 250-bp 也通过 PCR 扩增, 将这两个 PCR 片段与 pB1221 (Clontech) 的 *SacI*-*PstI* 片段组装, 所述的 pB1221 包含 pUC19 载体骨架和胭脂

氨酸合酶终止子。将这些片段组装起来形成 pSOG19, 其包含与内含子 6 序列融合的 35S 启动子, GUS 前导序列, DHFR 基因和胭脂氨酸合酶终止子。用来自玉米萎黄病斑纹病毒 (Maize Chlorotic Mottle Virus (MCMV)) 的前导序列替换 pSOG19 中的 GUS 前导序列产生载体 pSOG35。pSOG19 和 pSOG35 带有 pUC 基因用于氨苄青霉素抗性, 并具有用于克隆外来物质的 *HindIII*, *SphI*, *PstI* 和 *EcoRI* 位点。

3. 适合于叶绿体转化的载体

利用质体转化载体 pH143 (WO 97/32011, 实施例 36) 在植物质体中表达本发明的核苷酸序列。所述的核苷酸序列插入到 pH143 中, 由此替换 PROTOX 编码序列。这一载体用于质体转化, 并筛选壮观霉素抗性的转化子。或者, 所述的核苷酸序列插入到 pH143 中, 由此其替换 *aadH* 基因。在这种情况下, 则筛选对 PROTOX 抑制剂具有抗性的转化子。

实施例 7: 转化

一旦本发明的核酸克隆到表达载体系统中, 其就可以转化到植物细胞中。植物转化和再生的方法是本领域已知的。例如, Ti 质粒载体已被用于递送外源 DNA, 其它方法有直接 DNA 摄入, 脂质体, 电穿孔, 微注射和微粒轰击。另外, 来自土壤杆菌属的细菌可用于转化植物细胞。下面是对代表性的双子叶和单子叶植物转化技术, 及代表性的质体转化技术的描述。

1. 双子叶植物的转化

用于双子叶植物转化的技术是本领域已知的, 其包括基于土壤杆菌的技术和不需要土壤杆菌的技术。非土壤杆菌技术包括由原生质体或细胞直接摄入外源遗传物质。这可通过 PEG 或电穿孔介导的摄入, 粒子轰击-介导的递送, 或微注射来实现。这些技术的实例在下述文献中有描述, Paszkowski 等, EMBO J 3: 2717-2722 (1984),

Potrykus 等, Mol. Gen. Genet. 199: 169-177 (1985), Reich 等, Biotechnology 4: 1001-1004 (1986), 和 Klein 等, Nature 327: 70-73 (1987)。在每种情况下均可通过本领域已知的技术使转化的细胞再生成整个植株。

土壤杆菌介导的是转化双子叶植物的优选技术, 这是因为该技术转化效率高并可在不同的物种中广泛使用。土壤杆菌转化典型地包括将带有外源有用 DNA 的二元载体(如 pCIB200 或 pCIB2001)转移到合适的土壤杆菌菌株中, 其可能依赖于宿主土壤杆菌菌株共存的 Ti 质粒或染色体上带有的 *vir* 基因互补体(对于 pCIB200 和 pCIB2001 如菌株 CIB542(Uknes 等 Plant Cell 5: 159-169 (1993))。所述的二元载体向土壤杆菌的转移是利用带有重组二元载体的大肠杆菌, 带有质粒如 pRK2013 的能将重组二元载体移动到靶土壤杆菌菌株中的辅助大肠杆菌, 由三亲本交配进行的。或者, 重组的二元载体也可以通过 DNA 转化转移到土壤杆菌中(Höfgen & Willmitzer, Nucl. Acids Res. 16: 9877 (1988))。

通过重组土壤杆菌对靶植物种的转化通常包括, 用本领域已知的方法将土壤杆菌与植物外植体共同培养。经转化的组织在带有存在于二元质粒 T-DNA 边界之间的抗生素或除草剂抗性标记物的培养基上再生。

另一种用基因转化植物的方法包括将惰性或生物活性颗粒推进到植物组织和细胞中。这一技术在 U.S. Patent Nos. 4,945,050, 5,036,006, 和 5,100,792 中有描述。一般, 这一过程包括在能有效穿透细胞外表面并导入其内部的条件下, 向细胞推进惰性的或生物活性颗粒。当通过用惰性粒子时, 可通过用含有所需基因的载体包被所述颗粒, 将所述载体引入细胞。或者, 可用所述载体包围靶细胞, 通过粒子的活动(wake)将载体带入细胞。也可将生物活性颗粒(如, 干酵母细胞, 干细菌或噬菌体, 均含有待引入的 DNA)推进到植物细胞组织中。

2. 单子叶植物的转化

大多数单子叶植物物种的转化已经成为常规技术。优选的技术包括用 PEG 或电穿孔技术将基因直接转移到原生质体中，和粒子轰击到愈伤组织中。转化可以用单一 DNA 种类 (single DNA species) 或多 DNA 种类 (multiple DNA species) (即共转化) 转化，这两种技术均适用于本发明。当有相关的需要时，共转化具有下述的优势，即避免完全载体的构建，及产生带有非连接的有用基因和筛选标记的座位，使得能在后续子代中去除筛选标记。但是使用共转化也存在不足，即整合到所述基因组中的分离 DNA 种类转化频率低于 100% (Schocher 等 *Biotechnology* 4: 1093-1096 (1986))。

专利申请 EP 0 292 435, EP 0 392 225, 和 WO 93/07278 描述了用于由玉米原种自交系中制备愈伤组织和原生质体，通过 PEG 或电穿孔转化原生质体，和由转化的原生质体再生玉米植物的技术。Gordon-Kamm 等 (*Plant Cell* 2: 603-618 (1990)) 和 Fromm 等 (*Biotechnology* 8: 833-839 (1990)) 已公开了用粒子轰击转化 A188-衍生的玉米系的技术。另外，WO 93/07278 和 Koziel 等 (*Biotechnology* 11: 194-200 (1993)) 中描述了用粒子轰击转化玉米原种自交系的技术。这一技术利用由授粉后 14-15 天的玉米穗中切除的长 1.5-2.5 mm 的未成熟玉米胚和 PDS-1000He Biolistics 设备进行轰击。

稻的转化也可以利用原生质体或粒子轰击的直接基因转移技术进行。原生质体介导的转化已在 *Japonica*-型和 *Indica*-型 (Zhang 等 *Plant Cell Rep* 7: 379-384 (1988); Shimamoto 等 *Nature* 338: 274-277 (1989); Datta 等 *Biotechnology* 8: 736-740 (1990)) 中有描述。这两种类型也可使用粒子轰击进行常规转化 (Christou 等 *Biotechnology* 9: 957-962 (1991))。另外，WO 93/21335 描述了利用电穿孔的稻转化技术。

专利申请 EP 0 332 581 用于产生、转化和再生 Pooideae 原生质体的技术。这些技术可用于鸭茅属 (*Dactylis*) 和小麦的转化。而且

在 Vasil 等 (Biotechnology 10: 667-674 (1992)) 中也描述了小麦转化技术, 即, 利用粒子轰击到可长期再生的 C 型愈伤组织细胞中, Vasil 等 (Biotechnology 11: 1553-1558 (1993)) 和 Weeks 等 (Plant Physiol. 102: 1077-1084 (1993)) 描述了用粒子轰击未成熟的胚和由未成熟的胚衍生的愈伤组织的技术。但用于小麦转化的优选技术涉及通过粒子轰击未成熟的胚的玉米转化, 并包括在基因递送前的高蔗糖或高麦芽糖步骤。在进行轰击前, 将任意数量的胚 (0.75-1 mm 长) 置于含 3% 蔗糖 (Murashiga & Skoog, Physiologia Plantarum 15: 473-497 (1962)) 和 3mg/l 2,4-D 的 MS 培养基中以进行体细胞胚的诱导, 其在黑暗中进行。在选择进行轰击的那一天, 将胚从诱导培养基上移开并置于渗透剂上 (即, 添加了所需浓度, 典型地为 15%, 的蔗糖或麦芽糖的诱导培养基)。所述的胚进行 2-3h 的质壁分离然后进行轰击。典型的是每靶平皿上 20 个胚, 但这不是非常严格的。合适的带有基因的质粒 (如 pCIB3064 或 pSG35) 通过标准的技术沉淀到微米大小的金颗粒上。每一载有胚的板均通过 DuPont Biolistics® 氦设备射击, 所用的爆裂压为约 1000 psi, 使用标准的 80 网筛。轰击后, 将所述的胚重新置于黑暗中复苏 24 h (仍在渗透剂上)。24 hr 后, 将所述的胚从渗透剂上移走, 重新置于诱导培养基上, 并在再生前保持近一个月。大约一个月后, 将带有发育中的胚胎发生愈伤组织的胚外植体转移到再生培养基 (MS + 1 mg/l NAA, 5 mg/l GA), 所述的培养基进一步包含适当筛选试剂 (对于 pCIB3064 用 10 mg/l basta, 对于 pSOG35 用 2 mg/l 氨基嘌呤)。约一个月后, 发育的芽转移到更大的称为 "GA7s" 的无菌容器中, 其包含半强度的 MS, 2% 蔗糖, 相同浓度的筛选试剂。

利用土壤杆菌的单子叶植物的转化也已有所描述, 参见 WO 94/00977 和 U.S. Patent No. 5,591,616, 这两篇文献均引入此处作为参考。

3. 质体转化

Nicotiana tabacum c.v. 'Xanthi nc' 的种子以 1" 圆形阵列在 T 琼脂培养基上, 每皿萌发 7 个, 并在播种后的 12-14 天用 1 μ m 钨颗粒 (M10, Biorad, Hercules, CA) 进行轰击, 所述的钨颗粒基本上按先前的描述 (Svab, Z. 和 Maliga, P. (1993) *PNAS* 90, 913-917) 用质粒 pPH143 和 pPH145 的 DNA 包被。经过轰击的幼苗在 T 培养基上培养两天后将叶切除, 置在含 500 μ g/ml 壮观霉素二盐酸盐 (Sigma, St. Louis, MO) 的 RMOP 培养基上 (Svab, Z., Hajdukiewicz, P. 和 Maliga, P. (1990) *PNAS* 87, 8526-8530), 运轴侧向上, 给予强光 (350-500 μ mol 光子/ m^2/s)。将在轰击后 3 到 8 周在漂白叶下出现的抗性芽亚克隆到相同的选择培养基上以形成愈伤组织, 分离次级芽并亚克隆。在独立的亚克隆中经转化的质体基因组拷贝的完全分离 (homoplasmy) 可通过标准的 Southern 印迹技术 (Sambrook 等, (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor) 评估。BamHI/EcoRI-消化的总 DNA (Mettler, I. J. (1987) *Plant Mol Biol Reporter* 5, 346-349) 在 1% Tris-硼酸盐 (TBE) 的琼脂糖凝胶上分离, 转移到尼龙膜上 (Amersham), 用 ^{32}P -标记的随机引物 DNA 序列探测, 所述的随机引物 DNA 序列相应于来自含有 *rps7/12* 质体靶向序列一部分的 pC8 的 0.7 kb BamHI/HindIII DNA 片段。同质的芽于无菌条件下在含有壮观霉素的 MS/IBA 培养基 (McBride, K. E. 等 (1994) *PNAS* 91, 7301-7305) 上生根后转移到温室中。

实施例 8: 育种

经本发明的核酸序列转化得到的植物可以是任意的植物物种, 包括单子叶和双子叶植物; 但本发明的方法中使用的植物优选地选自上文所述在农学上重要的靶作物。可以通过育种使本发明的基因的表达与其它有关产量和品质的重要特性相结合一起引入植物系中。育种方法和技术是本领域已知的。参见 Welsh J. R., *Fundamentals of Plant Genetics and Breeding*, John Wiley & Sons, NY (1981);

Crop Breeding, Wood D. R. (编) American Society of Agronomy Madison, Wisconsin (1983); Mayo O., *The Theory of Plant Breeding*, 第2版, Clarendon Press, Oxford (1987); Singh, D.P., *Breeding for Resistance to Diseases and Insect Pests*, Springer-Verlag, NY (1986); Wricke 和 Weber, *Quantitative Genetics and Selection Plant Breeding*, Walter de Gruyter and Co., Berlin (1986)。

上述已改造到转基因种子和植物中的遗传特性通过有性繁殖和营养生长进行传递, 因而能保持下来, 并在后代植物中传播。一般说来用已开发适于特定目的的已知的农业方法进行所述维持和传播, 如耕种, 播种或收获。可以应用专门的方法, 如液体栽培和温室技术。生长中的作物易受昆虫的侵袭和破坏或感染, 对杂草的竞争也敏感。可通过控制杂草, 植物疾病和昆虫, 线虫和其它不利条件的方法提高收成。这些方法包括机械方法, 如土地的耕种或清除杂草或感染植物, 及应用农业化学制品如, 除草剂, 杀真菌剂, 杀配子剂, 杀线虫药, 生长调节剂, 催熟剂和杀虫剂。

利用本发明转基因植物和种子具有优势的遗传特性可进一步应用于植物育种, 目的是使植物具有改良的性状, 如对害虫, 除草剂或胁迫的耐性, 改善了的营养价值, 提高的产率, 或减少由倒伏或脱粒造成的损失。多种育种步骤以已有详细描述的人为介入为特征, 所述的人为介入如, 选择待杂交品系, 指导亲本株系的授粉或选择合适的子代植物。根据所希望的特性可应用不同的育种方法。相关的技术在本领域中是已知的, 其包括但不限于杂交, 近交, 回交育种, 多系育种 (multiline breeding), 混合变种 (variety blend), 种间杂交, 非整倍体技术等。杂交技术也包括通过机械, 化学或生物学手段使植物不育以获得雄性或雌性不育植物。雄性不育植物用不同系的花粉异花授粉保证雄性不育但雌性可育植物的基因组均一地获得两亲本系的特性。因此, 本发明的转基因种子和植物可用于培育改良的植物系, 如提高传统方法, 如除草剂或杀虫剂处理的效率, 或由于

其经修饰的遗传特性而允许人们省却所述的方法。或者，可以获得具有改良的耐胁迫性的新作物，它们由于具有经优化的遗传“装备”，与不能耐受可类比的不利发育条件的产品相比，可收获具有更优良品质的收获产品。

实施例 9： 种子的生产

在种子的生产中，萌发质量和种子的均一性是重要的产品特性，而由农民播种和收获的种子的萌发质量和均一性并不重要。由于保持某种作物与其它作物或杂草种子相分离，控制种子所带疾病，和生产发芽良好的种子是困难的，有经验的培养、调理和经销纯种子的种子生产者发展了非常详尽和明确的种子生产实践。因此农民普遍采用的方式是购买经鉴定满足特定品质标准的种子，而不使用自己收获的种子。用作种子的繁殖材料通常用经保护剂包被层，包括除草剂，杀虫剂，杀真菌剂，杀菌剂，杀线虫剂，杀螺剂或其混合物处理。常用的保护剂包被层包括，如化合物克菌丹，萎锈灵，福美双（thiram）（TMTD[®]），methalaxyl（Apron[®]），和安定磷（Actellic[®]）。如果需要，这些化合物可与载体，表面活性剂，或在制剂领域常规用来防止由细菌，真菌或动物害虫引起的损害的施用促进辅药一起配制。所述的保护剂包被层可通过用液体制剂浸渍繁殖材料进行，或通过用联合的干/湿制剂包被来进行。也可以使用其它的方法，例如直接对芽或果实进行处理。

实施例 10： 玉米植物分析

用质粒 pNOV1436, pNOV1441, 和 pNOV1313 通过土壤杆菌介导的转化转化的玉米植物给出 100% 抗欧洲玉米螟（*European cornborer*）和草地夜蛾（*fall armyworm*）死亡率。ELISA 数据如下：

事件编号	质粒	启动子	玉米基因型	T0/T1 ELISA (ng/mg)				
				叶	穗丝	壳	髓	皮壳(rind)
3275-2	pNOV1436	MTL	A188	125/299			4465/1913	4351/2611
3277-2	pNOV1436	MTL	A188	218/234	136	798	743/3251	613/3055
3279-1	pNOV1436	MTL	A188	108/398			1566/2505	1457/2514
3309-6	pNOV1436	MTL	A188	168/326			1164/1017	1527/2391
3324-1	pNOV1436	MTL	A188	192	0	203	1068	1437
3330-2	pNOV1436	MTL	A188	262/800	0	542	5565	3366
3331-1	pNOV1436	MTL	A188	236/347			1010	1341
3338-1	pNOV1436	MTL	A188	287/457	13		4578	1795
3357-1	pNOV1436	MTL	A188	349/551	61	780	3968	2022
3360-1	pNOV1436	MTL	A188	300/428	0	392	2026	1764
3717-2	pNOV1441	Mz Ubi	Hi II	2142	374	1719	NS	NS
3723-5	pNOV1441	Mz Ubi	Hi II	2302			13757	7215
3838-1	pNOV1441	Mz Ubi	Hi II	2188			24013	13564
3847-2	pNOV1441	Mz Ubi	Hi II	741	699	3707	NS	NS
3877-1	pNOV1441	Mz Ubi	Hi II	991	436	1349	15105	10904
3720-1	pNOV1441	Mz Ubi	Hi II	1437			3854	2719
3833-3	pNOV1441	Mz Ubi	Hi II	878	166	799		
4013-5	pNOV1441	Mz Ubi	Hi II	944	174	1918		
4029-4	pNOV1441	Mz Ubi	Hi II	1661				
4708-1	pNOV1313	Mz Ubi	Hill	832				
4709-2	pNOV1313	Mz Ubi	Hill	581				
4710-5	pNOV1313	Mz Ubi	Hill	625				
4711-2	pNOV1313	Mz Ubi	Hill	570				
4713-2	pNOV1313	Mz Ubi	Hill	962				
4717-1	pNOV1313	Mz Ubi	Hill	881				

MTL = 玉米金属硫蛋白样 (maize metallothionein-like)

Mz Ubi = 玉米遍在蛋白

实施例 11. 稻植物分析

由质粒 pNOV1305 通过土壤杆菌介导的转化转化的植物给出 100% 抗欧洲玉米螟 (*European cornborer*) 和草地夜蛾 (*fall armyworm*) 死亡率。 ELISA 数据如下:

事件编号	质粒	启动子	T0 ELISA (ng/mg) 叶
639	pNOV1305	MTL	294
640	pNOV1305	MTL	241
643	pNOV1305	MTL	153
650	pNOV1305	MTL	149
847	pNOV1305	MTL	173
871	pNOV1305	MTL	244
872	pNOV1305	MTL	252
886	pNOV1305	MTL	185
888	pNOV1305	MTL	160
893	pNOV1305	MTL	168
1148	pNOV1305	MTL	1816
1149	pNOV1305	MTL	224
1152	pNOV1305	MTL	173
1154	pNOV1305	MTL	142
1163	pNOV1305	MTL	139
1164	pNOV1305	MTL	138
1167	pNOV1305	MTL	284
1168	pNOV1305	MTL	137
1177	pNOV1305	MTL	167
1349	pNOV1305	MTL	164
1350	pNOV1305	MTL	115
1357	pNOV1305	MTL	132
1363	pNOV1305	MTL	119
1497	pNOV1305	MTL	94

MTL = 玉米金属硫蛋白样

实施例 12. 卷心菜植物分析

测试了由质粒 pZU578 (SEQ ID NO:17) 通过土壤杆菌介导的转化转化的卷心菜植物对菜蛾 (*Plutella xylostella*) 的抗性。用 16 只幼虫 (1-3 龄) 感染转基因可和对照植物, 用油漆刷从笼养的菜蛾 (*Plutella*) 培养物 (和卷心菜植物一起) 中转移, 4 片叶子中的每一片上放 4 只。经感染的植物转移到 1x1x1m 笼中进行持续的实验。对照植物包括未转化的卷心菜植物 s (易感染的对照) 和喷洒了可商购的 Bt 杀虫剂 Dipel 的未转化的卷心菜植物 (抗性对照)。评分为 (2 周后): - = 没有损害 (或仅有小孔 = 抗性); + = 在植物上有大孔 (= 易感染); ++ 有许多大孔且植物严重损坏 (= 易感染)。Dipel 植物总是评分为 -, 易感性对照评分总是 ++。转基因和对照植物的昆虫损害率和 ELISA 数据如下。

事件编号	质粒	启动子	损害率	T0 ELISA (ng/mg)
				叶

04-05-01-01	pZU578	Act2	++	0
04-05-01-02	pZU578	Act2	++	0
07-11-01	pZU578	Act2	-	921
10-25-05	pZU578	Act2	++	0
10-39-06	pZU578	Act2	-	270
304-F-07	pZU578	Act2	-	
304-F-11	pZU578	Act2	-	
304-F-15	pZU578	Act2	-	
304-F-16	pZU578	Act2	-	
304-F-38	pZU578	Act2	-	
304-g-07	pZU578	Act2	-	
304-g-08	pZU578	Act2	-	
304-g-12	pZU578	Act2	-	
304-g-21	pZU578	Act2	-	
304-g-24	pZU578	Act2	+	0
304-H-01	pZU578	Act2	-	
304-H-08	pZU578	Act2	-	
304-H-09	pZU578	Act2	-	
304-H-34	pZU578	Act2	-	
304-H-35	pZU578	Act2	-	
391-J-08	pZU578	Act2	-	
394-F-5	pZU578	Act2	-	
394-H-12	pZU578	Act2	-	

Act2 = 拟南芥属肌动蛋白 2

上述描述的实施例只是示例性的。本领域的技术人员可根据本发明的教导得到许多本发明的变化形式。所有这些显然的和可预见的变化形式均包含在本发明的范围内。

序列表

<110> Syngenta Participations AG

<120> 新的自来苏云金芽孢杆菌杀虫晶体蛋白的杀虫毒素

<130> Case S-31282A

<140>

<141>

<150> US 60/227956

<151> 2000-08-25

<160> 17

<170> PatentIn Ver. 2.1

<210> 1

<211> 3579

<212> DNA

<213> 人工序列

<220>

<223> 人工序列说明: 带 Cry1C 尾的 H04

<220>

<221> CDS

<222> (1)..(3579)

<223> 带 Cry1C 尾的 H04

<300>

<303> Appl. Environ. Microbiol.

<304> 62

<305> 5

<306> 1537-1543

<307> 1996

<300>

<310> 5,736,131

<400> 1

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	
gga agt att agg agt cca cat ttg atg gat ata ctt aac agt ata acc	912
Gly Ser Ile Arg Ser Pro His Leu Met Asp Ile Leu Asn Ser Ile Thr	
290 295 300	

atc tat acg gat gct cat aga gga gaa tat tat tgg tca ggg cat caa	960
Ile Tyr Thr Asp Ala His Arg Gly Glu 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 gca act ctt aca aat	1392
Arg Ala Pro Met Phe Ser Trp Ile His Arg Ser Ala Thr Leu Thr Asn	
450 455 460	
aca att gat cca gag aga att aat caa ata cct tta gtg aaa gga ttt	1440
Thr Ile Asp Pro Glu Arg Ile Asn Gln Ile Pro Leu Val Lys Gly Phe	
465 470 475 480	
aga gtt tgg ggg ggc acc tct gtc att aca gga cca gga ttt aca gga	1488
Arg Val Trp Gly Gly Thr Ser Val Ile Thr Gly Pro Gly Phe Thr Gly	
485 490 495	
ggg gat atc ctt cga aga aat acc ttt ggt gat ttt gta tct cta caa	1536
Gly Asp Ile Leu Arg Arg Asn Thr Phe Gly Asp Phe Val Ser Leu Gln	
500 505 510	
gtc aat att aat tca cca att acc caa aga tac cgt tta aga ttt cgt	1584
Val Asn Ile Asn Ser Pro Ile Thr Gln Arg Tyr Arg Leu Arg Phe Arg	
515 520 525	
tac gct tcc agt agg gat gca cga gtt ata gta tta aca gga gcg gca	1632
Tyr Ala Ser Ser Arg Asp Ala Arg Val Ile Val Leu Thr Gly Ala Ala	
530 535 540	
tcc aca gga gtg gga ggc caa gtt agt gta aat atg cct ctt cag aaa	1680

Ser Thr Gly Val Gly Gly Gln Val Ser Val Asn Met Pro Leu Gln Lys	
545 550 555 560	
act atg gaa ata ggg gag aac tta aca tct aga aca ttt aga tat acc	1728
Thr Met Glu Ile Gly Glu Asn Leu Thr Ser Arg Thr Phe Arg Tyr Thr	
565 570 575	
gat ttt agt aat cct ttt tca ttt aga gct aat cca gat ata att ggg	1776
Asp Phe Ser Asn Pro Phe Ser Phe Arg Ala Asn Pro Asp Ile Ile Gly	
580 585 590	
ata agt gaa caa cct cta ttt ggt gca ggt tct att agt agc ggt gaa	1824
Ile Ser Glu Gln Pro Leu Phe Gly Ala Gly Ser Ile Ser Ser Gly Glu	
595 600 605	
ctt tat ata gat aaa att gaa att att cta gca gat gca aca ttt gaa	1872
Leu Tyr Ile Asp Lys Ile Glu Ile Ile Leu Ala Asp Ala Thr Phe Glu	
610 615 620	
gca gaa tct gat tta gaa aga gca caa aag gcg gtg aat gcc ctg ttt	1920
Ala Glu Ser Asp Leu Glu Arg Ala Gln Lys Ala Val Asn Ala Leu Phe	
625 630 635 640	
act tct tcc aat caa atc ggg tta aaa acc gat gtg acg gat tat cat	1968
Thr Ser Ser Asn Gln Ile Gly Leu Lys Thr Asp Val Thr Asp Tyr His	
645 650 655	
att gat caa gta tcc aat tta gtg gat tgt tta tca gat gaa ttt tgt	2016
Ile Asp Gln Val Ser Asn Leu Val Asp Cys Leu Ser Asp Glu Phe Cys	
660 665 670	
ctg gat gaa aag cga gaa ttg tcc gag aaa gtc aaa cat gcg aag cga	2064
Leu Asp Glu Lys Arg Glu Leu Ser Glu Lys Val Lys His Ala Lys Arg	
675 680 685	
ctc agt gat gag cgg aat tta ctt caa gat cca aac ttc aga ggg atc	2112
Leu Ser Asp Glu Arg Asn Leu Leu Gln Asp Pro Asn Phe Arg Gly Ile	
690 695 700	
aat aga caa cca gac cgt ggc tgg aga gga agt aca gat att acc atc	2160
Asn Arg Gln Pro Asp Arg Gly Trp Arg Gly Ser Thr Asp Ile Thr Ile	
705 710 715 720	
caa gga gga gat gac gta ttc aaa gag aat tac gtc aca cta ccg ggt	2208
Gln Gly Gly Asp Asp Val Phe Lys Glu Asn Tyr Val Thr Leu Pro Gly	
725 730 735	
acc gtt gat gag tgc tat cca acg tat tta tat cag aaa ata gat gag	2256
Thr Val Asp Glu Cys Tyr Pro Thr Tyr Leu Tyr Gln Lys Ile Asp Glu	
740 745 750	
tcg aaa tta aaa gct tat acc cgt tat gaa tta aga ggg tat atc gaa	2304
Ser Lys Leu Lys Ala Tyr Thr Arg Tyr Glu Leu Arg Gly Tyr Ile Glu	
755 760 765	
gat agt caa gac tta gaa atc tat ttg atc cgt tac aat gca aaa cac	2352
Asp Ser Gln Asp Leu Glu Ile Tyr Leu Ile Arg Tyr Asn Ala Lys His	
770 775 780	
gaa ata gta aat gtg cca ggc acg ggt tcc tta tgg ccg ctt tca gcc	2400
Glu Ile Val Asn Val Pro Gly Thr Gly Ser Leu Trp Pro Leu Ser Ala	

785	790	795	800	
caa agt cca atc gga aag tgt gga gaa ccg aat cga tgc gcg cca cac				2448
Gln Ser Pro Ile Gly Lys Cys Gly Glu Pro Asn Arg Cys Ala Pro His				
	805	810	815	
ctt gaa tgg aat cct gat cta gat tgt tcc tgc aga gac ggg gaa aaa				2496
Leu Glu Trp Asn Pro Asp Leu Asp Cys Ser Cys Arg Asp Gly Glu Lys				
	820	825	830	
tgt gca cat cat tcc cat cat ttc acc ttg gat att gat gtt gga tgt				2544
Cys Ala His His Ser His His Phe Thr Leu Asp Ile Asp Val Gly Cys				
	835	840	845	
aca gac tta aat gag gac tta ggt gta tgg gtg ata ttc aag att aag				2592
Thr Asp Leu Asn Glu Asp Leu Gly Val Trp Val Ile Phe Lys Ile Lys				
	850	855	860	
acg caa gat ggc cat gca aga cta ggg aat cta gag ttt ctc gaa gag				2640
Thr Gln Asp Gly His Ala Arg Leu Gly Asn Leu Glu Phe Leu Glu Glu				
	865	870	875	880
aaa cca tta tta ggg gaa gca cta gct cgt gtg aaa aga gcg gag aag				2688
Lys Pro Leu Leu Gly Glu Ala Leu Ala Arg Val Lys Arg Ala Glu Lys				
	885	890	895	
aag tgg aga gac aaa cga gag aaa ctg cag ttg gaa aca aat att gtt				2736
Lys Trp Arg Asp Lys Arg Glu Lys Leu Gln Leu Glu Thr Asn Ile Val				
	900	905	910	
tat aaa gag gca aaa gaa tct gta gat gct tta ttt gta aac tct caa				2784
Tyr Lys Glu Ala Lys Glu Ser Val Asp Ala Leu Phe Val Asn Ser Gln				
	915	920	925	
tat gat aga tta caa gtg gat acg aac atc gcg atg att cat gcg gca				2832
Tyr Asp Arg Leu Gln Val Asp Thr Asn Ile Ala Met Ile His Ala Ala				
	930	935	940	
gat aaa cgc gtt cat aga atc cgg gaa gcg tat ctg cca gag ttg tct				2880
Asp Lys Arg Val His Arg Ile Arg Glu Ala Tyr Leu Pro Glu Leu Ser				
	945	950	955	960
gtg att cca ggt gtc aat gcg gcc att ttc gaa gaa tta gag gga cgt				2928
Val Ile Pro Gly Val Asn Ala Ala Ile Phe Glu Glu Leu Glu Gly Arg				
	965	970	975	
att ttt aca gcg tat tcc tta tat gat gcg aga aat gtc att aaa aat				2976
Ile Phe Thr Ala Tyr Ser Leu Tyr Asp Ala Arg Asn Val Ile Lys Asn				
	980	985	990	
ggc gat ttc aat aat ggc tta tta tgc tgg aac gtg aaa ggt cat gta				3024
Gly Asp Phe Asn Asn Gly Leu Leu Cys Trp Asn Val Lys Gly His Val				
	995	1000	1005	
gat gta gaa gag caa aac aac cac cgt tcg gtc ctt gtt atc cca gaa				3072
Asp Val Glu Glu Gln Asn Asn His Arg Ser Val Leu Val Ile Pro Glu				
	1010	1015	1020	
tgg gag gca gaa gtg tca caa gag gtt cgt gtc tgt cca ggt cgt ggc				3120
Trp Glu Ala Glu Val Ser Gln Glu Val Arg Val Cys Pro Gly Arg Gly				
	1025	1030	1035	1040

```

tat atc ctt cgt gtc aca gca tat aaa gag gga tat gga gag ggc tgc 3168
Tyr Ile Leu Arg Val Thr Ala Tyr Lys Glu Gly Tyr Gly Glu Gly Cys
      1045      1050      1055

gta acg atc cat gag atc gaa gac aat aca gac gaa ctg aaa ttc agc 3216
Val Thr Ile His Glu Ile Glu Asp Asn Thr Asp Glu Leu Lys Phe Ser
      1060      1065      1070

aac tgt gta gaa gag gaa gta tat cca aac aac aca gta acg tgt aat 3264
Asn Cys Val Glu Glu Glu Val Tyr Pro Asn Asn Thr Val Thr Cys Asn
      1075      1080      1085

aat tat act ggg act caa gaa gaa tat gag ggt acg tac act tct cgt 3312
Asn Tyr Thr Gly Thr Gln Glu Glu Tyr Glu Gly Thr Tyr Thr Ser Arg
      1090      1095      1100

aat caa gga tat gac gaa gcc tat ggt aat aac cct tcc gta cca gct 3360
Asn Gln Gly Tyr Asp Glu Ala Tyr Gly Asn Asn Pro Ser Val Pro Ala
1105      1110      1115      1120

gat tac gct tca gtc tat gaa gaa aaa tcg tat aca gat gga cga aga 3408
Asp Tyr Ala Ser Val Tyr Glu Glu Lys Ser Tyr Thr Asp Gly Arg Arg
      1125      1130      1135

gag aat cct tgt gaa tct aac aga ggc tat ggg gat tac aca cca cta 3456
Glu Asn Pro Cys Glu Ser Asn Arg Gly Tyr Gly Asp Tyr Thr Pro Leu
      1140      1145      1150

ccg gct ggt tat gta aca aag gat tta gag tac ttc cca gag acc gat 3504
Pro Ala Gly Tyr Val Thr Lys Asp Leu Glu Tyr Phe Pro Glu Thr Asp
      1155      1160      1165

aag gta tgg att gag atc gga gaa aca gaa gga aca ttc atc gtg gat 3552
Lys Val Trp Ile Glu Ile Gly Glu Thr Glu Gly Thr Phe Ile Val Asp
      1170      1175      1180

agc gtg gaa tta ctc ctt atg gag gaa 3579
Ser Val Glu Leu Leu Met Glu Glu
1185      1190

```

<210> 2

<211> 1193

<212> PRT

<213> 人工序列

<223> 人工序列说明：带 Cry1C 尾的 H04

<400> 2

```

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 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
 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
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
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
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

<210> 3
<211> 1896
<212> DNA
<213> 人工序列

<223> 人工序列说明：编码不带尾的 H04 毒素部分的合成基因

```
<220>  
<221> CDS  
<222> (1)..(1896)  
<223> 不带尾的H04毒素部分
```

71

Glu	Phe	Val	Pro	Gly	Ala	Gly	Phe	Val	Leu	Gly	Leu	Val	Asp	Ile	Ile		
50						55					60						
tgg	ggc	atc	ttc	ggc	ccc	agc	cag	tgg	gac	gcc	ttc	ctg	gtg	cag	atc	240	
Trp	Gly	Ile	Phe	Gly	Pro	Ser	Gln	Trp	Asp	Ala	Phe	Leu	Val	Gln	Ile		
65					70					75					80		
gag	cag	ttg	ata	aac	caa	cgc	ata	gag	gaa	ttc	gcc	cgc	aac	cag	gcc	288	
Glu	Gln	Leu	Ile	Asn	Gln	Arg	Ile	Glu	Glu	Phe	Ala	Arg	Asn	Gln	Ala		
				85					90					95			
atc	agc	cgc	ctg	gag	ggc	ctg	agc	aac	ctg	tac	caa	atc	tac	gcc	gag	336	
Ile	Ser	Arg	Leu	Glu	Gly	Leu	Ser	Asn	Leu	Tyr	Gln	Ile	Tyr	Ala	Glu		
			100					105					110				
agc	ttc	cgc	gag	tgg	gag	gcc	gac	ccc	acc	aac	ccc	gcc	ctg	cgc	gag	384	
Ser	Phe	Arg	Glu	Trp	Glu	Ala	Asp	Pro	Thr	Asn	Pro	Ala	Leu	Arg	Glu		
			115				120					125					
gag	atg	cgc	atc	cag	ttc	aac	gac	atg	aac	agc	gcc	ctg	acc	acc	gcc	432	
Glu	Met	Arg	Ile	Gln	Phe	Asn	Asp	Met	Asn	Ser	Ala	Leu	Thr	Thr	Ala		
			130			135					140						
atc	ccc	ctg	ttc	gcc	gtg	cag	aac	tac	cag	gtg	ccc	ctg	ctg	agc	gtg	480	
Ile	Pro	Leu	Phe	Ala	Val	Gln	Asn	Tyr	Gln	Val	Pro	Leu	Leu	Ser	Val		
145					150					155					160		
tac	gtg	cag	gcc	gcc	aac	ctg	cac	ctg	agc	gtg	ctg	cgc	gac	gtc	agc	528	
Tyr	Val	Gln	Ala	Ala	Asn	Leu	His	Leu	Ser	Val	Leu	Arg	Asp	Val	Ser		
				165				170						175			
gtg	ttc	ggc	cag	cgc	tgg	ggc	ttc	gac	gcc	gcc	acc	atc	aac	agc	cgc	576	
Val	Phe	Gly	Gln	Arg	Trp	Gly	Phe	Asp	Ala	Ala	Thr	Ile	Asn	Ser	Arg		
			180					185					190				
tac	aac	gac	ctg	acc	cgc	ctg	atc	ggc	aac	tac	acc	gac	cac	gcc	gtg	624	
Tyr	Asn	Asp	Leu	Thr	Arg	Leu	Ile	Gly	Asn	Tyr	Thr	Asp	His	Ala	Val		
			195				200					205					
cgc	tgg	tac	aac	acc	ggc	ctg	gag	cgc	gtg	tgg	ggc	ccc	gac	agc	cgc	672	
Arg	Trp	Tyr	Asn	Thr	Gly	Leu	Glu	Arg	Val	Trp	Gly	Pro	Asp	Ser	Arg		
			210			215					220						
gac	tgg	atc	agg	tac	aac	cag	ttc	cgc	cgc	gag	ctg	acc	ctg	acc	gtg	720	
Asp	Trp	Ile	Arg	Tyr	Asn	Gln	Phe	Arg	Arg	Glu	Leu	Thr	Leu	Thr	Val		
					230					235					240		
ctg	gac	atc	gtg	agc	ctg	ttc	ccc	aac	tac	gac	agc	cgc	acc	tac	ccc	768	
Leu	Asp	Ile	Val	Ser	Leu	Phe	Pro	Asn	Tyr	Asp	Ser	Arg	Thr	Tyr	Pro		
				245					250					255			
atc	cgc	acc	gtg	agc	cag	ctg	acc	cgc	gag	att	tac	acc	aac	ccc	gtg	816	
Ile	Arg	Thr	Val	Ser	Gln	Leu	Thr	Arg	Glu	Ile	Tyr	Thr	Asn	Pro	Val		
			260					265					270				
ctg	gag	aac	ttc	gac	ggc	agc	ttc	cgc	ggc	agc	gcc	cag	ggc	atc	gag	864	
Leu	Glu	Asn	Phe	Asp	Gly	Ser	Phe	Arg	Gly	Ser	Ala	Gln	Gly	Ile	Glu		
			275				280					285					
ggc	agc	atc	cgc	agc	ccc	cac	ctg	atg	gac	atc	ctg	aac	agc	atc	acc	912	
Gly	Ser	Ile	Arg	Ser	Pro	His	Leu	Met	Asp	Ile	Leu	Asn	Ser	Ile	Thr		

290	295	300	
atc tac acc gac gcc cac cgc ggc gag tac tac tgg agc ggc cac cag Ile Tyr Thr Asp Ala His Arg Gly Glu Tyr Tyr Trp Ser Gly His Gln 305 310 315 320			960
atc atg gcc agc ccc gtc ggc ttc agc ggc ccc gag ttc acc ttc ccc Ile Met Ala Ser Pro Val Gly Phe Ser Gly Pro Glu Phe Thr Phe Pro 325 330 335			1008
ctg tac ggc acc atg ggc aac gct gca cct cag cag cgc atc gtg gca Leu Tyr Gly Thr Met Gly Asn Ala Ala Pro Gln Gln Arg Ile Val Ala 340 345 350			1056
cag ctg ggc cag gga gtg tac cgc acc ctg agc agc acc ctg tac cgt Gln Leu Gly Gln Gly Val Tyr Arg Thr Leu Ser Ser Thr Leu Tyr Arg 355 360 365			1104
cga cct ttc aac atc ggc atc aac aac cag cag ctg agc gtg ctg gac Arg Pro Phe Asn Ile Gly Ile Asn Asn Gln Gln Leu Ser Val Leu Asp 370 375 380			1152
ggc acc gag ttc gcc tac ggc acc agc agc aac ctg ccc agc gcc gtg Gly Thr Glu Phe Ala Tyr Gly Thr Ser Ser Asn Leu Pro Ser Ala Val 385 390 395 400			1200
tac cgc aag agc ggc acc gtg gac agc ctg gac gag atc ccc cct cag Tyr Arg Lys Ser Gly Thr Val Asp Ser Leu Asp Glu Ile Pro Pro Gln 405 410 415			1248
aac aac aac gtg cca cct cga cag ggc ttc agc cac cgt ctg agc cac Asn Asn Asn Val Pro Pro Arg Gln Gly Phe Ser His Arg Leu Ser His 420 425 430			1296
gtg agc atg ttc cgc agt ggc ttc agc aac agc agc gtg agc atc atc Val Ser Met Phe Arg Ser Gly Phe Ser Asn Ser Ser Val Ser Ile Ile 435 440 445			1344
cgt gca ccc atg ttc agc tgg att cac cgc agc gcc acc ctg acc aac Arg Ala Pro Met Phe Ser Trp Ile His Arg Ser Ala Thr Leu Thr Asn 450 455 460			1392
acc atc gac ccc gag cgc atc aac cag atc ccc ctg gtg aag ggc ttc Thr Ile Asp Pro Glu Arg Ile Asn Gln Ile Pro Leu Val Lys Gly Phe 465 470 475 480			1440
cgg gtg tgg ggc ggc acc agc gtg atc acc ggc ccc ggc ttc acc gga Arg Val Trp Gly Gly Thr Ser Val Ile Thr Gly Pro Gly Phe Thr Gly 485 490 495			1488
ggc gac atc ctg cgc aga aac acc ttc ggc gac ttc gtg agc ctg cag Gly Asp Ile Leu Arg Arg Asn Thr Phe Gly Asp Phe Val Ser Leu Gln 500 505 510			1536
gtg aac atc aac agc ccc atc acc cag cgt tac cgc ctg cgc ttc cgc Val Asn Ile Asn Ser Pro Ile Thr Gln Arg Tyr Arg Leu Arg Phe Arg 515 520 525			1584
tac gcc agc agc cgc gac gcc cgt gtg atc gtg ctg act ggc gcc gct Tyr Ala Ser Ser Arg Asp Ala Arg Val Ile Val Leu Thr Gly Ala Ala 530 535 540			1632

```

agc acc ggt gtg ggc ggt cag gtg agc gtg aac atg ccc ctg cag aag 1680
Ser Thr Gly Val Gly Gly Gln Val Ser Val Asn Met Pro Leu Gln Lys
545                    550                    555                    560

```

```

act atg gag atc ggc gag aac ctg act agt cgc acc ttc cgc tac acc 1728
Thr Met Glu Ile Gly Glu Asn Leu Thr Ser Arg Thr Phe Arg Tyr Thr
                    565                    570                    575

```

```

gac ttc agc aac ccc ttc agc ttc cgc gcc aac ccc gac atc atc ggc 1776
Asp Phe Ser Asn Pro Phe Ser Phe Arg Ala Asn Pro Asp Ile Ile Gly
                    580                    585                    590

```

```

atc agc gag cag ccc ctg ttc ggt gcc ggc agc atc agc agc ggc gag 1824
Ile Ser Glu Gln Pro Leu Phe Gly Ala Gly Ser Ile Ser Ser Gly Glu
                    595                    600                    605

```

```

ctg tac atc gac aag atc gag atc atc ctg gcc gac gcc acc ttc gag 1872
Leu Tyr Ile Asp Lys Ile Glu Ile Ile Leu Ala Asp Ala Thr Phe Glu
        610                    615                    620

```

```

gcc gag agc gac ctg gag cgc taa 1896
Ala Glu Ser Asp Leu Glu Arg
625                    630

```

<210> 4

<211> 631

<212> PRT

<213> 人工序列

<223> 人工序列说明: 编码不带尾的 H04 毒素部分的合成基因

<400> 4

```

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 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
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
625      630

```

<210> 5

<211> 3582

<212> DNA

<213> 人工序列

<220>

<223> 人工序列说明: 编码带全长 Cry1Ab 尾的 H04 的合成基因

<220>

<221> CDS

<222> (1)..(3582)

<223> 带全长 Cry1Ab 尾的 H04

<400> 5

atg gac aac aac ccc aac atc aac gag tgc atc ccc tac aac tgc ctg	48
Met Asp Asn Asn Pro Asn Ile Asn Glu Cys Ile Pro Tyr Asn Cys Leu	
1 5 10 15	

agc aac ccc gag gtg gag gtg ctg ggc ggc gag cgc atc gag acc ggc	96
Ser Asn Pro Glu Val Glu Val Leu Gly Gly Glu Arg Ile Glu Thr Gly	
20 25 30	

tac acc ccc atc gac atc agc ctg agc ctg acc cag ttc ctg ctg agc	144
Tyr Thr Pro Ile Asp Ile Ser Leu Ser Leu Thr Gln Phe Leu Leu Ser	
35 40 45	

gag ttc gtg ccc ggc gcc ggc ttc gtg ctg ggc ctg gtg gac atc atc	192
Glu Phe Val Pro Gly Ala Gly Phe Val Leu Gly Leu Val Asp Ile Ile	
50 55 60	

tgg ggc atc ttc ggc ccc agc cag tgg gac gcc ttc ctg gtg cag atc	240
Trp Gly Ile Phe Gly Pro Ser Gln Trp Asp Ala Phe Leu Val Gln Ile	
65 70 75 80	

gag cag ttg ata aac caa cgc ata gag gaa ttc gcc cgc aac cag gcc	288
Glu Gln Leu Ile Asn Gln Arg Ile Glu Glu Phe Ala Arg Asn Gln Ala	
85 90 95	

atc agc cgc ctg gag ggc ctg agc aac ctg tac caa atc tac gcc gag	336
Ile Ser Arg Leu Glu Gly Leu Ser Asn Leu Tyr Gln Ile Tyr Ala Glu	
100 105 110	

agc ttc cgc gag tgg gag gcc gac ccc acc aac ccc gcc ctg cgc gag	384
Ser Phe Arg Glu Trp Glu Ala Asp Pro Thr Asn Pro Ala Leu Arg Glu	
115 120 125	

gag atg cgc atc cag ttc aac gac atg aac agc gcc ctg acc acc gcc	432
Glu Met Arg Ile Gln Phe Asn Asp Met Asn Ser Ala Leu Thr Thr Ala	
130 135 140	

atc ccc ctg ttc gcc gtg cag aac tac cag gtg ccc ctg ctg agc gtg	480
Ile Pro Leu Phe Ala Val Gln Asn Tyr Gln Val Pro Leu Leu Ser Val	
145 150 155 160	

tac gtg cag gcc gcc aac ctg cac ctg agc gtg ctg cgc gac gtc agc	528
Tyr Val Gln Ala Ala Asn Leu His Leu Ser Val Leu Arg Asp Val Ser	
165 170 175	

gtg ttc ggc cag cgc tgg ggc ttc gac gcc gcc acc atc aac agc cgc	576
Val Phe Gly Gln Arg Trp Gly Phe Asp Ala Ala Thr Ile Asn Ser Arg	
180 185 190	

tac aac gac ctg acc cgc ctg atc ggc aac tac acc gac cac gcc gtg	624
Tyr Asn Asp Leu Thr Arg Leu Ile Gly Asn Tyr Thr Asp His Ala Val	
195 200 205	

cgc tgg tac aac acc ggc ctg gag cgc gtg tgg ggt ccc gac agc cgc Arg Trp Tyr Asn Thr Gly Leu Glu Arg Val Trp Gly Pro Asp Ser Arg 210 215 220	672
gac tgg atc agg tac aac cag ttc cgc cgc gag ctg acc ctg acc gtg Asp Trp Ile Arg Tyr Asn Gln Phe Arg Arg Glu Leu Thr Leu Thr Val 225 230 235 240	720
ctg gac atc gtg agc ctg ttc ccc aac tac gac agc cgc acc tac ccc Leu Asp Ile Val Ser Leu Phe Pro Asn Tyr Asp Ser Arg Thr Tyr Pro 245 250 255	768
atc cgc acc gtg agc cag ctg acc cgc gag att tac acc aac ccc gtg Ile Arg Thr Val Ser Gln Leu Thr Arg Glu Ile Tyr Thr Asn Pro Val 260 265 270	816
ctg gag aac ttc gac ggc agc ttc cgc ggc agc gcc cag ggc atc gag Leu Glu Asn Phe Asp Gly Ser Phe Arg Gly Ser Ala Gln Gly Ile Glu 275 280 285	864
ggc agc atc cgc agc ccc cac ctg atg gac atc ctg aac agc atc acc Gly Ser Ile Arg Ser Pro His Leu Met Asp Ile Leu Asn Ser Ile Thr 290 295 300	912
atc tac acc gac gcc cac cgc ggc gag tac tac tgg agc ggc cac cag Ile Tyr Thr Asp Ala His Arg Gly Glu Tyr Tyr Trp Ser Gly His Gln 305 310 315 320	960
atc atg gcc agc ccc gtc ggc ttc agc ggc ccc gag ttc acc ttc ccc Ile Met Ala Ser Pro Val Gly Phe Ser Gly Pro Glu Phe Thr Phe Pro 325 330 335	1008
ctg tac ggc acc atg ggc aac gct gca cct cag cag cgc atc gtg gca Leu Tyr Gly Thr Met Gly Asn Ala Ala Pro Gln Gln Arg Ile Val Ala 340 345 350	1056
cag ctg ggc cag gga gtg tac cgc acc ctg agc agc acc ctg tac cgt Gln Leu Gly Gln Gly Val Tyr Arg Thr Leu Ser Ser Thr Leu Tyr Arg 355 360 365	1104
cga cct ttc aac atc ggc atc aac aac cag cag ctg agc gtg ctg gac Arg Pro Phe Asn Ile Gly Ile Asn Asn Gln Gln Leu Ser Val Leu Asp 370 375 380	1152
ggc acc gag ttc gcc tac ggc acc agc agc aac ctg ccc agc gcc gtg Gly Thr Glu Phe Ala Tyr Gly Thr Ser Ser Asn Leu Pro Ser Ala Val 385 390 395 400	1200
tac cgc aag agc ggc acc gtg gac agc ctg gac gag atc ccc cct cag Tyr Arg Lys Ser Gly Thr Val Asp Ser Leu Asp Glu Ile Pro Pro Gln 405 410 415	1248
aac aac aac gtg cca cct cga cag ggc ttc agc cac cgt ctg agc cac Asn Asn Asn Val Pro Pro Arg Gln Gly Phe Ser His Arg Leu Ser His 420 425 430	1296
gtg agc atg ttc cgc agt ggc ttc agc aac agc agc gtg agc atc atc Val Ser Met Phe Arg Ser Gly Phe Ser Asn Ser Ser Val Ser Ile Ile 435 440 445	1344

cgt gca ccc atg ttc agc tgg att cac cgc agc gcc acc ctg acc aac Arg Ala Pro Met Phe Ser Trp Ile His Arg Ser Ala Thr Leu Thr Asn 450 455 460	1392
acc atc gac ccc gag cgc atc aac cag atc ccc ctg gtg aag ggc ttc Thr Ile Asp Pro Glu Arg Ile Asn Gln Ile Pro Leu Val Lys Gly Phe 465 470 475 480	1440
cgg gtg tgg ggc ggc acc agc gtg atc acc ggc ccc ggc ttc acc gga Arg Val Trp Gly Gly Thr Ser Val Ile Thr Gly Pro Gly Phe Thr Gly 485 490 495	1488
ggc gac atc ctg cgc aga aac acc ttc ggc gac ttc gtg agc ctg cag Gly Asp Ile Leu Arg Arg Asn Thr Phe Gly Asp Phe Val Ser Leu Gln 500 505 510	1536
gtg aac atc aac agc ccc atc acc cag cgt tac cgc ctg cgc ttc cgc Val Asn Ile Asn Ser Pro Ile Thr Gln Arg Tyr Arg Leu Arg Phe Arg 515 520 525	1584
tac gcc agc agc cgc gac gcc cgt gtg atc gtg ctg act ggc gcc gct Tyr Ala Ser Ser Arg Asp Ala Arg Val Ile Val Leu Thr Gly Ala Ala 530 535 540	1632
agc acc ggt gtg ggc ggt cag gtg agc gtg aac atg ccc ctg cag aag Ser Thr Gly Val Gly Gly Gln Val Ser Val Asn Met Pro Leu Gln Lys 545 550 555 560	1680
act atg gag atc ggc gag aac ctg act agt cgc acc ttc cgc tac acc Thr Met Glu Ile Gly Glu Asn Leu Thr Ser Arg Thr Phe Arg Tyr Thr 565 570 575	1728
gac ttc agc aac ccc ttc agc ttc cgc gcc aac ccc gac atc atc ggc Asp Phe Ser Asn Pro Phe Ser Phe Arg Ala Asn Pro Asp Ile Ile Gly 580 585 590	1776
atc agc gag cag ccc ctg ttc ggt gcc ggc agc atc agc agc ggc gag Ile Ser Glu Gln Pro Leu Phe Gly Ala Gly Ser Ile Ser Ser Gly Glu 595 600 605	1824
ctg tac atc gac aag atc gag atc atc ctg gcc gac gcc acc ttc gag Leu Tyr Ile Asp Lys Ile Glu Ile Ile Leu Ala Asp Ala Thr Phe Glu 610 615 620	1872
gcc gag agc gac ctg gag cgc gcc cag aag gcc gtg aac gcc ctg ttc Ala Glu Ser Asp Leu Glu Arg Ala Gln Lys Ala Val Asn Ala Leu Phe 625 630 635 640	1920
acc agc agc aac cag atc ggc ctg aag acc gac gtg acc gac tac cac Thr Ser Ser Asn Gln Ile Gly Leu Lys Thr Asp Val Thr Asp Tyr His 645 650 655	1968
atc gac cag gtg agc aac ctg gtg gac tgc tta agc gac gag ttc tgc Ile Asp Gln Val Ser Asn Leu Val Asp Cys Leu Ser Asp Glu Phe Cys 660 665 670	2016
ctg gac gag aag aag gag ctg agc gag aag gtg aag cac gcc aag cgc Leu Asp Glu Lys Lys Glu Leu Ser Glu Lys Val Lys His Ala Lys Arg 675 680 685	2064
ctg agc gac gag cgc aac ctg ctg cag gac ccc aac ttc cgc ggc atc	2112

Leu 690	Ser	Asp	Glu	Arg	Asn	Leu 695	Leu	Gln	Asp	Pro	Asn 700	Phe	Arg	Gly	Ile	
aac 705	cgc	cag	ctg	gac	cgc	ggc	tgg	cga	ggc	agc	acc	gat	atc	acc	atc	2160
Asn	Arg	Gln	Leu	Asp	Arg	Gly 710	Trp	Arg	Gly	Ser 715	Thr	Asp	Ile	Thr	Ile 720	
cag	ggc	ggc	gac	gac	gtg	ttc	aag	gag	aac	tac	gtg	acc	ctg	cag	ggc	2208
Gln	Gly	Gly	Asp	Asp	Val	Phe	Lys	Glu	Asn 730	Tyr	Val	Thr	Leu	Gln	Gly 735	
acc	ttc	gac	gag	tgc	tac	ccc	acc	tac	ctg	tac	cag	ccg	atc	gac	gag	2256
Thr	Phe	Asp	Glu	Cys	Tyr	Pro	Thr	Tyr 745	Leu	Tyr	Gln	Pro	Ile	Asp	Glu 750	
agc	aag	ctg	aag	gcc	tac	acc	cgc	tac	cag	ctg	cgc	ggc	tac	atc	gag	2304
Ser	Lys	Leu	Lys	Ala	Tyr	Thr 755	Arg	Tyr	Gln	Leu	Arg	Gly 765	Tyr	Ile	Glu 760	
gac	agc	cag	gac	ctg	gaa	atc	tac	ctg	atc	cgc	tac	aac	gcg	aag	cac	2352
Asp	Ser	Gln	Asp	Leu	Glu 770	Ile 775	Tyr	Leu	Ile	Arg	Tyr 780	Asn	Ala	Lys	His	
gag	acc	gtg	aac	gtg	ccc	ggc	acc	ggc	agc	ctg	tgg	ccc	ccg	agc	gcc	2400
Glu	Thr	Val	Asn	Val	Pro 785	Gly 790	Thr	Gly	Ser	Leu 795	Trp	Pro	Pro	Ser	Ala 800	
ccc	agc	ccc	atc	ggc	aag	tgc	ggg	gag	ccg	aat	cga	tgc	gct	ccg	cac	2448
Pro	Ser	Pro	Ile	Gly 805	Lys	Cys	Gly	Glu	Pro 810	Asn	Arg	Cys	Ala	Pro 815	His	
ctg	gag	tgg	aac	ccg	gac	cta	gac	tgc	agc	tgc	agg	gac	ggg	gag	aag	2496
Leu	Glu	Trp	Asn 820	Pro	Asp	Leu	Asp	Cys 825	Ser	Cys	Arg	Asp	Gly 830	Glu	Lys	
tgc	gcc	cac	cac	agc	cac	cac	ttc	agc	ctg	gac	atc	gac	gtg	ggc	tgc	2544
Cys	Ala	His 835	His	Ser	His	His 840	Phe	Ser	Leu	Asp	Ile	Asp 845	Val	Gly	Cys	
acc	gac	ctg	aac	gag	gac	ctg	ggc	gtg	tgg	gtg	atc	ttc	aag	atc	aag	2592
Thr	Asp	Leu	Asn	Glu	Asp 850	Leu 855	Gly	Val	Trp	Val 860	Ile	Phe	Lys	Ile	Lys	
acc	cag	gac	ggc	cac	gcc	cgc	ctg	ggc	aat	cta	gag	ttc	ctg	gag	gag	2640
Thr	Gln	Asp	Gly	His	Ala 865	Arg 870	Leu	Gly	Asn	Leu 875	Glu	Phe	Leu	Glu	Glu 880	
aag	ccc	ctg	gtg	ggc	gag	gcc	ctg	gcc	cgc	gtg	aag	cgt	gct	gag	aag	2688
Lys	Pro	Leu	Val	Gly 885	Glu	Ala	Leu	Ala	Arg 890	Val	Lys	Arg	Ala	Glu 895	Lys	
aag	tgg	cgc	gac	aag	cgc	gag	aag	ctg	gag	tgg	gag	acc	aac	atc	gtg	2736
Lys	Trp	Arg	Asp 900	Lys	Arg	Glu	Lys	Leu 905	Glu	Trp	Glu	Thr	Asn 910	Ile	Val	
tac	aag	gag	gcc	aag	gag	agc	gtg	gac	gcc	ctg	ttc	gtg	aac	agc	cag	2784
Tyr	Lys	Glu	Ala 915	Lys	Glu	Ser	Val 920	Asp	Ala	Leu	Phe	Val 925	Asn	Ser	Gln	
tac	gac	cgc	ctg	cag	gcc	gac	acc	aac	atc	gcc	atg	atc	cac	gcc	gcc	2832
Tyr	Asp	Arg	Leu	Gln	Ala	Asp	Thr	Asn	Ile	Ala	Met	Ile	His	Ala	Ala	

930	935	940	
gac aag cgc gtg cac agc att cgc gag gcc tac ctg ccc gag ctg agc Asp Lys Arg Val His Ser Ile Arg Glu Ala Tyr Leu Pro Glu Leu Ser 945 950 955 960			2880
gtg atc ccc ggt gtg aac gcc gcc atc ttc gag gaa ctc gag ggc cgc Val Ile Pro Gly Val Asn Ala Ala Ile Phe Glu Glu Leu Glu Gly Arg 965 970 975			2928
atc ttc acc gcc ttc agc ctg tac gac gcc cgc aac gtg atc aag aac Ile Phe Thr Ala Phe Ser Leu Tyr Asp Ala Arg Asn Val Ile Lys Asn 980 985 990			2976
ggc gac ttc aac aac ggc ctg agc tgc tgg aac gtg aag ggc cac gtg Gly Asp Phe Asn Asn Gly Leu Ser Cys Trp Asn Val Lys Gly His Val 995 1000 1005			3024
gac gtg gag gag cag aac aac cac cgc agc gtg ctg gtg gtg ccc gag Asp Val Glu Glu Gln Asn Asn His Arg Ser Val Leu Val Val Pro Glu 1010 1015 1020			3072
tgg gag gcc gag gtg agc cag gag gtg cgc gtg tgc ccc ggc cgc ggc Trp Glu Ala Glu Val Ser Gln Glu Val Arg Val Cys Pro Gly Arg Gly 1025 1030 1035 1040			3120
tac atc ctg cgc gtg acc gcc tac aag gag ggc tac ggc gag ggc tgc Tyr Ile Leu Arg Val Thr Ala Tyr Lys Glu Gly Tyr Gly Glu Gly Cys 1045 1050 1055			3168
gtg acc atc cac gag atc gag aac aac acc gac gag ctc aag ttc agc Val Thr Ile His Glu Ile Glu Asn Asn Thr Asp Glu Leu Lys Phe Ser 1060 1065 1070			3216
aac tgc gtg gag gag gag gtt tac ccc aac aac acc gtg acc tgc aac Asn Cys Val Glu Glu Glu Val Tyr Pro Asn Asn Thr Val Thr Cys Asn 1075 1080 1085			3264
gac tac acc gcg acc cag gag gag tac gaa ggc acc tac acc tct cgc Asp Tyr Thr Ala Thr Gln Glu Glu Tyr Glu Gly Thr Tyr Thr Ser Arg 1090 1095 1100			3312
aac agg ggt tac gac ggc gcc tac gag tcc aac agc tcc gtg cca gct Asn Arg Gly Tyr Asp Gly Ala Tyr Glu Ser Asn Ser Ser Val Pro Ala 1105 1110 1115 1120			3360
gac tac gcc agc gcc cac gag gag aaa gcc tac acc gac ggt aga cgc Asp Tyr Ala Ser Ala His Glu Glu Lys Ala Tyr Thr Asp Gly Arg Arg 1125 1130 1135			3408
gac aac cca tgt gag agc aac aga ggc tac ggc gac tac acc ccc ctg Asp Asn Pro Cys Glu Ser Asn Arg Gly Tyr Gly Asp Tyr Thr Pro Leu 1140 1145 1150			3456
ccc gct gga tac gtg acc aag gag ctg gag tac ttc ccc gag acc gac Pro Ala Gly Tyr Val Thr Lys Glu Leu Glu Tyr Phe Pro Glu Thr Asp 1155 1160 1165			3504
aag gtg tgg atc gag att ggc gag acc gag ggc acc ttc atc gtg gac Lys Val Trp Ile Glu Ile Gly Glu Thr Glu Gly Thr Phe Ile Val Asp 1170 1175 1180			3552

agc gtg gag ctg ctg ctg atg gag gag tag
 Ser Val Glu Leu Leu Leu Met Glu Glu
 1185 1190

3582

<210> 6

<211> 1193

<212> PRT

<213> 人工序列

<223> 人工序列说明: 编码带全长 Cry1Ab 尾的 H04 的合成基因

<400> 6

```

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	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	
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	Lys	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	Leu	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	Gln	Gly
			725						730					735	
Thr	Phe	Asp	Glu	Cys	Tyr	Pro	Thr	Tyr	Leu	Tyr	Gln	Pro	Ile	Asp	Glu
		740						745					750		
Ser	Lys	Leu	Lys	Ala	Tyr	Thr	Arg	Tyr	Gln	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	Thr	Val	Asn	Val	Pro	Gly	Thr	Gly	Ser	Leu	Trp	Pro	Pro	Ser	Ala
785					790					795					800
Pro	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	Ser	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 Val Gly Glu Ala Leu Ala Arg Val Lys Arg Ala Glu Lys
                      885                      890                      895
Lys Trp Arg Asp Lys Arg Glu Lys Leu Glu Trp 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 Ala Asp Thr Asn Ile Ala Met Ile His Ala Ala
                      930                      935                      940
Asp Lys Arg Val His Ser 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 Phe Ser Leu Tyr Asp Ala Arg Asn Val Ile Lys Asn
                      980                      985                      990
Gly Asp Phe Asn Asn Gly Leu Ser Cys Trp Asn Val Lys Gly His Val
                      995                      1000                      1005
Asp Val Glu Glu Gln Asn Asn His Arg Ser Val Leu Val Val 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 Asn 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
Asp Tyr Thr Ala Thr Gln Glu Glu Tyr Glu Gly Thr Tyr Thr Ser Arg
1090                      1095                      1100
Asn Arg Gly Tyr Asp Gly Ala Tyr Glu Ser Asn Ser Ser Val Pro Ala
1105                      1110                      1115                      1120
Asp Tyr Ala Ser Ala His Glu Glu Lys Ala Tyr Thr Asp Gly Arg Arg
                      1125                      1130                      1135
Asp 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 Glu 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
Ser Val Glu Leu Leu Leu Met Glu Glu
1185                      1190

```

<210> 7

<211> 3582

<212> DNA

<213> 人工序列

<220>

<223> 人工序列说明: 编码带全长 Cry1Ab 尾的 H04 的合成基因

<220>

<221> CDS

<222> (1) .. (3582)

<223> 带全长 Cry1Ab 尾的 H04

<400> 7

atg gac aac aac ccc aac atc aac gag tgc atc ccc tac aac tgc ctg 48

Met	Asp	Asn	Asn	Pro	Asn	Ile	Asn	Glu	Cys	Ile	Pro	Tyr	Asn	Cys	Leu	
1				5					10					15		
agc	aac	ccc	gag	gtg	gag	gtg	ctg	ggc	ggc	gag	cgc	atc	gag	acc	ggc	96
Ser	Asn	Pro	Glu	Val	Glu	Val	Leu	Gly	Gly	Glu	Arg	Ile	Glu	Thr	Gly	
			20					25					30			
tac	acc	ccc	atc	gac	atc	agc	ctg	agc	ctg	acc	cag	ttc	ctg	ctg	agc	144
Tyr	Thr	Pro	Ile	Asp	Ile	Ser	Leu	Ser	Leu	Thr	Gln	Phe	Leu	Leu	Ser	
			35				40					45				
gag	ttc	gtg	ccc	ggc	gcc	ggc	ttc	gtg	ctg	ggc	ctg	gtg	gac	atc	atc	192
Glu	Phe	Val	Pro	Gly	Ala	Gly	Phe	Val	Leu	Gly	Leu	Val	Asp	Ile	Ile	
	50					55				60						
tgg	ggc	atc	ttc	ggc	ccc	agc	cag	tgg	gac	gcc	ttc	ctg	gtg	cag	atc	240
Trp	Gly	Ile	Phe	Gly	Pro	Ser	Gln	Trp	Asp	Ala	Phe	Leu	Val	Gln	Ile	
65					70				75					80		
gag	cag	ttg	ata	aac	caa	cgc	ata	gag	gaa	ttc	gcc	cgc	aac	cag	gcc	288
Glu	Gln	Leu	Ile	Asn	Gln	Arg	Ile	Glu	Glu	Phe	Ala	Arg	Asn	Gln	Ala	
				85				90					95			
atc	agc	cgc	ctg	gag	ggc	ctg	agc	aac	ctg	tac	caa	atc	tac	gcc	gag	336
Ile	Ser	Arg	Leu	Glu	Gly	Leu	Ser	Asn	Leu	Tyr	Gln	Ile	Tyr	Ala	Glu	
			100					105				110				
agc	ttc	cgc	gag	tgg	gag	gcc	gac	ccc	acc	aac	ccc	gcc	ctg	cgc	gag	384
Ser	Phe	Arg	Glu	Trp	Glu	Ala	Asp	Pro	Thr	Asn	Pro	Ala	Leu	Arg	Glu	
	115					120						125				
gag	atg	cgc	atc	cag	ttc	aac	gac	atg	aac	agc	gcc	ctg	acc	acc	gcc	432
Glu	Met	Arg	Ile	Gln	Phe	Asn	Asp	Met	Asn	Ser	Ala	Leu	Thr	Thr	Ala	
	130					135					140					
atc	ccc	ctg	ttc	gcc	gtg	cag	aac	tac	cag	gtg	ccc	ctg	ctg	agc	gtg	480
Ile	Pro	Leu	Phe	Ala	Val	Gln	Asn	Tyr	Gln	Val	Pro	Leu	Leu	Ser	Val	
145					150				155						160	
tac	gtg	cag	gcc	gcc	aac	ctg	cac	ctg	agc	gtg	ctg	cgc	gac	gtc	agc	528
Tyr	Val	Gln	Ala	Ala	Asn	Leu	His	Leu	Ser	Val	Leu	Arg	Asp	Val	Ser	
			165					170					175			
gtg	ttc	ggc	cag	cgc	tgg	ggc	ttc	gac	gcc	gcc	acc	atc	aac	agc	cgc	576
Val	Phe	Gly	Gln	Arg	Trp	Gly	Phe	Asp	Ala	Ala	Thr	Ile	Asn	Ser	Arg	
			180				185					190				
tac	aac	gac	ctg	acc	cgc	ctg	atc	ggc	aac	tac	acc	gac	cac	gcc	gtg	624
Tyr	Asn	Asp	Leu	Thr	Arg	Leu	Ile	Gly	Asn	Tyr	Thr	Asp	His	Ala	Val	
	195						200					205				
cgc	tgg	tac	aac	acc	ggc	ctg	gag	cgc	gtg	tgg	ggt	ccc	gac	agc	cgc	672
Arg	Trp	Tyr	Asn	Thr	Gly	Leu	Glu	Arg	Val	Trp	Gly	Pro	Asp	Ser	Arg	
	210					215					220					
gac	tgg	atc	agg	tac	aac	cag	ttc	cgc	cgc	gag	ctg	acc	ctg	acc	gtg	720
Asp	Trp	Ile	Arg	Tyr	Asn	Gln	Phe	Arg	Arg	Glu	Leu	Thr	Leu	Thr	Val	
	225				230				235					240		
ctg	gac	atc	gtg	agc	ctg	ttc	ccc	aac	tac	gac	agc	cgc	acc	tac	ccc	768
Leu	Asp	Ile	Val	Ser	Leu	Phe	Pro	Asn	Tyr	Asp	Ser	Arg	Thr	Tyr	Pro	

245										250										255										
atc	cgc	acc	gtg	agc	cag	ctg	acc	cgc	gag	att	tac	acc	aac	ccc	gtg															816
Ile	Arg	Thr	Val	Ser	Gln	Leu	Thr	Arg	Glu	Ile	Tyr	Thr	Asn	Pro	Val															
			260						265					270																
ctg	gag	aac	ttc	gac	ggc	agc	ttc	cgc	ggc	agc	gcc	cag	ggc	atc	gag															864
Leu	Glu	Asn	Phe	Asp	Gly	Ser	Phe	Arg	Gly	Ser	Ala	Gln	Gly	Ile	Glu															
			275					280					285																	
ggc	agc	atc	cgc	agc	ccc	cac	ctg	atg	gac	atc	ctg	aac	agc	atc	acc															912
Gly	Ser	Ile	Arg	Ser	Pro	His	Leu	Met	Asp	Ile	Leu	Asn	Ser	Ile	Thr															
			290				295					300																		
atc	tac	acc	gac	gcc	cac	cgc	ggc	gag	tac	tac	tgg	agc	ggc	cac	cag															960
Ile	Tyr	Thr	Asp	Ala	His	Arg	Gly	Glu	Tyr	Tyr	Trp	Ser	Gly	His	Gln															
			305			310					315				320															
atc	atg	gcc	agc	ccc	gtc	ggc	ttc	agc	ggc	ccc	gag	ttc	acc	ttc	ccc															1008
Ile	Met	Ala	Ser	Pro	Val	Gly	Phe	Ser	Gly	Pro	Glu	Phe	Thr	Phe	Pro															
				325					330					335																
ctg	tac	ggc	acc	atg	ggc	aac	gct	gca	cct	cag	cag	cgc	atc	gtg	gca															1056
Leu	Tyr	Gly	Thr	Met	Gly	Asn	Ala	Ala	Pro	Gln	Gln	Arg	Ile	Val	Ala															
			340					345					350																	
cag	ctg	ggc	cag	gga	gtg	tac	cgc	acc	ctg	agc	agc	acc	ctg	tac	cgt															1104
Gln	Leu	Gly	Gln	Gly	Val	Tyr	Arg	Thr	Leu	Ser	Ser	Thr	Leu	Tyr	Arg															
			355				360					365																		
cga	cct	ttc	aac	atc	ggc	atc	aac	aac	cag	cag	ctg	agc	gtg	ctg	gac															1152
Arg	Pro	Phe	Asn	Ile	Gly	Ile	Asn	Asn	Gln	Gln	Leu	Ser	Val	Leu	Asp															
			370			375					380																			
ggc	acc	gag	ttc	gcc	tac	ggc	acc	agc	agc	aac	ctg	ccc	agc	gcc	gtg															1200
Gly	Thr	Glu	Phe	Ala	Tyr	Gly	Thr	Ser	Ser	Asn	Leu	Pro	Ser	Ala	Val															
			385			390				395				400																
tac	cgc	aag	agc	ggc	acc	gtg	gac	agc	ctg	gac	gag	atc	ccc	cct	cag															1248
Tyr	Arg	Lys	Ser	Gly	Thr	Val	Asp	Ser	Leu	Asp	Glu	Ile	Pro	Pro	Gln															
				405					410					415																
aac	aac	aac	gtg	cca	cct	cga	cag	ggc	ttc	agc	cac	cgt	ctg	agc	cac															1296
Asn	Asn	Asn	Val	Pro	Pro	Arg	Gln	Gly	Phe	Ser	His	Arg	Leu	Ser	His															
			420					425					430																	
gtg	agc	atg	ttc	cgc	agt	ggc	ttc	agc	aac	agc	agc	gtg	agc	atc	atc															1344
Val	Ser	Met	Phe	Arg	Ser	Gly	Phe	Ser	Asn	Ser	Ser	Val	Ser	Ile	Ile															
			435				440					445																		
cgt	gca	ccc	atg	ttc	agc	tgg	att	cac	cgc	agc	gcc	acc	ctg	acc	aac															1392
Arg	Ala	Pro	Met	Phe	Ser	Trp	Ile	His	Arg	Ser	Ala	Thr	Leu	Thr	Asn															
			450				455					460																		
acc	atc	gac	ccc	gag	cgc	atc	aac	cag	atc	ccc	ctg	gtg	aag	ggc	ttc															1440
Thr	Ile	Asp	Pro	Glu	Arg	Ile	Asn	Gln	Ile	Pro	Leu	Val	Lys	Gly	Phe															
			465			470				475				480																
cgg	gtg	tgg	ggc	ggc	acc	agc	gtg	atc	acc	ggc	ccc	ggc	ttc	acc	gga															1488
Arg	Val	Trp	Gly	Gly	Thr	Ser	Val	Ile	Thr	Gly	Pro	Gly	Phe	Thr	Gly															
				485				490						495																

ggc gac atc ctg cgc aga aac acc ttc ggc gac ttc gtg agc ctg cag	1536
Gly Asp Ile Leu Arg Arg Asn Thr Phe Gly Asp Phe Val Ser Leu Gln	
500 505 510	
gtg aac atc aac agc ccc atc acc cag cgt tac cgc ctg cgc ttc cgc	1584
Val Asn Ile Asn Ser Pro Ile Thr Gln Arg Tyr Arg Leu Arg Phe Arg	
515 520 525	
tac gcc agc agc cgc gac gcc cgt gtg atc gtg ctg act ggc gcc gct	1632
Tyr Ala Ser Ser Arg Asp Ala Arg Val Ile Val Leu Thr Gly Ala Ala	
530 535 540	
agc acc ggt gtg ggc ggt cag gtg agc gtg aac atg ccc ctg cag aag	1680
Ser Thr Gly Val Gly Gly Gln Val Ser Val Asn Met Pro Leu Gln Lys	
545 550 555 560	
act atg gag atc ggc gag aac ctg act agt cgc acc ttc cgc tac acc	1728
Thr Met Glu Ile Gly Glu Asn Leu Thr Ser Arg Thr Phe Arg Tyr Thr	
565 570 575	
gac ttc agc aac ccc ttc agc ttc cgc gcc aac ccc gac atc atc ggc	1776
Asp Phe Ser Asn Pro Phe Ser Phe Arg Ala Asn Pro Asp Ile Ile Gly	
580 585 590	
atc agc gag cag ccc ctg ttc ggt gcc ggc agc atc agc agc ggc gag	1824
Ile Ser Glu Gln Pro Leu Phe Gly Ala Gly Ser Ile Ser Ser Gly Glu	
595 600 605	
ctg tac atc gac aag atc gag atc atc ctg gcc gac gcc acc ttc gag	1872
Leu Tyr Ile Asp Lys Ile Glu Ile Ile Leu Ala Asp Ala Thr Phe Glu	
610 615 620	
gcc gag agc gac ctg gag cgc gcc cag aag gcc gtg aac gcc ctg ttc	1920
Ala Glu Ser Asp Leu Glu Arg Ala Gln Lys Ala Val Asn Ala Leu Phe	
625 630 635 640	
acc agc agc aac cag atc ggc ctg aag acc gac gtg acc gac tac cac	1968
Thr Ser Ser Asn Gln Ile Gly Leu Lys Thr Asp Val Thr Asp Tyr His	
645 650 655	
atc gac cag gtg agc aac ctg gtg gac tgc tta agc gac gag ttc tgc	2016
Ile Asp Gln Val Ser Asn Leu Val Asp Cys Leu Ser Asp Glu Phe Cys	
660 665 670	
ctg gac gag aag aag gag ctg agc gag aag gtg aag cac gcc aag cgc	2064
Leu Asp Glu Lys Lys Glu Leu Ser Glu Lys Val Lys His Ala Lys Arg	
675 680 685	
ctg agc gac gag cgc aac ctg ctg cag gac ccc aac ttc cgc ggc atc	2112
Leu Ser Asp Glu Arg Asn Leu Leu Gln Asp Pro Asn Phe Arg Gly Ile	
690 695 700	
aac cgc cag ctg gac cgc ggc tgg cga ggc agc acc gat atc acc atc	2160
Asn Arg Gln Leu Asp Arg Gly Trp Arg Gly Ser Thr Asp Ile Thr Ile	
705 710 715 720	
cag ggc ggc gac gac gtg ttc aag gag aac tac gtg acc ctg cag ggc	2208
Gln Gly Gly Asp Asp Val Phe Lys Glu Asn Tyr Val Thr Leu Gln Gly	
725 730 735	

acc ttc gac gag tgc tac ccc acc tac ctg tac cag ccg atc gac gag Thr Phe Asp Glu Cys Tyr Pro Thr Tyr Leu Tyr Gln Pro Ile Asp Glu 740 745 750	2256
agc aag ctg aag gcc tac acc cgc tac cag ctg cgc ggc tac atc gag Ser Lys Leu Lys Ala Tyr Thr Arg Tyr Gln Leu Arg Gly Tyr Ile Glu 755 760 765	2304
gac agc cag gac ctg gaa atc tac ctg atc cgc tac aac gcg aag cac Asp Ser Gln Asp Leu Glu Ile Tyr Leu Ile Arg Tyr Asn Ala Lys His 770 775 780	2352
gag acc gtg aac gtg ccc ggc acc ggc agc ctg tgg ccc ctg agc gcc Glu Thr Val Asn Val Pro Gly Thr Gly Ser Leu Trp Pro Leu Ser Ala 785 790 795 800	2400
ccc agc ccc atc ggc aag tgc ggg gag ccg aat cga tgc gct ccg cac Pro Ser Pro Ile Gly Lys Cys Gly Glu Pro Asn Arg Cys Ala Pro His 805 810 815	2448
ctg gag tgg aac ccg gac cta gac tgc agc tgc agg gac ggg gag aag Leu Glu Trp Asn Pro Asp Leu Asp Cys Ser Cys Arg Asp Gly Glu Lys 820 825 830	2496
tgc gcc cac cac agc cac cac ttc agc ctg gac atc gac gtg ggc tgc Cys Ala His His Ser His His Phe Ser Leu Asp Ile Asp Val Gly Cys 835 840 845	2544
acc gac ctg aac gag gac ctg ggc gtg tgg gtg atc ttc aag atc aag Thr Asp Leu Asn Glu Asp Leu Gly Val Trp Val Ile Phe Lys Ile Lys 850 855 860	2592
acc cag gac ggc cac gcc cgc ctg ggc aat cta gag ttc ctg gag gag Thr Gln Asp Gly His Ala Arg Leu Gly Asn Leu Glu Phe Leu Glu Glu 865 870 875 880	2640
aag ccc ctg gtg ggc gag gcc ctg gcc cgc gtg aag cgt gct gag aag Lys Pro Leu Val Gly Glu Ala Leu Ala Arg Val Lys Arg Ala Glu Lys 885 890 895	2688
aag tgg cgc gac aag cgc gag aag ctg gag tgg gag acc aac atc gtg Lys Trp Arg Asp Lys Arg Glu Lys Leu Glu Trp Glu Thr Asn Ile Val 900 905 910	2736
tac aag gag gcc aag gag agc gtg gac gcc ctg ttc gtg aac agc cag Tyr Lys Glu Ala Lys Glu Ser Val Asp Ala Leu Phe Val Asn Ser Gln 915 920 925	2784
tac gac cgc ctg cag gcc gac acc aac atc gcc atg atc cac gcc gcc Tyr Asp Arg Leu Gln Ala Asp Thr Asn Ile Ala Met Ile His Ala Ala 930 935 940	2832
gac aag cgc gtg cac agc att cgc gag gcc tac ctg ccc gag ctg agc Asp Lys Arg Val His Ser Ile Arg Glu Ala Tyr Leu Pro Glu Leu Ser 945 950 955 960	2880
gtg atc ccc ggt gtg aac gcc gcc atc ttc gag gaa ctc gag ggc cgc Val Ile Pro Gly Val Asn Ala Ala Ile Phe Glu Glu Leu Glu Gly Arg 965 970 975	2928
atc ttc acc gcc ttc agc ctg tac gac gcc cgc aac gtg atc aag aac	2976

```

Ile Phe Thr Ala Phe Ser Leu Tyr Asp Ala Arg Asn Val Ile Lys Asn
      980                      985                      990

ggc gac ttc aac aac ggc ctg agc tgc tgg aac gtg aag ggc cac gtg      3024
Gly Asp Phe Asn Asn Gly Leu Ser Cys Trp Asn Val Lys Gly His Val
      995                      1000                      1005

gac gtg gag gag cag aac aac cac cgc agc gtg ctg gtg gtg ccc gag      3072
Asp Val Glu Glu Gln Asn Asn His Arg Ser Val Leu Val Val Pro Glu
      1010                      1015                      1020

tgg gag gcc gag gtg agc cag gag gtg cgc gtg tgc ccc ggc cgc ggc      3120
Trp Glu Ala Glu Val Ser Gln Glu Val Arg Val Cys Pro Gly Arg Gly
      1025                      1030                      1035                      1040

tac atc ctg cgc gtg acc gcc tac aag gag ggc tac ggc gag ggc tgc      3168
Tyr Ile Leu Arg Val Thr Ala Tyr Lys Glu Gly Tyr Gly Glu Gly Cys
      1045                      1050                      1055

gtg acc atc cac gag atc gag aac aac acc gac gag ctc aag ttc agc      3216
Val Thr Ile His Glu Ile Glu Asn Asn Thr Asp Glu Leu Lys Phe Ser
      1060                      1065                      1070

aac tgc gtg gag gag gag gtt tac ccc aac aac acc gtg acc tgc aac      3264
Asn Cys Val Glu Glu Glu Val Tyr Pro Asn Asn Thr Val Thr Cys Asn
      1075                      1080                      1085

gac tac acc gcg acc cag gag gag tac gaa ggc acc tac acc tct cgc      3312
Asp Tyr Thr Ala Thr Gln Glu Glu Tyr Glu Gly Thr Tyr Thr Ser Arg
      1090                      1095                      1100

aac agg ggt tac gac ggc gcc tac gag tcc aac agc tcc gtg cca gct      3360
Asn Arg Gly Tyr Asp Gly Ala Tyr Glu Ser Asn Ser Ser Val Pro Ala
      1105                      1110                      1115                      1120

gac tac gcc agc gcc tac gag gag aaa gcc tac acc gac ggt aga cgc      3408
Asp Tyr Ala Ser Ala Tyr Glu Glu Lys Ala Tyr Thr Asp Gly Arg Arg
      1125                      1130                      1135

gac aac cca tgt gag agc aac aga ggc tac ggc gac tac acc ccc ctg      3456
Asp Asn Pro Cys Glu Ser Asn Arg Gly Tyr Gly Asp Tyr Thr Pro Leu
      1140                      1145                      1150

ccc gct gga tac gtg acc aag gag ctg gag tac ttc ccc gag acc gac      3504
Pro Ala Gly Tyr Val Thr Lys Glu Leu Glu Tyr Phe Pro Glu Thr Asp
      1155                      1160                      1165

aag gtg tgg atc gag att ggc gag acc gag ggc acc ttc atc gtg gac      3552
Lys Val Trp Ile Glu Ile Gly Glu Thr Glu Gly Thr Phe Ile Val Asp
      1170                      1175                      1180

agc gtg gag ctg ctg ctg atg gag gag tag      3582
Ser Val Glu Leu Leu Leu Met Glu Glu
      1185                      1190

```

<210> 8

<211> 1193

<212> PRT

<213> 人工序列

<223> 人工序列说明：编码带全长 Cry1Ab 尾的 H04 的合成基因

```

<400> 8
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 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
 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 Lys 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 Leu 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 Gln Gly
 725 730 735
 Thr Phe Asp Glu Cys Tyr Pro Thr Tyr Leu Tyr Gln Pro Ile Asp Glu
 740 745 750
 Ser Lys Leu Lys Ala Tyr Thr Arg Tyr Gln 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 Thr Val Asn Val Pro Gly Thr Gly Ser Leu Trp Pro Leu Ser Ala
 785 790 795 800
 Pro 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 Ser 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 Val Gly Glu Ala Leu Ala Arg Val Lys Arg Ala Glu Lys
 885 890 895
 Lys Trp Arg Asp Lys Arg Glu Lys Leu Glu Trp 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 Ala Asp Thr Asn Ile Ala Met Ile His Ala Ala
 930 935 940
 Asp Lys Arg Val His Ser 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 Phe Ser Leu Tyr Asp Ala Arg Asn Val Ile Lys Asn
          980          985          990
Gly Asp Phe Asn Asn Gly Leu Ser Cys Trp Asn Val Lys Gly His Val
          995          1000          1005
Asp Val Glu Glu Gln Asn Asn His Arg Ser Val Leu Val Val 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 Asn 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
Asp Tyr Thr Ala Thr Gln Glu Glu Tyr Glu Gly Thr Tyr Thr Ser Arg
          1090          1095          1100
Asn Arg Gly Tyr Asp Gly Ala Tyr Glu Ser Asn Ser Ser Val Pro Ala
          1105          1110          1115          1120
Asp Tyr Ala Ser Ala Tyr Glu Glu Lys Ala Tyr Thr Asp Gly Arg Arg
          1125          1130          1135
Asp 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 Glu 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
Ser Val Glu Leu Leu Leu Met Glu Glu
          1185          1190

```

<210> 9
 <211> 2007
 <212> DNA
 <213> 人工序列

<220>

<223> 人工序列说明: 编码 H04 加 Cry1Ab 尾头 40 个氨基酸的合成基因

<220>

<221> CDS

<222> (1)..(2007)

<223> 带截短的 Cry1Ab 尾的 H04

<400> 9

```

atg gac aac aac ccc aac atc aac gag tgc atc ccc tac aac tgc ctg      48
Met Asp Asn Asn Pro Asn Ile Asn Glu Cys Ile Pro Tyr Asn Cys Leu
   1              5              10              15

```

```

agc aac ccc gag gtg gag gtg ctg ggc ggc gag cgc atc gag acc ggc      96
Ser Asn Pro Glu Val Glu Val Leu Gly Gly Glu Arg Ile Glu Thr Gly
          20              25              30

```

```

tac acc ccc atc gac atc agc ctg agc ctg acc cag ttc ctg ctg agc      144
Tyr Thr Pro Ile Asp Ile Ser Leu Ser Leu Thr Gln Phe Leu Leu Ser
          35              40              45

```

gag ttc gtg ccc ggc gcc ggc ttc gtg ctg ggc ctg gtg gac atc atc Glu Phe Val Pro Gly Ala Gly Phe Val Leu Gly Leu Val Asp Ile Ile 50 55 60	192
tgg ggc atc ttc ggc ccc agc cag tgg gac gcc ttc ctg gtg cag atc Trp Gly Ile Phe Gly Pro Ser Gln Trp Asp Ala Phe Leu Val Gln Ile 65 70 75 80	240
gag cag ttg ata aac caa cgc ata gag gaa ttc gcc cgc aac cag gcc Glu Gln Leu Ile Asn Gln Arg Ile Glu Glu Phe Ala Arg Asn Gln Ala 85 90 95	288
atc agc cgc ctg gag ggc ctg agc aac ctg tac caa atc tac gcc gag Ile Ser Arg Leu Glu Gly Leu Ser Asn Leu Tyr Gln Ile Tyr Ala Glu 100 105 110	336
agc ttc cgc gag tgg gag gcc gac ccc acc aac ccc gcc ctg cgc gag Ser Phe Arg Glu Trp Glu Ala Asp Pro Thr Asn Pro Ala Leu Arg Glu 115 120 125	384
gag atg cgc atc cag ttc aac gac atg aac agc gcc ctg acc acc gcc Glu Met Arg Ile Gln Phe Asn Asp Met Asn Ser Ala Leu Thr Thr Ala 130 135 140	432
atc ccc ctg ttc gcc gtg cag aac tac cag gtg ccc ctg ctg agc gtg Ile Pro Leu Phe Ala Val Gln Asn Tyr Gln Val Pro Leu Leu Ser Val 145 150 155 160	480
tac gtg cag gcc gcc aac ctg cac ctg agc gtg ctg cgc gac gtc agc Tyr Val Gln Ala Ala Asn Leu His Leu Ser Val Leu Arg Asp Val Ser 165 170 175	528
gtg ttc ggc cag cgc tgg ggc ttc gac gcc gcc acc atc aac agc cgc Val Phe Gly Gln Arg Trp Gly Phe Asp Ala Ala Thr Ile Asn Ser Arg 180 185 190	576
tac aac gac ctg acc cgc ctg atc ggc aac tac acc gac cac gcc gtg Tyr Asn Asp Leu Thr Arg Leu Ile Gly Asn Tyr Thr Asp His Ala Val 195 200 205	624
cgc tgg tac aac acc ggc ctg gag cgc gtg tgg ggt ccc gac agc cgc Arg Trp Tyr Asn Thr Gly Leu Glu Arg Val Trp Gly Pro Asp Ser Arg 210 215 220	672
gac tgg atc agg tac aac cag ttc cgc cgc gag ctg acc ctg acc gtg Asp Trp Ile Arg Tyr Asn Gln Phe Arg Arg Glu Leu Thr Leu Thr Val 225 230 235 240	720
ctg gac atc gtg agc ctg ttc ccc aac tac gac agc cgc acc tac ccc Leu Asp Ile Val Ser Leu Phe Pro Asn Tyr Asp Ser Arg Thr Tyr Pro 245 250 255	768
atc cgc acc gtg agc cag ctg acc cgc gag att tac acc aac ccc gtg Ile Arg Thr Val Ser Gln Leu Thr Arg Glu Ile Tyr Thr Asn Pro Val 260 265 270	816
ctg gag aac ttc gac ggc agc ttc cgc ggc agc gcc cag ggc atc gag Leu Glu Asn Phe Asp Gly Ser Phe Arg Gly Ser Ala Gln Gly Ile Glu 275 280 285	864

ggc agc atc cgc agc ccc cac ctg atg gac atc ctg aac agc atc acc Gly Ser Ile Arg Ser Pro His Leu Met Asp Ile Leu Asn Ser Ile Thr 290 295 300	912
atc tac acc gac gcc cac cgc ggc gag tac tac tgg agc ggc cac cag Ile Tyr Thr Asp Ala His Arg Gly Glu Tyr Tyr Trp Ser Gly His Gln 305 310 315 320	960
atc atg gcc agc ccc gtc ggc ttc agc ggc ccc gag ttc acc ttc ccc Ile Met Ala Ser Pro Val Gly Phe Ser Gly Pro Glu Phe Thr Phe Pro 325 330 335	1008
ctg tac ggc acc atg ggc aac gct gca cct cag cag cgc atc gtg gca Leu Tyr Gly Thr Met Gly Asn Ala Ala Pro Gln Gln Arg Ile Val Ala 340 345 350	1056
cag ctg ggc cag gga gtg tac cgc acc ctg agc agc acc ctg tac cgt Gln Leu Gly Gln Gly Val Tyr Arg Thr Leu Ser Ser Thr Leu Tyr Arg 355 360 365	1104
cga cct ttc aac atc ggc atc aac aac cag cag ctg agc gtg ctg gac Arg Pro Phe Asn Ile Gly Ile Asn Asn Gln Gln Leu Ser Val Leu Asp 370 375 380	1152
ggc acc gag ttc gcc tac ggc acc agc agc aac ctg ccc agc gcc gtg Gly Thr Glu Phe Ala Tyr Gly Thr Ser Ser Asn Leu Pro Ser Ala Val 385 390 395 400	1200
tac cgc aag agc ggc acc gtg gac agc ctg gac gag atc ccc cct cag Tyr Arg Lys Ser Gly Thr Val Asp Ser Leu Asp Glu Ile Pro Pro Gln 405 410 415	1248
aac aac aac gtg cca cct cga cag ggc ttc agc cac cgt ctg agc cac Asn Asn Asn Val Pro Pro Arg Gln Gly Phe Ser His Arg Leu Ser His 420 425 430	1296
gtg agc atg ttc cgc agt ggc ttc agc aac agc agc gtg agc atc atc Val Ser Met Phe Arg Ser Gly Phe Ser Asn Ser Ser Val Ser Ile Ile 435 440 445	1344
cgt gca ccc atg ttc agc tgg att cac cgc agc gcc acc ctg acc aac Arg Ala Pro Met Phe Ser Trp Ile His Arg Ser Ala Thr Leu Thr Asn 450 455 460	1392
acc atc gac ccc gag cgc atc aac cag atc ccc ctg gtg aag ggc ttc Thr Ile Asp Pro Glu Arg Ile Asn Gln Ile Pro Leu Val Lys Gly Phe 465 470 475 480	1440
cgg gtg tgg ggc ggc acc agc gtg atc acc ggc ccc ggc ttc acc gga Arg Val Trp Gly Gly Thr Ser Val Ile Thr Gly Pro Gly Phe Thr Gly 485 490 495	1488
ggc gac atc ctg cgc aga aac acc ttc ggc gac ttc gtg agc ctg cag Gly Asp Ile Leu Arg Arg Asn Thr Phe Gly Asp Phe Val Ser Leu Gln 500 505 510	1536
gtg aac atc aac agc ccc atc acc cag cgt tac cgc ctg cgc ttc cgc Val Asn Ile Asn Ser Pro Ile Thr Gln Arg Tyr Arg Leu Arg Phe Arg 515 520 525	1584
tac gcc agc agc cgc gac gcc cgt gtg atc gtg ctg act ggc gcc gct	1632

```

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

agc acc ggt gtg ggc ggt cag gtg agc gtg aac atg ccc ctg cag aag 1680
Ser Thr Gly Val Gly Gly Gln Val Ser Val Asn Met Pro Leu Gln Lys
545                               550                               555                               560

act atg gag atc ggc gag aac ctg act agt cgc acc ttc cgc tac acc 1728
Thr Met Glu Ile Gly Glu Asn Leu Thr Ser Arg Thr Phe Arg Tyr Thr
                               565                               570                               575

gac ttc agc aac ccc ttc agc ttc cgc gcc aac ccc gac atc atc ggc 1776
Asp Phe Ser Asn Pro Phe Ser Phe Arg Ala Asn Pro Asp Ile Ile Gly
                               580                               585                               590

atc agc gag cag ccc ctg ttc ggt gcc ggc agc atc agc agc ggc gag 1824
Ile Ser Glu Gln Pro Leu Phe Gly Ala Gly Ser Ile Ser Ser Gly Glu
                               595                               600                               605

ctg tac atc gac aag atc gag atc atc ctg gcc gac gcc acc ttc gag 1872
Leu Tyr Ile Asp Lys Ile Glu Ile Ile Leu Ala Asp Ala Thr Phe Glu
                               610                               615                               620

gcc gag agc gac ctg gag cgc gcc cag aag gcc gtg aac gcc ctg ttc 1920
Ala Glu Ser Asp Leu Glu Arg Ala Gln Lys Ala Val Asn Ala Leu Phe
625                               630                               635                               640

acc agc agc aac cag atc ggc ctg aag acc gac gtg acc gac tac cac 1968
Thr Ser Ser Asn Gln Ile Gly Leu Lys Thr Asp Val Thr Asp Tyr His
                               645                               650                               655

atc gac cag gtg agc aac ctg gtg gac tgc tta agc tag 2007
Ile Asp Gln Val Ser Asn Leu Val Asp Cys Leu Ser
                               660                               665

```

<210> 10

<211> 668

<212> PRT

<213> 人工序列

<223> 人工序列说明: 编码 H04 加 Cry1Ab 尾头 40 个氨基酸的合成基因

<400> 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	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	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	
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
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				
	660		665	

<210> 11

<211> 13269

<212> DNA

<213> 人工序列

<220>

<223> 人工序列说明: pNOV1308

<220>

<221> misc_feature

<222> (1)..(1896)

<223> 编码不带尾的 H04 毒素部分的合成核苷酸序列

<220>

<221> misc_feature

<222> (2102)..(4083)

<223> 玉米遍在蛋白启动子

<220>

<221> misc_feature

<222> (4180)..(5283)

<223> PMI 标记基因

<220>

<221> misc_feature

<222> (11247)..(12647)

<223> Zm Ubi 启动子

<400> 11

```

atggacaaca accccaacat caacgagtgat atccccctaca actgcctgag caacccccgag 60
gtggagggtgc tgggcggcga ggcgcacgag accggctaca ccccatcgat catcagcctg 120
agcctgaccc agttcctgct gacgcaggttc gtgcccggcg ccggttctgt gctggggcctg 180
gtggacatca tctggggcat cttcgccccc agccagtgagg acgcttctgt ggtgcagatc 240
gagcagttga taaaccaacg catagaggaa ttcgcccga accaggccat cagccgcctg 300
gaggggcctga gcaacctgta ccaaattctac gccgagagct tccgcgagtg ggagggcgcg 360
cccaccaacc ccgcccctgcg cgaggagatg cgcattccagt tcaacgacat gaacagcgcc 420
ctgaccaccg ccattcccct gtctcgccgtg cagaactacc aggtgcccct gctgagcgtg 480
tacgtgcagg ccgccaacct gcacctgagc gtgtctgcgcg acgtcagcgt gttcggccag 540
cgctgggggt tgcagcgccg caccatcaac agccgctaca acgacctgac ccgctgac 600
ggcaactaca ccgaccacgc cgtgcgctgg tacaacaccg gcctggagcg cgtgtgggg 660
cccgacagcc ggcactggat caggtacaac cagttccgcc gcgagctgac cctgaccgtg 720
ctggacatcg tgagcctgtt ccccaactac gacagccgca cctaccccat ccgcaccgtg 780
agccagctga ccgcgagat ttacaccaac ccggtgctgg agaacttcga cggcagcttc 840
cgcggcagcg ccaggggcat cgagggcagc atccgcagcg cgcggcgagt actactggag cgccaccag 960
aacagcatca ccattctacac cgacgcccac cgcggcgagt actactggag cgccaccag 1020
atcatggcca gcccgctcgg cttcagcggc cccgagttca ccttccccct gtacggcacc 1080
atgggcaacg ctgcacctca gcagcgcac gtggcacagc tgggcccagg agtgtaaccg 1140
accctgagca gcacctgta ccgtcgacct ttcaacatcg gcatcaacaa ccagcagctg 1200
agcgtgctgg acggcaccga gtctgcctac ggcaccagca gcaacctgcc cagcgccgtg 1260
taccgcaaga gcggcaccgt ggacagcctg gacgagatcc cccctcagaa caacaacgtg

```

```

ccacctcgac agggccttcag ccaccgtctg agccacgtga gcatgttccg cagtggcttc 1320
agcaacagca gcgtgagcat catccgtgca cccatgttca gctggattca ccgcagcgcc 1380
accctgacca acaccatcga ccccgagcgc atcaaccaga tccccctggg gaagggttc 1440
cgggtgtggg gcggcaccag cgtgatcacc ggccccgggt tcaccggagg cgacatcctg 1500
cgcagaaaca ccttcggcga cttcgtgagc ctgcaggtga acatcaacag ccccatcacc 1560
cagcgttacc gcctgcgcctt ccgctacgcc agcagccgcg acgcccgtgt gatcgtgctg 1620
actggcgccg ctagcaccgg tgtgggcggg caggtgagcg tgaacatgcc cctgcagaag 1680
actatggaga tcggcgagaa cctgactagt cgcaccttc cctacaccga cttcagcaac 1740
cccttcagct tcgcgcgcaa ccccgacatc atcggcatca gcgagcagcc cctgttcggg 1800
gccggcagca tcagcagcgg cgagctgtac atcgacaaga tcgagatcat cctggccgac 1860
gccaccttcg agggcgagag cgacctggag cgctaagatc tgttctgcac aaagtggagt 1920
agtcagtcac cgatcaggaa ccagacacca gacttttatt catacagtga agtgaagtga 1980
agtgcagtc agtgagttgc tgggttttgt acaacttagt atgtatttgt atttgtaaaa 2040
tacttctatc aataaaatct ctaattccta aaacccaaat ccaggggtac cagcttgcat 2100
gcctgcagtg cagcgtgacc cggtcgtgcc cctctctaga gataatgagc attgcatgtc 2160
taagttataa aaaattacca catatTTTTT ttgtcacact tgtttgaagt gcagtttatc 2220
tatctttata catatatTTT aactttactc tacgaataat ataactata gtactacaat 2280
aatatcagtg ttttagagaa tcataataat gaacagttag acatggtcta aaggacaatt 2340
gagtatTTTg acaacaggac totacagttt tatctTTTTa gtgtgcatgt gttctccttt 2400
TTTTTTgcaa atagcttcac ctatataata cttcatccat tttattagta catccattta 2460
gggtttaggg ttaatggttt ttatagacta atTTTTTTg tacatctatt ttattctatt 2520
ttagcctcta aattaagaaa actaaaactc tatttttagt tttttattta ataatttaga 2580
tataaaatag aataaaataa agtgactaaa aattaaacaa atacccttta agaaattaaa 2640
aaaactaagg aaacatTTTT cttgttttga gtagataatg ccagcctgtt aaacgcgcgc 2700
gacgagtcta acggacacca accagcgaac cagcagcgtc gcgtcggggc aagcaagca 2760
gacggcagcg catctctgtc gctgcctctg gacccctctc gagagtccg atccaccgtt 2820
ggacttgctc cgctgtcgcc atccagaaat tgcgtggcgg agcggcagac gtgagccggc 2880
acggcaggcg ggcctcctct cctctcaccg caccggcagc tacgggggat tcccttccca 2940
ccgctccttc gctttccctt cctcgcgcgc cgtaataaat agacacccc tccacacct 3000
ctttcccaa cctcgtgttg ttcgagcgc acacacacac aaccagatct ccccaaatc 3060
cacccgtcgg cacctcgcct tcaaggtaag ccgctcgtcc tcccccccc cccctctcta 3120
ccttctctag atcggcgctt cgggtccatgg ttagggcccg gtagttctac ttctgttcat 3180
gtttgtgtta gatccgtgtt tgtgttagat ccgtgtcgtc agcgttcgta cccggtcg 3240
acctgtacgt cagacacggt ctgattgcta acctgccagt gtttctcttt ggggaatcct 3300
gggatggctc tagcggttcc gcagacggga tcgatttcat gatttttttt gtttcgttgc 3360
atagggtttg gtttgcctt ttcctttatt tcaatatatg ccgtgcactt gtttgtcggg 3420
tcactctttc atgctTTTT ttgtcttggg tgtgatgatg tggctcgggt gggcggtcgt 3480
tctagatcgg agtagaatc tgtttcaaac tacctggtgg atttattaat tttggatctg 3540
tatgtgtgtg ccatacatat tcatagttac gaattgaaga tgatggatgg aaatatcgat 3600
ctaggatagg tatacatggt gatgcgggtt ttactgatgc atatacagag atgctTTTTg 3660
ttcgcttggt tgtgatgatg tgggtgtggt gggcggtcgt tcattcgttc tagatcggag 3720
tagaatactg tttcaacta cctggtgtat ttattaattt tggaaactga tgtgtgtgtc 3780
atacatcttc atagttacga gtttaagatg gatggaaata tcgatctagg ataggtatac 3840
atgttgatgt gggttttact gatgcatata catgatggca tatgcagcat ctattcatat 3900
gctctaacct tgagtacct tctattataa taaacaagta tgttttataa ttattttgat 3960
cttgatatac ttggatgatg gcatatgcag cagctatatg tggatttttt tagccctgcc 4020
ttcatacget atttattttgc ttggtactgt ttcttttgtc gatgctcacc ctgttgtttg 4080
gtgttacttc tgcagggatc cccgatcatg caaaaactca ttaactcagt gcaaaactat 4140
ccgatggcgg agctgtggat gggcgacat ccgaaaagca gttcacgagt gcagaatgcc 4200
gccggagata tcgtttcact gcgtgatgtg attgagagtg ataaatcgac tetgctcgga 4320
gaggcggttg ccaaacgctt tggcgaactg cctttcctgt tcaaagtatt atgcgcagca 4380
cagccactct ccattcaggt tcattccaaac aaacacaatt ctgaaatcgg ttttgccaaa 4440
gaaaatgccg caggtatccc gatggatgcc gccgagcgta actataaaga tccataaccac 4500
aagccggagg tggttttttgc gctgacgct ttctttgcga tgaacgcgtt tcgtgaattt 4560
tcagagattg tctccctact ccagccggtc gcaggtgcac atccggcgat tgctcacttt 4620
ttacaacagc ctgatgcga acgtttaagc gaactgttcg ccagcctgtt gaatatgcag 4680
ggtgaagaaa aatccgcgc gctggcgatt ttaaaactcg ccctcgatag caatcagggt 4740
gaaccgtggc aaacgattcg ttttaattct gaattttacc cggaagacag cggctcgttc 4800
tccccgtat tgetgaatgt ggtgaaattg aaccctggcg aagcgatgtt cctgttcgct 4860
gaaacaccgc acgcttaact gcaaggcgtg gcgctggaag tgatggcaaa ctccgataac 4920

```

gtgctgcgtg	cggtgtctgac	gcctaaatac	attgatattc	cggaactggt	tgccaatgtg	4980
aaattcgaag	ccaaaccggc	taaccagttg	ttgaccacgc	cggtgaaaca	aggtgcagaa	5040
ctggacttcc	cgattccagt	ggatgatttt	gccttctcgc	tgcatgacct	tagtgataaa	5100
gaaaccacca	ttagccagca	gagtgcgcgc	atthttgttct	gcgtcgaagg	cgatgcaacg	5160
ttgtggaaaag	gttctcagca	gttacagctt	aaaccgggtg	aatcagcggt	tattgccgcc	5220
aacgaatcac	cggtgactgt	caaaggccac	ggcgttttag	cgcgtgttta	caacaagctg	5280
taagagctta	ctgaaaaaat	taacatctct	tgctaagctg	ggagctcgat	ccgtcgacct	5340
gcagatcggt	caaacatttg	gcaataaagt	ttcttaagat	tgaatcctgt	tgccgggtctt	5400
gcgatgatta	tcatataatt	tctgttgaat	tacgttaagc	atgtaataat	taacatgtaa	5460
tgcatgacgt	tatttatgag	atgggttttt	atgattagag	tcccgcattt	atacatttaa	5520
tacgcgatag	aaaacaaaat	atagcgcgca	aactaggata	aattatcgcg	cgcggtgtca	5580
tctatgttac	tagatctgct	agccctgcag	gaaatttacc	ggtgcccggg	cggccagcat	5640
ggcgtatccc	gcaatgtgtt	attaagttgt	ctaagcgtca	atthgtttac	accacaatat	5700
atcctgccac	cagccagcca	acagctcccc	gaccggcagc	tcggcacaaa	atcaccactc	5760
gatacaggca	gcccatcaga	attaattctc	atgtttgaca	gcttatcatc	gactgcacgg	5820
tgaccaaatg	cttctggcgt	caggcagcca	tcggaagctg	tggtatgggt	gtgcaggtcg	5880
taaatacctg	cataattcgt	gtcgctcaag	gcgcactccc	gttctggata	atgttttttg	5940
cgccgacatc	ataacggttc	tggcaaatat	tctgaaatga	gctgttgaca	attaatcatc	6000
cggtctgtat	aatgtgtgga	attgtgagcg	gataacaatt	tcacacagga	aacagaccat	6060
gaggggaagcg	ttgatcgccg	aagtatcgac	tcaactatca	gaggtagttg	gcgtcatcga	6120
gcgccatctc	gaaccgacgt	tgctggccgt	acatttgtac	ggctccgcag	tggtatggcg	6180
cctgaagcca	cacagtgata	ttgattttgt	ggttacgggt	accgtaaggc	ttgatgaaac	6240
aacgcggcga	gctttgatca	acgacctttt	ggaaacttcg	gcttccccct	gagagagcga	6300
gattctccgc	gctgtagaag	tcaccattgt	tgtgcacgac	gacatcatte	cgtggcggtta	6360
tccagctaag	cgcgaaactgc	aatttggaga	atggcagcgc	aatgacattc	ttgcaggtat	6420
cttcgagcca	gccacgatcg	acattgatct	ggctatcttg	ctgacaaaag	caagagaaca	6480
tagcgtttgcc	ttggtaggtc	cagcggcgga	ggaactcttt	gatccgggtt	ctgaacagga	6540
tctatthttag	gcgctaaatg	aaaccttaac	gctatggaa	tcgccgccc	actgggctgg	6600
cgatgagcga	aatgtagtgc	ttacgtttgt	ccgcatttgg	tacagcgag	taaccggcaa	6660
aatcgcgccg	aaggatgtcg	ctgccgactg	ggcaatggag	cgccctgccg	cccagtatca	6720
gcccgtcata	cttgaagcta	ggcaggctta	tcttggacaa	gaagatcgct	tggcctcgcg	6780
cgcagatcag	ttggaagaat	ttgttcaact	cgtgaaaggc	gagatcacca	aagtagtcgg	6840
caaataaagc	tctagtggat	ctcgttacc	ccgggggac	tggctcgccg	cggacgcacg	6900
acgcgggggc	gagaccatag	gcgatctcct	aaatcaatag	tagctgtaac	ctcgaagcgt	6960
ttcactttgta	acaacgattg	agaatttttg	tcataaaaat	gaaataactg	gttcgcattt	7020
ttgtcatccg	cggtcagccg	caattctgac	gaactgccc	tttagctgga	gatgattgta	7080
catccttcac	gtgaaaattt	ctcaagcgt	gtgaacaagg	gttcagattt	tagattgaaa	7140
ggtgagccgt	tgaaacacgt	tcttcttgtc	gatgacgacg	tcgctatgcg	gcattcttatt	7200
attgaatacc	ttacgatcca	cgcttcaaaa	gtgaccgcgg	tagccgacag	caccaggttc	7260
acaagagtac	tctcttccgc	gacggtcgat	gtcgtgggtg	ttgatctaaa	tttaggtcgt	7320
gaagatgggc	tcgagatcgt	tcgtaatctg	gcggcaaaag	ctgatattcc	aatcataatt	7380
atcagtggcg	accgccttga	ggagacggat	aaagtgtgtg	cactcgagct	aggagcaagt	7440
gattttatcg	ctaagccgtt	cagtatcaga	gagtttctag	cacgcattcg	ggttgccttg	7500
cgcggtgcgc	ccaacgttgt	ccgctccaaa	gaccgacggg	ctttttgttt	tactgactgg	7560
acacttaatc	tcaggcaacg	tcgcttgatg	tccgaagctg	gcgggtgagg	gaaacttacg	7620
gcaggtgagt	tcaatcttct	cctcgcggtt	ttagagaaa	cccgcgacgt	tctatcgccg	7680
gagcaacttc	tcattgccag	tcgagtacgc	gacgaggagg	tttatgacag	gagtatagat	7740
gttctcattt	tgaggctgcg	ccgcaaaact	gaggcagatc	cgtcaagccc	tcaactgata	7800
aaaacagcaa	gaggtgccgg	ttattttctt	gacgcggacg	tgcaagtttc	gcacgggggg	7860
acgatggcag	cctgagccaa	ttcccagatc	cccaggaat	cgccgtgagc	ggtcgcaaac	7920
catccggccc	ggtacaaaac	ggcgcgccgc	tgggtgatga	cctgggtggg	aagttgaagg	7980
ccgcgcaggc	cgcccagcgg	caacgcacgc	aggcagaagc	acgccccggg	gaatcggtgg	8040
aagcggccgc	tgatcgaatc	cgcaagaat	cccggcaacc	gcccggcagc	ggtgcgccgt	8100
cgattaggaa	gccgcccaag	ggcgacgagc	aaccagattt	tttcgttccg	atgctctatg	8160
acgtgggcac	ccgcgatagt	cgcagcatca	tggacgtggc	cgthttccgt	ctgtcgaagc	8220
gtgaccgacg	agctggcgag	gtgatccgct	acgagcttcc	agacgggcac	gtagaggttt	8280
ccgcaggggc	ggccggcatg	gccagtgtgt	gggattacga	cctgggtactg	atggcggttt	8340
ccatctaac	cgaatccatg	aaccgatacc	gggaaggga	gggagacaa	cccggcccg	8400
tgthccgtcc	acacgttgcg	gacgtactca	agttctgcgc	gcgagccgat	ggcggaagc	8460
agaaagacga	cctggtagaa	acctgcattc	ggttaaacc	cacgcacgtt	gccatgcagc	8520
gtacgaagaa	ggccaagaac	ggccgcctgg	tgacggtatc	cgagggtgaa	gccttgatta	8580

```

gccgctacaa gatcgtaaag agcgaaaccg ggcggccgga gtacatcgag atcgagctag 8640
ctgattggat gtaccgcgag atcacagaag gcaagaaccc ggacgtgctg acggttcacc 8700
ccgattactt tttgatcgat cccggcatcg gccgttttct ctaccgcctg gcacgcccg 8760
ccgcaggcaa ggcagaagcc agatggttgt tcaagacgat ctacgaacgc agtggcagcg 8820
ccggagagtt caagaagttc tgtttcaccg tgcgcaagct gatcgggtca aatgacctgc 8880
cggagtacga tttgaaggag gaggcggggc aggcctggccc gatcctagtc atgcgctacc 8940
gcaacctgat cgagggcgaa gcatccgccc gttcctaattg tacggagcag atgctagggc 9000
aaattgcctt agcaggggaa aaaggtcgaa aaggtctctt tcctgtggat agcacgtaca 9060
ttgggaaccc aaagccgtac attgggaacc ggaacccgta cattgggaac ccaaagccgt 9120
acattgggaa ccggtcacac atgtaagtga ctgatataaa agagaaaaaa ggcgattttt 9180
ccgcctaaaa ctctttaaaa cttattaaaa ctcttaaaac ccgcctggcc tgtgcataac 9240
tgtctggcca gcgcacagcc gaagagctgc aaaaagcgcc tacccttcgg tcgctgcgct 9300
ccctacgccc cgccgcttcg cgtcggccta tcgcgccgc gccgtcgcca ctcgaccgcc 9420
gcctacggcc aggcgaatcta ccagggcgcg gacaagccgc gccgtcgcca ctcgaccgcc 9480
ggcgctgagg tctgcctcgt gaagaagggtg ttgctgactc ataccaggcc tgaatcgccc 9540
catcatccag ccagaaagtg agggagccac ggttgatgag agctttgttg taggtggacc 9600
agttggtgat tttgaacttt tgctttgcca cggaacggtc tgcgttgctg ggaagatgcg 9660
tgatctgatc cttcaactca gcaaaagttc gattttattca acaaagccgc cgtcccgtca 9720
agtcagcgta atgctctgcc agtgttacaa ccaattaacc aattctgatt agaaaaactc 9780
atcgagcatc aaatgaaact gcaattttatt catatcagga ttatcaatac catatttttg 9840
aaaaagccgt ttctgtaattg aaggagaaaa ctcaccaggg cagttccata ggaatggcaag 9900
atcctggtat cggctctgcga ttccgactcg tccaacatca atacaaccta ttaattttccc 9960
ctcgctcaaa ataagttat caagtgagaa atcaccatga gtgacgactg aatccgggtga 10020
gaatggcaaa agctctgcat taatgaatcg gccaacgcgc ggggagaggc ggtttgcgta 10080
ttgggcgctc ttccgcttcc tcgctcactg actcgctgcg ctcggtcggt cggctgcggc 10140
gagcgggtatc agctcactca aaggcggtaa tacggttatc cacagaatca ggggataacg 10200
caggaaagaa catgtgagca aaaggccagc aaaaggccag gaaccgtaaa aaggccgcgt 10260
tgctggcggt tttccatagg ctccgcccc ctgacgagca tcacaaaaat cgacgtcaa 10320
gtcagagggt gcgaaacccg acaggactat aaagatacca ggcgtttccc cctggaagct 10380
ccctcgctgc ctctcctgtt ccgacctgc cgtttaccgg atacctgtcc gctttctcc 10440
cttcgggaag cgtggcgctt tctcatagct cacgctgtag gtatctcagt tcagcccgcac 10500
tcgttcgctc caagctgggc tgtgtgcacg aaccccccg cagcccgcac cgacttatcg 10560
tatecggtaa ctatcgtctt gagtccaacc cggtagaca cgggtgctaca gagttcttga 10620
cagccactgg taacaggatt agcagagcga ggtatgtagg tggtatctgc gctctgctga 10680
agtgggtggc taactacggc tacactagaa gaacagttat ggtatctgc cgcaaaaca 10740
agccagttac cttcggaaaa agagttggta gctcttgatc cagaaaaaaa ggaatctcaag 10800
gtagcgggtg tttttttgtt tgcaagcagc agattacgcg gaacgaaaac tcaggttaag 10860
aagatccttt gatcttttct acggggtctg acgctcagtg gatecttttg atccggaatt 10920
ggattttggt catgagatta tcaaaaagga tcttcacctc ataaaccttt tcacgccctt 10980
aattcctgtg gttggcatgc acatacaaat ggacgaacgg ttttctctta ggtttaccgg ccaatatatc 11040
ttaaatatcc gattattcta ataaacgctc ttttctctta caatctgac atgagcggag 11100
ctgtcaaaca ctgatagttt aaactgaagg cgggaaacga gatgacgcgg gacaagccgt tttacgtttg 11160
aattaaggga gtcacgttat gacccccgc ctgaggaat tggccgcagc ggcattttaa atcaattggg 11220
gaactgacag aaccgcaacg ctgcaggaat ctgctcgca gtgcagctg acccggtcgt 11280
cgcccgcaat tcgagctcgg tacaagcttg catcctgca taaaaatta ccacatat 11340
gcccctctct agagataatg agcattgcat gctctaagtt atctatcttt atacatat 11400
ttttgtcac acttgtttga agtgcagttt atctatcttt aataatatca gtgttttaga 11460
ctctacgaat aatataatct atagtactac aataatatca ttgacacag gactctacag 11520
aatgaacagt tagacatggt ctaaaggaca tgtgttctcc ttttttttg caaatagctt cacctatata 11580
ttttatcttt ttagtgtgca gtacatccat ttagggttta ggggttaattg tttttataga 11640
atacttcac ctttttatta gtacatccat attttattct attttagcct ctaaattaa gaaaactaaa 11700
ctaatttttt tagtacatct attttattct agatatataa tagaataaaa taaagtgact 11760
ctctattttt gtttttttat ttaataattt aaaaaacta aggaaacatt tttcttgttt 11820
aaaaattaaa caaataccct ttaagaaatt aaaaaacta gtcgacgagt ctaacggaca ccaaccagcg 11880
cgagtagata atgccagcc gttaaaccgc gtcgacgag gcagacggca cggcatctct gtcgctgccc 11940
aaccagcagc gtcgcgtcgg gccaaagcga gttggacttg ctccgctgtc ggcattccaga 12000
ctggacccct ctcgagagtt ccgctccacc gtcgtgagcc ggcacggcag gcggcctcct cctcctctca 12060
aattgcgtgg cggagcggca gactgagcc attcctttcc caccgctcct tegettctcc ttctcggccc 12120
cggcacggca gctacggggg attcctttcc ctctttcccc aacctcgtgt tgttcggagc 12180
gccgtaataa atagacaccc cctccacacc ctctttcccc ggcacctcgg cttaaggtta 12240
gcacacacac acaaccagat ctcccccaaa tccaccgctc ggcacctcgg cttaaggtta

```

```

cgccgctcgt cctccccccc cccccctctc taccttctct agatcggcgt tccgggtccat 12300
ggttagggcc cggtagttct acttctgttc atgtttgtgt tagatccgtg tttgtgttag 12360
atccgtgctg ctageggttcg tacacggatg cgacctgtac gtcagacacg ttctgattgc 12420
taacttgcca gtgtttctct ttgggggaatc ctgggatggc tctagccgtt ccgcagacgg 12480
gatcgatttc atgatttttt ttgtttcgtt gcatagggtt tggtttgccc ttttccttta 12540
tttcaatata tgcggtgcac ttgtttgtcg gggtcatctt tcatgctttt ttttgtcttg 12600
gttgtgatga tgtggtctgg ttggggcggc gttctagatc ggagtagaat tctgtttcaa 12660
actacctggg ggattttatta attttggatc tgtatgtgtg tgccatacat attcatagtt 12720
acgaattgaa gatgatggat ggaaatatcg atctaggata ggtatacatg ttgatgcggg 12780
ttttactgat gcatatacag agatgctttt tgttcgcttg gtttgtgatga tgtggtgtgg 12840
ttggggcggc gttcatctcg tctagatcgg agtagaatac tgtttcaaac tacctgggtg 12900
atttattaat tttggaactg tatgtgtgtg tcatacatct tcatagttac gagtttaaga 12960
tggatggaaa tategatcta ggataggtat acatgttgat gtgggtttta ctgatgcata 13020
tacatgatgg catatgcagc atctattcat atgctctaac cttgagtacc tatctattat 13080
aataaacaag tatgttttat aattattttg atcttgatat acttggatga tggcatatgc 13140
agcagctata tgtggatttt tttageccctg ccttcatacg ctattttatt gcttggtact 13200
gtttcttttg tcgatgctca cctgtttgtt tgggtgttact tctgcaggtc gactctagag 13260
gatccaaca

```

<210> 12

<211> 16179

<212> DNA

<213> 人工序列

<220>

<223> 人工序列说明: pNOV1436

<220>

<221> misc_feature

<222> (1)..(3582)

<223> 编码 H04 毒素部分加全长 CryIAb 尾部分的合成核苷酸序列

<220>

<221> misc_feature

<222> Complement((10390)..(11598))

<223> 磷酸甘露糖异构酶 (PMI) 标记基因

<220>

<221> misc_feature

<222> Complement((12718)..(13608))

<223> 玉米泛在蛋白 (Zm Ubi) 启动子

<220>

<221> misc_feature

<222> (13613)..(16170)

<223> MTL 启动子

<400> 12

```

atggacaaca accccaacat caacgagtgc atcccctaca actgcctgag caaccccgag 60
gtggaggtgc tgggcggcga ggcgcatcgag accggctaca ccccatcga catcagcctg 120
agcctgaccc agttcctgct gagcgagttc gtgcccggcg ccggttctgt gctgggcctg 180
gtggacatca tctggggcat cttcggcccc agccagtggg acgcttctt ggtgcagatc 240
gagcagttga taaaccaacg catagaggaa ttgcgccgca accaggccat cagcgcctg 300
gagggcctga gcaacctgta ccaaactctac gccgagagct tccgcgagtg ggaggccgac 360
cccaccaacc ccgccctgcg cgaggagatg cgcattcagt tcaacgacat gaacagcgcc 420
ctgaccaccg ccateccccct gttcgcctgt cagaactacc aggtgcccc gctgagcgtg 480
tacgtgcagg ccgccaacct gcacctgagc gtgctgcgcg acgtcagcgt gttcggccag 540
cgctggggct tcgacgcgcg caccatcaac agccgctaca acgacctgac ccgcctgac 600

```



```

ggcaactaca ccgaccacgc cgtgcgctgg tacaacaccg gcctggagcg cgtgtggggt 660
cccgacagcc gcgactggat caggtacaac cagttccgcc gcgagctgac cctgaccgtg 720
ctggacatcg tgagcctggt ccccaactac gacagccgca cctaccccat ccgcaccgtg 780
agccagctga cccgcgagat ttacaccaac cccgtgctgg agaacttcga cggcagcttc 840
cgccggcagcg cccagggcat cgagggcagc atccgcagcc cccacctgat ggacatcctg 900
aacagcatca ccatctacac cgacgcccac cgccggcagt actactggag cggccaccag 960
atcatggcca gcccgcctgg cttcageggc cccgagttca ccttccccct gtacggcacc 1020
atggggcaacg ctgcacctca gcagcgcac gtggcacagc tgggccaggg agtgtaccgc 1080
accttgagca gcacctgta ccgtcgacct ttcaacatcg gcatacaaa ccagcagctg 1140
agcgtgctgg acggcaccga gttcgectac ggcaccagca gcaacctgcc cagcgcctg 1200
taccgcaaga gcggcaccgt ggacagcctg gacgagatcc cccctcagaa caacaacgtg 1260
ccacctcgac agggcttcag ccaccgtctg agccacgtga gcatgttcgc cagtggcttc 1320
agcaacagca gcgtgagcat catccgtgca cccatgttca gctggattca ccgcagcgcc 1380
acctgacca acaccatcga ccccgagcgc atcaaccaga tccccctggt gaagggttc 1440
cgggtgtggg gcggcaccag cgtgatcacc ggccccggct tcaccggagg cgacatcctg 1500
cgcagaaaca ccttcggcga cttcgtgagc ctgcaggtga acatcaacag ccccatcacc 1560
cagcgttacc gcctgcgctt ccgctacgcc agcagccgcg acgcccgtgt gatcgtgctg 1620
actggcgccg ctagcaccgg tgtgggcggt caggtgagcg tgaacatgcc cctgcagaag 1680
actatggaga tcggcgagaa cctgactagt cgcaccttc gctacaccga cttcagcaac 1740
cccttcagct tccgcgccaa ccccgacatc atcggcatca gcgagcagcc cctgttcggt 1800
gccggcagca tcagcagcgg cgagctgtac atcgacaaga tcgagatcat cctggccgac 1860
gccaccttcg aggcgagag cgacctggag cgcccccaga aggcctgtaa cgccctgttc 1920
accagcagca accagatcgg cctgaagacc gacgtgaccg actaccacat cgaccaggtg 1980
agcaacctgg tggactgctt aagcgacgag ttctgcctgg acgagaagaa ggagctgagc 2040
gagaaggtga agcacgccaa gcgctgagc gacgagcgca acctgctgca ggaccccaac 2100
ttccgcggca tcaaccgcca gctggaccgc gctggcgag tacgtgacc tgcagggcac cttcgacgag 2220
cagggcgcg acgacgtgtt caaggagaac gacgagagca agctgaaggc ctacaccgc 2280
tgctacccca ctaectgta ccagccgatc gacgagagca agctgaaggc ctacaccgc 2280
taccagctgc gcggtacat cgaggacagc caggacctgg aaatctacct gatccgctac 2340
aacgcgaagc acgagaccgt gaacgtgcc ggaccggca gcctgtggcc cccgagcgcc 2400
cccagcccca tcggcaagtgc cggggagccg aatcgatgcg ctccgcacct ggagtggaa 2460
ccggacctag actgcagctg cagggacggg gagaagtgcg cccaccacag ccaccacttc 2520
agcctggaca tcgacgtggg ctgcaccgac ctgaacgagg acctgggcgt gtgggtgatc 2580
ttcggatcca acgaccagga cggccacgcc cgctgggca atctagagtt cctggaggag 2640
aagccccctg tgggcgaggc cctggccgcg gtgaagcgtg ctgagaagaa gtggcgcgac 2700
aagcgcgaga agctggagtgc ggagaccaac atcgtgtaca aggaggccaa ggagagcgtg 2760
gacgccccgt tcgtgaacag ccagtacgac cgctgcagg ccgacaccaa catcgccatg 2820
atccacgccc cgacaagcg cgtgcacagc attcgcgagg cctacctgcc cgagctgagc 2880
gtgatccccg gtgtgaacgc cgccatcttc gaggaactcg agggccgcat cttcaccgcc 2940
ttcagcctgt acgacgccc caacgtgatc aagaacggcg acttcaacaa cggcctgagc 3000
tgctggaacc tgaaggcca cgtggacgtg gaggagcaga acaaccaccg cgccctgctg 3060
gtggtgcccc agtgggaggc cgaggtgagc caggaggtgc gcgtgtgcc cggccgcggt 3120
tacatcctgc gcgtgaccgc ctacaaggag ggctacggcg agggctgcgt gaccatccac 3180
gagatcgaga acaacaccga cgagctcaag ttcagcaact gcgtggagga ggaggtttac 3240
cccaacaaca ccgtgacctg caacgactac accgcgaccc aggaggagta cgaaggcacc 3300
tacacctctc gcaacagggg ttacgacggc gcctacgagt ccaacagctc cgtgccagct 3360
gactacgcca gcgcccacga ggagaaagcc tacaccgacg gtagacgcga caacccatgt 3420
gagagcaaca gaggctacgg cgactacacc cccgtgcccg ctggatacgt gaccaaggag 3480
ctggagtact tccccgagac cgacaagggt tggcgagac ttggcgagac cgagggcacc 3540
ttcatcgtgg acagcgtgga gctgctgctg atggaggagt agtagatctg ttctgcacaa 3600
agtggagtag tcagtcatcg atcaggaacc agacaccaga cttttattca tacagtgaag 3660
tgaagtgaag tgcagtgcag tgagttgctg gttttgttac cacttagtat gtatttgtat 3720
ttgtaaaata cttctatcaa taaaatttct aattcctaaa accaaaatcc agtgggtacc 3780
agcttgggct gagtggctcc ttcaacgttg cggttctgtc agttccaaac gtaaaacggc 3840
ttgtcccgcg tcatcggcgg ggtcataaac gtgactccct taattctccg ctcatgatca 3900
gattgtcggt tcccccttc agtttaaact atcagtgttt gacaggatat attggcgggt 3960
aaacctaaaga gaaaagagcg tttattagaa taaccgatat ttaaaagggc gtgaaaaggt 4020
ttatccgttc gtccatttgt atgtgcatgc caaccacag gtccccctcg ggagtgttg 4080
gcattccgta cgataatgac ttctgttcaa ccacccaaac gtcggaaagc ctgacgacgg 4140
agcagcattc caaaaagatc ccttggctcg tctgggtcgg ctagaaggtc gagtgggctg 4200
ctgtggcttg atccctcaac gcggtcgcg acgtagcgca gcgccgaaaa atcctcgatc 4260

```

```

gcaaataccga cgctgtcgaa aagcgtgac tgcttgtcgc tctttcggcc gacgtcctgg 4320
ccagtcacac cgcgccaaag ttccgtcaca ggatgatctg gcgcgagttg ctggatctcg 4380
ccttcaatcc gggtctgtgg cggaactcc acgaaaatat ccgaacgcag caagatcgtc 4440
gaccaattct tgaagacgaa agggcctcgt gatacgccta tttttatagg ttaatgtcat 4500
gataataatg gtttcttaga cgtcaggtgg cacttttcgg ggaaatgtgc gcggaacccc 4560
tatttgttta tttttctaaa tacattcaaa tatgtatccg ctcatgagac aataaccctg 4620
ataaatgctt caataatatt gaaaaaggaa gagtatgagt attcaacatt tccgtgtcgc 4680
ccttattccc ttttttgcgg cattttgcct tccgtttttt gctcaccag aaacgctggt 4740
gaaagtaaaa gatgtgaag atcagttggg tgcacgagtg gggtacatcg aactggatct 4800
caacagcggg aagatccttg agagttttcg ccccgagaa cgttttccaa tgatgagcac 4860
ttttaaagtt ctgctatgtg gcgcggtatt atcccggtgt gacgcggggc aagagcaact 4920
cggtcgcccgc atacactatt ctccagaatga cttgggttgag tactcaccag tcacagaaaa 4980
gcatcttacg gatggcatga cagtaagaga attatgcagt gctgccataa ccatgagtga 5040
taacactgcg gccaaactac ttctgacaac gatcggagga ccgaaggagc taaccgcttt 5100
tttgcaaac atgggggatc atgtaactcg ccttgatcgt tgggaaccgg agctgaatga 5160
agccatacca aacgacgagc gtgacaccac gatgcctgca gggggggggg ggggggggac 5220
atgaggttgc cccgtattca gtgtcgtga tttgtattgt ctgaagttgt ttttacgtta 5280
agttgatgca gatcaattaa tacgatacct gcgtcataat tgattatttg acgtggtttg 5340
atggcctcca cgcaggttgt gatattgtag tgataatcat taccacttta cgggtccttt 5400
ccggtgatcc gacaggttac ggggcggcga cctcgcgggt tttegtatt tatgaaaatt 5460
ttccggttta aggcgtttcc gttcttcttc gtcataactt aatgttttta tttaaaatac 5520
cctctgaaaa gaaaggaaac gacaggtgct gaaagcgagg ctttttggcc tctgtcgttt 5580
cctttctctg tttttgtcgg tggaaatgaa aatggaagtc cccccccccc cccccctg 5640
cagcaatggc aacaacgttg cgcaactat taactggcga actacttact ctagcttccc 5700
ggcaacaatt aatagactgg atggaggcgg ataaagtgc aggaccactt ctgcgtcgg 5760
cccttcggc ttgctggttt attgctgata aatctggagc cggtagagct gggctcgcg 5820
gtatcattgc agcactgggg ccagatggta agccctcccg tatcgtagtt atctacacga 5880
cggggagtcg ggcaactatg gatgaacgaa atagacagat cgctgagata ggtgcctcac 5940
tgattaagca ttggttaactg tcagaccaag ttactcata tatactttag attgatataa 6000
aacttcattt ttaatttaaa aggatctagg tgaagatcct ttttgataat ctcatgacca 6060
aaatccctta acgtgagttt tegtccact gagcgtcaga ccccgtagaa aagatcaaag 6120
gatcttcttg agatcctttt tttctgcgcg taactctgtg cttgcaaaca aaaaaccac 6180
cgctaccagc ggtggtttgt ttgcccgtac aagagctacc aactcttttt ccgaaggtaa 6240
ctggcttcag cagagcgcag ataccaata ctgtccttct agttagagcc tagttaggcc 6300
accacttcaa gaactctgta gcaccgccta catacctcgc tctgctaate ctgttaccag 6360
tggtgtctgc cagtggcgat aagtcgtgtc ttaccgggtt ggactcaaga cgatagttac 6420
cggataaggc gcagcggtcg ggtgaacgg ggggttcgtg cacacagccc agcttggagc 6480
gaacgacctc caccgaactg agatacctac agcgtgagct atgagaaagc gccacgcttc 6540
ccgaagggag aaaggcggac aggtatccgg taagcggcag ggtcggaaca ggagagcgca 6600
cgagggagct tccaggggga aacgcctggt atctttatag tctgtcggg tttcgccacc 6660
tctgacttga gcgtcgattt ttgtgatgct cgtcaggggg gcggagccta tggaaaaacg 6720
ccagcaacgc ggctttttta cggttccttg ccttttgcgt gccttttgc tcatgtttct 6780
ttctgtcgtt atccctgat tctgtggata accgtattac cgcctttgag tgagctgata 6840
ccgtcgcgcg cagccgaacg accgagcgca gcgagtcagt gagcgaggaa gcggaagagc 6900
gcctgatgcg gtattttctc cttacgcata tgtgcggtat ttcacaccgc atatggtgca 6960
ctctcagtac aatctgctct gatgccgat agttaagcca gtatacactc cgtatcgtct 7020
acgtgactgg gtcattggtg cgcgccgaca cccgccaaca cccgtgacg cgcctgacg 7080
ggcttgtctg ctcccgcac cgccttacag acaagctgtg accgtctccg ggagctgcat 7140
gtgtcagagg ttttcaccgt catcaccgaa acgcgcgagg cagcagatcc cccgatcaag 7200
tagatacact acatatactt acaatagaca tcgagccgga aggtgatgtt tactttctct 7260
aaatccccag caatttttag ccagttttta cccaagactt cgcctctaac ataaattata 7320
gttaccaaat ctggcaaaag ggttaacaag tggcagcaac ggattcgcaa acctgtcacg 7380
ccttttgtgc caaaagccgc gccaggtttg cgatccgctg tgccaggcgt taggcgtcat 7440
atgaagattt cggtgatccc tgagcaggtg gcggaacat tggatgctga gaaccatttc 7500
attgttcgtg aagtgttcga tgtgcacctc tccgaccaag gctttgaact atctaccaga 7560
agtgtgagcc cctaccggaa ggattacatg tccgatgatg actctgatga agactctgct 7620
tgctatggcg cattcatcga ccaagactt tcgcgggaaga ttgaactcaa ctcaactgg 7680
aacgatctag cctctatcga acacattgtt gtgtcgcaca cgcaccgagg caaaggagtc 7740
gcgcacagtc tcatcgaatt tgcgaaaaag tgggcactaa gcagacagct ccttggcata 7800
cgattagaga cacaaacgaa caatgtacct gcctgcaatt tgtacgcaaa atgtggcttt 7860
actctcggcg gcattgacct gttcacgtat aaaactagac ctcaagtctc gaacgaaaca 7920

```

```

gcgatgtact ggtactgggt ctcggggagca caggatgacg cctaacaatt cattcaagcc 7980
gacaccgctt cgcggcgcg ctttaattcag gagttaaaca tcatgaggga agcgggtgatc 8040
gccgaagtat cgactcaact atcagaggta gttggcgctca tcgagcgcca tctcgaaccg 8100
acgtttgctgg ccgtacattt gtacggctcc gcagtggatg gcggcctgaa gccacacagt 8160
gatattgatt tgctgggttac ggtgaccgta aggccttgatg aaacaacgcg gcgagccttg 8220
atcaacgacc ttttggaac ttcggcttcc cctggagaga gcgagattct ccgcgctgta 8280
gaagtcacca ttgttgtgca cgacgacatc attcctgggc gttatccagc taagcgcgaa 8340
ctgcaatttg gagaatggca gcgcaatgac attcttgcag gtatcttcga gccagccacg 8400
atcgacattg atctggctat cttgctgaca aaagcaagag aacatagcgt tgccttggtg 8460
ggccagcgcg cggaggaact ctttgatccg gttcctgaac aggatctatt tgaggcgcta 8520
aatgaaacct taacgctatg gaactcgccg cccgaactggg ctggcgatga gcgaaatgta 8580
gtgcttacgt tgtcccgcat ttggtacagc gcagtaaccg gcaaaatcgc gccgaaggat 8640
gtcgtgccc actgggcaat ggagcgccct cccgcccagt atcagcccgt catacttgaa 8700
gctaggcagg cttatcttgg acaagaagat cgccttggcct cgcgcgcaga tcagttggaa 8760
gaatttgttc actacgtgaa aggcgagatc accaaggtag tcggcaata atgtctaaca 8820
attcgttcaa gccgacgcgc cttcggggcg cggcttaact caagcgttag agagctgggg 8880
aagactatgc gcgatctgtt gaagggtggt ctaagcctcg tacttgcat ggcctgggg 8940
caggcacttg ctgactgccc aattgtttta ttggatgaag ctcgctctcc ctatgactac 9000
tccccatcca actacgacat ttctccaagc aactacgaca actccataag caattacgac 9060
aatagtccat caaattacga caactctgag agcaactacg ataatagttc atccaattac 9120
gacaatagtc gcaacggaaa tcgtaggctt atatatagcg caaatgggtc tcgcactttc 9180
gccggctact acgtcattgc caacaatggg acaacgaact tcttttccac atctggcaaa 9240
aggatgttct acaccccaaa agggggggcg ggcgtctatg gcggcaaaga tgggagcttc 9300
tgccgggcat tggctgctcat aaatggccaa ttttcgcttg cctgacaga taacggcctg 9360
aagatcatgt atctaagcaa ctagcctgct ctctaataaa atgttaggcc tcaacatcta 9420
gtcgcaagct gaggggaacc actagtgtca tacgaacctc caagagacgg ttacacaaac 9480
gggtacattg ttgatgtcat gtatgacaat cgcceaagta agtatccagc tgtgttcaga 9540
acgtacgtcc gaattaatte atcggggtac ggtcgacgat cgtcaacgtt cacttctaaa 9600
gaaatagcgc cactcagctt cctcagcggc tttatccagc gatttctctat tatgtcggca 9660
tagttctcaa gatcgacagc ctgtcacggt taagcgagaa atgaataaga aggctgataa 9720
ttcggatctc tcgaggggag atgatatttg atcacaggca gcaacgctct gtcacgtta 9780
caatcaacat gctaccctcc gcgagatcat ccgtgtttca aaccggcag cttagttgcc 9840
gttcttccga atagcatcg taacatgagc aaagtctgcc gccttacaac ggcctcccg 9900
gtgacgcgt cccggactga tgggctgcct gtatcgatg gtgattttgt gccgagctgc 9960
cggtcgggga gctgttggct ggctgggtggc aggatatatt gtggtgtaaa caaattgacg 10020
cttagacaac ttaataaac attgcgagc tttttaatgt actgaattgt ctagaccgg 10080
ggatctcatg tttgacagct tatcatcgga tctagtaaca tagatgacac cgcgcgcgat 10140
aatttatcct agtttgcgcg ctatatattt tttctatcg cgtattaaat gtataattgc 10200
gggactctaa tcataaaaa ccatctcata aataacgtca tgcattacat gttaattatt 10260
acatgcttaa cgtaattcaa cagaaattag atgataatca tcgcaagacc ggcaacagga 10320
ttcaacttta agaaacttta ttgccaaatg ttgaacgat ctctgcaggt cgacgatcg 10380
agctccagc tttagcaag agtttaattt tttcagtaag ctcttacagc ttgttgtaaa 10440
cacgcgctaa acggccgtgg cctttgacag tcaccggtga ttctgttggc gcaataaacg 10500
ctgattcacc cggtttaagc tgtaactgct gagaaccttt ccacaacgtt gcacgcctt 10560
cgacgcagaa caaaatggcg gcactctgct ggctaattgt ggtttcttta tctaaggt 10620
catgcagcga gaaggcaaaa tcatccactg gaatcgggaa gtccagttct gcacctgtt 10680
tcaccggctg ggtcaacaac tggttagccg gtttggttc gaatttcaca ttggcaacca 10740
gttcgggaat atcaatgtat ttaggcgtca gaccgcacg cagcacgtta tcgagtttg 10800
ccatcacttc cagcggccag ccttgcaggt aagcgtgcg tgtttcagc aacaggaaca 10860
tcgcttcgcc agggttcaat ttcaccacat tcagcaatag cggggagaa agaccgtgt 10920
cttcgggta aaattcagaa attaaacgaa tcgtttgcca cggttcacc tgctggctat 10980
cgagggcgga ttttaaaat gccagcgcg gggatttttc ttcacctgc atattcaaca 11040
ggctggcgaa cagttcgctt aaacgttcgg catcaggctg ttgtaaaaag tgagcaatcg 11100
ccgatgtgc acctgcgacc ggctggagta gggagacaat ctcgaaaat tcacgaaacg 11160
cgttcatcgc aaggaaaggc gtcagcgcaa aaaccagctc cggcttgttg ttaggatctt 11220
tatagttacg ctggcgggca tccatcgga tacctgcggc attttcttg gcaaaaccga 11280
tttcagaatt gtgtttgtt ggatgaacct gaatggagag tggctgtgt gcgcataata 11340
ctttgaacag gaaaggcagt tcgccaagc gttttgcaac ggcctctcc agcagagtc 11400
atztatcact ctcaatcaca tcacgcagt caacgatata tccggcgga ttctgactc 11460
gtgaactgct tttcgatgt gcgccatcc acagctcggc catcggttg ctggacggat 11520
tttccatacc ataaagtta gtcaacgcgt tttgctgcc caggcatagt tttgactga 11580

```

```

gttaatgagt ttttgcattga tcgggggatcc ctgcagaagt aacaccaaac aacagggtga 11640
gcatcgacaa aagaaacagt accaagcaaa taaatagcgt atgaaggcag ggctaaaaaa 11700
atccacatat agctgctgca tatgccatca tccaagtata tcaagatcaa aataattata 11760
aaacatactt gtttattata atagataggt actcaagggt agagcatatg aatagatgct 11820
gcatatgcca tcatgtatat gcatcagtaa aacccacatc aacatgtata cctatcctag 11880
atcgatattt ccatccatct taaactcgta actatgaaga tgtatgacac acacatacag 11940
ttccaaaatt aataaataca ccaggtagtt tgaaacggcg tctactccga tctagaacga 12000
atgaacgacc gcccaaccac accacatcat cacaaccaag cgaacaaaaa gcatctctgt 12060
atatgcatca gtaaaacccg catcaacatg tatacctatc ctagatcgat atttccatcc 12120
atcatcttca attcgtaact atgaatatgt atggcacaca catacagatc caaaattaat 12180
aaatccacca ggtagtttga aacagaattc tactccgata tagaacgacc gcccaaccag 12240
accacatcat cacaaccaag aaaaaaaaaa gcatgaaaag atgacccgac aaacaagtgc 12300
acggcatata ttgaaataaa ggaaaagggc aaaccaaac ctatgcaacg aaacaaaaaa 12360
aatcatgaaa tcgatcccggt ctgcggaacg gctagagcca tcccaggatt ccccaaagag 12420
aaacactggc aagtttagcaa tcagaacgtg tctgacgtac aggtcgcatc cgtgtacgaa 12480
cgctagcagc acggatctaa cacaacacg gatctaacac aaacatgaac agaagtagaa 12540
ctaccggggc ctaaccatgg accggaacgc cgatctagag aaggtagaga gggggggggg 12600
gggaggacga gcggcgatcc ttgaagcgga ggtgccgacg ggtggatttg ggggagatct 12660
ggttgtgtgt gtgtgcgctc cgaacaacac gaggttgggg aaagaggggt tggagggggg 12720
gtctatttat tacggcgggc gaggaaggga aagcgaagga gcggtgggaa aggaatcccc 12780
cgtagctgcc gtgccgtgag aggaggagga ggccgcctgc cgtgccggct cactctgcc 12840
gctccgccac gcaatttctg gatgccgaca gcggagcaag tccaacggtg gagcggaact 12900
ctcgagaggg gtccagaggc agcgacagag atgccgtgcc gtctgcttcg cttggcccg 12960
cgcgacgctg ctggttcgct ggttggtgtc cgttagactc gtcgacggcg tttaacaggc 13020
tggcattatc tactcgaaac aagaaaaatg ttcccttagt ttttttaatt tcttaaaggg 13080
tatttgttta atttttagtc actttatttt attctatttt atatctaaat tattaataa 13140
aaaaactaaa ataggctttt agttttctta atttagaggc taaaatagaa taaaatagat 13200
gtactaaaaa aattagtcta taaaaaccat taaccctaaa cctaaatgg atgtactaat 13260
aaaatggatg aagtattata taggtgaagc tatttgcaaa aaaaaaggag aacacatgca 13320
cactaaaaag ataaaactgt agagtcctgt tgtcaaaata ctcaattgtc ctttagacca 13380
tgtctaactg ttcatttata tgattctcta aaacactgat attattgtag tactatagat 13440
tatattatcc gtagagtaaa gtttaaatat atgtataaag atagataaac tgcacttcaa 13500
acaagtgtga aaaaaaaaaa atgtggtaat tttttataac ttagacatgc aatgctcatt 13560
atctctagta aggggcacga cgggtcagc attcactgca ggcattgcaa cttgcacatg 13620
acaacaattg taagaggatg gagaccacaa cgatccaaca atacttctgc gacgggctgt 13680
gaagtataga gaagttaaac gcccaaaagc cattgtgttt ggaattttta gttattctat 13740
ttttcatgat gtatcttctt ctaacatgcc ttaatttgca aatttggtat aactactgat 13800
tgaaaatata tgtatgtaaa aaaatactaa gcatatttgt gaagctaaac atgatgttat 13860
ttaagaaaat atgttggtta cagaataaga ttaatatcga aatggaaaca tctgtaaatt 13920
agaatcatct tacaagctaa gagatgttca cgctttgaga aacttcttca gatcatgacc 13980
gtagaagtga cctcacaaga ctcaacgaag gctggtgcaa ttccacaaat gcatgacatg 14040
catccttgta accctcgctc ccgctataaa cagcgtaaac tcaattccct gctccatcaa 14100
tttagaaatg agcaagcaag cacccgatcg ctcaccccat atgcaccaat ctgactccca 14160
agtctctgtt tcgcattagt accgccagca ctccacctat agctaccaat tgagaccttt 14220
ccagcctaag cagatcgatt gatcgttaga gtcaaagagt tgggtgtacg ggtactttaa 14280
ctaccatgga atgatggggc gtgatgtaga gcggaagcg cctccctacg cggaacaaca 14340
ccctcgccat gccgctcgac tacagcctcc tctctgctcg ccgccacaa cgagggagcc 14400
cgtggctgca gccaccgacc agcatgtctc tgtgtctctg tccgacctcg acatgtcatg 14460
gcaaacagtc ggacgccagc accagactga cgacatgagt ctctgaagag cccgccacct 14520
agaaagatcc gagccctgct gctggtagt gtaaccattt tcgtcgcgct gacgcggaga 14580
gcgagaggcc agaaatttat agcgactgac gctgtggcag gcacgctatc ggaggttacg 14640
acgtggcggg tctactcgacg cggagttcac aggtcctatc cttgcatcgc tcgggccgga 14700
gtttacggga cttatcttta cgacgtgctc taaggttgcg ataacggcg gaggaaggcg 14760
tgtggcgtgc ggagacgggt tatacacgta gtgtgcggga gtgtgtttcg tagacggcg 14820
aaagcacgac gacttacgaa ggttagtgga ggaggaggac aactaaaaat caggacgcaa 14880
gaaactcttc tattatagta gtagagaaga gattatagga gtgtgggttg attctaaaga 14940
aaatcgacgc aggacaaccg tcaaacggg tgctttaata tagtagatat atatatatag 15000
agagagagag aaagtacaaa ggatgcattt gtgtctgcat atgacggag tattactaac 15060
ggcgctcgta agaaggctca tcatgcgtgg agcgagccca tttggttggt tgtcaggccg 15120
cagttaaggc ctccatatat gattgtctgc gggcccataa cagcatctcc tccaccagtt 15180
tattgtaaga ataaattaag tagagatatt tgtcgtcggg cagaagaaac ttggacaaga 15240

```

```

agaagaagca agctaggcca atttcttgcc ggcaagagga agatagtggc ctctagttta 15300
tatatcggcg tgatgatgat gctcctagct agaaatgaga gaagaaaaac ggacgcgtgt 15360
ttggtgtgtg tcaatggcgt ccacccctcc atcagatcag aacgatgaaa aagtcaagca 15420
cggcatgcat agtatatgta tagcttggtt tagtgtggct ttgctgagac gaatgaaagc 15480
aacggcgggc atatTTTTca gtggctgtag ctttcaggct gaaagagacg tggcatgcaa 15540
taattcaggg aattcgtcag ccaattgagg tagctagtca acttgtacat tggcgcgagc 15600
aattttccgc actcaggagg gctagtttga gagtccaaaa actataggag attaaagagg 15660
ctaaaatcct ctcccttatt aattttaaat aagtagtgta tttgtatttt aactcctcca 15720
acccttccga ttttatggct ctcaaactag cattcagtct aatgcatgca tgcttggtta 15780
gaggctgcat ggggttggtt atagcatagc tagctacaag ttaaccgggt cttttatatt 15840
taataaggac aggcaaagta ttacttacia ataaagaata aagctaggac gaactcgtgg 15900
attattacta aatcgaaatg gacgtaatat tccaggcaag aataattgtt cgatcaggag 15960
acaagtgggg cattggaccg gttcttgcaa gcaagagcct atggcggtgg gacacggcgc 16020
gttgcccata catcatgcct ccacgatga tccatcctca cttgctataa aaagaggtgt 16080
ccatggtgct caagctcagc caagcaaata agacgacttg tttcattgat tcttcaagag 16140
atcgagcttc ttttgacca caagtcgag gatccaaca 16179

```

<210> 13

<211> 15643

<212> DNA

<213> 人工序列

<220>

<223> 人工序列说明: pNOV1441

<220>

<221> misc_feature

<222> (14)..(1414)

<223> 玉米泛在蛋白 (Mz Ubi) 启动子

<220>

<221> misc_feature

<222> (2037)..(5618)

<223> 编码 H04 毒素部分加全长 Cry1Ab 尾部分的合成核苷酸序列

<220>

<221> misc_feature

<222> (5821)..(6711)

<223> Mz Ubi 启动子

<220>

<221> misc_feature

<222> (7831)..(9039)

<223> PMI

<400> 13

```

aagctggtac aagcttgcat gctgcagtg cagcgtgacc cggctcgtgcc cctctctaga 60
gataatgagc attgcatgtc taagttataa aaaattacca catatttttt ttgtcacact 120
tgtttgaagt gcagtttatc tatctttata catatattta aactttactc tacgaataat 180
ataatctata gtactacaat aatatcagtg ttttagagaa tcatataaat gaacagttag 240
acatggctca aaggacaatt gagtatTTTg acaacaggac tctacagttt tatcttttta 300
gtgtgcatgt gttctccctt ttttttgcaa atagcttcac ctatataata cttcatccat 360
tttattagta catccattta gggtttaggg ttaatggttt ttatagacta atttttttag 420
tacatctatt ttattctatt ttgcctcta aattaagaaa actaaaactc tatttttagt 480
tttttatTTa ataatttaga tataaaatag aataaaataa agtgactaaa aattaaacaa 540
ataccctTTa agaaattaaa aaaactaagg aaacattttt cttgtttcga gtagataatg 600
ccagcctggt aaacgcgcgc gacgagtcta acggacacca accagcgaac cagcagcgtc 660
cgctcggggc aagcgaagca gacggcacgg catctctgtc gctgcctctg gacccctctc 720

```

```

gagagtccg cccaccgtt ggacttgctc cgtgtgcgc atccagaaat tgcgtggcgg 780
agcggcagac gtgagccggc acggcaggcg gcctcctcct cctctcacgg caccgcagct 840
acgggggatt cctttccac cgtcctctcg ctttcccttc ctgcccgc gtaataaata 900
gacacccct ccacaccctc tttccccaac ctggtgtgt tggagcgca cacacacaca 960
accagatctc ccccaaatcc acccgtcggc acctccgctt caaggtaacg cgtcgtcct 1020
ccccccccc cctctctac cttctctaga tcggcgttcc ggtccatggt tagggcccgg 1080
tagttctact tctgttcagt tttgtgttag atccgtgttt gtgttagatc cgtcgtgcta 1140
gcggttgtae acggatgcga cctgtacgtc agacacgttc tgattgctaa cttgccagtg 1200
tttctctttg gggaatcctg ggatggctct agccgttccg cagacgggat cgatttcctg 1260
attttttttg tttcgttgca taggttttg tttgcccttt tcttttattt caatatatgc 1320
cgtgcacttg tttgtcgggt catcttttca tgcctttttt tgtcttggtt gtgatgatgt 1380
ggctctgggt ggcggtcgtt ctagatcgga gtagaattct gtttcaaact acctgggtgga 1440
tttattaatt ttggatctgt atgtgtgtgc catacatatt catagttacg aattgaagat 1500
gatggatgga aatatcgatc taggataggt atacatgttg atcggggttt tactgatgca 1560
tatacagaga tgcctttttgt tcgcttggtt gtgatgatgt ggtgtgggtt ggcggtcgtt 1620
cattcgttct agatcggagt agaatactgt ttcaaactac ctggtgtatt tattaatttt 1680
ggaactgtat gtgtgtgtca tacatcttca tagttacgag tttaagatgg atggaatat 1740
cgatctagga taggtataca tgttgatgtg ggttttactg atgcatatac atgatggcat 1800
atgcagcatc tattcatatg ctctaaccct gagtacctat ctattataat aaacaagtat 1860
gttttataat tattttgatc ttgatatact tggatgatgg catatgcagc agctatatgt 1920
ggattttttt agccctgcct tcatacgtca ttattttgct tggtaactgt tcttttgtcg 1980
atgctcaccg tgttgttttg tgttacttct gcaggtcgac tctagaggat ccaacaatgg 2040
acaacaacc caacatcaac gagtgcaccc cctacaactg cctgagcaac cccgaggtgg 2100
aggtgctggg cggcgagcgc atcgagaccg gctacacccc catcgacatc agcctgagcc 2160
tgacccagtt cctgctgagc gaggctcgtc cggcgccggg cttcgtgctg ggccgtgggt 2220
acatcatctg gggcatcttc gggcccagcc agtgggacgc cttcctgggt cagatcgagc 2280
agttgataaa ccaacgcata gaggaattcg cccgcaacca ggccatcagc cgcctggagg 2340
gcctgagcaa cctgtacca aatctacgcg agagcttccg cgagtgggag gccgaccca 2400
ccaacccccc cctgcgcgag gagatgcgca tccagttcaa cgacatgaac agcgccctga 2460
ccaccgccat cccctgttc gccgtgcaga actaccaggt gcccctgctg agcgtgtacg 2520
tgcaggccgc caacctgcac ctgagcgtgc tgcgcgacgt cagcgtgttc ggccagcgt 2580
ggggcttga cgcgcacc atcaacagcc gctacaacga cctgaccgc ctgatcgga 2640
actacacga ccacgcgtg cgtggtaca acacggcct ggagcgcgtg tggggtcccg 2700
acagccgca ctggatcagg tacaaccagt tccgcgcga gctgaccctg agcgtgctg 2760
acatcgtgag cctgttcccc aactacgaca gccgcacct ccccatccgc accgtgagcc 2820
agctgacccg cgagatttac accaaccctg tctggagaa cttcgacggc agcttccgcg 2880
gcagcgcaca gggcatcgag ggcagcatcc gcagcccca cctgatggac atcctgaaca 2940
gcacaccat ctacaccgac gccaccgcg gcgagtacta ctggagcggc caccagatca 3000
tgccagccc cgtcggcttc agcggccccg agttcacctt cccctgtac ggcaccatgg 3060
gcaacgtgc acctcagcag cgcctgttg cacagctggg ccaggagtg taccgcacc 3120
tgagcagca cctgtaccgt cgaccttca acatcggcat caacaaccag cagctgagc 3180
tgcgtgagc caccagttc gctacggca ccagcagcaa cctgccagc gccgtgtacc 3240
gcaagagcgg caccgtggac agcctggac agatccccc tcagaacaac aacgtgccac 3300
ctcgacagg cttcagccac cgtctgagcc acgtgagcat gttccgcagt ggcttcagca 3360
acagcagcgt gagcatcatc cgtgcacca tgttcagctg gattcaccgc agcgcaccc 3420
tgaccaaac catcgacccc gagcgcata accagatccc cctggtgaag ggcttccggg 3480
tgtggggcgg caccagcgtg atcaccggcc cggcttcac cggagcgac atcctgcga 3540
gaaacacctt cggcgacttc gtgagcctgc aggtgaacat caacagccc atcaccagc 3600
gttaccgct gcgcttccgc tacgccagc gccgcagcc cctgtgtatc tgcgtactg 3660
gcgcgctag caccggtgtg ggcggtcagg tgagcgtgaa catgccctg cagaagacta 3720
tgagatcgg cgagaacctg actagtgcga cttccgcta caccgacttc agcaacccct 3780
tcagcttccg cgccaacccc gacatcatcg gcatcagcga gcagccctg ttccgtgccg 3840
gcagcatcag cagcggcgag ctgtacatcg acaagatcga gatcatcctg gccagccca 3900
ccttcgagc cgagagcgac ctggagcgc cccagaaggc cgtgaacgcc ctgttcacca 3960
gcagcaacca gatcggcctg aagaccgacg tgaccgacta ccacatcgac caggtgagca 4020
acctggtgga ctgcttaagc gacgagttct gcctggacga gaagaaggag ctgagcgaa 4080
aggtgaagca cgccaagcgc ctgagcgcag agcgcaacct gctgcaggac cccaacttcc 4140
gcggcatcaa ccgcagctg gaccgcgct ggcgaggcag caccgatata accatccagg 4200
gcggcgacga cgtgttcaag gagaactacg tgacctgca gggcaccttc gacgagtgct 4260
acccaccta cctgtaccag ccgatcgacg agagcaagct gaaggcctac acccgctacc 4320
agctgcgcgg ctacatcgag gacagccagg acctggaaat ctacctgatc cgctacaacg 4380

```

```

cgaagcacga gaccgtgaac gtgcccggca cgggcagcct gtggcccccg agcgccccca 4440
gccccatcgg caagtgcggg gagccgaatc gatgcgctcc gcacctggag tggaaacccg 4500
acctagactg cagctgcagg gacggggaga agtgcgcccc ccacagccac cacttcagcc 4560
tggacatcga cgtgggctgc accgacctga acgaggacct gggcgtgtgg gtgatcttca 4620
agatcaagac ccaggacggc caccgcccgc tgggcaatct agagtctctg gaggagaagc 4680
ccctggtggg cgaggccctg gcccgcgtga agcgtgctga gaagaagtgg cgcgacaagc 4740
gcgagaagct ggagtgggag accaaccatcg tgtacaagga ggccaaggag agcgtggacg 4800
ccctgttctg gaacagccag tacgaccgcc tgcaggccga caccaacatc gccatgatcc 4860
acgcgcccga caagcgcgtg cacagcattc gcgaggccta cctgcccagag ctgagcgtga 4920
tccccggtgt gaacgcccgc atcttcgagg aactcgaggg ccgcatcttc accgccttca 4980
gcctgtacga cgcgcccaac gtgatcaaga acggcgactt caacaacggc ctgagctgct 5040
ggaacgtgaa gggccacgtg gacgtggagg agcagaacaa ccaccgcagc gtgctggtgg 5100
tgcccagagt ggaggccgag gtgagccagg aggtgcgcgt gtgccccggc cgcggtaca 5160
tcctgcgcgt gaccgcctac aaggagggtc acggcgaggg ctgctgacc atccacgaga 5220
tcgagaacaa caccgacgag ctcaagtcca gcaactgcgt ggaggaggag gtttacccca 5280
acaacacagt gacctgaac gactacaccg cgaccagga ggagtacgaa ggcacctaca 5340
cctctcgcaa caggggttac gacggcgctc acgagtccaa cagctccgtg ccagctgact 5400
acgccagcgc ccacgaggag aaagcctaca ccgacggtag acgcgacaac ccagtgtgaga 5460
gcaacagagg ctacggcgac tacacccccg tgcccgtggg atacgtgacc aaggagctgg 5520
agtacttccc cgagaccgac aagggtgtga tcgagattgg cgagaccgag ggcaccttca 5580
tcgtggacag cgtggagctg ctgctgatgg aggagtagta gatctgttct gcacaaagtg 5640
gagtagtcag tcctcgatca ggaaccagac accagacttt tattcataca gtgaagtgaa 5700
gtgaagtgca gtgcagtga gtgtgtgttt ttgtaccact tagtatgtat ttgtatttgt 5760
aaaatacttc tatcaataaa atttctaat ctaaaacca aaatccagtg ggtaccagct 5820
tgcatgectg cagtgcagcg tgaccgggtc gtgcccctct ctagagataa tgagcattgc 5880
atgtctaaat tataaaaaat taccacatat tttttttgtc acacttgttt gaagtgcagt 5940
ttatctatct ttatacatat atttaaactt tactctacga ataataaat ctatagtact 6000
acaataatat cagtgtttta gagaatcata taaatgaaca gttagacatg gtctaaagga 6060
caattgagta ttttgacaac aggactctac agttttatct ttttagtgtg catgtgttct 6120
cctttttttt tgcaaatagc ttacacctata taatacttca tccattttat tagtacatcc 6180
atttagggtt taggggttaat ggtttttata gactaatttt ttttagtacct ctatttttatt 6240
ctattttagc ctctaaatta agaaaactaa aactctattt tagttttttt atttaataat 6300
ttagatataa aatagaataa aataaagtga ctaaaaatta aacaaatacc ctttaagaaa 6360
ttaaaaaaac taaggaaaca tttttcttgt ttcgagttaga taatgccagc ctgttaaacy 6420
ccgtcgacga gtctaaccga caccaaccag cgaaccagca gcgtcgcgtc gggccaagcg 6480
aagcagacgg caccgcatct ctgtcgctgc ctctggaccc ctctcgagag ttccgctcca 6540
ccgttggaact tgctccgctg tcggcatcca gaaattgcgt ggcggagcgg cagacgtgag 6600
ccggcacggc aggcggcctc ctccctctct caccgcacgg cagctacggg ggattccctt 6660
cccaccgctc ctctcgcttc ccttccctgc ccgccgtaat aaatagacac cccctccaca 6720
ccctctttcc ccaacctcgt gttgttcgga gcgcacacac acacaaccag atctccccc 6780
aatccaccgg tcggcacctc cgcttcaagg taccgcgctc gtcctcccc cccccccctc 6840
tctaccttct ctgacgcggc gttccggtcc atggttaggg cccggtagtt ctacttctgt 6900
tcattgttgt gttagatccg tgtttgtgtt agatccgtgc tgctagcgtt cgtacacgga 6960
tgcgacctgt acgtcagaca cgttctgatt gctaacttgc cagtgtttct ctttggggaa 7020
tcctgggatg gctctagccg ttccgcagac gggatcgatt tcattgattt ttttgtttcg 7080
ttgcataggg tttggtttgc ccttttccct tatttcaata tatgccgtgc acttgtttgt 7140
cgggtcatct tttcatgctt ttttttgtct tggttgtgat gatgtggtct ggttgggagg 7200
tcgttctaga tcggagtaga attctgtttc aaactacctg gtggatttat taattttgga 7260
tctgtatgtg tgtgccatac atattcatag tatcgaattg aagatgatgg atggaaatat 7320
cgatctagga taggtataca tgttgatgcg ggttttactg atgcatatac agagatgctt 7380
tttgttcgct tggttgtgat gatgtggtgt ggttgggagg tcgttcattc gttctagatc 7440
ggagttagac ccgtttcaaa ctacctggtg tatttattaa ttttggaact gtatgtgtgt 7500
gtcatacatc ttcatagtta cgagttaaag atggatggaa atatcgatct aggataggta 7560
tacatgttga tgtgggtttt actgatgcat atacatgatg gcattatgcag catctattca 7620
tatgctctaa ccttgagtac ctatctatta taataaacia gtatgtttta taattatttt 7680
gatcttgata tacttggtat atggcatatg cagcagctat atgtggattt ttttagccct 7740
gccttcatac gctattttat tgcttggtac tgttttttt gtcgatgctc accctgttgt 7800
ttggtgttac ttctgcaggg atccccgac atcgaaaaac tcattaaactc agtgcaaac 7860
tatgcctggg gcagcaaaac gcgttgactg aactttatgg tatggaaaat ccgtccagcc 7920
agccgatggc cgagctgtgg atgggcgcac atccgaaaag cagttcacga gtgcagaatg 7980
ccgccggaga tatcgtttca ctgcgtgatg tgattgagag tgataaatcg actctgctcg 8040

```

gagaggccgt	tgccaaacgc	tttggcgaac	tgcctttcct	gttcaaagta	ttatgcgcag	8100
cacagccact	ctccattcag	gttcatccaa	acaaacacaa	ttctgaaatc	ggttttgcc	8160
aagaaaatgc	cgcaggtatc	ccgatggatg	ccgccgagcg	taactataaa	gatacctaacc	8220
acaagccgga	gctgggtttt	gcgctgacgc	ctttccttgc	gatgaacgcg	tttcgtgaat	8280
tttccgagat	tgtctcccta	ctccagccgg	tgcaggtgc	acatccggcg	attgctcact	8340
ttttacaaca	gcctgatgcc	gaacgtttta	gcgaactgtt	cgccagcctg	ttgaatatgc	8400
agggtgaaga	aaaatcccgc	gcgctggcga	ttttaaaatc	ggccctcgat	agccagcagg	8460
gtgaaccgtg	gcaaacgatt	cgtttaattt	ctgaatttta	cccgaagac	agcggctctgt	8520
tctcccgcct	attgctgaat	gtggtgaaat	tgaaccctgg	cgaagcgatg	ttcctgttcg	8580
ctgaaacacc	gcacgcttac	ctgcaaggcg	tggcgctgga	agtgatggca	aactccgata	8640
acgtgctgcg	tgcgggtctg	acgcctaaat	acattgatat	tccggaactg	gttgccaatg	8700
tgaaattcga	agccaaaccg	gctaaccagt	tgttgaccca	gccggtgaaa	caaggtgcag	8760
aactggactt	cccgatccca	gtggatgatt	ttgccttctc	gctgcatgac	cttagtgata	8820
aagaaaccac	cattagccag	cagagtgcgc	ccattttgtt	ctgcgtcgaa	ggcgatgcaa	8880
cgttgtggaa	aggttctcag	cagttacagc	ttaaaccggg	tgaatcagcg	tttattgccg	8940
ccaacgaatc	accggtgact	gtcaaaggcc	acggccgttt	agcgcgtgtt	tacaacaagc	9000
tgtaagagct	tactgaaaaa	attaacatct	cttgctaagc	tgggagctcg	atccgtcgac	9060
ctgcagagat	cgttcaaaca	tttggcaata	aagtttctta	agattgaatc	ctggtgccgg	9120
tcttgcgatg	attatcatct	aattttctgtt	gaattacgtt	aagcatgtaa	taattaacat	9180
gtaatgcatg	acgttattta	tgagatgggt	ttttatgatt	agagtcccg	aattatacat	9240
ttaatacgcg	atagaaaaca	aaatatagcg	cgcaaactag	gataaattat	cgcgcgcggt	9300
gtcatctatg	ttactagatc	cgatgataag	ctgtcaaaca	tgagatcccc	gggtctagac	9360
aattcagtac	attaaaaacg	tccgcaatgt	gttattaagt	tgtctaagcg	tcaatttgtt	9420
tacaccacaa	tatatcctgc	caccagccag	ccaacagctc	cccgaaccgg	agctcggcac	9480
aaaatcacca	ctcgatacag	gcagcccatc	agtcggggac	ggcgctcagc	ggagagccgt	9540
tgtaaggcgg	cagactttgc	tcatgttacc	gatgctattc	ggaagaacgg	caactaagct	9600
gccgggtttg	aaacacggat	gatctcgcgg	agggtagcat	gttgattgta	acgatgacag	9660
agcgttgctg	cctgtgatca	aatatcatct	ccctcgcaga	gatccgaatt	atcagccttc	9720
ttattcattt	ctcgcttaac	cgtagacagg	tgtcgatctt	gagaactatg	ccgacataat	9780
aggaaatcgc	tggataaaag	cgctgaggaa	gctgagtggc	gctatttctt	tagaagtgaa	9840
cgttgacgat	cgtcgaccgt	accccgatga	attaattcgg	acgtacgttc	tgaacacagc	9900
tggatactta	cttggggcgat	tgtcatacat	gacatcaaca	atgtaccctg	ttgtgtaacc	9960
gtctcttgga	ggttcgtatg	acactagtgg	ttccctcag	cttgcgacta	gatgttgagg	10020
cctaacattt	tattagagag	caggctagt	gcttagatac	atgatcttca	ggcgtttatc	10080
tgtcagggca	agcgaataat	ggccatttat	gacgaccaat	gccccgcaga	agctcccatc	10140
tttgcgcgca	tagacgccgc	gcccccttt	tggggtgtag	aacatccttt	tgccagatgt	10200
ggaaaagaag	ttcgttgtcc	cattgttggc	aatgacgtag	tagccggcga	aagtgcgaga	10260
cccatttgcg	ctatatataa	gcctacgatt	tccgttgcga	ctattgtcgt	aattggatga	10320
actattatcg	tagttgctct	cagagttgtc	gtaatttgat	ggactattgt	cgtaattgct	10380
tatggagtgt	tcgtagtgtc	ttggagaaa	gtcgtagtgt	gatggggagt	agtcataggg	10440
aagacgagct	tcattccacta	aaacaattgg	caggtcagca	agtgcctgcc	ccgactctct	10500
cgcaagtacg	aggcttagaa	ccaccttcaa	cagatcgcgc	atagtcttcc	ccagctctct	10560
aacgcttgag	ttaagccgcg	ccgcgaagcg	gcgtcggctt	gaacgaattg	ttagacatta	10620
tttgcgcgact	accttggtga	tctcgccctt	cacgtagtga	acaaattctt	ccaactgatc	10680
tgcgcgcgag	gccaagcgat	cttcttgtcc	aagataagcc	tgcctagctt	caagtatgac	10740
gggctgatac	tgggcgggca	ggcgctccat	tgccagtcg	gcagcgacat	ccttcggcgc	10800
gattttgccg	gttactgcgc	tgtaccaa	gcgggacaac	gtaagcacta	catttcgcctc	10860
atgccagcc	cagtcggggc	gcgagttcca	tacggttaag	gtttcattta	cgcctcaaa	10920
tagatcctgt	tcaggaaccg	gatcaaagag	ttcctccgcc	gctggaccta	ccaaggcaac	10980
gctatgttct	cttgcttttg	tcagcaagat	agccagatca	atgtcgatcg	tggctggctc	11040
gaagatacct	gcaagaatgt	cattgcgctg	ccattctcca	aattgcagtt	cgcgcttagc	11100
tggataacgc	cacggaatga	tgtcgtcgtg	cacaacaatg	gtgacttcta	cagcgcggag	11160
aatctcgcct	tctccagggg	aagccgaagt	ttccaaaagg	tcgttgatca	aagctcgccg	11220
cgttggttca	tcaagcctta	cggtcaccgt	aaccagcaaa	tcaatatcac	tgtgtggctt	11280
caggccgcca	tccactgcgg	agccgtacaa	atgtacggcc	agcaacgctg	gttcgagatg	11340
gcgctcgatg	acgccaaacta	cctctgatag	ttgagtcgat	acttcggcga	tcaccgcttc	11400
cctcatgatg	tttaactcct	gaattaagcc	gcgcgcgaa	gcggtgtcgg	cttgaatgaa	11460
ttgttagggc	tcactctgtg	ctcccagaa	ccagtaaccag	tacatcgctg	ttcgttcga	11520
gacttgaggt	ctagctttat	acgtgaacag	gtcaatgccg	ccgagagtaa	agccacattt	11580
tgcgtacaaa	ttgcaggcag	gtacattgtt	cgtttgtgtc	tctaatacgta	tgccaaggag	11640
ctgtctgcct	agtgccact	ttttcgcaaa	ttcgatgaga	ctgtgcgcga	ctcctttgcc	11700

tccgtgctgtg	tgcgacacaa	caatgtgttc	gatagaggct	agatcgttcc	atgttgagtt	11760
gagttcaatc	ttcccgacaa	gctcttggtc	gatgaatgcg	ccatagcaag	cagagtcctc	11820
atcagagtc	tcacccgaga	tgtaatcctt	ccggtagggg	ctcacacttc	tggtagatag	11880
ttcaaagcct	tggtcggata	ggtgcacatc	gaacacttca	cgaacaatga	aatgggtctc	11940
agcatccaat	gtttccgcc	cctgctcagg	gatcacccga	atcttcata	gacgcctaac	12000
gcctggcaca	gcggatcgca	aacctggcgc	ggcttttggc	acaaaaggcg	tgacagggtt	12060
gcgaatccgt	tgttgcact	tgttaacct	tttgccagat	ttggtaacta	taatttatgt	12120
tagaggcgaa	gtcttggtta	aaaactggcc	taaaattgct	ggggatttca	ggaaagtaaa	12180
catcaccttc	cggctcgatg	tctattgtag	atatattgt	tgtatctact	tgatcggggg	12240
atctgctgcc	tgcgcggtt	cggatgatgac	ggtgaaaacc	tctgacacat	gcagctcccg	12300
gagacgggtc	cagcttggtc	gtaagcggat	gccgggagca	gacaagcccg	tcagggcgcg	12360
tcagcgggtg	ttggcgggtg	tcggggcgca	gccatgaccc	agtcacgtag	cgatagcgga	12420
gtgtatactg	gcttaactat	gcggcatcag	agcagattgt	actgagagtg	caccatatgc	12480
ggtgtgaaat	accgcacaga	tcgttaagga	gaaaataccg	catcaggcgc	tcttcgcctt	12540
cctcgctcac	tgactcgctg	cgtcgggtcg	ttcggctgcg	gcgagcggta	tcagctcact	12600
caaaggcgg	aatacggtta	tcacagaaat	caggggataa	cgcaggaaag	aacatgtgag	12660
caaaaggcca	gcaaaaggcc	aggaaccgta	aaaaggccgc	gttgcctggc	tttttccata	12720
ggctccgcgc	ccctgacgag	catcacaaaa	atcgacgtc	aagtcagagg	tggcgaaacc	12780
cgacaggact	ataaagatac	caggcgtttc	cccttggaag	ctccctcgtg	cgtctcctcg	12840
ttccgacct	gcgcgttacc	ggataacctgt	ccgcctttct	cccttcggga	agcgtggcgc	12900
tttctcatag	ctcacgctgt	aggtatctca	gttcgggtgta	ggtcgttcgc	tccaagctgg	12960
gctgtgtgca	cgaaccccc	gttcagcccg	accgctgcgc	cttatccgg	aactatcgct	13020
ttgagtccaa	cccggttaaga	cacgacttat	cgcactggc	agcagccact	ggtaacagga	13080
ttagcagagc	gaggtatgta	ggcgggtgta	cagagttctt	gaagtgggtg	cctaactacg	13140
gctacactag	aaggacagta	tttggtatct	gcgctctgct	gaagccagtt	accttcggaa	13200
aaagagttgg	tagctcttga	tcgggcaaac	aaaccaccgc	tggtagcgg	gggttttttg	13260
tttgcaagca	gcagattacg	cgcagaaaaa	aaggatctca	agaagatcct	ttgatctttt	13320
ctacggggtc	tgacgctcag	tggaacgaaa	actcacgtta	agggattttg	gtcatgagat	13380
tatcaaaaag	gatcttcacc	tagatccttt	taaattaaaa	atgaagtttt	aaatcaatct	13440
aaagtatata	tgagtaaact	tggtctgaca	gttaccaatg	cttaatcagt	gaggcaccta	13500
tctcagcgat	ctgtctattt	cgttcaccca	tagttgcctg	actccccgtc	gtgtagataa	13560
ctacgatacg	ggagggttta	ccatctggcc	ccagtgcctg	aatgataccg	cgagaccac	13620
gctcacccgc	tcagatttta	tcagcaataa	accagccagc	cggaaaggcc	gagcgcagaa	13680
gtggctcctg	aaactttacc	gcctccatcc	ttgttgccgg	gaagctagag	tggttgccgg	13740
taagttagttc	gccagttaat	agtttgcgca	acgttggttg	cattgctgca	gggggggggg	13800
ggggggggga	cttccattgt	tcattccacg	gacaaaaaca	gagaaaggaa	acgacagagg	13860
ccaaaaagcc	tcgcttttcag	cacctgtcgt	ttcctttctt	ttcagagggt	attttaataa	13920
aaaacattaa	gttatgacga	agaagaacgg	aaacgcctta	aaccggaaaa	ttttcataaa	13980
tagcgaaaac	ccgcgaggtc	gccgccccgt	aacctgtcgg	atcacccgaa	aggacccgta	14040
aagtgataat	gattatcacc	tacatatcac	aacgtgcgtg	gaggccatca	aaccacgtca	14100
aataatcaat	tatgacgcag	gtatcgatct	aattgatctg	catcaactta	acgtaaaaac	14160
aacttcagac	aatacaaatc	agcgacactg	aatacggggc	aacctcatgt	cccccccccc	14220
ccccccccctg	caggcatcgt	ggtgtcacgc	tcgtcgtttg	gtatggcttc	attcagctcc	14280
ggttcccaac	gatcaaggcg	agttacatga	ttccccatgt	tgtgcaaaaa	agcgggttagc	14340
tccttcgggtc	ctccgatcgt	tgtcagaagt	aagttggccg	cagtgttatc	actcatgggt	14400
atggcagcac	tgcataattc	tcttactgtc	atgccatccg	taagatgctt	ttctgtgact	14460
ggtgagtact	caaccaagtc	attctgagaa	tagtgtatgc	ggcgaccgag	ttgctcttgc	14520
ccggcgtcaa	cacgggataa	taccgcgcga	catagcagaa	ctttaaaagt	gtccatcatt	14580
ggaaaacggt	cttcggggcg	aaaactctca	aggactttac	cgtgtttgag	atccagttcg	14640
atgtaaccga	ctcgtgcacc	caactgatct	tcagcatctt	ttactttcac	cagcgtttct	14700
gggtgagcaa	aaacagggaag	gcaaaatgcc	gcaaaaaagg	gaataagggc	gacacggaaa	14760
tggtgaatac	tcatactctt	cctttttcaa	tattattgaa	gcatttatca	gggttattgt	14820
ctcatgagcg	gatacatatt	tgaatgtatt	tagaaaaata	aacaaatagg	ggttcgcgc	14880
acatttcccc	gaaaagtgcc	acctgacgtc	taagaaacca	ttattatcat	gacattaacc	14940
tataaaaaata	ggcgtatcac	gaggcccttt	cgtcttcaag	aattgggtcga	cgatcttgc	15000
gcgttcggat	attttcgtgg	agttcccgcc	acagaccggg	attgaaggcg	agatccagca	15060
actcgcgcca	gatcatectg	tgacggaact	ttggcgcgtg	atgactggcc	aggacgtcgg	15120
ccgaaagagc	gacaagcaga	tcacgctttt	cgacagcgtc	ggatttgca	tcgaggattt	15180
ttcggcgctg	cgtacgcctc	gcgaccgctg	tgagggatca	agccacagca	ccccactcga	15240
ccttctagcc	gacccagacg	agccaaggga	tctttttgga	atgctgctcc	gtcgtcaggc	15300
tttccgacgt	ttgggtggtt	gaacagaagt	cattatcgta	cggaatgcca	agcactcccg	15360

```

agggggaaccc tgtgggttggc atgcacatac aaatggacga acggataaac ctttttcacgc 15420
cctttttaaat atccgttatt ctaataaacg ctctttttctc ttaggtttac ccgccaatat 15480
atcctgtcaa aactgatag tttaaactga aggcgggaaa cgacaatctg atcatgagcg 15540
gagaattaag ggagtcacgt tatgaccccc gccgatgacg cgggacaagc cgtttttacgt 15600
ttggaactga cagaaccgca acgttgaagg agccactcag ccc 15643

```

<210> 14

<211> 15503

<212> DNA

<213> 人工序列

<220>

<223> 人工序列说明: pNOV1305

<220>

<221> misc_feature

<222> (1)..(3582)

<223> 编码 H04 毒素部分加全长 Cry1Ab 尾部分的合成核苷酸序列

<220>

<221> misc_feature

<222> (3790)..(5771)

<223> Zm Ubi 启动子

<220>

<221> misc_feature

<222> (5868)..(6971)

<223> PMI

<220>

<221> misc_feature

<222> (12934)..(15494)

<223> MTL 启动子

<400> 14

```

atggacaaca accccaacat caacgagtgc atcccctaca actgcctgag caacccccgag 60
gtggagggtgc tgggcggcga gcgcacgcag accggctaca ccccatcga catcagcctg 120
agcctgaccc agttcctgct gagcgagttc gtgcccgcg cgggcttcgt gctgggcctg 180
gtggacatca tctggggcat cttcggcccc agccagtggg acgccttcct ggtgcagatc 240
gagcagttga taaaccaacg catagaggaa ttcgcccgca accaggccat cagccgcctg 300
gagggcctga gcaacctgta ccaaactctac gccgagagct tccgcgagtg ggaggccgac 360
cccaccaacc ccgccttgcg cgaggagatg cgcattccagt tcaacgacat gaacagcgcc 420
ctgaccaccg ccatccccct gttcgccgtg cagaactacc aggtgccctt gctgagcgtg 480
tacgtgcagg ccgccaacct gcacctgagc gtgctgcgcg acgtcagcgt gttcggccag 540
cgctggggct tcgacgccgc caccatcaac agccgctaca acgacctgac ccgcctgatc 600
ggcaactaca ccgaccacgc cgtgcgctgg tacaacaccg gcctggagcg cgtgtggggg 660
cccgacagcc gcgactggat caggtacaac cagttccgcc gcgagctgac cctgaccgtg 720
ctggacatcg tgagcctgtt ccccaactac gacagccgca cctaccccat ccgcaccgtg 780
agccagctga ccgcgagat ttacaccaac ccgctgctgg agaacttcga cggcagcttc 840
cgcggcagcg ccaggggcat cgagggcagc atccgcagcc cccacctgat ggacatcctg 900
aacagcatca ccatctacac cgacgcccac cgcggcgagt actactggag cggccaccag 960
atcatggcca gcccgcctcg cttcagcggc ccgagttca ccttccccct gtacggcacc 1020
atgggcaacg ctgcacctca gcagcgcac gtggcacagc tgggccaggg agtgtaccgc 1080
accctgagca gcacctgtga ccgtcgacct ttcaacatcg gcatcaacaa ccagcagctg 1140
agcgtgctgg acggcaaccga gttcgcctac ggcaccagca gcaacctgcc cagcgcctg 1200
taccgcaaga gcggcaaccg ggacagcctg gacgagatcc cccctcagaa caacaacgtg 1260
ccacctcgac agggcttcag ccacctctg agccacgtga gcatgttcgg cagtggcttc 1320
agcaacagca gcgtgagcat catccgtgca cccatgttca gctggattca ccgcagcgcc 1380

```

```

accctgacca acaccatcga ccccgagcgc atcaaccaga tccccctggt gaagggcttc 1440
cggtgtggg ggggaccag cgtgatcacc ggccccggct tcaccggagg cgacatcctg 1500
cgcagaaaca ccttcggcga cttcgtgagc ctgcaggtga acatcaacag ccccatcacc 1560
cagcgttacc gcctgcgctt ccgctacgcc agcagccgcg acgcccgtgt gatcgtgctg 1620
actggcgccg ctagcacccg tgtgggcggt caggtgagcg tgaacatgcc cctgcagaag 1680
actatggaga tcggcgagaa cctgactagt cgcaccttc gctacaccga cttcagcaac 1740
cccttcagct tcgcgcgcaa ccccgacatc atcggcatca gcgagcagcc cctgttcggt 1800
gccggcagca tcagcagcgg cgagctgtac atcgacaaga tcgagatcat cctggccgac 1860
gccaccttcg agggcgagag cgacctggag cgcgccaga aggccgtgaa cgcctgttc 1920
accagcagca accagatcgg cctgaagacc gacgtgaccg actaccacat cgaccaggtg 1980
agcaacctgg tggactgctt aagcgacgag ttctgectgg acgagaagaa ggagctgagc 2040
gagaagggtg agcacgccaa gcgcctgagc gacgagcgca acctgctgca ggaccccaac 2100
ttccgcggca tcaaccgcca gctggaccgc ggctggcgag gcagcaccga tatcaccatc 2160
cagggcgccg acgacgtgtt caaggagaac tacgtgacct tgcagggcac cttcgacgag 2220
tgctacccca cctacctgta ccagccgac gacgagagca agctgaaggc ctacaccgc 2280
taccagctgc gcggctacat cgaggacagc caggacctgg aaatctacct gatecgctac 2340
aacgcgaagc acgagaccgt gaacgtgccc ggcaccggca gcctgtggcc cctgagcgcc 2400
cccagcccca tcggcaagtg cggggagccg aatcgatgcg ctccgcacct ggagtggaa 2460
ccggacctag actgcagctg cagggacggg gagaagtgcg cccaccacag ccaccattc 2520
agcctggaca tcgacgtggg ctgcaccgac ctgaacgagg acctgggcgt gtgggtgatc 2580
ttcaagatca agaccagga cggccacgcc cgctgggca atctagagtt cctggaggag 2640
aagcccctgg tgggcgaggc cctggcccgc gtgaagcgtg ctgagaagaa gtggcgcgac 2700
aagcgcgaga agctggagtg ggagaccaac atcgtgtaca aggaggccaa ggagagcgtg 2760
gacgccctgt tcgtgaacag ccagtacgac cgctgcagg ccgacacca catcgccatg 2820
atccacgcgc ccgacaagcg cgtgcacagc attcgcgagg cctacctgcc cgagctgagc 2880
gtgatccccg gtgtgaacgc cggcatcttc gaggaactcg agggccgcac cttcaccgc 2940
ttcagcctg acgacgcccg caacgtgatc aagaacggcg acttcaacaa cggcctgagc 3000
tgctggaacg tgaagggcca cgtggacgtg gaggagcaga acaaccaccg cagcgtgctg 3060
gtggtgcccg agtgggaggc cgaggtgagc caggaggtgc gcgtgtgccc cggccgcggc 3120
tacatcctgc gcgtgaccgc ctacaaggag ggctacggcg agggctgctg gacctccac 3180
gagatcgaga acaacaccga cgagctcaag ttcagcaact gcgtggagga ggaggtttac 3240
cccaacaaca ccgtgacctg caacgactac accgcgacc aggaggagta cgaaggcacc 3300
tacacctctc gcaacagggg ttacgacggc gcctacgagt ccaacagctc cgtgccagct 3360
gactagccca gcgcctacga ggagaaagcc tacaacgacg gtagacgcga caaccatgt 3420
gagagcaaca gaggctacgg cgactacacc cccctgcccg ctggatacgt gaccaaggag 3480
ctggagtact tccccgagac cgacaagggtg tggatcgaga ttggcgagac cgagggcacc 3540
ttcatcgtgg acagcgtgga gctgctgctg atggaggagt agtagatctg ttctgcacaa 3600
agtggagtag tcagtcatcg atcaggaacc agacaccaga cttttattca tacagtgaag 3660
tgaagtgaag tgcagtgcag tgagttgctg gttttgtac aacttagtat gtatttgtat 3720
ttgtaaaata cttctatcaa taaaatttct aattcctaaa accaaaatcc aggggtacca 3780
gcttgcatgc ctgcagtga gcgtgaccgc gtcctgcccc tctctagaga taatgagcat 3840
tgcatgtcta agttataaaa aattaccaca tatttttttt gtcacacttg tttgaagtgc 3900
agtttatcta tctttataca tatattttaa ctttactcta cgaataatat aatctatagt 3960
actacaataa tatcagtgtt ttagagaatc atataaatga acagttagac atggtctaaa 4020
ggacaattga gtattttgac aacaggactc tacagtttta tctttttagt gtgcatgtgt 4080
tctccttttt ttttgcaaat agcttcacct atataatact tcatccattt tattagtaca 4140
tccatttagg gtttaggggt aatggttttt atagactaat ttttttagta catctatttt 4200
attctatttt agcctctaaa ttaagaaaaa taaaactcta ttttagtttt tttatttaag 4260
aatttagata taaaatagaa taaaataaag tgactaaaaa ttaacaaat accctttaag 4320
aaattaaaaa aactaaggaa acatttttct tgttctgagt agataatgcc agcctgttaa 4380
acgcgctcga cgagtctaac ggacaccaac cagcgaacca gcagcgtcgc gtcgggcaa 4440
gcgaagcaga cggcacggca tctctgtcgc tgccctctgga cccctctcga gaggttccgt 4500
ccaccgttgg acttgctccg ctgtcggcat ccagaaattg cgtggcgagg cggcagacgt 4560
gagccggcac ggcaggcggc ctctcctctc tctcacggca cggcgagcta cgggggattc 4620
ctttccacc gctccttcgc tttcccttcc tcgcccgcgc taataaatag acacccctc 4680
cacacctct tttcccaacc tcgtgttgtt cggagcgcac acacacaaa ccagatctcc 4740
cccaatctca cccgtcggca cctccgcttc aaggtaacgc gctcgtctc cccccccc 4800
cctctctacc tctctagat cggcgttccg tccatgggtt agggcccggt agttctactt 4860
ctgttcatgt ttgtgttaga tccgtgtttg tgtagatcc gtgctgctag cgttcgtaca 4920
cggatgcgac ctgtacgtca gacacgttct gattgctaac ttgccagtg tttcttttgg 4980
ggaatcctgg gatggctcta gccgttccgc agacgggac gatttcatga tttttttgt 5040

```

```

ttcgttgcac aggggtttggt ttgccctttt cctttatttc aatatatgcc gtgcacttgt 5100
ttgtcgggtc atcttttcat gctttttttt gtcttggttg tgatgatgtg gtctgggttg 5160
gcggtcgttc tagatcggag tagaattctg tttcaaacta cctggtggat ttattaattt 5220
tggatctgta tgtgtgtgcc atacatatcc atagttacga attgaagatg atggatggaa 5280
atatcgatct aggataggta tacatgttga tgcgggtttt actgatgcac atacagagat 5340
gctttttgtt cgcttggttg tgatgatgtg gtgtgggttg gcggtcgttc attcgttcta 5400
gatcggagta gaatactgtt tcaaactacc tgggtgtatt attaattttg gaactgtatg 5460
tgtgtgtcat acatcttcat agttacgagt ttaagatgga tggaaatatc gatctaggat 5520
aggatatacat gttgatgtgg gttttactga tgcataatac tgatggcata tgcagcatct 5580
attcatatgc tctaaccctg agtacctatc tattataata aacaagtatg ttttataatt 5640
attttgatct tgatatactt ggatgatggc atatgcagca gctatatgtg gattttttta 5700
gccctgccct catagctatc ttatttgctt ggtactgttt cttttgtcga tgctcaccct 5760
gttgttttgt gttacttctg cagggatccc cgatcatgca aaaactcatt aactcagtgc 5820
aaaactatgc ctggggcagc aaaacggcgt tgactgaact ttatggtatg gaaaatccgt 5880
ccagccagcc gatggccgag ctgtggatgg gcgcacatcc gaaaagcagt tcacgagtgc 5940
agaatgccgc cggagatata gtttcaactg gtgatgtgat tgagagtgat aaatcgactc 6000
tgctcggaga ggccgttgcc aaacgctttg gcgaactgcc ttctctgttc aaagtattat 6060
gcgcagcaca gccactctcc attcaggttc atccaaacaa acacaattct gaaatcgggt 6120
ttgccaaaga aaatgccgca ggtatcccga tggatgccgc cgagcgtaac tataaagatc 6180
ctaaccacaa gccggagctg gtttttgccg tgacgccttt ccttgcatg aacgcgtttc 6240
gtgaattttc cgagattgtc tccctactcc agccggctgc aggtgcacat ccggcgattg 6300
ctcacttttt acaacagcct gatgccgaac gtttaagcga actgttcgcc agcctgttga 6360
atatgcaggg tgaagaaaaa tcccgcgcgc tggcgatttt aaaatcggcc ctcgatagcc 6420
agcaggggtg accgtggcaa acgattcgtt taatttctga attttaccgc gaagacagcg 6480
gtctgttctc cccgctattg ctgaatgtgg tgaattgaa cctggcgaa cgcgtattcc 6540
tgttcgctga aacacgcac gcttacctgc aaggcgtggc gctggaagtg atggcaaac 6600
ccgataacgt gctgcgtgcg ggtctgacgc ctaatacat tgatattccg gaactgggtg 6660
ccaatgtgaa attcgaagcc aaaccggcta accagtgtgt gacccagccg gtgaaacaag 6720
gtgcagaact ggacttcccg attccagtgg atgattttgc ctctcgcgtg catgacctta 6780
gtgataaaga aaccaccatt agccagcaga gtgccgccat tttgttctgc gtogaaggcg 6840
atgcaacggt gtggaaagggt tctcagcagt tacagcttaa accgggtgaa tcagcgttta 6900
ttgccgccaa cgaatcaccg gtgactgtca aaggccacgg ccgttttagcg cgtgtttaca 6960
acaagctgta agagcttact gaaaaaatta aagctctttg ctaagctggg agctcgatcc 7020
gtcgacctgc agatcgttca aacatttggc aataaagttt cttaagattg aatcctgttg 7080
ccggtcttgc gatgattatc atataatttc tgttgaatta cgtaagcat gtaataatta 7140
acatgtaatg catgacgtta tttatgagat gggtttttat gattagagtc ccgcaattat 7200
acatttaata cgcgatagaa aacaaaatat agcgcgcaaa ctaggataaa ttatcgcgcg 7260
cgggtgtcatc tatgttacta gatctgctag ccctgcagga aatttaccgc tgcccgggcg 7320
gccagcatgg ccgtatccgc aatgtgttat taagttgtct aagcgtcaat ttgtttacac 7380
cacaatatat cctgccacca gccagccaac agctccccga ccggcagctc ggcacaaaa 7440
caccactcga tacaggcagc ccatacgaat taattctcat gtttgacagc ttatcatcga 7500
ctgcacggtg caccaatgct tctggcgta ggcagccatc ggaagctgtg gtatggctgt 7560
gcaggctgta aatcactgca taattcgtgt cgtcaaggc gcaactccgt tctggataat 7620
gttttttgcg ccgacatcat aacggttctg gcaaatatc tgaatgagc tgttgacaat 7680
taatcatccg gctcgtataa tgtgtggaat tgtgagcgga taacaatttc acacaggaaa 7740
cagaccatga gggaagcggt gatcgccgaa gtatcgactc aactatcaga ggtagtgtgg 7800
gtcatcgagc gccatctcga accgacgttg ctggccgtac atttgtaagg ctccgcagtg 7860
gatggcgccg tgaagccaca cagtgatatt gatcttctgg ttacggtgac cgtaaagctt 7920
gatgaaacaa ccggcgagc tttgatcaac gaccttttgg aaacttcggc ttccccgtga 7980
gagagcgaga ttctccgcgc tgtagaagtc accattgttg tgcacgacga catcattccg 8040
tggcggttatc cagctaaggc cgaactgcaa tttggagaat ggcagcgcaa tgacattctt 8100
gcaggatatc tcgagccagc cagatcgac attgatctgg ctatcttgct gacaaaagca 8160
agagaacata gcgttgccct ggtaggtcca gcggcgagg aactctttga tccggttcc 8220
gaacaggatc tatttgaggc gctaaatgaa accttaacgc tatggaactc gccgccgac 8280
tgggctggcg atgagcgaaa tgtagtgtt acgttgtccc gcatttggtg cagcgcagta 8340
accggcaaaa tcgcgccgaa ggatgtcgct gccgactggg caatggagcg cctgccggcc 8400
cagtaacgag ccgtcatact tgaagctagg caggcttacc ttggacaaga agatcgcttg 8460
gcctcgcgcg cagatcagtt ggaagaattt gttcactacg tgaaaggcga gatcacaaa 8520
gtagtgcgca aataaagctc tagtggatct ccgtaccccc gggggatctg gctcgcggcg 8580
gacgcacgac gccggggcga gaccataggc gatctcctaa atcaatagta gctgtaacct 8640
cgaagcggtt cacttgtaac aacgattgag aatttttgtc ataaaattga aatacttggt 8700

```

tcgcattttt	gtcatccgcg	gtcagccgca	attctgacga	actgcccatt	tagctggaga	8760
tgattgtaca	tccttcacgt	gaaaattttc	caagcgctgt	gaacaagggg	tcagatttta	8820
gattgaaagg	tgagccgttg	aaacacgttc	ttcttgtcga	tgacgacgtc	gctatgcggc	8880
atcttattat	tgaatacctt	acgatccacg	ccttcaaagt	gaccgcggta	gccgacagca	8940
cccagttcac	aagagtactc	tcttcgcgga	cggtcgatgt	cgtggttggt	gatctaaatt	9000
taggtcgtga	agatgggctc	gagatcggtc	gtaatctggc	ggcaaagtct	gatattccaa	9060
tcataattat	cagtggcgac	cgcttgagg	agacggataa	agttgttgca	ctcgagctag	9120
gagcaagtga	ttttatcgct	aagccgttca	gtatcagaga	gtttctagca	cgcattcggg	9180
ttgccttgcg	cgtgcgcccc	aacgttggtc	gctccaaaaga	ccgacgggtc	ttttgtttta	9240
ctgactggac	acttaatctc	aggcaacgtc	gcttgatgtc	cgaagctggc	ggtgaggtga	9300
aacttacggc	agggtgagtc	aatctttctc	tcgcgttttt	agagaaaccc	cgcgacgttc	9360
tatcgcgga	gcaacttctc	attgccagtc	gagtacgga	cgaggagggt	tatgacagga	9420
gtatagatgt	tctcattttg	aggctgcgcc	gcaaacttga	ggcagatccg	tcaagccctc	9480
aactgataaa	aacagcaaga	ggtgccgggt	atttctttga	cgcggacgtg	caggttttcg	9540
acggggggac	gatggcagcc	tgagccaatt	cccagatccc	cgaggaatcg	gcgtgagcgg	9600
tcgcaaacca	tccggcccg	tacaaatcgg	cgcggcgctg	ggtgatgacc	tggtggagaa	9660
gttgaaggcc	gcgcaggccg	cccagcgga	acgcatecga	gcagaagcac	gccccggtga	9720
atcgtggcaa	gcggccgctg	atcgaatccg	caaagaatcc	cggcaaccgc	cggcagccgg	9780
tcgcccgtcg	attaggaagc	cgcccaaggg	cgacgagcaa	ccagattttt	tcgttccgat	9840
gctctatgac	gtgggcaccc	gcgatagtcg	cagcatcatg	gacgtggccg	ttttccgtct	9900
gtcgaagcgt	gaccgacgag	ctggcgaggt	gatccgctac	gagcttccag	acgggcacgt	9960
agaggtttcc	gcagggccgg	ccggcatggc	cagtgtgtgg	gattacgacc	tggtactgat	10020
ggcggtttcc	catctaaccg	aatccatgaa	ccgataaccg	gaagggaagg	gagacaagcc	10080
cgcccgctg	ttccgtccac	acgttgcgga	cgtactcaag	ttctgccggc	gagccgatgg	10140
cggaaagcag	aaagacgacc	tggtagaaac	ctgcattcgg	ttaaacacca	cgcccggtgc	10200
catgcaagcg	acgaagaagg	ccaagaacgg	ccgcctgggtg	acggtatccg	aggggtgaagc	10260
cttgattagc	cgctacaaga	tcgtaaagag	cgaaaccggg	cggccggagt	acatcgagat	10320
cgagctagct	gattggatgt	accgcgagat	cacagaaggc	aagaacccgg	acgtgctgac	10380
ggttcacccc	gattactttt	tgatcgatcc	cggcatcggc	cgttttctct	accgcctggc	10440
acgcgcgccc	gcaggcaagg	cagaagccag	atggttgttc	aagacgatct	acgaacgcag	10500
tggcagcgcc	ggagagttca	agaagttctg	tttcaccgtg	cgcaagctga	tcgggtcaaa	10560
tgacctgccg	gagtacgatt	tgaaggagga	ggcggggcag	gctggcccga	tcctagtcat	10620
gcgctaccgc	aacctgacgc	agggcgaaag	atccgcgggt	tcctaagtga	cggagcagat	10680
gctagggcaa	attgccctag	caggggaaaa	aggtcgaaaa	ggtctctttc	ctgtggatag	10740
cacgtacatt	gggaacccaa	agccgtacat	tgggaaccgg	aaccctgaca	ttgggaaccc	10800
aaagccgtac	attgggaacc	ggtcacacat	gtaagtgact	gatataaaaag	agaaaaaagg	10860
cgatttttcc	gcctaaaact	ctttaaaact	tattaaaact	cttaaaaccc	gcctggcctg	10920
tgcataactg	tctggccagc	gcacagccga	agagctgcaa	aaagcgccta	cccttcgggtc	10980
gctgcgctcc	ctacgccccg	ccgcttcgcg	tcggcctatc	gcggccgctg	gcccgtcaaa	11040
aatggctggc	ctacggccag	gcaatctacc	agggcgcgga	caagccgcgc	cctgcgcact	11100
cgaccgcggc	cgttgaggtc	tgccctcgta	agaagggtgt	gctgactcat	accaggcctg	11160
aatcgcccca	tcatccagcc	agaaagtga	ggagccacgg	ttgatgagag	ctttgttgta	11220
ggtggaccag	ttggtgattt	tgaacttttg	ctttgccacg	gaacgggtctg	cgttgctggg	11280
aagatgcgtg	atctgaccc	tcaactcagc	aaaagttcga	tttattcaac	aaagccgccc	11340
tcccgtcaag	tcagcgtaat	gctctgccag	tgttacaacc	aattaaccaa	ttctgattag	11400
aaaaactcat	cgagcatcaa	atgaaactgc	aatttattca	tatcaggatt	atcaatacca	11460
tatttttgaa	aaagccgttt	ctgtaatgaa	ggagaaaact	caccgaggca	gttccatagg	11520
atggcaagat	cctggtatcg	gtctgcgatt	ccgactcgte	caacatcaat	acaacctatt	11580
aattttccct	cgtcaaaaat	aagggttatc	agtgagaaat	caccatgagt	gacgactgaa	11640
tccggtgaga	atggcaaaaag	ctctgcatta	atgaatcggc	caacgcgcgg	ggagaggcgg	11700
tttgcgattt	gggcgctctt	ccgcttctct	gctcactgac	tcgctgcgct	cggctgcttc	11760
gctgcggcga	gcggtatcag	ctcactcaaa	ggcggttaata	cggttatcca	cagaatcagg	11820
ggataacgca	ggaaagaaca	tgtgagcaaa	aggccagcaa	aaggccagga	accgtaaaaa	11880
ggccgcggtt	ctggcggttt	tccataggct	ccgccccctt	gacgagcate	acaaaaatcg	11940
acgctcaagt	cagagggtgg	gaaacccgac	aggactataa	agataaccagg	cgtttccccc	12000
tggaaagctc	ctcgtgcgct	ctcctgttcc	gacctgcgcg	cttaccggat	actgtctcgc	12060
ctttctccct	tcgggaagcg	tggcgctttc	tcatatgctca	cgtgtgaggt	atctcagttc	12120
ggtgtagggtc	gttcgctcca	agctgggctg	tgtgcacgaa	cccccgcttc	agcccgaccg	12180
ctgcgcctta	tccggttaact	atcgtcttga	gtccaacccg	gtaagacacg	acttatcgcc	12240
actggcagca	gccactggta	acaggattag	cagagcgagg	tatgtaggcg	gtgctacaga	12300
gttcttgaag	tggtggccta	actacggcta	cactagaaga	acagtatttg	gtatctgcgc	12360

tctgctgaag	ccagttacct	tccgaaaaag	agttggtage	tcttgatccg	gcaaacaaac	12420
caccgctggt	agcgggtggt	tttttgtttg	caagcagcag	attacgcgca	gaaaaaaaagg	12480
atctcaagaa	gaccccttga	tcttttctac	ggggtctgac	gctcagtggg	acgaaaactc	12540
acgttaaggg	attttgggtca	tgagattatc	aaaaaggatc	ttcacctaga	tcctttttgat	12600
ccggaattaa	ttcctgtggt	tggcatgcac	atacaaatgg	acgaacggat	aaaccttttc	12660
acgccctttt	aaatatccga	ttattctaat	aaacgctctt	ttctcttagg	tttaccgcgc	12720
aatatatcct	gtcaaacact	gatagtttaa	actgaaggcg	ggaaacgaca	atctgatcat	12780
gagcggagaa	ttaagggagt	cacgttatga	ccccgcgcga	tgacgcggga	caagccggtt	12840
tacgttttga	actgacagaa	ccgcaacgct	gcaggaattg	gccgcagcgg	ccattttaat	12900
caattgggcg	cgccgaattc	gagctcggta	caagcttgca	catgacaaca	attgtaagag	12960
gatggagacc	acaacgatcc	aacaatactt	ctgcgacggg	ctgtgaagta	tagagaagtt	13020
aaacgcccaa	aagccattgt	gttttgaatt	tttagttatt	ctatttttca	tgatgtatct	13080
tcctctaaca	tgccttaatt	tgcaaatttg	gtataactac	tgattgaaaa	tatatgtatg	13140
taaaaaaata	ctaagcatat	ttgtgaagct	aaacatgatg	ttattttaaga	aaatatgttg	13200
ttaacagaat	aagattaata	tcgaaatgga	aacatctgta	aattagaatc	atcttacaag	13260
ctaagagatg	ttcacgcttt	gagaaacttc	ttcagatcat	gaccgtagaa	gtagctctcc	13320
aagactcaac	gaaggctgct	gcaattccac	aaatgcatga	catgcctcct	tgtaaccgctc	13380
gtcgcgcgta	taaacaacga	taactcaatt	ccctgctcca	tcaatttaga	aatgagcaag	13440
caagcacccg	atcgctccac	ccatatgcac	caatctgact	cccaagtctc	tgtttcgcct	13500
tagtaccgcc	agcactccac	ctatagctac	caattgagac	ctttccagcc	taagcagatc	13560
gattgatcgt	tagagtcaaa	gagttggtgg	tacgggtact	ttaactacca	tggaatgatg	13620
gggctgtgat	tagagcggaa	agcgctctcc	tacgcggaac	aacaccctcg	ccatgccgct	13680
cgactacagc	ctcctcctcg	tggcgcgcgc	acaacgaggg	agcccggtgt	cgcagccacc	13740
gaccagcatg	tctctgtgtc	ctcgctccgac	ctcgacatgt	catggcaaac	agtcggacgc	13800
cagcaccaga	ctgacgacat	gagtcctctga	agagcccgcg	acctagaaaag	atccgagccc	13860
tgctgctggt	agtggttaacc	attttcgtcg	cgctgacgcg	gagagcgaga	ggccagaaat	13920
ttatagcgac	tgacgctgtg	gcaggcacgc	tatcggaggt	tacgacgtgg	cgggtcactc	13980
gacgcggagt	tcacaggtcc	tatccttgca	tgcctcgggc	cggagtttac	gggacttatc	14040
cttacgacgt	gctctaaggt	tgcgataacg	ggcggaggaa	ggcgtgtggc	gtgcggagac	14100
ggtttatata	cgtagtgtgc	gggagtggtg	ttcgtagacg	cgggaaagca	cgacgactta	14160
cgaaggtttag	tggaggagga	ggacacacta	aaatcaggac	gcaagaaact	cttctattat	14220
agtagtagag	aagagattat	aggagtggtg	gttgattcta	aagaaaatcg	acgcaggaca	14280
accgtcaaaa	cgggtgcttt	aatatagtag	atatatatat	atagagagag	agagaaagta	14340
caaaggatgc	atttgtgtct	gcatatgac	ggagtattac	taacggccgt	cgtaagaagg	14400
tccatcatgc	gtggagcgag	cccatttggt	tggttgctcag	gccgcagtta	aggcctccat	14460
atatgattgt	cgctgggccc	ataacagcat	ctcctccacc	agtttattgt	aagaataaat	14520
taagtagaga	tatttgctgt	cgggcagaag	aaacttgga	aagaagaaga	agcaagctag	14580
gccaatttct	tgccggcaag	aggaagatag	tggcctctag	tttatatatc	ggcgtgatga	14640
tgatgctcct	agctagaaat	gagagaagaa	aaacggacgc	gtgtttggtg	tgtgtcaatg	14700
gcgtccatcc	ttccatcaga	tcagaacgat	gaaaaagtca	agcacggcat	gcatagtata	14760
tgtatagctt	gttttagtgt	ggctttgctg	agacgaatga	aagcaacggc	gggcatattt	14820
ttcatgtggt	gtagctttca	ggctgaaaga	gacgtggcat	gcaataattc	aggggaattcg	14880
tcagccaatt	gaggtagcta	gtcaacttgt	acattggtgc	gagcaatttt	ccgcactcag	14940
gagggctagt	ttgagagtcc	aaaaactata	ggagattaaa	gaggctaaaa	tcctctcctt	15000
atttaatttt	aaataagtag	tgtatttgta	ttttaactcc	tccaaccctt	ccgattttat	15060
ggctctcaaa	ctagcattca	gtctaattgca	tgcattgctg	gctagaggtc	gtatgggggt	15120
gttaatagca	tagctagcta	caagttaacc	gggtctttta	tatttaataa	ggacaggcaa	15180
agtattactt	acaaataaag	aataaaagcta	ggacgaactc	gtggattatt	actaaatcga	15240
aatggacgta	atattccagg	caagaataat	tgttcgatca	ggagacaagt	ggggcattgg	15300
accggttctt	gcaagcaaga	gcctatggcg	tggtgacacg	gcgcgttgcc	catacatcat	15360
gcctccatcg	atgatccatc	ctcacttgct	ataaaaagag	gtgtccatgg	tgctcaagct	15420
cagccaagca	aataagacga	cttgtttcat	tgattcttca	agagatcgag	cttcttttgc	15480
accacaaggt	cgaggatcca	aca				15503

<210> 15

<211> 14946

<212> DNA

<213> 人工序列

<220>

<223> 人工序列说明: pNOV1313

<220>

<221> misc_feature

<222> (12)..(1993)

<223> Zm Ubi 启动子

<220>

<221> misc_feature

<222> (2016)..(5597)

<223> 编码 H04 毒素部分加全长 CryIAb 尾部分的合成核苷酸序列

<220>

<221> misc_feature

<222> (5805)..(7786)

<223> Zm Ubi 启动子

<220>

<221> misc_feature

<222> (7883)..(8986)

<223> PMI

<400> 15

```

aagcttgcat gctgcagtg cagcgtgacc cggtcgtgcc cctctctaga gataatgagc 60
attgcatgtc taagttataa aaaattacca catatTTTTT ttgtcacact tgTTTTgaagt 120
gcagtttatc tatctttata catatatTTT aactttactc tacgaataat ataatctata 180
gtactacaat aatatcagtg ttttagagaa tcatataaat gaacagttag acatggctca 240
aaggacaatt gagtattttg acaacaggac tctacagttt tatcttttta gtgtgcatgt 300
gttctccttt ttttttgcaa atagcttcac ctatataata cttcatccat tttattagta 360
catccattta gggtttaggg ttaatggttt ttatagacta attttttagg tacatctatt 420
ttattctatt ttagcctcta aattaagaaa actaaaactc tatttttagtt tttttattta 480
ataattttaga tataaaatag aataaaataa agtgactaaa aattaaacaa atacccttta 540
agaaattaaa aaaactaagg aaacattttt cttgttttga gtagataatg ccagcctgtt 600
aaacgccgtc gacgagtcta acggacacca accagcgaac cagcagcgtc gcgtcggggc 660
aagcgacaat gacggcacgg catctctgtc gctgcctctg gacccctctc gagagttccg 720
ctccaccgtt ggacttgctc cgctgtcggc atccagaaat tgcgtggcgg agcggcagac 780
gtgagccggc acggcaggcg gcctcctcct cctctcacgg caccggcagc tacgggggat 840
tcctttccca ccgtccttct gctttccctt cctgcgccgc cgtaataaat agacaccccc 900
tcacaccctt ctttcccca cctcgtgttg ttccggagcg acacacacac aaccagatct 960
cccccaaatc caccgcgtcg cacctccgct tcaaggtaag ccgctcgtcc tcccccccc 1020
ccctctctta ccttctctag atcggcggtt cgggtccatg ttaggggccg gtagttctac 1080
ttctgttcat gtttgtgtta gatccgtgtt tgtgttagat ccgtgctgct agcgttcgta 1140
cacggatgag acctgtacgt cagacacggt ctgattgcta acttgccagt gtttctcttt 1200
ggggaatcct gggatggctc tagccgttcc gcagacggga tcgatttcat gatttttttt 1260
gtttcgttgc atagggtttg gtttgccctt ttcttttatt tcaatatatg ccgtgcactt 1320
gtttgtcggg tcatcttttc atgctttttt ttgtcttggg tgtgatgatg tggctcgggt 1380
gggcgggtcgt tctagatcgg agtagaatc tgtttcaaac tacctgggtg atttattaat 1440
tttggtatctg tatgtgtgtg ccatacatat tcatagttac gaattgaaga tgatggatgg 1500
aaatatcgat ctaggatagg tatacatgtt gatgccccgt ttactgatgc atatacagag 1560
atgctttttg ttgccttggt tgtgatgatg tgggtgtggt gggcggtcgt tcattcgttc 1620
tagatcggag tagaatactg ttcaaacta cctggtgtat ttattaattt tggaactgta 1680
tgtgtgtgtc atacatcttc atagttacga gtttaagatg gatggaaata tcgatctagg 1740
ataggatatac atgttgatgt ggggttttact gatgcatata catgatggca tatgcagcat 1800
ctattcatac gctctaacct tgagtacct tctattataa taaacaagta tgTTTTataa 1860
ttattttgat cttgatatac ttggatgatg gcatatgcag cagctatatg tggatttttt 1920
taccctgcc ttcatacgct atttatttgc ttggtactgt ttcttttgtc gatgctcacc 1980
ctgttgtttg gtgttacttc tgcaggggac caacaatgga caacaacccc aacatcaacg 2040
agtgcacccc ctacaactgc ctgagcaacc ccgaggtgga ggtgctgggc ggcgagcgca 2100
tcgagaccgg ctacaccccc atcgacatca gcctgagcct gaccagttc ctgctgagcg 2160

```

agttcgtgcc	cggcgcgcgc	ttcgtgctgg	gcctgggtgga	catcatctgg	ggcatcttcg	2220
gccccagcca	gtgggaagcc	tccctgggtg	agatcgagca	gttgataaac	caacgcatag	2280
aggaattcgc	ccgcaaccag	gccatcagcc	gcctggaggg	cctgagcaac	ctgtacccaa	2340
tctacgccga	gagcttccgc	gagtgaggag	ccgacccac	caaccccgcc	ctgcgcgagg	2400
agatgcgcac	ccagttcaac	gacatgaaca	gcgcctgac	caccgccatc	cccctgttcg	2460
ccgtgcagaa	ctaccagggt	cccctgctga	gcgtgtacgt	gcaggccgcc	aacctgcacc	2520
tgagcgtgct	gcgcgacgtc	agcgtgttcg	gccagcgtg	gggcttcgac	gccgccacca	2580
tcaacagccg	ctacaacgac	ctgacccgcc	tgatcggcaa	ctacaccgac	cacgcctgct	2640
gctggtacaa	caccggcctg	gagcgcgtgt	ggggctccga	cagccgcgac	tggtatcaggt	2700
acaaccagtt	ccgcgcgcgag	ctgaccctga	ccgtgctgga	catcgtgagc	ctgttcccca	2760
actacgacag	ccgcacctac	cccacccgca	ccgtgagcca	gctgaccgcg	gagatttaca	2820
ccaaccccg	gctggagaac	ttcgacggca	gcttcgcgg	cagcggccag	ggcatcgagg	2880
gcagatccg	cagcccccac	ctgatggaca	tcctgaacag	catcaccatc	tacacgcagc	2940
cccacgcgg	cgagtactac	tgagcggcc	accagatcat	ggccagcccc	gtcggcttca	3000
gcggcccgga	gttcaccttc	cccctgtacg	gcacctggg	caacgctgca	cctcagcagc	3060
gcacgtggc	acagctgggc	cagggagtg	accgcacct	gagcagcacc	ctgtaccgtc	3120
gacctttcaa	catcggcac	aacaaccagc	agctgagcgt	gctggacggc	accgagttcg	3180
cctacggcac	cagcagcaac	ctgcccagcg	ccgtgtaccg	caagagcggc	accgtggaca	3240
gcctggacga	gatccccctt	cagaacaaca	acgtgccacc	tcgacagggc	ttcagccacc	3300
gtctgagcca	cgtgagcatg	ttccgcagtg	gcttcagcaa	cagcagcgtg	agcatcatcc	3360
gtgcacccat	gttcagctgg	attcaccgca	gcgcacccct	gaccaacacc	atcgaccccg	3420
agcgcataca	ccagatcccc	ctgggtgaagg	gtttccgggt	gtggggcgcc	accagcgtga	3480
tcaccggccc	cggcttcacc	ggaggcgaca	tcctgcgcag	aaacaccttc	ggcgacttcg	3540
tgagcctgca	ggtgaacatc	aacagcccca	tcaccagcg	ttaccgcctg	cgcttccgct	3600
acgccagcag	ccgcgacgcc	cgtgtgatcg	tgctgactgg	cgccgctagc	accggtgtgg	3660
gcggtcaggt	gagcgtgaac	atgcccctgc	agaagactat	ggagatcggc	gagaacctga	3720
ctagtgcac	cttcgcgtac	accgacttca	gcaacccctt	cagcttccgc	gccaaccccg	3780
acatcatcgg	catcagcgag	cagccctgt	tcggtgccgg	cagcatcagc	agcggcgagc	3840
tgtacatcga	caagatcgag	atcatcctgg	ccgacgccac	cttcgaggcc	gagagcgacc	3900
tgagcgcgc	ccagaaggcc	gtgaacgccc	gcgcgcccg	cagcaaccag	atcggcctga	3960
agaccgacgt	gaccgactac	cacatcgacc	aggtgagcaa	cctggtggac	tgcttaagcg	4020
acgagttctg	cctggacgag	aagaaggagc	tgagcgagaa	ggtgaagcac	gccaaagcgc	4080
tgagcgacga	gcgcaacctg	ctgcaggacc	ccaacttccg	cggcatcaac	cgccagctgg	4140
accgcggctg	gcgaggcagc	accgatatca	ccatccaggg	cgcgacgac	gtgttcaagg	4200
agaactacgt	gacctgcag	ggcaccttcg	acgagtgtca	ccccacctac	ctgtaccagc	4260
cgatcgacga	gagcaagctg	aaggcctaca	cccgcctacca	gctgcgcggc	tacatcgagg	4320
acagccagga	cctggaaatc	tacctgatcc	gctacaacgc	gaagcacgag	accgtgaacg	4380
tgcccgccac	cggcagcctg	tgcccccctga	gcgcgcccg	ccccatcggc	aagtcggggg	4440
agccgaatcg	atgcgctccg	cacctggagt	ggaacccgga	cctagactgc	agctgcaggg	4500
acggggagaa	gtgcgcccac	cacagccacc	acttcagcct	ggacatcgac	gtgggctgca	4560
ccgacctgaa	cgaggacctg	ggcgtgtggg	tgatcttcaa	gatcaagacc	caggacggcc	4620
acgccgcct	gggcaatcta	gagttcctgg	aggagaagcc	cctggtgggc	gaggccctgg	4680
cccgcgtgaa	gcgtgctgag	aagaagtggc	gcgacaagcg	cgagaagctg	gagtgggaga	4740
ccaacatcgt	gtacaaggag	gccaaaggaga	gcgtggacgc	cctgttcgtg	aacagccagt	4800
acgaccgcct	gcaggccgac	accaacatcg	ccatgatcca	cgccgcgcgac	aagcgcgtgc	4860
acagcattcg	cgaggccctac	ctgcccagagc	tgagcgtgat	ccccgggtgtg	aacgcgcgca	4920
tcttcgagga	actcgagggc	cgcacttca	ccgccttcag	cctgtacgac	gcccgcacag	4980
tgatcaagaa	cggcgacttc	aacaacggcc	tgagctgctg	gaacgtgaag	ggccacgtgg	5040
acgtggagga	gcagaacaac	caccgcagcg	tgctggtggt	gcccagagtgg	gaggccgagg	5100
tgagccagga	ggtgcgcgtg	tgcccgcggc	gcggctacat	cctgcgcgtg	accgcctaca	5160
aggagggcta	cggcgagggc	tgctgacca	tccacgagat	cgagaacaac	accgacgagc	5220
tcaagttcag	caactgcgtg	gaggaggagg	tttaccctcaa	caacaccgtg	acctgcaacg	5280
actacaccgc	gaccagagg	gagtacgaag	gcacctacac	ctctcgcaac	aggggttacg	5340
acggcgcccta	cgagtcacaac	agctccgtgc	cagctgacta	cgccagcgcc	tacgaggaga	5400
aagcctacac	cgacggtaga	cgcgacaacc	catgtgagag	caacagaggc	tacggcgact	5460
acacccccct	gcccgcgtgga	tacgtgacca	aggagctgga	gtacttcccc	gagaccgaca	5520
aggtgtggat	cgagattggc	gagaccgagg	gcaccttcac	cgtggacagc	gtggagctgc	5580
tgctgatgga	ggagtagtag	atctgttctg	cacaaagtgg	agtagtcagt	catcgatcag	5640
gaaccagaca	ccagactttt	attcatacag	tgaagtgaag	tgaagtgcag	tgcatgtagt	5700
tgctggtttt	tgtacaactt	agtatgtatt	tgtatttgta	aaatacttct	atcaataaaa	5760
tttctaattc	ctaaaaccaa	aatccagggg	taccagcttg	catgcctgca	gtgcagcgtg	5820

accgcgtcgt	gccccctctct	agagataatg	agcattgcat	gtctaagtta	taaaaaatta	5880
ccacatatatt	tttttgtcac	acttgtttga	agtgcagttt	atctatcttt	atacatatat	5940
ttaaactttta	ctctacgaat	aatataatct	atagtactac	aataatatca	gtgtttttaga	6000
gaatcatata	aatgaacagt	tagacatggt	ctaaaggaca	attgagtatt	ttgacaacag	6060
gactctacag	ttttatcttt	ttagtgtgca	tgtgttctcc	tttttttttg	caaatagctt	6120
cacctatata	atacttcac	cattttatta	gtacatccat	ttagggttta	gggttaatgg	6180
tttttataga	ctaatttttt	tagtacatct	attttattct	attttagcct	ctaaattaa	6240
aaaaactaaaa	ctctatttta	gtttttttat	ttaataattt	agatataaaa	tagaataaaa	6300
taaagtgact	aaaaattaaa	caaataccct	ttaagaaatt	aaaaaaaacta	aggaaacatt	6360
tttcttggtt	cgagtagata	atgccagcct	gttaaacgcc	gtcgacgagt	ctaacggaca	6420
ccaaccagcg	aaccagcagc	gtcgcgtcgg	gccaaagcga	gcagacggca	cggcatctct	6480
gtcgcgtcct	ctggaccct	ctcgagagtt	ccgctccacc	gttggacttg	ctccgctgtc	6540
ggcatccaga	aattgcgtgg	cggagcggca	gacgtgagcc	ggcacggcag	gcggcctcct	6600
cctcctctca	cggcaccggc	agctacgggg	gattcctttc	ccaccgctcc	ttcgtcttcc	6660
cttcctcgcc	cgcgtaata	aatagacacc	ccctccacac	cctctttccc	caacctcgtg	6720
ttgttcggag	cgcacacaca	cacaaccaga	tctcccccaa	atccaccctg	cggcacctcc	6780
gcttcaagggt	acgcgcgtcg	tctccccccc	ccccccctct	ctaccttctc	tagatcggcg	6840
ttccggtcca	tggtagggc	ccggtagttc	tacttctggt	catgtttgtg	ttagatccgt	6900
gttttgtgtta	gatccgtgct	gctagcgttc	gtacacggat	gcgacctgta	cgtcagacac	6960
gttctgattg	ctaacttgcc	agtgtttctc	tttggggaat	cctgggatgg	ctctagccgt	7020
tccgcagacg	ggatcgattt	catgattttt	tttgtttcgt	tgcatagggt	ttggtttgcc	7080
cttttccctt	atltcaatat	atgccgtgca	cttgtttgtc	gggtcatctt	ttcatgcttt	7140
tttttgtctt	ggttgtgatg	atgtggtctg	gttgggcggg	cgttctagat	cggagtagaa	7200
ttctgtttca	aactacctgg	tggatttatt	aattttggat	ctgtatgtgt	gtgccataca	7260
tattcatagt	tacgaattga	agatgatgga	tggaaatata	gatctaggat	aggatatacat	7320
gttgatgcgg	gttttactga	tgcataataca	gagatgcttt	ttgttcgctt	ggttgtgatg	7380
atgtggtgtg	gttgggcggg	cgttcattcg	ttctagatcg	gagtagaata	ctgtttcaaa	7440
ctacctgggtg	tatttattaa	ttttggaact	gtatgtgtgt	gtcatacatc	ttcatagtta	7500
cgagtttaag	atggatggaa	atatcgatct	aggataggta	tacatgttga	tgtgggtttt	7560
actgatgcat	atacatgatg	gcataatgcag	catctattca	tatgctctaa	ccttgagtac	7620
ctatctatta	taataaacaa	gtatgtttta	taattatttt	gatcttgata	tacttggatg	7680
atggcatatg	cagcagctat	atgtggattt	ttttagccct	gccttcatac	ctattttatt	7740
tgccttggtac	tgtttctttt	gtcgatgctc	accctgttgt	ttggtgttac	ttctgcaggg	7800
atccccgac	atgcaaaaac	tcattaaactc	agtgcaaaac	tatgcctggg	gcagcaaaaac	7860
ggcgttgact	gaactttatg	gtatggaaaa	tccgtccagc	cagccgatgg	ccgagctgtg	7920
gatgggcgca	catccgaaaa	gcagttcacg	agtgcagaat	gccgcgggag	atatcgtttc	7980
actgcgtgat	gtgattgaga	gtgataaatc	gactctgctc	ggagaggccg	ttgccaaaacg	8040
ctttggcgaa	ctgcctttcc	tgttcaaagt	attatgcgca	gcacagccac	tctccattca	8100
ggttcatcca	aacaacacaca	attctgaaat	cggttttgcc	aaagaaaatg	ccgcagggtat	8160
ccgatggat	gccgcgagc	gtaactataa	agatcctaac	cacaagccgg	acgtggtttt	8220
tgcgtgacg	cctttccttg	cgatgaacgc	gtttcgtgaa	ttttccgaga	ttgtctccct	8280
actccagccg	gtcgcagggtg	cacatccggc	gattgtctac	tttttacaac	agcctgatgc	8340
cgaacgttta	agcgaactgt	tgcgccgct	gttgaatatg	caggggtgaag	aaaaatccccg	8400
cgcgttgccg	atlttaaaat	cggccctcga	tagccagcag	ggtgaaccgt	ggcaaacgat	8460
tgcgttaatt	tctgaatttt	acccggaaga	cagcgggtctg	ttctccccgc	tattgctgaa	8520
tgtggtgaaa	ttgaaccctg	gcgaagcgat	gttctctgtc	gctgaaacac	cgcacgctta	8580
cctgcaaggc	gtggcgctgg	aagtgatggc	aaactccgat	aacgtgctgc	gtgcgggtct	8640
gacgcctaaa	tacattgata	ttccggaact	ggttgccaat	gtgaaattcg	aagccaaacc	8700
ggctaaccag	ttgttgacct	agccgggtgaa	acaagggtgca	gaactggact	tcccgattcc	8760
agtggatgat	tttgcccttct	cgtgcgatga	ccttagtgat	aaagaaacca	ccattagcca	8820
gcagagtgcc	gccattttgt	tctgcgtcga	aggcgatgca	acgttgtgga	aaggttctca	8880
gcagttacag	cttaaacccg	gtgaatcagc	gtttattgcc	gccaacgaat	caccggtgac	8940
tgtcaaaggc	cacggccggt	tagcgcgtgt	ttacaacaag	ctgtaagagc	ttactgaaaa	9000
aattaacatc	tcttgctaag	ctgggagctc	gatccgtcga	cctgcagatc	gttcaaacat	9060
ttggcaataa	agtttctttaa	gattgaatcc	tgttgccggt	cttgcgatga	ttatcatata	9120
atttctgttg	aattacgtta	agcatgtaat	aattaacatg	taatgcatga	cgttatttat	9180
gagatgggtt	tttatgatta	gagtcgccga	attatacatt	taatacgcga	tagaaaaaaa	9240
aatatagcgc	gcaaaactagg	ataaattatc	gcgcgcgggtg	tcacttatgt	tactagatct	9300
gctagccctg	caggaaattt	accgggtgcc	gggcggccag	catggccgta	tccgcaatgt	9360
gttattaagt	tgtctaagcg	tcaattttgt	tacaccacaa	tatatcctgc	caccagccag	9420
ccaacagctc	ccgcaccggc	agctcggcac	aaaatcacca	ctcgatacag	gcagcccatc	9480

```

agaattaatt ctcattgtttg acagcttata atcgactgca cgggtgcacca atgcttcttg 9540
cgtcaggcag ccatcggaag ctgtggtatg gctgtgcagg tcgtaaatca ctgcataatt 9600
cgtgtcgtc aaggcgcact cccgttcttg ataattgttt ttgcgcgcac atcataacgg 9660
ttctggcaaa tattctgaaa tgagctgttg acaattaatc atccggctcg tataatgtgt 9720
ggaattgtga gcggaataca atttcacaca ggaaacagac catgagggaa gcgttgatcg 9780
ccgaagtatc gactcaacta tcagaggtag ttggcgtcat cgagcgccat ctgcaaccga 9840
cgttgctggc cgtacatttg tacggctccg cagtggatgg cggcctgaag ccacacagt 9900
atattgattt gctggttacg gtgaccgtaa ggcttgatga aacaacgcgg cgagctttga 9960
tcaacgacct ttggaact tcggcttccc ctggagagag cgagattctc cgcgctgtag 10020
aagtcaccat tgttgtgcac gacgacatca ttccgtggcg ttatccagct aagcgcgaac 10080
tgcaatttg agaatggcag cgcaatgaca ttcttgagg tatcttcgag ccagccacga 10140
tcgacattga tctggtatc ttgctgacaa aagcaagaga acatagcgtt gcctggtag 10200
gtccagcggc ggaggaactc ttgatccgg ttccctgaaca ggatctattt gaggcgctaa 10260
atgaaacctt aacgctatgg aactcgccgc ccgactgggc tggcgatgag cgaaatgtag 10320
tgcttacgtt gtccgcatt ttgtacagc cagtaaccgg caaaatcgcg ccgaaggatg 10380
tcgctgcga ctgggcaatg gagcgctgc cggcccgata tcagcccgct atacttgaag 10440
ctaggcaggc ttatcttgga caagaagatc gcttggcctc gcgcgcagat cagttggaag 10500
aatttgttca ctacgtgaaa ggcgagatca ccaaagtagt cggcaaataa agctctagt 10560
gatctccgta ccccggggg atctggctcg cggcggacgc acgacgcggg ggcgagacca 10620
taggcgatct cctaaatcaa tagtagctgt aacctcgagg cgtttcactt gtaacaacga 10680
ttgagaattt ttgtcataaa attgaaatac ttggttcgca tttttgtcat ccgcggtcag 10740
ccgcaattct gacgaactgc ccatttagct ggagatgatt gtacatcctt cacgtgaaaa 10800
tttctcaagc gctgtgaaca aggttccaga ttttagattg aaaggtgagc cgttgaaaca 10860
cgttctctct gtgatgacg acgtcgctat gcggcatctt attattgaat accttacgat 10920
ccacgccttc aaagtgaccg cggtagccga cagcaccag ttcaacagag tactctcttc 10980
cgcgacggtc gatgtcgtgg ttgttgatct aaatttaggt cgtgaagatg ggctcgagat 11040
cgttcgtaat ctggcggcaa agtctgatat tccaatcata attatcagtg gcgaccgct 11100
tgaggagacg gataaagtgt ttgcactcga gctaggagca agtgatttta tcgctaagcc 11160
gttcagtatc agagagtttc tagcacgat tcgggttgcc ttgcgcgtgc gcccacgt 11220
tgtccgctcc aaagaccgac ggtctttttg ttttactgac tggacactta atctcaggca 11280
acgtcgcttg atgtccgaag ctggcgggtga ggtgaaactt acggcagggt agttcaatct 11340
tctctcgcg ttttagaga aaccccgcg cgttctatcg cgcgagcaac ttctcattgc 11400
cagtcgagta cgcgacgagg aggtttatga caggagtata gatgttctca ttttgaggct 11460
gcgcgcgaaa cttgaggcag atccgtcaag cctcaactg ataaaaacag caagagggtgc 11520
cggttatttc tttagcgcgg acgtgcagggt ttgcacggg gggacgatgg cagcctgagc 11580
caattcccag atcccagg aatcggcgtg agcggtcgca aaccatccgg cccggtacaa 11640
atcggcgcgg cgctgggtga tgacctggtg gagaagttga aggcgcgcga ggcgcgccag 11700
cggcaacgca tcgaggcaga agcacgcccc ggtgaatcgt ggcaagcggc cgtgatcga 11760
atccgcaaag aatcccggca accgcgggca gccggtgcgc cgtcgattag gaagccgccc 11820
aagggcgacg agcaaccaga tttttctggt ccgatgctct atgacgtggg caccgcgcat 11880
agtcgcagca tcatggacgt ggccgttttc cgtctgtcga agcgtgaccg acgagctggc 11940
gaggtgatcc gctacgagct tccagacggg cagtagagg tttccgcagg gccggcggc 12000
atggccagtg tgtgggatta cgacctggtt ctgatggcgg tttcccatct aaccgaatcc 12060
atgaaccgat accgggaagg gaagggagac aagcccggc gcgtgttccg tccacaggtt 12120
gcggacgtac tcaagttctg ccggcgagcc gatggcgaa agcagaaaga cgacctggt 12180
gaaacctgca ttcggttaaa caccacgac gttgccatgc agcgtacgaa gaaggccaag 12240
aacggccgcc tggtagcgggt atccgagggt gaagccttga ttagccgcta caagatcgta 12300
aagagcgaag ccgggcggcc ggagtacatc gagatcgagc tagctgattg gatgtaccgc 12360
gagatcacag aaggcaagaa cccggacgtg ctgacggttc accccgatta ctttttgatc 12420
gatcccgcca tcggccgttt tctctaccgc ctggcacgcc gcgcgcgagg caaggcagaa 12480
gccagatggt tgttcaagac gatctacgaa cgcagtggca gcgcgggaga gttcaagaag 12540
ttctgtttca ccgtgcgcaa gctgatcggg tcaaatgacc tgccggagta cgatttgaag 12600
gaggaggcgg ggcaggctgg cccgatccta gctatgcgt accgcaacct gatcgagggc 12660
gaagcatecg ccggttctta atgtacggag cagatgctag ggcaaattgc ctagcagggt 12720
gaaaaaggtc gaaaaggctt ctttctgtg gatagcacgt acattgggaa ccaaagccg 12780
tacattggga accggaaccc gtacattggg aacccaaagc cgtacattgg gaaccggtca 12840
cacatgtaag tgactgatat aaaagagaaa aaaggcgatt tttccgcta aaactcttta 12900
aaacttatta aaactcttaa aaccgcctg gctgtgcat aactgtctgg ccagcgcaca 12960
gccgaagagc tgcaaaaage gctaccctt cggtcgctgc gctccctacg cccgcgcgt 13020
tcgctcggc ctatcgcggc cgtggccgc tcaaaaatgg ctggcctacg gccaggcaat 13080
ctaccagggc gcggacaagc cgcgcgctc cactcgacc gccggcgtg aggtctgcct 13140

```

```

cgtgaagaag gtgttgctga ctcataccag gctgaatcg ccccatcatc cagccagaaa 13200
gtgaggggagc cacgggtgat gagagctttg ttgtagggtgg accagttggt gatttttgaac 13260
tttttgctttg ccacggaacg gtctgcggttgc tccgggaagat gcgtagatctg atccttcaac 13320
tcagcaaaaag ttcgatttat tcaacaaaagc cgcgcgtccc tcaagtcagc gtaatgctct 13380
gccagtgtta caaccaatta accaattctg attagaaaaa ctcatcgagc atcaaatgaa 13440
actgcaattt attcatatca ggattatcaa taccatattt ttgaaaaagc cgtttctgta 13500
atgaaggaga aaactcaccg aggcagttcc ataggatggc aagatcctgg tatcgggtctg 13560
cgattccgac tcgtccaaca tcaatacaac ctattaattt cccctcgtca aaaataaggt 13620
tatcaagtga gaaatcacca tgagtgcga ctgaatccgg tgagaatggc aaaagctctg 13680
cattaatgaa tcggccaacg cgcggggaga ggcgggtttgc gtattgggag ctcttcgct 13740
tcctcgctca ctgactcgct gcgctcggtc gttcgggtgc ggcgagcgg atcagctcac 13800
tcaaaggcgg taatacgggt atccacagaa tcaggggata acgcaggaaa gaacatgtga 13860
gcaaaaggcc agcaaaaggc caggaaaccg aaaaaggccg cgttgctggc gtttttccat 13920
aggctccgcc cccctgacga gcatcacaaa aatcgacgct caagtcagag gtggcgaaac 13980
ccgacaggac tataaagata ccaggcgttt cccctgggaa gctccctcgt gcgctctcct 14040
gttccgaccc tgcgcgttac cggatacctg tccgccttcc tcccttcggg aagcgtggcg 14100
ctttctcata gctcagctg taggtatctc agttcgggtg aggtcgttcg ctccaagctg 14160
ggctgtgtgc acgaaccccc cgttcagccc gaccgctgcg ccttatccgg taactatcgt 14220
cttgagtcca acccggttaag acacgactta tcgccactgg cagcagccac tggtaacagg 14280
attagcagag cgaggtatgt aggcggtgct acagagttct tgaagtgggt gcctaactac 14340
ggctacacta gaagaacagt atttgggtatc tgcgctctgc tgaagccagt taccttcgga 14400
aaaagagtgt gtagctcttg atccggcaaa caaaccaccg ctggtagcgg tgggtttttt 14460
gtttgcaagc agcagattac gcgcagaaaa aaaggatctc aagaagatcc tttgatcttt 14520
tctacggggt ctgacgctca gtggaacgaa aactcacgtt aagggtttt gggtcatgaga 14580
ttatcaaaaa ggatcttcac ctagatcctt ttgatccgga attaatcct gtggttggca 14640
tgcacatata aatggacgaa cggataaacc ttttcacgcc cttttaataa tccgattatt 14700
ctaataaacg ctcttttctc ttaggtttac ccgccaatat atcctgtcaa acactgatag 14760
tttaaaactga aggcgggaaa cgacaatctg atcatgagcg gagaattaag ggagtcacgt 14820
tatgaccccc gccgatgacg cgggacaagc cgttttacgt ttggaactga cagaaccgca 14880
acgctgcagg aattggccgc agcggccatt taaatcaatt gggcgcccg aattcgagct 14940
cggtagc 14946

```

<210> 16

<211> 14603

<212> DNA

<213> 人工序列

<220>

<223> 人工序列说明: pNOV1435

<220>

<221> misc_feature

<222> (1)..(2007)

<223> 编码 H04 毒素部分加 Cry1Ab 尾头 40 个氨基酸的合成核苷酸序列

<220>

<221> misc_feature

<222> Complement((8814)..(10022))

<223> PMI

<220>

<221> misc_feature

<222> (11142)..(12032)

<223> 玉米泛在蛋白启动子

<220>

<221> misc_feature

<222> (12037)..(14594)

<223> MTL 启动子

<400> 16

```

atggacaaca accccaacat caacgagtgc atccccctaca actgcctgag caaccccgag 60
gtggaggtgc tgggcggcga ggcgcatcgag accgggtaca ccccatcgat catcagcctg 120
agcctgaccc agttcctgct gagecaggttc gtgccccggc cgggttcctg gctgggcctg 180
gtggacatca tctggggcat cttcggcccc agccagtggg acgccttctt ggtgcagatc 240
gagcagttga taaaccaacg catagaggaa ttgcgccgca accaggccat cagccgcctg 300
gagggcctga gcaacctgta ccaaatctac gccgagagct tccgcgagtg ggaggccgac 360
cccaccaacc ccgccttgcg cgaggagatg cgcattccagt tcaacgacat gaacagcgcc 420
ctgaccaccg ccatccccc gttegcctg cagaactacc aggtgcccc gctgagcgtg 480
tacgtgcagg ccgccaacct gcacctgagc gtgctgcgcg acgtcagcgt gttcggccag 540
cgctggggct tcgacgcgcg caccatcaac agcgcctaca acgacctgac ccgcctgatc 600
ggcaactaca ccgaccacgc cgtgcgctgg tacaacaccg gcctggagcg cgtgtggggt 660
cccgacagcc ggcactggat caggtacaac cagttccgcc gcgagctgac cctgaccgtg 720
ctggacatcg tggacatggt cccaactac gagagccgca cctaccccat ccgacccgtg 780
agccagctga ccgcgagat ttacaccaac cccgtgctgg agaacttcga cggcagcttc 840
cgcgccagcg ccagggcat cgagggcagc atccgcagcc cccacctgat ggacatcctg 900
aacagcatca ccatctacac cgacgccac cgcggcgagt actactggag cggccaccag 960
atcatggcca gccccgtcg cttcagcgcc cccgagttca ccttccccct gtacggcacc 1020
atgggcaacg ctgcacctca gcagcgcac gtggcacagc tgggccaggg agtgtaccgc 1080
accctgagca gcacctgta ccgtcgacct ttcaacatcg gcatcaacaa ccagcagctg 1140
agcgtgctgg acggcaccca gttcgectac ggcaccagca gcaacctgcc cagcgccgtg 1200
taccgcaaga cgggcaccgt ggacagcctg gagcagatcc cccctcagaa caacaacgtg 1260
ccacctcgac agggcttcag ccacctctg cccatgttca gctggattca ccgcagcgcc 1380
agcaacagca gcgtgagcat catcctgcca cccatgttca gctggattca ccgcagcgcc 1380
accctgacca acaccatcga ccccgagcgc atcaaccaga tccccctggt gaagggcttc 1440
cgggtgtggg gggccaccag cgtgatcacc ggccccggct tcaccggagg cgacatcctg 1500
cgcagaaaca ccttcggcga cttcgtgagc ctgcaggtga acatcaacag ccccatcacc 1560
cagcgttacc gctgcgctt tgtggcgctt caggtgagcg tgaacatgcc cctgcagaag 1680
actatggaga tcggcgagaa cctgactagt cgcaccttcc gctacaccga cttcagcaac 1740
cccttcagct tccgcgcaaa ccccgacatc atccgcatca gcgagcagcc cctgttcggt 1800
gccggcagca tcagcagcgg cgagctgtac atcgacaaga tcgagatcat cctggccgac 1860
gccaccttcg aggcgagag cgacctggag cgcgcccaga aggccgtgaa cgccctgttc 1920
accagcagca accagatcgg cctgaagacc gacgtgaccg actaccacat cgaccaggtg 1980
agcaacctgg tggactgctt aagctagaga tctgttctgc acaaagtgga gtatgcagtc 2040
atcgatcagg aaccagacac cagactttta ttcatacagt gaagtgaagt gaagtgcagt 2100
gcagtgaagt gctggttttt gtaccactta gtatgtattt gtatttgtaa aatacttcta 2160
tcaataaaat ttctaattcc taaaacaaa atccagtggg taccagcttg ggctgagtgg 2220
ctccttcaac gttgcggttc tgtcagttcc aaacgtaaaa cggcttgctc cgctcatcgc 2280
gcgggggtca taacgtgact ccttaatttc tccgctcatg atcagattgt cgtttcccg 2340
cttcagttta aactatcagt gtttgacagg atatatggc gggtaaacct aagagaaaag 2400
agcgtttatt agaataacgg atatttaaaa gggcgtgaaa aggtttatcc gttcgtccat 2460
ttgtatgtgc atgccaacca cagggttccc ctcggaagtg cttggcattc cgtacgataa 2520
tgactttctg tcaaccacc cagctctggg tcggctagaa ggtcgagtgg gctgctgtgg cttgatccct 2640
caacgcggtc gcggacgtag cgcagcgccg aaaaatcctc gatcgcaaat ccgacgctgt 2700
cgaaaagcgt gatctgcttg tcgctctttc ggccgagtc ctggccagtc atcacgcgcc 2760
aaagttccgt cacaggatga tctggcgcca gttgctggat ctgccttca atccgggtct 2820
gtggcgggaa ctccacgaaa atatccgaac gcagcaagat cgtcgaccaa ttcttgaaga 2880
cgaaagggcc tcgtgatacg cctattttta taggttaatg tcatgataat aatggtttct 2940
tagacgtcag gtggcacttt tcggggaaat gtgcgcggaa cccctatttg tttatttttc 3000
taaatacatt caaatatgta tccgctcatg agacaataac cctgataaat gcttcaataa 3060
tattgaaaaa ggaagagtat gagtattcaa catttccgtg tcgcccttat tccctttttt 3120
gcggcatttt gccttctgt ttttgctcac ccagaaacgc tggtgaaagt aaaagatgct 3180
gaagatcagt tgggtgcacg agtggtttac atcgaaactg atctcaacag cggtaagatc 3240
cttgagagtt tcgccccga agaactgttt ccaatgatga gcacttttaa agttctgcta 3300
tgtggcgcgg tattatcccc gtgtgacgcc gggcaagagc aactcggtcg ccgcatacac 3360
tattctcaga atgacttggg tgagtactca ccagtcacag aaaagcatct tacggatggc 3420
atgacagtaa gagaattatg cagtgtctgc ataaccatga gtgataacac tgcggccaac 3480

```

ttactttctga	caacgatcgg	aggaccgaag	gagctaaccg	ctttttttgca	caacatgggg	3540
gatcatgtaa	ctcgccttga	tcgttgggaa	cggagctga	atgaagccat	accaaacgac	3600
gagcgtgaca	ccacgatgcc	tgcagggggg	gggggggggg	ggacatgagg	ttgccccgta	3660
ttcagtgctg	ctgatttgta	ttgtctgaag	ttgtttttac	gttaagttga	tgcagatcaa	3720
ttaatacgat	acctgcgtca	taattgatta	tttgacgtgg	tttgatggcc	tccacgcacg	3780
ttgtgatatg	tagatgataa	tcattatcac	tttacgggtc	ctttccgggtg	atccgacagg	3840
ttacggggcg	gcgacctcgc	gggtttttcgc	tatttatgaa	aattttccgg	tttaaggcgt	3900
ttccgttctt	cttcgtcata	acttaatgtt	tttatttaaa	ataccctctg	aaaagaaaag	3960
aaacgacagg	tgtcgaaagc	gaggcttttt	ggcctctgtc	gtttcccttc	tctgtttttg	4020
tccgtggaat	gaacaatgga	agtccccccc	cccccccccc	cctgcagcaa	tggcaacaac	4080
gttgcgcaaa	ctattaactg	gcgaactact	tactctagct	tcccggaac	aattaataga	4140
ctggatggag	gcggataaag	ttgcaggacc	actctcgcgc	tcggcccttc	cggctggctg	4200
gtttattgct	gataaatctg	gagccggtga	gcgtgggtct	cgcggtatca	ttgcagcact	4260
ggggccagat	ggtaagccct	cccgtatcgt	agttatctac	acgacgggga	gtcaggcaac	4320
tatggatgaa	cgaaatagac	agatcgctga	gatagggtgc	tactgatta	agcattggta	4380
actgtcagac	caagtttact	catatatact	ttagattgat	ttaaaacttc	atttttaatt	4440
taaaaggatc	taggtgaaga	tcctttttga	taatctcatg	acccaaatcc	cttaacgtga	4500
gttttcgttc	cactgagcgt	cagaccccg	agaaaagatc	aaaggatctt	cttgagatcc	4560
ttttttctg	cgcgtaatct	gctgcttgca	aacaaaaaaa	ccaccgctac	cagcgggtgt	4620
ttgtttgccc	gatcaagagc	taccaactct	ttttccgaag	gttaactggc	tcagcagagc	4680
gcagatacca	aatactgtcc	ttctagtgtg	gccgtagtta	ggccaccact	tcaagaactc	4740
tgtagcaccg	cctacatacc	tcgctctgct	aatcctgtta	ccagtggctg	ctgccagtgg	4800
cgataagtcg	tgtcttaccg	ggttggactc	aagacgatag	ttaccggata	aggcgcagcg	4860
gtcgggctga	acgggggggt	cgtgcacaca	gcccagcttg	gagcgaacga	cctacaccga	4920
actgagatac	ctacagcgtg	agctatgaga	aagcgccacg	cttcccgaag	ggagaaaagg	4980
ggacagggtat	ccggttaagc	gcagggtcgg	aacaggagag	cgcacgaggg	agcttccagg	5040
gggaaacgcc	tggatatctt	atagtctctg	cgggtttcgc	cacctctgac	ttgagcgtcg	5100
atttttgtga	tgtctgtcag	gggggcggag	cctatggaaa	aacgccagca	acgcggcctt	5160
tttacggttc	ctggcctttt	gctggccttt	tgtcacatg	ttctttcctg	cgttatcccc	5220
tgattctgtg	gataaccgta	ttaccgcctt	tgagttagct	gataccgctc	gccgcagccg	5280
aacgaccgag	cgcagcgagt	cagttagcga	ggaagcggaa	gagcgcctga	tgcggtattt	5340
tctccttacg	catctgtgcg	gtatttcaca	cgcataatgg	tgcactctca	gtacaaatctg	5400
ctctgatgcc	gcatagttaa	gccagtatac	actccgctat	cgctacgtga	ctgggtcatg	5460
gctgcgcccc	gacacccgcc	aacacccgct	gacgcgccct	gacgggcttg	tctgtctccc	5520
gcacccgctt	acagacaagc	tgtgaccgtc	tcggggagct	gcattgtgtc	gaggttttca	5580
ccgtcatcac	cgaaacgcgc	gaggcagcag	atcccccgat	caagtagata	cactacatat	5640
atctacaata	gacatcgagc	cggaagggtga	tgtttacttt	cctgaaatcc	ccagcaattt	5700
taggccagtt	tttacccaag	acttcgcctc	taacataaat	tatagttacc	aaatctggca	5760
aaagggttaa	caagtggcag	caacggatcc	gcaaacctgt	cacgcctttt	gtgccaaaag	5820
ccgcgccagg	tttgcgatcc	gctgtgccag	gcgttaggcg	tcatatgaag	atttcggtga	5880
tccttgagca	ggtggcggaa	acattggatg	ctgagaacca	tttcattgtt	cgtgaagtgt	5940
tcgatgtgca	cctatccgac	caaggctttg	aactatctac	cagaagtgtg	agccccctacc	6000
ggaaggatta	catctcggat	gatgactctg	atgaagactc	tgcttgctat	ggcgatttca	6060
tcgaccaaga	gcttgctcgg	aagattgaac	tcaactcaac	atggaacgat	ctagcctcta	6120
tcgaacacat	tgttgtgtcg	cacacgcacc	gaggcaaagg	agtcgcgcac	agtctcatcg	6180
aattttgcgaa	aaagtgggca	ctaagcagac	agctccttgg	catacgatta	gagacacaaa	6240
cgaacaatgt	acctgcctgc	aattttgtacg	caaaatgtgg	ctttactctc	ggcggcattg	6300
acctgttcac	gtataaaaact	agacctcaag	tctcgaaacga	aacagcgatg	tactgggtact	6360
ggttctcggg	agcacaggat	gacgcctaac	aattcattca	agccgacacc	gcttcgcggc	6420
gcggtctaat	tcaggagtta	aacatcatga	gggaagcggg	gatecgccga	gtatcgactc	6480
aactatcaga	ggtagtgtgc	gtcatcgagc	gccatctcga	accgacgttg	ctggccgtac	6540
atttgtacgg	ctccgcagtg	gatggcggcc	tgaagccaca	cagtgatatt	gatttgctgg	6600
ttacggtgac	cgtaaggcct	gatgaaacaa	cgcggcgagc	tttgatcaac	gaccttttgg	6660
aaacttcggc	ttccccctga	gagagcgaga	ttctccgcgc	tgtagaagtc	accattgttg	6720
tgcacgacga	catcattccg	tggcgttatc	cagctaagcg	cgaactgcaa	tttgagaaag	6780
ggcagcgcaa	tgacattctt	gcaggatatc	tcgagccagc	cacgatcgac	attgatctgg	6840
ctatcttgct	gacaaaagca	agagaacata	cggttgccct	ggtaggtcca	gcggcggagg	6900
aactctttga	tcgggttcct	gaacaggatc	tatttgaggc	gctaaatgaa	accttaacgc	6960
tatggaactc	gccgcccagc	tgggctggcg	atgagcgaaa	tgtagtgtct	acgttgtccc	7020
gcatttggtg	cagcgcagta	accggcaaaa	tcgcgccgaa	ggatgtcgct	gccgactggg	7080
caatggagcg	cctgccggcc	cagtatcagc	ccgtcatact	tgaagctagg	caggcttata	7140

```

ttggacaaga agategcttg gcctcgcgcg cagatcagtt ggaagaattt gttcactacg 7200
tgaaaggcga gatcaccaag gtagtcggca aataatgtct aacaattcgt tcaagcgcac 7260
gccgcttcgc ggcgcggcctt aactcaagcg ttagagagct ggggaagact atgcgcgac 7320
tggtgaaggt ggttctaagc ctctgacttg cgatggcctc ggggcaggca cttgctgacc 7380
tgccaattgt tttagtggat gaagctcgtc ttccctatga ctactcccca tccaactacg 7440
acatttctcc aagcaactac gacaactcca taagcaatta cgacaatagt ccatcaaatt 7500
acgacaactc tgagagcaac tacgataata gttcatccaa ttacgacaat agtcgcaacg 7560
gaaatcgtag gcttatatat agcgcaaatg ggtctcgcac ttctcgccgc tactacgtca 7620
ttgccaacaa tgggacaacg aacttctttt ccacatctgg caaaaggatg ttctacaccc 7680
caaaaggggg gcgcggcgctc tatggcgcca aagatgggag cttctgcggg gcattggctc 7740
tcataaatgg ccaatttttcg cttgccctga cagataacgg cctgaagatc atgtatctaa 7800
gcaactagcc tgctctctaa taaaatgtta ggcccaaca tctagtcgca agctgagggg 7860
aaccactagt gtcatacgaa cctccaagag acggttacac aaacgggtac attggtgatg 7920
tcatgtatga caatcgcccc agtaagtatc cagctgtggt cagaacgtac gtccgaatta 7980
ttcatcggg gtagcggtcga cgatcgcca cgttcacttc taaagaaata gcgcactcca 8040
gcttctcag cggttttctc cagcgatttc ctattatgtc ggcatagttc tcaagatcga 8100
cagcctgtca cggttaagcg agaaatgaat aagaaggctg ataattcgga tctctgcgag 8160
ggagatgata tttgatcaca ggcagcaacg ctctgtcctc gttacaatca acatgctacc 8220
ctccgcgaga tcatccgtgt ttcaaaccgc gcagcttagt tgccgttctt ccgaatagca 8280
tcggtaacat gagcaagctc tgccgcctta caacggctct cccgctgacg ccgtcccggg 8340
ctgatgggct gcctgtatcg agtgggtgatt ttgtgcgcgag ctgcgggtcg gggagctgtt 8400
ggctggctgg tggcaggata tattgtgggtg taaacaaatt gacgcttaga caacttaata 8460
acacattgcy gacgttttta atgtactgaa ttgtctagac ccggggatct catgtttgac 8520
agcttatcat cggatctagt aacatagatg acaccgcgcg cgataattta tctagttttg 8580
cgcgctatat tttgttttct atcgcgtatt aaatgtataa ttgcgggact ctaatcataa 8640
aaaccatct cataaataac gtcatgcatt acatgttaat tattacatgc ttaacgtaat 8700
tcaacagaaa ttagatgata atcatcgcaa gaccggcaac aggattcaat cttaagaaac 8760
tttattgcca aatgtttgaa cgatctctgc aggtcgacgg atcgagctcc cagcttagca 8820
agagatgtta attttttcag taagctctta cagcttggtg taaacacgcg ctaaacggcc 8880
gtggcctttg acagtcaccg gtgattcggt ggcggcaata aacgctgatt caccgggttt 8940
aagctgtaac tgctgagaac ctttccacaa cgttgcctcg cttcgcacgc agaacaaat 9000
ggcggcactc tgcggctaa tgggtggttc tttatcacta aggtcatgca gcgagaaggc 9060
aaaatcatcc actggaatcg ggaagtccag ttctgcacct tgtttcacgc gctgggtcaa 9120
caactgggta gccgggtttg cttcgaattt cacattggca accagttccg gaatatcaat 9180
gtatttaggc gtcagaccgc cacgcagcac gttatcggag tttgccatca cttccagcgc 9240
cacgccttgc aggtaaagct gcggtgtttc agcgaacagg aacatcgctt cgccagggtt 9300
caatttcacc acattcagca atagcggggg gaacagaccg ctgtcttccg ggtaaaattc 9360
agaaattaaa cgaatcggtt gccacgggtc acctgtctgg ctatcgaggg ccgattttta 9420
aatgcgcgc gcgcgggatt tttcttcacc ctgcatttcc aacaggctgg cgaatgattc 9480
gcttaaacgt tcggcatcag gctgttgtaa aaagttagca atcgccggat gtgcacctgc 9540
gaccggctgg agtagggaga caatctcgga aaattcacga aacgcgttca tcgcaaggaa 9600
aggcgtcagc gcaaaaacca gctccggctt gtgggttagga tctttatagt tacgctcggc 9660
ggcatccatc gggatacctg cggcattttc tttggcaaaa ccgatttcag aattgtgttt 9720
gtttggatga acctgaatgg agagtggctg tgctgcgcct aatactttga acaggaaagg 9780
cagttcgcca aagcgttttg caacggcctc tccgagcaga gtcgatttat cactctcaat 9840
cacatcacgc agtgaaacga tatctccggc ggcattctgc actcgtgaac tgcctttcgg 9900
atgtgcgccc atccacagct cggccatcgg ctggctggac ggattttcca taccataaag 9960
ttcagtcacac gcgttttgcg gccccaggca tagttttgca ctgagttaat gagtttttgc 10020
atgatcgggg atccctgcag aagtaacacc aaacaacagg gtgagcatcg acaaaagaaa 10080
cagtaccaag caaataaata gcgtatgaag gcagggctaa aaaaatccac atatagctgc 10140
tgcatatgcc atcatccaag tatatcaaga tcaaaataat tataaacat acttgtttat 10200
tataatagat aggtactcaa ggttagagca tatgaataga tgctgcatat gccatcatgt 10260
atatgcatca gtaaaaccca catcaacatg tatacctatc ctagatcgat atttccatcc 10320
atcttaactc cgtaactatg aagatgtatg acacacatc acagttccaa aattaataaa 10380
tacaccaggt agtttgaaac ggcgtctact ccgattctaga acgaatgaac acgcgccaa 10440
ccacaccaca tcatcacaac caagcgaaca aaaagcatct ctgtatatgc atcgataaaa 10500
ccgcacatca catgtatacc tatcctagat cgatatttcc atccatcatc ttcaattcgt 10560
aactatgaat atgtatggca cacacataca gatccaaaat taataaatcc accaggtagt 10620
ttgaaacaga attctactcc gatctagaac gaccgcccac ccagaccaca tcatcacaac 10680
caagacaaaa aaaagcatga aaagatgacc cgacaaacaa gtgcacggca tatattgaaa 10740
taaaggaaaa gggcaaacca aaccctatgc aacgaaacaa aaaaaatcat gaaatcgatc 10800

```

ccgtctgcgg	aacggctaga	gccatcccag	gattccccaa	agagaaacac	tggcaagtta	10860
gcaatcagaa	cgtgtctgac	gtacagggtcg	catccgtgta	cgaacgctag	cagcacggat	10920
ctaacacaaa	cacggatcta	acacaaaacat	gaacagaagt	agaactaccg	ggccctaacc	10980
atggaccgga	acgccgatct	agagaaggta	gagagggggg	gggggggagg	acgagcggcg	11040
taccttgaag	cggagggtgcc	gacgggtgga	tttgggggag	atctggttgt	gtgtgtgtgc	11100
gctccgaaca	acacgagggt	gggaaaagag	ggtgtggagg	gggtgtctat	ttattacggc	11160
gggcgaggaa	gggaaagcga	aggagcgggtg	ggaaaggaat	ccccgtagc	tgccgtgccg	11220
tgagaggagg	aggaggccgc	ctgccgtgcc	ggctcacgtc	tgccgctccg	ccacgcaatt	11280
tctggatgcc	gacagcggag	caagtccaac	ggtggagcgg	aactctcgag	aggggtccag	11340
aggcagcgac	agagatgccg	tgccgtctgc	ttcgcttggc	ccgacgcgac	gctgctggtt	11400
cgctggttgg	tgtccgttag	actcgtcgac	ggcgtttaac	aggctggcat	tatctactcg	11460
aaacaagaaa	aatgtttcct	tagttttttt	aattttctta	agggtatttg	tttaattttt	11520
agtcacttta	ttttattcta	ttttatatct	aaattattaa	ataaaaaaac	taaaatagag	11580
ttttagtttt	cttaatttag	aggctaaaat	agaataaaat	agatgtacta	aaaaaattag	11640
tctataaaaa	ccattaaccc	taaaccctaa	atggatgtac	taataaaatg	gatgaagtat	11700
tataatagtg	aagctatttg	caaaaaaaaa	ggagaacaca	tgcacactaa	aaagataaaa	11760
ctgtagagtc	ctgtttgtcaa	aatactcaat	tgtccttttag	accatgtcta	actgttcatt	11820
tatatgattc	tctaaaaacac	tgatattatt	gtagtactat	agattatatt	attcgtagag	11880
taaagtttta	atatatgtat	aaagatagat	aaactgcact	tcaaacaagt	gtgacaaaaa	11940
aaatatgtgg	taatttttta	taacttagac	atgcaatgct	cattatctct	agagaggggc	12000
acgaccgggt	cacgctgcac	tgcaggcatg	caagcttgca	catgacaaca	attgtaagag	12060
gatggagacc	acaacgatcc	aacaatactt	ctgcgacggg	ctgtgaagta	tagagaagtt	12120
aaacgcccac	aagccattgt	gtttggaatt	tttagttatt	ctatttttca	tgatgtatct	12180
tctcttaaca	tgccttaatt	tgcaaatttg	gtataactac	tgattgaaaa	tatatgtatg	12240
taaaaaaaata	ctaagcatat	ttgtgaagct	aaacatgatg	ttatttaaga	aaatatgttg	12300
ttaacagaat	aagattaata	tcgaaatgga	aacatctgta	aattagaatc	atcttacaag	12360
ctaagagatg	ttcacgcttt	gagaaacttc	ttcagatcat	gaccgtagaa	gtagctctcc	12420
aagactcaac	gaaggctgct	gcaattccac	aaatgcatga	catgcatcct	tgtaaccgtc	12480
gtcgccgcta	taaacacgga	taactcaatt	ccctgctcca	tcaatttaga	aatgagcaag	12540
caagcacccg	atcgtccacc	ccatatgcac	caatctgact	cccaagtctc	tgtttcgcat	12600
tagtaccgcc	agcactccac	ctatagctac	caattgagac	ctttccagcc	taagcagatc	12660
gattgatcgt	tagagtcaaa	gagttggtgg	tacgggtact	ttaactacca	tggaatgatg	12720
gggcgtgatg	tagagcggaa	agcgccctcc	tacgcggaac	aacaccctcg	ccatgccgct	12780
cgactacagc	ctcctcctcg	tccggccgcc	acaacgaggg	agcccgtggg	cgcagccacc	12840
gaccagcatg	tctctgtgtc	ctcgtccgac	ctcgacatgt	catggcaaac	agtcggacgc	12900
cagcaccaga	ctgacgacat	gagtcctctga	agagcccgc	acctagaaag	atccgagccc	12960
tgctgctggt	agtggtaacc	attttcgtcg	cgctgacgcg	gagagcgaga	ggccagaaat	13020
ttatagcgac	tgacgctgtg	gcaggcacgc	tatcggaggt	tacgacgtgg	cgggtcactc	13080
gacgcggagt	tcacaggtcc	tatccttgca	ctcgtcgggc	cggagtttac	gggaattatc	13140
cttacgacgt	gctctaaggt	tgcgataacg	ggcggaggaa	ggcgtgtggc	gtgcggagac	13200
ggtttataca	cgtagtgtgc	gggagtgtgt	ttcgtagacg	cgggaaagca	cgacgactta	13260
cgaagggttag	tggaggagga	ggacacacta	aaatcaggac	gcaagaaact	cttctattat	13320
agtagtagag	aagagattat	aggagtgtgg	gttgattcta	aagaaaatcg	acgcaggaca	13380
accgtcaaaa	cgggtgcttt	aatatagtag	atatatataat	atagagagag	agagaaagta	13440
caaaggatgc	atltgtgtct	gcatatgac	ggagtattac	taacggccgt	cgtaagaagg	13500
tccatcatgc	gtggagcgag	cccatttggt	tggttgtcag	gccgcagtta	aggcctccat	13560
atatgattgt	cgtcggggcc	ataacagcat	ctcctccacc	agtttattgt	aagaataaat	13620
taagtagaga	tatttgtcgt	cgggcagaag	aaacttggac	aagaagaaga	agcaagctag	13680
gccaatttct	tgcgggcaag	aggaagatag	tggcctctag	tttatatata	ggcgtgatga	13740
tgatgctcct	agctagaaat	gagagaagaa	aaacggacgc	gtgttttggtg	tgtgtcaatg	13800
gcgtccatcc	ttccatcaga	tcagaacgat	gaaaaagtca	agcacggcat	gcatagtata	13860
tgtatagctt	gttttagtgt	ggctttgctg	agacgaatga	aagcaacggc	gggcatattt	13920
ttcagtggct	gtagctttca	ggctgaaaga	gacgtggcat	gcaataatcc	agggaaattcg	13980
tcagccaatt	gaggtagcta	gtcaacttgt	acattggtgc	gagcaatttt	ccgcactcag	14040
gagggctagt	ttgagagtcc	aaaaactata	ggagattaaa	gaggctaaaa	tccctctcct	14100
atttaatttt	aaataagtag	tgtatttgta	ttttaactcc	tccaaccctt	ccgattttat	14160
ggctctcaaa	ctagcattca	gtctaataga	tgcattgctt	gctagagggtc	gtatgggggt	14220
gttaatagca	tagctagcta	caagttaacc	gggtctttta	tatttaataa	ggacaggcaa	14280
agtattactt	acaaataaag	aataaagcta	ggacgaactc	gtggattatt	actaaatcga	14340
aatggacgta	atattccagg	caagaataat	tgttcgatca	ggagacaagt	ggggcatttg	14400
accggttctt	gcaagcaaga	gcctatggcg	tggtgacacg	gcgcgttgcc	catacatcat	14460

```

gcctccatcg atgatccatc ctcacttgct ataaaaagag gtgtccatgg tgctcaagct 14520
cagccaagca aataagacga cttgtttcat tgattcttca agagatcgag cttcttttgc 14580
accacaaggt cgaggatcca aca                                     14603

```

<210> 17

<211> 11127

<212> DNA

<213> 人工序列

<220>

<223> 人工序列说明: pZU578

<220>

<221> misc_feature

<222> (1485)..(3491)

<223> 编码 H04 毒素部分加 Cry1Ab 尾头 40 个氨基酸的合成核苷酸序列

<220>

<221> misc_feature

<222> (5052)..(6271)

<223> PMI

<220>

<221> misc_feature

<222> (3859)..(5030)

<223> SMAS 启动子

<220>

<221> misc_feature

<222> (56)..(1475)

<223> 肌动蛋白 2 启动子 U41998

<400> 17

```

ggccgcagcg gccatttaaa tcaattgggc ggcgcgaatt cgagctcggt accctgcatg      60
cctgcagggtc gacaaaattt agaacgaact taattatgat ctcaaataca ttgatacata      120
tctcatctag atctagggtta tcattatgta agaaaagtttt gacgaatatg gcacgacaaa      180
atggctagac tcgatgtaat tggatctca actcaacatt atacttatac caaacattag      240
ttagacaaaa tttaaacaac tattttttat gtatgcaaga gtcagcatat gtataattga      300
ttcagaatcg ttttgacgag ttcggatgta gtagtagcca ttattttaatg tacatactaa      360
tcgtgaatag tgaatatgat gaagcattgt atcttattgt ataaaatatcc ataaacacat      420
catgaaagac actttctttc acggtctgaa ttaattatga cacaattcta atagaaaacg      480
aattaaatta cgttgaattg tatgaaatct aattgaacaa gccaaaccacg acgacgacta      540
acgttgccctg gattgactcg gtttaagtta accactaaaa aaacggagct gtcatgtaac      600
acgcggatcg agcaggtcac agtcatgaag ccatcaaaagc aaaagaacta atccaagggc      660
tgagatgatt aattagttta aaaattagtt aacacgaggg aaaaggctgt ctgacagcca      720
ggtcacgtta tctttacctg tggtcgaaat gattcgtgtc tgtcgatttt aattattttt      780
ttgaaaggcc gaaaataaaag ttgtaagaga taaacccgcc tatataaatt catatatattt      840
cctctccgct ttgaattgtc tcgttgtcct cctcactttc atcagccggt ttgaatctcc      900
ggcgacttga cagagaagaa caaggaagaa gactaagaga gaaagtaaga gataatccag      960
gagattcatt ctccgttttg aatcttcctc aatctcatct tcttccgctc tttctttcca      1020
aggtaatagg aactttctgg atctacttta ttgctggat ctcgatcttg ttttctcaat      1080
ttccttgaga tctggaattc gtttaatttg gatctgtgaa cctccactaa atcttttggt      1140
tttactagaa tcgatctaag ttgaccgatc agttagctcg attatagcta ccagaatttg      1200
gcttgacctt gatggagaga tccatgttca tgttacctgg gaaatgattt gtatatgtga      1260
attgaaatct gaactgttga agttagattg aatctgaaca ctgtcaatgt tagattgaat      1320
ctgaacactg ttttaagttag atgaagtttg tgtatagatt cttecgaaact ttaggatttg      1380
tagtgctgta cgttgaacag aaagctatct ctgattcaat cagggttttat ttgactgtat      1440
tgaactcttt ttgtgtgttt gcagctcata aaaaggatcc aacaatggac aacaacccca      1500

```


acatcaacga	gtgcatcccc	tacaaetgcc	tgagcaaccc	cgaggtggag	gtgctgggag	1560
gcgagcgcat	cgagaccggc	tacaccccc	tcgacatcag	cctgagcctg	acccagttcc	1620
tgctgagcga	gttcgtgccc	ggcgccggct	tcgtgctggg	cctgggtggac	atcatctggg	1680
gcattcttcgg	ccccagccag	tgggacgcct	tcctggtgca	gatcgagcag	ttgataaacc	1740
aacgcataga	ggaatttcgcc	cgcaaccagg	ccatcagccg	cctggaggggc	ctgagcaacc	1800
tgtaccaa	ctacgccgag	agcttccgcg	agtgggaggc	cgacccccacc	aacccccgcc	1860
tgcgcgagga	gatgcgcate	cagttcaacg	acatgaacag	cgccttgacc	accgccatcc	1920
ccctgttcgc	cgtgcagaac	taccaggtgc	ccctgctgag	cgtgtacgtg	caggccgcca	1980
acctgcacct	gagcgtgctg	cgcgacgtca	gcgtgttcgg	ccagcgtggg	ggcttcgacg	2040
ccgccaccat	caacagccgc	tacaacgacc	tgacccgcct	gatcggcaac	tacaccgacc	2100
acgcctgctg	ctggtacaac	accggcctgg	agcgcgtgtg	gggtcccgcac	agccgcgact	2160
ggatcaggta	caaccagttc	cgcgcgcgag	tgacctgac	cgtgctggac	atcgtgagcc	2220
tgttccccaa	ctacgacagc	cgcacctacc	ccatccgcac	cgtgagccag	ctgacccgcg	2280
agattttacac	caaccccgctg	ctggagaact	tcgacggcag	cttccgcggc	agcgcgccagg	2340
gcatecgagg	cagcatecgc	agccccacc	tgatggacat	cctgaacagc	atcaccatct	2400
acaccgacgc	ccaccgcggc	gagtactact	ggagcggcca	ccagatcatg	gccagccccg	2460
tcggcttcag	cggccccgag	ttcaccttcc	ccctgtacgg	caccatgggc	aacgctgcac	2520
ctcagcagcg	catcgtggca	cagctggggc	agggagtgta	ccgcaccctg	agcagcacc	2580
tgtaccgtcg	acctttcaac	atcgccatca	acaaccagca	gctgagcgtg	ctggacggca	2640
ccgagttcgc	ctacggcacc	agcagcaacc	tgcccagcgc	cgtgtaccgc	aagagcggca	2700
ccgtggacag	cctggacgag	atccccctc	agaacaacaa	cgtgccacct	cgacagggct	2760
tcagccaccg	tctgagccac	gtgagcatgt	tcgcagtggt	cttcagcaac	agcagcgtga	2820
gcatecatcg	tgcacccatg	ttcagctgga	ttcaccgcag	cgccaccctg	accaacacca	2880
tcgaccccca	gogcatcaac	cagatcccc	tggtgaagg	cttcgggtg	tgggcgggca	2940
ccagcgtgat	caccggcccc	ggcttcaccg	gaggcgacat	cctgcgcaga	aacaccttcg	3000
gcgacttcgt	gagcctgcag	gtgaacatca	acagcccat	caccagcgt	taccgcctgc	3060
gcttccgcta	cgccagcagc	cgcgacgccc	gtgtgatcgt	gctgactggc	gccgctagca	3120
ccggtgtggg	cggtcagggtg	agcgtgaaca	tgccccctgca	gaagactatg	gagatcgggc	3180
agaacctgac	tagtcgcacc	ttccgctaca	ccgacttcag	caaccccttc	agcttccgcg	3240
ccaacccccga	catcatcggc	atcagcgcag	agccctgtt	cgggtgccggc	agcatcagca	3300
gcggcgagct	gtacatcgac	aagatcgaga	tcatectggc	cgacgccacc	ttcgaggccg	3360
agagcgacct	ggagcgcgc	cagaaggccg	tgaacgcct	gttcaccagc	agcaaccaga	3420
tcggcctgaa	gaccgacgtg	accgactacc	acatcgacca	ggtgagcaac	ctggtggact	3480
gcttaagcta	gagatcctct	agagtcgacc	atggtgatca	ctgcagatcg	ttcaaacatt	3540
tggcaataaa	gtttcttaag	attgaatcct	gttgccggtc	ttgcgatgat	tatcatataa	3600
tttctgttga	attacgttaa	gcattgtaata	attaacatgt	aatgcatgac	gttatattatg	3660
agatgggttt	ttatgattag	agtcocgcga	ttatacattt	aatacgcgat	agaaaacaaa	3720
atatagcgcg	caacctagga	taaattatcg	cgcgcgggtg	catctatgtt	actagatctc	3780
tagaaaagctt	cgtacgttaa	ttaatctgaa	tcgcggagcgg	ccgcagggct	agcagctatg	3840
gtaccgagct	cgagcatata	caggccaaat	tcgctcttag	ccgtacaata	ttactcaccg	3900
gtgcgatgcc	ccccatcgta	ggtgaagggtg	gaaattaatg	atccatcttg	agaccacagg	3960
cccacaacag	ctaccagttt	cctcaagggt	ccaccaaaaa	cgtaagcgt	tacgtacatg	4020
gtcgataaga	aaaggcaatt	tgtagatgtt	aacatccaac	gtcgctttca	gggatcccca	4080
attccaagct	tgggaattcg	gacccctacag	gccaaattcg	ctcttagccg	tacaatatta	4140
ctcaccgggtg	cgatgcccc	catcgtaggt	gaagggtgaa	attaatgatc	catcttgaga	4200
ccacaggccc	acaacagcta	ccagtttccct	caagggtcca	ccaaaaacgt	aagcgtttac	4260
gtacatgggtc	gataagaaaa	ggcaatttgt	agatgttaac	atccaaacgtc	gctttcagg	4320
atcccgaaatt	ccaagcttgg	aattcgggat	cctacaggcc	aaattcgctc	ttagccgtac	4380
aatattactc	accggtgcga	tccccccatc	gtaggtgaag	gtggaaatta	atgatccatc	4440
ttgagaccac	aggccacaaa	cagctaccag	tttctcaag	ggtccaccaa	aaacgtaagc	4500
gcttacgtac	atggtcgata	agaaaaggca	attttagat	gttaacatcc	aacgtcgctt	4560
tcagggatcc	cgaattccaa	gcttgggtcg	caggtaaatc	ccattgcttt	tgaagcagct	4620
caacattgat	ctctttctcg	agggagattt	ttcaaatcag	tgcgcaagac	gtgacgtaag	4680
tatccgagtc	agtttttatt	tttctactaa	tttggtcggt	tatttcggcg	tgtaggacat	4740
ggcaaccggg	cctgaatttc	gcgggtatcc	tgtttctatt	ccaacttttt	cttgatccgc	4800
agccattaac	gacttttgaa	tagatacgt	gacacgccaa	gcctcgctag	tcaaaaagtgt	4860
accaaacaa	gctttacagc	aagaacggaa	tgcgcgtgac	gctcgcgggtg	acgccatttc	4920
gccttttcag	aaatggataa	atagccttgc	ttcctattat	atcttcccaa	attaccaata	4980
cattacacta	gcattctgaat	ttcataacca	atctcgatac	accaaatacga	gatctgcagg	5040
gatccccgat	catgcaaaaa	ctcattaact	cagtgcacaaa	ctatgcctgg	ggcagcaaaa	5100
cggcggttgac	tgaactttat	ggtatggaaa	atccgtccag	ccagccgatg	gccgagctgt	5160

ggatggggcgc	acatccgaaa	agcagttcac	gagtgcagaa	tgccgccgga	gatatcgttt	5220
cactgcgtga	tgtgattgag	agtgataaat	cgactctgct	cggagaggcc	gttgccaaac	5280
gctttggcga	actgcctttc	ctgttcaaag	tattatgcgc	agcacagcca	ctctccattc	5340
aggttcatcc	aaacaaacac	aattctgaaa	tcggttttgc	caaagaaaat	gccgcaggta	5400
tcccgatgga	tgccgccgag	cgtaactata	aagatcctaa	ccacaagccg	gagctggttt	5460
ttgcgctgac	gcctttcctt	gcgatgaacg	cgtttcgtga	attttccgag	attgtctccc	5520
tactccagcc	ggtcgcaggt	gcacatccgg	cgattgctca	ctttttacaa	cagcctgatg	5580
ccgaacgttt	aagcgaactg	ttcgccagcc	tgttgaatat	gcagggtgaa	gaaaaatccc	5640
gcgcgctggc	gattttaaaa	tcggccctcg	atagccagca	gggtgaaccg	tggcaaacga	5700
ttcgtttaat	ttctgaattt	tacccggaag	acagcggctc	gttctccccg	ctattgctga	5760
atgtggtgaa	attgaaccct	ggcgaagcga	tgttcctggt	cgctgaaaca	ccgcacgctt	5820
acctgcaagg	cgtggcgctg	gaagtgatgg	caaactccga	taacgtgctg	cgtgcgggtc	5880
tgacgcctaa	atacattgat	attccggaac	tggttgccaa	tgtgaaattc	gaagccaaac	5940
cggctaacca	gttggtgacc	cagccgggtga	aacaagggtgc	agaactggac	ttcccgatcc	6000
cagtggatga	ttttgccttc	tcgctgcatg	accttagtga	taaagaaacc	accattagcc	6060
agcagagtgc	cgccattttg	ttctgcgtcg	aaggcgatgc	aacgltgtgg	aaaggttctc	6120
agcagttaca	gcttaaaccg	ggtgaatcag	cgtttatgtc	cgccaacgaa	tcaccggtga	6180
ctgtcaaagg	ccacggccgt	ttagcgcgtg	tttacaacaa	gctgtaagag	cttactgaaa	6240
aaattaacat	ctcttgctaa	gctgggagct	cgtcgacgga	tcgaattcct	gcagatcggt	6300
caaacatttg	gcaataaagt	ttcttaagat	tgaatcctgt	tgccggtctt	gcgatgatta	6360
tcatataatt	tctgttgaat	tacgttaagc	atgtaataat	taacatgtaa	tgcatgacgt	6420
tatttatgag	atgggttttt	atgattagag	tcccgcgaatt	atacatttaa	tacgcgatag	6480
aaaacaaaat	atagcgcgca	acctaggata	aattatcgcg	cgcggtgtca	tctatgttac	6540
tagatctcta	gaactagtgg	atctgctagc	cctgcaggaa	atttaccggt	gcccgggcgg	6600
ccagcatggc	cgtatccgca	atgtgttatt	aagtgtccta	agcgtcaatt	tggttacacc	6660
acaatatatc	ctggccaccg	ccagccaaca	gctcccgcac	cggcagctcg	gcacaaaatc	6720
accactgatc	acaggcagcc	catcagaatt	aattctcatg	tttgacagct	tatcatcgac	6780
tgcacgggtgc	accaatgctt	ctggcgctcag	gcagccatcg	gaagctgtgg	tatggctgtg	6840
caggctgtaa	atcactgcat	aattcgtgtc	gctcaaggcg	cactcccgtt	ctggataatg	6900
ttttttgctc	cgacatcata	acggttctg	caaataattct	gaaatgagct	gttgacaatt	6960
aatcatcggc	tcgtataatg	tgtggaattg	tgagcggata	acaatttcac	acaggaaaca	7020
gaccatgagg	gaagcgggtga	tcgccgaagt	atcgactcaa	ctatcagagg	tagttggcgt	7080
catcgagcgc	catctcgaac	cgacggttgct	ggccgtacat	ttgtacggct	ccgcagtgga	7140
tgggcgccat	aagccacaca	gtgatattga	tttgctgggt	acggtgaccg	taaggcttga	7200
tgaacaacag	cggcgagctt	tgatcaacga	ccttttgtaa	acttcggctt	cccctggaga	7260
gagcgagatt	ctccgcgctg	tagaagtcac	cattgttgtg	cacgacgaca	tcattccgtg	7320
gcgttatcca	gctaagcgcg	aactgcaatt	tggagaatgg	cagcgcaatg	acattcttgc	7380
aggtatcttc	gagccagcca	cgatcgacat	tgatctggct	atcttgctga	caaaagcaag	7440
agaacatagc	gttgcccttg	taggtccagc	ggcggaggaa	ctctttgatc	cggttcctga	7500
acaggatcta	tttgaggcgc	taaatgaaac	cttaacgcta	tggaactcgc	cgcccagctg	7560
ggctggcgat	gagccgaatg	tagtgcttac	tggttcccgc	atlttggtaca	gcgcagtaac	7620
cggcaaaaatc	gcgcggaagg	atgtcgtctg	cgactgggca	atggagcgcc	tgccggccca	7680
gtatcagccc	gtcatacttg	aagctaggca	ggcttatctt	ggacaagaag	atcgcttggc	7740
ctcgcgcgca	gatcagttgg	aagaatttgt	tcactacgtg	aaaggcgaga	tcaccaaggt	7800
agtcggcaaa	taaagctcta	gtggatcccc	gaggaatcgg	cgtgacggtc	gcaaaccatc	7860
cggcccggta	caaatcggcg	cggcgctggg	tgatgacctg	gtggagaagt	tgaaggccgc	7920
gcaggccgcc	cagcggcaac	gcacgcaggc	agaagcacgc	cccgggtgaat	cgtggcgaag	7980
ggcgcgtgat	cgaatccgca	agaatccccg	gcaaccgcgc	gcagccgggtg	cgccgtcgat	8040
taggaagccg	cccaaggcg	acgagcaacc	agattttttc	gttccgatgc	tctatgacgt	8100
gggcaccgcg	gatagtgcga	gcacatgga	cgtggccgtt	ttccgtctgt	cgaagcgtga	8160
ccgacgagct	ggcgagggtga	tccgctacga	gcttccagac	gggcacgtag	aggtttcagc	8220
agggccggcc	ggcatggcca	gtgtgtggga	ttacgacctg	gtactgatgg	cggtttccca	8280
tctaaccgaa	tccatgaacc	gataccggga	agggaaaggga	gacaagcccc	gccgcgtgtt	8340
ccgtccacac	gttgcggaag	tactcaagtt	ctgccggcga	gccgatggcg	gaaagcagaa	8400
agacgacctg	gtagaaacct	gcattcgggt	aaacaccacg	cacgttgcca	tgacgcgtac	8460
gaagaaggcc	aagaacggcc	gcctggtgac	ggttaaccgag	ggatgaagcct	tgattagccg	8520
ctacaagatc	gtaaagagcg	aaaccgggcg	gccggagtag	atcgagatcg	agctagctga	8580
ttggatgtac	cgcgagatca	cagaaggcaa	gaaccgggac	gtgctgacgg	ttcaccgccga	8640
ttactttttg	atcgatcccc	gcateggccg	ttttctctac	cgcttggcac	gccgcgcgcg	8700
aggcaaggca	gaagccagat	ggttggttcaa	gacgatctac	gaacgcagtg	gcagcgcgcg	8760
agagttcaag	aagttctgtt	tcaccgtgcg	caagctgacg	gggtcaaatg	acctgccgga	8820

gtacgatttg	aaggaggagg	cggggcaggc	tggcccgatc	ctagtcatgc	gctaccgcaa	8880
cctgatcgag	ggcgaagcat	ccgcgggttc	ctaattgtacg	gagcagatgc	tagggcaaat	8940
tgccttagca	ggggaaaaag	gtcgaaaagg	tctcttttct	gtggatagca	cgtacatttg	9000
gaacccaaag	ccgtacattg	ggaaccggaa	cccgtacatt	gggaacccaa	agccgtacat	9060
tgggaaccgg	tcacacatgt	aagtgactga	tataaaagag	aaaaaaggcg	atTTTTccgc	9120
ctaaaactct	ttaaaactta	ttaaaactct	taaaaccgcg	ctggcctgtg	cataactgtc	9180
tggccagcgc	acagccgaag	agctgcaaaa	agcgcctacc	cttcggtcgc	tgcgtccctt	9240
acgccccgcc	gcttcgcgtc	ggcctatcgc	ggcgcgtggc	cgtcaaaaa	tggctggcct	9300
acggccaggc	aatctaccag	ggcgcggaca	agccgcgcgc	tgcacctcg	accgccggcg	9360
ctgaggtctg	cctcgtgaag	aaggtgttgc	tgactcatac	caggcctgaa	tgcccccatc	9420
atccagccag	aaagtgaggg	agccacgggt	gatgagagct	ttgttgtagg	tggaccagtt	9480
ggtgatTTTT	aacttttgc	ttgccacgga	acggtctgcg	ttgtcgggaa	gatgcgtgat	9540
ctgatccctc	aactcagcaa	aagtctcgatt	tattcaacaa	agccgcgcgc	ccgtcaagtc	9600
agcgtaatgc	tctgccagtg	ttacaaccaa	ttaaccaatt	ctgattagaa	aaactcatcg	9660
agcatcaaat	gaaactgcaa	tttattcata	tcaggattat	caataccata	tttttgaaaa	9720
agccgtttct	gtaatgaagg	agaaaactca	ccgaggcagt	tccataggat	ggcaagatcc	9780
tggatcgggt	ctgcgattcc	gactcgtcca	acatcaatac	aacctattaa	tttccccctg	9840
tcaaaaaataa	ggttatcaag	tgagaaatca	ccatgagtga	cgaactgaatc	cggtgagaat	9900
ggcaaaagct	ctgcattaat	gaatcggcca	acgcgcgggg	agaggcgggt	tgcgtatttg	9960
gcgctcttcc	gcttctcgc	tcactgactc	gctgcgctcg	gtcgttcggc	tgcggcgagc	10020
ggtatcagct	cactcaaagg	cggtaatacg	gttatccaca	gaatcagggg	ataacgcagg	10080
aaagaacatg	tgagcaaaaag	gccagcaaaa	ggccagggaac	cgtaaaaaag	ccgcgttgct	10140
ggcgtttttc	cataggctcc	gccccctga	cgagcatcac	aaaaatcgac	gctcaagtca	10200
gaggtggcga	aacccgacag	gactataaag	ataccaggcg	tttccccctg	gaagctccct	10260
cgtgcgctct	cctgttccga	ccctgcgcgt	taccggatac	ctgtccgcct	ttctcccttc	10320
gggaagcgtg	gcgctttctc	aatgctcacg	ctgtagggtat	ctcagttcgg	tgtaggtcgt	10380
tgcctccaag	ctgggctgtg	tgcacgaacc	ccccgttcag	cccgaccgct	gcgccttatt	10440
cggtaactat	cgtcttgagt	ccaacccggg	aagacacgac	ttatcgccac	tggcagcagc	10500
cactggtaac	aggattagca	gagcgaggta	tgtaggcggg	gctacagagt	tcttgaagtg	10560
gtggcctaac	tacggctaca	ctagaaggac	agtatttggt	atctgcgctc	tgctgaagcc	10620
agttaccttc	ggaaaaagag	ttggtagctc	ttgatccggc	aaacaaacca	ccgctggtag	10680
cggtggtttt	tttgtttgca	agcagcagat	tacgcgcaga	aaaaaaggat	ctcaagaaga	10740
tcctttgatc	ttttctacgg	ggtctgacgc	tcagtggaac	gaaaactcac	gttaagggat	10800
tttggtcatg	agattatcaa	aaaggatctt	cacctagatc	cttttgatcc	ggaatttaatt	10860
cctgtggttg	gcatgcacat	acaaatggac	gaacggataa	acctttttcac	gcccttttaa	10920
atatccgatt	attctaataa	acgctctttt	ctcttagggt	taccgcgcaa	tatatcctgt	10980
caaacactga	tagtttaaac	tgaaggcggg	aaacgacaat	ctgatcatga	gcggagaatt	11040
aagggagtca	cgttatgacc	cccgccgatg	acgcgggaca	agccgtttta	cgtttggaac	11100
tgacagaacc	gcaacgctgc	aggaatt				11127