

[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 200410031491.2

[51] Int. Cl.

C12P 21/00 (2006.01)

C12N 15/80 (2006.01)

C12N 9/02 (2006.01)

C12N 9/10 (2006.01)

C12N 9/14 (2006.01)

C12N 9/88 (2006.01)

[45] 授权公告日 2009 年 7 月 8 日

[11] 授权公告号 CN 100510096C

[51] Int. Cl. (续)

C12N 9/90 (2006.01)

C12N 15/70 (2006.01)

C12N 9/00 (2006.01)

C12R 1/645 (2006.01)

[22] 申请日 2000.3.22

[21] 申请号 200410031491.2

分案原申请号 00805332.4

[30] 优先权

[32] 1999.3.22 [33] US [31] 09/274, 449

[73] 专利权人 诺沃奇梅兹有限公司

地址 美国加利福尼亚州

[72] 发明人 R·M·伯卡 M·W·雷伊

K·布朗 S·H·布朗

[56] 参考文献

WO9600787A 1996.1.11

审查员 孔佳音

[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

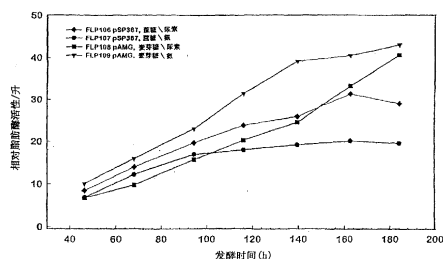
权利要求书 3 页 说明书 62 页 附图 24 页

[54] 发明名称

用于在真菌细胞中表达基因的启动子

[57] 摘要

本发明涉及制备多肽的方法，包括：(a) 在有利于产生该多肽的培养基中培养真菌宿主细胞，其中该真菌宿主细胞含有编码该多肽的第一核苷酸序列和与该第一核苷酸序列可操作地连接的、含有与第一核苷酸序列异源的启动子的第二核苷酸序列，其中该启动子含有选自下组的序列：SEQ ID NO.1 第 1-3949 位核苷酸、SEQ ID NO.2 第 1-938 位核苷酸和 SEQ ID NO.3 第 1-3060 位核苷酸，及其子序列；和它们的突变、杂合及串联启动子；和(b) 从该培养基分离该多肽。本发明还涉及该分离的启动子序列、以及含有与编码多肽的核苷酸序列可操作地连接的该启动子序列的结构、载体和真菌宿主细胞。



1. 制备多肽的方法，包括：

(a) 在适于该多肽产生的培养基中培养真菌宿主细胞，其中该真菌宿主细胞含有编码该多肽的第一核酸序列和与该第一核酸序列可操作地连接的、含有与第一核酸序列异源的启动子的第二核酸序列，其中该启动子由以下核酸序列构成：(i) SEQ ID NO.2 的第 1-938 位核苷酸；和

(b) 从该培养基分离该多肽。

2. 权利要求 1 的方法，其中该启动子由大肠杆菌 NRRL B-30067 中的质粒 pECO3 所包含的核酸序列构成。

3. 制备多肽的方法，包括：

(a) 在适于该多肽产生的培养基中培养真菌宿主细胞，其中该真菌宿主细胞含有编码该多肽的第一核酸序列和与该第一核酸序列可操作地连接的、含有与第一核酸序列异源的启动子的第二核酸序列，其中所述启动子是串联启动子，该串联启动子含有以下核酸序列的一个或多个拷贝：(i) SEQ ID NO.2 的第 1-938 位核苷酸；和

(b) 从该培养基分离该多肽。

4. 权利要求 3 的方法，其中该串联启动子还含有一个或多个不同启动子。

5. 权利要求 1 或 3 的方法，其中该真菌宿主细胞含有一个或多个拷贝的第一核酸序列。

6. 权利要求 1 或 3 的方法，其中所述第一核酸序列编码与该真菌宿主细胞异源的多肽。

7. 权利要求 1 或 3 的方法，其中该多肽是激素、酶、受体、抗体或其部分、或报道分子。
8. 权利要求 7 的方法，其中该酶是氧化还原酶、转移酶、水解酶、裂合酶、异构酶或连接酶。
9. 权利要求 1 或 3 的方法，其中该真菌宿主细胞是丝状真菌细胞或酵母细胞。
10. 权利要求 9 的方法，其中该丝状真菌细胞是 *Acremonium*、曲霉属、镰孢霉属、腐质霉属、毛霉属、毁丝霉属、脉孢菌属、青霉属、梭孢壳属、*Tolypocladium* 或木霉属的细胞。
11. 权利要求 9 的方法，其中该酵母细胞是假丝酵母属、汉逊酵母属、克鲁维酵母属、毕赤酵母属、酵母属、裂殖酵母属或 *Yarrowia* 的细胞。
12. 权利要求 9 的方法，其中该丝状真菌宿主细胞是曲霉属细胞。
13. 权利要求 9 的方法，其中该丝状真菌宿主细胞是镰孢霉属细胞。
14. 分离的启动子序列，其由以下核酸序列构成：(i) SEQ ID NO.2 的第 1-938 位核苷酸。
15. 权利要求 14 的分离启动子序列，其由大肠杆菌 NRRL B-30067 中的质粒 pECO3 所包含的核酸序列构成。
16. 核酸结构，其含有与权利要求 14 的启动子可操作地连接的编码多肽的核酸序列。

17. 含有权利要求 16 的核酸结构的重组表达载体。
18. 含有权利要求 16 的核酸结构的重组宿主细胞。

用于在真菌细胞中表达基因的启动子

发明背景

发明领域

本发明涉及制备多肽的方法。本发明还涉及分离的启动子，以及含有与编码多肽的核酸序列可操作连接的该启动子的核酸结构、载体和宿主细胞。

相关领域的描述

在真菌宿主细胞，尤其是丝状真菌细胞例如曲霉属 (*Aspergillus*) 中重组制备异源蛋白质可能为以商业上有意义的量生产蛋白质提供了更为理想的载体 (vehicle)。

异源蛋白质的重组制备通过构建如下表达盒来实现，在该表达盒中编码该蛋白质的 DNA 被置于适用于该宿主细胞的从被调节的基因中切下的启动子的表达控制之下。通常通过质粒介导的转化，将该表达盒导入该宿主细胞中。然后在表达盒所含启动子正常发挥功能所需的诱导条件下，通过培养该转化宿主细胞实现该异源蛋白质的产生。

开发用于蛋白质重组制备的新真菌宿主细胞通常要获得适于在该宿主细胞中控制蛋白质表达的启动子。已经显示，*Fusarium venenatum* 可以作为一种新的宿主细胞用于该类表达 (WO 96/00787, WO 97/26330)。而且，对于在 *Fusarium venenatum* 宿主细胞中表达异源基因有用的来自尖镰孢 (*Fusarium oxysporum*) 胰蛋白酶样蛋白酶基因的启动子，也已有过描述 (US 5,837,847)。

然而，在本领域中仍然需要用于控制异源基因表达的新启动子。

本发明的目的是，提供用于在真菌宿主细胞中制备多肽的改良方法和用于该制备的新启动子。

发明概述

本发明涉及制备多肽的方法，包括：(a)在有利于产生该多肽的培养基中培养真菌宿主细胞，其中该真菌宿主细胞含有编码该多肽的第一核酸序列和与该第一核酸序列可操作地连接的、含有与第一核酸序列异源的启动子的第二核酸序列，其中该启动子含有选自下组的序列：SEQ ID NO. 1 的第 1 - 3949 位核苷酸、SEQ ID NO. 2 的第 1 - 938 位核苷酸和 SEQ ID NO. 3 的第 1 - 3060 位核苷酸，及其子序列；和它们的突变、杂合及串联启动子；和(b)从该培养基分离该多肽。

本发明还涉及分离的启动子序列，以及含有与编码多肽的核酸序列可操作连接的一或多个该启动子的结构、载体和真菌宿主细胞。

本发明还涉及获得 SEQ ID NO. 1 第 1 - 3949 位核苷酸、SEQ ID NO. 2 第 1 - 938 位核苷酸和 SEQ ID NO. 3 第 1 - 3060 位核苷酸的突变启动子的方法。

本发明还涉及分离的核酸序列，其选自：

(a) 编码具有如下氨基酸序列的多肽的核酸序列，该氨基酸序列与 SEQ ID NO. 4 第 22 - 581 位氨基酸有至少 65% 的一致性、与 SEQ ID NO. 5 第 19 - 200 位氨基酸有至少 65% 的一致性、或与 SEQ ID NO. 6 第 1 - 187 位氨基酸有至少 65% 的一致性；

(b) 与 SEQ ID NO. 1 第 4013 - 5743 位核苷酸有至少 65% 的同源性、与 SEQ ID NO. 2 第 993 - 1593 位核苷酸有至少 65% 的同源性、或与 SEQ ID NO. 3 第 3061 - 3678 位核苷酸有至少 65% 的同源性的核酸序列；

(c) 在极低、低、中等、中高、高或极高严谨条件下与以下序列杂交的核酸序列：(i) SEQ ID NO. 1 第 4013 - 5743 位核苷酸、SEQ ID NO. 2 第 993 - 1593 位核苷酸、或 SEQ ID NO. 3 第 3061 - 3678 位核苷酸；(ii) 在 SEQ ID NO. 1 第 4013 - 5743 位核苷酸、SEQ ID NO. 2 第 993 - 1593 位核苷酸、或 SEQ ID NO. 3 第 3061 - 3678 位核苷酸中包含的 cDNA 序列；(iii) 具有至少 100 个核苷酸的(i)或(ii)的子序列；或(iv) (i)、(ii)或(iii)的互补链；

(d) 编码具有 SEQ ID NO. 4、SEQ ID NO. 5 或 SEQ ID NO. 6 之氨基酸

序列的多肽的变体的核酸序列，其中所述变体包含一或多个氨基酸的替代、缺失和/或插入；

(e) (a)、(b)或(c)的等位变体；和

(f) (a)、(b)、(c)或(e)的子序列。

本发明还涉及具有这些核酸序列的结构、载体、和重组宿主细胞。

附图简述

图 1A - F 显示 *Fusarium venenatum* 葡糖淀粉酶基因的基因组 DNA 序列和推导的氨基酸序列(分别是 SEQ ID NOs. 1 和 4)。

图 2A - C 显示命名为“Daria”的 *Fusarium venenatum* 基因的基因组 DNA 序列和推导的氨基酸序列(分别是 SEQ ID NOs. 2 和 5)。

图 3A - D 显示命名为“Quinn”的 *Fusarium venenatum* 基因的基因组 DNA 序列和推导的氨基酸序列(分别是 SEQ ID NOs. 3 和 6)。

图 4 显示 pDM181 的限制性图谱。

图 5 显示 pSheB1 的限制性图谱。

图 6 显示 pDM194 的限制性图谱。

图 7 显示 pJRoy35 的限制性图谱。

图 8 显示 pDM218 的限制性图谱。

图 9 显示 pEJG25A 的限制性图谱。

图 10 显示 pMWR60 的限制性图谱。

图 11 显示 pRaMB62 的限制性图谱。

图 12 显示 pRaMB64 的限制性图谱。

图 13 显示 pRaMB66 的限制性图谱。

图 14 显示在 *Fusarium Venenatum* 中处于 *Fusarium venenatum* 淀粉葡糖苷酶 (pAMG) 启动子和尖镰孢胰蛋白酶 (pSP387) 启动子控制下的 *Humicola lanuginosa* 脂酶报道基因表达的比较。

发明详述

本发明涉及制备多肽的方法，包括：(a)在有利于产生该多肽的培养

基中培养真菌宿主细胞，其中该真菌宿主细胞含有编码该多肽的第一核酸序列和与该第一核酸序列可操作地连接的、含有与第一核酸序列异源的启动子的第二核酸序列，其中该启动子含有选自下组的序列：SEQ ID NO. 1 第 1 - 3949 位核苷酸、SEQ ID NO. 2 第 1 - 938 位核苷酸和 SEQ ID NO. 3 第 1 - 3060 位核苷酸，及其子序列；和它们的突变、杂合及串联启动子；和 (b) 从该培养基分离该多肽。

在本发明的制备方法中，采用本领域已知的方法在适于所述多肽产生的营养培养基中培养所述细胞。例如，可以在适当培养基中、在允许该多肽表达和/或分离的条件下通过摇瓶培养、实验室或工业发酵罐中的小规模或大规模发酵（包括连续、批式、补料分批或固态发酵），来培养该细胞。该培养在含有碳和氮源以及无机盐的适合营养培养基中采用本领域已知的程序进行。适合的培养基可从商业厂家获得，也可根据（如美国典型培养物保藏中心目录中的）公开的组成制备。如果该多肽分泌至营养培养基中，则可以直接从该培养基中回收该多肽。如果该多肽是不分泌的，则可以从细胞裂解物中进行回收。

该多肽可采用特异针对该多肽的本领域已知方法来检测。这些检测方法可以包括特异性抗体的使用、酶产物的形成、或酶底物消失。

所获得的多肽可通过本领域已知的方法回收。例如，可通过传统方法，包括但不限于离心、过滤、提取、喷雾干燥、蒸发或沉淀，从营养培养基中回收所述多肽。

该多肽可以通过本领域已知的多种方法来纯化，这些方法包括但不限于层析（如离子交换、亲和、疏水、层析聚焦和大小排阻层析）、电泳方法（如制备型等电聚焦）、差异溶解度（如硫酸铵沉淀）、SDS-PAGE 或抽提（见例如，蛋白质纯化（Protein Purification），J.-C. Janson 和 Lars Ryden 编著，VCH Publishers，纽约，1989）。

启动子

术语“启动子”在本文中定义为结合 RNA 聚合酶并指导该聚合酶到达编码多肽的核酸序列的正确下游转录起始位点以起始转录的 DNA 序列。

RNA 聚合酶有效地催化与编码区适当 DNA 链互补的信使 RNA 的组装。术语“启动子”也应理解为包括在转录成 mRNA 之后用于翻译的(启动子和翻译起始位点之间的)5'非编码区、顺式作用转录控制元件例如增强子、和其它能够与转录因子相互作用的核苷酸序列。

术语“突变启动子”在本文中定义为具有其中含有一或多个核苷酸替代、缺失和/或插入的亲本启动子核苷酸序列的启动子,其中该突变启动子比相应的亲本启动子具有较强或较弱的启动子活性。术语“突变启动子”也将包括自然变体,和采用经典诱变、定点诱变和 DNA 改组等本领域熟知的方法获得的体外制备变体。

术语“杂合启动子”在本文中定义为两个或多个启动子的部分,它们融合在一起形成两个或更多个启动子的融合物序列,该序列与编码序列可操作地连接并介导该编码序列转录成 mRNA。

术语“串联启动子”在本文中定义为每一个均与一个编码序列可操作地连接并介导该编码序列转录成 mRNA 的两个或多个启动子序列。

术语“可操作地连接”在本文中定义为一种如下构象,在该构象中控制序列例如启动子序列被适当地置于相对于编码序列的一个位置上,以致该控制序列指导该编码序列所编码的多肽的产生。

术语“编码序列”在本文中定义为当被置于适当控制序列的控制之下时,可以转录为能翻译成多肽的 mRNA 的核酸序列。编码序列的边界一般由 mRNA 5'端刚好位于该开放阅读框上游的 ATG 起始密码子,和 mRNA 3'端刚好位于该开放阅读框下游的转录终止序列来确定。编码序列可以包括,但不限于基因组 DNA、cDNA、半合成的、合成的、和重组的核酸序列。

在一个优选的实施方案中,所述启动子具有 SEQ ID NO.1 第 1 - 3949 位核苷酸的核酸序列,或其子序列。该子序列优选含有至少大约 2100 个核苷酸,更优选至少大约 2400 个核苷酸,最优选至少大约 2700 个核苷酸。

在另一个优选实施方案中,所述启动子具有 SEQ ID NO.2 第 1 - 938 位核苷酸的核酸序列,或其子序列。该子序列优选含有至少大约 840 个核苷酸,更优选至少大约 870 个核苷酸,最优选至少大约 900 个核苷酸。

在另一个优选实施方案中,所述启动子具有 SEQ ID NO.3 第 1-3060 位核苷酸的核酸序列,或其子序列。该子序列优选含有至少大约 2100 个核苷酸,更优选至少大约 2400 个核苷酸,最优选至少大约 2700 个核苷酸。

在另一个优选实施方案中,所述启动子是大肠杆菌 (*Escherichia coli*) NRRL B-30067 中的质粒 pEC03 所包含的核酸序列。在另一个优选实施方案中,该启动子是大肠杆菌 NRRL B-30071 中的质粒 pFAMG 所包含的核酸序列。在另一个优选实施方案中,该启动子是大肠杆菌 NRRL B-30075 中的质粒 pQUINN 所包含的核酸。

在本发明的方法中,所述启动子还可以是在 SEQ ID NO.1 第 1-3949 位核苷酸、SEQ ID NO.2 第 1-938 位核苷酸、或 SEQ ID NO.3 第 1-3060 位核苷酸的核酸序列中含有一或多个核苷酸替代、缺失、和/或插入的上述启动子的突变体。

突变启动子可以含有一或多个突变。每个突变均是独立的核苷酸替代、缺失和/或插入。可以采用本领域已知的任何方法,例如经典诱变、定点诱变或 DNA 改组来实现向所述启动子中引入核苷酸替代、缺失和/或插入。尤其有用的一种方法是利用带有目的插入片段的超螺旋双链 DNA 载体和两个含有期望突变的合成引物来进行的。该寡核苷酸引物每一个均与该载体的相反链互补,在温度循环中通过 pfu DNA 聚合物的作用延伸。一旦这两个引物掺入后,就产生含有交错切口的突变质粒。在温度循环后,用对于甲基化和半甲基化 DNA 特异的 DpnI 处理该产物,以消化亲本 DNA 模板并选出含有突变的合成 DNA。也可以采用本领域已知的其它方法。

在一个优选实施方案中,所述启动子是 SEQ ID NO.1 第 1-3949 位核苷酸的突变体。在另一个优选实施方案中,所述启动子是 SEQ ID NO.2 第 1-938 位核苷酸的突变体。在另一个优选实施方案中,所述启动子是 SEQ ID NO.3 第 1-3060 位核苷酸的突变体。

在本发明的方法中,所述启动子还可以是杂合启动子,其含有一或多个本发明启动子的部分;一个本发明启动子的部分和另一个启动子的部

分, 例如一个启动子的前导序列和来自另一个启动子的转录起始位点; 或一或多个本发明启动子的部分和一或多个其它启动子的部分。所述其它启动子可以是在所选真菌宿主细胞中表现出转录活性的任何启动子序列, 包括突变的、截短的和杂合的启动子, 并且可以从编码与该宿主细胞同源或异源的细胞外或细胞内多肽的基因获得。该其它启动子对于编码所述多肽的核酸序列而言可以是自身的或外来的, 并且对于该细胞而言也可以是自身的或外来的。

用于和本发明的启动子一起构建杂合启动子的其它启动子的例子包括从米曲霉(*Aspergillus oryzae*) TAKA 淀粉酶、*Rhizomucor miehei* 天冬氨酸蛋白酶、黑曲霉(*Aspergillus niger*) 中性 α -淀粉酶、黑曲霉酸稳定性 α -淀粉酶、黑曲霉或泡盛曲霉(*Aspergillus awamori*) 葡糖淀粉酶(*glaA*)、*Rhizomucor miehei* 脂肪酶、米曲霉碱性蛋白酶、米曲霉丙糖磷酸异构酶、构巢曲霉(*Aspergillus nidulans*) 乙酰胺酶和尖镰孢胰蛋白酶样蛋白酶(WO 96/00787)的基因获得的启动子, 以及 NA2-tpi 启动子(黑曲霉中性 α -淀粉酶和米曲霉丙糖磷酸异构酶的基因启动子的杂合物), 和从酿酒酵母(*Saccharomyces cerevisiae*) 烯醇化酶(ENO-1)、酿酒酵母半乳糖激酶(GAL1)、酿酒酵母醇脱氢酶/甘油醛-3-磷酸脱氢酶(ADH2/GAP)和酿酒酵母 3-磷酸甘油酸激酶的基因获得的启动子, 及这些启动子的突变的、截短的和杂合的启动子。用于酵母宿主细胞的其它有用启动子描述于 Romanos 等, 1992, 酵母(Yeast) 8: 423-488。

所述启动子还可以是含有两个或多个本发明启动子或者是含有一或多个本发明启动子和一或多个诸如以上所列举的其它启动子的串联启动子。该串联启动子的两个或多个启动子序列可以同时启动所述核酸序列的转录。或者, 该串联启动子的一或多个启动子序列可以在细胞生长的不同阶段启动所述核酸序列的转录。

在本发明的方法中, 该启动子对于编码目的多肽的核酸序列而言是异源的, 但该启动子或核酸序列对于该真菌宿主细胞而言可以是天然的。即使野生型启动子对于编码多肽的核酸序列而言是自身的, 本发明的突变、杂合、或串联启动子也应理解为与该核酸序列异源。

本发明的突变、杂合或串联启动子具有 SEQ ID NO.1 第 1-3949 位核苷酸、SEQ ID NO.2 第 1-938 位核苷酸或 SEQ ID NO.3 第 1-3060 位核苷酸启动子的启动子活性的至少大约 20%、优选至少大约 40%、更优选至少大约 60%、更优选至少大约 80%、更优选至少大约 90%、更优选至少大约 100%、甚至更优选至少大约 200%、最优选至少大约 300%、甚至最优选至少大约 400%。

编码多肽的核酸序列

由所述核酸序列编码的多肽对于目的真菌宿主细胞而言可以是自身的或异源的。

术语“多肽”在此并非旨在指特定长度的编码产物，因此包括肽、寡肽和蛋白质。术语“异源多肽”在本文中定义为非该真菌细胞天然的多肽，已经过修饰改变了天然序列的天然多肽，或其表达由于重组 DNA 技术对宿主细胞的操作而在数量上改变的天然多肽。例如，天然多肽可以通过例如将编码该多肽的基因置于本发明的启动子控制之下重组产生以增强该多肽的表达、借助信号序列促进目的天然多肽向细胞外运输，和增加编码细胞正常所产生的该多肽的基因的拷贝数。该真菌细胞可以含有一或多个拷贝的编码该多肽的核酸序列。

优选地，该多肽是激素或其变体、酶、受体或其部分、抗体或其部分、或报道分子。在一个优选的实施方案中，该多肽是向细胞外分泌的。在一个更优选的实施方案中，该多肽是氧化还原酶、转移酶、水解酶、裂合酶、异构酶或连接酶。在一个甚至更优选的实施方案中，该多肽是氨肽酶、淀粉酶、糖酶、羧肽酶、过氧化氢酶、纤维素酶、几丁质酶、角质酶、环糊精糖基转移酶、脱氧核糖核酸酶、酯酶、 α -半乳糖苷酶、 β -半乳糖苷酶、葡糖淀粉酶、 α -葡萄糖苷酶、 β -葡萄糖苷酶、转化酶（invertase）、漆酶、脂肪酶、甘露糖苷酶、mutanase、氧化酶、果胶溶酶、过氧化物酶、磷脂酶、肌醇六磷酸酶、多酚氧化酶、蛋白水解酶、核糖核酸酶、谷氨酰胺转移酶（transglutaminase）或木聚糖酶。

编码目的多肽的核酸序列可以从任何原核、真核或其它来源获得。为

了本发明的目的，本文中与所述来源相连使用的术语“获自”是指该多肽是由该来源或已插入了该来源的基因的细胞产生的。

用于分离或克隆编码目的多肽的核酸序列的技术是本领域已知的，包括从基因组 DNA 分离，从 cDNA 制备，或它们的组合。从该基因组 DNA 克隆该核酸序列可以通过例如采用熟知的聚合酶链式反应（PCR）来实现。见例如，Innis 等，1990，PCR 操作：方法和应用指南（PCR Protocols: A Guide to Methods and Application），Academic Press，纽约。该克隆方法可能包括：切割和分离含有编码该多肽的核酸序列的期望核酸片段，将该片段插入载体分子中，并将该重组载体掺入可以复制出多个拷贝或克隆的该核酸序列的突变真菌细胞中。该核酸序列可以是基因组、cDNA、RNA、半合成、合成来源的或它们的任意组合。

在本发明方法中，所述多肽还可以包括其中在该多肽或其片段的 N 端或 C 端融合了另一个多肽的融合或杂合多肽。融合多肽是通过将编码一个多肽的核酸序列（或其部分）与编码另一个多肽的核酸序列（或其部分）融合在一起产生的。产生融合多肽的技术是本领域已知的，包括连接编码这些多肽的编码序列以使它们符合阅读框，并且使该融合多肽的表达处于相同的一个或多个启动子和相同的终止子的控制之下。所述杂合多肽可以含有从至少两个不同多肽获得的部分或完整多肽序列的组合，其中的一或多个多肽对于该突变真菌细胞可以是异源的。

核酸结构

本发明还涉及含有与本发明的启动子和一个或多个控制序列可操作地连接的编码多肽的核酸序列的核酸结构，其中所述控制序列可在适合的宿主细胞中在与该控制序列相容的条件下指导该编码序列的表达。表达将理解为包括多肽产生所涉及到的任何步骤，包括但不限于转录、转录后修饰、翻译、翻译后修饰和分泌。

“核酸结构”在本文中定义为单链或双链核酸分子，该核酸分子可以是天然存在的基因分离的，或是已被修饰而含有以天然不存在的方式组合和并置在一起的核酸片段。当核酸结构含有编码序列以及表达该

编码序列所必需的所有控制序列时，术语核酸结构与术语表达盒同义。

可以通过多种方式进一步操作编码多肽的分离核酸序列，以为该多肽的表达作准备。依据表达载体，在核酸序列插入载体之前对其进行操作可能是期望的或是必需的。利用重组 DNA 方法修饰核酸序列的技术是本领域熟知的。

在本发明的方法中，核酸序列可以含有一个或多个天然控制序列，或可以用该核酸序列外源的一或多个控制序列来取代该一个或多个天然控制序列，以改善编码序列在宿主细胞中的表达。

术语“控制序列”在本文中定义为包括所有对本发明多肽的表达而言必需或有利的成分。每个控制序列对于编码该多肽的核酸序列而言可以是自身的或是外源的。这些控制序列包括但不限于前导序列、聚腺苷酸化序列、前肽序列、本发明的启动子、信号肽序列和转录终止子。最少地，控制序列包括本发明的启动子，以及转录和翻译终止信号。控制序列可与接头一起提供，以便引入特异的限制性位点，便于控制序列与编码多肽的核酸序列编码区连接。

控制序列可以是适合的转录终止子序列，即被宿主细胞识别以终止转录的序列。终止子序列与编码所述多肽的核酸序列的 3' 末端可操作地连接。任何在所选宿主细胞中起作用的终止子都可用于本发明。

用于丝状真菌宿主细胞的优选终止子可获自米曲霉 TAKA 淀粉酶、黑曲霉葡萄糖淀粉酶、构巢曲霉邻氨基苯甲酸合成酶、黑曲霉 α -葡萄糖苷酶和尖镰孢胰蛋白酶样蛋白酶的基因。

用于酵母宿主细胞的优选终止子可获自酿酒酵母烯醇化酶、酿酒酵母细胞色素 C (CYC1) 和酿酒酵母甘油醛-3-磷酸脱氢酶的基因。用于酵母宿主细胞的其它有用的终止子描述于 Romanos 等，1992，同上。

控制序列也可以是适当的前导序列，即一种对宿主细胞翻译重要的 mRNA 非翻译区。前导序列与编码所述多肽的核酸序列的 5' 端可操作地连接。任何在所选宿主细胞中有功能的前导序列都可用于本发明。

用于丝状真菌宿主细胞的优选前导序列可从米曲霉 TAKA 淀粉酶和构巢曲霉丙糖磷酸异构酶的基因获得。

适用于酵母宿主细胞的前导序列可从酿酒酵母烯醇化酶(ENO-1)、酿酒酵母 3-磷酸甘油酸激酶、酿酒酵母 α -因子和酿酒酵母醇脱氢酶/甘油醛-3-磷酸脱氢酶(ADH2/GAP)的基因获得。

控制序列也可以是聚腺苷酸化序列,即一种与所述核酸序列 3'末端可操作地连接的、转录时被宿主细胞作为向转录的 mRNA 添加聚腺苷酸残基(polyadenosine residues)的信号而识别的序列。任何在所选宿主细胞中有作用的聚腺苷酸化序列均可用于本发明。

用于丝状真菌宿主细胞的优选聚腺苷酸化序列可从米曲霉 TAKA 淀粉酶、黑曲霉葡糖淀粉酶、构巢曲霉邻氨基苯甲酸合成酶、尖镰孢胰蛋白酶样蛋白酶和黑曲霉 α -葡糖苷酶的基因获得。

用于酵母宿主细胞的有用聚腺苷酸化序列描述于 Guo 和 Sherman, 1995, 分子细胞生物学(Molecular Cellular Biology) 15: 5983-5990。

控制序列还可以是编码与多肽氨基末端连接的氨基酸序列、并指导该编码多肽进入细胞分泌途径的信号肽编码区。所述核酸序列编码区的 5'末端本身可含有在翻译阅读框中天然与编码该分泌多肽的编码区片段连接的信号肽编码区。或者,编码序列的 5'末端可含有对该编码序列而言外源的信号肽编码区。在编码序列并不天然含有信号肽编码区的时候可能需要外源信号肽编码区。或者,为了增强多肽的分泌,可以简单地用外源信号肽编码区替代天然信号肽编码区。然而,任何指导表达多肽进入所选宿主细胞分泌途径的信号肽编码区均可用于本发明。

用于丝状真菌宿主细胞的有效信号肽编码区是从米曲霉 TAKA 淀粉酶、黑曲霉中性淀粉酶、黑曲霉葡糖淀粉酶、Rhizomucor miehei 天冬氨酸蛋白酶、Humicola insolens 纤维素酶和 Humicola lanuginosa 脂肪酶的基因获得的信号肽编码区。

用于酵母宿主细胞的有用信号肽可从酿酒酵母 α -因子和酿酒酵母转化酶(invertase)的基因获得。其它有用的信号肽编码区描述于 Romanos 等, 1992, 同上。

控制序列还可以是编码位于多肽氨基端的氨基酸序列的前肽编码区。所获得的多肽被称为酶原(proenzyme)和多肽原(有时也叫酶原

(zymogen))。多肽原通常是无活性的,但能通过前肽的催化切割或自身催化切割从多肽原转化成成熟活性多肽。前肽编码区可从枯草芽孢杆菌(*Bacillus subtilis*)碱性蛋白酶(aprE)、枯草芽孢杆菌中性蛋白酶(nprT)、酿酒酵母 α -因子、*Rhizomucor miehei*天冬氨酸蛋白酶和*Myceliophthora thermophila*漆酶(WO 95/33836)的基因获得。

当多肽的氨基端同时存在信号肽和前肽区时,前肽区紧接位于多肽的氨基端,而信号肽区紧接位于前肽区的氨基端。

也可以期望加入允许相对于宿主细胞生长调节所述多肽表达的调节序列。调节系统的例子是那些使基因表达响应化学或物理刺激包括调节化合物的存在而打开或关闭的系统。原核系统中的调节系统包括 lac、tac 和 trp 操纵子系统。在酵母中,可以使用 ADH2 系统或 GAL1 系统。在丝状真菌中,可以使用 TAKA α -淀粉酶启动子、黑曲霉葡萄糖淀粉酶启动子和米曲霉葡萄糖淀粉酶启动子作为调节序列。调节序列的其它例子是那些允许基因扩增的序列。在真核系统中,这些包括在氨甲蝶呤存在时扩增的二氢叶酸还原酶基因和在重金属存在时扩增的金属硫蛋白基因。在这些情况中,编码所述多肽的核酸序列将与调节序列可操作地连接。

本发明还涉及用于改变编码多肽的宿主细胞内源性基因的表达的核酸结构。这些结构可以含有改变该内源基因表达所必需的最少成分。在一个实施方案中,这些核酸结构优选含有(a)导向序列,(b)本发明的启动子,(c)外显子,和(d)剪接供体位点。一旦该核酸结构导入细胞后,该结构即通过同源重组在内源基因位点插入细胞的基因组中。导向序列指导元件(a)-(d)整合至内源基因中,以使元件(b)-(d)与该内源基因可操作地连接。在另一个实施方案中,这些核酸结构含有(a)导向序列,(b)本发明的启动子,(c)外显子,(d)剪接供体位点,(e)内含子,和(f)剪接受体位点,其中该导向序列指导元件(a)-(f)的整合,以使元件(b)-(f)与内源基因可操作地连接。然而,这些结构还可以含有其它成分如选择标记。

在这两个实施方案中,这些成分的引入导致产生了改变内源基因表达的新转录单元。大体上,此新转录单元是通过这些导向结构引入的序列

和内源基因的融合产物。在一个改变内源基因的实施方案中，该基因被激活。在该实施方案中，采用同源重组，通过插入使内源基因以比在相应的亲本细胞中更高的水平表达的调节序列，替代、破坏或失效正常与亲代细胞的该内源基因有关的调节区。采用本领域熟知的方法（见例如，美国专利 5,641,670），通过在该结构中包含可扩增的选择标记基因，可以进一步扩增该激活的基因。在另一个改变内源基因的实施方案中，该基因的表达被降低。

所述导向序列可位于所述内源基因内，紧邻该基因，在上游基因内，或在所述内源基因的上游并相隔一段距离。可使用一个或多个导向序列。例如，环状质粒或 DNA 片段优选使用单个导向序列，而线性质粒或 DNA 片段优选使用两个导向序列。

这些结构还含有所述内源基因的一个或多个外显子。外显子定义为可以被复制成 RNA 并在成熟 mRNA 分子中存在以致该外显子序列与该内源基因的编码区在阅读框上相符合的 DNA 序列。外显子可以任选地含有编码一个或多个氨基酸和/或部分编码氨基酸的 DNA。另一种情况是，外显子含有相应于 5'非编码区的 DNA。当一个或多个外源外显子编码一个或多个氨基酸和/或部分编码氨基酸时，设计该核酸结构，使得在转录和剪接后，该阅读框与内源基因的编码区在阅读框上相符合，以便来自第二外显子的 mRNA 部分的适当阅读框不改变。

这些结构的剪接供体位点指导一个外显子向另一个外显子的剪接。典型地，第一外显子位于第二外显子的 5'，而位于第一外显子 3'侧翼并与之重叠的剪接供体位点识别位于第二外显子 5'侧翼的剪接受体位点。象剪接供体位点一样，剪接受体位点也是指导一个外显子向另一个外显子剪接的序列。剪接装置联合剪接供体位点一起作用，使用剪接受体位点实现内含子的去除。

本发明还涉及制备多肽的方法，包括（a）在有益于产生该多肽的条件下，培养其中已掺入了如下新转录单元的同源重组细胞，该新转录单元含有与编码该多肽的内源核酸序列第二外显子可操作地连接的本发明启动子、外显子和/或剪接供体位点；和（b）回收该多肽。该方法基于

基因激活技术的使用，例如美国专利 5,641,670 所描述的技术。

表达载体

本发明还涉及包含本发明启动子、编码多肽的核酸序列以及转录和翻译终止信号的重组表达载体。可将以上描述的各种核酸和控制序列连接在一起，产生可包括一个或多个方便的限制性位点以允许该启动子和编码多肽的核酸序列在这些位点插入或替代的重组表达载体。或者，可以通过将该核酸序列或含有该启动子和/或序列的核酸结构插入适当的表达载体中来表达该核酸序列。在建立该表达载体时，将该编码序列定位于该载体中，使得该编码序列与本发明的启动子和一或多个适当的表达控制序列可操作地连接。

重组表达载体可以是任何能方便地用于重组 DNA 程序并能使所述核酸序列表达的载体（如质粒或病毒）。典型地，载体的选择取决于载体和载体将要导入的宿主细胞之间的相容性。该载体可以是线性或闭合环状质粒。

该载体可以是自主复制型载体，即作为染色体外实体而存在、其复制独立于染色体复制的载体，如质粒、染色体外因子、微型染色体或人工染色体。该载体可以含有任何保证自我复制的序列。或者，该载体可以是当导入宿主细胞后整合在基因组中并和它所整合的染色体一起复制的载体。而且，可使用单个载体或质粒，或一起含有待导入宿主细胞基因组的全部 DNA 的两个或多个载体或质粒，或转座子。

本发明的载体优选含有一个或多个允许方便地选择转化细胞的选择标记。选择标记是其产物提供杀生物剂或病毒抗性、重金属抗性、相对营养缺陷型的原养型等的基因。用于酵母宿主细胞的适当标记是 ADE2、HIS3、LEU2、LYS2、MET3、TRP1 和 URA3。用于丝状真菌宿主细胞的选择标记包括但不限于 amdS（乙酰胺酶）、argB（鸟氨酸氨甲酰基转移酶）、bar（蒴丝菌素乙酰基转移酶（phosphinothricin acetyltransferase））、hygB（潮霉素磷酸转移酶）、niaD（硝酸还原酶）、pyrG（乳清酸核苷-5'-磷酸脱羧酶）、sC（硫酸腺苷酰转移酶（sulfate adenylyltransferase））、

trpC (邻氨基苯甲酸合成酶), 以及它们的等同物。优选用于曲霉属细胞的是构巢曲霉或米曲霉的 amdS 和 pyrG 基因以及吸水链霉菌 (*Streptomyces hygroscopicus*) 的 bar 基因。

本发明的载体优选包含允许该载体稳定整合至宿主细胞基因组中或允许该载体在该细胞中独立于基因组而自主复制的一个或多个元件。

为了整合至宿主细胞基因组中, 该载体可以依赖于编码所述多肽的核酸序列或用于使该载体通过同源或非同源重组稳定整合至基因组中的任何其它载体元件。或者, 该载体可含有用于通过同源重组指导向宿主细胞基因组中整合的额外核酸序列。这些额外核酸序列使得该载体可以在一条或多条染色体的一个或多个精确位置整合至宿主细胞基因组中。为了增加在精确位置整合的可能性, 整合元件应优选含有足够数目的与相应靶序列同源的核酸, 如 100 - 1,500 个碱基对, 优选 400 - 1,500 个碱基对, 最优选 800 - 1,500 个碱基对, 以增强同源重组的可能性。整合元件可以是任何与宿主细胞基因组中的靶序列同源的序列。而且, 该整合元件可以是非编码或编码核酸序列。另一方面, 该载体可以通过非同源重组整合至宿主细胞的基因组中。

为了自主复制, 该载体可进一步包含使载体能在所述宿主细胞中自主复制的复制起点。用于酵母宿主细胞中的复制起点的例子是 2 μ 复制起点、ARS1、ARS4、ARS1 和 CEN3 的组合以及 ARS4 和 CEN6 的组合。复制起点可以是含有使该复制起点功能在宿主细胞中具有温度敏感性的突变的复制起点 (见例如, Ehrlich, 1978, 美国国家科学院院刊 (Proceedings of the National Academy of Sciences USA) 75: 1433)。

可在宿主细胞中插入一个拷贝以上的编码多肽的核酸序列, 以增加基因产物的产量。该核酸序列拷贝数的增加可通过如下方法获得: 在宿主细胞基因组中至少再整合一个额外拷贝的该序列; 或在细胞中包括带有该核酸序列的可扩增选择标记基因, 在适当选择剂存在时培养这些细胞, 能够选出含有扩增拷贝的选择标记基因、因此也含有额外拷贝的该核酸序列的细胞。

用于连接上述元件以构建本发明的重组表达载体的程序是本领域技术

人员熟知的（见例如，Sambrook 等，1989，同上）。

宿主细胞

本发明还涉及含有与编码多肽的核酸序列可操作地连接的本发明启动子的重组宿主细胞，这些细胞可以有利地用于该多肽的重组制备。将含有与编码多肽的核酸序列可操作地连接的本发明启动子的载体导入宿主细胞，以使该载体作为前面描述的染色体整合体或自我复制的染色体外载体而存在。术语“宿主细胞”包括任何由于复制过程中发生的突变而与亲代细胞不同的亲代细胞后代。宿主细胞的选择在很大程度上取决于编码所述多肽的基因及其来源。

宿主细胞可以是任何在本发明方法中有用的真菌细胞。本文所用“真菌”包括 phyla Ascomycota, Basidiomycota, Chytridiomycota 和 Zygomycota（定义在 Hawksworth 等，Ainsworth and Bisby's Dictionary of The Fungi, 第 8 版，1995, CAB International, University Press, Cambridge, UK 中）和以及 Oomycota（见 Hawksworth 等，1995，同上，第 171 页）和所有有丝分裂孢子真菌（Hawksworth 等，1995，见上）。

在一个优选的实施方案中，该真菌宿主细胞是酵母细胞。这里所用“酵母”包括产子囊酵母（内孢霉目（Endomycetales））、产担孢子（basidiosporogenous）酵母和属于半知菌类（Fungi Imperfecti）（芽孢纲（Blastomycetes））的酵母。因为酵母的分类将来可能改变，为了本发明的目的，酵母应按《酵母的生物学和活动》（Biology and Activities of Yeast）（Skinner, F.A., Passmore, S.M., 和 Davenport, R.R. 编著，Soc. App. Bacteriol. Symposium Series No. 9, 1980）中的描述来定义。

在一个更优选的实施方案中，酵母宿主细胞是假丝酵母属（*Candida*）、汉逊酵母属（*Hansenula*）、克鲁维酵母属（*Kluyveromyces*）、毕赤酵母属（*Pichia*）、酵母属（*Saccharomyces*）、裂殖酵母属（*Schizosaccharomyces*）或 *Yarrowia* 细胞。

在一个最优选的实施方案中，该酵母宿主细胞是卡尔酵母

(*Saccharomyces carlsbergensis*)、酿酒酵母、糖化酵母(*Saccharomyces diastaticus*)、*Saccharomyces douglasii*、克鲁弗酵母(*Saccharomyces kluyveri*)、诺地酵母(*Saccharomyces norbensis*)或 *Saccharomyces oviformis* 细胞。在另一个最优选的实施方案中,该酵母宿主细胞是乳酸克鲁维酵母(*Kluyveromyces lactis*)细胞。在另一个最优选的实施方案中,酵母宿主细胞是 *Yarrowia lipolytica* 细胞。

在另一个优选实施方案中,该真菌宿主细胞是丝状真菌细胞。“丝状真菌”包括真菌门(*Eumycota*)和 *Oomycota* 的所有丝状类型(正如 Hawksworth 等,1995,见上,所定义的)。丝状真菌的特征在于由几丁质、纤维素、葡聚糖、脱乙酰壳多糖、甘露聚糖和其它复合多糖构成的菌丝体壁。营养生长是通过菌丝延长来进行的,碳代谢为专性需氧的。相反,酵母如酿酒酵母的营养生长是通过单细胞菌体的出芽来进行的,而且碳代谢可以通过发酵进行。

在一个更优选的实施方案中,该丝状真菌宿主细胞是但不限于 *Acremonium*、曲霉属(*Aspergillus*)、镰孢霉属(*Fusarium*)、腐质霉属(*Humicola*)、毛霉属(*Mucor*)、毁丝霉属(*Myceliophthora*)、脉孢菌属(*Neurospora*)、青霉属(*Penicillium*)、梭孢壳属(*Thielavia*)、*Tolypocladium* 或木霉属(*Trichoderma*)的种的细胞。

在一个最优选的实施方案中,该丝状真菌宿主细胞是泡盛曲霉(*Aspergillus awamori*)、臭曲霉(*Aspergillus foetidus*)、日本曲霉(*Aspergillus japonicus*)、构巢曲霉(*Aspergillus nidulans*)、黑曲霉(*Aspergillus niger*)或米曲霉(*Aspergillus oryzae*)细胞。在另一个最优选的实施方案中,该丝状真菌宿主细胞是杆孢状镰孢(*Fusarium bactridioides*)、*Fusarium cerealis*、*Fusarium crookwellense*、大刀镰孢(*Fusarium culmorum*)、禾本科镰孢(*Fusarium graminearum*)、禾赤镰孢(*Fusarium graminum*)、异孢镰孢(*Fusarium heterosporum*)、合欢木镰孢(*Fusarium negundi*)、尖镰孢、多枝镰孢(*Fusarium reticulatum*)、粉红镰孢(*Fusarium roseum*)、接骨木镰孢(*Fusarium sambucinum*)、肤色镰孢(*Fusarium sarcochroum*)、拟分枝孢镰孢(*Fusarium*

sporotrichioides)、*Fusarium sulphureum*、*Fusarium torulosum*、*Fusarium trichothecioides* 或 *Fusarium venenatum* 细胞。在另一个最优的实施方案中,该丝状真菌宿主细胞是 *Humicola insolens*、*Humicola lanuginosa*、米黑毛霉(*Mucor miehei*)、*Myceliophthora thermophila*、粗糙脉孢菌(*Neurospora crassa*)、产紫青霉(*Penicillium purpurogenum*)、*Thielavia terrestris*、*Trichoderma harzianum*、康宁木霉(*Trichoderma koningii*)、*Trichoderma longibrachiatum*、*Trichoderma reesei* 或绿色木霉(*Trichoderma viride*)细胞。

在一个甚至最优的实施方案中,*Fusarium venenatum* 细胞是最初作为禾本科镰孢(*Fusarium graminearum*) ATCC 20334 保藏,最近由 Yoder 与 Christianson (1998, 真菌遗传学和生物学 (Fungal Genetics and Biology) 23:62-80) 和 O'Donnell 等 (1998, 真菌遗传学和生物学 23:57-67) 重新划分为 *Fusarium venenatum* 的 *Fusarium venenatum* A3/5; 以及 *Fusarium venenatum* 的分类学等同物,而无论它们目前被称为何种种名。在另一个优选实施方案中,正如 WO 97/26330 中所公开的,*Fusarium venenatum* 细胞是 *Fusarium venenatum* A3/5 或 *Fusarium venenatum* ATCC 20334 的形态学突变体。

可以通过涉及原生质体形成、原生质体转化和细胞壁再生的方法以本质上已知的方式来转化真菌细胞。用于转化曲霉属宿主细胞的适当程序描述于 EP 238 023 和 Yelton 等, 1984, 美国国家科学院院刊 81: 1470-1474。用于转化镰孢霉属的种的适当方法描述于 Malardier 等, 1989, 基因 (Gene) 78:147-156 和 WO 96/00787。酵母的转化可采用 Becker 和 Guarente, 由 Abelson, J.N. 和 Simon, M.I. 编, 酵母遗传学和分子生物学指南 (Guide to Yeast Genetics and Molecular Biology), 酶学方法 (Methods in Enzymology), 第 194 卷, 第 182-187 页, Academic Press, Inc., 纽约; Ito 等, 1983, 细菌学杂志 (Journal of Bacteriology) 153: 163; 和 Hinnen 等, 1978, 美国国家科学院院刊 75: 1920 中描述方法进行。

获得突变启动子的方法

本发明还涉及获得突变启动子的方法, 包括(a)在 SEQ ID NO.1 第 1 - 3949 位核苷酸、SEQ ID NO.2 第 1 - 938 位核苷酸或 SEQ ID NO.3 第 1 - 3060 位核苷酸的序列中引入至少一个突变, 其中该突变启动子具有启动子活性; 和(b)分离该突变启动子。

本发明还涉及通过以下步骤获得突变启动子的方法: (a)在极低、低、中等、中高、高或极高的严紧条件下使 DNA 与 (i)SEQ ID NO.1、SEQ ID NO.2 或 SEQ ID NO.3, (ii) SEQ ID NO.1、SEQ ID NO.2 或 SEQ ID NO.3 中所包含的 cDNA 序列, (iii) (i)或(ii)的子序列, 或(iv) (i)、(ii)或(iii)的互补链杂交; 和(b)从该 DNA 中分离出该突变启动子。本文中对严紧性和洗涤条件作了定义。

该突变启动子可以采用本领域已知的方法例如限制性酶消化或 PCR 扩增来分离。

核酸序列

本发明还涉及选自下组的分离核酸序列:

(a) 编码具有如下氨基酸序列的多肽的核酸序列, 该氨基酸序列与 SEQ ID NO.4 第 22 - 581 位氨基酸有至少 65% 的一致性、与 SEQ ID NO.5 第 19 - 200 位氨基酸有至少 65% 的一致性、或与 SEQ ID NO.6 第 1 - 187 位氨基酸有至少 65% 的一致性;

(b) 与 SEQ ID NO.1 第 4013 - 5743 位核苷酸有至少 65% 的同源性、与 SEQ ID NO.2 第 993 - 1593 位核苷酸有至少 65% 的同源性、或与 SEQ ID NO.3 第 3061 - 3678 位核苷酸有至少 65% 的同源性的核酸序列;

(c) 与 (i)SEQ ID NO.1 第 4013 - 5743 位核苷酸、SEQ ID NO.2 第 993 - 1593 位核苷酸、或 SEQ ID NO.3 第 3061 - 3678 位核苷酸; (ii) 在 SEQ ID NO.1 第 4013 - 5743 位核苷酸、SEQ ID NO.2 第 993 - 1593 位核苷酸、或 SEQ ID NO.3 第 3061 - 3678 位核苷酸中包含的 cDNA 序列; (iii) 具有至少 100 个核苷酸的 (i)或(ii)的子序列; 或(iv) (i)、(ii)或(iii)的互补链, 在极低、低、中等、中高、高或极高严紧条件下杂交的核酸序

列;

(d) 编码具有 SEQ ID NO. 4、SEQ ID NO. 5 或 SEQ ID NO. 6 之氨基酸序列的多肽的变体的核酸序列, 其中所述变体包含一或多个氨基酸的替代、缺失和/或插入;

(e) (a)、(b)或(c)的等位变体; 和

(f) (a)、(b)、(c)或(e)的子序列。

本文所用术语“分离的核酸序列”是指基本不含有其它核酸序列的核酸序列, 例如通过琼脂糖电泳确定具有至少约 20%的纯度, 优选至少约 40%的纯度, 更优选至少约 60%的纯度, 甚至更优选至少约 80%的纯度, 最优选至少约 90%的纯度的核酸序列。例如, 分离的核酸序列可以通过遗传工程中使用的标准克隆程序将该核酸序列从其天然位置重新置于其将进行复制的不同位置上来获得。该克隆程序可能涉及: 切割和分离包含编码该多肽的核酸序列的期望核酸片段, 将此片段插入载体分子中, 和将该重组载体掺入可以复制出多个拷贝或克隆的该核酸序列的宿主细胞中。该核酸序列可以是基因组、cDNA、RNA、半合成、合成来源的, 或它们的任意组合。

在第一个实施方案中, 本发明涉及编码具有与 SEQ ID NO. 4 第 22 - 581 位氨基酸、SEQ ID NO. 5 第 19 - 200 位氨基酸或 SEQ ID NO. 6 第 1 - 187 位氨基酸 (即成熟多肽) 有至少约 65%、优选至少约 70%、更优选至少约 80%、甚至更优选至少约 90%、最优选至少约 95%, 甚至最优选至少约 97%一致性程度的氨基酸序列的多肽 (以下称为“同源多肽”) 的分离核酸序列。在一个优选实施方案中, 该同源多肽的氨基酸序列与 SEQ ID NO. 4 第 22 - 581 位氨基酸、SEQ ID NO. 5 第 19 - 200 位氨基酸或 SEQ ID NO. 6 第 1 - 187 位氨基酸有 5 个氨基酸不同, 优选有 4 个氨基酸不同, 更优选有 3 个氨基酸不同, 甚至更优选有 2 个氨基酸不同, 最优选有 1 个氨基酸不同。为了本发明的目的, 两个氨基酸序列之间的一致性程度采用 Clustal 法 (Higgins, 1989, CABIOS 5: 151-153) 和 LASERGENE™ MEGALIGN™ 软件 (DNASTAR 公司, Madison, WI) 并使用同一性表和下列多序列对比 (multiple alignment) 参数来确定: 空位罚分 (gap penalty) 为 10,

空位长度罚分为 10。两两对比 (Pairwise alignment) 参数是: Ktuple=1, 空位罚分=3, 窗口 (windows) =5, 对角线 (diagonals) =5。

优选地, 本发明多核苷酸序列编码含有 SEQ ID NO. 4、SEQ ID NO. 5 或 SEQ ID NO. 6 的氨基酸序列; 或其等位变体; 或它们的片段的多肽。在一个更优选的实施方案中, 本发明核酸序列编码含有 SEQ ID NO. 4、SEQ ID NO. 5 或 SEQ ID NO. 6 的氨基酸序列的多肽。在另一个优选实施方案中, 本发明核酸序列编码含有 SEQ ID NO. 4 第 22-581 位氨基酸、SEQ ID NO. 5 第 19-200 位氨基酸或 SEQ ID NO. 6 第 1-187 位氨基酸; 或其等位变体; 或它们的片段的多肽。在另一个优选实施方案中, 本发明核酸序列编码含有 SEQ ID NO. 4 第 22-581 位氨基酸、SEQ ID NO. 5 第 19-200 位氨基酸或 SEQ ID NO. 6 第 1-187 位氨基酸的多肽。在另一个优选实施方案中, 本发明核酸序列编码由 SEQ ID NO. 4、SEQ ID NO. 5 或 SEQ ID NO. 6 的氨基酸序列; 或其等位变体; 或它们的片段组成的多肽。在另一个优选实施方案中, 本发明核酸序列编码由 SEQ ID NO. 4、SEQ ID NO. 5 或 SEQ ID NO. 6 的氨基酸序列组成的多肽。在另一个优选实施方案中, 本发明核酸序列编码由 SEQ ID NO. 4 第 22-581 位氨基酸、SEQ ID NO. 5 第 19-200 位氨基酸或 SEQ ID NO. 6 第 1-187 位氨基酸; 或其等位变体; 或它们的片段组成的多肽。在另一个优选实施方案中, 本发明核酸序列编码由 SEQ ID NO. 4 第 22-581 位氨基酸、SEQ ID NO. 5 第 19-200 位氨基酸或 SEQ ID NO. 6 第 1-187 位氨基酸组成的多肽。

本发明还包括编码具有 SEQ ID NO. 4、SEQ ID NO. 5 或 SEQ ID NO. 6 的氨基酸序列的多肽、但由于遗传密码简并性的原因又分别不同于 SEQ ID NO. 1、SEQ ID NO. 2 或 SEQ ID NO. 3 的核酸序列。本发明还涉及分别编码 SEQ ID NO. 4、SEQ ID NO. 5 或 SEQ ID NO. 6 的片段的 SEQ ID NO. 1、SEQ ID NO. 2 或 SEQ ID NO. 3 的子序列。

子序列是 SEQ ID NO. 1、SEQ ID NO. 2 或 SEQ ID NO. 3 所包含的、已从 5'和/或 3'端缺失了一个或多个核苷酸的核酸序列。优选地, SEQ ID NO. 1 的子序列含有至少 1410 个核苷酸, 更优选至少 1500 个核苷酸, 最优选至少 1590 个核苷酸。优选地, SEQ ID NO. 2 的子序列含有至少 450 个核

核苷酸,更优选至少 480 个核苷酸,最优选至少 510 个核苷酸。优选地,SEQ ID NO.3 的子序列含有至少 465 个核苷酸,更优选至少 495 个核苷酸,最优选至少 525 个核苷酸。SEQ ID NO.4、SEQ ID NO.6 或 SEQ ID NO.6 的片段是从该氨基酸序列的氨基和/或羧基端缺失了一或多个氨基酸的多肽。优选地,SEQ ID NO.4 的片段含有至少 470 个氨基酸残基,更优选至少 500 个氨基酸残基,最优选至少 530 个氨基酸残基。优选地,SEQ ID NO.5 的片段含有至少 150 个氨基酸残基,更优选至少 160 个氨基酸残基,最优选至少 170 个氨基酸残基。优选地,SEQ ID NO.6 的片段含有至少 155 个氨基酸残基,更优选至少 165 个氨基酸残基,最优选至少 175 个氨基酸残基。

等位变体是指占据相同染色体座位的一个基因的两种或多种不同形式中的任意一种。等位变异通过突变自然发生,并可导致种群内的多态现象。基因突变可以是沉默的(在编码的多肽中没有变化),也可编码具有改变的氨基酸序列的多肽。多肽的等位变体是由基因的等位变体编码的多肽。

在第二个实施方案中,本发明涉及与 SEQ ID NO.1 第 4013-5743 位核苷酸、SEQ ID NO.2 第 4013-1593 位核苷酸或 SEQ ID NO.3 第 3061-3678 位核苷酸(即成熟多肽编码区)有至少约 65%、优选约 70%、优选约 80%、更优选约 90%、甚至更优选约 95%、最优选约 97%同源性的、编码活性多肽的分离核酸序列;或其等位变体和子序列。为了本发明的目的,两个核酸序列之间的同源性程度通过 Wilbur-Lipman 方法(Wilbur 和 Lipman, 1983, 美国国家科学院院刊(Proceedings of the National Academy of Science USA) 80: 726-730)采用 LASERGENE™ MEGALIGN™软件(DNASTAR, Inc., Madison, WI)以及同一性表和下列多序列对比参数进行确定:空位罚分 10,空位长度罚分 10。两两对比参数为 Ktuple=3,空位罚分=3,窗口=20。

在第三个实施方案中,本发明涉及在极低严紧条件、优选低严紧条件、更优选中等严紧条件、更优选中高严紧条件、甚至更优选高严紧条件、最优选极高严紧条件下,和在相同条件下与下列序列杂交的核酸探针杂

交的编码多肽的分离核酸序列：(i) SEQ ID NO.1 第 4013 - 5743 位核苷酸、SEQ ID NO.2 第 993 - 1593 位核苷酸或 SEQ ID NO.3 第 3061 - 3678 位核苷酸；(ii) 在 SEQ ID NO.1 第 4013 - 5743 位核苷酸、SEQ ID NO.2 第 993 - 1593 位核苷酸或 SEQ ID NO.3 第 3061 - 3678 位核苷酸中包含的 cDNA 序列；(iii) (i) 或 (ii) 的子序列；或 (iv) (i)、(ii) 或 (iii) 的互补链 (J. Sambrook, E. F. Fritsch, 和 T. Maniatus, 1989, 分子克隆实验室手册 (Molecular Cloning, A Laboratory Manual), 第二版, Cold Spring Harbor, 纽约)。SEQ ID NO.1、SEQ ID NO.2 或 SEQ ID NO.3 的子序列可以是至少 100 个核苷酸, 优选至少 200 个核苷酸。而且, 该子序列可以编码多肽片段。

根据本领域熟知的方法, 可以利用 SEQ ID NO.1、SEQ ID NO.2 或 SEQ ID NO.3 的核酸序列或其子序列, 以及 SEQ ID NO.4、SEQ ID NO.5 或 SEQ ID NO.6 的氨基酸序列或其片段, 设计核酸探针以从不同属或种的株系中鉴定和克隆编码多肽的 DNA。具体地, 可以在标准 Southern 印迹程序之后, 采用该探针与感兴趣的属或种的基因组或 cDNA 杂交, 以鉴定和分离其中的相应基因。该探针可以比完整序列短得多, 但应长至少 15 个, 优选至少 25 个, 更优选至少 35 个核苷酸。也可使用更长的探针。既可使用 DNA 探针, 也可使用 RNA 探针。典型地, 为了检测相应基因对该探针进行标记 (例如用 ^{32}P 、 ^3H 、 ^{35}S 、生物素或抗生物素蛋白进行标记)。该探针包括在本发明的范围之内。

由此, 可以筛选从这些其它生物制备的基因组 DNA 或 cDNA 文库, 以寻找与上述探针杂交并编码多肽的 DNA。来自这些其它生物的基因组或其它 DNA 可以通过琼脂糖或聚丙烯酰胺凝胶电泳或其它分离技术分离。可以将来自这些文库的 DNA 或该分离的 DNA 转移并固定在硝酸纤维素或其它适当的载体材料上。为了鉴定与 SEQ ID NO.1 或其子序列同源的克隆或 DNA, 将该载体材料用于 Southern 印迹。为了本发明的目的, 杂交是指核酸序列与相应于 SEQ ID NO.1、SEQ ID NO.2 或 SEQ ID NO.3 所示核酸序列、它们的互补序列或其子序列的标记核酸探针在极低到极高的严谨条件下杂交。在这些条件下与该核酸探针杂交的分子用 X 光胶片检测。

在一个优选实施方案中, 该核酸探针是 SEQ ID NO.1、SEQ ID NO.2 或 SEQ ID NO.3 的核酸序列。在另一个优选实施方案中, 该核酸探针是编码 SEQ ID NO.4、SEQ ID NO.5 或 SEQ ID NO.6 的多肽的核酸序列, 或其子序列。在另一个优选实施方案中, 该核酸探针是 SEQ ID NO.1、SEQ ID NO.2 或 SEQ ID NO.3。在另一个优选实施方案中, 该核酸探针是编码成熟多肽的 SEQ ID NO.1 第 4013-5743 位核苷酸。在另一个优选实施方案中, 该核酸探针是编码成熟多肽的 SEQ ID NO.2 第 993-5743 位核苷酸。在另一个优选实施方案中, 该核酸探针是编码成熟多肽的 SEQ ID NO.3 第 3061-3678 位核苷酸。在另一个优选实施方案中, 该核酸探针是大肠杆菌 NRRL B-30067 中的质粒 pEC03 所包含的所述核酸序列。在另一个优选实施方案中, 该核酸探针是大肠杆菌 NRRL B-30071 中的质粒 pFAMG 所包含的所述核酸序列。在另一个优选实施方案中, 该核酸探针是大肠杆菌 NRRL B-30076 中的质粒 pQUINN 所包含的所述核酸序列。在另一个优选实施方案中, 该核酸探针是大肠杆菌 NRRL B-30067 中的质粒 pEC03 所包含的所述成熟多肽编码区。在另一个优选实施方案中, 该核酸探针是大肠杆菌 NRRL B-30071 中的质粒 pFAMG 所包含的所述成熟多肽编码区。在另一个优选实施方案中, 该核酸探针是大肠杆菌 NRRL B-30076 中的质粒 pQUINN 所包含的所述成熟多肽编码区。

对于至少 100 个核苷酸长的长探针, 极低到极高严紧条件定义为在标准 Southern 印迹程序之后, 在 $5\times$ SSPE、0.3% SDS、 $200\mu\text{g/ml}$ 剪切的变性鲑精 DNA 和或者 25%甲酰胺 (对于极低和低严紧条件) 或者 35%甲酰胺 (对于中等和中高严紧条件) 或者 50%甲酰胺 (对于高和极高严紧条件) 中于 42°C 进行预杂交和杂交。

对于至少 100 个核苷酸长的长探针, 最后载体材料用 $2\times$ SSC、0.2% SDS, 优选至少在 45°C (极低严紧条件), 更优选至少在 50°C (低严紧条件), 更优选至少在 55°C (中等严紧条件), 更优选至少在 60°C (中高严紧条件), 甚至更优选至少在 65°C (高严紧条件), 最优选至少在 70°C (极高严紧条件) 洗涤三次, 每次 15 分钟。

对于长约 15 个核苷酸到约 70 个核苷酸的短探针, 严紧条件定义为在

标准 Southern 印迹程序之后, 在使用 Bolton 和 McCarthy (1962, 美国国家科学院院刊 48:1390) 的计算方法计算的 T_m 值以下 5°C – 10°C , 于 0.9 M NaCl、0.09 M Tris-HCl pH 7.6、6 mM EDTA、0.5% NP-40、 $1\times$ Denhardt's 溶液、1 mM 焦磷酸钠、1 mM 磷酸二氢钠(sodium monobasic phosphate)、0.1 mM ATP 和 0.2 mg 酵母 RNA/ml 中进行预杂交、杂交和杂交后的洗涤。

对于长约 15 个核苷酸到约 70 个核苷酸的短探针, 载体材料在计算的 T_m 以下 5°C – 10°C 用 $6\times$ SSC 加 0.1% SDS 洗涤一次, 时间 15 分钟, 用 $6\times$ SSC 洗涤两次, 每次 15 分钟。

本发明还涉及通过如下步骤制备的分离核酸序列: (a) 在极低、低、中等、中高、高或极高的严紧条件下使 DNA 与 (i) SEQ ID NO.1 第 4013–5743 位核苷酸、SEQ ID NO.2 第 993–1593 位核苷酸或 SEQ ID NO.3 第 3061–3678 位核苷酸, (ii) SEQ ID NO.1 第 4013–5743 位核苷酸、SEQ ID NO.2 第 993–1593 位核苷酸或 SEQ ID NO.3 第 3061–3678 位核苷酸中所包含的 cDNA 序列, (iii) (i) 或 (ii) 的子序列, 或 (iv) (i)、(ii) 或 (iii) 的互补链杂交; 和 (b) 从该 DNA 中分离出该核酸序列。该子序列优选是具有至少 100 个核苷酸的序列, 例如编码多肽片段的序列。

在第四个实施方案中, 本发明涉及编码具有 SEQ ID NO.4、SEQ ID NO.5 或 SEQ ID NO.6 之氨基酸序列的多肽的变体的分离核酸序列, 其中所述变体包含了一或多个氨基酸的替代、缺失和/或插入。

该多肽变体的氨基酸序列与 SEQ ID NO.4、SEQ ID NO.5 或 SEQ ID NO.6, 或其成熟多肽的氨基酸序列的差别可以是一个或多个氨基酸残基的插入或缺失和/或一个或多个氨基酸残基被不同的氨基酸残基替代。优选地, 氨基酸的改变在性质上是较小的, 即不显著影响该蛋白质的折叠和/或活性的保守氨基酸替换; 小的缺失, 典型的是 1 到约 30 个氨基酸的缺失; 小的氨基或羧基末端突出, 如氨基末端的甲硫氨酸残基; 不超过约 20–25 个残基的小接头肽; 或通过改变净电荷或其它功能便于纯化的小突出, 如聚组氨酸片段、抗原表位或结合域。

保守替代的实例是以下各组内的替代: 碱性氨基酸组(精氨酸、赖氨酸和组氨酸), 酸性氨基酸组(谷氨酸和天冬氨酸), 极性氨基酸组(谷

氨酰胺和天冬酰胺), 疏水性氨基酸组(亮氨酸、异亮氨酸和缬氨酸), 芳香族氨基酸组(苯丙氨酸、色氨酸和酪氨酸)和小氨基酸组(甘氨酸、丙氨酸、丝氨酸、苏氨酸和甲硫氨酸)。通常不改变比活性的氨基酸替代是本领域已知的, 并描述于例如 H. Neurath 和 R.L. Hill, 1979, 蛋白质(The Proteins), Academic Press, 纽约。最常发生的交换是 Ala/Ser, Val/Ile, Asp/Glu, Thr/Ser, Ala/Gly, Ala/Thr, Ser/Asn, Ala/Val, Ser/Gly, Tyr/Phe, Ala/Pro, Lys/Arg, Asp/Asn, Leu/Ile, Leu/Val, Ala/Glu 和 Asp/Gly, 以及它们的相反替代。

本发明的核酸序列可以获自任何属的微生物。为了本发明的目的, 术语“获自”按本文中定义的方式使用。在一个优选实施方案中, 由本发明核酸序列编码的多肽被分泌至细胞外。

本发明的核酸序列可以获自真菌来源, 更优选获自酵母菌株, 如假丝酵母属、汉逊酵母属、克鲁维酵母属、毕赤酵母属、酵母属、裂殖酵母属或 *Yarrowia* 菌株; 或更优选获自丝状真菌菌株如 *Acremonium*、曲霉属、短梗霉属 (*Aureobasidium*)、隐球酵母属 (*Cryptococcus*)、*Filibasidium*、镰孢霉属、腐质霉属、*Magnaporthe*、毛霉属、毁丝霉属 (*Myceliophthora*)、*Neocallimastix*、脉孢菌属、拟青霉属 (*Paecilomyces*)、青霉属、*Piromyces*、裂褶菌属 (*Schizophyllum*)、*Talaromyces*、嗜热子囊菌属 (*Thermoascus*)、梭孢壳属、*Tolypocladium* 或木霉属 (*Trichoderma*) 菌株。

在一个优选实施方案中, 该核酸序列获自卡尔酵母、酿酒酵母、糖化酵母、*Saccharomyces douglasii*、克鲁弗酵母、诺地酵母或 *Saccharomyces oviformis* 菌株。

在另一个优选实施方案中, 该核酸序列获自棘孢曲霉 (*Aspergillus aculeatus*)、泡盛曲霉、臭曲霉、日本曲霉、构巢曲霉、黑曲霉、米曲霉、*Humicola insolens*、*Humicola lanuginosa*、米黑毛霉、*Myceliophthora thermophila*、粗糙脉孢菌、产紫青霉、*Trichoderma harzianum*、康宁木霉、*Trichoderma longibrachiatum*、*Trichoderma reesei* 或绿色木霉菌株。

在另一个优选实施方案中，该核酸序列获自杆孢状镰孢、*Fusarium cerealis*、*Fusarium crookwellense*、大刀镰孢、禾本科镰孢、禾赤镰孢、异孢镰孢、合欢木镰孢、尖镰孢、多枝镰孢、粉红镰孢、接骨木镰孢、肤色镰孢、拟分枝孢镰孢、*Fusarium sulphureum*、*Fusarium torulosum*、*Fusarium trichothecioides* 或 *Fusarium venenatum* 菌株。

在一个更优选的实施方案中，该核酸序列获自 *Fusarium venenatum*，最优选获自 *Fusarium venenatum* ATCC 20334，例如 SEQ ID NO.1、SEQ ID NO.2 或 SEQ ID NO.3 中所示的核酸序列。在另一个更优选的实施方案中，SEQ ID NO.1 的核酸序列是大肠杆菌 NRRL B-30067 中的质粒 pECO3 所包含的序列。在另一个更优选的实施方案中，SEQ ID NO.2 的核酸序列是大肠杆菌 NRRL B-30071 中的质粒 pFAMG 所包含的序列。在另一个更优选的实施方案中，SEQ ID NO.3 的核酸序列是大肠杆菌 NRRL B-30076 中的质粒 pQUINN 所包含的序列。在另一个优选实施方案中，该核酸序列是编码成熟多肽的 SEQ ID NO.1 第 4013 - 5743 位核苷酸。在另一个优选实施方案中，该核酸序列是编码成熟多肽的 SEQ ID NO.2 第 993 - 1593 位核苷酸。在另一个优选实施方案中，该核酸序列是编码成熟多肽的 SEQ ID NO.3 第 3061 - 3678 位核苷酸。

应该理解，对于前面提及的种，本发明既包括完全和不完全阶段，及其它的分类学等同物，例如无性型，而不论它们被称为何种种名。本领域技术人员将容易识别出适合的等同物。

公众可容易地从许多培养物保藏中心获得这些种的菌株，这些培养物保藏中心例如美国典型培养物保藏中心（ATCC）、德意志微生物保藏中心（DSM）、真菌菌种保藏中心（CBS）和农业研究机构保藏中心（NRRL）。

而且，采用上述探针可从其它来源鉴定并获得这类核酸序列，这些来源包括从自然环境（如土壤、堆肥、水等）分离的微生物。从自然生活环境中分离微生物的技术是本领域熟知的。然后可以通过对另一种微生物的基因组或 cDNA 文库的相似筛选获得核酸序列。一旦用探针检测到编码多肽的核酸序列，即可利用本领域普通技术人员已知的技术（见例如 Sambrook 等，1989，同上）分离或克隆该序列。

本发明还涉及在 SEQ ID NO. 1、SEQ ID NO. 2 或 SEQ ID NO. 3 的成熟多肽编码序列中含有至少一个突变的突变核酸序列，其中该突变核酸序列编码分别由 SEQ ID NO. 4 第 22-581 位氨基酸、SEQ ID NO. 5 第 19-200 位氨基酸或 SEQ ID NO. 6 第 1-187 位氨基酸组成的多肽。

用于分离或克隆编码多肽的核酸序列的技术在本文中已有过描述。

本发明还涉及含有上述核苷酸序列的核酸结构、重组载体和宿主细胞。对于它们的构建可以采用前面描述的相同方法来进行。

本发明还涉及由上述核酸序列编码的多肽，以及采用本文描述的方法产生该多肽的方法。

本发明通过如下实施例作进一步描述，这些实施例不应理解为是对本发明范围的限制。

实施例

用作缓冲液和底物的化学制品是至少试剂级别的商品产品。

培养基和溶液

COVE 痕量金属溶液每升由 0.04g $\text{NaB}_4\text{O}_7 \cdot 10\text{H}_2\text{O}$ 、0.4g $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ 、1.2g $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ 、0.7g $\text{MnSO}_4 \cdot \text{H}_2\text{O}$ 、0.8g $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$ 和 10g $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ 组成。

50× COVE 盐溶液每升由 26g KCl 、26g $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ 、76g KH_2PO_4 和 50ml COVE 痕量金属组成。

COVE 培养基每升由 342.3 g 蔗糖、20ml 50× COVE 盐溶液、10ml 1M 乙酰胺、10ml 1.5M CsCl_2 和 25g 纯净琼脂组成。

50× Vogels 培养基每升由 150g 柠檬酸钠、250g KH_2PO_4 、10g $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ 、10g $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$ 、2.5ml 生物素贮存液和 5.0ml AMG 痕量金属溶液组成。

痕量金属溶液每升由 14.3g $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ 、2.5g $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ 、0.5g NiCl_2 、13.8g FeSO_4 、8.5g MnSO_4 和 3.0g 柠檬酸组成。

COVE 顶层琼脂每升由 20ml 50× COVE 盐、0.8M 蔗糖、1.5M 氯化铯、1.0M 乙酰胺和 10 g 低熔点琼脂糖组成，pH 被调节至 6.0。

BASTA 顶层琼脂由补加了 10mg/ml 除草剂 Basta™ (Hoechst Schering, Rodovre, 丹麦) 的 COVE 顶层琼脂组成。

RA 孢子形成培养基每升由 50 g 琥珀酸、12.1 g NaNO₃、1 g 蔗糖、20ml 50× Vogels 和 0.5ml 10mg/ml NaMoSO₄ 贮存液组成, pH 被调节至 6.0。

YEPG 培养基每升由 10g 酵母提取物、20g 蛋白胨和 20g 葡萄糖组成。

STC 由 0.8 M 山梨糖醇、25mM Tris pH 8、25mM CaCl₂ 组成。

SPTC 由 40% PEG 4000、0.8 M 山梨糖醇、25 mM Tris pH 8、25 mM CaCl₂ 组成。

M400Da 培养基每升由 50 g 麦芽糖糊精、2 g MgSO₄·7H₂O、2g KH₂PO₄、4g 柠檬酸、8g 酵母提取物、2g 尿素和 1ml COVE 痕量金属溶液组成。

实施例 1: *Fusarium venenatum* 菌丝体组织的生产

采用补料分批发酵方案, 使用作为碳源的 NUTRIOSE™ (Roquette Freres, S.A., Beinheim, 法国) 和酵母提取物, 将 *Fusarium venenatum* ATCC 20334 的形态学突变体 *Fusarium venenatum* CC1-3 (Wiebe 等, 1991, Mycol. Research 95: 1284-1288) 培养在实验室规模的两升发酵罐中。pH 维持在 6-6.5, 温度保持在 30℃, 并含有活性溶解氧 (positive dissolved oxygen)。

在接种后第 2、4、6 和 8 天收获菌丝体样品, 并在液氮中迅速冷冻。在将其破碎用于 RNA 提取之前, 将这些样品保存在 -80℃。

实施例 2: cDNA 文库构建

根据 Timberlake 和 Barnard 的方法 (1981, 细胞 (Cell) 26: 29-37) 从实施例 1 中描述的菌丝体样品中提取总细胞 RNA, 从 1% 甲醛-琼脂糖凝胶上进行印迹之后通过 Northern 杂交分析该 RNA 样品 (Davis 等, 1986, 分子生物学基本方法 (Basic Methods in Molecular Biology), Elsevier Science Publishing Co., Inc., 纽约)。使用 mRNA 分离试剂盒 (mRNA Separator Kit™) (Clontech Laboratories, Inc., Palo Alto, CA) 根据厂商说明从总 RNA 分离聚腺苷酸化 mRNA 部分。除了使用 NotI-(dT)18

引物 (Pharmacia Biotech, Inc., Piscataway, NJ) 起始第一链合成之外, 根据 Gubler 和 Hoffman 的方法 (1983, 基因 (Gene) 25:263-269) 使用大约 5 μ g poly(A)⁺mRNA 合成双链 cDNA。用绿豆核酸酶 (Boehringer Mannheim Corporation, Indianapolis, IN) 处理该 cDNA, 并用 T4 DNA 聚合酶 (New England Biolabs, Beverly, MA) 使末端平端化。

用 NotI 消化 cDNA, 通过琼脂糖凝胶电泳按大小选择片段 (大约 0.7-4.5 kb), 并与用 NotI 和 EcoRV 切割并经小牛小肠碱性磷酸酶 (Boehringer Mannheim Corporation, Indianapolis, IN) 去酸化的 pZEr0-2.1 (Invitrogen Corporation, Carlsbad, CA) 连接。使用该连接混合物转化大肠杆菌 TOP10 感受态细胞 (Invitrogen Corporation, Carlsbad, CA)。在含有 50 μ g/ml 终浓度卡那霉素的 2YT 琼脂平板上筛选转化体 (Miller, 1992, 细菌遗传学简短教程。关于大肠杆菌和相关细菌的实验室指南和手册 (A Short Course in Bacterial Genetics. A laboratory Manual and Handbook for Escherichia coli and Related Bacteria), Cold Spring Harbor Press, Cold Spring Harbor, 纽约)。

使用质粒克隆载体 pZEr0-2.1 构建两个独立的定向 (directional) cDNA 文库。文库 A 使用第 4 天收获的菌丝体的 mRNA 制备, 文库 B 使用从第 6 天时间点获得的 mRNA 构建。为了检验细胞中基因表达谱的代表性“瞬态图 (snapshot)”, 两个 cDNA 文库均没有扩增。相反地, 将文库铺板, 标定滴度, 并通过 DNA 测序分析每个文库的独立克隆。

文库 A (第 4 天的细胞) 由大约 7.5×10^4 个独立克隆组成, 文库 B (第 6 天的细胞) 由大约 1.2×10^5 个克隆组成。从每个文库中的 40 个菌落分离小量制备的 DNA (Miniprep DNA), 并检查 cDNA 插入片段的存在和大小。在该分析中, 来自文库 A 的 40 个菌落中有 39 个 (97.5%) 含有大小在 600bp 至 2200bp 之间 (平均=1050bp) 的插入片段。相似地, 来自文库 B 的 40 个菌落中 39 个 (97.5%) 含有大小在 800bp 至 3600bp 之间 (平均=1380bp) 的插入片段。

实施例 3: 模板的制备和核苷酸序列测定

从实施例 2 中描述的每个 cDNA 文库中, 直接从转化平板挑取 1192 个转化体菌落至含有包含 50 μ g/ml 卡那霉素的 200 μ l 2YT 培养液 (Miller, 1992, 见上) 的 96 孔微量滴定板中。将微量滴定板在 37 $^{\circ}$ C 不摇动孵育过夜。孵育之后, 向每个孔加入 100 μ l 无菌的 50%甘油。将转化体复制到第二个深盘型 (deep dish) 96 孔微量培养板 (Advanced Genetic Technologies Corporation, Gaithersburg, MD) 中, 该微量培养板的每孔中含有 1ml 补充了 50 μ g/ml 卡那霉素的 Magnificent BrothTM (MacConnell Research, San Diego, CA)。原始的微量滴定板冷冻储存在 -80 $^{\circ}$ C。第二个深盘型培养板于 37 $^{\circ}$ C 旋转摇床上剧烈振荡 (300rpm) 孵育过夜。为防止溅出和交叉污染, 以及允许充分的通气, 用聚丙烯衬垫 (Advanced Genetic Technologies Corporation, Gaithersburg, MD) 和塑料微量滴定盘盖覆盖每个第二培养板。

使用 Utterback 等所修改的 Advanced Genetic Technologies 公司 (Gaithersburg, MD) 96 孔小量制备试剂盒程序 (1995, 基因组科学技术 (Genome Sci. Technol.) 1:1-8), 从每个孔分离 DNA。使用 Perkin-Elmer Applied Biosystems 377 XL 型自动 DNA 测序仪 (Perkin-Elmer Applied Biosystems 公司, Foster City, CA), 采用染料终止物化学法 (dye-terminator chemistry) (Giesecks 等, 1992, 病毒学方法杂志 (Journal of Virology Methods) 38:47-60) 以及反向 lac 测序引物进行单侧 (single-pass) DNA 测序。

实施例 4: DNA 序列数据分析

仔细检查核苷酸序列数据的质量, 给出少于或等于 9.2 的不适当间隔或超过 3% 的模糊水平的样品被丢弃或进行重新测序。借助于 FACTURATM 软件 (Perkin Elmer Applied Biosystems, Inc., Foster City, CA) 修剪载体序列。此外, 当模糊碱基信号数目增加时, 在每个样品的末端截短序列。使用 TIGR Assembler 软件 (Sutton, G.G. 等, 1995, 基因组科学和技术 (Genome Science and Technology) 1:9019), 将所有序列互相比对以构建重叠的毗连序列群, 从而确定每个文库中提供的各种 cDNA

种类的多重性。最后，按三种框架翻译所有序列，并使用采用修改的 Smith-Waterman 算法的 GeneAssist™ 软件（Perkin Elmer Applied Biosystems, Inc., Foster City, CA）对非冗余数据库（NRDB）进行搜索，其中所述算法使用阈值为 70 的 BLOSUM 62 矩阵。NRDB 是由 Genpept、Swiss-Prot 和 PIR 数据库组合而来。

实施例 5: 葡糖淀粉酶 cDNA 克隆的鉴定

通过使用 Applied Biosystems 377 XL 型自动 DNA 测序仪根据厂商说明进行随机 cDNA 克隆的部分测序，并按照实施例 4 所述将推导的氨基酸序列与 NRDB 中的序列进行比较，鉴定推测的葡糖淀粉酶克隆。从分析的 2000 多个 cDNA 序列中，来自文库 A 的 2 个克隆和来自文库 B 的 9 个克隆表现出与来自其它真菌和酵母的葡糖淀粉酶蛋白质具有氨基酸序列同源性。在以此方式发现的几个葡糖淀粉酶 cDNA 克隆中，基于其与粗糙脉孢菌 (Geneseq 蛋白质登录号 R71034) 和 *Humicola grisea* (Trembl 登录号 Q12623) 的葡糖淀粉酶氨基酸序列的比对，以及使用 signal-P 计算机程序 (Nielsen 等, 1997, 蛋白质工程 (Protein Engineering) 10:1-6) 所检测到的可能信号肽的存在，推测一个克隆是全长的。选择含有质粒 pFA0401、定名为大肠杆菌 FA0401 的克隆用作探针，以从 *Fusarium venenatum* 克隆相应的葡糖淀粉酶基因组 DNA 序列 (见实施例 7)。

实施例 6: 构建 *Fusarium venenatum* 的基因组 DNA 文库

按照在 Berka 等 (1998, Appl. Environ. Microbiol. 64, 4423-4427) 的程序中详述的方法，在 λ ZipLox 中构建基因组文库。简单地说，用 Tsp509I 部分消化 *Fusarium venenatum* 的总细胞 DNA，并在 1% 琼脂糖凝胶上进行大小分离。将在 3-7kb 大小范围内迁移的 DNA 片段切下，并采用 Prep-a-Gene 试剂 (BioRad, Hercules, CA) 将 DNA 从琼脂糖凝胶切块中洗脱出来。将洗脱的 DNA 片段与 EcoRI 切割并经去磷酸化的 λ ZipLox 载体臂 (Life Technologies, Gaithersburg, MD) 连接，连接混合物采用商业包装提取物 (Stratagene, La Jolla, CA) 进行包装。采用标准方法将该

包装的 DNA 文库铺板,并在大肠杆菌 Y1090ZL 细胞上扩增,然后储存于 4℃ (Davis, R.W., Botstein, D., 和 Roth, J.R., 1980, 高级细菌遗传学, 遗传工程手册 (Advanced Bacterial Genetics, A Manual for Genetic Engineering), Cold Spring Harbor Press, Cold Spring Harbor, NY)。

实施例 7: 编码 *Fusarium venenatum* 葡糖淀粉酶的基因组 DNA 片段的分离, 核苷酸测序和特性分析

采用含有来自 pFA0401 (实施例 5) 的克隆 cDNA 插入片段的放射性标记探针片段, 在高度严谨条件下 (即, 在 50% 甲酰胺、5×SSPE、0.3% SDS、200μg/ml 剪切的变性鲑精 DNA 中 45℃ 进行杂交; 滤膜在含有 0.1% SDS 的 0.2×SSPE 中 45℃ 洗涤一次, 之后在无 SDS 的 0.2×SSPE 中于相同温度下洗涤两次), 通过杂交 (Davis 等, 1980, 同上) 筛选来自 *Fusarium venenatum* 基因组 DNA 文库 (实施例 6) 的大约 50,000 噬菌斑。在大肠杆菌 Y1090ZL 细胞上两次纯化给出杂交信号的噬菌斑, 随后从 λZipLox 载体中切出单个克隆作为 pZL1 衍生物 (D'Alessio 等, 1992, Focus® 14:76)。DNA 序列分析揭示, 其中一个含有命名为 pFAMG 的质粒的克隆包含了完整的葡糖淀粉酶编码区以及分别包括了启动子和终止子区域的 3.9kb 5'侧翼 DNA 和 0.3kb 3'侧翼 DNA (见图 1)。

使用染料终止物化学法在 Applied Biosystems 377 XL 型自动 DNA 测序仪上对 pFAMG 中的克隆插入片段进行 DNA 序列测定。使用转座子插入策略 (Primer Island Transposition 试剂盒, Perkin-Elmer/Applied Biosystems, Inc., Foster City, CA) 产生连续序列。对来自 pFAMG 的葡糖淀粉酶基因组克隆进行测序, 达到平均冗余 4.2。

通过该基因组序列数据与葡糖淀粉酶 cDNA 序列的重叠群比较, 确定编码 *Fusarium venenatum* 葡糖淀粉酶的基因组 DNA 片段含有一个 1743bp 的开放阅读框, 该阅读框被一个 51bp 的内含子中断并编码具有 581 个氨基酸的多肽。该核苷酸序列 (SEQ ID NO. 1) 和推导的氨基酸序列 (SEQ ID NO. 4) 显示在图 1 中。采用 SignalP 程序 (Nielson 等, 1997, 蛋白质工程 10:1-6), 预测了一个 21 个残基的信号肽, 并经该葡糖淀粉酶蛋白质

氨基端测序获得了证实(实施例 8)。因此,成熟的葡糖淀粉酶由 560 个氨基酸组成。

使用 LASERGENE™ MEGALIGN™ 软件 (DNASTAR, Inc., Madison, WI) 和 Clustal 方法 (Higgins, 1989, CABIOS 5:151-153), 采用同一性表和如下多序列比对参数进行真菌葡糖淀粉酶蛋白质序列的比较性对比: 空位罚分为 10, 空位长度罚分为 10。两两对比参数是 Ktuple=1、空位罚分=3、窗口=5 和对角线=5。该对比显示, *Fusarium venenatum* 的葡糖淀粉酶与下列其它真菌的葡糖淀粉酶具有一致性(括弧内为一致残基的百分数): 粗糙脉孢霉 [Swissprot P14804] (47%)、黑曲霉 [Swissprot P04064] (46%)、*Humicola grisea* [Trembl Q12623] (44%)、*Hormoconis resinae* [Swissprot Q03045] (41%)、*Corticium rolfsii* [Q12596] (41%)、*Schizosaccharomyces pombe* [Trembl O60087] (31%)、和 *Rhizopus oryzae* [Swissprot P07683] (23%)。

实施例 8: *Fusarium venenatum* 葡糖淀粉酶的氨基端序列分析

通过 8-16 % Tris-甘氨酸 SDS-PAGE (Novex, San Diego, CA) 分离 *Fusarium venenatum* 发酵培养液 (实施例 1) 样品中的蛋白质, 然后采用 10 % 甲醇 (pH=11.0) 中的 10mM CAPS (3-[环己基氨基]-1-丙磺酸) 25 伏 2 小时将其电转印至 PVDF 膜上 (Novex, San Diego, CA)。用 40 % 甲醇 / 1 % 乙酸中的 0.1 % 考马斯亮兰 R250 染色该 PVDF 膜 20 秒, 并在 50 % 乙醇中脱色以观察蛋白质条带。

切下迁移在大约 65kDa 的主要多肽种类, 并在带有即时 HPLC 和液相三氟乙酸 (TFA) 递送的 Applied Biosystems 476A 蛋白质测序仪 (Perkin Elmer/Applied Biosystems Division, Foster City, CA) 上对其进行 N 端测序。测序试剂来自 Perkin Elmer/Applied Biosystems Division (Foster City, CA)。采用含有溶于水的 3.5 % 四氢呋喃的 A 缓冲液以及 18ml 含有乙酸、乙酸钠和己磺酸钠的 Premix 浓缩物 (Perkin Elmer/Applied Biosystems Division, Foster City, CA) 和含有乙腈的缓冲液 B, 通过即时 HPLC 检测乙内酰苯硫脲氨基酸。收集数据, 并在 Macintosh II si

上采用 Applied Biosystems 610 数据分析软件进行分析。对着光源观察色谱以鉴定氨基酸，并由操作员确定。N-端分析产生蛋白质序列 Ser-Pro-Ser-Lys-Asp-Asn-Ser-Leu-Glu-Arg-Phe-Ile-Asp-Lys-Gln-Ala-Asp-Ile-Ser (SEQ ID NO. 4)。

实施例 9: 鉴定编码推测的液泡相关蛋白质和未知的分泌基因产物的丰富 cDNA 克隆

基于其在文库 A 和 B (实施例 2) 的克隆中的相对丰度，再选择两种 cDNA 种类。

含有 pFA0035 的克隆 FA0035 编码具有 18 个氨基酸可能信号肽的 200 个氨基酸初级翻译产物，其中该可能信号肽是采用 Signal-P 计算机程序 (Nielsen 等, 1997, 同上) 检测到的。克隆 FA0035 占文库 A 中 cDNA 克隆的大约 1.9%，而占文库 B 中 cDNA 克隆的大约 0.8%。按实施例 4 所述将克隆 FA0035 的推导氨基酸序列与 NRDB 中的序列进行比较，结果显示与任何已知的蛋白质序列均没有显著的同源性，因此将其划为未知开放阅读框 (ORF)。该克隆被任意命名为 “Daria”。

含有 pFA0759 的克隆 FA0759 编码与粗糙脉孢霉液泡相关蛋白质的推测亚基 (Trembl 登录号 P87252) 有 72% 氨基酸序列一致性的具有 187 个氨基酸的多肽。由于采用 Signal-P 软件没有检测到信号肽，所以该基因产物似乎并不是分泌蛋白。克隆 FA0759 占文库 A 中 cDNA 克隆的大约 2.0%，而占文库 B 中 cDNA 克隆的大约 1.5%。

实施例 10: 编码未知基因 “Daria” 和推测的液泡相关蛋白质的基因组 DNA 片段的克隆和核苷酸序列分析

通过筛选实施例 6 和 7 所述的 λ ZipLox 文库分离编码未知分泌蛋白质 “Daria” 和推测的液泡相关蛋白质的 *Fusarium venenatum* 基因组 DNA 片段。对来自质粒 pFA0035 和 pFA0759 的 cDNA 插入片段进行放射性标记，将其用作探针筛选文库。将在实施例 7 的相同条件下与任一探针强杂交的噬菌斑在大肠杆菌 Y1090ZL 细胞上进行两次纯化，随后从 λ ZipLox 载

体中切下各个克隆作为 pZL1 衍生物 (D'Alessio 等, 1992, 同上)。

DNA 序列分析揭示, 其中一个含有名为 pEC03 的质粒的克隆含有完整的 “Daria” 编码区以及分别包括了启动子和终止子区域的 0.9kb 5'侧翼和 0.9kb 3'侧翼 DNA。该核苷酸序列 (SEQ ID NO. 2) 和推导的氨基酸序列 (SEQ ID NO. 5) 显示于图 2。该编码区被一个 55bp 内含子打断。

相似地, 用来源于 pFA0579 cDNA 插入片段的放射性标记探针, 通过筛选 λ ZipLox 文库分离基因组 DNA 克隆。命名为 pQUINN 的基因组克隆编码完整的液泡相关蛋白质亚基以及分别包含了启动子和终止子区域的 3.0kb 5'侧翼和 0.3kb 3'侧翼 DNA。该核苷酸序列 (SEQ ID NO. 3) 和推导的氨基酸序列 (SEQ ID NO. 6) 显示于图 3。该基因组克隆和 cDNA 克隆之间的序列比较揭示, 在该推测的液泡相关蛋白质亚基的 5'非翻译区中存在一个 594bp 的内含子。此外, 该编码还含有一个 77bp 的内含子。

实施例 11: pDM181 的构建

采用拼接重叠延伸技术构建质粒 pDM181, 以将 1.2kb 尖镰孢胰蛋白酶启动子与 1.1kb 尖镰孢胰蛋白酶终止子融合在一起。作为重叠 PCR 策略的一部分, 将含有 *Swa*I、*Kpn*I 和 *Pac*I 限制性位点的多位点接头插入该启动子和终止子之间。在启动子的 5'末端添加一个 *Xho*I 位点, 并保留天然的 *Eco*RI 位点。在终止子的 3'末端, 通过 PCR 反应引入 *Eco*RI、*Hind*III 和 *Nsi*I 位点。

使用如下引物从质粒 pJRoy20 (Royer 等, 1995, 生物技术 (Biotechnology) 13:1479-1483) 产生含有尖镰孢胰蛋白酶启动子第 -1208 至 -1 位序列和 25 个碱基对多位点接头的 PCR 片段:

引物 1 (有义): 5'-GAGCTCGAGGAATTCTTACAAACCTTCAAC-3' (SEQ ID NO. 7)
XhoI EcoRI

引物 2 (反义):

5'-TTAATTAAGGTACCTGAATTTAAATGGTGAAGAGATAGATATCCAAG-3' (SEQ ID NO. 8)
PacI KpnI SwaI

100 μ l PCR 反应物含有 10ng pJRoy20, 每个引物各 50pmol, 1 $\times$ Pwo

缓冲液 (Boehringer Mannheim, Indianapolis, IN), dATP、dCTP、dGTP 和 dTTP 各 200 μ M, 和 5 单位 Pwo DNA 聚合酶 (Boehringer Mannheim, Indianapolis, IN)。所用的 PCR 条件是 95 $^{\circ}$ C 3 分钟, 之后 25 个循环, 每个循环包括 95 $^{\circ}$ C 30 秒、50 $^{\circ}$ C 1 分钟和 72 $^{\circ}$ C 1 分钟。最后的延伸循环是 72 $^{\circ}$ C 5 分钟。

使用相同的 PCR 条件和如下引物, 从质粒 pJRoy20 产生含有尖镰孢胰蛋白酶启动子的第-5 到-1 位碱基对、25 个碱基对的多位点接头和尖镰孢胰蛋白酶基因 3'非翻译区的 1060 个碱基对 (终止子区) 的第二个 PCR 片段:

引物 3 (有义):

5'-TCACCATTTAAATTCAGGTACCTTAATTAAATTCCTTGTTGGAAGCGTCGA-3' (SEQ ID NO. 9)

SwaI KpnI PacI

引物 4 (反义):

5'-TGGTATGCATAAGCTTGAATTCAGGTAAACAAGATATAATTT-3' (SEQ ID NO. 10)

NsiI HindIII EcoRI

使用 0.2 μ l 第一次 PCR (启动子) 反应产物和 3 μ l 第二次 (终止子) 反应产物作为模板, 以及引物 1 和 4, 获得含有尖镰孢胰蛋白酶启动子第-1208 到-1 位、25 个碱基对的多位点接头和尖镰孢胰蛋白酶终止子的 1060 个碱基对的最终 2.3kb 重叠 PCR 片段。所用 PCR 条件是 95 $^{\circ}$ C 3 分钟, 之后 30 个循环, 每个循环包括 95 $^{\circ}$ C 30 秒、62 $^{\circ}$ C 1 分钟和 72 $^{\circ}$ C 3 分钟。最后的延伸循环是 72 $^{\circ}$ C 5 分钟。该反应也使用 Pwo DNA 聚合酶。

将所获的含有胰蛋白酶启动子、多位点接头和胰蛋白酶终止子的 2.3kb 片段用 EcoRI 消化, 并连入 EcoRI 消化的含有 bar 基因的载体 pMT1612 (WO 97/26330) 中产生 pDM181 (图 4)。

实施例 12: 质粒 pSheB1 的构建

通过修饰 pDM181 产生 *Fusarium venenatum* 表达载体 pSheB1 (图 5)。修饰包括 (a) 去除 pDM181 序列中的两个 NcoI 位点, 和 (b) 重新恢复尖镰

孢胰蛋白酶启动子的天然翻译起始（在 ATG 起始密码子处重新构建一个 NcoI 位点）。

使用 QuikChange™ 定点诱变试剂盒（Stratagene Cloning Systems, La Jolla, CA），根据厂商说明使用如下诱变引物对实现 pDM181 序列中两个 NcoI 位点的去除：

5'-dCAGTGAATTGGCCTCGATGGCCGCGGCCGCGAATT-3' (SEQ ID NO. 11) 和

5'-dAATTCGCGGCCGCGGCCATCGAGGCCAATTCCTG-3' (SEQ ID NO. 12)

5'-dCACGAAGGAAAGACGATGGCTTTCACGGTGTCTG-3' (SEQ ID NO. 13) 和

5'-dCAGACACCGTGAAAGCCATCGTCTTTCCTTCGTG-3' (SEQ ID NO. 14)

使用 QuikChange™ 定点诱变试剂盒以及如下诱变引物对也实现了尖镰孢胰蛋白酶启动子天然翻译起始的重新恢复：

5'-dCTATCTCTTCACCATGGTACCTTAATTAAATACCTTGTTGGAAGCG-3' (SEQ ID NO. 11) 和

5'-dCGCTTCCAACAAGGTATTTAATTAAAGGTACCATGGTGAAGAGATAG-3' (SEQ ID NO. 12)

所有定点改变均通过合适载体区的 DNA 序列分析确证。

实施例 13: pDM194 和 pDM218 的构建

用于在 *Fusarium venenatum* 中获得 *Thermomyces lanuginosus*（原称 *Humicola lanuginosa*）脂酶表达的 7.8kb 质粒 pDM194 (7.8kb) 具有如下基因元件：

含有尖镰孢胰蛋白酶基因启动子 (Royer 等, 1995, 同上) 的 1.2kb DNA 片段。

含有 *Thermomyces lanuginosus* 脂酶 cDNA (EP 305 216) 的 0.9kb DNA 片段。

含有尖镰孢胰蛋白酶基因终止子 (Royer 等, 1995, 同上) 的 1.1kb DNA 片段。

来自 pMT1612 (WO 98/11203) 的含有 2.8kb 大肠杆菌载体 pUC19 和一个 1.8kb 片段的 4.7kb DNA 片段，其中所述 1.8kb 片段含有构巢曲霉 *amdS*

基因启动子(Hynes 等, 1988, 分子细胞生物学(Mol. Cell Biol.) 8:2589-2596)、吸水链霉菌磷丝菌素乙酰转移酶(bar)基因(Thompson 等, 1987, EMBO Journal 6: 2519-2514)、和黑曲霉 AMG 终止子。

通过 PCR 产生 SwaI/PacI 脂酶片段。采用含有 *Thermomyces lanuginosus* 脂酶基因(EP 305 216) cDNA 的质粒 pMHan37 作为模板。以下显示的引物 5 和 6 用于在该脂酶编码区的 5'末端引入一个 SwaI 位点和在 3'末端引入一个 PacI 位点。

引物 5 (有义):

5'-ATTTAAATGATGAGGAGCTCCCTTGTGCTG-3' (SEQ ID NO. 17)

SwaI

引物 6 (反义):

5'-TTAATTAAGTAGAGTCGACCCAGCCGCGC-3' (SEQ ID NO. 18)

PacI

100 μ l PCR 反应物含有 10ng pMHan37, 每个引物各 50pmol, 1 $\times$ PCR 缓冲液 (Perkin-Elmer Corp., Branchburg, NJ), dATP、dCTP、dGTP 和 dTTP 各 250 μ M, 和 5 单位 Taq DNA 聚合酶 (Perkin-Elmer Corp., Branchburg, NJ)。所用的 PCR 条件是 95 $^{\circ}$ C 5 分钟一个循环, 之后 30 个循环, 每个循环包括 95 $^{\circ}$ C 1 分钟、55 $^{\circ}$ C 1 分钟和 72 $^{\circ}$ C 2 分钟。根据厂家说明将 0.9kb PCR 产物亚克隆在 TA 克隆试剂盒 (Invitrogen, Carlsbad, CA) 的 pCRII 中。采用 Qiagen Maxiprep 试剂盒 (Qiagen, Santa Clarita, CA) 分离质粒 DNA, 用 SwaI 和 PacI 消化, 之后在 1% 琼脂糖凝胶上 100 伏电泳 1 小时。切下该 0.9kb 片段, 采用 SpinBind 试剂盒 (FMC, Rockland, ME) 纯化, 并克隆至 pBANE6 (WO 98/11203) 中, 产生 pBANE8。

用 SwaI 和 PacI 消化 pBANE8, 并将该 0.9kb 脂酶片段与 SwaI/PacI 消化的 pDM181 连接, 产生脂酶表达质粒 pDM194 (图 6)。

从 pJRoy35 (图 7) 构建 pDM218。通过 PCR 在尖镰孢胰蛋白酶启动子 5'末端引入 NotI 和 PmeI 限制性位点。采用 pDM181 作为模板, 利用如下引物, 制备含有该启动子 5'末端的 362bp 扩增子:

multi 1: 5'-ATAAGAATGCGCCGCTAGTTTAACTTACAAACCTT CAACAGTG-3' (SEQ

ID NO. 19)

multi 2: 5'-TAGCATCTATCTCCGTCTT-3' (SEQ ID NO. 20)

100 μ l PCR 反应物含有 150ng 3kb EcoRI 片段、每个引物各 50pmol、1 $\times$ Pwo 缓冲液, dATP、dCTP、dGTP 和 dTTP 各 200 μ M, 和 5 单位 Pwo DNA 聚合酶。所用的 PCR 条件是 95 $^{\circ}$ C 3 分钟一个循环; 之后 30 个循环, 每个循环包括 95 $^{\circ}$ C 1 分钟、50 $^{\circ}$ C 1 分钟和 72 $^{\circ}$ C 1.5 分钟; 之后 72 $^{\circ}$ C 延伸 5 分钟。将该扩增子亚克隆至 pCR-Blunt 载体 (Invitrogen, Carlsbad, CA) 中。在 1% 琼脂糖凝胶上 100 伏电泳 0.25kb NotI/BcuI 片段 1 小时, 将其切下并采用 QIAquick 凝胶提取试剂盒 (Qiagen, Santa Clarita, CA) 进行纯化。

通过 PCR 在尖镰孢胰蛋白酶终止子 3' 末端引入 HpaI、SnaBI 和 PpuMI 位点。采用该 3kb EcoRI 片段作为模板, 利用如下引物, 制备含有该胰蛋白酶终止子 3' 末端的 714bp 扩增子:

multi3: 5'-GTGTGCAGTGACCCAGAAT-3' (SEQ ID NO. 21),

multi4: 5'-GATTGGGTCCCTACGTAGTTAACTATAGGCCATCGTTTAC-3' (SEQ ID NO. 22)

100 μ l PCR 反应物含有每个引物各 50pmol、150ng 3kb EcoRI 片段、1 $\times$ Pwo 缓冲液, dATP、dCTP、dGTP 和 dTTP 各 200 μ M, 和 5 单位 Pwo DNA 聚合酶。所用的 PCR 条件与上面列出的条件相同。将该扩增子亚克隆至 pCR-Blunt 载体中。按以上所述分离 0.62kb NheI/PpuMI 片段。

从 pDM194 分离含有尖镰孢胰蛋白酶启动子 3' 末端、*Humicola lanuginosa* 脂酶基因 (Geneseq 核苷酸登录号 N91076)、和尖镰孢胰蛋白酶终止子 5' 末端的 2.3kb BcuI/NheI 片段。

将该 0.25kb NotI/BcuI 片段、2.3kb BcuI/NheI 片段和 0.62kb NheI/PpuMI 片段一起克隆在用 NotI 部分消化和用 PpuMI 完全消化的 pDM194 中, 产生 pDM218 (图 8)。

实施例 14: 质粒 pMWR60 的构建

质粒 pMWR60 来源于 pEJG25A 表达载体。pEJG25A 是按如下步骤通过

在 pDM181 (实施例 11) 中插入一段来自 *Peniophora lycii* 的肌醇六磷酸酶编码序列 (WO 98/28408) 构建的:

首先, 设计了两个合成的寡核苷酸引物 (显示于下), 从质粒 pA1phy2 (WO 98/28408) 模板通过 PCR 扩增 *Peniophora lycii* 肌醇六磷酸酶编码序列。

正向引物: 5'-ATTTAAATATGGTTTCTTCGGCATTTCGC-3' (SEQ ID NO. 23)

反向引物: 5'-TTAATTAACCTATTCGACGGAACAAAGC-3' (SEQ ID NO. 24)

(粗体字母表示编码序列)

每个 100 μ l Pwo 聚合酶反应含有每个引物各 50pmol、1ng 模板 DNA、2 μ l 10mM dNTPs, 1 $\times$ Pwo 聚合酶缓冲液, 和 2.5 单位 Pwo 聚合酶。在 Perkin Elmer 9600 型热循环仪中按如下程序孵育这些反应物: 94 $^{\circ}$ C 2 分钟、55 $^{\circ}$ C 30 秒和 72 $^{\circ}$ C 1 分钟, 1 个循环; 94 $^{\circ}$ C 15 秒、55 $^{\circ}$ C 30 秒和 72 $^{\circ}$ C 1 分钟, 9 个循环; 94 $^{\circ}$ C 15 秒、55 $^{\circ}$ C 30 秒和 72 $^{\circ}$ C 1 分钟, 15 个循环, 每个循环延长 20 秒; 最后一个循环, 94 $^{\circ}$ C 15 秒、55 $^{\circ}$ C 30 秒和 72 $^{\circ}$ C 7 分钟; 之后为一个 4 $^{\circ}$ C 保温循环。在 1% 琼脂糖凝胶上 100 伏电泳该反应产物 1 小时。切下 1.3kb 条带, 并采用 Qiaex II 纯化。随后将该纯化的 PCR 产物克隆在质粒 pCR2.1 (Invitrogen, Carlsbad, CA) 中, 用于转化大肠杆菌 TOP10 细菌 (Invitrogen, Carlsbad, CA)。从几个转化菌落中分离质粒 DNA, 并通过 DNA 序列测定进行分析以确保在 PCR 期间没有引入突变。随后用限制性内切酶 *Swa*I 和 *Pac*I 消化一个肌醇六磷酸酶克隆。在 1% 琼脂糖凝胶上 100 伏电泳这些片段 1 小时, 将其切下并采用 QiaexII 纯化。将 1.3kb 肌醇六磷酸酶基因片段与预先经 *Swa*I 和 *Pac*I 切割的 pDM181 (实施例 11) 连接, 产生表达质粒 pEJG25A (图 9)。

为了除去 5'侧翼 DNA 中的外源接头序列, 采用 Quick-Change[™] 定点诱变试剂盒以及下列寡核苷酸引物诱变 pEJG25A, 获得载体 pMWR60-Int2:
引物 A: 5'-CTCTTGATATCTATCTCTTCACCATGGTTTCTTCGGCATTTCGC-3' (SEQ ID NO. 25)

引物 B: 5'-GCGAATGCCGAAGAAACCATGGTGAAGAGTAGATATCCAAGAG-3' (SEQ ID NO. 26)

通过 DNA 序列分析验证了这些设计的核苷酸变化。

然后用 SfuI 和 NheI 消化 pMWR60-Int2, 在 1% 琼脂糖凝胶上 100 伏电泳该 1.8kb 片段 1 小时, 将其切下并采用 QiaexII 纯化。将该分离的片段与 pDM218 的 SfuI-NheI 载体片段连接, 产生一个新的中间物, 称为 pMWR60-Int3。用 NotI 切割该中间物, 按以上所述纯化 5.35kb 片段。将该纯化片段与预先经 NotI 消化过的 pSheB1 (描述于实施例 12) 连接。所获中间物命名为 pMWR60-Int4a。然后用 BspLU11I 和 NheI 切割 pMWR60-Int4a, 并按以上所述纯化 5.25kb 片段。该分离的片段与也经 BspLU11I 和 NheI 消化、并按上述纯化的 pSheB1 连接。所获载体产物命名为 pMWR60 (图 10)。

实施例 15: 脂酶报道基因的制备

为了构建称为 LIPOLASE™ 的 *Humicola lanuginosa* 脂酶 (Novo Nordisk A/S, Bagsvard, 丹麦) 的变体, 根据厂家说明使用 Chameleon 双链定点诱变试剂盒。

从 pAHL (WO 92/05249) 获得编码该 LIPOLASE™ 酶的基因。根据厂家说明, 利用引物 7258 将 pAHL 氨苄青霉素基因的 ScaI 位点改变为 MluI 位点, 由此改变氨苄青霉素抗性基因中 ScaI 位点并换为 MluI 切割位点。

引物 7258: 5'-GAATGACTTGTTGACGCGTCACCAGTCAC-3' (SEQ ID NO. 27)

然后将含有 LIPOLASE™ 基因的 pAHL 载体作为模板, 和寡聚物 7258 和 7770 一起用于 DNA 聚合酶扩增, 由此改变 LIPOLASE™ 基因中的 ScaI 位点, 而又不改变氨基酸序列位置。

引物 7770: 5'-TCTAGCCCAGAATACTGGATCAAATC-3' (SEQ ID NO. 28)

通过加入含有期望突变的适合寡聚物, 在 LIPOLASE™ 基因中引入期望突变 (例如半胱氨酸的引入)。

使用上述的定点诱变和以下引物构建含有编码 LIPOLASE™ 变体 1S, E239C, Q249R 的基因的质粒:

以下显示的引物 106659 用于引入 E99N, N101S。

引物 107581:

5'-CTTAACTTTGACTTGAAAAACATATCTGACATTTGCTCC-3' (SEQ ID NO. 29)

以下显示的引物 101782 用于引入 SPPCGRRP (-E)。

引物 101782:

5'-GGACGGCCTTGGCTAGCCCTCCGTGCGGCCGCCGGCTCTCGCAGGATCTGTTTAAC-3' (SEQ ID NO. 30)

以下显示的引物 9639 用于引入 E239C。

引物 9639:

5'-ATATCGTGAAGATATGCGGCATTGATGCCACC-3' (SEQ ID NO. 31)

以下显示的引物 8829 用于引入 Q249R。

引物 8829:

5'-GGCGGCAATAACCGGCCGAACATTCCGGATATCCC-3' (SEQ ID NO. 32)

通过对该完整基因测序验证了这些突变。所获质粒命名为 pEVi1163。

实施例 16: pRaMB60 的构建

通过来自 pMWR60 的 6.5kb SfuI-NheI 载体片段与来自 pSheB1 的 0.5kb SfuI-NheI 片段连接, 构建质粒 pRaMB60。通过琼脂糖凝胶电泳分离所有片段, 然后采用 BioRad Prep-a-Gene 试剂 (BioRad Laboratories, Hercules, CA) 从凝胶切块中将其纯化出来。

实施例 17: 含有葡糖淀粉酶启动子的表达载体 pRaMB62 的构建

采用以下引物和 Quick-Change™ 诱变试剂盒, 通过质粒 pFAMG 的定点诱变产生单一 BspLU11I 位点:

引物 1: 5'-dCACTGCTATCACCAACATGTTTACTCAAGTCC-3' (SEQ ID NO. 33)

引物 2: 5'-dGGACTTGAGTAAACATGTTGGTGATAGCAGTG-3' (SEQ ID NO. 34)

通过 DNA 序列测定验证该定点改变, 所获质粒衍生物命名为 pMWR62-Int1。

接着, 采用质粒 pEVi1163 作为模板并利用下列引物, 通过 PCR 扩增实施例 15 中所述的脂酶报道基因, 所述引物在该脂酶编码区的起始密码子附近引入一个 BspHI 位点, 在终止密码子之后引入一个 PacI 位点:

正向引物: 5'-GACTCATGAGGAGCTCCCTTGTGCTGTTC-3' (SEQ ID NO. 35)

反向引物: 5'-TGATTAATTAACCTAAAGACATGTCCCAATTAAC - 3' (SEQ ID NO. 36)

该 PCR 反应物由 1 μ l 模板 DNA(10ng)、1 μ l 正向引物(77pmol)、1 μ l 反向引物(81pmol)、10 μ l 10 $\times$ Pwo 聚合酶缓冲液、16 μ l 1.25mM dNTP 混合物、和 1 μ l (2.5 个单位) Pwo 聚合酶组成。采用以下温度设置在 Perkin-Elmer 480 型热循环仪中孵育该反应物: 94 $^{\circ}$ C 5 分钟 1 个循环; 95 $^{\circ}$ C 1 分钟、47 $^{\circ}$ C 1 分钟和 68 $^{\circ}$ C 2 分钟, 30 个循环; 以及 4 $^{\circ}$ C 保温循环。

然后用 BspHI 和 PacI 消化该扩增产物, 通过琼脂糖凝胶电泳纯化(实施例 16), 之后和 pRaMB60 的 5.8kb PmeI-NcoI 片段以及 pMWR62-Int1 的 2.1kb StuI-BspLU11I 片段一起用于三部分连接。所获载体命名为 pRaMB62(图 11), 含有处于 *Fusarium venenatum* 葡糖淀粉酶启动子控制下的脂酶报道基因。

实施例 18: 含有“Daria”启动子的表达载体 pRaMB64 质粒的构建

采用 PCR 和以下设计分别在启动子片段的 5'和 3'末端引入 SwaI 和 BspLU11I 位点的引物对, 扩增称为“Daria”启动子的 pEC03 启动子区域(实施例 10):

引物 1: 5'-GCATTTAAATTACTACTGTGATGTG-3' (SEQ ID NO. 37)

引物 2: 5'-GATTGATGTGAAACACATGTTGATG-3' (SEQ ID NO. 38)

该 PCR 反应物由 1 μ l pEC03 (10ng)、1 μ l 正向引物(50 pmol)、1 μ l 反向引物(50 pmol)、10 μ l 10 $\times$ Pwo 聚合酶缓冲液、16 μ l 1.25mM dNTP 混合物、和 1 μ l (2.5 个单位) Pwo 聚合酶组成。采用以下温度设置在 Perkin-Elmer 480 型热循环仪中孵育该反应物: 95 $^{\circ}$ C 5 分钟 1 个循环; 95 $^{\circ}$ C 1 分钟、47 $^{\circ}$ C 1 分钟和 68 $^{\circ}$ C 2 分钟, 30 个循环; 以及 4 $^{\circ}$ C 保温循环。

该 0.9kb PCR 产物在 1% 琼脂糖凝胶上 100 伏电泳 1 小时, 切下并采用 Qiaex II 纯化。

将该扩增 DNA 片段亚克隆至 pCR2.1 (Invitrogen, Carlsbad, CA) 中, 然后通过 SwaI、BspLU11I、EcoRI 和 XhoI 限制性酶切割进行分析以验证其正确性。以此方式产生的质粒命名为 pEC04, 用 SwaI 和 BspLU11I 消化

之，并按实施例 16 所述方法通过凝胶电泳纯化该 0.9kb 启动子片段。将该纯化片段与 pRaMB60 的 5.8kb *PmeI*-*NcoI* 片段以及编码该脂酶报道基因的 0.9kb *BspHI*-*PacI* 片段混合进行三部分连接。该连接产物 pRaMB64 (图 12) 含有指导该脂酶报道基因表达的“Daria”启动子。

实施例 19: 含有来源于液泡相关蛋白质基因的启动子的表达载体 pRaMB66 质粒的构建

采用下列引物通过 PCR 扩增来自编码推测的液泡相关蛋白质的 pQUINN(实施例 10) 的启动子片段，所述引物在该启动子的 5'和 3'末端分别引入 *SmaI* 位点和 *NcoI* 位点:

引物 1: 5'-dCGACCCGGAATTAGAGAGGTTAGG-3' (SEQ ID NO. 39)

引物 2: 5'-dCGTATAACCCATGGTGGACTTGTCGGAC-3' (SEQ ID NO. 40)

该 PCR 反应物由 1 μ l pQUINN (10ng)、1 μ l 正向引物(50 pmol)、1 μ l 反向引物(50 pmol)、10 μ l 10 $\times$ Pwo 聚合酶缓冲液、16 μ l 1.25mM dNTP 混合物、和 1 μ l (2.5 个单位) Pwo 聚合酶组成。采用以下温度设置在 Perkin-Elmer 480 型热循环仪中孵育该反应物: 95 $^{\circ}$ C 5 分钟 1 个循环; 95 $^{\circ}$ C 1 分钟、47 $^{\circ}$ C 1 分钟和 68 $^{\circ}$ C 2 分钟, 30 个循环; 以及 4 $^{\circ}$ C 保温循环。

该 3.1kb 扩增 DNA 片段在 1% 琼脂糖凝胶上 100 伏电泳 1 小时, 切下并采用 Qiaex II 纯化。

将该 3.1kb 扩增 DNA 片段亚克隆至 pCR-Script (Stratagene, La Jolla, CA) 中, 产生中间质粒 pQUINN-启动子 A。从该质粒分离两个限制性片段: 一个 2.4kb *SmaI*-*NdeI* 片段, 和一个 0.7kb *NdeI*-*NcoI* 片段(这两个片段一起包括了该完整启动子区域)。将该分离片段与 pRaMB60 的 8kb *PmeI*-*NcoI* 片段和前面实施例中所述编码脂酶报道基因的 0.9kb *BspHI*-*PacI* 片段混合进行四部分连接。该连接产物 pRaMB66 (图 13) 含有处于推测的液泡相关蛋白质基因启动子转录控制之下的脂酶报道基因。

实施例 20: 在 *Fusarium venenatum* 中表达位于 AMG、“Daria”和液泡相关蛋白质的启动子控制下的脂酶报道基因

通过用从添加了 2.5%葡萄糖和 2.5mM 硝酸钠的 1 $\times$ Vogels 培养基平板(2.5% 纯净琼脂)获得的 10 个孢塞(plug)接种一个含有 500ml RA

孢子形成培养基的三角瓶中, 28℃、150rpm 孵育 2-3 天, 产生 *Fusarium venenatum* CC1-3 (MLY-3) 的孢子。通过 Miracloth (Calbiochem, San Diego, CA) 收获孢子, 并在 Sorvall RC-5B 离心机 (E. I. DuPont De Nemours and Co., Wilmington, DE) 内以 7000rpm 离心 20 分钟。用无菌蒸馏水洗涤沉淀的孢子两次, 重悬在少量水中, 然后采用血细胞计数器计数。

通过将 4×10^7 个 *Fusarium venenatum* CC1-3 孢子接种在 100ml YEPG 培养基中, 于 24℃、150rpm 孵育 16 小时, 制备原生质体。培养物在 Sorvall RT 6000D (E. I. DuPont De Nemours and Co., Wilmington, DE) 中以 3500rpm 离心 7 分钟。用 30ml 1M MgSO_4 洗涤沉淀两次, 并重悬在 15ml 含有 5mg/ml NOVOZYME 234™ (batch PPM 4356, Novo Nordisk A/S, Bagsværd, 丹麦) 的 1M MgSO_4 中。培养物于 24℃、150rpm 孵育, 直到原生质体形成。向原生质体消化物中加入 35ml 体积的 2M 山梨糖醇, 该混合物以 2500rpm 离心 10 分钟。重悬沉淀, 用 STC 洗涤两次, 并以 2000rpm 离心 10 分钟沉淀原生质体。用血细胞计数器计数原生质体, 并将其重悬在 8:2:0.1 的 STC:SPTC:DMSO 溶液中至终浓度 1.25×10^7 个原生质体/ml。在 Nalgene Cryo 1℃ 冷冻箱 (VWR Scientific, Inc., San Francisco, CA) 中控制速率地进行冷冻后, 将该原生质体保存在 -80℃。

在冰上融解冷冻的 *Fusarium venenatum* CC1-3 原生质体。在各个 50ml 无菌聚丙烯管中加入 5-10 μg pRaMB62、pRaMB64 和 pRaMB64 (描述于实施例 17-19 中) 以及 5 μl 肝素 (每 ml STC 含 5mg)。每管加入 100 μl 原生质体, 轻柔混合, 并在冰上孵育 30 分钟。加入 1ml SPTC 并在室温孵育 20 分钟。加入 25ml 40℃ COVE 顶层琼脂糖后, 将每管中的混合物铺在空的 150mm 直径平板上, 并于室温孵育过夜。大约 24 小时后, 在平板顶部再铺上含有 10mg BASTA™/ml 的另外 25ml 40℃ COVE 顶层琼脂糖, 并在室温孵育长达 14 天。除草剂 BASTA™ 中的活性成分是膦丝菌素。BASTA™ 从 AgrEvo (Hoechst Schering, Rodovre, 丹麦) 获得, 并在使用前用酚: 氯仿: 异戊醇 (25:24:1) 抽提两次, 用氯仿: 异戊醇 (24:1) 抽提一次。

从筛选平板 (COVE 底层, COVE-BASTA™ 覆盖层) 直接挑取转化体至含有 25ml 添加了 1mM CaCl_2 和 100 $\mu\text{g}/\text{ml}$ 氨基青霉素 (防止细菌污染) 的 M400Da 培养基的 125ml 摇瓶中, 并于 28℃、200rpm 在平台摇床上孵育 7 天。还包括未转化受体株作为阴性对照。

在第 7 天从摇瓶取样。通过离心除去细胞,然后将每个上清样品 10 μ l 与等体积 Tris-甘氨酸样品缓冲液 (Novex Experimental Technology, San Diego, CA) 一起加热至 95℃ 5 分钟。在 10-20% Tris-甘氨酸梯度凝胶 (Novex Experimental Technology, San Diego, CA) 上分离变性的上清液蛋白,并用考马斯兰染色。SDS-PAGE 分析显示,产生脂酶的转化体分泌一种表观分子量为大约 43kDa 的主要多肽。

相似地,采用对硝基苯丁酸作为底物分析来自每个转化体的无细胞培养液的脂酶活性(Royer 等, 1995, 同上)。

结果显示于表 1, 这些结果表明,采用存在于 pRaMB62、pRaMB64 和 pRaMB66 中的 *Fusarium venenatum* 启动子元件表达并分泌了活性脂酶。

表 1

所用载体	启动子	转化体	脂肪酶产量(LU/ml)
无	--	阴性对照	< 1.0
pRaMB62	葡糖淀粉酶	RaMB62.1	1158
"	"	RaMB62.2	500
"	"	RaMB62.3	1379
"	"	RaMB62.4	1678
"	"	RaMB62.5	702
"	"	RaMB62.6	616
"	"	RaMB62.7	473
"	"	RaMB62.8	894
"	"	RaMB62.9	564
"	"	RaMB62.10	1036
"	"	RaMB62.11	2731
"	"	RaMB62.12	1960

"	"	RaMB62.13	1682
"	"	RaMB62.14	572
"	"	RaMB62.15	1421
pRaMB64	"Daria"	RaMB64.1	1217
"	"	RaMB64.2	561
"	"	RaMB64.3	875
"	"	RaMB64.4	839
"	"	RaMB64.5	1449
"	"	RaMB64.6	354
"	"	RaMB64.7	377
"	"	RaMB64.8	184
"	"	RaMB64.9	1967
"	"	RaMB64.10	657
"	"	RaMB64.11	883
"	"	RaMB64.12	184
"	"	RaMB64.13	1935
"	"	RaMB64.14	1049
"	"	RaMB64.15	875
pRaMB66	液泡相关 蛋白质	RaMB66.1	1990
"	"	RaMB66.2	165
"	"	RaMB66.3	380
"	"	RaMB66.4	155
"	"	RaMB66.5	170
"	"	RaMB66.6	145
"	"	RaMB66.7	180
"	"	RaMB66.8	420
"	"	RaMB66.9	200
"	"	RaMB66.10	195
"	"	RaMB66.11	190
"	"	RaMB66.12	165

"	"	RaMB66.13	140
"	"	RaMB66.14	435
"	"	RaMB66.15	125

实施例 21: 比较处于 *Fusarium venenatum* 淀粉葡萄糖苷酶启动子和尖镰孢胰蛋白酶启动子控制之下的 *Humicola lanuginosa* 脂酶报道基因的表达

将按实施例 20 所述制备的含有 pRamB64 (实施例 17) 或 pDM218 (实施例 18) 的 *Fusarium venenatum* CC1-3 转化体, 于 pH6.25 含有适合碳源、氮源和痕量金属源并补加了尿素或磷酸铵作为饲料的 2 升发酵罐中, 29℃、1200rpm 培养 180 小时。采用对硝基苯丁酸作为底物 (Royer 等, 1995, 同上) 在 180 个小时内每隔 18-24 个小时分析每个发酵物的无细胞培养液的脂酶活性。

图 14 显示了在 *Fusarium venenatum* 中处于 *Fusarium venenatum* 淀粉葡萄糖苷酶 (pAMG) 启动子或尖镰孢胰蛋白酶 (pSP387) 启动子控制下的 *Humicola lanuginosa* 脂酶报道基因表达的比较。结果说明, 无论在饲料中采用尿素还是磷酸铵作为氮源, 采用 *Fusarium venenatum* 淀粉葡萄糖苷酶启动子比采用尖镰孢胰蛋白酶启动子可以获得更高水平的脂酶活性。而且, 当在饲料中采用磷酸铵时, 观察到更高的脂酶活性水平。

生物材料的保藏

如下生物材料已根据布达佩斯条约保藏在农业研究机构保藏中心, 1815 University Street, Peoria, Illinois, 61640, 并给予了如下保藏号:

保藏物	保藏号	保藏日期
大肠杆菌 TOP10 (pEC03)	NRRL B-30067	1998 年 10 月 27 日
大肠杆菌 DH10B (pFAMG)	NRRL B-30071	1998 年 10 月 27 日
大肠杆菌 DH10B (pQUINN)	NRRL B-30075	1998 年 10 月 27 日
大肠杆菌 TOP10 (pFB0346)	NRRL B-30073	1998 年 10 月 27 日

这些菌株在如下条件下保藏，即确保在本专利申请未决期间由专利和商标局局长依照 37 C.F.R. §1.14 和 35 U.S.C. §122 授权的人可获得该培养物。该保藏物代表该保藏菌株基本上纯的培养物。如果在本主题申请的等同申请或其派生申请提出的国家该外国专利法要求，可以获得该保藏物。但是，应当理解，可获得保藏物并不构成以损失政府法令所授予专利权利而实施本主题发明的许可。

本文描述和要求保护的本发明的范围并不被本文公开的具体实施方案限制，因为这些实施方案旨在用作本发明几个方面的举例说明。任何等同的实施方案都包括在本发明的范围中。实际上，从前面的描述来看，除了本文给出和描述的之外本发明的各种修改将为本领域技术人员所明了。这些修改也包括在所附权利要求的范围内。在冲突的情况下，包括定义在内的本公开将优先。

本文引用了各种参考文献，它们的公开以参考文献的方式完整地并入本文。

序列表

<110> Randy M. Berka
Michael W. Rey
Kimberly Brown

<120> 用于在真菌细胞中表达基因的启动子

<130> 5611.204-WO

<140> 等指定

<141> 2000-03-22

<150> 09/274,449

<151> 1999-03-22

<160> 40

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 6050

<212> DNA

<213> 镰孢霉属

<400> 1

aatttcgtcg	atagcgaggg	actcctggcc	ctcgaattta	gttagcgat	cagtgtaaag	60
tgctgggttc	tccaggcgta	agtaaattga	accagatgtt	agctcccaga	tttcgccccg	120
aagccgggtg	ggcagaccaa	cgcgataag	tttatggaaa	gttgggtggc	ggatcaatgt	180
aaagttcctg	ccattgtcac	gcagatactc	ggcccaaaga	cgcatctttg	ccctatcgcg	240
caactttttt	gggtccccag	gatatacgaa	gagcattcct	aagccagcat	ctggtgggag	300
atcggttctt	ttatcttcgg	ttcttaaaag	atggttcagag	taacactcag	cagcaactcg	360
acgcaacttg	gctacgttgc	caacgcctgc	tctaagaccc	ttcttgaggc	cgtcacagaa	420
acgttcgcag	gactgtctgg	ttccagcgag	atggattgtt	atgcgttggt	ctcgggggtc	480
tctcttgtct	ttagaggtat	cggcaagaag	gccattccac	gtagtcagag	cgagtgcgaa	540
ctataaccag	gcaacgtcag	aatttgtacc	atgcaagatt	tttataccac	atacctggaa	600
gttctggcta	ttcaagcgct	ctacccttcg	tatagcacat	aaaggaaatg	tgaatccatt	660
gccactgggt	cctcctccgt	gtgtctggcc	ggtaaaagcg	gtcgaggttg	aaaggctagc	720
agattgcaca	aagctagtgg	gtgttggtga	gaagcacatg	taatgctctg	acaggtgtag	780
cttgccagca	taatgccatc	ctcgatcgtg	atccttatca	ccgtgagtag	cgttggaggg	840
cgggatatga	agctcggcgt	tgatctcgta	gaggggcgcg	tgagaggcgg	gcaagcgaaa	900
ttggctctga	aatagcgatg	actttgaggg	attccgatct	gaactggata	gattgagccc	960
ctgggtagga	tcgataagct	gctgggcttt	ttgaacgatg	ttggtgaagt	tcgaaaacat	1020
gatttcgggc	agcggcgcat	gacgaggggg	gttccgggtc	aggagggagg	tcgcggtga	1080
gcttgaagga	gatgcaagac	acgaagcgaa	agacacgaag	agagcgcaag	agtctgagta	1140
tgtgcaacca	ggctcgaata	agtgaaggc	aggcagaagt	acggaataga	cgatagaatt	1200
gagtatagaa	aggctgaatg	gaagatggag	acgagttata	ggacgggtga	gatagagtgg	1260
agttgaagtt	gaacgaagct	gcgtcaggtc	cagatacggg	agactggcca	tcaactactg	1320
gccaggtagc	cagggcgcg	tgggcgggtg	ggcaggggtc	cgggggggac	ctcagggcat	1380
tcctttctcc	aagggcgcgt	ggggctatgg	acggggctgg	ctgaactcca	gccgtcatgg	1440
gatagcgggtg	caagagatca	ggtactaagt	ctaccatgat	aatttagggg	gcagagaaaa	1500
atgatataatt	tgttttagtag	taagcgggtt	tttacagttg	aggaaccaac	cttcttcatt	1560
tattttattct	ttctttctct	gcaattcagt	cctttttctt	aaatagaata	tctaccaatg	1620
gaacggcgtg	gctgaagtgg	ctgaagaata	tagctcgagc	tgtcaaaccg	ctcatcctac	1680
taccctaggt	ataaagctgg	gaactaagac	tcattttctat	ccaactcatc	atattgggag	1740
ttagtgtaga	cctgtcggcc	tagagaatat	gtgtatctgc	atactttcaa	ataccctacg	1800
tatacccact	atgttttagca	caatcattga	cctctcaagg	cctcaccat	ctcaacacct	1860
gtcgtgtgct	cacttgacta	cttctttgaa	ccagctcgcc	atcggactag	tcgaacaagc	1920

ttgtcgcccc	catacagatg	aatgtatggt	taaagctaca	tgatcagcct	gaaccgagca	1980
taactcgagt	gccgagactc	ctctgatgta	tatcgagatg	aatgacaaac	ctacgggtcc	2040
gttcttgaga	agtggcctga	gattttctcac	ttggtgagaa	aaaggacggg	cgagcgggag	2100
cctgagtcag	aagaaatacc	tgtctccttg	gatctcacat	gacggtgttg	tggaagagtg	2160
catctattgt	cattgctgga	gtgacggcag	agtaggggtc	taaagaaacc	catactgagt	2220
agagatggag	aagacaacaa	aagcccaaga	cgacagagac	gacagaagat	taaagctatc	2280
agagcgagac	tatatcacta	ttcgaaacct	gcgagtaatt	taacaagaag	tacacatcat	2340
cattgttatc	aattcgacga	agacatgggc	gaaaattctt	gcggtgtata	tgtctgttgt	2400
atatgggcct	gggcattgtt	atTTTTcgcc	gtctttatgt	gtactaacac	ttccattgat	2460
accccgagac	aaaagatgaa	cgcttaaaca	gcacccaaat	caggagaaga	atggcgctgc	2520
tctaggtatg	cttctgggat	aaaaagcgat	gttgatacct	ctcagaaaag	aagtgtattg	2580
aagttgaatc	aaacaaatag	ccgatggagc	gatctgaagg	ggtggcagac	ctgctacgcg	2640
catttaggca	aggcatcaac	tcggcagatg	attaagaaaag	gttttgtagg	ttcacgtgtt	2700
gtgttggtgt	ccattataag	tttataacct	tgctaagatg	caacgactct	gacctcaggg	2760
tgttagaaaa	attgaccact	aggagcataa	gtgacgaaat	tcggggatca	agacaataga	2820
tagtttcatt	ttcatgtgct	cctacgtctt	ttcacgtaat	gtttcttata	aaaaaaaaaga	2880
tagcattgtc	tctttggtga	aaagagaaaa	aaagatgtta	cgacgtggcc	ttgattcgaa	2940
cagacgcctc	cgaagagaat	agattttctag	tctatcgctg	tagaccactc	cgccaccacg	3000
ccttaacgtaa	tctgtgattg	ttgaaagtta	ctctcgtgtt	acggtctata	cgtgaagaat	3060
ctacacttga	cgagtctcga	ggtctggggg	cagttagacg	gaaatgggag	aacaaagaga	3120
cttggtgaca	ttgcaggcaa	ccgggtagat	gttgaggtca	ttgatcggac	aagattgttg	3180
cttcaaaagt	aacaggtatt	ctttttttta	atcaacagaa	acgttccatg	ttcatttggt	3240
aatccaatct	atTTgtgata	gcgtttgatg	acaaacaata	ataatgatgg	tctggcggct	3300
agtgatcggt	tgtaatgacg	tcgtcatata	tcctatcact	atacagttgc	tttgcacacg	3360
cactcacgtc	cttcattcgt	tgtcttcact	atTTgatggg	gatttgggtc	aacaacctac	3420
agaaataatg	acctgtgggtg	ttctccgaat	atggctagac	caacacaagc	ttgtaccgcg	3480
gcattcaaat	caccatgtga	tgcccatcat	cagatcatcc	accaacccaa	aaacagacca	3540
actactcaca	aaaaggcatc	tcatacaaga	aaaacggcca	actaacgtcc	aaaaggcccg	3600
aaaaacgtcc	atacgcgcgc	agccgagact	tcaatagact	gcacaagaag	gaccgatgag	3660
atcgaccaga	ctaaaccggg	gagagtgtca	aatatgcggg	ggattgggga	acttacccca	3720
gaaaagagaa	ggaggataaaa	ttccatgtct	ggggttgacg	tctctattgg	ttagacacga	3780
acgcctgctc	tcggcgtaat	ttataccata	gcgccaatga	gggcggaaac	tcctgttttg	3840
tcaagtgcgc	attgttggtt	gggtcatgat	atatagccag	taggtatccg	tcttggtgat	3900
tgaccagaca	tatcgctcat	cacagatcaa	catcactgct	atcaccaaca	tgcttactca	3960
agtcctttat	ggcttggtag	ccagtgcctt	ttggcaaggc	caagtcgttg	catcaccaag	4020
caaggacaat	tcactggagc	gcttcattga	caaacaagct	gatatttcta	tcaaggggtg	4080
ccttgctaata	attggcgctg	atggaaaaag	ggcacagggg	gcagcgccctg	gtgctgttgt	4140
ggcaagtcca	tcgaaagaag	atcctgattg	taagccagca	tcctaccttg	tccttgtccg	4200
catgctaata	atgggtctcag	attggtacac	ttggactcgt	gactctgctt	taacgtacaa	4260
agtgtcgtt	gagagattca	tccacggcga	caaatctctc	caacgaaaga	tagatgaata	4320
tgtctccgca	caagcgaaac	tgcaagggac	cacaaatcca	tcgggcagcc	cagagtcggg	4380
cggctccggc	gagccaaagt	tccatgtgaa	tctcactgct	ttcactggat	cttggggctg	4440
gcctcagcgc	gacggccctc	cgcttcgggc	taccgccttg	actctgtatg	cagaatggct	4500
catttcccac	ggcgaaagat	ccaaggcttt	gaacaaaagtc	tggccagtca	tcgagaaagg	4560
ccttgcgat	actaccaagt	tctggaatcg	cactggctat	gatctatggg	aggaggttaa	4620
tggatcttct	ttctttacac	tttcggcttc	gcacgtgct	cttgtcgaag	gtgccgctct	4680
ggctaagaaa	cttggcaaat	cttgtcctga	ctgtgtcacc	aaogctcctc	gcgttctgtg	4740
cttcttctag	actttctgga	ctgggtggct	cgttgactcc	aacattaacg	tcaaggatgg	4800
tcgcaagggt	ctcgatgtca	actccatcct	ctcgtccatt	catacattog	atcccaactc	4860
caagtgcacc	gactcgacgt	tcagcccttg	ttcaccacga	gctcttgoga	accacaaggc	4920
ggtcgtcgat	tctttcaggt	caatctatgg	tgtcaacaag	aatagagggtc	aaggcaaggc	4980
cgcggtgtt	ggtcgatata	gcgaggacgt	gtactatgat	ggcaaccott	ggtacctggc	5040
cactcttgct	gctgcagAAC	aactctacgc	tgcggtctac	cagtgggata	agcttggcgc	5100
tgttactgtt	gacgatgtat	ctttgtcttt	cttcaaggat	atcgttccca	aggtctccaa	5160
aggcacttat	gccaaagaaga	ccaagacata	caaggagatc	atcaaagcag	ccaagactta	5220
cgccgacggc	tttgtcgctg	tcgtgcagac	atacactccc	aaggacggct	cactagctga	5280
gcaatttgac	aagtcaactg	gagcccccaa	gtccgctgtt	cacctcacct	ggtcctacgc	5340
cgcttttgct	gccacaactg	aacgtcgoga	cggcatcctc	tctccctcct	ggggcgaaag	5400
cagcgccaac	aaggtecccg	cogtgtgtca	agctgcccc	gcattgtgaca	caacctcac	5460
cttcagtgtc	aagaacgtgc	aagtttctatc	cgacccaaag	gtttacgtgg	ttggctcagt	5520


```

gactgagctt tctaactggt cacctgatga tggcattgcg cttacgccat ctagttccgg 5580
agtgtggagc gtcaagggtta agattccttc tgatacaagc tttgagtaca agtatatcaa 5640
gaagactagc agtgggggatg ttacgtgggt gagtgatccc aacaaccggg ctattacggg 5700
tagcaagtgt ggaagtacaa gtactcttga tgatgagtgg aggtagtgga tgacagattt 5760
atcaagctat gtagttttgt gaatatataa ttatccaaat tatcagggtt cggttaagaat 5820
ataattcagt tcagcagctc gtacaagcaa gccatgattc acgcttcctt cgtttggaag 5880
ataagggttc ctcgccaccg tcaagataat tttctcgtgc taatatcacg taatccatct 5940
gcattaaacc cttgccggga aggtttcttc aaccagcaa cccagagta actcggagat 6000
aggggaagtct atttgcotta tcctcgtga cgaattccct gaaccgaatt 6050

```

<210> 2

<211> 200

<212> PRT

<213> 镰孢霉属

<400> 2

```

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

```

<210> 3

<211> 581

<212> PRT

<213> 镰孢霉属

<400> 3

```

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

```


565	570	575	
Asp Asp Glu Trp Arg			
580			
<210> 4			
<211> 4047			
<212> DNA			
<213> 镰孢霉属			
<400> 4			
aattagagag gttagggatt tcacatggcc accaatggga aggaggcaac catctgcacg			60
agcccaccaa gtcatctcct caaactgtgc tgcgactaag aatttgattc cggttctggc			120
ctggcctttg tatcagctag gtcattctcg actaccggag gccaggctga agcagtcagt			180
cacgcattgt cactttatcg gtccctgtcct catacggata cactaggcgt caatgggctt			240
caaacggaga tccagagatc tcatgaagag catcgacgat aaagtgagtg gttggggata			300
ctgtgcggtg ccgaccccgag cggcagccag gttccaccct tgattacatg gttgaaaagt			360
ggcgttactg ggcgagatca aatttggcat gtatgttcgt ccaatgacgc gagctctcca			420
tggtgtctcg aggttcagga accgaccagc atgggatcag tgaggtgaaa tccaaccgag			480
ggagatcgag atctttgtgc tcatatccat gctgccatgc tacgtgccga acaggccaga			540
tggcgttcaa ctacgtcgac caggtcgat gaacgcggag cggtgacgag atcgaagctt			600
catctatcgc ttacgggggtt atgttccact ttccattaac gtttgcgagt tgctgtttga			660
gagccatgtc gaaagcatgg accgtgtcac atctttcaag gtaaatctgg aggtgggaag			720
aagaattgag aggaacagga tgggaggata gcaggctgac gcggaagct aggtagctac			780
ctggctgatt actggctaaa gctggagagc aactaggtaa tatcaggcaa agagctccaa			840
gagctattgg gaggtcggct gattgtctct ggctgagacg caggagggaag ggtaaaatg			900
gccggcaggc caagaagggg ctgcaaaaaca cggagtggat ggtggggcct cccacatacg			960
ggattcgggc tgcggatcta acctcaattt ggcaagagggt aaataggacg acatgcaggc			1020
cccctgacat gtaacaaga caagtggtaa acaagccatc aacatcaaca agagcgaaca			1080
atcgacacac ccatgggggt gagatatggt agtaaggcag agagagatca gggcagcata			1140
cgtgggaaag ggctgggcaa gaaaggacac aacggatcaa cagaacgcag cgctaccgag			1200
ggagcaaac aagtacagta accgtcaca gaggcacaaac tegtccaatc ctgccccgt			1260
cttcaaaagc ccagtttcgt tctgagtcct gtccgggtccc tcttctcctc ctactccctc			1320
caattatcgc catccacatc gacatcgtea attcacaacc tcaccagac aagaagaaaa			1380
gaacgactga aggccttcgc tggccatcac ccgattcttt tccattctct tcgacttttg			1440
tttcgtagga acaagagcca gagaacttct tgtcatcctt tcgaatttcg gaaggttgta			1500
tgagaagctt ctctcgcgcc agcaaaagtc gcaaatctgg actttgaggc acgcgtcccc			1560
gttcccttca gcatcttccc atcgacatat cgggaatccg aatcccacac acagaccgtt			1620
accgaaacaa agatacacga agagggttag atcaaaacccc aacagccccga agccggacgg			1680
gaaggtgaaa tatcttctgt ctccgtcacc gccgaacagg tccctcctcc tcgtcaagag			1740
caagagttta tcgaagaaga ggtccatatt acgcgtgaag aagaacatta ccaccgtccc			1800
gggtgtccaaa aattcgagca cgaagacttt actatccgtg aagactcccg acggtagctt			1860
cgattttaca tttcctttca tctccattta ggtcgcatct ttttcgttac ttttttggtc			1920
aattacacgg gggatacgat tttcccacgg tcggagaaag ccctgcttgc tctctatgcc			1980
taggtctgta ttctctcatc cctctgcgtg gatctggcca tggagacgtg tgagaacaag			2040
actacaattc atcacatcat ttttcgctag gcgaaagcaa ttaccgttgt ccccgacctt			2100
ctcccaacca tcagttttca ctttcccttt tcttggtctg gcttgccctg accattaccc			2160
accgcgcacg gagcgcttca gtcccccagc atcccattct cacatcactt ctcatatcct			2220
ctcttcacac gcctcacaca cccaccccct gcatgctacc atgccaaccc acttcagctt			2280
ggctggatac ccaatttgct ttgcttcttc cccggctcac tagcgctct aagcctgctg			2340
gcctgagcaa ggcggtggag ctatctcagg ggccgcgcgc tcccgttgcc atatgatacg			2400
caaacgactt actatagaca tccatcagct aaccagaca aatctagacc tcaacctccc			2460
tctcaatacc aaccttccca gtaccaccaa ccttcccact accaaccacc tcccaaattc			2520
caaacttctc acactcacgt agagatcgac acccaccgtc atccctacta ctccaccccc			2580
attgatctcg ctgaacgtga ataccgccag cgttaccgcc ctgccccagc tttttccaca			2640
gaagaccctt cttcccactc tcatcctcac taocaaacct aagacaactt caaagccaac			2700
aactacaccg ttgaaggccg acccgctccc caattccatt cctctgagaa gactgaaatc			2760
aacaagttta ctgttgacga acactcctct cggcctcagt acaaccacac cgagaagacc			2820
gaattcaaca actacactgt tgacagccga tcttcccgct ctcaatacaa cactgtgag			2880
aagactgaga tcaacaattt cactgttgac gcccgctctt cccagccacg gtaccgcgac			2940
accaagacaa ctcaagtcaa cagctacgcc gttgacaagc ccgtttctcg tccatcttac			3000

aagaaggacg	tgagattttac	tgaacaaacc	gtcgaagcctt	caaagtcoga	caagtcgaag	3060
atgggttact	acgacgacga	gggtaagtga	aatctgtcac	ccagcgagcg	ccatcaagct	3120
ctctattcgt	gacgcaattc	aagctaacc	agtcaccagg	ttctttccgc	aacggcggca	3180
tccacaagct	cggtgacaag	tcccgcgaca	ttgaggttga	cattcgcgag	acttctcgtc	3240
ctgccaatga	ctgcgctccc	aacaccgtca	gcatcccctg	ccaccacatc	cgtctgggtg	3300
atttcctcat	gctccagggc	cgcccctgcc	aggatcatccg	catctccacc	tcctctgcca	3360
ctggccagta	ccgctacctt	ggtgtcgacc	tcttcaccaa	gcagcttcat	gaggagtctt	3420
ctttcatctc	caaccctgcc	cccagcgttg	togttcagtc	catgctcggc	cctgtcttca	3480
agcagtaccg	tgctcctcgat	atgcaggagg	gtcagatcgt	tgccatgacc	gagactggcg	3540
acgtcaagca	gggtctccct	gtcattgacc	agtcacaact	ctactctcgc	ctccacaacg	3600
ctttcgagtc	cggtcgtggc	tctgttcgcg	tctcgtcct	caacgacggg	ggccgtgagc	3660
ttgccgttga	catgaaggct	atccacggct	ctcgctgtga	agcgtgttca	actgttttct	3720
gaattcgggc	agccgcttgc	aatgcgactt	cttcccaatg	tttaattgag	tgaagggaca	3780
gcactaccag	tctcacctca	actgtgggga	gcgggtctgg	gctgtctcta	atcttacctg	3840
tacaatgtca	agtttcatag	gggacctgtt	gtgtcaagat	ggttcgagtt	ttgtttgtgt	3900
caagattgga	taaatgatat	tggctagctg	gaaataactg	agtcctttgt	gtagatggga	3960
gagttctgta	catgaactat	agtaattgac	aattgattcc	gcactactct	agcttttcat	4020
tggtgctcta	tgcccaacat	gtgaatt				4047

<210> 5

<211> 2517

<212> DNA

<213> 镰孢霉属

<400> 5

aattactact	gtgatgtgat	cacacctaac	taaataccta	actcaccoga	tggtatcgaca	60
aggaaatctc	acgcccttgt	cgagtctcct	ctttcgtctg	tctcctgggc	tcgctactgt	120
ccgattgtaa	ctctcgtctc	ccaacttgtt	caactccta	aagtgggtggc	acaacgtgaa	180
gatgtattgt	tgtgtgaggc	gggggggttg	gtggcattac	caaagagacc	aaaagtcccc	240
ctatgtcgta	ttgatgggtg	tgcgttgcca	tgatacggga	ccccgaatat	gttgatgca	300
tcattcggtac	agaaagctac	tgttcaaaac	gaacggcaaa	gcggattgat	caaccctgga	360
aagaccatgg	gtctctctca	gtccacaatc	ttctcttctc	gatcaaattt	atggatccaa	420
gcggccacaa	ttctagcgcc	atcatgggtc	cctttcctct	tttcgctcac	cccatgttcc	480
ctgtccaccc	tcattcagtg	gacctgatgg	atccctatcc	cccgatgagc	cggggggtgc	540
agccttggcg	ctctcttctt	gttagtgtga	cctactgttg	atttactca	gcagtcctag	600
agtccattta	gttgggcctg	gggtgatggg	gtctgagact	ttgcttcttg	cctgggtctg	660
tctagctcga	atctgtgggt	tgcctggcct	ggcctggcct	gacctgacct	gaggggggtg	720
cccctttgct	ctgttctgca	tatgttgcta	ttagctacct	actcgagaat	tcataaaagg	780
actgtccagc	ccgctctctt	actgacttct	ttcctttccc	tcttcaccct	cgttgtcata	840
tcaaactctgt	cactcgttag	accagactac	cattcccact	ttcgttttta	aactacttta	900
ctcaactaat	tctaatacca	actccaaaa	ccatcaacat	gcgtttcaca	tcaatcctcg	960
ctgccggcgc	tttcgccacc	atggccgctg	cccagagcaa	gaccgtctcc	ctcgacctgt	1020
ctcagcagtc	tcaggccgac	tgcctctccg	actgtgagcc	tggcgatgtc	aagtgcacgt	1080
cttactgcac	cactgtatgt	tacaacaacg	attccctgtg	catgtgtaga	aaactaacia	1140
tcccaatagg	ttccctctcc	tgacgagaag	aacatcgagg	aaaccacca	gtgttggttg	1200
ccgcctgccc	caagggcaag	ggctccgaag	gccgacactg	agaagtacac	cgtttgcatg	1260
aacgagtgtg	tcgccgacaa	ctactggaag	tccgttgatg	gtaccccccg	tggcaccgac	1320
gtccccgatg	tcaagagcaa	ggcctccgag	gctgcctcct	ccgctgctga	gaaggccacc	1380
gccaccggta	ctgctgctga	gtctgatgct	accgccactg	gtgcctccgc	tactgagtc	1440
gagtcgggct	ccgactccag	ctccgaggag	accggctctg	cctctggcac	tgccactggg	1500
accgctgctg	aggtctccga	gactggtaac	gccgcctctt	ccctcgttgg	tggtgtctcc	1560
ttcctcggtc	tcgttgccgc	tatcttcgct	ctgtaaattg	ggtttctctg	tttaggataa	1620
tctgatttgg	catgacggag	aaggatttaa	tgggttttat	tacagcggta	atgattggag	1680
tttgatttgc	aagatgtgac	acgttggaac	gcattgataag	gcctacgggt	ctgatcaatt	1740
tcattggacaa	atattgtttt	tttgggtaat	catttcgcgt	tcacatatgg	ctcgccatat	1800
gagcatgaat	acaatacctc	ttttttgcgc	ctcaattcat	tccaatttct	tgtgatctca	1860
cagtgattca	acttacaagt	tgccggcgca	ccactgaggt	cgtgtctgat	gtgggtcttc	1920
tgtttgtgat	tggctcatga	ttcccaatcg	ggtgcttcaa	acgttagttt	gtaaacaagc	1980
gaaatgaggg	tcttaggatg	catgttcaaa	gcgcaaaacc	caattgaatt	caaattgttaa	2040
agaatcatcg	agaagagcga	gttactgagg	tgaatttgtg	ctttcaactg	tcaatacctc	2100

```

cctcagaaca aatgaattga attattatttc acactcaatg cccaatattc taaacatggt 2160
cggttgtaaca gagttttaat tccttgacgc cacaatgttc cttggtaatt atcgcgcctg 2220
tcacatgaac tggctcctga acttaaacgt tggtgaccca gcaactcgtt taccaggctt 2280
agggtagctg tcatacaaca aacaaacttg tacaattgat gttattgatg aatcatgtat 2340
agaagagcac aattgattta aacacagata aactggctga accgatttta tcagggttggtg 2400
tgaacatgca ttgccgaatc agaaaccaga gtaagactat ctacagttcc atgaagacaa 2460
ttcacagact gccagaaagc aagggtactgg aagcacgaga gacaaaatta ttgaatt 2517

```

<210> 6

<211> 187

<212> PRT

<213> 镰孢霉属

<400> 6

```

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

```

<210> 7

<211> 30

<212> DNA

<213> 镰孢霉属

<400> 7

```

gagctcgagg aattcttaca aaccttcaac 30

```

<210> 8

<211> 47

<212> DNA

<213> 镰孢霉属

<400> 8

```

ttaattaagg tacctgaatt taaatggtga agagatagat atccaag 47

```

<210> 9

<211> 51

<212> DNA

<213> 镰孢霉属

<400> 9
 tcaccattta aattcaggta ccttaattaa attccttggt ggaagcgtcg a 51

 <210> 10
 <211> 42
 <212> DNA
 <213> 镰孢霉属

 <400> 10
 tggatatgcat aagcttgaat tcaggtaaac aagatataat tt 42

 <210> 11
 <211> 35
 <212> DNA
 <213> 镰孢霉属

 <400> 11
 cagtgaattg gcctcgatgg ccgcggccgc gaatt 35

 <210> 12
 <211> 35
 <212> DNA
 <213> 镰孢霉属

 <400> 12
 aattcgcggc cgcgccatc gaggccaatt cactg 35

 <210> 13
 <211> 34
 <212> DNA
 <213> 镰孢霉属

 <400> 13
 cacgaaggaa agacgatggc tttcacggtg tctg 34

 <210> 14
 <211> 34
 <212> DNA
 <213> 镰孢霉属

 <400> 14
 cagacaccgt gaaagccatc gtctttcctt cgtg 34

 <210> 15
 <211> 46
 <212> DNA
 <213> 镰孢霉属

 <400> 15
 ctatctcttc accatggtac ctttaattaa taccttggtg gaagcg 46

 <210> 16
 <211> 46
 <212> DNA
 <213> 镰孢霉属

 <400> 16
 cgcttccaac aaggtattta attaaggtac catgggtgaag agatag 46

 <210> 17

<211> 30
<212> DNA
<213> 镰孢霉属

<400> 17
attttaaata tgaggagctc ccttgctgctg 30

<210> 18
<211> 29
<212> DNA
<213> 镰孢霉属

<400> 18
ttaattaact agagtcgacc cagccgcgc 29

<210> 19
<211> 45
<212> DNA
<213> 镰孢霉属

<400> 19
ataagaatgc ggccgctagt ttaaacttac aaaccttcaa cagtg 45

<210> 20
<211> 19
<212> DNA
<213> 镰孢霉属

<400> 20
tagcatctat ctccgtctt 19

<210> 21
<211> 19
<212> DNA
<213> 镰孢霉属

<400> 21
gtgtgcagtg acccagaat 19

<210> 22
<211> 42
<212> DNA
<213> 镰孢霉属

<400> 22
gattgggtcc ctacgtagtt aacactatag gccatcgttt ac 42

<210> 23
<211> 28
<212> DNA
<213> 镰孢霉属

<400> 23
attttaaataat gggttcttcg gcattcgc 28

<210> 24
<211> 28
<212> DNA
<213> 镰孢霉属

<400> 24	
ttaattaact attccgacgg aacaaagc	28
<210> 25	
<211> 44	
<212> DNA	
<213> 镰孢霉属	
<400> 25	
ctcttgata tctatctctt caccatgggt tcttcggcat tgc	44
<210> 26	
<211> 43	
<212> DNA	
<213> 镰孢霉属	
<400> 26	
gcgaatgccg aagaaacat ggtgaagagt agatatccaa gag	43
<210> 27	
<211> 30	
<212> DNA	
<213> 镰孢霉属	
<400> 27	
gaatgacttg gttgacggt caccagtcac	30
<210> 28	
<211> 26	
<212> DNA	
<213> 镰孢霉属	
<400> 28	
tctagcccag aatactggat caaatc	26
<210> 29	
<211> 39	
<212> DNA	
<213> 镰孢霉属	
<400> 29	
cttaactttg acttgaaaaa catatctgac atttgctcc	39
<210> 30	
<211> 59	
<212> DNA	
<213> 镰孢霉属	
<400> 30	
ggacggcctt ggctagccct ccgtgcggcc gccggccggt ctgcaggat ctgtttaac	59
<210> 31	
<211> 32	
<212> DNA	
<213> 镰孢霉属	
<400> 31	
atatcgtgaa gatatgcggc attgatgcca cc	32
<210> 32	

<211> 35
<212> DNA
<213> 镰孢霉属

<400> 32
ggcggcaata accggccgaa cattccggat atccc 35

<210> 33
<211> 32
<212> DNA
<213> 镰孢霉属

<400> 33
cactgctatc accaacaatgt ttactcaagt cc 32

<210> 34
<211> 32
<212> DNA
<213> 镰孢霉属

<400> 34
ggacttgagt aaacatgttg gtgatagcag tg 32

<210> 35
<211> 29
<212> DNA
<213> 镰孢霉属

<400> 35
gactcatgag gagctccctt gtgctgttc 29

<210> 36
<211> 34
<212> DNA
<213> 镰孢霉属

<400> 36
tgattaatta acctaaagac atgtcccaat taac 34

<210> 37
<211> 25
<212> DNA
<213> 镰孢霉属

<400> 37
gcatttaaata tactactgtg atgtg 25

<210> 38
<211> 25
<212> DNA
<213> 镰孢霉属

<400> 38
gattgatgtg aaacacatgt tgatg 25

<210> 39
<211> 25
<212> DNA
<213> 镰孢霉属

<400> 39	
cgaccgcggga attagagagg ttagg	25
<210> 40	
<211> 28	
<212> DNA	
<213> 镰孢霉属	
<400> 40	
cgtataaccc atggtggact tgtcggac	28

```

AAATTCGTCTAGCGAGGGACTCCTGGCCCTCGAATTTAGTTAGCGTATCAGTGTAAAG 60
TGCTGGGTTCTCCAGGCGTAAGTAAATGAACAGATGTTAGTCTCCAGATTTCCGCCCG 120
AAGCCGGTTGGCAGACCAACGGGATAAGTTTATGGAAGTTGGTTGGCGGATCAATGT 180
AAAGTTCCTGCTCATTGTACGCGAGATACTCGGCCCAAGACGCATCTTTGCCCCATCGCG 240
CAACTTTTTTGGTCCCCAGGATATCGGAAGAGCATTCCTAAGCCAGCATCTGGTGGGAG 300
ATCGTTCTTTATCTTCGGTTCTTAAAGATGTTACAGAGTAACTCAGCAGCAACTCG 360
ACGCAACTTGGCTACGTTGCCAACGCCCTGCTCTAAGACCCCTTCTTGAGGCCGTACAGAA 420
ACGTTCCGAGGACTGTCTGGTTCAGCGAGATGGATGTTATGCGTTGTTCTCGGGGTC 480
TCCTTTGTCTTTAGAGGTATCGGCAAGAAAGCCATTCCACGTAGTCAGAGCGAGTGCAGAA 540
CTATAACAGGCAACGTCAGAAATTTGTACCATGCAAGATTTTATACCATACCTGGAA 600
GTTCTGGCTATTCAAGCGCTCTACCCCTTCGTATAGCACATAAAGGAAATGTGAATCCATT 660
GCCACTGGGTCTCTCCGTGTGTCTGCGCGGTAAAGCGGTGAGGTTGAAAGGCTAGC 720
AGATTGCACAAAGCTAGTGGGTGTTGTTGAGAAGCACATGTAATGCTCTGACAGGTGTAG 780
CTTGCCAGCATAAATGCCATCCTCGATCGTGATCCCTTATCACCGTGAGTAGCGTTGGAGG 840
CGGGATAGTAAGCTCGGCGTTGATCTCGTAGAGGGCGCTTGAGAGCGGCAAGCGAAA 900
TTGGCTCTGAAATAGCGATGACTTTGAGGGATTCGGATCTGAACCTGGATGATTTAGCCC 960
CTGGGTAGGATCGATAAGCTGCTGGGCTTTTGAACGATGTTGGTGAAGTTCGAAAACAT 1020
GATTTCCGTGAGCGCGCATGACGAGGGGGTTCCGGTCCAGGAGGGAGTCCGGCTGA 1080
GCTTGAAGGAGATGCAAGACACGAAGCAAGACACGAAGAGAGCGCAAGAGTCTGAGTA 1140
TGTGCAACCCAGGCTCGAATAAGTGCAAGGCAGGCAGAGTACGGAATAGACGATAGAATT 1200
GAGTATAGAAAGGCTGAATGGAAGATGGAGACGAGTTATAGGACGGTGGAGATAGAGTGG 1260
AGTTGAAGTTGAACGAAGCTGCGTCAGGTCCAGATACGGGAGACTGGCCATCAACTACTG 1320
GCCAGGTAGCCAGGGCGCGATGGCGGGTGGCAGGTCGCGGGGGGACCTCAGGGCAT 1380
TCCTTTCTCCAAGGGCCGCTGGGGCTATGGACGGGGCTGGCTGAACCTCCAGCCGTCATGG 1440

```

图 1A

GATAGCGGTGCAAGAGATCAGGTACTAAGTCTACCATGATAATTTAGGGGCGAGAGAAA 1500
 ATGATATATTTGTTTAGTAGTAAGCGGGTTTTTACAGTTGAGGAACCAACCTTCTTTCATT 1560
 TATTTATTTCTTTCTCTCTGCAATTGAGTCCCTTTTCTTAAATAGAATATCTACCAATG 1620
 GAACGGCGTGGCTGAAGTGGCTGAAGAATATAGCTCGAGCTGTCAAAACCGCTCATCCTAC 1680
 TACCCTAGGTATAAAGCTGGGAACTAAGACTCATTTCTATCCAACTCATATTTGGGAG 1740
 TTAGTGTAGACCTGTCCGGCCTAGAGAATATGTGTATCTGCATACTTCAAAATACCCCTACG 1800
 TATACCCACTATGTTTAGCACAATCATTTGACCTCTCAAGGCCCTCACCCATCTCAACACCT 1860
 GTCGTGTCTCACTTGACTACTTCTTTGAACCAAGCTCGCCATCGGACTAGTCGAACAAGC 1920
 TTGTCGCCCCCATACAGATGAATGTATGTTTAAAGCTACATGATCAGCCTGAACCGAGCA 1980
 TAACTCGAGTCCGAGACTCCTCTGATGTATATCGAGATGAATGACAAACCTACGGGTCC 2040
 GTTCTTGAGAAGTGGCCTGAGATTTCTCACTTGGTGAGAAAAGGACGGGCGAGCGGGAG 2100
 CCTGAGTCAGAAGAAATACTGTCTCCTTGGATCTCACATGACGGTGTGTGGAGAGATG 2160
 CATCTATTGTCATTGCTGGAGTGACGGCAGAGTAGGGTCTAAAGAAACCCATACTGAGT 2220
 AGAGATGGAGAAGACAAACAAAGCCCCAAGACGACAGAGACGACAGAAAGATTAAAGCTATC 2280
 AGAGCGAGACTATATCACTATTCGAAACCTGCGAGTAATTTAAACAAAGATACACATCAT 2340
 CATTTGTTATCAATTCGACGAAGACATGGTCGAAAATTTCTTGGGTGTATATGCTGTGT 2400
 ATATGGGCCCTGGGCATTGTTATTTTCGCCGCTCTTTATGTGTACTAACACTTCCATTGAT 2460
 ACCCAGAACAAAAGATGAACGCTTAAACAGCACCAAAATCAGGAGAAAGAAATGGCGCTGC 2520
 TCTAGGTATGCTTCTGGGATAAAAAGCGATGTTGTATACCTCTCAGAAAGAAAGTGTG 2580
 AAGTTGAATCAACAAATAGCCGATGGAGCGATCTGAAGGGGTGGCAGACCTGCTACGCG 2640
 CATTTAGGCAAGGCATCAACTCGGCAGATGATTAAAGAAAGTTTGTAGGTTACCGTGT 2700
 GTGTTGTGTTCCATTATAAGTTTATAACCTTGTCTAAGATGCAACGACTCTGACCTCAGGG 2760
 TGTTAGAAAAAATTGACCACCTAGGAGCATAAAGTGACGAAATTCGGGGATCAAGACAAATAGA 2820
 TAGTTTCATTTTCATGTGCTCCTACGTCCTTTTTCACGTAATGTTTCTTATAAAAAAAGA 2880

图 1B

```

TAGCATTGTCCTCTTTGGTGAAGAGAAAAAGATGTTACGACGTGGCCTTGATTCGAA 2940
CAGACGCTCCGAAGAGATAGATTTCTAGTCTATCGCGTTAGACCACTCCGCCACCACG 3000
CCTTACGTAATCTGTGATTGTTGAAAGTTACTCTCGTGTACGGTCTATACGTGAAGAAT 3060
CTACACTTGACGAGTCTCGAGGTCTGGGGTCAGTTAGACGGAAATGGGAGAACAAAGAGA 3120
CTTGGTGACATTGCAGGCAACCGGGTAGATGTTGAGGTCAATGATCGGACAAGATTGTTG 3180
CTTCAAAAGTAACAGGTAATCTTTTTTTTAAATCAACAGAAAAAGTTCATGTTTCTT 3240
AATCCAATCTATTTGTGATAGCGTTTGATGACAAACAATAATAATGATGGTCTGGCGGCT 3300
AGTGATCGTTTGTAAATGACGTCGTCAATATATCTATCCTATACATACAGTTGCTTTGCACACG 3360
CACTACGTCCTTCATTGCTTCTTCACTATTGATGGTGATTTGGTTCAACAACCTAC 3420
AGAAATAATGACCTGTGGTGTCTCCGAATATGGCTAGACCAACAAGCTTGTAACCGG 3480
GCATTCAAATCACCATGTGATGCCCATCATCAGATCATCCACCAACCCAAAACAGACCA 3540
ACTACTCAGAAAAGGCATCTCATCAAGAAAAACGGCCAACTAACGTCCAAAAGGCCCG 3600
AAAAACGTCCATCAGCCGCGAGAGACTTCAATAGACTGCACAAGAGACCGATGAG 3660
ATCGACAGACTAAACCCGGGAGAGTGTCAAATATGCGGGGGAATGGGGAACCTACCCCA 3720
GAAAAGAGAAAGAGGATAAATTCATGCTGGGGTTGACGTCTCTATTGGTTAGACACGA 3780
ACGCCTGCTCTCGGCGTAATTTATACCATAGCGCCAATGAGGGCGGAACTCCTGTTT 3840
TCAAGTCGTCATTGTTGGTGGTCAATATAGCCAGTAGGTATCCGTCCTTGGTGAT 3900
TGACCAGACATATCGCTCATCACAGATCAACATCACTGCTATCACCACATGCTTACTCA 3960
M L T Q
AGTCCTTTATGGCTTGGTAGCCAGTGCCCTTTGGCAAGGCCAAGTCGTTGCATCACCAG 4020
V L Y G L V A S A L W Q G Q V A S P S
CAAGGACAATTCACTGGAGCGCTTCATTGACAAAACAAGCTGATATTTCTATCAAGGGTGT 4080
K D N S L E R F I D K Q A D I S I K G V
CCTTGCTAATAATTGGCGCTGATGGAAGGACAGGGTGACAGCGCCTGGTGTGTTGT 4140
L A N I G A D G K R A Q G A A P G A V V

```

图 1C

```

GGCAAGTCCATCGAAAGAAGATCCTGATTGtaagccagcaccctaccttgctccttggtccg 4200
A S P S K E D P D
catgctaataatgatggtctcagATTGGTACACTTGGACTCGTGAAGTCTGCTTTAACGTACAA 4260
Y W Y T W T R D S A L T Y K
AGTGCTCGTTGAGAGATTTCATCCACGGCGACAAATCTCTCCAACGAAAGATAGTAATA 4320
V L V E R F I H G D K S L Q R K I D E Y
TGTCCTCCGACAAAGCGAAACTGCAAGGGACCAAAATCCATCGGGCAGCCAGAGTCGGG 4380
V S A Q A K L Q G T T N P S G S P E S G
CGGTCTCGGCGAGCCAAAGTTCCATGTGAATCTCACTGCTTTCACTGGATCTTGGGTCG 4440
G L G E P K F H V N L T A F T G S W G R
GCCTCAGCGCGACGGCCCTCCGGCTTCGGGCTACCGCCTTGACTCTGTATGCAGAAATGGCT 4500
P Q R D G P P L R A T A L T L Y A E W L
CATTTCACCGGAAAGATCCAAGGCTTTGAACAAAGTCTGGCCAGTCATCGAGAAGGA 4560
I S H G E R S K A L N K V W P V I E K D
CCTTGCGTATACTACCAAGTTCTGGAATCGCACTGGCTATGATCTATGGAGGAGGTTAA 4620
L A Y T T K F W N R T G Y D L W E E V N
TGGATCTTCTTTTACACTTTCGGCTTCGCATCGTGCTCTTGTGCGAAGGTGCCGCTCT 4680
G S S F F T L S A S H R A L V E G A A L
GGCTAAGAAACTTGGCAAATCTTGTCCTGACTGTGTACCAACGCTCCTCGCGTCTGTG 4740
A K K L G K S C P D C V T N A P R V L C
CTTCCTTCAGACTTTCTGGACTGGTGGCTACGTTGACTCCAACATTAAAGTCAAGGATGG 4800
F L Q T F W T G G Y V D S N I N V K D G
TCGCAAGGGTCTCGATGTCAACTCCATCCTCTCGTCCATTTCATACATTCGATCCCAACTC 4860
R K G L D V N S I L S S I H T F D P N S

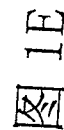
```

图 1D

```

CAAGTGCACCGACTCGACGTTCCAGCCTTGTTCACCCAGAGCTCTTGGCAACCAAGGC 4920
K C T D S T F Q P C S P R A L A N H K A
GGTCGTCGATTCTTTCAGGTCAATCTATGGTGTCAACAAGAATAGAGGTCAAGGCAAGGC 4980
V V D S F R S I Y G V N K N R G Q G K A
CGCGGCTGTTGGTCGATATAGCGAGGACGTGTACTATGATGGCAACCCCTTGGTACCTGGC 5040
A A V G R Y S E D V Y Y D G N P W Y L A
CACTCTTGTGCTGCAGAACAACTCTACGCTGCGGTCTACCAAGTGGGATAAGCTTGGCGC 5100
T L A A A E Q L Y A A V Y Q W D K L G A
TGTTACTGTTGACGATGTATCTTTGTCTTTCTTCAAGGATATCGTTCCCAAGGTCTCCAA 5160
V T V D D V S L S F F K D I V P K V S K
AGGCACCTTATGCCAAGAAGACCAAGACATACAAGGAGATCATCAAAGCAGCCAAAGACTTA 5220
G T Y A K K T K T Y K E I I K A A K T Y
CGCCGACGGCTTTGTGCTGTGTCGACACATACACTCCCAAGGACGGCTCACTAGCTGA 5280
A D G F V A V V Q T Y T P K D G S L A E
GCAATTTGACAAGTCAACTGGAGCCCCCAAGTCGGCTGTTACCTCACCTGGTCCTACGC 5340
Q F D K S T G A P K S A V H L T W S Y A
CGCCTTTGTGCGCCACAACCTGAACGTCGCGACGGCATCATCTCTCCCTCCTGGGCGAAAG 5400
A F V A T T E R R D G I I S P S W G E S
CAGCGCCAACAAGGTCCCGCGTGTGTCAAGCTGCCCGCCAGCATGTGACACACCAATCAC 5460
S A N K V P A V C Q A A P A C D T T I T
CTTCAGTGTCAAGAACGTGCAAGTTTCATCCGACCAAAAGTTTACGTGGTTGGCTCAGT 5520
F S V K N V Q V S S D Q K V Y V G S V
GACTGAGCTTTCTAACTGGTCACCTGATGATGGCATTGGCTTACGCCATCTAGTTCGG 5580
T E L S N W S P D D G I A L T P S S G

```



1E

AGTGTGGAGCGTCAAGGTTAAGATTCCCTTCTGATACAAGCTTTGAGTACAAGTATATCAA 5640
V W S V K V K I P S D T S F E Y K Y I K
GAAGACTAGCAGTGGGGATGTTACGTGGTTGAGTGATCCCAACAACCGGGCTATTACGGG 5700
K T S S G D V T W L S D P N N R A I T G
TAGCAAGTGTGGAAGTACAAGTACTCTTGATGATGAGTGGAGGTAGTGGATGACAGATTT 5760
S K C G S T S T L D D E W R .
ATCAAGCTATGTAGTTTGTGAATATATAATATCCAAATTATCAGGGTTCGGTAAGAAT 5820
ATAATTCAGTTCAGCAGTCTGTACAAGCAAGCCATGATTCACGCTTCCTTCGTTTGGAAG 5880
ATAAGGGTTCCTCGCCACCGTCAAGATAATTTCTCGTGCTAATATCACGTAATCCATCT 5940
GCATTAAACCCCTTGCCGGGAAGGTTTCTTCAACCCAGCAACCCAGAGTAACCTCGGAGAT 6000
AGGGAAGTCTATTTGCGCTTATCCTCCGTGACGAATCCCTGAACCCAATT 6050

图 1F


```

AATTACTGTGTGATCACACCTAACTAAATACCTAACTACCCGATGGATCGACA 60
AGGAAATCTCAGCCCTTGTGAGTCTCCTCTTTTCGTCTGTCTCCTGGGCTCGTACTGT 120
CCGATTGTAACTCTCGTCTCTCAACTTGTTCAACTCTAATAAGTGGTGGCACAAACGTGAA 180
GATGATTTGTTGTGAGGGGGGGTTCGTGGCATACCAAAGAGACCAAAAAGTCCCC 240
CTATGTCGATTTGATGGTGTGCGTTGCCATGATACGGGACCCCGAATATGTTGTATGCA 300
TCATGCGTACAGAAAGCTACTGTTCAAACGAAACGCAAGCGGATTGATCAACCCGTGA 360
AAGACCATGGGTCTCTCTCAGTCCACAATCTTCTCTCTCTGATCAAAATTTATGGATCCAA 420
GGGCCACAATTTAGGCCATCATGGGTCCCTTTCCTCTTTTCGCTCACCCCATGTTCC 480
CTGTCCCACCTCATTCAGTGGACCTGATGGATCCCTATCCCCCGATGAGCCGGGGTGC 540
AGCCTTGGCGCTCTCTTCTTGTAGTGTGACCTACTGTTGATTTCACTCAGCAGTCCCTAG 600
AGTCCATTTAGTTGGGCCCTGGGGTGTGGGTCTGAGACTTTGCTTCTTGCCTGGTCTTG 660
TCTAGCTCGAATCTGTGGTTGCCCTGGCCCTGGCTGACCTGACCTGAGGGGGTG 720
CCCCTTTGTCTGTCTGCATATGTTGCTATTAGCTACCTACTCGAGAATTCATAAAGG 780
ACTGTCCAGCCCCGTCTCTACTGACTTCTTTCTTTCCCTCTTCAACCCTGTTGTCATA 840
TCAAATCTGTCACTCGTTAGACCAGACTACCATTTCCCACTTTTCGCTTTTAAACTACTTA 900
CTCAACTAATTCTAATACCAACTCCAAAACCATCAACATGCGTTTCACATCAATCCTCG 960
                                     M R F T S I L
CTGCCGGCGCTTTCGCCACCATGGCCGCTGCCCAGAGCAAGACCGTCTCCCTCGACCCCTG 1020
A A G A F A T M A A A Q S K T V S L D P
CTCAGCAGTCTCAGGCCGACTGCCCTCTCCGACTGTGAGCCTGGCGATGTCAAGTGCCAGT 1080
A Q Q S Q A D C L S D C E P G D V K C Q
CTTACTGCATCACTgtatgttacaacaacgattccccctgtcatgtgtagaaaactaaca 1140
S Y C I T
tcccaatagGTTCCCTCTCCTGACGAGAAGAACATCGAGGAAACCAACCAAGTGTGTTTG 1200
V P S P D E K N I E E T T K C C L

```

图 2A

```

CCGCTGCCCCAAGGGCTCCGAAGCCGACACTGAGAAGTACACCGTTTGCATG 1260
P P A P R A R A P K A D T E K Y T V C M
AAGAGTGTATCGCCGACAACTACTGGAAGTCCGTTGATGGTACCCCGTGGCACCGAC 1320
N E C I A D N Y W K S V D G T P R G T D
GTCCCCGATGTCAAGAGCAAGCCCTCCGAGGCTGCCCTCCTCCGCTGCTGAGAGGCCACC 1380
V P D V K S K A S E A A S S A A E K A T
GCCACCGGTACTGCTGCTGAGTCTGATGCTACCGCCACTGGTGCCCTCCGCTACTGAGTCC 1440
A T G T A A E S D A T A T G A S A T E S
GAGTCCGGCTCCGACTCCAGCTCCGAGGAGACCGGCTCTGCCCTCTGGCACTGCCACTGGT 1500
E S G S D S S S E E T G S A S G T A T G
ACCGCTGCTGAGGTCTCCGAGACTGGTAACGCCGCCCTCTCCCTCGTTGGTGTCTCC 1560
T A A E V S E T G N A A S S L V G G V S
TTCCTCGGTCTCGTTGCCGCTATCTTCGCTCTGTAAATTGGGTTTCCTGCTTTAGGATAA 1620
F L G L V A A I F A L .
TCTGATTTGGCATGACGGAGAAGGATTTAATGGGTTTTATTACAGCGGTAATGATTGGAG 1680
TTTGGATTTCAAGATGTGACACCGTTGGACAGCATGATAAGCCCTACGGGTCTGATCAATT 1740
TCATGGACAAATTTGTTTTTTGGTAATCAATTCGCGTTCACATATGGCTCGGCATAT 1800
GAGCATGAATACAATACCTCTTTTGGCCCTCAATTCATCCAAATTTCTTGTGATCTCA 1860
CAGTGATTCAACTTACAAGTTGGGGCGGACCACTGAGGTCGTGCTGATGTGGGTCTTC 1920
TGTTTGTGATTGGCTCATGATTCCCAATCGGGTGCTTCAAACGTTAGTTTGTAAACAAGC 1980
GAAATGAGGGTCTTAGGATGCAATGTTCAAAGCGCAAAACCCAAATTGAATTCAAATGTTAA 2040
AGAAATCATCGAGAAAGCGAGTTACTGAGGTGAATTTGTGCTTTCAACTGTCAATACCTC 2100
CCTCAGAACAAATGAATTGAATTATTATTACACATCAATGCCCAATATTCTAAACATGTT 2160
CGTTGTAAACAGAGTTTAAATTCCTTGAACGCCACAATGTTCTTGGTAATTATCGCGCCTG 2220
TCACATGAACCTGGCTCCTGAACTTAAACGTTGGTGACCCAGCAACTCGTTTATCAGGCTT 2280

```

图 2B

AGGGTAGCTGTCAACAACAACAACCTTGTAACAATTGATGTTATTGATGAATCATGTAT 2340
AGAAAGAGCACAAATTGATTTAAACACAGATAAACTGGTCGAACCGATTTTATCAGGTTGTG 2400
TGAACATGCATTGCCGAATCAGAAACCCAGAGTAAGACTATCTACAGTTCATGAAGACAA 2460
TTCACAGACTGCCAGAAAAGCAAGGTA CTGGAGCACGAGAGACAAAATTATTGAATT 2517

图 2C

```

AATTAGAGAGGTTAGGGATTTCATATGGCCCAACCAATGGGAAGGAGGCAACCATCTGCACG 60
AGCCCAACCAAGTCATCTCCTCAAACTGTGCTGCGACTAAGAAATTGATTCGGTTCTGGC 120
CTGGCCTTTGTATCAGCTAGGTCAATCTCGACTACCGGAGGCCAGGCTGAAGCAGTCAGT 180
CACGCATTGTCACTTTATCGGTCCCTGTCCTCATACGGATACACTAGGCGTCAATGGGCTT 240
CAACGGAGATCCAGAGATCTCATGAAGAGCATCGACGATAAAGTGAGTGGTTGGGGATA 300
CTGTGCGGTGCCGACCCAGCGGCAGCCAGGTTCCACCCCTTGATTACATGGTTGAAAAGT 360
GGCGTTACTGGGCGAGATCAAAATTTGGCATGTATGTTCTGTCCTCAATGACGCGAGCTCTCCA 420
TGTGCTGCGAGGGTCAGGAACGGACCGCATGGGATCAGTGAGGTGAAATCCAAACCGAG 480
GGAGAGCGAGATCTTTGTGCTCATATCCATGCTGCCATGCTACGTGCCGAAACAGGCCAGA 540
TGGCGTTCAACTCAGTCGACCCAGGTCGGATGAACCGGGAGCGGTGACGAGATCGAAGCTT 600
CATCTATCGCTTACGGGTTATGTTCCACTTTCCATTAAAGTTTGCAGTTGCTGTTTGA 660
GAGCCATGTCGAAAGCATGGACCGTGTCAATCTTTCAAGGTAAATCTGGAGGTGGGAAG 720
AAGAAATGCGAGGAACAGGATGGGAGGATAGCAGGCTGACGCGGAAAGCTAGGTAGCTAC 780
CTGGCTGATTACTGGCTAAAGCTGGAGAGCAACTAGGTAATATCAGGCAAGAGCTCCAA 840
GAGCTATTGGGAGGCTGGCTGATTGTCTCTGGCTGAGACGCGAGGGAAGGTTAAATG 900
GCCGCAGGCCAAGAAAGGGCTGCAAAACACGGAGTGGATGGTGGGGCTCCACATACG 960
GGATTGGGGCTGCGGATCTAACCTCAATTTGGCAAGAGGTAAATAGGACGACATGCAGGC 1020
CCCCTGACATGTAAACAAGACAAGTGGTAAACAAGCCATCAACATCAACAAGAGCGAACA 1080
ATCGACACACCCCATGGGGTGAGATATGGTAGTAAGGCAGAGAGAGATCAGGGCAGCATA 1140
CGTGGGAAAGGGCTGGGCAAGAAAGGACACACACGATCAACAGAACGCGCGCTACCGAG 1200
GGAGCAACACAAGTACAGTAACCGCTCACAGAGGCAACAACCTCGTCCAAATCCTGCCCCGT 1260
CTTCAAAAGCCAGTTTCGTTCTGAGTCTGTCCGGTCCCTCTTCTCCTCCTACTCCCTC 1320
CAATTATCGCCATCCACATCGACATCGTCAATTCAACAACCTCACCCAGACAAGAAAGAAA 1380
GAACGACTGAAGGCCCTTCGCTCGCCATCACCCGATTCTTTTCCATTCTCTTCGACTTTTG 1440
TTTCGTAGGAACAAGAGCCAGAGAACTTCTTGTCTATCCTTTTCGAATTTTCGGAAGGTTGTA 1500

```

图 3A

```

TGAGAAAGCTTCTCTCGCGCCAGCAAAAAGTCGCAAAATCTGGAACTTTGAGGCACGCGTCCCC 1560
GTTCCCTTCAGCATCTTCCCATCGACATATCGGGAATCCGAATCCACACACAGACCGTT 1620
ACCGAAACAAAGATACACGAAGAGGTTGAGATCAAAACCCCAACAGCCCGAAGCCGACGG 1680
GAAGGTGAAATATCTTCTGTCTCGTCACCGCGGAACAGGTCCCTCCTCCTCGTCAAGAG 1740
CAAGAGTTTATCGAAGAGAGGTCCATATTACGCGTGAAGAAGAACATTACCAACCGTCCC 1800
GGTGTCAAAAATTTCGAGCACGAAGACTTTACTATCCGTGAAGACTCCCGACGgtacgtt 1860
cgattttacatttcctttcatctccattttaggtcgcatctttttcgttacttttttgggtc 1920
aattacacgggggatacgtattttccacaggtcggagaaaagccctgcttgcctctctatgcc 1980
taggtctgtattctctcatccctctgctgatctggccatggagacgtgtgagaacaag 2040
actacaattcatcacatcatttttcgctaggggaaagcaattaccgttgtcccgacctt 2100
ctcccaaccatcagttttcactttcccttttcttggtctggcttgccctgaccattaccc 2160
accgcgcaggagcgttcagtcgccagccatcccatctcacatcaccttctcatatcct 2220
ctcttcacacgcctcacacaccccccctgcatgctaccatgccaaacccacttcagctt 2280
ggctggatacccaatttgctttgcttcctcccggtcactagcgcctctaagcctgctg 2340
gcctgagcaaggggtggagctatctcagggggcgcgcctcccggttgccatatgatacg 2400
caaacgacttactatagacatccatcagctaaccacagacaaaatctagACCTCAACCTCCC 2460
TCTCAATACCAACCTTCCCAGTACCAACCAACCTTCCCACCTACCAACCAACCTCCCAAAATC 2520
CAAACTTCTCACACTCACGTAGAGATCGACACCCACCGTCACTCCCTACTACTCCACCCCC 2580
ATTGATCTCGCTGAACGTGAATACCGCCAGCGTTACCGCCCTGCCCCAAGCTTTTCCACA 2640
GAAGACCCCTTCTTCCCACCTCTCATCCCTCACTACCAACCTCAAGACAACTTCAAAGCCAAC 2700
AACTACACCGTTGAAGCCGACCCGCTCCCCAATTCATTCCTCTGAGAAGACTGAAATC 2760
AACAAAGTTTACTGTTGACGAACAACCTCCTCTCGCCCTCAGTACAACCCACCCGAGAAAGACC 2820
GAATTCAACAACACTACACTGTTGACAGCCGATCTTCCCGTCCCTCAATACAAACCTGTGAG 2880
AAGACTGAGATCAACAATTTCACTGTTGACGCCCGCTCTTCCCAGCCACCGTACCGCGAC 2940
ACCAAGACAACTCAAGTCAACAGCTACGCCGTTGACAGCCCGTTTCTCGTCCATCTTAC 3000

```

图 3B

```

AAGAAGGACGTGAGATTTACTGAACAAACCGTCGAAGCTTCAAAGTCCGACAAAGTCCAAG 3060
ATGGGTTACTACGACGACGAGGgtaagtgaatctgtcaccagcgagcgccatcaagct 3120
M G Y Y D D E
ctctattcgtgacgcaattcaagctaaccacccagtcaccagGTTCTTTCCGCAACGGCGGCA 3180
G S F R N G G
TCCACAAGCTCGGTGACAAGTCCCGCGACATTGAGGTTGACATTGCGGAGACTTCTCGTC 3240
I H K L G D K S R D I E V D I R E T S R
CTGCCAATGACTGCGCTCCCAACACCGTCAGCATCCCTGCGCACCATCCGTCTGGGTG 3300
P A N D C A P N T V S I P C H I R L G
ATTTCCTCATGCTCCAGGGCCGCCCTGCCAGGTCAATCCGCATCTCCACCTCCTCTGCCA 3360
D F L M L Q G R P C Q V I R I S T S S A
CTGGCCAGTACCGCTACCTTGGTGTGCGACCTCTTACCAAGCAGCTTCATGAGGAGTCTT 3420
T G Q Y R Y L G V D L F T K Q L H E E S
CTTTCATCTCCAACCCCTGCCCCAGCGTTGCTGTTCAGTCCATGCTCGGCCCTGTCTTCA 3480
S F I S N P A P S V V V Q S M L G P V F
AGCAGTACCGTGTCCTCGATATGCAGGAGGTCAGATCGTTGCCATGACCGAGACTGGCG 3540
K Q Y R V L D M Q E G Q I V A M T E T G
ACGTCAAGCAGGGTCTCCCTGTGTCATTGACCAAGTCCAACCTCTACTCTCGCCTCCACAACG 3600
D V K Q G L P V I D Q S N L Y S R L H N
CTTTCGAGTCCGGTCGTGGCTCTGTTTCGCGTCCTCGTCCCTCAACGACGGTGGCCGTGAGC 3660
A F E S G R G S V R V L V L N D G G R E
TTGCCGTTGACATGAAGGTCAATCCACGGCTCTCGCCTGTAAAGCGTGTCAACTGTTTCT 3720
L A V D M K V I H G S R L .
GAATTCGGGCGAGCCGCTTGCAATGCGACTTCTTCCCAATGTTTAATTGAGTGAAGGACA 3780
GCACTACAGTCTCACCTCAACTGTGGGGAGCGGGTCTGGGCTGTCTCTAATCTTACCTG 3840

```

图 3C

TACAATGTC AAGTTTC ATAGGGACCTGTTGTGTCAAGATGGTTCGAGTTTGTGTGT 3900
CAAGATTGGATAAATGATATTGGCTAGCTGGAAATACTGGAGTCTTTTGTGTAGATGGGA 3960
GAGTTCGTACATGAAC TATAGTAATTGACAATTGATTC CGCATCTACTTAGCTTTTCAT 4020
TGGTGCTCTATGCCCCAACATGTGAATT 4047

图 3D

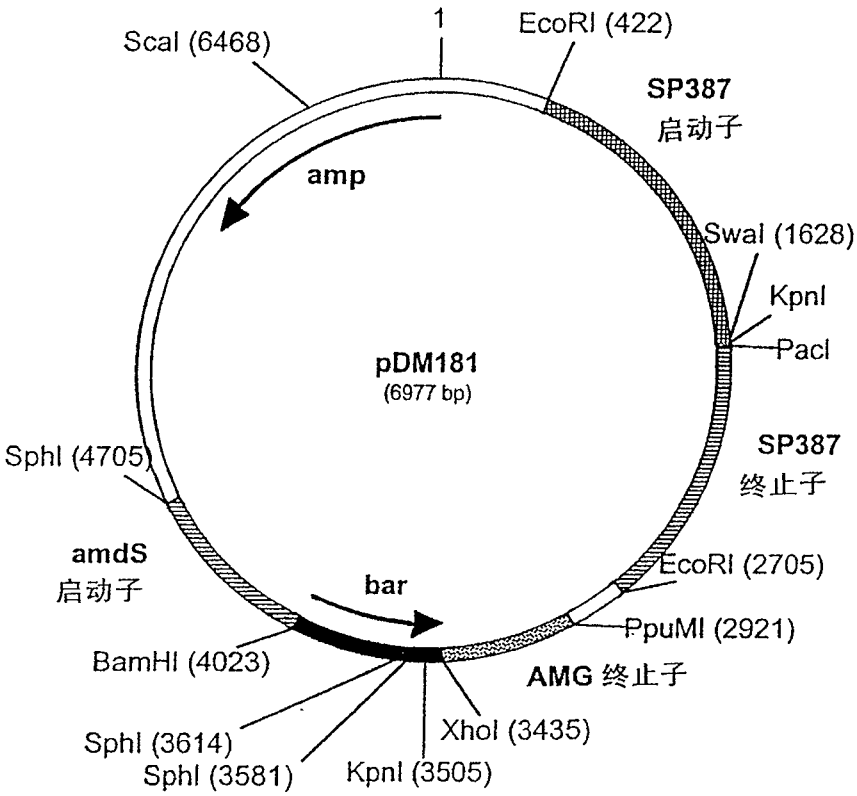


图 4

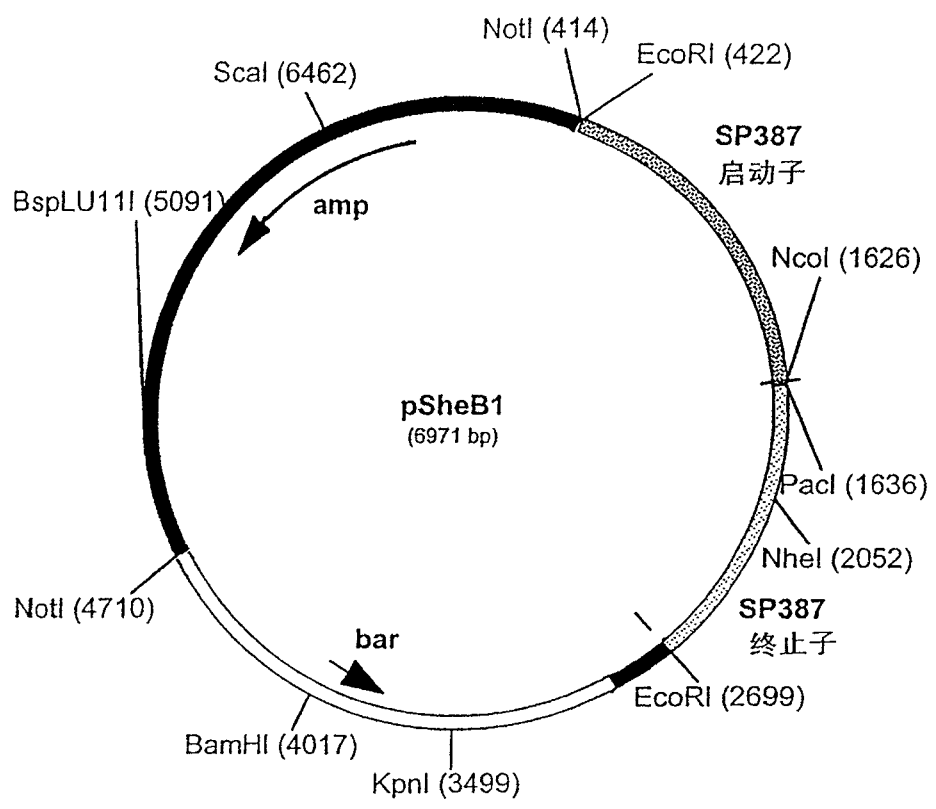


图 5

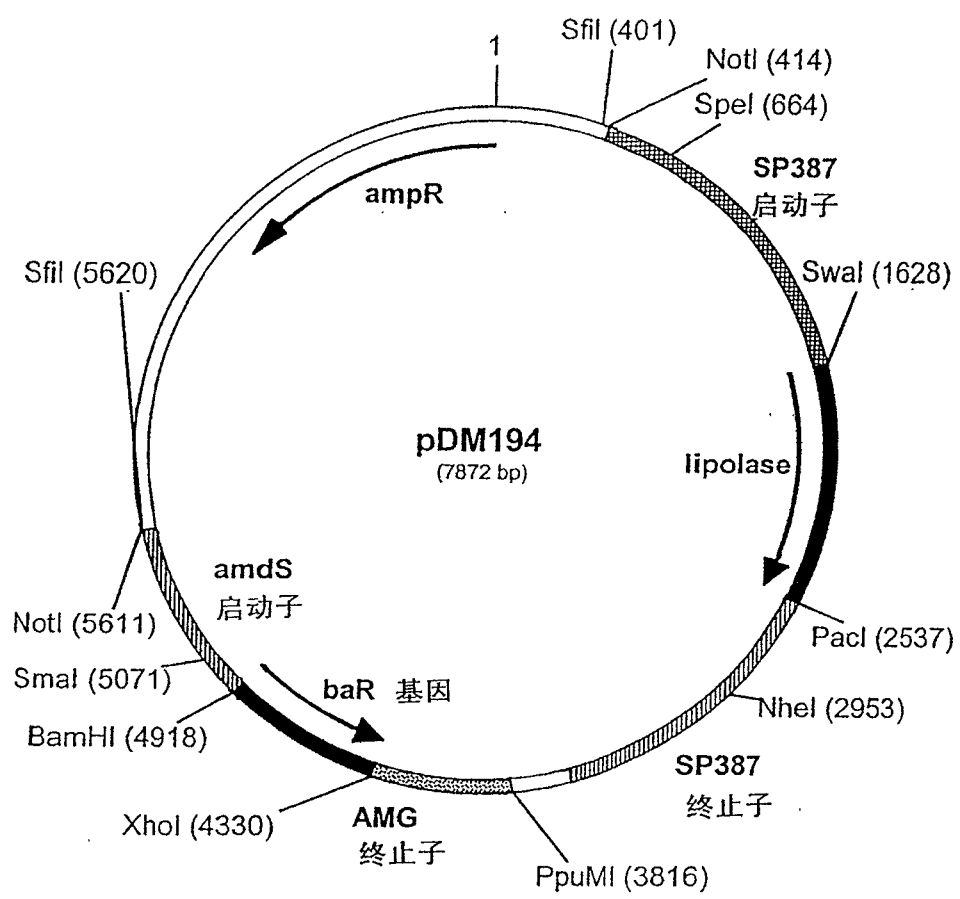


图 6

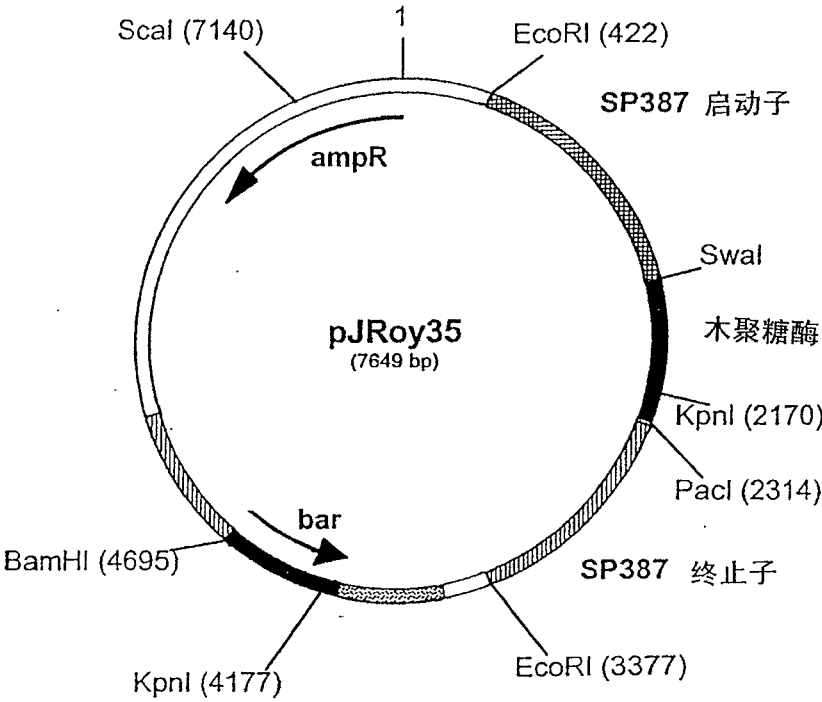


图 7

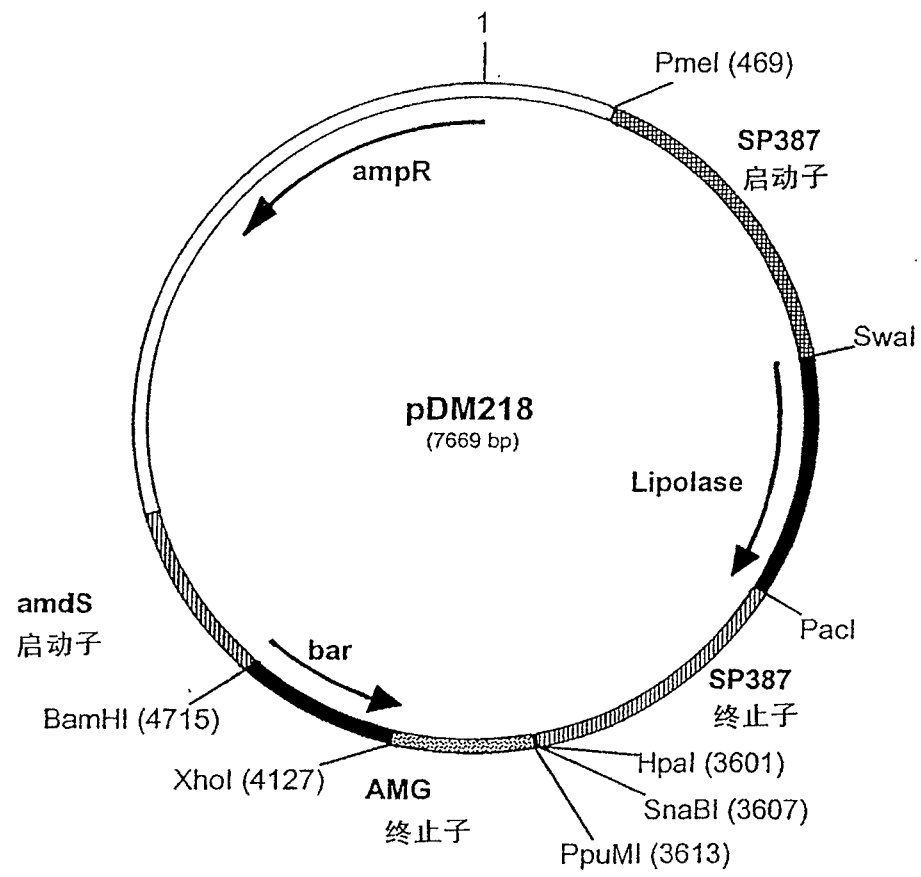


图 8

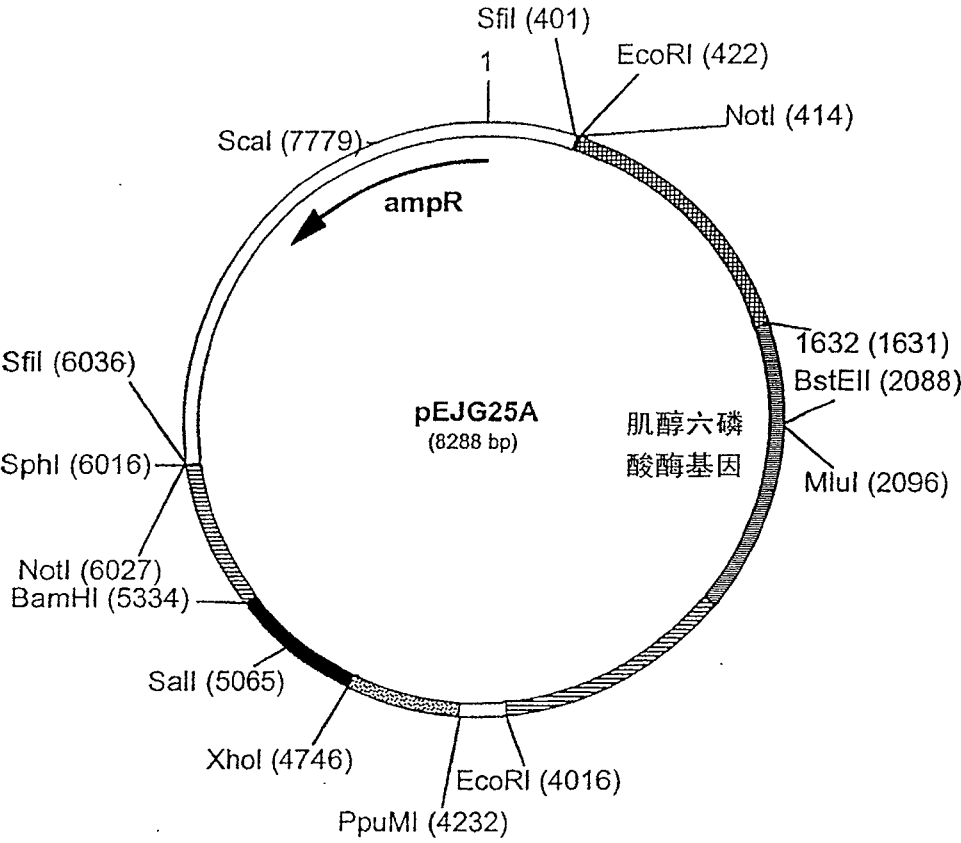


图 9

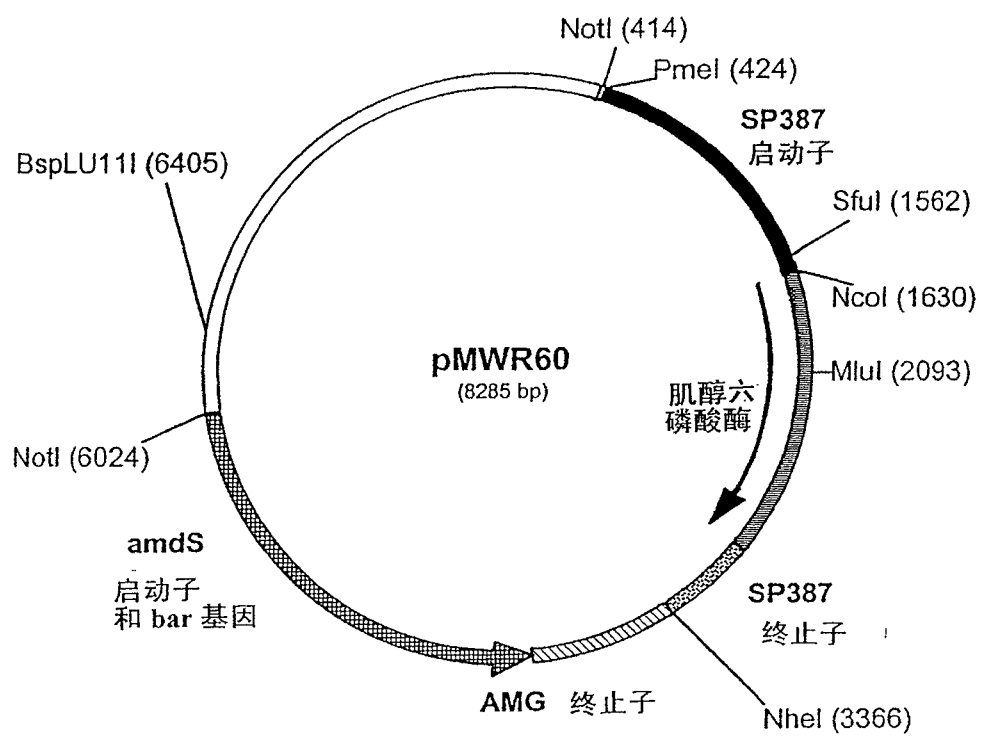


图 10

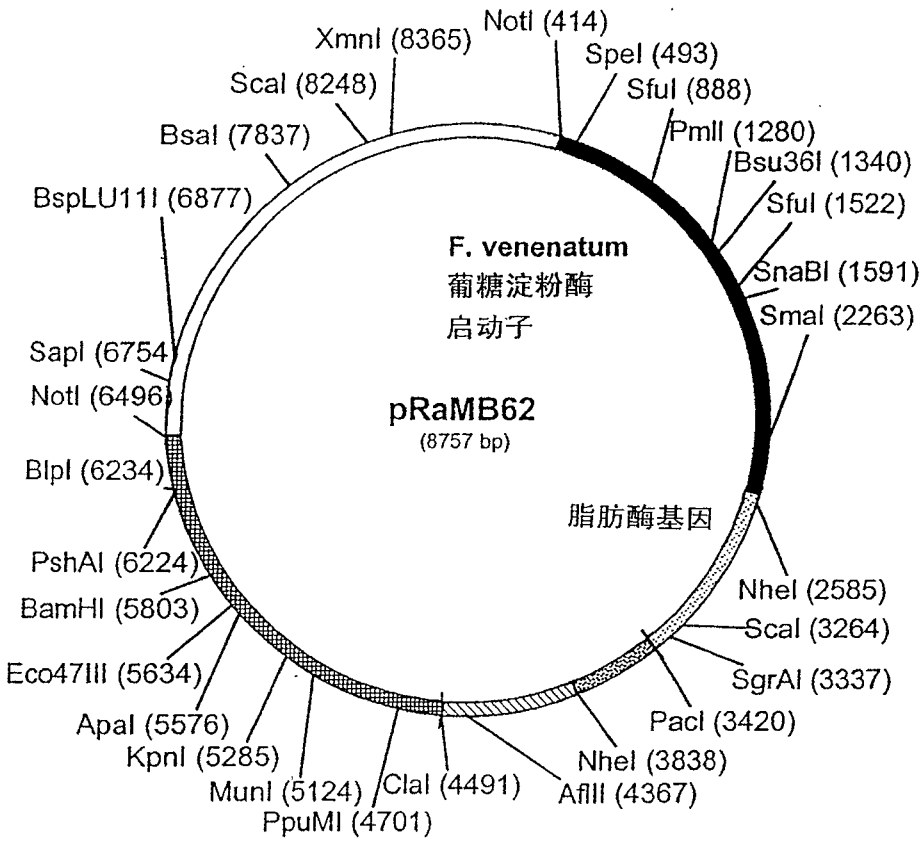


图 11

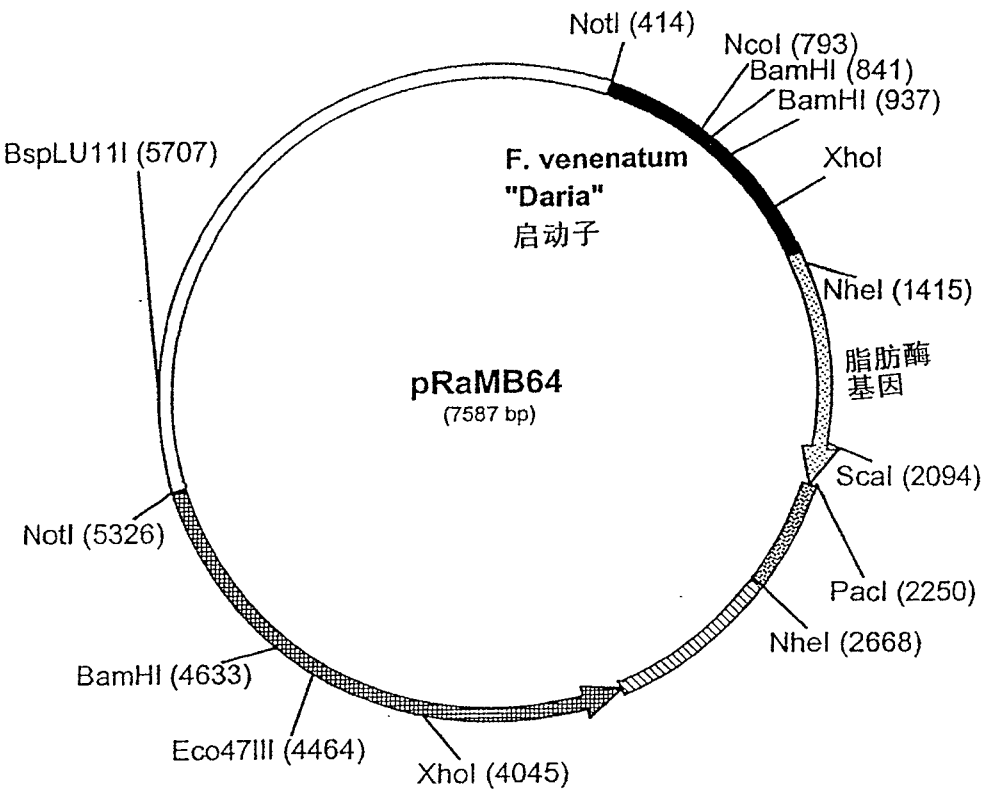


图 12

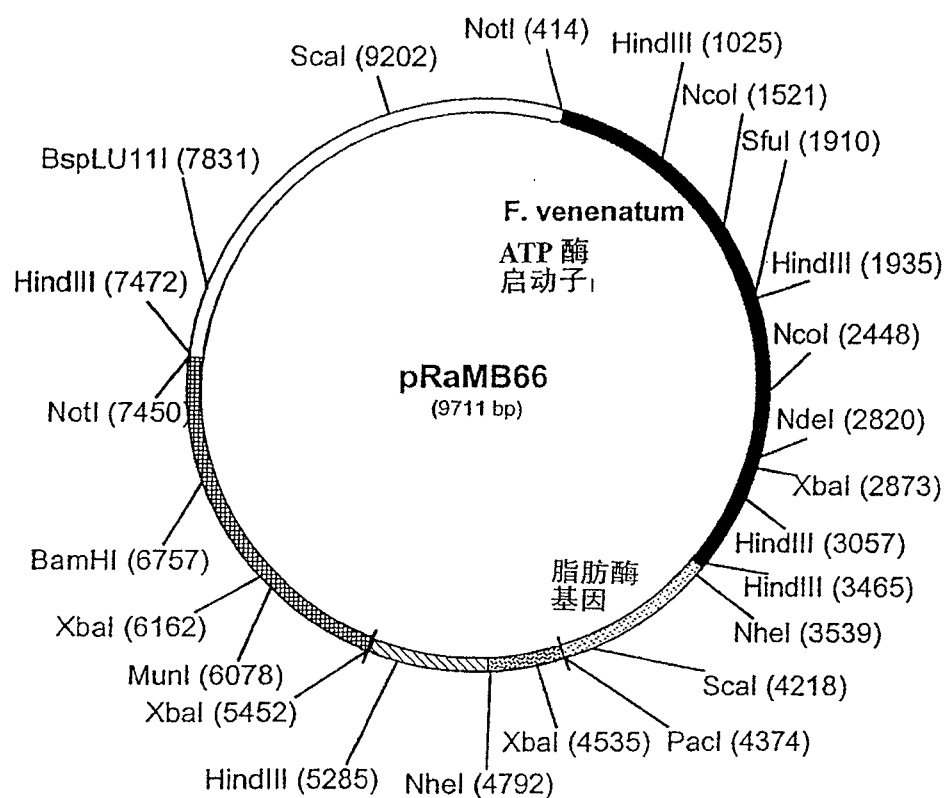


图 13

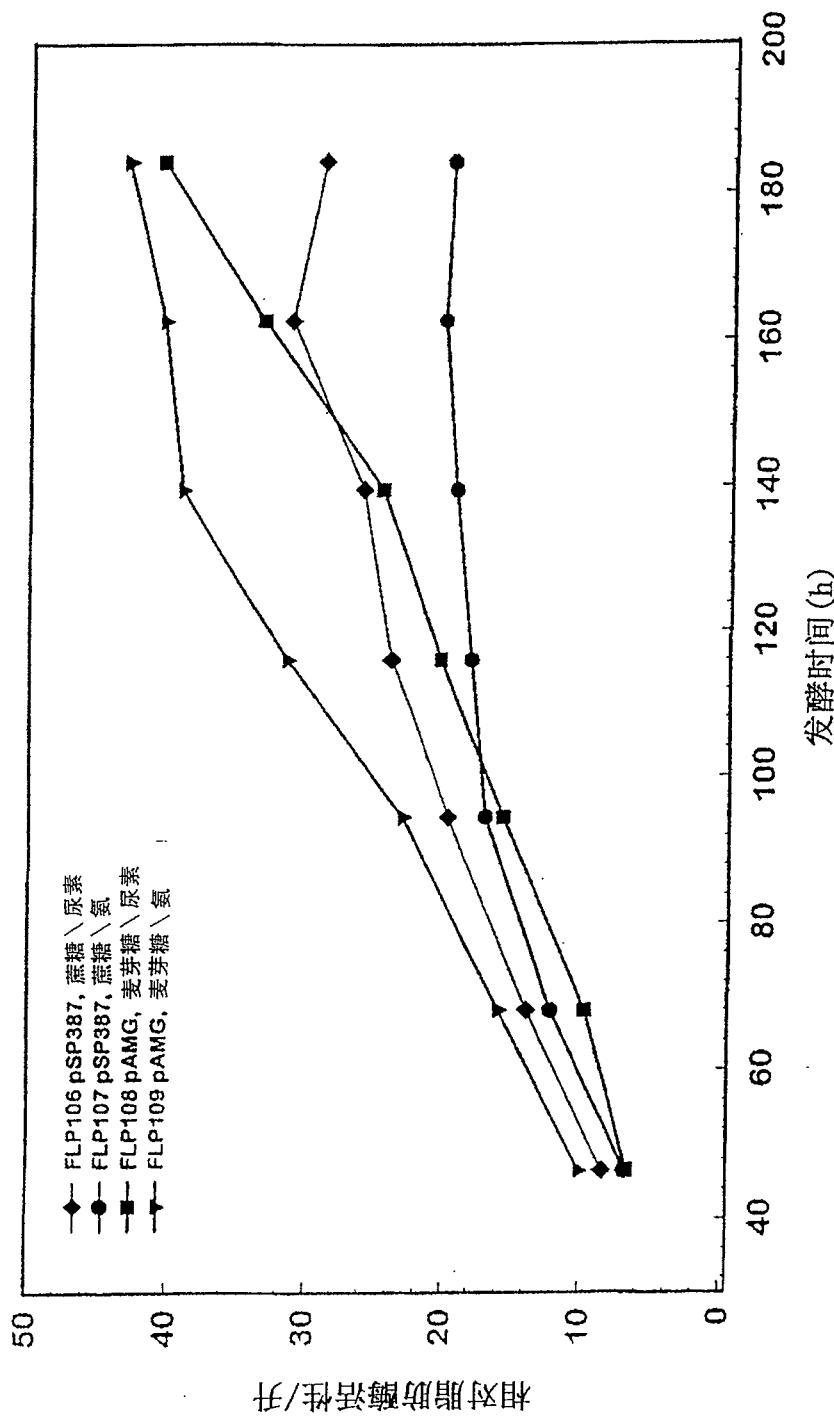


图 14