

[19] 中华人民共和国国家知识产权局

[51] Int. Cl⁷
C12N 9/10
C12N 15/52



[12] 发 明 专 利 说 明 书

[21] ZL 专利号 97193669.2

[43] 授权公告日 2003 年 6 月 4 日

[11] 授权公告号 CN 1110554C

[22] 申请日 1997.6.9 [21] 申请号 97193669.2
[30] 优先权
[32] 1996. 6. 10 [33] US [31] 60/019,399
[86] 国际申请 PCT/US97/09928 1997.6.9
[87] 国际公布 WO97/47736 英 1997.12.18
[85] 进入国家阶段日期 1998.10.7
[71] 专利权人 诺沃奇梅兹生物技术有限公司
地址 美国加利福尼亚州
[72] 发明人 S·L·爱尔罗德 J·R·查瑞
[56] 参考文献
US4902620A 1990.02.20 C12N15/00
审查员 沈志红

[74] 专利代理机构 中国国际贸易促进委员会专利
商标事务所
代理人 黄革生

权利要求书 3 页 说明书 35 页 附图 21 页

[54] 发明名称 米曲霉 5-氨基乙酰丙酸合成酶及
编码此酶的核酸

[57] 摘要

本发明涉及米曲霉 5-氨基乙酰丙酸合成酶及
含有编码此酶的核酸序列的分离核酸片段。 本发明
又涉及含有该核酸序列的核酸构建体、载体和重组
型宿主细胞。 本发明还涉及生产 5-氨基乙酰丙酸
合成酶的方法。

I S S N 1 0 0 8 - 4 2 7 4

1. 一种基本纯的 5-氨基乙酰丙酸合成酶，其特征在于：
 - (a) 得自一种米曲霉菌株；或
 - (b) 由在高严谨性条件下与 SEQ ID No:1 所示核酸序列可发生杂交的核酸序列所编码。
2. 按照权利要求 1 的 5-氨基乙酰丙酸合成酶，其得自米曲霉 IFO4177。
3. 按照权利要求 1 所述的 5-氨基乙酰丙酸合成酶，其具有 SEQ ID No:2 所示氨基酸序列。
4. 按照权利要求 1 所述的 5-氨基乙酰丙酸合成酶，其具有与 SEQ ID No:2 所示氨基酸序列之间不超过 3 个氨基酸不同的氨基酸序列。
5. 按照权利要求 1 所述的 5-氨基乙酰丙酸合成酶，其具有与 SEQ ID No:2 所示氨基酸序列之间不超过 2 个氨基酸不同的氨基酸序列。
6. 按照权利要求 1 所述的 5-氨基乙酰丙酸合成酶，其具有与 SEQ ID No:2 所示氨基酸序列之间不超过 1 个氨基酸不同的氨基酸序列。
7. 按照权利要求 1 所述的 5-氨基乙酰丙酸合成酶，其特征在于，它是由在高严谨性条件下，与 SEQ ID No:1 所示核酸序列可发生杂交的核酸序列所编码的。
8. 一种分离的核酸片段，其由编码权利要求 1 所述的 5-氨基乙酰丙酸合成酶的核酸序列所组成。
9. 按照权利要求 8 的核酸片段，其中所述核酸序列编码来自米曲霉 IFO4177 的 5-氨基乙酰丙酸合成酶。
10. 按照权利要求 9 的核酸片段，其中所述核酸序列示于 SEQ ID No:1 中。
11. 按照权利要求 8 的核酸片段，其特征在于，它是在高严谨性条件下与 SEQ ID No:1 所示核酸序列可发生杂交。
12. 一种核酸构建体，其特征在于，它含有权利要求 8-11 任一项所述的核酸片段，该核酸片段有效地与能在合适表达宿主内指导 5-氨基

乙酰丙酸合成酶表达的调控区相连接。

13. 一种重组载体，它包含权利要求 12 的核酸构建体。

14. 按照权利要求 13 的载体，其中所述核酸序列有效地与一段启动子序列连接。

15. 按照权利要求 14 所述的载体，它还含有转录终止信号。

16. 按照权利要求 14 所述的载体，它还含有选择性标记。

17. 一种重组宿主细胞，它含有权利要求 12 所述的核酸构建体。

18. 按照权利要求 17 的宿主细胞，其中所述核酸构建体位于载体上。

19. 按照权利要求 17 的宿主细胞，其中所述宿主细胞为细菌细胞。

20. 按照权利要求 17 的宿主细胞，其中所述宿主细胞为真菌细胞。

21. 按照权利要求 20 的宿主细胞，其中所述真菌细胞是一种丝状真菌细胞。

22. 按照权利要求 21 的宿主细胞，其中所述丝状真菌细胞是一种 *Acremonium* 属、曲霉属 (*Aspergillus*)、镰孢属 (*Fusarium*)、腐质霉属 (*Humicola*)、毁丝霉属 (*Myceliophthora*)、毛霉属 (*Mucor*)、脉孢菌属 (*Neurospora*)、青霉属 (*Penicillium*)、梭孢壳属 (*Thielavia*)、*Tolypocladium* 属或木霉属 (*Trichoderma*) 细胞。

23. 按照权利要求 20 的宿主细胞，其中所述真菌细胞是一种酵母细胞。

24. 按照权利要求 23 的宿主细胞，其中所述酵母细胞是酵母属 (*Saccharomyces*) 或裂殖酵母属 (*Schizosaccharomyces*) 的一个菌株。

25. 按照权利要求 17 的宿主细胞，其中所述核酸构建体被整合至宿主细胞基因组中。

26. 生产权利要求 1 所述的 5-氨基乙酰丙酸合成酶的方法，它包含以下步骤：

(a) 培养一种米曲霉菌株以生产该 5-氨基乙酰丙酸合成酶；以及

(b) 回收该 5-氨基乙酰丙酸合成酶。

27. 生产权利要求 1 所述的 5-氨基乙酰丙酸合成酶的方法，它包

含以下步骤:

(a) 在能促进该 5-氨基乙酰丙酸合成酶表达的条件下培养含有核酸构建体的宿主细胞, 该核酸构建体含有编码该 5-氨基乙酰丙酸合成酶的核酸序列; 以及

(b) 回收该 5-氨基乙酰合成酶。

米曲霉 5 - 氨基乙酰 丙酸合成酶及编码此酶的核酸

本发明涉及米曲霉 5 - 氨基乙酰丙酸合成酶以及含有编码该酶的核酸序列的分离核酸片段。本发明还涉及含有此核酸序列的核酸构建体、载体和宿主细胞。本发明也涉及生产 5 - 氨基乙酰丙酸合成酶 (5-aminolevulinic acid synthases) 的方法。

血红素是原卟啉 IX 与铁的螯合复合物, 可作为血红素蛋白的弥补基 (prosthetic group)。原卟啉 IX 包含一个卟啉环, 该卟啉环具有四个甲基、两个乙烯基和两个丙酸基的取代, 获得一个铁原子后可形成血红素。通过甘氨酸和琥珀酰辅酶 A 生物合成血红素包括 8 个酶反应步骤。此生物合成途径中的第一种酶是 5 - 氨基乙酰丙酸合成酶, 它能催化甘氨酸和琥珀酰辅酶 A 缩合形成 5 - 氨基乙酰丙酸。在肝细胞和分化红细胞内血红素的生物合成中, 5 - 氨基乙酰丙酸合成酶是一种关键的调控酶。

脱辅基蛋白转变为血红素蛋白依赖于血红素生物合成途径提供血红素的能力。血红素蛋白的脱辅基蛋白形式与血红素结合后, 产生活性血红素蛋白。活性血红素蛋白获得的构象使它在蛋白水解的攻击下比脱辅基蛋白更稳定。若微生物体内生成的血红素量低于脱辅基蛋白的生成量, 脱辅基蛋白将会累积并进行蛋白降解, 从而降低了活性血红素蛋白的产量。

为解决这个问题, Jesen 证明, 在发酵培养基中加入血红素或含血红素的物质, 能显著提高米曲霉生产过氧化物酶的产量 (WO 93/19195)。尽管发酵培养基中补加血红素能显著提高血红素蛋白产量, 但在应用于大规模生产中, 这种方法有悖常理, 既费钱, 又难于操作。

构巢曲霉中 5 - 氨基乙酰丙酸合成酶基因的克隆和序列测定已

被公开 (Bradshaw 等, 1993 年, 现代遗传学 (Current Genetics), 第 2233 卷, 501 - 507 页)。

本发明的一个目的是提供几种新的 5 - 氨基乙酰丙酸合成酶以及编码这些酶的基因。

本发明涉及来源于米曲霉的基本纯的 5 - 氨基乙酰丙酸合成酶, 以及含有编码米曲霉 5 - 氨基乙酰丙酸合成酶的核苷酸序列的分离核酸片段。本发明还提供了含有本发明核酸片段的核酸构建体、载体和重组宿主细胞。本发明也提供了生产 5 - 氨基乙酰丙酸合成酶的方法。

图 1 是质粒 pSE04 的限制性图谱。

图 2 是含有米曲霉 5 - 氨基乙酰丙酸合成酶基因的一段 4.2kb 基因组片段的限制性图谱。图下方给出了用千碱基对 (kb) 标示的刻度。箭头指示该基因开放阅读框的位置。

图 3 给出了米曲霉 5 - 氨基乙酰丙酸合成酶基因的核苷酸序列及由此推断出的氨基酸序列 (相应为 SEQ ID No:1 和 SEQ ID No:2)。公认的重要转录位点 CCAAT 盒和 TATA 盒由下划线标出。方框表示两个保守的称为 HRM 的基元 (motif)。圆圈表示参与磷酸吡哆醛辅助因子结合的甘氨酸环。星号指示重要的赖氨酸。

图 4 给出了多种 5 - 氨基乙酰丙酸合成酶基因中的保守血红素调节基元。方框表示戊肽 (pentapeptide) 基元。

图 5 给出了米曲霉、构巢曲霉、酿酒酵母 (*Saccharomyces cerevisiae*) 和人红细胞 5 - 氨基乙酰丙酸合成酶推断的氨基酸序列 (分别为 SEQ ID No:2, 16, 17 和 18) 的比较。保守的氨基酸用方框划出。

图 6 给出了质粒 pBANe6 的限制性图谱。

图 7 给出了质粒 pSE31 的限制性图谱。

图 8 给出了质粒 pJVi9 的构建过程。

图 9 给出了质粒 pJeRS6 的限制性图谱。

图 10 给出了质粒 pJRoC50 的限制性图谱。

如上述所提及, 本发明涉及得自一种米曲霉菌株的 5 - 氨基乙

酰丙酸合成酶。公众已能从一系列培养物保藏中心获得此米曲霉种的菌株，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），国际微生物研究所（IMI），农业研究机构专利培养物保藏中心，北方地区研究中心（NRRL），以及日本大阪发酵研究所（IFO）。

在一个优选实施方案中，本发明涉及得自米曲霉或其突变株的5-氨基乙酰丙酸合成酶。在一个更优选的实施方案中，本发明涉及得自米曲霉 IFO4177 或其突变株的5-氨基乙酰丙酸合成酶，如具有 SEQ ID No:2 所示氨基酸序列的5-氨基乙酰丙酸合成酶。

本发明还涉及在高严谨性条件下（即预杂交和杂交是在 45 °C 下，在 5 × SSPE，0.3 % SDS, 200µg/ml 已剪切和变性的鲑鱼精 DNA，以及 50 % 甲酰胺中进行），与在相同条件下可跟 SEQ ID No:1 所示核酸序列杂交上的探针也能杂交上的核酸序列所编码的5-氨基乙酰丙酸合成酶。该基因或以该基因为基础的寡核苷酸，可在 Southern 杂交中用作探针分离任何曲霉种的同源基因。特别是，这类探针可用于按照标准的 Southern 杂交方法对感兴趣的种类的基因组或 cDNA 进行的杂交中，以鉴定和分离其中相应的5-氨基乙酰丙酸合成酶基因。可将简并性 PCR 引物（寡核苷酸）与基因组 DNA 或 cDNA 片段一起用于扩增5-氨基乙酰丙酸合成酶特异性基因片段。

通过本文实施例中介绍的方法，利用公开可获得的曲霉菌株，可以从本文特别列举的例子以外的来源鉴定和分离5-氨基乙酰丙酸合成酶基因。

为了本发明目的，术语“得自”意指所说的5-氨基乙酰丙酸合成酶产生自一种特殊来源，如一种曲霉菌株，或产生自插入有该来源的编码5-氨基乙酰丙酸合成酶的基因的细胞。

本发明也包括与 SEQ ID No:2 所示氨基酸序列同源性至少约为 80 %，优选约为 85 %，更优选约为 90 %，最优选约为 95 % 的5-氨基乙酰丙酸合成酶变种，以及性质上仍保留有本文所述的5-氨基乙酰丙酸合成酶活性的5-氨基乙酰丙酸合成酶变种。本发明也包括

与 SEQ ID No: 2 所示氨基酸序列具有三个氨基酸, 优选两个氨基酸, 更优选一个氨基酸差异的氨基酸序列的 5-氨基乙酰丙酸合成酶。每种差异可以是一个氨基酸的插入或缺失, 或者是一个氨基酸被另一个不同的氨基酸所替代。上述种类中有用的变种包括如发生了保守氨基酸替代的变种, 这种替代不显著影响蛋白质活性。保守替代意指同类氨基酸可以由该类内的其它任何一种氨基酸所替代。例如, 非极性脂肪类残基 Ala、Val、Leu 和 Ile 可进行互变, 碱性残基 Lys 和 Arg 可互变, 酸性残基 Asp 和 Glu 也可互变。类似地, Ser 和 Thr 相互也为保守替代, Asn 和 Gln 也一样。

通过本领域熟知的各种技术, 包括但不局限于 SDS-PAGE、等电聚焦和交叉反应免疫同一性测定 (cross-reaction immunoidentity test), 可以测定本发明的 5-氨基乙酰丙酸合成酶的物理化学性质。利用本领域已知方法可对本发明的 5-氨基乙酰丙酸合成酶进行分析。

通过本领域所知的多种方法, 可纯化本发明的 5-氨基乙酰丙酸合成酶, 这些方法包括但不局限于层析法 (如离子交换, 亲和, 疏水性, 层析聚焦和尺寸排阻层析), 电泳法 (如制备性等电聚焦), 差异溶解力法 (differential solubility) (如硫酸铵沉淀法), 或抽提法 (参见如《蛋白质纯化》, J.-C. Janson 和 Lars Ryden 编, VCH 出版社, 纽约, 1989 年)。如本文所定义, “基本纯的” 5-氨基乙酰丙酸合成酶是指基本上不含有其它非 5-氨基乙酰丙酸合成酶蛋白质的 5-氨基乙酰丙酸合成酶, 例如, 按照 SDS-PAGE 所测定, 至少为约 20 % 纯, 优选为约 40 % 纯, 更优选为约 60 % 纯, 更更优选为约 80 % 纯, 最优选为约 90 % 纯, 最最优选为约 95 % 纯。

本发明还涉及含有编码本发明的 5-氨基乙酰丙酸合成酶的核酸序列的核酸片段, 以及含有本发明的核酸片段的核酸构建体。

在一个优选实施方案中, 此核酸序列编码得自米曲霉的 5-氨基乙酰丙酸合成酶。在一个更优选的实施方案中, 该核酸序列编码获自米曲霉 IFO4177 的 5-氨基乙酰丙酸合成酶, 如 SEQ ID No:1 所示核酸序列。本发明也包括编码具 SEQ ID No:2 所示氨基酸序列的 5

— 氨基乙酰丙酸合成酶、但由于遗传密码的简并性而不同于 SEQ ID No:1 的核酸序列。本发明的核酸序列包括本文所述的基因组序列，以及相应的 cDNA 和 RNA 序列。此处所用词语“核酸序列”应理解为包括合成 DNA 在内的所有这类变种。

本发明也涉及含有本发明核酸片段的核酸构建体。“核酸构建体”一般可理解为意指单链或双链的核酸分子，它分离自自然界存在的基因，或是已被修饰从而含有以自然界并不存在的方式进行组合和并置的几段核酸。在一个优选实施方案中，该核酸构建体有效地与能在适宜表达宿主内指导 5 — 氨基乙酰丙酸合成酶表达的调控区相连接。

本发明也提供了含有本发明的核酸构建体的重组载体。在一个优选实施方案中，核酸序列有效地与一段启动子序列连接。在另一个优选实施方案中，本发明的载体还含有转录终止信号和/或筛选标记。

本发明的重组载体在以活性形式表达米曲霉 5 — 氨基乙酰丙酸合成酶基因时是很有用的。一个有用的载体包括能允许该载体稳定整合至宿主细胞基因组、或在宿主细胞内能独立于该宿主细胞基因组而进行自主复制的元件，并优选还含有便于转化的宿主细胞筛选的一个或数个表型标记。该载体还可含有一些控制序列，如启动子，核糖体结合位点，翻译起始信号，另外，还可选择性地含有一个选择标记或多种激活子（activator）或阻遏子（repressor）序列。为使表达蛋白能分泌出来，也可在该基因的编码序列前插入一段编码信号序列的核酸。为了在控制序列的指导下进行表达，将按照本发明而使用的 5 — 氨基乙酰丙酸合成酶基因与该控制序列有效连接，以便能在与控制序列相容的条件下获得编码序列的表达。

含有本发明核酸构建体的载体可以是能便利地进行重组 DNA 操作的任何载体。载体的选择典型地取决于欲导入此载体的宿主细胞。载体可以是自主复制型载体，即以染色体外完整形式存在、独立于染色体复制而进行复制的载体，如质粒、染色体外元件、微小染色体或人工染色体。另外，载体也可以是导入至宿主细胞内时，能整合

至宿主细胞基因组中并与其整合入的染色体一起复制的载体。载体系统可以是单一载体或质粒,也可以是合在一起含有待整合至基因组内的总 DNA 的两个或多个载体或质粒。

在载体中, DNA 序列应与适当的启动子序列有效相连。所述启动子可以是在选定的宿主细胞内具有转录活性的任何 DNA 序列,并可得自编码与宿主细胞同源或异源的蛋白质的基因。能指导本发明的核酸构建体的转录尤其是在细菌宿主中转录的适当启动子的例子有大肠杆菌半乳糖 (lac) 操纵子的启动子、天蓝色链霉菌琼脂糖酶基因 *dagA* 启动子、地衣芽孢杆菌 α -淀粉酶基因 (*amyL*) 的启动子、嗜热脂肪芽孢杆菌产麦芽糖的淀粉酶基因 (*amyM*) 的启动子、解淀粉芽孢杆菌 α -淀粉酶基因 (*amyQ*) 的启动子、枯草芽孢杆菌 *xylA* 和 *xylB* 基因的启动子、原核性 β -内酰胺酶基因启动子 (Villa-Kamaroff 等, 1978, 美国国家科学院院报, 75: 3727 - 3731) 或 *tac* 启动子 (DeBoer 等, 1983, 美国国家科学院院报, 80: 21 - 25)。其它启动子以“重组细菌来源的有用蛋白质”为题记载于“科技美国”, 1980, 242: 74 - 94, 也记载于 Sambrook 等, “分子克隆, 实验室手册”, 第 2 版, 冷泉港, 纽约, 1989。在酵母宿主中, 有用的启动子是 *eno-1* 启动子。为了在真菌宿主中进行转录, 有用的启动子的例子有得自编码米曲霉 TAKA 淀粉酶、*Rhizomucor miehei* 天冬氨酸蛋白酶、黑曲霉中性 α -淀粉酶、黑曲霉酸稳定性 α -淀粉酶、黑曲霉或泡盛曲霉葡糖淀粉酶 (*glaA*)、*Rhizomucor miehei* 脂肪酶、米曲霉碱性蛋白酶、米曲霉磷酸丙糖异构酶或构巢曲霉乙酰胺酶 (*acetamidase*) 基因的启动子。特别优选的启动子是 TAKA 淀粉酶、NA2 - *tpi* (来源于编码黑曲霉中性 α -淀粉酶和米曲霉磷酸丙糖异构酶的基因的启动子的杂合体) 以及 *glaA* 启动子。

本发明的载体另外也可含有适当的转录终止子以及在真核生物内, 与编码本发明的 5-氨基乙酰丙酸合成酶的 DNA 序列有效连接的多聚腺苷酸化序列。终止序列和多聚腺苷酸化序列可与启动子来源相同。载体还可含有能促使载体在选定宿主细胞中复制的 DNA 序

列。这类序列的例子有质粒 pUC19 、 pACYC177 、 pUB110 、 pE194 、 pAMB1 和 pIJ702 的复制起始点。

本发明的载体优选地含有能便于对转化细胞进行筛选的一个或多个选择性标记。选择性标记是其产物能提供对杀生物剂或对病毒的抗性、重金属抗性、营养缺陷体的原养型等的基因。该选择性标记可选自下述组，所述组中含有但并不限于 amdS 、 pyrG 、 argB 、 niaD 、 sC 、 trpC 、 bar 和 hygB 。在曲霉细胞中，优选使用构巢曲霉或米曲霉的 amdS 和 pyrG 标记，以及吸水链霉菌（*Streptomyces hygroscopicus*）的 bar 标记。此外，也可通过共转化完成筛选，如 WO 91/17243 所述，其中选择性标记包含于另一个载体中。

本发明的载体优选地还含有信号肽编码区，所述编码区编码与血红素生物合成酶氨基末端相连的一段氨基酸序列，该氨基酸序列能将 5 - 氨基乙酰丙酸合成酶定位至特定的细胞区室中。信号肽编码序列可以是编码 5 - 氨基乙酰丙酸合成酶的第一种核酸序列的内部片段，也可以得自外源。第一种核酸序列的编码序列的 5' 末端本身可能含有信号肽编码区，所述信号肽编码区与编码定位的 5 - 氨基乙酰丙酸合成酶的编码区的片段在翻译读框内自然相连；或者，编码序列的 5' 末端可含有编码相对于编码定位的血红素生物合成酶的编码序列部分而言是外源的信号肽编码区的核酸。信号肽编码区可得自粗糙脉孢霉 ATPase 基因（Viebrock 等，1982，EMBO Journal 1:565-571）或酿酒酵母细胞色素 C 过氧化物酶基因（Kaput 等，1982，生物化学杂志，257:15054 - 15058）。然而，能够在选定的丝状真菌宿主中使 5 - 氨基乙酰丙酸合成酶定位的任何信号肽编码区在本发明中都可使用。

为避免在获得表达的 5 - 氨基乙酰丙酸合成酶时需破坏细胞，并使细胞内表达的 5 - 氨基乙酰丙酸合成酶可能的降解量减到最小，优选 5 - 氨基乙酰丙酸合成酶基因的表达产物能分泌至细胞外。为达此目的，本发明的 5 - 氨基乙酰丙酸合成酶可含有一个允许将此表达的蛋白质分泌至培养基中的前区域（preregion）。如果有必要

的话,此前区域可以是本发明的 5-氨基乙酰丙酸合成酶自身的,或者是由不同的前区域或信号肽所替代,后者可通过编码相应前区域的 DNA 序列的替代方便地完成。例如,所述前区域可以得自来源于曲霉属菌种的葡糖淀粉酶或淀粉酶基因、来源于芽孢杆菌属菌种的淀粉酶基因、来源于 *Rhizomucor miehei* 的脂肪酶或蛋白酶基因、来源于酿酒酵母的编码 α -因子的基因或是小牛前凝乳酶原 (preprochymosin) 基因。尤其优选的是米曲霉 TAKA 淀粉酶、黑曲霉中性淀粉酶、来源于芽孢杆菌 NCIB 11837 产麦芽糖的淀粉酶、嗜热脂肪芽孢杆菌 α -淀粉酶或地衣芽孢杆菌枯草蛋白酶的前区域。真菌宿主中有效的信号序列是米曲霉 TAKA 淀粉酶、*Rhizomucor miehei* 天冬氨酸蛋白酶或 *Rhizomucor miehei* 脂肪酶的信号序列。

本领域普通技术人员熟知用于连接本发明的核酸构建体、启动子、终止子和其它元件以及将它们插入至含有复制必需信息的适当载体中的方法 (参见如 Sambrook 等, 文献同上)。

本发明也涉及含有本发明核酸构建体或表达载体的宿主细胞,所述宿主细胞可有利地应用于本发明的 5-氨基乙酰丙酸合成酶的重组生产中。该细胞可利用本发明的核酸构建体、通过将此构建体整合至宿主染色体上而便利地进行转化。由于此时序列更有可能稳定地保持在细胞中,因此一般认为这种整合是更具优越性的。构建体在宿主染色体上的整合可按照常规方法进行,如同源或非同源重组。另外,对不同类型的宿主细胞,利用表达载体,可采用下述方法进行转化。

宿主细胞和载体的选择很大程度上取决于 5-氨基乙酰丙酸合成酶及其来源。宿主细胞可以选择原核细胞,如细菌细胞。适当细菌的例子有革兰氏阳性细菌,如枯草芽孢杆菌、地衣芽孢杆菌、迟缓芽孢杆菌、短芽孢杆菌、嗜热脂肪芽孢杆菌、嗜碱芽孢杆菌 (*Bacillus alkalophilus*)、解淀粉芽孢杆菌、凝结芽孢杆菌、环状芽孢杆菌、灿烂芽孢杆菌、巨大芽孢杆菌、苏云金芽孢杆菌或浅青紫链霉菌或鼠灰链霉菌,或是革兰氏阴性细菌,如大肠杆菌。细菌的转化可通过如

原生质体转化或利用感受态细胞以自身所知的方式进行。

宿主细胞优选真核细胞，如哺乳动物细胞、昆虫细胞、植物细胞，或优选为真菌细胞，包括酵母和丝状真菌。例如，有用的哺乳动物细胞包括 CHO 或 COS 细胞。酵母宿主细胞可选自酵母属或裂殖酵母属的种类，如酿酒酵母。有用的丝状真菌可选自曲霉属的种类，如米曲霉或黑曲霉。另外，镰孢属种类的菌株，如尖镰孢或禾本科镰孢也可用作宿主细胞。真菌细胞可通过包括原生质体形成、原生质体转化和细胞壁再生的方法，以其自身所知的方式进行转化。曲霉属宿主细胞转化的适当方法描述于 EP238023。镰孢属种类转化的适当方法描述于 Malardier 等，1989，基因，78：147 - 156，或共同未决的美国流水号 08/269,449。

在一个尤其优选的实施方案中，5 - 氨基乙酰丙酸合成酶基因的表达是在真菌宿主细胞如曲霉中完成的。将 5 - 氨基乙酰丙酸合成酶基因连接至质粒中，该质粒优选地含有米曲霉 TAKA 淀粉酶启动子或黑曲霉中性淀粉酶 NA2 启动子，并含有 amdS 或 pyrG 作为选择标记。另外，选择标记也可位于另一单独的质粒上，并用于共转化。可按照 Yelton 等，1984，美国国家科学院院报，81：1470 - 1474 记载的方法，利用这个（些）质粒对曲霉属宿主细胞如米曲霉或黑曲霉进行转化。

本发明也涉及生产本发明的 5 - 氨基乙酰丙酸合成酶的方法，所述方法包括：（a）在营养培养基中培养米曲霉菌株以生产 5 - 氨基乙酰丙酸合成酶，和（b）回收 5 - 氨基乙酰丙酸合成酶。

本发明还涉及重组生产本发明的 5 - 氨基乙酰丙酸合成酶的方法，所述方法包括：（a）在利于生产该酶的条件下，对含有核酸构建体的宿主细胞进行发酵，其中所述核酸构建体含有编码 5 - 氨基乙酰丙酸合成酶的核酸序列，和（b）回收 5 - 氨基乙酰丙酸合成酶。若表达系统将 5 - 氨基乙酰丙酸合成酶分泌至发酵培养基中，也可从培养基中直接回收此酶。若重组 5 - 氨基乙酰丙酸合成酶不被分泌，则从细胞裂解物中回收此酶。

能表达或分离本发明的 5 - 氨基乙酰丙酸合成酶的本领域已知。

的培养细胞的任何方法均可使用。例如,培养可理解为包括在实验室或工业发酵罐内适当的培养基中,允许5-氨基乙酰丙酸合成酶表达或分离的条件下进行的摇瓶培养、小规模或大规模发酵(包括连续、分批、分批补料或固相发酵)。培养是利用本领域已知方法(参见如Bennett J.W. 和 LaSure L.(编), *More Gene Manipulations in Fungi*, 学术出版社, CA, 1991)、在含有碳、氮源和无机盐的适当营养培养基中进行。适当的培养基可商购,或按照公开的组合物(如美国典型培养物保藏中心的目录中)配制而成。

按上述方法生产的5-氨基乙酰丙酸合成酶可通过常规方法从发酵培养基中回收,所述常规方法包括但不限于离心、过滤、雾化-干燥、蒸发或沉淀。此回收的蛋白质随后可通过多种层析方法进一步纯化,如离子交换层析、凝胶过滤层析、亲和层析等。

本发明也涉及5-氨基乙酰丙酸合成酶的使用方法。

可利用本发明的5-氨基乙酰丙酸合成酶将甘氨酸和琥珀酰辅酶A转化为可用作除草剂的5-氨基乙酰丙酸。

也可将本发明的5-氨基乙酰丙酸合成酶用于提高宿主细胞生产血红蛋白的产量,其中5-氨基乙酰丙酸合成酶是宿主细胞生产血红蛋白的限速步骤。这种血红蛋白产量的提高是通过在宿主细胞中过量表达编码5-氨基乙酰丙酸合成酶的核酸序列而完成的。该方法包括:

(a) 将编码5-氨基乙酰丙酸合成酶之核酸序列的一个或多个拷贝导入至能产生血红蛋白的宿主细胞中,其中所述核酸序列与能指导5-氨基乙酰丙酸合成酶表达的调控区有效相连;

(b) 在适于生产血红蛋白和5-氨基乙酰丙酸合成酶的营养培养基中培养该细胞;以及

(c) 从该细胞的营养培养基中回收血红蛋白。

本发明由下述实施例进一步描述。这些实施不应被认为是对本发明的范围的限制。

实施例1: 提取米曲霉菌株 A1560 基因组 DNA

于32℃, 250rpm下, 在25ml 0.5%酵母提取物-2%葡萄糖

(YEG) 培养基中将米曲霉 A1560 菌株 (IFO4177) 24 小时, 然后通过 Miracloth (Calbiochem, La Jolla, CA) 过滤收集菌丝, 用 25ml 10mM Tris-1mM EDTA(TE) 缓冲液洗涤 1 次, 吸干菌丝上多余的缓冲液, 随后将其在液氮中冷冻。在电动咖啡磨碎机中将冷冻的菌丝磨成细粉末, 将粉末加入一次性塑料离心管内的 20ml TE 缓冲液和 5ml 20% w/v 十二烷基硫酸钠 (SDS) 中, 将混合物轻轻颠倒几次以确保混匀。用等体积的苯酚: 氯仿: 异戊醇 (25: 24: 1 v/v/v) 抽提 2 次。加入醋酸钠 (3M 溶液) 使其终浓度为 0.3M, 用 2.5 倍体积的冰冷的乙醇沉淀核酸。以 $15,000 \times g$ 将此管离心 30 分钟得到核酸沉淀。将沉淀物风干 30 分钟, 然后重新悬浮于 0.5ml TE 缓冲液中。加入不含 DNA 酶的核糖核酸酶 A 至浓度为 $100\mu\text{g/ml}$, 将混合物在 37°C 下保温 30 分钟。然后加入蛋白酶 K 至浓度为 $200\mu\text{g/ml}$, 再在 37°C 下将混合物继续保温 1 小时。最后, 如前述, 用苯酚: 氯仿: 异戊醇 (25: 24: 1 v/v/v) 将混合物抽提 2 次, 再用醋酸钠和乙醇沉淀 DNA。真空干燥 DNA 沉淀物, 重新悬浮于 TE 缓冲液中, 贮于 4°C 以供后面的使用。

实施例 2: 构建质粒 pSE04

利用实施例 1 所述的同样方法提取构巢曲霉 A26 菌株 (真菌遗传学保藏中心, Kansas City, KS) 的基因组 DNA。将通过含有构巢曲霉 A26 基因组 DNA 的扩增反应获得的 PCR 片段连接而构建成质粒 pSE 04。扩增反应中含有下述组分: 构巢曲霉 A26 基因组 DNA 50ng, dATP、dCTP、dGTP、dTTP (Boehringer Mannheim, Indianapolis, IN) 各 $100\mu\text{M}$, 引物 ALAS3d 5' - TTTATGATGGAGGCCCTTCTCCAGCAGTCTC - 3' (SEQ ID No:3) 和引物 AIAS4e 5' - CTATGCATTTAAGCAGCAGCCGC GACTGG - 3' (SEQ ID No:4) 各 50pmol, Taq DNA 聚合酶 (Perkin-Elmer Corp., Branchburg, NJ) 2 单位, 以及 $1 \times$ Taq DNA 聚合酶缓冲液 (Perkin-Elmer Corp., Branchburg, NJ)。反应在 Perkin-Elmer 热循环仪 (Perkin -Elmer Corp., Branchburg, NJ) 中进行 30 个循环, 每个循环包括 95°C 1 分钟、 55°C 1 分钟以

及 72 °C 90 秒。在 40mM Tris - 乙酸 - 1mM 乙二胺四乙酸二钠 (TAE) 缓冲液中, 利用 1.1 % 低熔点琼脂糖凝胶 (FMC , Rockland, ME) 电泳后通过切割条带分离得到 2kb 的 PCR 产物, 并按照制造商的说明将其亚克隆至 pCRII 载体中 (Invitrogen, San Diego, CA) 以获得 pSE04(图 1)。

实施例 3 : 米曲霉 A1560 菌株 DNA 文库及 ALA 合成酶 (hema) 克隆的鉴定

按照制造商的说明, 将 *E. coli* Y1090 ZL 细胞用作平铺和纯化重组噬菌体的宿主, 将 *E. coli* DH10BZip 用作切割各个 pZL1-hema 克隆的宿主, 利用噬菌体克隆载体入 ZipLox(Life Technologies, Gaithersburg, MD)构建米曲霉 A1560 菌株的基因组 DNA 文库。用 Tsp509I 部分消化按实施例 1 制备的总细胞 DNA , 并在 50mM Tris - 50mM 硼酸盐 - 1mM 乙二胺四乙酸二钠 (TBE) 缓冲液中利用 1 % 琼脂糖凝胶进行片段大小分离。切下迁移范围在 4 - 7kb 的 DNA 片段, 并利用 Prep-a-Gene 试剂 (BioRad Laboratories, Hercules, CA) 从凝胶中将 DNA 片段洗脱下来。将洗脱下来的 DNA 片段与已经 EcoRI 切割并去磷酸化的入 ZipLox 载体臂进行连接, 并利用可商购的包装抽提物 (Stratagene, La Jolla, CA) 对连接混合物进行包装。在 *E.coli* Y1090ZL 细胞中平铺和扩增经包装的 DNA 文库。未扩增的基因组文库含有 1×10^6 pfu/ml 。

将 7×10^4 个噬斑的噬菌体 DNA 转移至双层圆形 Nytran Plus 膜 (Schleicher & Schuell, Keene, NH) 上, 并利用地高辛 (DIG) 标记的探针进行探测, 此探针是通过 PCR 扩增来自实施例 2 所述质粒 pSE04 的构巢曲霉 hema 基因组 DNA 制备而成。扩增反应含有下述组分: $1 \times$ DIG 探针合成混合液 (Boehringer Mannheim, Indianapolis, IN) , dATP 、 dCTP 、 dTTP 和 dGTP 各 100 μ M , 实施例 2 所述引物 ALAS3d 和引物 ALAS4e 各 50pmol, Taq DNA 聚合酶 2 单位, 以及 $1 \times$ Taq DNA 聚合酶缓冲液。反应在 Perkin-Elmer 热循环仪中进行 30 个循环, 每个循环包括 95 °C 1 分钟, 55 °C 1 分钟, 以及 72 °C 2 分钟。以 2ng/ml 的浓度将经变性的探针加入至杂交

缓冲液中，并与经预杂交的膜一起保温过夜。预杂交和杂交是在 42℃下，在 5 × SSC、0.1 % Sarkosyl、0.02 % SDS、1 % Genius 阻断试剂（Boehringer Mannheim, Indianapolis, IN）和 30 % 甲酰胺中进行的。膜在 5 × SSC - 0.1 % SDS 中洗涤两次，每次洗涤均是在室温下进行 15 分钟。按照厂商说明，于室温将洗涤好的膜对 Kodak X-OMAT AR 胶片曝光约 2 小时，再利用 Konica QX-70 自动胶片处理机进行显影。对初始噬斑进行再次纯化和筛选。鉴定出 5 个克隆，再按照厂商说明（Bethesda Research Laboratories Inc., Gaithersburg, MD）将其切割至 pZL 衍生物中。将所述 pZL 衍生物命名为 E. coli DH 5α pSE11, pSE13, pSE15, pSE17 和 pSE20。已发现这些克隆有所重叠，它们跨越一个 4.2kb 的区域，此区域的限制性图谱示于图 2。

实施例 4：利用 5 - 氨基乙酰丙酸合成酶（hemA）探针进行米曲霉 A1560 菌株基因组 DNA 的 Southern 杂交

将按实施例 1 所述的方法制备的米曲霉 A1560 菌株基因组 DNA（10μg）用 BamHI 或 EcoRI 进行限制性消化。在 1 % 琼脂糖 - TBE 凝胶上电泳分离片段。按照厂商说明，利用 TurboBlot 装置（Schleicher & Schuell, Keene, NH），在 0.4N NaOH 中将 DNA 转移至 NytranPlus 膜上。在 Hybaid 炉（Labnet, Woodbridge, NJ）内，将膜在 5 × SSC、0.1 % sarkosyl、0.02%SDS、1 % Genius 阻断试剂（Boehringer Mannheim, Indianapolis, IN）和 50 % 甲酰胺中于 42℃预杂交 2 小时。通过使用 DIG 标记的 hemA 探针进行杂交，该探针如实施例 3 所述通过 PCR 扩增获得，不同之处在于使用 hemA 克隆 pSE17 作为模板，并使用引物 hemA5' 5' - TCATTTAAATGATGGAGTCTCTTCTCC - 3'（SEQ ID No:5）和引物 hemA3' 5' - TCTTAATTAATCAGCTCACATGCGGG - 3'（SEQ ID No:6）。将 DIG 标记的 hemA 探针（1ng 探针/ml 溶液）加入新鲜的杂交缓冲液中，与膜一起在 42℃保温过夜。随后，将膜在 5 × SSC - 0.1 % SDS 中于室温洗涤 2 次，每次 15 分钟，再在 2 × SSC - 0.1 % SDS 中于相同的条件下洗涤 2 次。于室温下将洗涤后的膜

对 Kodak X-OMAT AR 胶片曝光约 2 小时，随后按照厂商说明，利用 Konica QX-70 自动胶片处理机进行显影。

利用米曲霉 hemA 探针与米曲霉基因组 DNA 一起进行的 Southern 印迹杂交结果表明杂交信号与单基因拷贝数一致。BamHI 泳道上观察到的 1.7kb 条带可从限制性图谱上推断出来（图 2）。

实施例 5：米曲霉 A1560 5-氨基乙酰丙酸合成酶（hemA）基因的特性

按下述方法，对实施例 3 中所述的 E. coli DH5 α pSE17 进行 DNA 测序。DNA 测序是在 Applied Biosystems 的 373A 型 DNA 自动测序仪（Applied Biosystems, Inc., Foster City, CA）上，使用含染料-终止剂化学的引物步行技术（Giesecke 等，1992，病毒学方法杂志，38：47-60）对两条链进行的，测序中使用了 M13 反向引物（-48）和 M13 正向引物（-20）（New England Biolabs, Beverly, MA），以及此待测序 DNA 的特异引物。

如图 3（SEQ ID No:1）所示，此克隆基因的核苷酸序列表明其中含有一个具 1911 个核苷酸的开放阅读框。该编码序列中不包含任何内含子，这一点可通过 cDNA 克隆和序列分析得到证实；与之相反，构巢曲霉 hemA 基因在其 5' 末端含有一个内含子（Bradshaw 等，1993，现代遗传学（Current Genetics），23:501-507）。如其它真菌基因一样，此克隆基因的 5' 非翻译序列含有一些嘧啶富含区和 AT 富含区（Gurr 等，1987，Kinghorn J.R.(编)，《真核微生物的基因结构》，pp93-139, IRL 出版社，Oxford），另外它还含有在 -249 位的 CCAAT 序列以及在 -35 位的公认 TATA 盒。此 CCAAT 序列是转录调节子的一致结合位点，所述的转录调节子能对氧作出反应从而调节转录，如酵母和人中的 Hap2/3/4 转录调节复合物（Olesen 和 Guarente, 1990，分子和细胞生物学，12：2302-2314）。此调节复合物在哺乳动物中也是保守的。在构巢曲霉中也检测到 CCAAT 结合活性（Davis 等，1993，Genetica 90:133-145）。在米曲霉 hemA 基因中此序列的重要性仍不清楚，并且，由于序列信息有限，构巢曲霉 hemA 5' 区内此序列的重要性也未能确证

(Bradshaw 等, 1993, 出处同前)。米曲霉 hemA 基因对氧作出反应从而受到的转录调控目前还不清楚, 但是甚至在氧限制条件下, 构巢曲霉 hemA 基因也没有显现出受到转录调控 (Bradshaw et al., 1993, 同处同前)。有趣的是, 酵母 HEM1 基因也是组成型表达, 但其表达是通过正调控位点与负调控位点间的平衡来控制的 (Keng 和 Guarente, 1987, 美国国家科学院院报, 84: 9113 - 9117)。此克隆基因的 3'非翻译区存在一个 (AC)₃₅ 重复基元 (motif)。在哺乳动物和酵母基因的亚端粒、内含子和启动子区也观察到类似的重复区, 尽管它们在基因扩增过程中有所参与, 但它们也不具有已知的功能 (Passanati 等, 1987, EMBO Journal 6:1967-1703)。

米曲霉 A1560 菌株之基因产物推断的氨基酸序列示于图 3 (SEQ ID No:2)。该核苷酸序列编码分子量为 68kDa、含 636 个氨基酸的预测蛋白质。由于此酶位于线粒体中, 推测其 N 末端含有线粒体前导序列。事实上, 此蛋白质前 35 个氨基酸富含丝氨酸、苏氨酸、赖氨酸和精氨酸残基, 这与作为线粒体前导序列的功能是一致的。在构巢曲霉和米曲霉 hemA 序列 (图 4) 推测的线粒体前导序列中都有一个可能的血红素调节基元 (HRM)。估计位于前导序列中的 HRM 能通过与血红素直接作用而阻止将 5-氨基乙酰丙酸合成酶蛋白输入小鼠线粒体中 (Lathrop 和 Timko, 1993, 科学, 259: 522 - 525; Zhang 和 Guarente, 1995, EMBO Journal 14: 313-320)。另一个可能的 HRM 也位于推定的成熟蛋白质序列的开始处。HRM 很可能在 5-氨基乙酰丙酸合成酶活性的调控中起着一定的作用。有趣的是, 酿酒酵母 5-氨基乙酰丙酸合成酶蛋白质序列不含有任何可能的 HRM, 并且, 它似乎并不是酵母血红素生物合成中的一个关键调控步骤 (Labbe-Bois 和 Labbe, 于 Daley, Harry A. 所编, 《血红素和叶绿素的生物合成》, McGraw Hill Publishers, 纽约, pp235-285)。

总的说来, 利用 Applied Biosystems 的 GeneAssist 程序 (blosum 62. mat matrix) 测定的结果表明图 5 所示的推断氨基酸序列与构巢曲霉 hemA 基因 (SEQ ID No:16) 有 81 % 的

同一性，与酿酒酵母 HEM1 基因（SEQ ID No:17；Urban-Grimal, 1986，生物化学欧洲杂志，156：511 - 519）有 57% 的同一性，与人红细胞 heml（ALAS2）基因（SEQ ID No:18；Bishop, 1990，核酸研究，18：7187 - 7188）有 51% 的同一性。然而，最高程度的保守位于蛋白质包括有催化结构域在内的 C 端 2/3 区。另外，在其它 5 - 氨基乙酰丙酸合成酶（Ferreira 等，1995，生物能和生物膜杂志（Journal of Bioenergetics and Biomembranes），27:151-159；Ferreira, 1995, 蛋白质科学，4：1001 - 1006）中对催化活性和吡哆醛磷酸辅助因子结合有重要作用的赖氨酸和甘氨酸环（loop）在此推断的氨基酸序列中也是高度保守的。

实施例 6：载体 pSE31 的构建

通过将 PCR 扩增的米曲霉 hemA DNA 定向克隆至 pBANe6 中构建成质粒 pSE31（图 6）。PCR 扩增反应中使用了来源于实施例 3 所述之 hemA 克隆 E.coli DH5 α pSE17 的 DNA。反应中含有下述组分：50ng pSE 17、2 单位 Vent DNA 聚合酶（New England Biolabs, Beverly, MA），1 $\times$ Vent DNA 聚合酶缓冲液（New England Biolabs, Beverly, MA），各 400 μ M 的 dATP、dCTP、dGTP 和 dTTP（Boehringer Mannheim, Indianapolis, IN），以及各 50pmol 的引物 hemA5' 5' - TCATTTAAATGATGGAGTCTCTTCTCC - 3'（SEQ ID No:5）和引物 hemA 3' 5' - TCTTAATTAATCAGCTCACATGCGGG - 3'（SEQ ID No:6）。反应在 Perkin-Elmer 热循环仪中进行 30 个循环，每个循环包括 95 $^{\circ}$ C 1 分钟、55 $^{\circ}$ C 1 分钟以及 72 $^{\circ}$ C 90 秒。引物 hemA 5' 含有一个 SwaI 位点（下划线），引物 hemA 3' 含有一个 PacI 位点（下划线），这两个位点用于将 PCR 产物克隆至经 SwaI 和 PacI 消化的 pBANe6 中以产生 pSE31（图 7）。

实施例 7：米曲霉 JRoC50.3.18A 菌株的构建

按下述方法构建含有质粒 pJROC50 的米曲霉 JRoC50.3.

18A 菌株。利用如下所示的特异性寡核苷酸引物 (Saiki 等, 1988, 科学, 239: 487 - 491), 通过 PCR 制备 *Coprinus cinereus* IFO 8371 过氧化物酶 cDNA 片段。这些引物是依据 *Coprinus macrorhizus* 过氧化物酶的氨基酸序列 (Baunsgaard 等, 1993, 生物化学欧洲杂志, 213: 605 - 611) 设计而成的:

1. 5'-GCGCGAATTCGTNGGNATNGGNATNAA(CT)CA (CT)GG-3'(SEQ ID No:7);

2. 3'-TACAGNTT(GA)AC(GA)GGNGGCCTAGGCG-5' (SEQ ID NO:8);

3. 5'-GCGAATTCACNCCNCA(GA)GTNTT(CT)GA(CT)AC-3' (SEQ ID NO:9);

4. 3'-GGNAA(GA)GGNCCNCT(CT)AA(GA)CCTAGGCG-5' (SEQ ID NO:10);

5. 5'-GCGCGAATTCTGGCA(GA) TCNAC-3' (SEQ ID NO:11);

6. 5'-GCGCGAATTCTGGCA(GA)AGNATG-3' (SEQ ID NO:12);

7. 3'-CGNTACCGNTT(CT)TACAGCCTAGG-5' (SEQ ID NO:13) .

按照厂商说明, 利用 Gene Amp 试剂盒和仪器 (Perkin-Elmer Cetus, Norwalk, CT) 进行 PCR。与厂商说明不同之处在于, 此反应中前 3 个循环是在 28 ℃ 下进行的, 以获得与第一链 cDNA (由来源于 *Coprinus cinereus* IFO8371 菌株的 mRNA 制备而成) 更好的杂交, 随后 PCR 在 65 ℃ 下进行 30 个循环。

对引物进行如下组合: 1 与 2; 3 与 4; 5 与 7; 6 与 7; 1 与 4; 以及 3 与 7。PCR 产物在 5' 末端延伸了一个 EcoRI 位点, 在 3' 末端延伸了一个 BamHI 位点。反应产物在 1 % 琼脂糖 - TBE 凝胶中进行分析, 发现所有反应中均含有预期大小的条带。为了证实这些片段都对应于过氧化物酶特异性序列, 对凝

胶进行 Southern 印迹，并与位于引物 3 和 4 之间的具下述序列的寡核苷酸探针进行杂交：

5'-GT(CT)TC(GA)AT(GA)TAGAA(CT)TG-3' (SEQ ID No:14)。

结果表明此探针能与大约 130bp、420bp、540bp 和 240bp 的条带进行杂交，由此证实观察到的 DNA 条带确实对应于过氧化物酶序列。

各个 PCR 反应来源的 DNA 经 EcoRI 和 BamHI 消化后，克隆至质粒 pUC19(New England Biolabs, Beverly, MA)中。利用上文所述寡核苷酸探针 (SEQ ID No:14)，通过杂交鉴定含有正确 PCR 片段的菌落。按照 Sanger 等所叙述的方法 (1977, 美国国家科学院院报, 74: 5463 - 5476)，通过限制性图谱和部分 DNA 序列分析，对阳性菌落的 DNA 进行分析。得自其中一个克隆的、通过利用引物 1 和 4 所获得的 430bp 片段按下述方法用于筛选 *Coprinus cinereus* cDNA 文库。

按照 Boel 等 (1984, EMBO Journal 3:1097-1102) 和 Chirgwin 等 (1979, 生物化学, 18: 5294 - 5299) 所叙述的方法，在最高过氧化物酶活性的时间点收集 *Coprinus cinereus* IFO8371 菌株的菌丝并进行匀浆，从中提取总 RNA。按照 Aviv 和 Leder (1972, 美国国家科学院院报, 69: 1408 - 1412) 所叙述的方法，在 oligo(dT)-纤维素柱上进行两轮亲和层析获得含有 poly(A)的 RNA。根据厂商说明，利用 cDNA 合成试剂盒 (Invitrogen, San Diego, CA) 合成 cDNA。将 *Coprinus cinereus* cDNA 文库中大约 50,000 个 E.coli 重组子转移至 Whatman 540 滤膜上。按照 Gerger 等 (1979, 核酸研究, 7: 2115 - 2135) 所叙述的方法裂解和固定菌落。在 $0.2 \times \text{SSC} - 0.1\% \text{SDS}$ 中将滤膜与 ^{32}P 标记的 430bp 过氧化物酶基因特异性探针一起杂交。在 65℃ 对滤膜进行杂交和洗涤，随后使用增感屏进行 24 小时放射自显影。放射自显影后，滤膜在逐渐递增的温度下洗涤并与增感屏一起放射自显影 24 小时。通过此

途径，鉴定出 50 多个阳性克隆。按照标准方法（Birnboim 和 Doly, 1979, 核酸研究, 7: 1513 - 1523），从杂交克隆中小量制备质粒 DNA，再通过 Sanger 双脱氧法（Sanger 等, 1977, 美国国家科学院院报, 74: 5463 - 5467）测定这些 cDNA 插入片段的 DNA 序列。从这些菌落中挑选出一个，并将此载体命名为 pCiP。经过 BamHI/XhoI 切割，从载体上切下过氧化物酶 cDNA 片段，通过琼脂糖凝胶电泳纯化此片段，进行电洗脱，并经各种处理以备连接反应之用。将此 cDNA 片段连接至经 BamHI/XhoI 消化的 pHD414 以产生 pJVi9，其中，如图 8 所示，cDNA 是在米曲霉来源的 TAKA 启动子和黑曲霉来源的 AMGTM（Novo Nordisk A/S, Bagsvard, Denmark）终止子的转录调控之下。

从质粒 pJVi9 中以 BamHI/XhoI 片段形式切下编码 *Coprinus cinereus* 过氧化物酶的 cDNA，再克隆至质粒 pJeRS6(图 9)中，形成质粒 pJRoC50（图 10）。质粒 pJRoC50 中含有作为选择性标记的 pyrG，以及 TAKA 启动子和 amdS 终止子。

利用 5 μ g 纯化的质粒 pJRoC50，按下述方法，并作如下变动，以制备米曲霉 HowB425 菌株的转化子。取消琼脂覆盖层，将原生质体直接铺板于基本培养基（Minimal Medium）平板上。转化反应使用浓度为 2×10^7 个/ml 的原生质体。将 100 μ l 原生质体与 5 μ g DNA 一起置于冰上 30 分钟，加入 1ml SPTC(40% PEG4000, 0.8M 山梨醇, 0.05M Tris pH8.0, 0.05M CaCl₂)，将原生质体置于 34℃ 保温 20 分钟。将转化产物直接铺板于含有基本培养基的平板上。每升所述基本培养基（pH6.5）含有下述组分：6g NaNO₃, 0.52g KCl, 1.52g KH₂PO₄, 1ml 微量金属溶液, 1g 葡萄糖, 500mg MgSO₄ - 7H₂O, 342.3g 蔗糖和 20g Noble 琼脂。每升微量金属溶液（1000x）含有下述组分：22g ZnSO₄-7H₂O, 11g H₃BO₃, 5g MnCl₂-4H₂O, 5g Fe SO₄-7H₂O, 1.6g CoCl₂-5H₂O, 1.6g (NH₄)₆Mo₇O₂₄ 以及 50g

Na₄EDTA。将平板置于 34 ℃ 保温 5 - 7 天, 再将转化子转移至含有同样培养基的平板上, 在 37 ℃ 继续保温 3 - 5 天。

按下述酶测定方法测定 66 个转化子中的过氧化物酶活性: 在按 1: 900 稀释的培养上清液 20μl 中加入 180μl 底物缓冲液{20ml 0.1M 磷酸钾 - 0.01 % 吐温 80 pH7.0, 250μl 2, 2'-azinobis (3-乙基苯酰噻唑啉 - 6 - 磺酸酯) (ABTS) 溶液 (22mg/ml) 及 2μl 30 % 过氧化氢}, 利用 Molecular Devices 的 Thermomax 微量板读数仪 (Molecular Devices, Sunnyvale, CA) 迅速在 25 ℃ 测定 405nm 处的光吸收值。在 2 分钟以上的时段内不断混合, 并每隔 10 秒钟记录一次测定值, 再利用 SOFTmax 程序 (Molecular Devices, Sunnyvale, CA) 计算 Vmax 值。以已知量的灰盖鬼伞过氧化物酶为标准绘制标准曲线, 利用此标准曲线估计每毫升的过氧化物酶单位 (POXU)。1POXU 定义为在 30 ℃ 下, 每分钟催化 0.88mM H₂O₂、1.67 mM ABTS、0.1M 磷酸盐 pH7.0 发生 1.0μ mole 转变所需的酶量。利用上述同样的平板, 在相同的条件下, 通过孢子划线, 挑选独立菌落, 从而对 4 个具最高表达水平的转化子进行孢子纯化。

在摇瓶中, 将每种转化子约 5×10^6 个孢子接种于 25ml MY25 培养基中, 进行最终分析。MY25 培养基含有下述组分: 1 % 酵母提取物, 2.5 % 麦芽糖, 0.2 % 尿素, 以及 1 × MY 盐溶液, pH6.5。每升 1 × MY 盐溶液含有下述组分: 2g MgSO₄-7H₂O, 2g K₂HPO₄、10g KH₂PO₄, 2g 柠檬酸, 0.5ml 微量金属溶液, 以及 1ml 10 % CaCl₂-2H₂O。每升微量金属溶液含有下述组分: 13.9g FeSO₄-7H₂O, 8.5g MnSO₄-H₂O, 14.28g ZnSO₄-7H₂O, 1.63g CuSO₄, 0.24g NiCl₂-6H₂O, 以及 3.0g 柠檬酸。将在 50mM NaOH 中新鲜制备的氯高铁血红素 (hemin) 10mg/ml 储存液加入摇瓶中至终浓度 0.01mg/ml。将摇瓶于 34 ℃、200rpm 保温 7 - 8 天。将能生产最大量过氧化物酶的转化子命名为 JRo C50.3.18A。

实施例 8：利用 pSE31 转化米曲霉 JRoC50.3.18A 菌株

利用 pSE31 转化米曲霉 JRoC50.3.18A 菌株以测定 hema 基因的过量表达是否会提高过氧化物酶的生产水平。

转化反应使用浓度为 2×10^7 个/ml 的原生质体。在 34 °C，将 100 µl 原生质体与 10 µg DNA 和 200 µl 60 % PEG4000 - 10mM HEPES - 10mM CaCl_2 溶液一起保温 30 分钟。加入 3ml SPTC (40 % PEG4000 , 0.8M 山梨醇, 0.05M Tris pH8.0, 0.05M CaCl_2) , 立即将原生质体平铺于 COVE 转化板 (每升含有: 0.52g KCl, 0.52g $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 1.52g KH_2PO_4 , 1ml 实施例 7 所述的微量金属溶液, 342.3g 蔗糖, 25g Noble 琼脂, 10ml 1M 乙酰胺, 以及 10ml 3M CsCl) 中以进行 amdS 转化。将平板于 34 °C 保温 5 - 7 天。再将转化子转移至含有同样培养基的平板上, 于 34 °C 保温 3 - 5 天。然后利用这些板, 在同样的条件下通过孢子划线、挑出独立菌落从而对转化子进行纯化。

实施例 9：hema 转化子中过氧化物酶的生产

在 24 孔微滴定板的各个孔中以约 1×10^5 个孢子/孔的接种量分别接种实施例 8 来源的转化子, 这些孔中含有 1ml 四分之一强度的 MY25 培养基和 $1 \times \text{MY}$ 盐 (参见实施例 1) , 所述四分之一强度的 MY25 培养基中含有 0.25 % 酵母提取物, 0.63 % 麦芽糖及 0.05 % 尿素, pH6.5 。在湿润摇床中, 于 34 °C、100rpm 将微滴定板保温 5 天。

利用实施例 7 所述的酶测定法测定过氧化物酶生产水平。微滴定板测试结果表明, hema 转化子的平均 POXU/ml 比仅转化载体的载化子平均水平高 1.4 倍, 而最好的 hema 转化子的过氧化物酶生产量提高了 1.6 倍。

小部分 hema 转化子 (39 %) 与大部分仅含载体的对照具有相似的过氧化物酶水平。利用按实施例 1 所述方法从各个转化子中分离的基因组 DNA 50ng , 按实施例 2 所述方法进行 PCR 。与实施例 2 所述方法不同之处在于, 此处使用引物 hema

3' (见实施例 4) 和引物 5'-TCTCTTCCTTCCTGAATCCTC-3' (SEQ ID NO: 15)。此分析结果表明 hemA 转化子含有表达盒。

在摇瓶中培养以上获得的 11 个最好的 hemA 转化子, 以便更好地分析对过氧化物酶生产的影响。为了进行摇瓶分析, 在 25ml MY25 培养基中接种约 5×10^6 个孢子的各个转化子。所述 MY25 培养基中含有 1% 酵母提取物, 2.5% 麦芽糖, 0.2% 尿素以及 $1 \times$ MY 盐溶液 pH6.5 (见实施例 7)。将摇瓶在 34℃、200rpm 保温 7 - 8 天。按照上述方法进行过氧化物酶分析。

结果表明, 5 个转化子: SE01 - 15、SE01 - 20、SE01 - 26、SE01 - 28 和 SE01 - 32 产生的过氧化物酶水平高于仅含载体的对照菌株, 其中 3 个转化子表达的过氧化物酶水平比对照的过氧化物酶平均水平高 1.9 倍。剩下的 6 个 hemA 转化子的过氧化物酶水平与对照相当。

以高水平麦芽糖浆作碳源, 利用标准的分批补料方法, 在 2 升发酵液中培养转化子 SE01 - 28 和对照菌株 SE05 - 18 (单独 pBANe6 载体的转化子)。在批料和补料中添加 FeCl_3 至约 0.4mM。在两种培养物中维持正溶解氧压, 并在第 3 - 8 天以每小时约 2g 糖化物/升的速率添加补料。在第 2 和 3 天后以逐步的方式达到此水平。发酵期间两种培养物的生物量大致相等。

据观察, SE01 - 28 的过氧化物酶活性化对照菌株 SE05 - 18 高 2 倍, SE01 - 28 的多肽水平也比对照菌株 SE05 - 18 高 2 倍。

所有结果表明 hemA 基因的过量表达能提高过氧化物酶产量约 2 倍, 这些数据进一步说明 hemA 可能在丝状真菌的血红素生物合成中代表一个关键的调控点, 在缺乏氯高铁血红素补给时, hemA 在遗传调控的水平上能提高血红素蛋白的产量。

如下菌株已按照布达佩斯条约保藏于农业研究机构专利培

养物保藏中心 (NRRL) , 北方地区研究实验室, 1815
University Street, Peoria, Illinois 61604, USA 。

菌株	保藏号	保藏日期
E.coli DH5 α (pSE17)	NRRL B-21563	1996 年 4 月 22 日

此菌株已在一定条件下被保藏, 能确保在此专利申请未决期间由专利和商标官员决定根据 37C.F.R. § 1.14 和 35U.S.C. § 122 授予可获得所述菌株之权利的人可得到所述菌株培养物。保藏物代表每种保藏菌株的基本上纯的培养物。当在外国提交本申请的副本或其子申请时, 如外国专利法需要, 也可得到该保藏物。然而, 应理解保藏物的可获得性在由政府行为授予的专利权的部分废除时不构成对操作本发明的许可。

本文所述和要求权利的发明并不局限于本文所公开的具体实施方案的范围, 因为这些实施方案只想阐明本发明的几个方面。任何等价的实施方案都在本发明的范围之内。实际上, 根据前文的描述, 除了本文所示和所述的以外, 本发明的多种修饰对于本领域技术人员而言都是显而易见的, 这种修饰也欲包括在所附权利要求书的范围内。

本文提及很多参考文献, 它们所公开内容都全文列入本文作为参考。

序列表

(1) 一般资料

(i) 申请人:

(A) 名称: Novo Nordisk Biotech, Inc.

(B) 街道: 1445 Drew Avenue

(C) 城市: Davis, California

(D) 国家: 美国

(E) 邮编代码 (邮编): 95616 - 4880

(F) 电话: (916) 757 - 8100

(G) 传真: (916) 758 - 0317

(ii) 发明名称: 米曲霉 5 - 氨基乙酰丙酸合成酶及编码此酶的核酸

(iii) 序列数: 18

(iv) 联系地址:

(A) 联系人: Novo Nordisk of North America, Inc.

(B) 街道: 405 Lexington Avenue, Suite 6400

(C) 城市: 纽约

(D) 州: 纽约

(E) 国家: 美国

(F) 邮编: 10174

(v) 计算机可读形式:

(A) 介质类型: 软盘

(B) 计算机: IBM 兼容机

(C) 操作系统: DOS

(D) 软件: Windows 2.0 版的 FastSEQ

(vii) 在先申请资料:

(A) 申请号: 60/019,399

(B) 申请日: 1996 年 6 月 10 日

(C) 分类号:

(viii) 代理人/代理机构资料:

(A) 姓名: Lambiris, Elias J.

(B) 登记号: 33, 728

(C) 参考文献/文档号: 4808.204-WO

(ix) 电信资料:

(A) 电话: 212 - 878 - 9652

(B) 传真: 212 - 878 - 9655

(C) 电传:

(2) SEQ ID No:1 的资料:

(i) 序列特征:

(A) 长度: 4157 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(xi) 序列描述: SEQ ID No: 1:

```

ACCATTGACT CTCAAGCTAT GGATCGTGCT CACCGTCTCG GCCAGACAAG ACAGGTCACG      60
GTGTATCGCC TGATTACTCG CGGCACCATT GAGGAGCGTA TTCGCAAGCG AGCTTTGCAG      120
AAGGAGGAAG TGCAGCGTGT CGTCATCTCA GGTGGCGCAG CTGGTGGGGT TGA CTTC AAT      180
ACTCGCAACC GCGAGAGCCG AACCAAGGAC ATCGCCATGT GGCTGGCAGA TGATGAACAG      240
GCGGAGCTTA TTGAGCAAAA GGAGAAGGAA GCGCTGGACC GAGGCGAAGT GTTTGGCGCT      300
AGTAAAGGCG GGAAGAAGGC TGCTCAGAAG AGAAAGAGAG ATATCACGCT GGATGATATG      360
TATCATGAAG GTATGTGAAT CTGATCAAAG CTCTTCGTTC CGGGGAGGCT TCTGGAAATA      420
GTACTAACCG CGTCAATCTA TAGGCGAAGG GAACTTTGAC GATGCCAGTG CAAAGCCATC      480
AGGAGCGGCC ACTCCTGTGT CGACTGCAGA GAATTTAGGC ACCCCATCCT CCACGCCAGT      540
TCCTAAACGA GGACGTGGAA GGGGGACAGG AAAGGGCACG TCTAAAAGAG CCAAAACTAC      600
CAAGGAGAGA TTACGTCTCA TTGATGGCGA CGGAGGCTTA GGGCCTAGTT GATTTAATCG      660
ATCTGTGCCT CAATAATGGA CACGGCTGGT TATGGTCATG GCGTTCAGAG ATTGCATTTT      720
TTTCCACCCC TTTATCTTTC TTTCTTTCCT CTAAACCCC TCTTTTTTGT TTTTCTTTT      780
ATCGGACTTT ACTTGTGGGC AGCTTACGTT CTGCCTTGTA TTAACAGCAT ATATTCCTGA      840
TTCCTGATGT ACGAAGCGAT TTAAGAGTCA TTGAAGACGA AGGATGAAAC CCGTGGTAAT      900
CAGCCGATAA TGGCAAAGAG AAGGAGAAGA AAAAAATCAA GTGCGAGTTT TGAAATTGAT      960
GGCAAGATAG ACATTGTATC CTGTACCTGT TCTTGGGCTG TGACGGGGGG GGTGAAATTG      1020
ACGGTCATCA CCCGGCTATT ATTACTATTG TTGTACTGTA CATCCGATC CTGCTGGTCT      1080
GTATCTAGTT AGGGCAATAT TCCCCGTCGC CAGGCCTCTT GGGTTATGAA TGATTCATA      1140
GGTGAAGTTT CGTATCCGTA CGCACCAGAG GATTTCTTAG TATTACTTGT ATTATGAAAA      1200
TGCACTTGCC GAGTTAAGTC CGCCGGCCAA TCACGGCGGA GGATATGGTA AGCCGAAAAG      1260
TCTCGCCGAA TCCCCGACT TACTCTTACT GGAAGTGGCT TAGTGCCCTC AGCGCCCCCT      1320
CGCCCTCAGT CCATCAGCCA GATTGACTCT TATTTCTCTC TCCTCTTCGC CGCGGGTGAC      1380
ATATCCCTCT CCTTCTCCCT CTCCCTCTTG ACAACATTTT ATCTTCGCTT CCTTTTGTGA      1440
TATAGTCAGT TTCGCTATCC ATTGAAGCAT CACTCATGGA GTCTCTTCTC CAGCAGTCCC      1500
GGGCGATGTG CCCGTTCCCT AAGCGCACAT CTCCATCTTC TCTGCGTACG CTGGCAACCG      1560
CGACTCGACC TAGCACTAGT TCCGGTGGAG GCACATATGTC TAATCTCCAG GTCATTGCCC      1620
GTCTGTGCC TAGCATGAGC AAGGCTCTGG CCGTGAGAG CGCTCGCATG GCCGGTACCA      1680
AAAGATTAC CTCATGTGCT GCCGCATCA CCGGTCTCGG CAACAAGCAT TGCCGTGCTC      1740
CTACTGGGAA GAGAACCCTG CACTCCACCT CCGGTAACGG CGCCAATGTG AGCGCAGAGA      1800
TCTACAAGAA CACCCAGCGA GATCCCGCCG GTTTCTCGAA GATCAAGACC CCTGCCAATG      1860

```

CTACCGCCGC	TGCCGCTACG	TCTGGCCCTC	GTCCAGAGGC	TCCCGTGGCG	AAGCCTTTCA	1920
ACTACAATTC	TTTCTACAAC	ACCGAATTGG	AAAAGAAACA	CAAGGACAAG	TCGTATCGCT	1980
ATTTCAACAA	CATCAATCGT	CTCGCTCAGG	AGTTTCCCCG	GGCTCACACC	ACATCTGCCG	2040
AGGAACGTGT	GACGGTCTGG	TGCTCGAACG	ATTATCTCGG	CATGGGCCGC	AACCCCGAGG	2100
TTCTGGCCAC	CATGCATAAG	ACATTGGACA	CCTACGGAGC	CGGTGCGGGA	GGTACTCGCA	2160
ACATTTTCAGG	TCACAATCAA	CATGCCGTGA	GCCTGGAGAA	CACCCTGGCC	AAATTGCACG	2220
GCAAGGAGGC	GGCATTAGTC	TTCAGCTCAT	GCTTCGTGGC	TAACGATGCC	ACCCTCGCAA	2280
CCCTGGGTAG	CAAGTTGCC	GACTGTGTTA	TTCTGTCCGA	TAGCCTGAAT	CATGCATCGA	2340
TGATTTCAGGG	TATTCGCCAT	TCAGGCGCCA	AGAAAATGGT	TTTCAAGCAT	AATGATCTGG	2400
TCGACCTTGA	GGCCAAGTTG	GCAGCTCTAC	CTCTTCATGT	CCCCAAGATT	ATTGCATTCTG	2460
AATCAGTTTA	TAGCATGTGC	GGATCTATTG	CCCCAATTGA	GAAGATCTGT	GATCTTGCAG	2520
ACAAGTACGG	TGCCATTACT	TTCCTGGATG	AAGTCCACGC	TGTGGGAATG	TACGGACCTC	2580
ACGGAGCAGG	TGTGGCAGAG	CACCTTGA	ATGACATCTA	TGCTTCCCAA	GATACGGTCA	2640
ACCCGCGCAG	TACTAAGGGA	ACCGTGATGG	ACCGAATCGA	TATTATCACC	GGTACTCTGG	2700
GCAAGGCCTA	CGGATGTGTC	GGGGGCTACA	TTGCTGGATC	CGCTGCGATG	GTTGACACCA	2760
TCCGCTCCCT	CGCCCCTGGC	TTCATCTTCA	CCACGTCCTT	GCCGCCCCGCC	ACCATGGCTG	2820
GTGCAGACAC	TGCTATCCAG	TACCAGGCTC	GTCACCGAGG	CGACCGCGTC	CTGCAGCAGT	2880
TGCACACCCG	CGCGGTCAAA	GCAGCTTTCA	AGGAGTTGGA	TATTCTGTGA	ATTCCCAACC	2940
CCTCCCATAT	CATTCCGCTC	CTGGTTGGGG	ATGCCGAGGT	TGCTAAGAAG	GCCTCGGACA	3000
AGCTTCTGGA	GGAGCATGGA	ATTTATGTAC	AAGCCATCAA	CTACCCAACC	GTGCCTCGGG	3060
TGTAAGAGCG	GCTTCGTATC	ACGCCCCACC	CGGGACATAT	CAAGGAGCAC	CGCGACCACC	3120
TGGTGCAAGC	CGTCCAAACA	GTCTGGAACG	AACTGGGCAT	CAAACGCACC	AGCGATTGGG	3180
AAGCGCAAGG	CGGCTTCGTC	GGCGTGGGTG	TGCATGGCGC	CGAGGCTGAG	AACCAGCCGA	3240
TTTGGAATGA	TGTGCAGCTG	GGGCTGAAGG	AAAACGAAGC	CATTGAGGCT	GCTGTGGAAC	3300
GCGAGTTTGC	CGAGGCCCCC	ATGCGGACCG	CCACCCGTCC	TGCCGCGGCT	GCTGCTTCGT	3360
CAATCCCGGT	GGGTGTGGCT	GCCTGAAGTG	GCTGCCCGCA	TGTGAGCTGA	AATCGACGTG	3420
GAATTCTATA	CACACACACA	CACACACACA	CACACACACA	CACACACACA	CACACACACA	3480
CACACACACA	CACACACACT	AACACACACT	ATGTTATAAA	TTCCACATCC	ACTCCTTTGT	3540
CCCTTGTTGG	ACGTAATTGG	TATTTGGACT	ATTAGTTAGA	ACCAGTCAGT	CGTTACCATG	3600
TGTTTTCGGTT	CGACTCGAAA	TCTGACATGT	TGTCTGCCCC	CATGCCACTT	CATCTCCTCC	3660
GTAACCGCAG	GGCTTCAAAT	ACACTGCCCA	GTAATTGTAG	TCAATATAGC	AGTTAACTAA	3720
CCTTCACCAA	TTTCCTAATA	ACAATAGAAG	GGGCCATACA	CGCAGTACCA	AAGATCACCT	3780
ACCTCCGATC	AATATCCGAA	CCTCAGGCTA	CATACATCAA	GTCGCATTAA	TCGATTCCGA	3840
CCTCTGTTTA	TCCCTGAAAA	TAACTAAGAT	CATGATCTAC	GTTTGGTAAG	TGGGACACCT	3900
ACCTACACTG	GGAGGTATTG	AATAAAGGCA	TCATTCA'TAT	AGTCACAAGA	TGCCAGGGCC	3960
AATTCATGAT	ATGGATAGCT	ACTTCCAAAC	ATAATTGAGA	GGTATCATTG	TGCTCTTCAG	4020
ACAGTTCTTC	TCGAAGATCA	GTAGGAGCCA	GTTTTGACCA	TTAACTTGTA	ATGTAATTGC	4080
GATTGTAGTA	GATCCGAGAT	CCATTCACCT	TCTAAGGGTT	AATTGATTCA	TTTTACTGAT	4140
ACCTCACCCA	CCATATT					4157

(2) SEQ ID No:2 的资料:

(i) 序列特征:

(A) 长度: 636 个碱基对

(B) 类型: 氨基酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(ii) 分子类型: 蛋白质

(v) 片断类型: 内部的

(xi) 序列描述: SEQ ID No:2

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

Gly	Ala	Asp	Thr	Ala	Ile	Gln	Tyr	Gln	Ala	Arg	His	Gln	Gly	Asp	Arg
	450					455					460				
Val	Leu	Gln	Gln	Leu	His	Thr	Arg	Ala	Val	Lys	Ala	Ala	Phe	Lys	Glu
465					470					475					480
Leu	Asp	Ile	Pro	Val	Ile	Pro	Asn	Pro	Ser	His	Ile	Ile	Pro	Leu	Leu
				485					490					495	
Val	Gly	Asp	Ala	Glu	Val	Ala	Lys	Lys	Ala	Ser	Asp	Lys	Leu	Leu	Glu
			500					505					510		
Glu	His	Gly	Ile	Tyr	Val	Gln	Ala	Ile	Asn	Tyr	Pro	Thr	Val	Pro	Arg
		515					520					525			
Gly	Glu	Glu	Arg	Leu	Arg	Ile	Thr	Pro	Thr	Pro	Gly	His	Ile	Lys	Glu
530						535					540				
His	Arg	Asp	His	Leu	Val	Gln	Ala	Val	Gln	Thr	Val	Trp	Asn	Glu	Leu
545					550					555					560
Gly	Ile	Lys	Arg	Thr	Ser	Asp	Trp	Glu	Ala	Gln	Gly	Gly	Phe	Val	Gly
				565					570					575	
Val	Gly	Val	Asp	Gly	Ala	Glu	Ala	Glu	Asn	Gln	Pro	Ile	Trp	Asn	Asp
			580					585					590		
Val	Gln	Leu	Gly	Leu	Lys	Glu	Asn	Glu	Ala	Ile	Glu	Ala	Ala	Val	Glu
		595				600						605			
Arg	Glu	Phe	Ala	Glu	Ala	Pro	Met	Arg	Thr	Ala	Thr	Arg	Pro	Ala	Ala
610						615					620				
Ala	Ala	Ala	Ser	Ser	Ile	Pro	Val	Gly	Val	Ala	Ala				
625					630					635					

(2) SEQ ID No:3 的资料:

(i) 序列特征:

(A) 长度: 31 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(xi) 序列描述: SEQ ID No:3 :

TTTATGATGG AGGCCCTTCT CCAGCAGTCT C 31

(2) SEQ ID No:4 的资料:

(i) 序列特征:

(A) 长度: 29 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(xi) 序列描述: SEQ ID No:4 :

CTATGCATTT AAGCAGCAGC CGCGACTGG 29

(2) SEQ ID No:5 的资料:

(i) 序列特征:

(A) 长度: 27 个碱基对

(B) 类型： 核酸

(C) 链型： 单链

(D) 拓扑结构： 线性

(xi) 序列描述： SEQ ID No:5 :

TCATTTAAAT GATGGAGTCT CTTCTCC 27

(2) SEQ ID No:6 的资料:

(i) 序列特征:

(A) 长度： 26 个碱基对

(B) 类型： 核酸

(C) 链型： 单链

(D) 拓扑结构： 线性

(xi) 序列描述： SEQ ID No:6 :

TCTTAATTAA TCAGCTCACA TGCGGG 26

(2) SEQ ID No:7 的资料:

(i) 序列特征:

(A) 长度： 33 个碱基对

(B) 类型： 核酸

(C) 链型： 单链

(D) 拓扑结构： 线性

(xi) 序列描述： SEQ ID No:7 :

GCGCGAATTC GTNGGNATNG GNATNAAYCA YGG 33

(2) SEQ ID No:8 的资料:

(i) 序列特征:

(A) 长度： 25 个碱基对

(B) 类型： 核酸

(C) 链型： 单链

(D) 拓扑结构： 线性

(xi) 序列描述： SEQ ID No:8 :

GCGGATCCGG NGGRCARTTN GACAT 25

(2) SEQ ID No:9 的资料:

(i) 序列特征:

(A) 长度: 28 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(xi) 序列描述: SEQ ID No:9 :

GCGAATTCAC NCCNCARGTN TTYGAYAC 28

(2) SEQ ID No:10 的资料:

(i) 序列特征:

(A) 长度: 26 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(xi) 序列描述: SEQ ID No:10 :

GCGGATCCRA AYTCNCCNGG RAANGG 26

(2) SEQ ID No:11 的资料:

(i) 序列特征:

(A) 长度: 21 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(xi) 序列描述: SEQ ID No:11

GCGCGAATTC TGGCARTCNA C 21

(2) SEQ ID No:12 的资料:

(i) 序列特征:

(A) 长度: 22 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(xi) 序列描述: SEQ ID No:12 :

GCGCGAATTC TGGCARAGNA TG	22
(2) SEQ ID No:13 的资料:	
(i) 序列特征:	
(A) 长度: 23 个碱基对	
(B) 类型: 核酸	
(C) 链型: 单链	
(D) 拓扑结构: 线性	
(xi) 序列描述: SEQ ID No:13 :	
GGATCCGACA TYTTNGCCAT NGC	23
(2) SEQ ID No:14 的资料:	
(i) 序列特征:	
(A) 长度: 17 个碱基对	
(B) 类型: 核酸	
(C) 链型: 单链	
(D) 拓扑结构: 线性	
(xi) 序列描述: SEQ ID No:14 :	
GTYTCRATRT AGAAYTG	17
(2) SEQ ID No:15 的资料:	
(i) 序列特征:	
(A) 长度: 21 个碱基对	
(B) 类型: 核酸	
(C) 链型: 单链	
(D) 拓扑结构: 线性	
(xi) 序列描述: SEQ ID No:15 :	
TCTCTTCCTT CCTGAATCCT C	21
(2) SEQ ID No:16 的资料:	
(i) 序列特征:	
(A) 长度: 649 个氨基酸	
(B) 类型: 氨基酸	
(C) 链型: 单链	

(D) 拓扑结构：线性

(ii) 分子类型：无

(xi) 序列描述：SEQ ID No:16：

```

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

```

```

Leu Glu Glu His Gly Ile Tyr Val Gln Ala Ile Asn Tyr Pro Thr Val
      515                      520                      525
Pro Arg Gly Glu Glu Arg Leu Arg Ile Thr Pro Thr Pro Gly His Thr
      530                      535                      540
Gln Glu Leu Arg Asp His Leu Val Glu Ala Val Asn Thr Val Trp Asn
545                      550                      555                      560
Asp Leu Gly Ile Lys Arg Ala Ser Asp Trp Lys Ala Met Gly Gly Phe
      565                      570                      575
Val Gly Val Gly Val Glu Ala Ala Glu Leu Glu Asn Gln Pro Ile Trp
      580                      585                      590
Thr Asp Ala Gln Leu Asn Met Arg Pro Asp Glu Thr Leu Glu Ala Ala
      595                      600                      605
Val Glu Arg Glu Phe Gln Ala Ala Val Pro Gly Met Lys Ala Gly Gly
      610                      615                      620
Ala Lys Ala Lys Pro Val Gly Ser Ile Ala Ala Asn Pro Ile Gly Ala
625                      630                      635                      640
Ser Ile Pro Val Ala Ala Ala Ala Glx
      645

```

(2) SEQ ID No:17 的资料:

(i) 序列特征:

(A) 长度: 548 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(ii) 分子类型: 无

(xi) 序列描述: SEQ ID No:17:

```

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

```

```

Ile Phe Lys His Asn Asp Leu Asn Glu Leu Glu Gln Leu Leu Gln Ser
225                230                235                240
Tyr Pro Lys Ser Val Pro Lys Leu Ile Ala Phe Glu Ser Val Tyr Ser
                245                250                255
Met Ala Gly Ser Val Ala Asp Ile Glu Lys Ile Cys Asp Leu Ala Asp
                260                265                270
Lys Tyr Gly Ala Leu Thr Phe Leu Asp Glu Val His Ala Val Gly Leu
                275                280                285
Tyr Gly Pro His Gly Ala Gly Val Ala Glu His Cys Asp Phe Glu Ser
290                295                300
His Arg Ala Ser Gly Ile Ala Thr Pro Lys Thr Asn Asp Lys Gly Gly
305                310                315                320
Ala Lys Thr Val Met Asp Arg Val Asp Met Ile Thr Gly Thr Leu Gly
                325                330                335
Lys Ser Phe Gly Ser Val Gly Gly Tyr Val Ala Ala Ser Arg Lys Leu
                340                345                350
Ile Asp Trp Phe Arg Ser Phe Ala Pro Gly Phe Ile Phe Thr Thr Thr
                355                360                365
Leu Pro Pro Ser Val Met Ala Gly Ala Thr Ala Ala Ile Arg Tyr Gln
370                375                380
Arg Cys His Ile Asp Leu Arg Thr Ser Gln Gln Lys His Thr Met Tyr
385                390                395                400
Val Lys Lys Ala Phe His Glu Leu Gly Ile Pro Val Ile Pro Asn Pro
                405                410                415
Ser His Ile Val Pro Val Leu Ile Gly Asn Ala Asp Leu Ala Lys Gln
                420                425                430
Ala Ser Asp Ile Leu Ile Asn Lys His Gln Ile Tyr Val Gln Ala Ile
                435                440                445
Asn Phe Pro Thr Val Ala Arg Gly Thr Glu Arg Leu Arg Ile Thr Pro
                450                455                460
Thr Pro Gly His Thr Asn Asp Leu Ser Asp Ile Leu Ile Asn Ala Val
465                470                475                480
Asp Asp Val Phe Asn Glu Leu Gln Leu Pro Arg Val Arg Asp Trp Glu
                485                490                495
Ser Gln Gly Gly Leu Leu Gly Val Gly Glu Ser Gly Phe Val Glu Glu
                500                505                510
Ser Asn Leu Trp Thr Ser Ser Gln Leu Ser Leu Thr Asn Asp Asp Leu
                515                520                525
Asn Pro Asn Val Arg Asp Pro Ile Val Lys Gln Leu Glu Val Ser Ser
530                535                540
Gly Ile Lys Gln
545

```

(2) SEQ ID No:18 的资料:

(i) 序列特征:

(A) 长度: 587 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(ii) 分子类型: 无

(xi) 序列描述: SEQ ID No:18:

```

Met Val Thr Ala Ala Met Leu Leu Gln Cys Cys Pro Val Leu Ala Arg
1          5          10          15
Gly Pro Thr Ser Leu Leu Gly Lys Val Val Lys Thr His Gln Phe Leu
20          25          30
Phe Gly Ile Gly Arg Cys Pro Ile Leu Ala Thr Gln Gly Pro Asn Cys
35          40          45
Ser Gln Ile His Leu Lys Ala Thr Lys Ala Gly Gly Asp Ser Pro Ser
50          55          60

```

Trp 65	Ala	Lys	Gly	His	Cys 70	Pro	Phe	Met	Leu	Ser 75	Glu	Leu	Gln	Asp	Gly 80
Lys	Ser	Lys	Ile	Val	Gln	Lys	Ala	Ala	Pro	Glu	Val	Gln	Glu	Asp	Val
			85						90					95	
Lys	Ala	Phe	Lys	Thr	Asp	Leu	Pro	Ser	105	Leu	Val	Ser	Val	Ser	Leu
			100										110		
Arg	Lys	Pro	Phe	Ser	Gly	Pro	Gln	Glu	Gln	Glu	Gln	Ile	Ser	Gly	Lys
		115					120					125			
Val	Thr	His	Leu	Ile	Gln	Asn	Asn	Met	Pro	Gly	Asn	Tyr	Val	Phe	Ser
						135					140				
Tyr	Asp	Gln	Phe	Phe	Arg	Asp	Lys	Ile	Met	Glu	Lys	Lys	Gln	Asp	His
145					150					155					160
Thr	Tyr	Arg	Val	Phe	Lys	Thr	Val	Asn	Arg	Trp	Ala	Asp	Ala	Tyr	Pro
				165					170					175	
Phe	Ala	Gln	His	Phe	Phe	Glu	Ala	Ser	Val	Ala	Ser	Lys	Asp	Val	Ser
			180					185					190		
Val	Trp	Cys	Ser	Asn	Asp	Tyr	Leu	Gly	Met	Ser	Arg	His	Pro	Gln	Val
		195					200					205			
Leu	Gln	Ala	Thr	Gln	Glu	Thr	Leu	Gln	Arg	His	Gly	Ala	Gly	Ala	Gly
		210				215					220				
Gly	Thr	Arg	Asn	Ile	Ser	Gly	Thr	Ser	Lys	Phe	His	Val	Glu	Leu	Glu
225					230					235					240
Gln	Glu	Leu	Ala	Glu	Leu	His	Gln	Lys	Asp	Ser	Ala	Leu	Leu	Phe	Ser
				245					250					255	
Ser	Cys	Phe	Val	Ala	Asn	Asp	Ser	Thr	Leu	Phe	Thr	Leu	Ala	Lys	Ile
			260					265					270		
Leu	Pro	Gly	Cys	Glu	Ile	Tyr	Ser	Asp	Ala	Gly	Asn	His	Ala	Ser	Met
		275					280					285			
Ile	Gln	Gly	Ile	Arg	Asn	Ser	Gly	Ala	Ala	Lys	Phe	Val	Phe	Arg	His
		290			295						300				
Asn	Asp	Pro	Asp	His	Leu	Lys	Lys	Leu	Leu	Glu	Lys	Ser	Asn	Pro	Lys
305					310					315					320
Ile	Pro	Lys	Ile	Val	Ala	Phe	Glu	Thr	Val	His	Ser	Met	Asp	Gly	Ala
			325						330					335	
Ile	Cys	Pro	Leu	Glu	Glu	Leu	Cys	Asp	Val	Ser	His	Gln	Tyr	Gly	Ala
			340					345					350		
Leu	Thr	Phe	Val	Asp	Glu	Val	His	Ala	Val	Gly	Leu	Tyr	Gly	Ser	Arg
		355					360					365			
Gly	Ala	Gly	Ile	Gly	Glu	Arg	Asp	Gly	Ile	Met	His	Lys	Ile	Asp	Ile
		370				375					380				
Ile	Ser	Gly	Thr	Leu	Gly	Lys	Ala	Phe	Gly	Cys	Val	Gly	Gly	Tyr	Ile
385					390					395					400
Ala	Ser	Thr	Arg	Asp	Leu	Val	Asp	Met	Val	Arg	Ser	Tyr	Ala	Ala	Gly
				405					410					415	
Phe	Ile	Phe	Thr	Ser	Leu	Pro	Pro	Met	Val	Leu	Ser	Gly	Ala	Leu	
			420				425					430			
Glu	Ser	Val	Arg	Leu	Leu	Lys	Gly	Glu	Glu	Gly	Gln	Ala	Leu	Arg	Arg
		435					440					445			
Ala	His	Gln	Arg	Asn	Val	Lys	His	Met	Arg	Gln	Leu	Leu	Met	Asp	Arg
		450				455					460				
Gly	Leu	Pro	Val	Ile	Pro	Cys	Pro	Ser	His	Ile	Ile	Pro	Ile	Arg	Val
465					470					475					480
Gly	Asn	Ala	Ala	Leu	Asn	Ser	Lys	Leu	Cys	Asp	Leu	Leu	Leu	Ser	Lys
				485					490					495	
His	Gly	Ile	Tyr	Val	Gln	Ala	Ile	Asn	Tyr	Pro	Thr	Val	Pro	Arg	Gly
			500					505					510		
Glu	Glu	Leu	Leu	Arg	Leu	Ala	Pro	Ser	Pro	His	His	Ser	Pro	Gln	Met
		515					520					525			
Met	Glu	Asp	Phe	Val	Glu	Lys	Leu	Leu	Leu	Ala	Trp	Thr	Ala	Val	Gly
		530				535					540				
Leu	Pro	Leu	Gln	Asp	Val	Ser	Val	Ala	Ala	Cys	Asn	Phe	Cys	Arg	Arg
545					550					555					560
Pro	Val	His	Phe	Glu	Leu	Met	Ser	Glu	Trp	Glu	Arg	Ser	Tyr	Phe	Gly
				565					570					575	
Asn	Met	Gly	Pro	Gln	Tyr	Val	Thr	Thr	Tyr	Ala					
			580						585						

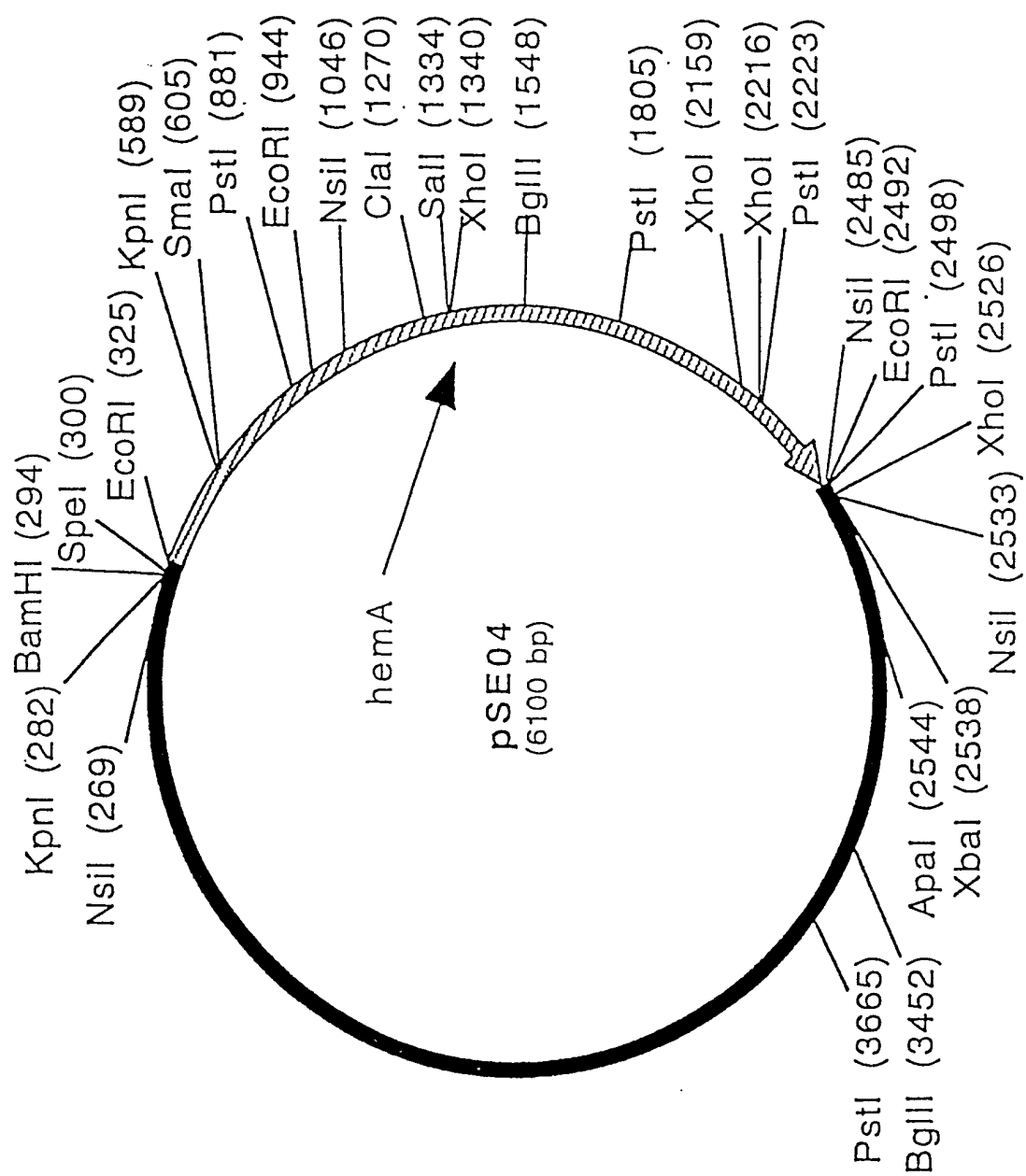


图 1

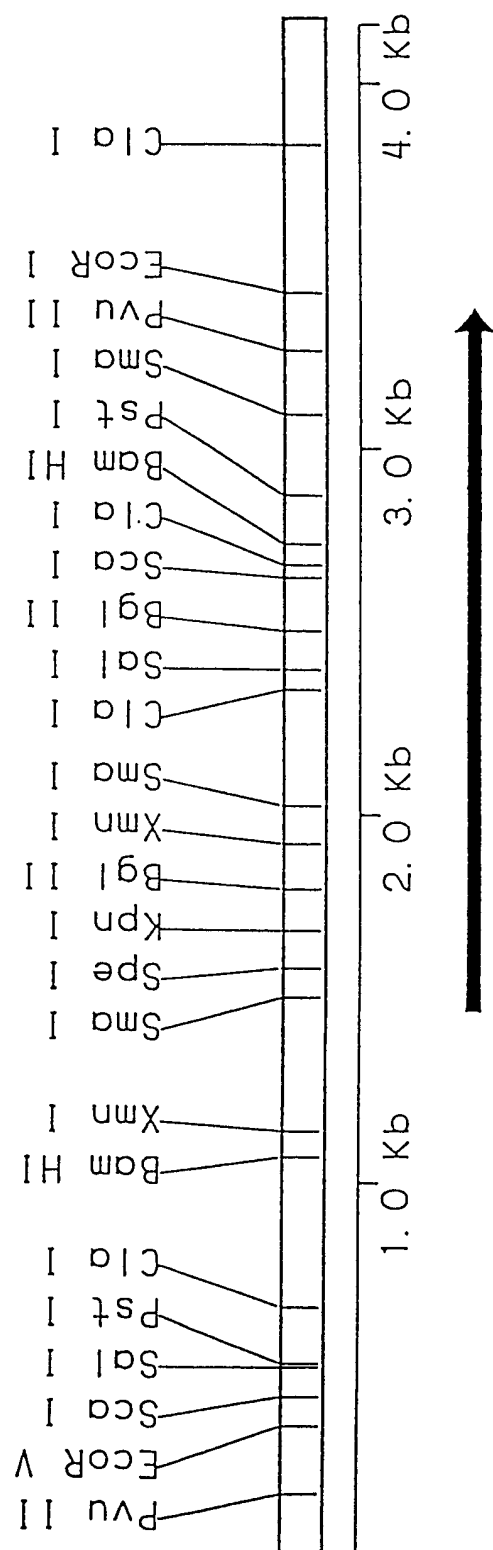


图 2

ACCATTGACTCTCAAGCTATGGATCGTGCTCACCGTCTCGGGCCAGACAAGACAGGTCACG 60
 GTGTATCGCCCTGATTACTCGCGGCACCAATTGAGGAGCGTATTCGCAAGCGAGCTTTGCGAG 120
 AAGGAGGAAGTGCAGCGTGTCTCATCTCAGGTGGCGCAGCTGGTGGGGTTGACTTCAAT 180
 ACTCGCAACCGGAGAGCCGAACCAAGGACATCGCCATGTGGCTGGCAGATGATGAACAG 240
 GCGGAGCTTATTGAGCAAAAGGAGAAAGGAGCGCTGGACCGAGGCGAAGTGTTTGGCGCT 300
 AGTAAAGGCGGGAAGAAGGCTGCTCAGAAGAGAAAGAGATATCACGCTGGATGATATG 360
 TATCATGAAGGTATGTGAATCTGATCAAGCTCTTCGTTCGGGGAGGCTTCTGGAAATA 420
 GTACTAACCGGCTCAATCTATAGCGGAAGGGAACTTTGACGATGCCAGTGCAAGCCATC 480
 AGGAGCGGCACCTCCTGTGTCGACTGCAGAGAAATTTAGGCACCCCATCCTCCACGCCAGT 540
 TCCTAAACGAGGACGTGGAAGGGGACAGGAAGGGCACGTCTAAAGAGCCAAACTAC 600
 CAAGGAGAGATTACGTCTCATTTGATGGCGACGGAGGCTTAGGGCTAGTTGATTTAATCG 660
 ATCTGTGCCCTCAATAATGGACACGGCTGGTTATGGTCA TGGCGTTCAGAGATTGCATTTT 720
 TTTCCCACTTTATCTTTCTTTCTTTCTTTAAACCCCTCTTTTGTGTTTCTTTT 780
 ATCGGACTTTACTTGTGGGCAGCTTACGTTCTGCCTTGATTAACAGCATATATTCCTGA 840
 TTCTTGATGTACGAAGCGATTTAAGAGTCA TTGAAGACGAAGGATGAACCCGTGGTAAT 900
 CAGCCGATAATGGCAAAGAGAAAGGAGAAAGAAATCAAGTGCAGT TTTGAAATTGAT 960
 GGCAAGATAGACATTGTATCCTGTACCTGTCTTTGGGCTGTGACGGGGGGTGAAATTG 1020
 ACGGTCATCACCCGGCTATTATTACTATTGTGTACTGTACATCCGGATCCTGCTGGTCT 1080
 GTATCTAGTTAGGGCAATATTCCCGTCGCCAGGCTCTTGGGTTATGAAATGATTTCAT 1140
 GGTGAAGTTTCGTATCCGTACGCACCGAGAGATTCTTAGTATTACTTGTAATTATGAAA 1200
 TGCAC TTGCCGAGTTAAGTCCGCCGCCCAATCACGGCGGAGGATATGGTAAGCCGAAAAG 1260
 TCTCGCCGAAGTCCCGACTTACTCTTACTGGAAGTGGCTTAGTGCCCTCAGCGCCCCCT 1320
 CGCCCTCAGTCCATCAGCCAGATTGACTCTTATTCTCTCTCTCTTCGCGCGGGTGAC 1380
 ATATCCCTCTCCTCTCCCTCTCCCTCTTGACAACATTTTCATCTTCGCTTCTTTTGTGA 1440
 TATAGTCAGTTTCGCTATCCATTGAAGCATCACTCATGGAGTCTCTTCTCCAGCAGTCCC 1500

M E S L L Q Q S
 L L Q Q S

图 3A

```

GGGCGAIGTGGCCGGIICCIIAAGCGCACATCTCCATCTTCTCTGCGTACGCTGGCAACCG 1560
R A M C P F L K R T S P S L R T L A T
CGACTCGACCTAGCACTAGTTCGGGTGGAGGCACATGTCTAATCTCCAGGTCAATTGCCCC 1620
A T R P S T S S G G T M S N L Q V I A
GTCGCGCGCCIGICATGAGCAAGGCTCTGGCCGTGCAGAGCGCTCGCATGGCCGGTACCA 1680
R R C P V M S K A L A V Q S A R M A G T
AAAGATTACCTCATGTGCTGCCGGCATCACCGGTCTCGGCAACAAGCATTGCCGTGCTC 1740
K R F T S C A A G I T G L G N K H C R A
CTACTGGGAAGAGAACCCCTGCACTCCACCTCCGGTAACGGCGCCAATGTGAGCGCAGAGA 1800
P T G K R T L H S T S G N G A N V S A E
TCTACAAGAACACCCAGCGAGATCCCGCGGTTTCTCGAAGATCAAGACCCCTGCCAATG 1860
I Y K N T Q R D P A G F S K I K T P A N
CTACCGCCGCTGCCGCTACGTCTGGCCCTCGTCCAGAGGCTCCCGTGGCGAAGCCTTCA 1920
A T A A A A T S G P R P E A P V A K P F
ACTACAATTCTTCTACAACACCCGAATTGGAAAAGAAACACAAGGACAAGTCGTATCGCT 1980
N Y N S F Y N T E L E K K H K D K S Y R

```

图 3B

ATTTCAACAACATCAATCGTCTCGCTCAGGAGTTTCCCGGGCTCACACCACATCTGCCG 2040
 Y F N N I N R L A Q E F P R A H T T S A
 AGGAACGTGTGACGGTCTGGTCTCGAACGATTATCTCGGCATGGGCCGCAACCCGAGG 2100
 E E R V T V W C S N D Y L G M G R N P E
 TTCTGGCCACCATGCATAAGACATTGGACACCTACGGAGCCGGTGGGAGGTACICGCA 2160
 V L A T M H K T L D T Y G A G A G G T R
 ACAITTCAGGTCACAATCAACATGCCCGTGAGCCCTGGAGAACAACCTGGCCAAATTGCACG 2220
 (N I) S G H N Q H A V S L E N T L A K L H
 GCAAGGAGGCGGCATTAGTCTTCAGCTCATGCTTCGTGGCTAACGATGCCACCCCTCGCAA 2280
 G K E A A L V F S S C F V A N D A T L A
 CCCTGGGTAGCAAGTTGCCGACTGTGTTATTCTGTCCGATAGCCTGAATCATGCATCGA 2340
 T L G S K L P D C V I L S D S L N H A S
 TGATTCAGGGTATTGCGCCATTTCAGGCGCCCAAGAAATGGTTTTCAAGCATAATGATCTGG 2400
 M I Q G I R H S G A K K M V F K H N D L

图 3C

TCGACCTTGAGGCCAAGTTGGCAGCTCTACCTCTTTCATGTCTCCCAAGATTATTGCAATTCG 2460
 V D L E A K L A A L P L H V P K I I A F
 AATCAGTTTATAGCATGTGCGGATCTATTGCCCAATTGAGAAGATCTGTGATCTTGCAG 2520
 E S V Y S M C G S I A P I E K I C D L A
 ACAAGTACGGTGCCATTACTTTCTGGATGAAGTCCACGCTGTGGGAATGTACGGACCTC 2580
 D K Y G A I T F L D E V H A V G M Y G P
 ACGGAGCAGGTGTGGCAGAGCACCTTGACTATGACATCTATGCTTCCCAAGATACGGTCA 2640
 H G A G V A E H L D Y D I Y A S Q D T V
 ACCCGCGCAGTACTAAGGGAACCGTGATGGACCGAATCGATATTATCACCGGTACTCTGG 2700
 N P R S T K G T V M D R I D I I T G T L
 GCAAGGCCTACGGATGTGTCGGGGCTACATTGCTGGATCCGCTGCGATGGTTGACACCA 2760
 G K* A Y G C V G G Y I A G S A A M V D T
 TCCGCTCCCTCGCCCTGGCTTCATCTTACCACGTCCTTGCCGCCCGCCACCATGGCTG 2820
 I R S L A P G F I F T T S L P P A T M A
 GTGCAGACACTGCTATCCAGTACCAGGCTCGTCAACGAGGCGACCGCGTCTCTGCAGCAGT 2880
 G A D T A I Q Y Q A R H Q G D R V L Q Q
 TGCACACCCGCGGTCAAAGCAGCTTTCAGGAGTTGGATATTCTCTGTAAATCCCAACC 2940
 L H T R A V K A A F K E L D I P V I P N
 CCTCCCATATCATTCGCTCCTGGTTGGGATGCCGAGGTTGCTAAGAAGGCCCTCGGACA 3000
 P S H I I P L L V G D A E V A K A S D
 AGCTTCTGGAGGAGCATGGAAATTTATGTACAAGCCATCAACTACCCAACCGTGCTCGGG 3060
 K L L E E H G I Y V Q A I N Y P T V P R
 GTGAAGAGCGGCTTCGTATCACGCCACCCCGGACATATCAAGGAGCACCGCGACCAACC 3120
 G E R L R I T P T P G H I K E H R D H
 TGGTGCAAGCCGTCCTGGAACGAACTGGGCATCAACGCACCAACCGGATTGGG 3180
 L V Q A V Q T V W N E L G I K R T S D W

图 3D

```

AAGCGCAAGGCGGCTTCGTGCGGCGTGCGGTGGGCGCGGCGGCTGAGAACAGCCGA 3240
E A Q G G F V G V G V D G A E A E N Q P
TTTGGAATGATGTGCAGCTGGGCTGAAGGAAACGAAGCCATTGAGGCTGCTGTGGAAC 3300
I W N D V Q L G L K E N E A I E A A V E
GCGAGTTTGCGGAGGCCCCCATGCGGACCGCCACCCGTCCTGCCGCGGCTGCTGCTTCGT 3360
R E F A E A P M R T A T R P A A A A S
CAATCCCGGTGGGTGTGGCTGCCTGAAGTGGCTGCCCGCATGTGAGCTGAAATCGACGTG 3420
S I P V G V A A
GAATTCTATACACACACACACACACACACACACACACACACACACACACACACACA 3480
CACACACACACACACACTAATGTTATATAATTCACATCCACTCCTTTGT 3540
CCCTTGTTGGACGTAATTGGTATTTGGACTATTAGTAACCAAGTCAGTCGTTACCATG 3600
TGTTTCGGTTCGACTCGAAATCTGACATGTTGTCGCCCCCATGCCACTTCATCTCCTCC 3660
GTAACCGCAGGCTTCAATAACACTGCCCCAGTAATTGTAGTCAATATAGCAGTTAACTAA 3720
CCTTCACCAATTTCCTAATAACAATAGAGGGCCATACACGCAGTACCAAAGATCACCT 3780
ACCTCCGATCAATAATCCGAACCTCAGGCTACATACATCAAGTCGCATTATCGATTCCGA 3840
CCTCTGTTTATCCCTGAAAATAACTAAGATCATGATCTACGTTTGGTAAGTGGGACACCT 3900
ACCTACACTGGGAGGTATTGAATAAAGGCATCATTCATATAGTCACAAGATGCCAGGGCC 3960
AATTCAATGATATGGATAGCTACTTCCAAACATAATTTCAGAGGTATCATTCCTGCTCTCAG 4020
ACAGTTCTTCTCGAAGATCAGTAGGAGCCAGTTTTCACCATTAACCTTGTAATGTAATGTC 4080
GATTGTAGTAGATCCGAGATCCATTTCACCTTCTAAGGGTTAATTGATTTCATTTTACTGAT 4140
ACCTCACCCACCATATT 4157

```

图 3E

[illegible]

4A

米曲霉	Q	V	I	A	R	R	C	P	V	M	S	-	-	-	-	-	-	-	-	-	-	K	A	L	A	V	Q	S	A	63	
构巢曲霉	Q	R	I	A	R	R	C	P	V	M	S	-	-	-	-	-	-	-	-	-	-	K	A	L	A	V	Q	S	A	63	
鸡红细胞	-	-	-	-	-	R	C	P	F	M	G	-	-	-	-	-	-	-	-	-	-	F	A	-	R	A	A	P	37		
人红细胞	-	-	-	-	-	R	C	P	I	L	A	T	Q	G	G	P	N	C	S	Q	I	H	-	-	-	-	-	G	60		
小鼠红细胞	-	-	-	-	-	R	C	P	I	L	A	T	Q	G	G	P	N	C	S	Q	I	H	-	-	-	-	-	G	60		
鸡肝	-	-	-	-	-	H	C	P	K	M	M	-	-	-	-	-	-	-	-	-	-	E	A	-	P	A	A	44			
人肝	-	-	-	-	-	N	C	P	K	M	M	-	-	-	-	-	-	-	-	-	-	E	A	-	P	A	A	44			
大鼠肝	-	-	-	-	-	N	C	P	K	M	M	-	-	-	-	-	-	-	-	-	-	E	A	-	P	A	A	44			
米曲霉	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	R	-	M	-	-	-	-	-	-	-	-	-	-	65	
构巢曲霉	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	R	-	M	-	-	-	-	-	-	-	-	-	-	65	
鸡红细胞	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	E	-	L	-	-	-	-	-	-	-	-	-	-	39	
人红细胞	D	S	P	S	W	A	K	G	H	C	P	F	M	L	S	E	-	-	L	Q	D	D	-	-	-	-	-	-	-	-	87
小鼠红细胞	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	E	-	-	L	Q	D	D	-	-	-	-	-	-	-	-	72
鸡肝	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	R	G	L	A	T	S	T	-	-	-	-	-	-	-	59
人肝	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	P	R	A	L	S	T	S	-	-	-	-	-	-	-	-	59
大鼠肝	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	P	R	A	L	S	T	S	-	-	-	-	-	-	-	-	59

4B

[illegible]

圖 4C

米曲霉	M	-	-	E	S	L	L	Q	Q	S	R	A	M	C	P	F	L	K	R	T	S	P	S	S	L	R	T	L	A	T	28
构巢曲霉	M	-	-	E	A	L	L	Q	Q	S	R	A	M	C	P	F	L	K	R	S	P	N	T	L	R	S	L	A	T	28	
人	M	V	T	A	A	M	L	L	Q	-	-	C	C	P	V	L	A	R	G	P	T	S	L	L	G	K	V	V	K	27	
酿酒酵母	M	-	-	Q	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	R	S	I	F	A	7	

米曲霉	A	T	R	P	S	T	S	S	G	G	G	T	M	S	N	L	Q	V	I	A	R	R	C	P	V	M	S	-	K	A	57
构巢曲霉	A	T	R	P	S	T	S	P	G	G	G	T	M	T	N	L	Q	R	I	A	R	R	C	P	V	M	S	-	K	A	57
人	T	H	Q	F	L	F	G	I	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	45	
酿酒酵母	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	7	

米曲霉	L	A	V	Q	S	A	R	M	A	G	T	K	R	F	T	S	C	A	A	G	I	T	G	L	G	N	-	-	-	K	84
构巢曲霉	L	A	V	Q	S	A	R	M	T	G	T	K	R	F	T	S	S	A	A	G	V	P	G	A	G	A	G	T	P	K	87
人	P	N	C	S	Q	I	H	L	K	A	T	K	A	G	G	D	S	P	S	W	A	K	G	H	C	P	F	M	L	S	75
酿酒酵母	-	-	-	-	-	-	-	-	-	-	-	R	F	G	N	S	S	A	V	S	T	L	N	R	L	S	T	T	25		

米曲霉	H	C	R	A	P	T	G	K	R	T	L	H	S	T	S	G	N	G	A	N	V	S	A	E	I	Y	K	N	T	Q	114
构巢曲霉	P	T	R	G	S	P	G	K	R	A	L	H	S	T	G	G	N	G	A	N	M	S	T	E	F	H	K	G	A	Q	117
人	E	L	Q	D	G	K	S	K	-	I	V	Q	K	A	A	P	E	V	Q	E	D	V	K	A	F	K	T	D	L	P	104
酿酒酵母	A	-	-	A	P	H	A	K	N	G	Y	A	T	A	T	G	A	G	A	A	A	T	A	-	-	-	-	-	-	47	

米曲霉	R	D	P	A	G	F	S	K	I	K	T	P	A	N	A	T	A	A	A	A	T	S	G	P	R	P	E	-	-	-	141
构巢曲霉	Q	I	H	P	G	L	S	N	A	-	T	R	S	H	V	G	A	S	A	T	V	S	G	P	T	P	R	-	-	-	143
人	S	S	L	V	S	V	S	L	R	K	P	S	G	P	Q	E	Q	E	Q	I	I	S	G	K	V	T	H	L	I	Q	134
酿酒酵母	-	-	-	-	-	-	-	-	-	-	T	A	S	S	T	H	A	A	A	A	A	A	A	A	A	N	H	-	-	64	

图 5A

米曲霉	-	A	P	V	A	K	P	F	N	Y	N	S	F	Y	N	T	E	L	E	K	K	H	K	D	K	S	Y	R	Y	169	
构巢曲霉	-	A	P	V	A	A	P	F	D	Y	D	A	F	Y	N	A	E	L	Q	K	K	H	Q	D	K	S	Y	R	Y	171	
人	N	M	P	G	N	Y	V	F	S	Y	D	Q	F	F	R	D	K	I	M	E	K	K	Q	D	H	T	Y	R	V	164	
酿酒酵母	-	S	T	Q	E	S	G	F	F	D	Y	E	G	L	I	D	S	E	L	Q	K	R	L	D	K	S	Y	R	Y	92	
米曲霉	F	N	I	N	R	L	A	Q	E	F	P	R	A	H	T	-	-	-	-	T	S	A	E	R	V	T	V	W	195		
构巢曲霉	F	N	I	N	R	L	A	Q	E	F	P	R	A	H	T	-	-	-	-	A	S	K	D	E	K	V	T	V	W	197	
人	F	K	T	V	N	R	W	A	D	A	Y	P	F	A	Q	H	F	F	E	A	S	V	A	S	K	D	V	S	V	W	194
酿酒酵母	F	N	I	N	R	L	A	K	E	F	P	L	A	H	R	-	-	-	-	Q	R	E	A	D	K	V	T	V	W	118	
米曲霉	C	S	N	D	Y	L	G	M	G	R	N	P	E	V	L	A	T	M	H	K	T	L	D	T	Y	G	A	G	A	G	225
构巢曲霉	C	S	N	D	Y	L	G	M	G	R	N	P	E	V	L	A	T	M	H	K	T	L	D	T	Y	G	A	G	A	G	227
人	C	S	N	D	Y	L	G	M	S	R	H	P	Q	V	L	Q	A	T	Q	E	T	L	Q	R	H	G	A	G	A	G	224
酿酒酵母	C	S	N	D	Y	L	A	L	S	K	H	P	E	V	L	D	A	M	H	K	T	I	D	K	Y	G	C	G	A	G	148
米曲霉	G	T	R	N	I	S	G	H	N	Q	H	A	V	S	L	E	N	T	L	A	K	L	H	G	K	E	A	A	L	V	255
构巢曲霉	G	T	R	N	I	S	G	H	N	Q	H	A	V	S	L	E	N	T	L	A	K	L	H	G	K	E	A	A	L	V	257
人	G	T	R	N	I	S	G	T	S	K	F	H	V	E	L	E	Q	E	L	A	E	L	H	Q	K	D	S	A	L	L	254
酿酒酵母	G	T	R	N	I	A	G	H	N	I	P	T	L	N	L	E	A	E	L	A	T	L	H	K	K	E	G	A	L	V	178

图 5B

米曲霉	F	S	S	C	F	V	A	N	D	A	T	L	A	T	L	G	S	K	L	P	D	C	V	I	L	S	D	S	L	N	285
构巢霉	F	S	S	C	F	V	A	N	D	A	T	L	A	T	L	G	S	K	M	P	D	C	V	I	L	S	D	S	L	N	287
人	F	S	S	C	F	V	A	N	D	S	T	L	F	T	L	A	K	I	L	P	G	C	E	I	Y	S	D	A	G	N	284
酿酒酵母	F	S	S	C	Y	V	A	N	D	A	V	L	S	L	G	Q	K	M	K	D	L	V	I	F	S	D	E	L	N	208	
米曲霉	H	A	S	M	I	Q	G	I	R	H	S	G	A	K	K	M	V	F	K	H	N	D	L	V	D	L	E	A	K	L	315
构巢霉	H	A	S	M	I	Q	G	I	R	H	S	G	R	K	K	M	V	F	K	H	N	D	L	V	D	L	E	T	K	L	317
人	H	A	S	M	I	Q	G	I	R	N	S	G	A	K	F	V	F	R	H	N	D	P	D	H	L	K	L	L	L	314	
酿酒酵母	H	A	S	M	I	V	G	I	K	H	A	N	V	K	K	I	F	K	H	N	D	L	N	E	L	E	Q	L	L	238	
米曲霉	A	A	L	P	L	H	V	P	K	I	I	A	F	E	S	V	Y	S	M	C	G	S	I	A	P	I	E	K	I	C	345
构巢霉	A	S	L	P	L	H	V	P	K	I	I	A	F	E	S	V	Y	S	M	C	G	S	I	A	P	I	E	A	I	C	347
人	E	K	S	N	P	K	I	P	K	I	V	A	F	E	T	V	H	S	M	D	G	A	I	C	P	L	E	E	L	C	344
酿酒酵母	Q	S	Y	P	K	S	V	P	K	L	I	A	F	E	S	V	Y	S	M	A	G	S	V	A	D	I	E	K	I	C	268

图5C

米曲霉 375
 构巢曲霉 377
 人 374
 酿酒酵母 298

D L A D K Y G A I T F L D E V H A V G M Y G P H G A G V A E
 D L A D K Y G A I T F L D E V H A V G M Y G P H G A G V A E
 D V S H Q Y G A L T F V D E V H A V G L Y G S R G A G I G E
 D L A D K Y G A L T F L D E V H A V G L Y G P H G A G V A E

米曲霉 401
 构巢曲霉 403
 人 382
 酿酒酵母 328

H L D Y D I Y A S Q D T V N P R S T - K G - - - T V M D R I
 H L D Y E I Y A S Q D T A N P L S T - K G - - - T V M D R I
 R - - - - - - - - - - - - - - - D G I M H K I
 H C D F E S H R A S G I A T P K T N D K G G A K T V M D R V

米曲霉 431
 构巢曲霉 433
 人 412
 酿酒酵母 358

D I I T G T L G K A Y G C V G G Y I A G S A A M V D T I R S
 N I I T G T L G K A Y G C V G G Y I A G S A A L V D T I R S
 D I I S G T L G K A F G C V G G Y I A S T R D L V D M V R S
 D M I T G T L G K S F G S V G G Y V A S R K L I D W F R S

米曲霉 461
 构巢曲霉 463
 人 442
 酿酒酵母 388

L A P G F I F T T S L P P A T M A G A D T A I Q Y Q A R H Q
 L A P G F I F T T S L P P A T M A G A D T A I R Y Q A R H Q
 Y A A G F I F T T S L P P M V L S G A L E S V R L L K G E E
 F A P G F I F T T T L P P S V M A G A T A A I R Y Q R C H I

图5D

米曲霉	G	D	-	-	R	V	L	Q	Q	L	H	T	R	A	V	K	A	A	F	K	E	L	D	I	P	V	I	P	N	P	489	
构巢霉	Q	D	-	-	R	I	L	Q	Q	L	H	T	R	A	V	K	Q	S	F	K	D	L	D	I	P	V	I	P	N	P	491	
人	G	Q	A	L	R	R	A	H	Q	R	N	V	K	H	M	R	Q	L	M	D	R	G	L	P	V	I	P	C	P	472		
酿酒酵母	D	L	-	-	R	T	S	Q	Q	K	H	T	M	Y	V	K	K	A	F	H	E	L	G	I	P	V	I	P	N	P	416	
米曲霉	S	H	I	I	P	L	L	V	G	D	A	E	V	A	K	K	A	S	D	K	L	L	E	E	H	G	I	Y	V	Q	519	
构巢霉	S	H	I	V	P	L	L	V	G	D	A	E	L	A	K	Q	A	S	D	K	L	L	E	E	H	G	I	Y	V	Q	521	
人	S	H	I	I	P	I	R	V	G	N	A	A	L	N	S	K	L	C	D	L	L	L	S	K	H	G	I	Y	V	Q	502	
酿酒酵母	S	H	I	V	P	V	L	I	G	N	A	D	L	A	K	Q	A	S	D	I	L	I	N	K	H	Q	I	Y	V	Q	446	
米曲霉	A	I	N	Y	P	T	V	P	R	G	E	E	R	L	R	I	T	P	T	P	G	H	I	K	E	H	R	D	H	L	549	
构巢霉	A	I	N	Y	P	T	V	P	R	G	E	E	R	L	R	I	T	P	T	P	G	H	T	Q	E	E	L	R	D	H	L	551
人	A	I	N	Y	P	T	V	P	R	G	E	E	L	L	R	L	A	P	S	P	H	H	S	P	Q	M	M	E	D	F	532.	
酿酒酵母	A	I	N	F	P	T	V	A	R	G	T	E	R	L	R	I	T	P	T	P	G	H	T	N	D	L	S	D	I	L	476	

图 5E

米曲霉	V	Q	A	V	Q	T	V	W	N	E	L	G	I	K	R	T	S	D	W	E	A	Q	G	G	F	V	G	V	G	V	579
构巢曲霉	V	E	A	V	N	T	V	W	N	D	L	G	I	K	R	A	S	D	W	K	A	M	G	G	F	V	G	V	G	V	581
人	V	E	K	L	L	L	A	W	T	A	V	G	L	P	-	-	-	-	-	-	-	-	-	L	Q	D	V	S	V	552	
酿酒酵母	I	N	A	V	D	D	V	F	N	E	L	Q	L	P	R	V	R	D	W	E	S	Q	G	G	L	L	G	V	G	-	505
米曲霉	D	G	A	E	A	E	N	Q	P	I	W	N	D	V	Q	L	G	L	K	E	N	E	A	I	E	A	A	V	E	R	609
构巢曲霉	E	A	A	E	L	E	N	Q	P	I	W	T	D	A	Q	L	N	M	R	P	D	E	T	L	E	A	A	V	E	R	611
人	A	A	C	N	F	C	R	R	P	V	-	-	-	-	H	F	E	L	M	S	E	-	-	-	-	-	-	W	E	R	572
酿酒酵母	E	S	G	F	V	E	E	S	N	L	W	T	S	S	Q	L	S	L	T	N	D	D	L	N	-	-	-	-	-	-	529
米曲霉	E	F	A	E	A	-	-	-	-	-	-	-	-	-	-	-	P	M	R	T	A	T	R	P	A	A	A	A	S	S	629
构巢曲霉	E	F	Q	A	A	V	P	G	M	K	A	G	G	A	K	A	K	P	V	G	S	I	A	A	N	P	I	G	A	S	641
人	S	Y	F	G	N	M	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	579	
酿酒酵母	-	-	-	-	-	-	P	N	V	R	D	P	I	V	K	Q	L	E	V	S	S	-	-	-	-	-	-	-	-	-	544
米曲霉	I	P	V	G	V	A	-	A	-	A	-	A	-	A	-	A	-	A	-	A	-	A	-	A	-	A	-	A	-	636	
构巢曲霉	I	P	V	A	A	A	A	Z	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	649	
人	P	Q	Y	V	T	T	Y	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	587	
酿酒酵母	-	-	-	-	-	-	G	I	K	Q	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	548	

图5F

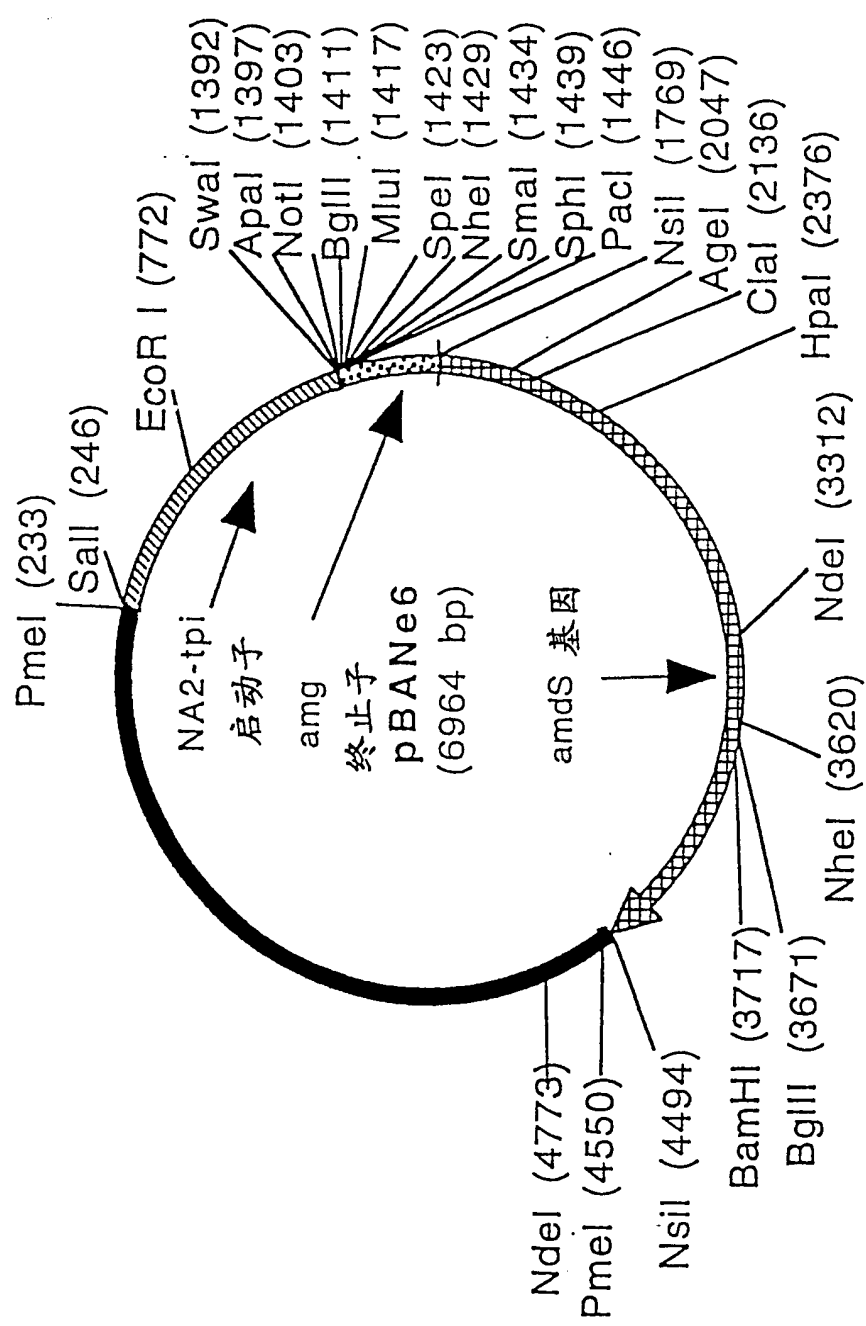


图 6

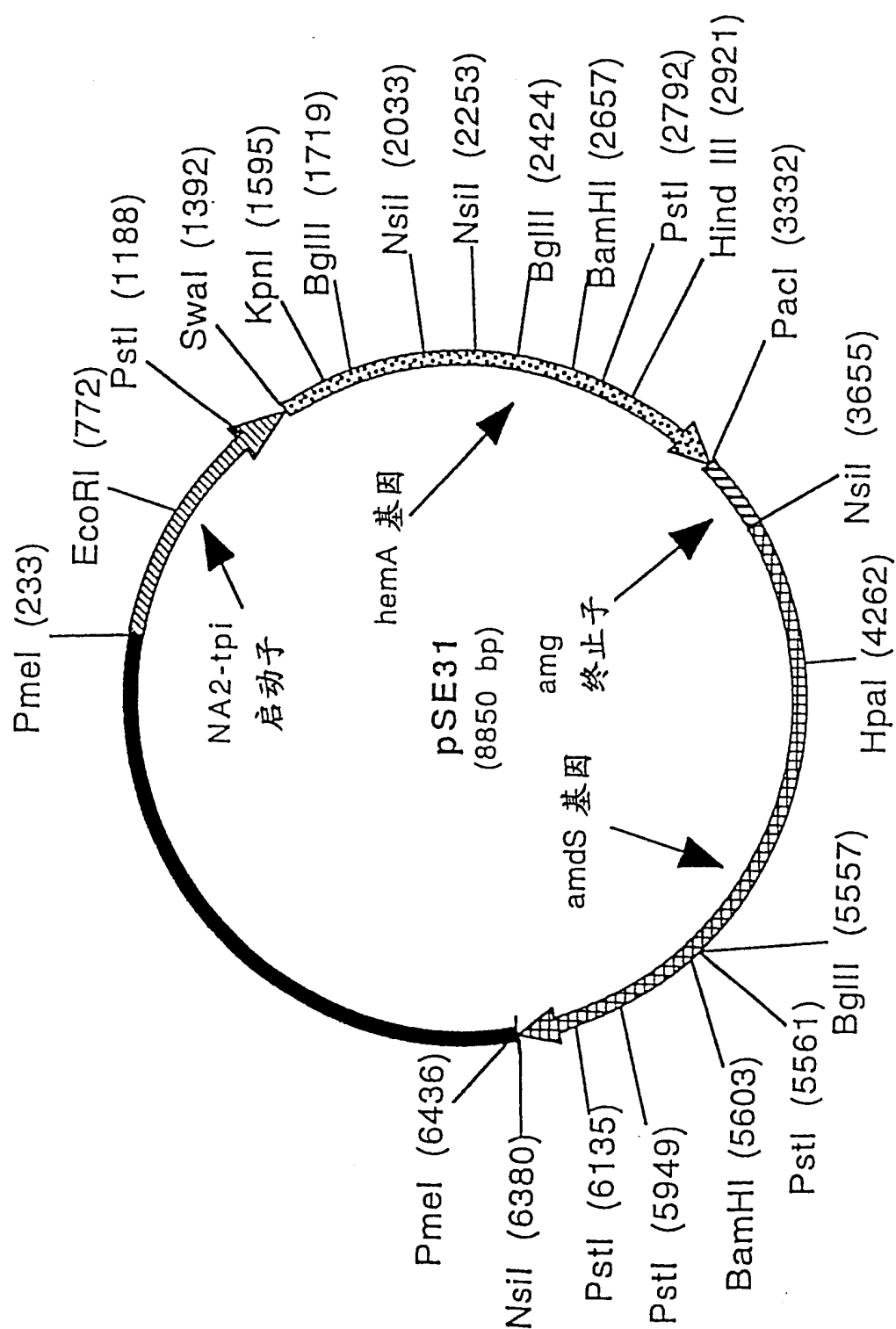


图 7

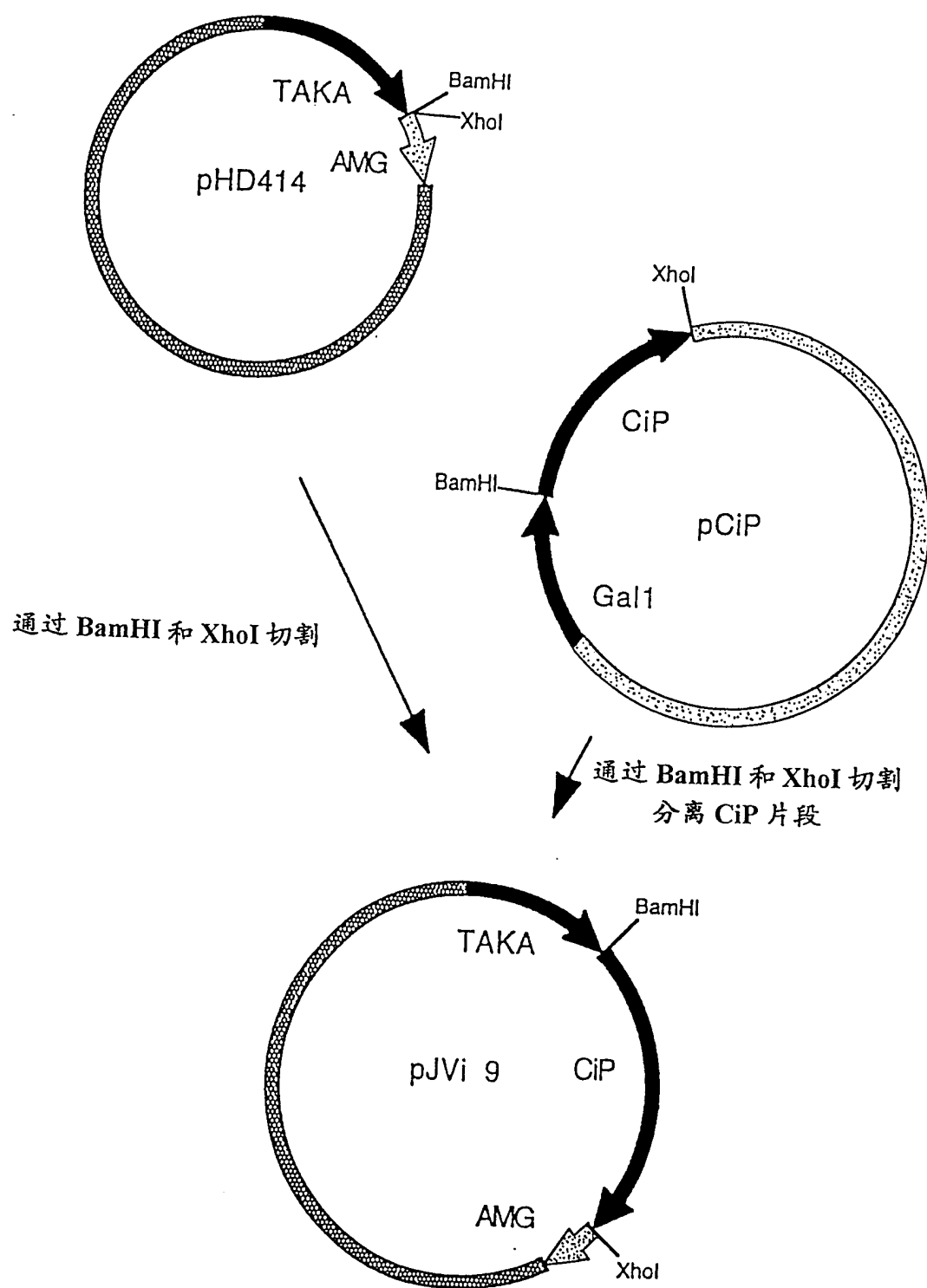


图 8

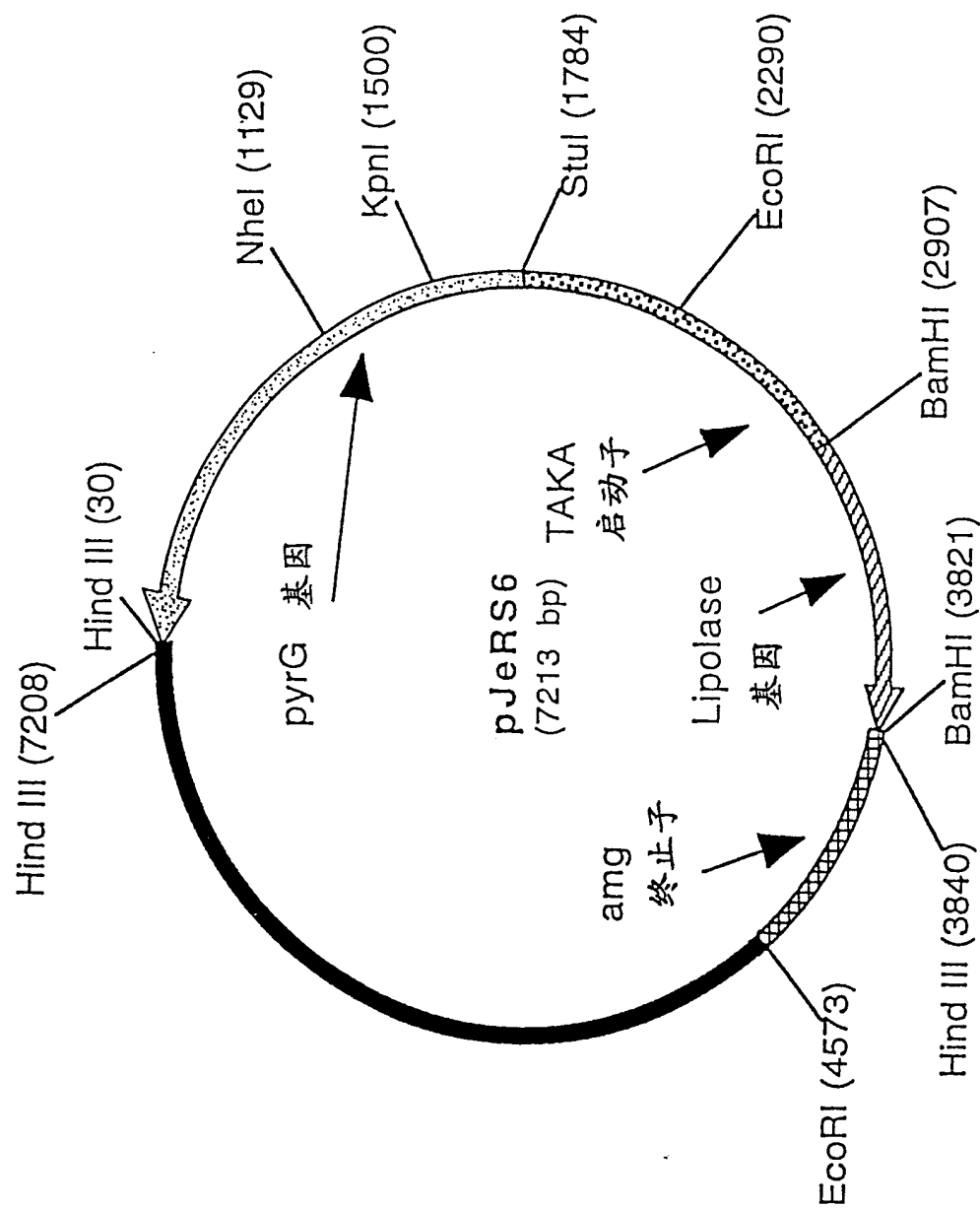


图 9

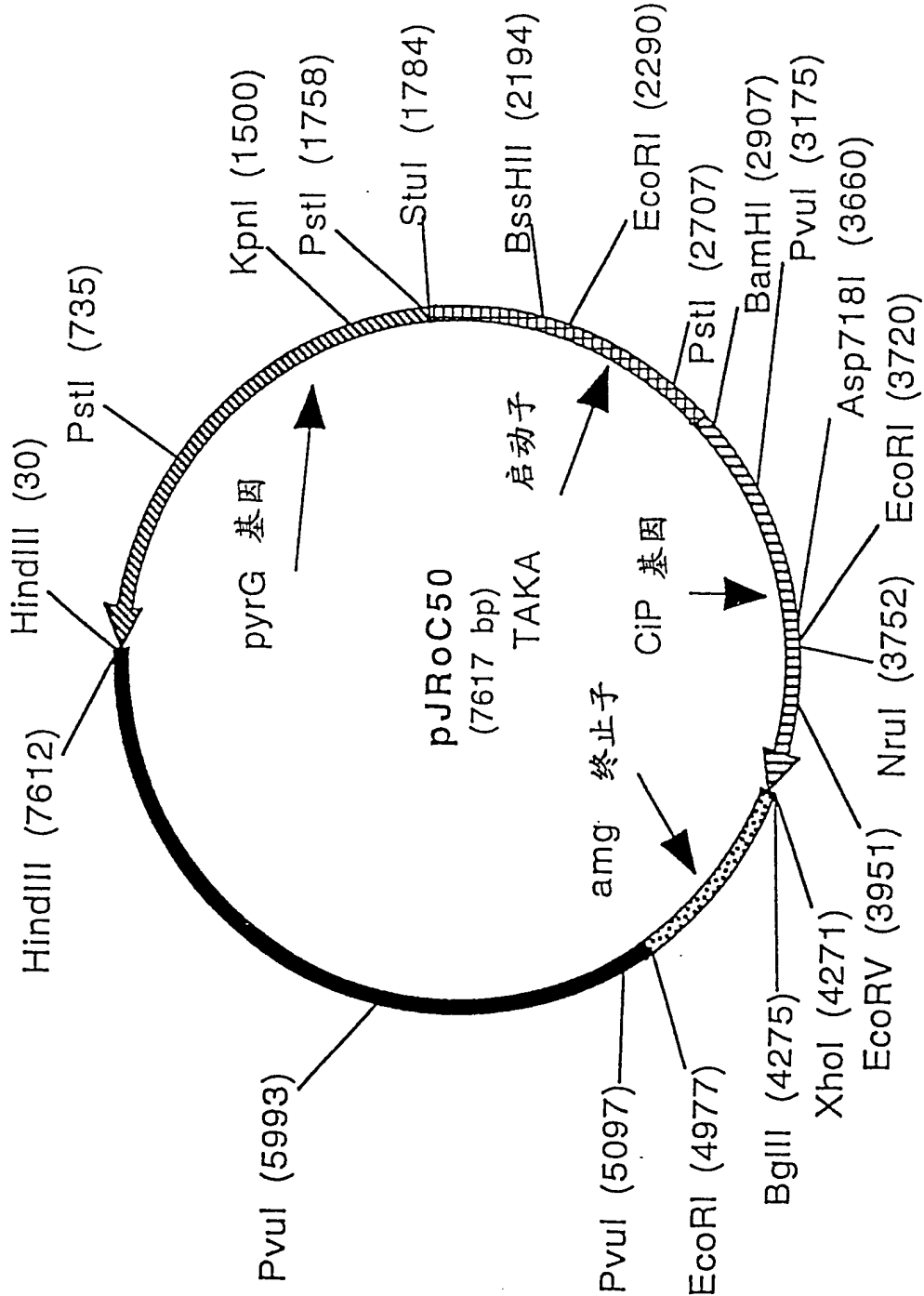


图 10