

[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 98804285.1

[51] Int. Cl.

C12N 15/64 (2006.01)

C12P 21/02 (2006.01)

C12N 1/14 (2006.01)

C12R 1/69 (2006.01)

[45] 授权公告日 2007 年 11 月 28 日

[11] 授权公告号 CN 100351383C

[22] 申请日 1998.3.17 [21] 申请号 98804285.1

[30] 优先权

[32] 1997. 3. 17 [33] US [31] 08/819,458

[86] 国际申请 PCT/US1998/005156 1998. 3. 17

[87] 国际公布 WO1998/041640 英 1998. 9. 24

[85] 进入国家阶段日期 1999. 10. 18

[73] 专利权人 诺沃奇梅兹有限公司

地址 丹麦鲍斯韦

[72] 发明人 E·B·詹森 J·R·彻里

S·L·埃尔罗德

[56] 参考文献

WO9609397 1996. 1. 1

US4902620A 1990. 1. 1

审查员 孙大龙

[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 5 页 说明书 36 页 附图 10 页

[54] 发明名称

在呼吸缺陷型细胞中生产多肽的方法

[57] 摘要

本发明涉及生产多肽的方法，包括：(a) 向细胞的呼吸缺陷型突变体中导入：(i) 一个或多个可弥补该呼吸缺陷的第一核酸序列，和(ii) 编码该多肽的第二核酸序列；(b) 在适合第一和第二核酸序列表达的需氧条件下，于培养基中培养含有第一和第二核酸序列的细胞；和(c) 从细胞培养基中分离该多肽。本发明也涉及破坏呼吸缺陷型突变细胞中基因的方法。本发明进一步涉及呼吸缺陷型细胞和获得这种突变细胞的方法。

1. 一种生产多肽的方法，包括：

(a) 在适合第一和第二核酸序列表达的需氧条件下，于培养基中培养真菌细胞的呼吸缺陷型突变体，其中该突变细胞内含有包含一个或多个第一核酸序列以及第二核酸序列的核酸构建体，其中第一核酸序列在表达后可弥补突变细胞的呼吸缺陷，第二核酸序列编码所述多肽，该第一和第二核酸序列包含在一种核酸构建体上或包含在分开的构建体上；及

(b) 从突变细胞的培养基中分离多肽。

2. 权利要求1的方法，其中该第一核酸序列编码的蛋白质是电子传递链中的成员。

3. 权利要求2的方法，其中该成员是NADH - Q还原酶。

4. 权利要求2的方法，其中该成员是细胞色素还原酶。

5. 权利要求4的方法，其中该细胞色素还原酶成员是细胞色素b或细胞色素c₁。

6. 权利要求2的方法，其中该成员是细胞色素c。

7. 权利要求2的方法，其中该成员是细胞色素氧化酶。

8. 权利要求7的方法，其中该细胞色素氧化酶成员是细胞色素a或细胞色素a₃。

9. 权利要求1的方法，其中第一核酸序列编码参与泛醌生物合成的酶。

10. 权利要求1的方法，其中该第一核酸序列编码参与黄素生物合成的酶。

11. 权利要求1的方法，其中该第一核酸序列编码参与血红素生物合成的酶。

12. 权利要求11的方法，其中该第一核酸序列编码选自下述组中的酶：5-氨基- γ -酮戊酸合成酶、胆色素原合成酶、胆色素原脱氨基酶、尿卟啉原合成酶、尿卟啉原脱羧酶、粪卟啉原氧化酶、原卟啉原

氧化酶、亚铁螯合酶、谷氨酸-tRNA_{glu}合成酶、谷氨酸-tRNA_{glu}还原酶、谷氨酸1-半醛氨基转移酶。

13. 权利要求1的方法，其中第一核酸序列和第二核酸序列被整合在所述细胞的染色体内。

14. 权利要求1的方法，其中第一核酸序列和第二核酸序列包含在染色体外元件上。

15. 权利要求1的方法，其中该多肽是所述突变细胞的天然或异源多肽。

16. 权利要求1的方法，其中该多肽是酶、激素、激素变体、受体或其部分、抗体或其部分，或者报道分子。

17. 权利要求16的方法，其中所述酶是氧化还原酶、转移酶、水解酶、裂解酶、异构酶或连接酶。

18. 权利要求17的方法，其中所述酶是氨肽酶、淀粉酶、糖酶、羧肽酶、过氧化氢酶、纤维素酶、壳多糖酶、环糊精糖基转移酶、角质酶、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 -葡糖苷酶、 β -葡糖苷酶、谷氨酰胺酶、盐过氧化物酶、转化酶、漆酶、脂酶、甘露糖苷酶、齿斑葡聚糖酶、氧化酶、果胶裂解酶、过氧化物酶、肌醇六磷酸酶、多酚氧化酶、蛋白水解酶、核糖核酸酶、转谷氨酰胺酶或木聚糖酶。

19. 权利要求1的方法，其中所述真菌细胞是丝状真菌细胞。

20. 权利要求19的方法，其中所述丝状真菌细胞是枝顶孢属(*Acremonium*)、曲霉属、镰孢属、腐质霉属、毁丝霉属、毛霉属、脉孢菌属、青霉属、梭孢壳属、*Tolypocladium*或木霉属细胞。

21. 权利要求1的方法，其中所述真菌细胞是酵母细胞。

22. 权利要求21的方法，其中所述酵母细胞是假丝酵母属、克鲁维酵母属、酵母属、裂殖酵母属、毕赤酵母属或*Yarrowia*属细胞。

23. 破坏真菌细胞的呼吸缺陷型突变体内的基因的方法，包括：

(a) 通过将核酸构建体导入呼吸缺陷型突变细胞中的基因或其部分内而取代该基因，所述核酸构建体含有包含该基因或其部分的第三核

酸序列及第一核酸序列,其中第一核酸序列表达后弥补了该呼吸缺陷型突变细胞的呼吸缺陷;和

(b)在适于第一核酸序列表达的需氧条件下,于培养基中培养含该第一和第三核酸序列的细胞。

24. 权利要求23的方法,其中该第一核酸序列编码的蛋白质是电子传递链中的成员。

25. 权利要求24的方法,其中该成员是NADH-Q还原酶。

26. 权利要求24的方法,其中该成员是细胞色素还原酶。

27. 权利要求26的方法,其中该细胞色素还原酶成员是细胞色素b或细胞色素c₁。

28. 权利要求24的方法,其中该成员是细胞色素c。

29. 权利要求24的方法,其中该成员是细胞色素氧化酶。

30. 权利要求29的方法,其中该细胞色素氧化酶成员是细胞色素a或细胞色素a₃。

31. 权利要求23的方法,其中第一核酸序列编码参与泛醌生物合成的酶。

32. 权利要求23的方法,其中该第一核酸序列编码参与黄素生物合成的酶。

33. 权利要求23的方法,其中该第一核酸序列编码参与血红素生物合成的酶。

34. 权利要求23的方法,其中所述真菌细胞是丝状真菌细胞。

35. 权利要求34的方法,其中所述丝状真菌细胞是枝顶孢属(Acremonium)、曲霉属、镰孢属、腐质霉属、毁丝霉属、毛霉属、脉孢菌属、青霉属、梭孢壳属、Tolypocladium或木霉属细胞。

36. 权利要求23的方法,其中所述真菌细胞是酵母细胞。

37. 权利要求36的方法,其中所述酵母细胞是假丝酵母属、克鲁维酵母属、酵母属、裂殖酵母属、毕赤酵母属或Yarrowia属细胞。

38. 亲本丝状真菌细胞的呼吸缺陷型突变体,其中含有第一核酸序列,其内至少一种氧化磷酸化作用必需基因具有修饰,其中,相同条

件下培养时,该突变体相比于所述亲本细胞而言是呼吸缺陷型的。

39. 权利要求38的呼吸缺陷型突变体,它是泛醌缺陷型的。

40. 权利要求38的呼吸缺陷型突变体,它是黄素缺陷型的。

41. 权利要求38的呼吸缺陷型突变体,它是血红素缺陷型的。

42. 权利要求38的呼吸缺陷型突变体,其中所述丝状真菌细胞是枝顶孢属(*Acremonium*)、曲霉属、镰孢属、腐质霉属、毁丝霉属、毛霉属、脉孢菌属、青霉属、梭孢壳属、*Tolypocladium*或木霉属细胞。

43. 获得亲本丝状真菌细胞的呼吸缺陷型突变体的方法,包括:

(a) 向该丝状真菌细胞内导入核酸序列,该核酸序列内含有至少一种氧化磷酸化作用必需基因的修饰,其中天然基因是通过所导入的经修饰基因同源重组到所述天然基因内而被破坏的;以及

(b) 从步骤(a)中鉴定出所述亲本细胞的含该核酸序列的突变体,其中突变体在与所述亲本细胞相同的条件下培养时是呼吸缺陷型的。

44. 权利要求43的方法,其中所述基因编码的蛋白质是电子传递链中的成员。

45. 权利要求44的方法,其中该成员是NADH-Q还原酶。

46. 权利要求44的方法,其中该成员是细胞色素还原酶。

47. 权利要求46的方法,其中该细胞色素还原酶成员是细胞色素b或细胞色素 c_1 。

48. 权利要求44的方法,其中该成员是细胞色素c。

49. 权利要求44的方法,其中该成员是细胞色素氧化酶。

50. 权利要求49的方法,其中该细胞色素氧化酶成员是细胞色素a或细胞色素 a_3 。

51. 权利要求43的方法,其中所述基因编码参与泛醌生物合成的一种酶。

52. 权利要求43的方法,其中所述基因编码参与黄素生物合成的酶。

53. 权利要求43的方法,其中所述基因编码参与血红素生物合成的酶。

54. 权利要求43的方法，其中所述真菌细胞是丝状真菌细胞。

55. 权利要求54的方法，其中所述丝状真菌细胞是枝顶孢属（*Acremonium*）、曲霉属、镰孢属、腐质霉属、毁丝霉属、毛霉属、脉孢菌属、青霉属、梭孢壳属、*Tolypocladium*或木霉属细胞。

在呼吸缺陷型细胞中生产多肽的方法

发明背景

发明领域

本发明涉及在呼吸缺陷型突变细胞中生产多肽的方法。本发明也涉及破坏呼吸缺陷型细胞中基因的方法。本发明进一步涉及呼吸缺陷型突变细胞和获得这种突变细胞的方法。

相关技术描述

在宿主细胞内选择和保持重组DNA分子有许多方法。一种方法包括一种用克隆的野生型基因补偿营养缺陷型突变。然而，在复杂的培养基中保持互补DNA是很困难的，因为在复杂的培养基中，可能存在某种氨基酸或核酸，它们可缓冲营养缺陷。另一种方法是利用能在野生型受体中被选择的标记基因，尤其是诸如抗生素抗性基因之类的抗性标记基因。抗生素抗性基因的使用通常需要在培养基中添加昂贵的抗生素，还需要随后从目标产品中将其去除。一种进一步的方法包括使用一种致死性染色体标记，这种标记被重组DNA克隆载体上携带的一种基因所抑制。这种方法已在细菌中成功地得到运用，但是，运用在其它宿主(例如真菌)中是有困难的。此外，该方法要求使用由质粒携带的一种阻遏蛋白基因，其不干扰驱动重组DNA分子表达的转录激活序列。

含有重组DNA分子的表达载体在宿主细胞内的不稳定性对于重组DNA分子的稳定高水平表达产生了一个严重的问题。转化体的亚培养中，重组DNA分子所编码产物的产量可能会急剧地和不可预测地下降。如果重组DNA分子的表达对宿主细胞具有有害的作用，例如该表达产物对细胞具有毒性时，尤为如此。理想的是筛选和维持含有重组DNA分子的转化体的微生物培养物，使得培养物中基本上所有的微生

物细胞都会有该重组DNA分子。

本领域亟需另外的选择和维持系统，以确保培养期间转化体的高水平表达和遗传稳定性。

美国专利No. 4, 902, 620中公开了将一种血红素生物合成酶(5-氨基- γ -酮戊酸合成酶)基因导入血红素缺陷型细胞中以生产5-氨基- γ -酮戊酸的方法。

本发明的一个目的是提供在基因的异源表达中选择和维持重组细胞的新方法。

发明概述

本发明涉及生产多肽的方法，包括：

(a) 将包括一个或多个第一核酸序列和一个第二核酸序列的一种核酸构建体导入一种呼吸缺陷型突变细胞中，其中第一核酸序列在表达后可弥补呼吸缺陷，第二核酸序列编码所述多肽；

(b) 在适合第一和第二核酸序列表达的需氧条件下，于培养基中培养含有该第一和第二核酸序列的细胞；及

(c) 从细胞的培养基中分离多肽。

本发明也涉及破坏呼吸缺陷型突变细胞中基因的方法。本发明进一步涉及呼吸缺陷型突变细胞和获得这种突变细胞的方法。

附图简要说明

图1显示质粒pSE04的限制性图谱。

图2显示含有米曲霉IFO 4177 5-氨基- γ -酮戊酸合成酶基因的4.2kb基因组片段的限制性图谱。

图3显示米曲霉IFO 4177 5-氨基- γ -酮戊酸合成酶基因的核苷酸序列和推导出的氨基酸序列(分别为SEQ ID NO:1和2)。

图4显示质粒pAJ005-1的限制性图谱。

图5显示米曲霉IFO 4177胆色素原合成酶基因的核苷酸序列和推导出的氨基酸序列(分别为SEQ ID NO:3和4)。

图6显示pSE52的构建。

图7显示质粒pJaL394的限制性图谱。

图8显示一种hemA缺失型等位基因的核苷酸序列和推导出的氨基酸序列(SEQ ID NO:15和16)。

发明详述

本发明涉及生产多肽的方法，包括：

(a) 将包括一个或多个第一核酸序列和一个第二核酸序列的一种核酸构建体导入一种细胞呼吸缺陷型突变细胞中，其中第一核酸序列在表达后可弥补呼吸缺陷，第二核酸序列编码所述多肽；

(b) 在适合第一和第二核酸序列表达的需氧条件下，于培养基中培养含有该第一和第二核酸序列的细胞；及

(c) 从细胞的培养基中分离多肽。

本发明方法基于利用氧化磷酸化作用在任何相关的工业发酵培养基中筛选细胞，其中唯一要求是细胞生长需要氧气。将一个或多个第一核酸序列导入呼吸缺陷型突变细胞中，可克服它的呼吸缺陷，并因此能基于细胞在氧气存在时的生长能力来筛选细胞。这样，那些含有一个或多个第一核酸序列的突变细胞能在培养基中培养，因而可从那些未发生成功转化的突变细胞中筛选出来。

术语“氧化磷酸化作用”是一个当电子通过一系列电子载体从NADH或FADH₂转移至O₂时形成ATP的过程。NADH和FADH₂在糖酵解、脂肪酸氧化及柠檬酸循环中形成，它们是高能分子，因为每一分子有一对具高转移势能的电子。氧化磷酸化作用是由呼吸链执行的，呼吸链位于例如真核生物的线粒体内膜或原核生物的细胞膜上。呼吸链中包括许多电子载体，例如细胞色素。电子通过这些载体从NADH或FADH₂逐步转移至O₂导致将质子泵出线粒体基质，产生一个质子动力。当质子通过一种酶复合体流回线粒体基质时可合成ATP。O₂充当唯一电子受体的ATP生成过程称为呼吸作用。不能通过这些载体转移电子的有机体是呼吸缺陷型的。

电子从NADH通过三种大蛋白质复合体的链转移至O₂，这三种蛋白质复合体被称为NADH-Q还原酶(也叫NADH脱氢酶)、细胞色素还原酶和

细胞色素氧化酶。这些酶中携带电子的基团是黄素、铁-硫簇、血红素和铜离子。电子从NADH-Q还原酶通过泛醌的还原形式传递至细胞色素还原酶。泛醌也可将电子从FADH₂携带至细胞色素还原酶。细胞色素C, 这种小蛋白质, 将电子从细胞色素还原酶转移至该链中的最后一个成分-细胞色素氧化酶。

第一个反应是由NADH-Q还原酶催化的NADH氧化作用, NADH-Q还原酶是一种黄素蛋白, 它包含一种紧密结合的黄素单核苷酸(FMN)作为辅基。两个电子从NADH传递至FMN, 产生还原形式, 即FMNH₂。然后电子从FMNH₂转移至一系列的铁-硫簇, 这是NADH-Q还原酶内的另一种辅基。NADH-Q还原酶的铁-硫簇内的电子接着转移至辅酶Q(又称泛醌)。泛醌是一种醌衍生物, 它有一长的类异戊二烯尾, 其取走一个电子则还原成泛醇(ubiquinol)。泛醇也是电子从FADH₂转移进入的位点。

然后泛醇将其两个电子之一转移至细胞色素还原酶内的铁-硫簇。细胞色素是一种电子转运蛋白, 其中包括一个血红素辅基。细胞色素还原酶的作用是催化电子从泛醇转移至细胞色素c。细胞色素还原酶中包含细胞色素b和c₁, 和一种铁-硫簇蛋白。细胞色素b、c₁和c的辅基是铁-原卟啉IX, 即通常所说的血红素。细胞色素b中, 血红素不与蛋白质共价键合, 而细胞色素c和c₁中, 血红素与蛋白质通过硫酯键共价连接。细胞色素a和a₃含有不同的铁卟啉辅基, 称为血红素A, 其与细胞色素c和c₁中的血红素不同在于其上一个甲酰基代替了一个甲基, 一条碳氢链代替了一个乙烯基。泛醇将其两个电子之一转移至还原酶的铁-硫复合体中, 再依次转移至细胞色素c₁和c, 将其带离复合体。细胞色素b使泛醇能与铁-硫簇相互作用。

细胞色素氧化酶由细胞色素a和a₃组成, 它催化电子从还原的细胞色素c转移至分子氧, 即最后的受体。

在本发明的方法中, 呼吸缺陷型突变细胞可以是一个或多个基因有缺陷而导致呼吸过程中一个或多个步骤有缺陷的任何细胞, 或是由于破坏了一个或多个基因而变成呼吸缺陷型的任何细胞。这些基因可包括但不限于编码作为电子传递链中一个成员的任何蛋白质的基

因，或者编码参与黄素生物合成、醌生物合成或血红素生物合成的酶的基因。

呼吸缺陷型突变体因缺乏功能性细胞色素而无氧化磷酸化作用。因此，破坏基因会整个消除专性需氧生物的生长。基因破坏后，呼吸缺陷型突变体实质上是一种营养缺陷型，它仅能在补充了可弥补缺陷并重建氧化磷酸化作用的合适物质的培养基中生长良好。

例如，呼吸缺陷型突变细胞也可以是其内在氧化磷酸化作用中起重要作用的一种调节基因或在相关途径中的一种基因(例如，编码介导血红素生物合成基因活化的激活子的基因)已被改变或破坏的细胞。例如，酿酒酵母粪卟啉原氧化酶基因的诱导在血红素缺乏下部分通过 Cyp19介导的活化作用进行(Amillet 等, 1995, 现代遗传学 28:503-511)。该基因的破坏可产生一种血红素缺陷型并最终是呼吸缺陷型的细胞。

此外，构建突变细胞可通过采用本领域中众所周知的方法，诸如在基因中插入、缺失或替换核酸残基等，破坏对于氧化磷酸化作用比较重要的基因进行。例如，这些基因之一可通过将含有一段与该基因同源的核酸片段的一个整合质粒插入该基因中、从而形成同源区的重复并在重复区之间掺入了载体DNA而进行破坏。如果插入的载体将该基因的启动子与编码区分开，或打断了编码序列以致产生无功能的基因产物，就会消除基因的表达。除此之外，可修饰对氧化磷酸化作用中比较重要的一个或多个基因的表达有利或必需的一个或多个控制序列(例如启动子)。或者，可通过基因转换方法(参见例如Iglesias和Trautner, 1983, 分子普通遗传学 189:73-76)或基因取代方法，降低或消除基因表达。在后一方法中，将该基因的突变形式与选择性标记结合导入至DNA的非复制性线性片段上。通过对标记的选择完成对质粒整合的选择。通过检查集落获得该突变基因的情况筛选出引起基因替代的重组事件。此外，降低或消除对氧化磷酸化作用比较重要的一个或多个基因的表达可以通过本领域中广为人知的随机诱变方法进行，包括但不限于转座和紫外线照射或化学剂诱变方法。

作为宿主细胞的呼吸缺陷型突变细胞的选择在很大程度上取决于一些因素。一种因素是在引入一个或多个第一核酸序列之前弥补突变细胞的呼吸缺陷、使突变细胞生长的能力。对突变细胞呼吸缺陷的互补可以通过在培养基中补加一种物质[例如一种血红素来源, 诸如亚铁原卟啉或高铁血红素(氯高铁血红素)]或者补加一种血红蛋白以克服涉及血红素生物合成的基因破坏问题而达到。然而, 这种选择应依赖于突变细胞被这些加入到培养基中的物质所拯救的能力。另一个主要因素是第一核酸序列能否在宿主细胞中表达。再一个因素是第二核酸序列能否在宿主细胞中表达。

宿主细胞可以是单细胞微生物, 例如原核生物, 或是一种非单细胞的微生物, 例如真核生物。

有用的单细胞是细菌细胞, 例如革兰氏阳性细菌, 包括但不限于芽孢杆菌属细胞, 例如嗜碱芽孢杆菌、液化淀粉芽孢杆菌、短芽孢杆菌、环状芽孢杆菌、凝结芽孢杆菌、灿烂芽孢杆菌、迟缓芽孢杆菌、地衣芽孢杆菌、巨大芽孢杆菌、嗜热脂肪芽孢杆菌、枯草芽孢杆菌和苏云金芽孢杆菌; 或链霉菌属细胞, 例如浅青紫链霉菌或鼠灰链霉菌; 或革兰氏阴性细菌, 诸如大肠杆菌和假单胞菌属细菌。一种优选的实施方案中, 该细菌宿主细胞是迟缓芽孢杆菌、地衣芽孢杆菌、嗜热脂肪芽孢杆菌或枯草芽孢杆菌细胞。

宿主细胞可以是一种真核细胞, 例如哺乳动物细胞, 昆虫细胞, 植物细胞或真菌细胞。有用的哺乳动物细胞包括中国仓鼠卵巢(CHO)细胞, Hela细胞, 幼仓鼠肾(BHK)细胞, COS细胞, 或其他可得自例如美国典型培养物保藏中心的永生化细胞系。

在一种优选的实施方案中, 宿主细胞是一种真菌细胞。此处所用的“真菌”包括子囊菌门(Ascomycota)、担子菌门(Basidiomycota)、壶菌门(Chytridiomycota)和接合菌门(Zygomycota)(Hawksworth等, Ainsworth和Bisby's Dictionary of The Fungi, 第8版, 1995, CAB International, University Press, 剑桥, 英国)以及卵菌门(Oomycota)(引用于Hawksworth等, 1995,

出处同前, 第171页)和所有有丝分裂孢子真菌(Hawksworth等, 1995, 出处同前)。子囊菌门的代表性组包括例如脉孢菌属、正青霉属(=青霉属)、*Emericella*(=曲霉属)、散囊菌属(=曲霉属), 和以下所列的真酵母。担子菌门的例子包括蘑菇、锈菌和黑粉菌。壶菌门的代表性组包括例如异水霉属、小芽枝霉属(*Blastocladiella*)、雕蚀菌属(*Coelomomyces*)和水生真菌。卵菌门的代表群包括例如水霉科(*Saprolegniomycetous*)水生真菌(水霉), 诸如绵霉属(*Achlya*)。有丝分裂孢子真菌的例子有曲霉属、青霉属、假丝酵母属和链格孢属。接合菌门的代表性群包括例如根霉属和毛霉属。

在一个优选的实施方案中, 真菌宿主细胞是一种酵母细胞。“酵母”用在这里包括子囊孢子酵母(内孢霉目)、担子孢子酵母, 和属于半知菌类(芽孢纲)的酵母。子囊孢子酵母分成蚀精霉科和酵母科两个科。后者包括四个亚科: 裂殖酵母亚科(*Schizosaccharomycoideae*)(例如裂殖酵母属)、拿逊酵母亚科(*Nadsonioideae*)、油脂酵母亚科(*Lipomycoideae*)和酵母亚科(*Saccharomycoideae*)(例如毕赤酵母属、克鲁维酵母属和酵母属)。担子孢子酵母包括白冬孢酵母属(*Leucosporidium*)、红冬孢酵母属(*Rhodospiridium*)、锁掷酵母属(*Sporidiobolus*)、*Filobasidium*和*Filobasidiella*。属于半知菌类的酵母分成两个科: 掷孢酵母科(*Sporobolomycetaceae*)(例如, *Sorobolomyces*属和布勒弹孢酵母属)和隐球酵母科(*Cryptococcaceae*)(例如假丝酵母属)。因为酵母的分类在将来可能会改变, 为了本发明的目的, 应按照《酵母的生物学和活性》(Skinner, F.A., Passmore, S.M., 和Davenport, R.R., 编, Soc. App. Bacteriol, 论文集系列No.9, 1980)中所述定义酵母。酵母的生物学和遗传学加工在本领域里已广为人知(见例如《酵母的生物学和遗传学》, Bacil, M., Horecker, B.J., Stopani, A.O.M. 编, 第2版, 1987; 《酵母》, Rose, A.H., 和Harrison, J.S., 编, 第2版, 1987; 《酵母属酵母的分子生物学》, Strathern等编, 1981)。

在一个较优选的实施方案中，酵母宿主细胞是假丝酵母属、克鲁维酵母属、酵母属、裂殖酵母属、毕赤酵母属或 *Yarrowia* 属内之种的细胞。

在一个最优选的实施方案中，酵母宿主细胞是卡尔酵母、酿酒酵母、糖化酵母、*Saccharomyces douglasii*、克鲁弗酵母、诺地酵母或 *Saccharomyces oviformis* 细胞。在另一个最优选的实施方案中，酵母宿主细胞是乳酸克鲁维酵母。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，真菌宿主细胞是一种丝状真菌细胞。

“丝状真菌”包括真菌门 (Eumycota) 和卵菌门亚支的所有丝状真菌形式 (由 Hawksworth 等定义, 1995, 出处同前)。丝状真菌的特征在于有由壳多糖、纤维素、葡聚糖、脱乙酰壳多糖、甘露聚糖和其他复杂多糖组成的菌丝壁。其营养生长是通过菌丝的延长进行的，碳分解代谢是专性需氧的。相比之下，酵母 (例如酿酒酵母) 的营养生长是通过单细胞菌体的出芽进行的，碳的分解代谢可能是发酵性的。在一个更优选的实施方案中，丝状真菌宿主细胞是下述种但不限于这些种的细胞：枝顶孢属 (*Acremonium*)、曲霉属、镰孢属、腐质霉属、毁丝霉属、毛霉属、脉孢菌属、青霉属、梭孢壳属、*Tolypocladium* 属及木霉属。

在一个甚至更优选的实施方案中，该丝状真菌宿主细胞是曲霉属细胞。在另一个更优选的实施方案中，该丝状真菌宿主细胞是枝顶孢属细胞。在另一个更优选的实施方案中，该丝状真菌宿主细胞是镰孢属细胞。在另一个更优选的实施方案中，该丝状真菌宿主细胞是腐质霉属细胞。在另一个更优选的实施方案中，该丝状真菌宿主细胞是毁丝霉属细胞。在另一个更优选的实施方案中，该丝状真菌宿主细胞是毛霉属细胞。在另一个更优选的实施方案中，该丝状真菌宿主细胞是脉孢菌属细胞。在另一个更优选的实施方案中，该丝状真菌宿主细胞是青霉属细胞。在另一个更优选的实施方案中，该丝状真菌宿主细胞是梭孢壳属细胞。在另一个更优选的实施方案中，该丝状真菌宿主细

胞是 *Tolypocladium* 属细胞。在另一个更优选的实施方案中，该丝状真菌宿主细胞是木霉属细胞。

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细胞、臭曲霉细胞、日本曲霉细胞、黑曲霉细胞或米曲霉细胞。在另一个最优选的实施方案中，该丝状真菌宿主细胞是 *Fusarium cerealis*, *Fusarium crookwellense*、尖镰孢、禾本科镰孢、*Fusarium sambucinum*, *Fusarium sulphureum*, 或 *Fusarium venenatum* 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 *Humicola insolens* 细胞或 *Humicola lanuginosa* 细胞。在另一个更优选的实施方案中，该丝状真菌宿主细胞是嗜热毁丝霉 (*Myceliophthora thermophilum*) 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米黑毛霉 (*Mucor miehei*) 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粗糙脉孢菌细胞。在另一个最优选的实施方案中，该丝状真菌宿主细胞是产紫青霉细胞。在另一个最优选的实施方案中，该丝状真菌宿主细胞是 *Thielavia terrestris* 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 *Trichoderma harzianum* 细胞，康宁木霉细胞、*Trichoderma longibrachiatum* 细胞、*Trichoderma reesei* 细胞，或绿色木霉细胞。

在一个优选实施方案中，将编码作为电子传递链中组分、使突变体能弥补呼吸缺陷的一种或多种蛋白质的一或多个第一核酸序列导入突变体中。这些蛋白质包括 NADH-Q 还原酶、细胞色素还原酶 (包括细胞色素 b 和 c₁)、细胞色素 c 和细胞色素氧化酶 (包括细胞色素 a 和 a₃)。

在另一优选的实施方案中，将编码使突变体能产生黄素的一种或多种黄素生物合成酶的一个或多个第一核酸序列导入突变体中。

在一个优选实施方案中，向突变体中导入编码一种或多种使突变体能产生泛醌的泛醌 (CoQ) 生物合成酶的一或多个第一核酸序列。

在另一个优选实施方案中，将编码使突变体能产生血红素的一种或多种血红素生物合成酶的一或多个第一核酸序列导入突变体中。血红素是原卟啉 IX 与铁的整合复合物，可作为血红素蛋白的辅基

(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，获得一个铁原子后形成血红素。

术语“血红素生物合成酶”在此处是指参与血红素生物合成的任何一种酶。这类血红素生物合成酶的例子包括但不限于：5-氨基- γ -酮戊酸合成酶(5-aminolevulinic acid synthase)(EC2.3.1.37)、胆色素原合成酶(EC4.2.1.24)、胆色素原脱氨基酶(EC4.3.1.8)、尿卟啉原III合成酶(EC4.2.11.75)、尿卟啉原III脱羧酶(EC4.1.1.37)、粪卟啉原III氧化酶(EC1.3.3.3)、原卟啉原IX氧化酶(EC1.3.3.4)以及亚铁螯合酶(EC4.99.1.1)。5-氨基- γ -酮戊酸合成酶催化甘氨酸和琥珀酰辅酶A缩合形成5-氨基- γ -酮戊酸(5-aminolevulinic acid)。胆色素原合成酶(也称为5-氨基- γ -酮戊酸脱水酶)可催化两个分子的5-氨基- γ -酮戊酸缩合形成胆色素原。胆色素原脱氨基酶(也称为羟甲基胆色烷合成酶或uro I 合成酶)可催化吡咯胆色素原四聚体化形成前尿卟啉原。尿卟啉原III合成酶(也称为uro III 合成酶或uro III 共合成酶)可催化前尿卟啉原第4个环的重排，随后通过环化形成尿卟啉原III。尿卟啉原III脱羧酶(也称为uro D 或尿卟啉原脱羧酶)可催化尿卟啉原III的全部4条乙酸侧链的脱羧反应形成甲基，从而产生粪卟啉原III。粪卟啉原III氧化酶(也称为粪卟啉原酶)可催化粪卟啉原III的A、B环上第2和4位的2个丙酸基发生氧化性脱羧反应形成乙烯基，从而产生原卟啉原IX。原卟啉原IX氧化酶可催化原卟啉原IX发生六电子氧化生成原卟啉IX。亚铁螯合酶(也称为亚铁裂解酶、血红素合成酶或原血红素亚铁裂解酶)可催化将铁插入到原卟啉中，生成血红素。

5-氨基- γ -酮戊酸的生物合成也可通过tRNA-Glu介导的途径由谷氨酸进行。在许多厌氧型细菌、植物及能进行光合成的生物体中都发现有谷氨酸途径。由谷氨酸生物合成5-氨基- γ -酮戊酸包括三个酶促步骤，它们由谷氨酸tRNA_{glu}合成酶(EC6.1.1.17)、谷氨酸tRNA_{glu}还原酶及谷氨酸1-半醛氨基转移酶(glutamate 1-semialdehyde

aminotransferase)催化。谷氨酸tRNA_{glu}合成酶(也称作谷氨酸tRNA_{glu}合成酶)在ATP和镁离子的存在下催化谷氨酸与tRNA的偶联。谷氨酸tRNA_{glu}还原酶(也称作谷氨酸-tRNA_{glu}脱氢酶)在NADPH的存在下催化与tRNA结合的谷氨酸发生还原,形成谷氨酸1-半醛。谷氨酸1-半醛氨基转移酶(也称作谷氨酸1-半醛氨基变位酶)催化谷氨酸1-半醛转化为5-氨基- γ -酮戊酸。

该第一核酸序列优选是编码选自下述的血红素生物合成酶的序列: 5-氨基- γ -酮戊酸合成酶、胆色素原合成酶、胆色素原脱氢酶、尿卟啉原合成酶、尿卟啉原脱羧酶、粪卟啉原氧化酶、原卟啉原氧化酶、亚铁螯合酶、谷氨酸tRNA_{glu}合成酶、谷氨酸tRNA_{glu}还原酶、谷氨酸1-半醛氨基转移酶。第一核酸序列可来源于任何微生物,可以与宿主细胞同源或异源。第一核酸序列来源的选择取决于待用作宿主细胞的血红素缺陷型突变细胞,但优选真菌来源,如酵母和丝状真菌。优选的丝状真菌来源包括但不限于枝顶孢属、曲霉属、镰孢属、腐质霉属、毁丝霉属、毛霉属、脉孢菌属、青霉属、*Phanerochaete*属、梭孢壳属、*Tolypocladium*属及木霉属。优选的酵母来源包括但不限于假丝酵母属、克鲁维酵母属、毕赤酵母属、酵母属、裂殖酵母属和*Yarrowia*属内之种。

编码血红素生物合成酶的第一核酸序列可以是下述序列之一:

1. 5-氨基- γ -酮戊酸合成酶基因:

a. 酿酒酵母(Urban-Grimal等, 1986, 欧洲生物化学杂志, 156: 511-59);

b. 构巢曲霉(Bradshaw等, 1993, 现代遗传学, 23: 501-507);

c. 类球红细菌(Tai等, 1988, 基因, 70: 139-152);

d. 荚膜红细菌(Hornberger等, 1990, 分子普通遗传学, 211: 371-378); 以及

e. 大肠杆菌(Drolet等, 1989, 分子普通遗传学, 216: 347-352)。

2. 胆色素原合成酶基因:

a. 酿酒酵母 (Myers等, 1987, 生物化学杂志, 262: 16822-16829);

b. 金黄色葡萄球菌 (Kafala和Sasarman, 1994, 加拿大微生物学杂志, 40: 651-657);

c. 类球红细菌 (Delaunay等, 1991, 细菌学杂志, 173: 2712-2715);

d. 大肠杆菌 (Echelard等, 1988, 分子普通遗传学, 214, 503-508); 以及

e. 枯草芽孢杆菌 (Hansson等, 1991, 细菌学杂志, 173: 2590-2599)。

3. 胆色素原脱氨基酶基因:

a. 酿酒酵母 (Keng等, 1992, 分子普通遗传学, 234: 33-433);

b. 人 (Yoo等, 1993, 基因组, 15: 221-29; Raich等, 1986, 核酸研究, 14: 5955-5968);

c. 大肠杆菌 (Thomas和Jordan, 1986, 核酸研究, 14: 6215-6226); 以及

d. 枯草芽孢杆菌 (Petricek等, 1990, 细菌学杂志, 172: 2250-2258)。

4. 尿卟啉原III合成酶基因:

a. 酿酒酵母 (Amillet和Labbe-Bois, 1995, 酵母, 11: 419-424);

b. 枯草芽孢杆菌 (Hansson等, 1991, 细菌学杂志, 173: 2590-2599); 以及

c. 大肠杆菌 (Jordan等, 1987, 核酸研究, 15: 10583)。

5. 尿卟啉原III脱羧酶基因:

a. 酿酒酵母 (Garey等, 1992, 欧洲生物化学杂志, 205: 1011-1016); 以及

b. 人(Romeo等, 1986, 生物化学杂志, 261: 9825-9831)。

6. 粪卟啉原III氧化酶基因:

a. 人(Martasek等, 1994, 美国国家科学院院报, 911: 3024-3028);

b. 大肠杆菌(Troup等, 1994, 细菌学杂志, 176: 673-680);
以及

c. 酿酒酵母(zaagorec等, 1986, 生物化学杂志, 263: 9718-9724)。

7. 原卟啉原IX氧化酶基因:

a. 人(Taketani等, 1995, 基因组, 29: 698-703);

b. 枯草芽孢杆菌(Dailey等, 1994, 生物化学杂志, 269: 813-815); 以及

c. 大肠杆菌(Sasarman等, 1993, 加拿大微生物学杂志, 39: 155-161)。

8. 亚铁螯合酶基因:

a. 酿酒酵母(Labbe-Bois, 1990, 生物化学杂志, 265: 7278-72883);

b. 牛(Shibuya等, 1995, 生物化学与生物物理学学报, 1231: 117-120);

c. 大豆慢生根瘤菌(Frustaci和O'Brian, 1993, 应用环境微生物学, 59: 2347-2351);

d. 大肠杆菌(Frustaci和O'Brian, 1993, 细菌学杂志, 175: 2154-2156); 以及

e. 枯草芽孢杆菌(Hansson和Hederstedt, 1992, 细菌学杂志, 174: 8081-88093)。

9. 谷氨酸-tRNA_{glu}合成酶基因

a. 热自养甲烷杆菌(Moore等, 1996, 生物化学与生物物理学学报, 1305: 113-116)。

b. 嗜热栖热菌(Nureki等, 1992, 欧洲生物化学杂志,

204:465-472)。

c. 枯草芽孢杆菌 (Breton 等, 1990, 生物化学杂志, 265:18248-18255)。

d. 嗜热脂肪芽孢杆菌 (Breton 等, 1990, 生物化学杂志, 265:18248-18255)。

e. 酿酒酵母 (Ludmerer 和 Schimmel, 1985, 细菌学杂志, 163:763-768)。

f. 大肠杆菌 (Breton 等, 1986, 生物化学杂志, 261:10610-10617)。

10. 谷氨酸 tRNA_{glu} 还原酶基因

a. 大麦 (Vothknecht 等, 1996, 美国国家科学院院报, 93:9287-9291)。

b. 铜绿假单胞菌 (Hungerer 等, 1995, 分子与普通遗传学, 248:375-380)。

c. 大肠杆菌 (Ikemi 等, 1992, 基因, 121:127-132)。

d. 枯草芽孢杆菌 (Petricek 等, 1990, 细菌学杂志, 172:2250-2258)。

e. 热自养甲烷杆菌 (Hungerer 等, 1996, Bioorg. Med. Chem. 4:1089-1095)。

11. 谷氨酸 1-半醛氨基转移酶基因

a. 鼠伤寒沙门氏菌 (Elliott 等, 1990, 细菌学杂志, 172:7071-7084)。

b. 大麦 (Grimm, 1990, 美国国家科学院院报, 87:4169-4173)。

c. 大肠杆菌 (Ikemi 等, 1992, 基因, 121:127-132)。

在一个更优选的实施方案中, 所说的第一核酸序列得自曲霉属之种。在一个甚至更优选的实施方案中, 所说的第一核酸序列得自无花果曲霉、臭曲霉、日本曲霉、黑曲霉、构巢曲霉或米曲霉。在另一个更优选的实施方案中, 第一核酸序列得自酵母属之种。在一个甚至更优选的实施方案中, 第一核酸序列得自酿酒酵母、卡尔酵母、糖化酵

母、*Saccharomyces douglasi*, 克鲁弗酵母、诺地酵母或 *Saccharomyces oviformis*.

在一个最优选的实施方案中, 编码5-氨基- γ -酮戊酸合成酶的第一核酸序列得自米曲霉A1560菌株(IF04177), 如SEQ ID No:1所示的核酸序列。该第一核酸序列也可以是编码具有SEQ ID No:2所示氨基酸序列的5-氨基- γ -酮戊酸合成酶的核酸序列, 该核酸序列由于遗传密码简并性而与SEQ ID No:1有所不同。在另一个最优选的实施方案中, 编码胆色素原合成酶的第一核酸序列得自米曲霉A1560菌株(IF04177), 如SEQ ID No:3所示的核酸序列。该第一核酸还可以是编码具有SEQ ID No:4所示氨基酸序列的胆色素原合成酶的核酸序列, 该核酸序列由于遗传密码简并性而与SEQ ID No:3有所不同。本发明的第一核酸序列还包括由SEQ ID No:1和SEQ ID No:3所示的基因组序列及相应的cDNA和RNA序列。应理解本文所用词语“核酸序列”包括合成DNA在内的所有这类变体。

为努力使质粒拷贝数最多, 可通过例如截短/突变启动子以降低转录水平, 使第一核酸序列的表达受到阻遏, 这样, 需要整合多拷贝的载体, 以支持生物体在无血红素或血红素前体存在下的快速生长。可以采用能降低所导入的第一核酸序列转录的任何突变。此外, 可以采用第一核酸序列所表达蛋白质的氨基酸序列中能使得该蛋白质活性有所降低的突变, 来使拷贝数尽量多。例如, 可以采用能使所导入的血红素生物合成酶基因的酶促活性降低的任何突变。

编码多肽的第二核酸序列可以是突变细胞的天然或异源序列。若是异源多肽, 则编码该异源多肽的核酸序列可来源于任何原核生物、真核生物, 或者其它来源, 如古生细菌(*archaeobacteria*)。此处术语“多肽”并非指特定长度的编码产物, 因而包括肽、寡肽和蛋白质。术语“多肽”还包括两或多个多肽结合在一起形成的编码产物。多肽还包括杂合多肽, 其中含有得自至少两种不同多肽(其中一或多种多肽是宿主细胞的异源多肽)的部分或全部多肽序列的组合。多肽还包括上述多肽天然存在的等位基因变体, 以及基因工程化变体。

多肽可以是酶、激素、激素变体、受体或其部分、抗体或其部分，或者报道分子。在一个更优选的实施方案中，多肽是氧化还原酶、转移酶、水解酶、裂解酶、异构酶，或连接酶。在一个还要更优选的实施方案中，多肽是氨肽酶、淀粉酶、糖酶、羧肽酶、过氧化氢酶、纤维素酶、壳多糖酶、环糊精糖基转移酶、角质酶、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、谷氨酰胺酶、盐过氧化物酶(haloperoxidase)、转化酶(invertase)、漆酶、脂酶、甘露糖苷酶、齿斑葡聚糖酶(mutanase)、氧化酶、果胶裂解酶、过氧化物酶、肌醇六磷酸酶、多酚氧化酶、蛋白水解酶、核糖核酸酶、转谷氨酰胺酶(transglutaminase)或木聚糖酶。

在术语“天然多肽”的范围内，本发明的方法还包括突变细胞的天然或内源多肽的重组产生方法，其中这种表达涉及非宿主细胞天然遗传因子的应用，或者已被加工而以非宿主细胞内正常存在的方式起作用的天然因子的应用。例如，可以通过将编码多肽的基因置于不同启动子的控制之下，以增强多肽的表达，或者将其置于不同信号序列的控制之下，以使天然多肽运输到细胞外，以及提高编码宿主细胞通常所产生的多肽的基因拷贝数，从而对天然多肽进行重组表达。

多肽也可以是血红素蛋白。本文将“血红素蛋白”定义为含有血红素作为辅基的一组蛋白质中的任一个成员。血红素蛋白可以是含有血红素作为辅基的珠蛋白、细胞色素、氧化还原酶或任何其它蛋白质。含有血红素的珠蛋白包括血红蛋白和肌红蛋白。含有血红素的细胞色素包括细胞色素P450、细胞色素b、细胞色素c1和细胞色素c。含有血红素的氧化还原酶包括但不限于过氧化氢酶、氧化酶、加氧酶、盐过氧化物酶(haloperoxidase)以及过氧化物酶。在一个优选实施方案中，所述氧化还原酶是过氧化氢酶。在另一个优选实施方案中，该氧化还原酶是氧化酶。在又一个优选实施方案中，该氧化还原酶是加氧酶。在又一个优选实施方案中，该氧化还原酶是盐过氧化物酶。在又一个优选实施方案中，该氧化还原酶是一种过氧化物酶。在

一更优选的实施方案中，所述过氧化物酶得自于鬼伞属菌株、*Arthromyces*属菌株或*Phanerochaete*属菌株。在一个甚至更优选的实施方案中，该过氧化物酶得自于灰盖鬼伞菌株如灰盖鬼伞 IF08371、长根鬼伞菌株，或*Arthromyces ramosus*菌株。在另一个更优选的实施方案中，所述过氧化氢酶得自于*Scytalidium*属菌株、曲霉属菌株或腐质霉属菌株。在另一个甚至更优选的实施方案中，该过氧化氢酶得自于*Scytalidium thermophilum*菌株如*Scytalidium thermophilum* CBS 117.65，黑曲霉菌株或*Humicola insolens*菌株。

第一和第二核酸序列优选包含在一个核酸构建体内。本文中术语“核酸构建体”意指单链或双链核酸分子，它分离自天然存在的基因，或是已被修饰而包含了以自然界并不存在的方式联合并置的数段核酸。当核酸构建体含有表达编码序列所需所有控制序列时，术语“核酸构建体”与“表达盒”同义。本文中术语“编码序列”定义为一旦置于上述控制序列的控制下即可转录成mRNA并翻译成多肽的序列。编码序列的边界确定为5'-末端翻译起始密码子ATG和3'-末端的翻译终止密码子。编码序列包括但不限于DNA、cDNA及重组核酸序列。

可以多种方式加工第一核酸序列，使之表达酶，或降低其表达，以加大第一核酸序列的拷贝数。在插入至载体之前加工核酸序列比较理想，或是必需的，这取决于表达载体。本领域中熟知利用诸如克隆方法和诱变对核酸序列进行修饰的技术。

本文中术语“控制序列”是指包括第一核酸序列编码序列的表达所必需的或有益的所有组分。控制序列对第一核酸序列而言可以是天然的，或可得自其它来源，或可以是天然和外源控制序列的组合。外源控制序列可简单替换或添加到天然控制序列中，以获得与编码序列正常相连的天然控制序列相比更高水平的第一核酸序列表达。这类控制序列包括但不限于前导序列、多聚腺苷酸化序列、前肽序列、启动子、信号序列以及转录终止子。为了能在控制序列的指导下进行表

达, 将根据本发明所用的第一核酸序列与控制序列可操作连接, 从而在与控制序列相容的条件下获得第一核酸序列编码序列的表达。本文中术语“编码序列”定义为一旦置于上述控制序列的控制下即可转录成mRNA并翻译成产物的序列。编码序列的边界由5'-末端翻译起始密码子和3'-末端的翻译终止密码子确定。编码序列包括但不限于DNA、cDNA及重组核酸序列。

第一控制序列可以是适当的启动子序列, 即可被宿主细胞识别而用于第一核酸序列表达的核酸序列。启动子序列含有可介导血红素生物合成酶表达的转录控制序列。该启动子可以是在选定的宿主细胞内具有转录活性的任何启动子序列, 包括突变的、截短的或杂合的启动子, 可得自与宿主细胞同源或异源的基因。

能指导本发明核酸构建体转录(特别是在细菌宿主细胞内)的合适启动子的例子是来源于以下的启动子: 大肠杆菌lac操纵子, 天蓝色链霉菌琼脂糖酶基因(dagA), 枯草芽孢杆菌左聚蔗糖酶(levansucrase)基因(sacB), 地衣芽孢杆菌 α -淀粉酶基因(amyL), 嗜热脂肪芽孢杆菌麦芽糖源淀粉酶基因(amyM), 解淀粉芽孢杆菌 α -淀粉酶基因(amyQ), 地衣芽孢杆菌青霉素酶基因(penP), 枯草芽孢杆菌xylA和xylB基因, 和原核 β -内酰胺酶基因(Villa-Kamaroff等, 1978, 美国国家科学院学报 75:3727-3731), 及tac启动子(DeBoer等, 1983, 美国国家科学院学报 80:21-25)。更多的启动区在下面文章中有描述: “来源于重组细菌的有用蛋白质”, 科学美国人, 1980, 242:74-94; 和Sambrook等, 分子克隆, 实验室手册, 第2版, 冷泉港, NY, 1989。

可在丝状真菌宿主细胞中指导第一核酸序列转录的适当启动子的实例有得自编码米曲霉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、构巢曲霉乙酰胺酶、尖镰孢胰蛋白酶样蛋白酶(美国专利4,288,627所述, 引入本文作为参考)等

的基因的启动子，或其突变的、截短的和杂合的启动子。特别优选用于丝状真菌宿主细胞的启动子是TAKA淀粉酶的启动子、NA2-tpi（来源于黑曲霉中性 α -淀粉酶和米曲霉磷酸丙糖异构酶编码基因的启动子的杂合体）以及glaA启动子。

在酵母宿主中，有用的启动子可得自酿酒酵母烯醇酶(ENO-1)基因、酿酒酵母半乳糖激酶基因(GAL1)、酿酒酵母乙醇脱氢酶/甘油醛-3-磷酸脱氢酶基因(ADH2/GAP)以及酿酒酵母3-磷酸甘油酸激酶基因。Romanos等，1992，酵母(Yeast)8:423-488中描述了酵母宿主细胞的其它有用启动子。在哺乳动物宿主细胞中，有用的启动子包括病毒启动子，如得自猴病毒40(SV40)、劳氏肉瘤病毒(RSV)、腺病毒及牛乳头瘤病毒(BPV)的启动子。

第一控制序列也可以是适当的转录终止序列，即可被宿主细胞识别而终止转录的一段序列。该终止序列与第一核酸序列的3'末端可操作连接。终止序列可能对第一核酸序列来说是天然的，也可以得自其它来源，即是一种外源终止序列。在选定的宿主细胞内有效的任何终止子在本发明中很可能有用处。

优选用于丝状真菌宿主细胞的终止子得自编码米曲霉TAKA淀粉酶、黑曲霉葡糖淀粉酶、构巢曲霉邻氨基苯甲酸盐合成酶、黑曲霉 α -葡糖苷酶及尖镰孢胰蛋白酶样蛋白酶的基因的终止子。

酵母宿主细胞的优选终止子得自编码酿酒酵母烯醇酶、酿酒酵母细胞色素C(CYC1)或酿酒酵母甘油醛-3-磷酸脱氢酶的基因。Romanos等，1992(出处同前)中描述了酵母宿主细胞的其它有用终止子。哺乳动物宿主细胞的终止子序列在本领域中是众所周知的。

该第一控制序列也可能是适当的前导序列，即mRNA中对宿主细胞的翻译具有重要作用的非翻译区。该前导序列与第一核酸序列的5'末端可操作连接。前导序列可能对第一核酸序列是天然的，或可得自其它来源，即是外源前导序列。在选定的宿主细胞内有功能的任何前导序列很可能在本发明中也是有用的。

丝状真菌宿主细胞的优选前导序列得自编码米曲霉TAKA淀粉酶和

米曲霉磷酸丙糖异构酶的基因。

酵母宿主细胞的合适前导序列得自酿酒酵母烯醇酶(ENO-1)基因、酿酒酵母3-磷酸甘油酸激酶基因、酿酒酵母 α -因子及酿酒酵母乙醇脱氢酶/甘油醛-3-磷酸脱氢酶基因(ADH2/GAP)。

该第一控制序列也可以是多聚腺苷酸化序列，即与第一核酸序列的3'末端可操作连接、并且转录时能被宿主细胞识别以将聚腺苷酸残基加到转录mRNA中的一段序列。多聚腺苷酸化序列可能对第一核酸序列来说是天然的，或可得自其它来源，即是外源多聚腺苷酸化序列。在选定的宿主细胞中有功能的任何多聚腺苷酸化序列很可能在本发明中也是有用的。

丝状真菌宿主细胞的优选多聚腺苷酸化序列得自编码米曲霉TAKA淀粉酶、黑曲霉葡糖淀粉酶、构巢曲霉邻氨基苯甲酸盐合成酶及黑曲霉 α -葡糖苷酶的基因。

Guo和Sherman, 1995, 分子细胞生物学(Molecular Cellular Biology) 15:5983-5990中描述了酵母宿主细胞的有用多聚腺苷酸化序列。哺乳动物宿主细胞可用的多聚腺苷酸化序列在本领域中是众所周知的。

第一控制序列也可以是信号肽编码区，其编码与所表达蛋白质氨基末端相连的氨基酸序列，从而将该蛋白质定位于特定细胞区室中。该信号肽编码区可能对第一核酸序列来说是天然的，或可以得自外源。第一核酸序列编码序列的5'末端本身可能含有信号肽编码区，该信号肽编码区与编码该定位蛋白质的编码区的片段在翻译读框内自然连接。或者，该编码序列的5'末端可含有该定位蛋白质编码序列部分的外源信号肽编码区的核酸。能够在选定的宿主细胞中促使该蛋白质定位的任何信号肽编码区在本发明中均有可能使用。

第一控制序列也可以是编码位于成熟并具生物化学活性的多肽氨基末端的氨基酸序列的肽原编码区，所获得的多肽公知为酶原或多肽原(或在某些情形下为酶原)。酶原通常是没有活性的，通过肽原的催化或自身催化切割可从酶原转变为成熟的活性多肽。本文将生物化

学活性的多肽限定为以可发挥其天然对应物的生物化学活性的活性形式生产的多肽。该肽原序列可以是编码血红蛋白生物合成酶的第一核酸序列的天然序列，也可得自其它来源，即是一种外源肽原序列。编码肽原的核酸序列可得自编码枯草芽孢杆菌碱性蛋白酶(aprE)、枯草芽孢杆菌中性蛋白酶(nprT)、酿酒酵母 α -因子和嗜热毁丝霉(*Myceliphthora thermophilum*)漆酶的基因。

在一个优选实施方案中，将与一个或多个第二控制序列可操作连接的第二核酸序列导入宿主细胞中。该第二控制序列可以是编码多肽的第二核酸序列的天然序列，也可以部分或全部来源于外源。外源控制序列可简单地替换天然控制序列，以获得相对于天然控制序列与编码序列正常相连时更高产量的目的多肽。第二控制序列可以是上述有关第一控制序列时例举的控制序列中的任何一种。

有多种信号肽编码区可用于将表达多肽引入细胞的分泌途径。细菌宿主细胞中有效的信号肽编码区是来源于以下基因的信号肽编码区：芽孢杆菌NCIB 11837的麦芽糖源淀粉酶基因，嗜热脂肪芽孢杆菌 α -淀粉酶基因，地衣芽孢杆菌枯草蛋白酶基因，地衣芽孢杆菌 β -内酰胺酶基因，嗜热脂肪芽孢杆菌中性蛋白酶基因(nprT, nprS, nprM)，和枯草芽孢杆菌prsA基因。更多的信号肽描述可见Simonen和Palva, 1993, 微生物学评论 57:109-137。

可用于丝状真菌宿主细胞的有效信号肽编码区可得自于米曲霉TAKA淀粉酶基因、黑曲霉中性淀粉酶基因、*Rhizomucor miehei*天冬氨酸蛋白酶基因、*Humicola Lanuginosa*纤维素酶基因或*Rhizomucor miehei*脂肪酶基因。

酵母宿主细胞的有用信号肽可得自酿酒酵母 α 因子和酿酒酵母转化酶编码基因。Romanos等(1992, 出处同前)描述了有用的其它信号肽编码区。

可利用本领域众所周知的方法，将第一和第二核酸序列导入突变体宿主细胞内。例如，可通过同源或非同源重组将序列导入并整合到宿主基因组内，其中一或多拷贝的所述序列被整合到了单一靶序列和

/或多处靶序列内。或者，可将所述序列作为非整合性表达载体（如自主复制性染色体外质粒）导入进去并维持下来。

第一和第二核酸序列优选包含在同一核酸构建体中，迫使同时对两种核酸序列进行筛选，然而，它们也可包含在不同的核酸构建体中。每一种核酸构建体都可含有整合元件，以指导通过同源重组而整合至宿主细胞基因组的精确位点。为了提高在精确位点发生整合的可能性，整合元件优选地应含有足够数目的核酸，如100-1500个碱基对，优选400-1500个碱基对，最优选为800-1500个碱基对，该核酸与相应的靶序列高度同源以提高同源重组的概率。整合元件可以是与宿主细胞基因组内的靶序列同源的任何序列。另外，整合元件可以是非编码核酸序列或编码核酸序列。另一方面，也可通过非同源重组将每种核酸构建体整合至宿主细胞的基因组中。

可将核酸构建体插入适当的载体中，或可将第一和第二核酸序列的编码区直接插入已含有控制序列的载体中。第一和第二核酸序列或者构建体也可包含在分开的载体中。这些载体可以是能便利地进行重组DNA操作并能导致本发明核酸序列表达的任何一种载体。载体的选择通常取决于该载体与欲导入该载体的宿主细胞之间的相容性。载体可以是自主复制型载体，即在染色体外完整存在、并能独立于染色体复制而进行复制的载体，诸如质粒、染色体外元件、微小染色体或人工染色体。或者，载体也可以是导入细胞时能整合至基因组中并与其整合入的染色体一起复制的一种载体。载体系统可以是单个载体或质粒，也可以是合在一起含有待导入宿主细胞基因组内的总DNA的两个或多个载体。

为了进行自主复制，载体中可进一步含有能使载体在目的宿主细胞内自主复制的复制起点。细菌复制起点的例子有：能使得在大肠杆菌内进行复制的质粒pBR322、pUC19、pACYC177和pACYC184的复制起点，能使得在芽孢杆菌内进行复制的pUB110、pE194、pTA1060和pAM β 1的复制起点。可在酵母宿主细胞内使用的复制起点的例子有2 μ 复制起点、CEN6与ARS4的组合、CEN3与ARS1的组合。复制起点内可以具

有能使其在宿主细胞内以温度敏感的方式起作用的突变(参见例如 Ehrlich, 1978, 美国国家科学院学报(PNAS), 75:1433)。

本领域普通技术人员熟知用于连接核酸构建体、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, 1989, 出处同前)。

细菌宿主细胞的转化可以通过例如原生质体转化(参见例如Chang和Cohen, 1979, 分子普通遗传学, 168: 111-115)、利用感受态细胞(参见例如Young和Spizizin, 1961, 细菌学杂志, 81:823-829, 或者Dubnau和Davidoff-Abelson, 1971, 分子生物学杂志, 56:209-221)、电穿孔(参见例如Shigekawa和Dower, 1988, 生物技术, 6:742-751)或者接合(参见例如Koehler和Thorne, 1987, 细菌学杂志, 169:5771-5278)进行。

本领域中将核酸序列导入丝状真菌细胞的标准方法包括以已知方式进行的原生质体形成、原生质体转化以及转化原生质体的细胞壁再生。EP238023和Yelton等, 1984, 美国国家科学院学报, 81:1470-1474中描述了转化曲霉属宿主细胞的合适操作。Malardier等, 1989, 基因, 78: 147-156或共同未决的美国流水号08/269449中描述了转化镰孢的合适方法。

可利用Becker和Guarente描述于酶学方法, 194卷, 酵母遗传学与分子生物学指南(Abelson, J.N. 和Simon, M.I. 编), 182-187页(Academic Press, Inc., New York); 及Ito等, 1983, 细菌学杂志, 153:163;和Hinnen等, 1978, 美国国家科学院学报, 75:1920中记载的操作转化酵母。

可利用Graham和Van der Eb(1978, 病毒学, 52:546)的磷酸钙沉淀法, 通过直接摄取而转化哺乳动物细胞。

在本发明方法中, 利用本领域已知方法, 在适于生产多肽和血红素生物合成酶的营养培养基中培养细胞。例如, 可在实验室或工业用发酵罐内, 在合适的培养基中、允许表达和/或分离该多肽的条件下经摇瓶培养、小规模或大规模发酵(包括连续、分批、分批补料或固

相发酵)培养该细胞。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), 在含碳、氮源和无机盐的适宜营养培养基中进行的。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。若多肽被分泌至营养培养基中, 则可直接从培养基中回收该多肽。若多肽不被分泌, 则可从细胞裂解物中回收。

可利用本领域熟知的、特异于多肽的方法检测多肽。这些检测方法包括特异抗体的使用、酶产物的形成或酶底物的消失。例如, 若多肽有酶活性, 则可利用酶测定法确定其活性。

可通过本领域已知方法回收所产生的多肽。例如, 可通过包括但不限于离心、过滤、提取、喷雾干燥、蒸发或沉淀等常规方法, 从营养培养基中回收多肽。然后可通过一系列层析方法如离子交换层析、凝胶过滤层析、亲和层析等对所回收的多肽进行进一步纯化。

本发明涉及破坏真菌细胞的呼吸缺陷型突变体内的基因的方法, 包括:

(a) 通过将核酸构建体导入呼吸缺陷型突变细胞中的基因或其部分内而取代该基因, 所述核酸构建体含有包含该基因或其部分的第三核酸序列及第一核酸序列, 其中第一核酸序列表达后弥补了该呼吸缺陷型突变细胞的呼吸缺陷; 和

(b) 在适于第一核酸序列表达的需氧条件下, 于培养基中培养含该第一和第三核酸序列的细胞。

本发明还涉及破坏血红素缺陷型突变细胞内基因的方法。该方法包括:

(a) 通过将一种核酸构建体导入血红素缺陷型突变细胞内的基因或其部分内而取代该基因, 所述核酸构建体含有包含该基因或其部分的第三核酸序列及编码一种血红素生物合成酶的第一核酸序列, 导入第一核酸序列后会导致突变体能产生血红素; 和

(b) 在适于第一核酸序列表达的条件下, 于无血红素和血红素中间体的培养基中培养含该第一和第三核酸序列的细胞。

细胞内源基因的破坏可通过本领域众所周知的方法完成, 如插入、缺失或取代基因内的核酸残基, 例如上述的那些方法。第三核酸序列可以是任何序列(需要用来产生缺陷或被标记的等位基因), 其通过整合至目标基因内可破坏该基因, 从而降低或消除其表达。第三核酸序列可以是上述任一种第二核酸序列的亚序列。为了确保整合至内源基因内, 构建体中应含有前述的整合元件。

本发明还涉及含有第一核酸序列的细胞的呼吸缺陷型突变体, 其

内至少一种氧化磷酸化必需基因具有修饰，其中，相同条件下培养时，突变体相比于所述细胞是呼吸缺陷型的。在一个优选实施方案中，该呼吸缺陷型突变体是血红素缺陷型的。

本发明还涉及获得细胞的呼吸缺陷型突变体的方法，所述方法包括：(a)向细胞内导入一种核酸序列，该核酸序列内含有至少一种氧化磷酸化必需基因的修饰；以及(b)从步骤(a)中鉴定出所述细胞含该核酸序列的突变体，其中突变体在与所述细胞相同的条件下培养时是呼吸缺陷型的。在一个优选实施方案中，该突变体是血红素缺陷型的。

本发明由下述实施例进一步描述，但这些实施例不应理解为是对本发明范围的限制。

实施例

实施例1：提取米曲霉A1560菌株基因组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(IF04177)培养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(25：24：1v/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倍体积的冰冷乙醇沉淀核酸。在15,000 x 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 μg/ml，将混合物在37℃下保温30分钟。然后加入蛋白酶K至浓度为200 μg/ml，再在37℃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℃备用。

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pSE04。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 μ M, 引物ALAS3d 5'-TTTATGATGGAGGCCCTTCTCCAGCAGTCTC-3' (SEQ ID No:5)和引物ALAS4e 5'-CTATGCATTT AAGCAGCAGCCGCGA CTGG-3' (SEQ ID No:6)各50pmol, Taq DNA聚合酶(Perkin-Elmer Corp., Branchburg, NJ)2单位, 以及1xTaq 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℃1分钟、55℃1分钟以及72℃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, 并按照生产商的说明将其亚克隆至pCRII载体 (Invitrogen, San Diego, CA)中, 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

按照生产商的说明, 将E. coli Y1090ZL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 λ 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 λ ZipLox载体臂进行连接, 利用可商购的包装提取物(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转移复制至圆形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:1xDIG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1xTaq 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 $^{\circ}$ C 1分钟、55 $^{\circ}$ C 1分钟以及72 $^{\circ}$ 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 $^{\circ}$ C下,在5xSSC、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。在5xSSC-0.1%SDS中将膜洗涤两次,随后在2xSSC-0.1%SDS中洗涤两次,每次洗涤均是在室温下进行15分钟。按照生产商说明,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,再按照生产商说明(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* DH5 α pSE11, pSE13, pSE15, pSE17和pSE20。已发现这些克隆有所重叠,覆盖一个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- γ -酮戊酸合成酶(hemA)探针对米曲霉A1560菌株基因组DNA进行Southern杂交

将按实施例1所述方法制备的米曲霉A1560菌株基因组DNA(10 μ g)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。按照生产商说明,利用TurboBlot装置(Schleicher & Schuell, Keene, NH),在0.4N NaOH中将DNA转移到NytranPlus膜上。在

Hybaid烘箱 (Labnet, Woodbridge, NJ) 内, 将膜在 5xSSC、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。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, 不同之处在于此处使用 hemA 克隆 pSE17 作为模板, 并使用引物 hemA5' 5'-TCATTTAAATGATGGAGTCTCTTCTCC-3' (SEQ ID No:7) 和引物 hemA3' 5'-TCTTAATTAATCAGCTCACATGCGGG-3' (SEQ ID No: 8)。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 5xSSC-0.1%SDS 中于室温洗涤 2 次, 每次 15 分钟, 再在 2xSSC-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, 随后按照生产商说明, 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- γ -酮戊酸合成酶 (hemA) 基因的特征

按下述方法, 对实施例 3 中所述的 E. coli DH5 α pSE17 进行 DNA 测序。利用 Applied Biosystems 的 373A 型 DNA 自动测序仪 (Applied Biosystems, Inc., Foster City, CA), 使用染料-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 DNA 测序, 测序中使用 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; 与之相比, 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)。

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,这一点与发挥线粒体前导序列功能相一致。

实施例6: 通过PCR制备基因组胆色素原合成酶基因(hemB)探针

依据米曲霉126bp hemB片段的侧翼氨基酸序列(Jesper Vind, 1994, 博士论文, 哥本哈根大学, 哥本哈根, 丹麦)和酵母及人hemB克隆的同源区(Myers等, 1987, 生物化学杂志, 262: 16822-16829; Wetmur等, 1986, 美国国家科学院院报, 83: 7703-7707), 设计简并性PCR引物。利用Applied Biosystems 394型DNA/RNA合成仪合成寡核苷酸引物。利用有义引物5'-GT(AGCT)GC(AGCT)CC(AGCT)(AT)(CG)(AGCT)GA(CT)ATGATGGA-3'(SEQ ID No:9)和反义引物5'-GC(AG)TC(AGCT)CG/T(AG)AA(AGCT)CC(AG)TA-3'(SEQ ID No:10), 将pJVi60(Vind, 1994, 出处同上)用作模板, PCR扩增hemB片段。

PCR反应体系(50 μ l)含有10mM Tris-HCl pH8.3, 50mM KCl, 1.5mM MgCl₂, 0.01% w/v 明胶, 各200 μ M的dATP、dCTP、dGTP和dTTP, 500ng pJVi60, 以及上述PCR引物每种各50pmol。将反应物在95℃保温3分钟, 再冷至80℃, 然后加入5单位的Taq聚合酶。将反应物置于Perkin-Elmer 9600热循环仪中, 反应程序设置为35个循环, 每个循环是95℃ 30秒、45℃ 1分钟和72℃ 1分钟。最后一个循环后, 将反应物在72℃保温5分钟。将预期的126bp hemB PCR产物克隆于pCRII载体中, 得到质粒pAJ005-1(图4)。

实施例7: 米曲霉A1560菌株DNA文库及胆色素原合成酶基因(hemB)克隆的鉴定

按照实施例3所述构建米曲霉A1560菌株基因组DNA文库。

将大约 8×10^4 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³²P标记的PCR产物

进行探测, 该PCR产物是按照Mertz和Rashtchian(1994, 分析生物化学, 221: 160-165)所述方法扩增pAJ005-1(参见实施例6)的hemB片段而得到的。扩增反应物(50 μ l)含有10mM Tris-HCl pH8.3, 50mM KCl, 1.5mM MgCl₂, 0.01%(w/v) 明胶, 各0.04mM的dATP、dCTP、dGTP和dTTP, 5 μ l ³²P-dCTP(3000Ci/mmol, 3.3 μ M; Amersham, Arlington Heights, IL), 以及各50pmole的有义引物5'-GTGGCTCCGAGTGATAT-3'(SEQ ID No:11)和反义引物5'-GCATCGCGAAAAGGACCG-3'(SEQ ID No:12)。反应物在95℃加热3分钟, 随后添加5单位Taq聚合酶。然后将反应物置于Perkin-Elmer热循环仪中, 反应程序设置为30个循环, 每个循环是95℃1分钟、55℃1分钟和72℃1分钟。将反应溶液过Sephadex G50柱(Pharmacia, Alameda, CA)以除去未掺入的核苷酸, 然后变性该反应溶液并加入到杂交缓冲液中。将变性探针(10⁶cpm/ml)加入杂交缓冲液中, 并与已经预杂交的膜一起保温过夜。预杂交和杂交是在42℃下, 于5xSSC, 50mM 磷酸钠 pH7, 5xDenhardt's溶液, 0.1%(w/v)SDS, 5mM EDTA pH8, 10 μ g/ml变性鲑精DNA, 以及50%甲酰胺中进行的。将膜在42℃于0.1xSSC、0.1%SDS中洗涤4次, 每次15分钟。按照生产商说明, 对产生阳性信号的初始噬斑再次筛选并纯化。按照生产商说明(Bethesda Research Laboratories, Inc., Bethesda, MD), 将产生阳性信号的10个基因组克隆作为pZL衍生物从 λ ZipLox载体中切出, 并按照Hattori和Sakaki(1986, 分析生物化学, 512:232-237)的方法进行测序。这些pZL衍生物被命名为pAJ007-1至pAJ007-10。根据限制性作图分析, 克隆E. coli DH52 pAJ007-6含有一个3.7kb 基因组片段, 对该克隆进一步分析。

实施例8: 胆色素原合成酶(hemB)基因的表征

按照实施例7所述的方法, 对实施例7所述的E. coli DH5 α pAJ007-6进行DNA测序。

如图5(SEQ ID No:3)所示, 克隆的米曲霉A1560 hemB基因的核苷酸序列表明其中含有一个具1308个核苷酸的开放阅读框, 该开放阅

读框编码如图5(SEQ ID No:4)所示、推断分子量为40kDa的含374个氨基酸的一个多肽。该核苷酸序列含有一个48bp的推测内含子,该内含子侧翼是剪接位点共有序列,并含有由(Unkles, 1992, 丝状真菌应用分子遗传学,第2章, J.R.Kinghorn和G.Turner编, Blackie Academic and Professional Publications)所推测的内部共有序列。3'剪接位点(TAG)位于Met下游254bp处,5'剪接位点(GTCCGC)位于3'剪接位点上游46bp处,内部共有序列(TCTAAC)位于5'剪接位点下游30bp处。5'非翻译区在-377和-233位含有两个CAAT基元,可能在转录调控中具有重要作用(Gurr等, 1987, 于Kinghorn J.R. 编的《真核微生物的基因结构》一书中, 93-139页, IRL出版社, 牛津)。另外,在3'非翻译区还发现了数个推测的TATA样盒(-117, -208, -650)。正如所预期的那样,由于hemB在除植物外的其它生物内是细胞质性的(Bottemley和Muller-Eberhard, 1988, 血液学讨论(Seminars in Hematology), 25: 282-302),因此它似乎在N-末端不含有前导序列。

实施例9: 含有hemAΔ::pyrG等位基因的pSE52的构建

图6示含hemAΔ::pyrG等位基因的质粒pSE52的构建。具体地说,采用Expand PCR试剂盒(Boehringer Mannheim, Indianapolis, IN),根据生产商的说明,并利用以下所示的hemAdel.A和hemAdel.B引物,从pJaL394上用PCR扩增而得到含米曲霉pyrG基因和5'及3'侧翼DNA的一个4.1kb片段,包括设计为便于重组和除去pyrG标记的直接重复。

hemAdel.A: 5'-ATCCATTGAAGCATCCAGGGTTATTGTCTC-3' (SEQ ID NO:13)

hemBdel.A: 5'-GGATTGACGAAGCAGAGGATGACGATGAGC-3' (SEQ ID NO:14)

在Perkin Elmer GeneAmp PCR系统9600中进行Expand PCR扩增反应,程序为94℃ 2分钟、62℃ 30秒、68℃ 3分钟进行1个循环; 94℃ 15秒、62℃ 30秒、68℃ 3分钟(在每个循环后增加30秒)进行10个循环; 94℃ 15秒、62℃ 30秒、72℃ 5分钟进行15个循环; 于4℃进行

一个保温循环。

根据生产商的说明，将产物克隆至pCRII (Invitrogen, San Diego, CA)中，获得pSE40，在两个独立的克隆中确证开放框(ORF)的核苷酸序列。将pSE40含有pyrG基因的EcoRI片段连接到pSE17的EcoRI位点内，这样将pyrG基因置于hemA基因的3'侧，得到pSE41。用平端切割酶MluNI和Eco47III切开pSE41(去掉hemA ORF的一个445bp片段)，进行凝胶纯化，并且重新连接得到质粒pSE48b。对含有该缺失的hemA基因区进行测序，以确证为hemAΔ::pyrG等位基因。示于图7的测序结果表明，该hemA缺失去除了编码序列，并破坏了基因的阅读框(SEQ ID NO: 15和16)。用Expand PCR试剂盒，根据生产商的说明，并利用以下所示的SE48up和SE48dwn两个引物，用与上述相同的条件，从pSE48b上PCR扩增出这个等位基因。

SE48up: 5'-AATGGTCAAACTGGCTCCTAC-3' (SEQ ID NO:17)

SE48dwn: 5'-TGTACCTGTTCTTGGGCTGTC-3' (SEQ ID NO:18)

然后将PCR扩增片段亚克隆至pCR2.1 (Invitrogen, San Diego, CA)中，产生pSE52，以便于用限制性内切酶SacI和NotI分离出含有hemAΔ::pyrG等位基因的一个6.3kb片段。这个6.3kb SacI-NotI片段含有接近700bp的5'和3'侧翼DNA，可用来促进与hemA基因的基因组拷贝发生同源重组。

实施例10: 用hemAΔ::pyrG等位基因片段转化米曲霉

为了用hemAΔ::pyrG缺失等位基因替代野生型的hemA基因，用大约20μg的SacI-NotI 6.3kb hemAΔ::pyrG片段转化米曲霉How B425。用原生质体进行转化，其中原生质体的浓度是 2×10^7 个/ml。将100μl原生质体与10μg DNA和200μl 60% PEG4000-10mM HEPES-10mM CaCl₂溶液在34℃温育30分钟。加入3ml的SPTC(40% PEG4000, 0.8M山梨糖醇, 0.05M Tris pH8.0, 0.05M CaCl₂)，并且将原生质体直接铺在补充有2.5mM或5mM的5-氨基-γ-酮戊酸的基本培养基上。通过评估原初转化体在34℃于缺少5-氨基-γ-酮戊酸的基

本培养基上的生长状况确定5-氨基- γ -酮戊酸营养缺陷。将在这些条件下的生长被判断为边缘状态的菌株划线分离，以进行单克隆分离。然后，对已纯化的单克隆分离物，利用上面就原始筛选所述的相同条件，进行5-氨基- γ -酮戊酸营养缺陷型二次筛选。

用缺失等位基因转化米曲霉HowB425，得到了265个集落。在缺少5-氨基- γ -酮戊酸的基本培养基上对240个一级转化体进行筛选，得到生长比较模糊的15株。其中11株在二次筛选中在基本培养基上的生长很差，选出其中6株：SE29-70、29-86、29-87、29-180、29-192和29-197进行进一步单克隆纯化和Southern分析。

按照实施例1所述分别制备六个转化体的基因组DNA（10 μ g），并用BamHI限制性消化。通过1%琼脂糖-TBE凝胶电泳分离片段。根据生产商的说明，用TurboBlot装置（Schleicher & Schuell, Keene, NH）在0.4N NaOH中将DNA转移至HybondN（Amersham, Arlington Heights, IL）尼龙膜上。将该膜在42℃下，于5 $\times$ SSC、0.1% Sarkosyl、0.02% SDS、1% Genius封闭剂和50% 甲酰胺中，在Hybaid温箱（Labnet, Woodbridge, NJ）中预杂交2小时。探针通过用下示的hemA delup1和hemAdeldwn1引物PCR扩增pSE17而制得。

hemAdelup1: 5'-AGGCCTCTTGGGTTATGAATG-3' (SEQ ID NO:19)

hemAdeldwn1: 5'-TGACCTGGAGATTAGACATAG-3' (SEQ ID NO:20)

扩增反应（100 μ l）中包括以下组分：1 μ g的pSE17、50pmol的hemAdelup1引物、50pmol的hemAdeldwn1引物，分别为200 μ M的dATP、dCTP、dGTP和dTTP，1 $\times$ Taq聚合酶缓冲液（Perkin-Elmer Corp., Branchburg, NJ），和5个单位的Taq聚合酶（Perkin-Elmer Corp., Branchburg, NJ）。按以下程序在PerkinElmer 480型热循环仪中进行反应：第1个循环-95℃ 5分钟，55℃ 2分钟，72℃ 5分钟；第2-30个循环-95℃ 1分钟，55℃ 1分钟，72℃ 1分钟；在4℃进行一个保温循环。

用凝胶分离PCR产物DNA，用QiaPure柱（Qiagen, Chatsworth,

CA) 纯化, 乙醇沉淀并重溶于25 μ l的TE中。根据生产商的说明, 用Prime-It II试剂盒(Invitrogen, San Diego, CA)和2.5 μ l纯化DNA, 通过随机引发法制备 α -³²P-dCTP标记的探针。该标记探针用G50 midi柱(Five Prime Three Prime, Boulder,) 纯化。纯化的探针(2 μ l, $\sim 1 \times 10^6$ cpm)先在0.5N NaOH中于37℃下2分钟变性, 然后加入到10ml上述新鲜杂交液中。在42℃下过夜杂交, 室温下用2 $\times$ SSC、0.1% SDS洗膜10分钟, 接着, 在室温下用0.2 $\times$ SSC、0.1% SDS洗10分钟, 再在68℃下于0.1 $\times$ SSC、0.1% SDS中洗15分钟两次。洗过的膜浸于2 $\times$ SSC中, 再对Kodak Xomat AR胶片曝光。

Southern分析这些菌株显示, 5个仅含hemA Δ ::pyrG等位基因, 1个含野生型和hemA Δ ::pyrG 等位基因。对SE29-70、29-87和29-197菌株的孢子纯化两轮, 得到不能在基本培养基上生长、从而证明为5-氨基- γ -酮戊酸营养缺陷型的分离物。

实施例11: 通过补充5-氨基- γ -酮戊酸或氯高铁血红素拯救致死的hemA缺失表型

5-氨基- γ -酮戊酸或氯高铁血红素拯救hemA缺失表型的能力由补充生长培养基来确定。

制备于水中的250mM 5-氨基- γ -酮戊酸(Porphyrin Products, Logan UT) 贮液, 用0.22微米的滤膜过滤除菌, 贮存在暗处-20℃, 并在用前或倒板前立即加入制好的培养基中。以0.005mg/ml-0.3mg/ml的浓度范围将5-氨基- γ -酮戊酸加至基本培养基琼脂或YEG液体培养基(每升含5g含酵母提取物和20g葡萄糖)中。

在50mM NaOH中新鲜制备10mg/ml的氯高铁血红素贮液, 用0.22微米滤膜过滤除菌, 存在冰中备用。以0.05mg/ml-0.2mg/ml的浓度范围将氯高铁血红素加至基本培养基琼脂中。

结果显示, 加至基本培养基琼脂上时, 浓度低至30 μ M (5.0 μ g/ml) 的5-氨基- γ -酮戊酸即足以拯救此营养缺陷表型。在液体YEG培养基中生长需浓度稍高一些(高达1mM)。

结果进一步显示, 氯高铁血红素加至基本培养基琼脂中时不能

支持所检测的任何缺失菌株的生长。氯高铁血红素的这些浓度并不抑制含有野生型hemA基因的菌株的生长。替换有氯高铁血红素的生长培养基和条件有可能拯救hemA表型。

实施例12: 用野生型hemA进行转化以拯救hemA缺失表型

对从米曲霉菌株SE29-70的第一次单克隆分离物制备的原生质体进行拯救转化。

用10 μ g的pSE17(其含有4.2kb的基因组hemA区)或pSE31(其含有与NA2-tpi启动子融合的hemA ORF),根据实施例8所述程序转化米曲霉菌株SE29-70的原生质体,铺在缺少5-氨基- γ -酮戊酸的基本培养基上,在34 $^{\circ}$ C下培养。培养两天后,转化克隆已比较明显,但也有很多更小的“流产”克隆,真的转化体可通过它们较大的集落大小和培养一天后能很好地产生孢子的能力而得以区分。用pSE17得到200个集落,用pSE31得到299个集落,而在“没有DNA”的对照平板上只出现28个集落。转化效率为20-30个转化体/ μ g DNA。

发现选出作进一步分析的所有集落均为5-氨基- γ -酮戊酸原养型。从8个pSE17转化体(命名为SE40-1-7, 9和10)分离出DNA,根据实施例10所述程序进行Southern分析,结果表明它们除含有该hemA缺失等位基因之外,还含有至少一个拷贝的完整hemA基因。这些结果证实了hemA缺失表型能被野生型hemA基因的转化所拯救。

微生物保藏

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
E. coli DH5 α (pAJ007-6)	NRRL B-21564	1996年4月22日

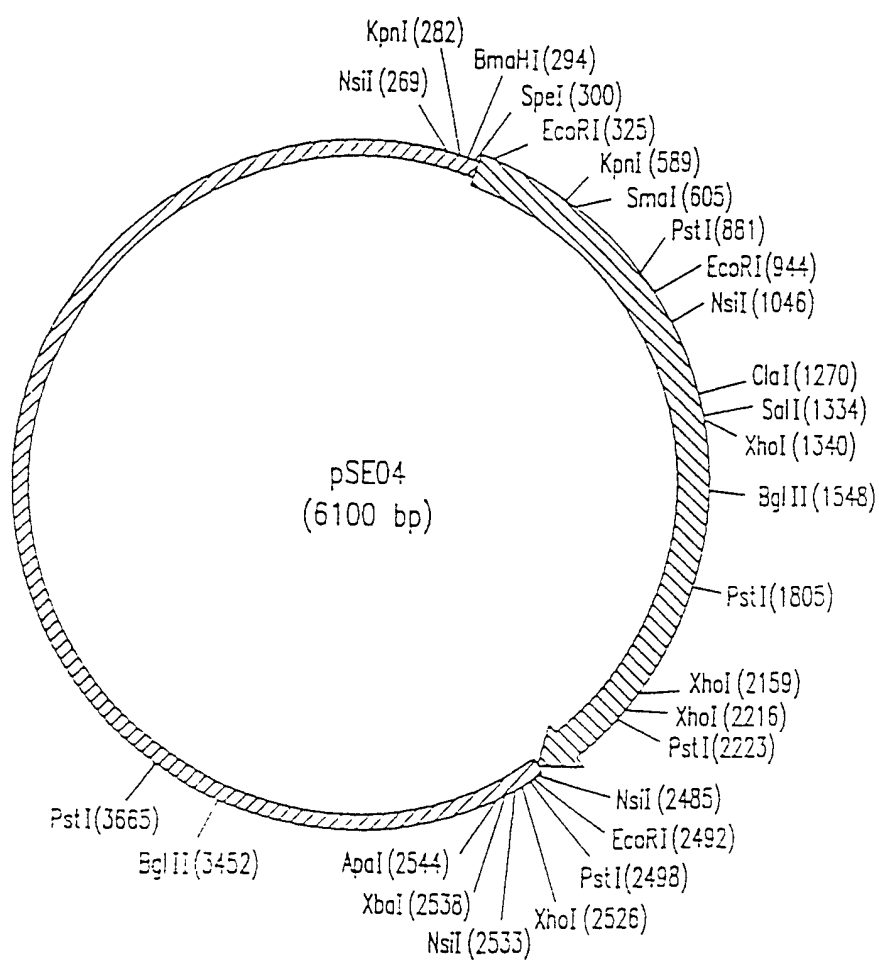
两菌株已在一定条件下被保藏,能确保在此专利申请未决期间由专利和商标官员决定根据37C.F.R. § 1.14和35U.S.C. § 122授予可获得所述菌株之权利的人可得到所述菌株。该保藏物代表每种保藏

菌株基本上纯的培养物。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并不意味着在由政府行为授予的专利权部分废除时是对实施本发明的许可。

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的实施方案的范围，因为这些实施方案仅仅旨在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。

图 1



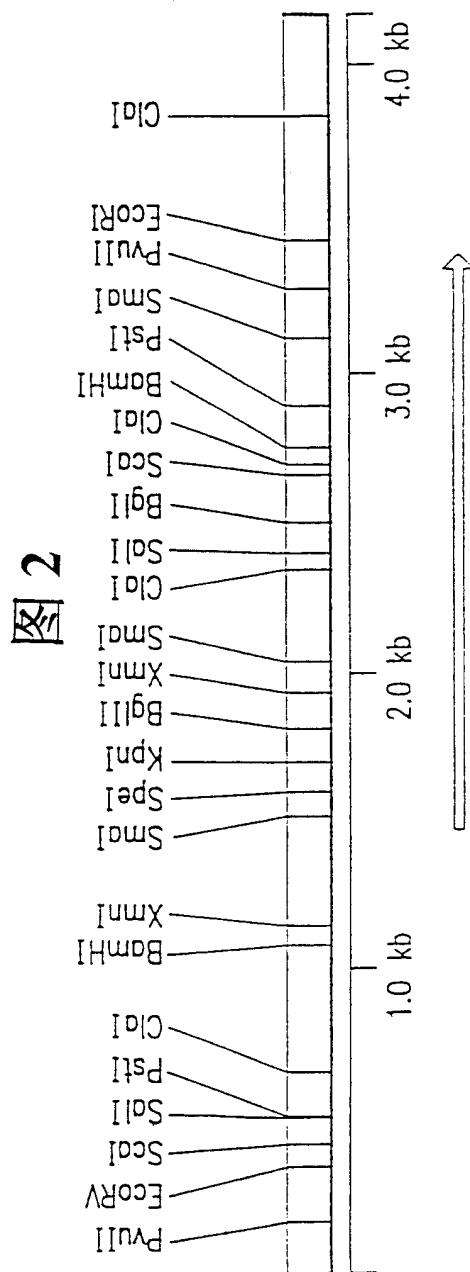


图 4

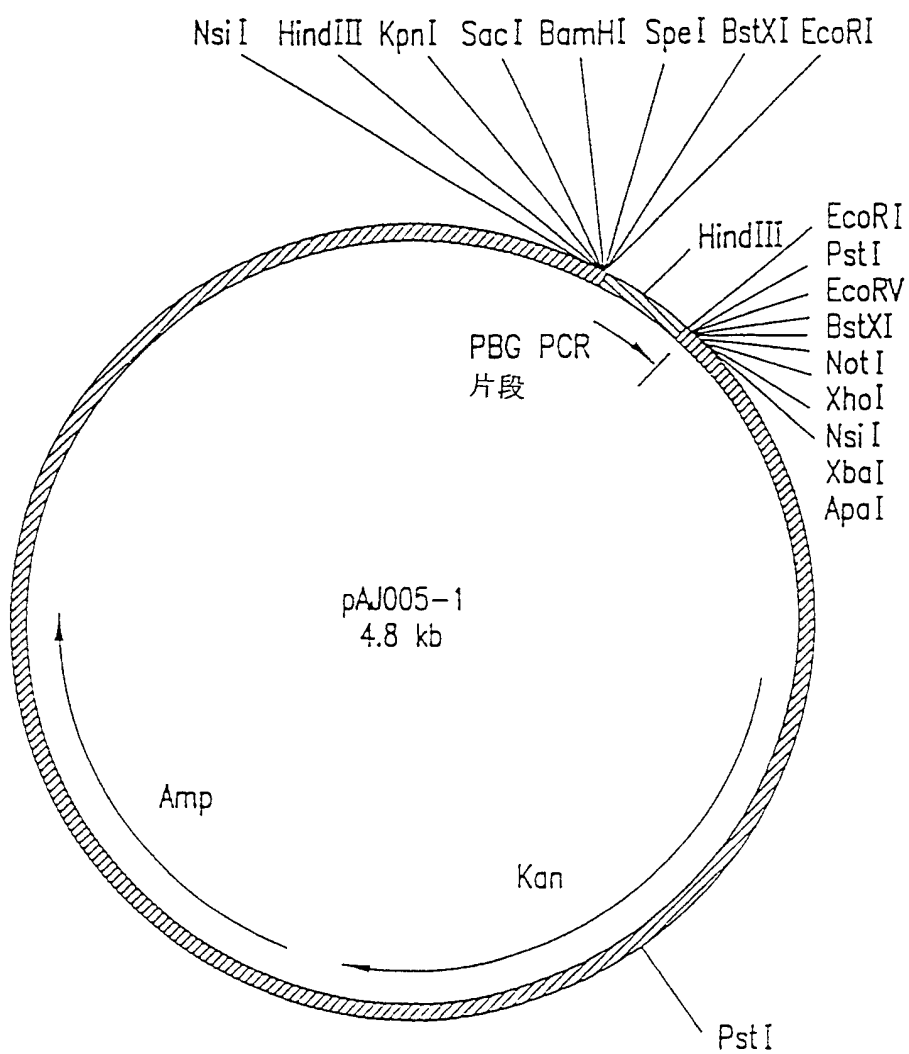


图 5

CTGGACCAATGGTAACCCCTCCGTAATTGCCTTACAGATTTAGCCCAGGGGGTTATGGTATCCTTGGGTA 70
 TTGAGGCGCTGGAAATTTTTTAGCCACCAGTTTACAGCCAGTTTCCGTTTGAAATATTTACATCCCCC 140
 GACCCGTGCCAATAACAATAATTTTTCGCTATATATAACGCCCTAGCGTTGTTTTATGATCCTTAAAT 210
 CCTTACTTGTACCTGAAAATTGCAACAAATGACTGACCTGGATCGTGGCCATTATATCATTGCCCTG 280
 CGAAGTCGTATTGTGCCAGTGGCACAGGCCGCTATTCTCTTTCTCCCTCCACCGGTTTCTATCTTCCA 350
 TAGCACCCCACTTGCTTGCCGCTCCTGTCATTATGCTCTTTCTAATCTCGTCTTGACCTCGCCTTCAG 420
 M S F S N L V S D L A F R
 AGATTCTCATGATGACCGAAGTTCTCAGATATCTCAGGTACAATCGCAAGCCACTGCACGATCGTATACA 490
 D S H D D R S S Q I S Q V Q S Q A T A R S Y T
 AGCAGAGCTGCCACAAGCGTCAGCATATCTGGGACATCTCAAGCCAGCTTCATTCCGTTACAGCCATC 560
 S T A A T S V S I S G D I S S Q L H S G Y S H
 CACTGAGCCGATCATGGCAGGCTGAAAGACAGTTGACTAAAGTCCGCAATTTCTTTTGTATTACTGAGC 630
 P L S R S W Q A E R Q L T K
 TGCTCTAACCCGAGATAGCAAATGCTTATTTATCCTCTCTTCATCACCGATAATCCCGATGAGGAGACT 700
 E M L I Y P L F I T D N P Q E E T
 CCTATCCCGTCTCTCCCTGGACAGTATCGTCGAGGATTAACCGTCTAGTTCTTTTCATCAAACTCTTG 770
 P I P S L P G Q Y R R G L N R L V P F I K P L
 CCCACAAGGGGCTACGCTCAGTCATCCTGTTTGGCGTCCCACTACACCCCTCTGCGAAGGATGCACTAGG 840
 A H K G L R S V I L F G V P L H P S A K D A L G
 TACCGCTGCAGACGATCCATCTGGACCGGTAATCAAGCTATTGCGTTGCTTAGGTGCGGTTTCTCAA 910
 T A A D D P S G P V I Q A I R L L R S R F P Q
 CTTTATATCGTGACAGATGTGTGCCTTTGCGAGTATACTTCGCATGGCCACTGTGGGATACTGCGAGAAG 980
 L Y I V T D V C L C E Y T S H G H C G I L R E
 ATGGGACTCTTGATAATACACAGTCTGTGGATCGGATTTCCGATGTTGCTCTGGCTTATGCTGCCGCCGG 1050
 D G T L D N T Q S V D R I S D V A L A Y A A A G
 AGCCCATGTGTGCTCGCTCGTCTGATATGAATGATGGCGAGTGGTGCTATAAACTGAAGCTTATTGAA 1120
 A H C V A P S D M N D G R V R A I K L K L I E
 GCGGGATGGCCACCGTGTCTACTGATGTCTACAGCGCCAAATTTAGCGGTTGTTGTACGGCCCTT 1190
 A G M A H R V L L M S Y S A K F S G C L Y G P
 TCCGTGATGCAGCGGGTCTGCCCATCATTCGGGGATCGCAGATGCTACCAGTTACCACCGGAGGCCG 1260
 F R D A A G S C P S F G D R R C Y Q L P P G G R
 TGGACTTGCTCGCGCGCTATACAGAGAGATATAGCGAAGGGGAGACATCATAATGGTAAAGCCGGCG 1330
 G L A R R A I Q R D I G E G A D I I M V [K] P A
 AGCAGCTACCTGGACATTATCAGAGACGCAAAAGAAATTGCCAAGACATTCCCATTTGCTGCTTACCAGG 1400
 S S Y L D I I R D A K E I A K O I P I A A Y Q
 TCAGCGTGAGTATGCTATGATACATGCTGGTGCCAAGGGCGGCTATTTGACTTGAAATCCATGGCCTT 1470
 V S G E Y A M I H A G A K A G V F D L K S M A F
 TGAAAGTACTGAAGGGATTATAAGGGCTGGTGGGATTATAGTAAGCTATTTGCTGCTGATTTTCTA 1540
 E S T E G I I R A G A G I I V S Y F V P D F L
 GATTGCGTTTGGAAATGATTTAGCTAGATGGAGCGTGATGAAAGCATCCACCAGATAAATAGCAGTGACC 1610
 D W L S K
 ATCGCGTTTGAAATCATACCTATTGGAGTAGAAGTCTCGGTATCTCGTTGGGGATTCTCTAGGTTGCTTAT 1680
 TTAACGTAATGCCACGCCATGTGTTATATATTCCTAAATACTTTATAAAAGATACCAAGCTGATGG 1750
 TGCCAAGTGACCACTTCTAATAAATAACAATTATACCAATTCCTCCGAAATATGCCGG 1807

图 6

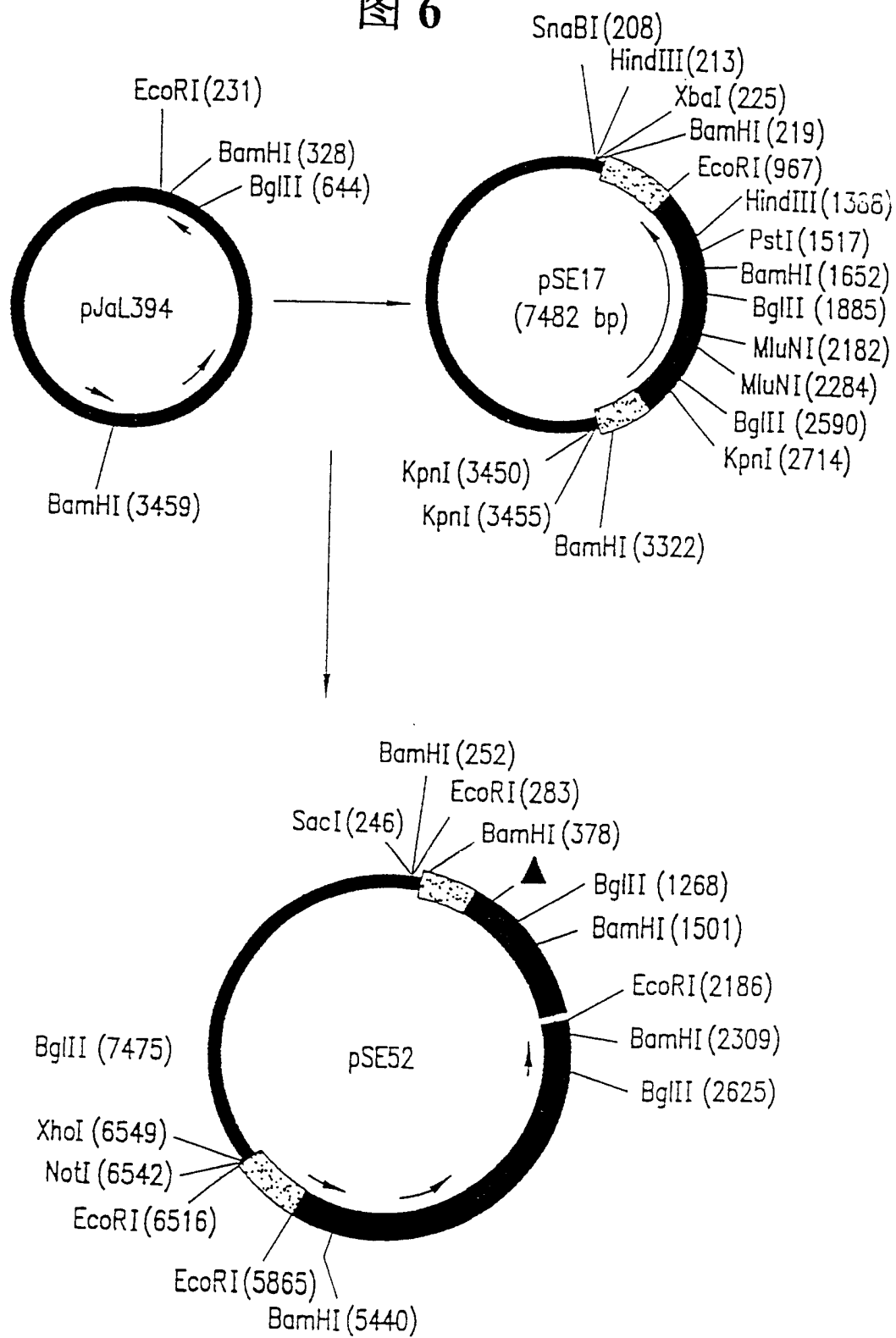


图 7

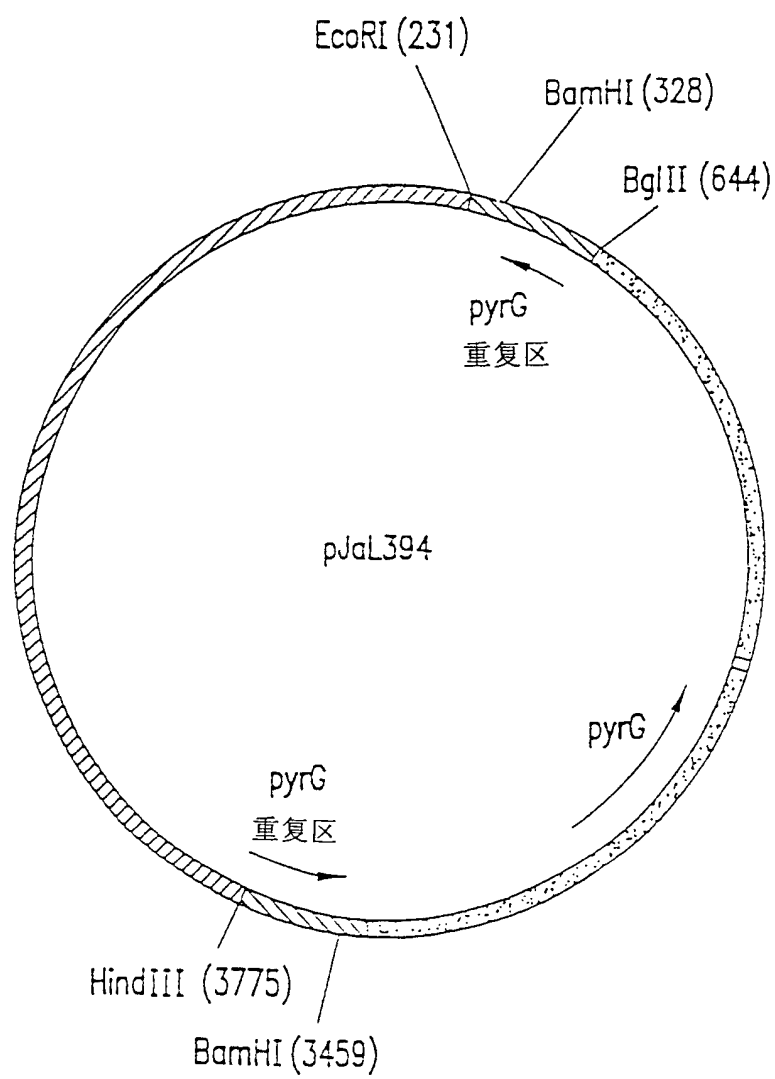


图 8A

AAATTGATGGCAAGATAGACATTGTATCCIGTACCTGTTCTTGGGCTGTGACGGGGGGTGAAATTGACGGTCATGACCCGGCTATTAT 90
 TACTATTGTTGTACTGTACATCCGGATCCGTGGTCTGTATCTAGTTAGGCAATATTCCCGTGCCAGGCCCTCTGGGTTATGAATG 180
 ATTTCATAGGTGAAGTTTCGTATCCGTACGACCCGAGAGATTCTTAGTATTACTTGATGAATATGACATTGCGGAGTTAAGTCCG 270
 CCGGCAATCACGGCGGAGGATATGGTAGCCGAAGTCTGCCGAAGTCCCGACTTACTTCTAGGAGTGGCTTAGTGCCTCAG 360
 CGCCCTCGCCCTCAGTCCATCAGCCAGATTGACICTTATTCTCTCTCTTTCGGCGGGTGACATATCCCTCTCTCTCCCTCT 450
 CCCTCTTGACAACATTCACTCTCGCTTCCCTTTTGATATAGTCAGTTTCGCTATCCATTGAAGCATCACTCATGGAGTCCTCTCTCCA 540
 GCAGTCCCGGGGATGIGCCCGTTCCTTAAGCGCACATCTCCATCTCTCTGCGTAGCTGGCAACCGCGACTCGACCTAGCACCTAGTTC 630
 Q S R A M C P F L K R T S P S S L R T L A T A T R P S T S S 720
 CGGTGGAGGCACTATGCTAATCTCCAGGTCAATGCCGTGGCTGCTGATGAGCAAGGCTCTGGCCGTGCGAGGCCCAAAATTGCAC 810
 G G T M S N L Q V I A R R C P V H S K A L A V Q S A P N C T 900
 GGCAAGGAGGCGGATTAGCTTCAGCTCATGCTTCGTGGCTAAGCATGCGACCCCTCGCAACCCCTGGGTAGCAAGTTGCCCGACTGTGT 1080
 A R R R H S S A H A S W L T M P P S Q P W V A S C P T V L 1170
 ATTCTGCCGATAGCCTGAATCATGCAATGATTCAGGGTATTCGCCATTAGGCGCCMAGAAATGGTTTTCAAGCATATGATCTG 900
 F C P I A I M H R F R V F A I Q A P R K W F S S I M I N 990
 GTGGACCTTGAGGCCAAGTTGGCAGCTCTACCTCTTCATGTCCCCAAGATTATTGCAATCAGTTTATAGCATGTGCGGATCTATT 1080
 S T L R P S W Q L Y L F M S P R L L H S N Q F I A C A D L L 1170
 GCCCCAATTGAGAAATCTGTGATCTTGCAGACAAGTACGGTGGCAATTACTTCTCTGGATGAAGTCCACGCTGTGGGAATGACGACCT 1260
 P Q L R R S V I L Q T S T V P L L S W M K S T L W E C T D L 1350
 CACGGAGCAGGTGTGGCAGAGCACCTTGACTATGACATCTAIGCTTCCCAAGATACGGTCMCCCGGCCAGTACTAAGGGAAACCGTGAIG 1440
 T E Q V W Q S T L T M T S M L P K I R S T R A V L R E P W 1530
 GACCGAATCGATATATCACCGGTACTCTGGGC/MGGGCTACGGATGTGTGGGGGCTACATTTGCTGGATCCGCTGCGATGGTTGACACC 1620
 T E S I L S P V L W A R P T D V S G A T L L D P L R W L T P 1350
 ATCCGCTCCCTCGCCCTTGGCTTTCATCTTACCACGTCTTGGCCGCCCGCCACCATGGCTGGTGGCAGACACTGCTATCCAGTACCGGCT 1440
 S A P S P L A S S S P R P C R P P W L V Q T L L S S T R L 1530
 CGTACACCGGCGGAGCGGCTGTGACAGTTCACACCCCGGGTCAAGCAGCTTTCMAGGAGTTGGATATTCCTGTATTTCCCAAC 1620
 V T R A T A S C S S C T P A R S K Q L S R S W I F L F P T 1350
 CCCTCCCATATCATTTCCGCTCCTTGGGTGGGATGCCGAGGTTCGTAAAGAGGCTTCGGACAMGCTTCGGAGGAGCATGGAAATTTATGTA 1440
 P P I S F R S W L G M P R L L R R P R T S F W R S M E F M Y 1530
 CAAGCCATCAACTACCC/MCCGTGCTCGGGTGAAGAGCGGCTTCGTATCACGCCACCCCGGAGACATATCAAGGAGCACCCCGGAGCAC 1620
 K P S T T Q P C L G V K S G F V S R P P R D I S R S T A T T

8B

CTGGTGCAGGCCGTCACACAGTCTGGAACGAACCTGGGCATCAACGCACACAGCGATTGGGAAGCGCAAGCGCGCTTCGTGCGCGTGGGT 1710
W C K P S K Q S G T N W A S N A P A I G K R K A A S A W V
GTGATGGCGCGAGGCTGAGAACACAGCCGATTGGAAATGATGTGCAGCTGGGCTGAGGAAACGAAGCCATTGAGGCTGCTGTGGAA 1800
S M A P R L R T S R F G M M C S W G . R K T K P L R L L W N
CGGAGTTTGGCAGAGGCCCCCATCGGAGACCGCACCGCTGCTGCCGCGGCTGCTGCTTGTGAATCCCGGTGGGTGGCTGGCTGCAAGT 1890
A S L P R P P C G P P V L P R L L L R Q S R W V W L P
GGCTGCCCCGCATGTGAGCTGAATCGACCGTGAATT 1926