

[19] 中华人民共和国国家知识产权局

[51] Int. Cl<sup>7</sup>

C12N 15/57

C12N 1/15

C07K 14/00 C12N 1/14

C12N 9/48

# [12] 发明专利说明书

[21] ZL 专利号 95191058.2

[45] 授权公告日 2002 年 2 月 27 日

[11] 授权公告号 CN 1079831C

[22] 申请日 1995.9.19 [24] 颁证日 2002.2.27

[21] 申请号 95191058.2

[30] 优先权

[32] 1994.9.20 [33] US [31] 08/309,341

[86] 国际申请 PCT/US95/11945 1995.9.19

[87] 国际公布 WO96/09397 英 1996.3.28

[85] 进入国家阶段日期 1996.6.18

[73] 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

[72] 发明人 D·S·亚夫尔 S·A·索姆普森

[56] 参考文献

APPL. ENNIRON. MICROBIOL. VOL. 58. NO. 7 1992. 1. 1 Dal

Degan F. et al " purification and characterization of Two serine carboxy

CELL VOL. 48 1987. 3. 13 L Vaus et al " Protein sorting in yeact; Localization determinant of yeact va

J. OF CHROMATOGRAPHY, VOL. 370 1986. 7. 1 S. Krihnan and M. VIJayalakshmi " Purification and some properties of three s

审查员 曾繁辉

[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 2 页 说明书 28 页 附图页数 10 页

[54] 发明名称 编码黑曲霉羧肽酶的基因

[57] 摘要

本发明涉及编码子囊菌或半知菌羧肽酶 Y 基因的基因, 以及经过修饰而产生减少量的羧肽酶的宿主细胞。

ISSN 1008-4274

# 权 利 要 求 书

---

1.一种核酸构建物，它含有

(a)编码黑曲霉羧肽酶Y的核酸序列；

(b)编码具有氨基酸序列 SEQ ID No:2 的羧肽酶的核酸序列，或在非常严格的条件下与 SEQ ID No:1 杂交的核酸序列；或

(c)编码具有氨基酸序列 SEQ ID No:4 的羧肽酶的核酸序列，或在非常严格的条件下与 SEQ ID No:3 杂交的核酸序列。

2.根据权利要求 1 的核酸构建物，它含有编码黑曲霉羧肽酶 Y 的核酸序列。

3. 根据权利要求 1 的核酸构建物，其中核酸序列编码具有 SEQ ID No:2 所示氨基酸序列的羧肽酶。

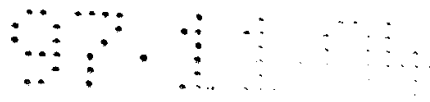
4. 根据权利要求 3 的核酸构建物，其中核酸序列为 SEQ ID No:1。

5. 根据权利要求 1 的核酸构建物，其中核酸序列编码具有 SEQ ID No:4 所示氨基酸序列的羧肽酶。

6.根据权利要求 5 的核酸构建物，其中核酸序列为 SEQ ID No:3。

7.根据权利要求 1 的核酸构建物，它含有在非常严格的条件下与 SEQ ID No:1 杂交的核酸序列。

8.根据权利要求 1 的核酸构建物，其中核酸序列得自曲霉属、镰孢属，青霉菌属，腐质霉属，木霉属，Scytalidium, 毁丝霉属或草根霉属。



9.根据权利要求 1 的核酸构建物,它含有在非常严格的条件下与 SEQ ID No:3 杂交的核酸序列。

10.根据权利要求 9 的核酸构建物,其中核酸序列得自曲霉属、镰孢属,其霉菌属,腐质聚属,木霉属, Scytalidium, 毁丝霉属或草根霉属。

11.根据权利要求 1 的核酸构建物,其中一选择标记被插入到羧肽酶序列中。

12.根据权利要 11 的核酸构建物,其中选择标记为 amds, pyrG, argB, niaD, sc 或 hygB。

13.一种丝状子囊菌或半知菌细胞,它含有权利要求 1 的核酸构建物。

14.根据权利要求 13 的丝状子囊菌或半知菌细胞,它为曲霉属、镰孢属,其霉菌属,腐质聚属,木霉属, Scytalidium, 毁丝霉属或草根霉属细胞。

# 说明书

## 编码黑曲霉羧肽酶的基因

本发明涉及编码真菌液泡蛋白酶的基因。特别是本发明涉及丝状子囊菌类或半知菌类真菌，例如曲霉属的羧肽酶基因。

真菌液泡是含有许多水解酶，包括数种蛋白酶的酸性器官，并且基本上等同于哺乳动物的溶酶体。已经鉴定并表征了特别是酵母的一种或多种真菌的水解酶；这些包括蛋白酶 A (PEPA 或 PrA)，蛋白酶 B (PrB) 或氨肽酶 (APE)，二肽基氨肽酶 B (DPAP B)，羧肽酶 Y (CPY) 和羧肽酶 S (CPS)。大多数液泡水解酶是作为非活性前体而合成的糖蛋白。事实上，除 APE 外，上述所有提到的蛋白酶都有导致过渡通过分泌途径的信号肽。在晚期高尔基体中，液泡蛋白与分泌蛋白分开，然后最终输送到液泡中。除信号肽外，大多数液泡蛋白还含有输送到液泡后被裂解以产生成熟活性酶的原肽。已经证实在原肽中存在 PrA 和 CPY 对于以液泡为靶所需的氨基酸信息 (Johnson et al., Cell 48:875-885, 1987; Rothman et al., PNAS USA 83:3248-3252, 1989; Valls et al., Cell 48:887-897; Valls et al. J. Cell Biol. 111:361-368, 1987)。对于 CPY，已经表明有四个氨基酸残基 (QRPL) 对于定位于液泡是必需的 (Valls et al. J. Cell Biol. 111:361-368, 1990)。一旦输送到液泡后，蛋白酶 A(pep4) 裂解 CPY 和 PrB 的原肽以便激活蛋白酶 (Ammerer et al., Mol. Cell. Biol. 6:2490-2499, 1986; Woolford et al., Mol. Cell. Biol.

6:2500-2510)。

在啤酒酵母中，已鉴定了错定位或错分类液泡蛋白的三类突变体(Bankaitis et al., PNAS USA 83:9075-9079,1986; Robinson et al., Mol.Cell. Biol., 8:4936-4948,1988; Rothman et al., EMBO J.8:2057-2065,1989; Rothman and Steven, Cell 47 :1041-1051,1986)。这些突变体被称为 vps 或液泡蛋白分类突变体。利用一种选择来分离数个这样的突变体，所述选择是基于观察到质粒上的高拷贝液泡蛋白酶的过表达会导致分泌液泡蛋白酶 ( Steven et al., J. Cell Biol.,102:1551-1557,1986;Rothman et al, PNAS USA 83:3248-3241,1986)。这表明在晚期高尔基体中可能有饱和分类机制。

在啤酒酵母中，还表明，缺失 PEP4，CPY 和 PrB 的菌株由于所需产物蛋白水解的减少生产较高水平的异源蛋白质。因此，由于生产所研究的液泡蛋白酶的真菌被用于重组生产异源蛋白，因而从商业角度看，所述液泡蛋白酶是很重要的。在发酵中这些蛋白酶的存在是很令人头痛的，因为它们还可以降解所需的蛋白质，从而明显降低生产率。通过破碎或缺失编码其的基因可以有利于消除任何已有蛋白质的功能；但是，这样做首先要鉴定并分离在所需宿主中的相应的基因。如上所述，从各种酵母菌株中只鉴定了很少的所述基因；然而，在丝状子囊菌或半知菌中编码液泡蛋白酶的基因很少有人知道。例如，已经分离了编码另两种黑曲霉液泡蛋白酶的 PEPC ( Frederich et al.,Gene 125:57-64,1993 )和 PEPE(Jarai et al., Gene· 145:171-178,1994 ) 基因。PEPC 编码蛋白酶 B ( PrB ) 同系物，PEPE 编码蛋白酶 A 的同系物。也已经克隆了编码 PrA 同系物的 *Neutospora Crassa* 的 PEP4 基因 ( Bowman ,17th funga;

Genetics Conference ,1993 )。本文首次描述了来自丝状子囊菌或半知菌的编码液泡 CPY 的基因。

本发明涉及含有编码丝状子囊菌或半知菌羧肽酶 Y 的核酸构建体，以及重组生产由其编码的蛋白质。如本文所用的，“核酸构建体”指任何 cDNA，基因组 DNA，合成的 DNA 或 RNA 源的核酸分子。术语“构建体”指核酸片段，所述核酸片段可以是单链或双链的，可以从天然基因中完全或部分地被分离，或可以被修饰以含有以在自然界不存在的方式组合且并置的 DNA 片段。所述构建体还可以含有其他核酸片段。在优选的实施方案中，所述序列编码曲霉属的羧肽酶。本发明还提供了生产不产生羧肽酶的丝状子囊菌或半知菌细胞的方法，包括破碎或缺失羧肽酶基因以防止功能酶的表达，或通过利用物理或化学处理的经典诱变以产生其生产 CPY 能力降低或没有了的细胞。另外，本发明还包含不能生产功能羧肽酶或以相对于野生型菌株生产的量而言，生产减少量的羧肽酶的丝状子囊菌或半知菌。所述生物提供了改进重组蛋白质生产方法的基础，其中用编码所述蛋白质的核酸构建体转化羧肽酶缺陷型微生物，然后在表达所述蛋白质的条件下培养。

#### 附图描述

- 图 1 表示黑曲霉 Bo-1 基因组 CPY 克隆的 DNA 序列及其翻译。
- 图 2 表示黑曲霉 SFAG2 CPYcDNA 的 DNA 序列及其翻译。标明了信号肽酶裂解的预期的位点以及成熟 CPY 的 N - 末端。
- 图 3 说明了在破坏 CPY 中所用的构建体。

利用啤酒酵母 CPY，微紫青霉羧肽酶 S1(Svedsen et al., FEBS

333:39-43,1993,和麦芽羧肽酶 - MIII(Sorensen et al., Carlsberg res. Commun .54:193-202,1993), 通过设计一系列的简并寡核苷酸来开始尝试分离曲霉属羧肽酶 Y。下文实施例中提供了所述寡核苷酸序列。在利用黑曲霉菌株 Bo - 1 基因组 DNA 为模板的 PCR 反应中, 用这些序列以各种组合为引物。其中的两个反应(用引物 1 - 1 和 2 - 1; 以及 1 - 2 和 2 - 2) 产生了一个 1100 bp 的扩增产物, 将其亚克隆并测序, 但是没有亚克隆与鉴定为所需基因的 CPY 明显同源。然后进行了用引物 3 - 1, 3 - 2, 4 - 1, 4 - 2, 2 - 1 和 2 - 2 的其他 PCR 反应。在其中的两个反应中(用引物 4 - 1 和 2 - 1; 以及 4 - 2 和 2 - 1) 得到了一个 600 bp 的扩增产物。将该产物亚克隆, 并将 11 个亚克隆测序; 其中的 9 个亚克隆是一致的, 而且与其他来源的羧肽酶 Y 基因同源。用一个亚克隆中的插入片段探测黑曲霉基因组 DNA; 用 Bamh I, Hind III, 和 SAI 消化产物观察到有一个带的杂交, 表明在黑曲霉中存在单个 CPY 基因。于 65 °C, 在 1.5 × SSPE, 1.0 % SDS, 0.5% 脱脂牛奶和 200 μg/ml 鲑精子 DNA 中进行杂交。

制备在 EMBL4 中的黑曲霉基因组 DNA 库, 然后用 PCRCPY - 衍生的基因片段 ( <sup>32</sup>P - 标记的 ) 检测, 以便分离全长基因。在约 28, 000 个噬菌体中, 检出 11 个阳性的; 将其中的 9 个在纯化后再用探针杂交。大多数这些克隆共有的 5.5 HindIII 片段被鉴定为黑曲霉 CPY 基因。将该片段亚克隆并测序; 在图 1 中提供了含有 CPY 编码区和预期的氨基酸序列的片段的序列。

随后, 再筛选来自不同黑曲霉菌株的 cDNA 库。从该库中还鉴定至少一个全长克隆。将该克隆测序并在图 2 中描述了所述序列。预期均有两个 557 个氨基酸长的 CPY 前体序列。以与来自啤酒酵

母的同源基因的比较为基础，来自黑曲霉的 CPY 似乎有一个 137 或 138 个氨基酸的前-原肽。所述基因含有一个 61 个碱基对的内含子。两个黑曲霉序列的比较表明，在氨基酸序列方面有些不同，这可能反映了所述基因组的和 cDNA 克隆是分离自不同的菌株。与啤酒酵母和 *C. albican* 的相应的 CPY 基因的比较表明分别有 65% 和 66% 的一致性。

本发明并不限于使用图 1 和 2 所公开的序列。首先，本发明还包括生产与图 1 或 2 中描述的氨基酸序列相同的，但是因遗传密码简并性而不同的核苷酸序列。另外，因相同种的两个菌株而得到不同的氨基酸序列说明在一个种内的序列上的变异是可以容忍的，而且利用本文描述的技术，可以很容易鉴定所述的变异体。因此在本文中提出“黑曲霉”时，应理解包括所有所述的变异体。具体地说，本发明还包括任何变异体核苷酸序列，及由其编码的蛋白质，其中所述蛋白质保留了与图 1 或 2 中所示的氨基酸序列至少约 80%，优选约 85%，最佳为至少约 90 - 95% 的同源性，而且在性质上保留了本文所述序列的活性。在上述定义的目录中有用的变异体包括例如，已完成的保守氨基酸取代，所述取代并未显著影响所述蛋白质的活性。所谓保守取代指相同类型的氨基酸可以由该类型的其他氨基酸取代。例如，可以交换非极性的脂肪族残基 Ala, Val, Leu, 和 Ile，碱性残基 Lys 和 Arg，或酸性残基 Asp 和 Glu。相似地，Ser 和 Thr 彼此可以进行保守取代，Asn 和 Gln 也可以。

另外，分离的基因提供了从其他丝状子囊菌或半知菌，例如曲霉种，如米曲霉，奥曲霉，日本曲霉，棘孢曲霉，或构巢曲霉中分离同源基因的方法。其他非曲霉丝状子囊菌包括镰孢属，例如，禾谷镰孢，尖孢镰孢，茄病镰孢大刀镰孢(或相应的 teleomorphs)



粗糙链孢霉, *Trichoderma reesei*, Tviridae, *T. harzianum* *T. longibranchiatum*, 微紫青霉, 点青霉, 产黄青霉, 沙门氏柏干酪青霉, 娄格法儿特氏青霉, *Humicola insolens*, 灰色腐质霉 *thermoidea* 变种, *H. lanuginosa*, *Scytalidium thermophilum* *yceliophthora thermophila*, 和 *Thielavia terrestris*。在 Southern 杂交中可以用所述基因或以其为基础的寡核苷酸作探针, 以分离这些其他种中的同源基因。具体地说, 用所述探针在低或高严谨条件下 (例如, 在 42 °C, 5 × SSPE, 0.3% SDS, 200 μg/ml 剪切并变性的鲑精 DNA, 分别为 50, 35 或 25 % 用于高, 中和低严谨条件下预杂交和杂交) 与所需种的基因组或 cDNA 进行杂交, 标准的 Southern 印迹方法后, 以鉴定并分离其中相应的 CPY 基因。用本文所述的简并探针及来自丝状真菌的基因组 DNA 或第一条链 cDNA 的 PCR 也可以得到 CPY 特异性产物, 然后可以用其作为探针克隆相应的基因组或 cDNA。

本发明基因在产生丝状子囊菌, 例如曲霉的羧肽酶-缺陷型突变体时特别有用。可以用很多方法达到该目的。在一种方法中, 如下文进一步描述的, 将一个选择性标记克隆到 CPY 基因个中部。用限制酶将破坏的片段从母本质粒中释放出。用所述线性化的 DNA 片段转化所选的宿主细胞。在所述的宿主细胞中, 同源末端与宿主染色体配对, 同源重组产生了染色体基因替代。与真菌细胞宿主可一起使用的可选择标记包括 *amdS*, *pyrG*; *argB*, *niaD*, *sC*, 和 *hygB*。另外, 用两种可选择标记可以使用两步法。在所述的一种方法中, 用在 CPY 配对内切割一次以便在转化过程中靶整合到 CPY 座位的限制酶消化含有截断的 CPY 基因和可选择标记基因的质粒。然后在选择丢失可选择标记的培养基上使转化体生长, 例如

若标记为 pyrG 时，培养基可含有 5 - fluorootic acid 。在野生型 CPY 和导入的截断的 CPY 基因之间重组时，常常发生可选择标记的丢失。与约 50 % 的所得菌株应只有截断的 CPY 基因，而另外 50 % 只含有野生型基因。在 Rothstein ,Meth. Enzymol.194,281-301,1991 中描述了所述方法。

所产生的 CPY 缺陷型突变体在表达异源蛋白质时，是特别有用的。在本发明中，“异源蛋白质”指在所述宿主中天然不存在的蛋白质，对天然序列已经进行了修饰的天然蛋白质，或用重组 DNA 技术修饰宿主后，其表达在数量上发生了变化的天然蛋白质。该术语还包括其突变体的表达使用了宿主中天然没有的遗传成分或使用了经加工后，以在宿主中不常见的方式发挥作用的天然成分的天然蛋白质。

正如已经提到的，由所选择的宿主细胞而进行的蛋白酶生产可以在其被回收前通过降解产物而严格限制了所需蛋白质的产率。因此由宿主生产的 CPY 的消除或量的减少实质上提高了所需蛋白质的产率，而且可以提供商业上可行的用于重组蛋白质生产的商业上通常不使用的宿主细胞。在优选的实施方案中，CPY 缺陷型细胞生产至少 25 %，优选至少 50 % 弱，且最佳为至少 70 % 弱的 CPY，与相应的野生型相比，最多可丧失全部的 CPY 功能。

可以用本发明的突变体真菌细胞以与野生型菌株相同的方法用于重组蛋白质的生产。本领域专业人员很容易从本文所述的具体实施方案中认识各种途径的变化，所述实施方案根据上述菌株而对方法加以改进。用表达载体可以表达活性形式的所需基因。有用的表达载体含有可以使所述载体稳定整合到宿主细胞的基因组中或使载体在宿主细胞基因组外独立自主复制的元件，以及优选的一种

或多种可以很容易地筛选转化的宿主细胞的表型标记。所述表达载体还可以包括编码启动子的序列，核糖体结合位点，翻译起始密码子和，可选择的阻遏物基因或激活物基因。为了在控制序列的指导下表达，根据本发明所用的基因优选的在适当的阅读框中与所述的控制序列可操作相连。

表达载体可以是任何可方便地利用重组 DNA 技术的载体，而且载体的选择通常取决于其所导入的宿主细胞。在本发明优选的实施方案中，宿主细胞是曲霉属的菌株。因此载体可以是自主复制载体，即可以作为染色体外个体而存在的载体，其复制独立于染色体复制，例如质粒或一种染色体外元件，小染色体，或人工染色体。另外，所述载体可以是当其导入宿主细胞中时，可整合到宿主细胞染色体并与其所整合上的染色体一起复制的载体。

在所述载体中，所需基因的序列应与适宜的启动子序列可操作相连。所述启动子可以是在所选的宿主细胞中有转录活性的任何 DNA 序列，并且可以得自编码与宿主细胞同源或异源的基因。为了在真菌宿主中转录，有用的启动子的例子是得自编码米曲霉 TAKA 淀粉酶，*Rhizomucor miehei* 天冬氨酸蛋白酶，黑曲霉中性  $\alpha$ -淀粉酶，黑曲霉酸稳定的  $\alpha$ -淀粉酶，黑曲霉或泡盛葡糖淀粉酶 (glaA)，*Rhizomucor miehei* 脂酶，米曲霉碱性蛋白酶，米曲霉丙糖磷酸异构酶或构巢曲霉乙酰胺酶基因的启动子。优选的是 TAKA - 淀粉酶和 glaA 启动子。

本发明的表达载体还可以含有适宜的转录终止子，和在真核细胞中与编码异源基因序列的 DNA 序列可操作相连的多聚腺苷酸序列。终止和多聚腺苷酸序列适宜的可从与启动子相同的来源得到。所述载体还可以含有能够使载体在所研究的宿主细胞中复制的

DNA 序列。所述序列的实例是质粒 pUC19,pACYC177,pUB110, pE194,pAMB1 和 pIJ702 的复制起点。

所述载体还可以含有选择性标记,例如其产物补充宿主细胞中的缺陷的基因,或赋予抗生素抗性,例如氨基青霉素,卡那霉素,氯霉素或四环素抗性的基因。曲霉选择标记的实例包括 amdS,PYrGmargB,niaD sC,和 hygB,产生 hygromycin 抗性的标记。在曲霉宿主细胞中优选使用构巢曲霉或米曲霉的 amdS 和 pyrG。此外,例如按 WO 91/17243 中的描述通过共转化来完成选择。

通常优选的是表达后得到细胞外产物。因此所述蛋白质含有可以将表达的蛋白质分泌到培养基中的前区。如果需要,所述前区可以是本发明蛋白质的天然的,或经编码相应前区的 DNA 序列一般取代而用不同前区或信号序列取代的。例如,所述前区可以得自曲霉种的葡糖淀粉酶或淀粉酶基因,杆菌种的淀粉酶基因,Rhizomucor miehei 的脂酶或蛋白酶基因,啤酒酵母  $\alpha$ -因子的基因或牛前原凝乳酶基因。当宿主是真菌细胞时,特别优选的是米曲霉 TAKA 淀粉酶,黑曲霉中性淀粉酶,杆菌 NCIB 11837 的 maltogenic 淀粉酶,嗜热脂肪芽孢杆菌  $\alpha$ -淀粉酶或地衣形芽孢杆菌蛋白酶的前区。一种有效的信号序列是米曲霉 TAKA 淀粉酶信号,Rhizomucor miehei 天冬氨酸蛋白酶信号和 Rhizomucor miehei 脂酶信号。

分别用于连接本发明的 DNA 构建体,启动子,终止子和其他成分然后将其插入含有复制所需的信息的适宜的载体中的方法,是本领域专业人员熟知的(参见例如,Sambrook et al.,Molecular Cloning,1989)。

可以用 CPY 缺陷型突变体表达任何所需的原核或真核蛋白

质，而且优选用于表达真核蛋白质。这些细胞特别有用的是适用于表达过氧化氢酶，漆酶，酚氧化酶，氧化酶，氧化还原酶，纤维素酶，木聚糖酶，过氧化物酶，脂酶，水解酶，酯酶，角质酶（cutinase）蛋白酶和其他蛋白水解酶，氨肽酶，羧肽酶，植酸酶，裂解酶，果胶酶和其他果胶水解（pectinolytic）酶，淀粉酶，葡糖淀粉酶，半乳糖苷酶，半乳糖苷酶， $\alpha$ -葡糖苷酶， $\beta$ -葡糖苷酶，甘露糖苷酶，异构酶，蔗糖酶，转移酶，核糖核酸酶，几丁质酶和脱氧核糖核酸酶。本领域专业人员应该理解，术语“真菌酶”不仅包括天然真菌，而且包括经氨基酸取代，缺失，加入或为提高其活性，热稳定性，pH耐受性等而修饰过的真菌酶。也可以用所述突变体表达药理学上的异源蛋白质，例如激素，生长因子，受体等。

将用下列实施例对本发明作进一步的说明。

## 实施例

### I. 分离黑曲霉 CPY 基因

#### A. 材料和方法

##### i. 菌株

在下文描述的方法中使用了下列生物材料。E. coli K802(ek4-(nrca),mcr B,hsdR2,galD2,GalT22,wupE44, metB1; E.coli SOLP(E14-(mcrA) (mcrCB-hsdSMR-mr<sup>r</sup>)171,sbcC,recB,recJ,uvrC,umuD::Tn5(kan<sup>r</sup>),lac gryS96,relA1,thi-1,endA1, $\lambda$ <sup>R</sup>[F'proAB lacIqZ  $\Delta$  M15]Su-, E. coli JM101supE,thi-1,  $\Delta$ (lac-proAB),[F'traD36,proAB, lacIqZ  $\Delta$  M15], E.coli XL-1 Blue rec A1,endA1,gyr A96,thi-1,hsdR17,supE44,relA1,lac [F'proAB lacIqZ  $\Delta$  M15,Tn10(tet<sup>R</sup>)],黑曲霉 Bo-1,黑曲霉 SFAG - 2。

## ii. PCR 扩增

用标准技术在 45 °C 退火来进行 PCR。用黑曲霉 Bo-1 基因组 DNA 作为模板，用下列简并寡核苷酸。

引物 1-1 (94-282) -GGIGGICCGGTTGYTC  
引物 1-2 (94-283) -GGIGGICCGGTTGYAG  
引物 2-1 (94-284) -CCIAGCCARTTRCADAT  
引物 2-2 (94-285) -CCYAACCARTTRCADAT  
引物 3-1 (94-331) -GTIGGITTYTCITAYTCIGG  
引物 3-2 (94-332) -GTIGGITTYAGYTAYAGYGG  
引物 4-1 (94-329) -GARTCITAYGCIGGICAYTA  
引物 2-1 (94-284) -GARAGYTAYGCIGGICAYTA

在上述引物中 I 代表肌苷，Y 代表 C 或 T，R 代表 A 或 G，D 代表 A，G 或 T。

## iii. 亚克隆 PCR 产物

按制造商的说明，用 TA Cloning 试剂盒 (Invitrogen) 将 PCR 产物亚克隆以进行测序。

#### iv. 从 $\lambda$ Zap II 体内切割

通过在大肠杆菌 SOLR 中传代, 然后按照 Strategene 提供的方法, 可从 CPY cDNA  $\lambda$  Zap 克隆获得含有在 pBluescript SK 载体中携带了 cDNA 插入子的质粒。

#### v. DNA 测序

用荧光标记的核苷酸, 利用聚合酶循环测序确定核苷酸测序。在 Applied Biosystems 自动 DNA 测序仪 ( Model 363 A ,1.20 版) 上电泳测序反应产物。使用下列 CPY 特异性引物, 再使用 M13 反向 (- 48 ) 和 M13 (- 20 ) 前向引物(Sanger et al., J. Mol. Biol. 143:163-178):

94-376	TCGCTGCCAGTCTATGATTGA
94-377	ACATCAACCGCAACTTCCTCT
94-378	TTGCCAATGAGAACGGACTGC
94-379	CGCACTTACCACGGACATCAT
94-503	CAAGCATCCTCAAACCTATCGT
94-504	GAGACGCATGAAGGTGAAGTT
94-505	GCCGTCCCTCCCTTCCAGCAG
94-506	GTGCCGACGGGTTCTCCAAGC
94-507	GCAGCGAGGAAGAGCGTTGTC
94-510	GGGTCATCTCGGGGTCATTG
94-511	GACCCCGAGAATGACCCTGTT
94-512	GTAGGGCTTCATCCAGTCACC
94-513	TCTCACCGTTCTCACCAGTAA
94-514	TCCCTCCCCAAGAAGCACAAC
94-528	AGCGTCTGGGTTACTGGTGAG
94-529	AAGATCGGCCAGGTCAAGTCC
94-530	GAGACGGTGGTAGGGCTTCAT
94-531	AACGTCCGGTACTCTTACAGC
94-532	GTGGTCCGGGCGGCGGTTGTG
94-533	TGTTTGAAGAAGAGGGTAAGC
94-575	CGCTGCTACTTGATTTTTCTA
94-576	CTCAGCGCCAACAGCCTCAAT
94-577	ACCTGCAGTCCGTTCTTATTG
94-634	TGCGATCGATTTCATTCTCATC
94-635	GGAGTAACCGACATTGACAGG
94-636	CCTGTCAATGTCGGTACTCC
94-637	GTCCCATGGCAACTTCACCTT
94-646	CTTCTCACCGTTCTCACCAGT
94-647	CGAGACTCGAAGAACCCTAAG

## B. 结果

用黑曲霉 Bo-1 基因组 DNA 为模板，使用各种组合的 CPY 特异性简并寡核苷酸，引物 1-1，1-2，2-1 和 2-2（图 1）完成 PCR 反应。所有的反应均以 95 °C 5 分钟，45 °C 1 分钟和 72 °C 2 分钟，接着进行 25 个下一循环，即 95 °C 1 分钟，45 °C 1 分钟和 72 °C 2 分钟。在琼脂糖凝胶上将反应物小样（10 μl）电泳，在两个反应中，一个使用引物 1-2 和 2-1，另一个使用引物 1-2 和 2-2，约 1100 bp 的扩增产物占大部分。使用假设在那儿的这些寡核苷酸组合而预期的产物的大小在 900 bp 的基因内没有内含子，使用前向和反向引物将 1100 bp 的扩增产物亚克隆并测序。7 个亚克隆被测序，但是，没有一个与 CPY 密码子同源。

使用与上述相同的条件，完成使用各种引物组合 3-1，3-2，4-1，4-2，2-1 和 2-2 的 PCR。在琼脂糖凝胶上电泳小样，在两个反应中，一个使用引物 4-1 和 2-1，另一个使用引物 4-2 和 2-1，约 600 bp 的扩增产物占大部分。根据与其它羧肽酶的同源性而预测的该扩增产物的大小为 600bp。将 600 bp 的扩增产物亚克隆，用前向和反向引物确定 11 个亚克隆的 DNA 序列。11 个克隆中的 9 个以与啤酒酵母有 69% 的一致性为基础，为黑曲霉 CPY 的密码子。所有 9 个均彼此相同，说明在黑曲霉中只有一个羧肽酶基因。将含有 600 bp 黑曲霉 CPY PCR 产物的亚克隆称为 pDSY17。

用 pDSY17 的插入子探测黑曲霉 Bo-1 基因组 DNA 的 Southern 印迹。用 Gibco-BRL 的缺口-翻译试剂盒放射性标记探针。所用的



杂交条件为于 60 °C, 在 1.5 × SSPE, 1 % SDS 和 0.5% 脱脂牛奶和 200 μ g/ml 鲑精 DNA。于 65 °C 用 0.2 × SSC, 1 % SDS 和 0.1 % 焦磷酸钠将印迹洗涤两次, 每次 15 分钟。在 BamHI, HindIII 和 Sall 的消化产物中, 约 10, 5.5 和 7 kb 的单带分别与 CPY 探针杂交。

为了分离 CPY 的全长基因, 用 PCR - 产生的 CPY 基因片段 (用 Gibco-BRL 的缺口 - 翻译试剂盒标记的), 探测含有约 26, 000 个重组体的黑曲霉 Bo-1 的 EMBL4 中的基因组库。将约 28, 000 个噬菌斑放在过滤器上并检测。从这些平皿中检出 11 个阳性的。纯化后, 9 个一级克隆仍与 CPY 探针杂交。从 9 个克隆中分离 DNA, 进行限制酶消化以对其进行表征。从限制图谱看, 9 个中有 7 个是相同的。其他两个克隆是特有的。从克隆的 Southern 产物, 确定 9 个中的 8 个含有与 CPY 探针杂交的约 5.5 kb 的相同 HindIII 片段。不含有相同 HindIII 片段的克隆含有较大 (> 12 kb) 的与 CPY 探针杂交的 HindIII 片段。亚克隆共有的 HindIII 片段以进行 DNA 测序。基因组 DNA 序列和推测的氨基酸序列示于图 1。

再筛选黑曲霉 SFAG - 2 的 λ ZAPII(Stratagene) 中的 cDNA 库。将 42, 000 个噬菌斑放到过滤器上, 然后用上述的 CPY 探针检测, 在严格的上述条件下, 112 个这样的噬菌斑进行了杂交。检出开始的 20 个, 然后进行再筛选, 纯化后, 18 个仍与 CPY 探针杂交。从 4 个阳性克隆中, 用配有 λ ZAP 试剂盒的体内切割方法分离 DNA。用 EcoRI 消化获救的质粒, 然后在琼脂糖凝胶上电泳以确定插入片段的大小。两个克隆 (2 - 1 和 3 - 2) 似乎有足够大的插入片段, 可含有全长 CPY 的 cDNA, 而且各含有约 1700 和 250 bp 的两个 EcoRI 片段。预计的全长 cDNA 的大小约为 1600 bp。

另两个 cDNA 克隆 (2-2 和 2-4) 含有较小的插入片段, 但是, 它们均含有 250 bp 的 EcoRI 片段。确定了克隆 3-2 和 2-2 的部分 DNA 序列, 3-2 含有全长 cDNA, 而克隆 2-2 在 5' 末端被截断约 200 bp。

确定了两个链的完整 cDNA 序列 (图 2)。认为 cDNA 编码 557 个氨基酸长的 CPY 前体。到目前为止, cDNA 和基因组克隆之间的大多数核苷酸差异均处于摆动范围内, 这点不令人惊奇, 因为它们是来自两个不同的黑曲霉菌株。以啤酒酵母的 CPY 排列为基础, 黑曲霉 CPY 看起来含有信号肽和前肽, 成熟 CPY 蛋白质为 419 或 420 个氨基酸长。黑曲霉 CPY 与啤酒酵母和 *C. albicans* 的 CPY (Mukhtar et al., Gene 121:173-177, 1992) 分别有约 65% 和 66% 的一致性。

## II 制备 CPY 缺陷型突变体

为了产生缺失了 CPY 的黑曲霉菌株, 制备其中米曲霉 *pyrG* 基因插入到 CPY 编码区中的构建体 (图 3)。将含有几乎完整 CPY 基因和约 6 kb 基因下游的约 6.5 kb HindIII 片段亚克隆到 pKS+(Stratagene) 衍生物中, 其中破坏了 PstI 位点。用 PstI 消化所得的重组体, 以便从 CPY 编码区缺失 815 bp 的片段, 加入 T4 DNA 聚合酶和所有 4 种 dNTP 而补平因 PstI 消化而产生的悬末端。将所得的平整末端载体与通过用 Hind III 消化, 然后用 Klenow 片段填平而得到的约 3.8 kb 平整末端的片段相连。用 Hind III 消化其中含有 *pyrG* 插入片段的最终构建体, 以得到用于转化黑曲霉 *pyrG* 菌株的线性片段, 在基本培养基上生长进行筛选。用 Southern 印迹筛选转化体以确定那些菌株含有破坏的 CPY 基因。用 Western 印迹分析转化体以寻找在细胞内丢失的 CPY。一旦鉴定了一个菌株含

有破坏了的 CPY，就可以确定对异源蛋白质的影响。

### 生物材料的保藏

已于 1994 年 9 月 13 日将下列生物材料保藏在 Agricultural Research Service Culture Collection (NRRL)1815 North University Street, Peoria, Illinois 61604。

细胞系

保藏号

含有 pDSY23 的  
E.coli(EMCC#0120)

NRRL B-21326

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

(D) 州: California

(E) 国家: US

(F) 邮编: 95616-4880

(G) 电话: (916) 757-8100

(H) 电传: (916) 758-0317

(ii) 发明题目: 编码黑曲霉羧肽酶的基因

(iii) 序列数目: 4

(iv) 相关地址:

(A) 收件人: Novo Nordisk of North America, Inc.

(B) 街道: 405 Lexington Avenue, Suite 6400

(C) 城市: New York

(D) 州: New York

(E) 国家: U.S.A.

(F) 邮编: 10174-6401

(v) 计算机可读形式:

(A) 介质类型: 软盘

(B)计算机: IBM PC

(C)操作系统: PC-DOS/MS-DOS

(D)软件: PatentIn Release #1.0, Version #1.25 (EPO)

(vi)目前的申请资料:

(A)申请号: US

(B)申请日: 19-9月-1995

(C)分类号:

(vii)在先申请资料:

(A)申请号: US 08/309, 341

(B)申请日: 20-9月-1994

(viii)代理人/代理资料:

(A)姓名: Lowney, Karen A.

(B)登记号: 31,274

(C)文件号: 4247.204-WO

(ix)通讯资料:

(A)电话: 212 867 0123

(B)电传: 212 867 0298

(2)SEQ ID NO:1 的资料:

(i)序列特征:

(A)长度: 2068 个碱基对

(B)类型: 核酸

(C)链型: 双链

(D)拓扑结构:线性

(ii)分子类型:基因组 DNA

(vi)来源:

(A)生物:黑曲霉

(ix)特征:

(A)名称/关键词:内含子

(B)位置:572...632

(ix)特征:

(A)名称/关键词:CDS

(B)位置:连接(571..633)

(xi)序列描述:SEQ ID NO:1:

```
TCCTCTGCCT ACTCATCCCA TCACCATCTC AATTCATACC GCCCCCGTGG GGTTCAGCA      60

CCA ATG AGA GTC CTT CCA GCT GCT ATG CTG GTT GGA GCG GCC ACG GCG      108
  Met Arg Val Leu Pro Ala Ala Met Leu Val Gly Ala Ala Thr Ala
   1                    5                    10                    15

GCC GTT CCT CCC TTC CAG CAG GTC CTT GGA GGT AAC GGT GCC AAG CAC      156
Ala Val Pro Pro Phe Gln Gln Val Leu Gly Gly Asn Gly Ala Lys His
                20                    25                    30

GGT GCC GAC CAT GCG GCC GAG GTC CCT GCG GAT CAC AGT GCC GAC GGG      204
Gly Ala Asp His Ala Ala Glu Val Pro Ala Asp His Ser Ala Asp Gly
                35                    40                    45

TTC TCC AAG CCG CTG CAC GCA TTC CAG GAG GAG CTG AAG TCT CTC TCT      252
Phe Ser Lys Pro Leu His Ala Phe Gln Glu Glu Leu Lys Ser Leu Ser
                50                    55                    60

GAC GAG GCT CGT AAG CTT TGG GAT GAG GTG GCC AGC TTC TTC CCG GAG      300
Asp Glu Ala Arg Lys Leu Trp Asp Glu Val Ala Ser Phe Phe Pro Glu
   65                    70                    75

AGC ATG GAT CAG AAC CCT CTC TTT TCC CTC CCC AAG AAG CAC AAC CGC      348
Ser Met Asp Gln Asn Pro Leu Phe Ser Leu Pro Lys Lys His Asn Arg
   80                    85                    90                    95

CGT CCC GAC TCG CAC TGG GAC CAC ATC GTC CGC GGC TCC GAC GTT CAG      396
Arg Pro Asp Ser His Trp Asp His Ile Val Arg Gly Ser Asp Val Gln
                100                    105                    110

AGC GTC TGG GTC ACT GGT GAG AAC GGT GAG AAG GAG CGC GAG GTC GAT      444
Ser Val Trp Val Thr Gly Glu Asn Gly Glu Lys Glu Arg Glu Val Asp
                115                    120                    125
```

GGC AAG CTG GAA GCC TAT GAT CTC AGG GTC AAG AAG ACC GAT CCT GGC Gly Lys Leu Glu Ala Tyr Asp Leu Arg Val Lys Lys Thr Asp Pro Gly 130 135 140	492
TCT CTT GGC ATC GAC CCC GGC GTG AAG CAG TAC ACC GGT TAT CTC GAT Ser Leu Gly Ile Asp Pro Gly Val Lys Gln Tyr Thr Gly Tyr Leu Asp 145 150 155	540
GAC AAC GAG AAT GAT AAG CAT TTG TTC TAC GTAAGCACAC CTTGGTTCAA Asp Asn Glu Asn Asp Lys His Leu Phe Tyr 160 165	590
GATCACGCTT TTTATATGCT CTGGATATCT AACGCAACTT AG TGG TTC TTC GAG Trp Phe Phe Glu 170	644
TCT CGC AAT GAC CCC GAG AAT GAT CCC GTT GTT CTG TGG CTG AAC GGT Ser Arg Asn Asp Pro Glu Asn Asp Pro Val Val Leu Trp Leu Asn Gly 175 180 185	692
GGC CCT GGG TGC TCT TCC CTC ACC GGT CTC TTC ATG GAG CTT GGC CCT Gly Pro Gly Cys Ser Ser Leu Thr Gly Leu Phe Met Glu Leu Gly Pro 190 195 200 205	740
AGC AGC ATC AAC AAG AAG ATC CAG CCG GTC TAC AAT GAC TAC GCT TGG Ser Ser Ile Asn Lys Lys Ile Gln Pro Val Tyr Asn Asp Tyr Ala Trp 210 215 220	788
AAC TCC AAC GCG TCC GTG ATC TTC CTT GAC CAG CCT GTC AAT GTC GGT Asn Ser Asn Ala Ser Val Ile Phe Leu Asp Gln Pro Val Asn Val Gly 225 230 235	836
TAC TCC TAC AGT AAC TCT GCT GTC AGC GAC ACG GTC GCT GCT GGC AAG Tyr Ser Tyr Ser Asn Ser Ala Val Ser Asp Thr Val Ala Ala Gly Lys 240 245 250	884
GAC GTC TAT GCC TTG CTT ACC CTC TTC TTC AAA CAA TTC CCC GAG TAT Asp Val Tyr Ala Leu Leu Thr Leu Phe Phe Lys Gln Phe Pro Glu Tyr 255 260 265	932
GCT AAG CAG GAC TTC CAC ATT GCC GGT GAA TCT TAT GCT GGT CAC TAT Ala Lys Gln Asp Phe His Ile Ala Gly Glu Ser Tyr Ala Gly His Tyr 270 275 280 285	980
ATC CCC GTC TTC GCT TCG GAG ATC CTG TCT CAC AAG AAG CGC AAC ATC Ile Pro Val Phe Ala Ser Glu Ile Leu Ser His Lys Lys Arg Asn Ile 290 295 300	1028
AAC CTG CAG TCC GTT CTC ATT GGC AAC GGT CTC ACC GAC GGA TAC ACC Asn Leu Gln Ser Val Leu Ile Gly Asn Gly Leu Thr Asp Gly Tyr Thr 305 310 315	1076
CAG TAC GAG TAC TAC CGT CCC ATG GCC TGC GGT GAC GGC GGT TAC CCA Gln Tyr Glu Tyr Tyr Arg Pro Met Ala Cys Gly Asp Gly Gly Tyr Pro 320 325 330	1124
GCT GTC TTG GAC GAG AGC TCC TGC CAG TCC ATG GAC AAC GCT CTT CCT Ala Val Leu Asp Glu Ser Ser Cys Gln Ser Met Asp Asn Ala Leu Pro 335 340 345	1172
CGC TGC CAG TCT ATG ATT GAG TCT TGC TAC AGT TCC GAG AGC GCT TGG Arg Cys Gln Ser Met Ile Glu Ser Cys Tyr Ser Ser Glu Ser Ala Trp 350 355 360 365	1220
GTT TGT GTC CCG GCC TCC ATC TAC TGT AAC AAC GCC CTC CTT GCC CCT Val Cys Val Pro Ala Ser Ile Tyr Cys Asn Asn Ala Leu Leu Ala Pro 370 375 380	1268

TAC CAG CGC ACT GGG CAG AAC GTC TAT GAT GTC CGT GGT AAG TGC GAG Tyr Gln Arg Thr Gly Gln Asn Val Tyr Asp Val Arg Gly Lys Cys Glu 385 390 395	1316
GAT AGC TCT AAC CTT TGC TAC TCG GCT ATG GGC TAC GTC AGC GAC TAC Asp Ser Ser Asn Leu Cys Tyr Ser Ala Met Gly Tyr Val Ser Asp Tyr 400 405 410	1364
CTG AAC AAG CCC GAA GTC ATC GAG GCT GTT GGC GCT GAG GTC AAC GGC Leu Asn Lys Pro Glu Val Ile Glu Ala Val Gly Ala Glu Val Asn Gly 415 420 425	1412
TAC GAC TCG TGC AAC TTT GAC ATC AAC CGC AAC TTC CTC TTC CAC GGT Tyr Asp Ser Cys Asn Phe Asp Ile Asn Arg Asn Phe Leu Phe His Gly 430 435 440 445	1460
GAC TGG ATG AAG CCC TAC CAC CGC CTC GTT CCG GGA CTC CTG GAG CAG Asp Trp Met Lys Pro Tyr His Arg Leu Val Pro Gly Leu Leu Glu Gln 450 455 460	1508
ATC CCT GTC TTG ATC TAT GCC GGT GAT GCT GAT TTC ATT TGC AAC TGG Ile Pro Val Leu Ile Tyr Ala Gly Asp Ala Asp Phe Ile Cys Asn Trp 465 470 475	1556
CTG GGC AAC AAG GCC TGG ACT GAA GCC CTG GAG TGG CCC GGA CAG GCT Leu Gly Asn Lys Ala Trp Thr Glu Ala Leu Glu Trp Pro Gly Gln Ala 480 485 490	1604
GAA TAT GCC TCC GCT GAG CTG GAG GAT CTG GTC ATT GTC GAC AAT GAG Glu Tyr Ala Ser Ala Glu Leu Glu Asp Leu Val Ile Val Asp Asn Glu 495 500 505	1652
CAC ACG GGC AAG AAG ATT GGC CAG GTT AAG TCC CAT GGC AAC TTC ACC His Thr Gly Lys Lys Ile Gly Gln Val Lys Ser His Gly Asn Phe Thr 510 515 520 525	1700
TTC ATG CGT CTC TAT GGT GGT GGC CAC ATG GTC CCG ATG GAC CAG CCC Phe Met Arg Leu Tyr Gly Gly Gly His Met Val Pro Met Asp Gln Pro 530 535 540	1748
GAG TCG AGT CTC GAG TTC TTC AAC CGC TGG TTG GGA GGT GAA TGG TTC Glu Ser Ser Leu Glu Phe Phe Asn Arg Trp Leu Gly Gly Glu Trp Phe 545 550 555	1796
TAA AGACGTGCTA CCACCGCATA TAGACTTTCT GGTCATTTTCG GTGACACTGC	1849
AGATATGTTT CTTAACGATA GTTTGAGCAT GCTTGTC AAT GCCCACTAGT CCCGATCCTT	1909
ATATGTTGCA TGGTATCTAT GAGTTTGTGTC ACTATAGTGC ATTATACATG TGTA CTTCGT	1969
ATGAGAATGA ATCGATCGCA TTTACACGCA TATAAATAGT ACCCACCTCC GCCTGGACAT	2029
GAATTAGGCC CGGCCAGTCG TTTACATACA GTGCTAGAA	2068

(2) SEQ ID NO:2 的资料:

(i) 序列特征:

- (A) 长度: 557 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:蛋白质

(vi)来源:

(A)生物:黑曲霉

(xi)序列描述: SEQ ID NO:2:

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

Phe Ala Ser Glu Ile Leu Ser His Lys Lys Arg Asn Ile Asn Leu Gln  
 290 295 300  
 Ser Val Leu Ile Gly Asn Gly Leu Thr Asp Gly Tyr Thr Gln Tyr Glu  
 305 310 315 320  
 Tyr Tyr Arg Pro Met Ala Cys Gly Asp Gly Gly Tyr Pro Ala Val Leu  
 325 330 335  
 Asp Glu Ser Ser Cys Gln Ser Met Asp Asn Ala Leu Pro Arg Cys Gln  
 340 345 350  
 Ser Met Ile Glu Ser Cys Tyr Ser Ser Glu Ser Ala Trp Val Cys Val  
 355 360 365  
 Pro Ala Ser Ile Tyr Cys Asn Asn Ala Leu Leu Ala Pro Tyr Gln Arg  
 370 375 380  
 Thr Gly Gln Asn Val Tyr Asp Val Arg Gly Lys Cys Glu Asp Ser Ser  
 385 390 395 400  
 Asn Leu Cys Tyr Ser Ala Met Gly Tyr Val Ser Asp Tyr Leu Asn Lys  
 405 410 415  
 Pro Glu Val Ile Glu Ala Val Gly Ala Glu Val Asn Gly Tyr Asp Ser  
 420 425 430  
 Cys Asn Phe Asp Ile Asn Arg Asn Phe Leu Phe His Gly Asp Trp Met  
 435 440 445  
 Lys Pro Tyr His Arg Leu Val Pro Gly Leu Leu Glu Gln Ile Pro Val  
 450 455 460  
 Leu Ile Tyr Ala Gly Asp Ala Asp Phe Ile Cys Asn Trp Leu Gly Asn  
 465 470 475 480  
 Lys Ala Trp Thr Glu Ala Leu Glu Trp Pro Gly Gln Ala Glu Tyr Ala  
 485 490 495  
 Ser Ala Glu Leu Glu Asp Leu Val Ile Val Asp Asn Glu His Thr Gly  
 500 505 510  
 Lys Lys Ile Gly Gln Val Lys Ser His Gly Asn Phe Thr Phe Met Arg  
 515 520 525  
 Leu Tyr Gly Gly Gly His Met Val Pro Met Asp Gln Pro Glu Ser Ser  
 530 535 540  
 Leu Glu Phe Phe Asn Arg Trp Leu Gly Gly Glu Trp Phe  
 545 550 555

(2)SEQ ID NO:3 的资料:

(i)序列特征:

(A)长度:2002 个碱基对

(B)类型:核酸

(C)链型:双链

(D)拓扑结构:线性

(ii)分子类型:cDNA

(vi)来源:

(A)生物:黑曲霉

(ix)特征:

(A)名称/关键词:内含子

(B)位置:349...411

(ix)特征:

(A)名称/关键词:CDS

(B)位置:连接(348..412)

(xi)序列描述:SEQ ID NO:3:

CGGGCCGCTG CTACTTGCTT TTTCTAATTT GATACTTTTG TGTCGGTACC GTACCTTCCA	60
GACCGCAAGG TACCCATCCT CTACCTACTC ATCCCATCAT CATCTCGATT TCATACCAAC	120
CCCGTTGGGT TTCAACACA ATG AGA GTT CTT CCA GCT GCT ATG CTG GTT GGA	172
Met Arg Val Leu Pro Ala Ala Met Leu Val Gly	
1 5 10	
GCG GGC ACT GCG GCC GTC CCT CCC TTC CAG CAG GTC CTT GGA GGT AAC	220
Ala Gly Thr Ala Ala Val Pro Pro Phe Gln Gln Val Leu Gly Gly Asn	
15 20 25	
GGT GCC AAG CAC GGT GCC GAC CAT GCG GCC GAG GTC CCT GCG GAT CAC	268
Gly Ala Lys His Gly Ala Asp His Ala Ala Glu Val Pro Ala Asp His	
30 35 40	
AGT GCC GAC GGG TTC TCC AAG CCG CTG CAC GCA TTC CAG GAG GAG CTG	316
Ser Ala Asp Gly Phe Ser Lys Pro Leu His Ala Phe Gln Glu Glu Leu	
45 50 55	
AAG TCT CTC TCT GAT GAG GCT CGT AAG CTC TGG GAT GAG GTT GCT AGC	364
Lys Ser Leu Ser Asp Glu Ala Arg Lys Leu Trp Asp Glu Val Ala Ser	
60 65 70 75	
TTC TTC CCG GAG AGC ATG GAT CAG AAC CCT CTC TTC TCC CTC CCC AAG	412
Phe Phe Pro Glu Ser Met Asp Gln Asn Pro Leu Phe Ser Leu Pro Lys	
80 85 90	
AAG CAC AAC CGC CGC CCC GAC CAC CAC TGG GAC CAC ATC GTC CGC GGC	460
Lys His Asn Arg Arg Pro Asp His His Trp Asp His Ile Val Arg Gly	
95 100 105	
TCC GAC GTT CAG AGC GTC TGG GTT ACT GGT GAG AAC GGT GAG AAG GAG	508
Ser Asp Val Gln Ser Val Trp Val Thr Gly Glu Asn Gly Glu Lys Glu	
110 115 120	

CGT Arg	GAG Glu	GTC Val	GAT Asp	GGC Gly	AAG Lys	CTG Leu	GAA Glu	GCC Ala	TAT Tyr	GAT Asp	CTC Leu	AGG Arg	GTC Val	AAG Lys	AAG Lys	556
125 130 135																
ACC Thr	GAT Asp	CCT Pro	AGC Ser	TCT Ser	CTT Leu	GGC Gly	ATC Ile	GAC Asp	CCT Pro	GGC Gly	GTA Val	AAG Lys	CAG Gln	TAC Tyr	ACC Thr	604
140 145 150																
GGT Gly	TAT Tyr	CTC Leu	GAT Asp	GAC Asp	AAC Asn	GAG Glu	AAC Asn	GAC Asp	AAG Lys	CAT His	CTG Leu	TTC Phe	TAC Tyr	TGG Trp	TTC Phe	652
160 165 170																
TTC Phe	GAG Glu	TCT Ser	CGC Arg	AAT Asn	GAC Asp	CCC Pro	GAG Glu	AAT Asn	GAC Asp	CCT Pro	GTT Val	GTT Val	CTG Leu	TGG Trp	CTG Leu	700
175 180 185																
AAC Asn	GGT Gly	GGC Gly	CCT Pro	GGA Gly	TGC Cys	TCT Ser	TCC Ser	CTC Leu	ACC Thr	GGT Gly	CTT Leu	TTC Phe	ATG Met	GAG Glu	CTC Leu	748
190 195 200																
GGC Gly	CCT Pro	AGC Ser	AGC Ser	ATC Ile	AAC Asn	AAG Lys	AAG Lys	ATC Ile	CAG Gln	CCG Pro	GTC Val	TAC Tyr	AAC Asn	GAC Asp	TAC Tyr	796
205 210 215																
GCT Ala	TGG Trp	AAC Asn	TCC Ser	AAC Asn	GCG Ala	TCC Ser	GTG Val	ATC Ile	TTC Phe	CTT Leu	GAC Asp	CAG Gln	CCT Pro	GTC Val	AAC Asn	844
220 225 230 235																
GTC Val	GGT Gly	TAC Tyr	TCT Ser	TAC Tyr	AGC Ser	AAC Asn	TCT Ser	GCT Ala	GTC Val	AGC Ser	GAC Asp	ACC Thr	GTT Val	GCT Ala	GCT Ala	892
240 245 250																
GGC Gly	AAG Lys	GAC Asp	GTC Val	TAT Tyr	GCC Ala	TTG Leu	CTT Leu	ACC Thr	CTC Leu	TTC Phe	TTC Phe	AAA Lys	CAA Gln	TTC Phe	CCC Pro	940
255 260 265																
GAG Glu	TAT Tyr	GCC Ala	AAG Lys	CAG Gln	GAC Asp	TTC Phe	CAC His	ATT Ile	GCC Ala	GGT Gly	GAA Glu	TCC Ser	TAT Tyr	GCT Ala	GGT Gly	988
270 275 280																
CAC His	TAT Tyr	ATC Ile	CCC Pro	GTC Val	TTT Phe	GCT Ala	TCG Ser	GAG Glu	ATT Ile	TTG Leu	TCT Ser	CAC His	AAG Lys	AAG Lys	CGC Arg	1036
285 290 295																
AAC Asn	ATC Ile	AAC Asn	CTG Leu	CAG Gln	TCC Ser	GTT Val	CTT Leu	ATT Ile	GGC Gly	AAC Asn	GGT Gly	CTC Leu	ACC Thr	GAC Asp	GGT Gly	1084
300 305 310 315																
CTC Leu	ACT Thr	CAG Gln	TAC Tyr	GAG Glu	TAC Tyr	TAC Tyr	CGT Arg	CCC Pro	ATG Met	GCC Ala	TGT Cys	GGT Gly	GAC Asp	GGT Gly	GGT Gly	1132
320 325 330																
TAC Tyr	CCA Pro	GCT Ala	GTC Val	TTG Leu	GAC Asp	GAG Glu	GGC Gly	TCC Ser	TGC Cys	CAG Gln	GCC Ala	ATG Met	GAC Asp	AAC Asn	GCC Ala	1180
340 345																
CTT Leu	CCT Pro	CGC Arg	TGC Cys	CAG Gln	TCT Ser	ATG Met	ATT Ile	GAG Glu	TCT Ser	TGC Cys	TAT Tyr	AGT Ser	TCC Ser	GAG Glu	AGC Ser	1228
350 355 360																
GCT Ala	TGG Trp	GTT Val	TGT Cys	GTC Val	CCG Pro	GCC Ala	TCC Ser	ATC Ile	TAC Tyr	TGT Cys	AAC Asn	AAC Asn	GCC Ala	CTC Leu	CTT Leu	1276
365 370 375																
GCC Ala	CCT Pro	TAC Tyr	CAG Gln	CGC Arg	ACC Thr	GGA Gly	CAG Gln	AAC Asn	GTC Val	TAC Tyr	GAT Asp	GTT Val	CGT Arg	GGT Gly	AAG Lys	1324
380 385 390 395																
TGC Cys	GAG Glu	GAT Asp	AGC Ser	TCC Ser	AAC Asn	CTC Leu	TGC Cys	TAC Tyr	TCG Ser	GCC Ala	ATG Met	GGC Gly	TAC Tyr	GTC Val	AGC Ser	1372
400 405 410																

GAC TAC CTG AAC AAG ACC GAG GTC ATT GAG GCT GTT GGC GCT GAG GTC Asp Tyr Leu Asn Lys Thr Glu Val Ile Glu Ala Val Gly Ala Glu Val 415 420 425	1420
AAC GGC TAC GAC TCG TGC AAC TTT GAC ATC AAC CGC AAC TTC CTC TTC Asn Gly Tyr Asp Ser Cys Asn Phe Asp Ile Asn Arg Asn Phe Leu Phe 430 435 440	1468
CAC GGT GAC TGG ATG AAG CCC TAC CAC CGT CTC GTT CCG GGA CTC CTG His Gly Asp Trp Met Lys Pro Tyr His Arg Leu Val Pro Gly Leu Leu 445 450 455	1516
GAG CAG ATC CCT GTC CTG ATC TAC GCT GGT GAC GCC GAT TTC ATC TGC Glu Gln Ile Pro Val Leu Ile Tyr Ala Gly Asp Ala Asp Phe Ile Cys 460 465 470 475	1564
AAC TGG CTG GGC AAC AAG GCC TGG ACT GAA GCC CTT GAG TGG CCC GGA Asn Trp Leu Gly Asn Lys Ala Trp Thr Glu Ala Leu Glu Trp Pro Gly 480 485 490	1612
CAG GCT GAA TAT GCC TCC GCT AAG CTG GAG GAC CTG GTC GTG GTC GAG Gln Ala Glu Tyr Ala Ser Ala Lys Leu Glu Asp Leu Val Val Val Glu 495 500 505	1660
AAT GAG CAC AAG GGC AAG AAG ATC GGC CAG GTC AAG TCC CAT GGC AAC Asn Glu His Lys Gly Lys Lys Ile Gly Gln Val Lys Ser His Gly Asn 510 515 520	1708
TTC ACC TTC ATG CGT CTC TAT GGC GGT GGC CAC ATG GTC CCG ATG GAC Phe Thr Phe Met Arg Leu Tyr Gly Gly Gly His Met Val Pro Met Asp 525 530 535	1756
CAA CCC GAG TCG AGT CTT GAA TTC TTC AAC CGC TGG TTG GGA GGT GAA Gln Pro Glu Ser Ser Leu Glu Phe Phe Asn Arg Trp Leu Gly Gly Glu 540 545 550 555	1804
TGG TTT TAA AGACGTGCTA TCACCGCATA TAGACTTTCC GGTCATTTCC GTGACACTGC Trp Phe	1863
AGATATGTTT CTTAACGATA GTTTGAGGAT GCTTGTCAAT GCCCACTAAT CCCGAGCCTT	1923
ATGTTACATG GTATCTATGA GTTTGTCAAT ATAGTGCATT ATGCATTTGT ACTCCGTACG	1983
AGAATGAATC AGCGGCCGC	2002

(2)SEQ ID NO:4 的资料:

(i)序列特征:

(A)长度:557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:蛋白质

(vi)来源:

(A)生物:黑曲霉

(xi)序列描述:SEQ ID NO:4:

Met Arg Val Leu Pro Ala Ala Met Leu Val Gly Ala Gly Thr Ala Ala  
1 5 10 15  
Val Pro Pro Phe Gln Gln Val Leu Gly Gly Asn Gly Ala Lys His Gly  
20 25 30  
Ala Asp His Ala Ala Glu Val Pro Ala Asp His Ser Ala Asp Gly Phe  
35 40 45  
Ser Lys Pro Leu His Ala Phe Gln Glu Glu Leu Lys Ser Leu Ser Asp  
50 55 60  
Glu Ala Arg Lys Leu Trp Asp Glu Val Ala Ser Phe Phe Pro Glu Ser  
65 70 75 80  
Met Asp Gln Asn Pro Leu Phe Ser Leu Pro Lys Lys His Asn Arg Arg  
85 90 95  
Pro Asp His His Trp Asp His Ile Val Arg Gly Ser Asp Val Gln Ser  
100 105 110  
Val Trp Val Thr Gly Glu Asn Gly Glu Lys Glu Arg Glu Val Asp Gly  
115 120 125  
Lys Leu Glu Ala Tyr Asp Leu Arg Val Lys Lys Thr Asp Pro Ser Ser  
130 135 140  
Leu Gly Ile Asp Pro Gly Val Lys Gln Tyr Thr Gly Tyr Leu Asp Asp  
145 150 155 160  
Asn Glu Asn Asp Lys His Leu Phe Tyr Trp Phe Phe Glu Ser Arg Asn  
165 170 175  
Asp Pro Glu Asn Asp Pro Val Val Leu Trp Leu Asn Gly Gly Pro Gly  
180 185 190  
Cys Ser Ser Leu Thr Gly Leu Phe Met Glu Leu Gly Pro Ser Ser Ile  
195 200 205  
Asn Lys Lys Ile Gln Pro Val Tyr Asn Asp Tyr Ala Trp Asn Ser Asn  
210 215 220  
Ala Ser Val Ile Phe Leu Asp Gln Pro Val Asn Val Gly Tyr Ser Tyr  
225 230 235 240  
Ser Asn Ser Ala Val Ser Asp Thr Val Ala Ala Gly Lys Asp Val Tyr  
245 250 255  
Ala Leu Leu Thr Leu Phe Phe Lys Gln Phe Pro Glu Tyr Ala Lys Gln  
260 265 270  
Asp Phe His Ile Ala Gly Glu Ser Tyr Ala Gly His Tyr Ile Pro Val  
275 280 285  
Phe Ala Ser Glu Ile Leu Ser His Lys Lys Arg Asn Ile Asn Leu Gln  
290 295 300  
Ser Val Leu Ile Gly Asn Gly Leu Thr Asp Gly Leu Thr Gln Tyr Glu  
305 310 315 320

Tyr Tyr Arg Pro Met Ala Cys Gly Asp Gly Gly Tyr Pro Ala Val Leu  
 325 330 335  
 Asp Glu Gly Ser Cys Gln Ala Met Asp Asn Ala Leu Pro Arg Cys Gln  
 340 345 350  
 Ser Met Ile Glu Ser Cys Tyr Ser Ser Glu Ser Ala Trp Val Cys Val  
 355 360 365  
 Pro Ala Ser Ile Tyr Cys Asn Asn Ala Leu Leu Ala Pro Tyr Gln Arg  
 370 375 380  
 Thr Gly Gln Asn Val Tyr Asp Val Arg Gly Lys Cys Glu Asp Ser Ser  
 385 390 395 400  
 Asn Leu Cys Tyr Ser Ala Met Gly Tyr Val Ser Asp Tyr Leu Asn Lys  
 405 410 415  
 Thr Glu Val Ile Glu Ala Val Gly Ala Glu Val Asn Gly Tyr Asp Ser  
 420 425 430  
 Cys Asn Phe Asp Ile Asn Arg Asn Phe Leu Phe His Gly Asp Trp Met  
 435 440 445  
 Lys Pro Tyr His Arg Leu Val Pro Gly Leu Leu Glu Gln Ile Pro Val  
 450 455 460  
 Leu Ile Tyr Ala Gly Asp Ala Asp Phe Ile Cys Asn Trp Leu Gly Asn  
 465 470 475 480  
 Lys Ala Trp Thr Glu Ala Leu Glu Trp Pro Gly Gln Ala Glu Tyr Ala  
 485 490 495  
 Ser Ala Lys Leu Glu Asp Leu Val Val Val Glu Asn Glu His Lys Gly  
 500 505 510  
 Lys Lys Ile Gly Gln Val Lys Ser His Gly Asn Phe Thr Phe Met Arg  
 515 520 525  
 Leu Tyr Gly Gly Gly His Met Val Pro Met Asp Gln Pro Glu Ser Ser  
 530 535 540  
 Leu Glu Phe Phe Asn Arg Trp Leu Gly Gly Glu Trp Phe  
 545 550 555

# 说 明 书 附 图

10	20	30	40	50	60
TCCTCTGCCT ACTCATCCCA TCACCATCTC AATTCATACC GCCCCGTGG GGTTCAGCA CCA					
69	78	87	96	105	114
> <u>ATG</u> <u>AGA</u> <u>GTC</u> <u>CTT</u> <u>CCA</u> <u>GCT</u> <u>GCT</u> <u>ATG</u> <u>CTG</u> <u>GTT</u> <u>GGA</u> <u>GCG</u> <u>GCC</u> <u>ACG</u> <u>GCG</u> <u>GCC</u> <u>GTT</u> <u>CCT</u> MET Arg Val Leu Pro Ala Ala MET Leu Val Gly Ala Ala Thr Ala Ala Val Pro					
123	132	141	150	159	168
<u>CCC</u> <u>TTC</u> <u>CAG</u> <u>CAG</u> <u>GTC</u> <u>CTT</u> <u>GGA</u> <u>GGT</u> <u>AAC</u> <u>GGT</u> <u>GCC</u> <u>AAG</u> <u>CAC</u> <u>GGT</u> <u>GCC</u> <u>GAC</u> <u>CAT</u> <u>GCG</u> Pro Phe Gln Gln Val Leu Gly Gly Asn Gly Ala Lys His Gly Ala Asp His Ala					
177	186	195	204	213	222
<u>GCC</u> <u>GAG</u> <u>GTC</u> <u>CCT</u> <u>GCG</u> <u>GAT</u> <u>CAC</u> <u>AGT</u> <u>GCC</u> <u>GAC</u> <u>GGG</u> <u>TTC</u> <u>TCC</u> <u>AAG</u> <u>CCG</u> <u>CTG</u> <u>CAC</u> <u>GCA</u> Ala Glu Val Pro Ala Asp His Ser Ala Asp Gly Phe Ser Lys Pro Leu His Ala					
231	240	249	258	267	276
<u>TTC</u> <u>CAG</u> <u>GAG</u> <u>GAG</u> <u>CTG</u> <u>AAG</u> <u>TCT</u> <u>CTC</u> <u>TCT</u> <u>GAC</u> <u>GAG</u> <u>GCT</u> <u>CGT</u> <u>AAG</u> <u>CTT</u> <u>TGG</u> <u>GAT</u> <u>GAG</u> Phe Gln Glu Glu Leu Lys Ser Leu Ser Asp Glu Ala Arg Lys Leu Trp Asp Glu					
285	294	303	312	321	330
<u>GTC</u> <u>GCC</u> <u>AGC</u> <u>TTC</u> <u>TTC</u> <u>CCG</u> <u>GAG</u> <u>AGC</u> <u>ATG</u> <u>GAT</u> <u>CAG</u> <u>AAC</u> <u>CCT</u> <u>CTC</u> <u>TTT</u> <u>TCC</u> <u>CTC</u> <u>CCC</u> Val Ala Ser Phe Phe Pro Glu Ser MET Asp Gln Asn Pro Leu Phe Ser Leu Pro					
339	348	357	366	375	384
<u>AAG</u> <u>AAG</u> <u>CAC</u> <u>AAC</u> <u>CGC</u> <u>CGT</u> <u>CCC</u> <u>GAC</u> <u>TCG</u> <u>CAC</u> <u>TGG</u> <u>GAC</u> <u>CAC</u> <u>ATC</u> <u>GTC</u> <u>CGC</u> <u>GGC</u> <u>TCC</u> Lys Lys His Asn Arg Arg Pro Asp Ser His Trp Asp His Ile Val Arg Gly Ser					
393	402	411	420	429	438
<u>GAC</u> <u>GTT</u> <u>CAG</u> <u>AGC</u> <u>GTC</u> <u>TGG</u> <u>GTC</u> <u>ACT</u> <u>GGT</u> <u>GAG</u> <u>AAC</u> <u>GGT</u> <u>GAG</u> <u>AAG</u> <u>GAG</u> <u>CGC</u> <u>GAG</u> <u>GTC</u> Asp Val Gln Ser Val Trp Val Thr Gly Glu Asn Gly Glu Lys Glu Arg Glu Val					

图 1A



447	456	465	474	483	492												
<u>GAT</u> <u>GGC</u> <u>AAG</u> <u>CTG</u> <u>GAA</u> <u>GCC</u> <u>TAT</u> <u>GAT</u> <u>CTC</u> <u>AGG</u> <u>GTC</u> <u>AAG</u> <u>AAG</u> <u>ACC</u> <u>GAT</u> <u>CCT</u> <u>GGC</u> <u>TCT</u>																	
Asp	Gly	Lys	Leu	Glu	Ala	Tyr	Asp	Leu	Arg	Val	Lys	Lys	Thr	Asp	Pro	Gly	Ser
501	510	519	528	537	546												
<u>CTT</u> <u>GGC</u> <u>ATC</u> <u>GAC</u> <u>CCC</u> <u>GGC</u> <u>GTG</u> <u>AAG</u> <u>CAG</u> <u>TAC</u> <u>ACC</u> <u>GGT</u> <u>TAT</u> <u>CTC</u> <u>GAT</u> <u>GAC</u> <u>AAC</u> <u>GAG</u>																	
Leu	Gly	Ile	Asp	Pro	Gly	Val	Lys	Gln	Tyr	Thr	Gly	Tyr	Leu	Asp	Asp	Asn	Glu
555	564		581	591	601												
<u>AAT</u> <u>GAT</u> <u>AAG</u> <u>CAT</u> <u>TTG</u> <u>TTC</u> <u>TAC</u> <u>T</u> <u>GTAAGCACAC</u> <u>CTTGGTTCAA</u> <u>GATCACGCTT</u>																	
Asn	Asp	Lys	His	Leu	Phe	Tyr	Trp										
611	621	631	641	650	659												
<u>TTTATATGCT</u> <u>CTGGATATCT</u> <u>AACGCAACTT</u> <u>AG</u> <u>GG</u> <u>TTC</u> <u>TTC</u> <u>GAG</u> <u>TCT</u> <u>CGC</u> <u>AAT</u> <u>GAC</u> <u>CCC</u>																	
			Phe	Phe	Glu	Ser	Arg	Asn	Asp	Pro							
668	677	686	695	704	713												
<u>GAG</u> <u>AAT</u> <u>GAT</u> <u>CCC</u> <u>GTT</u> <u>GTT</u> <u>CTG</u> <u>TGG</u> <u>CTG</u> <u>AAC</u> <u>GGT</u> <u>GGC</u> <u>CCT</u> <u>GGG</u> <u>TGC</u> <u>TCT</u> <u>TCC</u> <u>CTC</u>																	
Glu	Asn	Asp	Pro	Val	Val	Leu	Trp	Leu	Asn	Gly	Gly	Pro	Gly	Cys	Ser	Ser	Leu
722	731	740	749	758	767												
<u>ACC</u> <u>GGT</u> <u>CTC</u> <u>TTC</u> <u>ATG</u> <u>GAG</u> <u>CTT</u> <u>GGC</u> <u>CCT</u> <u>AGC</u> <u>AGC</u> <u>ATC</u> <u>AAC</u> <u>AAG</u> <u>AAG</u> <u>ATC</u> <u>CAG</u> <u>CCG</u>																	
Thr	Gly	Leu	Phe	MET	Glu	Leu	Gly	Pro	Ser	Ser	Ile	Asn	Lys	Lys	Ile	Gln	Pro
776	785	794	803	812	821												
<u>GTC</u> <u>TAC</u> <u>AAT</u> <u>GAC</u> <u>TAC</u> <u>GCT</u> <u>TGG</u> <u>AAC</u> <u>TCC</u> <u>AAC</u> <u>GCG</u> <u>TCC</u> <u>GTG</u> <u>ATC</u> <u>TTC</u> <u>CTT</u> <u>GAC</u> <u>CAG</u>																	
Val	Tyr	Asn	Asp	Tyr	Ala	Trp	Asn	Ser	Asn	Ala	Ser	Val	Ile	Phe	Leu	Asp	Gln
830	839	848	857	866	875												
<u>CCT</u> <u>GTC</u> <u>AAT</u> <u>GTC</u> <u>GGT</u> <u>TAC</u> <u>TCC</u> <u>TAC</u> <u>AGT</u> <u>AAC</u> <u>TCT</u> <u>GCT</u> <u>GTC</u> <u>AGC</u> <u>GAC</u> <u>ACG</u> <u>GTC</u> <u>GCT</u>																	
Pro	Val	Asn	Val	Gly	Tyr	Ser	Tyr	Ser	Asn	Ser	Ala	Val	Ser	Asp	Thr	Val	Ala

图 1B

884	893	902	911	920	929												
$\overline{\text{GCT}}$ $\overline{\text{GGC}}$ $\overline{\text{AAG}}$ $\overline{\text{GAC}}$ $\overline{\text{GTC}}$ $\overline{\text{TAT}}$ $\overline{\text{GCC}}$ $\overline{\text{TTG}}$ $\overline{\text{CTT}}$ $\overline{\text{ACC}}$ $\overline{\text{CTC}}$ $\overline{\text{TTC}}$ $\overline{\text{TTC}}$ $\overline{\text{AAA}}$ $\overline{\text{CAA}}$ $\overline{\text{TTC}}$ $\overline{\text{CCC}}$ $\overline{\text{GAG}}$																	
Ala	Gly	Lys	Asp	Val	Tyr	Ala	Leu	Leu	Thr	Leu	Phe	Phe	Lys	Gln	Phe	Pro	Glu
938	947	956	965	974	983												
$\overline{\text{TAT}}$ $\overline{\text{GCT}}$ $\overline{\text{AAG}}$ $\overline{\text{CAG}}$ $\overline{\text{GAC}}$ $\overline{\text{TTC}}$ $\overline{\text{CAC}}$ $\overline{\text{ATT}}$ $\overline{\text{GCC}}$ $\overline{\text{GGT}}$ $\overline{\text{GAA}}$ $\overline{\text{TCT}}$ $\overline{\text{TAT}}$ $\overline{\text{GCT}}$ $\overline{\text{GGT}}$ $\overline{\text{CAC}}$ $\overline{\text{TAT}}$ $\overline{\text{ATC}}$																	
Tyr	Ala	Lys	Gln	Asp	Phe	His	Ile	Ala	Gly	Glu	Ser	Tyr	Ala	Gly	His	Tyr	Ile
992	1001	1010	1019	1028	1037												
$\overline{\text{CCC}}$ $\overline{\text{GTC}}$ $\overline{\text{TTC}}$ $\overline{\text{GCT}}$ $\overline{\text{TCG}}$ $\overline{\text{GAG}}$ $\overline{\text{ATC}}$ $\overline{\text{CTG}}$ $\overline{\text{TCT}}$ $\overline{\text{CAC}}$ $\overline{\text{AAG}}$ $\overline{\text{AAG}}$ $\overline{\text{CGC}}$ $\overline{\text{AAC}}$ $\overline{\text{ATC}}$ $\overline{\text{AAC}}$ $\overline{\text{CTG}}$ $\overline{\text{CAG}}$																	
Pro	Val	Phe	Ala	Ser	Glu	Ile	Leu	Ser	His	Lys	Lys	Arg	Asn	Ile	Asn	Leu	Gln
1046	1055	1064	1073	1082	1091												
$\overline{\text{TCC}}$ $\overline{\text{GTT}}$ $\overline{\text{CTC}}$ $\overline{\text{ATT}}$ $\overline{\text{GGC}}$ $\overline{\text{AAC}}$ $\overline{\text{GGT}}$ $\overline{\text{CTC}}$ $\overline{\text{ACC}}$ $\overline{\text{GAC}}$ $\overline{\text{GGA}}$ $\overline{\text{TAC}}$ $\overline{\text{ACC}}$ $\overline{\text{CAG}}$ $\overline{\text{TAC}}$ $\overline{\text{GAG}}$ $\overline{\text{TAC}}$ $\overline{\text{TAC}}$																	
Ser	Val	Leu	Ile	Gly	Asn	Gly	Leu	Thr	Asp	Gly	Tyr	Thr	Gln	Tyr	Glu	Tyr	Tyr
1100	1109	1118	1127	1136	1145												
$\overline{\text{CGT}}$ $\overline{\text{CCC}}$ $\overline{\text{ATG}}$ $\overline{\text{GCC}}$ $\overline{\text{TGC}}$ $\overline{\text{GGT}}$ $\overline{\text{GAC}}$ $\overline{\text{GGC}}$ $\overline{\text{GGT}}$ $\overline{\text{TAC}}$ $\overline{\text{CCA}}$ $\overline{\text{GCT}}$ $\overline{\text{GTC}}$ $\overline{\text{TTG}}$ $\overline{\text{GAC}}$ $\overline{\text{GAG}}$ $\overline{\text{AGC}}$ $\overline{\text{TCC}}$																	
Arg	Pro	MET	Ala	Cys	Gly	Asp	Gly	Gly	Tyr	Pro	Ala	Val	Leu	Asp	Glu	Ser	Ser
1154	1163	1172	1181	1190	1199												
$\overline{\text{TGC}}$ $\overline{\text{CAG}}$ $\overline{\text{TCC}}$ $\overline{\text{ATG}}$ $\overline{\text{GAC}}$ $\overline{\text{AAC}}$ $\overline{\text{GCT}}$ $\overline{\text{CTT}}$ $\overline{\text{CCT}}$ $\overline{\text{CGC}}$ $\overline{\text{TGC}}$ $\overline{\text{CAG}}$ $\overline{\text{TCT}}$ $\overline{\text{ATG}}$ $\overline{\text{ATT}}$ $\overline{\text{GAG}}$ $\overline{\text{TCT}}$ $\overline{\text{TGC}}$																	
Cys	Gln	Ser	MET	Asp	Asn	Ala	Leu	Pro	Arg	Cys	Gln	Ser	MET	Ile	Glu	Ser	Cys
1208	1217	1226	1235	1244	1253												
$\overline{\text{TAC}}$ $\overline{\text{AGT}}$ $\overline{\text{TCC}}$ $\overline{\text{GAG}}$ $\overline{\text{AGC}}$ $\overline{\text{GCT}}$ $\overline{\text{TGG}}$ $\overline{\text{GTT}}$ $\overline{\text{TGT}}$ $\overline{\text{GTC}}$ $\overline{\text{CCG}}$ $\overline{\text{GCC}}$ $\overline{\text{TCC}}$ $\overline{\text{ATC}}$ $\overline{\text{TAC}}$ $\overline{\text{TGT}}$ $\overline{\text{AAC}}$ $\overline{\text{AAC}}$																	
Tyr	Ser	Ser	Glu	Ser	Ala	Trp	Val	Cys	Val	Pro	Ala	Ser	Ile	Tyr	Cys	Asn	Asn
1262	1271	1280	1289	1298	1307												
$\overline{\text{GCC}}$ $\overline{\text{CTC}}$ $\overline{\text{CTT}}$ $\overline{\text{GCC}}$ $\overline{\text{CCT}}$ $\overline{\text{TAC}}$ $\overline{\text{CAG}}$ $\overline{\text{CGC}}$ $\overline{\text{ACT}}$ $\overline{\text{GGG}}$ $\overline{\text{CAG}}$ $\overline{\text{AAC}}$ $\overline{\text{GTC}}$ $\overline{\text{TAT}}$ $\overline{\text{GAT}}$ $\overline{\text{GTC}}$ $\overline{\text{CGT}}$ $\overline{\text{GGT}}$																	
Ala	Leu	Leu	Ala	Pro	Tyr	Gln	Arg	Thr	Gly	Gln	Asn	Val	Tyr	Asp	Val	Arg	Gly

图 1C

1316	1325	1334	1343	1352	1361
AAG TGC GAG GAT AGC TCT AAC CTT TGC TAC TCG GCT ATG GGC TAC GTC AGC GAC					
Lys Cys Glu Asp Ser Ser Asn Leu Cys Tyr Ser Ala MET Gly Tyr Val Ser Asp					
1370	1379	1388	1397	1406	1415
TAC CTG AAC AAG CCC GAA GTC ATC GAG GCT GTT GGC GCT GAG GTC AAC GGC TAC					
Tyr Leu Asn Lys Pro Glu Val Ile Glu Ala Val Gly Ala Glu Val Asn Gly Tyr					
1424	1433	1442	1451	1460	1469
GAC TCG TGC AAC TTT GAC ATC AAC CGC AAC TTC CTC TTC CAC GGT GAC TGG ATG					
Asp Ser Cys Asn Phe Asp Ile Asn Arg Asn Phe Leu Phe His Gly Asp Trp MET					
1478	1487	1496	1505	1514	1523
AAG CCC TAC CAC CGC CTC GTT CCG GGA CTC CTG GAG CAG ATC CCT GTC TTG ATC					
Lys Pro Tyr His Arg Leu Val Pro Gly leu Leu Glu Gln Ile Pro Val Leu Ile					
1532	1541	1550	1559	1568	1577
TAT GCC GGT GAT GCT GAT TTC ATT TGC AAC TGG CTG GGC AAC AAG GCC TGG ACT					
Tyr Ala Gly Asp Ala Asp Phe Ile Cys Asn Trp Leu Gly Asn Lys Ala Trp Thr					
1586	1595	1604	1613	1622	1631
GAA GCC CTG GAG TGG CCC GGA CAG GCT GAA TAT GCC TCC GCT GAG CTG GAG GAT					
Glu Ala Leu Glu Trp Pro Gly Gln Ala Glu Tyr Ala Ser Ala Glu Leu Glu Asp					
1640	1649	1658	1667	1676	1685
CTG GTC ATT GTC GAC AAT GAG CAC ACG GGC AAG AAG ATT GGC CAG GTT AAG TCC					
Leu Val Ile Val Asp Asn Glu His Thr Gly Lys Lys Ile Gly Gln Val Lys Ser					
1694	1703	1712	1721	1730	1739
CAT GGC AAC TTC ACC TTC ATG CGT CTC TAT GGT GGT GGC CAC ATG GTC CCG ATG					
His Gly Asn Phe Thr Phe MET Arg Leu Tyr Gly Gly Gly His MET Val Pro MET					

图 1D

1748	1757	1766	1775	1784	1793												
<u>GAC</u>	<u>CAG</u>	<u>CCC</u>	<u>GAG</u>	<u>TCG</u>	<u>AGT</u>	<u>CTC</u>	<u>GAG</u>	<u>TTC</u>	<u>TTC</u>	<u>AAC</u>	<u>CGC</u>	<u>TGG</u>	<u>TTG</u>	<u>GGA</u>	<u>GGT</u>	<u>GAA</u>	<u>TGG</u>
Asp	Gln	Pro	Glu	Ser	Ser	Leu	Glu	Phe	Phe	Asn	Arg	Trp	Leu	Gly	Gly	Glu	Trp
<u>TTC</u>	<u>TAA</u>	1809	1819	1829	1839	1849											
		AGACGTGCTA	CCACCGCATA	TAGACTTTCT	GGTCATTCG	GTGACTGTC											
1859	1869	1879	1889	1899	1909												
AGATATGTTT	CTTAACGATA	GTTTGAGCAT	GCTTGTCAT	GCCCACTAGT	CCCGATCCTT												
1919	1929	1939	1949	1959	1969												
ATATGTTGCA	TGGTATCTAT	GAGTTTTGTC	ACTATAGTGC	ATTATACATG	TGTACTTCGT												
1979	1989	1999	2009	2019	2029												
ATGAGAATGA	ATCGATCGCA	TTTACACGCA	TATAAATAGT	ACCCACCTCC	GCCTGGACAT												
2039	2049	2059	2068														
GAATTAGGCC	CGGCCAGTCG	TTTACATACA	GTGCTAGAA														

图 1E

```

10          20          30          40          50          60
GCGGCCGCTG CTACTTGCTT TTTCTAATTT GATACTTTTG TGCCGTACC GTACCTCCA

70          80          90          100         110         120
GACCGCAAGG TACCCATCCT CTACCTACTC ATCCCATCAT CATCTCGATT TCATACCAAC

                                148          157          166          175
                                >
CCCGTTGGGT TTCAACACA ATG AGA GTT CTT CCA GCT GCT ATG CTG GTT GGA GCG GGC
                                MET Arg Val Leu Pro Ala Ala MET Leu Val Gly Ala Gly
START OF PROPEPTIDE
    1 ↓ 193                202          211          220          229

ACT GCG GCC GTC CCT CCC TTC CAG CAG GTC CTT GGA GGT AAC GGT GCC AAG CAC
Thr Ala Ala Val Pro Pro Phe Gln Gln Val Leu Gly Gly Asn Gly Ala Lys His
    238          247          256          265          274          283

GGT GCC GAC CAT GCG GCC GAG GTC CCT GCG GAT CAC AGT GCC GAC GGG TTC TCC
Gly Ala Asp His Ala Ala Glu Val Pro Ala Asp His Ser Ala Asp Gly Phe Ser
    292          301          310          319          328          337

AAG CCG CTG CAC GCA TTC CAG GAG GAG CTG AAG TCT CTC TCT GAT GAG GCT CGT
Lys Pro Leu His Ala Phe Gln Glu Glu Leu Lys Ser Leu Ser Asp Glu Ala Arg
    346          355          364          373          382          391

AAG CTC TGG GAT GAG GTT GCT AGC TTC TTC CCG GAG AGC ATG GAT CAG AAC CCT
Lys Leu Trp Asp Glu Val Ala Ser Phe Phe Pro Glu Ser MET Asp Gln Asn Pro
    400          409          418          427          436          445

CTC TTC TCC CTC CCC AAG AAG CAC AAC CGC CGC CCC GAC CAC CAC TGG GAC CAC
Leu Phe Ser Leu Pro Lys Lys His Asn Arg Arg Pro Asp His His Trp Asp His
    454          463          472          481          490          499

ATC GTC CGC GGC TCC GAC GTT CAG AGC GTC TGG GTT ACT GGT GAG AAC GGT GAG
Ile Val Arg Gly Ser Asp Val Gln Ser Val Trp Val Thr Gly Glu Asn Gly Glu

```

图 2A

预测的成熟CPY的N-末端

508	517	526	535	544	553
AAG GAG CGT GAG GTC GAT GGC AAG CTG GAA GCC TAT GAT CTC AGG GTC AAG AAG					
Lys Glu Arg Glu Val Asp Gly Lys Leu Glu Ala Tyr Asp Leu Arg Val Lys Lys					
562	571	580	589	598	607
ACC GAT CCT AGC TCT CTT GGC ATC GAC CCT GGC GTA AAG CAG TAC ACC GGT TAT					
Thr Asp Pro Ser Ser Leu Gly Ile Asp Pro Gly Val Lys Gln Tyr Thr Gly Tyr					
616	625	634	643	652	661
CTC GAT GAC AAC GAG AAC GAC AAG CAT CTG TTC TAC TGG TTC TTC GAG TCT CGC					
Leu Asp Asp Asn Glu Asn Asp Lys His Leu Phe Tyr Trp Phe Phe Glu Ser Arg					
670	679	688	697	706	715
AAT GAC CCC GAG AAT GAC CCT GTT GTT CTG TGG CTG AAC GGT GGC CCT GGA TGC					
Asn Asp Pro Glu Asn Asp Pro Val Val Leu Trp Leu Asn Gly Gly Pro Gly Cys					
724	733	742	751	760	769
TCT TCC CTC ACC GGT CTT TTC ATG GAG CTC GGC CCT AGC AGC ATC AAC AAG AAG					
Ser Ser Leu Thr Gly Leu Phe MET Glu Leu Gly Pro Ser Ser Ile Asn Lys Lys					
778	787	796	805	814	823
ATC CAG CCG GTC TAC AAC GAC TAC GCT TGG AAC TCC AAC GCG TCC GTG ATC TTC					
Ile Gln Pro Val Tyr Asn Asp Tyr Ala Trp Asn Ser Asn Ala Ser Val Ile Phe					
832	841	850	859	868	877
CTT GAC CAG CCT GTC AAC GTC GGT TAC TCT TAC AGC AAC TCT GCT GTC AGC GAC					
Leu Asp Gln Pro Val Asn Val Gly Tyr Ser Tyr Ser Asn Ser Ala Val Ser Asp					
886	895	904	913	922	931
ACC GTT GCT GCT GGC AAG GAC GTC TAT GCC TTG CTT ACC CTC TTC TTC AAA CAA					
Thr Val Ala Ala Gly Lys Asp Val Tyr Ala Leu Leu Thr Leu Phe Phe Lys Gln					

图 2B

940	949	958	967	976	985
TTC CCC GAG TAT GCC AAG CAG GAC TTC CAC ATT GCC GGT GAA TCC TAT GCT GGT					
Phe Pro Glu Tyr Ala Lys Gln Asp Phe His Ile Ala Gly Glu Ser Tyr Ala Gly					
994	1003	1012	1021	1030	1039
CAC TAT ATC CCC GTC TTT GCT TCG GAG ATT TTG TCT CAC AAG AAG CGC AAC ATC					
His Tyr Ile Pro Val Phe Ala Ser Glu Ile Leu Ser His Lys Lys Arg Asn Ile					
1048	1057	1066	1075	1084	1093
AAC CTG CAG TCC GTT CTT ATT GGC AAC GGT CTC ACC GAC GGT CTC ACT GAG TAC					
Asn Leu Gln Ser Val Leu Ile Gly Asn Gly Leu Thr Asp Gly Leu Thr Gln Tyr					
1102	1111	1120	1129	1138	1147
GAG TAC TAC CGT CCC ATG GCC TGT GGT GAC GGT GGT TAC CCA GCT GTC TTG GAC					
Glu Tyr Tyr Arg Pro MET Ala Cys Gly Asp Gly Gly Tyr Pro Ala Val Leu Asp					
1156	1165	1174	1183	1192	1201
GAG GGC TCC TGC CAG GCC ATG GAC AAC GCC CTT CCT CGC TGC CAG TCT ATG ATT					
Glu Gly Ser Cys Gln Ala MET Asp Asn Ala Leu Pro Arg Cys Gln Ser MET Ile					
1210	1219	1228	1237	1246	1255
GAG TCT TGC TAT AGT TCC GAG AGC GCT TGG GTT TGT GTC CCG GCC TCC ATC TAC					
Glu Ser Cys Tyr Ser Ser Glu Ser Ala Trp Val Cys Val Pro Ala Ser Ile Tyr					
1318	1327	1336	1345	1354	1363
GTT CGT GGT AAG TGC GAG GAT AGC TCC AAC CTC TGC TAC TCG GCC ATG GGC TAC					
Val Arg Gly Lys Cys Glu Asp Ser Ser Asn Leu Cys Tyr Ser Ala MET Gly Tyr					
1372	1381	1390	1399	1408	1417
GTC AGC GAC TAC CTG AAC AAG ACC GAG GTC ATT GAG GCT GTT GGC GCT GAG GTC					
Val Ser Asp Tyr Leu Asn Lys Thr Glu Val Ile Glu Ala Val Gly Ala Glu Val					

图 2C

1480	1489	1498	1507	1516	1525												
<u>GAC</u> <u>TGG</u> <u>ATG</u> <u>AAG</u> <u>CCC</u> <u>TAC</u> <u>CAC</u> <u>CGT</u> <u>CTC</u> <u>GTT</u> <u>CCG</u> <u>GGA</u> <u>CTC</u> <u>CTG</u> <u>GAG</u> <u>CAG</u> <u>ATC</u> <u>CCT</u>																	
Asp	Trp	MET	Lys	Pro	Tyr	His	Arg	Leu	Val	Pro	Gly	Leu	Leu	Glu	Gln	Ile	Pro
1534	1543	1552	1561	1570	1579												
<u>GTC</u> <u>CTG</u> <u>ATC</u> <u>TAC</u> <u>GCT</u> <u>GGT</u> <u>GAC</u> <u>GCC</u> <u>GAT</u> <u>TTC</u> <u>ATC</u> <u>TGC</u> <u>AAC</u> <u>TGG</u> <u>CTG</u> <u>GGC</u> <u>AAC</u> <u>AAG</u>																	
Val	Leu	Ile	Tyr	Ala	Gly	Asp	Ala	Asp	Phe	Ile	Cys	Asn	Trp	Leu	Gly	Asn	Lys
1588	1597	1606	1615	1624	1633												
<u>GCC</u> <u>TGG</u> <u>ACT</u> <u>GAA</u> <u>GCC</u> <u>CTT</u> <u>GAG</u> <u>TGG</u> <u>CCC</u> <u>GGA</u> <u>CAG</u> <u>GCT</u> <u>GAA</u> <u>TAT</u> <u>GCC</u> <u>TCC</u> <u>GCT</u> <u>AAG</u>																	
Ala	Trp	Thr	Glu	Ala	Leu	Glu	Trp	Pro	Gly	Gln	Ala	Glu	Tyr	Ala	Ser	Ala	Lys
1642	1651	1660	1669	1678	1687												
<u>CTG</u> <u>GAG</u> <u>GAC</u> <u>CTG</u> <u>GTC</u> <u>GTG</u> <u>GTC</u> <u>GAG</u> <u>AAT</u> <u>GAG</u> <u>CAC</u> <u>AAG</u> <u>GGC</u> <u>AAG</u> <u>AAG</u> <u>ATC</u> <u>GGC</u> <u>CAG</u>																	
Leu	Glu	Asp	Leu	Val	Val	Val	Glu	Asn	Glu	His	Lys	Gly	Lys	Lys	Ile	Gly	Gln
1696	1705	1714	1723	1732	1741												
<u>GTC</u> <u>AAG</u> <u>TCC</u> <u>CAT</u> <u>GGC</u> <u>AAC</u> <u>TTC</u> <u>ACC</u> <u>TTC</u> <u>ATG</u> <u>CGT</u> <u>CTC</u> <u>TAT</u> <u>GGC</u> <u>GGT</u> <u>GGC</u> <u>CAC</u> <u>ATG</u>																	
Val	Lys	Ser	His	Gly	Asn	Phe	Thr	Phe	MET	Arg	Leu	Tyr	Gly	Gly	Gly	His	MET
1750	1759	1768	1777	1786	1795												
<u>GTC</u> <u>CCG</u> <u>ATG</u> <u>GAC</u> <u>CAA</u> <u>CCC</u> <u>GAG</u> <u>TCG</u> <u>AGT</u> <u>CTT</u> <u>GAA</u> <u>TTC</u> <u>TTC</u> <u>AAC</u> <u>CGC</u> <u>TGG</u> <u>TTG</u> <u>GGA</u>																	
Val	Pro	MET	Asp	Gln	Pro	Glu	Ser	Ser	Leu	Glu	Phe	Phe	Asn	Arg	Trp	Leu	Gly
1804	1813	1823	1833	1843	1853												
<u>GGT</u> <u>GAA</u> <u>TGG</u> <u>TTT</u> <u>TAA</u> <sup>&gt;</sup> <u>AGACGTGCTA</u> <u>TCACCGCATA</u> <u>TAGACTTTCC</u> <u>GGTCATTTCC</u>																	
1863	1873	1883	1893	1903	1913												
<u>GTGACACTGC</u> <u>AGATATGTTT</u> <u>CTTAACGATA</u> <u>GTTTGAGGAT</u> <u>GCTTGTC AAT</u> <u>GCCCACTAAT</u>																	

图 2D



1923            1933            1943            2953            1963            1973  
 CCCGAGCCTT ATGTTACATG GTATCTATGA GTTTGTCATT ATAGTGCATT ATGCATTTGT  
 1983            1993            2002  
 ACTCCGTACG AGAATGAATC AGCGGCCGC

图 2E

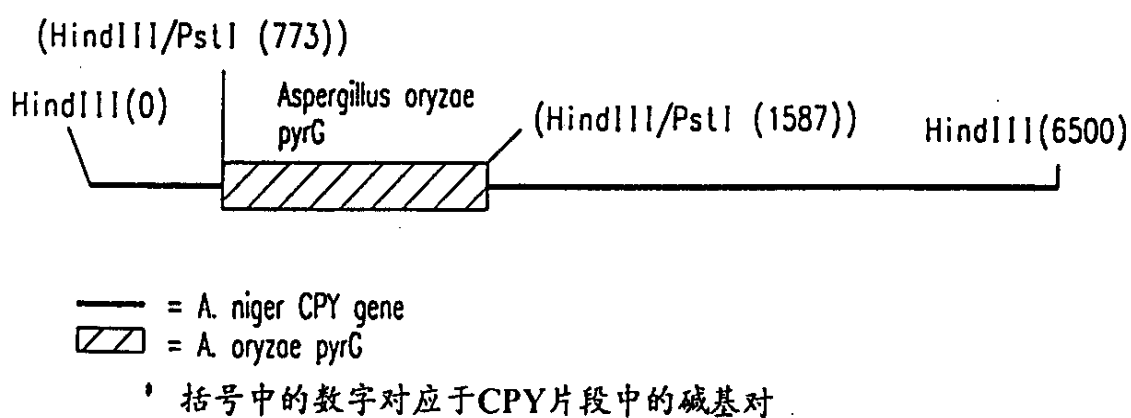


图 3