

[19] 中华人民共和国国家知识产权局

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

C12N 1/15

# [12] 发明专利申请公开说明书

C12N 15/10 C12P 9/00  
//(C12N1/15 C12R1: 66)

[21] 申请号 00804283.7

(C12N1/15, C12R1:  
77)

[43] 公开日 2002 年 3 月 20 日

[11] 公开号 CN 1341146A

[22] 申请日 2000.2.17 [21] 申请号 00804283.7

[30] 优先权

[32] 1999.2.24 [33] DK [31] PA199900253

[86] 国际申请 PCT/DK00/00063 2000.2.17

[87] 国际公布 WO00/50567 英 2000.8.31

[85] 进入国家阶段日期 2001.8.24

[71] 申请人 诺维信公司

地址 丹麦鲍斯韦

[72] 发明人 托本·V·博彻特

拉斯·克里斯琴森

[74] 专利代理机构 北京市柳沈律师事务所

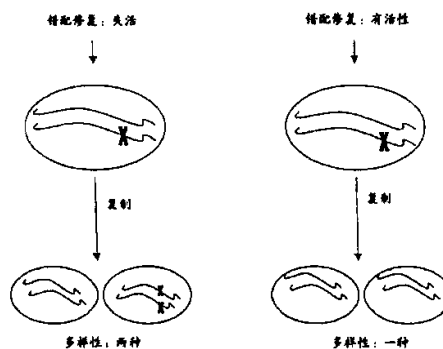
代理人 巫肖南

权利要求书 16 页 说明书 20 页 附图页数 8 页

[54] 发明名称 具有灭活的 DNA 错配修复系统的真菌细胞

[57] 摘要

使用参与丝状真菌细胞错配修复系统的新颖的克隆基因制备丝状真菌细胞 DNA 文库的方法。



ISSN 1008-4274

## 权 利 要 求 书

1、丝状真菌细胞，其中参与错配修复系统的基因已经被灭活，该参与错配修复系统的基因包含：

- 5 (a)编码 SEQ ID NO 2 之 683-758 位所示多肽序列的 DNA 序列；或  
(b)编码至少 70%等同于 SEQ ID NO 2 之 683-758 位的多肽序列的 DNA 序列。

2、丝状真菌细胞，其中参与错配修复系统的基因已经被灭活，该参与错配修复系统的基因包含：

- 10 (a)编码 SEQ ID NO 2 之 1-940 位所示多肽序列的 DNA 序列；或  
(b)编码至少 70%等同于 SEQ ID NO 2 之 1-940 位的多肽序列的 DNA 序列。

3、权利要求 1 或 2 的丝状真菌细胞，其中该参与错配修复的基因是有缺陷的。

- 15 4、权利要求 1 或 2 的丝状真菌细胞，其中该参与错配修复的基因已经被暂时灭活。

5、在先权利要求中任一项的丝状真菌细胞，其中该丝状真菌细胞是镰孢属菌株，或者更优选为曲霉属菌株。

- 20 6、制备丝状真菌细胞群体的方法，其中群体中的个体细胞分别包含代表目的 DNA 文库的不同目的 DNA 序列，该方法包括下列步骤：

(a)分别将不同的目的 DNA 序列放置在包含权利要求 1 或 2 的丝状真菌细胞的丝状真菌细胞群体中；

(b)将(a)群体生长一段时间，使群体中单个目的 DNA 序列在权利要求 1 或 2 的错配修复系统基因已被灭活的条件下复制至少一次。

- 25 7、权利要求 6 的方法，其中步骤(b)下的该错配修复系统是有缺陷的。

8、权利要求 6 的方法，其中步骤(b)下的该错配修复系统已经被暂时灭活。

9、权利要求 6-8 中任一项的方法，其中该丝状真菌细胞是镰孢属菌株，或者更优选为曲霉属菌株。

- 30 10、权利要求 6-9 中任一项的方法，其中在权利要求 6 的步骤(b)下，存在部分同源的目的 DNA 序列的体内基因间重组。

11、生产目的多肽的方法，其包括权利要求 6 的步骤，且其中所述目的 DNA 序列编码目的多肽，该方法进一步包括下列步骤：

(c)从权利要求 6 步骤(b)所得丝状真菌细胞群体中选择所需的目的是多肽。

5 12、权利要求 11 的方法，其进一步包括下列步骤：

(d)可选的步骤，包括根据进一步的特定需要修饰所需的所选择的目的多肽的氨基酸序列；

(e)将编码权利要求 11 步骤(c)的目的多肽或步骤(d)的修饰型目的多肽的 DNA 序列放置到适合于大规模制备目的多肽的丝状真菌细胞中；

10 (f)在允许目的多肽表达的条件下，在至少 10000m<sup>3</sup> 的发酵罐中培养步骤(e)的丝状真菌细胞；

(g)分离目的多肽。

15 13、权利要求 12 的方法，其中适合于大规模制备权利要求 12 步骤(e)的目的多肽的丝状真菌细胞是不同于权利要求 6 步骤(a)的丝状真菌细胞的另一种丝状真菌细胞。

14、权利要求 11-13 中任一项的方法，其中所述目的多肽是从丝状真菌细胞衍生的多肽。

15、权利要求 11-14 中任一项的方法，其中所述目的多肽是一种酶，例如淀粉酶、蛋白酶、纤维素酶、脂肪酶、木聚糖酶、磷脂酶。

# 说明书

## 具有灭活的 DNA 错配修复系统的真菌细胞

### 5 发明领域

一种利用新颖的克隆基因制备丝状真菌细胞 DNA 文库的方法，该基因参与丝状真菌细胞的错配修复系统。

### 发明背景

10 错配修复系统是细胞内识别新合成的双链 DNA 序列中的错配的系统。错配修复系统然后可利用甲基化的“旧”株系作为模板纠正被察觉为错误的错配，或者可以介导包含错配的双链 DNA 序列的降解。

独立于精确的机理之外，最终结果将是“错配修复系统”将限制细胞内的“多样性”，多样性表现为包含错配的双链 DNA 序列。

15 例如，包含单一错配的双链 DNA 序列代表细胞内两种不同的 DNA 序列的多样性。如果错配修复系统纠正错配，那么在细胞内将只有一种 DNA 序列。

或者，如果错配修复系统介导这样一种双链 DNA 序列的降解，那么多样性将消失。参见图 1，其以图解方式阐述错配修复系统在细胞内是如何工作的。

20 所以，如果包含错配的双链 DNA 序列代表目的 DNA 文库，那么当转化(放置)到具有活性错配修复系统的细胞内时，该文库的多样性可以受到限制。

本领域提供了一种该问题的解决方案，即制备其中错配系统已失活的

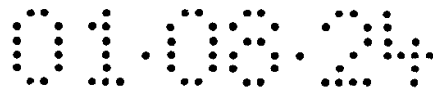
25 细胞。

EP 449923 描述了其中错配系统被灭活的细菌细胞。

WO 97/37011 描述了其中错配系统被灭活的酵母细胞。参见该文献的工作例。

WO 97/05268 描述了其中错配系统被灭活的小鼠细胞。参见该文献的

30 工作例。



## 发明概述

5 本发明所要解决的问题是提供制备丝状真菌细胞 DNA 文库的改进策略。包含这样一种 DNA 文库的丝状真菌细胞群体因而可以用于选择目的多肽。还可以选择具有特定性质的多核苷酸序列，例如具有改变/改进性质的启动子、终止子和其它调节元件。

解决方案的基础是发明人已经克隆了参与丝状真菌细胞错配修复系统的新基因。此外，该基因是第一种被克隆的参与丝状真菌细胞错配修复系统的基因。

10 通过灭活丝状细胞中的这种基因，有可能获得这样的丝状细胞，它在其错配修复系统中有缺陷，并且对制备丝状真菌细胞中的 DNA 文库来说是非常有用的。

该基因包含非常有特性的编码如 SEQ ID NO 2 之 683-758 位所示多肽序列的 DNA 序列。

15 该 DNA 已经用于克隆编码如 SEQ ID NO 2 之 1-940 位所示多肽序列的全长基因。参见本文的工作例(见下)。

如本文工作例所述被克隆的基因是从米曲霉丝状真菌细胞克隆的基因。

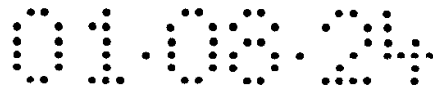
20 不过，基于本文所提供的新序列信息，对本领域技术人员来说从其他丝状真菌细胞克隆相似的同源基因是例行工作，例如使用标准的杂交或 PCR 技术，优选地使用编码 SEQ ID NO 2 之 683-758 位所示多肽序列的 DNA 序列作为制备杂交探针或 PCR 引物的基础。

因此，本发明在第一方面涉及丝状真菌细胞，其中参与错配修复系统的基因已经被灭活，该参与错配修复系统的基因包含：

- 25 (a)编码 SEQ ID NO 2 之 683-758 位所示多肽序列的 DNA 序列；或  
(b)编码至少 70%等同于 SEQ ID NO 2 之 683-758 位的多肽序列的 DNA 序列；

本发明在第二方面涉及丝状真菌细胞，其中参与错配修复系统的基因已经被灭活，该参与错配修复系统的基因包含：

- 30 (a)编码 SEQ ID NO 2 之 1-940 位所示多肽序列的 DNA 序列；或  
(b)编码至少 70%等同于 SEQ ID NO 2 之 1-940 位的多肽序列的 DNA 序列。



如上所述，本发明第一或第二方面的丝状真菌细胞非常适合于制备丝状真菌细胞中目的 DNA 文库。

因此，本发明在第三方面涉及制备丝状真菌细胞群体的方法，其中群体中的个体细胞分别包含代表目的 DNA 文库的不同目的 DNA 序列，该方法包括下列步骤：

(a) 分别将不同的目的 DNA 序列放置在包含本发明第一或第二方面丝状真菌细胞的丝状真菌细胞群体中；

(b) 将(a)群体生长一段时间，使群体中单个目的 DNA 序列在本发明第一或第二方面错配修复系统基因已被灭活的条件下复制至少一次。

图 1 阐述了如第三方面步骤(b)所述使单个错配修复灭活型丝状真菌细胞复制目的 DNA 至少一次的优点之一。图 1 可以看到，如本文所述使用丝状真菌错配修复灭活细胞的第三方面方法使其中最终异型双链 DNA 错配未被纠正的 DNA 文库的制备成为可行。所得 DNA 文库比在不具有灭活的错配修复系统的丝状真菌细胞中制得的 DNA 文库具有更高多样性(参见图 1)。目的 DNA 序列的复制意味着两股都被复制，使得分别产生的两套双链 DNA，每套均基于两条原始股之一。

其中群体中的个体细胞包含上述目的 DNA 文库的丝状真菌细胞群体可以用于选择目的多肽。

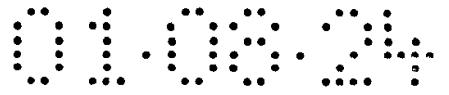
因此，本发明在第四方面涉及生产目的多肽的方法，其包括第三方面的步骤，且其中目的 DNA 序列编码目的多肽，该方法进一步包括下列步骤：

(c) 从第三方面步骤(b)所得丝状真菌细胞群体中选择所需的目的是多肽。

第四方面方法的优点是，目的是多肽是从表达该多肽的丝状真菌细胞中选择的。所以，可直接知道所述多肽可以从丝状真菌细胞表达，如果随后需要在丝状真菌细胞内大规模制备多肽，那么这是有用的。当 DNA 文库编码从丝状真菌细胞衍生的目的是多肽时，这一点是特别值得关注的，因为已知在工业上，丝状真菌多肽优选地是在丝状真菌细胞内制备的，且产量高。

这与类似的选择方法相反，所述类似方法例如使用酵母细胞的方法。这里唯一已知的是所选择的多肽能够在酵母中被表达，以后在丝状真菌细胞中的表达可能产生问题，尤其当需要高产量时。

定义：



下面提供上述本发明各方面技术特征的定义。

术语“基因”此处表示基因(DNA 序列)能够在所述细胞内被表达为多肽。因此,所述基因序列将被定义为开始于起始密码子(通常为“ATG”、“GTG”或“TTG”),结束于终止密码子(通常为“TAA”、“TAG”或“TGA”)的开发读码框。

如本领域已知,为了表达所述基因,必须有与该基因相连的元件,它们是该基因在细胞内表达所必需的。这样的标准元件可以包括启动子、核糖体结合位点、末端序列,还可以是本领域已知的其他元件。

按照本领域,术语“错配修复系统”此处应当被理解为细胞内识别双链 DNA 序列错配的系统。例如参见 WO 97/37011 第 1 页第 21-28 行。错配修复系统要么纠正被发现为错误的错配,例如通过使用甲基化“旧”链作为模板,要么它可以介导包含错配的双链 DNA 序列的降解。

独立于精确的机理之外,最终结果将是“错配修复系统”将限制细胞内的“多样性”,多样性表现为包含错配的双链 DNA 序列。

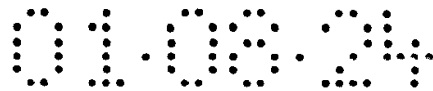
例如,包含单一错配的双链 DNA 序列代表细胞内两种不同的 DNA 序列的多样性。如果错配修复系统纠正该错配,那么在细胞内将只有一种 DNA 序列。或者,如果错配修复系统介导这样一种双链 DNA 序列的降解,那么多样性将消失。

被参与错配修复系统的基因编码的多肽通过与错配目的的机理识别该错配。

因此,检验如本文所述丝状真菌细胞的错配修复系统是否已灭活的测定法采用“凝胶移动(shift)试验”,或称“凝胶推迟试验”。这是本领域所用的一种标准试验。参见 WO 97/05268 第 16、17 和 25 页。

这类试验的原理是细胞提取物从下列制成:(a)丝状真菌细胞,其中如本文所述参与错配修复系统的基因是灭活的;和(b)相应的丝状真菌细胞,其中的基因不是灭活的。这些提取物然后与含有碱基对错配的 G:T、G:A、G:G、A:C 的寡核苷酸和超螺旋 TG 二核苷酸结合/混合,在非变性凝胶上进行测定。

如果凝胶移动(shift)试验证明所含基因未灭活的对照丝状真菌细胞包含与任意上述寡核苷酸结合的任意蛋白质,并且这些结合蛋白在所含如本文所述参与错配修复系统的基因是灭活的丝状真菌细胞中没有见到,那么确



认后者错配修复系统是灭活的。

本文工作例 1 提供了适合的凝胶移动(shift)试验的详细说明。

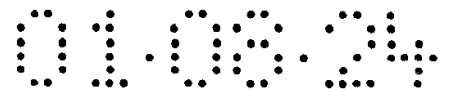
涉及术语“编码至少 70%等同于 SEQ ID NO 2 之 683-758 位所示多肽序列的多肽序列的 DNA 序列”和“编码至少 70%等同于 SEQ ID NO 2 之 1-940 5 位所示多肽序列的多肽序列的 DNA 序列”的序列同一性被确定为两种序列之间的等同程度，其说明第一种序列从第二种衍化而来。同一性可以适当借助本领域已知的计算机程序加以测定，例如 GCG 程序包中的 GAP (Program Manual for the Wisconsin Package, Version 8, August 1994, Genetics Computer Group, 575 Science Drive, Madison, Wisconsin, USA 53711) 10 (Needleman, S. B.和 Wunsch, C. D., (1970)《分子生物学杂志》48, 443453)。GAP 采用下列多肽序列比较设置：GAP 生成罚分 3.0，GAP 延长罚分 0.1，结果，按照发明的第一方面，针对 SEQ ID NO 2 之 683-758 位所示氨基酸序列，或者按照发明的第一方面，针对 SEQ ID NO 2 之 1-940 位所示氨基酸序列，被本发明相似 DNA 序列编码的多肽所表现的等同程度优选为至少 15 70%，更优选为至少 80%，更优选为至少 90%，更优选为至少 95%，尤其为至少 97%。

术语“DNA 文库”此处表示至少两种不同 DNA 序列的文库。出于多数实用目的，文库必须是更大的。因此，DNA 文库优选地包含至少 1000 种不同的 DNA 序列，更优选为至少 10000 种不同的 DNA 序列，进而更优 20 选为至少 100000 种不同的 DNA 序列。

涉及本发明第三方面方法步骤(a)的术语“分别将不同的目的 DNA 序列放置在丝状真菌细胞群体中”此处应当在“它不是等同于放置在丝状真菌细胞群体中的目的 DNA 序列”这一意义上加以广泛理解。在本文的上下文中，涉及使用错配修复缺陷型细胞制备 DNA 文库的方法时，该术语应当 25 优选地表示这样一种情形，其中丝状真菌细胞群体内的细胞包含至少两种不同的目的 DNA 序列，它们是部分同源的，以致它们能够在细胞内彼此杂交/重组。测定这类序列如何部分同源以获得所述细胞内的重组是本领域技术人员的常识。

实例可以是将与染色体 DNA 序列部分同源的单股寡核苷酸序列放置在 30 在细胞内，或者将包含错配(例如包含在载体内)的双链 DNA 序列放置在细胞内。见下关于这些例子的进一步说明。





将这些 DNA 序列放置在丝状细胞内的具体实验方法可以按照多种适当技术中的任一种进行，例如转化技术。

按照本发明第三方面步骤(b)，术语“将(a)群体生长一段时间，使群体中单个目的 DNA 序列在错配修复系统基因已被灭活的条件下复制至少一次”表示，在个体细胞已经自我复制至少一次之后，错配修复系统可以被重新激活，而不丧失该方法的优点。图 1 阐述了其中的技术原因。本例中，将包含单一错配的双链 DNA 序列放置在丝状细胞中。当细胞已经在错配修复系统基因已被灭活的条件下复制一次之后，双链 DNA 内两种不同的单股 DNA 序列分别被复制，得到两种不同的双链序列，每个复制细胞中含有其一，没有任何错配。因此，由于这样的细胞不包含具有错配的目的双链 DNA 序列，因此不存在维持错配修复系统灭活的技术需要。

下面描述本发明优选的实施方案，仅供例证。

#### 附图说明

15 图 1：

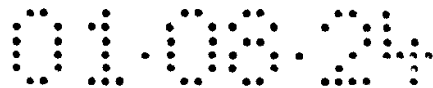
本图阐述将包含单一错配的双链 DNA 序列放置在丝状细胞中的实例。当细胞已经在错配修复系统基因已被灭活的条件下复制一次之后，双链 DNA 内两种不同的单股 DNA 序列分别被复制，得到两种不同的双链序列，每个复制细胞中含有其一，没有任何错配。相反，在错配修复系统是活性的细胞中，双链内的错配被纠正了。

图 2：

本图显示从所克隆的 PCR 片段衍生的三种不完全米曲霉多肽序列：‘msh2’ Ao-col10/13/15。将这三种不完全多肽序列与其他两种已知错配修复蛋白：人错配修复蛋白，即 msh2-human. p；和酿酒酵母错配修复蛋白，即 S. c. msh2 的不完全多肽序列对比排列。图中带下划线的序列衍生自 PCR 片段的构建。

图 3：

本图显示公认的(putative)米曲霉错配修复蛋白(Ao. MSH2)的推定(proposed)多肽序列与三种已知的错配修复蛋白：人类来源(msh2-human. p)、鼠类来源(msh2-mus. p)和酵母来源(S. c. msh2)多肽序列的排列。



## 发明的详细说明

如本文所述的丝状真菌细胞，其中如本文所述参与错配修复系统的基因已被灭活。

### 参与错配修复系统的基因的灭活：

5 如本文所述参与错配修复系统的新型基因可以通过技术人员已知的多种已知技术之一加以灭活。

本发明的实施方案涉及如本文所述的丝状真菌细胞，其中参与错配修复的基因是有缺陷的。

10 当 DNA 序列已知时，大量使基因缺陷的方法都是技术人员已知的。这些方法包括删除基因的一部分 DNA 序列；通过删除或插入核苷酸引入移码突变；引入终止密码子等。

本发明优选的实施方案涉及如本文所述的丝状真菌细胞，其中参与错配修复的基因已被暂时灭活。

15 与上述类似，可做到这一点的大量方法都是技术人员已知的，包括在基因上游插入可调节的启动子，或者例如永久删除染色体上该基因的一部分，然后将载体(例如质粒)插入到包含该基因的细胞中。质粒可以在基因上游包含可调节的启动子，或者当错配修复系统需被暂时灭活时，质粒可以从细胞中除去，然后当错配修复系统需被重新激活时，可以重新插入到细胞中。

20 出于具体技术目的而选择适当的策略是技术人员的常识。

制备如本文所述能够暂时灭活错配修复系统的丝状真菌细胞的优选方法是，首先如本文所述永久灭活细胞染色体上的错配修复基因，然后将质粒插入到包含该基因的细胞中，其中该质粒的特征在于它包含适合的复制起始序列和适合的选择性标记。

25 优选地，适合的复制起始序列是 AMA1 (Gems, D.等(1991)《基因》98:61-67)。

下面立即提供适合的复制起始序列和适合的选择性标记的更详细说明，本文工作例 4 提供了这种策略的实例，使用包含 AMA1 的质粒作为复制起始序列，使用 AmdS 作为选择性标记。

30 复制起始序列

本文所用的术语“真菌的复制起始序列”被定义为这样一种核酸序列，



它能够支持真菌宿主细胞中的染色体外分子、例如质粒或 DNA 载体的自我复制，通常不发生质粒的结构重排或向宿主细胞基因组的整合。复制起始序列可以是任意来源的，只要它能够介导真菌细胞中的复制引发活性即可。

5 优选地，复制起始序列来自丝状真菌细胞，更优选为曲霉属、镰孢属或链格孢属的菌株，进而更优选为构巢曲霉、米曲霉、黑曲霉、尖孢镰孢或互格链格孢。

复制起始序列可以通过本领域熟知的方法加以鉴定。例如，可以从选定生物衍生的基因组片段中鉴定出能够在酵母中的持续自我复制的序列 (Ballance 和 Turner 《基因》 36 (1985), 321-331)，其预示丝状真菌细胞中的自我复制能力。指定序列的真菌复制引发活性还可以这样测定，用所选质粒复制子转化真菌，选择具有不规则形态的菌落，其指示分段质粒的丧失，该分段质粒会引起在选择性培养基上筛选质粒上发现的基因时的生长缺乏 (Gems 等 《基因》 98 (1991) 61-67)。AMA1 是以这种方式分析的。分析复制起始序列的另一种方式是分析天然存在的质粒(例如 Tusge 等所公开的，15 《遗传学》 146 (1997) 111-120，关于互格链格孢)。

复制起始序列的实例包括但不限于构巢曲霉的 ANS1 和 AMA1 序列，例如分别描述于 Cullen, D.等(1987 《核酸研究》 15: 9163-9175)和 Gems, D.等(1991 《基因》 98: 61-67)。

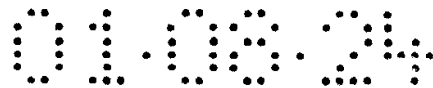
20 术语“复制引发活性”在本文中使用时其常规含义，也就是表示序列能够支持真菌细胞中的染色体外分子、例如质粒或 DNA 载体的自我复制。

术语“不发生质粒的结构重排”在本文中用于表示既没有删除质粒的一部分或插入该质粒的另一部分，也没有任何宿主基因组 DNA 插入到质粒中。

### 丝状真菌选择性标记

25 术语“选择性压力”在本文中被定义为在有或没有有效量的适当选择剂的存在下，培养含有 DNA 载体的丝状真菌细胞，该载体含有可操纵连接在目的多核苷酸序列上的真菌选择性标记基因。有效量的选择剂在本文中被定义为足以从不含选择性标记的细胞中选择出含有选择性标记的细胞的量。

30 在优选的实施方案中，真菌选择性标记选自，编码能够提供对杀生物剂或病毒毒性的抗性、对重金属毒性的抗性 or 原养型变为营养缺陷型的产



物的基因。

在更优选的实施方案中，原养型是从选自这样一组代谢途径的酶获得的，这些代谢途径是由核苷酸合成、辅因子合成、氨基酸合成、乙酰胺代谢、脯氨酸代谢、硫酸代谢和硝酸代谢组成的。

5 在进而更优选的实施方案中，在本发明的方法中，真菌选择性标记是这样一种基因，其选自 argB(鸟氨酸氨甲酰基转移酶)、amdS(乙酰胺酶)、bar(磷丝菌素乙酰基转移酶)、hemA(5-氨基乙酰丙酸合成酶)、hemB(胆色素原合成酶)、hygB(潮霉素磷酸转移酶)、niaD(硝酸盐还原酶)、prn(脯氨酸通透酶)、pyrG(乳清酸核苷-5'-磷酸脱羧酶)、pyroA、riboB、sC(硫酸腺苷酸转移酶)  
10 和 trpC(邻氨基苯甲酸合成酶)。

将真菌细胞在适合的培养基中、在适合筛选或选择携有变异的目的是多核苷酸序列的转化体的条件下培养，该序列具有或编码所需的特征。培养可以按照本领域熟知的用于筛选多核苷酸变体文库的方法进行。

#### 丝状真菌细胞

15 如本文所述的丝状真菌细胞包括真菌亚门(Eumycota)和卵菌亚门(Oomycota)的所有丝状形式。丝状真菌的特征是菌丝壁由壳多糖、纤维素、葡聚糖、脱乙酰壳多糖、甘露聚糖和其他复杂的多糖组成。营养生长借助于菌丝的伸长，碳的分解代谢为专性需氧型。相反，酿酒酵母等酵母的营养生长借助于单细胞原植体的芽殖，碳的代谢可以是厌氧发酵的。在优选  
20 的实施方案中，丝状真菌细胞是下列物种的细胞，支顶孢属、曲霉属、镰孢属、腐质霉属、毛霉属、毁丝霉属、链孢霉属、青霉属、小柱孢霉属(Scyrtidium)、草根霉属、Tolypocladium 和木霉属，但不限于此。

用在本发明中的丝状真菌细胞实例包括曲霉属细胞、支顶孢属细胞、镰孢属细胞、腐质霉属细胞、毛霉属细胞、毁丝霉属细胞、链孢霉属细胞、  
25 青霉属细胞、草根霉属细胞、Tolypocladium 细胞和木霉属细胞。

更具体地，丝状真菌细胞是泡盛曲霉、臭曲霉、日本曲霉、构巢曲霉、黑曲霉或米曲霉细胞；杆孢状镰孢、Fusarium cerealis、Fusarium crookwellense、大刀镰孢、禾谷镰孢、禾赤镰孢、异孢镰孢、合欢木镰孢、尖孢镰孢、多枝镰孢、玫瑰色镰孢、接骨木镰孢、肤色镰孢、拟分枝孢镰  
30 孢、硫色镰孢、Fusarium torulosum、Fusarium trichothecioides、Fusarium venenatum 的细胞或 Fusarium venenatum 细胞(Nirenberg sp. nov); Humicola

insolens 细胞或 Humicola lanuginosa 细胞；米赫毛霉细胞；嗜热毁丝霉细胞；粗糙链孢霉细胞；产紫青霉细胞；Thielavia terrestris 细胞；Trichoderma harzianum、康宁氏木霉、Trichoderma longibrachiatum、Trichoderma reesei 或绿色木霉细胞。

5        根据本发明第三方面制备包含 DNA 文库的丝状真菌细胞群体的方法：  
按照本发明第三方面方法的步骤(a)，分别将不同的目的 DNA 序列放置  
在丝状真菌细胞群体中；

10        如上所述，将这些 DNA 序列放置在丝状真菌细胞内的具体实验方法  
 可以按照多种适当技术之一进行，例如转化技术。参见普通的真菌教科书  
 《真菌遗传》(1996, ISBN 0-8247-9544-X)关于这些标准技术的进一步说明。

实例可以是将与染色体 DNA 序列部分同源的单股寡核苷酸序列放置在  
 细胞内。参见 Calissano 等《真菌遗传通讯(Fungal genetic newsletter)》43:  
 15-16 (1995)关于这一点的进一步说明。

15        另一种实例可以是将包含错配的双链 DNA 序列(例如图 1 所示，包含  
 在载体内)放置在细胞内。

在优选的实施方案中，将不同的目的 DNA 序列包含在质粒中，其中  
 该质粒的特征在于，它包含如上所述的适当复制起始序列和适合的选择性  
 标记。

20        优选地，适合的复制起始序列是 AMA1 (Gems, D.等(1991)《基因》98:  
 61-67)。

按照本发明第三方面的步骤(b)，将步骤(a)的群体生长一段时间使群体  
中的单个目的 DNA 序列在错配修复系统基因已如本文所述被灭活的条件下  
复制至少一次。

25        群体的生长可以在多种适合的已知用于生长丝状真菌细胞的培养基之  
 任一中进行。选择这样一种适合的培养基是技术人员的常识。

上文已经解释过，群体中的单个细胞必须能在错配修复系统基因已如  
 本文所述被灭活的条件下复制至少一次。

所述细胞当然可以在错配修复系统基因已被灭活的条件下复制一次以  
 上。

30        由于错配修复系统的灭活通常将导致细胞内染色体 DNA 上突变的累  
 积，由此可能成为细胞的致死突变，因此实际优选的上述复制周期次数将

取决于这些潜在的致死突变出现得有多快。

确定优选的复制周期有多少是技术人员的常识。

由于这些潜在的致死突变，优选步骤(b)下的错配修复系统已经被暂时灭活。

- 5 在根据第三方面步骤(b)的适合的复制周期之后，重新激活暂时灭活的错配修复系统，目的是避免丝状真菌细胞本身中的这些致死突变。这种暂时灭活的策略可以是任意上述策略。

- 10 另一种限制突变引入到染色体上的策略是在错配修复缺陷的条件下暂时停止染色体的复制，而复制染色体外元件。这可以通过在仅为染色体复制所必需的元件中引入突变加以实现。

优选的策略是使用这样的丝状真菌细胞，其中如本文所述参与错配修复系统的基因在细胞染色体上被永久灭活，然后插入质粒到包含该基因的细胞中，其中该质粒的特征在于它包含适合的复制起始序列和适合的选择性标记。见上关于这种策略的进一步解释。

- 15 优选地，适合的复制起始序列是 AMA1 (Gems, D.等(1991)《基因》98: 61-67)。

进一步的实施方案涉及本发明第三方面的方法，其中步骤(b)下的错配修复系统是有缺陷的。

- 20 在进一步的实施方案中，本发明涉及如本文所述的方法，其中在本发明第三方面的步骤(b)下，存在部分同源的目的 DNA 序列的体内基因间重组。

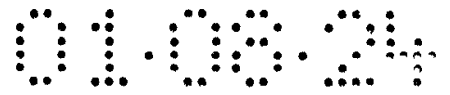
由于本发明的总体概念是提供涉及错配系统灭活的方法，当然优选所述部分同源的 DNA 序列能够体内形成包含错配的双链 DNA 序列。

- 25 按照本发明的第四方面制备目的多肽的方法，其包括本发明第三方面的步骤且其中目的 DNA 序列编码目的多肽：

按照第四方面的步骤(c)，从第三方面步骤(b)所得丝状真菌细胞群体中选择所需的目的是多肽。

- 30 所需的目的是多肽可以是包含所需技术特征的任意多肽，所述技术特征例如提高了的稳定性；所需的特异性活性；所需的最佳 pH；在洗涤剂中的改进的洗涤性能；等。

用于选择这种所需的目的是多肽的特殊策略可以是技术人员已知的大量



选择策略之任一，例如平板筛选试验、基于微滴板的试验等。

本发明的实施方案涉及本发明第四方面的方法，进一步包括下列步骤：

(d)可选的步骤，其包括根据进一步的特定需要修饰按需要选择的目的多肽的氨基酸序列；

5 (e)将编码第四方面步骤(c)的目的多肽或步骤(d)的修饰型目的多肽的DNA序列放置到适合于大规模制备该目的多肽的丝状真菌细胞中；

(f)在允许目的多肽表达的条件下，在至少 10000m<sup>3</sup> 的发酵罐中培养步骤(e)的丝状真菌细胞；

(g)分析目的多肽。

10 该实施方案涉及与工业化密切相关的方法，其中所选择的目的多肽是大规模制备的。

可选的步骤(d)涉及这样一种情形，其中选择例如所需的目的多肽，目的例如是鉴定根据本发明第三方面步骤(c)的、在洗涤剂中具有改良的洗涤性能的多肽。这种多肽尽管在洗涤剂中具有改进的洗涤性能，但作为商业  
15 产品可能不够稳定。因此，可能需要该选定的多肽中进行一些进一步的氨基酸取代，例如适合的脯氨酸取代，目的是获得足够的稳定性，以使这种多肽商品化。

进一步的实施方案涉及刚才的实施方案的方法，其中适合于大规模制备所述实施方案步骤(e)的目的多肽的丝状真菌细胞是不同于本发明第三方面  
20 步骤(a)的丝状真菌细胞的另一种丝状真菌细胞。

该实施方案涉及这样一种情形，其中用于选择目的多肽的丝状真菌细胞不同于用于大规模制备的丝状真菌细胞。

进一步的实施方案涉及如本文所述的方法，其中目的多肽是从丝状真菌细胞衍生的多肽。

25 术语“从丝状真菌细胞衍生”应当被理解为目的多肽的氨基酸序列信息是从得自丝状真菌细胞的多肽衍生的。

所以，它可以是野生型丝状真菌多肽的变体和/或可以是两种或更多不同的丝状真菌多肽的重组/改组产物。

在进而进一步的实施方案中，本发明涉及如本文所述的方法，其中目  
30 的多肽是一种酶，例如淀粉酶、蛋白酶、纤维素酶、脂肪酶、木聚糖酶、磷脂酶。

实施例：

材料

用作缓冲剂和底物的化学品是至少为试剂级的商业产品。

5 实施例 1：

适合于测定如本文所述的丝状真菌细胞是否在错配修复系统中被灭活的凝胶：

10 这种凝胶移动(shift)试验的原理是，细胞提取物从下列制成：(a)丝状真菌细胞，其中如本文所述参与错配修复系统的基因是灭活的；和(b)相应的丝状真菌细胞，其中的基因不是灭活的。

这些提取物然后与含有碱基对错配的 G:T、G:A、G:G、A:C 的寡核苷酸和超螺旋的 TG 的二核苷酸结合/混合，在非变性凝胶上进行测定。

15 如果凝胶移动(shift)试验证明所含基因未灭活的对照丝状真菌细胞包含与任意上述寡核苷酸结合的任意蛋白质，并且这些结合蛋白在所含如本文所述参与错配修复系统的基因已灭活的丝状真菌细胞中没有见到，那么确认后错配修复系统是灭活的。

实验方案：

细胞提取物的制备如 Nagata 等所述进行(Mol. Gen Genet (1993) 237: 251-260; 见材料和方法)。

20 基本上如文献所述进行寡核苷酸的退火、细胞提取物与含有错配的双链寡核苷酸的结合、以及非变性聚丙烯酰胺凝胶电泳(Stephenson 和 Karran; 来自人类细胞的蛋白质与 DNA 碱基对错配的选择性结合; 《生物化学杂志》264: 2177-21182 (1989))。

25 不过，凝胶电泳是在 TAE 缓冲液、而不是 TBE 缓冲液中进行的。为了获得双链寡核苷酸，使寡核苷酸 U 被放射性标记，并与任意未标记的寡核苷酸 L-G. T、L-G. A、L-G. C、L-A. C、L-T. G 或 L-HOM 退火。寡核苷酸序列是从 Aquilina 等《美国国家科学院院报》91: 8905-8909 (1994)获得的。

U:5'-GGGAAGCTGCCAGGCCCCAGTGTCAGCCTCCTATGCTC-3'  
(SEQ ID NO 3);

30 L-G.T:5'-GAGCATAGGAGGCTGACATTGGGGCCTGGCAGCTTCCC-3'  
(SEQ ID NO 4) (导致 G. T 错配);



L-G.A.:5'-GAGCATAGGAGGCTGACAATGGGGCCTGGCAGCTTCCC-3' (SEQ ID NO 5) (导致 G. A 错配);

L-G.G.:5'-AGCATAGGAGGCTGACAGTGGGGCCTGGCAGCTTCCC-3' (SEQ ID NO 6) (导致 G. G 错配);

5 L-A.C.:5'-GAGCATAGGAGGCTGACACCGGGGCCTGGCAGCTTCCC-3' (SEQ ID NO 7) (导致 A. C 错配);

L-TG:5'-GAGCATAGGAGGCTGACACTGTGGGGCCTGGCAGCTTCC C-3' (SEQ ID NO 8) (导致超螺旋的 TG 二核苷酸);

10 L-HOM:5'-GAGCATAGGAGGCTGACACCGGGGCCTGGCAGCTTCCC-3' (SEQ ID NO 9) (导致同源双链)。

所有实验均包括两倍过量的未标记的同型双链竞争性寡核苷酸。

### 实施例 2:

#### 参与米曲霉细胞错配修复系统的基因的克隆

如本例所述被克隆的基因显示在 SEQ ID NO 1(DNA 序列)和 SEQ ID NO 2(翻译的氨基酸序列)中。

来自不同生物的错配修复蛋白的若干序列是已知的, 下面仅用到其中的三种: 酿酒酵母(M84170)、H. sapiens (L47580)和小鼠(U21011)。

所示数字是公众可以获得的 GenBank 数据库的查询数。

20 基于已知错配修复蛋白之间的 C-端同源性, 设计了一组简异引物, 目的是扩增米曲霉同系物的不完全序列:

Pr 117858 (SEQ ID NO 10): P-GGCNCARATHGGNTGYTTYGTNCC

Pr 117859 (SEQ ID NO 11): P-GCCCANGCNARNCCRAANCC

25 利用来自米曲霉菌株 JaL 142(WO 96/29391)的染色体 DNA 作为模板以及上述引物, 在八种 MgSO<sub>4</sub> 浓度(0.5 mM 至 4.0 mM, 按照制造商的建议; Boehringer M.)下进行下列基于 50μl PWO 聚合酶的 PCR 反应。发现 1 mM MgSO<sub>4</sub> 是最优的, 得到约 230bp 的单一带, 是预计序列中不含内含子的结果。

PCR 循环为: [96°C 2 min — (94°C 15s; 50°C 15s; 72°C 30s)循环 30 次 — 72°C 7 min — 4°C 维持]。

30 230bp PCR 片段经钝端连接在 pUC19 的 BamHI 位点。pUC19 在牛小肠碱性磷酸酶的存在下经 BamHI 裂解, 然后通过 klenow 聚合酶和 dNTP 补

平粘性末端。从上述连接反应的大肠杆菌 XL1 转化体中分析携有插入片段的三个质粒，并测序。对克隆的 PCR 片段翻译的多肽进行排列揭示了它与已知的错配修复蛋白序列具有较高同源性(见图 2)。

图 2 中带下划线的序列是从上述共有 PCR 引物衍生的序列。

5 图 2 的三种曲霉属序列等于 SEQ ID NO 2 之 683-758 位所示序列，但其中 685 位在最终所克隆的序列中是 Thr (T)，而不是所示的 Ile (I)。这是由上述共有引物中的序列引起的。

图 2 所示排列清楚地证明，所克隆的片段来自错配修复蛋白的米曲霉同系物。

10 为了克隆完整的基因，通过 PCR 生成所克隆片段的放射性标记探针，即，使用 0.5 mg pUC19'msh2'-13(见上)作为模板，在 100 ml 中与 Taq 聚合酶、30 pmol pUC 正向与反向引物、0.2 mM dG-、dC-、dTTP 和 0.2 mM dATP + <sup>32</sup>P-dATP 反应。所生成的放射性标记探针通过 EcoR1-Hind3 消化作用从 pUC19 序列中释放出来，然后进行所得 293bp 片段的凝胶纯化。

15 使探针与来自米曲霉菌株 A1560(JaL 142 的父代)(WO 96/29391)基因组 DNA 的膜格点(grided)粘粒文库杂交。通磷酸成像仪(Phosphoimager)中分析而鉴定滤膜上的阳性克隆为λ31A2。

繁殖λ31A2 粘粒 DNA，使用上述相同的放射性标记引物进行 Southern 分析。鉴定了大约 9Kb Pst 片段，其被 BstX(以前在所克隆的 PCR 片段中发现)分裂为 5.8 和 3.2 kb 片段，均因携有探针而发出信号，将该 9kb 片段克隆至已用 Pst 切割的 pUC19，得到名为 pUC19msh2P 的质粒。从表明以前测定过的序列的引物开始按顺序对插入片段测序，以下是基于最后一轮所测序列的引物：

130740 (SEQ ID NO 12): GCTCGAAACATCCAACATCC  
 25 130741 (SEQ ID NO 13): GCTGTGAATCACTTGCACC  
 131928 (SEQ ID NO 14): C TTCATAAACTGCGACAAATCATGC  
 131929 (SEQ ID NO 15): GGAGGAGCATCTTCGC  
 131930 (SEQ ID NO 16): G GAACTTGAAGACTTTACTTCATCC  
 134608 (SEQ ID NO 17) CCAGAAACTCGCTAACACC  
 30 134609 (SEQ ID NO 18): GTGCTTTGCGGACGC  
 134610 (SEQ ID NO 19): CAGGACAGTAGGACGC

135320 (SEQ ID NO 20): CGAGCGATGAACTCTGC

135321 (SEQ ID NO 21): GCGTTGGTGGATTATCC

136105 (SEQ ID NO 22): CGTTGCATCTATCATATAACC

136106 (SEQ ID NO 23): GGTATATGATAGATGCAACGC

5 据此测定的 3825bp 序列(SEQ ID NO 1)在以前在 PCR 片段中测定的框架中翻译。所得蛋白质(SEQ ID NO 2)被称为 Ao. MSH2, 将其与图 3 中已知错配修复蛋白的蛋白质序列排列。从图 3 所示排列可以看出, 所克隆和测序的 DNA 清楚地涵盖了酵母、人和小鼠错配修复蛋白的同源物的编码序列, 其在 N-端部分有一个内含子。按照标准的剪接规则推断内含子的位置, 并且构成唯一的可能性。

### 实施例 3:

破坏米曲霉细胞染色体上实施例 1 所克隆的基因:

关于破坏实验, 通过 PCR 从 pUC19msh2P(见实施例 2)中删除 msh2 CDS, 引入 NotI 位点以代之。进行该反应的引物是:

15 137208 (SEQ ID NO 24): 5'P-CCGCGTCTCCAACAAGATGAATGG

137207 (SEQ ID NO 25): 5'P-CCGCTTTCTCGGGGTCATAGC

在与 2.5 mM MgSO<sub>4</sub> 和 50 pg pUC19msh2P 所进行的基于 Pwo 聚合酶的 PCR 反应(按照制造商建议的条件)中:

20 PCR 循环为: [96°C 2min — (94°C 30s; 52°C 30s; 72°C 3min)循环 4 次 — (94°C 30s; 59°C 30s; 72°C 3min)循环 25 次 — 72°C 10min]。

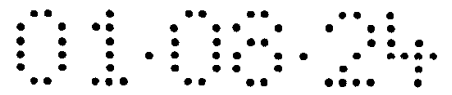
分析所得约 8.9 Kb 的 PCR 产物, 将其连接至 pUC19, 然后转化大肠杆菌 XL1。从所得转化体分析 pMsh2Δ, 通过测序[引物 138149 (SEQ ID NO 26): CCTTTCCACTTTAATCCTAAGC](X11/ pMsh2Δ: Lac3073)验证新接合及其周围点的正确性。

25 将米曲霉 pyrG (WO 96/29391)插入该构建体的 NotI 位点。

通过利用这种构建体, 可按本领域已知的关于通过同源重组在染色体上使这类删除基因交叉(crossing)的标准技术在米曲霉细胞中删除所述染色体基因(Miller, B. L.等, 1985 《分子与细胞生物学》 5: 1714-1721)。

### 实施例 4:

30 包含如 SEQ ID NO 1 所示错配修复基因、AMA1 复制起始序列及 AmdS 选择性标记的质粒的构建:



如下所述构建的质粒非常适合于制备所含错配修复系统已暂时灭活的丝状真菌细胞，其中当错配修复系统应当被激活时，可将这种质粒插入到实施例 3 的已错配破坏的菌株中，当错配修复系统应当被灭活时，可将这种质粒从菌株中删除。

5 错配修复基因的破坏可以导致新染色体突变的蓄积，因此这样的菌株可能是遗传上不稳定的。所以，决定在下述菌株中进行染色体破坏，所述菌株中错配修复基因由染色体外元件表达，当需要错配修复缺失型表型时该元件能很轻易地丢失。

10 这里，染色体外元件是包含 AMA1 作为复制起始序列和 AmdS 作为选择性标记的质粒。

为此，将错配修复基因(SEQ ID NO 1)克隆到携有一个 AMA1 重复序列的自我复制型构建体中。

从 pMT1505(见下实施例 5 关于 pMT1505 的说明)中分离下列片段，并将它们连接在一起：

15 5.16 kb NotI- [Hind3] \* + 3.515 kb [Sal] \*-BamHI + 757bp BamHI-NotI  
[] \*表示该位点已用 klenow-聚合酶和 dNTP 补平

通过该连接反应分析出 pMT1505DHS (LaC 3212)，将错配修复表达盒作为 BamHI-MunI 片段引入 pMT1505DHS 的相应位点，得到质粒 pAma-msh2 (LaC 3216)。

20 将米曲霉 JaL250(见实施例 5)用 pAma-msh2 转化为 AmdS<sup>+</sup>，检查未被选择的转化体(转化体的 50%)丢失 amdS 性状的能力，说明这种质粒是作为染色体元件维持的。(LaC3244 保持在乙酰胺+尿苷上)。

### 实施例 5：

#### 实施例 4 所用质粒 pMT1505 的构建：

25 质粒

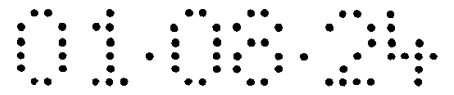
pMT1505：如下实施例 5 所述构建

pHelp1：含有米曲霉 pyrG 基因作为选择性标记，还含有能在构巢曲霉中自我复制的 AMA1 序列，如 Gems, D.等(1991.《基因》98: 61-67)所述

pToC68：如 EP 0 531 372 (Novo Nordisk A/S)所述

30 菌株

JaL250：米曲霉 A1560 的衍生物，其中 pyrG 基因已灭活，如 WO 98/01470



所述

DH5a: 大肠杆菌宿主细胞, 购自 GIBCO BRL (Life Technologies, Inc., Rockville MD)

5 pMT1466 通过将来自 pHelpI 的 SphI/NarI 片段插入到 pToC68 中而构建。pMT1489 通过用 SphI 和 StuI 消化 pMT1466, 然后重新连接而构建。pMT1500 通过用 AatII 和 NarI 消化 pMT1489, 再连接一个接头而构建。pMT1504 通过用 NheI 消化 pMT1500, 然后重新连接而构建。

10 pMT1505 通过将含有构巢曲霉基因组 DNA 中 amdS 编码基因的 2.7 kb XbaI 片段(Corrick, C. M.等 1987 《基因》 53: 63-71)插入已用 NheI 切割的 pMT1504 中而构建。

### 实施例 6:

#### 染色体上一部分 mshII 基因的删除

15 将质粒 p418MsHII(来自 lac3159)用 SalI 和 XhoI 切割, 再用牛小肠磷酸酶处理。以这种方式剪出 msHII 基因的一部分。从 1%琼脂糖凝胶中分析含有载体和大部分 msHII 基因的大带(6400bp)。

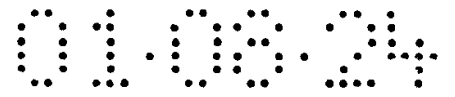
20 质粒 pJal554 通过将 pS02 的 SpeI/SspBI 切割片段(5330 bp)与 pS02 的 Asp718/NheI 切割片段(316 bp)连接而构建。将质粒 pJal554 用 SalI 切割, 在 1%琼脂糖凝胶上分析含有 pyrG 基因的 2350bp 带。将含有 pyrG 的 2350bp 带与切割的 p418MsHII 质粒连接, 转化到大肠杆菌中。通过限制性分析鉴定正确的大肠杆菌转化体, 从该转化体制备质粒制剂。

将所制备的质粒用 EcoRI 切割, 目的是使质粒在转化到例如米曲霉 Jal250 之前线性化。在基本平板上选择转化体。

25 发生双遗传交换的转化体是这样鉴定的, 制备曲霉属染色体 DNA 制剂, 然后利用适当引物对全长 mshII 基因进行 PCR 筛选。利用被适当的酶随机片段化的染色体 DNA 以及适当的针对被删除之 msHII 片段(不复存在)的探针以及阳性对照探针进行 Southern 印迹分析。

为了测定具有灭活的 msHII 基因的菌株中所增加的突变频率, 筛选 niaD 基因中的突变。这一点是通过在平板上培养母体菌株曲霉 Jal250 和 msHII 灭活型菌株来实现的。

30 制备孢子悬液, 将孢子试样接种在含有硝酸盐的平板上, 如 Unkles 等所述(S. E. Unkles, E. C. Campbell. Y. M. J. T. de Ruite Jacobs, M. Broekhuisen,



J. A. Macro, D. Carrez, R. Contreras, C. A. M. J. J. van den Hondel J. R. Kinghorn. 基于硝酸盐同化途径的米曲霉同源转化系统的开发：用于丝状真菌转化的方便和通用的选择系统,《分子与普通遗传学》V: 218 p 99-104 (1989))。

- 5 没有 MsHII 蛋白表达的菌株将比对照菌株具有更高频率的 *niaD* 突变(更多的氯酸盐抗性克隆)。

实施例 7:

使用 TAKA 启动子驱动反义 mRNA 的转录从而制备抑制 msHII mRNA 翻译的体内反义 msHII RNA

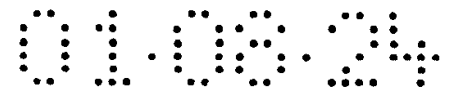
- 10 反义 RNA 表达是已知体内下调任意基因表达的方式(反义 RNA 的设计, Georg Sczakiel, 反义与核酸药物开发 V. 7 P. 439-444 (1997))。

使用下列 oligo: 000120j2 (SEQ ID NO 27): TCTGCGAATCGCTTGGAT CCCGAACGCGACAACAC、000120j4 (SEQ ID NO 28): GAGCTCAGATCTC TTAGGTTCTGGACGAGAAGA 并以 pUC19msh2P 作为模板制备 PCR 片段。该 PCR 片段含有 msHII 基因的 5' 末端, 包括 5' msHII mRNA 的假定部分。使用下列 oligo: 000120j3 (SEQ ID NO 29): GTTGTCGCGTTCGGG ATCCAAGCGATTCGCAGAAG、1298-TAKA (SEQ ID NO 30): GCAAGCGC GCGCAATACATGGTGT TTTTGATCAT 并以 pENI1298 作为模板 (PCT DK99/00552) 制备另一种 PCR 片段。两次 PCR 反应都是按照制造商手册 (Boehringer-Mannheim) 并使用 PWO 聚合酶进行的。

PCR 片段用 Qiagen PCR 纯化试剂盒(Qiagen)加以纯化。将两种 PCR 片段混合, 利用引物 1298-TAKA 和 000120j4 进行第三次 PCR 反应。以这种方式装配所述两种 PCR 片段。

25 装配后的 PCR 片段用 BssHII 和 BglII 切割, 从 1.5%琼脂糖凝胶上纯化, 再与 pENI1298 连接, 后者已用 BssHII 和 Bgl II 切割(从 1%琼脂糖凝胶上纯化)。将连接混合物转化到大肠杆菌中。制备所得每种大肠杆菌转化体的 DNA 制剂。对装配后的 PCR 片段进行测序, 以确认在操作期间没有引入所不需要的突变。正确的构建体含有 TAKA 启动子, 它驱动 msHII 反义 mRNA 的转录。

30 以 pENI1298 作为对照, 将所得质粒转化到例如米曲霉 J1250, 在基本平板上选择转化体。在基本平板上分析所得转化体, 在 37°C 下培养, 直到



它们形成孢子。

为了测定菌株(其中 msHII mRNA 的翻译因 msHII 反义 RNA 的表达而受到阻抗)中所增加的突变频率, 进行 niaD 基因中突变的筛选。制备对照转化体(pENil298)和 msHII 反义 RNA 转化体的孢子悬浮液, 将等量孢子接种在含有硝酸盐的平板上, 如 Unkles 等所述(S. E. Unkles, E. C. Campbell, Y. M. J. T. de Ruite Jacobs, M. Broekhuisen, J. A. Macro, D. Carrez, R. Contreras, C. A. M. J. J. van den Hondel J. R. Kinghorn. 基于硝酸盐同化途径的米曲霉同源转化系统的开发: 用于丝状真菌转化的方便和通用的选择系统, 《分子与普通遗传学》 V : 218 p 99-104 (1989))。

- 10 无或低 MsHII 蛋白表达的菌株将比对照菌株具有更高频率的 niaD 突变(更多的硝酸盐抗性克隆)。

## 说明书附图

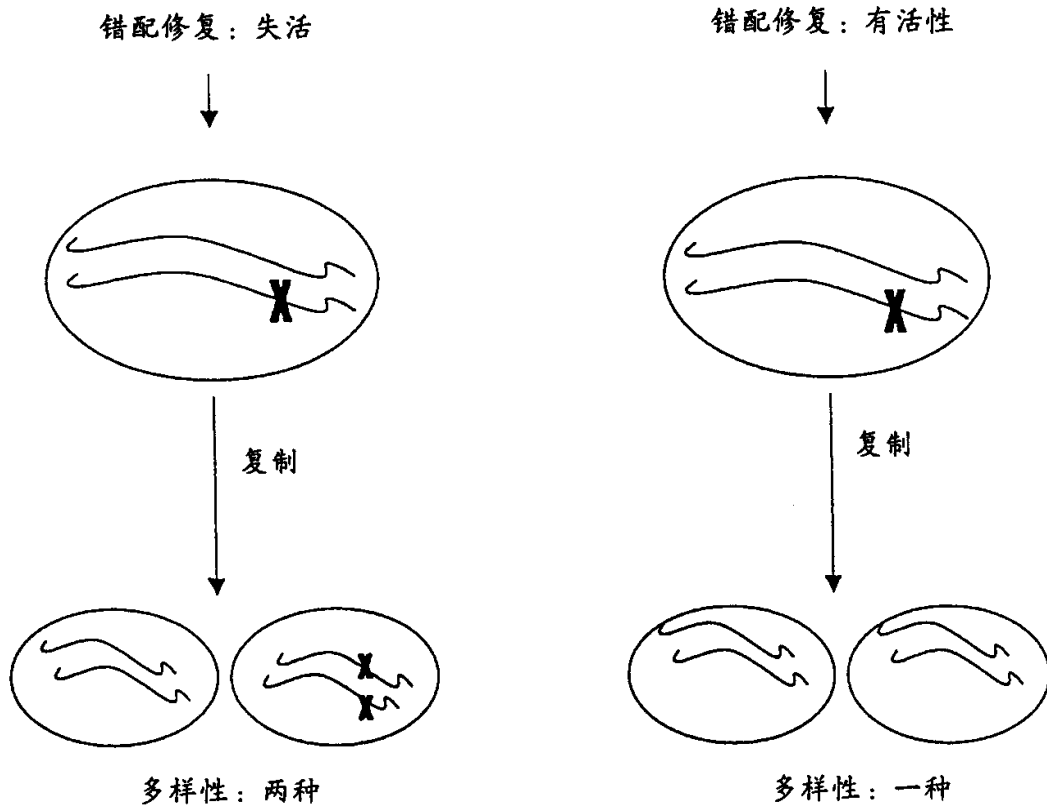


图 1

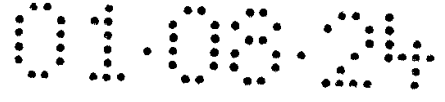




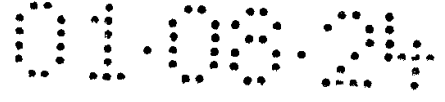
	10	20	30	40	50	60
Contig# 1						
S.c. msh2	o	o	o	o	o	o
Ao.MSH2	MSSTRPELKFSDVSEERNFYKKYTGLPKKPLK-TIRLVDKGDYTVIGSDAIFVADSVYHTQSVLK					
msh2 mus.p	MSS-RPELV-D--DEVGFI RFYRSLAANSND EIRVFD RGDWYSAHGAKAEFIARTVYKTT SILR					
msh2-human.p	MAVQPKETLQLEGA AEAGVRF FEGMPEKPT-TVRLFDRGDFYTAHGEDALLAAREVFKTQGVIK					

	70	80	90	100	110	120	130
Contig# 1							
S.c. msh2			o	o	o	o	o
Ao.MSH2	NCQLDPVTAKNFHEPTKYVTVSLQVLATLLKCLLDLGYKVEIY-----DKGWKLIKASASP						
msh2 mus.p	N--LGRSDSGGLPS---VTMSVTVFRNFLREALFRLNKR IEIW---GSVGTGKGHWKLVKQASP						
msh2-human.p	Y--MGPAGAKNLQS-----VVL SKMNFESFVKD LLLVLRQYRVEVYKKNRAGNKASKENEWYLA FKASP						

	140	150	160	170	180	190
Contig# 1						
S.c. msh2	oo	o	o	o	o	o
Ao.MSH2	GNIEQVNELM-----NMNIDSSIIIASLKVQWNSQDGN CII GVAFIDTTAYKVGMLDIVDNEVYSNL					
msh2 mus.p	GNLQDVEEELG SVGGLSMD SAPI ILAVKIS-AKAAEARSVGC FADASVRELGVSEFLDNDIYSNF					
msh2-human.p	GNLSQFEDIL--FGNNDMSASVGMGIKMA-VVDGQ-RHVG VGYVDSTQRKLG LGLCEFPENDQFSNL					







	400	410	420	430	440	450	460
Contig# 1							
S.c. msh2	oooo     o  oo  oo o     oo o   oo   o						
	YLIDQIELRQMLTSEYLPMPIDIRRLTKKLNKRG-NLEDVVKIYQFSKRIPEIVQVFTSFLEDDSP						
Ao.MSH2	AFVVNTELRQTMQEEHLRSIPDLYRLAKRFQRKANLEDDVVRVYQVAIRLPGFVNSLENVMDEEYQ						
msh2 mus.p	AFVEDSELRSLOEDLLRRFPDLNRLAKKFORQAANLQDCYRLYQGINQLPSVIALEKY - EGRHQ						
msh2-human.p	AFVEDAELRQTLQEDLLRRFPDLNRLAKKFORQAANLQDCYRLYQGINQLPNVIALEKH - EGKHQ						
	470	480	490	500	510	520	
Contig# 1							
S.c. msh2	o  o  oo ooo o o o oo  o o   o  o    o						
	TEPVNELVRVWLAPLSHHVEPLSKFEEMVETTVDLDAYEENNEFMIKVEFNEELGKIRSKLDTLR						
Ao.MSH2	T-----PLETEYTSNLRSHSDSLAKLEEMVETTVDLDALE -NHEFIIKPEFDESLRIIRKKLDKLR						
msh2 mus.p	A-----LLLAVFVTPLIDLRSDFSKFQEMIEETTLDMQVE -NHEFLVKPSFDPNISELREVMDGLE						
msh2-human.p	K-----LLLAVFVTPLTDLRSDFSKFQEMIEETTLDMQVE -NHEFLVKPSFDPNISELREIMNDLE						
	530	540	550	560	570	580	590
Contig# 1							
S.c. msh2	o  o  o  o  o    o  o  o  o     o       o  o  o  o						
	DEIHSIHLDSEAEDLGFDPDKKLENNHLLHGWCMRLTRNDAKELRKHKKYIELSTVKAGIFFSTKQ						
Ao.MSH2	HDMGVEHRRVARDLDQDIEKKLFLFNHRVHGWCFCRLTRNESGCIARNKREYQECSTQKNGVYFTTST						
msh2 mus.p	KKMQSTLINAARGGLDGPQKIKLDSSAQFGYYFRVTCKEEKVLRNNKNFSTVDIQKNGVKFTNSE						
msh2-human.p	KKMQSTLISAARDGLDGPQKIKLDSSAQFGYYFRVTCKEEKVLRNNKNFSTVDIQKNGVKFTNSK						

图 3 (续)

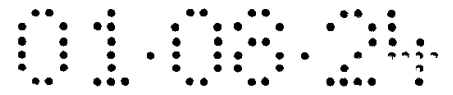


	800	810	820	830	840	850
Contig# 1						
S.c. msh2	oooooooooooo	o   o   oooooooo	o     o  o oo o			
Ao.MSH2	GTSTYDGFGLAWAIAEHIAASKIGCFALFATHFHELTELSEKLPN-VKNMHVVAHIEKNLKEQKHDD					
msh2 mus.p	GTSTYDGFGLAWAISEHIVTEIRCFGLFATHFHELTALADRYPKSVKNLHVVAFIGDGTDDDDSEDK					
msh2-human.p	GTSTYDGFGLAWAISDYIATKIGAFCMFATHFHELTALANQIP-TVNNLHVVTALTTEET-----					
	GTSTYDGFGLAWAISEYIATKIGAFCMFATHFHELTALANQIP-TVNNLHVVTALTTEET-----					
Contig# 1						
S.c. msh2	o oo o o ooooooooo	oo     oo oo				
Ao.MSH2	-----EDITLLYKVEPGISDQSFGIHVAEVVQFPEKIVKMAKRKANELDDLKTNNEDLK-----KA					
msh2 mus.p	KSKRNQVTLLYRVEPGICDQSFGIHVAELVRFPEKVVNMMARQKAELEDFTS-SEQDQSSMAID					
msh2-human.p	-----LTMLYQVKKGVCDQSFGIHVAELANFPRHVIAACAQKALELEEFQNGTSLGCDEAEPAA					
	-----LTMLYQVKKGVCDQSFGIHVAELANFPKHVIECAKQKALELEEFQYIGESQGYDIMEPAA					

图 3 (续)





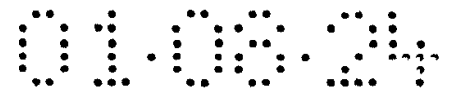


## 序列表

<110> Novo Nordisk A/S  
<120> 丝状真菌错配修复缺陷型细胞  
<130> 错配修复缺失型丝状真菌  
<140>  
<141>  
<160> 30  
<170> PatentIn Ver. 2.1  
<210> 1  
<211> 3823  
<212> DNA  
<213> 米曲霉  
<220>  
<221> CDS  
<222> (700)..(723)  
<220>  
<221> 内含子  
<222> (724)..(780)  
<220>  
<221> CDS  
<222> (781)..(3579)  
<400> 1  
gtgggtagtg gcccaaaagc tactgtggct gccagagga agccgtttcg cctagagatg 60  
atcgtacaga acgttcagga ctcatcgaag aatcctatct cagagaaaga ggtggaatc 120  
tgcgtcgaag tgcctgctcg gcccgacatt gctggacaat gggctgattt cgtcaccgtg 180  
aatcatatca aatcgggtggt tctgaaatcc tccgcggata tcaacctcaa ggatatcggt 240  
gcgaagggtgc gtagaactgaa gttcggcgag gacgagcctg cttcagcctc gaaccacctaa 300  
tcaagacttc tactgtttaa tgtgtgtttt ggatgtttgg tgttgctggg ttggataccc 360  
catctgtgga gtttcataca catgacttta ttatcacctt tgtgtagcat ctgtagcgt 420  
tgcattctatc atataccatt tttagctatta gagaatacat atcaatcatg atgaacatga 480  
agtacttcag ttctatgtg taggcgtttt ttggacctat cactttgtag agttetaacg 540  
ggtgctatta ttggccaatg atacttcaat atatcgcgaa cgcgacaaca cgtgaccgcg 600  
ttgccaegga gcgcgtctcc tgaattatc gaaatagatc gatcacggca gagctaattg 660  
tcagtcttcc attcatcttg ttggagacta actggcaag atg tct tct cgt cca 714  
Met Ser Ser Arg Pro  
1 5



gaa ctt aag gtaagttaaac aagacaaccg gtctctcgaa cattcaataa	763
Glu Leu Lys	
caattaaccc tggtag gtt gac gac gaa gtc ggc ttc att cgt ttt tac	813
Val Asp Asp Glu Val Gly Phe Ile Arg Phe Tyr	
10 15	
cgt tcc ctc gca gca aat agc aac gat gaa act att cgc gtt ttc gac	861
Arg Ser Leu Ala Ala Asn Ser Asn Asp Glu Thr Ile Arg Val Phe Asp	
20 25 30 35	
cgc ggt gac tgg tac tct gcc cat ggc gcc aaa gca gag ttc atc gct	909
Arg Gly Asp Trp Tyr Ser Ala His Gly Ala Lys Ala Glu Phe Ile Ala	
40 45 50	
cgc act gtg tac aag acc acc tct ata ctc cgc aat cta ggt cgc agc	957
Arg Thr Val Tyr Lys Thr Thr Ser Ile Leu Arg Asn Leu Gly Arg Ser	
55 60 65	
gac tca gga ggc ctt ccc tcc gtc acc atg agt gtc acc gtc ttc cgt	1005
Asp Ser Gly Gly Leu Pro Ser Val Thr Met Ser Val Thr Val Phe Arg	
70 75 80	
aac ttt ctc cgc gaa gct ctc ttc cga ctc aac aag cgc att gaa atc	1053
Asn Phe Leu Arg Glu Ala Leu Phe Arg Leu Asn Lys Arg Ile Glu Ile	
85 90 95	
tgg ggc tca gtc gga acg ggc aag ggt cat tgg aag ctg gta aag caa	1101
Trp Gly Ser Val Gly Thr Gly Lys Gly His Trp Lys Leu Val Lys Gln	
100 105 110 115	
gct agc ccg gga aac ctc caa gat gtg gaa gaa gag ttg ggc agc gtt	1149
Ala Ser Pro Gly Asn Leu Gln Asp Val Glu Glu Glu Leu Gly Ser Val	
120 125 130	
ggt gga tta tcc atg gac tcg gct cca att atc cta gca gtg aag atc	1197
Gly Gly Leu Ser Met Asp Ser Ala Pro Ile Ile Leu Ala Val Lys Ile	
135 140 145	
tcg gcc aag gcc gca gag gct agg agt gtg gga gtg tgc ttt gcg gac	1245
Ser Ala Lys Ala Ala Glu Ala Arg Ser Val Gly Val Cys Phe Ala Asp	
150 155 160	
gca agt gta cgg gaa ctc ggt gtt agc gag ttt ctg gat aac gat atc	1293
Ala Ser Val Arg Glu Leu Gly Val Ser Glu Phe Leu Asp Asn Asp Ile	
165 170 175	
tat tcc aac ttt gag tcg ctt att atc caa ctc ggg gtg aag gag tgt	1341
Tyr Ser Asn Phe Glu Ser Leu Ile Ile Gln Leu Gly Val Lys Glu Cys	
180 185 190 195	
ttg gtg cag atg gat gct aat aag aag gat gtt gag ctg gga aag att	1389
Leu Val Gln Met Asp Ala Asn Lys Lys Asp Val Glu Leu Gly Lys Ile	
200 205 210	
cgg gct att gcg gat agt tgt ggg atc gct atc tcc gag agg ccg gtg	1437
Arg Ala Ile Ala Asp Ser Cys Gly Ile Ala Ile Ser Glu Arg Pro Val	
215 220 225	



gct gat tat ggt gtc aag gat att gag cag gat ctg acg agg ttg ttg 1485  
Ala Asp Tyr Gly Val Lys Asp Ile Glu Gln Asp Leu Thr Arg Leu Leu  
230 235 240

agg gat gaa cgg tcc gct ggt acg ctg ccg cag acg gag cta aag ctt 1533  
Arg Asp Glu Arg Ser Ala Gly Thr Leu Pro Gln Thr Glu Leu Lys Leu  
245 250 255

gcg atg ggc tcc gcc tct gcg ttg atc aag tac ctt ggg gtt atg acg 1581  
Ala Met Gly Ser Ala Ser Ala Leu Ile Lys Tyr Leu Gly Val Met Thr  
260 265 270 275

gat cct aca aac ttc gcc cag tac cag ctc tat cag cat gat ttg tcc 1629  
Asp Pro Thr Asn Phe Gly Gln Tyr Gln Leu Tyr Gln His Asp Leu Ser  
280 285 290

cag ttt atg aag ttg gat tcc tcc gcc ctg cgt gct ctt aac ctt atg 1677  
Gln Phe Met Lys Leu Asp Ser Ser Ala Leu Arg Ala Leu Asn Leu Met  
295 300 305

cct ggt ccg cgg gac gga tcc aag tct atg agt ttg ttt ggt ttg ttg 1725  
Pro Gly Pro Arg Asp Gly Ser Lys Ser Met Ser Leu Phe Gly Leu Leu  
310 315 320

aat cac tgc aag acc cct gtt ggt agc cgg ttg ctt gcc cag tgg ctg 1773  
Asn His Cys Lys Thr Pro Val Gly Ser Arg Leu Leu Ala Gln Trp Leu  
325 330 335

aaa cag ccg ttg atg gat ctg gcc gag atc gag aag aga cag cag ctt 1821  
Lys Gln Pro Leu Met Asp Leu Ala Glu Ile Glu Lys Arg Gln Gln Leu  
340 345 350 355

gtt gag gcc ttt gtt gtt aac acg gag ctc aga cag act atg cag gag 1869  
Val Glu Ala Phe Val Val Asn Thr Glu Leu Arg Gln Thr Met Gln Glu  
360 365 370

gag cat ctt ccg tcc ata ccg gat ctg tat aga cta gcc aag cgg ttc 1917  
Glu His Leu Arg Ser Ile Pro Asp Leu Tyr Arg Leu Ala Lys Arg Phe  
375 380 385

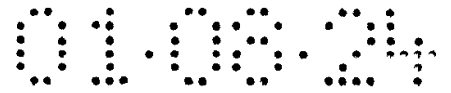
cag cgc aaa cag gca aac ttg gaa gac gtt gtg cgg gtg tac cag gtt 1965  
Gln Arg Lys Gln Ala Asn Leu Glu Asp Val Val Arg Val Tyr Gln Val  
390 395 400

gct att cct ttg cct ggt ttt gtc aac tct ctc gag aat gtt atg gat 2013  
Ala Ile Arg Leu Pro Gly Phe Val Asn Ser Leu Glu Asn Val Met Asp  
405 410 415

gaa gag tat cag acg ccc ctg gag acg gag tat act tcc aac ctc cgg 2061  
Glu Glu Tyr Gln Thr Pro Leu Glu Thr Glu Tyr Thr Ser Asn Leu Arg  
420 425 430 435

agt cac tct gat agc tta gcc aaa ctg gag gag atg gtt gag act acg 2109  
Ser His Ser Asp Ser Leu Ala Lys Leu Glu Glu Met Val Glu Thr Thr  
440 445 450

gtt gac ctt gat gcc ctg gag aac cac gag ttc atc atc aag cct gag 2157  
Val Asp Leu Asp Ala Leu Glu Asn His Glu Phe Ile Ile Lys Pro Glu  
455 460 465



ttt gac gag agt ctg cgg atc atc agg aag aag ctg gac aag ctc cgt 2205  
Phe Asp Glu Ser Leu Arg Ile Ile Arg Lys Lys Leu Asp Lys Leu Arg  
470 475 480

cat gat atg ggc gtt gag cac cgc agg gta gct cgg gac ctt gac caa 2253  
His Asp Met Gly Val Glu His Arg Arg Val Ala Arg Asp Leu Asp Gln  
485 490 495

gat att gag aag aag ttg ttc ctg gag aac cac agg gtg cac gga tgg 2301  
Asp Ile Glu Lys Lys Leu Phe Leu Glu Asn His Arg Val His Gly Trp  
500 505 510 515

tgc ttc cga ctt act cgc aac gag tgc gga tgc atc cgc aat aag aga 2349  
Cys Phe Arg Leu Thr Arg Asn Glu Ser Gly Cys Ile Arg Asn Lys Arg  
520 525 530

gag tac cag gaa tgt tct aca cag aag aac ggt gtc tac ttc act acg 2397  
Glu Tyr Gln Glu Cys Ser Thr Gln Lys Asn Gly Val Tyr Phe Thr Thr  
535 540 545

tcg act atg caa acc ttg cgc cgg gag cat gat caa ctg tcc tcg aac 2445  
Ser Thr Met Gln Thr Leu Arg Arg Glu His Asp Gln Leu Ser Ser Asn  
550 555 560

tac aat aga act cag acc ggc ctg gtg aat gag gtc gtt aac gtt gcc 2493  
Tyr Asn Arg Thr Gln Thr Gly Leu Val Asn Glu Val Val Asn Val Ala  
565 570 575

gcg tcc tac tgt cct gtt ttg gaa cga ctt gcc ggt gtc ata gca cac 2541  
Ala Ser Tyr Cys Pro Val Leu Glu Arg Leu Ala Gly Val Ile Ala His  
580 585 590 595

ctc gat gtc att gta agc ttc gct cat gct tct gtt cat gcg ccg acc 2589  
Leu Asp Val Ile Val Ser Phe Ala His Ala Ser Val His Ala Pro Thr  
600 605 610

ccc tat gct cgg ccc aag atg cac cca cga ggc acc gga aac aca gtt 2637  
Pro Tyr Ala Arg Pro Lys Met His Pro Arg Gly Thr Gly Asn Thr Val  
615 620 625

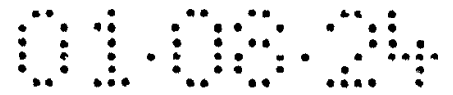
ctc aag gaa gcg cgc cac ccc tgt atg gaa atg cag gat gat att tca 2685  
Leu Lys Glu Ala Arg His Pro Cys Met Glu Met Gln Asp Asp Ile Ser  
630 635 640

ttc att act aat gat gtg gct ttg gtc cga gac gag tcc tcc ttc ctc 2733  
Phe Ile Thr Asn Asp Val Ala Leu Val Arg Asp Glu Ser Ser Phe Leu  
645 650 655

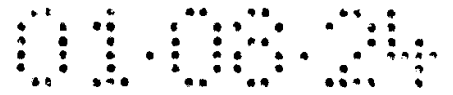
atc att act ggt cct aac atg gga ggt aaa tcg act tat att cgc caa 2781  
Ile Ile Thr Gly Pro Asn Met Gly Gly Lys Ser Thr Tyr Ile Arg Gln  
660 665 670 675

att ggt gtt atc gct ctc atg gct cag acg ggc tgc ttt gtg cct tgt 2829  
Ile Gly Val Ile Ala Leu Met Ala Gln Thr Gly Cys Phe Val Pro Cys  
680 685 690

aca gaa gca gaa ttg acc atc ttt gac tgt atc ctt gca cgt gtt ggt 2877  
Thr Glu Ala Glu Leu Thr Ile Phe Asp Cys Ile Leu Ala Arg Val Gly



695	700	705		
gca agc gat tea cag ctc aag gga gtt tcc act ttc atg gct gag atg 2925 Ala Ser Asp Ser Gln Leu Lys Gly Val Ser Thr Phe Met Ala Glu Met				
	715	720		
ctc gaa aca tcc aac atc ctc aag tcc gca acg tcc gag tct ctt atc 2973 Leu Glu Thr Ser Asn Ile Leu Lys Ser Ala Thr Ser Glu Ser Leu Ile				
	725	730	735	
atc atc gac gag ctt ggg cgc ggt aca agc acg tat gac gga ttc ggc 3021 Ile Ile Asp Glu Leu Gly Arg Gly Thr Ser Thr Tyr Asp Gly Phe Gly				
	740	745	750	755
cta gca tgg gcc atc tct gaa cac atc gtc aca gag att cgt tgc ttc 3069 Leu Ala Trp Ala Ile Ser Glu His Ile Val Thr Glu Ile Arg Cys Phe				
		760	765	770
ggc ctt ttc gct act cac ttc cat gaa ttg aca gct ctc gcc gat cga 3117 Gly Leu Phe Ala Thr His Phe His Glu Leu Thr Ala Leu Ala Asp Arg				
		775	780	785
tac ccc aag ttt gtc aag aac ctg cac gta gtc gcc ttc atc ggc gat 3165 Tyr Pro Lys Ser Val Lys Asn Leu His Val Val Ala Phe Ile Gly Asp				
		790	795	800
ggt act gat gat gac agt gaa gat aag aag tcc aag cgg aac cag gtc 3213 Gly Thr Asp Asp Asp Ser Glu Asp Lys Lys Ser Lys Arg Asn Gln Val				
	805	810	815	
act ctt ctg tac cgg gtc gaa cct ggc att tgt gac cag tca ttc ggt 3261 Thr Leu Leu Tyr Arg Val Glu Pro Gly Ile Cys Asp Gln Ser Phe Gly				
	820	825	830	835
atc cac gtt gcc gaa ttg gtc cgc ttc ccg gag aag gtg gtc aac atg 3309 Ile His Val Ala Glu Leu Val Arg Phe Pro Glu Lys Val Val Asn Met				
		840	845	850
gcc cgc cag aag gca gag gaa ctt gaa gac ttt act tca tcc gaa cag 3357 Ala Arg Gln Lys Ala Glu Glu Leu Glu Asp Phe Thr Ser Ser Glu Gln				
		855	860	865
caa gac cag cag tea tcc atg gcg atc gat aaa tac tcc cag gaa gaa 3405 Gln Asp Gln Gln Ser Ser Met Ala Ile Asp Lys Tyr Ser Gln Glu Glu				
		870	875	880
gtt gag gag gcc agt gcc ctt ctc aaa gcg atg ctg ctg aaa tgg aag 3453 Val Glu Glu Gly Ser Ala Leu Leu Lys Ala Met Leu Leu Lys Trp Lys				
	885	890	895	
tcc gag acc gag tcc tct ggt aag gag ttg aca gtg gaa gag aag cga 3501 Ser Glu Thr Glu Ser Ser Gly Lys Glu Leu Thr Val Glu Glu Lys Arg				
	900	905	910	915
cag atc atg cgt gat ctc gtc aaa gca gat gag aag ctg caa gca aac 3549 Gln Ile Met Arg Asp Leu Val Lys Ala Asp Glu Lys Leu Gln Ala Asn				
		920	925	930
aag gtc ttc cag ggt atc aag gct tta tag attagtattt gcgtcttttt 3599				



Lys Val Phe Gln Gly Ile Lys Ala Leu  
935 940

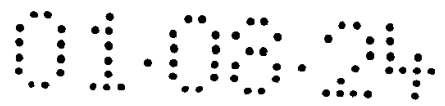
tctttctcgg ggtcatagcg gttcggcggt tggaaggtgt caatctgtgt atgtgtgatc 3659  
tacggacatg aggataaat gtgtaggga taatattatc caaaaatttt cgagtgattg 3719  
cttctttgga catatcgctt aggattaaag tggaagggga gaaatcccat tcaactatat 3779  
cgacataagt cacgtttaga tcgcgagtct agacgctcac cggg 3823

<210> 2  
<211> 940  
<212> PRT  
<213> 米曲霉

<400> 2  
Met Ser Ser Arg Pro Glu Leu Lys Val Asp Asp Glu Val Gly Phe Ile  
1 5 10 15  
Arg Phe Tyr Arg Ser Leu Ala Ala Asn Ser Asn Asp Glu Thr Ile Arg  
20 25 30  
Val Phe Asp Arg Gly Asp Trp Tyr Ser Ala His Gly Ala Lys Ala Glu  
35 40 45  
Phe Ile Ala Arg Thr Val Tyr Lys Thr Thr Ser Ile Leu Arg Asn Leu  
50 55 60  
Gly Arg Ser Asp Ser Gly Gly Leu Pro Ser Val Thr Met Ser Val Thr  
65 70 75 80  
Val Phe Arg Asn Phe Leu Arg Glu Ala Leu Phe Arg Leu Asn Lys Arg  
85 90 95  
Ile Glu Ile Trp Gly Ser Val Gly Thr Gly Lys Gly His Trp Lys Leu  
100 105 110  
Val Lys Gln Ala Ser Pro Gly Asn Leu Gln Asp Val Glu Glu Glu Leu  
115 120 125  
Gly Ser Val Gly Gly Leu Ser Met Asp Ser Ala Pro Ile Ile Leu Ala  
130 135 140  
Val Lys Ile Ser Ala Lys Ala Ala Glu Ala Arg Ser Val Gly Val Cys  
145 150 155 160  
Phe Ala Asp Ala Ser Val Arg Glu Leu Gly Val Ser Glu Phe Leu Asp  
165 170 175  
Asn Asp Ile Tyr Ser Asn Phe Glu Ser Leu Ile Ile Gln Leu Gly Val  
180 185 190  
Lys Glu Cys Leu Val Gln Met Asp Ala Asn Lys Lys Asp Val Glu Leu  
195 200 205  
Gly Lys Ile Arg Ala Ile Ala Asp Ser Cys Gly Ile Ala Ile Ser Glu  
210 215 220  
Arg Pro Val Ala Asp Tyr Gly Val Lys Asp Ile Glu Gln Asp Leu Thr  
225 230 235 240  
Arg Leu Leu Arg Asp Glu Arg Ser Ala Gly Thr Leu Pro Gln Thr Glu  
245 250 255  
Leu Lys Leu Ala Met Gly Ser Ala Ser Ala Leu Ile Lys Tyr Leu Gly  
260 265 270  
Val Met Thr Asp Pro Thr Asn Phe Gly Gln Tyr Gln Leu Tyr Gln His  
275 280 285  
Asp Leu Ser Gln Phe Met Lys Leu Asp Ser Ser Ala Leu Arg Ala Leu  
290 295 300  
Asn Leu Met Pro Gly Pro Arg Asp Gly Ser Lys Ser Met Ser Leu Phe  
305 310 315 320  
Gly Leu Leu Asn His Cys Lys Thr Pro Val Gly Ser Arg Leu Leu Ala



325 330 335  
Gln Trp Leu Lys Gln Pro Leu Met Asp Leu Ala Glu Ile Glu Lys Arg  
340 345 350  
Gln Gln Leu Val Glu Ala Phe Val Val Asn Thr Glu Leu Arg Gln Thr  
355 360 365  
Met Gln Glu Glu His Leu Arg Ser Ile Pro Asp Leu Tyr Arg Leu Ala  
370 375 380  
Lys Arg Phe Gln Arg Lys Gln Ala Asn Leu Glu Asp Val Val Arg Val  
385 390 395 400  
Tyr Gln Val Ala Ile Arg Leu Pro Gly Phe Val Asn Ser Leu Glu Asn  
405 410 415  
Val Met Asp Glu Glu Tyr Gln Thr Pro Leu Glu Thr Glu Tyr Thr Ser  
420 425 430  
Asn Leu Arg Ser His Ser Asp Ser Leu Ala Lys Leu Glu Glu Met Val  
435 440 445  
Glu Thr Thr Val Asp Leu Asp Ala Leu Glu Asn His Glu Phe Ile Ile  
450 455 460  
Lys Pro Glu Phe Asp Glu Ser Leu Arg Ile Ile Arg Lys Lys Leu Asp  
465 470 475 480  
Lys Leu Arg His Asp Met Gly Val Glu His Arg Arg Val Ala Arg Asp  
485 490 495  
Leu Asp Gln Asp Ile Glu Lys Lys Leu Phe Leu Glu Asn His Arg Val  
500 505 510  
His Gly Trp Cys Phe Arg Leu Thr Arg Asn Glu Ser Gly Cys Ile Arg  
515 520 525  
Asn Lys Arg Glu Tyr Gln Glu Cys Ser Thr Gln Lys Asn Gly Val Tyr  
530 535 540  
Phe Thr Thr Ser Thr Met Gln Thr Leu Arg Arg Glu His Asp Gln Leu  
545 550 555 560  
Ser Ser Asn Tyr Asn Arg Thr Gln Thr Gly Leu Val Asn Glu Val Val  
565 570 575  
Asn Val Ala Ala Ser Tyr Cys Pro Val Leu Glu Arg Leu Ala Gly Val  
580 585 590  
Ile Ala His Leu Asp Val Ile Val Ser Phe Ala His Ala Ser Val His  
595 600 605  
Ala Pro Thr Pro Tyr Ala Arg Pro Lys Met His Pro Arg Gly Thr Gly  
610 615 620  
Asn Thr Val Leu Lys Glu Ala Arg His Pro Cys Met Glu Met Gln Asp  
625 630 635 640  
Asp Ile Ser Phe Ile Thr Asn Asp Val Ala Leu Val Arg Asp Glu Ser  
645 650 655  
Ser Phe Leu Ile Ile Thr Gly Pro Asn Met Gly Gly Lys Ser Thr Tyr  
660 665 670  
Ile Arg Gln Ile Gly Val Ile Ala Leu Met Ala Gln Thr Gly Cys Phe  
675 680 685  
Val Pro Cys Thr Glu Ala Glu Leu Thr Ile Phe Asp Cys Ile Leu Ala  
690 695 700  
Arg Val Gly Ala Ser Asp Ser Gln Leu Lys Gly Val Ser Thr Phe Met  
705 710 715 720  
Ala Glu Met Leu Glu Thr Ser Asn Ile Leu Lys Ser Ala Thr Ser Glu  
725 730 735  
Ser Leu Ile Ile Ile Asp Glu Leu Gly Arg Gly Thr Ser Thr Tyr Asp  
740 745 750  
Gly Phe Gly Leu Ala Trp Ala Ile Ser Glu His Ile Val Thr Glu Ile  
755 760 765  
Arg Cys Phe Gly Leu Phe Ala Thr His Phe His Glu Leu Thr Ala Leu  
770 775 780  
Ala Asp Arg Tyr Pro Lys Ser Val Lys Asn Leu His Val Val Ala Phe  
785 790 795 800



```

Ile Gly Asp Gly Thr Asp Asp Asp Ser Glu Asp Lys Lys Ser Lys Arg
      805      810      815
Asn Gln Val Thr Leu Leu Tyr Arg Val Glu Pro Gly Ile Cys Asp Gln
      820      825      830
Ser Phe Gly Ile His Val Ala Glu Leu Val Arg Phe Pro Glu Lys Val
      835      840      845
Val Asn Met Ala Arg Gln Lys Ala Glu Glu Leu Glu Asp Phe Thr Ser
      850      855      860
Ser Glu Gln Gln Asp Gln Gln Ser Ser Met Ala Ile Asp Lys Tyr Ser
      865      870      875
Gln Glu Glu Val Glu Gln Gly Ser Ala Leu Leu Lys Ala Met Leu Leu
      885      890      895
Lys Trp Lys Ser Gln Thr Glu Ser Ser Gly Lys Glu Leu Thr Val Glu
      900      905      910
Glu Lys Arg Gln Ile Met Arg Asp Leu Val Lys Ala Asp Glu Lys Leu
      915      920      925
Gln Ala Asn Lys Val Phe Gln Gly Ile Lys Ala Leu
      930      935      940

```

<210> 3  
 <211> 38  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 人工序列的说明：寡聚# U

<400> 3  
 gggaaagctgc caggccccag tgtcagcctc ctatgctc

38

<210> 4  
 <211> 38  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 人工序列的说明：寡聚# L-G. T

<400> 4  
 gagcatagga ggctgacatt ggggacctggc agcttccc

38

<210> 5  
 <211> 38  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 人工序列的说明：寡聚# L-G. A

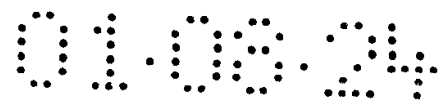
<400> 5  
 gagcatagga ggctgacaat ggggacctggc agcttccc

38

<210> 6

<211> 38  
 <212> DNA  
 <213> 人工序列  
  
 <220>  
 <223> 人工序列的说明：寡聚# L-G.G.  
  
 <400> 6  
 gagcatagga ggctgacagt ggggcctggc agcttccc 38  
  
 <210> 7  
 <211> 38  
 <212> DNA  
 <213> 人工序列  
  
 <220>  
 <223> 人工序列的说明：寡聚# L-A.C.  
  
 <400> 7  
 gagcatagga ggctgacacc ggggcctggc agcttccc 38  
  
 <210> 8  
 <211> 40  
 <212> DNA  
 <213> 人工序列  
  
 <220>  
 <223> 人工序列的说明：寡聚# L-TG  
  
 <400> 8  
 gagcatagga ggctgacact gtgggcctggc gcagcttccc 40  
  
 <210> 9  
 <211> 38  
 <212> DNA  
 <213> 人工序列  
  
 <220>  
 <223> 人工序列的说明：寡聚# L-HOM  
  
 <400> 9  
 gagcatagga ggctgacacc ggggcctggc agcttccc 38  
  
 <210> 10  
 <211> 24  
 <212> DNA  
 <213> 人工序列  
  
 <220>  
 <223> 人工序列的说明：寡聚# PR 117858  
  
 <400> 10





ggcncarath ggntgyttyg tncc 24

<210> 11  
<211> 20  
<212> DNA  
<213> 人工序列

<220>  
<223> 人工序列的说明：寡聚# Pr 117859

<400> 11  
gcccangena rnceraancc 20

<210> 12  
<211> 20  
<212> DNA  
<213> 人工序列

<220>  
<223> 人工序列的说明：寡聚# 130740

<400> 12  
gctcgaaaca tccaacatcc 20

<210> 13  
<211> 19  
<212> DNA  
<213> 人工序列

<220>  
<223> 人工序列的说明：寡聚# 130741

<400> 13  
gctgtgaate acttgcacc 19

<210> 14  
<211> 25  
<212> DNA  
<213> 人工序列

<220>  
<223> 人工序列的说明：寡聚# 131928

<400> 14  
cttcataaac tgcgacaaat catgc 25

<210> 15  
<211> 16  
<212> DNA  
<213> 人工序列

<220>  
<223> 人工序列的说明：寡聚# 131929

<400> 15  
 ggaggagcat cttcgc 16

<210> 16  
 <211> 25  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 人工序列的说明：寡聚# 131930

<400> 16  
 ggaacttgaa gactttactt catcc 25

<210> 17  
 <211> 19  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 人工序列的说明：寡聚# 134608

<400> 17  
 ccagaaactc gctaacacc 19

<210> 18  
 <211> 15  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 人工序列的说明：寡聚# 134609

<400> 18  
 gtgctttgcg gacgc 15

<210> 19  
 <211> 16  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 人工序列的说明：寡聚# 134610

<400> 19  
 caggacagta ggacgc 16

<210> 20  
 <211> 17  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 人工序列的说明：寡聚# 135320  
  
 <400> 20  
 cgagcgatga actccgc 17  
  
 <210> 21  
 <211> 17  
 <212> DNA  
 <213> 人工序列  
  
 <220>  
 <223> 人工序列的说明：寡聚# 135321  
  
 <400> 21  
 gcgttggtgg attatcc 17  
  
 <210> 22  
 <211> 20  
 <212> DNA  
 <213> 人工序列  
  
 <220>  
 <223> 人工序列的说明：寡聚# 136105  
  
 <400> 22  
 cgttgcatct atcatatacc 20  
  
 <210> 23  
 <211> 21  
 <212> DNA  
 <213> 人工序列  
  
 <220>  
 <223> 人工序列的说明：寡聚# 136106  
  
 <400> 23  
 ggtatatgat agatgcaacg c 21  
  
 <210> 24  
 <211> 24  
 <212> DNA  
 <213> 人工序列  
  
 <220>  
 <223> 人工序列的说明：寡聚# 137208  
  
 <400> 24  
 ccgctctcc aacaagatga atgg 24  
  
 <210> 25  
 <211> 21

<212> DNA  
 <213> 人工序列  
 <220>  
 <223> 人工序列的说明：寡聚# 137207  
 <400> 25  
 ccgctttctc ggggcatag c 21

<210> 26  
 <211> 22  
 <212> DNA  
 <213> 人工序列  
 <220>  
 <223> 人工序列的说明：引物 138149  
 <400> 26  
 cctttccact ttaattctaa gc 22

<210> 27  
 <211> 35  
 <212> DNA  
 <213> 人工序列  
 <220>  
 <223> 人工序列的说明：寡聚# 000120j2  
 <400> 27  
 tctgcgaatc gcttggatcc cgaacgcgac aacac 35

<210> 28  
 <211> 33  
 <212> DNA  
 <213> 人工序列  
 <220>  
 <223> 人工序列的说明：寡聚# 000120j4  
 <400> 28  
 gagctcagat ctcttaagggt ctggacgaga aga 33

<210> 29  
 <211> 35  
 <212> DNA  
 <213> 人工序列  
 <220>  
 <223> 人工序列的说明：寡聚# 000120j3  
 <400> 29  
 gttgtcgcgt ccggatcca agcgattcgc agaag 35

<210> 30  
<211> 33  
<212> DNA  
<213> 人工序列

<220>  
<223> 人工序列的说明：寡聚# 1298-TAKA

<400> 30  
gcaagcgcgc gcaatcatg gtgtttgat cat

33