



[12] 发明专利申请公开说明书

[21] 申请号 02810179.0

[43] 公开日 2004 年 6 月 30 日

[11] 公开号 CN 1509330A

[22] 申请日 2002.5.17 [21] 申请号 02810179.0
[30] 优先权
[32] 2001.5.18 [33] DK [31] PA200100798
[86] 国际申请 PCT/DK2002/000325 2002.5.17
[87] 国际公布 WO2002/095014 英 2002.11.28
[85] 进入国家阶段日期 2003.11.18
[71] 申请人 诺和酶股份有限公司
地址 丹麦拜格斯威尔德
[72] 发明人 M·许莱因 J·莱姆贝克

[74] 专利代理机构 北京市中咨律师事务所
代理人 黄革生 林柏楠

权利要求书 4 页 说明书 46 页 序列表 8 页

[54] 发明名称 具有纤维二糖酶活性的多肽和编码
其的多核苷酸

[57] 摘要

本发明涉及具有纤维二糖酶活性的多肽和具有
编码该多肽的核苷酸序列的多核苷酸。 本发明还
涉及核酸构建体、载体和包含该核酸构建体的宿主
细胞以及制备和使用该多肽的方法。

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

1、具纤维二糖酶活性的多肽，其选自：

(a) 包含这样的氨基酸序列的多肽，此氨基酸序列与 SEQ ID NO: 2 中 1 至 842 位氨基酸所示序列具有至少 80% 同一性或者与 SEQ ID NO: 2 中 1 至 351 位氨基酸所示序列具有至少 90% 同一性；

(b) 包含这样的氨基酸序列的多肽，此氨基酸序列与插在大肠杆菌 DSM 14240 的质粒中的核苷酸序列的纤维二糖酶编码部分所编码的多肽具有至少 80% 同一性；

(c) 由能够在中度严谨条件下与多核苷酸探针杂交的核苷酸序列编码的多肽，其中多核苷酸探针选自

(i) SEQ ID NO: 1 中 87 至 2612 位核苷酸的互补链，和 (ii) SEQ ID NO: 1 中 87 至 1139 位核苷酸的互补链；

(d) 具有纤维二糖酶活性的 (a)、(b) 或 (c) 的片断。

2、如权利要求 1 所述的多肽，其包含与 SEQ ID NO: 2 中 1 至 842 位氨基酸所示序列具有至少 85% 同一性，优选与 SEQ ID NO: 2 中 1 至 842 位氨基酸所示序列具有至少 90% 同一性、至少 95% 同一性或至少 99% 同一性的氨基酸序列。

3、如权利要求 2 所述的多肽，其包含 SEQ ID NO: 2 中 1 至 842 位氨基酸。

4、如权利要求 1-3 任一项所述的多肽，其由 SEQ ID NO: 2 中 1 至 842 位氨基酸组成。

5、如权利要求 1-3 任一项所述的多肽，其中所述多肽是含有这样的氨基酸序列的人工变体，此氨基酸序列与 SEQ ID NO: 2 中 1 至 842 位氨基酸所示序列相比具有至少一个氨基酸替代、缺失和/或插入。

6、如权利要求 1 所述的多肽，其包含这样的氨基酸序列，该序列与插在大肠杆菌 DSM 14240 的质粒内的核苷酸序列的纤维二糖酶编码部分所编码的多肽具有至少 85% 同一性，优选与插在大肠杆菌 DSM 14240 的

质粒内的核苷酸序列的纤维二糖酶编码部分所编码的多肽具有至少 90% 同一性、至少 95% 同一性 或至少 99% 同一性。

7、如权利要求 6 所述的多肽，其包含插在大肠杆菌 DSM 14240 的质粒内的核苷酸序列的纤维二糖酶编码部分所编码的氨基酸序列。

8、如权利要求 6 或 7 所述的多肽，其由插在大肠杆菌 DSM 14240 的质粒内的核苷酸序列的纤维二糖酶编码部分所编码的氨基酸序列组成。

9、如权利要求 6 或 7 所述的多肽，其中所述多肽是含有这样的氨基酸序列的人工变体，此氨基酸序列与插在大肠杆菌 DSM 14240 的质粒内的核苷酸序列的纤维二糖酶编码部分所编码的氨基酸序列相比有至少一个替代、缺失和/或插入的氨基酸。

10、如权利要求 1 所述的多肽，其是由在中度严紧条件下，优选在高度严紧条件下可与多核苷酸探针进行杂交的核苷酸序列编码的，所述多核苷酸探针选自：

(i) SEQ ID NO: 1 中核苷酸 87 至 2612 位的互补链，和

(ii) SEQ ID NO: 1 中核苷酸 87 至 1139 位的互补链。

11、如权利要求 10 所述的多肽，其是由在中度严紧条件下，优选在高度严紧条件下可与多核苷酸探针进行杂交的核苷酸序列编码的，所述多核苷酸探针为 SEQ ID NO: 1 中核苷酸 87 至 2612 位的互补链。

12、具有编码如权利要求 1-11 中任一项所述的多肽的核苷酸序列的多核苷酸。

13、含有如权利要求 12 所述的核苷酸序列的核酸构建体，其中所述核苷酸序列可操作地与一个或多个能够指导多肽在适宜宿主中产生的调控序列连接。

14、含有如权利要求 13 所述的核酸构建体的重组表达载体。

15、含有如权利要求 13 所述的核酸构建体的重组宿主细胞。

16、生产如权利要求 1-11 任一项所述的多肽的方法，该方法包括：

(a) 培养野生型能够产生所述多肽的菌株以产生所述多肽；以及

(b) 回收此多肽。

17、生产如权利要求 1-11 任一项所述的多肽的方法，该方法包括：

(a) 在有利于所述多肽产生的条件下培养如权利要求 16 所述的重组宿主细胞；以及

(b) 回收此多肽。

18、多核苷酸，其具有与 SEQ ID NO: 1 中核苷酸 87 至 2612 位所示序列至少 85 % 相同的核苷酸序列。

19、多核苷酸，其具有与插在大肠杆菌 DSM 14240 的质粒内的核苷酸序列的纤维二糖酶编码部分至少 85 % 相同的核苷酸序列。

20、含有以下核苷酸序列的多核苷酸，此核苷酸序列编码具纤维二糖酶活性的多肽，并且在中度严谨条件下能与选自(i)SEQ ID NO: 1 中 87 至 2612 位核苷酸的互补链、和 (ii) SEQ ID NO: 1 中 87 至 1139 位核苷酸的互补链的多核苷酸探针杂交。

21、含有修饰的核苷酸序列的多核苷酸，其中所述修饰的核苷酸序列在 SEQ ID NO: 1 的成熟多肽编码区中包含至少一个修饰，并且所述修饰的核苷酸序列编码由 SEQ ID NO: 2 中 1 至 842 位氨基酸组成的多肽。

22、含有如权利要求 1-11 任一项所述的多肽和表面活性剂的组合物。

23、用于从生物量生产乙醇的方法，其包括使生物量与如权利要求 1-11 任一项所述的多肽接触。

24、如权利要求 1-11 任一项所述的多肽在乙醇生产中的用途。

25、DNA 改组方法，其包括利用如权利要求 12 和 18-21 任一项所述的多核苷酸。

26、DNA 改组方法，其包括利用 SEQ ID NO: 1 的 87 至 1139 位核苷酸。

27、编码具纤维二糖酶活性的多肽的多核苷酸，其可利用如权利要求 25 或 26 所述的方法获得。

28、由如权利要求 27 所述的多核苷酸编码的具纤维二糖酶活性的多肽。

29、如权利要求 12 和 18 - 21 任一项所述的多核苷酸在 DNA 改组中的用途。

30、SEQ ID NO: 1 中 87 至 1139 位核苷酸在 DNA 改组中的用途。

31、转基因植物、植物部分或植物细胞，其已经被编码如权利要求 1 - 11 任一项所述的具纤维二糖酶活性的多肽的核苷酸序列转化。

具有纤维二糖酶活性的多肽和编码其的多核苷酸

发明领域

本发明涉及具有纤维二糖酶（亦称 β -葡糖苷酶）活性的多肽和具有编码本多肽的核苷酸序列的多核苷酸。本发明还涉及核酸构建体、载体和含本核酸构建体的宿主细胞，以及制备和使用本多肽的方法。

发明背景

纤维二糖酶是重要的降解生物量的酶。纤维二糖酶的这种用途是由植物材料生产乙醇过程中的一个关键步骤。考虑到环境因素，乙醇是一种吸引人的石油燃料代用品。纤维二糖酶的其他用途包括在果汁业中用于还原苦味化合物—如栎精。

本发明的一个目的是提供具有纤维二糖酶（亦称 β -葡糖苷酶）活性的多肽和编码本多肽的核苷酸序列。

发明概述

本发明第一方面涉及具有纤维二糖酶活性的多肽，其选自：

- (a) 包含这样一个氨基酸序列的多肽，此氨基酸序列与 SEQ ID NO: 2 中 1 至 842 位氨基酸所示序列具有至少 80% 同一性或者与 SEQ ID NO: 2 中 1 至 351 位氨基酸所示序列具有至少 90% 同一性；
- (b) 包含这样一个氨基酸序列的多肽，此氨基酸序列与插入大肠杆菌 (*E. coli*) DSM 14240 的质粒中的核苷酸序列的编码部分所编码的多肽具有至少 80% 同一性；
- (c) 由能够在中度严谨条件下与多核苷酸探针杂交的核苷酸序列编码的多肽，其中多核苷酸探针选自 (i) SEQ ID NO: 1 中 87 至 2612 位核苷

酸所示序列的互补链, 和 (ii) SEQ ID NO: 1 中 87 至 1139 位核苷酸所示序列的互补链;

(d) 具有纤维二糖酶活性的 (a)、(b) 或 (c) 的片断。

本发明第二方面涉及多核苷酸, 它具有能够编码本发明多肽的核苷酸序列。

本发明第三方面涉及核酸构建体, 它包含能够编码本发明多肽的核苷酸序列, 且该序列可操作地连接一个或多个指导此多肽在适当宿主中产生的调控序列。

本发明第四方面涉及包含本发明核酸构建体的重组表达载体。

本发明第五方面涉及包含本发明核酸构建体的重组宿主细胞。

本发明第六方面涉及用于产生本发明多肽的方法, 其包括:

- (a) 培养野生型能够生产此多肽的菌株以产生此多肽; 以及
- (b) 回收此多肽。

本发明第七方面涉及用于产生本发明多肽的方法, 其包括:

- (a) 在有助于本多肽产生的条件下培养重组宿主细胞; 以及
- (b) 回收本多肽。

本发明的其它方面可从下面的说明和所附的权利要求书中明显看出。

定义

进一步详细讨论本发明之前, 首先定义下列术语和常规用语。

基本上纯的多肽: 在本发明上下文中, 术语“基本上纯的多肽”指的是多肽制备物中包含至多 10% 重量的其它天然相关的多肽物质 (优选更低百分比——如至多 8% 重量, 至多 6% 重量, 至多 5% 重量, 至多 4% 重量, 至多 3% 重量, 至多 2% 重量, 至多 1% 重量和至多 1/2% 重量的其它多肽物质)。因此, 优选的是基本上纯的多肽具有至少 92% 的纯度, 即此多肽组成了制备物中总多肽物质重量的至少 92%, 优选更高的百分比纯度, 如至少 94% 纯, 至少 95% 纯, 至少 96% 纯, 至少 97% 纯, 至少 98% 纯, 至少 99% 纯, 以及至多 99.5% 纯。本文公开的多肽优选是基本上纯的形式。尤其, 优选的是本

文公开的多肽是“实质上的纯形式”，即多肽制备物中实质上没有其它天然相关多肽物质。举例来说，这种纯度可通过熟知的重组法制备多肽来得到。在本文中，术语“基本上纯的多肽”与术语“分离的多肽”和“分离形式的多肽”同义。

纤维二糖酶活性：术语“纤维二糖酶活性”在本文中被定义为 β -葡萄糖苷酶活性，如在酶分类 EC3.2.1.21 中的定义，它催化末端的非还原性 β -D-葡萄糖残基水解释放出 β -D-葡萄糖。对本发明来说，根据实施例部分中“纤维二糖酶试验”描述的步骤测定纤维二糖酶活性。一个单位的纤维二糖酶活性定义为在 40℃，pH5 下每分钟产生 2 μ mol 葡萄糖。

本发明多肽应优选具有由 SEQ ID NO: 2 中 1 至 842 位氨基酸所示氨基酸序列组成的多肽的纤维二糖酶活性的至少 20%。在特别优选的实施方案中，多肽应具有由 SEQ ID NO: 2 中 1 至 842 位氨基酸所示氨基酸序列组成的多肽的纤维二糖酶活性的至少 40%，如至少 50%，优选至少 60%，如至少 70%，更优选至少 80%，如至少 90%，最优选至少 95%，如大约或至少 100%。

同一性：在本发明上下文中，两个氨基酸序列之间或两个核苷酸序列之间的同源性通过参量“同一性”来描述。

对本发明来说，两个氨基酸序列之间的同一性程度通过可用于蛋白质和 DNA 序列比对 (align) 的 Needleman - Wunsch 序列比对法进行确定。对于蛋白质比对，使用的默认评分矩阵是 BLOSUM50，缺口 (gap) 中第一个残基的罚分是 -12，而缺口中其它残基的罚分是 -2。比对可利用 FASTA 程序包 v20u6 版 (W.R.Pearson 和 D.J.Lipman (1988)， “生物序列分析的改良工具”，PNAS 85: 2444 - 2448; 和 W.R.Pearson (1990) “利用 FASTP 和 FASTA 进行快速灵敏的序列比较”，酶学方法 (Methods in Enzymology), 183:63-98) 的比对软件来执行。

两个核苷酸序列之间的同一性程度可使用同一性矩阵为默认评分矩阵利用上述的相同计算法则和软件包进行确定。缺口中第一个残基的罚分是 -16，而缺口中其它残基的罚分是 -4。

片断：本文中，SEQ ID NO: 2 的“片断”是指从这个氨基酸序列的氨

基和/或羧基末端缺失了一个或多个氨基酸的多肽。片断优选包含至少 351 个氨基酸残基，例如 SEQ ID NO: 2 中的 1 至 351 位氨基酸。

等位变体：在本发明上下文中，术语“等位变体”表示占据同一染色体座位的一个基因的两种或多种可变形式中的任一种。等位变异可以由突变天然地引起，并可引起种群内多态性。基因突变可能是沉默突变（所编码的多肽没有变化）或者可能编码氨基酸序列发生改变的多肽。多肽等位变体是由基因等位变体所编码的多肽。

基本上纯的多核苷酸：本文使用的术语“基本上纯的多核苷酸”指的是一种多核苷酸制备物，其中此多核苷酸已经离开其天然遗传环境，从而它不含有其它外来或不想要的编码序列并且它存在的形式适于在遗传工程蛋白质生产系统内进行使用。因此，基本上纯的多核苷酸包含至多 10% 重量的其它天然相关多核苷酸物质（优选更低百分比，如至多 8% 重量，至多 6% 重量，至多 5% 重量，至多 4% 重量，至多 3% 重量，至多 2% 重量，至多 1% 重量和至多 1/2% 重量的其它多核苷酸物质）。然而，基本上纯的多核苷酸可以包括天然的 5' 和 3' 非翻译区，如启动子和终止子。优选的是基本上纯的多核苷酸具有至少 92% 的纯度，即此多核苷酸组成了制备物中总多核苷酸物质重量的至少 92%，优选更高的百分比纯度如至少 94% 纯，至少 95% 纯，至少 96% 纯，至少 97% 纯，至少 98% 纯，至少 99% 纯，以及至多 99.5% 纯。本文所公开的多核苷酸优选为基本上纯的形式。尤其，优选的是本文所公开的多核苷酸是“实质上纯的形式”，即多核苷酸制备物中实质上没有任何其它天然相关的多核苷酸物质。在本文中，术语“基本上纯的多核苷酸”与术语“分离的核苷酸”和“分离形式的多核苷酸”同义。

修饰：在本发明上下文中术语“修饰”指的是对由 SEQ ID NO: 2 中 1 至 842 位氨基酸所示氨基酸序列组成的多肽所进行的任何化学修饰，以及对编码此多肽的 DNA 所进行的遗传操作。修饰可以是在目的氨基酸内或其所在位置处进行的氨基酸侧链置换，以及氨基酸替代、缺失和/或插入。

人工变体：术语“人工变体”用于本文指的是具有纤维二糖酶活性的多肽，该多肽由表达修饰基因（对比 SEQ ID NO: 1 而言）的生物体产生。

修饰基因，即，在适当宿主中表达时产生所述变体的基因是经过人为干预对 SEQ ID NO: 1 公开的核苷酸序列进行修饰得到的。

cDNA: 术语“cDNA”用于本发明上下文意在覆盖这样的 DNA 分子，此 DNA 分子能够从成熟的、剪接的、真核细胞来源的 mRNA 分子的反转录制备得到。cDNA 缺少通常在相应遗传组 DNA 中存在的内含子序列。最初的初级 RNA 转录物是 mRNA 的前体，它经过一系列加工事件才成为成熟剪接的 mRNA。这些事件包括通过剪接过程去除内含子序列。cDNA 由于衍生自 mRNA 因此缺少内含子序列。

核酸构建体: 术语“核酸构建体”用于本文，既可以指单链的也可以指双链的核酸分子，它分离自天然基因或者它经修饰以非天然方式含有核酸片断。如果核酸构建体包含表达本发明编码序列所需的调控序列，则术语核酸构建体与术语“表达序列盒”同义。

调控序列: 术语“调控序列”在本文被定义为包括所有的对表达本发明多肽必需或有利的组分。对编码多肽的核苷酸序列来说每个调控序列可以是天然的或外来的。这些调控序列包括，但不限于：前导序列、聚腺苷酸化序列、前肽序列、启动子、信号肽序列和转录终止子。最少地调控序列包括启动子及转录和翻译终止信号。为引入特异性限制位点以便于调控序列与编码多肽的核苷酸序列的编码区进行连接，可以提供带接头的调控序列。

可操作地连接: 术语“可操作地连接”在本文定义为一种构型，在此构型中调控序列被置于相对于 DNA 序列的编码序列而言适当的位置，这样调控序列可指导多肽表达。

编码序列: 本文使用的术语“编码序列”意在覆盖这样的核苷酸序列，该序列直接规定了其蛋白质产物的氨基酸序列。编码序列的边界一般由开放阅读框确定，通常开始于 AGT 起始密码子。编码序列一般包括 DNA、cDNA、和重组核苷酸序列。

表达: 在本发明上下文中，术语“表达”包括多肽合成中涉及的任何步骤，其包括但不限于：转录，转录后修饰，翻译，翻译后修饰及分泌。

表达载体: 在本发明上下文中，术语“表达载体”包括这样的线状的

或环状的 DNA 分子，它包含编码本发明多肽的片断，而该片段与为其转录提供条件的其它片断可操作地连接。

宿主细胞:用于本文的术语“宿主细胞”包括任一种易被核酸构建体转化的细胞。

术语“多核苷酸探针”、“杂交”及各种严紧条件在标题为“具有纤维二糖酶活性的多肽”一节中进行定义。

发明详述

具有纤维二糖酶活性的多肽

在第一实施方案中，本发明涉及具有纤维二糖酶活性的多肽，此多肽包含这样的氨基酸序列，并优选由这样的氨基酸序列组成，此氨基酸序列与 SEQ ID NO: 2 中 1 至 842 位氨基酸（即成熟多肽）所示序列的同一性程度为至少 65%，优选至少 70%，如至少 75%，更优选至少 80%，如至少 85%，甚至更优选至少 90%，最优选至少 95%，如至 96%，至少 97%，甚至最优选至少 98%，如至 99%（以下称“同源多肽”）。在有意义的一个实施方案中，此氨基酸序列有至多十个氨基酸（例如，十个氨基酸），尤其是至多五个氨基酸（例如，五个氨基酸），如至多四个氨基酸（例如，四个氨基酸），举例来说至多三个氨基酸（例如，三个氨基酸）与 SEQ ID NO: 2 中 1 至 842 位氨基酸所示序列不同。在一个特别有意义的实施方案中，此氨基酸序列有至多两个氨基酸（例如，两个氨基酸），如一个氨基酸与 SEQ ID NO: 2 中 1 至 842 位氨基酸所示序列不同。

在另一实施方案中多肽包含这样的氨基酸序列，并优选由这样的氨基酸序列组成，此氨基酸序列与 SEQ ID NO: 2 中 1 至 351 位氨基酸（即催化核心）所示序列的同一性程度为至少 65%，优选至少 70%，如至少 75%，更优选至少 80%，如至少 85%，甚至更优选至少 90%，最优选至少 95%，如至少 96%，如至少 97%，甚至最优选至少 98%，如至 99%。

利用上文（参见“定义”一节）所述方法，对由 SEQ ID NO: 2 中 1 至 842 位氨基酸所示氨基酸序列组成的多肽与最新的现有技术进行序列比对，

得到下列同一性百分数:

棘孢曲霉 (*Aspergillus aculeatus*): 79.6 %

黑曲霉 (*Aspergillus niger*): 76.9 %

川地曲霉 (*Aspergillus kawasachi*): 77.0 %

对由 SEQ ID NO: 2 中 1 至 351 位氨基酸 (即催化核心) 所示氨基酸序列组成的多肽与本领域最新现有技术进行序列比对, 得到下列同一性百分数:

棘孢曲霉: 86.6 %

黑曲霉: 81.9 %

川地曲霉: 80.3 %

构巢曲霉 (*Aspergillus nidulans*): 25.1 %

优选地是, 本发明多肽包含 SEQ ID NO: 2 所示的氨基酸序列; 其等位变体; 或者其具有纤维二糖酶活性的片断。在另一优选实施方案中, 本发明多肽包含 SEQ ID NO: 2 中 1 至 842 位氨基酸。在另一优选实施方案中, 此多肽由 SEQ ID NO: 2 中 1 至 842 位氨基酸组成。

本发明多肽可以是天然源中鉴定分离出来的野生型纤维二糖酶。这种野生型多肽可通过本领域已知的标准技术进行特异性地筛选。此外, 本发明多肽可利用 DNA 改组技术 (如在 J.E.Ness 等, *Nature Biotechnology* 17, 893 - 896 (1999) 一文中的描述) 来制备。而且, 本发明多肽可以是人工变体, 它包含这样的氨基酸序列并优选由这样的氨基酸序列组成, 此氨基酸序列同 SEQ ID NO: 2 中 1 至 842 位氨基酸所示序列相比, 有至少一个替代、缺失和/或插入的氨基酸。这种人工变体可利用本领域已知的标准技术, 如通过对含有 SEQ ID NO: 2 中 1 至 842 位氨基酸所示氨基酸序列的多肽进行定点/随机诱变来构建。在本发明一个实施方案中, 氨基酸变化 (在人工变体和野生型多肽中) 是性质较小的, 即, 不会明显影响蛋白质折叠和/活性的保守性氨基酸替代; 小缺失, 一般有 1 至大约 30 个氨基酸; 小的氨基或羧基末端延伸, 如氨基末端的甲硫氨酸残基; 不超过大约 20 - 25 个残基的小接头肽; 或者可以通过改变净电荷或另一功能而利于纯化的小延伸, 如聚组氨酸序列段、抗原表位或结合域。

保守性替代的实例为碱性氨基酸组(精氨酸, 赖氨酸和组氨酸), 酸性氨基酸组(谷氨酸和天冬氨酸), 极性氨基酸组(谷氨酰胺和天冬酰胺), 疏水氨基酸组(亮氨酸, 异亮氨酸, 缬氨酸和甲硫氨酸), 芳香族氨基酸组(苯丙氨酸, 色氨酸和酪氨酸), 和小氨基酸组(甘氨酸, 丙氨酸, 丝氨酸和苏氨酸)内的替代。通常不改变特异性活性的氨基酸替代是本领域已知的并且在 H.Neurath 和 R.L.Hill, 1979 《The Protein》, Academic Press, New York 一书中进行过描述。最常发生的交换是 Ala/Ser, Val/Ile, Asp/Glu, Thr/Ser, Ala/Gly, Ala/Thr, Ser/Asn, Ala/Val, Ser/Gly, Tyr/Phe, Ala/Pro, Lys/Arg, Asp/Asn, Leu/Ile, Leu/Val, Ala/Glu 和 Asp/Gly 以及其反向交换。

在本发明一个有意义的实施方案中, 氨基酸变化具有如此特性以至改变了多肽的理化性质。举例来说, 可以实行氨基酸变化, 以提高多肽的热稳定性, 改变底物特异性, 改变 pH 最适值等。

优选地, 对比 SEQ ID NO: 2 中 1 至 842 位氨基酸, 这种替代、缺失和/或插入的数目至多为 10, 如至多为 9, 如至多为 8, 更优选至多为 7, 如至多为 6, 再如至多为 5, 最优选至多为 4, 如至多为 3, 再如至多为 2, 尤其至多为 1。

本发明人已经从米曲霉 (*Aspergillus oryzae*) 中分离出编码具纤维二糖酶活性的多肽的基因并将此基因插入质粒 pJaL621 (参见实施例 1) 中, 而该质粒被插入大肠杆菌中。根据国际承认用于专利程序保藏微生物的布达佩斯条约, 于 2001 年 4 月 19 日在德意志微生物保藏中心 (DSMZ - Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Mascheroder Weg 1B, D-38124 Braunschweig Germany) 保藏了携带此基因的大肠杆菌菌株, 保藏号为 DSM 14240。

这样, 在第二实施方案中, 本发明涉及包含如下氨基酸序列, 优选由如下氨基酸序列组成的多肽, 此氨基酸序列与插在大肠杆菌 DSM 14240 的质粒内的核苷酸序列的纤维二糖酶编码部分所编码的多肽具有至少 65% 的同一性。在本发明一个有意义的实施方案中, 多肽包含如下氨基酸序列并优选由如下氨基酸序列组成, 此氨基酸序列与插在大肠杆菌 DSM 14240 的质粒内

的核苷酸序列的纤维二糖酶编码部分所编码的多肽具有至少 70%，如至少 75%，优选至少 80%，如至少 85%，更优选至少 90%，最优选至少 95%，如至少 96%，如至少 97%，甚至最优选至少 98%，如至少 99% 的同一性（以下称“同源多肽”）。在一个有意义的实施方案中，此氨基酸序列有至多十个氨基酸（如十个氨基酸），尤其至多五个氨基酸（如五个氨基酸）如至多四个氨基酸（如四个氨基酸），至多三个氨基酸（如三个氨基酸）与插入在大肠杆菌 DSM 14240 的质粒内的核苷酸序列的纤维二糖酶编码部分所编码的多肽不同。在特别有意义的实施方案中，此氨基酸序列有至多两个氨基酸（如两个氨基酸），如一个氨基酸与插在大肠杆菌 DSM 14240 的质粒内的核苷酸序列的纤维二糖酶编码部分所编码的多肽不同。

优选地，本发明多肽包含插在大肠杆菌 DSM 14240 的质粒内的核苷酸序列的纤维二糖酶编码部分所编码的多肽的氨基酸序列。在另一个优选实施方案中，本发明多肽由插在大肠杆菌 DSM 14240 的质粒内的核苷酸序列的纤维二糖酶编码部分所编码的多肽的氨基酸序列组成。

与上文所述相似，本发明多肽可以是人工变体，它包含这样的氨基酸序列并优选由这样的氨基酸序列组成，此氨基酸序列与插在大肠杆菌 DSM 14240 的质粒内的核苷酸序列的纤维二糖酶编码部分所编码的氨基酸序列相比，具有至少一个替代、缺失和/或插入的氨基酸。

在第三实施方案中，本发明涉及具有纤维二糖酶活性的多肽，它是由在极低严紧条件下，优选在低度严紧条件下，更优选在中度严紧条件下，更优选在中高度严紧条件下，甚至更优选在高度严紧条件下，以及最优选在极高严紧条件下，可与选自 (i) SEQ ID NO: 1 中核苷酸 87 至 2612 位所示序列的互补链和 (ii) SEQ ID NO: 1 中核苷酸 87 至 1139 位所示序列的互补链的多核苷酸探针进行杂交的核苷酸序列编码的 (J. Sambrook, E. F. Fritsch, 和 T. Maniatus, 1989, 分子克隆实验室手册 (Molecular Cloning, A Laboratory Manual), 2d edition, Cold Spring Harbor, New York.)

按照本领域熟知的方法, SEQ ID NO: 1 所示的核苷酸序列或其亚序列, 以及 SEQ ID NO: 2 所示的氨基酸序列或其片断都可用于设计核苷酸探针以

便从不同属或种的菌株中识别和克隆出能编码具纤维二糖酶活性的多肽的 DNA。具体地,按照标准 DNA 印迹法利用这类探针与目的属或种的基因组或 cDNA 杂交,以便识别和分离出其中的相应基因。这类探针可能比整个序列短得多,但应该有至少 15 个,优选至少 25 个,更优选至少 35 个核苷酸长,如至少 70 个核苷酸长。然而,优选的是多核苷酸探针有至少 100 个核苷酸长。举例来说,多核苷酸探针可以是至少 200 个核苷酸长,至少 300 个核苷酸长,至少 400 个核苷酸长或者至少 500 个核苷酸长。甚至可利用更长的探针,如至少 600 个核苷酸长,至少 700 个核苷酸长,至少 800 个核苷酸长,或者至少 900 个核苷酸长的多核苷酸探针。DNA 和 RNA 探针都可利用。探针一般都进行标记用于检测相应基因(例如,用 ^{32}P 、 ^3H 、 ^{35}S 、生物素或抗生物素蛋白进行标记)。

这样,可从制备自这些其它生物体的基因组 DNA 或 cDNA 文库中筛选出与上述探针杂交的并且编码具纤维二糖酶活性的多肽的 DNA。来自这些其它生物体的基因组或其它 DNA 可通过琼脂糖或聚丙烯酰胺凝胶电泳,或其它分离技术进行分离。可以将来自文库的 DNA 或分离的 DNA 转移并固定在硝化纤维或其它适宜载体物质上。为鉴别出与 SEQ ID NO: 1 所示序列同源的克隆或 DNA,对载有固定化 DNA 的载体物质进行 DNA 印迹分析。

对于本发明,杂交指的是核苷酸序列杂交到标记的多核苷酸探针上,其中后者在极低到极高严紧条件下可与如 SEQ ID NO: 1 所示的核苷酸序列杂交。在这些条件下与多核苷酸探针杂交的分子可利用 X-射线胶片或通过任何其它本领域已知方法进行检测。只要在本发明上下文中使用术语“多核苷酸探针”,此术语应理解为是一种含有至少 15 个核苷酸的探针。

在一个有意义的实施方案中,多核苷酸探针是 SEQ ID NO: 1 之 87 至 1139 位核苷酸所示序列的互补链。

在另一有意义的实施方案中,多核苷酸探针是能够编码 SEQ ID NO: 2 多肽的核苷酸序列的互补链。在再一有意义的实施方案中,多核苷酸探针是如 SEQ ID NO: 1 所示序列的互补链。在再一个有意义的实施方案中,多核苷酸探针是 SEQ ID NO: 1 的成熟多肽编码区的互补链。在另一有意义的实

施方案中，多核苷酸探针是包含在大肠杆菌 DSM 14240 中的质粒 pJaL621 中的核苷酸序列的互补链。再一个有意义的实施方案中，多核苷酸探针是包含在大肠杆菌 DSM 14240 的质粒 pJaL621 中的成熟多肽编码区的互补链。

对于至少 100 个核苷酸长的长探针来说，极低至极高严紧条件被规定为按照标准 DNA 印迹分析法，在 42℃，在 5XSSPE、1.0%SDS、5X Denhardt 溶液、100 μg/ml 剪切变性的鲑鱼精子 DNA 中进行预杂交和杂交。优选地，至少 100 个核苷酸长的长探针不包括那些大于 1000 个核苷酸的探针。对于至少 100 个核苷酸长的长探针来说，载体物质最后利用 2xSSC，0.1%SDS 在 42℃（极低严紧条件）洗涤三次每次洗涤 15 分钟；优选利用 0.5x SSC，0.1%SDS 在 42℃（低度严紧条件）洗涤三次每次洗涤 15 分钟；更优选利用 0.2x SSC，0.1%SDS 在 42℃（中度严紧条件）洗涤三次每次洗涤 15 分钟；甚至更优选利用 0.2x SSC，0.1%SDS 在 55℃（中-高度严紧条件）洗涤三次每次洗涤 15 分钟；最优选利用 0.1x SSC，0.1%SDS 在 60℃（高度严紧条件）洗涤三次每次洗涤 15 分钟；尤其优选利用 0.1x SSC，0.1%SDS 在 68℃（极高严紧条件）洗涤三次每次洗涤 15 分钟。

尽管不是特别地优选，但也可考虑使用较短探针，可例举约 15 至 99 个核苷酸长，如约 15 至约 70 个核苷酸长的探针。对于这类短探针来说，严紧条件被规定为按照标准 DNA 印迹分析法，在低于计算出的 T_m 5℃至 10℃的温度下（利用 Bolton 和 McCarthy(1962, Proceedings of the National Academy of Sciences USA, 48: 1390) 的计算方法计算），在 0.9M NaCl、0.09M Tris-HCl pH7.6、6mM EDTA、0.5%NP-40、1XDenhardt 溶液、1mM 焦磷酸钠、1mM 一代磷酸钠、0.1mM ATP 及每 ml 0.2mg 的酵母 RNA 中进行预杂交、杂交及杂交后洗涤。

对于约 15 至 99 个核苷酸长的短探针，载体物质在低于计算出的 T_m 值 5℃至 10℃的温度下在 6XSSC 加 0.1%SDS 中洗涤一次洗 15 分钟，然后利用 6XSSC 洗涤两次每次洗 15 分钟。

具纤维二糖酶活性的多肽的来源

本发明多肽可从任何微生物属中获得。对于本发明来说，本处所用的术语“从……获得”应该指的是由核苷酸序列编码的多肽是通过这样的细胞产生的，其中此核苷酸序列天然存在于此细胞中或已经被插入到此细胞中。在优选实施方案中，多肽被分泌到细胞外。

本发明多肽可以是一种细菌多肽。例如，此多肽可以是革兰氏阳性细菌多肽，如芽孢杆菌属 (*Bacillus*) 多肽——可例举有嗜碱芽孢杆菌 (*Bacillus alkalophilus*)、解淀粉芽孢杆菌 (*Bacillus amyloliquefaciens*)、短芽孢杆菌 (*Bacillus brevis*)、环状芽孢杆菌 (*Bacillus circulans*)、凝结芽孢杆菌 (*Bacillus coagulans*)、灿烂芽孢杆菌 (*Bacillus lautus*)、迟缓芽孢杆菌 (*Bacillus lentus*)、地衣芽孢杆菌 (*Bacillus licheniformis*)、巨大芽孢杆菌 (*Bacillus megaterium*)、嗜热脂肪芽孢杆菌 (*Bacillus Stearothermophilus*)、枯草芽孢杆菌 (*Bacillus subtilis*) 或苏云金芽孢杆菌 (*Bacillus thuringiensis*) 多肽；或链霉菌属 (*Streptomyces*) 多肽，可例举有青紫链霉菌 (*Streptomyces lividans*) 或鼠灰链霉菌 (*Streptomyces murinus*) 多肽；或革兰氏阴性细菌多肽，例如大肠杆菌或假单胞菌属 (*Pseudomonas*) 的多肽。

本发明多肽可以是一种真菌多肽，并且更优选酵母多肽如假丝酵母属 (*Candida*)、克鲁维酵母属 (*Kluyveromyces*)、毕赤酵母属 (*Pichia*)、酵母属 (*Saccharomyces*)、裂殖酵母属 (*Schizosaccharomyces*) 或 *Yarrowia* 多肽；或更优选丝状真菌多肽如枝顶孢属 (*Acremonium*)、曲霉属、短梗霉属 (*Aureobasidium*)、隐球酵母属 (*Cryptococcus*)、线黑粉菌属 (*Filibasidium*)、镰孢属 (*Fusarium*)、腐质霉属 (*Humicola*)、*Magnaporthe*、毛霉属 (*Mucor*)、毁丝霉属 (*Myceliophthora*)、*Noecallimastix*、脉孢菌属 (*Neurospora*)、拟青霉属 (*Paecilomyces*)、青霉属 (*Penicillium*)、*Piromyces*、裂褶菌属 (*Schizophyllum*)、踝节菌属 (*Talaromyces*)、热子囊菌属 (*Thermoascus*)、梭孢壳属 (*Thielavia*)、*Tolypocladium* 或木霉属 (*Trichoderma*) 多肽。

在一个有意义的实施方案中，多肽是卡尔酵母 (*Saccharomyces calenbergensis*)、酿酒酵母 (*Saccharomyces cerevisiae*)、糖化酵母 (*Saccharomyces diastaticus*)、*Saccharomyces douglasii*、克鲁弗酵母 (*Saccharomyces kluyveri*)、

诺地酵母(*Saccharomyces norbensis*)或卵形酵母 (*Saccharomyces oviformis*) 多肽。

在另一个有意义的实施方案中,多肽是棘孢曲霉、泡盛曲霉(*Aspergillus awamori*)、臭曲霉(*Aspergillus foetidus*)、日本曲霉(*Aspergillus Japonicus*)、构巢曲霉、黑曲霉、米曲霉、杆孢状镰孢(*Fusarium bactridioides*)、*Fusarium cerealis*、*Fusarium crookwellense*、大刀镰孢(*Fusarium culmorum*)、禾谷镰孢(*Fusarium graminearum*)、禾赤镰孢(*Fusarium graminum*)、异孢镰孢(*Fusarium heterosporum*)、合欢木镰孢(*Fusarium negundi*)、尖镰孢(*Fusarium oxysporum*)、多枝镰孢(*Fusarium reticulatum*)、粉红镰孢(*Fusarium roseum*)、接骨木镰孢(*Fusarium sambucinum*)、肤色镰孢(*Fusarium sarcochroum*)、拟分枝孢镰孢(*Fusarium sporotrichioides*)、硫色镰孢(*Fusarium sulphureum*)、近念珠状镰孢(*Fusarium torulosum*)、*Fusarium trichothecioides*、有毒镰孢(*Fusarium venenatum*)、*Humicola insolens*、*Humicola lanuginosa*、米黑毛霉(*Mucor miehei*)、嗜热毁丝霉(*Myceliophthora thermophila*)、粗糙脉孢菌(*Neurospora crassa*)、产紫青霉(*Penicillium purpurogenum*)、*Trichoderma harzianum*、康宁木霉(*Trichoderma koningii*)、长枝木霉(*Trichoderma longibrachiatum*)、*Trichoderma reesei*或绿色木霉(*Trichoderma viride*)多肽。

在一个优选的实施方案中,多肽是米曲霉多肽,最优选大肠杆菌 DSM 14240 多肽,举例来说,由 SEQ ID NO: 2 中 1 至 842 位氨基酸序列组成的多肽。

对前面提及的菌种,应理解的是,不管已知的种名是什么,本发明既包括完全阶段和不完全阶段,还包括其它分类学上的等价物,如无性型。本领域技术人员能够很容易地识别出适当等价物的身份。

这些菌种的菌株可以容易地在许多培养物保藏机构为公众所得到,如美国典型培养物保藏中心(ATCC),德意志微生物保藏中心(DSM),真菌菌种保藏中心(CBS)和农业研究机构保藏中心(NRRL)。

此外,利用上述探针还可从其它来源,包括从自然界(如,土壤、堆肥、水域等)中分离出来的微生物中鉴别和获得这类多肽。从自然生境中分离微

生物的技术是本领域熟知的。这样核苷酸序列可类似地通过筛选另一种微生物的基因组或 cDNA 文库获得。一旦探针检测出编码多肽的核苷酸序列，此序列就可用那些本领域普通技术人员已知的技术（实例参见 Sambrook 等，1989，同上）进行分离或克隆。

由本发明核苷酸序列编码的多肽还包括融合多肽或可断裂的融合多肽，其中另一多肽在本多肽或其片段的 N-末端或 C-末端与本多肽或其片段融合。可通过将编码另一个多肽的核苷酸序列（或其部分）与本发明的核苷酸序列（或其部分）融合而产生融合多肽。产生融合多肽的技术是本领域已知的，其中包括连接编码这些多肽的编码序列以使这些序列处于阅读框中，并使融合多肽的表达受同一启动子和终止子的调控。

多核苷酸和核苷酸序列

本发明还涉及具有编码本发明多肽的核苷酸序列的多核苷酸。尤其，本发明涉及这样的多核苷酸，其由编码本发明多肽的核苷酸序列组成。在优选的实施方案中，此核苷酸序列为 SEQ ID NO:1 中所示序列。在更优选的实施方案中，此核苷酸序列是 SEQ ID NO:1 的成熟多肽编码区。在另一更优选的实施方案中，此核苷酸序列是包含在大肠杆菌 DSM 14240 的质粒 pJaL621 中的成熟多肽编码区。本发明还包括含有这样的核苷酸序列并优选由这样的核苷酸序列组成的多核苷酸，此核苷酸序列能够编码由 SEQ ID NO:2 所示的氨基酸序列组成的多肽或其成熟多肽，但此核苷酸序列与 SEQ ID NO:1 因遗传密码简并性而不同。

本发明还涉及含有 SEQ ID NO:1 亚序列，并优选由此亚序列组成的多核苷酸，所述亚序列能够编码 SEQ ID NO:2 中具有纤维二糖酶活性的片断。SEQ ID NO:1 的亚序列是 SEQ ID NO:1 所包含的核苷酸序列，但其在 5' 和/或 3' 末端有一个或多个核苷酸缺失。

本发明还涉及含有修饰核苷酸序列并优选由此序列组成的多核苷酸，所述修饰核苷酸序列在 SEQ ID NO:1 所示的成熟多肽编码序列中包含至少一个修饰，并且此修饰核苷酸序列编码的多肽由 SEQ ID NO:2 中的 1 至 842 位

氨基酸组成。

用于分离或克隆编码多肽的核苷酸序列的技术是本领域已知的，其包括从基因组 DNA 进行的分离，从 cDNA 进行的制备，或者两者的组合。举例来说，通过利用熟知的聚合酶链反应 (PCR) 或抗体筛选表达文库 检测出具有相同结构特征的克隆 DNA 片断，可以实现从基因组 DNA 克隆本发明核苷酸序列。实例参见 Innis 等，1990，PCR：方法和应用指南 (PCR: A Guide to Methods and Application)，Academic Press，New York。还可使用其它扩增方法如连接酶链反应 (LCR)，连接激活转录 (LAT) 及基于核苷酸序列的扩增 (NASBA)。核苷酸序列可从曲霉属菌株，或者另外或相关生物体中克隆得到，因而它可以，举例来说，是此核苷酸序列的多肽编码区的等位基因变体或物种变体。

核苷酸序列可利用遗传工程的标准克隆方法使天然位点的核苷酸序列重定位在一个可使此核苷酸序列复制的不同位点而获得。克隆方法可能涉及切除和分离出所需要的含有编码多肽的核苷酸序列的片断，将片断插入载体分子，然后将重组载体导入宿主细胞中，在宿主细胞中复制出多个拷贝或克隆的核苷酸序列。核苷酸序列可来源于基因组，cDNA，RNA，半合成，合成或其任意组合。

本发明还涉及含有这样一个核苷酸序列并优选由此序列组成的多核苷酸，所述核苷酸序列与 SEQ ID NO: 1 中 87 至 2612 位核苷酸所示序列具有至少 65% 的同一性。优选地，此核苷酸序列与 SEQ ID NO: 1 中 87 至 2612 位核苷酸所示序列具有至少 70% 的同一性，例如至少 80% 的同一性，如至少 90% 的同一性，更优选至少 95% 的同一性，如至少 96% 的同一性，例如至少 97% 的同一性，甚至更优选至少 98% 的同一性，如至少 99% 的同一性。优选地，此核苷酸序列编码具有纤维二糖酶活性的多肽。两个核苷酸序列的同一性程度可如前文所述进行确定(参见标题“定义”一节)。优选地，此核苷酸序列包含 SEQ ID NO: 1 中 87 至 2612 位核苷酸所示序列。在甚至更优选的实施方案中，此核苷酸序列由 SEQ ID NO: 1 中 87 至 2612 位核苷酸组成。

在另一有意义的方面，本发明涉及包含这样一个核苷酸序列并优选由此

序列组成的多核苷酸，所述核苷酸序列与插在大肠杆菌 DSM 14240 的质粒中的核苷酸序列的纤维二糖酶编码部分具有至少 65% 的同一性。在优选的实施方案中，与插在大肠杆菌 DSM 14240 的质粒中的核苷酸序列的纤维二糖酶编码部分具有的同一性是至少 70%，举例来说至少 80%，如至少 90%，更优选至少 95%，如至少 96%，再例如至少 97%，甚至更优选至少 98%，如至少 99%。优选地，此核苷酸序列包含插在大肠杆菌 DSM 14240 的质粒中的核苷酸序列的纤维二糖酶编码部分。在甚至更优选的实施方案中，此核苷酸序列由插在大肠杆菌 DSM 14240 的质粒中的核苷酸序列的纤维二糖酶编码部分组成。

对编码本发明多肽的核苷酸序列进行修饰可能是合成如下多肽所必需的，所述多肽包含与 SEQ ID NO: 2 中 1 至 842 位氨基酸所示序列相比具有至少一个替代、缺失和/或插入的氨基酸序列。这些人工变体可以以某种工程化方式区别于从天然源分离的多肽，如在比活性、热稳定性、pH 最适值等方面不同的变体。

对本领域技术人员来说，很明显这类修饰作用可在分子功能的关键区域之外进行且仍可得到活性多肽。对于本发明核苷酸序列编码的多肽的活性必需的氨基酸残基（因此优选不进行如替代这样的修饰），可按照本领域已知的方法——如定点诱变或丙氨酸扫描诱变进行鉴别（实例参见，Cunningham 和 Wells, 1989, Science 244: 1081 - 1085）。后一种技术中，向分子中每一个正电荷残基位置引入突变，然后检测所致的突变分子的纤维二糖酶活性以识别对分子活性起关键作用的氨基酸残基。底物-酶相互作用位点也可通过三维结构分析来确定，其中三维结构可通过核磁共振分析、晶体学或光亲和标记这类技术进行测定（实例参见 de Vos 等, 1992, Science 255: 306 - 312; Smith 等, 1992, Journal of Molecular Biology 224: 899 - 904; Wlodaver 等, 1992, FEBS Letters 309: 59 - 64）。

此外，编码本发明多肽的核苷酸序列可通过引入核苷酸替代来修饰，该替代不会导致此核苷酸序列编码的多肽具有另一氨基酸序列，但可使此核苷酸序列符合用来产酶的宿主生物体的密码子使用。

将突变引入核苷酸序列使一个核苷酸与另一个核苷酸交换可利用本领域任一已知方法通过定点诱变来完成。特别有用的是利用携带有目的插入物的超螺旋双链 DNA 和两个含有期望突变的合成引物的方法。寡聚核苷酸引物（每个引物各与载体的相反链互补）借助于 Pfu DNA 聚合酶在温度循环期间进行延伸。加入引物后产生出含有交错切口的突变质粒。温度循环之后，用特异针对甲基化和半甲基化 DNA 的 DpnI 处理产物，消化亲本 DNA 模板并筛选出含突变的合成 DNA。还可以使用其它本领域已知的方法。对核苷酸替代的一般说明可参见 Ford 等, 1991, 蛋白质表达和纯化(Protein Expression and Purification) 2: 95 - 107。

本发明还涉及具有这样一个核苷酸序列并优选由此序列组成的多核苷酸，所述核苷酸序列编码具有纤维二糖酶活性的多肽，并且可以在极低严紧条件下，优选在低度严紧条件下，更优选在中度严紧条件下，更优选在中 - 高度严紧条件下，甚至更优选在高度严紧条件下，最优选在极高严紧条件下与选自：(i) SEQ ID NO: 1 中核苷酸 87 至 2612 位所示序列的互补链，(ii) SEQ ID NO: 1 中核苷酸 87 至 1139 位所示序列的互补链的多核苷酸探针进行杂交。

应该知道，关于核苷酸序列杂交的细节与详情与标题为“具有纤维二糖酶活性的多肽”一节中论述的杂交方面相同或相似。

核酸构建体

本发明还涉及包含可操作地与一个或多个调控序列连接的本发明核苷酸序列的核酸构建体，所述调控序列能够指导此编码序列在适当宿主细胞中在与调控序列相适应的条件下进行表达。

可利用多种方法处理编码本发明多肽的核苷酸序列以备多肽表达。根据表达载体的不同，核苷酸序列在插入载体之前所进行的操作可能是需要的或必需的。利用重组 DNA 法修饰核苷酸序列的技术是本领域所熟知的。

调控序列可以是适当的启动子序列，即可被宿主细胞识别用于表达本发明核苷酸序列的核苷酸序列。启动子序列包含可调节多肽表达的转录调控序

列。启动子可以是任意的能够在所挑选的宿主细胞中表现出转录活性的核苷酸序列,包括突变启动子、截短启动子、和杂合启动子,并且可从与宿主细胞同源或异源的编码细胞外或细胞内多肽的基因中获得。

尤其在细菌宿主细胞中,可指导本发明核酸构建体转录的适用启动子的实例为来自大肠杆菌 *lac* 操纵子、天蓝色链霉菌(*Streptomyces coelicolor*)琼脂糖酶基因(*dagA*)、枯草芽孢杆菌果聚糖蔗糖酶基因(*sacB*)、地衣芽孢杆菌 α -淀粉酶基因(*amyL*)、嗜热脂肪芽孢杆菌生麦芽糖淀粉酶基因(*amyM*)、解淀粉芽孢杆菌 α -淀粉酶基因(*amyQ*)、地衣芽孢杆菌青霉素酶基因(*penP*)、枯草芽孢杆菌 *xylA* 和 *xylB* 基因及原核 β -内酰胺酶基因(Villa-Kamaroff 等, 1978, *Proceedings of the National Academy of Science USA*, 75:3727-3731)的启动子,以及 *tac* 启动子(DeBoer 等, 1983, *Proceedings of the National Academy of Science USA*, 80: 21 - 25)。对其它启动子的说明参见 *Scientific American*, 1980, 242: 74 - 94 中的“来自重组细菌的有用蛋白质”; 和 Sambrook 等, 1989, 同上引文。

在丝状真菌宿主细胞中,可指导本发明核酸构建体转录的适用启动子的实例是来自米曲霉 TKA 淀粉酶基因、米黑根毛霉(*Rhizomucor miehei*)的天冬氨酸蛋白酶基因、黑曲霉中性 α -淀粉酶基因、黑曲霉酸稳定 α -淀粉酶基因、黑曲霉或泡盛曲霉的葡糖淀粉酶(*glaA*)基因、米黑根毛霉脂肪酶基因、米曲霉碱性蛋白酶基因、米曲霉磷酸丙糖异构酶基因、构巢曲霉乙酰胺酶基因、和尖镰孢胰蛋白酶样蛋白酶(WO 96/00787)基因的启动子,以及 NA2-*tpi* 启动子(来自黑曲霉中性 α -淀粉酶和米曲霉磷酸丙糖异构酶基因的启动子杂合体),和它们的突变、截短及杂合启动子。

在酵母宿主中,有用的启动子来自酿酒酵母烯醇化酶(*ENO-1*)基因、酿酒酵母半乳糖激酶(*GAL1*)基因、酿酒酵母醇脱氢酶/甘油醛-3-磷酸脱氢酶(*ADH2/GAP*)基因和酿酒酵母 3-磷酸甘油酸激酶基因。Romanos 等, 1992, *Yeast* 8: 423 - 488 对其它适用于酵母宿主细胞的启动子进行了说明。

调控序列还可以是适当的转录终止序列,即可被宿主细胞识别进而终止

转录的序列。终止序列可操作地与编码多肽的核苷酸序列的 3'末端连接。任何在所选宿主细胞中起作用的终止子都可用于本发明。

对于丝状真菌宿主细胞，优选的终止子来自米曲霉 TAKA 淀粉酶基因、黑曲霉葡萄糖淀粉酶基因、构巢曲霉邻氨基苯甲酸合酶基因、黑曲霉 α -葡萄糖苷酶基因和尖镰孢胰蛋白酶样蛋白酶基因。

对于酵母宿主细胞，优选的终止子来自酿酒酵母烯醇化酶基因、酿酒酵母细胞色素 C (CYC1) 基因和酿酒酵母甘油醛-3-磷酸脱氢酶基因。Romanos 等, 1992, 同上引文中对其它适用于酵母宿主细胞的终止子进行了说明。

调控序列还可以是适当的前导序列，即对宿主细胞的翻译来说很重要的 mRNA 非翻译区。前导序列可操作地与编码多肽的核苷酸序列的 5'末端连接。任何在所选宿主细胞中起作用的前导序列都可用于本发明。

适用于酵母宿主细胞的前导序列可以获自酿酒酵母烯醇化酶基因 (ENO-1)、酿酒酵母 3-磷酸甘油酸激酶基因、酿酒酵母 α -因子基因和酿酒酵母醇脱氢酶/甘油醛-3-磷酸脱氢酶 (ADH2/GAP) 基因。

调控序列还可以是聚腺苷酸化序列，即可操作性地与核苷酸序列的 3'末端连接的序列，在转录中，此序列可被宿主细胞识别作为将聚腺苷残基添加到转录的 mRNA 上的信号。任何在所选宿主细胞中起作用的聚腺苷酸化序列都可用于本发明。

对于丝状真菌宿主细胞优选聚腺苷酸化序列获自米曲霉 TAKA 淀粉酶基因、黑曲霉葡萄糖淀粉酶基因、构巢曲霉邻氨基苯甲酸合酶基因、尖镰孢胰蛋白酶样蛋白酶基因和黑曲霉 α -葡萄糖苷酶基因。

Guo 和 Sherman, 1995, 分子细胞生物学 (Molecular Cellular Biology) 15: 5983-5990 对适用于酵母宿主细胞的聚腺苷酸化序列进行了说明。

调控序列还可以是信号肽编码区，它编码的氨基酸序列与多肽的氨基末端连接并且指导此编码多肽进入细胞分泌途径。核苷酸序列的编码序列的 5'末端可能固有地包含一个信号肽编码区，此信号肽编码区在翻译阅读框中与编码分泌多肽的编码区片断天然连接。或者，编码区的 5'末端可以包含一个

对编码序列而言外源的信号肽编码区。如果编码序列天然不含有信号肽编码区，则可能需要外源信号肽编码区。或者，外源信号肽编码区可以简单地替代天然信号肽编码区以增强多肽的分泌。然而，任何可以指导表达的多肽进入所选宿主细胞的分泌途径的信号肽编码区均可用于本发明。

信号肽编码区可以是编码 SEQ ID NO: 2 中 -19 至 -1 位氨基酸的 SEQ ID NO: 1 中 30 至 86 位核苷酸。

对于细菌宿主细胞有效的信号肽编码区有获自芽孢杆菌 NCIB 11837 生麦芽糖淀粉酶基因、嗜热脂肪芽孢杆菌 α -淀粉酶基因、地衣芽孢杆菌枯草杆菌蛋白酶基因、地衣芽孢杆菌 β -内酰胺酶基因、嗜热脂肪芽孢杆菌中性蛋白酶 (nprT, nprS, nprM) 基因和枯草芽孢杆菌 prsA 基因的信号肽编码区。Simonen 和 Palva, 1993 在 Microbiological Reviews 57: 109 - 137 中对其它信号肽进行了说明。

对于丝状真菌宿主细胞有效的信号肽编码区有来自米曲霉 TAKA 淀粉酶基因、黑曲霉中性淀粉酶基因、黑曲霉葡糖淀粉酶基因、米黑根毛霉天冬氨酸蛋白酶基因、*Humicola insolens* 纤维素酶基因和 *Humicola lanuginosa* 脂肪酶基因的信号肽编码区。

对于酵母宿主细胞适用的信号肽来自酿酒酵母 α -因子基因和酿酒酵母转化酶基因。Romanos 等, 1992, 同上引文中对其它适用的信号肽编码区进行了说明。

调控序列还可以是编码位于多肽氨基末端的氨基酸序列的前肽编码区。所得多肽被称作前酶或前多肽(或者有时称作酶原)。前多肽一般是无活性的，并可通过催化或自我催化从前多肽中切除前肽而转变成成熟活性多肽。前肽编码区可来自枯草芽孢杆菌碱性蛋白酶 (aprE) 基因、枯草芽孢杆菌中性蛋白酶 (nprT) 基因、酿酒酵母 α -因子基因、米黑根毛霉天冬氨酸蛋白酶基因和嗜热毁丝霉漆酶基因 (WO 95/33836)。

当信号肽和前肽区同时存在于多肽的氨基末端时，前肽区与多肽的氨基末端连接，信号肽区与前肽的氨基末端连接。

还可能需要加上调节序列以使多肽的表达可以相对于宿主细胞的生长

进行调节。可例举的调节系统有那些能够响应化学物理刺激（包括存在调节化合物）而引起基因表达开启或关闭的调节系统。原核生物系统中的调节系统包括 lac, tac 和 trp 操纵子系统。在酵母中，可使用 ADH2 系统或 GAL1 系统。在丝状真菌中，TAKA α -淀粉酶启动子、黑曲霉葡萄糖淀粉酶启动子和米曲霉葡萄糖淀粉酶启动子可用作调节序列。可例举的调节序列的其它实例是那些允许基因扩增的序列。在真核生物系统中，此类序列包括有氨甲蝶呤存在时进行扩增的二氢叶酸还原酶基因，以及有重金属存在时进行扩增的金属硫蛋白基因。在这些情况下，编码多肽的核苷酸序列与调节序列可操作地连接。

表达载体

本发明还涉及含有本发明核酸构建体的重组表达载体。可以将上文所述的各种核苷酸和调控序列连接起来产生重组表达载体，此重组表达载体可包括一个或多个方便的限制酶切位点使得编码多肽的核苷酸序列可以在这些位点进行插入或替代。或者，本发明核苷酸序列的表达可以通过将本核苷酸序列或含有本序列的核酸构建体插入适当表达载体中来进行。构建表达载体时，将编码序列置于载体中以使编码序列可操作地与适当的表达调控序列连接。

重组表达载体可以是任何便于接受重组 DNA 操作并能引起核苷酸序列表达的载体（如，质粒或病毒）。载体的选择一般取决于载体与引入此载体的宿主细胞之间的相容性。此载体可以是线状的或闭合环状的质粒。

载体可以是自主复制的载体，即载体以染色体外实体形式而存在，其复制独立于染色体的复制，举例来说，质粒、染色体外因子、微型染色体或人工染色体。

载体可以包含任何用于确保自主复制的手段。或者，载体可以是这样一个载体，它被引入宿主细胞之后，将整合进入基因组并与被整合的染色体一起复制。此外，可利用单一载体或质粒或者共同含有待引入宿主细胞基因组的全部 DNA 的两个或多个载体或质粒或者转座子。

本发明载体优选包含一个或多个选择标记以允许容易地挑选出转化细

胞。选择标记是一种基因，其产物提供抵抗杀生物剂或病毒的抗性、抗重金属抗性，对营养缺陷体的原养等。

可例举的细菌选择标记有枯草芽孢杆菌或地衣芽孢杆菌的 *dal* 基因，或提供抗生素抗性如氨苄青霉素、卡那霉素、氯霉素或四环素抗性的标记。适用于酵母宿主细胞的标记有 *ADE2*、*HIS3*、*LEU2*、*LYS2*、*MET3*、*TRP1* 和 *URA3*。适用于丝状真菌宿主细胞的选择标记包括，但不限于：*amdS*（乙酰胺酶）、*argB*（鸟氨酸氨基甲酰基转移酶）、*bar*（蒴丝菌素乙酰转移酶）、*hygB*（潮霉素磷酸转移酶）、*niaD*（硝酸还原酶）、*pyrG*（乳清酸核苷-5'-磷酸脱羧酶）、*sC*（硫酸腺苷酸转移酶）、*trpC*（邻氨基苯甲酸合酶）及其等价物。

优选用于曲霉属细胞的是构巢曲霉或米曲霉的 *amdS* 和 *pyrG* 基因及吸水链霉菌（*Streptomyces hygroscopicus*）的 *bar* 基因。

本发明载体优选包含允许载体稳定整合到宿主细胞的基因组中或者允许载体在细胞中不依赖于基因组自主复制的元件。

对整合到宿主细胞基因组中来说，载体可依靠编码多肽的核苷酸序列或任何其它载体元件以通过同源或非同源重组使载体稳定整合到基因组中。或者，载体可包含附加的核苷酸序列用来指导通过同源重组整合进入宿主细胞的基因组。此附加核苷酸序列能使载体整合到宿主细胞基因组中染色体的精确位置上。为提高整合在精确位置的可能性，整合元件应优选包含足够数目的与相应靶序列高度同源的核苷酸，如 100 至 1,500 个碱基对，优选 400 至 1,500 个碱基对，最优选 800 至 1,500 个碱基对，以提高同源重组的几率。整合元件可以是任何与宿主细胞基因组中的靶序列同源的序列。此外，整合元件可以是非编码或编码核苷酸序列。另一方面，载体可通过非同源重组整合到宿主细胞的基因组中。

对自主复制来说，载体还可包含复制起点以使载体在所研究的宿主细胞中自主复制。可例举的细菌复制起点的实例有允许在大肠杆菌中复制的质粒 *pBR322*、*pUC19*、*PACYC177* 和 *pACYC184* 的复制起点，和允许在芽孢杆菌中复制的质粒 *pUB110*、*pE194*、*pTA1060* 和 *pAM β 1* 的复制起点。可例举的用于酵母宿主细胞中的复制起点的实例是 *2 μ* 复制起点、*ARS1*、*ARS4*、

ARS1 和 CEN3 的组合, 以及 ARS4 和 CEN6 的组合。复制起点可具有使其可在宿主细胞中以温度敏感方式起作用的突变(实例参见, Ehrlich, 1978, *Proceedings of the National Academy of Science USA*, 75: 1433)。

可将一个以上拷贝的本发明核苷酸序列插入到宿主细胞中以提高基因产物的产量。增加核苷酸序列的拷贝数可通过将至少一个附加拷贝的序列整合进宿主细胞基因组中来得到, 或者通过同时包括核苷酸序列与可扩增的选择标记基因、并在适当的选择试剂存在时培养此细胞以筛选出含有扩增拷贝的选择标记基因及由此含有额外拷贝的核苷酸序列的细胞而得到。

用于连接上文所述的元件以构建本发明重组表达载体的方法是本领域技术人员熟知的(实例参见 Sambrook 等, 1989, 同上引文)。

宿主细胞

本发明还涉及重组的含有本发明核酸构建体的宿主细胞, 它们可有利地用于多肽的重组生产。如前所述, 可将含本发明核苷酸序列的载体引入宿主细胞以使载体作为染色体的整合体或作为自主复制的染色体外载体而存在。

宿主细胞可以是单细胞微生物, 如原核生物、或者是多细胞微生物, 如真核生物。

可适用的单细胞是细菌细胞如革兰氏阳性细菌, 包括, 但不限于: 芽孢杆菌属细胞, 例如嗜碱芽孢杆菌、解淀粉芽孢杆菌、短芽孢杆菌、环状芽孢杆菌、*Bacillus clausii*、凝结芽孢杆菌、灿烂芽孢杆菌、迟缓芽孢杆菌、地衣芽孢杆菌、巨大芽孢杆菌、嗜热脂肪芽孢杆菌、枯草芽孢杆菌和苏云金芽孢杆菌; 或链霉菌属细胞, 例如青紫链霉菌或鼠灰链霉菌; 或者革兰氏阴性细菌如大肠杆菌和假单胞菌属种。在一个优选实施方案中, 细菌宿主细胞是迟缓芽孢杆菌、地衣芽孢杆菌、嗜热脂肪芽孢杆菌或枯草芽孢杆菌细胞。在另一个优选实施方案中, 芽孢杆菌属细胞是嗜碱芽孢杆菌。

载体导入细菌宿主细胞可以例如通过原生质体转化(实例参见 Chang 和 Cohen, 1979, *Molecular General Genetics* 168: 111 - 115)、利用感受态细胞(实例参见 Young 和 Spizizn, 1961, *Journal of Bacteriology* 81:823-829,

或 Dubnau 和 Davidoff - Abelson, 1971, *Journal of Molecular Biology* 56: 209 - 221)、电穿孔(实例参见 Shigekawa 和 Dower, 1988, *Biotechniques* 6: 742 - 751)或接合(实例参见 Koehler 和 Thorne, 1987, *Journal of Bacteriology* 169:5771-5278)实现。

宿主细胞可以是真核生物如哺乳动物,昆虫,植物或真菌细胞。

在一个优选实施方案中,宿主细胞是真菌细胞。本文所用的“真菌”包括子囊菌门(Ascomycota)、担子菌门(Basidiomycota)、壶菌门(Chytridiomycota)、和接合菌门(Zygomycota)(如 Hawsksworth 等在 Ainsworth and Bisby' Dictionary of The Fungi, 8th edition, 1995, CAB international ,University Press, Cambridge,UK 中的定义)以及卵菌门(Oomycota)(如在 Hawsksworth 等, 1995, 上引书 171 页中的引用)以及所有有丝分裂孢子真菌(Hawsksworth 等, 1995, 上引书)。

在一个更优选的实施方案中,真菌宿主细胞是酵母细胞,本文所用的“酵母”包括产子囊酵母(内孢霉目(Endomycetales))、产担子孢酵母,以及属于半知菌类(芽孢纲(Blastomycetes))的酵母。由于酵母的分类将来还会变化,对于本发明来说,酵母应按照“酵母生物学及活性”(Skinner, F.A., Passmore, S.M., 和 Davenport, R.R., eds, Soc. App. Bacteriol. Symposium Series No.9, 1980)”中的描述进行定义。

在一个甚至更优选的实施方案中,酵母宿主细胞是假丝酵母属、汉逊酵母属(*Hansenula*)、克鲁维酵母属、毕赤酵母属、酵母属、裂殖酵母属或 *Yarrowia* 细胞。

在一个最优选的实施方案中,酵母宿主细胞是卡尔酵母、酿酒酵母、糖化酵母、*Saccharomyces douglasii*、克鲁弗酵母、诺第酵母或或卵形酵母细胞。在另一最优选的实施方案中,酵母宿主细胞是乳酸克鲁维酵母(*Kluyveromyces lactis*)细胞。在另一最优选的实施方案中,酵母宿主细胞是 *Yarrowia lipolytica* 细胞。

在另一个更优选的实施方案中,真菌宿主细胞是丝状真菌细胞。“丝状真菌”包括所有真菌门(Eumycota)和卵菌门亚门中的丝状真菌(如

Hawksworth 等, 1995, 上引文中定义)。丝状真菌的特征是菌丝体壁由几丁质, 纤维素, 葡聚糖, 脱乙酰壳多糖、甘露聚糖及其它复合多糖组成。营养生长表现为菌丝延长, 碳代谢是专性需氧的, 相反, 酵母如酿酒酵母的营养生长表现为单细胞菌体的出芽而碳代谢可以是发酵的。

在甚至更优选的实施方案中, 丝状真菌宿主细胞是, 但不限于枝顶孢属、曲霉属、镰孢属、腐质霉属、毛霉属、毁丝霉属、脉孢菌属、青霉属、梭孢壳属、*Tolypocladium* 或木霉属种的细胞。

在最优选的实施方案中, 丝状真菌宿主细胞是泡盛曲霉、臭曲霉、日本曲霉、构巢曲霉、黑曲霉、或米曲霉细胞。在另一最优选的实施方案中, 丝状真菌宿主细胞是杆孢状镰孢、*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、尖镰孢、多枝镰孢、粉红镰孢、接骨木镰孢、肤色镰孢、拟分枝孢镰孢、疏色镰孢、近念珠状镰孢、*Fusarium trichothecioides* 或有毒镰孢细胞。在甚至最优选的实施方案中, 丝状真菌亲本细胞是有毒镰孢 (Nirenberg 新种) 细胞。在另一最优选的实施方案中, 丝状真菌宿主细胞是 *Humicola insolens*、*Humicola lanuginosa*、米黑毛霉、嗜热毁丝霉、粗糙脉孢菌、产紫青霉、*Thielavia terrestris*、*Trichoderma harzianum*、康宁木霉、长枝木霉、*Trichoderma reesei* 或绿色木霉细胞。

真菌细胞可以以本身已知的方式通过涉及原生质形成、原生质转化和细胞壁重建的过程进行转化。在 EP238 023 和 Yelton 等, 1984, PNAS 81: 1470 - 1474 中描述了适合曲霉属宿主细胞转化的方法。Malardier 等, 1989, 基因(Gene)78: 147 - 156 和 WO 96/00787 中描述了适合转化镰孢属种的方法。酵母可利用 Becher 和 Guarente 在 Abelson, J.N.和 Simon, M.I.编著的酶学方法 (Methods in Enzymotgy) 一书中的酵母遗传学和分子生物学指南, 194 卷, 182 - 187 页, Academic Press, Inc., New York; Ito 等, 1983, 细菌学杂志 153: 163; 和 Hinnen 等, 1978, PNAS 75:1920 中描述的方法进行转化。

生产方法

本发明还涉及用于生产本发明多肽的方法，其包括（a）培养菌株，该菌株的野生型能够产生本多肽；以及（b）回收本多肽。优选地，所述菌株为曲霉属菌株，更优选米曲霉菌株。

本发明还涉及生产本发明多肽的方法，其包括（a）在有益于本多肽生产的条件下培养宿主细胞；以及（b）回收本多肽。

在本发明的生产方法中，利用本领域已知的方法在适于生产本多肽的营养培养基中对细胞进行培养。例如，细胞培养可在实验室或工业发酵罐中在适当的培养基中并在可使本多肽表达和/或分离的条件下，通过摇瓶培养、小型或大型发酵（包括连续、分批、补料分批或固态发酵）来进行。利用本领域已知的方法，培养在含有碳源、氮源和无机盐的适当营养培养基中进行。适用的培养基可从供应商购买或者根据已发表的（如，在美国典型培养物保藏中心的目录中）培养基组成制备。如果多肽分泌到营养培养基中，可直接从培养基回收多肽。如多肽不进行分泌，则可从细胞裂解物回收多肽。

多肽可利用本领域已知的特异针对本多肽的方法进行检测。这些检测方法包括特异性抗体的利用、酶产物的形成或酶底物的消失。例如，酶测定试验可用来确定本文所述的多肽的活性。

所得多肽可通过本领域已知的方法回收。例如，可以利用常规方法，包括，但不限于：离心、过滤、抽提、喷雾干燥、蒸发或沉淀从营养培养基中回收多肽。

本发明多肽可通过本领域已知的许多方法纯化，这些方法包括，但不限于：层析法（如离子交换层析、亲和层析、疏水层析，层析聚焦和分子筛排阻层析），电泳法（如制备性电点聚焦电泳），差别溶解性法（如硫酸铵沉淀），SDS - PAGES，或提取法（实例参见“蛋白质纯化”（Protein Purification），J.-C.Janson 和 Lars Ryden, editors, VCH Publishers, New York, 1989）。

从生物量生产乙醇

乙醇可通过生物量的酶降解并将释放出的多糖转化成乙醇来生产。这种乙醇通常称为生物乙醇或生物燃料。它可在混合物（燃料替代物）中用作燃

料添加剂或增量剂，浓度从不到1%直到100%。在一些国家如巴西，乙醇在很大程度上正在取代汽油。

生物量的初生细胞壁中占优势的多糖是纤维素，含量第二的是半纤维素，第三是果胶。次生细胞壁，即细胞停止生长之后产生的细胞壁，也含有多糖并且通过聚合的木质素与半纤维素的共价交联得到强化。纤维素是脱水纤维二糖的同聚物，即线性 $\beta - (1 - 4) - D -$ 葡聚糖，而半纤维素包括许多种化合物，如带有多种取代基的复杂分支结构的木聚糖、木葡聚糖、阿拉伯木聚糖和甘露聚糖。虽然纤维素一般来说是多形的，但在植物组织中发现的纤维素主要是以不溶性的平行葡聚糖链的晶体矩阵的形式存在。半纤维素通常与纤维素以及另外的半纤维素氢键结合，这有助于稳定细胞壁的矩阵结构。

有三大类别的纤维素酶被用于裂解生物量：

- “内切-1, 4- β -葡聚糖酶”或“1, 4- β -D-葡聚糖-4-葡聚糖水解酶”(EC 3.2.1.4)，它随机地作用于可溶的和不可溶的1, 4- β -葡聚糖底物。
- “外切-1, 4- β -D-葡聚糖酶”，既包括1, 4- β -D-葡聚糖葡聚糖水解酶(EC 3.2.1.74)，它可从1, 4- β -D-葡聚糖中释放出D-葡萄糖并缓慢水解D-纤维二糖；又包括1, 4- β -D-葡聚糖纤维二糖水解酶(EC 3.2.1.91)，又称为纤维二糖水解酶I，它可从1, 4- β -D-葡聚糖中释放出D-纤维二糖。
- “ β -D-葡萄糖苷酶”或 β -D-葡萄糖苷葡萄糖水解酶(EC 3.2.1.21)，它可从纤维二糖和可溶的纤维糊精以及一批配糖物中释放出D-葡萄糖单元。

这三类酶在复杂的相互影响中协同作用，导致有效的解晶作用并水解生物量中的天然纤维素，产生还原糖，进而可以通过发酵将这些糖转化成乙醇。

因此本发明还涉及从生物量生产乙醇的方法，其包括使生物量接触本发明多肽；并涉及本发明多肽在生产乙醇中的用途。

植物

本发明还涉及经编码本发明具纤维二糖酶活性的多肽的核苷酸序列转化以致可以表达和生产可回收量的本多肽的转基因植物、植株部分或植物细胞。本多肽可从植物或植株部分中回收。或者，将含有重组多肽的植物或植株部分就此用于改善食物或饲料的质量，如改良营养价值、适口性和流变学性质或破坏抗营养因子。

转基因的植物可以是双子叶的(dicot)或单子叶的(monocot)。可例举的单子叶植物有牧草如草地早熟禾(蓝草,早熟禾属(Poa));饲料草如羊茅属(Festuca)、黑麦草属(Lolium);温带牧草,如剪股颖属(Agrostis);和谷类如小麦、燕麦、黑麦、大麦、稻、高粱和玉米(corn)。

可例举的双子叶植物有烟草、马铃薯、甜菜、豆类如羽扇豆、豌豆、蚕豆、大豆;以及十字花科植物(Brassicaceae family)如花椰菜、菜籽油菜,和亲缘很近的模式生物体鼠耳芥(*Arabidopsis thaliana*)。

可例举的植株部分有茎、愈合组织、叶、根、果实、种子和块茎。特殊的植物组织,如叶绿体、质外体、线粒体、液泡、过氧化物酶体和细胞质也被认为是植株部分。此外,任何植物细胞,不论什么组织来源,也被认为是植株部分。

这些植物、植株部分和植物细胞的子代也包括在本发明的范围之内。

可按照本领域已知的方法构建能够表达本发明多肽的转基因植物或植物细胞。简而言之,转基因植物或植物细胞的构建可以通过将一个或多个编码本发明多肽的表达构建体并入植物宿主基因组中并繁殖所致的修饰植物或植物细胞得到转基因的植物或植物细胞来进行。

有利地,此表达构建体是这样一个核酸构建体,它含有编码本发明多肽的核苷酸序列,其中此核苷酸序列与在选定的植物或植株部分中表达此核苷酸序列所需的适当调节序列可操作地连接。而且,表达构建体可包含用于识别整合了此表达构建体的宿主细胞的选择标记,和将此构建体引入所研究的植物所需的DNA序列(后者取决于使用的DNA引入方法)。

对调控序列,如启动子和终止序列和任选地信号或转运序列的选择,举

例来说要根据需要多肽在何时、何处和如何进行表达来确定。例如，编码本发明多肽的基因表达可以是组成型的或可诱导的；或者可以是发育、阶段特异性的或组织特异性的；基因产物可以靶向特异性的组织或植株部分如种子或叶。例如，Tague 等，1988，植物生理学 (Plant Physiology) 86: 503 对调控序列进行了说明。

对于组成型表达来说，可使用 35S - CaMV 启动子 (Franck 等，1980，细胞 (Cell) 21: 285 - 294)。器官特异性启动子可以是例如，来自贮藏库组织如种子、马铃薯块茎和果实的启动子 (Edwards&Coruzzi, 1990，遗传学综述年鉴 (Ann.Rev.Genet.) 24: 275 - 303)，或者是来自代谢库组织如分生组织的启动子 (Ito 等，1994，植物分子生物学 (Plant Mol.Biol.) 24: 863 - 878)，种子特异性启动子如水稻的谷蛋白、谷醇溶蛋白、球蛋白或清蛋白启动子 (Wu 等，1998，植物和细胞生理学 (Plant and Cell Physiology) 39: 885 - 889)，豆球蛋白 B4 的蚕豆 (Vicia faba) 启动子和蚕豆中未知的种子蛋白基因的启动子 (Conrad 等，1998，植物生理学杂志 (Journal of Plant Physiology) 152: 708 - 711)，种子油体蛋白的启动子 (Chen 等，1998，植物和细胞生理学 39: 935-941)，欧洲油菜 (Brassica napus) 的贮藏蛋白 napA 启动子，或者如在 WO91/14772 中描述的本领域已知的任何其它种子特异性启动子。此外，启动子可以是叶特异性启动子如水稻或番茄的 rbcS 启动子 (Kyoizuka 等，1993，植物生理学 (Plant Physiology) 102: 991 - 1000)，小球藻病毒腺嘌呤甲基转移酶基因启动子 (Mitra 和 Higgins, 1994，植物分子生物学 (Plant Molecular Biology) 26:85-93)，或水稻 aldP 基因启动子 (Kagaya 等，1995，分子和普通遗传学 (Molecular and General Genetics) 248: 668 - 674)，或者伤口诱导的启动子如马铃薯 pin2 启动子 (Xu 等，1993，植物分子生物学 22: 573 - 588)。

还可利用启动子增强子元件在植物中获得酶的更高表达。举例来说，启动子增强子元件可以是位于启动子和编码本发明多肽的核苷酸序列之间的内含子。如 Xu 等 (1993，同上引文) 公开了利用水稻肌动蛋白 1 基因的第一个内含子增强表达。

选择标记基因和表达构建体的任何其它部分可从本领域可利用的那些中挑选。

利用本领域已知的常用技术可将核酸构建体并入植物基因组，这些技术包括土壤杆菌（*Agrobacterium*）介导的转化、病毒介导的转化、微注射、粒子轰击、生物轰击转化和电穿孔（Gasser 等，1990，*Science* 244: 1293; Potrykus, 1990, *Bio/Technology* 8: 535; Shimamoto 等，1989, *Nature* 338: 274）。

目前，根癌土壤杆菌（*Agrobacterium tumefaciens*）介导的基因转移是所选的用于产生转基因双子叶植物的方法（综述参见 Hooykas 和 Schilperoort, 1992, *植物分子生物学* 19: 15-38）。然而此方法也可以转化单子叶植物，虽然对于单子叶植物一般优选其它的转化方法。目前所选的产生转基因单子叶植物的方法是粒子轰击胚性愈伤组织或正在发育的胚胎（用转化 DNA 包被的金或钨微粒）（Christou, 1992, *植物杂志 (Plant Journal)* 2: 275 - 281; Shimamoto, 1994, *Current Opinion Biotechnology*, 5: 158 - 162; Vasil 等，1992, *Bio/Technology* 10:667-674）。如 Omirulleh 等，1993, *植物分子生物学* 21: 415 - 428 所述，转化单子叶植物的一个替代方法是基于原生质体转化来进行。

按照本领域熟知的方法，在转化之后，挑选出掺入表达构建体的转化体并使其再生成完整植株。

本发明还涉及用于生产本发明多肽的方法，其包括（a）在有利于本多肽产生的条件下，对含有编码本发明具纤维二糖酶活性的多肽的核苷酸序列的转基因植物或植物细胞进行培养；并且（b）回收本多肽。

组合物

在另一方面，本发明涉及包含本发明多肽的组合物。

本组合物可包含作为主要酶组分的本发明多肽，如单组分组合物。或者，本组合物可包含多种酶活性，如氨肽酶、淀粉酶、糖酶、羧肽酶、过氧化氢酶、纤维素酶、几丁质酶、角质酶（cutinase）、环糊精糖基转移酶、脱氧核糖核酸酶、酯酶、 α -半乳糖苷酶、 β -半乳糖苷酶、葡糖淀粉酶、 α -葡

糖苷酶、 β -葡萄糖苷酶、卤代过氧化物酶(haloperoxidase)、转化酶、漆酶、脂肪酶、甘露糖苷酶、氧化酶、果胶分解酶、肽谷氨酰胺酶(peptidoglutaminase)、过氧化物酶、植酸酶、多酚氧化酶、蛋白水解酶、核糖核酸酶、转谷氨酰胺酶或木聚糖酶。

本组合物可按照本领域已知的方法进行制备并且可以以液体或干组合物形式存在。举例来说,本多肽组合物可以是颗粒或微粒形式。包含在本组合物中的多肽可按照本领域已知的方法进行稳定。

下文给出本发明多肽组合物的优选用途实例。本发明多肽组合物的剂量和其它使用条件可根据本领域已知的方法进行确定。

去污剂组合物

本发明的纤维二糖酶可被添加到去污剂组合物中并因此成为其组分。

例如,可以将本发明的去污剂组合物配制成手洗或机洗洗衣去污剂组合物,包括适用于预处理污染织物的洗衣添加剂组合物和漂洗时添加的织物柔顺组合物,或配制成用于普通家庭硬表面清洁处理的去污剂组合物,或适用于手洗或机洗碗碟操作的组合物。

在一个特定方面,本发明提供了含本发明纤维二糖酶的去污剂添加剂。此去污剂添加剂以及去污剂组合物中可包含一种或多种其它的酶如蛋白酶、脂肪酶、角质酶、淀粉酶、糖酶、纤维素酶、果胶酶、甘露聚糖酶、阿拉伯聚糖酶、半乳聚糖酶、木聚糖酶,氧化酶,如漆酶和/或过氧化物酶。

一般来说,所选择的酶的性质应该与所选去污剂相兼容,(即最适 pH,与其它酶的或非酶的组分的相容性等),并且酶应该以有效量存在。

蛋白酶: 适宜的蛋白酶包括动物,植物或微生物来源的蛋白酶,优选微生物来源的蛋白酶。还包括化学修饰过的或蛋白质工程化的突变体。蛋白酶可以是丝氨酸蛋白酶或金属蛋白酶,尤其是碱性微生物蛋白酶或胰蛋白酶样蛋白酶。碱性蛋白酶的实例有枯草杆菌蛋白酶,尤其来源于芽孢杆菌的那些,如枯草杆菌蛋白酶 Novo,枯草杆菌蛋白酶 Carlsberg,枯草杆菌蛋白酶 309,枯草杆菌蛋白酶 147 和 枯草杆菌蛋白酶 168 (说明参见 WO 89/06279)。胰蛋白酶样蛋白酶的实例有胰蛋白酶(如猪或牛源的)和镰孢霉蛋白酶(对其

描述参见 WO 89/06270 和 WO 94/25583)。

有用的蛋白酶的实例是在 WO 92/19729, WO 98/20115, WO 98/20116 和 WO 98/34946 中描述的变体,尤其是在一个或多个如下位点发生替代的变体: 27,36,57,76,87,97,101,104, 120,123,167,170,194,206,218,222, 224,235 和 274。

优选的商业化的蛋白酶包括 ALCALASETM, SAVINASETM, PRIMASETM, EVerlaseTM, ESPERASETM 和 KANNASETM(Novo Zymes A/S), MAXATASETM, MAXACALTM, MAXAPEMTM, PROPERASETM, PURAFECTTM, PURAFECT OXPTM, FN2TM, 和 FN3TM(Genencor International Inc.)。

脂肪酶: 适宜的脂肪酶包括细菌和真菌来源的脂肪酶。还包括化学修饰过的或蛋白质工程化的突变体。有用的脂肪酶的实例包括腐质霉属(同义词 *Thermomyces*), (如 EP 258 068 和 EP 305 216 中描述的)*H. lanuginosa* (*T. lanuginosus*)或(如 WO 96/13580 中描述的)*H. insolens* 的脂肪酶;假单胞菌属,如产碱假单胞菌(*P. alcaligenes*)或类产碱假单胞菌(*P. pseudoalcaligenes*)(EP 218 272),洋葱假单胞菌(*P. cepacia*)(EP 331 376),施氏假单胞菌(*P. stutzeri*)(GB 1,372,034),荧光假单胞菌(*P. fluorescens*),假单胞菌属菌株 SD 705 (WO 95/06720 和 WO 96/27002), *P. wisconsinensis* (WO 96/12012) 的脂肪酶;芽孢杆菌属,如枯草芽孢杆菌 (Dartois 等(1993), *Biochemica et Biophysica Acta*, 1131,253-360),嗜热脂肪芽孢杆菌 (JP 64/744992)或短小芽孢杆菌 (*B. pumillus*)(WO 91/16422) 的脂肪酶。

其它实例有例如在 WO 92/05249, WO 94/01541, EP 407 225, EP 260 105,WO 95/35381, WO 96/00292, WO 95/30744, WO 94/25578, WO 95/14783, WO 95/22615, WO 97/04079 和 WO 97/07202 中描述的脂肪酶变体。

优选的商业化的脂肪酶包括 LIPOLASETM 和 LIPOLASE ULTRATM (Novo Zymes A/S)。

淀粉酶: 适宜的淀粉酶(α 和/或 β)包括细菌和真菌来源的淀粉酶。还包括化学修饰过的或蛋白质工程化的突变体。淀粉酶包括,举例来说,芽孢杆菌属如地衣芽孢杆菌的特定菌株的 α -淀粉酶,详细说明参见 GB

1,296,839.

有用的淀粉酶的实例有在 WO 94/02597, WO 94/18314, WO 96/23873 和 WO 97/43424 中描述的变体,特别是在一个或多个如下位点发生替代的变体: 15,23,105,106,124,128,133,154,156,181,188,190,197,202,208,209,243,264,304,305, 391,408 和 444。

商业化的淀粉酶有 DURAMYL™, TERMAMYL™, FUNGAMYL™ 和 BAN™ (Novo Zymes A/S), RAPIDASE™ 和 PURASTAR™ (来自 Genencor International Inc.)。

纤维素酶: 适宜的纤维素酶包括细菌和真菌来源的纤维素酶。还包括化学修饰过的或蛋白质工程化的突变体。适宜的纤维素酶包括芽孢杆菌属, 假单胞菌属, 腐质霉属, 镰孢属, 梭孢壳属, 枝顶孢属的纤维素酶, 例如在 US 4,435,307, US 5,648,263, US 5,691,178, US 5,776,757 和 WO 89/09259 中公开的由 *Humicola insolens*, 嗜热毁丝霉和尖镰孢产生的真菌纤维素酶。

尤其适宜的纤维素酶是有益于颜色维护的碱性或中性纤维素酶。这样的纤维素酶的实例有 EP 0 495 257, EP 0 531 372, WO 96/11262, WO 96/29397, WO 98/08940 中描述的纤维素酶。其它实例有如在 WO 94/07998, EP 0 531 315, US 5,457,046, US 5,686,593, US 5,763,254, WO 95/24471, WO 98/12307 和 PCT/DK98/00299 中描述的纤维素酶变体。

商业化的纤维素酶包括 CELLUZYME™, 和 CAREZYME™ (Novo Zymes A/S), CLAZINASE™, 和 PURADAX HA™ (Genencor International Inc.), 和 KAC-500 (B)™ (Kao Corporation)。

过氧化物酶/氧化酶: 适宜的过氧化物酶/氧化酶包括植物, 细菌或真菌来源的过氧化物酶/氧化酶, 还包括化学修饰过的或蛋白质工程化的突变体。有用的过氧化物酶的实例包括鬼伞属 (*Coprinus*) 如灰盖鬼伞 (*C. cinereus*) 的过氧化物酶, 及其变体, 如在 WO 93/24618, WO 95/10602, 和 WO 98/15257 中描述的那些。

可以通过添加含一种或多种酶的各单独添加剂, 或通过添加包含所有这些酶的组合添加剂, 使去污剂组合物中包含这些去污剂酶。本发明的去污剂

添加剂，即单独添加剂或组合添加剂可以以如粒状、液体、浆液等形式配制。优选的去污剂添加剂制剂是颗粒剂，尤其是无尘颗粒；液体，尤其是稳定液体；或浆液。

无尘颗粒可如 US4,106,991 和 4,661,452 中的公开进行生产并可任选地利用本领域已知的方法加包衣。蜡质包衣材料的实例有平均摩尔重量 1000 - 20000 的聚（环氧乙烷）产物（聚乙二醇，PEG）；具有 16 - 50 个环氧乙烷单位的乙氧基化壬基酚；具有 15 - 80 个环氧乙烷单位的乙氧基化脂肪醇，其中醇含 12 - 20 个碳原子；脂肪醇；脂肪酸；以及脂肪酸的单 - 和二 - 和三甘油酯。GB1483591 中给出了适于通过流化床技术来使用的成膜包衣材料的实例。举例说来，液体酶制备物可按照成熟方法通过添加多元醇如丙二醇、糖或糖醇、乳酸或硼酸来稳定化。受保护的酶可按照 EP238, 216 中公开的方法制备。

本发明的去污剂组合物可以采取任一种方便形式，如条状、片状、粉末、颗粒、糊或液体。液体去污剂可以是水性的，一般含高达 70% 的水和 0 - 30 % 的有机溶剂，或者是非水性的。

去污剂组合物可以包含一种或多种表面活性剂，其可以是非离子的（包括半极性的）和/或阴离子的和/或阳离子的和/或两性离子的表面活性剂。表面活性剂的含量按重量计一般是 0.1% 到 60%。

当去污剂中包括阴离子表面活性剂时，去污剂通常包含大约 1% 到大约 40% 的阴离子表面活性剂，如直链烷基苯磺酸盐， α -烯属磺酸盐，烷基硫酸盐（脂肪醇硫酸盐），脂肪醇乙氧基硫酸盐，仲链烷磺酸盐， α -磺基脂肪酸甲酯，烷基 - 或链烯基丁二酸或皂。

当去污剂中包含非离子表面活性剂时，去污剂通常含有约 0.2% 到约 40% 的非离子表面活性剂，例如乙氧基化醇、乙氧基化壬基酚、烷基多苷、烷基二甲基胺氧化物、乙氧基化脂肪酸单乙醇酰胺、脂肪酸单乙醇酰胺、多羟基烷基脂肪酸酰胺或葡糖胺的 N-酰基 N-烷基衍生物（“葡糖酰胺”）。

去污剂可以包含 0-65% 的去污剂助洗剂或络合剂，例如沸石、二磷酸盐、三磷酸盐、膦酸盐、碳酸盐、柠檬酸盐、次氨基三乙酸、乙二胺四乙酸、二

亚乙基三胺五乙酸、烷基-或链烯基琥珀酸、可溶性硅酸盐或层叠式硅酸盐(例如购自 Hoechst 的 SKS-6)。

去污剂可以包含一种或多种聚合物。实例为羧甲基纤维素、聚(乙烯吡咯烷酮)、聚(乙二醇)、聚(乙烯醇)、聚(乙烯基吡啶-N-氧化物)、聚(乙烯基咪唑)、聚羧酸酯如聚丙烯酸酯、马来酸/丙烯酸共聚物和甲基丙烯酸月桂醇酯/丙烯酸共聚物。

去污剂可包含漂白系统,该系统可包含 H2O2 源如过硼酸盐或过碳酸盐,并可以同形成过酸的漂白活化剂如四乙酰基乙二胺或壬酰氧基苯磺酸盐进行组合。或者,漂白系统可包含如酰胺,二酰亚胺,或矾类型的过氧酸。

本发明的去污剂组合物中的酶可以通过诸如以下的常规稳定剂稳定:多元醇如丙二醇或丙三醇、糖或糖醇、乳酸、硼酸或硼酸衍生物如芳香硼酸酯或苯基硼酸衍生物如 4-甲酰基苯基硼酸,且所述的组合物也可以按 WO 92/19709 和 WO 92/19708 中所述配制。

本去污剂还可包含其它常规去污剂成分,例如织物调理剂(包括黏土)、增泡剂、泡沫抑制剂、抗腐蚀剂、污物悬浮剂、抗污物再沉淀剂、染料、杀菌剂、荧光增白剂、水溶助长剂、酶抑制剂或香料。

目前认为在去污剂组合物中任何酶,尤其是本发明的纤维二糖酶都可以以相当于每升洗涤液中具有 0.01-100mg 酶蛋白质,优选每升洗涤液 0.05-5 mg 酶蛋白质,尤其是每升洗涤液中具有 0.1-1 mg 酶蛋白质的量加入。

此外,本发明的纤维二糖酶可以添加到如 WO 97/07202 所公开的去污剂制剂中,该文献特此引入作为参考。

DNA 重组(改组)

如 SEQ ID NO: 1 所示的核苷酸序列可用于 DNA 重组(或改组)过程。如此得到的新多核苷酸序列可编码具有如下改良性质的新纤维二糖酶活性多肽:如改良的稳定性(贮藏稳定性,热稳定性)、改良的比活性、改良的 pH 最适值,和/或改良的对特定化合物的耐受性。

两个或多个同源输入多核苷酸(起始点多核苷酸)之间的改组包括使多

核苷酸片段化并重组这些片断，获得与输入多核苷酸相比有许多核苷酸片断发生交换的输出多核苷酸（即已经过一个改组循环的多核苷酸）。

DNA 重组或改组可以从两个或多个起始基因产生出一个嵌合基因文库的（部分的）随机过程。可利用许多已知的方式来实现这样的改组或 DNA 重组。

此方法可涉及亲本 DNA 的随机片断化，然后通过 PCR 重新组装成新的全长基因，如在 US5605798, US5811238, US5830721, US6117679 中的说明。体外基因重组可按照如 US6159687, WO98/41623, US6159688, US5965408, US6153510 中的说明进行。此重组过程可在活细胞体内进行，如 WO97/07205 和 WO98/28416 中的说明。

如 Kikuchi 等（2000a, Gene 236: 159 - 167）的说明，亲本 DNA 可通过 DNA'se I 处理或限制性内切酶消化而片段化。两个亲本的改组可通过改组两亲本的单链亲本 DNA 来完成，如 Kikuchi 等（2000b, Gene 243: 133 - 137）的说明。

一个具体的改组方法是按照 Cramer 等，1998, Nature, 391: 288 - 291 和 Ness 等，Nature Biotechnology 17: 892 - 896 中说明的方法操作。另一方式是如 US6159687 中：实例 1 和 2 所述的方法。

本发明通过下列实施例进一步说明，这些实施例不应视为限制本发明的范围。

实施例

用作缓冲剂和底物的化学制品是至少试剂纯的上市产品。

纤维二糖酶试验

PNP 葡萄糖加终止试剂

底物溶液：5mM PNP β - D - 葡萄糖（Sigma N - 7006）底物置于 0.1M 醋酸钠缓冲液中，pH4.0；

终止试剂：0.1M 碳酸钠，pH 11.5。

50 μ l 纤维二糖酶溶液与 1ml 底物溶液混合并在 40℃ 下孵育 20 分钟。反

应通过加入 5ml 终止试剂而终止。测量 404nm 的吸光度。

纤维二糖的降解 (CBU 分析)

纤维二糖酶水解纤维二糖中的 β - 1, 4 键, 释放出两个葡萄糖分子。释放的葡萄糖量可利用适当的葡萄糖分析方法, 如己糖激酶法测定 (如 Bergmeyer 等: D - 葡萄糖, 利用己糖激酶和葡萄糖 - 6 - 磷酸脱氢酶的测定, 在 Bergmeyer HU, Gawehn K. 编《Methods of Enzymatic Analysis》New York: Academic Press; 1974: 196-201 中; 或 Kunst, A., Draeger, B. & Ziegenhorn, J. (1984)《Methods of Enzymatic Analysis》(Bergmeyer, HU, ed), 第 3 版, Vol. 6, pp. 163-172, VCH, Weinheim, W. Germany-Deer-field), 其可购自 Roche (# 127183)。一个纤维二糖酶单位 (CBU) 是以纤维二糖为底物, 在 40℃, pH 5 下每分钟释放出 2 μ mol 葡萄糖的酶量。

将纤维二糖酶样品在 0.1M 醋酸缓冲液, pH 5.0 中稀释到大约 0.05 - 0.25 CBU/ml。0.5ml 此溶液与 2.5ml 纤维二糖底物溶液 (0.2g D - (+) 纤维二糖, Sigma C - 7252, 于 0.1M 醋酸缓冲液中, pH 5.0) 混合并以 40℃ 加热 15 分钟。反应通过加入 300 μ l 1N 高氯酸终止并测量释放的葡萄糖。

米曲霉纤维二糖酶的比活性是每 mg 蛋白质 150 CBU, 黑曲霉纤维二糖酶的比活性是每 mg 蛋白质 10 CBU。

材料

菌株

BECh2: 本菌株的构建参见 WO00/39322。

JaL406: 本菌株的构建参见实施例 2。

质粒

pYES2.0: 购自 Invitrogen Corporation, San Diego, CA, USA;

pJaL621: 本质粒的构建参见实施例 1。

pJaL660: 本质粒的构建参见实施例 2。

pMT2188: 本质粒的构建参见实施例 2。

pCaHj527: 本质粒的构建参见 WO 00/70064。

实施例 1

米曲霉纤维二糖酶 (β - 葡糖苷酶) 的克隆

米曲霉菌株 IFO4177 的定向 cDNA 文库的构建

以尿素和酵母提取物为氮源使米曲霉菌株 IFO4177 以 10 升规模在 20 升实验室发酵罐中生长, 分批培养基中以右旋糖为碳源, 而补料中以 nutritose (麦芽糖浆) 为碳源。发酵罐中生长培养基的组成是 (g/l): 右旋糖 27.5, 酵母提取物 5.0, $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ 2.0, K_2SO_4 3.0, KH_2PO_4 2.0, 柠檬酸 4.0, 尿素 5.0 及微量元素 0.5ml/l。真菌菌丝体在 30℃ 下生长 68.3 小时后进行收获, 立即在液氮中冷冻并在 -80℃ 贮藏。

总 RNA 的提取

总 RNA 利用硫氰酸胍提取, 再利用超离心通过 5.7M CsCl 垫 (Chirgwin 等, 1979 Biochemistry 18: 5294 - 5299) 来制取, 该方法具有如下改良。在液氮中用研钵和研杵将冷冻菌丝体磨成细粉末, 然后在预冷的咖啡豆磨机中磨碎, 立即悬浮于 5 倍体积的 RNA 提取缓冲液中 (4M GuSCN, 0.5% 十二烷基肌氨酸钠, 25mM 柠檬酸钠, pH7.0, 0.1M β - 巯基乙醇)。在室温下搅拌混合物 30 分钟, 离心 (20 分钟, 10000rpm, Beckman) 沉淀细胞碎片。收集上清液, 按照每 12.0ml CsCl 垫取用 26.5ml 上清液, 小心地将上清液铺在 5.7M CsCl 垫层 (5.7M CsCl, 10mMEDTA, pH7.5, 0.1% DEPC; 使用前高压灭菌) 上分层, 然后离心得到总 RNA (Beckman, SW28 转头, 25000rpm, 室温, 24 小时)。离心后小心除去上清液, 切下含总 RNA 沉淀的试管底部并以 70% EtOH 漂洗。将总 RNA 沉淀转移到 Eppendorf 管, 悬浮于 500ml TE, pH7.6 中 (如果有难度, 可在 65℃ 下间断地加热 5 分钟), 用酚抽提并在 -20℃ 下用乙醇进行 12 小时的沉淀 (2.5 体积 EtOH, 0.1 体积 3M NaAc, pH5.2)。RNA 经离心收集, 在 70% 的 EtOH 中洗涤后重新悬浮于最小体积的 DEPC - DIW 中。RNA 浓度通过测量 OD260/280 来确定。

poly (A) + RNA 的分离

poly (A) + RNA 通过寡聚 (dT) - 纤维素亲和层析分离 (Aviv&Leder,1972, Proc.Natl.Acad.Sci.U.S.A.69:1408-1412)。通常,将 0.2g 寡聚 (dT) - 纤维素 (Boehringer Mannheim) 在 10 ml 的 1X 柱上样缓冲液 (20mM Tris - Cl, pH7.6, 0.5M NaCl, 1mM EDTA, 0.1% SDS) 中预溶胀,然后加到 DEPC - 处理的加盖塑料柱上 (Poly Prep Chromatography Column, Bio Rad), 并以 20ml 的 1x 上样缓冲液平衡。总 RNA (1 - 2mg) 在 65℃ 下加热 8 分钟, 在冰上淬火 5 分钟, 并将 1 体积 2x 柱上样缓冲液加入到 RNA 样品中之后, 上柱。收集洗脱物并重新加样 2 - 3 次, 而且每次加样前都如上述加热样品并在冰上淬火。寡聚 (dT) 柱用 10 体积的 1x 上样缓冲液洗涤, 再用 3 体积的中间盐缓冲液 (20mM Tris - Cl, pH7.6, 0.1M NaCl, 1mM EDTA, 0.1% SDS) 洗涤, 之后用预热到 + 65℃ 的 3 体积洗脱缓冲液 (10mM Tris - Cl, pH7.6, 1mM EDTA, 0.05% SDS) 洗脱 poly (A) + RNA, 最后收集 500 μ l 的洗脱级分。读取收集的每一级分的 OD260 值, 将含 mRNA 的级分合并并在 - 20℃ 用乙醇沉淀 12 小时。离心收集 poly (A) + RNA, 重新悬浮于 DEPC - DIW 并在 - 80℃ 以 5 - 10 μ g 等分试样贮藏。

cDNA 的合成

利用 F.S.Hagen(私人交流)研发的发夹改良法, 通过 RNA 酶 H 法 (Gulber&Hoffman 1983 Gene 25:263-269, Sambrook 等, 1989 分子克隆: 实验室手册. Cold Spring Harbor Lab., Cold Spring Harbor NY) 从 5 μ g 米曲霉 A1560 的 poly (A) + RNA 合成出双链 cDNA。将 poly (A) + RNA (5 μ l DEPC-处理过的水中 5 μ g) 置于预先硅化的无 RNA 酶的 Eppendorph 管中, 在 70℃ 下加热 8 分钟, 冰上淬火, 并与含 1mM dATP, dGTP 和 dTTP 以及 0.5mM 5 - 甲基 - dCTP (Pharmacia)、40 单位的人胎盘核糖核酸酶抑制剂 (Rnasin, Promega)、4.81 μ g 寡聚 (dT) 18 - NotI 引物 (Pharmacia) 和 1000 单位的 SuperScriptII RNase H 逆转录酶 (Bethesda Research Laboratories) 的逆转录酶缓冲液 (50mM Tris - Cl, pH8.3, 75mM KCl, 3mM MgCl₂, 10mM DTT, Bethesda Research Laboratories) 混合, 终体积为

50 μ l。反应混合物在 45℃ 下孵育 1 小时合成 cDNA 第一链。合成之后，按照厂商说明书对 mRNA: cDNA 杂合体混合物通过 MicroSpin S-400 HR (Pharmacia) 离心柱进行凝胶过滤。

凝胶过滤之后，杂合体在含 200 μ M 各种 dNTP、60 单位大肠杆菌 DNA 聚合酶 I (Pharmacia)、5.25 单位 RNase H (Promega) 和 15 单位大肠杆菌 DNA 连接酶 (Boehringer Mannheim) 的 250 μ l 第二链缓冲液 (20mM Tris - Cl, pH7.4, 90mM KCl, 4.6mM MgCl₂, 10mM (NH₄)₂SO₄, 0.16mM β NAD⁺) 中稀释。反应试管在 16℃ 下孵育 2 小时，25℃ 下再孵育 15 分钟进行第二链 cDNA 的合成。反应通过加入终浓度为 20mM 的 EDTA 终止，再进行苯酚和氯仿提取。

双链(ds)cDNA 通过加入 2 体积的 96% EtOH, 0.2 体积的 10M NH₄Ac, 在 -20℃ 下放置 12 小时进行乙醇沉淀，经离心回收，在 70% EtOH 中洗涤，干燥(Speed Vac),再重新悬浮于 30 μ l 的含 25 单位绿豆核酸酶 (Pharmacia) 的绿豆核酸酶缓冲液中 (30mM NaAc, pH4.6, 300mM NaCl, 1mM ZnSO₄, 0.35mM DTT, 2% 甘油)。单链发夹 DNA 的修剪通过下列步骤进行：使反应物在 30℃ 下孵育 30 分钟，然后加入 70 μ l 10mM Tris - Cl, pH7.5、1mM EDTA，苯酚提取，以 2 体积 96% EtOH 和 0.1 体积 3M NaAc, pH5.2 置于冰上 30 分钟进行乙醇沉淀。

ds cDNA 经离心 (20000rpm, 30 分钟) 回收，并与 30 μ l 的 T4 DNA 聚合酶缓冲液 (20mM Tris - 醋酸, pH7.9, 10mM MgAc, 50mM Kac, 1mM DTT) (其中包含 0.5mM 各种 dNTP 和 5 单位 T4 DNA 聚合酶) 中的 T4 DNA 聚合酶进行反应，反应混合物在 +16℃ 下孵育 1 小时，产生出平端。反应通过加入终浓度为 20mM 的 EDTA 而终止，然后进行苯酚和氯仿提取，加入 2 体积 96% EtOH 和 0.1 体积 3M NaAc, pH5.2 置于 -20℃ 下 12 小时进行乙醇沉淀。

补平反应之后，cDNA 经离心 (同上) 回收，在 70% EtOH 中洗涤，DNA 沉淀在 SpeedVac 中干燥。将 cDNA 沉淀重新悬浮于含 2 μ g EcoR I 衔接物 (0.2 μ g/ μ l, Pharmacia) 和 20 单位 T4 连接酶 (Promega) 的 25 μ l 连接缓冲液

(30mM Tris - Cl, pH7.8, 10mM MgCl₂, 10mM DTT, 0.5mM ATP) 中, 反应混合物在 +16℃ 下孵育 12 小时。反应通过 65℃ 加热 20 分钟, 然后在冰上放置 5 分钟而终止。带衔接物的 cDNA 通过加入 20 μl 高压灭菌水, 5 μl 的 10xNotI 限制性酶缓冲液 (New England Biolabs) 和 50 单位的 Not I (New England Biolabs), 然后 +37℃ 孵育 3 小时, 而实现 NotI 限制性酶消化。反应通过 65℃ 加热样品 15 分钟而终止。利用琼脂糖凝胶电泳 (在 1 × TBE (于高压灭菌水中) 中于 0.8 % SeaPlaque GTG 低熔点琼脂糖凝胶上) 对 cDNAs 进行大小分级, 分离出未连接的衔接物和小 cDNA。凝胶电泳在 15V 运行 12 小时, 通过在 0.7kb 位置处切下琼脂糖凝胶的下部对 cDNA 进行大小选择。然后将 1.5 % 琼脂糖凝胶倾倒在含 cDNA 的凝胶的前面, 反向进行凝胶电泳浓缩 ds cDNA, 直到其在凝胶上显示为一个压缩带。从凝胶上切下含 cDNA 的凝胶块并利用 GFX 凝胶带纯化试剂盒 (Amersham-Pharmacia) 按照如下步骤从凝胶中提取 cDNA。将修剪好的凝胶切块置于 2ml Biopure Eppendorf 管中称重, 然后以每 10mg 凝胶切块加入 10ml 的 Capture Buffer, 凝胶切块的溶解通过 60℃ 孵育 10 分钟直到琼脂糖完全溶解来实现, 样品经瞬时离心而位于管底部。将熔融样品转移到收集管中放置的 GFX 离心柱上, 25℃ 孵育 1 分钟, 然后在台式离心机中全速旋转 30 秒。除去流出液, 以 500 μl 的洗涤缓冲液洗涤柱子, 然后再全速离心 30 秒。

丢掉收集管, 将柱子放置在 1.5ml 的 Eppendorf 管中, 再将 50 μl TE, pH7.5 加到柱子中心, 25℃ 孵育 1 分钟, 最后以最大速度离心 1 分钟洗脱 cDNA。将洗脱得到的 cDNA 贮藏在 -20℃ 以待文库的构建。

制备 EcoRI/NotI 切割的载体用于文库构建

按厂商操作说明书利用 Qiagen Tip - 100 对用于制备含 EcoRI - NotI 插入片断的 pYES 2.0 cDNA 克隆的质粒 DNA 进行纯化。10 μg 纯化的质粒 DNA 在加有 6 μl 用于 EcoRI 的 10xNE 缓冲液 (New England Biolabs)、40 单位 NotI (New England Biolabs)、和 20 单位 EcoRI (New England Biolabs) 的总体积 60 μl 混合液中 37℃ 孵育 6 小时, 而实现 EcoRI 和 NotI 的完全消化。

反应通过 65℃加热样品 20 分钟而终止。消化的质粒 DNA 用苯酚-氯仿提取一次，然后用氯仿提取一次，再通过加入 2 体积 96% EtOH 和 0.1 体积 3M NaAc, pH 5.2 在 -20℃放置 12 小时进行乙醇沉淀。将沉淀 DNA 重新悬浮于 25 μ l 1xTE, pH 7.5 中，加载到 1xTBE（在高压灭菌水中）中的 0.8% SeaKem 琼脂糖凝胶上，在 60V 跑凝胶 3 小时。从凝胶上切下消化的载体，利用 GFX 凝胶带纯化试剂盒（Amersham - Pharmacia Biotech）按照使用说明从凝胶上提取 DNA。用 OD260/280 测量 DNA 浓度之后，将洗脱得到的载体贮藏在 -20℃以待文库构建。

米曲霉 IFO4177 cDNA 文库的构建

为建立 cDNA 文库的最适连接条件，在含 7 μ l ds cDNA（相当于 cDNA 样品的总体积的大约 1/10），2 单位 T4 连接酶（Promega）和分别 25ng、50ng 和 75ng 的 EcoRI - NotI 切割的 pYES2.0 载体（Invitrogen）的 10 μ l 连接缓冲液（30mM Tris - Cl, pH 7.8, 10mM MgCl₂, 10mM DTT, 0.5mM ATP）中进行了四个连接试验。载体背景对照连接反应中含有 75ng 的 EcoRI - NotI 切割的 pYES2.0 载体而不含 cDNA。连接反应通过 16℃孵育 12 小时，65℃加热 20 分钟来进行，然后向每个试管中加入 10 μ l 的高压灭菌水。1 μ l 的连接混合物经电穿孔（200W, 2.5kV, 25mF）进入 40 μ l 电感受态的大肠杆菌 DH10B 细胞（Bethesda Research Laboratories）中。向每个转化混合物中加入 1ml SOC 之后，使细胞在 +37℃生长 1 小时，从每个电穿孔反应物中取 50 μ l 和 5 μ l 铺在 LB + 氨苄青霉素平板（100mg/ml）上并在 37℃生长 12 小时。利用最适条件，在大肠杆菌中建立了含 2.5×10^7 个独立菌落形成单位的米曲霉（*A.oryzae*）A1560 cDNA 文库，载体背景为大约 1%。cDNA 文库以多个单独库（25000c.f.u./库）的形式在 20% 甘油中，-80℃进行贮藏。

利用 Qiaprep 系统从各单菌落中分离出质粒 DNA 并按照厂商使用说明在 ABI 3700 毛细管测序仪上进行序列测定。对比这些序列与 SWISSPROT 数据库，鉴定到命名为 pJaL621 的 cDNA 克隆能够编码纤维二糖酶。对 2771bp（碱基对）的 cDNA 克隆（SEQ ID NO: 1）测序揭示一个编码计算分子量为 93, 437 Da 的 861 个氨基酸的多肽（SEQ ID NO: 2）的开放阅读框。

实施例 2

在米曲霉中表达米曲霉纤维二糖酶

米曲霉纤维二糖酶表达质粒的构建

曲霉属表达质粒 pCaHj527 (公开于 WO 00/70064) 由一个表达序列盒组成, 此表达序列盒基于黑曲霉中性淀粉酶 II 启动子与构巢曲霉丙糖磷酸异构酶非翻译前导序列 (Pna2/tpi) 及黑曲霉淀粉糖苷酶终止子 (Tamg) 的融合。质粒中还存在使得菌株能以乙酰胺为唯一氮源进行生长的来自构巢曲霉的曲霉属选择标记 amdS 和能使 pyrF 缺陷型埃希氏大肠杆菌 (*Escherichia coli*) 菌株 DB6507 (ATCC 35673) 生长的酿酒酵母 URA3 标记。

利用酿酒酵母 (*S.cerevisiae*) URA3 基因为选择标记转化大肠杆菌 DB6507 的步骤如下:

利用 Mandel 和 Higa 方法 (Mandel, M. 和 A.Higa (1970) J.Mol.Biol.45, 154) 使大肠杆菌 DB6507 处于感受态。从补充有 1g/l 酪蛋白氨基酸、500 μ g/l 硫胺素和 10mg/l 卡那霉素的固体 M9 培养基 (Sambrook 等 (1989) 分子克隆, 实验室手册, 2 edition, Cold Spring Harbor Laboratory Press) 上挑选出转化体。

pCaHj527 的修饰步骤如下:

- 利用简单 PCR 法使存在于 pCaHj527 上的 Pna2/tpi 启动子发生定点突变;
- 利用突变引物 141223 (SEQ ID NO: 5) 使核苷酸第 134 - 144 位从 SEQ ID NO: 3 转变为 SEQ ID NO: 4;
- 利用突变引物 141222 (SEQ ID NO: 8) 使核苷酸第 423 - 436 位从 SEQ ID NO: 6 转变为 SEQ ID NO: 7。

所得质粒命名为 pMT2188。

将纤维二糖酶基因克隆到 pMT2188 的步骤如下:

利用下列两个寡聚核苷酸引物: B2902E12 (SEQ ID NO:9) 和 B2902F01 (SEQ ID NO:10) 通过 PCR 扩增米曲霉纤维二糖酶编码区。为便于克隆, 将限制性酶切位点插入每个引物的 5' 末端, 引物 B2902E12 含一个 BglII 位点, 引物 B2902F01 含一个 XhoI 位点。

米曲霉 cDNA 克隆 pJaL621 被用作 PCR 反应的模版。反应在含 2.5 单位 Taq 聚合酶, 100ng pJaL621, 250nM 的各种 dNTP, 以及如上所述两种引物各 10pmol 的 100 μ l 体积反应缓冲液 (50mM KCl, 10mM Tris - Cl pH8.0, 1.5mM MgCl₂) 中进行。

扩增在 Perkin - Elmer Cetus DNA Thermal 480 中进行, 由 94℃ 3 分钟一个循环, 之后 25 个循环每个循环为 94℃ 1 分钟、55℃ 30 秒和 72℃ 1 分钟组成。PCR 反应产生一个 2615bp 长的 DNA 片断。此片断用 BglII 和 XhoI 消化, 凝胶电泳分离, 纯化, 然后克隆到 BamHI 和 XhoI 消化的 pMT2188 中, 得到一个质粒, 命名为 pJaL660。这样, 通过质粒 pJaL660 的构建产生了米曲霉纤维二糖酶 cDNA 基因的真菌表达质粒。

在米曲霉中表达米曲霉纤维二糖酶

菌株 BECh2 (公开于 WO 00/30322) 利用 pJaL600 进行转化, 方法如 Christensen 等; Biotechnology 1988, 6, 1419 - 1422 一文所述。米曲霉的菌丝体一般在丰富营养肉汤中生长。菌丝体经过滤从肉汤中分离出来。将酶制品 Novozyme® (Novozyyme A/S) 添加到渗透压稳定化缓冲液 (如 1.2M MgSO₄, 以磷酸钠缓冲到 5.0) 中的菌丝体中。此悬浮液在 37℃ 搅拌孵育 60 分钟。通过微孔布 (mira - cloth) 过滤原生质体以除去菌丝体碎片。收获原生质体并用 STC (1.2M 山梨糖醇, 10mM CaCl₂, 10mM Tris - Cl pH7.5) 洗涤两次。最后将原生质体重新悬浮于 200 - 1000 μ l STC 中。

为进行转化, 将 5 μ g DNA 添加到 100 μ l 原生质体悬浮液中, 然后再加入 200 μ l PEG 溶液 (60% PEG4000, 10mM CaCl₂, 10mM Tris - Cl pH7.5) 并且混合物在室温下孵育 20 分钟。收获原生质体并以 1.2M 山梨糖醇洗涤两次。最后将原生质体重新悬浮于 200 μ l 1.2M 山梨糖醇中, 在选择性平板上 (基本培养基 + 10g/L 细菌用琼脂 (Bacto - Agar) (Difco)) 铺平板, 37℃ 培养。在 37℃ 下生长 3 - 4 天之后, 稳定转化体表现为旺盛生长和形成孢子的菌落。对转化体进行两次孢子分离。

使转化体在含 100ml YPM 培养基 (2g/l 酵母提取物, 2g/l 胨, 和 2% 麦芽糖) 的摇瓶中 30℃ 下生长 4 天。按照厂商使用说明在 SDS-PAGE 凝胶 (Novex

NuPAGE 10% Bis-Tris 凝胶) 上对上清液 (20 μ l) 进行分析。第四号转化体被命名为 JaL406。

实施例 3

米曲霉纤维二糖酶的生产

利用标准底物使转化的米曲霉 JAL406 在发酵罐中生长, 培养 5 天后收获发酵液。使整个发酵液通过 Whatman 公司的玻璃纤维滤器, 先经过 “D” 滤器然后经过 “F” 滤器, 最后使澄清酶溶液经过 Millipore 公司的孔径 22 μ m 的无菌滤器。除去菌丝体。利用 Filtron 交叉流动膜 (截留值 10kDa) 对澄清酶溶液浓缩。为获得纯酶, 使浓缩液经过 0.1M 醋酸钠 pH6 缓冲液平衡的 2l Sephacryl 200 柱。纯酶在 SDS-PAGE 中的分子量为 91kDa。在 DSC 中 pH6.0 时的熔解温度是 67 $^{\circ}$ C。比较来说, 纯化的黑曲霉纤维二糖酶的熔解温度是 62 $^{\circ}$ C。米曲霉纤维二糖酶的摩尔消光系数是 169540。

如果对硝基苯基 β -D-葡萄糖为底物, 在 pH4.0 和 40 $^{\circ}$ C 下黑曲霉纤维二糖酶的催化活性比米曲霉纤维二糖酶的催化活性低三倍。在 pH3.0 至 7.0 之间黑曲霉和米曲霉的纤维二糖酶都显示出大于 50% 的催化活性。

当将强启动子和终止子基因转化回米曲霉中时, 克隆的米曲霉纤维二糖酶获得高得多的产量。

生物材料的保藏

下列生物材料按照布达佩斯条约的条款保藏在德意志微生物保藏中心 (DSMZ - Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Mascheroder Weg 1B, D-38124 Braunschweig Germany), 授予如下保藏号:

保藏	保藏号	保藏日期
NN049573	DSM 14240	2001 - 04 - 19

0-1	表格-PCT/RO/134(EASY) 与保藏的微生物或其它生物材料有关的 声明 (PCT 细则第 13 条之 2)	
0-1-1	制备时利用	PCT-EASY 版本 2.92 (01.01.2002 更新)
0-2	国际申请号	
0-3	申请人或代理人的参考卷号	10128-WO
1	下面列出与说明书中提及的保藏的微生物或其它生物材料有关的声明:	
1-1	页	8
1-2	行	19 - 23
1-3	保藏鉴定	
1-3-1	保藏机构名称	DSMZ-德意志微生物保藏中心
1-3-2	保藏机构地址	Mascheroder weg 1b, D-38124 Braunschweig, 德国
1-3-3	保藏日期	2001 年 4 月 19 日 (19.04.2001)
1-3-4	保藏号	DSMZ 14240
1-4	其它声明	无
1-5	声明适用的指定国	所有指定国
1-6	声明的单独提供 这些声明之后将提交给国际局	无

```

<110> 诺和酶股份有限公司 (Novozymes A/S)

<120> 具有纤维二糖酶活性的多肽和编码其的多核苷酸

<130> 10128.000-DK

<160> 10

<170> PatentIn version 3.0

<210> 1
<211> 2771
<212> DNA
<213> 米曲霉(Aspergillus oryzae)

<220>
<221> CDS
<222> (30)..(2612)

<220>
<221> mat_peptide
<222> (87)..()

<400> 1
ctgttctgct gggtacctgc cacgttatc atg aag ctt ggt tgg atc gag gtg      53
                               Met Lys Leu Gly Trp Ile Glu Val
                               -15

gcc gca ttg gcg gct gcc tca gta gtc agt gcc aag gat gat ctc gcg      101
Ala Ala Leu Ala Ala Ala Ser Val Val Ser Ala Lys Asp Asp Leu Ala
-10                               -5                -1  1                5

tac tcc cct cct ttc tac cct tcc cca tgg gca gat ggt cag ggt gaa      149
Tyr Ser Pro Pro Phe Tyr Pro Ser Pro Trp Ala Asp Gly Gln Gly Glu
                10                15                20

tgg gcg gaa gta tac aaa cgc gct gta gac ata gtt tcc cag atg acg      197
Trp Ala Glu Val Tyr Lys Arg Ala Val Asp Ile Val Ser Gln Met Thr
                25                30                35

ttg aca gag aaa gtc aac tta acg act gga aca gga tgg caa cta gag      245
Leu Thr Glu Lys Val Asn Leu Thr Thr Gly Thr Gly Trp Gln Leu Glu
                40                45                50

agg tgt gtt gga caa act ggc agt gtt ccc aga ctc aac atc ccc agc      293
Arg Cys Val Gly Gln Thr Gly Ser Val Pro Arg Leu Asn Ile Pro Ser
                55                60                65

ttg tgt ttg cag gat agt cct ctt ggt att cgt ttc tcg gac tac aat      341
Leu Cys Leu Gln Asp Ser Pro Leu Gly Ile Arg Phe Ser Asp Tyr Asn
                70                75                80                85

tca gct ttc cct gcg ggt gtt aat gtc gct gcc acc tgg gac aag acg      389
Ser Ala Phe Pro Ala Gly Val Asn Val Ala Ala Thr Trp Asp Lys Thr
                90                95                100

ctc gcc tac ctt cgt ggt cag gca atg ggt gag gag ttc agt gat aag      437
Leu Ala Tyr Leu Arg Gly Gln Ala Met Gly Glu Glu Phe Ser Asp Lys
                105                110                115

```

ggt att gac gtt cag ctg ggt cct gct gct ggc cct ctc ggt gct cat	485
Gly Ile Asp Val Gln Leu Gly Pro Ala Ala Gly Pro Leu Gly Ala His	
120 125 130	
ccg gat ggc ggt aga aac tgg gaa ggt ttc tca cca gat cca gcc ctc	533
Pro Asp Gly Gly Arg Asn Trp Glu Gly Phe Ser Pro Asp Pro Ala Leu	
135 140 145	
acc ggt gta ctt ttt gcg gag acg att aag ggt att caa gat gct ggt	581
Thr Gly Val Leu Phe Ala Glu Thr Ile Lys Gly Ile Gln Asp Ala Gly	
150 155 160 165	
gtc att gcg aca gct aag cat tat atc atg aac gaa caa gag cat ttc	629
Val Ile Ala Thr Ala Lys His Tyr Ile Met Asn Glu Gln Glu His Phe	
170 175 180	
cgc caa caa ccc gag gct gcg ggt tac gga ttc aac gta agc gac agt	677
Arg Gln Gln Pro Glu Ala Ala Gly Tyr Gly Phe Asn Val Ser Asp Ser	
185 190 195	
ttg agt tcc aac gtt gat gac aag act atg cat gaa ttg tac ctc tgg	725
Leu Ser Ser Asn Val Asp Asp Lys Thr Met His Glu Leu Tyr Leu Trp	
200 205 210	
ccc ttc gcg gat gca gta cgc gct gga gtc ggt gct gtc atg tgc tct	773
Pro Phe Ala Asp Ala Val Arg Ala Gly Val Gly Ala Val Met Cys Ser	
215 220 225	
tac aac caa atc aac aac agc tac ggt tgc gag aat agc gaa act ctg	821
Tyr Asn Gln Ile Asn Asn Ser Tyr Gly Cys Glu Asn Ser Glu Thr Leu	
230 235 240 245	
aac aag ctt ttg aag gcg gag ctt ggt ttc caa ggc ttc gtc atg agt	869
Asn Lys Leu Leu Lys Ala Glu Leu Gly Phe Gln Gly Phe Val Met Ser	
250 255 260	
gat tgg acc gct cat cac agc ggc gta ggc gct gct tta gca ggt ctg	917
Asp Trp Thr Ala His His Ser Gly Val Gly Ala Ala Leu Ala Gly Leu	
265 270 275	
gat atg tcg atg ccc ggt gat gtt acc ttc gat agt ggt acg tct ttc	965
Asp Met Ser Met Pro Gly Asp Val Thr Phe Asp Ser Gly Thr Ser Phe	
280 285 290	
tgg ggt gca aac ttg acg gtc ggt gtc ctt aac ggt aca atc ccc caa	1013
Trp Gly Ala Asn Leu Thr Val Gly Val Leu Asn Gly Thr Ile Pro Gln	
295 300 305	
tgg cgt gtt gat gac atg gct gtc cgt atc atg gcc gct tat tac aag	1061
Trp Arg Val Asp Asp Met Ala Val Arg Ile Met Ala Ala Tyr Tyr Lys	
310 315 320 325	
gtt ggc cgc gac acc aaa tac acc cct ccc aac ttc agc tcg tgg acc	1109
Val Gly Arg Asp Thr Lys Tyr Thr Pro Pro Asn Phe Ser Ser Trp Thr	
330 335 340	
agg gac gaa tat ggt ttc gcg cat aac cat gtt tcg gaa ggt gct tac	1157
Arg Asp Glu Tyr Gly Phe Ala His Asn His Val Ser Glu Gly Ala Tyr	
345 350 355	
gag agg gtc aac gaa ttc gtg gac gtg caa cgc gat cat gcc gac cta	1205

Glu	Arg	Val	Asn	Glu	Phe	Val	Asp	Val	Gln	Arg	Asp	His	Ala	Asp	Leu	
		360					365					370				
atc	cgt	cgc	atc	ggc	gcg	cag	agc	act	gtt	ctg	ctg	aag	aac	aag	ggt	1253
Ile	Arg	Arg	Ile	Gly	Ala	Gln	Ser	Thr	Val	Leu	Leu	Lys	Asn	Lys	Gly	
		375				380					385					
gcc	ttg	ccc	ttg	agc	cgc	aag	gaa	aag	ctg	gtc	gcc	ctt	ctg	gga	gag	1301
Ala	Leu	Pro	Leu	Ser	Arg	Lys	Glu	Lys	Leu	Val	Ala	Leu	Leu	Gly	Glu	
		390			395				400						405	
gat	gcg	ggt	tcc	aac	tcg	tgg	ggc	gct	aac	ggc	tgt	gat	gac	cgt	ggt	1349
Asp	Ala	Gly	Ser	Asn	Ser	Trp	Gly	Ala	Asn	Gly	Cys	Asp	Asp	Arg	Gly	
			410					415						420		
tgc	gat	aac	ggt	acc	ctt	gcc	atg	gcc	tgg	ggt	agc	ggt	act	gcg	aat	1397
Cys	Asp	Asn	Gly	Thr	Leu	Ala	Met	Ala	Trp	Gly	Ser	Gly	Thr	Ala	Asn	
			425					430					435			
ttc	cca	tac	ctc	gtg	aca	cca	gag	cag	gcg	att	cag	aac	gaa	gtt	ctt	1445
Phe	Pro	Tyr	Leu	Val	Thr	Pro	Glu	Gln	Ala	Ile	Gln	Asn	Glu	Val	Leu	
		440					445					450				
cag	ggc	cgt	ggt	aat	gtc	ttc	gcc	gtg	acc	gac	agt	tgg	gcg	ctc	gac	1493
Gln	Gly	Arg	Gly	Asn	Val	Phe	Ala	Val	Thr	Asp	Ser	Trp	Ala	Leu	Asp	
		455				460				465						
aag	atc	gct	gcg	gct	gcc	cgc	cag	gcc	agc	gta	tct	ctc	gtg	ttc	gtc	1541
Lys	Ile	Ala	Ala	Ala	Ala	Arg	Gln	Ala	Ser	Val	Ser	Leu	Val	Phe	Val	
		470			475				480						485	
aac	tcc	gac	tca	gga	gaa	agc	tat	ctt	agt	gtg	gat	gga	aat	gag	ggc	1589
Asn	Ser	Asp	Ser	Gly	Glu	Ser	Tyr	Leu	Ser	Val	Asp	Gly	Asn	Glu	Gly	
			490					495						500		
gat	cgt	aac	aac	atc	act	ctg	tgg	aag	aac	ggc	gac	aat	gtg	gtc	aag	1637
Asp	Arg	Asn	Asn	Ile	Thr	Leu	Trp	Lys	Asn	Gly	Asp	Asn	Val	Val	Lys	
			505					510					515			
acc	gca	gcg	aat	aac	tgt	aac	aac	acc	gtg	gtc	atc	atc	cac	tcc	gtc	1685
Thr	Ala	Ala	Asn	Asn	Cys	Asn	Asn	Thr	Val	Val	Ile	Ile	His	Ser	Val	
		520					525					530				
gga	cca	gtt	ttg	atc	gat	gaa	tgg	tat	gac	cac	ccc	aat	gtc	act	ggt	1733
Gly	Pro	Val	Leu	Ile	Asp	Glu	Trp	Tyr	Asp	His	Pro	Asn	Val	Thr	Gly	
		535				540					545					
att	ctc	tgg	gct	ggt	ctg	cca	ggc	cag	gag	tct	ggt	aac	tcc	atc	gcc	1781
Ile	Leu	Trp	Ala	Gly	Leu	Pro	Gly	Gln	Glu	Ser	Gly	Asn	Ser	Ile	Ala	
		550			555				560						565	
gat	gtg	ctg	tac	ggt	cgt	gtc	aac	cct	ggc	gcc	aag	tct	cct	ttc	act	1829
Asp	Val	Leu	Tyr	Gly	Arg	Val	Asn	Pro	Gly	Ala	Lys	Ser	Pro	Phe	Thr	
			570					575						580		
tgg	ggc	aag	acc	cgg	gag	tcg	tat	ggt	tct	ccc	ttg	gtc	aag	gat	gcc	1877
Trp	Gly	Lys	Thr	Arg	Glu	Ser	Tyr	Gly	Ser	Pro	Leu	Val	Lys	Asp	Ala	
			585				590					595				
aac	aat	ggc	aac	gga	gcg	ccc	cag	tct	gat	ttc	acc	cag	ggt	gtt	ttc	1925
Asn	Asn	Gly	Asn	Gly	Ala	Pro	Gln	Ser	Asp	Phe	Thr	Gln	Gly	Val	Phe	
		600					605					610				

atc gat tac cgc cat ttc gat aag ttc aat gag acc cct atc tac gag Ile Asp Tyr Arg His Phe Asp Lys Phe Asn Glu Thr Pro Ile Tyr Glu 615 620 625	1973
ttt ggc tac ggc ttg agc tac acc acc ttc gag ctc tcc gac ctc cat Phe Gly Tyr Gly Leu Ser Tyr Thr Thr Phe Glu Leu Ser Asp Leu His 630 635 640 645	2021
gtt cag ccc ctg aac gcg tcc cga tac act ccc acc agt ggc atg act Val Gln Pro Leu Asn Ala Ser Arg Tyr Thr Pro Thr Ser Gly Met Thr 650 655 660	2069
gaa gct gca aag aac ttt ggt gaa att ggc gat gcg tcg gag tac gtg Glu Ala Ala Lys Asn Phe Gly Glu Ile Gly Asp Ala Ser Glu Tyr Val 665 670 675	2117
tat ccg gag ggg ctg gaa agg atc cat gag ttt atc tat ccc tgg atc Tyr Pro Glu Gly Leu Glu Arg Ile His Glu Phe Ile Tyr Pro Trp Ile 680 685 690	2165
aac tct acc gac ctg aag gca tcg tct gac gat tct aac tac ggc tgg Asn Ser Thr Asp Leu Lys Ala Ser Ser Asp Asp Ser Asn Tyr Gly Trp 695 700 705	2213
gaa gac tcc aag tat att ccc gaa ggc gcc acg gat ggg tct gcc cag Glu Asp Ser Lys Tyr Ile Pro Glu Gly Ala Thr Asp Gly Ser Ala Gln 710 715 720 725	2261
ccc cgt ttg ccc gct agt ggt ggt gcc gga gga aac ccc ggt ctg tac Pro Arg Leu Pro Ala Ser Gly Gly Ala Gly Gly Asn Pro Gly Leu Tyr 730 735 740	2309
gag gat ctt ttc cgc gtc tct gtg aag gtc aag aac acg ggc aat gtc Glu Asp Leu Phe Arg Val Ser Val Lys Val Lys Asn Thr Gly Asn Val 745 750 755	2357
gcc ggt gat gaa gtt cct cag ctg tac gtt tcc cta ggc ggc ccg aat Ala Gly Asp Glu Val Pro Gln Leu Tyr Val Ser Leu Gly Gly Pro Asn 760 765 770	2405
gag ccc aag gtg gta ctg cgc aag ttt gag cgt att cac ttg gcc cct Glu Pro Lys Val Val Leu Arg Lys Phe Glu Arg Ile His Leu Ala Pro 775 780 785	2453
tcg cag gag gcc gtg tgg aca acg acc ctt acc cgt cgt gac ctt gca Ser Gln Glu Ala Val Trp Thr Thr Thr Leu Thr Arg Arg Asp Leu Ala 790 795 800 805	2501
aac tgg gac gtt tcg gct cag gac tgg acc gtc act cct tac ccc aag Asn Trp Asp Val Ser Ala Gln Asp Trp Thr Val Thr Pro Tyr Pro Lys 810 815 820	2549
acg atc tac gtt gga aac tcc tca cgg aaa ctg ccg ctc cag gcc tcg Thr Ile Tyr Val Gly Asn Ser Ser Arg Lys Leu Pro Leu Gln Ala Ser 825 830 835	2597
ctg cct aag gcc cag taaggggcaa gtccctgattg tacagagcat ttcgagattt Leu Pro Lys Ala Gln 840	2652
atgatgtaca tgtttatgaa tgacctaggg tagggtaata cttagtaggg ttagttctaa	2712

ttcttggagt caagtattga ctactgggc cgataaaaaa aaaaaaaaaa aaaaaaaaaa 2771

<210> 2

<211> 861

<212> PRT

<213> 米曲霉

<400> 2

Met Lys Leu Gly Trp Ile Glu Val Ala Ala Leu Ala Ala Ala Ser Val
-15 -10 -5

Val Ser Ala Lys Asp Asp Leu Ala Tyr Ser Pro Pro Phe Tyr Pro Ser
-1 1 5 10

Pro Trp Ala Asp Gly Gln Gly Glu Trp Ala Glu Val Tyr Lys Arg Ala
15 20 25

Val Asp Ile Val Ser Gln Met Thr Leu Thr Glu Lys Val Asn Leu Thr
30 35 40 45

Thr Gly Thr Gly Trp Gln Leu Glu Arg Cys Val Gly Gln Thr Gly Ser
50 55 60

Val Pro Arg Leu Asn Ile Pro Ser Leu Cys Leu Gln Asp Ser Pro Leu
65 70 75

Gly Ile Arg Phe Ser Asp Tyr Asn Ser Ala Phe Pro Ala Gly Val Asn
80 85 90

Val Ala Ala Thr Trp Asp Lys Thr Leu Ala Tyr Leu Arg Gly Gln Ala
95 100 105

Met Gly Glu Glu Phe Ser Asp Lys Gly Ile Asp Val Gln Leu Gly Pro
110 115 120 125

Ala Ala Gly Pro Leu Gly Ala His Pro Asp Gly Gly Arg Asn Trp Glu
130 135 140

Gly Phe Ser Pro Asp Pro Ala Leu Thr Gly Val Leu Phe Ala Glu Thr
145 150 155

Ile Lys Gly Ile Gln Asp Ala Gly Val Ile Ala Thr Ala Lys His Tyr
160 165 170

Ile Met Asn Glu Gln Glu His Phe Arg Gln Gln Pro Glu Ala Ala Gly
175 180 185

Tyr Gly Phe Asn Val Ser Asp Ser Leu Ser Ser Asn Val Asp Asp Lys
190 195 200 205

Thr Met His Glu Leu Tyr Leu Trp Pro Phe Ala Asp Ala Val Arg Ala
210 215 220

Gly Val Gly Ala Val Met Cys Ser Tyr Asn Gln Ile Asn Asn Ser Tyr
225 230 235

Gly Cys Glu Asn Ser Glu Thr Leu Asn Lys Leu Leu Lys Ala Glu Leu
240 245 250

Gly	Phe	Gln	Gly	Phe	Val	Met	Ser	Asp	Trp	Thr	Ala	His	His	Ser	Gly	
255						260					265					
Val	Gly	Ala	Ala	Leu	Ala	Gly	Leu	Asp	Met	Ser	Met	Pro	Gly	Asp	Val	
270					275					280					285	
Thr	Phe	Asp	Ser	Gly	Thr	Ser	Phe	Trp	Gly	Ala	Asn	Leu	Thr	Val	Gly	
				290					295					300		
Val	Leu	Asn	Gly	Thr	Ile	Pro	Gln	Trp	Arg	Val	Asp	Asp	Met	Ala	Val	
			305					310				315				
Arg	Ile	Met	Ala	Ala	Tyr	Tyr	Lys	Val	Gly	Arg	Asp	Thr	Lys	Tyr	Thr	
		320					325					330				
Pro	Pro	Asn	Phe	Ser	Ser	Trp	Thr	Arg	Asp	Glu	Tyr	Gly	Phe	Ala	His	
	335					340					345					
Asn	His	Val	Ser	Glu	Gly	Ala	Tyr	Glu	Arg	Val	Asn	Glu	Phe	Val	Asp	
350					355					360					365	
Val	Gln	Arg	Asp	His	Ala	Asp	Leu	Ile	Arg	Arg	Ile	Gly	Ala	Gln	Ser	
				370					375					380		
Thr	Val	Leu	Leu	Lys	Asn	Lys	Gly	Ala	Leu	Pro	Leu	Ser	Arg	Lys	Glu	
			385					390					395			
Lys	Leu	Val	Ala	Leu	Leu	Gly	Glu	Asp	Ala	Gly	Ser	Asn	Ser	Trp	Gly	
	400						405					410				
Ala	Asn	Gly	Cys	Asp	Asp	Arg	Gly	Cys	Asp	Asn	Gly	Thr	Leu	Ala	Met	
	415					420					425					
Ala	Trp	Gly	Ser	Gly	Thr	Ala	Asn	Phe	Pro	Tyr	Leu	Val	Thr	Pro	Glu	
430					435					440					445	
Gln	Ala	Ile	Gln	Asn	Glu	Val	Leu	Gln	Gly	Arg	Gly	Asn	Val	Phe	Ala	
				450					455					460		
Val	Thr	Asp	Ser	Trp	Ala	Leu	Asp	Lys	Ile	Ala	Ala	Ala	Ala	Arg	Gln	
			465					470					475			
Ala	Ser	Val	Ser	Leu	Val	Phe	Val	Asn	Ser	Asp	Ser	Gly	Glu	Ser	Tyr	
	480					485						490				
Leu	Ser	Val	Asp	Gly	Asn	Glu	Gly	Asp	Arg	Asn	Asn	Ile	Thr	Leu	Trp	
	495					500					505					
Lys	Asn	Gly	Asp	Asn	Val	Val	Lys	Thr	Ala	Ala	Asn	Asn	Cys	Asn	Asn	
510					515					520					525	
Thr	Val	Val	Ile	Ile	His	Ser	Val	Gly	Pro	Val	Leu	Ile	Asp	Glu	Trp	
			530						535					540		
Tyr	Asp	His	Pro	Asn	Val	Thr	Gly	Ile	Leu	Trp	Ala	Gly	Leu	Pro	Gly	
			545					550					555			
Gln	Glu	Ser	Gly	Asn	Ser	Ile	Ala	Asp	Val	Leu	Tyr	Gly	Arg	Val	Asn	
	560						565					570				
Pro	Gly	Ala	Lys	Ser	Pro	Phe	Thr	Trp	Gly	Lys	Thr	Arg	Glu	Ser	Tyr	
	575					580					585					

Gly Ser Pro Leu Val Lys Asp Ala Asn Asn Gly Asn Gly Ala Pro Gln
 590 595 600 605
 Ser Asp Phe Thr Gln Gly Val Phe Ile Asp Tyr Arg His Phe Asp Lys
 610 615 620
 Phe Asn Glu Thr Pro Ile Tyr Glu Phe Gly Tyr Gly Leu Ser Tyr Thr
 625 630 635
 Thr Phe Glu Leu Ser Asp Leu His Val Gln Pro Leu Asn Ala Ser Arg
 640 645 650
 Tyr Thr Pro Thr Ser Gly Met Thr Glu Ala Ala Lys Asn Phe Gly Glu
 655 660 665
 Ile Gly Asp Ala Ser Glu Tyr Val Tyr Pro Glu Gly Leu Glu Arg Ile
 670 675 680 685
 His Glu Phe Ile Tyr Pro Trp Ile Asn Ser Thr Asp Leu Lys Ala Ser
 690 695 700
 Ser Asp Asp Ser Asn Tyr Gly Trp Glu Asp Ser Lys Tyr Ile Pro Glu
 705 710 715
 Gly Ala Thr Asp Gly Ser Ala Gln Pro Arg Leu Pro Ala Ser Gly Gly
 720 725 730
 Ala Gly Gly Asn Pro Gly Leu Tyr Glu Asp Leu Phe Arg Val Ser Val
 735 740 745
 Lys Val Lys Asn Thr Gly Asn Val Ala Gly Asp Glu Val Pro Gln Leu
 750 755 760 765
 Tyr Val Ser Leu Gly Gly Pro Asn Glu Pro Lys Val Val Leu Arg Lys
 770 775 780
 Phe Glu Arg Ile His Leu Ala Pro Ser Gln Glu Ala Val Trp Thr Thr
 785 790 795
 Thr Leu Thr Arg Arg Asp Leu Ala Asn Trp Asp Val Ser Ala Gln Asp
 800 805 810
 Trp Thr Val Thr Pro Tyr Pro Lys Thr Ile Tyr Val Gly Asn Ser Ser
 815 820 825
 Arg Lys Leu Pro Leu Gln Ala Ser Leu Pro Lys Ala Gln
 830 835 840

<210> 3
 <211> 11
 <212> DNA
 <213> 合成构建体

<400> 3
 gtactaaaac c

11

<210> 4
 <211> 11
 <212> DNA

<213> 合成构建体

<400> 4

ccgttaaatt t

11

<210> 5

<211> 45

<212> DNA

<213> 合成构建体

<400> 5

ggatgctgtt gactccggaa atttaacggt ttggtcttgc atccc

45

<210> 6

<211> 14

<212> DNA

<213> 合成构建体

<400> 6

atgcaattta aact

14

<210> 7

<211> 14

<212> DNA

<213> 合成构建体

<400> 7

cggcaattta acgg

14

<210> 8

<211> 44

<212> DNA

<213> 合成构建体

<400> 8

ggtattgtcc tgcagacggc aatttaacgg cttctgcgaa tcgc

44

<210> 9

<211> 27

<212> DNA

<213> 合成构建体

<400> 9

agatctacca tgaagcttgg ttggatc

27

<210> 10

<211> 24

<212> DNA

<213> 合成构建体

<400> 10

ctcgagttac tgggccttag gcag

24