



(12) 发明专利

(10) 授权公告号 CN 101253263 B

(45) 授权公告日 2014. 07. 02

(21) 申请号 200680023158. 4

C11D 3/386 (2006. 01)

(22) 申请日 2006. 04. 27

(56) 对比文件

(30) 优先权数据

60/675, 601 2005. 04. 27 US

(85) PCT国际申请进入国家阶段日

2007. 12. 27

(86) PCT国际申请的申请数据

PCT/US2006/016244 2006. 04. 27

(87) PCT国际申请的公布数据

W02006/116682 EN 2006. 11. 02

(73) 专利权人 诺维信股份有限公司

地址 美国加利福尼亚州

(72) 发明人 保罗·哈里斯

阿尔弗雷多·洛佩斯德利昂

米切尔·雷伊 丁晗澍

埃琳娜·弗拉森科

(74) 专利代理机构 北京市柳沈律师事务所

11105

代理人 封新琴

(51) Int. Cl.

C12N 9/42 (2006. 01)

C12N 15/00 (2006. 01)

C12N 9/24 (2006. 01)

Gerard J. Kleywegt et al. The crystal structure of the catalytic core domain of endoglucanase I from *trichoderma reesei* at 3.6 Å resolution, and a comparison with related enzymes. *J. Mol. Biol.* 272. 1997, 272383–397.

Janelle N. Van Arsdel et al. Cloning, characterization, and expression in *saccharomyces cerevisiae* of endoglucanase I from *trichoderma reesei*. *Bio/Technology* 5. 1987, 560–64.

审查员 陈中伟

权利要求书2页 说明书40页

序列表4页 附图8页

(54) 发明名称

具有内切葡聚糖酶活性的多肽和编码该多肽的多核苷酸

(57) 摘要

本发明涉及具有内切葡聚糖酶活性的分离的多肽和编码所述多肽的分离的多核苷酸。本发明还涉及包含所述多核苷酸的核酸构建体、载体和宿主细胞,以及用于制备和使用所述多肽的方法。

1. 具有内切葡聚糖酶活性的分离的土生梭孢霉 (*Thielavia terrestris*) 多肽,其选自下组:

- (a) 多肽,其由与 SEQ ID NO:2 的成熟多肽具有至少 99% 同一性的氨基酸序列组成;和
- (b) 变体,其由包含一个或几个氨基酸保守取代的 SEQ ID NO:2 的成熟多肽组成。

2. 权利要求 1 的多肽,所述多肽由质粒 pTter7F 中所含的多核苷酸编码,所述质粒包含在大肠杆菌 NRRL B-30837 中。

3. 权利要求 1 的多肽,其中所述多肽是变体,所述变体由包含一个或几个氨基酸保守取代的 SEQ ID NO:2 的氨基酸 18 至 336 组成。

4. 权利要求 1 的多肽,其中所述成熟多肽是 SEQ ID NO:2 的氨基酸 18 至 336。

5. 权利要求 1 的多肽,其中所述成熟多肽由 SEQ ID NO:1 的核苷酸 52 至 1008 编码。

6. 权利要求 1 的多肽,其中:

所述变体由包含 1、2、3、4、5、6、7、8、9 或 10 个氨基酸保守取代的 SEQ ID NO:2 的成熟多肽组成。

7. 权利要求 1-6 中任一项的多肽,还对一种或多种选自下组的底物具有酶活性:木聚糖、木葡聚糖、阿拉伯木聚糖、半乳甘露聚糖。

8. 具有内切葡聚糖酶活性的多肽,由 SEQ ID NO:2 组成。

9. 具有内切葡聚糖酶活性的多肽,由 SEQ ID NO:2 的成熟多肽组成。

10. 分离的多核苷酸,其由编码权利要求 1-9 中任一项的多肽的核苷酸序列组成。

11. 权利要求 10 的分离的多核苷酸,其在 SEQ ID NO:1 的成熟多肽编码序列中包含至少一个突变,其中突变的核苷酸序列编码 SEQ ID NO:2 的成熟多肽。

12. 分离的多核苷酸,其由 SEQ ID NO:1 或 SEQ ID NO:1 的成熟多肽编码序列组成。

13. 权利要求 12 的分离的多核苷酸,其由 SEQ ID NO:1 的核苷酸 52 至 1008 组成。

14. 核酸构建体,其包含权利要求 10-13 中任一项的多核苷酸,所述多核苷酸与一个或多个调控序列可操作地连接,所述调控序列指导多肽在表达宿主中产生。

15. 重组表达载体,包含权利要求 14 的核酸构建体。

16. 重组宿主细胞,包含权利要求 14 的核酸构建体。

17. 用于产生权利要求 1-9 中任一项的多肽的方法,其包括:(a) 在有益于所述多肽产生的条件下培养梭孢壳属 (*Theielavia*) 细胞,所述细胞以其野生型形式能够产生所述多肽;和 (b) 回收所述多肽。

18. 用于产生权利要求 1-9 中任一项的多肽的方法,其包括:(a) 在有益于所述多肽产生的条件下培养包含核酸构建体的宿主细胞,所述核酸构建体包含编码所述多肽的核苷酸序列;和 (b) 回收所述多肽。

19. 用于产生亲本细胞突变体的方法,所述方法包括破坏或缺失编码权利要求 1-9 中任一项的多肽的核苷酸序列,其导致突变体与亲本细胞相比产生较少的所述多肽。

20. 通过权利要求 19 的方法产生的突变细胞。

21. 权利要求 20 的突变细胞,所述突变细胞还包含编码天然或异源蛋白质的基因。

22. 用于产生权利要求 1-9 中任一项的具有内切葡聚糖酶活性的多肽的方法,其包括:(a) 在有益于多肽产生的条件下培养转基因植物或植物细胞,所述转基因植物或植物细胞包含编码权利要求 1-9 中任一项的具有内切葡聚糖酶活性的多肽的多核苷酸;和 (b) 回收

所述多肽。

23. 转基因植物部分或植物细胞,其已经使用多核苷酸转化,所述多核苷酸编码权利要求 1-9 中任一项的多肽,所述植物部分或植物细胞不属于植物品种。

24. 洗涤剂组合物,所述洗涤剂组合物包含权利要求 1-9 中任一项的多肽和表面活性剂。

25. 用于降解含有纤维素和半纤维素的生物质的方法,其包括用有效量的权利要求 1-9 中任一项的多肽处理生物质和回收降解的生物质。

26. 权利要求 25 的方法,还包括用有效量的内切 -1,4- β -葡聚糖酶、外切 -1,4-13-D-葡聚糖酶和 / 或 β -D-葡糖苷酶处理生物质。

27. 用于降解含有纤维素和半纤维素的生物质的方法,其包括用权利要求 16 的宿主细胞处理生物质和回收降解的生物质。

28. 权利要求 27 的方法,还包括用有效量的内切 -1,4- β -葡聚糖酶、外切 -1,4- β -D-葡聚糖酶和 / 或 β -D-葡糖苷酶处理生物质。

具有内切葡聚糖酶活性的多肽和编码该多肽的多核苷酸

[0001] 对于在联邦资助的研究和开发下所进行发明的权利声明

[0002] 本发明依据能源部 (Department of Energy) 授予的基本合同 (Prime Contract) DE-AC36-98G010337、NREL 转包合同 (Subcontract) No. ZC0-30017-02 在政府支持下完成。政府对本发明具有一定的权利。

[0003] 发明背景

发明领域

[0004] 本发明涉及具有内切葡聚糖酶活性的分离的多肽和编码所述多肽的分离的多核苷酸。本发明还涉及包含所述多核苷酸的核酸构建体、载体和宿主细胞,以及用于产生和使用所述多肽的方法。

[0005] 相关领域描述

[0006] 纤维素是单糖葡萄糖通过 β -1,4- 键共价连接的聚合物。许多微生物产生水解 β - 连接的聚糖的酶。这些酶包括内切葡聚糖酶、纤维二糖水解酶和 β - 葡糖苷酶。内切葡聚糖酶在随机位置消化纤维素聚合物,将其打开 (opening it) 以受到纤维二糖水解酶攻击 (attack)。纤维二糖水解酶继而从纤维素聚合物的末端释放纤维二糖的分子。纤维二糖水解酶 I 是 1,4-D- 葡聚糖纤维二糖水解酶 (E. C. 3. 2. 1. 91) 活性,其催化纤维素、cellotetrisose 或任何含有 β -1,4- 连接的葡萄糖的聚合物中 1,4- β -D- 糖苷键的水解,从链的还原性末端释放纤维二糖。纤维二糖水解酶 II 是 1,4-D- 葡聚糖纤维二糖水解酶 (E. C. 3. 2. 1. 91) 活性,其催化纤维素、cellotetrisose 或任何含有 β -1,4- 连接的葡萄糖的聚合物中 1,4- β -D- 糖苷键的水解,从链的非还原性末端释放纤维二糖。纤维二糖是水溶性的 β -1,4- 连接的葡萄糖二聚体。 β - 葡糖苷酶将纤维二糖水解成葡萄糖。

[0007] 将纤维素原料转化为乙醇具有以下优势:大量原料现成可用,避免燃烧或填埋材料的愿望,和乙醇燃料的清洁性。木材、农业残余物、草本作物和城市固体废物被认为是用于乙醇生产的原料。这些材料主要由纤维素、半纤维素和木质素组成。一旦将纤维素转化成葡萄糖,葡萄糖将容易地由酵母发酵成乙醇。

[0008] Kvesitadaze et al.,1995, Applied Biochemistry and Biotechnology 50 : 137-143 描述了来自土生梭孢霉 (*Thielavia terrestris*) 嗜热突变菌株的热稳定内切葡聚糖酶的分离和性质。Gilbert et al.,1992, Bioresource Technology 39 :147-154 描述了对土生梭孢霉 255B 的纤维素体系中存在的酶的特征。Breuil et al.,1986, Biotechnology Letters 8 :673-676 描述了来自土生梭孢霉菌株 C464 和 NRRL 8126 的纤维素酶和 β - 葡糖苷酶的产生和定位。

[0009] 本领域中有利的是鉴定新的具有改进性质的内切葡聚糖酶,例如改进的水解速率、更好的热稳定性、对木质素的吸附减少,和除水解纤维素之外水解生物质的非纤维素成分例如半纤维素的能力。对半纤维素具有广泛副活性 (side activity) 的内切葡聚糖酶能够特别有益于改进复杂的、富含半纤维素的生物质底物的总水解产率。

[0010] 本发明的目的是提供改进的具有内切葡聚糖酶活性的多肽和编码该多肽的多核

苷酸。

[0011] 发明概述

[0012] 本发明涉及具有内切葡聚糖酶活性的分离的多肽,所述多肽选自下组:

[0013] (a) 多肽,其包含与 SEQ ID NO:2 的成熟多肽编码序列具有至少 60% 同一性的氨基酸序列;

[0014] (b) 多肽,其由在至少中严紧条件下与以下序列杂交的核苷酸序列编码:(i) SEQ ID NO:1 的成熟多肽编码序列, (ii) 包含 SEQ ID NO:1 的成熟多肽编码序列的基因组 DNA 序列,或 (iii) (i) 或 (ii) 的互补链;和

[0015] (c) SEQ ID NO:2 的成熟多肽包含一个或多个氨基酸的保守取代、缺失和 / 或插入的变体。

[0016] 本发明还涉及分离的多核苷酸,其编码具有内切葡聚糖酶活性的多肽,所述多核苷酸选自下组:

[0017] (a) 多核苷酸,其编码的多肽包含与 SEQ ID NO:2 的成熟多肽具有至少 60% 同一性的氨基酸序列;

[0018] (b) 多核苷酸,其与 SEQ ID NO:1 的成熟多肽编码序列具有至少 60% 同一性;和

[0019] (c) 多核苷酸,其在至少低严紧条件下与以下序列杂交:(i) SEQ ID NO:1 的成熟多肽编码序列, (ii) 包含 SEQ ID NO:1 的成熟多肽编码序列的基因组 DNA 序列,或 (iii) (i) 或 (ii) 的互补链。

[0020] 在优选的方面,所述成熟多肽是 SEQ ID NO:2 的氨基酸 18 至 336。在另一个优选的方面,所述成熟多肽编码序列是 SEQ ID NO:1 的核苷酸 52 至 1008。

[0021] 本发明还涉及包含所述多核苷酸的核酸构建体、重组表达载体和重组宿主细胞。

[0022] 本发明还涉及用于产生这些具有内切葡聚糖酶活性的多肽的方法,其包括:(a) 在有益于产生该多肽的条件下培养包含核酸构建体的重组宿主细胞,所述核酸构建体包含编码所述多肽的多核苷酸;和 (b) 回收所述多肽。

[0023] 本发明还涉及在洗涤剂中和在纤维素到葡萄糖的转化中使用具有内切葡聚糖酶活性的多肽的方法。

[0024] 本发明进一步涉及包含编码蛋白质的基因的核酸构建体,其中将所述基因可操作地连接于编码信号肽的核苷酸序列,该信号肽包含 SEQ ID NO:2 的氨基酸 1 至 17 或由 SEQ ID NO:2 的氨基酸 1 至 17 组成,其中所述基因对于所述核苷酸序列是外源的。

[0025] 附图简述

[0026] 图 1A 和 1B 显示土生梭孢霉 NRRL 8126 内切葡聚糖酶 (CEL7F) 的 cDNA 序列和推导的氨基酸序列 (分别为 SEQ ID NO:1 和 2)。

[0027] 图 2 显示 pTter7F 的限制图谱。

[0028] 图 3 显示 pAlLo1 的限制图谱。

[0029] 图 4 显示 pBANE10 的限制图谱。

[0030] 图 5 显示 pAlLo2 的限制图谱。

[0031] 图 6 显示 pAlLo22 的限制图谱。

[0032] 图 7 显示在 pH5.5 和 60°C 水解 2 小时之后,β-葡聚糖 (1% w/v) 的相对转化率。

[0033] 图 8 显示在 pH5.5 和 60°C 水解 24 小时之后,β-葡聚糖 (1% w/v) 的相对转化率。

[0034] 定义

[0035] 内切葡聚糖酶活性:术语“内切葡聚糖酶活性”在本文中定义为内-1,4- β -D-葡聚糖 4-葡聚糖水解酶 (endo-1,4- β -D-glucan 4-glucanohydrolase) (E. C. No. 3. 2. 1. 4), 其催化纤维素、纤维素衍生物 (例如羧甲基纤维素和羟乙基纤维素)、地衣淀粉 (lichenin) 中的 1,4- β -D-糖苷键、混合的 β -1,3 葡聚糖例如谷类 β -D-葡聚糖或木葡聚糖中的 β -1,4 键和含有纤维素成分的其他植物材料的内水解 (endohydrolysis)。就本发明而言, 根据 Ghose, 1987, Pure and Appl. Chem. 59 :257-268 的方法, 使用羧甲基纤维素 (CMC) 水解来测定内切葡聚糖酶活性。将一单位的内切葡聚糖酶活性定义为:在 50℃、pH4. 8, 每分钟产生 1.0 微摩尔还原糖。

[0036] 在优选的方面, 具有内切葡聚糖酶活性的本发明的多肽还对选自下组的一种或多种底物具有酶活性:木聚糖、木葡聚糖、阿拉伯木聚糖、半乳聚糖、半乳甘露聚糖、葡聚糖和壳多糖。将具有内切葡聚糖酶活性的多肽对于这些多糖底物的活性测定为:将所述底物 (5g 每升) 与本发明的具有内切葡聚糖酶活性的多肽 (1mg 蛋白质每 g 底物) 一起在 pH5. 0 (50mM 乙酸钠) 和 50℃ 无搅拌温育 1 和 92 小时之后, 从不同 AZCL 染色的底物释放的染料的相对量。通过测量 590nm 的吸光度测定染料的释放。

[0037] 在更优选的方面, 具有内切葡聚糖酶活性的本发明的多肽进一步对木聚糖具有酶活性。在另外的更优选的方面, 具有内切葡聚糖酶活性的本发明的多肽进一步对木葡聚糖具有酶活性。在另外的更优选的方面, 具有内切葡聚糖酶活性的本发明的多肽进一步对阿拉伯木聚糖具有酶活性。在另外的更优选的方面, 具有内切葡聚糖酶活性的本发明的多肽进一步对半乳聚糖具有酶活性。在另外的更优选的方面, 具有内切葡聚糖酶活性的本发明的多肽进一步对半乳甘露聚糖具有酶活性。在另外的更优选的方面, 具有内切葡聚糖酶活性的本发明的多肽进一步对葡聚糖具有酶活性。在另外的更优选的方面, 具有内切葡聚糖酶活性的本发明的多肽进一步对壳多糖具有酶活性。在另外的更优选的方面, 具有内切葡聚糖酶活性的本发明的多肽进一步对木聚糖、木葡聚糖、阿拉伯木聚糖、半乳聚糖、半乳甘露聚糖、葡聚糖和壳多糖具有酶活性。

[0038] 本发明的多肽具有的内切葡聚糖酶活性是由 SEQ ID NO :2 的氨基酸 18 至 336 所示氨基酸序列组成的多肽的内切葡聚糖酶活性的至少 20%, 优选至少 40%, 更优选至少 50%, 更优选至少 60%, 更优选至少 70%, 更优选至少 80%, 甚至更优选至少 90%, 最优选至少 95%, 并且甚至最优选至少 100%。

[0039] 家族 7 糖苷水解酶或家族 GH7:术语“家族 7 糖苷水解酶”或“家族 GH7”或“CEL7F”在本文中定义为根据 Henrissat B., 1991, A classification of glycosylhydrolases based on amino-acid sequence similarities, Biochem. J. 280 :309-316, 和 Henrissat B., and Bairoch A., 1996, Updating the sequence-based classification of glycosyl hydrolases, Biochem. J. 316 :695-696 属于糖苷水解酶家族 7 的多肽。

[0040] 分离的多肽:术语“分离的多肽”用于本文中指, 如通过 SDS-PAGE 测定的, 其为至少 20% 纯, 优选至少 40% 纯, 更优选至少 60% 纯, 甚至更优选至少 80% 纯, 最优选至少 90% 纯, 并且甚至最优选至少 95% 纯的多肽。

[0041] 基本上纯的多肽:术语“基本上纯的多肽”在本文表示多肽制备物, 所述多肽制备物按重量计含有至多 10%, 优选至多 8%, 更优选至多 6%, 更优选至多 5%, 更优选至多

4%，更优选至多 3%，甚至更优选至多 2%，最优选至多 1%，并且甚至最优选至多 0.5% 的与其天然结合的 (associated) 的其它多肽材料。因此，优选所述基本上纯的多肽是按存在于制备物中的全部多肽材料的重量计至少 92% 纯，优选至少 94% 纯，更优选至少 95% 纯，更优选至少 96% 纯，更优选至少 96% 纯，更优选至少 97% 纯，更优选至少 98% 纯，甚至更优选至少 99% 纯，最优选至少 99.5% 纯，并且甚至最优选 100% 纯。

[0042] 本发明的多肽优选是基本上纯的形式。具体而言，优选所述多肽是“基本上 (essentially) 纯的形式”，即，所述多肽制备物基本上 (essentially) 不含与其天然结合的其它多肽材料。这能够通过以下方法实现，例如，通过公知的重组方法或由经典纯化方法制备多肽。

[0043] 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。

[0044] 成熟多肽：术语“成熟多肽”在本文中定义为具有内切葡聚糖酶活性的多肽，所述多肽以其在翻译和任何翻译后修饰之后的最终形式存在，所述修饰例如 N- 末端加工、C- 末端截断、糖基化等。

[0045] 同一性：参数“同一性”描述两个氨基酸序列之间或两个核苷酸序列之间的相关性。

[0046] 就本发明而言，两个氨基酸序列之间的同一性程度通过 Clustal 方法 (Higgins, 1989, CABIOS 5:151-153) 使用 LASERGENE™ MEGALIGN™ 软件 (DNASTAR, Inc., Madison, WI) 来测定，使用同一性表和以下多重比对参数：缺口罚分 (gap penalty) 10 和缺口长度罚分 (gap length penalty) 10。配对比对参数是 K 元组 (Ktuple) = 1，缺口罚分 = 3，窗口 (windows) = 5，并且对角线 (diagonals) = 5。

[0047] 就本发明而言，两个核苷酸序列之间的同一性程度通过 Wilbur-Lipman 方法 (Wilbur and Lipman, 1983, Proceedings of the National Academy of Science USA 80: 726-730) 使用 LASERGENE™ MEGALIGN™ 软件 (DNASTAR, Inc., Madison, WI) 来测定，使用同一性表和以下多重比对参数：缺口罚分 10 和缺口长度罚分 10。配对比对参数是 K 元组 = 3，缺口罚分 = 3，和窗口 = 20。

[0048] 多肽片段：术语“多肽片段”在本文中定义为从 SEQ ID NO:2 的氨基和 / 或羧基末端缺失一个或多个氨基酸的多肽，或其同源序列；其中所述片段具有内切葡聚糖酶活性。优选地，片段含有至少 270 个氨基酸残基，更优选至少 285 个氨基酸残基，并且最优选至少 300 个氨基酸残基，例如，SEQ ID NO:2 的氨基酸 18 至 336。

[0049] 亚序列：术语“亚序列 (subsequence)”在本文中定义为从 SEQ ID NO:1 的 5' 和 / 或 3' 端缺失一个或多个核苷酸的核苷酸序列，或其同源序列；其中所述亚序列编码具有内切葡聚糖酶活性的多肽片段。优选地，亚序列含有至少 810 个核苷酸，更优选至少 855 个核苷酸，并且最优选至少 900 个核苷酸。

[0050] 等位变体 (allelic variant)：术语“等位变体”在本文中表示占据相同染色体基因座的基因的任何两种或两种以上可选形式。等位变异通过突变天然地发生，并且可导致种群内的多态性。基因突变可以是沉默的（在编码的多肽中无变化）或可以编码具有改变的氨基酸序列的多肽。多肽的等位变体是由基因的等位变体编码的多肽。

[0051] 分离的多核苷酸：术语“分离的多核苷酸”用于本文中指，如通过琼脂糖电泳测定

的,其为至少 20%纯,优选至少 40%纯,更优选至少 60%纯,甚至更优选至少 80%纯,最优选至少 90%纯,并且甚至最优选至少 95%纯的多核苷酸。

[0052] 基本上纯的多核苷酸:术语“基本上纯的多核苷酸”用于本文指不含其它外来的或不期望的核苷酸的多核苷酸制备物,并且所述多核苷酸制备物处于适合于在遗传工程的蛋白质生产体系中使用的形式。因此,基本上纯的多核苷酸按重量计含有至多 10%,优选至多 8%,更优选至多 6%,更优选至多 5%,更优选至多 4%,更优选至多 3%,甚至更优选至多 2%,最优选至多 1%,并且甚至最优选至多 0.5%的与其天然结合的其它多核苷酸材料。然而,基本上纯的多核苷酸可以包括天然存在的 5' 和 3' 非翻译区,例如启动子和终止子。优选基本上纯的多核苷酸是按重量计至少 90%纯,优选至少 92%纯,更优选至少 94%纯,更优选至少 95%纯,更优选至少 96%纯,更优选至少 97%纯,甚至更优选至少 98%纯,最优选至少 99%,并且甚至最优选至少 99.5%纯的。本发明所述多核苷酸优选为基本上纯的形式。具体而言,优选在本文公开的多核苷酸是“基本上 (essentially) 纯的形式”,即,所述多核苷酸制备物基本上不含与其天然结合的其它多核苷酸材料。在本文中,术语“基本上纯的多核苷酸”与术语“分离的多核苷酸”和“分离形式的多核苷酸”同义。所述多核苷酸可以是基因组、cDNA、RNA、半合成、合成来源的,或它们的任何组合。

[0053] 成熟多肽编码序列:术语“成熟多肽编码序列”在本文中定义为核苷酸序列,其编码具有内切葡聚糖酶活性的成熟多肽。

[0054] cDNA:术语“cDNA”在本文中定义为能够通过反转录从得自真核细胞的成熟的、已剪接的 mRNA 分子制备的 DNA 分子。cDNA 缺少通常存在于相应基因组 DNA 中的内含子序列。起始的 (initial)、初级的 RNA 转录物是 mRNA 的前体,其通过一系列的步骤加工然后作为成熟的已剪接的 mRNA 出现。这些步骤包括通过称为剪接的过程去除内含子序列。因而源自 mRNA 的 cDNA 没有任何内含子序列。

[0055] 核酸构建体:术语“核酸构建体”用于本文指单链或双链的核酸分子,所述核酸分子分离自天然存在的基因,或将所述核酸分子以本来不存在于 (nototherwise exist) 自然界中的方式修饰以含有核酸的片段。当所述核酸构建体含有本发明的编码序列表达所需的调控序列时,术语核酸构建体与术语“表达盒”同义。

[0056] 调控序列 (control sequence):术语“调控序列”在本文定义为包括对编码本发明多肽的多核苷酸表达是必需的或有利的的所有成分。各种调控序列对于编码所述多肽的核苷酸序列可以是天然的或外源的,或各种调控序列对于彼此可以是天然的或外源的。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。最少的情况,调控序列包括启动子和转录和翻译的终止信号。调控序列可以和用于引入特异性限制位点的接头一起提供,所述特异性限制位点促进调控序列与编码多肽的核苷酸序列编码区的连接。

[0057] 可操作地连接:术语“可操作地连接”在本文表示这样的构型,其中将调控序列置于相对于多核苷酸序列的编码序列的合适位置,使得调控序列指导多肽编码序列的表达。

[0058] 编码序列:当用于本文时术语“编码序列”的意思是直接指定其蛋白产物的氨基酸序列的核苷酸序列。编码序列的边界通常由开读框决定,所述开读框通常以 ATG 起始密码子或可供选择的起始密码子例如 GTG 和 TTG 开始,并且以终止密码子例如 TAA、TAG 和 TGA 结束。编码序列可以是 DNA、cDNA 或重组核苷酸序列。

[0059] 表达:术语“表达”包括涉及多肽产生的任何步骤,其包括但不限于转录、转录后修饰、翻译、翻译后修饰和分泌。

[0060] 表达载体:术语“表达载体”在本文定义为线性的或环状的 DNA 分子,其包含编码本发明多肽的多核苷酸,并且所述多核苷酸与提供用于其表达的额外核苷酸可操作地连接。

[0061] 宿主细胞:如本文中所使用的术语“宿主细胞”包括任何细胞类型,所述细胞类型对于使用包含本发明多核苷酸的核酸构建体或表达载体的转化、转染、转导等是易感的(susceptible)。

[0062] 修饰:术语“修饰”在本文的意思是,对由 SEQ ID NO:2 的成熟多肽或其同源序列组成的多肽的任何化学修饰,以及对编码所述多肽的 DNA 的遗传操作。所述修饰可以是一个或多个氨基酸的取代、缺失和/或插入,以及一个或多个氨基酸侧链的置换。

[0063] 人工变体:当用在本文时,术语“人工变体”的意思是具有内切葡聚糖酶活性的多肽,所述多肽由表达 SEQ ID NO:1 或其同源序列的修饰的核苷酸序列或其成熟编码区的生物体产生。所述修饰的核苷酸序列通过人为干预(human intervention),通过修饰公开于 SEQ ID NO:1 或其同源序列的核苷酸序列或其成熟编码区来获得。

[0064] 发明详述

[0065] 具有内切葡聚糖酶活性的多肽

[0066] 在第一个方面,本发明涉及包含下述氨基酸序列的分离的多肽,所述氨基酸序列与 SEQ ID NO:2 的成熟多肽具有至少 60%,优选至少 65%,更优选至少 70%,更优选至少 75%,更优选至少 80%,更优选至少 85%,甚至更优选至少 90%,最优选至少 95%,并且甚至最优选至少 97%、98%或 99%的同一性程度,所述多肽具有内切葡聚糖酶活性(下文中的“同源多肽”)。在优选的方面,所述同源多肽具有的氨基酸序列与 SEQ ID NO:2 的成熟多肽相差十个氨基酸,优选相差五个氨基酸,更优选相差四个氨基酸,甚至更优选相差三个氨基酸,最优选相差两个氨基酸,并且甚至最优选相差一个氨基酸。

[0067] 本发明的多肽优选包含 SEQ ID NO:2 的氨基酸序列或其等位变体;或其具有内切葡聚糖酶活性的片段。在优选的方面,多肽包含 SEQ ID NO:2 的氨基酸序列。在另外的优选方面,多肽包含 SEQ ID NO:2 的成熟多肽。在另外的优选方面,多肽包含 SEQ ID NO:2 的氨基酸 18 至 336,或其等位变体;或其具有内切葡聚糖酶活性的片段。在另外的优选方面,多肽包含 SEQ ID NO:2 的氨基酸 18 至 336。在另外的优选方面,多肽由 SEQ ID NO:2 的氨基酸序列或其等位变体组成;或由其具有内切葡聚糖酶活性的片段组成。在另外的优选方面,多肽由 SEQ ID NO:2 的氨基酸序列组成。在另外的优选方面,多肽由 SEQ ID NO:2 的成熟多肽组成。在另外的优选方面,多肽由 SEQ ID NO:2 的氨基酸 18 至 335 或其等位变体组成;或由其具有内切葡聚糖酶活性的片段组成。在另外的优选方面,多肽由 SEQ ID NO:2 的氨基酸 18 至 336 组成。

[0068] 在第二个方面,本发明涉及具有内切葡聚糖酶活性的分离的多肽,所述分离的多肽由多核苷酸编码,所述多核苷酸在非常低严紧条件下,优选低严紧条件下,更优选中严紧条件下,更优选中-高严紧条件下,甚至更优选高严紧条件下,并且最优选非常高严紧条件下,与以下序列杂交:(i) SEQ ID NO:1 的成熟多肽编码序列,(ii) 包含 SEQ ID NO:1 的成熟多肽编码序列的基因组 DNA 序列,(iii) (i) 或 (ii) 的亚序列,或 (iv) (i)、(ii) 或

(iii) 的互补链 (J. Sambrook, E. F. Fritsch, and T. Maniatis, 1989, Molecular Cloning, A Laboratory Manual, 2nd edition, Cold Spring Harbor, New York)。SEQ ID NO:1 的亚序列含有至少 100 个连续的核苷酸或优选至少 200 个连续的核苷酸。此外,所述亚序列可编码具有内切葡聚糖酶活性的多肽片段。在优选的方面,所述成熟多肽编码序列是 SEQ ID NO:1 的核苷酸 52 至 1008。

[0069] SEQ ID NO:1 的核苷酸序列或其亚序列,以及 SEQ ID NO:2 的氨基酸序列或其片段,可用于设计核酸探针,以根据本领域内公知的方法从不同属和种的菌株鉴定和克隆编码具有内切葡聚糖酶活性的多肽的 DNA。具体而言,根据标准的 Southern 印迹方法,可将这些探针用于与感兴趣的属或种的基因组或 cDNA 杂交,以鉴定和从其中分离相应的基因。这些探针可明显短于完整序列,但长度上应为至少 14,优选至少 25,更优选至少 35,并且最优选至少 70 个核苷酸。然而,优选所述核酸探针是至少 100 个核苷酸长度。例如,所述核酸探针长度上可以是至少 200 个核苷酸,优选至少 300 个核苷酸,更优选至少 400 个核苷酸,或最优选至少 500 个核苷酸。甚至可以使用更长的探针,例如,长度是至少 600 个核苷酸,至少优选至少 700 个核苷酸,更优选至少 800 个核苷酸,或最优选至少 900 个核苷酸的核酸探针。DNA 和 RNA 探针二者均可使用。通常将探针标记以探测相应的基因(例如,用 ^{32}P 、 ^3H 、 ^{35}S 、生物素或抗生物素蛋白(avidin)标记)。将这些探针包含于本发明中。

[0070] 因而,可从由这些其它生物体制备的基因组 DNA 或 cDNA 文库中筛选 DNA,所述 DNA 与上述探针杂交并且编码具有内切葡聚糖酶活性的多肽。可以通过琼脂糖或聚丙烯酰胺凝胶电泳,或通过其它分离技术分离来自这些其它生物体的基因组或其它 DNA。可以将来自文库的 DNA 或分离的 DNA 转移至硝化纤维素(nitrocellulose)或其它合适的载体材料并且固定于其上。为了鉴定与 SEQ ID NO:1 或其亚序列同源的克隆或 DNA,将所述载体材料用在 Southern 印迹中。

[0071] 就本发明而言,杂交表示核苷酸序列在非常低至非常高的严紧条件下与标记的核酸探针杂交,所述核酸探针对应于 SEQ ID NO:1 所示的核苷酸序列、包含 SEQ ID NO:1 的基因组 DNA 序列、它的互补链或其亚序列。可使用例如 X 射线片(X-ray film)检测在这些条件下与核酸探针杂交的分子。

[0072] 在优选的方面,核酸探针是 SEQ ID NO:1 的成熟多肽编码序列。在另外的优选方面,核酸探针是 SEQ ID NO:1 的核苷酸 52 至 1008。在另外的优选方面,核酸探针是编码 SEQ ID NO:2 的多肽的多核苷酸序列,或其亚序列。在另外的优选方面,核酸探针是 SEQ ID NO:1。在另外的优选方面,核酸探针是 SEQ ID NO:1 的成熟多肽编码序列。在另外的优选方面,核酸探针是包含在大肠杆菌 NRRL B-30837 中的质粒 pTter7F 中含有的多核苷酸序列,其中所述其多核苷酸序列编码具有脂肪酶活性的多肽。在另外的优选方面,核酸探针是包含在大肠杆菌 NRRL B-30837 中的质粒 pTter7F 中含有的成熟多肽编码序列。

[0073] 对于长度至少 100 个核苷酸的长探针,将非常低至非常高的严紧条件定义为在 42°C ,在 5X SSPE、0.3% SDS、200 $\mu\text{g/ml}$ 已剪切并且变性的鲑精 DNA 中,并且对于非常低和低严紧性为 25% 的甲酰胺、对于中和中-高严紧性为 35% 的甲酰胺、或对于高和非常高严紧性为 50% 的甲酰胺,根据标准的 Southern 印迹法进行预杂交和杂交最佳 12 至 24 小时。

[0074] 对于长度为至少 100 个核苷酸的长探针,使用 2X SSC、0.2% SDS 优选至少在 45°C (非常低严紧性),更优选至少在 50°C (低严紧性),更优选至少在 55°C (中严紧性),

更优选至少在 60℃ (中-高严谨性), 甚至更优选至少在 65℃ (高严谨性), 并且最优选至少在 70℃ (非常高严谨性) 将载体材料最终洗涤三次, 每次 15 分钟。

[0075] 对于长度大约 15 个核苷酸至大约 70 个核苷酸的短探针, 将严谨条件定义为在比根据 Bolton 和 McCarthy 算法 (1962, Proceedings of the National Academy of Sciences USA 48:1390) 得出的 T_m 低大约 5℃ 至大约 10℃, 在 0.9M NaCl, 0.09M Tris-HCl pH7.6, 6mM EDTA, 0.5% NP-40, 1×Denhardt 溶液, 1mM 焦磷酸钠 (sodium pyrophosphate), 1mM 磷酸二氢钠 (sodium monobasic phosphate), 0.1mM ATP 和 0.2mg 每 ml 的酵母 RNA 中, 根据标准的 Southern 印迹步骤进行预杂交、杂交和杂交后洗涤最佳 12 至 24 小时。

[0076] 对于长度大约 15 个核苷酸至大约 70 个核苷酸的短探针, 将所述载体材料在 6×SSC 加 0.1% SDS 中洗涤一次 15 分钟, 并用 6×SSC 在比计算的 T_m 低 5℃ 至 10℃ 的温度下洗涤两次, 每次 15 分钟。

[0077] 在第三个方面, 本发明涉及人工变体, 所述人工变体包含保守取代、缺失和 / 或插入一个或多个氨基酸的 SEQ ID NO:2 或其同源序列; 或其成熟多肽。优选氨基酸改变对性质是较不重要的 (of a minor nature), 即保守的氨基酸取代或插入, 其不显著影响蛋白质的折叠和 / 或活性; 通常为 1 至大约 30 个氨基酸的小缺失; 小的氨基或羧基末端延伸, 例如氨基末端甲硫氨酸残基; 多至大约 20-25 个残基的小接头肽; 或通过改变净电荷或其它功能 (例如多组氨酸序列 (poly histidine tract)、抗原表位 (antigenic epitope) 或结合域 (binding domain)) 来促进纯化的小延伸。

[0078] 保守取代的实例是在以下组之内: 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)、疏水性氨基酸组 (亮氨酸、异亮氨酸和缬氨酸)、芳族氨基酸组 (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、苏氨酸和甲硫氨酸)。通常不改变比活性 (specific activity) 的氨基酸取代是本领域已知的, 并且由例如 H. Neurath and R. L. Hill, 1979, In, The Proteins, Academic Press, New York 描述。最普遍发生的交换是 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。

[0079] 除了 20 个基本氨基酸, 非基本氨基酸 (例如 4-羟脯氨酸、6-N-甲基赖氨酸、2-氨基异丁酸、异缬氨酸和 α -甲基丝氨酸) 可以取代野生型多肽的氨基酸残基。有限数量的非保守氨基酸、不由遗传密码编码的氨基酸和非天然氨基酸可以取代氨基酸残基。“非天然氨基酸”在蛋白质合成后已经过修饰, 和 / 或在它们的侧链具有不同于基本氨基酸的化学结构。非天然氨基酸能够以化学方法合成, 并且优选是商业上能够获得的, 包括六氢吡啶羧酸 (pipercolic acid)、噻唑烷羧酸 (thiazolidine carboxylic acid)、脱氢脯氨酸、3- 和 4- 甲基脯氨酸, 和 3,3- 二甲基脯氨酸。

[0080] 可供选择的是, 氨基酸改变具有这样的性质以使多肽的物理化学性质改变。例如, 氨基酸改变可改进多肽的热稳定性, 改变底物特异性, 改变最适 pH 等。

[0081] 能够根据本领域已知的方法, 例如定位诱变或丙氨酸分区诱变法 (Cunningham and Wells, 1989, Science 244:1081-1085) 来鉴定亲本多肽中的必需氨基酸。在最后一技术中, 将单一丙氨酸突变引入到分子中的每个残基, 并且测试所得突变分子的生物活性 (即, 内切葡聚糖酶活性) 以鉴定对于所述分子的活性关键的氨基酸残基。同样参见 Hilton et

al., 1996, J. Biol. Chem. 271 :4699-4708。酶的活性部位或其它的生物相互作用也能够通过结构的物理分析而测定,如通过以下这些技术:如核磁共振、晶体学、电子衍射或光亲和标记,连同推定的接触位点氨基酸的突变来测定。参见例如 de Vos et al., 1992, Science 255 :306-312; Smith et al., 1992, J. Mol. Biol. 224 :899-904; Wlodaver et al., 1992, FEBS Lett. 309 :59-64。必需氨基酸的同一性也能够从与多肽的同一性分析来推断,所述多肽与根据本发明的多肽相关。

[0082] 能够使用已知的诱变、重组和 / 或改组 (shuffling) 方法,然后是有关的筛选方法,例如那些由 Reidhaar-Olson and Sauer, 1988, Science 241 :53-57; Bowie and Sauer, 1989, Proc. Natl. Acad. Sci. USA 86 :2152-2156; WO 95/17413; 或 WO 95/22625 公开的那些方法来进行并测试单个或多个氨基酸取代。能够使用的其它方法包括易错 PCR、噬菌体展示 (例如, Lowman et al., 1991, Biochem. 30 :10832-10837; 美国专利号 5, 223, 409; WO 92/06204) 和区域定向的诱变 (Derbyshire et al., 1986, Gene 46 :145; Ner et al., 1988, DNA 7 :127)。

[0083] 诱变 / 改组方法能够与高通量、自动化的筛选方法组合以检测由宿主细胞表达的克隆的、诱变的多肽的活性 (Ness et al., 1999, Nature Biotechnology 17 :893-896)。能够从宿主细胞回收编码活性多肽的诱变的 DNA 分子,并且使用本领域内标准方法快速测序。这些方法允许快速测定感兴趣的多肽中单个氨基酸残基的重要性,并且能够应用于未知结构的多肽。

[0084] SEQ ID NO :2 的成熟多肽例如 SEQ ID NO :2 的氨基酸 18 至 336 的氨基酸取代、缺失和 / 或插入的总数是 10, 优选 9, 更优选 8, 更优选 7, 更优选至多 6, 更优选 5, 更优选 4, 甚至更优选 3, 最优选 2, 并且甚至最优选 1。

[0085] 具有内切葡聚糖酶活性的多肽的来源

[0086] 本发明的多肽可以获得自任何属的微生物。就本发明而言,用于本文与给定的来源有关的术语“获得自”,意思是核苷酸序列编码的多肽由所述来源产生,或由其中插入了来自所述来源的核苷酸序列的菌株产生。在优选的方面,获得自给定来源的多肽是胞外分泌的。

[0087] 本发明的多肽可以是细菌多肽。例如,所述多肽可以是革兰氏阳性细菌多肽例如芽孢杆菌属 (Bacillus) 多肽,例如具有内切葡聚糖酶活性的嗜碱芽孢杆菌 (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;或具有内切葡聚糖酶活性的链霉菌属 (Streptomyces) 多肽,例如,具有内切葡聚糖酶活性的浅青紫链霉菌 (Streptomyces lividans) 或鼠灰链霉菌 (Streptomyces murinus) 多肽;或革兰氏阴性细菌多肽,例如,具有内切葡聚糖酶活性的大肠杆菌或假单胞菌属菌种 (Pseudomonas sp.) 多肽。

[0088] 本发明的多肽也可以是真菌多肽,并且更优选酵母多肽例如具有内切葡聚糖酶活性的念珠菌属 (Candida)、克鲁维酵母属 (Kluyveromyces)、毕赤酵母属 (Pichia)、酵母

属 (*Saccharomyces*)、裂殖酵母属 (*Schizosaccharomyces*) 或西洋蓍霉属 (*Yarrowia*) 多肽;或更优选丝状真菌多肽例如具有内切葡聚糖酶活性的枝顶孢霉属 (*Acremonium*)、曲霉属 (*Aspergillus*)、短梗霉属 (*Aureobasidium*)、隐球菌属 (*Cryptococcus*)、*Filibasidium*、镰孢属 (*Fusarium*)、腐质霉属 (*Humicola*)、*Magnaporthe*、毛霉属 (*Mucor*)、毁丝霉属 (*Myceliophthora*)、*Neocallimastix*、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。

[0089] 在优选的方面,所述多肽是具有内切葡聚糖酶活性的卡尔酵母 (*Saccharomyces carlsbergensis*)、酿酒酵母 (*Saccharomyces cerevisiae*)、糖化酵母 (*Saccharomyces diastaticus*)、道格拉氏酵母 (*Saccharomyces douglasii*)、克鲁弗酵母 (*Saccharomyces kluyveri*)、诺地酵母 (*Saccharomyces norbensis*) 或卵形酵母 (*Saccharomyces oviformis*) 多肽。

[0090] 在另外的优选方面,所述多肽是具有内切葡聚糖酶活性的棘孢曲霉 (*Aspergillus aculeatus*)、泡盛曲霉 (*Aspergillus awamori*)、烟曲霉 (*Aspergillus fumigatus*)、臭曲霉 (*Aspergillus foetidus*)、日本曲霉 (*Aspergillus japonicus*)、构巢曲霉 (*Aspergillus nidulans*)、黑曲霉 (*Aspergillus niger*)、米曲霉 (*Aspergillus oryzae*)、杆孢状镰孢 (*Fusarium bactridioides*)、禾谷镰孢 (*Fusarium cerealis*)、库威镰孢 (*Fusarium crookwellense*)、大刀镰孢 (*Fusarium culmorum*)、禾本科镰孢 (*Fusarium graminearum*)、禾赤镰孢 (*Fusarium gramineum*)、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 sarcocrochrum*)、拟分枝孢镰孢 (*Fusarium sporotrichioides*)、硫色镰孢 (*Fusarium sulphureum*)、圆镰孢 (*Fusarium torulosum*)、拟丝孢镰孢 (*Fusarium trichothecioides*)、镶片镰孢 (*Fusarium venenatum*)、特异腐质霉 (*Humicola insolens*)、疏棉状腐质霉 (*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*)、哈茨木霉 (*Trichoderma harzianum*)、康宁木霉 (*Trichoderma koningii*)、长枝木霉 (*Trichoderma longibrachiatum*)、里氏木霉 (*Trichoderma reesei*) 或绿色木霉 (*Trichoderma viride*) 多肽。

[0091] 在另外的优选方面,所述多肽是具有内切葡聚糖酶活性的 *Thielavia achromatica*、*Thielavia albomyces*、*Thielavia albopilosa*、*Tgielavia australeinsis*、*Thielavia fimeti*、*Thielavia microspora*、*Thielavia ovispora*、*Tgielavia peruviana*、瘤孢梭孢壳 (*Tgielavia speditonum*)、毛梭孢壳 (*Tgielavia setosa*)、*Tgielavia subthermophila*、土生梭孢霉、*Thielavia terricola*、*Tgielavia thermophila*、*Thielavia variospora* 或 *Thielavia wareingii* 多肽。

[0092] 在更优选的方面,所述多肽是具有内切葡聚糖酶活性的土生梭孢霉多肽,并且最优选是具有内切葡聚糖酶活性的土生梭孢霉 NRRL 8126 多肽,例如 SEQ ID NO:2 的多肽或其成熟多肽。

[0093] 可以理解的是对于前述的种,本发明包含完全和不完全阶段 (perfect and imperfect states),和其它分类学的等同物 (equivalent),例如无性型 (anamorph),无论它们已知的种名。本领域熟练技术人员将轻易地识别适合的等同物的同一性。

[0094] 这些种的菌株在许多培养物保藏中心对于公众能够轻易地取得,所述保藏中心诸如美国典型培养物保藏中心 (the American Type Culture Collection) (ATCC)、德意志微生物和细胞培养物保藏中心 (Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH) (DSMZ)、真菌菌种保藏中心 (Centraalbureau Voor Schimmelcultures) (CBS) 和农业研究机构专利培养物保藏中心北区研究中心 (Agricultural Research Service Patent Culture Collection, Northern Regional Research Center) (NRRL)。

[0095] 此外,可以使用上述的探针从其它来源,包括从自然界 (例如,土壤、堆肥、水等) 分离的微生物鉴定和获得这些多肽。用于从天然生境 (habitat) 分离微生物的技术是本领域内公知的。随后可通过相似地筛选这种微生物的基因组或 cDNA 文库来获得所述多核苷酸。一旦用所述探针检测到编码多肽的多核苷酸序列,就能够使用本领域普通技术人员熟知的技术将所述多核苷酸分离或克隆 (参见,例如, Sambrook et al., 1989, 见上文)。

[0096] 本发明的多肽还包括融合多肽或可切割的融合多肽,其中将另外的多肽融合到所述多肽或其片段的 N 末端或 C 末端。通过将编码另一种多肽的核苷酸序列 (或其部分) 融合于本发明的核苷酸序列 (或其部分) 来产生融合的多肽。产生融合多肽的技术是本领域已知的,包括连接编码多肽的编码序列以使它们在阅读框中,并且使融合多肽的表达在相同启动子和终止子的控制下。

[0097] 多核苷酸

[0098] 本发明还涉及分离的多核苷酸,其包含编码具有内切葡聚糖酶活性的本发明的多肽的核苷酸序列,或由该核苷酸序列组成。

[0099] 在优选的方面,核苷酸序列包含 SEQ ID NO:1 或由其组成。在另外的更优选的方面,核苷酸序列包含大肠杆菌 NRRL B-30837 中所含质粒 pTter7F 中的序列,或由该序列其组成。在另外的优选方面,核苷酸序列包含 SEQ ID NO:1 的成熟多肽编码区或由其组成。在另外的优选方面,核苷酸序列包含 SEQ ID NO:1 的核苷酸 52 至 1008 或由其组成。在另外的更优选的方面,核苷酸序列包含大肠杆菌 NRRL B-30837 中所含质粒 pTter7F 中的成熟多肽编码区或由其组成。本发明还包含编码下述多肽的核苷酸序列,所述多肽包含 SEQ ID NO:2 的氨基酸序列或其成熟多肽,或由 SEQ ID NO:2 的氨基酸序列或其成熟多肽组成;由于遗传密码的简并性,所述核苷酸序列不同于 SEQ ID NO:1 或其成熟多肽编码序列。本发明还涉及 SEQ ID NO:1 的亚序列,所述亚序列编码具有内切葡聚糖酶活性的 SEQ ID NO:2 的片段。

[0100] 本发明还涉及突变多核苷酸,所述突变多核苷酸在 SEQ ID NO:1 的成熟多肽编码序列中包含至少一个突变,其中所述突变核苷酸序列编码 SEQ ID NO:2 的成熟多肽。在优选的方面,所述成熟多肽是 SEQ ID NO:2 的氨基酸 18 至 336。

[0101] 用于分离或克隆编码多肽的多核苷酸的技术是本领域内已知的,包括从基因组 DNA 分离,从 cDNA 制备,或其组合。可通过例如使用熟知的聚合酶链式反应 (PCR) 或表达文库的抗体筛选来检测具有共有结构特性的克隆 DNA 片段,从而实现从这种基因组 DNA 克隆本发明的多核苷酸。参见,例如, Innis et al., 1990, PCR: A Guide to Methods and

Application, Academic Press, New York。可以使用其它核酸扩增方法,例如连接酶链式反应(LCR)、连接活化转录(ligated activated transcription;LAT)和基于核苷酸序列的扩增(NASBA)。可以从梭孢壳属(*Thielavia*)的菌株,或从其它或相关的生物体克隆多核苷酸,并且因此可以是例如所述核苷酸序列的多肽编码区的等位基因变体或种变体(species variant)。

[0102] 本发明还涉及包含下述核苷酸序列的多核苷酸,所述核苷酸序列与 SEQ ID NO:1 的成熟多肽编码序列具有至少 60%,优选至少 65%,更优选至少 70%,更优选至少 75%,更优选至少 80%,更优选至少 85%,更优选至少 90%,甚至更优选至少 95%,并且最优选至少 97%同一性的同一性程度,所述多核苷酸编码活性多肽。在优选的方面,所述成熟多肽编码序列是 SEQ ID NO:1 的核苷酸 52 至 1008。

[0103] 修饰编码本发明多肽的核苷酸序列对于合成与所述多肽基本上相似的多肽可能是必需的。术语与所述多肽“基本上相似”指多肽的非天然存在的形式。这些多肽可能以一些工程改造的方式而不同于从其天然来源分离的多肽,例如,比活性、热稳定性、最适 pH 等方面不同的人工变体。可以在作为 SEQ ID NO:1 的多肽编码区存在的核苷酸序列,例如其亚序列的基础上,和/或通过引入如下核苷酸取代来构建变体序列:所述取代不产生由核苷酸序列编码的多肽的另外的氨基酸序列,但是符合意欲产生酶的宿主生物体的密码子选择;或者所述取代可产生不同的氨基酸序列。关于核苷酸取代的概述,参见,例如, Ford et al., 1991, Protein Expression and Purification 2:95-107。

[0104] 对于本领域技术人员显而易见的是,这些取代能够在对于分子功能重要的区域之外进行,并且仍然产生活性多肽。对于由本发明的分离的多核苷酸编码的多肽活性关键的并且因此优选不进行取代的氨基酸残基,可以根据本领域公知的方法,例如定位诱变或丙氨酸分区诱变法(参见,例如, Cunningham and Wells, 1989, Science 244:1081-1085)来鉴定。在后一的技术中,将突变引入到分子中的每个正电残基处,并且测试所得突变分子的内切葡聚糖酶活性,以鉴定对于所述分子的活性关键的氨基酸残基。底物-酶相互作用的位点也能够通过分析三维结构测定,通过如核磁共振分析、晶体学或光亲和标记这样的技术来测定(参见,例如, de Vos et al., 1992, Science 255:306-312; Smith et al., 1992, Journal of Molecular Biology 224:899-904; Wlodaver et al., 1992, FEBS Letters 309:59-64)。

[0105] 本发明还涉及编码本发明多肽的分离的多核苷酸,所述分离的多核苷酸在非常低严紧条件下,优选低严紧条件,更优选中等严紧条件,更优选中-高严紧条件,甚至更优选高严紧条件,并且最优选非常高的严紧条件下,与以下序列杂交:(i) SEQ ID NO:1 的成熟多肽编码序列, (ii) 包含 SEQ ID NO:1 的成熟多肽编码序列的基因组 DNA 序列,或 (iii) (i) 或 (ii) 的互补链;或它们的等位变体和亚序列(Sambrook et al., 1989, 见上文),如本文所定义的。在优选的方面,所述成熟多肽编码序列是 SEQ ID NO:1 的核苷酸 52 至 1008。

[0106] 本发明还涉及分离的多核苷酸,所述分离的多核苷酸通过以下方法获得:(a) 在非常低、低、中、中-高、高或非常高严紧条件下,将 DNA 的群体与以下序列杂交:(i) SEQ ID NO:1 的成熟多肽编码序列, (ii) 包含 SEQ ID NO:1 的成熟多肽编码序列的基因组 DNA 序列,或 (iii) (i) 或 (ii) 的互补链;和 (b) 分离杂交的多核苷酸,其编码具有内切葡聚糖酶活性的多肽。在优选的方面,所述成熟多肽编码序列是 SEQ ID NO:1 的核苷酸 52 至 1008。

[0107] 核酸构建体

[0108] 本发明还涉及包含本发明的分离的多核苷酸的核酸构建体,所述分离的多核苷酸与一个或多个调控序列可操作地连接,所述调控序列在合适的宿主细胞中在与该调控序列相容的条件下指导编码序列的表达。

[0109] 可以用许多方式操作编码本发明多肽的分离的多核苷酸以提供多肽的表达。依赖于表达载体,在将多核苷酸的序列插入载体之前对其进行操作可能是理想的或必需的。使用重组 DNA 方法修饰多核苷酸序列的技术是本领域熟知的。

[0110] 调控序列可以是合适的启动子序列,其是由用于表达编码本发明多肽的多核苷酸的宿主细胞识别的核苷酸序列。启动子序列含有介导多肽的表达的转录调控序列。启动子可以是在所选的宿主细胞中显示转录活性的任何核苷酸序列,包括突变的、截断的和杂合的启动子,并且可以从编码与宿主细胞同源或异源的胞外或胞内多肽的基因获得。

[0111] 用于指导本发明的核酸构建体转录,特别是在细菌宿主细胞中转录的合适启动子的实例是从下述获得的启动子: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)、嗜热脂肪芽孢杆菌产麦芽淀粉酶基因 (amyM)、解淀粉芽孢杆菌 α -淀粉酶基因 (amyQ)、地衣芽孢杆菌青霉素酶基因 (penP)、枯草芽孢杆菌 xylA 和 xylB 基因和原核 β -内酰胺酶基因 (Villa-Kamaroff et al., 1978, Proceedings of the National Academy of Sciences USA 75:3727-3731),以及 tac 启动子 (DeBoer et al., 1983, Proceedings of the National Academy of Sciences USA 80:21-25)。另外的启动子在 "Useful proteins from recombinant bacteria" in Scientific American, 1980, 242:74-94 中;和在 Sambrook et al., 1989, 见上文中有所描述。

[0112] 用于指导本发明的核酸构建体在丝状真菌宿主细胞中转录的合适启动子的实例是从下列酶的基因获得的启动子:米曲霉 TAKA 淀粉酶、曼赫根毛霉 (*Rhizomucor miehei*) 天冬氨酸蛋白酶、黑曲霉中性 α -淀粉酶、黑曲霉酸稳定性 α -淀粉酶、黑曲霉或泡盛曲霉葡糖淀粉酶 (glaA)、曼赫根毛霉脂肪酶、米曲霉碱性蛋白酶、米曲霉丙糖磷酸异构酶、构巢曲霉乙酰胺酶、镶片镰孢淀粉葡糖苷酶 (WO 00/56900)、镶片镰孢 Daria (WO 00/56900)、镶片镰孢 Quinn (W000/56900)、尖镰孢胰蛋白酶样蛋白酶 (WO 96/00787)、里氏木霉 β -葡糖苷酶、里氏木霉纤维二糖水解酶 I、里氏木霉纤维二糖水解酶 II、里氏木霉内切葡聚糖酶 I、里氏木霉内切葡聚糖酶 II、里氏木霉内切葡聚糖酶 III、里氏木霉内切葡聚糖酶 IV、里氏木霉内切葡聚糖酶 V、里氏木霉木聚糖酶 I、里氏木霉木聚糖酶 II、里氏木霉 β -木糖苷酶,以及 NA2-tpi 启动子(来自黑曲霉中性 α -淀粉酶基因和米曲霉丙糖磷酸异构酶基因的启动子的杂合体);和它们的突变的、截断的和杂合的启动子。

[0113] 在酵母宿主中,有用的启动子从如下酶的基因获得:酿酒酵母烯醇化酶 (ENO-1)、酿酒酵母半乳糖激酶 (GAL1)、酿酒酵母醇脱氢酶/甘油醛-3-磷酸脱氢酶 (ADH1, ADH2/GAP)、酿酒酵母丙糖磷酸异构酶 (TPI)、酿酒酵母金属硫蛋白 (CUP1) 和酿酒酵母 3-磷酸甘油酸激酶。对于酵母宿主细胞其它有用的启动子由 Romanos et al., 1992, Yeast 8: 423-488 描述。

[0114] 调控序列也可以是合适的转录终止子序列,是由宿主细胞识别以终止转录的序列。所述终止子序列与编码所述多肽的核苷酸序列的 3' 末端可操作地连接。可以将在所

选宿主细胞中有功能的任何终止子用在本发明中。

[0115] 对于丝状真菌宿主细胞优选的终止子从如下酶的基因获得：米曲霉 TAKA 淀粉酶、黑曲霉葡萄糖淀粉酶、构巢曲霉邻氨基苯甲酸合酶、黑曲霉 α -葡萄糖苷酶和尖镰孢胰蛋白酶样蛋白酶。

[0116] 对于酵母宿主细胞优选的终止子从如下酶的基因获得：酿酒酵母烯醇化酶、酿酒酵母细胞色素 C (CYC1) 和酿酒酵母甘油醛-3-磷酸脱氢酶。对于酵母宿主细胞其它有用的终止子由 Romanos et al., 1992, 见上文描述。

[0117] 调控序列还可以是合适的前导序列，其是对于宿主细胞的翻译重要的 mRNA 非翻译区。前导序列可操作地连接于编码多肽的核苷酸序列的 5'-末端。可以将所选宿主细胞中有功能的任何前导序列用在本发明中。

[0118] 对于丝状真菌宿主细胞优选的前导序列从如下酶的基因获得：米曲霉 TAKA 淀粉酶和构巢曲霉丙糖磷酸异构酶。

[0119] 对于酵母宿主细胞合适的前导序列从如下酶的基因获得：酿酒酵母烯醇化酶 (ENO-1)、酿酒酵母 3-磷酸甘油酸激酶、酿酒酵母 α 因子和酿酒酵母醇脱氢酶 / 甘油醛-3-磷酸脱氢酶 (ADH2/GAP)。

[0120] 调控序列也可以是聚腺苷酸化序列，其是与核苷酸序列的 3' 末端可操作地连接的序列，并且在转录时，宿主细胞将其识别为将聚腺苷残基添加至转录的 mRNA 的信号。可以将所选宿主细胞中有功能的任何聚腺苷酸化序列在本发明中使用。

[0121] 对于丝状真菌宿主细胞优选的聚腺苷酸化序列从如下酶的基因获得：米曲霉 TAKA 淀粉酶、黑曲霉葡萄糖淀粉酶、构巢曲霉邻氨基苯甲酸合酶、尖镰孢胰蛋白酶样蛋白酶和黑曲霉 α -葡萄糖苷酶。

[0122] 对于酵母宿主细胞有用的聚腺苷酸化序列由 Guo and Sherman, 1995, Molecular Cellular Biology 15 :5983-5990 描述。

[0123] 调控序列还可以是信号肽编码区，其编码与多肽的氨基末端相联的氨基酸序列，并且指导编码的多肽进入细胞分泌途径。核苷酸序列的编码序列 5' 端可固地包含信号肽编码区，其与编码分泌多肽的编码区片段一起天然地连接在翻译阅读框中。可供选择的是，编码序列 5' 端可含有对于所述编码序列是异源的信号肽编码区。异源信号肽编码区在编码序列不天然地含有信号肽编码区时可能是必需的。或者，外源信号肽编码区可以简单地取代天然信号肽编码区以增强多肽的分泌。然而，指导表达的多肽进入所选宿主细胞的分泌途径（即，分泌至培养基中）的任何信号肽编码区可在本发明中使用。

[0124] 对于细菌宿主细胞有效的信号肽编码区是从如下酶的基因获得的信号肽编码区：芽孢杆菌属 NCIB 11837 产麦芽糖淀粉酶、嗜热脂肪芽孢杆菌 α -淀粉酶、地衣芽孢杆菌枯草蛋白酶 (subtilisin)、地衣芽孢杆菌 β -内酰胺酶、嗜热脂肪芽孢杆菌中性蛋白酶 (nprT, nprS, nprM) 和枯草芽孢杆菌 prsA。另外的信号肽由 Simonen and Palva, 1993, Microbiological Reviews 57 :109-137 描述。

[0125] 对于丝状真菌宿主细胞有效的信号肽编码区是从如下酶的基因获得的信号肽编码区：米曲霉 TAKA 淀粉酶、黑曲霉中性淀粉酶、黑曲霉葡萄糖淀粉酶、曼赫根毛霉天冬氨酸蛋白酶、特异腐质霉纤维素酶、特异腐质霉内切葡聚糖酶 V 和疏棉状腐质霉脂肪酶。

[0126] 在优选的方面，信号肽是 SEQ ID NO :2 的氨基酸 1 至 17。在另外的优选方面，信

号肽编码区是 SEQ ID NO :1 的核苷酸 1 至 51,其编码 SEQ ID NO :2 的氨基酸 1 至 17。

[0127] 对于酵母宿主细胞有用的信号肽从酿酒酵母 α 因子和酿酒酵母转化酶的基因获得。其它有用的信号肽编码区由 Romanos et al.,1992,见上文,描述。

[0128] 调控序列还可以是前肽编码区,其编码位于多肽氨基末端的氨基酸序列。所得多肽称为酶原 (proenzyme) 或前多肽 (propeptide) (或在某些情况下称为酶原 (zymogen))。前多肽通常是无活性的并且能够通过前肽的催化或自催化切割从前多肽转化成成熟活性多肽。可以从枯草芽孢杆菌碱性蛋白酶 (aprE)、枯草芽孢杆菌中性蛋白酶 (nprT)、酿酒酵母 α 因子、曼赫根毛霉天冬氨酸蛋白酶和嗜热毁丝霉漆酶 (WO 95/33836) 的基因获得前肽编码区。

[0129] 当信号肽和前肽区二者均出现在多肽的氨基末端时,将前肽区置于紧接着 (next to) 多肽氨基末端,并且将信号肽区置于紧接着前肽区的氨基末端。

[0130] 同样理想的是添加调节序列,其允许相对于宿主细胞的生长来调节多肽的表达。调节系统的实例是引起基因表达响应化学或物理刺激物,包括调节化合物的存在而开启或关闭的那些系统。原核系统中的调节系统包括 lac、tac 和 trp 操纵基因系统。在酵母中,可以使用 ADH2 系统或 GAL1 系统。在丝状真菌中,可以使用 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作为调节序列。调节序列的其它实例是那些允许基因扩增的序列。在真核系统中,这些序列包括在氨甲蝶呤 (methotrexate) 存在下扩增的二氢叶酸还原酶基因,和以重金属 (with heavy metal) 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将与调节序列可操作地连接。

[0131] 表达载体

[0132] 本发明还涉及重组表达载体,所述重组表达载体包含本发明的多核苷酸、启动子和转录和翻译终止信号。本文所述的多种核酸和调控序列可以结合在一起以产生重组表达载体,所述表达载体可以包括一个或多个方便的限制位点以允许在这些位点插入或取代编码多肽的核苷酸序列。可供选择的是,可以通过在合适的用于表达的载体中插入包含所述序列的核苷酸序列或核酸构建体来表达本发明的核苷酸序列。在制备表达载体的过程中,将编码序列置于载体中,从而将该编码序列与合适的表达调控序列可操作地连接。

[0133] 重组表达载体可以是任何载体 (例如,质粒或病毒),其能够方便地进行重组 DNA 步骤,并且能够产生核苷酸序列的表达。载体的选择将通常依赖于载体与将引入该载体的宿主细胞的相容性。载体可以是线状或闭合环状质粒。

[0134] 载体可以是自主复制载体,即,作为染色体外实体 (entity) 存在的载体,其复制独立于染色体复制,例如,质粒、染色体外元件、微型染色体 (minichromosome) 或人工染色体。载体可以含有任何用于确保自复制的手段 (means)。或者,载体可以是一种当被引入宿主细胞中时,整合到基因组中并且与整合了该载体的染色体一起复制的载体。此外,可以使用单独的载体或质粒或两个或更多个载体或质粒,其共同含有待引入宿主细胞基因组的完整 DNA (total DNA),或可以使用转座子 (transposon)。

[0135] 本发明的载体优选地含有一个或多个选择性标记,其允许简单选择经转化、转染、转导等的细胞。选择性标记是基因,其产物提供杀生物剂或病毒抗性、对重金属的抗性、对营养缺陷型的原养性 (prototrophy to auxotrophs) 等。

[0136] 细菌选择性标记的实例是来自枯草芽孢杆菌或地衣芽孢杆菌的 *dal* 基因,或赋

予抗生素抗性的标记,所述抗生素抗性例如氨苄青霉素、卡那霉素、氯霉素或四环素抗性。对于酵母宿主细胞合适的标记是 ADE2、HIS3、LEU2、LYS2、MET3、TRP1 和 URA3。用于丝状真菌宿主细胞的选择性标记包括但不限于 amdS(乙酰胺酶)、argB(鸟氨酸氨甲酰基转移酶)、bar(草铵膦(phosphinothricin)乙酰转移酶)、hph(潮霉素磷酸转移酶)、niaD(硝酸还原酶)(nitrate reductase)、pyrG(乳清酸核苷-5'-磷酸脱羧酶)(orotidine-5'-phosphatedecarboxylase)、sC(硫酸腺苷酰转移酶)和 trpC(邻氨基苯甲酸合酶(anthranilate synthase))以及它们的等价物。优选用在曲霉属细胞中的是构巢曲霉或米曲霉的 amdS 和 pyrG 基因和吸水链霉菌(*Streptomyces hygroscopicus*)的 bar 基因。

[0137] 本发明的载体优选含有元件,其允许载体整合入宿主细胞基因组或载体在细胞中独立于基因组的自主复制。

[0138] 为了整合入宿主细胞基因组,载体可依赖编码多肽的多核苷酸的序列或用于通过同源或非同源重组整合入基因组的任何其它载体元件。或者,载体可以含有额外的核苷酸序列,用于指导通过同源重组整合入宿主细胞基因组染色体中的精确位置。为了增加在精确位置整合的可能性,整合元件应该优选含有足够数量的核酸,如 100 至 10,000 碱基对,优选 400 至 10,000 碱基对,并且最优选 800 至 10,000 碱基对,其与相应的目标序列具有高度同一性以增强同源重组的概率。整合元件可以是任何序列,其与宿主细胞基因组中的目标序列同源。此外,整合元件可以是非编码或编码的核苷酸序列。另一方面,可以将载体通过非同源重组整合到宿主细胞基因组中。

[0139] 为了自主复制,载体可以进一步包含复制起点,其使载体能够在所述的宿主细胞中自主地复制。复制起点可以是介导自主复制的任何质粒复制子(replicator),其在细胞中发挥功能。术语“复制起点”或“质粒复制子”在本文定义为能够使质粒或载体体内复制的核苷酸序列。

[0140] 细菌复制起点的实例是允许在大肠杆菌中复制的质粒 pBR322、pUC19、pACYC177 和 pACYC184 的复制起点,和允许在芽孢杆菌属中复制的质粒 pUB110、pE194、pTA1060 和 pAM β 1 的复制起点。

[0141] 用于酵母宿主细胞中的复制起点的实例是 2 微米复制起点,ARS1,ARS4,ARS1 和 CEN3 的组合,和 ARS4 和 CEN6 的组合。

[0142] 在丝状真菌细胞中有用的复制起点的实例是 AMA1 和 ANS1(Gems et al.,1991, Gene 98:61-67;Cullen et al.,1987, Nucleic Acids Research 15:9163-9175;WO 00/24883)。分离 AMA1 基因和构建包含该基因的质粒或载体能够根据公开于 WO 00/24883 中的方法完成。

[0143] 可以将多于一个拷贝的本发明的多核苷酸插入宿主细胞以增加基因产物的产生。多核苷酸拷贝数的增加可通过如下方法获得:将至少一个额外拷贝的序列整合入宿主细胞基因组,或包括与多核苷酸一起的可扩增的选择性标记基因,其中细胞含有选择性标记基因的扩增的拷贝,并且由此可通过在合适的选择剂(selectable agent)存在下培养细胞来选择多核苷酸的额外拷贝。

[0144] 用于连接上述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(参见,例如,Sambrook et al.,1989,见上文)。

[0145] 宿主细胞

[0146] 本发明还涉及重组宿主细胞,所述重组宿主细胞包含本发明的多核苷酸,将其有利地用于多肽的重组产生中。将包含本发明多核苷酸的载体引入宿主细胞中,从而将该载体保留作为染色体整合体(chromosomal integrant)或作为前述的自复制的染色体外载体。术语“宿主细胞”包含亲本细胞的任何子代,其由于复制过程中发生的突变而与亲本细胞不相同。宿主细胞的选择将很大程度依赖于编码多肽的基因和它的来源。

[0147] 宿主细胞可以是单细胞微生物,例如,原核生物,或非单细胞微生物,例如,真核生物。

[0148] 有用的单细胞微生物是细菌细胞,例如革兰氏阳性细菌,包括但不限于,芽孢杆菌属细胞,例如,嗜碱芽孢杆菌、解淀粉芽孢杆菌、短芽孢杆菌、环状芽孢杆菌、克劳氏芽孢杆菌(*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,浅青紫链霉菌和鼠灰链霉菌,或革兰氏阴性细菌例如大肠杆菌和假单胞菌属菌种。在优选的方面,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。在另外优选的方面,芽孢杆菌属细胞是嗜碱的芽孢杆菌属。

[0149] 可通过如下方法实现将载体引入到细菌宿主细胞:例如原生质体转化(参见,例如,Chang and Cohen,1979,Molecular General Genetics 168:111-115),使用感受态细胞(参见,例如,Young and Spizizen,1961,Journal of Bacteriology 81:823-829或Dubnau and Davidoff-Abelson,1971,Journal of Molecular Biology 56:209-221),电穿孔(参见,例如,Shigekawa and Dower,1988,Biotechniques 6:742-751)或接合(参见,例如,Koehler and Thorne,1987,Journal of Bacteriology 169:5771-5278)。

[0150] 宿主细胞还可以是真核生物,例如哺乳动物、昆虫、植物或真菌细胞。

[0151] 在优选的方面,宿主细胞是真菌细胞。“真菌”用在本文包括以下门:子囊菌门(Ascomycota)、担子菌门(Basidiomycota)、壶菌门(Chytridiomycota)和接合菌门(Zygomycota)(如由Hawksworth et al.,In,Ainsworth and Bisby's Dictionary of The Fungi,8th edition,1995,CAB International,University Press,Cambridge,UK所定义)以及卵菌门(Oomycota)(如Hawksworth et al.,1995,见上,171页中所引用),和所有有丝分裂孢子真菌(mitosporic fungi)(Hawksworth et al.,1995,见上)。

[0152] 在更优选的方面,真菌宿主细胞是酵母细胞。“酵母”用在本文包括产子囊酵母(ascosporogenous yeast)(内孢霉目(Endomycetales))、产担子酵母(basidiosporogenous yeast)和属于半知菌类(Fungi Imperfecti)(芽孢纲(Blastomycetes))的酵母。由于酵母的分类在未来可能改变,就本发明而言,将酵母定义为如Biology and Activities of Yeast(Skinner, F.A., Passmore, S.M., and Davenport, R.R., eds, Soc. App. Bacteriol. Symposium Series No. 9, 1980)中所述。

[0153] 在更加优选的方面,酵母宿主细胞是念珠菌属、汉逊酵母属(*Hansenula*)、克鲁维酵母属、毕赤酵母属、酵母属、裂殖酵母属或西洋蓍霉属细胞。

[0154] 在最优选的方面,酵母宿主细胞是卡尔酵母,酿酒酵母,糖化酵母,道格拉氏酵母,克鲁弗酵母,诺地酵母或卵形酵母细胞。在另外最优选的方面,酵母宿主细胞是乳酸克鲁维酵母(*Kluyveromyces lactis*)细胞。在另外最优选的方面,酵母宿主细胞是*Yarrowia lipolytica*细胞。

[0155] 在另外更优选的方面,真菌宿主细胞是丝状真菌细胞。“丝状真菌”包括真菌门(Eumycota)和卵菌门的亚门(如由Hawksworth et al.,1995,见上文,所定义)的所有丝状形式。丝状真菌通常的特征在于由壳多糖(chitin)、纤维素、葡聚糖、壳聚糖(chitosan)、甘露聚糖和其它复杂多糖组成的菌丝体壁。通过菌丝延伸进行营养生长,而碳分解代谢是专性需氧的。相反,酵母例如酿酒酵母的营养生长通过单细胞菌体的出芽生殖(budding)进行,而碳分解代谢可以是发酵的。

[0156] 在甚至更优选的方面,丝状真菌宿主细胞是枝顶孢霉属、曲霉属、短梗霉属、烟管霉属(Bjerkandera)、Ceriporiopsis、鬼伞属(Coprinus)、革盖菌属(Coriolus)、隐球菌属、Filibasidium、镰孢属、腐质霉属、梨孢菌属(Magnaporthe)、毛霉属、毁丝霉属、新考玛脂霉属(Neocallimastix)、脉孢菌属、拟青霉属、青霉属、平革菌属(Phanerochaete)、射脉菌属(Phlebia)、瘤胃壶菌属(Piromyces)、侧耳属(Pleurotus)、裂褶菌属、踝节菌属、嗜热子囊菌属、梭孢壳属、弯颈霉属(Tolypocladium)、栓菌属(Trametes)或木霉属细胞。

[0157] 在最优选的方面,丝状真菌宿主细胞是泡盛曲霉、烟曲霉、臭曲霉、日本曲霉、构巢曲霉、黑曲霉或米曲霉细胞。在另外的最优选方面,丝状真菌宿主细胞是杆孢状镰孢、禾谷镰孢、库威镰孢、大刀镰孢、禾本科镰孢、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、圆镰孢、拟丝孢镰孢或镶片镰孢细胞。在另外最优选的方面,丝状真菌宿主细胞是黑刺烟管菌(Bjerkanderaadusta)、Ceriporiopsis aneirina、Ceriporiopsis aneirina、Ceriporiopsis caregiea、Ceriporiopsis gilvescens、Ceriporiopsis pannocinta、Ceriporiopsis rivulosa、Ceriporiopsis subrufa、Ceriporiopsis subvermispora、灰盖鬼伞(Coprinuscinerus)、毛革盖菌(Coriolus hirsutus)、特异腐质霉、疏棉状腐质霉、米赫毛霉、嗜热毁丝霉、粗糙脉孢菌、产紫青霉、黄孢平革菌(Phanerochaetechrysosporium)、辐射射脉菌(Phlebia radiata)、Pleurotus eryngii、Thielaviaaachromatica、Thielavia albomyces、Thielavia albopilosa、Thielavia australeinsis、Thielavia fimeti、Thielavia microspora、Thielavia ovispora、Thielavia peruviana、瘤孢梭孢壳、毛梭孢壳、Thielavia subthermophila、土生梭孢霉、Thielaviaterricola、Thielavia thermophila、Thielavia variospora、Thielavia wareingii、Trametes villosa、Trametes versicolor、哈茨木霉、康宁木霉、长枝木霉、里氏木霉或绿色木霉细胞。

[0158] 可以将真菌细胞通过涉及原生质体形成、原生质体转化和细胞壁重建的方法以本身公知的方式转化。用于转化曲霉属和木霉属宿主细胞的合适方法在EP238023和Yelton et al.,1984,Proceedings of the National Academy of Sciences USA81:1470-1474中描述。用于转化镰孢属菌种的合适方法由Malardier et al.,1989,Genes 78:147-156和WO 96/00787描述。可以使用由如下文献描述的方法转化酵母:Becker and Guarente, In Abelson, J.N. and Simon, M.I., editors, Guide to Yeast Genetics and Molecular Biology, Methods in Enzymology, Volume 194, pp182-187, Academic Press, Inc., New York; Ito et al.,1983, Journal of Bacteriology 153:163;和Hinnen et al.,1978, Proceedings of the National Academy of Sciences USA 75:1920。

[0159] 产生方法

[0160] 本发明还涉及用于产生本发明多肽的方法,其包括:(a) 在有益于产生多肽的条

件下培养细胞,所述细胞以其野生型形式能够产生所述多肽;和(b)回收所述多肽。在优选的方面,所述细胞是梭孢壳属的细胞。在更优选的方面,所述细胞是土生梭孢霉。在最优选的方面,所述细胞是土生梭孢霉 NRRL8126。

[0161] 本发明还涉及用于产生本发明的多肽的方法,其包括:(a)在有益于产生多肽的条件下培养宿主细胞;和(b)回收所述多肽。

[0162] 本发明还涉及用于产生本发明的多肽的方法,包括:(a)在有益于产生多肽的条件下培养宿主细胞,其中所述宿主细胞包含突变核苷酸序列,其在 SEQ ID NO:1 的成熟多肽编码序列中具有至少一个突变,其中所述突变核苷酸序列编码由 SEQ ID NO:2 的成熟多肽组成的多肽,和(b)回收所述多肽。在优选的方面,所述成熟多肽是 SEQ ID NO:2 的氨基酸 18 至 336。

[0163] 在本发明的产生方法中,使用本领域熟知的方法在适合于产生所述多肽的营养培养基中培养细胞。例如,可以通过在合适培养基中和允许表达和/或分离所述多肽的条件下进行的摇瓶培养,和实验室或工业发酵罐中的小规模或大规模发酵(包括连续、分批、补料分批或固态发酵)来培养细胞。使用本领域已知的方法在合适的营养培养基中进行培养,所述营养培养基包含碳源和氮源和无机盐。合适的培养基能够从商业供应商获得或可以根据公布的成分制备(例如,在美国典型培养物保藏中心的目录中)。如果多肽分泌到营养培养基中,该多肽能够从所述培养基中直接回收。如果多肽不分泌到培养基中,其能够从细胞裂解物(lysate)回收。

[0164] 可以使用本领域已知的对于所述多肽是特异性的方法来检测多肽。这些检测方法可包括特异性抗体的使用、酶产物的形成或酶底物的消失。例如,酶试验(enzyme assay)可用于测定如本文所述的多肽的活性。

[0165] 所得多肽可以使用本领域已知的方法回收。例如,多肽可以通过常规方法从营养培养基中回收,所述常规方法包括但不限于离心、过滤、提取、喷雾干燥、蒸发或沉淀。

[0166] 本发明的多肽可以通过多种本领域已知的方法纯化以获得基本上纯的多肽,所述方法包括但不限于层析(例如,离子交换、亲和、疏水、层析聚焦和大小排阻)、电泳方法(例如,制备型(preparative)等电聚焦)、差示溶解度(例如,硫酸铵沉淀)、SDS-PAGE 或提取(参见,例如,Protein Purification, J.-C. Janson and Lars Ryden, editors, VCH Publishers, New York, 1989)。

[0167] 植物

[0168] 本发明还涉及转基因植物、植物部分或植物细胞,将其用编码本发明的具有内切葡聚糖酶活性的多肽的核苷酸序列转化,从而以可回收的量表达和产生所述多肽。多肽可从植物或植物部分回收。或者,同样可以将含有该重组多肽的植物或植物部分用于改进食品或饲料的质量,例如,改进营养价值、适口性(palatability)和流变性质(rheological properties),或用于破坏抗营养因子。

[0169] 转基因植物可以是双子叶的(双子叶植物)或单子叶的(单子叶植物)。单子叶植物的实例是草(grasses),如草地早熟禾(meadow grass)(蓝草(blue grass),早熟禾属(Poa));饲用牧草(forage grass)如羊茅属(Festuca)、黑麦草属(Lolium);寒地型牧草(temperate grass),如 Agrostis;和谷类,例如,小麦、燕麦、黑麦、大麦、稻(rice)、高粱和玉蜀黍(maize)(玉米)。

[0170] 双子叶植物的实例是烟草 (tobacco), 豆类 (legumes), 如羽扇豆 (lupins), 马铃薯, 糖甜菜 (sugar beet), 豌豆, 豆 (bean) 和大豆 (soybean) 和十字花科的 (cruciferous) 植物 (十字花科 (family Brassicaceae)), 如花椰菜 (cauliflower), 油菜籽 (rape seed) 和紧密相关的模型生物体拟南芥 (*Arabidopsis thaliana*)。

[0171] 植物部分的实例是茎 (stem)、愈伤组织 (callus)、叶 (leaf)、根 (root)、果实 (fruit)、种子 (seed) 和块茎 (tuber), 以及包含这些部分的独立组织, 例如, 表皮 (epidermis)、叶肉 (mesophyll)、薄壁组织 (parenchyma)、维管组织 (vascular tissue)、分生组织 (meristem)。具体的植物细胞区室 (compartments), 如叶绿体 (chloroplast)、质外体 (apoplast)、线粒体 (mitochondria)、液泡 (vacuole)、过氧化物酶体 (peroxisome) 和细胞质 (cytoplasm) 也被认为是植物部分。此外, 任何植物细胞, 无论什么组织来源, 都被认为是植物部分。同样地, 植物部分, 如分离以促进本发明的应用的具体组织和细胞也被认为是植物部分, 例如胚 (embryo)、胚乳 (endosperm)、糊粉 (aleurone) 和种皮 (seed coat)。

[0172] 同样包含于本发明范围内的还有这些植物、植物部分和植物细胞的子代。

[0173] 表达本发明多肽的转基因植物或植物细胞可以依照本领域已知方法构建。简而言之, 通过如下方法构建所述植物或植物细胞: 将编码本发明多肽的一个或多个表达构建体并入植物宿主基因组或叶绿体基因组, 并且繁殖所得的修饰植物或植物细胞成为转基因植物或植物细胞。

[0174] 表达载体便利地是包含编码本发明多肽的多核苷酸的核酸构建体, 所述多核苷酸与在选择的植物或植物部分中表达该多核苷酸序列所必需的适当的调节序列可操作地连接。此外, 表达构建体可以包含对于鉴定宿主细胞有用的选择性标记, 在所述宿主细胞中整合了表达构建体和将该构建体引入到所述植物中所必需的 DNA 序列 (后者依赖于使用的 DNA 引入方法)。

[0175] 调节序列的选择, 例如启动子和终止子序列和任选地信号或转运序列的选择, 举例来说, 基于期望何时、何处以及如何表达多肽而确定。例如, 编码本发明多肽的基因的表达可以是组成型的或诱导型的, 或可以是发育、阶段或组织特异性的, 并且基因产物可以靶向特定的组织或植物部分例如种子或叶。调节序列由例如 Tague et al., 1988, *Plant Physiology* 86:506 所述。

[0176] 对于组成性表达, 可以使用 35S-CaMV、玉米泛素 1 和稻肌动蛋白 1 启动子 (Franck et al., 1980, *Cell* 21:285-294, Christensen et al., 1992, *Plant Mo. Biol.* 18:675-689; Zhang et al., 1991, *Plant Cell* 3:1155-1165)。器官特异性启动子可以是例如来自贮藏库组织 (storage sink tissue) 例如种子、马铃薯块茎和果实的启动子 (Edwards&Coruzzi, 1990, *Ann. Rev. Genet.* 24:275-303), 或来自代谢库组织 (metabolic sink tissue) 例如分生组织的启动子 (Ito et al., 1994, *Plant Mol. Biol.* 24:863-878), 种子特异性启动子诸如来自稻的谷蛋白 (glutelin)、醇溶蛋白 (prolamin)、球蛋白 (globulin) 或白蛋白 (albumin) 启动子 (Wu 等, 1998, *Plant and Cell Physiology* 39:885-889), 来自豆球蛋白 (legumin) B4 和蚕豆 (*Vicia faba*) 的未知的种子蛋白基因的蚕豆启动子 (Conrad 等, 1998, *Journal of Plant Physiology* 152:708-711)、来自种子油体蛋白 (oil body protein) 的启动子 (Chen 等, 1998, *Plant and Cell Physiology* 39:935-941), 来自欧洲油菜 (*Brassica napus*) 的贮藏蛋白 napA 启动子, 或本技术领域公知的任何其他种子特异

性的启动子,例如,在 WO 91/14772 中所描述的。此外,启动子可为叶特异性的启动子,如来自稻或番茄的 *rbcs* 启动子 (Kyoizuka 等,1993, *Plant Physiology* 102 :991-1000),小球藻病毒 (*chlorella virus*) 腺嘌呤甲基转移酶 (adeninemethyltransferase) 基因启动子 (Mitra 和 Higgins,1994, *Plant Molecular Biology* 26 :85-93),或来自稻的 *aldP* 基因启动子 (Kagaya 等,1995, *Molecular and general genetics* 248 :668-674),或伤口诱导的启动子,如马铃薯 *pin2* 启动子 (Xu 等,1993, *Plant Molecular Biology* 22 :573-588)。同样地,所述启动子可通过非生物的处理诱导,所述非生物的处理诸如温度、干旱或盐度变化,或通过外源应用的激活所述启动子的物质诱导,例如乙醇、雌激素 (oestrogens)、植物激素 (plant hormones) 如乙烯、脱落酸 (abscisic acid)、赤霉素 (gibberellic acid) 和 / 或重金属。

[0177] 启动子增强子元件也可以用于实现本发明多肽在植物中的较高表达。例如,启动子增强子元件可以是内含子,将其置于启动子和编码本发明多肽的核苷酸序列之间。例如 Xu et al.,1993,见上,公开了使用稻肌动蛋白 1 基因的第一内含子以增强表达。

[0178] 选择性标记基因和表达构建体的任何其它部分可以选自本领域内可用的那些。

[0179] 将核酸构建体根据本领域已知的常规技术并入植物基因组,所述常规技术包括土壤杆菌属 (*Agrobacterium*) 介导的转化、病毒介导的转化、显微注射 (microinjection)、粒子轰击、生物射弹转化和电穿孔 (Gasser et al.,1990, *Science* 244 :1293;Potrykus,1990, *Bio/Technology* 8 :535;Shimamoto et al.,1989, *Nature* 338 :274)。

[0180] 目前,根癌土壤杆菌 (*Agrobacterium tumefaciens*) 介导的基因转移 (genetransfer),是产生转基因双子叶植物的优选方法 (为了参考,见 Hooykas 和 Schilperoort,1992, *Plant Molecular Biology* 19 :15-38),而且它也可以用于转化单子叶植物,虽然对于这些植物其他的转化方法是常用的。目前,产生转基因单子叶植物的优选的方法,是用粒子 (用转化 DNA 涂覆的微观的金或钨粒子) 轰击胚愈伤组织 (embryonic calli) 或发育中的胚 (developing embryos) (Christou,1992, *Plant Journal* 2 :275-281; Shimamoto,1994, *Current Opinion Biotechnology* 5 :158-162;Vasil et al.,1992, *Bio/Technology* 10 :667-674)。转化单子叶植物的可供选择的方法,是基于原生质体转化,如由 Omirulleh 等,1993, *Plant Molecular Biology* 21 :415-428 所描述的。

[0181] 转化之后,根据本领域熟知的方法选择具有并入的表达构建体的转化体并且再生成为完整植物。通常设计转化方法用于通过如下方法在再生期间或在后续世代中选择性消除选择基因:例如,使用带有两种独立的 T-DNA 构建体的共转化或通过特异性重组酶位点特异性地切除选择基因。

[0182] 本发明还涉及用于产生本发明多肽的方法,其包括:(a) 在有益于产生多肽的条件下培养转基因植物或植物细胞,其包含编码本发明具有内切葡聚糖酶活性的多肽的多核苷酸;和 (b) 回收所述多肽。

[0183] 去除或减少内切葡聚糖酶活性

[0184] 本发明还涉及用于产生亲本细胞突变体的方法,其包含破坏或缺失编码本发明的多肽的多核苷酸序列或其部分,所述方法导致在相同条件下培养时,与亲本细胞相比突变的细胞产生较少的所述多肽。

[0185] 可以使用本领域熟知的方法通过减少或消除编码本发明多肽的核苷酸序列的表

达,例如,插入、破坏、取代或缺失来构建突变细胞。在优选的方面,所述核苷酸序列是失活的。待修饰或失活的核苷酸序列可以是,例如,编码区或其对活性关键的部分,或表达编码区必需的调节元件。这种调节或调控序列的实例可以是启动子序列或其功能部分,即,足以影响核苷酸序列表达的部分。用于可能的修饰的其它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、信号肽序列、转录终止子和转录激活子。

[0186] 可以通过向亲本细胞施以诱变,并且选择其中已将所述核苷酸序列的表达减少或消除的突变细胞来进行核苷酸序列的修饰或失活。诱变可能是特异性的或随机的,可以通过例如使用合适的物理或化学诱变剂进行,通过使用合适的寡核苷酸进行,或通过所述 DNA 序列进行 PCR 产生的诱变。此外,可以通过使用这些诱变剂的任何组合来进行诱变。

[0187] 适合于本发明目的的物理或化学诱变剂的实例包括紫外线 (UV) 照射、羟胺、N-甲基-N'-硝基-N-亚硝基胍 (MNNG)、O-甲基羟胺、亚硝酸、乙基甲烷磺酸酯 (ethyl methane sulphonate) (EMS)、亚硫酸氢钠、甲酸和核苷酸类似物。

[0188] 当使用这些试剂时,通常通过如下方法来进行所述诱变:在合适条件下存在优选的诱变剂时孵育待诱变的亲本细胞,并筛选和/或选择显示基因表达减少的或无基因表达的突变体细胞。

[0189] 通过导入、取代,或去除基因中的一个或多个核苷酸或其转录或翻译所必需的调控元件可以实现所述核苷酸序列的修饰或失活。例如,可以插入或删除核苷酸从而导致引入终止密码子,去除起始密码子,或改变开放阅读框。按照本领域已知的方法通过定位诱变或 PCR 产生的诱变可以实现这种修饰或失活。尽管在理论上所述修饰可以在体内进行,即,直接在表达待修饰核苷酸序列的细胞上进行,但优选如下面所示例的那样在体外进行所述修饰。

[0190] 消除或减少核苷酸序列通过细胞表达的便利方法的例子是基于基因取代,基因缺失,或基因破坏的技术。例如,在基因破坏方法中,将相应于内源核苷酸序列的核酸序列在体外进行诱变以产生缺陷性的核苷酸序列,随后将其转化入亲本细胞中以产生缺陷基因。通过同源重组,所述缺陷性核酸序列替代了内源性核苷酸序列。理想的是所述缺陷性核苷酸序列还编码标记,其可用于选择其中核苷酸序列被修饰或破坏的转化子。在特别优选的方面,用可选择的标记,如本文所述的那些来破坏所述核苷酸序列。

[0191] 或者,可以使用与所述核苷酸序列互补的序列通过确定的反义或 RNAi 技术来进行所述核苷酸序列的修饰或失活。更具体地,通过导入与所述基因的核苷酸序列互补的序列可以减少或消除所述核苷酸序列通过细胞的表达,所述序列可以在细胞中转录并且能够与细胞中产生的 mRNA 杂交。在允许所述互补反义核苷酸序列与 mRNA 杂交的条件下,由此减少或消除翻译的蛋白质的量。

[0192] 本发明进一步涉及亲本细胞的突变体细胞,其包含编码多肽的核苷酸序列或其调控序列的破坏或缺失,这导致与亲本细胞相比突变体细胞产生更少的多肽或不产生多肽。

[0193] 这样生成的多肽缺陷型突变细胞作为表达同源和/或异源多肽的宿主细胞特别有用。所以,本发明进一步涉及生产同源或异源多肽的方法,其包括:(a) 在有益于生产多肽的条件下培养突变细胞;和(b) 回收所述多肽。术语“异源多肽”在本文中定义为对宿主细胞不是天然的多肽,进行了修饰以改变天然序列的天然蛋白,或作为通过重组 DNA 技术对宿主细胞操作的结果其表达在量上改变的天然蛋白。

[0194] 在另一方面,本发明涉及通过发酵可产生本发明的多肽以及目标蛋白产物的细胞来生产基本上无内切葡聚糖酶活性的蛋白质产品的方法:在发酵之前、之中,或发酵完成之后向发酵液(fermentation broth)中添加有效量的能够抑制内切葡聚糖酶活性的试剂,从发酵液中回收目标产物,并且任选地将回收的产物进行进一步纯化。

[0195] 在另一方面,本发明涉及如下生产基本上无内切葡聚糖酶活性的蛋白质产品的方法:在允许产物表达的条件下培养细胞,将得到的培养液进行组合的 pH 和温度处理以基本上减少内切葡聚糖酶活性,和从发酵液中回收产物。或者,可以将从培养液回收的酶制备物进行组合的 pH 和温度处理。所述组合的 pH 和温度处理可任选地与内切葡聚糖酶抑制剂处理组合使用。

[0196] 依照本发明的这个方面,可能去除至少 60%,优选至少 75%,更优选至少 85%,还更优选至少 95%,并且最优选至少 99%的内切葡聚糖酶活性。可使用此方法获得内切葡聚糖酶活性的完全去除。

[0197] 组合的 pH 和温度处理优选在 2-3 或 10-11 范围内的 pH 和至少 75-85°C 范围内的温度进行一段足够的时间以达到期望的效果,通常 1 至 3 小时是足够的。

[0198] 用于培养和纯化感兴趣的产物的方法可以通过本领域已知的方法进行。

[0199] 用于产生基本上无内切葡聚糖酶产物的本发明的方法在真核多肽,特别是真菌蛋白质例如酶的产生中是特别令人有兴趣的。所述酶可以选自,例如,淀粉分解酶(amylolytic enzyme)、脂肪分解酶、蛋白水解酶、纤维素分解酶(cellulolytic enzyme)、氧化还原酶或植物细胞壁降解酶。这些酶的实例包括氨肽酶、淀粉酶、淀粉葡萄糖苷酶、糖酶、羧肽酶、过氧化氢酶、纤维素酶、几丁质酶(chitinase)、角质酶(cutinase)、环糊精糖基转移酶、脱氧核糖核酸酶、酯酶、半乳糖苷酶、 β -半乳糖苷酶、葡糖淀粉酶、葡糖氧化酶、葡糖苷酶、卤素过氧化酶、半纤维素酶、转化酶、异构酶、漆酶、连接酶、脂肪酶、裂合酶、甘露糖苷酶、氧化酶、果胶分解酶、过氧化物酶、肌醇六磷酸酶、酚氧化酶、多酚氧化酶、蛋白水解酶、核糖核酸酶、转移酶、转谷氨酰胺酶或木聚糖酶。内切葡聚糖酶缺陷细胞也可以用于表达在制药上引起兴趣的异源蛋白质例如激素、生长因子、受体等。

[0200] 可理解的是术语“真核多肽”不仅包括天然多肽,也包括多肽例如酶,其通过氨基酸取代、缺失或添加修饰,或通过其它这样的修饰以增强活性、热稳定性、pH 耐受性等。

[0201] 在另外的方面,本发明涉及基本上无内切葡聚糖酶活性的蛋白质产物,其通过本发明的方法产生。

[0202] 组合物

[0203] 本发明还涉及包含本发明的多肽的组合物。优选地,所述组合物富含这种多肽。术语“富含”表示所述组合物的内切葡聚糖酶活性,例如,以至少 1.1 的富集因数(enrichment factor)增加。

[0204] 所述组合物可以包含本发明的多肽作为主要酶成分,例如,单成分组合物。或者,所述组合物可以包含多种酶活性,例如氨肽酶、淀粉酶、糖酶、羧肽酶、过氧化氢酶、纤维素酶、几丁质酶、角质酶、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 -葡糖苷酶、卤素过氧化酶、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、肌醇六磷酸酶、多酚氧化酶、蛋白水解酶、核糖核酸酶、转谷氨酰胺酶或木聚糖酶。额外的酶可

以通过例如属于以下属或种的微生物产生：曲霉属，优选棘孢曲霉、泡盛曲霉、烟曲霉、臭曲霉、日本曲霉、构巢曲霉、黑曲霉或米曲霉；镰孢属，优选杆孢状镰孢、禾谷镰孢、库威镰孢、大刀镰孢、禾本科镰孢、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、硫色镰孢、圆镰孢、拟丝孢镰孢或镶片镰孢；腐质霉属，优选特异腐质霉或疏棉状腐质霉；或木霉属，优选哈茨木霉、康宁木霉、长枝木霉、里氏木霉或绿色木霉。

[0205] 可以依照本领域内已知的方法制备多肽组合物，并且可以是液体或干组合物的形式。例如，所述多肽组合物可以是颗粒 (granulate) 或微粒 (microgranulate) 的形式。可以将包括于所述组合物中的多肽依照本领域内已知方法稳定。

[0206] 以下提供的实施例是本发明多肽组合物的优选的用途。本发明的多肽组合物的剂量和使用该组合物的其它条件可以基于本领域内已知的方法决定。

[0207] 用途

[0208] 本发明还涉及使用具有内切葡聚糖酶活性的多肽或其组合物的方法。

[0209] 将生物质降解为单糖、二糖和多糖

[0210] 本发明的具有内切葡聚糖酶活性的多肽和宿主细胞可用于产生单糖、二糖和多糖，它们作为来自生物质的化学或发酵原料用于产生乙醇、塑料或其它产品或中间物。具有内切葡聚糖酶活性的多肽可以是去除或不去除细胞的粗发酵液形式，或是半纯化或纯化的酶制剂形式。或者，本发明的宿主细胞可以在使用生物质的发酵方法中用作具有内切葡聚糖酶活性的多肽的来源。

[0211] 生物质可以包括但不限于木材资源、城市固体废物、废纸和作物残余物（参见，例如 Wiseloge et al., 1995, 在 Handbook on Bioethanol (编者 Charles E. Wyman) 中，第 105-118 页, Taylor&Francis, Washington D.C. ;Wyman, 1994, Bioresource Technology 50 :3-16 ;Lynd, 1990, Applied Biochemistry and Biotechnology 24/25 :695-719 ;Mosier et al., 1999, Recent Progress in Bioconversion of Lignocellulosics, 在 Advances in Biochemical Engineering/Biotechnology 中，总编 T. Scheper, 第 65 卷，第 23-40 页, Springer-Verlag, New York)。

[0212] 生物质初生细胞壁 (primary cell wall) 中的主要多糖是纤维素，其次最丰富的是半纤维素，而第三是果胶。次生细胞壁 (secondary cell wall) 在细胞停止生长后产生，其同样含有多糖并通过聚合木质素共价交联至半纤维素而加强。纤维素是脱水纤维二糖的均聚物，并且因此是线性 β -(1-4)-D- 葡聚糖，而半纤维素包括多种化合物，例如木聚糖、木葡聚糖 (xyloglucan)、阿拉伯木聚糖和甘露聚糖，具有系列取代基的复杂分支结构。尽管通常是多形的，存在于植物组织中的纤维素主要作为平行葡聚糖链的不溶晶体基质。半纤维素通常与纤维素以及其它半纤维素以氢键相连，其帮助稳定细胞壁基质。

[0213] 将三种主要类别的糖水解酶 (glycohydrolase) 用于破坏纤维素生物质：

[0214] (1) “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)，其随机作用于可溶和不溶的 1,4- β - 葡聚糖底物。

[0215] (2) “外切-1,4- β - 葡聚糖酶”包括 1,4- β -D- 葡聚糖葡萄糖水解酶 (EC 3. 2. 1. 74)，其从 1,4- β -D- 葡聚糖释放 D- 葡萄糖并且缓慢水解 D- 纤维二糖；和纤维二糖水解酶 (1,4- β -D- 葡聚糖纤维二糖水解酶, EC 3. 2. 1. 91)，其从 1,4- β - 葡聚糖释放 D- 纤维二糖。

[0216] (3) “ 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)，其作用于从纤

维二糖和可溶的纤维糊精以及一系列糖苷释放 D- 葡萄糖单位。

[0217] 这三个类别的酶一起协同地作用使来自生物质的天然纤维素有效解晶 (decrystallization) 和水解以产生还原糖。

[0218] 本发明具有内切葡聚糖酶活性的多肽可以与上述酶一同使用以进一步降解生物质底物的纤维素成分 (参见, 例如, Brigham et al., 1995, 在 Handbook on Bioethanol (编者 Charles E. Wyman) 中, 第 119-141 页, Taylor & Francis, Washington D.C.; Lee, 1997, Journal of Biotechnology 56:1-24)。

[0219] 能够通过酶降解生物质并将释放的糖转化为乙醇来产生乙醇。这种乙醇通常称为生物乙醇或生物燃料。可将其以少于 1% - 多至 100% (燃料替代品) 的混合物作为燃料添加剂或补充剂 (extender) 使用。

[0220] 洗涤剂 (detergent) 组合物

[0221] 可以将本发明的具有内切葡聚糖酶活性的多肽添加至洗涤剂组合物, 并且因此成为所述洗涤剂组合物的成分。

[0222] 可以将本发明的洗涤剂组合物例如配制为手洗或机洗的洗涤剂组合物, 包括适于预处理沾污织物的洗衣添加剂组合物和添加了漂洗剂的织物柔软剂组合物, 或将其配制为用于通常家庭硬表面清理操作的洗涤剂组合物, 或配制用于手洗或机洗的洗碟 (dishwashing) 操作。

[0223] 在具体的方面, 本发明提供包含本发明的具有内切葡聚糖酶活性的多肽的洗涤剂添加剂。所述洗涤剂添加剂以及洗涤剂组合物可以包含一个或多个其它酶, 例如蛋白酶、脂肪酶、角质酶、淀粉酶、糖酶、纤维素酶、果胶酶、甘露聚糖酶、阿拉伯糖酶 (arabinase)、半乳聚糖酶、木聚糖酶、氧化酶, 例如, 漆酶和 / 或过氧化物酶。

[0224] 通常酶成分的性质应该与选择的洗涤剂相容 (即, 最适 pH, 与其它酶和非酶成分的相容性等), 并且所述酶成分应该以有效量存在。

[0225] 蛋白酶: 合适的蛋白酶包括动物、植物或微生物来源的那些蛋白酶。微生物来源是优选的。其中包括化学修饰的或蛋白质工程化的突变体。所述蛋白酶可以是丝氨酸蛋白酶或金属蛋白酶, 优选碱性微生物蛋白酶或胰蛋白酶样蛋白酶。碱性蛋白酶的实例是枯草杆菌蛋白酶, 特别是源自芽孢杆菌属的那些, 例如, 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 (在 WO 89/06279 中描述)。胰蛋白酶样蛋白酶的实例是胰蛋白酶 (例如, 猪或牛来源) 和镰孢属蛋白酶, 在 WO 89/06270 和 WO 94/25583 中描述。

[0226] 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, 特别是在一个或多个以下位置具有取代的变体: 27、36、57、76、87、97、101、104、120、123、167、170、194、206、218、222、224、235 和 274。

[0227] 优选的商业上能够获得的蛋白酶包括 Alcalase™、Savinase™、Primase™、Duralase™、Esperase™ 和 Kannase™ (Novozymes A/S)、Maxatase™、Maxacal™、Maxapem™、Properase™、Purafect™、Purafect OxP™、FN2™ 和 FN3™ (Genencor International Inc.)。

[0228] 脂肪酶: 合适的脂肪酶包括细菌或真菌来源的那些。包括化学修饰的或蛋白质工程化的突变体。有用的脂肪酶的实例包括来自如下的脂肪酶: 腐质霉属 (同物异名嗜热霉属 (Thermomyces)), 例如, 来自疏棉状腐质霉 (细毛嗜热霉 (T. lanuginosus)) 如 EP

258068 和 EP 30526 中所述,或来自特异腐质霉如 WO 96/13580 中所述,假单胞菌属脂肪酶,例如,来自产碱假单胞菌 (*P. alcaligenes*) 或类产碱假单胞菌 (*P. pseudoalcaligenes*) (EP 218272)、洋葱假单胞菌 (*P. cepacia*) (EP 331376)、施氏假单胞菌 (*P. stutzeri*) (GB 1,372,034)、荧光假单胞菌 (*P. fluorescens*)、假单胞菌属菌种 (*Pseudomonas* sp.) 菌株 SD 705 (WO95/06720 和 WO 96/27002)、Pwiconsinensis (WO 96/12012)、芽孢杆菌属脂肪酶,例如,来自枯草芽孢杆菌 (Dartois et al., 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. pumilus*) (WO 91/16422) 的脂肪酶。

[0229] 其它实例是例如在 WO 92/05249、WO 94/01541、EP 407225、EP 260105、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 中描述的那些脂肪酶变体。

[0230] 优选的商业上能够获得的脂肪酶包括 Lipolase[™]、Lipex[™] 和 LipolaseUltra[™] (Novozymes A/S)。

[0231] 淀粉酶:合适的淀粉酶 (α 和 / 或 β) 包括细菌和真菌来源的那些。包括化学修饰的或蛋白质工程化的变体。淀粉酶包括例如从芽孢杆菌属获得的 α -淀粉酶,例如,地衣芽孢杆菌的特殊菌株,在 GB 1,296,839 中有更详细的描述。

[0232] 有用的淀粉酶的实例是在 WO 94/02597、WO 94/18314、WO 96/23873 和 WO 97/43424 中描述的变体,特别在一个或多个以下位置具有取代的变体:15、23、105、106、124、128、133、154、156、181、188、190、197、202、208、209、243、264、304、305、391、408 和 444。

[0233] 商业上能够获得的淀粉酶是 Duramyl[™]、Termamyl[™]、Fungamyl[™] 和 BAN[™] (Novozymes A/S)、Rapidase[™] 和 Purastar[™] (来自 Genencor International Inc.)。

[0234] 纤维素酶:合适的纤维素酶包括细菌或真菌来源的那些。包括化学修饰或蛋白质工程的突变体。合适的纤维素酶包括来自芽孢杆菌属、假单胞菌属、腐质霉属、镰孢属、梭孢壳属、枝顶孢霉属或木霉属菌属的纤维素酶,例如,产生自特异腐质霉、嗜热毁丝霉和尖镰孢的真菌纤维素酶,其公开于美国专利号 4,435,307、美国专利号 5,648,263、美国专利号 5,691,178、美国专利号 5,776,757 和 WO 89/09259。

[0235] 特别合适的纤维素酶是碱性或中性纤维素酶,其具有保护颜色的益处 (colour care benefits)。这种纤维素酶的实例是在 EP 0495257、EP 0531372、WO 96/11262、WO 96/29397、WO 98/08940 中描述的纤维素酶。其它实例是纤维素酶变体,例如在 WO 94/07998、EP 0531315、美国专利号 5,457,046、美国专利号 5,686,593、美国专利号 5,763,254、WO 95/24471、WO 98/12307 和 PCT/DK98/00299 中描述的那些。

[0236] 商业上能够获得的纤维素酶包括 Celluclast[®]、Celluzyme[™] 和 Carezyme[™] (Novozymes A/S)、Clazinase[™] 和 Puradax HA[™] (Genencor International Inc.) 和 KAC-500 (B)[™] (Kao Corporation)。

[0237] 过氧化物酶/氧化酶:合适的过氧化物酶/氧化酶包括植物、细菌或真菌来源的那些。包括化学修饰或蛋白质工程的突变体。有用的过氧化物酶的实例包括来自鬼伞属的过氧化物酶,例如,来自灰盖鬼伞及其变体,如在 WO93/24618、WO 95/10602 和 WO 98/15257 中描述的那些。

[0238] 商业上能够获得的过氧化物酶包括 Guardzyme™ (Novozymes A/S)。

[0239] 通过添加含有一种或多种酶的单独的添加剂,或通过添加包含所有这些酶的组合添加剂,可将酶成分包括在洗涤剂组合物中。可将本发明的洗涤剂添加剂(即单独的添加剂或组合的添加剂)配制成例如,颗粒、液体、浆等。优选的洗涤剂添加剂剂型是颗粒,具体为无粉尘(non-dusting)颗粒,液体,具体为稳定的液体,或浆。

[0240] 例如,可以如美国专利号 4,106,991 和 4,661,452 中所公开的产生无粉尘颗粒,并且可以任选地通过本领域已知的方法包覆。蜡制包覆材料(waxy coating material)的实例是具有 1000 至 20000 的平均摩尔量的聚(环氧乙烷)产品(聚乙二醇,PEG);具有从 16 至 50 个的环氧乙烷单元的乙氧基化壬基酚;乙氧基化脂肪醇,其中醇含有从 12 至 20 个碳原子,并且其中具有 15 至 80 个环氧乙烷单元;脂肪醇;脂肪酸;和脂肪酸的单酸甘油酯和甘油二酯和甘油三酯。适合于流化床技术应用的成膜涂覆材料的实例提供于 GB 1483591。例如,可根据已经建立的方法通过添加多元醇例如丙二醇、糖或糖醇、乳酸或硼酸来稳定液体酶制剂。可以根据公开于 EP 238,216 中的方法来制备保护酶。

[0241] 本发明的洗涤剂组合物可以是任何方便的形式,例如,条、片剂、粉剂、颗粒、糊剂或液体。液体洗涤剂可以是含水的,通常含有多至 70% 的水和 0-30% 的有机溶剂,或非水的(non-aqueous)。

[0242] 洗涤剂组合物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。所述表面活性剂通常以按重量计 0.1% 至 60% 的水平存在。

[0243] 当包括于此时,所述洗涤剂将通常含有大约 1% 至大约 40% 的阴离子表面活性剂,例如线性烷基苯磺酸盐、 α -烯烃磺酸盐(olefinsulfonate)、烷基硫酸盐(脂肪醇硫酸酯)、醇乙氧基硫酸盐(alcohol ethoxysulfate)、仲链烷磺酸盐(secondaryalkanesulfonate)、 α -磺基脂肪酸甲酯、烷基-或烯基琥珀酸,或肥皂(soap)。

[0244] 当包括于此时,洗涤剂将通常含有大约 0.2% 至大约 40% 的非离子表面活性剂,例如醇乙氧基化物、壬基苯酚乙氧基化物、烷基聚糖苷

[0245] (alkylpolyglycoside)、烷基二甲基氧化胺(alkyldimethylamineoxide)、乙氧基化脂肪酸单乙醇酰胺(ethoxylated fatty acid monoethanolamide)、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(glucamide)”)。

[0246] 洗涤剂可以含有 0-65% 的洗涤剂增清剂(builder)或复合剂例如沸石、二磷酸盐、三磷酸盐、膦酸酯(phosphonate)、碳酸盐(carbonate)、柠檬酸盐、氮川三乙酸(nitrilotriacetic acid)、乙二胺四乙酸、二亚乙基三胺五乙酸、烷基-或烯基琥珀酸、可溶硅酸盐或分层硅酸盐(例如来自 Hoechst 的 SKS-6)。

[0247] 洗涤剂可包含一种或多种聚合物。实例是羧甲基纤维素、聚(乙烯基吡咯烷酮)、聚(乙二醇)、聚(乙烯醇)、聚(乙烯吡啶-N-氧化物)、聚(乙烯咪唑)、聚羧酸酯(polycarboxylates)例如聚丙烯酸酯(polyacrylates)、马来酸/丙烯酸共聚物和甲基丙烯酸月桂酯/丙烯酸共聚物。

[0248] 洗涤剂可以含有漂白体系,其可以包含 H_2O_2 源例如过硼酸盐或过碳酸盐,其可以与形成过酸的漂白激活剂例如四乙酰乙二胺或壬酰氧苯磺酸酯(nonanoyloxybenzenesulfonate)组合。或者,漂白体系可以包含过氧酸,例如,酰胺、二酰

亚胺 (imide) 或砜类型的过氧酸。

[0249] 本发明的洗涤剂组合物的酶成分可以使用常规稳定剂稳定,例如,多元醇例如丙二醇或甘油、糖或糖醇、乳酸、硼酸或硼酸衍生物,例如,芳香硼酸酯,或苯基硼酸 (phenyl boronic acid) 衍生物例如 4-甲酰苯基硼酸 (4-formylphenylboronic acid),并可例如 WO 92/19709 和 WO 92/19708 中所述配制所述组合物。

[0250] 洗涤剂还可以含有其它常规洗涤剂成分,例如,织物整理剂 (fabric conditioner) 包括粘土、泡沫促进剂、抑泡剂、防腐蚀剂、悬污剂、防污再沉积剂、染料、杀菌剂、光亮剂 (optical brightener)、水溶助剂 (hydrotrope)、晦暗抑制剂 (tarnish inhibitors) 或香料。

[0251] 在洗涤剂组合物中的任何酶成分,特别是本发明的具有内切葡聚糖酶活性的多肽,可以按相当于每升洗涤液 0.01-100mg 酶蛋白,优选每升洗涤液 0.05-5mg 酶蛋白,特别是每升洗涤液 0.1-1mg 酶蛋白的量添加。

[0252] 本发明的具有内切葡聚糖酶活性的多肽可以额外地并入公开于 WO97/07202 的洗涤剂配制物,WO 97/07202 在本文中作为参考文献并入。

[0253] 信号肽

[0254] 本发明还涉及核酸构建体,所述核酸构建体包含编码蛋白质的基因,该基因与编码信号肽的核苷酸序列可操作地连接,所述信号肽包含 SEQ ID NO:2 的氨基酸 1 至 17 或由 SEQ ID NO:2 的氨基酸 1 至 17 组成,其允许蛋白质分泌到培养基中,其中所述基因对于该核苷酸序列是外源的。

[0255] 在优选的方面,所述核苷酸序列包含 SEQ ID NO:1 的核苷酸 1 至 51。在另外的优选方面,所述核苷酸序列由 SEQ ID NO:1 的核苷酸 1 至 51 组成。

[0256] 本发明还涉及包含这种核酸构建体的重组表达载体和重组宿主细胞。

[0257] 本发明还涉及用于产生蛋白质的方法,包括:(a) 在适合于产生所述蛋白质的条件下培养这样的重组细胞;和 (b) 回收所述蛋白质。

[0258] 所述蛋白质对于宿主细胞可以是天然的或异源的。术语“蛋白质”在本文的意思不是指特定长度的编码产物,并且因此包含肽、寡肽和蛋白质。术语“蛋白质”还包含组合以形成编码产物的两种或两种以上多肽。所述蛋白质还包括杂合多肽,其包含部分或全部多肽序列的组合,所述多肽序列从至少两种不同的蛋白质获得,其中一个或多个对于宿主细胞可以是异源或天然的。蛋白质进一步包括上述蛋白质和杂合蛋白质天然存在的等位基因变异和工程的变异。

[0259] 优选蛋白质是激素或其变体、酶、受体或其部分、抗体或其部分,或报道蛋白 (reporter)。在更优选的方面,所述蛋白质是氧化还原酶、转移酶、水解酶、裂合酶 (lyase)、异构酶或连接酶。在更加优选的方面,所述蛋白质是氨肽酶、淀粉酶、糖酶、羧肽酶、过氧化氢酶、纤维素酶、几丁质酶、角质酶、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 -葡糖苷酶、转化酶、漆酶、脂肪酶、甘露糖苷酶、变位酶 (mutanase)、氧化酶、果胶分解酶、过氧化物酶、肌醇六磷酸酶、多酚氧化酶、蛋白水解酶、核糖核酸酶、转谷氨酰胺酶或木聚糖酶。

[0260] 基因可以从任何原核、真核或其它来源获得。

[0261] 通过以下实施例进一步对本发明进行描述,但不应将其理解为对本发明范围的限

制。

[0262] 实施例

[0263] 材料

[0264] 作为缓冲液和底物使用的化学品是至少试剂等级的商业产品。

[0265] SDS-PAGE 凝胶、上样缓冲液和电泳缓冲液获得自 Invitrogen/Novex (Carlsbad, CA)。测序等级修饰的胰蛋白酶来自 Princeton Separations (Aldelphia, NJ)。BioSafe Commassie Blue G250 蛋白质染色剂 (protein stain) 获得自 BioRad Laboratories (Hercules, CA)。

[0266] 菌株

[0267] 将米曲霉 J1250 菌株 (WO 99/61651) 用于表达具有内切葡聚糖酶活性的土生梭孢霉多肽。将土生梭孢霉 NRRL 菌株 8126 用作具有内切葡聚糖酶活性的 7F 家族多肽 (Family 7F polypeptide) 的基因的来源。

[0268] 培养基

[0269] PDA 平板由每升 39 克马铃薯葡糖琼脂组成。

[0270] NNCYP 培养基由如下物质组成：每升 5.0g NH_4NO_3 、0.5g $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ 、0.3g CaCl_2 、2.5g 柠檬酸、1.0g 细菌用蛋白胨 (Bacto Peptone)、5.0g 酵母提取物、1ml COVE 痕量金属，和足够的 K_2HPO_4 以达到最终 pH 大约 5.4。

[0271] NNCYPmod 培养基由如下物质组成：每升 1.0g NaCl 、5.0g NH_4NO_3 、0.2g $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ 、0.2g CaCl_2 、2.0g 柠檬酸、1.0g 细菌用蛋白胨、5.0g 酵母提取物、1ml COVE 痕量金属溶液，和足够的 K_2HPO_4 以达到最终 pH 大约 5.4。

[0272] Cove 痕量金属溶液由如下物质组成：每升 0.04g $\text{Na}_2\text{B}_4\text{O}_7 \cdot 10\text{H}_2\text{O}$ 、0.4g $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ 、1.2g $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ 、0.7g $\text{MnSO}_4 \cdot \text{H}_2\text{O}$ 、0.8g $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$ 和 10g $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ 。

[0273] LB 平板由如下物质组成：每升 10g 胰蛋白胨、5g 酵母提取物、5g 氯化钠和 15g 细菌用琼脂 (Bacto Agar)。

[0274] MDU2BP 培养基由下述物质组成：每升 45g 麦芽糖、1g $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ 、1g NaCl 、2g K_2HSO_4 、12g KH_2PO_4 、2g 尿素和 500 μl AMG 痕量金属，将 pH 调节至 5.0 并且随后用 0.22 μm 过滤装置过滤除菌。

[0275] AMG 痕量金属由如下物质组成：每升 14.3g $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ 、2.5g $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ 、0.5g $\text{NiCl}_2 \cdot 6\text{H}_2\text{O}$ 、13.8g $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ 、8.5g $\text{MnSO}_4 \cdot 7\text{H}_2\text{O}$ 和 3g 柠檬酸。

[0276] SOC 培养基由如下物质组成：2% 胰蛋白胨、0.5% 酵母提取物、10mM NaCl 、2.5mM KCl 、10mM MgCl_2 和 10mM MgSO_4 ；通过高压灭菌并且随后添加过滤除菌的葡萄糖至 20mM。

[0277] 冷冻培养基由 60% SOC 和 40% 甘油组成。

[0278] 2X YT 培养基由如下物质组成：每升 16g 胰蛋白胨、10g 酵母提取物、5g NaCl 和 15g 细菌用琼脂，通过高压灭菌。

[0279] 实施例 1：表达序列标签 (EST) cDNA 文库构建

[0280] 将土生梭孢霉 NRRL 8126 在 250ml 摇瓶中补充 1% 葡萄糖的 50ml NNCYPmod 培养基中，在 45°C、200rpm 培养 24 小时。将来自所述 24 小时液体培养物的 2ml 等分试样用于接种 500ml 摇瓶，所述 500ml 摇瓶含有补充 2% Sigmacell-20 (Sigma Chemical Co., St. Louis, MO) 的 100ml NNCYPmod 培养基。将培养物在 45°C、200rpm 温育 3 天。通过过滤收获菌丝体，

所述过滤通过具有玻璃纤维预过滤器的布氏漏斗 (Buchner funnel) (Nalgene, Rochester NY) 进行,将菌丝体用 10mM Tris-HCl-1mM EDTA pH8(TE) 洗涤两次,并且在液氮中快速冷冻。

[0281] 使用如下方法分离总 RNA。将土生梭孢霉 NRRL 8126 的冷冻菌丝体在电动咖啡研磨机中研磨。在 50ml Falcon 管中将研磨的材料与 20ml Fenazol (Ambion, Inc., Austin, TX) 1 : 1v/v 混合。一旦将菌丝体悬浮,使用氯仿提取并且使用苯酚 - 氯仿 - 异戊醇 25 : 24 : 1v/v/v 的混合物提取三次。通过添加 1/10 体积的 3M 乙酸钠 pH 5.2 和 1.25 体积的异丙醇从所得水相沉淀 RNA。通过在 4°C 12,000xg 离心 30 分钟回收沉淀的 RNA。将最终的沉淀用冷的 70%乙醇洗涤,空气干燥,并且在 500ml 焦碳酸二乙酯处理的水 (DEPC 水) 中重悬。

[0282] 用 Agilent Bioanalyzer 2100 (Agilent Technologies, Inc., Palo Alto, CA) 评估纯化的 RNA 的质和量。将聚腺苷酸化 mRNA 从 360 μ g 总 RNA 中分离,所述分离借助 Poly (A) Purist Magnetic Kit (Ambion, Inc., Austin, TX) 根据制造商的说明书进行。

[0283] 为了产生 cDNA 文库,使用 CloneMiner™ Kit (Invitrogen, Carlsbad, CA) 来构建定向文库 (directional library),其不需要使用限制酶克隆,因此降低嵌合克隆的数目和大小偏差 (size bias)。

[0284] 为了确保 cDNA 的第一链合成成功,平行进行具有两种不同浓度 mRNA (2.2 和 4.4 μ g 的聚 (A)⁺mRNA) 的两个反应。将 mRNA 样品与 Biotin-attB2-Oligo(dt) 引物 (CloneMiner™ Kit, Invitrogen, Carlsbad, CA)、1X 第一链缓冲液 (Invitrogen, Carlsbad, CA)、2 μ l 的 0.1M 二硫苏糖醇 (DTT)、10mM 各种 dNTP 和水分别混合至终体积 18 和 16 μ l。将反应混合物小心地混合,随后添加 2 和 4 μ l 的 SuperScript™ 反转录酶 (Invitrogen, Carlsbad, CA),并且在 45°C 温育 60 分钟以合成第一互补链。

[0285] 为了第二链的合成,向各个第一链反应添加 30 μ l 的 5X 第二链缓冲液 (Invitrogen, Carlsbad, CA)、3 μ l 的 10 mM 各种 dNTP、10 单位大肠杆菌 DNA 连接酶 (Invitrogen, Carlsbad, CA)、40 单位大肠杆菌 DNA 聚合酶 I (Invitrogen, Carlsbad, CA) 和 2 单位大肠杆菌 RNase H (Invitrogen, Carlsbad, CA) 至总体积 150 μ l。随后将混合物在 16°C 温育 2 小时。两小时温育之后,向各个反应添加 2 μ l T4 DNA 聚合酶 (Invitrogen, Carlsbad, CA),并且在 16°C 温育 5 分钟以产生平端化的 cDNA。使用苯酚 - 氯仿 - 异戊醇 25 : 24 : 1v/v/v 的混合物提取 cDNA 反应物并且在 20 μ g 糖原、120 μ l 5M 乙酸铵和 660 μ l 乙醇存在下沉淀。在 4°C 12,000xg 离心 30 分钟之后,将 cDNA 沉淀用冷的 70%乙醇洗涤,在真空中干燥 2-3 分钟,并且在 18 μ l DEPC 水中重悬。向各个重悬的 cDNA 样品中添加 10 μ l 5X 连接 (adapted) 缓冲液 (Invitrogen, Carlsbad, CA)、10 μ g 如下所示的 attB1 衔接头 (adapter) (Invitrogen, Carlsbad, CA)、7 μ l 0.1M DTT 和 5 单位的 T4 DNA 连接酶 (Invitrogen, Carlsbad, CA)。

[0286] attB1 衔接头顶链 (top strand) :

[0287] 5' -TCGTCGGGGACAACTTTGTACAAAAAAGTTGG-3' (SEQ ID NO :3)

[0288] attB1 衔接头底链 (bottom strand) :

[0289] 3' -CCCCTGTTGAAACATGTTTTTCAACCP-5' (SEQ ID NO :4)

[0290] 连接反应在 16°C 温育过夜。过量的衔接头通过在 1ml Sephacryl™ S-500HR 树脂

(Amersham Biosciences, Piscataway, NJ) 中的大小排阻层析去除。根据 CloneMiner™ Kit 的说明书收集柱级分, 并使用 Agilent Bioanalyzer 分析级分 3 至 14 以测定其中 attB1 接头开始洗脱的级分。这种分析显示接头在大约级分 10 或 11 开始洗脱。对于第一文库汇集级分 6 至 11, 而对于第二文库汇集级分 4-11。

[0291] 通过根据 Gateway System 规程 (Invitrogen, Carlsbad, CA) 的同源 DNA 重组, 使用 BP Clonase™ (Invitrogen, Carlsbad, CA) 作为重组酶来进行 cDNA 的克隆。每个 BP Clonase™ 重组反应含有大约 70ng 的 attB 侧接的 cDNA (attB-flanked-cDNA)、250ng pDONR™222、2 μl 5X BP Clonase™ 缓冲液、2 μl TE 缓冲液和 3 μl BP Clonase™。全部试剂获得自 Invitrogen, Carlsbad, CA。将重组反应在 25℃ 温育过夜。

[0292] 随后将热失活的 BP 重组反应分成 6 个等分试样, 并且使用 BioRad GenePulser II (BioRad, Hercules, CA) 用如下参数电穿孔入 ElectroMax™ DH10B 电感受态细胞 (Invitrogen, Carlsbad, CA) 中: 电压: 2.0kV; 电阻: 200 Ω; 和电容 25 μF。将电穿孔的细胞在 1ml SOC 培养基中重悬, 并且在 37℃ 及 200rpm 的持续摇动下温育 60 分钟。温育期之后, 汇集转化的细胞并且与冷冻培养基 1:1 混合。将 200 μl 等分试样取出用于文库滴定, 并且随后将剩余的每种文库等分至 1.8ml 冷冻管 (Wheaton Science Products, Millville, NJ) 中, 并在 -80℃ 冷冻贮藏。

[0293] 制备每个文库的四个系列的稀释物: 1/100、1/1000、1/10⁴、1/10⁵。从每种稀释物将 100 μl 涂布于补充 50 μg 每 ml 卡那霉素的 150mm LB 平板上并且在 37℃ 温育过夜。计数每个稀释平板上的菌落数目并且用于计算每个文库中转化体的总数。

[0294] 显示第一文库具有 540 万独立克隆而第二文库显示具有 900 万独立克隆。

[0295] 实施例 2: 模板制备和 cDNA 克隆的核苷酸测序

[0296] 将来自两种文库的等分试样混合并且涂布到补充 50 μg 每 ml 卡那霉素的 25x25cm LB 平板上。将单独克隆借助于 Genetix QPix Robot (Genetix Inc., Boston, MA) 排列在 96 孔平板上, 所述平板含有 100 μl 补充 50 μg 每 ml 卡那霉素的 LB。从总共 4320 个单独克隆获得四十五个 96 孔平板。将平板在 37℃ 以 200rpm 摇动温育过夜。温育过后, 将 100 μl 无菌的 50% 甘油添加至每个孔。将转化体借助 96-针工具 (Boekel, Feasterville, PA) 复制到第二个深碟 96 孔微量培养平板 (Advanced Genetic Technologies Corporation, Gaithersburg, MD) 中, 所述深碟 96 孔微量培养平板在每个孔中含有 1ml 补充 50 μg 每 ml 卡那霉素的 Magnificent Broth™ (MacConnell Research, San Diego, CA)。将原始的微量滴定板在 -80℃ 冷冻贮藏。将第二深碟平板在旋转振荡器上以 300rpm 的剧烈搅拌于 37℃ 温育过夜。为了避免溢出和交叉污染, 并且使其充分通气, 将每个第二培养平板用聚丙烯垫 (Advanced Genetic Technologies Corporation, Gaithersburg, MD) 和塑料微量滴定碟盖覆盖。使用 MWG Robot-Smart 384 (MWG Biotech Inc., High Point, NC) 和 Montage Plasmid Miniprep Kits (Millipore, Billerica, MA) 制备质粒 DNA。

[0297] 使用 Big-Dye™ (Applied Biosystems, Inc., Foster City, CA) 终止子化学 (terminator chemistry) (Giesecke et al., 1992, Journal of Virology Methods 38: 47-60) 和如下所示的 M13 Forward (-20) 测序引物进行测序反应。

[0298] 5'-GTAAAACGACGGCCAG-3' (SEQ ID NO:5)

[0299] 使用 Robot-Smart 384 (MWG Biotech Inc., High Point, NC) 以 384 孔形

式进行测序反应,并且使用 Millipore MultiScreen Seq384 Sequencing Clean-up Kits(Millipore,Billerica,MA)进行终止子去除(terminator removal)。反应含有 6 μ l 质粒 DNA 和 4 μ l 测序主混合物(master-mix),所述主混合物含有 2 μ l 5x 测序缓冲液(Millipore,Billerica,MA)、1 μ l Big-Dye™ 终止子(Applied Biosystems, Inc., Foster City, CA)、1.6pmol M13 Forward 引物和 1 μ l 水。使用 ABI PRISMAutomated DNA Sequencer Model 3700(Applied Biosystems, Foster City, CA)进行单通道 DNA 测序。

[0300] 实施例 3 :cDNA 克隆的 DNA 序列数据分析

[0301] 在 PHRED/PHRAP 软件(University of Washington,Seattle,WA)的帮助下进行碱基判定(calling)、质量数值分配和载体修整(vector trimming)。使用 ParcelTranscript Assembler v.2.6.2.(Paracel, Inc., Pasadena, CA)进行 EST 的聚类分析(clustering analysis)。EST 聚类分析显示 395 个独立簇。

[0302] 使用 BLOSUM 62 矩阵(Henikoff,1992, Proc. Natl. Acad. Sci. USA 89 : 10915-10919)在 32 节点 Linux 群集(32-node Linux cluster)(Paracel, Inc., Pasadena, CA)上,以 Blastx 程序(Altschul et. al.,1990, J. Mol. Biol. 215 :403-410)进行装配的 EST 序列针对 PIR 数据库的序列同源性分析。从 395 个簇中,246 个对于公共的蛋白质数据库中的已知基因具有一致的 blast 结果(blast hit),并且 149 个对于这些数据库无显著一致的结果。在这 246 个基因中,13 个具有针对充分表征的糖基水解酶基因同源物的一致结果。

[0303] 实施例 4 :编码家族 7 内切葡聚糖酶(CEL7F)的 cDNA 克隆的鉴定

[0304] 编码家族 7 内切葡聚糖酶(CEL7F)的 cDNA 克隆最初通过其与来自长枝木霉(NREF NF00756647)的家族 7 内切葡聚糖酶 EG-1 蛋白质的同一性而鉴定。此分析显示两种蛋白质在蛋白质水平的 113 个氨基酸(339 个碱基对)的延伸(stretch)内是 44% 同一的。在此初始鉴定之后,从原始冷冻原种(stock)平板恢复(retrieve)克隆 Tter08C4,并且在补充 50 μ g 每 ml 卡那霉素的 LB 平板上划线。将平板在 37℃ 温育过夜,并且在次日使用来自平板的单菌落接种 3ml 补充 50 μ g 每 ml 卡那霉素的 LB。将液体培养物在 37℃ 温育过夜,并且使用 BioRobot 9600(QIAGEN, Inc., Valencia, CA)制备质粒 DNA。使用 M13 Forward 和如下所示的 Poly-T 引物测序克隆的 3' 端,采用如上所述的 Big-Dye™ 终止子化学再次测序克隆 Tter08C4 质粒 DNA。

[0305] 5' -TTTTTTTTTTTTTTTTTTTTTTTTTVN-3' (SEQ ID NO :6)

[0306] 其中 V = G、A、C 和 N = G、A、C、T

[0307] 序列信息的 Blastx 同源性分析显示克隆 Tter08C4 编码的蛋白质与里氏木霉 EG1 蛋白质(NREF NF00494331)相似。这些蛋白质在 365 个氨基酸的延伸段(over a 365 amino acid stretch)是 46% 同一的。

[0308] 使用 Interproscan 程序(Zdobnov and Apweiler,2001, Bioinformatics 17 : 847-8)分析克隆 Tter08C4 的推定蛋白质序列显示,所述基因含有家族 7 蛋白质的序列特征。从起始氨基酸甲硫氨酸的 18 个氨基酸存在这种称为 Pfam 模式 PF00840(Bateman et. al.,2002, Nucleic Acids Research 30 :276-280)的序列特征,这确认克隆 Tter08C4 编码家族 7 内切葡聚糖酶。

[0309] 土生梭孢霉内切葡聚糖酶的 cDNA 序列(SEQ ID NO :1)和推导的氨基酸序列(SEQ

ID NO:2) 示于图 1A 和 1B。cDNA 克隆编码 336 氨基酸的多肽。基因的 cDNA 克隆的 % G+C 含量是 67.5%，并且成熟蛋白质编码区域 (SEQ ID NO:1 的核苷酸 55 至 1011) 的 % G+C 含量是 67.5%。使用 SignalP 软件程序 (Nielsen et al., 1997, Protein Engineering 10: 1-6) 预测出 17 残基的信号肽。预测的成熟蛋白质含有 319 氨基酸具有 33.3 kDa 的分子量。

[0310] 通过 Clustal W 方法 (Higgins, 1989, 见上文) 测定内切葡聚糖酶家族 7 序列的比较比对 (comparative alignment), 所述方法使用 LASERGENE™MEGALIGN™ 软件 (DNASTAR, Inc., Madison, WI), 采用同一性表和下列多重比对参数: 缺口罚分 10 和缺口长度罚分 10。配对比对参数是 K 元组 (Ktuple) = 1、缺口罚分 = 3、窗口 = 5 和对角线 = 5。比对显示成熟土生梭孢霉 CEL7F 基因的推导的氨基酸序列与里氏木霉内切葡聚糖酶 I 基因 (NREF NF00494331, Uniprot Q5BMS5) 催化区的推导的氨基酸序列享有 46.7% 的同一性, 并且与全长里氏木霉内切葡聚糖酶 I 基因 (NREF NF00494331, Uniprot Q5BMS5) 的推导的氨基酸序列享有 44.9% 的同一性。对这些蛋白催化区的比对分析显示土生梭孢霉 CEL7F 内切葡聚糖酶缺少至少三个离散的序列基序, 所述序列基序在来自真菌的糖基水解酶家族 7 的所有其它已知成员中是保守的。这些序列基序中的两种由多于 10 个氨基酸残基组成, 并且含有高度保守的残基, 所述高度保守残基不存在于土生梭孢霉 CEL7F 内切葡聚糖酶中。

[0311] 一旦确定了克隆 Tter08C4 的性质, 将定名为 pTter7F (图 2) 的来自这种克隆的质粒 DNA 的 0.5 μl 等分试样转移到大肠杆菌 TOP10 细胞的小管 (Invitrogen, Carlsbad, CA) 中, 轻轻地混合, 并且在冰上温育 10 分钟。随后将细胞在 42°C 热休克 (heat-shock) 30 秒并且再次在冰上温育 2 分钟。将细胞在 250 μl SOC 培养基中重悬, 并且持续摇动 (200rpm) 在 37°C 温育 60 分钟。温育期之后, 将两个 30 μl 等分试样涂布于补充 50 μg 每 ml 卡那霉素的 LB 平板上, 并且在 37°C 温育过夜。次日挑取单菌落并且划线接种至三个 1.8ml 冷冻小管, 所述冷冻小管含有大约 1.5ml 补充 50 μg 每 ml 卡那霉素的 LB 琼脂糖。将小管用 PetriSeal™ (Diversified Biotech, Boston MA) 密封, 并作为 NRRL B-30802 保藏在农业研究机构专利培养物保藏中心 (Agricultural Research Service Patent Culture Collection), 北区研究中心 (Northern Regional Research Center), 1815 University Street, Peoria, Illinois, 61604, 保藏日期为 2005 年 4 月 11 日。

[0312] 实施例 5: pAlLo2 表达载体的构建

[0313] 通过修饰 pBAnE6 (美国专利号 6, 461, 837) 构建表达载体 pAlLo1, 所述 pBAnE6 包含来自黑曲霉中性 α-淀粉酶和米曲霉磷酸丙糖异构酶的基因启动子的杂合体 (NA2-tpi 启动子)、黑曲霉淀粉葡萄糖苷酶终止子序列 (AMG 终止子) 和构巢曲霉乙酰胺酶基因 (amdS)。使用所述的 Big-Dye™ 终止子化学通过测序验证所有诱变步骤。通过如下方法进行 pBAnE6 的修饰: 首先通过定位诱变从 amdS 选择标记消除在位置 2051、2722 和 3397bp 的三个 Nco I 限制位点。将所有改变设计为“沉默的”, 保留 amdS 基因产物的实际蛋白质序列不变。根据制造商的说明书采用 GeneEditor™ in vitro Site-Directed Mutagenesis Kit (Promega, Madison, WI), 使用如下引物 (加下划线的核苷酸代表改变的碱基) 同时进行这三个位点的去除:

[0314] AMDS3NcoMut (2050): 5' -GTGCCCCATGATACGCCTCCGG-3' (SEQ ID NO:7)

[0315] AMDS2NcoMut (2721): 5' -GAGTCGTATTTCCAAGGCTCCTGACC-3' (SEQ ID NO:8)

[0316] AMDS1NcoMut(3396):5'-GGAGGCCATGAAGTGGACCAACGG-3' (SEQ ID NO:9)

[0317] 随后使用 QuickChange™ Site-Directed Mutagenesis Kit(Stratagene, LaJolla, CA) 将包含所有三种期望的序列改变的质粒进行定位诱变,以消除在 AMG 终止子末端在位置 1643 的 Nco I 限制位点。下列引物(加下划线的核苷酸代表改变的碱基)用于诱变:

[0318] 诱变 AMG 终止子序列的 Upper Primer:

[0319] 5'-CACCGTGAAAGCCATGCTCTTTCCTTCGTGTAGAAGACCAGACAG-3' (SEQ ID NO:10)

[0320] 诱变 AMG 终止子序列的 Lower Primer:

[0321] 5'-CTGGTCTTCTACACGAAGGAAAGAGCATGGCTTTCACGGTGTCTG-3' (SEQ ID NO:11)

[0322] 修饰 pBAnE6 的最后步骤是使用 QuickChange™ Site-Directed Mutagenesis Kit(Stratagene, La Jolla, CA) 和以下引物(加下划线的核苷酸代表改变的碱基)在多接头的开始处添加新的 Nco I 限制位点以产生 pAlLo1(图 3)。

[0323] 诱变 NA2-tpi 启动子的 Upper Primer:

[0324] 5'-CTATATACACAACCTGGATTTACCATGGGCCCCGCGCCGCAGATC-3' (SEQ ID NO:12)

[0325] 诱变 NA2-tpi 启动子的 Lower Primer:

[0326] 5'-GATCTGCGGCCGCGGGCCCATGGTAAATCCAGTTGTGTATATAG-3' (SEQ ID NO:13)

[0327] 将 pAlLo1 的 amdS 基因与构巢曲霉 pyrG 基因交换 (swap)。将质粒 pBAnE10(图 4) 用作作为选择标记的 pyrG 基因的来源。pBAnE10 的序列分析显示 pyrG 标记包含在 Nsi I 限制片段之内,并且不含有 Nco I 或 Pac I 限制位点。因为 amdS 也与 Nsi I 限制位点在侧翼连接,转变选择标记的策略是简单的交换 Nsi I 限制片段。使用限制酶 Nsi I 消化来自 pAlLo1 和 pBAnE10 的质粒 DNA,并且通过琼脂糖凝胶电泳纯化产物。将含有 pyrG 基因的来自 pBAnE10 的 Nsi I 片段连接到 pAlLo1 的骨架中以取代原始含有 amdS 基因的 Nsi I DNA 片段。通过限制消化分析重组克隆以确定它们具有正确的插入物以及插入方向。选择具有以逆时针方向转录的 pyrG 基因的克隆。将新质粒定名为 pAlLo2(图 5)。

[0328] 实施例 6:将家族 CEL7F 内切葡聚糖酶基因克隆入米曲霉表达载体

[0329] 设计以下所示两种合成寡核苷酸引物用于 PCR 扩增来自土生梭孢霉 ESTTter08C4 的全长开读框,其编码家族 CEL7F 内切葡聚糖酶。使用 In-FusionCloning Kit(BD Biosciences, Palo Alto, CA) 将片段直接克隆到 pAlLo2 中。

[0330] In-Fusion Forward 引物:

[0331] 5'-ACTGGATTACCATGACCCTACGGCTCCCTGTCATCA-3' (SEQ ID NO:14)

[0332] In-Fusion Reverse 引物:

[0333] 5'-TCACCTCTAGTTAATTAAGTAGTTCTTCGTGGTAGACC-3' (SEQ ID NO:15)

[0334] 粗体字母表示编码序列。剩余序列与 pAlLo2 的插入位点相比包含序列同一性。

[0335] 将 50 皮摩尔的以上各种引物用在 PCR 反应中,所述反应在 50 μl 终体积中含有 50ng pTter11 C9 DNA、1X Pfx Amplification Buffer(Invitrogen, Carlsbad, CA)、6 μl 10mM dATP、dTTP、dGTP 和 dCTP 的混合物、2.5 单位 Platinum PfxDNA Polymerase(Invitrogen, Carlsbad, CA)、1 μl 50mM MgSO₄ 和 5 μl 10XpCRx Enhancer 溶液(Invitrogen, Carlsbad, CA)。使用 Eppendorf Mastercycler5333(Eppendorf Scientific, Inc., Westbury, NY) 扩增片段,将程序设定为在 98℃ 2 分钟的一个循环;和 94℃ 30 秒、

65°C 30 秒和 68°C 1.5 分钟的 35 个循环。在 35 个循环之后,将反应在 68°C 温育 10 分钟,并且随后在 10°C 冷却直到进一步处理。使用 40mM Tris 碱-20mM 乙酸钠-1mM EDTA 二钠 (TAE) 缓冲液和每 ml 0.1 μ g 溴化乙锭在 0.8% GTG 琼脂糖凝胶 (Cambrex Bioproducts OneMeadowlands Plaza East Rutherford, New Jersey 07073) 上分离 1.4kb 的 PCR 反应产物。在 Dark Reader™ (Clare Chemical Research, Dolores, CO) 的辅助下显现 DNA 条带以避免 UV 诱导的突变。使用一次性剃刀刀片 (razor blade) 切除 1.4kb DNA 条带并根据制造商的说明书使用 Ultrafree-DA 转杯 (spin cup) (Millipore, Billerica, MA) 将其纯化。

[0336] 通过用 NcoI 和 PacI 的消化将载体 pAlLo2 线性化。将片段通过如上所述的凝胶电泳和超滤纯化。将纯化的 PCR 片段克隆到线性化并且纯化的 pAlLo2 载体中,使用 In-Fusion Cloning Kit (BD Biosciences, Palo Alto, CA) 进行所述克隆。反应物 (20 μ l) 包含 1X In-Fusion Buffer (BD Biosciences, Palo Alto, CA)、1X BSA (BD Biosciences, Palo Alto, CA)、1 μ l In-Fusion 酶 (1 : 10 稀释的) (BD Biosciences, Palo Alto, CA)、100ng 用 NcoI 和 PacI 消化的 pAlLo2 和 50ng 土生梭孢霉 CEL7F 的纯化 PCR 产物。将反应在室温温育 30 分钟。将反应的 2 μ l 样品用于根据制造商的说明书转化大肠杆菌 XL10 SoloPac® Gold 细胞 (Stratagene, La Jolla, CA)。在恢复期 (recovery period) 之后,将来自转化反应的两个 100 μ l 等分试样涂布到补充 100 μ g 每 ml 的氨苄青霉素的 150mm 2X YT 平板上。将平板在 37°C 温育过夜。从选择平板中随机选择一组八个推定的重组克隆,并且使用 BioRobot 9600 从每个制备质粒 DNA。将克隆通过 XhoI 限制消化分析。随后将具有期望的限制消化模式 (restriction digest pattern) 的两个克隆测序以确认在克隆的插入物中不存在突变。选择克隆 #1 并定名为 pAlLo22 (图 6)。

[0337] 实施例 7 :土生梭孢霉家族 CEL7F 内切葡聚糖酶基因在米曲霉 JAL250 中的表达

[0338] 根据 Christensen et al., 1988, Bio/Technology 6 :1419-1422 的方法制备米曲霉 JAL250 (WO 99/61651) 原生质体。将 5 微克 pAlLo22 (以及作为载体对照的 pAlLo2) 用于转化米曲霉 JAL250 原生质体。

[0339] 使用 pAlLo2 转化米曲霉 JAL250 产生大约 50 个转化体。将 8 个转化体分离到单独的 PDA 平板并且在 34°C 温育 5 天。

[0340] 用 5ml 0.01% Tween80 洗涤汇合孢子平板 (confluent spore plate), 并将孢子悬浮液用于接种 125ml 玻璃摇瓶中的 25ml MDU2BP 培养基。将转化体培养物以 200rpm 的持续摇动在 34°C 温育。在接种后的第五天,将培养物在 6000xg 离心并且收集它们的上清液。将 5 μ l 的每种上清液与等体积的 2X 上样缓冲液 (10% β -巯基乙醇) 混合,上样到 1.5mm 8% -16% Tris-Glycine SDS-PAGE 凝胶上并用 Simply Blue SafeStain (Invitrogen, Carlsbad, CA) 染色。培养液的 SDS-PAGE 分布显示八个转化体中的六个具有大约 40kDa 的新蛋白质条带。选择转化体 7 号用于进一步研究并且将其指定为米曲霉 JAL250AlLo22。

[0341] 实施例 8 :米曲霉 JAL250AlLo22 的大摇瓶培养

[0342] 将米曲霉 JAL250AlLo22 孢子涂布到 PDA 平板上并且在 34°C 温育五天。用 5ml 0.01% Tween80 将汇合孢子平板洗涤两次以使收集的孢子数最大化。随后将孢子悬浮液用于接种 2 升 Fembach 烧瓶中的 500ml MDU2BP 培养基。将转化体培养物以持续摇动 (200rpm) 在 34°C 温育。在接种后的第五天,通过在孔大小 0.45 μ m 的 500ml, 75mm Nylon 过滤装置上过滤来收集培养液,所述过滤装置具有玻璃纤维预过滤器。对培养液的 5 μ l 样

品如上所述通过 SDS-PAGE 进行分析,以确认蛋白质分布与之前获得的相同。一旦培养液显示含有 40kDa 蛋白质条带,则对该培养液进行酶表征。

[0343] 实施例 9 :土生梭孢霉 CEL7F 内切葡聚糖酶的表征

[0344] 将实施例 8 中所述的米曲霉 J1250A1Lo22 培养液通过 0.22 μm 孔大小的滤器 (Millipore, Billerica, MA) 过滤,使用配备 PM10 膜的 Amicon 搅拌杯 (stirredcell) (Millipore, Billerica, MA) 进行浓缩,并且使用 Econo-Pac 10DG 柱 (BioRadLaboratories, Hercules, CA) 脱盐。

[0345] 将来自米曲霉 J1250 (单独的载体) 的培养液作为阴性对照进行与上述相同的处理。

[0346] 用于评估土生梭孢霉 CEL7F 内切葡聚糖酶底物特异性的经染色的底物包括: AZCL- 阿拉伯木聚糖 (小麦)、AZCL- β - 葡聚糖、AZCL- 葡聚糖、AZCL-HE- 纤维素、AZCL- 马铃薯半乳聚糖、AZCL- 半乳甘露聚糖 (Carob)、AZCL- 木聚糖 (Birchwood)、AZCL- 木葡聚糖 (Megazyme, Bray, Ireland) 和 Chitin Azure (Sigma, St Louis, MO)。

[0347] 在 96 深孔平板 (Axygen Scientific, Union City, CA) 中进行活性试验,所述平板由平板密封物 (ALPS-300, Abgene, Epsom, UK) 密封。将 800 μl 上述底物 (6.25g 每升 50mM 乙酸钠 pH5.0) 转移到 96 深孔平板的每个孔中,接着是 180 μl 50mM 乙酸钠 pH5.0 和 20 μl 土生梭孢霉 CEL7F 内切葡聚糖酶溶液 (0.25g/L) 以开始反应。最终反应混合物中的底物浓度和酶上样量分别是 5 克每升和 1mg 酶每克底物。将米曲霉 J1250 培养液作为阴性对照,与土生梭孢霉 CEL7F 内切葡聚糖酶培养液在相同条件下一起进行测试。将反应在 50°C 不进行混合地温育。取样前,将深孔平板在平板离心机 (Sorvall RT7, Global Medical Instrumentation, Ramsey, MN) 上以 3000rpm 离心 5 分钟。将 150 μl 上清液的样品转移到 96 孔过滤平板 (0.45 μm 孔大小, Millipore, Billerica, MA) 中,抽真空 (vacuum) 并且收集滤出液。将 100 μl 滤出液的样品转移到另外的 96 孔平板中,并且使用 Spectra MAX340 (Molecular Devices, Sunnyvale, CA) 测量 590nm 的吸光度。

[0348] 在 1 小时和 92 小时的温育之后,由土生梭孢霉 CEL7F 内切葡聚糖酶从不同经染色底物释放的染料作为相对的 590nm 数值示于表 1 (减去由米曲霉 J1250 释放的染料之后)。

[0349] 表 1. 土生梭孢霉内切葡聚糖酶与不同经染色底物一起在 50°C、pH5.0 温育 1 小时和 92 小时之后的相对 $A_{590\text{nm}}$

[0350]

时间 小时	AX	βG	Dex	HEC	Gal	GM	Xly	XG	Chitin Azure
1	0.00	0.22	0.00	0.16	0.00	0.00	0.01	0.01	0.00
92	0.58	0.78	0.00	1.00	0.00	0.08	0.32	0.75	0.00

[0351] AX :AZCL- 阿拉伯木聚糖 βG :AZCL- β - 葡聚糖

[0352] Dex :AZCL- 葡聚糖 HEC :AZCL-HE- 纤维素

[0353] Gal :AZCL- 马铃薯半乳聚糖 GM :AZCL- 半乳甘露聚糖

[0354] Xly :AZCL- 木聚糖 XG :AZCL- 木葡聚糖

[0355] 土生梭孢霉 CEL7F 内切葡聚糖酶在 1 小时后对 AZCL- β - 葡聚糖和 AZCL-HE- 纤维

素具有活性。92 小时温育之后,土生梭孢霉 CEL7F 内切葡聚糖酶还显示出对阿拉伯木聚糖、木聚糖和木葡聚糖染色底物的活性,和对半乳甘露聚糖染色底物的低活性。

[0356] 实施例 10 :用土生梭孢霉 CEL7F 内切葡聚糖酶水解预处理的玉米秸秆

[0357] 使用稀释的硫酸在 U.S.Department of Energy National Renewable EnergyLaboratory (NREL) 预处理玉米秸秆。将以下条件用于所述预处理:于 190℃ 0.048 克硫酸 / 克干的生物质和 25% w/w 干固体,持续大约 1 分钟。预处理的玉米秸秆 (PCS) 中的水不溶性固体含有 52% 纤维素、3.6% 半纤维素和 29.8% 木质素。通过两阶段硫酸水解和后续通过高效液相色谱的糖分析,使用 NRELStandard Analytical Procedure#002 测定纤维素和半纤维素。在以硫酸水解纤维素和半纤维素级分之后,使用 NREL Standard Analytical Procedure#003 以重力分析法测定木质素。在酶水解之前,将 PCS 用大体积的重蒸馏水洗涤直至 pH 高于 4.0,并且随后通过 100 目筛网筛滤,在 121℃ 高压灭菌 30 分钟。

[0358] PCS 的水解在 96 深孔平板中进行 (Axygen Scientific, Union City, CA),将所述平板用平板密封物 (ALPS-300, Abgene, Epsom, UK) 密封,总反应体积为 1.0ml。PCS (10mg/ml, 于 50mM 乙酸钠 pH5.0 缓冲液中) 的水解使用 1.25mg 土生梭孢霉 CEL7F 内切葡聚糖酶 (如实施例 9 中所述制备) 每克 PCS 来进行。将来自米曲霉 J1250 的培养液 (如实施例 9 中所述制备) 作为对照操作。PCS 水解在 50℃、pH5.0 进行。重复进行反应,并且在水解期间取得等分试样。通过将每种水解产物的 20 μl 等分试样与 180 μl 10.11M NaOH (终止试剂) 混合来停止 PCS 水解反应。对于每种样品产生适当的系列稀释物,并且使用如下所述适合 96 孔微量培养板形式的对羟基苯甲酸酰肼 (para-hydroxybenzoic acidhydrazide) (PHBAH, Sigma, St. Louis, MO) 试验测定还原糖含量。简而言之,将适当稀释的样品的 90 μl 等分试样置于 96 孔锥底微量培养板 (conicalbottomed microplate) 中。通过向每个孔中添加 60 μl 12% NaOH 中的 1.5% (w/v) PHBAH 来起始反应。将平板无覆盖地在 95℃ 加热 10 分钟。使平板冷却至室温 (RT),并且向每个孔添加 50 μl 蒸馏 H₂O。将来自每个孔的 100 μl 等分试样转移至平底 96 孔平板中,并且使用 SpectraMax Microplate Reader (MolecularDevices, Sunnyvale, CA) 测量 A_{410nm} 的吸光度。使用葡萄糖标准 (0.1-0.0125mg/ml, 用 0.4% 氢氧化钠稀释) 制备标准曲线,以将所得 A_{410nm} 数值转化成葡萄糖当量。将所得当量用于计算每个反应的 PCS 纤维素转化百分比。使用以下等式计算纤维素转化成还原糖的程度 (转化率, %):

[0359]
$$\text{转化率} (\%) = \text{RS}_{(\text{mg/ml})} * 100 * 162 / (\text{纤维素}_{(\text{mg/ml})} * 180)$$

[0360]
$$= \text{RS}_{(\text{mg/ml})} * 100 / (\text{纤维素}_{(\text{mg/ml})} * 1.111)$$

[0361] 在这个等式中, RS 是以葡萄糖当量 (mg/ml) 测量的溶液中还原糖的浓度,而因子 1.111 反映出将纤维素转化成葡萄糖的重量增加。

[0362] 土生梭孢霉 CEL7F 内切葡聚糖酶 (1.25mg/g PCS) 的 PCS 水解在 120 小时之后得到 2.1% 的纤维素转化率。米曲霉 J1250 (1.25mg/g PCS) 在 120 小时之后得到少于 1% 的转化率。

[0363] 实施例 11 :土生梭孢霉内切葡聚糖酶对来自大麦的可溶性 β-葡聚糖的水解

[0364] 将土生梭孢霉 Cel7F 内切葡聚糖酶以实施例 8 中所述的米曲霉 J1250A1Lo22 培养液的形式进行测试。将培养液浓缩,并且使用来自 Millipore (Bedford, MA) 的带有 Biomax-5 聚醚砜膜 (5000 NMWL) 的 Centricon Plus-20 离心滤器将培养液交换成 50mM 乙

酸钠 pH5.0。将来自米曲霉 J1250 的培养液同上处理。

[0365] 使用 Bicinchoninic Acid(BCA) 微量培养板试验, 依照 BCA Protein Assay Reagent Kit(Pierce Chemical Co., Rockford, IL) 的制造商说明书测定酶溶液中的蛋白质浓度。

[0366] 在各反应之前从贮藏在 -20℃ 的酶储液制备新鲜的酶稀释物。

[0367] 土生梭孢霉 Ce17F 内切葡聚糖酶对来自大麦的可溶性 β -葡聚糖(中等粘度, 230kDa, Megazyme International Ireland Ltd., Bray, Ireland) 的活性在 pH5.5(具有 0.02% 叠氮化钠的 50mM 乙酸钠) 和 60℃ 测定。将结果与里氏木霉 Ce17B(EGI) 内切葡聚糖酶的结果进行比较。重组里氏木霉 Ce17B(EGI) 内切葡聚糖酶能够依照 Takashima et al., 1998, Journal of Biotechnology 65:163-171 制备。

[0368] 水解反应中 β -葡聚糖的初始浓度是 1.0% (w/v)。将 1ml 反应在 Eppendorf96 DeepWell Plates(1.2ml, VWR Scientific, West Chester, PA) 中无搅拌地进行。将酶以三种蛋白质加载量 (loading) 使用: 0.05、0.1 和 0.2mg 每克葡聚糖。在对照试验中, 将内切葡聚糖酶用 50mM 含有 0.02% 叠氮化钠的乙酸钠 pH5.5 取代(缓冲液对照), 或用不含重组表达酶的经过浓缩并且以缓冲液交换的米曲霉 J1250 培养液取代(J1250 对照)。

[0369] 在 2 小时和 24 小时从水解反应中移出等分试样, 用去离子水稀释, 并且使用对羟基苯甲酸酰肼 (PHBAH) 试验如实施例 10 中所述分析还原糖。将 2 小时和 24 小时两个温育时间的 β -葡聚糖相对转化率作为蛋白质加载量的函数分别示于图 7 和 8。将所述相对转化率显示为在以土生梭孢霉 Ce17F 内切葡聚糖酶 (0.2mg 蛋白质每克葡聚糖) 水解 β -葡聚糖 24 小时之后获得的转化率的百分比。

[0370] 土生梭孢霉 Ce17F 内切葡聚糖酶与里氏木霉 Ce17B 内切葡聚糖酶相比显示较高的 β -葡聚糖转化率, 并且超过 2 小时温育时间之后继续产生新的还原端基团。与之相反, 里氏木霉 Ce17B 内切葡聚糖酶显示 2 小时水解之后在还原糖浓度上几乎没有额外的增加。

[0371] 生物材料的保藏

[0372] 依据布达佩斯条约的条款, 下述的生物材料已经保藏于农业研究机构专利培养物保藏中心 (Agricultural Research Service Patent Culture Collection), 北方区研究中心 (Northern Regional Research Center), 1815 University Street, Peoria, Illinois, 61604, 给出了下述的登录号:

[0373] 保藏物 登录号 保藏日期

[0374] 大肠杆菌 pTter7F NRRL B-30837 2005 年 4 月 11 日

[0375] 该菌株于下述条件下保藏: 确保在本专利申请未决期间, 由专利与商标委员依据 37C.F.R. § 1.14 和 35U.S.C. § 122 授权的人能够获得该培养物。该保藏物为所保藏菌株的基本上纯的培养物。在提交了该申请的副本, 或其后续文本的国家, 依据该外国专利法律的要求, 可以获得该保藏物。然而, 应当理解, 保藏物的获得并不构成对实施本发明的许可, 实施本发明是对政府行为所授予的专利权的侵犯。

[0376] 此处, 本发明中所描述的和要求的并非是要用本文所公开的具体方面来限定范围, 因为这些方面意欲作为本发明几个方面的说明。任何等价的方面意欲在本发明的范围之内。事实上, 从前面的说明中, 对于本领域技术人员来说, 除本文所显示和描述的之外, 本发明的多种修改是显而易见的。这些修改也意欲落入所附的权利要求的范围之内。在冲突

的情况下,将以包括定义部分的本公开为准。

[0377] 本文引用了许多参考文献,其中公开的全部内容并入作为参考。

[0378]

申请人或代理人档案号 10802.204-WO	国际申请号 待给出
-------------------------	-----------

关于微生物保藏的说明

(细则 13 之二)

A. 对说明书第 61 页, 第 18 行所述的微生物的说明。	
B. 保藏事项 其它保藏在补充页中 ☐	
保藏单位名称 农业研究机构专利培养物保藏中心(NRRL) (Agricultural Research Service Patent Culture Collection (NRRL))	
保藏单位地址(包括邮政编码和国名) 北方区研究中心(Northern Regional Research Center) 1815 University Street Peoria, IL, 61604, US	
保藏日期 2005 年 4 月 11 日	保藏编号 B-30837
C. 附加说明 (必要时) 本栏内容有补充页 ☐	
关于寻求欧洲专利和/或澳大利亚专利的这些指定, 在本专利申请的悬而未决期间, 所保藏的微生物样品将仅提供于由请求保藏此样品的人提名的独立专家(Rule 28(4) EPC/澳大利亚法规 1991 No. 71 的实施细则 3.25)。	
D. 本说明是为下列指定国作的 (如果说明不是为所有指定国而作的)	
E. 补充说明 (必要时)	
下列说明将随后向国际局提供 (写出说明的基本性质, 例如: “保藏物的登录号”)	

由受理局填写

☐ 本页已经和国际申请一起收到
授权的官员

由国际局填写

☐ 国际局收到本页日期:
授权的官员

PCT/RO/134 表(1992 年 7 月)

[0001]

序列表

<110> 诺维信股份有限公司(Novozymes, Inc.)

<120> 具有内切葡聚糖酶活性的多肽和编码该多肽的多核苷酸

<130> 10802.204-W0

<150> 60/675,601

<151> 2005-04-27

<160> 15

<170> PatentIn version 3.2

<210> 1

<211> 1011

<212> DNA

<213> 土生梭孢霉(*Thielavia terrestris*)

<400> 1

```

atgacctac ggctccctgt catcagcctg ctggcctcgc tggcagcagg cgccgtcgtc      60
gtccacaggc cggagtttca cccccctctc ccgacttgga aatgcacgac ctccgggggc      120
tgctgtcagc agaacaccag cgtcgtcctg gaccgtgact cgaagtacgc cgcacacagc      180
gccggctcgc ggacggaatc ggattacgcg gcaatgggag tgtccacttc gggcaatgcc      240
gtgacgctgt accactacgt caagaccaac ggcacctcgc tcccccttc gccgcgcac      300
tacctccttg gcgcggacgg caagtacgtg cttatggacc tcctcaacca ggagctgtcg      360
gtggacgtcg acttctcggc gctgccgtgc ggcgagaacg gggccttcta cctgtccgag      420
atggcggcgg acggcggggg cgacgcgggg gcggcgacgc ggtactgcga cgcgcagtgc      480
cagggtact gctgcaacga gatggacatc ctgaggcca actcgatggc gacggccatg      540
acgcgcgacc cgtgcaaggc caacaactgc gaccgcagcg gctgcggcta caaccgtac      600
gccagcggcc agecgggctt ctacggggcc ggcaagacgg tcgacacgag caagcccttc      660
accgtcgtca cgcagttcgc gccagcggc ggcaagciga ccagatcac ccgcaagtac      720
atccagaacg gccgggagat cggcggcggc ggcacatct ccagctgcgg ctccgagttt      780
tcgacgggcg gcctgaccgg catggcgag gcgctggggc gcggaatggt gctggccatg      840
agcatcttga acgacgcggc ccaggagatg gcatggctcg atccggcaa caacggccct      900
tgcgccagt gccaggcag cccgtccgtc attcagtcgc agcatccga caccacgtc      960
gttttctcca acatcagtg gggcgacatc gggtctacca cgaagaacta g      1011

```

<210> 2

<211> 336

<212> PRT

[0002]

<213> 土生梭孢霉(Thielavia terrestris)

<400> 2

```

Met Thr Leu Arg Leu Pro Val Ile Ser Leu Leu Ala Ser Leu Ala Ala
1           5           10           15

Gly Ala Val Val Val Pro Arg Ala Glu Phe His Pro Pro Leu Pro Thr
          20           25           30

Trp Lys Cys Thr Thr Ser Gly Gly Cys Val Gln Gln Asn Thr Ser Val
          35           40           45

Val Leu Asp Arg Asp Ser Lys Tyr Ala Ala His Ser Ala Gly Ser Arg
          50           55           60

Thr Glu Ser Asp Tyr Ala Ala Met Gly Val Ser Thr Ser Gly Asn Ala
65           70           75           80

Val Thr Leu Tyr His Tyr Val Lys Thr Asn Gly Thr Leu Val Pro Ala
          85           90           95

Ser Pro Arg Ile Tyr Leu Leu Gly Ala Asp Gly Lys Tyr Val Leu Met
          100          105          110

Asp Leu Leu Asn Gln Glu Leu Ser Val Asp Val Asp Phe Ser Ala Leu
          115          120          125

Pro Cys Gly Glu Asn Gly Ala Phe Tyr Leu Ser Glu Met Ala Ala Asp
          130          135          140

Gly Arg Gly Asp Ala Gly Ala Gly Asp Gly Tyr Cys Asp Ala Gln Cys
145          150          155          160

Gln Gly Tyr Cys Cys Asn Glu Met Asp Ile Leu Glu Ala Asn Ser Met
          165          170          175

Ala Thr Ala Met Thr Pro His Pro Cys Lys Gly Asn Asn Cys Asp Arg
          180          185          190

Ser Gly Cys Gly Tyr Asn Pro Tyr Ala Ser Gly Gln Arg Gly Phe Tyr
          195          200          205

Gly Pro Gly Lys Thr Val Asp Thr Ser Lys Pro Phe Thr Val Val Thr
          210          215          220

Gln Phe Ala Ala Ser Gly Gly Lys Leu Thr Gln Ile Thr Arg Lys Tyr
225          230          235          240

Ile Gln Asn Gly Arg Glu Ile Gly Gly Gly Thr Ile Ser Ser Cys
          245          250          255

Gly Ser Glu Ser Ser Thr Gly Gly Leu Thr Gly Met Gly Glu Ala Leu
          260          265          270

Gly Arg Gly Met Val Leu Ala Met Ser Ile Trp Asn Asp Ala Ala Gln
          275          280          285

```

[0003]

Glu Met Ala Trp Leu Asp Ala Gly Asn Asn Gly Pro Cys Ala Ser Gly
290 295 300

Gln Gly Ser Pro Ser Val Ile Gln Ser Gln His Pro Asp Thr His Val
305 310 315 320

Val Phe Ser Asn Ile Arg Trp Gly Asp Ile Gly Ser Thr Thr Lys Asn
325 330 335

<210> 3

<211> 32

<212> DNA

<213> 大肠杆菌(Escherichia coli)

<400> 3

tcgtcgggga caactttgta caaaaaagtt gg

32

<210> 4

<211> 27

<212> DNA

<213> 大肠杆菌(Escherichia coli)

<400> 4

ccccgttgga aacatgtttt ttcaacc

27

<210> 5

<211> 16

<212> DNA

<213> 大肠杆菌(Escherichia coli)

<400> 5

gtaaacgac ggccag

16

<210> 6

<211> 25

<212> DNA

<213> 大肠杆菌(Escherichia coli)

<220>

<221> misc_feature

<222> (25)..(25)

<223> N = A, C, G 或 T

<400> 6

tttttttttt tttttttttt tttvn

25

<210> 7

<211> 22

<212> DNA

<213> 构巢曲霉(Aspergillus nidulans)

<400> 7

gtgccccatg atacgcctcc gg

22

<210> 8

<211> 26

<212> DNA

<213> 构巢曲霉(Aspergillus nidulans)

[0004]

<400> 8	
gagtcgtatt tccaagctc ctgacc	26
<210> 9	
<211> 24	
<212> DNA	
<213> 构巢曲霉(<i>Aspergillus nidulans</i>)	
<400> 9	
ggaggccatg aagtggacca acgg	24
<210> 10	
<211> 45	
<212> DNA	
<213> 黑曲霉(<i>Aspergillus niger</i>)	
<400> 10	
caccgtgaaa gccatgctct ttccttcgtg tagaagacca gacag	45
<210> 11	
<211> 45	
<212> DNA	
<213> 黑曲霉(<i>Aspergillus niger</i>)	
<400> 11	
ctggtcttct acacgaagga aagagcatgg ctttcacggt gtctg	45
<210> 12	
<211> 44	
<212> DNA	
<213> 黑曲霉(<i>Aspergillus niger</i>)	
<400> 12	
ctatatacac aactggattt accatgggcc cgcggccgca gatc	44
<210> 13	
<211> 44	
<212> DNA	
<213> 黑曲霉(<i>Aspergillus niger</i>)	
<400> 13	
gatctgcggc cgcgggccca tggtaaatac agttgtgtat atag	44
<210> 14	
<211> 36	
<212> DNA	
<213> 土生梭孢霉(<i>Thielavia terrestris</i>)	
<400> 14	
actggattac catgacccta cggtccctg tcatca	36
<210> 15	
<211> 38	
<212> DNA	
<213> 土生梭孢霉(<i>Thielavia terrestris</i>)	
<400> 15	
tcacctctag ttaattaact agttcttcgt ggtagacc	38

M T L R L P V I S L L A S L A A G
1 ATGACCCTACGGCTCCCTGTCATCAGCCTGCTGGCCTCGCTGGCAGCAGGC
A V V V P R A E F H P P L P T W K
52 GCCGTCGTCGTCCCACGGGCGGAGTTTCACCCCCCTCTCCCGACTTGGA
C T T S G G C V Q Q N T S V V L D
103 TGCACGACCTCCGGGGGCTGCGTGCAGCAGAACACCAGCGTCGTCTCTGGAC
R D S K Y A A H S A G S R T E S D
154 CGTGA CT CGAAGTACGCCGCACACAGCGCCGGCTCGCGGACGGAATCGGAT
Y A A M G V S T S G N A V T L Y H
205 TACGCGGCAATGGGAGTGTCCACTTCGGGCAATGCCGTGACGCTGTACCAC
Y V K T N G T L V P A S P R I Y L
256 TACGTCAAGACCAACGGCACCCTCGTCCCCGCTTCGCCGCGCATCTACCTC
L G A D G K Y V L M D L L N Q E L
307 CTGGGCGCGGACGGCAAGTACGTGCTTATGGACCTCCTCAACCAGGAGCTG
S V D V D F S A L P C G E N G A F
358 TCGGTGGACGTGACTTCTCGGCGCTGCCGTGCGGCGAGAACGGGGCCTTC
Y L S E M A A D G R G D A G A G D
409 TACCTGTCCGAGATGGCGGCGGACGGGCGGGCGACGCGGGGGCGGGCGAC
G Y C D A Q C Q G Y C C N E M D I
460 GGGTACTGCGACGCGCAGTGCCAGGGCTACTGCTGCAACGAGATGGACATC
L E A N S M A T A M T P H P C K G
511 CTCGAGGCCAACTCGATGGCGACGGCCATGACGCCGCACCCGTGCAAGGGC
N N C D R S G C G Y N P Y A S G Q
562 AACAACTGCGACCGCAGCGGCTGCGGCTACAACCCGTACGCCAGCGGCCAG
R G F Y G P G K T V D T S K P F T
613 CGCGGCTTCTACGGGCCCCGGAAGACGGTCGACACGAGCAAGCCCTTCACC
V V T Q F A A S G G K L T Q I T R
664 GTCGTACGCAGTTCGCCGCCAGCGGCGGCAAGCTGACCCAGATCACCCGC
K Y I Q N G R E I G G G G T I S S
715 AAGTACATCCAGAACGGCCGGGAGATCGGCGGCGGCGGCACCATCTCCAGC
C G S E S S T G G L T G M G E A L
766 TGCGGCTCCGAGTCTTCGACGGGCGGCCTGACCGGCATGGGCGAGGCGCTG
G R G M V L A M S I W N D A A Q E
817 GGGCGCGGAATGGTGCTGGCCATGAGCATCTGGAACGACGCGGCCCAGGAG
M A W L D A G N N G P C A S G Q G

图 1A

```
868 ATGGCATGGCTCGATGCCGGCAACAACGGCCCTTGCGCCAGTGGCCAGGGC
    S P S V I Q S Q H P D T H V V F S
919 AGCCCGTCCGTCATTCAGTCGCAGCATCCCGACACCCACGTCGTCTTCTCC
    N I R W G D I G S T T K N *
970 AACATCAGGTGGGGCGACATCGGGTCTACCACGAAGAACTAG
```

图 1B

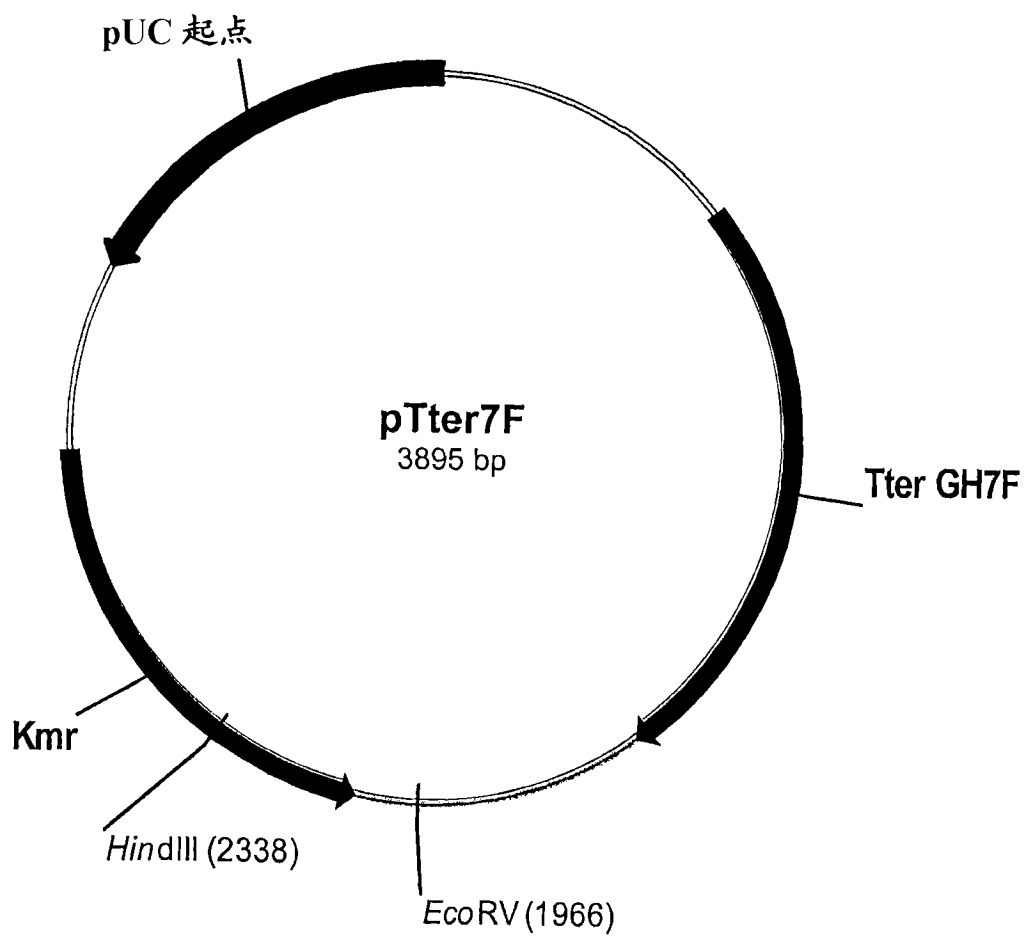


图 2

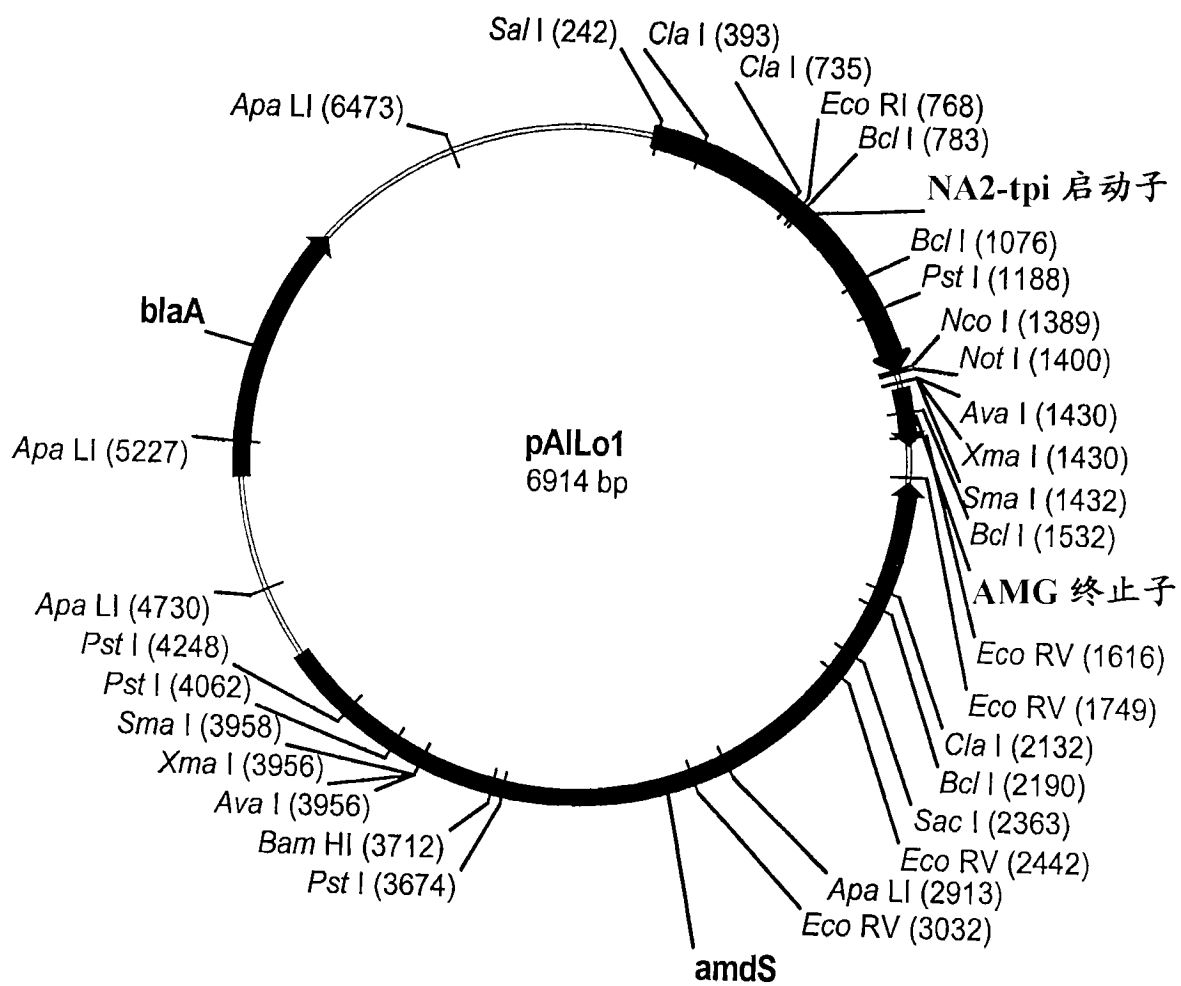


图 3

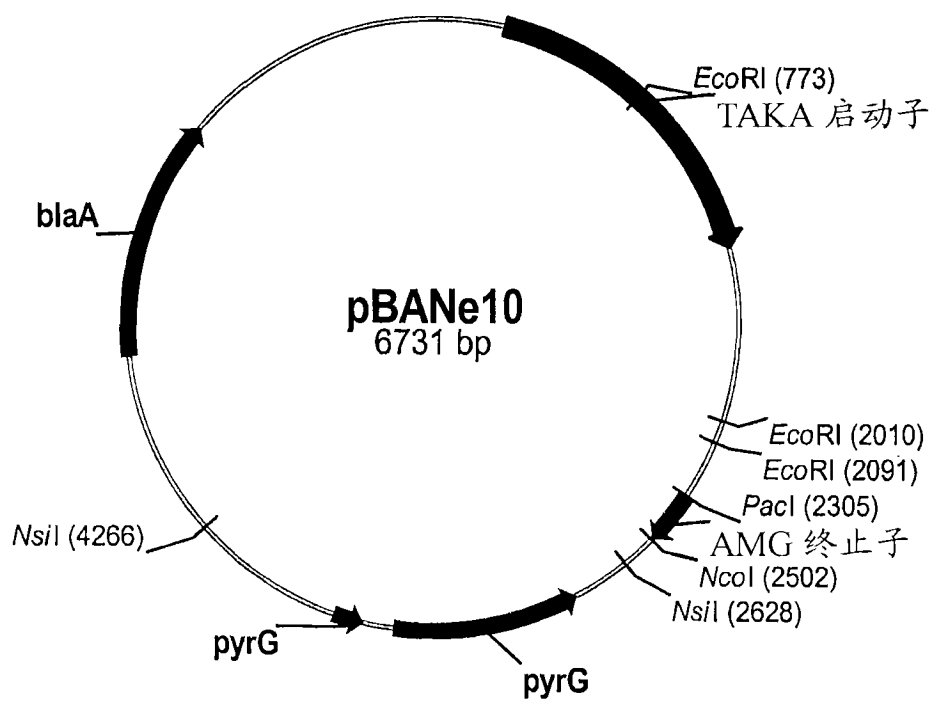


图 4

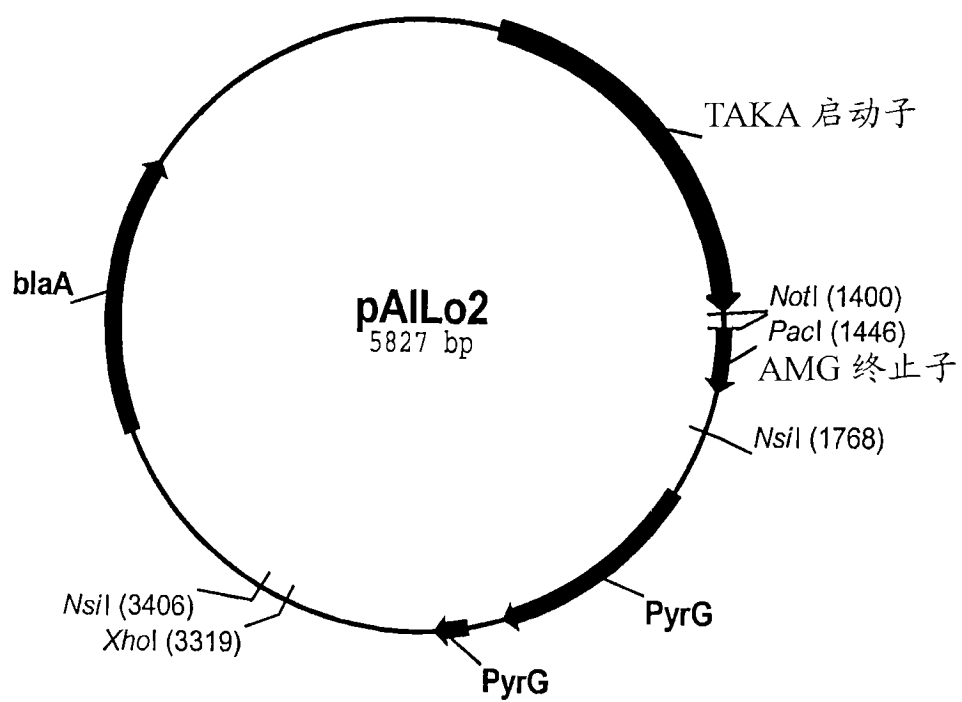


图 5

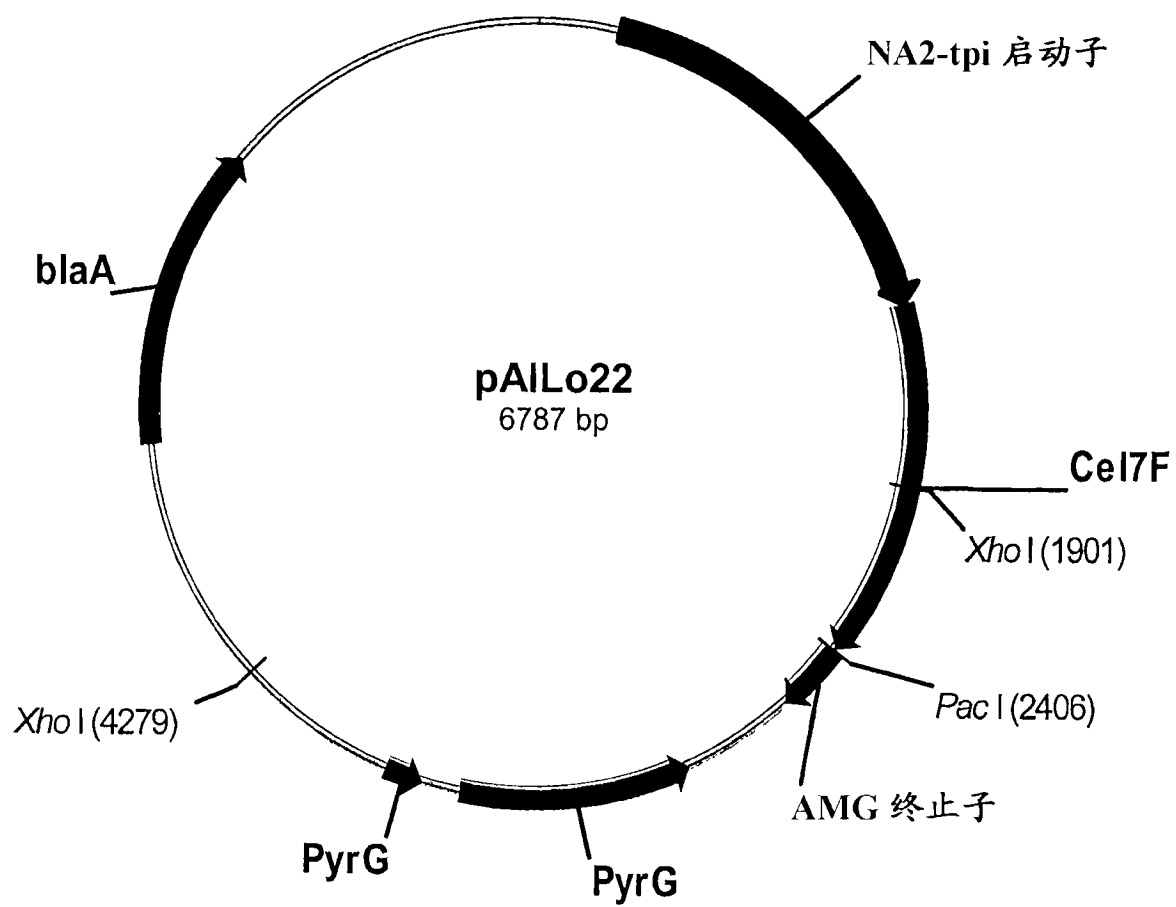


图 6

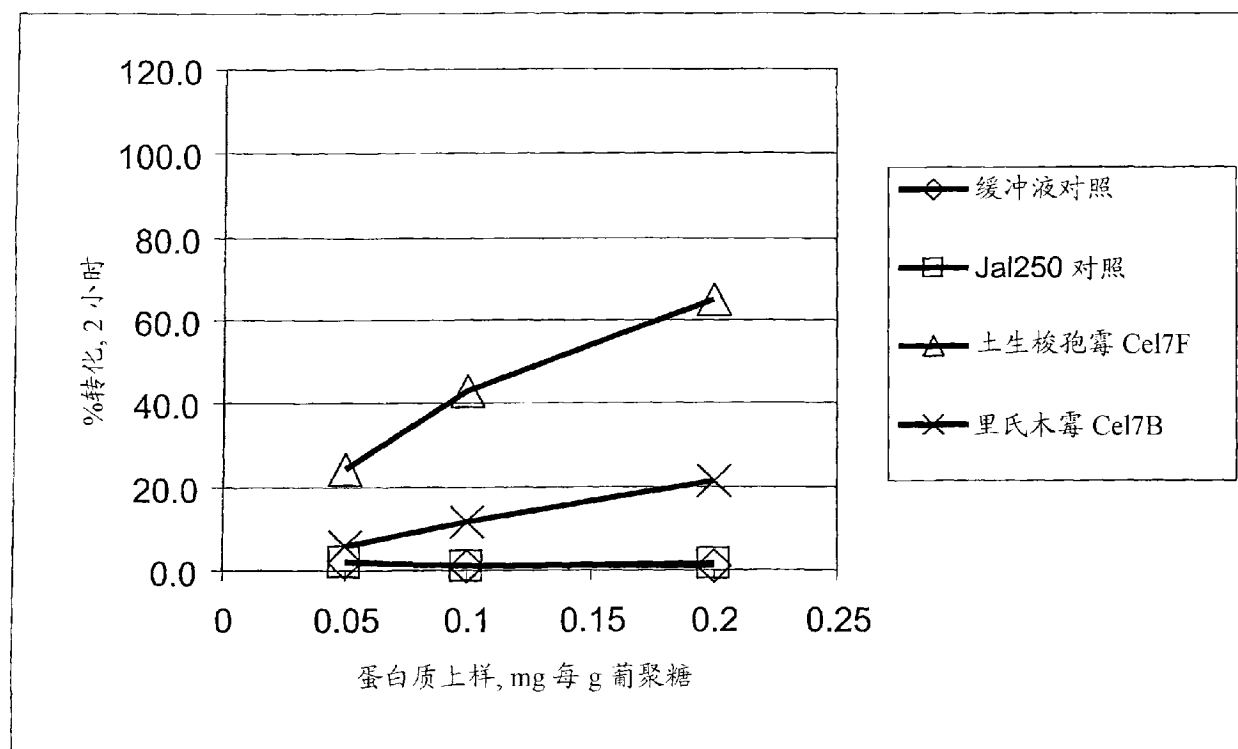


图 7

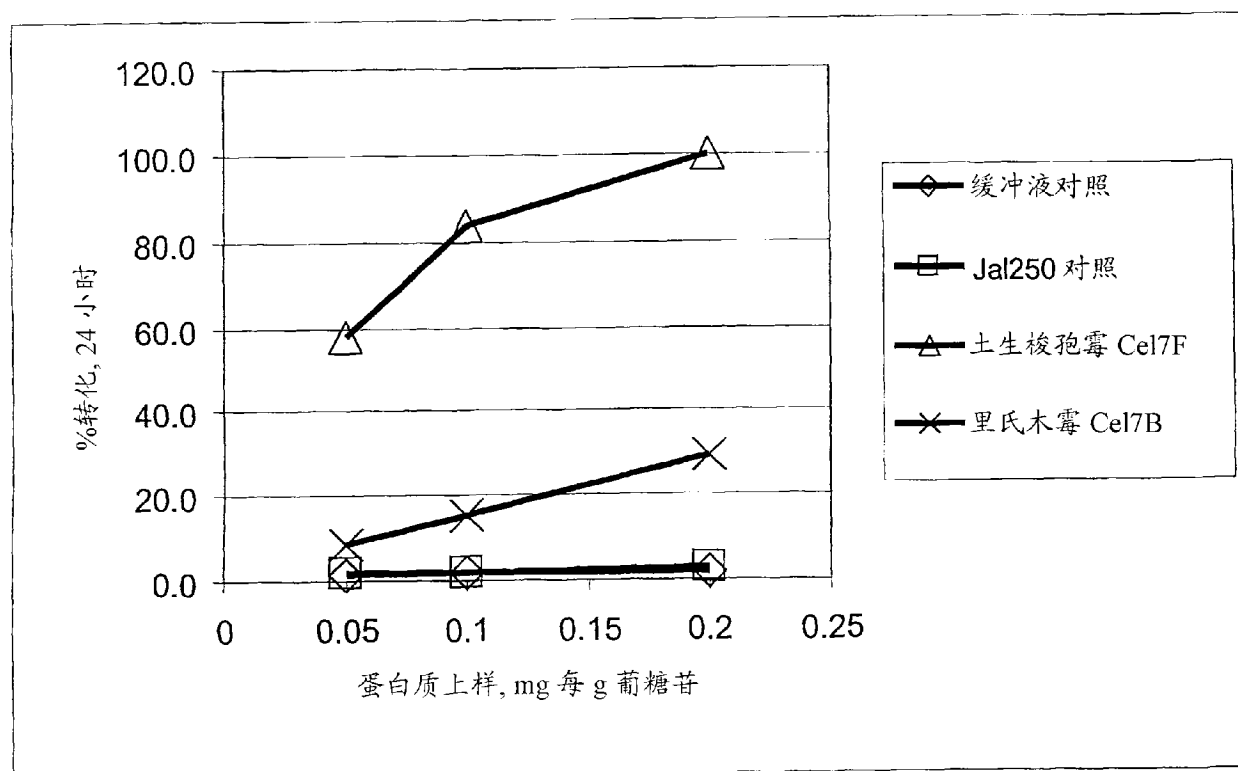


图 8