



(12) 发明专利

(10) 授权公告号 CN 102597243 B

(45) 授权公告日 2015. 05. 20

(21) 申请号 201080039731. 7

(22) 申请日 2010. 07. 07

(30) 优先权数据

61/223, 533 2009. 07. 07 US

(85) PCT国际申请进入国家阶段日

2012. 03. 07

(86) PCT国际申请的申请数据

PCT/US2010/041222 2010. 07. 07

(87) PCT国际申请的公布数据

W02011/005867 EN 2011. 01. 13

(83) 生物保藏信息

DSM 22711 2009. 06. 24

(73) 专利权人 诺维信股份有限公司

地址 美国加利福尼亚州

专利权人 诺维信公司

(72) 发明人 段俊欣 汤岚 刘晔 吴文平

J. 奎因兰 R. 克拉默

(74) 专利代理机构 北京市嘉元知识产权代理事

务所(特殊普通合伙) 11484

代理人 张永新

(51) Int. Cl.

C12N 15/82(2006. 01)

C12N 9/24(2006. 01)

(56) 对比文件

CN 1965078 A, 2007. 05. 16, 全文.

W0 2008140749 A2, 2008. 11. 20, 全文.

王卫国等. 混酸法降解植物纤维素技术研究. 《生物技术》. 2002, 第 12 卷(第 3 期), 第 11 - 13 页.

审查员 张蕊

权利要求书2页 说明书53页

序列表3页 附图4页

(54) 发明名称

具有纤维素分解增强活性的多肽和编码该多肽的多核苷酸

(57) 摘要

本发明涉及具有纤维素分解增强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. 一种具有纤维素分解增强活性的分离的多肽,其由氨基酸序列组成,所述氨基酸序列与 SEQ ID NO:2 的成熟多肽具有至少 100% 同一性。

2. 权利要求 1 的多肽,其由包含核苷酸序列的多核苷酸编码,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 100% 同一性。

3. 权利要求 1 的多肽,其是嗜松青霉 (*Penicillium pinophilum*) 多肽。

4. 多肽,其由 SEQ ID NO:2 的氨基酸序列组成。

5. 多肽,其由 SEQ ID NO:2 的氨基酸 22 至 322 组成。

6. 多肽,其由多核苷酸编码,所述多核苷酸由 SEQ ID NO:1 的核苷酸序列组成。

7. 多肽,其由多核苷酸编码,所述多核苷酸由 SEQ ID NO:1 的核苷酸 64 至 1018 组成。

8. 多肽,所述多肽由包含在大肠杆菌 DSM 22711 中的质粒中含有的多核苷酸编码。

9. 一种组合物,其包含权利要求 1-8 中任一项的多肽。

10. 一种分离的多核苷酸,其编码权利要求 1-8 中任一项的多肽。

11. 一种核酸构建体或表达载体,其包含权利要求 10 的多核苷酸,该多核苷酸可操作地连接于一个或多个指导多肽在表达宿主中产生的调控序列。

12. 一种重组宿主细胞,其包含可操作地连接于一个或多个指导多肽产生的调控序列的权利要求 10 的多核苷酸,其中所述重组宿主细胞不选自植物细胞。

13. 一种产生权利要求 1-8 中任一项的多肽的方法,其包括:(a) 在有助于所述多肽产生的条件下培养细胞,所述细胞以其野生型形式产生所述多肽;和 (b) 回收所述多肽。

14. 一种产生具有纤维素分解增强活性的多肽的方法,其包括:(a) 在有助于所述多肽产生的条件下培养权利要求 12 的宿主细胞;和 (b) 回收所述多肽。

15. 一种产生具有纤维素分解增强活性的多肽的方法,其包括:(a) 在有助于所述多肽产生的条件下培养转基因植物或植物细胞,其经编码权利要求 1-8 中任一项的多肽的多核苷酸转化;和 (b) 回收所述多肽。

16. 一种产生亲本细胞的突变体的方法,其包括使编码权利要求 1-8 中任一项的多肽的多核苷酸失活,其导致突变体与亲本细胞相比产生较少的所述多肽。

17. 由权利要求 16 的方法产生的突变细胞,其中所述突变细胞不选自植物细胞。

18. 权利要求 17 的突变细胞,其进一步包含编码天然或异源蛋白的基因。

19. 一种产生蛋白的方法,其包括:(a) 在有助于所述蛋白产生的条件下培养权利要求 17 或 18 的突变细胞;和 (b) 回收所述蛋白。

20. 一种用于降解或转化纤维素材料的方法,其包括:在权利要求 1-8 任一项的具有纤维素分解增强活性的多肽的存在下用酶组合物处理所述纤维素材料,其中所述酶组合物包含一种或多种选自下组的蛋白:纤维素酶、半纤维素酶、棒曲霉素、酯酶、木质素分解酶、果胶酶、过氧化物酶、蛋白酶和膨胀素。

21. 权利要求 20 的方法,其中所述纤维素材料经预处理。

22. 权利要求 20 的方法,其中所述纤维素酶为一种或多种选自下组的酶:内切葡聚糖酶、纤维二糖水解酶和  $\beta$ - 葡萄糖苷酶。

23. 权利要求 20 的方法,其中所述半纤维素酶为一种或多种选自下组的酶:乙酰甘露聚糖酯酶、乙酰木聚糖酯酶、阿拉伯聚糖酶、阿拉伯呋喃糖苷酶、香豆酸酯酶、阿魏酸酯酶、半乳糖苷酶、葡萄糖醛酸糖苷酶、葡萄糖醛酸酯酶、甘露聚糖酶、甘露糖苷酶、木聚糖酶和木糖苷

酶。

24. 权利要求 20-23 任一项的方法,进一步包括回收经降解的纤维素材料。

25. 权利要求 24 的方法,其中所述经降解的纤维素材料是糖。

26. 权利要求 25 的方法,其中所述糖选自下组:葡萄糖、木糖、甘露糖、半乳糖和阿拉伯糖。

27. 一种产生发酵产物的方法,其包括:(a) 在权利要求 1-8 中任一项的具有纤维素分解增强活性的多肽存在下,用酶组合物糖化纤维素材料,其中所述酶组合物包含选自下组的一种或多种蛋白:纤维素酶、半纤维素酶、棒曲霉素、酯酶、木质素分解酶、果胶酶、过氧化物酶、蛋白酶和膨胀素;(b) 用一种或多种发酵微生物发酵经糖化的纤维素材料以产生发酵产物;和(c) 从所述发酵回收所述发酵产物。

28. 权利要求 27 的方法,其中所述纤维素材料经预处理。

29. 权利要求 27 的方法,其中所述纤维素酶为一种或多种选自下组的酶:内切葡聚糖酶、纤维二糖水解酶和  $\beta$ -葡萄糖苷酶。

30. 权利要求 27 的方法,其中所述半纤维素酶为一种或多种选自下组的酶:乙酰甘露聚糖酯酶、乙酰木聚糖酯酶、阿拉伯聚糖酶、阿拉伯呋喃糖苷酶、香豆酸酯酶、阿魏酸酯酶、半乳糖苷酶、葡糖醛酸糖苷酶、葡糖醛酸酯酶、甘露聚糖酶、甘露糖苷酶、木聚糖酶和木糖苷酶。

31. 权利要求 27-30 中任一项的方法,其中步骤(a)和(b)在同步糖化和发酵中同时进行。

32. 权利要求 27-30 中任一项的方法,其中发酵产物是醇、有机酸、酮、氨基酸或气体。

33. 一种发酵纤维素材料的方法,其包括:用一种或多种发酵微生物发酵纤维素材料,其中在存在权利要求 1-8 中任一项的具有纤维素分解增强活性的多肽的存在下用酶组合物糖化所述纤维素材料,其中所述酶组合物包含选自下组的一种或多种蛋白:纤维素酶、半纤维素酶、棒曲霉素、酯酶、木质素分解酶、果胶酶、过氧化物酶、蛋白酶和膨胀素。

34. 权利要求 33 的方法,其中所述纤维素材料的发酵产生发酵产物。

35. 权利要求 34 的方法,其进一步包括从所述发酵回收发酵产物。

36. 权利要求 33-35 中任一项的方法,其中所述纤维素材料在糖化前经过预处理。

37. 权利要求 33 的方法,其中所述纤维素酶为一种或多种选自下组的酶:内切葡聚糖酶、纤维二糖水解酶和  $\beta$ -葡萄糖苷酶。

38. 权利要求 33 的方法,其中所述半纤维素酶为一种或多种选自下组的酶:乙酰甘露聚糖酯酶、乙酰木聚糖酯酶、阿拉伯聚糖酶、阿拉伯呋喃糖苷酶、香豆酸酯酶、阿魏酸酯酶、半乳糖苷酶、葡糖醛酸糖苷酶、葡糖醛酸酯酶、甘露聚糖酶、甘露糖苷酶、木聚糖酶和木糖苷酶。

39. 权利要求 34、35、37、38 中任一项的方法,其中发酵产物是醇、有机酸、酮、氨基酸,或气体。

## 具有纤维素分解增强活性的多肽和编码该多肽的多核苷酸

- [0001] 关于对联邦资助的研究和开发下完成的专利的权利
- [0002] 本发明是部分在能源部授予的合作协议 DE-FC36-08G018080 下完成的发明。政府对本发明拥有一定权利
- [0003] 涉及序列表
- [0004] 本申请包含计算机可读形式的序列表,其通过提述并入本文。
- [0005] 涉及生物材料的保藏
- [0006] 本申请包含对于生物材料保藏的引用,所述保藏通过提述并入本文。
- [0007] 发明背景

### 发明领域

[0008] 本发明涉及具有纤维素分解增强活性的多肽和编码所述多肽的多核苷酸。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。

### [0009] 相关领域描述

[0010] 纤维素是单糖葡萄糖通过  $\beta$ -1,4- 键连接的聚合物。许多微生物产生水解  $\beta$ - 连接的葡聚糖的酶。这些酶包括内切葡聚糖酶、纤维二糖水解酶和  $\beta$ - 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。纤维二糖是水溶性的  $\beta$ -1,4- 连接的葡萄糖二聚体。 $\beta$ - 葡糖苷酶将纤维二糖水解成葡萄糖。

[0011] 将含木素纤维素原料 (lignocellulosic feedstock) 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合意性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。

[0012] 本领域中有利的是改进酶法降解木素纤维素原料的能力。

[0013] WO 2005/074647 公开了来自土生梭孢霉 (*Thielavia terrestris*) 的具有纤维素分解增强活性的分离的多肽和它的多核苷酸。WO 2005/074656 公开了来自桔橙嗜热子囊菌 (*Thermoascus aurantiacus*) 的具有纤维素分解增强活性的分离的多肽和它的多核苷酸。WO 2007/089290 公开了来自里氏木霉 (*Trichoderma reesei*) 的具有纤维素分解增强活性的多肽和它的多核苷酸。

[0014] 本发明提供具有纤维素分解增强活性的多肽和编码该多肽的多核苷酸。

### [0015] 发明概述

[0016] 本发明涉及具有纤维素分解增强活性的分离的多肽,所述多肽选自下组:

[0017] (a) 多肽,其与 SEQ ID NO:2 的成熟多肽具有至少 80% 序列同一性;

[0018] (b) 多肽,其由多核苷酸编码,所述多核苷酸在高严格条件下与以下杂交:(i) SEQ ID NO:1 的成熟多肽编码序列,(ii) 包含于 SEQ ID NO:1 的成熟多肽编码序列中的 cDNA 序列,或 (iii) (i) 或 (ii) 的全长互补链;

[0019] (c) 多肽,其由多核苷酸编码,所述多核苷酸与 SEQ ID NO:1 的成熟多肽编码序列或其 cDNA 序列具有至少 80% 序列同一性;

[0020] (d) SEQ ID NO:2 的包含取代、缺失和 / 或插入一个或多个 (几个) 氨基酸的成熟多肽的变体;和

[0021] (e) (a)、(b)、(c) 或 (d) 的多肽具有纤维素分解增强活性的片段。

[0022] 本发明还涉及编码本发明的多肽的分离的多核苷酸;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,和产生所述多肽的方法。

[0023] 本发明还涉及用于降解或转化纤维素材料的方法,其包括:在存在本发明具有纤维素分解增强活性的多肽的情况下,用酶组合物处理纤维素材料。在一个优选的方面,本方法进一步包括回收经降解或转化的纤维素材料。

[0024] 本发明还涉及产生发酵产物的方法,其包括:(a) 在存在本发明的具有纤维素分解增强活性的多肽的情况下用酶组合物糖化纤维素材料;(b) 用一种或多种发酵微生物发酵经糖化的纤维素材料以产生发酵产物;和 (c) 从发酵回收发酵产物。

[0025] 本发明还涉及发酵纤维素材料的方法,其包括:用一种或多种发酵微生物发酵纤维素材料,其中在存在本发明的具有纤维素分解增强活性的多肽的情况下,用酶组合物糖化纤维素材料。在一个方面,所述纤维素材料的发酵产生发酵产物。在另一个方面,本发明进一步包括从发酵回收发酵产物。

[0026] 本发明还涉及编码信号肽的多核苷酸,所述信号肽包含 SEQ ID NO:2 的氨基酸 1 至 21,或由 SEQ ID NO:2 的氨基酸 1 至 21 组成,并可操作地连接于编码蛋白的基因;涉及包含所述多核苷酸的核酸构建体、表达载体和重组宿主细胞;并涉及产生蛋白的方法。

[0027] 附图简述

[0028] 图 1 显示编码具有纤维素分解增强活性的 GH61A 多肽的嗜松青霉 (*Penicillium pinophilum*) 株 NN046877 基因的基因组 DNA 序列和推定的氨基酸序列 (分别为 SEQ ID NO:1 和 2)。

[0029] 图 2 显示 PFJ0355 的限制图谱。

[0030] 图 3 显示 pPpin7 的限制图谱。

[0031] 图 4 显示 pGEM-T-Ppin7 的限制图谱。

[0032] 图 5 显示水解对添加的具有纤维素分解增强活性的嗜松青霉 GH61A 多肽的浓度。空心圈:3 日水解程度;实心圈:7 日水解程度。数据未就 PCS 液中存在的糖进行校正。数据以经修饰的非协同饱和结合模型 (modified, non-cooperative saturation-binding model) 进行拟合。

[0033] 定义

[0034] 纤维素分解增强活性:术语“纤维素分解增强活性”意指由 GH61 多肽催化的、增强通过具有纤维素分解活性的酶水解纤维素材料的生物学活性。就本发明而言,通过测量由纤维素分解酶在下述条件下水解纤维素材料所导致的还原糖增加或纤维二糖与葡萄糖的总量增加来确定纤维素分解增强活性:1-50mg 总蛋白 /g PCS 中的纤维素,其中总蛋白包含 50-99.5% w/w 的纤维素分解酶,及 0.5-50% w/w 的具有纤维素分解增强活性的 GH61 多肽的蛋白质,在 50°C 历时 1-7 天,与用等量的总蛋白加载量而无纤维素分解增强活性 (1-50mg 纤维素分解蛋白 /g PCS 中的纤维素) 所进行的对照水解相比。在一个优选的方面,使用在

总蛋白重量的 2-3% 的米曲霉  $\beta$ -葡糖苷酶 (根据 WO 02/095014 在米曲霉中重组产生) 或者总蛋白质量的 2-3% 的烟曲霉  $\beta$ -葡糖苷酶 (如 WO 2002/095014 所述在米曲霉中重组产生) 的纤维素酶蛋白加载量存在下的 CELLUCLAST® 1.5L (Novozymes A/S, Bagsvaerd, Denmark) 的混合物作为纤维素分解活性的来源。

[0035] 所述具有纤维素分解增强活性的 GH61 多肽通过降低达到相同水解水平所需的纤维素分解酶的量而增强由具有纤维素分解活性的酶催化的纤维素材料的水解, 优选降低至少 1.01 倍, 更优选至少 1.05 倍, 更优选至少 1.10 倍, 更优选至少 1.25 倍, 更优选至少 1.5 倍, 更优选至少 2 倍, 更优选至少 3 倍, 更优选至少 4 倍, 更优选至少 5 倍, 甚至更优选至少 10 倍, 并且最优选至少 20 倍。

[0036] 本发明的多肽具有 SEQ ID NO :2 的成熟多肽的纤维素分解增强活性的至少 20%, 例如至少 40%, 至少 50%, 至少 60%, 至少 70%, 至少 80%, 至少 90%, 至少 95% 和至少 100%。

[0037] 家族 61 糖苷水解酶: 术语“家族 61 糖苷水解酶”或“家族 GH61”或“GH61”意指根据 Henrissat B., 1991, A classification of glycosyl hydrolases based on amino-acid sequence similarities, Biochem. J. 280 :309-316, 及 Henrissat B., 和 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 61 的多肽。

[0038] 纤维素分解酶或纤维素酶: 术语“纤维素分解酶”或“纤维素酶”意指一种或多种 (几种) 水解纤维素材料的酶。此类酶包括内切葡聚糖酶、纤维二糖水解酶、 $\beta$ -葡糖苷酶或其组合。测量纤维素分解活性的两种基本方法包括: (1) 测量总纤维素分解活性, 和 (2) 测量单独的纤维素分解活性 (内切葡聚糖酶、纤维二糖水解酶和  $\beta$ -葡糖苷酶), 如 Zhang 等, Outlook for cellulase improvement: Screening and selection strategies, 2006, Biotechnology Advances 24 :452-481 所综述的。总纤维素分解活性通常是使用不溶性底物来测定的, 所述底物包括 Whatman No. 1 滤纸、微晶纤维素、细菌纤维素、藻类纤维素、棉花、经预处理的木素纤维素等。最常见的总纤维素分解活性测定法是使用 Whatman No. 1 滤纸作为底物的滤纸测定法。该测定法是由 International Union of Pure and Applied Chemistry (IUPAC) (Ghose, 1987, Measurement of cellulase activities, Pure Appl. Chem. 59 :257-68) 确立的。

[0039] 就本发明而言, 纤维素分解酶活性通过测量在下述条件下由纤维素分解酶进行的纤维素材料水解的增加来确定: 1-20mg 的纤维素分解酶蛋白 /g 的 PCS 中纤维素在 50°C 进行 3-7 日, 与未添加纤维素分解酶蛋白的对照水解相比较。通常条件为: 1ml 反应液, 经洗涤或未洗涤的 PCS, 5% 不溶性固形物, 50mM 乙酸钠 pH 5, 1mM  $MnSO_4$ , 50°C, 72 小时, 通过 AMINEX® HPX-87H 柱 (Bio-Rad Laboratories, Inc., Hercules, CA, USA) 进行糖分析。

[0040] 内切葡聚糖酶: 术语“内切葡聚糖酶”意指内切 -1,4-(1,3 ;1,4)- $\beta$ -D-葡聚糖 4-葡聚糖水解酶 (endo-1,4- $\beta$ -D-glucan 4-glucanohydrolase) (E. C. 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- $\beta$ -D-糖苷键、混合的  $\beta$ -1,3 葡聚糖例如谷类  $\beta$ -D-葡聚糖或木葡聚糖中的  $\beta$ -1,4 键和含有纤维素组分的其它植物材料的内水解 (endohydrolysis)。内切葡聚糖酶活性可通过测量底物粘度的减少或由还原糖测定法 (Zhang 等, 2006, Biotechnology Advances 24 :

452-481) 确定的还原端增加来确定。就本发明而言,根据 Ghose,1987, Pure and Appl. Chem. 59 :257-268 的方法,在 pH 5, 40°C, 使用羧甲基纤维素 (CMC) 作为底物来确定内切葡聚糖酶活性。

[0041] 纤维二糖水解酶:术语“纤维二糖水解酶”意指 1,4- $\beta$ -D-葡聚糖纤维二糖水解酶 (1,4-beta-D-glucan cellobiohydrolase) (E. C. No. 3. 2. 1. 91), 其催化纤维素、纤维素寡糖, 或任何包含  $\beta$ -1,4-连接的葡萄糖的聚合物中的 1,4- $\beta$ -D-糖苷键的水解, 从链的还原或非还原末端释放纤维二糖 (Teeri,1997, Crystalline cellulose degradation :New insight into the function of cellobiohydrolases, Trends in Biotechnology 15 : 160-167 ;Teeri 等,1998, Trichoderma reesei cellobiohydrolases :why so efficient on crystalline cellulose ? , Biochem. Soc. Trans. 26 :173-178)。就本发明而言,根据 Lever 等,1972, Anal. Biochem. 47 :273-279 ;van Tilbeurgh 等,1982, FEBS Letters 149 : 152-156 ;van Tilbeurgh 和 Claeyssens,1985, FEBS Letters 187 :283-288 和 Tomme 等, 1988, Eur. J. Biochem. 170 :575-581 描述的方法来测定纤维二糖水解酶活性。在本发明中,可采用 Lever 等的方法来评价玉米秸秆中纤维素的水解,而 van Tilbeurgh 等和 Tomme 等的方法可用于确定针对荧光二糖衍生物 4-甲基伞形基- $\beta$ -D-乳糖苷的纤维二糖水解酶活性。

[0042]  $\beta$ -葡糖苷酶:术语“ $\beta$ -葡糖苷酶”意指  $\beta$ -D-葡糖苷葡萄糖水解酶 (beta-D-glucoside glucohydrolase) (E. C. No. 3. 2. 1. 21), 其催化末端非还原  $\beta$ -D-葡萄糖残基的水解,并释放  $\beta$ -D-葡萄糖。就本发明而言,根据 Venturi 等,2002, Extracellular beta-D-glucosidase from Chaetomium thermophilum var. coprophilum :production, purification and some biochemical properties, J. Basic Microbiol. 42 :55-66 描述的基本方法确定  $\beta$ -葡糖苷酶活性。一个单位的  $\beta$ -葡糖苷酶定义为在 25°C, pH 4.8, 在含有 0.01% TWEEN® 20 的 50mM 柠檬酸钠中由作为底物 1mM 对硝基苯基- $\beta$ -D-葡萄糖吡喃糖苷每分钟产生 1.0 微摩尔对硝基苯酚阴离子。

[0043] 半纤维素分解酶或半纤维素酶:术语“半纤维素分解酶”或“半纤维素酶”意指一种或多种(几种)水解半纤维素材料的酶。参见,例如 Shallom, D. 和 Shoham, Y., Microbial hemicellulases. Current Opinion In Microbiology, 2003, 6(3) :219-228)。半纤维素酶是植物生物质降解中的关键组分。半纤维素酶的实例包括但不限于:乙酰甘露聚糖酯酶,乙酰木聚糖酯酶,阿拉伯聚糖酶,阿拉伯呋喃糖苷酶,香豆酸酯酶,阿魏酸酯酶,半乳糖苷酶,葡糖醛酸糖苷酶,葡糖醛酸酯酶,甘露聚糖酶,甘露糖苷酶,木聚糖酶和木糖苷酶。这些酶的底物,半纤维素,是支链和直链糖类的异质组,所述糖类在植物细胞壁中通过氢键结合于纤维素微纤维,交联为鲁棒的 (robust) 网络。半纤维素亦共价地附于木质素,与纤维素一同形成高度复杂的结构。半纤维素多变的结构和组织需要多种酶的协同作用才能使其完全降解。半纤维素酶的催化模块为水解糖苷键的糖基水解酶 (GH), 或糖酯酶 (CE), 其水解乙酸或阿魏酸侧基的酯连接。这些催化模块,基于其一级序列的同源性,可分配至通过数字标识的 GH 和 CE 家族。一些家族,由于总体类似的折叠,可进一步归类为宗族 (clan), 用字母编号 (例如, GH-A)。对于这些和其他糖活性酶的最具信息性和最新的分类可从 Carbohydrate-Active Enzymes (CAZy) 数据库获得。半纤维素分解酶活性可根据 Ghose 和 Bisaria, 1987, Pure & Appl. Chem. 59 :1739-1752 测量。

[0044] 木聚糖降解活性或木聚糖分解活性：术语“木聚糖降解活性”或“木聚糖分解活性”意指水解含木聚糖材料的生物学活性。两种测定木聚糖分解活性的基础方法包括：(1) 测定总木聚糖分解活性，和 (2) 测定单独的木聚糖分解活性（例如内切木聚糖酶、 $\beta$ -木糖苷酶、阿拉伯呋喃糖苷酶、 $\alpha$ -葡萄糖醛酸糖苷酶、乙酰木聚糖酯酶、阿魏酸酯酶和  $\alpha$ -葡萄糖醛酸酯酶 ( $\alpha$ -glucuronyl esterase))。最近在木聚糖分解酶测定法的进展总结于几个公开文献中，包括 Biely 和 Puchard, Recent progress in the assays of xylanolytic enzymes, 2006, Journal of the Science of Food and Agriculture 86(11):1636-1647; Spanikova 和 Biely, 2006, Glucuronoyl esterase—Novel carbohydrate esterase produced by Schizophyllum commune, FEBS Letters 580(19):4597-4601; Herrmann, Vrsanska, Jurickova, Hirsch, Biely 和 Kubicek, 1997, The beta-D-xylosidase of Trichoderma reesei is a multifunctional beta-D-xylan xylohydrolase, Biochemical Journal 321:375-381。

[0045] 总木聚糖降解活性可通过确定从多种类型的木聚糖形成的还原糖来测量，所述木聚糖包括例如燕麦小麦 (oat spelt)、山毛榉木 (beechwood) 和落叶松木 (larchwood) 木聚糖，或者可通过光度法确定从多种共价染色的木聚糖释放出的染色的木聚糖片段来测量。最常见的总木聚糖分解活性测定法基于从多聚的 4-O-甲基葡萄糖醛酸木聚糖产生还原糖，如 Bailey, Biely, Poutanen, 1992, Interlaboratory testing of methods for assay of xylanase activity, Journal of Biotechnology 23(3):257-270 中所述。木聚糖酶活性亦可在 37°C 用 0.2% AZCL-阿拉伯木聚糖作为底物在 0.01% Triton X-100 和 200mM 磷酸钠缓冲液 pH 6 中确定。一个单位的木聚糖酶活性定义为在 37°C, pH 6, 从 200mM 磷酸钠 pH 6 缓冲液中作为底物的 0.2% AZCL-阿拉伯木聚糖每分钟产生 1.0 微摩尔的天青精 (azurine)。

[0046] 就本发明而言，木聚糖降解活性是通过测量由木聚糖降解酶在下述通常条件下造成的桦木木聚糖 (Sigma Chemical Co., Inc., St. Louis, MO, USA) 水解的增加来确定的：1ml 反应液, 5mg/ml 底物 (总固形物), 5mg 木聚糖分解蛋白质 /g 底物, 50mM 乙酸钠, pH 5, 50°C, 24 小时, 如 Lever, 1972, A new reaction for colorimetric determination of carbohydrates, Anal. Biochem 47:273-279 所述使用对羟基苯甲酸酰肼 (PHBAH) 测定法进行糖分析。

[0047] 木聚糖酶：术语“木聚糖酶”意指 1,4- $\beta$ -D-木聚糖-木糖水解酶 (1,4- $\beta$ -D-xylan-xylohydrolase) (E. C. 3. 2. 1. 8), 其催化木聚糖中 1,4- $\beta$ -D-木糖苷键的内水解。就本发明而言，木聚糖酶活性是使用 0.2% AZCL-阿拉伯木聚糖作为底物在 37°C 在 0.01% Triton X-100 和 200mM 磷酸钠缓冲液 pH 6 中确定的。一个单位的木聚糖酶活性定义为在 37°C, pH 6, 从 200mM 磷酸钠 pH 6 缓冲液中作为底物的 0.2% AZCL-阿拉伯木聚糖每分钟产生 1.0 微摩尔的天青精 (azurine)。

[0048]  $\beta$ -木糖苷酶：术语“ $\beta$ -木糖苷酶”意指  $\beta$ -D-木糖苷木糖水解酶 ( $\beta$ -D-xyloside xylohydrolase) (E. C. 3. 2. 1. 37), 其催化短  $\beta$  (1  $\rightarrow$  4) 木寡糖 (xylooligosaccharide) 的外水解以从非还原端去除连续的 D-木糖残基。就本发明而言，一个单位的  $\beta$ -木糖苷酶定义为在 40°C, pH 5 从 1mM 作为底物的对硝基苯基- $\beta$ -D-木糖苷在含有 0.01% TWEEN® 20 的 100mM 柠檬酸钠中每分钟产生 1.0  $\mu$ mol 对硝基苯酚阴离子。

[0049] 乙酰木聚糖酯酶：术语“乙酰木聚糖酯酶”意指羧基酯酶活性 (EC3. 1. 1. 72),



其催化乙酰基从聚合木聚糖、乙酰化木糖、乙酰化葡萄糖、乙酸  $\alpha$ -萘酯 (alpha-naphthyl acetate) 和乙酸对硝基苯酯 (p-nitrophenyl acetate) 的水解。就本发明而言,乙酰木聚糖酯酶活性是使用 0.5mM 乙酸对硝基苯酯作为底物,在含有 0.01% TWEEN™ 20 的 50mM 乙酸钠 pH 5.0 中确定的。一个单位的乙酰木聚糖酯酶定义为能够在 pH 5, 25°C 每分钟释放 1 微摩尔对硝基苯酚阴离子 (p-nitrophenolate anion) 的酶量。

[0050] 阿魏酸酯酶:术语“阿魏酸酯酶 (feruloyl esterase)”意指 4-羟基-3-甲氧基肉桂酰-糖水解酶 (EC 3.1.1.73),其催化 4-羟基-3-甲氧基肉桂酰(阿魏酰)基团从酯化的糖(其在“天然”底物中通常为阿拉伯糖)的水解,以产生阿魏酸(4-羟基-3-甲氧基肉桂酸)。阿魏酸酯酶也称作阿魏酸酯酶 (ferulic acid esterase)、羟基肉桂酰基酯酶、FAE-III、肉桂酸酯水解酶、FAEA、cinnAE、FAE-I 或 FAE-II。就本发明而言,阿魏酸酯酶活性是使用 50mM 乙酸钠 pH 5.0 中的阿魏酸对硝基苯酯作为底物确定的。一个单位的阿魏酸酯酶等于能够在 pH 5, 25°C 每分钟释放出 1  $\mu$ mol 对硝基苯酚阴离子的酶量。

[0051]  $\alpha$ -葡糖醛酸糖苷酶:术语“ $\alpha$ -葡糖醛酸糖苷酶”意指  $\alpha$ -D-葡糖苷酸葡糖醛酸水解酶 (alpha-D-glucosiduronate glucuronohydrolase activity) (EC 3.2.1.139),其催化  $\alpha$ -D-葡糖醛酸糖苷水解为 D-葡糖醛酸和醇。就本发明而言, $\alpha$ -葡糖醛酸糖苷酶活性是根据 de Vries, 1998, J. Bacteriol. 180 :243-249 确定的。一个单位的  $\alpha$ -葡糖醛酸糖苷酶等于能够在 pH 5, 40°C 每分钟释放出 1  $\mu$ mol 葡糖醛酸或 4-O-甲基葡糖醛酸的酶量。

[0052]  $\alpha$ -L-阿拉伯呋喃糖苷酶:术语“ $\alpha$ -L-阿拉伯呋喃糖苷酶”意指  $\alpha$ -L-阿拉伯呋喃糖苷阿拉伯呋喃水解酶 (EC 3.2.1.55),其催化对  $\alpha$ -L-阿拉伯糖苷中的末端非还原性  $\alpha$ -L-阿拉伯呋喃糖苷残基的水解。该酶对  $\alpha$ -L-阿拉伯呋喃糖苷、含有 (1,3)-和/或 (1,5)-键的  $\alpha$ -L-阿拉伯聚糖、阿拉伯木聚糖和阿拉伯半乳聚糖起作用。 $\alpha$ -L-阿拉伯呋喃糖苷酶也称为阿拉伯糖苷酶、 $\alpha$ -阿拉伯糖苷酶、 $\alpha$ -L-阿拉伯糖苷酶、 $\alpha$ -阿拉伯呋喃糖苷酶、多糖  $\alpha$ -L-阿拉伯呋喃糖苷酶、 $\alpha$ -L-阿拉伯呋喃糖苷水解酶、L-阿拉伯糖苷酶或  $\alpha$ -L-阿拉伯聚糖酶。就本发明而言, $\alpha$ -L-阿拉伯呋喃糖苷酶活性是使用总体积 200  $\mu$ l 中的每 ml 的 100mM 乙酸钠 pH 5 中 5mg 的中等粘性小麦阿拉伯木聚糖 (Megazyme International Ireland, Ltd., Bray, Co. Wicklow) 在 40°C 进行 30 分钟,接着通过 AMINEX® HPX-87H 柱层析 (Bio-Rad Laboratories, Inc., Hercules, CA, USA) 的阿拉伯糖分析来确定的。

[0053] 纤维素材料:纤维素材料可以是包含纤维素的任何材料。生物质的初生细胞壁 (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共价交联至半纤维素的聚合木质素而加强。纤维素是脱水纤维二糖的均聚物,并且因此是直链  $\beta$ -(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是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。

[0054] 纤维素通常见于例如植物的茎、叶、壳、皮和穗轴,或树的叶、枝和木材。纤维素材料可以是,但不限于,草本材料、农业残余物、林业残余物、城市固体废物、废纸和纸浆与造纸厂残余物(参见,例如, Wiseloge 等, 1995, 于 Handbook on Bioethanol (Charles

E. Wyman 编), pp. 105-118, Taylor & Francis, Washington D. C. ;Wyman, 1994, Bioresource Technology 50 :3-16 ;Lynd, 1990, Applied Biochemistry and Biotechnology 24/25 : 695-719 ;Mosier 等, 1999, Recent Progress in Bioconversion of Lignocellulosics, 于 Advances in Biochemical Engineering/Biotechnology, T. Scheper 主编, Volume 65, pp. 23-40, Springer-Verlag, New York)。在本文中应理解的是,纤维素可以是任何形式的木素纤维素,在混合基质中包含木质素、纤维素和半纤维素的植物细胞壁材料。在一个优选的方面,纤维素材料是木素纤维素,其包含纤维素、半纤维素和木质素。

[0055] 在一个方面,纤维素材料是草本材料。在另一个方面,纤维素材料是农业残余物。在另一个方面,纤维素材料是林业残余物。在另一个方面,纤维素材料是城市固体废物。在另一个方面,纤维素材料是废纸。在另一个方面,纤维素材料是纸浆和造纸厂残余物。

[0056] 在另一个方面,纤维素材料是玉米秸秆。在另一个方面,纤维素材料是玉米纤维。在另一个方面,纤维素材料是玉米穗轴。在另一个方面,纤维素材料是橙皮。在另一个方面,纤维素材料是稻秆。在另一个方面,纤维素材料是麦秆。在另一个方面,纤维素材料是柳枝稷 (switch grass)。在另一个方面,纤维素材料是芒草属。在另一个方面,纤维素材料是甘蔗渣。

[0057] 在另一个方面,纤维素材料是微晶纤维素。在另一个方面,纤维素材料是细菌纤维素。在另一个方面,纤维素材料是藻类纤维素。在另一个方面,纤维素材料是棉线头 (cotton linter)。在另一个方面,纤维素材料是无定形的磷酸处理的纤维素。在另一个方面,纤维素材料是滤纸。

[0058] 纤维素材料可以直接使用或进行预处理,使用本领域已知的常规方法,如本文所述。在一个优选的方面,预处理纤维素材料。

[0059] 预处理的玉米秸秆:术语“PCS”或“预处理的玉米秸秆”意指通过用热和稀硫酸处理的源自玉米秸秆的纤维素材料。

[0060] 含木聚糖材料:术语“含木聚糖材料”意指任何包含含有  $\beta$ -(1-4) 连接的木糖残基骨架的植物细胞壁多糖的材料。陆生植物的木聚糖是具有  $\beta$ -(1-4)-吡喃木糖骨架的杂聚物,其由短的糖链分支。它们包含 D-葡萄糖醛酸或其 4-O-甲基醚(或酯), L-阿拉伯糖和/或多种包含 D-木糖、L-阿拉伯糖、D-或 L-半乳糖和 D-葡萄糖的寡糖。木聚糖类型的多糖可分为均木聚糖 (homoxylan) 和杂木聚糖 (heteroxylan),后者包括葡萄糖醛酸木聚糖,(阿拉伯)葡萄糖醛酸木聚糖,(葡萄糖醛酸)阿拉伯木聚糖,阿拉伯木聚糖和复合杂木聚糖。参见,例如 Ebringerova 等, 2005, Adv. Polym. Sci. 186 :1-6。

[0061] 在本发明的方法中,可使用任何含有木聚糖的材料。在一个优选的方面,所述含木聚糖材料是木素纤维素。

[0062] 分离或纯化:术语“分离”或“纯化”意指从与其天然结合的至少一种成分移除的多肽或多核苷酸。举例而言,多肽可为至少 1% 纯,例如至少 5% 纯,至少 10% 纯,至少 20% 纯,至少 40% 纯,至少 60% 纯,至少 80% 纯,至少 90% 纯或至少 95% 纯,如通过 SDS-PAGE 确定,而多核苷酸可为至少 1% 纯,例如至少 5% 纯,至少 10% 纯,至少 20% 纯,至少 40% 纯,至少 60% 纯,至少 80% 纯,至少 90% 纯或至少 95% 纯,如通过琼脂糖电泳确定。

[0063] 成熟多肽:术语“成熟多肽”意指以其在翻译和任何翻译后修饰之后的最终形式存在的多肽,所述修饰例如 N-末端加工、C-末端截短、糖基化、磷酸化等。在一个方面,

SignalP 软件 (Nielsen 等,1997,Protein Engineering 10 :1-6) 预测 SEQ ID NO :2 的氨基酸 1 至 21 是信号肽,根据该预测,成熟多肽是 SEQ ID NO :2 的氨基酸 22 至 322。本领域中已知宿主细胞可产生由相同多核苷酸表达的两种或更多种不同成熟多肽的混合物(即,具有不同的 C 端和 / 或 N 端氨基酸)。

[0064] 成熟多肽编码序列:术语“成熟多肽编码序列”意指编码具有纤维素分解增强活性的成熟多肽的多核苷酸。在一个方面,SignalP 软件 (Nielsen 等,1997,见上) 预测 SEQ ID NO :1 的核苷酸 1 至 63 编码信号肽,根据该预测,成熟多肽编码序列是 SEQ ID NO :1 的核苷酸 64 至 1018。在另一个方面,所述成熟多肽编码序列是包含于 SEQ ID NO :1 的核苷酸 64 至 1018 中的 cDNA 序列。

[0065] 序列同一性:参数“序列同一性”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。

[0066] 就本发明而言,两个氨基酸序列之间的序列同一性程度使用如在 EMBOSS 软件包 (EMBOSS:The European Molecular Biology Open Software Suite,Rice 等,2000,Trends Genet. 16 :276-277) 的 Needle 程序,优选 3.0.0 版或更高版本中所执行的 Needleman-Wunsch 算法 (Needleman 和 Wunsch,1970,J. Mol. Biol. 48 :443-453) 来测定。使用的可选参数为缺口罚分 (gap penalty)10,缺口延伸罚分 (gap extension penalty)0.5 和 EBLOSUM62 (BLOSUM62 的 EMBOSS 版) 取代矩阵。使用 Needle 标记为“最高同一性 (longest identity)”的输出结果 (使用 -nobrief 选项获得) 作为同一性百分比,并计算如下:

[0067]  $(\text{同样的残基} \times 100) / (\text{比对长度} - \text{比对中缺口的总数})$

[0068] 就本发明而言,两个核苷酸序列之间的序列同一性程度使用如在 EMBOSS 软件包 (EMBOSS:The European Molecular Biology Open Software Suite,Rice 等,2000,见上文) 的 Needle 程序,优选 3.0.0 版或更高版本中所执行的 Needleman-Wunsch 算法 (Needleman 和 Wunsch,1970,见上文) 来测定。使用的可选参数为缺口罚分 10,缺口延伸罚分 0.5 和 EDNAFULL (NCBI NUC4.4 的 EMBOSS 版) 取代矩阵。使用 Needle 标记为“最高同一性”的输出结果 (使用 -nobrief 选项获得) 作为同一性百分比,并计算如下:

[0069]  $(\text{同样的脱氧核糖核苷酸} \times 100) / (\text{比对长度} - \text{比对中缺口的总数})$

[0070] 片段:术语“片段”意指从的成熟多肽的氨基和 / 或羧基端缺失一个或多个 (几个) 氨基酸的多肽;其中所述片段具有纤维素分解增强活性。在一个方面,所述片段含有 SEQ ID NO :2 的成熟多肽的至少 255 个氨基酸残基,更优选至少 270 个氨基酸残基,并且最优选至少 285 个氨基酸残基。

[0071] 亚序列:术语“亚序列 (subsequence)”意指从成熟多肽编码序列的 5' 和 / 或 3' 端缺失一个或多个 (几个) 核苷酸的多核苷酸;其中所述亚序列编码具有纤维素分解增强活性的片段。在一个方面,所述亚序列含有 SEQ ID NO :1 的成熟多肽编码序列的至少 765 个核苷酸,更优选至少 810 个核苷酸,并且最优选至少 855 个核苷酸。

[0072] 等位变体 (allelic variant):术语“等位变体”意指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。

[0073] 编码序列:术语“编码序列”意指直接指定多肽氨基酸序列的多核苷酸。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、合成的或重组的多核苷酸。

[0074] cDNA :术语“cDNA”意指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(包括剪接)加工然后作为成熟的已剪接的 mRNA 出现。

[0075] 核酸构建体 :术语“核酸构建体”意指单链或双链的核酸分子,其分离自天然存在的基因,或将其以本来不存在于(not otherwise exist)自然界中的方式修饰以含有核酸的区段,或其为合成的。当所述核酸构建体含有表达本发明的编码序列所需的调控序列时,术语核酸构建体与术语“表达盒”同义。

[0076] 调控序列(control sequence) :术语“调控序列”意指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。各个调控序列对于编码所述多肽的多核苷酸可以是天然的或外源的,或各个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便于调控序列与编码多肽的多核苷酸编码区的连接。

[0077] 可操作地连接 :术语“可操作地连接”意指这样的构型,其中将调控序列置于相对于多核苷酸的编码序列的适当位置,使得调控序列指导编码序列的表达。

[0078] 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。

[0079] 表达载体 :术语“表达载体”意指线性的或环状的 DNA 分子,其包含编码多肽的多核苷酸,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。

[0080] 宿主细胞 :术语“宿主细胞”意指任何细胞类型,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)。术语“宿主细胞”涵盖任何由于在复制过程中发生的突变而与亲本细胞不完全相同的亲本细胞后代。

[0081] 变体 :术语“人工变体”意指具有纤维素分解增强活性的多肽,所述多肽包含改变,即在一个或多个(几个)位置的取代、插入和/或缺失。取代意指用另一个氨基酸替代占据一个位置的氨基酸;缺失意指去除占据一个位置的氨基酸;而插入意指在占据一个位置的氨基酸邻接处添加一个或多个(几个)氨基酸例如 1-5 个氨基酸。

[0082] 发明详述

[0083] 具有纤维素分解增强活性的多肽

[0084] 本发明涉及具有纤维素分解增强活性的分离的多肽,所述多肽选自下组:

[0085] (a) 多肽,其与 SEQ ID NO :2 的成熟多肽具有至少 80% 序列同一性;

[0086] (b) 多肽,其由多核苷酸编码,所述多核苷酸在高严格条件下与以下杂交:(i) SEQ ID NO :1 的成熟多肽编码序列,(ii) 包含于 SEQ ID NO :1 的成熟多肽编码序列中的 cDNA 序列,或 (iii) (i) 或 (ii) 的全长互补链;

[0087] (c) 多肽,其由多核苷酸编码,所述多核苷酸与 SEQ ID NO:1 的成熟多肽编码序列或其 cDNA 序列具有至少 80% 序列同一性;

[0088] (d) SEQ ID NO:2 的包含取代、缺失和 / 或插入一个或多个 (几个) 氨基酸的成熟多肽的变体;和

[0089] (e) (a)、(b)、(c) 或 (d) 的多肽具有纤维素分解增强活性的片段。

[0090] 本发明涉及与 SEQ ID NO:2 的成熟多肽具有至少 80%, 例如至少 85%, 至少 90%, 至少 91%, 至少 92%, 至少 93%, 至少 94%, 至少 95%, 至少 96%, 至少 97%, 至少 98%, 至少 99% 或 100% 的序列同一性程度的分离的多肽, 所述多肽具有纤维素分解增强活性。在一个方面, 所述同源多肽与 SEQ IDNO:2 的成熟多肽相差十个氨基酸, 例如相差五个氨基酸, 相差四个氨基酸, 相差三个氨基酸, 相差两个氨基酸, 和相差一个氨基酸。

[0091] 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; 或它们的具有纤维素分解增强活性的片段, 或者所述多肽由包含 SEQ ID NO:2 的氨基酸序列或其等位变体; 或它们的具有纤维素分解增强活性的片段组成。在另一个方面, 多肽包含 SEQ ID NO:2 的成熟多肽或由 SEQ ID NO:2 的成熟多肽组成。在另一个优选的方面, 所述多肽包含 SEQ ID NO:2 的氨基酸 22 至 322, 或由 SEQ ID NO:2 的氨基酸 22 至 322 组成。

[0092] 本发明还涉及具有纤维素分解增强活性的分离的多肽, 所述分离的多肽由多核苷酸编码, 所述多核苷酸在高严格条件下或非常高严格条件下, 与以下杂交: (i) SEQ ID NO:1 的成熟多肽编码序列, (ii) 包含于 SEQ ID NO:1 的成熟多肽编码序列中的 cDNA 序列, 或 (iii) (i) 或 (ii) 的全长互补链 (J. Sambrook, E. F. Fritsch, 和 T. Maniatis, 1989, Molecular Cloning, A Laboratory Manual, 第 2 版, Cold Spring Harbor, New York)。

[0093] SEQ ID NO:1 的多核苷酸或其亚序列, 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纤维素分解增强活性的多肽的 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 DNA 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, 例如至少 25, 至少 35, 或至少 70 个核苷酸。优选地, 所述核酸探针是至少 100 个核苷酸的长度, 例如至少 200 个核苷酸, 至少 300 个核苷酸, 至少 400 个核苷酸, 至少 500 个核苷酸, 至少 600 个核苷酸, 至少 700 个核苷酸, 至少 800 个核苷酸, 或至少 900 个核苷酸的长度。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  $^{32}\text{P}$ 、 $^3\text{H}$ 、 $^{35}\text{S}$ 、生物素或抗生物素蛋白 (avidin) 标记)。这些探针涵盖于本发明中。

[0094] 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纤维素分解增强活性的多肽。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, 将所述载体材料优选用在 Southern 印迹中。

[0095] 就本发明而言, 杂交表示多核苷酸在非常低至非常高的严格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; SEQ ID NO:1 的成熟多肽编码序列; 包含于 SEQ ID NO:1 的成熟多肽编码序列中的 cDNA 序列; 其全长互补链; 或它们的亚序列。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。

[0096] 在一个方面,核酸探针是 SEQ ID NO :1 的成熟多肽编码序列或其 cDNA 序列。在另一个方面,核酸探针是编码 SEQ ID NO :2 的多肽或其成熟多肽,或它们的片段的多核苷酸。在另一个优选方面,核酸探针是 SEQ ID NO :1 或其 cDNA 序列。在另一个方面,核酸探针是包含在大肠杆菌 (E. coli) DSM22711 中的质粒 pGEM-T-Ppin7 中含有的多核苷酸,其中所述多核苷酸序列编码具有纤维素分解增强活性的多肽。在另一个方面,核酸探针是包含在大肠杆菌 DSM 22711 中的质粒 pGEM-T-Ppin7 中含有的成熟多肽编码区。

[0097] 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格条件定义为在 42°C,在 5X SSPE、0.3% SDS、200 微克 /ml 已剪切并且变性的鲑精 DNA 中,并且对于非常低和低严格性为 25% 的甲酰胺、对于中和中 - 高严格性为 35% 的甲酰胺、或对于高和非常高严格性为 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。使用 2X SSC、0.2% SDS 至少在 45°C (非常低严格性),至少在 50°C (低严格性),至少在 55°C (中严格性),至少在 60°C (中 - 高严格性),至少在 65°C (高严格性),至少在 70°C (非常高严格性) 将载体材料最终洗涤三次,每次 15 分钟。

[0098] 对于长度大约 15 个核苷酸至大约 70 个核苷酸的短探针,将严格条件定义为在比根据 Bolton 和 McCarthy 计算法 (1962, Natl. Acad. Sci. USA 48 :1390) 得出的  $T_m$  低大约 5°C 至大约 10°C,在 0.9M NaCl,0.09M Tris-HCl pH 7.6,6mMEDTA,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和杂交最佳 12 至 24 小时。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°C 至 10°C 的温度最终洗涤两次,每次 15 分钟。

[0099] 本发明还涉及由多核苷酸编码的具有纤维素分解增强活性的分离的多肽,所述多核苷酸与 SEQ ID NO :1 的成熟多肽编码序列或其 cDNA 序列具有至少 80%,例如至少 85%,至少 90%,至少 91%,至少 92%,至少 93%,至少 94%,至少 95%,至少 96%,至少 97%,至少 98%,至少 99% 或 100% 的序列同一性。

[0100] 本发明还涉及 SEQ ID NO :2 的成熟多肽或其同源序列的包含取代、缺失和 / 或插入一个或多个 (几个) 氨基酸的变体。优选地,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,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来促进纯化的小延伸,如多组氨酸序列 (poly histidine tract)、抗原表位 (antigenic epitope) 或结合域 (binding domain)。

[0101] 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 和 R. L. Hill, 1979, 于 The Proteins, Academic Press, New York 中描述。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。

[0102] 或者,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。

[0103] 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 和 Wells,1989, Science 244 :1081-1085) 来鉴定亲本多肽中的必需氨基酸。在上一技术中,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纤维素分解增强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 等,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 等,1992, Science 255 :306-312 ;Smith 等,1992, J. Mol. Biol. 224 :899-904 ;Wlodaver 等,1992, FEBS Lett. 309 :59-64。必需氨基酸的身份 (identity) 也能够从与多肽的同一性分析来推断,所述多肽与亲本多肽相关。

[0104] 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 和 Sauer,1988, Science 241 :53-57 ;Bowie 和 Sauer, 1989, Proc. Natl. Acad. Sci. USA 86 :2152-2156 ;WO 95/17413 ;或 W095/22625 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 等,1991, Biochemistry. 30 :10832-10837 ;美国专利 No. 5, 223, 409 ;WO 92/06204) 和区域定向的诱变 (Derbyshire 等,1986, Gene 46 :145 ;Ner 等,1988, DNA 7 :127)。

[0105] 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 等,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多肽中单个氨基酸残基的重要性。

[0106] SEQ ID NO :2 的成熟多肽的氨基酸取代、缺失和 / 或插入的总数不大于 10, 例如 1, 2, 3, 4, 5, 6, 7, 8 或 9。

[0107] 所述多肽可为杂合多肽,其中一个多肽的一部分融合于另一个多肽一部分的 N 端或 C 端。

[0108] 所述多肽可为融合多肽或可切割的融合多肽,其中另一个多肽融合于本发明的多肽的 N 端或 C 端。融合多肽是通过将编码另一个多肽的多核苷酸融合于本发明的多核苷酸产生的。用于产生融合多肽的技术在本领域是已知的,且包括将编码多肽的编码序列相连接从而使得其符合读框 (in frame),且融合多肽的表达置于相同的启动子和终止子调控下。融合多肽亦可使用内蛋白技术构建,其中融合蛋白是在翻译后构建的 (Cooper 等, 1993, EMBO J. 12 :2575-2583 ;Dawson et al., 1994, Science 266 :776-779)。

[0109] 融合多肽可进一步包含两个多肽之间的切割位点。当融合多肽分泌时,该位点受切割,释放出两个多肽。切割位点的实例包括但不限于 Martin 等,2003, J. Ind. Microbiol. Biotechnol. 3 :568-576 ;Svetina 等,2000, J. Biotechnol. 76 :245-251 ;Rasmussen-Wilson 等,1997, Appl. Environ. Microbiol. 63 :3488-3493 ;Ward 等,1995, Biotechnology 13 :498-503 ; 和 Contreras 等,1991, Biotechnology 9 :378-381 ;Eaton 等,1986, Biochemistry 25 :505-512 ;Collins-Racie 等,1995, Biotechnology 13 :982-987 ;Carter

等,1989,Proteins :Structure,Function,and Genetics 6 :240-248 ;以及 Stevens,2003,Drug Discovery World 4 :35-48 中公开的位点。

[0110] 具有纤维素分解增强活性的多肽的来源

[0111] 本发明的具有纤维素分解增强活性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”,意思是多核苷酸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多核苷酸的菌株产生。在一个方面,获得自给定来源的多肽是胞外分泌的。

[0112] 本发明具有纤维素分解增强活性的多肽可以是细菌多肽。例如,所述多肽可以是具有纤维素分解增强活性的革兰氏阳性细菌多肽例如芽孢杆菌属 (*Bacillus*)、梭菌属 (*Clostridium*)、肠球菌属 (*Enterococcus*)、地芽孢杆菌属 (*Geobacillus*)、乳杆菌属 (*Lactobacillus*)、乳球菌属 (*Lactococcus*)、海洋芽孢杆菌属 (*Oceanobacillus*)、葡萄球菌属 (*Staphylococcus*)、链球菌属 (*Streptococcus*) 或链霉菌属 (*Streptomyces*) 多肽;或具有纤维素分解增强活性的革兰氏阴性细菌多肽,如弯曲杆菌属 (*Campylobacter*)、大肠杆菌 (*E. coli*)、黄杆菌属 (*Flavobacterium*)、梭杆菌属 (*Fusobacterium*)、螺杆菌属 (*Helicobacter*)、泥杆菌属 (*Ilyobacter*)、奈瑟氏菌属 (*Neisseria*)、假单胞菌属 (*Pseudomonas*)、沙门氏菌属 (*Salmonella*) 或脲原体属 (*Ureaplasma*) 属多肽,其具有纤维素分解增强活性。

[0113] 在一个方面,所述多肽是嗜碱芽孢杆菌 (*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*)、克劳氏芽孢杆菌 (*Bacillus clausii*)、凝结芽孢杆菌 (*Bacillus coagulans*)、坚强芽孢杆菌 (*Bacillus firmus*)、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*)、短小芽孢杆菌 (*Bacillus pumilus*)、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。

[0114] 在另一个方面,所述多肽是似马链球菌 (*Streptococcus equisimilis*)、酿脓链球菌 (*Streptococcus pyogenes*)、乳房链球菌 (*Streptococcus uberis*) 或马链球菌兽瘟亚种 (*Streptococcus equi* subsp. *Zooepidemicus*) 多肽。

[0115] 在另一个方面,所述多肽是不产色链霉菌 (*Streptomyces achromogenes*)、除虫链霉菌 (*Streptomyces avermitilis*)、天蓝链霉菌 (*Streptomyces coelicolor*)、灰色链霉菌 (*Streptomyces griseus*) 或浅青紫链霉菌 (*Streptomyces lividans*) 多肽。

[0116] 所述多肽也可以是真菌多肽。例如,所述多肽可为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*) 或西洋蓍霉属 (*Yarrowia*) 多肽;或丝状真菌多肽如枝顶孢霉属 (*Acremonium*)、伞菌属 (*Agaricus*)、链格孢属 (*Alternaria*)、曲霉属 (*Aspergillus*)、短梗霉属 (*Aureobasidium*)、*Botryosphaeria*、拟蜡菌属 (*Ceriporiopsis*)、毛喙壳属 (*Chaetomidium*)、金孢子菌属 (*Chrysosporium*)、*Claviceps*、*Cochliobolus*、鬼伞属 (*Coprinopsis*)、*Coptotermes*、棒囊壳属 (*Corynascus*)、隐丛赤壳菌属 (*Cryphonectria*)、隐球菌属 (*Cryptococcus*)、色二孢属 (*Diplodia*)、黑



耳属 (*Exidia*)、*Filibasidium*、镰孢属 (*Fusarium*)、赤霉属 (*Gibberella*)、全鞭毛虫属 (*Holomastigotoides*)、腐质霉属 (*Humicola*)、耙齿菌属 (*Irpex*)、蘑菇属 (*Lentinula*)、*Leptosphaeria*、梨孢菌属 (*Magnaporthe*)、*Melanocarpus*、多孔菌属 (*Meripilus*)、毛霉属 (*Mucor*)、毁丝霉属 (*Myceliophthora*)、新考玛脂霉属 (*Neocallimastix*)、脉孢菌属 (*Neurospora*)、拟青霉属 (*Paecilomyces*)、平革菌属 (*Phanerochaete*)、瘤胃壶菌属 (*Piromyces*)、*Poitrasia*、假黑盘菌属 (*Pseudoplectania*)、*Pseudotriconympha*、根毛霉属 (*Rhizomucor*)、裂褶菌属 (*Schizophyllum*)、柱顶孢属 (*Scytalidium*)、踝节菌属 (*Talaromyces*)、嗜热子囊菌属 (*Thermoascus*)、梭孢壳属 (*Thielavia*)、弯颈霉属 (*Tolyposcladium*)、木霉属 (*Trichoderma*)、长毛盘菌属 (*Trichophaea*)、轮枝孢属 (*Verticillium*)、包脚菇属 (*Volvariella*) 或炭角菌属 (*Xylaria*) 多肽。

[0117] 在另一个方面,所述多肽是青霉属 (*Penicillium*) 多肽。

[0118] 在另一个方面,所述多肽是卡尔酵母 (*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。

[0119] 在另一个方面,所述多肽是解纤维枝顶孢霉 (*Acremonium cellulolyticus*)、棘孢曲霉 (*Aspergillus aculeatus*)、泡盛曲霉 (*Aspergillus awamori*)、臭曲霉 (*Aspergillus foetidus*)、烟曲霉 (*Aspergillus fumigatus*)、日本曲霉 (*Aspergillus japonicus*)、构巢曲霉 (*Aspergillus nidulans*)、黑曲霉 (*Aspergillus niger*)、米曲霉 (*Aspergillus oryzae*)、*Chrysosporium inops*、嗜角质金孢子菌 (*Chrysosporium keratinophilum*)、*Chrysosporium lucknowense*、*Chrysosporium merdarium*、毡金孢子菌 (*Chrysosporium pannicola*)、*Chrysosporium queenslandicum*、热带金孢子菌 (*Chrysosporium tropicum*)、*Chrysosporium zonatum*、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 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 sarcochromum*)、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*)、灰腐质霉 (*Humicola grisea*)、特异腐质霉 (*Humicola insolens*)、疏棉状腐质霉 (*Humicola lanuginosa*)、白耙齿菌 (*Irpex lacteus*)、米黑毛霉 (*Mucor miehei*)、嗜热毁丝霉 (*Myceliophthora thermophila*)、粗糙脉孢菌 (*Neurospora crassa*)、绳状青霉 (*Penicillium funiculosum*)、产紫青霉 (*Penicillium purpurogenum*)、黄孢平革菌 (*Phanerochaete chrysosporium*)、无色梭孢壳 (*Thielavia achromatica*)、*Thielavia albomyces*、*Thielavia albopilosa*、澳洲梭孢壳 (*Thielavia australeinsis*)、*Thielavia fimeti*、小孢梭孢壳 (*Thielavia microspora*)、卵孢梭孢壳 (*Thielavia ovispora*)、*Thielavia peruviana*、毛梭孢壳 (*Thielavia setosa*)、瘤孢梭孢壳 (*Thielavia spededonium*)、*Thielavia subthermophila*、土生梭孢霉 (*Thielavia terrestris*)、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。

[0120] 在另一个方面,所述多肽是嗜松青霉多肽。

[0121] 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),而无论它们已知的种名。本领域熟练技术人员将容易地识别适合的等同物的同一性。

[0122] 这些种的菌株在许多培养物保藏中心对于公众能够容易地取得,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)。

[0123] 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所述多肽。用于从天然生境 (habitat) 分离微生物的技术是本领域内公知的。然后可通过相似地筛选另一种微生物的基因组或 cDNA 文库或混合 DNA 样品来获得编码所述多肽的多核苷酸。一旦用所述探针检测到编码多肽的多核苷酸,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 等,1989, 见上文)。

[0124] 多核苷酸

[0125] 本发明还涉及编码本发明的多肽的分离的多核苷酸。

[0126] 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 克隆多核苷酸。参见,例如, Innis 等,1990, PCR: A Guide to Methods and Application, Academic Press, New York。可以使用其它核酸扩增方法,如连接酶链式反应 (LCR)、连接活化转录 (ligation activated transcription; LAT) 和基于多核苷酸的扩增 (NASBA)。可以从青霉属菌株,或其它或相关生物体克隆多核苷酸,并且因此可为例如所述多核苷酸的多肽编码区的等位基因变体或种变体 (species variant)。

[0127] 本发明还涉及分离的多核苷酸,其包含如下的多核苷酸或由其组成:与 SEQ ID NO:1 的成熟多肽编码序列或其 cDNA 具有至少 80%,例如至少 85%,至少 90%,至少 91%,至少 92%,至少 93%,至少 94%,至少 95%,至少 96%,至少 97%,至少 98%,至少 99% 或 100% 的序列同一性程度,其编码具有纤维素分解增强活性的多肽。

[0128] 修饰编码本发明多肽的多核苷酸对于合成与所述多肽基本上相似的多肽可为必需的。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 等方面不同的变体。可以在作为 SEQ ID NO:1 的多肽编码序列或其 cDNA 序列存在的多核苷酸,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:所述取代不导

致多肽的氨基酸序列变化,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 等,1991, Protein Expression and Purification 2:95-107。

[0129] 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高严格条件或非常高的严格条件下,与以下序列杂交:(i)SEQ ID NO:1的成熟多肽编码序列,(ii)包含于SEQ ID NO:1的成熟多肽编码序列中的cDNA序列,或(iii)(i)或(ii)的全长互补链;或它们的等位变体和亚序列(Sambrook等,1989,见上文),如本文所定义的。

[0130] 在一个方面,所述多核苷酸包含SEQ ID NO:1,SEQ ID NO:1的成熟多肽编码序列或包含在大肠杆菌DSM 22711的质粒pGEM-T-Ppin7中含有的序列,或编码SEQ ID NO:2具有纤维素分解增强活性的片段的SEQ ID NO:1的亚序列,或所述多核苷酸由上述组成。

[0131] 核酸构建体

[0132] 本发明还涉及包含本发明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(几个)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。

[0133] 可以用许多方式操作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是本领域熟知的。

[0134] 调控序列可以是启动子序列,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多核苷酸。启动子序列含有介导多肽的表达的转录调控序列。启动子可以在所选的宿主细胞中显示转录活性的任何多核苷酸,包括突变的、截短的和杂合的启动子,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。

[0135] 用于指导本发明的核酸构建体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:解淀粉芽孢杆菌 $\alpha$ -淀粉酶基因(amyQ)、地衣芽孢杆菌 $\alpha$ -淀粉酶基因(amyL)、地衣芽孢杆菌青霉素酶基因(penP)、嗜热脂肪芽孢杆菌产麦芽淀粉酶基因(amyM)、枯草芽孢杆菌果聚糖蔗糖酶基因(sacB)、枯草芽孢杆菌xylA和xylB基因、大肠杆菌lac操纵子、天蓝链霉菌琼脂糖酶基因(dagA)和原核 $\beta$ -内酰胺酶基因(Villa-Kamaroff等,1978, Proc. Natl. Acad. Sci. USA 75:3727-3731),以及tac启动子(DeBoer等,1983, Proc. Natl. Acad. Sci. USA 80:21-25)。另外的启动子描述于"Useful proteins from recombinant bacteria" Gilbert等, Scientific American, 1980, 242: 74-94;和 Sambrook等, 1989, 见上文。

[0136] 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:构巢曲霉乙酰胺酶、黑曲霉中性 $\alpha$ -淀粉酶、黑曲霉酸稳定性 $\alpha$ -淀粉酶、黑曲霉或泡盛曲霉葡糖淀粉酶(glaA)、米曲霉TAKA淀粉酶、米曲霉碱性蛋白酶、米曲霉丙糖磷酸异构酶、尖镰孢胰蛋白酶样蛋白酶(WO 96/00787)、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(WO 00/56900)、曼赫根毛霉脂肪酶、曼赫根毛霉(Rhizomucor miehei)天冬氨酸蛋白酶、里氏木霉 $\beta$ -葡糖苷酶、里氏木霉纤维二糖水解脱酶I、里氏木霉纤维二糖水解脱酶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 $\beta$ -木糖苷酶,以及

NA2-tpi 启动子（一种修饰的启动子，其来自在曲霉属中编码中性  $\alpha$ -淀粉酶的基因，其中未翻译的前导序列由在曲霉属中编码丙糖磷酸异构酶的基因的未翻译的前导序列所替代；非限制性实例包括修饰的启动子，其来自在黑曲霉中编码中性  $\alpha$ -淀粉酶的基因，其中未翻译的前导序列由在构巢曲霉和米曲霉中编码丙糖磷酸异构酶的基因的未翻译的前导序列所替代）；和它们的突变的、截短的和杂合的启动子。

[0137] 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 等, 1992, *Yeast* 8 :423-488 描述。

[0138] 调控序列也可以是合适的转录终止子序列，其由宿主细胞识别以终止转录。所述终止子序列与编码所述多肽的多核苷酸的 3' 末端可操作地连接。可以将所选宿主细胞中有功能的任何终止子用在本发明中。

[0139] 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：构巢曲霉邻氨基苯甲酸合酶、黑曲霉葡糖淀粉酶、黑曲霉  $\alpha$ -葡糖苷酶、米曲霉 TAKA 淀粉酶和尖镰孢胰蛋白酶样蛋白酶。

[0140] 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 等, 1992, 见上文描述。

[0141] 调控序列还可以是合适的前导序列，当转录时其为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多核苷酸的 5' 端。可以使用在所选宿主细胞中有功能的任何前导序列。

[0142] 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。

[0143] 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  $\alpha$  因子和酿酒酵母醇脱氢酶 / 甘油醛 -3- 磷酸脱氢酶 (ADH2/GAP)。

[0144] 调控序列也可以是聚腺苷酸化序列，其是与多核苷酸的 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。可以使用在所选宿主细胞中有功能的任何聚腺苷酸化序列。

[0145] 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  $\alpha$ -葡糖苷酶。

[0146] 对于酵母宿主细胞有用的聚腺苷酸化序列由 Guo 和 Sherman, 1995, *Mol. Cellular Biol.* 15 :5983-5990 描述。

[0147] 调控序列还可为信号肽编码区，其编码与多肽的 N 端相连的信号肽，并且指导多肽进入细胞分泌途径。多核苷酸的编码序列 5' 端可固有地包含信号肽编码区，其与编码多肽的编码序列的区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异源的信号肽编码序列。异源信号肽编码序列在编码序列不天然地含

有信号肽编码序列时可为必需的。或者,外源信号肽编码序列可以简单地取代天然信号肽编码序列以增强多肽的分泌。然而,可使用指导表达的多肽进入所选宿主细胞的分泌途径的任何信号肽编码序列。

[0148] 对于细菌宿主细胞有效的信号肽编码序列是从如下酶的基因获得的信号肽编码序列:芽孢杆菌属 NCIB 11837 产麦芽糖淀粉酶、地衣芽孢杆菌枯草杆菌蛋白酶 (subtilisin)、地衣芽孢杆菌  $\beta$ -内酰胺酶、嗜热脂肪芽孢杆菌  $\alpha$ -淀粉酶、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 和 Palva, 1993, *Microbiological Reviews* 57:109-137 描述。

[0149] 对于丝状真菌宿主细胞有效的信号肽编码序列是从如下酶的基因获得的信号肽编码序列:黑曲霉中性淀粉酶、黑曲霉葡糖淀粉酶、米曲霉 TAKA 淀粉酶、特异腐质霉纤维素酶、特异腐质霉内切葡聚糖酶 V、疏棉状腐质霉脂肪酶和曼赫根毛霉天冬氨酸蛋白酶。

[0150] 对于酵母宿主细胞有用的信号肽从酿酒酵母  $\alpha$  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序列由 Romanos 等, 1992, 见上文描述。

[0151] 调控序列还可以是前肽编码序列,其编码位于多肽 N 端的前肽。所得多肽称为酶原 (proenzyme) 或前多肽 (propolypeptide) (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为活性多肽。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)、嗜热毁丝霉漆酶 (WO 95/33836)、曼赫根毛霉天冬氨酸蛋白酶和酿酒酵母  $\alpha$  因子的基因获得前肽编码序列。

[0152] 当信号肽和前肽区二者均出现在多肽的 N 端时,将前肽区置于紧接着 (next to) 多肽 N 端,并且将信号肽区置于紧接着前肽区的 N 端。

[0153] 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,可以使用黑曲霉葡糖淀粉酶启动子、TAKA  $\alpha$ -淀粉酶启动子和米曲霉葡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,编码多肽的多核苷酸将与调节序列可操作地连接。

[0154] 表达载体

[0155] 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。多种核苷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 (几个) 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多核苷酸。可供选择的是,所述多核苷酸可以通过在适当的用于表达的载体中插入包含所述序列的多核苷酸或核酸构建体来表达。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适当的表达调控序列可操作地连接。

[0156] 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,并且能够产生多核苷酸的表达。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。

[0157] 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)。

[0158] 所述载体优选地含有一个或多个 (几个) 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。

[0159] 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、氯霉素、卡那霉素或四环素抗性。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'-phosphate decarboxylase)、*sC* (硫酸腺苷酰转移酶) 和 *trpC* (邻氨基苯甲酸合酶 (anthranilate synthase)) 以及它们的等同物。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。

[0160] 所述载体优选含有元件,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。

[0161] 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多核苷酸,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,整合元件应含有足够数量的核酸,如 100 至 10,000 碱基对,400 至 10,000 碱基对,和 800 至 10,000 碱基对,其与相应的目标序列具有高度序列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多核苷酸。另一方面,可以将载体通过非同源重组整合到宿主细胞的基因组中。

[0162] 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”意指能够使质粒或载体体内复制的多核苷酸。

[0163] 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 $\beta$ 1 的复制起点。

[0164] 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。

[0165] 在丝状真菌细胞中有用的复制起点的实例是 AMA1 和 ANS1 (Gems 等,1991, Gene 98 :61-67 ;Cullen 等,1987, Nucleic Acids Res. 15 :9163-9175 ;W000/24883)。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。

[0166] 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，或将可扩增的选择性标记基因包括于多核苷酸，其中可通过在合适的选择剂 (selectable agent) 存在下培养细胞来选择含有选择性标记基因的扩增拷贝，且由此含有多核苷酸的额外拷贝的细胞。

[0167] 用于连接上述元件以构建本发明的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 等, 1989, 见上文)。

[0168] 宿主细胞

[0169] 本发明还涉及重组宿主细胞, 其包含可操作地连接于一种或多种指导本发明多肽产生的调控序列的本发明的多核苷酸。将包含多核苷酸的构建体或载体导入宿主细胞, 使所述构建体或载体如前所述作为染色体整体或者作为自复制的染色体外载体维持。术语“宿主细胞”包括亲本细胞的任何后代, 其由于复制过程中发生的突变而不同于亲本细胞。宿主细胞的选择将在很大程度上依赖于编码多肽的基因及其来源。

[0170] 宿主细胞可以是在本发明的多肽的重组产生中有用的任何细胞, 例如, 原核或真核细胞。

[0171] 原核宿主细胞可以是任何革兰氏阳性细菌或革兰氏阴性细菌。革兰氏阳性细菌包括但不限于, 芽孢杆菌属、梭菌属、肠球菌属、地芽孢杆菌属、乳杆菌属、乳球菌属、海洋芽孢杆菌属、葡萄球菌属、链球菌属和链霉菌属。革兰氏阴性细菌包括但不限于, 弯曲杆菌属、大肠杆菌、黄杆菌属、梭杆菌属、螺杆菌属、泥杆菌属、奈瑟氏菌属、假单胞菌属、沙门氏菌属和脉原体属。

[0172] 细菌宿主细胞可以是任何芽孢杆菌属细胞, 包括但不限于, 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、凝结芽孢杆菌、坚强芽孢杆菌、灿烂芽孢杆菌、迟缓芽孢杆菌、地衣芽孢杆菌、巨大芽孢杆菌、短小芽孢杆菌、嗜热脂肪芽孢杆菌、枯草芽孢杆菌和苏云金芽孢杆菌细胞。

[0173] 细菌宿主细胞还可以是任何链球菌属细胞, 包括但不限于, 似马链球菌、酿脓链球菌、乳房链球菌和马链球菌兽瘟亚种细胞。

[0174] 细菌宿主细胞还可以是任何链霉菌属细胞, 包括但不限于, 不产色链霉菌、除虫链霉菌、天蓝链霉菌、灰色链霉菌和浅青紫链霉菌细胞。

[0175] 可通过如下方法实现将 DNA 引入到芽孢杆菌属细胞: 例如原生质体转化 (参见, 例如, Chang 和 Cohen, 1979, Mol. Gen. Genet. 168 :111-115), 使用感受态细胞 (参见, 例如, Young 和 Spizizen, 1961, J. Bacterio. 81 :823-829 或 Dubnau 和 Davidoff-Abelson, 1971, J. Mol. Biol. 56 :209-221), 电穿孔 (参见, 例如, Shigekawa 和 Dower, 1988, Biotechniques 6 :742-751) 或接合 (参见, 例如, Koehler 和 Thorne, 1987, J. Bacteriol. 169 :5771-5278)。可通过如下方法实现将 DNA 引入到大肠杆菌细胞: 例如原生质体转化 (参见, 例如, Hanahan, 1983, J. Mol. Biol. 166 :557-580) 或电穿孔 (参见, 例如, Dower 等, 1988, Nucleic Acids Res. 16 :6127-6145)。可通过如下方法实现将 DNA 引入到链霉菌属细胞: 例如原生质体转化和电穿孔 (参见, 例如, Gong 等, 2004, Folia Microbiol. (Praha) 49 :399-405), 接合 (参见, 例如, Mazodier 等, 1989, J. Bacteriol. 171 :3583-3585), 或转导 (参见, 例如,

Burke 等, 2001, Proc. Natl. Acad. Sci. USA 98 :6289-6294)。可通过如下方法实现将 DNA 引入到假单胞菌属细胞:例如电穿孔(参见,例如, Choi 等, 2006, J. Microbiol. Methods 64 :391-397)或接合(参见,例如, Pinedo 和 Smets, 2005, Appl. Environ. Microbiol. 71 :51-57)。可通过如下方法实现将 DNA 引入到链球菌属细胞:例如天然感受态 (natural competence) (参见,例如, Perry 和 Kuramitsu, 1981, Infect. Immun. 32 :1295-1297), 原生质体转化(参见,例如, Catt 和 Jollick, 1991, Microbios. 68 :189-207), 电穿孔(参见,例如, Buckley 等, 1999, Appl. Environ. Microbiol. 65 :3800-3804)或接合(参见,例如, Clewell, 1981, Microbiol. Rev. 45 :409-436)。然而,可以使用本领域已知的将 DNA 引入宿主细胞的任何方法。

[0176] 宿主细胞还可以是真核生物,如哺乳动物、昆虫、植物或真菌细胞。

[0177] 所述宿主细胞可为真菌细胞。“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 等, 于 Ainsworth and Bisby's Dictionary of The Fungi, 第 8 版, 1995, CAB International, University Press, Cambridge, UK 中所定义) 以及卵菌门 (Oomycota) (如 Hawksworth 等, 1995, 见上, 171 页中所引用), 和所有有丝分裂孢子真菌 (mitosporic fungi) (Hawksworth 等, 1995, 见上文)。

[0178] 真菌宿主细胞可为酵母细胞。“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., 和 Davenport, R. R. 编, Soc. App. Bacteriol. Symposium Series No. 9, 1980) 中所述。

[0179] 所述酵母宿主细胞可为假丝酵母属、汉逊酵母属 (Hansenula)、克鲁维酵母属、毕赤酵母属、酵母属、裂殖酵母属或西洋薯霉属细胞,如卡尔酵母、酿酒酵母、糖化酵母、道格拉氏酵母、克鲁弗酵母、诺地酵母、卵形酵母或解脂西洋薯霉 (*Yarrowia lipolytica*) 细胞。

[0180] 所述真菌宿主细胞可为丝状真菌细胞。“丝状真菌”包括真菌门 (Eumycota) 和卵菌门的亚门(如由 Hawksworth 等, 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。

[0181] 所述丝状真菌宿主细胞可为枝顶孢霉属、曲霉属、短梗霉属、烟管霉属 (Bjerkandera)、拟蜡菌属、金孢子菌属、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、毛霉属、毁丝霉属、新考玛脂霉属、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、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、栓菌属 (Trametes) 或木霉属细胞。

[0182] 例如,所述丝状真菌宿主细胞可为泡盛曲霉、臭曲霉、烟曲霉、日本曲霉、构巢曲霉、黑曲霉、米曲霉、黑刺烟管菌 (*Bjerkandera adusta*)、干拟蜡菌 (*Ceriporiopsis*



aneirina)、*Ceriporiopsis caregiea*、*Ceriporiopsis gilvescens*、*Ceriporiopsis pannocinta*、*Ceriporiopsis rivulosa*、*Ceriporiopsis subrufa*、虫拟蜡菌 (*Ceriporiopsis subvermispota*)、*Chrysosporium inops*、嗜角质金孢子菌、*Chrysosporium lucknowense*、*Chrysosporium merdarium*、毡金孢子菌、*Chrysosporium queenslandicum*、热带金孢子菌、*Chrysosporium zonatum*、灰盖鬼伞 (*Coprinus cinereus*)、毛革盖菌 (*Coriolus hirsutus*)、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、硫色镰孢、圆镰孢、拟丝孢镰孢、镶片镰孢细胞、特异腐质霉、疏棉状腐质霉、米黑毛霉、嗜热毁丝霉、粗糙脉孢菌、产紫青霉、黄孢平革菌 (*Phanerochaete chrysosporium*)、辐射射脉菌 (*Phlebia radiata*)、刺芹侧耳 (*Pleurotus eryngii*)、土生梭孢霉、长绒毛栓菌 (*Trametes villosa*)、变色栓菌 (*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。

[0183] 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、Yelton 等, 1984, *Proc. Natl. Acad. Sci. USA* 81:1470-1474 以及 Christensen 等, 1988, *Bio/Technology* 6:1419-1422 中描述。用于转化镰孢属菌种的合适方法由 Malardier 等, 1989, *Gene* 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 和 Guarente, 于 Abelson, J. N. 和 Simon, M. I. 编, *Guide to Yeast Genetics and Molecular Biology, Methods in Enzymology, Volume 194*, pp 182-187, Academic Press, Inc., New York; Ito 等, 1983, *J. Bacteriol.* 153:163; 和 Hinnen 等, 1978, *Proc. Natl. Acad. Sci. USA* 75:1920。

[0184] 产生方法

[0185] 本发明还涉及用于产生本发明多肽的方法, 其包括: (a) 在有助于产生多肽的条件下培养细胞, 所述细胞以其野生型形式产生所述多肽; 和 (b) 回收所述多肽。在一个优选的方面, 所述细胞是青霉属细胞。在更优选的方面, 所述细胞是嗜松青霉。在最优选的方面, 所述细胞是嗜松青霉 NN46877。

[0186] 本发明还涉及用于产生本发明的多肽的方法, 其包括: (a) 在有助于产生多肽的条件下培养本发明的重组宿主细胞; 和 (b) 回收所述多肽。

[0187] 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宿主细胞。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开的组成制备 (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, 其能够从细胞裂解物 (lysate) 回收。

[0188] 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, 酶测定法 (enzyme assay) 可用于测定多肽活性。

[0189] 所述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。

[0190] 所述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 和 Lars Ryden 编, VCH Publishers, New York, 1989)。

[0191] 在可供替代的方面,不回收所述多肽,而是将表达所述多肽的本发明的宿主细胞用作多肽来源。

[0192] 植物

[0193] 本发明还涉及分离的植物,例如,转基因植物、植物部分或植物细胞,其包含本发明分离的多核苷酸,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。

[0194] 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(翦股颖属);和谷类,例如,小麦、燕麦、黑麦、大麦、稻(rice)、高粱和玉蜀黍(maize)(玉米)。

[0195] 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(pea),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)。

[0196] 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(parenchyme)、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)。

[0197] 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后代。

[0198] 表达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:将编码多肽的一个或多个(几个)表达构建体并入植物宿主基因组或叶绿体基因组,并且将所得的修饰植物或植物细胞繁殖为转基因植物或植物细胞。

[0199] 表达构建体便利地是包含编码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所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）。

[0200] 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 等,1988,Plant Physiology 86 : 506 所述。

[0201] 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 等,1980,Cell 21 :285-294,Christensen 等,1992,Plant Mo. Biol. 18 :675-689 ;Zhang 等,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 和 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 等,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. Cell Physiol. 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, J. Plant Physiol. 152 : 708-711)、来自种子油体蛋白 (oil body protein) 的启动子 (Chen 等,1998,Plant Cell Physiol. 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 (Kyojuka 等,1993,Plant Physiol. 102 :991-1000),小球藻病毒 (*chlorella virus*) 腺嘌呤甲基转移酶 (adenine methyltransferase) 基因启动子 (Mitra 和 Higgins,1994,Plant Biol. 26 :85-93),或来自稻的 aldP 基因启动子 (Kagaya 等,1995,Molecular Gen. Genet. 248 :668-674),或伤口诱导的启动子,如马铃薯 pin2 启动子 (Xu 等,1993,Plant Mol. Biol. 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,或通过外源施加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) 和赤霉酸 (gibberellic acid),和重金属。

[0202] 启动子增强子元件也可以用于实现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,其置于启动子和编码多肽的多核苷酸之间。例如 Xu 等,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。

[0203] 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。

[0204] 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 等,1990, Science 244 :1293 ;Potrykus,1990, Bio/Technology 8 :535 ;Shimamoto 等,1989, Nature 338 :274)。

[0205] 目前,根癌土壤杆菌 (*Agrobacterium tumefaciens*) 介导的基因转移 (gene transfer),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 Mol. Biol. 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, *Curr. Opin. Biotechnol.* 5 :158-162 ;Vasil 等, 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 Mol. Biol.* 21 :415-428 所描述的。其他依照本公开使用的转化方法包括那些描述于美国专利号 6, 395, 966 和 7, 151, 204 中的方法 (两者均通过全文提述并入本文)。

[0206] 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, 使用带有两个独立的 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。

[0207] 除了直接用根据本发明制备的构建体直接转化具体植物基因型之外, 还可通过将具有构建体的植物与缺乏该构建体的第二植物杂交来制备转基因植物。举例而言, 可将编码多肽的构建体或其部分通过杂交而引入特定植物品种, 而根本无需直接转化该给定品种的植物。因此, 本发明不仅涵盖从依照本发明经转化的细胞直接再生的植物, 还涵盖此类植物的后代 (progeny)。如用于本文, 后裔可指依照本发明制备的亲本植物任何世代的后裔 (offspring)。此种后裔可包含依据本发明制备的 DNA 构建体, 或依据本发明制备的 DNA 构建体的一部分。杂交导致转基因通过将起始种系与供体植物种系交叉授粉而引入植物种系。此类步骤的非限制性实例进一步阐述于美国专利 7, 151, 204 号。

[0208] 植物可通过回交转化方法生成。举例而言, 植物包括称作回交转化的基因型、种系、近交体 (inbred) 或杂合体 (hybrid) 的植物。

[0209] 可使用遗传标记以协助本发明的一种或多种转基因从一个遗传背景基因渗入 (introgression) 至另一个。标记协助的选择提供了相对于常规育种的优点, 在于其可用于避免由表型变异导致的错误。进一步, 遗传标记可在特定杂交的个体后裔中提供有关良种种质相对程度的数据。举例而言, 当本不 (otherwise) 具有非农艺学所需的遗传背景但具有所需性状的植物与良种亲本杂交时, 可使用遗传标记来选择不仅具有目标性状, 还具有相对较大比例所需种质的后裔。以此方式, 使一种或多种性状基因渗入特定遗传背景所需的世代数得到最小化。

[0210] 本发明还涉及产生本发明的多肽的方法, 包括: (a) 在有助于所述多肽产生的条件下, 培养包含编码所述多肽的多核苷酸的转基因植物或植物细胞; 和 (b) 回收所述多肽。

[0211] 去除或减少纤维素分解增强活性

[0212] 本发明还涉及用于产生亲本细胞突变体的方法, 其包括破坏或缺失编码本发明的多肽的多核苷酸或其部分, 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。

[0213] 可以使用本领域熟知的方法 (例如, 插入、破坏、替代或缺失) 通过减少或消除所述多核苷酸的表达来构建突变细胞。在一个优选的方面, 所述多核苷酸是失活的。待修饰或失活的多核苷酸可以是, 例如, 编码区或其对活性关键的部分, 或表达编码区所需的调节元件。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, 足以影响多核苷酸表达的部分。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。

[0214] 可以通过向亲本细胞施以诱变,并且选择其中已将所述多核苷酸的表达减少或消除的突变细胞来进行多核苷酸的修饰或失活。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。

[0215] 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。

[0216] 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温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。

[0217] 通过导入、取代或去除基因中的一个或多个 (几个) 核苷酸或其转录或翻译所需的调节元件可以实现所述多核苷酸的修饰或失活。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,直接在表达待修饰多核苷酸的细胞上进行,但优选如下面所示例的那样在体外进行所述修饰。

[0218] 消除或减少多核苷酸表达的便利方式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,将相应于内源多核苷酸的核酸序列在体外进行诱变以产生缺陷性的核酸序列,然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多核苷酸。理想的是所述缺陷性多核苷酸还编码标记,其可用于选择其中多核苷酸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) 来破坏所述多核苷酸。

[0219] 本发明还涉及抑制细胞中具有纤维素分解增强活性的多肽的表达的方法,包括对所述细胞施用或在所述细胞中表达双链 RNA (dsRNA) 分子,其中所述 dsRNA 包含本发明多核苷酸的亚序列。在一个优选的方面,所述 dsRNA 长度为约 15、16、17、18、19、20、21、22、23、24、25 或更多个双链体核苷酸。

[0220] 所述 dsRNA 优选为小干扰 RNA (siRNA) 或微 RNA (miRNA)。在一个优选的方面,所述 dsRNA 是用于抑制转录的小干扰 RNA (siRNA)。在另一个优选的方面,所述 dsRNA 是用于抑制翻译的微 RNA (miRNA)。

[0221] 本发明还涉及此类双链 RNA (dsRNA) 分子,其包含 SEQ ID NO :1 的成熟多肽编码序列的一部分以供抑制所述多肽在细胞中的表达。尽管本发明不限于任何作用机制,但所述 dsRNA 可进入细胞并导致相似或相同序列的单链 RNA (ssRNA) 包括内源 mRNA 的降解。当细胞暴露于 dsRNA 时,来自同源基因的 mRNA 通过称为 RNA 干扰 (RNAi) 的过程被选择性降解。

[0222] 本发明的 dsRNA 可用于基因静默。在一个方面,本发明提供了使用本发明的 dsRNAi 选择性降解 RNA 的方法。所述方法可在体外、离体或体内进行。在一个方面,所述 dsRNA 可用于在细胞、器官或动物中生成功能缺失突变。制备和使用 dsRNA 分子以选择性降解 RNA 的方法在本领域是公知的:参见例如美国专利号 6,489,127 ;6,506,559 ;6,511,824 ;和 6,515,109。

[0223] 本发明进一步涉及亲本细胞的突变体细胞,其包含编码多肽的多核苷酸或其调控

序列的破坏或缺失,或者静默的编码多肽的基因,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。

[0224] 多肽缺陷型突变细胞作为表达天然和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天然或异源多肽的方法,其包括:(a) 在有助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”意指对宿主细胞为非天然的多肽,例如天然蛋白的变体。所述宿主细胞可包含多于一个拷贝的编码所述天然或异源多肽的多核苷酸。

[0225] 用于培养和纯化感兴趣的产物的方法可通过本领域已知的方法进行。

[0226] 本发明用于产生基本上无纤维素分解增强活性的产物的方法在真核多肽,特别是真菌蛋白质例如酶的产生中是特别令人有兴趣的。纤维素分解增强-缺陷细胞也可用于表达在制药上感兴趣的异源蛋白质如激素、生长因子、受体等。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、缺失或添加或其它这样的修饰而被修饰以增强活性、热稳定性、pH 耐受性等。

[0227] 在另外的方面,本发明涉及基本上无纤维素分解增强活性的蛋白质产物,其通过本发明的方法产生。

[0228] 组合物

[0229] 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”表示所述组合物的纤维素分解增强活性,例如,以至少 1.1 的富集因数(enrichment factor) 增加。

[0230] 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,所述组合物亦可包含一种或多种(几种)其他酶,如乙酰木聚糖酯酶、淀粉酶、 $\alpha$ -L-阿拉伯呋喃糖苷酶、过氧化氢酶、纤维二糖水解酶、纤维素酶、 $\alpha$ -半乳糖苷酶、 $\beta$ -半乳糖苷酶、内切葡聚糖酶、 $\alpha$ -葡糖醛酸糖苷酶、酯酶、阿魏酸酯酶、葡糖淀粉酶、 $\alpha$ -葡糖苷酶、 $\beta$ -葡糖苷酶、漆酶、脂肪酶、甘露糖苷酶、过氧化物酶、多酚氧化酶、木聚糖酶或  $\beta$ -木糖苷酶。

[0231] 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。可以依照本领域内已知方法将包括于所述组合物中的多肽稳定化。

[0232] 下面给出的是本发明多肽组合物的优选的用途的实例。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确定。

[0233] 用途

[0234] 本发明亦涉及下述使用具有纤维素分解增强活性的多肽或其组合物的方法。

[0235] 本发明还涉及降解或转化纤维素材料的方法,其包括:在存在本发明的具有纤维素分解增强活性的多肽的情况下,用酶组合物处理纤维素材料。在一个优选方面,所述方法还包括回收已降解或转化的纤维素材料。

[0236] 本发明还涉及产生发酵产物的方法,其包括:(a) 在存在本发明的具有纤维素分解增强活性的多肽的条件下,用酶组合物糖化纤维素材料;(b) 用一种或多种(几种)发酵微生物发酵经糖化的纤维素材料以产生发酵产物;和(c) 从发酵回收发酵产物。

[0237] 本发明还涉及发酵纤维素材料的方法,其包括:用一种或多种(几种)发酵微生物发酵纤维素材料,其中在存在具有纤维素分解增强活性的多肽的条件下,用酶组合物糖化

纤维素材料。在一个优选的方面,纤维素材料的发酵产生发酵产物。在另一个优选的方面,方法进一步包括从发酵回收发酵产物。

[0238] 本发明的方法可以用于将纤维素材料糖化成可发酵糖,并且将可发酵糖转化成很多有用的物质,例如燃料、饮用乙醇和/或发酵产物(例如酸、醇、酮、气体等)。从纤维素材料产生期望的发酵产物通常涉及预处理、酶水解(糖化)和发酵。

[0239] 根据本发明的纤维素材料的处理可以使用本领域的常规过程完成。此外,本发明的方法能使用经配置以依照发明操作的任何常规生物质处理设备进行处理。

[0240] 水解(糖化)和发酵,分别或同时,包括但不限于,分离的水解和发酵(SHF)、同步糖化和发酵(SSF)、同步糖化和共发酵(SSCF)、混合的水解和发酵(HHF)、分离的水解和共发酵(SHCF)、混合的水解和共发酵(HHCF),和直接微生物转化(DMC)。SHF使用分离的处理步骤以首先将纤维素材料酶水解为可发酵糖,例如,葡萄糖,纤维二糖、纤维三糖和戊糖,然后将可发酵糖发酵成为乙醇。在SSF中,纤维素材料的酶水解和糖变为乙醇的发酵在一个步骤中组合(Philippidis, G. P., 1996, Cellulose bioconversion technology, 于 Handbook on Bioethanol: Production and Utilization, Wyman, C. E 编, Taylor & Francis, Washington, DC, 179-212)。SSCF包括多种糖的共发酵(Sheehan, J., 和 Himmel, M., 1999, Enzymes, energy and the environment: A strategic perspective on the U. S. Department of Energy's research and development activities for bioethanol, Biotechnol. Prog. 15: 817-827)。HHF在同步糖化和水解步骤之外,还涉及单独的水解步骤,所述步骤可以在同一个反应器中进行。HHF过程中的步骤可以在不同的温度,即,高温酶糖化,然后在发酵菌株能够耐受的较低温度进行SSF。DMC在一个或多个(几个)步骤中组合了所有三个过程(酶产生、水解和发酵),其中使用相同的生物体产生用于将木素纤维素转化成可发酵糖并将可发酵糖转化成终产物的酶(Lynd, L. R., Weimer, P. J., van Zyl, W. H., 和 Pretorius, I. S., 2002, Microbial cellulose utilization: Fundamentals and biotechnology, Microbiol. Mol. Biol. Reviews 66: 506-577)。在本文可以理解的是,本领域中任何已知的方法,包括预处理、酶水解(糖化)、发酵,或它们的组合,可用于实施本发明的方法。

[0241] 常规设备包括补料批式搅拌反应器、批式搅拌反应器、具有超滤的连续流搅拌反应器和/或连续活塞流柱式反应器(Fernanda de Castilhos Corazza, Flávio Faria de Moraes, Gisella Maria Zanin 和 Ivo Neitzel, 2003, Optimal control in fed-batch reactor for the cellobiose hydrolysis, Acta Scientiarum. Technology 25: 33-38; Gusakov, A. V., 和 Sinitsyn, A. P., 1985, Kinetics of the enzymatic hydrolysis of cellulose: 1. A mathematical model for a batch reactor process, Enz. Microb. Technol. 7: 346-352)、研磨反应器(Ryu, S. K., 和 Lee, J. M., 1983, Bioconversion of waste cellulose by using an attrition bioreactor, Biotechnol. Bioeng. 25: 53-65), 或者具有由电磁场引起的强烈搅拌的反应器(Gusakov, A. V., Sinitsyn, A. P., Davydkin, I. Y., Davydkin, V. Y., Protas, O. V., 1996, Enhancement of enzymatic cellulose hydrolysis using a novel type of bioreactor with intensive stirring induced by electromagnetic field, Appl. Biochem. Biotechnol. 56: 141-153)。其它反应器类型包括: 流化床、升流层(upflow blanket)、固定化和用于水解和/或发酵的挤出机型的反应器。

[0242] 预处理。在本发明的方法的实施中,可以使用本领域已知的任何预处理过程破坏植物细胞壁的纤维素材料组分 (Chandra 等,2007, Substrate pretreatment: The key to effective enzymatic hydrolysis of lignocellulosics? *Adv. Biochem. Engin. / Biotechnol.* 108 :67-93 ;Galbe 和 Zacchi,2007, Pretreatment of lignocellulosic materials for efficient bioethanol production, *Adv. Biochem. Engin. / Biotechnol.* 108 :41-65 ;Hendriks 和 Zeeman,2009, Pretreatments to enhance the digestibility of lignocellulosic biomass, *Bioresource Technol.* 100 :10-18 ;Mosier 等,2005, Features of promising technologies for pretreatment of lignocellulosic biomass, *Bioresource Technol.* 96 :673-686 ;Taherzadeh 和 Karimi,2008, Pretreatment of lignocellulosic wastes to improve ethanol and biogas production: A review, *Int. J. of Mol. Sci.* 9 :1621-1651 ;Yang 和 Wyman,2008, Pretreatment: the key to unlocking low-cost cellulosic ethanol, *Biofuels Bioproducts and Biorefining-Biofpr.* 2 :26-40)。

[0243] 纤维素材料也可以在预处理之前使用本领域中已知的方法进行预浸泡、润湿、洗涤或调节。

[0244] 常规的预处理包括但不限于,蒸汽预处理(伴随或不伴随爆炸)、稀酸预处理、热水预处理、碱性预处理、石灰预处理、湿氧化、湿爆炸、氨纤维爆炸、有机溶剂预处理和生物预处理。其它预处理包括氨渗滤、超声、电穿孔、微波、超临界 CO<sub>2</sub>、超临界 H<sub>2</sub>O、臭氧和  $\gamma$  辐射预处理。

[0245] 可以在水解和 / 或发酵之前预处理纤维素材料。预处理优选在水解前进行。或者,预处理可以与酶水解同时进行以释放可发酵糖,如葡萄糖、木糖和 / 或纤维二糖。在大多数情况下,预处理步骤本身使一些生物质转化成可发酵糖(甚至在不存在酶的情况下)。

[0246] 蒸汽预处理。在蒸汽预处理中,加热纤维素材料以破坏植物细胞壁成分,包括木质素、半纤维素和纤维素,使酶可接触纤维素和其它部分,例如,半纤维素。将纤维素材料通过或穿过反应容器,其中注入蒸汽以增加温度至需要的温度和压力,并且在其中保持期望的反应时间。蒸汽预处理优选在 140-230 °C,更优选 160-200 °C,并且最优选 170-190 °C 进行,其中最优的温度范围依赖于任何化学催化剂的添加。蒸汽预处理的停留时间优选 1-15 分钟,更优选 3-12 分钟,并且最优选 4-10 分钟,其中最优的停留时间依赖于温度范围和任何化学催化剂的添加。蒸汽预处理允许相对较高的固体加载量,使纤维素材料在预处理过程中通常仅仅变得潮湿。蒸汽预处理经常与预处理后的物质的爆炸放料 (explosive discharge) 组合,这称为蒸汽爆炸,即,快速闪变至大气压和物质的湍流,以通过破碎增加可接触的表面积 (Duff 和 Murray,1996, *Bioresource Technology* 855 : 1-33 ;Galbe 和 Zacchi,2002, *Appl. Microbiol. Biotechnol.* 59 :618-628 ;美国专利申请 No. 20020164730)。在蒸汽预处理过程中,切割半纤维素乙酰基团,并且得到的酸自催化纤维素部分水解成单糖和寡糖。去除木质素仅至有限的程度。

[0247] 经常在蒸汽预处理之前加入催化剂如 H<sub>2</sub>SO<sub>4</sub> 或 SO<sub>2</sub> (通常 0.3 至 3% w/w),其可减少时间,降低温度,增加回收率,并改进酶水解 (Ballesteros 等,2006, *Appl. Biochem. Biotechnol.* 129-132 :496-508 ;Varga 等,2004, *Appl. Biochem. Biotechnol.* 113-116 : 509-523 ;Sassner 等.,2006, *Enzyme Microb. Technol.* 39 :756-762)。



[0248] 化学预处理:术语“化学处理”指能促进纤维素、半纤维素和 / 或木质素分离和 / 或释放的任何化学处理。合适的化学预处理步骤的实例包括例如稀酸预处理、石灰预处理、湿氧化、氨纤维 / 冷冻爆炸 (AFEX)、氨渗滤 (APR) 和有机溶剂预处理。

[0249] 在稀酸预处理中,将纤维素材料与稀酸 (通常是  $H_2SO_4$ ) 和水混合以形成浆料,由蒸汽加热至期望的温度,并在一段停留时间后闪变至大气压。可以用很多反应器设计进行稀酸预处理,例如,活塞流式反应器、逆流反应器或连续逆流收缩床反应器 (Duff 和 Murray, 1996, *supra*; Schell 等, 2004, *Bioresource Technol.* 91 :179-188 ;Lee 等, 1999, *Adv. Biochem. Eng. Biotechnol.* 65 :93-115)。

[0250] 还可以使用碱性条件下的几种预处理方法。这些碱性预处理包括,但不限于,石灰预处理、湿氧化、氨渗滤 (APR) 和氨纤维 / 冷冻爆炸 (AFEX)。

[0251] 用碳酸钙、氢氧化钠或氨,在 85-150 °C 的低温进行石灰预处理,停留时间从 1 小时到几天 (Wyman 等, 2005, *Bioresource Technol.* 96 :1959-1966 ;Mosier 等, 2005, *Bioresource Technol.* 96 :673-686)。WO 2006/110891、WO 2006/110899、WO 2006/110900 和 WO 2006/110901 公开了使用氨的预处理方法。

[0252] 湿法氧化是热预处理,通常在 180-220 °C 进行 5-15 分钟,加入氧化剂如过氧化氢或过压氧 (Schmidt 和 Thomsen, 1998, *Bioresource Technol.* 64 :139-151 ;Palonen 等, 2004, *Appl. Biochem. Biotechnol.* 117 :1-17 ;Varga 等, 2004, *Biotechnol. Bioeng.* 88 :567-574 ;Martin 等, 2006, *J. Chem. Technol. Biotechnol.* 81 :1669-1677)。预处理以优选 1-40 % 干物质,更优选 2-30 % 干物质,并且最优性 5-20 % 干物质进行,并且由于加入碱如碳酸钠,初始 pH 经常会增加。

[0253] 湿法氧化预处理方法的修改方法,称为湿爆炸 (湿氧化和蒸汽爆炸的组合),能够处理高达 30 % 的干物质。在湿爆炸中,在预处理过程中,在一定的停留时间后引入氧化剂。然后通过闪变至大气压而结束预处理 (WO 2006/032282)。

[0254] 氨纤维爆炸 (AFEX) 涉及在温和温度如 90-100 °C 和高压如 17-20 bar,用液体或气体氨处理纤维素材料 5-10 分钟,其中干物质含量可高达 60 % (Gollapalli 等, 2002, *Appl. Biochem. Biotechnol.* 98 :23-35 ;Chundawat 等, 2007, *Biotechnol. Bioeng.* 96 :219-231 ;Alizadeh 等, 2005, *Appl. Biochem. Biotechnol.* 121 :1133-1141 ;Teymour 等, 2005, *Bioresource Technol.* 96 :2014-2018)。AFEX 预处理导致纤维素解聚,和半纤维素的部分水解。木质素-糖复合物得到切割。

[0255] 有机溶剂预处理通过用含水乙醇 (40-60 % 乙醇) 在 160-200 °C 提取 30-60 分钟而将纤维素材料去木质素化 (Pan 等, 2005, *Biotechnol. Bioeng.* 90 :473-481 ;Pan 等, 2006, *Biotechnol. Bioeng.* 94 :851-861 ;Kurabi 等, 2005, *Appl. Biochem. Biotechnol.* 121 :219-230)。经常加入硫酸作为催化剂。在有机溶剂预处理中,去除大部分半纤维素。

[0256] 合适的预处理方法的其他实例如 Schell 等, 2003, *Appl. Biochem and Biotechnol.* Vol. 105-108 :69-85, 和 Mosier 等, 2005, *Bioresource Technology* 96 :673-686, 和美国专利公开申请 2002/0164730 所述。

[0257] 在一个方面,化学预处理优选作为酸处理,并且更优选作为连续稀酸和 / 或弱酸 (mild acid) 处理进行。酸通常是硫酸,但也可以使用其它酸,如乙酸、柠檬酸、硝酸、磷酸、酒石酸、琥珀酸、氯化氢或其混合物。弱酸处理在优选 1-5,更优选 1-4,并且最优选 1-3 的

pH 范围内进行。在一个方面,酸浓度在优选 0.01 至 20wt% 酸,更优选 0.05 至 10wt% 酸,甚至更优选 0.1 至 5wt% 酸,并且最优选 0.2 至 2.0wt% 酸的范围内。酸与纤维素材料接触,并在优选 160-220°C,和更优选 165-195°C 范围内的温度保持数秒到数分钟,例如 1 秒-60 分钟的时间。

[0258] 在另一个方面,预处理作为氨纤维爆炸步骤 (AFEX 预处理步骤) 进行。

[0259] 在另一个方面,预处理发生在含水浆料中。在优选的方面,在预处理过程中纤维素材料以优选 10-80wt%,更优选 20-70wt%,并且最优性 30-60wt%,如约 50wt% 的量存在。预处理的纤维素材料可以不洗涤或者使用本领域任何已知的方法洗涤,例如,用水洗涤。

[0260] 机械预处理:术语“机械预处理”指各种类型的磨制 (grinding) 或粉碎 (milling) (例如,干磨、湿磨或振动球磨)。

[0261] 物理预处理:术语“物理预处理”指促进纤维素、半纤维素和 / 或木质素从纤维素材料分离和 / 或释放的任何预处理。例如,物理预处理可涉及辐射 (例如,微波辐射)、汽蒸 / 蒸汽爆炸、水热解 (hydrothermolysis) 和它们的组合。

[0262] 物理预处理可以涉及高压和 / 或高温 (蒸汽爆炸)。在一个方面,高压指优选约 300 至约 600psi,更优选约 350 至约 550psi,并且最优选约 400 至约 500psi,如约 450psi 的压力。在另一个方面,高温指约 100 至约 300°C,优选约 140 至约 235°C 的温度。在一个优选的方面,机械预处理在使用如上所定义的高温和高压的分批过程、蒸汽枪水解器系统,例如来自 Sunds Defibrator AB, Sweden 的 Sunds Hydrolyzer 中进行。

[0263] 组合的物理和化学预处理:可以对纤维素材料进行物理和化学预处理。例如,预处理步骤可以涉及稀酸或弱酸处理和高的温度和 / 或压力处理。根据需要,可以顺序或同时进行物理和化学预处理。还可以包括机械预处理。

[0264] 因此,在一个优选的方面,对纤维素材料进行机械、化学或物理预处理,或者它们的任何组合,以促进纤维素、半纤维素和 / 或木质素的分离和 / 或释放。

[0265] 生物预处理:术语“生物预处理”指可以促进纤维素、半纤维素和 / 或木质素从纤维素材料分离和 / 或释放的任何生物预处理。生物预处理技术可以包括应用溶解木质素的微生物 (参见,例如, Hsu, T.-A., 1996, Pretreatment of biomass, 于 Handbook on Bioethanol: Production and Utilization, Wyman, C. E 编, Taylor & Francis, Washington, DC, 179-212; Ghosh 和 Singh, 1993, Physicochemical and biological treatments for enzymatic/microbial conversion of lignocellulosic biomass, Adv. Appl. Microbiol. 39 :295-333; McMillan, J. D., 1994, Pretreating lignocellulosic biomass: a review, 于 Enzymatic Conversion of Biomass for Fuel Production, Himmel, M. E., Baker, J. O., 和 Overend, R. P., 编, ACS Symposium Series 566, American Chemical Society, Washington, DC, 第 15 章; Gong, C. S., Cao, N. J., Du, J., 和 Tsao, G. T., 1999, Ethanol production from renewable resources, 于 Advances in Biochemical Engineering/Biotechnology, Scheper, T., 编, Springer-Verlag Berlin Heidelberg, Germany, 65 :207-241; Olsson 和 Hahn-Hagerdal, 1996, Fermentation of lignocellulosic hydrolysates for ethanol production, Enz. Microb. Tech. 18 :312-331; 和 Vallander 和 Eriksson, 1990, Production of ethanol from lignocellulosic materials: State of the art, Adv. Biochem. Eng. / Biotechnol. 42 :63-95)。

[0266] 糖化。在水解（也称作糖化）步骤中，将例如经预处理的纤维素材料水解以将纤维素和任选地也将半纤维素分解成可发酵糖，如葡萄糖、纤维二糖、木糖、木酮糖、阿拉伯糖、甘露糖、半乳糖和 / 或可溶的寡糖。水解由酶组合物以酶法在本发明具有纤维素分解增强活性的多肽的存在下进行。组合物的酶还可以顺序加入。

[0267] 酶水解优选在易于由本领域技术人员确定的条件下，在合适的含水环境中进行。在一个优选的方面，水解在适于酶的活性，即对于酶最优的条件下进行。水解可以以补料批次或连续的过程进行，其中将预处理的纤维素材料（底物）逐渐补入，例如，含酶的水解溶液中。

[0268] 糖化通常在搅拌釜反应器或发酵罐中，在受控的 pH、温度和混合条件下进行。合适的处理时间、温度和 pH 条件可以由本领域技术人员容易地确定。例如，糖化可以持续长达 200 小时，但是通常优选进行约 12 至约 96 小时，更优选约 16 至约 72 小时，并且最优选约 24 至约 48 小时。温度优选约 25°C 至约 70°C，更优选约 30°C 至约 65°C，并且最优选约 40°C 至约 60°C，特别是约 50°C。pH 优选约 3 至约 8，更优选约 3.5 至约 7，并且最优选约 4 至约 6，特别是约 pH 5。干燥固体含量优选约 5 至约 50wt%，更优选约 10 至约 40wt%，并且最优选约 20 至约 30wt%。

[0269] 在一个方面，所述酶组合物包含或进一步包含一种或多种（几种）选自下组的蛋白：纤维素酶、半纤维素酶、棒曲霉素、酯酶、木质素分解酶、果胶酶、过氧化物酶、蛋白酶和膨胀素 (swollenin)。在另一个方面，所述纤维素酶优选为一种或多种（几种）选自下组的酶：内切葡聚糖酶、纤维二糖水解酶和  $\beta$ - 葡糖苷酶。在另一个方面，所述半纤维素酶优选为一种或多种（几种）选自下组的酶：乙酰甘露聚糖酯酶、乙酰木聚糖酯酶、阿拉伯聚糖酶、阿拉伯呋喃糖苷酶、香豆酸酯酶、阿魏酸酯酶、半乳糖苷酶、葡糖醛酸糖苷酶、葡糖醛酸酯酶、甘露聚糖酶、甘露糖苷酶、木聚糖酶和木糖苷酶。

[0270] 在另一个方面，所述酶组合物包含一种或多种（几种）纤维素分解酶。在另一个方面，所述酶组合物包含或进一步包含一种或多种（几种）半纤维素分解酶。在另一个方面，所述酶组合物包含一种或多种（几种）纤维素分解酶和一种或多种（几种）半纤维素分解酶。在另一个方面，所述酶组合物包含一种或多种（几种）选自纤维素分解酶和半纤维素分解酶的组的酶。在另一个方面，所述酶组合物包含内切葡聚糖酶。在另一个方面，所述酶组合物包含纤维二糖水解酶。在另一个方面，所述酶组合物包含  $\beta$ - 葡糖苷酶。在另一个方面，所述酶组合物包含乙酰甘露聚糖酯酶。在另一个方面，所述酶组合物包含乙酰木聚糖酯酶。在另一个方面所述酶组合物包含阿拉伯聚糖酶（例如  $\alpha$ -L- 阿拉伯聚糖酶）。在另一个方面，所述酶组合物包含阿拉伯呋喃糖苷酶（例如  $\alpha$ -L- 阿拉伯呋喃糖苷酶）。在另一个方面，所述酶组合物包含香豆酸酯酶。在另一个方面，所述酶组合物包含阿魏酸酯酶。在另一个方面，所述酶组合物包含半乳糖苷酶（例如， $\alpha$ - 半乳糖苷酶和 / 或  $\beta$ - 半乳糖苷酶）。在另一个方面，所述酶组合物包含葡糖醛酸糖苷酶（例如， $\alpha$ -D- 葡糖醛酸糖苷酶）。在另一个方面，所述酶组合物包含葡糖醛酸酯酶。在另一个方面，所述酶组合物包含甘露聚糖酶。在另一个方面，所述酶组合物包含甘露糖苷酶（例如  $\beta$ - 甘露糖苷酶）。在另一个方面，所述酶组合物包含木聚糖酶。在一个优选的方面，所述木聚糖酶为家族 10 木聚糖酶。在另一个方面，所述酶组合物包含木糖苷酶。

[0271] 在另一个方面，所述酶组合物包含棒曲霉素。在另一个方面，所述酶组合物包含酯

酶。在另一个方面,所述酶组合物包含木质素分解酶。在一个优选的方面,所述木质素分解酶为漆酶。在另一个优选的方面,所述木质素分解酶是锰过氧化物酶。在另一个优选的方面,所述木质素分解酶是木质素过氧化物酶。在另一个优选的方面,所述木质素分解酶是H<sub>2</sub>O<sub>2</sub>产生酶。在另一个方面,所述酶组合物包含果胶酶。在另一个方面,所述酶组合物包含过氧化物酶。在另一个方面,所述酶组合物包含蛋白酶。在另一个方面,所述酶组合物包含膨胀素(swollenin)。

[0272] 在本发明的方法中,酶可在发酵之前或过程中添加,例如在糖化过程中或在发酵微生物繁殖过程中或之后添加。

[0273] 所述酶组合物的一种或多种(几种)组分可为野生型蛋白、重组蛋白或野生型蛋白和重组蛋白的组合。举例而言,一种或多种(几种)组分可为细胞的天然蛋白,其用作宿主细胞以重组表达一种或多种(几种)酶组合物的其他组分。酶组合物的一种或多种(几种)组分可作为单组分产生,然后将其组合以形成酶组合物。所述酶组合物可为多组分和单组分蛋白制备物的组合。

[0274] 用于本发明方法中的酶可为任何适用于本文中所述工艺的形式,如例如去除或未去除细胞的粗发酵液,含有或不含细胞碎片的细胞裂解物,半纯化或纯化的酶制备物,或宿主细胞作为酶的来源。所述酶组合物可为干粉或颗粒,无粉尘的颗粒,液体,稳定化液体或稳定化受保护的酶。液体酶制备物可根据确立的工艺,例如通过添加稳定剂如糖、糖醇或其他多元醇,和/或乳酸或其他有机酸来稳定化。

[0275] 具有纤维素分解增强活性的酶和多肽的最适量取决于几个因素,其包括但不限于,组分纤维素分解酶的混合物、含纤维素底物、含纤维素底物的浓度、含纤维素底物的预处理、温度、时间、pH和包括发酵生物体(例如,用于同步糖化和发酵的酵母)。

[0276] 在一个优选的方面,纤维素分解酶对于纤维素材料的有效量是约0.5至约50mg,优选约0.5至约40mg,更优选约0.5至约25mg,更优选约0.75至约20mg,更优选约0.75至约15mg,甚至更优选约0.5至约10mg,并且最优选约2.5至约10mg每g纤维素材料。

[0277] 在另一个优选的方面,具有纤维素分解增强活性的多肽对于纤维素材料的有效量是约0.01至约50.0mg,优选约0.01至约40mg,更优选约0.01至约30mg,更优选约0.01至约20mg,更优选约0.01至约10mg,更优选约0.01至约5mg,更优选约0.025至约1.5mg,更优选约0.05至约1.25mg,更优选约0.075至约1.25mg,更优选约0.1至约1.25mg,甚至更优选约0.15至约1.25mg,并且最优选约0.25至约1.0mg每g纤维素材料。

[0278] 在另一个优选的方面,具有纤维素分解增强活性的多肽对于纤维素分解酶的有效量是约0.005至约1.0g,优选约0.01至约1.0g,更优选约0.15至约0.75g,更优选约0.15至约0.5g,更优选约0.1至约0.5g,甚至更优选约0.1至约0.25g,并且最优选约0.05至约0.2g每g纤维素分解酶。

[0279] 酶可源自或获得自任何合适的来源,包括细菌、真菌、酵母、植物或哺乳动物来源。术语“获得”在本文中意指该酶可从天然产生该酶作为天然酶的生物分离。术语“获得”在本文中还意指该酶可在宿主生物中使用本文中所述的方法重组产生,其中经重组产生的酶对于宿主生物是天然的或异源的,或具有修饰的氨基酸序列,例如,具有一个或多个(几个)缺失、插入和/或取代的氨基酸,即重组产生的酶,其为天然氨基酸序列的片段和/或突变体或通过本领域已知的氨基酸改组方法产生的酶。天然酶的含义中涵盖的是天然变体,而

外来酶的含义中涵盖的是重组（如通过定位诱变或重排）获得的变体。

[0280] 具有酶活性的多肽可以是细菌多肽。例如，所述多肽可以是具有酶活性的革兰氏阳性细菌多肽如芽孢杆菌属、链球菌属、链霉菌属、葡萄球菌属、肠球菌属、乳杆菌属、乳球菌属、梭菌属、地芽孢杆菌属或海洋芽孢杆菌属多肽；或具有酶活性的革兰氏阴性细菌多肽，如大肠杆菌、假单胞菌属、沙门氏菌属、弯曲杆菌属、螺杆菌属、黄杆菌属、梭杆菌属、泥杆菌属、奈瑟氏菌属或脲原体属多肽。

[0281] 在一个优选的方面，所述多肽是具有酶活性的嗜碱芽孢杆菌、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、凝结芽孢杆菌、坚强芽孢杆菌、灿烂芽孢杆菌、迟缓芽孢杆菌、地衣芽孢杆菌、巨大芽孢杆菌、短小芽孢杆菌、嗜热脂肪芽孢杆菌、枯草芽孢杆菌或苏云金芽孢杆菌多肽。

[0282] 在另一个优选的方面，所述多肽是具有酶活性的似马链球菌、酿脓链球菌、乳房链球菌或马链球菌兽瘟亚种多肽。

[0283] 在另一个优选的方面，所述多肽是具有酶活性的不产色链霉菌、除虫链霉菌、天蓝链霉菌、灰色链霉菌或浅青紫链霉菌多肽。

[0284] 具有酶活性的多肽也可以是真菌多肽，并且更优选具有酶活性的酵母多肽如假丝酵母属、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多肽；或更优选具有酶活性的丝状真菌多肽如枝顶孢霉属、伞菌属、链格孢属、曲霉属、短梗霉属、Botryosphaeria、拟蜡菌属、毛喙壳属、金孢子菌属、Claviceps、Cochliobolus、鬼伞属、Coptotermes、棒囊壳属、隐丛赤壳菌属、隐球菌属、色二孢属、黑耳属、Filibasidium、镰孢属、赤霉属、全鞭毛虫属、腐质霉属、耙齿菌属、蘑菇属、Leptosphaeria、梨孢菌属、Melanocarpus、多孔菌属、毛霉属、毁丝霉属、新考玛脂霉属、脉孢菌属、拟青霉属、青霉属、平革菌属、瘤胃壶菌属、Poitrasia、假黑盘菌属、Pseudotriconympha、根毛霉属、裂褶菌属、柱顶孢属、踝节菌属、嗜热子囊菌属、梭孢壳属、弯颈霉属、木霉属、长毛盘菌属、轮枝孢属、包脚菇属或炭角菌属多肽。

[0285] 在一个优选的方面，所述多肽是具有酶活性的卡尔酵母、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多肽。

[0286] 在另一个优选的方面，所述多肽是具有酶活性的解纤维枝顶孢霉、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、米曲霉、嗜角质金孢子菌、Chrysosporium lucknowense、热带金孢子菌、Chrysosporium merdarium、Chrysosporium inops、毡金孢子菌、Chrysosporium queenslandicum、Chrysosporium zonatum、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、硫色镰孢、圆镰孢、拟丝孢镰孢、镶片镰孢、灰腐质霉、特异腐质霉、疏棉状腐质霉、白耙齿菌、米黑毛霉、嗜热毁丝霉、粗糙脉孢菌、绳状青霉、产紫青霉、黄孢平革菌、无色梭孢壳、Thielavia albomyces、Thielavia albopilosa、澳洲梭孢壳、Thielavia fimeti、小孢梭孢壳、卵孢梭孢壳、Thielavia peruviana、瘤孢梭孢壳、毛梭孢壳、Thielavia subthermophila、土生梭孢霉、哈茨木霉、康宁木霉、长枝木霉、里氏木霉、绿色木霉或 Trichophaea saccata 多肽。

[0287] 还可使用具有酶活性的多肽的经化学修饰或蛋白质工程改造的突变体。

[0288] 所述纤维素分解酶组合物的一种或多种组分可以是重组组分，亦即，通过克隆编

码所述单独组分的 DNA 序列并随后用该 DNA 序列转化细胞并在宿主中表达（参见，例如，W091/17243 和 W091/17244）来产生。所述宿主优选是异源宿主（酶对宿主是异源的），但该宿主在一定条件下也可以是同源宿主（酶对宿主是同源的）。单组分纤维素分解蛋白还可以通过从发酵液中纯化这样的蛋白质来制备。

[0289] 在一个方面，所述一种或多种（几种）酶包括商业性纤维素分解酶制备物。适用于本发明的商业性的纤维素分解酶制备物的实例包括，例如，CELLIC™ Ctec(Novozymes A/S)、CELLUCLAST™(Novozymes A/S)、NOVOZYM™ 188(Novozymes A/S)、CELLUZYME™(Novozymes A/S)、CEREFLO™(Novozymes A/S) 和 ULTRAFLO™(Novozymes A/S)，ACCELERASE™(Genencor Int.)、LAMINEX™(Genencor Int.)、SPEZYME™ CP(Genencor Int.)，ROHAMENT™7069W( Röh m GmbH)，FIBREZYME® LDI(Dyadic International, Inc.)、FIBREZYME® LBR(Dyadic International, Inc.) 或 VISCOSTAR® 150L(Dyadic International, Inc.)。所述纤维素酶以固体的约 0.001 到约 5.0wt. %，更优选固体的 0.025 到约 4.0wt. %，且最优选固体的约 0.005 到约 2.0wt. % 的有效量添加。所述纤维素酶以固体的约 0.001 到约 5.0wt. %，更优选固体的 0.025 到约 4.0wt. %，且最优选固体的约 0.005 到约 2.0wt. % 的有效量添加。

[0290] 可以用于本发明的方法的细菌内切葡聚糖酶的实例包括但不限于，解纤维热酸菌 (*Acidothermus cellulolyticus*) 内切葡聚糖酶 (WO 91/05039; W093/15186; 美国专利 5,275,944; WO 96/02551; 美国专利 5,536,655, W000/70031, WO 05/093050); *Thermobifida fusca* 内切葡聚糖酶 III (W005/093050); 和 *Thermobifida fusca* 内切葡聚糖酶 V (WO 05/093050)。

[0291] 可以用于本发明的方法的真菌内切葡聚糖酶的实例包括但不限于，里氏木霉内切葡聚糖酶 I (Penttila 等, 1986, Gene 45 :253-263; GENBANK™ 登录号 M15665); 里氏木霉内切葡聚糖酶 II (Saloheimo 等, 1988, Gene 63 :11-22; GENBANK™ 登录号 M19373); 里氏木霉内切葡聚糖酶 III (Okada 等, 1988, Appl. Environ. Microbiol. 64 :555-563; GENBANK™ 登录号 AB003694); 以及里氏木霉内切葡聚糖酶 V (Saloheimo 等, 1994, Molecular Microbiology 13 :219-228; GENBANK™ 登录号 Z33381); 棘孢曲霉内切葡聚糖酶 (Ooi 等, 1990, Nucleic Acids Research 18 :5884); 川地曲霉 (*Aspergillus kawachii*) 内切葡聚糖酶 (Sakamoto 等, 1995, Current Genetics 27 :435-439); 胡萝卜软腐欧文氏菌 (*Erwinia carotovora*) 内切葡聚糖酶 (Saarilahti 等, 1990, Gene 90 :9-14); 尖镰孢内切葡聚糖酶 (GENBANK™ 登录号 L29381); 灰腐质霉 *thermoidea* 变种内切葡聚糖酶 (GENBANK™ 登录号 AB003107); *Melanocarpus albomyces* 内切葡聚糖酶 (GENBANK™ 登录号 MAL515703); 粗糙脉孢菌内切葡聚糖酶 (GENBANK™ 登录号 XM\_324477); 特异腐质霉内切葡聚糖酶 V; 嗜热毁丝霉 CBS 117.65 内切葡聚糖酶; 担子菌纲 (*basidiomycete*) CBS 495.95 内切葡聚糖酶; 担子菌纲 CBS 494.95 内切葡聚糖酶; 土生梭孢霉 NRRL 8126 CEL6B 内切葡聚糖酶; 土生梭孢霉 NRRL 8126 CEL6C 内切葡聚糖酶; 土生梭孢霉 NRRL 8126 CEL7C 内切葡聚糖酶; 土生梭孢霉 NRRL8126 CEL7E 内切葡聚糖酶; 土生梭孢霉 NRRL 8126 CEL7F 内切葡聚糖酶; *Cladorrhinum foecundissimum* ATCC 62373 CEL7A 内切葡聚糖酶; 以及里氏木霉菌株 VTT-D-80133 内切葡聚糖酶 (GENBANK™ 登录号 M15665)。

[0292] 可用于本发明的方法的纤维二糖水解酶的示例包括但不限于，里氏木霉纤维

二糖水解酶 I ;里氏木霉纤维二糖水解酶 II ;特异腐质霉纤维二糖水解酶 I、嗜热毁丝霉纤维二糖水解酶 II、土生梭孢霉纤维二糖水解酶 II (CEL6A)、嗜热毛壳菌 (*Chaetomium thermophilum*) 纤维二糖水解酶 I 以及嗜热毛壳菌纤维二糖水解酶 II。

[0293] 可用于本发明的方法的  $\beta$ -葡糖苷酶的实例包括但不限于米曲霉  $\beta$ -葡糖苷酶 ;烟曲霉  $\beta$ -葡糖苷酶 ;巴西青霉 (*Penicillium brasiliense*) IBT 20888  $\beta$ -葡糖苷酶 ;黑曲霉  $\beta$ -葡糖苷酶 ;以及棘孢曲霉  $\beta$ -葡糖苷酶。

[0294] 具有  $\beta$ -葡糖苷酶活性的米曲霉多肽可根据 WO 2002/095014 获取。具有  $\beta$ -葡糖苷酶活性的烟曲霉多肽可根据 WO 2005/047499 获取。具有  $\beta$ -葡糖苷酶活性的巴西青霉多肽可根据 WO 2007/019442 获取。具有  $\beta$ -葡糖苷酶活性的黑曲霉多肽可根据 Dan 等, 2000, *J. Biol. Chem.* 275 :4973-4980 获取。具有  $\beta$ -葡糖苷酶活性的棘孢曲霉多肽可根据 Kawaguchi 等, 1996, *Gene* 173 :287-288 获取。

[0295] 所述  $\beta$ -葡糖苷酶可以是融合蛋白。在一个方面, 所述  $\beta$ -葡糖苷酶是根据 WO 2008/057637 得到的米曲霉  $\beta$ -葡糖苷酶融合蛋白或米曲霉  $\beta$ -葡糖苷酶变体 BG 融合蛋白。

[0296] 其它内切葡聚糖酶、纤维二糖水解酶和  $\beta$ -葡糖苷酶使用根据 Henrissat B., 1991, *A classification of glycosyl hydrolases based on amino-acid sequence similarities*, *Biochem. J.* 280 :309-316 和 Henrissat B. 和 Bairoch A., 1996, *Updating the sequence-based classification of glycosyl hydrolases*, *Biochem. J.* 316 :695-696 的分类公开于许多糖基水解酶家族中。

[0297] 其它可用于本发明的纤维素分解酶描述于 EP 495, 257、EP 531, 315、EP531, 372、WO 89/09259、WO 94/07998、WO 95/24471、WO 96/11262、W096/29397、WO 96/034108、WO 97/14804、WO 98/08940、WO 98/012307、W098/13465、WO 98/015619、WO 98/015633、WO 98/028411、WO 99/06574、WO 99/10481、WO 99/025846、WO 99/025847、WO 99/031255、W02000/009707、WO 2002/050245、WO 2002/0076792、WO 2002/101078、W02003/027306、WO 2003/052054、WO 2003/052055、WO 2003/052056、W02003/052057、WO 2003/052118、WO 2004/016760、WO 2004/043980、W02004/048592、WO 2005/001065、WO 2005/028636、WO 2005/093050、W02005/093073、WO 2006/074005、WO 2006/117432、WO 2007/071818、W02007/071820、WO 2008/008070、WO 2008/008793、美国专利 No. 4, 435, 307、美国专利 No. 5, 457, 046、美国专利 No. 5, 648, 263、美国专利 No. 5, 686, 593、美国专利 No. 5, 691, 178、美国专利 No. 5, 763, 254 以及美国专利 No. 5, 776, 757。

[0298] 在一个方面, 所述一种或多种 (几种) 半纤维素分解酶包括商业性半纤维素分解酶制备物。适用于本发明的商业性半纤维素分解酶制备物的实例包括, 例如 SHEARZYME™ (Novozymes A/S)、CELLIC™ Htec (Novozymes A/S)、VISCOZYME® (Novozymes A/S)、ULTRAFLO® (Novozymes A/S)、PULPZYME® HC (Novozymes A/S)、MULTIFECT® Xylanase (Genencor)、ECOPULP® TX-200A (AB Enzymes)、HSP 6000 Xylanase (DSM)、DEPOL™333P (Biocatalysts Limit, Wales, UK)、DEPOL™ 740L (Biocatalysts Limit, Wales, UK) 和 DEPOL™ 762P (Biocatalysts Limit, Wales, UK)。

[0299] 可用于本发明方法的木聚糖酶的实例包括但不限于棘孢曲霉 (*Aspergillus aculeatus*) 木聚糖酶 (GeneSeqP:AAR63790;WO 94/21785)、烟曲霉 (*Aspergillus*

fumigatus) 木聚糖酶 (WO 2006/078256) 和土生梭孢霉 (*Thielavia terrestris*) NRRL 8126 木聚糖酶 (WO 2009/079210)。

[0300] 可用于本发明方法的  $\beta$ -木糖苷酶的实例包括但不限于里氏木霉 (*Trichoderma reesei*)  $\beta$ -木糖苷酶 (UniProtKB/TrEMBL 登录号 Q92458), 埃默森踝节菌 (*Talaromyces emersonii*) (SwissProt 登录号 Q8X212) 和粗糙脉孢菌 (*Neurospora crassa*) (SwissProt 登录号 Q7S0W4)。

[0301] 可用于本发明方法的乙酰木聚糖酯酶的实例包括但不限于红褐肉座菌 (*Hypocrea jecorina*) 乙酰木聚糖酯酶 (WO 2005/001036)、粗糙脉孢菌乙酰木聚糖酯酶 (UniProt 登录号 q7s259)、土生梭孢霉 NRRL 8126 乙酰木聚糖酯酶 (WO2009/042846)、球毛壳菌 (*Chaetomium globosum*) 乙酰木聚糖酯酶 (Uniprot 登录号 Q2GWX4)、细丽毛壳菌 (*Chaetomium gracile*) 乙酰木聚糖酯酶 (GeneSeqP 登录号 AAB82124)、颖枯壳针孢 (*Phaeosphaeria nodorum*) 乙酰木聚糖酯酶 (Uniprot 登录号 Q0UHJ1) 和特异腐质霉 (*Humicola insolens*) DSM 1800 乙酰木聚糖酯酶 (WO 2009/073709)。

[0302] 可用于本发明方法的阿魏酸酯酶的实例包括但不限于特异腐质霉 DSM 1800 阿魏酸酯酶 (WO 2009/076122)、粗糙脉孢菌阿魏酸酯酶 (UniProt 登录号 Q9HGR3) 和费希新萨托菌 (*Neosartorya fischeri*) 阿魏酸酯酶 (UniProt 登录号 A1D9T4)。

[0303] 可用于本发明方法的阿拉伯呋喃糖苷酶的实例包括但不限于特异腐质霉 (*Humicola insolens*) DSM 1800 阿拉伯呋喃糖苷酶 (WO 2009/073383) 和黑曲霉 (*Aspergillus niger*) 阿拉伯呋喃糖苷酶 (GeneSeqP 登录号 AAR94170)。

[0304] 可用于本发明方法的  $\alpha$ -葡糖醛酸糖苷酶的实例包括但不限于棒曲霉 (*Aspergillus clavatus*)  $\alpha$ -葡糖醛酸糖苷酶 (UniProt 登录号 alcc12)、里氏木霉 (*Trichoderma reesei*)  $\alpha$ -葡糖醛酸糖苷酶 (Uniprot 登录号 Q99024)、埃默森踝节菌 (*Talaromyces emersonii*)  $\alpha$ -葡糖醛酸糖苷酶 (UniProt 登录号 Q8X211)、黑曲霉 (*Aspergillus niger*)  $\alpha$ -葡糖醛酸糖苷酶 (Uniprot 登录号 Q96WX9)、土曲霉 (*Aspergillus terreus*)  $\alpha$ -葡糖醛酸糖苷酶 (SwissProt 登录号 Q0CJP9) 和烟曲霉 (*Aspergillus fumigatus*)  $\alpha$ -葡糖醛酸糖苷酶 (SwissProt 登录号 Q4WW45)。

[0305] 用于本发明方法的酶和蛋白可通过在含有合适碳源和氮源和无机盐, 使用本领域已知方法 (参见, 例如 Bennett, J. W. 和 LaSure, L. (编), *More Gene Manipulations in Fungi*, Academic Press, CA, 1991) 的营养培养基上发酵上述指出的微生物菌株来产生。合适的培养基可从供应商获得, 或可根据已公开组合物制备 (例如美国典型培养物保藏中心的目录)。适于生长和酶产生的温度范围和其他条件在本领域是已知的 (参见, 例如 Bailey, J. E. 和 Ollis, D. F., *Biochemical Engineering Fundamentals*, McGraw-Hill Book Company, NY, 1986)。

[0306] 所述发酵可以是任何导致酶的表达或分离的培养细胞的方法。因此, 发酵可以理解为包括在合适的培养基中并在允许所述酶得以表达或分离的条件下进行的摇瓶培养, 或在实验室或工业发酵罐中的小-或大规模发酵 (包括连续、分批、补料分批或固态发酵)。通过上述方法产生的所得的酶可从发酵培养基回收并通过常规方法纯化。

[0307] **发酵**。可通过一种或多种 (几种) 能将糖直接或间接发酵成所需发酵产物的发酵微生物发酵自经水解的纤维素材料获得的可发酵糖。“发酵”或“发酵方法”指任何发酵方法



或包含发酵步骤的任何方法。发酵方法还包括用于消费品醇工业（例如，啤酒和葡萄酒）、乳品业（例如，发酵乳制品）、皮革业和烟草业的发酵方法。发酵条件依赖于期望的发酵产物和发酵生物体，并且能由本领域的技术人员容易地确定。

[0308] 在发酵步骤中，作为预处理和酶水解步骤的结果从纤维素材料释放的糖，通过发酵生物体（如酵母）发酵成为产物，例如，乙醇。如上所述，水解（糖化）和发酵可以是单独或同时的。

[0309] 在实施本发明时在发酵步骤中可以使用任何合适的经水解的纤维素材料。通常根据所需发酵产品（即，要从发酵获得的物质）和使用的方法来选择所述材料，如本领域中所公知的。

[0310] 术语“发酵培养基”在本文中可理解为指加入发酵微生物之前的培养基，如，由糖化过程产生的培养基，以及同步的糖化和发酵方法（SSF）中使用的培养基。

[0311] “发酵微生物”指适用于理想的发酵方法产生发酵产物的任何微生物，包括细菌和真菌生物体。发酵生物体可以是  $C_6$  和 / 或  $C_5$  发酵生物体，或它们的组合。 $C_6$  和  $C_5$  发酵生物体均在本领域公知。合适的发酵微生物能将糖（如葡萄糖、木糖、木酮糖、阿拉伯糖、麦芽糖、甘露糖、半乳糖或寡糖）直接或间接地发酵（即，转化）成所需的发酵产品。

[0312] 可产生乙醇的细菌和真菌发酵生物体的实例如 Lin 等，2006, Appl. Microbiol. Biotechnol. 69 :627-642 所述。

[0313] 能发酵  $C_6$  糖的发酵微生物的实例包括细菌和真菌生物体，如酵母。优选的酵母包括酵母属菌种，优选酿酒酵母。

[0314] 能发酵  $C_5$  糖的发酵生物体的实例包括细菌和真菌生物体，如酵母。优选的  $C_5$  发酵酵母包括毕赤酵母属，优选树干毕赤酵母 (*Pichia stipitis*) 的菌株，如树干毕赤酵母 CBS 5773；假丝酵母属，优选博伊丁假丝酵母 (*Candida boidinii*)、芸薹假丝酵母 (*Candida brassicae*)、休哈塔假丝酵母 (*Candida sheatae*)、迪丹斯假丝酵母 (*Candida diddensii*)、假热带假丝酵母 (*Candida pseudotropicalis*) 或产朊假丝酵母 (*Candida utilis*) 的菌株。

[0315] 其它发酵生物体包括发酵单胞菌属 (*Zymomonas*)，如运动发酵单胞菌 (*Zymomonas mobilis*)；汉逊酵母属，如异常汉逊酵母 (*Hansenula anomala*)；克鲁维酵母属，如脆壁克鲁维酵母；裂殖酵母属，如粟酒裂殖酵母 (*S. pombe*)；和大肠杆菌的菌株，特别是已经经过遗传修饰而改进乙醇产量 (yield) 的大肠杆菌菌株。

[0316] 在一个优选的方面，酵母是酵母属菌种。在一个更优选的方面，酵母是酿酒酵母。在另一个更优选的方面，酵母是糖化酵母 (*Saccharomyces distaticus*)。在另一个更优选的方面，酵母是葡萄汁酵母 (*Saccharomyces uvarum*)。在另一个优选的方面，酵母是克鲁维酵母属。在另一个更优选的方面，酵母是马克斯克鲁维酵母 (*Kluyveromyces marxianus*)。在另一个更优选的方面，酵母是脆壁克鲁维酵母。在另一个优选的方面，酵母是假丝酵母属。在另一个更优选的方面，酵母是博伊丁假丝酵母。在另一个更优选的方面，酵母是芸薹假丝酵母。在另一个更优选的方面，酵母是迪丹斯假丝酵母。在另一个更优选的方面，酵母是假热带假丝酵母。在另一个更优选的方面，酵母是产朊假丝酵母。在另一个优选的方面，酵母是棒孢酵母属 (*Clavispora*)。在另一个更优选的方面，酵母是葡萄牙棒孢酵母 (*Clavispora lusitaniae*)。在另一个更优选的方面，酵母是仙人掌棒孢酵

母 (*Clavispora opuntiae*)。在另一个优选的方面,酵母是管囊酵母属 (*Pachysolen*)。在另一个更优选的方面,酵母是嗜鞣管囊酵母 (*Pachysolen tannophilus*)。在另一个优选的方面,酵母是毕赤酵母属。在另一个更优选的方面,酵母是树干毕赤酵母。在另一个优选的方面,酵母是酒香酵母属 (*Bretannomyces*)。在另一个更优选的方面,酵母是克劳森酒香酵母 (*Bretannomyces clausenii*) (Philippidis, G. P., 1996, Cellulose bioconversion technology, 于 Handbook on Bioethanol :Production and Utilization, Wyman, C. E. 编, Taylor & Francis, Washington, DC, 179-212)。

[0317] 能有效地将己糖和戊糖发酵成乙醇的细菌包括,例如,运动发酵单胞菌和热纤维梭菌 (*Clostridium thermocellum*) (Philippidis, 1996, 见上文)。

[0318] 在一个优选的方面,细菌是发酵单胞菌属。在更优选的方面,细菌是运动发酵单胞菌。在另一个优选的方面,细菌是梭菌属。在另一个更优选的方面,细菌是热纤维梭菌。

[0319] 商业上可得到的适合乙醇产生的酵母包括,例如 ETHANOL RED™ 酵母 (可从 Red Star/Lesaffre, USA 获得)、FALI™ (可从 Fleischmann's Yeast, Burns Philp Food Inc., USA 的部门获得)、SUPERSTART™ 和 THERMOSACC™ 新鲜酵母 (可从 Ethanol Technology, WI, USA 获得)、BIOFERM™ AFT 和 XR (可从 NABC-North American Bioproducts Corporation, GA, USA 获得)、GERT STRAND™ (可从 Gert Strand AB, Sweden 获得) 和 FERMIOL™ (可从 DSM Specialties 获得)。

[0320] 在一个优选的方面,发酵微生物已经经过遗传修饰以提供发酵戊糖的能力,如利用木糖、利用阿拉伯糖和共同利用木糖和阿拉伯糖的微生物。

[0321] 通过将异源基因克隆入多种发酵微生物已经构建了能将己糖和戊糖转化成乙醇 (共发酵) 的生物体 (Chen 和 Ho, 1993, Cloning and improving the expression of *Pichia stipitis* xylose reductase gene in *Saccharomyces cerevisiae*, Appl. Biochem. Biotechnol. 39-40 :135-147 ;Ho 等, 1998, Genetically engineered *Saccharomyces* yeast capable of effectively cofermenting glucose and xylose, Appl. Environ. Microbiol. 64 :1852-1859 ;Kotter 和 Ciriacy, 1993, Xylose fermentation by *Saccharomyces cerevisiae*, Appl. Microbiol. Biotechnol. 38 :776-783 ;Walfridsson 等, 1995, Xylose-metabolizing *Saccharomyces cerevisiae* strains overexpressing the TKL1 and TAL1 genes encoding the pentose phosphate pathway enzymes transketolase and transaldolase, Appl. Environ. Microbiol. 61 :4184-4190 ;Kuyper 等, 2004, Minimal metabolic engineering of *Saccharomyces cerevisiae* for efficient anaerobic xylose fermentation :a proof of principle, FEMS Yeast Research 4 : 655-664 ;Beall 等, 1991, Parametric studies of ethanol production from xylose and other sugars by recombinant *Escherichia coli*, Biotech. Bioeng. 38 :296-303 ;Ingram 等, 1998, Metabolic engineering of bacteria for ethanol production, Biotechnol. Bioeng. 58 :204-214 ;Zhang 等, 1995, Metabolic engineering of a pentose metabolism pathway in ethanologenic *Zymomonas mobilis*, Science 267 :240-243 ;Deanda 等, 1996, Development of an arabinose-fermenting *Zymomonas mobilis* strain by metabolic pathway engineering, Appl. Environ. Microbiol. 62 :4465-4470 ;WO 2003/062430, xylose isomerase)。

[0322] 在一个优选的方面,经过遗传修饰的发酵微生物是酿酒酵母。在另一个优选的方面,经过遗传修饰的发酵微生物是运动发酵单胞菌。在另一个优选的方面,经过遗传修饰的发酵微生物是大肠杆菌。在另一个优选的方面,经过遗传修饰的发酵微生物是产酸克雷伯氏菌 (*Klebsiella oxytoca*)。在另一个优选的方面,所述经遗传修饰的发酵微生物是克鲁维酵母属菌种。

[0323] 本领域中公知的是,上述生物体还能用于产生其它物质,如本文所述。

[0324] 通常向降解的木素纤维素或水解物加入发酵微生物,并进行约 8 至约 96 小时,如约 24 至约 60 小时发酵。温度通常为约 26°C 至约 60°C,特别是约 32°C 或 50°C,并且在约 pH 3 至约 pH 8,如约 pH 4-5、6 或 7。

[0325] 在一个优选的方面,对降解的纤维素材料施用酵母和 / 或另一种微生物,并进行约 12 至约 96 小时,如通常为 24-60 小时发酵。在一个优选的方面,温度优选为约 20°C 至约 60°C,更优选约 25°C 至约 50°C,并且最优选约 32°C 至约 50°C,特别是约 32°C 或 50°C,并且 pH 通常为约 pH 3 至约 pH 7,优选约 pH 4-7。然而,一些发酵生物体例如细菌,具有更高的最适发酵温度。酵母或另一种微生物优选以约  $10^5$ - $10^{12}$ , 优选约  $10^7$ - $10^{10}$ , 特别是约  $2 \times 10^8$  活细胞计数每 ml 发酵液的量施用。关于使用酵母进行发酵的进一步指导可以在例如“*The Alcohol Textbook*”(K. Jacques, T. P. Lyons 和 D. R. Kelsall 编, Nottingham University Press, United Kingdom 1999) 中找到,其通过提述并入本文。

[0326] 对于乙醇生产,在发酵后蒸馏发酵的浆料以提取乙醇。根据本发明的方法获得的乙醇可以用作,例如燃料乙醇;饮料乙醇,即,中性饮料酒,或工业乙醇。

[0327] 发酵刺激剂可以与本文所述的任何方法组合使用,以进一步改进发酵工艺,而且特定地,改进发酵微生物的性能,如,速率增加和乙醇得率。“发酵刺激剂”指用于发酵微生物(特别是酵母)生长的刺激剂。优选的用于生长的发酵刺激剂包括维生素和矿物质。维生素的实例包括多种维生素、生物素、泛酸(盐)、烟酸、内消旋肌醇(meso-inositol)、硫胺素、吡哆醇(pyridoxine)、对氨基苯甲酸、叶酸、核黄素和维生素 A、B、C、D 和 E。参见,例如,Alfenore 等, *Improving ethanol production and viability of Saccharomyces cerevisiae by a vitamin feeding strategy during fed-batch process*, Springer-Verlag (2002), 其通过提述并入本文。矿物质的实例包括能够提供营养物的矿物质和矿物质盐,所述营养物包括 P、K、Mg、S、Ca、Fe、Zn、Mn 和 Cu。

[0328] **发酵产物:** 发酵产物可以是源自发酵的任何物质。发酵产物可以是,不限于,醇(例如,阿拉伯醇、丁醇、乙醇、甘油、甲醇、1,3-丙二醇、山梨醇和木糖醇);有机酸(例如,乙酸、醋酮酸、己二酸、抗坏血酸、柠檬酸、2,5-二酮-D-葡糖酸、甲酸、反丁烯二酸、葡糖二酸、葡糖酸、葡糖醛酸、戊二酸、3-羟基丙酸、衣康酸、乳酸、苹果酸、丙二酸、草酸、草酰乙酸、丙酸、琥珀酸和木糖酸);酮(例如,丙酮);氨基酸(例如,天冬氨酸、谷氨酸、甘氨酸、赖氨酸、丝氨酸和苏氨酸);和气体(例如,甲烷、氢气(H<sub>2</sub>)、二氧化碳(CO<sub>2</sub>)和一氧化碳(CO))。发酵产物还可以是作为高价值产品的蛋白质。

[0329] 在一个优选的方面,发酵产物是醇。可理解的是,术语“醇”包括包含一个或多个羟基基团的物质。在更优选的方面,所述醇是阿拉伯醇。在另一个更优选的方面,所述醇是丁醇。在另一个更优选的方面,所述醇是乙醇。在另一个更优选的方面,所述醇是甘油。在另一个更优选的方面,所述醇是甲醇。在另一个更优选的方面,所述醇是 1,3-丙二醇。在另一

个更优选的方面,所述醇是山梨醇。在另一个更优选的方面,所述醇是木糖醇。参见,例如, Gong, C. S., Cao, N. J., Du, J., 和 Tsao, G. T., 1999, Ethanol production from renewable resources, 于 *Advances in Biochemical Engineering/Biotechnology*, Scheper, T. 编, Springer-Verlag Berlin Heidelberg, Germany, 65:207-241; Silveira, M. M., 和 Jonas, R., 2002, The biotechnological production of sorbitol, *Appl. Microbiol. Biotechnol.* 59:400-408; Nigam, P., 和 Singh, D., 1995, Processes for fermentative production of xylitol—a sugar substitute, *Process Biochemistry* 30(2):117-124; Ezeji, T. C., Qureshi, N. 和 Blaschek, H. P., 2003, Production of acetone, butanol and ethanol by *Clostridium beijerinckii* BA101 and in situ recovery by gas stripping, *World Journal of Microbiology and Biotechnology* 19(6):595-603。

[0330] 在另一个优选的方面,所述物质是有机酸。在另一个更优选的方面,所述有机酸是乙酸。在另一个更优选的方面,所述有机酸是醋酮酸。在另一个更优选的方面,所述有机酸是己二酸。在另一个更优选的方面,所述有机酸是抗坏血酸。在另一个更优选的方面,所述有机酸是柠檬酸。在另一个更优选的方面,所述有机酸是 2,5-二酮-D-葡萄糖酸。在另一个更优选的方面,所述有机酸是甲酸。在另一个更优选的方面,所述有机酸是反丁烯二酸。在另一个更优选的方面,所述有机酸是葡糖二酸。在另一个更优选的方面,所述有机酸是葡糖酸。在另一个更优选的方面,所述有机酸是葡糖醛酸。在另一个更优选的方面,所述有机酸是戊二酸。在另一个优选的方面,所述有机酸是 3-羟基丙酸。在另一个更优选的方面,所述有机酸是衣康酸。在另一个更优选的方面,所述有机酸是乳酸。在另一个更优选的方面,所述有机酸是苹果酸。在另一个更优选的方面,所述有机酸是丙二酸。在另一个更优选的方面,所述有机酸是草酸。在另一个更优选的方面,所述有机酸是丙酸。在另一个更优选的方面,所述有机酸是琥珀酸。在另一个更优选的方面,所述有机酸是木糖酸。参见,例如, Chen, R., 和 Lee, Y. Y., 1997, Membrane-mediated extractive fermentation for lactic acid production from cellulosic biomass, *Appl. Biochem. Biotechnol.* 63-65:435-448。

[0331] 在另一个优选的方面,所述物质是酮。可理解的是术语“酮”包括含有一个或多个酮基的物质。在另一个更优选的方面,所述酮是丙酮。参见,例如, Qureshi 和 Blaschek, 2003, 见上文。

[0332] 在另一个优选的方面,所述物质是氨基酸。在另一个更优选的方面,所述有机酸是天冬氨酸。在另一个更优选的方面,所述氨基酸是谷氨酸。在另一个更优选的方面,所述氨基酸是甘氨酸。在另一个更优选的方面,所述氨基酸是赖氨酸。在另一个更优选的方面,所述氨基酸是丝氨酸。在另一个更优选的方面,所述氨基酸是苏氨酸。参见,例如, Richard, A., 和 Margaritis, A., 2004, Empirical modeling of batch fermentation kinetics for poly(glutamic acid) production and other microbial biopolymers, *Biotechnology and Bioengineering* 87(4):501-515。

[0333] 在另一个优选的方面,所述物质是气体。在另一个更优选的方面,所述气体是甲烷。在另一个更优选的方面,所述气体是  $H_2$ 。在另一个更优选的方面,所述气体是  $CO_2$ 。在另一个更优选的方面,所述气体是 CO。参见,例如, Kataoka, N., A. Miya, 和 K. Kiriya, 1997, Studies on hydrogen production by continuous culture system of hydrogen-producing anaerobic bacteria, *Water Science and Technology* 36(6-7):

41-47 ;和 Gunaseelan V.N. 于 Biomass and Bioenergy, Vol.13(1-2), pp. 83-114,1997, Anaerobic digestion of biomass for methane production :A review.

[0334] 回收 可以使用本领域已知的任何方法,任选地从发酵培养基回收发酵产物,所述方法包括,但不限于,层析、电泳方法、差示溶解度、蒸馏或提取。例如,通过常规蒸馏方法从发酵的纤维素材料分离并纯化醇。可以获得纯度高达约 96vol% 的乙醇,其能用作,例如,燃料乙醇、饮用乙醇,即,中性饮料酒,或工业乙醇。

[0335] 信号肽

[0336] 本发明还涉及编码信号肽的分离的多核苷酸,所述信号肽包含 SEQ IDNO :2 的氨基酸 1 至 21,或由 SEQ ID NO :2 的氨基酸 1 至 21 组成。所述多核苷酸可进一步包含编码蛋白的基因,其可操作连接于信号肽。所述蛋白优选对于所述信号肽和 / 或前肽是外源的。

[0337] 本发明还涉及包含这种多核苷酸的核酸构建体、表达载体和重组宿主细胞。

[0338] 本发明还涉及用于产生蛋白质的方法,包括:(a) 培养包含此种多核苷酸的重组宿主细胞;和 (b) 回收所述蛋白质。

[0339] 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,并且因此涵盖肽、寡肽和多肽。术语“蛋白质”还包含组合以形成编码产物的两种以上多肽。所述蛋白质还包括杂合多肽和融合多肽。

[0340] 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,或报告蛋白(reporter)。举例而言,所述蛋白质可为氧化还原酶、转移酶、水解酶、裂合酶(lyase)、异构酶或连接酶,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、 $\alpha$ -半乳糖苷酶、 $\beta$ -半乳糖苷酶、葡糖淀粉酶、 $\alpha$ -葡糖苷酶、 $\beta$ -葡糖苷酶、转化酶、漆酶、另外的脂肪酶、甘露糖苷酶、变聚糖酶(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。

[0341] 基因可以从任何原核、真核或其它来源获得。

[0342] 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。

## 实施例

[0343] 培养基

[0344] PDA 平板由 39.0 克的马铃薯右旋糖琼脂和蒸馏水加至 1 升组成。

[0345] YG 琼脂平板由 5.0g 的酵母提取物、10.0g 的葡萄糖、20.0g 的琼脂和蒸馏水加至 1 升组成。

[0346] NNCYP-PCS 培养基由 5.0g 的  $\text{NaNO}_3$ , 3.0g 的  $\text{NH}_4\text{Cl}$ , 2.0g 的 MES, 2.5g 的柠檬酸, 0.2g 的  $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$ , 1.0g 的细菌蛋白胨, 5.0g 的酵母提取物, 0.2g 的  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ , 4.0g 的  $\text{K}_2\text{HPO}_4$ , 1.0ml 的 COVE 痕量元素溶液, 2.5g 的葡萄糖, 25.0g 的 PCS, 和蒸馏水加至 1 升组成。

[0347] COVE 痕量元素溶液由 0.04g 的  $\text{Na}_2\text{B}_4\text{O}_7 \cdot 10\text{H}_2\text{O}$ , 0.4g 的  $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ , 1.2g 的  $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ , 0.7g 的  $\text{MnSO}_4 \cdot \text{H}_2\text{O}$ , 0.8g 的  $\text{Na}_2\text{MoO}_2 \cdot 2\text{H}_2\text{O}$ , 10g 的  $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ , 和蒸馏水加至 1 升组成。

[0348] LB 琼脂平板由 10g 的胰蛋白胨, 5g 的酵母提取物, 10g 的氯化钠, 15g 的琼脂, 和 1

升蒸馏水组成。

[0349] SOC培养基由2%胰蛋白胨,0.5%酵母提取物,10mM NaCl,2.5mM KCl,10mM MgCl<sub>2</sub>,10mM MgSO<sub>4</sub>,和20mM葡萄糖组成。

[0350] YPM培养基由1%酵母提取物,2%蛋白胨,和2%麦芽糖组成。

[0351] 基本培养基平板由6g的NaNO<sub>3</sub>,0.52g的KCl,1.52g的KH<sub>2</sub>PO<sub>4</sub>,1ml的COVE微量元素溶液,20g的Noble琼脂,20ml的50%葡萄糖,2.5ml的20% MgSO<sub>4</sub>·7H<sub>2</sub>O,20ml的0.02%生物素溶液,和蒸馏水加至1升组成。

[0352] 实施例1:从土壤样品分离嗜松青霉菌株 NN046877

[0353] 嗜松青霉 NN046877 是从来自中国湖南的土壤分离的,即直接将土壤样品铺板于PDA平板,接着在37°C温育5日。然后通过将菌丝体转移至YG琼脂平板上来纯化菌株。菌株 NN046877 基于形态学和分子(ITS测序)表征鉴定为嗜松青霉。

[0354] 实施例2:嗜松青霉菌株 NN046877RNA的分离

[0355] 将嗜松青霉菌株 NN46877 接种至PDA平板上,并在黑暗中在37°C温育4日。将几个菌丝体-PDA栓接种入各含有100ml NNCYP-PCS培养基的500ml摇瓶。将烧瓶在37°C以160rpm振荡温育5日。在第4日和第5日收集菌丝体,然后将来自每日的菌丝体冻结于液氮中,并在-80°C冷冻器中储藏直至使用。

[0356] 将冷冻的菌丝体转移至经液氮预冷的杵和研钵,并磨碎至细粉。通过用TRIZOL®试剂(Invitrogen, Carlsbad, CA, USA)提取从第4日和第5日的粉末状菌丝体制备总RNA。使用mTRAP™ Total Kit(Active Motif, Carlsbad, CA, USA)分离富polyA的RNA。

[0357] 实施例3:嗜松青霉菌株 cDNA文库的构建

[0358] 使用SMART™ cDNA Library Construction Kit(Takara Bio Inc., Otsu, Shiga, Japan)合成第4日和第5日的双链cDNA。使用标准方法用Sfi I切割cDNA,并通过使用44mM Tris碱,44mM硼酸,0.5mM EDTA(TBE)缓冲液的0.8%琼脂糖凝胶电泳对cDNA进行大小分级。将500bp或更大的cDNA级分从凝胶切出,并使用GFX® PCR DNA和Gel Band Purification Kit(GE Healthcare, Buckinghamshire, UK)根据生产商的指示进行纯化。然后汇集等量的来自第4日和第5日的cDNA以供文库构建。

[0359] 然后对所述cDNA进行定向克隆,即使用T4连接酶(New England Biolabs, Ipswich, MA, USA)依照生产商的指示将其连接入经Sfi I切割的pMHas7(W02009/037253)。将连接混合物使用GENE PULSER®和Pulse Controller (Bio-Rad Laboratories, Inc., Hercules, CA, USA)以25 μF, 25mA, 1.8kV和1mm间隙(gap)小杯依照生产商的指示电穿孔入大肠杆菌ELECTROMAX™DH10B™细胞(Invitrogen, Carlsbad, CA, USA)。

[0360] 将经电穿孔的细胞铺板于补充50mg卡那霉素每升的LB琼脂平板。从起始pMHas7载体连接物的总共60000个转化体制备cDNA质粒汇集物(pool)。使用QIAGEN®Plasmid Kit(QIAGEN Inc., Valencia, CA, USA)直接从菌落的汇集物制备质粒DNA。

[0361] 实施例4:构建含有β-内酰胺酶报道基因的SigA4转座子

[0362] 从含有W0 01/77315中描述的质粒的pSigA2转座子构建含有命名为pSigA4的质粒的转座子,以创建具有减少的选择背景的信号俘获转座子(signal trapping transposon)pSigA2的改进形式。所述pSigA2转座子含有在该转座子自身上编码的无信号

的  $\beta$ -内酰胺酶构建体。使用 PROOFSTART™DNA 聚合酶 (QIAGEN GmbH Corporation, Hilden, Germany) 和下述 5' 磷酸化的引物 (TAG Copenhagen, Denmark) 采用 PCR 以产生存在于质粒骨架上的完整  $\beta$ -内酰胺酶基因的缺失：

[0363] SigA2NotU-P :5' -TCGCGATCCGTTTTTCGCATTTATCGTGAAACGCT-3' (SEQ ID NO :3)

[0364] SigA2NotD-P :5' -CCGCAAACGCTGGTGAAAGTAAAAGATGCTGAA-3' (SEQ ID NO :4)

[0365] 所述扩增反应物由 1  $\mu$ l 的 pSigA2(10ng 每  $\mu$ l)、5  $\mu$ l 的 10X PROOFSTART™Buffer(QIAGEN GmbH Corporation, Hilden, Germany)、2.5  $\mu$ l 的 dNTP 混合物 (20mM)、0.5  $\mu$ l 的 SigA2NotU-P(10mM)、0.5  $\mu$ l 的 SigA2NotD-P(10mM)、10  $\mu$ l 的 Q 溶液 (QIAGEN GmbH Corporation, Hilden, Germany) 和 31.25  $\mu$ l 的去离子水。使用 DNA ENGINE™ Thermal Cycler (MJ Research Inc., Waltham, MA, USA) 进行扩增,其程序为:一个循环,在 95°C 进行 5 分钟;和 20 个循环,每循环在 94°C 进行 30 秒,62°C 进行 30 秒和 72°C 进行 4 分钟。

[0366] 通过 0.8% 琼脂糖凝胶电泳使用 40mM Tris 碱-20mM 乙酸钠-1mM EDTA 二钠 (TAE) 缓冲液和每 ml 0.1  $\mu$ g 的溴乙锭分离 3.9kb PCR 反应产物。DNA 条带用 EAGLE EYE® Imaging System (Stratagene, La Jolla, CA, USA) 协助在 360nm 显现。将 3.9kb DNA 条带从凝胶切出,并使用 GFX® PCR DNA 和 Gel Band Purification Kit 依照生产商的指示进行纯化。

[0367] 将该 3.9kb 片段在 16°C 用 10 单位 T4DNA 连接酶 (New England Biolabs, Inc., Ipswich, MA, USA)、9  $\mu$ l 的该 3.9kb PCR 片段和 1  $\mu$ l 的 10X 连接缓冲液 (New England Biolabs, Inc., Ipswich, MA, USA) 自连接过夜。将连接反应物在 65°C 热失活 10 分钟,然后用 Dpn I 在 37°C 消化 2 小时。在温育之后,使用 GFX® PCR DNA 和 Gel Band Purification Kit 纯化消化液。

[0368] 然后将纯化的物质依照生产商的指示转化入大肠杆菌 TOP10™感受态细胞 (TIANGEN Biotech (Beijing) Co. Ltd., Beijing, China)。将转化混合物铺板于补充每 ml 25  $\mu$ g 的氯霉素的 LB 琼脂平板上。从几个转化体制备质粒小量制备物,并用 Bgl II 消化。选择一个具有正确构建的质粒。将质粒命名为 pSigA4。质粒 pSigA4 含有与 WO 01/77315 中公开的相同的侧翼为 Bgl II 的转座子 SigA2。

[0369] 将质粒 pSigA4DNA (0.3  $\mu$ g/ $\mu$ l) 的 60  $\mu$ l 样品用 Bgl II 消化,并使用 TBE 缓冲液通过 0.8% 琼脂糖凝胶电泳进行分离。将 2kb 的 SigA2 转座子 DNA 条带用 200  $\mu$ l 的 EB 缓冲液 (QIAGEN GmbH Corporation, Hilden, Germany) 洗脱,并使用 GFX® PCR DNA 和 Gel Band Purification Kit 依照生产商的指示进行纯化,并在 200  $\mu$ l 的 EB 缓冲液中洗脱。将 SigA2 用于转座子协助的信号俘获。

[0370] 实施例 5:嗜松青霉菌株的转座子协助的信号俘获

[0371] 对于转座子协助的信号俘获的完整描述可见于 WO 01/77315。用转座子 SigA2 和 HYPERMU™ MuA 转座酶 (Epicenter Biotechnologies, Inc., Madison, WI, USA) 依照生产商的指示处理所述质粒汇集物。

[0372] 对于嗜松青霉 cDNA 文库的体外转座子标记,将含有大约 100ng 的 DNA 的 2  $\mu$ l SigA2 转座子与含有 1  $\mu$ g DNA 的嗜松青霉 cDNA 文库的 1  $\mu$ l 质粒 DNA 汇集物,1  $\mu$ l 的 HYPERMU™ MuA 转座酶和 2  $\mu$ l 的 10X 缓冲液 (Epicenter Biotechnologies, Inc., Madison,

WI, USA) 在总体积 20  $\mu$ l 中混合,并在 37 $^{\circ}$ C 温育 3 小时,然后添加 2  $\mu$ l 的终止缓冲液 (Epicenter Biotechnologies, Inc., Madison, WI, USA) 并在 75 $^{\circ}$ C 热失活 10 分钟。将 DNA 通过添加 2  $\mu$ l 的 3M 乙酸钠 pH 5 和 55  $\mu$ l 的 96% 乙醇沉淀,并以 10000 $\times$ g, 4 $^{\circ}$ C 离心 30 分钟。将沉淀在 70% 乙醇中洗涤,并在室温空气干燥,并重悬于 10  $\mu$ l 去离子水。

[0373] 将 2  $\mu$ l 体积的转座子标记质粒汇集物使用 GENE PULSER<sup>®</sup> 和 Pulse Controller 以 25  $\mu$ F, 25mAmp, 1.8kV 用 1mm 间隔小杯根据生产商的指示电穿孔入 50  $\mu$ l 的大肠杆菌 ELECTROMAX<sup>™</sup> DH10B<sup>™</sup> 细胞。

[0374] 将经电穿孔的细胞在 37 $^{\circ}$ C 在 SOC 培养基中以 225rpm 振荡温育 1 小时,然后铺板于下述选择培养基:补充了 50  $\mu$ g 每 ml 卡那霉素的 LB 琼脂培养基;补充了 50  $\mu$ g 每 ml 卡那霉素和 15  $\mu$ g 每 ml 氯霉素的 LB 琼脂培养基;和补充了 50  $\mu$ g 每 ml 卡那霉素、15  $\mu$ g 每 ml 氯霉素和 30  $\mu$ g 每 ml 氨苄青霉素的 LB 琼脂培养基。

[0375] 通过将电穿孔液铺板于补充了卡那霉素、氯霉素和氨苄青霉素的 LB 琼脂培养基,在 30 $^{\circ}$ C 3 日之后观察到每 50  $\mu$ l 大约 200 个菌落。将所有菌落平板复制 (replica plate) 至补充了 50  $\mu$ g 每 ml 卡那霉素、15  $\mu$ g 每 ml 氯霉素和 100  $\mu$ g 每 ml 氨苄青霉素的 LB 琼脂培养基上。在该选择条件下回收了五百个菌落。用下述所示的转座子正向和反向引物 (引物 A 和 B) 依照 WO 01/77315 公开的步骤对质粒 DNA 进行测序。

[0376] 引物 A :5' -agcggtttgcggcccgatcc-3' (SEQ ID NO :5)

[0377] 引物 B :5' -ttattcgggtcgaaaaggatcc-3' (SEQ ID NO :6)

[0378] 实施例 6 :序列汇编 (assembly) 和注释 (annotataion)

[0379] 从 SinoGenoMax Co., Ltd., Beijing, China 获取 DNA 序列。修整对于每个质粒读取的 (reads for each plasmid) 引物 A 和引物 B 序列以去除载体和转座子序列。将经汇编的序列通过使用程序 PhredPhrap (Ewing 等, 1998, Genome Research 8 :175-185; Ewing 和 Green, 1998, Genome Research 8 :186-194) 将其分组为重叠群 (contig)。随后将所有重叠群与在公共 DNA 和蛋白序列数据库 (TrEMBL, SWALL, PDB, EnsemblPep, GeneSeqP) 可获得的序列通过使用程序 BLASTX 2.0a19MP-WashU [14-Jul-1998] [Build linux-x86 18:51:44 30-Jul-1998] (Gish 等, 1993, Nat. Genet. 3 :266-72) 进行比较。通过分析 BlastX 结果直接鉴定了家族 GH61 蛋白。

[0380] 实施例 7 :从基因组 DNA 克隆嗜松青霉 GH61A 多肽基因

[0381] 将嗜松青霉 NN046877 在 37 $^{\circ}$ C 生长于 PDA 平板上 4-5 日。将菌丝体直接从琼脂平板收集至经灭菌的研钵,并在液氮下冷冻。将冷冻的菌丝体通过杵和研钵磨碎至细粉,并使用 DNEASY<sup>®</sup> Plant Mini Kit (QIAGEN Inc., Valencia, CA, USA) 分离基因组 DNA。

[0382] 基于实施例 6 中获得的 cDNA 序列,设计了如下所示的寡核苷酸引物以从嗜松青霉 NN046877 的基因组 DNA 扩增 gh61 基因。使用 IN-FUSION<sup>®</sup> CF Dry-down Cloning Kit (Clontech Laboratories, Inc., Mountain View, CA, USA) 将片段直接克隆入表达载体 pPFJ0355 (图 2), 而无需进行限制性消化和连接。

[0383] 有义引物 :

[0384] 5' -ACACAACTGGGGATCCACCATGCCTTCTACTAAAGTCGCTG-3' (SEQ ID NO :7)

[0385] 反义引物 :

[0386] 5' -GTCACCCTCTAGATCTTCAAAGGACAGTAGTGGTGATGAC-3' (SEQ ID NO :8)



[0387] 粗体字母代表编码序列,而剩余序列与 pPFJ0355 的插入位点同源。

[0388] 表达载体 pPFJ0355 含有来源于米曲霉的 TAKA-淀粉酶启动子,黑曲霉葡萄糖淀粉酶终止子元件,用于选择和大肠杆菌中繁殖的 PUC19 衍生序列,和构巢曲霉 pyrG 基因,其编码乳清酸核苷脱羧酶以供选择 pyrG 突变体曲霉属菌株的转化体。

[0389] 将上述每种引物各二十皮摩尔用于 PCR 反应,反应物由嗜松青霉 NN046877 基因组 DNA,10  $\mu$ l 的 5X GC Buffer(Finnzymes Oy, Espoo, Finland),1.5  $\mu$ l 的 DMSO, dATP、dTTP、dGTP 和 dCTP 各 2.5mM,以及 0.6 单位的 PHUSION™ High-Fidelity DNA Polymerase(Finnzymes Oy, Espoo, Finland),最终体积为 50  $\mu$ l 组成。使用 Peltier Thermal Cyclers(MJ Research Inc., South San Francisco, CA, USA) 进行扩增,程序如下:在 98°C 变性 1 分钟,5 个循环,每循环在 98°C 变性 15 秒,在 56°C 退火 30 秒,其中每循环增加 1°C,并在 72°C 延伸 75 秒;25 个循环,每循环在 98°C 进行 15 秒,65°C 进行 30 秒,和 72°C 进行 75 秒;以及在 72°C 最终延伸 10 分钟。然后加热块进入 4°C 浸泡循环。

[0390] 通过使用 TBE 缓冲液的 1.0% 琼脂糖凝胶电泳分离反应产物,其中将约 1.0kb 的产物条带从凝胶切出,并使用 Illustra GFX® PCR DNA 和 Gel Band Purification Kit(GE Healthcare, Buckinghamshire, UK) 依照生产商的指示进行纯化。

[0391] 将质粒 pPFJ0355 用 Bam HI 和 Bgl II 消化,通过使用 TBE 缓冲液的 1.0% 琼脂糖凝胶电泳进行分离,然后使用 Illustra GFX® PCR DNA 和 Gel Band Purification Kit 进行纯化。

[0392] 使用 IN-FUSION® CF Dry-down PCR Cloning Kit 将基因片段和经消化的载体连接在一起,得到 pPpin7(图 3),其中嗜松青霉 GH61A 多肽基因的转录处于来自米曲霉  $\alpha$ -淀粉酶基因的启动子的调控之下。克隆操作是依照生产商的指示进行的。简言之,用 Bam HI 和 Bgl II 消化 30ng 的 pPFJ0355,然后将 60ng 的纯化的嗜松青霉 GH61A 多肽 PCR 产物添加至反应小瓶,并通过添加去离子水重悬在 10  $\mu$ l 的最终体积中。将反应物在 37°C 温育 15 分钟,然后在 50°C 温育 15 分钟。使用三  $\mu$ l 的反应物转化大肠杆菌 TOP10™ 感受态细胞。通过菌落 PCR 检测出含有 pPpin7 的大肠杆菌转化体,并使用 QIAPREP™ Spin Miniprep Kit(QIAGEN Inc., Valencia, CA, USA) 制备质粒 DNA。插入 pPpin7 的嗜松青霉 GH61A 多肽基因通过使用 3730XL DNA Analyzer(AppliedBiosystems Inc, Foster City, CA, USA) 的 DNA 测序得到确证。

[0393] 使用 pGEM-T Vector System 将相同的 PCR 片段克隆入 pGEM-T(Promega Corporation, Madison, WI, USA) 以生成 pGEM-T-Ppin7(图 4)。插入 pGEM-T-Ppin7 的嗜松青霉 GH61A 多肽基因通过使用 3730XL DNA Analyzer 的 DNA 测序得到确证。将含有 pGEM-T-Ppin7 的大肠杆菌菌株 T-Ppin7 于 2009 年 6 月 24 日保藏于德意志微生物和细胞培养保藏中心(DSM), Braunschweig, Germany, 并分配登录号为 DSM22711。

[0394] 实施例 8:嗜松青霉 GH61A 多肽基因的特征

[0395] 就其品质审视了核苷酸序列数据,并在 PHRED/PHRAP 软件(University of Washington, Seattle, WA, USA) 的协助下将所有序列相互比较。

[0396] 嗜松青霉 GH61A 多肽基因组序列的核苷酸序列(SEQ ID NO:1)和推定的氨基酸序列(SEQ ID NO:2)示于图 1A 和 1B。基因组片段编码 322 个氨基酸的多肽,由一个 52 碱基对(核苷酸 102-153)的内含子间断。全长编码序列和成熟编码序列的 % G+C 含量分别为

53.42%和53.38%。使用SignalP软件程序(Nielsen等,1997,Protein Engineering 10:1-6),预测了21个残基的信号肽。预测的成熟蛋白含有301个氨基酸,具有32.68kDa的预测分子量。

[0397] 使用如EMBOSS的Needle程序中所执行的Needleman-Wunsch算法(Needleman和Wunsch,1970,J.Mol.Biol.48:443-453)及缺口开放罚分10、缺口延伸罚分0.5和EBLOSUM62矩阵确定氨基酸序列的比较配对全局比对。比对显示嗜松青霉GH61A多肽基因的成熟多肽的推定氨基酸序列与橙橘热子囊菌GH61A多肽基因(GeneSeq登录号AUP68836)的推定氨基酸序列享有75%同一性(排除缺口)。

[0398] 实施例9:在米曲霉中表达具有纤维素分解增强活性的嗜松青霉GH61A多肽

[0399] 依照Christensen等,1988,Bio/Technology 6:1419-1422的方法制备米曲霉HowB101(WO 95/35385)原生质体,并用三 $\mu$ g的pPpin7对其进行转化。该转化得到约50个转化体。将四个转化体分离至单独的基本培养基平板。

[0400] 将四个转化体分别接种入24孔板中的3ml YPM培养基,并在30 $^{\circ}$ C以150rpm振荡温育。在温育3日之后,通过使用含MES的NUPAGE<sup>®</sup> NOVEX<sup>®</sup> 4-12% Bis-Tris Gel(Invitrogen Corporation, Carlsbad, CA, USA)的SDS-PAGE依照生产商的指示分析来自每个培养物的20 $\mu$ l上清。将所得的凝胶用INSTANT BLUE<sup>™</sup>(Expedeon Ltd., Babraham Cambridge, UK)染色。SDS-PAGE显示大多数转化体具有约64kDa的主要条带。表达菌株命名为米曲霉EXP02769。

[0401] 将米曲霉EXP02769的斜面培养物用10ml的YPM培养基洗涤,并接种入含各400ml YPM培养基的2升烧瓶中以生成用于表征该酶的培养液。在第3日收获培养物,并使用0.45 $\mu$ m DURAPORE<sup>®</sup> Membrane(Millipore, Bedford, MA, USA)过滤。

[0402] 实施例10:具有纤维素分解增强活性的重组嗜松青霉GH61A多肽的纯化

[0403] 将1升体积的米曲霉Exp02769上清用硫酸铵(80%饱和)沉淀,并重新溶解于100ml的25mM Bis-Tris pH 6.0缓冲液,然后针对相同缓冲液进行透析,并通过0.45mm过滤器过滤;最终体积为200ml。将溶液上样于在25mMBis-Tris pH 6.0缓冲液中平衡的40ml Q SEPAHROSE<sup>®</sup> Fast Flow柱(GE Healthcare, Buckinghamshire, UK),并用线性NaCl梯度(0-0.25M)洗脱重组GH61A蛋白。如实施例9中所述通过SDS-PAGE分析来自柱的级分。汇集含有大约64kDa的条带的级分。然后通过超滤浓缩汇集的溶液。

[0404] 实施例11:具有纤维素分解增强活性的嗜松青霉GH61A多肽的浓缩和定量

[0405] 将来自实施例10的纯化样品使用10kDa MWC0 Amicon Ultracentrifuge浓缩器(Millipore, Bedford, MA, USA)进一步浓缩至大约1/10体积。将浓缩的滤过物缓冲液交换至20mM三(羟甲基)氨基甲烷pH 8.0,并使用经20ml的20mM三(羟甲基)氨基甲烷pH 8.0(Sigma Chemical Co., Inc., St. Louis, MO, USA)预平衡的BIO-GEL<sup>®</sup> P-6脱盐柱(Bio-Rad Laboratories, Inc., Hercules, CA, USA)进行脱盐,即添加3ml的样品,并用3ml的相同缓冲液洗脱。经浓缩、脱盐的嗜松青霉GH61A蛋白使用Microplate BCA<sup>™</sup> Protein Assay Kit(Thermo Fisher Scientific, Waltham, MA, USA)以0至0.8mg每ml的浓度的牛血清白蛋白(Pierce, Rockford, IL, USA)作为标准物进行定量。定量重复进行三次。使用8-16%梯度SDS-PAGE在200V进行1小时,并用Coomassie BIO-SAFE<sup>™</sup>染色剂(Bio-Rad Laboratories, Inc., Hercules, CA, USA)进行染色来确证酶纯度。

[0406] 实施例 12 :玉米秸秆的预处理

[0407] 在 U. S. Department of Energy National Renewable Energy Laboratory (NREL) 使用稀硫酸预处理玉米秸秆。将下述条件用于所述预处理 :1. 4wt% 硫酸在 165°C 和 107psi 进行 8 分钟。经预处理的玉米秸秆中的不溶于水的固形物含有 57. 5% 纤维素、4. 6% 半纤维素和 28. 4% 木质素。测定纤维素和半纤维素, 即进行两阶段硫酸水解, 随后通过使用 NREL Standard Analytical Procedure#002 的高效液相色谱对糖进行分析。在用硫酸使用 NREL Standard Analytical Procedure#003 水解纤维素和半纤维素级分之后, 通过重量分析法测定木质素。

[0408] 在使用之前, 将经预处理的玉米秸秆磨碎并用水洗涤。制备经磨碎和洗涤的经预处理的玉米秸秆 (起始干重量为 32. 35%), 即在 Cosmos ICMG 40 湿式多功能研磨器 (wet multi-utility grinder) (EssEmm Corporation, Tamil Nadu, India) 中进行研磨, 然后用去离子水反复洗涤, 并倾去上清级分。发现经磨碎和水洗涤的经预处理的玉米秸秆的干重量为 7. 114%。

[0409] 实施例 13 :具有纤维素分解增强活性的嗜松青霉 GH61A 多肽对经预处理的玉米秸秆的酶水解的作用

[0410] 使用 2. 2mL 含有 1g 的总反应质量的 96 深孔板 (Axygen, Union City, CA, USA) 进行经预处理的玉米秸秆的水解。使用含有 1mM 硫酸锰的 50mM 乙酸钠 pH 5. 0 缓冲液中的 5% 总固形物的经洗涤、预处理的玉米秸秆 (相当于 28. 75mg 纤维素每 ml) 和 4mg 每 g 纤维素的里氏木霉纤维素酶组合物 (补充了可从 Novozymes A/S, Bagsvaerd, Denmark 获得的米曲霉  $\beta$ -葡萄糖苷酶的 CELLUCLAST®; 该纤维素酶组合物在本文中在实施例中称作“里氏木霉纤维素酶组合物”) 进行水解。将具有纤维素分解增强活性的嗜松青霉 GH61A 多肽以总蛋白的 0 至 93% (w/w) 的浓度添加。使用 ALPS-300™ 平板热密封器 (plate heat sealer) (Abgene, Epsom, United Kingdom) 密封, 并在 50°C 以 150rpm 振荡温育 0-168 小时。所有的反应重复进行两次或三次。

[0411] 在温育 24 至 168 小时之间的数个时间点, 移出 100  $\mu$ l 等分试样, 并通过使用下述方案的高效液相色谱 (HPLC) 测定水解程度。

[0412] 对于 HPLC 分析, 用 0. 45  $\mu$ m MULTISCREEN® 96 孔过滤板 (Millipore, Bedford, MA, USA) 过滤样品, 并如下所述就糖含量对滤过物进行分析。稀释于 0. 005M  $\mu$ 2SO<sub>4</sub> 中的样品的糖浓度使用 4. 6x250mm AMINEX® HPX-87H 柱 (Bio-Rad Laboratories, Inc., Hercules, CA, USA) 进行测量, 即用 0. 5 % w/w 苯甲酸 -5mM H<sub>2</sub>SO<sub>4</sub> 以每分钟 0. 6ml 的流速在 65°C 洗脱 11 分钟, 并通过积分来自用纯糖样品校准的折光率检测 (CHEMSTATION®, AGILENT® 1100HPLC, Agilent Technologies, Santa Clara, CA, USA) 的葡萄糖和纤维二糖信号进行定量。所得的当量用于计算对于每个反应纤维素转化的百分率。每个水解的程度测定为总的转化的纤维素对纤维二糖 + 葡萄糖的分数, 并且不对经预处理的玉米秸秆液中存在可溶性糖进行校正。

[0413] 所有 HPLC 数据加工使用 Kaleidagraph 软件 (Synergy software, Reading, PA, USA) 进行。对测得的糖浓度就适当的稀释因子进行调整。将葡萄糖和纤维二糖用色谱法分离, 积分, 并独立地确定其各自的浓度。然而, 为了计算总转化, 将葡萄糖和纤维二糖值合并。水解分数报道为总的转化质量比, 即 [葡萄糖 + 纤维二糖]/[总纤维素]。将三次重复

的数据点平均并计算标准偏差。

[0414] 将水解分数作为嗜松青霉 GH61A 蛋白浓度的函数作图,并以 Kaleidagraph 使用修饰的饱和结合模型拟合。如图 5 中所示的数据表明通过添加具有纤维素分解增强活性的嗜松青霉 GH61A 多肽增强了通过里氏木霉纤维素酶组合物的水解。添加 0-93% (w/w) 的 GH61A 多肽使得通过里氏木霉纤维素酶组合物的水解在 3 日水解之后增加了  $23.4 \pm 1.6\%$ , 葡聚糖转化率从  $60.0 \pm 1.3\%$  变为  $73.3 \pm 2.1\%$ ; 而水解在 7 日水解之后增加了  $27.7 \pm 1.9\%$ , 葡聚糖转化率从  $71.3 \pm 1.3\%$  变为  $99.0 \pm 2.3\%$ 。

[0415] 生物材料的保藏

[0416] 下述的生物材料已经依据布达佩斯条约的条款保藏于德意志微生物和细胞培养保藏中心 (DSM), Mascheroder Weg 1B, D-38124 Braunschweig, Germany, 并给予下述的登录号:

保藏物	登录号	保藏日期
[0417] 大肠杆菌 (pGEM-T-Ppin7)	DSM 22711	2009 年 6 月 24 日

[0419] 所述菌株于下述条件下保藏: 确保在本专利申请未决期间, 依据该外国专利法律的授权的人能够获得所述培养物。所述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, 依据专利法律可以获得所述保藏物。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。

[0420] 通过下述编号段落进一步描述本发明:

[0421] [1] 具有纤维素分解增强活性的分离的多肽, 其选自下组: (a) 与 SEQ ID NO: 2 的成熟多肽具有至少 80% 序列同一性的多肽; (b) 由多核苷酸编码的多肽, 所述多核苷酸在高严格条件下与以下杂交: (i) SEQ ID NO: 1 的成熟多肽编码序列, (ii) 包含于 SEQ ID NO: 1 的成熟多肽编码序列中的 cDNA 序列, 或 (iii) (i) 或 (ii) 的全长互补链; (c) 由多核苷酸编码的多肽, 所述多核苷酸与 SEQ ID NO: 1 的成熟多肽编码序列具有至少 80% 序列同一性; (d) SEQ ID NO: 2 的成熟多肽的包含取代、缺失和 / 或插入一个或多个 (几个) 氨基酸的变体; 和 (e) (a)、(b)、(c) 或 (d) 的多肽具有纤维素分解增强活性的片段。

[0422] [2] 段 1 的多肽, 其包含氨基酸序列, 所述氨基酸序列与 SEQ ID NO: 2 的氨基酸序列具有至少 80% 同一性。

[0423] [3] 段 2 的多肽, 其包含氨基酸序列, 所述氨基酸序列与 SEQ ID NO: 2 的氨基酸序列具有至少 85% 同一性。

[0424] [4] 段 3 的多肽, 其包含氨基酸序列, 所述氨基酸序列与 SEQ ID NO: 2 的氨基酸序列具有至少 90% 同一性。

[0425] [5] 段 4 的多肽, 其包含氨基酸序列, 所述氨基酸序列与 SEQ ID NO: 2 的氨基酸序列具有至少 95% 同一性。

[0426] [6] 段 5 的多肽, 其包含氨基酸序列, 所述氨基酸序列与 SEQ ID NO: 2 的氨基酸序列具有至少 97% 同一性。

[0427] [7] 段 1 的多肽, 其包含 SEQ ID NO: 2 的氨基酸序列其具有纤维素分解增强活性的片段或由 SEQ ID NO: 2 的氨基酸序列或其具有纤维素分解增强活性的片段组成。

[0428] [8] 段 7 的多肽, 其包含 SEQ ID NO: 2 的氨基酸序列或由 SEQ ID NO: 2 的氨基酸

序列组成。

[0429] [9] 段 7 的多肽,其包含 SEQ ID NO:2 的成熟多肽或由 SEQ ID NO:2 的成熟多肽组成。

[0430] [10] 段 1 的多肽,其由多核苷酸编码,所述多核苷酸在高严格条件下与下述杂交:(i)SEQ ID NO:1 的成熟多肽编码序列,(ii)包含在 SEQ ID NO:1 的成熟多肽编码序列中的 cDNA 序列,或 (iii)(i) 或 (ii) 的全长互补链。

[0431] [11] 段 10 的多肽,其由多核苷酸编码,所述多核苷酸在非常高严格条件下与下述杂交:(i)SEQ ID NO:1 的成熟多肽编码序列,(ii)包含在 SEQ ID NO:1 的成熟多肽编码序列中的 cDNA 序列,或 (iii)(i) 或 (ii) 的全长互补链。

[0432] [12] 段 1 的多肽,其由包含核苷酸序列的多核苷酸编码,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 80% 同一性。

[0433] [13] 段 12 的多肽,其由包含核苷酸序列的多核苷酸编码,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 85% 同一性。

[0434] [14] 段 13 的多肽,其由包含核苷酸序列的多核苷酸编码,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 90% 同一性。

[0435] [15] 段 14 的多肽,其由包含核苷酸序列的多核苷酸编码,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 95% 同一性。

[0436] [16] 段 15 的多肽,其由包含核苷酸序列的多核苷酸编码,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 97% 同一性。

[0437] [17] 段 1 的多肽,其由多核苷酸编码,所述多核苷酸包含 SEQ ID NO:1 的核苷酸序列或其编码具有纤维素分解增强活性的片段的亚序列,或由 SEQ ID NO:1 的核苷酸序列或其编码具有纤维素分解增强活性的片段的亚序列组成。

[0438] [18] 段 17 的多肽,其由多核苷酸编码,所述多核苷酸包含 SEQ ID NO:1 的核苷酸序列或由 SEQ ID NO:1 的核苷酸序列组成。

[0439] [19] 段 17 的多肽,其由多核苷酸编码,所述多核苷酸包含 SEQ ID NO:1 的成熟多肽编码序列或由 SEQ ID NO:1 的成熟多肽编码序列组成。

[0440] [20] 段 1 的多肽,其中所述多肽是 SEQ ID NO:2 的成熟多肽的包含一个或多个(几个)氨基酸的取代、缺失和/或插入的变体。

[0441] [21] 段 1 的多肽,所述多肽由质粒 pGEM-T-Ppin7 中含有的多核苷酸编码,所述质粒 pGEM-T-Ppin7 包含在大肠杆菌 DSM 22711 中。

[0442] [22] 段 1-21 中任一项的多肽,其中成熟多肽是 SEQ ID NO:2 的氨基酸 22 至 322。

[0443] [23] 段 1-22 中任一项的多肽,其中成熟多肽编码序列是 SEQ ID NO:1 的核苷酸 64 至 1018。

[0444] [24] 组合物,其包含段 1-23 中任一项的多肽。

[0445] [25] 分离的多核苷酸,其编码段 1-23 中任一项的多肽。

[0446] [26] 核酸构建体或表达载体,其包含可操作地连接于一个或多个(几个)指导多肽在表达宿主中产生的调控序列的段 25 的多核苷酸。

[0447] [27] 重组宿主细胞,其包含可操作地连接于一个或多个指导多肽产生的调控序列的段 25 的多核苷酸。

[0448] [28] 产生段 1-23 中任一项的多肽的方法,其包括:(a) 在有助于所述多肽产生的条件下培养细胞,所述细胞以其野生型形式产生所述多肽;和(b) 回收所述多肽。

[0449] [29] 产生具有纤维素分解增强活性的多肽的方法,其包括:(a) 在有助于所述多肽产生的条件下培养段 27 的宿主细胞;和(b) 回收所述多肽。

[0450] [30] 转基因植物、植物部分和植物细胞,其经编码段 1-23 中任一项的多肽的多核苷酸转化。

[0451] [31] 产生具有纤维素分解增强活性的多肽的方法,其包括:(a) 在有助于所述多肽产生的条件下培养段 30 的转基因植物或植物细胞;和(b) 回收所述多肽。

[0452] [32] 产生亲本细胞的突变体的方法,其包括使编码段 1-23 中任一项的多肽的多核苷酸失活,其导致突变体与亲本细胞相比产生较少的所述多肽。

[0453] [33] 由段 32 的方法产生的突变细胞。

[0454] [34] 段 33 的突变细胞,其进一步包含编码天然或异源蛋白的基因。

[0455] [35] 产生蛋白的方法,其包括:(a) 在有助于所述蛋白产生的条件下培养段 33 或 34 的突变细胞;和(b) 回收所述蛋白。

[0456] [36] 双链抑制 RNA(dsRNA) 分子,其包含段 25 的多核苷酸的亚序列,其中 dsRNA 任选地为 siRNA 或 miRNA 分子。

[0457] [37] 段 36 的双链抑制 RNA(dsRNA) 分子,其为长约 15、16、17、18、19、20、21、22、23、24、25 或更长的双链体核苷酸。

[0458] [38] 抑制细胞中具有纤维素分解增强活性的多肽的表达的方法,其包括对细胞施用或者在细胞中表达段 36 的双链抑制 RNA(dsRNA) 分子。

[0459] [39] 通过段 38 的方法产生的细胞。

[0460] [40] 段 39 的细胞,进一步包含编码同源或异源蛋白的基因。

[0461] [41] 产生蛋白的方法,其包括:(a) 在有助于所述蛋白产生的条件下培养段 39 或 40 的细胞;和(b) 回收所述蛋白。

[0462] [42] 分离的多核苷酸,其编码信号肽,所述信号肽包含 SEQ ID NO:2 的氨基酸 1 至 21 或由 SEQ ID NO:2 的氨基酸 1 至 21 组成。

[0463] [43] 核酸构建体或表达载体,其包含可操作地连接于段 42 的多核苷酸的编码蛋白质的基因,其中所述基因对于编码所述信号肽的多核苷酸是外源的。

[0464] [44] 重组宿主细胞,其包含段 42 的多核苷酸,其中所述基因对于编码信号肽的多核苷酸是外源的。

[0465] [45] 产生蛋白的方法,其包括:(a) 在有助于所述蛋白产生的条件下培养重组宿主细胞,所述宿主细胞包含段 42 的多核苷酸,其中所述基因对于编码信号肽的多核苷酸是外源的;和(b) 回收所述多肽。

[0466] [46] 用于降解或转化纤维素材料的方法,其包括:在段 1-23 任一项的具有纤维素分解增强活性的多肽的存在下用酶组合物处理所述纤维素材料。

[0467] [47] 段 46 的方法,其中所述纤维素材料经预处理。

[0468] [48] 段 46 或 47 的方法,其中所述酶组合物包含一种或多种选自下组的酶:纤维素酶、半纤维素酶、棒曲霉素、酯酶、木质素分解酶、果胶酶、过氧化物酶、蛋白酶和膨胀素。

[0469] [49] 段 48 的方法,其中所述纤维素酶为一种或多种选自下组的酶:内切葡聚糖

酶、纤维二糖水解酶和  $\beta$ -葡糖苷酶。

[0470] [50] 段 48 的方法,其中所述半纤维素酶为一种或多种选自下组的酶:乙酰甘露聚糖酯酶、乙酰木聚糖酯酶、阿拉伯聚糖酶、阿拉伯呋喃糖苷酶、香豆酸酯酶、阿魏酸酯酶、半乳糖苷酶、葡糖醛酸糖苷酶、葡糖醛酸酯酶、甘露聚糖酶、甘露糖苷酶、木聚糖酶和木糖苷酶。

[0471] [51] 段 46-50 任一项的方法,进一步包括回收经降解的纤维素材料。

[0472] [52] 段 51 的方法,其中所述经降解的纤维素材料是糖。

[0473] [53] 段 52 的方法,其中所述糖选自下组:葡萄糖、木糖、甘露糖、半乳糖和阿拉伯糖。

[0474] [54] 产生发酵产物的方法,其包括:(a) 在段 1-23 中任一项的具有纤维素分解增强活性的多肽存在下,用酶组合物糖化纤维素材料;(b) 用一种或多种发酵微生物发酵经糖化的纤维素材料以产生发酵产物;和(c) 从所述发酵回收所述发酵产物。

[0475] [55] 段 54 的方法,其中所述纤维素材料经预处理。

[0476] [56] 段 54 或 55 的方法,其中所述酶组合物包含选自下组的一种或多种酶:纤维素酶、半纤维素酶、棒曲霉素、酯酶、木质素分解酶、果胶酶、过氧化物酶、蛋白酶和膨胀素。

[0477] [57] 段 56 的方法,其中所述纤维素酶为一种或多种选自下组的酶:内切葡聚糖酶、纤维二糖水解酶和  $\beta$ -葡糖苷酶。

[0478] [58] 段 56 的方法,其中所述半纤维素酶为一种或多种选自下组的酶:乙酰甘露聚糖酯酶、乙酰木聚糖酯酶、阿拉伯聚糖酶、阿拉伯呋喃糖苷酶、香豆酸酯酶、阿魏酸酯酶、半乳糖苷酶、葡糖醛酸糖苷酶、葡糖醛酸酯酶、甘露聚糖酶、甘露糖苷酶、木聚糖酶和木糖苷酶。

[0479] [59] 段 54-58 中任一项的方法,其中步骤(a)和(b)在同步糖化和发酵中同时进行。

[0480] [60] 段 54-59 中任一项的方法,其中发酵产物是醇、有机酸、酮、氨基酸或气体。

[0481] [61] 发酵纤维素材料的方法,其包括:用一种或多种发酵微生物发酵纤维素材料,其中在存在段 1-23 中任一项的具有纤维素分解增强活性的多肽的存在下用酶组合物糖化所述纤维素材料。

[0482] [62] 段 61 的方法,其中所述纤维素材料的发酵产生发酵产物。

[0483] [63] 段 62 的方法,其进一步包括从所述发酵回收发酵产物。

[0484] [64] 段 61-63 中任一项的方法,其中所述纤维素材料在糖化前经过预处理。

[0485] [65] 段 61-64 中任一项的方法,其中所述酶组合物包含选自下组的一种或多种酶:纤维素酶、半纤维素酶、棒曲霉素、酯酶、木质素分解酶、果胶酶、过氧化物酶、蛋白酶和膨胀素。

[0486] [66] 段 65 的方法,其中所述纤维素酶为一种或多种选自下组的酶:内切葡聚糖酶、纤维二糖水解酶和  $\beta$ -葡糖苷酶。

[0487] [67] 段 65 的方法,其中所述半纤维素酶为一种或多种选自下组的酶:乙酰甘露聚糖酯酶、乙酰木聚糖酯酶、阿拉伯聚糖酶、阿拉伯呋喃糖苷酶、香豆酸酯酶、阿魏酸酯酶、半乳糖苷酶、葡糖醛酸糖苷酶、葡糖醛酸酯酶、甘露聚糖酶、甘露糖苷酶、木聚糖酶和木糖苷酶。

[0488] [68] 段 62-67 中任一项的方法,其中发酵产物是醇、有机酸、酮、氨基酸,或气体。

[0489] 本文描述和要求保护的本发明并不局限于本文公开的具体方面的范围内,因为这些方面旨在作为本发明几个方面的说明。旨在将任何等同的方面包含于本发明的范围内。实际上,从前面的说明中,除本文所显示和描述的之外,本发明的多种修改对于本领域的技术人员来说是显而易见的。这些修改也旨在落入所附的权利要求的范围内。在冲突的情况下,将以包括定义部分的本公开为准。



[0001]

序列表

<110> 诺维信股份有限公司(Novozymes, Inc.)  
 诺维信公司(Novozymes A/S)  
 Duan, Junxin  
 Liu, Ye  
 Tang, Lan  
 Quinlan, Jason  
 Kramer, Randall  
 Wu, Wenping

<120> 具有纤维素分解增强活性的多肽和编码该多肽的多核苷酸

<130> 11629-WO-PCT

<150> US 61/223,533

<151> 2009-07-07

<160> 8

<170> PatentIn version 3.5

<210> 1

<211> 1021

<212> DNA

<213> 嗜松青霉(Penicillium pinophilum)

<400> 1

atgectteta ctaaagtgc tgeccittet getgttctag ctttggcctc cacggttget	60
ggccatggtt ttgtcaaaa catcgttate gacggtaaat cgtaagcagt gatgcatcca	120
ttattaaact agacatgctt acaaaaaaat cagttactct ggataccttg tgaatcagtt	180
ccccctacag tccaaccac cagctgttat tgggtgggca acaactgcaa ccgacctggg	240
attgctgct cccagtgagt acaccaatge agacattate tgccacaaga acgccacacc	300
tggcgcgctt tctgctccag ttgctgcagg gggcactgtc gagctccagt ggactacatg	360
gcccgatagt catcacggte ctgtcatcag ctacctgccc aactgcaatg gcaattgttc	420
taccgtgat aagactaage tagactttgt caagattgac caaggtggtt tgatecacga	480
tactaccccc cgggtacat gggttccga caaacttate gtgccaaca acagctggac	540
tgtaactate cctccacca tegegcctgg aaactacgtt ttgegccacg aaatcattgc	600
tcttactcc gctgaaaacg cagacgggtc caaaactac cctcaatgca tcaacttggg	660
gateaccgce agcggaaacc egectccctc tggtaaccgt ggcgaaaage tctacacctc	720
tactgacccc ggtatcttgg teaatatcta ccaatccttg tcgacctacg ttattcccgg	780
accaactctg tggageggtg ctgccaatgg cgtgttgc actggtttctg ctactgoggt	840
tgtctgact gccactgctt ctgcgaccgc tactcctacc acacltltla cctctgtcgc	900
tccagettca tetaccttg ccaactgtgt tgtgaccact gtcgctctg cagtaactga	960
tgtcgtgact gtcaccgatg tagttaccgt gaccaccgtc atcaccacta ctgtcctttg	1020
a	1021

<210> 2

<211> 322

<212> PRT

<213> 嗜松青霉(Penicillium pinophilum)

<400> 2

Met Pro Ser Thr Lys Val Ala Ala Leu Ser Ala Val Leu Ala Leu Ala

[0002]

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

[0003]

Val	Thr	Val	Thr	Asp	Val	Val	Thr	Val	Thr	Thr	Val	Ile	Thr	Thr	Thr	
305					310					315					320	
Val Leu																
<210>	3															
<211>	34															
<212>	DNA															
<213>	大肠杆菌( <i>Escherichia coli</i> )															
<400>	3															
	tgcgatccg	ttttgcatt	tatcgtgaaa	cgct												34
<210>	4															
<211>	33															
<212>	DNA															
<213>	大肠杆菌( <i>Escherichia coli</i> )															
<400>	4															
	ccgcaaacgc	tggtgaaagt	aaaagatgct	gaa												33
<210>	5															
<211>	20															
<212>	DNA															
<213>	嗜松青霉( <i>Penicillium pinophilum</i> )															
<400>	5															
	agcgtttgcg	gccgcgatcc														20
<210>	6															
<211>	21															
<212>	DNA															
<213>	嗜松青霉( <i>Penicillium pinophilum</i> )															
<400>	6															
	ttattcggtc	gaaaaggatc	c													21
<210>	7															
<211>	41															
<212>	DNA															
<213>	嗜松青霉( <i>Penicillium pinophilum</i> )															
<400>	7															
	acacaactgg	ggatccacca	tgccttctac	taaagtcgct	g											41
<210>	8															
<211>	40															
<212>	DNA															
<213>	嗜松青霉( <i>Penicillium pinophilum</i> )															
<400>	8															
	gtcacctct	agatcttcaa	aggacagtag	tggatgatgc												40

```

      M P S T K V A A L S A V L A L A S T V A
1  ATGCCTTCTA CTAAAGTCGC TGCCTTTCT GCTGTTCTAG CTTGGCCTC CACGGTTGCT
      G H G F V Q N I V I D G K S
61  GGCCATGGTT TTGTGCAAAA CATCGTTATC GACGGTAAAT CGTAAGCAGT GATGCATCCA
                                     Y S G Y L V N Q F ·
121 TTATTAAACT AGACATGCTT ACAAAAAAT CAGTTACTCT GGATACCTTG TGAATCAGTT
      · P Y E S N P P A V I G W A T T A T D L G ·
181 CCCCTACGAG TCCAACCCAC CAGCTGTTAT TGGGTGGGCA ACAACTGCAA CCGACCTGGG
      · F V A P S E Y T N A D I I C H K N A T P ·
241 ATTCGTCGCT CCCAGTGAGT ACACCAATGC AGACATTATC TGCCACAAGA ACGCCACACC
      · G A L S A P V A A G G T V E L Q W T T W ·
301 TGGCGCGCTT TCTGCTCCAG TTGTGCAGG GGGCACTGTC GAGCTCCAGT GGACTACATG
      · P D S H H G P V I S Y L A N C N G N C S ·
361 GCCCATAGT CATCACGGTC CTGTCAATCAG CTACCTCGCC AACTGCAATG GCAATGTGTC
      · T V D K T K L D F V K I D Q G G L I D D ·
421 TACCGTGGAT AAGACTAAGC TAGACTTTGT CAAGATTGAC CAAGGTGGTT TGATCGACGA
      · T T P P G T W A S D K L I A A N N S W T ·
481 TACTACCCCC CCGGTACAT GGGCTCCGA CAACTTATC GGTGCCAACA ACAGCTGGAC
      · V T I P S T I A P G N Y V L R H E I I A ·
541 TGTAATATC CCCTCCACCA TCGCGCCTGG AACTACGTT TTGCGCCACG AAATCATTGC
      · L H S A G N A D G A Q N Y P Q C I N L E ·
601 TCTTCACTCC GCTGGAAACG CAGACGGTGC CCAAACTAC CCTCAATGCA TCAACTGGA
      · I T G S G T A A P S G T A G E K L Y T S ·
661 GATCACCGGC AGCGGAACCG CCGCTCCCTC TGGTACCGCT GCGGAAAAGC TCTACACCTC
      · T D P G I L V N I Y Q S L S T Y V I P G ·
721 TACTGACCCC GGTATCTTGG TCAATATCTA CCAATCCTTG TCGACCTACG TTATTCCCGG
      · P T L W S G A A N G A V A T G S A T A V ·
781 ACCAACTCTG TGGAGCGGTG CTGCCAATGG CGCTGTTGCC ACTGGTTCTG CTAATGCGGT
      · A T T A T A S A T A T P T T L V T S V A ·
841 TGCTACGACT GCCACTGCTT CTGCGACCGC TACTCCTACC AACTTGTTA CCTCTGTGCG
      · P A S S T F A T A V V T T V A P A V T D ·
901 TCCAGCTTCA TCTACCTTTG CCACTGCTGT TGTGACCACT GTCGCTCCTG CAGTAACTGA
      · V V T V T D V V T V T T V I T T T V L * ·
961 TGTCGTGACT GTCACCGATG TAGTTACCGT GACCACCGTC ATCACCCTA CTGTCCTTTG
      · *
1021 A

```

图 1

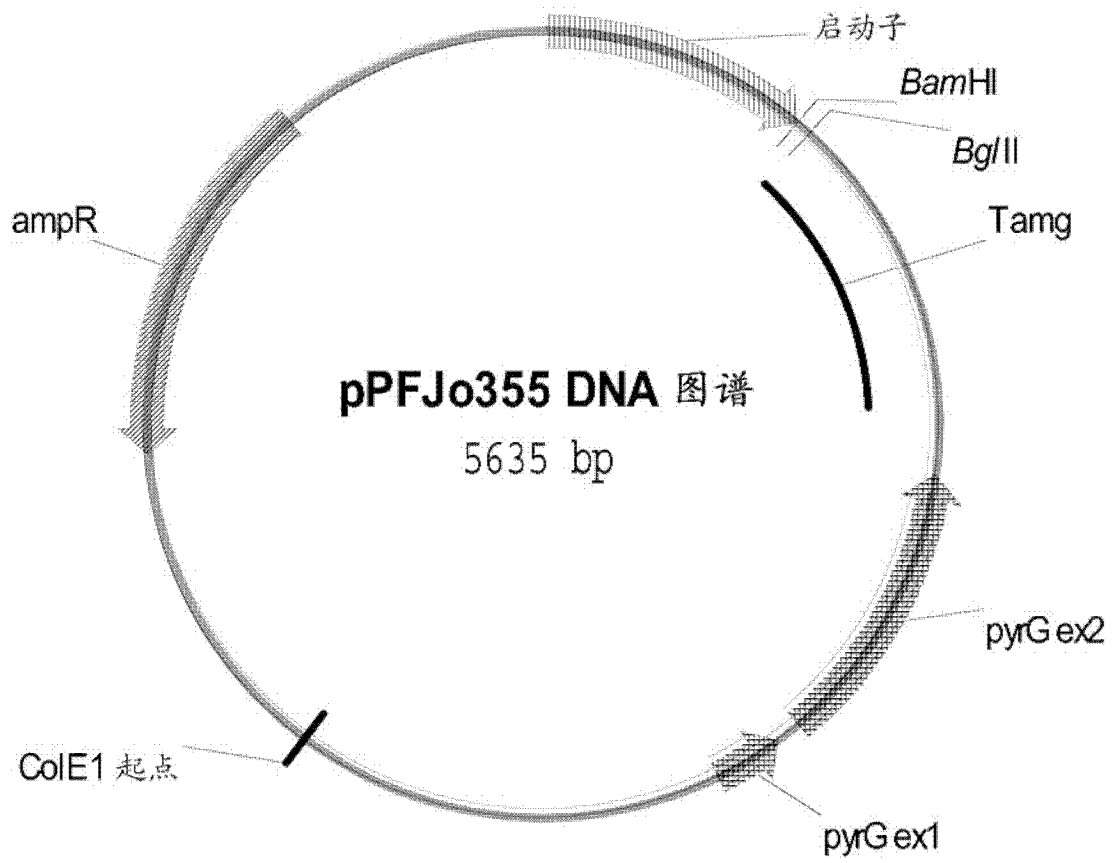


图 2

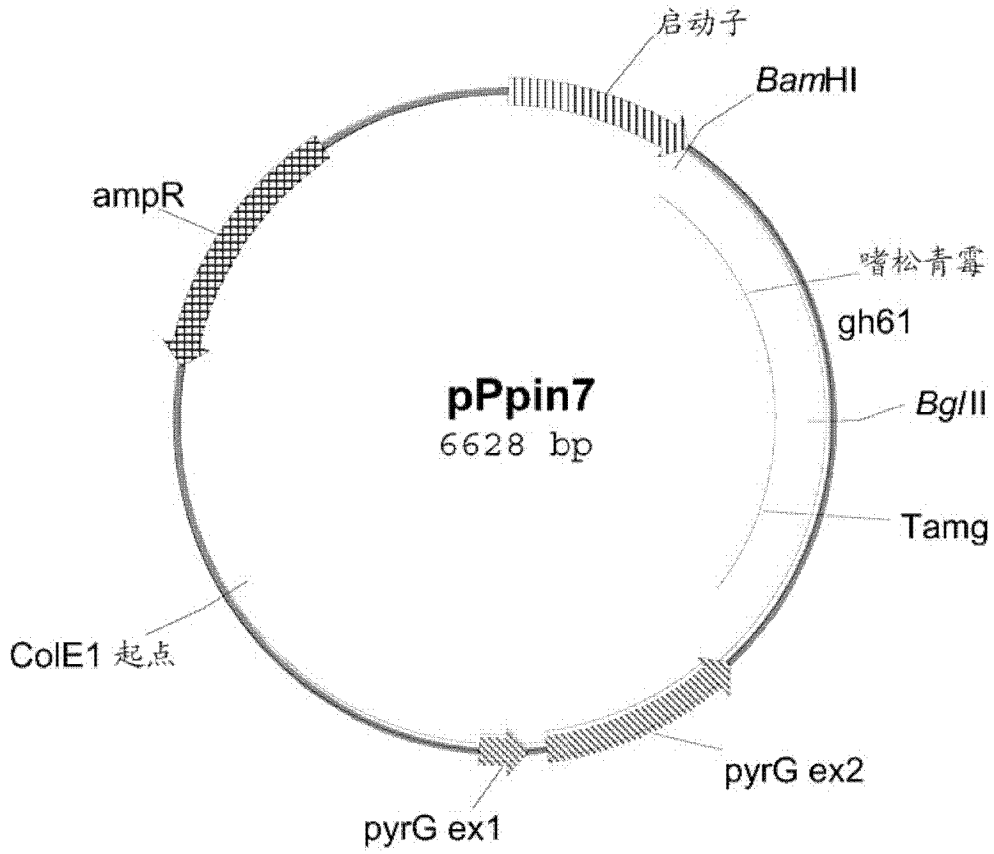


图 3

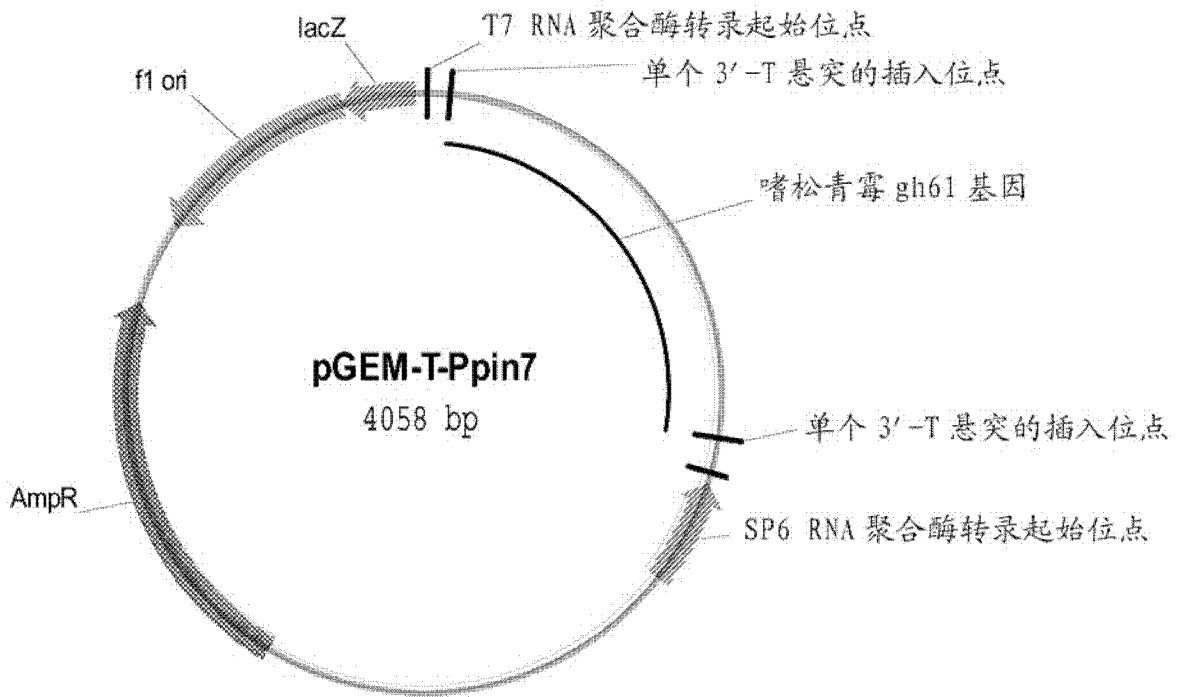


图 4

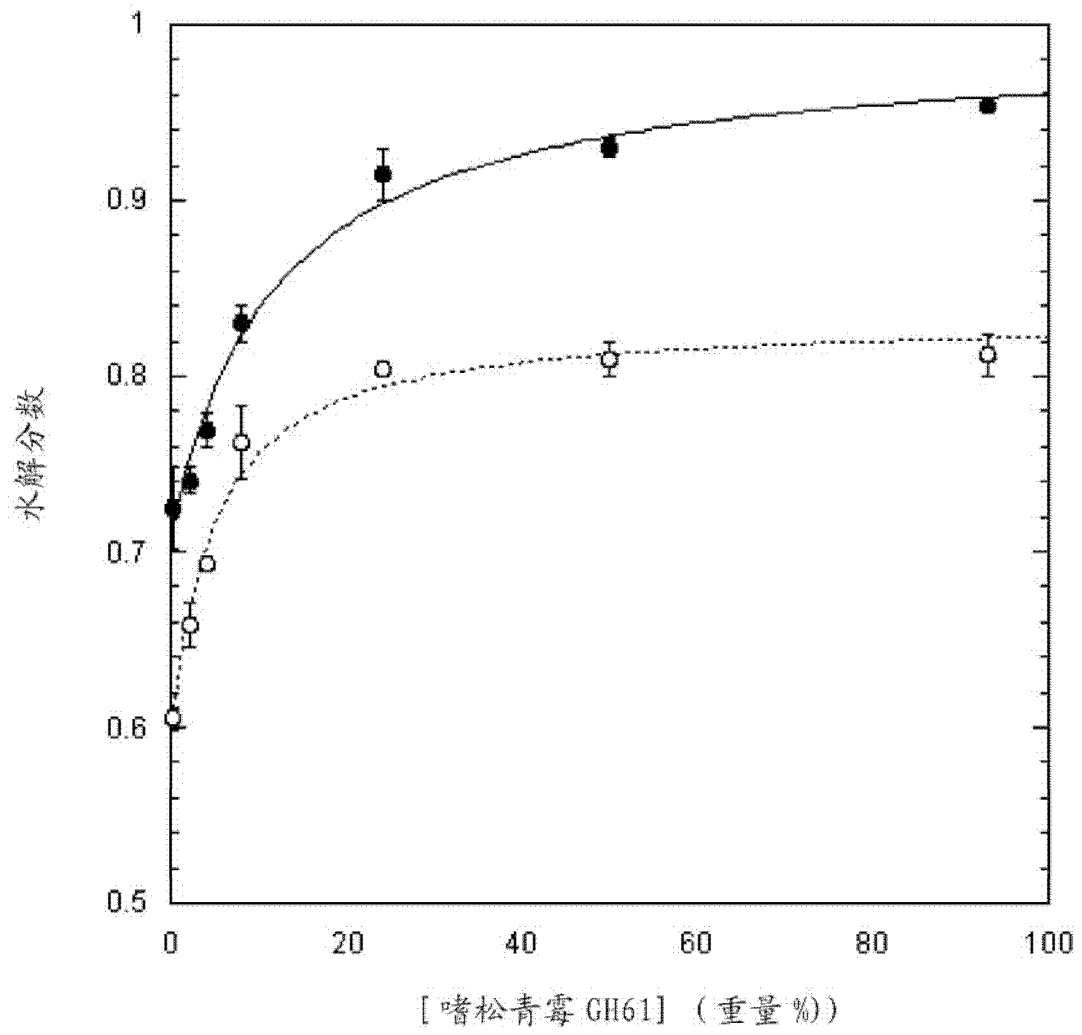


图 5