



(12) 发明专利申请

(10) 申请公布号 CN 102918151 A

(43) 申请公布日 2013.02.06

(21) 申请号 201180027233.5

(22) 申请日 2011.03.29

(30) 优先权数据

61/319,672 2010.03.31 US

(85) PCT申请进入国家阶段日

2012.11.30

(86) PCT申请的申请数据

PCT/US2011/030352 2011.03.29

(87) PCT申请的公布数据

W02011/123450 EN 2011.10.06

(71) 申请人 诺维信股份有限公司

地址 美国加利福尼亚州

(72) 发明人 M. 沃古利斯

(74) 专利代理机构 北京市柳沈律师事务所

11105

代理人 谢梦欣

(51) Int. Cl.

C12N 9/42 (2006.01)

权利要求书 3 页 说明书 62 页

序列表 5 页 附图 2 页

(54) 发明名称

纤维二糖水解酶变体及编码其的多核苷酸

(57) 摘要

本发明涉及亲本纤维二糖水解酶的变体。本发明还涉及编码所述纤维二糖水解酶变体的多核苷酸；包含所述多核苷酸的核酸构建体、载体和宿主细胞；和使用所述纤维二糖水解酶变体的方法。

1. 一种亲本纤维二糖水解酶的分离的变体,其在对应于SEQ ID NO:2的位置254、285、286、330、342和360的一个或多个(几个)位置包含取代,其中所述变体具有纤维二糖水解酶活性。

2. 权利要求1的变体,其中所述亲本纤维二糖水解酶是:

(a) 多肽,其与SEQ ID NO:2的成熟多肽具有至少60%序列同一性;

(b) 多肽,其由多核苷酸编码,所述多核苷酸在低严格条件下与以下杂交:(i)SEQ ID NO:1的成熟多肽基因组编码序列,(ii)包含于SEQ ID NO:1的成熟多肽cDNA编码序列,或(iii)(i)或(ii)的全长互补链;

(c) 多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:1的成熟多肽基因组编码序列,或包含于SEQ ID NO:1的成熟多肽cDNA编码序列具有至少60%同一性;或

(d)SEQ ID NO:2的成熟多肽的片段,其具有纤维二糖水解酶活性。

3. 权利要求1或2的变体,其中所述亲本纤维二糖水解酶包含SEQ IDNO:2的成熟多肽或由SEQ ID NO:2的成熟多肽组成。

4. 权利要求1-3任一项的变体,其中对应于位置254的取代是用Leu的取代,对应于位置285的取代是用Ile的取代,对应于位置286的取代是用Gln的取代,对应于位置330的取代是用Asn的取代,对应于位置342的取代是用Phe的取代,而对应于位置360的取代是用Gly的取代。

5. 权利要求1-3任一项的变体,其中对应于位置254的取代是C254L,对应于285的取代是L285I,对应于位置286的取代是G286Q,对应于位置330的取代是D330N,对应于位置342的取代是M342F,而对应于位置360的取代是S360G。

6. 权利要求1-5任一项的变体,进一步在对应于SEQ ID NO:2的位置245、382、420、437和440的一个或多个(几个)位置包含取代。

7. 权利要求6的变体,其中对应于位置245的取代是用Ser的取代,对应于位置382的取代是用Cys的取代,对应于位置420的取代是用Ile的取代,对应于位置437的取代是用Gln的取代,对应于位置440的取代是用Cys的取代。

8. 权利要求6的变体,其中对应于位置245的取代是A245S,对应于位置382的取代是G382C,对应于位置420的取代是L420I,对应于位置437的取代是T437Q,而对应于位置440的取代是Q440C。

9. 权利要求1-8任一项的变体,其相对于亲本纤维二糖水解酶具有改善的热稳定性。

10. 权利要求2-9任一项的变体,其中SEQ ID NO:2的成熟多肽是SEQ IDNO:2的氨基酸20-454。

11. 权利要求2-9任一项的变体,其中SEQ ID NO:1的成熟多肽编码序列是SEQ ID NO:1的核苷酸58-1710。

12. 一种具有纤维二糖水解酶活性的分离的多肽,其中所述多肽:

(a) 与SEQ ID NO:2的成熟多肽具有至少60%序列同一性;

(b) 由多核苷酸编码,所述多核苷酸在低严格条件下与以下杂交:(i)SEQID NO:1的成熟多肽基因组编码序列,(ii)包含于SEQ ID NO:1的成熟多肽cDNA编码序列,或(iii)(i)或(ii)的全长互补链;

(c) 由多核苷酸编码,所述多核苷酸与SEQ ID NO:1的成熟多肽基因组编码序列或包含

于 SEQ ID NO:1 的成熟多肽 cDNA 编码序列具有至少 60% 同一性 ;或

(d) 是 SEQ ID NO:2 的成熟多肽的片段 ;

且其中多肽氨基酸序列的一个或多个 (几个) 位置与 SEQ ID NO:2 的相应位置 254、285、286、330、342 和 360 不同。

13. 权利要求 12 的多肽,其中对应于 SEQ ID NO:2 的位置 254 的位置是 Leu,对应于 SEQ ID NO:2 的位置 285 的位置是 Ile,对应于 SEQ ID NO:2 的位置 286 的位置是 Gln,对应于 SEQ ID NO:2 的位置 330 的位置是 Asn,对应于 SEQ ID NO:2 的位置 342 的位置是 Phe,或对应于 SEQ ID NO:2 的位置 360 的位置是 Gly。

14. 权利要求 12 或 13 的多肽,进一步在对应于 SEQ ID NO:2 的位置 245、382、420、437 和 440 的一个或多个 (几个) 位置包含取代。

15. 权利要求 14 的多肽,其中对应于 SEQ ID NO:2 的位置 245 的位置是 Ser,对应于 SEQ ID NO:2 的位置 382 的位置是 Cys,对应于 SEQ ID NO:2 的位置 420 的位置是 Ile,对应于 SEQ ID NO:2 的位置 437 的位置是 Gln,或对应于 SEQ ID NO:2 的位置 440 的位置是 Cys。

16. 权利要求 11-15 任一项的多肽,其相对于 SEQ ID NO:2 的成熟多肽具有改善的热稳定性。

17. 一种分离的多核苷酸,其编码权利要求 1-11 任一项的变体或权利要求 12-16 任一项的多肽。

18. 一种核酸构建体,其包含权利要求 17 的多核苷酸。

19. 一种表达载体,其包含权利要求 17 的多核苷酸。

20. 一种宿主细胞,其包含权利要求 17 的多核苷酸。

21. 一种产生亲本纤维二糖水解酶的变体的方法,其包括 :

(a) 在适于表达所述变体的条件下培养权利要求 20 的宿主细胞 ;和

(b) 回收所述变体。

22. 一种用于获得权利要求 1-11 任一项的变体的方法,其包括 :

(a) 在对应于 SEQ ID NO:2 的位置 254、285、286、330、342 和 360 的一个或多个 (几个) 位置将取代导入亲本纤维二糖水解酶,其中所述变体具有纤维二糖水解酶活性 ;和

(b) 回收所述变体。

23. 权利要求 22 的方法,进一步包括在对应于 SEQ ID NO:2 的位置 245、382、420、437 和 440 的一个或多个 (几个) 位置将取代导入亲本纤维二糖水解酶。

24. 一种酶组合物,其包含权利要求 1-11 任一项的变体或权利要求 12-16 任一项的多肽。

25. 权利要求 24 的酶组合物,进一步包括选自下组的纤维素分解酶 :内切葡聚糖酶、纤维二糖水解酶和  $\beta$ - 葡糖苷酶,并且任选地进一步包含具有纤维素分解增强活性的多肽、木聚糖酶、半纤维素酶、酯酶、蛋白酶、漆酶或过氧化物酶。

26. 一种用于降解或转化纤维素材料的方法,其包括 :用权利要求 24 或 25 的酶组合物处理所述纤维素材料。

27. 一种用于从纤维素材料产生发酵产物的方法,其包括 :

(a) 用权利要求 24 或 25 的酶组合物糖化所述纤维素材料以产生经糖化的纤维素材料 ;

- (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) 回收所述发酵产物。

28. 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所述纤维素材料,其中所述纤维素材料用权利要求 24 或 25 的酶组合物糖化。

## 纤维二糖水解酶变体及编码其的多核苷酸

[0001] 关于对在联邦资助的研究和开发下完成的发明的权利的声明

[0002] 本发明是在能源部授予的合作协议 DE-FC36-08G018080 下以政府支持完成的。政府对本发明具有一定权利。

[0003] 涉及序列列表

[0004] 本申请包含计算机可读形式的序列列表,其通过提述并入本文。

[0005] 发明背景

### 技术领域

[0006] 本发明涉及纤维二糖水解酶的变体,编码所述变体的多核苷酸,产生所述变体的方法和使用所述变体的方法。

### 背景技术

[0007] 纤维素是单糖葡萄糖通过  $\beta$ -1, 4- 键共价连接的聚合物。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,使其暴露于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。

[0008] 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。一旦将木素纤维素转化成可发酵糖,例如葡萄糖,可发酵糖容易地由酵母发酵成乙醇。

[0009] WO 2006/074005 公开了红褐肉座菌 (*Hypocrea jecorina*) 纤维二糖水解酶 II 的变体。Heinzelman 等, 2009, Proceedings of the National Academy of Sciences USA 106:5610-5615 公开了由结构导向的重组所构建的热稳定性真菌纤维素酶的家族。Heinzelman 等, 2009, Journal of Biological Chemistry 284, 26229-26233 公开了有助于真菌纤维素酶稳定性的单个突变。

[0010] 在本领域中,提高具有纤维二糖水解酶活性的多肽增强木素纤维素原料的酶降解的能力是有利的。

[0011] 本发明提供了与其亲本相比具有改善的特性的亲本纤维二糖水解酶的变体。

### 发明内容

[0012] 本发明涉及亲本纤维二糖水解酶的分离的变体,其在一个或多个 (几个) 对应于 SEQ ID NO:2 的位置 254、285、286、330、342 和 360 的位置处包含取代,其中所述变体具有纤维二糖水解酶活性。在一个方面,所述分离的变体进一步在对应于 SEQ ID NO:2 的位置 245、382、420、437 和 440 的一个或多个 (几个) 位置包含取代。

[0013] 本发明还涉及编码所述变体的分离的多核苷酸；包含所述多核苷酸的核酸构建体、载体和宿主细胞；和产生所述变体的方法。

[0014] 本发明还涉及植物，其包含编码此种具有纤维二糖水解酶活性的变体的分离的多核苷酸。

[0015] 本发明还涉及产生此种具有纤维二糖水解酶活性的变体的方法，包括：(a) 在有助于产生所述变体的条件下培养包含编码具有纤维二糖水解酶活性的变体的多核苷酸的转基因植物或植物细胞；和 (b) 回收所述变体。

[0016] 本发明还涉及降解或转化纤维素材料的方法，包括：在此种具有纤维二糖水解酶活性的变体的存在下用酶组合物处理纤维素材料。在一个方面，所述方法进一步包括回收经降解或转化的纤维素材料。

[0017] 本发明还涉及产生发酵产物的方法，包括：(a) 在此种具有纤维二糖水解酶活性的变体的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步骤 (a) 中经糖化的纤维素材料以产生发酵产物；和 (c) 从所述发酵回收发酵产物。

[0018] 本发明还涉及发酵纤维素材料的方法，包括：用一种或多种发酵微生物发酵纤维素材料，其中将所述纤维素材料在此种具有纤维二糖水解酶活性的变体的存在下用酶组合物糖化。

#### 附图说明

[0019] 图 1A 显示野生型烟曲霉 (*Aspergillus fumigatus*) 家族 GH6A 纤维二糖水解酶和几种单独的烟曲霉家族 GH6A 纤维二糖水解酶的变体的残余活性的比较。

[0020] 图 1B 显示野生型烟曲霉家族 GH6A 纤维二糖水解酶和几种组合的烟曲霉家族 GH6A 纤维二糖水解酶的变体的残余活性的比较。

[0021] 图 2 显示野生型烟曲霉家族 GH6A 纤维二糖水解酶和几种组合的烟曲霉家族 GH6A 纤维二糖水解酶的变体的水解活性的比较。

[0022] 发明详述

[0023] 本发明涉及亲本纤维二糖水解酶的分离的变体，其在一个或多个（几个）对应于 SEQ ID NO:2 的位置 254、285、286、330、342 和 360 的位置处包含取代，其中所述变体具有纤维二糖水解酶活性。在一些方面，所述变体具有改善的热稳定性。

[0024] 定义

[0025] 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. 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 van Tilbeurgh 等, 1982, FEBS Letters, 149:152-156 和 van Tilbeurgh 和 Claeyssens, 1985 描述的方法使用荧光二糖衍生物 4-甲基伞形基- $\beta$ -D-乳糖苷确定纤维二糖水解酶活性。“纤维二糖水解酶活性”意指由纤维二糖水解酶催化的对纤维素材料的水

解活性。

[0026] 纤维素分解活性：术语“纤维素分解活性”意指水解纤维素材料的生物活性。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。

[0027] 就本发明而言，纤维素分解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 的纤维素分解蛋白 /g 的 PCS 中纤维素在 50-65°C 进行 3-7 日，与未添加纤维素分解蛋白的对照水解相比较。典型条件为：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  $MnSO_4$ ，50-65°C，72 小时，通过 **AMINEX®** HPX-87H 柱 (Bio-Rad Laboratories, Inc., Hercules, CA, USA) 进行糖分析。

[0028] 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-葡聚糖或木葡聚糖和含有纤维素组分的其它植物材料中的  $\beta$ -1, 4 键的内水解 (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，使用羧甲基纤维素 (CMC) 作为底物来确定内切葡聚糖酶活性。

[0029]  $\beta$ -葡糖苷酶：术语“ $\beta$ -葡糖苷酶”意指  $\beta$ -D-葡糖苷葡糖水解酶 (beta-D-glucoside glucohydrolase) (E. C. 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 40°C，pH 5，在含有 0.01% **TWEEN®** 20 的 100mM 柠檬酸钠中从 1mM 对硝基苯基- $\beta$ -D-葡糖吡喃糖苷每分钟产生 1.0 微摩尔对硝基苯酚。

[0030] 纤维素分解增强活性：术语“纤维素分解增强活性”意指由 GH61 多肽催化的生物活性，该多肽增强具有纤维素分解活性的酶对纤维素材料的水解。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 中纤维素，其中总蛋白的构成为 50-99.5%w/w 的纤维素分解蛋白，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 50-65°C 进行 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 3% 的米曲霉 (*Aspergillus oryzae*)  $\beta$ -葡糖苷酶 (根据 W002/095014 在米曲霉中重组产生) 或者总蛋白重量的 3% 的烟曲霉 (*Aspergillus fumigatus*)  $\beta$ -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。

[0031] 具有纤维素分解增强活性的 GH61 多肽具有 GH61 多肽的成熟多肽的纤维素分解增强活性的至少 20%, 优选至少 40%, 更优选至少 50%, 更优选至少 60%, 更优选至少 70%, 更优选至少 80%, 甚至更优选至少 90%, 最优选至少 95%, 并且甚至最优选至少 100%。

[0032] 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程度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。

[0033] 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 glycosylhydrolases, *Biochem. J.* 316:695-696 属于糖苷水解酶家族 61 的多肽。

[0034] 家族 6 糖苷水解酶: 术语“家族 6 糖苷水解酶”或“家族 GH6”或“GH6”意指根据 Henrissat B., 1991, 见上及 Henrissat B., 和 Bairoch A., 1996, 见上而属于糖苷水解酶家族 6 的多肽。

[0035] 木聚糖降解活性: 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本方法包括: (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 (内切木聚糖酶、 $\beta$ -木糖苷酶、阿拉伯呋喃糖苷酶、 $\alpha$ -葡糖醛酸糖苷酶、乙酰木聚糖酯酶、阿魏酸酯酶和  $\alpha$ -葡糖醛酸酯酶 ( $\alpha$ -glucuronoyl esterase))。最近在本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。

[0036] 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。



[0037] 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典型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,5mg/ml 底物(总固形物),5mg 木聚糖分解蛋白/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。

[0038] 木聚糖酶:术语“木聚糖酶”意指催化木聚糖中 1, 4- $\beta$ -D-木糖苷键的内水解的 1, 4- $\beta$ -D-木聚糖-木糖水解酶(E. C. 3. 2. 1. 8)。就本发明而言,木聚糖酶活性使用桦木木聚糖作为底物确定。一个单位的木聚糖酶定义为从 2g 桦木木聚糖每升作为底物在含有 0.01% **TWEEN®** 20 的 50mM 乙酸钠中在 50 °C, pH 5 在水解的起始阶段中每分钟产生 1.0 微摩尔还原糖(如 Lever, 1972, A new reaction for colorimetric determination of carbohydrates, Anal. Biochem 47:273-279 所述作为葡萄糖当量测量)。

[0039]  $\beta$ -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-xylosidexylohydrolase)(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 从含有 0.01% **TWEEN®** 20 的 100mM 柠檬酸钠中的 1mM 对硝基苯基- $\beta$ -D-木糖苷作为底物每分钟产生 1.0 微摩尔对硝基苯酚。

[0040] 乙酰木聚糖酯酶:术语“乙酰木聚糖酯酶”意指羧基酯酶(EC 3. 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。

[0041] 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)、羟基肉桂酰酯酶(hydroxycinnamoyl esterase)、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 0.5mM 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 微摩尔对硝基苯酚阴离子的酶量。

[0042]  $\alpha$ -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)(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 微摩尔葡糖醛酸或 4-O-甲基葡糖醛酸的酶量。

[0043]  $\alpha$ -L-阿拉伯呋喃糖苷酶:术语“ $\alpha$ -L-阿拉伯呋喃糖苷酶活性”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 的中等粘度小麦阿拉伯木聚糖 (Megazyme International Ireland, Ltd., Bray, Co. Wicklow, Ireland) 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。

[0044] 纤维素材料: 所述纤维素材料可为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键合, 其帮助稳定细胞壁基质。

[0045] 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, 纤维素可以是木素纤维素的形式, 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。

[0046] 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。

[0047] 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 (miscanthus)。在另一个方面, 纤维素材料是甘蔗渣。

[0048] 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经磷酸处理的纤维素。在另一个方面, 纤维素材料是滤纸。

[0049] 纤维素材料可以按原样 (as is) 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,纤维素材料经预处理。

[0050] 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。

[0051] 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) 连接木糖残基骨架的植物细胞壁多肽的材料。陆生植物的木聚糖是具有  $\beta$ -(1-4)-D-吡喃木糖骨架(其由短糖链分支)的杂聚物。其包含 D-葡萄糖醛酸或其 4-O-甲基醚、L-阿拉伯糖和/或多种寡糖,由 D-木糖、L-阿拉伯糖, D-或 L-半乳糖和 D-葡萄糖组成。木聚糖类型的多糖可分为同木聚糖(homoxylan)和杂木聚糖(heteroxylan),其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(complex heteroxylan)。参见例如 Ebringerova 等,2005, Adv. Polym. Sci. 186:1-67。

[0052] 在本发明的方法中,可使用任何含木聚糖材料。在一个优选的方面,所述含木聚糖材料是木素纤维素。

[0053] 变体:术语“变体”意指具有纤维二糖水解酶活性的多肽,其在一个或多个(几个)位置包含改变,即取代、插入和/或缺失。取代意指用不同的氨基酸替代占据某位置的氨基酸;缺失意指去除占据某位置的氨基酸;而插入意指邻接于占据某位置的氨基酸添加 1-3 个氨基酸。

[0054] 突变体:术语“突变体”意指编码变体的多核苷酸。

[0055] 野生型:术语“野生型”(例如用于野生型蛋白、野生型多肽或野生型纤维二糖水解酶)描述其表示的多肽由天然存在的微生物,如见于自然界的细菌、酵母或丝状真菌表达。

[0056] 亲本或亲本纤维二糖水解酶:术语“亲本”或“亲本纤维二糖水解酶”意指对其进行了改变以产生本发明的纤维二糖水解酶变体的纤维二糖水解酶。所述亲本可为天然存在的(野生型)多肽或其变体。举例而言,亲本多肽可为野生型多肽的变体。亲本亦可为等位变体,其是由任何占据相同染色体位置的基因的两种或更多种可选形式编码的多肽。

[0057] 分离的:术语“分离的”和“纯化的”意指从其天然相关的至少一种组分移出的多肽或多核苷酸。例如,变体或多肽如通过 SDS-PAGE 测定的,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,和至少 95% 纯;而多核苷酸如通过琼脂糖电泳测定的,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,和至少 95% 纯。

[0058] 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,基于 Sigcleave 程序预测 SEQ ID NO:2 的氨基酸 1 至 19 是信号肽(von Heijne G., 1986, Nucleic Acids Res. 14:4683-4690),成熟多肽是 SEQ ID NO:2 的氨基酸 20 至 454。

[0059] 成熟多肽编码序列:术语“成熟多肽编码序列”意指编码具有纤维二糖水解酶活性的成熟多肽的多核苷酸(例如, cDNA 或基因组 DNA)。在一个方面,基于 Sigcleave 程序预测 SEQ ID NO:1 的核苷酸 1 至 57 编码信号肽(von Heijne G., 1986, 见上),成熟多肽编码序列是 SEQ ID NO:1 的核苷酸 58 至 1710。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 58 至 1710 中的 cDNA 序列。

[0060] 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。

[0061] 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EMBOSS软件包(EMBOSS:The European Molecular Biology Open Software Suite,Rice等,2000,Trends Genet.16:276-277),优选3.0.0版或更高版本的Needle程序中所执行的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开放罚分(gap open penalty)10,缺口延伸罚分(gap extension penalty)0.5和EBLUSUM62(BLUSUM62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)作为百分比同一性,并计算如下:

[0062]  $(\text{同样的残基} \times 100) / (\text{比对长度} - \text{比对中缺口的总数})$

[0063] 就本发明而言,两个脱氧核糖核苷酸序列之间的序列同一性程度使用如EMBOSS软件包(EMBOSS:The European Molecular Biology Open Software Suite,Rice等,2000,见上文),优选3.0.0版或更高版本的Needle程序中所执行的Needleman-Wunsch算法(Needleman和Wunsch,1970,见上文)来测定。使用的可选参数为缺口开放罚分10,缺口延伸罚分0.5和EDNAFULL(NCBIUC4.4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)作为百分比同一性,并计算如下:

[0064]  $(\text{同样的脱氧核糖核苷酸} \times 100) / (\text{比对长度} - \text{比对中缺口的总数})$

[0065] 片段:术语“片段”意指从成熟多肽的氨基和/或羧基末端缺失一个或多个(几个)氨基酸的多肽;其中所述片段具有纤维二糖水解酶活性。在一个方面,片段含有至少370个氨基酸残基,至少390个氨基酸残基,和至少410个氨基酸残基。

[0066] 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二糖水解酶活性的片段。在一个方面,亚序列含有至少1110个核苷酸,至少1170个核苷酸,和至少1230个核苷酸。

[0067] 等位变体(allelic variant):术语“等位变体”在本文中占据相同染色体基因座的基因的任何两种或更多种可选形式。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。

[0068] 编码序列:术语“编码序列”意指直接指定其多肽产物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起始密码子或其他起始密码子例如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。

[0069] 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,其通过一系列的包括剪接的步骤加工然后作为成熟的已剪接的mRNA出现。

[0070] 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,或受修饰以本来不存在于(not otherwise exist)自然界中的方式含有核酸的区段,或是合成的。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。

[0071] 调控序列 (control sequence) :术语“调控序列”意指对编码本发明变体的多核苷酸表达是必需的所有成分。各个调控序列对于编码所述变体的多核苷酸可以是天然的或异源的,或各个调控序列对于彼此可以是天然的或异源的。这些调控序列包括但不限于前导序列、聚腺苷酸化序列、前肽序列、启动子、信号肽序列和转录终止子。最少的情况,调控序列包括启动子和转录和翻译的终止信号。调控序列可以和用于引入特异性限制位点的接头一起提供,所述特异性限制位点促进调控序列与编码变体的多核苷酸编码区的连接。

[0072] 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,使得调控序列指导该编码序列的表达。

[0073] 表达 :术语“表达”包括涉及变体产生的任何步骤,其包括但不限于转录、转录后修饰、翻译、翻译后修饰和分泌。

[0074] 表达载体 :术语“表达载体”意指线性的或环状的 DNA 分子,其包含编码变体的多核苷酸,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。

[0075] 宿主细胞 :“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”涵盖亲本细胞的任何后代,所述后代由于在复制中发生的突变与亲本细胞不完全相同。

[0076] 改善的特性 :术语“改善的特性”意指与变体相关并相对于亲本改善的特征。此类改善的特性包括但不限于热活性、热稳定性、pH 活性、pH 稳定性、底物 / 辅因子特异性、改善的表面特性、产物特异性、在预处理的生物质存在下增加的稳定性或溶解性、在储藏条件下改善的稳定性和化学稳定性。

[0077] 改善的热稳定性 :术语“改善的热稳定性”意指变体在升高的温度在缓冲液中或在如那些在产物储藏 / 运输过程中存在的条件或类似于在变体的产业应用过程中存在的那些条件下温育一段时间之后相对于亲本显示纤维二糖水解酶活性的保持。变体相对于亲本可能会或不会显示改变的热活性概貌。举例而言,变体在升高的温度温育之后相对于亲本可具有改善的重新折叠的能力。

[0078] 在一个方面,当使用生物质水解测定比较在 100mM NaCl-50mM 乙酸钠 pH 5.0 缓冲液中在 67°C 热攻击 20 分钟之后的残余活性时,具有纤维二糖水解酶活性的变体的热稳定性与亲本相比至少 1.05 倍,例如至少 1.1 倍,至少 1.5 倍,至少 1.8 倍,至少 2 倍,至少 5 倍,至少 10 倍,至少 15 倍,至少 20 倍,和至少 25 倍更稳定,在该测定中,将表示为 mg 蛋白每克纤维素的具有纤维二糖水解酶活性的变体或亲本酶在 50°C 的固定温度在 pH 5 以相同量添加至 PASC 进行 30 分钟的总水解时间,在此时点测量生物质水解。

[0079] 变体命名规则 :

[0080] 就本发明而言,使用公开于 SEQ ID NO:2 中的多肽以确定在其他纤维二糖水解酶中对应的氨基酸残基。将其他纤维二糖水解酶的氨基酸序列与公开于 SEQ ID NO:2 中的多肽进行比对,并基于该比对,使用如 EMBOSS 软件包 (EMBOSS:The European Molecular Biology Open Software Suite,Rice 等,2000,TrendsGenet. 16:276-277),优选 3.0.0 版或更高版本的 Needle 程序中所执行的 Needleman-Wunsch 算法 (Needleman 和 Wunsch,1970,J. Mol. Biol. 48:443-453) 确定 SEQ ID NO:2 所公开的多肽中任何氨基酸残基

对应的氨基酸位置编号。

[0081] 在另一个纤维二糖水解酶中对应氨基酸残基的鉴定可通过使用“ClustalW” (Larkin 等, 2007, *Bioinformatics* 23:2947-2948) 进行的多个多肽序列的比对来确证。

[0082] 当其他酶与 SEQ ID NO:2 的多肽歧异到传统的基于序列的比较无法检测出其关系时 (Lindhahl 和 Elofsson, 2000, *J. Mol. Biol.* 295:613-615), 可使用其他配对序列比较算法。在基于序列的搜索中较高的敏感度可使用采用多肽家族的概率表现 (probabilistic representation) (序型) 搜索数据库的搜索程序来获得。例如, PSI-BLAST 程序通过迭代的 (iterative) 数据库搜索过程来产生序型, 并能够检测出较远的同源物 (Atschul 等, 1997, *Nucleic Acids Res.* 25:3389-3402)。当多肽的家族或超家族在蛋白质结构数据库中具有一个或多个代表时, 甚至可达成更高的敏感度。程序如 GenTHREADER (Jones 1999, *J. Mol. Biol.* 287:797-815; McGuffin 和 Jones, 2003, *Bioinformatics* 19:874-881) 使用来自不同来源 (PSI-BLAST、二级结构预测 (secondary structure prediction)、结构比对序型 (structural alignment profile) 以及溶解势 (solvation potential)) 的信息作为预测所查询序列 (query sequence) 结构折叠 (structural fold) 的神经网络的输入。类似地, Gough 等, 2000, *J. Mol. Biol.* 313:903-919 的方法可用于将未知结构的序列与存在于 SCOP 数据库中的超家族模型进行比对。这些比对可接着用于生成所述多肽的同源性模型, 且上述模型可就其准确度使用多种为此目的开发的工具加以评价。

[0083] 对于已知结构的蛋白质, 可获取数种工具和资源以供找回 (retrieve) 和生成结构比对。例如, 已将 SCOP 超家族的蛋白质进行了结构比对, 且这些比对是可获取并可下载的。两个或更多蛋白质结构可使用多种算法, 如距离比对矩阵 (distance alignment matrix) (Holm 和 Sander, 1998, *Proteins* 33:88-96) 或组合延伸 (combinatorial extension) (Shindyalov 和 Bourne, 1998, *Protein Engineering*, 11:739-747) 进行比对, 且这些算法的实施可另外用于查询具有目标结构的结构数据库, 以发现可能的结构同源物 (例如 Holm 和 Park, 2000, *Bioinformatics* 16:566-567)。

[0084] 在描述本发明的纤维二糖水解酶变体时, 为了参照方便起见, 采用了下面所述的命名法。使用通用的 IUPAC 单字母或三字母氨基酸缩写。

[0085] 取代。对于氨基酸取代, 使用下述命名法: 初始氨基酸, 位置, 取代氨基酸。相应地, 在 226 位用丙氨酸取代苏氨酸命名为“Thr226Ala”或“T226A”。多重突变可以用加号 (“+”) 分开, 例如, “Gly205Arg+Ser411Phe”或“G205R+S411F”, 分别代表 205 和 411 位用精氨酸 (R) 取代甘氨酸 (G), 以及用苯丙氨酸 (F) 取代丝氨酸 (S)。

[0086] 缺失。对于氨基酸缺失, 使用了如下命名法: 初始氨基酸, 位置\*。相应地, 在位置 195 缺失甘氨酸命名为“Gly195\*”或“G195\*”。多个缺失通过加法符 (“+”) 分开, 例如, “Gly195\*+Ser411\*”或“G195\*+S411\*”。

[0087] 插入。对于氨基酸插入, 使用了如下的命名法: 初始氨基酸, 位置, 初始氨基酸, 插入的氨基酸。因此, 在位置 195 的甘氨酸之后插入赖氨酸命名为“Gly195GlyLys”或“G195GK”。多个氨基酸的插入命名为 [ 初始氨基酸, 位置, 初始氨基酸, 插入的氨基酸 #1, 插入的氨基酸 #2 ; 等等 ]。例如, 在位置 195 的甘氨酸之后插入赖氨酸和丙氨酸记为“Gly195GlyLysAla”或“G195GKA”。

[0088] 在此情况下,通过在插入的氨基酸残基前的氨基酸残基的位置号添加小写字母而对插入的氨基酸残基进行编号。因此,在上一例子中序列为:

[0089]

亲本:	变体:

[0090]

195	195195a 195b
G	G-K-A

[0091] 多重改变。包含多重改变的变体由加法符号(“+”)分隔,例如“Arg170Tyr+Gly195Glu”或“R170Y+G195E”代表分别在位置 170 用酪氨酸取代精氨酸,和在位置 195 用谷氨酸取代甘氨酸。

[0092] 不同的改变。当可在一个位置导入不同的改变时,所述不同改变由逗号分隔,例如“Arg170Tyr,Glu”代表在位置 170 用酪氨酸或谷氨酸取代精氨酸,因此,“Tyr167Gly,Ala+Arg170Gly,Ala”指下述变体:

[0093] “Tyr167Gly+Arg170Gly”、“Tyr167Gly+Arg170Ala”、“Tyr167Ala+Arg170Gly”和“Tyr167Ala+Arg170Ala”。

[0094] 亲本纤维二糖水解酶

[0095] 所述亲本纤维二糖水解酶可为 (a) 多肽,其与 SEQ ID NO:2 的成熟多肽具有至少 60% 序列同一性;(b) 多肽,其由多核苷酸编码,所述多核苷酸在低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的 cDNA 序列,或 (iii) (i) 或 (ii) 的全长互补链;或 (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具有至少 60% 序列同一性。

[0096] 在第一个方面,所述亲本与 SEQ ID NO:2 的成熟多肽具有至少 60%,例如至少 65%,至少 70%,至少 75%,至少 80%,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,并具有纤维二糖水解酶活性。在一个方面,亲本的氨基酸序列与 SEQ ID NO:2 的成熟多肽相差不超过十个氨基酸,例如相差不超过五个氨基酸,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。

[0097] 在一个方面,所述亲本包含 SEQ ID NO:2 的氨基酸序列;或由 SEQ ID NO:2 的氨基酸序列组成。在另一个方面,所述亲本包含 SEQ ID NO:2 的成熟多肽,或由 SEQ ID NO:2 的成熟多肽组成。在另一个方面,所述亲本包含 SEQ ID NO:2 的氨基酸 20 至 454,或由 SEQ ID NO:2 的氨基酸 20 至 454 组成。

[0098] 在一个实施方案中,所述亲本为 SEQ ID NO:2 的成熟多肽含有至少 370 个氨基酸残基,例如至少 390 和至少 410 个氨基酸残基的片段。

[0099] 在另一个实施方案中,所述亲本是 SEQ ID NO:2 的成熟多肽的等位变体。

[0100] 在第二个方面,所述亲本由多核苷酸编码,所述多核苷酸在非常低严格条件、低严格条件、中等严格条件、中-高严格条件、高严格条件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)。

[0101] 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或种的菌株鉴定和克隆编码亲本的 DNA。具体而言, 根据标准的 Southern 印迹方法, 可将这些探针用于与感兴趣的属或种的基因组或 cDNA 杂交, 以从其中鉴定和分离相应的基因。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。

[0102] 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亲本。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 或其亚序列杂交的克隆或 DNA, 将所述载体材料用在 Southern 印迹中。

[0103] 就本发明而言, 杂交表示多核苷酸在低至非常高的严格条件下与标记的核苷酸探针杂交, 所述核苷酸探针对应于 SEQ ID NO:1 所示的多核苷酸, 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) 或其他任何本领域中已知的检测手段检测与探针杂交的分子。

[0104] 在一个方面, 核酸探针是 SEQ ID NO:1 的成熟多肽编码序列。在另一个方面, 核酸探针是 SEQ ID NO:1 的核苷酸 58 至 1710。在另一个方面, 所述核酸探针是编码 SEQ ID NO:2 的多肽, 或它们的片段的多核苷酸。在另一个方面, 所述核酸探针是 SEQ ID NO:1。

[0105] 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  $42^\circ\text{C}$ , 在  $5\times\text{SSPE}$ 、 $0.3\%\text{SDS}$ 、 $200\text{ 微克}/\text{ml}$  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  $2\times\text{SSC}$ 、 $0.2\%\text{SDS}$  在  $45^\circ\text{C}$  (非常低严格性),  $50^\circ\text{C}$  (低严格性),  $55^\circ\text{C}$  (中严格性),  $60^\circ\text{C}$  (中 - 高严格性),  $65^\circ\text{C}$  (高严格性), 或  $70^\circ\text{C}$  (非常高严格性) 将载体材料最终洗涤三次, 每次 15 分钟。

[0106] 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使用根据 Bolton 和 McCarthy 算法 (1962, Proc. Natl. Acad. Sci. USA 48:1390) 计算的  $T_m$  低大约  $5^\circ\text{C}$  至大约  $10^\circ\text{C}$ , 在  $0.9\text{M NaCl}$ 、 $0.09\text{M Tris-HCl}$  pH 7.6,  $6\text{mM EDTA}$ 、 $0.5\%\text{NP-40}$ 、 $1\times\text{Denhardt}$  溶液,  $1\text{mM}$  焦磷酸钠 (sodium pyrophosphate),  $1\text{mM}$  磷酸二氢钠 (sodium monobasic phosphate),  $0.1\text{mM ATP}$  和  $0.2\text{mg}$  每  $\text{ml}$  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  $6\times\text{SSC}$  加  $0.1\%\text{SDS}$  中最终洗涤一次 15 分钟, 并用  $6\times\text{SSC}$  洗涤两次, 每次 15 分钟, 在比计算的  $T_m$  低  $5^\circ\text{C}$  至  $10^\circ\text{C}$  的温度下。



[0107] 在第三个方面,所述亲本由多核苷酸编码,所述多核苷酸与SEQ ID NO:1的成熟多肽编码序列具有至少 60%,例如至少 65%,至少 70%,至少 75%,至少 80%,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,并编码具有纤维二糖水解酶活性的多肽。在一个方面,所述成熟多肽编码序列是SEQ ID NO:1的核苷酸 58 至 1710。在一个实施方案中,所述亲本由多核苷酸编码,所述多核苷酸包含SEQ ID NO:1,或由SEQ ID NO:1组成。

[0108] 所述亲本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”,意思应由多核苷酸编码的亲本由所述来源产生,或由其中插入了来自所述来源的多核苷酸的细胞产生。在一个方面,所述亲本是胞外分泌的。

[0109] 所述亲本可以是细菌纤维二糖水解酶。例如,所述亲本可为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*) 纤维二糖水解酶;或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*) 纤维二糖水解酶。

[0110] 在一个方面,所述亲本是嗜碱芽孢杆菌 (*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*) 纤维二糖水解酶。

[0111] 在另一个方面,所述亲本是似马链球菌 (*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*) 纤维二糖水解酶。

[0112] 在另一个方面,所述亲本是不产色链霉菌 (*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*) 纤维二糖水解酶。

[0113] 所述亲本可为真菌纤维二糖水解酶。例如,所述亲本可为酵母纤维二糖水解酶如假丝酵母属 (*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*) 纤维二糖水解酶。例如,所述亲本可为丝状真菌纤维二糖水解酶如枝顶孢霉属 (*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

(Coprinosia)、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*)、青霉属 (*Penicillium*)、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 (*Tolyocladium*)、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*) 纤维二糖水解酶。

[0114] 在另一个方面,所述亲本是卡尔酵母 (*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*) 纤维二糖水解酶。

[0115] 在另一个方面,所述亲本是解纤维枝顶孢霉 (*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)、Thielaviaperuviana、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) 纤维二糖水解酶。

[0116] 在另一个方面,所述亲本是曲霉属纤维二糖水解酶,如烟曲霉纤维二糖水解酶,例如 SEQ ID NO:2 的纤维二糖水解酶或其成熟多肽。

[0117] 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。本领域技术人员将容易地识别适合的等同物的身份。

[0118] 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)。

[0119] 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)分离的或从直接从自然材料(例如,土壤、堆肥、水等)获得的 DNA 样品分离的微生物鉴定和获得所述亲本。用于直接从天然生境 (habitat) 分离微生物和 DNA 的技术是本领域内公知的。随后可通过相似地筛选另一种微生物的基因组或 cDNA 文库或混合的 DNA 样品来得到编码亲本的多核苷酸。一旦用所述探针检测到编码亲本的多核苷酸,就能够使用本领域普通技术人员已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)。

[0120] 所述亲本可为杂合多肽,其中一个多肽的一部分融合于另一个多肽的一部分的 N 端或 C 端。

[0121] 所述亲本亦可为融合多肽或可切割的融合多肽,其中一个多肽融合于另一个多肽的 N 端或 C 端。通过将编码一个多肽的多核苷酸融合于编码另一个多肽的多核苷酸来产生融合的多肽。产生融合多肽的技术是本领域已知的,并包括连接编码多肽的编码序列以使它们处于同一阅读框 (in frame),并且使融合多肽的表达在相同启动子和终止子的控制下。融合蛋白亦可使用内蛋白 (intein) 技术构建,其中融合物在翻译后产生 (Cooper 等,1993,EMBO J. 12:2575-2583;Dawson 等,1994,Science 266:776-779)。

[0122] 融合多肽还可以在两个多肽之间包含切割位点。一旦分泌了融合多肽,就切割所述位点,释放所述两个多肽。切割位点的实例包括,但不限于,公开于 Martin 等,2003, J. Ind. Microbiol. Biotechnol. 3:568-76;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, Biochem. 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 中的位点。

[0123] 变体的制备

[0124] 本发明还涉及用于获得具有纤维二糖水解酶的变体的方法,所述方法包括:(a) 在对应于 SEQ ID NO:2 的成熟多肽的位置 254、285、286、330、342 和 360 的一个或多个(几个)位置将取代导入亲本纤维二糖水解酶,其中所述变体具有纤维二糖水解酶活性;和(b) 回收所述变体。在一个方面,所述方法进一步包括在对应于 SEQ ID NO:2 的位置 254、285、286、330、342 和 360 的一个或多个(几个)位置将取代导入亲本纤维二糖水解酶。

[0125] 所述变体可使用本领域任何已知的任何诱变方法,如定位诱变、合成基因构建(synthetic gene construction)、半合成基因构建(semi synthetic geneconstruction)、随机诱变、改组(shuffling)等来制备。

[0126] 可使用已知的诱变、重组和/或改组方法,然后进行相关的筛选过程,如由 Reidhaar-Olson 和 Sauer,1988,Science 241:53-57;Bowie 和 Sauer,1989,Proc. Natl. Acad. Sci. USA 86:2152-2156;WO 95/17413;或者 WO 95/22625 所公开的那些,进行一个或多个氨基酸取代、缺失和/或插入并加以测试。其他可使用的方法包括易错 PCR、噬菌体展示(例如 Lowman 等,1991,Biochemistry 30:10832-10837;美国专利号 5,223,409;WO 92/06204)和区域定向诱变(region-directed mutagenesis)(Derbyshire 等,1986, Gene 46:145;等,1988, DNA 7:127)。

[0127] 定位诱变是在编码亲本的多核苷酸上的一个或多个确定位点创建一个或多个(几个)突变的技术。

[0128] 定位诱变可在体外通过 PCR 达成,涉及使用含有所期望突变的寡核苷酸引物。定位诱变亦可在体外通过表达盒诱变进行,涉及用限制酶在包含编码亲本的多核苷酸的质粒中的位点进行切割,然后将含有突变的寡核苷酸连接到所述多核苷酸中。通常在质粒上和寡核苷酸上进行消化的限制酶是相同的,使所述质粒和插入物的粘性末端能够彼此连接。参见,例如, Scherer 和 Davis, 1979, Proc. Natl. Acad. Sci. USA 76:4949-4955;以及 Barton 等,1990, Nucleic Acids Res. 18:7349-4966。

[0129] 定位诱变亦可在体内通过本领域已知方法达成。参见,例如,美国专利申请公开号 2004/0171154;Storici 等,2001, Nature Biotechnol. 19:773-776;Kren 等,1998, Nat. Med. 4:285-290;以及 Calissano 和 Macino,1996, FungalGenet. Newslett. 43:15-16。

[0130] 任何定位诱变方法可用于本发明。有许多可用的市售试剂盒可用于制备变体。

[0131] 合成基因构建涉及在体外合成设计为编码目标多肽的多核苷酸分子。基因合成可使用多种技术进行,如由 Tian 等(2004, Nature 432:1050-1054)所述基于多通道微芯片(multiplex microchip-based)的技术,以及类似的技术,其中合成寡核苷酸并在光可编程的(photo-programmable)微流芯片(microfluidic chip)上组装。

[0132] 诱变/改组方法可与高通量、自动筛选方法组合以检测由宿主细胞表达的经克隆、诱变的多肽的活性(Ness 等,1999, Nature Biotechnology 17:893-896)。编码活性多肽的经诱变的 DNA 分子可自宿主细胞回收并使用本领域标准方法迅速测序。这些方法允许快速确定多肽中单个氨基酸残基的重要性。

[0133] 半合成基因构建通过组合合成基因构建、和/或定位诱变、和/或随机诱变、和/或改组的方面而达成。半合成构建通常为使用合成的多核苷酸片段与 PCR 技术组合的方法。基因的确切区域可因而从头开始合成,而其他区域可使用位点特异性诱变引物扩增,还可对另外的其他区域进行易错 PCR 或非易错 PCR(non-error prone PCR)扩增。多核苷酸

亚序列可随后进行改组。

[0134] 变体

[0135] 本发明还提供了亲本纤维二糖水解酶的变体,其在对应于 SEQ ID NO:2 的位置 254、285、286、330、342 和 360 的一个或多个(几个)位置包含取代,其中所述变体具有纤维二糖水解酶活性。

[0136] 在一个方面,所述变体与亲本纤维二糖水解酶的氨基酸序列具有至少 60%,例如,至少 65%,至少 70%,至少 75%,至少 80%,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%,或至少 99%,但少于 100% 的序列同一性。

[0137] 在另一个方面,所述变体与 SEQ ID NO:2 的成熟多肽具有至少 60%,例如,至少 65%,至少 70%,至少 75%,至少 80%,至少 85%,至少 90%,至少 91%,至少 92%,至少 93%,至少 94%,至少 95%,如至少 96%,至少 97%,至少 98%,和至少 99%,但少于 100% 的序列同一性。

[0138] 在一个方面,本发明的变体包含 1-5 个取代,如 1、2、3、4 或 5 个取代。在另一个方面,本发明的变体包含 1-3 个取代,如 1、2 或 3 个取代。

[0139] 在一个方面,变体在对应于 SEQ ID NO:2 的位置 254、285、286、330、342 和 360 的一个或多个(几个)位置包含取代。在另一个方面,所述变体包含或组成为一个或多个(几个)选自对应于 SEQ ID NO:2 的 C254L、L285I、G286Q、D330N、M342F 和 S360G 的取代。

[0140] 在一个方面,变体在对应于 SEQ ID NO:2 的位置 254、285、286、330、342 和 360 的任何一个位置包含取代。在另一个方面,所述变体包含或组成为一个选自对应于 SEQ ID NO:2 的 C254L、L285I、G286Q、D330N、M342F 和 S360G 的取代。

[0141] 在一个方面,所述变体包含或组成为在对应于 SEQ ID NO:2 的位置 254 的位置的取代。在另一个方面,将在对应于位置 254 的位置的氨基酸取代为 Ala、Arg、Asn、Asp、Cys、Gln、Glu、Gly、His、Ile、Leu、Lys、Met、Phe、Pro、Ser、Thr、Trp、Tyr 或 Val,优选 Leu。在另一个方面,所述变体包含或组成为对应于 SEQ ID NO:2 的取代 C254L。

[0142] 在一个方面,所述变体包含或组成为在对应于 SEQ ID NO:2 的位置 285 的位置的取代。在另一个方面,将在对应于位置 285 的位置的氨基酸取代为 Ala、Arg、Asn、Asp、Cys、Gln、Glu、Gly、His、Ile、Leu、Lys、Met、Phe、Pro、Ser、Thr、Trp、Tyr 或 Val,优选 Ile。在另一个方面,所述变体包含或组成为对应于 SEQ ID NO:2 的取代 L285I。

[0143] 在一个方面,所述变体包含或组成为在对应于 SEQ ID NO:2 的位置 286 的位置的取代。在另一个方面,将在对应于位置 286 的位置的氨基酸取代为 Ala、Arg、Asn、Asp、Cys、Gln、Glu、Gly、His、Ile、Leu、Lys、Met、Phe、Pro、Ser、Thr、Trp、Tyr 或 Val,优选 Gln。在另一个方面,所述变体包含或组成为对应于 SEQ ID NO:2 的取代 G286Q。

[0144] 在一个方面,所述变体包含或组成为在对应于 SEQ ID NO:2 的位置 286 的位置的取代,以及在对应于 SEQ ID NO:2 的位置 286 的位置处的取代。在另一个方面,将在对应于位置 285 和 286 的位置的氨基酸取代为 Ala、Arg、Asn、Asp、Cys、Gln、Glu、Gly、His、Ile、Leu、Lys、Met、Phe、Pro、Ser、Thr、Trp、Tyr 或 Val,对于位置 285 和 286 分别优选为 Ile 和 Gln。在另一个方面,所述变体包含或组成为对应于 SEQ ID NO:2 的取代 L285I 以及取代 G286Q。

[0145] 在另一个方面,所述变体包含或组成为在对应于 SEQ ID NO:2 的位置 330 的位置的取代。在另一个方面,将在对应于位置 330 的位置的氨基酸取代为 Ala、Arg、Asn、Asp、

Cys、Gln、Glu、Gly、His、Ile、Leu、Lys、Met、Phe、Pro、Ser、Thr、Trp、Tyr 或 Val, 优选 Asn。在另一个方面, 所述变体包含或组成为对应于 SEQ ID NO:2 的取代 D330N。

[0146] 在另一个方面, 所述变体包含或组成为在对应于 SEQ ID NO:2 的位置 342 的位置的取代。在另一个方面, 将在对应于位置 342 的位置的氨基酸取代之为 Ala、Arg、Asn、Asp、Cys、Gln、Glu、Gly、His、Ile、Leu、Lys、Met、Phe、Pro、Ser、Thr、Trp、Tyr 或 Val, 优选 Phe。在另一个方面, 所述变体包含或组成为对应于 SEQ ID NO:2 的取代 M342F。

[0147] 在另一个方面, 所述变体包含或组成为在对应于 SEQ ID NO:2 的位置 360 的位置的取代。在另一个方面, 将在对应于位置 360 的位置的氨基酸取代之为 Ala、Arg、Asn、Asp、Cys、Gln、Glu、Gly、His、Ile、Leu、Lys、Met、Phe、Pro、Ser、Thr、Trp、Tyr 或 Val, 优选 Gly。在另一个方面, 所述变体包含或组成为对应于 SEQ ID NO:2 的取代 S360G。

[0148] 在一个方面, 所述变体包含或组成为在对应于 SEQ ID NO:2 的位置 254、285、286、330、342 和 360 的任何两个位置的取代 (例如 285 和 286), 如上述的那些。在一个方面, 所述变体包含或组成为在对应于 SEQ ID NO:2 的位置 254 和 285 的位置的取代, 如上述的那些。在另一个方面, 所述变体包含或组成为在对应于位置 254 和 286 的位置的取代, 如上述的那些。在另一个方面, 所述变体包含或组成为在对应于位置 254 和 330 的位置的取代, 如上述的那些。在另一个方面, 所述变体包含或组成为在对应于位置 254 和 342 的位置的取代, 如上述的那些。在另一个方面, 所述变体包含或组成为在对应于位置 254 和 360 的位置的取代, 如上述的那些。在另一个方面, 所述变体包含或组成为在对应于位置 285 和 286 的位置的取代, 如上述的那些。在另一个方面, 所述变体包含或组成为在对应于位置 285 和 330 的位置的取代, 如上述的那些。在另一个方面, 所述变体包含或组成为在对应于位置 285 和 342 的位置的取代, 如上述的那些。在另一个方面, 所述变体包含或组成为在对应于位置 285 和 360 的位置的取代, 如上述的那些。在另一个方面, 所述变体包含或组成为在对应于位置 286 和 330 的位置的取代, 如上述的那些。在另一个方面, 所述变体包含或组成为在对应于位置 286 和 342 的位置的取代, 如上述的那些。在另一个方面, 所述变体包含或组成为在对应于位置 286 和 360 的位置的取代, 如上述的那些。

[0149] 在一个方面, 所述变体包含或组成为选自对应于 SEQ ID NO:2 的 C254L、L285I、G286Q、D330N、M342F 和 S360G 的两个取代。在一个方面, 所述变体包含或组成为取代 C254L+L285I。在另一个方面, 所述变体包含或组成为取代 C254L+G286Q。在另一个方面, 所述变体包含或组成为取代 C254L+D330N。在另一个方面, 所述变体包含或组成为取代 C254L+M342F。在另一个方面, 所述变体包含或组成为取代 C254L+S360G。在另一个方面, 所述变体包含或组成为取代 L285I+G286Q。在另一个方面, 所述变体包含或组成为取代 L285I+D330N。在另一个方面, 所述变体包含或组成为取代 L285I+M342F。在另一个方面, 所述变体包含或组成为取代 L285I+S360G。在另一个方面, 所述变体包含或组成为取代 G286Q+D330N。在另一个方面, 所述变体包含或组成为取代 G286Q+M342F。在另一个方面, 所述变体包含或组成为取代 G286Q+S360G。

[0150] 在一个方面, 所述变体包含或组成为在对应于 SEQ ID NO:2 的位置 254、285、286、330、342 和 360 的任何三个位置的取代, 如上述的那些。在一个方面, 所述变体包含或组成为在对应于 SEQ ID NO:2 的位置 254、285 和 286 的位置的取代, 如上述的那些。在另一个方面, 所述变体包含或组成为在对应于位置 254、285 和 330 的位置的取代, 如上述的那些。

在另一个方面,所述变体包含或组成为在对应于位置 254、285 和 342 的位置的取代,如上述的那些。在另一个方面,所述变体包含或组成为在对应于位置 254、285 和 360 的位置的取代,如上述的那些。在另一个方面,所述变体包含或组成为在对应于位置 254、286 和 330 的位置的取代,如上述的那些。在另一个方面,所述变体包含或组成为在对应于位置 254、286 和 342 的位置的取代,如上述的那些。在另一个方面,所述变体包含或组成为在对应于位置 254、286 和 360 的位置的取代,如上述的那些。在另一个方面,所述变体包含或组成为在对应于位置 254、330 和 342 的位置的取代,如上述的那些。在另一个方面,所述变体包含或组成为在对应于位置 254、330 和 360 的位置的取代,如上述的那些。在另一个方面,所述变体包含或组成为在对应于位置 254、342 和 360 的位置的取代,如上述的那些。在另一个方面,所述变体包含或组成为在对应于位置 285、286 和 330 的位置的取代,如上述的那些。在另一个方面,所述变体包含或组成为在对应于位置 285、286 和 342 的位置的取代,如上述的那些。在另一个方面,所述变体包含或组成为在对应于位置 285、286 和 360 的位置的取代,如上述的那些。在另一个方面,所述变体包含或组成为在对应于位置 285、330 和 342 的位置的取代,如上述的那些。在另一个方面,所述变体包含或组成为在对应于位置 285、330 和 360 的位置的取代,如上述的那些。在另一个方面,所述变体包含或组成为在对应于位置 285、342 和 360 的位置的取代,如上述的那些。在另一个方面,所述变体包含或组成为在对应于位置 286、330 和 342 的位置的取代,如上述的那些。在另一个方面,所述变体包含或组成为在对应于位置 286、330 和 360 的位置的取代,如上述的那些。在另一个方面,所述变体包含或组成为在对应于位置 286、342 和 360 的位置的取代,如上述的那些。在另一个方面,所述变体包含或组成为在对应于位置 330、342 和 360 的位置的取代,如上述的那些。

[0151] 在另一个方面,所述变体包含或组成为选自对应于 SEQ ID NO:2 的 C254L、L285I、G286Q、D330N、M342F 和 S360G 的三个取代。在一个方面,所述变体包含或组成为取代 C254L+L285I+G286Q。在另一个方面,所述变体包含或组成为取代 C254L+L285I+D330N。在另一个方面,所述变体包含或组成为取代 C254L+L285I+M342F。在另一个方面,所述变体包含或组成为取代 C254L+L285I+S360G。在另一个方面,所述变体包含或组成为取代 C254L+G286Q+D330N。在另一个方面,所述变体包含或组成为取代 C254L+G286Q+M342F。在另一个方面,所述变体包含或组成为取代 C254L+G286Q+S360G。在另一个方面,所述变体包含或组成为取代 C254L+D330N+M342F。在另一个方面,所述变体包含或组成为取代 C254L+D330N+S360G。在另一个方面,所述变体包含或组成为取代 C254L+M342F+S360G。在另一个方面,所述变体包含或组成为取代 L285I+G286Q+D330N。在另一个方面,所述变体包含或组成为取代 L285I+G286Q+M342F。在另一个方面,所述变体包含或组成为取代 L285I+G286Q+S360G。在另一个方面,所述变体包含或组成为取代 L285I+D330N+M342F。在另一个方面,所述变体包含或组成为取代 L285I+D330N+S360G。在另一个方面,所述变体包含或组成为取代 L285I+M342F+S360G。在另一个方面,所述变体包含或组成为取代 G286Q+D330N+M342F。在另一个方面,所述变体包含或组成为取代 G286Q+D330N+S360G。在另一个方面,所述变体包含或组成为取代 G286Q+M342F+S360G。在另一个方面,所述变体包含或组成为取代 D330N+M342F+S360G。

[0152] 在一个方面,所述变体包含或组成为在对应于 SEQ ID NO:2 的位置 254、285、286、330、342 和 360 的任何四个位置的取代,如上述的那些。在一个方面,所述变体包含或组成

为在对应于 SEQ ID NO:2 的位置 254、285、286 和 330 的位置的取代,如上述的那些。在另一个方面,所述变体包含或组成为在对应于位置 254、285、286 和 342 的位置的取代,如上述的那些。在另一个方面,所述变体包含或组成为在对应于位置 254、285、330 和 342 的位置的取代,如上述的那些。在另一个方面,所述变体包含或组成为在对应于位置 254、286、330 和 342 的位置的取代,如上述的那些。在另一个方面,所述变体包含或组成为在对应于位置 285、286、330 和 342 的位置的取代,如上述的那些。在另一个方面,所述变体包含或组成为在对应于位置 254、285、286 和 360 的位置的取代,如上述的那些。在另一个方面,所述变体包含或组成为在对应于位置 254、285、330 和 360 的位置的取代,如上述的那些。在另一个方面,所述变体包含或组成为在对应于位置 254、286、330 和 360 的位置的取代,如上述的那些。在另一个方面,所述变体包含或组成为在对应于位置 285、286、330 和 360 的位置的取代,如上述的那些。在另一个方面,所述变体包含或组成为在对应于位置 254、285、342 和 360 的位置的取代,如上述的那些。在另一个方面,所述变体包含或组成为在对应于位置 254、286、342 和 360 的位置的取代,如上述的那些。在另一个方面,所述变体包含或组成为在对应于位置 285、286、342 和 360 的位置的取代,如上述的那些。在另一个方面,所述变体包含或组成为在对应于位置 254、330、342 和 360 的位置的取代,如上述的那些。在另一个方面,所述变体包含或组成为在对应于位置 285、330、342 和 360 的位置的取代,如上述的那些。在另一个方面,所述变体包含或组成为在对应于位置 286、330、342 和 360 的位置的取代,如上述的那些。

[0153] 在另一个方面,所述变体包含或组成为选自对应于 SEQ ID NO:2 的 C254L、L285I、G286Q、D330N、M342F 和 S360G 的四个取代。在一个方面,所述变体包含或组成为取代 C254L+L285I+G286Q+D330N。在另一个方面,所述变体包含或组成为取代 C254L+L285I+G286Q+M342F。在另一个方面,所述变体包含或组成为取代 C254L+L285I+D330N+M342F。在另一个方面,所述变体包含或组成为取代 C254L+G286Q+D330N+M342F。在另一个方面,所述变体包含或组成为取代 L285I+G286Q+D330N+M342F。在另一个方面,所述变体包含或组成为取代 C254L+L285I+G286Q+S360G。在另一个方面,所述变体包含或组成为取代 C254L+L285I+D330N+S360G。在另一个方面,所述变体包含或组成为取代 C254L+G286Q+D330N+S360G。在另一个方面,所述变体包含或组成为取代 L285I+G286Q+D330N+S360G。在另一个方面,所述变体包含或组成为取代 C254L+L285I+M342F+S360G。在另一个方面,所述变体包含或组成为取代 C254L+G286Q+M342F+S360G。在另一个方面,所述变体包含或组成为取代 L285I+G286Q+M342F+S360G。在另一个方面,所述变体包含或组成为取代 C254L+D330N+M342F+S360G。在另一个方面,所述变体包含或组成为取代 L285I+D330N+M342F+S360G。在另一个方面,所述变体包含或组成为取代 G286Q+D330N+M342F+S360G。

[0154] 在一个方面,所述变体包含或组成为在对应于 SEQ ID NO:2 的位置 254、285、286、330、342 和 360 的任何五个位置的取代,如上述的那些。在一个方面,所述变体包含或组成为在对应于 SEQ ID NO:2 的位置 285、286、330、342 和 360 的位置的取代,如上述的那些。在另一个方面,所述变体包含或组成为在对应于位置 254、286、330、342 和 360 的位置的取



代,如上述的那些。在另一个方面,所述变体包含或组成为在对应于位置 254、285、330、342 和 360 的位置的取代,如上述的那些。在另一个方面,所述变体包含或组成为在对应于位置 254、285、286、342 和 360 的位置的取代,如上述的那些。在另一个方面,所述变体包含或组成为在对应于位置 254、285、286、330 和 360 的位置的取代,如上述的那些。在另一个方面,所述变体包含或组成为在对应于位置 254、285、286、330 和 342 的位置的取代,如上述的那些。

[0155] 在另一个方面,所述变体包含或组成为选自对应于 SEQ ID NO:2 的 C254L、L285I、G286Q、D330N、M342F 和 S360G 的五个取代。在一个方面,所述变体包含或组成为取代 L285I+G286Q+D330N+M342F+S360G。在另一个方面,所述变体包含或组成为取代 C254L+G286Q+D330N+M342F+S360G。在另一个方面,所述变体包含或组成为取代 C254L+L285I+D330N+M342F+S360G。在另一个方面,所述变体包含或组成为取代 C254L+L285I+G286Q+M342F+S360G。在另一个方面,所述变体包含或组成为取代 C254L+L285I+G286Q+D330N+S360G。在另一个方面,所述变体包含或组成为取代 C254L+L285I+G286Q+D330N+M342F。

[0156] 在另一个方面,所述变体包含或组成为对应于 SEQ ID NO:2 的位置 254、285、286、330、342 和 360 的取代,如上述的那些。在另一个方面,所述变体包含或组成为对应于 SEQ ID NO:2 的取代 C254L+L285I+G286Q+D330N+M342F+S360G。

[0157] 所述变体可进一步包含在一个或多个(几个)其他位置的改变。在一个方面,本发明的变体包含 1-5 个进一步的改变,如 1、2、3、4 或 5 个改变。在另一个方面,本发明的变体进一步包含 1-5 个取代,如 1、2、3、4 或 5 个取代。

[0158] 在一个方面,变体包含或组成为在对应于 SEQ ID NO:2 的位置 245、382、420、437 和 440 的一个或多个(几个)位置的取代。在另一个方面,所述变体包含或组成为一个或多个(几个)选自对应于 SEQ ID NO:2 的 A245S、G382C、L420I、T437Q 和 Q440C 的取代。

[0159] 在一个方面,变体包含或组成为在对应于 SEQ ID NO:2 的位置 245、382、420、437 和 440 的任何一个位置的取代。在另一个方面,所述变体包含或组成为选自对应于 SEQ ID NO:2 的 A245S、G382C、L420I、T437Q 和 Q440C 的一个取代。

[0160] 在一个方面,所述变体进一步包含在对应于 SEQ ID NO:2 的位置 245 的位置的取代。在另一个方面,将在对应于位置 245 的位置的氨基酸取代为 Ala、Arg、Asn、Asp、Cys、Gln、Glu、Gly、His、Ile、Leu、Lys、Met、Phe、Pro、Ser、Thr、Trp、Tyr 或 Val, 优选 Ser。在另一个方面,所述变体进一步包含或组成为对应于 SEQ ID NO:2 的取代 A245S。

[0161] 在一个方面,所述变体进一步包含在对应于 SEQ ID NO:2 的位置 382 的位置的取代。在另一个方面,将在对应于位置 382 的位置的氨基酸取代为 Ala、Arg、Asn、Asp、Cys、Gln、Glu、Gly、His、Ile、Leu、Lys、Met、Phe、Pro、Ser、Thr、Trp、Tyr 或 Val, 优选 Cys。在另一个方面,所述变体进一步包含或组成为对应于 SEQ ID NO:2 的取代 G382C。

[0162] 在一个方面,所述变体进一步包含在对应于 SEQ ID NO:2 的位置 420 的位置的取代。在另一个方面,将在对应于位置 420 的位置的氨基酸取代为 Ala、Arg、Asn、Asp、Cys、Gln、Glu、Gly、His、Ile、Leu、Lys、Met、Phe、Pro、Ser、Thr、Trp、Tyr 或 Val, 优选 Ile。在另一个方面,所述变体进一步包含或组成为对应于 SEQ ID NO:2 的取代 L420I。

[0163] 在一个方面,所述变体进一步包含在对应于 SEQ ID NO:2 的位置 437 的位置的取

代。在另一个方面,将在对应于位置 437 的位置的氨基酸取代为 Ala、Arg、Asn、Asp、Cys、Gln、Glu、Gly、His、Ile、Leu、Lys、Met、Phe、Pro、Ser、Thr、Trp、Tyr 或 Val, 优选 Gln。在另一个方面,所述变体进一步包含或组成为对应于 SEQ ID NO:2 的取代 T437Q。

[0164] 在一个方面,所述变体进一步包含在对应于 SEQ ID NO:2 的位置 440 的位置的取代。在另一个方面,将在对应于位置 440 的位置的氨基酸取代为 Ala、Arg、Asn、Asp、Cys、Gln、Glu、Gly、His、Ile、Leu、Lys、Met、Phe、Pro、Ser、Thr、Trp、Tyr 或 Val, 优选 Cys。在另一个方面,所述变体进一步包含或组成为对应于 SEQ ID NO:2 的取代 Q440C。在一个方面,所述变体进一步包含或组成为对应于 SEQ ID NO:2 的取代 G382C 和 Q440C。

[0165] 在一个方面,所述变体进一步包含或组成为在对应于 SEQ ID NO:2 的位置 245、382、420、437 和 440 的任何两个位置的取代,如上述的那些。在一个方面,所述变体进一步包含或组成为在对应于 SEQ ID NO:2 的位置 420 和 437 的位置的取代,如上述的那些。在另一个方面,所述变体进一步包含或组成为在对应于位置 245 和 420 的位置的取代,如上述的那些。在另一个方面,所述变体进一步包含或组成为在对应于位置 382 和 420 的位置的取代,如上述的那些。在另一个方面,所述变体进一步包含或组成为在对应于位置 420 和 440 的位置的取代,如上述的那些。在另一个方面,所述变体进一步包含或组成为在对应于位置 245 和 437 的位置的取代,如上述的那些。在另一个方面,所述变体进一步包含或组成为在对应于位置 382 和 437 的位置的取代,如上述的那些。在另一个方面,所述变体进一步包含或组成为在对应于位置 437 和 440 的位置的取代,如上述的那些。在另一个方面,所述变体进一步包含或组成为在对应于位置 245 和 382 的位置的取代,如上述的那些。在另一个方面,所述变体进一步包含或组成为在对应于位置 245 和 440 的位置的取代,如上述的那些。在另一个方面,所述变体进一步包含或组成为在对应于位置 382 和 440 的位置的取代,如上述的那些。

[0166] 在另一个方面,所述变体进一步包含或组成为选自对应于 SEQ ID NO:2 的 A245S、G382C、L420I、T437Q 和 Q440C 的任何两个取代。在一个方面,所述变体进一步包含或组成为对应于 SEQ ID NO:2 的取代 L420I+T437Q。在另一个方面,所述变体进一步包含或组成为取代 A245S+L420I。在另一个方面,所述变体进一步包含或组成为取代 G382C+L420I。在另一个方面,所述变体进一步包含或组成为取代 L420I+Q440C。在另一个方面,所述变体进一步包含或组成为取代 A245S+T437Q。在另一个方面,所述变体进一步包含或组成为取代 G382C+T437Q。在另一个方面,所述变体进一步包含或组成为取代 T437Q+Q440C。在另一个方面,所述变体进一步包含或组成为取代 A245S+G382C。在另一个方面,所述变体进一步包含或组成为取代 A245S+Q440C。在另一个方面,所述变体进一步包含或组成为取代 G382C+Q440C。

[0167] 在一个方面,所述变体进一步包含或组成为在对应于 SEQ ID NO:2 的位置 245、382、420、437 和 440 的任何三个位置的取代,如上述的那些。在一个方面,所述变体进一步包含或组成为在对应于 SEQ ID NO:2 的位置 245、420 和 437 的位置的取代,如上述的那些。在另一个方面,所述变体进一步包含或组成为在对应于位置 382、420 和 437 的位置的取代,如上述的那些。在另一个方面,所述变体进一步包含或组成为在对应于位置 420、437 和 440 的位置的取代,如上述的那些。在另一个方面,所述变体进一步包含或组成为在对应于位置 245、382 和 420 的位置的取代,如上述的那些。在另一个方面,所述变体进一步包含或组成

为在对应于位置 245、420 和 440 的位置的取代,如上述的那些。在另一个方面,所述变体进一步包含或组成为在对应于位置 382、420 和 440 的位置的取代,如上述的那些。在另一个方面,所述变体进一步包含或组成为在对应于位置 245、382 和 437 的位置的取代,如上述的那些。在另一个方面,所述变体进一步包含或组成为在对应于位置 245、437 和 440 的位置的取代,如上述的那些。在另一个方面,所述变体进一步包含或组成为在对应于位置 382、437 和 440 的位置的取代,如上述的那些。在另一个方面,所述变体进一步包含或组成为在对应于位置 245、382 和 440 的位置的取代,如上述的那些。

[0168] 在另一个方面,所述变体进一步包含或组成为选自对应于 SEQ ID NO:2 的 A245S、G382C、L420I、T437Q 和 Q440C 的任何三个取代。在一个方面,所述变体进一步包含或组成为对应于 SEQ ID NO:2 的取代 A245S+L420I+T437Q。在另一个方面,所述变体进一步包含或组成为取代 G382C+L420I+T437Q。在另一个方面,所述变体进一步包含或组成为取代 L420I+T437Q+Q440C。在另一个方面,所述变体进一步包含或组成为取代 A245S+G382C+L420I。在另一个方面,所述变体进一步包含或组成为取代 A245S+L420I+Q440C。在另一个方面,所述变体进一步包含或组成为取代 G382C+L420I+Q440C。在另一个方面,所述变体进一步包含或组成为取代 A245S+G382C+T437Q。在另一个方面,所述变体进一步包含或组成为取代 A245S+T437Q+Q440C。在另一个方面,所述变体进一步包含或组成为取代 G382C+T437Q+Q440C。在另一个方面,所述变体进一步包含或组成为取代 A245S+G382C+Q440C。

[0169] 在一个方面,所述变体进一步包含或组成为在对应于 SEQ ID NO:2 的位置 245、382、420、437 和 440 的任何四个位置的取代,如上述的那些。在一个方面,所述变体进一步包含或组成为在对应于 SEQ ID NO:2 的成熟多肽的位置 245、382、437 和 440 的位置的取代,如上述的那些。在另一个方面,所述变体进一步包含或组成为在对应于位置 245、382、420 和 440 的位置的取代,如上述的那些。在另一个方面,所述变体进一步包含或组成为在对应于位置 382、420、437 和 440 的位置的取代,如上述的那些。在另一个方面,所述变体进一步包含或组成为在对应于位置 245、420、437 和 440 的位置的取代,如上述的那些。在另一个方面,所述变体进一步包含或组成为在对应于位置 245、382、420 和 437 的位置的取代,如上述的那些。

[0170] 在另一个方面,所述变体进一步包含或组成为选自对应于 SEQ ID NO:2 的 A245S、G382C、L420I、T437Q 和 Q440C 的任何四个取代。在一个方面,所述变体进一步包含或组成为对应于 SEQ ID NO:2 的取代 A245S+G382C+T437Q+Q440C。在另一个方面,所述变体进一步包含或组成为取代 A245S+G382C+L420I+Q440C。在另一个方面,所述变体进一步包含或组成为取代 G382C+L420I+T437Q+Q440C。在另一个方面,所述变体进一步包含或组成为取代 A245S+L420I+T437Q+Q440C。在另一个方面,所述变体进一步包含或组成为取代 A245S+G382C+L420I+T437Q。

[0171] 在一个方面,所述变体进一步包含或组成为在对应于 SEQ ID NO:2 的位置 245、382、420、437 和 440 的位置的取代,如上述的那些。在一个方面,所述变体进一步包含或组成为对应于 SEQ ID NO:2 的取代 A245S+G382C+L420I+T437Q+Q440C。

[0172] 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) 来鉴定亲本中的必需氨基酸。在后一技术中,将单一丙氨酸突变引入到分子中的每个残基,并且测试所得突变分子的纤维二糖水解酶活性以鉴定对于所述分子的活性关键的氨基酸残基。亦参见 Hilton 等, 1996, J.

Biol. Chem. 271:4699-4708。纤维二糖水解酶的活性部位,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确定。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)也能够从与同亲本相关的多肽的同一性分析来推断。

[0173] 所述变体可由 370 至 450 个氨基酸,例如,370 至 379、380 至 389、390 至 399、400 至 409、410 至 419 或 420 至 440 个氨基酸组成。

[0174] 本发明还涵盖具有纤维二糖水解酶活性的多肽,其中其氨基酸序列的一个或多个(几个)位置与 SEQ ID NO:2 的相应位置 254、285、286、330、342 和 360 不同。本文中所述的变体方面亦涵盖涉及具有纤维二糖水解酶活性的多肽的类似方面。

[0175] 多肽

[0176] 本发明还涉及编码任何本发明的变体的分离的多核苷酸。

[0177] 核酸构建体

[0178] 本发明还涉及包含编码本发明的变体的多核苷酸的核酸构建体,所述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。

[0179] 可以用许多方式操作所述多核苷酸以提供变体的表达。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 方法修饰多核苷酸的技术是本领域熟知的。

[0180] 调控序列可为启动子序列,其由用于表达多核苷酸的宿主细胞所识别。启动子序列含有介导变体的表达的转录调控序列。启动子可以是在宿主细胞中显示转录活性的任何核酸序列,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。

[0181] 用于在细菌宿主细胞中指导本发明的核酸构建体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, Proceedings of the National Academy of Sciences USA 75:3727-3731), 以及 tac 启动子(DeBoer 等,1983, Proc. Natl. Acad. Sci. USA 80:21-25)。另外的启动子在 "Useful proteins from recombinant bacteria" 于 Gilbert 等,1980, Scientific American, 242:74-94 中;和在 Sambrook 等,1989, 见上文中描述。

[0182] 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(Rhizomucor miehei) 脂肪酶、曼赫根毛霉天冬氨酸蛋白酶、里氏木霉  $\beta$ -葡糖苷

酶、里氏木霉纤维二糖水解酶 I、里氏木霉纤维二糖水解酶 II、里氏木霉内切葡聚糖酶 I、里氏木霉内切葡聚糖酶 II、里氏木霉内切葡聚糖酶 III、里氏木霉内切葡聚糖酶 IV、里氏木霉内切葡聚糖酶 V、里氏木霉木聚糖酶 I、里氏木霉木聚糖酶 II、里氏木霉  $\beta$ -木糖苷酶,以及 NA2-tpi 启动子(一种修饰的启动子,其包括在曲霉属 (*Aspergilli*) 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,其包括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来自在构巢曲霉或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。

[0183] 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 (TP1)、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。

[0184] 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变体的多核苷酸的 3' 末端可操作地连接。可以使用在宿主细胞中有功能的任何终止子。

[0185] 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  $\alpha$ -葡糖苷酶、黑曲霉葡糖淀粉酶、米曲霉 TAKA 淀粉酶和尖镰孢胰蛋白酶样蛋白酶。

[0186] 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。

[0187] 调控序列还可以是合适的前导序列,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变体的多核苷酸的 5'- 末端。可使用在宿主细胞中有功能的任何前导序列。

[0188] 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。

[0189] 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)。

[0190] 调控序列也可以是聚腺苷酸化序列,其是与变体编码序列的 3' 末端可操作地连接的序列,并且在转录时,宿主细胞将其识别为将聚腺苷残基添加至转录的 mRNA 的信号。可使用在宿主细胞中有功能的任何聚腺苷酸化序列。

[0191] 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: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。

[0192] 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。

[0193] 调控序列还可以是信号肽编码区,其编码与变体的 N 端相连的信号肽,并且指导

所述变体进入细胞分泌途径。多核苷酸的编码序列 5' 端可固有地包含信号肽编码区, 其与编码所述变体的编码区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区。异源信号肽编码区在编码序列不天然地含有信号肽编码区时可为必需的。或者, 异源信号肽编码区可以简单地取代天然信号肽编码区以增强变体的分泌。然而, 可使用指导表达的变体进入宿主细胞的分泌途径的任何信号肽编码区。

[0194] 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。

[0195] 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。

[0196] 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。

[0197] 调控序列还可以是前肽编码区, 其编码位于变体 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区。

[0198] 当信号肽和前肽区二者均出现在变体的 N 端时, 将前肽区置于紧接着 (next to) 变体 N 端, 并且将信号肽区置于紧接着前肽区的 N 端。

[0199] 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变体的表达。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、米曲霉 TAKA  $\alpha$ -淀粉酶启动子和米曲霉葡糖淀粉酶启动子。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, 这些调节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, 编码变体的多核苷酸将与调节序列可操作地连接。

[0200] 表达载体

[0201] 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变体的多核苷酸。可供选择的是, 可以通过在适当的用于表达的载体中插入多核苷酸或包含所述多核苷酸的核酸构建体来表达所述多核苷酸。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。

[0202] 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。

[0203] 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)。

[0204] 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(proto-trophy to auxotrophs)等。

[0205] 细菌选择性标记的实例是来自地衣芽孢杆菌或枯草芽孢杆菌的 *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*(鸟氨酸氨甲基转移酶)、*bar*(草铵膦(phosphinothricin)乙酰转移酶)、*hph*(潮霉素磷酸转移酶)、*niaD*(硝酸还原酶)(nitrate reductase)、*pyrG*(乳清酸核苷-5'-磷酸脱羧酶)(orotidine-5'-phosphatedecarboxylase)、*sC*(硫酸腺苷酰转移酶)和 *trpC*(邻氨基苯甲酸合酶(anthranilate synthase))以及它们的等同物。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。

[0206] 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。

[0207] 为了整合入宿主细胞基因组,载体可依赖编码变体的多核苷酸的序列或用于通过同源或非同源重组整合入基因组的任何其它载体元件。或者,载体可以含有额外的核苷酸序列,用于指导通过同源重组整合入宿主细胞基因组染色体中的精确位置。为了增加在精确位置整合的可能性,整合元件应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同一性以增强同源重组的概率。整合元件可以是任何序列,其与宿主细胞基因组中的目标序列同源。此外,整合元件可以是非编码或编码的核苷酸序列。另一方面,可以将载体通过非同源重组整合到宿主细胞的基因组中。

[0208] 为了自主复制,载体可以进一步包含复制起点,其使载体能够在关注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核苷酸序列。

[0209] 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。

[0210] 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*、*ARS 1* 和 *CEN3* 的组合,和 *ARS4* 和 *CEN6* 的组合。

[0211] 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; WO00/24883)。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。

[0212] 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变体的产生。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。

[0213] 用于连接上述元件以构建本发明的重组表达载体以获得分离的变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)。

[0214] 宿主细胞

[0215] 本发明还涉及重组宿主细胞,其包含本发明的多核苷酸可操作地连接于一个或多个(几个)指导本发明变体的产生的调控序列。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变体的基因及其来源。

[0216] 宿主细胞可以是在变体的重组产生中有用的任何细胞,例如,原核或真核细胞。

[0217] 原核宿主细胞可以是任何革兰氏阳性或革兰氏阴性细菌。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、大肠杆菌属、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、沙门氏菌属和脲原体属。

[0218] 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。

[0219] 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 (*Streptococcus uberis*) 和马链球菌兽瘟亚种细胞。

[0220] 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。

[0221] 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. Bacteriol. 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。

[0222] 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。

[0223] 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) 以及卵菌门 (Oomycota) 和所有有丝分裂孢子真菌 (mitosporicfungi) (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)。

[0224] 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。

[0225] 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、裂殖酵母属或西洋蓍霉属细胞, 如乳酸克鲁维酵母 (*Kluyveromyces lactis*)、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、或解脂西洋蓍霉 (*Yarrowia lipolytica*) 细胞。

[0226] 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。

[0227] 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。

[0228] 例如, 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 subvermispora*)、*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、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。在一个方面, 宿主细胞为曲霉属菌种, 例如米曲霉。

[0229] 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再生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 EP 238 023 和 Yelton 等, 1984, *Proc. Natl. Acad. Sci. USA* 81:1470-1474 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。

[0230] 产生方法

[0231] 本发明还涉及用于产生变体的方法, 其包括: (a) 在适于所述变体表达的条件下培养本发明的宿主细胞; 和 (b) 回收所述多肽。

[0232] 使用本领域已知的方法在适合于产生所述变体的营养培养基中培养细胞。例如, 可以通过在合适培养基中和允许表达和 / 或分离所述多肽的条件下进行的摇瓶培养, 或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)。如果变体分泌到营养培养基中, 该变体能够从所述培养基中直接回收。如果变体不分泌, 则其能够从细胞裂解物 (lysate) 回收。

[0233] 可以使用本领域已知的对于所述变体是特异性的方法来检测变体。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 (enzyme assay) 可用于变体的活性。

[0234] 变体可以通过本领域已知的方法回收。例如, 变体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收集、离心、过滤、提取、喷雾干燥、蒸发或沉淀。

[0235] 变体可以通过多种本领域已知的方法纯化以获得分离的变体, 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)。

[0236] 在可选的方面,不回收变体,而将表达变体的本发明宿主细胞作为变体的来源。

[0237] 组合物

[0238] 本发明还涉及包含本发明的变体的组合物。优选地,所述组合物富含此种变体。术语“富含”意指所述组合物的纤维二糖水解酶活性以例如 1.1 的富集因子增加。

[0239] 所述组合物可包含变体作为主要酶组分,例如单组分组合物。或者,所述组合物可包含多种酶活性,如氨肽酶、淀粉酶、糖酶、羧肽酶、过氧化氢酶、纤维素酶、几丁质酶(chitinase)、角质酶(cutinase)、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、卤素过氧化物酶、转化酶、漆酶、脂肪酶、甘露糖苷酶、氧化酶、果胶分解酶、肽谷氨酰胺酶、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。所述其他酶可通过例如属于曲霉属,例如棘孢曲霉(*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*)或米曲霉(*Aspergillus oryzae*);镰孢属,例如杆孢状镰孢(*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*)、硫色镰孢(*Fusarium sulphureum*)、圆镰孢(*Fusarium torulosum*)、拟丝孢镰孢(*Fusarium trichothecioides*)或镶片镰孢(*Fusarium venenatum*);腐质霉属,例如特异腐质霉(*Humicola insolens*)或疏棉状腐质霉(*Humicolalanuginosa*);或木霉属,例如哈茨木霉(*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*)的微生物产生。

[0240] 所述组合物可依照本领域已知方法制备,且可为液体或干组合物的形式。举例而言,所述组合物可为颗粒或微粒的形式。所述变体可依照本领域已知方法稳定化。

[0241] 纤维素材料的加工

[0242] 本发明还涉及降解或转化纤维素材料的方法,包括:在本发明的具有纤维二糖水解酶活性的变体的存在下用酶组合物处理纤维素材料。在一个优选的方面,所述方法还包括回收经降解或转化的纤维素材料。

[0243] 本发明还涉及产生发酵产物的方法,其包括:(a)在本发明的具有纤维二糖水解酶活性的变体存在的条件下,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。

[0244] 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,其中在本发明的具有纤维二糖水解酶活性的变体存在的条件下,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,所述方法进一步包括从发酵回收发酵产物。

[0245] 本发明的方法可用于将纤维素材料糖化为可发酵的糖,并将可发酵的糖转化为许多有用物质,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。

[0246] 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执行。

[0247] 水解(糖化)和发酵,分别或同时,包括但不限于,分开的水解和发酵(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)、分开的水解和共发酵(SHCF)、混合的水解和发酵(HHCF),和直接微生物转化(DMC)。SHF使用分开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组合在一个步骤中(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进行,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纤维素材料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。

[0248] 常规设备包括补料分批搅拌反应器、分批搅拌反应器、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、升流层式(upflow blanket)、固定化和挤出机型反应器以供水解和 / 或发酵。

[0249] **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,Substratepretreatment:The key to effective enzymatic hydrolysis of lignocellulosics?Adv. Biochem. Engin./Biotechnol.108:67-93;Galbe和Zacchi,2007,Pretreatment oflignocellulosic 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 andBiorefining-Biofpr.2:26-40)。

[0250] 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粒度减小、预浸泡、润湿、洗涤或调理(conditioning)。

[0251] 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γ辐射预处理。

[0252] 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)。

[0253] **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。使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℃,更优选160-200℃,并且最优选170-190℃进行,其中最优的温度范围依赖于任何化学催化剂的加入。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加入。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半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。

[0254] 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)。

[0255] 化学预处理:术语“化学处理”指促进纤维素、半纤维素和 / 或木质素分离和 / 或释放的任何化学预处理。合适的化学预处理过程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。

[0256] 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)。

[0257] 还可以使用碱性条件下的几种预处理方法。这些碱预处理包括,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)。

[0258] 用碳酸钙、氢氧化钠或氨,在 85-150 °C 的低温进行石灰预处理,并且停留时间从 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。

[0259] 湿法氧化是热预处理,通常在 180-200 °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,并且最优选 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。

[0260] 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)。

[0261] 氨纤维爆炸 (AFEX) 涉及在温和温度如 90-100 °C 和高压如 17-20bar,用液体或气体氨将纤维素材料处理 5-10 分钟,其中干物质含量可以高达 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; Teymouri 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。

[0262] 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)。通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。

[0263] 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, 和美国已公开的申请 2002/0164730 所述。

[0264] 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。

[0265] 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。

[0266] 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%, 并且最优选 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。

[0267] 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)。

[0268] 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。

[0269] 物理预处理可以涉及高压和 / 或高温 (蒸汽爆炸)。在一个方面, 高压指范围在优选约 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, 高温指范围在约 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 SundsDefibrator AB, Sweden 的 Sunds Hydrolyzer 中进行。

[0270] 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。

[0271] 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, 或者它们的任意组合, 以促进纤维素、半纤维素和 / 或木质素的分离和 / 或释放。

[0272] 生物预处理: 术语“生物预处理”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 Fuels 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)。

[0273] **糖化**。在水解（也称作糖化）步骤中，将纤维素材料（例如经预处理的）水解以将纤维素或者还有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二糖水解酶活性的变体的存在下进行。所述组合物可进一步包含一种或多种（几种）半纤维素分解或木聚糖降解酶。组合物的酶还可以顺序加入。

[0274] 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。水解可以以补料分批或连续的工艺进行，其中将纤维素材料（底物）（例如，经预处理的）逐渐补入，例如，含酶的水解溶液中。

[0275] 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。温度优选的范围是约 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%。

[0276] 所述酶组合物优选包含具有纤维素分解活性和 / 或木聚糖降解活性的酶。在一个方面，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，所述酶组合物包含一种或多种（几种）木聚糖降解酶。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）木聚糖降解酶。

[0277] 所述一种或多种（几种）纤维素分解酶优选选自下组：内切葡聚糖酶、纤维二糖水解酶和  $\beta$ -葡糖苷酶。所述一种或多种（几种）木聚糖降解酶优选选自下组：木聚糖酶、乙酰木聚糖酯酶、阿魏酸酯酶、阿拉伯呋喃糖苷酶、木糖苷酶和葡糖醛酸糖苷酶。

[0278] 在另一个方面，所述酶组合物另外或甚至还包含具有纤维素分解增强活性的多肽（参见，例如 WO 2005/074647、WO 2005/074656 和 WO 2007/089290）。在另一个方面，所述酶组合物可进一步或甚至进一步包含一种或多种（几种）其他酶活性以改善含纤维素材料的降解。其他优选的酶为半纤维素酶（例如  $\alpha$ -D-葡糖醛酸糖苷酶、 $\alpha$ -L-阿拉伯呋喃糖苷酶、内切甘露聚糖酶、 $\beta$ -甘露糖苷酶、 $\alpha$ -半乳糖苷酶、内切- $\alpha$ -L-阿拉伯聚糖酶、 $\beta$ -半乳糖苷酶），糖酯酶（例如乙酰木聚糖酯酶、乙酰甘露聚糖酯酶、阿魏酸酯酶、香豆酸酯酶、葡糖醛酸酯酶 (glucuronoyl esterases)），果胶酶，蛋白酶，木质素水解酶 (ligninolytic enzyme)（例如漆酶、锰过氧化物酶、木质素过氧化物酶、 $H_2O_2$  产生酶、氧化还原酶），棒曲霉素 (expansin)、膨胀素 (swollenin) 或其组合。在本发明的方法中，其他酶可在发酵之前或过程中添加，例如在糖化之中或在发酵微生物繁殖过程中或之后添加。

[0279] 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，所述细胞用作宿主细胞以重组表达酶组合物的一种或多种（几种）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。

[0280] 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。

[0281] 具有纤维二糖水解酶活性的酶和变体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、纤维素底物、纤维素底物的浓度、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）。

[0282] 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。

[0283] 在另一个优选的方面，具有纤维二糖水解酶活性的变体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。

[0284] 在另一个优选的方面，具有纤维二糖水解酶活性的变体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 0.5g，并且最优选约 0.05 至约 0.2g 每 g 纤维素分解酶。

[0285] 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”在本文中还可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（如通过定位诱变或改组）获得的变体。

[0286] 具有纤维素分解酶活性或木聚糖降解活性的多肽可以是细菌多肽。例如，所述多肽可以是革兰氏阳性细菌多肽如芽孢杆菌属 (*Bacillus*)、链球菌属 (*Streptococcus*)、链霉菌属 (*Streptomyces*)、葡萄球菌属 (*Staphylococcus*)、肠球菌属 (*Enterococcus*)、乳杆菌属 (*Lactobacillus*)、乳球菌属 (*Lactococcus*)、梭菌属 (*Clostridium*)、地芽孢杆菌属 (*Geobacillus*) 或海洋芽孢杆菌属 (*Oceanobacillus*) 多肽，所述多肽具有纤维素分解酶活性或木聚糖降解活性；或革兰氏阴性细菌多肽，如大肠杆菌、假单胞菌属 (*Pseudomonas*)、沙门氏菌属 (*Salmonella*)、弯曲杆菌属 (*Campylobacter*)、螺杆菌属 (*Helicobacter*)、黄杆菌属 (*Flavobacterium*)、梭杆菌属 (*Fusobacterium*)、泥杆菌属 (*Ilyobacter*)、奈瑟氏菌属 (*Neisseria*) 或脲原体属 (*Ureaplasma*) 多肽，所述多肽具有纤维素分解酶活性或木聚糖降解活性。

[0287] 在一个优选的方面，所述多肽是具有纤维素分解酶活性或木聚糖降解活性的嗜碱

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。

[0288] 在另一个优选的方面,所述多肽是具有纤维素分解酶活性或木聚糖降解活性的似马链球菌、酿脓链球菌、乳房链球菌或马链球菌兽瘟亚种多肽。

[0289] 在另一个优选的方面,所述多肽是具有纤维素分解酶活性或木聚糖降解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。

[0290] 具有纤维素分解酶活性或木聚糖降解活性的多肽也可以是真菌多肽,并且更优选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,其具有纤维素分解活性或木聚糖降解活性;或更优选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、Chaetomidium、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(Holomastigotoides)、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、Pseudotrichonympha、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,其具有纤维素分解酶活性或木聚糖降解活性。

[0291] 在一个优选的方面,所述多肽是具有纤维素分解酶活性或木聚糖降解活性的卡尔酵母、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。

[0292] 在另一个优选的方面,所述多肽是具有纤维素分解酶活性或木聚糖降解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、Thielavia achromatica、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 Trichohaea saccata 多肽。

[0293] 还可以使用具有纤维素分解酶活性或木聚糖降解活性的多肽经化学修饰或蛋白质工程改造的突变体。

[0294] 所述酶组合物的一种或多种(几种)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)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(酶对宿主是天然的)。单组分纤维素分解蛋白还可以通过从发酵液中提纯这样的蛋白质来制备。

[0295] 适用于本发明的商业的纤维素分解蛋白制备物的实例包括,例如,CELLIC™Ctec(Novozymes A/S)、CELLUCLAST™(Novozymes A/S)、NOVOZYM™ 188(Novozymes A/S)、CELLUZYME™(Novozymes A/S)、CEREFLO™(NovozymesA/S)和 ULTRAFLO™(Novozymes A/S),ACCELERASE™(Genencor Int.)、LAMINEX™(Genencor Int.)、SPEZYME™CP(Genencor Int.),ROHAMENT™7069W (Röhm GmbH), FIBREZYME® LDI(Dyadic International, Inc.)、FIBREZYME® LBR(Dyadic International, Inc.)或 VISCOSTAR® 150L (DyadicInternational, Inc.)。所述纤维素酶以固形物的约 0.001 到约 5.0wt.%,更优选固形物的约 0.025 到约 4.0wt.%,且最优选固形物的约 0.005 到约 2.0wt.%的有效量添加。纤维素酶以固形物的约 0.001 至约 5.0wt%,更优选固形物的约 0.025 至约 4.0wt%,且最优选固形物的约 0.005 至约 2.0wt%的有效量添加。

[0296] 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)。

[0297] 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 8126CEL6B 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 NRRL 8126 CEL7E 内切葡聚糖酶;土生梭孢霉 NRRL 8126 CEL7F 内切葡聚糖酶;Cladorrhinumfoecundissimum 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)。

[0298] 可用于本发明的方法的纤维二糖水解酶的实例包括但不限于,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。

[0299] 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 brasilianum*)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。

[0300] 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。

[0301] 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 获得的米曲霉  $\beta$ -葡糖苷酶变体 BG 融合蛋白或米曲霉  $\beta$ -葡糖苷酶融合蛋白。

[0302] 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 sequences 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。

[0303] 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。

[0304] 在本发明的方法中,可使用任何具有纤维素分解增强活性的多肽。

[0305] 在一个方面,所述具有纤维素分解增强活性的多肽包含下述基序:

[0306] [ILMV]-P-X(4,5)-G-X-Y-[ILMV]-X-R-X-[EQ]-X(4)-[HNQ] 和 [FW]-[TF]-K-[AIV],

[0307] 其中 X 为任意氨基酸,X(4,5) 为在 4 或 5 个连续位置的任意氨基酸,而 X(4) 是在 4 个连续位置的任意氨基酸。

[0308] 具有上述所示的基序的多肽可进一步包含:

[0309] H-X(1,2)-G-P-X(3)-[YW]-[AILMV],

[0310] [EQ]-X-Y-X(2)-C-X-[EHQN]-[FILV]-X-[ILV] 或

[0311] H-X(1,2)-G-P-X(3)-[YW]-[AILMV] 和 [EQ]-X-Y-X(2)-C-X-[EHQN]-[FILV]-X-[ILV],

[0312] 其中 X 为任意氨基酸,X(1,2) 为在 1 个位置或 2 个连续位置的任意氨基酸,X(3) 为 3 个连续位置的任意氨基酸,而 X(2) 为 2 个连续位置的任意氨基酸。在上述基序中,采用公认的 IUPAC 单字母氨基酸缩写。

[0313] 在一个优选的方面,所述具有纤维素分解增强活性的多肽进一步包含 H-X(1, 2)-G-P-X(3)-[YW]-[AILMV]。在另一个优选的方面,具有纤维素分解增强活性的分离的多肽进一步包含 [EQ]-X-Y-X(2)-C-X-[EHQN]-[FILV]-X-[ILV]。在另一个优选的方面,具有纤维素分解增强活性的多肽进一步包含 H-X(1, 2)-G-P-X(3)-[YW]-[AILMV] 和 [EQ]-X-Y-X(2)-C-X-[EHQN]-[FILV]-X-[ILV]。

[0314] 在第二个方面,所述具有纤维素分解增强活性的多肽包含下述基序:

[0315] [ILMV]-P-x(4, 5)-G-x-Y-[ILMV]-x-R-x-[EQ]-x(3)-A-[HNQ],

[0316] 其中 x 为任意氨基酸, x(4, 5) 为在 4 或 5 个连续位置的任意氨基酸,而 x(3) 为 3 个连续位置的任意氨基酸。在上述基序中,采用公认的 IUPAC 单字母氨基酸缩写。

[0317] 可用于本发明的方法的具有纤维素分解增强活性的多肽的实例但不限于来自土生梭孢霉的具有纤维素分解增强活性的多肽(WO 2005/074647);来自桔橙热子囊菌的具有纤维素分解增强活性的多肽(WO 2005/074656);和来自里氏木霉的具有纤维素分解增强活性的多肽(WO 2007/089290)。

[0318] 适用于本发明的商业性木聚糖降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®木聚糖酶(Genencor)、ECOPULP® TX-200A(AB Enzymes)、HSP 6000 木聚糖酶(DSM)、DEPOL™333P(Biocatalysts Limit, Wales, UK)、DEPOL™740L(Biocatalysts Limit, Wales, UK) 和 DEPOL™ 762P(Biocatalysts Limit, Wales, UK)。

[0319] 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)。

[0320] 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 Q7SOW4)。

[0321] 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)。

[0322] 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 A 1D9T4)。

[0323] 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)。

[0324] 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)。

[0325] 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的营养培养基上,使用本领域已知方法(参见,例如 Bennett, J. W. 和 LaSure, L. (编), *MoreGene Manipulations in Fungi*, Academic Press, CA, 1991) 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)。

[0326] 所述发酵可以是任何其结果为酶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。

[0327] **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,发酵乳产品)、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。

[0328] 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。

[0329] 在实施例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。通常根据期望的发酵产品(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。

[0330] 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。

[0331] “发酵微生物”指适用于期望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)成期望的发酵产品。

[0332] 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。

[0333] 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。

[0334] 能发酵 C<sub>5</sub> 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<sub>5</sub> 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 (*Candidaboidinii*)、芸薹假丝酵母 (*Candida brassicae*)、休哈塔假丝酵母 (*Candidasheatae*)、迪丹斯假丝酵母 (*Candida diddensii*)、假热带假丝酵母 (*Candidapseudotropicalis*) 或产朊假丝酵母 (*Candida utilis*) 的菌株。

[0335] 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的大肠杆菌菌株。

[0336] 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)。

[0337] 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)。

[0338] 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。

[0339] 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 Fermentis/Lesaffre, USA 获得)、FALI™(可从 Fleischmann's Yeast, USA 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)。

[0340] 在一个优选的方面,发酵微生物已经经过遗传修饰,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。

[0341] 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)。

[0342] 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菌种。

[0343] 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。

[0344] 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。

[0345] 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。

[0346] 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。



[0347] 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 avitamin 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。

[0348] 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-葡萄糖酸、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。

[0349] 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 - asugar 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 feijerinckii* BA101 and in situ recovery by gas stripping, *World Journal of Microbiology and Biotechnology* 19(6):595-603。

[0350] 在另一个优选的方面,所述发酵产物是有机酸。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。在另一个更优选的方面,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。

[0351] 在另一个优选的方面,所述发酵产物是酮。可理解的是术语“酮”涵盖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。

[0352] 在另一个优选的方面,所述发酵产物是氨基酸。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。

[0353] 在另一个优选的方面,所述发酵产物是气体。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<sub>2</sub>。在另一个更优选的方面,所述气体是 CO<sub>2</sub>。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。

[0354] 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。

[0355] 植物

[0356] 本发明还涉及植物,例如,转基因植物、植物部分或植物细胞,其包含本发明的多核苷酸,从而以可回收的量表达和产生所述变体。变体可从植物或植物部分回收。或者,同样可以将含有所述变体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。

[0357] 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)。

[0358] 双子叶植物的实例是烟草 (tobacco),豆类 (legumes),如羽扇豆 (lupins),马铃薯

薯,糖甜菜 (sugar beet),豌豆,豆 (bean) 和大豆 (soybean) 和十字花科的 (cruciferous) 植物 (十字花科 (family Brassicaceae)),如花椰菜 (cauliflower),油菜籽 (rape seed) 和紧密相关的模型生物体拟南芥 (*Arabidopsis thaliana*)。

[0359] 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)。

[0360] 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。

[0361] 表达变体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:将编码变体的一个或多个 (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。

[0362] 表达构建体便利地是包含编码变体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整合了表达构建体的植物细胞有用的选择性标记,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)。

[0363] 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、何处以及如何表达变体而确定。例如,编码变体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 Physiol.* 86:506 所述。

[0364] 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. of 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 W0 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 *rbes* 启动子 (Kyojuka 等,1993, *Plant Physiology* 102:991-1000),小球藻病

毒 (chlorella virus) 腺嘌呤甲基转移酶 (adeninemethyltransferase) 基因启动子 (Mitra 和 Higgins, 1994, *Plant Mol. Biol.* 26:85-93), 来自稻的 aldP 基因启动子 (Kagaya 等, 1995, *Mol. 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) 和赤霉素 (gibberellic acid), 和重金属。

[0365] 启动子增强子元件也可以用于实现变体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变体的多核苷酸之间。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。

[0366] 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。

[0367] 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)。

[0368] 目前, 根癌土壤杆菌 (*Agrobacterium tumefaciens*) 介导的基因转移 (genetransfer), 是产生转基因双子叶植物的优选方法 (为了参考, 见 Hooykas 和 Schilperoort, 1992, *Plant 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 J.* 2:275-281; Shimamoto, 1994, *Current Opin. Biotech.* 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。其他用于依照本公开使用的转化方法包括那些描述于美国专利 6, 395, 966 和 7, 151, 204 的那些 (两者均通过全文提述并入本文)。

[0369] 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。

[0370] 除了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, 还可通过将具有所述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。例如, 可将编码变体的构建体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, 还包括此类植物的后代 (progeny)。如用于本文, 后代可指依照本发明制备的亲本植物任何世代的后裔 (offspring)。此种后代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。

[0371] 植物可通过回交转化方法生成。例如, 植物包括称作回交转化的基因型、种系、近交体 (inbred) 或杂交体 (hybrid) 的植物。

[0372] 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代中提供有关良种种质相对程度的数据。例如, 当具有期望性状但除此之外 (otherwise) 具有非农艺学期望的遗传背景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该目标性状, 还具有相对较大比例期望种质的后代。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。

[0373] 本发明还涉及产生本发明变体的方法, 包括: (a) 在有助于产生所述变体的条件下培养包含编码所述变体的多核苷酸的转基因植物或植物细胞; 和 (b) 回收所述变体。

[0374] 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

[0375] 材料

[0376] 用作缓冲剂和底物的化学品为至少试剂级的商品。

[0377] 菌株

[0378] 将烟曲霉 (NN051616) 用作编码家族 6A 纤维二糖水解酶 II 的 DNA 的来源。将米曲霉 JaL355 (WO 2002/40694) 用于表达烟曲霉纤维二糖水解酶 II 及其变体。

[0379] 培养基

[0380] PDA 平板包含 39 克的马铃薯右旋糖琼脂并添加去离子水至 1 升。

[0381] MDU2BP 包含 45g 的麦芽糖, 1g 的  $MgSO_4 \cdot 7H_2O$ , 1g 的 NaCl, 2g 的  $K_2SO_4$ , 12g 的  $KH_2PO_4$ , 7g 的酵母提取物, 2g 的尿素, 0.5ml 的 AMG 微量金属溶液, 和去离子水加至 1 升; pH 至 5.0。

[0382] AMG 微量金属溶液包含 14.3g 的  $ZnSO_4 \cdot 7H_2O$ , 2.5g 的  $CuSO_4 \cdot 5H_2O$ , 0.5g 的  $NiCl_2 \cdot 6H_2O$ , 13.8g 的  $FeSO_4 \cdot 7H_2O$ , 8.5g 的  $MnSO_4 \cdot H_2O$ , 3g 的柠檬酸, 和去离子水加至 1 升。

[0383] LB 培养基包含 10g 的胰蛋白胨, 5g 的酵母提取物, 5g 的氯化钠, 和去离子水加至 1 升。

[0384] 2X YT 平板包含 16g 的胰蛋白胨, 10g 的酵母提取物, 5g 的氯化钠, 15g 的 Noble 琼脂, 和去离子水加至 1 升。

[0385] 实施例 1: 烟曲霉基因组 DNA 提取。

[0386] 将烟曲霉在带挡板的烧瓶中 250ml 的马铃薯右旋糖培养基中在 37°C 和 240rpm 下生长。将菌丝体通过过滤收获, 在 10mM Tris-1mM EDTA (TE) 缓冲液中洗涤两次 (并在液氮下冻结)。将冻结的菌丝体通过研钵和杵磨碎至细微粉末, 将其重悬于含有 10mM Tris, 100mM EDTA, 1% Triton X-100, 0.5M 胍-HCl 和 200mM NaCl 的 pH 8.0 缓冲液中。将不含 DNase 的 RNaseA 以 20mg/ 升的浓度添加, 并将裂解物在 37°C 温育 30 分钟。细胞碎片通过离心去除, 且 DNA 通过使用 QIAGEN® Maxi 500 柱 (QIAGEN Inc., Valencia, CA, USA) 分离。将柱在 10ml 的经 30ml 的 QC 洗涤的 QBT 中平衡, 并用 15ml 的 QF (所有缓冲液来自 QIAGEN Inc., Valencia, CA, USA) 洗脱。将 DNA 在异丙醇中沉淀, 在 70% 乙醇中洗涤, 并通过离心回收。将 DNA 重悬于 TE 缓冲液。

[0387] 实施例 2: 用于烟曲霉家族 GH6A 纤维二糖水解酶 II 基因的表达载体的构建

[0388] 设计了下文中所示的两个合成寡核苷酸引物以从基因组 DNA PCR 扩增烟曲霉 GH6A 纤维二糖水解酶 II 的全长开读框。使用 TOPO Cloning Kit (Invitrogen, Carlsbad, CA, USA) 克隆 PCR 产物。使用 IN-FUSION™ Cloning Kit (BDBiosciences, Palo Alto, CA, USA) 将该片段克隆入 pAllo2 (WO 2004/099228)。

[0389] In-Fusion 正向引物：

[0390] 5' -ACTGGATTTACC **ATGAAGCACCTTGCATCTTCCATCG** -3' (SEQ IDNO:3)

[0391] In-Fusion 反向引物：

[0392] 5' -TCACCTCTAGTTAATTA **AAAGGACGGGTTAGCGT** -3' (SEQ ID NO:4)

[0393] 粗体字母代表编码序列。剩余序列与 pAllo2 的插入位点相比包含序列同一性。

[0394] 将五十皮摩尔的上述各引物用于 PCR 反应,所述反应包含 500ng 的烟曲霉基因组 DNA, 1X ThermoPol 反应缓冲液 (New England Biolabs, Ipswich, MA, USA), 6 μl 的 dATP、dTTP、dGTP 和 dCTP 的 10mM 混合物,和 0.1 单位的 Taq DNA 聚合酶 (New England Biolabs, Ipswich, MA, USA), 最终体积为 50 μl。扩增反应在 **EPPENDORF® MASTERCYCLER®** 5333 (Eppendorf Scientific, Inc., Westbury, NY, USA) 中进行,程序为一个循环在 98°C 进行 2 分钟;和 35 个循环,每循环在 96°C 进行 30 秒,61°C 进行 30 秒,和 72°C 进行 2 分钟。在 35 个循环之后,将反应物在 72°C 温育 10 分钟,然后在 10°C 冷却直至进一步处理。为了去除由 Taq DNA 聚合酶产生的 A-尾,将反应在 1 单位的 Pfx DNA 聚合酶 (Invitrogen, Carlsbad, CA, USA) 存在下在 68°C 温育 10 分钟。

[0395] 将 1.7kb PCR 反应产物使用 0.8%GTG-琼脂糖凝胶 (Cambrex Bioproducts, East Rutherford, NJ, USA) 用 40mM Tris 碱 -20mM 乙酸钠 -1mM EDTA 二钠盐 (TAE) 缓冲液和 0.1 μg 的溴乙锭每 ml 来分离。DNA 条带用 DARK READER™ (Clare Chemical Research, Dolores, CO, USA) 协助显影以避免 UV 诱导的突变。将 1.3kb DNA 条带用一次性刀片从凝胶切出,并使用 **ULTRAFREE®-DA** 旋转杯 (spin cup) (Millipore, Billerica, MA, USA) 根据生产商的指示进行纯化。

[0396] 将纯化的 1.7kb PCR 产物克隆入载体 **pCR®4Blunt-TOPO®** (Invitrogen, Carlsbad, CA, USA)。将两微升的纯化的 PCR 产物与 1 μl 的 2M 氯化钠溶液和 1 μl 的 **TOPO®** 载体混合。将反应在室温温育 15 分钟,然后使用 2 μl 的反应物根据生产商的指示 (Invitrogen, Carlsbad, CA, USA) 转化 TOP10 大肠杆菌感受态细胞。将 100 μl 各转化反应的两份等分试样铺板于两个补充 100 μg 每 ml 的氨苄青霉素的 150mm 2X YT 平板上,并在 37°C 温育过夜。

[0397] 将八个重组菌落分别接种于 3ml 的补充 100 μg 每 ml 的氨苄青霉素的 LB 培养基。使用 **BIOROBOT®** 9600 (QIAGEN Inc., Valencia, CA, USA) 从这些培养物制备质粒 DNA。克隆通过选择性消化进行分析。将来自每个克隆的质粒 DNA 用 Eco RI 消化,并通过使用 TAE 缓冲液的 1% 琼脂糖凝胶电泳进行分析。八个克隆中的六个具有预期的限制性消化样式。选择克隆 2、4、5、6、7 和 8 进行测序以确证在克隆的插入中不存在突变。对其 5' 和 3' 端的序列分析表明克隆 2、6 和 7 具有正确的序列。选择这三个克隆重新克隆入 pAllo2。将一微升的各克隆与 17 μl 的稀释 10 倍的 TE 缓冲液混合,并使用 1 μl 的这种混合物重新扩增烟曲霉 GH6A 纤维二糖水解酶编码区。

[0398] 将五十皮摩尔的上述各引物用于 PCR 反应,所述反应含有 1 μl 的克隆 2、6 和 7 的稀释混合物,1X Pfx Amplification Buffer (Invitrogen, Carlsbad, CA, USA),

6  $\mu$  l 的 dATP、dTTP、dGTP 和 dCTP 的 10mM 混合物, 2.5 单位的 **PLATINUM®** Pfx DNA 聚合酶 (Invitrogen, Carlsbad, CA, USA) 和 1  $\mu$  l 的 50mM  $MgSO_4$ , 最终体积为 50  $\mu$  l。使用 **EPPENDORF® MASTERCYCLER®** 5333 以扩增片段, 其程序为一个循环在 98°C 进行 2 分钟, 和 35 个循环, 每循环在 94°C 进行 30 秒, 61°C 进行 30 秒, 和 68°C 进行 1.5 分钟。在 35 个循环之后, 将反应在 68°C 温育 10 分钟, 然后在 10°C 冷却直至进一步处理。将 1.3kb PCR 反应产物使用 0.8%GTG- 琼脂糖凝胶用 TAE 缓冲液和 0.1  $\mu$  g 每 ml 的溴乙锭进行分离。DNA 条带以 DARK READER™ 的协助显影以避免 UV 诱导的突变。将 1.7kbDNA 条带用一次性刀片从凝胶切出, 并使用 **ULTRAFREE®-DA** 旋转杯根据生产商的指示进行纯化。

[0399] 将载体 pAlLo2 通过用 Nco I 和 Pac I 的消化直链化。将片段通过如上所述的琼脂糖凝胶电泳和超滤进行纯化。将纯化的 PCR 片段克隆入直链化和纯化的 pAlLo2 载体是使用 IN-FUSION™ Cloning Kit (BD Biosciences, Palo Alto, CA, USA) 进行的。反应 (20  $\mu$  l) 包含 1X IN-FUSION™ Buffer (BD Biosciences, Palo Alto, CA, USA), 1X BSA (BD Biosciences, Palo Alto, CA, USA), 1  $\mu$  l 的 IN-FUSION™ 酶 (1:10 稀释) (BD Biosciences, Palo Alto, CA, USA), 100ng 的用 Nco I 和 Pac I 消化的 pAlLo2, 和 50ng 的烟曲霉纯化 1.7kb PCR 产物。将反应在室温温育 30 分钟。将 2  $\mu$  l 的反应样品用于根据生产商的指示转化 TOP10 大肠杆菌感受态细胞。在恢复期之后, 将来自转化反应的两份 100  $\mu$  l 等分试样铺板于补充 100  $\mu$  g 每 ml 的氨苄青霉素的 150mm 2X YT 平板上。将平板在 37°C 温育过夜。从选择平板随机选择八个推定的重组克隆, 并从每一个使用 **BIOROBOT®** 9600 制备质粒 DNA。克隆通过 Pst I 限制性消化进行分析。八个克隆中的七个具有预期的限制性消化样式。然后对克隆 1、2 和 3 进行测序以确证在克隆的插入中不存在任何突变。选择克隆 #2 并命名为 pAlLo33。

[0400] 实施例 3 : 烟曲霉家族 GH6A 纤维二糖水解酶 II 基因变体的构建

[0401] 烟曲霉 GH6A 纤维二糖水解酶 II 的变体通过对 pAlLo33 (对于 pMaWo55、pMaWo 58、pMaWo78 和 pAhyG90), 或对 pAhyG90 (对于 pMaWo70 and pMaWo71)、pMaWo55 (对于 pMaWo72)、pMaWo58 (对于 pAhyG108)、pMaWo70 (对于 pMaWo73) 或 pMaWo71 (对于 pAhyG111) 使用 **QUIKCHANGE®** XLSite-Directed Mutagenesis Kit (Stratagene, La Jolla, CA, USA) 进行定点突变来构建。用于定点突变的寡聚物和获得的变体的总结示于表 1。

[0402] 所得的变体质粒 DNA 使用 **BIOROBOT®** 9600 制备。将变体质粒构建体使用 3130xl Genetic Analyzer (Applied Biosystems, Inc., Foster City, CA, USA) 测序以验证变化。

[0403] 表 1

[0404]

氨基酸变化	引物名称	序列	克隆质粒名称
D330N	MaWo167	cctacaccagggaAaccccaactgcgacg (SEQ ID NO: 5)	pMaWo55
	MaWo168	cgtcgcagttggggTtccctgggtgtagg (SEQ ID NO: 6)	
C254L	MaWo173	gagcgcctacctggagCTtgtcgactatgctctgaagc (SEQ ID NO: 7)	pMaWo58
	MaWo174	gcttcagagcatagtcgacaAGctccaggtaggcgctc (SEQ ID NO: 8)	
M342F	Af cel6a 3f	gaagaagtacateaacgccTtTgcgccttctcaaggaagccg (SEQ ID NO: 9)	pAhyG90
	Af cel6a 3r	cggcttccttgagaagagggcgcAaAggcggtgatgtacttctc (SEQ ID NO: 10)	
M342F + D330N	MaWo167	cctacaccagggaAaccccaactgcgacg (SEQ ID NO: 5)	pMaWo70
	MaWo168	cgtcgcagttggggTtccctgggtgtagg (SEQ ID NO: 6)	
M342F + C254L	MaWo173	gagcgcctacctggagCTtgtcgactatgctctgaagc (SEQ ID NO: 7)	pMaWo71
	MaWo174	gcttcagagcatagtcgacaAGctccaggtaggcgctc (SEQ ID NO: 8)	
C254L + D330N	MaWo173	gagcgcctacctggagCTtgtcgactatgctctgaagc (SEQ ID NO: 7)	pMaWo72
	MaWo174	gcttcagagcatagtcgacaAGctccaggtaggcgctc (SEQ ID NO: 8)	

[0405]



C254L + D330N + M342F	MaWo173	gagcgcctacctggagCTtgtcgactatgctctgaagc (SEQ ID NO: 7)	pMaWo73
	MaWo174	gcttcagagcatagtcgacaAGctccaggtagggcctc (SEQ ID NO: 8)	
S360G + C254L	MaWo181	gcccacttcatcatggataccGGGcggaatggcgtccagccc (SEQ ID NO: 11)	pAhyG108
	MaWo182	gggctggacgccattccgCCCggtatccatgatgaagtgggc (SEQ ID NO: 12)	
S360G + C254L + M342F	MaWo181	gcccacttcatcatggataccGGGcggaatggcgtccagccc (SEQ ID NO: 11)	pAhyG111
	MaWo182	gggctggacgccattccgCCCggtatccatgatgaagtgggc (SEQ ID NO: 12)	
L285I + G286Q	MaWo205	ggctcggatggccccccaacAtCCAAccgcccgaacactcttcgcc (SEQ ID NO: 13)	pMaWo78
	MaWo206	ggcgaagagtgttgcggcgggTTGGaTgttggcgggcatccgagcc (SEQ ID NO: 14)	

[0406] 实施例 4: 编码家族 GH6A 纤维二糖水解酶 II 变体的烟曲霉 cDNA 在米曲霉 JaL355 中的表达

[0407] 米曲霉 JaL355 原生质体根据 Christensen 等, 1988, Bio/Technology 6:1419-1422 的方法制备。将五  $\mu$ g 的表达载体 (pMaWo55、pMaWo58、pAhyG90、pMaWo70、pMaWo71、pMaWo72、pMaWo73、pAhyG108、pAhyG111 或 pMaWo78) 用于转化米曲霉 JaL355。将表达载体 pAlLo33 转化入米曲霉 JaL355 以供表达烟曲霉家族 GH6A 纤维二糖水解酶 II 基因。

[0408] 米曲霉 JaL355 用 pAlLo33、pMaWo55、pMaWo58、pAhyG90、pMaWo70、pMaWo71、pMaWo72、pMaWo73、pAhyG108、pAhyG111 或 pMaWo78 的转化对于每种载体生成了约 1-10 个转化体, 对于每次转化分离了多至四个转化体至单独 PDA 平板。

[0409] 将转化体的汇集 PDA 平板用 8ml 的 0.01% **TWEEN®** 20 洗涤, 并分别接种入无菌的 24 孔组织培养板中的 1ml 的 MDU2BP 培养基, 并在 34°C 温育。在温育三日之后, 将来自每个培养的 20  $\mu$ l 的收获的培养液使用 8-16%Tris-GlycineSDS-PAGE 凝胶 (Invitrogen, Carlsbad, CA, USA) 根据生产商的指示进行分析。培养物的 SDS-PAGE 概貌显示几个转化体具有大约 75kDa 的新的主要条带。

[0410] 将对于每次转化的一个转化体的汇集平板 (在 PDA 上生长) 用 8ml 的 0.01% **TWEEN®** 20 洗涤, 并接种入含有 25ml 的 MDU2BP 培养基的 125ml 塑料摇瓶, 并在

34℃静止或以 200rpm 温育,以生成培养液以供表征该酶。将烧瓶在第 3 日收获,并使用 0.22 μm GP Express plus Membrane (Millipore, Bedford, MA, USA) 过滤。

[0411] 实施例 5:测量烟曲霉家族 GH6A 纤维二糖水解酶 II 变体的热稳定性

[0412] 将 3ml 的过滤的来自实施例的测试培养的培养液使用 **ECONO-PAC®** 10DG Desalting Columns (Bio-Rad Laboratories, Inc., Hercules, CA, USA) 脱盐入 100mM NaCl-50mM 乙酸钠 pH 5.0。将在脱盐的培养液中的蛋白使用 **VIVASPIN®** 6 (5kDa 截取值) 超滤器 (Argos Technology, Elgin, IL, USA) 浓缩至 0.5ml 的体积。

[0413] 将浓缩的培养液使用 100mM NaCl-50mM 乙酸钠 pH 5.0 稀释至 1mg/ml 蛋白浓度。将各 1mg/ml 蛋白样品的两个 25 μl 等分试样添加至 **THERMOWELL®** 试管条 (tube strip) PCR 试管 (Corning, Corning, NY, USA)。将一份等分试样保持在冰上,而将另一份等分试样在 **EPPENDORF® MASTERCYCLER®** ep 梯度 S 热循环仪 (Eppendorf Scientific, Inc., Westbury, NY, USA) 中在 68℃加热 20 分钟,然后冷却至 4℃,然后置于冰上。然后将两个样品均用 175 μl 的 0.0114% **TWEEN®** 20-100mM NaCl-50mM 乙酸钠 pH 5.0 稀释。

[0414] 然后测量经加热的样品的残余活性,即确定经加热的样品和保持在冰上的样品在水解磷酸泡胀的纤维素 (PASC) 中的活性。将 10 微升的各样品一式三份添加至 96 孔 PCR 板 (Eppendorf, Westbury, NY, USA)。然后将 190 μl 的 0.01% **TWEEN®** 20-50mM 乙酸钠 pH 5.0 缓冲液中的 2.1g/l PASC 添加至 10 μl 的样品,并混合。将 50mM 乙酸钠 pH 5.0 缓冲液中的 100、75、50、25、12.5 和 0mg 每升的葡萄糖标样一式两份以 200 μl 每孔添加。将所得的混合物在 50℃在 **EPPENDORF® MASTERCYCLER®** ep 梯度 S 热循环仪中温育 30 分钟。反应通过将 50 μl 的 0.5M NaOH 添加至每个孔,包括葡萄糖标样,来终止。然后将平板在具有配置微板载具 (microplate carrier) 的 **SORVALL®** 1000B 转子的 **SORVALL®** RT 6000D 离心机 (Thermo Scientific, Waltham, MA, USA) 中以 2000rpm 离心 2 分钟。

[0415] 对 PASC 的活性通过在 50℃测量在 30 分钟水解过程中释放的还原端来确定。将 100 微升的来自各经离心的平板的上清转移至单独的 96 孔 PCR 平板。将 50 微升的 0.5M NaOH 中的 1.5%(w/v) PHBAH(4-羟基-二苯甲醇, SigmaChemical Co., St. Louis, MO, USA) 添加至每个孔。然后将平板在 **EPPENDORF® MASTERCYCLER®** ep 梯度 S 热循环仪中在 95℃加热 15 分钟,然后在 15℃进行 5 分钟。将总共 100 μl 的各样品转移至透明的平底 96 孔板 (Corning, Inc., Corning, NY, USA)。然后使用 **SPECTRAMAX®** 340pc 分光光度读板器 (Molecular Devices, Sunnyvale, CA, USA) 测量在 410nm 的吸光度。释放的还原端的浓度根据对释放的还原端的浓度对葡萄糖标样在 410nm 处的吸光度的直线拟合来确定。然后通过将由经加热的样品水解的 PASC 所释放的还原端除以由保持在冰上的样品水解的 PASC 所释放还原端来计算残余活性。将纤维二糖水解酶 II 变体的活性与亲本酶的活性相比较。

[0416] 图 1A 和 1B 中显示的结果说明了热稳定性增加,即所示变体相对于亲本酶的较高残余活性。

[0417] 实施例 6:测量烟曲霉家族 GH6A 纤维二糖水解酶 II 变体的水解活性

[0418] 将 150ml 含有烟曲霉家族 GH6A 纤维二糖水解酶 II 野生型或变体 pAhyG108、pAhyG111 或 pMaWo73 的培养液上清使用 10kDa 截取值 Vivaspin 超滤器

(Millipore, Billerica, MA, USA) 浓缩。将高至 13ml 的浓缩样品加载于用 100mM NaCl-20mM 乙酸钠 pH 5.0 缓冲液平衡的 **HiLoad®** 26/60Superdex 200 柱 (GE Healthcare Life Sciences, Piscataway, NJ, USA) 上。将蛋白用 100mM 氯化钠 -20mM 乙酸钠 pH 5.0 缓冲液洗脱。对于每个样品, 收集 7ml 的级分, 并在 8-16%Bior-Rad Criterion 无染色 Tris-HCl 凝胶 (Bio-Rad Laboratories, Inc., Hercules, CA, USA) 上对其进行 SDS-PAGE 分析。汇集在 230 至 240ml 洗脱的级分, 其显示具有大于 95% 的纯度。

[0419] 将纯化的样品使用 50mM 乙酸钠 pH 5.0 稀释至 125、112.5、101.3、91.1、82.0、73.8 和 66.4  $\mu$ g/ml 蛋白浓度。然后通过水解磷酸泡胀的纤维素 (PASC) 来测量样品的水解活性。将 10 微升的每种样品一式三份添加至 96 孔 PCR 平板 (Eppendorf, Westbury, NY, USA)。然后将 190  $\mu$ l 的 0.01% **TWEEN®** 20-50mM 乙酸钠 pH 5.0 缓冲液中的 2.1g/l PASC 添加至 10  $\mu$ l 的样品并混合。将 50mM 乙酸钠 pH 5.0 缓冲液中的 100、75、50、25、12.5 和 0mg 每升的葡萄糖标样一式两份以 200  $\mu$ l 每孔添加。将所得的混合物在 50°C 在 **EPPENDORF® MASTERCYCLER®** ep 梯度 S 热循环仪中温育 30 分钟。反应通过将 50  $\mu$ l 的 0.5M NaOH 添加至每个孔, 包括葡萄糖标样, 来终止。然后将平板在具有配置微板载具的 **SORVALL®** 1000B 转子的 **SORVALL®** RT 6000D 离心机 (ThermoScientific, Waltham, MA, USA) 中以 2000rpm 离心 2 分钟。

[0420] 对 PASC 的活性通过在 50°C 测量在 30 分钟水解过程中释放的还原端来确定。将 100 微升的来自各经离心的平板的上清转移至单独的 96 孔 PCR 平板。将 50 微升的 0.5M NaOH 中的 1.5% (w/v) PHBAH (4-羟基-二苯甲醇, SigmaChemical Co., St. Louis, MO, USA) 添加至每个孔。然后将平板在 **EPPENDORF® MASTERCYCLER®** ep 梯度 S 热循环仪中在 95°C 加热 15 分钟, 然后在 15°C 进行 5 分钟。将总共 100  $\mu$ l 的各样品转移至透明的平底 96 孔板 (Corning, Inc., Corning, NY, USA)。然后使用 **SPECTRAMAX®** 340pc 分光光度读板器 (Molecular Devices, Sunnyvale, CA, USA)。测量在 410nm 的吸光度。释放的还原端的浓度根据对释放的还原端的浓度对葡萄糖标样在 410nm 处的吸光度的直线拟合来确定。将纤维二糖水解酶 II 变体的活性与亲本酶的活性相比较。

[0421] 图 2 中显示的结果说明了每种含有 S360G 突变的变体的水解活性相对于亲本酶增加。

[0422] 通过下述编号段落可进一步描述本发明：

[0423] [1] 一种亲本纤维二糖水解酶的分离的变体, 其在对应于 SEQ ID NO:2 的位置 254、285、286、330、342 和 360 的一个或多个 (几个) 位置包含取代, 其中所述变体具有纤维二糖水解酶活性。

[0424] [2] 段 1 的变体, 其中所述亲本纤维二糖水解酶是：

[0425] a. 多肽, 其与 SEQ ID NO:2 的成熟多肽具有至少 60% 序列同一性；

[0426] b. 多肽, 其由多核苷酸编码, 所述多核苷酸在低严格条件下与以下杂交：(i) SEQ ID NO:1 的成熟多肽基因组编码序列, (ii) 包含于 SEQ ID NO:1 的成熟多肽 cDNA 编码序列, 或 (iii) (i) 或 (ii) 的全长互补链；

[0427] c. 多肽, 其由多核苷酸编码, 所述多核苷酸与 SEQ ID NO:1 的成熟多肽基因组编码序列, 或包含于 SEQ ID NO:1 的成熟多肽 cDNA 编码序列具有至少 60% 同一性；

[0428] d. SEQ ID NO:2 的成熟多肽的片段, 其具有纤维二糖水解酶活性。

[0429] [3] 段 1 或 2 的变体,其中所述亲本纤维二糖水解酶与 SEQ ID NO:2 的成熟多肽具有至少 60%,例如至少 65%,至少 70%,至少 75%,至少 80%,至少 85%,至少 90%,至少 95%,至少 96%,至少 97%,至少 98%,至少 99% 或 100% 的序列同一性。

[0430] [4] 段 1-3 任一项的变体,其中所述亲本纤维二糖水解酶由多核苷酸编码,所述多核苷酸在低严格条件下与以下杂交:(i)SEQ ID NO:1 的成熟多肽基因组编码序列,(ii) 包含于 SEQ ID NO:1 的成熟多肽 cDNA 编码序列,或 (iii) (i) 或 (ii) 的全长互补链。

[0431] [5] 段 1-4 任一项的变体,其中所述亲本纤维二糖水解酶由多核苷酸编码,所述多核苷酸与 SEQ ID NO:1 的成熟多肽基因组编码序列或包含于 SEQ ID NO:1 的成熟多肽 cDNA 编码序列具有至少 60%,例如,至少 65%,至少 70%,至少 75%,至少 80%,至少 85%,至少 90%,至少 95%,至少 96%,至少 97%,至少 98%,至少 99%,或 100% 序列同一性。

[0432] [6] 段 1-5 任一项的变体,其中所述亲本纤维二糖水解酶包含 SEQ ID NO:2 的成熟多肽或由 SEQ ID NO:2 的成熟多肽组成。

[0433] [7] 段 1-5 任一项的变体,其中所述亲本纤维二糖水解酶是 SEQ ID NO:2 的成熟多肽的片段,其中所述片段具有纤维二糖水解酶活性。

[0434] [8] 段 1-7 任一项的变体,其与亲本纤维二糖水解酶的氨基酸序列具有至少 60%,例如至少 65%,至少 70%,至少 75%,至少 80%,至少 85%,至少 90%,至少 95% 同一性至少 96%,至少 97%,至少 98%,至少 99%,但少于 100% 序列同一性。

[0435] [9] 段 1-8 任一项的变体,其与 SEQ ID NO:2 的成熟多肽具有至少 60%,例如至少 65%,至少 70%,至少 75%,至少 80%,至少 85%,至少 90%,至少 95%,至少 96%,至少 97%,至少 98%,和至少 99%,但少于 100% 序列同一性。

[0436] [10] 段 1-9 任一项的变体,其中所述变体包含至少 370 个氨基酸残基,例如至少 390 或至少 410 个氨基酸残基。

[0437] [11] 段 1-10 任一项的变体,其中所述变体包含 1-5 个取代,如 1、2、3、4 或 5 个取代。

[0438] [12] 段 1-11 任一项的变体,其在对应于 SEQ ID NO:2 的位置 254 的位置包含取代。

[0439] [13] 段 12 的变体,其中所述取代为用 Leu 的取代。

[0440] [14] 段 1-13 任一项的变体,其在对应于 SEQ ID NO:2 的位置 330 的位置包含取代。

[0441] [15] 段 14 的变体,其中所述取代为用 Asn 的取代。

[0442] [16] 段 1-15 任一项的变体,其在对应于 SEQ ID NO:2 的位置 342 的位置包含取代。

[0443] [17] 段 16 的变体,其中所述取代为用 Phe 的取代。

[0444] [18] 段 1-17 任一项的变体,其在对应于 SEQ ID NO:2 的位置 360 的位置包含取代。

[0445] [19] 段 18 的变体,其中所述取代为用 Gly 的取代。

[0446] [20] 段 1-19 任一项的变体,其在对应于 SEQ ID NO:2 的位置 285 的位置包含取代。

[0447] [21] 段 20 的变体,其中所述取代为用 Ile 的取代。

- [0448] [22] 段 1-21 任一项的变体,其在对应于 SEQ ID NO:2 的位置 286 的位置包含取代。
- [0449] [23] 段 22 的变体,其中所述取代为用 Gln 的取代。
- [0450] [24] 段 1-23 任一项的变体,其在对应于 SEQ ID NO:2 的位置 254、285、286、330、342 和 360 的任何位置的至少两个位置包含取代。
- [0451] [25] 段 1-23 任一项的变体,其在对应于 SEQ ID NO:2 的位置 254、285、286、330、342 和 360 的任何位置的至少三个位置包含取代。
- [0452] [26] 段 1-23 任一项的变体,其在对应于 SEQ ID NO:2 的位置 254、285、286、330、342 和 360 的任何位置的至少四个位置包含取代。
- [0453] [27] 段 1-23 任一项的变体,其在对应于 SEQ ID NO:2 的位置 254、285、286、330、342 和 360 的任何位置的至少五个位置包含取代。
- [0454] [27] 段 1-23 任一项的变体,其在对应于 SEQ ID NO:2 的位置 254、285、286、330、342 和 360 的任何位置的每个位置均包含取代。
- [0455] [28] 段 1-27 任一项的变体,其包含一个或多个(几个)选自 C254L、L285I、G286Q、D330N、M342F 和 S360G 的取代。
- [0456] [29] 段 28 的变体,其包含取代 C254L。
- [0457] [30] 段 28 的变体,其包含取代 L285I。
- [0458] [31] 段 28 的变体,其包含取代 G286Q。
- [0459] [32] 段 28 的变体,其包含取代 D330N。
- [0460] [33] 段 28 的变体,其包含取代 M342F。
- [0461] [34] 段 28 的变体,其包含取代 S360G。
- [0462] [35] 段 28 的变体,其包含选自 C254L、L285I、G286Q、D330N、M342F 和 S360G 的两个取代。
- [0463] [36] 段 35 的变体,其包含取代 C254L+L285I。
- [0464] [37] 段 35 的变体,其包含取代 C254L+G286Q。
- [0465] [38] 段 35 的变体,其包含取代 C254L+D330N。
- [0466] [39] 段 35 的变体,其包含取代 C254L+M342F。
- [0467] [40] 段 35 的变体,其包含取代 C254L+S360G。
- [0468] [41] 段 35 的变体,其包含取代 L285I+G286Q。
- [0469] [42] 段 35 的变体,其包含取代 L285I+D330N。
- [0470] [43] 段 35 的变体,其包含取代 L285I+M342F。
- [0471] [44] 段 35 的变体,其包含取代 L285I+S360G。
- [0472] [45] 段 35 的变体,其包含取代 G286Q+D330N。
- [0473] [46] 段 35 的变体,其包含取代 G286Q+M342F。
- [0474] [47] 段 35 的变体,其包含取代 G286Q+S360G。
- [0475] [48] 段 28 的变体,其包含选自 C254L、L285I、G286Q、D330N、M342F 和 S360G 的三个取代。
- [0476] [49] 段 48 的变体,其包含取代 C254L+L285I+G286Q。
- [0477] [50] 段 48 的变体,其包含取代 C254L+L285I+D330N。

- [0478] [51] 段 48 的变体,其包含取代 C254L+L285I+M342F。
- [0479] [52] 段 48 的变体,其包含取代 C254L+L285I+S360G。
- [0480] [53] 段 48 的变体,其包含取代 C254L+G286Q+D330N。
- [0481] [54] 段 48 的变体,其包含取代 C254L+G286Q+M342F。
- [0482] [55] 段 48 的变体,其包含取代 C254L+G286Q+S360G。
- [0483] [56] 段 48 的变体,其包含取代 C254L+D330N+M342F。
- [0484] [57] 段 48 的变体,其包含取代 C254L+D330N+S360G。
- [0485] [58] 段 48 的变体,其包含取代 C254L+M342F+S360G。
- [0486] [59] 段 48 的变体,其包含取代 L285I+G286Q+D330N。
- [0487] [60] 段 48 的变体,其包含取代 L285I+G286Q+M342F。
- [0488] [61] 段 48 的变体,其包含取代 L285I+G286Q+S360G。
- [0489] [62] 段 48 的变体,其包含取代 L285I+D330N+M342F。
- [0490] [63] 段 48 的变体,其包含取代 L285I+D330N+S360G。
- [0491] [64] 段 48 的变体,其包含取代 L285I+M342F+S360G。
- [0492] [65] 段 48 的变体,其包含取代 G286Q+D330N+M342F。
- [0493] [66] 段 48 的变体,其包含取代 G286Q+D330N+S360G。
- [0494] [67] 段 48 的变体,其包含取代 G286Q+M342F+S360G。
- [0495] [68] 段 48 的变体,其包含取代 D330N+M342F+S360G。
- [0496] [69] 段 28 的变体,其包含选自 C254L、L285I、G286Q、D330N、M342F 和 S360G 的四个取代。
- [0497] [70] 段 69 的变体,其包含取代 C254L+L285I+G286Q+D330N。
- [0498] [71] 段 69 的变体,其包含取代 C254L+L285I+G286Q+M342F。
- [0499] [72] 段 69 的变体,其包含取代 C254L+L285I+D330N+M342F。
- [0500] [73] 段 69 的变体,其包含取代 C254L+G286Q+D330N+M342F。
- [0501] [74] 段 69 的变体,其包含取代 L285I+G286Q+D330N+M342F。
- [0502] [75] 段 69 的变体,其包含取代 C254L+L285I+G286Q+S360G。
- [0503] [76] 段 69 的变体,其包含取代 C254L+L285I+D330N+S360G。
- [0504] [77] 段 69 的变体,其包含取代 C254L+G286Q+D330N+S360G。
- [0505] [78] 段 69 的变体,其包含取代 L285I+G286Q+D330N+S360G。
- [0506] [79] 段 69 的变体,其包含取代 C254L+L285I+M342F+S360G。
- [0507] [80] 段 69 的变体,其包含取代 C254L+G286Q+M342F+S360G。
- [0508] [81] 段 69 的变体,其包含取代 L285I+G286Q+M342F+S360G。
- [0509] [82] 段 69 的变体,其包含取代 C254L+D330N+M342F+S360G。
- [0510] [83] 段 69 的变体,其包含取代 L285I+D330N+M342F+S360G。
- [0511] [84] 段 69 的变体,其包含取代 G286Q+D330N+M342F+S360G。
- [0512] [85] 段 28 的变体,其包含选自 C254L、L285I、G286Q、D330N、M342F 和 S360G 的五个取代。
- [0513] [86] 段 85 的变体,其包含取代 L285I+G286Q+D330N+M342F+S360G。
- [0514] [87] 段 85 的变体,其包含取代 C254L+G286Q+D330N+M342F+S360G。

- [0515] [88] 段 85 的变体,其包含取代 C254L+L285I+D330N+M342F+S360G。
- [0516] [89] 段 85 的变体,其包含取代 C254L+L285I+G286Q+M342F+S360G。
- [0517] [90] 段 85 的变体,其包含取代 C254L+L285I+G286Q+D330N+S360G。
- [0518] [91] 段 85 的变体,其包含取代 C254L+L285I+G286Q+D330N+M342F。
- [0519] [92] 段 28 的变体,其包含取代 C254L+L285I+G286Q+D330N+M342F+S360G。
- [0520] [93] 段 1-92 任一项的变体,进一步在对应于 SEQ ID NO:2 的位置 245、382、420、437 和 440 的任何位置包含一个或多个(几个)取代。
- [0521] [94] 段 93 的变体,其在对应于 SEQ ID NO:2 的位置 245、382、420、437 和 440 的任何位置包含两个取代。
- [0522] [95] 段 93 的变体,其在对应于 SEQ ID NO:2 的位置 245、382、420、437 和 440 的任何位置包含三个取代。
- [0523] [96] 段 93 的变体,其在对应于 SEQ ID NO:2 的位置 245、382、420、437 和 440 的任何位置包含四个取代。
- [0524] [97] 段 93 的变体,其在对应于 SEQ ID NO:2 的位置 245、382、420、437 和 440 的每个位置均包含取代。
- [0525] [98] 段 93-97 任一项的变体,其在对应于 SEQ ID NO:2 的位置 245 的位置包含取代。
- [0526] [99] 段 98 的变体,其中所述取代是 Ser。
- [0527] [100] 段 93-99 任一项的变体,其在对应于 SEQ ID NO:2 的位置 382 的位置包含取代。
- [0528] [101] 段 100 的变体,其中所述取代是 Cys。
- [0529] [102] 段 93-101 任一项的变体,其在对应于 SEQ ID NO:2 的位置 420 的位置包含取代。
- [0530] [103] 段 102 的变体,其中所述取代是 Ile。
- [0531] [104] 段 93-103 任一项的变体,其在对应于 SEQ ID NO:2 的位置 437 的位置包含取代。
- [0532] [105] 段 104 的变体,其中所述取代是 Gln。
- [0533] [106] 段 93-105 任一项的变体,其在对应于 SEQ ID NO:2 的位置 440 的位置包含取代。
- [0534] [107] 段 106 的变体,其中所述取代是 Cys。
- [0535] [108] 段 93-107 任一项的变体,包含一个或多个(几个)选自 A245S、G382C、L420I、T437Q 和 Q440C 的取代。
- [0536] [109] 段 108 的变体,包含选自 A245S、G382C、L420I、T437Q 和 Q440C 的两个取代。
- [0537] [110] 段 109 的变体,包含取代 L420I+T437Q。
- [0538] [111] 段 109 的变体,包含取代 A245S+L420I。
- [0539] [112] 段 109 的变体,包含取代 G382C+L420I。
- [0540] [113] 段 109 的变体,包含取代 L420I+Q440C。
- [0541] [114] 段 109 的变体,包含取代 A245S+T437Q。
- [0542] [115] 段 109 的变体,包含取代 G382C+T437Q。

- [0543] [116] 段 109 的变体, 包含取代 T437Q+Q440C。
- [0544] [117] 段 109 的变体, 包含取代 A245S+G382C。
- [0545] [118] 段 109 的变体, 包含取代 A245S+Q440C。
- [0546] [119] 段 109 的变体, 包含取代 G382C+Q440C。
- [0547] [120] 段 108 的变体, 包含选自 A245S、G382C、L420I、T437Q 和 Q440C 的三个取代。
- [0548] [121] 段 120 的变体, 包含取代 A245S+L420I+T437Q。
- [0549] [122] 段 120 的变体, 包含取代 G382C+L420I+T437Q。
- [0550] [123] 段 120 的变体, 包含取代 L420I+T437Q+Q440C。
- [0551] [124] 段 120 的变体, 包含取代 A245S+G382C+L420I。
- [0552] [125] 段 120 的变体, 包含取代 A245S+L420I+Q440C。
- [0553] [126] 段 120 的变体, 包含取代 G382C+L420I+Q440C。
- [0554] [127] 段 120 的变体, 包含取代 A245S+G382C+T437Q。
- [0555] [128] 段 120 的变体, 包含取代 A245S+T437Q+Q440C。
- [0556] [129] 段 120 的变体, 包含取代 G382C+T437Q+Q440C。
- [0557] [130] 段 120 的变体, 包含取代 A245S+G382C+Q440C。
- [0558] [131] 段 108 的变体, 包含选自 A245S、G382C、L420I、T437Q 和 Q440C 的四个取代
- [0559] [132] 段 131 的变体, 包含取代 A245S+G382C+T437Q+Q440C。
- [0560] [133] 段 131 的变体, 包含取代 A245S+G382C+L420I+Q440C。
- [0561] [134] 段 131 的变体, 包含取代 G382C+L420I+T437Q+Q440C。
- [0562] [135] 段 131 的变体, 包含取代 A245S+L420I+T437Q+Q440C。
- [0563] [136] 段 131 的变体, 包含取代 A245S+G382C+L420I+T437Q。
- [0564] [137] 段 108 的变体, 包含取代 A245S+G382C+L420I+T437Q+Q440C。
- [0565] [138] 段 1-137 任一项的变体, 其相对于亲本纤维二糖水解酶具有改善的热稳定性。
- [0566] [139] 段 1-138 任一项的变体, 其中所述成熟多肽是 SEQ ID NO:2 的氨基酸 20-454。
- [0567] [140] 段 1-139 任一项的变体, 其中所述成熟多肽编码序列是 SEQ IDNO:1 的核苷酸 58-1710。
- [0568] [141] 一种分离的多核苷酸, 其编码段 1-140 任一项的变体。
- [0569] [142] 一种核酸构建体, 其包含段 141 的多核苷酸。
- [0570] [143] 一种表达载体, 其包含段 141 的多核苷酸。
- [0571] [144] 一种宿主细胞, 其包含段 141 的多核苷酸。
- [0572] [145] 一种产生亲本纤维二糖水解酶的变体的方法, 其包括:
- [0573] (a) 在适于表达所述变体的条件下培养段 144 的宿主细胞; 和
- [0574] (b) 回收所述变体。
- [0575] [146] 一种转基因植物、植物部分或植物细胞, 其用段 141 的多核苷酸转化。
- [0576] [147] 一种产生段 1-140 任一项的变体的方法, 其包括:
- [0577] (a) 在有助于所述变体产生的条件下培养包含编码所述变体的多核苷酸的转基因植物或植物细胞; 和



- [0578] (b) 回收所述变体。
- [0579] [148] 一种用于获得段 1-140 任一项的变体的方法,其包括:
- [0580] (a) 在对应于 SEQ ID NO:2 的成熟多肽的位置 254、285、286、330、342 和 360 的一个或多个(几个)位置将取代导入亲本纤维二糖水解酶,其中所述变体具有纤维二糖水解酶活性;和
- [0581] (b) 回收所述变体。
- [0582] [149] 段 148 的方法,进一步包括在对应于 SEQ ID NO:2 的成熟多肽的位置 245、382、420、437 和 440 的一个或多个(几个)位置将取代导入亲本纤维二糖水解酶。
- [0583] [150] 一种酶组合物,其包含段 1-140 任一项的变体。
- [0584] [151] 段 150 的酶组合物,进一步包括选自下组的纤维素分解酶:内切葡聚糖酶、纤维二糖水解酶和  $\beta$ -葡萄糖苷酶。
- [0585] [152] 段 150 或 151 的酶组合物,进一步包含具有纤维素分解增强活性的多肽。
- [0586] [153] 段 150-152 任一项的酶组合物,进一步包含选自下组的酶:木聚糖酶、半纤维素酶、酯酶、蛋白酶、漆酶和过氧化物酶。
- [0587] [154] 一种用于降解或转化纤维素材料的方法,其包括用段 150-153 任一项的酶组合物处理所述纤维素材料。
- [0588] [155] 段 154 的方法,其中所述纤维素材料经预处理。
- [0589] [156] 段 154 或 155 的方法,进一步包括回收经降解的纤维素材料。
- [0590] [157] 段 156 的方法,其中所述经降解的纤维素材料是糖。
- [0591] [158] 段 157 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。
- [0592] [159] 一种用于从纤维素材料产生发酵产物的方法,其包括:
- [0593] (a) 用段 150-153 任一项的酶组合物糖化所述纤维素材料以产生经糖化的纤维素材料;
- [0594] (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
- [0595] (c) 回收所述发酵产物。
- [0596] [160] 段 159 的方法,其中所述纤维素材料经预处理。
- [0597] [161] 段 159 或 160 的方法,其中步骤 (a) 和 (b) 同时进行。
- [0598] [162] 段 159-161 任一项的方法,其中所述发酵产物是醇、有机酸、酮、氨基酸或气体。
- [0599] [163] 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所述纤维素材料,其中所述纤维素材料用段 150-153 任一项的酶组合物糖化。
- [0600] [164] 段 163 的方法,其中所述纤维素材料的发酵产生发酵产物。
- [0601] [165] 段 163 或 164 的方法,进一步包括回收所述发酵产物。
- [0602] [166] 段 163-165 任一项的方法,其中所述纤维素材料在糖化之前经预处理。
- [0603] [167] 段 163-166 任一项的方法,其中所述发酵产物是醇、有机酸、酮、氨基酸或气体。
- [0604] 本文中描述和要求保护的方面的范围并不受本文中公开的具体方面的限制,因为这些方面旨在作为本发明几个方面的说明。任何等同的方面旨在落于本发明的范围内。事

实上,根据前述描述,除了本文中显示和描述的那些之外,本发明多种修饰对于本领域技术人员会是显而易见的。这些修饰亦旨在落在所附权利要求的范围内。在冲突的情况下,以包括定义的本公开为准。

[0001]

## 序列表

<110> 诺维信公司  
 Wogulis, Mark  
 <120> 纤维二糖水解酶变体及编码其的多核苷酸  
 <130> 11730-WO-PCT  
 <150> US 61/319,672  
 <151> 2010-03-31  
 <160> 14  
 <170> PatentIn version 3.5  
 <210> 1  
 <211> 1713  
 <212> DNA  
 <213> 烟曲霉(*Aspergillus fumigatus*)  
 <400> 1  
 atgaagcacc ttgcatcttc catcgcaatg actctactgt tgccctgccg gcaggcccag 60  
 cagaccgtat ggggccaatg tatgttcigg ctgtcactgg aataagactg tatcaactgc 120  
 fgatatgctt ctaggtggcg gccaaagctg gtctggcccg acgagctgtg ttgccggcgc 180  
 agcctgtagc acactgaate cctgtatggt agatategtc ctgagtggag acttatactg 240  
 acttccttag actacgetca gtgtatcccg ggagcccacg cgacgtccac taccctcaeg 300  
 acgacgacgg cggcgacgac gacatcccag accaccacca aacctaccac gaetggcca 360  
 actacatccg caccaccctg gaccgcatcc ggtaaccctt tcagoggcta ccagctgtat 420  
 gccaacccct actactcttc cgaggtccat actctggcca tgccttctct gccagctcg 480  
 ctgcagccca aggetagtgc tgttgcigaa gtgccctcat ttgtttgget gtaagtggcc 540  
 ttatcccaat actgagacca actctctgac agtcgtagcg acgttgccgc caagtgccc 600  
 actatgggaa cctacctgga cgacattcag gccaagaaca aggcgggcgc caaccctcct 660  
 atcgctggta tcttctgggt ctacgacttg cggaccctg actgcgccgc tctggccagt 720  
 aatggcgagt actcaattgc caacaacggt gtggccaact acaaggcgta cattgacgcc 780  
 atccgtgctc agctggtgaa gtactctgac gttcacacca tctctgtcat cgttaggccg 840  
 tacacctccg ttgcgcgccg cttttctctg acatcttga gaaccgaca gcttggccaa 900  
 cctggtgacc aacctcaacg tcgccaatg cgccaatgag cagagccct acctggagtg 960  
 tgtgactat getetgaagc agctcaacct gcccaacgtc gccatgtacc tcgacgcagg 1020  
 tatgcctcac ttcccgcatt ctgtatccct tccagacact aactcatcag gccatgcggg 1080  
 ctggctcgga tggcccgcga acttgggccc cgccgcaaca ctcttcgcca aagtctaac 1140  
 cgacgcgggt tccccgcgg ctgttcgtgg cctggccacc aacgtcgcca actacaacgc 1200  
 ctggtcgctc agtacctgcc cctctacac ccaggagac cccaactgag acgagaagaa 1260  
 gtacatcaac gccatggcgc ctcttctcaa ggaagccggc ttgatgccc acttcatcat 1320  
 ggatacctgt aagtgttat tccaategcc gatgtgtgac gactaatcaa tgtttcagcc 1380  
 cggaatggcg tccagcccac gaagenaaac gcctggggtg actggtgcaa cgtcctcgcc 1440

[0002]

accggetteg ggttgcgcc cfcgactaac accggcgatc cgtccagga tgcctttgtg 1500  
 tggatcaagc ccggtggaga gagtgatgc acgtccaact cgaattcccc ccggtatgac 1560  
 ggcactgag gatatagtga tgcctgcag cctgctctg aggettggtac ttggttccag 1620  
 gtatgtcacc cattagccag atgagggata agtgactgac ggacctagge ctactttgag 1680  
 cagcttctga ccaacgctaa ccgctccttt taa 1713

<210> 2  
 <211> 454  
 <212> PRT  
 <213> 烟曲霉(Aspergillus fumigatus)

<400> 2

Met Lys His Leu Ala Ser Ser Ile Ala Leu Thr Leu Leu Leu Pro Ala  
 1 5 10 15

Val Gln Ala Gln Gln Thr Val Trp Gly Gln Cys Gly Gly Gln Gly Trp  
 20 25 30

Ser Gly Pro Thr Ser Cys Val Ala Gly Ala Ala Cys Ser Thr Leu Asn  
 35 40 45

Pro Tyr Tyr Ala Gln Cys Ile Pro Gly Ala Thr Ala Thr Ser Thr Thr  
 50 55 60

Leu Thr Thr Thr Thr Ala Ala Thr Thr Thr Ser Gln Thr Thr Thr Lys  
 65 70 75 80

Pro Thr Thr Thr Gly Pro Thr Thr Ser Ala Pro Thr Val Thr Ala Ser  
 85 90 95

Gly Asn Pro Phe Ser Gly Tyr Gln Leu Tyr Ala Asn Pro Tyr Tyr Ser  
 100 105 110

Ser Glu Val His Thr Leu Ala Met Pro Ser Leu Pro Ser Ser Leu Gln  
 115 120 125

Pro Lys Ala Ser Ala Val Ala Glu Val Pro Ser Phe Val Trp Leu Asp  
 130 135 140

Val Ala Ala Lys Val Pro Thr Met Gly Thr Tyr Leu Ala Asp Ile Gln  
 145 150 155 160

Ala Lys Asn Lys Ala Gly Ala Asn Pro Pro Ile Ala Gly Ile Phe Val  
 165 170 175

Val Tyr Asp Leu Pro Asp Arg Asp Cys Ala Ala Leu Ala Ser Asn Gly  
 180 185 190

Glu Tyr Ser Ile Ala Asn Asn Gly Val Ala Asn Tyr Lys Ala Tyr Ile  
 195 200 205

[0003]

Asp Ala Ile Arg Ala Gln Leu Val Lys Tyr Ser Asp Val His Thr Ile  
 210 215 220  
 Leu Val Ile Glu Pro Asp Ser Leu Ala Asn Leu Val Thr Asn Leu Asn  
 225 230 235 240  
 Val Ala Lys Cys Ala Asn Ala Gln Ser Ala Tyr Leu Glu Cys Val Asp  
 245 250 255  
 Tyr Ala Leu Lys Gln Leu Asn Leu Pro Asn Val Ala Met Tyr Leu Asp  
 260 265 270  
 Ala Gly His Ala Gly Trp Leu Gly Trp Pro Ala Asn Leu Gly Pro Ala  
 275 280 285  
 Ala Thr Leu Phe Ala Lys Val Tyr Thr Asp Ala Gly Ser Pro Ala Ala  
 290 295 300  
 Val Arg Gly Leu Ala Thr Asn Val Ala Asn Tyr Asn Ala Trp Ser Leu  
 305 310 315 320  
 Ser Thr Cys Pro Ser Tyr Thr Gln Gly Asp Pro Asn Cys Asp Glu Lys  
 325 330 335  
 Lys Tyr Ile Asn Ala Met Ala Pro Leu Leu Lys Glu Ala Gly Phe Asp  
 340 345 350  
 Ala His Phe Ile Met Asp Thr Ser Arg Asn Gly Val Gln Pro Thr Lys  
 355 360 365  
 Gln Asn Ala Trp Gly Asp Trp Cys Asn Val Ile Gly Thr Gly Phe Gly  
 370 375 380  
 Val Arg Pro Ser Thr Asn Thr Gly Asp Pro Leu Gln Asp Ala Phe Val  
 385 390 395 400  
 Trp Ile Lys Pro Gly Gly Glu Ser Asp Gly Thr Ser Asn Ser Thr Ser  
 405 410 415  
 Pro Arg Tyr Asp Ala His Cys Gly Tyr Ser Asp Ala Leu Gln Pro Ala  
 420 425 430  
 Pro Glu Ala Gly Thr Trp Phe Gln Ala Tyr Phe Glu Gln Leu Leu Thr  
 435 440 445  
 Asn Ala Asn Pro Ser Phe  
 450

<210> 3  
 <211> 37  
 <212> DNA

[0004]

<213> 烟曲霉( <i>Aspergillus fumigatus</i> )	
<400> 3	
actggattta ccatgaagca ccttgcattct tccatcg	37
<210> 4	
<211> 34	
<212> DNA	
<213> 烟曲霉( <i>Aspergillus fumigatus</i> )	
<400> 4	
tcacctctag ttaattaa ggaegggtta gcgt	34
<210> 5	
<211> 30	
<212> DNA	
<213> 烟曲霉( <i>Aspergillus fumigatus</i> )	
<400> 5	
cetacaccca gggaaacccc aactgcgacg	30
<210> 6	
<211> 30	
<212> DNA	
<213> 烟曲霉( <i>Aspergillus fumigatus</i> )	
<400> 6	
cgtcgcagtt ggggtttccc tgggtgtagg	30
<210> 7	
<211> 38	
<212> DNA	
<213> 烟曲霉( <i>Aspergillus fumigatus</i> )	
<400> 7	
gagcgcctac ctggagcttg tgcactatgc tctgaage	38
<210> 8	
<211> 38	
<212> DNA	
<213> 烟曲霉( <i>Aspergillus fumigatus</i> )	
<400> 8	
gettcaagc atagtcgaca agctccaggt aggegetc	38
<210> 9	
<211> 44	
<212> DNA	
<213> 烟曲霉( <i>Aspergillus fumigatus</i> )	
<400> 9	
gaagaagtac atcaacgect ttgcgcctct tctcaaggaa gccg	44
<210> 10	
<211> 44	
<212> DNA	
<213> 烟曲霉( <i>Aspergillus fumigatus</i> )	
<400> 10	
cggcttcett gagaagagc gcaaaggcgt tgatgtactt cttc	44

[0005]

<210>	11	
<211>	42	
<212>	DNA	
<213>	烟曲霉( <i>Aspergillus fumigatus</i> )	
<400>	11	
	gcccacttea teatggatac cgggcggaat ggcgtccagc cc	42
<210>	12	
<211>	42	
<212>	DNA	
<213>	烟曲霉( <i>Aspergillus fumigatus</i> )	
<400>	12	
	gggctggacg ccattccgcc cggatccat gatgaagtgg gc	42
<210>	13	
<211>	47	
<212>	DNA	
<213>	烟曲霉( <i>Aspergillus fumigatus</i> )	
<400>	13	
	ggctcggatg gccgcacaac atccaaccg ccgaacact cttcgc	47
<210>	14	
<211>	47	
<212>	DNA	
<213>	烟曲霉( <i>Aspergillus fumigatus</i> )	
<400>	14	
	ggcgaagagt gttcggcgg gttgatggtt ggcggccat ccgagcc	47

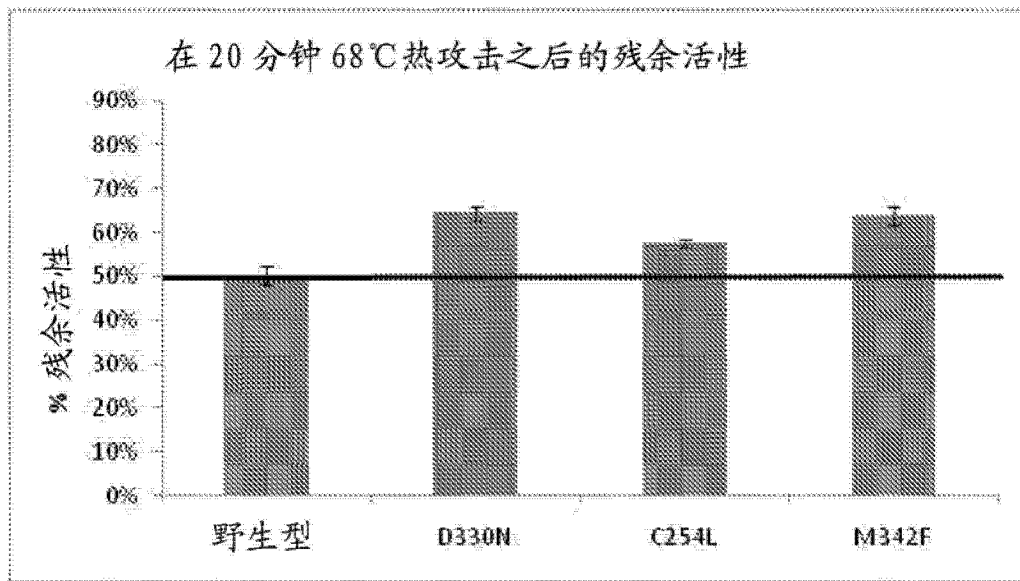


图 1A

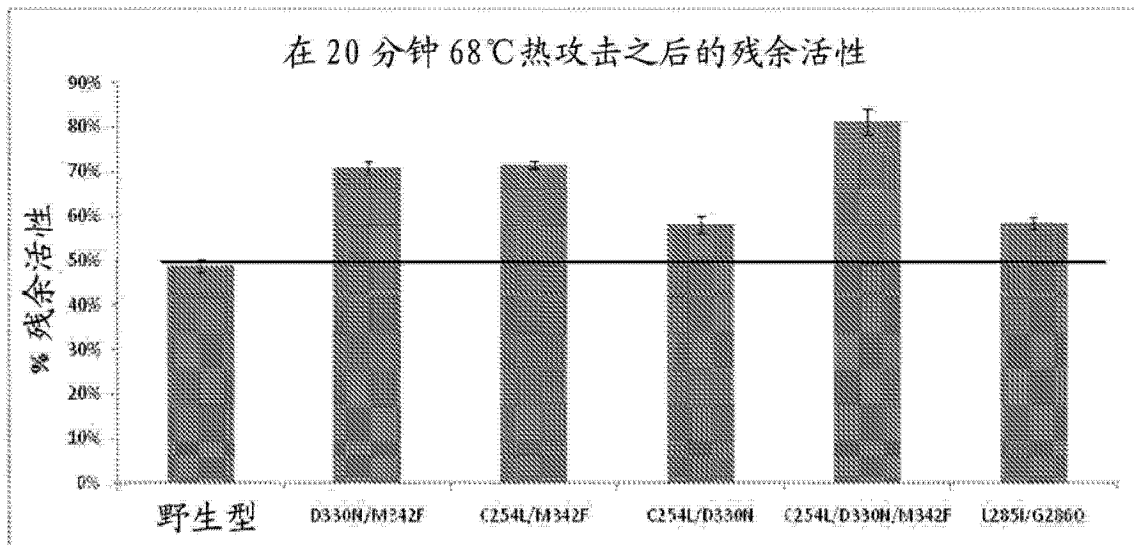


图 1B



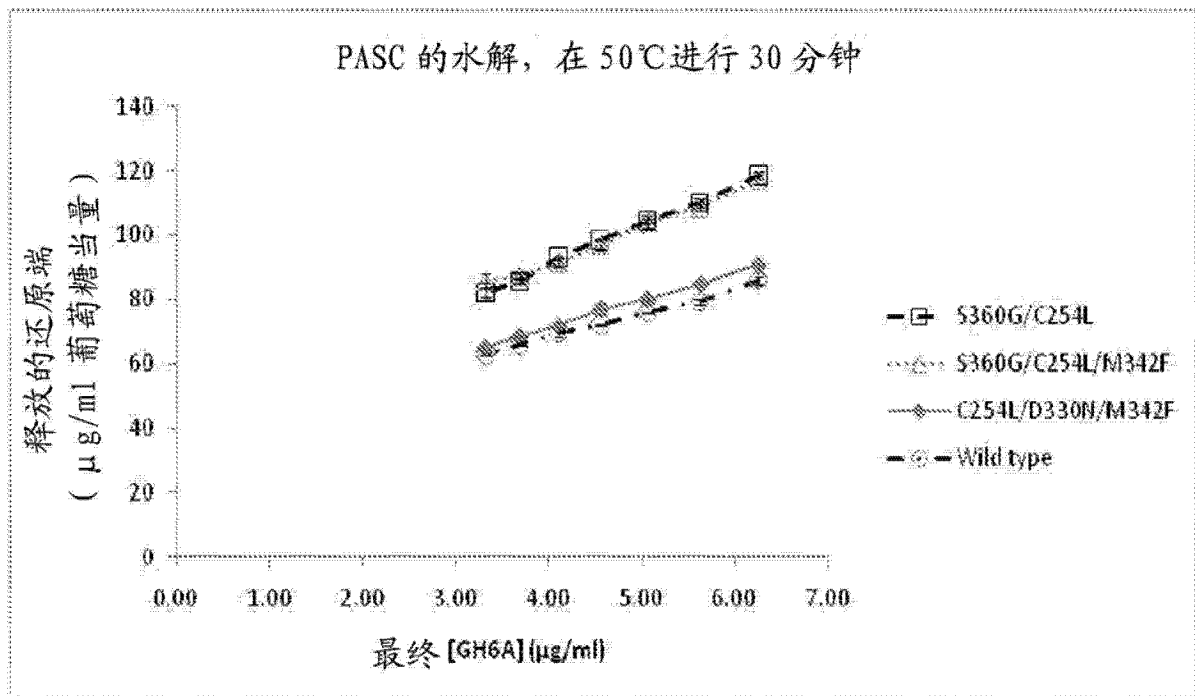


图 2