

[19] 中华人民共和国国家知识产权局



[12] 发明专利申请公布说明书

[21] 申请号 200780019316.3

[51] Int. Cl.

C12N 9/42 (2006.01)

C12N 15/82 (2006.01)

C12N 15/56 (2006.01)

C07K 14/37 (2006.01)

[43] 公开日 2009 年 6 月 17 日

[11] 公开号 CN 101460616A

[22] 申请日 2007.3.30

[21] 申请号 200780019316.3

[30] 优先权

[32] 2006.3.30 [33] US [31] 60/788,389

[86] 国际申请 PCT/US2007/065682 2007.3.30

[87] 国际公布 WO2007/115201 英 2007.10.11

[85] 进入国家阶段日期 2008.11.26

[71] 申请人 诺维信股份有限公司

地址 美国加利福尼亚州

共同申请人 诺维信公司

[72] 发明人 保罗·哈罗斯 埃琳娜·维拉森科

马库斯·S·考皮南

伊丽莎白·扎雷特斯基 萨拉·泰特

金伯利·布朗

[74] 专利代理机构 北京市柳沈律师事务所

代理人 史悦

权利要求书 6 页 说明书 85 页 序列表 10 页
附图 19 页

[54] 发明名称

具有内切葡聚糖酶活性的多肽和编码该多肽的多核苷酸

[57] 摘要

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

1. 具有内切葡聚糖酶活性的分离的多肽，其选自下组：

(a)包含氨基酸序列的多肽，所述氨基酸序列与 SEQ ID NO: 2 的成熟多肽具有至少 80%的同一性，与 SEQ ID NO: 4 的成熟多肽具有至少 70%的同一性，或与 SEQ ID NO: 6 的成熟多肽具有至少 85%的同一性；

(b)由多核苷酸编码的多肽，所述多核苷酸在至少高严紧条件下与下列杂交：(i) SEQ ID NO: 1 或 SEQ ID NO: 5 的成熟多肽编码序列，(ii)包含于 SEQ ID NO: 1 或 SEQ ID NO: 5 的成熟多肽编码序列中的 cDNA 序列，或(iii) (i)或(ii)的互补链，或者在至少中-高严紧条件下与下列杂交：(i) SEQ ID NO: 3 的成熟多肽编码序列，(ii)包含 SEQ ID NO: 3 的成熟多肽编码序列的基因组 DNA 序列，或(iii) (i)或(ii)的互补链；

(c)由包含核苷酸序列的多核苷酸编码的多肽，所述核苷酸序列与 SEQ ID NO: 1 的成熟多肽编码序列具有至少 80%同一性，与 SEQ ID NO: 3 的成熟多肽编码序列具有至少 70%同一性，或者与 SEQ ID NO: 5 的成熟多肽编码序列具有至少 85%同一性；和

(d)包含一个或几个氨基酸的取代、缺失和/或插入的 SEQ ID NO: 2、SEQ ID NO: 4 或 SEQ ID NO: 6 的成熟多肽的变体。

2. 权利要求1的多肽，其包含：与SEQ ID NO: 2的成熟多肽具有优选至少80%同一性、更优选至少85%同一性、甚至更优选至少90%同一性、最优选至少95%同一性的氨基酸序列；与SEQ ID NO: 4的成熟多肽具有优选至少70%同一性、更优选至少75%同一性、甚至更优选至少80%同一性、甚至更优选至少85%同一性、最优选至少90%同一性的氨基酸序列、甚至最优选至少95%同一性的氨基酸序列；或与SEQ ID NO: 6的成熟多肽具有优选至少85%同一性、更优选至少90%同一性、最优选至少95%同一性的氨基酸序列。

3. 权利要求1的多肽，其包含SEQ ID NO: 2、SEQ ID NO: 4或SEQ ID NO: 6的氨基酸序列，或其具有内切葡聚糖酶活性的片段；或由SEQ ID NO: 2、SEQ ID NO: 4或SEQ ID NO: 6的氨基酸序列，或其具有内切葡聚糖酶活性的片段组成。

4. 权利要求3的多肽，其包含SEQ ID NO: 2、SEQ ID NO: 4或SEQ ID NO: 6的氨基酸序列或由SEQ ID NO: 2、SEQ ID NO: 4或SEQ ID NO: 6的氨

氨基酸序列组成。

5. 权利要求3的多肽，其包含SEQ ID NO: 2、SEQ ID NO: 4或SEQ ID NO: 6的成熟多肽或由SEQ ID NO: 2、SEQ ID NO: 4或SEQ ID NO: 6的成熟多肽组成。

6. 权利要求1的多肽，其由多核苷酸所编码，所述多核苷酸在至少高严紧条件与下列杂交：(i) SEQ ID NO: 1或SEQ ID NO: 5的成熟多肽编码序列，(ii)包含于SEQ ID NO: 1或SEQ ID NO: 5的成熟多肽编码序列中的cDNA序列，或(iii) (i)或(ii)的互补链，或者在优选至少中-高严紧条件和更优选至少高严紧条件下与下列杂交：(i) SEQ ID NO: 3的成熟多肽编码序列，(ii)包含SEQ ID NO: 3的成熟多肽编码序列的基因组DNA序列，或(iii) (i)或(ii)的互补链。

7. 权利要求1的多肽，其由包含如下核苷酸序列的多核苷酸所编码：与SEQ ID NO: 1的成熟多肽编码序列具有优选至少80%同一性、更优选至少85%同一性、甚至更优选至少90%同一性、最优选至少95%同一性的核苷酸序列；与SEQ ID NO: 3的成熟多肽编码序列具有优选至少70%同一性、更优选至少75%同一性、甚至更优选至少80%同一性、甚至更优选至少85%同一性、最优选至少90%同一性、甚至最优选至少95%同一性的氨基酸序列；或者与SEQ ID NO: 5的成熟多肽编码序列具有优选至少85%同一性、更优选至少90%同一性、最优选至少95%同一性的氨基酸序列。

8. 权利要求1的多肽，其由多核苷酸所编码，所述多核苷酸包含SEQ ID NO: 1、SEQ ID NO: 3或SEQ ID NO: 5，或其编码具有内切葡聚糖酶活性的多肽片段的亚序列；或由SEQ ID NO: 1、SEQ ID NO: 3或SEQ ID NO: 5，或其编码具有内切葡聚糖酶活性的多肽片段的亚序列组成。

9. 权利要求8的多肽，其由多核苷酸所编码，所述多核苷酸包含SEQ ID NO: 1、SEQ ID NO: 3或SEQ ID NO: 5或由SEQ ID NO: 1、SEQ ID NO: 3或SEQ ID NO: 5组成。

10. 权利要求8的多肽，其由多核苷酸所编码，所述多核苷酸包含SEQ ID NO: 1、SEQ ID NO: 3或SEQ ID NO: 5的成熟多肽编码序列或由SEQ ID NO: 1、SEQ ID NO: 3或SEQ ID NO: 5的成熟多肽编码序列组成。

11. 权利要求1的多肽，其中所述多肽为包含一个或几个氨基酸的取代、缺失和/或插入的SEQ ID NO: 2、SEQ ID NO: 4或SEQ ID NO: 6的成熟多肽的

变体。

12. 权利要求1的多肽，其由包含于质粒pPH50、pPH38或pPH46中的多核苷酸所编码，其中质粒pPH50包含于大肠杆菌NRRL B-30899，pPH38包含于大肠杆菌NRRL B-30896，pPH46包含于大肠杆菌NRRL B-30897。

13. 权利要求1、2、5或11的多肽，其中所述成熟多肽是SEQ ID NO: 2的氨基酸23至464、SEQ ID NO: 4的氨基酸20至423或SEQ ID NO: 6的氨基酸19至440。

14. 权利要求1、6、7或10的多肽，其中所述成熟多肽编码序列是SEQ ID NO: 1的核苷酸79至1461、SEQ ID NO: 3的核苷酸74至1349或SEQ ID NO: 5的核苷酸107至1372。

15. 权利要求1-14任一项的多肽，进一步具有针对选自下组的一种或多种底物的酶活性：木聚糖、木葡聚糖、阿拉伯木聚糖、1,4-β-D-甘露聚糖和半乳甘露聚糖。

16. 包含核苷酸序列的分离的多核苷酸，所述核苷酸序列编码权利要求1-15中任一项的多肽。

17. 权利要求16的分离的多核苷酸，其在SEQ ID NO: 1、SEQ ID NO: 3或SEQ ID NO: 5的成熟多肽编码序列中包含至少一个突变，其中所述突变的核苷酸序列分别编码SEQ ID NO: 2、SEQ ID NO: 4或SEQ ID NO: 6的成熟多肽。

18. 权利要求17的分离的多核苷酸，其中所述成熟多肽编码序列是SEQ ID NO: 1的核苷酸79至1461、SEQ ID NO: 3的核苷酸74至1349或SEQ ID NO: 5的核苷酸107至1372，所述成熟多肽是SEQ ID NO: 2的氨基酸23至464、SEQ ID NO: 4的氨基酸20至423或SEQ ID NO: 6的氨基酸19至440。

19. 包含权利要求16-18中任一项的多核苷酸的核酸构建体，所述多核苷酸可操作地连接到一个或几个调控序列，所述调控序列指导多肽在表达宿主中产生。

20. 重组表达载体，其包含权利要求19的核酸构建体。

21. 重组宿主细胞，其包含权利要求19的核酸构建体。

22. 产生权利要求1-15中任一项的多肽的方法，其包括：(a)在有益于产生所述多肽的条件下培养细胞，所述细胞以其野生型形式产生所述多肽；和(b)回收所述多肽。

23. 产生权利要求1-15中任一项的多肽的方法，其包括：(a)在有益于产

生所述多肽的条件下培养包含核酸构建体的宿主细胞,所述核酸构建体含有编码所述多肽的核苷酸序列;和(b)回收所述多肽。

24. 产生亲本细胞的突变体的方法,其包括破坏或缺失编码权利要求1-15中任一项的多肽的核苷酸序列,得到与亲本细胞相比产生更少所述多肽的突变体。

25. 由权利要求24的方法产生的突变细胞。

26. 权利要求25的突变细胞,其还包含编码天然或异源蛋白质的基因。

27. 产生蛋白质的方法,其包括:(a)在有益于产生所述蛋白质的条件下培养权利要求26的突变细胞;和(b)回收所述蛋白质。

28. 权利要求16-18中任一项的分离的多核苷酸,其通过如下获得:(a)在至少高严谨条件下将DNA群体与下列杂交:(i) SEQ ID NO: 1或SEQ ID NO: 5的成熟多肽编码序列,(ii)包含于SEQ ID NO: 1或SEQ ID NO: 5的成熟多肽编码序列中的cDNA序列,或(iii) (i)或(ii)的互补链,或者在优选至少中-高严谨条件和更优选至少高严谨条件下将DNA群体与下列杂交:(i) SEQ ID NO: 3的成熟多肽编码序列,(ii)包含SEQ ID NO: 3的成熟多肽编码序列的基因组DNA序列,或(iii) (i)或(ii)的互补链杂交;和(b)分离杂交的多核苷酸,其编码具有内切葡聚糖酶活性的多肽。

29. 权利要求28的分离的多核苷酸,其中所述成熟多肽编码序列是SEQ ID NO: 1的核苷酸79至1461、SEQ ID NO: 3的核苷酸74至1349或SEQ ID NO: 5的核苷酸107至1372。

30. 产生包含突变的核苷酸序列的多核苷酸的方法,所述突变的核苷酸序列编码具有内切葡聚糖酶活性的多肽,所述方法包括:(a)将至少一个突变引入到SEQ ID NO: 1、SEQ ID NO: 3或SEQ ID NO: 5的成熟多肽编码序列中,其中所述突变的核苷酸序列编码由SEQ ID NO: 2、SEQ ID NO: 4或SEQ ID NO: 6的成熟多肽组成的多肽;和(b)回收包含所述突变的核苷酸序列的多核苷酸。

31. 权利要求30的方法,其中所述成熟多肽编码序列是SEQ ID NO: 1的核苷酸79至1461、SEQ ID NO: 3的核苷酸74至1349或SEQ ID NO: 5的核苷酸107至1372,所述成熟多肽是SEQ ID NO: 2的氨基酸23至464、SEQ ID NO: 4的氨基酸20至423或SEQ ID NO: 6的氨基酸19至440。

32. 由权利要求30或31的方法所产生的突变的多核苷酸。

33. 产生多肽的方法，其包括：(a)在有益于产生所述多肽的条件下培养包含权利要求32的突变的多核苷酸的细胞，所述突变的多核苷酸编码所述多肽；和(b)回收所述多肽。

34. 核酸构建体，其包含可操作地连接到编码信号肽的核苷酸序列的编码蛋白质的基因，所述信号肽包含SEQ ID NO: 2的氨基酸1到22、SEQ ID NO: 4的氨基酸1到19、SEQ ID NO: 6的氨基酸1到18或由SEQ ID NO: 2的氨基酸1到22、SEQ ID NO: 4的氨基酸1到19、SEQ ID NO: 6的氨基酸1到18组成，其中所述基因对于所述核苷酸序列来说是外源的。

35. 重组表达载体，其包含权利要求34的核酸构建体。

36. 重组宿主细胞，其包含权利要求34的核酸构建体。

37. 产生蛋白质的方法，其包括：(a)在有益于产生所述蛋白质的条件下培养权利要求36的重组宿主细胞；和(b)回收所述蛋白质。

38. 产生权利要求1-15中任一项的多肽的方法，其包括：(a)在有益于产生所述多肽的条件下培养转基因植物或植物细胞，所述转基因植物或植物细胞包含编码所述具有内切葡聚糖酶活性的多肽的多核苷酸；和(b)回收所述多肽。

39. 转基因植物、植物部分或植物细胞，其已经用编码权利要求1-15中任一项的多肽的多核苷酸转化。

40. 降解或转化纤维素材料的方法，其包括：用组合物处理所述纤维素材料，所述组合物包含有效量的权利要求1-15中任一项的具有内切葡聚糖酶活性的多肽或权利要求20的宿主细胞。

41. 权利要求40的方法，其中所述组合物还包含有效量的内切-1,4- β -葡聚糖酶、外切-1,4- β -D-葡聚糖酶和/或 β -D-葡萄糖苷酶。

42. 权利要求40或41的方法，其中所述组合物还包含有效量的选自下组的一种或多种酶：半纤维素酶、酯酶、蛋白酶、漆酶和过氧化物酶。

43. 权利要求40-42中任一项的方法，其中所述方法是预处理方法；同时糖化和发酵方法(SSF)中的步骤；或杂合水解和发酵方法(HHF)中的步骤。

44. 权利要求40-43中任一项的方法，进一步包括回收经降解或经转化的纤维素材料。

45. 产生物质的方法，其包括：(a)用组合物糖化纤维素材料，所述组合物包含有效量的权利要求1-15中任一项的具有内切葡聚糖酶活性的多肽或

权利要求20的宿主细胞, (b)用一种或多种发酵微生物对步骤(a)的经糖化的纤维索材料进行发酵; 和(c)自发酵回收所述物质。

46. 权利要求45的方法, 其中所述组合物还包含有效量的内切-1,4- β -葡聚糖酶、外切-1,4- β -D-葡聚糖酶和/或 β -D-葡萄糖苷酶。

47. 权利要求45或46的方法, 其中所述组合物还包含有效量的选自下组的一种或多种酶: 半纤维素酶、酯酶、蛋白酶、漆酶和过氧化物酶。

48. 权利要求45-47中任一项的方法, 其中在同时糖化和发酵中同时实施步骤(a)和(b)。

49. 权利要求45-48中任一项的方法, 其中所述物质是醇、有机酸、酮、氨基酸或气体。

50. 抑制多肽在细胞中表达的方法, 其包括将双链RNA (dsRNA)分子施用于所述细胞或者在所述细胞中表达双链RNA (dsRNA)分子, 其中所述dsRNA包含编码SEQ ID NO: 2、SEQ ID NO: 4或SEQ ID NO: 6的成熟多肽的多核苷酸的亚序列或部分。

51. 权利要求50的方法, 其中所述dsRNA长度为大约15、16、17、18、19、20、21、22、23、24、25或更多个双链体核苷酸。

52. 用于抑制多肽在细胞中表达的双链RNA (dsRNA)分子, 其中所述dsRNA包含编码SEQ ID NO: 2、SEQ ID NO: 4或SEQ ID NO: 6的成熟多肽的多核苷酸的亚序列或部分。

53. 权利要求52的双链RNA (dsRNA)分子, 其长度为大约15、16、17、18、19、20、21、22、23、24、25或更多个双链体核苷酸。

具有内切葡聚糖酶活性的多肽和编码该多肽的多核苷酸

涉及序列表

本申请包含计算机可读形式的序列表。该计算机可读形式通过引用并入本文。

涉及生物材料的保藏

本申请包含对生物材料的保藏的涉及，所述保藏已经依据布达佩斯条约在北区研究中心(Northern Regional Research Center) (NRRL)进行，并指定了登录号 NRRL B-30896、NRRL B-30897 和 NRRL B-30899，所述微生物保藏通过引用并入本文。

发明背景

发明领域

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

相关领域描述

纤维素是单糖葡萄糖通过 β -1,4-键共价连接的聚合物。许多微生物产生水解 β -连接的葡聚糖的酶。这些酶包括内切葡聚糖酶、纤维二糖水解酶和 β -葡糖苷酶。内切葡聚糖酶在随机位置消化纤维素聚合物，将其打开(opening it)以受到纤维二糖水解酶攻击(attack)。纤维二糖水解酶继而从纤维素聚合物的末端释放纤维二糖的分子。纤维二糖水解酶 I 是 1,4-D-葡聚糖纤维二糖水解酶(E.C. 3.2.1.91)活性，其催化纤维素、纤维四糖(cellotriose)或任何含有 β -1,4-连接的葡萄糖的聚合物中 1,4- β -D-糖苷键的水解，从链的还原性末端释放纤维二糖。纤维二糖水解酶 II 是 1,4- β -D-葡聚糖纤维二糖水解酶(E.C. 3.2.1.91)活性，其催化纤维素、纤维四糖或任何含有 β -1,4-连接的葡萄糖的聚合物中 1,4- β -D-糖苷键的水解，从链的非还原性末端释放纤维二糖。纤维二

糖是水溶性的 β -1,4-连接的葡萄糖二聚体。 β -葡糖苷酶将纤维二糖水解成葡萄糖。

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

Kvesitadaze 等, 1995, *Applied Biochemistry and Biotechnology* 50: 137-143 描述了来自土生梭孢霉(*Thielavia terrestris*)嗜热突变菌株的热稳定内切葡聚糖酶的分离和性质。Gilbert 等, 1992, *Bioresource Technology* 39: 147-154 描述了对土生梭孢霉 255B 的纤维素体系中存在的酶的特征。Breuil 等, 1986, *Biotechnology Letters* 8: 673-676 描述了来自土生梭孢霉菌株 C464 和 NRRL 8126 的纤维素酶和 β -葡糖苷酶的产生和定位。Kumar 等, 2000, *Bioresource Technology* 75: 95-97 公开了用 *Cladorrhinum foecundissimum* 生产内切-1,4- β -葡聚糖酶。

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

本发明的目的是提供改进的具有内切葡聚糖酶活性的多肽和编码该多肽的多核苷酸。

发明概述

本发明涉及具有内切葡聚糖酶活性的分离的多肽，其选自下组：

(a)包含氨基酸序列的多肽，所述氨基酸序列与 SEQ ID NO: 2 的成熟多肽具有至少 80% 的同一性，与 SEQ ID NO: 4 的成熟多肽具有至少 70% 的同一性，或与 SEQ ID NO: 6 的成熟多肽具有至少 85% 的同一性；

(b)由多核苷酸编码的多肽，所述多核苷酸在至少高严紧条件下与下列杂交：(i) SEQ ID NO: 1 或 SEQ ID NO: 5 的成熟多肽编码序列，(ii)包含于 SEQ ID NO: 1 或 SEQ ID NO: 5 的成熟多肽编码序列中的 cDNA 序列，或(iii) (i)

或(ii)的互补链, 或者在至少中-高严谨条件下与下列杂交: (i) SEQ ID NO: 3 的成熟多肽编码序列, (ii)包含 SEQ ID NO: 3 的成熟多肽编码序列的基因组 DNA 序列, 或(iii) (i)或(ii)的互补链;

(c)由包含核苷酸序列的多核苷酸编码的多肽, 所述核苷酸序列与 SEQ ID NO: 1 的成熟多肽编码序列具有至少 80%同一性, 与 SEQ ID NO: 3 的成熟多肽编码序列具有至少 70%同一性, 或者与 SEQ ID NO: 5 的成熟多肽编码序列具有至少 85%同一性; 和

(d)包含一个或几个氨基酸的取代、缺失和/或插入的 SEQ ID NO: 2、SEQ ID NO: 4 或 SEQ ID NO: 6 的成熟多肽的变体。

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

(a)编码多肽的多核苷酸, 该多肽包含: 与 SEQ ID NO: 2 的成熟多肽具有至少 80%同一性、与 SEQ ID NO: 4 的成熟多肽具有至少 70%同一性、与 SEQ ID NO: 6 的成熟多肽具有至少 85%同一性的氨基酸序列;

(b)多核苷酸, 其在至少高严谨条件下与以下序列杂交: (i) SEQ ID NO: 1 或 SEQ ID NO: 5 的成熟多肽编码序列, (ii) 包含于 SEQ ID NO: 1 或 SEQ ID NO: 5 的成熟多肽编码序列中的 cDNA 序列, 或(iii) (i)或(ii)的互补链; 或者其由在至少中-高严谨条件下与以下序列杂交: (i) SEQ ID NO: 3 的成熟多肽编码序列, (ii) 包含 SEQ ID NO: 3 的成熟多肽编码序列的基因组 DNA 序列, 或(iii) (i)或(ii)的互补链;

(c)包含核苷酸序列的多核苷酸, 所述核苷酸序列与 SEQ ID NO: 1 的成熟多肽编码序列具有至少 80%同一性、与 SEQ ID NO: 3 的成熟多肽编码序列具有至少 70%同一性, 或与 SEQ ID NO: 5 的成熟多肽编码序列具有至少 85%同一性; 和

(d)多核苷酸, 其编码包含一个或几个氨基酸的取代、缺失和/或插入的 SEQ ID NO: 2、SEQ ID NO: 4 或 SEQ ID NO: 6 的成熟多肽变体。

在优选的方面, 所述成熟多肽是 SEQ ID NO: 2 的氨基酸 23 至 464。在另一优选的方面, 所述成熟多肽是 SEQ ID NO: 4 的氨基酸 20 至 423。在另一优选的方面, 所述成熟多肽是 SEQ ID NO: 6 的氨基酸 19 至 440。在另一优选的方面, 所述成熟多肽编码序列是 SEQ ID NO: 1 的核苷酸 79 至 1461。在另一优选的方面, 所述成熟多肽编码序列是 SEQ ID NO: 3 的核苷酸 74 至

1349。在另一优选的方面,所述成熟多肽编码序列是 SEQ ID NO: 5 的核苷酸 107 至 1372。

本发明还涉及包含所述多核苷酸的核酸构建体、重组表达载体和重组宿主细胞,和产生具有内切葡聚糖酶活性的多肽的方法。

本发明还涉及抑制所述多肽在细胞中表达的方法,其包括:对细胞施用双链 RNA (dsRNA) 分子或在细胞中表达双链 RNA (dsRNA) 分子,其中 dsRNA 包括本发明的多核苷酸的亚序列。本发明还涉及这种双链抑制性 RNA (dsRNA) 分子,其中任选地 dsRNA 是 siRNA 或 miRNA 分子。

本发明还涉及在纤维素到葡萄糖和各种物质的转化中使用具有内切葡聚糖酶活性的多肽的方法。

本发明还涉及包含分离的多核苷酸的植物,所述多核苷酸编码这样的具有内切葡聚糖酶活性的多肽。

本发明还涉及用于产生这样的具有内切葡聚糖酶活性的多肽的方法,其包括:(a)在有益于产生该多肽的条件下培养包含多核苷酸的转基因植物或植物细胞,所述多核苷酸编码这样的具有内切葡聚糖酶的多肽;和(b)回收所述多肽。

本发明进一步涉及包含编码蛋白质的基因的核酸构建体,其中所述基因可操作地连接至编码信号肽的核苷酸序列,所述信号肽包含 SEQ ID NO: 2 的氨基酸 1 至 22、SEQ ID NO: 4 的氨基酸 1 至 19 或 SEQ ID NO: 6 的氨基酸 1 至 18 或由 SEQ ID NO: 2 的氨基酸 1 至 22、SEQ ID NO: 4 的氨基酸 1 至 19 或 SEQ ID NO: 6 的氨基酸 1 至 18 组成,其中所述基因相对于所述核苷酸序列来说是外源的。

附图简述

图 1 显示 pPH32 的限制图谱。

图 2 显示 pPH37 的限制图谱。

图 3 显示 pPH50 的限制图谱。

图 4 显示 pPH38 的限制图谱。

图 5A 和 5B 显示土生梭孢霉 NRRL 8126 CEL7C 内切葡聚糖酶的基因组 DNA 序列和推导的氨基酸序列(分别为 SEQ ID NO: 1 和 2)。

图 6A 和 6B 显示土生梭孢霉 NRRL 8126 CEL7E 内切葡聚糖酶的基因组

序列和推导的氨基酸序列(分别为 SEQ ID NO: 3 和 4)。

图 7 显示 pAIlLo1 的限制图谱。

图 8 显示 pBANe10 的限制图谱。

图 9 显示 pAIlLo2 的限制图谱。

图 10 显示 pEJG109 的限制图谱。

图 11 显示 pAIlLo25 的限制图谱。

图 12 显示 pPH46 的限制图谱。

图 13A 和 13B 显示 *Cladorrhinum foecundissimum* ATCC 62373 CEL7A 内切葡聚糖酶的 cDNA 序列和推导的氨基酸序列(分别为 SEQ ID NO: 5 和 6)。

图 14 显示在 pH 5.0 和 50°C, 土生梭孢霉 CEL7C 内切葡聚糖酶(5 mg 蛋白质每 g 固体)水解不同底物(5 mg/ml)的时间进程(time course)。

图 15 显示在 pH 5.5 和 60°C 水解 2 小时之后, β -葡聚糖(1% w/v)的相对转化率。

图 16 显示在 pH 5.5 和 60°C 水解 24 小时之后, β -葡聚糖(1% w/v)的相对转化率。

定义

内切葡聚糖酶活性: 术语“内切葡聚糖酶活性”在本文中定义为内切-1,4- β -D-葡聚糖 4-葡聚糖水解酶(endo-1,4- β -D-glucan 4-glucanohydrolase) (E.C. No. 3.2.1.4), 其催化纤维素、纤维素衍生物(例如羧甲基纤维素和羟乙基纤维素)、木质纤维素、木质纤维素衍生物、地衣淀粉(lichenin)中的 1,4- β -D-糖苷键、混合的 β -1,3 葡聚糖例如谷类 β -D-葡聚糖或木葡聚糖中的 β -1,4 键和含有纤维素成分的其它植物材料的内水解(endohydrolysis)。就本发明而言, 根据 Ghose, 1987, *Pure and Appl. Chem.* 59: 257-268 的方法, 使用羧甲基纤维素(CMC)水解来测定内切葡聚糖酶活性。将一单位的内切葡聚糖酶活性定义为: 在 50°C、pH 5.0, 每分钟产生 1.0 微摩尔还原糖。

在优选的方面, 具有内切葡聚糖酶活性的本发明的多肽还对选自下组的一种或多种底物具有酶活性: 木聚糖、木葡聚糖(xyloglucan)、阿拉伯木聚糖(arabinoxylan)、1,4- β -D-甘露聚糖和半乳甘露聚糖(galactomannan)。将具有内切葡聚糖酶活性的多肽对于这些多糖底物的活性测定为: 将所述底物(5 mg 每升)与本发明的具有内切葡聚糖酶活性的多肽(5 mg 蛋白质每 g 底物)一起

在 pH 5.0 (50 mM 乙酸钠)和 50°C 及间歇搅拌温育 24 小时之后,底物水解为还原糖的百分比。水解混合物中的还原糖通过对羟基苯甲酸酰肼(PHBAH)方法确定。

在更优选的方面,本发明具有内切葡聚糖酶活性的多肽进一步对木聚糖具有酶活性。在另一个更优选的方面,本发明具有内切葡聚糖酶活性的多肽进一步对木葡聚糖具有酶活性。在另一个更优选的方面,本发明具有内切葡聚糖酶活性的多肽进一步对阿拉伯木聚糖具有酶活性。在另一个更优选的方面,本发明具有内切葡聚糖酶活性的多肽进一步对 1,4- β -D-甘露聚糖具有酶活性。在另一个更优选的方面,本发明具有内切葡聚糖酶活性的多肽进一步对半乳甘露聚糖具有酶活性。在另一个更优选的方面,本发明具有内切葡聚糖酶活性的多肽进一步对木聚糖、木葡聚糖、阿拉伯木聚糖、1,4- β -D-甘露聚糖和/或半乳甘露聚糖具有酶活性。

本发明的多肽具有 SEQ ID NO: 2 或 SEQ ID NO: 4 或 SEQ ID NO: 6 的成熟多肽的内切葡聚糖酶活性的至少 20%,优选至少 40%,更优选至少 50%,更优选至少 60%,更优选至少 70%,更优选至少 80%,甚至更优选至少 90%,最优选至少 95%,并且甚至最优选至少 100%。

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

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

基本上纯的多肽: 术语“基本上纯的多肽”在本文表示多肽制备物,所述多肽制备物含有按重量计至多 10%,优选至多 8%,更优选至多 6%,更优选至多 5%,更优选至多 4%,更优选至多 3%,甚至更优选至多 2%,最优选至多 1%,并且甚至最优选至多 0.5%的与其天然或重组结合的(associated)的其它多肽材料。因此,优选所述基本上纯的多肽是按存在于制备物中的全部多肽材料的重量计至少 92%纯,优选至少 94%纯,更优选至少 95%纯,更优

选至少 96%纯, 更优选至少 96%纯, 更优选至少 97%纯, 更优选至少 98%纯, 甚至更优选至少 99%纯, 最优选至少 99.5%纯, 并且甚至最优选 100%纯。

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

在本文中, 术语“基本上纯的多肽”与术语“分离的多肽”和“分离形式的多肽”同义。

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

成熟多肽编码序列: 术语“成熟多肽编码序列”在本文中定义为编码具有内切葡聚糖酶活性的成熟多肽的核苷酸序列。

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

就本发明而言, 两个氨基酸序列之间的同一性程度通过 Needleman-Wunsch 算法(Needleman 和 Wunsch, 1970, *J. Mol. Biol.* 48: 443-453)使用 EMBOSS 的 Needle 程序来测定, 使用缺口罚分(gap penalty) 10, 缺口延伸罚分(gap extension penalty) 0.5 和 EBLOSUM62 矩阵。使用 Needle 标记为“最高同一性(longest identity)”的输出结果作为同一性百分比, 并计算如下:

$$(\text{同样的残基} \times 100) / (\text{比对长度} - \text{比对中缺口的数目})$$

就本发明而言, 两个核苷酸序列之间的同一性程度通过 Needleman-Wunsch 算法(Needleman 和 Wunsch, 1970, *J. Mol. Biol.* 48: 443-453)使用 EMBOSS 的 Needle 程序来测定, 使用参数缺口罚分 10, 缺口延伸罚分 0.5 和 EDNAFULL 矩阵。使用 Needle 标记为“最高同一性”的输出结果作为同一性百分比, 并计算如下:

$$(\text{同样的残基} \times 100) / (\text{比对长度} - \text{比对中缺口的数目})$$

同源序列: 术语“同源序列”在本文中定义为用土生梭孢霉 CEL7C 或 CEL7E 内切葡聚糖酶(分别为 SEQ ID NO: 2 或 SEQ ID NO: 4)或 *Cladorrhinum foecundissimum* CEL7A 内切葡聚糖酶(SEQ ID NO: 6)作为查询序列, 在 fasta

检索(Pearson, W.R., 1999, 于 *Bioinformatics Methods and Protocols* 中, S. Misener 和 S. A. Krawetz, ed., pp. 185-219)中给出小于 0.001 的 E 值(或预期分数)的预测蛋白质。

多肽片段: 术语“多肽片段”本文定义为从 SEQ ID NO: 2、SEQ ID NO: 4 或 SEQ ID NO: 6 或其同源序列的成熟多肽的氨基和/或羧基末端缺失了一个或几个氨基酸的多肽; 其中所述片段具有内切葡聚糖酶活性。在优选的方面, 片段包含 SEQ ID NO: 2 或其同源序列的成熟多肽的至少 380 个氨基酸残基、更优选至少 400 个氨基酸残基、最优选至少 420 个氨基酸残基。在另一优选的方面, 片段包含 SEQ ID NO: 4 或其同源序列的成熟多肽的至少 340 个氨基酸残基、更优选至少 360 个氨基酸残基、最优选至少 380 个氨基酸残基。在另一优选的方面, 片段包含 SEQ ID NO: 6 或其同源序列的成熟多肽的至少 360 个氨基酸残基、更优选至少 380 个氨基酸残基、最优选至少 400 个氨基酸残基。

亚序列: 术语“亚序列(subsequence)”在本文中定义为从 SEQ ID NO: 1、SEQ ID NO: 3 或 SEQ ID NO: 5 或其同源序列的成熟多肽编码序列的 5'和/或 3'端缺失了一个或几个核苷酸的核苷酸序列; 其中所述亚序列编码具有内切葡聚糖酶活性的多肽片段。在优选的方面, 亚序列包含 SEQ ID NO: 1 或其同源序列的成熟多肽编码序列的至少 1140 个核苷酸, 更优选至少 1200 个核苷酸, 最优选至少 1260 个核苷酸。在另一优选的方面, 亚序列包含 SEQ ID NO: 3 或其同源序列的成熟多肽编码序列的至少 1020 个核苷酸, 更优选至少 1080 个核苷酸, 最优选至少 1140 个核苷酸。在另一优选的方面, 亚序列包含 SEQ ID NO: 5 或其同源序列的成熟多肽编码序列的至少 1080 个核苷酸, 更优选至少 1140 个核苷酸, 最优选至少 1200 个核苷酸。

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

分离的多核苷酸: 术语“分离的多核苷酸”用于本文中指, 如通过琼脂糖电泳测定的, 其为至少 20%纯, 优选至少 40%纯, 更优选至少 60%纯, 甚至更优选至少 80%纯, 最优选至少 90%纯, 并且甚至最优选至少 95%纯的多

核苷酸。

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

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

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

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

核酸构建体：术语“核酸构建体”用于本文指单链或双链的核酸分子，

所述核酸分子分离自天然存在的基因，或将所述核酸分子以本来不存在于(not otherwise exist)自然界中的方式修饰以含有核酸的区段。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。

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

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

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

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

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

修饰: 术语“修饰”在本文的意思是，对由 SEQ ID NO: 2、SEQ ID NO: 4 或 SEQ ID NO: 6 的成熟多肽组成的多肽或其同源序列的任何化学修饰，以及对编码所述多肽的 DNA 的遗传操作。所述修饰可以是一个或几个氨基酸的取代、缺失和/或插入，以及一个或几个氨基酸侧链的置换。

人工变体: 当用在本文时，术语“人工变体”的意思是具有内切葡聚糖酶活性的多肽，所述多肽由表达 SEQ ID NO: 1、SEQ ID NO: 3 或 SEQ ID NO: 5 的成熟多肽编码序列或其同源序列的修饰的核苷酸序列的生物体产生。所述修饰的核苷酸序列通过人为干预(human intervention)，通过修饰公开于

SEQ ID NO: 1、SEQ ID NO: 3 或 SEQ ID NO: 5 或它们的同源序列的核苷酸序列来获得。

发明详述

具有内切葡聚糖酶活性的多肽

在第一个方面，本发明涉及包含下述氨基酸序列的分离的多肽，所述氨基酸序列与 SEQ ID NO: 2、SEQ ID NO: 4 或 SEQ ID NO: 6 的成熟多肽具有至少 60%，优选至少 65%，更优选至少 70%，更优选至少 75%，更优选至少 80%，更优选至少 85%，甚至更优选至少 90%，最优选至少 95%，并且甚至最优选至少 96%、97%、98%或 99%的同一性程度，所述多肽具有内切葡聚糖酶活性(下文中的“同源多肽”)。在优选的方面，所述同源多肽具有的氨基酸序列与 SEQ ID NO: 2、SEQ ID NO: 4 或 SEQ ID NO: 6 的成熟多肽相差十个氨基酸，优选相差五个氨基酸，更优选相差四个氨基酸，甚至更优选相差三个氨基酸，最优选相差两个氨基酸，并且甚至最优选相差一个氨基酸。

本发明的多肽优选包含 SEQ ID NO: 2 的氨基酸序列或其等位变体；或它们的具有内切葡聚糖酶活性的片段。在优选的方面，多肽包含 SEQ ID NO: 2 的氨基酸序列。在另一优选方面，多肽包含 SEQ ID NO: 2 的成熟多肽。在另一优选方面，多肽包含 SEQ ID NO: 2 的氨基酸 23 至 464，或其等位变体；或它们的具有内切葡聚糖酶活性的片段。在另一优选方面，多肽包含 SEQ ID NO: 2 的氨基酸 23 至 464。在另一优选方面，多肽由 SEQ ID NO: 2 的氨基酸序列或其等位变体，或它们的具有内切葡聚糖酶活性的片段组成。在另一优选方面，多肽由 SEQ ID NO: 2 的氨基酸序列组成。在另一优选方面，多肽由 SEQ ID NO: 2 的成熟多肽组成。在另一优选方面，多肽由 SEQ ID NO: 2 的氨基酸 23 至 464 或其等位变体，或它们的具有内切葡聚糖酶活性的片段组成。在另一优选方面，多肽由 SEQ ID NO: 2 的氨基酸 23 至 464 组成。

本发明的多肽还优选包含 SEQ ID NO: 4 的氨基酸序列或其等位变体；或其具有内切葡聚糖酶活性的片段。在优选的方面，多肽包含 SEQ ID NO: 4 的氨基酸序列。在另一优选的方面，多肽包含 SEQ ID NO: 4 的成熟多肽。在另一优选的方面，多肽包含 SEQ ID NO: 4 的氨基酸 20 至 423，或其等位变体；或它们的具有内切葡聚糖酶活性的片段。在另一优选的方面，多肽包含 SEQ ID NO: 4 的氨基酸 20 至 423。在另一优选的方面，多肽由 SEQ ID NO: 4

的氨基酸序列或其等位变体,或它们的具有内切葡聚糖酶活性的片段组成。在另一优选的方面,多肽由 SEQ ID NO: 4 的氨基酸序列组成。在另一优选的方面,多肽由 SEQ ID NO: 4 的成熟多肽组成。在另一优选的方面,多肽由 SEQ ID NO: 4 的氨基酸 20 至 423 或其等位变体,或它们的具有内切葡聚糖酶活性的片段组成。在另一优选的方面,多肽由 SEQ ID NO: 4 的氨基酸 20 至 423 组成。

本发明的多肽还优选包含 SEQ ID NO: 6 的氨基酸序列或其等位变体;或它们的具有内切葡聚糖酶活性的片段。在优选的方面,多肽包含 SEQ ID NO: 6 的氨基酸序列。在另一优选的方面,多肽包含 SEQ ID NO: 6 的成熟多肽。在另一优选的方面,多肽包含 SEQ ID NO: 6 的氨基酸 19 至 440 或其等位变体;或它们的具有内切葡聚糖酶活性的片段。在另一优选的方面,多肽包含 SEQ ID NO: 6 的氨基酸 19 至 440。在另一优选的方面,多肽由 SEQ ID NO: 6 的氨基酸序列或其等位变体,或它们的具有内切葡聚糖酶活性的片段组成。在另一优选的方面,多肽由 SEQ ID NO: 6 的氨基酸序列组成。在另一优选的方面,多肽由 SEQ ID NO: 6 的成熟多肽组成。在另一优选的方面,多肽由 SEQ ID NO: 6 的氨基酸 19 至 440 或其等位变体,或它们的具有内切葡聚糖酶活性的片段组成。在另一优选的方面,多肽由 SEQ ID NO: 6 的氨基酸 19 至 440 组成。

在第二个方面,本发明涉及具有内切葡聚糖酶活性的分离的多肽,所述分离的多肽由多核苷酸编码,所述多核苷酸在非常低严紧条件下,优选低严紧条件下,更优选中严紧条件下,更优选中-高严紧条件下,甚至更优选高严紧条件下,并且最优选非常高严紧条件下,与以下序列杂交: (i) SEQ ID NO: 1、SEQ ID NO: 3 或 SEQ ID NO: 5 的成熟多肽编码序列, (ii) 包含于 SEQ ID NO: 1 或 SEQ ID NO: 5 的成熟多肽编码序列中的 cDNA 序列或者包含 SEQ ID NO: 1 的成熟多肽编码序列的基因组 DNA 序列, (iii) (i) 或 (ii) 的亚序列, 或 (iv) (i)、(ii) 或 (iii) 的互补链(J. Sambrook, E.F. Fritsch, 和 T. Maniatis, 1989, *Molecular Cloning, A Laboratory Manual*, 2d edition, Cold Spring Harbor, New York)。SEQ ID NO: 1、SEQ ID NO: 3 或 SEQ ID NO: 5 的成熟多肽编码序列的亚序列含有至少 100 个连续的核苷酸或优选至少 200 个连续的核苷酸。此外,所述亚序列可编码具有内切葡聚糖酶活性的多肽片段。在优选的方面,所述成熟多肽编码序列是 SEQ ID NO: 1 的核苷酸 79 至 1461。在另一优选的方面,所述成

成熟多肽编码序列是 SEQ ID NO: 3 的核苷酸 74 至 1349。在另一优选的方面,所述成熟多肽编码序列是 SEQ ID NO: 5 的核苷酸 107 至 1372。在另一优选的方面,所述互补链是 SEQ ID NO: 1、SEQ ID NO: 3 或 SEQ ID NO: 5 的成熟多肽编码序列的全长互补链。

SEQ ID NO: 1、SEQ ID NO: 3 或 SEQ ID NO: 5 的核苷酸序列或其亚序列;以及 SEQ ID NO: 2、SEQ ID NO: 4 或 SEQ ID NO: 6 的氨基酸序列或其片段,可用于设计核酸探针,以根据本领域内公知的方法从不同属和种的菌株鉴定和克隆编码具有内切葡聚糖酶活性的多肽的 DNA。具体而言,根据标准的 Southern 印迹方法,可将这些探针用于与感兴趣的属或种的基因组或 cDNA 杂交,以鉴定和从其中分离相应的基因。这些探针可明显短于完整序列,但长度上应为至少 14,优选至少 25,更优选至少 35,并且最优选至少 70 个核苷酸。然而,优选所述核酸探针是至少 100 个核苷酸长度。例如,所述核酸探针长度上可以是至少 200 个核苷酸,优选至少 300 个核苷酸,更优选至少 400 个核苷酸,或最优选至少 500 个核苷酸。甚至可以使用更长的探针,例如,长度是至少 600 个核苷酸,至少优选至少 700 个核苷酸,更优选至少 800 个核苷酸,或最优选至少 900 个核苷酸的核酸探针。DNA 和 RNA 探针二者均可使用。通常将探针标记以探测相应的基因(例如,用 ^{32}P 、 ^3H 、 ^{35}S 、生物素或抗生物素蛋白(avidin)标记)。将这些探针包含于本发明中。

因而,可从由这些其它菌株制备的基因组 DNA 或 cDNA 文库中筛选 DNA,所述 DNA 与上述探针杂交并且编码具有内切葡聚糖酶活性的多肽。可以通过琼脂糖或聚丙烯酰胺凝胶电泳,或通过其它分离技术分离来自这些其它菌株的基因组或其它 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、SEQ ID NO: 3 或 SEQ ID NO: 5 或其亚序列同源的克隆或 DNA,将所述载体材料优选用在 Southern 印迹中。

就本发明而言,杂交表示核苷酸序列在非常低至非常高的严紧条件下与标记的核酸探针杂交,所述核酸探针对应于 SEQ ID NO: 1、SEQ ID NO: 3 或 SEQ ID NO: 5 的成熟多肽编码序列;包含于 SEQ ID NO:1 或 SEQ ID NO:5 的成熟多肽编码序列中的 cDNA 序列;包含 SEQ ID NO:3 的成熟多肽编码序列的基因组 DNA 序列;它的互补链或其亚序列。可使用例如 X 射线片(X-ray film)检测在这些条件下与核酸探针杂交的分子。

在优选的方面，核酸探针是 SEQ ID NO: 1 的成熟多肽编码序列。在另一优选方面，核酸探针是 SEQ ID NO: 1 的核苷酸 79 至 1461。在另一优选方面，核酸探针是编码 SEQ ID NO: 2 的多肽的多核苷酸序列，或其亚序列。在另一优选方面，核酸探针是 SEQ ID NO: 1。在另一优选方面，核酸探针是包含在大肠杆菌(*E. coli*) NRRL B-30899 中的质粒 pPH50 中含有的多核苷酸序列，其中所述其多核苷酸序列编码具有内切葡聚糖酶活性的多肽。在另一优选方面，核酸探针是包含在大肠杆菌 NRRL B-30899 中的质粒 pPH50 中含有的成熟多肽编码区。

在另一个优选的方面，核酸探针是 SEQ ID NO: 3 的成熟多肽编码序列。在另一优选方面，核酸探针是 SEQ ID NO: 3 的核苷酸 74 至 1349。在另一优选方面，核酸探针是编码 SEQ ID NO: 4 的多肽的多核苷酸序列，或其亚序列。在另一优选方面，核酸探针是 SEQ ID NO: 3。在另一优选方面，核酸探针是包含在大肠杆菌 NRRL B-30896 中的质粒 pPH38 中含有的多核苷酸序列，其中所述其多核苷酸序列编码具有内切葡聚糖酶活性的多肽。在另一优选方面，核酸探针是包含在大肠杆菌 NRRL B-30896 中的质粒 pPH38 中含有的成熟多肽编码区。

在另一个优选的方面，核酸探针是 SEQ ID NO: 5 的成熟多肽编码序列。在另一优选方面，核酸探针是 SEQ ID NO: 5 的核苷酸 107 至 1372。在另一优选方面，核酸探针是编码 SEQ ID NO: 6 的多肽的多核苷酸序列，或其亚序列。在另一优选方面，核酸探针是 SEQ ID NO: 5。在另一优选方面，核酸探针是包含在大肠杆菌 NRRL B-30897 中的质粒 pPH46 中含有的多核苷酸序列，其中所述其多核苷酸序列编码具有内切葡聚糖酶活性的多肽。在另一优选方面，核酸探针是包含在大肠杆菌 NRRL B-30897 中的质粒 pPH46 中含有的成熟多肽编码区。

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

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

在 55°C (中严紧性), 更优选至少在 60°C (中-高严紧性), 甚至更优选至少在 65°C (高严紧性), 并且最优选至少在 70°C (非常高严紧性)将载体材料最终洗涤三次, 每次 15 分钟。

对于长度大约 15 个核苷酸至大约 70 个核苷酸的短探针, 将严紧条件定义为在比根据 Bolton 和 McCarthy 计算法(1962, *Proceedings of the National Academy of Sciences USA* 48:1390)得出的 T_m 低大约 5°C 至大约 10°C, 在 0.9 M NaCl, 0.09 M Tris-HCl pH 7.6, 6 mM EDTA, 0.5% NP-40, 1 × Denhardt 溶液, 1 mM 焦磷酸钠(sodium pyrophosphate), 1 mM 磷酸二氢钠(sodium monobasic phosphate), 0.1 mM ATP 和 0.2 mg 每 ml 的酵母 RNA 中, 根据标准的 Southern 印迹步骤进行预杂交、杂交和杂交后洗涤最佳 12 至 24 小时。

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

在第三个方面, 本发明涉及由多核苷酸编码的分离的多肽, 所述多核苷酸包含与 SEQ ID NO: 1、SEQ ID NO: 3 或 SEQ ID NO: 5 的成熟多肽编码序列具有至少 60%、优选至少 65%、更优选至少 70%、更优选至少 75%、更优选至少 80%、更优选至少 85%、更优选至少 90%、甚至更优选至少 95%、最优选至少 97% 同一性程度的核苷酸序列或由所述核苷酸序列组成, 其编码活性多肽。在优选的方面, 所述成熟多肽编码序列是 SEQ ID NO: 1 的核苷酸 79 至 1461。在另一优选的方面, 所述成熟多肽编码序列是 SEQ ID NO: 3 的核苷酸 74 至 1349。在另一优选的方面, 所述成熟多肽编码序列是 SEQ ID NO: 5 的核苷酸 107 至 1372。参见本文多核苷酸部分。

在第四个方面, 本发明涉及人工变体, 所述人工变体涉及包含取代、缺失和/或插入一个或多个(或几个)氨基酸的 SEQ ID NO: 2、SEQ ID NO: 4 或 SEQ ID NO: 6 的成熟多肽; 或其同源序列。优选地, 氨基酸改变对性质是较不重要的(of a minor nature), 即保守的氨基酸取代或插入, 其不显著影响蛋白质的折叠和/或活性; 通常为 1 至大约 30 个氨基酸的小缺失; 小的氨基或羧基末端延伸, 例如氨基末端甲硫氨酸残基; 多至大约 20-25 个残基的小接头肽; 或通过改变净电荷或其它功能来促进纯化的小延伸, 如多组氨酸序列(poly histidine tract)、抗原表位(antigenic epitope)或结合域(binding domain)。

保守取代的实例是在以下组之内: 碱性氨基酸组(精氨酸、赖氨酸和组氨

酸)、酸性氨基酸组(谷氨酸和天冬氨酸)、极性氨基酸组(谷氨酰胺和天冬酰胺)、疏水性氨基酸组(亮氨酸、异亮氨酸和缬氨酸)、芳族氨基酸组(苯丙氨酸、色氨酸和酪氨酸)和小氨基酸组(甘氨酸、丙氨酸、丝氨酸、苏氨酸和甲硫氨酸)。通常不改变比活性(specific activity)的氨基酸取代是本领域已知的,并且由例如 H. Neurath 和 R.L. Hill, 1979, 于 *The Proteins*, Academic Press, New York 中描述。最普遍发生的交换是 Ala/Ser、Val/Ile、Asp/Glu、Thr/Ser、Ala/Gly、Ala/Thr、Ser/Asn、Ala/Val、Ser/Gly、Tyr/Phe、Ala/Pro、Lys/Arg、Asp/Asn、Leu/Ile、Leu/Val、Ala/Glu 和 Asp/Gly。

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

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

能够根据本领域已知的方法,例如定位诱变或丙氨酸分区诱变法(Cunningham 和 Wells, 1989, *Science* 244: 1081-1085)来鉴定亲本多肽中的必需氨基酸。在后一技术中,将单一丙氨酸突变引入到分子中的每个残基,并且测试所得突变分子的生物活性(即,内切葡聚糖酶活性)以鉴定对于所述分子的活性关键的氨基酸残基。同样参见 Hilton 等, 1996, *J. Biol. Chem.* 271: 4699-4708。酶的活性部位或其它的生物相互作用也能够通过结构的物理分析而测定,如通过以下这些技术:如核磁共振、晶体学、电子衍射或光亲和标记,连同推定的接触位点氨基酸的突变来测定。参见例如 de Vos 等, 1992, *Science* 255: 306-312; Smith 等, 1992, *J. Mol. Biol.* 224: 899-904; Wlodaver 等, 1992, *FEBS Lett.* 309: 59-64。必需氨基酸的同一性也能够从与多肽的同一性分析来推断,所述多肽与根据本发明的多肽相关。

能够使用已知的诱变、重组和/或改组(shuffling)方法,然后是有关的筛

选方法,例如那些由 Reidhaar-Olson 和 Sauer, 1988, *Science* 241: 53-57; Bowie 和 Sauer, 1989, *Proc. Natl. Acad. Sci. USA* 86: 2152-2156; WO 95/17413; 或 WO 95/22625 公开的那些方法来进行并测试单个或多个氨基酸取代、缺失、和/或插入。能够使用的其它方法包括易错 PCR、噬菌体展示(例如, Lowman 等, 1991, *Biochem.* 30: 10832-10837; 美国专利号 5,223,409; WO 92/06204) 和区域定向的诱变(Derbyshire 等, 1986, *Gene* 46: 145; Ner 等, 1988, *DNA* 7: 127)。

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

SEQ ID NO: 2、SEQ ID NO: 4 或 SEQ ID NO: 6 的成熟多肽例如 SEQ ID NO: 2 的氨基酸 23 至 464、SEQ ID NO: 4 的氨基酸 20 至 423、SEQ ID NO: 6 的氨基酸 19 至 440 的氨基酸取代、缺失和/或插入的总数是 10, 优选 9, 更优选 8, 更优选 7, 更优选至多 6, 更优选 5, 更优选 4, 甚至更优选 3, 最优选 2, 并且甚至最优选 1。

具有内切葡聚糖酶活性的多肽的来源

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

本发明具有内切葡聚糖酶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例如芽孢杆菌属(*Bacillus*)、链球菌属(*Streptococcus*)、链霉菌属(*Streptomyces*)、葡萄球菌属(*Staphylococcus*)、肠球菌属(*Enterococcus*)、乳酸菌属(*Lactobacillus*)、乳球菌属(*Lactococcus*)、梭菌属(*Clostridium*)、*Geobacillus* 或 *Oceanobacillus* 多肽, 所述多肽具有内切葡聚糖酶活性; 或革兰氏阴性细菌多肽, 如大肠杆菌(*E. coli*)、假单胞菌(*Pseudomonas*)、沙门氏菌(*Salmonella*)、弯曲杆菌属(*Campylobacter*)、螺杆菌(*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*)属多肽, 其

具有内切葡聚糖酶活性。

在优选方面,所述多肽是嗜碱芽孢杆菌(*Bacillus alkalophilus*)、解淀粉芽孢杆菌(*Bacillus amyloliquefaciens*)、短芽孢杆菌(*Bacillus brevis*)、环状芽孢杆菌(*Bacillus circulans*)、克劳氏芽孢杆菌(*Bacillus clausii*)、凝结芽孢杆菌(*Bacillus coagulans*)、坚强芽孢杆菌(*Bacillus firmus*)、灿烂芽孢杆菌(*Bacillus lautus*)、迟缓芽孢杆菌(*Bacillus lentus*)、地衣芽孢杆菌(*Bacillus licheniformis*)、巨大芽孢杆菌(*Bacillus megaterium*)、短小芽孢杆菌(*Bacillus pumilus*)、嗜热脂肪芽孢杆菌(*Bacillus stearothermophilus*)、枯草芽孢杆菌(*Bacillus subtilis*)或苏云金芽孢杆菌(*Bacillus thuringiensis*)具有内切葡聚糖酶活性的多肽。

在另一个优选的方面,所述多肽是具有内切葡聚糖酶活性的似马链球菌(*Streptococcus equisimilis*)、酿脓链球菌(*Streptococcus pyogenes*)、乳房链球菌(*Streptococcus uberis*)或马链球菌兽瘟亚种(*Streptococcus equi* subsp. *Zooepidemicus*)多肽。

在另一个优选的方面,所述多肽是不产色链霉菌(*Streptomyces achromogenes*)、阿维链霉菌(*Streptomyces avermitilis*)、天蓝链霉菌(*Streptomyces coelicolor*)、灰色链霉菌(*Streptomyces griseus*)或浅青紫链霉菌(*Streptomyces lividans*)多肽,其具有内切葡聚糖酶活性。

本发明具有内切葡聚糖酶活性的多肽也可以是真菌多肽,并且更优选酵母多肽如念珠菌属(*Candida*)、克鲁维酵母属(*Kluyveromyces*)、毕赤酵母属(*Pichia*)、酵母属(*Saccharomyces*)、裂殖酵母属(*Schizosaccharomyces*)或西洋著霉属(*Yarrowia*)多肽,其具有内切葡聚糖酶活性;或更优选丝状真菌多肽如枝顶孢霉属(*Acremonium*)、曲霉属(*Aspergillus*)、短梗霉属(*Aureobasidium*)、金孢子菌属(*Chrysosporium*)、隐球菌属(*Cryptococcus*)、*Filibasidium*、镰孢属(*Fusarium*)、腐质霉属(*Humicola*)、梨孢菌属(*Magnaporthe*)、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*)、瘤胃壶菌属(*Piromyces*)、裂褶菌属(*Schizophyllum*)、踝节菌属(*Talaromyces*)、热子囊菌属(*Thermoascus*)、梭孢壳属(*Thielavia*)、弯颈霉属(*Tolypocladium*)或木霉属(*Trichoderma*)多肽,其具有内切葡聚糖酶活性。

在优选的方面,所述多肽是卡尔酵母(*Saccharomyces carlsbergensis*)、酿酒酵母(*Saccharomyces cerevisiae*)、糖化酵母(*Saccharomyces diastaticus*)、道

格拉氏酵母(*Saccharomyces douglasii*)、克鲁弗酵母(*Saccharomyces kluyveri*)、诺地酵母(*Saccharomyces norbensis*)或卵形酵母(*Saccharomyces oviformis*)多肽,其具有内切葡聚糖酶活性。

在另外的优选方面,所述多肽是棘孢曲霉(*Aspergillus aculeatus*)、泡盛曲霉(*Aspergillus awamori*)、烟曲霉(*Aspergillus fumigatus*)、臭曲霉(*Aspergillus foetidus*)、日本曲霉(*Aspergillus japonicus*)、构巢曲霉(*Aspergillus nidulans*)、黑曲霉(*Aspergillus niger*)、米曲霉(*Aspergillus oryzae*)、嗜角质金孢子菌(*Chrysosporium keratinophilum*)、*Chrysosporium lucknowense*、热带金孢子菌(*Chrysosporium tropicum*)、*Chrysosporium merdarium*、*Chrysosporium inops*、毡金孢子菌(*Chrysosporium pannicola*)、*Chrysosporium queenslandicum*、*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 sarcochroum*)、拟分枝孢镰孢(*Fusarium sporotrichioides*)、硫色镰孢(*Fusarium sulphureum*)、圆镰孢(*Fusarium torulosum*)、拟丝孢镰孢(*Fusarium trichothecioides*)、镶片镰孢(*Fusarium venenatum*)、特异腐质霉(*Humicola insolens*)、疏棉状腐质霉(*Humicola lanuginosa*)、米黑毛霉(*Mucor miehei*)、嗜热毁丝霉(*Myceliophthora thermophila*)、粗糙脉孢菌(*Neurospora crassa*)、巴西青霉(*Penicillium brasilianum*)、沙门柏干酪青霉(*Penicillium camembertii*)、胶囊青霉(*Penicillium capsulatum*)、产黄青霉(*Penicillium chrysogenum*)、*Penicillium citreonigrum*、桔青霉(*Penicillium citrinum*)、棒形青霉(*Penicillium claviforme*)、顶青霉(*Penicillium corylophilum*)、皮落青霉(*Penicillium crustosum*)、指状青霉(*Penicillium digitatum*)、*Penicillium expansum*、*Penicillium funiculosum*、平滑青霉(*Penicillium glabrum*)、粒状青霉(*Penicillium granulatum*)、灰黄青霉(*Penicillium griseofulvum*)、岛青霉(*Penicillium islandicum*)、意大利青霉(*Penicillium italicum*)、微紫青霉(*Penicillium janthinellum*)、铅色青霉(*Penicillium lividum*)、大孢青霉(*Penicillium megasporum*)、梅林青霉(*Penicillium melinii*)、特异青霉(*Penicillium*

notatum)、草酸青霉(*Penicillium oxalicum*)、软毛青霉(*Penicillium puberulum*)、变紫青霉(*Penicillium purpurescens*)、产紫青霉(*Penicillium purpurogenum*)、娄地青霉(*Penicillium roquefortii*)、皱褶青霉(*Penicillium rugulosum*)、小刺青霉(*Penicillium spinulosum*)、瓦克曼青霉(*Penicillium waksmanii*)、哈茨木霉(*Trichoderma harzianum*)、康宁木霉(*Trichoderma koningii*)、长枝木霉(*Trichoderma longibrachiatum*)、里氏木霉(*Trichoderma reesei*)或绿色木霉(*Trichoderma viride*)多肽,其具有内切葡聚糖酶活性。

在另一优选的方面,所述多肽是 *Thielavia achromatica*、*Thielavia albomyces*、*Thielavia albopilosa*、*Thielavia australeinsis*、*Thielavia fimeti*、小孢梭孢壳(*Thielavia microspora*)、*Thielavia ovispora*、*Thielavia peruviana*、瘤孢梭孢壳(*Thielavia spededonium*)、毛梭孢壳(*Thielavia setosa*)、*Thielavia subthermophila*、土生梭孢霉(*Thielavia terrestris*)、栖土梭孢壳(*Thielavia terricola*)、嗜热梭孢壳(*Thielavia thermophila*)、*Thielavia variospora* 或 *Thielavia wareingii* 多肽,其具有内切葡聚糖酶活性。

在更优选的方面,所述多肽是土生梭孢霉多肽,并且最优选是土生梭孢霉 NRRL 8126 多肽,例如 SEQ ID NO: 2 或 SEQ ID NO: 4 的多肽或其成熟多肽。

在另一更优选的方面,所述多肽是 *Cladorrhinum foecundissimum* 多肽,最优选是 *Cladorrhinum foecundissimum* ATCC 62373 多肽,例如,SEQ ID NO: 6 的多肽或其成熟多肽。

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

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) (DSM)、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)。

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

本发明的多肽还包括融合多肽或可切割的融合多肽,其中将另外的多肽融合到所述多肽或其片段的 N 末端或 C 末端。通过将编码另一个多肽的核苷酸序列(或其部分)融合于本发明的核苷酸序列(或其部分)来产生融合的多肽。产生融合多肽的技术是本领域已知的,并包括连接编码多肽的编码序列以使它们在阅读框中,并且使融合多肽的表达在相同启动子和终止子的控制下。

融合多肽还可以包括切割位点。一旦分泌了融合多肽,就切割所述位点,从融合蛋白质释放具有内切葡聚糖酶活性的多肽。

切割位点的实例包括,但不限于,编码二肽 Lys-Arg 的 Kex2 位点(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); Ile-(Glu 或 Asp)-Gly-Arg 位点,其在精氨酸残基后通过 Factor Xa 蛋白酶切割(Eaton 等, 1986, *Biochem.* 25: 505-512); Asp-Asp-Asp-Asp-Lys 位点,其在赖氨酸后通过肠激酶切割(Collins-Racie 等, 1995, *Biotechnology* 13: 982-987); His-Tyr-Glu 位点或 His-Tyr-Asp 位点,其通过 Genenase I 切割(Carter 等, 1989, *Proteins: Structure, Function, and Genetics* 6: 240-248); Leu-Val-Pro-Arg-Gly-Ser 位点,其在 Arg 后通过凝血酶切割(Stevens, 2003, *Drug Discovery World* 4: 35-48); Glu-Asn-Leu-Tyr-Phe-Gln-Gly 位点,其在 Gln 后通过 TEV 蛋白酶切割(Stevens, 2003, 见上文); 和 Leu-Glu-Val-Leu-Phe-Gln-Gly-Pro 位点,其在 Gln 后通过基因工程形式的人鼻病毒 3C 蛋白酶切割(Stevens, 2003, 见上文)。

多核苷酸

本发明还涉及分离的多核苷酸,其包含编码本发明具有内切葡聚糖酶活性的多肽的核苷酸序列,或由编码本发明具有内切葡聚糖酶活性的多肽的核

核苷酸序列组成。

在优选的方面,核苷酸序列包含 SEQ ID NO: 1 或由 SEQ ID NO: 1 组成。在另一更优选的方面,核苷酸序列包含大肠杆菌 NRRL B-30899 中所含质粒 pPH50 中的序列,或由该序列其组成。在另一优选方面,核苷酸序列包含 SEQ ID NO: 1 的成熟多肽编码区或由 SEQ ID NO: 1 的成熟多肽编码区组成。在另一优选方面,核苷酸序列包含 SEQ ID NO: 1 的核苷酸 79 至 1461 或由 SEQ ID NO: 1 的核苷酸 79 至 1461 组成。在另一更优选的方面,核苷酸序列包含大肠杆菌 NRRL B-30899 中所含质粒 pPH50 中的成熟多肽编码区或由大肠杆菌 NRRL B-30899 中所含质粒 pPH50 中的成熟多肽编码区组成。本发明还包含编码多肽的核苷酸序列,所述多肽包含 SEQ ID NO: 2 的氨基酸序列或其成熟多肽,或由 SEQ ID NO: 2 的氨基酸序列或其成熟多肽组成;由于遗传密码的简并性,所述核苷酸序列不同于 SEQ ID NO: 1 或其成熟多肽编码序列。本发明还涉及 SEQ ID NO: 1 的亚序列,所述亚序列编码具有内切葡聚糖酶活性的 SEQ ID NO: 2 的片段。

在另一优选的方面,核苷酸序列包含 SEQ ID NO: 3 或由 SEQ ID NO: 3 组成。在另一更优选的方面,核苷酸序列包含大肠杆菌 NRRL B-30896 中所含质粒 pPH38 中的序列,或由该序列其组成。在另一优选的方面,核苷酸序列包含 SEQ ID NO: 3 的成熟多肽编码区或由 SEQ ID NO: 3 的成熟多肽编码区组成。在另一的优选方面,核苷酸序列包含 SEQ ID NO: 3 的核苷酸 74 至 1349 或由 SEQ ID NO: 3 的核苷酸 74 至 1349 组成。在另一的更优选的方面,核苷酸序列包含大肠杆菌 NRRL B-30896 中所含质粒 pPH38 中的成熟多肽编码区或由大肠杆菌 NRRL B-30896 中所含质粒 pPH38 中的成熟多肽编码区组成。本发明还包含编码多肽的核苷酸序列,所述多肽包含 SEQ ID NO: 4 的氨基酸序列或其成熟多肽,或由 SEQ ID NO: 4 的氨基酸序列或其成熟多肽组成;由于遗传密码的简并性,所述核苷酸序列不同于 SEQ ID NO: 3 或其成熟多肽编码序列。本发明还涉及 SEQ ID NO: 3 的亚序列,所述亚序列编码具有内切葡聚糖酶活性的 SEQ ID NO: 4 的片段。

在另一优选的方面,核苷酸序列包含 SEQ ID NO: 5 或由 SEQ ID NO: 5 组成。在另一更优选的方面,核苷酸序列包含大肠杆菌 NRRL B-30897 中所含质粒 pPH46 中的序列,或由该序列其组成。在另一优选方面,核苷酸序列包含 SEQ ID NO: 5 的成熟多肽编码区或由 SEQ ID NO: 5 的成熟多肽编码区

组成。在另一优选方面,核苷酸序列包含 SEQ ID NO: 5 的核苷酸 107 至 1372 或由 SEQ ID NO: 5 的核苷酸 107 至 1372 组成。在另一更优选的方面,核苷酸序列包含大肠杆菌 NRRL B-30897 中所含质粒 pPH46 中的成熟多肽编码区或由大肠杆菌 NRRL B-30897 中所含质粒 pPH46 中的成熟多肽编码区组成。本发明还包含编码下述多肽的核苷酸序列,所述多肽包含 SEQ ID NO: 6 的氨基酸序列或其成熟多肽,或由 SEQ ID NO: 6 的氨基酸序列或其成熟多肽组成;由于遗传密码的简并性,所述核苷酸序列不同于 SEQ ID NO: 5 或其成熟多肽编码序列。本发明还涉及 SEQ ID NO: 5 的亚序列,所述亚序列编码具有内切葡聚糖酶活性的 SEQ ID NO: 6 的片段。

本发明还涉及突变多核苷酸,所述突变多核苷酸在 SEQ ID NO:1、SEQ ID NO:3 或 SEQ ID NO:5 的成熟多肽编码序列中包含至少一个突变或由具有至少一个突变的 SEQ ID NO:1、SEQ ID NO:3 或 SEQ ID NO:5 的成熟多肽编码序列组成,其中所述突变核苷酸序列编码 SEQ ID NO: 2、SEQ ID NO:4 或 SEQ ID NO:6 的成熟多肽。在优选的方面,所述成熟多肽是 SEQ ID NO: 2 的氨基酸 23 至 464。在另一优选的方面,所述成熟多肽是 SEQ ID NO: 4 的氨基酸 20 至 423。在另一优选的方面,所述成熟多肽是 SEQ ID NO: 6 的氨基酸 19 至 440。

用于分离或克隆编码多肽的多核苷酸的技术是本领域内已知的,包括从基因组 DNA 分离,从 cDNA 制备,或其组合。可通过例如使用熟知的聚合酶链式反应(PCR)或表达文库的抗体筛选来检测具有共有结构特性的克隆 DNA 片段,从而实现从这种基因组 DNA 克隆本发明的多核苷酸。参见,例如,Innis 等,1990, *PCR: A Guide to Methods and Application*, Academic Press, New York。可以使用其它核酸扩增方法,如连接酶链式反应(LCR)、连接活化转录(ligated activated transcription; LAT)和基于核苷酸序列的扩增(NASBA)。可以从嗜热毁丝霉 CBS 117.65、担子菌 CBS 494.95 或担子菌 CBS 495.95 菌株,或其它或相关生物体克隆多核苷酸,并且因此可以是例如所述核苷酸序列的多肽编码区的等位基因变体或种变体(species variant)。

本发明还涉及分离的多核苷酸,其包含下述核苷酸序列或由其组成,所述核苷酸序列与 SEQ ID NO: 1、SEQ ID NO: 3 或 SEQ ID NO: 5 的成熟多肽编码序列具有至少 60%,优选至少 65%,更优选至少 70%,更优选至少 75%,更优选至少 80%,更优选至少 85%,更优选至少 90%,甚至更优选至少 95%,

并且最优选至少 97%同一性的同一性程度,所述多核苷酸编码活性多肽。在优选的方面,所述成熟多肽编码序列是 SEQ ID NO: 1 的核苷酸 79 至 1461。在另一优选的方面,所述成熟多肽编码序列是 SEQ ID NO: 3 的核苷酸 74 至 1349。在另一优选的方面,所述成熟多肽编码序列是 SEQ ID NO: 5 的核苷酸 107 至 1372。

修饰编码本发明多肽的核苷酸序列对于合成与所述多肽基本上相似的多肽可为必需的。术语与所述多肽“基本上相似”指多肽的非天然存在的形式。这些多肽可能以一些工程改造的方式而不同于从其天然来源分离的多肽,例如,比活性、热稳定性、最适 pH 等方面不同的人工变体。可以在作为 SEQ ID NO:1、SEQ ID NO: 3 或 SEQ ID NO: 5 的多肽编码区存在的核苷酸序列,例如其亚序列的基础上,和/或通过引入如下核苷酸取代来构建变体序列:所述取代不产生由核苷酸序列编码的多肽的另外的氨基酸序列,但是符合意欲产生酶的宿主生物体的密码子选择;或者所述取代可产生不同的氨基酸序列。关于核苷酸取代的概述,参见,例如, Ford 等, 1991, *Protein Expression and Purification* 2: 95-107。

对于本领域技术人员显而易见的是,这些取代能够在对于分子功能重要的区域之外进行,并且仍然产生活性多肽。对于由本发明的分离的多核苷酸编码的多肽活性关键的并且因此优选不进行取代的氨基酸残基,可以根据本领域公知的方法,例如定位诱变或丙氨酸分区诱变法(参见,例如, Cunningham 和 Wells, 1989, 见上文)来鉴定。在另一技术中,将突变引入到分子中的每个荷正电的残基处,并且测试所得突变分子的内切葡聚糖酶活性,以鉴定对于所述分子的活性关键的氨基酸残基。底物-酶相互作用的位点也能够通过分析三维结构测定,通过如核磁共振分析、晶体学或光亲和标记这样的技术来测定(参见,例如, de Vos 等, 1992, 见上文; Smith 等, 1992, 见上文; Wlodaver 等, 1992, 见上文)。

本发明还涉及编码本发明多肽的分离的多核苷酸,所述分离的多核苷酸在非常低严紧条件下,优选低严紧条件,更优选中等严紧条件,更优选中高严紧条件,甚至更优选高严紧条件,并且最优选非常高的严紧条件下,与以下序列杂交: (i) SEQ ID NO: 1、SEQ ID NO: 3 或 SEQ ID NO: 5 的成熟多肽编码序列, (ii) 包含于 SEQ ID NO: 1 或 SEQ ID NO: 5 的成熟多肽编码序列中的 cDNA 序列或者包含 SEQ ID NO: 3 的成熟多肽编码序列的基因组

DNA 序列, 或(iii) (i)或(ii)的互补链; 或它们的等位变体和亚序列(Sambrook 等, 1989, 同上), 如本文所定义的。在优选的方面, 所述成熟多肽编码序列是 SEQ ID NO: 1 的核苷酸 79 至 1461。在另一优选的方面, 所述成熟多肽编码序列是 SEQ ID NO: 3 的核苷酸 74 至 1349。在另一优选的方面, 所述成熟多肽编码序列是 SEQ ID NO: 5 的核苷酸 107 至 1372。在另一优选的方面, 所述互补链是 SEQ ID NO: 1、SEQ ID NO: 3 或 SEQ ID NO: 5 的成熟多肽编码序列的全长互补链。

本发明还涉及分离的多核苷酸, 所述分离的多核苷酸通过以下方法获得: (a)在非常低、低、中、中-高、高或非常高严谨条件下, 将 DNA 的群体与以下杂交: (i) SEQ ID NO: 1、SEQ ID NO: 3 或 SEQ ID NO: 5 的成熟多肽编码序列, (ii)包含于 SEQ ID NO: 1 或 SEQ ID NO: 5 的成熟多肽编码序列中的 cDNA 序列或者包含 SEQ ID NO: 3 的成熟多肽编码序列的基因组 DNA 序列, 或(iii) (i)或(ii)的互补链; 和(b)分离杂交的多核苷酸, 其编码具有内切葡聚糖酶活性的多肽。在优选的方面, 所述成熟多肽编码序列是 SEQ ID NO: 1 的核苷酸 79 至 1461。在另一优选的方面, 所述成熟多肽编码序列是 SEQ ID NO: 3 的核苷酸 74 至 1349。在另一优选的方面, 所述成熟多肽编码序列是 SEQ ID NO: 5 的核苷酸 107 至 1372。在另一优选的方面, 所述互补链是 SEQ ID NO: 1、SEQ ID NO: 3 或 SEQ ID NO: 5 的成熟多肽编码序列的全长互补链。

核酸构建体

本发明还涉及包含本发明的分离的多核苷酸的核酸构建体, 所述分离的多核苷酸与一个或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。

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

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

用于指导本发明的核酸构建体转录,特别是在细菌宿主细胞中转录的合适启动子的实例是从下述获得的启动子:大肠杆菌 *lac* 操纵子、天蓝色链霉菌(*Streptomyces coelicolor*)琼脂糖酶基因(*dagA*)、枯草芽孢杆菌果聚糖蔗糖酶基因(*sacB*)、地衣芽孢杆菌 α -淀粉酶基因(*amyL*)、嗜热脂肪芽孢杆菌产麦芽淀粉酶基因(*amyM*)、解淀粉芽孢杆菌 α -淀粉酶基因(*amyQ*)、地衣芽孢杆菌青霉素酶基因(*penP*)、枯草芽孢杆菌 *xylA* 和 *xylB* 基因和原核 β -内酰胺酶基因(Villa-Kamaroff 等, 1978, *Proceedings of the National Academy of Sciences USA* 75: 3727-3731), 以及 *tac* 启动子(DeBoer 等, 1983, *Proceedings of the National Academy of Sciences USA* 80: 21-25)。另外的启动子在"Useful proteins from recombinant bacteria"于 *Scientific American*, 1980, 242: 74-94 中; 和在 Sambrook 等, 1989, 见上文中描述。

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

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

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

连接。可以将在所选宿主细胞中有功能的任何终止子用在本发明中。

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

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。

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

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。

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

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

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

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

调控序列还可以是信号肽编码区，其编码与多肽的氨基末端相连的氨基酸序列，并且指导编码的多肽进入细胞分泌途径。核苷酸序列的编码序列 5'端可固有地包含信号肽编码区，其与编码分泌多肽的编码区片段一起天然地连接在翻译阅读框中。可供选择的是，编码序列 5'端可含有对于所述编码序列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，外源信号肽编码区可以简单地取代天然信号肽编码区以增强多肽的分泌。然而，指导表达的多肽进入所选宿主细胞的分泌

途径(即,分泌至培养基中)的任何信号肽编码区可在本发明中使用。

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

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

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

在优选的方面,信号肽是 SEQ ID NO: 2 的氨基酸 1 至 22。在另一优选的方面,信号肽编码区是 SEQ ID NO: 1 的核苷酸 13 至 79。

在另一优选的方面,信号肽是 SEQ ID NO: 4 的氨基酸 1 至 19。在另一优选的方面,信号肽编码区是 SEQ ID NO: 3 的核苷酸 17 至 73。

在另一优选的方面,信号肽是 SEQ ID NO: 6 的氨基酸 1 至 18。在另一优选的方面,信号肽编码区是 SEQ ID NO: 5 的核苷酸 53 至 106。

调控序列还可以是前肽编码区,其编码位于多肽氨基末端的氨基酸序列。所得多肽称为酶原(proenzyme)或前多肽(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成熟活性多肽。可以从枯草芽孢杆菌碱性蛋白酶(*aprE*)、枯草芽孢杆菌中性蛋白酶(*nprT*)、酿酒酵母 α 因子、曼赫根毛霉天冬氨酸蛋白酶和嗜热毁丝霉漆酶(WO 95/33836)的基因获得前肽编码区。

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

同样理想的是添加调节序列,其允许相对于宿主细胞的生长来调节多肽的表达。调节系统的实例是引起基因表达响应化学或物理刺激物,包括调节化合物的存在而开启或关闭的那些系统。原核系统中的调节系统包括 *lac*、*tac* 和 *trp* 操纵基因系统。在酵母中,可以使用 ADH2 系统或 GAL1 系统。在丝状真菌中,可以使用 TAKA α -淀粉酶启动子、黑曲霉葡糖淀粉酶启动子和米

曲霉葡萄糖淀粉酶启动子作为调节序列。调节序列的其它实例是那些允许基因扩增的序列。在真核系统中, 这些序列包括在氨甲蝶呤(methotrexate)存在下扩增的二氢叶酸还原酶基因, 和以重金属(with heavy metal)扩增的金属硫蛋白基因。在这些情况下, 编码多肽的核苷酸序列将与调节序列可操作地连接。

表达载体

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

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

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)。

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

细菌选择性标记的实例是来自枯草芽孢杆菌或地衣芽孢杆菌的 *dal* 基因, 或赋予抗生素抗性的标记, 所述抗生素抗性例如氨苄青霉素、卡那霉素、氯霉素或四环素抗性。对于酵母宿主细胞合适的标记是 ADE2、HIS3、LEU2、

LYS2、MET3、TRP1 和 URA3。用于丝状真菌宿主细胞的选择性标记包括但不限于 *amdS* (乙酰胺酶)、*argB* (鸟氨酸氨甲酰基转移酶)、*bar* (草铵膦(phosphinothricin)乙酰转移酶)、*hph* (潮霉素磷酸转移酶)、*niaD* (硝酸还原酶)(nitrate reductase)、*pyrG* (乳清酸核苷-5'-磷酸脱羧酶)(orotidine-5'-phosphate decarboxylase)、*sC* (硫酸腺苷酰转移酶)和 *trpC* (邻氨基苯甲酸合酶(anthranilate synthase))以及它们的等价物。优选用在曲霉属细胞中的是构巢曲霉或米曲霉的 *amdS* 和 *pyrG* 基因和吸水链霉菌(*Streptomyces hygroscopicus*)的 *bar* 基因。

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

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

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

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 pAMB1 的复制起点。

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

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 Research* 15: 9163-9175; WO 00/24883)。分离 AMA1 基因和构建包含该基因的质粒或载体能够根据公

开于 WO 00/24883 中的方法完成。

可以将多于一个拷贝的本发明的多核苷酸插入宿主细胞以增加基因产物的产生。多核苷酸拷贝数的增加可通过如下方法获得：将至少一个额外拷贝的序列整合入宿主细胞基因组，或将可扩增的选择性标记基因包括于多核苷酸，其中可通过在合适的选择剂(selectable agent)存在下培养细胞来选择含有选择性标记基因的扩增拷贝的细胞，和由此得到多核苷酸的额外拷贝。

用于连接上述元件以构建本发明的重组表达载体的方法是本领域技术人员熟知的(参见，例如，Sambrook 等, 1989, 见上文)。

宿主细胞

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

原核宿主细胞可以是任何革兰氏阳性细菌或革兰氏阴性细菌。革兰氏阳性细菌包括但不限于，芽孢杆菌属、链球菌属、链霉菌属、葡萄球菌属、肠球菌属、乳酸菌属、乳球菌属、梭菌属、*Geobacillus* 和 *Oceanobacillus*。革兰氏阴性细菌包括但不限于，大肠杆菌、假单胞菌属、沙门氏菌属、弯曲杆菌属、螺杆菌属、黄杆菌属、梭杆菌属、泥杆菌属、奈瑟氏菌属和脲原体属。

细菌宿主细胞可以是任何芽孢杆菌属细胞。在本发明的实施中有用的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。

在优选的方面，细菌宿主细胞是解淀粉芽孢杆菌、迟缓芽孢杆菌、地衣芽孢杆菌、嗜热脂肪芽孢杆菌或枯草芽孢杆菌细胞。在更优选的方面，细菌宿主细胞是解淀粉芽孢杆菌细胞。在另一个更优选的方面，细菌宿主细胞是克劳氏芽孢杆菌细胞。在另一个更优选的方面，细菌宿主细胞是地衣芽孢杆菌细胞。在另一个更优选的方面，细菌宿主细胞是枯草芽孢杆菌细胞。

细菌宿主细胞还可以是任何链球菌细胞。在本发明的实施中有用的链球菌细胞包括但不限于，似马链球菌、酿脓链球菌、乳房链球菌和马链球菌兽瘟亚种。

在优选的方面，细菌宿主细胞是似马链球菌细胞。在另一个优选的方面，

细菌宿主细胞是酿脓链球菌细胞。在另一个优选的方面，细菌宿主细胞是乳房链球菌细胞。在另一个优选的方面，细菌宿主细胞是马链球菌兽瘟亚种细胞。

细菌宿主细胞还可以是任何链霉菌细胞。在本发明的实施中有用的链霉菌细胞包括但不限于，不产色链霉菌、阿维链霉菌、天蓝链霉菌、灰色链霉菌和浅青紫链霉菌。

在优选的方面，细菌宿主细胞是不产色链霉菌细胞。在另一个优选的方面，细菌宿主细胞是阿维链霉菌细胞。在另一个优选的方面，细菌宿主细胞是天蓝链霉菌细胞。在另一个优选的方面，细菌宿主细胞是灰色链霉菌细胞。在另一个优选的方面，细菌宿主细胞是浅青紫链霉菌细胞。

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, *Molecular General Genetics* 168: 111-115)，使用感受态细胞(参见，例如，Young 和 Spizizen, 1961, *Journal of Bacteriology* 81: 823-829 或 Dubnau 和 Davidoff-Abelson, 1971, *Journal of Molecular Biology* 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, *Journal of Bacteriology* 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: 189-2070)，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，可以使用任何已知的将

DNA 引入宿主细胞的方法。

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。

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

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

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

在最优选的方面，酵母宿主细胞是卡尔酵母、酿酒酵母、糖化酵母、道格拉氏酵母、克鲁弗酵母、诺地酵母或卵形酵母细胞。在另一个最优选的方面，酵母宿主细胞是乳酸克鲁维酵母(*Kluyveromyces lactis*)细胞。在另一个最优选的方面，酵母宿主细胞是 *Yarrowia lipolytica* 细胞。

在另一个更优选的方面，真菌宿主细胞是丝状真菌细胞。“丝状真菌”包括真菌门(Eumycota)和卵菌门的亚门(如由 Hawksworth 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。

在甚至更优选的方面，丝状真菌宿主细胞是枝顶孢霉属、曲霉属、短梗

霉属、烟管霉属(*Bjerkandera*)、*Ceriporiopsis*、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。

在最优选的方面，丝状真菌宿主细胞是泡盛曲霉、烟曲霉、臭曲霉、日本曲霉、构巢曲霉、黑曲霉或米曲霉细胞。在另一个最优选方面，丝状真菌宿主细胞是杆孢状镰孢、禾谷镰孢、库威镰孢、大刀镰孢、禾本科镰孢、禾赤镰孢、异孢镰孢、合欢木镰孢、尖镰孢、多枝镰孢、粉红镰孢、接骨木镰孢、肤色镰孢、拟分枝孢镰孢、疏色镰孢、圆镰孢、拟丝孢镰孢或镶片镰孢细胞。在另一个最优选的方面，丝状真菌宿主细胞是黑刺烟管菌(*Bjerkandera adusta*)、*Ceriporiopsis aneirina*、*Ceriporiopsis aneirina*、*Ceriporiopsis caregiea*、*Ceriporiopsis gilvescens*、*Ceriporiopsis pannocinta*、*Ceriporiopsis rivulosa*、*Ceriporiopsis subrufa*、*Ceriporiopsis subvermispora*、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*、灰盖鬼伞(*Coprinus cinereus*)、毛革盖菌(*Coriolus hirsutus*)、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、巴西青霉、产紫青霉、黄孢平革菌(*Phanerochaete chrysosporium*)、辐射射脉菌(*Phlebia radiata*)、*Pleurotus eryngii*、土生梭孢霉、*Trametes villosa*、*Trametes versicolor*、哈茨木霉、康宁木霉、长枝木霉、里氏木霉或绿色木霉细胞。

可以将真菌细胞通过涉及原生质体形成、原生质体转化和细胞壁重建的方法以本身公知的方式转化。用于转化曲霉属和木霉属宿主细胞的合适方法在 EP 238 023 和 Yelton 等, 1984, *Proceedings of the National Academy of Sciences 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.(编者), *Guide to Yeast Genetics and Molecular Biology, Methods in Enzymology*, Volume 194, pp 182-187, Academic Press, Inc., New York; Ito 等,

1983, *Journal of Bacteriology* 153: 163; 和 Hinnen 等, 1978, *Proceedings of the National Academy of Sciences USA* 75: 1920。

产生方法

本发明还涉及用于产生本发明多肽的方法, 其包括: (a)在有益于产生多肽的条件下培养细胞, 所述细胞以其野生型形式产生所述多肽; 和(b)回收所述多肽。在优选的方面, 所述细胞是梭孢壳属的细胞。在另一优选的方面, 所述细胞是枝孢菌属(*Cladorrhinum*)属细胞。在更优选的方面, 所述细胞是土生梭孢霉。在另一更优选的方面, 所述细胞是 *Cladorrhinum foecundissimum*。在最优选的方面, 所述细胞是土生梭孢霉 NRRL 8126。在另一最优选的方面, 所述细胞是 *Cladorrhinum foecundissimum* ATCC 62373。

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

本发明还涉及用于产生本发明的多肽的方法, 包括: (a)在有益于产生多肽的条件下培养宿主细胞, 其中所述宿主细胞包含突变核苷酸序列, 其在 SEQ ID NO:1、SEQ ID NO:3 或 SEQ ID NO:5 的成熟多肽编码序列中具有至少一个突变, 其中所述突变核苷酸序列编码多肽, 所述多肽包含 SEQ ID NO: 2、SEQ ID NO:4 或 SEQ ID NO:6 的成熟多肽或由该成熟多肽组成, 和(b)回收所述多肽。

在优选的方面, 所述成熟多肽是 SEQ ID NO: 2 的氨基酸 23 至 464。在另一优选的方面, 所述成熟多肽是 SEQ ID NO: 4 的氨基酸 20 至 423。在另一优选的方面, 所述成熟多肽是 SEQ ID NO: 6 的氨基酸 19 至 440。在另一优选的方面, 所述成熟多肽编码序列是 SEQ ID NO: 1 的核苷酸 79 至 1461。在另一优选的方面, 所述成熟多肽编码序列是 SEQ ID NO: 3 的核苷酸 74 至 1349。在另一优选的方面, 所述成熟多肽编码序列是 SEQ ID NO: 5 的核苷酸 107 至 1372。

在本发明的产生方法中, 使用本领域熟知的方法在适合于产生所述多肽的营养培养基中培养细胞。例如, 可以通过在合适培养基中和允许表达和/或分离所述多肽的条件下进行的摇瓶培养, 和实验室或工业发酵罐中的小规模或大规模发酵(包括连续、分批、补料分批或固态发酵)来培养细胞。使用本领域已知的方法在合适的营养培养基中进行培养, 所述营养培养基包含碳

源和氮源和无机盐。合适的培养基能够从商业供应商获得或可以根据公开的组成制备(例如,在美国典型培养物保藏中心的目录中)。如果多肽分泌到营养培养基中,该多肽能够从所述培养基中直接回收。如果多肽不分泌到培养基中,其能够从细胞裂解物(lysate)回收。

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

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

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

植物

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

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)。

双子叶植物的实例是烟草(tobacco),豆类(legumes),如羽扇豆(lupins),马铃薯,糖甜菜(sugar beet),豌豆,豆(bean)和大豆(soybean)和十字花科的(cruciferous)植物(十字花科(family Brassicaceae)),如花椰菜(cauliflower),油

菜籽(rape seed)和紧密相关的模型生物体拟南芥(*Arabidopsis thaliana*)。

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)。

同样包含于本发明范围内的还有这些植物、植物部分和植物细胞的子代。

表达本发明多肽的转基因植物或植物细胞可以依照本领域已知方法构建。简而言之, 通过如下方法构建所述植物或植物细胞: 将编码本发明多肽的一个或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。

表达载体便利地是包含编码本发明多肽的多核苷酸的核酸构建体, 所述多核苷酸与在选择的植物或植物部分中表达该多核苷酸序列所需的适当的调节序列可操作地连接。此外, 表达构建体可以包含对于鉴定宿主细胞有用的选择性标记, 在所述宿主细胞中整合了表达构建体并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)。

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

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 & 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 and Cell Physiology* 39: 885-889), 来自豆球蛋白(legumin) B4 和蚕豆(*Vicia faba*)的未知的种子蛋白基因的蚕豆启动子(Conrad 等, 1998, *Journal of Plant Physiology* 152:708-711)、来自种子油体蛋白(oil body protein)的启动子(Chen 等, 1998, *Plant and Cell Physiology* 39: 935-941), 来自欧洲油菜(*Brassica napus*)的贮藏蛋白 napA 启动子, 或本技术领域公知的任何其他种子特异性的启动子, 例如, 在 WO 91/14772 中所描述的。此外, 启动子可为叶特异性的启动子, 如来自稻或番茄的 *rbcS* 启动子(Kyozuka 等, 1993, *Plant Physiology* 102: 991-1000), 小球藻病毒(chlorella virus)腺嘌呤甲基转移酶(adenine methyltransferase)基因启动子(Mitra 和 Higgins, 1994, *Plant Molecular Biology* 26: 85-93), 或来自稻的 *aldP* 基因启动子(Kagaya 等, 1995, *Molecular and general genetics* 248: 668-674), 或伤口诱导的启动子, 如马铃薯 *pin2* 启动子(Xu 等, 1993, *Plant Molecular Biology* 22: 573-588)。同样地, 所述启动子可通过非生物的处理诱导, 所述非生物的处理诸如温度、干旱或盐度变化, 或通过外源施加的激活所述启动子的物质诱导, 例如乙醇、雌激素(oestrogens)、植物激素(plant hormones)如乙烯、脱落酸(abscisic acid)和赤霉酸(gibberellic acid), 和重金属。

启动子增强子元件也可以用于实现本发明多肽在植物中的较高表达。例如, 启动子增强子元件可以是内含子, 其置于启动子和编码本发明多肽的核苷酸序列之间。例如 Xu 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。

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。

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)。

目前, 根癌土壤杆菌(*Agrobacterium tumefaciens*)介导的基因转移(gene

transfer), 是产生转基因双子叶植物的优选方法(为了参考, 见 Hooykas 和 Schilperoort, 1992, *Plant Molecular Biology* 19: 15-38), 而且它也可以用于转化单子叶植物, 虽然对于这些植物其他的转化方法是常用的。目前, 产生转基因单子叶植物的优选的方法, 是用粒子(用转化 DNA 涂覆的微观的金或钨粒子)轰击胚愈伤组织(embryonic calli)或发育中的胚(developing embryos) (Christou, 1992, *Plant Journal* 2: 275-281; Shimamoto, 1994, *Current Opinion Biotechnology* 5: 158-162; Vasil 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 Molecular Biology* 21: 415-428 所描述的。

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

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

去除或减少内切葡聚糖酶活性

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

可以使用本领域熟知的方法通过减少或消除编码本发明多肽的核苷酸序列的表达, 所述方法例如, 插入、破坏、替代或缺失来构建突变细胞。在优选的方面, 所述核苷酸序列是失活的。待修饰或失活的核苷酸序列可以是, 例如, 编码区或其对活性关键的部分, 或表达编码区所需的调节元件。这种调节或调控序列的实例可以是启动子序列或其功能部分, 即, 足以影响核苷酸序列表达的部分。用于可能的修饰的其它调控序列包括但不限于前导序列、聚腺苷酸化序列、前肽序列、信号肽序列、转录终止子和转录激活子。

可以通过向亲本细胞施以诱变, 并且选择其中已将所述核苷酸序列的表达减少或消除的突变细胞来进行核苷酸序列的修饰或失活。诱变可能是特异性的或随机的, 可以通过例如使用合适的物理或化学诱变剂进行, 通过使用

合适的寡核苷酸进行,或通过将所述DNA序列进行PCR产生的诱变。此外,可以通过使用这些诱变剂的任何组合来进行诱变。

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

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

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

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

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

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

这样生成的多肽缺陷型突变细胞作为表达同源和/或异源多肽的宿主细

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

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

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

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

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

用于培养和纯化感兴趣的产物的方法可以通过本领域已知的方法进行。

用于产生基本上无内切葡聚糖酶产物的本发明的方法在真核多肽,特别是真菌蛋白质例如酶的产生中是特别令人有兴趣的。所述酶可以选自,例如,淀粉分解酶(amylytic enzyme)、脂肪分解酶、蛋白水解酶、纤维素分解酶(cellulolytic enzyme)、氧化还原酶或植物细胞壁降解酶。这些酶的实例包括氨肽酶、淀粉酶、淀粉葡萄糖苷酶、糖酶、羧肽酶、过氧化氢酶、纤维素酶、几丁质酶(chitinase)、角质酶(cutinase)、环糊精糖基转移酶、脱氧核糖核酸酶、酯酶、半乳糖苷酶、 β -半乳糖苷酶、葡萄糖淀粉酶、葡萄糖氧化酶、葡萄糖苷酶、卤素过氧化酶、半纤维素酶、转化酶、异构酶、漆酶、连接酶、脂肪酶、裂合酶、甘露糖苷酶、氧化酶、果胶分解酶、过氧化物酶、肌醇六磷酸酶、酚

氧化酶、多酚氧化酶、蛋白水解酶、核糖核酸酶、转移酶、转谷氨酰胺酶或木聚糖酶。内切葡聚糖酶缺陷细胞也可以用于表达在制药上引起兴趣的异源蛋白质例如激素、生长因子、受体等。

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

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

抑制多肽表达的方法

本发明还涉及抑制多肽在细胞中表达的方法，其包括对细胞施用双链RNA (dsRNA)分子或者在细胞中表达双链RNA (dsRNA)分子，其中dsRNA包含本发明多核苷酸的亚序列或部分。在优选的方面，dsRNA长约15、16、17、18、19、20、21、22、23、24、25或更多的双链体核苷酸。在另一个优选的方面，多肽具有内切葡聚糖活性。

dsRNA优选是小干扰RNA (siRNA)或微RNA (miRNA)。在优选的方面，dsRNA是用于抑制转录的小干扰RNA (siRNA)。在另一个优选的方面，dsRNA是用于抑制翻译的微RNA (miRNA)。

本发明还涉及用于抑制细胞中多肽表达的这些双链RNA (dsRNA)分子，其中dsRNA包含编码SEQ ID NO: 2、SEQ ID NO: 4或SEQ ID NO: 6的成熟多肽的多核苷酸的亚序列或部分。本发明不受到任何特定作用机制的限制，dsRNA能进入细胞，并引起相似或相同序列的单链RNA (ssRNA)的降解，所述RNA包括内源mRNA。当细胞接触dsRNA时，来自同源基因的mRNA选择性地受到称为RNA干扰(RNAi)的过程的降解。

本发明的dsRNA可以用于基因沉默疗法。在一个方面，本发明提供使用本发明的dsRNAi选择性地降解RNA的方法。所述过程可以在体外、在活体外(*ex vivo*)或在体内进行。在一个方面，dsRNA分子能够用于产生细胞、器官或动物中的功能缺失突变(loss-of-function mutation)。制备或使用选择性降解RNA的dsRNA分子的方法在本领域中公知，参见，例如，美国专利号6,506,559；美国专利号6,511,824；美国专利号6,515,109；和美国专利号6,489,127。

组合物

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

所述组合物可以包含本发明的多肽作为主要酶成分，例如，单成分组合物。或者，所述组合物可以包含多种酶活性，例如氨肽酶、淀粉酶、糖酶、羧肽酶、过氧化氢酶、纤维素酶、几丁质酶、角质酶、环糊精糖基转移酶、脱氧核糖核酸酶、酯酶、 α -半乳糖苷酶、 β -半乳糖苷酶、葡糖淀粉酶、 α -葡糖苷酶、 β -葡糖苷酶、卤素过氧化物酶、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(peptidoglutaminase)、过氧化物酶、肌醇六磷酸酶、多酚氧化酶、蛋白水解酶、核糖核酸酶、转谷氨酰胺酶或木聚糖酶。额外的酶可以通过例如属于以下属或种的微生物产生：曲霉属，优选棘孢曲霉、泡盛曲霉、烟曲霉、臭曲霉、日本曲霉、构巢曲霉、黑曲霉或米曲霉；镰孢属，优选杆孢状镰孢、禾谷镰孢、库威镰孢、大刀镰孢、禾本科镰孢、禾赤镰孢、异孢镰孢、合欢木镰孢、尖镰孢、多枝镰孢、粉红镰孢、接骨木镰孢、肤色镰孢、硫色镰孢、圆镰孢、拟丝孢镰孢或镶片镰孢；腐质霉属，优选特异腐质霉或疏棉状腐质霉；或木霉属，优选哈茨木霉、康宁木霉、长枝木霉、里氏木霉或绿色木霉。

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

下面给出的是本发明多肽组合物的优选的用途的实例。本发明的多肽组合物的剂量和使用该组合物的其它条件可以基于本领域内已知的方法确定。

用途

本发明还涉及降解或转化含纤维素物质的方法，其包括：用组合物处理含纤维素物质，所述组合物包含有效量的本发明具有内切葡聚糖酶活性的多肽。在优选方面，所述方法还包括回收已降解或转化的含纤维素物质。

本发明的多肽和宿主细胞可用于单糖、二糖和多糖的生产中，所述糖类

作为来自含纤维素生物质的化学或发酵原料用于产生乙醇、塑料或其它产品或中间物。包含具有内切葡聚糖酶活性的多肽的组合物可以是去除或不去除细胞的粗发酵液形式,或是半纯化或纯化的酶制剂形式。或者,组合物可以包含本发明的宿主细胞,所述宿主细胞在使用生物质的发酵方法中作为具有内切葡聚糖酶活性的多肽的来源。宿主细胞还可以包含编码在生物质加工中有用的上述其它蛋白质和酶的天然或异源基因。具体地,本发明的多肽和宿主细胞可以通过部分或全部降解纤维素或半纤维素而增加加工残余物(干燥的蒸馏残渣、酿酒酒糟、甘蔗渣等)的价值。组合物还可以包含在处理生物质中有用的其它蛋白质和酶,例如,纤维二糖水解酶、 β -葡糖苷酶、半纤维素分解酶、增强剂(WO 2005/074647, WO 2005/074656)等。

在本发明的方法中,可以使用任何含纤维素物质,如生物质。在本文中可理解的是,术语“含纤维素物质”包括木质纤维素。生物质可以包括但不限于木材资源、城市固体废物、废纸、作物和作物残余物(参见,例如 Wiselogle 等, 1995, 在 *Handbook on Bioethanol* (编者 Charles E. Wyman) 中, 第 105-118 页, Taylor & Francis, Washington D.C.; Wyman, 1994, *Bioresource Technology* 50: 3-16; Lynd, 1990, *Applied Biochemistry and Biotechnology* 24/25: 695-719; Mosier 等, 1999, *Recent Progress in Bioconversion of Lignocellulosics*, 在 *Advances in Biochemical Engineering/Biotechnology* 中, 总编 T. Scheper, 第 65 卷, 第 23-40 页, Springer-Verlag, New York)。

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

将三种主要类别的酶用于分解含纤维素生物质:

(1) “内切-1,4- β -葡聚糖酶”或 1,4- β -D-葡聚糖-4-葡聚糖水解酶(EC 3.2.1.4),其随机作用于可溶和不溶的 1,4- β -葡聚糖底物。

(2) “外切-1,4- β -葡聚糖酶”包括 1,4- β -D-葡聚糖葡糖水解酶(EC 3.2.1.74),

其从 1,4- β -D-葡聚糖释放 D-葡萄糖并且缓慢水解 D-纤维二糖;和纤维二糖水解脱酶(1,4- β -D-葡聚糖纤维二糖水解脱酶, EC 3.2.1.91), 其从 1,4- β -葡聚糖释放 D-纤维二糖。

(3) “ β -D-葡糖苷酶”或 β -D-葡糖苷葡糖水解脱酶(EC 3.2.1.21), 其作用于从纤维二糖和可溶的纤维糊精以及一系列糖苷释放 D-葡萄糖单位。

本发明具有内切葡聚糖酶活性的多肽可以与其它纤维素分解蛋白质, 例如, 外切-1,4- β -D-葡聚糖酶和 β -D-葡糖苷酶, 一同使用以降解生物质底物的纤维素成分(参见, 例如, Brigham 等, 1995, 于 *Handbook on Bioethanol* (编者 Charles E. Wyman)中, 第 119-141 页, Taylor & Francis, Washington D.C.; Lee, 1997, *Journal of Biotechnology* 56: 1-24)。术语“纤维素分解蛋白质”在本文中定义为显示出能够在测试条件下水解或转化或降解纤维素的那些蛋白质或蛋白质的混合物。

外切-1,4- β -D-葡聚糖酶和 β -D-葡糖苷酶可以通过本领域已知的任何已知方法产生(参见, 例如, Bennett, J.W.和 LaSure, L. (编者), *More Gene Manipulations in Fungi*, Academic Press, CA, 1991)。

具有内切葡聚糖酶活性的多肽和其它纤维素分解蛋白质的最适量取决于几个因素, 其包括但不限于, 组分纤维素分解蛋白质的混合物、含纤维素底物、含纤维素底物的浓度、含纤维素底物的预处理、温度、时间、pH 和包括发酵生物体(例如, 同步糖化和发酵的酵母)。

在优选的方面, 每克含纤维素物质的具有内切葡聚糖酶活性的多肽的量是约 0.5 至约 50 mg, 优选约 0.5 至约 40 mg, 更优选约 0.5 至约 25 mg, 更优选约 0.75 至约 20 mg, 更优选约 0.75 至约 15 mg, 甚至更优选约 0.5 至约 10 mg, 并且最优选约 2.5 至约 10 mg 每 g 含纤维素物质。

在另一个优选的方面, 每克含纤维素物质的纤维素分解蛋白质的量是约 0.5 至约 50 mg, 优选约 0.5 至约 40 mg, 更优选约 0.5 至约 25 mg, 更优选约 0.75 至约 20 mg, 更优选约 0.75 至约 15 mg, 甚至更优选约 0.5 至约 10 mg, 并且最优选约 2.5 至约 10 mg 每 g 含纤维素物质。

在本发明的方法中, 组合物可补充一种或几种的其它酶活性, 以促进含纤维素物质的降解。优选的其它酶是半纤维素酶、酯酶(例如, 脂肪酶、磷脂酶和/或角质酶)、蛋白酶、漆酶、过氧化物酶或它们的混合物。

在本发明的方法中, 可以在发酵前或过程中(包括在发酵微生物的繁殖的

过程中或之后)加入其它酶。

酶可以源自或获得自任何合适的来源,包括细菌、真菌、酵母或哺乳动物来源。术语“获得”在本文中表示酶可以从自然地产生所述酶作为天然酶的生物体分离。术语“获得”在本文中还表示酶可以在宿主生物体中重组产生,其中重组产生的酶对宿主生物体是天然或外源的,或具有修饰的氨基酸序列,例如,具有缺失的、插入的和/或取代的一个或几个氨基酸,即,突变的重组产生的酶和/或天然氨基酸序列的片段或由本领域已知的核苷酸改组方法产生的酶。天然酶的意义中包含天然变体,而外源酶的意义中包含通过重组获得的变体,如通过定向诱变或改组。

还可以将酶纯化。本文中使用的术语“纯化”包括不含来自其所源自的生物体中其它成分的酶。术语“纯化”还包括不含来自其所获得的天然生物体的组分的酶。可将酶纯化,仅存在少量的其它蛋白质。表述“其它蛋白质”具体涉及其它酶。本文使用的术语“纯化”还指去除其它组分,特别是存在于本发明的酶的起源细胞中的其它蛋白质,最特别是存在于本发明的酶的起源细胞中的其它酶。酶可为“基本上纯的”,即,不含来自产生所述酶的生物体的其它组分,即,例如,用于重组产生酶的宿主生物体。在优选的方面,酶是至少 75% (w/w) 纯,优选至少 80% 纯,更优选至少 85% 纯,更优选至少 90% 纯,更优选至少 95% 纯,更优选至少 96% 纯,更优选至少 97% 纯,甚至更优选至少 98% 纯的,或者最优选至少 99% 纯。在另一个优选的方面,酶是 100% 纯的。

本发明使用的酶可以是示于本文所述方法的任何形式,如,例如,含有或者不含细胞的粗发酵液、干粉或颗粒、无粉尘颗粒、液体、稳定化的液体或受保护的酶。颗粒可以通过,例如美国专利号 4,106,991 和 4,661,452 中公开的方法产生,或者可任选地通过本领域已知的方法包覆。可以,例如,根据已确立的方法通过加入稳定剂(如糖、糖醇或另一种多羟基化合物(polyol),和/或乳酸或另一种有机酸)将液体酶制剂稳定化。受保护的酶可以根据 EP 238,216 中公开的方法制备。

可以使用本发明的方法将含纤维素物质处理成为很多有用的有机产品、化学品和燃料。除乙醇外,一些可以由纤维素产生的用品和专门化学品包括木糖、丙酮、乙酸、甘氨酸、赖氨酸、有机酸(例如,乳酸)、1,3-丙二醇、丁二醇、甘油、乙二醇、糠醛、多羟基链烷酸(polyhydroxyalkanoates)、顺式、

顺式-粘康酸和动物饲料(Lynd, L. R., Wyman, C. E.,和 Gerngross, T. U., 1999, *Biocommodity Engineering, Biotechnol. Prog.*, 15: 777-793; Philippidis, G. P., 1996, Cellulose bioconversion technology, 于 *Handbook on Bioethanol: Production and Utilization* 中, Wyman, C. E.,编, Taylor & Francis, Washington, DC, 179-212; 及 Ryu, D. D. Y.,和 Mandels, M., 1980, Cellulases: biosynthesis and applications, *Enz. Microb. Technol.*, 2: 91-102)。可能的共生产有助于扩展从可发酵糖合成的多种有机产物的范围。生物处理后残留的富含木质素的残余物可以转化成木质素衍生的化学品, 或者用于产生能量。

根据本发明的方法处理含纤维素物质使用的常规方法为本领域那些技术人员充分理解。可以使用任何常规生物质处理设备来实施本发明的方法, 所述设备设定为根据本发明操作。

这种设备可以包括批式搅拌反应器、具有超滤的连续流搅拌反应器、连续活塞流柱式反应器(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)。

常规方法包括但不限于, 糖化、发酵、分离的水解和发酵(SHF)、同步糖化和发酵(SSF)、同步糖化和共发酵(SSCF)、混合的水解和发酵(HHF), 和直接微生物转化(DMC)。

SHF 使用分离的处理步骤, 首先将纤维素用酶水解为葡萄糖, 然后将葡萄糖发酵成为乙醇。在 SSF 中, 纤维素的酶水解和葡萄糖到乙醇的发酵组合在一个步骤中(Philippidis, G. P., 1996, Cellulose bioconversion technology, in *Handbook on Bioethanol: Production and Utilization*, Wyman, C. E., ed., 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 包括在同一个反应器但不同的温度进行的两个分开的步骤, 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 在一个步骤中组合了所有三个过程(纤维素酶产生、纤维素水解和发酵)(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)。

“发酵”或“发酵过程”指任何发酵过程或包含发酵步骤的任何过程。发酵过程无限制地包括, 用于产生发酵产品的发酵过程, 所述发酵产品包括醇(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、丙酮酸、己二酸、抗坏血酸、柠檬酸、2,5-二酮-D-葡萄糖酸、蚁酸、反丁烯二酸、葡萄糖二酸、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、丙酸、丁二酸和木质酸); 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); 气体(例如, 甲烷、氢气(H₂)、二氧化碳(CO₂)和一氧化碳(CO))。发酵过程还包括在消费品醇工业(例如, 啤酒和酒)、乳制品业(例如, 发酵乳制品)、皮革业和烟草业中使用的发酵过程。

本发明进一步涉及产生物质的方法, 其包括: (a)用组合物糖化含纤维素物质, 所述组合物包含有效量的具有内切葡聚糖酶活性的多肽; (b)用一种或几种(more)的发酵微生物发酵步骤(a)的糖化的含纤维素物质; 和(c)从发酵回收物质。包含具有内切葡聚糖酶活性的多肽的组合物可以是去除或不去除细胞的粗发酵液的形式, 或者是半纯化或纯化的酶制剂形式, 或者组合物可以包含本发明的宿主细胞, 作为具有生物物质的发酵过程中有内切葡聚糖酶活性的多肽的来源。

物质可以是任何源自发酵的物质。在优选的实施方案中, 物质是醇。可理解的是, 术语“醇”包括包含一个或几个羟基基团的物质。在更优选的实施方案中, 所述醇是阿拉伯醇。在另一个更优选的实施方案中, 所述醇是丁醇。在另一个更优选的实施方案中, 所述醇是乙醇。在另一个更优选的实施方案中, 所述醇是甘油。在另一个更优选的实施方案中, 所述醇是甲醇。在另一个更优选的实施方案中, 所述醇是1,3-丙二醇。在另一个更优选的实施方案中, 所述醇是山梨醇。在另一个更优选的实施方案中, 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, in *Advances in Biochemical Engineering/Biotechnology*, Scheper, T., ed., Springer-Verlag Berlin Heidelberg, Germany, 65: 207-241; Silveira, M. M.,和 Jonas, R., 2002, The biotechnological production of sorbitol, *Appl. Microbiol. Biotechnol.* 59: 400-408; Nigam, P.,和 Singh, D., 1995, Processes for fermentative production of xylitol – a sugar substitute, *Process Biochemistry* 30 (2): 117-124; Ezeji, T. C., Qureshi, N.和 Blaschek, H. P., 2003, Production of acetone, butanol and ethanol by *Clostridium beijerinckii* BA101 and *in situ* recovery by gas stripping, *World Journal of Microbiology and Biotechnology* 19 (6): 595-603.

在另一个优选的实施方案中,所述物质是有机酸。在另一个更优选的实施方案中,所述有机酸是乙酸。在另一个更优选的实施方案中,所述有机酸是丙酮酸。在另一个更优选的实施方案中,所述有机酸是己二酸。在另一个更优选的实施方案中,所述有机酸是抗坏血酸。在另一个更优选的实施方案中,所述有机酸是柠檬酸。在另一个更优选的实施方案中,所述有机酸是2,5-二酮-D-葡萄糖酸。在另一个更优选的实施方案中,所述有机酸是蚁酸。在另一个更优选的实施方案中,所述有机酸是反丁烯二酸。在另一个更优选的实施方案中,所述有机酸是葡萄糖二酸。在另一个更优选的实施方案中,所述有机酸是葡糖酸。在另一个更优选的实施方案中,所述有机酸是葡糖醛酸。在另一个更优选的实施方案中,所述有机酸是戊二酸。在另一个优选的实施方案中,所述有机酸是3-羟基丙酸。在另一个更优选的实施方案中,所述有机酸是衣康酸。在另一个更优选的实施方案中,所述有机酸是乳酸。在另一个更优选的实施方案中,所述有机酸是苹果酸。在另一个更优选的实施方案中,所述有机酸是丙二酸。在另一个更优选的实施方案中,所述有机酸是草酸。在另一个更优选的实施方案中,所述有机酸是丙酸。在另一个更优选的实施方案中,所述有机酸是丁二酸。在另一个更优选的实施方案中,所述有机酸是木质酸。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.

在另一个优选的实施方案中,所述物质是酮。可理解的是术语“酮”包括含有一个或几个酮基的物质。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。

在另一个优选的实施方案中,所述物质是氨基酸。在另一个更优选的实施方案中,所述有机酸是天冬氨酸。在另一个更优选的实施方案中,所述氨基酸是谷氨酸。在另一个更优选的实施方案中,所述氨基酸是甘氨酸。在另一个更优选的实施方案中,所述氨基酸是赖氨酸。在另一个更优选的实施方案中,所述氨基酸是丝氨酸。在另一个更优选的实施方案中,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。

在另一个优选的实施方案中,所述物质是气体。在另一个更优选的实施方案中,所述气体是甲烷。在另一个更优选的实施方案中,所述气体是 H_2 。在另一个更优选的实施方案中,所述气体是 CO_2 。在另一个更优选的实施方案中,所述气体是 CO 。参见,例如, Kataoka, N., A. Miya,和 K. Kiriyama, 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. in *Biomass and Bioenergy*, Vol. 13 (1-2), pp. 83-114, 1997, Anaerobic digestion of biomass for methane production: A review。

从含纤维素物质产生物质通常需要四个主要步骤。这四个步骤是预处理、酶水解、发酵和回收。下面示例的是产生乙醇的过程,但可理解的是,可以使用相似的过程产生其它物质,例如,上述物质。

预处理。在预处理或预水解步骤中,将含纤维素物质加热以分解木质素和碳水化合物结构,溶解大部分半纤维素,并使纤维素分解酶能够到达纤维素部分。加热或者直接用蒸汽进行,或者在浆料中进行,可以向物质加入催化剂以加速反应的进行。催化剂包括强酸,如硫酸和 SO_2 , 或碱,如氢氧化钠。预处理阶段的目的是促进酶和微生物的渗透。含纤维素生物质还可以经过热蒸汽喷发的预处理(参见美国专利申请号 20020164730)。

糖化。在酶水解步骤中,所述步骤也称作糖化,将本文所述酶加入预处理的物质中,将纤维素部分转化成葡萄糖和/或其它糖。糖化通常在搅拌釜反应器或发酵罐中,在控制 pH、温度和混合条件下进行。糖化步骤可以持续长达 200 小时。糖化可以在约 $30^{\circ}C$ 至约 $65^{\circ}C$, 特别是约 $50^{\circ}C$ 的温度,和在约 4-约 5, 尤其是约 pH 4.5 的 pH 进行。为了产生能由酵母代谢的葡萄糖,通常在存在 β -葡糖苷酶时进行水解。

发酵。在发酵步骤中，作为预处理和酶水解步骤的结果从含纤维素物质释放的糖，通过发酵生物体(如酵母)发酵成为乙醇。发酵还可与酶水解在同一个容器中，同样在控制 pH、温度和混合条件下同时进行。当糖化和发酵在同一个容器中同时进行时，过程通常称作同步的糖化和发酵或 SSF。

在本发明的发酵过程中可以使用任何合适的含纤维素底物或原材料。通常根据理想的发酵产品(即，要从发酵获得的物质)和使用的方法来选择底物，如本领域中所公知。适用于本发明的方法中的底物的实例包括含纤维素物质，如木材或植物残渣或获得自能够由发酵微生物代谢的经处理的含纤维素物质的低分子量糖 DP1-3，所述物质可以通过直接向发酵培养基中加入而提供。

术语“发酵培养基”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过程(SSF)中使用的培养基。

“发酵微生物”指适用于理想的发酵过程中的任何微生物。根据本发明的合适的发酵微生物能将糖(如葡萄糖、木糖、阿拉伯糖、甘露糖、半乳糖或寡糖)直接或间接地发酵(即，转化)成理想的发酵产品。发酵微生物的实例包括真菌生物体，如酵母。优选的酵母包括酵母属菌种，且具体为酿酒酵母。商业上可得到的酵母包括，例如 Red Star®/™/Lesaffre Ethanol Red (可从 Red Star/Lesaffre, USA 获得)、FALI (可从 Fleischmann's Yeast, Burns Philp Food Inc., USA 的部门获得)、SUPERSTART (可从 Alltech 获得)、GERT STRAND (可从 Gert Strand AB, Sweden 获得)和 FERMIOL (可从 DSM Specialties 获得)。

在优选的实施方案中，所述酵母是酵母属。在更优选的实施方案中，所述酵母是酿酒酵母。在另一个更优选的实施方案中，所述酵母是糖化酵母(*Saccharomyces distaticus*)。在另一个更优选的实施方案中，所述酵母是 *Saccharomyces uvarum*。在另一个优选的实施方案中，所述酵母是克鲁维酵母。在另一个更优选的实施方案中，所述酵母是 *Kluyveromyces marxianus*。在另一个更优选的实施方案中，所述酵母是 *Kluyveromyces fragilis*。在另一个优选的实施方案中，所述酵母是念珠菌属。在另一个更优选的实施方案中，所述酵母是假热带念珠菌(*Candida pseudotropicalis*)。在另一个更优选的实施方案中，所述酵母是 *Candida brassicae*。在另一个优选的实施方案中，所述酵母是 *Clavispora*。在另一个更优选的实施方案中，所述酵母是 *Clavispora lusitaniae*。在另一个更优选的实施方案中，所述酵母是 *Clavispora opuntiae*。

在另一个优选的实施方案中,所述酵母是管囊酵母属(*Pachysolen*)。在另一个更优选的实施方案中,所述酵母是管囊酵母(*Pachysolen tannophilus*)。在另一个优选的实施方案中,所述酵母是酒香酵母属(*Bretannomyces*)。在另一个更优选的实施方案中,所述酵母是克劳森酒香酵母(*Bretannomyces clausenii*) (Philippidis, G. P., 1996, Cellulose bioconversion technology, in *Handbook on Bioethanol: Production and Utilization*, Wyman, C. E., ed., Taylor & Francis, Washington, DC, 179-212)。

可以有效地将葡萄糖发酵成乙醇的细菌包括,例如,运动发酵单胞菌(*Zymomonas mobilis*)和热纤维梭菌(*Clostridium thermocellum*) (Philippidis, 1996, 见上文)。

在本领域中公知上述生物体还能用于产生其它物质,如本文所述。

在酿酒酵母(Chen, Z., Ho, N. W. Y., 1993, Cloning and improving the expression of *Pichia stipitis* xylose reductase gene in *Saccharomyces cerevisiae*, *Appl. Biochem. Biotechnol.* 39-40: 135-147; Ho, N. W. Y., Chen, Z., Brainard, A. P., 1998, Genetically engineered *Saccharomyces* yeast capable of effectively cofermenting glucose and xylose, *Appl. Environ. Microbiol.* 64: 1852-1859)或细菌如大肠杆菌(Beall, D. S., Ohta, K., Ingram, L. O., 1991, Parametric studies of ethanol production from xylose and other sugars by recombinant *Escherichia coli*, *Biotech. Bioeng.* 38: 296-303)、产酸克雷伯氏菌(*Klebsiella oxytoca*) (Ingram, L. O., Gomes, P. F., Lai, X., Moniruzzaman, M., Wood, B. E., Yomano, L. P., York, S. W., 1998, Metabolic engineering of bacteria for ethanol production, *Biotechnol. Bioeng.* 58: 204-214)和运动发酵单胞菌(Zhang, M., Eddy, C., Deanda, K., Finkelstein, M.,和 Picataggio, S., 1995, Metabolic engineering of a pentose metabolism pathway in ethanologenic *Zymomonas mobilis*, *Science* 267: 240-243; Deanda, K., Zhang, M., Eddy, C.,和 Picataggio, S., 1996, Development of an arabinose-fermenting *Zymomonas mobilis* strain by metabolic pathway engineering, *Appl. Environ. Microbiol.* 62: 4465-4470)中克隆异源基因导致构建能将己糖和戊糖转化成乙醇(共发酵)的生物体。

通常向降解的纤维素或水解产物中加入酵母或另一种微生物,进行约 24 至 96 小时的发酵,如约 35 至约 60 小时。温度通常在约 26°C-约 40°C,特别是约 32°C,并在约 pH 3 至约 pH 6,特别是约 pH 4-5 进行。

在优选的实施方案中，将酵母或另一种微生物施用于降解的纤维素或水解产物，发酵进行约 24 至约 96 小时，如通常 35-60 小时。在优选的实施方案中，温度通常在约 26 至约 40℃ 之间，特别是约 32℃，而 pH 通常在约 pH 3 至约 pH 6，优选约 pH 4-5。优选以大约 10^5 至 10^{12} ，优选大约 10^7 至 10^{10} ，特别是大约 5×10^7 活菌数每 ml 发酵液的量加入酵母或另一种微生物。在乙醇产生过程阶段中，酵母细胞计数应该优选在大约 10^7 至 10^{10} ，尤其是约 2×10^8 的范围中。关于使用酵母进行发酵的其它指导可见于例如“The Alcohol Textbook” (Editors K. Jacques, T.P. Lyons 和 D.R. Kelsall, Nottingham University Press, United Kingdom 1999) 中，所述文献通过引用并入本文。

本领域最广泛使用的方法是同步的糖化和发酵(SSF)方法，其中没有糖化的保留阶段，意味着酵母和酶一起添加。

对于乙醇产生，在发酵后蒸馏醪(mash)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。

发酵刺激剂可以与本文所述的任何酶过程结合使用，以进一步改进发酵过程，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、烟酸、内消旋肌醇(meso-inositol)、硫胺、维生素 B6 (pyridoxine)、对氨基苯甲酸、叶酸、核黄素和维生素 A、B、C、D 和 E。参见，例如，Alfenore 等, Improving ethanol production and viability of *Saccharomyces cerevisiae* by a vitamin feeding strategy during fed-batch process, Springer-Verlag (2002)，其通过引用并入本文。矿物质的实例包括能够提供营养的矿物质和矿物质盐，所述营养包括 P、K、Mg、S、Ca、Fe、Zn、Mn 和 Cu。

回收 从发酵的含纤维素物质中分离醇，并通过蒸馏的常规方法纯化。能够获得纯度高达约 96 vol. % 的乙醇，其可以用作，例如，燃料乙醇、饮料乙醇(即，中性饮料酒)或工业乙醇。

对于其它物质，可以使用本领域已知的任何方法，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，制备型等电聚焦)、差示溶解度(例如，硫酸铵沉淀)、SDS-PAGE、蒸馏或提取。

在本发明的方法中，可以向具有内切葡聚糖酶活性的多肽和其它纤维素分解蛋白质补充一种或几种其它酶活性，以改进含纤维素物质的降解。优选的其它酶是半纤维素酶、酯酶(例如，脂肪酶、磷脂酶和/或角质酶)、蛋白酶、漆酶、过氧化物酶，或它们的混合物。

在本发明的方法中，可以在发酵前或发酵过程中(包括在繁殖发酵微生物的过程中或之后)加入其它酶。

信号肽

本发明还涉及核酸构建体，所述核酸构建体包含编码蛋白质的基因，其中所述基因与编码信号肽的核苷酸序列可操作地连接，所述信号肽包含 SEQ ID NO: 2 的氨基酸 1 至 22、SEQ ID NO: 6 的氨基酸 1 至 19 或 SEQ ID NO: 8 的氨基酸 1 至 18，或由 SEQ ID NO: 2 的氨基酸 1 至 22、SEQ ID NO: 6 的氨基酸 1 至 19 或 SEQ ID NO: 8 的氨基酸 1 至 18 组成，其中所述基因对于该核苷酸序列是外源的。

在优选的方面，所述核苷酸序列包括 SEQ ID NO: 1 的核苷酸 13 至 79 或由 SEQ ID NO: 1 的核苷酸 13 至 79 组成。在另一优选的方面，所述核苷酸序列包括 SEQ ID NO: 3 的核苷酸 17 至 73 或由 SEQ ID NO: 3 的核苷酸 17 至 73 组成。在另一优选的方面，所述核苷酸序列包括 SEQ ID NO: 5 的核苷酸 53 至 106 或由 SEQ ID NO: 5 的核苷酸 53 至 106 组成。

本发明还涉及包含这种核酸构建体的重组表达载体和重组宿主细胞。

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

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

优选蛋白质是激素或其变体、酶、受体或其部分、抗体或其部分，或报告蛋白(reporter)。在更优选的方面，所述蛋白质是氧化还原酶、转移酶、水

解酶、裂合酶(lyase)、异构酶或连接酶。在更加优选的方面,所述蛋白质是氨肽酶、淀粉酶、糖酶、羧肽酶、过氧化氢酶、纤维素酶、几丁质酶、角质酶、环糊精糖基转移酶、脱氧核糖核酸酶、酯酶、 α -半乳糖苷酶、 β -半乳糖苷酶、葡萄糖淀粉酶、 α -葡萄糖苷酶、 β -葡萄糖苷酶、转化酶、漆酶、另外的脂肪酶、甘露糖苷酶、变位酶(mutanase)、氧化酶、果胶分解酶、过氧化物酶、肌醇六磷酸酶、多酚氧化酶、蛋白水解酶、核糖核酸酶、转谷氨酰胺酶或木聚糖酶。

基因可以从任何原核、真核或其它来源获得。

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

材料

作为缓冲液和底物使用的化学品是至少试剂等级的商业产品。

菌株

将土生梭孢霉 NRRL 8126 和 *Cladorrhinum foecundissimum* ATCC 62373 用作具有内切葡聚糖酶活性的糖基水解酶家族 7 (CEL7)多肽的来源。米曲霉 JaL250 菌株(WO 99/61651)用于从土生梭孢霉表达 CEL7 多肽,米曲霉菌株 HowB104 (α -淀粉酶阴性)用于表达来自 *cladorrhinum foecundissimum* 的 CEL7 多肽。

培养基

YEG 培养基由 0.5%酵母提取物和 2%葡萄糖组成。

马铃薯右旋糖(dextrose)培养基每升由 39 克马铃薯右旋糖(Difco)组成。

PDA 平板由每升 39 克马铃薯右旋糖琼脂组成。

M400 培养基由每升 50 g 麦芽糖糊精、2 g $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ 、2 g KH_2PO_4 、4 g 柠檬酸、8 g 酵母提取物、2 g 尿素、0.5 ml AMG 痕量金属溶液和 0.5 g 氯化钙组成。

MDU2BP 培养基由每升 45 g 麦芽糖、1 g $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ 、1 g NaCl、2 g K_2SO_4 、12 g KH_2PO_4 、7 g 酵母提取物、2 g 尿素和 0.5 ml AMG 痕量金属溶

液组成, pH 调至 5.0。

AMG 痕量金属溶液由如下组成: 每升 14.3 g $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ 、2.5 g $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ 、0.5 g $\text{NiCl}_2 \cdot 6\text{H}_2\text{O}$ 、13.8 g $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ 、8.5 g $\text{MnSO}_4 \cdot 7\text{H}_2\text{O}$ 和 3 g 柠檬酸。

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

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

LB 培养基由如下物质组成: 每升 10 g 胰蛋白胨、5 g 酵母提取物和 5 g 氯化钠。

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

SOC 培养基由如下物质组成: 2%胰蛋白胨、0.5%酵母提取物、10 mM NaCl 、2.5 mM KCl 、10 mM MgCl_2 、10 mM MgSO_4 , 高压灭菌后添加过滤除菌的葡萄糖至 20 mM。

冷冻培养基由 60% SOC 培养基和 40%甘油组成。

2X YT 培养基由如下物质组成: 每升 16 g 胰蛋白胨、10 g 酵母提取物、5 g NaCl 和 15 g 细菌用琼脂。

含纤维素的 PD 培养基由每升 24 克马铃薯右旋糖(Difco)和 30 克 Solcafloc (Diacel, 可由 Dicalie-Europe-Nord, Gent, Belgium 获得)组成。

每升 SC-琼脂由 SC-URA 培养基(含有指定的葡萄糖或半乳糖)和 20 g 琼脂组成。

每升含有半乳糖的 0.1% AZCL HE 纤维素 SC 琼脂平板由含半乳糖的 SC-URA 培养基、20 g 琼脂和 0.1% AZCL HE 纤维素(Megazyme, Wicklow, Ireland)组成。

实施例 1: 鉴定来自土生梭孢霉 NRRL 8126 的 CEL7C 内切葡聚糖酶

将来自生长于添加有 1% Sigmacell (Sigma Chemical Co., St. Louis, MO, USA)的 NNCYPmod 培养基上的土生梭孢霉 NRRL 8126 的新鲜平板的琼脂糖

圆柱状样本(plug)接种到添加有 1%葡萄糖的 50 ml NNCYPmod 培养基中,并在 45°C 和 200 rpm 温育 25 个小时。将两 ml 的这种培养物用于接种添加有 2% Sigmacell-20 的 15 x 100 ml (500 ml 烧瓶)和 2 x 50 ml (250 ml 烧瓶)的 NNCYPmod 培养基并在 45°C、200 rpm 温育 4 天。汇集培养物, 3000 x g 离心 10 分钟, 通过 Nalgene 281-5000 玻璃纤维预滤器(Nalge Nunc Int'l., Rochester, NY, USA)过滤上清。滤液冷却至 4°C 用于存储。

然后将滤液进一步过滤(GP Express 膜、聚醚砜(polyethersulfone), 0.22 μ m, Millipore, Bedford, MA, USA), 用 50 mM 乙酸钠 pH 5.0 (Pall Filtron, North Borough, MA, 10 kDa 聚醚砜膜, 大约 10-20 psi)进行缓冲液交换, 用 Amicon 超滤装置(Millipore, Bedford, MA, 10 kDa, 40 psi, 4°C)浓缩。将浓缩样品脱盐, 通过流经用 20 mM Tris-HCl pH 8.2 平衡和操作的 10DG EconoPAC 柱 (Bio-Rad, Hercules, CA, USA)来进行缓冲液交换。使用 BCA 蛋白质测试试剂盒(Pierce, Rockford, IL, USA)测定蛋白质浓度, 在该试剂盒中将牛血清白蛋白用作标准蛋白质。将含蛋白质的级分上样到用 20 mM Tris-HCl, pH 8.2 平衡的 5-ml Hi Trap Q Sepharose™ Fast Flow 柱(Amersham Pharmacia, Uppsala Sweden)上, 在 AKTA FPLC 系统(Amersham Pharmacia, Uppsala Sweden)中以 2 ml 每分钟的流速操作。洗脱前, 用五个柱体积的起始缓冲液洗涤柱子。用 20 mM Tris-HCl pH 8.2 中 0 至 1.0 M NaCl (20 个柱子体积; 2 ml 级份)的线性梯度洗脱结合的物质。根据 280 nm 的 UV 谱, 汇集各级分用于进一步分析。如上所述, 用 Amicon 装置浓缩所汇集的在大约 10-210 mM NaCl 之间洗脱的级分。将此物质的等分试样在 8-16% Tris-Glycine SDS-PAGE 凝胶(BioRad Laboratories, Hercules, CA, USA)上电泳, 200 V 电泳 1 小时。包含精确的分子量标准(BioRad, Hercules, CA, USA)用于分子量测定。依照制造商建议的规程, 采用 Biosafe 考马斯染料(Coomassie Stain) (BioRad, Hercules, CA, USA)用于凝胶的蛋白质染色。

从凝胶切下迁移至大约 61 kDa 的条带, 进行凝胶内消化(in-gel digestion)并用串联质谱法(tandem mass spectrometry)从头测序。

用于肽测序的多肽凝胶内消化。用 MultiPROBE® II Liquid Handling Robot (PerkinElmer Life and Analytical Sciences, Boston, MA, USA)进行凝胶内消化。含目标多肽的二维凝胶斑在室温用溶于 100 mM 碳酸氢铵 pH 8.0 中的 50 μ l 10 mM 二硫苏糖醇(DTT)还原 30 分钟。还原后, 用溶于 100 mM 碳

酸氢铵 pH 8.0 中的 50 μ l 55 mM 碘乙酰胺将凝胶片烷基化 20 分钟。将干燥的凝胶片在室温在胰蛋白酶消化溶液(溶于 50 mM 碳酸氢铵 pH 8 中的 6 ng/ μ l 测序级胰蛋白酶(Promega, Madison, WI, USA))中溶胀 30 分钟, 之后在 40°C 消化 8 小时。每个所述反应步骤后都按照制造商的标准规程用适当的溶液进行多次洗涤和预洗涤。在反应之间用 50 μ l 的乙腈将凝胶脱水, 并在步骤之间将凝胶片风干。肽用 HPLC 级水中的 1%甲酸/2%乙腈提取两次, 30 分钟。将肽提取溶液转移到 96 孔带缘的(skirted) PCR 型平板(PCR type plate) (ABGene, Rochester, NY, USA), 所述带缘的 PCR 型平板已经冷却至 10–15°C 并用 96 孔平板盖(PerkinElmer Life and Analytical Sciences, Boston, MA, USA) 覆盖以防蒸发。再次于 4°C 存储平板, 直至可以进行质谱分析。

通过串联质谱法进行肽测序。为了通过串联质谱法进行肽测序, 使用 Q-ToF *micro*TM 混成正交式四极飞行时间质谱仪(hybrid orthogonal quadrupole time-of-flight mass spectrometer) (Waters Micromass[®] MS Technologies, Milford, MA, USA)进行 LC-MS/MS 分析。Q-ToF *micro*TM 质谱仪装配有 UltimateTM 毛细管和纳米流(nano-flow) HPLC 系统(Dionex, Sunnyvale, CA, USA), 其已经与 FAMOS 微型自动进样器(Dionex, Sunnyvale, CA, USA)和 Switchos II 柱转换装置(Dionex, Sunnyvale, CA, USA)偶联, 用于浓缩样品并对样品脱盐。将 6 μ l 由凝胶内消化回收的肽溶液上样到装配在注射环(injection loop)中的保护柱(guard column) (300 μ m ID x 5 cm, C18 PepMapTM) (Dionex, Sunnyvale, CA, USA)上, 并用 Switchos II 泵(Dionex, Sunnyvale, CA, USA)以 40 μ l/分钟用溶于水中的 0.1%甲酸洗涤 2 分钟。在 75 μ m ID x 15 cm、C18、3 μ m、100Å PepMapTM 纳米流融合的毛细管柱(Dionex, Sunnyvale, CA, USA)上, 用 NAN-75 校准器(Dionex, Sunnyvale, CA, USA)自 175 μ l/分钟的分流(split flow)以 175 nl/分钟的流速分离肽。在 45 分钟的时间施加的线性洗脱梯度为溶于 0.1%甲酸中的 5%至 60%乙腈。于 215 nm 监测柱洗脱液, 并通过装配有 nanospray 界面的电喷雾离子源将其导入到 Q-ToF *micro*TM 质谱仪。Q-ToF *micro*TM 质谱仪完全是用 MassLynxTM 软件 3.5 版(Waters Micromass[®] MS Technologies, Milford, MA, USA)控制的微处理器。以测量扫描(survey scan)模式并从 50 至 2000 m/z 的质量范围以 MS 至 MS/MS 的切换标准(switching criteria)获得数据, 以包括大于 10.0 计数/秒的离子强度和+2、+3 和+4 的电荷状态。可以获得多至 4 个共洗脱类别的分析谱, 扫描时间为 1.9 秒, 扫描间

期(inter-scan time)为 0.1 秒。通常使用 65 伏的锥形电压(cone voltage), 并对碰撞能量进行编程使之根据洗脱肽的质量和电荷状态而变化, 且处于 10-60 伏的范围内。获得的谱是组合的、平滑的且以自动方式居中的, 并产生峰列表。用 ProteinLynx™ Global Server 1.1 软件(Waters Micromass® MS Technologies, Milford, MA, USA)针对选择的数据库检索产生的峰列表。评估 ProteinLynx™检索的结果, 进一步通过评估每个感兴趣离子的 MS/MS 谱来分析未鉴定的蛋白质, 通过鉴定 y 和 b 离子系列并将质量差异与适当的氨基酸进行匹配来测定从头序列。

对于大约 61 kDa 的多肽凝胶条带, 从几个多电荷离子通过质谱分析由从头测序获得土生梭孢霉 CEL7C 内切葡聚糖酶的肽序列。双倍荷电的 933.49 m/z 胰蛋白酶肽离子(tryptic peptide ion)测定为 Phe-[Ile 或 Leu]-Thr-Asp-Asp-Gly-Thr-Thr-Ser-Gly-Thr-[Ile-Leu]-Asn-[Gln-Lys]-[Ile-Leu]-[Gln-Lys]-Arg (SEQ ID NO: 2 的氨基酸 256 至 272)。第二个双倍带电的 1108.58 胰蛋白酶肽离子也进行从头测序, 且部分序列测定为 Tyr-Gly-Pro-Gly-[Ile-Leu]-Thr-Val-Asp-Thr-Ser-[Lys-Gln] (SEQ ID NO: 2 的氨基酸 238 至 248)。

实施例 2: 土生梭孢霉 NRRL 8126 基因组 DNA 提取

土生梭孢霉 NRRL 8126 于 37°C、250 rpm 在 25 ml YEG 培养基中培养 24 小时。然后通过 Miracloth™ (Calbiochem, La Jolla, CA, USA)过滤收集菌丝体并用 25 ml 的 10 mM Tris-1 mM EDTA (TE)缓冲液洗涤一次。从菌丝体制备物排干多余的缓冲液, 随后冷冻在液氮中。在电动咖啡磨碎机中将冰冻的菌丝体制备物粉碎为细粉, 将粉末添加到含有 20 ml TE 缓冲液和 5 ml 20% w/v 十二烷基硫酸钠(SDS)的一次性(disposable)塑料离心管。混合物轻轻颠倒数次以确保混合, 用等体积苯酚:氯仿:异戊醇(25:24:1 v/v/v)提取两次。向提取的样品添加乙酸钠(3 M 溶液)至终浓度 0.3 M, 之后添加 2.5 体积的冰冷乙醇以沉淀 DNA。将管在 15,000 x g 离心 30 分钟以使 DNA 沉淀(pellet)。在将 DNA 沉淀风干 30 分钟然后重悬在 0.5 ml TE 缓冲液中。被无脱氧核糖核酸酶的核糖核酸酶 A 添加到重悬的 DNA 沉淀至 100 µg 每 ml 的浓度, 然后将混合物于 37°C 温育 30 分钟。添加蛋白酶 K (200 µg/ml), 将管于 37°C 再温育一小时。最后, 将样品于 12,000 x g 离心 15 分钟, 将上清加到 Qiaprep® 8 Manifold (QIAGEN Inc., Valencia, CA, USA)。柱子在真空下用 1 ml PB

(QIAGEN Inc., Valencia, CA, USA)和 1 ml PE (QIAGEN Inc., Valencia, CA, USA)洗涤两次。用 100 μ l TE 缓冲液洗脱分离的 DNA, 用乙醇沉淀, 用 70% 乙醇洗涤, 真空下干燥, 重悬于 TE 缓冲液中, 4°C 存储。

为了产生用于 PCR 扩增的基因组 DNA, 在带挡板的摇瓶中在添加有 1% 葡萄糖的 50 ml NNCYP 培养基中, 在 42°C 和 200 rpm 将土生梭孢霉 NRRL 8126 培养 24 小时。通过过滤收获菌丝体, 在 TE 缓冲液中洗涤两次, 液氮冷冻。将豌豆大小的冷冻菌丝体小块悬浮在 0.7 ml 溶于 TE 缓冲液中的 1% 十二烷基硫酸锂中, 在 FastPrep FP120 (ThermoSavant, Holbrook, NY, USA)中用等体积的 0.1 mm 氧化锆/二氧化硅珠(Biospec Products, Inc., Bartlesville, OK, USA)搅动 45 秒将其破碎。通过在 13,000 $\times g$ 离心 10 分钟去除碎片, 将澄清的上清加至 2.5 M 乙酸铵并在冰上温育 20 分钟。温育期后, 通过添加 2 体积的乙醇沉淀核酸。在微型离心机中在 4°C 离心 15 分钟后, 在 70%乙醇中洗涤沉淀并风干。将 DNA 重悬在 120 μ l 的 0.1X TE 缓冲液中, 在 37°C 与 1 μ l 无脱氧核糖核酸酶的核糖核酸酶 A 一起温育 20 分钟。添加乙酸铵至 2.5 M, 用 2 体积的乙醇沉淀 DNA。在 70%乙醇中洗涤沉淀, 风干, 重悬在 TE 缓冲液中。

实施例 3: 从土生梭孢霉 NRRL 8126 克隆编码 CEL7C 内切葡聚糖酶的基因

如实施例 1 所述的质谱分析显示质量为 933.49 的双倍带电的肽的部分序列是 **TDDGTTSGT**[I/L]**NQ**[I/L]**QR** (SEQ ID NO: 2 的氨基酸 258 至 272)。对上面以粗体显示的部分序列(SEQ ID NO: 2 的氨基酸 259 至 267)设计反义链寡核苷酸 CODEHOP 引物(Rose 等, 1998, *Nucleic Acids Res.* 26 : 1628-35)。引物序列如下所示。

5'-AGGGTGCCGCTGGTNGTNCRCRTC-3' (SEQ ID NO: 7)

根据存在于家族 7 的许多糖基水解酶中的保守序列(特别是 **AGAKYGTGYCD** (SEQ ID NO: 2 的氨基酸 164 至 173; 发现以粗体显示的残基在 CEL7C 中是保守的))设计第二个 CODEHOP 有义链引物。引物序列如下所示。

5'-AGCTGGTGCTAAATATGGTACTGGNTAYTGYGA-3' (SEQ ID NO: 8)

在含 1X AmpliTaq 缓冲液(Applied Biosystems, Inc., Foster City, CA,

USA)、1.5 单位的 AmpliTaq DNA 聚合酶(Applied Biosystems, Inc., Foster City, CA, USA)、有义和反义引物各 1 μ M、和大约 1 μ g 来自土生梭孢霉 NRRL 8126 的基因组 DNA 的 30 μ l 体积中进行 PCR 扩增。在 RoboCycler® (Stratagene, La Jolla, CA, USA)中进行扩增, 程序为: 96°C 3 分钟、72°C 3 分钟(在此期间添加 DNA 聚合酶)的一个循环; 35 个循环, 每个循环为 94°C 45 秒、52°C、55°C 或 58°C 45 秒、72°C 1 分钟; 之后在 72°C 最终延伸 7 分钟。

用 40 mM Tris 碱-20 mM 乙酸钠-1 mM EDTA 二钠(TAE)缓冲液通过 3% 琼脂糖凝胶电泳对反应产物进行分级, 切下大约 300 bp 的条带, 用 QIAEX® II 凝胶提取试剂盒(QIAGEN Inc., Valencia, CA, USA)纯化, 用 TOPO TA 试剂盒(Invitrogen, Carlsbad, CA, USA)亚克隆。对来自一个大肠杆菌转化体的质粒测序, 发现含有编码预测的糖基水解酶家族 7 蛋白(CEL7C)的 315 bp 插入。该质粒被命名为 pPH32 (图 1)。

实施例 4: 鉴定来自土生梭孢霉 NRRL 8126 的 CEL7E 内切葡聚糖酶

将来自在添加有 1% Sigmacell 的 NNCYPmod 培养基上培养的土生梭孢霉 NRRL 8126 的新鲜平板的琼脂糖圆柱状样本(plug)接种到添加有 2% Sigmacell 的 25 ml NNCYPmod 培养基中, 并在 42°C 和 150 rpm 温育 3 天。通过 Nalgene 281-5000 玻璃纤维预滤器过滤培养液。滤液冷却至 4°C 以储存。

二维聚丙烯酰胺凝胶电泳。通过添加 30 μ l β -巯基乙醇和 300 μ l 饱和三氯乙酸(4°C 的饱和水溶液)在冰上温育 10 分钟, 之后添加 30 ml 冰冷的丙酮, 并进一步在冰上温育 30 分钟, 来沉淀来自实施例 1 中所述的汇集的培养物的三毫升滤液。沉淀的溶液于 4°C, 10,000 $\times$ g 离心 10 分钟, 倾析上清, 将沉淀用冰冷的丙酮漂洗两次并风干。将干燥的沉淀溶解在 0.2 ml 等电聚焦(IEF)样品缓冲液(9.0 M 尿素、3.0% (wt/v) 3-[(3-胆酰胺丙基)二甲基-铵]-1-丙烷磺酸盐(CHAPS, Pierce Chemical Co. Rockford, IL, USA)、1% (v/v) pH 4-7 两性电解质(ampholyte)、1% β -巯基乙醇和 0.005% 溴酚蓝, 溶于蒸馏水中)中。用来自 BioRad 实验室(Hercules, CA, USA)的 AG 501-X8 (D)、20-5-筛目(mesh)、混合床树脂对尿素储液去离子化。将去离子化的溶液储存于-20°C。在 LabQuake™摇床(Lab Industries, Berkeley, CA, USA)上以轻柔混合使所得混合物溶解(solubilize)几小时。在 IPG 复水托盘(Amersham Biosciences, Piscataway, NJ, USA)中, 将样品缓冲液-蛋白混合物施加于 11 cm IPG 条带

(strip) (BioRad Laboratories, Hercules, CA, USA)。在 20°C, 将 750 μ l 的等份干燥-条带包封液体(Amersham Biosciences, Piscataway, NJ, USA)层叠在 IPG 条带上, 以阻止蒸发并使之复水 12 小时, 同时用 IPGPhor 等电聚焦单元(Isoelectric Focusing Unit) (Amersham Biosciences, Piscataway, NJ, USA)施加 30 伏电压。将 IPGPhor 单元设置为恒定电压但每个条带最大电流为 50 μ A。复水 12 小时后, 等电聚焦条件如下: 200 伏 1 小时, 500 伏 1 小时和 1000 伏 1 小时。然后施加 1000 伏至 8000 伏的梯度 30 分钟, 设定等电聚焦使之在 8000 伏运行, 当达到>30,000 伏小时时完成。在第二维分析(second dimension analysis)之前, 通过如下将 IPG 凝胶条带还原并烷基化: 首先在 100 mg 二硫苏糖醇每 10 ml SDS-平衡缓冲液(50 mM Tris HCl pH 8.8、6.0 M 尿素、2% (w/v)十二烷基硫酸钠(SDS)、30%甘油、和 0.002% (w/v)溴酚蓝)中还原 15 分钟, 之后在 250 mg 碘乙酰胺每 10 ml 缓冲液的平衡缓冲液中在黑暗中烷基化 15 分钟。在 SDS-PAGE 电泳缓冲液(Invitrogen/Novex, Carlsbad, CA, USA)中快速漂洗 IPG 条带, 放置在 11 cm、1 孔 8-16% Tris-甘氨酸 SDS-PAGE 凝胶(BioRad Laboratories, Hercules, CA, USA)上用 Criterion 电泳单元(BioRad Laboratories, Hercules, CA, USA)于 50 伏电泳, 直至样品进入凝胶, 然后将电压增加到 200 伏, 使之运行直至溴酚蓝染料到达凝胶的底部。

多肽检测。用荧光 SYPRO Orange Protein Stain (Molecular Probes, Eugene, OR, USA)对二维凝胶染色。荧光染色法由 Malone 等, 2001, *Electrophoresis*, 22, 919-932 优化和改编而来。将 SDS-PAGE 凝胶在 40%乙醇、2%乙酸和 0.0005% SDS 中在平台摇移器(platform rocker)上固定 1 小时至过夜。采用由 2%乙酸和 0.0005% SDS 组成的三个重复洗涤步骤去除和替换固定液, 每次 30 分钟。黑暗中用 2%乙酸、0.0005% SDS 和 0.02% SYPRO Orange Protein 染料对凝胶染色 1.5 小时至过夜。在 Hoefer Processor Plus Staining Unit (Amersham Biosciences, Piscataway, NJ, USA)上进一步优化染色和脱色以改进再现性和自动化程度。利用蓝色荧光和 200 μ m 像素尺寸(pixel size)和 800 V 的光电倍增管增益(gain), 通过在 Molecular Dynamics STORM 860 成像系统(Amersham Biosciences, Piscataway, NJ, USA)上扫描来获得荧光染色的 SDS-PAGE 凝胶的图像。利用 ImageQuant 软件 5.0 版(Amersham Biosciences, Piscataway, NJ, USA)观察并修正图像。在 Dark Reader Blue transilluminator 上用橙色滤镜(Clare Chemical Co, Denver, CO, USA)进一步显示凝胶。用 2 mm Acu-Punch

活组织切片冲床(Acuderm Inc., Ft. Lauderdale, FL, USA)切下观察到的蛋白凝胶斑点并保存在 96 孔平板中,所述平板用 60%乙腈中的 0.1%三氟乙酸(TFA)预洗涤然后用 HPLC 级水再洗涤两次。在-20°C,将染色的二维凝胶斑保存在预洗涤过的平板中 25-50 μ l 的水中直至消化。

用胰蛋白酶凝胶内消化相应于分子量约为 50 kDa 且等电点约为 5.0 的 2D 凝胶斑并进行从头测序,如实施例 1 所述。将双倍荷电的 1114.516 m/z 胰蛋白酶肽离子测定为 Ser-Pro-Leu-Asn-Pro-Ala-Gly-Ala-Thr-Tyr-Gly-Thr-Gly-Tyr-Cam-Asp-Ala-Gln-Cam-Pro-Lys (SEQ ID NO: 5 的氨基酸 156 至 176,其中 Cam 是羧基氨基甲基半胱氨酸(carboxyamidomethylcysteine))。

针对上面以粗体显示的序列部分设计有义链寡核苷酸 CODEHOP 引物 (SEQ ID NO: 4 的 168 至 176)。引物序列如下所示。

5'-GGCTACTGCGACGCCARTGYCNAA-3' (SEQ ID NO: 9)

根据存在于许多家族 7 糖基水解酶中的保守序列(特别是 CCNEMDIWEAN)(随后发现所有氨基酸在 CEL7E 中在 SEQ ID NO: 4 的 193 至 203 都是保守的)设计第二个 CODEHOP 反义链引物。所述引物序列是:

5'-CCTCCCAGATRTCCATYTCGTTRCARCA-3' (SEQ ID NO: 10)

在包含 1X AmpliTaq 缓冲液、1.5 单位的 Taq DNA 聚合酶(New England Biolabs, Ipswich, MA, USA)、有义和反义引物各 1 μ M 和大约 1 μ g 的来自土生梭孢霉 NRRL 8126 的基因组 DNA(如实施例 2 所述制备)的 30 μ l 体积中进行 PCR。在 Stratagene RoboCycler®中进行扩增,程序为: 96°C 3 分钟、72°C 3 分钟(在此期间添加 DNA 聚合酶)的一个循环; 35 个循环,每个循环 94°C 45 秒、53°C、56°C 或 59°C 45 秒、72°C 1 分钟; 之后在 72°C 最终延伸 7 分钟。

用 TAE 缓冲液在 3%琼脂糖凝胶上对反应产物分级,切下大约 180 bp 的条带,用 QIAEX® II 凝胶提取试剂盒(QIAGEN Inc., Valencia, CA, USA)纯化,并使用 TOPO TA 试剂盒(Invitrogen, Carlsbad, CA, USA)亚克隆。对来自一个大肠杆菌转化体的质粒测序,发现含有编码预测的糖基水解酶家族 7 蛋白(CEL7E)的 169 bp 插入。该质粒被命名为 pPH37 (图 2)。

实施例 5: 土生梭孢霉 NRRL 8126 基因组 DNA 文库构建

利用噬菌体克隆载体 λ ZipLox (Life Technologies, Gaithersburg, MD, USA)构建基因组 DNA 文库,以大肠杆菌 Y1090ZL 细胞(Life Technologies,

Gaithersburg, MD, USA)作为宿主用于重组噬菌体的涂布和纯化, 大肠杆菌 DH10Bzip (Life Technologies, Gaithersburg, MD, USA)用于含 GH61B 基因的单个 pZL1 克隆的切除。

用 *Tsp* 509I 部分消化如实施例 2 所述制备的土生梭孢霉 NRRL 8126 基因组 DNA 并用 TAE 缓冲液在 1%琼脂糖凝胶上进行大小分级。切下在 3-7 kb 大小范围内迁移的 DNA 片段, 并使用 Prep-a-Gene 试剂(BioRad Laboratories, Hercules, CA, USA)将其从凝胶洗脱。将洗脱的 DNA 片段与 *Eco* RI-切割且去磷酸化的 λ ZipLox 载体臂(Life Technologies, Gaithersburg, MD, USA)连接在一起, 用商业包装提取物(Stratagene, La Jolla, CA, USA)将连接混合物包装。涂布包装的 DNA 文库并在大肠杆菌 Y1090ZL 细胞中扩增。未扩增的基因组 DNA 文库含 3.1×10^6 pfu/ml (无 DNA 的背景效价为 2.0×10^4 pfu/ml)。

实施例 6: 鉴定土生梭孢霉 NRRL 8126 *cel7c* 和 *cel7e* 克隆

如下所示, 利用与 TOPO 载体同源的引物和 Herculase® DNA 聚合酶 (Stratagene, La Jolla, CA, USA)分别从 pPH32 和 pPH37 扩增土生梭孢霉 *cel7c* 和 *cel7e* 基因探针片段。

5'-CTTGGTACCGAGCTCGGATCCACTA-3' (SEQ ID NO: 11)

5'-ATAGGGCGAATTGGGCCCTCTAGAT-3' (SEQ ID NO: 12)

PCR 反应物中每种引物使用 50 皮摩尔, 所述反应物在 50 μ l 的最终体积中含 10 ng pPH32 或 pPH37、1X Herculase®扩增缓冲液(Stratagene, La Jolla, CA, USA)、1 μ l 10 mM dATP、dTTP、dGTP 和 dCTP 的混合物、以及 2.5 单位的 Herculase® DNA 聚合酶。在 Stratagene Robocycler®中进行扩增, 程序为: 94°C 1 分钟的 1 个循环; 20 个循环, 每个循环 94°C 30 秒、55°C 30 秒和 72°C 1 分钟。然后加热块进入 4°C 的保温循环(soak cycle)。用 TAE 缓冲液在 1.0%琼脂糖凝胶上分离反应产物, 其中从凝胶切下三个<400 bp 的产物条带, 根据制造商的说明用 QIAquick®凝胶提取试剂盒(QIAGEN Inc., Valencia, CA, USA)将其纯化。利用 Prime-It® II 试剂盒(Stratagene, La Jolla, CA, USA)以 32 P 放射性标记 25 ng 的每个片段。

使用两个标记的 PCR 片段作为探针, 通过噬菌斑杂交从实施例 5 所述的文库筛选大约 90,000 个噬菌斑。利用 UV Stratalinker (Stratagene, La Jolla, CA, USA)将 DNA 交联到膜(Hybond N+, Amersham, Arlington Heights, IL,

USA)上。添加氢氧化钠至终浓度 0.1 M 使每个 ^{32}P 放射性标记的基因片段变性, 添加到活性为约 1×10^6 cpm 每 ml 杂交溶液的含 6X SSPE、7% SDS 的杂交溶液。每种混合物于 55°C 在振荡水浴中温育过夜。温育后, 将所述膜于 65°C 在含 0.1% SDS 的 0.2X SSC 中洗涤 3 次, 15 分钟。将所述膜在吸水纸上干燥 15 分钟, 包裹在 SaranWrap™ 中, 在 70°C 以增光屏(Kodak, Rochester, NY, USA)暴露于 X 射线片过夜。

基于与上述探针产生的强杂交信号, 选择几个噬菌斑用于进一步研究。噬菌斑在大肠杆菌 Y1090ZL 细胞中纯化两次, 随后通过大肠杆菌 DH10BZL 细胞(Life Technologies, Gaithersburg, MD, USA)的感染用体内切除法从 λ ZipLox 载体切除插入的基因和 pZL1 质粒作为 pZL1-衍生物(D'Alessio 等, 1992, Focus® 14:76)。将菌落接种到添加有 50 μg 氨苄青霉素每 ml 的 3 ml LB 培养基, 并在 37°C 培养过夜。用 BioRobot 9600 (QIAGEN Inc., Valencia, CA, USA)从每种这些培养物制备 Miniprep DNA。通过 DNA 测序显示命名为 pPH50 的克隆(图 3)包含 *cel7c* 的全长基因组基因, 通过 DNA 测序显示命名为 pPH38 的克隆(图 4)包含 *cel7e* 的全长基因。

包含质粒 pPH50 (图 3)的大肠杆菌 PaHa50 和包含质粒 pPH38 (图 4)的大肠杆菌 PaHa38 保藏在农业研究机构专利培养物保藏中心(Agricultural Research Service Patent Culture Collection)北区研究中心, 1815 University Street, Peoria, Illinois, 61604, 分别为 NRRL B-30899 和 NRRL B-30896, 保藏日期为 2006 年 2 月 23 日。

实施例 7: 表征编码 CEL7C 和 CEL7E 内切葡聚糖酶的土生梭孢霉基因组序列

利用 3.1 版 BigDye™终止子化学和 dGTP 化学(Applied Biosystems, Inc., Foster City, CA, USA)和引物步移(primer walking)策略, 用 Applied Biosystems 3700 型自动 DNA 测序仪对土生梭孢霉 *cel7c* 和 *cel7e* 基因组克隆进行 DNA 测序。仔细检查核苷酸序列数据质量, 在 PHRED/PHRAP 软件(University of Washington, Seattle, WA, USA)的辅助下, 将所有序列相互比较。

根据与来自里氏木霉(登录号 Q5BMS5)和尖镰孢(登录号 P46237)的同源基因的相似性, 构建了土生梭孢霉 *cel7c* 和 *cel7e* 基因组 DNA 序列的基因模型。

土生梭孢霉 *cel7C* 基因的核苷酸序列(SEQ ID NO: 1)和推导的氨基酸序列(SEQ ID NO: 2)如图 5A 和 5B 所示。编码序列为 1452 bp, 包括终止密码子, 被 57 bp 的内含子所打断。编码的预测蛋白为 464 个氨基酸。所述基因的编码序列的%G+C 为 63.7%, 成熟多肽编码序列的%G+C 为 63.8%。用 SignalP 程序(Nielsen 等, 1997, *Protein Engineering* 10 : 1-6)预测到 22 个残基的信号肽。预测的成熟蛋白包含 442 个氨基酸, 分子量为 46.5 kDa。用 InterProScan 程序(Zdobnov 和 Apweiler, 2001, *Bioinformatics* 17 : 847-848)分析 *cel7c* 基因推导的氨基酸序列, 显示 CEL7C 多肽含有糖苷水解酶家族 7 的结构域标记(domain signature) (InterPro 登录号 IPR001722), 其为成熟多肽的大约氨基酸 2 到 376。CEL7C 还包含真菌纤维素结合结构域的序列标记。发现称为 Prosite 模式 PS00562 (Sigrist 等, 2002, *Brief Bioinform.* 3 : 265-274)的该序列标记为成熟多肽的大约氨基酸 414 到 441。

使用 Needleman-Wunsch 算法(Needleman 和 Wunsch, 1970, *J. Mol. Biol.* 48: 443-453)如在具有缺口开放罚分 10、缺口延伸罚分 0.5 和 EBLOSUM62 矩阵的 EMBOSS 的 Needle 软件中所实施的, 来确定氨基酸序列的比较配对全局比对(comparative pairwise global alignment)。比对显示, 编码具有内切葡聚糖酶活性的 CEL7C 成熟多肽的土生梭孢霉基因的推导的氨基酸序列与来自烟曲霉和里氏木霉的两个糖基水解酶家族 7 蛋白的推导的氨基酸序列(登录号分别为 Q4WCM9 和 P07981)分别共享 76%和 70%同一性(缺口除外)。

土生梭孢霉 *cel7e* 基因的核苷酸序列(SEQ ID NO: 3)和推导的氨基酸序列(SEQ ID NO: 4)如图 6A 和 6B 所示。编码序列为 1336 bp, 包括终止密码子, 被 64 bp 的内含子所打断。编码的预测蛋白为 423 个氨基酸。所述基因的编码序列的%G+C 为 66.6%, 成熟多肽编码序列的%G+C 为 66.5%。利用 SignalP 程序(Nielsen 等, 1997, 同上)预测到了 19 个残基的信号肽。预测的成熟蛋白包含 404 个氨基酸, 分子量为 43.4 kDa。用 InterProScan 程序(Zdobnov 和 Apweiler, 2001, 同上)分析 *cel7e* 基因的推导的氨基酸序列, 显示 CEL7E 多肽含有糖苷水解酶家族 7 的结构域标记(InterPro 登录号 IPR001722), 其为成熟多肽的大约氨基酸 2 到 400。

使用 Needleman-Wunsch 算法(Needleman 和 Wunsch, 1970, *J. Mol. Biol.* 48: 443-453)如在具有缺口开放罚分 10、缺口延伸罚分 0.5 和 EBLOSUM62 矩阵的 EMBOSS 的 Needle 软件中所实施的, 来确定氨基酸序列的比较配对

全局比对。比对显示, 编码具有内切葡聚糖酶活性的 CEL7E 成熟多肽的土生梭孢霉基因的推导的氨基酸序列与来自粗糙脉孢菌、*Talaromyces emersonii* 和米曲霉(登录号分别为 Q7RXC7、CAC94521.1 和 O13455)的三个糖基水解酶家族 7 蛋白分别共享 65%、63%和 59%的同一性(缺口除外)。

实施例 8: pAILo2 表达载体的构建

通过修饰 pBANe6 (美国专利 6,461,837)构建表达载体 pAILo1, pBANe6 包含来自黑曲霉中性 α -淀粉酶和米曲霉磷酸丙糖异构酶的基因启动子的杂合体(NA2-tpi 启动子)、黑曲霉淀粉葡萄糖苷酶终止子序列(AMG 终止子)和构巢曲霉乙酰胺酶基因(*amdS*)。用 Big-Dye™终止子化学(Applied Biosystems, Inc., Foster City, CA, USA)通过测序验证所有诱变步骤。通过定点诱变, 首先从 *amdS* 选择标记去除位置 2051、2722 和 3397 bp 处的三个 *Nco* I 限制位点, 来进行 pBANe6 的修饰。将所有改变都设计为“沉默的”, 使 *amdS* 基因产物的实际蛋白质序列不变。利用下述引物(加下划线的核苷酸代表改变的碱基), 根据制造商的说明用 GeneEditor™体外定点诱变试剂盒(Promega, Madison, WI, USA)同时进行这三个位点的去除:

AMDS3NcoMut (2050):

5'-GTGCCCCATGATACGCCTCCGG-3' (SEQ ID NO: 13)

AMDS2NcoMut (2721):

5'-GAGTCGTATTTCCAAGGCTCCTGACC-3' (SEQ ID NO: 14)

AMDS1NcoMut (3396):

5'-GGAGGCCATGAAGTGGACCAACGG-3' (SEQ ID NO: 15)

然后用 QuickChange™定点诱变试剂盒(Stratagene, La Jolla, CA, USA)对包含所有三个预期序列改变的质粒进行定点诱变, 以去除在 AMG 终止子末端在位点 1643 的 *Nco* I 限制位点。下述引物(加下划线的核苷酸代表改变的碱基)用于诱变:

用于诱变 AMG 终止子序列的上游引物:

5'-CACCGTGAAAGCCATGCTCTTTTCCTTCGTGTAGAAGACCAGACAG-3' (SEQ ID NO: 16)

用于诱变 AMG 终止子序列的下游引物:

5'-CTGGTCTTCTACACGAAGGAAAGAGCATGGCTTTCACGGTGTCT

G-3' (SEQ ID NO: 17)

pBANE6 修饰的最后步骤是用 QuickChange™定点诱变试剂盒和以下引物(加下划线的核苷酸代表改变的碱基)在多接头(polylinker)的起始处添加新的 *Nco* I 限制位点, 以得到 pAILo1 (图 7)。

用于诱变 NA2-tpi 启动子的上游引物:

5'-CTATATACACAACCTGGATTTACCATTGGGCCCCGCGGCCGCAGATC-3' (SEQ ID NO: 18)

用于诱变 NA2-tpi 启动子的下游引物:

5'-GATCTGCGGCCGCGGGCCCATGGTAAATCCAGTTGTGTATATAG-3' (SEQ ID NO: 19)

将 pAILo1 的 *amdS* 基因与构巢曲霉 *pyrG* 基因交换(swap)。质粒 pBANE10 (图 8)用作作为选择标记的 *pyrG* 基因的来源。分析 pBANE10 的序列显示 *pyrG* 标记包含在 *Nsi* I 限制片段之内, 且不包含 *Nco* I 或 *Pac* I 限制位点。由于 *Nsi* I 限制位点也位于 *amdS* 的侧翼, 因而转换选择标记的策略是 *Nsi* I 限制片段的简单交换。用限制酶 *Nsi* I 消化来自 pAILo1 和 pBANE10 的质粒 DNA, 通过琼脂糖凝胶电泳纯化产物。将来自含 *pyrG* 基因的 pBANE10 的 *Nsi* I 片段连接到 pAILo1 的骨架以替换含 *amdS* 基因的原始的 *Nsi* I DNA 片段。通过限制酶消化分析重组克隆, 以确定它们具有正确插入序列及其方向。选择 *pyrG* 基因以逆时针方向转录的克隆。新质粒命名为 pAILo2 (图 9)。

实施例 9: 土生梭孢霉 NRRL 8126 *cel7c* 和 *cel7e* 基因的米曲霉表达载体的构建

设计了如下所示的两个合成的寡核苷酸引物, 用于从基因组克隆 PCR 扩增土生梭孢霉 NRRL 8126 *cel7c* 基因。用 InFusion 克隆试剂盒(BD Biosciences, Palo Alto, CA, USA)将所述片段直接克隆到表达载体 pAILo2 中, 而无需限制消化和连接。

正向引物:

5'-ACTGGATTACCATGGGCCAGAAGACGCTG-3' (SEQ ID NO: 20)

反向引物:

5'-AGTCACCTCTAGTTAGAGGCACTGGTAGTAC-3' (SEQ ID NO: 21)

粗体字母代表编码序列。其余序列与 pAILo2 的插入位点同源。

PCR 反应物中使用 50 皮摩尔的上述每种引物,所述反应物在 50 μ l 的最终体积中包含 100 ng 的土生梭孢霉基因组 DNA (如实施例 2 所述制备)、1X Pfx 扩增缓冲液、dATP、dTTP、dGTP 和 dCTP 的 10 mM 混合物 1.5 μ l、2.5 单位的 Platinum Pfx DNA 聚合酶(Invitrogen, Carlsbad, CA, USA)、1 μ l 的 50 mM MgSO₄ 和 5 μ l 的 10X pCRx 增强剂溶液(Invitrogen, Carlsbad, CA, USA)。在 Stratagene Robocycler®中进行扩增,程序为: 94°C 2 分钟的 1 个循环; 和 30 个循环, 每个循环 94°C 15 秒、55°C 30 秒和 68°C 3 分钟。然后加热模块进入 4°C 保温循环。

用 TAE 缓冲液在 1.0%琼脂糖凝胶上分离反应产物,从凝胶切下约 3 kb 的产物条带,根据制造商的说明用 QIAquick 凝胶提取试剂盒纯化。

然后用 InFusion 克隆试剂盒将所述片段克隆到 pAllo2 中。用 *Nco* I 和 *Pac* I 消化所述载体(使用制造商指定的条件)。通过用 TAE 缓冲液进行 1%琼脂糖凝胶电泳和 QIAquick 凝胶纯化试剂盒纯化所述片段。在反应中将所述基因片段和经消化的载体连接在一起,产生表达质粒 pEJG109 (图 10),其中 *cel7c* 基因的转录处于 NA2-tpi 启动子的控制之下。连接反应物(20 μ l)由如下组成: 1X InFusion 缓冲液(BD Biosciences, Palo Alto, CA, USA)、1X BSA (BD Biosciences, Palo Alto, CA, USA)、1 μ l InFusion 酶(1:10 稀释) (BD Biosciences, Palo Alto, CA)、用 *Nco* I 和 *Pac* I 消化的 100 ng 的 pAllo2 和 100 ng 土生梭孢霉 *cel7c* 纯化的 PCR 产物。反应物在室温下温育 30 分钟。用 1 μ l 所述反应物转化大肠杆菌 XL10 Solopac Gold 细胞(Stratagene, La Jolla, CA, USA)。通过用 *Xho* I 限制酶消化检测含 pEJG109 (*cel7c* 基因)的大肠杆菌转化体,并用 QIAGEN BioRobot 9600 制备质粒 DNA。

利用下述引物,以与上述相同的方式得到土生梭孢霉 *cel7e* 基因的表达构建体。

In-Fusion 正向引物:

5'-ACTGGATTACCATGGCGCCCAAGTCTACAGTTCTGG-3' (SEQ ID NO: 22)

In-Fusion 反向引物:

5'-TCACCTCTAGTTAATTAAGTAGTGGCTGCACTCGCTCT-3' (SEQ ID NO: 23)

粗体字母代表编码序列。其余序列与 pAllo2 的插入位点同源。

用 TAE 缓冲液和每 ml 0.1 μ g 的溴化乙锭在 0.8% GTG-琼脂糖凝胶 (Cambrex Bioproducts One Meadowlands Plaza East Rutherford, New Jersey 07073) 上分离 1.3 kb 的 PCR 反应产物。借助于 Dark Reader™ (Clare Chemical Research, Dolores, CO) 显示 DNA 条带, 以避免 UV-诱导的突变。根据制造商的说明, 用一次性刀片切下 1.3 kb 的 DNA 条带, 用 Ultrafree-DA 旋杯 (spin cup) (Millipore, Billerica, MA) 纯化。如上所述, 用 In-Fusion 克隆试剂盒 (BD Biosciences, Palo Alto, CA) 将纯化的 PCR 片段克隆入线性化且纯化的 pAILo2 载体中, 以产生 pAILo25 (图 11)。

通过限制酶消化分析鉴定含 pAILo25 的大肠杆菌转化体, 并用 QIAGEN BioRobot 9600 制备质粒 DNA。

实施例 10: 编码 CEL7C 和 CEL7E 内切葡聚糖酶的土生梭孢霉 NRRL 8126 基因在米曲霉 JaL250 中的表达

根据 Christensen 等, 1988, *Bio/Technology* 6: 1419-1422 的方法制备米曲霉 JaL250 原生质体。用 5 μ g 的 pEJG109 (同时将 pAILo2 作为载体对照) 转化米曲霉 JaL250 (WO 99/61651)。

用 pEJG109 (*cel7c* 基因) 转化米曲霉 JaL250 得到约 100 个转化体。将十个转化体分离至单独的 PDA 平板。

用 5 ml 0.01% Tween 80 洗涤 5 个转化体的汇合的 PDA 平板, 各别接种到 125 ml 玻璃摇瓶中 25 ml 的 MDU2BP 培养基中, 在 34°C、200 rpm 温育。温育 5 天后, 根据制造商的说明, 用 8-16% Tris-甘氨酸 SDS-PAGE 凝胶 (Invitrogen, Carlsbad, CA, USA) 分析来自每种培养物的 5 μ l 上清样品。培养物上清的 SDS-PAGE 谱显示在 5 个转化体中, 5 个都具有新的约 74 kDa 的主带。

用 10 ml 0.01% Tween 20 洗涤转化体 10 (培养在 PDA 上) 汇合的平板, 接种到含 500 ml MDU2BP 培养基的 2 升 Fernbach 中, 以产生发酵液用于酶的特征。在第 5 天收获烧瓶, 用 0.22 μ m GP Express plus 膜 (Millipore, Bedford, MA, USA) 过滤。

用如上所述的相同规程在米曲霉 JaL250 中表达质粒 pAILo25 (*cel7e* 基因)。转化米曲霉 JaL250 得到约 35 个转化体。将八个转化体分离至单独的 PDA 平板, 34°C 温育 5 天。培养物上清的 SDS-PAGE 谱显示, 八个 pAILo25 转化体中八个都具有新的约 55 kDa 的弥散蛋白质条带。选择#4 转化体用于进

一步研究，命名为米曲霉 JaL250 pAlLo25。

实施例 11: 编码 CEL7A 内切葡聚糖酶的 *Cladorrhinum foecundissimum* ATCC 62373 的克隆和表达

将 *Cladorrhinum foecundissimum* ATCC 62373 在 200 ml 含纤维素的 PD 培养基中在 30°C、200 rpm 培养五天。将内容物通过衬有 Miracloth™ 的漏斗过滤，以收获来自摇瓶培养的菌丝体。然后将菌丝体夹在两个 Miracloth™ 片之间，用吸水纸巾吸干。然后将菌丝团转移到 Falcon® 1059 塑料离心管(BD Biosciences, Palo Alto, CA, USA)，并冷冻在液氮中。将冷冻的菌丝体存储在 -80°C 冰箱中直至使用。

用硫氰酸胍(guanidinium thiocyanate)提取总 RNA，之后通过 5.7 M CsCl 缓冲(cushion)超速离心，并用 WO 94/14953 中所述的方法通过寡(dT)-纤维素亲和层析分离多聚(A)+RNA。

通过 RNase (核糖核酸酶) H 法(Gubler 和 Hoffman, 1983, *Gene* 25: 263-269, Sambrook 等, 1989, *Molecular cloning: A laboratory manual*, Cold Spring Harbor lab., Cold Spring Harbor, NY)由 5 µg 多聚(A)+RNA 合成双链 cDNA。将多聚(A)⁺ RNA (5 µl DEPC (0.1%焦碳酸二乙酯)-处理过的水中 5 µg) 在预硅化的无核糖核酸酶 Eppendorf®管中在 70°C 加热 8 分钟，在冰上淬火，与逆转录酶缓冲液组合在 50 µl 的最终体积中，所述逆转录酶缓冲液由下列组成：50 mM Tris-HCl, pH 8.3、75 mM KCl、3 mM MgCl₂、10 mM 二硫苏糖醇(DTT) (Bethesda Research Laboratories, Bethesda, MD, USA)、1 mM 的 dATP、dGTP 和 dTTP、0.5 mM 的 5-甲基-dCTP (Pharmacia, Uppsala, Sweden)、40 个单位的人胎盘核糖核酸酶抑制剂(RNasin, Promega, Madison, WI, USA)、1.45 µg 的寡(dT)₁₈-Not I 引物(Pharmacia, Uppsala, Sweden)和 1000 个单位的 SuperScript® II RNase H 逆转录酶(Bethesda Research Laboratories, Bethesda, MD, USA)。通过将反应混合物于 45°C 温育 1 小时来合成第一链 cDNA。合成后，根据制造商的说明通过 MicroSpin S-400 HR 旋转柱(Pharmacia, Uppsala, Sweden)凝胶过滤 mRNA:cDNA 杂合体混合物。

凝胶过滤后，将杂合体稀释于 250 µl 含 200 µM 的每种 dNTP、60 个单位的大肠杆菌 DNA 聚合酶 I (Pharmacia, Uppsala, Sweden)、5.25 个单位的 RNase H (Promega, Madison, WI, USA)和 15 个单位的大肠杆菌 DNA 连接酶

(Boehringer Mannheim, Mannheim, Germany)的第二链缓冲液(20 mM Tris-HCl, pH 7.4、90 mM KCl、4.6 mM MgCl₂、10 mM (NH₄)₂SO₄、0.16 mM NAD)。通过将反应管在 16°C 温育 2 小时,在 25°C 再温育 15 分钟来合成第二链 cDNA。添加 EDTA 至终浓度 20 mM 以终止反应,之后用苯酚和氯仿提取。

通过添加 2 体积的 96%乙醇和 0.2 体积的 10 M 乙酸铵以将双链 cDNA 在-20°C 沉淀 12 小时,通过在 13,000 $\times$ g 离心回收,在 70%乙醇中洗涤,干燥,并重悬于含 25 个单位的绿豆核酸酶(Pharmacia, Uppsala, Sweden)的 30 μ l 绿豆核酸酶缓冲液(30 mM 乙酸钠 pH 4.6、300 mM NaCl、1 mM ZnSO₄、0.35 mM DTT、2%甘油)。通过如下方法修剪单链发夹 DNA:将所述反应物于 30°C 温育 30 分钟,之后添加 70 μ l 的 10 mM Tris-HCl-1 mM EDTA pH 7.5,苯酚提取,并用 2 体积的 96%乙醇和 0.1 体积的 3 M 乙酸钠 pH 5.2 在冰上沉淀 30 分钟。

通过在 13,000 $\times$ g 离心回收双链 cDNA,并通过将所述反应混合物于 16°C 温育 1 小时以将其在含 0.5 mM 的每种 dNTP 和 5 个单位的 T4 DNA 聚合酶(New England Biolabs, Ipswich, MA, USA)的 30 μ l T4 DNA 聚合酶缓冲液(20 mM Tris-乙酸, pH 7.9、10 mM 乙酸镁、50 mM 乙酸钾、1 mM DTT)中末端钝化。添加 EDTA 至终浓度 20 mM 以终止反应,之后进行苯酚和氯仿提取,并通过添加 2 体积的 96%乙醇和 0.1 体积的 3 M 乙酸钠 pH 5.2 在-20°C 沉淀 12 小时。

所述填充反应之后,通过在 13,000 $\times$ g 离心回收 cDNA,在 70%乙醇中洗涤,并干燥。将 cDNA 沉淀重悬在含 2.5 μ g 如下所示的非回文(non-palindromic) *Bst* XI 接头(Invitrogen, Carlsbad, CA, USA)和 30 个单位的 T4 连接酶(Promega, Madison, WI, USA)的 25 μ l 连接缓冲液(30 mM Tris-HCl, pH 7.8、10 mM MgCl₂、10 mM DTT、0.5 mM ATP)中,然后 16°C 温育 12 小时。通过在 65°C 加热 20 分钟以终止反应,然后在冰上冷却 5 分钟。

5'-CTTTCCAGCACA-3' (SEQ ID NO: 24)

3'-GAAAGGTC-5'

用 *Not* I 消化修饰的(adapted) cDNA,之后在 37°C 温育 2.5 小时。通过在 65°C 加热 10 分钟以终止反应。通过如下方法对 cDNA 进行大小分级:在 44 mM Tris 碱、44 mM 硼酸、0.5 mM EDTA (TBE)缓冲液中,在 0.8% SeaPlaque® GTG 低溶解温度琼脂糖凝胶(Cambrex Corporation, East

Rutherford, NJ, USA)上进行凝胶电泳以分离未连接的衔接头和小 cDNA。以 0.7 kb 为截止点(cut-off)根据大小选择(size-selected) cDNA, 根据制造商的说明利用 β -琼脂糖酶(New England Biolabs, Ipswich, MA, USA Biolabs)从凝胶挽救(rescue)所述 cDNA, 并通过添加两个体积的 96%乙醇和 0.1 体积的 3M 乙酸钠 pH 5.2 将其在 -20°C 沉淀 12 小时。

通过在 13,000 $\times g$ 离心回收根据方向、大小选择的 cDNA, 在 70%乙醇中洗涤, 干燥, 然后重悬在 30 μ l 的 10 mM Tris-HCl-1 mM EDTA pH 7.5 中。根据制造商的说明, 通过经 MicroSpin S-300 HR 旋转柱凝胶过滤将所述 cDNAs 脱盐。在含有 5 μ l 双链 cDNA (#1 和 #2 反应管)、15 个单位的 T4 连接酶、30 ng (#1 管)、40 ng (#2 管)、和 40 ng (#3 管, 载体背景对照) *Bst* XI-*Not* I 切割的 pYES2.0 载体(Invitrogen, Carlsbad, CA, USA)的 10 μ l 连接缓冲液(30 mM Tris-HCl, pH 7.8、10 mM $MgCl_2$ 、10 mM DTT、0.5 mM ATP)中进行三个检验连接。通过在 16°C 温育 12 小时, 然后在 70°C 加热 20 分钟, 最后向每管添加 10 μ l 水来进行连接反应。如 Sambrook 等, 1989, 同上所述, 将 1 μ l 的每种连接混合物电穿孔到 40 μ l 电感受态(electrocompetent)大肠杆菌 DH10B 细胞(Bethesda Research Laboratories, Bethesda, MD, USA)中。

将 *Cladorrhinum foecundissimum* ATCC 62373 cDNA 文库建立为大肠杆菌 DH10B 中的汇集物(pool)。通过在添加有 100 μ g 氨苄青霉素每 ml 的 LB 平板上涂布转化的大肠杆菌来制备每个汇集物, 在 37°C 温育 24 小时后得到 15,000-30,000 菌落/平板。向平板添加 20 ml 补充有 100 μ g 氨苄青霉素每 ml 的 LB 培养基, 并将细胞悬浮于其中。将细胞悬液在 37°C 在 50 ml 管中在 100 rpm 摇动 1 小时。

所得 *Cladorrhinum foecundissimum* ATCC 62373 cDNA 文库由约 10^6 个单一克隆组成, 具有 1%的载体背景。根据制造商的说明, 利用 Plasmid Midi 试剂盒(QIAGEN Inc., Valencia, CA, USA)从一些文库汇集物分离质粒 DNA, 并储存于 -20°C。

通过电穿孔(Becker 和 Guarante, 1991, *Methods Enzymol.* 194: 182-187)将来自一些文库汇集物(实施例 1)的 1 ml 等份纯化质粒 DNA (100 ng/ml)转化到酿酒酵母 W3124 中, 将转化体涂布在含 2%葡萄糖的 SC 琼脂上并在 30°C 温育。共计从每个汇集物获得含 250-400 酵母菌落的 50-100 个平板。

温育 3-5 天后, 将所述 SC 琼脂平板复制涂布(replica plate)到一组含半乳

糖的 0.1% AZCL HE 纤维素 SC URA 琼脂平板上。将平板在 30°C 温育 2-4 天, 将由蓝色晕圈环绕的菌落鉴定为内切葡聚糖酶阳性菌落。

将表达内切葡聚糖酶的酵母菌落接种到 50 ml 玻璃试管中的 20 ml YPD 培养基中。将试管在 30°C 在 200 rpm 摇动 2 天。通过在具有 75002252 转子的 Heraeus Megafuge 1.0R 离心机(Hanau, Germany)中在 3000 rpm 离心 10 分钟来收获细胞。

依照 WO 94/14953 分离 DNA, 并将其溶解在 50 μ l 去离子水中。通过根据 Sambrook 等, 1989, 同上的标准方法将 DNA 转化到大肠杆菌 DH10B 细胞中。随后显示一个大肠杆菌转化体含有 *Cladorrhinum foecundissimum* ATCC 62373 *cel7a* 基因, 该菌株中的质粒命名为 pCIC521。

实施例 12: *Cladorrhinum foecundissimum* ATCC 62373 *cel7a* 基因在米曲霉中的表达

利用标准方法(Sambrook 等, 1989, 同上)用 *Bam* HI 和 *Xba* I 从 pCIC521 切下 *Cladorrhinum foecundissimum* ATCC 62373 *cel7a* 基因, 并连接到曲霉属表达载体 pHD464 (Dalboge 和 Heldt-Hansen, 1994, *Molecular and General Genetics* 243 : 253-260)中。所得质粒命名为 pA2CIC521。

如 WO 95/02043 所述制备米曲霉 HowB104 的原生质体。将 100 μ l 的原生质体悬浮液与 5-25 μ g 的曲霉属表达载体 pA2CIC521 在 10 μ l STC 中混合, STC 由 1.2 M 山梨醇、10 mM Tris-HCl, pH 7.5、10 mM CaCl_2 组成, 并进一步与 5-25 μ g 携带构巢曲霉 *amdS* 基因的质粒 p3SR2 (Christensen 等, 1988, *Bio/Technology* 6 : 1419-1422)混合。将混合物在室温下放置 25 分钟。添加 200 μ l 60% PEG 4000 (BDH, Poole, England) (聚乙二醇,分子量 4,000)、10 mM CaCl_2 和 10 mM Tris-HCl pH 7.5, 轻轻混合, 最后添加 0.85 ml 的相同溶液, 轻轻混合。将所述混合物在室温下放置 25 分钟, 在 2,500 x g 离心 15 分钟, 将沉淀重悬在 2 ml 的 1.2 M 山梨醇中。重复该沉积过程, 将原生质体涂布在 COVE 平板上。在 37°C 温育 4-7 天后, 挑选并涂布孢子以分离单菌落。重复该过程, 在第二次再分离后储存单菌落的孢子。

将每个转化体接种在 10 ml 的 YPM 培养基中。在 30°C、200 rpm 温育 2-5 天后去除上清。通过将 20 μ l 培养液添加到在 0.1% AZCL HE 纤维素 SC-琼脂平板中穿出的 4 mm 直径孔, 并在 30°C 温育过夜, 来鉴定内切葡聚糖酶

活性。内切葡聚糖酶活性的存在在菌落周围产生蓝色晕圈。几个转化体培养液具有的内切葡聚糖酶活性明显大于来自未转化的米曲霉背景对照的培养液，这显示来自 *Cladorrhinum foecundissimum* ATCC 62373 的 CEL7A 内切葡聚糖酶在米曲霉中有效表达。将一个米曲霉转化体命名为 A2.11C521 4/3。

实施例 13: 米曲霉 A2.11C521 4/3 基因组 DNA 提取

用带挡板的摇瓶将表达 *Cladorrhinum foecundissimum* 的 CEL7A 蛋白的米曲霉菌株 A2.11C521 4/3 在 300 ml 培养体积中的 M400 培养基中在 34°C 200 rpm 培养 7 天。将生物物质在液氮中冷冻并用研钵和杵粉碎为粉末。将粉末悬浮在 15 ml 的 0.1 M CAPS-NaOH pH 11.0、1 mM EDTA、0.5% 十二烷基硫酸锂中，通过倒转以周期性混合在 60°C 温育 60 分钟。添加等体积的中和苯酚，并将管于 37°C 轻轻摇动 1 hr。添加 5 ml 的氯仿，并用力搅动试管 1 分钟。在 1300 x g 离心 10 分钟后，用等体积苯酚:氯仿(1:1)通过搅动 5 分钟来再次提取上部水相。重复离心，将水相转到 2.5 M 乙酸铵，并在 -20°C 储存 20 分钟。在 4°C 17,000 x g 离心 20 分钟后，通过添加 0.7 体积的异丙醇来沉淀上清中的上清核酸。在 17,000 x g 离心 10 分钟后，倾析上清并用 70% 乙醇漂洗沉淀，风干。将沉淀溶解在 950 µl 去离子水中，之后添加 50 µl 的 Promega 细胞重悬液(Promega Corporation, Madison WI, USA)，室温下温育 5 分钟。添加乙酸铵至 1.0 M，并通过添加 2 体积的乙醇沉淀核酸。在 13,000 x g 离心 10 分钟后，将沉淀溶解在 300 µl 的 1 mM Tris-HCl、0.1 mM EDTA, pH 8.0 中，储存于 -20°C。

实施例 14: 从米曲霉表达菌株 A2.11C521 4/3 的基因组 DNA PCR 扩增 *Cladorrhinum foecundissimum cel7a* cDNA

为了测序和克隆保藏物，从米曲霉菌株 A2.11C521 4/3 基因组 DNA 扩增 *cel7a* cDNA。设计了与 pHD464 表达载体同源的两个合成寡核苷酸引物，用于从米曲霉菌株 A2.11C521 4/3 基因组 DNA PCR 扩增 *cel7a* 基因。

正向引物:

5'-CCACACTTCTCTTCCTTCCTC-3' (SEQ ID NO: 25)

反向引物:

5'-CCCCATCCTTTAACTATAGCG-3' (SEQ ID NO: 26)

PCR 反应物中使用 17 皮摩尔的上述每种引物,所述反应物在 50 μ l 的最终体积中包含 100 ng 的米曲霉 A2.11C521 4/3 基因组 DNA (如实施例 13 所述制备)、1X Pfx 扩增缓冲液(Invitrogen, Carlsbad CA, USA)、1.7 μ l 的 dATP、dTTP、dGTP 和 dCTP 的 10 mM 混合物、2.5 个单位的 Platinum Pfx DNA 聚合酶和 1 μ l 的 50 mM MgSO₄。在 Stratagene Robocycler®中进行扩增,程序编制为 96°C 3 分钟和 72°C 3 分钟(在此期间添加 DNA 聚合酶)的 1 个循环;和 35 个循环,每个循环 94°C 50 秒、5°C 50 秒、68°C 2 分钟;之后为在 68°C 最终延伸 7 分钟。

用 TAE 缓冲液在 1.0%琼脂糖凝胶上分离反应产物,从凝胶切下约 1.7 kb 的产物条带,根据制造商的说明用 QIAEX® II 凝胶提取试剂盒纯化。

用 Zero Blunt TOPO PCR 克隆试剂盒将 1.7 kb 片段克隆到 pCR-Blunt II-TOPO 载体中,并根据制造商的说明(Invitrogen, Carlsbad CA, USA)转化到大肠杆菌 TOP10 细胞中。用 QIAGEN BioRobot 9600 从几个转化体制备质粒 DNA。对来自两个转化体的质粒测序,发现是相同的且包含 *cel7a* cDNA。将一个质粒命名为 pPH46 (图 12)。

将含有质粒 pPH46 (图 12)的大肠杆菌 PaHa46 保藏于农业研究机构专利培养物保藏中心北区研究中心,1815 University Street, Peoria, Illinois, 61604, 保藏号为 NRRL B-30897, 保藏日期为 2006 年 2 月 23 日。

实施例 15: 编码 CEL7A 内切葡聚糖酶的 *Cladorrhinum foecundissimum* ATCC 62373 cDNA 序列的表征

利用 3.1 版 BigDye™终止子化学和 dGTP 化学和引物步移策略,用 3700 型 Applied Biosystems 自动 DNA 测序仪对质粒 pPH46 中的 *Cladorrhinum foecundissimum* ATCC 62373 *cel7a* cDNA 进行 DNA 测序。仔细检查核苷酸序列数据质量,在 PHRED/PHRAP 软件(University of Washington, Seattle, WA, USA)的辅助下,将所有序列相互比较。

Cladorrhinum foecundissimum cel7a cDNA 的核苷酸序列(SEQ ID NO: 5)和推导的氨基酸序列(SEQ ID NO: 6)如图 13A 和 13B 所示。所述编码序列为 1323 bp, 包含终止密码子。编码的预测蛋白为 440 个氨基酸。所述基因的编码序列的 %G+C 为 59.2%,成熟多肽编码序列的 %G+C 为 59.0%。利用 SignalP 程序(Nielsen 等, 1897, 同上)预测了 18 个残基的信号肽。预测的成熟蛋白包含

422 个氨基酸, 分子量为 45.7 kDa。用 InterProScan 程序(Zdobnov 和 Apweiler, 2001, 同上)分析 *cel7a* 基因的推导的氨基酸序列, 显示 CEL7A 多肽含有糖苷水解酶家族 7 的结构域标签(InterPro 登录号 IPR001722), 其为成熟多肽的大约氨基酸 2 到 405。

使用 Needleman-Wunsch 算法(Needleman 和 Wunsch, 1970, *J. Mol. Biol.* 48: 443-453)如在具有缺口开放罚分 10、缺口延伸罚分 0.5 和 EBLOSUM62 矩阵的 EMBOSS 的 Needle 软件中所实施的, 来确定氨基酸序列的比较配对全局比对。比对显示, 编码具有内切葡聚糖酶活性的 CEL7A 成熟多肽的 *Cladorrhinum foecundissimum* cDNA 的推导的氨基酸序列与来自粗糙链孢菌(登录号为 Q7RXC7)和 *Talaromyces emersonii* (登录号为 CAC94521.1)的两个糖基水解酶家族 7 蛋白的推导的氨基酸序列分别共享 79%和 55%的同一性(缺口除外)。其也分别与土生梭孢霉 CEL7E 和 CEL7C 成熟多肽共享 64%和 57%同一性。

如实施例 1 所述, 通过 10-20% Tris-Glycine SDS-PAGE 分离含有 *Cladorrhinum foecundissimum* 家族 7A 的培养液。如实施例 1 中所述, 利用串联质谱法将约 50 kDa 的凝胶条带凝胶内消化并从头测序。

将来自 701.85 m/z 序列的双倍荷电的胰蛋白肽离子的部分序列测定为 Gly-Thr-Val-Val-Pro-Glu-Ser-His-Pro-Lys, 其对应于序列 GTVVPE SHPK (SEQ ID NO: 6 的氨基酸 22 至 31)。将来自第二个 739.35 的双倍荷电的胰蛋白肽离子的部分序列测定为 Gly-Glu-Ala-Asn-[Ile 或 Leu]-Asp-[Gln 或 Lys]-Lys, 其中 Xaa 为任何的氨基酸, 其对应于序列 GEANIDQK (SEQ ID NO: 6 的氨基酸 202 至 209)。将第三个 919.83 的双倍荷电的胰蛋白肽离子测定为 [CAM (羧基氨基甲基半胱氨酸)或 Phe]-Glu-Gly-Glu-Asp-Glu-[Cam 或 Phe]-Gly-[Gln 或 Lys]-Pro-Val-Gly-Val-[Cam 或 Phe]-Asp-Lys, 其对应于序列 CEGEDECGQPVGVC DK (SEQ ID NO: 6 的氨基酸 242 至 257)。将第四个 970.45 的双倍荷电的胰蛋白肽离子部分序列测定为 Pro-Ser-Thr-Pro-Cam-Val-Val-Gly-Gly-Pro-[Ile 或 Leu]-Cam-Pro-Asp-Ala-Lys, 其对应于序列 PSTPCVVGGPLCPDAK (SEQ ID NO: 6 的氨基酸 65 至 80)。

实施例 16: 从 *Cladorrhinum foecundissimum* ATCC 62373 和土生梭孢霉 NRRL 8126 制备重组内切葡聚糖酶

用基本上如 Otzen 等, 1999, Protein Sci. 8 : 1878-87 所述的规程, 将如实施例 12 所述在米曲霉中重组产生的 *Cladorrhinum foecundissimum* CEL7 内切葡聚糖酶纯化至均质。

以来自宿主米曲霉表达的培养液的形式检测如实施例 10 所述在米曲霉中重组产生的土生梭孢霉 CEL7C 和土生梭孢霉 CEL7E 内切葡聚糖酶。用带有 Biomax-5 聚醚砜膜(5000 NMWL)的 Centricon Plus-20 离心滤器(Millipore, Bedford, MA, USA)浓缩培养液并交换至 50 mM 乙酸钠缓冲液 pH 5.0。

根据制造商对 BCA 蛋白测试试剂盒(Pierce Chemical Co., Rockford, IL, USA)的说明, 用二金鸡纳酸(Bicinchoninic Acid) (BCA)微板测试测定酶制备物中的蛋白质浓度。

在每个试验之前从储存于-20°C 的储备酶溶液新鲜制备酶稀释液。

实施例 17: 底物的制备

由美国能源部可再生能源国家实验室(NREL, Golden, CO, USA)利用稀硫酸制备预处理的玉米秸(PCS)。使用下述条件进行预处理: 1.4 wt %硫酸、在 165°C 和 10 psi 进行 8 分钟。在 NREL 进行组成分析。用 NREL 标准分析方法#002 通过二阶段硫酸水解及随后通过高效液相层析(HPLC)分析糖来测定纤维素和半纤维素。用硫酸水解纤维素和半纤维素级分后, 以重量分析法测定木质素(NREL 标准分析方法#003)。将预处理的玉米秸(PCS)中不溶于水的固体测定为 56.5%纤维素、4.6%半纤维素和 28.4%木质素。

用粗孔隙玻璃滤器(Fisher Scientific, Pittsburg, PA, USA)在 Kimax 漏斗上用大量去离子水洗涤 PCS。在咖啡磨碎机中磨碎水洗涤过的 PCS, 在具有 6P Express Membrane 的 22 μ m Millipore Filter (Millipore, Bedford, MA, USA)上再次用去离子水洗涤。磨碎的 PCS 的干重为 32.4%。

用下述方法制备溶于去离子水中的 10 mg/ml 磷酸溶胀纤维素(PASC)储存悬浮液。将 150 ml 冰冷的 85% *o*-磷酸添加到用水润湿的 5 g Avicel (艾维素) PH101 (FMC Corp., Philadelphia, PA, USA)。悬浮液在冰浴中缓慢搅动一小时, 在恒定搅动时将 100 ml 冰冷的丙酮添加到所述悬浮液。把浆料转到具有粗孔隙玻璃滤器的 Kimax 漏斗, 用 100 ml 冰冷的丙酮洗涤三次, 在每次洗涤后尽可能完全排干。最后用 500 ml 水将所述浆料洗涤两次, 每次洗涤后同样尽可能完全排干。将 PASC 与水混合至总体积 500 ml。添加叠氮化钠至终

浓度 0.02% 以阻止微生物生长。用搅拌器将浆料均质化，并在 4°C 储存多至一个月。

从 Hercules Inc., Wilmington, DE, USA 的 Aqualon Division 获得平均取代度(DS)为 0.7 的羧甲基纤维素(CMC, 钠盐, 7L2 型)。通过将 CMC 缓慢加到用力搅动的缓冲液来制备溶于 50 mM 乙酸钠 pH 5.0 中的 6.25 mg/ml CMC 溶液。在持续搅动下把浆料加热到约 60°C，直至 CMC 完全溶解。

从 Sigma, St. Louis, MO, USA 获得对硝基苯基- β -D-纤维二糖糖苷(PNPC)和对硝基苯基- β -D-乳糖苷(PNPL)。

如 Boisset 等, 1999, *Biochemical Journal*, 340 : 829-835 中所述，由食品级商业纤维素 Nata de Coco (Fujicco Co., Kobe, Japan) 制备细菌纤维素(BC)。在 4°C 保存溶于含 0.01% (w/v) 叠氮钠的去离子水中的 1 mg/ml 细菌纤维素悬浮液。

Avicel PH101 从 FMC Corporation, Philadelphia, PA, USA 获得。

来自桦木的木聚糖从 Sigma, St. Louis, MO 获得。来自罗望子种子的木葡聚糖(淀粉状, lot 00401)、小麦阿拉伯木聚糖(中粘度, 27 cSt, lot 90601)、1,4- β -D-甘露聚糖(氢硼化物还原的, Man:Gal=97:3, 聚合度 DP~15, lot 90302)、和角豆(carob)半乳甘露聚糖(低粘度, 氢硼化物还原的, lot 90301)从 Megazyme, Bray, Ireland 获得。

实施例 18: 用于测定还原糖的对羟苯甲酸酞肼测试

通过对羟苯甲酸酞肼(PHBAH)测试(Lever, 1972, *Anal. Biochem.* 47, 273-279)测定还原糖(RS)，该测试方法适合于 96-孔微板阵列。

将稀释样品的 90- μ l 等分试样置于 96-孔锥底微板(Costar, 透明聚碳酸酯, Corning Inc., Acton, MA)的每个孔中。通过将 60 μ l 溶于 2% 氢氧化钠中的 1.25% PHBAH 添加到每个孔中来启动所述测试。未覆盖的平板于 95°C 在定制的加热块上加热 10 分钟。加热之后将微板冷却至室温，向每个孔添加 35 μ l 去离子水。从每个孔取出 100 μ l 等分试样，并转移到平底 96-孔平板(Costar, 中等结合聚苯乙烯, Corning Inc., Acton, MA)。用 SpectraMAX 微板读数器(Molecular Devices, Sunnyvale, CA)测量 410 nm 处的吸光度(A_{410})。用标准曲线将 A_{410} 值折算为葡萄糖当量。

用六个葡萄糖标准(0.005、0.010、0.025、0.050、0.075 和 0.100 mg/ml)

获得所述标准曲线，对其进行与样品类似的处理。通过用去离子水稀释 10 mg/ml 储备葡萄糖溶液来制备葡萄糖标准。

对于除木聚糖和阿拉伯木聚糖之外的所有底物，利用下述等式计算转化的程度(%):

$$\begin{aligned}\text{转化率}(\%) &= \text{RS}_{(\text{mg/ml})} \times 100 \times 162 / (\text{起始底物浓度}_{(\text{mg/ml})} \times 180) \\ &= \text{RS}_{(\text{mg/ml})} \times 100 / (\text{起始底物浓度}_{(\text{mg/ml})} \times 1.111)\end{aligned}$$

对于木聚糖和阿拉伯木聚糖，用下述等式计算底物水解为 RS 的百分数:

$$\begin{aligned}\text{转化率}(\%) &= \text{RS}_{(\text{mg/ml})} \times 100 \times 132 / (\text{起始底物浓度}_{(\text{mg/ml})} \times 150) \\ &= \text{RS}_{(\text{mg/ml})} \times 100 / (\text{起始底物浓度}_{(\text{mg/ml})} \times 1.136)\end{aligned}$$

在这些等式中，RS 是以葡萄糖当量测量的溶液中还原糖的浓度(mg/ml)，因数 1.111 和 1.136 反映在相应的多糖至己糖(MW 180)或戊糖(MW 150)糖转化中的重量增加。

实施例 19: *Cladorrhinum foecundissimum* ATCC 62373 CEL7A 内切葡聚糖酶对不同底物的相对活性

Cladorrhinum foecundissimum ATCC 62373 CEL7A 内切葡聚糖酶对四种底物的相对活性示于表 1。将相对活性显示为 *Cladorrhinum foecundissimum* CEL7A 内切葡聚糖酶对羧甲基纤维素(CMC)的活性的百分数。通过测量产物浓度随时间线性增加范围内水解的初始速率来测定活性。稀释 *Cladorrhinum foecundissimum* CEL7A 内切葡聚糖酶以给出酶浓度与测得的活性之间的线性关系。

通过测量在 50°C 在 50 mM 乙酸钠 pH 5.0 中水解 30 分钟后自 CMC (5 mg/ml)产生的还原糖(RS)的浓度来测定 *Cladorrhinum foecundissimum* CEL7A 内切葡聚糖酶对于羧甲基纤维素(CMC)的可溶性钠盐的活性。在存在 0.5 mg/ml 牛血清白蛋白(BSA)且不搅动的情况下进行水解。利用实施例 18 中描述的对羧苯甲酸酞肼(PHBAH)测试方法测定还原糖。

通过测量在 50°C、50 mM 乙酸钠 pH 5.0 中磷酸溶胀纤维素(PASC) (2 mg/ml)的初始水解期间释放的还原糖(RS)的浓度来测定 *Cladorrhinum foecundissimum* CEL7A 内切葡聚糖酶对 PASC 的活性。在存在 0.5 mg/ml BSA 且不搅动的情况下进行水解。将酶稀释以使在水解的最初 30 至 90 分钟期间 RS 浓度会线性增加，在此期间 PASC 转化的程度将不超过 2%。利用实施例

18 中描述的对羟苯甲酸酰肼(PHBAH)测试方法测定还原糖。

利用修饰为 96-孔微板形式的 Deshpande 等的方法(Deshpande 等, 1984, Anal. Biochem. 138, 481-487)测定内切葡聚糖酶对于产色底物对硝基苯基- β -D-纤维二糖糖苷(PNPC)和对硝基苯基- β -D-乳糖苷 (PNPL)的活性。PNPC (2.5 mM)或 PNPL (2.5 mM)在 40°C 在 50 mM 乙酸钠 pH 5.0 中水解 30-分钟后, 通过分光光度测量于 405 nm 处测定对硝基苯酚(PNP)。水解前, 将酶稀释在 50 mM 乙酸钠 pH 5.0 中, 在指定条件下两种底物均得到小于 8%的转化率。

表 1. 在 pH 5.0 和 50°C *Cladorrhinum foecundissimum* CEL7A 纤维素酶的相对活性

底物	相对活性, %
羧甲基纤维素(CMC)	100
磷酸溶胀纤维素(PASC)	14.5
对硝基苯基- β -D-纤维二糖糖苷(PNPC)	0.58
对硝基苯基- β -D-乳糖苷(PNPL)	0.77

实施例 20: *Cladorrhinum foecundissimum* ATCC 62373 CEL7A 内切葡聚糖酶的热稳定性

通过将酶溶液在五个温度(40°C、50°C、60°C、70°C 和 80°C)温育并测量对于羧甲基纤维素(CMC)的残余活性, 从而测定纯化的 *Cladorrhinum foecundissimum* CEL7 内切葡聚糖酶的热稳定性。

将酶在含 3.0 mg/ml BSA 的 50 mM 乙酸钠 pH 5.0 中稀释, 并在以 8x12 微板形式(Pierce, Rockford, IL)放置的 1.1-ml ImmunoWare 微型试管(Microtubes)中温育 3 小时。添加 BSA 以防止酶可能吸附在微型试管的塑料壁上。选择温育混合物中的蛋白质浓度使得在随后的 CMCase 活性测试中获得小于 1%的 CMC 转化率。

温育 3 小时后, 用 8-通道移液管取出 15 μ l 的等份试样, 添加到 96-孔锥底微板(Costar, 透明聚碳酸酯, Corning Inc., Acton, MA)中的 75 μ l CMC 溶液(在 50 mM 乙酸钠 pH 5.0 中 6 mg/ml)。然后如实施例 19 所述测量残留的 CMCase 活性, 并将其表示为初始 CMCase 活性的百分数(表 2)。

温育 3 小时后, *Cladorrhinum foecundissimum* CEL7A 内切葡聚糖酶在 40°C 和 50°C 保留了超过 90%的初始 CMCase 活性, 在 60°C 保留了初始

CMCase 活性的三分之一, 在 70°C 和 80°C 保留了初始 CMC-ase 活性的 0%。

表 2. 在 pH 5.0 温育三个小时后 *Cladorrhinum foecundissimum* CEL7 内切葡聚糖酶的残留 CMCase 活性

温度, °C	残留 CMCase 活性, 初始活性的%
40°C	95.0
50°C	92.9
60°C	29.2
70°C	0.0
80°C	0.0

实施例 21: *Cladorrhinum foecundissimum* ATCC 62373 和土生梭孢霉 NRRL 8126 内切葡聚糖酶对不同多糖底物的表征

在 pH 5.0 (50 mM 乙酸钠缓冲液)和 50°C, 在不同多糖的水解中评估 *Cladorrhinum foecundissimum* ATCC 62373 CEL7 和土生梭孢霉 NRR 8126 CEL7C 内切葡聚糖酶。将结果与重组里氏木霉 CEL7B (EGI)内切葡聚糖酶的结果相比较。重组里氏木霉 CEL7B (EGI)内切葡聚糖酶可以根据包括本文的参考文献制备。

多糖包括预处理的玉米秸(PCS)、磷酸溶胀纤维素(PASC)、羧甲基纤维素(CMC)、细菌纤维素(BC)、Avicel、木聚糖、木葡聚糖、阿拉伯木聚糖、甘露聚糖和半乳甘露聚糖。除细菌纤维素以 0.9 mg/ml 使用之外, 所有底物都以 5 mg/ml 使用。

在覆盖有 Eppendorf® 96 DeepWell Mats (VWR Scientific, West Chester, PA, USA)的 Eppendorf® 96 DeepWell 平板(1.2 ml, VWR Scientific, West Chester, PA, USA)中在间歇搅动的情况下以 1 ml 的初始体积进行 24 小时反应。除非另有说明, 酶以每 g 固体 5 mg 蛋白质装载。

24 小时后, 用 8-通道移液器从水解反应中移出 20 µl 的等份试样, 添加到 MultiScreen HV 96-孔过滤平板(Millipore, Bedford, MA, USA)中 180 µl 的 102 mM Na₂CO₃-58 mM NaHCO₃ 以终止水解。将样品真空过滤到平底微板中。适当稀释后, 如实施例 18 所述用对羟苯甲酸酰肼(PHBAH)测试分析滤液的还原糖。

表3显示24-小时温育后内切葡聚糖酶对不同多糖的相对转化率。该相对转化率计算为用土生梭孢霉 CEL7C 内切葡聚糖酶水解磷酸溶胀纤维素 (PASC) 24-小时后获得的转化率的百分数。表3的结果显示三种内切葡聚糖酶对木聚糖、木葡聚糖和阿拉伯木聚糖均具有相对高的活性,但对甘露聚糖和半乳甘露聚糖的活性低。内切葡聚糖酶显示对 PASC (不溶性未取代无定形纤维素)的水解好于对 CMC (可溶性取代的纤维素衍生物)的水解。内切葡聚糖酶对具有高结晶度的不溶性底物具有低活性: 细菌纤维素、Avicel 和 PCS。

表3. 内切葡聚糖酶(每 g 固体 5 mg 蛋白质)对不同多糖底物(5 mg/ml)的相对转化率; pH 5.0, 50°C, 24 小时

底物	<i>Cladorrhinum foecundissimum</i> CEL7A	土生梭孢霉 CEL7C	里氏木霉 CEL7B
预处理的玉米秸(PCS)	5.2	ND	8.7
磷酸溶胀纤维素(PASC)	21.1**	100.0	32.4
羧甲基纤维素(CMC)	13.2**	22.5	11.6
细菌纤维素(BC)*	1.3	8.1***	3.9
Avicel (微晶纤维素)	1.4	6.4	4.2
桦木木聚糖	14.2	35.8	42.8
罗望子木葡聚糖	40.0	67.5	73.0
小麦阿拉伯木聚糖	47.5	42.5	68.2
1,4-β-D-甘露聚糖	0.5	1.4	1.8
角豆半乳甘露聚糖	0.9	1.4	2.6

*细菌纤维素的初始浓度为 0.9 mg/ml

**水解 PASC 和 CMC 时以每 g 固体 0.25 mg 蛋白质使用 *Cladorrhinum foecundissimum* CEL7A

***水解细菌纤维素时以每 g 固体 25 mg 蛋白质使用土生梭孢霉 CEL7C

实施例 22: 在 50°C 用土生梭孢霉 NRRL 8126 CEL7C 水解不同多糖

重复实施例 21 的方法,只是将水解进行 71 个小时且于不同的时间点从反应物获取等份试样以跟踪水解的时间进程。用八种多糖检测了土生梭孢霉 NRRL 8126 CEL7C 内切葡聚糖酶。多糖的相对转化程度作为水解时间的函数示于图 14。将相对转化率显示为磷酸溶胀纤维素(PASC) 71-小时水解后获

得的转化率百分数。所有反应中底物浓度均为 5 mg/ml, 且酶载荷为每 g 固体 5 mg 蛋白质。

土生梭孢霉 CEL7C 内切葡聚糖酶对 PASC、木聚糖、木葡聚糖和阿拉伯木聚糖显示出高活性, 但对甘露聚糖和半乳甘露聚糖显示出低活性。与不溶性未取代的纤维素(PASC)相比, 土生梭孢霉 CEL7C 内切葡聚糖酶对可溶性取代的纤维素(CMC)具有低得多的活性。对微晶纤维素(Avicel)的活性明显低于对无定形纤维素(PASC)的活性。

实施例 23: 用 *Cladorrhinum foecundissimum* ATCC 62373 土生梭孢霉 NRRL 8126 内切葡聚糖酶从大麦水解可溶性 β -葡聚糖

在 pH 5.5 (含 0.02% 叠氮钠的 50 mM 乙酸钠)和 60°C 测定 *Cladorrhinum foecundissimum* ATCC 62373 CEL7A 和土生梭孢霉 NRRL 8126 CEL7C 和 CEL7E 内切葡聚糖酶对来自大麦的可溶性 β -葡聚糖(中粘度, 230 kDa, Megazyme International Ireland Ltd., Bray, Ireland)的活性。将该结果与里氏木霉 CEL7B (EGI)内切葡聚糖酶的那些结果相比较。重组里氏木霉 CEL7B (EGI)内切葡聚糖酶可以如实施例 21 所述制备。

水解反应中 β -葡聚糖的初始浓度为 1.0% (w/v)。不搅动的情况下在 Eppendorf® 96 DeepWell 平板(1.2 ml, VWR Scientific, West Chester, PA)中进行 1 ml 反应。以三个蛋白质负载 0.05、0.1 和 0.2 mg 每 g 葡聚糖使用酶。在对照反应中, 用含 0.02% 叠氮钠的 50 mM 乙酸钠 pH 5.5 (缓冲液对照)或用不含重组表达的酶的浓缩并缓冲交换的米曲霉 J1250 培养液(米曲霉 J1250 对照)代替内切葡聚糖酶。

在 2 小时和 24 小时从水解反应物移出等份试样, 用去离子水稀释, 并用如实施例 18 所述的对羧苯甲酸酐肼(PHBAH)测试分析还原糖。在两个温育时间(2 小时和 24 小时) β -葡聚糖的相对转化率作为蛋白质负载的函数分别示于图 15 和 16。将相对转化率显示为用 *Cladorrhinum foecundissimum* CEL7A 内切葡聚糖酶(每 g 葡聚糖 0.2 mg 蛋白质)水解 β -葡聚糖 24-小时后所获得的转化率的百分数。

与其它内切葡聚糖酶相比, 里氏木霉 CEL7B 内切葡聚糖酶显示低的 β -葡聚糖转化率, 水解 2 小时后还原糖浓度几乎没有额外的增加。相比之下, 超过 2 小时温育时间后, *Cladorrhinum foecundissimum* CEL7A、土生梭孢霉

CEL7C 和土生梭孢霉 CEL7E 内切葡聚糖酶继续产生新的还原性端基。与土生梭孢霉 CEL7C 和土生梭孢霉 CEL7E 内切葡聚糖酶相比,在水解 β -葡聚糖方面 *Cladorrhinum foecundissimum* CEL7A 内切葡聚糖酶显示出更好的性能。

生物材料保藏

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

保藏物	保藏号	保藏日期
大肠杆菌 PaHa50	NRRL B-30899	2006 年 2 月 23 日
大肠杆菌 PaHa38	NRRL B-30896	2006 年 2 月 23 日
大肠杆菌 PaHa46	NRRL B-30897	2006 年 2 月 23 日

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

本文描述和要求保护的本发明并不受限于本文所公开的具体方面的范围, 因为这些方面意欲作为本发明几个方面的说明。任何等价的方面意欲在本发明的范围之内。实际上, 从前面的说明中, 除本文所显示和描述的之外, 本发明的多种修改对于本领域的技术人员来说是显而易见的。这些修改也意欲落入所附的权利要求的范围内。在冲突的情况下, 将以包括定义部分的本公开为准。

本文引用了多篇参考文献, 其公开的内容通过引用整体并入。

<110>	诺维信股份有限公司(Novozymes, Inc.) 诺维信公司(Novozymes A/S)	
<120>	具有内切葡聚糖酶活性的多肽和编码该多肽的多核苷酸	
<130>	10892.204-CN	
<150>	60/788,389	
<151>	2006-03-30	
<150>	PCT/US2007/065682	
<151>	2007-03-30	
<160>	26	
<170>	PatentIn version 3.5	
<210>	1	
<211>	1501	
<212>	DNA	
<213>	土生梭孢霉(<i>Thielavia terrestris</i>)	
<400>	1	60
	gccgttgtca agatgggccga gaagacgtg cacggattcg ccgccacggc ttggccgtt	
	ctcccccttg tgaaggctca gcagcccgcc aacttcacgc cggaggtgca cccgcaactg	120
	ccaacgtgga agtgcacgac cgccggcgcc tgcgttcagc aggacacitc ggtggtgctc	180
	gactggaact accgttggat ccacaatgcc gacggcaccg cctcgtgcac gacgtccagc	240
	ggggctgacc acacgctgtg tccagatgag gcgacctgcg cgaagaactg cttcgtggaa	300
	ggcgtcaact acacgagcag cgggtgtacc acatccggca gttcgtgac gatgaggcag	360
	tatttcaagg ggagcaacgg gcagaccaac agcgtttcgc ctcgtctcta cctgctcgcc	420
	tcggatggaa actacgtaat gctcaagctg ctcggccagg agctgagctt cgatgtcgat	480
	ctctccacgc tcccctgcgg cgagaacggc gcgctgtacc tgtccgagat ggacgcgacc	540
	ggtggcagga accagtacaa caccggcggt gccaaactac gctcgggcta ctgtacgcc	600
	cagtgtcccc tgcagacgtg gatgaacggc acgctgaaca ccaacgggca gggctactgc	660
	tgcaacgaga tggacatcct cgaggccaac tcccgcgcca acgcgatgac acctaccccc	720
	tgcgccaacg gcagctgcga caagagcggg tgcggactca acccctacgc cgagggtac	780
	aagagctact acggaccggg cctcacggtt gacacgtcga agcccttcac catcattacc	840
	cgcttcatca ccgacgacgg cagcaccagc ggcaacctca accagatcca gcggatctat	900
	gtgcagaatg gcaagacggg cgcgtcggct gcgtccggag gcgacatcat cacggcatcc	960
	ggctgcacct cggcccaggc gttcggcggg ctggccaaca tgggcgcggc gcttggacgg	1020
	ggcatggtgc tgaccttcag catctggaac gacgctgggg gctacatgaa ctggctcgac	1080
	agcggcaaca acggcccgtg cagcagcacc gagggcaacc cgtccaacat cctggccaac	1140
	tacccgga caaacgtggt cttctccaac atccgctggg gagacatcgg ctcgacggtc	1200
	caggctctcg gaggcggcaa cggcggctcg accaccacca cgtcgaccac cacgctgagg	1260
	acctcgacca cgaccaccac caccgccccg acggccactg ccacgcactg gggacaatgc	1320
	ggcggaatcg gggtagtca accgcctcct gcattctgtt gaggaagtta actaacgtgg	1380
	cctacgcagt ggactggacc gaccgtctgc gaatcgccgt acgcatgcaa ggagctgaac	1440
	ccctggtact accagtgcct ctaaagtatt gcagtgaagc catactccgt gctcggcatg	1500
	g	1501

<210> 2
 <211> 464
 <212> PRT
 <213> 土生梭孢霉

<400> 2

```

Met Gly Gln Lys Thr Leu His Gly Phe Ala Ala Thr Ala Leu Ala Val
1          5          10          15

Leu Pro Phe Val Lys Ala Gln Gln Pro Gly Asn Phe Thr Pro Glu Val
20          25          30

His Pro Gln Leu Pro Thr Trp Lys Cys Thr Thr Ala Gly Gly Cys Val
35          40          45

Gln Gln Asp Thr Ser Val Val Leu Asp Trp Asn Tyr Arg Trp Ile His
50          55          60

Asn Ala Asp Gly Thr Ala Ser Cys Thr Thr Ser Ser Gly Val Asp His
65          70          75          80

Thr Leu Cys Pro Asp Glu Ala Thr Cys Ala Lys Asn Cys Phe Val Glu
85          90          95

Gly Val Asn Tyr Thr Ser Ser Gly Val Thr Thr Ser Gly Ser Ser Leu
100         105         110

Thr Met Arg Gln Tyr Phe Lys Gly Ser Asn Gly Gln Thr Asn Ser Val
115         120         125

Ser Pro Arg Leu Tyr Leu Leu Gly Ser Asp Gly Asn Tyr Val Met Leu
130         135         140

Lys Leu Leu Gly Gln Glu Leu Ser Phe Asp Val Asp Leu Ser Thr Leu
145         150         155         160

Pro Cys Gly Glu Asn Gly Ala Leu Tyr Leu Ser Glu Met Asp Ala Thr
165         170         175

Gly Gly Arg Asn Gln Tyr Asn Thr Gly Gly Ala Asn Tyr Gly Ser Gly
180         185         190

Tyr Cys Asp Ala Gln Cys Pro Val Gln Thr Trp Met Asn Gly Thr Leu
195         200         205

Asn Thr Asn Gly Gln Gly Tyr Cys Cys Asn Glu Met Asp Ile Leu Glu
210         215         220

Ala Asn Ser Arg Ala Asn Ala Met Thr Pro His Pro Cys Ala Asn Gly
225         230         235         240

Ser Cys Asp Lys Ser Gly Cys Gly Leu Asn Pro Tyr Ala Glu Gly Tyr
245         250         255

Lys Ser Tyr Tyr Gly Pro Gly Leu Thr Val Asp Thr Ser Lys Pro Phe
260         265         270

Thr Ile Ile Thr Arg Phe Ile Thr Asp Asp Gly Thr Thr Ser Gly Thr
275         280         285

Leu Asn Gln Ile Gln Arg Ile Tyr Val Gln Asn Gly Lys Thr Val Ala
290         295         300

Ser Ala Ala Ser Gly Gly Asp Ile Ile Thr Ala Ser Gly Cys Thr Ser
305         310         315         320

Ala Gln Ala Phe Gly Gly Leu Ala Asn Met Gly Ala Ala Leu Gly Arg
325         330         335

Gly Met Val Leu Thr Phe Ser Ile Trp Asn Asp Ala Gly Gly Tyr Met
340         345         350

Asn Trp Leu Asp Ser Gly Asn Asn Gly Pro Cys Ser Ser Thr Glu Gly
355         360         365
  
```

Asn Pro Ser Asn Ile Leu Ala Asn Tyr Pro Asp Thr His Val Val Phe
 370 375 380
 Ser Asn Ile Arg Trp Gly Asp Ile Gly Ser Thr Val Gln Val Ser Gly
 385 390 395 400
 Gly Gly Asn Gly Gly Ser Thr Thr Thr Thr Ser Thr Thr Thr Leu Arg
 405 410 415
 Thr Ser Thr Thr Thr Thr Thr Thr Ala Pro Thr Ala Thr Ala Thr His
 420 425 430
 Trp Gly Gln Cys Gly Gly Ile Gly Trp Thr Gly Pro Thr Val Cys Glu
 435 440 445
 Ser Pro Tyr Ala Cys Lys Glu Leu Asn Pro Trp Tyr Tyr Gln Cys Leu
 450 455 460

<210> 3
 <211> 1368
 <212> DNA
 <213> 土生梭孢霉

<400> 3
 accgatccgc tcgaagatgg cgccaagtc tacagtcttg gccgcctggc tgctctcctc 60
 gctggccgcg gccacgaga tcggcaaagc cgtgcccag gtcacccca aactgacaac 120
 gcagaagtgc actctccgcg gcgggtgcaa gcctgtccgc acctcggtcg tgctcgactc 180
 gtccgcgcgc tcgtgcaca aggtcgggga cccaacacc agctgcagcg tcggcggcga 240
 cctgtgctcg gacgcgaagt cgtgcggcaa gaactgcgcg ctcgaggcg tcgactacgc 300
 ggcccacggc gtggcgacca agggcgacgc cctcacgtg caccagtggc tcaagggggc 360
 cgacggcacc tacaggaccg tctcgccgcg cgtatacctc ctgggcgagg acggaagaa 420
 ctacaggac ttcaagtgc tcaacgccga gctcagcttc gacgtcgacg tgtcccagct 480
 cgtctgcgcg atgaacggcg ccctgtactt ctccgagatg gagatggacg gcggccgcag 540
 cccgctgaac ccggcggcg ccacgtacgg caggggctac tgcgacgcg agtgcccaa 600
 gttgacttt atcaacggcg aggtatttct tctctcttct gttttcttt tccatcgctt 660
 tttctgaccg gaatccgcc tcttagctca acaccaacca cagtacggg gcgtgctgca 720
 acgagatgga catctgggag gccaacgcgc tggcgcagge gctcacgcc caccctgca 780
 acgcgacgcg ggtgtacaag tgcgacacgg cggacgagtg cgggcagccg gtgggcgtgt 840
 gcgacgaatg ggggtgctcg tacaaccctt ccaacttcgg ggtcaaggac tactacgggc 900
 gcaacctgac ggtggacacg aaccgcaagt tcacgtgac gacgcagttc gtgacgtcca 960
 acgggcggcg gacggcgag ctgaccgaga tccggcggt gtacgtgcag gacggcgtgg 1020
 tgatccagaa ccacgcggtc acggcggcg gggcgacgta cgacagcatc acggacggct 1080
 tctgcaacgc gacggccacc tggacgcagc agcggggcgg gctcgcgcgc atgggcgagg 1140
 ccatcgcccg cggcatggtg ctcatcttca gcctgtgggt tgacaacggc ggcttcatga 1200
 actggctcga cagcggcaac gccgggccct gcaacgccac cgaggcgac ccggccctga 1260
 tcctgcagca gcacccggac gccagcgtca ccttctccaa catccgatgg ggcgagatcg 1320
 gcagcacgta caagagcag tgcagccact agagtagagc ttgtaatt 1368

<210> 4
 <211> 423
 <212> PRT
 <213> 土生梭孢霉

<400> 4

Met Ala Pro Lys Ser Thr Val Leu Ala Ala Trp Leu Leu Ser Ser Leu
 1 5 10 15
 Ala Ala Ala Gln Gln Ile Gly Lys Ala Val Pro Glu Val His Pro Lys
 20 25 30
 Leu Thr Thr Gln Lys Cys Thr Leu Arg Gly Gly Cys Lys Pro Val Arg
 35 40 45
 Thr Ser Val Val Leu Asp Ser Ser Ala Arg Ser Leu His Lys Val Gly
 50 55 60
 Asp Pro Asn Thr Ser Cys Ser Val Gly Gly Asp Leu Cys Ser Asp Ala
 65 70 75 80
 Lys Ser Cys Gly Lys Asn Cys Ala Leu Glu Gly Val Asp Tyr Ala Ala
 85 90 95
 His Gly Val Ala Thr Lys Gly Asp Ala Leu Thr Leu His Gln Trp Leu
 100 105 110
 Lys Gly Ala Asp Gly Thr Tyr Arg Thr Val Ser Pro Arg Val Tyr Leu
 115 120 125
 Leu Gly Glu Asp Gly Lys Asn Tyr Glu Asp Phe Lys Leu Leu Asn Ala
 130 135 140
 Glu Leu Ser Phe Asp Val Asp Val Ser Gln Leu Val Cys Gly Met Asn
 145 150 155 160
 Gly Ala Leu Tyr Phe Ser Glu Met Glu Met Asp Gly Gly Arg Ser Pro
 165 170 175
 Leu Asn Pro Ala Gly Ala Thr Tyr Gly Thr Gly Tyr Cys Asp Ala Gln
 180 185 190
 Cys Pro Lys Leu Asp Phe Ile Asn Gly Glu Leu Asn Thr Asn His Thr
 195 200 205
 Tyr Gly Ala Cys Cys Asn Glu Met Asp Ile Trp Glu Ala Asn Ala Leu
 210 215 220
 Ala Gln Ala Leu Thr Pro His Pro Cys Asn Ala Thr Arg Val Tyr Lys
 225 230 235 240
 Cys Asp Thr Ala Asp Glu Cys Gly Gln Pro Val Gly Val Cys Asp Glu
 245 250 255
 Trp Gly Cys Ser Tyr Asn Pro Ser Asn Phe Gly Val Lys Asp Tyr Tyr
 260 265 270
 Gly Arg Asn Leu Thr Val Asp Thr Asn Arg Lys Phe Thr Val Thr Thr
 275 280 285
 Gln Phe Val Thr Ser Asn Gly Arg Ala Asp Gly Glu Leu Thr Glu Ile
 290 295 300
 Arg Arg Leu Tyr Val Gln Asp Gly Val Val Ile Gln Asn His Ala Val
 305 310 315 320
 Thr Ala Gly Gly Ala Thr Tyr Asp Ser Ile Thr Asp Gly Phe Cys Asn
 325 330 335
 Ala Thr Ala Thr Trp Thr Gln Gln Arg Gly Gly Leu Ala Arg Met Gly
 340 345 350
 Glu Ala Ile Gly Arg Gly Met Val Leu Ile Phe Ser Leu Trp Val Asp
 355 360 365
 Asn Gly Gly Phe Met Asn Trp Leu Asp Ser Gly Asn Ala Gly Pro Cys
 370 375 380
 Asn Ala Thr Glu Gly Asp Pro Ala Leu Ile Leu Gln Gln His Pro Asp
 385 390 395 400

Ala Ser Val Thr Phe Ser Asn Ile Arg Trp Gly Glu Ile Gly Ser Thr
405 410 415

Tyr Lys Ser Glu Cys Ser His
420

<210> 5
<211> 1480
<212> DNA
<213> Cladorrhinum foecundissimum

<400> 5
gatccgaatt cctcctctcg ttcttttagtc acagaccaga catctgcccc cgatggttca 60
caagttcgcc ctccctaccg gcctcgccgc ctccctcgca tctgcccagc agatcgggcac 120
cgtcgtcccc gagtctcacc ccaagcttcc caccaagcgc tgcactctcg ccggtggctg 180
ccagaccgtc gacacctcca tcgtcatcga cgccttcag cgtccctcc acaagatcgg 240
cgaccttcc actccttgcg tcgtcgccgg cctctctgc cccgacgcca agtcttgcgc 300
tgagaactgc gcgtcagagg gtgtcgacta tgcctcctgg ggcatcaaga ccgagggcga 360
cgccctaaact ctcaaccagt ggatgccccga cccggcgaac cctggccagt acaagacgac 420
tactccccgt acttaccttg ttgctgagga cggcaagaac tacgaggatg tgaagctcct 480
ggctaaggag atctcgtttg atgccgatgt cagcaacctt cctgcccga tgaacggtgc 540
tttctacttg tctgagatgt tgatggatgg tggacgtggc gacctcaacc ctgctggtgc 600
cgagtatggt accggttact gtgatgcgca gtgcttcaag ttggatttca tcaacggcga 660
ggccaacatc gaccaaagc acggcgccctg ctgcaacgaa atggacattt tcgaatccaa 720
ctcgcgccgc aagaccttcg tccccaccc ctgcaacatc acgcaggctt acaagtgcga 780
aggcgaagac gagtgcggcc agcccgtcgg cgtgtgcgac aagtgggggt gcggcttcaa 840
cgagtacaaa tggggcgctg agtccttcta cggccggggc tcgcagttcg ccatcgactc 900
ctccaagaag ttcaccgtca ccacgcagtt cctgaccgac aacggcaagg aggacggcgt 960
cctcgctgag atccgcccgt tgtggcacca ggatggcaag ctgatcaaga acaccgctat 1020
ccaggttgag gagaactaca gcacggactc ggtgagcacc gagttctgcg agaagactgc 1080
ttctttcacc atgcagcgcg gtggtctcaa ggcgatgggc gaggctatcg gtcgtggtat 1140
ggtgctggtt ttcagcatct gggcggatga ttccgggttt atgaactggt tggatgcgga 1200
gggtaatggc ccttgacgag cgactgaggg cgatccgaag gagattgtca agaataagcc 1260
ggatgctagg gttacgttct caaacattag gattggtgag gttggtagca cgtatgtccc 1320
gggtgggaag tgcggtgtta agagcagggg tgcctagggg cttactgctt ctttaagggg 1380
gtgtgaagag aggaggaggt gttgttgggg gttggagatg ataattgggc gagatggtgt 1440
agagcgggtt ggttgatat gaatacgttg aattggatgt 1480

<210> 6
<211> 440
<212> PRT
<213> Cladorrhinum foecundissimum

<400> 6

Met Val His Lys Phe Ala Leu Leu Thr Gly Leu Ala Ala Ser Leu Ala
1 5 10 15

Ser Ala Gln Gln Ile Gly Thr Val Val Pro Glu Ser His Pro Lys Leu
20 25 30

Pro Thr Lys Arg Cys Thr Leu Ala Gly Gly Cys Gln Thr Val Asp Thr
 35 40 45
 Ser Ile Val Ile Asp Ala Phe Gln Arg Pro Leu His Lys Ile Gly Asp
 50 55 60
 Pro Ser Thr Pro Cys Val Val Gly Gly Pro Leu Cys Pro Asp Ala Lys
 65 70 75 80
 Ser Cys Ala Glu Asn Cys Ala Leu Glu Gly Val Asp Tyr Ala Ser Trp
 85 90 95
 Gly Ile Lys Thr Glu Gly Asp Ala Leu Thr Leu Asn Gln Trp Met Pro
 100 105 110
 Asp Pro Ala Asn Pro Gly Gln Tyr Lys Thr Thr Thr Pro Arg Thr Tyr
 115 120 125
 Leu Val Ala Glu Asp Gly Lys Asn Tyr Glu Asp Val Lys Leu Leu Ala
 130 135 140
 Lys Glu Ile Ser Phe Asp Ala Asp Val Ser Asn Leu Pro Cys Gly Met
 145 150 155 160
 Asn Gly Ala Phe Tyr Leu Ser Glu Met Leu Met Asp Gly Gly Arg Gly
 165 170 175
 Asp Leu Asn Pro Ala Gly Ala Glu Tyr Gly Thr Gly Tyr Cys Asp Ala
 180 185 190
 Gln Cys Phe Lys Leu Asp Phe Ile Asn Gly Glu Ala Asn Ile Asp Gln
 195 200 205
 Lys His Gly Ala Cys Cys Asn Glu Met Asp Ile Phe Glu Ser Asn Ser
 210 215 220
 Arg Ala Lys Thr Phe Val Pro His Pro Cys Asn Ile Thr Gln Val Tyr
 225 230 235 240
 Lys Cys Glu Gly Glu Asp Glu Cys Gly Gln Pro Val Gly Val Cys Asp
 245 250 255
 Lys Trp Gly Cys Gly Phe Asn Glu Tyr Lys Trp Gly Val Glu Ser Phe
 260 265 270
 Tyr Gly Arg Gly Ser Gln Phe Ala Ile Asp Ser Ser Lys Lys Phe Thr
 275 280 285
 Val Thr Thr Gln Phe Leu Thr Asp Asn Gly Lys Glu Asp Gly Val Leu
 290 295 300
 Val Glu Ile Arg Arg Leu Trp His Gln Asp Gly Lys Leu Ile Lys Asn
 305 310 315 320
 Thr Ala Ile Gln Val Glu Glu Asn Tyr Ser Thr Asp Ser Val Ser Thr
 325 330 335
 Glu Phe Cys Glu Lys Thr Ala Ser Phe Thr Met Gln Arg Gly Gly Leu
 340 345 350
 Lys Ala Met Gly Glu Ala Ile Gly Arg Gly Met Val Leu Val Phe Ser
 355 360 365
 Ile Trp Ala Asp Asp Ser Gly Phe Met Asn Trp Leu Asp Ala Glu Gly
 370 375 380
 Asn Gly Pro Cys Ser Ala Thr Glu Gly Asp Pro Lys Glu Ile Val Lys
 385 390 395 400
 Asn Lys Pro Asp Ala Arg Val Thr Phe Ser Asn Ile Arg Ile Gly Glu
 405 410 415
 Val Gly Ser Thr Tyr Ala Pro Gly Gly Lys Cys Gly Val Lys Ser Arg
 420 425 430
 Val Ala Arg Gly Leu Thr Ala Ser
 435 440

<210> 7
 <211> 26
 <212> DNA
 <213> 土生梭孢霉

<220>
 <221> misc_feature
 <222> (15)..(15)
 <223> N=A, C, G 或 T

<220>
 <221> misc_feature
 <222> (18)..(18)
 <223> N=A, C, G 或 T

<220>
 <221> misc_feature
 <222> (21)..(21)
 <223> R=A 或 G

<220>
 <221> misc_feature
 <222> (24)..(24)
 <223> R=A 或 G

<400> 7
 aggggtgccgc tggtngtnc rtrctc

26

<210> 8
 <211> 33
 <212> DNA
 <213> 土生梭孢霉

<220>
 <221> misc_feature
 <222> (25)..(25)
 <223> N= A, C, G 或 T

<220>
 <221> misc_feature
 <222> (28)..(28)
 <223> Y=T 或 C

<220>
 <221> misc_feature
 <222> (31)..(31)
 <223> Y=T 或 C

<400> 8
 agctgggtgct aaatatggta ctggntaytg yga

33

<210> 9
 <211> 25
 <212> DNA
 <213> 土生梭孢霉

<220>
 <221> misc_feature
 <222> (18)..(18)
 <223> R=A 或 G

<220>
 <221> misc_feature
 <222> (21)..(21)
 <223> Y=T 或 C

<220>
 <221> misc_feature
 <222> (23)..(23)
 <223> N=A, C, G 或 T

<400> 9 ggctactgcg acgcccartg ycnaa	25
<210> 10 <211> 28 <212> DNA <213> 土生梭孢霉	
<220> <221> misc_feature <222> (11)..(11) <223> R=A 或 G	
<220> <221> misc_feature <222> (17)..(17) <223> Y=T 或 C	
<220> <221> misc_feature <222> (23)..(23) <223> R=A 或 G	
<220> <221> misc_feature <222> (26)..(26) <223> R=A 或 G	
<400> 10 cctcccagat rtccatytcg ttrcarca	28
<210> 11 <211> 25 <212> DNA <213> 土生梭孢霉	
<400> 11 cttggtagcg agctcggatc cacta	25
<210> 12 <211> 25 <212> DNA <213> 土生梭孢霉	
<400> 12 atagggcgaa ttgggccctc tagat	25
<210> 13 <211> 22 <212> DNA <213> 构巢曲霉(Aspergillus nidulans)	
<400> 13 glgccccatg atacgcctcc gg	22
<210> 14 <211> 26 <212> DNA <213> 构巢曲霉	
<400> 14 gagtcgtatt tccaaggctc ctgacc	26
<210> 15 <211> 24 <212> DNA <213> 构巢曲霉	
<400> 15	

ggaggccatg aagtggacca acgg	24
<210> 16	
<211> 45	
<212> DNA	
<213> 黑曲霉(<i>Aspergillus niger</i>)	
<400> 16	
caccgtgaaa gccatgctct ttccttcgtg tagaagacca gacag	45
<210> 17	
<211> 45	
<212> DNA	
<213> 黑曲霉	
<400> 17	
ctggtcttct acacgaagga aagagcatgg ctttcacggt gtctg	45
<210> 18	
<211> 44	
<212> DNA	
<213> 黑曲霉	
<400> 18	
ctatatacac aactggattt accatgggcc cgcggccgca gatc	44
<210> 19	
<211> 44	
<212> DNA	
<213> 黑曲霉	
<400> 19	
gatctgcggc cgcgggcca tggtaaattcc agttgtgtat atag	44
<210> 20	
<211> 29	
<212> DNA	
<213> 米曲霉(<i>Aspergillus oryzae</i>)	
<400> 20	
actggattac catgggccag aagacgctg	29
<210> 21	
<211> 31	
<212> DNA	
<213> 米曲霉	
<400> 21	
agtcacctct agttagaggc actggtagta c	31
<210> 22	
<211> 36	
<212> DNA	
<213> 土生梭孢霉	
<400> 22	
actggattac catggcgccc aagtctacag ttctgg	36
<210> 23	
<211> 38	
<212> DNA	
<213> 土生梭孢霉	
<400> 23	
tcacctctag ttaattaact agtggctgca ctcgctct	38
<210> 24	
<211> 12	

<212>	DNA	
<213>	Cladorrhinum foecundissimum	
<400>	24	
	ctttccagca ca	12
<210>	25	
<211>	21	
<212>	DNA	
<213>	Cladorrhinum foecundissimum	
<400>	25	
	ccacacttct cttccttcct c	21
<210>	26	
<211>	21	
<212>	DNA	
<213>	Cladorrhinum foecundissimum	
<400>	26	
	ccccatcctt taactatagc g	21

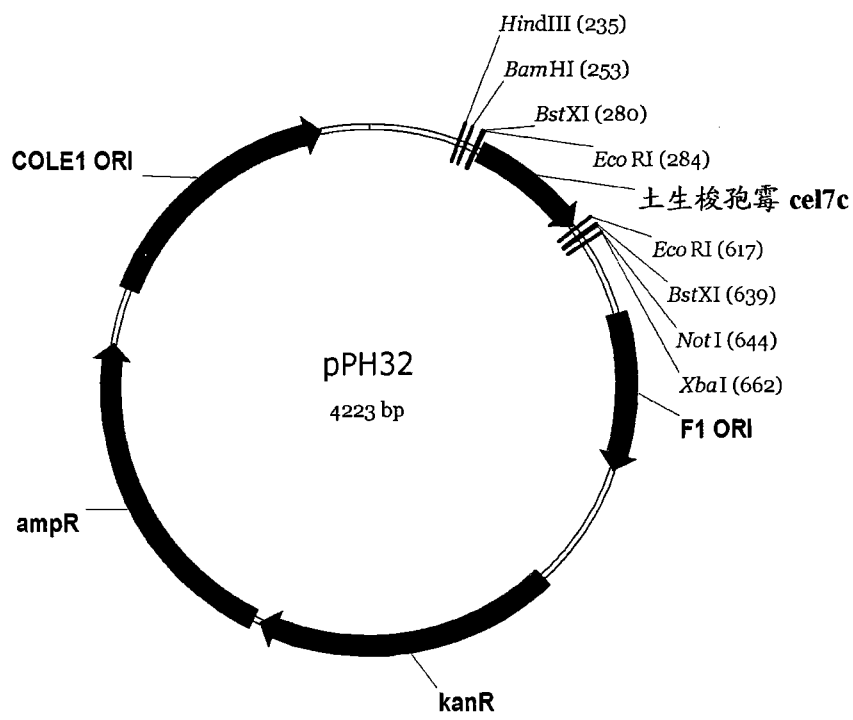


图 1

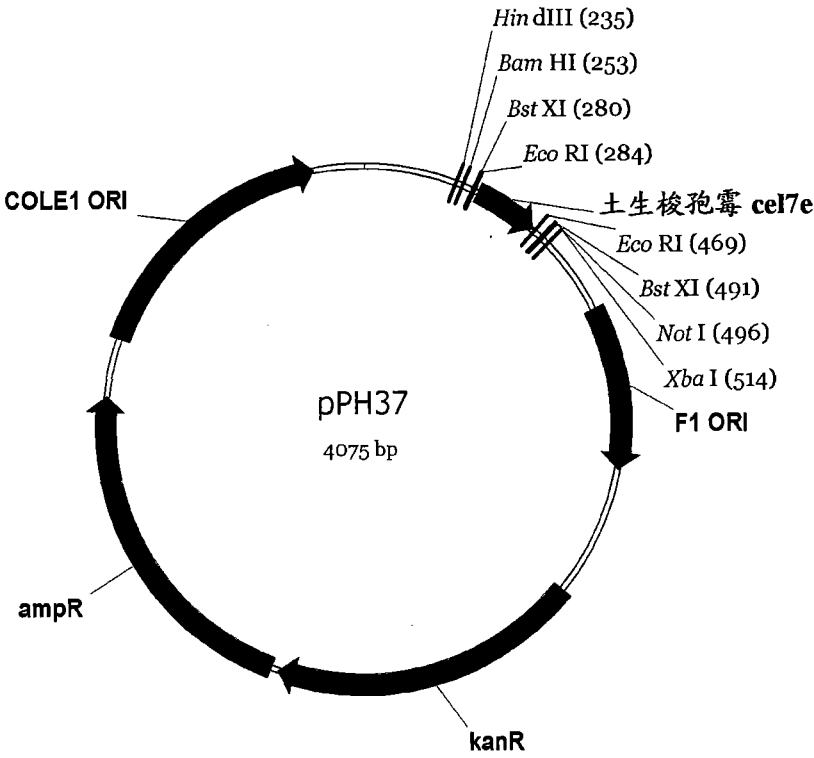


图 2

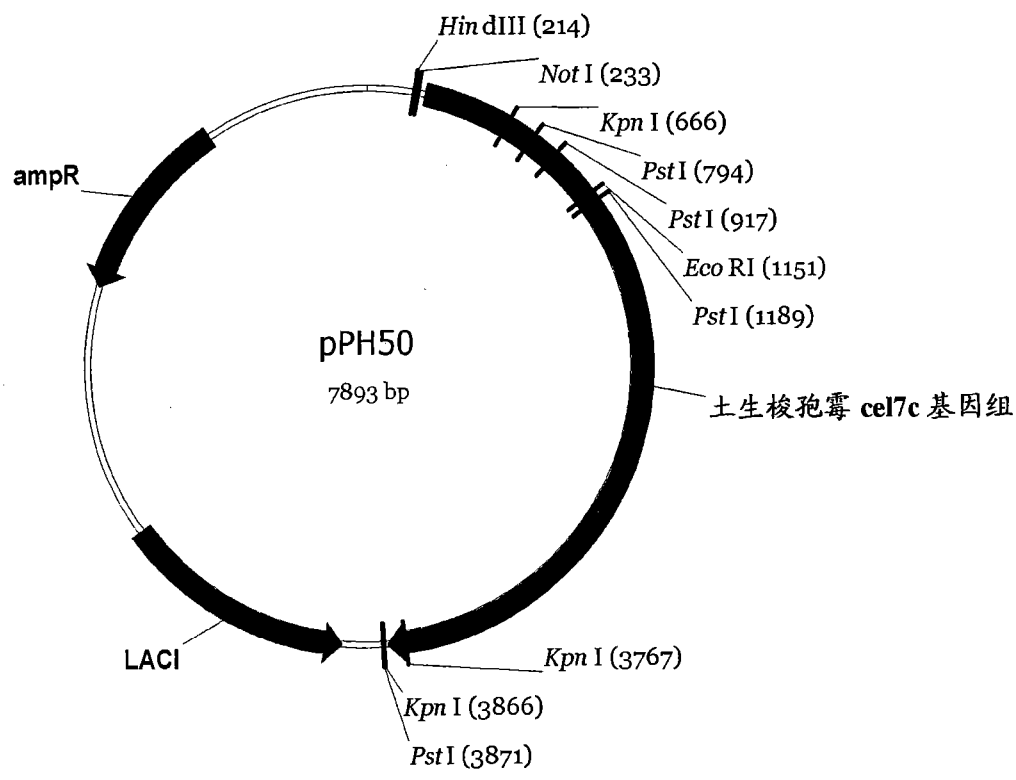


图 3

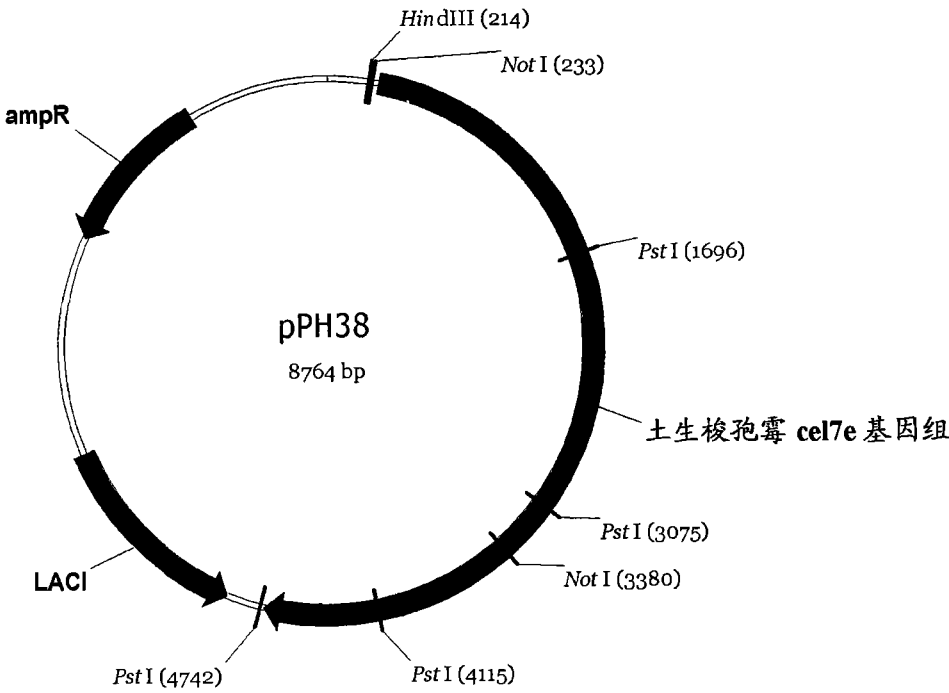


图 4

GCCGTTGTCAAGATGGGCCAGAAGACGCTGCACGGATTCCGCCGCCACGGCTTTGGCCGTTCTCCCTTTGTGAAGGCTC
 M G Q K T L H G F A A T A L A V L P F V K A
 AGCAGCCCGCAACTTCACGCCGGAGGTGCACCCGCAACTGCCAACGTGGAAGTGCACGACCGCGGGCTGCGTTCA
 Q Q P G N F T P E V H P Q L P T W K C T T A G G C V Q
 GCAGGACACTTCGGTGGTCTCGACTGGAACACTACCGTTGGATCCACAATGCCAGCGCACCGCTCGTGACGACGTCC
 Q D T S V V L D W N Y R W I H N A D G T A S C T T S
 AGCGGGTGCACACACGCTGTGTCCAGATGAGGCGACCTGCCGGAAGAACTGCTTCGTGGAAGGCGTCAACTACACGA
 S G V D H T L C P D E A T C A K N C F V E G V N Y T
 GCAGCGGTGTACACACATCCGGCAGTTCGCTGACGATGAGGCAGTATTTCAGGGGAGCAACGGGCAGACCAACAGCGT
 S S G V T T S G S S L T M R Q Y F K G S N G Q T N S V
 TTCGCCTCGTCTACCTGCTCGGCTCGGATGGAACACTACGTAATGCTCAAGCTGCTCGGCCAGGAGCTGAGCTTCGAT
 S P R L Y L L G S D G N Y V M L K L L G Q E L S F D
 GTCGATCTCCACGCTCCCTGCGGCGAGAACGGCGCTGTACCTGTCCGAGATGGACGCGACCGGTGGCAGGAACC
 V D L S T L P C G E N G A L Y L S E M D A T G G R N
 AGTACAACACCGCGGTGCCAACTACGGCTCGGGCTACTGTGACGCCCAAGTGTCCCGTGCAGACGTGGATGAACGGCAC
 Q Y N T G G A N Y G S G Y C D A Q C P V Q T W M N G T
 GCTGAACACCAACGGCAGGGCTACTGCTGCAACGAGATGGACATCCTCGAGGCCAACTCCCGCGCCAAACGGATGACA
 L N T N G Q G Y C C N E M D I L E A N S R A N A M T
 CCTCACCCCTGCGCCAACGGCAGCTGCGACAAGAGCGGTGCGGACTCAACCCCTACGCCGAGGGCTACAAGAGCTACT
 P H P C A N G S C D K S G C G L N P Y A E G Y K S Y
 ACGGACCGGGCTCACGGTTGACACGTCGAAGCCCTTACCACATCATACCCGCTTCATCACCGACGACGACGACCCAG
 Y G P G L T V D T S K P F T I I T R F I T D D G T T S
 CGCACCCCTCAACCAGATCCAGCGGATCTATGTGCAAGATGGCAAGACGTCGCGTCCGTCCGGAGGCGACATC
 G T L N Q I Q R I Y V Q N G K T V A S A S G G D I
 ATCAGGGCATCCGGCTGCACCTCGGCCACGGGCTCGCGGGCTGGCCAAACATGGCGCGGCGCTTGGACGGGCATGG
 I T A S G C T S A Q A F G G L A N M G A A L G R G M
 TGCTGACCTTCAGCATCTGGAACGACGCTGGGGCTACATGAACCTGGCTCGACAGCGGCAACAACGGCCCCGTGCAGCAG
 V L T F S I W N D A G G Y M N W L D S G N N G P C S S

图 5A

CACCGAGGGCAACCCGTCCAAACATCCTGGCCAACTACCCGGACACCCACGTGGTCTTCTCCAACATCCCGCTGGGGAGAC
T E G N P S N I L A N Y P D T H V V F S N I R W G D
ATCGGCTCGACGGTCCAGGTCTCGGGAGGCGGCAACGGCGGCTCGACCAACCAACGTCGACCAACGCTGAGGACCT
I G S T V Q V S G G N G G S T T T S T T L R T
CGACCACGACCAACCAACCGCCCCGACGGCCACTGCCACGCACTGGGGACAATCGGGCGGAATCGGGGTACGTCAAACC
S T T T T A P T A T A T H W G Q C G G I G
GCCTCCTGCATTCTGTTGAGGAAGTTAACTAACGTGGCCCTACGCAGTGGACTGGACCGACCGTCTGCCGAATCGCCGTAC
W T G P T V C E S P Y
GCATGCAAGGAGCTGAACCCCTGGTACTACCAGTGCCTCTAAAGTATTGCAGTGAAGCCATACTCCGTGCTCGGCATGG
A C K E L N P W Y Y Q C L .

图 5B

ACCGATCCGGCTCGAAGATGGCGGCCAAGTCTACAGTTCTGGCCGCTGGCTGCTCTCTCGCTGGCGGCCCGCAG
 M A P K S T V L A A W L L S S L A A Q
 CAGATCGGCAAGCCGTGCCGAGGTCCACCCAACTGACAACGCAGAGTGCACCTCTCCGCGGGGTGCAAGC
 Q I G K A V P E V H P K L T T Q K C T L R G G C K
 CTGTCCGCACCTCGGTGCTCGACTCGTCCGCGCTCGCTGCACAAGTGGGACCCCAACACCAAGTGCAG
 P V R T S V V L D S S A R S L H K V G D P N T S C S
 CGTCGGCGGCGACCTGTGCTCGGACGCGAAGTCGTGCGGCAAGAACTGCGCGCTCGAGGGCGTCGACTACGCGGCC
 V G G D L C S D A K S C G K N C A L E G V D Y A A
 CACGGCGTGGGACCAAGGGGACGCCCTCACGCTGCACCAAGTGGCTCAAGGGGCGGACGGCACCTACAGGACCG
 H G V A T K G D A L T L H Q W L K G A D G T Y R T
 TCTCGCCGCGGTATACCTCCTGGCGGAGGACGGGAAGAACTACGAGGACITCAAGCTGCTCAACGCCGAGCTCAG
 V S P R V Y L L G E D G K N Y E D F K L L N A E L S
 CTTGACGTCGACGTGTCCAGCTCGTCTGCGGCAATGAACGGGCCCTGTACTTCTCCGAGATGGAGATGGACGGC
 F D V D V S Q L V C G M N G A L Y F S E M E M D G
 GGCGCAGCCCGCTGAACCCGCGGGCGCCACGTACGGCACGGGCTACTGCGACGCGCAGTGCCCCAAGTTGGACT
 G R S P L N P A G A T Y G T G Y C D A Q C P K L D
 TTATCAACGGCGAGGTTTCTCTCTCTCTGTTTTTCCATCGCTTTTCTGACCGGAATCCGCCCTCTT
 F I N G E
 AGCTCAACACCAACACACGTACGGGGCGTGTGCAACGAGATGGACATCTGGGAGGCCAACGCGCTGGCGCAGGC
 L N T N H T Y G A C C N E M D I W E A N A L A Q A
 GCTCAGCCCGCACCCGTGCAACGCGACGCGGGTGTACAAGTGCACACGCGGACGAGTGGGGCAGCCGGTGGGC
 L T P H P C N A T R V Y K C D T A D E C G Q P V G
 GTGTGCGACGAATGGGGTGTCTGTACAACCCGTCCAACCTCGGGGTCAAGGACTACTACGGGCGCAACCTGACGG
 V C D E W G C S Y N P S N F G V K D Y Y G R N L T
 TGGACACGAACCGCAAGTTACGGTGACGACGCGAGTTCGTGACGTCCAACGGGCGGCGGACGGCGAGCTGACCCGA
 V D T N R K F T V T T Q F V T S N G R A D G E L T E

图 6A

GATCCGGCGGCTGTACGTGCAGGACGGCGTGGTGATCCAGAACCAACGCGGTACGCGGGGGCGGACGTTACGAC
I R R L Y V Q D G V V I Q N H A V T A G G A T Y D
AGCATCAGGACGGCTTCTGCAACGCGACGGCCACCTGGACGCAGCAGCGGGGGCTCGCGGCATGGGCGAGG
S I T D G F C N A T A T W T Q Q R G G L A R M G E
CCATCGGCGGCGCATGTGCTCATCTTCAGCCTGTGGTTGACAAACGGCGGCTTCATGAACCTGGCTCGACAGCGG
A I G R G M V L I F S L W V D N G G F M N W L D S G
CAACGCGGGCCCTGCAACGCCACCGAGGGCGACCCGCCCTGATCCTGCAGCAGCAACCCGGACGCCAGCGTCACC
N A G P C N A T E G D P A L I L Q Q H P D A S V T
TTCTCCAACATCCGATGGGGCGAGATCGGCAGCACGTACAAGAGCGAGTGCAGCCACTAGAGTAGAGCTTGTAAATT
F S N I R W G E I G S T Y K S E C S H .

图 6B

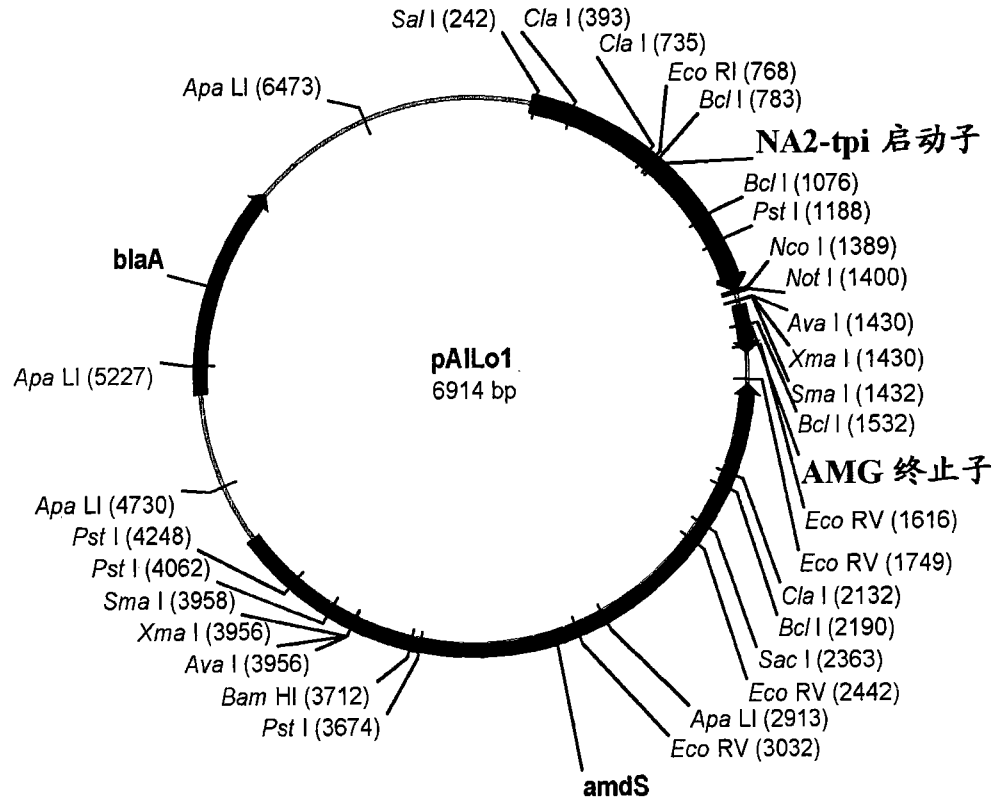


图 7

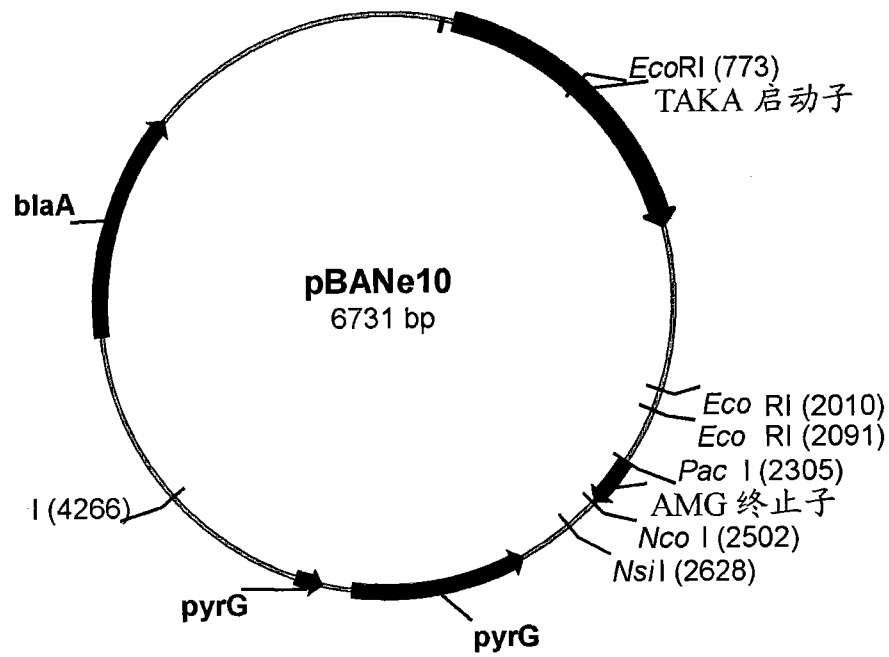


图 8

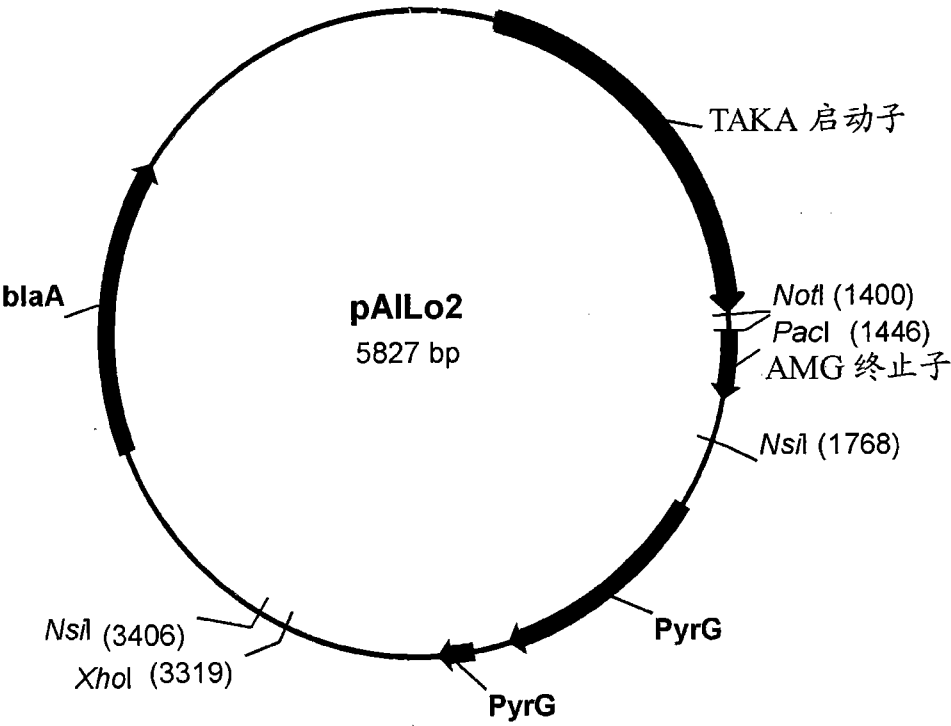


图 9

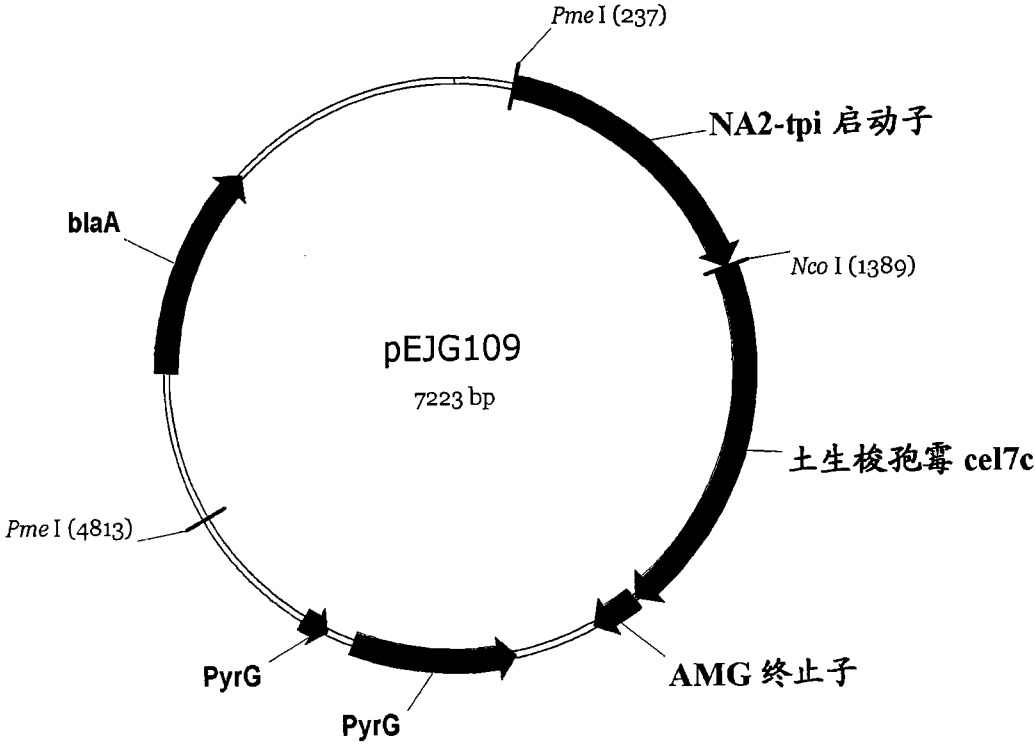


图 10

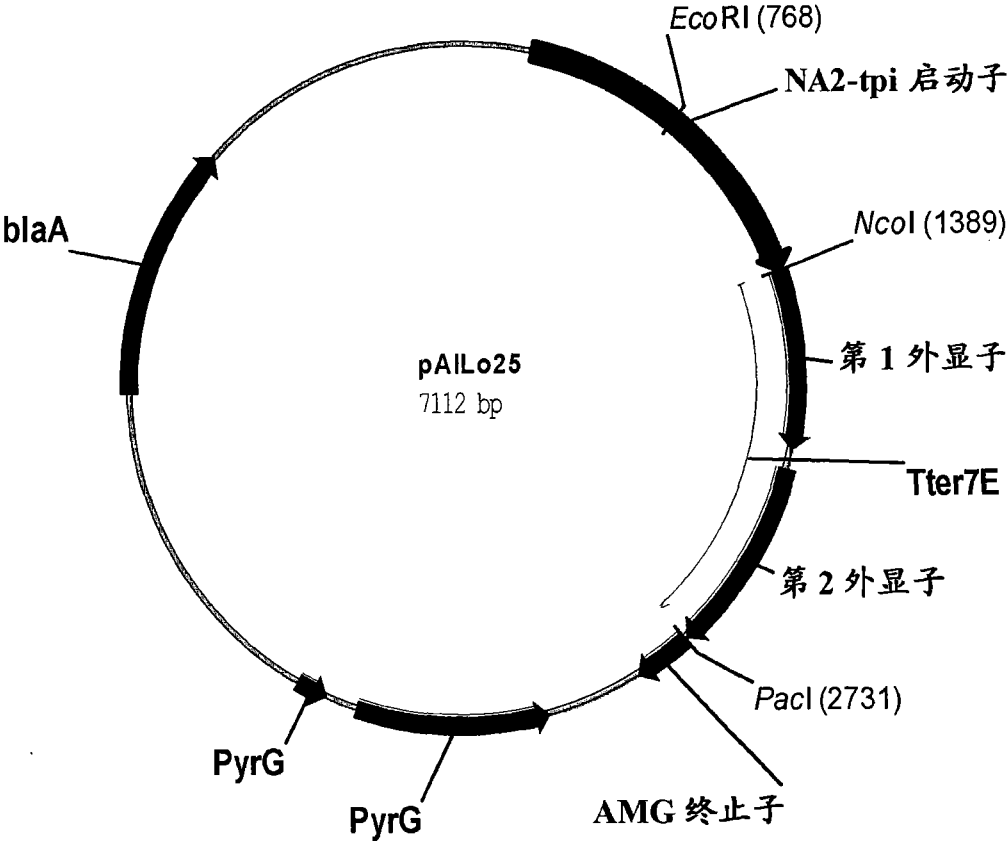


图 11

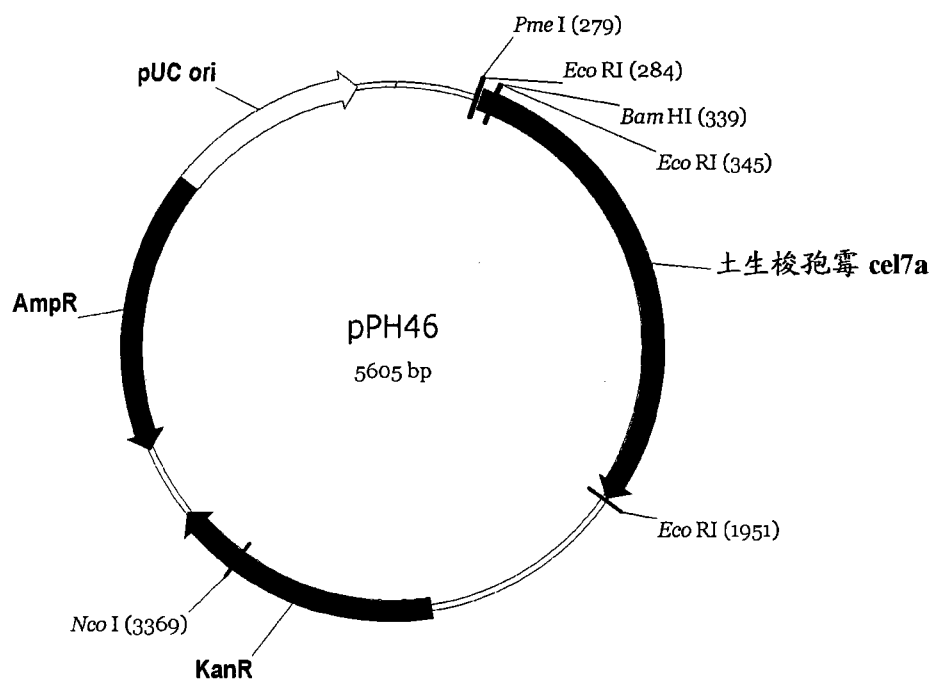


图 12

TACAGCACGGACTCGGTGAGCACCGAGTTCTGCGAGAAGACTGCTTCTTTCACCATGCAGCGCGGTGCTCTCAA
Y S T D S V S T E F C E K T A S F T M Q R G G L K
GGCGATGGCGAGGCTATCGGTCGTGGTAATGGTGTCTGGTTTTCAGCATCTGGCGGATGATTCGGGTTTATGA
A M G E A I G R G M V L V F S I W A D D S G F M
ACTGGTTGGATGCGGAGGTAATGGCCCTTGCAGCGCGACTGAGGGCGATCCGAAGGAGATTGTCAAGAATAAG
N W L D A E G N G P C S A T E G D P K E I V K N K
CCGATGCTAGGTTACGTTCTCAAACATTAGGATTGGTGAGGTTGGTAGCACGTATGCTCCGGGTGGGAAGTG
P D A R V T F S N I R I G E V G S T Y A P G K C
CGGTGTTAAGAGCAGGTTGCTAGGGGCTTACTGCTTCTTAAGGGGGTGTGAAGAGAGAGGAGGTGTTGTT
G V K S R V A R G L T A S .
GGGGTTGGAGATGATAATTGGCGGAGATGGTGTAGAGCGGGTTGGTTGGATATGAATACGTTGAATTGGATGT

图 13B

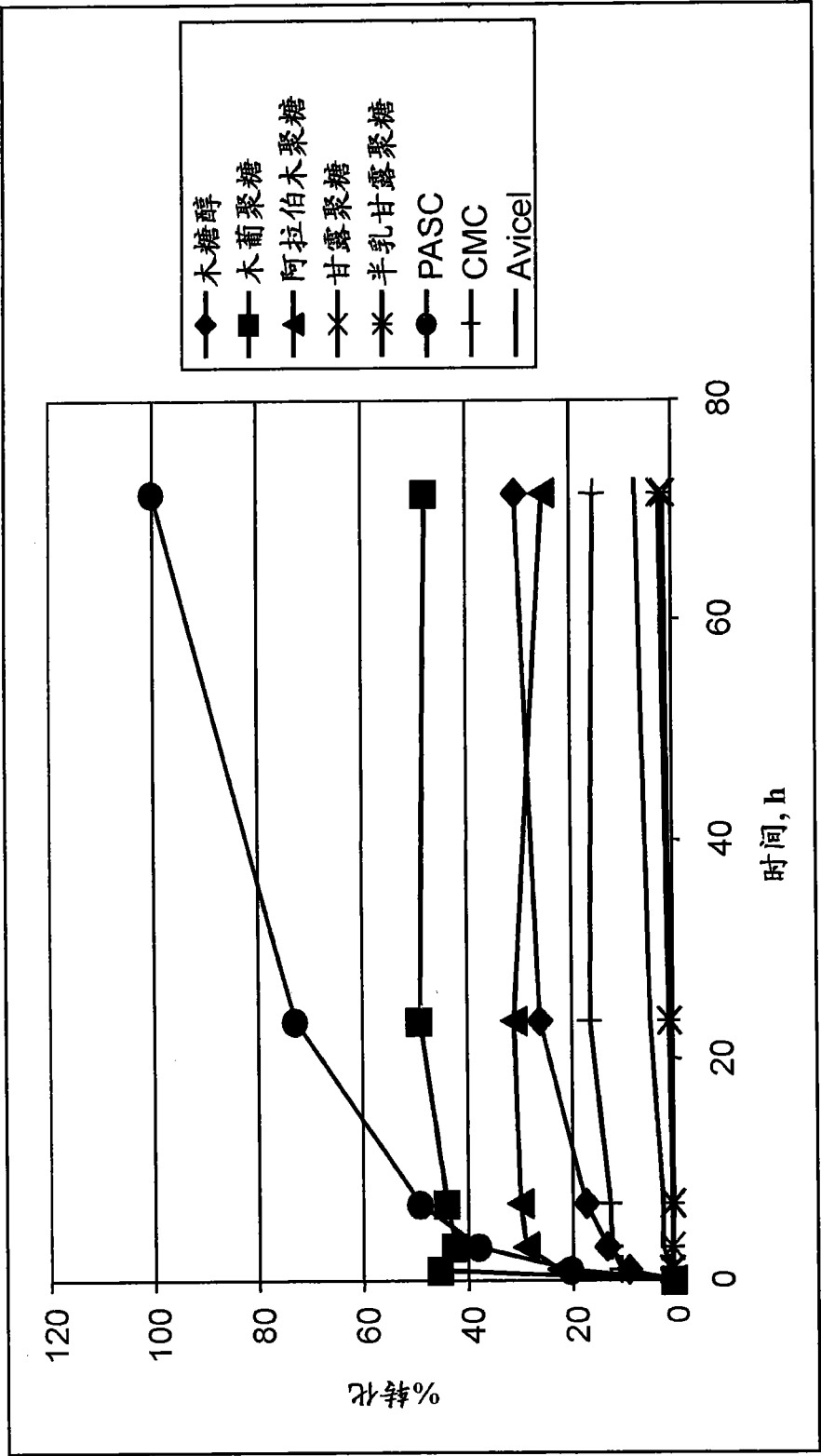


图 14

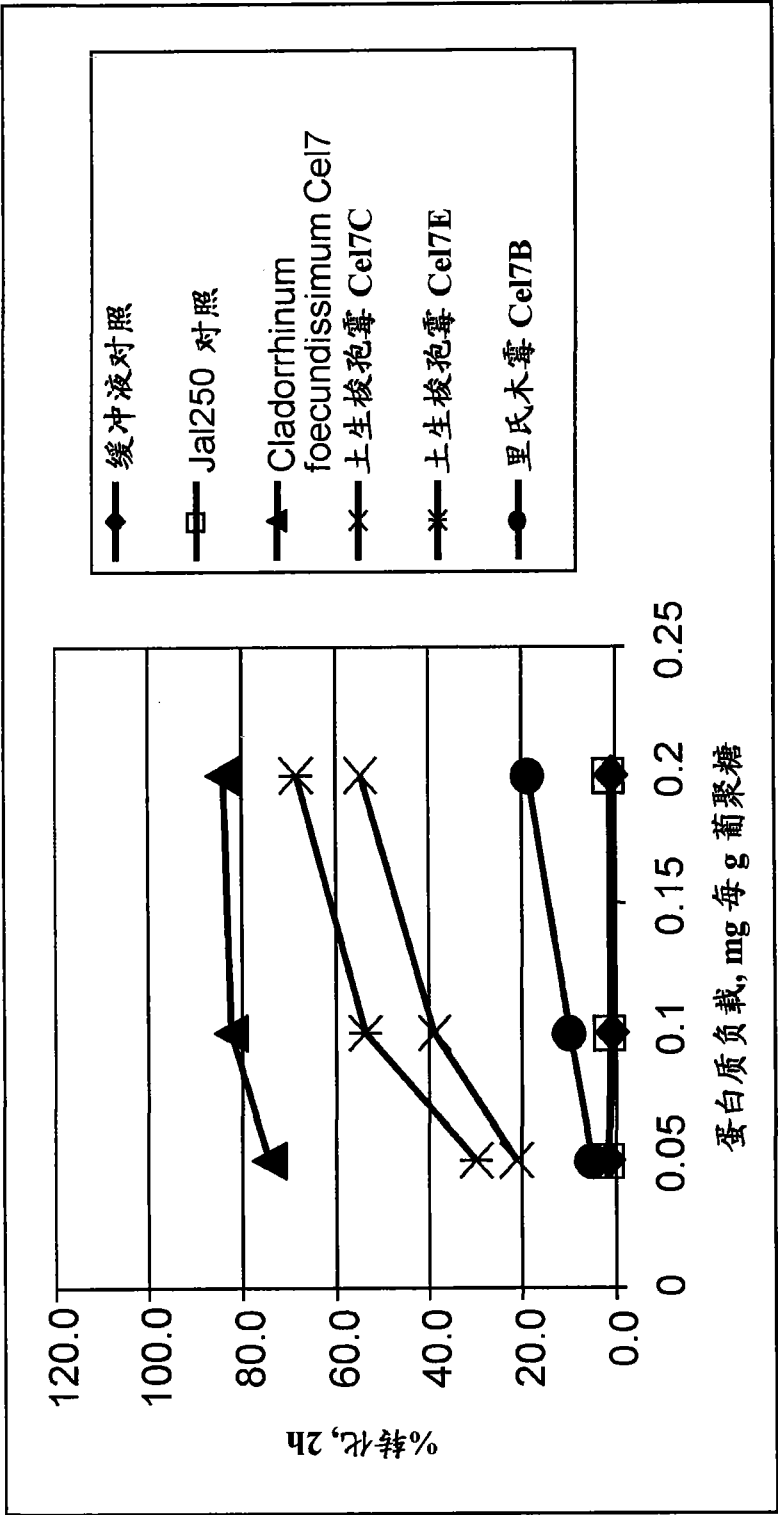


图 15

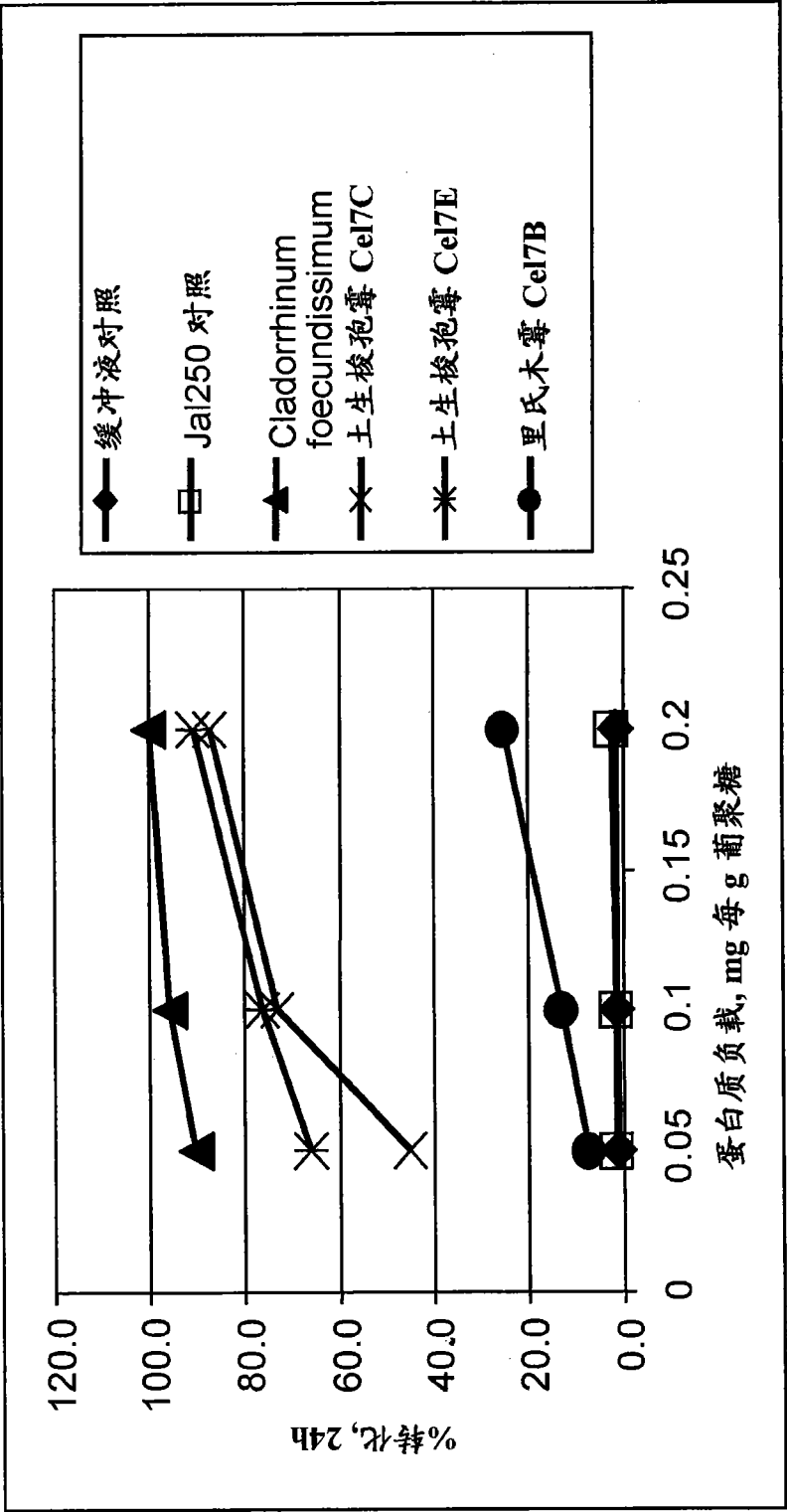


图 16