(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 107073052 A
(43)申请公布日 2017.08.18

(21)申请号 201580055388.8

(22)申请日 2015.09.25

(30)优先权数据
62/055,702 2014.09.26 US

(85)PCT国际申请进入国家阶段日
2017.04.12

(86)PCT国际申请的申请数据
PCT/US2015/052200 2015.09.25

(87)PCT国际申请的公布数据
WO2016/049443 EN 2016.03.31

(71)申请人 希乐克公司
地址 美国马萨诸塞州

(72)发明人 M·梅多夫 N·科瑞德 J·林驰 S·兰德里 A·吉田 D·彭泽伦 T·C·马斯特曼

(74)专利代理机构 北京市铸成律师事务所 11313
代理人 王建秀 郗名悦

(51)Int.Cl.
A61K 36/00(2006.01)
A63J 1/00(2006.01)
C12N 15/74(2006.01)
C12N 9/14(2006.01)

权利要求书4页 说明书41页
序列表32页 附图3页

(54)发明名称

被增溶酶及其用途

(57)摘要

本发明涉及包含从包涵体增溶，并且保留生物质降解活性的具有生物质降解活性的一种多肽或多种多肽的混合物，以及用于产生和使用所述混合物的方法。本文所述的本发明提供用于增加可从宿主细胞分离的具有生物质降解活性的重组蛋白的产率的方法。

CN 107073052 A

1.一种混合物，其包含多种具有生物质降解活性的多肽和增溶剂，其中相较于具有生物质降解活性的天然多肽，所述多肽具有至少8-10％的所述生物质降解活性。

2.如权利要求1所述的混合物，其进一步包含一种或多种与包涵体相关的蛋白质。

3.如权利要求1所述的混合物，其中所述混合物不包含一种或多种与包涵体相关的蛋白质。

4.如前述权利要求中任一项所述的混合物，其进一步包含细胞碎片、一种或多种核糖体组分、一种或多种宿主蛋白质和/或包括DNA和/或RNA的宿主核酸。

5.如前述权利要求中任一项所述的混合物，其中所述生物质降解活性是纤维二糖酶活性、木质素酶活性、内切葡聚糖酶活性、纤维二糖水解酶活性或木聚糖酶活性。

6.如前述权利要求中任一项所述的混合物，其中所述多肽是部分未折叠的，部分错误折叠的，或部分变性的。

7.如权利要求1所述的混合物，其中所述多肽包含与SEQ ID NO:1具有至少90％同一性的氨基酸序列。

8.如前述权利要求中任一项所述的混合物，其中所述多肽包括来自里氏木霉的Cel3A酶或其功能性变体或片段。

9.如权利要求8所述的混合物，其中所述Cel3A酶包含氨基酸序列SEQ ID NO:1或与其具有至少90％同一性的氨基酸序列。

10.如前述权利要求中任一项所述的混合物，其中所述多肽由包含与SEQ ID NO:2或SEQ ID NO:3的至少90％同一性的核酸序列编码。

11.如前述权利要求中任一项所述的混合物，其中所述多肽是无糖基化的。

12.如权利要求1-3或7中任一项所述的混合物，其中所述增溶剂包括尿素，并且任选地，以介于0.2M-6M之间的浓度存在。

13.如前述权利要求中任一项所述的混合物，其进一步包含至少一种具有生物质降解活性的额外多肽或产生一种或多种具有生物质降解活性的酶的微生物。

14.如权利要求13所述的混合物，其中所述额外多肽选自木质素酶、内切葡聚糖酶、纤维二糖水解酶、纤维二糖酶和木聚糖酶或其任何组合。

15.如权利要求13或14所述的混合物，其中所述额外多肽选自：

a.包含与SEQ ID NO:1具有至少90％同一性的氨基酸序列的多肽；

b.来自里氏木霉的Cel3A酶或其功能性变体或片段；或

c.由包含SEQ ID NO:2或SEQ ID NO:3(例如由SEQ ID NO:2或SEQ ID NO:3组成)的核酸序列编码的多肽。

16.如权利要求13-15中任一项所述的混合物，其中所述额外多肽是无糖基化的。

17.如权利要求13-15中任一项所述的混合物，其中所述额外多肽是糖基化的。

18.一种混合物，其包含具有与SEQ ID NO:1具有至少90％同一性的氨基酸序列的多种多肽和增溶剂，其中所述多种多肽具有至少20％-40％的包含SEQ ID NO:1的天然多肽的活性。

19.如权利要求18所述的混合物，其进一步包含一种或多种与包涵体相关的蛋白质。

20.如权利要求18所述的混合物，其中所述混合物不包含一种或多种与包涵体相关的蛋白质。

21.如权利要求18-20中任一项所述的混合物，其进一步包含细胞碎片、一种或多种核糖体组分、一种或多种宿主蛋白质和/或包括DNA和/或RNA的宿主核酸。

22.如权利要求18-21中任一项所述的混合物，其中所述多肽是部分未折叠的，部分错误折叠的，或部分变性的。

23.如权利要求18-22中任一项所述的混合物，其中所述多肽由包含与SEQ ID NO:2或SEQ ID NO:3的至少90％同一性的核酸序列编码。

24.如权利要求18-23中任一项所述的混合物，其中所述多肽是无糖基化的。

25.如权利要求18-20中任一项所述的混合物，其中所述增溶剂包括尿素，并且任选地，以介于0.2M-6M之间的浓度存在。

26.如权利要求18-25中任一项所述的混合物，其进一步包含至少一种具有生物质降解活性的额外多肽或产生一种或多种具有生物质降解活性的酶的微生物。

27.如权利要求26所述的混合物，其中所述额外多肽选自木质素酶、内切葡聚糖酶、纤维二糖水解酶、纤维二糖酶和木聚糖酶或其任何组合。

28.如权利要求26或27所述的混合物，其中所述额外多肽选自：

a.包含与SEQ ID NO:1具有至少90％同一性的氨基酸序列的多肽；

b.来自里氏木霉的Ce13A酶或其功能性变体或片段；或

c.由包含SEQ ID NO:2或SEQ ID NO:3(例如由SEQ ID NO:2或SEQ ID NO:3组成)的核酸序列编码的多肽。

29.如权利要求26-28中任一项所述的混合物，其中所述额外多肽是无糖基化的。

30.如权利要求26-28中任一项所述的混合物，其中所述额外多肽是糖基化的。

31.一种用于产生如权利要求1-30中任一项所述的混合物的方法，其包括使表达具有生物质降解活性的所述多肽的细胞或其溶解产物与适于增溶所述多肽的浓度的增溶剂接触。

32.如权利要求31所述的方法，其进一步包括溶解所述细胞以获得溶解产物，使所述溶解产物的可溶性部分与不溶性部分分离，以及将所述不溶性部分再混悬于所述增溶剂中。

33.如权利要求31或32所述的方法，其中所述增溶剂是尿素，并且任选地，其中所述增溶剂的所述浓度介于0.2M-6M之间。

34.如权利要求31-33中任一项所述的方法，其中所述生物质降解活性是纤维二糖酶活性、木质素酶活性、内切葡聚糖酶活性、纤维二糖水解酶或木聚糖酶活性。

35.如权利要求31-34中任一项所述的方法，其中所述多肽包含与SEQ ID NO:1具有至少90％同一性的氨基酸序列。

36.如权利要求31-35中任一项所述的方法，其中所述多肽包括来自里氏木霉的Ce13A或其功能性变体或片段。

37.如权利要求31-36中任一项所述的方法，其中所述多肽是无糖基化的。

38.一种用于产生具有生物质降解活性的多肽的方法，其包括在细胞中表达所述多肽，以及使所述细胞或其溶解产物与适于增溶所述多肽的浓度的增溶剂接触。

39.一种用于产生具有生物质降解活性的多肽的方法，其包括提供已被遗传修饰来产生至少一种具有生物质降解活性的多肽的细胞，其中至少一部分具有生物质降解活性的所述多肽见于包涵体中，以及使含有所述包涵体的所述细胞或其溶解产物与适于增溶所述多

肽的浓度的增溶剂接触。

40.如权利要求38或39所述的方法，其中所述增溶剂包括尿素。

41.如权利要求38-40中任一项所述的方法，其中所述增溶剂的所述浓度介于0.2M-6M之间。

42.如权利要求38-41中任一项所述的方法，其进一步包括溶解所述细胞以获得溶解产物，使所述溶解产物的可溶性部分与不溶性部分分离，以及将所述不溶性部分再混悬于所述增溶剂中。

43.如权利要求38-42中任一项所述的方法，其中所述生物质降解活性是纤维二糖酶活性、木质素酶活性、内切葡聚糖酶活性、纤维二糖水解酶活性或木聚糖酶活性。

44.如权利要求38或39中任一项所述的方法，其中所述多肽包含与SEQ ID NO:1具有至少90％同一性的氨基酸序列。

45.如权利要求38-44中任一项所述的方法，其中所述多肽包括来自里氏木霉的Cel3A或其功能性变体或片段。

46.如权利要求38-45中任一项所述的方法，其中所述细胞是原核或细菌细胞，例如大肠杆菌细胞、origami大肠杆菌细胞。

47.如权利要求38-46中任一项所述的方法，其中所述多肽是无糖基化的。

48.一种从生物质产生产物的方法，其包括使生物质与如权利要求1-30中任一项所述的混合物，以及任选地，产生一种或多种生物质降解酶的微生物和/或包含生物质降解酶的酶混合物在适于产生所述产物的条件下接触。

49.如权利要求48所述的方法，其进一步包括在使所述生物质与所述混合物接触之前用电子束处理所述生物质。

50.如权利要求48或49所述的方法，其中所述产物是糖产物。

51.如权利要求50所述的方法，其中所述糖产物是葡萄糖和/或木糖。

52.如权利要求48-51中任一项所述的方法，其进一步包括分离所述产物。

53.如权利要求52所述的方法，其中所述产物的所述分离包括沉淀、结晶、色谱法、离心和/或提取。

54.如权利要求48-53中任一项所述的方法，其中所述酶混合物包含选自以下的酶中的至少两种:B2AF03、CIP1、CIP2、Cel1a、Cel3a、Cel5a、Cel6a、Cel7a、Cel7b、Cel12a、Cel45a、Cel74a、paMan5a、paMan26a和膨胀蛋白。

55.如权利要求48-54中任一项所述的方法，其中所述生物质包括以下中的一种或多种:农业产品或废弃物、纸产品或废弃物、林业产品或一般性废弃物或其任何组合，其中:

a)农业产品或废弃物包括甘蔗、黄麻、大麻、亚麻、竹子、剑麻、苜蓿、干草、秘鲁胡萝卜、荞麦、香蕉、大麦、木薯、野葛、圆齿酢酱草、西米、高粱、马铃薯、甜薯、芋头、山药、菜豆、蚕豆、扁豆、豌豆、草、柳枝稷、芒草、绳草、草芦、谷物残渣、菜籽秆、小麦秆、大麦秆、燕麦秆、稻草、玉米穗轴、玉米秸杆、玉米纤维、椰子毛、甜菜浆、甘蔗渣、大豆秸杆、谷物残渣、稻壳、燕麦壳、小麦谷壳、大麦壳或细麸或其组合;

b)纸产品或废弃物包括纸、着色纸、装料纸、涂覆纸、填充纸、杂志、印刷品、打印纸、涂塑纸、卡片坯料、卡纸板、纸板或纸浆或其组合;

c)林业产品包括白杨木、创花板、木片或锯屑或其组合;并且

d)一般性废弃物包括粪肥、污水或垃圾或其组合。

56.如权利要求48-55中任一项所述的方法，其进一步包括在引入所述微生物或所述酶混合物之前处理所述生物质以降低所述生物质的不顺应性的步骤，其中所述处理包括用电子进行轰击、声波处理、氧化、热解、蒸汽爆炸、化学处理、机械处理或冷冻研磨。

57.如权利要求48-56中任一项所述的方法，其中产生生物质降解酶的所述微生物来自选自以下属的种：芽孢杆菌属、鬼伞属、毁丝霉属、头孢霉属、柱顶孢霉属、青霉属、曲霉属、假单胞菌属、腐质霉属、镰刀菌属、梭孢壳属、枝顶孢属、金孢子菌属或木霉属。

58.如权利要求48-57中任一项所述的方法，其中产生生物质降解酶的所述微生物选自曲霉属、特异腐质霉(嗜热柱顶孢霉)灰盖鬼伞、尖孢镰刀菌、嗜热毁丝霉、大型亚灰树花菌、太瑞斯梭孢壳霉、桃色枝顶孢、Acremonium acremonium、Acremonium brachypenium、Acremonium dichromosporum、Acremonium obclavatum、Acremonium pinkertoniae、灰粉枝顶孢、Acremonium incoloratum、棕色枝顶孢、Chrysosporium lucknowense、绿色木霉、里氏木霉或康宁木霉。

59.如权利要求48-58中任一项所述的方法，其中已通过将所述微生物与诱导生物质样品在相较于未诱导微生物，适于使生物质降解酶的产生增加的条件下组合来诱导所述微生物以产生生物质降解酶。

60.如权利要求48-59中任一项所述的方法，其中所述诱导生物质样品包括以下中的一种或多种：农业产品或废弃物、纸产品或废弃物、林业产品或一般性废弃物或其任何组合，其中：

a)农业产品或废弃物包括甘蔗黄麻、大麻、亚麻、竹子、剑麻、苜蓿、干草、秘鲁胡萝卜、荞麦、香蕉、大麦、木薯、野葛、圆齿酢酱草、西米、高粱、马铃薯、甜薯、芋头、山药、菜豆、蚕豆、扁豆、豌豆、草、柳枝稷、芒草、绳草、草芦、谷物残渣、菜籽秆、小麦秆、大麦秆、燕麦秆、稻草、玉米穗轴、玉米秸杆、玉米纤维、椰子毛、甜菜浆、甘蔗渣、大豆秸杆、谷物残渣、稻壳、燕麦壳、小麦谷壳、大麦壳或细麸或其组合；

b)纸产品或废弃物包括纸、着色纸、装料纸、涂覆纸、填充纸、杂志、印刷品、打印纸、涂塑纸、卡片坯料、卡纸板、纸板或纸浆或其组合；

c)林业产品包括白杨木、创花板、木片或锯屑或其组合；并且

d)一般性废弃物包括粪肥、污水或垃圾或其组合。

被增溶酶及其用途

[0001] 相关申请

[0002] 本申请要求2014年9月26日提交的美国临时申请号62/055,702的权益;所述美国临时申请的整个内容据此以引用的方式并入。

[0003] 序列表

[0004] 本申请含有已以电子方式以ASCII格式提交,并且据此以引用的方式整体并入的序列表。所述ASCII拷贝于2015年9月1日创建,命名为X2002-7003WO_SL.txt,并且大小是76,946个字节。

发明领域

[0005] 本发明大体上涉及包含从包涵体增溶并具有生物质降解活性的具有生物质降解活性的多肽的混合物,以及用于产生本文所述的混合物的方法。本发明也提供用于使用此类混合物例如来加工生物质材料的方法。

[0006] 发明背景

[0007] 诸如纤维素酶、木聚糖酶和木质素酶的生物质降解酶对于诸如原料的生物质的降解是重要的。纤维素和木质纤维素材料在许多应用中被大量生产、加工和使用。此类材料常常被使用一次,接着作为废弃物丢弃,或仅被视为是废弃材料,例如污水、甘蔗渣、锯屑和秸杆。

[0008] 发明概述

[0009] 重组蛋白在诸如大肠杆菌(E.coli)的宿主细胞中高水平表达可导致重组蛋白在宿主细胞内积聚成不溶性聚集物。这些不溶性聚集物被称为包涵体,并且也可含有其它组分,诸如宿主细胞内源性蛋白质、核糖体组分、核酸和细胞碎片。从包涵体增溶重组蛋白可通过用高浓度的破坏氢键和疏水性相互作用的增溶剂诸如尿素处理来实现。然而,用诸如尿素的增溶剂处理可导致蛋白质变性和酶活性丧失。因此,重组蛋白聚集成包涵体可降低可从宿主细胞分离的具有酶活性的重组蛋白的产率。

[0010] 本发明至少部分地基于以下惊人发现:已通过诸如尿素的增溶剂来从包涵体增溶的异源性表达的纤维二糖酶保留纤维二糖酶活性。因此,本文所述的用于增溶异源性表达的纤维二糖酶或其它生物质降解酶的方法适用于使具有生物质降解活性的异源性表达的酶的产率增加例如30-40%。此外,由于添加被增溶生物质降解酶例如纤维二糖酶而存在例如尿素的增溶剂并不会不利地影响用于使生物质转化成糖产物的糖化反应和/或产物的产率。

[0011] 因此,在一个方面,本公开的特征在于一种包含具有生物质降解活性的一种多肽或多种多肽和例如尿素的增溶剂的混合物,其中相较于天然多肽,所述多肽或多种所述多肽具有至少8-10%生物质降解活性。

[0012] 在一个实施方案中,混合物进一步包含一种或多种与包涵体相关的蛋白质。或者,在一个实施方案中,混合物不包含一种或多种与包涵体相关的蛋白质。在一个实施方案中,混合物进一步包含细胞碎片、一种或多种核糖体组分、一种或多种宿主蛋白质例如由宿主

细胞内源性表达的蛋白质、和/或宿主核酸例如DNA和/或RNA。

[0013] 在一个实施方案中，生物质降解活性是纤维二糖酶活性、木质素酶活性、内切葡聚糖酶活性、纤维二糖水解酶活性或木聚糖酶活性。

[0014] 在一个实施方案中，多肽是部分未折叠的，部分错误折叠的，或部分变性的。

[0015] 在另一方面，本公开的特征在于一种包含具有与SEQ ID NO:1具有至少90％同一性的氨基酸序列的一种多肽或多种多肽和例如尿素的增溶剂的混合物，其中所述多肽或多种所述多肽具有至少20％的天然多肽例如SEQ ID NO:1或来自里氏木霉(T.reesei)的Cel3a的活性。举例来说，混合物进一步包含一种或多种与包涵体相关的蛋白质。或者，混合物不包含一种或多种与包涵体相关的蛋白质。混合物可进一步包含以下中的一种或多种：细胞碎片、一种或多种核糖体组分、一种或多种宿主蛋白质例如由宿主细胞内源性表达的蛋白质、和/或宿主核酸例如DNA和/或RNA。与SEQ ID NO:1具有至少90％同一性的多肽可为部分未折叠的，部分错误折叠的，或部分变性的。

[0016] 在一个实施方案中，多肽包含与SEQ ID NO:1具有至少90％同一性的氨基酸序列。在一个实施方案中，多肽包括来自里氏木霉的Cel3A酶或其功能性变体或片段。在一个实施方案中，Cel3A酶包含氨基酸序列SEQ ID NO:1(例如由所述氨基酸序列SEQ ID NO:1组成)。在一个实施方案中，多肽由包含与SEQ ID NO:2或SEQ ID NO:3的至少90％同一性(例如由SEQ ID NO:2或SEQ ID NO:3组成)的核酸序列编码。

[0017] 在一个实施方案中，多肽是无糖基化的。

[0018] 在一个实施方案中，例如尿素的增溶剂以0.2M-6M之间的浓度存在于混合物中。

[0019] 在一个实施方案中，混合物进一步包含至少一种具有生物质降解活性的额外多肽或产生一种或多种具有生物质降解活性的酶的微生物。在一个实施方案中，额外多肽选自木质素酶、内切葡聚糖酶、纤维二糖水解酶、纤维二糖酶和木聚糖酶或其任何组合。在一个实施方案中，额外多肽选自：

[0020] a.包含与SEQ ID NO:1具有至少90％同一性的氨基酸序列(例如由所述氨基酸序列组成)的多肽；

[0021] b.来自里氏木霉的Cel3A酶或其功能性变体或片段；或

[0022] c.由包含SEQ ID NO:2或SEQ ID NO:3(例如由SEQ ID NO:2或SEQ ID NO:3组成)的核酸序列编码的多肽。

[0023] 在一个实施方案中，额外多肽是无糖基化的。

[0024] 在一个实施方案中，额外多肽是糖基化的。

[0025] 在一个方面，本公开的特征在于一种用于产生本文所述的包含具有生物质降解活性的多肽、一种或多种与包涵体相关的蛋白质、和例如尿素的增溶剂的混合物的方法，其中所述方法包括使表达具有生物质降解活性的所述多肽的细胞或其溶解产物与适于增溶所述多肽的浓度的例如尿素的增溶剂接触。在一个实施方案中，方法进一步包括溶解细胞以获得溶解产物，使所述溶解产物的可溶性部分与不溶性部分分离，以及将所述不溶性部分再混悬于例如尿素的增溶剂中。在一个实施方案中，例如尿素的增溶剂的浓度介于0.2M-6M之间，例如是6M。

[0026] 在一个实施方案中，生物质降解活性是纤维二糖酶活性、木质素酶活性、内切葡聚糖酶活性、纤维二糖水解酶或木聚糖酶活性。

[0027] 在一个实施方案中，多肽包含与SEQ ID NO:1具有至少90％同一性的氨基酸序列。在一个实施方案中，多肽包括来自里氏木霉的Cel3A或其功能性变体或片段。

[0028] 在一个实施方案中，多肽是无糖基化的。

[0029] 在一个方面，本公开的特征在于一种用于产生具有生物质降解活性的多肽的方法，其包括在细胞中表达所述多肽，以及使所述细胞或其溶解产物与适于增溶所述多肽的浓度的例如尿素的增溶剂接触。

[0030] 在另一方面，本公开的特征在于一种用于产生具有生物质降解活性的多肽的方法，其包括提供已被遗传修饰来产生至少一种具有生物质降解活性的多肽的细胞，其中至少一部分具有生物质降解活性的所述多肽见于包涵体中，以及使含有所述包涵体的所述细胞或其溶解产物与适于增溶所述多肽的浓度的例如尿素的增溶剂接触。

[0031] 在一个实施方案中，本文公开的方法进一步包括溶解细胞以获得溶解产物，使所述溶解产物的可溶性部分与不溶性部分分离，以及将所述不溶性部分再混悬于例如尿素的增溶剂中。在一个实施方案中，例如尿素的增溶剂的浓度介于0.2M-6M之间，例如是6M。

[0032] 在一个实施方案中，生物质降解活性是纤维二糖酶活性、木质素酶活性、内切葡聚糖酶活性、纤维二糖水解酶活性或木聚糖酶活性。

[0033] 在一个实施方案中，无糖基化多肽包含与SEQ ID NO:1具有至少90％同一性的氨基酸序列(例如由所述氨基酸序列组成)。在一个实施方案中，无糖基化多肽包括来自里氏木霉的Cel3A或其功能性变体或片段。

[0034] 在一个实施方案中，细胞是原核或细菌细胞，例如大肠杆菌细胞、origami大肠杆菌细胞。

[0035] 在一个实施方案中，多肽是无糖基化的。

[0036] 在一个方面，本公开的特征在于一种从生物质产生产物(例如氢气、糖、醇等)(或使生物质转化成产物)的方法，其包括使生物质与本文所述的包含具有生物质降解活性的多肽、一种或多种与包涵体相关的蛋白质、和例如尿素的增溶剂的混合物，以及任选地，与产生一种或多种生物质降解酶的微生物和/或包含生物质降解酶的酶混合物在适于产生所述产物的条件下接触。

[0037] 在一个实施方案中，方法进一步包括在使生物质与本文所述的包含具有生物质降解活性的多肽、一种或多种与包涵体相关的蛋白质、和例如尿素的增溶剂的混合物接触之前，用电子束处理生物质的步骤。

[0038] 在一个实施方案中，产物是糖产物。在一个实施方案中，糖产物是葡萄糖和/或木糖。

[0039] 在一个实施方案中，方法进一步包括分离产物的步骤。在一个实施方案中，分离产物的步骤包括沉淀、结晶、色谱法、离心和/或提取。

[0040] 在一个实施方案中，酶混合物包含选自以下的酶中的至少两种：B2AF03、CIP1、CIP2、Cel1a、Cel3a、Cel5a、Cel6a、Cel7a、Cel7b、Cel12a、Cel45a、Cel74a、paMan5a、paMan26a、膨胀蛋白(Swollenin)。

[0041] 在一个实施方案中，生物质包括淀粉材料、甘蔗、农业废弃物、纸、纸产品、纸废弃物、纸浆、着色纸、装料纸、涂覆纸、填充纸、杂志、印刷品、打印纸、涂塑纸(polycoated paper)、卡片坯料、卡纸板、纸板、棉花、木材、创花板、林业废弃物、锯屑、白杨木、木片、草、

柳枝稷、芒草、绳草、草芦(reed canary grass)、谷物残渣、稻壳、燕麦壳、小麦谷壳、大麦壳、农业废弃物、青贮饲料、菜籽秆(canola straw)、小麦秆、大麦秆、燕麦秆、稻草、黄麻、大麻、亚麻、竹子、剑麻、蕉麻(abaca)、玉米穗轴、玉米秸杆、大豆秸杆、玉米纤维、苜蓿、干草、椰子毛、糖加工残渣、甘蔗渣、甜菜浆、龙舌兰渣(agave bagasse)、海藻、海草、粪肥、污水、垃圾(offal)、农业或工业废弃物、秘鲁胡萝卜(arracacha)、荞麦、香蕉、大麦、木薯、野葛(kudzu)、圆齿酢酱草(oca)、西米、高粱、马铃薯、甜薯、芋头、山药、菜豆、蚕豆、扁豆、豌豆或其任何组合。

[0042] 在一个实施方案中,生物质包括淀粉材料或包括纤维素组分的淀粉材料。在一些实施方案中,生物质包括以下中的一种或多种:农业产品或废弃物、纸产品或废弃物、林业产品或一般性废弃物或其任何组合;其中:a)农业产品或废弃物包括甘蔗、黄麻、大麻、亚麻、竹子、剑麻、苜蓿、干草、秘鲁胡萝卜、荞麦、香蕉、大麦、木薯、野葛、圆齿酢酱草、西米、高粱、马铃薯、甜薯、芋头、山药、菜豆、蚕豆、扁豆、豌豆、草、柳枝稷、芒草、绳草、草芦、谷物残渣、菜籽秆、小麦秆、大麦秆、燕麦秆、稻草、玉米穗轴、玉米秸杆、玉米纤维、椰子毛、甜菜浆、甘蔗渣、大豆秸杆、谷物残渣、稻壳、燕麦壳、小麦谷壳、大麦壳或细麸(beeswing)或其组合;b)纸产品或废弃物包括纸、着色纸、装料纸、涂覆纸、填充纸、杂志、印刷品、打印纸、涂塑纸、卡片坯料、卡纸板、纸板或纸浆或其组合;c)林业产品包括白杨木、刨花板、木片或锯屑或其组合;并且d)一般性废弃物包括粪肥、污水或垃圾或其组合。

[0043] 在一个实施方案中,方法进一步包括在引入微生物或酶混合物之前处理生物质以降低生物质的不顺应性的步骤,例如通过用电子进行轰击、声波处理、氧化、热解、蒸汽爆炸、化学处理、机械处理和/或冷冻研磨来处理生物质。

[0044] 在一个实施方案中,产生生物质降解酶的微生物来自选自以下属的种:芽孢杆菌属(Bacillus)、鬼伞属(Coprinus)、毁丝霉属(Myceliophthora)、头孢霉属(Cephalosporium)、柱顶孢霉属(Scytalidium)、青霉属(Penicillium)、曲霉属(Aspergillus)、假单胞菌属(Pseudomonas)、腐质霉属(Humicola)、镰刀菌属(Fusarium)、梭孢壳属(Thielavia)、枝顶孢属(Acremonium)、金孢子菌属(Chrysosporium)或木霉属(Trichoderma)。在一个实施方案中,产生生物质降解酶的微生物选自曲霉属、特异腐质霉(Humicola insolens)(嗜热柱顶孢霉(Scytalidium thermophilum))、灰盖鬼伞(Coprinus cinereus)、尖孢镰刀菌(Fusarium oxysporum)、嗜热毁丝霉(Myceliophthora thermophila)、大型亚灰树花菌(Meripilus giganteus)、太瑞斯梭孢壳霉(Thielavia terrestris)、桃色枝顶孢(Acremonium persicinum)、Acremonium acremonium、Acremonium brachypenium、Acremonium dichromosporum、Acremonium obclavatum、Acremonium pinkertoniae、灰粉枝顶孢(Acremonium roseogriseum)、Acremonium incoloratum、棕色枝顶孢(Acremonium furatum)、Chrysosporium lucknowense、绿色木霉(Trichoderma viride)、里氏木霉(Trichoderma reesei)或康宁木霉(Trichoderma koningii)。

[0045] 在一个实施方案中,已通过将微生物与诱导生物质样品在相较于未诱导微生物,适于使生物质降解酶的产生增加的条件下组合来诱导微生物产生生物质降解酶。在一个实施方案中,诱导生物质样品包括淀粉材料、甘蔗、纸、纸产品、纸废弃物、纸浆、着色纸、装料纸、涂覆纸、填充纸、杂志、印刷品、打印纸、涂塑纸、卡片坯料、卡纸板、纸板、棉花、木材、创

花板、林业废弃物、锯屑、白杨木、木片、草、柳枝稷、芒草、绳草、草芦、谷物残渣、稻壳、燕麦壳、小麦谷壳、大麦壳、农业废弃物、青贮饲料、菜籽秆、小麦秆、大麦秆、燕麦秆、稻草、黄麻、大麻、亚麻、竹子、剑麻、蕉麻、玉米穗轴、玉米秸杆、大豆秸杆、玉米纤维、苜蓿、干草、椰子毛、糖加工残渣、甘蔗渣、甜菜浆、龙舌兰渣、海藻、海草、粪肥、污水、垃圾、农业或工业废弃物、秘鲁胡萝卜、荞麦、香蕉、大麦、木薯、野葛、圆齿酢酱草、西米、高粱、马铃薯、甜薯、芋头、山药、菜豆、蚕豆、扁豆、豌豆或其任何组合。

[0046] 在一个实施方案中，诱导生物质包括淀粉材料或包括纤维素组分的淀粉材料。在一些实施方案中，诱导生物质包括以下中的一种或多种：农业产品或废弃物、纸产品或废弃物、林业产品或一般性废弃物或其任何组合；其中：a)农业产品或废弃物包括甘蔗、黄麻、大麻、亚麻、竹子、剑麻、苜蓿、干草、秘鲁胡萝卜、荞麦、香蕉、大麦、木薯、野葛、圆齿酢酱草、西米、高粱、马铃薯、甜薯、芋头、山药、菜豆、蚕豆、扁豆、豌豆、草、柳枝稷、芒草、绳草、草芦、谷物残渣、菜籽秆、小麦秆、大麦秆、燕麦秆、稻草、玉米穗轴、玉米秸杆、玉米纤维、椰子毛、甜菜浆、甘蔗渣、大豆秸杆、谷物残渣、稻壳、燕麦壳、小麦谷壳、大麦壳或细麸或其组合；b)纸产品或废弃物包括纸、着色纸、装料纸、涂覆纸、填充纸、杂志、印刷品、打印纸、涂塑纸、卡片坯料、卡纸板、纸板或纸浆或其组合；c)林业产品包括白杨木、刨花板、木片或锯屑或其组合；并且d)一般性废弃物包括粪肥、污水或垃圾或其组合。

[0047] 在一个实施方案中，相对于本领域中使用的当前方法，本发明提供优势。这些优势包括提供对将通常被丢弃的不溶性酶的获取，从而增加保留酶活性的所需蛋白质的产率，酶被纯化以达成较清洁的下游加工，以及生物体选择(例如，增加可已由于倾向于产生包涵体而先前从使用排除的生物体的可用性)。

[0048] 附图简述

[0049] 图1是显示对被增溶Cel3a的IMAC纯化的结果的色谱图。纯化的被增溶Cel3a峰用箭号指示。

[0050] 图2是SDS-PAGE胶凝的照片，其显示IMAC纯化的不同部分中的蛋白质。泳道1显示分子量标准物。泳道2显示来自可溶性部分的纯化Cel3a。泳道3显示由对不溶性部分进行的IMAC纯化获得的流过物。泳道4显示来自不溶性部分的纯化的被增溶Cel3a。

[0051] 图3是比较纯化的可溶性Cel3a和来自不溶性部分的纯化的被增溶Cel3a的纤维二糖酶活性的图。

[0052] 图4是比较纯化的可溶性Cel3a、不溶性部分的洗涤部分、和从不溶性部分增溶的未经纯化的Cel3a的纤维二糖酶活性的图。

[0053] 详细描述

[0054] 定义

[0055] 除非另外定义，否则本文所用的所有技术和科学术语都具有与由本发明所属领域中的普通技术人员通常理解相同的含义。

[0056] 术语"一个/一种(a/an)"是指冠词的一个或超过一个(即至少一个)语法对象。举例来说，"一种要素"意指一种要素或超过一种要素。

[0057] 如本文所用的术语"无糖基化"是指例如多肽的分子在一个或多个当所述分子在它的天然环境中产生时连接有聚糖的位点处未被糖基化(即它包含不连接于糖基化基团的羟基或其它官能团)。在一些实施方案中，无糖基化分子不具有任何连接的聚糖。在一个实

施方案中，已使分子改变或突变以致分子不能被糖基化，例如使一个或多个糖基化位点突变以致聚糖不能连接于糖基化位点。在另一实施方案中，可例如通过酶促过程，例如通过与移除聚糖或具有去糖基化活性的酶一起孵育来从分子移除连接的聚糖。在另一实施方案中，可例如通过使用糖基化抑制剂(其抑制糖基化酶)来抑制分子的糖基化。在另一实施方案中，例如多肽的分子可由不进行糖基化的宿主细胞(例如大肠杆菌)产生。举例来说，当蛋白质中的一个或多个在里氏木霉中产生Ce13A酶时通常具有连接于它的聚糖基团的位点不具有连接在那个位点处的聚糖时，所述Ce13A酶是无糖基化的。

[0058] 如本文所用的术语"生物质"是指任何非化石化有机物质。各种类型的生物质包括植物生物质(例如木质纤维素和纤维素生物质)、微生物生物质、动物生物质(任何动物副产物、动物废弃物等)和城市废弃物生物质(诸如金属和玻璃的可再循环物被移除的住宅和轻型商业废物)。植物生物质是指任何植物源性有机物质(木质或非木质)。植物生物质可包括但不限于农业或粮食作物(例如甘蔗、糖用甜菜或玉米粒)或由其获得的提取物(例如来自甘蔗的糖和来自玉米的玉米淀粉)、农业作物废弃物和残渣，诸如玉米秸杆、小麦秆、稻草、甘蔗渣等。植物生物质进一步包括但不限于树木、木质能源作物、木材废弃物和残渣诸如软木林疏伐物、含有树皮的废弃物、锯屑、造纸和纸浆行业废弃物物流、木材纤维等。另外，诸如柳枝稷等的草类作物具有作为另一植物生物质来源加以大规模产生的潜力。对于市区，最具潜力的植物生物质原料包括庭院废弃物(例如草修剪物、叶子、树木修剪物和灌丛)和蔬菜加工废弃物。

[0059] 如本文所用的术语"生物质降解酶"是指将本文所述的生物质物质的组分分解成中间体或最终产物的酶。举例来说，生物质降解酶包括至少木质素酶、内切葡聚糖酶、纤维二糖酶、木聚糖酶和纤维二糖水解酶。生物质降解酶由广泛多种微生物产生，并且可从诸如里氏木霉的微生物分离。

[0060] 如本文所用的术语"生物质降解活性"是指将本文所述的生物质物质的组分分解成中间体或最终产物的酶活性。生物质降解活性包括至少木质素酶活性、内切葡聚糖酶活性、纤维二糖酶活性、纤维二糖水解酶活性和木聚糖酶活性。举例来说，具有生物质降解活性的多肽是纤维二糖酶，诸如来自里氏木霉的Ce13a。

[0061] 如本文所用的术语"纤维二糖酶"是指催化葡萄糖的二聚体、三聚体、四聚体、五聚体、六聚体、七聚体、八聚体或寡聚体;或葡萄糖和木糖的寡聚体水解成葡萄糖和/或木糖的酶。举例来说，纤维二糖酶是β-葡萄糖苷酶，其催化纤维二糖中的β-1,4键以释放两个葡萄糖分子。

[0062] 如本文所用的术语"纤维二糖酶活性"是指一个种类的纤维素酶催化纤维二糖水解成葡萄糖，例如催化β-D-葡萄糖残基水解以释放β-D-葡萄糖的活性。可根据本文例如在实施例4中所述的测定来测定纤维二糖酶活性。一个单位的纤维二糖酶活性可定义为[葡萄糖]g/L/[Ce13a]g/L/30分钟。

[0063] 如本文所用的术语"纤维二糖水解酶"是指水解纤维素中的糖苷键的酶。举例来说，纤维二糖水解酶是1,4-β-D-葡聚糖纤维二糖水解酶，其催化纤维素、纤维寡糖或任何含有β-1,4连接的葡萄糖的聚合物中的1,4-β-D-葡萄糖苷键的水解，从而从聚合物链释放寡糖。

[0064] 如本文所用的术语"纤维二糖水解酶活性"是指催化纤维素中的糖苷键的水解，具

体来说纤维素、纤维寡糖或任何含有β-1,4连接的葡萄糖的聚合物中的1,4-β-D-葡萄糖苷键的水解,以从糖链的末端,例如从链的还原性或非还原性末端释放纤维二糖的酶的活性。可根据本文所述的测定来测定纤维二糖水解酶活性。一个单位的纤维二糖水解酶活性可例如定义为每分钟从底物(例如阿维塞尔(Avicel))释放1μM葡萄糖当量的酶的量。

[0065] 如本文所用的术语"内切葡聚糖酶"是指催化内部β-1,4糖苷键的水解的酶。举例来说,内切葡聚糖酶是内切-1,4-(1,3;1,4)-β-D-葡聚糖4-葡聚糖水解酶,其催化纤维素、纤维素衍生物(诸如羧甲基纤维素和羟乙基纤维素)、地衣淀粉(lichenan)中的1,4-β-D-糖苷键;混合β-1,3葡聚糖(诸如谷类β-D-葡聚糖或木葡聚糖)和其它含有纤维素组分的植物材料中的β-1,4键的内水解。

[0066] 如本文所用的术语"内切葡聚糖酶活性"是指催化纤维素、纤维素衍生物(诸如羧甲基纤维素和羟乙基纤维素)、地衣淀粉的内部糖苷键例如内部β-1,4糖苷键;混合β-1,3葡聚糖(诸如谷类β-D-葡聚糖或木葡聚糖)和其它含有纤维素组分的植物材料中的β-1,4键的内水解的酶的活性。可根据本文所述的测定来测定内切葡聚糖酶活性。一个单位的内切葡聚糖酶活性可例如定义为每分钟从底物使还原性末端的浓度增加1μM的酶的量。

[0067] 如本文所用的术语"酶混合物"是指至少两种不同酶或某一酶的两种不同变体(例如某一酶的糖基化形式和无糖基化形式)的组合。本文提及的酶混合物包括至少本文所述的具有纤维二糖酶活性的无糖基化多肽。在一个实施方案中,酶混合物包括纤维二糖酶、内切葡聚糖酶、纤维二糖水解酶、木质素酶和/或木聚糖酶中的一种或多种。在一些实施方案中,酶混合物包括表达和例如分泌一种或多种酶的细胞,例如微生物。举例来说,酶混合物可包括本文所述的无糖基化多肽和表达和例如分泌一种或多种额外酶和/或所述多肽的变体的细胞,例如微生物。

[0068] 如本文所用的术语"包涵体"是指由例如宿主细胞的微生物产生的不溶性聚集物,其含有以下中的一种或多种:异源性表达的多肽例如具有生物质降解活性的多肽、细胞碎片、一种或多种核糖体组分、一种或多种由所述宿主细胞内源性表达的蛋白质、一种或多种核酸(RNA和/或DNA)或其任何组合。包涵体通常在高水平表达重组蛋白期间存在于宿主细胞例如细菌细胞中。见于包涵体中的异源性表达的多肽可为部分未折叠的,部分错误折叠的,或部分变性的。

[0069] 如本文所用的术语"木质素酶"是指诸如通过氧化反应来催化通常见于植物的细胞壁中的木质素的分解的酶。木质素酶包括木质素修饰酶、木质素过氧化酶和漆酶。

[0070] 如本文所用的术语"木质素酶活性"是指通过氧化反应来催化木质素和木质素样聚合物的分解的酶的活性。可根据本文所述的测定来测定木质素酶活性。

[0071] 术语"核酸"或"多核苷酸"可互换使用,并且是指呈单链或双链形式的脱氧核糖核酸(DNA)或核糖核酸(RNA)及其聚合物。除非明确限制,否则所述术语涵盖含有天然核苷酸的已知类似物的核酸,其与参照核酸具有类似结合性质,并且以与天然存在的核苷酸类似的方式进行代谢。除非另外指示,否则特定核酸序列也隐含地涵盖其保守修饰的变体(例如简并密码子取代)、等位基因、直系同源物、SNP和互补序列以及明确指示的序列。具体来说,简并密码子取代可通过产生其中一个或多个所选(或所有)密码子的第三位置被混合碱基和/或脱氧肌苷残基取代的序列来实现(Batzer等,Nucleic Acid Res.19:5081(1991);Ohtsuka等,J.Biol.Chem.260:2605-2608(1985);以及Rossolini等,Mol.Cell.Probes 8:

91-98(1994))。

[0072] 如本文所用的术语"可操作地连接"是指以下构型:其中控制或调控序列相对于编码多肽的核酸序列被放置在使得所述控制序列影响多肽(由DNA序列编码)的表达的位置处。在一个实施方案中,控制或调控序列在编码具有纤维二糖酶活性的多肽的核酸序列的上游。在一个实施方案中,控制或调控序列在编码具有纤维二糖酶活性的多肽的核酸序列的下游。

[0073] 术语"肽"、"多肽"和"蛋白质"可互换使用,并且是指包含由肽键共价连接的氨基酸残基组成的化合物。蛋白质或肽必须含有至少两个氨基酸,并且关于可构成蛋白质的序列或肽的序列的氨基酸的最大数目不设限制。多肽包括包含两个或更多个由肽键彼此接合的氨基酸的任何肽或蛋白质。"多肽"包括例如生物活性片段、大致上同源性多肽、寡肽、同二聚体、异二聚体、多肽的变体、修饰的多肽、衍生物、类似物、融合蛋白以及其它多肽。多肽包括天然肽、重组肽或其组合。"多种多肽"是指两种或更多种多肽,例如2、3、4、5、6、7、8、9、10、20、50、100、200或500种或更多种多肽。

[0074] 如本文所用的术语"启动子"是指由细胞的合成机构或引入的合成机构识别,为启始多核苷酸序列的特异性转录所需的DNA序列。

[0075] 如在本文中可互换使用的术语"调控序列"或"控制序列"是指为表达核酸产物所需的核酸序列。在一些情况下,这个序列可为启动子序列,并且在其它情况下,这个序列也可包括增强子序列和为表达基因产物所需的其它调控元件。调控/控制序列可例如是以调控方式(例如可诱导方式)来表达核酸产物的序列。

[0076] 术语"组成型"启动子是指当与编码多肽的多核苷酸可操作地连接时,在细胞的大多数或所有生理条件下导致在所述细胞中产生所述多肽的核苷酸序列。在一个实施方案中,多肽是具有纤维二糖酶活性的多肽。

[0077] 术语"诱导型"启动子是指当与编码多肽的多核苷酸可操作地连接时,仅当对应于启动子的诱导剂存在于细胞中时导致在所述细胞中大量产生所述多肽的核苷酸序列。在一个实施方案中,多肽是具有纤维二糖酶活性的多肽。

[0078] 术语"阻遏型"启动子是指当与编码多肽的多核苷酸可操作地连接时,仅直至对应于启动子的阻遏子存在于细胞中时导致在所述细胞中大量产生所述多肽的核苷酸序列。在一个实施方案中,多肽是具有纤维二糖酶活性的多肽。

[0079] 术语"增溶剂"是指具有破坏非共价键例如氢键、疏水性相互作用、范德华(van der Waals)相互作用、偶极-偶极相互作用、离子相互作用、π堆积作用或其任何组合的能力的试剂。破坏非共价键导致先前不溶性物质增溶或溶解至溶液中。具体来说,本文所用的增溶剂使已聚集成包涵体的本文所述的具有生物质降解活性的多肽溶解至溶液例如基于水的溶液或缓冲液中的能力增加。本文描述适合的增溶剂的实例。

[0080] 如本文所用的术语"木聚糖酶"是指水解含有木聚糖的材料的酶。木聚糖是包含木糖单元的多糖。木聚糖酶可为内切木聚糖酶、β-木糖苷酶、阿拉伯呋喃糖苷酶、α-葡萄糖醛酸苷酶、乙酰木聚糖酯酶、阿魏酸酯酶(feruloyl esterase)或α-葡萄糖醛酸基酯酶。

[0081] 如本文所用的术语"木聚糖酶活性"是指催化木聚糖和木聚糖样聚合物中的1,4-β-D-木糖苷键的内水解的酶的活性。可根据本文所述的测定来测定木聚糖酶活性。一个单位的木聚糖酶活性将每分钟从木聚糖释放1uM木糖当量。

[0082] 描述

[0083] 重组蛋白在诸如大肠杆菌的宿主细胞中高水平表达常常导致重组蛋白在宿主细胞内积聚成非活性、错误折叠以及不溶性聚集物。这些不溶性聚集物被称为包涵体，并且也可含有其它宿主细胞内源性组分，诸如蛋白质、核糖体组分、核酸和细胞碎片。多达70-80％的通过重组技术产生的蛋白质可形成包涵体，由此显著降低可易于从宿主细胞分离的活性重组蛋白的产率。

[0084] 从包涵体增溶重组蛋白可通过用破坏氢键和疏水性相互作用的离液剂，例如高浓度的尿素处理来实现。然而，所述增溶方法常常导致蛋白质变性和天然功能或酶活性损失。在移除离液剂之后，可溶性变性蛋白可再折叠成它们的天然状态，然而，使重组蛋白再折叠成具有酶活性的生物活性形式可为麻烦的，昂贵的，并且导致最终产物的回收率较低。

[0085] 本发明至少部分地基于以下惊人发现：已通过尿素来从包涵体增溶的异源性表达的纤维二糖酶保留纤维二糖酶活性。从包涵体回收异源性表达的纤维二糖酶使纤维二糖酶的总产率增加30-40％。此外，由于添加被增溶生物质降解酶例如纤维二糖酶而存在例如尿素的增溶剂并不会不利影响用于使生物质转化成糖产物的糖化反应和/或产物的产率。

[0086] 因此，本发明提供用于从包涵体增溶具有生物质降解活性的多肽的方法，其中所得被增溶多肽保留生物质降解活性，借此不需要再折叠所述多肽和移除增溶剂的额外处理步骤。本发明提供用于在保留酶活性的同时使从包涵体对异源性表达的生物质降解酶的回收增加，以及在用于例如通过糖化来使生物质转化成产物的方法中使用回收的生物质降解酶的方法。

[0087] 具有生物质降解活性的多肽

[0088] 本公开提供具有生物质降解活性的一种多肽、多种多肽。在各实施方案中，具有生物质降解活性的多肽或多种所述多肽与一种或多种增溶剂以混合物形式存在。一些混合物也可含有一种或多种与包涵体相关的蛋白质。在其它实施方案中，混合物不含有一种或多种与包涵体相关的蛋白质，例如从一种或多种与包涵体相关的蛋白质纯化具有生物质降解活性的多肽或多种所述多肽。

[0089] 举例来说，多肽具有纤维二糖酶活性、木质素酶活性、内切葡聚糖酶活性、纤维二糖水解酶活性或木聚糖酶活性。

[0090] 在一个实施方案中，多肽是纤维二糖酶。纤维二糖酶是水解它的底物例如纤维二糖中的β-1，4键以释放两个葡萄糖分子的酶。纤维二糖是葡萄糖的水溶性1，4-连接的二聚体。在一个实施方案中，多肽是Cel3a。Cel3a(也称为Bg1I)是在里氏木霉中鉴定的纤维二糖酶。以下提供Cel3a的氨基酸序列(GenBank登记号NW_006711153)：

[0091]

MGDSHSTSGASAEAVVPPAGTPWGTAYDKAKAALAKLNLQDKVGIVSGVGWNGGPCVGNTSPASKISYPSLCLQDGP
LGVRYSTGSTAFTPGVQAASTWDVNLIRERGQFIGEEVKASGIHVILGPVAGPLGKTPQGGRNWEGFGVDPYLTGIA
MGQTINGIQSVGVQATAKHYILNEQELNRETISSNPDDRTLHELYTWPFADAVQANVASVMCSYNKVNTTWACEDQY
TLQTVLKDQLGFPGYVMTDWNAQHTTVQSANSGLDMSMPGTDFNGNNRLwGPALTNAVNSNQVPTSRVDDMVTRILA
AWYLTGQDQAGYPSFNISRNVQGNHKTNVRAIARDGIVLLKNDANILPLKKPASIAVVGSAAIIGNHARNSPSCNDK
GCDDGALGMGWGSGAVNYPYFVAPYDAINTRASSQGTQVTLSNTDNTSSGASAARGKDVAIVFITADSGEGYITVEG
NAGDRNNLDPWHNGNALVQAVAGANSNVIVVVHSVGAIILEQILALPQVKAVVWAGLPSQESGNALVDVLWGDVSPS

GKLVYTIAKSPNDYNTRIVSGGSDSFSEGLFIDYKHFDDANITPRYEFGYGLSYTKFNYSRLSVLSTAKSGPATGAV
VPGGPSDLFQNVATVTVDIANSGQVTGAEVAQLYITYPSSAPRTPPKQLRGFAKLNLTPGQSGTATFNIRRRDLSYW
DTASQKWVVPSGSFGISVGASSRDIRLTSTLSVAGSGS

[0092] (SEQ ID NO:1)

[0093] 在一个实施方案中，多肽是木质素酶。木质素酶是分解木质素的酶，所述木质素是称为单木质醇的芳族醇的复杂聚合物，并且是植物和一些藻类的次生细胞壁的主要组成部分。木质素酶包括木质素过氧化物酶1,2-双(3,4-二甲氧基苯基)丙烷-1,3-二醇:过氧化氢氧化还原酶、二芳基丙烷加氧酶、木质素酶I、二芳基丙烷过氧化物酶、LiP、过氧化氢氧化还原酶(裂解C-C键)以及一些漆酶。木质素酶的实例包括来自里氏木霉的CIP2;来自黄孢原毛平革菌(Phanerochaete chrysosporium)的LPOA、GLG2、GLG4、LIPA、GLG5、GLG3、GLG6和LIPB;来自射纹革菌(Phelbia radiate)的木质素酶-3;来自云芝(Coriolus versicolor)的木质素酶A和B;以及云芝LPG I和LPGIV。

[0094] 在一个实施方案中，多肽是内切葡聚糖酶。内切葡聚糖酶是催化纤维素的水解的酶。具体来说，内切葡聚糖酶裂解纤维素链的内部键。内切葡聚糖酶由真菌、细菌和原虫产生。内切葡聚糖酶也称为β-1-4内切葡聚糖酶、4-β-D-葡聚糖纤维二糖水解酶、外切-纤维二糖水解酶、β-1,4-葡聚糖纤维二糖水解酶、β-1,4-葡聚糖纤维二糖基水解酶、1,4-β-葡聚糖纤维二糖苷酶、外切葡聚糖酶、微晶纤维素酶、CBH 1、C1纤维素酶、纤维二糖水解酶I、纤维二糖水解酶、外切-β-1,4-葡聚糖纤维二糖水解酶、1,4-β-D-葡聚糖纤维二糖水解酶或纤维二糖苷酶。内切葡聚糖酶的实例包括来自里氏木霉的Cel5A、Cel5B、Cel7B、Cel12A、Cel45A、Cel61A、Cel61B和Cel74A。

[0095] 在一个实施方案中，多肽是纤维二糖水解酶，也称为外切葡聚糖酶。纤维二糖水解酶催化含有1-4-β-D-葡萄糖苷键的寡糖例如纤维素和纤维四糖中的那个键的水解，由此从链的非还原性末端释放纤维二糖。纤维二糖水解酶的实例包括来自里氏木霉的纤维二糖水解酶(CBHI)和纤维二糖水解酶II(CBHII)。

[0096] 在一个实施方案中，多肽是木聚糖酶。木聚糖酶也称为内切-(1-4)-β-木聚糖4-木聚糖水解酶、内切-1,4-木聚糖酶、内切-1,4-β-木聚糖酶、β-1,4-木聚糖酶、内切-1,4-β-D-木聚糖酶、1,4-β-木聚糖木聚糖水解酶、β-木聚糖酶、β-1,4-木聚糖木聚糖水解酶、β-D-木聚糖酶。木聚糖酶分解植物细胞壁的称为半纤维素的组分，例如降解多糖，诸如木聚糖例如β-1,4-木聚糖、葡糖醛酸木聚糖、阿糖基木聚糖、葡甘露聚糖和木葡聚糖，以释放木糖。木聚糖酶的实例包括来自里氏木霉的Xyn1、Xyn2和Xyn3;以及来自船蛆杆菌(Terendinibacter turnerae)T7901的TERTU_1599、TERTU_3603、TERTU_2546和TERTU_4506。

[0097] 本公开也提供本文所述的具有生物质降解活性的多肽的功能性变体。在一个实施方案中，功能性变体具有与本文所述的生物质降解酶或其功能性片段具有至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%同一性，例如与本文所述的生物质降解酶或其功能性片段具有至少80%、至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%同一性的氨基酸序列。

[0098] 在一个实施方案中，功能性变体具有与由里氏木霉产生的Cel3a或SEQ ID NO:1或

其功能性片段具有至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少91%同一性、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%同一性的氨基酸序列。

[0099] 在两个或更多个氨基酸或核酸序列的情形下的同一性百分比是指相同的两个或更多个序列。如使用以下序列比较算法中的一种或通过手动比对和目视检查所测量，如果当在比较窗或指定区域上进行比较和对准以达成最大对应性时，两个序列具有指定百分比的相同氨基酸残基或核苷酸(例如在指定区域上，或当未指定时在整个序列上，60%同一性，任选70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%同一性)，那么两个序列"大致上同一"。任选地，在长度是至少约50个核苷酸、100个核苷酸、150个核苷酸的区域上存在同一性。更优选地，在长度是至少约200或更多个氨基酸、或至少约500或1000或更多个核苷酸的区域上存在同一性。

[0100] 对于序列比较，通常一个序列充当一个或多个测试序列与其进行比较的参照序列。当使用序列比较算法时，将测试序列和参照序列输入计算机中，必要时指定子序列坐标，并且指定序列算法程序参数。可使用默认程序参数，或可指定替代参数。序列比较算法接着基于程序参数计算测试序列相对于参照序列的序列同一性百分比。使序列对准以供比较的方法在本领域中是熟知的。可例如通过Smith和Waterman，(1970)Adv.Appl.Math.2:482c的局部同源性算法，通过Needleman和Wunsch，(1970)J.Mol.Biol.48:443的同源性比对算法，通过Pearson和Lipman，(1988)Proc.Nat'l.Acad.Sci.USA 85:2444的类似性搜索方法，通过这些算法的计算机化执行程序(Wisconsin Genetics软件包，Genetics Computer Group，575Science Dr.，Madison，WI)中的GAP、BESTFIT、FASTA和TFASTA)，或通过手动比对和目视检查(参见例如Brent等，(2003)Current Protocols in Molecular Biology)来进行供比较的序列的最优比对。

[0101] 适于确定序列同一性百分比和序列类似性百分比的算法的两个实例是BLAST和BLAST 2.0算法，其分别描述于Altschul等，(1977)Nuc.Acids Res.25:3389-3402；以及Altschul等，(1990)J.Mol.Biol.215:403-410中。用于进行BLAST分析的软件可通过国家生物技术信息中心(National Center for Biotechnology Information)公开获得。

[0102] 功能性变体可包含一个或多个突变，以使变体保留比由酶所源于的微生物产生的本文所述的生物质降解酶的生物质降解活性更好的生物质降解活性。在一个实施方案中，功能性变体具有如同如由大肠杆菌产生的生物质降解酶的生物质降解活性的至少10%、至少20%、至少30%、至少40%、至少50%、至少60%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%或至少99%(例如至少80%、至少85%、至少90%、至少95%或至少99%)。在各实施方案中，功能性变体具有如同由大肠杆菌或酶所源于的微生物产生的生物质降解酶的生物质降解活性的至少200%、至少300%、至少400%、至少500%、至少1000%或更多。可使用本文所述的功能性测定测试生物质降解活性。在一个实施方案中，功能性变体保留比如由里氏木霉产生的Cel3a的纤维二糖酶活性更好的纤维二糖酶活性。在另一实施方案中，功能性变体具有如同Cel3a或如由大肠杆菌产生的包含SEQ ID NO:1的酶的纤维二糖酶活性的至少10%、至少20%、至少30%、至少40%、至少50%、至少60%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%或至少99%(例如至少80%、至少

85％、至少90％、至少95％或至少99％)。在各实施方案中，相较于本文所述的生物质降解酶，功能性变体具有增加的生物质降解活性，例如是本文所述的生物质降解酶的生物质降解活性(例如是如同Ce13a或由大肠杆菌或酶所源于的微生物产生的包含SEQ ID NO:1的酶的纤维二糖酶活性)的至少200％、至少300％、至少400％、至少500％、至少1000％或更多。

[0103] 存在于功能性变体中的突变包括氨基酸取代、添加和缺失。可通过本领域中已知的标准技术，诸如定点诱变和PCR介导的诱变来引入突变。突变可为保守性氨基酸取代，其中氨基酸残基被具有类似侧链的氨基酸残基替换。本领域中已定义具有类似侧链的氨基酸残基的家族。这些家族包括具有碱性侧链的氨基酸(例如赖氨酸、精氨酸、组氨酸)、具有酸性侧链的氨基酸(例如天冬氨酸、谷氨酸)、具有不带电荷极性侧链的氨基酸(例如甘氨酸、天冬酰胺、谷氨酰胺、丝氨酸、苏氨酸、酪氨酸、半胱氨酸、色氨酸)、具有非极性侧链的氨基酸(例如丙氨酸、缬氨酸、亮氨酸、异亮氨酸、脯氨酸、苯丙氨酸、甲硫氨酸)、具有β分支侧链的氨基酸(例如苏氨酸、缬氨酸、异亮氨酸)和具有芳族侧链的氨基酸(例如酪氨酸、苯丙氨酸、色氨酸、组氨酸)。因此，本公开的具有纤维二糖酶活性的多肽内的一个或多个氨基酸残基可被来自同一侧链家族的其它氨基酸替换，并且所得多肽保留与野生型多肽的纤维二糖酶活性类似的纤维二糖酶活性(例如至少80％、85％、90％、95％或99％的纤维二糖酶活性)。或者，突变可为其中氨基酸残基被具有不同侧链的氨基酸残基替换的氨基酸取代。

[0104] 所述突变可改变或影响生物质降解酶例如纤维二糖酶、木质素酶、内切葡聚糖酶或纤维二糖水解酶的各种酶特征。举例来说，所述突变可改变或影响生物质降解酶的生物质降解活性、热稳定性、最优反应pH、酶动力学或底物识别。在一些实施方案中，相较于生物质降解酶例如由里氏木霉产生的纤维二糖酶和/或在大肠杆菌中产生的SEQ ID NO:1，突变使变体的生物质降解活性增加。在一些实施方案中，相较于野生型生物质降解酶例如纤维二糖酶和/或在大肠杆菌中产生的SEQ ID NO:1，突变使变体的热稳定性增加或降低。在一个实施方案中，相较于野生型生物质降解酶例如野生型纤维二糖酶和/或在大肠杆菌中产生的SEQ ID NO:1，突变改变变体最优进行生物质降解反应所处的pH范围。在一个实施方案中，相较于野生型生物质降解酶例如野生型纤维二糖酶和/或在大肠杆菌中产生的SEQ ID NO:1，突变使生物质降解反应的动力学(例如kcat、KM或KD)增加或降低。在一个实施方案中，相较于野生型纤维二糖酶和/或在大肠杆菌中产生的SEQ ID NO:1，突变使纤维二糖酶识别或结合底物(例如纤维二糖)的能力增加或降低。

[0105] 本发明也提供如本文所述的具有生物质降解活性例如纤维二糖酶活性的多肽(例如Ce13a或SEQ ID NO:1)的功能性月段。本领域普通技术人员可易于设想如本文所述的具有生物质降解活性的多肽的含有导致酶活性的功能性结构域的片段将保留例如生物质降解活性的功能活性，并且因此，所述片段涵盖在本发明中。在一个实施方案中，功能性月段的长度是至少700个氨基酸、至少650个氨基酸、至少600个氨基酸、至少550个氨基酸、至少500个氨基酸、至少450个氨基酸、至少400个氨基酸、至少350个氨基酸、至少300个氨基酸、至少250个氨基酸、至少200个氨基酸、至少150个氨基酸、至少100个氨基酸或至少50个氨基酸。在一个实施方案中，功能性片段是700至744个氨基酸、650至699个氨基酸、600至649个氨基酸、550至599个氨基酸、500至549个氨基酸、450至499个氨基酸、400至449个氨基酸、350至399个氨基酸、300至349个氨基酸、250至299个氨基酸、200至249个氨基酸、150至199个氨基酸、100至149个氨基酸或50至99个氨基酸。关于上述氨基酸长度的范围，氨基酸长度

的最低值和最高值包括在各公开范围内。在一个实施方案中，功能性片段具有如同本文所述的野生型生物质降解酶或在大肠杆菌中产生的生物质降解酶的生物质降解活性的至少10％、至少20％、至少30％、至少40％、至少50％、至少60％、至少70％、至少75％、至少80％、至少85％、至少90％、至少95％或至少99％。在一个实施方案中，功能性片段具有如同野生型Cel3a或在大肠杆菌中产生的包含SEQ ID NO:1的多肽的纤维二糖酶活性的至少10％、至少20％、至少30％、至少40％、至少50％、至少60％、至少70％、至少75％、至少80％、至少85％、至少90％、至少95％或至少99％。

[0106] 用于检测纤维二糖酶活性的测定在本领域中是已知的。举例来说，可通过使纯化纤维二糖酶与例如纤维二糖D-(+)-纤维二糖的底物一起孵育，以及在反应完成之后检测游离葡萄糖的所得量来确定对从纤维二糖释放的葡萄糖的量的检测。可使用本领域中已知的多种方法测定游离葡萄糖的量。举例来说，在含有50mM柠檬酸钠的缓冲液(pH 5.0,NaOH)中制备纯化纤维二糖酶的稀释液。以使得纤维二糖在反应混合物中的最终浓度是30mM的量将纤维二糖底物添加至纯化纤维二糖酶中。在适于反应发生的条件下，例如在振荡器(700rpm)中，在48℃下持续30分钟孵育反应混合物。为终止反应，在100℃下加热反应混合物5分钟。使反应混合物经0.45μm过滤器过滤，分析滤液以定量葡萄糖和/或纤维二糖酶的量。测量诸如葡萄糖的分析物的YSI仪器可用于测定由反应产生的葡萄糖的浓度。或者，UPLC(超高效液相色谱法)可用于测定来自反应的葡萄糖和纤维二糖的浓度。可将这个测定排列成单反应或多反应形式，例如96孔形式。在一些实施方案中，多反应形式可优选用以产生关于不同浓度的纯化纤维二糖酶来表示纤维二糖酶活性的活性曲线。可使用标准布拉德福德(Bradford)测定来测定纯化纤维二糖酶的浓度。制备纯化纤维二糖酶测定的稀释液，例如2倍稀释液，并且等分至96孔板中，例如12个孔的2倍稀释液。如先前所述添加纤维二糖底物以使纤维二糖酶在反应中的最终浓度是30mM。将板密封并在足以发生纤维二糖酶反应的条件下，接着在足以终止反应的条件下处理。接着使反应经96孔形式0.45μm膜(例如Durapore)过滤，并且通过YSI和/或HPLC方法(例如UPLC)来分析。

[0107] 这个活性测定也可用于通过产生已知浓度的纤维二糖酶的活性的标准曲线以外推未知浓度样品的浓度来确定浓度未知的样品中纤维二糖酶的浓度或效价。举例来说，在96孔板的一行中制备具有已知浓度的纤维二糖酶的两倍连续稀释液，例如12个两倍连续稀释液。其它行含有效价待测定的其它剩余样品(例如粗制溶解产物样品或被增溶包涵体样品)的两倍连续稀释液。使稀释液与D-(+)-纤维二糖(Fluka)底物溶液一起在48℃下在50mM柠檬酸二氢钠缓冲液中在pH 5.0下孵育30分钟。在30分钟之后，加热样品至100℃，持续10分钟以终止反应。使用YSI生物化学分析仪(YSI Life Sciences)和/或HPLC方法分析样品中的葡萄糖和纤维二糖。使用浓度已知的样品，利用在测定的线性范围内的数据点产生标准曲线。可将从效价未知的样品检测的纤维二糖酶活性与标准曲线进行比较以确定这些样品中的纤维二糖酶效价。

[0108] 如果使用在测定的线性范围内的值进行计算，那么活性单位数仅是相对的。测定的线性范围如使用小于30％的原始可溶性底物负载量的葡萄糖值所定义。此外，低于0.05g/L的葡萄糖值由于仪器报告水平而被忽略。一个单位的纤维二糖酶活性定义为葡萄糖的量/Cel3a的量/30分钟:[葡萄糖]g/L/[Cel3a]g/L/30分钟。

[0109] 在其它实施方案中，可使用比色/荧光测定。在用于发生反应的条件下使纯化纤维

二糖酶与底物纤维二糖一起孵育。如下检测产物葡萄糖。向混合物中添加葡萄糖氧化酶，其使葡萄糖(产物)氧化成葡萄糖酸和过氧化氢。接着添加过氧化物酶和邻联茴香胺。邻联茴香胺与过氧化氢在过氧化物酶存在下反应以形成有色产物。添加硫酸，其与氧化邻联茴香胺反应以形成更稳定有色产物。当例如通过分光光度计或色度计例如在540nm下测量时，颜色的强度与葡萄糖浓度成正比。所述比色/荧光葡萄糖测定可例如从Sigma Aldrich(目录号GAGO-20)商购获得。

[0110] 用于检测木质素酶活性的测定在本领域中是已知的。可通过测定藜芦醇氧化成藜芦醛的速率来测量木质素酶活性(将藜芦醇氧化(veratryl alcohol oxidation)缩写为VAO)。制备反应混合物，并且其含有各种稀释度的酶、2mM藜芦醇、0.4mM H_2O_2和20或100mM酒石酸钠，pH 2.9，最终体积0.5ml。反应通过添加H_2O_2来起始，并且通过分光光度测定法在310nm下监测。根据Bradford,M.M.,(1976)Anal.Biochem.72:248-254，使用牛血清白蛋白(Sigma Chemical Co.,St.Louis,MO)作为标准物，或通过使用蛋白质溶液的409nm吸光度以及根据木质素酶的消光系数计算蛋白质量来测定蛋白质。

[0111] 用于检测内切葡聚糖酶活性的测定在本领域中是已知的。举例来说，可通过测量底物羧甲基纤维素(CMC)的水解以及通过BCA方法定量还原性末端的浓度来测定内切葡聚糖酶活性，其中还原性末端的总浓度通过样品溶液的与还原性末端的浓度成比例的颜色变化来展现。首先，在pH 4.8的50mM柠檬酸盐缓冲液中稀释具有生物质降解活性的多肽。将CMC溶液(于柠檬酸钠缓冲液中的0.05％w/v CMC)添加至反应管中并在50℃下平衡。将稀释的酶样品添加至反应中，并且在50℃下孵育10分钟。添加BCA试剂，并且在75℃下孵育30分钟。在减去酶空白和底物空白的读数之后，在560nm下读取吸光度。可基于还原性末端浓度与酶浓度之间的线性范围计算酶活性。其它内切葡聚糖酶活性测定在本领域中是已知的，例如通过测定底物粘度的降低(Zhang等,Biotechnol Adv,2006,24:452-481)。

[0112] 用于检测纤维二糖水解酶活性的测定在本领域中是已知的。可通过在苯酚-硫酸测定中测量从底物阿维塞尔释放的可溶性底物来测定纤维二糖水解酶活性。将阿维塞尔溶液(1.25％w/v于乙酸盐缓冲液中)等分至反应管中，并且制备酶的稀释液。使底物与酶溶液两者在50℃下平衡。将稀释的酶溶液添加至底物中，并且孵育一段足以发生反应的时间，例如在50℃下持续2小时。通过将样品浸没至冰冷水浴中来终止反应。将样品离心以使样品分成可溶性部分和不溶性部分。通过苯酚-硫酸测定来测定可溶性部分中可溶性糖的总浓度。具体来说，使可溶性部分的等分试样与5％苯酚混合，并且添加浓硫酸。使反应冷却至室温(约20-30分钟)，并且在490nm下读取样品的吸光度。基于总可溶性糖释放量与酶稀释度之间的线性关系计算酶活性。其它纤维二糖水解酶活性测定描述于Zhang等，Biofuels:Methods and Protocols，第581卷.第213-231页中。

[0113] 用于检测木聚糖酶活性的测定在本领域中是已知的。可通过比色测定对从木聚糖底物释放的木糖的水平进行测量来测定木聚糖酶活性。将木聚糖底物制备成于50mM乙酸钠缓冲液(pH 4.5)中的1.0％w/v溶液。制备酶的稀释液。混合木聚糖和酶稀释液，并且在足以发生反应的条件下孵育，例如30℃持续10分钟。接着将含有16mM硫酸铜、1.3M硫酸钠、226mM碳酸钠、190mM碳酸氢钠和43mM酒石酸钾钠的溶液添加至反应中。接着将反应煮沸10分钟，并且使其冷却至室温。添加含有40mM钼酸、19mM砷酸和756mM硫酸的溶液。将反应振荡或涡旋直至起泡停止，并且存在的任何沉淀得以溶解。将反应离心以进行澄清，接着通过分光光

度计在540nM下读取溶液，并且基于总可溶性糖释放量与酶稀释度之间的线性关系计算酶活性。

[0114] 无糖基化多肽

[0115] 任何本文所述的具有生物质降解活性例如纤维二糖酶活性的多肽都可为糖基化的或无糖基化的。如本文所述，可从包涵体增溶具有生物质降解活性的无糖基化多肽。或者，可将具有生物质降解活性的无糖基化多肽添加至包含已从包涵体增溶的具有生物质降解活性的多肽的混合物中，其中从包涵体增溶的所述多肽可为糖基化的或无糖基化的。

[0116] 糖基化是碳水化合物连接于糖基接受体所采用的酶促过程，所述糖基接受体例如精氨酸或天冬酰胺侧链的氮或丝氨酸、苏氨酸或酪氨酸侧链的羟基氧。有两种类型的糖基化：N连接的糖基化和O连接的糖基化。N连接的糖基化发生在共有位点Asn-X-Ser/Thr处，其中X可为除脯氨酸之外的任何氨基酸。O连接的糖基化发生在Ser/Thr残基处。可使用本领域中已知的各种算法预测糖基化位点，所述算法诸如可通过瑞士生物信息学研究所(Swiss Institute of Bioinformatics)公开获得的Prosite和可通过生物序列分析中心(Center for Biological Sequence Analysis)公开获得的NetNGlyc 1.0或NetOGlyc 4.0。

[0117] 本发明提供用于产生具有生物质降解活性的无糖基化多肽的方法。在一个实施方案中，已使编码多肽的核酸改变或突变以致多肽不能被糖基化，例如使一个或多个糖基化位点突变以致聚糖不能连接于糖基化位点。举例来说，由本文所述的核酸序列编码的具有生物质降解活性的无糖基化多肽在一个或多个糖基化位点处含有一个或多个突变，所述糖基化位点已被突变以致聚糖不再能附接或连接于糖基化位点。在另一实例中，由本文所述的核酸序列编码的具有生物质降解活性的多肽含有一个或多个邻近于一个或多个糖基化位点的突变，所述糖基化位点已被突变以致聚糖不再能附接或连接于糖基化位点。举例来说，邻近于糖基化位点的突变使由糖基化酶识别的共有基序突变，或改变多肽的构象以使多肽不能被糖基化，例如糖基化位点被隐藏或归因于新构象的空间位阻阻止糖基化酶接近糖基化位点。邻近于具有生物质降解活性的多肽中的糖基化位点的突变直接邻近于糖基化位点或离糖基化位点有至少2、至少3、至少4、至少5、至少6、至少7、至少8、至少9、至少10、至少15、至少20、至少30或至少40个氨基酸，所述糖基化位点由于邻近突变而将不被糖基化。

[0118] 在一个实施方案中，使纤维二糖酶例如Cel3a或SEQ ID NO:1的一个或多个以下糖基化位点突变：氨基酸位置78处的苏氨酸、氨基酸位置241处的苏氨酸、氨基酸位置343处的丝氨酸、氨基酸位置450处的丝氨酸、氨基酸位置599处的苏氨酸、氨基酸位置616处的丝氨酸、氨基酸位置691处的苏氨酸、氨基酸位置21处的丝氨酸、氨基酸位置24处的苏氨酸、氨基酸位置25处的丝氨酸、氨基酸位置28处的丝氨酸、氨基酸位置38处的苏氨酸、氨基酸位置42处的苏氨酸、氨基酸位置303处的苏氨酸、氨基酸位置398处的丝氨酸、氨基酸位置435处的丝氨酸、氨基酸位置436处的丝氨酸、氨基酸位置439处的苏氨酸、氨基酸位置442处的苏氨酸、氨基酸位置446处的苏氨酸、氨基酸位置451处的丝氨酸、氨基酸位置619处的丝氨酸、氨基酸位置622处的丝氨酸、氨基酸位置623处的苏氨酸、氨基酸位置626处的丝氨酸或氨基酸位置630处的苏氨酸或其任何组合。在各实施方案中，使糖基化位点从丝氨酸或苏氨酸突变成丙氨酸。举例来说，本文所述的无糖基化多肽具有一个或多个以下突变：T78A、T241A、S343A、S450A、T599A、S616A、T691A、S21A、T24A、S25A、S28A、T38A、T42A、T303A、T398A、S435A、S436A、T439A、T442A、T446A、S451A、S619A、S622A、T623A、S626A或T630A或其任何组

合。或者，使一个或多个邻近于上述糖基化位点的氨基酸突变。

[0119] 用以检测多肽是否由聚糖修饰(例如多肽是否被糖基化或无糖基化)的测定在本领域中是已知的。多肽可被纯化或分离，并且可被染色以检测和定量聚糖部分，或可通过质谱测定法来分析多肽，并且与相应参照多肽进行比较。参照多肽与测试多肽(其糖基化状态待测定)具有相同一级序列，但被糖基化或无糖基化。

[0120] 相较于相应糖基化多肽例如糖基化Ce13a多肽，本文所述的无糖基化多肽可具有增加的生物质降解活性，例如纤维二糖酶活性。举例来说，相较于糖基化多肽，具有生物质降解活性例如纤维二糖酶活性的无糖基化多肽具有至少1％、2％、5％、10％、15％、20％、25％、30％、40％、50％、60％、70％、80％、90％、100％或200％生物质降解活性，例如纤维二糖酶活性。

[0121] 核酸、表达载体和宿主细胞

[0122] 在宿主细胞中表达如本文所述的具有生物质降解活性的多肽。在一个实施方案中，将包含编码任何本文所述的具有生物质降解活性的多肽的核酸序列的表达载体引入宿主细胞中，并且在适于表达具有生物质降解活性的多肽的条件下培养所述宿主细胞。在各实施方案中，具有生物质降解活性的多肽的表达处于使得形成包含具有生物质降解活性的多肽的包涵体或聚集物的水平下。本文也描述用于表达和分离在宿主细胞中表达的具有生物质降解活性的可溶性多肽的方法。用于从包涵体增溶和分离具有生物质降解活性的多肽的方法进一步描述于题为"从包涵体进行增溶"的章节中。

[0123] 本发明也提供一种编码具有生物质降解活性的多肽的核酸序列。在一个实施方案中，核酸序列编码本文所述的木质素酶、内切葡聚糖酶、纤维二糖水解酶、木聚糖酶或纤维二糖酶。可将编码具有生物质降解活性的多肽的核酸序列进行密码子优化以达成在宿主细胞中的表达增加。密码子优化包括改变核酸序列以考虑包括密码子使用偏好、隐蔽剪接位点、mRNA二级结构、过早多聚腺苷酸位点、密码子和抗密码子的相互作用、以及RNA不稳定性基序在内因素来增加编码多肽在宿主中的表达。用于密码子优化的各种算法和商业服务在本领域中是已知的和可用的。

[0124] 在一个实施方案中，核酸序列编码来自里氏木霉的具有本文所述的氨基酸序列例如SEQ ID NO:1的Ce13a酶。在一个实施方案中，以下提供编码Ce13a的核酸序列：

[0125] ATGCGTTACCGAACAGCAGCTGCGCTGGCACTTGCCACTGGGCCCTTTGCTAGGGCAGACAGTCACTCA
ACATCGGGGGCCTCGGCTGAGGCAGTTGTACCTCCTGCAGGGACTCCATGGGGAACCGCGTACGACAAGGCGAAGGC
CGCATTGGCAAAGCTCAATCTCCAAGATAAGGTCGGCATCGTGAGCGGTGTCGGCTGGAACGGCGGTCCTTGCGTTG
GAAACACATCTCCGGCCTCCAAGATCAGCTATCCATCGCTATGCCTTCAAGACGGACCCCTCGGTGTTCGATACTCG
ACAGGCAGCACAGCCTTTACGCCGGGCGTTCAAGCGGCCTCGACGTGGGATGTCAATTTGATCCGCGAACGTGGACA
GTTCATCGGTGAGGAGGTGAAGGCCTCGGGGATTCATGTCATACTTGGTCCTGTGGCTGGGCCGCTGGGAAAGACTC
CGCAGGGCGGTCGCAACTGGGAGGGCTTCGGTGTCGATCCATATCTCACGGGCATTGCCATGGGTCAAACCATCAAC
GGCATCCAGTCGGTAGGCGTGCAGGCGACAGCGAAGCACTATATCCTCAACGAGCAGGAGCTCAATCGAGAAACCAT
TTCGAGCAACCCAGATGACCGAACTCTCCATGAGCTGTATACTTGGCCATTTGCCGACGCGGTTCAGGCCAATGTCG
CTTCTGTCATGTGCTCGTACAACAAGGTCAATACCACCTGGGCCTGCGAGGATCAGTACACGCTGCAGACTGTGCTG
AAAGACCAGCTGGGGTTCCCAGGCTATGTCATGACGGACTGGAACGCACAGCACACGACTGTCCAAAGCGCGAATTC
TGGGCTTGACAIGTCAATGCCTGGCACAGACTTCAACGGTAACAATCGGCTCIGGGGTCCAGCTCTCACCAATGCGG

TAAATAGCAATCAGGTCCCCACGAGCAGAGTCGACGATATGGTGACTCGTATCCTCGCCGCATGGTACTTGACAGGC
CAGGACCAGGCAGGCTATCCGTCGTTCAACATCAGCAGAAATGTTCAAGGAAACCACAAGACCAATGTCAGGGCAAT
TGCCAGGGACGGCATCGTTCTGCTCAAGAATGACGCCAACATCCTGCCGCTCAAGAAGCCCGCTAGCATTGCCGTCG
TTGGATCTGCCGCAATCATTGGTAACCACGCCAGAAACTCGCCCTCGTGCAACGACAAAGGCTGCGACGACGGGGCC
TTGGGCATGGGTTGGGGTTCCGGCGCCGTCAACTATCCGTACTTCGTCGCGCCCTACGATGCCATCAATACCAGAGC
GTCTTCGCAGGGCACCCAGGTTACCTTGAGCAACACCGACAACACGTCCTCAGGCGCATCTGCAGCAAGAGGAAAGG
ACGTCGCCATCGTCTTCATCACCGCCGACTCGGGTGAAGGCTACATCACCGTGGAGGGCAACGCGGGCGATCGCAAC
AACCTGGATCCGTGGCACAACGGCAATGCCCTGGTCCAGGCGGTGGCCGGTGCCAACAGCAACGTCATTGTTGTTGT
CCACTCCGTTGGCGCCATCATTCTGGAGCAGATTCTTGCTCTTCCGCAGGTCAAGGCCGTTGTCTGGGCGGGTCTTC
CTTCTCAGGAGAGCGGCAATGCGCTCGTCGACGTGCTGTGGGGAGATGTCAGCCCTTCTGGCAAGCTGGTGTACACC
ATTGCGAAGAGCCCCAATGACTATAACACTCGCATCGTTTCCGGCGGCAGTGACAGCTTCAGCGAGGGACTGTTCAT
CGACTATAAGCACTTCGACGACGCCAATATCACGCCGCGGTACGAGTTCGGCTATGGACTGTCTTACACCAAGTTCA
ACTACTCACGCCTCTCCGTCTTGTCGACCGCCAAGTCTGGTCCTGCGACTGGGGCCGTTGTGCCGGGAGGCCCGAGT
GATCTGTTCCAGAATGTCGCGACAGTCACCGTTGACATCGCAAACTCTGGCCAAGTGACTGGTGCCGAGGTAGCCCA
GCTGTACATCACCTACCCATCTTCAGCACCCAGGACCCCTCCGAAGCAGCTGCGAGGCTTTGCCAAGCTGAACCTCA
CGCCTGGTCAGAGCGGAACAGCAACGTTCAACATCCGACGACGAGATCTCAGCTACTGGGACACGGCTTCGCAGAAA
TGGGTGGTGCCGTCGGGGTCGTTTGGCATCAGCGTGGGAGCGAGCAGCCGGGATATCAGGCTGACGAGCACTCTGTC
GGTAGCG

[0126] (SEQ ID NO:2)

[0127] 以下提供编码Ce13a的密码子优化的核酸序列二

[0128] ATGCGTTATCGTACAGCCGCAGCCCTGGCACTGGCCACAGGTCCGTTCGCACGTGCCGATAGTCACAGT
ACCAGCGGTGCCAGCGCAGAAGCCGTGGTTCCGCCGGCAGGCACACCGTGGGGCACAGCCTATGATAAAGCCAAAGC
CGCCCTGGCCAAGCTGAATCTGCAGGATAAAGTGGGCATCGTGAGTGGCGTGGGCTGGAACGGTGGTCCGTGCGTTG
GCAACACCAGCCCGGCAAGCAAGATCAGCTATCCGAGCTTATGCCTGCAGGATGGTCCGCTGGGCGTGCGCTATAGC
ACCGGTAGTACCGCCTTTACACCTGGTGTGCAGGCCGCCAGTACCTGGGACGTTAACCTGATCCGCGAACGTGGCCA
ATTTATCGGCGAAGAAGTTAAAGCCAGCGGCATTCATGTTATTCTGGGTCCGGTGGCCGGTCCTCTGGGTAAAACCC
CGCAGGGCGGCCGTAATTGGGAAGGCTTCGGCGTTGATCCGTATTTAACCGGCATCGCAATGGGCCAGACCATTAAT
GGCATCCAGAGCGTGGGTGTTCAAGCCACCGCCAAACACTACATATTAAACGAACAGGAACTGAATCGTGAAACCAT
CAGCAGCAATCCGGATGATCGCACCCTGCATGAGCTGTATACATGGCCTTTTGCCGACGCAGTTCAGGCCAACGTGG
CAAGTGTGATGTGTAGCTATAACAAGGTGAACACCACCTGGGCCTGCGAAGACCAGTACACCCTGCAGACCGTTTTA
AAAGACCAACTGGGCTTCCCTGGTTACGTGATGACAGATTGGAATGCCCAGCACACAACCGTTCAGAGCGCAAACAG
TGGCCTGGATATGAGCATGCCGGGCACCGACTTCAACGGCAATAATCGTCTGTGGGGTCCGGCACTGACCAATGCCG
TTAACAGCAACCAGGTGCCGACCAGTCGTGTGGACGATATGGTTACCCGTATTCTGGCCGCCTGGTACCTGACAGGT
CAAGACCAGGCCGGCTACCCGAGCTTCAACATCAGCCGCAACGTGCAGGGTAATCACAAGACCAACGTTCGCGCAAT
CGCACGCGATGGTATCGTGCTGTTAAAGAACGATGCCAACATTCTGCCGCTGAAAAAACCGGCCAGCATCGCCGTTG
TTGGTAGCGCAGCCATCATTGGCAACCACGCCCGTAACAGTCCGAGCTGCAATGATAAAGGCTGTGACGACGGTGCC
CTGGGCATGGGTTGGGGTAGTGGTGCCGTGAACTACCCGTATTTCGTGGCCCCGTACGACGCCATTAACACCCGTGC
AAGTAGCCAGGGTACCCAGGTTACCCTGAGCAACACCGACAACACAAGCAGCGGTGCCAGTGCAGCACGTGGTAAGG
ATGTGGCCATCGTGTTCATCACCGCCGACAGCGGCGAAGGCTACATTACCGTGGAGGGTAATGCCGGTGATCGCAAT

AATCTGGACCCGTGGCATAACGGCAACGCCCTGGTTCAGGCAGTGGCAGGCGCAAATAGCAACGTGATCGTTGTGGT
GCATAGCGTGGGTGCCATCATTCTGGAGCAGATCCTGGCCCTGCCGCAAGTTAAGGCAGTTGTGTGGGCAGGTCTGC
CGAGCCAAGAAAGTGGCAATGCCCTGGTGGACGTTCTGTGGGGCGATGTTAGTCCGAGCGGCAAGCTGGTGTATACA
ATCGCCAAGAGCCCGAACGACTATAACACCCGCATCGTTAGCGGCGGCAGTGATAGCTTCAGCGAGGGCCTGTTTAT
CGACTACAAGCATTTCGATGATGCCAATATTACCCCGCGCTACGAATTTGGTTATGGCCTGAGCTATACCAAGTTCA
ACTACAGCCGCCTGAGCGTTTTAAGTACCGCCAAGAGTGGTCCGGCAACAGGTGCCGTGGTTCCTGGTGGTCCGAGT
GATCTGTTTCAGAATGTGGCCACCGTGACCGTGGATATCGCCAACAGTGGTCAGGTTACCGGCGCCGAAGTGGCACA
GCTGTACATCACCTATCCGAGCAGTGCACCGCGCACCCCGCCGAAACAGCTGCGTGGCTTCGCCAAATTAAACCTGA
CCCCGGGCCAGAGCGGTACAGCAACCTTCAATATTCGCCGCCGTGATCTGAGCTATTGGGACACCGCCAGCCAAAAA
TGGGTGGTGCCGAGCGGCAGCTTTGGCATTAGTGTGGGTGCAAGTAGCCGCGACATTCGCTTAACAAGCACCCTGAG
TGTTGCC

[0129] (SEQ ID NO:3)

[0130] 在一个实施方案中，编码Cel3a酶或其功能性变体的核酸序列包含与SEQ ID NO:2的至少50％、至少55％、至少60％、至少65％、至少70％、至少75％、至少80％、至少85％、至少90％、至少91％、至少92％、至少93％、至少94％、至少95％、至少96％、至少97％、至少98％或至少99％同一性。在一个实施方案中，编码Cel3a酶或其功能性变体的核酸序列包含与SEQ ID NO:2或SEQ ID NO:3的至少50％、至少55％、至少60％、至少65％、至少70％、至少75％、至少80％、至少85％、至少90％、至少91％、至少92％、至少93％、至少94％、至少95％、至少96％、至少97％、至少98％或至少99％同一性。

[0131] 本文也提供一种编码本文所述的具有生物质降解活性的无糖基化多肽例如Cel3a多肽的核酸序列，其中一个或多个存在于多肽中的糖基化位点已被突变以致聚糖不再能附接或连接于糖基化位点。在另一实施方案中，本文所述的核酸序列编码如上所述的无糖基化多肽，例如Cel3a多肽，其中一个或多个突变邻近于一个或多个存在于多肽中的已被突变的糖基化位点以致聚糖不再能附接或连接于糖基化位点，如先前所述。

[0132] 用于分离或克隆编码多肽的核酸序列的技术在本领域中是已知的，并且包括从基因组DNA分离、从cDNA制备或其组合。可例如通过使用熟知的聚合酶链式反应(PCR)或抗体表达文库筛选以检测具有共有结构特征的克隆DNA片段来实现从所述基因组DNA克隆本发明的核酸序列。参见例如Innis等，1990，PCR:A Guide to Methods and Application，Academic Press，New York。可使用其它扩增程序，诸如连接酶链式反应(LCR)、连接活化转录(LAT)和基于核苷酸序列的扩增(NASBA)。核酸序列可从里氏木霉菌株(例如野生型里氏木霉或里氏木霉RUTC30)或另一或相关生物体克隆，并且因此，例如可为核酸序列的多肽编码区域的等位基因或物种变体。

[0133] 可通过在遗传工程改造中用于使核酸序列从它的天然位置重新定位于它将被再现所处的不同位点的标准克隆程序来获得核酸序列。克隆程序可涉及切除和分离包含编码多肽的核苷酸序列的所需片段，将所述片段插入载体分子中，以及将重组载体并入宿主细胞中，在所述宿主细胞中，多个拷贝或克隆的核苷酸序列将被复制。核苷酸序列可为基因组、cDNA、RNA、半合成、合成来源或其任何组合。

[0134] 如本文所用，"表达载体"是用于向宿主细胞中引入和表达目标核酸序列的核酸构建体。在一些实施方案中，载体包含可操作地连接于本文所述的核酸序列，并且能够实现由

本文所述的核酸序列编码的多肽的表达的适合控制序列。控制序列可为由用于表达核酸序列的宿主细胞识别的适当启动子序列。在一个实施方案中，目标核酸序列是编码如本文所述的具有生物质降解活性例如纤维二糖酶活性的多肽的核酸序列。

[0135] 本文所述的表达载体中的启动子可包括从编码细胞外或细胞内多肽的基因获得的对于宿主细胞是同源性或异源性的启动子、突变启动子、截短启动子和杂合启动子。

[0136] 适于指导本发明的核酸构建体在细菌宿主细胞中转录的启动子的实例是从大肠杆菌lac操纵元、大肠杆菌tac启动子(杂合启动子，DeBoer等，PNAS，1983，80:21-25)、大肠杆菌rec A、大肠杆菌araBAD、大肠杆菌tetA和原核β-内酰胺酶获得的启动子。适合启动子的其它实例包括病毒启动子，诸如来自噬菌体的启动子，包括T7启动子、T5启动子、T3启动子、M13启动子和SP6启动子。在一些实施方案中，超过一个启动子控制目标核酸序列的表达，例如大肠杆菌lac启动子和T7启动子。可适用于本发明中的其它启动子描述于"Useful proteins from recombinant bacteria"，Scientific American，1980，242:74-94；以及Sambrook等，Molecular Cloning:A Laboratory Manual，1989中。在一些优选实施方案中，启动子是诱导型的，其中添加分子会刺激下游阅读框的转录和表达。

[0137] 适于指导本发明的核酸构建体在真核宿主细胞中，例如在真菌或酵母细胞中转录的启动子的实例是从里氏木霉、甲醇诱导性醇氧化酶(AOX启动子)、构巢曲霉(Aspergillus nidulans)色氨酸生物合成(trpC启动子)、黑曲霉泡盛变种(Aspergillus niger Var.awamori)葡萄糖淀粉酶(glaA)、酿酒酵母(Saccharomyces cerevisiae)半乳糖激酶(GAL1)的基因获得的启动子或乳酸克鲁维酵母(Kluyveromyces lactis)Plac4-PBI启动子。

[0138] 存在于本文所述的表达载体中的控制序列也可为编码连接于多肽的氨基末端的氨基酸序列，并且将所编码多肽引导至细胞的分泌路径中的信号序列，例如分泌信号序列。信号序列可为内源性信号序列，例如其中当由目标多肽所源于的生物体内源性表达时，信号序列存在于野生型多肽的N末端。信号序列可为外来或异源性信号肽，其中信号序列来自不同生物体或来自不同于所表达的目标多肽的多肽。将所表达多肽引导至宿主细胞的分泌路径中的任何信号序列都可用于本发明中。通常，信号序列由6个与136个之间的碱性和/或疏水性氨基酸组成。

[0139] 适于本发明的信号序列的实例包括来自酿酒酵母α因子的信号序列。

[0140] 融合标签也可用于本文所述的表达载体以有助于检测和纯化所表达多肽。适合融合标签的实例包括His标签(例如3xHis、6x His(SEQ ID NO:22)或8xHis(SEQ ID NO:21))、GST标签、HSV标签、S标签、T7标签。其它适合融合标签包括myc标签、血凝素(HA)标签和荧光蛋白标签(例如绿色荧光蛋白)。通常使融合标签可操作地连接于待表达的多肽的N末端或C末端。在一些实施方案中，在融合标签序列与待表达的多肽的N末端或C末端之间可存在接头区域。在一个实施方案中，接头区域包含具有1个至20个的氨基酸，不影响或改变所表达多肽的表达或功能的序列。

[0141] 利用本文所述的融合标签允许例如通过使用特异性识别标签的抗体进行蛋白质印迹来检测所表达蛋白质。标签也允许例如通过亲和色谱法来从宿主细胞纯化所表达多肽。举例来说，可通过使用镍亲和色谱法来纯化融合于His标签的所表达多肽。His标签对镍离子具有亲和力，并且镍柱将截留his标记的多肽，同时允许所有其它蛋白质和细胞碎片流

动通过柱。使用例如含有咪唑的洗脱缓冲液洗脱His标记的多肽使His标记的多肽从柱释放，从而产生大致上纯化的多肽。

[0142] 本文所述的表达载体可进一步包含如为本领域技术人员通常所知的用以使得能够分离用构建体转化的经遗传修饰的微生物的可选择标记基因。可选择标记基因可赋予对抗生素的抗性或在缺乏为宿主生物体所特定的营养物的培养基上生长的能力，所述宿主生物体否则不可在这些条件下生长。本发明不受对可选择标记基因的选择的限制，并且本领域技术人员可易于确定适当基因。举例来说，可选择标记基因可赋予对氨苄青霉素、氯霉素、四环素、卡那霉素、潮霉素、腐草霉素、遗传霉素或G418的抗性，或可弥补宿主微生物在trD、arg、leu、pyr4、pyr、ura3、ura5、his或ade基因中的一个基因方面的缺陷，或可赋予在乙酰胺作为唯一氮源的情况下生长的能力。

[0143] 本文所述的表达载体可进一步包含如为本领域技术人员通常所知的其它核酸序列，例如额外控制序列，例如转录终止子、用以将各种其它核酸序列连接在一起的合成序列、复制起点、核糖体结合位点、多克隆位点(或多接头位点)、多聚腺苷酸化信号等。适于表达载体的核糖体结合位点取决于所用的宿主细胞，例如对于在原核宿主细胞中表达，可使用原核RBS，例如T7噬菌体RBS。多克隆位点或多接头位点含有一个或多个优选不存在于表达载体的其余序列中的限制酶位点。限制酶位点用于插入编码具有纤维二糖酶活性的多肽的核酸序列或其它所需控制序列。本发明的实施不受这些其它核酸序列例如其它控制序列中的任何一个或多个的存在的限制。

[0144] 适用于本发明中的表达载体的实例包括用于在原核生物中表达的载体，例如细菌表达载体。适用于本发明中的一种细菌表达载体是pET载体(Novagen)，其含有以下：病毒T7启动子，其仅对T7RNA聚合酶(而非细菌RNA聚合酶)具有特异性，并且也不存在于原核基因组中任何地方；lac操纵子，其包含lac启动子和lac阻遏蛋白的编码序列(lacI基因)；多接头；f1复制起点(以使当与M13辅助噬菌体共感染时，可产生单链质粒)；氨苄青霉素抗性基因和ColE1复制起点(Blaber，1998)。启动子与lac操纵子两者均位于其中插入编码本文所述的多肽的核酸序列的多接头的5′或上游。lac操纵子赋予编码具有纤维二糖酶活性的多肽的核酸序列的可诱导表达。添加乳糖代谢物IPTG(异丙基β-D-1-硫代半乳糖吡喃糖苷)会触发lac操纵元的转录，并且诱导核酸序列在lac操纵子的控制下的蛋白质表达。使用这个系统要求向宿主细胞中添加T7RNA聚合酶以达成载体表达。T7RNA聚合酶可通过第二表达载体来引入，或可使用被遗传工程改造以表达T7RNA聚合酶的宿主细胞株。

[0145] 供与本发明一起使用的示例性表达载体是可从Novagen商购获得的pET载体。pET表达系统描述于美国专利号4,952,496；5,693,489；和5,869,320中。在一个实施方案中，pET载体是可从Novagen商购获得的pET-DUET载体，例如pET-Duet1。适用于本发明中的其它载体包括含有His标签序列的载体，诸如美国专利号5,310,663和5,284,933；以及欧洲专利号282042中所述的那些。

[0146] 本发明也涉及一种包含本发明的核酸序列或表达载体的宿主细胞，其用于重组产生具有生物质降解活性的多肽。

[0147] 将包含本发明的核酸序列的表达载体引入宿主细胞中以使所述载体得以维持(例如通过染色体整合或作为自我复制性染色体外载体)，以便表达多肽。

[0148] 宿主细胞可为原核生物或真核生物。宿主细胞可为细菌，诸如大肠杆菌菌株，例如

K12菌株NovaBlue、NovaBlue T1R、JM109和DH5α。优选地，细菌细胞能够折叠或部分折叠外源性表达的蛋白质，诸如大肠杆菌Origami菌株，例如Origami B、Origami B(DE3)、Origami 2和Origami 2(DE3)菌株。在一些实施方案中，可优选的是使用具有糖基化缺陷，或具有受损糖基化途径的宿主细胞以使由所述宿主细胞表达的蛋白质不被显著糖基化。

[0149] 宿主细胞可为酵母或丝状真菌，特别是分类为子囊菌门(Ascomycota)的那些。适用作用于表达本发明的经修饰的TrCel3Aβ-葡萄糖苷酶的宿主微生物的酵母的属包括酵母属(Saccharomyces)、毕赤酵母属(Pichia)、汉逊酵母属(Hansenula)、克鲁维酵母属(Kluyveromyces)、耶罗威亚酵母属(Yarrowia)和Arxula。适用作用于表达本发明的多肽的微生物的真菌的属包括木霉属、肉座菌属(Hypocrea)、曲霉属、镰刀菌属、腐质霉属、链孢霉属(Neurospora)、金孢子菌属、毁丝霉属、梭孢壳属、侧孢霉属(Sporotrichum)和青霉属。举例来说，宿主细胞可为巴斯德毕赤酵母。举例来说，宿主细胞可为里氏木霉的工业菌株或其突变体，例如里氏木霉RUTC30。通常，宿主细胞是不表达亲本生物质降解酶例如纤维二糖酶或Cel3a的细胞。

[0150] 对例如细菌细胞或真菌细胞的特定宿主细胞的选择取决于用于产生本发明的无糖基化多肽的表达载体(例如控制序列)和/或方法，如以下进一步详细所述。

[0151] 可通过由微生物转化领域技术人员所知的许多方法来将本发明的表达载体引入宿主细胞中，所述方法包括但不限于转化、用$CaCl_2$处理细胞、电穿孔、基因枪轰击、脂质体转染和PEG介导的原生质体融合(例如White等，WO 2005/093072，其以引用的方式并入本文)。在选择含有表达载体的重组宿主细胞(例如通过利用表达载体的可选择标记进行选择)之后，可在诱导本发明的具有生物质降解活性的多肽的表达的条件下培养重组宿主细胞。

[0152] 用于回收由原核生物和真核生物细胞表达的具有生物质降解活性的可溶性多肽的方法在本领域中是已知的。在各实施方案中，用于回收多肽的方法包括例如通过离心或过滤来收集细胞，以及例如通过机械、化学或酶促手段来溶解细胞。举例来说，可例如通过声波处理、研磨(与珠粒一起振荡)或剪切力来将细胞物理破坏分开。可处理细胞膜以使它们可被渗透以便释放细胞的内含物，诸如用例如曲通(Triton)、NP-40或SDS的清洁剂处理。可使用诸如溶菌酶(lysozyme)或溶解酶(lysonase)的酶使具有细胞壁的细胞例如细菌细胞可被渗透。在本发明的情形下，上述机械、化学和酶促技术的任何组合也适于从宿主细胞回收所表达目标多肽。举例来说，当在例如大肠杆菌细胞的细菌细胞中表达本文所述的具有生物质降解活性的多肽时，通常通过离心和沉淀细胞培养物来收集细胞，并且通过将细胞沉淀再混悬于含有溶菌酶的溶解缓冲液中来使细胞溶解。为确保完全溶解，使再混悬细胞经受以下方法中的一种：声波处理、研磨或均质化。在离心之后，具有生物质降解活性的可溶性多肽存在于上清液中，而例如在包涵体中的具有生物质降解活性的不溶性多肽见于沉淀中。用于例如从包涵体回收具有生物质降解活性的不溶性多肽的方法以下进一步描述于题为"从包涵体进行增溶"的章节中。

[0153] 可接着使用本领域中已知的标准方法从细胞溶解产物纯化或分离具有生物质降解活性的可溶性多肽。对于包含例如His标签的标签的具有生物质降解活性的多肽，亲和色谱法可用于使可溶性多肽与溶解产物的可溶性部分的其余部分分离。

[0154] 在一个实施方案中，在添加至生物质中以进行糖化反应之前，不使表达本文所述

的具有生物质降解活性的多肽的宿主细胞溶解。在一些情况下，用于溶解宿主细胞以及提取具有生物质降解活性的多肽的方法可导致蛋白质变性和/或酶活性降低，此导致下游加工成本增加。因此，本发明也提供用于在糖化步骤之前将表达本文所述的具有生物质降解活性的无糖基化多肽的宿主细胞直接添加至生物质中的方法。

[0155] 在一个实施方案中，例如通过离心来分离表达本文所述的具有生物质降解活性的多肽的宿主细胞例如大肠杆菌细胞，并且添加至糖化反应，例如含有生物质的糖化反应器中。通过来自生物质的剪切力、叶轮和增加温度的组合来使细胞溶解。在一个实施方案中，将表达本文所述的具有生物质降解活性的多肽的宿主细胞例如大肠杆菌细胞的培养物从发酵罐直接添加至糖化罐中，并且消除对通过离心来沉淀细胞的需要。在一个实施方案中，多肽是糖基化的或无糖基化的。

[0156] 从包涵体进行增溶

[0157] 在各实施方案中，已使用本文所述的方法遗传修饰细胞例如本文公开的微生物以产生至少一种具有生物质降解活性的多肽。至少一部分，例如至少5％、10％、20％、30％、40％、50％、60％、70％、80％或90％的具有生物质降解活性的多肽见于经遗传修饰的细胞中的包涵体中。本文公开用于从包涵体增溶至少一种具有生物质降解活性的多肽的方法。

[0158] 包涵体是当在高水平下表达时于宿主细胞中的不溶性聚集物，其包含异源性表达的蛋白质，例如具有生物质降解活性的多肽。包涵体可见于核或细胞质中。包涵体也可含有其它组分，诸如其它宿主细胞内源性蛋白质例如宿主蛋白质、核糖体组分、核酸(例如RNA和/或DNA)和细胞碎片(例如细胞壁碎片、脂质、代谢物)。宿主细胞内源性蛋白质包括由宿主细胞的基因组DNA编码的任何蛋白质。宿主细胞内源性蛋白质可定位于细胞质，或可与异源性表达的蛋白质相互作用。可见于包涵体中的核糖体组分的实例包括核糖体、核糖体亚基(例如50S亚基或30S亚基)的片段、部分翻译的多肽、转运RNA、延伸因子和/或信使RNA。可见于包涵体中的核酸的实例包括宿主的基因组DNA、外源性DNA(例如来自引入宿主细胞中的质粒或表达载体)、信使RNA、转运RNA、核糖体RNA或其任何片段。可见于包涵体中的细胞碎片的实例包括细胞壁或细胞膜碎片(例如细胞壁或细胞膜的片段或组分)、核膜碎片(例如核膜的片段或组分)、其它宿主细胞器的片段或组分、内毒素、脂质和/或代谢物。

[0159] 用于降低包涵体的聚集的额外方法包括声波处理、在不同温度下孵育、酸/碱处理、蛋白酶处理、电处理、机械处理和添加产生蛋白酶的生物体。举例来说，在-20℃至0℃、0℃至4℃、4℃至20℃、20℃至40℃以及40℃至80℃的范围内的温度下孵育含有包涵体的细胞或其溶解产物。

[0160] 为分离包涵体，首先使用本领域中的标准方法溶解表达具有生物质降解活性的多肽的宿主细胞，所述方法诸如通过溶菌酶或其它变性剂来溶解、超声处理、声波处理或高压均质化。在不导致包涵体增溶的条件下使宿主细胞溶解。使用本领域中已知的技术从宿主细胞分离包涵体。举例来说，分离细胞溶解产物以使含有具有生物质降解活性的多肽和其它不溶性物质的包涵体存在于不溶性部分中，而可溶性部分含有具有生物质降解活性的可溶性多肽。所述分离可通过离心来实现，借此包涵体见于沉淀例如不溶性部分中，而可溶性多肽见于上清液例如可溶性部分中。适于使不溶性部分与可溶性部分分离的其它方法包括过滤。

[0161] 使包涵体增溶以释放具有生物质降解活性的多肽包括将增溶剂添加至不溶性部

分或包涵体中。在一些实施方案中，增溶剂可为阻止蛋白质聚集或沉淀，或溶解蛋白质聚集物的试剂。在一些实施方案中，增溶剂包括破坏范德华相互作用、疏水性相互作用、氢键合、偶极-偶极相互作用、离子相互作用、π堆积或其任何组合的试剂。

[0162] 在一些实施方案中，增溶剂可为破坏疏水性相互作用的试剂，例如像清洁剂。示例性清洁剂包括非离子型、两性离子型、阴离子型和阳离子型清洁剂。在一些实施方案中，增溶剂可为非离子型的，例如NP-40和曲通X-100。在一些实施方案中，增溶剂可为两性离子型的，例如CHAPS和磺基甜菜碱，例如SB3-10或ASB 14。在一些实施方案中，蛋白质试剂可为阴离子型的，例如十二烷基硫酸钠(SDS)。

[0163] 在一些实施方案中，增溶剂可为使二硫键还原的试剂，例如硫醇还原剂。示例性硫醇还原剂包括2-巯基乙醇βME和二硫苏糖醇(DTT)。在一些实施方案中，使二硫键还原的增溶剂可为膦，例如三丁基膦(TBP)或三羧乙基膦(TCEP)。

[0164] 在一些实施方案中，增溶剂可为破坏氢键合和疏水性相互作用的试剂。在一些实施方案中，增溶剂可为离液化合物，例如尿素和取代的尿素(例如硫脲)以及盐酸胍。

[0165] 在一些实施方案中，增溶剂可为是非极性溶剂的试剂。非极性溶剂在具有类似电负性的原子诸如碳和氢之间含有键，并且具有极低介电常数。举例来说，非极性溶剂具有小于5的介电常数。非极性溶剂的实例包括戊烷、己烷、环己烷、苯、甲苯、氯仿、乙醚。

[0166] 在一些实施方案中，增溶剂可为是极性溶剂的试剂。极性溶剂的特征在于具有大偶极矩(或"部分电荷")；它们在具有极其不同电负性的原子诸如氧和氢之间含有键。在一个实施方案中，本文中适用于本发明中的极性溶剂具有至少5或至少20的介电常数。在一个优选实施方案中，极性溶剂具有高介电常数，例如大于25的介电常数。在一些实施方案中，增溶剂是质子性极性溶剂，其具有O-H或N-H键，具有例如大于20、大于25的高介电常数，并且是良好氢键供体，例如甲酸、正丁醇、异丙醇、正丙醇、乙醇、甲醇或硝基甲烷。在一些实施方案中，增溶剂可为非质子性极性溶剂，其缺乏O-H或N-H键，并且具有在5与20之间的介电常数，例如二甲亚砜(DMSO)、二氯甲烷(DCM)、四氢呋喃(THF)、乙酸乙酯、丙酮、二甲基甲酰胺(DMF)或乙腈(MeCN)。

[0167] 在一些实施方案中，增溶剂可为具有正电荷的试剂，其可适于破坏带净负电荷分子的离子相互作用。示例性带正电荷增溶剂包括N-甲基D-还原葡糖胺、胆碱、精氨酸、赖氨酸、普鲁卡因(procaine)、缓血酸胺(TRIS)、精胺(spermine)、N-甲基-吗啉、葡糖胺、N,N-双2-羟乙基甘氨酸、二氮杂双环十一烯、肌酸(creatine)、精氨酸乙酯、金刚烷胺(amantadine)、金刚烷乙胺(rimantadine)、鸟氨酸、牛磺酸和瓜氨酸。阳离子部分可另外包括钠、钾、钙、镁、铵、单乙醇胺、二乙醇胺、三乙醇胺、缓血酸胺、赖氨酸、组氨酸、精氨酸、吗啉、甲基还原葡糖胺和葡糖胺。

[0168] 在一些实施方案中，增溶剂可为具有负电荷的试剂，其可适于破坏带净正电荷分子的离子相互作用。示例性带负电荷的增溶剂包括乙酸盐、丙酸盐、丁酸盐、戊酸盐、己酸盐、庚酸盐、乙酰丙酸盐、氯化物、溴化物、碘化物、柠檬酸盐、丁二酸盐、顺丁烯二酸盐、乙醇酸盐葡萄糖酸盐、葡糖醛酸盐、3-羟基异丁酸盐、2-羟基异丁酸盐、乳酸盐、苹果酸盐、丙酮酸盐、反丁烯二酸盐、酒石酸盐、丙醇二酸盐、硝酸盐、磷酸盐、苯磺酸盐、甲烷磺酸盐、硫酸盐、磺酸盐、乙酸、金刚烷甲酸、α酮戊二酸、D-天冬氨酸或L-天冬氨酸、苯磺酸、苯甲酸、10-樟脑磺酸、柠檬酸、1,2-乙烷二磺酸、反丁烯二酸、D-葡萄糖酸、D-葡萄糖醛酸、葡糖二酸、D-

谷氨酸或L-谷氨酸、戊二酸、乙醇酸、马尿酸(hippuric acid)、氢溴酸、盐酸、1-羟基-2-萘甲酸、乳糖酸、顺丁烯二酸、L-苹果酸、扁桃酸、甲烷磺酸、粘液酸、1,5-萘二磺酸四水合物、2-萘磺酸、硝酸、油酸、帕莫酸、磷酸、对甲苯磺酸水合物、D-糖二酸单钾盐、水杨酸、硬脂酸、丁二酸、硫酸、丹宁酸和D-酒石酸或L-酒石酸。

[0169] 将增溶剂以足以从包涵体增溶具有生物质降解活性的多肽的浓度，例如以约0.01-10M、约0.05-10M、约0.1-10M、约0.2-10M、约0.5-10M、约1-10M、约2-10M、约5-10M、约8-10M、约0.01-6M、约0.05-6M、约0.1-6M、约0.2-6M、约0.5-6M、约1-6M、约2-6M、约4-6M或约5-6M的浓度添加至包涵体或含有包涵体的部分中。在一个实施方案中，将增溶剂以约0.01M、约0.02M、约0.05M、约0.1M、约0.2M、约0.5M、约1M、约2M、约3M、约4M、约5M、约6M、约7M、约8M、约9M或约10M的浓度添加至包涵体或含有包涵体的部分中。

[0170] 在一个优选实施方案中，增溶剂是尿素，并且以约0.01-10M、约0.05-10M、约0.1-10M、约0.2-10M、约0.5-10M、约1-10M、约2-10M、约5-10M、约8-10M、约0.01-6M、约0.05-6M、约0.1-6M、约0.2-6M、约0.5-6M、约1-6M、约2-6M、约4-6M或约5-6M的浓度添加至包涵体或含有包涵体的部分中。在一个实施方案中，将尿素以约0.01M、约0.02M、约0.05M、约0.1M、约0.2M、约0.5M、约1M、约2M、约3M、约4M、约5M、约6M、约7M、约8M、约9M或约10M的浓度添加至包涵体或含有包涵体的部分中。在一个优选实施方案中，将尿素以6M的浓度添加至包涵体或含有包涵体的部分中。

[0171] 在使用增溶剂进行增溶之后，所得混合物含有如本文所述的具有生物质降解活性的被增溶多肽。所得混合物可直接用于酶促过程，诸如用于产生产物的反应例如糖化反应中，如在题为"使用被增溶酶产生产物的方法"的章节中进一步详细所述。在这个实施方案中，混合物可含有包涵体的其它组分，诸如其它宿主细胞其它内源性蛋白质、核糖体组分、核酸(例如RNA和/或DNA)和细胞碎片(例如细胞壁碎片、脂质、代谢物)。

[0172] 在其它实施方案中，进一步处理含有具有生物质降解活性的被增溶多肽的所得混合物以从包涵体的其它被增溶组分纯化或分离具有生物质降解活性的所述被增溶多肽。适于分离或纯化具有生物质降解活性的被增溶多肽的方法包括亲和纯化技术。举例来说，具有生物质降解活性的多肽优选含有可用于亲和纯化的标签或融合肽。在一个实施方案中，具有生物质降解活性的多肽含有His标签，例如8X His标签(SEQ ID NO:21)，并且可使用镍亲和色谱法例如固定化金属离子亲和色谱法(IMAC)系统纯化具有生物质降解活性的被增溶多肽。在一个实施方案中，所有纯化步骤都在用于从包涵体增溶多肽的增溶剂存在下发生，包括在纯化过程的洗涤和洗脱步骤中。因此，在一个实施方案中，所得具有生物质降解活性的纯化被增溶多肽也含有增溶剂。

[0173] 本发明提供一种包含具有生物质降解活性的多肽和增溶剂的混合物。混合物是通过如上所述使包涵体增溶来获得。所得混合物也可进一步包含一种或多种与包涵体相关的蛋白质。具有生物质降解活性的被增溶多肽也可通过亲和纯化技术纯化。在这种情况下，所得混合物不包含一种或多种与包涵体相关的蛋白质。混合物可包含见于包涵体中的其它组分，诸如其它宿主细胞内源性蛋白质、核糖体组分、核酸(例如RNA和/或DNA)和细胞碎片(例如细胞壁碎片、脂质、代谢物)。

[0174] 在一个实施方案中，具有生物质降解活性的被增溶多肽可为部分未折叠的，部分错误折叠的，或部分变性的。在一个实施方案中，相较于具有生物质降解活性的天然多肽，

具有生物质降解活性的被增溶多肽具有至少1%、至少2%、至少3%、至少4%、至少5%、至少6%、至少7%、至少8%、至少9%、至少10%、至少8-10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%或至少100%生物质降解活性。在一个实施方案中，相较于具有生物质降解活性的天然多肽，具有生物质降解活性的被增溶多肽具有约1-10%、1-20%、1-30%、1-40%、1-50%、1-60%、1-70%、1-80%、1-90%、1-100%、10-20%、10-30%、10-40%、10-50%、10-60%、10-70%、10-80%、10-90%、10-100%、20-30%、20-40%、20-50%、20-60%、20-70%、20-80%、20-90%、20-100%、30-40%、30-50%、30-60%、30-/0%、30-80%、30-90%、30-100%、40-50%、40-60%、40-70%、40-80%、40-90%、40-100%、50-60%、50-70%、50-80%、50-90%、50-100%、60-70%、60-80%、60-90%、60-100%、70-80%、70-90%、70-100%、80-90%、80-100%或90-100%的生物质降解活性。在一个优选实施方案中，具有生物质降解活性的被增溶多肽具有至少8-10%的天然多肽的活性。具有生物质降解活性的天然多肽是指例如从可溶性部分分离的具有生物质降解活性的相应多肽、以它的天然形式适当折叠(由此具有100%生物质降解活性)的具有生物质降解活性的相应多肽、或由多肽所源于的微生物内源性表达的具有生物质降解活性的相应多肽。生物质降解活性可通过任何本文所述的测定来测定，例如木质素酶活性测定、内切葡聚糖酶活性测定、纤维二糖水解酶活性测定、纤维二糖酶活性测定或木聚糖酶活性测定。

[0175] 在一个方面，混合物包含具有纤维二糖酶活性的多肽，例如来自里氏木霉的Ce13a或其功能性变体，例如与SEQ ID NO:1具有至少90%同一性的多肽；以及例如尿素的增溶剂。在一个实施方案中，混合物包含与SEQ ID NO:1具有至少90%同一性的多肽和例如尿素的增溶剂，其中相较于天然多肽例如SEQ ID NO:1或来自里氏木霉的Ce13a，所述多肽具有至少20%的纤维二糖酶活性。混合物是通过如上所述使包涵体增溶来获得。所得混合物也可进一步包含一种或多种与包涵体相关的蛋白质。具有纤维二糖酶活性的被增溶多肽，例如与SEQ ID NO:1具有至少90%同一性的多肽也可通过亲和纯化技术纯化。在这种情况下，所得混合物不包含一种或多种与包涵体相关的蛋白质。混合物可包含见于包涵体中的其它组分，诸如其它宿主细胞内源性蛋白质、核糖体组分、核酸(例如RNA和/或DNA)和细胞碎片(例如细胞壁碎片、脂质、代谢物)。在一个实施方案中，增溶剂是尿素，并且尿素以0.2-6M存在。

[0176] 在一个实施方案中，具有纤维二糖酶活性或与SEQ ID NO:1具有至少90%同一性的被增溶多肽可为部分未折叠的，部分错误折叠的，或部分变性的。在一个实施方案中，相较于天然多肽，具有纤维二糖酶活性或与SEQ ID NO:1具有至少90%同一性的被增溶多肽具有至少1%、至少2%、至少3%、至少4%、至少5%、至少6%、至少7%、至少8%、至少9%、至少10%、至少8-10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%或至少100%纤维二糖酶活性。在一个实施方案中，相较于天然多肽，具有纤维二糖酶活性或与SEQ ID NO:1具有至少90%同一性的被增溶多肽具有约1-10%、1-20%、1-30%、1-40%、1-50%、1-60%、1-70%、1-80%、1-90%、1-100%、10-20%、10-30%、10-40%、10-50%、10-60%、10-70%、10-80%、10-90%、10-100%、20-30%、20-

40％、20-50％、20-60％、20-70％、20-80％、20-90％、20-100％、30-40％、30-50％、30-60％、30-70％、30-80％、30-90％、30-100％、40-50％、40-60％、40-70％、40-80％、40-90％、40-100％、50-60％、50-70％、50-80％、50-90％、50-100％、60-70％、60-80％、60-90％、60-100％、70-80％、70-90％、70-100％、80-90％、80-100％或90-100％的纤维二糖酶活性。天然多肽是例如适当折叠(例如100％折叠)的SEQ ID NO:1、从里氏木霉分离的Cel3a或其功能性变体。纤维二糖酶活性可使用本文所述的测定来测量，并且可定量为在30分钟之后释放的葡萄糖的浓度(g/L)或在30分钟内转化成葡萄糖的纤维二糖的％。

[0177] 用于产生无糖基化多肽的方法

[0178] 本发明进一步提供用于在宿主细胞中产生具有生物质降解活性的无糖基化多肽的方法，其中如本文所述用增溶剂以适于增溶所述无糖基化多肽的浓度处理所述宿主细胞或其溶解产物。方法包括在适于表达多肽的条件下培养表达具有生物质降解活性的多肽的宿主细胞。方法也可包括从宿主细胞回收具有生物质降解活性的无糖基化多肽。在以下进一步详细描述的方法中，具有生物质降解活性的多肽具有例如木质素酶活性、内切葡聚糖酶活性、纤维二糖水解酶活性、纤维二糖酶活性或木聚糖酶活性。在一个实施方案中，具有纤维二糖酶活性的多肽包括来自里氏木霉的Cel3a或其功能性片段。在另一实施方案中，具有纤维二糖酶活性的多肽包括SEQ ID NO:1。

[0179] 使用在糖基化方面缺陷的宿主细胞

[0180] 在各实施方案中，表达载体包含编码可操作地连接于融合标签的本文所述的具有生物质降解活性的多肽的核酸序列，所述表达载体被引入不使在细胞中表达的蛋白质显著糖基化的所述细胞(例如细菌宿主细胞)中，并且在所述细胞中表达。在用于表达多肽的条件下培养重组宿主细胞，从而导致产生具有生物质降解活性的无糖基化多肽。可使用针对如本文所述的融合标签的亲和色谱方法从宿主细胞纯化或分离无糖基化多肽。

[0181] 举例来说，在这个实施方案中，表达载体在编码具有生物质降解活性的多肽的核酸序列的上游含有lac操纵子和T7启动子，并且宿主细胞能够表达T7RNA聚合酶。通过添加IPTG来诱导具有生物质降解活性的多肽的表达。优选地，宿主细胞是大肠杆菌细胞，优选是大肠杆菌Origami细胞。在这个实施方案中，融合标签是His标签，并且对所表达无糖基化多肽的纯化包括镍亲和色谱法。

[0182] 使用具有糖基化能力的宿主细胞

[0183] 在另一实施方案中，在宿主细胞中表达包含编码如本文所述的多肽的核酸序列的表达载体，所述多肽包含一个或多个糖基化位点突变以使所述多肽不被糖基化，其中所述宿主细胞能够使在所述细胞内表达的蛋白质糖基化，例如酵母或真菌宿主细胞。或者，宿主细胞不能够使在所述细胞例如细菌宿主细胞内表达的蛋白质糖基化。在这个实施方案中，使多肽可操作地连接于融合标签。可使用针对如本文所述的融合标签的亲和色谱方法从细菌宿主细胞纯化或分离无糖基化多肽。

[0184] 在另一实施方案中，在宿主细胞中表达包含编码本文所述的具有生物质降解活性的多肽的核酸序列的表达载体，其中所述宿主细胞能够使在所述细胞内表达的蛋白质糖基化。在足以达成多肽的表达和糖基化的条件下培养细胞。在这个实施方案中，使多肽可操作地连接于融合标签。可使用针对如本文所述的融合标签的亲和色谱方法从细菌宿主细胞纯化或分离糖基化多肽。在从宿主细胞和其它内源性宿主酶(例如糖基化酶)纯化之后，可通

过与去糖基化酶一起孵育来移除所分离的糖基化多肽的聚糖。去糖基化酶包括PNGase F、PNGase A、EndoH(内切糖苷酶H)、EndoS(内切糖苷酶S)、EndoD(内切糖苷酶D)、EndoF(内切糖苷酶F)、EndoF1(内切糖苷酶F1)或EndoF2(内切糖苷酶F2)。含有足以完全移除聚糖的酶的蛋白质去糖基化混合物可例如从New England Biolabs商购获得。使所分离多肽与一种或多种去糖基化酶在足以从多肽移除所有聚糖的条件下一起孵育。本领域中已知用于从多肽移除聚糖的其它方法,例如用弱碱进行β消除或温和肼解。可使用本领域中已知的用于染色和观察聚糖的方法或质谱测定法来确定对多肽的糖基化状态的评估。

[0185] 在另一实施方案中,在宿主细胞中表达包含编码本文所述的具有生物质降解活性的多肽的核酸序列的表达载体,其中所述宿主细胞能够使在所述细胞内表达的蛋白质糖基化。在足以表达多肽的条件下,但在糖基化抑制剂存在下培养细胞。糖基化抑制剂以使得所表达多肽呈无糖基化的浓度并持续足够时间存在。在这个实施方案中,使多肽可操作地连接于融合标签。可使用针对如本文所述的融合标签的亲和色谱方法从细菌宿主细胞纯化或分离所得无糖基化多肽。

[0186] 适用于这个实施方案中的糖基化抑制剂的实例包括衣霉素、苯甲基-GalNAc(苯甲基2-乙酰胺基-2-脱氧-α-D-吡喃半乳糖苷)、2-氟-2-脱氧-D-葡萄糖和5'CDP(5'二磷酸胞苷)。在一些实施方案中,使用糖基化抑制剂的组合。优选地,用于这个实施方案中的糖基化抑制剂的浓度足以抑制多肽的糖基化,但不导致细胞毒性或对宿主细胞的蛋白质表达的抑制。

[0187] 使生物质转化成产物的方法

[0188] 本发明提供用于使用如本文所述的具有纤维二糖酶活性的无糖基化多肽将生物质转化或加工成产物的方法和组合物。用于使生物质转化成诸如糖产物的产物的方法在本领域中是已知的,例如如美国专利申请2014/0011258中所述,所述专利的内容以引用的方式整体并入本文。简要来说,将生物质进行最优预处理例如以降低不顺应性,并且通过涉及使所处理生物质与生物质降解酶或纤维素分解酶一起孵育以产生糖(例如葡萄糖和/或木糖)的糖化过程来糖化。可接着例如通过发酵或蒸馏来进一步加工糖产物以产生最终产物。最终产物包括醇(例如乙醇、异丁醇或正丁醇)、糖醇(例如赤藻糖醇、木糖醇或山梨糖醇)或有机酸(例如乳酸、丙酮酸、丁二酸)。

[0189] 使用本文所述的方法,可使生物质材料转化成一种或多种产物,诸如能量、燃料、食物和材料。产物的具体实例包括但不限于氢气、糖(例如葡萄糖、木糖、阿拉伯糖、甘露糖、半乳糖、果糖、纤维二糖、二糖、寡糖和多糖)、醇(例如一元醇或二元醇,诸如乙醇、正丙醇、异丁醇、仲丁醇、叔丁醇或正丁醇)、水合醇或含水醇(例如含有大于10%、20%、30%或甚至大于40%水)、生物柴油、有机酸、烃(例如甲烷、乙烷、丙烷、异丁烯、戊烷、正己烷、生物柴油、生物汽油及其混合物)、副产物(例如蛋白质,诸如纤维素分解蛋白质(酶)或单细胞蛋白)、以及呈任何组合或相对浓度,并且任选地与任何添加剂(例如燃料添加剂)组合的任何这些产物的混合物。其它实例包括羧酸、羧酸的盐、羧酸和羧酸的盐和羧酸的酯(例如甲酯、乙酯和正丙酯)的混合物、酮(例如丙酮)、醛(例如乙醛)、α和β不饱和酸(例如丙烯酸)和烯烃(例如乙烯)。其它醇和醇衍生物包括丙醇、丙二醇、1,4-丁二醇、1,3-丙二醇、糖醇和多元醇(例如二醇、甘油、赤藻糖醇、苏糖醇、阿糖醇、木糖醇、核糖醇、甘露糖醇、山梨糖醇、半乳糖醇、艾杜糖醇、肌醇、庚七醇、异麦芽酮糖醇、麦芽糖醇、乳糖醇、麦芽三糖醇、麦芽四糖醇

和聚糖醇(polyglycitol)以及其它多元醇)、以及任何这些醇的甲酯或乙酯。其它产物包括丙烯酸甲酯、甲基丙烯酸甲酯、乳酸、柠檬酸、甲酸、乙酸、丙酸、丁酸、丁二酸、戊酸、己酸、3-羟基丙酸、棕榈酸、硬脂酸、草酸、丙二酸、戊二酸、油酸、亚油酸、乙醇酸、γ-羟基丁酸及其混合物、任何这些酸的盐、任何酸和它们的相应盐的混合物。

[0190] 生物质

[0191] 待使用本文所述的方法加工的生物质是包含纤维素的淀粉材料和/或纤维素材料,例如木质纤维素材料。生物质也可包含半纤维素和/或木质素。生物质可包括以下中的一种或多种:农业产品或废弃物、纸产品或废弃物、林业产品或一般性废弃物或其任何组合。农业产品或废弃物包括可被栽培、收获或加工以供例如由人或动物使用或消耗的材料;或由栽培、收获或加工方法产生的任何中间体、副产品或废弃物。农业产品或废弃物包括但不限于甘蔗、黄麻、大麻、亚麻、竹子、剑麻、苜蓿、干草、秘鲁胡萝卜、荞麦、香蕉、大麦、木薯、野葛、圆齿酢酱草、西米、高粱、马铃薯、甜薯、芋头、山药、菜豆、蚕豆、扁豆、豌豆、草、柳枝稷、芒草、绳草、草芦、谷物残渣、菜籽秆、小麦秆、大麦秆、燕麦秆、稻草、玉米穗轴、玉米秸杆、玉米纤维、椰子毛、甜菜浆、甘蔗渣、大豆秸杆、谷物残渣、稻壳、燕麦壳、小麦谷壳、大麦壳或细麸或其组合。纸产品或废弃物包括用于制造纸产品的材料;任何纸产品;或由制造或分解纸产品产生的任何中间体、副产品或废弃物。纸产品或废弃物包括但不限于纸、着色纸、装料纸、涂覆纸、瓦楞纸、填充纸、杂志、印刷品、打印纸、涂塑纸、卡片坯料、卡纸板、纸板或纸浆或其组合。林业产品或废弃物包括通过栽培、收获或加工木材产生的材料;或由栽培、收获或加工木材产生的任何中间体、副产品或废弃物。林业产品或废弃物包括但不限于白杨木、来自任何树属或树种的木材、创花板、木片或锯屑或其组合。一般性废弃物包括但不限于粪肥、污水或垃圾或其组合。

[0192] 生物质可包括但不限于淀粉材料、甘蔗、农业废弃物、纸、纸产品、纸废弃物、纸浆、着色纸、装料纸、涂覆纸、填充纸、杂志、印刷品、打印纸、涂塑纸、卡片坯料、卡纸板、纸板、棉花、木材、创花板、林业废弃物、锯屑、白杨木、木片、草、柳枝稷、芒草、绳草、草芦、谷物残渣、稻壳、燕麦壳、小麦谷壳、大麦壳、农业废弃物、青贮饲料、菜籽秆、小麦秆、大麦秆、燕麦秆、稻草、黄麻、大麻、亚麻、竹子、剑麻、蕉麻、玉米穗轴、玉米秸杆、大豆秸杆、玉米纤维、苜蓿、干草、椰子毛、糖加工残渣、甘蔗渣、甜菜浆、龙舌兰渣、海藻、海草、浮游生物粪肥、污水、垃圾、农业或工业废弃物、秘鲁胡萝卜、荞麦、香蕉、大麦、木薯、野葛、圆齿酢酱草、西米、高粱、马铃薯、甜薯、芋头、山药、菜豆、蚕豆、扁豆、豌豆或任何这些的混合物。在一个优选实施方案中,生物质包括农业废弃物,诸如玉米穗轴,例如玉米秸杆。在另一实施方案中,生物质包括草。

[0193] 在一个实施方案中,在与本文所述的组合物接触之前处理生物质。举例来说,处理生物质以降低生物质的不顺应性,降低它的堆密度,和/或增加它的表面积。适合生物质处理方法可包括但不限于:用电子进行轰击、声波处理、氧化、热解、蒸汽爆炸、化学处理、机械处理和冷冻研磨。优选地,处理方法是用电子进行轰击。

[0194] 在一些实施方案中,进行电子轰击直至生物质接受至少0.5Mrad,例如至少5、10、20、30或至少40Mrad的总剂量。在一些实施方案中,进行处理直至生物质接受约0.5Mrad至约150Mrad、约1Mrad至约100Mrad、约5Mrad至约75Mrad、约2Mrad至约75Mrad、约10Mrad至约50Mrad,例如约5Mrad至约50Mrad、约20Mrad至约40Mrad、约10Mrad至约35Mrad、或约20Mrad

至约30Mrad的剂量。在一些执行中，25至35Mrad的总剂量是优选的，在理想情况下历经几秒进行施加，例如在每次穿过5Mrad下，其中各次穿过被施加约1秒。在一些情况下，施加每次穿过大于7至9Mrad的剂量可导致原料材料的热降解。

[0195] 生物质材料(例如植物生物质、动物生物质、纸和城市废弃物生物质)可用作原料以产生有用中间体和产物，诸如有机酸、有机酸的盐、酸酐、有机酸的酯和燃料，例如用于内燃机的燃料或用于燃料电池的原料。本文描述可使用可易于获得，但常常可难以加工的纤维素和/或木质纤维素材料作为原料的系统和方法，所述材料例如城市废弃物物流和废纸物流，诸如包括报纸、牛皮纸(kraft paper)、瓦楞纸或这些的混合物的物流。

[0196] 为了将原料转化成可被容易加工的形式，可通过糖化剂(例如酶或酸)将原料中的含有葡聚糖或木聚糖的纤维素水解成低分子量碳水化合物诸如糖，所过过程被称为糖化。然后，低分子量碳水化合物可用于例如现有制造厂中，如单细胞蛋白质厂、酶制造厂或燃料厂，例如乙醇制造设施。

[0197] 可使用酶，例如通过使材料和所述酶在溶剂中，例如在水溶液中组合来水解原料。可根据本文所述的方法制备/诱导酶。

[0198] 具体来说，酶可由能够分解生物质(诸如生物质的纤维素和/或木质素部分)，或含有或制造各种纤维素分解酶(纤维素酶)、木质素酶或各种小分子生物质降解代谢物的生物体供给。这些酶可为协同作用降解生物质的结晶纤维素或木质素部分的酶复合物。纤维素分解酶的实例包括：内切葡聚糖酶、纤维二糖水解酶和纤维二糖酶(β-葡萄糖苷酶)。

[0199] 在糖化期间，纤维素底物可通过内切葡聚糖酶在随机位置初步水解，从而产生低聚中间体。这些中间体随后被作为外切葡聚糖酶诸如纤维二糖水解酶的底物，以从纤维素聚合物的末端产生纤维二糖。纤维二糖是水溶性的1,4-连接的葡萄糖二聚体。最后，纤维二糖酶裂解纤维二糖以得到葡萄糖。此过程的效率(例如水解时间和/或水解完全性)取决于纤维素材料的不顺应性。

[0200] 糖化

[0201] 通常通过使不顺应性降低的生物质和生物质降解酶在例如水溶液的流体介质中组合来用以上讨论的生物质降解酶处理不顺应性降低的生物质。在一些情况下，在糖化之前在热水中煮沸、浸渍或蒸煮原料，如2012年4月26日公开的由Medoff和Masterman完成的美国专利申请公布2012/0100577 A1中所述，所述美国专利申请公布的整个内容被并入本文。

[0202] 本文提供用于本文所述的糖化过程中的能够降解生物质的酶的混合物，例如生物质降解酶的酶混合物。

[0203] 糖化过程可在制造厂中的储罐(例如具有至少4000L、40,000L、500,000L、2,000,000L、4,000,000L或6,000,000L或更大体积的储罐)中部分或完全地进行，和/或可在转运中，例如在轨道车、油罐卡车中或在高级油轮或船舱中部分或完全地进行。完全糖化所需要的时间将取决于工艺条件和所使用的生物质材料和酶。如果糖化是在受控的条件下在制造厂中进行，那么纤维素可在约12-96小时内大致上完全转化成糖，例如葡萄糖。如果糖化是在转运中部分或完全地进行，那么糖化可能花费较长时间。

[0204] 在一个优选实施方案中，在对于发生酶促反应最优的pH下，例如在对于生物质降解酶的活性最优的pH下发生糖化反应。优选地，糖化反应的pH在pH 4-4.5下。在一个优选实

施方案中，在对于发生酶促反应最优的温度下，例如在对于生物质降解酶的活性最优的温度下发生糖化反应。优选地，糖化反应的温度在42℃-52℃下。

[0205] 通常优选在糖化期间例如使用喷射混合对储罐内含物进行混合，如在2010年5月18日提交的国际申请号PCT/US2010/035331中所描述，所述申请以英语公布为WO 2010/135380并且指定美国，所述申请的全部公开内容以引用的方式并入本文。

[0206] 表面活性剂的添加可提高糖化速率。表面活性剂的实例包括非离子型表面活性剂(诸如Tween®20或Tween®80聚乙二醇表面活性剂)、离子型表面活性剂或两性表面活性剂。

[0207] 通常优选的是由糖化得到的糖溶液的浓度相对较高，例如大于5重量%、7.5重量%、10重量%、10.5重量%，或大于40重量%，或大于50重量%、60重量%、70重量%，或甚至大于80重量%。可例如通过蒸发去除水以增加糖溶液的浓度。这减小了待装运的体积并且还抑制了溶液中的微生物生长。

[0208] 或者，可使用较低浓度的糖溶液，在这种情况下，可能希望以低浓度(例如50至150ppm)添加抗微生物添加剂，例如广谱抗生素。其它适合抗生素包括两性霉素B、氨苄青霉素、氯霉素、环丙沙星、庆大霉素、潮霉素B、卡那霉素、新霉素、青霉素、嘌呤霉素、链霉素。抗生素将在运输和储存期间抑制微生物的生长，并且可以适当浓度(例如按重量计在15与10,000ppm之间，例如在25与500ppm之间，或在50与150ppm之间)使用。如果希望，则即使糖浓度相对较高也可包括抗生素。或者，可使用具有抗微生物防腐性质的其它添加剂。优选地，抗微生物添加剂是食品级的。

[0209] 可通过限制与酶一起添加到生物质材料中的水量来获得相对较高浓度的溶液。可例如通过控制糖化发生到何种程度来控制浓度。举例来说，可通过向溶液中添加更多生物质材料来增加浓度。为保持所产生的糖处于溶解状态，可添加表面活性剂，例如以上讨论的那些中的一种。也可通过增加溶液的温度来增加溶解性。举例来说，可将溶液维持在40℃-50℃、60℃-80℃或甚至更高的温度下。

[0210] 在本文所述的方法中，例如在糖化之后，可分离糖(例如葡萄糖和木糖)。举例来说，可通过沉淀、结晶、色谱法(例如模拟移动床色谱法、高压色谱法)、离心、提取、本领域中已知的任何其它分离方法及其组合来分离糖。

[0211] 用于糖化中的混合物

[0212] 在一方面，本发明的特征在于一种用于糖化中的混合物，其包含如本文所述的已从包涵体增溶的具有生物质降解活性的多肽、一种或多种与包涵体相关的蛋白质、和增溶剂。在一个实施方案中，混合物可含有包涵体的其它组分，诸如其它宿主细胞内源性蛋白质、核糖体组分、核酸(例如RNA和/或DNA)和细胞碎片(例如细胞壁碎片、脂质、代谢物)。具有生物质降解活性的多肽可为糖基化的或无糖基化的。

[0213] 在一个实施方案中，混合物包含具有纤维二糖酶活性的多肽和尿素。在一个实施方案中，尿素以0.2-6M存在。在一个实施方案中，混合物包含来自里氏木霉的Ce13a或其功能性变体或片段。在一个实施方案中，混合物包含以下多肽：所述多肽包含与来自里氏木霉的Ce13a的至少50%、55%、60%、65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%同一性。在另一实施方案中，混合物包含以下多肽：所述多肽包含与SEQ ID NO:1的至少50%、55%、60%、65%、70%、75%、80%、85%、90%、91%、

92%、93%、94%、95%、96%、97%、98%或99%同一性。具有纤维二糖酶活性的多肽可为糖基化的或无糖基化的。

[0214] 在各实施方案中，本文所述的混合物进一步包含至少一种源于微生物的额外酶，其中所述额外酶具有生物质或纤维素基材料降解活性。举例来说，额外酶是木质素酶、内切葡聚糖酶、纤维二糖水解酶、木聚糖酶或纤维二糖酶。在一个实施方案中，混合物进一步包含一种或多种木质素酶、一种或多种内切葡聚糖酶、一种或多种纤维二糖水解酶、一种或多种木聚糖酶或一种或多种纤维二糖酶。在各实施方案中，额外生物质降解酶是糖基化的。在各实施方案中，酶混合物进一步包含至少2种、至少3种、至少4种、至少5种、至少6种、至少7种、至少8种、至少9种、至少10种或至少20种或更多种本文所述的额外生物质降解酶。以下显示若干生物质降解酶的典型一级氨基酸序列。

[0215] 举例来说，混合物进一步包含由微生物产生的额外生物质降解酶的混合物，所述微生物例如真菌细胞，诸如野生型里氏木霉或其突变体，例如里氏木霉RUTC30。在一个实施方案中，从微生物分离额外生物质降解酶。在一个实施方案中，混合物包含一种或多种以下生物质降解酶：B2AF03、CIP1、CIP2、Cel1a、Cel3a、Cel5a、Cel6a、Cel7a、Cel7b、Cel12a、Cel45a、Cel74a、paMan5a、paMan26a或膨胀蛋白或其任何组合。例如以上所列的额外生物质降解酶可从酶所源于的微生物(例如真菌细胞)内源性表达和分离(以下列于表1中)。或者，可使用类似表达方法在本文所述的宿主细胞中异源性表达例如以上所列的额外生物质降解酶，并且从所述宿主细胞分离。在一个实施方案中，异源性表达的额外生物质降解酶在酶的C末端或N末端用His标签进行标记，并且使用本领域中已知的镍亲和色谱法技术进行分离。举例来说，额外生物质降解酶选自以下表1。

[0216] 表1.额外生物质降解酶的实例

蛋白质	MW，kDa	氨基酸数目	理论等电点	半胱氨酸数目	生物体
B2AF03	87.1	800	5.94	10	鹅柄孢壳菌(*Podospora anserina*)
CIP1	32.9	316	4.93	8	里氏木霉
CIP2	48.2	460	7.0	12	里氏木霉
Cel1a	52.2	466	5.3	5	里氏木霉
Cel3a	78.4	744	6.3	6	里氏木霉

[0217]

[0218]

Cel5a	44.1	418	4.9	12	里氏木霉
Cel6a	49.6	471	5.1	12	里氏木霉
Cel7a	54.1	514	4.6	24	里氏木霉
Cel7b	48.2	459	4.7	22	里氏木霉
Cel12a	25.1	234	6.6	2	里氏木霉
Cel45a	24.4	242	4.2	16	里氏木霉
Cel74a	87.1	838	5.4	4	里氏木霉
paMan5a	41.1	373	7.0	6	鹅柄孢壳菌
paMan26a	51.7	469	4.7	1	鹅柄孢壳菌
膨胀蛋白	51.5	493	4.8	28	里氏木霉

[0219] 以下提供表1中所列的生物质降解酶的氨基酸序列。

[0220] B2AF03(鹅柄孢壳菌)

[0221]

MKSSVFWGASLTSAVVRAIDLPFQFYPNCVDDLLSTNQVCNTTLSPPERAAALVAALTPEEKLQNIVSKSLGAPRIG
LPAYNWWSEALHGVAYAPGTQFWQGDGPFNSSTSFPMPLLMAATFDDELLEKIAEVIGIEGRAFGNAGFSGLDYWTP
NVNPFKDPRWGRGSETPGEDVLLVKRYAAAMIKGLEGPVPEKERRVVATCKHYAANDFEDWNGATRHNFNAKISLQD
MAEYYFMPFQQCVRDSRVGSIMCAYNAVNGVPSCASPYLLQTILREHWNWTEHNNYITSDCEAVLDVSLNHKYAATN
AEGTAISFEAGMDTSCEYEGSSDIPGAWSQGLLKESTVDRALLRLYEGIVRAGYFDGKQSLYSSLGWADVNKPSAQK
LSLQAAVDGTVLLKNDGTLPLSDLLDKSRPKKVAMIGFWSDAKDKLRGGYSGTAAYLHTPAYAASQLGIPFSTASGP
ILHSDLASNQSWTDNAMAAAKDADYILYFGGIDTSAAGETKDRYDLDWPGAQLSLINLLTTLSKPLIVLQMGDQLDN
TPLLSNPKINAILWANWPGQDGGTAVMELVTGLKSPAGRLPVTQYPSNFTELVPMTDMALRPSAGNSQLGRTYRWYK
TPVQAFGFGLHYTTFSPKFGKKFPAVIDVDEVLEGCDDKYLDTCPLPDLPVVVENRGNRTSDYVALAFVSAPGVGPG
PWPIKTLGAFTRLRGVKGGEKREGGLKWNLGNLARHDEEGNTVVYPGKYEVSLDEPPKARLRFEIVRGGKGKGKVKG
KGKAAQKGGVVLDRWPKPPKGQEPPAIERV(SEQ ID NO:6)

[0222] CIP1(里氏木霉)

[0223]

MVRRTALLALGALSTLSMAQISDDFESGWDQTKWPISAPDCNQGGTVSLDTTVAHSGSNSMKVVGGPNGYCGHIFFG
TTQVPTGDVYVRAWIRLQTALGSNHVTFIIMPDTAQGGKHLRIGGQSQVLDYNRESDDATLPDLSPNGIASTVTLPT
GAFQCFEYHLGTDGTIETWLNGSLIPGMTVGPGVDNPNDAGWTRASYIPEITGVNFGWEAYSGDVNTVWFDDISIAS
TRVGCGPGSPGGPGSSTTGRSSTSGPTSTSRPSTTIPPPTSRTTTATGPTQTHYGQCGGIGYSGPTVCASGTTCQVL
NPYYSQCL(SEQ ID NO:7)

[0224] CIP2(里氏木霉)

[0225]

MASRFFALLLLAIPIQAQSPVWGQCGGIGWSGPTTCVGGATCVSYNPYYSQCIPSTQASSSIASTTLVTSFTTTTAT
RTSASTPPASSTGAGGATCSALPGSITLRSNAKLNDLFTMFNGDKVTTKDKFSCRQAEMSELIQRYELGTLPGRPST
LTASFSGNTLTINCGEAGKSISFTVTITYPSSGTAPYPAIIGYGGGSLPAPAGVAMINFNNDNIAAQVNTGSRGQGK
FYDLYGSSHSAGAMTAWAWGVSRVIDALELVPGARIDTTKIGVTGCSRNGKGAMVAGAFEKRIVLTLPQESGAGGSA
CWRISDYLKSQGANIQTASEIIGEDPWFSTTFNSYVNQVPVLPFDHHSLAALIAPRGLFVIDNNIDWLGPQSCFGCM

TAAHMAWQALGVSDHMGYSQIGAHAHCAFPSNQQSQLTAFVQKFLLGQSTNTAIFQSDFSANQSQWIDWTTPTLS (SEQ ID NO:8)

[0226] Ce11a(里氏木霉)

[0227]

MLPKDFQWGFATAAYQIEGAVDQDGRGPSIWDTFCAQPGKIADGSSGVTACDSYNRTAEDIALLKSLGAKSYRFSIS WSRIIPEGGRGDAVNQAGIDHYVKFVDDLLDAGITPFITLFHWDLPEGLHQRYGGLLNRTEFPLDFENYARVMFRAL PKVRNWITFNEPLCSAIPGYGSGTFAPGRQSTSEPWTVGHNILVAHGRAVKAYRDDFKPASGDGQIGIVLNGDFTYP WDAADPADKEAAERRLEFFTAWFADPIYLGDYPASMRKQLGDRLPTFTPEERALVHGSNDFYGMNHYTSNYIRHRSS PASADDTVGNVDVLFTNKQGNCIGPETQSPWLRPCAAGFRDFLVWISKRYGYPPIYVTENGTSIKGESDLPKEKILE DDFRVKYYNEYIRAMVTAVELDGVNVKGYFAWSLMDNFEWADGYVTRFGVTYVDYENGQKRFPKKSAKSLKPLFDEL IAAA(SEQ ID NO:9)

[0228] Ce13a(里氏木霉)

[0229]

MRYRTAAALALATGPFARADSHSTSGASAEAVVPPAGTPWGTAYDKAKAALAKLNLQDKVGIVSGVGWNGGPCVGNT SPASKISYPSLCLQDGPLGVRYSTGSTAFTPGVQAASTWDVNLIRERGQFIGEEVKASGIHVILGPVAGPLGKTPQG GRNWEGFGVDPYLTGIAMGQTINGIQSVGVQATAKHYILNEQELNRETISSNPDDRTLHELYTWPFADAVQANVASV MCSYNKVNTTWACEDQYTLQTVLKDQLGFPGYVMTDWNAQHTTVQSANSGLDMSMPGTDFNGNNRLWGPALTNAVNS NQVPTSRVDDMVTRILAAWYLTGQDQAGYPSFNISRNVQGNHKTNVRAIARDGIVLLKNDANILPLKKPASIAVVGS AAIIGNHARNSPSCNDKGCDDGALGMGWGSGAVNYPYFVAPYDAINTRASSQGTQVTLSNTDNTSSGASAARGKDVA IVFITADSGEGYITVEGNAGDRNNLDPWHNGNALVQAVAGANSNVIVVVHSVGAIILEQILALPQVKAVVWAGLPSQ ESGNALVDVLWGDVSPSGKLVYTIAKSPNDYNTRIVSGGSDSFSEGLFIDYKHFDDANITPRYEFGYGLSYTKFNYS RLSVLSTAKSGPATGAVVPGGPSDLFQNVATVTVDIANSGQVTGAEVAQLYITYPSSAPRTPPKQLRGFAKLNLTPG QSGTATFNIRRRDLSYWDTASQKWVVPSGSFGISVGASSRDIRLTSTLSVA(SEQ ID NO:10)

[0230] Ce15a(里氏木霉)

[0231]

MNKSVAPLLLAASILYGGAAAQQTVWGQCGGIGWSGPTNCAPGSACSTLNPYYAQCIPGATTITTSTRPPSGPTTTT RATSTSSSTPPTSSGVRFAGVNIAGFDFGCTTDGTCVTSKVYPPLKNFTGSNNYPDGIGQMQHFVNDDGMTIFRLPV GWQYLVNNNLGGNLDSTSISKYDQLVQGCLSLGAYCIVDIHNYARWNGGIIGQGGPTNAQFTSLWSQLASKYASQSR VWFGIMNEPHDVNINTWAATVQEVVTAIRNAGATSQFISLPGNDWQSAGAFISDGSAAALSQVTNPDGSTTNLIFDV HKYLDSDNSGTHAECTTNNIDGAFSPLATWLRQNNRQAILTETGGGNVQSCIQDMCQQIQYLNQNSDVYLGYVGWGA GSFDSTYVLTETPTGSGNSWTDTSLVSSCLARK(SEQ ID NO:11)

[0232] Ce16a(里氏木霉)

[0233]

MIVGILTTLATLATLAASVPLEERQACSSVWGQCGGQNWSGPTCCASGSTCVYSNDYYSQCLPGAASSSSSTRAAST TSRVSPTTSRSSSATPPPGSTTTRVPPVGSGTATYSGNPFVGVTPWANAYYASEVSSLAIPSLTGAMATAAAAVAKV PSFMWLDTLDKTPLMEQTLADIRTANKNGGNYAGQFVVYDLPDRDCAALASNGEYSIADGGVAKYKNYIDTIRQIVV EYSDIRTLLVIEPDSLANLVTNLGTPKCANAQSAYLECINYAVTQLNLPNVAMYLDAGHAGWLGWPANQDPAAQLFA NVYKNASSPRALRGLATNVANYNGWNITSPPSYTQGNAVYNEKLYIHAIGPLLANHGWSNAFFITDQGRSGKQPTGQ QQWGDWCNVIGTGFGIRPSANTGDSLLDSFVWVKPGGECDGTSDSSAPRFDSHCALPDALQPAPQAGAWFQAYFVQL

LTNANPSFL(SEQ ID NO:12)

[0234] Ce17a(里氏木霉)

[0235]

MYRKLAVISAFLATARAQSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTL
CPDNETCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQKNVGARLYLMASDTTYQEFTLLGNEFSFDVDVSQLP
CGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMD
IWEANSISEALTPHPCTTVGQEICEGDGCGGTYSDNRYGGTCDPDGCDWNPYRLGNTSFYGPGSSFTLDTTKKLTVV
TQFETSGAINRYYVQNGVTFQQPNAELGSYSGNELNDDYCTAEEAEFGGSSFSDKGGLTQFKKATSGGMVLVMSLWD
DYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSPNAKVTFSNIKFGPIGSTGNPSGGNPPGGNPPGT
TTTRRPATTTGSSPGPTQSHYGQCGGIGYSGPTVCASGTTCQVLNPYYSQCL(SEQ ID NO:13)

[0236] Ce17b(里氏木霉)

[0237]

MAPSVTLPLTTAILAIARLVAAQQPGTSTPEVHPKLTTYKCTKSGGCVAQDTSVVLDWNYRWMHDANYNSCTVNGGV
NTTLCPDEATCGKNCFIEGVDYAASGVTTSGSSLTMNQYMPSSSGGYSSVSPRLYLLDSDGEYVMLKLNGQELSFDV
DLSALPCGENGSLYLSQMDENGGANQYNTAGANYGSGYCDAQCPVQTWRNGTLNTSHQGFCCNEMDILEGNSRANAL
TPHSCTATACDSAGCGFNPYGSGYKSYYGPGDTVDTSKTFTIITQFNTDNGSPSGNLVSITRKYQQNGVDIPSAQPG
GDTISSCPSASAYGGLATMGKALSSGMVLVFSIWNDNSQYMNWLDSGNAGPCSSTEGNPSNILANNPNTHVVFSNIR
WGDIGSTTNSTAPPPPPASSTTFSTTRRSSTTSSSPSCTQTHWGQCGGIGYSGCKTCTSGTTCQYSNDYYSQCL
(SEQ ID NO:14)

[0238] Ce112a(里氏木霉)

[0239]

MKFLQVLPALIPAALAQTSCDQWATFTGNGYTVSNNLWGASAGSGFGCVTAVSLSGGASWHADWQWSGGQNNVKSYQ
NSQIAIPQKRTVNSISSMPTTASWSYSGSNIRANVAYDLFTAANPNHVTYSGDYELMIWLGKYGDIGPIGSSQGTVN
VGGQSWTLYYGYNGAMQVYSFVAQTNTTNYSGDVKNFFNYLRDNKGYNAAGQYVLSYQFGTEPFTGSGTLNVASWTA
SIN(SEQ ID NO:15)

[0240] Ce145a(里氏木霉)

[0241]

MKATLVLGSLIVGAVSAYKATTTRYYDGQEGACGCGSSSGAFPWQLGIGNGVYTAAGSQALFDTAGASWCGAGCGKC
YQLTSTGQAPCSSCGTGGAAGQSIIVMVTNLCPNNGNAQWCPVVGGTNQYGYSYHFDIMAQNEIFGDNVVVDFEPIA
CPGQAASDWGTCLCVGQQETDPTPVLGNDTGSTPPGSSPPATSSSPPSGGGQQTLYGQCGGAGWTGPTTCQAPGTCK
VQNQWYSQCLP(SEQ ID NO:16)

[0242] Ce174a(里氏木霉)

[0243]

MKVSRVLALVLGAVIPAHAAFSWKNVKLGGGGGFVPGIIFHPKTKGVAYARTDIGGLYRLNADDSWTAVTDGIADNA
GWHNWGIDAVALDPQDDQKVYAAVGMYTNSWDPSNGAIIRSSDRGATWSFTNLPFKVGGNMPGRGAGERLAVDPANS
NIIYFGARSGNGLWKSTDGGVTFSKVSSFTATGTYIPDPSDSNGYNSDKQGLMWVTFDSTSSTTGGATSRIFVGTAD
NITASVYVSTNAGSTWSAVPGQPGKYFPHKAKLQPAEKALYLTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVSGS
DLYFGFGGLGLDLQKPGTLVVASLNSWWPDAQLFRSTDSGTTWSPIWAWASYPTETYYYSISTPKAPWIKNNFIDVT
SESPSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTRHNVSIQSLADGIEEFSVQDLASAPGGSE

LLAAVGDDNGFTFASRNDLGTSPQTVWATPTWATSTSVDYAGNSVKSVVRVGNTAGTQQVAISSDGGATWSIDYAAD TSMNGGTVAYSADGDTILWSTASSGVQRSQFQGSFASVSSLPAGAVIASDKKTNSVFYAGSGSTFYVSKDTGSSFTR GPKLGSAGTIRDIAAHPTTAGTLYVSTDVGIFRSTDSGTTFGQVSTALTNTYQIALGVGSGSNWNLYAFGTGPSGAR LYASGDSGASWTDIQGSQGFGSIDSTKVAGSGSTAGQVYVGTNGRGVFYAQGTVGGGTGGTSSSTKQSSSSTSSASS STTLRSSVVSTTRASTVTSSRTSSAAGPTGSGVAGHYAQCGGIGWTGPTQCVAPYVCQKQNDYYYQCV(SEQ ID NO:17)

[0244] paMan5a(鹅柄孢壳菌)

[0245]

MKGLFAFGLGLLSLVNALPQAQGGGAAASAKVSGTRFVIDGKTGYFAGTNSYWIGFLTNNRDVDTTLDHIASSGLKI LRVWGFNDVNNQPSGNTVWFQRLASSGSQINTGPNGLQRLDYLVRSAETRGIKLIIALVNYWDDFGGMKAYVNAFGG TKESWYTNARAQEQYKRYIQAVVSRYVNSPAIFAWELANEPRCKGCNTNVIFNWATQISDYIRSLDKDHLITLGDEG FGLPGQTTYPYQYGEGTDFVKNLQIKNLDFGTFHMYPGHWGVPTSFGPGWIKDHAAACRAAGKPCLLEEYGYESDRC NVQKGWQQASRELSRDGMSGDLFWQWGDQLSTGQTHNDGFTIYYGSSLATCLVTDHVRAINALPA(SEQ ID NO: 18)

[0246] paMan26a(鹅柄孢壳菌)

[0247]

MVKLLDIGLFALALASSAVAKPCKPRDGPVTYEAEDAILTGTTVDTAQVGYTGRGYVTGFDEGSDKITFQISSATTK LYDLSIRYAAIYGDKRTNVVLNNGAVSEVFFPAGDSFTSVAAGQVLLNAGQNTIDIVNNWGWYLIDSITLTPSAPRP PHDINPNLNNPNADTNAKKLYSYLRSVYGNKIISGQQELHHAEWIRQQTGKTPALVAVDLMDYSPSRVERGTTSHAV EDAIAHHNAGGIVSVLWHWNAPVGLYDTEENKWWSGFYTRATDFDIAATLANPQGANYTLLIRDIDAIAVQLKRLEA AGVPVLWRPLHEAEGGWFWWGAKGPEPAKQLWDILYERLTVHHGLDNLIWVWNSILEDWYPGDDTVDILSADVYAQG NGPMSTQYNELIALGRDKKMIAAAEVGAAPLPGLLQAYQANWLWFAVWGDDFINNPSWNTVAVLNEIYNSDYVLTLD EIQGWRS(SEQ ID NO:19)

[0248] 膨胀蛋白(里氏木霉)

[0249]

MAGKLILVALASLVSLSIQQNCAALFGQCGGIGWSGTTCCVAGAQCSFVNDWYSQCLASTGGNPPNGTTSSSLVSRT SSASSSVGSSSPGGNSPTGSASTYTTTDTATVAPHSQSPYPSIAASSCGSWTLVDNVCCPSYCANDDTSESCSGCGT CTTPPSADCKSGTMYPEVHHVSSNESWHYSRSTHFGLTSGGACGFGLYGLCTKGSVTASWTDPMLGATCDAFCTAYP LLCKDPTGTTLRGNFAAPNGDYYTQFWSSLPGALDNYLSCGECIELIQTKPDGTDYAVGEAGYTDPITLEIVDSCPC SANSKWCCGPGADHCGEIDFKYGCPLPADSIHLDLSDIAMGRLQGNGSLTNGVIPTRYRRVQCPKVGNAYIWLRNGG GPYYFALTAVNTNGPGSVTKIEIKGADTDNWVALVHDPNYTSSRPQERYGSWVIPQGSGPFNLPVGIRLTSPTGEQI VNEQAIKTFTPPATGDPNFYYIDIGVQFSQN(SEQ ID NO:20)

[0250] 适用于本发明的酶混合物中的生物质降解酶的其它实例包括来自芽孢杆菌属、鬼伞属、毁丝霉属、头孢霉属、柱顶孢霉属、青霉属、曲霉属、假单胞菌属、腐质霉属、镰刀菌属、梭孢壳属、枝顶孢属、金孢子菌属和木霉属中的种的酶,特别是由选自以下种的菌株产生的那些:曲霉属(参见例如欧洲公布号0 458 162)、特异腐质霉(重新分类为嗜热柱顶孢霉,参见例如美国专利号4,435,307)、灰盖鬼伞、尖孢镰刀菌、嗜热毁丝霉、大型亚灰树花菌、太瑞斯梭孢壳霉、枝顶孢属种(包括但不限于桃色枝顶孢、A.acremonium、A.brachypenium、A.dichromosporum、A.obclavatum、A.pinkertoniae、灰粉枝顶孢、A.incoloratum和棕色枝

顶孢)。优选菌株包括特异腐质霉DSM 1800、尖孢镰刀菌DSM 2672、嗜热毁丝霉CBS 117.65、头孢霉属种RYM-202、枝顶孢属种CBS 478.94、枝顶孢属种CBS 265.95、桃色枝顶孢CBS 169.65、Acremonium acremonium AHU 9519、头孢霉属种CBS 535.71、Acremonium brachypenium CBS 866.73、Acremonium dichromosporum CBS 683.73、Acremonium obclavatum CBS 311.74、Acremonium pinkertoniae CBS 157.70、灰粉枝顶孢CBS 134.56、Acremonium incoloratum CBS 146.62和棕色枝顶孢CBS 299.70H。生物质降解酶也可从金孢子菌属,优选Chrysosporium lucknowense菌株获得。可使用的额外菌株包括但不限于木霉属(特别是绿色木霉、里氏木霉和康宁木霉)、嗜碱性芽孢杆菌属(alkalophilic Bacillus)(参见例如美国专利号3,844,890和欧洲公布号0 458 162)和链霉菌属(参见例如欧洲公布号0 458 162)。

[0251] 在各实施方案中,在相较于未诱导微生物,适于使生物质降解酶的产生增加的条件下诱导微生物来产生本文所述的生物质降解酶。举例来说,使包含如本文所述的生物质的诱导生物质样品与微生物一起孵育以增加生物质降解酶的产生。对诱导过程的进一步描述可见于US 2014/0011258中,所述专利的内容据此以引用的方式整体并入。

[0252] 由以上提及的微生物产生和/或分泌的生物质降解酶可被分离并添加至本发明的混合物中,或直接添加至糖化反应中。或者,在一个实施方案中,在添加至糖化反应中之前,不使以上提及的表达本文和以上所述的生物质降解酶的微生物或宿主细胞溶解。

[0253] 在一个实施方案中,包含表达一种或多种如本文所述的额外生物质降解酶的宿主细胞的酶混合物可与包含本文所述的具有生物质降解活性的被增溶多肽的混合物一起使用。

[0254] 相较于不添加被增溶多肽的糖化,本文所述的包含从包涵体增溶的具有生物质降解活性的多肽和增溶剂的混合物的使用不抑制、妨碍或降低由糖化获得的糖产物的产率。在一些实施方案中,在使用本文所述的包含从包涵体增溶的具有生物质降解活性的多肽和增溶剂的混合物后,糖产物的产率增加。相较于当将生物质降解酶的标准混合物添加至糖化中而无含有被增溶多肽和被增溶试剂的混合物时,糖产物的产率增加至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少60%、至少70%、至少80%、至少90%、至少100%。

[0255] 进一步加工

[0256] 可对由糖化产生的糖进行进一步加工步骤以产生替代性产物。举例来说,糖可被氢化、发酵或用其它化学物质处理以产生其它产物。

[0257] 葡萄糖可被氢化成山梨糖醇。木糖可被氢化成木糖醇。氢化可通过使用催化剂(例如Pt/γ-Al_2O_3、Ru/C、雷尼镍(Raney Nickel)或本领域中已知的其它催化剂)与H_2组合在高压(例如10至12000psi)下来实现。可使用本领域中已知的方法分离和纯化山梨糖醇和/或木糖醇产物。

[0258] 由糖化获得的糖产物也可被发酵以产生醇;糖醇,诸如赤藻糖醇;或有机酸,例如乳酸、谷氨酸或柠檬酸或氨基酸。

[0259] 酵母和发酵单胞菌属(Zymomonas)细菌,例如可用于将一种或多种糖发酵或转化成一种或多种醇。其它微生物在以下进行讨论。最优发酵pH是约pH 4至7。举例来说,对于酵母最优的pH是约pH 4至5,而对于发酵单胞菌属最优的pH是约pH 5至6。典型发酵时间是约

24至168小时(例如24至96小时),其中温度在20℃至40℃(例如26℃至40℃)的范围内,然而,嗜热微生物偏好较高温度。

[0260] 在一些实施方案中,例如当使用厌氧生物体时,至少一部分发酵是在不存在氧的情况下,例如在惰性气体诸如N_2、Ar、He、CO_2或其混合物的覆盖层下进行。另外,混合物可具有在部分或全部发酵期间流经储罐的惰性气体的恒定吹扫。在一些情况下,可通过发酵期间的二氧化碳产生来实现或维持厌氧条件而不需要额外惰性气体。

[0261] 在一些实施方案中,可在低分子量糖完全转化成产物(例如乙醇)之前中断全部或部分发酵过程。中间体发酵产物包括高浓度的糖和碳水化合物。糖和碳水化合物可通过本领域中已知的任何手段进行分离。这些中间体发酵产物可用于制备用于人或动物消耗的食品。另外或可替代地,可在不锈钢实验室磨机中将中间体发酵产物研磨成细小粒子尺寸以产生面粉状物质。

[0262] 可在发酵期间使用射流混合,并且在一些情况下在同一储罐中进行糖化和发酵。

[0263] 可在糖化和/或发酵期间添加微生物的营养物,例如在2011年7月15日提交的美国专利申请公布2012/0052536中所述的基于食品的营养物包,所述专利申请公布的完整公开内容以引用的方式并入本文。

[0264] "发酵"包括2012年12月22日提交的美国临时申请号61/579,559和2012年12月22日提交的美国临时申请号61/579,576中公开的方法和产物,所述美国临时申请两者的内容均以引用的方式整体并入本文。

[0265] 可利用移动发酵罐,如在国际申请号PCT/US2007/074028(其在2007年7月20日提交,以英语公布为WO 2008/011598并且指定美国)中所描述,所述申请的内容以引用的方式整体并入本文。类似地,糖化设备可为移动的。此外,糖化和/或发酵可在转运期间部分或完全地进行。

[0266] 用于发酵中的微生物可为天然存在的微生物和/或工程改造的微生物。举例来说,微生物可为细菌(包括但不限于例如纤维素分解细菌)、真菌(包括但不限于例如酵母)、植物、原生生物例如原虫或真菌样原生生物(包括但不限于例如粘液霉菌)、或海藻。当生物体可相容时,可利用生物体的混合物。

[0267] 适合发酵微生物能够使碳水化合物(诸如葡萄糖、果糖、木糖、阿拉伯糖、甘露糖、半乳糖、寡糖或多糖)转化成发酵产物。发酵微生物包括以下种属的菌株:酵母属菌种(Saccharomyces spp.)(包括但不限于酿酒酵母(S.cerevisiae)(面包酵母)、糖化酵母(S.distaticus)、葡萄汁酵母(S.uvarum))、克鲁维酵母(Kluyveromyces)属(包括但不限于马克斯克鲁维酵母(K.marxianus)、脆壁克鲁维酵母(K.fragilis))、假丝酵母(Candida)属(包括但不限于假热带假丝酵母(C.pseudotropicalis)和芸薹假丝酵母(C.brassicae))、树干毕赤酵母(Pichia stipitis)(休哈塔假丝酵母(Candida shehatae)的亲缘菌)、棒孢酵母(Clavispora)属(包括但不限于葡萄牙棒孢酵母(C.lusitaniae)和仙人掌棒孢酵母(C.opuntiae))、管囊酵母(Pachysolen)属(包括但不限于嗜鞣管囊酵母(P.tannophilus))、酒香酵母(Bretannomyces)属(包括但不限于,例如铁红梅氏酒香酵母(B.clausenii)(Handbook on Bioethanol:Production and Utilization,Wyman,C.E.编,Taylor&Francis,Washington,D.C.,179-212中的Philippidis,G.P.,1996,Cellulose bioconversion technology))。其它适合微生物包括例如运动发酵单胞菌(Zymomonas

mobilis)、梭菌属菌种(Clostridium spp.)(包括但不限于热纤维梭菌(C.thermocellum)(Philippidis,1996,同上)、糖丁基丙酮梭菌(C.saccharobutylacetonicum)、糖丁酸梭菌(C.saccharobutylicum)、略紫色梭菌(C.Puniceum)、拜氏梭菌(C.beijerncki)以及丙酮丁醇梭菌(C.acetobutylicum))、丛梗孢酵母(Moniliella pollinis)、Moniliella megachiliensis、乳酸杆菌属种(Lactobacillus spp.)、解脂耶氏酵母(Yarrowia lipolytica)、短梗霉属菌种(Aureobasidium sp.)、三型孢菌属菌种(Trichosporonoides sp.)、变异三角酵母(Trigonopsis variabilis)、毛孢子菌属菌种(Trichosporon sp.)、丛梗孢酵母属菌种(Moniliellaacetoabutans sp.)、变异核瑚菌(Typhula variabilis)、木兰假丝酵母(Candida magnoliae)、黑粉菌纲属菌种(Ustilaginomycetes sp.)、筑波拟酵母(Pseudozyma tsukubaensis)、接合酵母属(Zygosaccharomyces)的酵母种、德巴利酵母属(Debaryomyces)、汉逊酵母属(Hansenula)和毕赤酵母属(Pichia)、以及暗丛梗孢形圆酵母属(Torula)的真菌。

[0268] 举例来说,梭菌属种可用于产生乙醇、丁醇、丁酸、乙酸和丙酮。乳酸杆菌属种可用于产生乳酸。

[0269] 许多所述微生物菌株可公开商购获得抑或通过储藏所获得,所述储藏所如诸如ATCC(美国典型培养物保藏中心(American Type Culture Collection),Manassas,Va.,USA)、NRRL(农业研究服务培养物保藏中心(Agricultural Research Service Culture Collection),Peoria,Ill.,USA)或DSMZ(德意志微生物保藏中心(Deutsche Sammlung von Mikroorganismen and Zellkulturen GmbH),Braunschweig,Germany)。

[0270] 可商购获得的酵母包括例如RedStar®/乐斯福乙醇红(Lesaffre Ethanol Red)(可从Red Star/Lesaffre,USA获得)、FALI®(可从Fleischmann's Yeast(Bums Philip Food Inc.的部门),USA获得)、SUPERSTART®(可从Alltech,现在的Lalemand获得)、GERT STRAND®(可从Gert Strand AB,Sweden获得)以及FERMOL®(可从DSM Specialties获得)。

[0271] 可用于糖化生物质材料和产生糖的许多微生物也可用于使那些糖发酵和转化成有用产物。

[0272] 在发酵之后,可使用例如"醪塔"蒸馏所得流体以使乙醇和其它醇与大部分水和残余固体分离。出醪塔的蒸汽可为例如35重量%乙醇并且可被供应至精馏塔中。可使用汽相分子筛将来自精馏塔的接近共沸的(92.5%)乙醇与水的混合物纯化为纯(99.5%)乙醇。可将醪塔底部物传送至三效蒸发器的第一效。精馏塔回流冷凝器可为这个第一效提供热量。在第一效之后,可使用离心机分离固体并且在旋转干燥器中干燥。可将离心机流出液的一部分(25%)再循环至发酵,并且将其余部分传送至第二效蒸发器和第三效蒸发器中。大部分蒸发器冷凝液可作为相当干净的冷凝液返回所述过程中,其中分离一小部分至废水处理以防止低沸点化合物的积聚。

[0273] 可使用由本文所述的方法获得的产物的其它类型的化学转化,例如产生有机糖衍生产物(例如糠醛和糠醛衍生产物)。糖衍生产物的化学转化描述于2012年7月3日提交的美国临时申请号61/667,481中,所述美国临时申请的公开内容以引用的方式整体并入本文。

实施例

[0274] 通过参照以下实验实施例来进一步详述本发明。除非另外规定，否则这些实施例仅出于说明的目的提供，并且不意图具有限制性。因此，本发明决不应解释为限于以下实施例，而是应解释为涵盖由于本文提供的教义而变得明显的任何和所有变化形式。

[0275] 在不进一步描述的情况下，据信本领域普通技术人员可使用先前描述和以下说明性实施例制备和利用本发明化合物以及实施要求保护的方法。以下操作实施例明确指出本发明的各个方面，并且不应解释为以任何方式限制本公开的其余部分。

[0276] 实施例1：在大肠杆菌中表达Cel3a-C'His

[0277] 成熟Cel3a序列(氨基酸20-744)由Genewiz合成并针对大肠杆菌表达进行密码子优化。在以下实施例中提及的Cel3a-C'His是指在C末端具有8xHis(SEQ ID NO:21)标签的密码子优化的成熟Cel3a序列(aa 20-744)。以下引物用于将Cel3a-C'His克隆至pET-Duet(Novagen，目录号71146)中：

[0278] 正向5'-CATGCC**ATG**GGCGATAGTCACAGTACCAGC(SEQ ID NO:4)

[0279] 反向3'-CCCAAGCTT**TCATTA***GTGATGATGATGATGATGATGATGGCTGCCGCTGCC*GGCAACACTCAGGGTGC(SEQ ID NO:5)

[0280] (NcoI和HindIII位点被加下划线；起始密码子和终止密码子用粗体表示；聚组氨酸(8-His(SEQ ID NO:21)标签；和甘氨酸-丝氨酸(GSGS(SEQ ID NO:23))接头用斜体表示。)

[0281] 使用PfuUltra II Fusion HS聚合酶(Agilent，目录号600672)进行扩增反应。

[0282] 通过使用NcoI限制酶(New England Biolabs，R3193)和HindIII限制酶(New England Biolabs，R3104)在由制造商建议的条件下进行限制消化来克隆扩增的DNA。使用T4DNA连接酶(New England Biolabs，M0202)将消化的扩增DNA连接至pETDuet载体中的NcoI-HindIII位点中，随后转化大肠杆菌克隆宿主Top10 One Shot(Invitrogen)。使用Qiagen的质粒纯化试剂盒进行质粒纯化。

[0283] 将Cel3A-C'His构建体转化至大肠杆菌表达宿主Origami B(DE3)(EMD Millipore，目录号70837)中，并且在含有LB培养基和100μg/ml氨苄青霉素(Fisher Scientific，目录号BP1760)、15μg/ml卡那霉素(kanamysin)(Fisher Scientific，目录号BP906)和12.5μg/ml四环素(Fisher Scientified，目录号BP912)的板上划线。从板挑选携带重组DNA的菌落以接种2ml起子培养物(starter culture)，并且在37℃下生长过夜，接着随后用于接种100ml含有适当抗生素的LB培养基。使培养物在37℃下生长直至OD600达到0.8。

[0284] 为诱导蛋白质表达，添加500μM IPTG(异丙基-b-D-硫代半乳糖吡喃糖苷；Fisher Scientific，目录号BP1755)。使表达培养物在37℃下进一步再生长4小时。通过使用Sorvall St16转子TX400，在室温(RT)下在4200下离心30分钟来收集细胞。

[0285] 实施例2：从不溶性部分增溶Cel3a

[0286] 培养表达具有生物质降解活性的酶Cel3a的大肠杆菌培养物，并且诱导酶表达，如实施例1中所述。如下进行从可溶性部分和不溶性部分分离在C末端用His标签标记的Cel3a(Cel3a-C'His)。在4200rpm下离心细胞培养物30分钟。丢弃上清液，并且将细胞沉淀再混悬于具有1mg/mL溶菌酶的溶解缓冲液中。添加溶解酶，例如每克细胞糊状物10μl溶解酶生物处理试剂(Lysonase Bioprocessing Reagent)(EMD Millipore 71320)，并且在环境温度

下孵育样品1小时。在1小时之后，以30秒间隔对样品进行声波处理总计2分钟。在声波处理之后，在10000rpm下离心样品30分钟。上清液或可溶性部分含有被增溶Ce13a，而剩余沉淀或不溶性部分含有包涵体和不溶性Ce13a。

[0287] 将不溶性部分再混悬于含有增溶剂的缓冲液中持续15分钟，并且在室温下涡旋，所述缓冲液具体来说是含6M尿素的pH 8IMAC结合缓冲液。接着将样品经0.45μm膜过滤以为IMAC纯化做准备。添加的含6M尿素的pH 8IMAC结合缓冲液的量与细胞团块的量成比例，例如2或3体积的缓冲液对1体积的细胞团块。结合缓冲液的量随着细胞团块增加而增加以使得对样品的过滤是可能的。

[0288] 实施例3:Ce13a的纯化

[0289] 可溶性Ce13a的纯化

[0290] 将来自实施例2的可溶性部分转移至含有100ul预平衡Bio-Scale™ Profinity (Biorad)Ni装料IMAC树脂浆液(BioRad，目录号732-4614)的新管中。非变性结合缓冲液含有50mM Tris HCl(pH 7.5)、150mM NaCl、0.1％曲通X-100和5μM咪唑。使蛋白质在室温(RT)下分批结合1小时，接着用含有25μM咪唑的非变性缓冲液洗涤。将蛋白质洗脱于300ul含有200μM咪唑的非变性缓冲液中。

[0291] 不溶性Ce13a的纯化

[0292] 用5个柱体积的含6M尿素的pH 8IMAC结合缓冲液以5mL/min的流速使5mL Bio-Scale™小型Profinity IMAC柱筒(BioRad，目录号732-4614)平衡。在柱平衡之后，将来自实施例2的再混悬不溶性部分以1mL/min的流速上样。柱接着接受15个柱体积的含6M尿素的pH 8IMAC结合缓冲液在5mL/min的流速下洗涤。接着用10个柱体积的含6M尿素的pH 4IMAC洗脱缓冲液以5mL/min的流速从柱洗脱被增溶Ce13a。所得被增溶Ce13a样品含有6M尿素。

[0293] 进行IMAC色谱分析(使用IMAC柱，5mL Bio-Scale™小型Profinity™ IMAC柱筒，目录号732-4614)，并且如图1中所示，检测纯化的被增溶Ce13a。

[0294] 进行SDS-PAGE分析以评估以下各部分中由上述纯化获得的Ce13a的量:纯化的可溶性Ce13a(图2中的泳道2)；由对不溶性部分的IMAC纯化获得的流过物(图2中的泳道3)；和来自不溶性部分的纯化Ce13a(被增溶Ce13a)(图2中的泳道4)。如图2中所示，使用上述方法从不溶性部分的包涵体成功分离Ce13a。

[0295] 实施例4:分析被增溶Ce13a的纤维二糖酶活性

[0296] 如实施例3中所述，使用IMAC技术纯化Ce13a。在进行活性测定之前，使用布拉德福德测定和/或nanodrop测定纯化Ce13a的量(效价)。对于nanodrop定量，通过将目标形式的Ce13a的氨基酸序列插入ExPASy ProtParam在线工具中来估计摩尔消光系数。

[0297] 对于活性测定，使用50mM柠檬酸钠，pH 5.0NaOH作为缓冲液制备含有纯化Ce13a的样品的两倍连续稀释液。跨越96孔板的一行来等分稀释液。使稀释液与D-(+)-纤维二糖(Fluka)底物溶液一起在48℃下在50mM柠檬酸二氢钠缓冲液中在pH 5.0下孵育30分钟。使用粘性板密封膜(adhesive plate seal)将板立刻密封，并且放置在设置在48℃、700rpm下的微板孵育振荡器上。在30分钟之后，在加热干燥浴上在100℃下加热样品5分钟以终止反应。接着将板经96孔形式0.45μm Durapore膜过滤。使用YSI生物化学分析仪(YSI Life Sciences)和/或HPLC(UPLC)方法分析滤液样品中的葡萄糖和纤维二糖。将来自可溶性部分和不溶性部分的纯化Ce13a的稀释液的纤维二糖酶活性绘制在图上，并且结果显示于图3

中。图3显示即使在尿素存在下，被增溶Cel3a具有纤维二糖酶活性。

[0298] 也评估尚未通过实施例3中所述的IMAC纯化方法纯化的被增溶Cel3a的纤维二糖酶活性。用溶解缓冲液洗涤表达Cel3a的细胞的细胞沉淀，随后用含有6M尿素的增溶缓冲液增溶。通过上述纤维二糖酶测定来测定于6M尿素缓冲液中含有被增溶Cel3a的粗制溶解产物样品的纤维二糖酶活性。在可溶性Cel3a、可溶性洗涤物和来自粗制溶解产物的被增溶Cel3a之间比较在30分钟内转化成葡萄糖的纤维二糖的百分比(图4)。未经纯化的被增溶Cel3a也具有纤维二糖酶活性。

[0299] 等效案

[0300] 本文引用的各个和每个专利、专利申请和公布的公开内容都据此以引用的方式整体并入本文。尽管已参照特定方面来公开本发明，但明显的是本发明的其它方面和变化形式可由本领域其它技术人员在不脱离本发明的真实精神和范围下进行设计。随附权利要求意图解释为包括所有所述方面和等效变化形式。

序列表

<110> XYLECO, INC.

<120> 被增溶酶及其用途

<130> X2002-7003WO

<140>

<141>

<150> 62/055,702

<151> 2014-09-26

<160> 23

<170> PatentIn 3.5版

<210> 1

<211> 731

<212> PRT

<213> 里氏木霉

<400> 1

Met Gly Asp Ser His Ser Thr Ser Gly Ala Ser Ala Glu Ala Val Val
1 5 10 15
Pro Pro Ala Gly Thr Pro Trp Gly Thr Ala Tyr Asp Lys Ala Lys Ala
20 25 30
Ala Leu Ala Lys Leu Asn Leu Gln Asp Lys Val Gly Ile Val Ser Gly
35 40 45
Val Gly Trp Asn Gly Gly Pro Cys Val Gly Asn Thr Ser Pro Ala Ser
50 55 60
Lys Ile Ser Tyr Pro Ser Leu Cys Leu Gln Asp Gly Pro Leu Gly Val
65 70 75 80
Arg Tyr Ser Thr Gly Ser Thr Ala Phe Thr Pro Gly Val Gln Ala Ala
85 90 95
Ser Thr Trp Asp Val Asn Leu Ile Arg Glu Arg Gly Gln Phe Ile Gly
100 105 110
Glu Glu Val Lys Ala Ser Gly Ile His Val Ile Leu Gly Pro Val Ala
115 120 125
Gly Pro Leu Gly Lys Thr Pro Gln Gly Gly Arg Asn Trp Glu Gly Phe
130 135 140
Gly Val Asp Pro Tyr Leu Thr Gly Ile Ala Met Gly Gln Thr Ile Asn
145 150 155 160
Gly Ile Gln Ser Val Gly Val Gln Ala Thr Ala Lys His Tyr Ile Leu
165 170 175
Asn Glu Gln Glu Leu Asn Arg Glu Thr Ile Ser Ser Asn Pro Asp Asp

180 185 190

Arg Thr Leu His Glu Leu Tyr Thr Trp Pro Phe Ala Asp Ala Val Gln

195 200 205

Ala Asn Val Ala Ser Val Met Cys Ser Tyr Asn Lys Val Asn Thr Thr

210 215 220

Trp Ala Cys Glu Asp Gln Tyr Thr Leu Gln Thr Val Leu Lys Asp Gln

225 230 235 240

Leu Gly Phe Pro Gly Tyr Val Met Thr Asp Trp Asn Ala Gln His Thr

245 250 255

Thr Val Gln Ser Ala Asn Ser Gly Leu Asp Met Ser Met Pro Gly Thr

260 265 270

Asp Phe Asn Gly Asn Asn Arg Leu Trp Gly Pro Ala Leu Thr Asn Ala

275 280 285

Val Asn Ser Asn Gln Val Pro Thr Ser Arg Val Asp Asp Met Val Thr

290 295 300

Arg Ile Leu Ala Ala Trp Tyr Leu Thr Gly Gln Asp Gln Ala Gly Tyr

305 310 315 320

Pro Ser Phe Asn Ile Ser Arg Asn Val Gln Gly Asn His Lys Thr Asn

325 330 335

Val Arg Ala Ile Ala Arg Asp Gly Ile Val Leu Leu Lys Asn Asp Ala

340 345 350

Asn Ile Leu Pro Leu Lys Lys Pro Ala Ser Ile Ala Val Val Gly Ser

355 360 365

Ala Ala Ile Ile Gly Asn His Ala Arg Asn Ser Pro Ser Cys Asn Asp

370 375 380

Lys Gly Cys Asp Asp Gly Ala Leu Gly Met Gly Trp Gly Ser Gly Ala

385 390 395 400

Val Asn Tyr Pro Tyr Phe Val Ala Pro Tyr Asp Ala Ile Asn Thr Arg

405 410 415

Ala Ser Ser Gln Gly Thr Gln Val Thr Leu Ser Asn Thr Asp Asn Thr

420 425 430

Ser Ser Gly Ala Ser Ala Ala Arg Gly Lys Asp Val Ala Ile Val Phe

435 440 445

Ile Thr Ala Asp Ser Gly Glu Gly Tyr Ile Thr Val Glu Gly Asn Ala

450 455 460

Gly Asp Arg Asn Asn Leu Asp Pro Trp His Asn Gly Asn Ala Leu Val

465 470 475 480

Gln Ala Val Ala Gly Ala Asn Ser Asn Val Ile Val Val Val His Ser

485 490 495

Val Gly Ala Ile Ile Leu Glu Gln Ile Leu Ala Leu Pro Gln Val Lys
500 505 510
Ala Val Val Trp Ala Gly Leu Pro Ser Gln Glu Ser Gly Asn Ala Leu
515 520 525
Val Asp Val Leu Trp Gly Asp Val Ser Pro Ser Gly Lys Leu Val Tyr
530 535 540
Thr Ile Ala Lys Ser Pro Asn Asp Tyr Asn Thr Arg Ile Val Ser Gly
545 550 555 560
Gly Ser Asp Ser Phe Ser Glu Gly Leu Phe Ile Asp Tyr Lys His Phe
565 570 575
Asp Asp Ala Asn Ile Thr Pro Arg Tyr Glu Phe Gly Tyr Gly Leu Ser
580 585 590
Tyr Thr Lys Phe Asn Tyr Ser Arg Leu Ser Val Leu Ser Thr Ala Lys
595 600 605
Ser Gly Pro Ala Thr Gly Ala Val Val Pro Gly Gly Pro Ser Asp Leu
610 615 620
Phe Gln Asn Val Ala Thr Val Thr Val Asp Ile Ala Asn Ser Gly Gln
625 630 635 640
Val Thr Gly Ala Glu Val Ala Gln Leu Tyr Ile Thr Tyr Pro Ser Ser
645 650 655
Ala Pro Arg Thr Pro Pro Lys Gln Leu Arg Gly Phe Ala Lys Leu Asn
660 665 670
Leu Thr Pro Gly Gln Ser Gly Thr Ala Thr Phe Asn Ile Arg Arg Arg
675 680 685
Asp Leu Ser Tyr Trp Asp Thr Ala Ser Gln Lys Trp Val Val Pro Ser
690 695 700
Gly Ser Phe Gly Ile Ser Val Gly Ala Ser Ser Arg Asp Ile Arg Leu
705 710 715 720
Thr Ser Thr Leu Ser Val Ala Gly Ser Gly Ser
725 730

<210> 2
<211> 2232
<212> DNA
<213> 里氏木霉
<400> 2
atgcgttacc gaacagcagc tgcgctggca cttgccactg ggcccttgc tagggcagac 60
agtcactcaa catcgggggc ctcggctgag gcagttgtac ctcctgcagg gactccatgg 120
ggaaccgcgt acgacaaggc gaaggccgca ttggcaaagc tcaatctcca agataaggtc 180
ggcatcgtga gcggtgtcgg ctggaacggc ggtccttgcg ttggaaacac atctccggcc 240

tccaagatca gctatccatc gctatgcctt caagacggac ccctcggtgt tcgatactcg 300
acaggcagca cagcctttac gccgggcgtt caagcggcct cgacgtggga tgtcaatttg 360
atccgcgaac gtggacagtt catcggtgag gaggtgaagg cctcggggat tcatgtcata 420
cttggtcctg tggctgggcc gctgggaaag actccgcagg gcggtcgcaa ctgggagggc 480
ttcggtgtcg atccatatct cacgggcatt gccatgggtc aaaccatcaa cggcatccag 540
tcggtaggcg tgcaggcgac agcgaagcac tatatcctca acgagcagga gctcaatcga 600
gaaaccattt cgagcaaccc agatgaccga actctccatg agctgtatac ttggccattt 660
gccgacgcgg ttcaggccaa tgtcgcttct gtcatgtgct cgtacaacaa ggtcaatacc 720
acctgggcct gcgaggatca gtacacgctg cagactgtgc tgaaagacca gctggggttc 780
ccaggctatg tcatgacgga ctggaacgca cagcacacga ctgtccaaag cgcgaattct 840
gggcttgaca tgtcaatgcc tggcacagac ttcaacggta acaatcggct ctggggtcca 900
gctctcacca atgcggtaaa tagcaatcag gtccccacga gcagagtcga cgatatggtg 960
actcgtatcc tcgccgcatg gtacttgaca ggccaggacc aggcaggcta tccgtcgttc 1020
aacatcagca gaaatgttca aggaaaccac aagaccaatg tcagggcaat tgccagggac 1080
ggcatcgttc tgctcaagaa tgacgccaac atcctgccgc tcaagaagcc cgctagcatt 1140
gccgtcgttg gatctgccgc aatcattggt aaccacgcca gaaactcgcc ctcgtgcaac 1200
gacaaaggct gcgacgacgg ggccttgggc atgggttggg gttccggcgc cgtcaactat 1260
ccgtacttcg tcgcgcccta cgatgccatc aataccagag cgtcttcgca gggcacccag 1320
gttaccttga gcaacaccga caacacgtcc tcaggcgcat ctgcagcaag aggaaaggac 1380
gtcgccatcg tcttcatcac cgccgactcg ggtgaaggct acatcaccgt ggagggcaac 1440
gcgggcgatc gcaacaacct ggatccgtgg cacaacggca atgccctggt ccaggcggtg 1500
gccggtgcca acagcaacgt cattgttgtt gtccactccg ttggcgccat cattctggag 1560
cagattcttg ctcttccgca ggtcaaggcc gttgtctggg cgggtcttcc ttctcaggag 1620
agcggcaatg cgctcgtcga cgtgctgtgg ggagatgtca gcccttctgg caagctggtg 1680
tacaccattg cgaagagccc caatgactat aacactcgca tcgtttccgg cggcagtgac 1740
agcttcagcg agggactgtt catcgactat aagcacttcg acgacgccaa tatcacgccg 1800
cggtacgagt tcggctatgg actgtcttac accaagttca actactcacg cctctccgtc 1860
ttgtcgaccg ccaagtctgg tcctgcgact ggggccgttg tgccgggagg cccgagtgat 1920
ctgttccaga atgtcgcgac agtcaccgtt gacatcgcaa actctggcca agtgactggt 1980
gccgaggtag cccagctgta catcacctac ccatcttcag cacccaggac ccctccgaag 2040
cagctgcgag gctttgccaa gctgaacctc acgcctggtc agagcggaac agcaacgttc 2100
aacatccgac gacgagatct cagctactgg gacacggctt cgcagaaatg ggtggtgccg 2160
tcggggtcgt ttggcatcag cgtgggagcg agcagccggg atatcaggct gacgagcact 2220
ctgtcggtag cg 2232

<210> 3
<211> 2232
<212> DNA
<213> 人工序列
<220>

<223> 人工序列的描述:合成

多核苷酸

<400> 3

atgcgttatc gtacagccgc agccctggca ctggccacag gtccgttcgc acgtgccgat 60
agtcacagta ccagcggtgc cagcgcagaa gccgtggttc cgccggcagg cacaccgtgg 120
ggcacagcct atgataaagc caaagccgcc ctggccaagc tgaatctgca ggataaagtg 180
ggcatcgtga gtggcgtggg ctggaacggt ggtccgtgcg ttggcaacac cagcccggca 240
agcaagatca gctatccgag cttatgcctg caggatggtc cgctgggcgt gcgctatagc 300
accggtagta ccgcctttac acctggtgtg caggccgcca gtacctggga cgttaacctg 360
atccgcgaac gtggccaatt tatcggcgaa gaagttaaag ccagcggcat tcatgttatt 420
ctgggtccgg tggccggtcc tctgggtaaa accccgcagg gcggccgtaa ttgggaaggc 480
ttcggcgttg atccgtattt aaccggcatc gcaatgggcc agaccattaa tggcatccag 540
agcgtgggtg ttcaagccac cgccaaacac tacatattaa acgaacagga actgaatcgt 600
gaaaccatca gcagcaatcc ggatgatcgc accctgcatg agctgtatac atggcctttt 660
gccgacgcag ttcaggccaa cgtggcaagt gtgatgtgta gctataacaa ggtgaacacc 720
acctgggcct gcgaagacca gtacaccctg cagaccgttt taaaagacca actgggcttc 780
cctggttacg tgatgacaga ttggaatgcc cagcacacaa ccgttcagag cgcaaacagt 840
ggcctggata tgagcatgcc gggcaccgac ttcaacggca ataatcgtct gtggggtccg 900
gcactgacca atgccgttaa cagcaaccag gtgccgacca gtcgtgtgga cgatatggtt 960
acccgtattc tggccgcctg gtacctgaca ggtcaagacc aggccggcta cccgagcttc 1020
aacatcagcc gcaacgtgca gggtaatcac aagaccaacg ttcgcgcaat cgcacgcgat 1080
ggtatcgtgc tgttaaagaa cgatgccaac attctgccgc tgaaaaaacc ggccagcatc 1140
gccgttgttg gtagcgcagc catcattggc aaccacgccc gtaacagtcc gagctgcaat 1200
gataaaggct gtgacgacgg tgccctgggc atgggttggg gtagtggtgc cgtgaactac 1260
ccgtatttcg tggccccgta cgacgccatt aacacccgtg caagtagcca gggtacccag 1320
gttaccctga gcaacaccga caacacaagc agcggtgcca gtgcagcacg tggtaaggat 1380
gtggccatcg tgttcatcac cgccgacagc ggcgaaggct acattaccgt ggagggtaat 1440
gccggtgatc gcaataatct ggacccgtgg cataacggca acgccctggt tcaggcagtg 1500
gcaggcgcaa atagcaacgt gatcgttgtg gtgcatagcg tgggtgccat cattctggag 1560
cagatcctgg ccctgccgca agttaaggca gttgtgtggg caggtctgcc gagccaagaa 1620
agtggcaatg ccctggtgga cgttctgtgg ggcgatgtta gtccgagcgg caagctggtg 1680
tatacaatcg ccaagagccc gaacgactat aacacccgca tcgttagcgg cggcagtgat 1740
agcttcagcg agggcctgtt tatcgactac aagcatttcg atgatgccaa tattaccccg 1800
cgctacgaat ttggttatgg cctgagctat accaagttca actacagccg cctgagcgtt 1860
ttaagtaccg ccaagagtgg tccggcaaca ggtgccgtgg ttcctggtgg tccgagtgat 1920
ctgtttcaga atgtggccac cgtgaccgtg gatatcgcca acagtggtca ggttaccggc 1980
gccgaagtgg cacagctgta catcacctat ccgagcagtg caccgcgcac cccgccgaaa 2040
cagctgcgtg gcttcgccaa attaaacctg accccgggcc agagcggtac agcaaccttc 2100
aatattcgcc gccgtgatct gagctattgg gacaccgcca gccaaaaatg ggtggtgccg 2160

agcggcagct ttggcattag tgtgggtgca agtagccgcg acattcgctt aacaagcacc 2220
ctgagtgttg cc 2232
<210> 4
<211> 30
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述:合成
引物
<400> 4
catgccatgg gcgatagtca cagtaccagc 30
<210> 5
<211> 68
<212> DNA
<213> 人工序列
<220>
<223> 人工序列的描述:合成
引物
<400> 5
cgtgggactc acaacggccg tcgccgtcgg tagtagtagt agtagtagta gtgattactt 60
tcgaaccc 68
<210> 6
<211> 800
<212> PRT
<213> 鹅柄孢壳菌
<400> 6

Met	Lys	Ser	Ser	Val	Phe	Trp	Gly	Ala	Ser	Leu	Thr	Ser	Ala	Val	Val
1				5				10					15		
Arg	Ala	Ile	Asp	Leu	Pro	Phe	Gln	Phe	Tyr	Pro	Asn	Cys	Val	Asp	Asp
			20					25				30			
Leu	Leu	Ser	Thr	Asn	Gln	Val	Cys	Asn	Thr	Thr	Leu	Ser	Pro	Pro	Glu
		35					40					45			
Arg	Ala	Ala	Ala	Leu	Val	Ala	Ala	Leu	Thr	Pro	Glu	Glu	Lys	Leu	Gln
	50					55					60				
Asn	Ile	Val	Ser	Lys	Ser	Leu	Gly	Ala	Pro	Arg	Ile	Gly	Leu	Pro	Ala
65					70					75					80
Tyr	Asn	Trp	Trp	Ser	Glu	Ala	Leu	His	Gly	Val	Ala	Tyr	Ala	Pro	Gly
				85					90					95	
Thr	Gln	Phe	Trp	Gln	Gly	Asp	Gly	Pro	Phe	Asn	Ser	Ser	Thr	Ser	Phe

100 105 110

Pro Met Pro Leu Leu Met Ala Ala Thr Phe Asp Asp Glu Leu Leu Glu

115 120 125

Lys Ile Ala Glu Val Ile Gly Ile Glu Gly Arg Ala Phe Gly Asn Ala

130 135 140

Gly Phe Ser Gly Leu Asp Tyr Trp Thr Pro Asn Val Asn Pro Phe Lys

145 150 155 160

Asp Pro Arg Trp Gly Arg Gly Ser Glu Thr Pro Gly Glu Asp Val Leu

165 170 175

Leu Val Lys Arg Tyr Ala Ala Ala Met Ile Lys Gly Leu Glu Gly Pro

180 185 190

Val Pro Glu Lys Glu Arg Arg Val Val Ala Thr Cys Lys His Tyr Ala

195 200 205

Ala Asn Asp Phe Glu Asp Trp Asn Gly Ala Thr Arg His Asn Phe Asn

210 215 220

Ala Lys Ile Ser Leu Gln Asp Met Ala Glu Tyr Tyr Phe Met Pro Phe

225 230 235 240

Gln Gln Cys Val Arg Asp Ser Arg Val Gly Ser Ile Met Cys Ala Tyr

245 250 255

Asn Ala Val Asn Gly Val Pro Ser Cys Ala Ser Pro Tyr Leu Leu Gln

260 265 270

Thr Ile Leu Arg Glu His Trp Asn Trp Thr Glu His Asn Asn Tyr Ile

275 280 285

Thr Ser Asp Cys Glu Ala Val Leu Asp Val Ser Leu Asn His Lys Tyr

290 295 300

Ala Ala Thr Asn Ala Glu Gly Thr Ala Ile Ser Phe Glu Ala Gly Met

305 310 315 320

Asp Thr Ser Cys Glu Tyr Glu Gly Ser Ser Asp Ile Pro Gly Ala Trp

325 330 335

Ser Gln Gly Leu Leu Lys Glu Ser Thr Val Asp Arg Ala Leu Leu Arg

340 345 350

Leu Tyr Glu Gly Ile Val Arg Ala Gly Tyr Phe Asp Gly Lys Gln Ser

355 360 365

Leu Tyr Ser Ser Leu Gly Trp Ala Asp Val Asn Lys Pro Ser Ala Gln

370 375 380

Lys Leu Ser Leu Gln Ala Ala Val Asp Gly Thr Val Leu Leu Lys Asn

385 390 395 400

Asp Gly Thr Leu Pro Leu Ser Asp Leu Leu Asp Lys Ser Arg Pro Lys

405 410 415

Lys Val Ala Met Ile Gly Phe Trp Ser Asp Ala Lys Asp Lys Leu Arg
420 425 430
Gly Gly Tyr Ser Gly Thr Ala Ala Tyr Leu His Thr Pro Ala Tyr Ala
435 440 445
Ala Ser Gln Leu Gly Ile Pro Phe Ser Thr Ala Ser Gly Pro Ile Leu
450 455 460
His Ser Asp Leu Ala Ser Asn Gln Ser Trp Thr Asp Asn Ala Met Ala
465 470 475 480
Ala Ala Lys Asp Ala Asp Tyr Ile Leu Tyr Phe Gly Gly Ile Asp Thr
485 490 495
Ser Ala Ala Gly Glu Thr Lys Asp Arg Tyr Asp Leu Asp Trp Pro Gly
500 505 510
Ala Gln Leu Ser Leu Ile Asn Leu Leu Thr Thr Leu Ser Lys Pro Leu
515 520 525
Ile Val Leu Gln Met Gly Asp Gln Leu Asp Asn Thr Pro Leu Leu Ser
530 535 540
Asn Pro Lys Ile Asn Ala Ile Leu Trp Ala Asn Trp Pro Gly Gln Asp
545 550 555 560
Gly Gly Thr Ala Val Met Glu Leu Val Thr Gly Leu Lys Ser Pro Ala
565 570 575
Gly Arg Leu Pro Val Thr Gln Tyr Pro Ser Asn Phe Thr Glu Leu Val
580 585 590
Pro Met Thr Asp Met Ala Leu Arg Pro Ser Ala Gly Asn Ser Gln Leu
595 600 605
Gly Arg Thr Tyr Arg Trp Tyr Lys Thr Pro Val Gln Ala Phe Gly Phe
610 615 620
Gly Leu His Tyr Thr Thr Phe Ser Pro Lys Phe Gly Lys Lys Phe Pro
625 630 635 640
Ala Val Ile Asp Val Asp Glu Val Leu Glu Gly Cys Asp Asp Lys Tyr
645 650 655
Leu Asp Thr Cys Pro Leu Pro Asp Leu Pro Val Val Val Glu Asn Arg
660 665 670
Gly Asn Arg Thr Ser Asp Tyr Val Ala Leu Ala Phe Val Ser Ala Pro
675 680 685
Gly Val Gly Pro Gly Pro Trp Pro Ile Lys Thr Leu Gly Ala Phe Thr
690 695 700
Arg Leu Arg Gly Val Lys Gly Gly Glu Lys Arg Glu Gly Gly Leu Lys
705 710 715 720
Trp Asn Leu Gly Asn Leu Ala Arg His Asp Glu Glu Gly Asn Thr Val

725 730 735

Val Tyr Pro Gly Lys Tyr Glu Val Ser Leu Asp Glu Pro Pro Lys Ala

740 745 750

Arg Leu Arg Phe Glu Ile Val Arg Gly Gly Lys Gly Lys Gly Lys Val

755 760 765

Lys Gly Lys Gly Lys Ala Ala Gln Lys Gly Gly Val Val Leu Asp Arg

770 775 780

Trp Pro Lys Pro Pro Lys Gly Gln Glu Pro Pro Ala Ile Glu Arg Val

785 790 795 800

<210> 7

<211> 316

<212> PRT

<213> 里氏木霉

<400> 7

Met Val Arg Arg Thr Ala Leu Leu Ala Leu Gly Ala Leu Ser Thr Leu

1 5 10 15

Ser Met Ala Gln Ile Ser Asp Asp Phe Glu Ser Gly Trp Asp Gln Thr

20 25 30

Lys Trp Pro Ile Ser Ala Pro Asp Cys Asn Gln Gly Gly Thr Val Ser

35 40 45

Leu Asp Thr Thr Val Ala His Ser Gly Ser Asn Ser Met Lys Val Val

50 55 60

Gly Gly Pro Asn Gly Tyr Cys Gly His Ile Phe Phe Gly Thr Thr Gln

65 70 75 80

Val Pro Thr Gly Asp Val Tyr Val Arg Ala Trp Ile Arg Leu Gln Thr

85 90 95

Ala Leu Gly Ser Asn His Val Thr Phe Ile Ile Met Pro Asp Thr Ala

100 105 110

Gln Gly Gly Lys His Leu Arg Ile Gly Gly Gln Ser Gln Val Leu Asp

115 120 125

Tyr Asn Arg Glu Ser Asp Asp Ala Thr Leu Pro Asp Leu Ser Pro Asn

130 135 140

Gly Ile Ala Ser Thr Val Thr Leu Pro Thr Gly Ala Phe Gln Cys Phe

145 150 155 160

Glu Tyr His Leu Gly Thr Asp Gly Thr Ile Glu Thr Trp Leu Asn Gly

165 170 175

Ser Leu Ile Pro Gly Met Thr Val Gly Pro Gly Val Asp Asn Pro Asn

180 185 190

Asp Ala Gly Trp Thr Arg Ala Ser Tyr Ile Pro Glu Ile Thr Gly Val

195 200 205

Asn Phe Gly Trp Glu Ala Tyr Ser Gly Asp Val Asn Thr Val Trp Phe
210 215 220

Asp Asp Ile Ser Ile Ala Ser Thr Arg Val Gly Cys Gly Pro Gly Ser
225 230 235 240

Pro Gly Gly Pro Gly Ser Ser Thr Thr Gly Arg Ser Ser Thr Ser Gly
245 250 255

Pro Thr Ser Thr Ser Arg Pro Ser Thr Thr Ile Pro Pro Pro Thr Ser
260 265 270

Arg Thr Thr Thr Ala Thr Gly Pro Thr Gln Thr His Tyr Gly Gln Cys
275 280 285

Gly Gly Ile Gly Tyr Ser Gly Pro Thr Val Cys Ala Ser Gly Thr Thr
290 295 300

Cys Gln Val Leu Asn Pro Tyr Tyr Ser Gln Cys Leu
305 310 315

<210> 8

<211> 460

<212> PRT

<213> 里氏木霉

<400> 8

Met Ala Ser Arg Phe Phe Ala Leu Leu Leu Leu Ala Ile Pro Ile Gln
1 5 10 15

Ala Gln Ser Pro Val Trp Gly Gln Cys Gly Gly Ile Gly Trp Ser Gly
20 25 30

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

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

Thr Thr Leu Val Thr Ser Phe Thr Thr Thr Thr Ala Thr Arg Thr Ser
65 70 75 80

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

Ser Ala Leu Pro Gly Ser Ile Thr Leu Arg Ser Asn Ala Lys Leu Asn
100 105 110

Asp Leu Phe Thr Met Phe Asn Gly Asp Lys Val Thr Thr Lys Asp Lys
115 120 125

Phe Ser Cys Arg Gln Ala Glu Met Ser Glu Leu Ile Gln Arg Tyr Glu
130 135 140

Leu Gly Thr Leu Pro Gly Arg Pro Ser Thr Leu Thr Ala Ser Phe Ser

145 150 155 160
Gly Asn Thr Leu Thr Ile Asn Cys Gly Glu Ala Gly Lys Ser Ile Ser
165 170 175
Phe Thr Val Thr Ile Thr Tyr Pro Ser Ser Gly Thr Ala Pro Tyr Pro
180 185 190
Ala Ile Ile Gly Tyr Gly Gly Gly Ser Leu Pro Ala Pro Ala Gly Val
195 200 205
Ala Met Ile Asn Phe Asn Asn Asp Asn Ile Ala Ala Gln Val Asn Thr
210 215 220
Gly Ser Arg Gly Gln Gly Lys Phe Tyr Asp Leu Tyr Gly Ser Ser His
225 230 235 240
Ser Ala Gly Ala Met Thr Ala Trp Ala Trp Gly Val Ser Arg Val Ile
245 250 255
Asp Ala Leu Glu Leu Val Pro Gly Ala Arg Ile Asp Thr Thr Lys Ile
260 265 270
Gly Val Thr Gly Cys Ser Arg Asn Gly Lys Gly Ala Met Val Ala Gly
275 280 285
Ala Phe Glu Lys Arg Ile Val Leu Thr Leu Pro Gln Glu Ser Gly Ala
290 295 300
Gly Gly Ser Ala Cys Trp Arg Ile Ser Asp Tyr Leu Lys Ser Gln Gly
305 310 315 320
Ala Asn Ile Gln Thr Ala Ser Glu Ile Ile Gly Glu Asp Pro Trp Phe
325 330 335
Ser Thr Thr Phe Asn Ser Tyr Val Asn Gln Val Pro Val Leu Pro Phe
340 345 350
Asp His His Ser Leu Ala Ala Leu Ile Ala Pro Arg Gly Leu Phe Val
355 360 365
Ile Asp Asn Asn Ile Asp Trp Leu Gly Pro Gln Ser Cys Phe Gly Cys
370 375 380
Met Thr Ala Ala His Met Ala Trp Gln Ala Leu Gly Val Ser Asp His
385 390 395 400
Met Gly Tyr Ser Gln Ile Gly Ala His Ala His Cys Ala Phe Pro Ser
405 410 415
Asn Gln Gln Ser Gln Leu Thr Ala Phe Val Gln Lys Phe Leu Leu Gly
420 425 430
Gln Ser Thr Asn Thr Ala Ile Phe Gln Ser Asp Phe Ser Ala Asn Gln
435 440 445
Ser Gln Trp Ile Asp Trp Thr Thr Pro Thr Leu Ser
450 455 460

<210> 9
<211> 466
<212> PRT
<213> 里氏木霉
<400> 9

Met Leu Pro Lys Asp Phe Gln Trp Gly Phe Ala Thr Ala Ala Tyr Gln
1 5 10 15
Ile Glu Gly Ala Val Asp Gln Asp Gly Arg Gly Pro Ser Ile Trp Asp
20 25 30
Thr Phe Cys Ala Gln Pro Gly Lys Ile Ala Asp Gly Ser Ser Gly Val
35 40 45
Thr Ala Cys Asp Ser Tyr Asn Arg Thr Ala Glu Asp Ile Ala Leu Leu
50 55 60
Lys Ser Leu Gly Ala Lys Ser Tyr Arg Phe Ser Ile Ser Trp Ser Arg
65 70 75 80
Ile Ile Pro Glu Gly Gly Arg Gly Asp Ala Val Asn Gln Ala Gly Ile
85 90 95
Asp His Tyr Val Lys Phe Val Asp Asp Leu Leu Asp Ala Gly Ile Thr
100 105 110
Pro Phe Ile Thr Leu Phe His Trp Asp Leu Pro Glu Gly Leu His Gln
115 120 125
Arg Tyr Gly Gly Leu Leu Asn Arg Thr Glu Phe Pro Leu Asp Phe Glu
130 135 140
Asn Tyr Ala Arg Val Met Phe Arg Ala Leu Pro Lys Val Arg Asn Trp
145 150 155 160
Ile Thr Phe Asn Glu Pro Leu Cys Ser Ala Ile Pro Gly Tyr Gly Ser
165 170 175
Gly Thr Phe Ala Pro Gly Arg Gln Ser Thr Ser Glu Pro Trp Thr Val
180 185 190
Gly His Asn Ile Leu Val Ala His Gly Arg Ala Val Lys Ala Tyr Arg
195 200 205
Asp Asp Phe Lys Pro Ala Ser Gly Asp Gly Gln Ile Gly Ile Val Leu
210 215 220
Asn Gly Asp Phe Thr Tyr Pro Trp Asp Ala Ala Asp Pro Ala Asp Lys
225 230 235 240
Glu Ala Ala Glu Arg Arg Leu Glu Phe Phe Thr Ala Trp Phe Ala Asp
245 250 255
Pro Ile Tyr Leu Gly Asp Tyr Pro Ala Ser Met Arg Lys Gln Leu Gly
260 265 270

Asp Arg Leu Pro Thr Phe Thr Pro Glu Glu Arg Ala Leu Val His Gly
275 280 285
Ser Asn Asp Phe Tyr Gly Met Asn His Tyr Thr Ser Asn Tyr Ile Arg
290 295 300
His Arg Ser Ser Pro Ala Ser Ala Asp Asp Thr Val Gly Asn Val Asp
305 310 315 320
Val Leu Phe Thr Asn Lys Gln Gly Asn Cys Ile Gly Pro Glu Thr Gln
325 330 335
Ser Pro Trp Leu Arg Pro Cys Ala Ala Gly Phe Arg Asp Phe Leu Val
340 345 350
Trp Ile Ser Lys Arg Tyr Gly Tyr Pro Pro Ile Tyr Val Thr Glu Asn
355 360 365
Gly Thr Ser Ile Lys Gly Glu Ser Asp Leu Pro Lys Glu Lys Ile Leu
370 375 380
Glu Asp Asp Phe Arg Val Lys Tyr Tyr Asn Glu Tyr Ile Arg Ala Met
385 390 395 400
Val Thr Ala Val Glu Leu Asp Gly Val Asn Val Lys Gly Tyr Phe Ala
405 410 415
Trp Ser Leu Met Asp Asn Phe Glu Trp Ala Asp Gly Tyr Val Thr Arg
420 425 430
Phe Gly Val Thr Tyr Val Asp Tyr Glu Asn Gly Gln Lys Arg Phe Pro
435 440 445
Lys Lys Ser Ala Lys Ser Leu Lys Pro Leu Phe Asp Glu Leu Ile Ala
450 455 460
Ala Ala
465

<210> 10
<211> 744
<212> PRT
<213> 里氏木霉
<400> 10

Met Arg Tyr Arg Thr Ala Ala Ala Leu Ala Leu Ala Thr Gly Pro Phe
1 5 10 15
Ala Arg Ala Asp Ser His Ser Thr Ser Gly Ala Ser Ala Glu Ala Val
20 25 30
Val Pro Pro Ala Gly Thr Pro Trp Gly Thr Ala Tyr Asp Lys Ala Lys
35 40 45
Ala Ala Leu Ala Lys Leu Asn Leu Gln Asp Lys Val Gly Ile Val Ser
50 55 60

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

370 375 380

Ser Ala Ala Ile Ile Gly Asn His Ala Arg Asn Ser Pro Ser Cys Asn
385 390 395 400

Asp Lys Gly Cys Asp Asp Gly Ala Leu Gly Met Gly Trp Gly Ser Gly
405 410 415

Ala Val Asn Tyr Pro Tyr Phe Val Ala Pro Tyr Asp Ala Ile Asn Thr
420 425 430

Arg Ala Ser Ser Gln Gly Thr Gln Val Thr Leu Ser Asn Thr Asp Asn
435 440 445

Thr Ser Ser Gly Ala Ser Ala Ala Arg Gly Lys Asp Val Ala Ile Val
450 455 460

Phe Ile Thr Ala Asp Ser Gly Glu Gly Tyr Ile Thr Val Glu Gly Asn
465 470 475 480

Ala Gly Asp Arg Asn Asn Leu Asp Pro Trp His Asn Gly Asn Ala Leu
485 490 495

Val Gln Ala Val Ala Gly Ala Asn Ser Asn Val Ile Val Val Val His
500 505 510

Ser Val Gly Ala Ile Ile Leu Glu Gln Ile Leu Ala Leu Pro Gln Val
515 520 525

Lys Ala Val Val Trp Ala Gly Leu Pro Ser Gln Glu Ser Gly Asn Ala
530 535 540

Leu Val Asp Val Leu Trp Gly Asp Val Ser Pro Ser Gly Lys Leu Val
545 550 555 560

Tyr Thr Ile Ala Lys Ser Pro Asn Asp Tyr Asn Thr Arg Ile Val Ser
565 570 575

Gly Gly Ser Asp Ser Phe Ser Glu Gly Leu Phe Ile Asp Tyr Lys His
580 585 590

Phe Asp Asp Ala Asn Ile Thr Pro Arg Tyr Glu Phe Gly Tyr Gly Leu
595 600 605

Ser Tyr Thr Lys Phe Asn Tyr Ser Arg Leu Ser Val Leu Ser Thr Ala
610 615 620

Lys Ser Gly Pro Ala Thr Gly Ala Val Val Pro Gly Gly Pro Ser Asp
625 630 635 640

Leu Phe Gln Asn Val Ala Thr Val Thr Val Asp Ile Ala Asn Ser Gly
645 650 655

Gln Val Thr Gly Ala Glu Val Ala Gln Leu Tyr Ile Thr Tyr Pro Ser
660 665 670

Ser Ala Pro Arg Thr Pro Pro Lys Gln Leu Arg Gly Phe Ala Lys Leu
675 680 685

Asn Leu Thr Pro Gly Gln Ser Gly Thr Ala Thr Phe Asn Ile Arg Arg
690 695 700
Arg Asp Leu Ser Tyr Trp Asp Thr Ala Ser Gln Lys Trp Val Val Pro
705 710 715 720
Ser Gly Ser Phe Gly Ile Ser Val Gly Ala Ser Ser Arg Asp Ile Arg
725 730 735
Leu Thr Ser Thr Leu Ser Val Ala
740

<210> 11
<211> 418
<212> PRT
<213> 里氏木霉
<400> 11

Met Asn Lys Ser Val Ala Pro Leu Leu Leu Ala Ala Ser Ile Leu Tyr
1 5 10 15
Gly Gly Ala Ala Ala Gln Gln Thr Val Trp Gly Gln Cys Gly Gly Ile
20 25 30
Gly Trp Ser Gly Pro Thr Asn Cys Ala Pro Gly Ser Ala Cys Ser Thr
35 40 45
Leu Asn Pro Tyr Tyr Ala Gln Cys Ile Pro Gly Ala Thr Thr Ile Thr
50 55 60
Thr Ser Thr Arg Pro Pro Ser Gly Pro Thr Thr Thr Thr Arg Ala Thr
65 70 75 80
Ser Thr Ser Ser Ser Thr Pro Pro Thr Ser Ser Gly Val Arg Phe Ala
85 90 95
Gly Val Asn Ile Ala Gly Phe Asp Phe Gly Cys Thr Thr Asp Gly Thr
100 105 110
Cys Val Thr Ser Lys Val Tyr Pro Pro Leu Lys Asn Phe Thr Gly Ser
115 120 125
Asn Asn Tyr Pro Asp Gly Ile Gly Gln Met Gln His Phe Val Asn Asp
130 135 140
Asp Gly Met Thr Ile Phe Arg Leu Pro Val Gly Trp Gln Tyr Leu Val
145 150 155 160
Asn Asn Asn Leu Gly Gly Asn Leu Asp Ser Thr Ser Ile Ser Lys Tyr
165 170 175
Asp Gln Leu Val Gln Gly Cys Leu Ser Leu Gly Ala Tyr Cys Ile Val
180 185 190
Asp Ile His Asn Tyr Ala Arg Trp Asn Gly Gly Ile Ile Gly Gln Gly
195 200 205

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

<210> 12
<211> 471
<212> PRT
<213> 里氏木霉
<400> 12

Met Ile Val Gly Ile Leu Thr Thr Leu Ala Thr Leu Ala Thr Leu Ala
1 5 10 15
Ala Ser Val Pro Leu Glu Glu Arg Gln Ala Cys Ser Ser Val Trp Gly
20 25 30
Gln Cys Gly Gly Gln Asn Trp Ser Gly Pro Thr Cys Cys Ala Ser Gly
35 40 45
Ser Thr Cys Val Tyr Ser Asn Asp Tyr Tyr Ser Gln Cys Leu Pro Gly

50 55 60

Ala Ala Ser Ser Ser Ser Ser Thr Arg Ala Ala Ser Thr Thr Ser Arg
65 70 75 80

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

Ser Thr Thr Thr Arg Val Pro Pro Val Gly Ser Gly Thr Ala Thr Tyr
100 105 110

Ser Gly Asn Pro Phe Val Gly Val Thr Pro Trp Ala Asn Ala Tyr Tyr
115 120 125

Ala Ser Glu Val Ser Ser Leu Ala Ile Pro Ser Leu Thr Gly Ala Met
130 135 140

Ala Thr Ala Ala Ala Ala Val Ala Lys Val Pro Ser Phe Met Trp Leu
145 150 155 160

Asp Thr Leu Asp Lys Thr Pro Leu Met Glu Gln Thr Leu Ala Asp Ile
165 170 175

Arg Thr Ala Asn Lys Asn Gly Gly Asn Tyr Ala Gly Gln Phe Val Val
180 185 190

Tyr Asp Leu Pro Asp Arg Asp Cys Ala Ala Leu Ala Ser Asn Gly Glu
195 200 205

Tyr Ser Ile Ala Asp Gly Gly Val Ala Lys Tyr Lys Asn Tyr Ile Asp
210 215 220

Thr Ile Arg Gln Ile Val Val Glu Tyr Ser Asp Ile Arg Thr Leu Leu
225 230 235 240

Val Ile Glu Pro Asp Ser Leu Ala Asn Leu Val Thr Asn Leu Gly Thr
245 250 255

Pro Lys Cys Ala Asn Ala Gln Ser Ala Tyr Leu Glu Cys Ile Asn Tyr
260 265 270

Ala Val Thr Gln Leu Asn Leu Pro Asn Val Ala Met Tyr Leu Asp Ala
275 280 285

Gly His Ala Gly Trp Leu Gly Trp Pro Ala Asn Gln Asp Pro Ala Ala
290 295 300

Gln Leu Phe Ala Asn Val Tyr Lys Asn Ala Ser Ser Pro Arg Ala Leu
305 310 315 320

Arg Gly Leu Ala Thr Asn Val Ala Asn Tyr Asn Gly Trp Asn Ile Thr
325 330 335

Ser Pro Pro Ser Tyr Thr Gln Gly Asn Ala Val Tyr Asn Glu Lys Leu
340 345 350

Tyr Ile His Ala Ile Gly Pro Leu Leu Ala Asn His Gly Trp Ser Asn
355 360 365

Ala Phe Phe Ile Thr Asp Gln Gly Arg Ser Gly Lys Gln Pro Thr Gly
370 375 380
Gln Gln Gln Trp Gly Asp Trp Cys Asn Val Ile Gly Thr Gly Phe Gly
385 390 395 400
Ile Arg Pro Ser Ala Asn Thr Gly Asp Ser Leu Leu Asp Ser Phe Val
405 410 415
Trp Val Lys Pro Gly Gly Glu Cys Asp Gly Thr Ser Asp Ser Ser Ala
420 425 430
Pro Arg Phe Asp Ser His Cys Ala Leu Pro Asp Ala Leu Gln Pro Ala
435 440 445
Pro Gln Ala Gly Ala Trp Phe Gln Ala Tyr Phe Val Gln Leu Leu Thr
450 455 460
Asn Ala Asn Pro Ser Phe Leu
465 470

<210> 13
<211> 514
<212> PRT
<213> 里氏木霉
<400> 13

Met Tyr Arg Lys Leu Ala Val Ile Ser Ala Phe Leu Ala Thr Ala Arg
1 5 10 15
Ala Gln Ser Ala Cys Thr Leu Gln Ser Glu Thr His Pro Pro Leu Thr
20 25 30
Trp Gln Lys Cys Ser Ser Gly Gly Thr Cys Thr Gln Gln Thr Gly Ser
35 40 45
Val Val Ile Asp Ala Asn Trp Arg Trp Thr His Ala Thr Asn Ser Ser
50 55 60
Thr Asn Cys Tyr Asp Gly Asn Thr Trp Ser Ser Thr Leu Cys Pro Asp
65 70 75 80
Asn Glu Thr Cys Ala Lys Asn Cys Cys Leu Asp Gly Ala Ala Tyr Ala
85 90 95
Ser Thr Tyr Gly Val Thr Thr Ser Gly Asn Ser Leu Ser Ile Gly Phe
100 105 110
Val Thr Gln Ser Ala Gln Lys Asn Val Gly Ala Arg Leu Tyr Leu Met
115 120 125
Ala Ser Asp Thr Thr Tyr Gln Glu Phe Thr Leu Leu Gly Asn Glu Phe
130 135 140
Ser Phe Asp Val Asp Val Ser Gln Leu Pro Cys Gly Leu Asn Gly Ala
145 150 155 160

Leu Tyr Phe Val Ser Met Asp Ala Asp Gly Gly Val Ser Lys Tyr Pro
165 170 175
Thr Asn Thr Ala Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp Ser Gln
180 185 190
Cys Pro Arg Asp Leu Lys Phe Ile Asn Gly Gln Ala Asn Val Glu Gly
195 200 205
Trp Glu Pro Ser Ser Asn Asn Ala Asn Thr Gly Ile Gly Gly His Gly
210 215 220
Ser Cys Cys Ser Glu Met Asp Ile Trp Glu Ala Asn Ser Ile Ser Glu
225 230 235 240
Ala Leu Thr Pro His Pro Cys Thr Thr Val Gly Gln Glu Ile Cys Glu
245 250 255
Gly Asp Gly Cys Gly Gly Thr Tyr Ser Asp Asn Arg Tyr Gly Gly Thr
260 265 270
Cys Asp Pro Asp Gly Cys Asp Trp Asn Pro Tyr Arg Leu Gly Asn Thr
275 280 285
Ser Phe Tyr Gly Pro Gly Ser Ser Phe Thr Leu Asp Thr Thr Lys Lys
290 295 300
Leu Thr Val Val Thr Gln Phe Glu Thr Ser Gly Ala Ile Asn Arg Tyr
305 310 315 320
Tyr Val Gln Asn Gly Val Thr Phe Gln Gln Pro Asn Ala Glu Leu Gly
325 330 335
Ser Tyr Ser Gly Asn Glu Leu Asn Asp Asp Tyr Cys Thr Ala Glu Glu
340 345 350
Ala Glu Phe Gly Gly Ser Ser Phe Ser Asp Lys Gly Gly Leu Thr Gln
355 360 365
Phe Lys Lys Ala Thr Ser Gly Gly Met Val Leu Val Met Ser Leu Trp
370 375 380
Asp Asp Tyr Tyr Ala Asn Met Leu Trp Leu Asp Ser Thr Tyr Pro Thr
385 390 395 400
Asn Glu Thr Ser Ser Thr Pro Gly Ala Val Arg Gly Ser Cys Ser Thr
405 410 415
Ser Ser Gly Val Pro Ala Gln Val Glu Ser Gln Ser Pro Asn Ala Lys
420 425 430
Val Thr Phe Ser Asn Ile Lys Phe Gly Pro Ile Gly Ser Thr Gly Asn
435 440 445
Pro Ser Gly Gly Asn Pro Pro Gly Gly Asn Pro Pro Gly Thr Thr Thr
450 455 460
Thr Arg Arg Pro Ala Thr Thr Thr Gly Ser Ser Pro Gly Pro Thr Gln

465 470 475 480
Ser His Tyr Gly Gln Cys Gly Gly Ile Gly Tyr Ser Gly Pro Thr Val
485 490 495
Cys Ala Ser Gly Thr Thr Cys Gln Val Leu Asn Pro Tyr Tyr Ser Gln
500 505 510
Cys Leu

<210> 14
<211> 459
<212> PRT
<213> 里氏木霉
<400> 14

Met Ala Pro Ser Val Thr Leu Pro Leu Thr Thr Ala Ile Leu Ala Ile
1 5 10 15
Ala Arg Leu Val Ala Ala Gln Gln Pro Gly Thr Ser Thr Pro Glu Val
20 25 30
His Pro Lys Leu Thr Thr Tyr Lys Cys Thr Lys Ser Gly Gly Cys Val
35 40 45
Ala Gln Asp Thr Ser Val Val Leu Asp Trp Asn Tyr Arg Trp Met His
50 55 60
Asp Ala Asn Tyr Asn Ser Cys Thr Val Asn Gly Gly Val Asn Thr Thr
65 70 75 80
Leu Cys Pro Asp Glu Ala Thr Cys Gly Lys Asn Cys Phe Ile Glu Gly
85 90 95
Val Asp Tyr Ala Ala Ser Gly Val Thr Thr Ser Gly Ser Ser Leu Thr
100 105 110
Met Asn Gln Tyr Met Pro Ser Ser Ser Gly Gly Tyr Ser Ser Val Ser
115 120 125
Pro Arg Leu Tyr Leu Leu Asp Ser Asp Gly Glu Tyr Val Met Leu Lys
130 135 140
Leu Asn Gly Gln Glu Leu Ser Phe Asp Val Asp Leu Ser Ala Leu Pro
145 150 155 160
Cys Gly Glu Asn Gly Ser Leu Tyr Leu Ser Gln Met Asp Glu Asn Gly
165 170 175
Gly Ala Asn Gln Tyr Asn Thr Ala Gly Ala Asn Tyr Gly Ser Gly Tyr
180 185 190
Cys Asp Ala Gln Cys Pro Val Gln Thr Trp Arg Asn Gly Thr Leu Asn
195 200 205
Thr Ser His Gln Gly Phe Cys Cys Asn Glu Met Asp Ile Leu Glu Gly
210 215 220

Asn Ser Arg Ala Asn Ala Leu Thr Pro His Ser Cys Thr Ala Thr Ala
225 230 235 240
Cys Asp Ser Ala Gly Cys Gly Phe Asn Pro Tyr Gly Ser Gly Tyr Lys
245 250 255
Ser Tyr Tyr Gly Pro Gly Asp Thr Val Asp Thr Ser Lys Thr Phe Thr
260 265 270
Ile Ile Thr Gln Phe Asn Thr Asp Asn Gly Ser Pro Ser Gly Asn Leu
275 280 285
Val Ser Ile Thr Arg Lys Tyr Gln Gln Asn Gly Val Asp Ile Pro Ser
290 295 300
Ala Gln Pro Gly Gly Asp Thr Ile Ser Ser Cys Pro Ser Ala Ser Ala
305 310 315 320
Tyr Gly Gly Leu Ala Thr Met Gly Lys Ala Leu Ser Ser Gly Met Val
325 330 335
Leu Val Phe Ser Ile Trp Asn Asp Asn Ser Gln Tyr Met Asn Trp Leu
340 345 350
Asp Ser Gly Asn Ala Gly Pro Cys Ser Ser Thr Glu Gly Asn Pro Ser
355 360 365
Asn Ile Leu Ala Asn Asn Pro Asn Thr His Val Val Phe Ser Asn Ile
370 375 380
Arg Trp Gly Asp Ile Gly Ser Thr Thr Asn Ser Thr Ala Pro Pro Pro
385 390 395 400
Pro Pro Ala Ser Ser Thr Thr Phe Ser Thr Thr Arg Arg Ser Ser Thr
405 410 415
Thr Ser Ser Ser Pro Ser Cys Thr Gln Thr His Trp Gly Gln Cys Gly
420 425 430
Gly Ile Gly Tyr Ser Gly Cys Lys Thr Cys Thr Ser Gly Thr Thr Cys
435 440 445
Gln Tyr Ser Asn Asp Tyr Tyr Ser Gln Cys Leu
450 455

<210> 15
<211> 234
<212> PRT
<213> 里氏木霉
<400> 15

Met Lys Phe Leu Gln Val Leu Pro Ala Leu Ile Pro Ala Ala Leu Ala
1 5 10 15
Gln Thr Ser Cys Asp Gln Trp Ala Thr Phe Thr Gly Asn Gly Tyr Thr
20 25 30

Val Ser Asn Asn Leu Trp Gly Ala Ser Ala Gly Ser Gly Phe Gly Cys
35 40 45
Val Thr Ala Val Ser Leu Ser Gly Gly Ala Ser Trp His Ala Asp Trp
50 55 60
Gln Trp Ser Gly Gly Gln Asn Asn Val Lys Ser Tyr Gln Asn Ser Gln
65 70 75 80
Ile Ala Ile Pro Gln Lys Arg Thr Val Asn Ser Ile Ser Ser Met Pro
85 90 95
Thr Thr Ala Ser Trp Ser Tyr Ser Gly Ser Asn Ile Arg Ala Asn Val
100 105 110
Ala Tyr Asp Leu Phe Thr Ala Ala Asn Pro Asn His Val Thr Tyr Ser
115 120 125
Gly Asp Tyr Glu Leu Met Ile Trp Leu Gly Lys Tyr Gly Asp Ile Gly
130 135 140
Pro Ile Gly Ser Ser Gln Gly Thr Val Asn Val Gly Gly Gln Ser Trp
145 150 155 160
Thr Leu Tyr Tyr Gly Tyr Asn Gly Ala Met Gln Val Tyr Ser Phe Val
165 170 175
Ala Gln Thr Asn Thr Thr Asn Tyr Ser Gly Asp Val Lys Asn Phe Phe
180 185 190
Asn Tyr Leu Arg Asp Asn Lys Gly Tyr Asn Ala Ala Gly Gln Tyr Val
195 200 205
Leu Ser Tyr Gln Phe Gly Thr Glu Pro Phe Thr Gly Ser Gly Thr Leu
210 215 220
Asn Val Ala Ser Trp Thr Ala Ser Ile Asn
225 230

<210> 16
<211> 242
<212> PRT
<213> 里氏木霉
<400> 16

Met Lys Ala Thr Leu Val Leu Gly Ser Leu Ile Val Gly Ala Val Ser
1 5 10 15
Ala Tyr Lys Ala Thr Thr Thr Arg Tyr Tyr Asp Gly Gln Glu Gly Ala
20 25 30
Cys Gly Cys Gly Ser Ser Ser Gly Ala Phe Pro Trp Gln Leu Gly Ile
35 40 45
Gly Asn Gly Val Tyr Thr Ala Ala Gly Ser Gln Ala Leu Phe Asp Thr
50 55 60

Ala Gly Ala Ser Trp Cys Gly Ala Gly Cys Gly Lys Cys Tyr Gln Leu
65 70 75 80
Thr Ser Thr Gly Gln Ala Pro Cys Ser Ser Cys Gly Thr Gly Gly Ala
85 90 95
Ala Gly Gln Ser Ile Ile Val Met Val Thr Asn Leu Cys Pro Asn Asn
100 105 110
Gly Asn Ala Gln Trp Cys Pro Val Val Gly Gly Thr Asn Gln Tyr Gly
115 120 125
Tyr Ser Tyr His Phe Asp Ile Met Ala Gln Asn Glu Ile Phe Gly Asp
130 135 140
Asn Val Val Val Asp Phe Glu Pro Ile Ala Cys Pro Gly Gln Ala Ala
145 150 155 160
Ser Asp Trp Gly Thr Cys Leu Cys Val Gly Gln Gln Glu Thr Asp Pro
165 170 175
Thr Pro Val Leu Gly Asn Asp Thr Gly Ser Thr Pro Pro Gly Ser Ser
180 185 190
Pro Pro Ala Thr Ser Ser Ser Pro Pro Ser Gly Gly Gly Gln Gln Thr
195 200 205
Leu Tyr Gly Gln Cys Gly Gly Ala Gly Trp Thr Gly Pro Thr Thr Cys
210 215 220
Gln Ala Pro Gly Thr Cys Lys Val Gln Asn Gln Trp Tyr Ser Gln Cys
225 230 235 240
Leu Pro

<210> 17
<211> 838
<212> PRT
<213> 里氏木霉
<400> 17

Met Lys Val Ser Arg Val Leu Ala Leu Val Leu Gly Ala Val Ile Pro
1 5 10 15
Ala His Ala Ala Phe Ser Trp Lys Asn Val Lys Leu Gly Gly Gly Gly
20 25 30
Gly Phe Val Pro Gly Ile Ile Phe His Pro Lys Thr Lys Gly Val Ala
35 40 45
Tyr Ala Arg Thr Asp Ile Gly Gly Leu Tyr Arg Leu Asn Ala Asp Asp
50 55 60
Ser Trp Thr Ala Val Thr Asp Gly Ile Ala Asp Asn Ala Gly Trp His
65 70 75 80
Asn Trp Gly Ile Asp Ala Val Ala Leu Asp Pro Gln Asp Asp Gln Lys

85 90 95

Val Tyr Ala Ala Val Gly Met Tyr Thr Asn Ser Trp Asp Pro Ser Asn

100 105 110

Gly Ala Ile Ile Arg Ser Ser Asp Arg Gly Ala Thr Trp Ser Phe Thr

115 120 125

Asn Leu Pro Phe Lys Val Gly Gly Asn Met Pro Gly Arg Gly Ala Gly

130 135 140

Glu Arg Leu Ala Val Asp Pro Ala Asn Ser Asn Ile Ile Tyr Phe Gly

145 150 155 160

Ala Arg Ser Gly Asn Gly Leu Trp Lys Ser Thr Asp Gly Gly Val Thr

165 170 175

Phe Ser Lys Val Ser Ser Phe Thr Ala Thr Gly Thr Tyr Ile Pro Asp

180 185 190

Pro Ser Asp Ser Asn Gly Tyr Asn Ser Asp Lys Gln Gly Leu Met Trp

195 200 205

Val Thr Phe Asp Ser Thr Ser Ser Thr Thr Gly Gly Ala Thr Ser Arg

210 215 220

Ile Phe Val Gly Thr Ala Asp Asn Ile Thr Ala Ser Val Tyr Val Ser

225 230 235 240

Thr Asn Ala Gly Ser Thr Trp Ser Ala Val Pro Gly Gln Pro Gly Lys

245 250 255

Tyr Phe Pro His Lys Ala Lys Leu Gln Pro Ala Glu Lys Ala Leu Tyr

260 265 270

Leu Thr Tyr Ser Asp Gly Thr Gly Pro Tyr Asp Gly Thr Leu Gly Ser

275 280 285

Val Trp Arg Tyr Asp Ile Ala Gly Gly Thr Trp Lys Asp Ile Thr Pro

290 295 300

Val Ser Gly Ser Asp Leu Tyr Phe Gly Phe Gly Gly Leu Gly Leu Asp

305 310 315 320

Leu Gln Lys Pro Gly Thr Leu Val Val Ala Ser Leu Asn Ser Trp Trp

325 330 335

Pro Asp Ala Gln Leu Phe Arg Ser Thr Asp Ser Gly Thr Thr Trp Ser

340 345 350

Pro Ile Trp Ala Trp Ala Ser Tyr Pro Thr Glu Thr Tyr Tyr Tyr Ser

355 360 365

Ile Ser Thr Pro Lys Ala Pro Trp Ile Lys Asn Asn Phe Ile Asp Val

370 375 380

Thr Ser Glu Ser Pro Ser Asp Gly Leu Ile Lys Arg Leu Gly Trp Met

385 390 395 400

Ile Glu Ser Leu Glu Ile Asp Pro Thr Asp Ser Asn His Trp Leu Tyr
405 410 415
Gly Thr Gly Met Thr Ile Phe Gly Gly His Asp Leu Thr Asn Trp Asp
420 425 430
Thr Arg His Asn Val Ser Ile Gln Ser Leu Ala Asp Gly Ile Glu Glu
435 440 445
Phe Ser Val Gln Asp Leu Ala Ser Ala Pro Gly Gly Ser Glu Leu Leu
450 455 460
Ala Ala Val Gly Asp Asp Asn Gly Phe Thr Phe Ala Ser Arg Asn Asp
465 470 475 480
Leu Gly Thr Ser Pro Gln Thr Val Trp Ala Thr Pro Thr Trp Ala Thr
485 490 495
Ser Thr Ser Val Asp Tyr Ala Gly Asn Ser Val Lys Ser Val Val Arg
500 505 510
Val Gly Asn Thr Ala Gly Thr Gln Gln Val Ala Ile Ser Ser Asp Gly
515 520 525
Gly Ala Thr Trp Ser Ile Asp Tyr Ala Ala Asp Thr Ser Met Asn Gly
530 535 540
Gly Thr Val Ala Tyr Ser Ala Asp Gly Asp Thr Ile Leu Trp Ser Thr
545 550 555 560
Ala Ser Ser Gly Val Gln Arg Ser Gln Phe Gln Gly Ser Phe Ala Ser
565 570 575
Val Ser Ser Leu Pro Ala Gly Ala Val Ile Ala Ser Asp Lys Lys Thr
580 585 590
Asn Ser Val Phe Tyr Ala Gly Ser Gly Ser Thr Phe Tyr Val Ser Lys
595 600 605
Asp Thr Gly Ser Ser Phe Thr Arg Gly Pro Lys Leu Gly Ser Ala Gly
610 615 620
Thr Ile Arg Asp Ile Ala Ala His Pro Thr Thr Ala Gly Thr Leu Tyr
625 630 635 640
Val Ser Thr Asp Val Gly Ile Phe Arg Ser Thr Asp Ser Gly Thr Thr
645 650 655
Phe Gly Gln Val Ser Thr Ala Leu Thr Asn Thr Tyr Gln Ile Ala Leu
660 665 670
Gly Val Gly Ser Gly Ser Asn Trp Asn Leu Tyr Ala Phe Gly Thr Gly
675 680 685
Pro Ser Gly Ala Arg Leu Tyr Ala Ser Gly Asp Ser Gly Ala Ser Trp
690 695 700
Thr Asp Ile Gln Gly Ser Gln Gly Phe Gly Ser Ile Asp Ser Thr Lys

705 710 715 720
Val Ala Gly Ser Gly Ser Thr Ala Gly Gln Val Tyr Val Gly Thr Asn
725 730 735
Gly Arg Gly Val Phe Tyr Ala Gln Gly Thr Val Gly Gly Gly Thr Gly
740 745 750
Gly Thr Ser Ser Ser Thr Lys Gln Ser Ser Ser Ser Thr Ser Ser Ala
755 760 765
Ser Ser Ser Thr Thr Leu Arg Ser Ser Val Val Ser Thr Thr Arg Ala
770 775 780
Ser Thr Val Thr Ser Ser Arg Thr Ser Ser Ala Ala Gly Pro Thr Gly
785 790 795 800
Ser Gly Val Ala Gly His Tyr Ala Gln Cys Gly Gly Ile Gly Trp Thr
805 810 815
Gly Pro Thr Gln Cys Val Ala Pro Tyr Val Cys Gln Lys Gln Asn Asp
820 825 830
Tyr Tyr Tyr Gln Cys Val
835

<210> 18
<211> 373
<212> PRT
<213> 鹅柄孢壳菌
<400> 18
Met Lys Gly Leu Phe Ala Phe Gly Leu Gly Leu Leu Ser Leu Val Asn
1 5 10 15
Ala Leu Pro Gln Ala Gln Gly Gly Gly Ala Ala Ala Ser Ala Lys Val
20 25 30
Ser Gly Thr Arg Phe Val Ile Asp Gly Lys Thr Gly Tyr Phe Ala Gly
35 40 45
Thr Asn Ser Tyr Trp Ile Gly Phe Leu Thr Asn Asn Arg Asp Val Asp
50 55 60
Thr Thr Leu Asp His Ile Ala Ser Ser Gly Leu Lys Ile Leu Arg Val
65 70 75 80
Trp Gly Phe Asn Asp Val Asn Asn Gln Pro Ser Gly Asn Thr Val Trp
85 90 95
Phe Gln Arg Leu Ala Ser Ser Gly Ser Gln Ile Asn Thr Gly Pro Asn
100 105 110
Gly Leu Gln Arg Leu Asp Tyr Leu Val Arg Ser Ala Glu Thr Arg Gly
115 120 125
Ile Lys Leu Ile Ile Ala Leu Val Asn Tyr Trp Asp Asp Phe Gly Gly

130 135 140

Met Lys Ala Tyr Val Asn Ala Phe Gly Gly Thr Lys Glu Ser Trp Tyr
145 150 155 160
Thr Asn Ala Arg Ala Gln Glu Gln Tyr Lys Arg Tyr Ile Gln Ala Val
165 170 175
Val Ser Arg Tyr Val Asn Ser Pro Ala Ile Phe Ala Trp Glu Leu Ala
180 185 190
Asn Glu Pro Arg Cys Lys Gly Cys Asn Thr Asn Val Ile Phe Asn Trp
195 200 205
Ala Thr Gln Ile Ser Asp Tyr Ile Arg Ser Leu Asp Lys Asp His Leu
210 215 220
Ile Thr Leu Gly Asp Glu Gly Phe Gly Leu Pro Gly Gln Thr Thr Tyr
225 230 235 240
Pro Tyr Gln Tyr Gly Glu Gly Thr Asp Phe Val Lys Asn Leu Gln Ile
245 250 255
Lys Asn Leu Asp Phe Gly Thr Phe His Met Tyr Pro Gly His Trp Gly
260 265 270
Val Pro Thr Ser Phe Gly Pro Gly Trp Ile Lys Asp His Ala Ala Ala
275 280 285
Cys Arg Ala Ala Gly Lys Pro Cys Leu Leu Glu Glu Tyr Gly Tyr Glu
290 295 300
Ser Asp Arg Cys Asn Val Gln Lys Gly Trp Gln Gln Ala Ser Arg Glu
305 310 315 320
Leu Ser Arg Asp Gly Met Ser Gly Asp Leu Phe Trp Gln Trp Gly Asp
325 330 335
Gln Leu Ser Thr Gly Gln Thr His Asn Asp Gly Phe Thr Ile Tyr Tyr
340 345 350
Gly Ser Ser Leu Ala Thr Cys Leu Val Thr Asp His Val Arg Ala Ile
355 360 365
Asn Ala Leu Pro Ala
370

<210> 19
<211> 469
<212> PRT
<213> 鹅柄孢壳菌
<400> 19

Met Val Lys Leu Leu Asp Ile Gly Leu Phe Ala Leu Ala Leu Ala Ser
1 5 10 15
Ser Ala Val Ala Lys Pro Cys Lys Pro Arg Asp Gly Pro Val Thr Tyr

20 25 30

Glu Ala Glu Asp Ala Ile Leu Thr Gly Thr Thr Val Asp Thr Ala Gln

35 40 45

Val Gly Tyr Thr Gly Arg Gly Tyr Val Thr Gly Phe Asp Glu Gly Ser

50 55 60

Asp Lys Ile Thr Phe Gln Ile Ser Ser Ala Thr Thr Lys Leu Tyr Asp

65 70 75 80

Leu Ser Ile Arg Tyr Ala Ala Ile Tyr Gly Asp Lys Arg Thr Asn Val

85 90 95

Val Leu Asn Asn Gly Ala Val Ser Glu Val Phe Phe Pro Ala Gly Asp

100 105 110

Ser Phe Thr Ser Val Ala Ala Gly Gln Val Leu Leu Asn Ala Gly Gln

115 120 125

Asn Thr Ile Asp Ile Val Asn Asn Trp Gly Trp Tyr Leu Ile Asp Ser

130 135 140

Ile Thr Leu Thr Pro Ser Ala Pro Arg Pro Pro His Asp Ile Asn Pro

145 150 155 160

Asn Leu Asn Asn Pro Asn Ala Asp Thr Asn Ala Lys Lys Leu Tyr Ser

165 170 175

Tyr Leu Arg Ser Val Tyr Gly Asn Lys Ile Ile Ser Gly Gln Gln Glu

180 185 190

Leu His His Ala Glu Trp Ile Arg Gln Gln Thr Gly Lys Thr Pro Ala

195 200 205

Leu Val Ala Val Asp Leu Met Asp Tyr Ser Pro Ser Arg Val Glu Arg

210 215 220

Gly Thr Thr Ser His Ala Val Glu Asp Ala Ile Ala His His Asn Ala

225 230 235 240

Gly Gly Ile Val Ser Val Leu Trp His Trp Asn Ala Pro Val Gly Leu

245 250 255

Tyr Asp Thr Glu Glu Asn Lys Trp Trp Ser Gly Phe Tyr Thr Arg Ala

260 265 270

Thr Asp Phe Asp Ile Ala Ala Thr Leu Ala Asn Pro Gln Gly Ala Asn

275 280 285

Tyr Thr Leu Leu Ile Arg Asp Ile Asp Ala Ile Ala Val Gln Leu Lys

290 295 300

Arg Leu Glu Ala Ala Gly Val Pro Val Leu Trp Arg Pro Leu His Glu

305 310 315 320

Ala Glu Gly Gly Trp Phe Trp Trp Gly Ala Lys Gly Pro Glu Pro Ala

325 330 335

Lys Gln Leu Trp Asp Ile Leu Tyr Glu Arg Leu Thr Val His His Gly
340 345 350
Leu Asp Asn Leu Ile Trp Val Trp Asn Ser Ile Leu Glu Asp Trp Tyr
355 360 365
Pro Gly Asp Asp Thr Val Asp Ile Leu Ser Ala Asp Val Tyr Ala Gln
370 375 380
Gly Asn Gly Pro Met Ser Thr Gln Tyr Asn Glu Leu Ile Ala Leu Gly
385 390 395 400
Arg Asp Lys Lys Met Ile Ala Ala Ala Glu Val Gly Ala Ala Pro Leu
405 410 415
Pro Gly Leu Leu Gln Ala Tyr Gln Ala Asn Trp Leu Trp Phe Ala Val
420 425 430
Trp Gly Asp Asp Phe Ile Asn Asn Pro Ser Trp Asn Thr Val Ala Val
435 440 445
Leu Asn Glu Ile Tyr Asn Ser Asp Tyr Val Leu Thr Leu Asp Glu Ile
450 455 460
Gln Gly Trp Arg Ser
465

<210> 20
<211> 493
<212> PRT
<213> 里氏木霉
<400> 20

Met Ala Gly Lys Leu Ile Leu Val Ala Leu Ala Ser Leu Val Ser Leu
1 5 10 15
Ser Ile Gln Gln Asn Cys Ala Ala Leu Phe Gly Gln Cys Gly Gly Ile
20 25 30
Gly Trp Ser Gly Thr Thr Cys Cys Val Ala Gly Ala Gln Cys Ser Phe
35 40 45
Val Asn Asp Trp Tyr Ser Gln Cys Leu Ala Ser Thr Gly Gly Asn Pro
50 55 60
Pro Asn Gly Thr Thr Ser Ser Ser Leu Val Ser Arg Thr Ser Ser Ala
65 70 75 80
Ser Ser Ser Val Gly Ser Ser Ser Pro Gly Gly Asn Ser Pro Thr Gly
85 90 95
Ser Ala Ser Thr Tyr Thr Thr Thr Asp Thr Ala Thr Val Ala Pro His
100 105 110
Ser Gln Ser Pro Tyr Pro Ser Ile Ala Ala Ser Ser Cys Gly Ser Trp
115 120 125

Thr Leu Val Asp Asn Val Cys Cys Pro Ser Tyr Cys Ala Asn Asp Asp
130 135 140
Thr Ser Glu Ser Cys Ser Gly Cys Gly Thr Cys Thr Thr Pro Pro Ser
145 150 155 160
Ala Asp Cys Lys Ser Gly Thr Met Tyr Pro Glu Val His His Val Ser
165 170 175
Ser Asn Glu Ser Trp His Tyr Ser Arg Ser Thr His Phe Gly Leu Thr
180 185 190
Ser Gly Gly Ala Cys Gly Phe Gly Leu Tyr Gly Leu Cys Thr Lys Gly
195 200 205
Ser Val Thr Ala Ser Trp Thr Asp Pro Met Leu Gly Ala Thr Cys Asp
210 215 220
Ala Phe Cys Thr Ala Tyr Pro Leu Leu Cys Lys Asp Pro Thr Gly Thr
225 230 235 240
Thr Leu Arg Gly Asn Phe Ala Ala Pro Asn Gly Asp Tyr Tyr Thr Gln
245 250 255
Phe Trp Ser Ser Leu Pro Gly Ala Leu Asp Asn Tyr Leu Ser Cys Gly
260 265 270
Glu Cys Ile Glu Leu Ile Gln Thr Lys Pro Asp Gly Thr Asp Tyr Ala
275 280 285
Val Gly Glu Ala Gly Tyr Thr Asp Pro Ile Thr Leu Glu Ile Val Asp
290 295 300
Ser Cys Pro Cys Ser Ala Asn Ser Lys Trp Cys Cys Gly Pro Gly Ala
305 310 315 320
Asp His Cys Gly Glu Ile Asp Phe Lys Tyr Gly Cys Pro Leu Pro Ala
325 330 335
Asp Ser Ile His Leu Asp Leu Ser Asp Ile Ala Met Gly Arg Leu Gln
340 345 350
Gly Asn Gly Ser Leu Thr Asn Gly Val Ile Pro Thr Arg Tyr Arg Arg
355 360 365
Val Gln Cys Pro Lys Val Gly Asn Ala Tyr Ile Trp Leu Arg Asn Gly
370 375 380
Gly Gly Pro Tyr Tyr Phe Ala Leu Thr Ala Val Asn Thr Asn Gly Pro
385 390 395 400
Gly Ser Val Thr Lys Ile Glu Ile Lys Gly Ala Asp Thr Asp Asn Trp
405 410 415
Val Ala Leu Val His Asp Pro Asn Tyr Thr Ser Ser Arg Pro Gln Glu
420 425 430
Arg Tyr Gly Ser Trp Val Ile Pro Gln Gly Ser Gly Pro Phe Asn Leu

435 440 445

Pro Val Gly Ile Arg Leu Thr Ser Pro Thr Gly Glu Gln Ile Val Asn

450 455 460

Glu Gln Ala Ile Lys Thr Phe Thr Pro Pro Ala Thr Gly Asp Pro Asn

465 470 475 480

Phe Tyr Tyr Ile Asp Ile Gly Val Gln Phe Ser Gln Asn

485 490

<210> 21

<211> 8

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述:合成

8xHis标签

<400> 21

His His His His His His His His

1 5

<210> 22

<211> 6

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述:合成

6xHis标签

<400> 22

His His His His His His

1 5

<210> 23

<211> 4

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述:合成

肽

<400> 23

Gly Ser Gly Ser

1

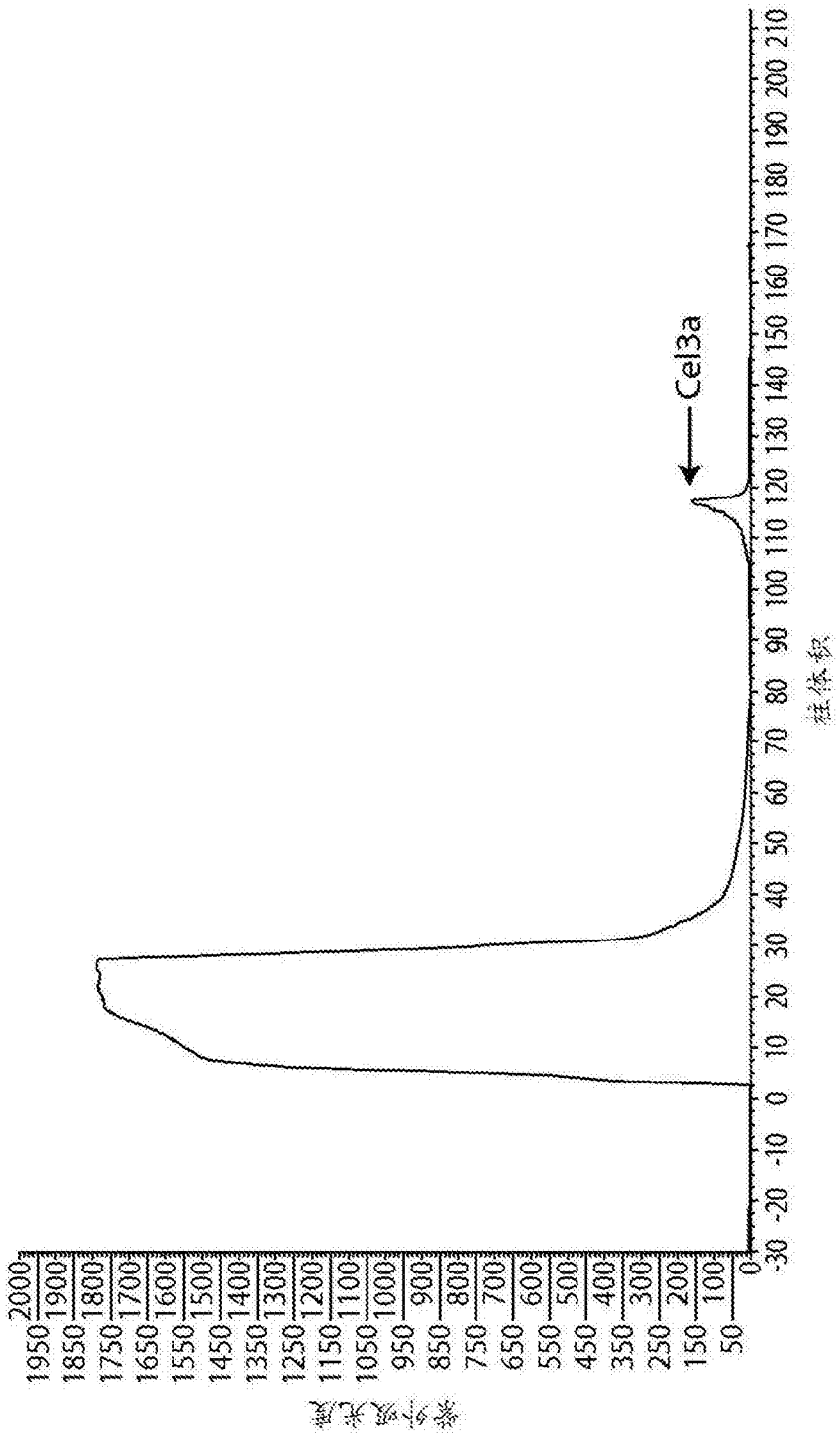


图1

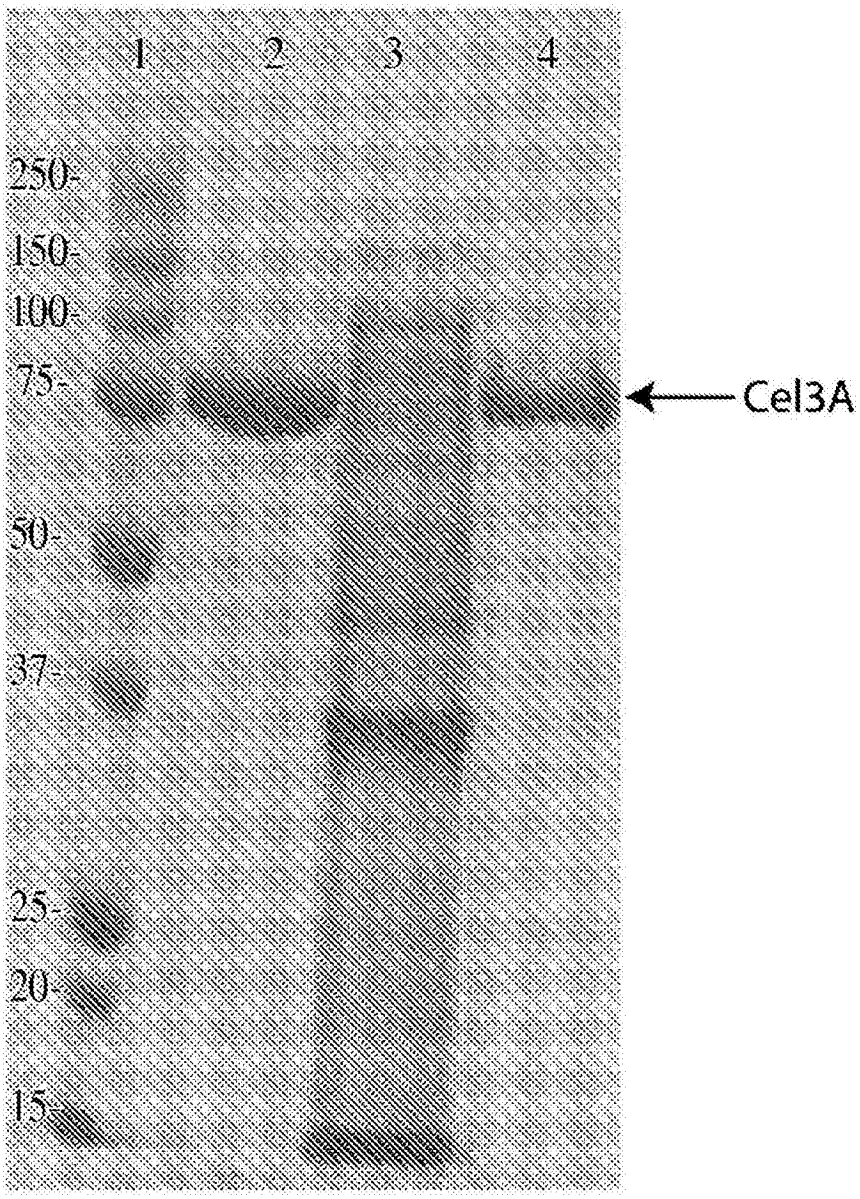


图2

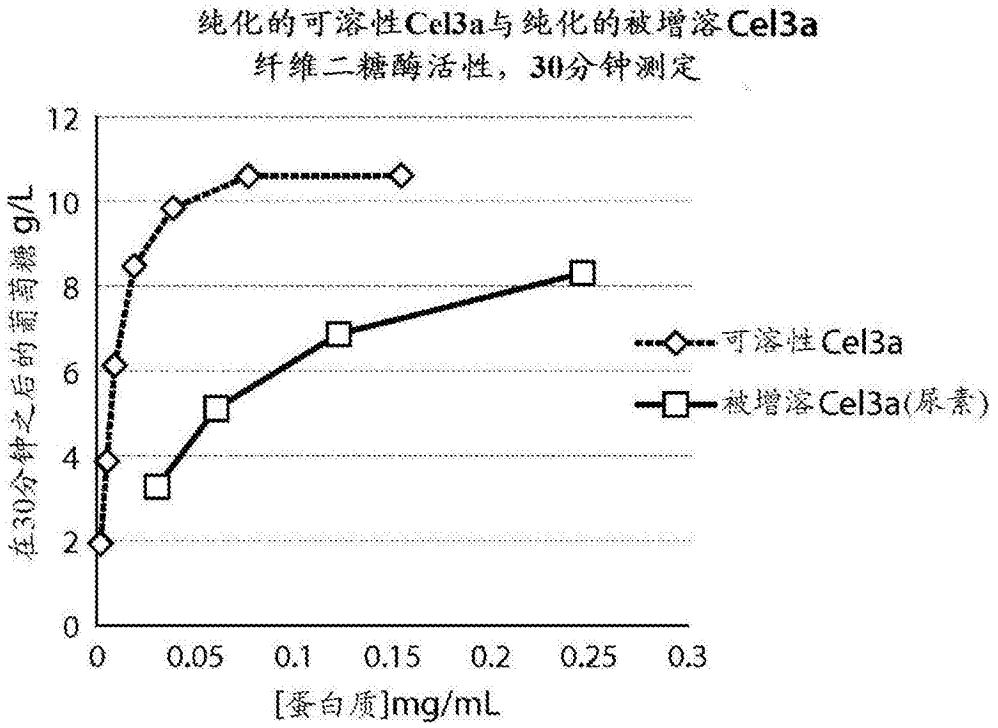

纯化的可溶性Cel3a与纯化的被增溶Cel3a
纤维二糖酶活性，30分钟测定
在30分钟之后的葡萄糖 g/L
12
10
8
6
4
2
0
0
0.05
0.1
0.15
0.2
0.25
0.3
[蛋白质]mg/mL
可溶性Cel3a
被增溶Cel3a(尿素)

图3

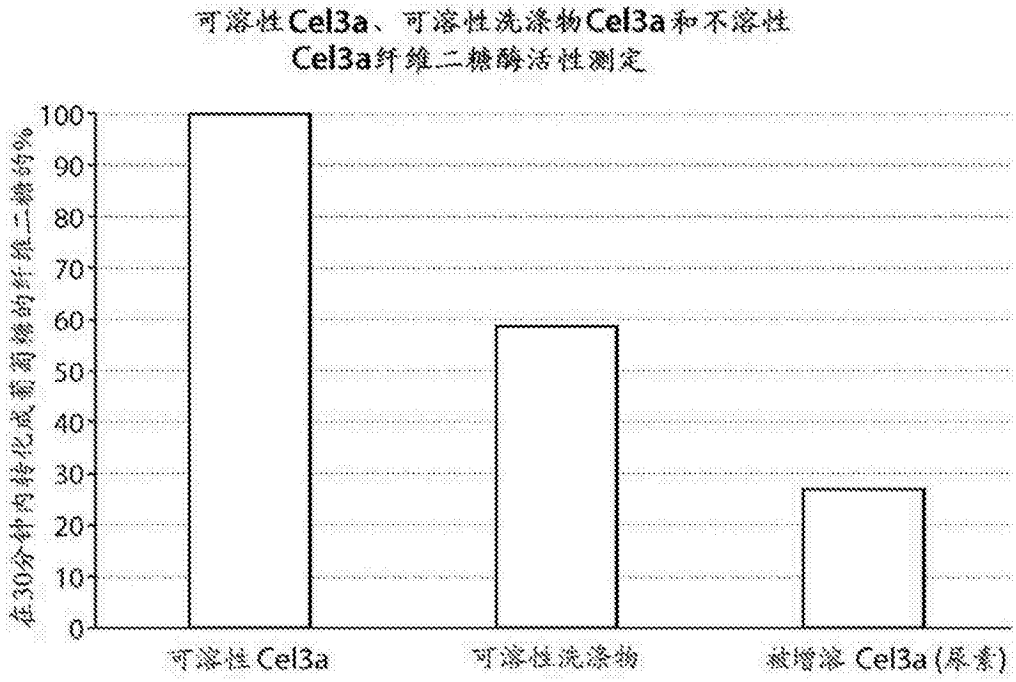

可溶性Cel3a、可溶性洗涤物Cel3a和不溶性
Cel3a纤维二糖酶活性测定
在30分钟内转化成葡萄糖的纤维二糖的%
100
90
80
70
60
50
40
30
20
10
0
可溶性Cel3a
可溶性洗涤物
被增溶 Cel3a (尿素)

图4