(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 109371002 A
(43)申请公布日 2019.02.22

(21)申请号 201811241955.0

(22)申请日 2012.03.16

(30)优先权数据
61/453,918 2011.03.17 US

(62)分案原申请数据
201280013801.0 2012.03.16

(71)申请人 丹尼斯科美国公司
地址 美国加利福尼亚州

(72)发明人 T•卡佩尔 I•尼古拉耶夫
S•兰茨 M•K•福达拉
M•Y•席

(74)专利代理机构 北京市中咨律师事务所 11247
代理人 凌立 黄革生

(51)Int.Cl.
C12N 9/42(2006.01)
C12N 15/56(2006.01)
C12P 19/14(2006.01)

权利要求书2页 说明书108页 附图115页

(54)发明名称

纤维素酶组合物和使用相同组合物用于改善木质纤维素生物质转化成可发酵糖的方法

(57)摘要

本发明涉及可以用于水解生物质的组合物(诸如包含具有β-葡糖苷酶活性的多肽的组合物)、用于水解生物质的方法以及用于改善包含这类β-葡糖苷酶多肽和/或活性的组合物的稳定性和糖化效率的方法。

编号	NT或AA	名称
1.	核苷酸	来自轮枝镰刀菌的GH3酶Fv3A的核苷酸序列
2.	氨基酸	Fv3A的蛋白质序列
3.	核苷酸	来自绳状青霉的GH43酶Pf43A的核苷酸序列
4.	氨基酸	Pf43A的蛋白质序列
5.	核苷酸	来自轮枝镰刀菌的GH43酶Fv43E的核苷酸序列
6.	氨基酸	Fv43E的蛋白质序列
7.	核苷酸	来自轮枝镰刀菌的GH39酶Fv39A的核苷酸序列
8.	氨基酸	Fv39A的蛋白质序列
9.	核苷酸	来自轮枝镰刀菌的GH43酶Fv43A的核苷酸序列
10.	氨基酸	Fv43A的蛋白质序列
11.	核苷酸	来自轮枝镰刀菌的GH43酶Fv43B的核苷酸序列
12.	氨基酸	Fv43B的蛋白质序列
13.	核苷酸	来自鹅掌柄孢霉的GH51酶Pa51A的核苷酸序列
14.	氨基酸	Pa51A的蛋白质序列
15.	核苷酸	来自玉米赤霉的GH43酶Gz43A的核苷酸序列
16.	氨基酸	Gz43A的蛋白质序列
17.	核苷酸	来自尖孢镰刀菌的GH43酶Fo43A的核苷酸序列
18.	氨基酸	Fo43A的蛋白质序列
19.	核苷酸	来自烟曲霉的GH43酶Af43A的核苷酸序列
20.	氨基酸	Af43A的蛋白质序列
21.	核苷酸	来自绳状青霉的GH51酶Pf51A的核苷酸序列
22.	氨基酸	Pf51A的蛋白质序列
23.	核苷酸	来自烟曲霉的GH11酶AfuXyn2的核苷酸序列
24.	氨基酸	AfuXyn2的蛋白质序列
25.	核苷酸	来自烟曲霉的GH11酶AfuXyn5的核苷酸序列
26.	氨基酸	AfuXyn5的蛋白质序列

CN 109371002 A

1.一种分离的多肽,其包含:

a)与SEQ ID NO:135具有至少约70%同一性的氨基酸序列;或者

b)N末端序列和C末端序列,其中所述N末端序列包含衍生自第一β-葡糖苷酶的第一氨基酸序列,长度为至少200个残基,并且包含SEQ ID NOs:164-169的一种或多种或全部,并且其中所述C末端序列包含衍生自第二β-葡糖苷酶的第二氨基酸序列,

长度为至少50个残基,并且包含SEQ ID NO:170,

其中所述多肽具有β-葡糖苷酶活性。

2.根据权利要求1所述的分离的多肽,其包含与SEQ ID NO:135具有至少约80%同一性的氨基酸序列。

3.根据权利要求1或2中任一项所述的分离的多肽,其包含与SEQ ID NO:135具有至少约90%同一性的氨基酸序列。

4.根据权利要求1所述的分离的多肽,其包含衍生自第一β-葡糖苷酶的N末端序列和衍生自第二β-葡糖苷酶的C末端序列,其中所述第一β-葡糖苷酶和所述第二β-葡糖苷酶彼此不同。

5.根据权利要求1或4所述的分离的多肽,其中所述N末端序列和所述C末端序列并不直接连接,而是通过接头结构域功能性地连接。

6.根据权利要求5所述的分离的多肽,其中所述N末端序列、所述C末端序列或所述接头结构域包含长度为3、4、5、6、7、8、9、10或11个氨基酸残基的环区序列,所述环区序列包含SEQ ID NO:171或172的氨基酸序列。

7.根据权利要求1-6中任一项所述的分离的多肽,其与第一β-葡糖苷酶或第二β-葡糖苷酶相比具有改善的稳定性。

8.根据权利要求7所述的分离的多肽,其中所述改善的稳定性是在储存条件或生产条件下增加的蛋白水解切割抵抗性。

9.根据权利要求4-8中任一项所述的分离的多肽,其中所述N末端序列包含与SEQ ID NO:54、56、58、60、62、64、66、68、70、72、74、76、78或79的等长度序列具有至少90%序列同一性的氨基酸序列,其中所述C末端序列包含SEQ ID NO:170的序列基序。

10.根据权利要求4-8中任一项所述的分离的多肽,其中所述N末端序列包含SEQ ID NOs:164-169的序列基序的一种或多种或全部,并且所述C末端序列包含与SEQ ID NO:54、56、58、60、62、64、66、68、70、72、74、76、78或79的等长度序列具有至少90%序列同一性的氨基酸序列。

11.根据权利要求9或10所述的分离的多肽,其中所述N末端序列接在SEQ ID NOs:136-148的序列基序中的3种或更多种、4种或更多种、5种或更多种之后,并且其中所述C末端序列接在SEQ ID NOs:149-156的序列基序中的2种或更多种、3种或更多种、4种或更多种之后。

12.一种组合物,其包含根据权利要求1-11中任一项所述的分离的多肽。

13.根据权利要求12所述的组合物,其还包含一种或多种纤维素酶。

14.根据权利要求13所述的组合物,其中所述一种或多种纤维素酶选自内切葡聚糖酶、GH61/内切葡聚糖酶、纤维二糖水解酶和其他β-葡糖苷酶。

15.根据权利要求12-14中任一项所述的组合物,其中所述组合物还包含一种或多种半

纤维素酶。

16.根据权利要求15所述的组合物，其中所述一种或多种半纤维素酶选自木聚糖酶、β-木糖苷酶或L-α-阿拉伯呋喃糖苷酶。

17.根据权利要求12-16中任一项所述的组合物，其中所述β-葡糖苷酶以相对于所述组合物中蛋白质的总量的1重量％至75重量％的量存在。

18.根据权利要求12-17中任一项所述的组合物，其为培养物混合物或发酵液。

19.根据权利要求18所述的组合物，其为全发酵液制剂。

20.一种分离的多核苷酸：

a)包含与SEQ ID NO:83具有至少70％序列同一性的核苷酸序列；

或者

b)包含能够在高严格性条件下与SEQ ID NO:83或其互补序列杂交的核苷酸序列；或者

c)编码具有β-葡糖苷酶活性的分离的多肽，所述分离的多肽包含与SEQ ID NO:135具有至少约70％同一性的氨基酸序列；或编码具有β-葡糖苷酶活性的分离的多肽，所述分离的多肽包含N末端序列和C末端序列，其中所述N末端序列包含衍生自第一β-葡糖苷酶的第一氨基酸序列，长度为至少200个残基，并且包含SEQ ID NOs:164-169的一种或多种或全部，并且其中所述C末端序列包含衍生自第二β-葡糖苷酶的第二氨基酸序列，长度为至少50个残基，并且包含SEQ ID NO:170。

21.根据权利要求20所述的分离的多核苷酸，其包含与SEQ ID NO:83具有至少90％同一性的核苷酸序列。

22.一种载体，其包含根据权利要求20或21所述的多核苷酸。

23.一种重组宿主细胞，其经工程化以表达根据权利要求20或21所述的多核苷酸。

24.根据权利要求23所述的重组宿主细胞，其为细菌或真菌细胞。

25.根据权利要求24所述的重组宿主细胞，其选自芽孢杆菌或大肠杆菌。

26.根据权利要求24所述的重组宿主细胞，其选自木霉属、曲霉属、金孢属或酵母细胞。

27.一种发酵液或培养物混合物组合物，其通过对根据权利要求23-26中任一项所述重组宿主细胞进行发酵而制备。

28.一种水解纤维素类生物质的方法，其包括使所述生物质与根据权利要求1-11中任一项所述的多肽，或与根据权利要求12-19中任一项所述的组合物，或与根据权利要求27所述的发酵液或培养物混合物组合物接触。

29.根据权利要求28所述的方法，其中所述生物质选自种子、谷粒、块茎、植物废料或食品加工或工业加工的副产物、茎秆、玉米芯、秸秆、叶、草、多年生木质茎、木材、纸材、纸浆和回收纸、马铃薯、大豆、大麦、裸麦、燕麦、小麦、甜菜和甘蔗渣。

30.根据权利要求28或29所述的方法，其中所述生物质经受预处理。

31.根据权利要求30所述的方法，其中所述预处理包括酸预处理或碱预处理，或酸预处理和碱预处理的组合。

32.一种应用根据权利要求1-11中任一项所述的多肽、或根据权利要求12-19中任一项所述的组合物、或根据权利要求27所述的发酵液或培养物混合物组合物的方法、或根据权利要求28-31中任一项所述的水解方法，其在商业环境或工业环境中实施，其中所述方法遵循商业酶供应模式策略或现场生物炼制模式策略。

纤维素酶组合物和使用相同组合物用于改善木质纤维素生物质转化成可发酵糖的方法

[0001] 本申请是申请人于2012年3月16日提交的题为“纤维素酶组合物和使用相同组合物用于改善木质纤维素生物质转化成可发酵糖的方法”的中国专利申请201280013801.0的分案申请。

[0002] 相关申请的交叉引用

[0003] 本专利申请要求提交于2011年3月17日的美国临时申请No. 61/453,918的权益，该申请据此全文以引用方式并入。

技术领域

[0004] 本发明通常涉及某些β-葡糖苷酶和工程化的β-葡糖苷酶组合物、β-葡糖苷酶发酵液组合物，以及包含这类β-葡糖苷酶的其他组合物，以及制备前者或在研究、工业或商业环境中使用(例如，用于将包含半纤维素以及任选的纤维素的生物质糖化或转化成可发酵糖)的方法。

背景技术

[0005] 自二十世纪七十年代石油危机发生以来，研究人员便一直高度关注生物转化可再生性木质纤维素生物质成可发酵糖，所述可发酵糖随后经发酵以产生作为液体燃料替代物的醇(如乙醇)(Bungay,H.R.,“Energy:the biomass options”,NY:Wiley;1981(Bungay,H.R.,《能源:生物质选择》,纽约威利出版社,1981年);Olsson L,Hahn-Hagerdal B.Enzyme Microb Technol 1996,18:312-31(Olsson L,Hahn-Hagerdal B.,《酶微生物技术》,1996年,第18卷,第312-331页);Zaldivar,J et al.,Appl Microbiol Biotechnol 2001,56:17-34(Zaldivar,J等人,《应用微生物学与生物技术》,2001年,第56卷,第17-34页);Galbe,M et al.,Appl Microbiol Biotechnol 2002,59:618-28(Galbe,M等人,《应用微生物学与生物技术》,2002年,第59卷,第618-628页))。在过去几十年，乙醇已经在美国用作汽油的10％掺合物，或者在巴西作为车辆的纯燃料。燃料生物乙醇的重要性将伴随石油价格上涨及其来源逐渐枯竭而增加。另外，可发酵糖越来越多地用来生产塑料、聚合物和基于生物的其他产品。因此，对于可用于替代石油基燃料原料的大量低成本可发酵糖的需求快速增长。

[0006] 有用的可再生生物质当中主要是纤维素和半纤维素(木聚糖)，其可以转化成可发酵糖。将这些多糖酶促转化成可溶性糖(例如葡萄糖、木糖、阿拉伯糖、半乳糖、甘露糖和/或其他己糖和戊糖因多种酶的联合作用而发生。例如，内切-1,4-β-葡聚糖酶(EG)和外切纤维二糖水解酶(CBH)催化不溶性纤维素水解成纤维低聚糖(例如，以纤维二糖作为主要产物)的，而β-葡糖苷酶(BGL)将低聚糖转化成葡萄糖。木聚糖酶与其他辅助蛋白 (半纤维素酶；其非限制性例子包括L-α-阿拉伯呋喃糖苷酶、阿魏酰和乙酰木聚糖酯酶、葡糖苷酸酶和β-木糖苷酶)一起催化半纤维素的水解。

[0007] 植物的细胞壁由借助共价和非共价方式相互作用的复杂多糖的异质混合物组成。高等植物细胞壁的复杂多糖包括，例如，纤维素(β-1,4葡聚糖)，其通常构成细胞壁成分中

存在的碳的35-50％。纤维素聚合物通过氢键、范德华力和疏水相互作用自缔合以形成半结晶纤维素微纤维。这些微纤维还包括非结晶区，一般称为非晶态纤维素。纤维素微纤维包埋于由半纤维素(包括，例如，木聚糖、阿拉伯聚糖和甘露聚糖)、果胶(例如，聚半乳糖醛酸和半乳聚糖)以及多种其他β-1,3和β-1,4葡聚糖形成的基质内。这些基质聚合物通常例如被阿拉伯糖、半乳糖和/或木糖残基置换以生成高度复杂的阿糖基木聚糖、阿糖基半乳聚糖、半乳甘露聚糖和葡聚木糖。半纤维素基质转而由多酚木质素包围。

[0008] 为从生物质获得有用的可发酵糖，通常将木质素透化并且破坏半纤维素以允许纤维素水解酶抵达。酶活性的组合可能是必需的，以便在可以获得可发酵糖之前分解生物质的复杂基质。

[0009] 无论纤维素原料的类型如何，酶的成本和水解效率是限制生物质的生物转化过程商业化的主要因素。微生物产生的酶的生产成本与产酶菌株的生产能力和发酵液中的最终活性收率紧密联系。多酶复合体的水解效率可能取决于多个因素，例如，各种酶的特性、它们之间的协同作用以及它们在所述多酶掺合物中的比率。

[0010] 在本领域中存在着对于确定如下酶和/或酶组合物的需求：所述酶和/ 或酶组合物能够以足够或改善的功效、改善的可发酵糖产率和/或作用于更多种类的纤维素或半纤维素材料的改善能力将植物和/或其他纤维素或半纤维素材料转化成可发酵糖。本文所述的改进方法和组合物提供这类酶组合物，它们能够以低成本并且从可再生来源产生可发酵糖。

[0011] 本文引用的专利、专利申请、文献、核苷酸/蛋白质序列数据库登录号和文章以引用方式全文并入本文。

发明内容

[0012] 本文提供多种β-葡糖苷酶多肽，包括变体、突变体、杂合/嵌合/融合酶、编码这些多肽的核酸、包含这类多肽的组合物以及使用这些组合物的方法。本文的组合物在某些方面为非天然存在的纤维素酶组合物。这些组合物还可以包含一种或多种半纤维素酶，并且其自身为半纤维素酶组合物。在某些方面，所述组合物可以用于将多种生物质转化成可发酵糖的糖化过程中。在某些方面，本文的组合物提供改善的糖化功效或效率以及其他优点。本文还提供细胞(例如，重组工程化宿主细胞)、衍生自这些细胞的发酵液以及使用这些细胞或发酵液的方法或过程。此外，本发明中描述和设想了使用这些多肽、编码这些多肽的核酸以及包含这些多肽的组合物的行业方法。

[0013] 在某些方面，本发明提供一种非天然存在的纤维素酶组合物，其包含β-葡糖苷酶多肽，所述β-葡糖苷酶多肽为至少两种β-葡糖苷酶序列的嵌合体(或杂合体，或融合体，这些术语在本文可互换地用来指称相同的概念)。在某些方面，非天然存在的纤维素酶组合物包含β-葡糖苷酶活性。该组合物还可以包含一种或多种木聚糖酶、β-木糖苷酶和/或L-α-阿拉伯呋喃糖苷酶活性。因此，该组合物可以是半纤维素酶组合物。非天然存在的纤维素酶/半纤维素酶组合物包含衍生自至少两种不同来源的成分。在某些方面，非天然存在的纤维素酶/半纤维素酶组合物包含一种或多种天然存在的半纤维素酶。该组合物中的β-葡糖苷酶多肽还可以包含一个或多个糖基化位点。在某些方面，β-葡糖苷酶多肽包含N末端序列和C末端序列，其中N末端序列或C末端序列中的每一个包含衍生自不同β-葡糖苷酶的一个或

多个子序列。在某些方面,N末端和C末端序列衍生自不同来源。在一些实施例中,N末端和C末端序列的一个或多个子序列中的至少两个衍生自不同来源。在某些方面,N末端序列或C末端序列还包含长度为约3、4、5、6、7、8、9、10或11个氨基酸残基的环区序列。在某些实施例中,N末端序列和C末端序列是紧邻的或直接连接的。在其他实施例中,N末端和C末端序列并非紧邻,而是通过接头结构域功能性地连接。在某些实施例中,该接头结构域位于该嵌合多肽的中央(例如,并非位于N末端或C末端)。在某些实施例中,杂合多肽的N末端序列或C末端序列均不包含环序列。相反,接头结构域包含环序列。在某些方面,N末端序列包含β-葡糖苷酶或其变体的第一氨基酸序列,其长度为至少约200(例如,约200、250、300、350、400、450、500、550或600)个残基。在某些方面,N末端序列包含SEQ ID NOs:136-148所代表的多肽序列基序的一种或多种或全部。在某些方面,C末端序列包含β-葡糖苷酶或其变体的第二氨基酸序列,其长度为至少约50(例如,约50、75、100、125、150、175或200)个氨基酸残基。在某些方面,C末端序列包含SEQ ID NOs:149-156所代表的多肽序列基序的一种或多种或全部。特别地,所述两种或更多种β-葡糖苷酶序列的第一种是具有至少约200个氨基酸残基长度并且包含SEQ ID NOs:164-169的氨基酸序列基序中的至少2种(例如,至少2、3、4或全部)的序列,并且所述两种或更多种β-葡糖苷酶序列的第二种具有至少50个氨基酸残基长度并且包含SEQ ID NO:170。在某些方面,C末端或N末端序列包含环序列,所述环序列包含约3、4、5、6、7、8、9、10或11个氨基酸残基,所述序列包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。在某些方面,C末端或N末端序列均不包含环序列。在一些实施例中,C末端序列与N末端序列通过包含环序列的接头结构域连接,所述环序列包含约3、4、5、6、7、8、9、10或11个氨基酸残基,包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。在某些实施例中,所述β-葡糖苷酶多肽包含与SEQ ID NO:135具有至少约65%(例如,至少约65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)同一性的序列。在一些实施例中,具有β-葡糖苷酶活性的多肽(即,β-葡糖苷酶多肽)由与SEQ ID NO:83具有至少约65%(例如,至少约65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%)同一性的核苷酸所编码,或由能够在高严格性条件下与SEQ ID NO:83或其互补物杂交的多核苷酸所编码。在某些方面,非天然存在的纤维素酶或半纤维素酶组合物中的β-葡糖苷酶多肽具有优于衍生所述嵌合多肽的每个C末端和/或N末端序列的任何天然酶的改善稳定性。在某些方面,改善的稳定性包括在储存、表达或生产过程期间蛋白水解稳定性的改善。在某些方面,改善的稳定性包括在储存或生产条件下相关的酶活性损失的速度或程度降低,其中酶活性损失优选地低于约50%,低于约40%,低于约30%或低于约20%,更优选地低于15%,或低于10%。

[0014] 本发明的多肽可以适合地以“基本纯净”的形式获得和/或使用。例如,本发明的多肽在给定的组合物中构成至少约80重量%(例如至少约85重量%、90重量%、91重量%、92重量%、93重量%、94重量%、95重量%、96重量%、97重量%、98重量%或99重量%)的总蛋白,所述组合物还包括其他成分,诸如缓冲剂或溶液。

[0015] 在某些方面,本发明提供核酸,其编码β-葡糖苷酶多肽,包括变体、突变体和杂合/融合/嵌合多肽。例如,本发明提供分离的核酸,其编码β-葡糖苷酶多肽,其中所述核酸是与SEQ ID NO:83具有至少约65%(例如,至少约65%、70%、75%、80%、85%、90%、91%、

92％、93％、 94％、95％、96％、97％、98％、99％或100％)同一性的核酸,或是能够在高严格性条件下与SEQ ID NO:83或其互补物杂交的核酸。本发明还提供了包含这类核酸分子的宿主细胞。在一些实施例中,本发明还提供了适合与核酸分子和宿主细胞一起使用的启动子和载体。在某些方面,本发明提供通过发酵宿主细胞而制备的组合物,所述组合物包括纤维素酶组合物或半纤维素酶组合物。因此,本发明提供发酵液组合物。

[0016] 在某些方面,本发明提供使用所述组合物、多肽、细胞或编码本文多肽的核酸来实现生物质底物/材料糖化的方法。在某些实施例中,将所述生物质底物/材料适当地预处理或使其经历合适的预处理方法。在一些实施例中,本发明还提供与本文所述的组合物、多肽、细胞或核酸相关的某些商业或行业方法。

附图说明

[0017] 下述的图和表旨在举例说明而不是限制本文公开说明书或权利要求的范围和内容。

[0018] 图1:提供多种酶和编码这些酶中某些者的核苷酸的本发明中所用序列标识符的汇总。

[0019] 图2提供某些β-葡糖苷酶(例如,Fv3C)同源物之间的保守残基,所述保守残基基于在-1亚位点中与葡萄糖复合的新阿波罗栖热袍菌(T. neapolitana)Bgl3B的晶体结构(蛋白质数据库登录号pdb:2X41的晶体结构)预测。

[0020] 图3:提供了由里氏木霉(T.reesei)整合型菌株H3A产生的发酵液的酶组合物。

[0021] 图4A-4E:图4A列出实例2中单独添加至每一份样品的酶(纯化或未纯化),以及这些酶的母液蛋白浓度。图4B示出了通过添加酶组合物糖化经稀氨水预处理的玉米芯后释放的葡萄糖的量,所述酶组合物包含图 4A的多种纯化或未纯化的酶,其根据实例2而添加至里氏木霉整合型菌株 H3A。图4C示出了通过添加酶组合物糖化经稀氨水预处理的玉米芯后释放的纤维二糖的量,所述酶组合物包含图4A的多种纯化或未纯化的酶,其根据实例2而添加至里氏木霉整合型菌株H3A。图4D示出了通过添加酶组合物糖化经稀氨水预处理的玉米芯后释放的木二糖的量,所述酶组合物包含图4A的多种纯化或未纯化的酶,其根据实例2而添加至里氏木霉整合型菌株H3A。图4E示出了通过添加酶组合物糖化经稀氨水预处理的玉米芯后释放的木糖的量,所述酶组合物包含图4A的多种纯化或未纯化的酶,其根据实例2而添加至里氏木霉整合型菌株H3A。

[0022] 图5A-5B:图5A列出了多种β-葡糖苷酶同源物的β-葡糖苷酶活性,所述β-葡糖苷酶同源物包括里氏木霉Bgl1(Tr3A)、黑曲霉(A.niger)Bglu (An3A)、Fv3C、Fv3D和Pa3C。根据实例4,对纤维二糖和CNPG底物测量活性;图5B将另一组β-葡糖苷酶同源物相对于里氏木霉Bgl1对纤维二糖和CNPG底物的活性根据实例5A进行比较。

[0023] 图6:列出了在实例5B-D中测试的酶混合物/组合物中的酶的相对重量。

[0024] 图7:提供酶组合物对经稀氨水预处理的玉米芯的作用的比较。

[0025] 图8A-8B:图8A示出了Fv3A的核苷酸序列(SEQ ID NO:1)。图8B示出了Fv3A的氨基酸序列(SEQ ID NO:2)。预测的信号序列以下划线标出。预测的保守结构域以粗体标出。

[0026] 图9A-9B:图9A示出了Pf43A的核苷酸序列(SEQ ID NO:3)。图9B 示出了Pf43A的氨基酸序列(SEQ ID NO:4)。预测的信号序列以下划线标出,预测的保守结构域以粗体标出,

预测的碳水化合物结合模块 (“CBM”)以大写字母表示,分隔CD和CBM的预测接头以斜体表示。

[0027] 图10A-10B:图10A示出了Fv43E的核苷酸序列(SEQ ID NO:5)。图 10B示出了Fv43E的氨基酸序列(SEQ ID NO:6)。预测的信号序列以下划线标出。预测的保守结构域以粗体标出。

[0028] 图11A-11B:图11A示出了Fv39A的核苷酸序列(SEQ ID NO:7)。图11B示出了Fv39A的氨基酸序列(SEQ ID NO:8)。预测的信号序列以下划线标出。预测的保守结构域以粗体标出。

[0029] 图12A-12B:图12A示出了Fv43A的核苷酸序列(SEQ ID NO:9)。图 12B示出了Fv43A的氨基酸序列(SEQ ID NO:10)。预测的信号序列以下划线标出。预测的保守结构域以粗体标出,预测的CBM以大写字母表示,分隔保守结构域和CBM的预测接头以斜体表示。

[0030] 图13A-13B:图13A示出了Fv43B的核苷酸序列(SEQ ID NO:11)。图 13B示出了Fv43B的氨基酸序列(SEQ ID NO:12)。预测的信号序列以下划线标出。预测的保守结构域以粗体标出。

[0031] 图14A-14B:图14A示出了Pa51A的核苷酸序列(SEQ ID NO:13)。图 14B示出了Pa51A的氨基酸序列(SEQ ID NO:14)。预测的信号序列以下划线标出。预测的L-α-阿拉伯呋喃糖苷酶保守结构域以粗体标出。为了在里氏木霉中表达,将基因组DNA作密码子优化(参见图27C)。

[0032] 图15A-15B:图15A示出了Gz43A的核苷酸序列(SEQ ID NO:15)。图 15B示出了Gz43A的氨基酸序列(SEQ ID NO:16)。预测的信号序列以下划线标出,并且预测的保守结构域以粗体标出。为了在里氏木霉中表达,将预测的信号序列在里氏木霉中替换为里氏木霉CBH1信号序列 (MYRKLAVISAFLATARA(SEQ ID NO:159))。

[0033] 图16A-16B:图16A示出了Fo43A的核苷酸序列(SEQ ID NO:17)。图 16B示出了Fo43A的氨基酸序列(SEQ ID NO:18)。预测的信号序列以下划线标出。预测的保守结构域以粗体标出。为了在里氏木霉中进行表达,将预测的信号序列替换为里氏木霉CBH1信号序列(MYRKLAVISAFLATARA (SEQ ID NO:159))。

[0034] 图17A-17B:图17A示出了Af43A的核苷酸序列(SEQ ID NO:19)。图 17B示出了Af43A的氨基酸序列(SEQ ID NO:20)。预测的保守结构域以粗体标出。

[0035] 图18A-18B:图18A示出了Pf51A的核苷酸序列(SEQ ID NO:21)。图 18B示出了Pf51A的氨基酸序列(SEQ ID NO:22)。预测的信号序列以下划线标出。预测的L-α-阿拉伯呋喃糖苷酶保守结构域以粗体标出。为了在里氏木霉中表达,将预测的Pf51A信号序列替换为里氏木霉CBH1信号序列 (MYRKLAVISAFLATARA(SEQ ID NO:159))并且将所述Pf51A核苷酸序列针对在里氏木霉中的表达作密码子优化。

[0036] 图19A-19B:图19A示出了AfuXyn2的核苷酸序列(SEQ ID NO:23)。图19B示出了AfuXyn2的氨基酸序列(SEQ ID NO:24)。预测的信号序列以下划线标出。预测的GH11保守结构域以粗体标出。

[0037] 图20A-20B:图20A示出了AfuXyn5的核苷酸序列(SEQ ID NO:25)。图20B示出了AfuXyn5的氨基酸序列(SEQ ID NO:26)。预测的信号序列以下划线标出。预测的GH11保守结构域以粗体标出。

[0038] 图21A-21B:图21A示出了Fv43D的核苷酸序列(SEQ ID NO:27)。图 21B示出了Fv43D的氨基酸序列(SEQ ID NO:28)。预测的信号序列以下划线标出。预测的保守结构域以粗体标出。

[0039] 图22A-22B:图22A示出了Pf43B的核苷酸序列(SEQ ID NO:29)。图 22B示出了Pf43B的氨基酸序列(SEQ ID NO:30)。预测的信号序列以下划线标出。预测的保守结构域以粗体标出。

[0040] 图23A-23B:图23A示出了核苷酸序列(SEQ ID NO:31)。图23B示出了Fv51A的氨基酸序列(SEQ ID NO:32)。预测的信号序列以下划线标出。预测的L-α-阿拉伯呋喃糖苷酶保守结构域以粗体标出。

[0041] 图24A-24B:图24A示出了里氏木霉Xyn3的核苷酸序列(SEQ ID NO:41)。图24B示出了里氏木霉Xyn3的氨基酸序列(SEQ ID NO:42)。预测的信号序列以下划线标出。预测的保守结构域以粗体标出。

[0042] 图25A-25B:图25A示出了里氏木霉Xyn2的氨基酸序列(SEQ ID NO:43)。信号序列以下划线标出。预测的保守结构域以粗体标出。图25B 示出了里氏木霉Xyn2的核苷酸序列(SEQ ID NO:162)。该编码序列可见于Törrönen et al.Biotechnology,1992,10:1461-65(Törrönen等人,生物技术,1992年,第10卷,第1461-1465页)。

[0043] 图26A-26B:图26A示出了里氏木霉Bxl1的氨基酸序列(SEQ ID NO:44)。信号序列以下划线标出。预测的保守结构域以粗体标出。图26B 示出了里氏木霉Bxl1的核苷酸序列(SEQ ID NO:163)。该编码序列可见于 Margolles-Clark et al.Appl.Environ.Microbiol.1996,62(10):3840-46 (Margolles-Clark等人,《应用环境微生物学》,1996年,第62卷,第 10期,第3840-3846页)。

[0044] 图27A-27F:图27A示出了里氏木霉Bgl1的氨基酸序列(SEQ ID NO:45)。信号序列以下划线标出。该编码序列可见于Barnett et al.Bio- Technology,1991,9(6):562-567(Barnett等人,《生物技术》,1991年,第 9卷,第6期,第562-567页)。图27B示出了Pa51A的推定cDNA(SEQ ID NO:46)。图27C示出了Pa51A的密码子优化cDNA(SEQ ID NO:47)。图 27D:包含编码成熟Gz43A的基因组DNA上游的CBH1信号序列(以下划线标出)的构建体的编码序列(SEQ ID NO:48)。图27E:包含编码成熟 Fo43A的基因组DNA上游的CBH1信号序列(以下划线标出)的构建体的编码序列(SEQ ID NO:49)。图27F:包含编码Pf51A的密码子优化DNA 上游的CBH1信号序列(以下划线标出)的构建体的编码序列(SEQ ID NO:50)。

[0045] 图28A-28B:图28A示出了里氏木霉Eg4的核苷酸序列(SEQ ID NO:51)。图28B示出了里氏木霉Eg4的氨基酸序列(SEQ ID NO:52)。预测的信号序列以下划线标出。预测保守结构域以粗体标出。预测接头以斜体表示。

[0046] 图29A-29B:图29A示出了Pa3D的核苷酸序列(SEQ ID NO:53)。图 29B示出了Pa3D的氨基酸序列(SEQ ID NO:54)。预测的信号序列以下划线标出。预测保守结构域以粗体标出。

[0047] 图30A-30B:图30A示出了Fv3G的核苷酸序列(SEQ ID NO:55)。图 30B示出了Fv3G的氨基酸序列(SEQ ID NO:56)。预测的信号序列以下划线标出。预测保守结构域以粗体标出。

[0048] 图31A-31B:图31A示出了Fv3D的核苷酸序列(SEQ ID NO:57)。图 31B示出了Fv3D

的氨基酸序列(SEQ ID NO:58)。预测的信号序列以下划线标出。预测保守结构域以粗体标出。

[0049] 图32A-32B:图32A示出了Fv3C的核苷酸序列(SEQ ID NO:59)。图 32B示出了Fv3C的氨基酸序列(SEQ ID NO:60)。预测的信号序列以下划线标出。预测保守结构域以粗体标出。

[0050] 图33A-33B:图33A示出了Tr3A的核苷酸序列(SEQ ID NO:61)。图33B示出了Tr3A的氨基酸序列(SEQ ID NO:62)。预测的信号序列以下划线标出。预测保守结构域以粗体标出。

[0051] 图34A-34B:图34A示出了Tr3B的核苷酸序列(SEQ ID NO:63)。图 34B示出了Tr3B的氨基酸序列(SEQ ID NO:64)。预测的信号序列以下划线标出。预测保守结构域以粗体标出。

[0052] 图35A-35B:图35A示出了Te3A经密码子优化的核苷酸序列(SEQ ID NO:65)。图35B示出了Te3A的氨基酸序列(SEQ ID NO:66)。预测的信号序列以下划线标出。预测保守结构域以粗体标出。

[0053] 图36A-36B:图36A示出了An3A的核苷酸序列(SEQ ID NO:67)。图 36B示出了An3A的氨基酸序列(SEQ ID NO:68)。预测的信号序列以下划线标出。预测保守结构域以粗体标出。

[0054] 图37A-37B:图37A示出了Fo3A的核苷酸序列(SEQ ID NO:69)。图 37B示出了Fo3A的氨基酸序列(SEQ ID NO:70)。预测的信号序列以下划线标出。预测保守结构域以粗体标出。

[0055] 图38A-38B:图38A示出了Gz3A的核苷酸序列(SEQ ID NO:71)。图 38B示出了Gz3A的氨基酸序列(SEQ ID NO:72)。预测的信号序列以下划线标出。预测保守结构域以粗体标出。

[0056] 图39A-39B:图39A示出了Nh3A的核苷酸序列(SEQ ID NO:73)。图 39B示出了Nh3A的氨基酸序列(SEQ ID NO:74)。预测的信号序列以下划线标出。预测保守结构域以粗体标出。

[0057] 图40A-40B:图40A示出了Vd3A的核苷酸序列(SEQ ID NO:75)。图 40B示出了Vd3A的氨基酸序列(SEQ ID NO:76)。预测的信号序列以下划线标出。预测保守结构域以粗体标出。

[0058] 图41A-41B:图41A示出了Pa3G的核苷酸序列(SEQ ID NO:77)。图 41B示出了Pa3G的氨基酸序列(SEQ ID NO:78)。预测的信号序列以下划线标出。预测保守结构域以粗体标出。

[0059] 图42:示出了Tn3B的氨基酸序列(SEQ ID NO:79)。标准的信号预测程序Signal P未提供预测的信号序列。

[0060] 图43A-43B:图43A示出了某些β-葡糖苷酶同源物的氨基酸序列比对结果。图43B示出了β-葡糖苷酶同源物的比对结果,其中已知一些同源物对蛋白水解剪切敏感,而其他则不敏感。第一下划线区域包含大致处于这类酶的位于中央的环序列内的残基。该第一下划线区域下游的第二下划线区域包含通常对于初始蛋白水解消化或剪切敏感的残基。

[0061] 图44:示出了具有Fv3C可读框的pENTR/D-TOPO载体。

[0062] 图45A-45B:图45A示出了pTrex6g载体。图45B示出了pExpression 构建体

pTrex6g/Fv3C。

[0063] 图46A-46C:图46A示出了Fv3C基因组DNA序列的预测编码区。图46B示出了Fv3C的N末端氨基酸序列。箭头示出推定的信号肽切割位点。成熟蛋白质的起始处以下划线标出。图46C示出了从注释的(1)和替代的(2)起始密码子表达Fv3C的里氏木霉转化体的SDS-PAGE凝胶电泳结果。

[0064] 图47:比较多种完整纤维素酶和β-葡糖苷酶混合物在50℃糖化经磷酸溶胀的纤维时的性能。在本实验中,将完整纤维素酶(10mg蛋白/g纤维素)与5mg/g的β-葡糖苷酶掺合并且将酶混合物用来在0.7%纤维素,pH 5.0水解经磷酸溶胀的纤维素。在该图中标记为背景的样品是从单独未添加β-葡糖苷酶的10mg/g完整纤维素酶获得的转化。反应在50℃在微量滴定平板中进行2小时。对所述样品进行三组重复测试。这根据实例5A进行。

[0065] 图48:比较多种完整纤维素酶与β-葡糖苷酶混合物在50℃糖化经酸预处理的玉米秸秆(PCS)时的性能。在本实验中,将完整纤维素酶(10mg 蛋白/g纤维素)与5mg/g的β-葡糖苷酶掺合并且将酶混合物用来以13%固形物,pH 5.0水解PCS。在该图中标记为背景的样品是从单独未添加β-葡糖苷酶的10mg/g完整纤维素酶获得的转化。反应在50℃在微量滴定平板中进行48小时。对所述样品进行三组重复测试。实验细节在实例5B中描述。

[0066] 图49:比较多种完整纤维素酶和β-葡糖苷酶混合物在50℃糖化经稀氨水预处理的玉米芯时的性能。在本实验中,将完整纤维素酶(10mg蛋白/g纤维素)与8mg/g的半纤维素酶和5mg/g的β-葡糖苷酶掺合并且将酶混合物以20%固形物,pH 5.0水解氨水预处理的玉米芯。在该图中标记为背景的样品是从单独未添加β-葡糖苷酶的10mg/g完整纤维素酶+8mg/g半纤维素混合物获得的转化。反应在50℃在微量滴定平板中进行48小时。对所述样品进行三组重复测试。实验细节在实例5C中描述。

[0067] 图50:比较完整纤维素酶和β-葡糖苷酶混合物在50℃糖化经氢氧化钠(NaOH)预处理的玉米芯时的性能。在本实验中,将完整纤维素酶 (10mg蛋白/g纤维素)和5mg/g的β-葡糖苷酶掺合并且将酶混合物用于以17%固形物,pH 5水解经NaOH预处理的玉米芯。在该图中标记为背景的样品从单独未添加β-葡糖苷酶的单独10mg/g完整纤维素酶混合物获得的转化。反应在50℃在微量滴定平板中进行48小时。对每份样品进行4 次重复测试。这根据实例5D进行。

[0068] 图51:比较完整纤维素酶和β-葡糖苷酶混合物在50℃糖化经稀氨水预处理的柳枝稷时的性能。在本实验中,将完整纤维素酶(10mg蛋白/g纤维素)与5mg/g的β-葡糖苷酶掺合并且将酶混合物用来以17%固形物,pH 5.0水解柳枝稷。在该图中标记为背景的样品从单独未添加β-葡糖苷酶的单独10mg/g完整纤维素酶混合物获得的转化。反应在50℃在微量滴定平板中进行48小时。对每份样品进行4次重复测试。实验细节在实例5E中描述。

[0069] 图52:比较完整纤维素酶和β-葡糖苷酶混合物在50℃糖化AFEX玉米秸秆时的性能。在本实验中,将完整纤维素酶(10mg蛋白/g纤维素)与 5mg/g的β-葡糖苷酶掺合并且将酶混合物用来以14%固形物,pH 5.0水解 AFEX玉米秸秆。在该图中标记为背景的样品是从单独未添加β-葡糖苷酶的10mg/g完整纤维素酶混合物获得的转化。反应在50℃在微量滴定平板中进行48小时。对每份样品进行4次重复测试。实验细节在实例5F中描述。

[0070] 图53A-53C:示出了在β-葡糖苷酶与完整纤维素酶(0至50%的量) 的变动比例下,来自经稀氨水预处理的玉米芯(20%固形物)的葡聚糖转化%。所述酶剂量对每个实验保持

恒定。图53A示出了使用里氏木霉Bgl1 进行的实验。图53B示出了使用Fv3C进行的实验。图53C示出了使用黑曲霉Bglu(An3A)进行的实验。

[0071] 图54:示出了根据实例7,借助以2.5-40mg/g葡聚糖水平投予的三种不同酶组合物,来自经稀氨水预处理的玉米芯(20%固形物)的葡聚糖转化%。Δ标示出使用Accellerase 1500+Multifect木聚糖酶(Multifect Xylanase)所观测到的葡聚糖转化率,◇标示出使用来自里氏木霉整合型菌株H3A的完整纤维素酶所观测到的葡聚糖转化率,◆标示出使用使用包含 75重量%的来自里氏木霉整合型菌株H3A的完整纤维素酶加25重量%的 Fv3C的酶组合物所观测到的葡聚糖转化率。

[0072] 图55A-55I:图55A示出了用于黑曲霉中表达的pRAX2-Fv3C表达质粒的图谱。图55B示出了pENTR-TOPO-Bgl1-943/942质粒。图55C示出了pTrex3g 943/942表达载体。图55D示出了pENTR/里氏木霉Xyn3质粒。图55E示出了pTrex3g/里氏木霉Xyn3表达载体。图55F示出了 pENTR-Fv3A质粒。图55G示出了pTrex6g/Fv3A表达载体。图55H示出了TOPO平端/Pegl1-Fv43D质粒。图55I示出了TOPO平端/Pegl1-Fv51A 质粒。

[0073] 图56:示出了里氏木霉β-木糖苷酶Bxl1和Fv3A之间的氨基酸比对结果。

[0074] 图57:示出了某些GH43家族水解酶的氨基酸序列比对结果。在该家族的成员中保守的氨基酸残基以下划线标出并以粗体标出。

[0075] 图58:示出了某些GH51家族酶的氨基酸序列比对结果。在该家族的成员中保守的氨基酸残基以下划线标出并以粗体标出。

[0076] 图59A-59B:示出了大量GH10和GH11家族内切木聚糖酶的氨基酸序列比对结果。图59A:GH10家族木聚糖酶的比对结果。加下划线的以粗体标出的残基是催化亲核残基(在比对上方标记为"N")。图59B: GH11家族木聚糖酶的比对结果。加下划线的以粗体标出的残基是催化亲核残基以及通用的酸碱残基(在比对上方分别标记为"N"和"A")。

[0077] 图60A-60C:图60A示出了编码Fv3C/里氏木霉Bgl3("FB")嵌合/ 融合多肽的基因的示意图。图60B示出了编码融合/嵌合多肽Fv3C/里氏木霉Bgl3("FB")的核苷酸序列(SEQ ID NO:82)。图60C示出了编码融合/ 嵌合多肽Fv3C/里氏木霉Bgl3的氨基酸序列(SEQ ID NO:159)。粗体表示的序列来自里氏木霉Bgl3。

[0078] 图61:示出了pTTT-pyrG13-Fv3C/Bgl3融合质粒的图谱。

[0079] 图62:在糖化经稀氨水预处理的玉米中比较里氏木霉Bgl1(实心菱形)和在黑曲霉中产生的Fv3C(空心菱形)。在本实验中,将里氏木霉 Bgl1和Fv3C(0-10mg蛋白/g纤维素)与恒定水平的10mg/g H3A-5一起加载,并且将这些混合物用于在5%纤维素,pH 5.0水解经稀氨水预处理的玉米芯。反应在50℃在微量滴定平板中进行2天。对每份样品进行5个重复测定。实验细节在实例13中示出。

[0080] 图63:以90℃/r的扫描速度(25℃-110℃)在50mM乙酸钠缓冲液 (pH 5)中采集的β-葡糖苷酶里氏木霉Bglu1(Tr3A)、Fv3C和 Fv3C/Te3A/Bgl3("FAB")嵌合多肽的DSC图谱。

[0081] 图64A-64E:图64A:完整纤维素酶:里氏木霉Bgl3混合物在50℃糖化经磷酸溶胀的纤维素时的性能。图64B:在37℃糖化经磷酸溶胀的纤维素时的里氏木霉Bgl3混合物。图64C:在50℃糖化经酸预处理的玉米秸秆的里氏木霉Bgl3混合物。图64D:在37℃糖化经酸预处理的玉米秸秆的里氏木霉Bgl3混合物。

[0082] 图65A-65B图65A:在糖化经磷酸溶胀的纤维素方面比较里氏木霉 Bgl1(实心菱

形)和里氏木霉Bgl3(空心菱形)。图65B:在糖化经磷酸溶胀的纤维素方面比较里氏木霉Bgl1(左栏)和里氏木霉Bgl3(右栏)产生的纤维二糖(黑色柱形)和葡萄糖(白色柱形)。

[0083] 图66:示出了多种引物的核苷酸序列。

[0084] 图67A-67B:图67A示出了Fv3C/Te3A/里氏木霉Bgl3("FAB")的全长氨基酸序列(SEQ ID NO:135)(Te3A为粗体斜体大写字母,里氏木霉 Bgl3为下划线标出的大写字母)。图67B示出了编码Fv3C/Te3A/里氏木霉Bgl3("FAB")嵌合体的核酸序列(SEQ ID NO:83)。

[0085] 图68A-68C:图68A是列出在某些嵌合β-葡糖苷酶多肽的N末端和C 末端结构域中存在的结构基序的表。图68B是列出用于设计本发明的适当β-葡糖苷酶多肽杂合体/嵌合体的某些氨基酸序列基序的表。图68C是 GH61/内切葡聚糖酶的氨基酸序列基序的表。

[0086] 图69:示出了Pa3C的核苷酸和蛋白质序列(分别为SEQ ID NOs:80 和81)。

[0087] 图70A-G:图70A示出了Fv3C和Te3A以及里氏木霉Bgl1的3-D叠合结构,所述结构以第一角度观察,使得"插入物1"的结构可见。图 70B示出了相同的叠合结构,所述结构以第二角度观察,使得"插入物 2"的结构可见。图70C示出了相同的叠合结构,所述结构以第三角度观察,使得"插入物3"的结构可见。图70D示出了相同的叠合结构,所述结构以第四角度观察,使得"插入物4"的结构可见。FIG 70E是里氏木霉 Bgl1(Q12715_TRI)、Te3A(ABG2_T_eme)和Fv3C(FV3C)的序列比对结果,以插入物1-4标出,其全部为环状结构。图70F示出了Fv3C(浅灰色)、Te3A(深灰色)和里氏木霉Bgl1(黑色)的结构的叠合部分,显示残基W59/W33与W355/W325(Fv3C/Te3A)之间保守的相互作用。图70G 示出了Fv3C(浅灰色)、Te3A(深灰色)和里氏木霉Bgl1(黑色)的结构的叠合部分,显示第一对残基S57/31和N291/261(Fv3C/Te3A)以及第二组残基Y55/29、P775/729和A778/732(Fv3C/Te3A)之间保守的相互作用。图70H示出了Fv3C(深灰色)和里氏木霉Bgl1(黑色)的结构的叠合部分,显示了Fv3C在K162处与"插入物2"中的V409的主链氧原子的氢键相互作用,该相互作用在Te3A中是保守的,但是不存在于里氏木霉 Bgl1中。图70I(a)-(b)示出了SEQ ID NO:168内部的保守的糖基化位点,其在Fv3C、Te3A和SEQ ID NO:135的嵌合/杂合β-葡糖苷酶之间共有,(a) 示出了与Te3A(深灰色)和里氏木霉Bgl1(黑色)叠合的相同区域;(b) 示出了与SEQ ID NO:135的嵌合/杂合β-葡糖苷酶(浅灰色)、Te3A(深灰色)和里氏木霉Bgl1(黑色)叠合的相同区域。黑色箭头显示了Te3A 中"插入物3"的环状结构(也存在于SEQ ID NO:135的杂合β-葡糖苷酶中),其似乎显示包埋糖基化聚糖。图70J示出了Fv3C(浅灰色)、 Te3A(深灰色)和里氏木霉Bgl1(黑色)结构的叠合部分,表明残基 W386/355之间的保守相互作用与Fv3C和Te3A的"插入物2"的W95/68 (Fv3C/Te3A)相互作用。这种相互作用从里氏木霉Bgl1中缺失。

[0088] 图71A-71C:图71A:示出了根据实例13,在50℃孵育44小时后可溶级分(上清液)中测定的未结合蛋白的量。图71B:示出了根据实例13, 在50℃孵育44小时后浆液中(结合和未结合)的总蛋白。图71C:示出了根据实例13在缓冲液中额外孵育30分钟后浆液中未结合的蛋白质。

具体实施方式

[0089] 酶传统上借助底物特异性和反应产物分类。在基因组前时代,将功能视为用于比较酶的最可操作(并且可能最有用)的基础,并且用于多种酶活性的测定法已经充分形成多年,产生熟知的EC分类方案。根据这种分类方案,作用于两个碳水化合物部分(或碳水化合

物和非碳水化合物部分—如存在于硝基酚-配醣衍生物中)之间糖苷键的纤维素酶和其他糖基水解酶命名为EC 3.2.1.-,最后的数字表明所切割键的确切类型。例如,根据这种方案,具有内切作用的纤维素酶(1,4-β-内切葡聚糖酶)命名为EC 3.2.1.4。

[0090] 随着广泛的基因组测序计划进展,测序数据已经辅助分析和比较对相关的基因和蛋白质。另外,已将日益增多的能够作用于碳水化合物部分的酶(即糖酶)进行结晶并且解析了其3D结构。这类分析已经鉴定了具有相关序列的酶的离散家族,所述酶包含可以根据它们的氨基酸序列被预测的保守三维折叠。此外,已经显示具有相同或相似三维折叠的酶表现出相同或相似的水解立体特异性,甚至在催化不同反应时也是如此(Henrissat et al.,FEBS Lett 1998,425(2):352-4(Henrissat等人,《欧洲生化学会联盟通讯》,1998年,第425卷,第2期,第352-354页);Coutinho and Henrissat,Genetics,biochemistry and ecology of cellulose degradation,1999,T. Kimura.Tokyo,Uni Publishers Co:15-23(Coutinho和Henrissat,《纤维素降解的遗传学、生物化学和生态学》,1999年,T.Kimura编辑,东京, Uni出版社,第15-23页))。

[0091] 这些研究结果形成了基于序列的糖酶模块分类的基础,所述分类以互联网数据库形式在www.cazy.org可获得:碳水化合物-活性酶服务器(CAZy) ()(参见Cantarel et al.,2009,The Carbohydrate-Active EnZymes database (CAZy):an expert resource for Glycogenomics.Nucleic Acids Res.37(Database issue):D233-38(Cantarel等人,2009年,碳水化合物-活性酶数据库 (CAZy):糖基因组学的专业资源,《核酸研究》,37(数据库专辑): D233-38))。

[0092] CAZy定义了四个主要类别的通过所催化的反应类型可区别的糖酶:糖基水解酶(GH)、糖基转移酶(GT)、多糖裂解酶(PL)和糖酯酶(CE)。本发明的酶是糖基水解酶。GH是一组水解两个或更多个碳水化合物之间的糖苷键或碳水化合物与非碳水化合物部分之间的糖苷键的酶。通过序列相似性来分组的糖基水解酶分类系统已经导致对超过120个不同家族的定义。这种分类在CAZy网站可获得。本发明的酶属于糖基水解酶家族3 (GH3)。

[0093] GH3酶包括,例如,β-葡糖苷酶(EC:3.2.1.21);β-木糖苷酶 (EC:3.2.1.37);N-乙酰基β-氨基葡糖苷酶(EC:3.2.1.52);葡聚糖β-1,3-葡糖苷酶(EC:3.2.1.58);纤维糊精酶(EC:3.2.1.74);外切-1,3-1,4-葡聚糖酶 (EC:3.2.1);以及β-半乳醣苷酶(EC 3.2.1.23)。例如,GH3酶可以是具有β- 葡糖苷酶、β-木糖苷酶、N-乙酰基β-氨基葡糖苷酶、葡聚糖β-1,3-葡糖苷酶、纤维糊精酶、外切-1,3-1,4-葡聚糖酶和/或β-半乳醣苷酶活性的那些酶。一般来讲,GH3酶是球状蛋白并且可以由两个或更多的亚结构域组成。已经鉴定一个催化残基为天冬氨酸残基,其在β-葡糖苷酶中位于该肽的N末端第三位并且处于SDW氨基酸片段中(Li et al.2001,Biochem.J. 355:835-840(Li等人,2001年,《生化杂志》,第355卷,第835-840页))。在来自里氏木霉的Bgl1中的对应序列是T266D267W268(从起始位置处的甲硫氨酸计数),所述催化残基天冬氨酸为D267。所述羟基/天冬氨酸序列在测试的GH3β-木糖苷酶中也是保守的。例如,里氏木霉Bxl1 中的对应序列是S310D311,而在Fv3A中的对应序列是S290D291。

[0094] 本发明的多肽

[0095] 纤维素酶

[0096] 本发明的组合物可以包含一种或多种纤维素酶。纤维素酶是水解纤维素(β-1,4-

葡聚糖或者βD-葡糖苷键)而导致葡萄糖、纤维二糖、纤维寡糖等形成的酶类。传统上将纤维素酶分为三大类:内切葡聚糖酶(EC 3.2.1.4)(“EG”)、外切葡聚糖酶或纤维二糖水解酶(EC 3.2.1.91)(“CBH”) 和β-葡糖苷酶(β-D-葡萄糖苷葡萄糖水解酶;EC 3.2.1.21)(“BG”)(Knowles et al.,1987,Trends in Biotechnology 5(9):255-261(Knowles等人, 1987年,《生物技术动向》,第5卷,第9期,第255-261页);Shulein, 1988,Methods in Enzymology,160:234-242(Shulein,1988年,《酶学方法》,第160卷,第234-242页))。

[0097] 用于本发明方法和组合物的纤维素酶可以获自或以重组方式产生自下列生物的一种或多种(不限于此):勒克瑙金孢菌(Chrysosporium lucknowense)、柄毛皮伞(Crinipellis scapella)、菜豆壳球孢(Macrophomina phaseolina)、嗜热毁丝菌(Myceliophthora thermophila)、粪生粪壳菌 (Sordaria fimicola)、拟刺盘孢周刺座霉(Volutella colletotrichoides)、太瑞斯梭孢壳霉(Thielavia terrestris)、支顶孢属(Acremonium sp.)、黑耳(Exidia glandulosa)、木蹄层孔菌(Fomes fomentarius)、绵皮孔菌属(Spongipellis sp.)、红根囊壶菌(Rhizophlyctis rosea)、微小根毛霉(Rhizomucor pusillus)、闪光须霉(Phycomyces niteus)、弗雷生刺枝霉(Chaetostylum fresenii)、棉色二孢(Diplodia gossypina)、异色尾孢藻(Ulospora bilgramii)、集孢粪盘菌 (Saccobolus dilutellus)、疣孢青霉(Penicillium verruculosum)、产黄青霉(Penicillium chrysogenum)、疣状顶孢霉(Thermomyces verrucosus)、合生腐皮壳菌(Diaporthe syngenesia)、黄瓜炭疽菌(Colletotrichum lagenarium)、黑孢霉属(Nigrospora sp.)、团炭角菌(Xylaria hypoxylon)、松色丛赤壳菌 (Nectria pinea)、大孢粪壳菌(Sordaria macrospora)、嗜热梭孢壳霉 (Thielavia thermophila)、突孢毛壳(Chaetomium mororum)、绿毛壳(Chaetomium virscens)、巴西毛壳(Chaetomium brasiliensis)、链丝毛壳菌 (Chaetomium cunicolorum)、Syspastospora boninensis、多生枝鼻菌 (Cladorrhinum foecundissimum)、嗜热柱霉(Scytalidium thermophila)、链孢粘帚菌(Gliocladium catenulatum)、尖孢镰孢菌番茄亚种(Fusarium oxysporum ssp.lycopersici)、尖孢镰孢菌西番莲亚种(Fusarium oxysporum ssp.passiflora)、镶素对镰孢菌(Fusarium solani)、蛇形镰孢菌(Fusarium anguioides)、梨孢镰孢菌(Fusarium poae)、黑腐质霉(Humicola nigrescens)、灰腐质霉(Humicola grisea)、网纹斑褶菌(Panaeolus retirugis)、血红栓菌(Trametes sanguinea)、裂褶菌(Schizophyllum commune)、粉红单端孢(Trichothecium roseum)、小球壳孢菌(Microsphaeropsis sp.)、粪盘菌(Acsobolus stictoideus spej.)、点孔座壳 (Poronia punctata)、多节孢属(Nodulisporum sp.)、木霉属(Trichoderma sp.) (例如,里氏木霉)以及柱孢菌属(Cylindrocarpon sp)。纤维素酶还可以获自或按重组方式产自细菌,或可以按重组方式产自酵母。

[0098] 例如,用于本发明的方法和/或组合物的纤维素酶是完整纤维素酶并且 /或者如卡尔科弗卢尔测定法所测定的能够实现至少0.1(例如0.1至0.4) 分率产物。

[0099] β-葡糖苷酶

[0100] β-葡糖苷酶(或在本文中可互换地为“β-葡糖苷酶多肽”)催化β-D- 葡糖苷中的末端非还原性残基的水解,释放出葡萄糖。β-葡糖苷酶多肽的例子包括具有至少一种β-葡糖苷酶多肽活性的多肽、多肽片段、肽和融合多肽。β-葡糖苷酶多肽和核酸的例子包括来自本

文所述的任何来源生物的天然存在的多肽(包括,例如变体)和核酸,以及来自本文所述的任何来源生物的突变多肽和核酸,其具有至少一种β-葡糖苷酶多肽的活性。

[0101] 本发明的组合物可以包含一种或多种β-葡糖苷酶多肽。如本文所使用,术语"β-葡糖苷酶"是指β-D-葡糖苷葡糖水解酶(归类为EC 3.2.1.21)和/或GH家族3的成员,其催化纤维二糖水解以释放β-D-葡萄糖。本发明的GH3β-葡糖苷酶包括但不限于,Fv3C、Pa3D、Fv3G、Fv3D、Tr3A(亦称为"里氏木霉Bgl1"或"里氏木霉Bglu1")、Tr3B (亦称为"里氏木霉Bgl3")、Te3A、An3A(亦称为"黑曲霉Bglu")、Fo3A、Gz3A、Nh3A、Vd3A、Pa3G或Tn3B多肽。在一些实施例中,本文的GH3β-葡糖苷酶多肽具有β-葡糖苷酶多肽的至少一种活性。

[0102] 合适的β-葡糖苷酶多肽可以从多种微生物通过重组手段获得或从商业来源购买。来自微生物的β-葡糖苷酶的例子包括但不限于来自细菌和真菌的酶。例如,本发明的β-葡糖苷酶适当地获自丝状真菌。

[0103] β-葡糖苷酶多肽可以尤其获自或按重组方式产自棘孢曲霉(A.aculeatus)(Kawaguchi et al.Gene 1996,173:287-288(Kawaguchi等人,《基因》, 1996年,第173卷,第287-288页))、A.kawachi(Iwashita et al.Appl. Environ.Microbiol.1999,65:5546-5553(Iwashita等人,《应用与环境微生物学》,1999年,第65卷,第5546-5553页))、米曲霉(A.oryzae) (WO 2002/095014)、双氮纤维单胞菌(C.biazotea)(Wong et al.Gene,1998,207:79-86(Wong等人,《基因》,1998年,第207卷,第79-86 页))、绳状青霉(P.funiculosum)(WO 2004/078919)、扣囊复膜孢酵母(S. fibuligera)(Machida et al.Appl.Environ.Microbiol.1988,54:3147-3155 (Machida等人,《应用与环境微生物学》,1988年,第54卷,第3147- 3155页))、粟酒裂殖酵母(S.pombe)(Wood et al.Nature 2002,415:871- 880(Wood等人,《自然》,2002年,第415卷,第871-880页))、里氏木霉(例如β-葡糖苷酶1(美国专利No.6,022,725)、β-葡糖苷酶3(美国专利No.6,982,159)、β-葡糖苷酶4(美国专利No.7,045,332)、β-葡糖苷酶5(美国专利No.7,005,289)、β-葡糖苷酶6(美国专利公布No. 20060258554)、β-葡糖苷酶7(美国专利公布No.20060258554))、鹅掌柄孢霉(P.anserina)(例如Pa3D)、轮枝镰刀菌(F.verticillioides)(例如 Fv3G、Fv3D或Fv3C)、里氏木霉(例如Tr3A或Tr3B)、埃默森篮状菌 (T.emersonii)(例如Te3A)、黑曲霉(例如An3A)、尖孢镰刀菌(F. oxysporum)(例如Fo3A)、玉米赤霉(G.zeae)(例如Gz3A)、鞭毛藻丛赤壳菌(N.haematococca)(例如Nh3A)、大丽轮枝菌(V.dahliae)(例如 Vd3A)、鹅掌柄孢霉(例如Pa3G)或新阿波罗栖热袍菌(例如Tn3B)。

[0104] 该β-葡糖苷酶多肽可以通过量表达编码β-葡糖苷酶、变体、杂合体/ 嵌合体/融合体或突变体的内源/外源基因来产生。例如,β-葡糖苷酶多肽可以分泌进入胞外间隙,例如,由革兰氏阳性菌生物(诸如芽孢杆菌属 (Bacillus)或放线菌属(Actinomycetes))或由真核宿主(诸如真菌(例如木霉属(Trichoderma)、金孢属(Chrysosporium、曲霉属(Aspergillus)、酵母菌属 (Saccharomyces)、毕赤酵母属(Pichia))分泌。β-葡糖苷酶多肽可以在酵母如酿酒酵母(Saccharomyces cerevisiae)中表达。该β-葡糖苷酶多肽可能过量表达或不足表达。

[0105] 该β-葡糖苷酶多肽还可以获自商业来源。适用于本发明的市售β-葡萄糖苷酶制品的例子包括,例如,以Accellerase®BG(丹尼斯科美国有限公司,杰能科(Danisco US Inc.,Genencor))商购获得的里氏木霉β-葡糖苷酶;NOVOZYM™188(来自黑曲霉(A.niger)的β-葡

糖苷酶);农杆菌属 (Agrobacterium sp.)的β-葡糖苷酶以及海栖热袍菌(T.maritima)的β-葡糖苷酶,其来自麦格酶(爱尔兰的麦格酶国际爱尔兰有限公司(Megazyme International Ireland Ltd.,Ireland))。

[0106] 此外,该β-葡糖苷酶多肽可以是纤维素酶组合物、完整细胞纤维素酶组合物、纤维素酶发酵液或全发酵液制剂纤维素酶组合物的成分。

[0107] β-葡糖苷酶活性可以通过本领域中已知的多种合适方法测定,所述方法以非限定性例子的形式包括Chen et al.,Biochimica et Biophysica Acta 1992,121:54-60(Chen等人,《生物化学与生物物理学报》,1992年,第 121卷,第54-60页)中描述的测定法,其中1pNPG代表在50℃和pH 4.8 在10分钟内从4-硝基苯基-β-D-吡喃葡萄糖苷中释放1μmoL硝基酚。

[0108] β-葡糖苷酶多肽在适当情况下构成本发明的纤维素酶组合物中酶总重量的约0重量%至约75重量%。任何酶对彼此的比例可以根据本发明轻易地计算。设想了以从本文公开的重量百分比可导出的任何重量比的包含酶的纤维素酶组合物。β-葡糖苷酶含量可以处于这样一个范围内,其中下限为所述纤维素酶组合物中的酶总重量的约0重量%、1重量%、2重量%、3 重量%、4重量%、5重量%、6重量%、7重量%、8重量%、9重量%、10 重量%、12重量%、15重量%、17%、20重量%、25重量%、30重量%、 40重量%、45重量%或50重量%并且上限为所述纤维素酶组合物中的酶总重量的约10重量%、12重量%、15重量%、17重量%、20重量%、25重量%、30重量%、35重量%、40重量%、50重量%、55重量%、60重量%、65重量%或70重量%。例如,该β-葡糖苷酶合适地为所述纤维素酶组合物中的酶总重量的约0.1重量%至约40重量%、约1重量%至约35重量%、约2重量%至约30重量%;约5重量%至约25重量%、约7重量%至约20重量%、约9重量%至约17重量%、约10重量%至约20重量%;或约5重量%至约10重量%。

[0109] 突变β-葡糖苷酶多肽:本发明提供突变β-葡糖苷酶多肽。突变β-葡糖苷酶多肽包括这些酶,所述酶中的一个或多个氨基酸残基已经历氨基酸置换,同时维持β-葡糖苷酶活性(即,催化β-D-葡糖苷中的末端非还原性残基水解而释放葡萄糖的能力)。因此,突变β-葡糖苷酶多肽构成特定类型的“β-葡糖苷酶多肽”,如本文定义这个术语。突变β-葡糖苷酶多肽可以通过将一个或多个氨基酸置换到所述多肽的天然或野生型氨基酸序列中来制备。在某些方面,本发明包括多肽,其与前体酶氨基酸序列相比包含改变的氨基酸序列,其中突变酶维持前体酶的特征性纤维素水解性质,但是与前体酶相比较可能在一些特定方面具有改变的特性,例如,增加或降低的pH最佳值,增加或降低的氧化稳定性;增加或降低的热稳定性以及对一种或多种底物的增加或降低的比活性水平。可以使用本领域中已知的计算机程序(例如,LASERGENE软件(DNASTAR)),找到确定可以置换、插入或缺失哪些氨基酸残基而不影响生物活性的指导。氨基酸置换可以是保守或非保守的,并且这类置换的氨基酸残基可以是或可以不是由遗传密码编码的氨基酸残基。氨基酸置换可以位于多肽的碳水化合物结合模块 (CBM)内、位于多肽的催化结构域(CD)内和/或同时位于CBM和CD内。标准的二十种氨基酸的“字母表”已经根据其侧链的相似性而划分成为化学家族。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)、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、苯基丙氨酸、甲硫氨酸、色氨酸)、具有β分支侧链的氨基酸(例如,苏氨酸、缬氨酸、异亮氨酸) 以及具有芳香侧链的氨基酸(例如,酪氨酸、苯基丙氨酸、色氨酸、组氨酸)。“保守氨基酸置换”是其中氨基酸残基由具有化学上相似的侧链的氨基酸残基替换的置换(即,将具有碱性侧链的氨基酸替换为具有碱性侧链的另一种氨基酸)。“非保守氨基酸置换”是其中氨基酸残基由具有化学上不同的侧链的氨基酸残基替换的置换(即,将具有碱性侧链的氨基酸替换为具有芳香侧链的另一种氨基酸)。

[0110] 嵌合多肽:本发明还提供杂合/融合/嵌合蛋白质,其包含与一种或多种融合片段连接的本发明蛋白质的结构域,所述融合片段通常对于所述蛋白质是异源的(即,衍生自与本发明蛋白质不同的来源)。这些杂合/融合 /嵌合酶还可认为是一种类型的突变β-葡糖苷酶,这是因为它们在序列上与野生型参照β-葡糖苷酶不同,尽管具有与天然或野生型参照β-葡糖苷酶不同的其他特性,但是维持β-葡糖苷酶活性。合适的嵌合片段包括但不限于,可以增强蛋白质稳定性、提供其他所需生物活性或增强水平的所需生物活性,和/或有助于纯化该蛋白质(例如,通过亲和层析法)的片段。合适的嵌合片段可以是任何大小的结构域,其具有所需的功能(例如,赋予改善的稳定性、溶解度、作用或生物活性;和/或简化蛋白质纯化)。本发明的嵌合蛋白可以从两个或更多个嵌合片段构建,所述嵌合片段每一者或者至少两者衍生自不同来源或微生物。嵌合片段可以连接到本发明蛋白质的结构域的氨基末端和/或羧基末端。嵌合片段可以对切割敏感。具有这种敏感性可能存在优点,例如,这可能实现对目的蛋白的直接回收。嵌合蛋白质优选地通过培养重组细胞来产生,其中所述重组细胞用编码蛋白质的嵌合核酸转染,所述蛋白质包括与蛋白质或其结构域的羧基末端或氨基末端之一连接的嵌合片段或同时与所述羧基默末端和氨基末端连接的嵌合片段。

[0111] 因此,本发明的β-葡糖苷酶多肽还包括诱变基因(例如,具有增强基因转录和翻译的密码子修饰的基因)和截短基因(例如移除信号序列或代之以异源信号序列的基因)的基因融合物的表达产物(例如过量表达的、可溶并且活性形式的重组蛋白)。

[0112] 利用不溶底物的糖基水解酶通常是模块化酶。它们一般包含与一个或多个非催化性碳水化合物结合模块(CBM)附接的催化模块。实际上,认为 CBM促进糖基水解酶与其靶底物多糖的相互作用。因此,本发明提供具有改变的底物特异性的嵌合酶;包括,例如,具有多种底物的嵌合酶,这是由于“剪接进入”的异源CBM所致。本发明嵌合酶的异源CBM还可以设计为模块化的,从而它们与催化模块或催化结构域(“CD”,例如,在活性部位处)附接,所述催化模块或催化结构域可类似地相对于所述糖基水解酶是异源或同源的。

[0113] 因此,本发明提供肽和多肽,其由CBM/CD模块组成或包含 CBM/CD模块,所述CBM/CD模块可以同源地配对或接合以形成嵌合(异源)CBM/CD对。因此,这些嵌合多肽/肽可以用来改进或改变目的酶的性能。因此,在某些方面,本发明提供嵌合酶,其包含例如,SEQ ID NO:2、4、6、8、10、12、14、16、18、20、22、24、26、28、30、32、 34、36、38、40、42、43、44、52、54、56、58、60、62、64、66、68、 70、72、74、76、78或79的酶的至少一种CBM(如果可获得的话)。本发明的多肽,例如,包含氨基酸序列,所述氨基酸序列含有SEQ ID NO:2、4、6、8、10、12、14、16、18、20、22、24、26、28、30、32、 34、36、38、40、42、43、44、52、54、56、58、60、62、64、66、68、70、72、74、76、78或79的多肽序列的CD和/或CBM。本发明的多肽由此可以适当地为融合蛋白,其包含来自两种或更多种不同蛋白质的功能结构域(例如,来自一种蛋白质的CBM与来自另一种蛋白质的CD连接)。

[0114] 本发明还提供非天然存在的纤维素酶组合物，所述纤维素酶组合物包含β-葡糖苷酶多肽，其为至少两种β-葡糖苷酶序列的嵌合体。在某些方面，非天然存在的纤维素酶组合物包含β-葡糖苷酶活性。该组合物还可以包含一种或多种木聚糖酶、β-木糖苷酶和/或L-α-阿拉伯呋喃糖苷酶活性。因此，该组合物是半纤维素酶组合物。在某些方面，非天然存在的纤维素酶/半纤维素酶组合物包含衍生自至少两种不同来源的酶成分或多肽。在某些方面，非天然存在的纤维素酶/半纤维素酶组合物包含一种或多种天然存在的半纤维素酶。

[0115] 在某些方面，该组合物中的β-葡糖苷酶多肽还包含一个或多个糖基化位点。在某些方面，β-葡糖苷酶多肽包含N末端序列和C末端序列，其中 N末端序列或C末端序列中的每一个可以包含衍生自不同β-葡糖苷酶的一个或多个子序列。在某些方面，N末端和C末端序列衍生自不同来源。在一些实施例中，N末端和C末端序列的一个或多个子序列中的至少两个衍生自不同来源。在某些方面，N末端序列或C末端序列还包含长度为约 3、4、5、6、7、8、9、10或11个氨基酸残基的环区序列。在某些实施例中，N末端序列和C末端序列是紧邻的或直接连接的。在其他实施例中， N末端和C末端序列并非紧邻，而是通过接头结构域功能性地连接。接头结构域可以位于该嵌合多肽的中央(例如，并非位于N末端或C末端)。在某些实施例中，杂合多肽的N末端序列或C末端序列均不包含环序列。相反，接头结构域包含环序列。在某些方面，N末端序列包含β-葡糖苷酶或其变体的第一氨基酸序列，其长度为至少约200(例如，约200、250、 300、350、400、450、500、550或600)个残基。在某些方面，N末端序列包含SEQ ID NOs:136-148所代表的多肽序列基序的一种或多种或全部。在某些方面，C末端序列包含β-葡糖苷酶或其变体的第二氨基酸序列，其长度为至少约50(例如，约50、75、100、125、150、175或200)个氨基酸残基。在某些方面，C末端序列包含SEQ ID NOs:149-156所代表的多肽序列基序的一种或多种或全部。特别地，所述两种或更多种β-葡糖苷酶序列的第一种是具有至少约200个氨基酸残基长度并且包含SEQ ID NOs:164-169的氨基酸序列基序中的至少2种(例如，至少2、3、4或全部)的序列，并且所述两种或更多种β-葡糖苷酶序列的第二种具有至少50 个氨基酸残基长度并且包含SEQ ID NO:170。在某些方面，C末端或N末端序列包含环序列，所述环序列包含约3、4、5、6、7、8、9、10或11个氨基酸残基，以及FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。在某些方面，C末端或N末端序列均不包含环序列。在一些实施例中，C末端序列与N末端序列通过接头结构域连接，所述接头结构域包含环序列，所述序列包含约3、4、5、6、7、8、9、10或 11个氨基酸残基，以及FDRRSPG的序列(SEQ ID NO:171)或 FD(R/K)YNIT的序列(SEQ ID NO:172)。在某些方面，非天然存在的纤维素酶或半纤维素酶组合物中的β-葡糖苷酶多肽具有优于衍生所述嵌合多肽的每个C末端和/或N末端序列的任何天然酶的改善稳定性。在某些方面，改善的稳定性包括在储存、表达或生产过程期间蛋白水解稳定性的改善。在某些方面，改善的稳定性包括在储存或生产条件下的酶活性损失的速度或程度的相关降低，其中酶活性损失优选地低于约50％，低于约40％，低于约30％或低于约20％，更优选地低于15％，或低于10％。

[0116] 本发明的多肽可以适合地以“基本纯净”的形式获得和/或使用。例如，本发明的多肽在给定的组合物中构成至少约80重量％(例如至少约 85重量％、90重量％、91重量％、92重量％、93重量％、94重量％、95重量％、96重量％、97重量％、98重量％或99重量％)的总蛋白，所述组合物还包括其他成分，诸如缓冲剂或溶液。

[0117] 发酵液：另外，本发明的多肽可以合适地获自和/或用于发酵液(例如，丝状真菌发

酵液)。发酵液可以是工程化的酶组合物,例如,该发酵液可以通过经工程化以表达目的异源多肽的重组宿主细胞产生,或通过经工程化以比内源表达水平更高或更低的量(例如,以比内源表达水平高或低约1、2、3、4、5倍或更多倍的量)表达本发明内源多肽的重组宿主细胞产生。本发明的发酵液还可以通过某些“整合型”宿主细胞株来产生,所述宿主细胞株经工程化而以所需的比例表达本发明的多种多肽。例如,编码目的多肽的一种或多种或全部基因可以整合进入所述宿主细胞株的遗传物质。

[0118] Fv3C

[0119] Fv3C的氨基酸序列(SEQ ID NO:60)在图32B和43中示出。SEQ ID NO:60是未成熟Fv3C的序列。Fv3C具有与SEQ ID NO:60的第1位至第 19位置(以下划线标出)相对应的预测的信号序列;对该信号序列的切割预测产生具有与SEQ ID NO:60的第20至899位置相对应的序列的成熟蛋白。使用SignalP-NN算法进行信号序列预测。预测的保守结构域在图32B中以粗体标出。结构域预测基于Pfam、SMART或NCBI数据库进行。Fv3C的残基E536和D307据预测分别作为催化性酸-碱和亲核物质发挥功能,这是基于上文所述的来自以下的GH3葡糖苷酶的序列比对:例如鹅掌柄孢霉(登录号XP_001912683)、大丽轮枝菌、鞭毛藻丛赤壳菌(登录号XP_003045443)、玉米赤霉(登录号XP_386781)、尖孢镰刀菌(登录号BGL FOXG_02349)、黑曲霉(登录号CAK48740)、埃默森篮状菌 (登录号AAL69548)、里氏木霉(登录号AAP57755)、里氏木霉(登录号AAA18473)、轮枝镰刀菌和新阿波罗栖热袍菌(登录号Q0GC07)等等(参见图43)。如本文所用,“Fv3C多肽”在某些方面是指包含下述序列的多肽和/或其变体,所述序列包含与SEQ ID NO:60的第20位至第 899位残基中的至少50、75、100、125、150、175、200、250、300、 350、400、450、500、550、600、650、700、750或800个连续氨基酸残基具有至少85%、86%、87%、88%、89%、90%、91%、92%、93%、 94%、95%、96%、97%、98%、99%或100%的序列同一性。与天然Fv3C 相比,Fv3C多肽优选地在残基E536和D307处未改变。Fv3C多肽优选地在本文所述的GH3家族β-葡糖苷酶之间保守的至少70%、80%、90%、95%、98%或99%的氨基酸残基处未改变,如图43的比对结果中示出。Fv3C多肽合适地包含图32B中示出的天然Fv3C的整个预测保守结构域。示例性的Fv3C多肽包含与图32B中示出的成熟Fv3C序列具有至少 85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、 96%、97%、98%、99%或100%同一性的序列。本发明的Fv3C多肽优选地具有β-葡糖苷酶活性。

[0120] 因此,本发明的Fv3C多肽合适地包含与SEQ ID NO:60的氨基酸序列或与SEQ ID NO:60第(i)20-327、(ii)22-600、(iii)20-899、(iv)428-899或 (v)428-660位残基具有至少80%、85%、90%、91%、92%、93%、94%、 95%、96%、97%、98%、99%或100%序列同一性的氨基酸序列。该多肽合适地具有β-葡糖苷酶活性。

[0121] 在某些方面,本发明的“Fv3C多肽”可以指突变Fv3C多肽。可以向 Fv3C多肽引入氨基酸置换以改善该分子的β-葡糖苷酶活性和/或稳定性。例如,可以向Fv3C多肽中引入氨基酸置换,所述氨基酸置换增强Fv3C多肽对其底物的结合亲和力或改善Fv3C催化β-D-葡糖苷中的末端非还原性残基水解的能力。在某些方面,突变Fv3C多肽包含一种或多种保守氨基酸置换。在某些方面,突变Fv3C多肽包含一种或多种非保守的氨基酸置换。在某些方面,一种或多种氨基酸置换位于Fv3C多肽的CD中。或者一种或多种氨基酸置换位于Fv3C多肽的CBM中。一种或多种氨基酸置换可以同时位于CD和CBM中。在某些方面,Fv3C多肽氨基酸置换

可以在氨基酸E536和/或D307处发生。在某些方面，Fv3C多肽氨基酸置换可以在氨基酸D119、R125、L168、R183、K216、H217、R227、M272、Y275、D307、W308、S477和/或E536的一个或多个或全部处发生。突变 Fv3C多肽合适地具有β-葡糖苷酶活性。

[0122] 在某些方面，Fv3C多肽包括嵌合体/融合体/杂合体，或者两种β-葡糖苷酶序列的嵌合构建体，其中第一序列衍生自第一β-葡糖苷酶，具有至少约200个氨基酸残基长度并且包含与Fv3C的等长度序列(SEQ ID NO:60) 约60％、65％、70％、75％、80％或更高的同一性，并且其中第二序列衍生自第二β-葡糖苷酶，长度为至少约50个氨基酸残基，并且包含与SEQ ID NOs:54、56、58、62、64、66、68、70、72、74、76、78和79中任何一种的等长度序列约60％、65％、70％、75％、80％或更高的同一性，或者包含SEQ ID NOs:170的氨基酸序列基序。在某些方面，第一β-葡糖苷酶序列包含SEQ ID NO:60的至少约200个连续氨基酸残基的N末端序列，并且第二β-葡糖苷酶序列包含SEQ ID NOs:54、56、58、62、64、66、68、 70、72、74、76、78和79中任何一种的至少约50个连续氨基酸残基的C 末端序列，或者包含SEQ ID NO:170的氨基酸序列基序。

[0123] 在某些方面，Fv3C多肽可以是嵌合体/杂合体/融合体或两种β-葡糖苷酶序列的嵌合构建体，其中第一序列衍生自第一β-葡糖苷酶，具有至少约 200个氨基酸残基长度并且包含与SEQ ID NOs:54、56、58、62、64、66、 68、70、72、74、76、78和79中任何一种的等长度序列约60％、65％、 70％、75％、80％或更高的同一性，或包含SEQ ID NOs:164-169的氨基酸序列基序的一种或多种或全部，其中第二序列衍生自第二β-葡糖苷酶，长度为至少约50个氨基酸残基，并且包含与Fv3C的等长度序列(SEQ ID NO: 60)约60％、65％、70％、75％、80％或更高的同一性。在某些方面，第一β-葡糖苷酶序列包含SEQ ID NOs:54、56、58、62、64、66、68、70、 72、74、76、78或79的至少200个连续氨基酸残基的N末端序列，或者包含SEQ ID NOs:164-169的氨基酸序列基序的一种或多种或全部，并且第二β-葡糖苷酶序列包含SEQ ID NO:60的至少约50个连续氨基酸残基的 C末端序列。

[0124] 在某些方面，第一β-葡糖苷酶序列位于嵌合β-葡糖苷酶多肽的N末端处，而第二β-葡糖苷酶序列位于嵌合β-葡糖苷酶多肽的C末端处。在一些实施例中，第一、第二或两种β-葡糖苷酶序列还包含一个或多个糖基化位点。在某些实施例中，第一种和第二β-葡糖苷酶序列互相紧邻或直接互相连接。在其他实施例中，第一种和第二β-葡糖苷酶序列并非紧邻，而是通过接头结构域连接。在某些方面，第一或第二β-葡糖苷酶序列包含环区或代表环样结构的序列，其包含约3、4、5、6、7、8、9、10或11个氨基酸残基，所述序列包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT 的序列(SEQ ID NO:172)。在某些方面，第一或第二β-葡糖苷酶序列均不包含环序列。在一些实施例中，接头结构域包含环区，所述环区包含约3、4、5、6、7、8、9、10或11个氨基酸残基，所述环区包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。在一些实施例中，连接第一β-葡糖苷酶序列和第二β-葡糖苷酶序列的接头结构域位于中央(即，并非位于嵌合多肽的N末端或C末端)。在某些方面，嵌合β-葡糖苷酶的N末端序列包含长度至少200、250、300、350、400、450、 500、550或600个残基的来自Fv3C多肽或其变体的序列。在某些方面， N末端序列包含SEQ ID NOs:136-148所代表的多肽序列基序的一种或多种或全部。在某些方面，C末端序列包含长度至少50、75、100、125、 150、175或200个氨基酸残基的来自β-葡糖苷酶多肽或其变体的序列。在某些方面，C末端序列包含SEQ ID NOs:149-156所代表的多肽序列基序的一种或多种或全部。特

别地，所述两种或更多种β-葡糖苷酶序列的第一种是具有至少约200个氨基酸残基长度并且包含SEQ ID NOs:164-169的氨基酸序列基序中的至少2种(例如，至少2、3、4或全部)的序列，并且所述两种或更多种β-葡糖苷酶序列的第二种具有至少50个氨基酸残基长度并且包含SEQ ID NO:170。在某些实施例中，β-葡糖苷酶多肽、其变体或其杂合体/嵌合体还包含一个或多个糖基化位点。一个或多个糖基化位点可以位于C末端序列内部、N末端序列内部，或位于两者内部。

[0125] 在某些方面，本发明的非天然存在的纤维素酶或半纤维素酶组合物还包含一种或多种天然存在的半纤维素酶。在某些方面，该非天然存在的纤维素酶组合物具有优于天然酶的改善的稳定性，包括优于衍生嵌合β-葡糖苷酶的C末端或N末端序列的Fv3C。在某些方面，改善的稳定性包括在储存、表达或生产过程期间蛋白水解稳定性的改善。在某些方面，改善的稳定性包括在储存或生产条件下的酶活性损失的速度或程度的相关降低，其中酶活性损失的速度或程度优选地低于约50％、低于约40％、低于约 20％，更优选地低于约15％或甚至更优选地低于约10％。在某些方面，β- 葡糖苷酶多肽是嵌合或融合酶，其包含有效连接于里氏木霉Bgl3序列的 Fv3C多肽序列。在某些实施例中，β-葡糖苷酶多肽包含衍生自Fv3C多肽的N末端序列，以及衍生自里氏木霉Bgl3多肽的C末端序列。在某些方面，N末端序列或C末端序列可包含长度约3、4、5、6、7、8、9、10或 11个氨基酸残基的环序列，所述环序列包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。N末端和C末端序列可以互相紧邻或直接互相连接。在其他方面，N末端序列和C末端序列可以通过接头结构域连接。在某些实施例中，接头结构域包含长度约3、4、 5、6、7、8、9、10或11个氨基酸残基的环序列，所述环序列包含 FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。在某些方面，非天然存在的纤维素酶组合物包含β-葡糖苷酶活性。非天然存在的纤维素酶组合物还可以包含一种或多种木聚糖酶、β-木糖苷酶和/或L-α-阿拉伯呋喃糖苷酶活性。

[0126] Pa3D：

[0127] Pa3D的氨基酸序列(SEQ ID NO:54)在图29B和43中示出。SEQ ID NO:54是未成熟Pa3D的序列。Pa3D具有对应于SEQ ID NO:2的第1位至第17位残基(以下划线标出)的预测信号序列；对该信号序列的切割预计会产生成熟的蛋白质，该蛋白质具有对应于SEQ ID NO:54的第18位至第 733位残基的序列。对本发明的这种多肽和其他多肽的信号序列的预测使用SignalP-NN算法(www.cbs.dtu.dk)来进行。预测的保守结构域在图29B 中以粗体标出。对本发明的这种多肽和其他多肽的结构域预测基于Pfam、 SMART或NCBI数据库进行。Pa3D的残基E463和D262据预测分别作为催化性酸碱和亲核物质发挥功能，这是基于上文所述的来自例如以下的多个GH3家族的β-葡糖苷酶的序列比对结果：鹅掌柄孢霉(登录号 XP_001912683)、大丽轮枝菌、鞭毛藻丛赤壳菌(登录号 XP_003045443)、玉米赤霉(登录号XP_386781)、尖孢镰刀菌(登录号 BGL FOXG_02349)、黑曲霉(登录号CAK48740)、埃默森篮状菌(登录号AAL69548)、里氏木霉(登录号AAP57755)、里氏木霉(登录号AAA18473)、轮枝镰刀菌和新阿波罗栖热袍菌(登录号Q0GC07)等等，(参见图43)。如本文所用，“Pa3D多肽”在某些方面是指包含以下序列的多肽和/或其变体，所述序列与SEQ ID NO:54的第18位至第733 位残基中的至少50、75、100、125、150、175、200、250、300、350、 400、450、500、550、600、650或700个连续氨基酸残基具有至少85％、 86％、87％、88％、89％、90％、91％、92％、93％、

94％、95％、96％、 97％、98％、99％或100％序列同一性。与天然Pa3D相比，Pa3D多肽优选地在残基E463和D262处未改变。Pa3D多肽优选地在本文所述的GH3家族β-葡糖苷酶之间保守的至少70％、80％、90％、95％、98％或99％的氨基酸残基处未改变，如图43的比对结果中示出。Pa3D多肽合适地包含图 29B中示出的天然Pa3D的整个预测保守结构域。示例性的Pa3D多肽包含与图29B中示出的成熟Pa3D序列具有至少85％、86％、87％、88％、 89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或 100％同一性的序列。本发明的Pa3D多肽优选地具有β-葡糖苷酶活性。

[0128] 因此，本发明的Pa3D多肽合适地包含与SEQ ID NO:54的氨基酸序列或与SEQ ID NO:54第(i)18-282、(ii)18-601、(iii)18-733、(iv)356-601或 (v)356-733位残基具有至少80％、85％、90％、91％、92％、93％、94％、 95％、96％、97％、98％、99％或100％序列同一性的氨基酸序列。该多肽合适地具有β-葡糖苷酶活性。

[0129] 本发明的“Pa3D多肽”还可以指突变Pa3D多肽。可向Pa3D多肽引入氨基酸置换以改善β-葡糖苷酶活性和/或其他特性。例如，可以引入氨基酸置换，所述氨基酸置换增强Pa3D多肽对其底物的结合亲和力或改善 Pa3D催化β-D-葡糖苷中的末端非还原性残基水解的能力。在某些方面，突变Pa3D多肽包含一种或多种保守氨基酸置换。或者，突变Pa3D多肽可以包含一种或多种非保守的氨基酸置换。在某些方面，一种或多种氨基酸置换处于Pa3D多肽的CD中。或者一种或多种氨基酸置换位于Pa3D多肽的CBM中。一种或多种氨基酸置换可以同时位于CD和CBM中。在某些方面，Pa3D多肽氨基酸置换可以在氨基酸E463和/或D262处发生。在某些方面，Pa3D多肽氨基酸置换可以在氨基酸D87、R93、L136、R151、K184、H185、R195、M227、Y230、D262、W263、S406和/或E463的一个或多个或全部处发生。突变Pa3D多肽合适地具有β-葡糖苷酶活性。

[0130] 在某些方面，Pa3D多肽可以是两种β-葡糖苷酶序列的嵌合体/杂合体/ 融合体，其中第一序列衍生自第一β-葡糖苷酶，具有至少约200个氨基酸残基长度并且包含与Pa3D的等长度序列(SEQ ID NO:54)约60％(例如，约60％、65％、70％、75％或80％)或更高的同一性，并且其中第二序列衍生自第二β-葡糖苷酶，长度为至少约50个氨基酸残基，并且与SEQ ID NOs:56、58、60、62、64、66、68、70、72、74、76、78或79中任何一种的等长度序列具有约60％、70％、75％、80％或更高的同一性，或者包含SEQ ID NO:170的氨基酸序列基序。在某些方面，第一β-葡糖苷酶序列包含SEQ ID NO:54的至少约200个连续氨基酸残基的N末端序列，并且第二β-葡糖苷酶序列包含SEQ ID NOs:56、58、60、62、64、66、68、 70、72、74、76、78和79中任何一种的至少约50个连续氨基酸残基的C 末端序列，或者包含SEQ ID NO:170的氨基酸序列基序。

[0131] 在某些方面，本发明的Pa3D多肽包含嵌合体/杂合体/融合体，或者β- 葡糖苷酶序列的嵌合构建体，其中第一序列衍生自第一β-葡糖苷酶，长度为至少约200个氨基酸残基并且与SEQ ID NOs:56、58、60、62、64、 66、68、70、72、74、76、78和79中任何一种的等长度序列具有约60％ (例如，60％、65％、70％、75％或80％)或更高的同一性，或包含SEQ ID NOs:164-169的氨基酸序列基序的一种或多种或全部，并且第二序列衍生自第二β-葡糖苷酶，长度为至少约50个氨基酸残基，并且与Pa3D的等长度序列(SEQ ID NO:54)具有约60％、65％、70％、75％、80％或更高的同一性。例如，第一β-葡糖苷酶序列包含SEQ ID NOs:56、58、60、62、 64、66、68、70、72、74、76、78或79的至少200个连续氨基酸残基的 N末端序列，或包含

SEQ ID NOs:164-169的氨基酸序列基序的一种或多种或全部，并且第二β-葡糖苷酶序列包含SEQ ID NO:54的至少50个连续氨基酸残基的C末端序列。

[0132] 在某些方面，第一β-葡糖苷酶序列位于嵌合β-葡糖苷酶多肽的N末端处，而第二β-葡糖苷酶序列位于嵌合β-葡糖苷酶多肽的C末端处。在某些实施例中，第一、第二或两种β-葡糖苷酶序列还包含一个或多个糖基化位点。在某些实施例中，第一种和第二β-葡糖苷酶序列互相紧邻或直接互相连接。在其他实施例中，第一种和第二β-葡糖苷酶序列并非紧邻，而是通过接头结构域连接。在某些方面，第一或第二β-葡糖苷酶序列包含环区或代表环样结构的序列，其包含约3、4、5、6、7、8、9、10或11个氨基酸残基，所述序列包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT 的序列(SEQ ID NO:172)。在某些方面，第一或第二β-葡糖苷酶序列均不包含环序列。在一些实施例中，接头结构域包含环区，所述环区包含约3、4、5、6、7、8、9、10或11个氨基酸残基，所述环区包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。在一些实施例中，连接第一β-葡糖苷酶序列和第二β-葡糖苷酶序列的接头结构域位于中央(即，并非位于嵌合多肽的N末端或C末端)。在某些方面，嵌合β-葡糖苷酶的N末端序列包含长度至少200、250、300、350、400、450、 500、550或600个残基的来自Pa3D多肽或其变体的序列。在某些方面， N末端序列包含SEQ ID NOs:136-148所代表的多肽序列基序的一种或多种或全部，或优选地包含一种或多种或全部序列基序SEQ ID NOs:164-169。在某些方面，C末端序列包含长度至少50、75、100、125、150、175或200个氨基酸残基的来自β-葡糖苷酶多肽或其变体的序列。在某些方面，C 末端序列包含SEQ ID NOs:149-156所代表的多肽序列基序的一种或多种或全部，或优选地包含多肽序列基序SEQ ID NO:170。在某些实施例中，β- 葡糖苷酶多肽、其变体或其杂合体或嵌合体还包含一个或多个糖基化位点。一个或多个糖基化位点可以位于C末端序列内部、N末端序列内部或位于两者内部。

[0133] 在某些方面，本发明的非天然存在的纤维素酶或半纤维素酶组合物还包含一种或多种天然存在的半纤维素酶。在某些方面，该非天然存在的纤维素酶组合物具有优于天然酶的改善的稳定性，包括优于衍生嵌合β-葡糖苷酶的C末端或N末端序列的Pa3D。在某些方面，改善的稳定性包括在储存、表达或生产过程期间蛋白水解稳定性的改善。在某些方面，改善的稳定性包括在储存或生产条件下的酶活性损失的速度或程度的相关降低，其中酶活性损失优选地低于约50％，低于约40％，低于约20％，更优选地低于约15％，或甚至更优选地低于约10％。在某些方面，N末端序列或C 末端序列可包含长度约3、4、5、6、7、8、9、10或11个氨基酸残基的环序列，所述环序列包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。N末端和C末端序列可以互相紧邻或直接互相连接。在其他方面，N末端序列和C末端序列可以通过接头结构域连接。在某些实施例中，接头结构域包含长度约3、4、5、6、7、8、9、10或11个氨基酸残基的环序列，所述环序列包含FDRRSPG的序列(SEQ ID NO:171) 或FD(R/K)YNIT的序列(SEQ ID NO:172)。在某些方面，非天然存在的纤维素酶组合物包含β-葡糖苷酶活性。在某些方面，非天然存在的纤维素酶组合物还包含一种或多种木聚糖酶、β-木糖苷酶和/或L-α-阿拉伯呋喃糖苷酶活性。

[0134] Fv3G

[0135] Fv3G的氨基酸序列(SEQ ID NO:56)在图30B和43中示出。SEQ ID NO:56是未成熟Fv3G的序列。Fv3G具有与SEQ ID NO:56的第1位至第 21位(以下划线标出)相对应的预测的

信号序列;对该信号序列的切割据预测产生具有与SEQ ID NO:56的第22至780位相对应的序列的成熟蛋白。如上所述,信号序列的预测使用SignalP-NN算法 (http://www.cbs.dtu.dk)来进行,正如对本发明的其他多肽所进行的预测。预测的保守结构域在图30B中以粗体标出。正如对本发明的其他多肽进行的预测,结构域预测基于Pfam、SMART或NCBI数据库进行。Fv3G残基 E509和D272据预测分别作为催化性酸碱和亲核物质发挥功能,这是基于上文所述的GH3葡糖苷酶的序列比对结果来自例如以下的:鹅掌柄孢霉 (登录号XP_001912683)、大丽轮枝菌、鞭毛藻丛赤壳菌(登录号 XP_003045443)、玉米赤霉(登录号XP_386781)、尖孢镰刀菌(登录号 BGL FOXG_02349)、黑曲霉(登录号CAK48740)、埃默森篮状菌(登录号AAL69548)、里氏木霉(登录号AAP57755)、里氏木霉(登录号 AAA18473)、轮枝镰刀菌和新阿波罗栖热袍菌(登录号Q0GC07)等等 (参见图43)。如本文所用,"Fv3G多肽"在某些方面是指包含以下序列的多肽和/或其变体,所述序列与SEQ ID NO:56第20位至第780位残基中的至少50、75、100、125、150、175、200、250、300、350、400、 450、500、550、600、650、700或750个连续氨基酸残基具有至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、 97%、98%、99%或100%序列同一性。与天然Fv3G相比,Fv3G多肽优选地在残基E509和D272处未改变。Fv3G多肽优选地在本文所述的GH3家族β-葡糖苷酶之间保守的至少70%、80%、90%、95%、98%或99%的氨基酸残基处未改变,如图43的比对结果中示出。Fv3G多肽合适地包含图 30B中示出的天然Fv3G的整个预测保守结构域。示例性的Fv3G多肽包含与图30B中示出的成熟Fv3G序列具有至少85%、86%、87%、88%、 89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或 100%同一性的序列。本发明的Fv3G多肽优选地具有β-葡糖苷酶活性。

[0136] 因此,本发明的Fv3G多肽合适地包含与SEQ ID NO:56的氨基酸序列或与SEQ ID NO:56第(i)22-292、(ii)22-629、(iii)22-780、(iv)373-629或 (v)373-780位残基具有至少80%、85%、90%、91%、92%、93%、94%、 95%、96%、97%、98%、99%或100%序列同一性的氨基酸序列。该多肽合适地具有β-葡糖苷酶活性。

[0137] 在某些方面,本发明的"Fv3G多肽"还可以指突变Fv3G多肽。可以向Fv3G多肽引入氨基酸置换以改善该分子的β-葡糖苷酶活性。例如,可以向Fv3G多肽引入氨基酸置换,所述氨基酸置换增强该Fv3G多肽对其底物的结合亲和力或改善Fv3G催化β-D-葡糖苷中的末端非还原性残基水解的能力。在某些方面,突变Fv3G多肽包含一种或多种保守氨基酸置换。在某些方面,突变Fv3G多肽包含一种或多种非保守的氨基酸置换。在某些方面,一种或多种氨基酸置换位于Fv3G多肽的CD中。在某些方面,一种或多种氨基酸置换位于Fv3G多肽的CBM中。在某些方面,一种或多种氨基酸置换同时位于CD和CBM中。在某些方面,Fv3G多肽的氨基酸置换可以在氨基酸E509和/或D272处发生。在某些方面,Fv3G多肽的氨基酸置换可以在氨基酸D101、R107、L150、R165、K198、H199、 R209、M237、Y240、D272、W273、S455和/或E509中的一个或多个处发生。突变Fv3G多肽合适地具有β-葡糖苷酶活性。

[0138] 在某些方面,Fv3G多肽包含两种β-葡糖苷酶序列的嵌合体,其中第一β-葡糖苷酶序列具有至少约200个氨基酸残基长度,并且包含与Fv3G 的等长度序列(SEQ ID NO:56)约60%、65%、70%、75%或80%或更高的序列同一性,并且其中第二β-葡糖苷酶序列具有至少约50个氨基酸残基长度并且包含与SEQ ID NOs:54、58、60、62、64、66、68、70、72、74、 76、78和79中任何一种的等长度序列至少约60%、65%、70%、75%、 80%或更高的序列同一

性，或包含多肽序列基序SEQ ID NO:170。在某些方面，第一β-葡糖苷酶序列包含SEQ ID NO:56的至少200个氨基酸残基的N末端序列，并且第二β-葡糖苷酶序列包含SEQ ID NOs:54、58、60、62、64、66、68、70、72、74、76、78和79中任何一种的至少约50个连续氨基酸残基的C末端序列，或包含SEQ ID NO:170的基序。

[0139] 在某些方面，本发明的Fv3G多肽包含两种β-葡糖苷酶序列的嵌合体或嵌合构建体，其中第一β-葡糖苷酶序列具有至少约200个氨基酸残基长度，并且包含与SEQ ID NOs:54、58、60、62、64、66、68、70、72、74、76、78和79中任何一种的等长度系列约60％、65％、70％、75％、80％或更高的序列同一性，或包含SEQ ID NOs:164-169的基序中的一种或多种或全部，而第二β-葡糖苷酶序列具有至少约50个氨基酸残基长度，包含与Fv3G的等长度序列(SEQ ID NO:56)约60％、65％、70％、75％、80％或更高的序列同一性。在某些方面，第一β-葡糖苷酶序列包含SEQ ID NOs:54、58、60、62、64、66、68、70、72、74、76、78和79中任何一种的至少200个氨基酸残基的N末端序列，或包含SEQ ID NOs:164-169 的序列基序的一种或多种或全部，并且第二β-葡糖苷酶序列包含SEQ ID NO:56的至少50个连续氨基酸残基的C末端序列。

[0140] 在某些方面，第一β-葡糖苷酶序列位于嵌合β-葡糖苷酶多肽的N末端处，而第二β-葡糖苷酶序列位于嵌合β-葡糖苷酶多肽的C末端处。在某些实施例中，第一、第二或两种β-葡糖苷酶序列还包含一个或多个糖基化位点。在某些实施例中，第一种和第二β-葡糖苷酶序列互相紧邻或直接互相连接。在其他实施例中，第一种和第二β-葡糖苷酶序列并非紧邻，而是通过接头结构域连接。在某些方面，第一或第二β-葡糖苷酶序列包含环区或代表环样结构的序列，其包含约3、4、5、6、7、8、9、10或11个氨基酸残基，所述序列包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT 的序列(SEQ ID NO:172)。在某些方面，第一或第二β-葡糖苷酶序列均不包含环序列。在一些实施例中，接头结构域包含环区，所述环区包含约3、4、5、6、7、8、9、10或11个氨基酸残基，所述环区包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。在一些实施例中，连接第一β-葡糖苷酶序列和第二β-葡糖苷酶序列的接头结构域位于中央(即，并非位于嵌合多肽的N末端或C末端)。在某些方面，嵌合β-葡糖苷酶的N末端序列包含长度至少200、250、300、350、400、450、500、550或600个残基的来自Fv3G多肽或其变体的序列。在某些方面，N末端序列包含SEQ ID NOs:136-148所代表的多肽序列基序的一种或多种或全部，或优选地包含SEQ ID NOs:164-169中的一种或多种或全部。在某些方面，C末端序列包含长度至少50、75、100、125、150、175或200个氨基酸残基的来自β-葡糖苷酶多肽或其变体的序列。在某些方面，C末端序列包含SEQ ID NOs:149-156(或优选地SEQ ID NO:170)所代表的多肽序列基序的一种或多种或全部。该β-葡糖苷酶多肽、其变体或其杂合体或嵌合体还可以包含一个或多个糖基化位点。一个或多个糖基化位点可以位于C末端序列内部、N末端序列内部或位于两者内部。

[0141] 在某些方面，本发明的非天然存在的纤维素酶或半纤维素酶组合物还包含一种或多种天然存在的半纤维素酶。在某些方面，非天然存在的纤维素酶组合物具有优于天然酶的改善的稳定性，所述天然酶包括衍生嵌合β- 葡糖苷酶的C末端或N末端序列的Fv3G。在某些方面，改善的稳定性包括在储存、表达或生产过程期间蛋白水解稳定性的改善。在某些方面，改善的稳定性包括在储存或生产条件下的酶活性损失的速度或程度的相关降低，其中酶活性损失优选地低于约50％，低于约40％，低于约20％，更优选地低于约15％，或甚至更

优选地低于约10％。在某些方面，N末端序列或C末端序列可包含长度约3、4、5、6、7、8、9、10或11个氨基酸残基的环序列，所述环序列包含FDRRSPG的序列(SEQ ID NO:171)或 FD(R/K)YNIT的序列(SEQ ID NO:172)。N末端和C末端序列可以互相紧邻或直接互相连接。在其他方面，N末端序列和C末端序列可以通过接头结构域连接。在某些实施例中，接头结构域包含长度约3、4、5、6、7、8、 9、10或11个氨基酸残基的环序列，所述环序列包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。在某些方面，非天然存在的纤维素酶组合物包含β-葡糖苷酶活性。在某些方面，非天然存在的纤维素酶组合物还包含一种或多种木聚糖酶、β-木糖苷酶和/或L-α- 阿拉伯呋喃糖苷酶活性。

[0142] Fv3D

[0143] Fv3D的氨基酸序列(SEQ ID NO:58)在图31B和43中示出。SEQ ID NO:58是未成熟Fv3D的序列。Fv3D具有与SEQ ID NO:58的第1位至第 19位(以下划线标出)相对应的预测的信号序列；对该信号序列的切割据预测产生具有与SEQ ID NO:58的第20至811位相对应的序列的成熟蛋白。使用SignalP-NN算法进行信号序列预测。预测的保守结构域在图31B 中以粗体标出。结构域预测基于Pfam、SMART或NCBI数据库进行。 Fv3D的残基E534和D301据预测分别作为催化性酸-碱和亲核物质发挥功能，这是基于上文所述的来自例如以下的GH3葡糖苷酶的序列比对结果：鹅掌柄孢霉(登录号XP_001912683)、大丽轮枝菌、鞭毛藻丛赤壳菌(登录号XP_003045443)、玉米赤霉(登录号XP_386781)、尖孢镰刀菌(登录号BGL FOXG_02349)、黑曲霉(登录号CAK48740)、埃默森篮状菌(登录号AAL69548)、里氏木霉(登录号AAP57755)、里氏木霉(登录号AAA18473)、轮枝镰刀菌和新阿波罗栖热袍菌(登录号Q0GC07)等等(参见图43)。如本文所用，“Fv3D多肽”在某些方面是指包含以下序列的多肽和/或其变体，所述序列与SEQ ID NO:58第20位至第811位残基中的至少50、75、100、125、150、175、200、250、300、 350、400、450、500、550、600、650、700或750个连续氨基酸残基具有至少85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、 95％、96％、97％、98％、99％或100％序列同一性。与天然Fv3D相比， Fv3D多肽优选地在残基E534和D301处未改变。Fv3D多肽优选地在本文所述的GH3家族β-葡糖苷酶之间保守的至少70％、80％、90％、95％、98％或99％的氨基酸残基处未改变，如图43的比对结果中示出。Fv3D多肽合适地包含图31B中示出的天然Fv3D的整个预测保守结构域。示例性的Fv3D多肽包含与图31B中示出的成熟Fv3D序列具有至少85％、86％、 87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、 98％、99％或100％同一性的序列。本发明的Fv3D多肽优选地具有β-葡糖苷酶活性。

[0144] 因此，本发明的Fv3D多肽合适地包含与SEQ ID NO:58的氨基酸序列或与SEQ ID NO:58第(i)20-321、(ii)20-651、(iii)20-811、(iv)423-651或 (v)423-811位残基具有至少80％、85％、90％、91％、92％、93％、94％、 95％、96％、97％、98％、99％或100％序列同一性的氨基酸序列。该多肽合适地具有β-葡糖苷酶活性。

[0145] 在某些方面，本发明的“Fv3D多肽”还可以指突变Fv3D多肽。可以向Fv3D多肽引入氨基酸置换以改善该分子的β-葡糖苷酶活性。例如，可以向Fv3D多肽引入氨基酸置换，所述氨基酸置换增强Fv3D多肽对其底物的结合亲和力或改善Fv3D催化β-D-葡糖苷中的末端非还原性残基水解的能力。在某些方面，突变Fv3D多肽包含一种或多种保守氨基酸置换。在某些方面，突变Fv3D多肽包含一种或多种非保守的氨基酸置换。在某些方面，一种或多种氨基酸置换位于Fv3G多肽的CD中。在某些方面，一种或多种氨基酸置换位于Fv3D多肽的CBM中。

在某些方面，一种或多种氨基酸置换同时位于CD和CBM中。在某些方面，Fv3D多肽的氨基酸置换可以在氨基酸E534和/或D301处发生。在某些方面，Fv3D多肽的氨基酸置换可以在氨基酸D111、R117、L160、R175、K208、H209、R219、 M266、Y269、D301、W302、S472和/或E534中的一个或多个处发生。突变Fv3D多肽合适地具有β-葡糖苷酶活性。

[0146] 在某些方面，Fv3D多肽包含两种β-葡糖苷酶序列的嵌合体，其中第一β-葡糖苷酶序列具有至少约200个氨基酸残基长度并且包含与Fv3D的等长度序列(SEQ ID NO:58)约60％、65％、70％、75％或80％或更高的序列同一性，并且其中第二β-葡糖苷酶序列具有至少约50个氨基酸残基长度，并且包含与SEQ ID NOs:54、56、60、62、64、66、68、70、72、 74、76、78和79中的至少一个的等长度序列至少约60％、65％、70％、 75％、80％或更高的序列同一性。在某些方面，第一β-葡糖苷酶序列包含 SEQ ID NO:58的至少200个氨基酸残基的N末端序列，并且第二β-葡糖苷酶序列包含SEQ ID NOs:54、56、60、62、64、66、68、70、72、74、76、78和79中任何一种的至少约50个连续氨基酸残基的C末端序列。

[0147] 在某些方面，本发明的Fv3D多肽包含杂合体/融合体/嵌合体或者两种β-葡糖苷酶序列的嵌合构建体，其中第一β-葡糖苷酶序列具有至少约200 个氨基酸残基长度，并且包含与SEQ ID NOs:54、56、60、62、64、66、 68、70、72、74、76、78和79中任何一种的等长度序列约60％、65％、 70％、75％、80％或更高的序列同一性，或包含SEQ ID NOs:164-169多肽序列基序的一种或多种或全部，而第二β-葡糖苷酶序列具有至少约50个氨基酸残基长度并且包含与Fv3D的等长度序列(SEQ ID NO:58)约60％、 65％、70％、75％、80％或更高的序列同一性。在某些方面，第一β-葡糖苷酶序列包含SEQ ID NOs:54、56、60、62、64、66、68、70、72、74、 76、78和79中任何一种的等长度序列约60％、65％、70％、75％、80％或更高的序列同一性，或包含SEQ ID NOs:164-169多肽序列基序的一种或多种或全部，并且第二β-葡糖苷酶序列包含SEQ ID NO:58的至少50个连续氨基酸残基的C末端序列。

[0148] 在某些方面，第一β-葡糖苷酶序列位于嵌合β-葡糖苷酶多肽的N末端处，而第二β-葡糖苷酶序列位于嵌合β-葡糖苷酶多肽的C末端处。在某些实施例中，第一、第二或两种β-葡糖苷酶序列还包含一个或多个糖基化位点。在某些实施例中，第一种和第二β-葡糖苷酶序列互相紧邻或直接互相连接。在其他实施例中，第一种和第二β-葡糖苷酶序列并非紧邻，而是通过接头结构域连接。在某些方面，第一或第二β-葡糖苷酶序列包含环区或代表环样结构的序列，其包含约3、4、5、6、7、8、9、10或11个氨基酸残基，所述序列包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT 的序列(SEQ ID NO:172)。在某些方面，第一或第二β-葡糖苷酶序列均不包含环序列。在一些实施例中，接头结构域包含环区，所述环区包含约3、4、5、6、7、8、9、10或11个氨基酸残基，所述环区包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。在一些实施例中，连接第一β-葡糖苷酶序列和第二β-葡糖苷酶序列的接头结构域位于中央(即，并非位于嵌合多肽的N末端或C末端)。在某些方面，嵌合β-葡糖苷酶的N末端序列包含长度至少200、250、300、350、400、450、 500、550或600个残基的来自Fv3D多肽或其变体的序列。在某些方面， N末端序列包含SEQ ID NOs:136-148所代表的多肽序列基序的一种或多种或全部，或优选地包含序列基序SEQ ID NOs:164-169。在某些方面，C末端序列包含长度至少50、75、100、125、150、175或200个氨基酸残基的来自β-葡糖苷酶多肽或其变体的序列。在某些方面，C末端序列包含SEQ ID NOs:149-156所代表的多肽序列基序的一种或多种或全部，或优选地包含基序SEQ ID NO:170。在某些实

施例中，β-葡糖苷酶多肽、其变体或其杂合体或嵌合体还包含一个或多个糖基化位点。一个或多个糖基化位点可以位于C末端序列内部、N末端序列内部或位于两者内部。

[0149] 在某些方面，本发明的非天然存在的纤维素酶或半纤维素酶组合物还包含一种或多种天然存在的半纤维素酶。在某些方面，非天然存在的纤维素酶组合物具有优于天然酶的改善的稳定性，所述天然酶包括衍生嵌合β- 葡糖苷酶的C末端或N末端序列的Fv3D。在某些方面，改善的稳定性包括在储存、表达或生产过程期间蛋白水解稳定性的改善。在某些方面，改善的稳定性包括在储存或生产条件下的酶活性损失的速度或程度的相关降低，其中酶活性损失优选地低于约50％，低于约40％，低于约20％，更优选地低于约15％，或甚至更优选地低于约10％。在某些方面，N末端序列或C末端序列可包含长度约3、4、5、6、7、8、9、10或11个氨基酸残基的环序列，所述环序列包含FDRRSPG的序列(SEQ ID NO:171)或 FD(R/K)YNIT的序列(SEQ ID NO:172)。N末端和C末端序列可以互相紧邻或直接互相连接。在其他方面，N末端序列和C末端序列可以通过接头结构域连接。在某些实施例中，接头结构域包含长度约3、4、5、6、7、8、 9、10或11个氨基酸残基的环序列，所述环序列包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。在某些方面，非天然存在的纤维素酶组合物包含β-葡糖苷酶活性。在某些方面，非天然存在的纤维素酶组合物还包含一种或多种木聚糖酶、β-木糖苷酶和/或L-α- 阿拉伯呋喃糖苷酶活性。

[0150] Tr3A

[0151] Tr3A的氨基酸序列(SEQ ID NO:62)在图33B和43中示出。Tr3A也称为里氏木霉Bgl1。SEQ ID NO:62是未成熟Tr3A的序列。Tr3A具有与 SEQ ID NO:62的第1位至第19位(以下划线标出)相对应的预测的信号序列；对该信号序列的切割据预测产生具有与SEQ ID NO:62的第20至 744位相对应的序列的成熟蛋白。使用SignalP-NN算法进行信号序列预测。预测的保守结构域在图33B中以粗体标出。结构域预测基于Pfam、SMART或NCBI数据库进行。Tr3A的残基E472和D267据预测分别作为催化性酸碱和亲核物质发挥功能，这是基于上文所述的来自例如以下的 GH3葡糖苷酶的序列比对结果：鹅掌柄孢霉(登录号XP_001912683)、大丽轮枝菌、鞭毛藻丛赤壳菌(登录号XP_003045443)、玉米赤霉(登录号XP_386781)、尖孢镰刀菌(登录号BGL FOXG_02349)、黑曲霉 (登录号CAK48740)、埃默森篮状菌(登录号AAL69548)、里氏木霉 (登录号AAP57755)、里氏木霉(登录号AAA18473)、轮枝镰刀菌和新阿波罗栖热袍菌(登录号Q0GC07)等等(参见图43)。如本文所用，“Tr3A多肽”在某些方面是指包含以下序列的多肽和/或其变体，所述序列与SEQ ID NO:62的第20位至第744位残基中的至少50、75、100、 125、150、175、200、250、300、350、400、450、500、550、600、650 或700个连续氨基酸残基具有至少85％、86％、87％、88％、89％、90％、 91％、92％、93％、94％、95％、96％、97％、98％、99％或100％序列同一性。与天然Tr3A相比，Tr3A多肽优选地在残基E472和D267处未改变。 Tr3A多肽优选地在本文所述的GH3家族β-葡糖苷酶之间保守的至少70％、80％、90％、95％、98％或99％的氨基酸残基处未改变，如图43的比对结果中示出。Tr3A多肽合适地包含图33B中示出的天然Tr3A的整个预测保守结构域。示例性的Tr3A多肽包含与图33B中示出的成熟Tr3A序列具有至少85％、86％、87％、88％、89％、90％、91％、92％、93％、 94％、95％、96％、97％、98％、99％或100％同一性的序列。本发明的 Tr3A多肽优选地具有β-葡糖苷酶活性。

[0152] 因此，本发明的Tr3A多肽合适地包含与SEQ ID NO:62的氨基酸序列或与SEQ ID

NO:62第(i)20-287、(ii)22-611、(iii)20-744、(iv)362-611或 (v)362-744位残基具有至少80％、85％、90％、91％、92％、93％、94％、 95％、96％、97％、98％、99％或100％序列同一性的氨基酸序列。该多肽合适地具有β-葡糖苷酶活性。

[0153] 在某些方面，本发明的“Tr3A多肽”还可以指突变Tr3A多肽。可以向Tr3A多肽引入氨基酸置换以改善该分子的β-葡糖苷酶活性。例如，可以向Tr3A多肽引入氨基酸置换，所述氨基酸置换增强Tr3A多肽对其底物的结合亲和力或改善Tr3A催化β-D-葡糖苷中的末端非还原性残基水解的能力。在某些方面，突变Tr3A多肽包含一种或多种保守氨基酸置换。在某些方面，突变Tr3A多肽包含一种或多种非保守的氨基酸置换。在某些方面，一种或多种氨基酸置换位于Tr3A多肽的CD中。在某些方面，一种或多种氨基酸置换位于Tr3A多肽的CBM中。在某些方面，一种或多种氨基酸置换同时位于CD和CBM中。在某些方面，Tr3A多肽的氨基酸置换可以在氨基酸E472和/或D267处发生。在某些方面，Tr3A多肽的氨基酸置换可以在氨基酸D92、R98、L141、R156、K189、H190、R200、 M232、Y235、D267、W268、S415和/或E472中的一个或多个处发生。突变Tr3A多肽合适地具有β-葡糖苷酶活性。

[0154] 在某些方面，Tr3A多肽包含两种β-葡糖苷酶序列的嵌合体/融合体/杂合体，其中第一β-葡糖苷酶序列具有至少约200个氨基酸残基长度，并且包含与Tr3A的等长度序列(SEQ ID NO:62)约60％、65％、70％、75％或 80％或更高的序列同一性，并且其中第二β-葡糖苷酶序列具有至少约50个氨基酸残基长度并且包含与SEQ ID NOs:54、56、58、60、64、68、70、 72、74、76、78和79中任何一种的等长度序列至少约60％、65％、70％、 75％、80％或更高的序列同一性，或包含多肽序列基序SEQ ID NO:170。在某些方面，第一β-葡糖苷酶序列包含SEQ ID NO:62的至少200个氨基酸残基的N末端序列，并且第二β-葡糖苷酶序列包含SEQ ID NOs:54、 56、58、60、64、66、68、70、72、74、76、78和79中任何一种的至少约50个连续氨基酸残基的C末端序列，或包含多肽序列基序SEQ ID NO:170。

[0155] 在某些方面，本发明的Tr3A多肽包含两种β-葡糖苷酶序列的嵌合体或嵌合构建体，其中第一β-葡糖苷酶序列具有至少约200个氨基酸残基长度，并且包含与SEQ ID NOs:54、56、58、60、64、66、68、70、72、 74、76、78和79中任何一种的等长度序列约60％、65％、70％、75％、 80％或更高的序列同一性，或包含SEQ ID NOs:164-169多肽序列基序的一种或多种或全部，而第二β-葡糖苷酶序列具有至少约50个氨基酸残基长度并且包含与Tr3A的等长度序列(SEQ ID NO:62)约60％、65％、70％、 75％、80％或更高的序列同一性。在某些方面，第一β-葡糖苷酶序列包含 SEQ ID NOs:54、56、58、60、64、66、68、70、72、74、76、78和79中任何一种的等长度序列约60％、65％、70％、75％、80％或更高的序列同一性，或包含SEQ ID NOs:164-169多肽序列基序的一种或多种或全部，并且第二β-葡糖苷酶序列包含SEQ ID NO:62的至少50个连续氨基酸残基的 C末端序列。

[0156] 在某些方面，第一β-葡糖苷酶序列位于嵌合β-葡糖苷酶多肽的N末端处，而第二β-葡糖苷酶序列位于嵌合β-葡糖苷酶多肽的C末端处。在某些实施例中，第一、第二或两种β-葡糖苷酶序列还包含一个或多个糖基化位点。在某些实施例中，第一种和第二β-葡糖苷酶序列互相紧邻或直接互相连接。在其他实施例中，第一种和第二β-葡糖苷酶序列并非紧邻，而是通过接头结构域连接。在某些方面，第一或第二β-葡糖苷酶序列包含环区或代表环样结构的序列，其包含约3、4、5、6、7、8、9、10或11个氨基酸残基，所述序列包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT 的序列(SEQ ID NO:172)。在某些方面，第一或第二β-葡

糖苷酶序列均不包含环序列。在一些实施例中，接头结构域包含环区，所述环区包含约3、4、5、6、7、8、9、10或11个氨基酸残基，所述环区包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。在一些实施例中，连接第一β-葡糖苷酶序列和第二β-葡糖苷酶序列的接头结构域位于中央(即，并非位于嵌合多肽的N末端或C末端)。在某些方面，嵌合β-葡糖苷酶的N末端序列包含长度至少200、250、300、350、400、450、 500、550或600个残基的来自Tr3A或其变体的序列。在某些方面，N末端序列包含SEQ ID NOs:136-148所代表的多肽序列基序的一种或多种或全部，或优选地包含序列基序SEQ ID NOs:164-169。在某些方面，C末端序列包含长度至少50、75、100、125、150、175或200个氨基酸残基的来自β-葡糖苷酶多肽或其变体的序列。在某些方面，C末端序列包含SEQ ID NOs:149-156所代表的多肽序列基序的一种或多种或全部，或优选地包含序列基序SEQ ID NO:170。在某些实施例中，β-葡糖苷酶多肽、其变体或其杂合体或嵌合体还包含一个或多个糖基化位点。一个或多个糖基化位点可以位于C末端序列内部、N末端序列内部或位于两者内部。

[0157] 在某些方面，本发明的非天然存在的纤维素酶或半纤维素酶组合物还包含一种或多种天然存在的半纤维素酶。在某些方面，非天然存在的纤维素酶组合物具有优于天然酶的改善的稳定性，所述天然酶包括衍生嵌合β- 葡糖苷酶的C末端或N末端序列的Tr3A。在某些方面，改善的稳定性包括在储存、表达或生产过程期间蛋白水解稳定性的改善。在某些方面，改善的稳定性包括在储存或生产条件下的酶活性损失的速度或程度的相关降低，其中酶活性损失优选地低于约50％，低于约40％，低于约20％，更优选地低于约15％，或甚至更优选地低于约10％。在某些方面，N末端序列或C末端序列可包含长度约3、4、5、6、7、8、9、10或11个氨基酸残基的环序列，所述环序列包含FDRRSPG的序列(SEQ ID NO:171)或 FD(R/K)YNIT的序列(SEQ ID NO:172)。N末端和C末端序列可以互相紧邻或直接互相连接。在其他方面，N末端序列和C末端序列可以通过接头结构域连接。在某些实施例中，接头结构域包含长度约3、4、5、6、7、8、 9、10或11个氨基酸残基的环序列，所述环序列包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。非天然存在的纤维素酶组合物包含β-葡糖苷酶活性。非天然存在的纤维素酶组合物还可以包含一种或多种木聚糖酶、β-木糖苷酶和/或L-α-阿拉伯呋喃糖苷酶活性。

[0158] Tr3B

[0159] Tr3B的氨基酸序列(SEQ ID NO:64)在图34B和43中示出。Tr3B也称为“里氏木霉Bgl3”或“里氏木霉Cel3B”。SEQ ID NO:64是未成熟 Tr3B的序列。Tr3B具有与SEQ ID NO:64的第1位至第18位(以下划线标出)相对应的预测的信号序列;对该信号序列的切割据预测产生具有与 SEQ ID NO:64的第19至874位相对应的序列的成熟蛋白。使用SignalP- NN算法进行信号序列预测。预测的保守结构域在图34B中以粗体标出。结构域预测基于Pfam、SMART或NCBI数据库进行。Tr3B的残基E516 和D287据预测分别作为催化性酸碱和亲核物质发挥功能，这是基于上文所述的来自例如以下的GH3葡糖苷酶的序列比对结果:鹅掌柄孢霉(登录号XP_001912683)、大丽轮枝菌、鞭毛藻丛赤壳菌(登录号 XP_003045443)、玉米赤霉(登录号XP_386781)、尖孢镰刀菌(登录号 BGL FOXG_02349)、黑曲霉(登录号CAK48740)、埃默森篮状菌(登录号AAL69548)、里氏木霉(登录号AAP57755)、里氏木霉(登录号AAA18473)、轮枝镰刀菌和新阿波罗栖热袍菌(登录号Q0GC07)等等 (参见图43)。如本文所用，“Tr3B多肽”在某些方面是指包含以下序列的多肽和/或其变体，所述序列与SEQ ID NO:64的第19位

至第874位残基中的至少50、75、100、125、150、175、200、250、300、350、400、 450、500、550、600、650、700、750、800或850个连续氨基酸残基具有至少85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、 95％、96％、97％、98％、99％或100％序列同一性。与天然Tr3B相比， Tr3B多肽优选地在残基E516和D287处未改变。Tr3B多肽优选地在本文所述的GH3家族β-葡糖苷酶之间保守的至少70％、80％、90％、95％、 98％或99％的氨基酸残基处未改变，如图43的比对结果中示出。Tr3B多肽合适地包含图34B中示出的天然Tr3B的整个预测保守结构域。示例性的Tr3A多肽包含与图34B中示出的成熟Tr3B序列具有至少85％、86％、 87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、 98％、99％或100％同一性的序列。本发明的Tr3B多肽优选地具有β-葡糖苷酶活性。

[0160] 因此，本发明的Tr3B多肽合适地包含与SEQ ID NO:64的氨基酸序列或与SEQ ID NO:64第(i)19-307、(ii)19-640、(iii)19-874、(iv)407-640或 (v)407-874位残基具有至少80％、85％、90％、91％、92％、93％、94％、 95％、96％、97％、98％、99％或100％序列同一性的氨基酸序列。该多肽合适地具有β-葡糖苷酶活性。

[0161] 在某些方面，本发明的“Tr3B多肽”还可以指突变Tr3B多肽。可以向Tr3B多肽引入氨基酸置换以改善该分子的β-葡糖苷酶活性。例如，可以向Tr3B多肽中引入氨基酸置换，所述氨基酸置换增强Tr3B多肽对其底物的结合亲和力或改善Tr3B催化β-D-葡糖苷中的末端非还原性残基水解的能力。在某些方面，突变Tr3B多肽包含一种或多种保守氨基酸置换。在某些方面，突变Tr3B多肽包含一种或多种非保守的氨基酸置换。在某些方面，一种或多种氨基酸置换位于Tr3B多肽的CD中。在某些方面，一种或多种氨基酸置换位于Tr3B多肽的CBM中。在某些方面，一种或多种氨基酸置换同时位于CD和CBM中。在某些方面，Tr3B多肽的氨基酸置换可以在氨基酸E516和/或D287处发生。在某些方面，Tr3B多肽的氨基酸置换可以在氨基酸D99、R105、L148、R163、K196、H197、R207、 M252、Y255、D287、W288、S457和/或E516中的一个或多个处发生。突变Tr3B多肽合适地具有β-葡糖苷酶活性。

[0162] 在某些方面，Tr3B多肽包含两种β-葡糖苷酶序列的嵌合体/杂合体/融合体，其中第一β-葡糖苷酶序列具有至少约200个氨基酸残基长度，并且包含与Tr3B的等长度序列(SEQ ID NO:64)约60％、65％、70％、75％或 80％或更高的序列同一性，并且其中第二β-葡糖苷酶序列具有至少约50个氨基酸残基长度并且包含与SEQ ID NOs:54、56、58、60、62、66、68、 70、72、74、76、78和79中任何一种的等长度序列至少约60％、65％、 70％、75％、80％或更高的序列同一性，或包含多肽序列基序SEQ ID NO:170。在某些方面，第一β-葡糖苷酶序列包含SEQ ID NO:64的至少 200个氨基酸残基的N末端序列，并且第二β-葡糖苷酶序列包含SEQ ID NOs:54、56、58、60、62、68、70、72、74、76、78和79中任何一种的至少约50个连续氨基酸残基的C末端序列，或包含多肽序列基序SEQ ID NO:170。

[0163] 在某些方面，本发明的Tr3B多肽包含两种β-葡糖苷酶序列的嵌合体或嵌合构建体，其中第一β-葡糖苷酶序列具有至少约200个氨基酸残基长度，并且包含与SEQ ID NOs:54、56、58、60、62、66、68、70、72、 74、76、78和79中任何一种的等长度序列约60％、65％、70％、75％、 80％或更高的序列同一性，或包含SEQ ID NOs:164-169的多肽序列基序的一种或多种，而第二β-葡糖苷酶序列具有至少约50个氨基酸残基长度并且包含与Tr3B的等长度序列(SEQ ID NO:64)约60％、65％、70％、75％、 80％或更高的序列同一性。在某些方面，第一β-葡糖苷酶序列包含SEQ ID NOs:54、56、58、60、62、66、68、70、72、74、76、78和79中任

何一种的至少200个氨基酸残基的N末端序列，或包含SEQ ID NOs:164-169多肽序列基序的一种或多种或全部，并且第二β-葡糖苷酶序列包含SEQ ID NO:64的至少50个连续氨基酸残基的C末端序列。

[0164] 在某些方面，第一β-葡糖苷酶序列位于嵌合β-葡糖苷酶多肽的N末端处，而第二β-葡糖苷酶序列位于嵌合β-葡糖苷酶多肽的C末端处。在某些实施例中，第一、第二或两种β-葡糖苷酶序列还包含一个或多个糖基化位点。在某些实施例中，第一种和第二β-葡糖苷酶序列互相紧邻或直接互相连接。在其他实施例中，第一种和第二β-葡糖苷酶序列并非紧邻，而是通过接头结构域连接。在某些方面，第一或第二β-葡糖苷酶序列包含环区或代表环样结构的序列，其包含约3、4、5、6、7、8、9、10或11个氨基酸残基，所述序列包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT 的序列(SEQ ID NO:172)。在某些方面，第一或第二β-葡糖苷酶序列均不包含环序列。在一些实施例中，接头结构域包含环区，所述环区包含约3、4、5、6、7、8、9、10或11个氨基酸残基，所述环区包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。在一些实施例中，连接第一β-葡糖苷酶序列和第二β-葡糖苷酶序列的接头结构域位于中央(即，并非位于嵌合多肽的N末端或C末端)。在某些方面，嵌合β-葡糖苷酶的N末端序列包含长度至少200、250、300、350、400、450、 500、550或600个残基的来自Tr3B或其变体的序列。在某些方面，N末端序列包含SEQ ID NOs:136-148所代表的多肽序列基序的一种或多种或全部，或优选地包含基序SEQ ID NOs:164-169。在某些方面，C末端序列包含长度至少50、75、100、125、150、175或200个氨基酸残基的来自β- 葡糖苷酶多肽或其变体的序列。在某些方面，C末端序列包含SEQ ID NOs:149-156所代表的多肽序列基序的一种或多种或全部，或优选地包含序列基序SEQ ID NO:170。在某些实施例中，β-葡糖苷酶多肽、其变体或其杂合体或嵌合体还包含一个或多个糖基化位点。一个或多个糖基化位点可以位于C末端序列内部、N末端序列内部或位于两者内部。

[0165] 在某些方面，本发明的非天然存在的纤维素酶或半纤维素酶组合物还包含一种或多种天然存在的半纤维素酶。在某些方面，非天然存在的纤维素酶组合物具有优于天然酶的改善的稳定性，所述天然酶包括衍生嵌合β- 葡糖苷酶的C末端或N末端序列的Tr3B。在某些方面，改善的稳定性包括在储存、表达或生产过程期间蛋白水解稳定性的改善。在某些方面，改善的稳定性包括在储存或生产条件下的酶活性损失的速度或程度的相关降低，其中酶活性损失优选地低于约50％，低于约40％，低于约20％，更优选地低于约15％，或甚至更优选地低于约10％。在某些方面，N末端序列或C末端序列可包含长度约3、4、5、6、7、8、9、10或11个氨基酸残基的环序列，所述环序列包含FDRRSPG的序列(SEQ ID NO:171)或 FD(R/K)YNIT的序列(SEQ ID NO:172)。N末端和C末端序列可以互相紧邻或直接互相连接。在其他方面，N末端序列和C末端序列可以通过接头结构域连接。在某些实施例中，接头结构域包含长度约3、4、5、6、7、8、 9、10或11个氨基酸残基的环序列，所述环序列包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。在某些方面，非天然存在的纤维素酶组合物包含β-葡糖苷酶活性。在某些方面，非天然存在的纤维素酶组合物还包含一种或多种木聚糖酶、β-木糖苷酶和/或L-α- 阿拉伯呋喃糖苷酶活性。

[0166] Te3A

[0167] Te3A的氨基酸序列(SEQ ID NO:66)在图35B和43中示出。Te3A也称为“Abg2”。SEQ ID NO:66是未成熟Te3A的序列。Te3A具有与SEQ ID NO:66的第1位至第19位(以下划线标

出)相对应的预测的信号序列;对该信号序列的切割据预测产生具有与SEQ ID NO:66的第20至857位相对应的序列的成熟蛋白。使用SignalP-NN算法进行信号序列预测。预测的保守结构域在图35B中以粗体标出。结构域预测基于Pfam、SMART或 NCBI数据库进行。Te3A的残基E505和D277据预测分别作为催化性酸碱和亲核物质发挥功能,这是基于上文所述的来自例如以下的GH3葡糖苷酶的序列比对结果:鹅掌柄孢霉(登录号XP_001912683)、大丽轮枝菌、鞭毛藻丛赤壳菌(登录号XP_003045443)、玉米赤霉(登录号 XP_386781)、尖孢镰刀菌(登录号BGL FOXG_02349)、黑曲霉(登录号CAK48740)、埃默森篮状菌(登录号AAL69548)、里氏木霉(登录号AAP57755)、里氏木霉(登录号AAA18473)、轮枝镰刀菌和新阿波罗栖热袍菌(登录号Q0GC07)等等(参见图43)。如本文所用,“Te3A 多肽”在某些方面是指包含以下序列的多肽和/或其变体,所述序列与SEQ ID NO:66的第20位至第857位残基中的至少50、75、100、125、150、 175、200、250、300、350、400、450、500、550、600、650、700、750 或800个连续氨基酸残基具有至少85%、86%、87%、88%、89%、90%、 91%、92%、93%、94%、95%、96%、97%、98%、99%或100%序列同一性。与天然Te3A相比,Te3A多肽优选地在残基E505和D277处未改变。Te3A多肽优选地在本文所述的GH3家族β-葡糖苷酶之间保守的至少70%、80%、90%、95%、98%或99%的氨基酸残基处未改变,如图43的比对结果中示出。Te3A多肽合适地包含图35B中示出的天然Te3A的整个预测保守结构域。示例性的Te3A多肽包含与图35B中示出的成熟Te3A 序列具有至少85%、86%、87%、88%、89%、90%、91%、92%、93%、 94%、95%、96%、97%、98%、99%或100%同一性的序列。本发明的 Te3A多肽优选地具有β-葡糖苷酶活性。

[0168] 因此,本发明的Te3A多肽合适地包含与SEQ ID NO:66的氨基酸序列或与SEQ ID NO:66第(i)20-297、(ii)20-629、(iii)20-857、(iv)396-629或 (v)396-857位残基具有至少80%、85%、90%、91%、92%、93%、94%、 95%、96%、97%、98%、99%或100%序列同一性的氨基酸序列。该多肽合适地具有β-葡糖苷酶活性。

[0169] 在某些方面,本发明的“Te3A多肽”还可以指突变Te3A多肽。可以向Te3A多肽引入氨基酸置换以改善该分子的β-葡糖苷酶活性。例如,可以向Te3A多肽引入氨基酸置换,所述氨基酸置换增强Te3A多肽对其底物的结合亲和力或改善Te3A催化β-D-葡糖苷中的末端非还原性残基水解的能力。在某些方面,突变Te3A多肽包含一种或多种保守氨基酸置换。在某些方面,突变Te3A多肽包含一种或多种非保守的氨基酸置换。在某些方面,一种或多种氨基酸置换位于Te3A多肽的CD中。在某些方面,一种或多种氨基酸置换位于Te3A多肽的CBM中。在某些方面,一种或多种氨基酸置换同时位于CD和CBM中。在某些方面,Te3A多肽的氨基酸置换可以在氨基酸E505和/或D277处发生。在某些方面,Te3A多肽的氨基酸置换可以在氨基酸D92、R98、L141、R156、K189、H190、R200、 M242、Y245、D277、W278、S447和/或E505中的一个或多个处发生。突变Te3A多肽合适地具有β-葡糖苷酶活性。

[0170] 在某些方面,Te3A多肽包含两种β-葡糖苷酶序列的嵌合体/融合体/杂合体,其中第一β-葡糖苷酶序列具有至少约200个氨基酸残基长度,并且包含与Te3A的等长度序列(SEQ ID NO:66)约60%、65%、70%、75%或 80%或更高的序列同一性,并且其中第二β-葡糖苷酶序列具有至少约50个氨基酸残基长度并且包含与SEQ ID NOs:54、56、58、60、62、64、68、70、72、74、76、78和79中任何一种的等长度序列至少约60%、65%、 70%、75%、80%或更高的序列同一性,或包含多肽序列基序SEQ ID NO:170。在某些方面,第一β-葡糖苷酶序

列包含SEQ ID NO:66的至少 200个氨基酸残基的N末端序列，并且第二β-葡糖苷酶序列包含SEQ ID NOs:54、56、58、60、62、64、68、70、72、74、76、78和79中任何一种的至少约50个连续氨基酸残基的C末端序列，或包含多肽序列基序 SEQ ID NO:170。

[0171] 在某些方面，本发明的Te3A多肽包含嵌合体/杂合体/融合体或者两种β-葡糖苷酶序列的嵌合构建体，其中第一β-葡糖苷酶序列具有至少约200 个氨基酸残基长度，并且包含与SEQ ID NOs:54、56、58、60、62、64、68、70、72、74、76、78和79中任何一种的等长度序列约60％、65％、 70％、75％、80％或更高的序列同一性，或包含SEQ ID NOs:164-169的多肽序列基序中的一种或多种或全部，而第二β-葡糖苷酶序列具有至少约50 个氨基酸残基长度，并且包含与Te3A的等长度序列(SEQ ID NO:66)约 60％、65％、70％、75％、80％或更高的序列同一性。在某些方面，第一β- 葡糖苷酶序列包含SEQ ID NOs:54、56、58、60、62、64、68、70、72、 74、76、78和79中任何一种的至少200个氨基酸残基的N末端序列，或包含SEQ ID NOs:164-169多肽序列基序的一种或多种或全部，并且第二β- 葡糖苷酶序列包含SEQ ID NO:66的至少50个连续氨基酸残基的C末端序列。

[0172] 在某些方面，第一β-葡糖苷酶序列位于嵌合β-葡糖苷酶多肽的N末端处，而第二β-葡糖苷酶序列位于嵌合β-葡糖苷酶多肽的C末端处。在某些实施例中，第一、第二或两种β-葡糖苷酶序列还包含一个或多个糖基化位点。在某些实施例中，第一种和第二β-葡糖苷酶序列互相紧邻或直接互相连接。在其他实施例中，第一种和第二β-葡糖苷酶序列并非紧邻，而是通过接头结构域连接。在某些方面，第一或第二β-葡糖苷酶序列包含环区或代表环样结构的序列，其包含约3、4、5、6、7、8、9、10或11个氨基酸残基，所述序列包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT 的序列(SEQ ID NO:172)。在某些方面，第一或第二β-葡糖苷酶序列均不包含环序列。在一些实施例中，接头结构域包含环区，所述环区包含约3、4、5、6、7、8、9、10或11个氨基酸残基，所述环区包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。在一些实施例中，连接第一β-葡糖苷酶序列和第二β-葡糖苷酶序列的接头结构域位于中央(即，并非位于嵌合多肽的N末端或C末端)。在某些方面，嵌合β-葡糖苷酶的N末端序列包含长度至少200、250、300、350、400、450、 500、550或600个残基的来自Te3A多肽或其变体的序列。在某些方面， N末端序列包含SEQ ID NOs:136-148所代表的多肽序列基序的一种或多种或全部，或优选地包含基序SEQ ID NOs:164-169。在某些方面，C末端序列包含长度至少50、75、100、125、150、175或200个氨基酸残基的来自β-葡糖苷酶多肽或其变体的序列。在某些方面，C末端序列包含SEQ ID NOs:149-156所代表的多肽序列基序的一种或多种或全部，或优选地SEQ ID NO:170的基序。在某些实施例中，β-葡糖苷酶多肽、其变体或其杂合体或嵌合体还包含一个或多个糖基化位点。一个或多个糖基化位点可以位于C末端序列内部、N末端序列内部或位于两者内部。

[0173] 在某些方面，本发明的非天然存在的纤维素酶或半纤维素酶组合物还包含一种或多种天然存在的半纤维素酶。在某些方面，非天然存在的纤维素酶组合物具有优于天然酶的改善的稳定性，所述天然酶包括衍生嵌合β- 葡糖苷酶的C末端或N末端序列的Te3A。在某些方面，改善的稳定性包括在储存、表达或生产过程期间蛋白水解稳定性的改善。在某些方面，改善的稳定性包括在储存或生产条件下的酶活性损失的速度或程度的相关降低，其中酶活性损失优选地低于约50％，低于约40％，低于约20％，更优选地低于约15％，或甚至更优选地低于约10％。在某些方面，N末端序列或C末端序列可包含长度约3、4、5、6、7、8、9、10

或11个氨基酸残基的环序列，所述环序列包含FDRRSPG的序列(SEQ ID NO:171)或 FD(R/K)YNIT的序列(SEQ ID NO:172)。N末端和C末端序列可以互相紧邻或直接互相连接。在其他方面，N末端序列和C末端序列可以通过接头结构域连接。在某些实施例中，接头结构域包含长度约3、4、5、6、7、8、 9、10或11个氨基酸残基的环序列，所述环序列包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。在某些方面，非天然存在的纤维素酶组合物包含β-葡糖苷酶活性。在某些方面，非天然存在的纤维素酶组合物还包含一种或多种木聚糖酶、β-木糖苷酶和/或L-α- 阿拉伯呋喃糖苷酶活性。

[0174] An3A

[0175] An3A的氨基酸序列(SEQ ID NO:68)在图36B和43中示出。An3A也称为“黑曲霉Bglu”。SEQ ID NO:68是未成熟An3A的序列。An3A具有与SEQ ID NO:68的第1位至第19位(以下划线标出)相对应的预测的信号序列；对该信号序列的切割据预测产生具有与SEQ ID NO:68的第20至 860位相对应的序列的成熟蛋白。使用SignalP-NN算法进行信号序列预测。预测的保守结构域在图36B中以粗体标出。结构域预测基于Pfam、 SMART或NCBI数据库进行。An3A的残基E509和D277据预测分别作为催化性酸碱和亲核物质发挥功能，这是基于上文所述的来自例如以下的 GH3葡糖苷酶序列比对结果：鹅掌柄孢霉(登录号XP_001912683)、大丽轮枝菌、鞭毛藻丛赤壳菌(登录号XP_003045443)、玉米赤霉(登录号XP_386781)、尖孢镰刀菌(登录号BGL FOXG_02349)、黑曲霉(登录号CAK48740)、埃默森篮状菌(登录号AAL69548)、里氏木霉(登录号AAP57755)、里氏木霉(登录号AAA18473)、轮枝镰刀菌和新阿波罗栖热袍菌(登录号Q0GC07)等等(参见图43)。如本文所用，“An3A多肽”在某些方面是指包含以下序列的多肽和/或其变体，所述序列与SEQ ID NO:68的第20位至第860位残基中的至少50、75、100、 125、150、175、200、250、300、350、400、450、500、550、600、 650、700、750或800个连续氨基酸残基具有至少85％、86％、87％、 88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、 99％或100％序列同一性。与天然An3A相比，An3A多肽优选地在残基 E509和D277处未改变。An3A多肽优选地在本文所述的GH3家族β-葡糖苷酶之间保守的至少70％、80％、90％、95％、98％或99％的氨基酸残基处未改变，如图43的比对结果中示出。An3A多肽合适地包含图36B中示出的天然An3A的整个预测保守结构域。示例性的An3A多肽包含与图36B 中示出的成熟An3A序列具有至少85％、86％、87％、88％、89％、90％、 91％、92％、93％、94％、95％、96％、97％、98％、99％或100％同一性的序列。本发明的An3A多肽优选地具有β-葡糖苷酶活性。

[0176] 因此，本发明的An3A多肽合适地包含与SEQ ID NO:68的氨基酸序列或与SEQ ID NO:68第(i)20-300、(ii)20-634、(iii)20-860、(iv)400-634 或(v)400-860位残基具有至少80％、85％、90％、91％、92％、93％、 94％、95％、96％、97％、98％、99％或100％序列同一性的氨基酸序列。该多肽合适地具有β-葡糖苷酶活性。

[0177] 在某些方面，本发明的“An3A多肽”还可以指突变An3A多肽。可以向An3A多肽引入氨基酸置换以改善该分子的β-葡糖苷酶活性。例如，可以向An3A多肽引入氨基酸置换，所述氨基酸置换增强An3A多肽对其底物的结合亲和力或改善An3A催化β-D-葡糖苷中的末端非还原性残基水解的能力。在某些方面，突变An3A多肽包含一种或多种保守氨基酸置换。在某些方面，突变An3A多肽包含一种或多种非保守的氨基酸置换。在某些方面，一种或多种氨基酸置换位于An3A多肽的CD中。在某些方面，一种或多种氨基酸置换位于An3A多肽的CBM中。

在某些方面，一种或多种氨基酸置换同时位于CD和CBM中。在某些方面，An3A多肽的氨基酸置换可以在氨基酸E509和/或D277处发生。在某些方面，An3A多肽的氨基酸置换可以在氨基酸D92、R98、L141、R156、K189、H190、 R200、M245、Y248、D277、W278、S451和/或E509中的一个或多个处发生。突变An3A多肽合适地具有β-葡糖苷酶活性。

[0178] 在某些方面，An3A多肽包含两种β-葡糖苷酶序列的嵌合体/杂合体/融合体，其中第一β-葡糖苷酶序列具有至少约200个氨基酸残基长度，并且包含与An3A的等长度序列(SEQ ID NO:68)约60％、65％、70％、75％或 80％或更高的序列同一性，并且其中第二β-葡糖苷酶序列具有至少约50个氨基酸残基长度并且包含与SEQ ID NOs:54、56、58、60、62、64、66、 70、72、74、76、78和79中任何一种的等长度序列至少约60％、65％、 70％、75％、80％或更高的序列同一性，或包含多肽序列基序SEQ ID NO:170。在某些方面，第一β-葡糖苷酶序列包含SEQ ID NO:68的至少 200个氨基酸残基的N末端序列，并且第二β-葡糖苷酶序列包含SEQ ID NOs:54、56、58、60、62、64、66、70、72、74、76、78和79中任何一种的至少约50个连续氨基酸残基的C末端序列，或包含多肽序列基序 SEQ ID NO:170。

[0179] 在某些方面，本发明的An3A多肽包含两种β-葡糖苷酶序列的嵌合体或嵌合构建体，其中第一β-葡糖苷酶序列具有至少约200个氨基酸残基长度，并且包含与SEQ ID NOs:54、56、58、60、62、64、66、70、72、 74、76、78和79中任何一种的等长度序列约60％、65％、70％、75％、 80％或更高的序列同一性，或包含SEQ ID NOs:164-169的多肽序列基序中的一种或多种或全部，而第二β-葡糖苷酶序列具有至少约50个氨基酸残基长度，并且包含与An3A的等长度序列(SEQ ID NO:68)约60％、65％、 70％、75％、80％或更高的序列同一性。在某些方面，第一β-葡糖苷酶序列包含SEQ ID NOs:54、56、58、60、62、64、66、70、72、74、76、78 和79中任何一种的至少200个氨基酸残基的N末端序列，或包含SEQ ID NOs:164-169多肽序列基序的一种或多种或全部，并且第二β-葡糖苷酶序列包含SEQ ID NO:68的至少50个连续氨基酸残基的C末端序列。

[0180] 在某些方面，第一β-葡糖苷酶序列位于嵌合β-葡糖苷酶多肽的N末端处，而第二β-葡糖苷酶序列位于嵌合β-葡糖苷酶多肽的C末端处。在某些实施例中，第一、第二或两种β-葡糖苷酶序列还包含一个或多个糖基化位点。在某些实施例中，第一种和第二β-葡糖苷酶序列互相紧邻或直接互相连接。在其他实施例中，第一种和第二β-葡糖苷酶序列并非紧邻，而是通过接头结构域连接。在某些方面，第一或第二β-葡糖苷酶序列包含环区或代表环样结构的序列，其包含约3、4、5、6、7、8、9、10或11个氨基酸残基，所述序列包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT 的序列(SEQ ID NO:172)。在某些方面，第一或第二β-葡糖苷酶序列均不包含环序列。在一些实施例中，接头结构域包含环区，所述环区包含约3、4、5、6、7、8、9、10或11个氨基酸残基，所述环区包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。在一些实施例中，连接第一β-葡糖苷酶序列和第二β-葡糖苷酶序列的接头结构域位于中央(即，并非位于嵌合多肽的N末端或C末端)。在某些方面，嵌合β-葡糖苷酶的N末端序列包含长度至少200、250、300、350、400、450、 500、550或600个残基的来自An3A多肽或其变体的序列。在某些方面， N末端序列包含SEQ ID NOs:136-148所代表的多肽序列基序的一种或多种或全部，优选地包含基序SEQ ID NOs:164-169。在某些方面，C末端序列包含长度至少50、75、100、125、150、175或200个氨基酸残基的来自β-葡糖苷酶多肽或其变体的序列。在某些方面，C末端序列包含SEQ ID NOs:149-156所代表

的多肽序列基序的一种或多种或全部，优选地包含基序SEQ ID NO:170。在某些实施例中，β-葡糖苷酶多肽、其变体或其杂合体或嵌合体还包含一个或多个糖基化位点。一个或多个糖基化位点可以位于C末端序列内部、N末端序列内部或位于两者内部。

[0181] 在某些方面，本发明的非天然存在的纤维素酶或半纤维素酶组合物还包含一种或多种天然存在的半纤维素酶。在某些方面，非天然存在的纤维素酶组合物具有优于天然酶的改善的稳定性，所述天然酶包括衍生嵌合β- 葡糖苷酶的C末端或N末端序列的An3A。在某些方面，改善的稳定性包括在储存、表达或生产过程期间蛋白水解稳定性的改善。在某些方面，改善的稳定性包括在储存或生产条件下的酶活性损失的速度或程度的相关降低，其中酶活性损失优选地低于约50％，低于约40％，低于约20％，更优选地低于约15％，或甚至更优选地低于约10％。在某些方面，N末端序列或C末端序列可包含长度约3、4、5、6、7、8、9、10或11个氨基酸残基的环序列，所述环序列包含FDRRSPG的序列(SEQ ID NO:171)或 FD(R/K)YNIT的序列(SEQ ID NO:172)。N末端和C末端序列可以互相紧邻或直接互相连接。在其他方面，N末端序列和C末端序列可以通过接头结构域连接。在某些实施例中，接头结构域包含长度约3、4、5、6、7、8、 9、10或11个氨基酸残基的环序列，所述环序列包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。在某些方面，非天然存在的纤维素酶组合物包含β-葡糖苷酶活性。在某些方面，非天然存在的纤维素酶组合物还包含一种或多种木聚糖酶、β-木糖苷酶和/或L-α- 阿拉伯呋喃糖苷酶活性。

[0182] Fo3A

[0183] Fo3A的氨基酸序列(SEQ ID NO:70)在图37B和43中示出。SEQ ID NO:70是未成熟Fo3A的序列。Fo3A具有与SEQ ID NO:70的第1位至第 19位(以下划线标出)相对应的预测的信号序列；对该信号序列的切割据预测产生具有与SEQ ID NO:70的第20至899位相对应的序列的成熟蛋白。使用SignalP-NN算法进行信号序列预测。预测的保守结构域在图37B 中以粗体标出。结构域预测基于Pfam、SMART或NCBI数据库进行。Fo3A残基E536和D307据预测分别作为催化性酸碱和亲核物质发挥功能，这是基于上文所述的来自例如以下的GH3葡糖苷酶的序列比对结果：鹅掌柄孢霉(登录号XP_001912683)、大丽轮枝菌、鞭毛藻丛赤壳菌(登录号XP_003045443)、玉米赤霉(登录号XP_386781)、尖孢镰刀菌(登录号BGL FOXG_02349)、黑曲霉(登录号CAK48740)、埃默森篮状菌(登录号AAL69548)、里氏木霉(登录号AAP57755)、里氏木霉(登录号AAA18473)、轮枝镰刀菌和新阿波罗栖热袍菌(登录号Q0GC07)等等(参见图43)。如本文所用，“Fo3A多肽”在某些方面是指包含以下序列的多肽和/或其变体，所述序列与SEQ ID NO:70第20位至第899位残基中的至少50、75、100、125、150、175、200、250、300、 350、400、450、500、550、600、650、700、750、800或850个连续氨基酸残基具有至少85％、86％、87％、88％、89％、90％、91％、92％、 93％、94％、95％、96％、97％、98％、99％或100％序列同一性。与天然 Fo3A相比，Fo3A多肽优选地在残基E536和D307处未改变。Fo3A多肽优选地在本文所述的GH3家族β-葡糖苷酶之间保守的至少70％、80％、 90％、95％、98％或99％的氨基酸残基处未改变，如图43的比对结果中示出。Fo3A多肽合适地包含图37B中示出的天然Fo3A的整个预测保守结构域。示例性的Fo3A多肽包含与图37B中示出的成熟Fo3A序列具有至少 85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、 96％、97％、98％、99％或100％同一性的序列。本发明的Fo3A多肽优选地具有β-葡糖苷酶活性。

[0184] 因此，本发明的Fo3A多肽合适地包含与SEQ ID NO:70的氨基酸序列或与SEQ ID NO:70第(i)20-327、(ii)20-660、(iii)20-899、(iv)428-660或 (v)428-899位残基具有至少80％、85％、90％、91％、92％、93％、94％、 95％、96％、97％、98％、99％或100％序列同一性的氨基酸序列。该多肽合适地具有β-葡糖苷酶活性。

[0185] 在某些方面，本发明的“Fo3A多肽”还可以指突变Fo3A多肽。可以向Fo3A多肽引入氨基酸置换以改善该分子的β-葡糖苷酶活性。例如，可以向Fo3A多肽引入氨基酸置换，所述氨基酸置换增强Fo3A多肽对其底物的结合亲和力或改善Fo3A催化β-D-葡糖苷中的末端非还原性残基水解的能力。在某些方面，突变Fo3A多肽包含一种或多种保守氨基酸置换。在某些方面，突变Fo3A多肽包含一种或多种非保守的氨基酸置换。在某些方面，一种或多种氨基酸置换位于Fo3A多肽的CD中。在某些方面，一种或多种氨基酸置换位于Fo3A多肽的CBM中。在某些方面，一种或多种氨基酸置换同时位于CD和CBM中。在某些方面，Fo3A多肽的氨基酸置换可以在氨基酸E536和/或D307处发生。在某些方面，Fo3A多肽的氨基酸置换可以在氨基酸D119、R125、L168、R183、K216、H217、R227、 M272、Y275、D307、W308、S477和/或E536中的一个或多个处发生。突变Fo3A多肽合适地具有β-葡糖苷酶活性。

[0186] 在某些方面，Fo3A多肽包含两种β-葡糖苷酶序列的嵌合体/杂合体/融合体，其中第一β-葡糖苷酶序列具有至少约200个氨基酸残基长度，并且包含与Fo3A的等长度序列(SEQ ID NO:70)约60％、65％、70％、75％或 80％或更高的序列同一性，并且其中第二β-葡糖苷酶序列具有至少约50个氨基酸残基长度并且包含与SEQ ID NOs:54、56、58、60、62、64、66、 68、72、74、76、78和79中任何一种的等长度序列至少约60％、65％、 70％、75％、80％或更高的序列同一性，或包含多肽序列基序SEQ ID NO:170。在某些方面，第一β-葡糖苷酶序列包含SEQ ID NO:70的至少 200个氨基酸残基的N末端序列，并且第二β-葡糖苷酶序列包含SEQ ID NOs:54、56、58、60、62、64、66、68、72、74、76、78和79中任何一种的至少约50个连续氨基酸残基的C末端序列，或包含多肽序列基序 SEQ ID NO:170。

[0187] 在某些方面，本发明的Fo3A多肽包含两种β-葡糖苷酶序列的嵌合体或嵌合构建体，其中第一β-葡糖苷酶序列具有至少约200个氨基酸残基长度，并且包含与SEQ ID NOs:54、56、58、60、62、64、66、68、72、 74、76、78和79中任何一种的等长度序列约60％、65％、70％、75％、 80％或更高的序列同一性，或包含SEQ ID NOs:164-169的多肽序列基序中的一种或多种或全部，而第二β-葡糖苷酶序列具有至少约50个氨基酸残基长度，并且包含与Fo3A的等长度序列(SEQ ID NO:70)约60％、65％、 70％、75％、80％或更高的序列同一性。在某些方面，第一β-葡糖苷酶序列包含SEQ ID NOs:54、56、58、60、62、64、66、68、72、74、76、78 和79中任何一种的至少200个氨基酸残基的N末端序列，或包含SEQ ID NOs:164-169多肽序列基序的一种或多种或全部，并且第二β-葡糖苷酶序列包含SEQ ID NO:70的至少50个连续氨基酸残基的C末端序列。

[0188] 在某些方面，第一β-葡糖苷酶序列位于嵌合β-葡糖苷酶多肽的N末端处，而第二β-葡糖苷酶序列位于嵌合β-葡糖苷酶多肽的C末端处。在某些实施例中，第一、第二或两种β-葡糖苷酶序列还包含一个或多个糖基化位点。在某些实施例中，第一种和第二β-葡糖苷酶序列互相紧邻或直接互相连接。在其他实施例中，第一种和第二β-葡糖苷酶序列并非紧邻，而是通过接头结构域连接。在某些方面，第一或第二β-葡糖苷酶序列包含环区或代表环样结构的序列，其包含约3、4、5、6、7、8、9、10或11个氨基酸残基，所述序列包含FDRRSPG的序列

(SEQ ID NO:171)或FD(R/K)YNIT 的序列(SEQ ID NO:172)。在某些方面,第一或第二β-葡糖苷酶序列均不包含环序列。在一些实施例中,接头结构域包含环区,所述环区包含约3、4、5、6、7、8、9、10或11个氨基酸残基,所述环区包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。在一些实施例中,连接第一β-葡糖苷酶序列和第二β-葡糖苷酶序列的接头结构域位于中央(即,并非位于嵌合多肽的N末端或C末端)。在某些方面,嵌合β-葡糖苷酶的N末端序列包含长度至少200、250、300、350、400、450、 500、550或600个残基的来自Fo3A多肽或其变体的序列。在某些方面, N末端序列包含SEQ ID NOs:136-148所代表的多肽序列基序的一种或多种或全部,优选地包含基序SEQ ID NOs:164-169。在某些方面,C末端序列包含长度至少50、75、100、125、150、175或200个氨基酸残基的来自β-葡糖苷酶多肽或其变体的序列。在某些方面,C末端序列包含SEQ ID NOs:149-156所代表的多肽序列基序的一种或多种或全部,优选地包含基序SEQ ID NO:170。在某些实施例中,β-葡糖苷酶多肽、其变体或其杂合体或嵌合体还包含一个或多个糖基化位点。一个或多个糖基化位点可以位于C末端序列内部、N末端序列内部或位于两者内部。

[0189] 在某些方面,本发明的非天然存在的纤维素酶或半纤维素酶组合物还包含一种或多种天然存在的半纤维素酶。在某些方面,非天然存在的纤维素酶组合物具有优于天然酶的改善的稳定性,所述天然酶包括衍生嵌合β- 葡糖苷酶的C末端或N末端序列的Fo3A。在某些方面,改善的稳定性包括在储存、表达或生产过程期间蛋白水解稳定性的改善。在某些方面,改善的稳定性包括在储存或生产条件下的酶活性损失的速度或程度的相关降低,其中酶活性损失优选地低于约50%,低于约40%,低于约20%,更优选地低于约15%,或甚至更优选地低于约10%。在某些方面,N末端序列或C末端序列可包含长度约3、4、5、6、7、8、9、10或11个氨基酸残基的环序列,所述环序列包含FDRRSPG的序列(SEQ ID NO:171)或 FD(R/K)YNIT的序列(SEQ ID NO:172)。N末端和C末端序列可以互相紧邻或直接互相连接。在其他方面,N末端序列和C末端序列可以通过接头结构域连接。在某些实施例中,接头结构域包含长度约3、4、5、6、7、8、 9、10或11个氨基酸残基的环序列,所述环序列包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。在某些方面,非天然存在的纤维素酶组合物包含β-葡糖苷酶活性。在某些方面,非天然存在的纤维素酶组合物还包含一种或多种木聚糖酶、β-木糖苷酶和/或L-α- 阿拉伯呋喃糖苷酶活性。

[0190] Gz3A

[0191] Gz3A的氨基酸序列(SEQ ID NO:72)在图38B和43中示出。SEQ ID NO:72是未成熟Gz3A的序列。Gz3A具有与SEQ ID NO:72的第1位至第 18位(以下划线标出)相对应的预测的信号序列;对该信号序列的切割据预测产生具有与SEQ ID NO:72的第19至886位相对应的序列的成熟蛋白。使用SignalP-NN算法进行信号序列预测。预测的保守结构域在图38B 中以粗体标出。结构域预测基于Pfam、SMART或NCBI数据库进行。Gz3A的残基E523和D294据预测分别作为催化性酸碱和亲核物质发挥功能,这是基于上文所述的来自例如以下的GH3葡糖苷酶的序列比对结果:鹅掌柄孢霉(登录号XP_001912683)、大丽轮枝菌、鞭毛藻丛赤壳菌(登录号XP_003045443)、玉米赤霉(登录号XP_386781)、尖孢镰刀菌 (登录号BGL FOXG_02349)、黑曲霉(登录号CAK48740)、埃默森篮状菌(登录号AAL69548)、里氏木霉(登录号AAP57755)、里氏木霉(登录号AAA18473)、轮枝镰刀菌和新阿波罗栖热袍菌(登录号Q0GC07)等等(参见图43)。如本文所用,“Gz3A多肽”在某些方面是指包含以下序列的多肽和/或其变

体，所述序列与SEQ ID NO:72的第19位至第886位残基中的至少50、75、100、125、150、175、200、250、300、350、400、 450、500、550、600、650、700、750、800或850个连续氨基酸残基具有至少85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、 95％、96％、97％、98％、99％或100％序列同一性。与天然Gz3A相比， Gz3A多肽优选地在残基E536和D307处未改变。Gz3A多肽优选地在本文所述的GH3家族β-葡糖苷酶之间保守的至少70％、80％、90％、95％、98％或99％的氨基酸残基处未改变，如图43的比对结果中示出。Gz3A多肽合适地包含图38B中示出的天然Gz3A的整个预测保守结构域。示例性的Gz3A多肽包含与图38B中示出的成熟Gz3A序列具有至少85％、 86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％同一性的序列。本发明的Gz3A多肽优选地具有β-葡糖苷酶活性。

[0192] 因此，本发明的Gz3A多肽合适地包含与SEQ ID NO:72的氨基酸序列或与SEQ ID NO:72第(i)19-314、(ii)19-647、(iii)19-886、(iv)415-647 或(v)415-886位残基具有至少80％、85％、90％、91％、92％、93％、 94％、95％、96％、97％、98％、99％或100％序列同一性的氨基酸序列。该多肽合适地具有β-葡糖苷酶活性。

[0193] 在某些方面，本发明的“Gz3A多肽”还可以指突变Gz3A多肽。可以向Gz3A多肽引入氨基酸置换以改善该分子的β-葡糖苷酶活性。例如，可以向Gz3A多肽引入氨基酸置换，所述氨基酸置换Gz3A多肽对其底物的结合亲和力或改善Gz3A催化β-D-葡糖苷中的末端非还原性残基水解的能力。在某些方面，突变Gz3A多肽包含一种或多种保守氨基酸置换。在某些方面，突变Gz3A多肽包含一种或多种非保守的氨基酸置换。在某些方面，一种或多种氨基酸置换位于Gz3A多肽的CD中。在某些方面，一种或多种氨基酸置换位于Gz3A多肽的CBM中。在某些方面，一种或多种氨基酸置换同时位于CD和CBM中。在某些方面，Gz3A多肽的氨基酸置换可以在氨基酸E536和/或D307处发生。在某些方面，Gz3A多肽的氨基酸置换可以在氨基酸D106、R112、L155、R170、K203、H204、R214、M259、Y262、D294、W295、S464和/或E523中的一个或多个处发生。突变Gz3A多肽合适地具有β-葡糖苷酶活性。

[0194] 在某些方面，Gz3A多肽包含两种β-葡糖苷酶序列的嵌合体/融合体/杂合体，其中第一β-葡糖苷酶序列具有至少约200个氨基酸残基长度，并且包含与Gz3A的等长度序列(SEQ ID NO:72)约60％、65％、70％、75％或 80％或更高的序列同一性，并且其中第二β-葡糖苷酶序列具有至少约50个氨基酸残基长度并且包含与SEQ ID NOs:54、56、58、60、62、64、66、 68、70、74、76、78和79中任何一种的等长度序列至少约60％、65％、 70％、75％、80％或更高的序列同一性，或包含多肽序列基序SEQ ID NO:170。在某些方面，第一β-葡糖苷酶序列包含SEQ ID NO:72的至少 200个氨基酸残基的N末端序列，并且第二β-葡糖苷酶序列包含SEQ ID NOs:54、56、58、60、62、64、66、68、70、74、76、78和79中任何一种的至少约50个连续氨基酸残基的C末端序列，或包含多肽序列基序 SEQ ID NO:170。

[0195] 在某些方面，本发明的Gz3A多肽包含两种β-葡糖苷酶序列的嵌合体或嵌合构建体，其中第一β-葡糖苷酶序列具有至少约200个氨基酸残基长度，并且包含与SEQ ID NOs:54、56、58、60、62、64、66、68、70、 74、76、78和79中任何一种的等长度序列约60％、65％、70％、75％、 80％或更高的序列同一性，或包含SEQ ID NOs:164-169多肽序列基序的一种或多种或全部，而第二β-葡糖苷酶序列具有至少约50个氨基酸残基长度并且包含与Gz3A的等长度序列(SEQ ID NO:72)约60％、65％、70％、 75％、80％或更高的序列同一性。在某些方面，第一β-葡糖苷酶序列包含 SEQ ID NOs:54、56、58、60、62、64、66、68、70、74、76、78和

79中任何一种的至少200个氨基酸残基的N末端序列，或者包含SEQ ID NOs: 164-169的氨基酸序列基序的一种或多种或全部，并且第二β-葡糖苷酶序列包含SEQ ID NO:72的至少50个连续氨基酸残基的C末端序列。

[0196] 在某些方面，第一β-葡糖苷酶序列位于嵌合β-葡糖苷酶多肽的N末端处，而第二β-葡糖苷酶序列位于嵌合β-葡糖苷酶多肽的C末端处。在某些实施例中，第一、第二或两种β-葡糖苷酶序列还包含一个或多个糖基化位点。在某些实施例中，第一种和第二β-葡糖苷酶序列互相紧邻或直接互相连接。在其他实施例中，第一种和第二β-葡糖苷酶序列并非紧邻，而是通过接头结构域连接。在某些方面，第一或第二β-葡糖苷酶序列包含环区或代表环样结构的序列，其包含约3、4、5、6、7、8、9、10或11个氨基酸残基，所述序列包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT 的序列(SEQ ID NO:172)。在某些方面，第一或第二β-葡糖苷酶序列均不包含环序列。在一些实施例中，接头结构域包含环区，所述环区包含约3、4、5、6、7、8、9、10或11个氨基酸残基，所述环区包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。在一些实施例中，连接第一β-葡糖苷酶序列和第二β-葡糖苷酶序列的接头结构域位于中央(即，并非位于嵌合多肽的N末端或C末端)。在某些方面，嵌合β-葡糖苷酶的N末端序列包含长度至少200、250、300、350、400、450、 500、550或600个残基的来自Gz3A多肽或其变体的序列。在某些方面， N末端序列包含SEQ ID NOs:136-148所代表的多肽序列基序的一种或多种或全部，优选地包含序列基序SEQ ID NOs:164-169。在某些方面，C末端序列包含长度至少50、75、100、125、150、175或200个氨基酸残基的来自β-葡糖苷酶多肽或其变体的序列。在某些方面，C末端序列包含SEQ ID NOs:149-156所代表的多肽序列基序的一种或多种或全部，或优选地包含序列基序SEQ ID NO:170。在某些实施例中，β-葡糖苷酶多肽、其变体或其杂合体或嵌合体还包含一个或多个糖基化位点。一个或多个糖基化位点可以位于C末端序列内部、N末端序列内部或位于两者内部。

[0197] 在某些方面，本发明的非天然存在的纤维素酶或半纤维素酶组合物还包含一种或多种天然存在的半纤维素酶。在某些方面，非天然存在的纤维素酶组合物具有优于天然酶的改善的稳定性，所述天然酶包括衍生嵌合β- 葡糖苷酶的C末端或N末端序列的Gz3A。在某些方面，改善的稳定性包括在储存、表达或生产过程期间蛋白水解稳定性的改善。在某些方面，改善的稳定性包括在储存或生产条件下的酶活性损失的速度或程度的相关降低，其中酶活性损失优选地低于约50％，低于约40％，低于约20％，更优选地低于约15％，或甚至更优选地低于约10％。在某些方面，N末端序列或C末端序列可包含长度约3、4、5、6、7、8、9、10或11个氨基酸残基的环序列，所述环序列包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。N末端和C末端序列可以互相紧邻或直接互相连接。在其他方面，N末端序列和C末端序列可以通过接头结构域连接。在某些实施例中，接头结构域包含长度约3、4、5、6、7、8、 9、10或11个氨基酸残基的环序列，所述环序列包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。在某些方面，非天然存在的纤维素酶组合物包含β-葡糖苷酶活性。在某些方面，非天然存在的纤维素酶组合物还包含一种或多种木聚糖酶、β-木糖苷酶和/或L-α- 阿拉伯呋喃糖苷酶活性。

[0198] Nh3A

[0199] Nh3A的氨基酸序列(SEQ ID NO:74)在图39B和43中示出。SEQ ID NO:74是未成熟Nh3A的序列。Nh3A具有与SEQ ID NO:74的第1位至第 19位(以下划线标出)相对应的预测的

信号序列;对该信号序列的切割据预测产生具有与SEQ ID NO:74的第20至880位相对应的序列的成熟蛋白。使用SignalP-NN算法进行信号序列预测。预测的保守结构域在图39B 中以粗体标出。结构域预测基于Pfam、SMART或NCBI数据库进行。 Nh3A的残基E523和D294据预测分别作为催化性酸碱和亲核物质发挥功能,这是基于上文所述的来自例如以下的GH3葡糖苷酶序列比对结果:鹅掌柄孢霉(登录号XP_001912683)、大丽轮枝菌、鞭毛藻丛赤壳菌(登录号XP_003045443)、玉米赤霉(登录号XP_386781)、尖孢镰刀菌 (登录号BGL FOXG_02349)、黑曲霉(登录号CAK48740)、埃默森篮状菌(登录号AAL69548)、里氏木霉(登录号AAP57755)、里氏木霉 (登录号AAA18473)、轮枝镰刀菌和新阿波罗栖热袍菌(登录号Q0GC07)等等(参见图43)。如本文所用,“Nh3A多肽”在某些方面是指包含以下序列的多肽和/或其变体,所述序列与SEQ ID NO:74的第20 位至第880位残基中的至少50、75、100、125、150、175、200、250、 300、350、400、450、500、550、600、650、700、750、800或850个连续氨基酸残基具有至少85%、86%、87%、88%、89%、90%、91%、 92%、93%、94%、95%、96%、97%、98%、99%或100%序列同一性。与天然Nh3A相比,Nh3A多肽优选地在残基E523和D294处未改变。Nh3A 多肽优选地在本文所述的GH3家族β-葡糖苷酶之间保守的至少70%、80%、90%、95%、98%或99%的氨基酸残基处未改变,如图43的比对结果中示出。Nh3A多肽合适地包含图39B中示出的天然Nh3A的整个预测保守结构域。示例性的Nh3A多肽包含与图39B中示出的成熟Nh3A序列具有至少85%、86%、87%、88%、89%、90%、91%、92%、93%、 94%、95%、96%、97%、98%、99%或100%同一性的序列。本发明的 Nh3A多肽优选地具有β-葡糖苷酶活性。

[0200] 因此,本发明的Nh3A多肽合适地包含与SEQ ID NO:74的氨基酸序列或与SEQ ID NO:74第(i)20-295、(ii)20-647、(iii)20-880、(iv)414-647 或(v)414-880位残基具有至少80%、85%、90%、91%、92%、93%、 94%、95%、96%、97%、98%、99%或100%序列同一性的氨基酸序列。该多肽合适地具有β-葡糖苷酶活性。

[0201] 在某些方面,本发明的“Nh3A多肽”还可以指突变Nh3A多肽。可以向Nh3A多肽引入氨基酸置换以改善该分子的β-葡糖苷酶活性。例如,可以向Nh3A多肽引入氨基酸置换,所述氨基酸置换增强Nh3A多肽对其底物的结合亲和力或改善Nh3A催化β-D-葡糖苷中的末端非还原性残基水解的能力。在某些方面,突变Nh3A多肽包含一种或多种保守氨基酸置换。在某些方面,突变Nh3A多肽包含一种或多种非保守的氨基酸置换。在某些方面,一种或多种氨基酸置换位于Nh3A多肽的CD中。在某些方面,一种或多种氨基酸置换位于Nh3A多肽的CBM中。在某些方面,一种或多种氨基酸置换同时位于CD和CBM中。在某些方面,Nh3A多肽的氨基酸置换可以在氨基酸E523和/或D294处发生。在某些方面,Nh3A多肽的氨基酸置换可以在氨基酸D106、R112、L155、R170、K203、H204、 R214、M259、Y262、D294、W295、S464和/或E523中的一个或多个处发生。突变Nh3A多肽合适地具有β-葡糖苷酶活性。

[0202] 在某些方面,Nh3A多肽包含两种β-葡糖苷酶序列的嵌合体/融合体/杂合体,其中第一β-葡糖苷酶序列具有至少约200个氨基酸残基长度,并且包含与Nh3A的等长度序列(SEQ ID NO:74)约60%、65%、70%、75%或 80%或更高的序列同一性,并且其中第二β-葡糖苷酶序列具有至少约50个氨基酸残基长度并且包含与SEQ ID NOs:54、56、58、60、62、64、66、 68、70、72、76、78和79中任何一种的等长度序列至少约60%、65%、70%、75%、80%或更高的序列同一性,或包含多肽序列基序SEQ ID NO:170。在某些方面,第一β-葡糖苷酶序

列包含SEQ ID NO:74的至少 200个氨基酸残基的N末端序列，并且第二β-葡糖苷酶序列包含SEQ ID NOs:54、56、58、60、62、64、66、68、70、72、76、78和79中任何一种的至少约50个连续氨基酸残基的C末端序列，或包含多肽序列基序 SEQ ID NO:170。

[0203] 在某些方面，本发明的Nh3A多肽包含两种β-葡糖苷酶序列的嵌合体或嵌合构建体，其中第一β-葡糖苷酶序列具有至少约200个氨基酸残基长度，并且包含与SEQ ID NOs:54、56、58、60、62、64、66、68、70、 72、76、78和79中任何一种的等长度序列约60％、65％、70％、75％、 80％或更高的序列同一性，或包含SEQ ID NOs:164-169多肽序列基序的一种或多种或全部，而第二β-葡糖苷酶序列具有至少约50个氨基酸残基长度并且包含与Nh3A的等长度序列(SEQ ID NO:74)约60％、65％、70％、 75％、80％或更高的序列同一性。在某些方面，第一β-葡糖苷酶序列包含 SEQ ID NOs:54、56、58、60、62、64、66、68、70、72、76、78和79中任何一种的至少200个氨基酸残基的N末端序列，或者包含SEQ ID NOs: 164-169多肽序列基序的一种或多种或全部，并且第二β-葡糖苷酶序列包含SEQ ID NO:74的至少50个连续氨基酸残基的C末端序列。

[0204] 在某些方面，第一β-葡糖苷酶序列位于嵌合β-葡糖苷酶多肽的N末端处，而第二β-葡糖苷酶序列位于嵌合β-葡糖苷酶多肽的C末端处。在某些实施例中，第一、第二或两种β-葡糖苷酶序列还包含一个或多个糖基化位点。在某些实施例中，第一种和第二β-葡糖苷酶序列互相紧邻或直接互相连接。在其他实施例中，第一种和第二β-葡糖苷酶序列并非紧邻，而是通过接头结构域连接。在某些方面，第一或第二β-葡糖苷酶序列包含环区或代表环样结构的序列，其包含约3、4、5、6、7、8、9、10或11个氨基酸残基，所述序列包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT 的序列(SEQ ID NO:172)。在某些方面，第一或第二β-葡糖苷酶序列均不包含环序列。在一些实施例中，接头结构域包含环区，所述环区包含约3、4、5、6、7、8、9、10或11个氨基酸残基，所述环区包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。在一些实施例中，连接第一β-葡糖苷酶序列和第二β-葡糖苷酶序列的接头结构域位于中央(即，并非位于嵌合多肽的N末端或C末端)。在某些方面，嵌合β-葡糖苷酶的N末端序列包含长度至少200、250、300、350、400、450、 500、550或600个残基的来自Nh3A多肽或其变体的序列。在某些方面， N末端序列包含SEQ ID NOs:136-148所代表的多肽序列基序的一种或多种或全部，优选地包含序列基序SEQ ID NOs:164-169。在某些方面，C末端序列包含长度至少50、75、100、125、150、175或200个氨基酸残基的来自β-葡糖苷酶多肽或其变体的序列。在某些方面，C末端序列包含SEQ ID NOs:149-156所代表的多肽序列基序的一种或多种或全部，或优选地包含序列基序SEQ ID NO:170。在某些实施例中，β-葡糖苷酶多肽、其变体或其杂合体或嵌合体还包含一个或多个糖基化位点。一个或多个糖基化位点可以位于C末端序列内部、N末端序列内部或位于两者内部。

[0205] 在某些方面，本发明的非天然存在的纤维素酶或半纤维素酶组合物还包含一种或多种天然存在的半纤维素酶。在某些方面，非天然存在的纤维素酶组合物具有优于天然酶的改善的稳定性，所述天然酶包括衍生嵌合β- 葡糖苷酶的C末端或N末端序列的Nh3A。在某些方面，改善的稳定性包括在储存、表达或生产过程期间蛋白水解稳定性的改善。在某些方面，改善的稳定性包括在储存或生产条件下的酶活性损失的程度或速度的相关降低，其中酶活性损失优选地低于约50％，低于约40％，低于约20％，更优选地低于约15％，或甚至更优选地低于约10％。在某些方面，N末端序列或C末端序列可包含长度约3、4、5、6、7、8、9、10

或11个氨基酸残基的环序列，所述环序列包含FDRRSPG的序列(SEQ ID NO:171)或 FD(R/K)YNIT的序列(SEQ ID NO:172)。N末端和C末端序列可以互相紧邻或直接互相连接。在其他方面，N末端序列和C末端序列可以通过接头结构域连接。在某些实施例中，接头结构域包含长度约3、4、5、6、7、8、 9、10或11个氨基酸残基的环序列，所述环序列包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。在某些方面，非天然存在的纤维素酶组合物包含β-葡糖苷酶活性。在某些方面，非天然存在的纤维素酶组合物还包含一种或多种木聚糖酶、β-木糖苷酶和/或L-α- 阿拉伯呋喃糖苷酶活性。

[0206] Vd3A

[0207] Vd3A的氨基酸序列(SEQ ID NO:76)在图40B和43中示出。SEQ ID NO:76是未成熟Vd3A的序列。Vd3A具有与SEQ ID NO:76的第1位至第 18位(以下划线标出)相对应的预测的信号序列；对该信号序列的切割据预测产生具有与SEQ ID NO:76的第19至890位相对应的序列的成熟蛋白。使用SignalP-NN算法进行信号序列预测。预测的保守结构域在图40B 中以粗体标出。结构域预测基于Pfam、SMART或NCBI数据库进行。显示Vd3A在例如使用cNPG和纤维二糖的酶分析法中以及在稀氨水预处理的玉米芯作为底物的水解中具有β-葡糖苷酶活性。Vd3A的残基E524和 D295据预测分别作为催化性酸碱和亲核物质发挥功能，这是基于上文所述的来自例如以下的GH3葡糖苷酶的序列比对结果：鹅掌柄孢霉(登录号 XP_001912683)、大丽轮枝菌、鞭毛藻丛赤壳菌(登录号 XP_003045443)、玉米赤霉(登录号XP_386781)、尖孢镰刀菌(登录号 BGL FOXG_02349)、黑曲霉(登录号CAK48740)、埃默森篮状菌(登录号AAL69548)、里氏木霉(登录号AAP57755)、里氏木霉(登录号 AAA18473)、轮枝镰刀菌和新阿波罗栖热袍菌(登录号Q0GC07)等等 (参见图43)。如本文所用，“Vd3A多肽”在某些方面是指包含以下序列的多肽和/或其变体，所述序列与SEQ ID NO:76的残基19至890中的至少50、75、100、125、150、175、200、250、300、350、400、450、 500、550、600、650、700、750、800或850个连续氨基酸残基具有至少 85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、 96％、97％、98％、99％或100％序列同一性。与天然Vd3A相比，Vd3A多肽优选地在残基E524和D295处未改变。Vd3A多肽优选地在本文所述的 GH3家族β-葡糖苷酶之间保守的至少70％、80％、90％、95％、98％或 99％的氨基酸残基处未改变，如图43的比对结果中示出。Vd3A多肽合适地包含图40B中示出的天然Vd3A的整个预测保守结构域。示例性的 Vd3A多肽包含与图40B中示出的成熟Vd3A序列具有至少85％、86％、 87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、 98％、99％或100％同一性的序列。本发明的Vd3A多肽优选地具有β-葡糖苷酶活性。

[0208] 因此，本发明的Vd3A多肽合适地包含与SEQ ID NO:76的氨基酸序列或与SEQ ID NO:76第(i)19-296、(ii)19-649、(iii)19-890、(iv)415-649 或(v)415-890位残基具有至少80％、85％、90％、91％、92％、93％、 94％、95％、96％、97％、98％、99％或100％序列同一性的氨基酸序列。该多肽合适地具有β-葡糖苷酶活性。

[0209] 在某些方面，本发明的“Vd3A多肽”还可以指突变Vd3A多肽。可以向Vd3A多肽引入氨基酸置换以改善该分子的β-葡糖苷酶活性。例如，可以向Vd3A多肽引入氨基酸置换，所述氨基酸置换增强Vd3A多肽对其底物的结合亲和力或改善Vd3A催化β-D-葡糖苷中的末端非还原性残基水解的能力。在某些方面，突变Vd3A多肽包含一种或多种保守氨基酸置换。在某些方面，突变Vd3A多肽包含一种或多种非保守的氨基酸置换。在某些方面，一种或多种氨基

酸置换位于Vd3A多肽的CD中。在某些方面，一种或多种氨基酸置换位于Vd3A多肽的CBM中。在某些方面，一种或多种氨基酸置换同时位于CD和CBM中。在某些方面，Vd3A多肽的氨基酸置换可以在氨基酸E524和/或D295处发生。在某些方面，Vd3A多肽的氨基酸置换可以在氨基酸D107、R113、L156、R171、K204、H205、 R215、M260、Y263、D295、W296、S465和/或E524中的一个或多个处发生。突变Vd3A多肽合适地具有β-葡糖苷酶活性。

[0210] 在某些方面，Vd3A多肽包含两种β-葡糖苷酶序列的嵌合体/杂合体/融合体，其中第一β-葡糖苷酶序列具有至少约200个氨基酸残基长度，并且包含与Vd3A的等长度序列(SEQ ID NO:76)约60％、65％、70％、75％或 80％或更高的序列同一性，并且其中第二β-葡糖苷酶序列具有至少约50个氨基酸残基长度并且包含与SEQ ID NOs:54、56、58、60、62、64、66、 68、70、72、74、78和79中任何一种的等长度序列约60％、65％、70％、 75％、80％或更高的序列同一性，或包含多肽序列基序SEQ ID NO:170。在某些方面，第一β-葡糖苷酶序列包含SEQ ID NO:76的至少200个氨基酸残基的N末端序列，并且第二β-葡糖苷酶序列包含SEQ ID NOs:54、 56、58、60、62、64、66、68、70、72、74、78和79中任何一种的至少约50个连续氨基酸残基的C末端序列，或包含多肽序列基序SEQ ID NO: 170。

[0211] 在某些方面，本发明的Vd3A多肽包含两种β-葡糖苷酶序列的嵌合体或嵌合构建体，其中第一β-葡糖苷酶序列具有至少约200个氨基酸残基长度并且包含与SEQ ID NOs:54、56、58、60、62、64、66、68、70、72、 74、78和79中任何一种的等长度序列约60％、65％、70％、75％、80％或更高的序列同一性，或包含多肽序列基序SEQ ID NOs:164-169的一种或多种或全部，而第二β-葡糖苷酶序列具有至少约50个氨基酸残基长度并且包含与Vd3A的等长度序列(SEQ ID NO:76)约60％、65％、70％、75％、 80％或更高的序列同一性。在某些方面，第一β-葡糖苷酶序列包含SEQ ID NOs:54、56、58、60、62、64、66、68、70、72、74、78和79中任何一种的至少200个氨基酸残基的N末端，或包含多肽序列基序SEQ ID NOs: 164-169的一种或多种或全部，并且第二β-葡糖苷酶序列包含SEQ ID NO:76的至少50个连续氨基酸残基的C末端序列。

[0212] 在某些方面，第一β-葡糖苷酶序列位于嵌合β-葡糖苷酶多肽的N末端处，而第二β-葡糖苷酶序列位于嵌合β-葡糖苷酶多肽的C末端处。在某些实施例中，第一、第二或两种β-葡糖苷酶序列还包含一个或多个糖基化位点。在某些实施例中，第一种和第二β-葡糖苷酶序列互相紧邻或直接互相连接。在其他实施例中，第一种和第二β-葡糖苷酶序列并非紧邻，而是通过接头结构域连接。在某些方面，第一种和第二β-葡糖苷酶序列包含环区或代表环样结构的序列，其包含约3、4、5、6、7、8、9、10或11个氨基酸残基，所述序列包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT 的序列(SEQ ID NO:172)。在某些方面，第一或第二β-葡糖苷酶序列均不包含环序列。在一些实施例中，接头结构域包含环区，所述环区包含约3、4、5、6、7、8、9、10或11个氨基酸残基，所述环区包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。在一些实施例中，连接第一β-葡糖苷酶序列和第二β-葡糖苷酶序列的接头结构域位于中央(即，并非位于嵌合多肽的N末端或C末端)。在某些方面，嵌合β-葡糖苷酶的N末端序列包含长度至少200、250、300、350、400、450、 500、550或600个残基的来自Vd3A多肽或其变体的序列。在某些方面， N末端序列包含SEQ ID NOs:136-148所代表的多肽序列基序的一种或多种或全部，或优选地包含基序SEQ ID NOs:164-169。在某些方面，C末端序列包含长度至少50、75、100、125、150、175或200个氨基酸残基的来自

β-葡糖苷酶多肽或其变体的序列。在某些方面,C末端序列包含SEQ ID NOs:149-156所代表的多肽序列基序的一种或多种或全部,或优选地包含序列基序SEQ ID NO:170。在某些实施例中,β-葡糖苷酶多肽、其变体或其杂合体或嵌合体还包含一个或多个糖基化位点。一个或多个糖基化位点可以位于C末端序列内部、N末端序列内部或位于两者内部。

[0213] 在某些方面,本发明的非天然存在的纤维素酶或半纤维素酶组合物还包含一种或多种天然存在的半纤维素酶。在某些方面,非天然存在的纤维素酶组合物具有优于天然酶的改善的稳定性,所述天然酶包括衍生嵌合β- 葡糖苷酶的C末端或N末端序列的Vd3A。在某些方面,改善的稳定性包括在储存、表达或生产过程期间蛋白水解稳定性的改善。在某些方面,改善的稳定性包括在储存或生产条件下的酶活性损失的速度或程度的相关降低,其中酶活性损失优选地低于约50%,低于约40%,低于约20%,更优选地低于约15%,或甚至更优选地低于约10%。在某些方面,N末端序列或C末端序列可包含长度约3、4、5、6、7、8、9、10或11个氨基酸残基的环序列,所述环序列包含FDRRSPG的序列(SEQ ID NO:171)或 FD(R/K)YNIT的序列(SEQ ID NO:172)。N末端和C末端序列可以互相紧邻或直接互相连接。在其他方面,N末端序列和C末端序列可以通过接头结构域连接。在某些实施例中,接头结构域包含长度约3、4、5、6、7、8、 9、10或11个氨基酸残基的环序列,所述环序列包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。在某些方面,非天然存在的纤维素酶组合物包含β-葡糖苷酶活性。在某些方面,非天然存在的纤维素酶组合物还包含一种或多种木聚糖酶、β-木糖苷酶和/或L-α- 阿拉伯呋喃糖苷酶活性。

[0214] Pa3G

[0215] Pa3G的氨基酸序列(SEQ ID NO:78)在图41B和43中示出。SEQ ID NO:78是未成熟Pa3G的序列。Pa3G具有与SEQ ID NO:78的第1位至第19位(以下划线标出)相对应的预测的信号序列;对该信号序列的切割据预测产生具有与SEQ ID NO:78的第20至805位相对应的序列的成熟蛋白。使用SignalP-NN算法进行信号序列预测。预测的保守结构域在图41B 中以粗体标出。结构域预测基于Pfam、SMART或NCBI数据库进行。Pa3G的残基E517和D289据预测分别作为催化性酸碱和亲核物质发挥功能,这是基于上文所述的来自例如以下的GH3葡糖苷酶序列比对结果:鹅掌柄孢霉(登录号XP_001912683)、大丽轮枝菌、鞭毛藻丛赤壳菌(登录号XP_003045443)、玉米赤霉(登录号XP_386781)、尖孢镰刀菌 (登录号BGL FOXG_02349)、黑曲霉(登录号CAK48740)、埃默森篮状菌(登录号AAL69548)、里氏木霉(登录号AAP57755)、里氏木霉 (登录号AAA18473)、轮枝镰刀菌和新阿波罗栖热袍菌(登录号Q0GC07)等等(参见图43)。如本文所用,“Pa3G多肽”在某些方面是指多肽或其变体,其包含与SEQ ID NO:78的第20位至第805位残基中的至少50、75、100、125、150、175、200、250、300、350、400、450、 500、550、600、650、700或750个连续氨基酸残基具有至少85%、 86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、 97%、98%、99%或100%序列同一性。与天然Pa3G相比,Pa3G多肽优选地在残基E517和D289处未改变。Pa3G多肽优选地在本文所述的GH3家族β-葡糖苷酶之间保守的至少70%、80%、90%、95%、98%或99%的氨基酸残基处未改变,如图43的比对结果中示出。Pa3G多肽合适地包含图 41B中示出的天然Pa3G的整个预测保守结构域。示例性的Pa3G多肽包含与图41B中示出的成熟Pa3G序列具有至少85%、86%、87%、88%、 89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性的序列。本发明的Pa3G多肽优选地具有β-葡糖苷酶活性。

[0216] 因此，本发明的Pa3G多肽合适地包含与SEQ ID NO:78的氨基酸序列或与SEQ ID NO:78第(i)20-354、(ii)20-660、(iii)20-805、(iv)449-660或 (v)449-805位残基具有至少80％、85％、90％、91％、92％、93％、94％、 95％、96％、97％、98％、99％或100％序列同一性的氨基酸序列。该多肽合适地具有β-葡糖苷酶活性。

[0217] 在某些方面，本发明的“Pa3G多肽”还可以指突变Vd3A多肽。可以向a3G多肽引入氨基酸置换P以改善该分子的β-葡糖苷酶活性。例如，可以向Pa3G多肽中引入氨基酸置换，所述氨基酸置增强Pa3G多肽对其底物的结合亲和力或改善其催化β-D-葡糖苷中的末端非还原性残基水解的能力。在某些方面，突变Pa3G多肽包含一种或多种保守氨基酸置换。在某些方面，突变Pa3G多肽包含一种或多种非保守的氨基酸置换。在某些方面，一种或多种氨基酸置换位于Pa3G多肽的CD中。在某些方面，一种或多种氨基酸置换位于Pa3G多肽的CBM中。在某些方面，一种或多种氨基酸置换同时位于CD和CBM中。在某些方面，Pa3G多肽的氨基酸置换可以在氨基酸E517和/或D289处发生。在某些方面，Pa3G多肽的氨基酸置换可以在氨基酸D101、R107、L150、R165、K199、H209、R215、 M254、Y257、D289、W290、S458和/或E517中的一个或多个处发生。突变Pa3G多肽合适地具有β-葡糖苷酶活性。

[0218] 在某些方面，Pa3G多肽包含两种β-葡糖苷酶序列的嵌合体/融合体/杂合体，其中第一β-葡糖苷酶序列具有至少约200个氨基酸残基长度，并且包含与Pa3G的等长度序列(SEQ ID NO:78)约60％、65％、70％、75％或 80％或更高的序列同一性，并且其中第二β-葡糖苷酶序列具有至少约50个氨基酸残基长度并且包含与SEQ ID NOs:54、56、58、60、62、64、66、 68、70、72、74、76和79中任何一种的等长度序列至少约60％、65％、 70％、75％、80％或更高的序列同一性，或包含多肽序列基序SEQ ID NO:170。在某些方面，第一β-葡糖苷酶序列包含SEQ ID NO:78的至少 200个氨基酸残基的N末端序列，并且第二β-葡糖苷酶序列包含SEQ ID NOs:54、56、58、60、62、64、66、68、70、72、74、76和79中任何一种的至少约50个连续氨基酸残基的C末端序列，或包含多肽序列基序 SEQ ID NO:170。

[0219] 在某些方面，本发明的Pa3G多肽包含两种β-葡糖苷酶序列的嵌合体或嵌合构建体，其中第一β-葡糖苷酶序列具有至少约200个氨基酸残基长度，并且包含与SEQ ID NOs:54、56、58、60、62、64、66、68、70、 72、74、76和79中任何一种的等长度序列约60％、65％、70％、75％、 80％或更高的序列同一性，或包含SEQ ID NOs:164-169的多肽序列基序中的一种或多种或全部，而第二β-葡糖苷酶序列具有至少约50个氨基酸残基长度，并且包含与Pa3G的等长度序列(SEQ ID NO:78)约60％、65％、 70％、75％、80％或更高的序列同一性。在某些方面，第一β-葡糖苷酶序列包含SEQ ID NOs:54、56、58、60、62、64、66、68、70、72、74、76 和79中任何一种的至少200个氨基酸残基的N末端序列，或包含多肽序列基序SEQ ID NOs:164-169的一种或多种或全部，并且第二β-葡糖苷酶序列包含SEQ ID NO:78的至少50个连续氨基酸残基的C末端序列。

[0220] 在某些方面，第一β-葡糖苷酶序列位于嵌合β-葡糖苷酶多肽的N末端处，而第二β-葡糖苷酶序列位于嵌合β-葡糖苷酶多肽的C末端处。在某些实施例中，第一、第二或两种β-葡糖苷酶序列还包含一个或多个糖基化位点。在某些实施例中，第一种和第二β-葡糖苷酶序列互相紧邻或直接互相连接。在其他实施例中，第一种和第二β-葡糖苷酶序列并非紧邻，而是通过接头结构域连接。在某些方面，第一或第二β-葡糖苷酶序列包含环区或代表环样结构的序列，其包含约3、4、5、6、7、8、9、10或11个氨基酸残基，所述序列包含FDRRSPG的序列

(SEQ ID NO:171)或FD(R/K)YNIT 的序列(SEQ ID NO:172)。在某些方面,第一或第二β-葡糖苷酶序列均不包含环序列。在一些实施例中,接头结构域包含环区,所述环区包含约3、4、5、6、7、8、9、10或11个氨基酸残基、STYLE 1>FDRRSPG的序列 (SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。在一些实施例中,连接第一β-葡糖苷酶序列和第二β-葡糖苷酶序列的接头结构域位于中央(即,并非位于嵌合多肽的N末端或C末端)。在某些方面,嵌合β-葡糖苷酶的N末端序列包含长度至少200、250、300、350、400、450、 500、550或600个残基的来自Pa3G多肽或其变体的序列。在某些方面, N末端序列包含SEQ ID NOs:136-148所代表的多肽序列基序的一种或多种或全部,或优选地包含基序SEQ ID NOs:164-169。在某些方面,C末端序列包含长度至少50、75、100、125、150、175或200个氨基酸残基的来自β-葡糖苷酶多肽或其变体的序列。在某些方面,C末端序列包含SEQ ID NOs:149-156所代表的多肽序列基序的一种或多种或全部,或优选地包含基序SEQ ID NO:170。在某些实施例中,β-葡糖苷酶多肽、其变体或其杂合体或嵌合体还包含一个或多个糖基化位点。一个或多个糖基化位点可以位于C末端序列内部、N末端序列内部或位于两者内部。

[0221] 在某些方面,本发明的非天然存在的纤维素酶或半纤维素酶组合物还包含一种或多种天然存在的半纤维素酶。在某些方面,非天然存在的纤维素酶组合物具有优于天然酶的改善的稳定性,所述天然酶包括衍生嵌合β- 葡糖苷酶的C末端或N末端序列的Pa3G。在某些方面,改善的稳定性包括在储存、表达或生产过程期间蛋白水解稳定性的改善。在某些方面,改善的稳定性包括在储存或生产条件下的酶活性损失的速度或程度的相关降低,其中酶活性损失优选地低于约50%,低于约40%,低于约20%,更优选地低于约15%,或甚至更优选地低于约10%。在某些方面,N末端序列或C末端序列可包含长度约3、4、5、6、7、8、9、10或11个氨基酸残基的环序列,所述环序列包含FDRRSPG的序列(SEQ ID NO:171)或 FD(R/K)YNIT的序列(SEQ ID NO:172)。N末端和C末端序列可以互相紧邻或直接互相连接。在其他方面,N末端序列和C末端序列可以通过接头结构域连接。在某些实施例中,接头结构域包含长度约3、4、5、6、7、8、 9、10或11个氨基酸残基的环序列,所述环序列包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。在某些方面,非天然存在的纤维素酶组合物包含β-葡糖苷酶活性。在某些方面,非天然存在的纤维素酶组合物还包含一种或多种木聚糖酶、β-木糖苷酶和/或L-α- 阿拉伯呋喃糖苷酶活性。

[0222] Tn3B

[0223] Tn3B的氨基酸序列(SEQ ID NO:79)在图42和43中示出。SEQ ID NO:79是未成熟Tn3B的序列。SignalP-NN算法(http://www.cbs.dtu.dk)并未提供预测的信号序列。Tn3B的残基E458和D242据预测分别作为催化性酸碱和亲核物质发挥功能,这是基于上文所述的来自例如以下的GH3葡糖苷酶的序列比对结果:鹅掌柄孢霉(登录号XP_001912683)、大丽轮枝菌、鞭毛藻丛赤壳菌(登录号XP_003045443)、玉米赤霉(登录号 XP_386781)、尖孢镰刀菌(登录号BGL FOXG_02349)、黑曲霉(登录号CAK48740)、埃默森篮状菌(登录号AAL69548)、里氏木霉(登录号AAP57755)、里氏木霉(登录号AAA18473)、轮枝镰刀菌和新阿波罗栖热袍菌(登录号Q0GC07)等等(参见图43)。如本文所用,“Tn3B 多肽”在某些方面是指包含以下序列的多肽和/或其变体,所述序列与SEQ ID NO:79的至少50、75、100、125、150、175、200、250、300、350、 400、450、500、550、600、650、700或750个连续氨基酸残基具有至少 85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%序列同

一性。与天然Tn3B相比，Tn3B多肽优选地在残基E458和D242处未改变。Tn3B多肽优选地在本文所述的 GH3家族β-葡糖苷酶之间保守的至少70％、80％、90％、95％、98％或 99％的氨基酸残基处未改变，如图43的比对结果中示出。Tn3B多肽合适地包含图43中示出的天然Tn3B的整个预测保守结构域。示例性的Tn3B 多肽包含与图42中示出的成熟Tn3B序列具有至少85％、86％、87％、 88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、 99％或100％同一性的序列。本发明的Tn3B多肽优选地具有β-葡糖苷酶活性。

[0224] 因此，本发明的Tn3B多肽合适地包含与SEQ ID NO:79的氨基酸序列具有至少80％、85％、90％、91％、92％、93％、94％、95％、96％、 97％、98％、99％或100％序列同一性的氨基酸序列。该多肽合适地具有β- 葡糖苷酶活性。

[0225] 在某些方面，本发明的“Tn3B多肽”还可以指突变Tn3B多肽。可以向Tn3B多肽引入氨基酸置换以改善该分子的β-葡糖苷酶活性。例如，可以向Tn3B多肽引入氨基酸置换，所述氨基酸置换增强Tn3B多肽对其底物的结合亲和力或改善Tn3B催化β-D-葡糖苷中的末端非还原性残基水解的能力。在某些方面，突变Tn3B多肽包含一种或多种保守氨基酸置换。在某些方面，突变Tn3B多肽包含一种或多种非保守的氨基酸置换。在某些方面，一种或多种氨基酸置换位于Tn3B多肽的CD中。在某些方面，一种或多种氨基酸置换位于Tn3B多肽的CBM中。在某些方面，一种或多种氨基酸置换同时位于CD和CBM中。在某些方面，Tn3B多肽的氨基酸置换可以在氨基酸E458和/或D242处发生。在某些方面，Tn3B多肽的氨基酸置换可以在氨基酸D58、R64、L116、R130、K163、H164、R174、 M207、Y210、D242、W243、S370和/或E458中的一个或多个处发生。突变Tn3B多肽合适地具有β-葡糖苷酶活性。

[0226] 在某些方面，Tn3B多肽包含两种β-葡糖苷酶序列的嵌合体/融合体/杂合体，其中第一β-葡糖苷酶序列具有至少约200个氨基酸残基长度，并且包含与Tn3B的等长度序列(SEQ ID NO:79)约60％、65％、70％、75％或 80％或更高的序列同一性，并且其中第二β-葡糖苷酶序列具有至少约50个氨基酸残基长度并且包含与SEQ ID NOs:54、56、58、60、62、64、66、 68、70、72、74、76和78中任何一种的等长度序列至少约60％、65％、 70％、75％、80％或更高的序列同一性，或包含多肽序列基序SEQ ID NO:170。在某些方面，第一β-葡糖苷酶序列包含SEQ ID NO:79的至少 200个氨基酸残基的N末端序列，并且第二β-葡糖苷酶序列包含SEQ ID NOs:54、56、58、60、62、64、66、68、70、72、74、76和78中任何一种的至少约50个连续氨基酸残基的C末端序列，或包含多肽序列基序 SEQ ID NO:170。

[0227] 在某些方面，本发明的Tn3B多肽包含两种β-葡糖苷酶序列的嵌合体或嵌合构建体，其中第一β-葡糖苷酶序列具有至少约200个氨基酸残基长度，并且包含与SEQ ID NOs:54、56、58、60、62、64、66、68、70、 72、74、76和78中任何一种的等长度序列约60％、65％、70％、75％、 80％或更高的序列同一性，或包含多肽序列基序SEQ ID NOs:164-169的一种或多种或全部，而第二β-葡糖苷酶序列具有至少约50个氨基酸残基长度并且包含与Tn3B的等长度序列(SEQ ID NO:79)约60％、65％、70％、 75％、80％或更高的序列同一性。在某些方面，第一β-葡糖苷酶序列包含 SEQ ID NOs:54、56、58、60、62、64、66、68、70、72、74、76和78中任何一种的至少200个氨基酸残基的N末端序列，或者包含多肽序列基序 SEQ ID NOs:164-169的一种或多种或全部，并且第二β-葡糖苷酶序列包含 SEQ ID NO:79的至少50个连续氨基酸残基的C末端序列。

[0228] 在某些方面，第一β-葡糖苷酶序列位于嵌合β-葡糖苷酶多肽的N末端处，而第二β-

葡糖苷酶序列位于嵌合β-葡糖苷酶多肽的C末端处。在某些实施例中，第一、第二或两种β-葡糖苷酶序列还包含一个或多个糖基化位点。在某些实施例中，第一种和第二β-葡糖苷酶序列互相紧邻或直接互相连接。在其他实施例中，第一种和第二β-葡糖苷酶序列并非紧邻，而是通过接头结构域连接。在某些方面，第一或第二β-葡糖苷酶序列包含环区或代表环样结构的序列，其包含约3、4、5、6、7、8、9、10或11个氨基酸残基，所述序列包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT 的序列(SEQ ID NO:172)。在某些方面，第一或第二β-葡糖苷酶序列均不包含环序列。在一些实施例中，接头结构域包含环区，所述环区包含约3、4、5、6、7、8、9、10或11个氨基酸残基。在一些实施例中，连接第一β-葡糖苷酶序列和第二β-葡糖苷酶序列的接头结构域位于中央(即，并非位于嵌合多肽的N末端或C末端)。在某些方面，嵌合β-葡糖苷酶的N 末端序列包含长度至少200、250、300、350、400、450、500、550或600 个残基的来自Tn3B多肽或其变体的序列。在某些方面，N末端序列包含 SEQ ID NOs:136-148所代表的多肽序列基序的一种或多种或全部，或优选地包含基序SEQ ID NOs:164-169。在某些方面，C末端序列包含长度至少 50、75、100、125、150、175或200个氨基酸残基的来自β-葡糖苷酶多肽或其变体的序列。在某些方面，C末端序列包含SEQ ID NOs:149-156所代表的多肽序列基序的一种或多种或全部，或优选地包含基序SEQ ID NO:170。在某些实施例中，β-葡糖苷酶多肽、其变体或其杂合体或嵌合体还包含一个或多个糖基化位点。一个或多个糖基化位点可以位于C末端序列内部、N末端序列内部或位于两者内部。

[0229] 在某些方面，本发明的非天然存在的纤维素酶或半纤维素酶组合物还包含一种或多种天然存在的半纤维素酶。在某些方面，非天然存在的纤维素酶组合物具有优于天然酶的改善的稳定性，所述天然酶包括衍生嵌合β- 葡糖苷酶的C末端或N末端序列的Tn3B。在某些方面，改善的稳定性包括在储存、表达或生产过程期间蛋白水解稳定性的改善。在某些方面，改善的稳定性包括在储存或生产条件下的酶活性损失的速度或程度的相关降低，其中酶活性损失优选地低于约50％，低于约40％，低于约20％，更优选地低于约15％，或甚至更优选地低于约10％。在某些方面，N末端序列或C末端序列可包含长度约3、4、5、6、7、8、9、10或11个氨基酸残基的环序列，所述环序列包含FDRRSPG的序列(SEQ ID NO:171)或 FD(R/K)YNIT的序列(SEQ ID NO:172)。N末端和C末端序列可以互相紧邻或直接互相连接。在其他方面，N末端序列和C末端序列可以通过接头结构域连接。在某些实施例中，接头结构域包含长度约3、4、5、6、7、8、 9、10或11个氨基酸残基的环序列，所述环序列包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。在某些方面，非天然存在的纤维素酶组合物包含β-葡糖苷酶活性。在某些方面，非天然存在的纤维素酶组合物还包含一种或多种木聚糖酶、β-木糖苷酶和/或L-α- 阿拉伯呋喃糖苷酶活性。

[0230] 核酸

[0231] 示例性β-葡糖苷酶核酸包括编码具有β-葡糖苷酶多肽的至少一种活性的多肽、多肽片段、肽或融合多肽的核酸。示例性β-葡糖苷酶多肽和核酸包括来自本文所述的任何来源生物的天然存在的多肽和核酸以及源自本文所述的任何来源生物的突变多肽和核酸。示例性β-葡糖苷酶核酸包括，例如，分离自以下一种或多种生物(但不限于这些生物)的β-葡糖苷酶：柄毛皮伞、菜豆壳球孢、嗜热毁丝菌、粪生粪壳菌、拟刺盘孢周刺座霉、太瑞斯梭孢壳霉、支顶孢属、黑耳、木蹄层孔菌、绵皮孔菌属、红根囊壶菌、微小根毛霉、闪光须霉、弗雷生刺枝霉、棉色二孢、异色尾孢藻、集孢粪盘菌、疣孢青霉、产黄青霉、疣状顶孢霉、合生腐皮

壳菌、黄瓜炭疽菌、黑孢霉属、团炭角菌、松色丛赤壳菌、大孢粪壳菌、嗜热梭孢壳霉、突孢毛壳、绿毛壳、巴西毛壳、链丝毛壳菌、Syspastospora boninensis、多生枝鼻菌、嗜热柱霉、链孢粘帚菌、尖孢镰孢菌番茄亚种、尖孢镰孢菌西番莲亚种、镰素对镰孢菌、蛇形镰孢菌、梨孢镰孢菌、黑腐质霉、灰腐质霉、网纹斑褶菌、血红栓菌、裂褶菌、粉红单端孢、小球壳孢菌、粪盘菌、点孔座壳、多节孢属、木霉属(例如,里氏木霉)以及柱孢菌属 (Cylindrocarpon sp)。

[0232] 本发明提供了分离的、合成的或重组的核酸,其包含与SEQ ID NO: 1、3、5、7、9、11、13、15、17、19、21、23、25、27、29、31、33、 35、37、39、41、46、47、48、49、50、51、53、57、59、61、63、65、 67、69、71、73、75或77的核酸在至少约10个核苷酸(例如,至少约 15、20、25、30、35、40、45、50、75、100、150、200、250、300、 350、400、450、500、550、600、650、700、750、800、850、900、 950、1000、1050、1100、1150、1200、1250、1300、1350、1400、 1450、1500、1550、1600、1650、1700、1750、1800、1850、1900、1950 或2000个核苷酸)的区域范围内具有至少约70%的序列同一性,例如至少约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%或完全(100%)序列同一性的核酸序列。本发明还提供了编码至少一种多肽的核酸,所述多肽具有半纤维素分解活性(例如,木聚糖酶、β-木糖苷酶和/或L-α-阿拉伯呋喃糖苷酶活性)。此外,本发明提供了编码具有纤维素分解活性(例如,β-葡糖苷酶活性或内切葡聚糖酶活性)的多肽的核酸。

[0233] 本发明的核酸还包括分离的、合成的或重组的核酸,其编码酶或酶的成熟部分,所述酶包含SEQ ID NO:2、4、6、8、10、12、14、16、18、 20、22、24、26、28、30、32、34、36、38、40、42、43、44、52、54、 56、58、60、62、64、66、68、70、72、74、76、78或79的序列,或是 GH61内切葡聚糖酶或该酶的成熟部分,所述酶包含如下多肽序列基序: (1)SEQ ID NOs:84和88;(2)SEQ ID NOs:85和88;(3)SEQ ID NO:86;(4) SEQ ID NO:87;(5)SEQ ID NOs:84、88和89;(6)SEQ ID NOs:85、88和 89;(7)SEQ ID NOs:84、88和90;(8)SEQ ID NOs:85、88和90;(9)SEQ ID NOs:84、88和91;(10)SEQ ID NOs:85、88和91;(11)SEQ ID NOs: 84、88、89和91;(12)SEQ ID NOs:84、88、90和91;(13)SEQ ID NOs: 85、88、89和91和(14)SEQ ID NOs:85、88、90和91,及其子序列(例如,保守结构域或碳水化合物结合结构域(“CBM”)及其变体。

[0234] 本发明具体地提供了编码Fv3A、Pf43A、Fv43E、Fv39A、Fv43A、 Fv43B、Pa51A、Gz43A、Fo43A、Af43A、Pf51A、AfuXyn2、AfuXyn5、 Fv43D、Pf43B、Fv43B、Fv51A、里氏木霉Xyn3、里氏木霉Xyn2、里氏木霉Bxl1、里氏木霉Bgl1(Tr3A)、里氏木霉Eg4、里氏木霉Bgl3(Tr3B)、Pa3D、Fv3G、Fv3D、Fv3C、Te3A、An3A、Fo3A、Gz3A、 Nh3A、Vd3A、Pa3G或Tn3B多肽、其变体、突变体或杂合或嵌合多肽的核酸。在某些方面,本发明提供了编码嵌合或融合酶的核酸,所述嵌合或融合酶包含例如第一β-葡糖苷酶序列和第二β-葡糖苷酶序列,其中第一β-葡糖苷酶序列和第二β-葡糖苷酶序列衍生自不同的生物。在某些方面,第一β-葡糖苷酶序列位于杂合或嵌合β-葡糖苷酶多肽的N末端,而第二β-葡糖苷酶位于其C末端。在某些方面,第一β-葡糖苷酶序列,或更具体地讲第一β-葡糖苷酶序列的C末端直接邻接于或连接于第二β-葡糖苷酶序列,或更具体地讲第二β-葡糖苷酶序列的N末端。在一些实施例中,第一β-葡糖苷酶序列和第二β-葡糖苷酶并非直接邻接或连接,相反地,第一β-葡糖苷酶序列通过接头序列或结构域有效连接或相接于第二β-葡糖苷酶序列。在一些例子中,第一β-葡糖苷酶序列具有至少约200个氨基酸残基长度,并且包含SEQ ID NOs:136-148所代表的多肽序

列基序的一种或多种或全部，而所述第二β-葡糖苷酶序列具有至少约50个氨基酸残基长度并且包含SEQ ID NOs:149-156所代表的多肽序列基序的一种或多种或全部。特别地，所述两种或更多种β-葡糖苷酶序列的第一种是具有至少约200个氨基酸残基长度并且包含SEQ ID NOs:164-169的氨基酸序列基序中的至少2 种(例如，至少2、3、4或全部)的序列，并且所述两种或更多种β-葡糖苷酶序列的第二种具有至少50个氨基酸残基长度并且包含SEQ ID NO:170。在某些方面，第一β-葡糖苷酶序列和第二β-葡糖苷酶序列直接互相连接或互相紧邻。在某些方面，第一β-葡糖苷酶序列并非与第二β-葡糖苷酶序列直接连接或紧邻，相反地，第一种和第二β-葡糖苷酶通过接头序列连接。在某些实施例中，接头序列位于中央。在某些具体的例子中，第一β-葡糖苷酶序列包含例如Fv3C多肽的至少200个氨基酸残基长度的N 末端序列的序列。在一些实施例中，第二β-葡糖苷酶序列包含例如里氏木霉Bgl3多肽的至少50个氨基酸残基长度的C末端序列的序列。在具体的例子中，该β-葡糖苷酶多肽是杂合或嵌合Fv3C多肽或里氏木霉Bgl3 (Tr3B)多肽，并且包含SEQ ID NO:159的氨基酸序列。又如，该β-葡糖苷酶多肽是杂合或嵌合Fv3C多肽或里氏木霉Bgl3多肽，其任选地包含衍生自第三β-葡糖苷酶多肽序列的接头序列，其中所述β-葡糖苷酶多肽包含 SEQ ID NO:135的氨基酸序列。在某些方面，该嵌合或融合酶合适地还包含接头序列，并且因此本发明提供了编码嵌合酶的核酸，其中可以将所述嵌合酶认为是衍生生N末端序列、C末端序列或其子序列中任一个的β-葡糖苷酶多肽。例如，可以将杂合Fv3C/Bgl3多肽认为是Fv3C多肽、其变体、里氏木霉Bgl3多肽、其变体或嵌合Fv3C/Bgl3多肽或其变体。又如，可以将杂合Fv3C/Te3A/Bgl3多肽认为是Fv3C多肽或其变体、里氏木霉 Bgl3多肽或其变体、Te3A多肽或其变体、或嵌合Fv3C/Te3A/Bgl3多肽或其变体。

[0235] 在多核苷酸序列的上下文中使用时，术语“变体”可以涵盖相关于基因或其编码序列相关的多核苷酸序列。该定义还可以包括，例如，“等位的”、“剪接的”、“物种的”或“多态性的”变体。剪接变体可以与参比多核苷酸具有明显的同一性，但通常由于mRNA加工期间对外显子的可变剪接而具有更多或更少数目的残基。对应的多肽可以具有额外的功能结构域或结构域缺失。物种变体是在一个物种与另一个物种之间变动的多核苷酸序列。所得的多肽通常相互之间具有明显的氨基酸同一性，如其中进一步详述。多态性变体是在给定物种的个体之间特定基因的多核苷酸序列中的变化。

[0236] 例如，本发明提供了分离的核酸分子，其中该核酸分子编码：

[0237] (1)包含与SEQ ID NO:54的氨基酸序列或与SEQ ID NO:54第(i)18- 282、(ii)18-601、(iii)18-733、(iv)356-601或(v)356-733位残基之间具有至少80％、85％、90％、91％、92％、93％、94％、 95％、96％、97％、98％、99％或100％序列同一性的氨基酸序列的多肽；或者

[0238] (2)包含与SEQ ID NO:56的氨基酸序列或与SEQ ID NO:56第(i)22- 292、(ii)22-629、(iii)22-780、(iv)373-629或(v)373-780位残基之间具有至少80％、85％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％序列同一性的氨基酸序列的多肽；或者

[0239] (3)包含与SEQ ID NO:58的氨基酸序列或与SEQ ID NO:58第(i)20- 321、(ii)20-651、(iii)20-811、(iv)423-651或(v)423-811位残基之间具有至少80％、85％、90％、91％、92％、93％、94％、 95％、96％、97％、98％、99％或100％序列同一性的氨基酸序列的多肽；

或者

[0240] (4)包含与SEQ ID NO:60的氨基酸序列或与SEQ ID NO:60第(i)20- 327、(ii)22-600、(iii)20-899、(iv)428-899或(v)428-660位残基之间具有至少80％、85％、90％、91％、92％、93％、94％、 95％、96％、97％、98％、99％或100％序列同一性的氨基酸序列的多肽；或者

[0241] (5)包含与SEQ ID NO:62的氨基酸序列或与SEQ ID NO:62第(i)20- 287、(ii)22-611、(iii)20-744、(iv)362-611或(v)362-744位残基之间具有至少80％、85％、90％、91％、92％、93％、94％、 95％、96％、97％、98％、99％或100％序列同一性的氨基酸序列的多肽；或者

[0242] (6)包含与SEQ ID NO:64的氨基酸序列或与SEQ ID NO:64第(i)19- 307、(ii)19-640、(iii)19-874、(iv)407-640或(v)407-874位残基之间具有至少80％、85％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％序列同一性的氨基酸序列的多肽;或者

[0243] (7)包含与SEQ ID NO:66的氨基酸序列或与SEQ ID NO:66第(i)20- 297、(ii)20-629、(iii)20-857、(iv)396-629或(v)396-857位残基之间具有至少80％、85％、90％、91％、92％、93％、94％、 95％、96％、97％、98％、99％或100％序列同一性的氨基酸序列的多肽；或者

[0244] (8)包含与SEQ ID NO:68的氨基酸序列或与SEQ ID NO:68第(i)20- 300、(ii)20-634、(iii)20-860、(iv)400-634或(v)400-860位残基之间具有至少80％、85％、90％、91％、92％、93％、94％、 95％、96％、97％、98％、99％或100％序列同一性的氨基酸序列的多肽；或者

[0245] (9)包含与SEQ ID NO:70的氨基酸序列或与SEQ ID NO:70第(i)20- 327、(ii)20-660、(iii)20-899、(iv)428-660或(v)428-899位残基之间具有至少80％、85％、90％、91％、92％、93％、94％、 95％、96％、97％、98％、99％或100％序列同一性的氨基酸序列的多肽；或者

[0246] (10)包含与SEQ ID NO:72的氨基酸序列或与SEQ ID NO:72第(i)19- 314、(ii)19-647、(iii)19-886、(iv)415-647或(v)415-886位残基之间具有至少80％、85％、90％、91％、92％、93％、94％、 95％、96％、97％、98％、99％或100％序列同一性的氨基酸序列的多肽;或者

[0247] (11)包含与SEQ ID NO:74的氨基酸序列或与SEQ ID NO:74第(i)20- 295、(ii)20-647、(iii)20-880、(iv)414-647或(v)414-880位残基之间具有至少80％、85％、90％、91％、92％、93％、94％、 95％、96％、97％、98％、99％或100％序列同一性的氨基酸序列的多肽;或者

[0248] (12)包含与SEQ ID NO:76的氨基酸序列或与SEQ ID NO:76第(i)19- 296、(ii)19-649、(iii)19-890、(iv)415-649或(v)415-890位残基之间具有至少80％、85％、90％、91％、92％、93％、94％、 95％、96％、97％、98％、99％或100％序列同一性的氨基酸序列的多肽;或者

[0249] (13)包含与SEQ ID NO:78的氨基酸序列或与SEQ ID NO:78第(i)20- 354、(ii)20-660、(iii)20-805、(iv)449-660或(v)449-805位残基之间具有至少80％、85％、90％、

91％、92％、93％、94％、 95％、96％、97％、98％、99％或100％序列同一性的氨基酸序列的多肽;或者

[0250] (14)包含与SEQ ID NO:79的氨基酸序列之间具有至少80％、85％、 90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％序列同一性的氨基酸序列的多肽。

[0251] 本发明还提供:

[0252] (1)与SEQ ID NO:53具有至少90％(例如,至少90％、91％、 92％、93％、94％、95％、96％、97％、98％、99％或更高)序列同一性的核酸,或能够与SEQ ID NO:53的互补物或与其片段在高严格性条件下杂交的核酸;或者

[0253] (2)与SEQ ID NO:55具有至少90％(例如,至少90％、91％、 92％、93％、94％、95％、96％、97％、98％、99％或更高)序列同一性的核酸,或能够与SEQ ID NO:55的互补物或与其片段在高严格性条件下杂交的核酸;或者

[0254] (3)与SEQ ID NO:57具有至少90％(例如,至少90％、91％、 92％、93％、94％、95％、96％、97％、98％、99％或更高)序列同一性的核酸,或能够与SEQ ID NO:57的互补物或与其片段在高严格性条件下杂交的核酸;或者

[0255] (4)与SEQ ID NO:59具有至少90％(例如,至少90％、91％、 92％、93％、94％、95％、96％、97％、98％、99％或更高)序列同一性的核酸,或能够与SEQ ID NO:59的互补物或与其片段在高严格性条件下杂交的核酸;或者

[0256] (5)与SEQ ID NO:61具有至少90％(例如,至少90％、91％、 92％、93％、94％、95％、96％、97％、98％、99％或更高)序列同一性的核酸,或能够与SEQ ID NO:61的互补物或与其片段在高严格性条件下杂交的核酸;或者

[0257] (6)与SEQ ID NO:63具有至少90％(例如,至少90％、91％、 92％、93％、94％、95％、96％、97％、98％、99％或更高)序列同一性的核酸,或能够与SEQ ID NO:63的互补物或与其片段在高严格性条件下杂交的核酸;或者

[0258] (7)与SEQ ID NO:65具有至少90％(例如,至少90％、91％、 92％、93％、94％、95％、96％、97％、98％、99％或更高)序列同一性的核酸,或能够与SEQ ID NO:65的互补物或与其片段在高严格性条件下杂交的核酸;或者

[0259] (8)与SEQ ID NO:67具有至少90％(例如,至少90％、91％、 92％、93％、94％、95％、96％、97％、98％、99％或更高)序列同一性的核酸,或能够与SEQ ID NO:67的互补物或与其片段在高严格性条件下杂交的核酸;或者

[0260] (9)与SEQ ID NO:69具有至少90％(例如,至少90％、91％、 92％、93％、94％、95％、96％、97％、98％、99％或更高)序列同一性的核酸,或能够与SEQ ID NO:69的互补物或与其片段在高严格性条件下杂交的核酸;或者

[0261] (10)与SEQ ID NO:71具有至少90％(例如,至少90％、91％、 92％、93％、94％、95％、96％、97％、98％、99％或更高)序列同一性的核酸,或能够与SEQ ID NO:71的互补物或与其片段在高严格性条件下杂交的核酸;或者

[0262] (11)与SEQ ID NO:73具有至少90％(例如,至少90％、91％、 92％、93％、94％、95％、96％、97％、98％、99％或更高)序列同一性的核酸,或能够与SEQ ID NO:73的互补物或与其片段在高严格性条件下杂交的核酸;或者

[0263] (12)与SEQ ID NO:75具有至少90％(例如,至少90％、91％、 92％、93％、94％、

95％、96％、97％、98％、99％或更高)序列同一性的核酸,或能够与SEQ ID NO:75的互补物或与其片段在高严格性条件下杂交的核酸;或者

[0264] (13)与SEQ ID NO:77具有至少90％(例如,至少90％、91％、 92％、93％、94％、95％、96％、97％、98％、99％或更高)序列同一性的核酸,或能够与SEQ ID NO:77的互补物或与其片段在高严格性条件下杂交的核酸。

[0265] 如本文所用,术语“在低严格性条件下、中度严格性条件下、高严格性条件下或极高严格性条件下杂交”描述用于杂交和洗涤的条件。对于进行杂交反应的指导可见于Current Protocols in Molecular Biology,John Wiley &Sons,N.Y.(1989),6.3.1-6.3.6(《最新分子生物学实验方法汇编》,纽约约翰威立出版社,1989年,6.3.1-6.3.6)。该参考文献中描述了含水和无水的方法,并且两种方法均可以使用。本文提及的具体杂交条件如下：1)低严格性杂交条件是在约45℃于6X氯化钠/柠檬酸钠(SSC)中,随后至少在50℃于0.2X SSC,0.1％SDS中洗涤两次(对于低严格性条件,所述洗涤的温度可以提高至55℃);2)中等严格性杂交条件是在约45℃于6X SSC中,随后在60℃于0.2X SSC,0.1％SDS中洗涤一次或多次;3)高严格性杂交条件是在约45℃于6X SSC中,随后在65℃于0.2X SSC,0.1％SDS中洗涤一次或多次;并且优选地4)极高严格性杂交条件是在65℃于 0.5M磷酸钠,7％SDS中,随后在65℃于0.2X SSC,1％SDS中洗涤一次或多次。除非另外指明,否则极高严格性条件(4)是优选的条件。

[0266] 分离核酸的方法的例子

[0267] 本发明的β-葡糖苷酶和其他核酸可以使用标准方法分离。从目的来源生物(例如细菌基因组)获得所需核酸的方法是分子生物学领域常见和众所周知的。分离核酸的标准方法,包括已知序列的PCR扩增、核酸的合成、基因组文库的筛选、粘粒文库的筛选,在国际专利公布No.WO 2009/076676A2和美国专利申请No.12/335,071中描述。

[0268] 宿主细胞的例子

[0269] 本发明提供了宿主细胞,其经过工程化以表达本发明的一种或多种酶。合适的宿主细胞包括任何生物的细胞(例如,细菌、原生生物、藻类、真菌(例如,酵母或丝状真菌)或其他微生物的细胞),并且优选地是细菌、酵母或丝状真菌的细胞。

[0270] 细菌菌属的合适宿主细胞包括但不限于,埃希杆菌属(Escherichia)、芽孢杆菌属(Bacillus)、乳酸杆菌属(Lactobacillus)、假单胞菌属 (Pseudomonas)和链霉菌属(Streptomyces)的细胞。合适的细菌菌种的细胞包括但不限于大肠杆菌(Escherichia coli)、枯草芽孢杆菌(Bacillus subtilis)、地衣芽孢杆菌(Bacillus licheniformis)、短乳杆菌(Lactobacillus brevis)、铜绿假单胞菌(Pseudomonas aeruginosa)和浅青紫链霉菌(Streptomyces lividans)的细胞。

[0271] 合适的酵母菌属的宿主细胞包括但不限于酵母菌属、裂殖酵母属(Schizosaccharomyces)、假丝酵母属(Candida)、汉逊酵母属(Hansenula)、毕赤酵母属、克鲁维酵母菌属(Kluyveromyces)和法夫酵母属(Phaffia)的细胞。合适的酵母菌种的细胞包括但不限于酿酒酵母、粟酒裂殖酵母 (Schizosaccharomyces pombe)、白假丝酵母(Candida albicans)、多形汉逊酵母(Hansenula polymorpha)、巴斯德毕赤酵母(Pichia pastoris)、加拿大毕赤酵母(P.canadensis)、马克斯克鲁维酵母(Kluyveromyces marxianus)和红法夫酵母(Phaffia rhodozyma)的细胞。

[0272] 丝状真菌的合适宿主细胞包括真菌(Eumycotina)亚门的全部丝状形式。丝状真菌属的合适细胞包括但不限于支顶孢属、曲霉属、出芽短梗霉属(Aureobasidium)、黑管菌属(Bjerkandera)、拟蜡菌属(Ceriporiopsis)、金孢属(Chrysoporium)、鬼伞属(Coprinus)、革盖菌属(Coriolus)、棒囊壳属 (Corynascus)、毛壳菌属(Chaertomium)、隐球酵母属(Cryptococcus)、线黑粉酵母属(Filobasidium)、镰孢属(Fusarium)、赤霉菌属(Gibberella)、腐质霉属(Humicola)、大角间座壳菌属(Magnaporthe)、毛霉菌属(Mucor)、毁丝菌属(Myceliophthora)、毛霉菌属、新美鞭菌属(Neocallimastix)、脉孢菌属(Neurospora)、拟青霉属(Paecilomyces)、青霉属(Penicillium)、平革菌属(Phanerochaete)、射脉菌属(Phlebia)、瘤胃壶菌属(Piromyces)、侧耳属 (Pleurotus)、柱顶孢属(Scytaldium)、裂褶菌属(Schizophyllum)、孢子丝菌属 (Sporotrichum)、篮状菌属(Talaromyces)、嗜热子囊菌属(Thermoascus)、梭孢壳属(Thielavia)、弯颈霉属(Tolypocladium)、栓菌属(Trametes)和木霉属的细胞。

[0273] 丝状真菌菌种的合适细胞包括但不限于,泡盛曲霉(Aspergillus awamori)、烟曲霉(Aspergillus fumigatus)、臭曲霉(Aspergillus foetidus)、日本曲霉(Aspergillus japonicus)、构巢曲霉(Aspergillus nidulans)、黑曲霉 (Aspergillus niger)、米曲霉(Aspergillus oryzae)、勒克瑙金孢菌、杆孢状镰孢(Fusarium bactridioides)、谷类镰孢(Fusarium cerealis)、克地镰刀菌 (Fusarium crookwellense)、黄色镰孢菌(Fusarium culmorum)、禾谷镰孢菌 (Fusarium graminearum)、禾赤镰孢菌(Fusarium graminum)、异孢镰刀菌 (Fusarium heterosporum)、合欢木镰孢(Fusarium negundi)、尖孢镰刀菌(Fusarium oxysporum)、网状镰孢(Fusarium reticulatum)、粉红镰刀菌 (Fusarium roseum)、接骨木镰刀菌(Fusarium sambucinum)、肤色镰孢 (Fusarium sarcochroum)、拟分枝孢镰刀菌(Fusarium sporotrichioides)、硫色镰刀菌(Fusarium sulphureum)、念珠镰孢菌(Fusarium torulosum)、拟丝孢镰刀菌(Fusarium trichothecioides)、镰孢霉(Fusarium venenatum)、烟管菌 (Bjerkandera adusta)、干拟蜡菌(Ceriporiopsis aneirina)、干拟蜡菌、 Ceriporiopsis caregiea、浅黄拟蜡菌(Ceriporiopsis gilvescens)、Ceriporiopsis pannocinta、Ceriporiopsis rivulosa、Ceriporiopsis subrufa、虫拟蜡菌 (Ceriporiopsis subvermispora)、灰盖鬼伞(Coprinus cinereus)、毛革盖菌 (Coriolus hirsutus)、特异腐质霉(Humicola insolens)、柔毛腐质霉(Humicola lanuginosa)、米黑毛霉(Mucor miehei)、嗜热毁丝菌(Myceliophthora thermophila)、粗糙脉孢菌(Neurospora crassa)、间型脉孢菌(Neurospora intermedia)、产紫青霉菌(Penicillium purpurogenum)、变灰青霉(Penicillium canescens)、离生青霉(Penicillium solitum)、绳状青霉(Penicillium funiculosum)、黄孢原毛平革菌(Phanerochaete chrysosporium)、射脉齿菌 (Phlebia radiate)、杏鲍菇(Pleurotus eryngii)、黄色蠕形霉(Talaromyces flavus)、太瑞斯梭孢壳霉、长绒毛栓菌(Trametes villosa)、变色栓菌 (Trametes versicolor)、哈茨木霉(Trichoderma harzianum)、康氏木霉 (Trichoderma koningii)、长枝木霉(Trichoderma longibrachiatum)、里氏木霉(Trichoderma reesei)和绿色木霉(Trichoderma viride)的细胞。

[0274] 本发明还提供了重组宿主细胞,其经工程化以表达Fv3A、Pf43A、 Fv43E、Fv39A、Fv43A、Fv43B、Pa51A、Gz43A、Fo43A、Af43A、 Pf51A、AfuXyn2、AfuXyn5、Fv43D、Pf43B、Fv43B、

Fv51A、里氏木霉 Xyn3、里氏木霉Xyn2、里氏木霉Bxl1、里氏木霉Bgl1(Tr3A)、GH61内切葡聚糖酶、里氏木霉Eg4、Pa3D、Fv3G、Fv3D、Fv3C、Tr3B、Te3A、 An3A、Fo3A、Gz3A、Nh3A、Vd3A、Pa3G或Tn3B多肽或其变体中的一种或更多种、两种或更多种、三种或更多种、四种或更多种或者五种或更多种。

[0275] 在某些实施例中，设想了表达衍生自两种或更多种纤维素酶序列和/或半纤维素酶序列的杂合酶或嵌合酶的重组宿主细胞。在某些方面，杂合酶或嵌合酶包含两种或更多种β-葡糖苷酶序列。在某些方面，第一β-葡糖苷酶序列具有至少约200个氨基酸残基长度，并且包含多肽基序SEQ ID NOs:136-148的一种或多种或全部，并且第二β-葡糖苷酶序列具有至少约 50个氨基酸残基长度并且包含选自SEQ ID NOs:149-156的多肽序列基序的一种或多种或全部。特别地，所述两种或更多种β-葡糖苷酶序列的第一种是具有至少约200个氨基酸残基长度并且包含SEQ ID NOs:164-169的氨基酸序列基序中至少2种(例如，至少2、3、4或全部)的序列，并且所述两种或更多种β-葡糖苷酶序列的第二种具有至少50个氨基酸残基长度并且包含SEQ ID NO:170。在某些实施例中，第一b葡糖苷酶序列位于杂合或嵌合多肽的N末端，而第二β-葡糖苷酶序列位于其C末端。在某些实施例中，第一种和第二β-葡糖苷酶序列互相紧邻或直接互相连接。在其他实施例中，第一种和第二β-葡糖苷酶序列并非紧邻或直接连接，而是通过接头结构域连接。在某些实施例中，接头结构域位于中央。在某些方面，第一或第二β-葡糖苷酶序列包含环序列，其长度为约3、4、5、6、7、8、 9、10或11个氨基酸残基，包含FDRRSPG的序列(SEQ ID NO:171)或 FD(R/K)YNIT的序列(SEQ ID NO:172)，与未修饰的对应多肽相比或与衍生杂合或嵌合多肽的嵌合部分的多肽相比，对所述环序列的修饰改善杂合或嵌合多肽的稳定性。在某些实施例中，第一或第二β-葡糖苷酶序列均不包含环序列，相反地，接头结构域包含环序列。在一些实施例中，对所述环序列的修饰(例如，截短、加长、缺失、替换、置换或其他方式修饰该序列)减少对该环序列中残基的切割。在其他实施例中，对该环序列的修饰减少对该环序列外部的位点处残基的切割。

[0276] 在某些实施例中，设想了表达衍生自两种或更多种纤维素酶序列和/或半纤维素酶序列的杂合酶或嵌合酶的重组宿主细胞。在某些方面，杂合酶或嵌合酶包含两种或更多种β-葡糖苷酶序列。在一些实施例中，设想了表达杂合或嵌合酶的重组宿主细胞，所述杂合或嵌合酶包含第一序列，其具有至少约200个连续氨基酸残基长度并且与SEQ ID NO:60的等长度序列具有至少60％、70％、80％、90％、91％、92％、93％、94％、95％、 96％、97％、98％、99％或更高的序列同一性；并且包含第二序列，其具有至少约50个连续氨基酸残基长度并且与SEQ ID NOs:54、56、58、62、 64、66、68、70、72、74、76、78和79中任何一种的等长度序列具有至少约60％、70％、80％、90％、91％、92％、93％、94％、95％、96％、 97％、98％、99％或更高的序列同一性。在可选的实施例中，设想了表达杂合或嵌合酶的重组宿主细胞，所述杂合或嵌合酶包含第一序列，其具有至少约200个连续氨基酸残基长度并且与SEQ ID NOs:54、56、58、60、 62、64、66、68、70、72、74、76、78和79中任何一种的等长度序列具有至少60％、70％、80％、90％、91％、92％、93％、94％、95％、96％、 97％、98％、99％或更高的序列同一性；并且包含第二序列，其具有至少约50个连续氨基酸残基长度并且与SEQ ID NO:60具有至少约60％、 70％、80％、90％、91％、92％、93％、94％、95％、96％、97％、98％、 99％或更高的序列同一性。在某些实施例中，第一β葡糖苷酶序列位于杂合或嵌合多肽的N末端，而第二β-葡糖苷酶序列位于其C末端。在某些实施例中，第一种和第二β-葡糖苷酶序列互相

紧邻或直接互相连接。在其他实施例中，第一种和第二β-葡糖苷酶序列并非紧邻或直接连接，而是通过接头结构域连接。在某些实施例中，接头结构域位于中央。在某些方面，第一或第二β-葡糖苷酶序列包含环序列，其长度为约3、4、5、6、7、8、9、10或11个氨基酸残基，包含FDRRSPG的序列(SEQ ID NO:171)或 FD(R/K)YNIT的序列(SEQ ID NO:172)，与未修饰的对应多肽相比或与衍生杂合或嵌合多肽的嵌合部分的多肽相比，对所述环序列的修饰改善杂合或嵌合多肽的稳定性。在某些实施例中，第一或第二β-葡糖苷酶序列均不包含环序列，相反地，接头结构域包含环序列。在一些实施例中，对所述环序列的修饰(例如，截短、加长、缺失、替换、置换或其他方式修饰该序列)减少对该环序列中残基的切割。在其他实施例中，对该环序列的修饰减少对该环序列外部的位点处残基的切割。

[0277] 在某些方面，该重组宿主细胞表达一种或多种嵌合酶，例如Fv3C融合酶、里氏木霉Bgl3融合酶、Fv3C/Bgl3融合酶、Te3A融合酶或 Fv3C/Te3A/Bgl3融合酶。对于本发明，术语“某某融合酶”、“某某嵌合酶”和“某某杂合酶”可互换地用于指具有来自某某酶的至少一个嵌合部分的酶。例如，Fv3C融合或嵌合酶可以指Fv3C/Bgl3杂合酶(其也是 Bgl3嵌合酶)或指Fv3C/Te3A/Bgl3杂合酶(其也是Te3A或Bgl3嵌合酶)。

[0278] 该重组宿主细胞是，例如，重组里氏木霉宿主细胞。在一个特定的例子中，本发明提供了重组真菌，如重组里氏木霉，所述重组真菌经工程化以表达Fv3A、Pf43A、Fv43E、Fv39A、Fv43A、Fv43B、Pa51A、 Gz43A、Fo43A、Af43A、Pf51A、AfuXyn2、AfuXyn5、Fv43D、Pf43B、Fv43B、Fv51A、里氏木霉Xyn3、里氏木霉Xyn2、里氏木霉Bxl1、里氏木霉Bgl1(Tr3A)、里氏木霉Bgl3(Tr3B)、GH61内切葡聚糖酶、里氏木霉 Eg4、Pa3D、Fv3G、Fv3D、Fv3C、Fv3C融合/嵌合酶、Fv3C/Bgl3、 Fv3C/Te3A/Bgl3融合/嵌合酶、Te3A、An3A、Fo3A、Gz3A、Nh3A、 Vd3A、Pa3G或Tn3B多肽或它们的变体(包括，例如，其杂合或嵌合多肽)中的1种或更多种、2种或更多种、3种或更多种、4种或更多种或者 5种或更多种。

[0279] 本发明提供了宿主细胞，例如重组真菌宿主细胞或重组丝状真菌，所述宿主细胞经工程化以重组表达至少一种木聚糖酶、至少一种β-木糖苷酶以及一种L-α-阿拉伯呋喃糖苷酶。本发明还提供了宿主细胞，例如重组真菌宿主细胞或重组丝状真菌(诸如重组里氏木霉)，除里氏木霉Xyn3、里氏木霉Xyn2、里氏木霉Bxl1、里氏木霉Bgl1、GH61内切葡聚糖酶、里氏木霉Eg4或其变体中的一种或多种外，所述宿主细胞还经工程化以表达 Fv3A、Pf43A、Fv43E、Fv39A、Fv43A、Fv43B、Pa51A、Gz43A、 Fo43A、Af43A、Pf51A、AfuXyn2、AfuXyn5、Fv43D、Pf43B、Fv43B、 Fv51A、Pa3D、Fv3G、Fv3D、Fv3C、Fv3C融合酶、里氏木霉Bgl3 (Tr3B)、里氏木霉Bgl3融合酶、Fv3C/Bgl3融合酶、Tr3A、Te3A、Te3A 融合酶、Fv3C/Te3A/Bgl3融合酶、An3A、Fo3A、Gz3A、Nh3A、Vd3A、 Pa3G或Tn3B多肽中的1种、2种、3种、4种、5种或更多种。该重组宿主细胞是，例如，里氏木霉宿主细胞。

[0280] 本发明还提供了重组宿主细胞，例如，重组真菌宿主细胞或重组生物，例如，丝状真菌(诸如重组里氏木霉)，所述重组宿主细胞经过工程化以重组表达里氏木霉Xyn3、里氏木霉Bgl1、里氏木霉Bgl3(Tr3B)、里氏木霉Bgl3融合酶、Fv3A、Fv43D和Fv51A多肽。例如，该重组宿主细胞合适地是里氏木霉宿主细胞。该重组真菌合适地是重组里氏木霉。本发明提供了，例如，里氏木霉宿主细胞，其经工程化以重组表达里氏木霉 Xyn3、里氏木霉Bgl1、里氏木霉Bgl3融合酶、Fv3A、Fv43D和Fv51A多肽。

[0281] 启动子和载体的例子

[0282] 本发明还提供了包含上述核酸的表达盒和/或载体。合适地，编码本发明的酶的核酸有效连接于启动子。启动子是本领域所熟知的。在宿主细胞内发挥功能的任何启动子均可以用于表达本发明的β-葡糖苷酶和/或任何其他核酸。可用于在多种宿主细胞中驱动本发明的β-葡糖苷酶核酸和/或任何其他核酸表达的起始控制区域或启动子数量繁多并且是本领域技术人员熟悉的(参见，例如WO 2004/033646和其引用的参考文献)。事实上可以使用能够驱动这些核酸的任何启动子。

[0283] 具体地，在期望在丝状真菌宿主中进行重组表达的情况下，该启动子可以是丝状真菌启动子。核酸可以处于，例如，异源启动子的控制下。核酸还可以在组成型或诱导型启动子控制下表达。可使用的启动子的例子包括但不限于纤维素酶启动子、木聚糖酶启动子、1818启动子(此前通过对木霉属进行EST作图鉴定为高表达的蛋白质)。例如，该启动子可以合适地为纤维二糖水解酶、内切葡聚糖酶或β-葡糖苷酶启动子。特别合适的启动子可以是，例如里氏木霉纤维二糖水解酶、内切葡聚糖酶或β-葡糖苷酶启动子。例如，该启动子是纤维二糖水解酶I(cbh1)启动子。启动子的非限制性例子包括cbh1、cbh2、egl1、egl2、egl3、egl4、egl5、pki1、gpd1、 xyn1或xyn2启动子。启动子的额外非限制性例子包括里氏木霉的cbh1、 cbh2、egl1、egl2、egl3、egl4、egl5、pki1、gpd1、xyn1或xyn2启动子。

[0284] 如本文所用，术语“有效地连接”是指选定的核苷酸序列(例如，编码本文所述的多肽)处于启动子附近以允许该启动子调控选定DNA的表达。此外，该启动子就转录和翻译的方向而言位于选定的核苷酸序列的上游。“有效地连接”是指核苷酸序列和调控序列以合适分子(例如，转录激活因子蛋白)结合于所述调控序列时允许基因表达的方式连接。

[0285] 本文所述的任何β-葡糖苷酶和/或其他核酸均可以包含于一种或多种载体中。因此，本文还描述了具有编码本发明的任何β-葡糖苷酶和/或其他核酸的一种或多种核酸的载体。在某些方面，载体含有受表达控制序列控制的核酸。在某些方面，表达控制序列是天然表达控制序列。在某些方面，表达控制序列是非天然表达控制序列。在某些方面，载体包含选择标记或可选标记。在某些方面，一种或多种β-葡糖苷酶在没有可选标记的情况下整合至细胞的染色体中。

[0286] 合适的载体是与所采用的宿主细胞相容的那些。合适的载体可以源自例如细菌、病毒(例如源自噬菌体的噬菌体T7或M-13)、粘粒、酵母或植物。合适的载体可以在宿主细胞中以低、中或高拷贝数维持。获得和使用这些载体的方案是本领域的技术人员已知的(参见例如Sambrook et al., Molecular Cloning:A Laboratory Manual,2nd ed.,Cold Spring Harbor,1989 (Sambrook等人，《分子克隆实验指南》，第二版，冷泉港实验室出版社，1989年))。

[0287] 在某些方面，表达载体还包括终止序列。终止控制区域还可源自宿主细胞中天然存在的多种基因。在某些方面，终止序列和启动子序列源自相同来源。

[0288] β-葡糖苷酶核酸可以使用标准技术(Sambrook et al.,Molecular Cloning: A Laboratory Manual,Cold Spring Harbor,1982(Sambrook等人，《分子克隆实验指南》，冷泉港实验室出版社，1982年))整合进入载体，诸如表达载体。

[0289] 在某些方面，可能想要以远高于天然存在细胞中当前存在的水平来过量表达本发明中所述的一种或多种β-葡糖苷酶和/或一种或多种任何其他核酸。在一些实施例中，可能想要以远低于天然存在细胞中当前存在的水平来不足表达(例如，突变、失活或缺失)本发

明中所述的β-葡糖苷酶和/或一种或多种任何其他核酸。

[0290] 转化方法的例子

[0291] 可以使用将DNA构建体或载体引入宿主细胞内的标准技术将β-葡糖苷酶核酸或含有它们的载体插入宿主细胞(例如,本文所述的植物细胞、真菌细胞、酵母细胞或细菌细胞)内,所述标准技术例如是转化、电穿孔、细胞核显微注射、转导、转染(例如,微脂粒感染介导的或DEAE-糊精介导的转染,或使用重组噬菌体病毒的转染)、使用磷酸钙DNA沉淀物孵育、用DNA涂覆的微抛射体进行高速轰击以及原生质体融合。通用的转化技术是本领域已知的(参见例如Current Protocols in Molecular Biology(F.M.Ausubel et al.(eds) Chapter 9,1987(《最新分子生物学实验方法汇编》,F.M.Ausubel等人编辑,第9章,1987年);Sambrook et al., Molecular Cloning:A Laboratory Manual,2nd ed.,Cold Spring Harbor,1989 (Sambrook等人,《分子克隆实验指南》,第二版,冷泉港实验室出版社,1989年);以及Campbell et al.,Curr.Genet.16:53-56,1989(Campbell 等人,《当代遗传学》,第16卷,第53-56页,1989年))。引入的核酸可以整合至染色体DNA内或作为染色体外复制型序列维持。可以通过本领域中已知的任何方法选择转化体。

[0292] 细胞培养基的例子

[0293] 一般来讲,将微生物在适用于产生本文所述的多肽的细胞培养基中培养。培养使用本领域已知的步骤和变化在合适的营养培养基中进行,所述培养基包含碳源和氮源及无机盐。适于生长和纤维素酶产生的培养基、温度范围和其他条件是本领域已知的。作为非限制性例子,用于里氏木霉产生纤维素酶的典型温度范围为24℃至28℃。

[0294] 细胞培养条件的例子

[0295] 适于维持和生长细菌培养物的材料和方法是本领域众所周知的。示例性的技术可以参见Manual of Methods for General Bacteriology Gerhardt et al., eds), American Society for Microbiology,Washington,D.C.(1994)(《普通细菌学方法手册》,Gerhardt等人编,美国微生物学会,华盛顿,1994年) 或Brock in Biotechnology:A Textbook of Industrial Microbiology,Second Edition(1989)Sinauer Associates, Inc.,Sunderland,MA(Brock,《生物技术:工业微生物学教材》,第二版(1989年),Sinauer Associates公司,马萨诸塞州桑德兰)。在某些方面,将细胞在容许由插入宿主细胞内的核酸所编码的一种或多种β-葡糖苷酶多肽表达的条件下在培养基中培养。可以使用标准细胞培养条件来培养细胞。在某些方面,将细胞在合适的温度、气体混合物和pH培育和维持。在某些方面,在合适的细胞培养基中培育细胞。

[0296] 本发明的组合物

[0297] 本发明提供了工程化的酶组合物(例如,纤维素酶组合物)或富含一种或多种上述多肽的发酵液。在某些方面,该组合物是纤维素酶组合物。该纤维素酶组合物可以是,例如,丝状真菌纤维素酶组合物,诸如木霉纤维素酶组合物。在某些方面,该组合物是包含编码一种或多种纤维素酶多肽的一种或多种核酸的细胞。在某些方面,该组合物是包含纤维素酶活性的发酵液,其中所述发酵液能够将生物质样品中存在的超过约50重量%的纤维素转化成糖。如本文所用,术语“发酵液”是指通过发酵产生的酶制品,所述酶制品在发酵后不经历或经历最最少回收和/或纯化。发酵液可以是丝状真菌(例如,木霉属、腐质霉属、镰孢属、曲霉属、脉孢菌属、青霉属、头孢霉属(Cephalosporium)、绵霉属(Achlya)、柄孢壳菌属

(Podospora)、内座壳属(Endothia)、毛霉菌属、旋孢腔菌属(Cochliobolus)、梨孢属(Pyricularia)或金孢属的发酵液。特别地,发酵液可以是,例如木霉属物种(诸如里氏木霉)或是青霉属物种(诸如绳状青霉)之一。发酵液还可以合适地是无细胞发酵液。在一个方面,本发明的任何纤维素酶、细胞、或发酵液组合物还可以包含一种或多种半纤维素酶。在一个方面,该发酵液包含完整纤维素酶。在某些实施例中,该发酵液可以在有限的生产后加工情况下使用,所述生产后加工包括例如,纯化、超滤、过滤或细胞杀灭步骤,并且因此,将这种发酵液称为以全发酵液配方使用。在某些方面,该完整纤维素酶组合物在里氏木霉中表达。在某些方面,该完整纤维素酶组合物在里氏木霉整合型菌株H3A中表达。在某些方面,该完整纤维素酶组合物在里氏木霉整合型菌株H3A中表达,其中在里氏木霉整合型菌株H3A中表达的多肽的一种或多种成分已被缺失。在某些方面,该完整纤维素酶组合物在黑曲霉或其工程化菌株中表达。在某些方面,该纤维素酶组合物通过卡尔科弗卢尔测定法进行测定能够获得至少0.1至0.4的分率产物。在某些方面,该纤维素酶组合物构成所述组合物的总酶重量的0.1至 25重量%。在某些方面,该纤维素酶组合物还包含一种或多种半纤维素酶。在某些方面,该纤维素酶组合物能够将生物质中存在的超过约70%、 75%、80%、85%、90%重量的纤维素转化成糖。在某些方面,该纤维素酶组合物包含多肽,其中生物质样品中被转化成糖的纤维素的重量百分比相对于并不包含所述多肽的纤维素酶组合物改善。

[0298] 在某些方面,该组合物是纤维素酶组合物,其包含与SEQ ID NOs:54、56、58、60、62、64、66、68、70、72、74、76、78和79中任何一种氨基酸序列具有至少约60%(例如,至少约65%、70%、75%、 80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%)序列同一性的多肽。在某些方面,该纤维素酶组合物包含与SEQ ID NOs:54、56、58、60、62、64、66、68、70、72、74、76、78和79中任何一种氨基酸序列具有至少约60%(例如,至少约65%、70%、75%、 80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%)序列同一性的多肽,其中该纤维素酶组合物能够将生物质底物中存在的超过约30重量%(例如,超过约40重量%、45重量%、50重量%、 55重量%、60重量%、65重量%、70重量%、75重量%或80重量%)的纤维素转化成糖。在某些实施例中,该生物质底物是固体、凝胶、半液体或液体形式的混合物,这通常由于使该生物质底物经历某些合适的预处理过程,诸如本文所述的预处理过程所致。在某些方面,该纤维素酶组合物,其包含与SEQ ID NO:54、56、58、60、62、64、66、68、70、72、74、 76、78和79的氨基酸序列具有至少约60%(例如,至少约65%、70%、 75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%)序列同一性的多肽并且能够将生物质样品中存在的超过约30 重量%(例如,超过约40重量%、45重量%、50重量%、55重量%、60重量%、65重量%、70重量%、75重量%或80重量%)的纤维素转化成糖,是完整细胞组合物。在某些方面,该纤维素酶组合物是发酵液,其包含与 SEQ ID NOs:54、56、58、60、62、64、66、68、70、72、74、76、78和 79中任何一种的氨基酸序列具有至少约60%(例如,至少约65%、70%、 75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、 98%或99%)序列同一性的多肽,其中所述纤维素酶组合物能够将生物质样品中存在的超过约30重量%(例如,超过约40重量%、45重量%、50 重量%、55重量%、60重量%、65重量%、70重量%、75重量%或80重量%)的纤维素转化成糖。在某些方面,该发酵液包含完整纤维素酶。在某些方面,该发酵液是无细胞发酵液。在某些方面,包含与SEQ ID NO:54、56、58、60、62、64、66、68、70、72、74、76、78和79的氨基酸序列

具有至少约60％(例如,至少约65％、70％、75％、80％、85％、 90％、91％、92％、93％、94％、95％、96％、97％、98％或99％)序列同一性的多肽的纤维素酶组合物在里氏木霉中表达。在某些方面,包含与 SEQ ID NOs:54、56、58、60、62、64、66、68、70、72、74、76、78和 79中任何一种氨基酸序列具有至少约60％(例如,至少约65％、70％、 75％、80％、85％、90％、91％、92％、93％、94％、95％、96％、97％、 98％或99％)序列同一性的多肽的纤维素酶组合物在里氏木霉整合型菌株 H3A中表达。在某些方面,在里氏木霉整合型菌株H3A中表达的多肽的一种或多种成分已被缺失。在某些方面,包含与SEQ ID NOs:54、56、 58、60、62、64、66、68、70、72、74、76、78和79的至少一个氨基酸序列具有至少约60％(例如,至少约65％、70％、75％、80％、85％或 90％)序列同一性的多肽的纤维素酶组合物在黑曲霉或其工程化菌株中表达。在某些方面,包含与SEQ ID NOs:54、56、58、60、62、64、66、 68、70、72、74、76、78和79中任何一种氨基酸序列具有至少约60％ (例如,至少约65％、70％、75％、80％、85％或90％)序列同一性的多肽的纤维素酶组合物能够获得至少0.1至0.4的分率产物,如通过卡尔科弗卢尔测定法测定。在某些方面,包含与SEQ ID NOs:54、56、58、60、62、64、66、68、70、72、74、76、78和79的至少一个氨基酸序列具有至少约60％(例如,至少约65％、70％、75％、80％、85％或90％)序列同一性的多肽的纤维素酶组合物构成所述组合物的蛋白总重量的0.1至25重量％(例如,0.5至22重量％、1至20重量％、5至19重量％、7至18重量％、9至17重量％、10至15重量％)。在某些方面,包含与SEQ ID NOs:54、56、58、60、62、64、66、68、70、72、74、76、78和79的至少一个氨基酸序列具有至少约60％(例如,至少约65％、70％、75％、80％、85％或90％)序列同一性的多肽的纤维素酶组合物还包含一种或多种半纤维素酶。在某些方面,包含与SEQ ID NOs:54、56、58、60、62、 64、66、68、70、72、74、76、78和79的至少一个氨基酸序列具有至少约60％(例如,至少约65％、70％、75％、80％、85％或90％)序列同一性的多肽的纤维素酶组合物能够将生物质中存在的超过约50％(例如,超过约55％、60％、65％、70％、75％、80％、85％或90％)重量的纤维素转化成糖。在某些方面,该纤维素酶组合物包含与SEQ ID NOs:54、56、 58、60、62、64、66、68、70、72、74、76、78和79的至少一个氨基酸序列具有至少约60％(例如,至少约65％、70％、75％、80％、85％或 90％)序列同一性的多肽,其中生物质样品中被转化成糖的纤维素的重量百分比相对于并不包含所述多肽的纤维素酶组合物改善。

[0299] 在某些方面,该纤维素酶组合物是非天然存在的纤维素酶组合物,其包含两种或更多种β-葡糖苷酶序列的嵌合体/杂合体/融合体,其中第一β- 葡糖苷酶序列具有至少约200个氨基酸残基长度并且包含与Fv3C的等长度(等于第一β-葡糖苷酶序列)连续序列(SEQ ID NO:60)约60％(例如,约65％、70％、75％、80％)或更高的序列同一性,并且其中第二β-葡糖苷酶序列具有至少约50个氨基酸残基长度并且包含与SEQ ID NOs:54、 56、58、62、64、66、68、70、72、74、76、78和79中任何一种的等长度(等于第二β-葡糖苷酶序列)连续序列至少60％(例如,至少约65％、 70％、75％、80％)的序列同一性,或包含多肽序列基序SEQ ID NO:170。在某些方面,第一β-葡糖苷酶序列在嵌合多肽的N末端处,而第二β-葡糖苷酶序列在嵌合多肽的C末端处。在某些方面,该纤维素酶组合物是完整细胞组合物。在某些方面,该纤维素酶组合物是发酵液。在某些方面,该发酵液包含完整纤维素酶。在某些方面,该发酵液是无细胞发酵液。

[0300] 在某些方面,该纤维素酶组合物是非天然存在的纤维素酶组合物,其包含两种或

更多种β-葡糖苷酶序列的嵌合体或杂合体，其中第一β-葡糖苷酶序列具有至少约200个氨基酸残基长度并且包含与SEQ ID NOs:54、56、58、62、64、66、68、70、72、74、76、78和79中任何一种的等长度(等于第一β-葡糖苷酶序列)连续序列约60％(例如，约65％、70％、75％、80％)或更高的序列同一性，或包含多肽序列基序SEQ ID NOs:164- 169的一种或多种或全部，并且其中第二β-葡糖苷酶序列具有至少约50个氨基酸残基长度，并且包含与Fv3C的等长度(等于第二β-葡糖苷酶序列)连续序列(SEQ ID NO:60)至少60％(例如，至少约65％、70％、75％、80％)的序列同一性。在某些方面，第一β-葡糖苷酶序列在嵌合多肽的N末端处，而第二β-葡糖苷酶序列在嵌合多肽的C末端处。在某些方面，该纤维素酶组合物是发酵液。在某些方面，该发酵液包含完整纤维素酶。在某些方面，该发酵液是无细胞发酵液。

[0301] 在某些实施例中，第一β-葡糖苷酶序列和第二β-葡糖苷酶序列直接相邻或连接。在一些实施例中，第一β-葡糖苷酶序列和第二β-葡糖苷酶序列并非直接相邻，而是通过接头结构域连接。在某些实施例中，接头结构域位于杂合或嵌合β-葡糖苷酶多肽的中央(即，并非位于N末端或C末端)。在某些实施例中，第一β-葡糖苷酶序列或第二β-葡糖苷酶序列或这两种序列均包含一个或多个糖基化位点。在某些实施例中，第一β-葡糖苷酶序列或第二β-葡糖苷酶序列包含环序列，所述环序列具有例如，约3、4、5、6、7、8、9、10或11个氨基酸残基长度，包含FDRRSPG的序列 (SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。在某些实施例中，该环序列提供了连接第一种和第二β-葡糖苷酶序列的接头序列。在某些方面，该纤维素酶组合物是完整细胞组合物。在某些方面，该纤维素酶组合物是发酵液。在某些方面，该发酵液包含完整纤维素酶。在某些方面，该发酵液是无细胞发酵液。

[0302] 在某些方面，该纤维素酶组合物是非天然存在的纤维素酶组合物，其包含两种或更多种β-葡糖苷酶序列的嵌合体或杂合体，其中第一β-葡糖苷酶序列具有至少约200个氨基酸残基长度并且包含与Fv3C的等长度(等于第一β-葡糖苷酶序列)连续序列(SEQ ID NO:60)约60％(例如，约 65％、70％、75％、80％)或更高的序列同一性，并且其中第二β-葡糖苷酶序列具有至少约50个氨基酸残基长度并且包含与SEQ ID NOs:54、56、 58、62、64、66、68、70、72、74、76、78和79中任何一种的等长度 (等于第二β-葡糖苷酶序列)连续序列至少60％(例如，至少约65％、 70％、75％、80％)的序列同一性，或包含多肽序列基序SEQ ID NO:170。在某些方面，第一β-葡糖苷酶序列在嵌合多肽的N末端处，而第二β-葡糖苷酶序列在嵌合多肽的C末端处。在某些实施例中，第一β-葡糖苷酶序列和第二β-葡糖苷酶序列直接相邻或连接。在一些实施例中，第一β-葡糖苷酶序列和第二β-葡糖苷酶序列并非直接相邻，而是通过接头结构域连接。在某些实施例中，接头结构域位于杂合或嵌合β-葡糖苷酶多肽的中央 (即，并非位于N末端或C末端)。在某些实施例中，第一β-葡糖苷酶序列或第二β-葡糖苷酶序列或这两种序列均包含一个或多个糖基化位点。在某些实施例中，第一β-葡糖苷酶序列或第二β-葡糖苷酶序列包含环序列，所述环序列具有例如约3、4、5、6、7、8、9、10或11个氨基酸残基长度，包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。在某些实施例中，该环序列提供了连接第一种和第二β-葡糖苷酶序列的接头序列。在某些方面，该纤维素酶组合物是完整细胞组合物。在某些方面，该纤维素酶组合物是发酵液。在某些方面，该发酵液包含完整纤维素酶。

[0303] 在某些方面，该发酵液是无细胞发酵液。在某些方面，该纤维素酶组合物是非天然存在的纤维素酶组合物，其包含两种或更多种β-葡糖苷酶序列的嵌合体或杂合体，其中第

一β-葡糖苷酶序列是长度为至少约200(例如,至少约250、300、350、400或450)个连续氨基酸残基的序列,其包含SEQ ID NOs:136-148的氨基酸序列基序的一种或多种或全部;而第二β- 葡糖苷酶序列是长度为至少约50(例如,至少约50、75、100、120、 150、180、200、220或250)个连续氨基酸残基的序列,其包含SEQ ID NOs:149-156的氨基酸序列基序的一种或多种或全部。特别地,所述两种或更多种β-葡糖苷酶序列的第一种是具有至少约200个氨基酸残基长度并且包含SEQ ID NOs:164-169的氨基酸序列基序中的至少2种(例如,至少 2、3、4或全部)的序列,并且所述两种或更多种β-葡糖苷酶序列的第二种具有至少50个氨基酸残基长度并且包含SEQ ID NO:170。在某些方面,第一β-葡糖苷酶序列在嵌合多肽的N末端处,而第二β-葡糖苷酶序列在嵌合多肽的C末端处。在某些实施例中,第一β-葡糖苷酶序列和第二β-葡糖苷酶序列直接相邻或连接。在一些实施例中,第一β-葡糖苷酶序列和第二β-葡糖苷酶序列并非直接相邻,而是通过接头结构域连接。在某些实施例中,接头结构域位于杂合或嵌合β-葡糖苷酶多肽的中央(即,并非位于N 末端或C末端)。在某些实施例中,第一β-葡糖苷酶序列或第二β-葡糖苷酶序列或这两种序列均包含一个或多个糖基化位点。在某些实施例中,第一β-葡糖苷酶序列或第二β-葡糖苷酶序列包含环序列,所述环序列具有例如约3、4、5、6、7、8、9、10或11个氨基酸残基长度,包含FDRRSPG 的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。在某些实施例中,该环序列提供了连接第一种和第二β-葡糖苷酶序列的接头序列。在某些方面,该纤维素酶组合物是完整细胞组合物。在某些方面,该纤维素酶组合物是发酵液。在某些方面,该发酵液包含完整纤维素酶。在某些方面,该发酵液是无细胞发酵液。

[0304] 半纤维素酶组合物

[0305] 在某些方面,本发明的任何纤维素酶组合物还包含一种或多种半纤维素酶。在这种情况下,该纤维素酶组合物也是半纤维素酶组合物。在某些方面,本发明的半纤维素酶组合物包含选自木聚糖酶、β-木糖苷酶、L-α- 阿拉伯呋喃糖苷酶及其组合的半纤维素酶。在某些方面,本发明的半纤维素酶组合物包含至少一种木聚糖酶。在某些方面,所述至少一种木聚糖酶选自里氏木霉Xyn2、里氏木霉Xyn3、AfuXyn2和AfuXyn5。在某些方面,本发明的半纤维素酶组合物包含至少一种β-木糖苷酶。在某些方面,β-木糖苷酶包含1类β-木糖苷酶,其选自诸如Fv3A和Fv43A这样的β-木糖苷酶。在某些方面,β-木糖苷酶包含2类β-木糖苷酶,其选自诸如 Pf43A、Fv43D、Fv39A、Fv43E、Fo43E、Fv43B、Pa51A、Gz43A和里氏木霉Bxl1这样的β-木糖苷酶。在某些方面,本发明的纤维素酶组合物包含单一的β-木糖苷酶,其选自1类或2类β-木糖苷酶。在某些方面,本发明的纤维素酶组合物包含两种β-木糖苷酶,其中一种β-木糖苷酶选自1类而另一种选自2类。在某些方面,本发明的半纤维素酶组合物包含至少一种L-α-阿拉伯呋喃糖苷酶。在某些方面,所述至少一种L-α-阿拉伯呋喃糖苷酶选自Af43A、Fv43B、Pf51A、Pa51A和Fv51A。

[0306] 木聚糖酶:在某些方面,该纤维素酶组合物是半纤维素酶组合物,其包含至少一种合适的木聚糖酶。在某些方面,所述至少一种木聚糖酶选自里氏木霉Xyn2、里氏木霉Xyn3、AfuXyn2和AfuXyn5。

[0307] 任何木聚糖酶(EC 3.2.1.8)可以作为所述一种或多种木聚糖酶使用。合适的木聚糖酶包括,例如,解糖热纤维菌(Caldocellum saccharolyticum)木聚糖酶(Luthi et al.1990,Appl.Environ.Microbiol.56(9):2677-2683(Luthi 等人,1990年,《应用环境微

生物学》,第56卷,第9期,第2677-2683 页))、海栖热袍菌(Thermatoga maritima)木聚糖酶(Winterhalter&Liebel, 1995,Appl.Environ.Microbiol.61(5):1810-1815(Winterhalter和Liebel, 1995年,《应用环境微生物学》,第61卷,第5期,第1810-1815页))、热袍菌属(Thermatoga Sp.)菌株FJSS-B.1木聚糖酶(Simpson et al. 1991,Biochem.J.277,413-417(Simpson等人,1991年,《生物化学杂志》,第277卷,第413-417页))、环状芽胞杆菌(Bacillus circulans)木聚糖酶(BcX)(美国专利No.5,405,769)、黑曲霉木聚糖酶(Kinoshita et al. 1995,Journal of Fermentation and Bioengineering 79(5):422-428(Kinoshita等人,1995年,《发酵与生物工程杂志》,第79卷,第5期,第422-428 页))、浅青紫链霉菌木聚糖酶(Shareck et al.1991,Gene 107:75-82 (Shareck等人,1991年,《基因》,第107卷,第75-82页);Morosoli et al.1986Biochem.J.239:587-592(Morosoli等人,1986年,《生物化学杂志》,第239卷,第587-592页);Kluepfel et al.1990,Biochem.J. 287:45-50(Kluepfel等人,1990年,《生物化学杂志》,第287卷,第 45-50页))、枯草芽孢杆菌木聚糖酶(Bernier et al.1983,Gene 26(1):59- 65(Bernier等人,1983年,《基因》,第26卷,第1期,第59-65 页))、粪碱纤维单胞菌(Cellulomonas fimi)木聚糖酶(Clarke et al.,1996, FEMS Microbiology Letters 139:27-35(Clarke等人,1996年,《FEMS微生物学通讯》,第139卷,第27-35页))、荧光假单胞菌(Pseudomonas fluorescens)木聚糖酶(Gilbert et al.1988,Journal of General Microbiology 134:3239-3247(Gilbert等人,1988年,《普通微生物学杂志》,第134 卷,第3239-3247页))、热纤维梭菌(Clostridium thermocellum)木聚糖酶 (Dominguez et al.,1995,Nature Structural Biology 2:569-576(Dominguez等人,1995年,《自然—结构生物学》,第2卷,第569-576页))、短小芽孢杆菌(Bacillus pumilus)木聚糖酶(Nuyens et al.Applied Microbiology and Biotechnology 2001,56:431-434(Nuyens等人,《应用微生物学和生物技术》,2001年,第56卷,第431-434页);Yang et al.1998,Nucleic Acids Res.16(14B):7187(Yang等人,1998年,《核酸研究》,第16卷,第14B期,第7187页))、丙酮丁醇梭杆菌(Clostridium acetobutylicum) P262木聚糖酶(Zappe et al.1990,Nucleic Acids Res.18(8):2179(Zappe等人,1990年,《核酸研究》,第18卷,第8期,第2179页))或哈茨木霉木聚糖酶(Rose et al.1987,J.Mol.Biol.194(4):755-756(Rose等人,1987 年,《分子生物学杂志》,第194卷,第4期,第755-756页))。

[0308] Xyn2:在某些方面,本发明的纤维素酶组合物还包含Xyn2。里氏木霉Xyn2的氨基酸序列(SEQ ID NO:43)在图25和59B中示出。SEQ ID NO:43是未成熟里氏木霉Xyn2的序列。里氏木霉Xyn2具有对应于SEQ ID NO:43的第1位至第33位残基(在图25中以下划线标出)的预测前肽原序列;切割第16位和第17位之间预测的信号序列据预测产生前肽,其由kexin样蛋白酶在第32和33位之间加工,产生具有与SEQ ID NO:43的第33位至第222位残基对应的序列的成熟蛋白。预测的保守结构域在图 25中以粗体标出。当里氏木霉Xyn2作用于预处理的生物质或分离的半纤维素时,通过观察到该酶在木二糖酶存在下催化增加的木糖单体产生能力,间接地显示里氏木霉Xyn2具有木聚糖内切酶活性。保守的酸性残基包括E118、E123和E209。如本文所用,“里氏木霉Xyn2多肽”是指包含以下序列的多肽和/或其变体,所述序列与SEQ ID NO:43的第33位至第 222位残基中至少50、75、100、125、150或175个连续氨基酸残基具有至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、 95%、

96％、97％、98％、99％或100％序列同一性。与天然的里氏木霉 Xyn2相比，里氏木霉Xyn2多肽优选地在残基E118、E123和E209处未改变。里氏木霉Xyn2多肽优选地在里氏木霉Xyn2、AfuXyn2和AfuXyn5之间保守的至少70％、80％、90％、95％、98％或99％的氨基酸残基处未改变，如图59B的比对结果中示出。里氏木霉Xyn2多肽合适地包含图25中示出的天然里氏木霉Xyn2的整个预测保守结构域。示例性的里氏木霉 Xyn2多肽包含与图25中示出的成熟里氏木霉Xyn2序列具有至少85％、 86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、 97％、98％、99％或100％同一性的序列。本发明的里氏木霉Xyn2多肽优选地具有木聚糖酶活性。

[0309] Xyn3：在某些方面，本发明的纤维素酶组合物还包含Xyn3。里氏木霉Xyn3的氨基酸序列(SEQ ID NO:42)在图24B中示出。SEQ ID NO:42是未成熟里氏木霉Xyn3的序列。里氏木霉Xyn3具有对应于SEQ ID NO:42 的第1位至第16位残基的预测信号序列(在图24中以下划线标出)；对该信号序列的切割预计会产生成熟的蛋白质，该蛋白质具有对应于SEQ ID NO:42的第17位至第347位残基的序列。预测的保守结构域在图24B中以粗体标出。当里氏木霉Xyn3作用于预处理的生物质或分离的半纤维素时，通过观察到该酶在木二糖酶存在下催化增加的木糖单体产生能力，间接地显示里氏木霉Xyn3具有木聚糖内切酶活性。保守的催化残基包括 E91、E176、E180、E195和E282，如通过与另一个GH10家族的酶(来自霍耳斯特德链霉菌(Streptomyces halstedii)的Xys1δ(Canals et al.,2003,Act Crystalogr.D Biol.59:1447-53(Canals等人，2003年，《晶体学报》D卷生物类，第59卷，第1447-1453页))比对所测定，所述酶与里氏木霉 Xyn3具有33％的序列同一性。如本文所用，“里氏木霉Xyn3多肽”是指包含以下序列的多肽和/或其变体，所述序列与SEQ ID NO:42的第17位至第347位残基中至少50、75、100、125、150、175、200、250或300个连续氨基酸残基具有至少85％、86％、87％、88％、89％、90％、91％、 92％、93％、94％、95％、96％、97％、98％、99％或100％序列同一性。与天然的里氏木霉Xyn3相比，里氏木霉Xyn3多肽优选地在残基E91、 E176、E180、E195和E282处未改变。里氏木霉Xyn3多肽优选地在里氏木霉Xyn3和Xys1δ之间保守的至少70％、80％、90％、95％、98％或99％的氨基酸残基处未改变。里氏木霉Xyn3多肽合适地包含图24B中示出的天然里氏木霉Xyn3的整个预测保守结构域。示例性的里氏木霉Xyn3多肽包含与图24B中示出的成熟里氏木霉Xyn3序列具有至少85％、86％、 87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、 98％、99％或100％同一性的序列。本发明的里氏木霉Xyn3多肽优选地具有木聚糖酶活性。

[0310] AfuXyn2：在某些方面，本发明的纤维素酶组合物还包含AfuXyn2。AfuXyn2的氨基酸序列(SEQ ID NO:24)在图19B和59B中示出。SEQ ID NO:24是未成熟AfuXyn2的序列。AfuXyn2具有对应于SEQ ID NO:24的第1位至第18位残基的预测信号序列(在图19B中以下划线标出)；对该信号序列的切割预计会产生成熟的蛋白质，该蛋白质具有对应于SEQ ID NO:24的第19位至第228位残基的序列。预测的GH11保守结构域在图 19B中以粗体标出。当AfuXyn2作用于预处理的生物质或分离的半纤维素时，通过观察到该酶在木二糖酶存在下催化增加的木糖单体产生的能力，间接地显示AfuXyn2具有木聚糖内切酶活性。保守的催化性残基包括 E124、E129和E215。如本文所用，“AfuXyn2多肽”是指包含以下序列的多肽和/或其变体，所述序列与SEQ ID NO:24第19至228位残基中至少 50、75、100、125、150、175或200个连续氨基酸残基具有至少85％、 86％、87％、88％、89％、90％、91％、92％、93％、

94％、95％、96％、 97％、98％、99％或100％序列同一性。与天然的AfuXyn2相比，AfuXyn2多肽优选地在残基E124、E129和E215处未改变。AfuXyn2多肽优选地在 AfuXyn2、AfuXyn5和里氏木霉Xyn2之间保守的至少70％、80％、90％、 95％、98％或99％的氨基酸残基处未改变，如图59B中示出。AfuXyn2多肽合适地包含图19B中示出的天然AfuXyn2的整个预测保守结构域。示例性的AfuXyn2多肽包含与图19B中示出的成熟AfuXyn2序列具有至少 85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、 96％、97％、98％、99％或100％同一性的序列。本发明的AfuXyn2多肽优选地具有木聚糖酶活性。

[0311] AfuXyn5：在某些方面，本发明的纤维素酶组合物还包含AfuXyn5。AfuXyn5的氨基酸序列(SEQ ID NO:26)在图20B和59B中示出。SEQ ID NO:26是未成熟AfuXyn5的序列。AfuXyn5具有对应于SEQ ID NO:26的第1位至第19位残基的预测信号序列(在图20B中以下划线标出)；对该信号序列的切割预计会产生成熟的蛋白质，该蛋白质具有对应于SEQ ID NO:26的第20位至第313位残基的序列。预测的GH11保守结构域在图 20B中以粗体标出。当AfuXyn5作用于预处理的生物质或分离的半纤维素时，通过观察到该酶在木二糖酶存在下催化增加的木糖单体产生的能力，间接地显示AfuXyn5具有木聚糖内切酶活性。保守的催化性残基包括 E119、E124和E210。所预测的CBM处于C末端附近，特征在于大量的疏水残基，并且接在长的富含丝氨酸和苏氨酸的氨基酸串之后。该区域在图59B中以下划线标出。如本文所用，“AfuXyn5多肽”是指包含以下序列的多肽和/或其变体，所述序列与SEQ ID NO:26的第20位至第313位残基中至少50、75、100、125、150、175、200、250或275个连续氨基酸残基具有至少85％、86％、87％、88％、89％、90％、91％、92％、93％、 94％、95％、96％、97％、98％、99％或100％序列同一性。与天然的 AfuXyn5相比，AfuXyn5多肽优选地在残基E119、E120和E210处未改变。AfuXyn5多肽优选地在AfuXyn5、AfuXyn2和里氏木霉Xyn2之间保守的至少70％、80％、90％、95％、98％或99％的氨基酸残基处未改变，如图59B中示出。AfuXyn5多肽合适地包含天然AfuXyn5的整个预测CBM 和/或天然AfuXyn5的整个预测保守结构域(以下划线标出)，其在图20B 中示出。示例性的AfuXyn5多肽包含与图20B中示出的成熟AfuXyn5序列具有至少85％、86％、87％、88％、89％、90％、91％、92％、93％、 94％、95％、96％、97％、98％、99％或100％同一性的序列。本发明的 AfuXyn5多肽优选地具有木聚糖酶活性。

[0312] 该木聚糖酶合适地构成本发明的纤维素酶组合物的约0.05重量％至约 50重量％，其中所述重量％表示相对于给定组合物中全部酶的合并重量而言木聚糖酶的合并重量。该木聚糖酶可以一定范围存在，其中下限是0.05 重量％、1重量％、1.5重量％、2重量％、3重量％、4重量％、5重量％、6 重量％、7重量％、8重量％、9重量％、10重量％、12重量％、15重量％、 20重量％、25重量％、30重量％、40重量％或45重量％，并且上限是5重量％、10重量％、15重量％、20重量％、25重量％、30重量％、35重量％、40重量％或50重量％。合适地，本发明的酶组合物中的一种或多种木聚糖酶的合并重量可以构成所述酶组合物中全部酶的总重量的例如约 0.05重量％至约50重量％(例如，0.05重量％、1重量％、2重量％、3重量％至50重量％、3重量％至40重量％、3重量％至30重量％、3重量％至 20重量％、5重量％至20重量％、10重量％至30重量％、15重量％至35 重量％、20重量％至40重量％、20重量％至50重量％等等)。

[0313] 该木聚糖酶可以通过量表达编码木聚糖酶的内源或外源基因来产生。该木聚糖酶在某些情况下可以过量表达或不足表达。

[0314] β-木糖苷酶:在某些方面,本发明的纤维素酶组合物包含至少一种β- 木糖苷酶。在某些方面,该纤维素酶组合物包含至少一种1类β-木糖苷酶,其选自,例如,Fv3A和Fv43A。在某些方面,该纤维素酶组合物包含至少一种2类β-木糖苷酶,其选自,例如,Pf43A、Fv43D、Fv39A、 Fv43E、Fo43E、Fv43B、Pa51A、Gz43A和里氏木霉Bxl1。在某些方面,该纤维素酶组合物包含单一β-木糖苷酶,并且该β-木糖苷酶选自1类或2 类。在某些方面,该纤维素酶组合物包含两种β-木糖苷酶,其中一种β-木糖苷酶选自1类而另一种选自2类。

[0315] 任何β-木糖苷酶(EC 3.2.1.37)均可以作为合适的β-木糖苷酶使用。合适的β-木糖苷酶包括,例如埃默森篮状菌Bxl1(Reen et al.2003,Biochem Biophys Res Commun.305(3):579-85(Reen等人,2003年,《生物化学与生物物理研究通讯》,第305卷,第3期,第579-585页)、嗜热脂肪地芽孢杆菌(G.stearothermophilus)β-木糖苷酶(Shallom et al.2005, Biochemistry 44:387-397(Shallom等人,2005年,《生物化学》,第44 卷,第387-397页))、嗜热柱霉(S.thermophilum)β-木糖苷酶(Zanoelo et al.2004, J.Ind.Microbiol.Biotechnol.31:170-176(Zanoelo等人,2004年,《工业微生物学与生物技术杂志》,第31卷,第170-176页))、木素木霉(T.lignorum)β-木糖苷酶(Schmidt,1998, Methods Enzymol.160:662-671 (Schmidt,1998年,《酶学方法》,第160卷,第662-671页))、泡盛曲霉(A.awamori)β-木糖苷酶(Kurakake et al.2005,Biochim.Biophys.Acta 1726:272-279(Kurakake等人,2005年,《生物化学与生物物理学报》,第1726卷,第272-279页))、花斑曲霉(A.versicolor)β-木糖苷酶 (Andrade et al.2004,Process Biochem.39:1931-1938(Andrade等人,2004 年,《过程生化》,第39卷,第1931-1938页))、链霉菌属(Streptomyces sp.)β-木糖苷酶(Pinphanichakarn et al.2004,World J. Microbiol.Biotechnol.20:727-733(Pinphanichakarn等人,2004年,《世界微生物学与生物技术杂志》,第20卷,第727-733页))、海栖热袍菌β- 木糖苷酶(Xue and Shao,2004, Biotechnol.Lett.26:1511-1515(Xue和 Shao,2004年,《生物技术通讯》,第26卷,第1511-1515页))、木霉属SYβ-木糖苷酶(Kim et al.2004,J.Microbiol.Biotechnol.14:643-645(Kim等人,2004年,《微生物学与生物技术杂志》,第14卷,第643- 645页))、黑曲霉β-木糖苷酶(Oguntimein and Reilly,1980,Biotechnol. Bioeng.22:1143-1154(Oguntimein和Reilly,《生物技术与生物工程》,第22卷,第1143-1154页))或者沃特曼青霉(P.wortmanni)β-木糖苷酶 (Matsuo et al.1987,Agric.Biol.Chem.51:2367-2379(Matsuo等人,1987年,《农业生物化学》,第51卷,第2367-2379页))。合适的β-木糖苷酶可以由宿主生物内源产生,或可以由宿主生物重组克隆和/或表达。此外,可以将合适的β-木糖苷酶添加至纯化或分离形式的纤维素酶组合物。

[0316] Fv3A:在某些方面,本发明的纤维素酶组合物包含Fv3A多肽。Fv3A 的氨基酸序列(SEQ ID NO:2)在图8B和56中示出。SEQ ID NO:2是未成熟 Fv3A的序列。Fv3A具有对应于SEQ ID NO:2第1位至第23位残基(以下划线标出)的预测信号序列;对该信号序列的切割预计会产生成熟的蛋白质,该蛋白质具有对应于SEQ ID NO:2的第24位至第766位残基的序列。预测的保守结构域在图8B中以粗体标出。在例如,使用对硝基苯基-β-木吡喃糖苷、木二糖、混合的线型木聚糖低聚物、来自半纤维素的具支链的阿糖基木聚糖低聚物或者经稀氨水预处理的玉米芯作为底物的酶测定法中,显示Fv3A具有β-木糖苷酶活性。预测的催化残基是D291,而旁侧残基S290和C292据预测参与底物结合。E175和E213在其他GH3和GH39 酶

中是保守的并且据预测具有催化功能。如本文所用，“Fv3A多肽”是指包含与SEQ ID NO:2第24至766位残基中至少50个(例如至少75、100、125、150、175、200、250、300、350、400、450、500、550、600、650或700个)连续氨基酸残基具有至少85％(例如，至少86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％、或100％)序列同一性的序列的多肽和/或其变体。与天然的 Fv3A相比，Fv3A多肽优选地在残基D291、S290、C292、E175和E213处未改变。Fv3A多肽优选地在Fv3A与里氏木霉Bxl1之间保守的至少 70％、75％、80％、85％、90％、95％、98％或99％的氨基酸残基处未改变，如图56的比对结果中示出。Fv3A多肽合适地包含图8B中示出的天然Fv3A的整个预测保守结构域。本发明的示例性Fv3A多肽包含与图8B 所示的成熟Fv3A序列具有至少85％、86％、87％、88％、89％、90％、 91％、92％、93％、94％、95％、96％、97％、98％、99％或100％同一性的序列。本发明的Fv3A多肽优选地具有β-木糖苷酶活性。

[0317] 因此，本发明的Fv3A多肽合适地包含与SEQ ID NO:2的氨基酸序列或与SEQ ID NO:2第(i)24-766、(ii)73-321、(iii)73-394、(iv)395-622、(v) 24-622或(vi)73-622位残基具有至少90％、91％、92％、93％、94％、 95％、96％、97％、98％、99％或100％序列同一性的氨基酸序列。该多肽合适地具有β-木糖苷酶活性。

[0318] Fv43A:在某些方面，本发明的纤维素酶组合物包含Fv43A多肽。Fv43A的氨基酸序列(SEQ ID NO:10)在图12B和57中提供。SEQ ID NO:10 是未成熟Fv43A的序列。Fv43A具有对应于SEQ ID NO:10第1位至第22 位残基的预测信号序列(在图12B中以下划线标出)；对该信号序列的切割预计会产生成熟的蛋白质，该蛋白质具有对应于SEQ ID NO:10的第23位至第449位残基的序列。在图12B中预测的保守结构域以粗体标出，预测CBM以大写字母表示，并且分隔CD和CBM的预测接头以斜体表示。在例如使用4-硝基苯基-β-D-木吡喃糖苷、木二糖、混合的线型木聚糖低聚物、来自半纤维素的支链的阿糖基木聚糖低聚物和/或线型木聚糖低聚物作为底物的酶测定法中，显示Fv43A具有β-木糖苷酶活性。预测的催化残基包括D34或D62之一、D148和E209。如本文所用，“Fv43A多肽”是指包含以下序列的多肽和/或其变体，所述序列与SEQ ID NO:10的第23位至第449位残基中至少50、75、100、125、150、175、200、250、300、350 或400个连续氨基酸残基具有至少85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％序列同一性。与天然的Fv43A相比，Fv43A多肽优选地在残基D34或D62、D148 和E209处未改变。Fv43A多肽优选地在包括Fv43A和图57比对结果中 1、2、3、4、5、6、7、8或全部9种其他氨基酸序列的酶家族之间保守的至少70％、80％、90％、95％、98％或99％的氨基酸残基处未改变。Fv43A 多肽合适地包含天然Fv43A的整个预测CBM和/或天然Fv43A的整个预测保守结构域和/或Fv43A的接头，如图12B中示出。示例性Fv43A多肽包含与图12B示出的成熟Fv43A序列具有至少85％、86％、87％、88％、 89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％、或 100％同一性的序列。本发明的Fv43A多肽优选地具有β-木糖苷酶活性。

[0319] 因此，本发明的Fv43A多肽合适地包含与SEQ ID NO:10的氨基酸序列或与SEQ ID NO:10第(i)23-449、(ii)23-302、(iii)23-320、(iv)23-448、 (v)303-448、(vi)303-449、(vii)321-448或(viii)321-449位残基具有至少 90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％序列同一性的氨基酸序列。该多肽合适地具有β-木糖苷酶活性。

[0320] Pf43A:在某些方面，本发明的纤维素酶组合物包含Pf43A多肽。Pf43A的氨基酸序

列(SEQ ID NO:4)在图9B和57中示出。SEQ ID NO:4是未成熟Pf43A的序列。Pf43A具有对应于SEQ ID NO:4的第1位至第20位残基的预测信号序列(在图9B中以下划线标出);对该信号序列的切割预计会产生成熟的蛋白质,该蛋白质具有对应于SEQ ID NO:4的第21位至第445位残基的序列。在图9B中,预测的保守结构域以粗体标出,预测CBM以大写字母表示,并且分隔CD和CBM的预测接头以斜体表示。在例如使用对硝基苯基-β-木吡喃糖苷、木二糖、混合的线型木聚糖低聚物或者经稀氨水预处理的玉米芯作为底物的酶测定法中、已经显示Pf43A具有β-木糖苷酶活性。预测的催化残基包括D32或D60之一、D145和 E206。在图57中以下划线标出的C末端区域是预测的CBM。如本文所用,"Pf43A多肽"是指包含以下序列的多肽和/或其变体,所述序列与 SEQ ID NO:4第21至445位残基中至少50、75、100、125、150、175、200、250、300、350或400个连续氨基酸残基具有至少85%、86%、 87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、 98%、99%或100%序列同一性。与天然的Pf43A相比,Pf43A多肽优选地在残基D32或D60、D145和E206处未改变。Pf43A多肽优选地在包括了Pf43A和图57的比对结果中1、2、3、4、5、6、7或全部8种其他氨基酸序列的蛋白家族之间发现是保守的至少70%、80%、90%、95%、98%或 99%的氨基酸残基处未改变。本发明的Pf43A多肽合适地包含两种或更多种或全部以下结构域:(1)预测的CBM,(2)预测的保守结构域,以及(3) Pf43A的接头,如图9B中示出。本发明的示例性Pf43A多肽包含与图9B 中所示的成熟Pf43A序列具有至少85%、86%、87%、88%、89%、90%、 91%、92%、93%、94%、95%、96%、97%、98%、99%、或100%同一性的序列。本发明的Pf43A多肽优选地具有β-木糖苷酶活性。

[0321] 因此,本发明的Pf43A多肽合适地包含与SEQ ID NO:4的氨基酸序列或与SEQ ID NO:4第(i)21-445、(ii)21-301、(iii)21-323、(iv)21-444,(v) 302-444,(vi)302-445、(vii)324-444或(viii)324-445位残基具有至少90%、 91%、92%、93%、94%、95%、96%、97%、98%、99%或100%序列同一性的氨基酸序列。该多肽合适地具有β-木糖苷酶活性。

[0322] Fv43D:在某些方面,本发明的纤维素酶组合物还包含Fv43D多肽。 Fv43D的氨基酸序列(SEQ ID NO:28)在图21B和57中示出。SEQ ID NO:28 是未成熟Fv43D的序列。Fv43D具有对应于SEQ ID NO:28的第1位至第 20位残基的预测信号序列(在图21B中以下划线标出);对该信号序列的切割预计会产生成熟的蛋白质,该蛋白质具有对应于SEQ ID NO:28的第 21位至第350位残基的序列。预测的保守结构域在图21B中以粗体标出。在例如使用对硝基苯基-β-木吡喃糖苷、木二糖和/或混合的线型木聚糖低聚物作为底物的酶测定法中,显示Fv43D具有β-木糖苷酶活性。预测的催化残基包括D37或D72之一、D159和E251。如本文所用,"Fv43D多肽"是指包含以下序列的多肽和/或其变体,所述序列与SEQ ID NO:28第 21至350位残基中至少50、75、100、125、150、175、200、250、300或 320个连续氨基酸残基具有至少85%、86%、87%、88%、89%、90%、 91%、92%、93%、94%、95%、96%、97%、98%、99%或100%序列同一性。与天然的Fv43D相比,Fv43D多肽优选地在残基D37或D72、D159 和E251处未改变。Fv43D多肽优选地在包括Fv43D和图57的比对结果中 1、2、3、4、5、6、7、8或全部9种其他氨基酸序列的酶群体之间保守的至少70%、80%、90%、95%、98%或99%的氨基酸残基处未改变。Fv43D 多肽合适地包含图21B中示出的天然Fv43D的整个预测CD。示例性的Fv43D多肽包含与图21B中示出的成熟Fv43D序列具有至少85%、86%、 87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、 98%、99%或100%同一性的序列。本发明

的Fv43D多肽优选地具有β-木糖苷酶活性。

[0323] 因此，本发明的Fv43D多肽合适地包含与SEQ ID NO:28的氨基酸序列或与SEQ ID NO:28第(i)20-341、(ii)21-350、(iii)107-341或(iv)107-350 位残基具有至少90％、91％、92％、93％、94％、95％、96％、97％、 98％、99％或100％序列同一性的氨基酸序列。该多肽合适地具有β-木糖苷酶活性。

[0324] Fv39A：在某些方面，本发明的纤维素酶组合物包含Fv39A多肽。Fv39A的氨基酸序列(SEQ ID NO:8)在图11B中示出。SEQ ID NO:8是未成熟Fv39A的序列。Fv39A具有对应于SEQ ID NO:8的第1位至第19位残基的预测信号序列(在图11B中以下划线标出)；对该信号序列的切割预计会产生成熟的蛋白质，该蛋白质具有对应于SEQ ID NO:8的第20位至第439位残基的序列。预测的保守结构域在图11B中以粗体标出。在例如使用对硝基苯基-β-木吡喃糖苷、木二糖或混合的线型木聚糖低聚物作为底物的酶测定法中，显示Fv39A具有β-木糖苷酶活性。Fv39A的残基E168 和E272据预测分别作为催化性酸碱和亲核物质发挥功能，这是基于上文所述的来自解糖热厌氧杆菌(Thermoanaerobacterium saccharolyticum)(Uniprot登录号P36906)和嗜热脂肪地芽孢杆菌(Geobacillus stearothermophilus)(Uniprot登录号Q9ZFM2)的GH39木糖苷酶与Fv39A 的序列比对结果。如本文所用，“Fv39A多肽”是指包含以下序列的多肽和/或其变体，所述序列与SEQ ID NO:8第20至439位残基中至少50、 75、100、125、150、175、200、250、300、350或400个连续氨基酸残基具有至少85％、86％、87％、88％、89％、90％、91％、92％、93％、 94％、95％、96％、97％、98％、99％或100％序列同一性。与天然的 Fv39A相比，Fv39A多肽优选地在残基E168和E272处未改变。Fv39A多肽优选地在包括Fv39A和来自高温产氢菌和嗜热脂肪地芽孢杆菌的木糖苷酶的酶家族之间保守的至少70％、80％、90％、95％、98％或99％的氨基酸残基处未改变(参见上文)。Fv39A多肽合适地包含图11B中示出的天然 Fv39A的整个预测保守结构域。示例性Fv39A多肽包含与图11B示出的成熟Fv39A序列具有至少85％、86％、87％、88％、89％、90％、91％、 92％、93％、94％、95％、96％、97％、98％、99％、或100％同一性的序列。本发明的Fv39A多肽优选地具有β-木糖苷酶活性。

[0325] 因此，本发明的Fv39A多肽合适地包含与SEQ ID NO:8的氨基酸序列或与SEQ ID NO:8第(i)20-439、(ii)20-291、(iii)145-291或(iv)145-439位残基具有至少90％、91％、92％、93％、94％、95％、96％、97％、98％、 99％或100％序列同一性的氨基酸序列。该多肽合适地具有β-木糖苷酶活性。

[0326] Fv43E：在某些方面，本发明的纤维素酶组合物包含Fv43E多肽。Fv43E的氨基酸序列(SEQ ID NO:6)在图10B和57中示出。SEQ ID NO:6 是未成熟Fv43E的序列。Fv43E具有对应于SEQ ID NO:6的第1位至第18 位残基的预测信号序列(在图10B中以下划线标出)；对该信号序列的切割预计会产生成熟的蛋白质，该蛋白质具有对应于SEQ ID NO:6的第19 位至第530位残基的序列。预测的保守结构域在图10B中以粗体标出。在例如使用4-硝基苯基-β-D-木吡喃糖苷、木二糖和混合的线型木聚糖低聚物或者经稀氨水预处理的玉米芯作为底物的酶测定法中，显示Fv43E具有β- 木糖苷酶活性。预测的催化残基包括D40或D71之一、D155和E241。如本文所用，“Fv43E多肽”是指包含以下序列的多肽和/或其变体，所述序列与SEQ ID NO:6第19至530位残基中至少50、75、100、125、150、 175、200、250、300、350、400、450或500个连续氨基酸残基具有至少 85％、86％、87％、88％、89％、90％、91％、92％、93％、

94％、95％、 96％、97％、98％、99％或100％序列同一性。与天然的Fv43E相比， Fv43E多肽优选地在残基D40或D71、D155和E241处未改变。Fv43E多肽优选地在包括Fv43E和图57的比对结果中1、2、3、4、5、6、7或全部8种其他氨基酸序列的酶家族中已经发现为保守的至少70％、80％、 90％、95％、98％或99％的氨基酸残基处未改变。Fv43E多肽合适地包含图 10B中示出的天然Fv43E的整个预测保守结构域。示例性Fv43E多肽包含与图10B中示出的成熟Fv43E序列具有至少85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％、或 100％同一性的序列。本发明的Fv43E多肽优选地具有β-木糖苷酶活性。

[0327] 因此，本发明的Fv43E多肽合适地包含与SEQ ID NO:6的氨基酸序列或与SEQ ID NO:6第(i)19-530、(ii)29-530、(iii)19-300或(iv)29-300位残基具有至少90％、91％、92％、93％、94％、95％、96％、97％、98％、 99％或100％序列同一性的氨基酸序列。该多肽合适地具有β-木糖苷酶活性。

[0328] Fv43B：在某些方面，本发明的纤维素酶组合物包含Fv43B多肽。Fv43B的氨基酸序列(SEQ ID NO:12)在图13B和57中示出。SEQ ID NO:12 是未成熟Fv43B的序列。Fv43B具有对应于SEQ ID NO:12的第1位至第 16位残基的预测信号序列(在图13B中以下划线标出)；对该信号序列的切割预计会产生成熟的蛋白质，该蛋白质具有对应于SEQ ID NO:12的第17位至第574位残基的序列。预测的保守结构域在图13B中以粗体标出。在例如使用4-硝基苯基-β-D-木吡喃糖苷和对硝基苯基-α-L-阿拉伯呋喃糖苷作为底物的第一酶测定法中，显示Fv43B具有β-木糖苷酶和L-α-阿拉伯呋喃糖苷酶活性。在第二酶测定法中，显示该酶催化阿拉伯糖从支链的阿拉伯糖-木聚糖低聚物中的释放并且在其他木糖苷酶存在的情况下催化木糖从低聚物混合物中释放的增加。预测的催化残基包括D38或D68之一、D151和E236。如本文所用，“Fv43B多肽”是指包含以下序列的多肽和/ 或其变体，所述序列与SEQ ID NO:12第17至574位残基中至少50、75、 100、125、150、175、200、250、300、350、400、450、500或550个连续氨基酸残基具有至少85％、86％、87％、88％、89％、90％、91％、 92％、93％、94％、95％、96％、97％、98％、99％或100％序列同一性。与天然的Fv43B相比，Fv43B多肽优选地在残基D38或D68、D151和E236 处未改变。Fv43B多肽优选地在包括Fv43B和图57的比对结果中1、2、 3、4、5、6、7、8或全部9种其他氨基酸序列的酶家族之间保守的至少 70％、80％、90％、95％、98％或99％的氨基酸残基处未改变。Fv43B多肽合适地包含图13B和57中示出的天然Fv43B的整个预测保守结构域。示例性的Fv43B多肽包含与图13B中示出的成熟Fv43B序列具有至少 85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％同一性的序列。本发明的Fv43B多肽优选地具有β-木糖苷酶活性、L-α-阿拉伯呋喃糖苷酶活性或同时具有β-木糖苷酶和L-α-阿拉伯呋喃糖苷酶活性。

[0329] 因此，本发明的Fv43B多肽合适地包含与SEQ ID NO:12的氨基酸序列或与SEQ ID NO:12第(i)17-574、(ii)27-574、(iii)17-303或(iv)27-303位残基具有至少90％、91％、92％、93％、94％、95％、96％、97％、98％、 99％或100％序列同一性的氨基酸序列。该多肽合适地具有β-木糖苷酶活性、L-α-阿拉伯呋喃糖苷酶活性，或同时具有β-木糖苷酶和L-α-阿拉伯呋喃糖苷酶活性。

[0330] Pa51A：在某些方面，本发明的纤维素酶组合物包含Pa51A多肽。Pa51A的氨基酸序列(SEQ ID NO:14)在图14B和58中示出。SEQ ID NO:14 是未成熟Pa51A的序列。Pa51A具有

对应于SEQ ID NO:14的第1位至第 20位残基的预测信号序列(在图14B中以下划线标出)；对该信号序列的切割预计会产生成熟的蛋白质，该蛋白质具有对应于SEQ ID NO:14的第2位至第676位残基的序列。预测的L-α-阿拉伯呋喃糖苷酶保守结构域在图 14B以粗体标出。在例如使用人工底物对硝基苯基-β-木吡喃糖苷和对硝基苯基-α-L-阿拉伯呋喃糖苷进行的酶测定法中，显示Pa51A同时具有β-木糖苷酶活性和L-α-阿拉伯呋喃糖苷酶活性。显示该酶催化了拉伯糖从支链的阿拉伯糖-木聚糖低聚物中的释放并且在其他木糖苷酶存在的情况下催化木糖从低聚物混合物中的释放增加。保守的酸性残基包括E43、D50、 E257、E296、E340、E370、E485和E493。如本文所用，“Pa51A多肽”是指包含以下序列的多肽和/或其变体，所述序列与SEQ ID NO:14第21至 676位残基中的至少50、75、100、125、150、175、200、250、300、 350、400、450、500、550、600或650个连续氨基酸残基具有至少85％、 86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、 97％、98％、99％或100％序列同一性。与天然的Pa51A相比，Pa51A多肽优选地在残基E43、D50、E257、E296、E340、E370、E485和E493处未改变。Pa51A多肽优选地在包括Pa51A、Fv51A和Pf51A的酶群体之间保守的至少70％、80％、90％、95％、98％或99％的氨基酸残基处未改变，如图58的比对结果中示出。Pa51A多肽合适地包含如图14B中示出的天然Pa51A的预测保守结构域。示例性Pa51A多肽包含与图14B中示出的成熟 Pa51A序列具有至少85％、86％、87％、88％、89％、90％、91％、92％、 93％、94％、95％、96％、97％、98％、99％、或100％同一性的序列。本发明的Pa51A多肽优选地具有β-木糖苷酶活性、L-α-阿拉伯呋喃糖苷酶活性，或同时具有β-木糖苷酶和L-α-阿拉伯呋喃糖苷酶活性。

[0331] 因此，本发明的Pa51A多肽合适地包含与SEQ ID NO:14的氨基酸序列或与SEQ ID NO:14第(i)21-676、(ii)21-652、(iii)469-652或(iv)469-676 位残基具有至少90％、91％、92％、93％、94％、95％、96％、97％、 98％、99％或100％序列同一性的氨基酸序列。该多肽合适地具有β-木糖苷酶活性、L-α-阿拉伯呋喃糖苷酶活性，或同时具有β-木糖苷酶和L-α-阿拉伯呋喃糖苷酶活性。

[0332] Gz43A：在某些方面，本发明的纤维素酶组合物包含Gz43A多肽。Gz43A的氨基酸序列(SEQ ID NO:16)在图15B和57中示出。SEQ ID NO:16是未成熟的Gz43A序列。Gz43A具有对应于SEQ ID NO:16的第1 位至第18位残基的预测信号序列(在图15B中以下划线标出)；对该信号序列的切割预计会产生成熟的蛋白质，该蛋白质具有对应于SEQ ID NO:16的第19位至第340位残基的序列。预测的保守结构域在图15B中以粗体标出。在例如使用对硝基苯基-β-木吡喃糖苷、木二糖或混合和/或线型木聚糖低聚物作为底物的酶测定法中，显示Gz43A具有β-木糖苷酶活性。预测的催化残基包括D33或D68之一、D154和E243。如本文所用，“Gz43A多肽”是指包含以下序列的多肽和/或其变体，所述序列与SEQ ID NO:16第19至340位残基中至少50、75、100、125、150、175、200、 250或300个连续氨基酸残基具有至少85％、86％、87％、88％、89％、 90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％序列同一性。与天然的Gz43A相比，Gz43A多肽优选地在残基D33或 D68、D154和E243处未改变。Gz43A多肽优选地在包括Gz43和图57的比对结果中A1、2、3、4、5、6、7、8或全部9种其他氨基酸序列的酶群体之间保守的至少70％、80％、90％、95％、98％或99％的氨基酸残基处未改变。Gz43A多肽合适地包含如图15B中示出的天然Gz43A的预测保守结构域。示例性的Gz43A多肽包含与如图15B中示出的成熟Gz43A序列具有至少85％、86％、87％、88％、89％、

90％、91％、92％、93％、 94％、95％、96％、97％、98％、99％或100％同一性的序列。本发明的Gz43A多肽优选地具有β-木糖苷酶活性。

[0333] 因此，本发明的Gz43A多肽合适地包含与SEQ ID NO:16的氨基酸序列或与SEQ ID NO:16第(i)19-340、(ii)53-340、(iii)19-383或(iv)53-383位残基具有至少90％、91％、92％、93％、94％、95％、96％、97％、98％、 99％或100％序列同一性的氨基酸序列。该多肽合适地具有β-木糖苷酶活性。

[0334] 该β-木糖苷酶合适地构成本发明的纤维素酶或半纤维素酶组合物中的酶总重量的约0重量％至约75重量％(例如，约0.1重量％至约50重量％、约1重量％至约40重量％、约2重量％至约35重量％、约5重量％至约30重量％、约10重量％至约25重量％)。任何蛋白质对子的相互比例可以基于本发明轻易地计算。设想了以通过本文公开的重量百分比可导出的任何重量比包含酶的组合物。该β-木糖苷酶含量可以处于这样一个范围内，其中下限为所述掺合物/组合物中酶总重量的约0重量％、0.05重量％、0.5重量％、1重量％、2重量％、3重量％、4重量％、5重量％、6重量％、7重量％、8重量％、9重量％、10重量％、12重量％、15重量％、20 重量％、25重量％、30重量％、40重量％、45重量％或50重量％，并且上限为所述组合物中酶总重量的约10重量％、15重量％、20重量％、25重量％、30重量％、35重量％、40重量％、50重量％、55重量％、60重量％、65重量％或70重量％。例如，该β-木糖苷酶合适地占所述组合物中酶总重量的约2重量％至约30重量％；约10重量％至约20重量％；约3 重量％至约10重量％，或约5重量％至约9重量％。

[0335] 该β-木糖苷酶可以通过量表达编码β-木糖苷酶的内源或外源基因而产生。该β-木糖苷酶在某些情况下可以过量表达或不足表达。作为另外一种选择，该β-木糖苷酶可以对宿主生物是异源的，所述β-木糖苷酶由该宿主生物重组表达。此外，可以将该β-木糖苷酶添加至纯化或分离形式的本发明的纤维素酶或半纤维素酶组合物。

[0336] L-α-阿拉伯呋喃糖苷酶：在某些方面，本发明的纤维素酶组合物包含至少一种L-α-阿拉伯呋喃糖苷酶。在某些方面，所述至少一种L-α-阿拉伯呋喃糖苷酶选自Af43A、Fv43B、Pf51A、Pa51A和Fv51A。在某些方面，Pa51A、Fv43A同时具有L-α-阿拉伯呋喃糖苷酶和β-木糖苷酶活性。

[0337] 来自任何合适生物的L-α-阿拉伯呋喃糖苷酶(EC 3.2.1.55)可以作为一种或多种L-α-阿拉伯呋喃糖苷酶使用。合适的L-α-阿拉伯呋喃糖苷酶包括例如，米曲霉的L-α-阿拉伯呋喃糖苷酶(Numan&Bhosle,J.Ind.Microbiol. Biotechnol.2006,33:247-260(Numan&Bhosle,《工业微生物学与生物技术杂志》,2006年，第33卷，第247-260页))、酱油曲霉(A.sojae) (Oshima et al.J.Appl.Glycosci.2005,52:261-265(Oshima等人,《应用糖类物质科学》,2005年，第52卷，第261-265页))、短芽孢杆菌(B. brevis)(Numan&Bhosle,J.Ind.Microbiol.Biotechnol.2006,33:247-260 (Numan&Bhosle,《工业微生物学和生物技术杂志》,2006年，第33 卷，第247-260页))、嗜热脂肪芽胞杆菌(B.stearothermophilus) (Kim et al.,J.Microbiol.Biotechnol.2004,14:474-482(Kim等人,《微生物学和生物技术杂志》,2004年，第14卷，第474-482页))、短双歧杆菌(B. breve)(Shin et al.,Appl.Environ.Microbiol.2003,69:7116-7123(Shin等人,《应用环境微生物学》,2003年，第69卷，第7116-7123页))、长双歧杆菌(B.longum)(Margolles et al.,Appl.Environ.Microbiol.2003, 69:5096-5103(Margolles等人,《应用环境微生物学》,

2003年,第69 卷,第5096-5103页))、热纤梭菌(C.thermocellum)(Taylor et al., Biochem.J.2006,395:31-37(Taylor等人,《生物化学杂志》,2006年,第395卷,第31-37页))、尖孢镰刀菌(Panagiotou et al.,Can.J. Microbiol.2003,49:639-644(Panagiotou等人,《加拿大微生物学杂志》, 2003年,第49卷,第639-644页))、尖孢镰刀菌石竹专化型(F. oxysporum f.sp.dianthi)(Numan&Bhosle,J.Ind.Microbiol.Biotechnol.2006, 33:247-260(Numan&Bhosle,《工业微生物学和生物技术杂志》,2006 年,第33卷,第247-260页))、嗜热脂肪地芽孢杆菌T-6(Shallom et al., J.Biol.Chem.2002,277:43667-43673(Shallom等人,《生物化学杂志》, 2002年,第277卷,第43667-43673页))、大麦(H.vulgare)(Lee et al.,J. Biol.Chem.2003,278:5377-5387(Lee等人,《生物化学杂志》,2003 年,第278卷,第5377-5387页))、产黄青霉菌(P.chrysogenum) (Sakamoto et al.,Biophys.Acta 2003,1621:204-210(Sakamoto等人,《生物物理学报》,2003年,第1621卷,第204-210页))、青霉属 (Rahman et al.,Can.J.Microbiol.2003,49:58-64(Rahman等人,《加拿大微生物学期刊》,2003年,第49卷,第58-64页))、P.cellulosa (Numan&Bhosle, J.Ind.Microbiol.Biotechnol.2006,33:247-260(Numan &Bhosle,《工业微生物学和生物技术》,2006年,第33卷,第247-260 页))、微小根毛霉(R.pusillus)(Rahman et al., Carbohydr.Res.2003, 338:1469-1476(Rahman等人,《碳水化合物研究》,2003年,第338卷,第1469-1476页))、教酒链霉菌(S.chartreusis)、高温紫链霉菌(S. thermoviolacus)、嗜热厌氧乙醇菌(T.ethanolicus)、T.xylanilyticus(Numan &Bhosle, J.Ind.Microbiol.Biotechnol.2006,33:247-260(Numan&Bhosle,《工业微生物学和生物技术》,2006年,第33卷,第247-260页))、禾草腥黑粉菌(T.fusca)(Tuncer and Ball,Folia Microbiol.2003,(Praha) 48:168-172(Tuncer和Ball,《Folia微生物学》,2003年,(Praha)第48 卷,第168-172页))、海栖热袍菌(Miyazaki,Extremophiles 2005,9:399- 406(Miyazaki,《极端微生物》,2005年,第9卷,第399-406页))、木霉属SY(Jung et al.Agric.Chem.Biotechnol.2005,48:7-10(Jung等人,《农业化学生物技术》,2005年,第48卷,第7-10页))、白曲霉菌(A. kawachii)(Koseki et al.,Biochim.Biophys.Acta 2006, 1760:1458-1464 (Koseki等人,《生物化学与生物物理学报》,2006年,第1760卷,第 1458-1464页))、尖孢镰刀菌石竹专化型(Chacon-Martinez et al., Physiol.Mol.Plant Pathol.2004,64:201-208(Chacon-Martinez等人,《生理分子植物病理学》,2004年,第64卷,第201-208页))、T. xylanilyticus(Debeche et al.,Protein Eng.2002,15:21-28(Debeche等人,《蛋白质工程》,2002年,第15卷,第21-28页))、特异腐质霉、巨芒(M.giganteus)(Sorensen et al.,Biotechnol.Prog.2007,23:100-107(Sorensen 等人,《生物技术进展》,2007年,第23卷,第100-107页))或萝卜(R. sativus)(Kotake et al.J.Exp.Bot.2006,57:2353-2362(Kotake等人,《实验植物学杂志》,2006年,第57卷,第2353-2362页))。合适的L-α-阿拉伯呋喃糖苷酶可以由宿主生物内源产生,或可以由宿主生物重组克隆和/ 或表达。此外,可以将合适的L-α-阿拉伯呋喃糖苷酶添加至纯化或分离形式的纤维素酶组合物。

[0338] Af43A:在某些方面,本发明的纤维素酶组合物包含Af43A多肽。Af43A的氨基酸序列(SEQ ID NO:20)在图17B和57中示出。SEQ ID NO:20 是未成熟Af43A的序列。预测的保守结构域在图17B中以粗体标出。在例如使用对硝基苯基-α-L阿拉伯呋喃糖苷作为底物的酶

测定法中，显示 Af43A具有L-α-阿拉伯呋喃糖苷酶活性。Af43A据显示催化阿拉伯糖从一组低聚物中的释放，该组低聚物因内切木聚糖酶的作用从半纤维素中释放。预测的催化残基包括D26或D58之一、D139和E227。如本文所用，“Af43A多肽”是指包含以下序列的多肽和/或其变体，所述序列与SEQ ID NO:20的至少50、75、100、125、150、175、200、250或300个连续氨基酸残基具有至少85％、86％、87％、88％、89％、90％、91％、92％、 93％、94％、95％、96％、97％、98％、99％或100％序列同一性。与天然的 Af43A相比，Af43A多肽优选地在残基D26或D58、D139和E227处未改变。Af43A多肽优选地在包括Af43A和图57的比对结果中1、2、3、4、 5、6、7、8或全部9种其他氨基酸序列的酶群体之间保守的至少70％、 80％、90％、95％、98％或99％的氨基酸残基处未改变。Af43A多肽合适地包含如图17B中示出的天然Af43A的预测保守结构域。示例性的Af43A 多肽包含与SEQ ID NO:20具有至少85％、86％、87％、88％、89％、 90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％序列同一性。本发明的Af43A多肽优选地具有L-α-阿拉伯呋喃糖苷酶活性。

[0339] 因此，本发明的Af43A多肽合适地包含与SEQ ID NO:20的氨基酸序列或与SEQ ID NO:20的残基(i)15-558或(ii)15-295具有至少90％、91％、 92％、93％、94％、95％、96％、97％、98％、99％或100％序列同一性的氨基酸序列。该多肽合适地具有L-α-阿拉伯呋喃糖苷酶活性。

[0340] Pf51A：在某些方面，本发明的纤维素酶组合物包含Pf51A多肽。 Pf51A的氨基酸序列(SEQ ID NO:22)在图18B和58中示出。SEQ ID NO:22 是未成熟Pf51A的序列。Pf51A具有对应于SEQ ID NO:22的第1位至第 20位残基的预测信号序列(在图18B中以下划线标出)；对该信号序列的切割预计会产生成熟的蛋白质，该蛋白质具有对应于SEQ ID NO:22的第21至第642位残基的序列。预测的L-α-阿拉伯呋喃糖苷酶保守结构域在图 18B中以粗体标出。在例如使用4-硝基苯基-α-L阿拉伯呋喃糖苷作为底物的酶测定法中，显示Pf51A具有L-α-阿拉伯呋喃糖苷酶活性。Pf51A据显示催化阿拉伯糖从一组低聚物中的释放，该组低聚物通过内切木聚糖酶的作用从半纤维素中释放。预测的保守酸性残基包括E43、D50、E248、E287、E331、E360、E472和E480。如本文所用，“Pf51A多肽”是指包含以下序列的多肽和/或其变体，所述序列与SEQ ID NO:22的残基21至 642中至少50、75、100、125、150、175、200、250、300、350、400、 450、500、550或600个连续氨基酸残基具有至少85％、86％、87％、 88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、 99％或100％序列同一性。与天然的Pf51A相比，Pf51A多肽优选地在残基 E43、D50、E248、E287、E331、E360、E472和E480处未改变。Pf51A多肽优选地在Pf51A、Pa51A和Fv51A之间保守的至少70％、80％、90％、 95％、98％或99％的氨基酸残基处未改变，如图58的比对结果中示出。 Pf51A多肽合适地包含如图18B中示出的天然Pf51A的预测保守结构域。示例性的Pf51A多肽包含与图18B中示出的成熟Pf51A序列具有至少 85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％同一性的序列。本发明的Pf51A多肽优选地具有L-α-阿拉伯呋喃糖苷酶活性。

[0341] 因此，本发明的Pf51A多肽合适地包含与SEQ ID NO:22的氨基酸序列或与SEQ ID NO:22第(i)21-632、(ii)461-632、(iii)21-642或(iv)461-642 位残基具有至少90％、91％、92％、93％、94％、95％、96％、97％、 98％、99％或100％序列同一性的氨基酸序列。该多肽具有L-α-阿拉伯呋喃糖苷酶活性。

[0342] Fv51A:在某些方面,本发明的纤维素酶组合物包含Fv51A多肽。Fv51A的氨基酸序列(SEQ ID NO:32)在图23B和58中示出。SEQ ID NO:32 是未成熟Fv51A的序列。Fv51A具有对应于SEQ ID NO:32的第1位至第 19位残基的预测信号序列(在图23B中以下划线标出);对该信号序列的切割预计会产生成熟的蛋白质,该蛋白质具有对应于SEQ ID NO:32的第20位至第660位残基的序列。预测的L-α-阿拉伯呋喃糖苷酶保守结构域在图23B以粗体标出。在例如使用4-硝基苯基-α-L阿拉伯呋喃糖苷作为底物的酶测定法中,显示Fv51A具有L-α-阿拉伯呋喃糖苷酶活性。Fv51A据显示催化阿拉伯糖从一组低聚物中的释放,该组低聚物通过内切木聚糖酶的作用从半纤维素中释放。保守残基包括E42、D49、E247、E286、E330、E359、E479和E487。如本文所用,"Fv51A多肽"是指包含以下序列的多肽和/或其变体,所述序列与SEQ ID NO:32的第20至660位残基中至少 50、75、100、125、150、175、200、250、300、350、400、450、500、 550、600或625个连续氨基酸残基具有至少85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或 100%序列同一性。与天然的Fv51A相比,Fv51A多肽优选地在残基E42、 D49、E247、E286、E330、E359、E479和E487处未改变。Fv51A多肽优选地在Fv51A、Pa51A和Pf51A之间保守的至少70%、80%、90%、 95%、98%或99%的氨基酸残基处未改变,如图58的比对结果中示出。 Fv51A多肽合适地包含如图23B中示出的天然Fv51A的预测保守结构域。示例性的Fv51A多肽包含与图23B中示出的成熟Fv51A序列具有至少 85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、或100%同一性的序列。本发明的Fv51A多肽优选地具有L-α-阿拉伯呋喃糖苷酶活性。

[0343] 因此,本发明的Fv51A多肽合适地包含与SEQ ID NO:32的氨基酸序列或与SEQ ID NO:32的第(i)21-660、(ii)21-645、(iii)450-645或(iv)450- 660位残基具有至少90%、91%、92%、93%、94%、95%、96%、97%、 98%、99%或100%序列同一性的氨基酸序列。该多肽合适地具有L-α-阿拉伯呋喃糖苷酶活性。

[0344] 该L-α-阿拉伯呋喃糖苷酶合适地构成本发明的纤维素酶或半纤维素酶组合物中的酶总重量的约0.05重量%至约30重量%(例如,约0.1重量%至约25重量%、约0.5重量%至约20重量%、约1重量%至约10重量%),其中所述重量%表示相对于给定组合物中的全部酶的合并重量而言L-α-阿拉伯呋喃糖苷酶的合并重量。该L-α-阿拉伯呋喃糖苷酶可以按这样的范围存在,其中下限是0.05重量%、0.5重量%、1重量%、2重量%、3重量%、4重量%、5重量%、6重量%、7重量%、8重量%、9重量%、10重量%、12重量%、15重量%、20重量%、25重量%或28重量%,并且上限是5重量%、10重量%、15重量%、20重量%、25重量%或30重量%。例如,所述一种或多种L-α-阿拉伯呋喃糖苷酶可以合适地构成本发明的纤维素酶或半纤维素酶组合物中的酶总重量的约2重量%至约30重量%(例如,约2重量%至约30重量%、约5重量%至约30重量%、约5重量%至约10重量%、约10重量%至约30重量%、约20重量%至约30重量%、约 25重量%至约30重量%、约2重量%至约10重量%、约5重量%至约15 重量%、约10重量%至约25重量%、约20重量%至约30重量%,等等)。

[0345] 该L-α-阿拉伯呋喃糖苷酶可以通过量表达编码L-α-阿拉伯呋喃糖苷酶的内源或外源基因来产生。该L-α-阿拉伯呋喃糖苷酶在某些情况下可以过量表达或不足表达。作为另外一种选择,该L-α-阿拉伯呋喃糖苷酶可以对宿主生物是异源的,所述L-α-阿拉伯呋喃糖苷酶由该宿主生物重组表达。此外,可以将该L-α-阿拉伯呋喃糖苷酶添加至纯化或分离

形式的本发明的纤维素酶或半纤维素酶组合物。

[0346] 细胞组合物

[0347] 在某些方面，本发明设想了细胞，其包含编码具有纤维素酶活性的多肽的核酸。在某些方面，所述细胞是里氏木霉细胞。在某些方面，所述细胞是黑曲霉细胞。在某些方面，所述细胞包括任何微生物的细胞(例如，细菌、原生生物、藻类、真菌(例如，酵母或丝状真菌)或其他微生物的细胞)，并且优选地是细菌、酵母或丝状真菌的细胞。细菌菌属的合适宿主细胞包括但不限于，埃希杆菌属(Escherichia)、芽孢杆菌属(Bacillus)、乳酸杆菌属(Lactobacillus)、假单胞菌属(Pseudomonas)和链霉菌属 (Streptomyces)的细胞。合适的细菌菌种的细胞包括但不限于大肠杆菌 (Escherichia coli)、枯草芽孢杆菌(Bacillus subtilis)、地衣芽孢杆菌(Bacillus licheniformis)、短乳杆菌(Lactobacillus brevis)、铜绿假单胞菌(Pseudomonas aeruginosa)和浅青紫链霉菌(Streptomyces lividans)的细胞。合适的酵母菌属的宿主细胞包括但不限于酵母菌属、裂殖酵母属(Schizosaccharomyces)、假丝酵母属(Candida)、汉逊酵母属(Hansenula)、毕赤酵母属、克鲁维酵母菌属(Kluyveromyces)和法夫酵母属(Phaffia)的细胞。合适的酵母菌种的细胞包括但不限于酿酒酵母、粟酒裂殖酵母(Schizosaccharomyces pombe)、白假丝酵母(Candida albicans)、多形汉逊酵母(Hansenula polymorpha)、巴斯德毕赤酵母(Pichia pastoris)、加拿大毕赤酵母(P.canadensis)、马克斯克鲁维酵母(Kluyveromyces marxianus)和红法夫酵母(Phaffia rhodozyma)的细胞。丝状真菌的合适宿主细胞包括真菌(Eumycotina)亚门的全部丝状形式。丝状真菌属的合适细胞包括但不限于支顶孢属、曲霉属、出芽短梗霉属(Aureobasidium)、黑管菌属(Bjerkandera)、拟蜡菌属(Ceriporiopsis)、金孢属(Chrysoporium)、鬼伞属(Coprinus)、革盖菌属(Coriolus)、棒囊壳属 (Corynascus)、毛壳菌属(Chaertomium)、隐球酵母属(Cryptococcus)、线黑粉酵母属(Filobasidium)、镰孢属(Fusarium)、赤霉菌属(Gibberella)、腐质霉属(Humicola)、大角间座壳菌属(Magnaporthe)、毛霉菌属(Mucor)、毁丝菌属(Myceliophthora)、毛霉菌属、新美鞭菌属(Neocallimastix)、脉孢菌属 (Neurospora)、拟青霉属(Paecilomyces)、青霉属(Penicillium)、平革菌属 (Phanerochaete)、射脉菌属(Phlebia)、瘤胃壶菌属(Piromyces)、侧耳属 (Pleurotus)、柱顶孢属(Scytaldium)、裂褶菌属(Schizophyllum)、孢子丝菌属 (Sporotrichum)、篮状菌属(Talaromyces)、嗜热子囊菌属(Thermoascus)、梭孢壳属(Thielavia)、弯颈霉属(Tolypocladium)、栓菌属(Trametes)和木霉属的细胞。丝状真菌菌种的合适细胞包括但不限于，泡盛曲霉、烟曲霉、臭曲霉、日本曲霉、构巢曲霉、黑曲霉、米曲霉、勒克瑙金孢菌、杆孢状镰孢、谷类镰孢、克地镰刀菌、黄色镰孢菌、禾谷镰孢菌、禾赤镰孢菌、异孢镰刀菌、合欢木镰孢、尖孢镰刀菌、网状镰孢、粉红镰刀菌、接骨木镰刀菌、肤色镰孢、拟分枝孢镰刀菌、硫色镰刀菌、念珠镰孢菌、拟丝孢镰刀菌、镰孢霉、烟管菌、干拟蜡菌、干拟蜡菌、Ceriporiopsis caregiea、浅黄拟蜡菌、Ceriporiopsis pannocinta、Ceriporiopsis rivulosa、Ceriporiopsis subrufa、虫拟蜡菌、灰盖鬼伞、毛革盖菌、特异腐质霉、柔毛腐质霉、米黑毛霉、嗜热毁丝菌、粗糙脉孢菌、间型脉孢菌、产紫青霉菌、变灰青霉、离生青霉、绳状青霉、黄孢原毛平革菌、射脉齿菌、杏鲍菇、黄色蠕形霉、太瑞斯梭孢壳霉、长绒毛栓菌、变色栓菌、哈茨木霉、康氏木霉、长枝木霉、里氏木霉和绿色木霉的细胞。在某些方面，所述细胞是里氏木霉细胞。在某些方面，所述细胞是黑曲霉细胞。在某些方面，所

述细胞还包含编码一种或多种半纤维素酶的一种或多种核酸。在某些方面，所述细胞包含非天然存在的纤维素酶组合物，其包含β-葡糖苷酶，所述β-葡糖苷酶是至少两种β-葡糖苷酶的嵌合体。

[0348] 在某些方面，本发明设想了包含编码多肽的核酸的细胞，所述多肽与 SEQ ID NOs:60、54、56、58、62、64、66、68、70、72、74、76、78和 79中任何一种具有至少约60％(例如，至少约65％、70％、75％、80％、 85％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％)的序列同一性。在某些方面，所述细胞还包含编码多肽的核酸，所述多肽具有至少一种半纤维素酶活性，诸如β-木糖苷酶、L-α-阿拉伯呋喃糖苷酶或木聚糖酶活性。在某些方面，本发明还设想了包含两种或更多种β-葡糖苷酶序列的嵌合体的细胞，其中第一β-葡糖苷酶序列具有至少约200个氨基酸残基长度并且包含与SEQ ID NO:60的等长度连续片段约60％(例如，约65％、约70％、约75％或约80％)或更高的序列同一性，并且其中第二β-葡糖苷酶序列具有至少约50个氨基酸残基长度并且包含与选自SEQ ID NOs:54、56、58、62、64、66、68、70、72、74、76、78和79的氨基酸序列之一的等长度连续片段约60％(例如，约65％、约65％、约70％、约 75％、约80％)或更高的序列同一性。在某些方面，本发明设想了包含两种或更多种β-葡糖苷酶序列的嵌合体或杂合体的细胞，其中第一β-葡糖苷酶序列具有至少约200个氨基酸残基长度并且包含与选自SEQ ID NOs:54、 56、58、62、64、66、68、70、72、74、76、78和79的氨基酸序列之一的等长度连续片段约60％(例如，约65％、约65％、约70％、约75％、约 80％)或更高的序列同一性或包含多肽序列基序SEQ ID NOs:164-169的一种或多种或全部，并且其中第二β-葡糖苷酶序列具有至少约50个氨基酸残基长度并且包含与SEQ ID NO:60的等长度连续片段约60％(例如，约 65％、约65％、约70％、约75％、约80％)或更高的序列同一性。在某些实施例中，第一β-葡糖苷酶序列、第二β-葡糖苷酶序列或第一种和第二β- 葡糖苷酶序列均包含一个或多个糖基化位点。在某些实施例中，第一β-葡糖苷酶序列或第二β-葡糖苷酶序列包含环区或编码环样结构的序列，其具有约3、4、5、6、7、8、9、10或11个氨基酸残基长度，包含FDRRSPG 的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。在某些实施例中，第一β-葡糖苷酶序列和第二β-葡糖苷酶序列直接相邻或连接。在一些实施例中，第一β-葡糖苷酶序列和第二β-葡糖苷酶序列并非直接相邻，而是通过接头结构域连接。在某些实施例中，接头结构域可以包含环区，其中所述环区具有约3、4、5、6、7、8、9、10或11个氨基酸残基长度，其包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列 (SEQ ID NO:172)。.在某些实施例中，接头结构域位于中央(即，并非位于所述嵌合分子的N末端或附近或位于其C末端或附近)。

[0349] 在某些方面，本发明设想了包含两种或更多种β-葡糖苷酶序列的嵌合体或杂合体的细胞，其中第一β-葡糖苷酶序列具有至少约200个氨基酸残基(例如，长度约250、300、350或400个氨基酸残基)长度并且包含 SEQ ID NOs:136-148的氨基酸序列基序的一种或多种或全部，而第二β-葡糖苷酶序列具有至少约50个氨基酸残基(例如，长度约120、150、170、200或220个氨基酸残基)长度并且包含SEQ ID NOs:149-156的氨基酸序列基序的一种或多种或全部。特别地，所述两种或更多种β-葡糖苷酶序列的第一种是具有至少约200个氨基酸残基长度并且包含SEQ ID NOs:164- 169的氨基酸序列基序中的至少2种(例如，至少2、3、4或全部)的序列，并且所述两种或更多种β-葡糖苷酶序列的第二种具有至少50个氨基酸残基长度并且包含SEQ ID NO:170。在某些实施例中，第一β-葡糖苷酶序列、第二β-葡糖苷酶

序列或第一种和第二β-葡糖苷酶序列均包含一个或多个糖基化位点。在某些实施例中，第一β-葡糖苷酶序列或第二β-葡糖苷酶序列包含环区或编码环样结构的序列，其具有约3、4、5、6、7、8、 9、10或11个氨基酸残基长度，包含FDRRSPG的序列(SEQ ID NO:171)或 FD(R/K)YNIT的序列(SEQ ID NO:172)。在某些实施例中，第一β-葡糖苷酶序列和第二β-葡糖苷酶序列直接相邻或连接。在一些实施例中，第一β- 葡糖苷酶序列和第二β-葡糖苷酶序列并非直接相邻，而是通过接头结构域连接。在某些实施例中，接头结构域可以包含环区，其中所述环区具有约 3、4、5、6、7、8、9、10或11个氨基酸残基长度，其包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。.在某些实施例中，接头结构域位于中央(即，并非位于所述嵌合分子的N末端或附近或位于其C末端或附近)。

[0350] 发酵液组合物

[0351] 在某些方面，本发明设想了包含一种或多种纤维素酶活性的发酵液，其中所述发酵液能够将生物质样品中存在的超过约50重量％的纤维素转化成可发酵糖。在某些方面，该发酵液能够将生物质样品中存在的超过约55 重量％(例如，超过约60重量％、65重量％、70重量％、75重量％、80重量％、85重量％或90重量％)的纤维素转化成可发酵糖。在某些方面，该发酵液还可以包含一种或多种半纤维素酶活性。在某些方面，本发明设想了包含至少一种β-葡糖苷酶多肽的发酵液，所述多肽与SEQ ID NOs:54、 56、58、60、62、64、66、68、70、72、74、76、78和79中的任何一种具有至少约60％(例如，至少约65％、70％、75％、80％、85％、90％、 91％92％、83％、94％、95％、96％、97％、98％、99％)序列同一性。在某些方面，本发明设想了包含杂合或嵌合β-葡糖苷酶的发酵液，所述酶是至少两种β-葡糖苷酶序列的嵌合体。

[0352] 在某些方面，本发明设想了包含至少一种β-葡糖苷酶活性的发酵液，其中所述发酵液能够将生物质样品中存在的超过约50重量％(例如，超过约55重量％、60重量％、65重量％、70重量％、75重量％或80重量％) 的纤维素转化成可发酵糖。在某些实施例中，该发酵液包含Fv3C纤维素酶活性、Pa3D纤维素酶活性、Fv3G活性、Fv3D活性、Tr3A活性、Tr3B 活性、Te3A活性、An3A活性、Fo3A活性、Gz3A活性、Nh3A活性、 Vd3A活性、Pa3G活性和/或Tn3B活性，其中所述发酵液能够将生物质样品中存在的超过约50重量％(例如，超过约55重量％、60重量％、65重量％、70重量％、75重量％或甚至80重量％)的纤维素转化成糖。

[0353] 在某些方面，本发明设想了包含两种β-葡糖苷酶序列的嵌合体或杂合体的发酵液，其中第一β-葡糖苷酶序列具有至少200个氨基酸残基长度并且包含与SEQ ID NO:60的等长度序列约60％(例如，约65％、约70％、约75％或约80％)或更高的序列同一性，并且其中第二β-葡糖苷酶序列具有至少50个氨基酸残基长度并且包含与SEQ ID NOs:54、56、58、62、 64、66、68、70、72、74、76、78和79之一的等长度序列至少约60％ (例如，约65％、约70％、约75％或约80％)或更高的序列同一性。在某些方面，本发明设想了包含两种β-葡糖苷酶序列的嵌合体或杂合体的发酵液，其中第一β-葡糖苷酶序列具有至少200个氨基酸残基长度并且包含与 SEQ ID NOs:54、56、58、62、64、66、68、70、72、74、76、78和79之一的等长度序列约60％(例如，约65％、约70％、约75％或约80％)或更高的序列同一性，并且其中第二β-葡糖苷酶序列具有至少50个氨基酸残基长度并且包含与SEQ ID NO:60的等长度序列至少约60％(例如，约 65％、约70％、约75％或约80％)或更高的序列同一性。在某些实施例中，第一β-葡糖苷酶序列、第二β-葡糖苷酶序列或第一种和第二β-葡糖苷酶序列均包含一

个或多个糖基化位点。在某些实施例中，第一β-葡糖苷酶序列或第二β-葡糖苷酶序列包含环区或编码环样结构的序列，其具有约 3、4、5、6、7、8、9、10或11个氨基酸残基长度，包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。在某些实施例中，第一β-葡糖苷酶序列和第二β-葡糖苷酶序列直接相邻或连接。在一些实施例中，第一β-葡糖苷酶序列和第二β-葡糖苷酶序列并非直接相邻，而是通过接头结构域连接。在某些实施例中，接头结构域可以包含环区，其中所述环区具有约3、4、5、6、7、8、9、10或11个氨基酸残基长度，其包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。.在某些实施例中，接头结构域位于中央(即，并非位于所述嵌合分子的N末端或附近或位于其C末端或附近)。

[0354] 本发明的方法

[0355] 在某些方面，本文提供了产生嵌合酶主链(例如，纤维素酶如内切葡聚糖酶、纤维二糖水解酶以及β-葡糖苷酶，和半纤维素酶如木聚糖酶、α- 阿拉伯呋喃糖苷酶、β-木糖苷酶)以改善稳定性的方法。在某些方面，改善的稳定性是改善的蛋白水解稳定性，因此所得的酶在适当或通常地使用所述酶的某些标准条件下对于蛋白水解切割较不敏感。在某些方面，蛋白水解的稳定性是针对储存期间的稳定性，而在其他方面，蛋白水解的稳定性是针对表达和生产期间的稳定性，这允许更有效产生酶。因此，改善的稳定性是在标准储存条件下或标准的表达或生产条件下，与未修饰的酶 (其为所述嵌合酶的来源酶(即，其序列或其变体序列构成所述嵌合酶的一部分的酶))相比，降低的蛋白水解切割水平。在某些方面，改善的稳定性同时反映在改善的储存稳定性和表达及生产期间改善的蛋白水解稳定性方面。因此，改善的稳定性是在标准的储存条件下以及表达和生产条件下降低的蛋白水解切割水平。

[0356] 在某些方面，本文提供用于将生物质转化成糖的方法，所述方法包括将该生物质与某个量的本文公开的有效地使生物质转化成可发酵糖的任何组合物接触。在某些方面，本文提供一种糖化过程，所述过程包括使用多肽处理生物质，其中所述多肽具有纤维素酶活性并且其中所述过程导致至少约50重量％(例如，至少约55重量％、至少约60重量％、至少约65重量％、至少约70重量％、至少约75重量％或至少约80重量％)的生物质转化成可发酵糖。在某些方面，本文提供了销售本文公开的任何组合物的方法，其中将所述组合物供应或出售给乙醇炼制厂或其他生物化学或生物材料制造商，并且任选地其中所述组合物在位于所述乙醇炼制厂或其他生物化学或生物材料制造商处或其附近的制造设施中制造。

[0357] 产生嵌合主链的方法

[0358] 在某些方面，本发明提供了某些β-葡糖苷酶多肽的改善的稳定性。在某些方面，改善的稳定性是改善的蛋白水解稳定性，其反映为例如在通常使用β-葡糖苷酶多肽的标准条件下β-葡糖苷酶多肽的较低程度的蛋白水解降解或切割。在某些方面，改善的蛋白水解稳定性是储存、表达或生产期间改善的稳定性。因此，改善的蛋白水解稳定性反映在通常使用或应用β- 葡糖苷酶多肽的标准储存、表达和/或生产条件下蛋白水解切割的较低水平 (例如，如反应在降低程度或水平的活性损失方面)。

[0359] 与其他异源表达的蛋白并无不同，某些β-葡糖苷酶在生产和储存期间易遭受外源性蛋白酶、细菌或真菌宿主细胞表达的蛋白酶或在生产和储存过程期间其他外力引起的蛋白水解切割。在传统上，这类蛋白水解降解可以通过以下方式降低：鉴定某蛋白质的一级氨

基酸序列中已知的蛋白水解共有序列或切割位点并且突变这些氨基酸,从而使蛋白酶不再能够在该位点处切割该蛋白质。这种方法存在缺点,其在于该多肽可能遭受超过一种蛋白酶的蛋白水解切割,或者这种切割作用可能并非酶促蛋白水解的结果。这种方法还不足以解决蛋白水解切割在多个位点处发生同时对所述多个位点具有多层次偏好水平(tiered preference levels)的状况。例如,原始蛋白(例如目的β-葡糖苷酶多肽)可能最初通过蛋白水解切割机制在特定位点上受到切割。但是一旦该初始切割位点被鉴定、修饰或突变并且不再对相同的蛋白水解切割机制敏感,则发现相同的酶通过相同或有些不同的蛋白水解切割机制在不同于这个初始切割位点的位点处受到切割。当然,还可以将第二位点鉴定、修饰或突变,以便对蛋白水解切割不再敏感,但是该酶仍可能通过如上所述的相同或不同机制在又一位点处受到蛋白水解切割。

[0360] 申请人已经发现,可以基于进化上相关的酶的二级结构间之间的比较而鉴定异源表达的多肽上的切割位点。比较在异源表达、生产和/或储存期间不遭受切割的相关酶的氨基酸序列和预测的二级结构,可以导致鉴定到蛋白质的二级结构中存在的环序列。然而,所述环序列可能是或可能不是发生切割的地方。在一些实施例中,实际的蛋白水解切割可以在该环序列的上游或下游发生。并非如采用常规方法那样突变各个氨基酸和/或突变在各个氨基酸残基或所述切割位点附近的氨基酸残基,将本发明设计为修饰环结构域,例如,替换这种环结构域,或者修饰这种环结构域的长度和/或序列,从而获得在表达、生产和/或储存期间具有优良稳定性的多肽。在某些实施例中,修饰可包括,例如,移除、加长、缩短或替换在参考不受切割的进化相关酶时所鉴定的环。此外,可以对多种异源表达的多肽实施这种方法,并随后将这些多肽融合成单一嵌合主链,与未改变以除去易切割的二级结构的嵌合多肽相比,所述嵌合主链有其整体上更优良的蛋白水解稳定性。据测定,某些氨基酸序列基序(例如,图68A中列出的基序)对于构建完全活性和高性能的β-葡糖苷酶杂合/嵌合/融合分子可能是重要的。

[0361] 申请人还比较了对剪切敏感或对剪切抵抗的某些GH3家族β-葡糖苷酶的已知3D结构,并且使用常规的3-D酶结构工具,诸如名为“Coot”的建模方法,如在,例如,Acta Cryst.(2010)D66,486-501(《晶体学报》,2010年,D66,第486-501页)中所述。例如,发现Fv3C和Te3A 对多种纤维素底物均具有比里氏木霉Bgl1更佳的β-葡糖苷酶活性和性能。还发现Fv3C在标准的储存条件或生产条件下受到蛋白水解切割,这使得包含该酶作为商业或工业酶组合物的成分较为低效或较不理想。使用诸如Coot这样的建模技术,与里氏木霉Bgl1相比,研究Te3A、Fv3C的共有特征,并且发现了四种插入物,如图70E中标出。通过这些插入物,进一步找到指示保守性相互作用(例如氢键)的残基和氨基酸序列基序,所述残基和氨基酸序列基序存在于Fv3C和Te3A中、但是不存在于里氏木霉Bgl1中的糖基化位点(如图70F-J中标出)。因此确定某些氨基酸序列基序(包括图68B中列出的基序)在决定给定的天然存在的β-葡糖苷酶或其变体或其杂合/嵌合/融合分子是否将具有改善的性能/活性以及稳定性方面是关键。

[0362] 不受理论的约束,改善的蛋白稳定性可以降低酶活性。酶活性的降低优选地低于20%,更优选地低于15%,并且甚至更优选地低于10%。因此,本文提供了通过修饰酶(例如,纤维素酶或半纤维素酶)中的环序列来改善蛋白稳定性的方法。在某些实施例中,该环序列自身对蛋白水解切割敏感。在其他实施例中,该环序列自身并不对蛋白水解切割敏感,

但是对该环序列的修饰可影响在酶中该环序列上游或下游的位点处的切割。

[0363] 在某些实施例中,该环序列存在于杂合或嵌合酶中,例如,杂合或嵌合的β-葡糖苷酶,其包含两种或更多种β-葡糖苷酶序列,每个序列衍生自不同的β-葡糖苷酶。例如,该杂合或嵌合β-葡糖苷酶可以包含两种β-葡糖苷酶序列,其中第一β-葡糖苷酶序列具有至少200个氨基酸残基长度,并且与SEQ ID NO:60的等长度序列为至少约60%(例如,至少约65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%)的序列同一性,其中第二β-葡糖苷酶具有至少50个氨基酸残基长度,并且与SEQ ID NOs:54、56、58、62、64、66、68、70、72、74、76、78或79中任何一种的等长度序列为至少约60%(例如,至少约65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%)的序列同一性。又如,该杂合或嵌合β-葡糖苷酶可以包含两种β-葡糖苷酶序列,其中第一β-葡糖苷酶序列具有至少 200个氨基酸残基长度,并且与SEQ ID NOs:54、56、58、62、64、66、68、70、72、74、76、78或79中任何一种的等长度序列为至少约60% (例如,至少约65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%)的序列同一性,其中第二β- 葡糖苷酶具有至少约50个氨基酸残基长度,并且对于SEQ ID NO:60的等长度序列为至少约60%(例如,至少约65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%)的序列同一性。在一些实施例中,长度为至少约200个氨基酸残基的第一β-葡糖苷酶序列处于该杂合酶的N末端,而长度为至少约50个氨基酸残基的第二β-葡糖苷酶序列处于该杂合酶的C末端。在某些实施例中,β-葡糖苷酶序列的N末端或C末端之一包含环序列。在某些实施例中,该环序列具有约 3、4、5、6、7、8、9、10或11个氨基酸残基长度,其包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。在某些实施例中,β-葡糖苷酶序列的N末端或C末端互相紧邻或直接互相连接。在其他实施例中,β-葡糖苷酶序列的N末端或C末端并非互相紧邻,而是通过接头结构域连接。在某些实施例中,接头结构域位于中央。在一些实施例中,接头结构域包含该环序列。在某些实施例中,对该环序列的修饰 (包括,例如,对该环序列的加长、缩短、突变、缺失(整体或部分)或替换)使所得的杂合或嵌合酶对于蛋白水解切割较不敏感。因此,所得的多肽或嵌合多肽有利地获得优于其天然对应物(例如,对于嵌合多肽而言,天然对应物是指各嵌合部分所来自的天然酶)的改善的稳定性。改善的稳定性可以由标准储存、表达、生产或使用条件下分解产物的减少或更低水平来反映。

[0364] 可以通过测试储存、表达或其他生产过程期间以及在使用这类多肽的过程中蛋白水解稳定性的改善来确定异源表达的多肽和嵌合多肽的改善的稳定性。

[0365] 在某些实施例中,该环序列存在于杂合或嵌合酶中,例如,杂合或嵌合的β-葡糖苷酶,其包含两种或更多种β-葡糖苷酶序列,每个序列衍生自不同的β-葡糖苷酶。例如,该杂合或嵌合β-葡糖苷酶可以包含两种β-葡糖苷酶序列,其中第一β-葡糖苷酶序列具有至少200个氨基酸残基长度,并且包含SEQ ID NOs:136-148的氨基酸序列的一种或多种或全部,其中第二β-葡糖苷酶具有至少约50个氨基酸残基长度,并且包含氨基酸序列基序 SEQ ID NOs:149-156的一种或多种或全部。特别地,所述两种或更多种β- 葡糖苷酶序列的第一种是具有至少约200个氨基酸残基长度并且包含SEQ ID NOs:164-169的氨基酸序列基序中至少2种(例如,至少2、3、4或全部)的序列,并且所述两种或更多种β-葡糖苷酶序列的第二种具有至少50 个氨基酸残基长度并且包含SEQ ID NO:170。在一些实施例中,长度为至少

约200个氨基酸残基的第一β-葡糖苷酶序列处于该杂合酶的N末端，而长度为至少约50个氨基酸残基的第二β-葡糖苷酶序列处于该杂合酶的C 末端。在某些实施例中，β-葡糖苷酶序列的N末端或C末端之一包含环序列。在某些实施例中，该环序列具有约3、4、5、6、7、8、9、10或11个氨基酸残基长度，其包含FDRRSPG的序列(SEQ ID NO:171)或 FD(R/K)YNIT的序列(SEQ ID NO:172)。在某些实施例中，β-葡糖苷酶序列的N末端或C末端互相紧邻或直接互相连接。在其他实施例中，β-葡糖苷酶序列的N末端或C末端并非互相紧邻，而是通过接头结构域连接。在某些实施例中，接头结构域位于中央。在一些实施例中，接头结构域包含该环序列。在某些实施例中，对该环序列的修饰(包括，例如，对该环序列的加长、缩短、突变、缺失(整体或部分)或替换)使所得的杂合或嵌合酶对于蛋白水解切割较不敏感。因此，所得的多肽或嵌合多肽有利地获得优于其天然对应物(例如，对于嵌合多肽而言，天然对应物是指各嵌合部分所来自的天然酶)的改善的稳定性。改善的稳定性可以由标准储存、表达、生产或使用条件下分解产物的减少或更低水平来反映。

[0366] 在某些方面，该环序列存在于杂合或嵌合酶中，例如杂合或嵌合β-葡糖苷酶，其包含两种或更多种酶序列，其中至少一种是β-葡糖苷酶序列，而另一种不是另一种酶的序列，也不是β-葡糖苷酶的序列。例如，嵌合酶的至少一个嵌合部分的非β-葡糖苷酶序列可以选自其他半纤维素酶或纤维素酶，例如，木聚糖酶、内切葡聚糖酶、木糖苷酶、阿拉伯呋喃糖苷酶，以及其他酶。该嵌合多肽的N末端结构域和C末端结构域可以直接互相邻接。作为另外一种选择，所述N末端结构域和C末端结构域并非直接邻接或连接，而是通过接头序列连接。在某些实施例中，β-葡糖苷酶序列的N 末端或C末端之一包含环序列。在某些实施例中，该环序列具有约3、4、 5、6、7、8、9、10或11个氨基酸残基长度，其包含FDRRSPG的序列 (SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。在某些实施例中，接头结构域位于中央。在一些实施例中，接头结构域包含该环序列。在某些实施例中，对该环序列的修饰(包括，例如，对该环序列的加长、缩短、突变、缺失(整体或部分)或替换)使所得的杂合或嵌合酶对于蛋白水解切割较不敏感。因此，所得的多肽或嵌合多肽有利地获得优于其天然对应物(例如，对于嵌合多肽而言，天然对应物是指各嵌合部分所来自的天然酶)的改善的稳定性。改善的稳定性可以由标准储存、表达、生产或使用条件下分解产物的减少或更低水平来反映。在某些实施例中，嵌合或杂合多肽可以具有双重纤维素酶和/或半纤维素酶活性。例如，本发明的嵌合或杂合多肽可以同时具有β-葡糖苷酶活性和木聚糖酶活性。在一些实施例中，该嵌合或杂合多肽可以具有优于其嵌合部分的天然对应物的改善的稳定性。例如，包含修饰的环序列的嵌合β-葡糖苷酶-木聚糖酶多肽可以具有改善的稳定性，例如，在标准的储存、表达、生产或使用条件下优于该嵌合多肽从中衍生其β-葡糖苷酶序列和其木聚糖酶序列的β-葡糖苷酶和木聚糖酶的改善的蛋白水解稳定性。

[0367] 在某些方面，本发明涉及改善纤维素酶或半纤维素酶的稳定性的方法，其中在标准的储存、表达、生产或使用条件下所述稳定性改善例如 5％或更高、10％或更高、15％或更高、20％或更高、25％或更高或者甚至 30％或更高。可以通过在特定的标准储存、表达、生产或使用条件下测定在特定的一段时间后受到切割的该酶的量来确定稳定性的改善。例如，可以在，例如标准储存条件下(例如，在常温下或在约40℃、45℃、50℃的高温下或甚至更高的温度下)在约1(例如，约1、2、3、4、5、6、8、 10、12、15、18、20、24)小时或更久后，通过切割产物的量来测量所述稳定性的改善。在某些实施例中，可以在，例如，在标准的生产条

件下 (例如,在超过50℃的温度(例如,超过50℃、超过55℃、超过60℃或甚至超过65℃))在约1(例如,约1、2、3、4、5、6、8、10、12、 15、18、20、24)小时或更久后,通过检测并确定剩余的完整产物的量来测量所述稳定性的改善。

[0368] 将生物质转化成糖的方法

[0369] 在某些方面,本文提供用于将生物质转化成糖的方法,所述方法包括将该生物质与某个量的本文公开的有效使生物质转化成可发酵糖的任何组合物接触。在某些方面,该方法还包括使用酸和/或碱预处理该生物质。在某些方面,酸包括磷酸。在某些方面,碱包括氢氧化钠或氨。

[0370] 生物质:本发明提供了用于生物质糖化的方法和过程,其使用本发明的纤维素酶或非天然存在的半纤维素酶组合物。如本文所用,术语"生物质"是指包含纤维素和/或半纤维素的任何组合物(任选地还有木质纤维素生物质中的木质素)。如本文所用,生物质包括但不限于种子、谷粒、块茎、植物废料或食品加工或工业加工的副产物(例如,茎秆)、玉米(包括,例如,玉米芯、秸秆等等)、草料(包括,例如,印度草,诸如黄假高粱(Sorghastrum nutans);或者柳枝稷,(例如,黍属(Panicum)物种,诸如柳枝稷(Panicum virgatum))、多年生木质茎(例如芦竹)、木材(包括例如,木屑、加工废料)、纸材、纸浆和回收纸(包括,例如,报纸、打印纸等等)。其他生物质包括但不限于,马铃薯、大豆(例如油菜籽)、大麦、黑麦、燕麦、小麦、甜菜和甘蔗渣。

[0371] 本发明提供糖化方法,所述方法包括使包含生物质(例如,包含木聚糖、半纤维素、纤维素和/或可发酵糖的材料)的组合物与本发明的多肽,或本发明核酸编码的多肽,或本发明的任何一种纤维素酶或非天然存在的半纤维素酶组合物或制造产物接触。

[0372] 糖化的生物质(例如由本发明的酶加工的木质纤维素材料)可以通过,诸如微生物发酵和/或化学合成的过程制成众多基于生物的产品。如本文所用,"微生物发酵"是指在合适的条件下培育和收集发酵微生物的过程。发酵微生物可以是适用于生产基于生物的产品所需的发酵过程的任何微生物。合适的发酵微生物包括但不限于丝状真菌、酵母和细菌。可以通过发酵和/或通过化学合成将经糖化的生物质制成燃料(例如,生物燃料,如生物乙醇、生物丁醇、生物甲醇、生物丙醇、生物柴油、喷气发动机燃料等)。糖化的生物质还可以例如通过发酵和/或通过化学合成制成日用化学药品(例如,抗坏血酸、异戊二烯、1,3-丙二醇)、脂质、氨基酸、蛋白质和酶。

[0373] 预处理:在糖化之前,生物质(例如,木质纤维素材料)优选地进行一个或多个预处理步骤以使木聚糖、半纤维素、纤维素和/或木质素材料对于酶而言更可接近或更敏感,并且因此而更易于由本发明的酶和/或纤维素酶或非天然存在的半纤维素酶组合物水解。

[0374] 在示例性的实施例中,预处理需要使生物质经历催化剂,所述催化剂包含反应器中强酸和金属盐的稀溶液。该生物质可以是,例如,原材料或干燥材料。这种预处理可以降低纤维素水解的活化能或温度,最终允许可发酵糖的更高产量。参见,例如,美国专利No.6,660,506;6,423,145。

[0375] 另一个示例性预处理方法需要使生物质在含水介质中在所选择的温度和压力经历第一水解步骤而水解生物质,以便主要实现半纤维素的解聚,而不实现纤维素显著解聚为葡萄糖。这个步骤产生浆液,在所述浆液中液态水相包含因半纤维素解产生的溶解单糖,以及包含纤维素和木质素的固相。随后该浆液在允许主要部分的纤维素解聚的条件下经历

第二水解步骤，产生包含纤维素的溶解/可溶解解聚产物的液态水相。参见，例如美国专利No.5,536,325。

[0376] 又一个示例性方法涉及使用约0.4％至约2％的强酸，通过一个或多个阶段的稀酸水解来处理生物质；随后使用碱性脱木质法处理酸水解材料的未反应固态木质纤维素成分。参见，例如，美国专利No.6,409,841。

[0377] 另一个示例性预处理方法包括在预水解反应器中预水解生物质(例如木质纤维素材料)；将酸性液体添加至固态木质纤维素材料以制得混合物；将该混合物加热至反应温度；保持反应温度一段时间，所述时间足够使木质纤维素材料分解成溶解部分和固体级分，所述溶解部分包含至少约 20％来自木质纤维素材料的木质素，所述固体级分包含纤维素；将溶解的部分与固体级分分离，并且在反应温度或接近反应温度移出溶解部分；以及回收溶解部分。固体级分中的纤维素变得更易于进行酶消化。参见，例如，美国专利No.5,705,369。

[0378] 其他预处理方法可以涉及过氧化氢H_2O_2的使用。参见Gould,1984, Biotech,and Bioengr.26:46-52(Gould,1984年，《生物技术和生物能》，第26卷，第46-52页)。

[0379] 预处理还可以包括使生物质化学计量的量的极低浓度氢氧化钠和氢氧化铵与接触。参见，Teixeira et al.,1999,Appl.Biochem.and Biotech.77- 79:19-34(Teixeira等人，1999年，《应用生物化学和生物技术》，第77- 79卷，第19-34页)。

[0380] 预处理还可以包括在约9至约14的pH、适当温度、压力和pH，使木质纤维素与化学品(例如碱，如碳酸钠或氢氧化钾)接触。参见，PCT公布WO2004/081185。

[0381] 氨可以例如用于优选的预处理方法中。这种预处理方法包括在高固形物含量的条件下使生物质与低浓度的氨接触。参见，例如，美国专利公布 No.20070031918和PCT公布WO06110901。

[0382] 糖化过程

[0383] 在某些方面，本文提供一种糖化过程，所述过程包括用多肽处理生物质，其中所述多肽具有纤维素酶活性并且其中所述过程导致至少约50重量％(例如，至少约55重量％、60重量％、65重量％、70重量％、75重量％或80重量％)的生物质转化成可发酵糖。在某些方面，该生物质包含木质素。在某些方面，该生物质包含纤维素。在某些方面，该生物质包含半纤维素。在某些方面，包含纤维素的生物质还包含木聚糖、半乳聚糖或阿拉伯聚糖中的一者或多者。在某些方面，生物质包括但不限于，种子、谷粒、块茎、植物废料或食品加工或工业加工的副产物(例如，茎秆)、玉米(包括，例如，玉米芯、秸秆等等)、草(包括，例如，印度草，诸如黄假高粱；或柳枝稷(例如，黍属物种，诸如柳枝稷)、多年生木质茎 (例如芦竹)、木材(包括，例如，木屑、加工废料)、纸材、纸浆和回收纸(包括，例如，报纸、打印纸等等)、马铃薯、大豆(例如油菜籽)、大麦、黑麦、燕麦、小麦、甜菜和甘蔗渣。在某些方面，在用多肽处理前，用酸和/或碱来处理包含生物质的材料。在某些方面，该酸是磷酸。在某些方面，该碱是氨或氢氧化钠。在某些方面，糖化过程还包括使用纤维素酶和/或半纤维素酶处理该生物质。在某些方面，使用完整纤维素酶处理该生物质。在某些方面，糖化过程导致至少约50重量％、55重量％、60重量％、65重量％、70重量％、75重量％、80重量％、85重量％或90重量％的生物质转化成糖。在某些方面，纤维素酶组合物或半纤维素酶组合物包含多肽，所述多肽为杂合或嵌合β-葡糖苷酶，它为至少两种β- 葡糖苷酶序列的嵌合体。

[0384] 在某些方面，提供了包括用包含多肽的组合物处理生物质的糖化过程，其中所述多肽与SEQ ID NOs:60、54、56、58、62、64、66、68、70、72、74、76、78和79中的任何一种具有至少约60％(例如，至少约 65％、70％、75％、80％、85％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％)序列同一性，并且其中所述过程导致至少约50 重量％(例如，至少约55重量％、60重量％、65重量％、70重量％、75重量％、80重量％、85重量％或90重量％)的生物质转化成可发酵糖。在某些方面，该糖化过程包括用多肽处理生物质，其中所述多肽与SEQ ID NOs:60、54、56、58、62、64、66、68、70、72、74、76、78和79中的任何一种具有至少约60％(例如，至少约65％、70％、75％、80％、85％、 90％、91％、92％、93％、94％、95％、96％、97％、98％、99％)序列同一性，并且导致至少约60重量％、70重量％、75重量％、80重量％、85重量％或90重量％的生物质转化成糖。在某些方面，在用多肽处理前，用用酸和/或碱处理包含生物质的材料，同时所述多肽与SEQ ID NOs:60、54、 56、58、62、64、66、68、70、72、74、76、78和79中任何一种具有至少80％、至少90％、至少95％或至少97％的序列同一性。在某些方面，该酸是磷酸。

[0385] 在某些方面，提供一种糖化过程，所述过程包括用包含β-葡糖苷酶的非天然存在的纤维素酶组合物或半纤维素酶组合物处理生物质，其中所述所述葡糖苷酶是至少两种β-葡糖苷酶序列的嵌合体或杂合体。

[0386] 在某些方面，该糖化过程包括使用包含至少两种β-葡糖苷酶序列的嵌合体的非天然存在的纤维素酶组合物或半纤维素酶组合物处理生物质，其中第一β-葡糖苷酶序列具有至少约200个氨基酸残基长度，并且包含与 Fv3C的氨基酸序列(SEQ ID NO:60)的等长度序列约60％(例如，约 65％、70％、75％、或80％)或更高的序列同一性，并且其中第二v葡糖苷酶序列具有至少约50个氨基酸残基长度，并且包含与选自SEQ ID NOs:54、56、68、62、64、66、68、70、72、74、76、78和79的氨基酸序列之一的等长度序列至少约60(例如，至少约65％、70％、75％或 80％)的序列同一性。在某些方面，该糖化过程包括用非天然存在的纤维素酶组合物或半纤维素酶组合物处理生物质，所述组合物包含至少两种β- 葡糖苷酶序列的嵌合体，其中第一β-葡糖苷酶序列具有至少约200个氨基酸残基长度，并且包含与选自SEQ ID NOs:54、56、68、62、64、66、68、70、72、74、76、78或79中任何一种氨基酸序列的氨基酸序列等长度序列约60％(例如，约65％、70％、75％或80％)或更高的序列同一性，并且其中第二β-葡糖苷酶序列具有至少约50个氨基酸残基长度，并且包含与SEQ ID NO:60的等长度序列至少约60％(例如，至少约65％、 70％、75％或80％)的序列同一性。在某些方面，该糖化过程包括用非天然存在的纤维素酶组合物或半纤维素酶组合物处理生物质，所述组合物包含至少两种β-葡糖苷酶序列的嵌合体，其中第一β-葡糖苷酶序列具有至少约200个氨基酸残基长度，并且包含氨基酸序列基序SEQ ID NOs:136-148 的一种或多种或全部，并且其中第二β-葡糖苷酶序列具有至少约50个氨基酸残基长度并且包含SEQ ID NOs:149-156的氨基酸序列基序的一种或多种或全部。特别地，所述两种或更多种β-葡糖苷酶序列的第一种是具有至少约200个氨基酸残基长度并且包含SEQ ID NOs:164-169的氨基酸序列基序中的至少2种(例如，至少2、3、4或全部)的序列，并且所述两种或更多种β-葡糖苷酶序列的第二种具有至少50个氨基酸残基长度并且包含 SEQ ID NO:170。在某些实施例中，第一b葡糖苷酶序列位于杂合或嵌合多肽的N末端，而第二β-葡糖苷酶序列位于杂合或嵌合多肽的C末端。在某些实施例中，第一种和第二β-葡糖苷酶序列互相紧邻或直接互相连接。在其他实施例中，

第一种和第二β-葡糖苷酶序列并非紧邻，而是通过接头结构域连接。在某些方面，第一或第二β-葡糖苷酶序列包含环序列，所述环序列具有例如约3、4、5、6、7、8、9、10或11个氨基酸残基长度，包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。在一些实施例中，如此修饰该环序列，从而所述杂合或嵌合酶对于环序列内的位点处或在环序列之外的残基处的蛋白水解切割较不敏感。在某些实施例中，第一或第二β-葡糖苷酶序列均不包含环序列，相反地，接头结构域包含环序列。在一些实施例中，接头结构域位于该杂合或嵌合多肽的中央。在某些方面，在用包含至少两种β-葡糖苷酶的嵌合体的非天然存在的纤维素酶组合物或半纤维素酶组合物处理前，用酸和/或碱处理包含生物质的材料。在某些方面，该酸是磷酸。在某些方面，该碱是氨或氢氧化钠。在某些方面，糖化过程还包括使用半纤维素酶处理该生物质。在某些方面，使用完整纤维素酶处理该生物质。在某些方面，包括用包含至少两种β-葡糖苷酶序列的嵌合体或杂合体的非天然存在的纤维素酶组合物或半纤维素酶组合物处理生物质的糖化过程导致至少约50重量％、60重量％、70重量％、75重量％、80重量％、85重量％或90重量％的生物质转化成糖，其中第一β-葡糖苷酶序列具有至少约200个氨基酸残基长度，并且包含与SEQ ID NO:60的等长度序列约60％(例如，约65％、约 70％、约75％或约80％)或更高的序列同一性，并且其中第二β-葡糖苷酶序列具有至少约50个氨基酸残基长度，并且包含与选自SEQ ID NOs:54、56、58、62、64、66、68、70、72、74、76、78和79中任何一种氨基酸序列的等长度序列至少约60％(例如，至少约65％、70％、75％或80％) 的序列同一性。在某些方面，包括用包含至少两种β-葡糖苷酶序列的嵌合体或杂合体的非天然存在的纤维素酶组合物或半纤维素酶组合物处理生物质的糖化过程导致至少约50重量％、60重量％、70重量％、75重量％、80 重量％、85重量％或90重量％的生物质转化成糖，其中第一β-葡糖苷酶序列具有至少约200个氨基酸残基长度，并且包含与选自SEQ ID NOs:54、56、58、62、64、66、68、70、72、74、76、78和79中任何一种氨基酸序列的等长度序列约60％(例如，约65％、约70％、约75％或约80％)或更高的序列同一性，并且其中第二β-葡糖苷酶序列长度为至少约50个氨基酸残基并包含与SEQ ID NO:60的等长度序列至少约60％(例如，至少约65％、70％、75％或80％)的序列同一性。在某些方面，包括用包含至少两种β-葡糖苷酶序列的嵌合体或杂合体的非天然存在的纤维素酶组合物或半纤维素酶组合物处理生物质的糖化过程导致至少约50重量％、60重量％、70重量％、75重量％、80重量％、85重量％或90重量％的生物质转化成糖，其中第一β-葡糖苷酶序列具有至少约200个氨基酸残基长度，并且包含SEQ ID NOs:136-148的氨基酸序列基序的一种或多种或全部，或优选地包含基序SEQ ID NOs:164-169的一种或多种或全部，并且其中第二β-葡糖苷酶序列具有至少约50个氨基酸残基长度，并且包含SEQ ID NOs:149-156的氨基酸序列基序的一种或多种或全部，或优选地包含序列基序SEQ ID NO:170。在某些方面，第一β-葡糖苷酶序列位于嵌合或杂合β-葡糖苷酶多肽的N末端并且第二β-葡糖苷酶序列位于其C末端。在某些实施例中，第一种和第二β-葡糖苷酶序列紧邻或直接连接。在其他实施例中，第一种和第二β-葡糖苷酶序列并非紧邻，而是通过接头结构域连接。在某些方面，第一或第二β-葡糖苷酶序列包含环序列，其中所述环序列包含约3、4、5、6、7、8、9、10或11个氨基酸残基，其包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)，并且其中对该环序列的修饰导致改善的稳定性，这可以反映为所述杂合或嵌合多肽的较低程度的切割或分解。在某些实施例中，改善的稳定性由环序列残基处的切割减少或消除来反映。

在一些实施例中，改善的稳定性由环区域之外的残基处的切割减少或消除来反映。在某些实施例中，第一或第二β-葡糖苷酶序列均不包含环区域，而接头结构域包含环序列，其具有约3、4、 5、6、7、8、9、10或11个氨基酸残基长度，包含FDRRSPG的序列(SEQ ID NO:171)或FD(R/K)YNIT的序列(SEQ ID NO:172)。在一些实施例中，糖化过程导致至少约50重量％、60重量％、70重量％、75重量％、80重量％、85重量％或90重量％的生物质转化成糖。

[0387] 行业方法

[0388] 本发明的纤维素酶和/或半纤维素酶组合物还可以用于工业和/或商业环境。因此，还设想了制造、销售或商业化本发明的纤维素酶和非天然存在的半纤维素酶组合物的方法。

[0389] 在具体的实施例中，本发明的纤维素酶和非天然存在的半纤维素酶组合物可以供应或出售给某些乙醇(生物乙醇)炼制厂或其他生物化学或生物材料制造商。在第一例子中，非天然存在的纤维素酶和/或半纤维素酶组合物可以在专门以工业规模制造酶的酶制造设施中制造。可以随后将非天然存在的纤维素酶和/或半纤维素酶组合物包装或出售给酶制造商的客户。这种运营策略在本文中称为“商业酶供应模式”。

[0390] 在另一项运营策略中，本发明的非天然存在的纤维素酶和/或半纤维素酶组合物可以在现有技术的酶生产系统中生产，所述酶生产系统由酶制造商在位于生物乙醇炼制厂或生物化学或生物材料制造商处或其附近(现场)的地点建设。在一些实施例中，酶供货协议由酶制造商和生物乙醇炼制厂或生物化学/生物材料制造商执行。酶制造商设计、控制并运行该现场酶生产系统，所述酶生产系统利用如本文所述的宿主细胞、表达以及生产方法来生产非天然存在的纤维素酶和/或半纤维素酶组合物。在某些实施例中，可以在生物乙醇炼制厂或生物化学/生物材料制造设施处或其附近使用本文的糖化方法以及酶和/或酶组合物，水解合适的生物质(优选地如本文所述经受适当的预处理)。所得的可发酵糖可以随后在同一设施处或附近的设施处发酵。这种运营策略在本文中称为“现场生物炼制模式”。

[0391] 现场生物炼制模式提供优于商业酶供应模式的某些优点，包括，例如，供应自足式运行，这允许最少依赖来自商业酶供应商的酶供应。这继而使该生物乙醇炼制厂或生物化学/生物材料制造商根据实时或近乎实时的需求更好地控制酶供应。在某些实施例中，设想了现场酶生产设施可在位置相互邻近的两个生物乙醇炼制厂和/或生物化学/生物材料制造商之间或在两个或更多个生物乙醇炼制厂和/或生物化学/生物材料制造商之间共享，这降低运输和储存酶的成本。此外，这实现现场酶生产设施更直接的“嵌入式”技术改进，减少了酶组合物的改进至更高产量的可发酵糖以及最终更高产量的生物乙醇或生物化学品之间的时间滞后。

[0392] 现场生物炼制模式在生物乙醇和生物化学品的工业生产和商业化中具有更普遍的应用性，这是因为它不仅可以用来制造、供应和生产本发明的纤维素酶和非天然存在的半纤维素酶组合物，还可以用来制造、供应和生产出加工淀粉(例如，玉米)以允许淀粉更高效及有效地转化成生物乙醇或生物化学品的那些酶和酶组合物。所述淀粉加工用酶可以在某些实施例中以现场生物炼制方式生产，随后迅速和容易地整合到生物乙醇炼制厂或生物化学/生物材料制造设施中以生产生物乙醇。

[0393] 因此，在某些方面，本发明还涉及在制造和销售某些生物乙醇、生物燃料、生物化学品或其他生物材料时应用本文的酶(例如，纤维素酶、半纤维素酶)、细胞、组合物和过程

的某些商业方法。在一些实施例中，本发明涉及以现场生物炼制模式应用这类酶、细胞、组合物和过程。在其他实施例中，本发明涉及以商业酶供应模式应用这类酶、细胞、组合物和过程。

[0394] 相关地，本发明提供了本发明的酶和/或酶组合物在商业环境下的用途。例如，本发明的酶和/或酶组合物可以在合适的市场环境中与使用该酶和/或组合物的典型或优选方法的说明一起出售。因此，本发明的酶和/或酶组合物可以按商业酶供应商模型使用或商业化，其中将本发明的酶和/或酶组合物销售给生物乙醇制造商、燃料炼制厂或生产燃料或生物产品的行业中的生物化学或生物材料制造商。在某些方面，本发明的酶和/或酶组合物可以使用现场生物炼制模式进行销售或商业化，其中所述酶和/或酶组合物在燃料炼制厂或生物化学/生物材料制造商的设施处或其附近的设施中生产或制备，并且将本发明的酶和/或酶组合物以实时为基础针对该燃料炼制厂或生物化学/生物材料制造商的具体需要定制。此外，本发明涉及向这些制造商提供技术支持和/或使用所述酶和/或酶组合物的说明，从而可以制造并销售所需的生物产品(例如生物燃料、生物化学品、生物材料等等)。

[0395] 通过参照以下实例能进一步理解本发明，这些实例以举例说明的方式提供而非旨在进行限制。

[0396] 实例

[0397] 实例1:测定法/方法

[0398] 以下的测定法/方法通常用于下文所述的实例中。与下文提供的方案的任何偏离均在具体实例中指出。

[0399] A.生物质底物的预处理

[0400] 将玉米芯、玉米秸秆和柳枝稷在酶促水解前根据WO06110901A(除非另外指明)中描述的方法和加工范围预处理。这些用于预处理的参考文献还包括在US-2007-0031918-A1、US-2007-0031919-A1、US-2007- 0031953-A1和/或US-2007-0037259-A1的公开内容中。

[0401] 经氨纤维爆破处理(AFEX)的玉米秸秆可购自密歇根生物技术国际研究所(Michigan Biotechnology Institute International(MBI))。该玉米秸秆的组成由MBI(Teymouri,F et al.Applied Biochemistry and Biotechnology,2004, 113:951-963(Teymouri,F等人,《应用生物化学与生物技术》,2004年,第113卷,第951-963页))使用国家可再生能源实验室(National Renewable Energy Laboratory(NREL))程序(NREL LAP-002)确定。NREL程序可得自:http://www.nrel.gov/biomass/analytical_procedures.html。

[0402] B.生物质的组成分析

[0403] 在Determination of structural carbohydrates and lignin in the biomass(生物质中的结构性碳水化合物和木质素测定)(科罗拉多州高登市的国家可再生能源实验室,2008年(National Renewable Energy Laboratory,Golden, CO 2008)http://www.nrel.gov/biomass/pdfs/42618.pdf)中所述的2步骤酸水解方法用于测量生物质底物的组成。使用这种方法，相对于底物中起始纤维素和木聚糖含量的理论产量以百分比转化率为单位在本文中报道酶促水解结果。

[0404] C.总蛋白测定法

[0405] BCA蛋白测定法是以分光光度计测量蛋白质浓度的比色测定法。根据制造商的建

议使用BCA蛋白分析试剂盒(Pierce Chemical(皮尔斯化学))。使用50mM乙酸钠pH5缓冲液在试管中制备酶稀释物。将稀释的酶溶液(各0.1mL)分别添加至含有1mL 15％三氯乙酸(TCA)的2mL Eppendorf离心管。将该管涡旋混合并置于冰浴中10分钟。将该管以 14,000rpm离心6分钟。弃去上清液,将沉淀物分别重悬于1mL 0.1N NaOH中,并且再次将所述管涡旋混合直至沉淀物溶解。从2mg/mL母液制备BSA标准溶液。通过将BCA蛋白分析试剂盒的0.5mL试剂B与 25mL试剂A混合来制备BCA工作溶液。将重悬的酶样品分别以0.1mL体积添加至3个Eppendorf离心管中。将二(2)mL皮尔斯BCA工作溶液添加至每份样品和BSA标准物的管。在37℃水浴中孵育该管30分钟。将样品冷却至室温(15分钟)并且测量每份样品在562nm的吸光度。

[0406] 计算每份标准物的蛋白吸光度的平均值。对平均蛋白标准物作图,吸光度在x轴上,浓度(mg/mL)在y轴上。将数据点拟合到线性方程:y＝mx +b。通过将吸光度代入x值来计算酶样品的原始浓度。通过乘以稀释倍数计算总蛋白浓度。

[0407] 纯化样品的总蛋白通过A280测定(Pace,CN,et al.Protein Science, 1995、4:2411-2423(Pace,CN等人,《蛋白质科学》,1995年,第4卷,第2411-2423页))。

[0408] 有时使用Kjeldahl法(rtech laboratories(rtech实验室))或使用 DUMAS法(TruSpec CN)(Sader,A.P.O.et al.,Archives of Veterinary Science, 2004,9(2):73-79(Sader,A.P.O.等人,《兽医科学资料》,2004年,第9 卷,第2期,第73-79页)),通过燃烧、捕获和测量释放的氮以总氮来测量发酵产物的总蛋白含量。对于复杂样品,例如,发酵液,使用平均16％的N含量以及氮至蛋白质的换算系数6.25来计算。在一些情况下,考虑到干扰性的非蛋白氮,测量总的可沉淀蛋白。在这些情况下,将12.5％的TCA浓度用于测量,并将含有蛋白质的TCA沉淀物重悬于0.1M NaOH 中。

[0409] 在一些情况下,根据制造商的建议,使用Coomassie Plus(考马斯加强型),其亦称为Better Bradford Assay(更优布拉德福德测定法) (Thermo Scientific,Rockford,IL(伊利诺伊州罗克福德市的赛默飞世尔科技公司))。在其他情况下,使用如Weichselbaum和Gornall改进的缩二脲法,使用牛血清白蛋白作为校准物,测量总蛋白(Weichselbaum,T.Amer.J.Clin.Path.1960,16:40(Weichselbaum,T.Amer.,《临床病理学杂志》,1960年,第16卷,第40页);Gornall,A.et al.J.Biol.Chem.1949, 177:752(Gornall,A.等人,《生物化学杂志》,1949年,第177卷,第 752页))。

[0410] D.使用ABTS的葡萄糖测定法

[0411] 用于葡萄糖测定的ABTS(2,2'-联氮-双(3-乙烯二氢噻唑-6)-磺酸)测定法基于以下原理:在O_2存在的情况下葡萄糖氧化酶催化葡萄糖的氧化,同时产生化学计量的量的过氧化氢(H_2O_2)。这种反应后是辣根过氧化物酶 (HRP)催化的ABTS氧化,其与H_2O_2的浓度线性相关。由绿色的演变指示氧化的ABTS的出现,其在OD 405nm处定量。在50mM乙酸钠缓冲液(pH5.0)中制备2.74mg/mL ABTS粉末(Sigma(西格玛))、0.1U/mL HRP (Sigma(西格玛))和1U/mL的葡萄糖氧化酶(OxyGO® HP L5000, Genencor,Danisco USA(OxyGO® HP L5000,杰能科,美国丹尼斯克公司))并保持于暗处。在50mM乙酸钠缓冲液(pH 5.0)中制备葡萄糖标准物 (0、2、4、6、8、10nmol)。将十(10)μL标准物以三组重复分别添加至 96孔的平底微量滴定平板中。还将十(10)μL连续稀释样品添加至平板中。向每个孔添加一百(100)μL的ABTS底物溶液,并将该平板置于分光光度酶标仪上。在405nm读取ABTS的氧化5分钟。

[0412] 作为另外一种选择，在孵育15-30分钟并接下来使用包含50mM乙酸钠缓冲液(pH 5.0)和2％SDS的猝灭混合物来猝灭该反应，此后测量在 405nm的吸光度。

[0413] E.通过HPLC分析糖

[0414] 通过以下方式制备来自玉米芯糖化水解的样品：使用离心移除不溶性物质，经0.22μm尼龙Spin-X离心管滤器(Corning,Corning,NY(纽约州康宁市的康宁公司))过滤，并且使用蒸馏水稀释至所需浓度的可溶性糖。在Shodex Sugar SH-G SH1011(8×300mm，具有6×50mm SH-1011P 保护柱)(www.shodex.net)上测定单体糖。所用溶剂为0.01N H_2SO_4，并且色谱以0.6mL/min的流速运行。将柱温维持在50℃并且通过折射率进行检测。作为另外一种选择，使用Biorad Aminex HPX-87H柱(配有Waters 2410折射率检测器)分析糖的量。分析时间为约20分钟，进样体积为 20μL，移动相是0.01N硫酸，其通过0.2μm滤器过滤并脱气，流速为 0.6mL/min，并且将柱温维持于60℃。葡萄糖、木糖和阿拉伯糖的外部标准物与每个样品组一起运行。

[0415] 尺寸排阻层析法用于分离和鉴定低聚糖。使用Tosoh Biosep G2000PW 柱(7.5mm×60cm)。使用蒸馏水洗脱糖。使用0.6mL/min的流速，并且在室温运行该柱。六碳糖标准物包括水苏糖、棉子糖、纤维二糖和葡萄糖；五碳糖标准物包括木六糖、木五糖、木四糖、木三糖、木二糖和木糖。木聚糖低聚物标准物是购得的(Megazyme(麦格酶))。通过折射率进行检测。使用峰面积单位或相对峰面积百分比来报告结果。

[0416] 总可溶性糖通过水解经离心和过滤澄清的样品(上文)而测定。澄清的样品使用0.8N H_2SO_4进行1:1稀释。所得的溶液在封盖的小瓶内于 121℃高压消毒1小时。结果在不对水解期间单体糖的损失进行校正的情况下报告。

[0417] F.来自玉米芯的低聚物制备和酶测定法

[0418] 来自里氏木霉Xyn3水解玉米芯的低聚物通过将8mg里氏木霉Xyn3 每g葡聚糖+木聚糖与250g干重的经稀氨水预处理的玉米芯在50mM乙酸钠缓冲液(pH 5.0)中孵育而制备。该反应在48℃进行72小时，伴以180rpm 旋转振荡。上清液以9,000×G离心，随后通过0.22μm Nalgene滤器过滤以回收可溶性糖。

[0419] G.生物质糖化测定法

[0420] 对于本文的典型实例，玉米芯糖化测定以微量滴定平板形式根据以下程序进行，除非特定的实例标明具体的改动。该生物质底物，例如，经稀氨水预处理的玉米芯，在水中稀释并且使用硫酸调节pH以产生pHμ5的 7％纤维素浆液，所述纤维素浆液不经进一步加工情况下用于本测定法。酶样品基于玉米芯底物中mg总蛋白/g的纤维素或/g木聚糖或/g合并的纤维素和木聚糖(使用常规组成分析方法，见上文)来加载。将所述酶在pH 5.0的50mM乙酸钠中稀释以获得所需的加载浓度。将四十(40)μL的酶溶液添加至70mg的经稀氨水预处理的玉米芯，其为7％纤维素每孔(等同于 4.5％纤维素每孔)。随后以铝箔板封膜覆盖测定平板，在室温混合并且在 50℃以200rpm孵育3天。在孵育期结束时，通过向每孔添加100μL的100mM甘氨酸缓冲液(pH10.0)来猝灭糖化反应，并且将该平板以3,000rpm 离心5分钟。将十(10)μL的上清液添加至96孔HPLC平板中的200μL MilliQ水，并且通过HPLC测量可溶性糖。

[0421] H.微量滴定平板糖化测定法

[0422] 将纯化的纤维素酶和完整纤维素酶株系无细胞产物以基于总蛋白(mg) /g底物中的纤维素的量引入所述糖化测定法。将纯化的半纤维素酶基于所述底物中的木聚糖含量加

载。生物质底物(包括,例如,经稀酸预处理的玉米秸秆(PCS)、经氨纤维爆破处理(AFEX)的玉米秸秆、经稀氨水预处理的玉米芯、经氢氧化钠(NaOH)预处理的玉米芯和经稀氨水预处理的柳枝稷)以所标示的%固形物水平混合并且将所述混合物的pH调节至5.0。将平板用铝箔平板封膜覆盖并且置于50℃的培养箱内。孵育在振荡时进行2 天。通过向各个孔添加100μL 100mM甘氨酸(pH 10)来终止反应。彻底混合后,将该平板离心并且将上清液以10倍稀释至包含100μL 10mM甘氨酸缓冲液,pH 10的HPLC平板中。如纤维二糖水解测定法(下文)所述,使用HPLC测量产生的可溶性糖的浓度。葡聚糖转化百分比定义为[mg葡萄糖+(mg纤维二糖×1.056+mg纤维三糖×1.056)]/[mg底物中的纤维素× 1.111];木聚糖转化百分比定义为[mg木糖+(mg木二糖×1.06)]/[mg底物中的木聚糖×1.136]。

[0423] I.纤维二糖水解测定法

[0424] 使用Ghose,T.K.Pure and Applied Chemistry,1987,59(2),257-268 (Ghose,T.K.,《纯化学与应用化学》,1987年,第59卷,第2期,第 257-268页)的方法测定纤维二糖酶活性。将纤维二糖单位(如Ghose所述而产生)定义为0.815除以在分析条件下释放0.1mg葡萄糖所需的酶的量。

[0425] J.氯-硝基-苯基-葡糖苷(CNPG)水解测定法

[0426] 将二百(200)μL的50mM乙酸钠缓冲液(pH 5)添加至微量滴定平板的各个孔。覆盖该平板,并使其在Eppendorf热混合仪中在37℃平衡15分钟。还将稀释于50mM乙酸钠缓冲液(pH 5)中的五(5)μL酶添加至各个孔。再次覆盖该平板,并使其在37℃平衡5分钟。将Millipore水中制备的二十(20)μL的2mM 2-氯-4-硝基苯基-β-D-吡喃葡萄糖苷(CNPG,加利福尼亚州埃德蒙顿市的罗斯科技有限公司(Rose Scientific Ltd.,Edmonton, CA))添加至各个孔并且将该平板迅速转移至分光光度计(SpectraMax 250,分子仪器公司(Molecular Devices))。在OD 405nm进行15分钟的动力学读值并且将数据记录为V_{max}。使用CNP的消光系数,将V_{max}从OD/ 秒的单位换算成μM CNP/秒。比活性(μM CNP/秒/mg蛋白)通过μM CNP/秒除以该测定法中所用的酶蛋白的毫克数来测定。

[0427] K.卡尔科弗卢尔测定法

[0428] 使用的全部化学品均为分析纯。Avicel PH-101购自宾夕法尼亚州费城的富美实公司生物聚合物部(FMC BioPolymer(Philadelphia,PA))。纤维二糖和卡尔科弗卢尔白购自密苏里州圣路易斯市的西格玛公司(Sigma(St. Louise,MO))。经磷酸溶胀的纤维素(PASC)使用如下文献的修订方案由 Avicel PH-101制备:Walseth,TAPPI 1971,35:228(Walseth,《纸浆与造纸工业技术协会》,1971年,第35卷,第228页)和Wood,Biochem.J.1971,121:353-362(Wood,《生物化学杂志》,1971年,第121卷,第 353-362页)。简而言之,将Avicel溶解于浓磷酸中,然后使用冷去离子水沉淀。在收集纤维素并用更多的水洗涤以中和pH之后,将其在50mM 乙酸钠(pH 5)中稀释至1%固形物。

[0429] 将全部酶稀释物在50mM乙酸钠缓冲液(pH 5.0)中制备。将GC220纤维素酶(丹尼斯科美国有限公司,杰能科(Danisco US Inc.,Genencor))稀释至2.5、5、10和15mg蛋白/GPASC,以产生线性校准曲线。将待测试的样品稀释至校准曲线的范围内,即得到0.1至0.4的分率产物的响应。将 150μL的冰冷1%PASC添加至96孔微量滴定平板中的20μL酶溶液内。将该平板覆盖并于50℃,200rpm在Innova培养箱/摇床内孵育2小时。使用 100μM甘氨酸(pH 10)中的100μL 50μg/mL卡尔科弗卢尔来猝灭反应。在荧光酶标仪上以Ex=365nm的激发波

长和Em＝435nm的发射波长读取荧光 (SpectraMax M5,分子仪器公司(Molecular Devices))。根据以下公式得到结果,以分率产物表示:

[0430] FP＝1-(F1样品-含纤维二糖的F1缓冲液)/(F1起点酶-含纤维二糖的F1缓冲液),

[0431] 其中FP为分率产物,并且F1为荧光单位。

[0432] 实例2:里氏木霉整合型表达菌株的构建

[0433] 构建了里氏木霉整合型表达菌株,所述表达菌株共表达五种基因:里氏木霉β-葡糖苷酶基因bgl1、里氏木霉内切木聚糖酶基因xyn3、轮枝镰刀菌β-木糖苷酶基因fv3A、轮枝镰刀菌β-木糖苷酶基因fv43D和轮枝镰刀菌α-阿拉伯呋喃糖苷酶基因fv51A。

[0434] 下文描述这些不同基因的表达盒的构建以及里氏木霉菌株的转化。

[0435] A.β-葡糖苷酶表达载体的构建

[0436] 将天然里氏木霉β-葡糖苷酶基因bgl1的N末端部分进行密码子优化 (DNA 2.0,加利福尼亚州门洛帕克市(Menlo Park,CA))。这种合成的部分包含该酶编码区的最初447个碱基。随后通过PCR使用引物SK943和 SK941(见下文)扩增这种片段。从提取自里氏木霉菌株RL-P37的基因组 DNA样品PCR扩增天然bgl1基因的剩余区域(Sheir-Neiss,G et al.Appl. Microbiol.Biotechnol.1984,20:46-53(Sheir-Neiss,G等人,《应用微生物学和生物技术》,1984年,第20卷,第46-53页)),其使用引物SK940 和SK942(见下文)。bgl1基因的这两个PCR片段在融合PCR反应中融合在一起,其使用引物SK943和SK942:

[0437] 正向引物SK943:(5'-CACCATGAGATATAGAACAGCTGCCGCT-3') (SEQ ID NO:92)

[0438] 反向引物SK941:(5'- CGACCGCCCTGCGGAGTCTTGCCCAGTGGTCCCGCGACAG-3')(SEQ ID NO:93)

[0439] 正向引物(SK940):(5'- CTGTCGCGGGACCACTGGGCAAGACTCCGCAGGGCGGTCG-3')(SEQ ID NO:94)

[0440] 反向引物(SK942):(5'-CCTACGCTACCGACAGAGTG-3')(SEQ ID NO:95)

[0441] 将所得的融合PCR片段克隆至Gateway® 入门载体pENTR™/D- TOPO®(Gateway® Entry vector pENTR™/D-TOPO®),并且转化至中大肠杆菌 One *Shot*® TOP10化学感受态细胞(英杰公司(Invitrogen)),这产生中间载体pENTR TOPO-Bgl1(943/942)(图55B)。测定插入DNA的核苷酸序列。将具有正确的bgl1序列的pENTR-943/942载体与pTrex3g使用LR clonase®反应加以重组(参见英杰公司(Invitrogen)描述的方案)。将LR clonase反应混合物转化至大肠杆菌One *Shot*® TOP10(E.coli One Shot® TOP10)化学感受态细胞(英杰公司(Invitrogen))中,产生表达载体 pTrex3g 943/942(图谱参见图55C)。该载体还包含构巢曲霉amdS基因,其编码乙酰胺酶作为里氏木霉转化的可选标记。使用引物SK745和 SK771(见下文)对表达盒进行PCR扩增以生成用于转化的产物。

[0442] 正向引物SK771:(5'-GTCTAGACTGGAAACGCAAC-3')(SEQ ID NO:96)

[0443] 反向引物SK745:(5'-GAGTTGTGAAGTCGGTAATCC-3')(SEQ ID NO:97)

[0444] 1)内切木聚糖酶表达盒的构建

[0445] 天然里氏木霉内切木聚糖酶基因xyn3使用引物xyn3F-2和xyn3R-2从提取自里氏木霉的基因组DNA样品进行PCR扩增。

[0446] 正向引物xyn3F-2:(5'- CACCATGAAAGCAAACGTCATCTTGTGCCTCCTGG-3')(SEQ ID

NO:98)

[0447] 反向引物xyn3R-2:(5'- CTATTGTAAGATGCCAACAATGCTGTTATATGCCG GCTTGGGG-3')(SEQ ID NO:99)

[0448] 将所得的PCR片段克隆至Gateway®入门载体pENTR™/D- TOPO®(Gateway® Entry vector pENTR™/D-TOPO®)中,并且转化至大肠杆菌 One *Shot*® TOP10化学感受态细胞中,产生如图55D中示出的载体。测定插入DNA的核苷酸序列。将具有正确的xyn3序列的pENTR/Xyn3载体与 pTrex3g使用LR clonase®反应方案(英杰公司(Invitrogen))加以重组。随后将该LR clonase®反应混合物转化至大肠杆菌One *Shot*® TOP10(E.coli One Shot® TOP10)化学感受态细胞(英杰公司(invitrogen))中,产生最终表达载体pTrex3g/Xyn3(参见图55E)。该载体还包含构巢曲霉amdS基因,其编码乙酰胺酶作为里氏木霉转化的可选标记。使用引物SK745和 SK822(见下文)对表达盒进行PCR扩增以生成用于转化的产物。

[0449] 正向引物SK745:(5'-GAGTTGTGAAGTCGGTAATCC-3')(SEQ ID NO:100)

[0450] 反向引物SK822:(5'-CACGAAGAGCGGCGATTC-3')(SEQ ID NO:101)

[0451] 2)β-木糖苷酶Fv3A表达载体的构建

[0452] 轮枝镰刀菌β-木糖苷酶fv3A基因使用引物MH124和MH125从轮枝镰刀菌基因组DNA样品进行扩增。

[0453] 正向引物MH124:(5'-CACCCATGCTGCTCAATCTTCAG-3')(SEQ ID NO:102)

[0454] 反向引物MH125:(5'-TTACGCAGACTTGGGGTCTTGAG-3')(SEQ ID NO:103)

[0455] 将所述PCR片段克隆至Gateway®入门载体pENTR™/D-TOPO®(Gateway® Entry vector pENTR™/D-TOPO®)中,并且转化至大肠杆菌One *Shot*® TOP10化学感受态细胞(英杰公司(Invitrogen))中,这产生中间载体 pENTR-Fv3A(参见图55F)。测定插入DNA的核苷酸序列。将具有正确的fv3A序列的pENTR-Fv3A载体与pTrex6g使用LR clonase®反应方案(英杰公司(Invitrogen))加以重组。将LR clonase®反应混合物转化至大肠杆菌One *Shot*® TOP10(E.coli One *Shot*® TOP10)化学感受态细胞(英杰公司 (invitrogen))中,产生最终表达载体pTrex6g/Fv3A(参见图55G)。该载体还包含具有氯嘧磺隆抗性的天然里氏木霉乙酰乳酸合酶(als)的突变体基因alsR,根据国际公布WO2008/039370A1中描述的方法,所述突变体基因与其天然启动子和终止子共同用作转化里氏木霉的可选标记,这是进行的。使用引物SK1334、SK1335和SK1299(见下文)对表达盒进行PCR 扩增以产生用于转化的产物。

[0456] 正向引物SK1334:(5'-GCTTGAGTGTATCGTGTAAG-3')(SEQ ID NO:104)

[0457] 正向引物SK1335:(5'-GCAACGGCAAAGCCCCACTTC-3')(SEQ ID NO:105)

[0458] 反向引物SK1299:(5'- GTAGCGGCCGCCTCATCTCATCTCATCCATCC-3')(SEQ ID NO:106)

[0459] 3)β-木糖苷酶Fv43D表达盒的构建

[0460] 对于轮枝镰刀菌β-木糖苷酶Fv43D表达盒的构建,使用引物SK1322 和SK1297(见下文)从轮枝镰刀菌基因组DNA样品扩增fv43D基因产物。内切葡聚糖酶基因egl1的启动子

区域使用引物SK1236和SK1321(见下文)从提取自菌株RL-P37的里氏木霉基因组DNA样品进行PCR扩增。将这些PCR扩增的DNA片段随后在融合PCR反应中使用引物 SK1236和SK1297(见下文)进行融合。将所得的融合PCR片段克隆至 pCR-平端II-TOPO载体(英杰公司(Invitrogen))中,以产生质粒TOPO平端/Pegl1-Fv43D(参见,图55H)。随后使用这种质粒转化大肠杆菌One Shot® TOP10(E.coli One Shot® TOP10)化学感受态细胞(英杰公司(invitrogen))。从多个大肠杆菌克隆中提取质粒DNA并通过限制酶切消化证实其序列。

[0461] 正向引物SK1322:(5'-CACCATGCAGCTCAAGTTTCTGTC-3')(SEQ ID NO:107)

[0462] 反向引物SK1297:(5'- GGTTACTAGTCAACTGCCCGTTCTGTAGCGAG-3')(SEQ ID NO:108)

[0463] 正向引物SK1236:(5'-CATGCGATCGCGACGTTTTGGTCAGGTCG- 3')(SEQ ID NO:109)

[0464] 反向引物SK1321:(5'- GACAGAAACTTGAGCTGCATGGTGTGGGACAACAAGAAGG-3')(SEQ ID NO:110)

[0465] 该表达盒使用引物SK1236和SK1297(见上文)从TOPO平端/Pegl1- Fv43D进行PCR扩增以产生用于转化的产物。

[0466] 4)α-阿拉伯呋喃糖苷酶表达盒的构建

[0467] 对于轮枝镰刀菌α-阿拉伯呋喃糖苷酶基因fv51A表达盒的构建,使用引物SK1159和SK1289(见下文)由轮枝镰刀菌基因组DNA样品扩增 fv51A基因产物。内切葡聚糖酶基因egl1的启动子区域使用引物SK1236和 SK1262(见下文)从提取自菌株RL-P37(见上文)的里氏木霉基因组 DNA样品进行PCR扩增。所述PCR扩增的DNA片段随后在融合PCR反应中使用引物SK1236和SK1289(见下文)进行融合。将所得的融合PCR 片段克隆至pCR-平端II-TOPO载体(英杰公司(invitrogen))中以产生质粒 TOPO平端/Pegl1-Fv51A(参见,图55I)并且使用这种质粒,转化大肠杆菌One Shot® TOP10(E.coli One Shot® TOP10)化学感受态细胞(英杰公司 (invitrogen))。

[0468] 正向引物SK1159:(5'-CACCATGGTTCGCTTCAGTTCAATCCTAG- 3')(SEQ ID NO:111)

[0469] 反向引物SK1289:(5'-GTGGCTAGAAGATATCCAACAC-3')(SEQ ID NO:112)

[0470] 正向引物SK1236:(5'-CATGCGATCGCGACGTTTTGGTCAGGTCG- 3')(SEQ ID NO:113)

[0471] 反向引物SK1262:(5'- GAACTGAAGCGAACCATGGTGTGGGACAACAAGAAGGAC-3')(SEQ ID NO:114)

[0472] 使用引物SK1298和SK1289(见上文)对表达盒进行PCR扩增以产生用于转化的产物。

[0473] 正向引物SK1298:(5'-GTAGTTATGCGCATGCTAGAC-3')(SEQ ID NO:115)

[0474] 反向引物SK1289:(5'-GTGGCTAGAAGATATCCAACAC-3')(SEQ ID NO:112)

[0475] 5)使用β-葡糖苷酶和内切木聚糖酶表达盒共转化里氏木霉

[0476] 衍生自RL-P37(Sheir-Neiss,G et al.Appl.Microbiol.Biotechnol.1984, 20:46-53(Sheir-Neiss,G等人,《应用微生物学和生物技术》,1984年,第20卷,第46-53页))并针对高纤维素酶产量进行选择的里氏木霉突变菌株用β-葡糖苷酶表达盒(cbh1启动子、里氏木霉β-葡糖苷酶1基因、cbh1终止子和amdS标记)和内切木聚糖酶表达盒(cbh1启动子、里氏木霉xyn3和cbh1终止子)以PEG-介导转化方法(参见,Penttila,M et al. Gene 1987, 61(2):155-64(Penttila,M等人,《基因》,1987年,第61卷,第2期,第155-164页))共转化。分

离出许多转化体并对它们检测β-葡萄糖苷酶和内切木聚糖酶生产。选择称为里氏木霉菌株#229的转化体以便用其他表达盒的转化。

[0477] 6)用两种β-木糖苷酶和α-阿拉伯呋喃糖苷酶表达盒共转化里氏木霉菌株#229

[0478] 使用β-木糖苷酶fv3A表达盒(cbh1启动子、fv3A基因、cbh1终止子和alsR标记)、β-木糖苷酶fv43D表达盒(egl1启动子、fv43D基因、天然fv43D终止子)以及fv51Aα-阿拉伯呋喃糖苷酶表达盒(egl1启动子、fv51A基因、fv51A天然终止子),根据例如国际公布WO2008153712A2,通过电穿孔法将里氏木霉菌株#229共转化。在包含氯嘧磺隆(80ppm)的Vogels琼脂平板上选择转化体。

[0479]

50×Vogels母液(配方)	20mL
BBL琼脂	20g
以去离子水定容至	980mL
在消毒后添加:50%葡萄糖	20mL

[0480] 50×Vogels母液,每升:

[0481] 在750mL去离子水中依次溶解:

[0482]

二水合柠檬酸三钠	125g
KH_2PO_4(无水)	250g

[0483]

NH_4NO_3(无水)	100g
$MgSO_4*7H_2O$	10g
$CaCl_2*2H_2O$	5g
Vogels微量元素溶液(配方见下文)	5mL
d-生物素	0.1g
以去离子水	定容至1L

[0484] Vogels微量元素溶液:

[0485]

柠檬酸	50g
$ZnSO_4.*7H_2O$	50g
$Fe(NH_4)2SO_4.*6H_2O$	10g
$CuSO_4.5H_2O$	2.5g
$MnSO_4.4H_2O$	0.5g
H_3BO_3	0.5g
$Na_2MoO_4.2H_2O$	0.5g

[0486] 分离出许多转化体并对它们检验β-木糖苷酶和L-α-阿拉伯呋喃糖苷酶生产。还对转化体根据实例1中所述的玉米芯糖化测定法筛选生物质转化性能。本文所述的里氏木霉整合型表达菌株的例子选自H3A、39A、A10A、11A和G9A,其以不同比率表达编码β-葡糖苷酶1、Xyn3的里氏木霉基因以及编码Fv3A、Fv51A和Fv43D的镰刀菌基因。与其他H3A菌株相比,将表达较低水平的里氏木霉Bgl1的特定H3A菌株#5("H3A-5") 用于本文如下所述的实验中。将表达降低水平的里氏木霉Bgl1的另一种 H3A菌株用于实例5中所述的实验。如Western印迹法所测定,在其他菌株中,一个里氏木霉菌株缺少过量表达的里氏木霉Xyn3;另一个菌株缺少 Fv51A,并且另外两个菌株缺少Fv3A。

[0487] 7)里氏木霉整合型菌株H3A的组成

[0488] 里氏木霉整合型菌株H3A的发酵和组成测确定以下基因产物的存在:里氏木霉Xyn3、里氏木霉Bgl1、Fv3A、Fv51A和Fv43D,其比率如本文图3中示出。

[0489] 8)通过HPLC进行的蛋白分析

[0490] 进行液相色谱(LC)和质谱(MS)以分离并定量发酵液中包含的酶。首先使用从皱褶链霉菌(S.plicatus)重组表达的的endoH糖苷酶(例如NEB P0702L)处理酶样品。EndoH以0.01-0.03mg endoH/mg样品中总蛋白的量使用。在HPLC分析之前,将所述混合物在37℃,pH 4.5-6.0孵育3小时以酶促移除N-联糖基化。随后使用HIC-苯基柱和35分钟范围内由高至低的盐梯度,将约50μg的蛋白进行疏水相互作用色谱(Agilent 1100HPLC)。使用高盐缓冲液A:4M硫酸铵,其包含20mM磷酸钾,pHμ6.75;以及低盐缓冲液B:20mM磷酸钾,pH 6.75实现所述梯度。在UV 222nm检测峰。收集级分并且使用质谱法分析。蛋白质比例以每个峰面积相对于样品的总积分面积的百分比来报告。

[0491] 9)向里氏木霉整合型菌株H3A的发酵液中添加纯化蛋白对于糖化经稀氨水预处理的玉米芯的影响

[0492] 本实验评估多种酶(多数经纯化,但还有一种未纯化的酶)赋予经预处理的生物质的糖化的有益效果。将纯化的蛋白质和一种未纯化的蛋白质从母液连续稀释并且添加至里氏木霉整合型菌株H3A的发酵液中。将经稀氨水预处理的玉米芯以20%固形物(w/w)(约5mg纤维素每孔)在pH 5加载到96孔微量滴定平板的孔内。将H3A发酵液以20mg蛋白/g纤维素添加至每个孔。将10、5、2和1μL体积的每种稀释的蛋白质(图4A)添加至各个孔,并且还添加水,从而向单个孔添加的液体总计10μL。参照孔包含添加10μL水或另外H3A的稀释物。将微量滴定平板用用铝箔密封并且于50℃在Innova恒温振荡培养箱内以200rpm速度振荡孵育3天。使用 100μL 100μM甘氨酸(pHμ10)猝灭样品。随后将该平板用塑料封膜覆盖并且在4℃以3,000rpm离心5分钟。将猝灭反应混合物的5μL等分试样用 100μL的水稀释。使用HPLC测定该反应中产生的葡萄糖的浓度。测量随添加到20mg/g H3A的蛋白质的浓度变化而变化的葡萄糖产率。结果在图 4B-4E中示出。

[0493] 实例3:FV3C的克隆、表达和纯化

[0494] A.Fv3C的克隆和表达

[0495] 通过在博德研究所(Broad Institute)数据库(http://www.broadinstitute.org/) 中搜索轮枝镰刀菌基因组中的GH3β-葡糖苷酶同源物而获得Fv3C的序列 (SEQ ID NO:60)。使用来自轮枝镰刀菌的纯化基因组DNA作为模板,通过 PCR扩增Fv3C可读框。所用的PCR热循环仪是DNA Engine Tetrad 2 Peltier热循环仪(伯乐实验室(Bio-Rad Laboratories))。所用的DNA聚合酶是PfuUltra II Fusion HS DNA聚合酶(Stratagene)。用于扩增可读框的引物如下:

[0496] 正向引物MH234(5'-CACCATGAAGCTGAATTGGGTCGC-3')(SEQ ID NO: 116)

[0497] 反向引物MH235(5'-TTACTCCAACTTGGCGCTG-3')(SEQ ID NO:117)

[0498] 正向引物在5'末端包括四个额外的核苷酸(序列-CACC)以辅助定向克隆至pENTR/D-TOPO(加利福尼亚州卡尔斯巴德市的英杰公司(Invitrogen,Carlsbad,CA))中。用于扩增该可读框的PCR条件如下:步骤 1:94℃进行2分钟。步骤2:94℃进行30秒。步骤3:57℃进行30秒。步骤4:72℃进行60秒。对步骤2、3和4另外重复29个循环。步骤5: 72℃进行2分钟。使用Qiaquick PCR纯化试剂盒(凯杰公司(Qiagen))纯化Fv3C可读框的PCR产物。将纯化的PCR产物首先克隆至pENTR/D- TOPO载体中,转化至TOP10化学感受态大肠杆菌细胞(英杰公司(invitrogen))中并且涂布在包含50ppm卡那霉素的LA平板上。使用 QIAspin质粒制备试剂

盒(凯杰公司(Qiagen))从大肠杆菌转化体中获得质粒DNA。使用M13正向和反向引物以及如下额外的测序引物,证实 pENTR/D-TOPO载体中插入的DNA的序列:

[0499] MH255(5'-AAGCCAAGAGCTTTGTGTCC-3')(SEQ ID NO:118)

[0500] MH256(5'-TATGCACGAGCTCTACGCCT-3')(SEQ ID NO:119)

[0501] MH257(5'-ATGGTACCCTGGCTATGGCT-3')(SEQ ID NO:120)

[0502] MH258(5'-CGGTCACGGTCTATCTTGGT-3')(SEQ ID NO:121)

[0503] 将具有正确的Fv3C可读框的DNA序列的pENTR/D-TOPO载体(图 44)与pTrex6g(图45A)目的载体使用LRclonase®反应混合物(英杰公司(invitrogen))进行重组。

[0504] 随后将LRclonase®反应的产物转化至TOP10化学感受态大肠杆菌细胞(英杰公司(invitrogen))中,然后涂布到包含50ppm羧苄青霉素的LA 平板上。所得的pExpression构建体是pTrex6g/Fv3C(图45B),其包含 Fv3C可读框和里氏木霉突变型乙酰乳酸合酶选择标记(als)。使用Qiagen 小量提取试剂盒,分离包含Fv3C可读框的pExpression构建体的DNA并且用于对里氏木霉孢子的生物射弹转化。

[0505] 使用包含合适的Fv3C可读框的pTrex6g表达载体实施对里氏木霉的生物射弹转化。具体地,根据制造商说明(参见US 2006/0003408)使用Biolistic®PDS-1000/he粒子递送系统(伯乐公司(Bio-Rad)),通过氦-轰击法转化里氏木霉菌株,其中cbh1、cbh2、eg1、eg2、eg3和bgl1已缺失 (即,六重缺失菌株,参见国际公布WO 05/001036)。将转化体转移至新鲜的氯嘧磺隆选择平板。将稳定的转化体接种到滤器微量滴定平板(康宁 (Corning))中,所述平板包含200μL/孔的甘氨酸基本培养基(包含6.0g/L 甘氨酸;4.7g/L $(NH_4)_2SO_4$;5.0g/L KH_2PO_4;1.0g/L $MgSO_4 \cdot 7H_2O$;33.0g/L PIPPS,pH 5.5),灭菌后添加约2%的葡萄糖/槐糖混合物作为碳源, 10mL/L的100g/L $CaCl_2$、2.5mL/L的400X里氏木霉微量元素溶液,所述溶液包含:175g/L无水柠檬酸;200g/L $FeSO_4 \cdot 7H_2O$;16g/L $ZnSO_4 \cdot 7H_2O$; 3.2g/L $CuSO_4 \cdot 5H_2O$;1.4g/L $MnSO_4 \cdot H_2O$;0.8g/L H_3BO_3。在安置于28℃培养箱中的富氧室内将转化体在液体培养物中培养五天。在真空歧管上收集来自所述滤器微量滴定平板的上清液样品。上清液样品在4-12%NuPAGE 凝胶上电泳并用Simply Blue染液(英杰公司(invitrogen))染色。

[0506] B.Fv3C的纯化

[0507] 将来自摇瓶浓缩物的Fv3C对25mM TES缓冲液(pH 6.8)进行透析。将透析的酶溶液以1mL/min的流速加载至SEC HiLoad Superdex 200制备级交联琼脂糖和葡聚糖柱(通用电气医疗集团(GE Healthcare)),所述柱已经用25mM TES、0.1M氯化钠(pH 6.8)预平衡。使用SDS-PAGE来鉴定和确认Fv3C在来自SEC分离的级分中的存在。汇集并浓缩包含Fv3C的级分。还使用SEC纯化将Fv3C与低分子量和高分子量污染物分离。使用考马斯亮蓝染色的SDS/PAGE测定酶制品的纯度。SDS/PAGE在97kDa处显示了一条单一主带。

[0508] C.Fv3C的可变翻译

[0509] 为表达Fv3C基因,使用了包含如镰刀菌数据库中所注释的ORF的基因组序列。http://www.broadinstitute.org/annotation/ genome/fusarium_group/MultiHome.html。预测的编码区包含3个内含子,而第一内含子打断信号肽序列(图46A)。

[0510] 但是,在其3'部分,第一内含子包含与成熟序列同框的可变ORF,其也被预测为编

码信号肽(图46B)。在全部两种翻译中,如通过N末端序列分析所确定,成熟蛋白的起始位点(图46B中以下划线标出)在两种推定信号肽切割位点的下游起始(以箭头示出)。据显示,可以通过使用两个ATG之一作为推定性翻译起点有效地表达Fv3C(图46C)。

[0511] 实例4:纤维二糖和CNPG的β-葡糖苷酶活性

[0512] 在本实验中,测试了里氏木霉Bgl1、黑曲霉Bglu(An3A)(爱尔兰威克洛的麦格酶国际爱尔兰有限公司(Megazyme International Ireland Ltd., Wicklow,Ireland))、Fv3C(SEQ ID NO:60)、Fv3D(SEQ ID NO:58)和Pa3C (SEQ ID NO:80)对于纤维二糖和CNPG的β-葡糖苷酶活性。里氏木霉 Bgl1、黑曲霉Bglu(“An3A”)、Fv3C、Fv3C/Te3A/Bgl3(FAB)嵌合体、Fv3C/Bgl3(FB)嵌合体、里氏木霉Bgl3和Te3A是纯化的蛋白质。Fv3D和 Pa3C不是纯化的蛋白质。它们在里氏木霉六重缺失菌株(如上文定义)中表达,但是仍然存在某些背景蛋白活性。如图5A中示出,发现Fv3C对纤维二糖具有约两倍于里氏木霉Bgl1的活性,而发现黑曲霉Bglu活性超过里氏木霉Bgl1约12倍。

[0513] Fv3C对CNPG底物的活性约等于里氏木霉Bgl1的活性,但是黑曲霉 Bglu的活性为里氏木霉Bgl1活性的约14%(图5A)。Fv3D,以与Fv3C 类似的方式表达的另一种轮枝镰刀菌β-葡糖苷酶,不具有可测量的纤维二糖酶活性,但其对于CNPG的活性约5倍于里氏木霉Bgl1活性。此外,以类似方式产生的鹅掌柄孢霉β-葡糖苷酶同源物Pa3C对于纤维二糖或 CNPG底物不具有可测量的活性。这些研究显示,Fv3C对于纤维二糖和 CNPG的活性归因于分子自身而并非是归因于背景蛋白活性。

[0514] 实例5:Fv3C对于多种生物质底物的糖化

[0515] A.Fv3C对PASC的糖化性能

[0516] 在本实验中,测试了里氏木霉Bgl1、Fv3C和多种Fv3C同源物增强 PASC糖化的能力。在96孔HPLC平板中,将二十(20)μL的每种β-葡糖苷酶以5mg蛋白/g纤维素的量添加至含10mg蛋白/g纤维素的完整纤维素酶 (来自里氏木霉bgl1-简化菌株)中。将一百五十(150)μL PASC的0.7%固形物浆液加至每个孔,将平板用铝箔平板封膜覆盖,然后置于设定为50℃的培养箱中振荡2小时。通过向各个孔内添加100μL的100mM甘氨酸缓冲液(pH 10)来终止反应。在彻底混合后,对平板离心并且将上清液10倍稀释入另一块HPLC平板,其在各个孔内包含100μL的10mM甘氨酸(pH 10)。使用HPLC测量所产生的可溶性糖的浓度(图47)。

[0517] 观察到在相同的条件下包含Fv3C的混合物产生比包含里氏木霉Bgl1 的混合物更高比例的葡萄糖。这显示Fv3C具有比里氏木霉Bgl1更高的纤维二糖酶活性(还参见,图5B)。Fv3G、Pa3D和Pa3G对于PASC水解并没有可观测的影响,这显示对于PASC水解而言缺乏所述六重缺失背景的贡献(其中克隆并表达多种Fv3C同源物)。

[0518] B.Fv3C在经稀酸预处理的玉米秸秆(PCS)上的糖化性能

[0519] 在本实验中,使用微量滴定平板糖化测定法(见上文)中所述的方法测试里氏木霉Bgl1、Fv3C和多种Fv3C同源物增强PCS(13%固形物)糖化的能力。对于各种测试的酶,将5mg蛋白/g纤维素的β-葡糖苷酶添加至 10mg蛋白/g纤维素的完整纤维素酶(来自里氏木霉-Bgl1简化菌株)。

[0520] 具体地,将5mg蛋白/g纤维素的各个β-葡糖苷酶(Bgl1、Fv3C和同源物)添加至10mg蛋白/g纤维素的完整纤维素酶(来自里氏木霉Bgl1简化菌株)或添加至8mg蛋白/g纤维素的纯化半纤维素酶混合物(其成分在图6中示出)。在酶混合物与底物在50℃孵育2天后,测量

葡聚糖转化率%。

[0521] 结果在图48中示出。还观察到与里氏木霉Bgl1相比就葡聚糖转化率%而言,Fv3C产生明显的有益效果。此外,Fv3C还促进比里氏木霉 Bgl1更高的葡萄糖产率和总糖产率。

[0522] 所述结果显示,来自宿主细胞背景蛋白的有限的贡献(如果有的话)。

[0523] C.Fv3C在经稀氨水预处理的玉米芯上的糖化性能

[0524] 在本实验中,根据微量滴定平板糖化测定法(见上文)中所述的方法测试里氏木霉Bgl1、Fv3C和黑曲霉Bglu(An3A)增强糖化经氨预处理的玉米芯(20%固形物)的能力。具体地,将5mg蛋白/g纤维素的β-葡糖苷酶 (例如里氏木霉Bgl1、Fv3C和同源物)添加至经稀氨水预处理的玉米芯底物,并且还添加10mg蛋白/g纤维素的完整纤维素酶(来自里氏木霉Bgl1-简化菌株)。此外,还向该混合物中添加8mg蛋白/g纤维素的纯化半纤维素酶混合物(图6),其包含Xyn3、Fv3A、Fv43D和Fv51A。将酶混合物与底物在50℃孵育2天后测量葡聚糖转化率%。

[0525] 结果在图49中示出。还观察到Fv3C似乎比包括里氏木霉Bgl1(Tr3A) 在内的其他β-葡糖苷酶表现得更好。还观察到向所述酶混合物添加黑曲霉 Bglu(An3A)到超过2.5mg/g纤维素的水平阻碍糖化。

[0526] D.Fv3C在经氢氧化钠(NaOH)预处理的玉米芯上的糖化性能

[0527] 为测试多种底物预处理方法对Fv3C性能的影响,根据微量滴定平板糖化测定法(见上文)中所述的方法测量里氏木霉Bgl1(亦称为 Tr3A)、Fv3C和黑曲霉Bglu(An3A)增强糖化经NaOH预处理的玉米芯 (12%固形物)的能力。对玉米芯的氢氧化钠预处理如下进行:将1,000g 玉米芯研磨至约2mm大小,并且随后悬浮于4L 5%的氢氧化钠水溶液中,并且加热至110℃持续16小时。在实验室真空下热过滤深棕色液体。用水洗涤滤器上的固体残留物直至没有更多的颜色洗脱。该固体在实验室真空下干燥24小时。将一百(100)g的所述样品悬浮于700mL水中并搅拌。测量该溶液的pH为11.2。添加柠檬酸水溶液(10%)以降低pH至5.0并且搅拌该悬浮液30分钟。随后将该固体过滤,用水洗涤并且在真空下于室温干燥24小时。干燥后获得86.2g的富多糖生物质。这种物质的含水量为约 7.3重量%。在氢氧化钠处理之前和之后测量葡聚糖、木聚糖、木质素和总碳水化合物含量,如通过用于碳水化合物分析的NREL方法所测定。这种预处理导致生物质的脱木质化,同时将葡聚糖/木聚糖重量比维持在未处理生物质的该重量比的15%之内。

[0528] 将约5mg蛋白/g纤维素的β-葡糖苷酶(Fv3C和同源物)添加至经 NaOH预处理的底物,另外还添加8.7mg蛋白/g纤维素的完整纤维素酶 (来自整合型里氏木霉菌株H3A,专门针对其低水平的Bgl1表达而选择 ("H3A-5菌株"))。在本实验中,并未向完整纤维素酶背景添加另外的纯化半纤维素酶(例如图6的混合物)。将酶混合物与底物在50℃孵育 2天后测量葡聚糖转化率%。

[0529] 结果在图50中示出。观察到Fv3C似乎比包括里氏木霉Bgl1 (Tr3A)、An3A、和Te3A在内的其他β-葡糖苷酶表现得多少更好。还观察到添加黑曲霉Bglu(An3A)至超过4mg/g纤维素的水平导致更低的转化率。

[0530] E.Fv3C在经稀氨水预处理的柳枝稷上的糖化性能

[0531] 在本实验中,根据在微量滴定平板糖化测定法(见上文)中所述的方法测试里氏木霉Bgl1、Fv3C和黑曲霉Bglu(An3A增强糖化经稀氨水预处理的柳枝稷(17%固形物)的能力。

经稀氨水预处理的柳枝稷获自杜邦 (DuPont)。使用国家可再生能源实验室(National Renewable Energy Laboratory(NREL))的程序(NREL LAP-002)测定组成,其可获自:http://www.nrel.gov/biomass/analytical_procedures.html。

[0532] 基于干重的组成为葡聚糖(36.82%)、木聚糖(26.09%)、阿拉伯聚糖 (3.51%)、木质素-酸不溶物(24.7%)和乙酰基(2.98%)。将这种原料进行刀研磨以通过1mm筛网。将经研磨的材料在6重量%(干固形物)氨水存在的情况下在约160℃预处理90分钟。初始的固形物含量是约50%干物质。处理的生物质在使用前储存在4℃。

[0533] 在本实验中,将5mg蛋白/g纤维素的β-葡糖苷酶(例如,里氏木霉 Bgl1、Fv3C和同源物)在10mg蛋白/g纤维素的完整纤维素酶(来自整合里氏木霉菌株(H3A),针对低β-葡糖苷酶表达而选择)存在的情况下添加至经稀氨水预处理的柳枝稷。将所述酶混合物与底物在50℃孵育2天后测量葡聚糖转化率%,并且结果在图51中示出。

[0534] 显示采用柳枝稷底物时,Fv3C比里氏木霉Bgl1和黑曲霉Bglu表现得更好。

[0535] F.Fv3C在AFEX玉米秸秆上的糖化性能

[0536] 在本实验中,根据在微量滴定平板糖化测定法(见上文)中所述的方法测试里氏木霉Bgl1、Fv3C和黑曲霉Bglu增强糖化AFEX玉米秸秆 (14%固形物)的能力。经AFEX预处理的玉米秸秆购自国家可再生能源实验室(National Renewable Energy Laboratory(NREL))。使用国家可再生能源实验室(National Renewable Energy Laboratory(NREL))的程序LAP-002测定玉米秸秆的组成,其可获自:http://www.nrel.gov/biomass/analytical_procedures.html。

[0537] 基于干重的组成为葡聚糖(31.7%)、木聚糖(19.1%)、半乳聚糖(1.83%) 和阿拉伯聚糖(3.4%)。将这种原料在5加仑的压力反应器(帕尔(Parr))中在90℃、60%的含水量、1:1的生物质与氨加载量比率下AFEX处理30分钟。从该反应器中取出经处理的生物质并且将其留在通风橱中挥发残余的氨。处理的生物质在使用前储存在4℃。

[0538] 在本实验中,将约5mg蛋白/g纤维素的β-葡糖苷酶(Fv3C和同源物)在10mg蛋白/g纤维素的完整纤维素酶(源自低表达β-葡糖苷酶的整合型里氏木霉菌株)存在的情况下添加至预处理的底物(参见图3)。将所述酶混合物与底物在50℃孵育2天后测量葡聚糖转化率%,并且结果在图52中示出。

[0539] 观察到在葡聚糖转化率方面,Fv3C比里氏木霉Bgl1表现得更好。还注意到,10mg/g纤维素的Fv3C和10mg/g纤维素的H3A完整纤维素酶在上述条件下产生完全的或表观上完全的葡聚糖转化。在低于1mg/g纤维素的水平下,黑曲霉Bglu(An3A)似乎给出比Fv3C和里氏木霉Bgl1更高的葡萄糖转化率和总葡聚糖转化率,但是在超过2.5mg/g纤维素的水平下,观察到Fv3C和里氏木霉Bgl1具有比黑曲霉Bglu更高的葡萄糖转化率和葡聚糖转化率。

[0540] 实例6:针对经稀氨水预处理的玉米芯糖化的优化FV3C与完整纤维素酶的比率

[0541] 在本实验中改变Fv3C对完整纤维素酶的比率,以确定半纤维素酶组合物中Fv3C对完整纤维素酶的最佳比率。使用经稀氨水预处理的玉米芯作为底物。半纤维素酶组合物中β-葡糖苷酶(例如,里氏木霉Bgl1、Fv3C、黑曲霉Bglu)对源自里氏木霉整合型菌株(H3A)的完整纤维素酶的比率在0至50%之间变动。将所述混合物以20mg蛋白/g纤维素添加以便水解经氨预处理的玉米芯(20%固形物)。结果在图53A-53C中示出。

[0542] 里氏木霉Bgl1对完整纤维素酶的最佳比率是广泛的,以约10%为中心,其中50%

的混合物产生与相同加载量的单独完整纤维素酶相似的性能。相比之下，黑曲霉Bglu在约5％达到最佳，并且峰形更为尖锐。在峰值/最佳水平，黑曲霉Bglu给出比包含里氏木霉Bglu的最佳混合物更高的转化率。

[0543] 确定Fv3C对完整纤维素酶的最佳比率为约25％，同时该混合物在 20mg总蛋白/g纤维素下产生超过96％的葡聚糖转化率。因此，完整纤维素酶中25％的酶可以替换为单一的酶Fv3C，导致改善的糖化性能。

[0544] 实例7：不同的酶掺合物对经氨预处理的玉米芯的糖化。

[0545] 将25％Fv3C/75％来自里氏木霉整合型菌株(H3A)的完整纤维素酶的混合物在剂量反应实验中与其他高性能纤维素酶混合物比较。将单独的来自里氏木霉整合型菌株(H3A)的完整纤维素酶、25％Fv3C/75％来自里氏木霉整合型菌株(H3A)的完整纤维素酶的混合物和Accellerase® 1500 + Multifect®木聚糖酶(Multifect® Xylanase)就它们对经稀氨水预处理的玉米芯(20％固形物)的糖化性能进行比较。在该反应中从2.5至40mg蛋白/g纤维素投予酶掺合物。结果在图54中示出。

[0546] 25％Fv3C/75％来自里氏木霉整合型菌株(H3A)的完整纤维素酶的混合物表现得比Accellerase® 1500 + Multifect®木聚糖酶(Multifect® Xylanase)掺合物好得多，并且表现出优于来自里氏木霉整合型菌株(H3A)的完整纤维素酶的大幅改善。从每种酶混合物中为达到70、80或90％葡聚糖转化率所需的剂量列于图7中。在70％葡聚糖转化率时，与Accellerase® 1500 + Multifect®木聚糖酶(Multifect® Xylanase)掺合物比较时，25％Fv3C/75％来自里氏木霉整合型菌株(H3A)的完整纤维素酶的混合物产生3.2倍的剂量减低。在70、80或90％葡聚糖转化率时，25％Fv3C/75％来自里氏木霉整合型菌株(H3A)的完整纤维素酶的混合物需要比单用来自里氏木霉整合型菌株(H3A)的完整纤维素酶少约1.8倍的酶。

[0547] 实例8：FV3C在黑曲霉菌株中的表达

[0548] 为了在黑曲霉中表达Fv3C，使用Gateway LR重组反应(英杰公司 (invitrogen))，将pENTR-Fv3C质粒与目的载体pRAXdest2进行重组(如美国专利No.7459299中所述)。该表达质粒包含处于黑曲霉葡糖淀粉酶启动子和终止子控制下的Fv3C基因组序列、作为选择标记的构巢曲霉 pyrG基因以及用于真菌细胞内自主复制的构巢曲霉ama1序列。将生成的重组产物转化至大肠杆菌Max Efficiency DH5α(英杰公司(invitrogen)) 中，并且在2xYT琼脂平板上选择包含表达构建体pRAX2-Fv3C(图 55A)的克隆，所述平板使用16g/L Bacto蛋白胨(迪飞科(Difco))、10g/L Bacto酵母提取物(迪飞科(Difco))、5g/L NaCl、16g/L Bacto琼脂(迪飞科(Difco))和100μg/mL氨苄青霉素制备。

[0549] 将约50-100mg的表达质粒转化至黑曲霉泡盛变种(var awamori)菌株中 (参见，美国专利No.7459299)。内源性葡糖淀粉酶glaA基因从这种菌株中缺失，并且其在pyrG基因中携带突变，所述突变允许针对尿苷原养型选择转化体。将黑曲霉转化体在MM培养基(与用于里氏木霉转化相同的基本培养基，但是使用10mM NH_4Cl代替乙酰胺作为氮源)上于37℃培育4-5天，并且使用来自不同转化平板的全体孢子(约10^6个孢子/mL)接种摇瓶，所述摇瓶包含生产培养基(每1升)：12g胰蛋白胨；8g大豆蛋白胨；15g $(NH_4)_2SO_4$；12.1g $NaH_2PO_4xH_2O$；2.19g $Na_2HPO_4x2H_2O$；1g $MgSO_4x7H_2O$；1mL Tween 80；150g麦芽糖；pH 5.8。在30℃并以

200rpm 振荡发酵3天后,通过SDS-PAGE来证实转化体中的Fv3C表达。

[0550] 实例9:里氏木霉BGL3(Tr3B)的性能

[0551] A.使用完整纤维素酶/里氏木霉Bgl3掺合物对PASC和PCS的糖化

[0552] 在这些实验的背景中使用来自里氏木霉突变体菌株的经澄清的完整纤维素酶发酵液,所述突变体菌株衍生自RL-P37(Sheir-Neiss,G.et al.Appl.Microbiol.Biotechnol.1984,20:46-53(Sheir-Neiss,G.等人,《应用微生物学和生物技术》,1984年,第20卷,第46-53页))并且针对高纤维素酶产量进行选择。基于mg总蛋白/g底物中的纤维素,将完整纤维素酶和纯化的里氏木霉Bgl3(Tr3B)加载入糖化测定法。纯化的里氏木霉Bgl3以0- 100%Bgl3的水平与完整纤维素酶掺合。该混合物以20mg蛋白/g纤维素加载。对每份样品进行三组重复测试。

[0553] 经磷酸溶胀的纤维素(PASC)使用如下文献的修订方案由Avicel PH- 101制备:Walseth,TAPPI 1971,35:228(Walseth,《纸浆与造纸工业技术协会》,1971年,第35卷,第228页)和Wood,Biochem.J.1971, 121:353-362(Wood,《生物化学杂志》,1971年,第121卷,第353-362 页)。简而言之,将25Avicel溶解于浓磷酸中,然后使用冷去离子水沉淀。在收集纤维素并用更多的水洗涤以中和pH之后,将其在50mM乙酸钠缓冲液(pH 5.0)中稀释至1%固形物。将二十(20)μL经稀释的酶混合物添加平底微量滴定平板的各个孔内。使用重复移液器将150μL底物添加每个孔中并以2层铝箔平板封膜覆盖该平板。

[0554] 将经稀酸预处理的玉米秸秆(见上文)在50mM乙酸钠pH 5缓冲液中稀释至7%的纤维素,并将该混合物的pH调节至5.0。使用重复移液器,将150μL的底物添加平底微量滴定平板的各个孔内。将二十(20)μL经稀释的酶混合物添加至各个孔并以2层铝箔平板封膜覆盖该平板。

[0555] 将这些平板在37℃或50℃孵育,伴以700rpm混合。将PASC孵育2 小时而将PCS平板孵育48小时。通过向各个孔添加100mL的100mM甘氨酸缓冲液(pH 10)来终止反应。彻底混合后,过滤所述平板的内容物并且将上清液以6倍稀释至包含100mL 10mM甘氨酸缓冲液(pH 10)的HPLC平板中。随后使用维持在85℃的HPLC(Agilent 1100系列,装配以除尘/保护柱(Biorad#125-0118))和Aminex HPX-87P碳水化合物柱,测量所产生的可溶性糖的浓度。移动相是具有0.6mL/分钟流速的水。葡聚糖转化率百分比在此定义为100×[mg葡萄糖+(mg纤维二糖×1.056)]/[mg底物中的纤维素×1.111]。因此,%转化率针对水解的水进行校正。完整纤维素酶:里氏木霉Bgl3混合物在50℃糖化PASC时的性能结果在图64A中示出。完整纤维素酶:里氏木霉Bgl3混合物在37℃糖化PASC时的性能结果在图 64B中示出。完整纤维素酶:里氏木霉Bgl3混合物在50℃糖化经酸再处理的玉米秸秆时的性能在图64C中示出。完整纤维素酶:里氏木霉Bgl3 混合物在37℃糖化经酸再处理的玉米秸秆时的性能在图64D中示出。

[0556] B.具有完整纤维素酶背景的Bgl3对于PASC的剂量反应

[0557] 在这些实验的背景中使用来自里氏木霉突变体菌株的经澄清的完整纤维素酶发酵液,所述突变体菌株衍生自RL-P37(Sheir-Neiss,G.Microbiol. Biotechnol.1984,20:46-53(Sheir-Neiss,G.等人,《应用微生物学和生物技术》,1984年,第20卷,第46-53页))并且针对高纤维素酶产量进行选择。

[0558] 基于mg总蛋白/g底物中的纤维素,将完整纤维素酶和纯化的里氏木霉Bgl3加载入

糖化测定。以0-10mg蛋白/g纤维素的量加载纯化的里氏木霉Bgl3。还向各样品添加恒定水平的10mg完整纤维素酶蛋白/g纤维素。对每份样品进行三组重复测试。

[0559] 将磷酸溶胀的纤维素底物在50mM乙酸钠pH 5缓冲液中稀释至1％的纤维素，并将pH调节至5.0。将二十(20)μL经稀释的酶混合物添加平底微量滴定平板的各个孔内。使用重复移液器，将150μL底物添加至各个孔并以2层铝箔平板封膜覆盖该平板。随后将这些平板在50℃以700rpm混合时孵育1小时。

[0560] 通过向各个孔添加100μL的100mM甘氨酸缓冲液(pH 10)来终止反应。彻底混合后，过滤所述平板的内容物并且将上清液以6倍稀释至包含 100μL 10mM甘氨酸缓冲液(pH 10)的HPLC平板中。随后使用维持在85℃的HPLC(Agilent 1100系列，装配以除尘/保护柱(Biorad#125-0118))和 Aminex HPX-87P碳水化合物柱，测量所产生的可溶性糖的浓度。移动相是具有0.6mL/分钟流速的水。

[0561] 葡聚糖转化率百分比在此定义为100×[mg葡萄糖+(mg纤维二糖×1.056)]/[mg底物中的纤维素×1.111]。因此，％转化率针对水解的水进行校正。里氏木霉Bgl1和里氏木霉Bgl3在糖化经磷酸溶胀的纤维素时的剂量反应比较在图65A中示出。由里氏木霉Bgl1与里氏木霉Bgl3在糖化经磷酸溶胀的纤维素时产生的纤维二糖和葡萄糖的比较在图65B中示出。

[0562] 实例10：嵌合β-葡糖苷酶

[0563] A.在里氏木霉中的表达

[0564] 将野生型Fv3C的部分的C末端序列替换为来自里氏木霉β-葡糖苷酶 Bgl3(Tr3B)的C末端序列。具体地，代表Fv3C的第1-691位残基的连续片段与代表Bgl3的第668-874位残基的连续片段融合。图60A中示出了编码Fv3C/Bgl3的嵌合/融合多肽的基因的示意图。图60B和60C中示出了氨基酸序列和编码融合/嵌合多肽Fv3C/Bgl3的多核苷酸序列。

[0565] 该嵌合/融合分子使用融合PCR构建。使用基因组Fv3C和Bgl3编码序列的pENTR的克隆作为PCR模板。两种入门克隆均构建于pDonor221 载体中(英杰公司(invitrogen))。融合产物以两个步骤装配。首先，在 PCR反应中扩增Fv3C嵌合部分，所述PCR反应使用pENTR Fv3C克隆作为模板和以下寡核苷酸引物：

[0566] pDonor正向：5’- GCTAGCATGGATGTTTTCCCAGTCACGACGTTGTAAAACGACGGC-3’ (SEQ ID NO:122)

[0567] Fv3C/Bgl3反向：5’- GGAGGTTGGAGAACTTGAACGTCGACCAAGATAGACCGTGA CCGAACTCGTAG 3’(SEQ ID NO:123)

[0568] 使用以下寡核苷酸引物，通过pENTR Bgl3载体扩增Bgl3嵌合部分：pDonor反向：5’- TGCCAGGAAACAGCTATGACCATGTAATACGACTCACTATAGG-3’ (SEQ ID NO:124)

[0569] Fv3C/Bgl3正向：5’- CTACGAGTTCGGTCACGGTCTATCTTGGTCGACGTTCAAGTTCTCCAACCTCC-3’(SEQ ID NO:125)

[0570] 在第二步骤中，将等摩尔量的所述PCR产物(分别为约1μL和0.2μL 的初始PCR反应物)作为模板添加，以便用于随后的融合PCR反应，所述融合PCR反应使用如下的一组套叠引物：

[0571] Att L1正向：5’ TAAGCTCGGGCCCCAAATAATGATTTTATTTTGACTGATAGT 3’(SEQ ID NO:126)

[0572] AttL2反向：5’GGGATATCAGCTGGATGGCAAATAATGATTTTATTTTGACTGATA 3’ (SEQ ID NO:127)

[0573] 所述PCR反应使用高保真的Phusion DNA聚合酶进行(芬兰生化酶公司(Finnzymes OY))。所得的融合PCR产物在末端包含完整的Gateway 特异性attL1、attL2重组位点，这实现了通过Gateway LR重组反应(英杰公司(invitrogen))直接克隆至最终目的载体中。

[0574] 在0.8％的琼脂糖凝胶上分离DNA片段之后，使用Nucleospin®提取 PCR纯化试剂盒(Nucleospin®Extract PCR clean-up kit)(Macherey-Nagel GmbH&Co.KG)纯化该片段，并且使用pTTT-pyrG13目的载体和LR clonase™II酶混合物(英杰公司(invitrogen))将100ng的各种片段重组。将所得的重组产物转化至大肠杆菌Max Efficiency DH5α(英杰公司 (invitrogen))中，并且在2xYT琼脂平板上选择包含表达构建体pTTT- pyrG13-Fv3C/Bgl3融合体(图61)的克隆，所述融合体包含嵌合β-葡糖苷酶，所述平板使用16g/L Bacto蛋白胨(迪飞科(Difco))、10g/L Bacto酵母提取物(迪飞科(Difco))、5g/L NaCl、16g/L Bacto琼脂(迪飞科 (Difco))和100μg/mL氨苄青霉素制备。在包含100μg/mL氨苄青霉素的2x YT培养基上培育细菌。此后将质粒分离并且使用BglI或EcoRV进行限制酶切消化。使用ABI3100序列分析仪(应用生物系统公司(Applied Biosystems))对所得的Fv3C/Bgl3区测序来证实。经具有经证实的限制模式和正确序列的质粒作为模板用于进一步的PCR反应以产生DNA片段，其使用高保真的Phusion DNA聚合酶(芬兰生化酶公司(Finnzymes OY)) 和以下引物：

[0575] Cbh1正向:5’GAGTTGTGAAGTCGGTAATCCCGCTG 3’(SEQ ID NO:128 AmdS反向:5’CCTGCACGAGGGCATCAAGCTCACTAACCG 3’(SEQ ID NO:129)

[0576] 所得片段涵盖处于cbh1启动子和终止子控制下的Fv3C/Bgl3编码区。具体地，使用具有如下所述的轻微修改形式的PEG-原生质体方法将0.5- 1mg的该片段转化至里氏木霉六重缺失菌株(参见上文)中。对于原生质体制备，将孢子于24℃在木霉基本培养基MM中以150rpm振荡培育16- 24小时，该培养基包含20g/L葡萄糖、15g/L KH_2PO_4(pH 4.5)、5g/L $(NH_4)_2SO_4$、0.6g/L $MgSO_4x7H_2O$、0.6g/L $CaCl_2x2H_2O$、1mL的1000X里氏木霉微量元素溶液(其包含5g/L $FeSO_4x7H_2O$、1.4g/L $ZnSO_4x7H_2O$、 1.6g/L $MnSO_4x\ H_2O$、3.7g/L $CoCl_2x\ 6H_2O$)。通过离心收集正在萌发的孢子并且使用50mg/mL的Glucanex G200(诺维信公司(Novozymes AG))溶液裂解真菌细胞壁。对原生质体的进一步制备根据Penttilä et al.Gene 61(1987) 155-164(Penttilä等人，《基因》，第61卷，1987年，第155- 164页)描述的方法进行。

[0577] 在200μL总体积中包含约1μg DNA和1-5×μ10^7原生质体的转化混合物各自用2mL的25％PEG溶液处理，用2体积的1.2M山梨醇/10mM Tris (pH 7.5)，10mM $CaCl_2$稀释，与包含5mM尿苷和20mM乙酰胺的3％选择性顶层琼脂糖MM混合。将所得的混合物倾倒于包含尿苷和乙酰胺的2％选择性琼脂糖平板上。在重新选取单一转化体到包含尿苷和乙酰胺的新鲜MM平板上之前，将平板在28℃进一步孵育7-10天。使用来自独立克隆的孢子在96孔微量滴定平板或摇瓶中接种发酵培养基。

[0578] 使用表达Fv3C/Bgl3嵌合体的里氏木霉转化体的孢子悬浮液接种(超过10^4个孢子/孔)包含250μL甘氨酸生产培养基的96孔滤板(康宁公司 (Corning))，所述培养基包含4.7g/L $(NH_4)_2SO_4$、33g/L 1,4-哌嗪双(丙磺酸)(pH 5.5)、6.0g/L甘氨酸、5.0g/L KH_2PO_4、

1.0g/L $CaCl_2x2H_2O$、1.0g/L $MgSO_4x7H_2O$、2.5mL/L 400X里氏木霉微量元素溶液、20g/L葡萄糖和 6.5g/L槐糖。将平板在28℃和约80％湿度孵育6-8天。通过真空过滤收集培养物上清液并将其用于测试所述杂合体的性能及其表达水平。通过 PAGE电泳测定全发酵液样品的蛋白质表达谱。将二十(20)μL的培养物上清液与8μL不含还原剂的4X上样缓冲液混合。使用MES SDS运行缓冲液 (英杰公司(invitrogen))在NuPAGE®Novex 10％Bis-Tris凝胶上分离样品。

[0579] 这产生Fv3C/Bgl3(FB)嵌合β-葡糖苷酶,其在里氏木霉中表达时或在储存期间对蛋白酶降解较不敏感。在微量滴定平板中发酵8天之后,与可比较条件下的Fv3Cβ-葡糖苷酶相比,对Fv3C/Bgl3(FB)嵌合体观察到表达的v葡糖苷酶分解显著更低。

[0580] B.Fv3C和FAB在勒克瑙金孢菌宿主细胞中的表达

[0581] 表达盒的构建

[0582] 针对里氏木霉(pTrex6g/Fv3c,实例3,图45B)和针对黑曲霉菌 (pRAX2-Fv3C,实例8,图55A)所描述的Fv3C表达载体用来在勒克瑙金孢菌中表达Fv3C或FAB。使用天然的Fv3C信号序列。pRAX2-Fv3C载体包含处于黑曲霉葡糖淀粉酶启动子和终止子序列控制下的fv3C基因序列、作为选择标记的构巢曲霉pyrG基因以及用于在真菌细胞内自主复制的构巢曲霉ama1序列。pTrex6g/Fv3c载体包含处于里氏木霉cbhI启动子和终止子序列控制下的Fv3C可读框,以及里氏木霉突变型乙酰乳酸合酶选择标记(als)及其天然启动子和终止子。作为另外一种选择,还可以使用诸如腐草霉素或潮霉素抗性的选择标记物,或者营养型选择标记物乙酰胺酶(amdS)。

[0583] 勒克瑙金孢菌的转化

[0584] 通过原生质体融合用pTrex6g/Fv3C转化勒克瑙金孢菌宿主细胞,如Penttilä et al.Gene 61(1987)155-164(Penttilä等人,《基因》,第61卷, 1987年,第155-164页)所描述,并具有本领域中已知的修改形式,如描述于,例如,美国专利6,573,086中的修改形式。随后可以在新鲜的氯嘧磺隆平板上选择抗性转化体。作为另外一种选择,可以通过原生质体融合用 pRAX2-Fv3C转化pyrG-(尿苷营养缺陷型)勒克瑙金孢菌宿主细胞并且针对尿苷原养型进行选择,如实例8中所述,见上文。

[0585] 培养勒克瑙金孢菌转化体用于蛋白质生产。

[0586] 通过在27-40℃,pH 5-10培养勒克瑙金孢菌转化体,在例如,WO 98/15633中描述的培养基中振荡约5天,使用纤维素或乳糖诱导CBHI启动子或使用麦芽糖、玛尔特灵或淀粉诱导葡糖淀粉酶启动子,生产Fv3C 和FAB。

[0587] 实例11:嵌合β-葡糖苷酶

[0588] SDS-PAGE和肽图分析显示,Fv3C/Bgl3嵌合体在里氏木霉产生时被剪切成两个片段。N末端测序示出在全长Fv3C的第674位和第683位残基之间的剪切位点。

[0589] 构建了第二种嵌合β-葡糖苷酶,其包含衍生自Fv3C的N末端序列、衍生自第二β-葡糖苷酶(来自埃默森篮状菌Te3A)的序列的环区以及来自里氏木霉Bgl3(或Tr3B)的C末端部分序列。这是通过替换Fv3C/Bgl3 嵌合体(参见,上文实例10)的环区而实现的。具体地,Fv3C/Bgl3嵌合体的Fv3C第665-683位残基(具有RRSPSTDGKSSPNN TAAPL(SEQ ID NO:157))由Te3A第634-640位残基(KYNITPI(SEQ ID NO:158))替换。使用融合PCR方法构建这种杂合分子,如上文实例10中所述。

[0590] 将两个N糖基化位点(即,S725N和S751N)引入Fv3C/Bgl3主链。将这些糖基化突变使用如上所述的融合PCR扩增技术引入Fv3C/Bgl3主链中,利用pTTT-pyrG13-Fv3C/Bgl3融合质粒(图61)作为模板以产生初始 PCR片段。将以下引物对添加至独立的PCR反应:

[0591] Pr CbhI正向:5'CGGAATGAGCTAGTAGGCAAAGTCAGC 3'(SEQ ID NO:130)和 725/751反向:5'- CTCCTTGATGCGGCGAACGTTCTTGGGGAAGCCATAGTCCTTAAGGTTCTTGCTGAAGTTGCCCAGAGAG 3'(SEQ ID NO:131) 725/751正向:5'- GGCTTCCCCAAGAACGTTCGCCGCATCAAGGAGTTTATCTACC CCTACCTGAACACCACTACCTC 3'(SEQ ID NO:132)和 Ter CbhI反向:5'GATACACGAAGAGCGGCGATTCTACGG 3'(SEQ ID NO:133)。

[0592] 接下来,使用Pr CbhI正向和Ter CbhI引物融合PCR片段。所得的融合产物包括两个所需的糖基化位点,还包含完整的attB1和attB2位点,这允许使用Gateway BP重组反应(英杰公司(invitrogen))与pDonor221载体重组。这产生pENTR-Fv3C/Bgl3/S725N S751N克隆,所述克隆随后作为主链用于构建三元杂合分子Fv3C/Te3A/Bgl3。

[0593] 为将Fv3C/Bgl3杂合体在第665-683位残基处的环替换为来自Te3A 的环序列,使用以下引物组进行初级PCR反应:

[0594] 组1:pDonor正向:5'- GCTAGCATGGATGTTTTCCCAGTCACGACGTTGTAAA ACGACGGC 3'(SEQ ID NO:122)和

[0595] Te3A反向:5'-GATAGACCGTGACCGAACTCGTAGATAGGCGTGATGTT GTACTTGTCGAAGTGACGGTAGTCGATGAAGAC 3'(SEQ ID NO:160);

[0596] 组2:Te3A2正向:5'- GTCTTCATCGACTACCGTCACTTCGACAAGTACAACATCACGCCTATCTACGAGTTCGGTCACGGTCTATC-3'(SEQ ID NO:161);以及 pDonor反向:5' TGCCAGGAAACAGCTATGACCATGTAATACGACTCACTATAGG 3' (SEQ ID NO:124)

[0597] 随后使用以下引物,将初级PCR反应中获得的片段融合:

[0598] Att L1正向:5' TAAGCTCGGGCCCCAAATAATGATTTTATTTTGACTGATAGT 3'(SEQ ID NO:126)和

[0599] AttL2反向: 5'GGGATATCAGCTGGATGGCAAATAATGATTTTATTTTGACTGATA 3' (SEQ ID NO:127)。

[0600] 所得的PCR产物在末端包含完整的Gateway特异性attL1、attL2重组位点,这允许使用Gateway LR重组反应(英杰公司(invitrogen))直接克隆至最终目的载体中。

[0601] Fv3C/Te3A/Bgl3编码基因的DNA序列在SEQ ID NO:83中列出。Fv3C/Te3A/Bgl3(FAB)杂合体的氨基酸序列在SEQ ID NO:135中列出。将编码Fv3C/Te3A/Bgl3嵌合体的基因序列克隆于pTTT-pyrG13载体中并在里氏木霉受体菌株中表达,如上文实例10中所述。

[0602] 实例12:嵌合β-葡糖苷酶的改善的稳定性

[0603] 本实验使用差示扫描量热法(DSC)测定多种β-葡糖苷酶的热变性温度。具体地,测定了纯化的酶Fv3C/Te3A/Bgl3嵌合体、Fv3C和里氏木霉 Bgl1的热转变温度。将所述酶在50mM乙酸钠缓冲液(pH 5.0)中稀释至 500ppm。以500mL各份稀释的酶样品加载DSC96孔微量滴定平板(微凯尔公司(MicroCal))。还包括水和缓冲液空白。将DSC(Auto VP-DSC,微凯尔公司(MicroCal))的参数设定为90℃/小时的扫描速度;在25℃初始温度以及110℃的最终温度。热谱曲线在图63中示出。Fv3C和 Fv3C/Te3A/Bgl3嵌合体的T_m似乎类似于并且可能某种程度上低于里氏木霉Bgl1。

[0604] 实例13:黑曲霉表达的FV3C在糖化经稀氨水预处理的玉米芯时的活性

[0605] 基于mg总蛋白/g底物中的纤维素,将整合型菌株H3A-5(低β-葡糖苷酶产生株)、黑曲霉中产生的Fv3C(参见实例8)和纯化的里氏木霉 Bgl1(本文中亦称为“里氏木霉Bglu1”或“Tr3A”)加载导致糖化测定法中。以0-10mg蛋白/g纤维素加载β-葡糖苷酶。向各个样品添加恒定水平的10mg/g H3A-5。对每份样品进行5个重复测定。

[0606] 将经稀氨水预处理的玉米芯底物在50mM乙酸钠(pH 5)缓冲液中稀释至7%纤维素,并将pH调节至5.0。将底物送入96孔微量滴定平板 (65mg/孔)中。向该96孔板每孔添加三十(30)μL的适当稀释的酶混合物。在添加酶混合物后,底物经计算包含5%纤维素。将所述平板用2层铝箔平板封膜覆盖。随后将全部平板置于50℃和200rpm的培养箱中48小时。

[0607] 通过向每个孔添加100μL的100mM甘氨酸缓冲液(pH 10)来终止反应。彻底混合后,将所述平板的内容物离心并且将上清液以11倍稀释至包含100μL 10mM甘氨酸缓冲液(pH 10)的HPLC平板中。随后通过HPLC测量所产生的可溶性糖的浓度。Agilent 1100系列HPLC配有除尘/保护柱 (Biorad#125-0118)和Aminex铅基碳水化合物柱(Aminex HPX-87P),维持在85℃。移动相是具有0.6mL/分钟流速的水。

[0608] 葡聚糖转化率百分比定义为100×[mg葡萄糖+(mg纤维二糖×1.056)]/ [mg底物中的纤维素×1.111]。以这种方式,在图62中示出了针对水解的水加以校正的%转化率。

[0609] 实例13:比较FV3C、FAB和里氏木霉BGL1的底物结合

[0610] 本实验比较了Fv3C、嵌合β-葡糖苷酶分子FAB和里氏木霉Bgl1各自对于特定的典型生物质底物的结合。

[0611] 木质素,苯基丙酸类的复杂生物聚合物,是木材中的主要非碳水化合物组分,其结合纤维素纤维以硬化和强化植物的细胞壁。因为其与其他细胞壁成分交联,所以木质素使得纤维素和半纤维素对纤维素降解酶的可及性最小化。因此,木质素通常与所有植物生物质的消化性降低相关。具体地讲,纤维素酶与木质素的结合降低纤维素酶对纤维素的降解。木质素是疏水并且表观上带负电荷。在FAB、Bgl1和Fv3C当中,Fv3C具有最低的 pI并且带有最低的正电荷,而Bglu1具有最高的pI并且带有最高的正电荷,并研究了其与木质纤维素底物的结合。

[0612] 在使用包含100mg/g纤维素的Accellerase和8mg Multifect木聚糖酶(Multifect Xylanase)/g纤维素的糖化混合物广泛糖化经稀氨水预处理的玉米芯(DACC)或玉米秸秆(DACS)或对经酸预处理的玉米秸秆(PCS或 whPCS)后,回收木质素。在糖化后接着通过添加非特异性的丝氨酸蛋白酶水解所述纤维素酶。将0.1N HCl添加所述混合物以使蛋白酶失活,随后使用乙酸盐缓冲液(50mM乙酸钠,pH 5)重复洗涤以使样品回到pH 5。

[0613] 将一百(100)μL的DACS(约5%的葡聚糖)、DACC(约5%的葡聚糖)、whPCS(约5%的葡聚糖)、由DACC制备的木质素(相当于5%的葡聚糖)、由PCS制备的木质素(相当于5%的葡聚糖)或50mM乙酸钠 (pH 5)缓冲液对照与100μL 150mg/mL的FAB、里氏木霉Bgl1或Fv3C在微量滴定平板中混合,随后将所述平板密封并在50℃孵育44小时。将该微量滴定平板以高速离心,以便将可溶性物质与不溶性物质分离。测量可溶性级分中的酶活性。简而言之,将上清液稀释5倍,随后将20μL添加至80μL 2mM的2-氯-4-硝基苯基β-D-吡喃葡萄糖苷(CNPG)中并在室温孵育6分钟。添加一百(100)μL的500mM Na2CO3(pH 9.5)以猝灭反应。读取OD405。通过使用可溶级性分中β-葡糖苷酶活性的OD405除以对照样品的OD405来计算未结

合β-葡糖苷酶的百分比，所述对照样品在缺乏木质素和生物质底物的情况下以相同的方式孵育。

[0614] 测量了结合和未结合的β-葡糖苷酶的总活性。将该微量滴定平板重新混合，将20μL等分试样各自添加至80μL乙酸钠缓冲液(pH 5)中，将20μL 的稀释混合物添加至80μL的2mM 2-氯-4-硝基苯基β-D-吡喃葡萄糖苷 (CNPG)中并且在室温孵育6分钟，然后添加100μL 500mM的Na2CO3(pH 9.5)以猝灭反应。将反应混合物离心下来并且转移出100μL的上清液至新的微量滴定平板中。测量OD405。通过使用总混合物的OD405除以对照样品的OD405来计算在生物质或木质素存在下的相对总β-葡糖苷酶活性，所述对照样品在缺乏木质素和生物质底物的情况下以相同的方式孵育。

[0615] 为证实结合的β-葡糖苷酶没有在测量的时间范围内解离，从重新混合的微量滴定平板中取20μL等分试样至新微量滴度平板中的80μL的乙酸钠缓冲液(pH 5)内，将该平板在室温振荡孵育半小时以使β-葡糖苷酶从生物质或木质素解离。随后将该平板离心并且如上所述，测量上清液中的β-葡糖苷酶活性。再次计算未结合的β-葡糖苷酶。

[0616] Fv3C显示对生物质底物或木质素的最低结合，而FAB和里氏木霉1 显示对生物质底物和木质素的高水平结合(图71A)。这三种β-葡糖苷酶均不与DACC结合，但是里氏木霉和FAB均与通过完全糖化DACC所制备的木质素结合。令人吃惊的是，与游离的FAB或Bg11相比，结合的 FAB或里氏木霉Bg11保持约50-80％活性(图71B)。还观察到结合的 FAB不从生物质或木质素上解离，但是约20％的Bg11在30分钟的孵育期间从结合状态解离成为未结合的状态(图71C)。

编号	NT或AA	名称
1.	核苷酸	来自轮枝镰刀菌的GH3酶Fv3A的核苷酸序列
2.	氨基酸	Fv3A的蛋白质序列
3.	核苷酸	来自绳状青霉的GH43酶Pf43A的核苷酸序列
4.	氨基酸	Pf43A的蛋白质序列
5.	核苷酸	来自轮枝镰刀菌的GH43酶Fv43E的核苷酸序列
6.	氨基酸	Fv43E的蛋白质序列
7.	核苷酸	来自轮枝镰刀菌的GH39酶Fv39A的核苷酸序列
8.	氨基酸	Fv39A的蛋白质序列
9.	核苷酸	来自轮枝镰刀菌的GH43酶Fv43A的核苷酸序列
10.	氨基酸	Fv43A的蛋白质序列
11.	核苷酸	来自轮枝镰刀菌的GH43酶Fv43B的核苷酸序列
12.	氨基酸	Fv43B的蛋白质序列
13.	核苷酸	来自鹅掌柄孢霉的GH51酶Pa51A的核苷酸序列
14.	氨基酸	Pa51A的蛋白质序列
15.	核苷酸	来自玉米赤霉的GH43酶Gz43A的核苷酸序列
16.	氨基酸	Gz43A的蛋白质序列
17.	核苷酸	来自尖孢镰刀菌的GH43酶Fo43A的核苷酸序列
18.	氨基酸	Fo43A的蛋白质序列
19.	核苷酸	来自烟曲霉的GH43酶Af43A的核苷酸序列
20.	氨基酸	Af43A的蛋白质序列
21.	核苷酸	来自绳状青霉的GH51酶Pf51A的核苷酸序列
22.	氨基酸	Pf51A的蛋白质序列
23.	核苷酸	来自烟曲霉的GH11酶AfuXyn2的核苷酸序列
24.	氨基酸	AfuXyn2的蛋白质序列
25.	核苷酸	来自烟曲霉的GH11酶AfuXyn5的核苷酸序列
26.	氨基酸	AfuXyn5的蛋白质序列

图1A

编号	NT或AA	名称
27.	核苷酸	来自轮枝镰刀菌的GH43酶Fv43D的核苷酸序列
28.	氨基酸	Fv43D的蛋白质序列
29.	核苷酸	来自绳状青霉的GH43酶Pf43B的核苷酸序列
30.	氨基酸	Pf43B的蛋白质序列
31.	核苷酸	来自轮枝镰刀菌的GH51酶Fv51A的核苷酸序列
32.	氨基酸	Fv51A的蛋白质序列
33.	核苷酸	来自球毛壳菌(*C. globosum*)的GH51酶Cg51B的核苷酸序列
34.	氨基酸	Cg51B的蛋白质序列
35.	核苷酸	来自轮枝镰刀菌的GH43酶Fv43C的核苷酸序列
36.	氨基酸	Fv43C的蛋白质序列
37.	核苷酸	来自轮枝镰刀菌的GH30酶Fv30A的核苷酸序列
38.	氨基酸	Fv30A的蛋白质序列
39.	核苷酸	来自轮枝镰刀菌的GH43酶Fv43F的核苷酸序列
40.	氨基酸	Fv43F的蛋白质序列
41.	核苷酸	来自里氏木霉的GH10木聚糖酶Xyn3的核苷酸序列
42.	氨基酸	Xyn3的蛋白质序列
43.	氨基酸	来自里氏木霉的GH11木聚糖酶Xyn2的蛋白质序列
44.	氨基酸	来自里氏木霉的GH3 β-木糖苷酶Bx11的蛋白质序列
45.	氨基酸	来自里氏木霉的GH3 β-葡糖苷酶Bg11的蛋白质序列
46.	核苷酸	Pa51A的推定的cDNA。
47.	核苷酸	Pa51A的推定的cDNA。
48.	核苷酸	编码成熟Gz43A的基因组DNA上游的CBH1信号序列的编码序列。
49.	核苷酸	编码成熟Fo43A的基因组DNA上游的CBH1信号序列的编码序列。
50.	核苷酸	编码PF51A的密码子优化DNA上游的CBH1信号序列的核苷酸序列
51.	核苷酸	来自里氏木霉的内切葡聚糖酶Eg4的核苷酸序列

图1B

编号	NT或AA	名称
52.	氨基酸	Eg4的蛋白质序列
53.	核苷酸	来自鹅掌柄孢霉的GH3 β-葡糖苷酶Pa3D的核苷酸序列
54.	氨基酸	Pa3D的蛋白质序列
55.	核苷酸轮枝镰刀菌	来自轮枝镰刀菌的GH3 β-葡糖苷酶Fv3G的核苷酸序列
56.	氨基酸	Fv3G的蛋白质序列
57.	核苷酸轮枝镰刀菌	来自轮枝镰刀菌的GH3 β-葡糖苷酶Fv3D的核苷酸序列
58.	氨基酸	Fv3D的蛋白质序列
59.	核苷酸轮枝镰刀菌	来自轮枝镰刀菌的GH3 β-葡糖苷酶Fv3C的核苷酸序列
60.	氨基酸	Fv3C的蛋白质序列
61.	核苷酸	来自里氏木霉的GH3 β-葡糖苷酶Tr3A的核苷酸序列
62.	氨基酸	Tr3A的蛋白质序列
63.	核苷酸	来自里氏木霉的GH3 β-葡糖苷酶Tr3B的核苷酸序列
64.	氨基酸	Tr3B的蛋白质序列
65.	核苷酸	来自埃默森篮状菌的针对在里氏木霉中的表达进行了密码子优化的GH3 β-葡糖苷酶Te3A的核苷酸序列
66.	氨基酸	Te3A的蛋白质序列
67.	核苷酸黑曲霉	来自黑曲霉的GH3 β-葡糖苷酶An3A的核苷酸序列
68.	氨基酸	An3A的蛋白质序列
69.	核苷酸尖孢镰刀菌	来自尖孢镰刀菌的GH3 β-葡糖苷酶Fo3A的核苷酸序列
70.	氨基酸	Fo3A的蛋白质序列
71.	核苷酸玉米赤霉	来自玉米赤霉的GH3 β-葡糖苷酶Gz3A的核苷酸序列
72.	氨基酸	Gz3A的蛋白质序列
73.	核苷酸鞭毛藻丛赤壳菌	来自鞭毛藻丛赤壳菌的GH3 β-葡糖苷酶Nh3A的核苷酸序列
74.	氨基酸	Nh3A的蛋白质序列
75.	核苷酸轮枝菌	来自大丽轮枝菌的GH3 β-葡糖苷酶Vd3A的核苷酸序列
76.	氨基酸	Vd3A的蛋白质序列

图1C

编号	NT或AA	名称
77.	核苷酸	来自鹅掌柄孢霉的GH3 β-葡糖苷酶Pa3G的核苷酸序列
78.	氨基酸	Pa3G的蛋白质序列
79.	氨基酸	来自新阿波罗柄热袍菌的GH3 β-葡糖苷酶Tn3B的蛋白质序列
80.	氨基酸	来自鹅掌柄孢霉的GH3酶Pa3C的蛋白质序列
81.	核苷酸	来自鹅掌柄孢霉的Pa3C的核苷酸序列
82.	核苷酸	Fv3C/里氏木霉Bg13融合体/嵌合体的核苷酸序列
83.	核苷酸	Fv3C/Te3A/里氏木霉Bg13融合嵌合体的核苷酸序列
84.	氨基酸	GH61家族内切葡聚糖酶的蛋白质序列基序1
85.	氨基酸	GH61家族内切葡聚糖酶的蛋白质序列基序2
86.	氨基酸	GH61家族内切葡聚糖酶的蛋白质序列基序3
87.	氨基酸	GH61家族内切葡聚糖酶的蛋白质序列基序4
88.	氨基酸	GH61家族内切葡聚糖酶的蛋白质序列基序5
89.	氨基酸	GH61家族内切葡聚糖酶的蛋白质序列基序6
90.	氨基酸	GH61家族内切葡聚糖酶的蛋白质序列基序7
91.	氨基酸	GH61家族内切葡聚糖酶的蛋白质序列基序8
135.	氨基酸	Fv3C/Te3A/Bg13 (FAB)嵌合体的蛋白质序列
159.	氨基酸	Fv3C/Bg13 (FB)嵌合体的蛋白质序列
162.	核苷酸	里氏木霉Xyn2的核苷酸序列
163.	核苷酸	里氏木霉Bx11的核苷酸序列

图1D

酶	Tn3B	Fv3G	Fv3D	Tr3A	Pa3D	Te3A	An3A	Tr3B	Nh3A	Gz3A	Fv3C	Fo3A	Pa3G	Vd3A
底物相互作用	D58	D101	D111	D92	D87	D92	D92	D99	D106	D106	D119	D119	D101	D107
	R64	R107	R117	R98	R93	R98	R98	R105	R112	R112	R125	R125	R107	R113
	L116	L150	L160	L141	L136	L141	L141	L148	L155	L155	L168	L168	L150	L156
	R130	R165	R175	R156	R151	R156	R156	R163	R170	R170	R183	R183	R165	R171
	K163	K198	K208	K189	K184	K189	K189	K196	K203	K203	K216	K216	K198	K204
	H164	H199	H209	H190	H185	H190	H190	H197	H204	H204	H217	H217	K199	H205
	R174	R209	R219	R200	R195	R200	R200	R207	R214	R214	R227	R227	R209	R215
	M207	M237	M266	M232	M227	M242	M245	M252	M259	M259	M272	M272	M254	M260
	Y210	Y240	Y269	Y235	Y230	Y245	Y248	Y255	Y262	Y262	Y275	Y275	Y257	Y263
亲核物质	D242	D272	D301	D267	D262	D277	D277	D287	D294	D294	D307	D307	D289	D295
底物相互作用	W243	W273	W302	W268	W263	W278	W278	W288	W295	W295	W308	W308	W290	W296
	S370	S455	S472	S415	S406	S447	S451	S457	S464	S464	S477	S477	S458	S465
酸/碱	E458	E509	E534	E472	E463	E505	E509	E516	E523	E523	E536	E536	E517	E524

图2

里氏木霉整合型菌株H3A的蛋白质组成	
蛋白	总面积的%
Fv3A	9.6
Fv51A+Fv43D	14.8
Xyn 3	12.6
Bgl 1	7.5
CBH1	36.4
EGLs	5.6
CBH2	9.5
其他	4.0

图3

添加至里氏木霉整合型菌株H3A的蛋白质		
	蛋白	母液蛋白浓度 (mg/mL)
1	纯化的里氏木霉CBH1	7.4
2	纯化的里氏木霉CBH2	3.0
3	纯化的里氏木霉EGI	3.9
4	未纯化的Fv3C	2.1
5	水	
6	纯化的里氏木霉EG4	1.1
7	H3A UF浓缩物	102.8
8	纯化的里氏木霉Bg11	3.9
9	纯化的里氏木霉Xyn2	2.6
10	纯化的里氏木霉Xyn3	4.6
11	纯化的轮枝镰刀菌Fv43D	6.8
12	纯化的轮枝镰刀菌Fv51A	7.8

图4A

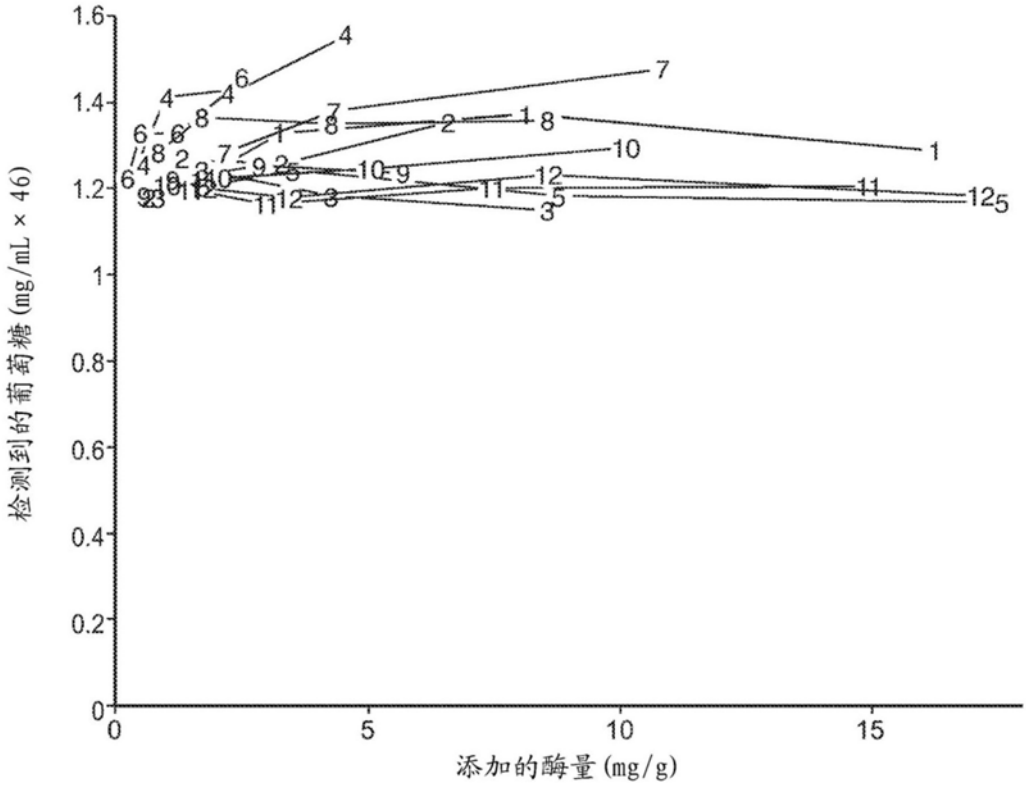

检测到的葡萄糖(mg/mL × 46)
添加的酶量(mg/g)
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
0
5
10
15

图4B

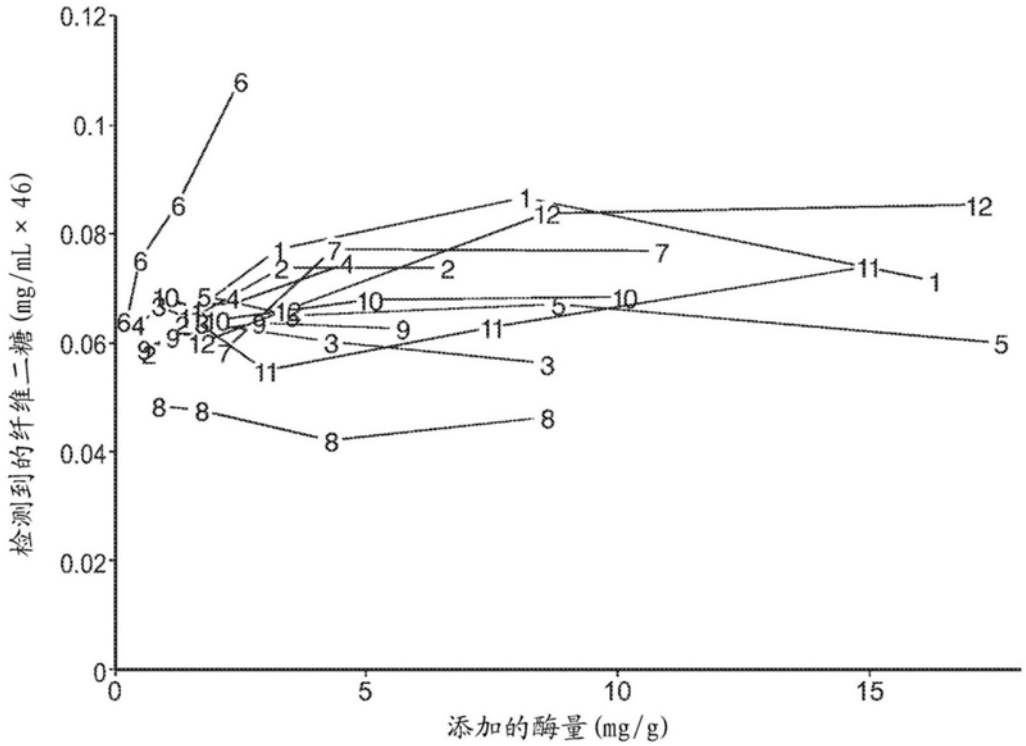

检测到的纤维二糖(mg/mL × 46)
添加的酶量(mg/g)
0
0.02
0.04
0.06
0.08
0.1
0.12
0
5
10
15

图4C

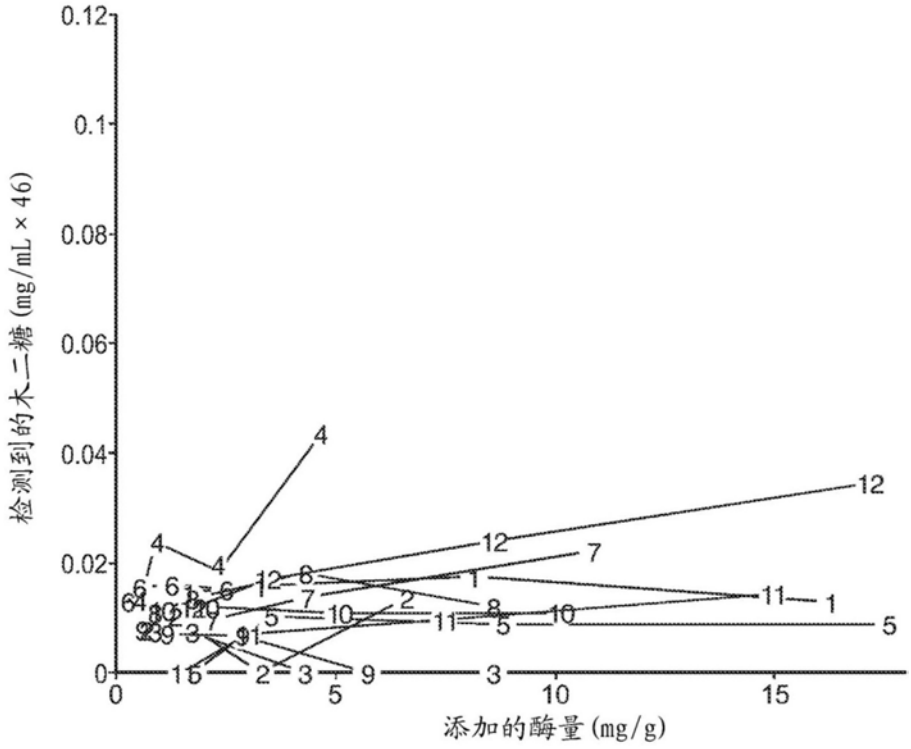

检测到的木二糖(mg/mL × 46)
添加的酶量(mg/g)
0.12
0.1
0.08
0.06
0.04
0.02
0
0
5
10
15

图4D

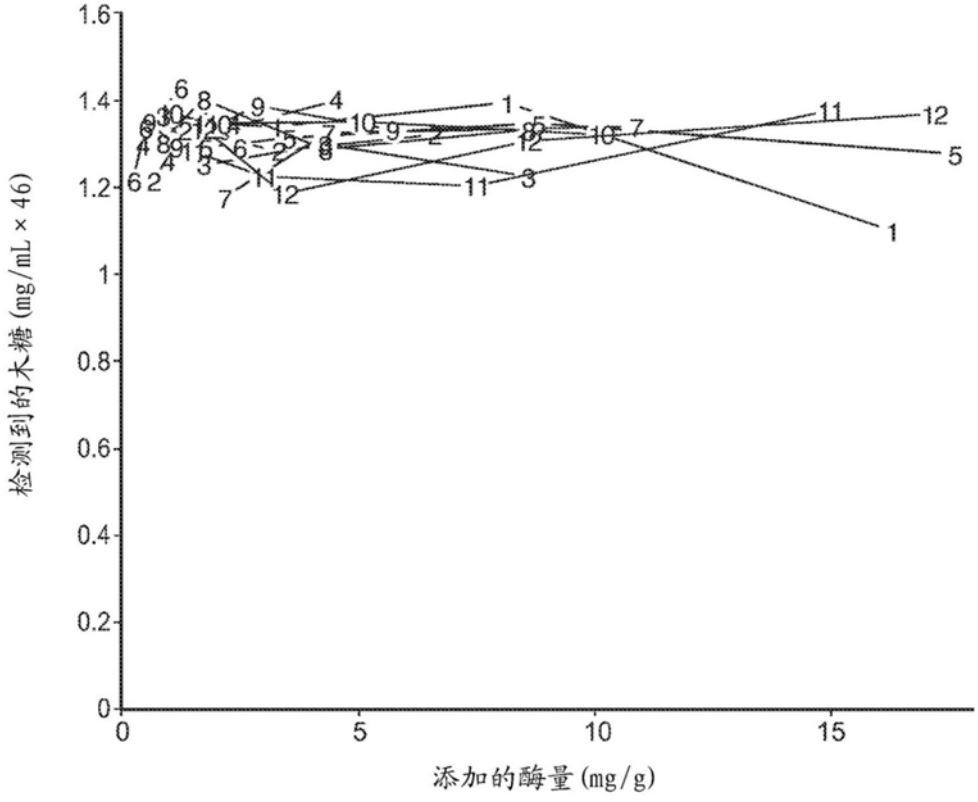

检测到的木糖(mg/mL × 46)
添加的酶量(mg/g)
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
5
10
15

图4E

酶	浓度(mg/mL)	纤维二糖酶活性		CNPG nM CNP/秒/mg 蛋白
		U/ml	U/mg	
纯化的里氏木霉Bg1u1	2.3	19.4	8.4	1242
Fv3C摇瓶	2.4	42.7	18	1156
Fv3D摇瓶	2.9	0.0	0.0	6221
Pa3C摇瓶	1.9	0.0	0.0	2
纯化的黑曲霉菌Bg1u1	2.4	244	102	168

图5A

酶	相对于里氏木霉Bg11的纤维二糖酶活性	相对于里氏木霉Bg11的CNPG活性
纯化的里氏木霉Bg11	1	1
纯化的Fv3C	1.8	1
未纯化的Fv3D	0	5
未纯化的Pa3C	0	0
纯化的黑曲霉Bg1u	12.1	0.1
纯化的FAB	1.3	1
纯化的FB	2.4	1.3
纯化的里氏木霉Bg13	5.6	1.5
纯化的Te3A	0.7-1.4	4.2-7.7

图5B

酶	重量/重量%
Xyn 3	45.0
Fv3A	15.0
Fv43D	5.0
Fv51A	35.0

图6

酶混合物	达到如下葡聚糖转化率的剂量（mg蛋白/g纤维素）：		
	70%	80%	90%
Accellerase 1500 + Multifect 木聚糖酶(Multifect Xylanase)	38	-	-
来自里氏木霉整合型菌株(H3A)的完整纤维素酶	21	28	36
75%来自里氏木霉整合型菌株(H3A)的完整纤维素酶/25% Fv3C	12	15	19

图7

SEQ ID NO:1

来自轮枝镰刀菌的GH3家族酶Fv3A的核苷酸序列

atgctgctcaatcttcaggtcgctgccagcgctttgtcgctttctcttttaggtggattggctgaggctg
ctacgccatataccttccggactgtaccaaaggacctttgagcaagaatggaatctgcgatacttcgtt
atctccagctaaaagagcggctgctctagttgctgctctgacgcccgaagagaaggtgggcaatctggtc
aggtaaaatatacccccccccataatcactattcggagattggagctgacttaacgcagcaatgcaactg
gtgcaccaagaatcggacttccaaggtacaactggtggaacgaagcccttcatggcctcgctggatctcc
aggtggtcgctttgccgacactcctccctacgacgcggccacatcatttcccatgcctcttctcatggcc
gctgctttcgacgatgatctgatccacgatatcggcaacgtcgtcggcaccgaagcgcgtgcgttcacta
acggcggttggcgcggagtcgacttctggacacccaacgtcaacccttttaaagatcctcgctggggtcg
tggctccgaaactccaggtgaagatgcccttcatgtcagccggtatgctcgctatatcgtcaggggtctc
gaaggcgataaggagcaacgacgtattgttgctacctgcaagcactatgctggaaacgactttgaggact
ggggaggcttcacgcgtcacgactttgatgccaagattactcctcaggacttggctgagtactacgtcag
gcctttccaggagtgcacccgtgatgcaaaggttggttccatcatgtgcgcctacaatgccgtgaacggc
attcccgcatgcgcaaactcgtatctgcaggagacgatcctcagagggcactggaactggacgcgcgata
acaactggatcactagtgattgtggcgccatgcaggatatctggcagaatcacaagtatgtcaagaccaa
cgctgaaggtgcccaggtagcttttgagaacggcatggattctagctgcgagtatactactaccagcgat
gtctccgattcgtacaagcaaggcctcttgactgagaagctcatggatcgttcgttgaagcgccttttcg
aagggcttgttcatactggtttctttgacggtgccaaagcgcaatggaactcgctcagttttgcggatgt
caacaccaaggaagctcaggatcttgcactcagatctgctgtggagggtgctgttcttcttaagaatgac
ggcactttgcctctgaagctcaagaagaaggatagtgttgcaatgatcggattctgggccaacgatactt
ccaagctgcagggtggttacagtggacgtgctccgttcctccacagcccgctttatgcagctgagaagct
tggtcttgacaccaacgtggcttggggtccgacactgcagaacagctcatctcatgataactggaccacc
aatgctgttgctgcggcgaagaagtctgattacattctctactttggtggtcttgacgcctctgctgctg
gcgaggacagagatcgtgagaaccttgactggcctgagagccagctgacccttcttcagaagctctctag
tctcggcaagccactggttgttatccagcttggtgatcaagtcgatgacaccgctcttttgaagaacaag
aagattaacagtattctttgggtcaattaccctggtcaggatggcggcactgcagtcatggacctgctca
ctggacgaaagagtcctgctggccgactacccgtcacgcaatatcccagtaaatacactgagcagattgg
catgactgacatggacctcagacctaccaagtcgttgccagggagaacttatcgctggtactcaactcca
gttcttccctacggctttggcctccactacaccaagttccaagccaagttcaagtccaacaagttgacgt
ttgacatccagaagcttctcaagggctgcagtgctcaatactccgatacttgcgcgctgcccccatcca
agttagtgtcaagaacaccggccgcattacctccgactttgtctctctggtctttatcaagagtgaagtt
ggacctaagccttaccctctcaagacccttgcggcttatggtcgcttgcatgatgtcgcgccttcatcga
cgaaggatatctcactggagtggacgttggataacattgcgcgacggggagagaatggtgatttggttgt
ttatcctgggacttacactctgttgctggatgagcctacgcaagccaagatccaggttacgctgactgga
aagaaggctattttggataagtggcctcaagaccccaagtctgcgtaa

图8A

SEQ ID NO:2

Fv3A的蛋白质序列

mllnlqvaasalslsllgglaeaatpytlpdctkgplskngicdtslspakraaalvaaltpeekvgnlv
snATGAPRIGLPRYNWWNEALHGLAGSPGGRFADTPPYDAATSFPMPLLMAAAFDDDLIHDIGNVVGTEA
RAFTNGGWRGVDFWTPNVNPFKDPRWGRGSETPGEDALHVSRYARYIVRGLEGDKEQRRIVATCKHYAGN
DFEDWGGFTRHDFDAKITPQDLAEYYVRPFQECTRDAKVGSIMCAYNAVNGIPACANSYLQETILRGHWN
WTRDNNWITSDCGAMQDIWQNHKYVKTNAEGAQVAFENGMDssceytttsdvsdsykqgllteklmdrsl
krlfeglvhtgffdgakaqwnslsfadvntkeaqdlalrsavegAVLLKNDGTLPLKLKKKDSVAMIGFW
ANDTSKLQGGYSGRAPFLHSPLYAAEKLGLDTNVAWGPTLQNSSSHDNWTTNAVAAAKKSDYILYFGGLD
ASAAGEDRDRENLDWPESQLTLLQKLSSLGKPLVVIQLGDQVDDTALLKNKKINSILWVNYPGQDGGTAV
MDLLTGRKSPAGRLPVTQYPSKYTEQIGMTDMDLRPTKSLPGRTYRWYSTPVLPYGFGLHYTkfqakfks
nkltfdiqkllkgcsaqysdtcalppiqvsvkntgritsdfvslvfiksevgpkpyplktlaaygrlhdv
apsstkdislewtldniarrgengdlvvypgtytllldeptqakiqvtltgkkaildkwpqdpksa

图8B

SEQ ID NO:3

来自绳状青霉的GH43家族酶Pf43A的核苷酸序列

atgcttcagcgatttgcttatattttaccactggctctattgagtgttggagtgaaagccgacaacccct
ttgtgcagagcatctacaccgctgatccggcaccgatggtatacaatgaccgcgtttatgtcttcatgga
ccatgacaacaccggagctacctactacaacatgacagactggcatctgttctcgtcagcagatatggcg
aattggcaagatcatggcattccaatgagcctggccaatttcacctgggccaacgcgaatgcgtgggccc
cgcaagtcatccctcgcaacggccaattctacttttatgctcctgtccgacacaacgatggttctatggc
tatcggtgtgggagtgagcagcaccatcacaggtccataccatgatgctatcggcaaaccgctagtagag
aacaacgagattgatcccaccgtgttcatcgacgatgacggtcaggcatacctgtactggggaaatccag
acctgtggtacgtcaaattgaaccaagatatgatatcgtacagcgggagccctactcagattccactcac
cacggctggatttggtactcgaacgggcaatgctcaacggccgaccacttttgaagaagctccatgggta
tacaaacgcaacggcatctactatatcgcctatgcagccgattgttgttctgaggatattcgctactcca
cgggaaccagtgccactggtccgtggacttatcgaggcgtcatcatgccgacccaaggtagcagcttcac
caatcacgagggtattatcgacttccagaacaactcctacttttttctatcacaacggcgctcttcccggc
ggaggcggctaccaacgatctgtatgtgtggagcaattcaaatacaatgcagatggaaccattccgacga
tcgaaatgaccaccgccggtccagctcaaattgggactctcaacccttacgtgcgacaggaagccgaaac
ggcggcatggtcttcaggcatcactacggaggtttgtagcgaaggcggaattgacgtcgggtttatcaac
aatggcgattacatcaaagttaaaggcgtagctttcggttcaggagcccattctttctcagcgcgggttg
cttctgcaaatagcggcggcactattgcaatacacctcggaagcacaactggtacgctcgtgggcacttg
tactgtccccagcactggcggttggcagacttggactaccgttacctgttctgtcagtggcgcatctggg
acccaggatgtgtattttgttttcggtggtagcggaacaggatacctgttcaactttgattattggcagt
tcgcataa

图9A

SEQ ID NO:4

Pf43A的蛋白质序列

mlqrfavilplallsvqvka**dnpfvqsiytadpapmvyndrvyvfmdhdntgatyynmtdwhlfssadma**
nwqdhgipmslanftwananawapqviprngqfyfyapvrhndgsmaigvgvsstitgpyhdaigkplve
nneidptvfidddgqaylywgnpdlwyvklnqdmisysgsptqiplttagfgtrtgnaqrpttfeeapwv
ykrngiyyiayaadccsedirystgtsatgpwtyrgvimptqgssftnhegiidfqnnsyffyhngalpg
gggyqrsvcveqfkynadgti*ptiemttagpaqigtlnpyvrq*EAETAAWSSGITTEVCSEGGIDVGFIN
NGDYIKVKGVAFGSGAHSFSARVASANSGGTIAIHLGSTTGTLVGTCTVPSTGGWQTWTTVTCSVSGASG
TQDVYFVFGGSGTGYLFNFDYWQFa

图9B

SEQ ID NO:5

来自轮枝镰刀菌的GH43家族酶Fv43E的核苷酸序列

atgaaggtatactggctcgtggcgtgggccacttctttgacgccggcactggctggcttgattggacacc
gtcgcgccaccaccttcaacaatcctatcatctactcagactttccagataacgatgtattcctcggtcc
agataactactactacttctctgcttccaacttccacttcagcccaggagcacccgttttgaagtctaaa
gatctgctaaactgggatctcatcggccattcaattccccgcctgaactttggcgacggctatgatcttc
ctcctggctcacgttattaccgtggaggtacttgggcatcatccctcagatacagaaagagcaatggaca
gtggtactggatcggctgcatcaacttctggcagacctgggtatacactgcctcatcgccggaaggtcca
tggtacaacaagggaaacttcggtgataacaattgctactacgacaatggcatactgatcgatgacgatg
ataccatgtatgtcgtatacggttccggtgaggtcaaagtatctcaactatctcaggacggattcagcca
ggtcaaatctcaggtagttttcaagaacactgatattggggtccaagacttggagggtaaccgcatgtac
aagatcaacgggctctactatatcctaaacgatagcccaagtggcagtcagacctggatttggaagtcga
aatcaccctggggcccttatgagtctaaggtcctcgccgacaaagtcaccccgcctatctctggtggtaa
ctcgccgcatcagggtagtctcataaagactcccaatggtggctggtacttcatgtcattcacttgggcc
tatcctgccggccgtcttccggttcttgcaccgattacgtggggtagcgatggtttccccattcttgtca
agggtgctaatggcggatggggatcatcttacccaacacttcctggcacggatggtgtgacaaagaattg
gacaaggactgataccttccgcggaacctcacttgctccgtcctgggagtggaaccataatccggacgtc
aactccttcactgtcaacaacggcctgactctccgcactgctagcattacgaaggatatttaccaggcga
ggaacacgctatctcaccgaactcatggtgatcatccaacaggaatagtgaagattgatttctctccgat
gaaggacggcgaccgggccgggctttcagcgtttcgagaccaaagtgcatacatcggtattcatcgagat
aacggaaagttcacaatcgctacgaagcatgggatgaatatggatgagtggaacggaacaacaacagacc
tgggacaaataaaagccacagctaatgtgccttctggaaggaccaagatctggctgagacttcaacttga
taccaacccagcaggaactggcaacactatcttttcttacagttgggatggagtcaagtatgaaacactg
ggtcccaacttcaaactgtacaatggttgggcattctttattgcttaccgattcggcatcttcaacttcg
ccgagacggctttaggaggctcgatcaaggttgagtctttcacagctgcatag

图10A

SEQ ID NO:6

Fv43E的蛋白质序列

mkvywlvawatsltpalaglighrrattfnnpiiysdfpdndvflgpdnyyyfsasnfhfspgapvlksk
dllnwdlighsiprlnfgdgydlppgsryyrggtwasslryrksngqwywigcinfwqtwvytasspegp
wynkgnfgdnncyydngilidddtmyvvygsgevkvsqlsqdgfsqvksqvvfkntdigvqdlegnrmy
kinglyyilndspsgsqtwiwkskspwgpyeskvladkvtppisggnsphqgsliktpnggwyfmsftwa
ypagrlpvlapitwgsdgfpilvkganggwgssyptlpgtdgvtknwtrtdtfrgtslapswewnhnpdv
nsftvnngltlrtasitkdiyqarntlshrthgdhptgivkidfspmkdgdraglsafrdqsayigihrd
ngkftiatkhgmnmdewngtttdlgqikatanvpsgrtkiwlrlqldtnpagtgntifsyswdgvkyetl
gpnfklyngwaffiayrfgifnfaetalggsikvesftaa

图10B

SEQ ID NO:7

来自轮枝镰刀菌的GH39家族酶Fv39A的核苷酸序列

atgcactacgctaccctcaccactttggtgctggctctgaccaccaacgtcgctgcacagcaaggcacag
caactgtcgacctctccaaaaatcatggaccggcgaaggcccttggttcaggcttcatatacggctggcc
tgacaacggaacaagcgtcgacacctccataccagatttcttggtaactgacatcaaattcaactcaaac
cgcggcggtggcgcccaaatcccatcactgggttgggccagaggtggctatgaaggatacctcggccgct
tcaactcaaccttatccaactatcgcaccacgcgcaagtataacgctgactttatcttgttgcctcatga
cctctggggtgcggatggcgggcagggttcaaactccccgtttcctggcgacaatggcaattggactgag
atggagttattctggaatcagcttgtgtctgacttgaaggctcataatatgctggaaggtcttgtgattg
atgtttggaatgagcctgatattgatatcttttgggatcgcccgtggtcgcagtttcttgagtattacaa
tcgcgcgaccaaactacttcggtgagtctactactgatccatacgtatttacagtgagctgactggtcga
attagaaaaacacttcccaaaactcttctcagtggcccagccatggcacattctcccattctgtccgatg
ataaatggcatacctggcttcaatcagtagcgggtaacaagacagtccctgatatttactcctggcatca
gattggcgcttgggaacgtgagccggacagcactatccccgactttaccaccttgcgggcgcaatatggc
gttcccgagaagccaattgacgtcaatgagtacgctgcacgcgatgagcaaaatccagccaactccgtct
actacctctctcaactagagcgtcataaccttagaggtcttcgcgcaaactggggtagcggatctgacct
ccacaactggatgggcaacttgatttacagcactaccggtacctcggaggggacttactaccctaatggt
gaatggcaggcttacaagtactatgcggccatggcagggcagagacttgtgaccaaagcatcgtcggact
tgaagtttgatgtctttgccactaagcaaggccgtaagattaagattatagccggcacgaggaccgttca
agcaaagtataacatcaaaatcagcggtttggaagtagcaggacttcctaagatgggtacggtaaaggtc
cggacttatcggttcgactgggctgggccgaatggaaaggttgacgggcctgttgatttgggggagaaga
agtatacttattcggccaatacggtgagcagcccctctacttga

图11A

SEQ ID NO:8

Fv39A的蛋白质序列

mhyatlttlvlalttnvaaqqgtatvdlsknhgpakalgsgfiygwpdngtsvdtsipdflvtdikfnsn
rgggaqipslgwarggyegylgrfnstlsnyrttrkynadfillphdlwgadgggqgsnspfpgdngnwte
melf**wnqlvsdlkahnmleglvidvwnepdidifwdrpwsqfleyynratkllrktlpktllsgpamahs**
pilsddkwhtwlqsvagnktvpdiyswhqigawerepdstipdfttlraqygvpekpidvneyaardeqn
pansvyylsqlerhnlrglranwgsgsdlhnwmgnliysttgtsegtyypngewqaykyyaamagqrlvt
kassdlkfdvfatkqgrkikiiagtrtvqakynikisglevaglpkmgtvkvrtyrfdwagpngkvdgpv
dlgekkytysantvsspst

图11B

SEQ ID NO:9

来自轮枝镰刀菌的GH43家族酶Fv43A的核苷酸序列

atgtggctgacctccccattgctgttcgccagcaccctcctgggcctcactggcgttgctctagcagaca
accccatcgtccaagacatctacaccgcagacccagcaccaatggtctacaatggccgcgtctacctctt
cacaggccatgacaacgacggctctaccgacttcaacatgacagactggcgtctcttctcgtcagcagac
atggtcaactggcagcaccatggtgtccccatgagcttaaagaccttcagctgggccaacagcagagcct
gggctggtcaagtcgttgcccgaaacggaaagttttacttctatgttcctgtccgtaatgccaagacggg
tggaatggctattggtgtcggtgttagtaccaacatccttgggccctacactgatgcccttggaaagcca
ttggtcgagaacaatgagatcgacccaactgtctacatcgacactgatggccaggcctatctctactggg
gcaaccctggattgtactacgtcaagctcaaccaagacatgctctcctacagtggtagcatcaacaaagt
atcgctcacaacagctggattcggcagccgcccgaacaacgcgcagcgtcctactactttcgaggaagga
ccgtggctgtacaagcgtggaaatctctactacatgatctacgcagccaactgctgttccgaggacattc
gctactcaactggacccagcgccactggaccttggacttaccgcggtgtcgtgatgaacaaggcgggtcg
aagcttcaccaaccatcctggcatcatcgactttgagaacaactcgtacttcttttaccacaatggcgct
cttgatggaggtagcggttatactcggtctgtggctgtcgagagcttcaagtatggttcggacggtctga
tccccgagatcaagatgactacgcaaggcccagcgcagctcaagtctctgaacccatatgtcaagcagga
ggccgagactatcgcctggtctgagggtatcgagactgaggtctgcagcgaaggtggtctcaacgttgct
ttcatcgacaatggtgactacatcaaggtcaagggagtcgactttggcagcaccggtgcaaagacgttca
gcgcccgtgttgcttccaacagcagcggaggcaagattgagcttcgacttggtagcaagaccggtaagtt
ggttggtacctgcacggtaacgactacgggaaactggcagacttataagactgtggattgccccgtcagt
ggtgctactggtacgagcgatctattctttgtcttcacgggctctgggtctggctctctgttcaacttca
actggtggcagtttagctaa

图12A

SEQ ID NO:10

Fv43A的蛋白质序列

mwltspllfastllqltqvala**dnpivqdiytadpapmvyngrvylftghdndgstdfnmtdwrlfssad**
mvnwqhhgvpmslktfswansrawagqvvarngkfyfyvpvrnaktggmaigvgvstnilgpytdalgkp
lvenneidptvyidtdgqaylywgnpglyyvklnqdmlsysgsinkvslttagfgsrpnnaqrpttfeeg
pwlykrgnlyymiyaanccsedirystgpsatgpwtyrgvvmnkagrsftnhpgiidfennsyffyhnga
ldggsgytrsvavesfkygsdg*lipeikmttqgpaqlksl*NPYVKQEAETIAWSEGIETEVCSEGGLNVA
FIDNGDYIKVKGVDFGSTGAKTFSARVASNSSGGKIELRLGSKTGKLVGTCTVTTTGNWQTYKTVDCPVS
GATGTSDLFFVFTGSGSGSLFNFNWWQFs

图12B

SEQ ID NO:11

来自轮枝镰刀菌的GH43家族酶Fv43B的核苷酸序列

atgcgcttctcttggctattgtgccccttctagcgatgggaagtgctcttcctgaaacgaagacggatg
tttcgacatacaccaaccctgtccttccaggatggcactcggatccatcgtgtatccagaaagatggcct
ctttctctgcgtcacttcaacattcatctccttcccaggtcttcccgtctatgcctcaagggatctagtc
aactggcgtctcatcagccatgtctggaaccgcgagaaacagttgcctggcattagctggaagacggcag
gacagcaacagggaatgtatgcaccaaccattcgataccacaagggaacatactacgtcatctgcgaata
cctgggcgttggagatattattggtgtcatcttcaagaccaccaatccgtgggacgagagtagctggagt
gaccctgttaccttcaagccaaatcacatcgaccccgatctgttctgggatgatgacggaaaggtttatt
gtgctacccatggcatcactctgcaggagattgatttggaaactggagagcttagcccggagcttaatat
ctggaacggcacaggaggtgtatggcctgagggtccccatatctacaagcgcgacggttactactatctc
atgattgccgagggtggaactgccgaagaccacgctatcacaatcgctcgggcccgcaagatcaccggcc
cctatgaagcctacaataacaacccaatcttgaccaaccgcgggacatctgagtacttccagactgtcgg
tcacggtgatctgttccaagataccaagggcaactggtggggtctttgtcttgctactcgcatcacagca
cagggagtttcacccatgggccgtgaagctgttttgttcaatggcacatggaacaagggcgaatggccca
agttgcaaccagtacgaggtcgcatgcctggaaacctcctcccaaagccgacgcgaaacgttcccggaga
tgggcccttcaacgctgacccagacaactacaacttgaagaagactaagaagatccctcctcactttgtg
caccatagagtcccaagagacggtgccttctctttgtcttccaagggtctgcacatcgtgcctagtcgaa
acaacgttaccggtagtgtgttgccaggagatgagattgagctatcaggacagcgaggtctagctttcat
cggacgccgccaaactcacactctgttcaaatatagtgttgatatcgacttcaagcccaagtccgatgat
caggaagctggaatcaccgttttccgcacgcagttcgaccatatcgatcttggcattgttcgtcttccta
caaaccaaggcagcaacaagaaatctaagcttgccttccgattccgggccacaggagctcagaatgttcc
tgcaccgaaggtagtaccggtccccgatggctgggagaagggcgtaatcagtctacatatcgaggcagcc
aacgcgacgcactacaaccttggagcttcgagccacagaggcaagactctcgacatcgcgacagcatcag
caagtcttgtgagtggaggcacgggttcatttgttggtagtttgcttggaccttatgctacctgcaacgg
caaaggatctggagtggaatgtcccaagggaggtgatgtctatgtgacccaatggacttataagcccgtg
gcacaagagattgatcatggtgtttttgtgaaatcagaattgtag

图13A

SEQ ID NO:12

Fv43B的蛋白质序列

mrfswllcpllamqsalpetktdvst**ytnpvlpgwhsdpsciqkdglflcvtstfisfpglpvyasrdlv**
nwrlishvwnrekqlpgiswktagqqqgmyaptiryhkgtyyviceylgvgdiigvifkttnpwdesswsdpvtfkpnhidpdlfwdddgkvycathgitlqeidletgelspelniwngtggvwpegphiykrdgyyyl
miaeggtaedhaitiararkitgpyeaynnnpiltnrgtseyfqtvghgdlfqdtkgnwwglclatrita
qgvspmgreavlfngtwnkgewpklqpvrgrmpgnllpkptrnvpgdgpfnadpdnynlkktkkipphfv
hhrvprdgafslsskglhivpsrnnvtgsvlpgdeielsgqrglafigrrqthtlfkysvdidfkpksdd
qeagitvfrtqfdhidlgivrlptnqgsnkksklafrfratgaqnvpapkvvpvpdgwekgvislhieaa
nathynlgasshrgktldiatasaslvsggtgsfvgsllgpyatcngkgsgvecpkggdvyv
tqwtykpvaqeidhgvfvksel

图13B

SEQ ID NO:13

来自鹅掌柄孢霉的GH51家族酶Pa51A的核苷酸序列

atgatccacctcaagccagccctcgcggcgttgttggcgctgtcgacgcaatgtgtggctattgatttgt
ttgtcaagtcttcgggggggaataagacgactgatatcatgtatggtcttatgcacgaggtatgtgtttt
gcgagatctccctttgtttttgcgcactgctgacatggagactgcaaacaggatatcaacaactccggc
gacggcggcatctacgccgagctaatctccaaccgcgcgttccaagggagtgagaagttcccctccaacc
tcgacaactggagccccgtcggtggcgctacccttacccttcagaagcttgccaagccccttcctctgc
gttgccttactccgtcaatgttgccaaccccaaggagggcaagggcaagggcaaggacaccaaggggaag
aaggttggcttggccaatgctgggttttggggtatggatgtcaagaggcagaagtacactggtagcttcc
acgttactggtgagtacaagggtgactttgaggttagcttgcgcagcgcgattaccggggagacctttgg
caagaaggtggtgaagggtgggagtaagaaggggaagtggaccgagaaggagtttgagttggtgcctttc
aaggatgcgcccaacagcaacaacacctttgttgtgcagtgggatgccgaggtatgtgcttctttgatat
tggctgagatagaagttgggttgacatgatgtggtgcagggcgcaaaggacggatctttggatctcaact
tgatcagcttgttccctccgacattcaagggaaggaagaatgggctgagaattgatcttgcgcagacgat
ggttgagctcaagccggtaagtcctctctagtcagaaaagtagagcctttgttaacgcttgacagacctt
cttgcgcttccccggtggcaacatgctcgagggtaacaccttggacacttggtggaagtggtacgagacc
attggccctctgaaggatcgcccgggcatggctggtgtctgggagtaccagcaaacccttggcttgggtc
tggtcgagtacatggagtgggccgatgacatgaacttggagcccagtatgtgatcccattttctggagtg
acttctcttgctaacgtatccacagttgtcggtgtcttcgctggtcttgccctcgatggctcgttcgttc
ccgaatccgagatgggatgggtcatccaacaggctctcgacgaaatcgagttcctcactggcgatgctaa
gaccaccaaatggggtgccgtccgcgcgaagcttggtcaccccaagccttggaaggtcaagtgggttgag
atcggtaacgaggattggcttgccggacgccctgctggcttcgagtcgtacatcaactaccgcttcccca
tgatgatgaaggccttcaacgaaaagtaccccgacatcaagatcatcgcctcgccctccatcttcgacaa
catgacaatccccgcgggtgctgccggtgatcaccacccgtacctgactcccgatgagttcgttgagcga
ttcgccaagttcgataacttgagcaaggataacgtgacgctcatcggcgaggctgcgtcgacgcatccta
acggtggtatcgcttgggagggagatctcatgcccttgccttggtggggcggcagtgttgctgaggctat
cttcttgatcagcactgagagaaacggtgacaagatcatcggtgctacttacgcgcctggtcttcgcagc
ttggaccgctggcaatggagcatgacctgggtgcagcatgccgccgacccggccctcaccactcgctcga
ccagttggtatgtctggagaatcctcgcccaccacatcatccgtgagacgctcccggtcgatgccccggc
cggcaagcccaactttgaccctctgttctacgttgccggaaagagcgagagtggcaccggtatcttcaag
gctgccgtctacaactcgactgaatcgatcccggtgtcgttgaagtttgatggtctcaacgagggagcgg
ttgccaacttgacggtgcttactgggccggaggatccgtatggatacaacgacccccttcactggtatcaa
tgttgtcaaggagaagaccaccttcatcaaggccggaaagggcggcaagttcaccttcaccctgccgggc
ttgagtgttgctgtgttggagacggccgacgcggtcaagggtggcaagggaaagggcaagggcaagggaa
agggtaactga

图14A

SEQ ID NO:14

Pa51A的蛋白质序列

mihlkpalaallalstqcvaidlfvkssggnkttdimyglmhedinnsgdggiyaelisnrafqgsekfp
snldnwspvggatltlqklakplssalpysvnvanpkegkgkgkdtkgkkvglanagfwgmdvkrqkytg
sfhvtgeykgdfevslrsaitgetfgkkvvkggskkgkwtekefelvpfkdapnsnntfvvqwdaegakd
gsldlnlislfpptfkgrknglridlaqtmvelkptflrfpggnmlegntldtwwkwyetigplkdrpgm
agvweyqqtlglglveymewaddmnlepivgvfaglaldgsfvpesemgwviqqaldeiefltgdakttk
wgavraklghpkpwkvkwveignedwlagrpagfesyinyrfpmmmkafnekypdikiiaspsifdnmti
pagaagdhhpyltpdefverfakfdnlskdnvtligeaasthpnggia**wegdlmplpwwggsvaeaifli**
sterngdkiigatyapglrsldrwqwsmtwvqhaadpalttrstswyvwrilahhiiretlpvdapagkp
nfdplfyvagksesgtgifkaavynstesipvslkfdglnegavanltvltgpedpygyndpftginvvk
ekttfikagkggkftftlpglsvavletadavkggkgkgkgkgkgn

图14B

SEQ ID NO:15

来自玉米赤霉的GH43家族酶Gz43A的核苷酸序列

atgaagtccaagttgttattcccactcctctctttcgttggtcaaagtcttgccaccaacgacgactgtc
ctctcatcactagtagatggactgcggatccttcggctcatgtctttaacgacaccttgtggctctaccc
gtctcatgacatcgatgctggatttgagaatgatcctgatggaggccagtacgccatgagagattaccat
gtctactctatcgacaagatctacggttccctgccggtcgatcacggtacggccctgtcagtggaggatg
tcccctgggcctctcgacagatgtgggctcctgacgctgcccacaagaacggcaaatactacctatactt
ccctgccaaagacaaggatgatatcttcagaatcggcgttgctgtctcaccaacccccggcggaccattc
gtccccgacaagagttggatccctcacactttcagcatcgaccccgccagtttcgtcgatgatgatgaca
gagcctacttggcatggggtggtatcatgggtggccagcttcaacgatggcaggataagaacaagtacaa

图15A

SEQ ID NO:16

Gz43A的蛋白质序列

mkskllfpllsfvgqslatnddcplitsrwtadpsahvfndtlwlypshdid**agfendpdggqyamrdyh**
vysidkiygslpvdhgtalsvedvpwasrqmwapdaahkngkyylyfpakdkddifrigvavsptpggpf
vpdkswiphtfsidpasfvdddraylawggimggqlqrwqdknkynesgtepgngtaalspqiaklskd
mhtlaekprdmlildpktgkpllsededrrffegpwihkrnkiyyltystgtthylvyatsktpygpyty
qgrilepvdgwtthssivkyqgqwwlfyhdaktsgkdylrqvkakkiwydskgkiltkkp

图15B

SEQ ID NO:17
来自尖孢镰刀菌的GH43家族酶Fo43A的核苷酸序列

atgcagctcaagtttctgtcttcagcattgctgttctctctgaccagcaaatgcgctgcgcaagacacta
atgacattcctcccctgatcaccgacctctggtccgcagatccctcggctcatgttttcgaaggcaagct
ctgggtttacccatctcacgacatcgaagccaatgttgtcaacggcacaggaggcgctcaatacgccatg
agggattaccatacctactccatgaagagcatctatggtaaagatcccgttgtcgaccacggcgtcgctc
tctcagtcgatgacgttccctgggcgaagcagcaaatgtgggctcctgacgcagctcataagaacggcaa
atattatctgtacttccccgccaaggacaaggatgagatcttcagaattggagttgctgtctccaacaag
cccagcggtcctttcaaggccgacaagagctggatccctggcacgtacagtatcgatcctgctagctacg
tcgacactgataacgaggcctacctcatctggggcggtatctggggcggccagctccaagcctggcagga
taaaaagaactttaacgagtcgtggattggagacaaggctgctcctaacggcaccaatgccctatctcct
cagatcgccaagctaagcaaggacatgcacaagatcaccgaaacaccccgcgatctcgtcattctcgccc
ccgagacaggcaagcctcttcaggctgaggacaacaagcgacgattcttcgagggcccttggatccacaa
gcgcggcaagctttactacctcatgtactccaccggtgatacccacttccttgtctacgctacttccaag
aacatctacggtccttatacctaccggggcaagattcttgatcctgttgatgggtggactactcatggaa
gtattgttgagtataagggacagtggtggcttttctttgctgatgcgcatacgtctggtaaggattacct
tcgacaggtgaaggcgaggaagatctggtatgacaagaacggcaagatcttgcttcaccgtccttag

图16A

SEQ ID NO:18
Fo43A的蛋白质序列

mqlkflssallfsltskcaaqdtndipplitdlwsadpsahvfegklwvypshdieanvvngtggaqyam
rdyhtysmksiygkdpvvdhgvalsvddvpwakqqm**wapdaahkngkyylyfpakdkdeifrigvavsnk**
psgpfkadkswipgtysidpasyvdtdneayliwggiwggqlqawqdknfneswigdkaapngtnalsp
qiaklskdmhkitetprdlvilapetgkplqaednkrrffegpwihkrgklyylmystgdthflvyatsk
niygpytyrgkildpvdgwtthgsiveykgqwwlffadahtsgkdylrqvkarkiwydkngkillhrp

图16B

SEQ ID NO:19

来自烟曲霉的GH43家族酶Af43A的核苷酸序列

```
atggcagctccaagtttatcctaccccacaggtatccaatcgtataccaatcctctcttccctggttggc
actccgatcccagctgtgcctacgtagcggagcaagacacctttttctgcgtgacgtccactttcattgc
cttccccggtcttcctctttatgcaagccgagatctgcagaactggaaactggcaagcaatattttcaat
cggcccagccagatccctgatcttcgcgtcacggatggacagcagtcgggtatctatgcgcccactctgc
gctatcatgagggccagttctacttgatcgtttcgtacctgggcccgcagactaagggcttgctgttcac
ctcgtctgatccgtacgacgatgccgcgtggagcgatccgctcgaattcgcggtacatggcatcgacccg
gatatcttctgggatcacgacgggacggtctatgtcacgtccgccgaggaccagatgattaagcagtaca
cactcgatctgaagacgggggcgattggcccggttgactacctctggaacggcaccggaggagtctggcc
cgagggcccgcacatttacaagagagacggatactactacctcatgatcgcagagggaggtaccgagctc
ggccactcggagaccatggcgcgatctagaacccggacaggtccctgggagccatacccgcacaatccgc
tcttgtcgaacaagggcacctcggagtacttccagactgtgggccatgcggacttgttccaggatgggaa
cggcaactggtgggccgtggcgttgagcacccgatcagggcctgcatggaagaactatcccatgggtcgg
gagacggtgctcgcccccgccgcttgggagaagggtgagtggcctgtcattcagcctgtgagaggccaaa
tgcaggggccgtttccaccaccaaataagcgagttcctcgcggcgagggcggatggatcaagcaacccga
caaagtggatttcaggcccggatcgaagataccggcgcacttccagtactggcgatatcccaagacagag
gattttaccgtctcccctcggggccacccgaatactcttcggctcacaccctccttttacaacctcaccg
gaactgcggacttcaagccggatgatggcctgtcgcttgttatgcgcaaacagaccgacaccttgttcac
gtacactgtggacgtgtcttttgaccccaaggttgccgatgaagaggcgggtgtgactgttttccttacc
cagcagcagcacatcgatcttggtattgtccttctccagacaaccgaggggctgtcgttgtccttccggt
tccgcgtggaaggccgcggtaactacgaaggtcctcttccagaagccaccgtgcctgttcccaaggaatg
gtgtggacagaccatccggcttgagattcaggccgtgagtgacaccgagtatgtctttgcggctgccccg
gctcggcaccctgcacagaggcaaatcatcagccgcgccaactcgttgattgtcagtggtgatacgggac
ggtttactggctcgcttgttggcgtgtatgccacgtcgaacgggggtgccggatccacgcccgcatatat
cagcagatggagatacgaaggacggggccagatgattgattttggtcgagtggtcccgagctactga
```

图17A

SEQ ID NO:20

Af43A的蛋白质序列

```
maapslsyptgiqsytnplfpgwhsdpscayvaeqdtffcvtstfiafpglplyasrdlqnwklasnifn
rpsqipdlrvtdgqqsgiyaptlryhegqfylivsylgpqtkgllftssdpyddaawsdplefavhgidp
difwdhdgtvyvtsaedqmikqytldlktgaigpvdylwngtggvwpegphiykrdgyyylmiaeggtel
ghsetmarsrtrtgpwepyphnpllsnkgtseyfqtvghadlfqdgngnwwavalstrsgpawknypmgr
etvlapaawekgewpviqpvrgqmqgpfpppnkrvprgeggwikqpdkvdfrpgskipahfqywrypkte
dftvsprghpntlrltpsfynltgtadfkpddglslvmrkqtdtlftytvdvsfdpkvadeeagvtvflt
qqqhidlgivllqtteglslsfrfrvegrgnyegplpeatvpvpkewcgqtirleiqavsdteyvfaaap
arhpaqrqiisranslivsgdtgrftgslvgvyatsnggagstpayisrwryegrgqmidfgrvvpsy
```

图17B

SEQ ID NO:21

来自绳状青霉的GH51家族酶Pf51A的核苷酸序列

atgggaaagatgtggcattcgatcttggttgtgttgggcttattgtctgtcgggcatgccatcactatca
acgtgtcccaaagtggcggcaataagaccagtcctttgcaatatggtctgatgttcgaggtaatccttct
cttataccacatataaaagttgcgtcatttctaagacaagtcaaggacataaatcacggcggtgatggcg
gtctgtatgcagagcttgttcgaaaccgagcattccaaggtagcaccgtctatccagcaaacctcgatgg
atacgactcggtcaatggagcaatcctagcgcttcagaatttgacaaaccctctatcaccctccatgcct
agctctctcaacgtcgccaaggggtccaacaatggaagcatcggtttcgcaaatgaaggctggtgggggga
tagaagtcaagccgcaaagatacgcgggctcattctacgtccaggggactatcaaggagatttcgacat
ctctcttcagtcgaaattgacacaagaagtcttcgcaacggcaaaagtcaggtcctcgggcaaacacgag
gactgggttcaatacaagtacgagttggtgcccaaaaaggcagcatcaaacaccaataacactctgacca
ttacttttgactcaaaggtatgttaaattttgggtttagttcgatgtctggcaattgtcttacgagaaac
gtagggattgaaagacggatccttgaacttcaacttgatcagcctatttcccccaacttacaacaatcgg
cccaatggcctaagaatcgacctggttgaagctatggctgaactagaggggtaagctcttacaaatcaa
ctttatctttacgaagactaatgtgaaaacttagaaatttctgcggtttccaggcggtagcgatgtggaa
ggtgtacaagctccttactggtataagtggaatgaaacggtaggagatctcaaggaccgttatagtaggc
ccagtgcatggacgtacgaagaaagcaatggaattggcttgattgagtacatgaattggtgtgatgacat
ggggcttgagccgagtgagtgtattccattcagcgtcaaatccagtgttctaatcatacacatcagttct
tgccgtatgggatggacattacctttcgaacgaagtgatatcggaaaacgatttgcagccatatatcgac
gacaccctcaaccaactggaattcctgatgggtgccccagatacgccatatggtagttggcgtgcgtctc
tgggctatccgaagccgtggacgattaactacgtcgagattggaaacgaagacaatctatacgggggact
agaaacatacatcgcctaccggtttcaggcatattacgacgctataacagctaaatatccccatatgacg
gtcatggaatctttgacggagatgcctggtccggcggccgctgcaagcgattaccatcaatattctactc
ctgatgggtttgtttcccagttcaactactttgatcagatgccagtcactaatagaacactgaacggtat
gaaaaccccccctttttaaatatgcttttaatggtattaaccatctttcataggagagattgcaaccgt
ttatccaaataatcctagtaattcggtggcctggggaagcccattcccctttgtatccttggtggattggg
tccgttgcagaagctgttttcctaattggtgaagagaggaattcgccaaagataatcggtgctagctacg
tacggaattctacttttcgagattttaacattggataagaaggactaacctcaatacaggctccaatgtt
cagaaatatcaacaattggcagtggtctccaacactcatcgcttttgacgctgactcgtcgcgtacaagt
cgttcaacaagctggcatgtgatcaaggtatgctaattttcctcctcattcaaacccgcagatgtgagct
aactttccgaagcttctctcgacaaacaaaatcacgcaaaatttacccacgacttggagtggcggtgaca
taggtccattatactgggtagctggacgaaacgacaatacaggatcgaacatattcaaggccgctgtttta
caacagcacctcagacgtccctgtcaccgttcaatttgcaggatgcaacgcaaagagcgcaaatttgacc
atcttgtcatccgacgatccgaacgcatcgaactaccctgggggggcccgaagttgtgaagactgagatcc
agtctgtcactgcaaatgctcatggagcatttgagttcagtctcccgaacctaagtgtggctgttctcaa
aacggagtaa

图18A

SEQ ID NO:22

Pf51A的蛋白质序列

mqkmwhsilvvlqllsvqhaitinvsqsggnktsplqyglmfedinhggdgglyaelvrnrafqgstvyp
anldgydsvngailalqnltnplspsmpsslnvakgsnngsigfanegwwgievkpqryagsfyvqgdyq
gdfdislqskltqevfatakvrssgkhedwvqykyelvpkkaasntnntltitfdskglkdgslnfnlis
lfpptynnrpnglridlveamaelegkflrfpggsdvegvqapywykwnetvgdlkdrysrpsawtyees
ngiglieymnwcddmglepilavwdghylsnevisendlqpyiddtlnqleflmgapdtpygswraslgy
pkpwtinyveignednlyggletyiayrfqayydaitakyphmtvmesltempgpaaaasdyhqystpdg
fvsqfnyfdqmpvtnrtlngeiatvypnnpsnsvawgspfplyp**wwigsvaeavfligeernspkiigas**
yapmfrninnwqwsptliafdadssrtsrstswhviklllstnkitqnlpttwsggdigplywvagrndnt
gsnifkaavynstsdvpvtvqfagcnaksanltilssddpnasnypggpevvkteiqsvtanahgafefs
lpnlsvavlkte

图18B

SEQ ID NO:23

来自烟曲霉的GH11家族酶AfuXyn2的核苷酸序列

atggtttctttctcctacctgctgctggcgtgctccgccattggagctctggctgcccccgtcgaacccg
agaccacctcgttcaatgagactgctcttcatgagttcgctgagcgcgccggcaccccaagctccaccgg
ctggaacaacggctactactactccttctggactgatggcggcggcgacgtgacctacaccaatggcgcc
ggtggctcgtactccgtcaactggaggaacgtgggcaactttgtcggtggaaagggctggaaccctggaa
gcgctaggtaccgagctttgtcaacgtcggatgtgcagacctgtggctgacagaagtagaaccatcaact
acggaggcagcttcaaccccagcggcaatggctacctggctgtctacggctggaccaccaacccttgat
tgagtactacgttgttgagtcgtatggtacatacaaccccggcagcggcggtaccttcaggggcactgtc
aacaccgacggtggcacttacaacatctacacggccgttcgctacaatgctccctccatcgaaggcacca
agaccttcacccagtactggtctgtgcgcacctccaagcgtaccggcggcactgtcaccatggccaacca
cttcaacgcctggagcagactgggcatgaacctgggaactcacaactaccagattgtcgccactgagggt
taccagagcagcggatctgcttccatcactgtctactag

图19A

SEQ ID NO:24

AfuXyn2的蛋白质序列

mvsfsylllacsaigalaapvepettsfnetalhefaeragtpsst**gwnngyyysfwtdgggdvtytnga**
ggsysvnwrnvgnfvggkgwnpgsartinyggsfnpsgngylavygwttnplieyyvvesygtynpgsgg
tfrgtvntdggtyniytavrynapsiegtktftqywsvrtskrtggtvtmanhfnawsrlgmnlgthnyq
ivategyqssgsasitvy

图19B

SEQ ID NO:25

来自烟曲霉的GH11家族酶AfuXyn5的核苷酸序列

```
atgatctccatttcctcgctcagctttggactcgccgctatcgccggcgcatatgctcttccgagtgaca
aatccgtcagcttagcggaacgtcagacgatcacgaccagccagacaggcacaaacaatggctactacta
ttccttctggaccaacggtgccggatcagtgcaatatacaaatggtgctggtggcgaatatagtgtgacg
tgggcgaaccagaacggtggtgactttacctgtgggaagggctggaatccagggagtgaccagtaggcaa
cgcccgagaactatagaagaggacgcaaagaaagcactaaactctctactagtgacattaccttctctgg
cagcttcaatccttccggaaatgcttacctgtccgtgtatggatggactaccaacccctagtcgaatac
tacatcctcgagaactatggcagttacaatcctggctcgggcatgacgcacaagggcaccgtcaccagcg
atggatccacctacgacatctatgagcaccaacaggtcaaccagccttcgatcgtcggcacggccacctt
caaccaatactggtccatccgccaaaacaagcgatccagcggcacagtcaccaccgcgaatcacttcaag
gcctgggctagtctggggatgaacctgggtacccataactatcagattgtttccactgagggatatgaga
gcagcggtacctcgaccatcactgtctcgtctggtggttcttcttctggtggaagtggtggcagctcgtc
tactacttcctcaggcagctcccctactggtggctccggcagtgtaagtcttcttccatatggttgtggc
tttatgtgtattctgactgtgatagtgctctgctttgtggggccagtgcggtggaattggctggtctggt
cctacttgctgctcttcgggcacttgccaggtttcgaactcgtactactcccagtgcttgtagtaccttc
ttgcagggttatatccaagtga
```

图20A

SEQ ID NO:26

AfuXyn5的蛋白质序列

MISISSLSFGLAAIAGAYALPSDKSVSLAERQTITTSQT**GTNNGYYYSFWTNGAGSVQYTNGAGGEYSVT**
WANQNGGDFTCGKGWNPGSDHDITFSGSFNPSGNAYLSVYGWTTNPLVEYYILENYGSYNPGSGMTHKGT
VTSDGSTYDIYEHQQVNQPSIVGTATFNQYWSIRQNKRSSGTVTTANHFKAWASLGMNLGTHNYQIVSTE
GYESSGTSTITV*SSGGSSSGGSGGSSSTTSSGSSPTGGSGSCSA***LWGQCGGIGWSGPTCCSSGTCQVSNS**
YYSQCL

图20B

SEQ ID NO:27

来自轮枝镰刀菌的GH43家族酶Fv43D的核苷酸序列

```
atgcagctcaagtttctgtcttcagcattgttgctgtctttgaccggcaattgcgctgcgcaagacacta
atgatatccctcctctgatcaccgacctctggtctgcggatccctcggctcatgttttcgagggcaaact
ctgggtttacccatctcacgacatcgaagccaatgtcgtcaacggcaccggaggcgctcagtacgccatg
agagattatcacacctattccatgaagaccatctatggaaaagatcccgttatcgaccatggcgtcgctc
tgtcagtcgatgatgtcccatgggccaagcagcaaatgtgggctcctgacgcagcttacaagaacggcaa
atattatctctacttccccgccaaggataaagatgagatcttcagaattggagttgctgtctccaacaag
cccagcggtcctttcaaggccgacaagagctggatccccggtacttacagtatcgatcctgctagctatg
tcgacactaatggcgaggcatacctcatctggggcggtatctggggcggccagcttcaggcctggcagga
tcacaagacctttaatgagtcgtggctcggcgacaaagctgctcccaacggcaccaacgccctatctcct
cagatcgccaagctaagcaaggacatgcacaagatcaccgagacaccccgcgatctcgtcatcctggccc
ccgagacaggcaagccccttcaagcagaggacaataagcgacgattttcgaggggccctgggttcacaa
gcgcggcaagctgtactacctcatgtactctaccggcgacacgcacttcctcgtctacgcgacttccaag
aacatctacggtccttatacctatcagggcaagattctcgaccctgttgatgggtggactacgcatggaa
gtattgttgagtacaagggacagtggtggttgttctttgcggatgcgcatacttctggaaaggattatct
gagacaggttaaggcgaggaagatctggtatgacaaggatggcaagattttgcttactcgtcctaagatt
tag
```

图21A

SEQ ID NO:28

Fv43D的蛋白质序列

```
mqlkflssalllsltqncaaqdtndipplitdlwsadpsahvfegklwvypshdieanvvngtggaqyam
rdyhtysmktiygkdpvidhgvalsvddvpwakqqmwapdaaykngkyylyfpakdkdeifrigvavsnk
psgpfkadkswipgtysidpasyvdtngeayliwggiwggqlqawqdhktfneswlgdkaapngtnalsp
qiaklskdmhkitetprdlvilapetgkplqaednkrrffegpwvhkrgklyylmystgdthflvyatsk
niygpytyqgkildpvdgwtthgsiveykgqwwlffadahtsgkdylrqvkarkiwydkdgkilltrpki
```

图21B

SEQ ID NO:29
来自绳状青霉的GH43家族酶Pf43B的核苷酸序列

atgagtcgcagcatccttccgtacgcctctgttttcgccctcctgggcggggctatcgccgaaccgtttt
tggttctcaatagcgattttcccgatcccagtctcatagagacatccagcggatactatgcattcggtac
caccggaaacggagtcaatgcgcaggttgcttcttcaccagactttaatacctggactttgctttccggc
acagatgccctcccgggaccatttccgtcatgggtagcttcgtctccacaaatctgggcgccagatgttt
tggttaaggtatgttcttatggaataacagttttaggagtaggtcagccaggatattgacaaaattataa
taggccgatggtacctatgtcatgtacttttcggcatctgctgcgagtgactcgggcaaacactgcgttg
gtgccgcaactgcgacctcaccggaaggaccttacaccccggtcgatagcgctgttgcctgtccattaga
ccagggaggagctattgatgccaatggatttattgacaccgacggcactatatacgttgtatacaaaatt
gatggaaacagtctagacggtgatggaaccacacatcctacccccatcatgcttcaacaaatggaggcag
acggaacaaccccaaccggcagcccaatccaactcattgaccgatccgacctcgacggacctttgatcga
ggctcctagtttgctcctctccaatggaatctactacctcagtttctcttccaactactacaacactaat
tactacgacacttcatacgcctatgcctcgtcgattactggtccttggaccaaacaatctgcgccttatg
caccttgttggttactggaaccgagactagcaatgacggcgcattgagcgcccctggtggtgccgatttt
ctccgtcgatggcaccaagatgttgttccacgcaaacctcaatggacaagatatctcgggcggacgcgcc
ttatttgctgcgtcaattactgaggccagcgatgtggttacattgcagtag

图22A

SEQ ID NO:30
Pf43B的蛋白质序列

msrsilpyasvfallqqaiaepfl**vlnsdfpdpslietssgyyafgttgngvnaqvasspdfntwtllsg**
tdalpgpfpswvasspqiwapdvlvkadgtyvmyfsasaasdsgkhcvgaatatspegpytpvdsavacp
ldqggaidangfidtdgtiyvvykidgnsldgdgtthptpimlqqmeadgttptgspiqlidrsdldgpl
ieapslllsngiyylsfssnyyntnyydtsyayassitgpwtkqsapyapllvtgtetsndgalsapgga
dfsvdgtkmlfhanlngqdisggralfaasiteasdvvtlq

图22B

SEQ ID NO:31

来自轮枝镰刀菌的GH51家族酶Fv51A的核苷酸序列

atggttcgcttcagttcaatcctagcggctgcggcttgcttcgtggctgttgagtcagtcaacatcaagg
tcgacagcaagggcggaaacgctactagcggtcaccaatatggcttccttcacgaggttggtattgacac
accactggcgatgattgggatgctaacttggagctaggatatcaacaattccggtgatggtggcatctac
gctgagctcatccgcaatcgtgctttccagtacagcaagaaataccctgtttctctatctggctggagac
ccatcaacgatgctaagctctccctcaaccgtctcgacactcctctctccgacgctctccccgtttccat
gaacgtgaagcctggaaagggcaaggccaaggagattggtttcctcaacgagggttactggggaatggat
gtcaagaagcaaaagtacactggctctttctgggttaagggcgcttacaagggccactttacagcttctt
tgcgatctaaccttaccgacgatgtctttggcagcgtcaaggtcaagtccaaggccaacaagaagcagtg
ggttgagcatgagtttgtgcttactcctaacaagaatgcccctaacagcaacaacacttttgctatcacc
tacgatcccaaggtgagtaacaatcaaaactgggacgtgatgtatactgacaatttgtagggcgctgatg
gagctcttgacttcaacctcattagcttgttccctcccacctacaagggccgcaagaacggtcttcgagt
tgatcttgccgaggctctcgaaggtctccacccgtaaggtttaccgtctcacgtgtatcgtgaacagtc
gctgacttgtagaaaagagcctgctgcgcttccccggtggtaacatgctcgagggcaacaccaacaagac
ctggtgggactggaaggataccctcggacctctccgcaaccgtcctggtttcgagggtgtctggaactac
cagcagacccatggtcttggaatcttggagtacctccagtgggctgaggacatgaaccttgaaatcagta
ggttctataaaattcagtgacggttatgtgcatgctaacagatttcagttgtcggtgtctacgctggcct
ctccctcgacggctccgtcacccccaaggaccaactccagcccctcatcgacgacgcgctcgacgagatc
gaattcatccgaggtcccgtcacttcaaagtggggaaagaagcgcgctgagctcggccaccccaagcctt
tcagactctcctacgttgaagtcggaaacgaggactggctcgctggttatcccactggctggaactctta
caaggagtaccgcttccccatgttcctcgaggctatcaagaaagctcaccccgatctcaccgtcatctcc
tctggtgcttctattgaccccgttggtaagaaggatgctggtttcgatattcctgctcctggaatcggtg
actaccacccttaccgcgagcctgatgttcttgttgaggagttcaacctgtttgataacaataagtatgg
tcacatcattggtgaggttgcttctacccaccccaacggtggaactggctggagtggtaaccttatgcct
tacccctggtggatctctggtgttggcgaggccgtcgctctctgcggttatgagcgcaacgccgatcgta
ttcccggaacattctacgctcctatcctcaagaacgagaaccgttggcagtgggctatcaccatgatcca
attcgccgccgactccgccatgaccacccgctccaccagctggtatgtctggtcactcttcgcaggccac
cccatgacccatactctccccaccaccgccgacttcgacccccctctactacgtcgctggtaagaacgagg
acaagggaactcttatctggaagggtgctgcgtataacaccaccaagggtgctgacgttcccgtgtctct
gtccttcaagggtgtcaagcccggtgctcaagctgagcttactcttctgaccaacaaggagaaggatcct
tttgcgttcaatgatcctcacaagggcaacaatgttgttgatactaagaagactgttctcaaggccgatg
gaaagggtgctttcaacttcaagcttcctaacctgagcgtcgctgttcttgagaccctcaagaagggaaa
gccttactctagctag

图23A

SEQ ID NO:32

Fv51A的蛋白质序列

mvrfssilaaaacfvavesvnikvdskggnatsghqygflhedinnsgdggiyaelirnrafqyskkypv
slsgwrpindaklslnrldtplsdalpvsmnvkpgkgkakeigflnegywgmdvkkqkytgsfwvkgayk
ghftaslrsnltddvfgsvkvkskankkqwvehefvltpnknapnsnntfaitydpkgadgaldfnlisl
fpptykgrknglrvdlaealeglhpsllrfpggnmlegntnktwwdwkdtlgplrnrpgfegvwnyqqth
glgileylqwaedmnleiivgvyaglsldgsvtpkdqlqpliddaldeiefirgpvtskwgkkraelghp
kpfrlsyvevgnedwlagyptgwnsykeyrfpmfleaikkahpdltvissgasidpvgkkdagfdipapg
igdyhpyrepdvlveefnlfdnnkyghii**gevasthpnggtgwsgnlmpypwwisgvgeavalcgyerna**
dripgtfyapilknenrwqwaitmiqfaadsamttrstswyvwslfaghpmthtlpttadfdplyyvagk
nedkgtliwkgaaynttkgadvpvslsfkgvkpgaqaeltlltnkekdpfafndphkgnnvvdtkktvlk
adgkgafnfklpnlsvavletlkkgkpyss

图23B

SEQ ID NO:41

来自里氏木霉GH10木聚糖酶Xyn3的核苷酸序列

atgaaagcaaacgtcatcttgtgcctcctggccccctggtcgccgctctccccaccgaaaccatccacc
tcgaccccgagctcgccgctctccgcgccaacctcaccgagcgaacagccgacctctgggaccgccaagc
ctctcaaagcatcgaccagctcatcaagagaaaaggcaagctctactttggcaccgccaccgaccgcggc
ctcctccaacgggaaaagaacgcggccatcatccaggcagacctcggccaggtgacgccggagaacagca
tgaagtggcagtcgctcgagaacaaccaaggccagctgaactggggagacgccgactatctcgtcaactt
tgcccagcaaaacggcaagtcgatacgcggccacactctgatctggcactcgcagctgcctgcgtgggtg
aacaatatcaacaacgcggatactctgcggcaagtcatccgcacccatgtctctactgtggttgggcggt
acaagggcaagattcgtgcttgggtgagttttgaacaccacatgccccttttcttagtccgctcctcctc
ctcttggaacttctcacagttatagccgtatacaacattcgacaggaaatttaggatgacaactactgac
tgacttgtgtgtgtgatggcgataggacgtggtcaatgaaatcttcaacgaggatggaacgctgcgctct
tcagtcttttccaggctcctcggcgaggagtttgtctcgattgcctttcgtgctgctcgagatgctgacc
cttctgcccgtctttacatcaacgactacaatctcgaccgcgccaactatggcaaggtcaacgggttgaa
gacttacgtctccaagtggatctctcaaggagttcccattgacggtattggtgagccacgacccctaaat
gtcccccattagagtctctttctagagccaaggcttgaagccattcagggactgacacgagagccttctc
tacaggaagccagtcccatctcagcggcggcggaggctctggtacgctgggtgcgctccagcagctggca
acggtacccgtcaccgagctggccattaccgagctggacattcaggggcaccgacgacggattacaccc
aagttgttcaagcatgcctgagcgtctccaagtgcgtcggcatcaccgtgtggggcatcagtgacaaggt
aagttgcttcccctgtctgtgcttatcaactgtaagcagcaacaactgatgctgtctgtctttacctagg
actcgtggcgtgccagcaccaaccctcttctgtttgacgcaaacttcaaccccaagccggcatataacag
cattgttggcatcttacaatag

图24A

SEQ ID NO:42
Xyn3的蛋白质序列

mkanvilcllaplvaalptetihldpelaalranltertadlwdrqasq**sidqlikrkgklyfgtatdrg**
llqreknaaiiqadlgqvtpensmkwqslennqgqlnwgdadylvnfaqqngksirghtliwhsqlpawv
nninnadtlrqvirthvstvvgrykgkirawdvvneifnedgtlrssvfsrllgeefvsiafraardadp
sarlyindynldranygkvnglktyvskwisqgvpidgigsqshlsggggsgtlgalqqlatvpvtelai
teldiqgapttdytqvvqaclsvskcvgitvwgisdkdswrastnpllfdanfnpkpaynsivgilq

图24B

SEQ ID NO:43
来自里氏木霉的GH11家族木聚糖酶的Xyn2的蛋白质序列

mvsftsllaasppsrascrpaaevesvavekrqtiqpgt**gynngyfysywndghggvtytngpggqfsvn**
wsnsgnfvggkgwqpgtknkvinfsgsynpngnsylsvygwsrnplieyyivenfgtynpstgatklgev
tsdgsvydiyrtqrvnqpsiigtatfyqywsvrrnhrssgsvntanhfnawaqqgltlgtmdyqivaveg
yfssgsasitvs

图25A

SEQ ID NO:162
来自里氏木霉的Xyn2的核苷酸序列

ATGGTCTCCTTCACCTCCCTCCTCGCCGGCGTCGCCGCCATCTCGGGCGTCTTGGCCGCTCCCGCCGCCG
AGGTCGAATCCGTGGCTGTGGAGAAGCGCCAGACGATTCAGCCCGGCACGGGCTACAACAACGGCTACTT
CTACTCGTACTGGAACGATGGCCACGGCGGCGTGACGTACACCAATGGTCCCGGCGGGCAGTTCTCCGTC
AACTGGTCCAACTCGGGCAACTTTGTCGGCGGCAAGGGATGGCAGCCCGGGACCAAGAACAAGTAAGACT
ACCTACTCTTACCCCCTTTGACCAACACAGCACAACACAATACAACACATGTGACTACCAATCATGGAAT
CGGATCTAACAGCTGTGTTTTAAAAAAAAGGGTCATCAACTTCTCGGGAAGCTACAACCCCAACGGCAAC
AGCTACCTCTCCGTGTACGGCTGGTCCCGCAACCCCCTGATCGAGTACTACATCGTCGAGAACTTTGGCA
CCTACAACCCGTCCACGGGCGCCACCAAGCTGGGCGAGGTCACCTCCGACGGCAGCGTCTACGACATTTA
CCGCACGCAGCGCGTCAACCAGCCGTCCATCATCGGCACCGCCACCTTTTACCAGTACTGGTCCGTCCGC
CGCAACCACCGCTCGAGCGGCTCCGTCAACACGGCGAACCACTTCAACGCGTGGGCTCAGCAAGGCCTGA
CGCTCGGGACGATGGATTACCAGATTGTTGCCGTGGAGGGTTACTTTAGCTCTGGCTCTGCTTCCATCAC
CGTCAGCTAA

图25B

SEQ ID NO:44

来自里氏木霉的GH3家族β-木糖苷酶Bx11的蛋白质序列

mvnnaallaalsallptalaqnnqtyanysaqgqpdlypetlatltlsfpdcehgplknnlvcdssagyv
eraqalislftleelilntqn**sgpgvprlglpnyqvwnealhgldranfatkggqfewatsfpmpiltta**
alnrtlihqiadiistqarafsnsgrygldvyapnvngfrsplwgrgqetpgedafflssaytyeyitgi
qggvdpehlkvaatvkhfagydlenwnnqsrlgfdaiitqqdlseyytpqflaaaryaksrslmcaynsv
ngvpscansfflqtllreswgfpewgyvssdcdavynvfnphdyasnqssaaasslragtdidcgqtypw
hlnesfvagevsrgeiersvtrlyanlvrlgyfdkknqyrslgwkdvvktdawnisyeaavegivllknd
gtlplskkvrsialigpwanattqmqgnyygpapylispleaakkagyhvnfelgteiagnsttgfakai
aaakksdaiiylggidntieqegadrtdiawpgnqldlikqlsevgkplvvlqmgggqvdssslksnkkv
nslvwggypgqsggvalfdilsgkrapagrlvttqypaeyvhqfpqndmnlrpdgksnpgqtyiwytgkp
vyefgsglfyttfketlashpkslkfntssilsaphpgytyseqipvftfeaniknsgktespytamlfv
rtsnagpapypnkwlvgfdrladikpghssklsipipvsalarvdshgnrivypgkyelalntdesvkle
felvgeevtienwpleeqqikdatpda

图26A

SEQ ID NO:163

来自里氏木霉的Bx11的核苷酸序列

atggtgaataacgcagctcttctcgccgccctgtcggctctcctgcccacggccctggcgcagaacaatc
aaacatacgccaactactctgctcagggccagcctgatctctaccccgagacacttgccacgctcacact
ctcgttccccgactgcgaacatggcccctcaagaacaatctcgtctgtgactcatcggccggctatgta
gagcgagcccaggccctcatctcgctcttcaccctcgaggagctcattctcaacacgcaaaactcgggcc
ccggcgtgcctcgcctgggtcttccgaactaccaagtctggaatgaggctctgcacggcttggaccgcgc
caacttcgccaccaagggcggccagttcgaatgggcgacctcgttccccatgcccatcctcactacggcg
gccctcaaccgcacattgatccaccagattgccgacatcatctcgacccaagctcgagcattcagcaaca
gcggccgttacggtctcgacgtctatgcgccaaacgtcaatggcttccgaagccccctctggggccgtgg
ccaggagacgcccggcgaagacgccttttcctcagctccgcctatacttacgagtacatcacgggcatc
cagggtggcgtcgaccctgagcacctcaaggttgccgccacggtgaagcactttgccggatacgacctcg
agaactggaacaaccagtcccgtctcggtttcgacgccatcataactcagcaggacctctccgaatacta
cactccccagttcctcgctgcggcccgttatgcaaagtcacgcagcttgatgtgcgcatacaactccgtc
aacggcgtgcccagctgtgccaacagcttcttcctgcagacgcttttgcgcgagagctggggcttccccg
aatggggatacgtctcgtccgattgcgatgccgtctacaacgttttcaaccctcatgactacgccagcaa
ccagtcgtcagccgccgccagctcactgcgagccggcaccgatatcgactgcggtcagacttacccgtgg
cacctcaacgagtcctttgtggccggcgaagtctcccgcggcgagatcgagcggtccgtcacccgtctgt
acgccaacctcgtccgtctcggatacttcgacaagaagaaccagtaccgctcgctcggttggaaggatgt
cgtcaagactgatgcctggaacatctcgtacgaggctgctgttgagggcatcgtcctgctcaagaacgat
ggcactctccctctgtccaagaaggtgcgcagcattgctctgatcggaccatgggccaatgccacaaccc
aaatgcaaggcaactactatggccctgccccatacctcatcagccctctggaagctgctaagaaggccgg
ctatcacgtcaactttgaactcggcacagagatcgccggcaacagcaccactggctttgccaaggccatt
gctgccgccaagaagtcggatgccatcatctacctcggtggaattgacaacaccattgaacaggagggcg
ctgaccgcacggacattgcttggcccggtaatcagctggatctcatcaagcagctcagcgaggtcggcaa
acccccttgtcgtcctgcaaatgggcggtggtcaggtagactcatcctcgctcaagagcaacaagaaggtc
aactccctcgtctggggcggatatcccggccagtcgggaggcgttgccctcttcgacattctctctggca
agcgtgctcctgccggccgactggtcaccactcagtacccggctgagtatgttcaccaattcccccagaa
tgacatgaacctccgacccgatggaaagtcaaaccctggacagacttacatctggtacaccggcaaaccc
gtctacgagtttggcagtggtctcttctacaccaccttcaaggagactctcgccagccaccccaagagcc
tcaagttcaacacctcatcgatcctctctgctcctcaccccggatacacttacagcgagcagattcccgt
cttcaccttcgaggccaacatcaagaactcgggcaagacggagtccccatatacggccatgctgtttgtt
cgcacaagcaacgctggcccagccccgtacccgaacaagtggctcgtcggattcgaccgacttgccgaca
tcaagcctggtcactcttccaagctcagcatccccatccctgtcagtgctctcgcccgtgttgattctca
cggaaaccggattgtataccccggcaagtatgagctagccttgaacaccgacgagtctgtgaagcttgag
tttgagttggtgggagaagaggtaacgattgagaactggccgttggaggagcaacagatcaaggatgcta
cacctgacgcataa

图26B

SEQ ID NO:45

Bg11（得自里氏木霉的GH3家族β-葡萄糖苷酶）的蛋白序列

mryrtaaalalatgpfaradshstsgasaeavvppagtpwgtaydkakaalaklnlqdkvgivsgv**gwng**
gpcvgntspaskisypslclqdgplgvrystgstaftpgvqaastwdvnlirergqfigeevkasgihvi
lgpvagplgktpqggrnwegfgvdpyltgiamgqtingiqsvgvqatakhyilneqelnretissnpddr
tlhelytwpfadavqanvasvmcsynkvnttwacedqytlqtvlkdqlgfpgyvmtdwnaqhttvqsans
gldmsmpgtdfngnnrlwgpaltnavnsnqvptsrvddmvtrilaawyltgqdqagypsfnisrnvqgnh
ktnvraiardgivllk**ndanilplkkpasiavvgsaaiignharnspscndkgcddgalgmgwgsgavny**
pyfvapydaintrassqgtqvtlsntdntssgasaargkdvaivfitadsgegyitvegnagdrnnldpw
hngnalvqavagansnvivvvhsvgaiileqilalpqvkavvwaglpsqesgnalvdvlwgdvspsgklv
ytiakspndyntrivsggsdsfseglfidykhfddanitpryefgyglsytkfnysrlsvlstaksgpat
gavvpggpsdlfqnvatvtvdiansgqvtgaevaqlyitypssaprtppkqlrgfaklnltpgqsgtatf
nirrrdlsywdtasqkwvvpsgsfgisvgassrdirltstlsva

图27A

SEQ ID NO:46:

Pa51A（得自柄孢霉的GH51家族酶）的核苷酸序列

atgatccacctcaagccagccctcgcggcgttgttggcgctgtcgacgcaatgtgtggctattgatttgt
ttgtcaagtcttcgggggggaataagacgactgatatcatgtatggtcttatgcacgaggatatcaacaa
ctccggcgacggcggcatctacgccgagctaatctccaaccgcgcgttccaagggagtgagaagttcccc
tccaacctcgacaactggagccccgtcggtggcgctacccttacccttcagaagcttgccaagccccttt
cctctgcgttgccttactccgtcaatgttgccaaccccaaggagggcaagggcaagggcaaggacaccaa
ggggaagaaggttggcttggccaatgctgggttttggggtatggatgtcaagaggcagaagtacactggt
agcttccacgttactggtgagtacaagggtgactttgaggttagcttgcgcagcgcgattaccggggaga
cctttggcaagaaggtggtgaagggtgggagtaagaaggggaagtggaccgagaaggagtttgagttggt
gcctttcaaggatgcgcccaacagcaacaacacctttgttgtgcagtgggatgccgagggcgcaaaggac
ggatctttggatctcaacttgatcagcttgttccctccgacattcaagggaaggaagaatgggctgagaa
ttgatcttgcgcagacgatggttgagctcaagccgaccttcttgcgcttccccggtggcaacatgctcga
gggtaacaccttggacacttggtggaagtggtacgagaccattggccctctgaaggatcgcccgggcatg
gctggtgtctgggagtaccagcaaacccttggcttgggtctggtcgagtacatggagtgggccgatgaca
tgaacttggagcccattgtcggtgtcttcgctggtcttgccctcgatggctcgttcgttcccgaatccga
gatgggatgggtcatccaacaggctctcgacgaaatcgagttcctcactggcgatgctaagaccaccaaa
tggggtgccgtccgcgcgaagcttggtcaccccaagccttggaaggtcaagtgggttgagatcggtaacg
aggattggcttgccggacgccctgctggcttcgagtcgtacatcaactaccgcttccccatgatgatgaa
ggccttcaacgaaaagtaccccgacatcaagatcatcgcctcgccctccatcttcgacaacatgacaatc
cccgcgggtgctgccggtgatcaccacccgtacctgactcccgatgagttcgttgagcgattcgccaagt
tcgataacttgagcaaggataacgtgacgctcatcggcgaggctgcgtcgacgcatcctaacggtggtat
cgcttgggagggagatctcatgcccttgccttggtggggcggcagtgttgctgaggctatcttcttgatc
agcactgagagaaacggtgacaagatcatcggtgctacttacgcgcctggtcttcgcagcttggaccgct
ggcaatggagcatgacctgggtgcagcatgccgccgacccggccctcaccactcgctcgaccagttggta
tgtctggagaatcctcgcccaccacatcatccgtgagacgctcccggtcgatgccccggccggcaagccc
aactttgaccctctgttctacgttgccggaaagagcgagagtggcaccggtatcttcaaggctgccgtct
acaactcgactgaatcgatcccggtgtcgttgaagtttgatggtctcaacgagggagcggttgccaactt
gacggtgcttactgggccggaggatccgtatggatacaacgacccccttcactggtatcaatgttgtcaag
gagaagaccaccttcatcaaggccggaaagggcggcaagttcaccttcaccctgccgggcttgagtgttg
ctgtgttggagacggccgacgcggtcaagggtggcaagggaaagggcaagggcaagggaaagggtaactga

图27B

SEQ ID NO:47

来自鹅掌柄孢霉的GH51家族酶Pa51A的密码子优化cDNA

```
atgatccacctcaagcccgccctcgccgccctcctcgccctcagcacccaatgcgtcgccatcgacctct
tcgtcaagagcagcggcggcaacaagaccaccgacatcatgtacggcctcatgcacgaggacatcaacaa
cagcggcgacggcggcatctacgccgagctgatcagcaaccgcgccttccagggcagcgagaagttcccc
agcaacctcgacaactggtcccccgtcggcggcgccaccctcaccctccagaagctcgccaagcccctgt
cctctgccctccctactccgtcaacgtcgccaaccccaaggagggtaagggtaagggcaaggacaccaa
gggcaagaaggtcggcctcgccaacgccggcttttggggcatggacgtcaagcgccagaaatacaccggc
agcttccacgtcaccggcgagtacaagggcgacttcgaggtcagcctccgcagcgccattaccggcgaga
ccttcggcaagaaggtcgtcaagggcggcagcaagaagggcaagtggaccgagaaggagttcgagctggt
cccccttcaaggacgcccccaacagcaacaacaccttcgtcgtccagtgggacgccgagggcgccaaggac
ggcagcctcgacctcaacctcatcagcctcttcccgcccaccttcaagggccgcaagaacggcctccgca
tcgacctcgcccagaccatggtcgagctgaagcccaccttcctccgctttcccggcggcaacatgctcga
gggcaacaccctcgacacctggtggaagtggtacgagaccatcggcccccctgaaggaccgccctggcatg
gccggcgtctgggagtaccagcagacgctgggcctcggcctggtcgagtacatggagtgggccgacgaca
tgaacctcgagcccatcgtcggcgtctttgctggcctggccctggatggcagctttgtccccgagagcga
gatgggctgggtcatccagcaggctctcgatgagatcgagttcctcaccggcgacgccaagaccaccaag
tggggcgccgtccgcgccaagctcggccaccctaagccctggaaggtcaaatgggtcgagatcggcaacg
aggactggctcgccggccgacctgccggcttcgagagctacatcaactaccgcttccccatgatgatgaa
ggccttcaacgagaaataccccgacatcaagatcattgccagcccctccatcttcgacaacatgaccatt
ccagccggtgctgccggtgaccaccaccctacctcacccccgacgaatttgtcgagcgcttcgccaagt
tcgacaacctcagcaaggacaacgtcaccctcattggcgaggccgccagcacccaccccaacggcggcat
tgcctgggagggcgacctcatgcccctgccctggtggggcggcagcgtcgccgaggccatcttcctcatc
agcaccgagcgcaacggcgacaagatcatcggcgccacctacgcccctggcctccgatctctcgaccgct
ggcagtggagcatgacctgggtccagcacgccgccgaccctgccctcaccacccgcagcaccagctggta
cgtctggcgcatcctcgcccaccacatcattcgcgagaccctccccgtcgacgcccccgccggcaagccc
aacttcgacccccctcttctacgtcgctggcaagtcggagagcggcaccggcatcttcaaggccgccgtct
acaacagcaccgagagcatccccgtcagcctcaagttcgacggcctcaacgagggcgccgtcgccaacct
caccgtcctcaccggccccgaggacccctacggctacaacgacccccttcaccggcatcaacgtcgtcaag
gaaaagaccaccttcatcaaggccggcaagggcggcaagttcacctttaccctccccggcctctctgtcg
ccgtcctcgagaccgccgacgccgtgaagggtggcaagggaaagggaaagggcaagggtaagggtaacta
a
```

图27C

SEQ ID NO:48

来自玉米赤霉的GH43家族酶Gz43A的核苷酸序列

atgtatcggaagttggccgtcatctcggccttcttggccacagctcgtgctaccaacgacgactgtcctc
tcatcactagtagatggactgcggatccttcggctcatgtctttaacgacaccttgtggctctacccgtc
tcatgacatcgatgctggatttgagaatgatcctgatggaggccagtacgccatgagagattaccatgtc
tactctatcgacaagatctacggttccctgccggtcgatcacggtacggccctgtcagtggaggatgtcc
cctgggcctctcgacagatgtgggctcctgacgctgcccacaagaacggcaaatactacctatacttccc
tgccaaagacaaggatgatatcttcagaatcggcgttgctgtctcaccaacccccggcggaccattcgtc
cccgacaagagttggatccctcacactttcagcatcgaccccgccagtttcgtcgatgatgatgacagag
cctacttggcatggggtggtatcatgggtggccagcttcaacgatggcaggataagaacaagtacaacga
atctggcactgagccaggaaacggcaccgctgccttgagccctcagattgccaagctgagcaaggacatg
cacactctggcagagaagcctcgcgacatgctcattcttgaccccaagactggcaagccgctcctttctg
aggatgaagaccgacgcttcttcgaaggaccctggattcacaagcgcaacaagatttactacctcaccta
ctctactggcacaacccactatcttgtctatgcgacttcaaagacccccctatggtccttacacctaccag
ggcagaattctggagccagttgatggctggactactcactctagtatcgtcaagtaccagggtcagtggt
ggctattttatcacgatgccaagacatctggcaaggactatcttcgccaggtaaaggctaagaagatttg
gtacgatagcaaaggaaagatcttgacaaagaagccttga

图27D

SEQ ID NO:49

来自尖孢镰刀菌的GH43家族酶Fo43A的核苷酸序列

atgtatcggaagttggccgtcatctcggccttcttggccacagctcgtgctcaagacactaatgacattc
ctcccctgatcaccgacctctggtccgcagatccctcggctcatgttttcgaaggcaagctctgggttta
cccatctcacgacatcgaagccaatgttgtcaacggcacaggaggcgctcaatacgccatgagggattac
catacctactccatgaagagcatctatggtaaagatcccgttgtcgaccacggcgtcgctctctcagtcg
atgacgttccctgggcgaagcagcaaatgtgggctcctgacgcagctcataagaacggcaaatattatct
gtacttccccgccaaggacaaggatgagatcttcagaattggagttgctgtctccaacaagcccagcggt
cctttcaaggccgacaagagctggatccctggcacgtacagtatcgatcctgctagctacgtcgacactg
ataacgaggcctacctcatctggggcggtatctggggcggccagctccaagcctggcaggataaaaagaa
ctttaacgagtcgtggattggagacaaggctgctcctaacggcaccaatgccctatctcctcagatcgcc
aagctaagcaaggacatgcacaagatcaccgaaacaccccgcgatctcgtcattctcgcccccgagacag
gcaagcctcttcaggctgaggacaacaagcgacgattcttcgagggcccttggatccacaagcgcggcaa
gctttactacctcatgtactccaccggtgatacccacttccttgtctacgctacttccaagaacatctac
ggtccttatacctaccggggcaagattcttgatcctgttgatgggtggactactcatggaagtattgttg
agtataagggacagtggtggcttttctttgctgatgcgcatacgtctggtaaggattaccttcgacaggt
gaaggcgaggaagatctggtatgacaagaacggcaagatcttgcttcaccgtccttag

图27E

SEQ ID NO:50

来自绳状青霉的GH51家族酶Pf51A的核苷酸序列

atgtaccggaagctcgccgtgatcagcgccttcctggcgactgctcgcgccatcaccatcaacgtcagcc
agagcggcggcaacaagaccagcccgctccagtacggcctcatgttcgaggacatcaaccacggcggcga
cggcggcctctacgccgagctggtccggaaccgggccttccagggcagcaccgtctacccggccaacctc
gacggctacgactcggtgaacggcgcgattctcgcgctccagaacctcaccaacccgctcagcccgagca
tgccctcgtcgctgaacgtcgccaagggctcgaacaacggcagcatcggcttcgccaacgaggggtggtg
gggcatcgaggtcaagccgcagcggtacgccggcagcttctacgtccagggcgactaccagggcgacttc
gacatcagcctccagagcaagctcacccaggaggtcttcgcgacggcgaaggtccggtcgagcggcaagc
acgaggactgggtccagtacaagtacgagctggtcccgaagaaggccgccagcaacaccaacaacaccct
caccatcaccttcgacagcaagggcctcaaggacggcagcctcaacttcaacctcatcagcctcttcccg
ccgacctacaacaaccggccgaacggcctccggatcgacctcgtcgaggccatggcggagctggagggca
agttcctccgcttccccggcggctcggacgtggagggcgtccaggccccgtactggtacaagtggaacga
gaccgtcggcgacctcaaggaccgctactcgcgcccgagcgcctggacctacgaggagagcaacggcatc
ggcctcatcgagtacatgaactggtgcgacgacatgggcctcgagccgatcctcgccgtctgggacggcc
actacctcagcaacgaggtcatcagcgagaacgacctccagccgtacatcgacgacaccctcaaccagct
cgagttcctcatgggcgccccggacactccctacgggtcttggagggctagcctcggctacccgaagccg
tggaccatcaactacgtcgagatcggcaacgaggacaacctctacggcggcctcgagacctacatcgcct
accggttccaggcctactacgacgccatcaccgccaagtacccgcacatgaccgtcatggagagcctcac
cgagatgcccggccccgctgccgcggcgtcggactaccaccagtactcgacgcccgacggcttcgtcagc
cagttcaactacttcgaccagatgccggtcaccaaccgcacgctgaacggcgagatcgccaccgtctacc
ccaacaacccgagcaactcggtggcgtggggcagcccgttcccgctctacccgtggtggatcgggtccgt
ggctgaggccgtcttcctcatcggcgaggagcggaacagcccgaagatcatcggcgccagctacgccccc
atgttccgcaacattaacaactggcagtggagcccgaccctgatcgccttcgacgccgacagcagccgga
cgtcgcgctctacttcctggcacgtcatcaagctcctcagcaccaacaagatcacccagaacctgcccac
gacgtggtctgggggggacatcggcccgctctactgggtcgccggccggaacgacaacaccggcagcaac
atcttcaaggccgccgtctacaacagcaccagcgacgtcccggtcaccgtccagttcgccggctgcaacg
ccaagagcgccaacctcaccatcctctcgtcggacgaccccaacgccagcaactacccgggcggccccga
ggtcgtcaagaccgagatccagagcgtcaccgccaacgcccacggcgccttcgagttcagcctcccgaac
ctgtcggtggctgtgctgaagacggagtag

图27F

SEQ ID NO:51
来自里氏木霉的内切葡聚糖酶Eg4的核酸序列

```
atgatccagaagctttccaaccttcttctcaccgcactagcggtggcaaccggtgttgttggacacggac
acatcaacaacattgtcgtcaacggagtgtactaccagggatatgatcctacatcgttcccatatgaatc
tgacccgcccatagtggtgggctggacggctgccgatcttgacaacggcttcgtctcacccgacgcatat
cagagcccggacatcatctgccacaagaatgccaccaacgccaaaggacacgcgtccgtcaaggccggag
acactattcccctccagtgggtgccagttccttggccgcacccaggccccatcgtcgactacctggccaa
ctgcaacggcgactgcgagaccgtggacaagacgtcccttgagttcttcaagattgacggcgtcggtctc
atcagcggcggagatccgggcaactgggcctcggacgtgttgattgccaacaacaacacctgggttgtca
agatccccgaggatctcgccccgggcaactacgtgcttcgccacgagatcatcgccttgcacagcgccgg
gcaggcggacggcgctcagaactaccctcagtgcttcaacctcgccgtcccaggctccggatctctgcag
ccgagcggcgtcaagggaaccgcgctctaccactccgatgaccccggtgtcctcatcaacatctacacca
gccctcttgcgtacaccattcctggaccttccgtggtatcaggcctccccacgagtgtcgcccagggcag
ctccgccgcgacggccactgccagcgccactgttcctggcggtagcggaccgggaaacccgaccagtaag
actacgacgacggcgaggacgacacaggcctcctctagcagggccagctctactcctcctgctactacgt
cggcacctggtggaggcccaacccagactttgtacggccagtgtggtggcagcggctacagtggtcctac
tcgatgcgcgccgccggccacttgctctaccttgaacccatactacgcccagtgccttaactag
```

图28A

SEQ ID NO:52
来自里氏木霉的内切葡聚糖酶Eg4的蛋白质序列

MIQKLSNLLVTALAVATGVVG**hghindivingvwyqaydpttfpyesnppivvgwtaadldngfvspday**
qnpdiichknatnakghasvkagdtilfqwvpvpwphpgpivdylancngdcetvdkttleffkidgvgl
lsggdpgtwasdvlisnnntwvvkipdnlapgnyvlrheiialhsagqangaqnypqcfniavsgsgslq
psgvlgtdlyhatdpgvliniytsplnyiipgptvvsglptsvaq*gssaatatasatvpgggsgptsrtt*
*ttarttqassrpsstppattsapagg***ptqtlygqcggsgysgptrcappatcstlnpyyaqcl**n

图28B

SEQ ID NO:53

来自鹅掌柄孢霉的GH3家族β-葡糖苷酶Pa3D的核酸序列

atggctcttcaaaccttcttcctgctggcggcagccatgctggccaacgcagagacaacaggcgaaaagg
tctctcggcaagcaccgtctggcgctcaagcatgggccgccgcccactcccaggctgccgccactctggc
cagaatgtcacagcaagacaagatcaacatggtcacgggcattggctgggacagagggccttgcgtggga
aacacagctgccatcagctccatcaactatcctcaaatctgtcttcaggatggaccattgggcattcgct
tcggcactggtaccaccgccttcacacctggcgtccaagctgcttcgacatgggacgttgatctgatccg
gcagcgcggtgcttacctgggcgccgaagccaagggctgcggcattcacatcctttggggcccgttgcc
ggtgccctgggcaagattccccacggcggtcgcaactgggagggatttggcgccgacccctaccttgccg
gtattgccatgaaggagaccatcgagggtattcagtcagcaggcgtccaggccaacgccaagcactacat
tgcaaacgaacaagagctcaaccgcgagaccatgagcagcaatgtggatgaccgcactcagcacgagctc
tacctctggccctttgccgacgccgtgcacgccaacgtcgccagcgtcatgtgcagttacaacaagctca
atggcacgtgggcttgcgagaatgacaaggctctgaatcagatcttgaagaaggagctcggattccaggg
ctacgttctcagcgactggaatgctcagcacagcactgctctgtctgctaacagtggtctggacatgact
atgcccggtaccgatttcaacggccgcaatgtctactggggccctcaactgaacaacgctgtcaacgccg
gccaggttcagagatccagactagacgacatgtgcaagagaatcttggctggctggtacttgctcggtca
gaaccagggctatcccgccatcaacatcagggccaacgttcagggcaaccataaggagaacgtacgtgct
gttgccagagacggcatcgtcttgctgaagaacgatggaattctgccgctttccaagccgagaaagattg
ctgtcgtgggctcccactccgtcaacaatccccagggaatcaacgcctgtgttgacaagggctgcaatgt
tggcacccttggcatgggctggggttcaggcagcgtcaactacccctatctcgtgtccccgtacgatgct
ctccggactcgtgctcaggccgatggcacacaaatcagcctccacaacactgacagcaccaacggtgtgt
caaacgttgtgtctgacgctgatgctgttgttgttgtcatcactgccgattctggtgaagggtacatcac
tgtcgagggccacgctggcgaccgcagccaccttgacccgtggcacaatggcaaccaacttgttcaggct
gccgcggctgccaacaagaacgtcatcgttgttgtgcacagtgttggccagatcaccctggagactatcc
tcaacaccaatggagtccgcgcgattgtgtgggctggtcttccgggccaagagaatggcaacgctcttgt
tgatgttctctacggcttggtttcgccatctggaaagcttccctacaccattggcaagagggagtcggac
tatggcacagccgttgttcgtggggatgataacttcagggagggccttttttgttgactaccgtcactttg
acaatgccaggatcgagccgcgctatgagtttggctttggtctttgtaagttccagcggcggagttgggt
ttgatttcaagctttcctaacctgataaaacagcttacaccaatttcaccttctccgacatcaagattac
ttccaatgtcaagccggggcccgctactggccagaccattcccggcggacctgccgacctgtgggaggac
gttgcgacagtcactgcaaccatcaccaactcgggtgctgtcgagggcgctgaggttgcccagctttaca
tcggcctgccgtcctcggctcctgcctctcccccgaagcagctgcgtggattttccaagctgaagctggc
cccgggtgccagcggcactgccacattcaacctcagacgcagagatctcagctattgggatacccgcctc
cagaactgggtcgtgcccagcggcaactttgtcgtcagcgtcggcgccagctcgagagatatccgcttga
cgggcaccatcacggcgtag

图29A

SEQ ID NO:54

来自鹅掌柄孢霉的GH3家族β-葡糖苷酶Pa3D的蛋白质序列

malqtfflaaamlanaettgekvsrqapsgaqawaaahsqaaatlarmsqqdkinmvtgigwdrgpcvg
ntaaissinypqiclqdgplgirfgtgttaftpgvqaastwdvdlirqrgaylgaeakgcgihillgpva
galgkiphggrnwegfgadpylagiamketiegiqsagvqanakhyianeqelnretmssnvddrtqhel
ylwpfadavhanvasvmcsynklngtwacendkalnqilkkelgfqgyvlsdwnaqhstalsansgldmt
mp*gtdfngrnvywgpqlnnavnaggvqrsrlddmckrilagwyllgqnqgypainiranvqgnhkenvra*
*vardg***ivllkndgilpl**skprkiavvgshsvnnpqginacvdkgcnvgtlgmgwgsgsvnypylvspyda
lrtraqadgtqislhntdstngvsnvvsdadavvvvitadsgegyitveghagdrshldpwhngnqlvqa
aaaanknvivvvhsvgqitletilntngvraivwaglpgqengnalvdvlyglvspsgklpytigkresd
ygtavvrgddnfreglfvdyrhfdnariepryefgfglsytnftfsdikitsnvkpgpatgqtipggpad
lwedvatvtatitnsgavegaevaqlyiglpssapasppkqlrgfsklklapgasgtatfnlrrrdlsyw
dtrlqnwvvpsgnfvvsvgassrdirltgtita

图29B

SEQ ID NO:55

来自轮枝镰刀菌的GH3家族β-葡糖苷酶Fv3G的核苷酸序列

atgtttccttcttccatatcttgtttggcggccctgagtctgatgagccagggtctactagctcagagcc
aaccggaaaatgtcatcaccgatgatacctacttctacggtcaatcgccaccagtgtatcctacacgtaa
gcactctctctgatttcccaacgaaagcaatactgatctcttgaccagcggaacaggtagacaccggctc
atgggctgccgctgtagccaaagccaagaacttggtgtcccagttgactcttgaagagaaagtcaacttg
actacaggaggccagacgaccaccggctgctctggcttcatccctggcattccccgtgtaggctttccag
gactgtgtttagcagacgctggcaacggtgtccgcaacacagattatgtgagctcgtttccctccgggat
tcatgtcggtgcaagctggaatccggagttgacctacagccggagctactacatgggtgctgaggccaaa
gccaagggcgttaacatccttctcggtccagtatttggacctttgggccgagtagttgaaggtggacgca
actgggaggggttttccaatgatccctacctggcgggtaaattagggcatgaagctgtcgccggtatcca
agacgccggagttgttgcatgcggaaaacatttccttgctcaagagcaggagacccatagacttgcggcg
tctgtcactggggctgatgcaatctcatcaaatctcgatgacaagacactccatgaattatatctctggt
aagcacatcatatcttggctgagtagatgaaccttactaacacccgaactgggcttttcgctgatgcagt
ccacgccggacttgccagtgtgatgtgcagctacaacagagcaaacaattcacacgcctgccaaaactcg
aagcttctcaatggccttctcaagggcgagttaggattccagggttttgtcgtctcggactggggcgcac
agcaatctggtatggcttcagcattggctggcctggatgttgtcatgcccagctcgatcttgtggggtgc
caaccttacccttggtgtgaacaacggaactattcccgagtcacaggttgacaatatggttacacggtac
gcgaagtctcagccttacttctcaattcttttgaactgacaatcgtgtaggctccttgcaacttggtatc
agttgaaccaggaccaagacaccgaagccccaggtcacggactcgctgccaagctttgggagcctcaccc
agtagtcgacgctcgcaacgcaagctccaagcctactatctgggacggtgcagtcgagggccatgttctt
gttaagaacaccaacaacgcactgccattcaagcccaacatgaaactcgtttctttgttcggatactctc
acaaagctcctgataagaacatcccagaccccgcccaaggcatgttctccgcttggtctatcggtgccca
atccgccaacatcactgagctgaacctcggctttctcggaaatttgagtctcacatactccgccatcgcg
cccaacggaaccatcatctcgggtggaggctcgggtgccagcgcttggactctgttcagctcacccttcg
atgcattcgtttctcgggcgaagaaagagggtactgcgcttttctgggattttgagagctgggatcctta
tgtgaaccctacatctgaagcttgcatcgttgctggtaatgcatgggctagcgaaggctgggatagacct
gcaacctatgatgcctatactgatgagctcatcaataacgtcgctgacaagtgcgctaacactattgttg
ttcttcacaatgctggaacacgacttgtggatggcttctttggtcaccccaacgtcaccgctattatcta
cgctcatctcccaggtcaggatagtggagatgctctggtatctttgctctatggcgatgagaacccatct
ggtcgcctcccttacaccgttgcccgcaacgagacggattatggtcacctgctgaagccagacttgactc
tcgcccccaaccagtaccaacactttccccagtccgacttctccgagggtattttcattgactaccgaca
tttcgatgctaagaacatcacgcctcgcttcgagtttggtttcggcttgagctacacaacctttgagtac
gctagtctccagatctcaaagtcccaggcccagacaccggaatacccagctggtgctcttaccgagggag
gccgttcagatttgtgggacgtcgttgctactgtcacagcaagcgtcaggaacactgggtctgtcgacgg
caaggaggttgcacagctatacgttggtgttccaggtggtcctatgagacagctacgtggctttacgaaa
ccagctattaaggctggagagacggctacagtgacctttgagcttactcgccgcgacttgagtgtctggg
atgttaatgcgcaggagtggcaacttcagcaaggcaactatgctatctacgttggccgaagtagtcgaga
tttgcctctgcaaagtaccttgagcatctag

图30A

SEQ ID NO:56

来自轮枝镰刀菌的GH3家族β-葡糖苷酶Fv3G的蛋白质序列

mfpssisclaalslmsqqllaqsqpenvitddtyfygqsppvypthtgswaaavakaknlvsqltleekv
nlttggqtttgcsgf**ipgiprvgfpglcladagngvrntdyvssfpsgihvgaswnpeltysrsyymgae**
akakgvnillgpvfgplgrvveggrnwegfsndpylagklgheavagiqdagvvacgkhflaqeqethrl
aasvtgadaissnlddktlhelylcvmcsynrannshacqnskllngllkgelgfqgfvvsdwgaqqsgm
asalagldvvmp*ssilwganltlgvnngtipesqvdnmvtrllatwyqlnqdqdteapghglaaklweph*
pvvdarnasskptiwdgaveghvlvkntnnalpfkpnmklvslfgyshkapdknipdpaqgmfsawsiga
qsanitelnlgflgnlsltysaiapngtiisgggsgasawtlfsspfdafvsrakkegtalfwdfeswdp
yvnptseacivagnawasegwdrpatydaytdelinnvadkcantivvlhnagtrlvdgffghpnvtaii
yahlpgqdsgdalvsllygdenpsgrlpytvarnetdyghllkpdltlapnqyqhfpqsdfsegifidyr
hfdaknitprfefgfglsyttfeyaslqisksqaqtpeypagalteggrsdlwdvvatvtasvrntgsvd
gkevaqlyvgvpggpmrqlrgftkpaikagetatvtfeltrrdlsvwdvnaqewqlqqgnyaiyvgrssr
dlplqstlsi

图30B

SEQ ID NO:57

来自轮枝镰刀菌的GH3家族β-葡糖苷酶Fv3D的核苷酸序列

atggctagcattcgatctgtgttggtctcgggtcttttggccgcgggtgtcaatgcccaagcctacgatg
cgagtgatcgcgctgaagatgctttcagctgggtccagcccaagaacaccactattcttggacagtacgg
ccattcgcctcattaccctgccagtatgttcaccaactacaccaagtgacactgaggctgtactgacatt
ctagacaatgctactggcaagggctgggaagatgccttcgccaaggctcaaaactttgtctcccaactaa
ccctcgaggaaaaggccgacatggtcacaggaactccaggtccttgcgtcggcaacatcgtcgccattcc
ccgtctcaacttcaacggtctctgtcttcacgacggcccctcgccatccgagtagcagactacgccagt
gttttccccgctggtgtatcagccgcttcatcgtgggacaaggacctcctctaccagcgcggtctcgcca
tgggtcaagagttcaaggccaagggtgctcacatcctcctcggccccgtcgccggtcctcttggccgctc
ggcatactctggtcgtaactgggagggtttctcgccggacccttacctcactggtattgcgatggaggag
actatcatgggacatcaagatgctggtgttcaggctactgcgaagcactttatcggtaatgagcaggagg
tcatgcgaaaccctacttttgtcaaggatgggtatattggtgaggttgacaaggaggctctttcgtctaa
catggatgatcgaaccatgcacgagctttacctctggccctttgccaatgctgttcatgccaaggcttcc
agcatgatgtgctcgtaccagcgtctcaacggctcctacgcctgccagaactcaaaggtcctcaacggaa
ttctgcgtgatgagcttggtttccagggctacgtcatgtcagattggggtgccacccacgccggtgttgc
tgccatcaacagcggtctcgacatggacatgcccggtggtatcggtgcctacggaacatactttaccaag
tccttcttcggcggcaacctcacccgcgccgtcaccaacggcaccctcgacgagacccgcgtcaacgaca
tgatcacccgcatcatgactccctacttctggctcggccaggacaaggactatccctccgtcgacccctc
cagcggtgatctcaacaccttcagccccaagagctcctggttccgcgagttcaacctcaccggcgagcgc
agccgtgacgtccgcggtaaccacggcgacttgatccgcaagcacggcgccgagtctaccgtccttctca
agaacgagaagaacgcccttcccctcaagaagcccaagtccatcgctgtctttggcaacgatgctggtga
tatcactgagggtttctacaaccagaatgactacgaatttggcactcttgttgctggtggtggctctgga
actggtcgtttgacataccttgtttcgcctctagccgccatcaatgctcgtgctaagcaggacggtactc
ttgttcagcagtggatgaacaacactcttattgctaccaccaacgtcactgatctctggatccctgctac
tcccgatgtctgcctcgttttcttgaagacttgggctgaggaggctgctgatcgtgagcacctctccgtt
gactgggacggtaatgatgttgttgagtctgttgccaagtactgcaataacactgtcgtcgtcactcact
cttctggtatcaacactcttccttgggctgaccaccccaacgtcaccgctattctcgctgcccacttccc
cggtcaggagtctggcaactccctcgttgacctcctctacggcgatgtcaacccctctggtcgtcttccc
tacaccatcgccttcaacggcaccgactacaacgctccccccaccactgccgtcaacaccaccggcaagg
aggactggcagtcttggttcgacgagaagctcgagattgactaccgctacttcgacgcgcacaacatctc
cgtccgctacgaattcggcttcggtctctcctactccaccttcgaaatctccgacatctccgctgagcca
ctcgcatccgacattacctcccagcccgaggatctccccgtgcagcccggcggcaaccccgccctctggg
agaccgtctacaacgtgaccgtctccgtctccaacacgggcaaggtcgacggcgccactgtcccccagct
atacgtgacattccccgacagcgcgcctgccggtacaccacccaagcagctccgtgggttcgacaaggtc
ttccttgaggctggcgagagcaagagtgtcagctttgagctgatgcgccgtgatctgagctactgggata
tcatttctcagaagtggctcatccctgagggagagtttactattcgtgttggattcagcagtcgggactt
gaaggaggagacaaaggttactgttgttgaggcgtaa

图31A

SEQ ID NO:58

来自轮枝镰刀菌的GH3家族β-葡糖苷酶Fv3D的蛋白质序列

masirsvlvsqllaaqvnaqaydasdraedafswvqpknttilgqyghsphypannatgkgwedafaka
qnfvsqltleekadmvtgtpgpcvgn**ivaiprlnfnglclhdgplairvadyasvfpagvsaasswdkd**
llyqrglamgqefkakgahillgpvagplgrsaysgrnwegfspdpyltgiameetimghqdagvqata
khfigneqevmrnptfvkdgyigevdkealssnmddrtmhelylwpfanavhakassmmcsyqrlngsy
acqnskvlngilrdelgfqgyvmsdwgathagvaainsgldmdmp*ggigaygtyftksffggnltravt*
ngtldetrvndmitrimtpyfwlgqdkdypsvdpssgdlntfspksswfrefnltgersrdvrgnhgdl
*irkhgaes***tvllkneknalplkkpksiavfgndagditegfynqndyefgtlvagggsgtgrltylvsp**
laainarakqdgtlvqqwmnntliattnvtdlwipatpdvclvflktwaeeaadrehlsvdwdgndvve
svakycnntvvvthssgintlpwadhpnvtailaahfpgqesgnslvdllygdvnpsgrlpytiafngt
dynappttavnttgkedwqswfdekleidyryfdahnisvryefgfglsystfeisdisaeplasdits
qpedlpvqpggnpalwetvynvtvsvsntgkvdgatvpqlyvtfpdsapagtppkqlrgfdkvfleage
sksvsfelmrrdlsywdiisqkwlipegeftirvgfssrdlkeetkvtvvea

图31B

SEQ ID NO:59

来自轮枝镰刀菌的GH3家族β-葡糖苷酶Fv3C的蛋白质序列

atgaagctgaattgggtcgccgcagccctgtctataggtgctgctggcactgacagcgcagttgctcttg
cttctgcagttccagacactttggctggtgtaaaggtcagtttttttttcaccatttcctcgtctaatctc
agccttgttgccatatcgcccttgttcgctcggacgccacgcaccagatcgcgatcatttcctcccttgc
agccttggttcctcttacgatcttccctccgcaattatcagcgcccttagtctacacaaaaaccccgag
acagtctttcattgagtttgtcgacatcaagttgcttctcaactgtgcatttgcgtggctgtctacttct
gcctctagacaaccaaatctgggcgcaattgaccgctcaaaccttgttcaaataacctttttttattcgag
acgcacatttataaatatgcgccttcaataataccgactttatgcgcggcggctgctgtggcggttgat
cagaaagctgacgctcaaaaggttgtcacgagagatacactcgcatactcgccgcctcattatccttcac
catggatggaccctaatgctgttggctgggaggaagcttacgccaaagccaagagctttgtgtcccaact
cactctcatggaaaaggtcaacttgaccactggtgttgggtaagcagctccttgcaaacagggtatctca
atcccctcagctaacaacttctcagatggcaaggcgaacgctgtgtaggaaacgtgggatcaattcctcg
tctcggtatgcgaggtctctgtctccaggatggtcctcttggaattcgtctgtccgactacaacagcgct
tttcccgctggcaccacagctggtgcttcttggagcaagtctctctggtatgagagaggtctcctgatgg
gcactgagttcaaggagaagggtatcgatatcgctcttggtcctgctactggacctcttggtcgcactgc
tgctggtggacgaaactgggaaggcttcaccgttgatccttatatggctggccacgccatggccgaggcc
gtcaagggtattcaagacgcaggtgtcattgcttgtgctaagcattacatcgcaaacgagcagggtaagc
cacttggacgatttgaggaattgacagagaactgaccctcttgtagagcacttccgacagagtggcgagg
tccagtcccgcaagtacaacatctccgagtctctctcctccaacctggatgacaagactatgcacgagct
ctacgcctggcccttcgctgacgccgtccgcgccggcgtcggttccgtcatgtgctcgtacaaccagatc
aacaactcgtacggttgccagaactccaagctcctcaacggtatcctcaaggacgagatgggcttccagg
gtttcgtcatgagcgattgggcggcccagcataccggtgccgcttctgccgtcgctggtctcgatatgag
catgcctggtgacactgccttcgacagcggatacagcttctggggcggaaacttgactctggctgtcatc
aacggaactgttcccgcctggcgagttgatgacatggctctgcgaatcatgtctgccttcttcaaggttg
gaaagacgatagaggatcttcccgacatcaacttctcctcctggacccgcgacaccttcggcttcgtgca
tacatttgctcaagagaaccgcgagcaggtcaactttggagtcaacgtccagcacgaccacaagagccac
atccgtgaggccgctgccaagggaagcgtcgtgctcaagaacaccgggtcccttcccctcaagaacccaa
agttcctcgctgtcattggtgaggacgccggtcccaaccctgctggacccaatggttgtggtgaccgtgg
ttgcgataatggtaccctggctatggcttggggctcgggaacttcccaattcccttacttgatcacccc
gatcaagggctctctaatcgagctactcaagacggaactcgatatgagagcatcttgaccaacaacgaat
gggcttcagtacaagctcttgtcagccagcctaacgtgaccgctatcgttttcgccaatgccgactctgg
tgagggatacattgaagtcgacggaaactttggtgatcgcaagaacctcaccctctggcagcagggagac
gagctcatcaagaacgtgtcgtccatatgccccaacaccattgtagttctgcacaccgtcggccctgtcc
tactcgccgactacgagaagaaccccaacatcactgccatcgtctgggctggtcttcccggccaagagtc
aggcaatgccatcgctgatctcctctacggcaaggtcagccctggccgatctcccttcacttggggccgc
acccgcgagagctacggtactgaggttctttatgaggcgaacaacggccgtggcgctcctcaggatgact
tctctgagggtgtcttcatcgactaccgtcacttcgaccgacgatctccaagcaccgatggaaagagctc
tcccaacaacaccgctgctcctctctacgagttcggtcacggtctatcttggtccacctttgagtactct
gacctcaacatccagaagaacgtcgagaaccctactctcctcccgctggccagaccatccccgccccaa
cctttggcaacttcagcaagaacctcaacgactacgtgttccccaagggcgtccgatacatctacaagtt
catctacccttcctcaacacctcctcatccgccagcgaggcatccaacgatggtggccagtttggtaag
actgccgaagagttcctccctcccaacgccctcaacggctcagcccagcctcgtcttcccgcctctggtg
ccccaggtggtaaccctcaattgtgggacatcttgtacaccgtcacagccacaatcaccaacacaggcaa
cgccacctccgacgagattccccagctgtatgtcagcctcggtggcgagaacgagcccatccgtgttctc
cgcggtttcgaccgtatcgagaacattgctcccggccagagcgccatcttcaacgctcaattgacccgtc
gcgatctgagtaactgggatacaaatgcccagaactgggtcatcactgaccatcccaagactgtctgggt
tggaagcagctctcgcaagctgcctctcagcgccaagttggagtaagaaagccaaacaagggttgttttt
tggactgcaatttttgggaggacatagtagccgcgcgccagttacgtc

图32A

SEQ ID NO:60

来自轮枝镰刀菌的GH3家族β-葡糖苷酶Fv3C的蛋白质序列

mklnwvaaalsiqaaqtdsavalasavpdtlagvkkadaqkvvtrdtlayspphypspwmdpnavgweea
yakaksfvsqltlmekvnlttgvgwqgercvgnv**gsiprlgmrglclqdgplgirlsdynsafpagttag**
aswskslwyergllmgtefkekgidialgpatgplgrtaaggrnwegftvdpymaghamaeavkgiqdag
viacakhyianeqehfrqsgevqsrkyniseslssnlddktmhelyawpfadavragvgsvmcsynqinn
sygcqnskllngilkdemgfqgfvmsdwaaqhtgaasavagldmsmp*gdtafdsgysfwggnltlaving*
tvpawrvddmalrimsaffkvgktiedlpdinfsswtrdtfgfvhtfaqenreqvnfgvnvqhdhkshir
*eaaakgs***vvlkntgslplknpkflavigedagpnpagpngcgdrgcdngtlamawgsgtsqfpylitpdq**
glsnratqdgtryesiltnnewasvqalvsqpnvtaivfanadsgegyievdgnfgdrknltlwqqgdel
iknvssicpntivvlhtvgpvlladyeknpnitaivwaglpgqesgnaiadllygkvspgrspftwgrtr
esygtevlyeanngrgapqddfsegvfidyrhfdrrspstdgksspnntaaplyefghglswstfeysdl
niqknvenpysppagqtipaptfgnfsknlndyvfpkgvryiykfiypflntsssaseasndggqfgkta
eeflppnalngsaqprlpasgapggnpqlwdilytvtatitntgnatsdeipqlyvslggenepirvlrg
fdrieniapgqsaifnaqltrrdlsnwdtnaqnwvitdhpktvwvgsssrklplsakle

图32B

SEQ ID NO:61

来自里氏木霉的GH3家族β-葡糖苷酶Tr3A的核苷酸序列

atgcgttaccgaacagcagctgcgctggcacttgccactgggccctttgctagggcagacagtcagtata
gctggtcccatactgggatgtgatatgtatcctggagacaccatgctgactcttgaatcaaggtagctca
acatcgggggcctcggctgaggcagttgtacctcctgcagggactccatggggaaccgcgtacgacaagg
cgaaggccgcattggcaaagctcaatctccaagataaggtcggcatcgtgagcggtgtcggctggaacgg
cggtccttgcgttggaaacacatctccggcctccaagatcagctatccatcgctatgccttcaagacgga
cccctcggtgttcgatactcgacaggcagcacagcctttacgccgggcgttcaagcggcctcgacgtggg
atgtcaatttgatccgcgaacgtggacagttcatcggtgaggaggtgaaggcctcggggattcatgtcat
acttggtcctgtggctgggccgctgggaaagactccgcagggcggtcgcaactgggagggcttcggtgtc
gatccatatctcacgggcattgccatgggtcaaaccatcaacggcatccagtcggtaggcgtgcaggcga
cagcgaagcactatatcctcaacgagcaggagctcaatcgagaaaccatttcgagcaacccagatgaccg
aactctccatgagctgtatacttggccatttgccgacgcggttcaggccaatgtcgcttctgtcatgtgc
tcgtacaacaaggtcaataccacctgggcctgcgaggatcagtacacgctgcagactgtgctgaaagacc
agctggggttcccaggctatgtcatgacggactggaacgcacagcacacgactgtccaaagcgcgaattc
tgggcttgacatgtcaatgcctggcacagacttcaacggtaacaatcggctctggggtccagctctcacc
aatgcggtaaatagcaatcaggtccccacgagcagagtcgacgatatggtgactcgtatcctcgccgcat
ggtacttgacaggccaggaccaggcaggctatccgtcgttcaacatcagcagaaatgttcaaggaaacca
caagaccaatgtcagggcaattgccagggacggcatcgttctgctcaagaatgacgccaacatcctgccg
ctcaagaagcccgctagcattgccgtcgttggatctgccgcaatcattggtaaccacgccagaaactcgc
cctcgtgcaacgacaaaggctgcgacgacggggccttgggcatgggttggggttccggcgccgtcaacta
tccgtacttcgtcgcgccctacgatgccatcaataccagagcgtcttcgcagggcacccaggttaccttg
agcaacaccgacaacacgtcctcaggcgcatctgcagcaagaggaaaggacgtcgccatcgtcttcatca
ccgccgactcgggtgaaggctacatcaccgtggagggcaacgcgggcgatcgcaacaacctggatccgtg
gcacaacggcaatgccctggtccaggcggtggccggtgccaacagcaacgtcattgttgttgtccactcc
gttggcgccatcattctggagcagattcttgctcttccgcaggtcaaggccgttgtctgggcgggtcttc
cttctcaggagagcggcaatgcgctcgtcgacgtgctgtggggagatgtcagcccttctggcaagctggt
gtacaccattgcgaagagccccaatgactataacactcgcatcgtttccggcggcagtgacagcttcagc
gagggactgttcatcgactataagcacttcgacgacgccaatatcacgccgcggtacgagttcggctatg
gactgtgtaagtttgctaacctgaacaatctattagacaggttgactgacggatgactgtggaatgatag
cttacaccaagttcaactactcacgcctctccgtcttgtcgaccgccaagtctggtcctgcgactggggc
cgttgtgccgggaggcccgagtgatctgttccagaatgtcgcgacagtcaccgttgacatcgcaaactct
ggccaagtgactggtgccgaggtagcccagctgtacatcacctacccatcttcagcacccaggacccctc
cgaagcagctgcgaggctttgccaagctgaacctcacgcctggtcagagcggaacagcaacgttcaacat
ccgacgacgagatctcagctactgggacacggcttcgcagaaatgggtggtgccgtcggggtcgtttggc
atcagcgtgggagcgagcagccgggatatcaggctgacgagcactctgtcggtagcgtag

图33A

SEQ ID NO:62

来自里氏木霉的GH3家族β-葡糖苷酶Tr3A的蛋白质序列

MRYRTAAALALATGPFARADSHSTSGASAEAVVPPAGTPWGTAYDKAKAALAKLNLQDKVGIVSGVGWNG
GPCVGN**TSPASKISYPSLCLQDGPLGVRYSTGSTAFTPGVQAASTWDVNLIRERGQFIGEEVKASGIHVI**
LGPVAGPLGKTPQGGRNWEGFGVDPYLTGIAMGQTINGIQSVGVQATAKHYILNEQELNRETISSNPDDR
TLHELYTWPFADAVQANVASVMCSYNKVNTTWACEDQYTLQTVLKDQLGFPGYVMTDWNAQHTTVQSANS
GLDMSMP*GTDFNGNNRLWGPALTNAVNSNQVPTSRVDDMVTRILAAWYLTGQDQAGYPSFNISRNVQGNH*
*KTNVRAIARDG***IVLLKNDANILPLKKPASIAVVGSAAIIGNHARNSPSCNDKGCDDGALGMGWGSGAVNY**
PYFVAPYDAINTRASSQGTQVTLSNTDNTSSGASAARGKDVAIVFITADSGEGYITVEGNAGDRNNLDPW
HNGNALVQAVAGANSNVIVVVHSVGAIILEQILALPQVKAVVWAGLPSQESGNALVDVLWGDVSPSGKLV
YTIAKSPNDYNTRIVSGGSDSFSEGLFIDYKHFDDANITPRYEFGYGLSYTKFNYSRLSVLSTAKSGPAT
GAVVPGGPSDLFQNVATVTVDIANSGQVTGAEVAQLYITYPSSAPRTPPKQLRGFAKLNLTPGQSGTATF
NIRRRDLSYWDTASQKWVVPSGSFGISVGASSRDIRLTSTLSVA

图33B

SEQ ID NO:63

来自里氏木霉的GH3家族β-葡糖苷酶Tr3B的核苷酸序列

atgaagacgttgtcagtgtttgctgccgccctttttggcggccgtagctgaggccaatccctacccgcctc
ctcactccaaccaggcgtactcgcctcctttctaccttcgccatggatggaccccagtgctccaggctg
ggagcaagcctatgcccaagctaaggagttcgtctcgggcttgactctcttggagaaggtcaacctcacc
accggtgttggctggatgggtgagaagtgcgttggaaacgttggtaccgtgcctcgcttgggcatgcgaa
gtctttgcatgcaggacggcccctgggtctccgattcaacacgtacaacagcgctttcagcgttggctt
gacggccgccgccagctggagccgacacctttgggttgaccgcggtaccgctctgggctccgaggcaaag
ggcaagggtgtcgatgttcttctcggacccgtggctggccctctcggtcgcaaccccaacggaggccgta
acgtcgagggtttcggctcggatccctatctggcgggtttggctctggccgataccgtgaccggaatcca
gaacgcgggcaccatcgcctgtgccaagcacttcctcctcaacgagcaggagcatttccgccaggtcggc
gaagctaacggttacggataccccatcaccgaggctctgtcttccaacgttgatgacaagacgattcacg
aggtgtacggctggcccttccaggatgctgtcaaggctggtgtcgggtccttcatgtgctcgtacaacca
ggtcaacaactcgtacgcttgccaaaactccaagctcatcaacggcttgctcaaggaggagtacggtttc
caaggctttgtcatgagcgactggcaggcccagcacacgggtgtcgcgtctgctgttgccggtctcgata
tgaccatgcctggtgacaccgccttcaacaccggcgcatcctactttggaagcaacctgacgcttgctgt
tctcaacggcaccgtccccgagtggcgcattgacgacatggtgatgcgtatcatggctcccttcttcaag
gtgggcaagacggttgacagcctcattgacaccaactttgattcttggaccaatggcgagtacggctacg
ttcaggccgccgtcaatgagaactgggagaaggtcaactacggcgtcgatgtccgcgccaaccatgcgaa
ccacatccgcgaggttggcgccaagggaactgtcatcttcaagaacaacggcatcctgcccccttaagaag
cccaagttcctgaccgtcattggtgaggatgctggcggcaaccctgccggccccaacggctgcggtgacc
gcggctgtgacgacggcactcttgccatggagtggggatctggtactaccaacttcccctacctcgtcac
ccccgacgcggccctgcagagccaggctctccaggacggcacccgctacgagagcatcctgtccaactac
gccatctcgcagacccaggcgctcgtcagccagcccgatgccattgccattgtctttgccaactcggata
gcggcgagggctacatcaacgtcgatggcaacgagggcgaccgcaagaacctgacgctgtggaagaacgg
cgacgatctgatcaagactgttgctgctgtcaaccccaagacgattgtcgtcatccactcgaccggcccc
gtgattctcaaggactacgccaaccaccccaacatctctgccattctgtgggccggtgctcctggccagg
agtctggcaactcgctggtcgacattctgtacggcaagcagagcccgggccgcactcccttcacctgggg
cccgtcgctggagagctacggagttagtgttatgaccacgcccaacaacggcaacggcgctccccaggat
aacttcaacgagggcgccttcatcgactaccgctactttgacaaggtggctcccggcaagcctcgcagct
cggacaaggctcccacgtacgagtttggcttcggactgtcgtggtcgacgttcaagttctccaacctcca
catccagaagaacaatgtcggccccatgagcccgcccaacggcaagacgattgcggctccctctctgggc
agcttcagcaagaaccttaaggactatggcttccccaagaacgttcgccgcatcaaggagtttatctacc
cctacctgagcaccactacctctggcaaggaggcgtcgggtgacgctcactacggccagactgcgaagga
gttcctccccgccggtgccctggacggcagccctcagcctcgctctgcggcctctggcgaacccggcggc
aaccgccagctgtacgacattctctacaccgtgacggccaccattaccaacacgggctcggtcatggacg
acgccgttccccagctgtacctgagccacggcggtcccaacgagccgcccaaggtgctgcgtggcttcga
ccgcatcgagcgcattgctcccggccagagcgtcacgttcaaggcagacctgacgcgccgtgacctgtcc
aactgggacacgaagaagcagcagtgggtcattaccgactaccccaagactgtgtacgtgggcagctcct
cgcgcgacctgccgctgagcgcccgcctgccatga

图34A

SEQ ID NO:64

来自里氏木霉的GH3家族β-葡糖苷酶Tr3B的蛋白质序列

mktlsvfaaallaavaeanpyppphsnqaysppfypspwmdpsapgweqayaqakefvsgltllekvnl
ttgvgwmgekcvgnvgt**vprlgmrslcmqdgplglrfntynsafsvgltaaaswsrhlwvdrgtalgse**
akgkgvdvllgpvagplgrnpnggrnvegfgsdpylaglaladtvtgiqnagtiacakhfllneqehfr
qvgeangygypitealssnvddktihevygwpfqdavkagvgsfmcsynqvnnsyacqnsklingllke
eygfqgfvmsdwqaqhtgvasavagldmtmp*gdtafntgasyfgsnltlavlngtvpewriddmvmrim*
*apffkvgktvdslidtnfdswtngeygyvqaavnenwekvnygvdvranhanhirevgakg***tvifknng**
ilplkkpkfltvigedaggnpagpngcgdrgcddgtlamewgsgttnfpylvtpdaalqsqalqdgtry
esilsnyaisqtqalvsqpdaiaivfansdsgegyinvdgnegdrknltlwkngddliktvaavnpkti
vvihstgpvilkdyanhpnisailwagapgqesgnslvdilygkqspgrtpftwgpslesygvsvmttp
nngngapqdnfnegafidyryfdkvapgkprssdkaptyefgfglswstfkfsnlhiqknnvgpmsppn
gktiaapslgsfsknlkdygfpknvrrikefiypylstttsgkeasgdahygqtakeflpagaldgspq
prsaasgepggnrqlydilytvtatitntgsvmddavpqlylshggpneppkvlrgfdrieriapgqsv
tfkadltrrdlsnwdtkkqqwvitdypktvyvgsssrdlplsarlp

图34B

SEQ ID NO:65

来自埃默森篮状菌的针对在里氏木霉中的表达进行了密码子优化的GH3家族β-葡糖苷酶Te3A的核苷酸序列

atgcgcaacggcctcctcaaggtcgccgccttagccgctgccagcgccgtcaacggcgagaacctcgcct
acagccccccttctaccccagcccctgggccaacggccagggcgactgggccgaggcctaccagaaggc
cgtccagttcgtcagccagctcaccctcgccgagaaggtcaacctcaccaccggcaccggctgggagcag
gaccgctgcgtcggccaggtcggcagcatccccgcttaggcttccccggcctctgcatgcaggacagcc
ccctcggcgtccgcgacaccgactacaacagcgccttccctgccggcgttaacgtcgccgccacctggga
ccgcaacttagcctaccgcagaggcgtcgccatgggcgaggaacaccgcggcaagggcgtcgacgtccag
ttaggccccgtcgccggcccccttaggccgctctcctgatgccggccgcaactgggagggcttcgccccg
accccgtcctcaccggcaacatgatggccagcaccatccagggcatccaggatgctggcgtcattgcctg
cgccaagcacttcatcctctacgagcaggaacacttccgccagggcgcccaggacggctacgacatcagc
gacagcatcagcgccaacgccgacgacaagaccatgcacgagttatacctctggcccttcgccgatgccg
tccgcgccggtgtcggcagcgtcatgtgcagctacaaccaggtcaacaacagctacgcctgcagcaacag
ctacaccatgaacaagctcctcaagagcgagttaggcttccagggcttcgtcatgaccgactggggcggc
caccacagcggcgtcggctctgccctcgccggcctcgacatgagcatgcccggcgacattgccttcgaca
gcggcacgtctttctggggcaccaacctcaccgttgccgtcctcaacggctccatccccgagtggcgcgt
cgacgacatggccgtccgcatcatgagcgcctactacaaggtcggccgcgaccgctacagcgtccccatc
aacttcgacagctggaccctcgacacctacggccccgagcactacgccgtcggccagggccagaccaaga
tcaacgagcacgtcgacgtccgcggcaaccacgccgagatcatccacgagatcggcgccgcctccgccgt
cctcctcaagaacaagggcggcctccccctcactggcaccgagcgcttcgtcggtgtctttggcaaggat
gctggcagcaacccctggggcgtcaacggctgcagcgaccgcggctgcgacaacggcaccctcgccatgg
gctggggcagcggcaccgccaactttccctacctcgtcacccccgagcaggccatccagcgcgaggtcct
cagccgcaacggcaccttcaccggcatcaccgacaacggcgccttagccgagatggccgctgccgcctct
caggccgacacctgcctcgtctttgccaacgccgactccggcgagggctacatcaccgtcgatggcaacg
agggcgaccgcaagaacctcaccctctggcagggcgccgaccaggtcatccacaacgtcagcgccaactg
caacaacaccgtcgtcgtcttacacaccgtcggccccgtcctcatcgacgactggtacgaccaccccaac
gtcaccgccatcctctgggccggtttacccggtcaggaaagcggcaacagcctcgtcgacgtcctctacg
gccgcgtcaaccccggcaagacccccttcacctggggcagagcccgcgacgactatggcgcccctctcat
cgtcaagcctaacaacggcaagggcgcccccccagcaggacttcaccgagggcatcttcatcgactaccgc
cgcttcgacaagtacaacatcacccccatctacgagttcggcttcggcctcagctacaccaccttcgagt
tcagccagttaaacgtccagcccatcaacgcccctccctacacccccgccagcggctttacgaaggccgc
ccagagcttcggccagccctccaatgccagcgacaacctctaccctagcgacatcgagcgcgtcccctc
tacatctacccctggctcaacagcaccgacctcaaggccagcgccaacgaccccgactacggcctcccca
ccgagaagtacgtcccccccaacgccaccaacggcgacccccagcccattgaccctgccggcggtgcccc
tggcggcaaccccagcctctacgagcccgtcgcccgcgtcaccaccatcatcaccaacaccggcaaggtc
accggcgacgaggtcccccagctctatgtcagcttaggcggccctgacgacgcccccaaggtcctccgcg
gcttcgaccgcatcaccctcgcccctggccagcagtacctctggaccaccaccctcactcgccgcgacat
cagcaactgggaccccgtcacccagaactgggtcgtcaccaactacaccaagaccatctacgtcggcaac
agcagccgcaacctcccccctccaggcccccctcaagccctaccccggcatctgatga

图35A

SEQ ID NO:66

来自埃默森篮状菌的GH3家族β-葡糖苷酶Te3A的蛋白质序列

mrnqllkvaalaaasavnqenlaysppfypspwangqgdwaeayqkavqfvsqltlaekvnlttgtgweq
drcvgqv**gsiprlgfpglcmqdsplgvrdtdynsafpagvnvaatwdrnlayrrgvamgeehrgkgvdvq**
lgpvagplgrspdagrnwegfapdpvltgnmmastiqgiqdagviacakhfilyeqehfrqgaqdgydis
dsisanaddktmhelylwpfadavragvgsvmcsynqvnnsyacsnsytmnkllkselgfqgfvmtdwgg
hhsgvgsalagldmsmp*gdiafdsgtsfwgtnltvavlngsipewrvddmavrimsayykvgrdrysvpi*
*nfdswtldtygpehyavgggqtkinehvdvrgnhaeiiheigaas***avllknkgglpltgterfvgvfgkd**
agsnpwgvngcsdrgcdngtlamgwgsgtanfpylvtpeqaiqrevlsrngtftgitdngalaemaaaas
qadtclvfanadsgegyitvdgnegdrknltlwqgadqvihnvsancnntvvvlhtvgpvliddwydhpn
vtailwaglpgqesgnslvdvlygrvnpgktpftwgrarddygaplivkpnngkgapqqdftegifidyr
rfdkynitpiyefgfglsyttfefsqlnvqpinappytpasgftkaaqsfgqpsnasdnlypsdiervpl
yiypwlnstdlkasandpdyglptekyvppnatngdpqpidpaggapggnpslyepvarvttiitntgkv
tgdevpqlyvslggpddapkvlrgfdritlapgqqylwtttltrrdisnwdpvtqnwvvtnytktiyvgn
ssrnlplqaplkpypgi

图35B

SEQ ID NO:67

来自黑曲霉的GH3家族β-葡糖苷酶An3A的核苷酸序列

atgcgcttcaccagcatcgaggccgtcgccctcaccgccgtcagcctcgccagcgccgacgagttagcct
acagccccccctactaccccagcccctgggccaacggccagggcgactgggccgaggcctaccagcgcgc
cgtcgacatcgtcagccagatgaccctcgccgagaaggtcaacctcaccaccggcaccggctgggagtta
gagttatgcgtcggccagactggtggcgtcccccgcctcggcatccccggcatgtgcgcccaggacagcc
ccctcggcgtccgcgacagcgactacaacagcgccttccctgccggcgtcaacgtcgccgccacctggga
caagaacctcgcctacctccgcggccaggccatgggccaggaattcagcgacaagggcgccgacatccag
ttaggccccgctgccggccctttaggccgctctcccgacggcggcagaaactgggagggcttcagccccg
accccgctctcagcggcgtcctcttcgccgagactatcaagggcatccaggatgctggcgtcgtcgccac
cgccaagcactacattgcctacgagcaggaacacttccgccaggcccccgaggcccagggctacggcttc
aacatcaccgagagcggcagcgccaacctcgacgacaagaccatgcacgagttatacctctggcccttcg
ccgacgccattagagctggcgctggtgctgtcatgtgcagctacaaccagatcaacaacagctacggctg
ccagaacagctacaccctcaacaagctcctcaaggccgagttaggcttccagggcttcgtcatgtccgac
tgggccgcccaccacgccggcgtcagcggcgccttagccggcctcgacatgagcatgcccggcgacgtcg
actacgacagcggcaccagctactggggcaccaacctcaccatcagcgtcctcaacggcaccgtccccca
gtggcgcgtcgacgacatggccgtccgcatcatggccgcctactacaaggtcggccgcgaccgcctctgg
acccccccccaacttcagcagctggacccgcgacgagtacggcttcaagtactactacgtcagcgagggcc
cctatgagaaggtcaaccagttcgtcaacgtccagcgcaaccacagcgagttaatccgccgcatcggcgc
cgacagcaccgtcctcctcaagaacgacggcgccctccccctcaccggcaaggaacgcctcgtcgccctc
atcggcgaggacgccggcagcaacccctacggcgccaacggctgcagcgaccgcggctgcgacaacggca
ccctcgccatgggctggggcagcggcaccgccaacttcccttacctcgtcacccccgagcaggccatcag
caacgaggtcctcaagaacaagaacggcgtctttaccgccaccgacaactgggccatcgaccagatcgag
gccttagccaagaccgcctctgtcagcctcgtctttgtcaacgccgacagcggcgagggctacatcaacg
tcgacggcaacctcggcgaccgccgcaacctcaccctctggcgcaacggcgacaacgtcatcaaggccgc
cgccagcaactgcaacaacaccatcgtcatcatccacagcgtcggccccgtcctcgtcaacgagtggtac
gacaaccccaacgtcaccgccatcctctggggcggcttacccggccaggaaagcggcaacagcctcgccg
acgtcctctacggccgcgtcaaccctggcgccaagagccccttcacctggggcaagacccgcgaggcctа
tcaggactacctctacaccgagcccaacaacggcaacggcgcccccaggaagatttcgtcgagggcgtc
tttatcgactaccgcggctttgacaagcgcaacgagactcccatctacgagttcggctacggcctcagct
acaccaccttcaactacagcaacctccaggtcgaggtcctcagcgcccctgcctacgagcccgccagcgg
cgagactgaggccgcccccaccttcggcgaggtcggcaacgccagcgactacttatccccgacggcctc
cagcgcatcaccaagttcatctacccctggctcaacagcaccgacctcgaggccagcagcggcgacgcct
cttacggccaggacgcctccgactacctccccgagggtgccaccgacggcagcgctcagcccatcttacc
tgccggtggcggtgctggcggcaaccccagactctacgacgagctgatccgcgtcagcgtcaccatcaag
aacaccggcaaggtcgctggtgacgaggtcccccagctctacgtcagcttaggcggccctaacgagccca
agatcgtcctccgccagttcgagcgcatcaccctccagcccagcaaggaaactcagtggagcaccaccct
cactcgccgcgacctcgccaactggaacgtcgagactcaggactgggagatcaccagctaccccaagatg
gtctttgccggcagcagcagccgcaagctcccctccgcgccagcctccccaccgtccactgatga

图36A

SEQ ID NO:68

来自黑曲霉的GH3家族β-葡糖苷酶An3A的蛋白质序列

mrftsieavaltavslasadelaysppyypspwangqgdwaeayqravdivsqmtlaekvnlttgtgwel
elcvgq**tggvprlgipgmcaqdsplgvrdsdynsafpagvnvaatwdknlaylrgqamgqefsdkgadiq**
lgpaagplgrspdggrnwegfspdpalsgvlfaetikgiqdagvvatakhyiayeqehfrqapeaqgygf
nitesgsanlddktmhelylwpfadairagagavmcsynqinnsygcqnsytlnkllkaelgfqgfvmsd
waahhagvsgalagldmsmp*gdvdydsgtsywgtnltisvlngtvpqwrvddmavrimaayykvgrdrlw*
*tppnfsswtrdeygfkyyyvsegpyekvnqfvnvqrnhselirrigads***tvllkndgalpltgkerlval**
igedagsnpygangcsdrgcdngtlamgwgsgtanfpylvtpeqaisnevlknkngvftatdnwaidqie
alaktasvslvfvnadsgegyinvdgnlgdrrnltlwrngdnvikaaasncnntiviihsvgpvlvnewy
dnpnvtailwgglpgqesgnsladvlygrvnpgakspftwgktreayqdylytepnngngapqedfvegv
fidyrgfdkrnetpiyefgyglsyttfnysnlqvevlsapayepasgeteaaptfgevgnasdylypdgl
qritkfiypwlnstdleassgdasygqdasdylpegatdgsaqpilpagggaggnprlydelirvsvtik
ntgkvagdevpqlyvslggpnepkivlrqferitlqpsketqwsttltrrdlanwnvetqdweitsypkm
vfagsssrklplraslptvh

图36B

SEQ ID NO:69

来自轮枝镰刀菌的GH3家族β-葡糖苷酶Fo3A的核苷酸序列

atgaagctgaactgggtcgccgcagccctctctataggtgctgctggcactgatggtgcagttgctcttg
cttctgaagttccaggcactttggctggtgtaaaggtcggtttttttaccatttcctcacctaatctcag
ccttgttgccatatcgcccttattcgctcggacgctacgcaccaaatcgcgatcatttcctcccttgcag
ccttgttttctttttcgatcttccctccgcaatcgccagcacccttagcctacacaaaaaccccgaga
cagtctcattgagtttgtcgacatcaagttgcttctcaagtgtgcatttgcgtggctgtctacttctgcc
tctagaccaccaaatctgggcgcaattgatcgctcaaaccttgttcgaataagccttttattcgagacgt
ccaatttttacagagaatgtacctttcaataataccgacgttatgcgcggcggtggctgctgtgatggtt
gttgatcagaatactgacgctcaaaaggttgtcacgagagatacactcgcacactcacctcctcactatc
cttcaccatggatggatcctaatgccattggctgggaggaagcttacgccaaagcaaagaactttgtgtc
ccagctcactctcctcgaaaaggtcaacttgaccactggtgttgggtaagtagctccttgcgaacagtgc
atctcggtctccttgactaacgactctctcaggtggcaaggcgaacgctgtgtaggaaacgtgggatcaa
ttcctcgtcttggtatgcgaggtctttgtcttcaggatggtcctcttggaattcgtctgtccgattacaa
cagtgcttttcccgctggcaccacagctggtgcttcttggagcaagtctctctggtatgagaggggtctt
ctgatgggaactgagttcaaggggaagggtatcgatatcgctcttggccctgctactggtcctcttggcc
gcactgctgctggtggacgaaactgggagggctttaccgttgatccttatatggctggccatgccatggc
cgaggccgtcaagggcatccaagacgcaggtgtcattgcttgtgctaagcattacatcgcaaacgagcaa
ggtaagccaattggacggtttgggaaatcgacagagaactgacccccttgtagagcacttccgacagagt
ggcgaggtccagtcccgcaagtacaacatctccgagtctctctcctccaacctggacgacaagactttgc
acgagctctacgcctggccctttgctgatgccgtccgcgctggcgtcggttcagtcatgtgctcttacaa
tcagatcaacaactcgtacggttgccagaactccaagctcctcaacggtatcctcaaggacgagatgggt
ttccagggcttcgtcatgagcgattgggcggcccagcacaccggtgctgcttctgccgtcgctggtcttg
atatgagcatgcctggtgacaccgcgttcgacagtggatatagcttctggggtggaaacctgactcttgc
tgtcatcaacggaactgttcccgcctggcgagttgatgacatggctctgcgaatcatgtcggccttcttc
aaggttggaaagacggtagaggacctccccgacatcaacttctcctcctggacccgcgacaccttcggct
tcgtccaaacatttgctcaagagaaccgcgaacaagtcaactttggagttaacgtccagcacgaccacaa
gaaccacatccgtgagtctgccgccaagggaagcgtcatcctcaagaacaccggctcccttccccctcaac
aatcccaagttcctcgctgtcattggtgaggacgccggtcccaaccctgctggacccaatggttgcggcg
accgtggttgcgacaatggtaccctggctatggcttggggctcgggaacttctcaattcccttacttgat
cacacccgaccaaggtctccagaaccgagctgcccaagacggaactcgatatgagagcatcttgaccaac
aacgaatgggcccagacacaggctcttgtcagccaacccaacgtgaccgctatcgtttttgccaacgccg
actctggtgagggttacattgaagtcgacggaaacttcggtgatcgcaagaacctcaccctctggcaaca
gggagacgagctcatcaagaacgtctcgtccatctgccccaacaccattgtcgttctgcataccgtcggc
cctgtcctgctcgccgactacgagaagaaccccaacatcaccgccatcgtctgggctggtcttcccggcc
aagagtctggcaatgccatcgctgatctcctctacggcaaggtaagccctggccgatctcccttcacttg
gggccgcacccgtgagagctacggtaccgaggttctttatgaggcgaacaacggccgtggcgctcctcag
gatgacttctcggagggtgtcttcattgactaccgtcactttgatcgacgatctcccagcaccgatggca
agagcgctcccaacaacaccgctgctcctctctacgagttcggtcatggtctgtcttggactacctttga
gtattcagacctcaacatccagaagaacgttaactccacctactctcctcctgctggtcagaccattcct
gccccaacctttggcaacttcagcaagaacctcaacgactacgtgttccctaagggtgtccgatacatct
acaagttcatctacccctttcctgaacacttcctcatccgccagcgaggcatctaacgacggcggccagtt
tggtaagactgccgaagagttcctacctccaaacgccctcaacggctcagcccagcctcgtcttccctct
tctggtgccccaggcggtaaccctcaattgtgggatatcctgtacaccgtcacagccacaatcaccaaca
caggcaacgccacctccgacgagattccccagctgtatgtcagcctcggtggcgagaacgaacccgttcg
tgtcctccgcggtttcgaccgtatcgagaacattgctcccggccagagcgccatcttcaacgctcaattg
acccgtcgcgatctgagcaactgggatgtggatgcccagaactgggttatcaccgaccatccaaagacgg
tgtgggttggaagtagttctcgcaagctgcctctcagcgccaagttggaataa

图37A

SEQ ID NO:70

来自轮枝镰刀菌的GH3家族β-葡糖苷酶Fo3A的蛋白质序列

mklnwvaaalsiqaaqtdqavalasevpgtlagvkntdaqkvvtrdtlahspphypspwmdpnaigweea
yakaknfvsqltllekvnlttgvgwqgercvgnv**gsiprlgmrglclqdgplgirlsdynsafpagttag**
aswskslwyergllmgtefkgkgidialgpatgplgrtaaggrnwegftvdpymaghamaeavkgiqdag
viacakhyianeqehfrqsgevqsrkyniseslssnlddktlhelyawpfadavragvgsvmcsynqinn
sygcqnskllngilkdemgfqgfvmsdwaaqhtgaasavagldmsmp*gdtafdsgysfwggnltlaving*
tvpawrvddmalrimsaffkvgktvedlpdinfsswtrdtfgfvqtfaqenreqvnfgvnvqhdhknhir
*esaakgs***vilkntgslplnnpkflavigedagpnpagpngcgdrgcdngtlamawgsgtsqfpylitpdq**
glqnraaqdgtryesiltnnewaqtqalvsqpnvtaivfanadsgegyievdgnfgdrknltlwqqgdel
iknvssicpntivvlhtvgpvlladyeknpnitaivwaglpgqesgnaiadllygkvspgrspftwgrtr
esygtevlyeanngrgapqddfsegvfidyrhfdrrspstdgksapnntaaplyefghglswttfeysdl
niqknvnstysppagqtipaptfgnfsknlndyvfpkgvryiykfiypflntsssaseasndggqfgkta
eeflppnalngsaqprlpssgapggnpqlwdilytvtatitntgnatsdeipqlyvslggenepvrvlrg
fdrieniapgqsaifnaqltrrdlsnwdvdaqnwvitdhpktvwvgsssrklplsakle

图37B

SEQ ID NO:71

来自玉米赤霉的GH3家族β-葡糖苷酶Gz3A的核苷酸序列

ATGAAGGCCAATTGGCTTGCCGCGGCCGTTTATTTGGCTGCTGGCACCGATGCTGCAGTCCCTGACACTT
TGGCAGGAGTCAATGTAAGCTACTCTTCAATTTCATCTCATCTCAACTTTGCCAGGCCACAACAACTTTT
CTTCACTCACGATCTTTTCACCATAAACGCAACAGTTTCACAAAAAATAAAGCCCAAATCATGTCTCTGA
TCGTTGAACTCGCCATCTTCGTTTACATCGCGGTTGTCTTTTTCTTCTTGTACTTCTCATTCGTTGTTGT
TCTCTACATTTTCGACTGGCTGTTTAGCCTTGAGATTCTTCTCACTCCCCGTGATGCCTAGATCACTCTC
TGAGGCGTTTAATCTACTTGTAGAGATGCGCCTCTCATTTGTTGTGTCGCTAGTCGCGATAGTTGCTGGA
ATTGCAGTCCTTGATCTTCCTACTGACACTCAAAAGCTCGTTGCGCGGGACACACTCGCTCACTCTCCTC
CTCACTATCCCTCGCCATGGATGGACCCTAACGCTGTCGGCTGGGAGGACGCCTACGCCAAGGCCAAGGA
CTTTGTCTCCCAGATGACTCTCCTAGAAAAGGTCAACTTGACCACTGGTGTTGGGTAAGTAACGAGCGAC
AAGACGTCTACAATCCACTAACACGATCTCTAGATGGCAGGGCGAACGTTGTGTTGGAAACGTGGGATCT
ATCCCTCGTCTCGGTATGCGAGGCCTCTGTCTCCAGGATGGTCCTCTCGGAATTCGCTTCTCCGACTACA
ACAGCGCTTTCCCTACTGGTGTCACCGCTGGTGCTTCTTGGAGTAAGGCCCTTTGGTACGAGCGAGGACG
ATTGATGGGTACCGAGTTTAAGGAGAAGGGTATCGATATTGCTCTCGGCCCTGCAACTGGTCCTCTCGGT
CGCCACGCTGCTGGTGGACGAAACTGGGAAGGCTTCACTGTCGACCCCTACGCCGCTGGCCATGCTATGG
CTGAGACTGTCAAGGGTATCCAAGATTCTGGAGTCATTGCTTGTGCTAAGCATTACATCGCAAACGAGCA
AGGTATGTACAGGCCCATTCAATGGCTTCAGGAACGAAAACTAACTCTTAATAGAACACTTCCGTCAACG
AGGCGATGTCATGTCTCAAAAGTTCAACATTTCCGAGTCTCTGTCTTCCAACCTTGACGATAAGACTATG
CACGAGCTCTACAACTGGCCTTTCGCCGACGCCGTCCGCGCCGGTGTTGGCTCCATTATGTGCTCTTACA
ACCAGGTCAACAACTCATATGCTTGCCAGAACTCCAAGCTCCTCAACGGCATCCTCAAGGACGAGATGGG
TTTCCAGGGTTTCGTCATGAGCGATTGGCAGGCTCAGCACACCGGTGCCGCCTCCGCTGTTGCCGGTCTT
GACATGACCATGCCTGGTGACACCGAGTTCAACACTGGCTTCAGCTTCTGGGGTGGAAACCTGACCCTCG
CTGTTATCAACGGTACTGTTCCCGCCTGGAGAATCGACGACATGGCTACCCGAATTATGGCTGCTTTCTT
CAAGGTTGGCCGATCTGTTGAGGAGGAACCCGACATCAACTTCTCAGCTTGGACTCGTGATGAGTATGGC
TTCGTCCAGACCTACGCCCAAGAGAACCGAGAAAAGGTCAACTTTGCTGTTAATGTCCAGCACGACCACA
AGCGCCACATTCGCGAGGCTGGCGCAAAGGGATCCGTCGTCCTCAAGAACACTGGCTCACTTCCTCTTAA
GAAGCCCCAGTTCCTCGCTGTCATTGGAGAGGACGCTGGTTCCAACCCTGCCGGACCCAACGGTTGCGCT
GACCGTGGATGCGACAACGGTACTCTTGCCATGGCATGGGGTTCCGGAACCTCTCAATTCCCCTACCTTG
TCACCCCCGACCAAGGCATCTCGCTCCAGGCTATTCAGGACGGTACTCGTTATGAGAGCATCCTCAACAA
CAACCAGTGGCCCCAGACACAAGCTCTTGTCAGCCAGCCCAACGTCACCGCCATTGTCTTTGCCAATGCC
GATTCTGGTGAGGGCTACATCGAGGTTGACGGCAACTACGGCGACCGCAAGAACCTCACTCTGTGGAAGC
AAGGCGATGAGCTCATCAAGAACGTCTCTGCTATCTGCCCCAACACCATTGTGGTCCTTCACACCGTTGG
CCCCGTCCTTCTAACCGAGTGGCACAACAACCCCAACATCACCGCCATTGTTTGGGCTGGTGTGCCTGGA
CAGGAGTCCGGTAACGCCATCGCCGACATCCTCTACGGCAAGACCAGCCCTGGACGTTCTCCCTTCACCT
GGGGTCGCACTTATGACAGCTATGGCACCAAGGTTCTCTACAAGGCCAACAATGGAGAGGGTGCCCCTCA
AGAGGACTTTGTCGAGGGCAACTTCATCGACTACCGCCACTTTGACCGACAATCCCCCAGCACCAACGGA
AAGAGTGCCACCAACGACTCTTCTGCTCCTCTCTACGAGTTCGGTTTCGGTCTGTCCTGGACTACCTTTG
AGTACTCTGATCTCAAAGTCGAGTCTGTCAGCAACGCCTCTTACAGCCCCTCTGTCGGAAACACCATTCC
TGCCCCTACCTACGGCAACTTCAGCAAGAACCTGGACGATTACACATTCCCCTCAGGTGTCCGATACCTC
TACAAGTTCATCTACCCCTACCTCAACACCTCTTCCTCCGCTGAGAAGGCTTCCGGCGATGTCAAGGGCA
GATTTGGTGAGACCGGCGACGAGTTCCTCCCTCCCAACGCTCTCAACGGTTCATCGCAGCCTCGTCTTCC
TTCCAGTGGTGCTCCCGGCGGTAACCCTCAGCTCTGGGACATTATGTACACCGTCACTGCCACCATCACC
AACACTGGTGACGCTACCTCGGATGAGGTTCCCCAGCTGTACGTCAGCCTCGGTGGTGAGGGCGAGCCTG
TCCGTGTCCTCCGTGGCTTCGAGCGTCTTGAAAACATTGCTCCTGGTGAGAGTGCCACATTCACCGCTCA
GCTTACTCGCCGTGACCTGAGCAACTGGGACGTCAACGTCCAGAACTGGGTCATCACCGATCACGCCAAG
AAGATCTGGGTCGGCAGCAGCTCTCGCAATCTGCCCCTCAGCGCCGACCTGTAG

图38A

SEQ ID NO:72

来自玉米赤霉的GH3家族β-葡糖苷酶Gz3A的蛋白质序列

mkanwlaaavylaaqtdaavpdtlagvnlvardtlahspphypspwmdpnavgwedayakakdfvsqmtl
lekvnlttgvgwqgercvgnv**gsiprlgmrglclqdgplgirfsdynsafptgvtagaswskalwyergr**
lmgtefkekgidialgpatgplgrhaaggrnwegftvdpyaaghamaetvkgiqdsgviacakhyianeq
ehfrqrgdvmsqkfniseslssnlddktmhelynwpfadavragvgsimcsynqvnnsyacqnskllngi
lkdemgfqgfvmsdwqaqhtgaasavagldmtmp*gdtefntgfsfwggnltlavingtvpawriddmatr*
*imaaffkvgrsveeepdinfsawtrdeygfvqtyaqenrekvnfavnvqhdhkrhireagakgs***vvlknt**
gslplkkpqflavigedagsnpagpngcadrgcdngtlamawgsgtsqfpylvtpdqgislqaiqdgtry
esilnnnqwpqtqalvsqpnvtaivfanadsgegyievdgnygdrknltlwkqgdeliknvsaicpntiv
vlhtvgpvlltewhnnpnitaivwagvpgqesgnaiadilygktspgrspftwgrtydsygtkvlykann
gegapqedfvegnfidyrhfdrqspstngksatndssaplyefgfglswttfeysdlkvesvsnasysps
vgntipaptygnfsknlddytfpsgvrylykfiypylntsssaekasgdvkgrfgetgdeflppnalngs
sqprlpssgapggnpqlwdimytvtatitntgdatsdevpqlyvslggegepvrvlrgferleniapges
atftaqltrrdlsnwdvnvqnwvitdhakkiwvgsssrnlplsadl

图38B

SEQ ID NO:73

来自鞭毛藻丛赤壳菌的GH3家族β-葡糖苷酶Nh3A的核苷酸序列

atgcggttcaccgtccttctcgcggcattttcggggcttgtccccatggttggttcgcaagctgaccaga
aaccactacagctcggtgtgaacaataacactctggcgcattcacctcctcactatccttcgccatggat
ggatcctgctgctcctggctgggaggaagcctatctcaaggcgaaagattttgtttcacagcttacctt
cttgaaaaggtcaacttgaccactggtgttgggtgagtcacttgttttcctctctcctgacgtgacactt
tgctttggcctgcttcctatatcgtctactagcattgctaacactcgaggcagatggatgggcgaacgtt
gcgtcggcaacgtgggttcactccctcgttttggaatgcgtggtctctgcatgcaggatggcccctcgg
catccgcttgtctgactataactctgcctttcctactggtattacagctggtgcctcttggagccgtgcc
ctttggtaccaacgtggcctcctgatgggcaccgagcatcgtgaaaaaggcatcgacgttgcacttgggc
ctgctactggtcctcttggtcgtactcctactggcggccgcaactgggagggtttctcggttgatcccta
cgttgctggcgttgccatggccgagactgttagcggcattcaagatggtggtactatcgcctgtgctaag
cactacatcggcaacgaacaaggtatgcctcttcacttctcctcgctgataaatctgctcacaacaacct
agagcaccatcgccaagcccccgaatccattggccgcggctacaacatcaccgagtccctgtcgtcgaac
gttgatgacaagaccctccacgagctctatctctggccgttcgcagatgccgtcaaggctggtgttggtg
ctatcatgtgttcctaccagcagctgaacaactcttacggttgccaaaactctaagcttctcaacggaat
tctcaaggacgagctaggattccagggcttcgtcatgagtgactggcaagcccaacatgctggagctgct
accgctgttgcaggccttgacatgaccatgcccggtgacactttgttcaacaccggatacagcttctggg
gtggtaacctgaccctcgctgtagtcaatggcactgttcccgactggcgtattgacgacatggctatgag
aatcatggcagctttcttcaaggttggcaagactgttgaggaccttcctgacatcaacttttcttcttgg
tctcgagacacttttggctacgttcaagccgctgcccaagagaactgggaacagatcaacttcggagttg
atgttcgtcacgaccacagcgaacacattcgactctcggccgccaagggcaccgtcctccttaagaactc
tggctcattgcctctgaagaagcccaagttccttgccgtcgttggcgaggacgccggcccgaaccctgct
ggccccaacggctgtaacgaccgcggatgtaacaacggcactctggccatgtcctggggctcaggaacag
cccagttcccttacctcgttactcccgactcagcgctacagaaccaggctgtcctcgacggcactcgcta
cgagagtgtcttgcggaacaaccagtgggaacagacacgcagtctcattagccaacctaacgtgacggct
attgtgtttgccaatgccaattccggagagggatatatcgatgttgacggcaacgaaggcgatcggaaga
atttgaccttgtggaacgagggtgatgacctaattaagaacgtctcctcaatctgccccaacaccattgt
tgttctgcacactgttggccctgtcatcctgacggaatggtatgacaacccgaacattaccgccatagtg
tgggctggtgtacctggacaggagtccggcaatgctcttgtggacatcctttatggcaaaacaagccctg
gtcgctctcccttcacatggggtcgcacccgaaagagttacggcactgatgtcctatacgagcccaacaa
tggtcagggtgctcctcaagatgatttcacggagggagtctttatcgactatcgtcattttgaccaggtt
tctcctagcaccgacggcagcaagtctaatgatgagtccagtcccatctacgagtttggccatggtctgt
cctggaccacgtttgagtactctgaactcaacattcaagctcacaacaagattcccttcgatcctcctat
tggcgagacgattgccgctccggtccttggcaactacagtaccgaccttgccgattacacgttccccgat
ggaattcgctacatctaccagttcatctatccctggttgaatacttcttcttccggaagagaggcttctg
gcgatcccgactacggaaagacggccgaagagttcctgccccccggagctctcgacgggtcagctcagcc
gcgacctccatcctctggtgctccaggtggaaaccctcatctttgggatgtgttgtacactgttagtgct
atcatcaccaacactggcaacgccacctcggacgagatcccgcagctctacgttagtctcggtggcgaga
acgagcccgtccgcgtccttcgcgggttcgaccgaattgagaacattgcgcctggccagagtgtcagatt
cacaactgacatcactcgccgcgacctgagcaactgggacgtcgtctctcagaactgggtcattacagac
tacgagaagaccgtatatgtcgggagcagctcccgcaacctgcctctcaaggcaaccctgaagtaa

图39A

SEQ ID NO:74

来自鞭毛藻丛赤壳菌的GH3家族β-葡糖苷酶Nh3A的蛋白质序列

mrftvllaafsglvpmvqsqadqkplqlgvnnntlahspphypspwmdpaapgweeaylkakdfvsqltl
lekvnlttgvgwmgercvgnv**gslprfgmrglcmqdgplgirlsdynsafptgitagaswsralwyqrgl**
lmgtehrekgidvalgpatgplgrtptggrnwegfsvdpyvagvamaetvsgiqdggtiacakhyigneq
ehhrqapesigrgyniteslssnvddktlhelylwpfadavkagvgaimcsyqqlnnsygcqnskllngi
lkdelgfqgfvmsdw*qaqhagaatavagldmtmpgdtlfntgysfwggnltlavvngtvpdwriddmamr*
*imaaffkvgktvedlpdinfsswsrdtfgyvqaaaqenweqinfgvdvrhdhsehirlsaakg***tvllkns**
gslplkkpkflavvgedagpnpagpngcndrgcnngtlamswgsgtaqfpylvtpdsalqnqavldgtry
esvlrnnqweqtrslisqpnvtaivfananSgegyidvdgnegdrknltlwnegddliknvssicpntiv
vlhtvgpviltewydnpnitaivwagvpgqesgnalvdilygktspgrspftwgrtrksygtdvlyepnn
gqgapqddftegvfidyrhfdqvspstdgsksndesspiyefghglswttfeyselniqahnkipfdppi
getiaapvlgnystdladytfpdgiryiyqfiypwlntsssgreasgdpdygktaeeflppgaldgsaqp
rppssgapggnphlwdvlytvsaiitntgnatsdeipqlyvslggenepvrvlrgfdrieniapgqsvrf
ttditrrdlsnwdvvsqnwvitdyektvyvgsssrnlplk

图39B

SEQ ID NO:75

来自大丽轮枝菌的GH3家族β-葡糖苷酶Vd3A的核苷酸序列

ATGAAGCTGACCCTCGCTACTGCCTTACTGGCAGCCAGCGGGTGTGTCTCTGCGGGACAACCCAAGCTCA
AGGTACGTACTTGCCTCTTTTTCACAAGGAAACCAAACCCGCACCATAATGGTGATTGAGCAGTCGTGCT
TTCCTCAACCCGAATCAAACCCATGCCGTGTTCGCGCATGCCCTTTCGATCGTCTGTTGTGTGTGAACCC
ACGCTCTTCAAGCATCGCACATAGCACCACTCCATCTTCATTTTCGAGCAATTTCGGGCCGCAGAGAGCG
GTCTTTCACTTCACCACAATCGTTCATGCCTCGTGCCCCACTGCCATGTTTCTTCCCAGTATTCTACTTC
TGAGAGCCTTGACCACCGTTGTCGACATCTCGTCGCCAAGGCTCGTTGACACGGACTCTGTTTCCCTTGG
AATTAATATTCGAAACAATGCTGACCAGCATCCTCAGCGCCAGACTAACAGCTCTAGCGAGCTCGCCTTT
TCCCCTCCGCACTACCCTTCTCCATGGATGAACCCCCAAGCGACTGGGTGGGAGGACGCCTACGCCCGTG
CCAGAGAGGTGGTAGAGCAGATGACTCTGCTCGAAAAGGTCAACCTGACGACAGGTGTCGGGTAAGCTTC
ACAGACCCCGTCTTGCCATCCAAAGTCATCTGACAGAATCCTAGCTGGAGCGGTGATCTCTGCGTCGGAA
ACGTCGGCTCGATCCCCCGAATCGGCTGGAGGGGGCTTTGTTTGCAGGATGGCCCACAGGGTATCCGTTT
CGCGGACTACGTCTCGTACTTCACTTCGAGCCAGACAGCCGGCGCTACCTGGGACCGAGGGCTTCTGTAC
CAGCGCGCTCACGCCATTGGCGCCGAAGGAGTAGCCAAGGGCGTCGACGTCGTCCTCGGGCCCGCCATTG
GCCCTCTAGGTCGCCTTCCCGCCGGAGGTCGTAACTGGGAGGGTTTCGCCGTGGACCCTTACCTCAGTGG
CGTTGCTGTCGCCGAATCCGTCAGGGGCATCCAGGATGCTGGTGCTATTGCCAACGTCAAGCACTACATC
GTCAATGAGCAGGAACATTTCCGCCAGGCTGGCGAGGCTCAAGGTTACGGCTACGATGTCGACGAGGCAT
TATCGTCGAACGTTGACGACAAGACCATGCATGAGCTTTACCTTTGGCCATTTGCAGACGCTGTCCGTGC
TGGAGCCGGCAGTGTCATGTGTTCTTATCAACAGGTGGGGGCAATACCATTCTCTCCTCTTTCCTTGCAG
ACAGTGCACTGACCGACCTTTTTTGCCCAAGATCAACAACAGTTACGGCTGTCAAAACTCACATCTTCTG
AATGGGCTCCTCAAGGACGAACTCGGCTTTCAGGGGTTCGTCCTCAGCGATTGGCAAGCGCAGCATGCTG
GTGCTGCCACTGCCGTTGCTGGACTTGACATGGCCATGCCCGGTGACACTCGCTTCAACACCGGAGTCGC
CTTCTGGGGCGCTAACCTTACCAATGCCATTTTGAACGGCACCGTTCCCGAATATCGGCTCGATGACATG
GCCATGCGTATTATGGCGGCCTTTTTCAAAGTTGGAAAGACCCTGGACGATGTTCCTGACATCAACTTCT
CGTCTTGGACAAAAGACACCATCGGCCCGCTGCACTGGGCGGCCCAGGACAATGTGCAGGTCATCAACCA
ACACGTTGATGTCCGTCAAGACCACGGCGCCCTCATTCGCACCATCGCTGCCCGCGGTACTGTCTTACTA
AAAAATGAGGGATCACTGCCTCTGAACAAGCCGAAATTTGTTGCTGTCATTGGTGAAGATGCTGGCCCTC
GTCCTGTTGGTCCCAATGGCTGCCCTGATCAGGGTTGCAATAACGGCACTCTGGCTGCTGGATGGGGATC
TGGCACCGCCAGTTTCCCTTATCTCATCACTCCTGATAGTGCTCTTCAGTTTCAAGCCGTTTCGGATGGC
TCGCGATACGAAAGCATCCTCAGCAACTGGGATTATGAGCGCACAGAGGCCTTGGTTTCCCAGGCGGATG
CTACTGCTCTGGTTTTCGTCAATGCAAACTCTGGCGAAGGATATATCAGCGTTGATGGAAACGAAGGTGA
TCGCAAGAACCTCACTCTCTGGAATGGAGGAGACGAGCTTATTCAACGAGTCGCTGCGGCCAACAACAAC
ACCATCGTCATCATCCATTCGGTTGGTCCCGTTCTAGTCACTGACTGGTACGAGAATCCCAATATCACGG
CTATCATCTGGGCCGGCTTACCCGGACAGGAGTCTGGCAACTCTATCGCCGATATTCTTTACGGCCGCGT
GAACCCTGGTGGCAAGACACCTTTCACCTGGGGTCCAACTGTTGAGAGCTACGGCGTTGACGTCCTGAGA
GAGCCCAACAATGGCAATGGTGCTCCCCAGAGCGATTTCGACGAGGGAGTCTTCATCGATTACCGTTGGT
TTGACCGGCAGTCGGGTGTTGATAACAATGCATCAGCGCCGAGGAACAGCAGCAGCAGCCACGCCCCAAT
CTTCGAGTTTGGCTATGGCCTTTCGTACACAACCTTTGAATTCTCCAATCTTCAGATTGAGAGGCATGAC
GTTCACGATTACGTCCCTACCACTGGGCAGACGAGCCCTGCGCCGAGATTTGGTGCTAACTACAGTACGA
ACTACGACGACTACGTCTTTCCCGAGGGCGAAATCCGTTACATCTATCAACACATCTACCCATACCTCAA
TTCCTCAGACCCAAAGGAGGCATTGGCTGATCCTAAATACGGCCAAACTGCAGAAGAGTTCCTCCCAGAG
GGCGCTCTTGATGCCTCACCGCAGCCTAGGCTCCCAGCTTCTGGAGGGCCCGGAGGCAACCCAATGCTTT
GGGACGTCATATTCACGGTCACCGCGACCGTGACCAACACGGGTAAGGTTGCTGGGGACGAAGTGGCACA
GCTTTACGTTTCTCTTGGTGGACCTGACGATCCGATTCGAGTCCTCCGTGGGTTCGACCGCATTCACATC
GCGCCTGGAGCCTCGCAAACCTTCCGTGCGGAACTCACGCGCCGGGACCTCAGCAACTGGGATGTTGTCA
CGCAAAATTGGTTCATCAGCCAGTACGAAAAGACGGTCTTTGTCGGGAGCTCATCCCGAAACCTCCCTCT
CAGCACTCGCCTCGAATAG

图40A

SEQ ID NO:76

来自大丽轮枝菌的GH3家族β-葡糖苷酶Vd3A的蛋白质序列

mkltlatallaasqcvsaqqpklkhpqrqtnssselafspphypspwmnpqatgwedayararevveqmt
llekvnlttgvgwsgdlcvgnv**gsiprigwrglclqdgpqgirfadyvsyftssqtagatwdrgllyqra**
haigaegvakgvdvvlgpaigplgrlpaggrnwegfavdpylsgvavaesvrgiqdagaianvkhyivne
qehfrqageaqgygydvdealssnvddktmhelylwpfadavragagsvmcsyqqinnsygcqnshllng
llkdelgfqgfvlsdw*qaqhagaatavagldmampgdtrfntgvafwganltnailngtvpeyrlddmam*
*rimaaffkvgktlddvpdinfsswtkdtigplhwaaqdnvqvinqhvdvrqdhgalirtiaarg***tvllkn**
egslplnkpkfvavigedagprpvgpngcpdqgcnngtlaagwgsgtasfpylitpdsalqfqavsdgsr
yesilsnwdyertealvsqadatalvfvnansgegyisvdgnegdrknltlwnggdeliqrvaaannnti
viihsvgpvlvtdwyenpnitaiiwaglpgqesgnsiadilygrvnpggktpftwgptvesygvdvlrep
nngngapqsdfdegvfidyrwfdrqsgvdnnasaprnsssshapifefgyglsyttfefsnlqierhdvh
dyvpttgqtspaprfganystnyddyvfpegeiryiyqhiypylnssdpkealadpkygqtaeeflpega
ldaspqprlpasggpggnpmlwdviftvtatvtntgkvagdevaqlyvslggpddpirvlrgfdrihiap
gasqtfraeltrrdlsnwdvvtqnwfisqyektvfvgsssrnlplstrle

图40B

SEQ ID NO:77

来自鹅掌柄孢霉的GH3家族β-葡糖苷酶Pa3G的核苷酸序列

ATGAAACTCAATAAGCCATTCCTGGCCATTTATTTGGCTTTCAACTTGGCCGAGGCTTCGAAAACTCCGG
ATTGCATCAGTGGTCCGCTGGCAAAGACCTTGGCATGTGATACAACGGCGTCACCTCCTGCGCGAGCAGC
TGCTCTTGTGCAGGCTTTAAATATCACGGAAAAGCTTGTGAATCTAGTGGAGTATGTCAAGTCAAGAGAA
GCTCCTTTAGGGATTTCAATTCAGCTAATCACTCCTCATAGCATGAGCCTCGGTGCAGAAAGGATCGGCC
TTCCAGCTTATGCTTGGTGGAACGAAGCTCTTCATGGTGTTGCCGCGTCGCCTGGGGTCTCCTTCAATCA
GGCCGGACAAGAATTCTCACACGCTACTTCATTTGCGAATACTATTACGCTAGCAGCCGCCTTTGACAAT
GACCTGGTTTACGAGGTGGCGGATACCATCAGCACTGAAGCGCGAGCGTTCAGCAATGCCGAGCTCGCTG
GACTGGATTACTGGACGCCTAACATCAACCCGTACAAAGATCCGAGATGGGGGAGGGGCCATGAGGTTTG
TTACCTTAGCCTTCTTTTCCGTGCCGTGCAGTTGCTGAGAACTCAAAAGACACCCGGAGAAGATCCGGTA
CACATCAAAGGCTACGTCCAAGCACTTCTCGAGGGTCTAGAAGGGAGAGACAAGATCAGAAAGGTGATTG
CCACTTGTAAACACTTTGCAGCCTATGATTTGGAGAGATGGCAAGGGGCTCTTAGATACAGGTTCAATGC
TGTTGTGACCTCGCAGGATCTTTCGGAGTACTACCTCCAACCGTTTCAACAATGCGCTCGAGACAGCAAG
GTCGGGTCTTTCATGTGCTCATATAATGCGCTCAACGGAACACCGGCATGTGCAAGCACGTATTTGATGG
ACGACATCCTTCGAAAACACTGGAATTGGACCGAGCACAACAACTATATAACGAGCGACTGTAATGCTAT
TCAGGACTTCCTCCCCAACTTTCACAACTTCAGCCAAACTCCAGCTCAAGCCGCCGCTGATGCTTATAAC
GCCGGTACAGACACCGTCTGTGAGGTGCCTGGATACCCCCCACTCACAGATGTAATCGGAGCATACAATC
AGTCTCTGCTGTCAGAGGAAATTATCGACCGAGCACTTCGCAGATTATACGAAGGCCTCATCCGAGCTGG
CTATCTCGACTCAGCCTCCCCACATCCATACACCAAAATCTCATGGTCCCAAGTAAACACCCCCAAAGCC
CAAGCCCTGGCTCTCCAGTCCGCCACCGACGGGATAGTCCTTCTCAAAAACAACGGCCTCCTTCCCCTAG
ACCTCACCAACAAAACCATAGCCCTCATAGGCCACTGGGCCAATGCAACCCGCCAAATGCTAGGCGGCTA
CAGCGGTATCCCCCCTTACTACGCCAACCCAATCTATGCAGCCACCCAGCTCAACGTCACTTTTCATCAC
GCCCCAGGACCGGTGAACCAGTCATCTCCCTCCACAAATGACACCTGGACCTCCCCCGCCCTCTCCGCGG
CTTCCAAATCGGATATCATCCTCTACCTCGGCGGCACCGACCTCTCCATCGCAGCCGAAGACCGAGACAG
AGACTCCATCGCCTGGCCATCCGCTCAACTTTCCTTGTTAACCTCCCTCGCCCAGATGGGAAAACCCACA
ATCGTAGCAAGACTAGGCGACCAAGTAGACGACACCCCCCTGCTCTCCAACCCAAACATCTCCTCCATCC
TATGGGTAGGCTACCCAGGCCAATCAGGCGGAACAGCCCTCTTGAACATCATCACCGGAGTCAGCTCCCC
CGCCGCTCGACTGCCCGTCACAGTCTACCCAGAAACTTACACCTCCCTCATCCCCCTGACAGCCATGTCC
CTCCGCCCAACCTCCGCCCGCCCAGGCCGGACTTACAGGTGGTACCCCTCCCCCGTGCTCCCCTTCGGCC
ACGGCCTCCACTACACAACCTTTACCGCCAAATTCGGCGTCTTTGAGTCCCTCACCATCAACATTGCCGA
ACTCGTTTCCAACTGTAACGAACGATACCTCGACCTCTGCCGGTTCCCGCAGGTGTCCGTCTGGGTGTCG
AATACGGGAGAACTCAAATCTGACTATGTCGCCCTTGTTTTTGTCAGGGGTGAGTACGGACCGGAGCCGT
ACCCGATCAAGACGCTGGTGGGGTACAAGCGGATAAGGGATATCGAGCCGGGGACTACGGGGGCGGCGCC
GGTGGGGGTGGTGGTGGGGGATTTGGCTAGGGTGGATTTGGGGGGGAATAGGGTTTTGTTTCCGGGGAAG
TATGAGTTTCTGCTGGATGTGGAGGGGGGGAGGGATAGGGTTGTGATCGAGTTGGTTGGGGAGGAGGTGG
TGTTGGAGAAGTTCCCTCAGCCGCCTGCGGCGGGTTGA

图41A

SEQ ID NO:78

来自鹅掌柄孢霉的GH3家族β-葡糖苷酶Pa3G的蛋白质序列

mklnkpflaiylafnlaeasktpdcisgplaktlacdttaspparaaalvqalniteklvnlveyvksre
aplgisiqlitphsmslgaeriglpayawwnealhgvaaspgvsfnqagqefshatsfantitlaaafdn
dlvyevadtistearafsnaelagldywtpninpykdprwgrghevcylsllfravqllrtqktpgedpv
hikgyvqalleglegrdkirkviatckhfaaydlerwqgalryrfnavvtsqdlseyylqpfqqcardsk
vgsfmcsynalngtpacastylmddilrkhwnwtehnnyitsdcnaiqdflpnfhnfsqtpaqaaadayn
agtdtvcevpgyppltdvigayngsllseeiidralrrlyegliragyldsasphpytkiswsqvntpka
qalalqsatdgivllknngllpldltnktialighwanatrqmlggysgippyyanpiyaatqlnvtfhh
apgpvnqsspstndtwtspalsaasksdiilylggtdlsiaaedrdrdsiawpsaqlslltslaqmgkpt
ivarlgdqvddtpllsnpnissilwvgypgqsggtallniitgvsspaarlpvtvypetytslipltams
lrptsarpgrtyrwypspvlpfghglhyttftakfgvfesltiniaelvsncneryldlcrfpqvsvwvs
ntgelksdyvalvfvrgeygpepypiktlvgykrirdiepgttgaapvgvvvgdlarvdlggnrvlfpgk
yeflldveggrdrvvielvgeevvlekfpqppaag

图41B

SEQ ID NO:79

来自新阿波罗栖热袍菌的GH3家族β-葡糖苷酶Tn3B的蛋白质序列

MEKVNEILSQLTLEEKVKLVVGVGLPGLFGNPHSRVAGAAGETHPVPRVGLPAFVLADGPAGLRINPTRE
NDENTYYTTAFPVEIMLASTWNRELLEEVGKAMGEEVREYGVDVLLAPAMNIHRNPLCGRNFEYYSEDPV
LSGEMASSFVKGVQSQGVGACIKHFVANNQETNRMVVDTIVSERALREIYLRGFEIAVKKSKPWSVMSAY
NKLNGKYCSQNEWLLKKVLREEWGFEGFVMSDWYAGDNPVEQLKAGNDLIMPGKAYQVNTERRDEIEEIM
EALKEGKLSEEVLDECVRNILKVLVNAPSFKNYRYSNKPDLEKHAKVAYEAGAEGVVLLRNEEALPLSEN
SKIALFGTGQIETIKGGTGSGDTHPRYAISILEGIKERGLNFDEELAKTYEDYIKKMRETEEYKPRRDSW
GTIIKPKLPENFLSEKEIHKLAKKNDVAVIVISRISGEGYDRKPVKGDFYLSDDETDLIKTVSREFHEQG
KKVIVLLNIGSPVEVVSWRDLVDGILLVWQAGQETGRIVADVLTGRINPSGKLPTTFPRDYSDVPSWTFP
GEPKDNPQKVVYEEDIYVGYRYYDTFGVEPAYEFGYGLSYTTFEYSDLNVSFDGETLRVQYRIENTGGRA
GKEVSQVYIKAPKGKIDKPFQELKAFHKTRLLNPGESEEVVLEIPVRDLASFNGEEWVVEAGEYEVRVGA
SSRNIKLKGTFSVGEERRFKP

图42

Fv3C同源物的氨基酸序列的比对结果 比对在Muscle（根据Edger R.C., BMC bioinformatics, 2004, 5: 113（Edger R.C., BMC生物信息学, 2004年, 第5卷, 第113页））中使用默认参数进行。

```
                *         20          *         40          *
Tn3B : ------------------------------------------------- :   -
Fv3G : ---MFPSSISCLA--ALSLMSQGLLAQSQPENV----------ITDDTY :  34
Fv3D : MASIRSVLVSGLL--AAGVNAQAYDASDRAEDAFSWVQP---KNTTILGQ :  45
Tr3A : --MRYRTAAALAL--ATGPFARADSHS----------------TSGASA :  29
Pa3D : ----MALQTFFLL--AAAMLANAE--------------------TTGEKV :  24
Te3A : --MRNGLLKVAAL--AAASA-----------------------VNGENL :  22
An3A : --MRFTSIEAVAL--TAVSL-----------------------ASADEL :  22
Tr3B : -MKTLSVFAAALL--AAVAEANPYPPP----------------HSNQ :  28
Nh3A : --MRFTVLLAAFSGLVPMVGSQADQKPLQLG------------VNNNTL :  35
Gz3A : --MKANWLAAAVYL-AAGTDAA------VPDTLAGV------NLVARDTL :  35
Fv3C : --MKLNWVAAALSIGAAGTDSAVALASAVPDTLAGVKKADAQKVVTRDTL :  48
Fo3A : --MKLNWVAAALSIGAAGTDGAVALASEVPGTLAGVKNTDAQKVVTRDTL :  48
Pa3G : --MKFSVVVAAAL--ASGALATPQYPPK----------------LIKRDL :  30
Vd3A : --MKLTLATALLA--ASGCVSAGQPKLKHPQRQT----------NSSSEL :  36

              60          *         80          *        100
Tn3B : ----------------------MEKVNEILSQLTLEEKVKLVVGVGLP :  26
Fv3G : FYGQSPPVYP-----THTGSWAAAVAKAKNLVSQLTLEEKVNLTTG-GQT :  78
Fv3D : YG--HSPHYPAN--NATGKGWEDAFAKAQNFVSQLTLEEKADMVTGT--- :  88
Tr3A : EA--VVP--------PAGTPWGTAYDKAKAALAKLNLQDKVGIVSGVGWN :  69
Pa3D : SR--QAP--------SGAQAWAAAHSQAAATLARMSQQDKINMVTGIGWD :  64
Te3A : AY--SPPFYPSPWANGQGD-WAEAYQKAVQFVSQLTLAEKVNLTTGTGWE :  69
An3A : AY--SPPYYPSPWANGQGD-WAEAYQRAVDIVSQMTLAEKVNLTTGTGWE :  69
Tr3B : AY--SPPFYPSPWMDPSAPGWEQAYAQAKEFVSGLTLLEKVNLTTGVGWM :  76
Nh3A : AH--SPPHYPSPWMDPAAPGWEEAYLKAKDFVSQLTLLEKVNLTTGVGWM :  83
Gz3A : AH--SPPHYPSPWMDPNAVGWEDAYAKAKDFVSQMTLLEKVNLTTGVGWQ :  83
Fv3C : AY--SPPHYPSPWMDPNAVGWEEAYAKAKSFVSQLTLMEKVNLTTGVGWQ :  96
Fo3A : AH--SPPHYPSPWMDPNAIGWEEAYAKAKNFVSQLTLLEKVNLTTGVGWQ :  96
Pa3G : AY--SPPVYPSPWMNPEADGWAEAYVKAREFVSQMTLLEKVNLTTGTGWA :  78
Vd3A : AF--SPPHYPSPWMNPQATGWEDAYARAREVVEQMTLLEKVNLTTGVGWS :  84

                *       120          *       140          *
Tn3B : GLFGNPHSRVAGAAGETHPVPRVGLPAFVLADGPAGLRINPTRENDENTY :  76
Fv3G : T---------TGCSGFIPGIPRVGFPGLCLADAGNGVRNTD--------- : 110
Fv3D : P---------GPCVGNIVAIPRLNFNGLCLHDGPLAIRVAD--------- : 120
Tr3A : G---------GPCVGNTSPASKISYPSLCLQDGPLGVRYST--------- : 101
Pa3D : R---------GPCVGNTAAISSINYPQICLQDGPLGIRFGT--------- :  96
Te3A : Q---------DRCVGQVGSIPRLGFPGLCMQDSPLGVRDTD--------- : 101
An3A : L---------ELCVGQTGGVPRLGIPGMCAQDSPLGVRDSD--------- : 101
Tr3B : G---------EKCVGNVGTVPRLGMRSLCMQDGPLGLRFNT--------- : 108
Nh3A : G---------ERCVGNVGSLPRFGMRGLCMQDGPLGIRLSD--------- : 115
Gz3A : G---------ERCVGNVGSIPRLGMRGLCLQDGPLGIRFSD--------- : 115
Fv3C : G---------ERCVGNVGSIPRLGMRGLCLQDGPLGIRLSD--------- : 128
Fo3A : G---------ERCVGNVGSIPRLGMRGLCLQDGPLGIRLSD--------- : 128
Pa3G : S---------EQCVGQVGAIPRLGLRSLCMHDAPLGIRGTD--------- : 110
Vd3A : G---------DLCVGNVGSIPRIGWRGLCLQDGPQGIRFAD--------- : 116
```

图43A-1

```
                160         *       180         *       200
Tn3B : YTTAFPVEIMLASTWNRELLEEVGKAMGEEVREYGVDVLLAPAMN-IHRN : 125
Fv3G : YVSSFPSGIHVGASWNPELTYSRSYYMGAEAKAKGVNILLGPVFGPLGRV : 160
Fv3D : YASVFPAGVSAASSWDKDLLYQRGLAMGQEFKAKGAHILLGPVAGPLGRS : 170
Tr3A : GSTAFTPGVQAASTWDVNLIRERGQFIGEEVKASGIHVILGPVAGPLGKT : 151
Pa3D : GTTAFTPGVQAASTWDVDLIRQRGAYLGAEAKGCGIHILLGPVAGALGKI : 146
Te3A : YNSAFPAGVNVAATWDRNLAYRRGVAMGEEHRGKGVDVQLGPVAGPLGRS : 151
An3A : YNSAFPAGVNVAATWDKNLAYLRGQAMGQEFSDKGADIQLGPAAGPLGRS : 151
Tr3B : YNSAFSVGLTAAASWSRHLWVDRGTALGSEAKGKGVDVLLGPVAGPLGRN : 158
Nh3A : YNSAFPTGITAGASWSRALWYQRGLLMGTEHREKGIDVALGPATGPLGRT : 165
Gz3A : YNSAFPTGVTAGASWSKALWYERGRLMGTEFKEKGIDIALGPATGPLGRH : 165
Fv3C : YNSAFPAGTTAGASWSKSLWYERGLLMGTEFKEKGIDIALGPATGPLGRT : 178
Fo3A : YNSAFPAGTTAGASWSKSLWYERGLLMGTEFKGKGIDIALGPATGPLGRT : 178
Pa3G : YNSAFPSGQTAAATWDRQLMYRRGYAIGKEAKGKGINVILGPVAGPLGRM : 160
Vd3A : YVSYFTSSQTAGATWDRGLLYQRAHAIGAEGVAKGVDVVLGPAIGPLGRL : 166

                  *       220         *       240         *
Tn3B : PLCGRNFEYYSEDPVLSGEMASSFVKGVQSQGVGACIKHFVANNQETNRM : 175
Fv3G : VEGGRNWEGFSNDPYLAGKLGHEAVAGIQDAGVVACGKHFLAQEQETHRL : 210
Fv3D : AYSGRNWEGFSPDPYLTGIAMEETIMGHQDAGVQATAKHFIGNEQEVMRN : 220
Tr3A : PQGGRNWEGFGVDPYLTGIAMGQTINGIQSVGVQATAKHYILNEQELNR- : 200
Pa3D : PHGGRNWEGFGADPYLAGIAMKETIEGIQSAGVQANAKHYIANEQELNR- : 195
Te3A : PDAGRNWEGFAPDPVLTGNMMASTIQGIQDAGVIACAKHFILYEQEHFRQ : 201
An3A : PDGGRNWEGFSPDPALSGVLFAETIKGIQDAGVVATAKHYIAYEQEHFRQ : 201
Tr3B : PNGGRNVEGFGSDPYLAGLALADTVTGIQNAGTIACAKHFLLNEQEHFRQ : 208
Nh3A : PTGGRNWEGFSVDPYVAGVAMAETVSGIQDGGTIACAKHYIGNEQEHHRQ : 215
Gz3A : AAGGRNWEGFTVDPYAAGHAMAETVKGIQDSGVIACAKHYIANEQEHFRQ : 215
Fv3C : AAGGRNWEGFTVDPYMAGHAMAEAVKGIQDAGVIACAKHYIANEQEHFRQ : 228
Fo3A : AAGGRNWEGFTVDPYMAGHAMAEAVKGIQDAGVIACAKHYIANEQEHFRQ : 228
Pa3G : PAAGRNWEGFSPDPVLTGVGMAETVKGHQDAGVIACAKHFIGNEQEHFRQ : 210
Vd3A : PAGGRNWEGFAVDPYLSGVAVAESVRGIQDAGAIANVKHYIVNEQEHFRQ : 216

                260         *       280         *       300
Tn3B : V---------------VDTIVSERALREIYLRGFEIAVKKSKPWSVMSAY : 210
Fv3G : AASVTG-------ADAISSNLDDKTLHELYL-------------CVMCSY : 240
Fv3D : PTFVKDGYIGEVDKEALSSNMDDRTMHELYLWPFANAVHAKAS-SMMCSY : 269
Tr3A : --------------ETISSNPDDRTLHELYTWPFADAVQANVA-SVMCSY : 235
Pa3D : --------------ETMSSNVDDRTQHELYLWPFADAVHANVA-SVMCSY : 230
Te3A : GAQD------GYDISDSISANADDKTMHELYLWPFADAVRAGVG-SVMCSY : 245
An3A : APEAQG--YGFNITESGSANLDDKTMHELYLWPFADAIRAGAG-AVMCSY : 248
Tr3B : VGEANG--YGYPITEALSSNVDDKTIHEVYGWPFQDAVKAGVG-SFMCSY : 255
Nh3A : APESIG--RGYNITESLSSNVDDKTLHELYLWPFADAVKAGVG-AIMCSY : 262
Gz3A : RGDVMS--QKFNISESLSSNLDDKTMHELYNWPFADAVRAGVG-SIMCSY : 262
Fv3C : SGEVQS--RKYNISESLSSNLDDKTMHELYAWPFADAVRAGVG-SVMCSY : 275
Fo3A : SGEVQS--RKYNISESLSSNLDDKTLHELYAWPFADAVRAGVG-SVMCSY : 275
Pa3G : VGEARG--YGFNISETLSSNIDDKTMHELYLWPFADAVRAGAG-SFMCSY : 257
Vd3A : AGEAQG--YGYDVDEALSSNVDDKTMHELYLWPFADAVRAGAG-SVMCSY : 263
```

图43A-2

```
                *       320         *       340         *
Tn3B : NKLNGKYCSQNEWLLKKVLREEWGFEGFVMSDWYAGDNPVEQLKAGNDLI : 260
Fv3G : NRANNSHACQNSKLLNGLLKGELGFQGFVVSDWGAQQSGMASALAGLDVV : 290
Fv3D : QRLNGSYACQNSKVLNGILRDELGFQGYVMSDWGATHAGVAAINSGLDMD : 319
Tr3A : NKVNTTWACEDQYTLQTVLKDQLGFPGYVMTDWNAQHTTVQSANSGLDMS : 285
Pa3D : NKLNGTWACENDKALNQILKKELGFQGYVLSDWNAQHSTALSANSGLDMT : 280
Te3A : NQVNNSYACSNSYTMNKLLKSELGFQGFVMTDWGGHHSGVGSALAGLDMS : 295
An3A : NQINNSYGCQNSYTLNKLLKAELGFQGFVMSDWAAHHAGVSGALAGLDMS : 298
Tr3B : NQVNNSYACQNSKLINGLLKEEYGFQGFVMSDWQAQHTGVASAVAGLDMT : 305
Nh3A : QQLNNSYGCQNSKLLNGILKDELGFQGFVMSDWQAQHAGAATAVAGLDMT : 312
Gz3A : NQVNNSYACQNSKLLNGILKDEMGFQGFVMSDWQAQHTGAASAVAGLDMT : 312
Fv3C : NQINNSYGCQNSKLLNGILKDEMGFQGFVMSDWAAQHTGAASAVAGLDMS : 325
Fo3A : NQINNSYGCQNSKLLNGILKDEMGFQGFVMSDWAAQHTGAASAVAGLDMS : 325
Pa3G : QQVNNSYGCQNSKLMNGLLKDELGFQGFVLSDWQAQHTGAAAAAAGLDMS : 307
Vd3A : QQINNSYGCQNSHLLNGLLKDELGFQGFVLSDWQAQHAGAATAVAGLDMA : 313

              360         *       380         *       400
Tn3B : MPGKAYQVNTERRDEI--EEIMEALKEGKLSEEVLDECVRNILKVL---- : 304
Fv3G : MPSS----------ILWGANLTLGVNNGTIPESQVDNMVTRLLATWYQLN : 330
Fv3D : MPGGIGAYGTYFTKSFFGGNLTRAVTNGTLDETRVNDMITRIMTPYFWLG : 369
Tr3A : MPG----TDFNGNNRLWGPALTNAVNSNQVPTSRVDDMVTRILAAWYLTG : 331
Pa3D : MPG----TDFNGRNVYWGPQLNNAVNAGQVQRSRLDDMCKRILAGWYLLG : 326
Te3A : MPGD---IAFDSGTSFWGTNLTVAVLNGSIPEWRVDDMAVRIMSAYYKVG : 342
An3A : MPGD---VDYDSGTSYWGTNLTISVLNGTVPQWRVDDMAVRIMAAYYKVG : 345
Tr3B : MPGD---TAFNTGASYFGSNLTLAVLNGTVPEWRIDDMVMRIMAPFFKVG : 352
Nh3A : MPGD---TLFNTGYSFWGGNLTLAVVNGTVPDWRIDDMAMRIMAAFFKVG : 359
Gz3A : MPGD---TEFNTGFSFWGGNLTLAVINGTVPAWRIDDMATRIMAAFFKVG : 359
Fv3C : MPGD---TAFDSGYSFWGGNLTLAVINGTVPAWRVDDMALRIMSAFFKVG : 372
Fo3A : MPGD---TAFDSGYSFWGGNLTLAVINGTVPAWRVDDMALRIMSAFFKVG : 372
Pa3G : MPGD---TEFNTGVSFWGTNLTVAVLNGTVPAYRIDDMAMRIMAAFFKVE : 354
Vd3A : MPGD---TRFNTGVAFWGANLTNAILNGTVPEYRLDDMAMRIMAAFFKVG : 360

                *       420         *       440         *
Tn3B : --------VNAPSFKNYRYSNKPDL----------------EKHAKVAYE : 330
Fv3G : QDQ-------------DTEAPGHGLAAKL-WEPHPVVDARNASSKPTIWD : 366
Fv3D : QDK-DYPSVDPSSGDLNTFSPKSSWFREF-NLTGERSRDVRGNHGDLIRK : 417
Tr3A : QDQAGYPSFNISR------------------------NVQGNHKTNVRA : 356
Pa3D : QNQ-GYPAINIRA------------------------NVQGNHKENVRA : 350
Te3A : RDRYSVP-INFDSWTLDTYGPEHYAVGQG-QTKINEHVDVRGNHAEIIHE : 390
An3A : RDRLWTP-PNFSSWTRDEYGFKYYYVSEGPYEKVNQFVNVQRNHSELIRR : 394
Tr3B : KTVDSLIDTNFDSWTNGEYGYVQAAVNEN-WEKVNYGVDVRANHANHIRE : 401
Nh3A : KTVEDLPDINFSSWSRDTFGYVQAAAQEN-WEQINFGVDVRHDHSEHIRL : 408
Gz3A : RSVEEEPDINFSAWTRDEYGFVQTYAQEN-REKVNFAVNVQHDHKRHIRE : 408
Fv3C : KTIEDLPDINFSSWTRDTFGFVHTFAQEN-REQVNFGVNVQHDHKSHIRE : 421
Fo3A : KTVEDLPDINFSSWTRDTFGFVQTFAQEN-REQVNFGVNVQHDHKNHIRE : 421
Pa3G : KSIELDP-INFSFWSLDTYGPIHWAAGEG-HQQINYHVDVRADHANLIRE : 402
Vd3A : KTLDDVPDINFSSWTKDTIGPLHWAAQDN-VQVINQHVDVRQDHGALIRT : 409
```

图43A-3

```
                460         *       480         *       500
Tn3B : AGAEGVVLLRNEE-ALPLS-ENSKIALFG-----TGQ------------ : 360
Fv3G : GAVEGHVLVKNTNNALPFKPNMKLVSLFGYSHKAPDKNIPDPAQGMFSAW : 416
Fv3D : HGAESTVLLKNEKNALPLK-KPKSIAVFG-----NDA--GDITEGFYNQN : 459
Tr3A : IARDGIVLLKNDANILPLK-KPASIAVVG-----SAAIIGNHARNSPSCN : 400
Pa3D : VARDGIVLLKNDG-ILPLS-KPRKIAVVG-----SHS--VNNPQGINACV : 391
Te3A : IGAASAVLLKNKG-GLPLTGTERFVGVFG-----KDA--GSNPWGVNGCS : 432
An3A : IGADSTVLLKNDG-ALPLTGKERLVALIG-----EDA--GSNPYGANGCS : 436
Tr3B : VGAKGTVIFKNNG-ILPLK-KPKFLTVIG-----EDA--GGNPAGPNGCG : 442
Nh3A : SAAKGTVLLKNSG-SLPLK-KPKFLAVVG-----EDA--GPNPAGPNGCN : 449
Gz3A : AGAKGSVVLKNTG-SLPLK-KPQFLAVIG-----EDA--GSNPAGPNGCA : 449
Fv3C : AAAKGSVVLKNTG-SLPLK-NPKFLAVIG-----EDA--GPNPAGPNGCG : 462
Fo3A : SAAKGSVILKNTG-SLPLN-NPKFLAVIG-----EDA--GPNPAGPNGCG : 462
Pa3G : IAAKGTVLLKNTG-SLPLN-KPKFVAVIG-----EDA--GPNPNGPNSCA : 443
Vd3A : IAARGTVLLKNEG-SLPLN-KPKFVAVIG-----EDA--GPRPVGPNGCP : 450

                *       520         *       540         *
Tn3B : ----------------------------IETIKGGTGSGDTHPRYAISI : 381
Fv3G : SIGAQSANITELNLGFLGNLSLTYSAIAPNGTIISGGGSGASAWTLFSSP : 466
Fv3D : D--YE-----------------------FGTLVAGGGSGTGRLTYLVSP : 483
Tr3A : DKGCD-----------------------DGALGMGWGSGAVNYPYFVAP : 426
Pa3D : DKGCN-----------------------VGTLGMGWGSGSVNYPYLVSP : 417
Te3A : DRGCD-----------------------NGTLAMGWGSGTANFPYLVTP : 458
An3A : DRGCD-----------------------NGTLAMGWGSGTANFPYLVTP : 462
Tr3B : DRGCD-----------------------DGTLAMEWGSGTTNFPYLVTP : 468
Nh3A : DRGCN-----------------------NGTLAMSWGSGTAQFPYLVTP : 475
Gz3A : DRGCD-----------------------NGTLAMAWGSGTSQFPYLVTP : 475
Fv3C : DRGCD-----------------------NGTLAMAWGSGTSQFPYLITP : 488
Fo3A : DRGCD-----------------------NGTLAMAWGSGTSQFPYLITP : 488
Pa3G : DRGCN-----------------------NGTLAMGWGSGTANFPYLITP : 469
Vd3A : DQGCN-----------------------NGTLAAGWGSGTASFPYLITP : 476

                560         *       580         *       600
Tn3B : LEGIKERGLNFDEELAKTYEDYIKKMRETEEYKPRRDSWGTIIKPKLPEN : 431
Fv3G : FDAFVSRAK----------------------------KEGTALF-----W : 483
Fv3D : LAAINARAK----------------------------QDGTLVQQWMNNT : 505
Tr3A : YDAINTRAS----------------------------SQGTQVT--LSNT : 446
Pa3D : YDALRTRAQ----------------------------ADGTQIS--LHNT : 437
Te3A : EQAIQREVL----------------------------SRNGTFTGITDNG : 480
An3A : EQAISNEVL----------------------------KNKNGVFTATDNW : 484
Tr3B : DAALQSQAL----------------------------QDGTRYESILSNY : 490
Nh3A : DSALQNQAV----------------------------LDGTRYESVLRNN : 497
Gz3A : DQGISLQAI----------------------------QDGTRYESILNNN : 497
Fv3C : DQGLSNRAT----------------------------QDGTRYESILTNN : 510
Fo3A : DQGLQNRAA----------------------------QDGTRYESILTNN : 510
Pa3G : DAALQAQAI----------------------------KDGSRYESILTNY : 491
Vd3A : DSALQFQAV----------------------------SDGSRYESILSNW : 498
```

图43A-4

```
                 *       620         *       640         *
Tn3B : FLSEKEIHK--LAKKNDVAVIVISRISGEGY----DRKPVKGDFYLSDDE : 475
Fv3G : DFESWDPY---VNPTSEACIVAGNAWASEGW--------DRPATY-DAYT : 521
Fv3D : LIATTNVTDLWIPATPDVCLVFLKTWAEE--------AADREHLSVDWDG : 547
Tr3A : DNTSSGAS---AARGKDVAIVFITADSGEGYITVEGNAGDRNNLDPWHNG : 493
Pa3D : DSTNGVSN---VVSDADAVVVVITADSGEGYITVEGHAGDRSHLDPWHNG : 484
Te3A : ALAEMAA----AASQADTCLVFANADSGEGYITVDGNEGDRKNLTLWQGA : 526
An3A : AIDQIEA----LAKTASVSLVFVNADSGEGYINVDGNLGDRRNLTLWRNG : 530
Tr3B : AISQTQAL---VSQPDAIAIVFANSDSGEGYINVDGNEGDRKNLTLWKNG : 537
Nh3A : QWEQTRSL---ISQPNVTAIVFANANSGEGYIDVDGNEGDRKNLTLWNEG : 544
Gz3A : QWPQTQAL---VSQPNVTAIVFANADSGEGYIEVDGNYGDRKNLTLWKQG : 544
Fv3C : EWASVQAL---VSQPNVTAIVFANADSGEGYIEVDGNFGDRKNLTLWQQG : 557
Fo3A : EWAQTQAL---VSQPNVTAIVFANADSGEGYIEVDGNFGDRKNLTLWQQG : 557
Pa3G : AASQTRAL---VSQDNVTAIVFVNADSGEGYINFEGNMGDRNNLTLWRGG : 538
Vd3A : DYERTEAL---VSQADATALVFVNANSGEGYISVDGNEGDRKNLTLWNGG : 545

              660         *       680         *       700
Tn3B : TDLIKTVSREFHEQGKKVIVLLNIGSPVEVVSWRDLVDGILLVWQA--GQ : 523
Fv3G : DELINNVA----DKCANTIVVLHNAGTRLVDGFFGHPNVTAIIYAHLPGQ : 567
Fv3D : NDVVESVA----KYCNNTVVVTHSSGINTLP-WADHPNVTAILAAHFPGQ : 592
Tr3A : NALVQAVA----GANSNVIVVVHSVGAIILEQILALPQVKAVVWAGLPSQ : 539
Pa3D : NQLVQAAA----AANKNVIVVVHSVGQITLETILNTNGVRAIVWAGLPGQ : 530
Te3A : DQVIHNVS----ANCNNTVVVLHTVGPVLIDDWYDHPNVTAILWAGLPGQ : 572
An3A : DNVIKAAA----SNCNNTIVIIHSVGPVLVNEWYDNPNVTAILWGGLPGQ : 576
Tr3B : DDLIKTVA----AVNPKTIVVIHSTGPVILKDYANHPNISAILWAGAPGQ : 583
Nh3A : DDLIKNVS----SICPNTIVVLHTVGPVILTEWYDNPNITAIVWAGVPGQ : 590
Gz3A : DELIKNVS----AICPNTIVVLHTVGPVLLTEWHNNPNITAIVWAGVPGQ : 590
Fv3C : DELIKNVS----SICPNTIVVLHTVGPVLLADYEKNPNITAIVWAGLPGQ : 603
Fo3A : DELIKNVS----SICPNTIVVLHTVGPVLLADYEKNPNITAIVWAGLPGQ : 603
Pa3G : DDLVKNVS----SWCSNTIVVIHSTGPVLISEWYDSPNITAILWAGLPGQ : 584
Vd3A : DELIQRVA----AANNNTIVIIHSVGPVLVTDWYENPNITAIIWAGLPGQ : 591

                 *       720         *       740         *
Tn3B : ETGRIVADVLTGRINPSGKLPTTFPRDYSD------VPSWTFPG-EPKDN : 566
Fv3G : DSGDALVSLLYGDENPSGRLPYTVARNETDYGHLLKPDLTLAPN-QYQHF : 616
Fv3D : ESGNSLVDLLYGDVNPSGRLPYTIAFNGTDYN----APPTTAVNTTGKED : 638
Tr3A : ESGNALVDVLWGDVSPSGKLVYTIAKSPNDYN----TRIVS----GGSD- : 580
Pa3D : ENGNALVDVLYGLVSPSGKLPYTIGKRESDYG----TAVV-----RGDD- : 570
Te3A : ESGNSLVDVLYGRVNP-GKTPFTWGRARDDYG----APLIVKPN-NGKGA : 616
An3A : ESGNSLADVLYGRVNPGAKSPFTWGKTREAYQ----DYLYTEPN-NGNGA : 621
Tr3B : ESGNSLVDILYGKQSP-GRTPFTWGPSLESYG----VSVMTTPN-NGNGA : 627
Nh3A : ESGNALVDILYGKTSP-GRSPFTWGRTRKSYG----TDVLYEPN-NGQGA : 634
Gz3A : ESGNAIADILYGKTSP-GRSPFTWGRTYDSYG----TKVLYKAN-NGEGA : 634
Fv3C : ESGNAIADLLYGKVSP-GRSPFTWGRTRESYG----TEVLYEAN-NGRGA : 647
Fo3A : ESGNAIADLLYGKVSP-GRSPFTWGRTRESYG----TEVLYEAN-NGRGA : 647
Pa3G : ESGNSITDVLYGKVNPSGKSPFTWGATREGYG----ADVLYTPN-NGEGA : 629
Vd3A : ESGNSIADILYGRVNPGGKTPFTWGPTVESYG----VDVLREPN-NGNGA : 636
```

图43A-5

```
                760         *       780         *       800
Tn3B : PQKVVYEEDIYVGYRYYDT---------------FGVEPAYEFGYGLSYT : 601
Fv3G : PQS-DFSEGIFIDYRHFDA---------------KNITPRFEFGFGLSYT : 650
Fv3D : WQS-WFDEKLEIDYRYFDA---------------HNISVRYEFGFGLSYS : 672
Tr3A : ----SFSEGLFIDYKHFDD---------------ANITPRYEFGYGLSYT : 611
Pa3D : ----NFREGLFVDYRHFDN---------------ARIEPRYEFGFGLSYT : 601
Te3A : PQQ-DFTEGIFIDYRRFDK---------------YNITPIYEFGFGLSYT : 650
An3A : PQE-DFVEGVFIDYRGFDK---------------RNETPIYEFGYGLSYT : 655
Tr3B : PQD-NFNEGAFIDYRYFDKVAPGKPRS-------SDKAPTYEFGFGLSWS : 669
Nh3A : PQD-DFTEGVFIDYRHFDQVSPSTDGSKSN----DESSPIYEFGHGLSWT : 679
Gz3A : PQE-DFVEGNFIDYRHFDRQSPSTNGKSATN---DSSAPLYEFGFGLSWT : 680
Fv3C : PQD-DFSEGVFIDYRHFDRRSPSTDGKSSPN---NTAAPLYEFGHGLSWS : 693
Fo3A : PQD-DFSEGVFIDYRHFDRRSPSTDGKSAPN---NTAAPLYEFGHGLSWT : 693
Pa3G : PQQ-DFSEGVFIDYRYFDK---------------ANTSVIYEFGHGLSYT : 663
Vd3A : PQS-DFDEGVFIDYRWFDRQSGVDNNASAPRNSSSSHAPIFEFGYGLSYT : 685

                  *       820         *       840         *
Tn3B : TFEYSDLNVSF------------------------------------------ : 612
Fv3G : TFEYASLQISK------------------------------------------ : 661
Fv3D : TFEISDISAEP------------------------------------------ : 683
Tr3A : KFNYSRLSVLSTAKS--------------G------------------------ : 627
Pa3D : NFTFSDIKITSNVKP--------------G------------------------ : 617
Te3A : TFEFSQLNVQPINAPPYTPASGFTKAAQSFG--QPSNASDNLYPSD-IER : 697
An3A : TFNYSNLQVEVLSAPAYEPASGETEAAPTFG--EVGNASDYLYPDG-LQR : 702
Tr3B : TFKFSNLHIQKNNVGPMSPPNGKTIAAPSLG-SFSKNLKDYGFPKN-VRR : 717
Nh3A : TFEYSELNIQAHNKIPFDPPIGETIAAPVLG-NYSTDLADYTFPDG-IRY : 727
Gz3A : TFEYSDLKVESVSNASYSPSVGNTIPAPTYG-NFSKNLDDYTFPSG-VRY : 728
Fv3C : TFEYSDLNIQKNVENPYSPPAGQTIPAPTFG-NFSKNLNDYVFPKG-VRY : 741
Fo3A : TFEYSDLNIQKNVNSTYSPPAGQTIPAPTFG-NFSKNLNDYVFPKG-VRY : 741
Pa3G : TFEYSNIQVTKKNAGPYKPTTGQTAPAPTFG-NFSTDLSDYLFPDEEFPY : 712
Vd3A : TFEFSNLQIERHDVHDYVPTTGQTSPAPRFGANYSTNYDDYVFPEGEIRY : 735

                860         *       880         *       900
Tn3B : ----------------------------------------DGET------- : 616
Fv3G : ----------------------------------------SQAQTPEYPAG : 672
Fv3D : ----------------------------------------LASDITSQPED : 694
Tr3A : ------------------------------------------------PAT : 630
Pa3D : ------------------------------------------------PAT : 620
Te3A : VPLYIYPWLNSTDL-KASAND--PDYGLPTEKYVPPNATNGDPQPIDPAG : 744
An3A : ITKFIYPWLNSTDL-EASSGD--ASYGQDASDYLPEGATDGSAQPILPAG : 749
Tr3B : IKEFIYPYLSTTTSGKEASGD--AHYGQTAKEFLPAGALDGSPQPRSAAS : 765
Nh3A : IYQFIYPWLNTSSSGREASGD--PDYGKTAEEFLPPGALDGSAQPRPPSS : 775
Gz3A : LYKFIYPYLNTSSSAEKASGDVKGRFGETGDEFLPPNALNGSSQPRLPSS : 778
Fv3C : IYKFIYPFLNTSSSASEASND-GGQFGKTAEEFLPPNALNGSAQPRLPAS : 790
Fo3A : IYKFIYPFLNTSSSASEASND-GGQFGKTAEEFLPPNALNGSAQPRLPSS : 790
Pa3G : VYQYIYPYLNTTDP-RNASGD--PHFGQTAEEFMPPHAIDDSPQPLLPSS : 759
Vd3A : IYQHIYPYLNSSDP-KEALAD--PKYGQTAEEFLPEGALDASPQPRLPAS : 782
```

图43A-6

```
                 *       920         *       940         *
Tn3B : ----------------LRVQYRIENTGGRAGKEVSQVYIKAPKGKI-DKPF : 650
Fv3G : A-LTEGGRSDLWDVVATVTASVRNTGSVDGKEVAQLYV----GVP-GGPM : 716
Fv3D : LPVQPGGNPALWETVYNVTVSVSNTGKVDGATVPQLYVTFPDSAPAGTPP : 744
Tr3A : GAVVPGGPSDLFQNVATVTVDIANSGQVTGAEVAQLYITYPSSAP-RTPP : 679
Pa3D : GQTIPGGPADLWEDVATVTATITNSGAVEGAEVAQLYIGLPSSAP-ASPP : 669
Te3A : G--APGGNPSLYEPVARVTTIITNTGKVTGDEVPQLYVSL--GGP-DDAP : 789
An3A : G--GAGGNPRLYDELIRVSVTIKNTGKVAGDEVPQLYVSL--GGP-NEPK : 794
Tr3B : G--EPGGNRQLYDILYTVTATITNTGSVMDDAVPQLYLSH--GGP-NEPP : 810
Nh3A : G--APGGNPHLWDVLYTVSAIITNTGNATSDEIPQLYVSL--GGE-NEPV : 820
Gz3A : G--APGGNPQLWDIMYTVTATITNTGDATSDEVPQLYVSL--GGE-GEPV : 823
Fv3C : G--APGGNPQLWDILYTVTATITNTGNATSDEIPQLYVSL--GGE-NEPI : 835
Fo3A : G--APGGNPQLWDILYTVTATITNTGNATSDEIPQLYVSL--GGE-NEPV : 835
Pa3G : GKNSPGGNRALYDILYEVTADITNTGEIVGDEVVQLYVSL--GGP-DDPK : 806
Vd3A : G--GPGGNPMLWDVIFTVTATVTNTGKVAGDEVAQLYVSL--GGP-DDPI : 827

              960         *       980         *      1000
Tn3B : QELKAFHKTRLLNPGESEEVVLEIPVRDLASFN--GEEWVVEAGEYEVRV : 698
Fv3G : RQLRGFTKP-AIKAGETATVTFELTRRDLSVWDVNAQEWQLQQGNYAIYV : 765
Fv3D : KQLRGFDKV-FLEAGESKSVSFELMRRDLSYWDIISQKWLIPEGEFTIRV : 793
Tr3A : KQLRGFAKL-NLTPGQSGTATFNIRRRDLSYWDTASQKWVVPSGSFGISV : 728
Pa3D : KQLRGFSKL-KLAPGASGTATFNLRRRDLSYWDTRLQNWVVPSGNFVVSV : 718
Te3A : KVLRGFDRI-TLAPGQQYLWTTTLTRRDISNWDPVTQNWVVTNYTKTIYV : 838
An3A : IVLRQFERI-TLQPSKETQWSTTLTRRDLANWNVETQDWEITSYPKMVFA : 843
Tr3B : KVLRGFDRIERIAPGQSVTFKADLTRRDLSNWDTKKQQWVITDYPKTVYV : 860
Nh3A : RVLRGFDRIENIAPGQSVRFTTDITRRDLSNWDVVSQNWVITDYEKTVYV : 870
Gz3A : RVLRGFERLENIAPGESATFTAQLTRRDLSNWDVNVQNWVITDHAKKIWV : 873
Fv3C : RVLRGFDRIENIAPGQSAIFNAQLTRRDLSNWDTNAQNWVITDHPKTVWV : 885
Fo3A : RVLRGFDRIENIAPGQSAIFNAQLTRRDLSNWDVDAQNWVITDHPKTVWV : 885
Pa3G : VVLRDFGKL-RIEPGQTAKFRGLLTRRDLSNWDVVSQDWVISEHTKTVFV : 855
Vd3A : RVLRGFDRI-HIAPGASQTFRAELTRRDLSNWDVVTQNWFISQYEKTVFV : 876

                 *      1020
Tn3B : GASSRNIKLKGTFSVGEERRFKP : 721
Fv3G : GRSSRDLPLQSTLSI-------- : 780
Fv3D : GFSSRDLKEETKVTVVEA----- : 811
Tr3A : GASSRDIRLTSTLSVA------- : 744
Pa3D : GASSRDIRLTGTITA-------- : 733
Te3A : GNSSRNLPLQAPLKPYPGI---- : 857
An3A : GSSSRKLPLRASLPTVH------ : 860
Tr3B : GSSSRDLPLSARLP--------- : 874
Nh3A : GSSSRNLPLKATLK--------- : 884
Gz3A : GSSSRNLPLSADL---------- : 886
Fv3C : GSSSRKLPLSAKLE--------- : 899
Fo3A : GSSSRKLPLSAKLE--------- : 899
Pa3G : GKSSRDLGLSAVLE--------- : 869
Vd3A : GSSSRNLPLSTRLE--------- : 890
```

图43A-7

```
                          *         20         *         40         *
Te3A            ------------MRNGLLKVAALAAA--------------SAVNGENLAY :  24
Tr3B            ------MKTLSVFAAALLAAVAEANPYPP--------------PHSNQAY :  30
Fv3C            MKLNWVAAALSIGAAGTDSAVALASAVPDTLAGVKKADAQKVVTRDTLAY :  50
Fv3C/Tr3B       MKLNWVAAALSIGAAGTDSAVALASAVPDTLAGVKKADAQKVVTRDTLAY :  50
Fv3C/Te3A/Tr3B MKLNWVAAALSIGAAGTDSAVALASAVPDTLAGVKKADAQKVVTRDTLAY :  50

                         60         *         80         *        100
Te3A            SPPFYPSPWANGQG-DWAEAYQKAVQFVSQLTLAEKVNLTTGTGWEQDRC :  73
Tr3B            SPPFYPSPWMDPSAPGWEQAYAQAKEFVSGLTLLEKVNLTTGVGWMGEKC :  80
Fv3C            SPPHYPSPWMDPNAVGWEEAYAKAKSFVSQLTLMEKVNLTTGVGWQGERC : 100
Fv3C/Tr3B       SPPHYPSPWMDPNAVGWEEAYAKAKSFVSQLTLMEKVNLTTGVGWQGERC : 100
Fv3C/Te3A/Tr3B SPPHYPSPWMDPNAVGWEEAYAKAKSFVSQLTLMEKVNLTTGVGWQGERC : 100

                          *       120         *       140         *
Te3A            VGQVGSIPRLGFPGLCMQDSPLGVRDTDYNSAFPAGVNVAATWDRNLAYR : 123
Tr3B            VGNVGTVPRLGMRSLCMQDGPLGLRFNTYNSAFSVGLTAAASWSRHLWVD : 130
Fv3C            VGNVGSIPRLGMRGLCLQDGPLGIRLSDYNSAFPAGTTAGASWSKSLWYE : 150
Fv3C/Tr3B       VGNVGSIPRLGMRGLCLQDGPLGIRLSDYNSAFPAGTTAGASWSKSLWYE : 150
Fv3C/Te3A/Tr3B VGNVGSIPRLGMRGLCLQDGPLGIRLSDYNSAFPAGTTAGASWSKSLWYE : 150

                        160         *       180         *       200
Te3A            RGVAMGEEHRGKGVDVQLGPVAGPLGRSPDAGRNWEGFAPDPVLTGNMMA : 173
Tr3B            RGTALGSEAKGKGVDVLLGPVAGPLGRNPNGGRNVEGFGSDPYLAGLALA : 180
Fv3C            RGLLMGTEFKEKGIDIALGPATGPLGRTAAGGRNWEGFTVDPYMAGHAMA : 200
Fv3C/Tr3B       RGLLMGTEFKEKGIDIALGPATGPLGRTAAGGRNWEGFTVDPYMAGHAMA : 200
Fv3C/Te3A/Tr3B RGLLMGTEFKEKGIDIALGPATGPLGRTAAGGRNWEGFTVDPYMAGHAMA : 200

                          *       220         *       240         *
Te3A            STIQGIQDAGVIACAKHFILYEQEHFRQGAQ---DGYDISDSISANADDK : 220
Tr3B            DTVTGIQNAGTIACAKHFLLNEQEHFRQVGEANGYGYPITEALSSNVDDK : 230
Fv3C            EAVKGIQDAGVIACAKHYIANEQEHFRQSGEVQSRKYNISESLSSNLDDK : 250
Fv3C/Tr3B       EAVKGIQDAGVIACAKHYIANEQEHFRQSGEVQSRKYNISESLSSNLDDK : 250
Fv3C/Te3A/Tr3B EAVKGIQDAGVIACAKHYIANEQEHFRQSGEVQSRKYNISESLSSNLDDK : 250

                        260         *       280         *       300
Te3A            TMHELYLWPFADAVRAGVGSVMCSYNQVNNSYACSNSYTMNKLLKSELGF : 270
Tr3B            TIHEVYGWPFQDAVKAGVGSFMCSYNQVNNSYACQNSKLINGLLKEEYGF : 280
Fv3C            TMHELYAWPFADAVRAGVGSVMCSYNQINNSYGCQNSKLLNGILKDEMGF : 300
Fv3C/Tr3B       TMHELYAWPFADAVRAGVGSVMCSYNQINNSYGCQNSKLLNGILKDEMGF : 300
Fv3C/Te3A/Tr3B TMHELYAWPFADAVRAGVGSVMCSYNQINNSYGCQNSKLLNGILKDEMGF : 300

                          *       320         *       340         *
Te3A            QGFVMTDWGGHHSGVGSALAGLDMSMPGDIAFDSGTSFWGTNLTVAVLNG : 320
Tr3B            QGFVMSDWQAQHTGVASAVAGLDMTMPGDTAFNTGASYFGSNLTLAVLNG : 330
Fv3C            QGFVMSDWAAQHTGAASAVAGLDMSMPGDTAFDSGYSFWGGNLTLAVING : 350
Fv3C/Tr3B       QGFVMSDWAAQHTGAASAVAGLDMSMPGDTAFDSGYSFWGGNLTLAVING : 350
Fv3C/Te3A/Tr3B QGFVMSDWAAQHTGAASAVAGLDMSMPGDTAFDSGYSFWGGNLTLAVING : 350
```

图43B-1

```
                      360         *       380         *       400
Te3A            SIPEWRVDDMAVRIMSAYYKVGRDRYSVP-INFDSWTLDTYGPEHYAVGQ : 369
Tr3B            TVPEWRIDDMVMRIMAPFFKVGKTVDSLIDTNFDSWTNGEYGYVQAAVNE : 380
Fv3C            TVPAWRVDDMALRIMSAFFKVGKTIEDLPDINFSSWTRDTFGFVHTFAQE : 400
Fv3C/Tr3B       TVPAWRVDDMALRIMSAFFKVGKTIEDLPDINFSSWTRDTFGFVHTFAQE : 400
Fv3C/Te3A/Tr3B TVPAWRVDDMALRIMSAFFKVGKTIEDLPDINFSSWTRDTFGFVHTFAQE : 400

                       *       420         *       440         *
Te3A            GQTKINEHVDVRGNHAEIIHEIGAASAVLLKNKGGLPLTGTERFVGVFGK : 419
Tr3B            NWEKVNYGVDVRANHANHIREVGAKGTVIFKNNGILPLK-KPKFLTVIGE : 429
Fv3C            NREQVNFGVNVQHDHKSHIREAAAKGSVVLKNTGSLPLK-NPKFLAVIGE : 449
Fv3C/Tr3B       NREQVNFGVNVQHDHKSHIREAAAKGSVVLKNTGSLPLK-NPKFLAVIGE : 449
Fv3C/Te3A/Tr3B NREQVNFGVNVQHDHKSHIREAAAKGSVVLKNTGSLPLK-NPKFLAVIGE : 449

                      460         *       480         *       500
Te3A            DAGSNPWGVNGCSDRGCDNGTLAMGWGSGTANFPYLVTPEQAIQREVLSR : 469
Tr3B            DAGGNPAGPNGCGDRGCDDGTLAMEWGSGTTNFPYLVTPDAALQSQALQD : 479
Fv3C            DAGPNPAGPNGCGDRGCDNGTLAMAWGSGTSQFPYLITPDQGLSNRATQD : 499
Fv3C/Tr3B       DAGPNPAGPNGCGDRGCDNGTLAMAWGSGTSQFPYLITPDQGLSNRATQD : 499
Fv3C/Te3A/Tr3B DAGPNPAGPNGCGDRGCDNGTLAMAWGSGTSQFPYLITPDQGLSNRATQD : 499

                       *       520         *       540         *
Te3A            NGTFTGITDNGALAEMAAAASQAD-TCLVFANADSGEGYITVDGNEGDRK : 518
Tr3B            GTRYESILSNYAISQTQALVSQPDAIAIVFANSDSGEGYINVDGNEGDRK : 529
Fv3C            GTRYESILTNNEWASVQALVSQPNVTAIVFANADSGEGYIEVDGNFGDRK : 549
Fv3C/Tr3B       GTRYESILTNNEWASVQALVSQPNVTAIVFANADSGEGYIEVDGNFGDRK : 549
Fv3C/Te3A/Tr3B GTRYESILTNNEWASVQALVSQPNVTAIVFANADSGEGYIEVDGNFGDRK : 549

                      560         *       580         *       600
Te3A            NLTLWQGADQVIHNVSANCNNTVVVLHTVGPVLIDDWYDHPNVTAILWAG : 568
Tr3B            NLTLWKNGDDLIKTVAAVNPKTIVVIHSTGPVILKDYANHPNISAILWAG : 579
Fv3C            NLTLWQQGDELIKNVSSICPNTIVVLHTVGPVLLADYEKNPNITAIVWAG : 599
Fv3C/Tr3B       NLTLWQQGDELIKNVSSICPNTIVVLHTVGPVLLADYEKNPNITAIVWAG : 599
Fv3C/Te3A/Tr3B NLTLWQQGDELIKNVSSICPNTIVVLHTVGPVLLADYEKNPNITAIVWAG : 599

                       *       620         *       640         *
Te3A            LPGQESGNSLVDVLYGRVNPGKTPFTWGRARDDYGAPLIVKPNNGKGAPQ : 618
Tr3B            APGQESGNSLVDILYGKQSPGRTPFTWGPSLESYGVSVMTTPNNGNGAPQ : 629
Fv3C            LPGQESGNAIADLLYGKVSPGRSPFTWGRTRESYGTEVLYEANNGRGAPQ : 649
Fv3C/Tr3B       LPGQESGNAIADLLYGKVSPGRSPFTWGRTRESYGTEVLYEANNGRGAPQ : 649
Fv3C/Te3A/Tr3B LPGQESGNAIADLLYGKVSPGRSPFTWGRTRESYGTEVLYEANNGRGAPQ : 649

                      660         *       680         *       700
Te3A            QDFTEGIFIDYRRFDKYNIT----------PIYEFGFGLSYTTFEFSQ : 656
Tr3B            DNFNEGAFIDYRYFDKVAPG----KPRSSDKAPTYEFGFGLSWSTFKFSN : 675
Fv3C            DDFSEGVFIDYRHFDRRSPSTDGKSSPNNTAAPLYEFGHGLSWSTFEYSD : 699
Fv3C/Tr3B       DDFSEGVFIDYRHFDRRSPSTDGKSSPNNTAAPLYEFGHGLSWSTFKFSN : 699
Fv3C/Te3A/Tr3B DDFSEGVFIDYRHFDKYNIT----------PIYEFGHGLSWSTFKFSN : 687
```

图43B-2

```
                        *       720         *       740         *
Te3A            LNVQPINAPPYTPASGFTKAAQSFGQ-PSNASDNLYPSDIERVPLYIYPW : 705
Tr3B            LHIQKNNVGPMSPPNGKTIAAPSLGSFSKNLKDYGFPKNVRRIKEFIYPY : 725
Fv3C            LNIQKNVENPYSPPAGQTIPAPTFGNFSKNLNDYVFPKGVRYIYKFIYPF : 749
Fv3C/Tr3B       LHIQKNNVGPMSPPNGKTIAAPSLGSFSKNLKDYGFPKNVRRIKEFIYPY : 749
Fv3C/Te3A/Tr3B LHIQKNNVGPMSPPNGKTIAAPSLGNFSKNLKDYGFPKNVRRIKEFIYPY : 737

                      760         *       780         *       800
Te3A            LNSTDL-KASAND-PDYGLPTEKYVPPNATNGDPQPIDPAGGAPGGNPSL : 753
Tr3B            LSTTTSGKEASGD-AHYGQTAKEFLPAGALDGSPQPRSAASGEPGGNRQL : 774
Fv3C            LNTSSSASEASNDGGQFGKTAEEFLPPNALNGSAQPRLPASGAPGGNPQL : 799
Fv3C/Tr3B       LSTTTSGKEASGD-AHYGQTAKEFLPAGALDGSPQPRSAASGEPGGNRQL : 798
Fv3C/Te3A/Tr3B LNTTTSGKEASGD-AHYGQTAKEFLPAGALDGSPQPRSAASGEPGGNRQL : 786

                        *       820         *       840         *
Te3A            YEPVARVTTIITNTGKVTGDEVPQLYVSLGGPDDAPKVLRGFDRI-TLAP : 802
Tr3B            YDILYTVTATITNTGSVMDDAVPQLYLSHGGPNEPPKVLRGFDRIERIAP : 824
Fv3C            WDILYTVTATITNTGNATSDEIPQLYVSLGGENEPIRVLRGFDRIENIAP : 849
Fv3C/Tr3B       YDILYTVTATITNTGSVMDDAVPQLYLSHGGPNEPPKVLRGFDRIERIAP : 848
Fv3C/Te3A/Tr3B YDILYTVTATITNTGSVMDDAVPQLYLSHGGPNEPPKVLRGFDRIERIAP : 836

                      860         *       880         *       900
Te3A            GQQYLWTTTLTRRDISNWDPVTQNWVVTNYTKTIYVGNSSRNLPLQAPLK : 852
Tr3B            GQSVTFKADLTRRDLSNWDTKKQQWVITDYPKTVYVGSSSRDLPLSARLP : 874
Fv3C            GQSAIFNAQLTRRDLSNWDTNAQNWVITDHPKTVWVGSSSRKLPLSAKLE : 899
Fv3C/Tr3B       GQSVTFKADLTRRDLSNWDTKKQQWVITDYPKTVYVGSSSRDLPLSARLP : 898
Fv3C/Te3A/Tr3B GQSVTFKADLTRRDLSNWDTKKQQWVITDYPKTVYVGSSSRDLPLSARLP : 886

Te3A            PYPGI : 857
Tr3B            ----- :   -
Fv3C            ----- :   -
Fv3C/Tr3B       ----- :   -
Fv3C/Te3A/Tr3B ----- :   -
```

Te3A: 总是完整产生
Tr3B: 总是完整产生
Fv3C: 在比对结果中的位置760周围剪切
Fv3C/Tr3B: 在比对结果中的位置680周围剪切
Fv3C/Te3A/Tr3B: 总是完整产生

图43B-3

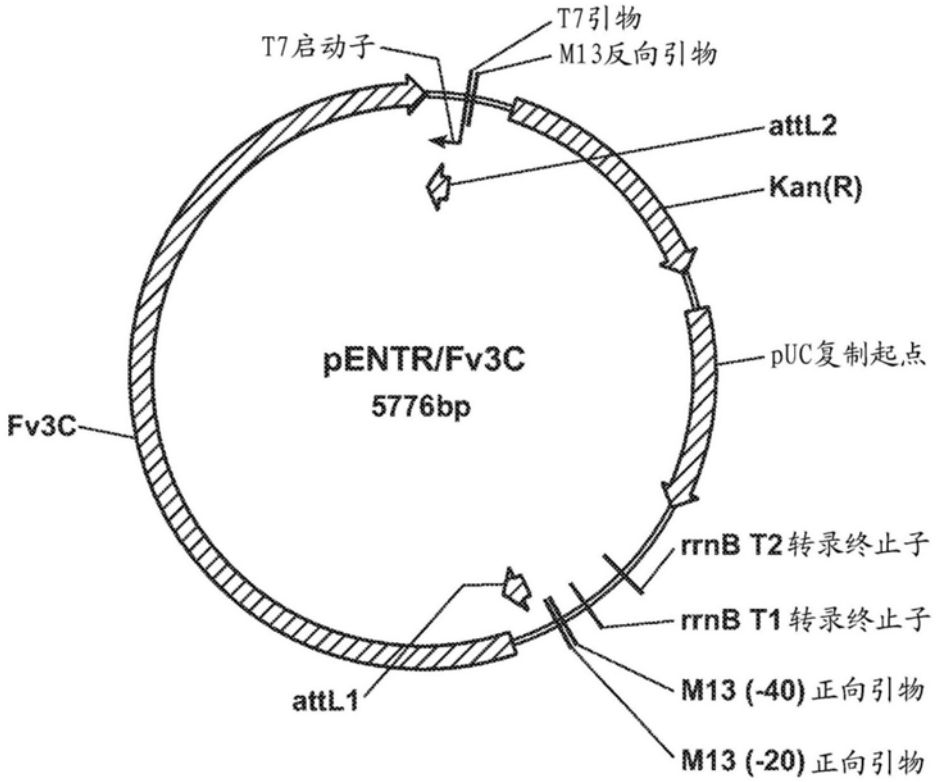

T7引物
T7启动子
M13反向引物
attL2
Kan(R)
pENTR/Fv3C
5776bp
pUC复制起点
Fv3C
rrnB T2转录终止子
rrnB T1转录终止子
attL1
M13 (-40)正向引物
M13 (-20)正向引物

图44

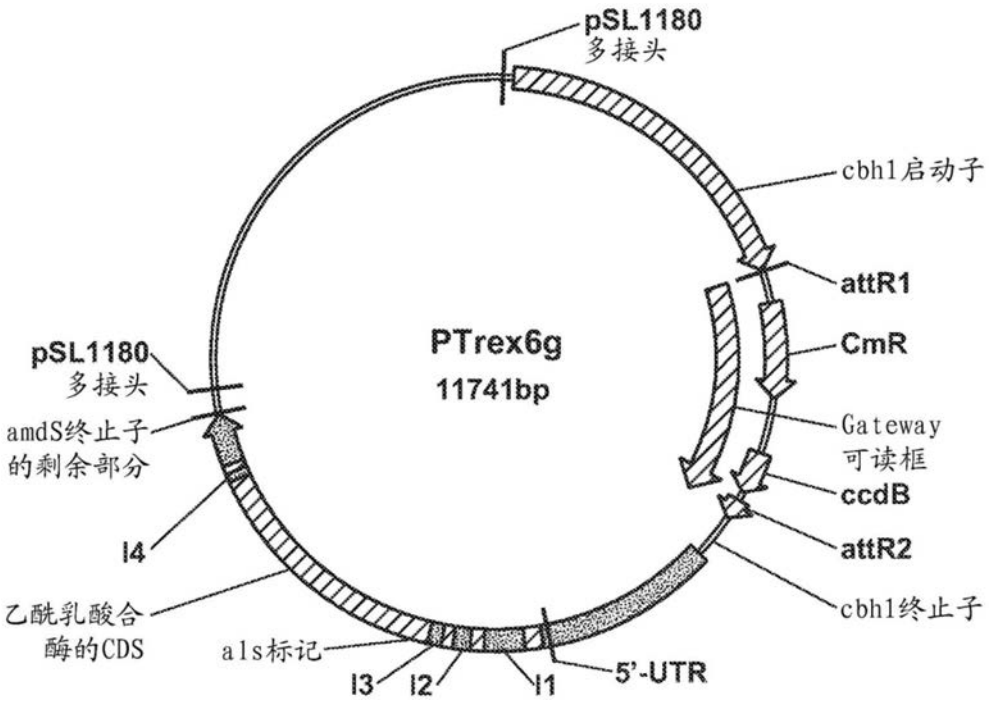

pSL1180
多接头
cbh1启动子
attR1
PTrex6g
11741bp
CmR
pSL1180
多接头
Gateway
可读框
amdS终止子
的剩余部分
ccdB
attR2
I4
cbh1终止子
乙酰乳酸合
酶的CDS
als标记
5'-UTR
I3
I2
I1

图45A

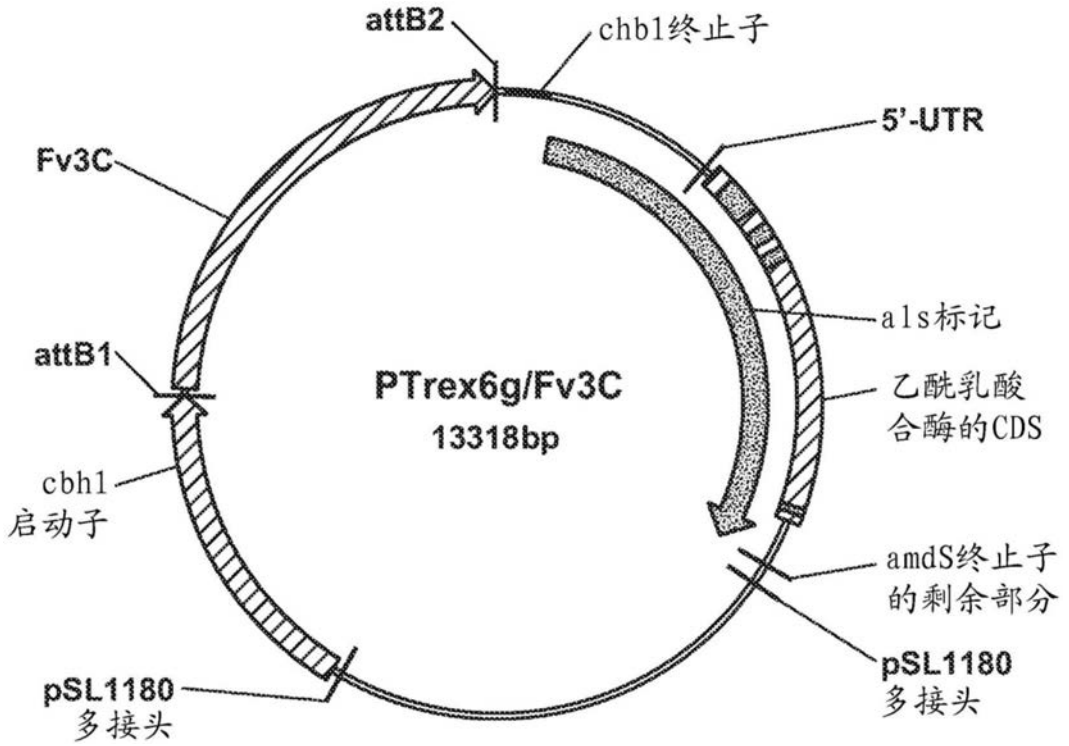


图45B

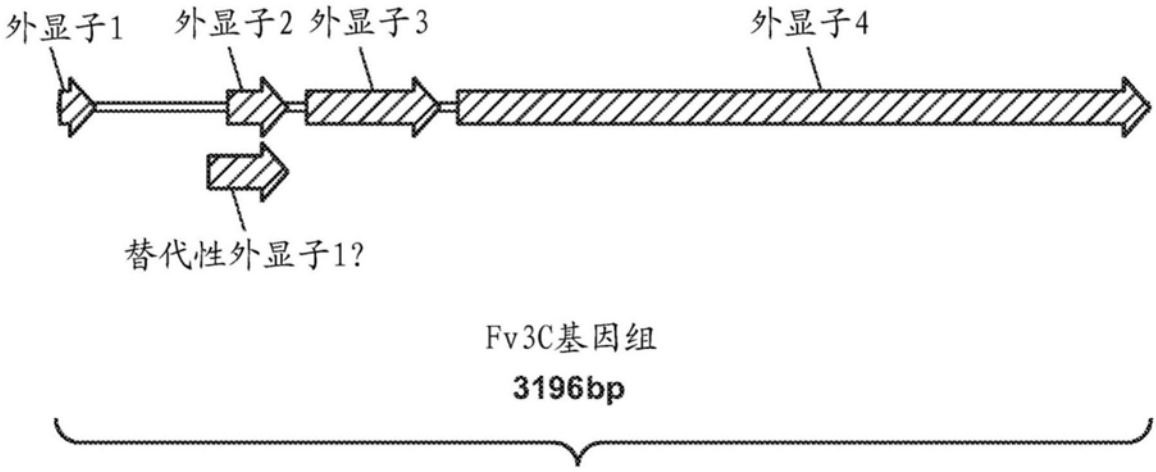


图46A

图46A

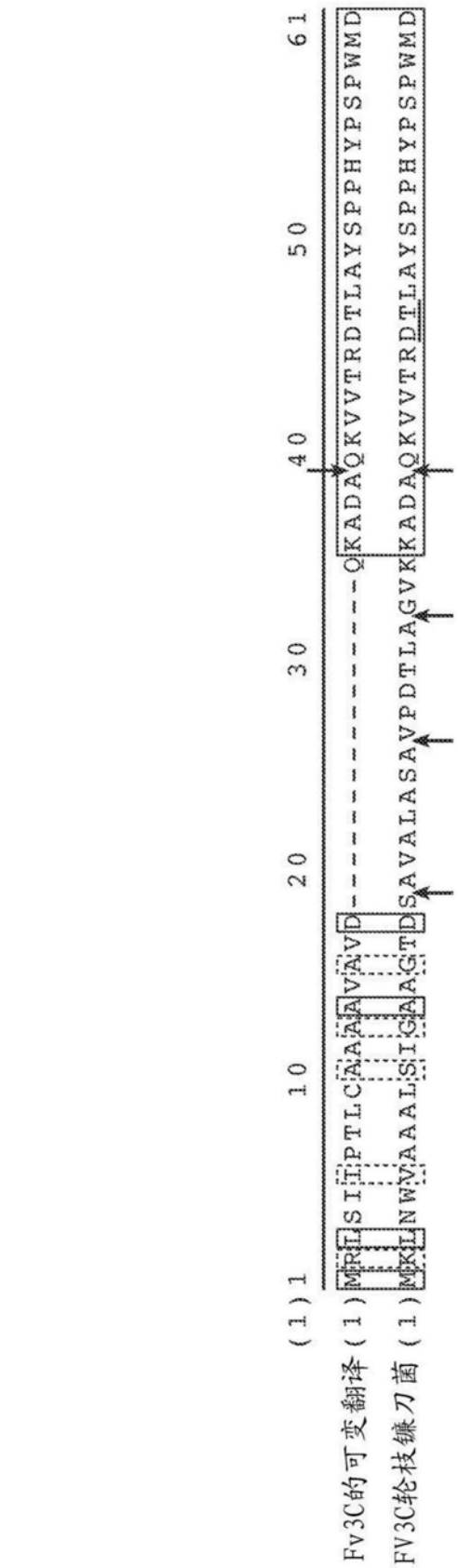

(1)1
10
20
30
40
50
61
Fv3C的可变翻译(1) MRLSIIPTLCAAAAVAVD----------QKADAQKVVTRDTLAYSPPHYPSPWMD
FV3C轮枝镰刀菌(1) MKLNWVAAALSIGAAGTDSAVALASAVPDTLAGVKKADAQKVVTRDTLAYSPPHYPSPWMD

图46B

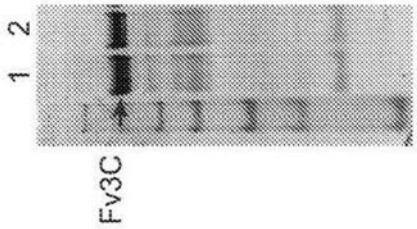

2
1
Fv3C

图46C

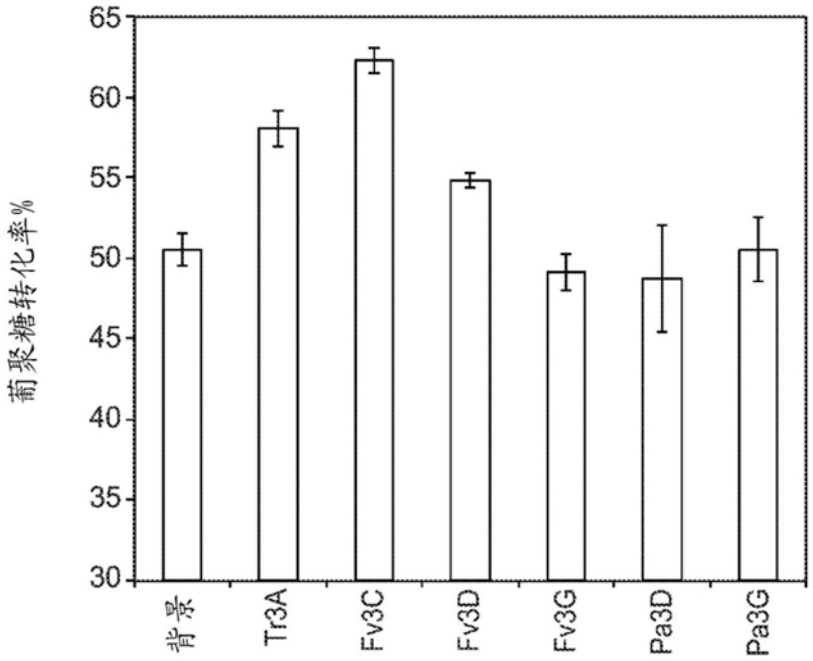

葡聚糖转化率%
65
60
55
50
45
40
35
30
背景
Tr3A
Fv3C
Fv3D
Fv3G
Pa3D
Pa3G

图47

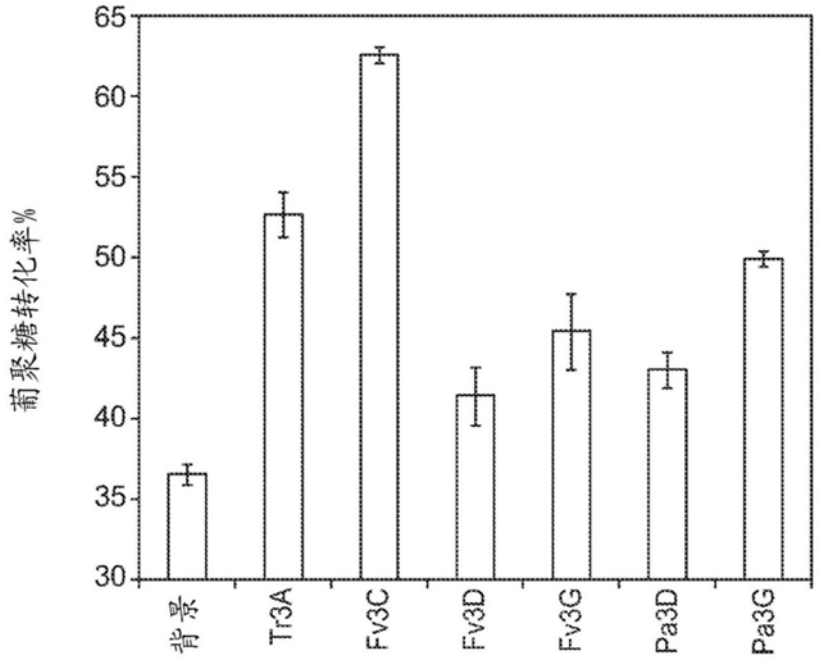

葡聚糖转化率%
65
60
55
50
45
40
35
30
背景
Tr3A
Fv3C
Fv3D
Fv3G
Pa3D
Pa3G

图48

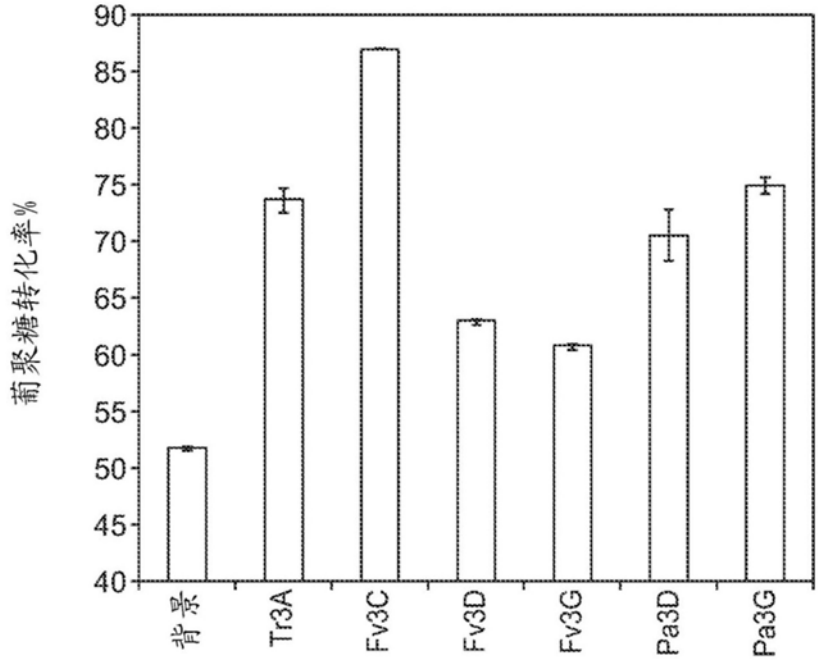

葡聚糖转化率%
90
85
80
75
70
65
60
55
50
45
40
背景
Tr3A
Fv3C
Fv3D
Fv3G
Pa3D
Pa3G

图49

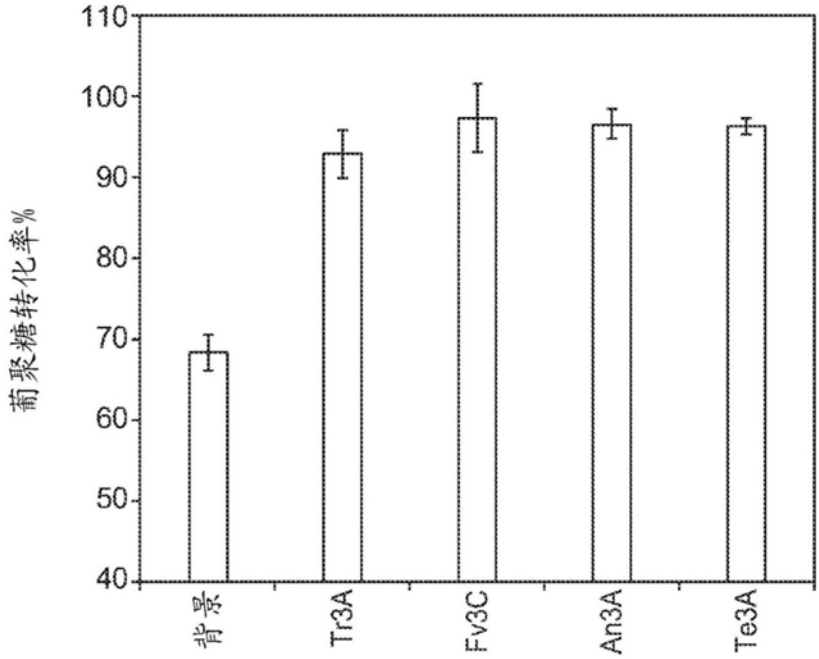

葡聚糖转化率%
110
100
90
80
70
60
50
40
背景
Tr3A
Fv3C
An3A
Te3A

图50

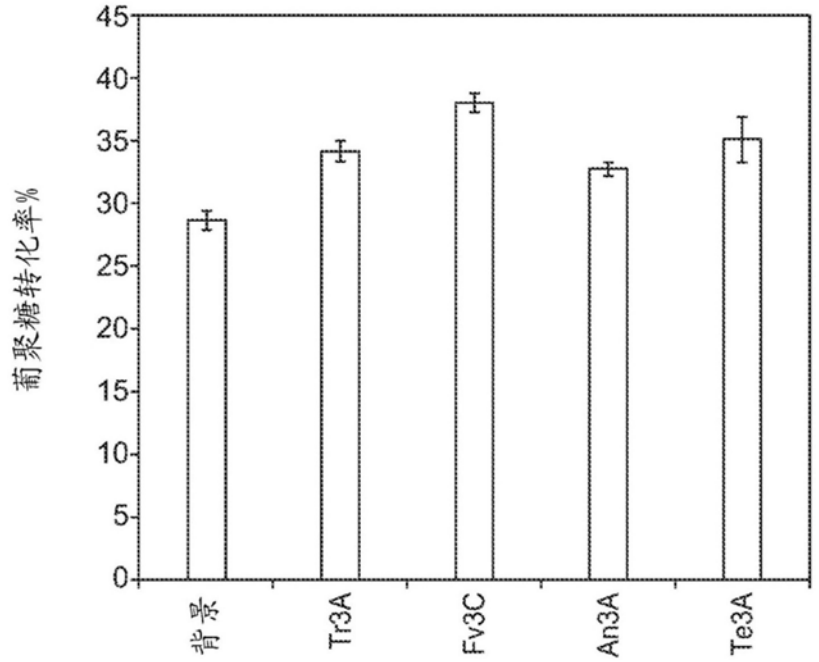

葡聚糖转化率%
45
40
35
30
25
20
15
10
5
0
背景
Tr3A
Fv3C
An3A
Te3A

图51

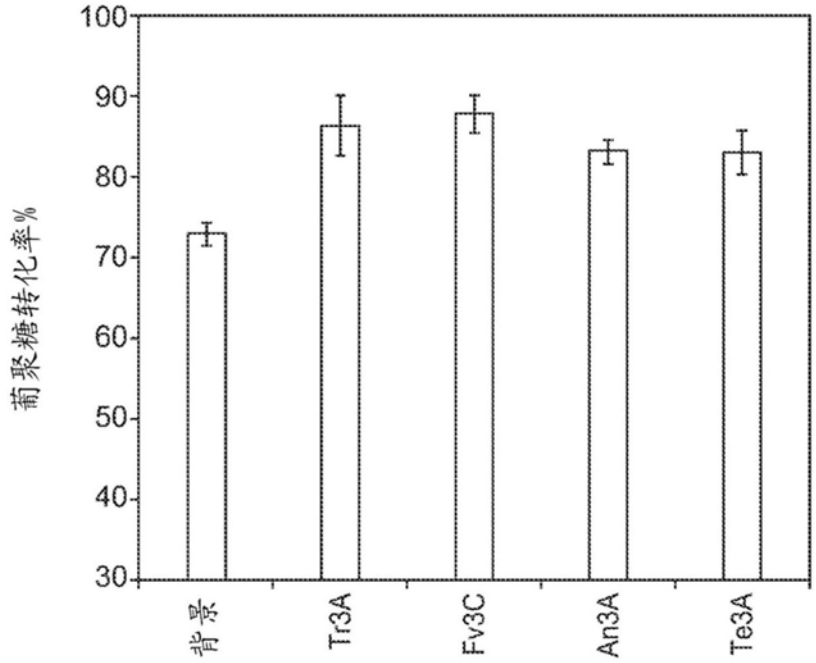

葡聚糖转化率%
100
90
80
70
60
50
40
30
背景
Tr3A
Fv3C
An3A
Te3A

图52

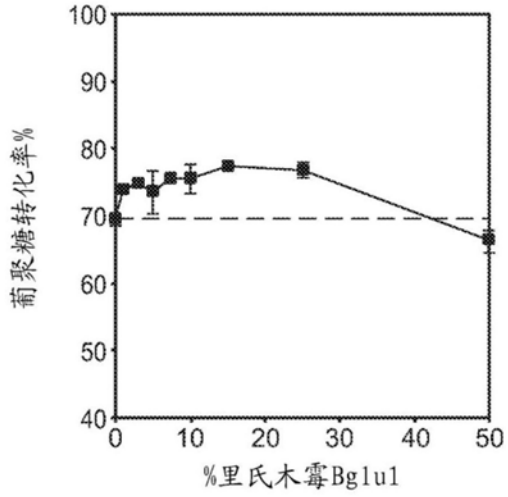

葡聚糖转化率%
100
90
80
70
60
50
40
0
10
20
30
40
50
%里氏木霉Bglu1

图53A

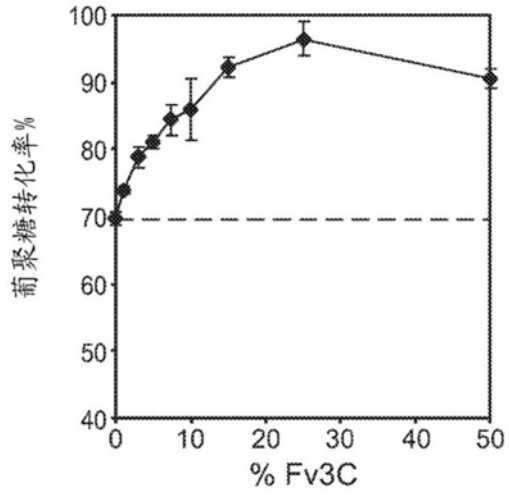

葡聚糖转化率%
100
90
80
70
60
50
40
0
10
20
30
40
50
% Fv3C

图53B

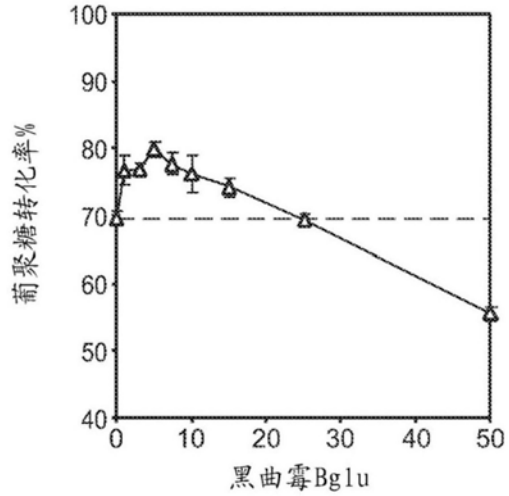

葡聚糖转化率%
100
90
80
70
60
50
40
0
10
20
30
40
50
黑曲霉Bglu

图53C

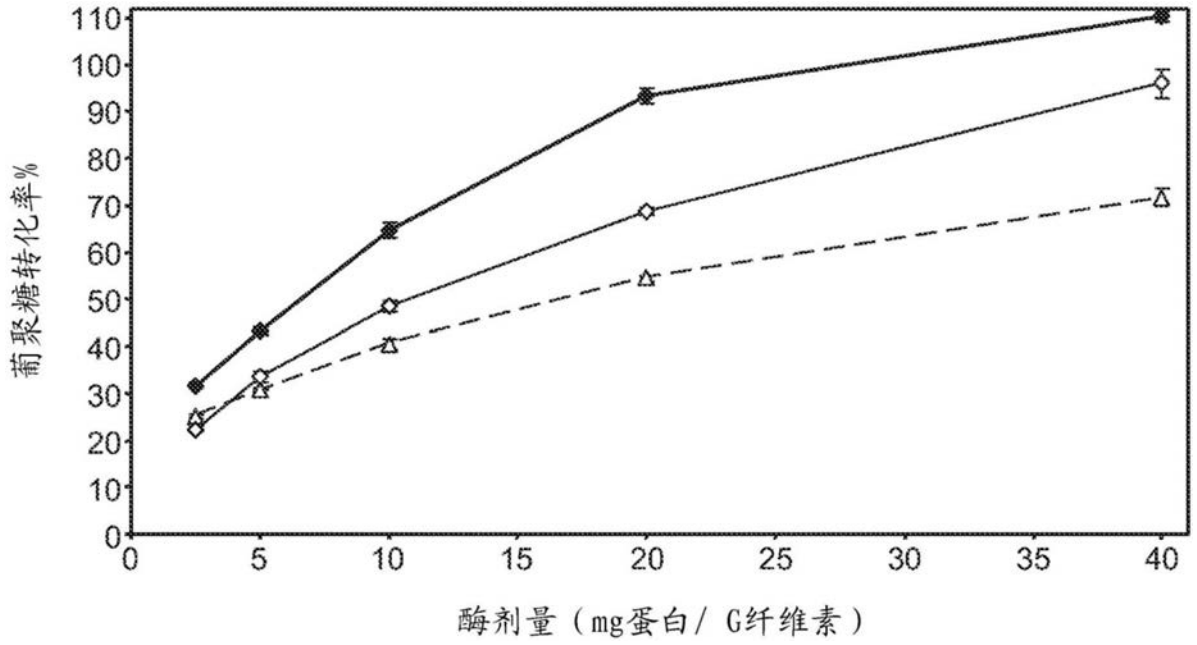

葡聚糖转化率%
110
100
90
80
70
60
50
40
30
20
10
0
0
5
10
15
20
25
30
35
40
酶剂量（mg蛋白/ G纤维素）

图54

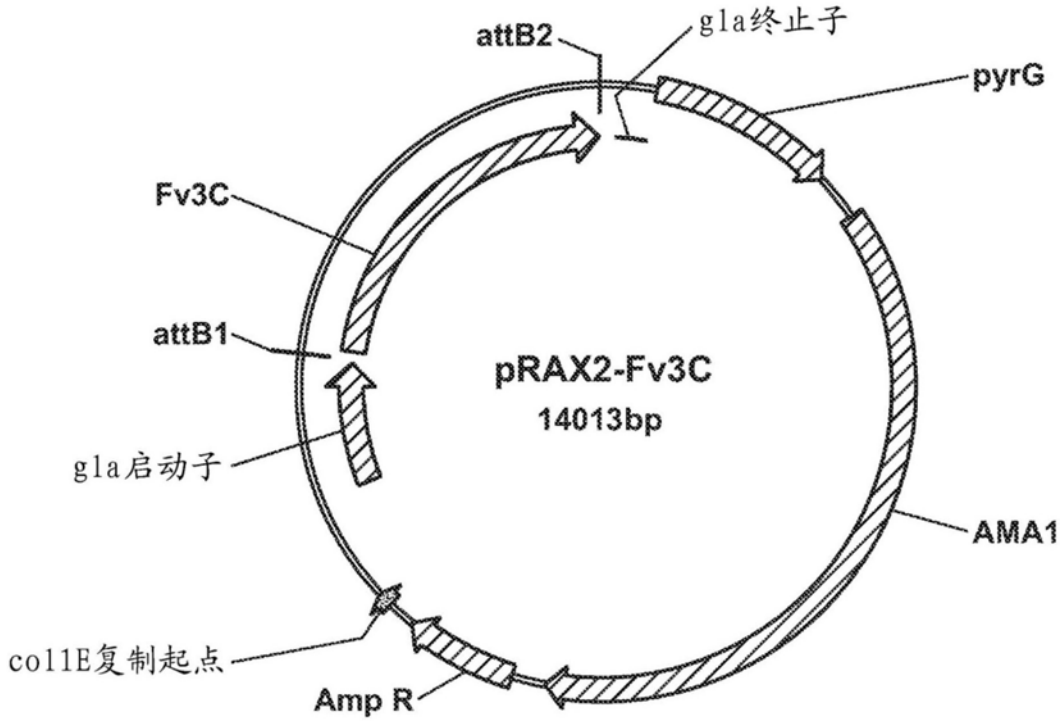

attB2
gla终止子
pyrG
Fv3C
attB1
pRAX2-Fv3C
14013bp
gla启动子
AMA1
col1E复制起点
Amp R

图55A

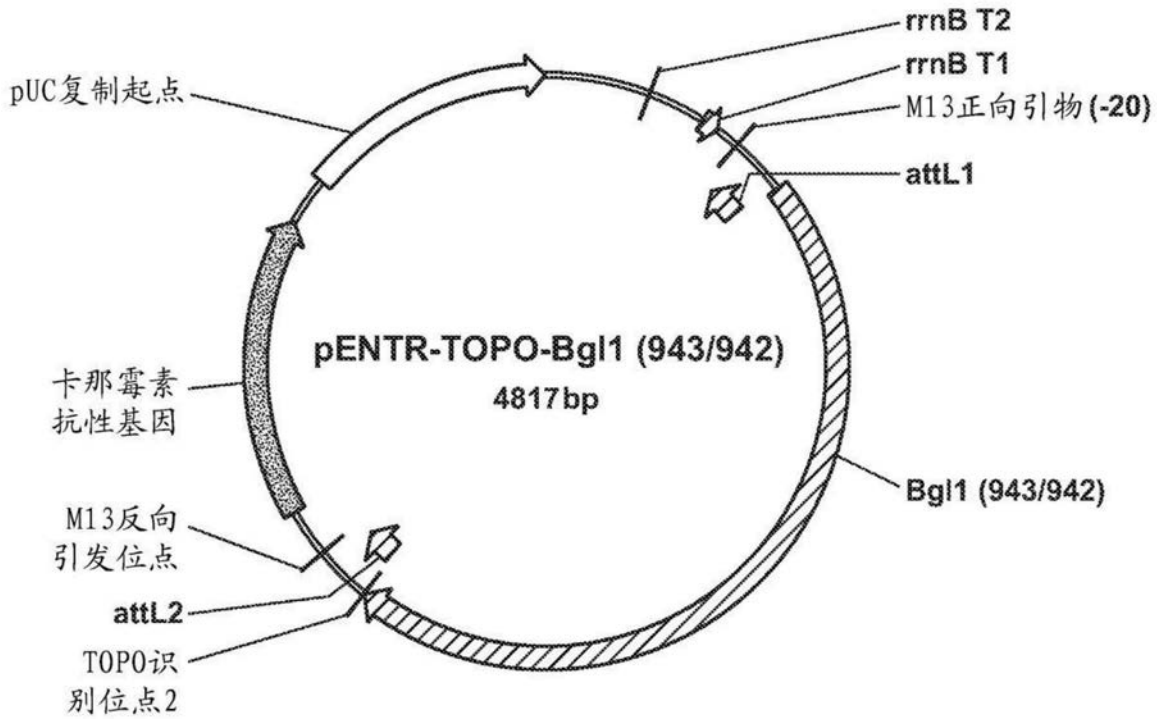

rrnB T2
rrnB T1
M13正向引物(-20)
attL1
pUC复制起点
pENTR-TOPO-Bgl1 (943/942)
4817bp
卡那霉素
抗性基因
Bgl1 (943/942)
M13反向
引发位点
attL2
TOPO识
别位点2

图55B

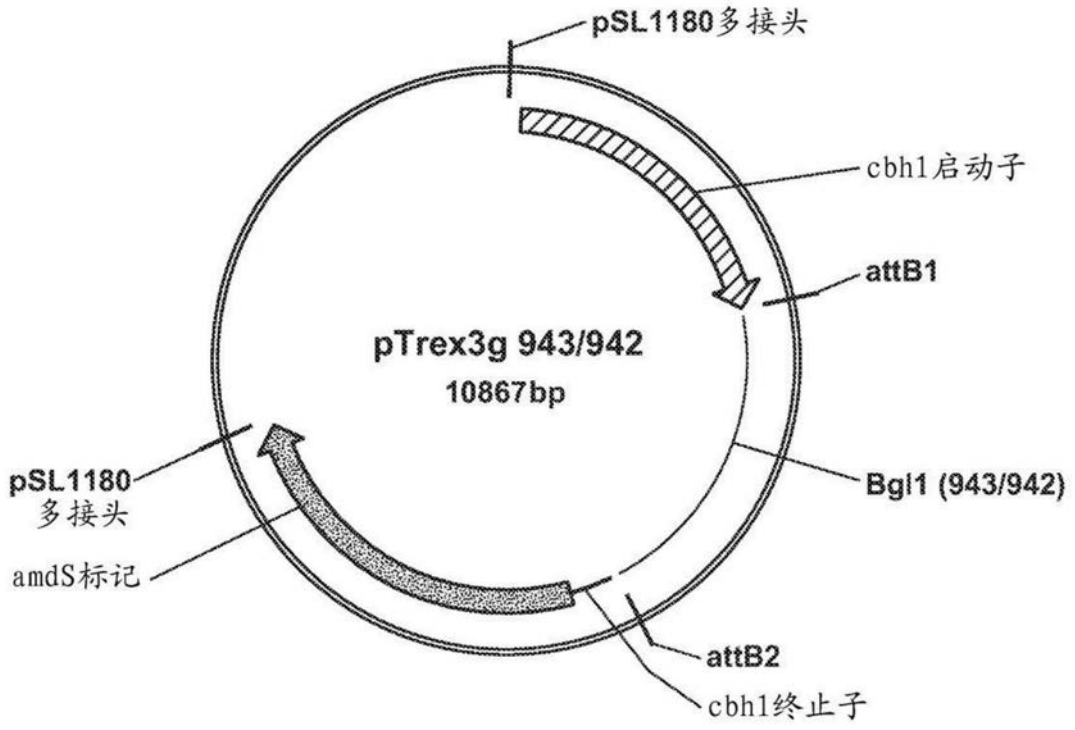

pSL1180多接头
cbh1启动子
attB1
pTrex3g 943/942
10867bp
pSL1180
多接头
Bgl1 (943/942)
amdS标记
attB2
cbh1终止子

图55C

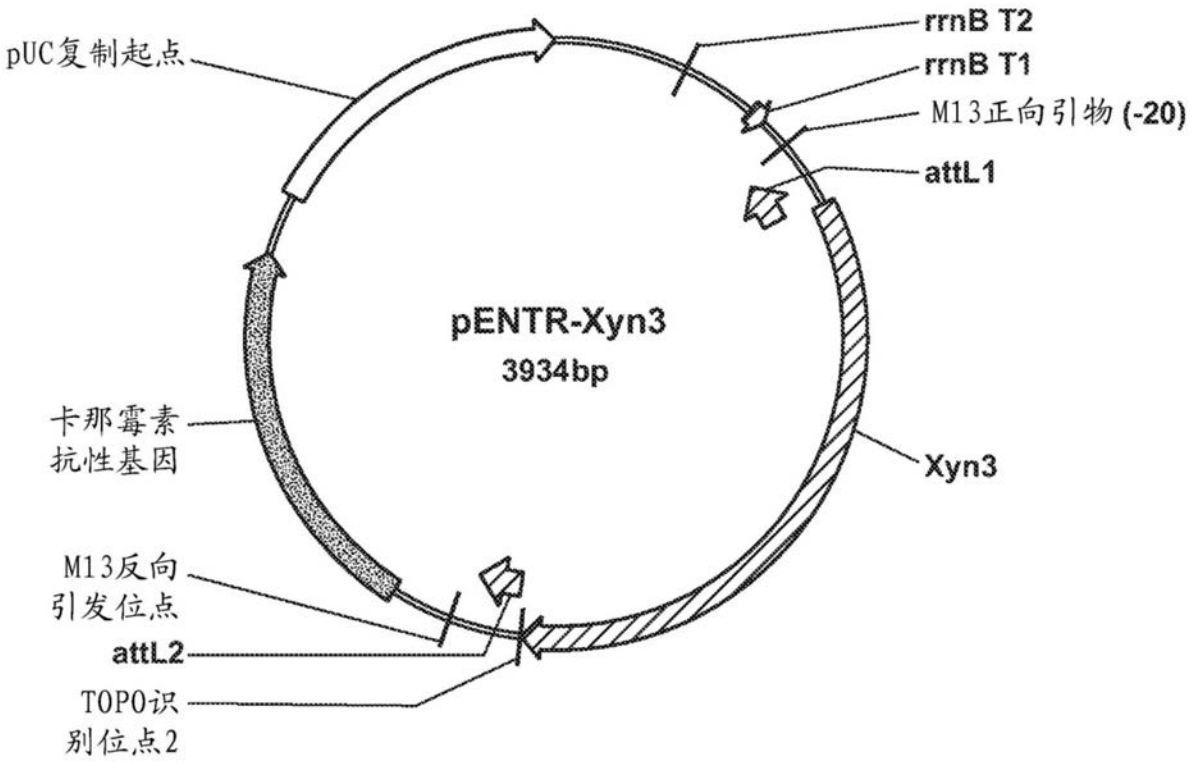

pUC复制起点
rrnB T2
rrnB T1
M13正向引物 (-20)
attL1
pENTR-Xyn3
3934bp
卡那霉素
抗性基因
Xyn3
M13反向
引发位点
attL2
TOPO识
别位点2

图55D

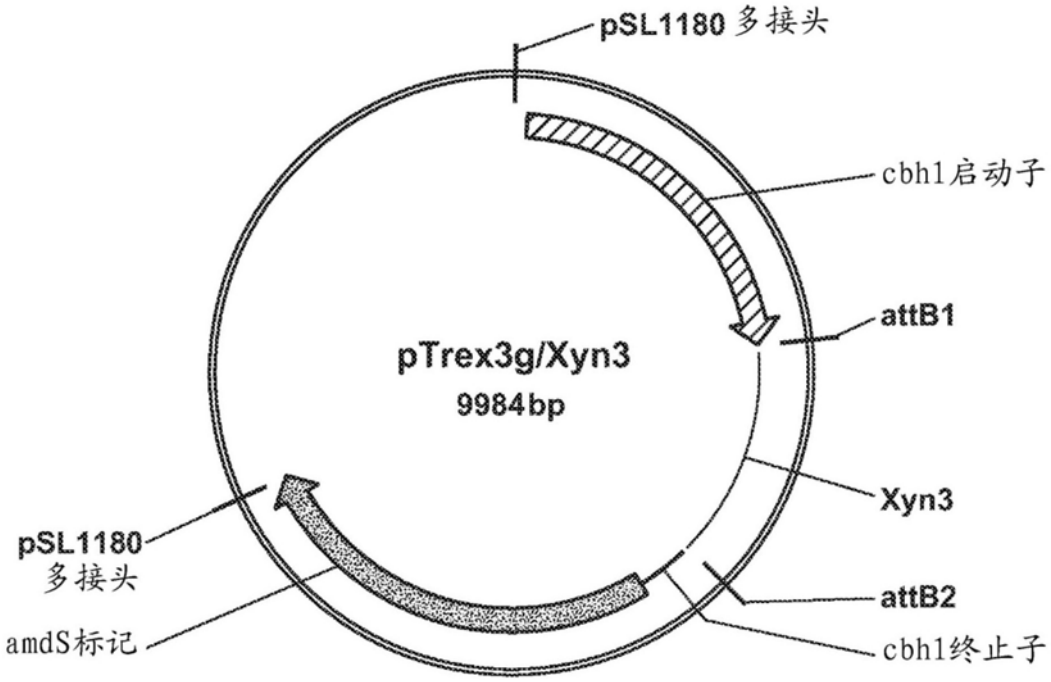

pSL1180 多接头
cbh1启动子
attB1
pTrex3g/Xyn3
9984bp
Xyn3
pSL1180
多接头
attB2
amdS标记
cbh1终止子

图55E

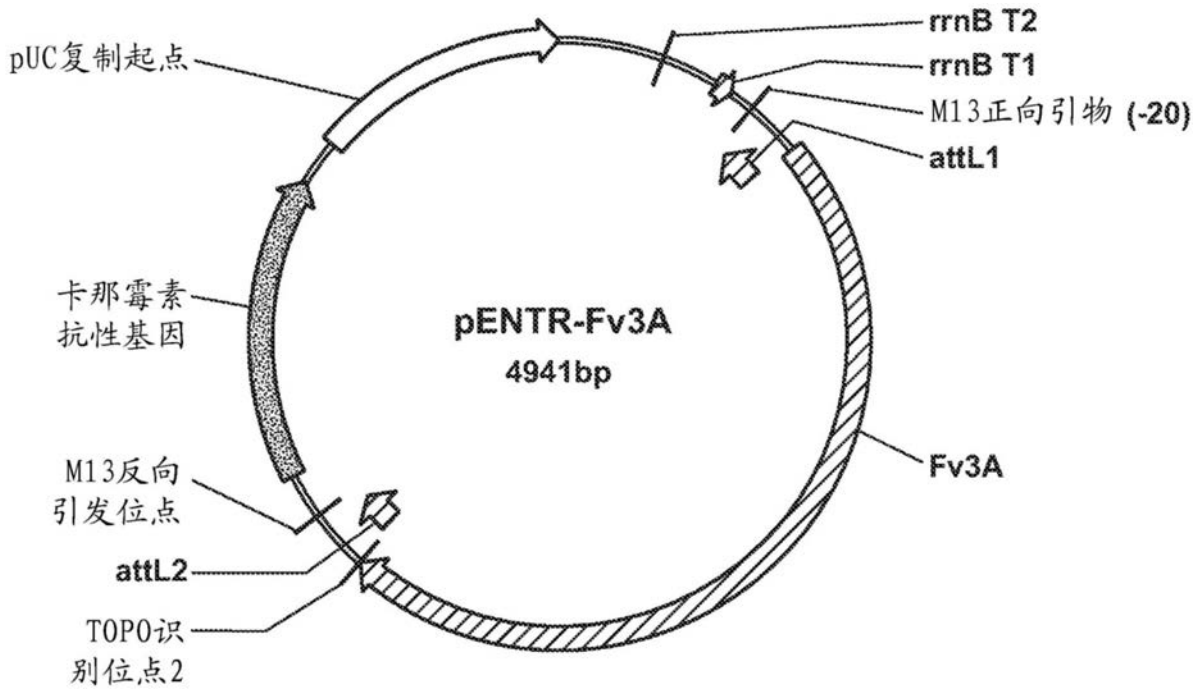

rrnB T2
rrnB T1
M13正向引物 (-20)
attL1
pUC复制起点
卡那霉素
抗性基因
pENTR-Fv3A
4941bp
Fv3A
M13反向
引发位点
attL2
TOPO识
别位点2

图55F

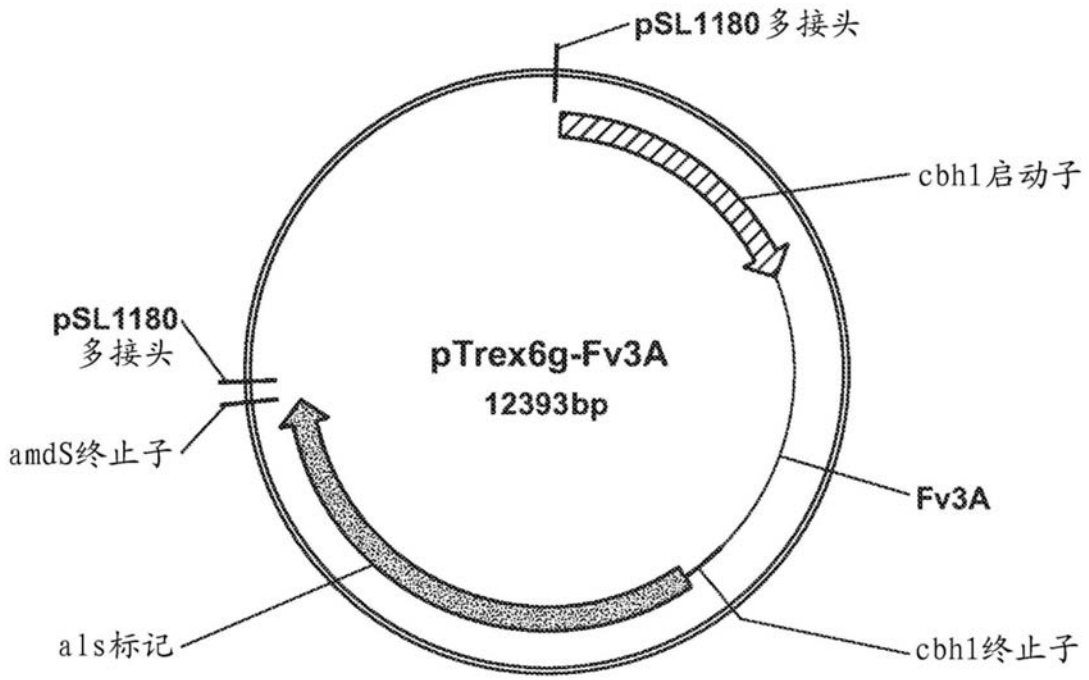

pSL1180多接头
cbh1启动子
pSL1180
多接头
pTrex6g-Fv3A
12393bp
amdS终止子
Fv3A
als标记
cbh1终止子

图55G

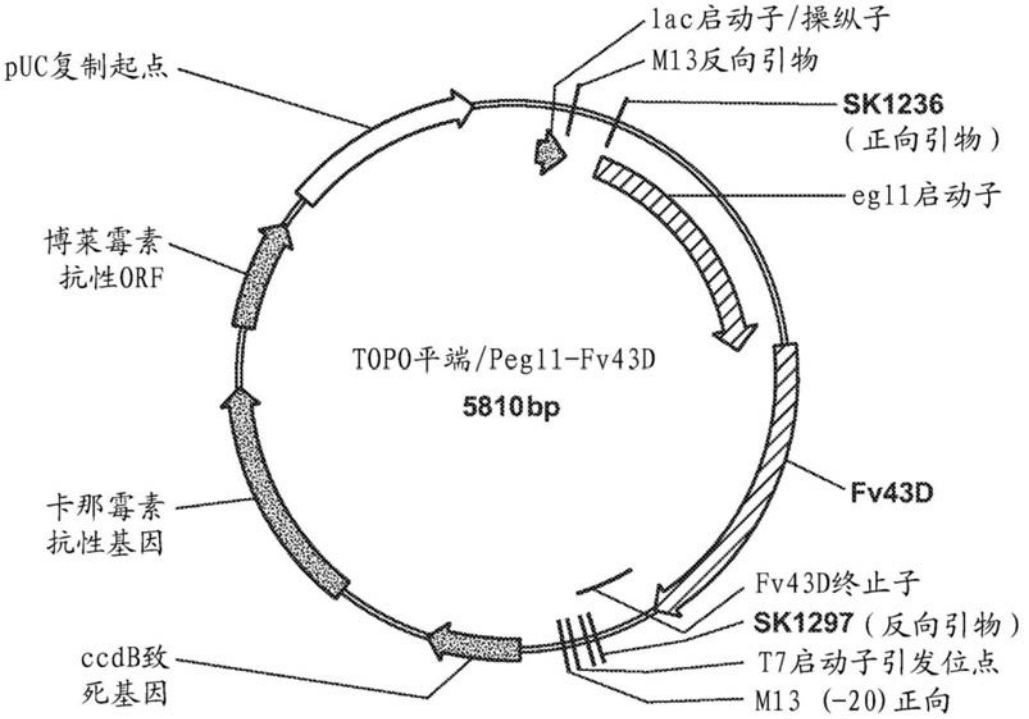

lac启动子/操纵子
M13反向引物
SK1236
（正向引物）
egl1启动子
pUC复制起点
博莱霉素
抗性ORF
TOPO平端/Pegl1-Fv43D
5810bp
Fv43D
卡那霉素
抗性基因
Fv43D终止子
SK1297（反向引物）
T7启动子引发位点
ccdB致
死基因
M13 (-20)正向

图55H

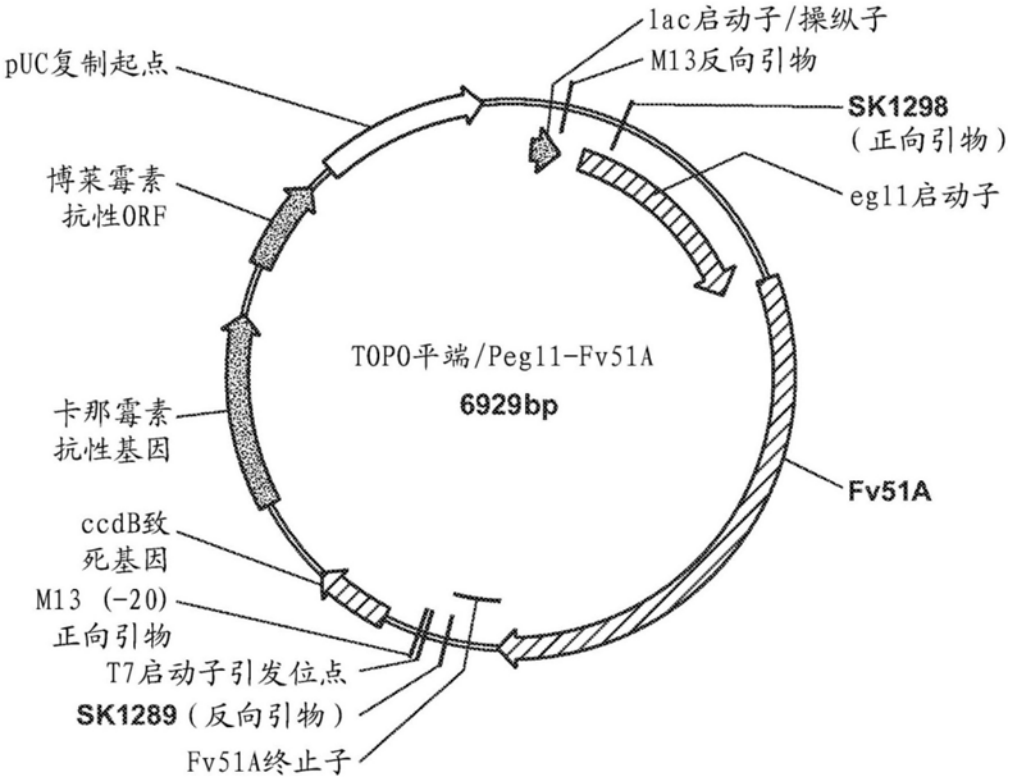

lac启动子/操纵子
M13反向引物
SK1298
（正向引物）
egl1启动子
pUC复制起点
博莱霉素
抗性ORF
TOPO平端/Pegl1-Fv51A
6929bp
卡那霉素
抗性基因
Fv51A
ccdB致
死基因
M13 (-20)
正向引物
T7启动子引发位点
SK1289（反向引物）
Fv51A终止子

图55I

```
gi|reesei|Bxl1  -MVNNAALLAALSALLPTALAQNNQTYANYSAQGQPDLYPETLATLTLSFPDCEHGPLKN
gi|fvert| Fv3A  MLLNLQVAASALSLSLLGGLAEAATPYT----------------------LPDCTKGPLSK
                 ::*  .  :***  *  .**:   .*:                      :*** :***.:
gi|reesei|      NLVCDSSAGYVERAQALISLFTLEELILNTQNSGPGVPRLGLPNYQVWNEALHGLDRAN-
gi|fvert|       NGICDTSLSPAKRAAALVAALTPEEKVGNLVSNATGAPRIGLPRYNWWNEALHGLAGSPG
                * :**:* . .:** **:: :* ** : *  ....*.**:***.*: ********  :
gi|reesei|      --FATKGGQFEWATSFPMPILTTAALNRTLIHQIADIISTQARAFSNSGRYGLDVYAPNV
gi|fvert|       GRFADTP-PYDAATSFPMPLLMAAAFDDDLIHDIGNVVGTEARAFTNGGWRGVDFWTPNV
                  ** .   :: *******:* :**::  ***:*.::::.*:****:*.*  *:*.::***
gi|reesei|      NGFRSPLWGRGQETPGEDAFFLSSAYTYEYITGIQGGVDPEHLKVAATVKHFAGYDLENW
gi|fvert|       NPFKDPRWGRGSETPGEDALHVS-RYARYIVRGLEG--DKEQRRIVATCKHYAGNDFEDW
                * *:.* ****.*******:.:*  *:   : *::*  * *: ::.** **:** *:*:*
gi|reesei|      NNQSRLGFDAIITQQDLSEYYTPQFLAAARYAKSRSLMCAYNSVNGVPSCANSFFLQTLL
gi|fvert|       GGFTRHDFDAKITPQDLAEYYVRPFQECTRDAKVGSIMCAYNAVNGIPACANSYLQETIL
                .. :* .*** ** ***:***.  *  .:* **  *:*****:***:*:****:: :*:*
gi|reesei|      RESWGFP-EWGYVSSDCDAVYNVFNPHDYASNQSSAAASSLRAGTDIDCGQTYPWHLNES
gi|fvert|       RGHWNWTRDNNWITSDCGAMQDIWQNHKYVKTNAEGAQVAFENGMDSSCEYTTTSDVSDS
                *  *.:. : .:::***.*: :::: *.*...::..*  ::. * * .*  * . .:.:*
gi|reesei|      FVAGEVSRGEIERSVTRLYANLVRLGYFD-KKNQYRSLGWKDVVKTDAWNISYEAAVEGI
gi|fvert|       YKQGLLTEKLMDRSLKRLFEGLVHTGFFDGAKAQWNSLSFADVNTKEAQDLALRSAVEGA
                :  * ::.  ::**:.**: .**: *:**  * *:.**.: ** ..:* ::: .:****
gi|reesei|      VLLKNDGTLPLSKKVR-SIALIGPWANATTQMQGNYYGPAPYLISPLEAAKKAGYHVNFE
gi|fvert|       VLLKNDGTLPLKLKKKDSVAMIGFWANDTSKLQGGYSGRAPFLHSPLYAAEKLGLDTNVA
                ***********. * : *:*:** *** *:::**.* * **:* *** **:* * ..*.
gi|reesei|      LGTEIAGNSTTG--FAKAIAAAKKSDAIIYLGGIDNTIEQEGADRTDIAWPGNQLDLIKQ
gi|fvert|       WGPTLQNSSSHDNWTTNAVAAAKKSDYILYFGGLDASAAGEDRDRENLDWPESQLTLLQK
                 *. : ..*: .   ::*:******* *:*:**:* :   *. ** :: ** .** *:::
gi|reesei|      LSEVGKPLVVLQMGGGQVDSSSLKSNKKVNSLVWGGYPGQSGGVALFDILSGKRAPAGRL
gi|fvert|       LSSLGKPLVVIQLG-DQVDDTALLKNKKINSILWVNYPGQDGGTAVMDLLTGRKSPAGRL
                **.:******:*:* .***.::* .***:**::* .****.**.*::*:*:*:::*****
gi|reesei|      VTTQYPAEYVHQFPQNDMNLRPDGKSNPGQTYIWYTGKPVYEFGSGLFYTTFKETLASHP
gi|fvert|       PVTQYPSKYTEQIGMTDMDLRPT-KSLPGRTYRWYS-TPVLPYGFGLHYTKFQAKFKSN-
                 .****::*..*:  .**:***  ** **:** **: .**  :* **.**.*: .: *:
gi|reesei|      KSLKFNTSSILSAPHPGYTYSEQIPVFTFEANIKNSGKTESPYTAMLFVRTSNAGPAPYP
gi|fvert|       -KLTFDIQKLLKG--CSAQYSDTCALPPIQVSVKNTGRITSDFVSLVFIKS-EVGPKPYP
                 .*.*: ..:*..   .  **:  .: .::..:**:*:  * :.:::*::: :.** ***
gi|reesei|      NKWLVGFDRLADIKPGHSSKLSIPIPVSALARVDSHGNRIVYPGKYELALNTDESVKLEF
gi|fvert|       LKTLAAYGRLHDVAPSSTKDISLEWTLDNIARRGENGDLVVYPGTYTLLLDEPTQAKIQV
                 * *..:.** *: *. :..:*:  .:. :** ..:*: :****.* * *:   ..*::.
gi|reesei|      ELVGEEVTIENWPLEEQQIKDATPDA
gi|fvert|       TLTGKKAILDKWPQDPKSA--------
```

图56

```
Fv43D ---MQLKFLSSALLLSLTGNCAAQDTNDIPPLITDLWSADPSAHVFEGKLWVYPSHDIEA
Fo43A ---MQLKFLSSALLFSLTSKCAAQDTNDIPPLITDLWSADPSAHVFEGKLWVYPSHDIEA
Gz43A ---MKSKLLFP--LLSFVG--QSLATNDDCPLITSRWTADPSAHVFNDTLWLYPSHDIDA
Pf43A --MLQRFAYILPLALLSVG--VKADN----PFVQSIYTADPAPMVYNDRVYVFMDHDNTG
Fv43A -MWLTSPLLFASTLLGLTGVALADN-----PIVQDIYTADPAPMVYNGRVYLFTGHDNDG
Fv43B --MRFSWLLCPLLAMGSALPETKTDVSTYTNPVLPGWHSDPSC-IQKDGLFLCVTSTFIS
Af43A -------------MAAPSLSYPTGIQSYTNPLFPGWHSDPSCAYVAEQDTFFCVTSTFI
Pf43B ----MSRSILPYASVFALLGGAIAEP----FLVLNSDFPDPSLIETSSGYYAFGTTGNGV
Fv43E MKVYWLVAWATSLTPALAGLIGHRRATTFNNPIIYSDFPDNDVFLGPDNYYYFSASNFHF

Fv43D NVVNGTGGAQYAMRDYHTYSMKTIYGKDPVIDHGVALSVDDVPWAKQQMWAPDAAYK--N
Fo43A NVVNGTGGAQYAMRDYHTYSMKSIYGKDPVVDHGVALSVDDVPWAKQQMWAPDAAHK--N
Gz43A GFENDPDGGQYAMRDYHVYSIDKIYGSLP-VDHGTALSVEDVPWASRQMWAPDAAHK--N
Pf43A -------ATYYNMTDWHLFSSADMANWQD---HGIPMSLANFTWANANAWAPQVIPR--N
Fv43A -------STDFNMTDWRLFSSADMVNWQH---HGVPMSLKTFSWANSRAWAGQVVAR--N
Fv43B FP----GLPVYASRDLVNWRLISHVWNRE---KQLPGISWKTAGQQQGMYAPTIRYH--K
Af43A AFP---GLPLYASRDLQNWKLASNIFNRP---SQIPDLR-VTDGQQSGIYAPTLRYH--E
Pf43B N------AQVASSPDFNTWTLLSGT-------DALPGPFPSWVASSPQIWAPDVLVKA-D
Fv43E SP----GAPVLKSKDLLNWDLIGHSIPRLNFGDGYDLPPGSRYYRG-GTWASSLRYRKSN

Fv43D GKYYLYFPAK-DK-DEIFRIGVAVSNKPSGPFK---ADK-SWIPGTYSIDPASYVDTNGE
Fo43A GKYYLYFPAK-DK-DEIFRIGVAVSNKPSGPFK---ADK-SWIPGTYSIDPASYVDTDNE
Gz43A GKYYLYFPAK-DK-DDIFRIGVAVSPTPGGPFV---PDK-SWIPHTFSIDPASFVDDDDR
Pf43A GQFYFYAPVR-HN-DGSMAIGVGVSSTITGPYH---DAIGKPLVENNEIDPTVFIDDDGQ
Fv43A GKFYFYVPVRNAK-TGGMAIGVGVSTNILGPYT---DALGKPLVENNEIDPTVYIDTDGQ
Fv43B GTYYVICEYLGVG-DIIGVIFKTTNPWDESSWS---DPV---TFKPNHIDPDLFWDDDGK
Af43A GQFYLIVSYLGP--QTKGLLFTSSDPYDDAAWS---DPL---EFAVHGIDPDIFWDHDGT
Pf43B GTYVMYFSASAASDSGKHCVGAATATSPEGPYTPVDSAVACPLDQGGAIDANGFIDTDGT
Fv43E GQWYWIGCIN------FWQTWVYTASSPEGPWY----NKGNFGDNNCYYDNGILIDDDDT

Fv43D AYLIWGGI-WGGQLQAWQDHKTFNESWLGDKAAPNGTNALSPQIAKLSKDMHKITETPRD
Fo43A AYLIWGGI-WGGQLQAWQDKKNFNESWIGDKAAPNGTNALSPQIAKLSKDMHKITETPRD
Gz43A AYLAWGGI-MGGQLQRWQDKNKYNES--GTEPG-NGTAALSPQIAKLSKDMHTLAEKPRD
Pf43A AYLYWG-------------------------------NPDLWYVKLNQDMISYSGSPTQ
Fv43A AYLYWG-------------------------------NPGLYYVKLNQDMLSYSGSINK
Fv43B VYCATHG----ITLQEIDLETGELSPELNIWNGTGGVWPEGPHIYKRDGYYYLMIAEGGT
Af43A VYVTSAED-QMIKQYTLDLKTGAIGPVDYLWNGTGGVWPEGPHIYKRDGYYYLMIAEGGT
Pf43B IYVVYKID-----------------------------GNSLDGDGTTHPTPIMLQQMEADGT
Fv43E MYVVYGSGEVKVSQLSQDGFSQVKSQVVFKNTDIGVQDLEGNRMYKING-------LYYI

Fv43D LVILAPETGKPLQAEDNKRRFFEGP-----WVHKRGKLYYLMYSTG--------------
Fo43A LVILAPETGKPLQAEDNKRRFFEGP-----WIHKRGKLYYLMYSTG--------------
Gz43A MLILDPKTGKPLLSEDEDRRFFEGP-----WIHKRNKIYYLTYSTG--------------
Pf43A IPLTTAGFGTRTGNAQRPTTFEEAP-----WVYKRNGIYYIAYAAD--------------
Fv43A VSLTTAGFGSRPNNAQRPTTFEEGP-----WLYKRGNLYYMIYAAN--------------
Fv43B ----AEDHAITIARARKITGPYEAYNNNPILTNRGTSEYFQTVGHGDLFQDTKGNWWGLC
Af43A ----ELGHSETMARSRTRTGPWEPYPHNPLLSNKGTSEYFQTVGHADLFQDGNGNWWAVA
Pf43B --TPTGSPIQLIDRSDLDGPLIEAP-----SLLLSNGIYYLSFSSN--------------
Fv43E LNDSPSGSQTWIWKSKSPWGPYESKVLADKVTPPISGGNSPHQGSLIKTPNGGWY-----

Fv43D -DTHFLVYATSKN---IYGPYT-------------YQGKILDPVDG---------WTTHG
Fo43A -DTHFLVYATSKN---IYGPYT-------------YRGKILDPVDG---------WTTHG
Gz43A -TTHYLVYATSKT---PYGPYT-------------YQGRILEPVDG---------WTTHS
Pf43A CCSEDIRYSTGTS---ATGPWT-------------YRGVIMPTQGSS--------FTNHE
```

图57A

```
Fv43A CCSEDIRYSTGPS---ATGPWT-------------YRGVVMNKAGRS---------FTNHP
Fv43B LATRITAQGVSPMGREAVLFNGTWNKGEWPKLQPVRGRMPGNLLPKPTRN-------VPGD
Af43A LSTRSGPAWKNYPMGRETVLAPAAWEKGEWPVIQPVRGQMQG-PFPPPNKR----VPRGE
Pf43B YYNTNYYDTSYAYASSITGPWT-----------KQSAPYAPLLVTGT---------ETSND
Fv43E FMSFTWAYPAGRLPVLAPITWG------------SDGFPILVKGANGGWGSSYPTLPGT

Fv43D SIVEYKGQWWLFFAD-AHTSGKDYLRQVKARKIWYDKDG-----KILLTRPKI-------
Fo43A SIVEYKGQWWLFFAD-AHTSGKDYLRQVKARKIWYDKNG-----KILLHRP---------
Gz43A SIVKYQGQWWLFYHD-AKTSGKDYLRQVKAKKIWYDSKG-----KILTKKP---------
Pf43A GIIDFQNNSYFFYHNGALPGGGGYQRSVCVEQFKYNADG-----TIPTIEMTTAG-----
Fv43A GIIDFENNSYFFYHNGALDGGSGYTRSVAVESFKYGSDG-----LIPEIKMTTQG-----
Fv43B GPFNADPDNYNLKKTKKIPPHFVHHRVPRDGAFSLSSKG----LHIVPSRNNVTGSVLPG
Af43A GGWIKQPDKVDFRPGSKIPAHFQYWRYPKTEDFTVSPRGHPNTLRLTPSFYNLTG-----
Pf43B GALSAPGGADFSVDGTKMLFHANLNGQDISGGRALFAAS------ITEASDVVTLQ----
Fv43E DGVTKNWTRTDTFRGTSLAPSWEWNHNPDVNSFTVNNGLTLRTASITKDIYQARN-----

Fv43D ------------------------------------------------------------
Fo43A ------------------------------------------------------------
Gz43A ------------------------------------------------------------
Pf43A ------------------PAQIGTLNPYVRQEAETAAWSSGITTEVCSEGGIDVGFINNG
Fv43A ------------------PAQLKSLNPYVKQEAETIAWSEGIETEVCSEGGLNVAFIDNG
Fv43B DEIELSGQRGLAFIGRRQTHTLFKYSVDIDFKPKSDDQEAGITVFRTQFDHIDLGIVRLP
Af43A -TADFKPDDGLSLVMRKQTDTLFTYTVDVSFDPKVADEEAGVTVFLTQQQHIDLGIVLLQ
Pf43B ------------------------------------------------------------
Fv43E ------------TLSHRTHGDHPTGIVKIDFSPMKDGDRAGLSAFRDQSAYIGIHRDNGK


Fv43D ------------------------------------------------------------
Fo43A ------------------------------------------------------------
Gz43A ------------------------------------------------------------
Pf43A DYIK------------VKGVAFGS-GAHSFSARVASANSGGTIAIHLGSTTGTLVGTCTV
Fv43A DYIK------------VKGVDFGSTGAKTFSARVASNSSGGKIELRLGSKTGKLVGTCTV
Fv43B TNQGSNKKSKLAFRFRATGAQNVPAPK---VVPVPDGWEKGVISLHIEAANATHYNLGAS
Af43A TTEG----LSLSFRFRVEGRGNYEGPLPEATVPVPKEWCGQTIRLEIQAVSDTEYVFAAA
Pf43B ------------------------------------------------------------
Fv43E FTIAT----KHGMNMDEWNGTTTDLGQIKATANVPSGRTKIWLRLQLDTNPAGTGNTIFS

Fv43D ------------------------------------------------------------
Fo43A ------------------------------------------------------------
Gz43A ------------------------------------------------------------
Pf43A PSTGGWQTWTTVTCSVSGASGTQ--------DVYFVFGGSGTGYLFN------FDYWQFA
Fv43A TTTGNWQTYKTVDCPVSGATGTS--------DLFFVFTGSGSGSLFN------FNWWQFS
Fv43B --SHRGKTLDIATASASLVSGGTGSFVGSLLGPYATCNGKGSGVECPKGGDVYVTQWTYK
Af43A PARHPAQRQIISRANSLIVSGDTGRFTGSLVGVYATSNG-GAGSTP-----AYISRWRYE
Pf43B ------------------------------------------------------------
Fv43E YSWDGVKYETLGPNFKLYNG-----------WAFFIAYRFGIFNFAETALGGSIKVESFT

Fv43D ----------------
Fo43A ----------------
Gz43A ----------------
Pf43A ----------------
Fv43A ----------------
Fv43B PVAQEIDHGVFVKSEL
Af43A GRGQMIDFGRVVPSY-
Pf43B ----------------
Fv43E AA--------------
```

图57B

```
Pa51A    MIHLKPALAALLALSTQCVAIDLFVKSSGGNKTTDIMYGLMHEDINNSGDGGIYAELISN
Fv51A    MVRFSSILAAAACF-VAVESVNIKVDSKGGNATSGHQYGFLHEDINNSGDGGIYAELIRN
Pf51A    MGKMWHSILVVLGLLSVGHAITINVSQSGGNKTSPLQYGLMFEDINHGGDGGLYAELVRN

Pa51A    RAFQGSEKFPSNLDNWSPVGGATLTLQKLAKPLSSALPYSVNVANPKEGKGKGKDTKGKK
Fv51A    RAFQYSKKYPVSLSGWRPINDAKLSLNRLDTPLSDALPVSMNVK---PGKGK------AKE
Pf51A    RAFQGSTVYPANLDGYDSVNGAILALQNLTNPLSPSMPSSLNVA------KGS----NNGS

Pa51A    VGLANAGFWGMDVKRQKYTGSFHVTGEYKGDFEVSLRSAITGETFGKKVVKGGSKKGKWT
Fv51A    IGFLNEGYWGMDVKKQKYTGSFWVKGAYKGHFTASLRSNLTDDVFGSVKVKSKANKKQWV
Pf51A    IGFANEGWWGIEVKPQRYAGSFYVQGDYQGDFDISLQSKLTQEVFATAKVRSSGKHEDWV

Pa51A    EKEFELVPFKDAPNSNNTFVVQWDAEGAKDGSLDLNLISLFPPTFKGRKNGLRIDLAQTM
Fv51A    EHEFVLTPNKNAPNSNNTFAITYDPKGA-DGALDFNLISLFPPTYKGRKNGLRVDLAEAL
Pf51A    QYKYELVPKKAASNTNNTLTITFDSKGLKDGSLNFNLISLFPPTYNNRPNGLRIDLVEAM

Pa51A    VELKPTFLRFPGGNMLEGNTLDTWWKWYETIGPLKDRPGMAGVWEYQQTLGLGLVEYMEW
Fv51A    EGLHPSLLRFPGGNMLEGNTNKTWWDWKDTLGPLRNRPGFEGVWNYQQTHGLGILEYLQW
Pf51A    AELEGKFLRFPGGSDVEGVQAPYWYKWNETVGDLKDRYSRPSAWTYEESNGIGLIEYMNW

Pa51A    ADDMNLEPIVGVFAGLALDGSFVPESEMGWVIQQALDEIEFLTGDAKTTKWGAVRAKLGH
Fv51A    AEDMNLEIIVGVYAGLSLDGSVTPKDQLQPLIDDALDEIEFIRG-PVTSKWGKKRAELGH
Pf51A    CDDMGLEPILAVWDGHYLSNEVISENDLQPYIDDTLNQLEFLMG-APDTPYGSWRASLGY

Pa51A    PKPWKVKWVEIGNEDWLAGRPAGFESYINYRFPMMMKAFNEKYPDIKIIASPSIFD-----
Fv51A    PKPFRLSYVEVGNEDWLAGYPTGWNSYKEYRFPMFLEAIKKAHPDLTVISSGASIDPVGK
Pf51A    PKPWTINYVEIGNEDNLYG---GLETYIAYRFQAYYDAITAKYPHMTVMESLTEMPG----

Pa51A    ---NMTIPAGAAGDHHPYLTPDEFVERFAKFDNLSKDNVTLIGEAASTHPNG---GIAWE
Fv51A    KDAGFDIPAPGIGDYHPYREPDVLVEEFNLFDNNKYG--HIIGEVASTHPNG---GTGWS
Pf51A    --------PAAAASDYHQYSTPDGFVSQFNYFDQMPVTNRTLNGEIATVYPNNPSNSVAWG

Pa51A    GDLMPLPWWGGSVAEAIFLISTERNGDKIIGATYAPGLRSLDRWQWSMTWVQHAADPALT
Fv51A    GNLMPYPWWISGVGEAVALCGYERNADRIPGTFYAPILKNENRWQWAITMIQFAADSAMT
Pf51A    SPFPLYPWWIGSVAEAVFLIGEERNSPKIIGASYAPMFRNINNWQWSPTLIAFDADSSRT

Pa51A    TRSTSWYVWRILAHHIIRETLPVDAPAGKPNFDPLFYVAGKSES-GTGIFKAAVYNSTES
Fv51A    TRSTSWYVWSLFAGHPMTHTLPTTA------DFDPLYYVAGKNEDKGTLIWKGAAYNTTKG
Pf51A    SRSTSWHVIKLLSTNKITQNLPTTWSGG--DIGPLYWVAGRNDNTGSNIFKAAVYNSTSD

Pa51A    --IPVSLKFDGLNEGAVANLTVLTGPE-DPYGYNDPFTGINVVKEKTTFIKAGKGGKFTF
Fv51A    ADVPVSLSFKGVKPGAQAELTLLTNKEKDPFAFNDPHKGNNVVDTKKTVLKADGKGAFNF
Pf51A    --VPVTVQFAGCN-AKSANLTILSSDD--PNASNYPG-GPEVVKTEIQSVTANAHGAFEF

Pa51A    TLPGLSVAVLETADAVKGGKGKGKGKGKGN
Fv51A    KLPNLSVAVLETLK----------KGKPYSS
Pf51A    SLPNLSVAVLKTE----------------
```

图58

```
                  *         20          *         40          *
xyn3    : -MKANVILC--LLAPLVAALPTETIHLDPELAALRANLTERTADLWDRQA :  47
P56588  : ------------------------------------------------QA :   2
P23360  : MVRPTILLTSLLLAPFAAASPI----------------------LEERQA :  28

                 60          *         80          *        100
xyn3    : SQSIDQLIKRKGKLYFGTATDRGLLQRE-KNAAIIQADLGQVTPENSMKW :  96
P56588  : SVSIDAKFKAHGKKYLGTIGDQYTLTKNTKNPAIIKADFGQLTPENSMKW :  52
P23360  : AQSVDQLIKARGKVYFGVATDQNRLTTG-KNAAIIQADFGQVTPENSMKW :  77

                  *        120          *        140          *
xyn3    : QSLENNQGQLNWGDADYLVNFAQQNGKSIRGHTLIWHSQLPAWVNNINNA : 146
P56588  : DATEPNRGQFTFSGSDYLVNFAQSNGKLIRGHTLVWHSQLPGWVSSITDK : 102
P23360  : DATEPSQGNFNFAGADYLVNWAQQNGKLIRGHTLVWHSQLPSWVSSITDK : 127

                160          *        180          *        200
xyn3    : DTLRQVIRTHVSTVVGRYKGKIRAWDVVNEIFNEDGTLRSSVFSRLLGEE : 196
P56588  : NTLISVLKNHITTVMTRYKGKIYAWDVLNEIFNEDGSLRNSVFYNVIGED : 152
P23360  : NTLTNVMKNHITTLMTRYKGKIRAWDVVNEAFNEDGSLRQTVFLNVIGED : 177

                  *        220          *        240          *
xyn3    : FVSIAFRAARDADPSARLYINDYNLDRANYGKVNGLKTYVSKWISQGVPI : 246
P56588  : YVRIAFETARSVDPNAKLYINDYNLDSAGYSKVNGMVSHVKKWLAAGIPI : 202
P23360  : YIPIAFQTARAADPNAKLYINDYNLDSASYPKTQAIVNRVKQWRAAGVPI : 227

                260          *        280     N    *        300
xyn3    : DGIGSQSHLSGGGGSGTLGALQQLATVPVTELAITELDIQGAPTTDYTQV : 296
P56588  : DGIGSQTHLGAGAGSAVAGALNALASAGTKEIAITELDIAGASSTDYVNV : 252
P23360  : DGIGSQTHLSAGQGAGVLQALPLLASAGTPEVAITELDVAGASPTDYVNV : 277

                  *        320          *        340          *
xyn3    : VQACLSVSKCVGITVWGISDKDSWRASTNPLLFDANFNPKPAYNSIVGIL : 346
P56588  : VNACLNQAKCVGITVWGVADPDSWRSSSSPLLFDGNYNPKAAYNAIANAL : 302
P23360  : VNACLNVQSCVGITVWGVADPDSWRASTTPLLFDGNFNPKPAYNAIVQDL : 327


xyn3    : Q- : 347
P56588  : -- :   -
P23360  : QQ : 329
```

图59A

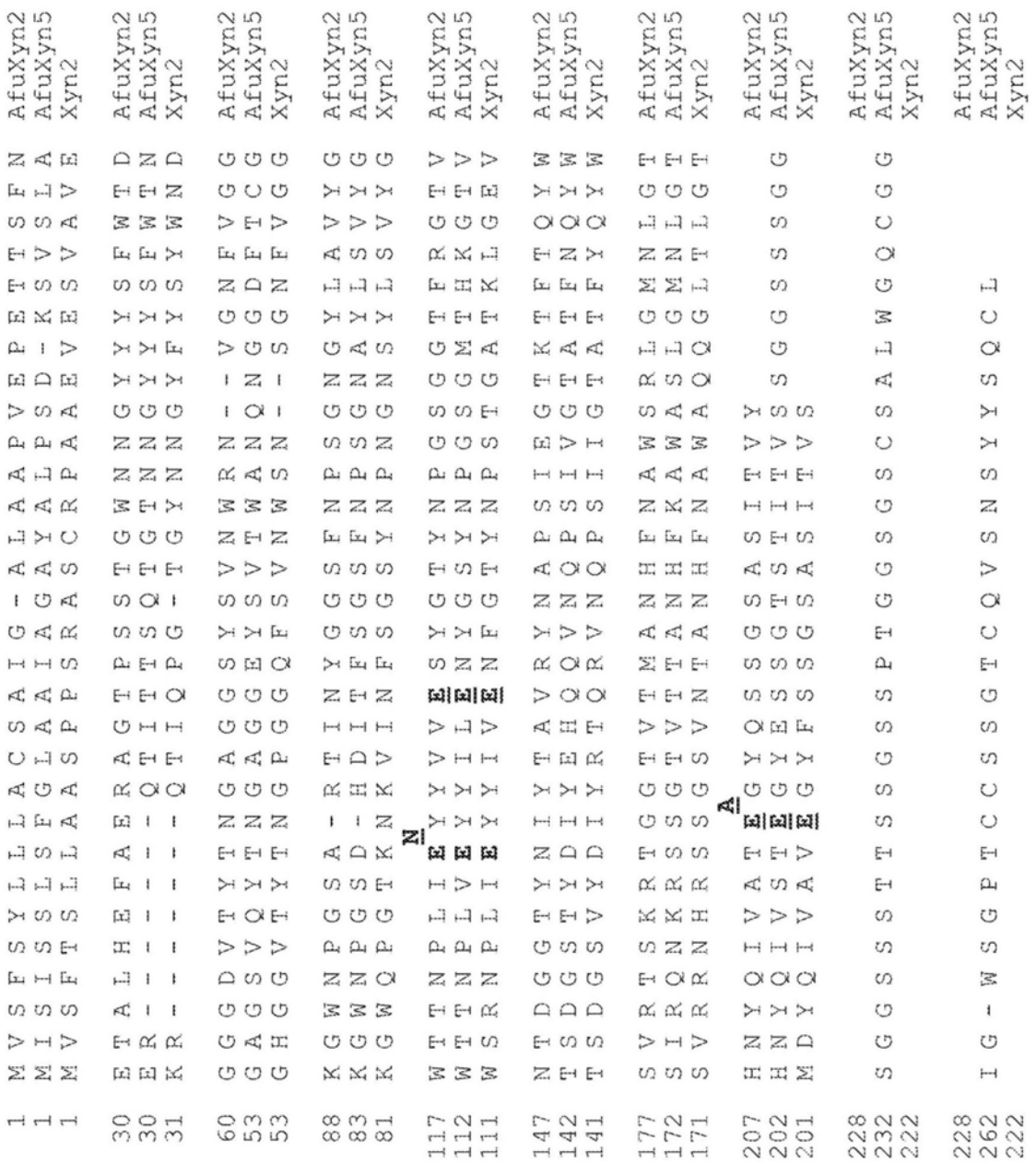


图59B

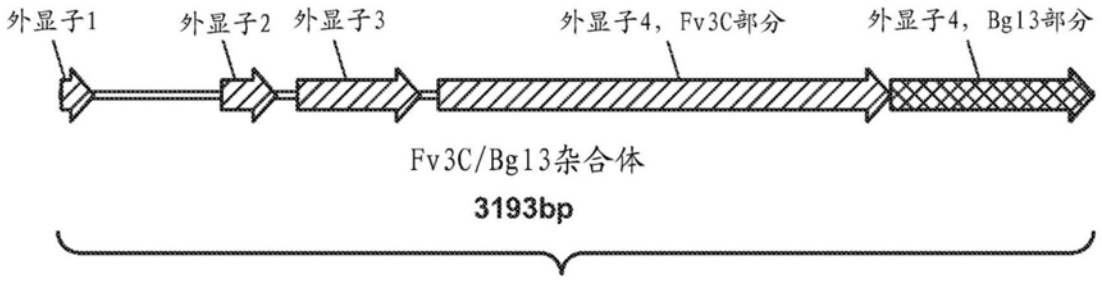


图60A

图60A

SEQ ID NO: 82

编码Fv3C/Bg13的核苷酸序列

atgaagctgaattgggtcgccgcagccctgtctataggtgctgctggcactgacagcgca
gttgctcttgcttctgcagttccagacactttggctggtgtaaaggtcagtttttttca
ccatttcctcgtctaatctcagccttgttgccatatcgcccttgttcgctcggacgccac
gcaccagatcgcgatcatttcctcccttgcagccttggttcctcttacgatcttccctcc
gcaattatcagcgcccttagtctacacaaaaaccccgagacagtctttcattgagtttg
tcgacatcaagttgcttctcaactgtgcatttgcgtggctgtctacttctgcctctagac
aaccaaatctgggcgcaattgaccgctcaaaccttgttcaaataacctttttattcgag
acgcacatttataaatatgcgcctttcaataataccgactttatgcgcggcggctgctgt
ggcggttgatcagaaagctgacgctcaaaaggttgtcacgagagatacactcgcatactc
gccgcctcattatccttcaccatggatggaccctaatgctgttggctgggaggaagctta
cgccaaagccaagagctttgtgtcccaactcactctcatggaaaaggtcaacttgaccac
tggtgttgggtaagcagctccttgcaaacagggtatctcaatcccctcagctaacaactt
ctcagatggcaaggcgaacgctgtgtaggaaacgtgggatcaattcctcgtctcggtatg
cgaggtctctgtctccaggatggtcctcttggaattcgtctgtccgactacaacagcgct
tttcccgctggcaccacagctggtgcttcttggagcaagtctctctggtatgagagaggt
ctcctgatgggcactgagttcaaggagaagggtatcgatatcgctcttggtcctgctact
ggacctcttggtcgcactgctgctggtggacgaaactgggaaggcttcaccgttgatcct
tatatggctggccacgccatggccgaggccgtcaagggtattcaagacgcaggtgtcatt
gcttgtgctaagcattacatcgcaaacgagcagggtaagccacttggacgatttgaggaa
ttgacagagaactgaccctcttgtagagcacttccgacagagtggcgaggtccagtcccg
caagtacaacatctccgagtctctctcctccaacctggatgacaagactatgcacgagct
ctacgcctggcccttcgctgacgccgtccgcgccggcgtcggttccgtcatgtgctcgta
caaccagatcaacaactcgtacggttgccagaactccaagctcctcaacggtatcctcaa
ggacgagatgggcttccagggtttcgtcatgagcgattgggcggcccagcataccggtgc
cgcttctgccgtcgctggtctcgatatgagcatgcctggtgacactgccttcgacagcgg
atacagcttctggggcggaaacttgactctggctgtcatcaacggaactgttcccgcctg
gcgagttgatgacatggctctgcgaatcatgtctgccttcttcaaggttggaaagacgat
agaggatcttcccgacatcaacttctcctcctggacccgcgacaccttcggcttcgtgca
tacatttgctcaagagaaccgcgagcaggtcaactttggagtcaacgtccagcacgacca
caagagccacatccgtgaggccgctgccaagggaagcgtcgtgctcaagaacaccgggtc
ccttcccctcaagaacccaaagttcctcgctgtcattggtgaggacgccggtcccaaccc
tgctggacccaatggttgtggtgaccgtggttgcgataatggtaccctggctatggcttg

图60B-1

gggctcgggaacttcccaattcccttacttgatcacccccgatcaagggctctctaatcg
agctactcaagacggaactcgatatgagagcatcttgaccaacaacgaatgggcttcagt
acaagctcttgtcagccagcctaacgtgaccgctatcgttttcgccaatgccgactctgg
tgagggatacattgaagtcgacggaaactttggtgatcgcaagaacctcaccctctggca
gcagggagacgagctcatcaagaacgtgtcgtccatatgccccaacaccattgtagttct
gcacaccgtcggccctgtcctactcgccgactacgagaagaaccccaacatcactgccat
cgtctgggctggtcttcccggccaagagtcaggcaatgccatcgctgatctcctctacgg
caaggtcagccctggccgatctcccttcacttggggccgcacccgcgagagctacggtac
tgaggttctttatgaggcgaacaacggccgtggcgctcctcaggatgacttctctgaggg
tgtcttcatcgactaccgtcacttcgaccgacgatctccaagcaccgatggaaagagctc
tcccaacaacaccgctgctcctctctacgagttcggtcacggtctatcttggtcgacgtt
caagttctccaacctccacatccagaagaacaatgtcggccccatgagcccgcccaacgg
caagacgattgcggctccctctctgggcagcttcagcaagaaccttaaggactatggctt
ccccaagaacgttcgccgcatcaaggagtttatctacccctacctgagcaccactacctc
tggcaaggaggcgtcgggtgacgctcactacggccagactgcgaaggagttcctccccgc
cggtgccctggacggcagccctcagcctcgctctgcggcctctggcgaacccggcggcaa
ccgccagctgtacgacattctctacaccgtgacggccaccattaccaacacgggctcggt
catggacgacgccgttccccagctgtacctgagccacggcggtcccaacgagccgcccaa
ggtgctgcgtggcttcgaccgcatcgagcgcattgctcccggccagagcgtcacgttcaa
ggcagacctgacgcgccgtgacctgtccaactgggacacgaagaagcagcagtgggtcat
taccgactaccccaagactgtgtacgtgggcagctcctcgcgcgacctgccgctgagcgc
ccgcctgccatga

图60B-2

SEQ ID NO:159
Fv3C/Bg13嵌合多肽序列（Bg13嵌合部分以粗体和大写标出）

mklnwvaaalsigaagtdsavalasavpdtlagvkkadaqkvvtrdtlayspphypspwmdpna
vgweeayakaksfvsqltlmekvnlttgvgwqgercvgnvgsiprlgmrglclqdgplgirlsd
ynsafpagttagaswskslwyergllmgtefkekgidialgpatgplgrtaaggrnwegftvdp
ymaghamaeavkgiqdagviacakhyianeqehfrqsgevqsrkyniseslssnlddktmhely
awpfadavragvgsvmcsynqinnsygcqnskllngilkdemgfqgfvmsdwaaqhtgaasava
gldmsmpgdtafdsgysfwggnltlavingtvpawrvddmalrimsaffkvgktiedlpdinfs
swtrdtfgfvhtfaqenreqvnfgvnvqhdhkshireaaakgsvvlkntgslplknpkflavig
edagpnpagpngcgdrgcdngtlamawgsgtsqfpylitpdqglsnratqdgtryesiltnnew
asvqalvsqpnvtaivfanadsgegyievdgnfgdrknltlwqqgdeliknvssicpntivvlh
tvgpvlladyeknpnitaivwaglpgqesgnaiadllygkvspgrspftwgrtresygtevlye
anngrgapqddfsegvfidyrhfdrrspstdgksspnntaaplyefghgls**WSTFKFSNLHIQK**
NNVGPMSPPNGKTIAAPSLGSFSKNLKDYGFPKNVRRIKEFIYPYLSTTTSGKEASGDAHYGQT
AKEFLPAGALDGSPQPRSAASGEPGGNRQLYDILYTVTATITNTGSVMDDAVPQLYLSHGGPNE
PPKVLRGFDRIERIAPGQSVTFKADLTRRDLSNWDTKKQQWVITDYPKTVYVGSSSRDLPLSARLP

图60C

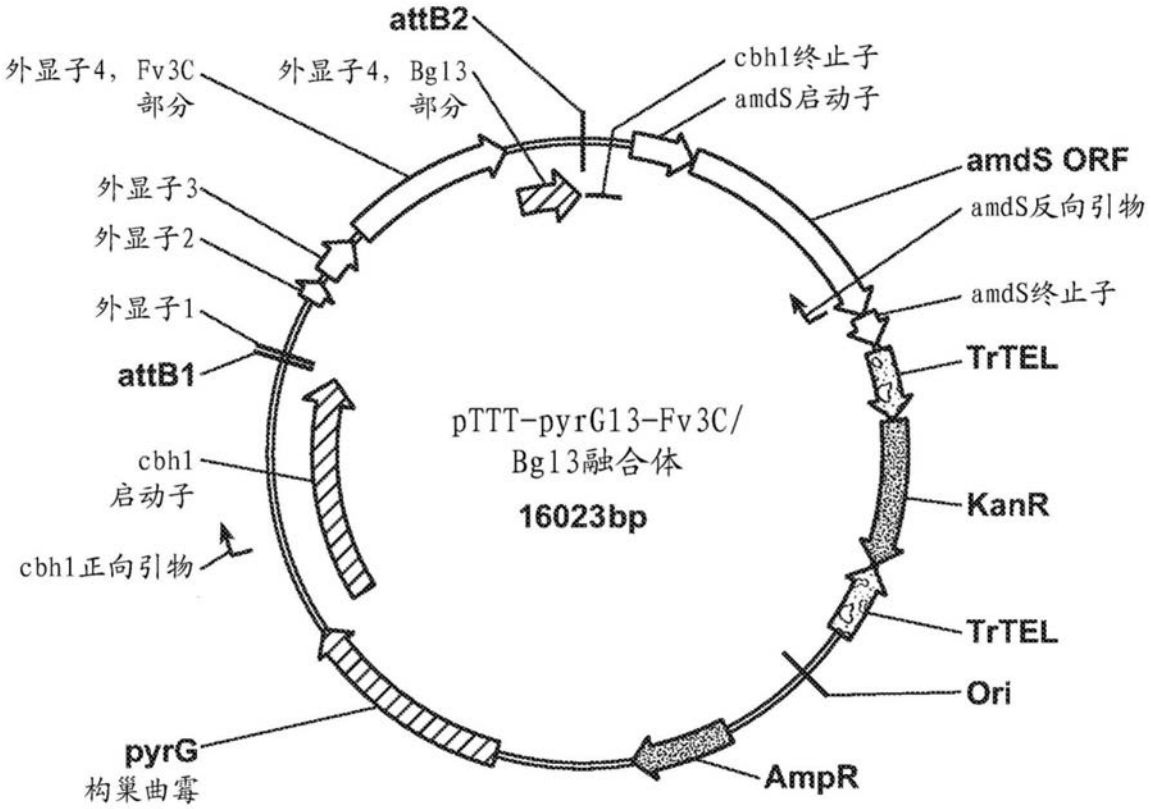

attB2
外显子4，Fv3C
部分
外显子4，Bg13
部分
cbh1终止子
amdS启动子
amdS ORF
amdS反向引物
amdS终止子
TrTEL
KanR
TrTEL
Ori
AmpR
pyrG
构巢曲霉
外显子3
外显子2
外显子1
attB1
cbh1
启动子
cbh1正向引物
pTTT-pyrG13-Fv3C/
Bg13融合体
16023bp

图61

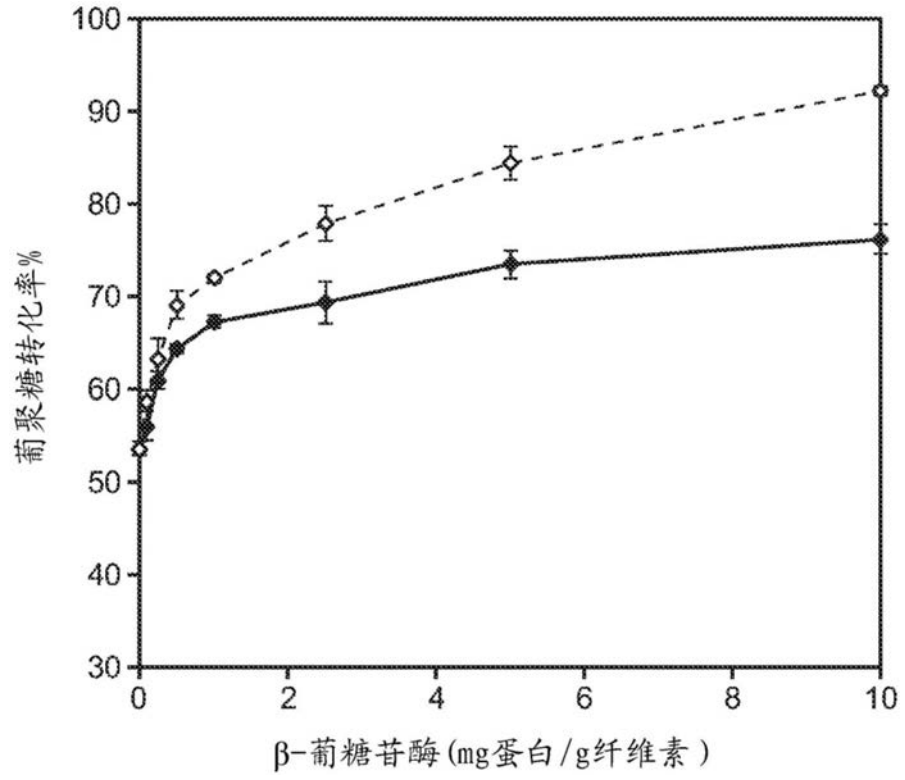

葡聚糖转化率%
β-葡糖苷酶(mg蛋白/g纤维素)
100
90
80
70
60
50
40
30
0
2
4
6
8
10

图62

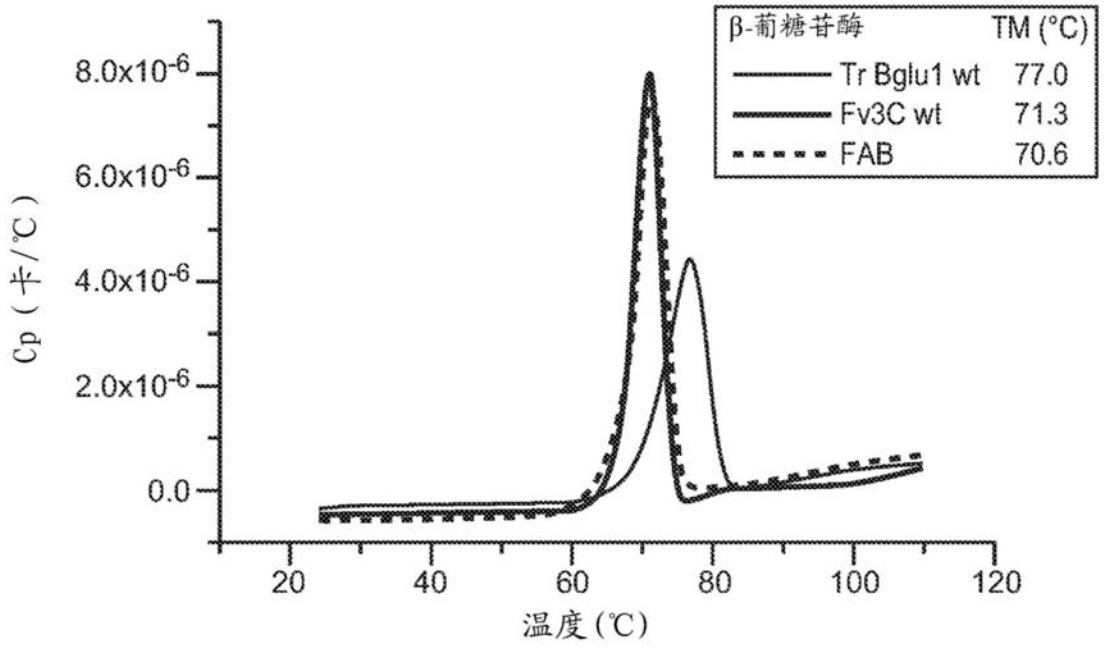

β-葡糖苷酶 TM (°C)
Tr Bglu1 wt 77.0
Fv3C wt 71.3
FAB 70.6
8.0x10-6
6.0x10-6
4.0x10-6
2.0x10-6
0.0
Cp (卡/℃)
20
40
60
80
100
120
温度(℃)

图63

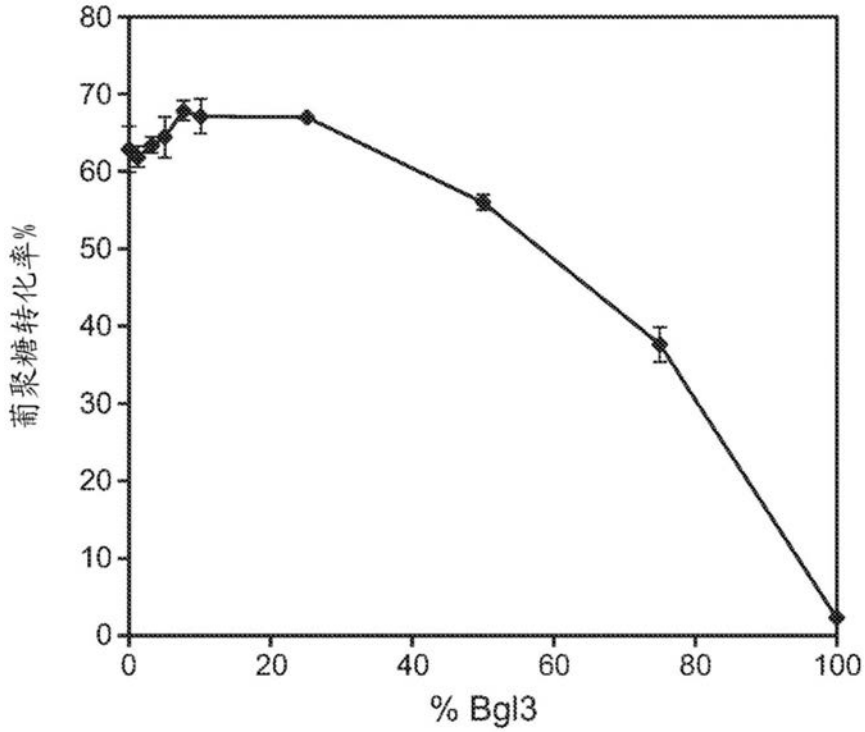

80
70
60
50
40
30
20
10
0
葡聚糖转化率%
0
20
40
60
80
100
% Bgl3

图64A

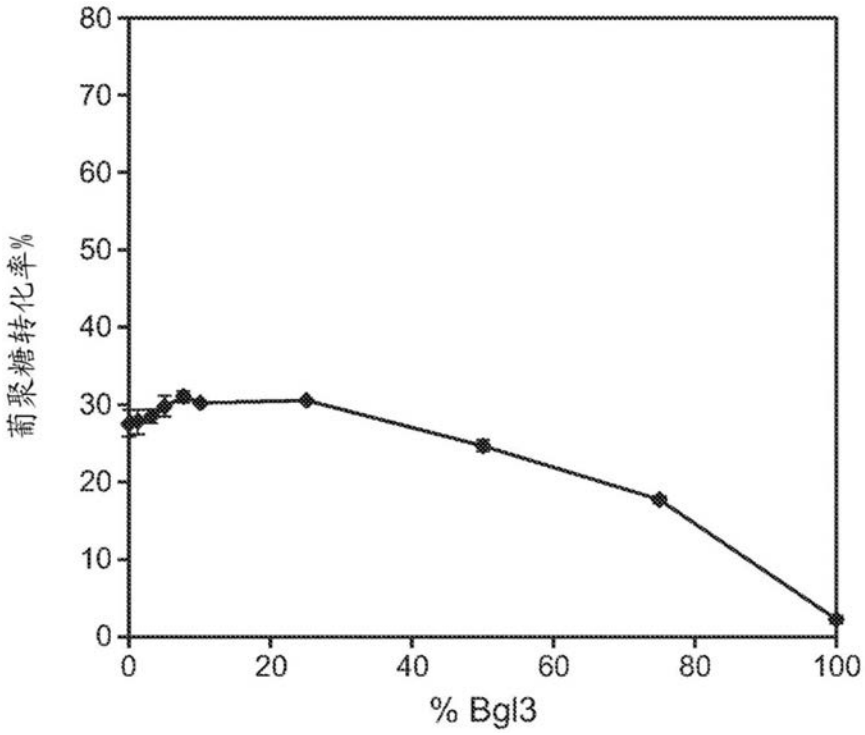

葡聚糖转化率%
% Bgl3
0
10
20
30
40
50
60
70
80
0
20
40
60
80
100

图64B

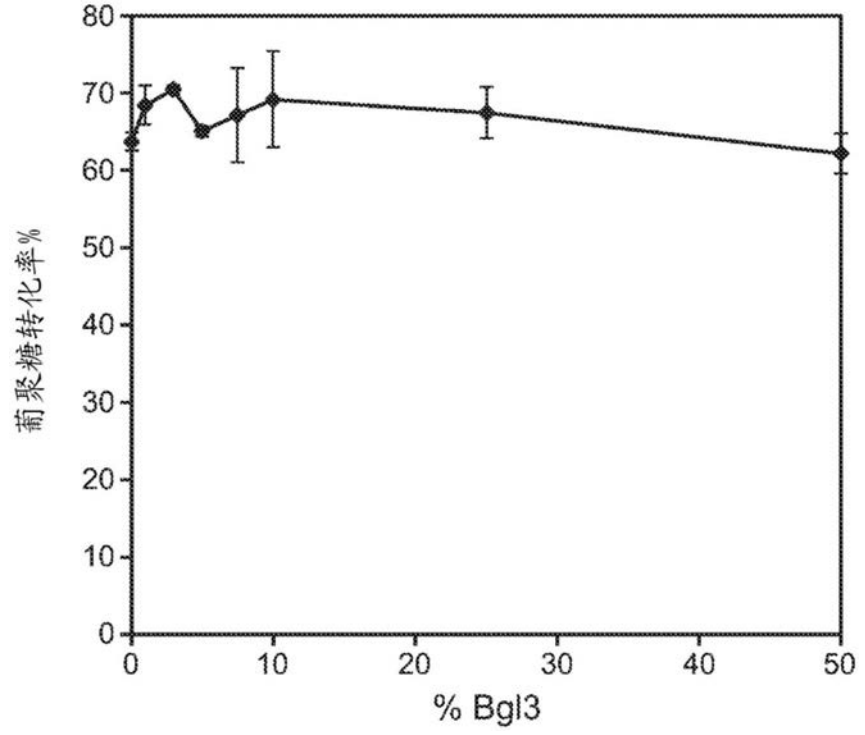

葡聚糖转化率%
% Bgl3
0
10
20
30
40
50
60
70
80
0
10
20
30
40
50

图64C

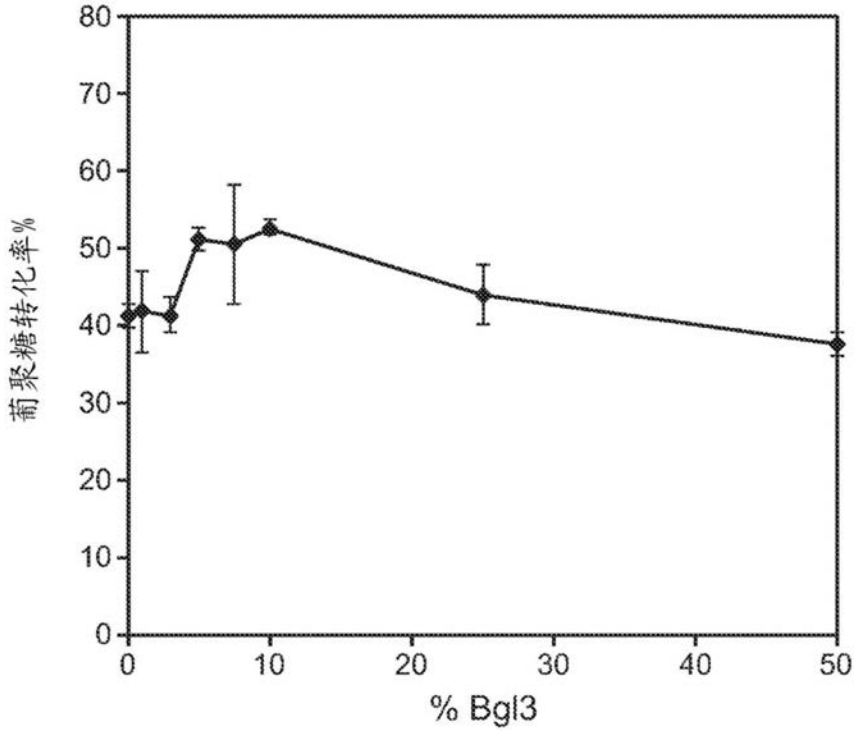

葡聚糖转化率%
% Bgl3
0
10
20
30
40
50
60
70
80

图64D

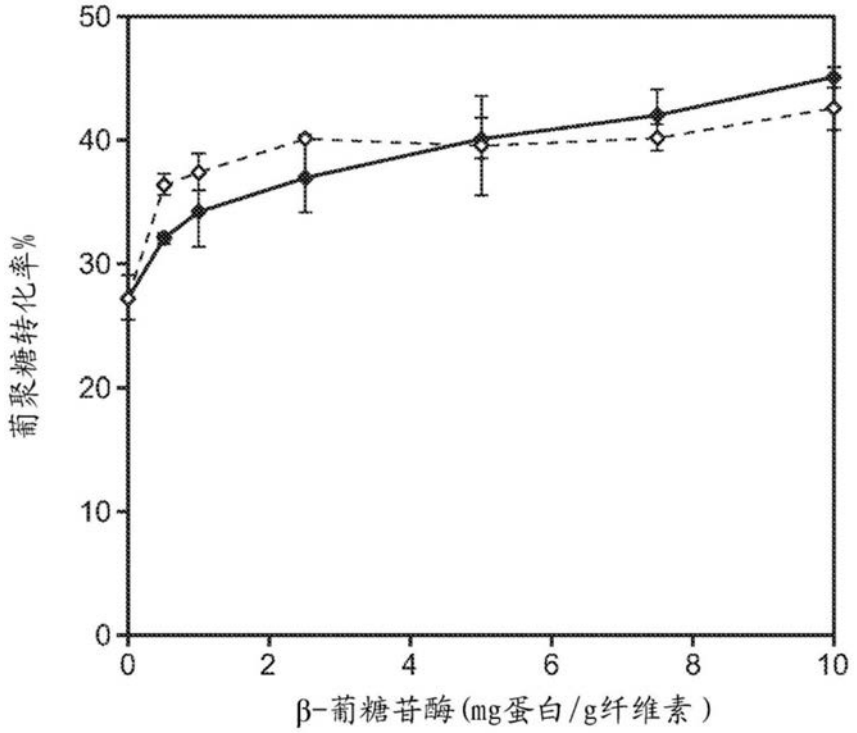

葡聚糖转化率%
β-葡糖苷酶(mg蛋白/g纤维素)
0
2
4
6
8
10
20
30
40
50

图65A

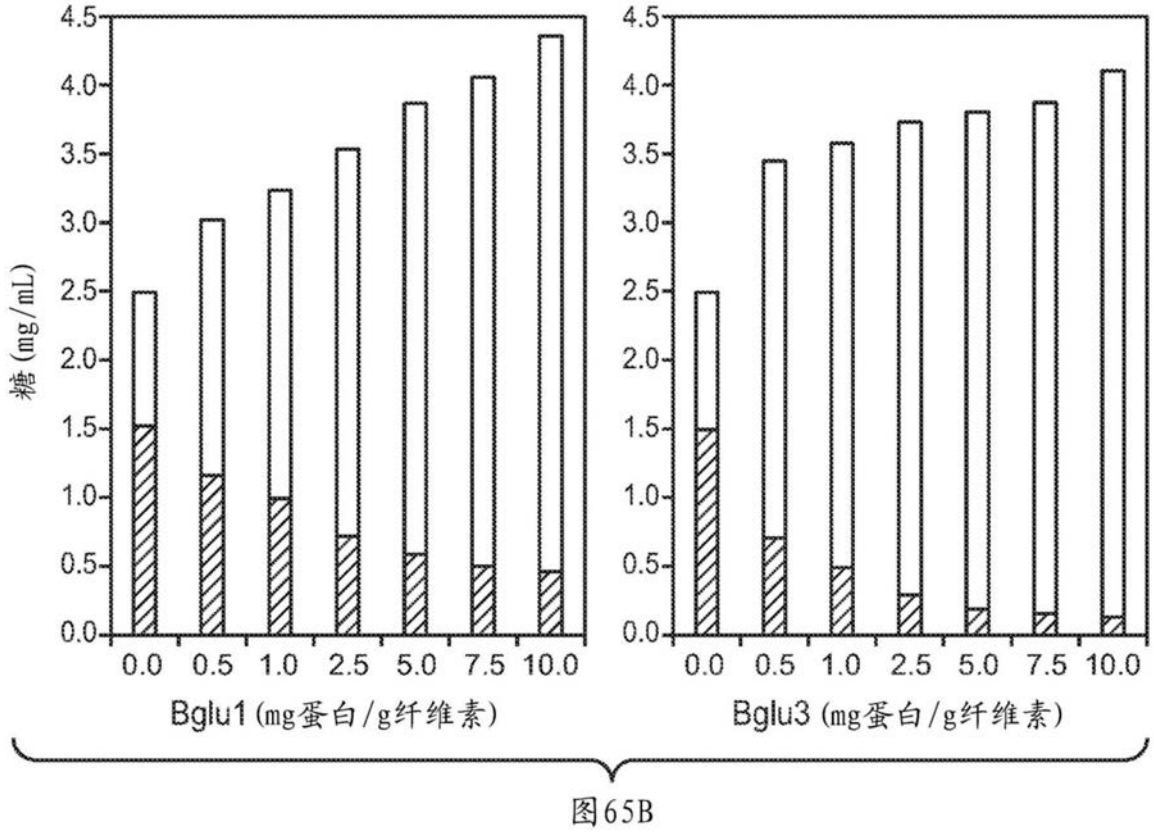

4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
糖(mg/mL)
0.0 0.5 1.0 2.5 5.0 7.5 10.0
Bglu1(mg蛋白/g纤维素)
0.0 0.5 1.0 2.5 5.0 7.5 10.0
Bglu3(mg蛋白/g纤维素)

图65B

图65B

SEQ ID NO:	引物	序列
92	正向 SK943	5'– CACCATGAGATATAGAACAGCTGCCGCT-3'
93	反向 SK941	5'-CGACCGCCCTGCGGAGTCTTGCCCAGTGGTCCCGCGACAG-3'
94	正向 (SK940)	5'–CTGTCGCGGGACCACTGGGCAAGACTCCGCAGGGCGGTCG-3'
95	反向 (SK942)	5'– CCTACGCTACCGACAGAGTG-3'
96	正向 SK771	5' – GTCTAGACTGGAAACGCAAC -3'
97	反向 SK745	5' – GAGTTGTGAAGTCGGTAATCC -3'
98	正向 xyn3F-2	5'–CACCATGAAAGCAAACGTCATCTTGTGCCTCCTGG-3
99	反向 (xyn3R-2)	5'-CTATTGTAAGATGCCAACAATGCTGTTATATGCCGGCTTGGGG-3'
100	正向 SK745	5' – GAGTTGTGAAGTCGGTAATCC-3'
101	反向 SK822	5' – CACGAAGAGCGGCGATTC-3'
102	正向 MH124	5' – CAC CCA TGC TGC TCA ATC TTC AG -3'
103	反向 MH125	5' - TTA CGC AGA CTT GGG GTC TTG AG -3'
104	正向 SK1334	5' – GCTTGAGTGTATCGTGTAAG -3'
105	正向引物 SK1335	5' – GCAACGGCAAAGCCCCACTTC -3'
106	反向 SK1299	5' - GTAGCGGCCGCCTCATCTCATCTCATCCATCC -3'
107	正向 SK1322	5'–CACCATGCAGCTCAAGTTTCTGTC-3
108	反向 SK1297	5'–GGTTACTAGTCAACTGCCCGTTCTGTAGCGAG-3'
109	正向 SK1236	5'–CATGCGATCGCGACGTTTTGGTCAGGTCG-3
110	反向 SK1321	5'-GACAGAAACTTGAGCTGCATGGTGTGGGACAACAAGAAGG-3'
111	正向 SK1159	5'–CACCATGGTTCGCTTCAGTTCAATCCTAG-3'
112	反向 SK1289	5'-GTGGCTAGAAGATATCCAACAC-3'
113	正向 SK1236	5'–CATGCGATCGCGACGTTTTGGTCAGGTCG-3'
114	反向 SK1262	5'–GAACTGAAGCGAACCATGGTGTGGGACAACAAGAAGGAC-3'
115	正向 SK1298	5'-GTAGTTATGCGCATGCTAGAC-3'
116	正向 MH234	5'-CACCATGAAGCTGAATTGGGTCGC-3'
117	反向 MH235	5'-TTACTCCAACTTGGCGCTG-3'

图66A

SEQ ID NO:	引物	序列
118	MH255	5'-AAGCCAAGAGCTTTGTGTCC-3'
119	MH256	5'-TATGCACGAGCTCTACGCCT-3'
120	MH257	5'-ATGGTACCCTGGCTATGGCT-3'
121	MH258	5'-CGGTCACGGTCTATCTTGGT-3'
122	pDonor正向	5'- GCTAGCATGGATGTTTTCCCAGTCACGACGTTGTAAAACGACGGC- 3'
123	Fv3C/Bgl3反向	5'-GGAGGTTGGAGAACTTGAACGTCGACCAAGATAGACCGTGACCGAAC TCGTAG-3'
124	pDonor反向	5'-TGCCAGGAAACAGCTATGACCATGTAATACGACTCACTATAGG-3'
125	Fv3C/Bgl3正向	5'- CTACGAGTTCGGTCACGGTCTATCTTGGTCGACGTTCAAGTTCTCCA ACCTCC-3
126	Att L1正向	5'-TAAGCTCGGGCCCCAAATAATGATTTTATTTTGACTGATAGT-3'
127	Attl2反向	5'-GGGATATCAGCTGGATGGCAAATAATGATTTTATTTTGACTGATA-3'
128	Cbh1正向	5'-GAGTTGTGAAGTCGGTAATCCCGCTG-3'
129	AmdS反向	5'-CCTGCACGAGGGCATCAAGCTCACTAACCG-3'
130	Pr Cbhl正向	5'-CGGAATGAGCTAGTAGGCAAAGTCAGC-3'
131	725/751反向	5'-CTCCTTGATGCGGCGAACGTTCTTGGGGAAGCCATAGTCCTTAAGGTTCTT GCTGAAGTTGCCCAGAGAG-3'
132	725/751正向	5'-GGCTTCCCCAAGAACGTTCGCCGCATCAAGGAGTTTATCTACCCCTACC TGAACACCACTACCTC-3'
133	Ter Cbhl反向	5'-GATACACGAAGAGCGGCGATTCTACGG-3'
134	正向MH234	5'-CACCATGAAGCTGAATTGGGTCGC-3'
160	Te3A反向	5'-GATAGACCGTGACCGAACTCGTAGATAGGCGTGATGTT GTACTTGTCGAAG TGACGGTAGTCGATGAAGAC-3'
161	Te3A2正向	5'-GTCTTCATCGACTACCGTCACTTCGACAAGTACAACATCAC GCCTATCTACG AGTTCGGTCACGGTCTATC-3'

图66B

SEQ ID NO:135

Fv3C/Te3A/里氏木霉Bg13 (FAB)嵌合体的氨基酸序列

mklnwvaaalsigaagtdsavalasavpdtlagvkkadaqkvvtrdtlayspphypspwmdpna
vgweeayakaksfvsqltlmekvnlttgvgwqgercvgnvgsiprlgmrglclqdgplgirlsd
ynsafpagttagaswskslwyergllmgtefkekgidialgpatgplgrtaaggrnwegftvdp
ymaghamaeavkgiqdagviacakhyianeqehfrqsgevqsrkyniseslssnlddktmhely
awpfadavragvgsvmcsynqinnsygcqnskllngilkdemgfqgfvmsdwaaqhtgaasava
gldmsmpgdtafdsgysfwggnltlavingtvpawrvddmalrimsaffkvgktiedlpdinfs
swtrdtfgfvhtfaqenreqvnfgvnvqhdhkshireaaakgsvvlkntgslplknpkflavig
edagpnpagpngcgdrgcdngtlamawgsgtsqfpylitpdqglsnratqdgtryesiltnnew
asvqalvsqpnvtaivfanadsgegyievdgnfgdrknltlwqqgdeliknvssicpntivvlh
tvgpvlladyeknpnitaivwaglpgqesgnaiadllygkvspgrspftwgrtresygtevlye
anngrgapqddfsegvfidyrhfd***KYNITPI***yefghglsWSTFKFSNLHIQKNNVGPMSPPNGK
TIAAPSLGNFSKNLKDYGFPKNVRRIKEFIYPYLNTTTSGKEASGDAHYGQTAKEFLPAGALDG
SPQPRSAASGEPGGNRQLYDILYTVTATITNTGSVMDDAVPQLYLSHGGPNEPPKVLRGFDRIE
RIAPGQSVTFKADLTRRDLSNWDTKKQQWVITDYPKTVYVGSSSRDLPLSARLP

图67A

SEQ ID NO :83

编码Fv3C/Te3A/里氏木霉Bg13 (FAB)嵌合体的核酸序列:

atgaagctgaattgggtcgccgcagccctgtctataggtgctgctggcactgacagcgcagttg
ctcttgcttctgcagttccagacactttggctggtgtaaaggtcagtttttttcaccatttcc
tcgtctaatctcagccttgttgccatatcgcccttgttcgctcggacgccacgcaccagatcgc
gatcatttcctcccttgcagccttggttcctcttacgatcttcctccgcaattatcagcgccc
ttagtctacacaaaaacccccgagacagtctttcattgagtttgtcgacatcaagttgcttctc
aactgtgcatttgcgtggctgtctacttctgcctctagacaaccaaatctgggcgcaattgacc
gctcaaaccttgttcaaataacctttttattcgagacgcacatttataaatatgcgcctttca
ataataccgactttatgcgcggcggctgctgtggcggttgatcagaaagctgacgctcaaaagg
ttgtcacgagagatacactcgcatactcgccgcctcattatccttcaccatggatggaccctaa
tgctgttggctgggaggaagcttacgccaaagccaagagctttgtgtcccaactcactctcatg
gaaaaggtcaacttgaccactggtgttgggtaagcagctccttgcaaacagggtatctcaatcc
cctcagctaacaacttctcagatggcaaggcgaacgctgtgtaggaaacgtgggatcaattcct
cgtctcggtatgcgaggtctctgtctccaggatggtcctcttggaattcgtctgtccgactaca
acagcgcttttcccgctggcaccacagctggtgcttcttggagcaagtctctctggtatgagag
aggtctcctgatgggcactgagttcaaggagaagggtatcgatatcgctcttggtcctgctact
ggacctcttggtcgcactgctgctggtggacgaaactgggaaggcttcaccgttgatccttata
tggctggccacgccatggccgaggccgtcaagggtattcaagacgcaggtgtcattgcttgtgc
taagcattacatcgcaaacgagcagggtaagccacttggacgatttgaggaattgacagagaac
tgaccctcttgtagagcacttccgacagagtggcgaggtccagtcccgcaagtacaacatctcc
gagtctctctcctccaacctggatgacaagactatgcacgagctctacgcctggcccttcgctg
acgccgtccgcgccggcgtcggttccgtcatgtgctcgtacaaccagatcaacaactcgtacgg
ttgccagaactccaagctcctcaacggtatcctcaaggacgagatgggcttccagggtttcgtc
atgagcgattgggcggcccagcataccggtgccgcttctgccgtcgctggtctcgatatgagca
tgcctggtgacactgccttcgacagcggatacagcttctggggcggaaacttgactctggctgt
catcaacggaactgttcccgcctggcgagttgatgacatggctctgcgaatcatgtctgccttc
ttcaaggttggaaagacgatagaggatcttcccgacatcaacttctcctcctggacccgcgaca
ccttcggcttcgtgcatacatttgctcaagagaaccgcgagcaggtcaactttggagtcaacgt
ccagcacgaccacaagagccacatccgtgaggccgctgccaagggaagcgtcgtgctcaagaac
accgggtcccttcccctcaagaacccaaagttcctcgctgtcattggtgaggacgccggtccca
accctgctggacccaatggttgtggtgaccgtggttgcgataatggtaccctggctatggcttg
gggctcgggaacttcccaattcccttacttgatcacccccgatcaagggctctctaatcgagct
actcaagacggaactcgatatgagagcatcttgaccaacaacgaatgggcttcagtacaagctc
ttgtcagccagcctaacgtgaccgctatcgttttcgccaatgccgactctggtgagggatacat
tgaagtcgacggaaactttggtgatcgcaagaacctcaccctctggcagcagggagacgagctc
atcaagaacgtgtcgtccatatgccccaacaccattgtagttctgcacaccgtcggccctgtcc
tactcgccgactacgagaagaaccccaacatcactgccatcgtctgggctggtcttcccggcca
agagtcaggcaatgccatcgctgatctcctctacggcaaggtcagccctggccgatctcccttc
acttggggccgcacccgcgagagctacggtactgaggttctttatgaggcgaacaacggccgtg
gcgctcctcaggatgacttctctgagggtgtcttcatcgactaccgtcacttcgacaagtacaa
catcacgcctatctacgagttcggtcacggtctatcttggtcgacgttcaagttctccaacctc
cacatccagaagaacaatgtcggccccatgagcccgcccaacggcaagacgattgcggctccct
ctctgggcaacttcagcaagaaccttaaggactatggcttccccaagaacgttcgccgcatcaa
ggagtttatctacccctacctgaacaccactacctctggcaaggaggcgtcgggtgacgctcac
tacggccagactgcgaaggagttcctccccgccggtgccctggacggcagccctcagcctcgct
ctgcggcctctggcgaacccggcggcaaccgccagctgtacgacattctctacaccgtgacggc
caccattaccaacacgggctcggtcatggacgacgccgttccccagctgtacctgagccacggc
ggtcccaacgagccgcccaaggtgctgcgtggcttcgaccgcatcgagcgcattgctcccggcc
agagcgtcacgttcaaggcagacctgacgcgccgtgacctgtccaactgggacacgaagaagca
gcagtgggtcattaccgactaccccaagactgtgtacgtgggcagctcctcgcgcgacctgccg
ctgagcgcccgcctgccatga

图67B

SEQ ID NO:	末端	序列基序
136	N	A-x-S-P-P-x-Y-P-S-P-W-M-D-P-x-A-x-G-W-E-x-A-Y
137	N	A-K-x-F-V-S-x-x-T-L-x-E-K-V-N-L-T-T-G-V-G-W-x-G-E-x-C-V-G-N-V-G
138	N	P-R-x-G-M-R-x-L-C-x-Q-D-G-P-L-G-x-R
139	N	Y-N-S-A-F-x-x-G-x-T-A-x-A-S-W-S
140	N	G-x-I-A-C-A-K-H-x-x-x-N-E-Q-E-H-x-R-Q
141	N	L-S-S-N-x-D-D-K-T-x-H-E-x-Y-x-W-P-F-x-D-A-V-x-A-G-V-G
142	N	M-C-S-Y-x-Q-x-N-N-S-Y-x-C-Q-N-S-K-L-x-N-G
143	N	G-F-Q-G-F-V-M-S-D-W-x-A-Q-H-x-G-x-A-x-A-V-A-G-L-D-M-x-M-P-G-D-T
144	N	N-L-T-L-A-V-x-N-G-T-V-P-x-W-R-x-D-D-M
145	N	P-x-F-L-x-V-x-G-E-D-A-G-x-N-P-A-G-P-N-G-C-x-D-R-G-C
146	N	G-T-L-A-M-x-W-G-S-G-T-x-F-P-Y-L
147	N	A-I-V-F-A-N-x-x-S-G-E-G-Y-I-x-V-D-G-N-x-G-D-R-K-N-L-T-L-W
148	N	D-x-L-Y-G-K-x-S-P-G-R-x-P-F-T-W-G
149	C	P-x-Y-E-F-G-x-G-L-S-W-x-T-F-x-x-S-x-L
150	C	L-x-D-Y-x-F-P
151	C	E-F-L-P-x-x-A-L-x-G-S-x-Q-P-R
152	C	S-G-x-P-G-G-N-x-x-L-x-D
153	C	Y-T-V-x-A-x-I-T-N-T-G
154	C	V-L-R-G-F-x-R-x-E-x-I-A-P-G-x-S
155	C	T-R-R-D-L-S-N-W-D-x-x-x-Q-x-W-V-I-T-D
156	C	V-G-S-S-S-R-x-L-P-L-x-A-x-L

图68A

用于设计本发明的适当β-葡糖苷酶
多肽杂合体/嵌合体的氨基酸序列基序

SEQ ID NO:	末端	序列基序
164	N	Y-P-S-P-W-M-D-P
165	N	E-K-V-N-L-T-T-G-V-G-W
166	N	K-G-(I/V)-D-(V/I)
167	N	C-Q-N-S-K-L-x-N-G
168	N	N-L-T-L-A-V-(L/I/V)-N-G-(S/T)-(V/I)-P-x-W
169	N	S-W-(T/S)-x-D-T-(Y/F)-G
170	C	E-F-L-P-x-x-A-L-x-G-S-x-Q-P-R

图68B

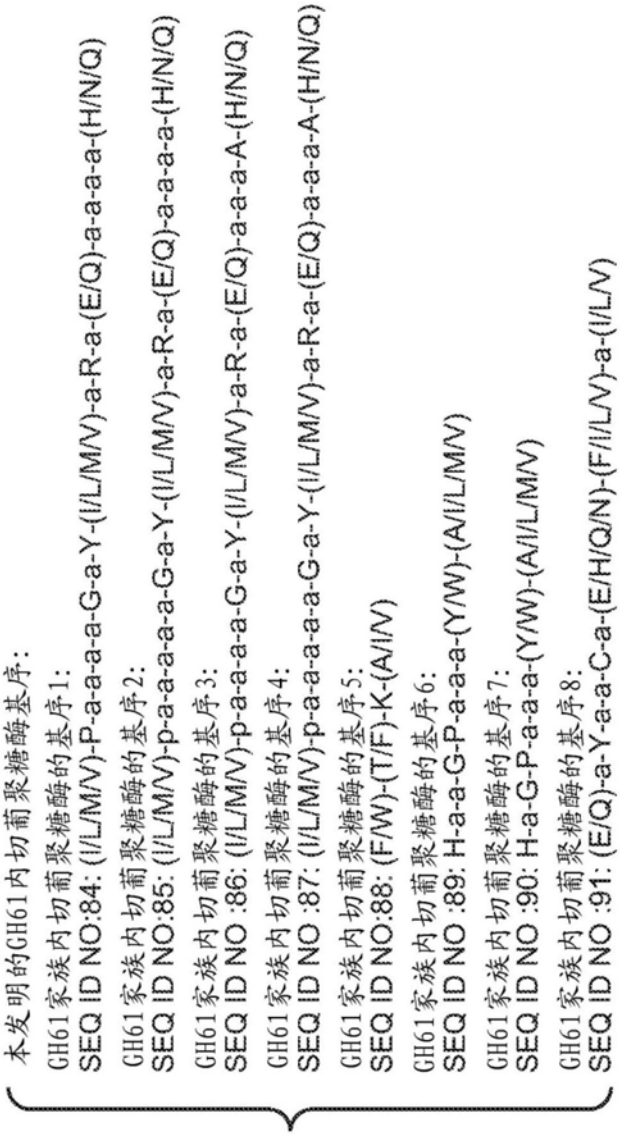

本发明的GH61内切葡聚糖酶基序:
GH61家族内切葡聚糖酶的基序1:
SEQ ID NO:84: (I/L/M/V)-P-a-a-a-a-G-a-Y-(I/L/M/V)-a-R-a-(E/Q)-a-a-a-a-(H/N/Q)
GH61家族内切葡聚糖酶的基序2:
SEQ ID NO:85: (I/L/M/V)-p-a-a-a-a-G-a-Y-(I/L/M/V)-a-R-a-(E/Q)-a-a-a-a-(H/N/Q)
GH61家族内切葡聚糖酶的基序3:
SEQ ID NO :86: (I/L/M/V)-p-a-a-a-a-G-a-Y-(I/L/M/V)-a-R-a-(E/Q)-a-a-a-A-(H/N/Q)
GH61家族内切葡聚糖酶的基序4:
SEQ ID NO :87: (I/L/M/V)-p-a-a-a-a-G-a-Y-(I/L/M/V)-a-R-a-(E/Q)-a-a-a-A-(H/N/Q)
GH61家族内切葡聚糖酶的基序5:
SEQ ID NO:88: (F/W)-(T/F)-K-(A/I/V)
GH61家族内切葡聚糖酶的基序6:
SEQ ID NO :89: H-a-a-G-P-a-a-a-(Y/W)-(A/I/L/M/V)
GH61家族内切葡聚糖酶的基序7:
SEQ ID NO :90: H-a-G-P-a-a-a-(Y/W)-(A/I/L/M/V)
GH61家族内切葡聚糖酶的基序8:
SEQ ID NO :91: (E/Q)-a-Y-a-a-C-a-(E/H/Q/N)-(F/I/L/V)-a-(I/L/V)

图68C

SEQ NO. ID : 81

鹅掌柄孢霉Pa 3C的核苷酸序列

atggcataccgctcattagtcttgggcgccttcgcctccacctctcttgccgccagcgtcgtgacgcct
cgagatcctgttccgcctggattcgtcgctgccccatactatccagcgcctcatggaggatgggtcgct
tcgtgggaagaggcttacagcaaggccgaagccttggtctcgcagatgaccttggctgaaaagaccaac
atcacctcaggcattggcatctttatgggtgagttattaaccagacatggcttatataaaagcacaaga
gactgactgacatgtgaatagggtcagtgccaccaccctaatgagacgtttttctgattttgactaaca
catgatacgctagtccatgcgtaggaaatactggaagcgcagaaagattggggttcccgcgcatgtgtc
ttcaggactctgcgttgggtgtgtcgtcggctgacaacgtcactgcgtttcctgctggcatcaccactg
gtgcaacgtttgacaagaagctgatctatgctcgtggtgttgctattggtgaagagcatcgcggcaagg
gcacaaatgtctatctgggtccttccgtaggccctcttgggcggaagcctttgggtggccgcaactggg
agggctttggatctgacccagttcttcaagccaaggctgctgccctgacgatcaagggcgttcaggaac
aaggcatcattgctactatcaagcatctgatcggcaacgagcaggagatgtatagaatgtacaaccccct
tccagcctggatatagcgccaatattggtgagtggactcttgctctttgacggactaaaaggctgactc
cccacagatgatcggactctgcacgagctctacctgtggccctttgccgaatccgtccatgccggtgtt
gggtcggcaatgacagcttacaatgctgtaaacgggtctgcttgctctcagcacagctatctcatcaac
ggtattttgaaggatgagcttggattccagggcttcgtcatgtctgactggctgtcccacatctccgga
gtcgactccgcgttggcaggtctcgacatgaacatgccaggtgacaccaacattcccctatttggtttc
agcaactggcactatgagctcagcagatcggttctcaacgggtctgtgcctcttgacagactgaacgac
atggtcaccagaatcgtcgcgacatggtacaagttcggtcaggatagggaccacccaaggcctaacttc
tcgtcaaacacccgtgaccgtgacggtctgctttatcctgcagctcttctccccaagggtcaggtg
aactggtttgtcaatgttcaggctgatcattatttgatcgccagagaggtcgcccaggatgccatcacc
cttctcaagaacaatgggagcttccttcccctgacgacttcgcagtctctccatgtcttcggtactgct
gcccaggtcaaccccgatgggcccaacgcttgcatgaaccgcgcctgcaacaaaggaacacttggcatg
ggctggggttctggtgttgccgattatccttacttggatgacccgatctcggctatcaggaagcgggtt
cccgacgtcaagttcttcaacaccgacggcttcccttggttccaccctacaccgtcgcccgatgacgtt
gccatcgtgttcatcacctccgatgctggagagaactcgttcactgttgagggcaacaacggtgatcgc
aacagtgccaagctggctgcgtggcataacggtgacgagctggtcaggaagactgccgagaagtacaac
aacgttattgtggtagctcaaaccgtcggccctctcgatctcgaatcctggatcgacaaccctcgcgtc
aagggcgtcctgtttcagcaccttcccggtcaagaagcgggcgagtcgttggccaacattctctttggc
gatgtctcccctagcggtcaccttccctactccatcaccaagcgcgccaacgacttccccgacagcatc
gccaacctccgtggctttgcctttggtcaggtccaggacacgtacagcgagggcctgtacattgactac
cgctggctcaacaaggagaagatcaggccccgctttgcttttggccacggtctcagctacaccaacttc
tcgtttgatgccaccatcgagtctgtcactccactgtctctggttcctcctgcccgtgcccccaagggc
tcaacgccggtgtactcgaccgaaatcccccccgcctcagaggcgtactggccggaagggttcaacagg
atctggcggtaccctctactcctggctcaacaagaacgacgcggataacgcctacgctgttggtatcgcc
ggggtgaagaagtataactatcccgctgggtacagcaccgcccagaagcccggtcccgcagccggtggc
ggggagggggtaatcctgcgctttgggatattgctttccgtgtcccagttacggtcaagaacactggg
gatacgttctcgggacgggcttcggtgcaggcttatgttcagtatcctgaggggatcccgtatgatacg
cctgttgtgcagctgagggactttgagaagacgagggttttggctccgggggaggaggagacggtgacg
gttgagctgaccaggaaggacttgagcgtgtgggacacggagctgcagaactgggttgtgccgggggtt
gggggaagaggtatacggtttggattggggaggcgagcgataggttgtttacggcttgttatacggat
acgggggtttgtgaggggggagggtgccgcctgtttaa

图69A

SEQ NO. ID : 80

鹅掌柄孢霉Pa3C的蛋白质序列

MAYRSLVLGAFASTSLAASVVTPRDPVPPGFVAAPYYPAPHGGWVASWEEAYSKAEALVSQMTLAEKTN
ITSGIGIFMGNTGSAERLGFPRMCLQDSALGVSSADNVTAFPAGITTGATFDKKLIYARGVAIGEEHRG
KGTNVYLGPSVGPLGRKPLGGRNWEGFGSDPVLQAKAAALTIKGVQEQGIIATIKHLIGNEQEMYRMYN
PFQPGYSANIDDRTLHELYLWPFAESVHAGVGSAMTAYNAVNGSACSQHSYLINGILKDELGFQGFVMS
DWLSHISGVDSALAGLDMNMPGDTNIPLFGFSNWHYELSRSVLNGSVPLDRLNDMVTRIVATWYKFGQD
RDHPRPNFSSNTRDRDGLLYPAALFSPKGQVNWFVNVQADHYLIAREVAQDAITLLKNNGSFLPLTTSQ
SLHVFGTAAQVNPDGPNACMNRACNKGTLGMGWGSGVADYPYLDDPISAIRKRVPDVKFFNTDGFPWFH
PTPSPDDVAIVFITSDAGENSFTVEGNNGDRNSAKLAAWHNGDELVRKTAEKYNNVIVVAQTVGPLDLE
SWIDNPRVKGVLFQHLPGQEAGESLANILFGDVSPSGHLPYSITKRANDFPDSIANLRGFAFGQVQDTY
SEGLYIDYRWLNKEKIRPRFAFGHGLSYTNFSFDATIESVTPLSLVPPARAPKGSTPVYSTEIPPASEA
YWPEGFNRIWRYLYSWLNKNDADNAYAVGIAGVKKYNYPAGYSTAQKPGPAAGGGEGGNPALWDIAFRV
PVTVKNTGDTFSGRASVQAYVQYPEGIPYDTPVVQLRDFEKTRVLAPGEEETVTVELTRKDLSVWDTEL
QNWVVPGVGGKRYTVWIGEASDRLFTACYTDTGVCEGGRVPPV*

图69B

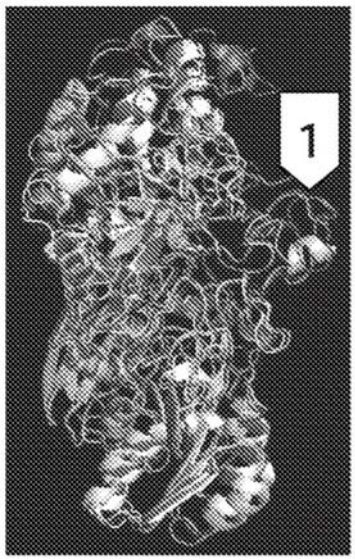


图70A

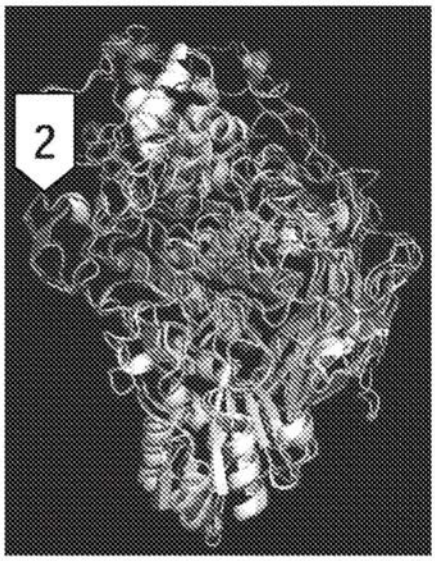


图70B

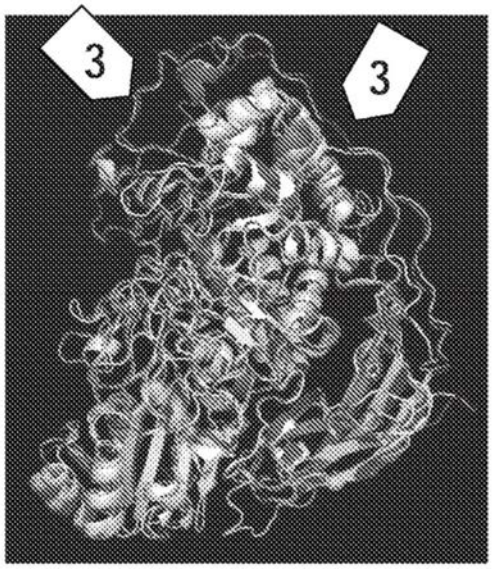


图70C

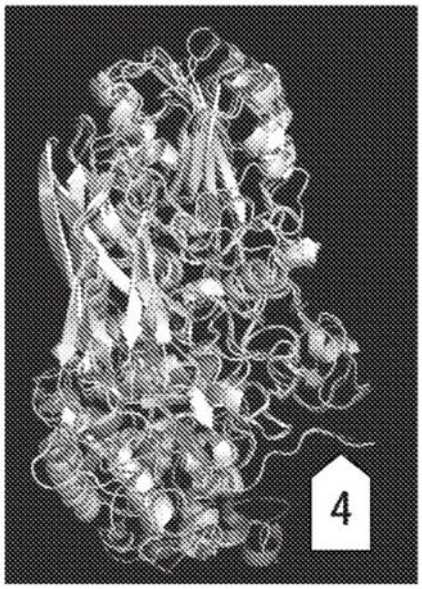


图70D

里氏木霉BGL1、Te3A和Fv3C的比对结果。示出了成熟的里氏木霉BGL1序列，成熟Fv3C和ABG2的起始点以'Mat'标出

```
                                  *         20        *         40        *         60        *
Q12715_TRI : ------------------------------------------------------------VVPAGTPWGTA :  12
ABG2_T_eme : ------------------------MRNGLLKVAALAAASAVNGENLAYSPPFYPSPWANGQG-DWAEA :  43
Fv3C       : MKLNWVAAALSIGAAGTDSAVALASAVPDTLAGVKKADAQKVVTRDTLAYSPPHYPSPWMDPNAVGWEEA :  70
                                                                  Mat  Ins4
                                  80        *        100        *        120        *        140
Q12715_TRI : YDKAKAALAKLNLQDKVGIVSGVGWNGGPCVGNTSPASKISYPSLCLQDGPLGVRYSTGSTAFTPGVQAA :  82
ABG2_T_eme : YQKAVQFVSQLTLAEKVNLTTGTGWEQDRCVGQVGSIPRLGFPGLCMQDSPLGVRDTDYNSAFPAGVNVA : 113
Fv3C       : YAKAKSFVSQLTLMEKVNLTTGVGWQGERCVGNVGSIPRLGMRGLCLQDGPLGIRLSDYNSAFPAGTTAG : 140
                                  *        160        *        180        *        200        *
Q12715_TRI : STWDVNLIRERGQFIGEEVKASGIHVILGPVAGPLGKTPQGGRNWEGFGVDPYLTGIAMGQTINGIQSVG : 152
ABG2_T_eme : ATWDRNLAYRRGVAMGEEHRGKGVDVQLGPVAGPLGRSPDAGRNWEGFAPDPVLTGNMMASTIQGIQDAG : 183
Fv3C       : ASWSKSLWYERGLLMGTEFKEKGIDIALGPATGPLGRTAAGGRNWEGFTVDPYMAGHAMAEAVKGIQDAG : 210
                                  220       *        240       *        260       *        280
Q12715_TRI : VQATAKHYILNEQELNR----------------ETISSNPDDRTLHELYTWPFADAVQANVASVMCSYNKVNT : 209
ABG2_T_eme : VIACAKHFILYEQEHFRQGAQ---DGYDISDSISANADDKTMHELYLWPFADAVRAGVGSVMCSYNQVNN : 250
Fv3C       : VIACAKHYIANEQEHFRQSGEVQSRKYNISESLSSNLDDKTMHELYAWPFADAVRAGVGSVMCSYNQINN : 280
                                                  插入物1
                                  *        300       *        320       *        340       *
Q12715_TRI : TWACEDQYTLQTVLKDQLGFPGYVMTDWNAQHTTVQSANSGLDMSMPG-TDFNGNNRLWGPALTNAVNSN : 278
ABG2_T_eme : SYACSNSYTMNKLLKSELGFQGFVMTDWGGHHSGVGSALAGLDMSMPGDIAFDSGTSFWGTNLTVAVLNG : 320
Fv3C       : SYGCQNSKLLNGILKDEMGFQGFVMSDWAAQHTGAASAVAGLDMSMPGDTAFDSGYSFWGGNLTLAVING : 350
                                  360       *        380       *        400       *        420
Q12715_TRI : QVPTSRVDDMVTRILAAWYLTGQDQAGYPSFNISR----------------------NVQGNHKTNVR : 324
ABG2_T_eme : SIPEWRVDDMAVRIMSAYYKVGRDRYSVP-INFDSWTLDTYGPEHYAVGQGQTKINEHVDVRGNHAEIIH : 389
Fv3C       : TVPAWRVDDMALRIMSAFFKVGKTIEDLPDINFSSWTRDTFGFVHTFAQENREQVNFGVNVQHDHKSHIR : 420
                                                                  插入物2
                                  *        440       *        460       *        480       *
Q12715_TRI : AIARDGIVLLKNDANILPLK-KPASIAVVGSAAIIGNHARNSPSCNDKGCDDGALGMGWGSGAVNYPYFV : 393
ABG2_T_eme : EIGAASAVLLKNKGG-LPLTGTERFVGVFGKDA--GSNPWGVNGCSDRGCDNGTLAMGWGSGTANFPYLV : 456
Fv3C       : EAAAKGSVVLKNTGS-LPLK-NPKFLAVIGEDA--GPNPAGPNGCGDRGCDNGTLAMAWGSGTSQFPYLI : 486
```

图70E-1

里氏木霉BGL1、Te3A和Fv3C的比对结果。示出了成熟的里
氏木霉BGL1序列，成熟Fv3C和ABG2的起始点以‘Mat’标出

```
                500       *       520       *       540       *       560
Q12715_TRI : APYDAINTRASSQGTQVT--LSNTDNTSSGASAARGKDVAIVFITADSGEGYITVEGNAGDRNNLDPWHN : 461
ABG2_T_eme : TPEQAIQREVLSRNGTFTG-ITDNGALAEMAAAASQADTCLVFANADSGEGYITVDGNEGDRKNLTLWQG : 525
Fv3C___ : TPDQGLSNRATQDGTRYESILTNNEWASVQALVSQPNVTAIVFANADSGEGYIEVDGNFGDRKNLTLWQQ : 556
                    *       580       *       600       *       620       *
Q12715_TRI : GNALVQAVAGANSNVIVVVHSVGAILLEQILALPQVKAVVWAGLPSQESGNALVDVLWGDVSPSGKLVYT : 531
ABG2_T_eme : ADQVIHNVSANCNNTVVVLHTVGPVLIDDWYDHPNVTAILWAGLPGQESGNSLVDVLYGRVNP-GKTPFT : 594
Fv3C___ : GDELIKNVSSICPNTIVVLHTVGPVLLADYEKNPNITAIVWAGLPGQESGNAIADLLYGKVSP-GRSPFT : 625
                640       *       660       *       680       *       700
Q12715_TRI : IAKSPNDYNTRIVSGGSDS-------FSEGLFIDYKHFDDAN----------ITPRYEFGYGLSYTKF : 582
ABG2_T_eme : WGRARDDYGAPLIVKPNNGKGAPQQDFTEGIFIDYRRFDKYN----------ITPIYEFGFGLSYTTF : 652
Fv3C___ : WGRTRESYGTEVLYEANNGRGAPQDDFSEGVFIDYRHFDRRSPSTDGKSSPNNTAAPLYEFGHGLSWSTF : 695
                    *       720       *       740       *       760       *
Q12715_TRI : NYSRLSV----------LSTAKSGPA : 598
ABG2_T_eme : EFSQLNVQ-PINAPPYTPASGFTKAAQSFGQ-PSNASDNLYPSDIERVPLYIYPWLN--STDLKASANDPD : 719
Fv3C___ : EYSDLNIQ-KNVENPYSPPAGQTIPAPTFGNFSKNLNDYVFPKGVRYIYKFTYPFLNTSSSASEASNDGGQ : 765
                                                插入物3
                780       *       800       *       820       *       840
Q12715_TRI : TGAVV----------PGGPSDLFQNVATVTVDIANSGQVTGAEVAQLYITYPSSAPRT : 646
ABG2_T_eme : YGLPTEKYVPPNATNGDPQPIDPAGGAPGGNPSLYEPVARVTTIITNTGKVTGDEVPQLYVSL--GGPDD : 787
Fv3C___ : FGKTAEEFLPPNALNGSAQPRLPASGAPGGNPQLWDILYTVTATITNTGNATSDEIPQLYVSL--GGENE : 833
                                    环3
                    *       860       *       880       *       900       *
Q12715_TRI : PPKQLRGFAKL-NLTPGQSGTATFNIRRRDLSYWDTASQKWVVPSGSFGISVGASSRDIRLTSTLSVA-- : 713
ABG2_T_eme : APKVLRGFDRI-TLAPGQQYLWTTTLTRRDISNWDPVTQNWVVTNYTKTIYVGNSSRNLPLQAPLKPYPG : 856
Fv3C___ : PIRVLRGFDRIENIAPGQSAIFNAQLTRRDLSNWDTNAQNWVITDHPKTVWVGSSSRKLPLSAKLE---- : 899

Q12715_TRI : - : -
ABG2_T_eme : I : 857
Fv3C___ : - : -
```

图70E-2

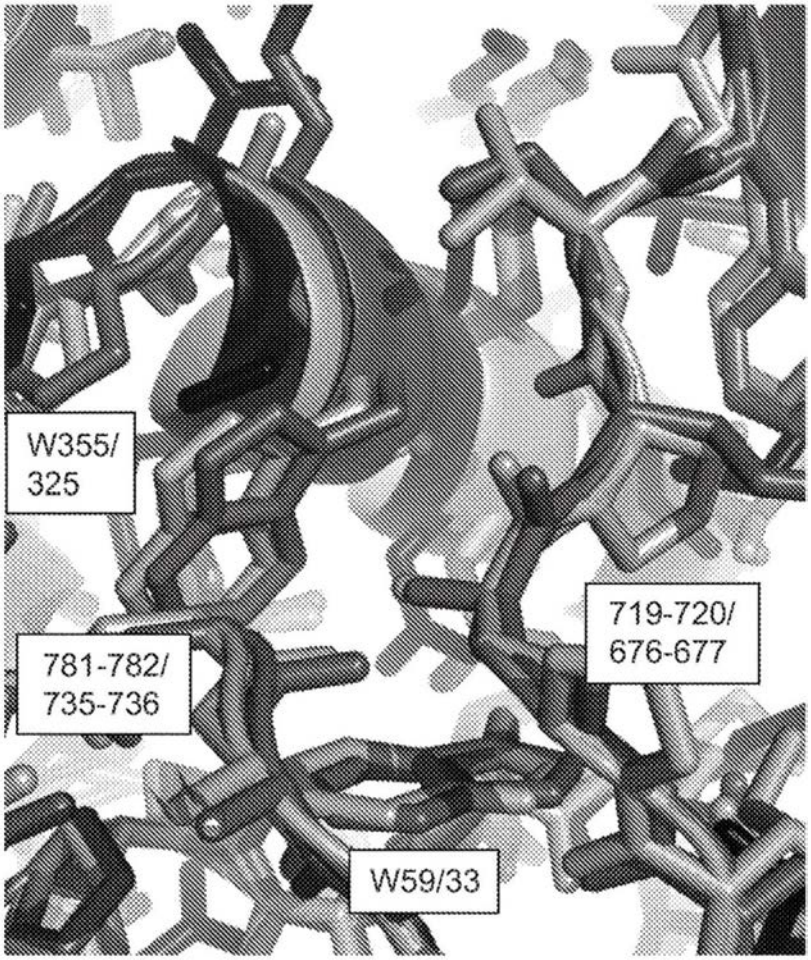


图70F

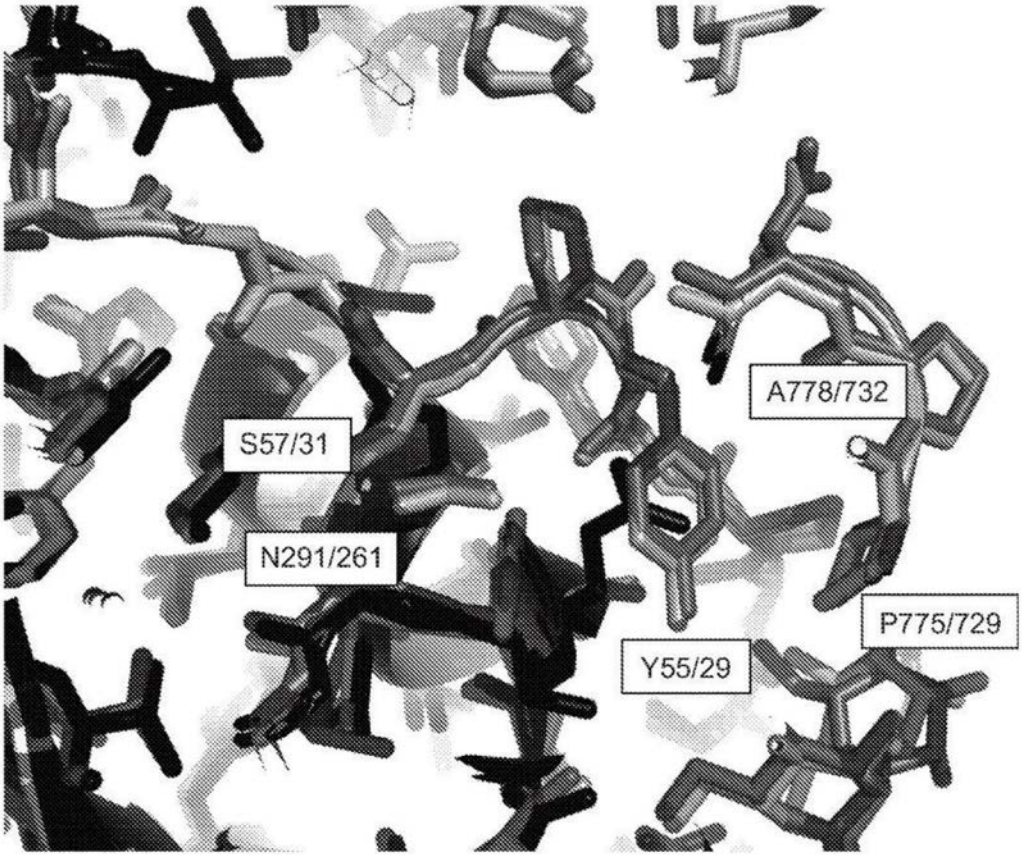


图70G

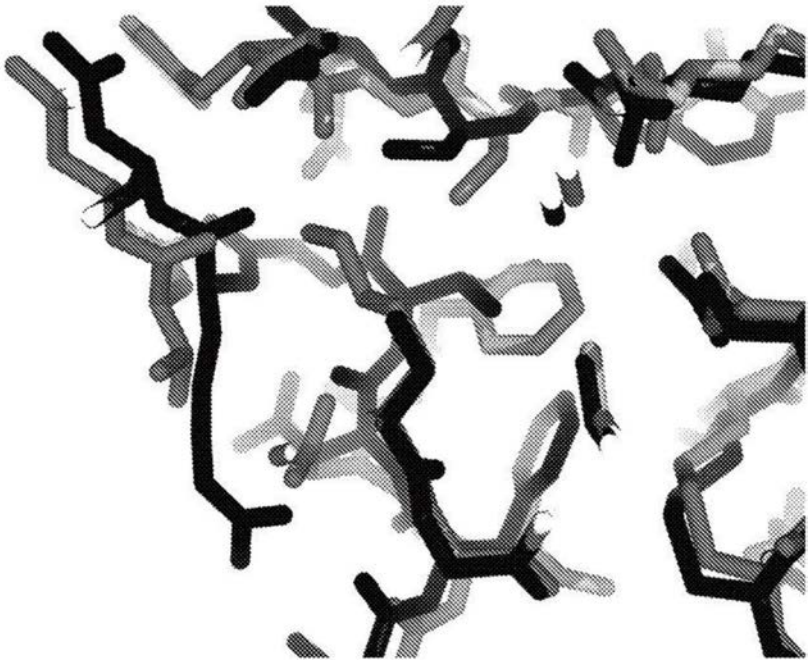


图70H

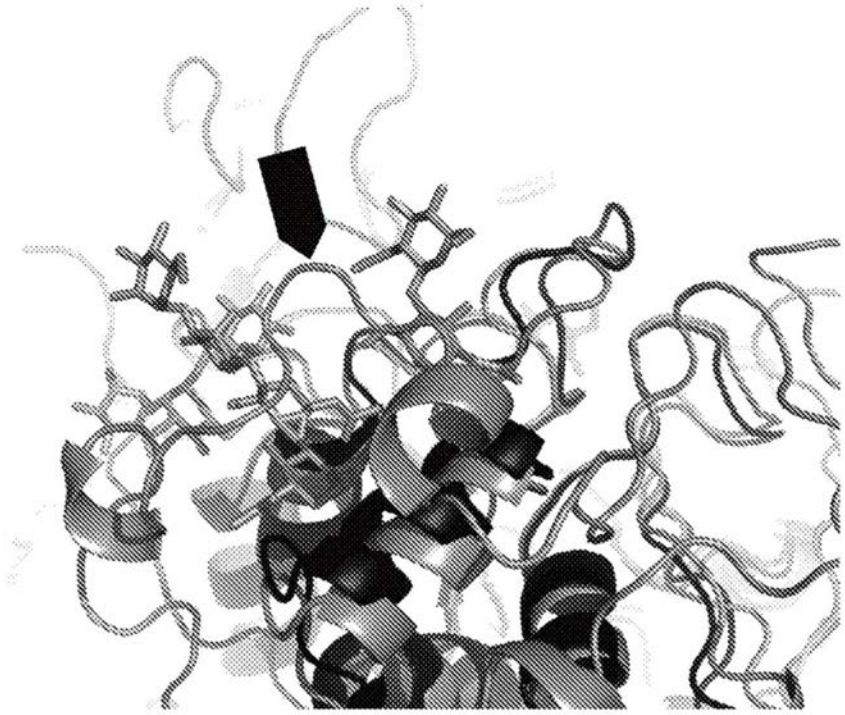


图70I(a)

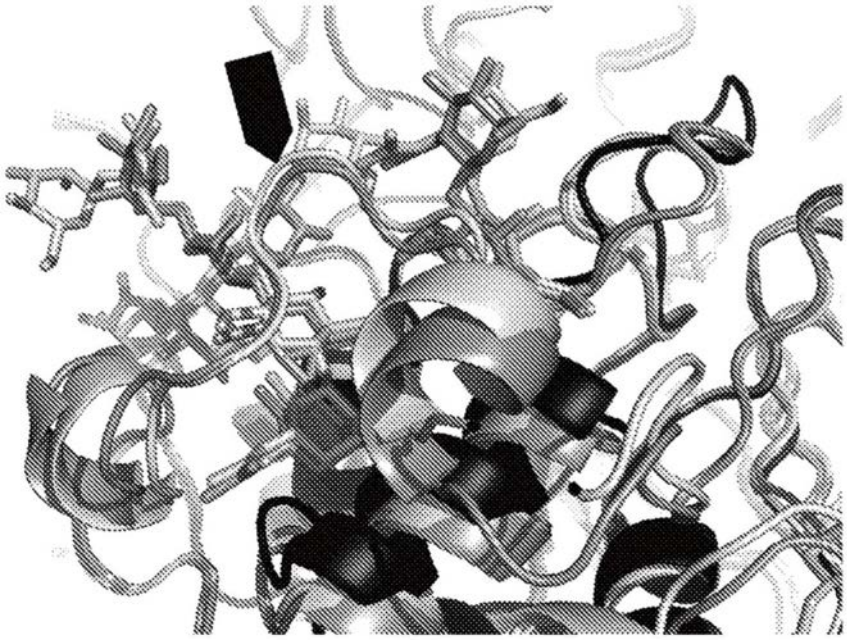


图70I(b)

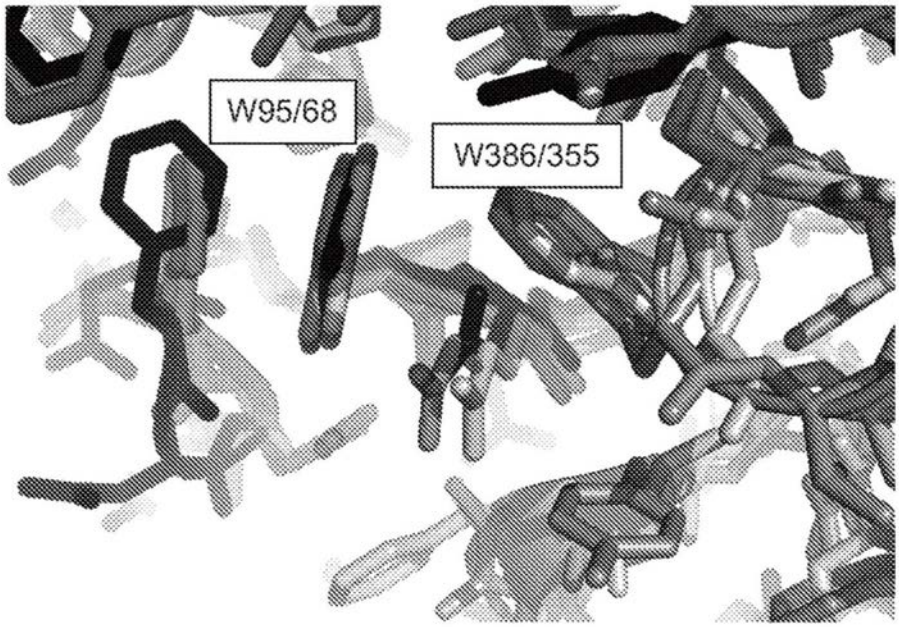

W95/68
W386/355

图70J

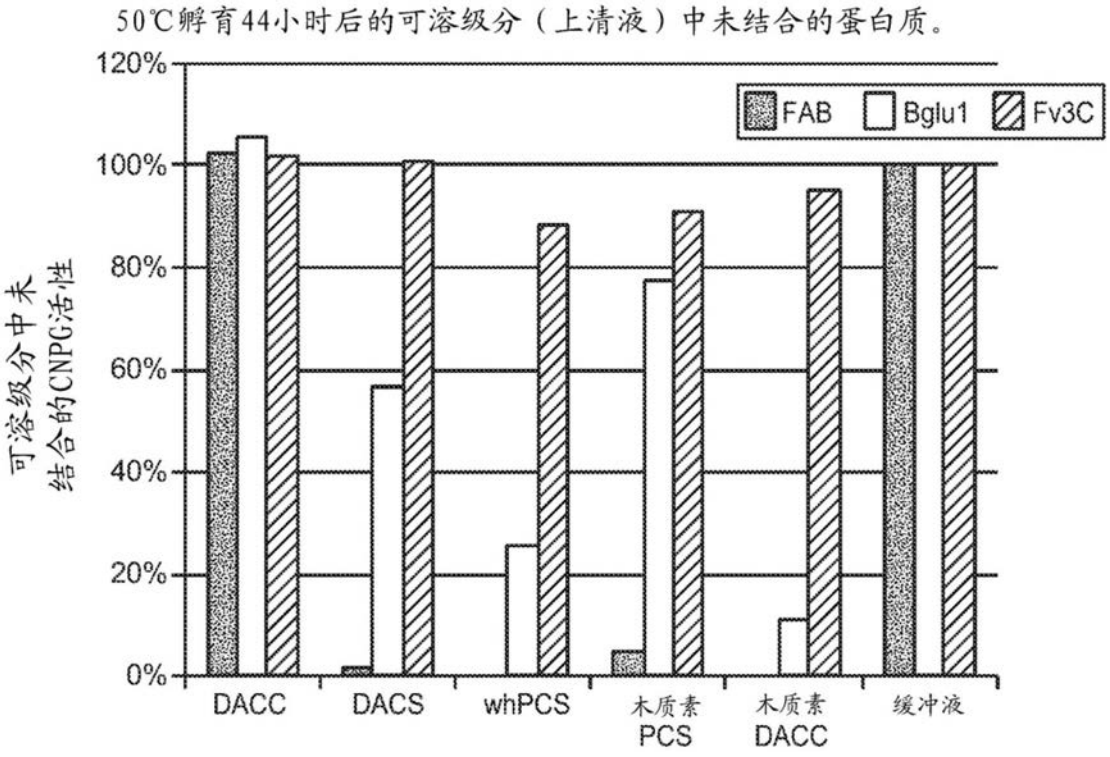

50℃孵育44小时后的可溶级分（上清液）中未结合的蛋白质。
可溶级分中未结合的CNPG活性
120%
100%
80%
60%
40%
20%
0%
FAB
Bglu1
Fv3C
DACC
DACS
whPCS
木质素PCS
木质素DACC
缓冲液

图71A

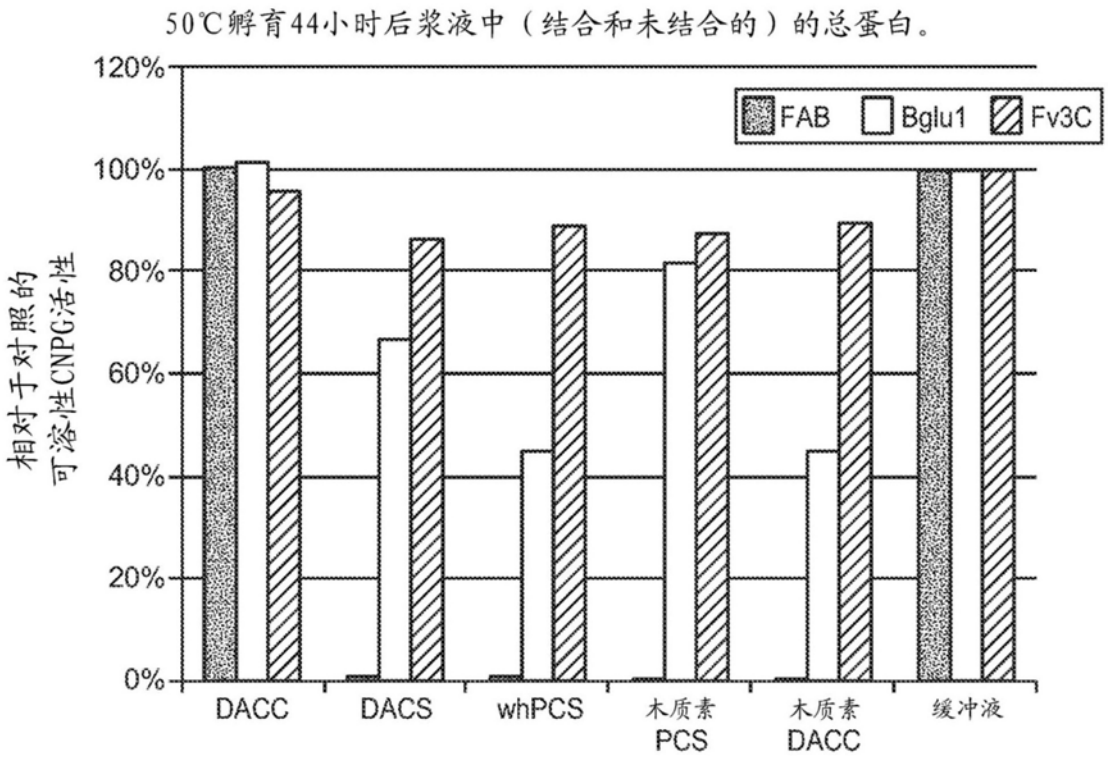

50℃孵育44小时后浆液中（结合和未结合的）的总蛋白。
120%
100%
80%
60%
40%
20%
0%
相对于对照的
可溶性CNPG活性
FAB
Bglu1
Fv3C
DACC
DACS
whPCS
木质素
PCS
木质素
DACC
缓冲液

图71B

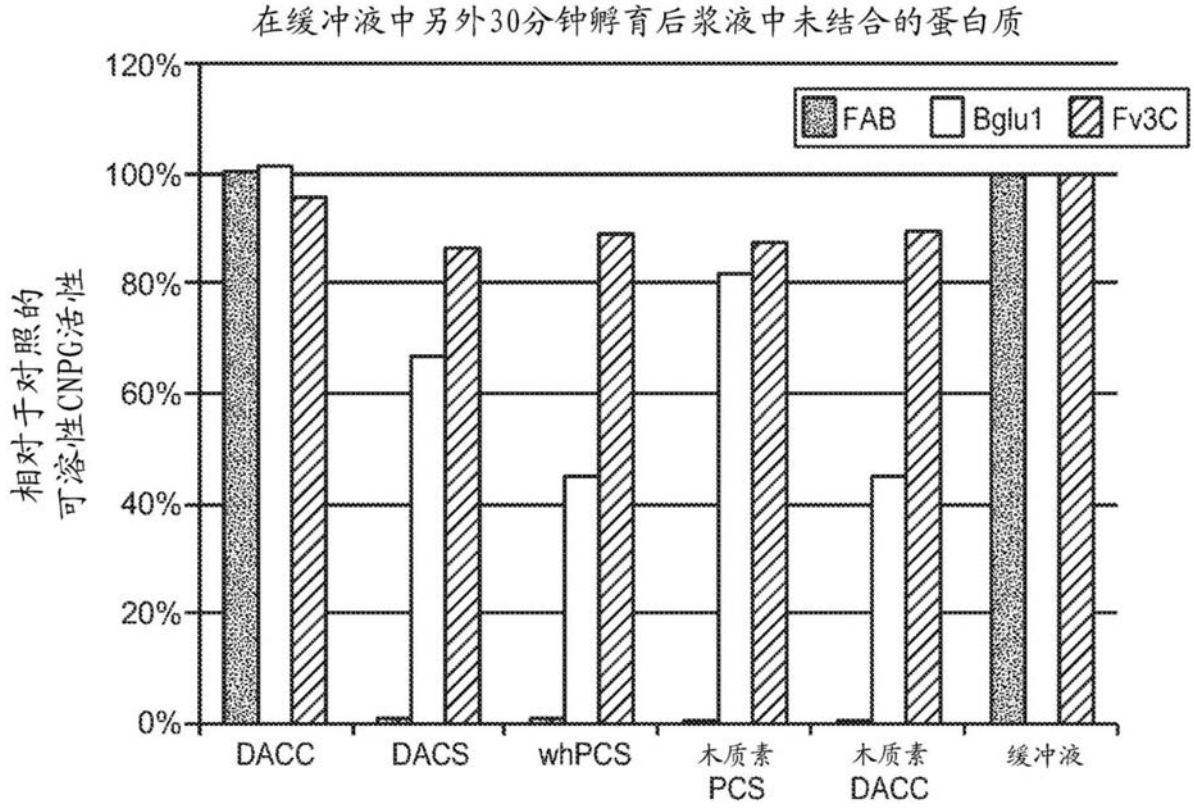

在缓冲液中另外30分钟孵育后浆液中未结合的蛋白质
120%
100%
80%
60%
40%
20%
0%
相对于对照的
可溶性CNPG活性
FAB
Bglu1
Fv3C
DACC
DACS
whPCS
木质素
PCS
木质素
DACC
缓冲液

图71C