



(12)发明专利申请

(10)申请公布号 CN 106085988 A

(43)申请公布日 2016.11.09

(21)申请号 201610349186.0

NRRL B-50324 2009.09.18

(22)申请日 2010.09.16

NRRL B-50325 2009.09.18

NRRL B-50326 2009.09.18

(30)优先权数据

61/243,397 2009.09.17 US

61/243,531 2009.09.18 US

61/243,543 2009.09.18 US

61/243,679 2009.09.18 US

(71)申请人 诺维信股份有限公司

地址 美国加利福尼亚州

(72)发明人 S.梅于兰 R.克雷默 P.哈里斯

(74)专利代理机构 北京坤瑞律师事务所 11494

代理人 史悦

(62)分案原申请数据

201080052019.0 2010.09.16

(51)Int.Cl.

C12N 9/42(2006.01)

C12N 15/56(2006.01)

C12N 5/10(2006.01)

C12P 19/14(2006.01)

C12P 19/02(2006.01)

C12P 7/10(2006.01)

C11D 3/386(2006.01)

(83)生物保藏信息

NRRL B-50300 2009.07.15

NRRL B-50301 2009.07.15

NRRL B-50302 2009.07.15

NRRL B-50303 2009.07.15

NRRL B-50320 2009.09.18

NRRL B-50321 2009.09.18

NRRL B-50322 2009.09.18

NRRL B-50323 2009.09.18

权利要求书2页 说明书83页

序列表19页 附图13页

(54)发明名称

具有纤维素分解增强活性的多肽及编码其的多核苷酸

(57)摘要

本发明涉及具有纤维素分解增强活性的多肽及编码其的多核苷酸,具体地涉及具有纤维素分解增强活性的分离的多肽和编码所述多肽的分离的多核苷酸。本发明还涉及包含所述多核苷酸的核酸构建体、载体和宿主细胞以及产生和使用所述多肽的方法。

1. 一种具有纤维素分解增强活性的分离的多肽,其选自下组:

(a)多肽,其与SEQ ID NO:20的成熟多肽具有至少90%,例如,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%序列同一性;

(b)多肽,其由多核苷酸编码,所述多核苷酸在非常高严格条件下与SEQ ID NO:19的成熟多肽编码序列或其cDNA序列,或它们的全长互补链杂交,其中所述非常高严格条件定义为:在42℃,在5X SSPE、0.3%SDS、200微克/ml已剪切并且变性的鲑精DNA和50%的甲酰胺中进行预杂交和杂交,并使用2X SSC、0.2%SDS在70℃洗涤三次,每次15分钟;

(c)一种多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:19的成熟多肽编码序列或其cDNA序列具有至少90%,例如,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%序列同一性;

(d)变体,其包含SEQ ID NO:20的成熟多肽的一个或几个氨基酸的取代、缺失和/或插入;和

(e)(a)、(b)、(c)或(d)的多肽具有纤维素分解增强活性的片段。

2. 权利要求1的多肽,其由包含于pAG78的多核苷酸编码,所述pAG78包含于大肠杆菌NRRL B-50325。

3. 权利要求1的多肽,其中所述成熟多肽是SEQ ID NO:20的氨基酸17至230。

4. 一种组合物,其包含权利要求1-3中任一项的多肽。

5. 一种分离的多核苷酸,其编码权利要求1-3中任一项的多肽。

6. 一种重组宿主细胞,其用与一个或多个指导所述多肽的产生的调控序列可操作地连接的权利要求5的多核苷酸转化。

7. 一种产生权利要求1-3中任一项的多肽的方法,其包括:

(a)在有助于所述多肽产生的条件下培养梭孢壳属(*Thielavia*)细胞;和

(b)回收所述多肽。

8. 一种产生具有纤维素分解增强活性的多肽的方法,其包括:

(a)在有助于所述多肽产生的条件下培养权利要求6的重组宿主细胞;和

(b)回收所述多肽。

9. 一种转基因植物细胞,其用编码权利要求1-3中任一项的多肽的多核苷酸转化。

10. 一种产生具有纤维素分解增强活性的多肽的方法,其包括:

(a)在有助于所述多肽产生的条件下培养权利要求9的转基因植物细胞;和

(b)回收所述多肽。

11. 一种产生亲本细胞的突变体的方法,包括使编码权利要求1-3中任一项的多肽的多核苷酸失活,其导致所述突变体与亲本细胞相比较产生较少的多肽。

12. 一种用于降解或转化纤维素材料的方法,其包括用包含权利要求1-3中任一项的具有纤维素分解增强活性的多肽的酶组合物处理所述纤维素材料。

13. 权利要求12的方法,进一步包括回收经降解的或转化的纤维素材料。

14. 权利要求12或13的方法,其中所述经降解的或转化的纤维素材料是糖,例如,葡萄糖、木糖、甘露糖、半乳糖和阿拉伯糖。

15. 一种用于产生发酵产物的方法,包括:

(a)用包含权利要求1-3中任一项的具有纤维素分解增强活性的多肽的酶组合物糖化所述纤维素材料;

(b)用一种或多种发酵微生物发酵经糖化的纤维素材料以产生发酵产物;和

(c)从发酵回收所述发酵产物。

16.一种发酵纤维素材料的方法,包括:用一种或多种发酵微生物发酵所述纤维素材料,其中所述纤维素材料是用包含权利要求1-3中任一项的具有纤维素分解增强活性的多肽的酶组合物糖化的。

17.权利要求16的方法,其中所述纤维素材料的发酵产生发酵产物。

18.权利要求17的方法,进一步包括从发酵回收发酵产物。

19.权利要求12-18中任一项的方法,其中所述酶组合物还包含一种或几种选自下组的酶:纤维素酶、半纤维素酶、棒曲霉素、酯酶、漆酶、木质素分解酶、果胶酶、过氧化物酶、蛋白酶和膨胀素。

20.一种洗涤剂组合物,其包含权利要求1-3中任一项的多肽,和表面活性剂。

具有纤维素分解增强活性的多肽及编码其的多核苷酸

[0001] 本发明申请是基于申请日为2010年9月16日,申请号为201080052019.0(国际申请号为PCT/US2010/049124)、名称为“具有纤维素分解增强活性的多肽及编码其的多核苷酸”的发明专利申请的分案申请。

[0002] 关于对在联邦资助的研究和开发下完成的发明的权利的声明

[0003] 本发明是在能源部授予的合作协议DE-FC36-08G018080下以政府支持完成的。政府对本发明具有一定权利。

[0004] 涉及序列表

[0005] 本申请包含计算机可读形式的序列表,其通过提述并入本文。

[0006] 涉及生物材料的保藏

[0007] 本申请包含对于生物材料保藏的引用,所述保藏通过提述并入本文。

[0008] 发明背景

技术领域

[0009] 本发明涉及具有纤维素分解增强活性的多肽和编码所述多肽的多核苷酸。本发明还涉及包含所述多核苷酸的核酸构建体、载体和宿主细胞以及产生和使用所述多肽的方法。

背景技术

[0010] 纤维素是单糖葡萄糖通过 β -1,4-键共价连接的聚合物。许多微生物产生水解 β -连接的葡聚糖的酶。这些酶包括内切葡聚糖酶、纤维二糖水解酶和 β -葡糖苷酶。内切葡聚糖酶在随机位置消化纤维素聚合物,使其暴露于纤维二糖水解酶攻击(attack)。纤维二糖水解酶继而从纤维素聚合物的末端释放纤维二糖的分子。纤维二糖是水溶性的 β -1,4-连接的葡萄糖二聚体。 β -葡糖苷酶将纤维二糖水解成葡萄糖。

[0011] 将含木素纤维素原料(lignocellulosic feedstock)转化为乙醇具有以下优势:大量原料现成可用,避免燃烧或填埋材料的合意性和乙醇燃料的清洁性。木材、农业残余物、草本作物和城市固体废物被认为是用于乙醇生产的原料。这些材料主要由纤维素、半纤维素和木质素组成。一旦将纤维素转化成葡萄糖,葡萄糖容易地由酵母发酵成乙醇。

[0012] WO 2005/074647公开了来自土生梭孢壳(*Thielavia terrestris*)的具有纤维素分解增强活性的多肽。WO 2005/074656公开了来自橙橘嗜热子囊菌(*Thermoascus aurantiacus*)的具有纤维素分解增强活性的多肽。WO 2007/089290公开了来自里氏木霉(*Trichoderma reesei*)的具有纤维素分解增强活性的多肽。WO 2009/085935;WO 2009/085859;WO 2009/085864;和WO 2009/085868公开了来自嗜热毁丝霉(*Myceliophthora thermophila*)的的具有纤维素分解增强活性的多肽。

[0013] 本领域中有对于供用于降解纤维素材料、具有改善性质的具有纤维素分解增强活性的多肽的需要。

[0014] 本发明提供了具有纤维素分解增强活性的多肽的编码所述多肽的多核苷酸。

发明内容

[0015] 本发明涉及具有纤维素分解增强活性的分离的多肽,其选自下组:

[0016] (a)多肽,其具有与SEQ ID NO:2,SEQ ID NO:6,或SEQ ID NO:12的成熟多肽至少60%序列同一性;与SEQ ID NO:4的成熟多肽至少65%序列同一性;与SEQ ID NO:18的成熟多肽至少70%序列同一性;与SEQ ID NO:10,SEQ ID NO:16,或SEQ ID NO:22的成熟多肽至少75%序列同一性;与SEQ ID NO:8的成熟多肽至少80%序列同一性;与SEQ ID NO:14的成熟多肽至少85%序列同一性;或者与SEQ ID NO:20的成熟多肽至少90%序列同一性;

[0017] (b)多肽,其由多核苷酸编码,所述多核苷酸在中等-高,高或非常高严格条件下与以下杂交:(i)SEQ ID NO:1,SEQ ID NO:3,SEQ ID NO:5,SEQ ID NO:11,或SEQ ID NO:17的成熟多肽编码序列,(ii)包含于SEQ ID NO:1,SEQ ID NO:3,SEQ ID NO:5,SEQ ID NO:11,或SEQ ID NO:17的成熟多肽编码序列中的cDNA序列,或(iii)(i)或(ii)的全长互补链;在高或非常高严格条件下与以下杂交:(i)SEQ ID NO:7,SEQ ID NO:9,SEQ ID NO:13,SEQ ID NO:15,或SEQ ID NO:21的成熟多肽编码序列,(ii)包含于SEQ ID NO:7,SEQ ID NO:9,SEQ ID NO:13,SEQ ID NO:15,或SEQ ID NO:21的成熟多肽编码序列中的cDNA序列,或(iii)(i)或(ii)的全长互补链;或者在非常高严格条件下与以下杂交:i)SEQ ID NO:19的成熟多肽编码序列,(ii)包含于SEQ ID NO:19的成熟多肽编码序列中的cDNA序列,或(iii)(i)或(ii)的全长互补链;

[0018] (c)多肽,其由多核苷酸编码,所述多核苷酸具有与SEQ ID NO:1,SEQ ID NO:5,或SEQ ID NO:11的成熟多肽编码序列至少60%序列同一性;与SEQ ID NO:3的成熟多肽编码序列至少65%序列同一性;与SEQ ID NO:17的成熟多肽编码序列至少70%序列同一性;与SEQ ID NO:9,SEQ ID NO:15,或SEQ ID NO:21的成熟多肽编码序列至少75%序列同一性;与SEQ ID NO:7的成熟多肽编码序列至少80%序列同一性;与SEQ ID NO:13的成熟多肽编码序列至少85%序列同一性;或者与SEQ ID NO:19的成熟多肽编码序列至少90%序列同一性;或者其cDNA序列;

[0019] (d)变体,其包含SEQ ID NO:2,SEQ ID NO:4,SEQ ID NO:6,SEQ ID NO:8,SEQ ID NO:10,SEQ ID NO:12,SEQ ID NO:14,SEQ ID NO:16,SEQ ID NO:18,SEQ ID NO:20,或SEQ ID NO:22的成熟多肽的一个或多个(几个)取代,缺失,和/或插入;和

[0020] (e)(a)、(b)、(c)或(d)的多肽具有纤维素分解增强活性的片段。

[0021] 本发明还涉及编码本发明的多肽的分离的多核苷酸;包含所述多核苷酸的核酸构建体,重组表达载体和重组宿主细胞;以及产生所述多肽的方法。

[0022] 本发明还涉及降解或转化纤维素材料的方法,包括:在本发明具有纤维素分解增强活性的多肽的存在下用酶组合物处理所述纤维素材料。

[0023] 本发明还涉及产生发酵产物的方法,包括:(a)在本发明具有纤维素分解增强活性的多肽的存在下用酶组合物糖化所述纤维素材料;(b)用一种或多种(几种)发酵微生物发酵糖化的纤维素材料以产生发酵产物;和(c)从发酵回收所述发酵产物。

[0024] 本发明还涉及发酵纤维素材料的方法,包括:用一种或多种(几种)发酵微生物发酵所述纤维素材料,其中所述纤维素材料经在本发明具有纤维素分解增强活性的多肽的存在下用酶组合物的糖化。

[0025] 本发明还涉及编码信号肽的多核苷酸,所述信号肽包含或组成为(consisting of)SEQ ID NO:2的氨基酸1至17,SEQ ID NO:4的氨基酸1至19,SEQ ID NO:6的氨基酸1至17,SEQ ID NO:8的氨基酸1至19,SEQ ID NO:10的氨基酸1至21,SEQ ID NO:12的氨基酸1至24,SEQ ID NO:14的氨基酸1至16,SEQ ID NO:16的氨基酸1至18,SEQ ID NO:18的氨基酸1至22,SEQ ID NO:20的氨基酸1至16,或SEQ ID NO:22的氨基酸1至19,其可操作地连接于编码蛋白的基因;涉及包含所述多核苷酸的核酸构建体,表达载体,和重组宿主细胞;和产生蛋白的方法。

[0026] 具体地,本发明涉及如下各项:

[0027] 1.一种分离的多肽,其具有纤维素分解增强活性,选自下组:

[0028] (a)多肽:其具有与SEQ ID NO:2,SEQ ID NO:6,或SEQ ID NO:12的成熟多肽至少60%,例如,至少65%,至少70%,至少75%,至少80%,至少81%,至少82%,至少83%,至少84%,至少85%,至少86%,至少87%,至少88%,至少89%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%序列同一性;与SEQ ID NO:4的成熟多肽至少65%,例如,至少70%,至少75%,至少80%,至少81%,至少82%,至少83%,至少84%,至少85%,至少86%,至少87%,至少88%,至少89%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%序列同一性;与SEQ ID NO:18的成熟多肽至少70%,例如,至少75%,至少80%,至少81%,至少82%,至少83%,至少84%,至少85%,至少86%,至少87%,至少88%,至少89%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%序列同一性;与SEQ ID NO:10,SEQ ID NO:16,或SEQ ID NO:22的成熟多肽至少75%,例如,至少80%,至少81%,至少82%,至少83%,至少84%,至少85%,至少86%,至少87%,至少88%,至少89%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%序列同一性;与SEQ ID NO:8的成熟多肽至少80%,例如,至少81%,至少82%,至少83%,至少84%,至少85%,至少86%,至少87%,至少88%,至少89%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%序列同一性;与SEQ ID NO:14的成熟多肽至少85%,例如,至少86%,至少87%,至少88%,至少89%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%的序列同一性;

[0029] (b)多肽,其由多核苷酸编码,所述多核苷酸在中等-高,高或非常高严格条件下与以下杂交:(i)SEQ ID NO:1,SEQ ID NO:3,SEQ ID NO:5,SEQ ID NO:11,或SEQ ID NO:17的成熟多肽编码序列,(ii)包含于SEQ ID NO:1,SEQ ID NO:3,SEQ ID NO:5,SEQ ID NO:11,或SEQ ID NO:17的成熟多肽编码序列中的cDNA序列,或(iii)(i)或(ii)的全长互补链;在高或非常高严格条件下与以下杂交:(i)SEQ ID NO:7,SEQ ID NO:9,SEQ ID NO:13,SEQ ID NO:15,或SEQ ID NO:21的成熟多肽编码序列,(ii)包含于SEQ ID NO:7,SEQ ID NO:9,SEQ ID NO:13,SEQ ID NO:15,或SEQ ID NO:21的成熟多肽编码序列中的cDNA序列,或(iii)(i)或(ii)的全长互补链;或者在非常高严格条件下与以下杂交:i)SEQ ID NO:19的成熟多肽

编码序列,(ii)包含于SEQ ID NO:19的成熟多肽编码序列中的cDNA序列,或(iii)(i)或(ii)的全长互补链;

[0030] (c)一种多肽,其由多核苷酸编码,所述多核苷酸具有与SEQ ID NO:1,SEQ ID NO:5,或SEQ ID NO:11的成熟多肽编码序列至少60%,至少60%,例如,至少65%,至少70%,至少75%,至少80%,至少81%,至少82%,至少83%,至少84%,至少85%,至少86%,至少87%,至少88%,至少89%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%序列同一性;与SEQ ID NO:3的成熟多肽编码序列至少65%,例如,至少70%,至少75%,至少80%,至少81%,至少82%,至少83%,至少84%,至少85%,至少86%,至少87%,至少88%,至少89%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%序列同一性;与SEQ ID NO:17的成熟多肽编码序列至少70%,例如,至少75%,至少80%,至少81%,至少82%,至少83%,至少84%,至少85%,至少86%,至少87%,至少88%,至少89%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%序列同一性;与SEQ ID NO:9,SEQ ID NO:15,或SEQ ID NO:21的成熟多肽编码序列至少75%,例如,至少80%,至少81%,至少82%,至少83%,至少84%,至少85%,至少86%,至少87%,至少88%,至少89%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%序列同一性;与SEQ ID NO:7的成熟多肽编码序列至少80%,例如,至少81%,至少82%,至少83%,至少84%,至少85%,至少86%,至少87%,至少88%,至少89%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%序列同一性;与SEQ ID NO:13的成熟多肽编码序列至少85%,例如,至少86%,至少87%,至少88%,至少89%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%序列同一性;或者与SEQ ID NO:19的成熟多肽编码序列至少90%,例如,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%的序列同一性;或其cDNA序列;

[0031] (d)变体,其包含SEQ ID NO:2,SEQ ID NO:4,SEQ ID NO:6,SEQ ID NO:8,SEQ ID NO:10,SEQ ID NO:12,SEQ ID NO:14,SEQ ID NO:16,SEQ ID NO:18,SEQ ID NO:20,或SEQ ID NO:22的成熟多肽的一个或多个(几个)氨基酸的取代,缺失,和/或插入;和

[0032] (e)(a)、(b)、(c)或(d)的多肽具有纤维素分解增强活性的片段。

[0033] 2.项1的多肽,其由多核苷酸编码,所述多核苷酸包含于pSMai216,其包含于大肠杆菌NRRL B-50301;所述多核苷酸包含于pSMai217,其包含于大肠杆菌NRRL B-50302;所述多核苷酸包含于pSMai218,其包含于大肠杆菌NRRL B-50303;所述多核苷酸包含于pSMai213,其包含于大肠杆菌NRRL B-50300;所述多核苷酸包含于pAG68,其包含于大肠杆菌NRRL B-50320;所述多核苷酸包含于pAG69,其包含于大肠杆菌NRRL B-50321;所述多核苷酸包含于pAG75,其包含于大肠杆菌NRRL B-50322;所述多核苷酸包含于pAG76,其包含于大肠杆菌NRRL B-50323;所述多核苷酸包含于pAG77,其包含于大肠杆菌NRRL B-50324;所述多核苷酸包含于pAG78,其包含于大肠杆菌NRRL B-50325;或所述多核苷酸包含于pAG79,其包含于大肠杆菌NRRL B-50326。

[0034] 3.一种组合物,其包含项1或2的多肽。

- [0035] 4.一种分离的多核苷酸,其编码项1或2的多肽。
- [0036] 5.一种重组宿主细胞,其包含项4的多核苷酸,所述多核苷酸可操作地连接于一个或多个调控序列,所述调控序列指导所述多肽在表达宿主中的产生。
- [0037] 6.一种产生项1或2的多肽的方法,其包括:
- [0038] (a)在有助于所述多肽产生的条件下培养细胞,所述细胞以其野生型形式产生所述多肽;和
- [0039] (b)回收所述多肽。
- [0040] 7.一种产生具有纤维素分解增强活性的多肽的方法,其包括:
- [0041] (a)在有助于所述多肽产生的条件下培养包含项4的多核苷酸的重组宿主细胞;和
- [0042] (b)回收所述多肽。
- [0043] 8.一种转基因植物、植物部分或植物细胞,其用编码项1或2的多肽的多核苷酸转化。
- [0044] 9.一种产生具有纤维素分解增强活性的多肽的方法,其包括:
- [0045] (a)在有助于所述多肽产生的条件下培养项8的转基因植物或植物细胞;和
- [0046] (b)回收所述多肽。
- [0047] 10.一种产生亲本细胞的突变体的方法,包括使编码项1或2的多肽的多核苷酸失活,其导致所述突变体与亲本细胞相比较产生较少的多肽。
- [0048] 11.一种双链抑制RNA(dsRNA)分子,其包含项4的多核苷酸的亚序列,其中dsRNA任选地为siRNA或miRNA分子。
- [0049] 12.一种抑制具有纤维素分解增强活性的多肽在细胞中表达的方法,其包括对细胞施用或在细胞中表达项11的双链抑制RNA(dsRNA)分子。
- [0050] 13.一种分离的多核苷酸,其编码信号肽,所述信号肽包含或组成为SEQ ID NO:2的氨基酸1至17,SEQ ID NO:4的氨基酸1至19,SEQ ID NO:6的氨基酸1至17,SEQ ID NO:8的氨基酸1至19,SEQ ID NO:10的氨基酸1至21,SEQ ID NO:12的氨基酸1至24,SEQ ID NO:14的氨基酸1至16,SEQ ID NO:16的氨基酸1至18,SEQ ID NO:18的氨基酸1至22,SEQ ID NO:20的氨基酸1至16,或SEQ ID NO:22的氨基酸1至19。
- [0051] 14.一种产生蛋白的方法,其包括:(a)在有助于所述蛋白产生的条件下培养包含项13的多核苷酸的重组宿主细胞,其中所述基因对于编码信号肽的多核苷酸是外源的;和(b)回收所述蛋白。
- [0052] 15.一种用于降解或转化纤维素材料的方法,其包括:在项1或2的具有纤维素分解增强活性的多肽的存在下用酶组合物处理所述纤维素材料。
- [0053] 16.项15的方法,进一步包括回收经降解的纤维素材料。
- [0054] 17.一种用于产生发酵产物的方法,包括:
- [0055] (a)在项1或2的具有纤维素分解增强活性的多肽的存在下用酶组合物糖化纤维素材料;
- [0056] (b)用一种或多种发酵微生物发酵糖化的纤维素材料以产生发酵产物;和
- [0057] (c)从发酵回收所述发酵产物。
- [0058] 18.一种发酵纤维素材料的方法,包括:用一种或多种发酵微生物发酵所述纤维素材料,其中所述纤维素材料经在项1或2的具有纤维素分解增强活性的多肽的存在下用酶组

合物的糖化。

[0059] 19. 项18的方法,其中所述纤维素材料的发酵产生发酵产物。

[0060] 20. 项19的方法,进一步包括从发酵回收发酵产物。

[0061] 21. 一种洗涤剂组合物,其包含项1或2的多肽,和表面活性剂。

附图说明

[0062] 图1显示编码具有纤维素分解增强活性的土生梭孢壳NRRL 8126 GH61J多肽的基因的不含内含子的基因组DNA序列和推导的氨基酸序列。含有内含子的全长基因组DNA序列示于SEQ ID NO:1,而推导的氨基酸序列示于SEQ ID NO:2。

[0063] 图2显示编码具有纤维素分解增强活性的土生梭孢壳NRRL 8126 GH61K多肽的基因的不含内含子的基因组DNA序列和推导的氨基酸序列。含有内含子的全长基因组DNA序列示于SEQ ID NO:3,而推导的氨基酸序列示于SEQ ID NO:4。

[0064] 图3显示编码具有纤维素分解增强活性的土生梭孢壳NRRL 8126 GH61L多肽的基因的不含内含子的基因组DNA序列和推导的氨基酸序列。含有内含子的全长基因组DNA序列示于SEQ ID NO:5,而推导的氨基酸序列示于SEQ ID NO:6。

[0065] 图4显示在不同浓度的具有纤维素分解增强活性的土生梭孢壳NRRL 8126 GH61J, GH61K和GH61L多肽的存在下用里氏木霉纤维素酶混合物水解经预处理的玉米秸秆(PCS)。

[0066] 图5显示pSMai197的限制图。

[0067] 图6显示编码具有纤维素分解增强活性的土生梭孢壳NRRL 8126 GH61M多肽的基因的基因组DNA序列和推导的氨基酸序列(分别为SEQ ID NO:7和8)。

[0068] 图7显示在不同浓度的具有纤维素分解增强活性的土生梭孢壳NRRL 8126 GH61M多肽的存在下用里氏木霉纤维素酶混合物水解经预处理的玉米秸秆(PCS)。

[0069] 图8显示编码具有纤维素分解增强活性的土生梭孢壳NRRL 8126 GH61N多肽的基因的基因组DNA序列和推导的氨基酸序列(分别为SEQ ID NO:9和10)。

[0070] 图9显示在不同浓度的具有纤维素分解增强活性的土生梭孢壳NRRL 8126 GH61N多肽的存在下用里氏木霉纤维素酶混合物水解经预处理的玉米秸秆(PCS)。

[0071] 图10显示编码具有纤维素分解增强活性的土生梭孢壳NRRL 8126 GH61O多肽的基因的基因组DNA序列和推导的氨基酸序列(分别为SEQ ID NO:11和12)。

[0072] 图11显示编码具有纤维素分解增强活性的土生梭孢壳NRRL 8126 GH61P多肽的基因的基因组DNA序列和推导的氨基酸序列(分别为SEQ ID NO:13和14)。

[0073] 图12显示编码具有纤维素分解增强活性的土生梭孢壳NRRL 8126 GH61R多肽的基因的基因组DNA序列和推导的氨基酸序列(分别为SEQ ID NO:15和16)。

[0074] 图13显示编码具有纤维素分解增强活性的土生梭孢壳NRRL 8126 GH61S多肽的基因的基因组DNA序列和推导的氨基酸序列(分别为SEQ ID NO:17和18)。

[0075] 图14显示编码具有纤维素分解增强活性的土生梭孢壳NRRL 8126 GH61T多肽的基因的基因组DNA序列和推导的氨基酸序列(分别为SEQ ID NO:19和20)。

[0076] 图15显示编码具有纤维素分解增强活性的土生梭孢壳NRRL 8126 GH61U多肽的基因的基因组DNA序列和推导的氨基酸序列(分别为SEQ ID NO:21和22)。

[0077] 定义

[0078] 具有纤维素分解增强活性的多肽:术语“具有纤维素分解增强活性的多肽”意指增强具有纤维素分解活性的酶水解纤维素材料的GH61。就本发明而言,通过测量由纤维素分解酶在下述条件下水解纤维素材料所导致的还原糖增加或纤维二糖与葡萄糖的总量增加来确定纤维素分解增强活性:1-50mg总蛋白/g PCS中的纤维素,其中总蛋白包含50-99.5% w/w的纤维素分解酶蛋白,及0.5-50%w/w的具有纤维素分解增强活性的GH61多肽的蛋白质,在50℃历时1-7天,与用等量的总蛋白加载量而无纤维素分解增强活性(1-50mg纤维素分解蛋白/g PCS中的纤维素)所进行的对照水解相比。在一个优选的方面,使用在总蛋白重量的2-3%的米曲霉(*Aspergillus oryzae*)β-葡糖苷酶(根据WO 02/095014在米曲霉中重组产生)或者总蛋白重量的2-3%的烟曲霉(*Aspergillus fumigatus*)β-葡糖苷酶(如WO 2002/095014所述在米曲霉中重组产生)的纤维素酶蛋白加载量存在下的CELLUCLAST®1.5L(Novozymes A/S, Bagsværd, Denmark)的混合物作为纤维素分解活性的来源。

[0079] 具有纤维素分解增强活性的GH61多肽通过降低达到相同水解程度所需的纤维素分解酶的量而增强由具有纤维素分解活性的酶催化的纤维素材料的水解,优选降低至少1.01倍,更优选至少1.05倍,更优选至少1.10倍,更优选至少1.25倍,更优选至少1.5倍,更优选至少2倍,更优选至少3倍,更优选至少4倍,更优选至少5倍,甚至更优选至少10倍,并且最优选至少20倍。

[0080] 本发明的多肽具有SEQ ID NO:2,SEQ ID NO:4,SEQ ID NO:6,SEQ ID NO:8,SEQ ID NO:10,SEQ ID NO:12,SEQ ID NO:14,SEQ ID NO:16,SEQ ID NO:18,SEQ ID NO:20,或SEQ ID NO:22的成熟多肽的纤维素分解增强活性的至少20%,例如,至少40%,至少50%,至少60%,至少70%,至少80%,至少90%,至少95%,和至少100%。

[0081] 家族61糖苷水解酶:术语“家族61糖苷水解酶”或“家族GH61”或“GH61”意指根据Henrissat B.,1991,A classification of glycosyl hydrolases based on amino-acid sequence similarities,Biochem.J.280:309-316,及Henrissat B.,和Bairoch A.,1996,Updating the sequence-based classification of glycosyl hydrolases,Biochem.J.316:695-696属于糖苷水解酶家族61的多肽。

[0082] 纤维素分解酶或纤维素酶:术语“纤维素分解酶”或“纤维素酶”意指一种或多种(几种)水解纤维素材料的酶。此类酶包括内切葡聚糖酶、纤维二糖水解酶、β-葡糖苷酶或其组合。测量纤维素分解活性的两种基本方法包括:(1)测量总纤维素分解活性,和(2)测量单独的纤维素分解活性(内切葡聚糖酶、纤维二糖水解酶和β-葡糖苷酶),如Zhang等,Outlook for cellulase improvement:Screening and selection strategies,2006,Biotechnology Advances 24:452-481所综述的。总纤维素分解活性通常是使用不溶性底物来测定的,所述底物包括Whatman No.1滤纸、微晶纤维素、细菌纤维素、藻类纤维素、棉花、经预处理的木素纤维素等。最常见的总纤维素分解活性测定法是使用Whatman No.1滤纸作为底物的滤纸测定法。该测定法是由International Union of Pure and Applied Chemistry(IUPAC)(Ghose,1987,Measurement of cellulase activities,Pure Appl.Chem.59:257-68)确立的。

[0083] 就本发明而言,纤维素分解酶活性通过测量在下述条件下由纤维素分解酶进行的纤维素材料水解的增加来确定:1-20mg的纤维素分解酶蛋白/g的PCS中纤维素在50℃进行

3-7日,与未添加纤维素分解酶蛋白的对照水解相比较。典型条件为:1ml反应液,经洗涤或未洗涤的PCS,5%不溶性固形物,50mM乙酸钠pH 5,1mM MnSO₄,50℃,72小时,通过AMINEX®HPX-87H柱(Bio-Rad Laboratories, Inc., Hercules, CA, USA)进行糖分析。

[0084] 内切葡聚糖酶:术语“内切葡聚糖酶”意指内切-1,4-(1,3;1,4)-β-D-葡聚糖4-葡聚糖水解酶(endo-1,4-β-D-glucan 4-glucanohydrolase)(E.C.3.2.1.4),其催化纤维素、纤维素衍生物(例如羧甲基纤维素和羟乙基纤维素)、地衣淀粉(lichenin)中的1,4-β-D-糖苷键、混合的β-1,3葡聚糖例如谷类β-D-葡聚糖或木葡聚糖和含有纤维素组分的其它植物材料中的β-1,4键的内水解(endohydrolysis)。内切葡聚糖酶活性可通过测量底物粘度的减少或由还原糖测定法(Zhang等,2006,Biotechnology Advances 24:452-481)确定的还原端增加来确定。就本发明而言,根据Ghose,1987,Pure and Appl.Chem.59:257-268的方法,在pH 5,40℃,使用羧甲基纤维素(CMC)作为底物来确定内切葡聚糖酶活性。

[0085] 纤维二糖水解酶:术语“纤维二糖水解酶”意指1,4-β-D-葡聚糖纤维二糖水解酶(1,4-beta-D-glucan cellobiohydrolase)(E.C.No.3.2.1.91),其催化纤维素、纤维素寡糖,或任何包含β-1,4-连接的葡萄糖的聚合物中的1,4-β-D-糖苷键的水解,从链的还原或非还原末端释放纤维二糖(Teeri,1997,Crystalline cellulose degradation:New insight into the function of cellobiohydrolases,Trends in Biotechnology 15:160-167;Teeri等,1998,Trichoderma reesei cellobiohydrolases:why so efficient on crystalline cellulose?,Biochem.Soc.Trans.26:173-178)。就本发明而言,根据Lever等,1972,Anal.Biochem.47:273-279;van Tilbeurgh等,1982,FEBS Letters,149:152-156;van Tilbeurgh和Claeyssens,1985,FEBS Letters,187:283-288;以及Tomme等,1988,Eur.J.Biochem.170:575-581描述的方法确定纤维二糖水解酶活性。在本发明中,可采用Lever等的方法来评价玉米秸秆中的纤维素水解,而van Tilbeurgh等和Tomme等的方法可用于确定对荧光二糖衍生物4-甲基伞形基-β-D-乳糖苷(4-methylumbelliferyl-β-D-lactoside)的纤维二糖水解酶活性。

[0086] β-葡糖苷酶:术语“β-葡糖苷酶”意指β-D-葡糖苷葡萄糖水解酶(beta-D-glucoside glucohydrolase)(E.C.No.3.2.1.21),其催化末端非还原β-D-葡萄糖残基的水解,并释放β-D-葡萄糖。就本发明而言,β-葡糖苷酶活性是根据由Venturi等,2002,Extracellular beta-D-glucosidase from Chaetomium thermophilum var.coprophilum:production, purification and some biochemical properties,J.Basic Microbiol.42:55-66所述的基本步骤确定的。一单位的β-葡糖苷酶定义为在25℃,pH 4.8从作为底物的1mM对硝基苯基-β-D-葡糖吡喃糖苷在含有0.01% TWEEN®20的50mM柠檬酸钠中每分钟产生1.0微摩尔的对硝基苯酚阴离子。

[0087] 半纤维素分解酶或半纤维素酶:术语“半纤维素分解酶”或“半纤维素酶”意指一种或多种(几种)水解半纤维素材料的酶。参见,例如Shallom,D.和Shoham,Y.Microbial hemicellulases.Current Opinion In Microbiology,2003,6(3):219-228)。半纤维素酶是植物生物质降解中的关键成分。半纤维素酶的实例包括但不限于乙酰甘露聚糖酯酶、乙酰木聚糖酯酶、阿拉伯聚糖酶、阿拉伯呋喃糖苷酶、香豆酸酯酶、阿魏酸酯酶、半乳糖苷酶、葡糖醛酸糖苷酶、葡糖醛酸酯酶、甘露聚糖酶、甘露糖苷酶、木聚糖酶和木糖苷酶。这些酶的底物,半纤维素,是支化和直链多糖的混杂集团,这些多糖通过氢键键合于植物细胞壁中的

纤维素微纤维,将其交联为鲁棒(robust)的网络。半纤维素亦共价地附于木质素,与纤维素一同形成高度复杂的结构。半纤维素的变性的结构和组织形式需要许多酶的协同作用使其完全降解。半纤维素酶的催化模块为水解糖苷键的糖苷水解酶(GH),或水解乙酸或阿魏酸侧基酯连接的糖酯酶(CE)。这些催化模块,基于其一级结构的同源性,可指派为以数字标记的GH和CE家族。一些家族,具有总体上类似的折叠,可进一步归类为宗族(clan),以字母标记(例如,GH-A)。最具信息性和最新的这些和其他糖活性酶的分类可在Carbohydrate-Active Enzymes(CAZy)数据库获得。半纤维素分解酶活性可根据Ghose和Bisaria,1987, Pure&Appl.Chem.59:1739-1752测量。

[0088] 木聚糖降解活性或木聚糖分解活性:术语“木聚糖降解活性”或“木聚糖分解活性”意指水解含木聚糖材料的生物学活性。两种测定木聚糖分解活性的基础方法包括:(1)测定总木聚糖分解活性,和(2)测定单独的木聚糖分解活性(例如,内切木聚糖酶、 β -木糖苷酶、阿拉伯呋喃糖苷酶、 α -葡糖醛酸糖苷酶、乙酰木聚糖酯酶、阿魏酸酯酶和 α -葡糖醛酸酯酶(α -glucuronoyl esterase))。最近在木聚糖分解酶测定法的进展总结于几个公开文献中,包括Biely和Puchard,Recent progress in the assays of xylanolytic enzymes,2006, Journal of the Science of Food and Agriculture 86(11):1636-1647;Spanikova和Biely,2006,Glucuronoyl esterase-Novel carbohydrate esterase produced by *Schizophyllum commune*,FEBS Letters 580(19):4597-4601;Herrmann,Vrsanska, Jurickova,Hirsch,Biely和Kubicek,1997,The beta-D-xylosidase of *Trichoderma reesei* is a multifunctional beta-D-xylan xylohydrolase,Biochemical Journal 321:375-381。

[0089] 总木聚糖降解活性可通过确定从多种类型的木聚糖形成的还原糖来测量,所述木聚糖包括例如燕麦小麦(oat spelt)、山毛榉木(beechwood)和落叶松木(larchwood)木聚糖,或者可通过光度法确定从多种共价染色的木聚糖释放出的染色的木聚糖片段来测量。最常见的总木聚糖分解活性测定法基于从多聚的4-O-甲基葡糖醛酸木聚糖产生还原糖,如Bailey,Biely,Poutanen,1992,Interlaboratory testing of methods for assay of xylanase activity,Journal of Biotechnology 23(3):257-270中所述。木聚糖酶活性亦可用0.2%AZCL-阿拉伯木聚糖作为底物在0.01% TRITON®X-100和200mM磷酸钠缓冲液pH 6在37°C中确定。一单位的木聚糖酶活性定义为在37°C,pH 6从作为底物的0.2%AZCL阿拉伯木聚糖在200mM磷酸钠pH 6缓冲液中每分钟产生1.0微摩尔的天青精。

[0090] 就本发明而言,木聚糖降解活性是通过测量由木聚糖降解酶在下述典型条件下造成的桦木木聚糖(Sigma Chemical Co.,Inc.,St.Louis,MO,USA)水解的增加来确定的:1ml反应,5mg/ml底物(总固形物),5mg木聚糖分解蛋白/g底物,50mM乙酸钠,pH 5,50°C,24小时,如Lever,1972,A new reaction for colorimetric determination of carbohydrates,Anal.Biochem 47:273-279所述使用对羟基苯甲酸酰肼(PHBAH)测定法进行糖分析。

[0091] 木聚糖酶:术语“木聚糖酶”意指催化木聚糖中1,4- β -D-木糖苷连接的内水解的1,4- β -D-木聚糖-木糖水解酶(E.C.3.2.1.8)。就本发明而言,木聚糖酶活性是使用0.2% AZCL-阿拉伯木聚糖作为底物在37°C在0.01% TRITON®X-100和200mM磷酸钠缓冲液pH 6中进行确定的。一个单位的木聚糖酶活性定义为在37°C,pH 6从200mM磷酸钠pH 6缓冲液中

的0.2%AZCL-阿拉伯木聚糖作为底物每分钟产生1.0微摩尔的天青精(azurine)。

[0092] β -木糖苷酶:术语“ β -木糖苷酶”意指 β -D木糖苷木糖水解酶(β -D-xyloside xylohydrolase)(E.C.3.2.1.37),其催化短 β (1 \rightarrow 4)木寡糖(xylooligosaccharide)的外水解以从非还原端去除连续的D-木糖残基。就本发明而言,一个单位的 β -木糖苷酶定义为在40 $^{\circ}$ C,pH 5从1mM对硝基苯基- β -D-木糖苷作为底物在含有0.01%TWEEN[®]20的100mM柠檬酸钠中每分钟产生1.0微摩尔对硝基苯酚阴离子。

[0093] 乙酰木聚糖酯酶:术语“乙酰木聚糖酯酶”意指羧基酯酶(EC 3.1.1.72),其催化乙酰基从聚合木聚糖、乙酰化木糖、乙酰化葡萄糖、乙酸 α -萘酯(alpha-naphthyl acetate)和乙酸对硝基苯酯(p-nitrophenyl acetate)的水解。就本发明而言,乙酰木聚糖酯酶活性是使用0.5mM乙酸对硝基苯酯作为底物,在含有0.01%TWEEN[™]20的50mM乙酸钠pH 5.0中确定的。一个单位的乙酰木聚糖酯酶定义为能够在pH 5,25 $^{\circ}$ C每分钟释放1微摩尔对硝基苯酚阴离子(p-nitrophenolate anion)的酶量。

[0094] 阿魏酸酯酶:术语“阿魏酸酯酶(feruloyl esterase)”意指4-羟基-3-甲氧基肉桂酰-糖水水解酶(EC 3.1.1.73),其催化4-羟基-3-甲氧基肉桂酰(阿魏酰)基团从酯化的糖(其在“天然”底物中通常为阿拉伯糖)的水解,以产生阿魏酸(4-羟基-3-甲氧基肉桂酸)。阿魏酸酯酶也称作阿魏酸酯酶(ferulic acid esterase)、羟基肉桂酸酯酶(hydroxycinnamoyl esterase)、FAE-III、肉桂酸酯水解酶、FAEA、cinnAE、FAE-I或FAE-II。就本发明而言,阿魏酸酯酶是使用50mM乙酸钠pH 5.0中的0.5mM阿魏酸对硝基苯酯作为底物确定的。一个单位的阿魏酸酯酶活性等于能够在pH 5,25 $^{\circ}$ C每分钟释放出1微摩尔对硝基苯酚阴离子的酶量。

[0095] α -葡糖醛酸糖苷酶:术语“ α -葡糖醛酸糖苷酶”意指 α -D-葡糖苷酸葡糖醛酸水解酶(alpha-D-glucosiduronate glucuronohydrolase)(EC 3.2.1.139),其催化 α -D-葡糖醛酸糖苷水解为D-葡糖醛酸和醇。就本发明而言, α -葡糖醛酸糖苷酶活性是根据de Vries, 1998, J. Bacteriol. 180:243-249确定的。一个单位的 α -葡糖醛酸糖苷酶等于能够在pH 5, 40 $^{\circ}$ C每分钟释放出1微摩尔葡糖醛酸或4-O-甲基葡糖醛酸的酶量。

[0096] α -L-阿拉伯呋喃糖苷酶:术语“ α -L-阿拉伯呋喃糖苷酶活性”意指 α -L-阿拉伯呋喃糖苷阿拉伯呋喃水解酶(EC 3.2.1.55),其催化对 α -L-阿拉伯糖苷中的末端非还原性 α -L-阿拉伯呋喃糖苷残基的水解。该酶对 α -L-阿拉伯呋喃糖苷、含有(1,3)-和/或(1,5)-键的 α -L-阿拉伯聚糖、阿拉伯木聚糖和阿拉伯半乳聚糖起作用。 α -L-阿拉伯呋喃糖苷酶也称为阿拉伯糖苷酶、 α -阿拉伯糖苷酶、 α -L-阿拉伯糖苷酶、 α -阿拉伯呋喃糖苷酶、多糖 α -L-阿拉伯呋喃糖苷酶、 α -L-阿拉伯呋喃糖苷水解酶、L-阿拉伯糖苷酶或 α -L-阿拉伯聚糖酶。就本发明而言, α -L-阿拉伯呋喃糖苷酶活性是使用总体积200 μ l中的每ml的100mM乙酸钠pH 5中5mg的中等粘性小麦阿拉伯木聚糖(Megazyme International Ireland, Ltd., Bray, Co. Wicklow, Ireland)在40 $^{\circ}$ C进行30分钟,接着通过AMINEX[®]HPX-87H柱层析(Bio-Rad Laboratories, Inc., Hercules, CA, USA)的阿拉伯糖分析来确定的。

[0097] 纤维素材料:术语“纤维素材料”意指包含纤维素的任何材料。生物质的初生细胞壁(primary cell wall)中的主要多糖是纤维素,其次最丰富的是半纤维素,而第三是果胶。次生细胞壁(secondary cell wall)在细胞停止生长后产生,其同样含有多糖并通过共价交联至半纤维素的聚合木质素而加强。纤维素是脱水纤维二糖的均聚物,并且因此是直链 β -(1-4)-D-葡聚糖,而半纤维素包括多种化合物,例如木聚糖、木葡聚糖(xyloglucan)、

阿拉伯木聚糖和甘露聚糖,具有系列取代基的复杂分支结构。尽管通常是多形的,存在于植物组织中的纤维素主要是平行葡聚糖链的不溶晶体基质。半纤维素通常与纤维素以及其它半纤维素以氢键键合,其帮助稳定细胞壁基质。

[0098] 纤维素通常见于例如植物的茎、叶、壳、皮和穗轴,或树的叶、枝和木材。纤维素材料可以是,但不限于,草本材料、农业残余物、林业残余物、城市固体废物、废纸和纸浆与造纸厂残余物(参见,例如,Wiselogel等,1995,于Handbook on Bioethanol(Charles E. Wyman编),pp.105-118,Taylor&Francis,Washington D.C.;Wyman,1994,Bioresource Technology 50:3-16;Lynd,1990,Applied Biochemistry and Biotechnology 24/25:695-719;Mosier等,,1999,Recent Progress in Bioconversion of Lignocellulosics,于Advances in Biochemical Engineering/Biotechnology,T.Scheper主编,Volume 65,pp.23-40,Springer-Verlag,New York)。在本文中应理解的是,纤维素可以是任何形式的木素纤维素,在混合基质中包含木质素、纤维素和半纤维素的植物细胞壁材料。在一个优选的方面,纤维素材料是木素纤维素,其包含纤维素、半纤维素和木质素。

[0099] 在一个方面,纤维素材料是草本材料。在另一个方面,纤维素材料是农业残余物。在另一个方面,纤维素材料是林业残余物。在另一个方面,纤维素材料是城市固体废物。在另一个方面,纤维素材料是废纸。在另一个方面,纤维素材料是纸浆和造纸厂残余物。

[0100] 在另一个方面,纤维素材料是玉米秸秆。在另一个方面,纤维素材料是玉米纤维。在另一个方面,纤维素材料是玉米穗轴。在另一个方面,纤维素材料是橙皮。在另一个方面,纤维素材料是稻秆。在另一个方面,纤维素材料是麦秆。在另一个方面,纤维素材料是柳枝稷(switch grass)。在另一个方面,纤维素材料是芒草属(miscanthus)。在另一个方面,纤维素材料是甘蔗渣。

[0101] 在另一个方面,纤维素材料是微晶纤维素。在另一个方面,纤维素材料是细菌纤维素。在另一个方面,纤维素材料是藻类纤维素。在另一个方面,纤维素材料是棉线头(cotton linter)。在另一个方面,纤维素材料是无定形的磷酸处理的纤维素。在另一个方面,纤维素材料是滤纸。

[0102] 纤维素材料可以直接使用或进行预处理,使用本领域已知的常规方法,如本文所述。在一个优选的方面,预处理纤维素材料。

[0103] 预处理的玉米秸秆:术语“PCS”或“预处理的玉米秸秆”意指通过用热和稀硫酸处理的源自玉米秸秆的纤维素材料。

[0104] 含木聚糖材料:术语“含木聚糖材料”意指任何包含含有 β -(1-4)连接木糖残基骨架的植物细胞壁多肽的材料。陆生植物的木聚糖是具有 β -(1-4)-D-吡喃木糖骨架、具有短糖链分支的杂聚物。其包含D-葡糖醛酸或其4-O-甲基醚、L-阿拉伯糖和/或多种寡糖,由D-木糖、L-阿拉伯糖,D-或L-半乳糖和D-葡萄糖组成。木聚糖类型的多糖可分为同木聚糖(homoxylan)和杂木聚糖(heteroxylan),其包括葡糖醛酸木聚糖,(阿拉伯)葡糖醛酸木聚糖,(葡糖醛酸)阿拉伯木聚糖,阿拉伯木聚糖和复合杂木聚糖(complex heteroxylan)。参见例如Ebringerova等,2005,Adv.Polym.Sci.186:1-67。

[0105] 在本发明的方法中,可使用任何含木聚糖材料。在一个优选的方面,所述含木聚糖材料是木素纤维素。

[0106] 分离的或纯化的:术语“分离的”和“纯化的”意指从与其天然联系的至少一种组分

移出的多肽或多核苷酸。举例而言,多肽如通过SDS-PAGE测定的,可为至少1%纯,例如至少5%纯,至少10%纯,至少20%纯,至少40%纯,至少60%纯,至少80%纯,至少90%纯,或至少95%纯,而多核苷酸如通过琼脂糖电泳测定的,可为至少1%纯,例如至少5%纯,至少10%纯,至少20%纯,至少40%纯,至少60%纯,至少80%纯,至少90%纯,或至少95%纯。

[0107] 成熟多肽:术语“成熟多肽”意指以其在翻译和任何翻译后修饰之后的最终形式存在的多肽,所述修饰例如N-末端加工、C-末端截短、糖基化、磷酸化等。在一个方面,根据预测SEQ ID NO:2的氨基酸1至17是信号肽的SignalP程序(Nielsen等,1997,Protein Engineering 10:1-6),成熟多肽是SEQ ID NO:2的氨基酸18至246。在另一个方面,根据预测SEQ ID NO:4的氨基酸1至19是信号肽的SignalP程序,成熟多肽是SEQ ID NO:4的氨基酸20至334。在另一个方面,根据预测SEQ ID NO:6的氨基酸1至17是信号肽的SignalP程序,成熟多肽是SEQ ID NO:6的氨基酸18至227。在另一个方面,根据预测SEQ ID NO:8的氨基酸1至19是信号肽的SignalP程序,成熟多肽是SEQ ID NO:8的氨基酸20至223。在另一个方面,根据预测SEQ ID NO:10的氨基酸1至21是信号肽的SignalP程序,成熟多肽是SEQ ID NO:10的氨基酸22至368。在另一个方面,根据预测SEQ ID NO:12的氨基酸1至24是信号肽的SignalP程序,成熟多肽是SEQ ID NO:12的氨基酸25至330。在另一个方面,根据预测SEQ ID NO:14的氨基酸1至16是信号肽的SignalP程序,成熟多肽是SEQ ID NO:14的氨基酸17至236。在另一个方面,根据预测SEQ ID NO:16的氨基酸1至18是信号肽的SignalP程序,成熟多肽是SEQ ID NO:16的氨基酸19至250。在另一个方面,根据预测SEQ ID NO:18的氨基酸1至22是信号肽的SignalP程序,成熟多肽是SEQ ID NO:18的氨基酸23至478。在另一个方面,根据预测SEQ ID NO:20的氨基酸1至16是信号肽的SignalP程序,成熟多肽是SEQ ID NO:20的氨基酸17至230。在另一个方面,根据预测SEQ ID NO:22的氨基酸1至19是信号肽的SignalP程序,成熟多肽是SEQ ID NO:22的氨基酸20至257。本领域中已知宿主细胞可产生两种或更多种由相同多核苷酸表达的不同成熟多肽(即,具有不同的C端和/或N端氨基酸)的混合物。

[0108] 成熟多肽编码序列:术语“成熟多肽编码序列”意指编码具有纤维素分解增强活性的成熟多肽的多核苷酸。在一个方面,根据预测SEQ ID NO:1的核苷酸1至51编码信号肽的SignalP程序(Nielsen等,1997,见上),成熟多肽编码序列是SEQ ID NO:1的核苷酸52至875。在另一个方面,成熟多肽编码序列是包含于SEQ ID NO:1的核苷酸52至875中的cDNA序列。在另一个方面,根据预测SEQ ID NO:3的核苷酸1至57编码信号肽的SignalP程序,成熟多肽编码序列是SEQ ID NO:3的核苷酸58至1250。在另一个方面,成熟多肽编码序列是包含于SEQ ID NO:3的核苷酸58至1250中的cDNA序列。在另一个方面,根据预测SEQ ID NO:5的核苷酸1至51编码信号肽的SignalP程序,成熟多肽编码序列是SEQ ID NO:5的核苷酸52至795。在另一个方面,成熟多肽编码序列是包含于SEQ ID NO:5的核苷酸52至795中的cDNA序列。在另一个方面,根据预测SEQ ID NO:7的核苷酸1至57编码信号肽的SignalP程序,成熟多肽编码序列是SEQ ID NO:7的核苷酸58至974。在另一个方面,成熟多肽编码序列是包含于SEQ ID NO:7的核苷酸58至974中的cDNA序列。在另一个方面,根据预测SEQ ID NO:9的核苷酸1至63编码信号肽的SignalP程序,成熟多肽编码序列是SEQ ID NO:9的核苷酸64至1104。在另一个方面,成熟多肽编码序列是包含于SEQ ID NO:9的核苷酸64至1104中的cDNA序列。在另一个方面,根据预测SEQ ID NO:11的核苷酸1至72编码信号肽的SignalP程序,成

成熟多肽编码序列是SEQ ID NO:11的核苷酸73至990。在另一个方面,成熟多肽编码序列是包含于SEQ ID NO:11的核苷酸73至990中的cDNA序列。在另一个方面,根据预测SEQ ID NO:13的核苷酸1至48编码信号肽的SignalP程序,成熟多肽编码序列是SEQ ID NO:13的核苷酸49至1218。在另一个方面,成熟多肽编码序列是包含于SEQ ID NO:13的核苷酸49至1218中的cDNA序列。在另一个方面,根据预测SEQ ID NO:15的核苷酸1至54编码信号肽的SignalP程序,成熟多肽编码序列是SEQ ID NO:15的核苷酸55至930。在另一个方面,成熟多肽编码序列是包含于SEQ ID NO:15的核苷酸55至930中的cDNA序列。在另一个方面,根据预测SEQ ID NO:17的核苷酸1至66编码信号肽的SignalP程序,成熟多肽编码序列是SEQ ID NO:17的核苷酸67至1581。在另一个方面,成熟多肽编码序列是包含于SEQ ID NO:17的核苷酸67至1581中的cDNA序列。在另一个方面,根据预测SEQ ID NO:19的核苷酸1至48编码信号肽的SignalP程序,成熟多肽编码序列是SEQ ID NO:19的核苷酸49至865。在另一个方面,成熟多肽编码序列是包含于SEQ ID NO:19的核苷酸49至865中的cDNA序列。在另一个方面,根据预测SEQ ID NO:21的核苷酸1至57编码信号肽的SignalP程序,成熟多肽编码序列是SEQ ID NO:21的核苷酸58至1065。在另一个方面,成熟多肽编码序列是包含于SEQ ID NO:21的核苷酸58至1065中的cDNA序列。

[0109] 序列同一性:参数“序列同一性”描述两个氨基酸序列之间或两个核苷酸序列之间的相关性。

[0110] 就本发明而言,两个氨基酸序列之间的序列同一性程度使用如EMBOSS软件包(EMBOSS:The European Molecular Biology Open Software Suite,Rice等,2000,Trends Genet.16:276-277),优选3.0.0版或更高版本的Needle程序中所执行的Needleman-Wunsch算法(Needleman和Wunsch,1970,J.Mol.Biol.48:443-453)来测定。使用的可选参数为缺口开放罚分(gap open penalty)10,缺口延伸罚分(gap extension penalty)0.5和EBLOSUM62(BLOSUM62的EMBOSS版)取代矩阵。使用Needle标记为“最高同一性(longest identity)”的输出结果(使用-nobrief选项获得)作为同一性百分比,并计算如下:

[0111] (同样的残基×100)/(比对长度-比对中缺口的总数)

[0112] 就本发明而言,两个脱氧核苷酸序列之间的序列同一性程度使用如EMBOSS软件包(EMBOSS:The European Molecular Biology Open Software Suite,Rice等,2000,见上文),优选3.0.0版或更高版本的Needle程序中所执行的Needleman-Wunsch算法(Needleman和Wunsch,1970,见上文)来测定。使用的可选参数为缺口开放罚分10,缺口延伸罚分0.5和EDNAFULL(NCBI NUC4.4的EMBOSS版)取代矩阵。使用Needle标记为“最高同一性”的输出结果(使用-nobrief选项获得)作为同一性百分比,并计算如下:

[0113] (同样的脱氧核糖核苷酸×100)/(比对长度-比对中缺口的总数)

[0114] 片段:术语“片段”意指从成熟多肽的氨基和/或羧基末端缺失一个或多个(几个)氨基酸的多肽;其中所述片段具有纤维素分解增强活性。在一个方面,片段含有SEQ ID NO:2的成熟多肽的至少190个氨基酸残基,例如至少200个氨基酸残基,或至少210个氨基酸残基。在另一个方面,片段含有SEQ ID NO:4的成熟多肽的至少265个氨基酸残基,例如至少280个氨基酸残基,或至少295个氨基酸残基。在另一个方面,片段含有SEQ ID NO:6的成熟多肽的至少180个氨基酸残基,例如至少190个氨基酸残基,或至少200个氨基酸残基。在另一个方面,片段含有SEQ ID NO:4的成熟多肽的至少265个氨基酸残基,例如至少280个氨基

酸残基,或至少295个氨基酸残基。在另一个方面,片段含有SEQ ID NO:8的成熟多肽的至少170个氨基酸残基,例如至少180个氨基酸残基,或至少190个氨基酸残基。在另一个方面,片段含有SEQ ID NO:10的成熟多肽的至少305个氨基酸残基,例如至少320个氨基酸残基,或至少335个氨基酸残基。在另一个方面,片段含有SEQ ID NO:12的成熟多肽的至少255个氨基酸残基,例如至少270个氨基酸残基,或至少285个氨基酸残基。在另一个方面,片段含有SEQ ID NO:14的成熟多肽的至少190个氨基酸残基,例如至少200个氨基酸残基,或至少210个氨基酸残基。在另一个方面,片段含有SEQ ID NO:16的成熟多肽的至少200个氨基酸残基,例如至少210个氨基酸残基,或至少220个氨基酸残基。在另一个方面,片段含有SEQ ID NO:18的成熟多肽的至少390个氨基酸残基,例如至少410个氨基酸残基,或至少430个氨基酸残基。在另一个方面,片段含有SEQ ID NO:20的成熟多肽的至少180个氨基酸残基,例如至少190个氨基酸残基,或至少200个氨基酸残基。在另一个方面,片段含有SEQ ID NO:22的成熟多肽的至少210个氨基酸残基,例如至少220个氨基酸残基,或至少230个氨基酸残基。

[0115] 亚序列:术语“亚序列(subsequence)”意指从成熟多肽编码序列的5'和/或3'端缺失一个或多个(几个)核苷酸的多核苷酸;其中所述亚序列编码具有纤维素分解增强活性的片段。在一个方面,亚序列含有SEQ ID NO:1的成熟多肽编码序列的至少560个核苷酸,例如至少600个核苷酸,或至少630个核苷酸。在另一个优选的方面,亚序列含有SEQ ID NO:3的成熟多肽编码序列的至少795个核苷酸,例如至少840个核苷酸,或至少885个核苷酸。在另一个优选的方面,亚序列含有SEQ ID NO:5的成熟多肽编码序列的至少540个核苷酸,例如至少570个核苷酸,或至少600个核苷酸。在另一个优选的方面,亚序列含有SEQ ID NO:7的成熟多肽编码序列的至少510个核苷酸,例如至少540个核苷酸,或至少570个核苷酸。在另一个优选的方面,亚序列含有SEQ ID NO:9的成熟多肽编码序列的至少915个核苷酸,例如至少960个核苷酸,或至少1005个核苷酸。在另一个优选的方面,亚序列含有SEQ ID NO:11的成熟多肽编码序列的至少765个核苷酸,例如至少810个核苷酸,或至少855个核苷酸。在另一个优选的方面,亚序列含有SEQ ID NO:3的成熟多肽编码序列的至少795个核苷酸,例如至少840个核苷酸,或至少885个核苷酸。在另一个优选的方面,亚序列含有SEQ ID NO:13的成熟多肽编码序列的至少570个核苷酸,例如至少600个核苷酸,或至少630个核苷酸。在另一个优选的方面,亚序列含有SEQ ID NO:15的成熟多肽编码序列的至少600个核苷酸,例如至少630个核苷酸,或至少660个核苷酸。在另一个优选的方面,亚序列含有SEQ ID NO:17的成熟多肽编码序列的至少1170个核苷酸,例如至少1230个核苷酸,或至少1290个核苷酸。在另一个优选的方面,亚序列含有SEQ ID NO:19的成熟多肽编码序列的至少540个核苷酸,例如至少570个核苷酸,或至少600个核苷酸。在另一个优选的方面,亚序列含有SEQ ID NO:21的成熟多肽编码序列的至少630个核苷酸,例如至少660个核苷酸,或至少690个核苷酸。

[0116] 等位变体(allelic variant):术语“等位变体”在本文中表示占据相同染色体基因座的基因的任何两种或更多种可选形式。等位变异通过突变天然地发生,并且可导致种群内的多态性。基因突变可以是沉默的(在编码的多肽中无变化)或可以编码具有改变的氨基酸序列的多肽。多肽的等位变体是由基因的等位变体编码的多肽。

[0117] 编码序列:术语“编码序列”意指直接指定多肽氨基酸序列的多核苷酸。编码序列的边界通常由开读框决定,所述开读框通常以ATG起始密码子或其他起始密码子例如GTG和TTG开始,并且以终止密码子例如TAA、TAG和TGA结束。编码序列可以是DNA、cDNA、合成的或

重组的多核苷酸。

[0118] cDNA:术语“cDNA”意指能够通过反转录从得自真核细胞的成熟的、已剪接的mRNA分子制备的DNA分子。cDNA缺少通常存在于相应基因组DNA中的内含子序列。起始的(initial)、初级的RNA转录物是mRNA的前体,其通过一系列的包括剪接的步骤加工然后作为成熟的已剪接的mRNA出现。

[0119] 核酸构建体:术语“核酸构建体”意指分离自天然存在的基因的,或受修饰以本来不存在于(not otherwise exist)自然界中的方式含有核酸的区段的,或是合成的单链或双链的核酸分子。当所述核酸构建体含有表达本发明的编码序列所需的调控序列时,术语核酸构建体与术语“表达盒”同义。

[0120] 调控序列(control sequence):术语“调控序列”意指对编码本发明多肽的多核苷酸表达是必需的所有成分。各个调控序列对于编码所述多肽的多核苷酸可以是天然的或外源的,或各个调控序列对于彼此可以是天然的或外源的。这些调控序列包括但不限于前导序列、聚腺苷酸化序列、前肽序列、启动子、信号肽序列和转录终止子。最少的情况,调控序列包括启动子和转录和翻译的终止信号。调控序列可以和用于引入特异性限制位点的接头一起提供,所述特异性限制位点促进调控序列与编码多肽的多核苷酸编码区的连接。

[0121] 可操作地连接:术语“可操作地连接”意指这样的构型,其中将调控序列置于相对于多核苷酸的编码序列的适当位置,使得调控序列指导编码序列的表达。

[0122] 表达:术语“表达”包括涉及多肽产生的任何步骤,其包括但不限于转录、转录后修饰、翻译、翻译后修饰和分泌。

[0123] 表达载体:术语“表达载体”意指线性的或环状的DNA分子,其包含编码多肽的多核苷酸,并且所述多核苷酸与提供用于其表达的额外核苷酸可操作地连接。

[0124] 宿主细胞:“宿主细胞”意指任何细胞类型,所述细胞类型对于使用包含本发明多核苷酸的核酸构建体或表达载体的转化、转染、转导等是易感的(susceptible)。术语“宿主细胞”涵盖亲本细胞的任何后代,所述后代由于在复制中发生的突变与亲本细胞不完全相同。

[0125] 变体:术语“变体”意指具有纤维素分解增强活性的多肽,其包含改变,即在一个或多个(几个)位置的取代、插入和/或缺失一个或多个(几个)氨基酸残基。取代意指用别的氨基酸取代占据某位置的氨基酸;缺失意指去除占据某位置的氨基酸;而插入意指邻接于占据某位置的氨基酸添加一个或多个(几个)氨基酸,例如1-5个氨基酸。

[0126] 发明详述

[0127] 具有纤维素分解增强活性的多肽

[0128] 本发明涉及具有纤维素分解增强活性的分离的多肽,其选自下组:

[0129] (a)多肽,其具有与SEQ ID NO:2,SEQ ID NO:6,或SEQ ID NO:12的成熟多肽至少60%序列同一性;与SEQ ID NO:4的成熟多肽至少65%序列同一性;与SEQ ID NO:18的成熟多肽至少70%序列同一性;与SEQ ID NO:10,SEQ ID NO:16,或SEQ ID NO:22的成熟多肽至少75%序列同一性;与SEQ ID NO:8的成熟多肽至少80%序列同一性;与SEQ ID NO:14的成熟多肽至少85%序列同一性;或者与SEQ ID NO:20的成熟多肽至少90%序列同一性;

[0130] (b)多肽,其由多核苷酸编码,所述多核苷酸在中等-高,高或非常高严格条件下与以下杂交:(i)SEQ ID NO:1,SEQ ID NO:3,SEQ ID NO:5,SEQ ID NO:11,或SEQ ID NO:17的

成熟多肽编码序列,(ii)包含于SEQ ID NO:1,SEQ ID NO:3,SEQ ID NO:5,SEQ ID NO:11,或SEQ ID NO:17的成熟多肽编码序列中的cDNA序列,或(iii)(i)或(ii)的全长互补链;在高或非常高严格条件下与以下杂交:(i)SEQ ID NO:7,SEQ ID NO:9,SEQ ID NO:13,SEQ ID NO:15,或SEQ ID NO:21的成熟多肽编码序列,(ii)包含于SEQ ID NO:7,SEQ ID NO:9,SEQ ID NO:13,SEQ ID NO:15,或SEQ ID NO:21的成熟多肽编码序列中的cDNA序列,或(iii)(i)或(ii)的全长互补链;或者在非常高严格条件下与以下杂交:i)SEQ ID NO:19的成熟多肽编码序列,(ii)包含于SEQ ID NO:19的成熟多肽编码序列中的cDNA序列,或(iii)(i)或(ii)的全长互补链;

[0131] (c)多肽,其由多核苷酸编码,所述多核苷酸具有与SEQ ID NO:1,SEQ ID NO:5,或SEQ ID NO:11的成熟多肽编码序列至少60%序列同一性;与SEQ ID NO:3的成熟多肽编码序列至少65%序列同一性;与SEQ ID NO:17的成熟多肽编码序列至少70%序列同一性;与SEQ ID NO:9,SEQ ID NO:15,或SEQ ID NO:21的成熟多肽编码序列至少75%序列同一性;与SEQ ID NO:7的成熟多肽编码序列至少80%序列同一性;与SEQ ID NO:13的成熟多肽编码序列至少85%序列同一性;或者与SEQ ID NO:19的成熟多肽编码序列至少90%序列同一性;或者其cDNA序列;

[0132] (d)变体,其包含SEQ ID NO:2,SEQ ID NO:4,SEQ ID NO:6,SEQ ID NO:8,SEQ ID NO:10,SEQ ID NO:12,SEQ ID NO:14,SEQ ID NO:16,SEQ ID NO:18,SEQ ID NO:20,或SEQ ID NO:22的成熟多肽的一个或多个(几个)取代,缺失,和/或插入;和

[0133] (e)(a)、(b)、(c)或(d)的多肽具有纤维素分解增强活性的片段。

[0134] 本发明涉及分离的多肽,其具有与SEQ ID NO:2,SEQ ID NO:6,或SEQ ID NO:12的成熟多肽至少60%,例如,至少65%,至少70%,至少75%,至少80%,至少81%,至少82%,至少83%,至少84%,至少85%,至少86%,至少87%,至少88%,至少89%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%;与SEQ ID NO:4的成熟多肽至少65%,例如,至少70%,至少75%,至少80%,至少81%,至少82%,至少83%,至少84%,至少85%,至少86%,至少87%,至少88%,至少89%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%;与SEQ ID NO:18的成熟多肽至少70%,例如,至少75%,至少80%,至少81%,至少82%,至少83%,至少84%,至少85%,至少86%,至少87%,至少88%,至少89%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%;与SEQ ID NO:10,SEQ ID NO:16,或SEQ ID NO:22的成熟多肽至少75%,例如,至少80%,至少81%,至少82%,至少83%,至少84%,至少85%,至少86%,至少87%,至少88%,至少89%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%;与SEQ ID NO:8的成熟多肽至少80%,例如,至少81%,至少82%,至少83%,至少84%,至少85%,至少86%,至少87%,至少88%,至少89%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%;与SEQ ID NO:14的成熟多肽至少85%,例如,至少86%,至少87%,至少88%,至少89%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%;或者与SEQ ID NO:20的成熟多肽至少90%,例如,至少91%,至少92%,至少93%,至少94%,至少95%,至

少96%，至少97%，至少98%，至少99%，或100%的序列同一性，并具有纤维素分解增强活性。

[0135] 在一个方面，所述多肽与SEQ ID NO:2,SEQ ID NO:4,SEQ ID NO:6,SEQ ID NO:8,SEQ ID NO:10,SEQ ID NO:12,SEQ ID NO:14,SEQ ID NO:16,SEQ ID NO:18,SEQ ID NO:20,或SEQ ID NO:22的成熟多肽相差不超过十个氨基酸，例如相差五个氨基酸，相差四个氨基酸，相差三个氨基酸，相差两个氨基酸，和相差一个氨基酸。

[0136] 本发明的多肽优选包含或组成为SEQ ID NO:2,SEQ ID NO:4,SEQ ID NO:6,SEQ ID NO:8,SEQ ID NO:10,SEQ ID NO:12,SEQ ID NO:14,SEQ ID NO:16,SEQ ID NO:18,SEQ ID NO:20,或SEQ ID NO:22的氨基酸序列，或其等位变体；或为它们具有纤维素分解增强活性的片段。在另一个方面，多肽包含或组成为SEQ ID NO:2,SEQ ID NO:4,SEQ ID NO:6,SEQ ID NO:8,SEQ ID NO:10,SEQ ID NO:12,SEQ ID NO:14,SEQ ID NO:16,SEQ ID NO:18,SEQ ID NO:20,或SEQ ID NO:22的成熟多肽。在另一个优选的方面，多肽包含或组成为SEQ ID NO:2的氨基酸18至246。在另一个优选的方面，多肽包含或组成为SEQ ID NO:4的氨基酸20至334。在另一个优选的方面，多肽包含或组成为SEQ ID NO:6的氨基酸18至227。在另一个优选的方面，多肽包含或组成为SEQ ID NO:8的氨基酸20至223。在另一个优选的方面，多肽包含或组成为SEQ ID NO:10的氨基酸22至368。在另一个优选的方面，多肽包含或组成为SEQ ID NO:12的氨基酸25至330。在另一个优选的方面，多肽包含或组成为SEQ ID NO:14的氨基酸17至236。在另一个优选的方面，多肽包含或组成为SEQ ID NO:16的氨基酸19至250。在另一个优选的方面，多肽包含或组成为SEQ ID NO:18的氨基酸23至478。在另一个优选的方面，多肽包含或组成为SEQ ID NO:20的氨基酸17至230。在另一个优选的方面，多肽包含或组成为SEQ ID NO:22的氨基酸20至257。

[0137] 本发明还涉及具有纤维素分解增强活性的分离的多肽，其由多核苷酸编码，所述多核苷酸在中等-高，高或非常高严格条件下与以下杂交：(i)SEQ ID NO:1,SEQ ID NO:3,SEQ ID NO:5,SEQ ID NO:11,或SEQ ID NO:17的成熟多肽编码序列，(ii)包含于SEQ ID NO:1,SEQ ID NO:3,SEQ ID NO:5,SEQ ID NO:11,或SEQ ID NO:17的成熟多肽编码序列中的cDNA序列，或(iii)(i)或(ii)的全长互补链；在高或非常高严格条件下与以下杂交：(i)SEQ ID NO:7,SEQ ID NO:9,SEQ ID NO:13,SEQ ID NO:15,或SEQ ID NO:21的成熟多肽编码序列，(ii)包含于SEQ ID NO:7,SEQ ID NO:9,SEQ ID NO:13,SEQ ID NO:15,或SEQ ID NO:21的成熟多肽编码序列中的cDNA序列，或(iii)(i)或(ii)的全长互补链；或者在非常高严格条件下与以下杂交：i)SEQ ID NO:19的成熟多肽编码序列，(ii)包含于SEQ ID NO:19的成熟多肽编码序列中的cDNA序列，或(iii)(i)或(ii)的全长互补链(J. Sambrook, E.F. Fritsch, 和T. Maniatis, 1989, Molecular Cloning, A Laboratory Manual, 第2版, Cold Spring Harbor, New York)。

[0138] SEQ ID NO:1,SEQ ID NO:3,SEQ ID NO:5,SEQ ID NO:7,SEQ ID NO:9,SEQ ID NO:11,SEQ ID NO:13,SEQ ID NO:15,SEQ ID NO:17,SEQ ID NO:19,或SEQ ID NO:21的多核苷酸，或其亚序列，以及SEQ ID NO:2,SEQ ID NO:4,SEQ ID NO:6,SEQ ID NO:8,SEQ ID NO:10,SEQ ID NO:12,SEQ ID NO:14,SEQ ID NO:16,SEQ ID NO:18,SEQ ID NO:20,或SEQ ID NO:22的氨基酸序列，或其片段，可用于设计核酸探针，以根据本领域内公知的方法从不同属或种的菌株鉴定和克隆编码具有纤维素分解增强活性的多肽的DNA。具体而言，根据标

准的Southern印迹方法,可将这些探针用于与感兴趣的属或种的基因组DNA或cDNA杂交,以鉴定和分离其中相应的基因。这些探针可明显短于完整序列,但长度上应为至少14,例如至少25,至少35,或至少70个核苷酸。优选地,所述核酸探针是至少100个核苷酸的长度,例如,至少200个核苷酸,至少300个核苷酸,至少400个核苷酸,至少500个,至少600个核苷酸,至少700个核苷酸,至少800个核苷酸,或至少900个核苷酸的长度。DNA和RNA探针二者均可使用。通常将探针标记以检测相应的基因(例如,用³²P、³H、³⁵S、生物素或抗生物素蛋白(avidin)标记)。这些探针涵盖于本发明中。

[0139] 可从由这些其它株(strain)制备的基因组DNA或cDNA文库中筛选DNA,所述DNA与上述探针杂交并且编码具有纤维素分解增强活性的多肽。可以通过琼脂糖或聚丙烯酰胺凝胶电泳,或通过其它分离技术分离来自这些其它株的基因组或其它DNA。可以将来自文库的DNA或分离的DNA转移至硝化纤维素(nitrocellulose)或其它合适的载体材料并且固定于其上。为了鉴定与SEQ ID NO:1,SEQ ID NO:3,SEQ ID NO:5,SEQ ID NO:7,SEQ ID NO:9,SEQ ID NO:11,SEQ ID NO:13,SEQ ID NO:15,SEQ ID NO:17,SEQ ID NO:19,或SEQ ID NO:21,或其亚序列同源的克隆或DNA,将所述载体材料优选用在Southern印迹中。

[0140] 就本发明而言,杂交表示多核苷酸在非常低至非常高的严格条件下与标记的核酸探针杂交,所述核酸探针对应于SEQ ID NO:1,SEQ ID NO:3,SEQ ID NO:5,SEQ ID NO:7,SEQ ID NO:9,SEQ ID NO:11,SEQ ID NO:13,SEQ ID NO:15,SEQ ID NO:17,SEQ ID NO:19,或SEQ ID NO:21;SEQ ID NO:1,SEQ ID NO:3,SEQ ID NO:5,SEQ ID NO:7,SEQ ID NO:9,SEQ ID NO:11,SEQ ID NO:13,SEQ ID NO:15,SEQ ID NO:17,SEQ ID NO:19,或SEQ ID NO:21的成熟多肽编码序列;包含于SEQ ID NO:1,SEQ ID NO:3,SEQ ID NO:5,SEQ ID NO:7,SEQ ID NO:9,SEQ ID NO:11,SEQ ID NO:13,SEQ ID NO:15,SEQ ID NO:17,SEQ ID NO:19,或SEQ ID NO:21的成熟多肽编码序列中的cDNA序列;其全长互补链;或它们的亚序列。可使用例如X射线片(X-ray film)检测在这些条件下与核酸探针杂交的分子。

[0141] 在一个方面,核酸探针是的成熟多肽编码序列SEQ ID NO:1,SEQ ID NO:3,SEQ ID NO:5,SEQ ID NO:7,SEQ ID NO:9,SEQ ID NO:11,SEQ ID NO:13,SEQ ID NO:15,SEQ ID NO:17,SEQ ID NO:19,或SEQ ID NO:21,或其cDNA序列。在另一个方面,核酸探针是SEQ ID NO:1的核苷酸52至875,SEQ ID NO:3的核苷酸58至1250,SEQ ID NO:5的核苷酸52至795,SEQ ID NO:7的核苷酸58至974,SEQ ID NO:9的核苷酸64至1104,SEQ ID NO:11的核苷酸73至990,SEQ ID NO:13的核苷酸49至1218,SEQ ID NO:15的核苷酸55至930,SEQ ID NO:17的核苷酸67至1581,SEQ ID NO:19的核苷酸49至865,或SEQ ID NO:21的核苷酸58至1065。在另一个方面,核酸探针是多核苷酸,其编码SEQ ID NO:2,SEQ ID NO:4,SEQ ID NO:6,SEQ ID NO:8,SEQ ID NO:10,SEQ ID NO:12,SEQ ID NO:14,SEQ ID NO:16,SEQ ID NO:18,SEQ ID NO:20,或SEQ ID NO:22的多肽,或其成熟多肽;或其片段。在另一个优选的方面,核酸探针是SEQ ID NO:1,SEQ ID NO:3,SEQ ID NO:5,SEQ ID NO:7,SEQ ID NO:9,SEQ ID NO:11,SEQ ID NO:13,SEQ ID NO:15,SEQ ID NO:17,SEQ ID NO:19,或SEQ ID NO:21,或其cDNA序列。在另一个方面,核酸探针是包含在大肠杆菌(E.coli)NRRL B-50301中的质粒pSMai216中含有的多核苷酸,其中所述多核苷酸序列编码具有纤维素分解增强活性的多肽。在另一个方面,核酸探针是包含在大肠杆菌NRRL B-50301中的质粒pSMai21中含有的成熟多肽编码区。在另一个方面,核酸探针是包含在大肠杆菌NRRL B-50302中的质粒pSMai217中含有

的多核苷酸,其中所述多核苷酸序列编码具有纤维素分解增强活性的多肽。在另一个方面,核酸探针是包含在大肠杆菌NRRL B-50302中的质粒pSMai217中含有的成熟多肽编码区。在另一个方面,核酸探针是包含在大肠杆菌NRRL B-50303中的质粒pSMai218中含有的多核苷酸,其中所述多核苷酸序列编码具有纤维素分解增强活性的多肽。在另一个方面,核酸探针是包含在大肠杆菌NRRL B-50303中的质粒pSMai218中含有的成熟多肽编码区。在另一个方面,核酸探针是包含在大肠杆菌NRRL B-50300中的质粒pSMai213中含有的多核苷酸,其中所述多核苷酸序列编码具有纤维素分解增强活性的多肽。在另一个方面,核酸探针是包含在大肠杆菌NRRL B-50300中的质粒pSMai213中含有的成熟多肽编码区。在另一个方面,核酸探针是包含在大肠杆菌NRRL B-50320中的质粒pAG68中含有的多核苷酸,其中所述多核苷酸序列编码具有纤维素分解增强活性的多肽。在另一个方面,核酸探针是包含在大肠杆菌NRRL B-50320中的质粒pAG68中含有的成熟多肽编码区。在另一个方面,核酸探针是包含在大肠杆菌NRRL B-50321中的质粒pAG69中含有的多核苷酸,其中所述多核苷酸序列编码具有纤维素分解增强活性的多肽。在另一个方面,核酸探针是包含在大肠杆菌NRRL B-50321中的质粒pAG69中含有的成熟多肽编码区。在另一个方面,核酸探针是包含在大肠杆菌NRRL B-50322中的质粒pAG75中含有的多核苷酸,其中所述多核苷酸序列编码具有纤维素分解增强活性的多肽。在另一个方面,核酸探针是包含在大肠杆菌NRRL B-50322中的质粒pAG75中含有的成熟多肽编码区。在另一个方面,核酸探针是包含在大肠杆菌NRRL B-50323中的质粒pAG76中含有的多核苷酸,其中所述多核苷酸序列编码具有纤维素分解增强活性的多肽。在另一个方面,核酸探针是包含在大肠杆菌NRRL B-50323中的质粒pAG76中含有的成熟多肽编码区。在另一个方面,核酸探针是包含在大肠杆菌NRRL B-50324中的质粒pAG77中含有的多核苷酸,其中所述多核苷酸序列编码具有纤维素分解增强活性的多肽。在另一个方面,核酸探针是包含在大肠杆菌NRRL B-50324中的质粒pAG77中含有的成熟多肽编码区。在另一个方面,核酸探针是包含在大肠杆菌NRRL B-50325中的质粒pAG78中含有的多核苷酸,其中所述多核苷酸序列编码具有纤维素分解增强活性的多肽。在另一个方面,核酸探针是包含在大肠杆菌NRRL B-50325中的质粒pAG78中含有的成熟多肽编码区。在另一个方面,核酸探针是包含在大肠杆菌NRRL B-50326中的质粒pAG79中含有的多核苷酸,其中所述多核苷酸序列编码具有纤维素分解增强活性的多肽。在另一个方面,核酸探针是包含在大肠杆菌NRRL B-50326中的质粒pAG79中含有的成熟多肽编码区。

[0142] 对于长度至少100个核苷酸的长探针,将非常低至非常高的严格条件定义为在42℃,在5X SSPE、0.3%SDS、200微克/ml已剪切并且变性的鲑精DNA中,并且对于非常低和低严格性为25%的甲酰胺、对于中和中-高严格性为35%的甲酰胺、或对于高和非常高严格性为50%的甲酰胺,根据标准的Southern印迹法进行预杂交和杂交最佳12至24小时。使用2X SSC、0.2%SDS在45℃(非常低严格性),在50℃(低严格性),在55℃(中严格性),在60℃(中-高严格性),在65℃(高严格性),和在70℃(非常高严格性)将载体材料最终洗涤三次,每次15分钟。

[0143] 对于长度大约15个核苷酸至大约70个核苷酸的短探针,将严格条件定义为在比使用根据Bolton和McCarthy算法(1962,Proc.Natl.Acad.Sci.USA48:1390)计算的 T_m 低大约5℃至大约10℃,在0.9M NaCl,0.09M Tris-HCl pH7.6,6mM EDTA,0.5%NP-40,1×Denhardt溶液,1mM焦磷酸钠(sodium pyrophosphate),1mM磷酸二氢钠(sodium monobasic

phosphate), 0.1mM ATP和0.2mg每ml的酵母RNA中,根据标准的Southern印迹步骤进行预杂交和杂交最佳12至24小时。将所述载体材料在6×SSC加0.1%SDS中最终洗涤一次15分钟,并用6×SSC在比计算的 T_m 低5℃至10℃的温度洗涤两次,每次15分钟。

[0144] 本发明还涉及由多核苷酸编码的具有纤维素分解增强活性的分离的多肽,所述多核苷酸具有与SEQ ID NO:1,SEQ ID NO:5,或SEQ ID NO:11的成熟多肽编码序列至少60%,至少60%,例如,至少65%,至少70%,至少75%,至少80%,至少81%,至少82%,至少83%,至少84%,至少85%,至少86%,至少87%,至少88%,至少89%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%;与SEQ ID NO:3的成熟多肽至少65%,例如,至少70%,至少75%,至少80%,至少81%,至少82%,至少83%,至少84%,至少85%,至少86%,至少87%,至少88%,至少89%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%;与SEQ ID NO:17的成熟多肽至少70%,例如,至少75%,至少80%,至少81%,至少82%,至少83%,至少84%,至少85%,至少86%,至少87%,至少88%,至少89%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%;与SEQ ID NO:9,SEQ ID NO:15,或SEQ ID NO:21的成熟多肽至少75%,例如,至少80%,至少81%,至少82%,至少83%,至少84%,至少85%,至少86%,至少87%,至少88%,至少89%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%;与SEQ ID NO:7的成熟多肽至少80%,例如,至少81%,至少82%,至少83%,至少84%,至少85%,至少86%,至少87%,至少88%,至少89%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%;与SEQ ID NO:13的成熟多肽至少85%,例如,至少86%,至少87%,至少88%,至少89%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%;或者与SEQ ID NO:19的成熟多肽至少90%,例如,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%的序列同一性。

[0145] 本发明还涉及SEQ ID NO:2,SEQ ID NO:4,SEQ ID NO:6,SEQ ID NO:8,SEQ ID NO:10,SEQ ID NO:12,SEQ ID NO:14,SEQ ID NO:16,SEQ ID NO:18,SEQ ID NO:20,或SEQ ID NO:22的成熟多肽或其同源序列的包含取代、缺失和/或插入一个或多个(几个)氨基酸的变体。优选地,氨基酸改变对性质是较不重要的(of a minor nature),即保守的氨基酸取代或插入,其不显著影响蛋白质的折叠和/或活性;通常为1至大约30个氨基酸的小缺失;小的氨基或羧基末端延伸,如氨基末端甲硫氨酸残基;多至大约20-25个残基的小接头肽;或通过改变净电荷或其它功能来促进纯化的小延伸,如多组氨酸序列(poly histidine tract)、抗原表位(antigenic epitope)或结合域(binding domain)。

[0146] 保守取代的实例是在以下组之内:碱性氨基酸组(精氨酸、赖氨酸和组氨酸)、酸性氨基酸组(谷氨酸和天冬氨酸)、极性氨基酸组(谷氨酰胺和天冬酰胺)、疏水性氨基酸组(亮氨酸、异亮氨酸和缬氨酸)、芳族氨基酸组(苯丙氨酸、色氨酸和酪氨酸)和小氨基酸组(甘氨酸、丙氨酸、丝氨酸、苏氨酸和甲硫氨酸)。通常不改变比活性(specific activity)的氨基酸取代是本领域已知的,并且由例如H.Neurath和R.L.Hill,1979,于The Proteins, Academic Press, New York中描述。最普遍发生的交换是Ala/Ser、Val/Ile、Asp/Glu、Thr/

Ser、Ala/Gly、Ala/Thr、Ser/Asn、Ala/Val、Ser/Gly、Tyr/Phe、Ala/Pro、Lys/Arg、Asp/Asn、Leu/Ile、Leu/Val、Ala/Glu和Asp/Gly。

[0147] 可供选择的是,氨基酸改变具有这样的性质:使多肽的物理化学性质改变。例如,氨基酸改变可改进多肽的热稳定性,改变底物特异性,改变最适pH等。

[0148] 能够根据本领域已知的方法,例如定位诱变或丙氨酸分区诱变法(Cunningham和Wells,1989,Science 244:1081-1085)来鉴定亲本多肽中的必需氨基酸。在后一技术中,将单一丙氨酸突变引入到分子中的每个残基,并且测试所得突变分子的纤维素分解增强活性以鉴定对于所述分子的活性关键的氨基酸残基。同样参见Hilton等,1996,J.Biol.Chem.271:4699-4708。酶的活性部位或其它的生物相互作用也能够通过结构的物理分析而测定,如通过以下这些技术:如核磁共振、晶体学、电子衍射或光亲和标记,连同推定的接触位点氨基酸的突变来测定。参见例如de Vos等,1992,Science 255:306-312;Smith等,1992,J.Mol.Biol.224:899-904;Wlodaver等,1992,FEBS Lett.309:59-64。必需氨基酸的身份(identity)也能够从与多肽的同一性分析来推断,所述多肽与亲本多肽相关。

[0149] 能够使用已知的诱变、重组和/或改组(shuffling)方法,然后是有关的筛选方法,例如那些由Reidhaar-Olson和Sauer,1988,Science 241:53-57;Bowie和Sauer,1989,Proc.Natl.Acad.Sci.USA 86:2152-2156;WO 95/17413;或WO 95/22625公开的那些方法进行并测试单个或多个氨基酸取代、缺失和/或插入。能够使用的其它方法包括易错PCR、噬菌体展示(例如,Lowman等,1991,Biochemistry 30:10832-10837;美国专利No.5,223,409;WO 92/06204)和区域定向的诱变(Derbyshire等,1986,Gene 46:145;Ner等,1988,DNA 7:127)。

[0150] 诱变/改组方法能够与高通量、自动化的筛选方法组合以检测由宿主细胞表达的克隆的、诱变的多肽的活性(Ness等,1999,Nature Biotechnology 17:893-896)。能够从宿主细胞回收编码活性多肽的诱变的DNA分子,并且使用本领域内标准方法快速测序。这些方法允许快速测定多肽中单个氨基酸残基的重要性。

[0151] SEQ ID NO:2,SEQ ID NO:4,SEQ ID NO:6,SEQ ID NO:8,SEQ ID NO:10,SEQ ID NO:12,SEQ ID NO:14,SEQ ID NO:16,SEQ ID NO:18,SEQ ID NO:20,或SEQ ID NO:22的成熟多肽的氨基酸取代、缺失和/或插入的总数不多于10,例如1,2,3,4,5,6,7,8或9。

[0152] 所述多肽可为杂合多肽,其中一种多肽的一部分融合于另一种多肽的一部分的N端或C端。

[0153] 所述多肽可为融合多肽或可切割的融合多肽,其中另一种多肽融合于本发明多肽的N端或C端。通过将编码另一个多肽的多核苷酸融合于本发明的多核苷酸来产生融合的多肽。产生融合多肽的技术是本领域已知的,并包括连接编码多肽的编码序列以使它们在阅读框中,并且使融合多肽的表达在相同启动子和终止子的控制下。融合蛋白亦可使用内蛋白(intein)技术构建,其中融合物在翻译后产生(Cooper等,1993,EMBO J.12:2575-2583;Dawson等,1994,Science 266:776-779)。

[0154] 融合多肽还可以包含在两个多肽之间的切割位点。一旦分泌了融合多肽,就切割所述位点,释放所述两个多肽。切割位点的实例包括,但不限于,公开于Martin等,2003,J.Ind.Microbiol.Biotechnol.3:568-76;Svetina等,2000,J.Biotechnol.76:245-251;

Rasmussen-Wilson等,1997,Appl.Environ.Microbiol.63:3488-3493;Ward等,1995,Biotechnology 13:498-503;和Contreras等,1991,Biotechnology 9:378-381;Eaton等,1986,Biochem.25:505-512);Collins-Racie等,1995,Biotechnology 13:982-987;Carter等,1989,Proteins:Structure,Function,and Genetics 6:240-248;以及Stevens,2003,Drug Discovery World 4:35-48中的位点。

[0155] 具有纤维素分解增强活性的多肽的来源

[0156] 本发明的具有纤维素分解增强活性的多肽可以获得自任何属的微生物。就本发明而言,用于本文与给定的来源有关的术语“获得自”,意思应为多核苷酸编码的多肽由所述来源产生,或由其中插入了来自所述来源的多核苷酸的菌株产生。在一个方面,获得自给定来源的多肽是胞外分泌的。

[0157] 所述多肽可以是细菌多肽。例如,所述多肽可以是具有纤维素分解增强活性的革兰氏阳性细菌多肽例如芽孢杆菌属(*Bacillus*)、梭菌属(*Clostridium*)、肠球菌属(*Enterococcus*)、地芽孢杆菌属(*Geobacillus*)、乳杆菌属(*Lactobacillus*)、乳球菌属(*Lactococcus*)、海洋芽孢杆菌属(*Oceanobacillus*)、葡萄球菌属(*Staphylococcus*)、链球菌属(*Streptococcus*)、或链霉菌属(*Streptomyces*)多肽;或革兰氏阴性细菌多肽,如弯曲杆菌属(*Campylobacter*)、大肠杆菌(*E.coli*)、黄杆菌属(*Flavobacterium*)、梭杆菌属(*Fusobacterium*)、螺杆菌属(*Helicobacter*)、泥杆菌属(*Ilyobacter*)、奈瑟氏菌属(*Neisseria*)、假单胞菌属(*Pseudomonas*)、沙门氏菌属(*Salmonella*)或脲原体属(*Ureaplasma*)多肽。

[0158] 在一个方面,所述多肽是嗜碱芽孢杆菌(*Bacillus alkalophilus*)、解淀粉芽孢杆菌(*Bacillus amyloliquefaciens*)、短芽孢杆菌(*Bacillus brevis*)、环状芽孢杆菌(*Bacillus circulans*)、克劳氏芽孢杆菌(*Bacillus clausii*)、凝结芽孢杆菌(*Bacillus coagulans*)、坚强芽孢杆菌(*Bacillus firmus*)、灿烂芽孢杆菌(*Bacillus lautus*)、迟缓芽孢杆菌(*Bacillus lentus*)、地衣芽孢杆菌(*Bacillus licheniformis*)、巨大芽孢杆菌(*Bacillus megaterium*)、短小芽孢杆菌(*Bacillus pumilus*)、嗜热脂肪芽孢杆菌(*Bacillus stearothermophilus*)、枯草芽孢杆菌(*Bacillus subtilis*)或苏云金芽孢杆菌(*Bacillus thuringiensis*)多肽。

[0159] 在另一个方面,所述多肽是似马链球菌(*Streptococcus equisimilis*)、酿脓链球菌(*Streptococcus pyogenes*)、乳房链球菌(*Streptococcus uberis*)或马链球菌兽瘟亚种(*Streptococcus equi* subsp.*Zooepidemicus*)多肽。

[0160] 在另一个方面,所述多肽是不产色链霉菌(*Streptomyces achromogenes*)、除虫链霉菌(*Streptomyces avermitilis*)、天蓝链霉菌(*Streptomyces coelicolor*)、灰色链霉菌(*Streptomyces griseus*)或浅青紫链霉菌(*Streptomyces lividans*)多肽。

[0161] 所述多肽可以是真菌多肽。例如,所述多肽可为酵母多肽如假丝酵母属(*Candida*)、克鲁维酵母属(*Kluyveromyces*)、毕赤酵母属(*Pichia*)、酵母属(*Saccharomyces*)、裂殖酵母属(*Schizosaccharomyces*)或西洋蓍霉属(*Yarrowia*)多肽;或丝状真菌多肽如枝顶孢霉属(*Acremonium*)、伞菌属(*Agaricus*)、链格孢属(*Alternaria*)、曲霉属(*Aspergillus*)、短梗霉属(*Aureobasidium*)、*Botryosphaeria*、拟蜡菌属(*Ceriporiopsis*)、毛喙壳属(*Chaetomidium*)、金孢子菌属(*Chrysosporium*)、*Claviceps*、

Cochliobolus、鬼伞属(*Coprinopsis*)、*Coptotermes*、棒囊壳属(*Corynascus*)、隐丛赤壳菌属(*Cryphonectria*)、隐球菌属(*Cryptococcus*)、色二孢属(*Diplodia*)、黑耳属(*Exidia*)、*Filibasidium*、镰孢属(*Fusarium*)、赤霉属(*Gibberella*)、全鞭毛虫属(*Holomastigotoides*)、腐质霉属(*Humicola*)、耙齿菌属(*Irpex*)、蘑菇属(*Lentinula*)、*Leptosphaeria*、梨孢菌属(*Magnaporthe*)、*Melanocarpus*、多孔菌属(*Meripilus*)、毛霉属(*Mucor*)、毁丝霉属(*Myceliophthora*)、新考玛脂霉属(*Neocallimastix*)、脉孢菌属(*Neurospora*)、拟青霉属(*Paecilomyces*)、青霉属(*Penicillium*)、平革菌属(*Phanerochaete*)、瘤胃壶菌属(*Piromyces*)、*Poitrasia*、假黑盘菌属(*Pseudoplectania*)、*Pseudotrichonympha*、根毛霉属(*Rhizomucor*)、裂褶菌属(*Schizophyllum*)、柱顶孢属(*Scytalidium*)、踝节菌属(*Talaromyces*)、嗜热子囊菌属(*Thermoascus*)、梭孢壳属(*Thielavia*)、弯颈霉属(*Tolyposcladium*)、木霉属(*Trichoderma*)、长毛盘菌属(*Trichophaea*)、轮枝孢属(*Verticillium*)、包脚菇属(*Volvariella*)或炭角菌属(*Xylaria*)多肽。

[0162] 在另一个方面,所述多肽是卡尔酵母(*Saccharomyces carlsbergensis*)、酿酒酵母(*Saccharomyces cerevisiae*)、糖化酵母(*Saccharomyces diastaticus*)、道格拉氏酵母(*Saccharomyces douglasii*)、克鲁弗酵母(*Saccharomyces kluyveri*)、诺地酵母(*Saccharomyces norbensis*)或卵形酵母(*Saccharomyces oviformis*)多肽。

[0163] 在另一个优选的方面,所述多肽是解纤维枝顶孢霉(*Acremonium cellulolyticum*)、棘孢曲霉(*Aspergillus aculeatus*)、泡盛曲霉(*Aspergillus awamori*)、臭曲霉(*Aspergillus foetidus*)、烟曲霉(*Aspergillus fumigatus*)、日本曲霉(*Aspergillus japonicus*)、构巢曲霉(*Aspergillus nidulans*)、黑曲霉(*Aspergillus niger*)、米曲霉(*Aspergillus oryzae*)、*Chrysosporium inops*、嗜角质金孢子菌(*Chrysosporium keratinophilum*)、*Chrysosporium lucknowense*、*Chrysosporium merdarium*、毡金孢子菌(*Chrysosporium pannicola*)、*Chrysosporium queenslandicum*、热带金孢子菌(*Chrysosporium tropicum*)、*Chrysosporium zonatum*、杆孢状镰孢(*Fusarium bactridioides*)、禾谷镰孢(*Fusarium cerealis*)、库威镰孢(*Fusarium crookwellense*)、大刀镰孢(*Fusarium culmorum*)、禾本科镰孢(*Fusarium graminearum*)、禾赤镰孢(*Fusarium graminum*)、异孢镰孢(*Fusarium heterosporum*)、合欢木镰孢(*Fusarium negundi*)、尖镰孢(*Fusarium oxysporum*)、多枝镰孢(*Fusarium reticulatum*)、粉红镰孢(*Fusarium roseum*)、接骨木镰孢(*Fusarium sambucinum*)、肤色镰孢(*Fusarium sarcochromum*)、拟分枝孢镰孢(*Fusarium sporotrichioides*)、硫色镰孢(*Fusarium sulphureum*)、圆镰孢(*Fusarium torulosum*)、拟丝孢镰孢(*Fusarium trichothecioides*)、镶片镰孢(*Fusarium venenatum*)、灰腐质霉(*Humicola grisea*)、特异腐质霉(*Humicola insolens*)、疏棉状腐质霉(*Humicola lanuginosa*)、白耙齿菌(*Irpex lacteus*)、米黑毛霉(*Mucor miehei*)、嗜热毁丝霉(*Myceliophthora thermophila*)、粗糙脉孢菌(*Neurospora crassa*)、绳状青霉(*Penicillium funiculosum*)、产紫青霉(*Penicillium purpurogenum*)、黄孢平革菌(*Phanerochaete chrysosporium*)、无色梭孢壳(*Thielavia achromatica*)、*Thielavia albomyces*、*Thielavia albopilosa*、澳洲梭孢壳(*Thielavia australeinsis*)、*Thielavia fimeti*、小孢梭孢壳(*Thielavia microspora*)、卵孢梭孢壳(*Thielavia*

ovispora)、*Thielavia peruviana*、毛梭孢壳(*Thielavia setosa*)、瘤孢梭孢壳(*Thielavia spededonium*)、*Thielavia subthermophila*、哈茨木霉(*Trichoderma harzianum*)、康宁木霉(*Trichoderma koningii*)、长枝木霉(*Trichoderma longibrachiatum*)、里氏木霉(*Trichoderma reesei*)或绿色木霉(*Trichoderma viride*)多肽。

[0164] 在另一个方面,所述多肽是具有纤维素分解增强活性的土生梭孢壳多肽。在另一个方面,所述多肽是具有纤维素分解增强活性的棘孢曲霉NRRL 8126多肽,例如包含SEQ ID NO:2,SEQ ID NO:4,SEQ ID NO:6,SEQ ID NO:8,SEQ ID NO:10,SEQ ID NO:12,SEQ ID NO:14,SEQ ID NO:16,SEQ ID NO:18,SEQ ID NO:20,或SEQ ID NO:22的成熟多肽的多肽。

[0165] 可以理解的是对于前述的种,本发明包含完全和不完全阶段(perfect and imperfect states),和其它分类学的等同物(equivalent),例如无性型(anamorph),而无论它们已知的种名。本领域技术人员将容易地识别适合的等同物的身份。

[0166] 这些种的菌株在许多培养物保藏中心对于公众能够容易地取得,所述保藏中心诸如美国典型培养物保藏中心(the American Type Culture Collection)(ATCC)、德意志微生物和细胞培养物保藏中心(Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH)(DSMZ)、真菌菌种保藏中心(Centraalbureau Voor Schimmelcultures)(CBS)和农业研究机构专利培养物保藏中心北区研究中心(Agricultural Research Service Patent Culture Collection,Northern Regional Research Center)(NRRL)。

[0167] 可以使用上述的探针从其它来源,包括从自然界(例如,土壤、堆肥、水等)分离的微生物鉴定和获得所述多肽。用于从天然生境(habitat)分离微生物的技术是本领域内公知的。随后可通过相似地筛选另一种微生物的基因组DNA或cDNA文库或混合的DNA样品来获得编码所述多肽的多核苷酸。一旦用所述探针检测到编码多肽的多核苷酸,就能够使用本领域普通技术人员熟知的技术将所述多核苷酸分离或克隆(参见,例如,Sambrook等,1989,见上文)。

[0168] 多核苷酸

[0169] 本发明还涉及编码本发明多肽的分离的多核苷酸。

[0170] 用于分离或克隆编码多肽的多核苷酸的技术是本领域内已知的,包括从基因组DNA分离,从cDNA制备,或其组合。可通过例如使用熟知的聚合酶链式反应(PCR)或表达文库的抗体筛选来检测具有共有结构特性的克隆DNA片段,从而实现从这种基因组DNA克隆多核苷酸。参见,例如,Innis等,1990,PCR:A Guide to Methods and Application,Academic Press,New York。可以使用其它核酸扩增方法,如连接酶链式反应(LCR)、连接活化转录(ligated activated transcription;LAT)和基于多核苷酸的扩增(NASBA)。可以从土生梭孢壳菌株,或相关生物体克隆所述多核苷酸,并且因此可为例如所述多核苷酸的多肽编码区的等位基因变体或种变体(species variant)。

[0171] 本发明还涉及分离的多核苷酸,所述多核苷酸具有与SEQ ID NO:1,SEQ ID NO:5,或SEQ ID NO:11的成熟多肽编码序列至少60%,例如,至少65%,至少70%,至少75%,至少80%,至少81%,至少82%,至少83%,至少84%,至少85%,至少86%,至少87%,至少88%,至少89%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%;与SEQ ID NO:3的成熟多肽至少65%,例如,至少70%,

至少75%，至少80%，至少81%，至少82%，至少83%，至少84%，至少85%，至少86%，至少87%，至少88%，至少89%，至少90%，至少91%，至少92%，至少93%，至少94%，至少95%，至少96%，至少97%，至少98%，至少99%，或100%；与SEQ ID NO:17的成熟多肽至少70%，例如，至少75%，至少80%，至少81%，至少82%，至少83%，至少84%，至少85%，至少86%，至少87%，至少88%，至少89%，至少90%，至少91%，至少92%，至少93%，至少94%，至少95%，至少96%，至少97%，至少98%，至少99%，或100%；与SEQ ID NO:9, SEQ ID NO:15, 或SEQ ID NO:21的成熟多肽至少75%，例如，至少80%，至少81%，至少82%，至少83%，至少84%，至少85%，至少86%，至少87%，至少88%，至少89%，至少90%，至少91%，至少92%，至少93%，至少94%，至少95%，至少96%，至少97%，至少98%，至少99%，或100%；与SEQ ID NO:7的成熟多肽至少80%，例如，至少81%，至少82%，至少83%，至少84%，至少85%，至少86%，至少87%，至少88%，至少89%，至少90%，至少91%，至少92%，至少93%，至少94%，至少95%，至少96%，至少97%，至少98%，至少99%，或100%；与SEQ ID NO:13的成熟多肽至少85%，例如，至少86%，至少87%，至少88%，至少89%，至少90%，至少91%，至少92%，至少93%，至少94%，至少95%，至少96%，至少97%，至少98%，至少99%，或100%；或者与SEQ ID NO:19的成熟多肽至少90%，例如，至少91%，至少92%，至少93%，至少94%，至少95%，至少96%，至少97%，至少98%，至少99%，或100%，其编码具有纤维素分解增强活性的多肽。

[0172] 修饰编码本发明多肽的多核苷酸对于合成与所述多肽基本上相似的多肽可为必需的。术语与所述多肽“基本上相似”指多肽的非天然存在的形式。这些多肽可能以一些工程改造的方式而不同于从其天然来源分离的多肽，例如，比活性、热稳定性、最适pH等方面不同的变体。可以在作为SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, 或SEQ ID NO:21的成熟多肽编码序列或其cDNA序列存在的多核苷酸，例如其亚序列的基础上，和/或通过引入如下核苷酸取代来构建变体：所述取代不导致多肽氨基酸序列的改变，但是符合意欲产生酶的宿主生物体的密码子选择；或者所述取代可产生不同的氨基酸序列。关于核苷酸取代的概述，参见，例如，Ford等，1991, Protein Expression and Purification 2:95-107。

[0173] 本发明还涉及编码本发明多肽的分离的多核苷酸，所述分离的多核苷酸在中等-高，高或非常高严格条件下与以下杂交：(i) SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:11, 或SEQ ID NO:17的成熟多肽编码序列，(ii) 包含于SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:11, 或SEQ ID NO:17的成熟多肽编码序列中的cDNA序列，或(iii) (i)或(ii)的全长互补链；在高或非常高严格条件下与以下杂交：(i) SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:13, SEQ ID NO:15, 或SEQ ID NO:21的成熟多肽编码序列，(ii) 包含于SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:13, SEQ ID NO:15, 或SEQ ID NO:21的成熟多肽编码序列中的cDNA序列，或(iii) (i)或(ii)的全长互补链；或者在非常高严格条件下与以下杂交：i) SEQ ID NO:19的成熟多肽编码序列，(ii) 包含于SEQ ID NO:19的成熟多肽编码序列中的cDNA序列，或(iii) (i)或(ii)的全长互补链；或它们的等位变体和亚序列(Sambrook等，1989, 见上文)，如本文所定义的。

[0174] 在一个方面，所述多核苷酸包含SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID

NO:19,或SEQ ID NO:21,或SEQ ID NO:1,SEQ ID NO:3,SEQ ID NO:5,SEQ ID NO:7,SEQ ID NO:9,SEQ ID NO:11,SEQ ID NO:13,SEQ ID NO:15,SEQ ID NO:17,SEQ ID NO:19,或SEQ ID NO:21的亚序列,其分别编码SEQ ID NO:2,SEQ ID NO:4,SEQ ID NO:6,SEQ ID NO:8,SEQ ID NO:10,SEQ ID NO:12,SEQ ID NO:14,SEQ ID NO:16,SEQ ID NO:18,SEQ ID NO:20,或SEQ ID NO:22具有纤维素分解增强活性的片段,如SEQ ID NO:1的核苷酸52至875,SEQ ID NO:3的核苷酸58至1250,SEQ ID NO:5的核苷酸52至795,SEQ ID NO:7的核苷酸58至974,SEQ ID NO:9的核苷酸64至1104,SEQ ID NO:11的核苷酸73至990,SEQ ID NO:13的核苷酸49至1218,SEQ ID NO:15的核苷酸55至930,SEQ ID NO:17的核苷酸67至1581,SEQ ID NO:19的核苷酸49至865,或SEQ ID NO:21的核苷酸58至1065。

[0175] 在另一个方面,所述多核苷酸包含或组成为SEQ ID NO:1或其成熟多肽编码序列,其包含于质粒pSMai216,所述质粒包含于大肠杆菌NRRL B-50301,其中所述多核苷酸编码具有纤维素分解增强活性的多肽。在另一个方面,所述多核苷酸包含或组成为SEQ ID NO:3或其成熟多肽编码序列,其包含于质粒pSMai217,所述质粒包含于大肠杆菌NRRL B-50302,其中所述多核苷酸编码具有纤维素分解增强活性的多肽。在另一个方面,所述多核苷酸包含或组成为SEQ ID NO:5或其成熟多肽编码序列,其包含于质粒pSMai218,所述质粒包含于大肠杆菌NRRL B-50303,其中所述多核苷酸编码具有纤维素分解增强活性的多肽。在另一个方面,所述多核苷酸包含或组成为SEQ ID NO:7或其成熟多肽编码序列,其包含于质粒pSMai213,所述质粒包含于大肠杆菌NRRL B-50300,其中所述多核苷酸编码具有纤维素分解增强活性的多肽。在另一个方面,所述多核苷酸包含或组成为SEQ ID NO:9或其成熟多肽编码序列,其包含于质粒pAG68,所述质粒包含于大肠杆菌NRRL B-50320,其中所述多核苷酸编码具有纤维素分解增强活性的多肽。在另一个方面,所述多核苷酸包含或组成为SEQ ID NO:11或其成熟多肽编码序列,其包含于质粒pAG69,所述质粒包含于大肠杆菌NRRL B-50321,其中所述多核苷酸编码具有纤维素分解增强活性的多肽。在另一个方面,所述多核苷酸包含或组成为SEQ ID NO:13或其成熟多肽编码序列,其包含于质粒pAG75,所述质粒包含于大肠杆菌NRRL B-50322,其中所述多核苷酸编码具有纤维素分解增强活性的多肽。在另一个方面,所述多核苷酸包含或组成为SEQ ID NO:15或其成熟多肽编码序列,其包含于质粒pAG76,所述质粒包含于大肠杆菌NRRL B-50323,其中所述多核苷酸编码具有纤维素分解增强活性的多肽。在另一个方面,所述多核苷酸包含或组成为SEQ ID NO:17或其成熟多肽编码序列,其包含于质粒pAG77,所述质粒包含于大肠杆菌NRRL B-50324,其中所述多核苷酸编码具有纤维素分解增强活性的多肽。在另一个方面,所述多核苷酸包含或组成为SEQ ID NO:19或其成熟多肽编码序列,其包含于质粒pAG78,所述质粒包含于大肠杆菌NRRL B-50325,其中所述多核苷酸编码具有纤维素分解增强活性的多肽。在另一个方面,所述多核苷酸包含或组成为SEQ ID NO:21或其成熟多肽编码序列,其包含于质粒pAG79,所述质粒包含于大肠杆菌NRRL B-50326,其中所述多核苷酸编码具有纤维素分解增强活性的多肽。

[0176] 核酸构建体

[0177] 本发明还涉及包含本发明的多核苷酸的核酸构建体,所述多核苷酸与一个或多个(几个)调控序列可操作地连接,所述调控序列在合适的宿主细胞中在与该调控序列相容的条件下指导编码序列的表达。

[0178] 可以用许多方式操作所述多核苷酸以提供多肽的表达。依赖于表达载体,在将多

核苷酸插入载体之前对其进行操作可能是理想的或必需的。使用重组DNA方法修饰多核苷酸的技术是本领域熟知的。

[0179] 调控序列可为启动子序列,其是由用于表达编码本发明多肽的多核苷酸的宿主细胞所识别的多核苷酸。启动子序列含有介导多肽的表达的转录调控序列。启动子可以是在所选的宿主细胞中显示转录活性的任何多核苷酸,包括突变的、截短的和杂合的启动子,并且可以从编码与宿主细胞同源或异源的胞外或胞内多肽的基因获得。

[0180] 用于在细菌宿主细胞中指导本发明的核酸构建体转录的合适启动子的实例是从下述获得的启动子:解淀粉芽孢杆菌 α -淀粉酶基因(*amyQ*)、地衣芽孢杆菌 α -淀粉酶基因(*amyL*)、地衣芽孢杆菌青霉素酶基因(*penP*)、嗜热脂肪芽孢杆菌产麦芽淀粉酶基因(*amyM*)、枯草芽孢杆菌果聚糖蔗糖酶基因(*sacB*)、枯草芽孢杆菌*xylA*和*xylB*基因、大肠杆菌*lac*操纵子、天蓝链霉菌琼脂糖酶基因(*dagA*)和原核 β -内酰胺酶基因(Villa-Kamaroff等,1978, *Proceedings of the National Academy of Sciences USA* 75:3727-3731),以及*tac*启动子(DeBoer等,1983, *Proc. Natl. Acad. Sci. USA* 80:21-25)。另外的启动子在“Useful proteins from recombinant bacteria”于Gilbert等,1980, *Scientific American*, 242:74-94中;和在Sambrook等,1989,见上文中描述。

[0181] 用于指导本发明的核酸构建体在丝状真菌宿主细胞中转录的合适启动子的实例是从下列酶的基因获得的启动子:构巢曲霉乙酰胺酶、黑曲霉中性 α -淀粉酶、黑曲霉酸稳定性 α -淀粉酶、黑曲霉或泡盛曲霉葡糖淀粉酶(*glaA*)、米曲霉TAKA淀粉酶、米曲霉碱性蛋白酶、米曲霉丙糖磷酸异构酶、尖镰孢胰蛋白酶样蛋白酶(WO 96/00787)、镶片镰孢淀粉葡萄糖苷酶(WO 00/56900)、镶片镰孢Daria(WO 00/56900)、镶片镰孢Quinn(WO 00/56900)、曼赫根毛霉(*Rhizomucor miehei*)脂肪酶、曼赫根毛霉天冬氨酸蛋白酶、里氏木霉 β -葡糖苷酶、里氏木霉纤维二糖水分解酶I、里氏木霉纤维二糖水分解酶II、里氏木霉内切葡聚糖酶I、里氏木霉内切葡聚糖酶II、里氏木霉内切葡聚糖酶III、里氏木霉内切葡聚糖酶IV、里氏木霉内切葡聚糖酶V、里氏木霉木聚糖酶I、里氏木霉木聚糖酶II、里氏木霉 β -木糖苷酶,以及NA2-*tpi*启动子(一种修饰的启动子,其来自在曲霉属中编码中性 α -淀粉酶的基因,其中未翻译的前导序列由在曲霉属(*Aspergilli*)中编码丙糖磷酸异构酶的基因的未翻译的前导序列所替代;非限制性实例包括修饰的启动子,其来自在黑曲霉中编码中性 α -淀粉酶的基因,其中未翻译的前导序列由在构巢曲霉或米曲霉中编码丙糖磷酸异构酶的基因的未翻译的前导序列所替代);和它们的突变的、截短的和杂合的启动子。

[0182] 在酵母宿主中,有用的启动子从如下酶的基因获得:酿酒酵母烯醇化酶(*ENO-1*)、酿酒酵母半乳糖激酶(*GAL1*)、酿酒酵母醇脱氢酶/甘油醛-3-磷酸脱氢酶(*ADH1*, *ADH2/GAP*)、酿酒酵母丙糖磷酸异构酶(*TPI*)、酿酒酵母金属硫蛋白(*CUP1*)和酿酒酵母3-磷酸甘油酸激酶。对于酵母宿主细胞其它有用的启动子由Romanos等,1992, *Yeast* 8:423-488描述。

[0183] 调控序列也可以是合适的转录终止子序列,其由宿主细胞识别以终止转录。所述终止子序列与编码所述多肽的多核苷酸的3'末端可操作地连接。可以将所选宿主细胞中有功能的任何终止子用在本发明中。

[0184] 对于丝状真菌宿主细胞优选的终止子从如下酶的基因获得:构巢曲霉邻氨基苯甲酸合酶、黑曲霉葡糖淀粉酶、黑曲霉 α -葡糖苷酶、米曲霉TAKA淀粉酶和尖镰孢胰蛋白酶样蛋白酶。

[0185] 对于酵母宿主细胞优选的终止子从如下酶的基因获得：酿酒酵母烯醇化酶、酿酒酵母细胞色素C(CYC1)和酿酒酵母甘油醛-3-磷酸脱氢酶。对于酵母宿主细胞其它有用的终止子由Romanos等,1992,见上文描述。

[0186] 调控序列还可以是合适的前导序列,当被转录时其为对于宿主细胞的翻译重要的mRNA非翻译区。前导序列可操作地连接于编码多肽的多核苷酸的5'-末端。可使用在所选宿主细胞中有功能的任何前导序列。

[0187] 对于丝状真菌宿主细胞优选的前导序列从如下酶的基因获得：米曲霉TAKA淀粉酶和构巢曲霉丙糖磷酸异构酶。

[0188] 对于酵母宿主细胞合适的前导序列从如下酶的基因获得：酿酒酵母烯醇化酶(ENO-1)、酿酒酵母3-磷酸甘油酸激酶、酿酒酵母 α 因子和酿酒酵母醇脱氢酶/甘油醛-3-磷酸脱氢酶(ADH2/GAP)。

[0189] 调控序列也可以是聚腺苷酸化序列,其是与多核苷酸的3'末端可操作地连接的序列,并且在转录时,宿主细胞将其识别为将聚腺苷残基添加至转录的mRNA的信号。可使用在所选宿主细胞中有功能的任何聚腺苷酸化序列。

[0190] 对于丝状真菌宿主细胞优选的聚腺苷酸化序列从如下酶的基因获得：米曲霉TAKA淀粉酶、黑曲霉葡糖淀粉酶、构巢曲霉邻氨基苯甲酸合酶、尖镰孢胰蛋白酶样蛋白酶和黑曲霉 α -葡糖苷酶。

[0191] 对于酵母宿主细胞有用的聚腺苷酸化序列由Guo和Sherman,1995,Mol.Cellular Biol.15:5983-5990描述。

[0192] 调控序列还可以是信号肽编码区,其编码与多肽的N端相连的信号肽,并且指导所述多肽进入细胞分泌途径。多核苷酸的编码序列5'端可固有地包含信号肽编码序列,其与编码所述多肽的编码序列的区段一起天然地连接在翻译阅读框中。可供选择的是,编码序列5'端可含有对于所述编码序列异源的信号肽编码序列。异源信号肽编码序列在编码序列不天然地含有信号肽编码序列时可为必需的。或者,外源信号肽编码序列可以简单地取代天然信号肽编码序列以增强多肽的分泌。然而,可使用指导表达的多肽进入所选宿主细胞的分泌途径的任何信号肽编码序列。

[0193] 对于细菌宿主细胞有效的信号肽编码序列是从如下酶的基因获得的信号肽编码序列：芽孢杆菌属NCIB 11837产麦芽糖淀粉酶、地衣芽孢杆菌枯草杆菌蛋白酶(subtilisin)、地衣芽孢杆菌 β -内酰胺酶、嗜热脂肪芽孢杆菌 α -淀粉酶、嗜热脂肪芽孢杆菌中性蛋白酶(nprT,nprS,nprM)和枯草芽孢杆菌prSA。另外的信号肽由Simonen和Palva,1993,Microbiological Reviews 57:109-137描述。

[0194] 对于丝状真菌宿主细胞有效的信号肽编码序列是从如下酶的基因获得的信号肽编码序列：黑曲霉中性淀粉酶、黑曲霉葡糖淀粉酶、米曲霉TAKA淀粉酶、特异腐质霉纤维素酶、特异腐质霉内切葡聚糖酶V、疏棉状腐质霉脂肪酶和曼赫根毛霉天冬氨酸蛋白酶。

[0195] 对于酵母宿主细胞有用的信号肽从酿酒酵母 α 因子和酿酒酵母转化酶的基因获得。其它有用的信号肽编码序列由Romanos等,1992,见上文描述。

[0196] 调控序列还可以是前肽编码序列,其编码位于多肽N端的前肽。所得多肽称为酶原(proenzyme)或前多肽(propolypeptide)(或在某些情况下称为酶原(zymogen))。前多肽通常是无活性的,并且能够通过前肽的催化或自催化切割从前多肽转化为活性多肽。可以从

枯草芽孢杆菌碱性蛋白酶(aprE)、枯草芽孢杆菌中性蛋白酶(nprT)、嗜热毁丝霉漆酶(WO 95/33836)、曼赫根毛霉天冬氨酸蛋白酶和酿酒酵母 α 因子的基因获得前肽编码序列。

[0197] 当信号肽和前肽序列二者均出现在多肽的N端时,将前肽序列置于紧接着(next to)多肽N端,并且将信号肽序列置于紧接着前肽序列的N端。

[0198] 同样理想的是添加调节序列,其允许相对于宿主细胞的生长来调节多肽的表达。调节系统的实例是引起基因表达响应化学或物理刺激物,包括调节化合物的存在而开启或关闭的那些系统。原核系统中的调节系统包括lac、tac和trp操纵基因系统。在酵母中,可使用ADH2系统或GAL1系统。在丝状真菌中,可以使用黑曲霉葡萄糖淀粉酶启动子、米曲霉TAKA α -淀粉酶启动子和米曲霉葡萄糖淀粉酶启动子。调节序列的其它实例是那些允许基因扩增的序列。在真核系统中,这些调节序列包括在氨甲蝶呤(methotrexate)存在下扩增的二氢叶酸还原酶基因,和以重金属(with heavy metal)扩增的金属硫蛋白基因。在这些情况下,编码多肽的多核苷酸将与调节序列可操作地连接。

[0199] 表达载体

[0200] 本发明还涉及重组表达载体,所述重组表达载体包含本发明的多核苷酸、启动子和转录和翻译终止信号。多种核苷酸和调控序列可以结合在一起以产生重组表达载体,所述表达载体可以包括一个或多个(几个)方便的限制位点以允许在这些位点插入或取代编码多肽的多核苷酸。可供选择的是,可以通过在适当的用于表达的载体中插入包含所述序列的多核苷酸或核酸构建体来表达所述多核苷酸。在制备表达载体的过程中,将编码序列置于载体中,从而将该编码序列与适当的表达调控序列可操作地连接。

[0201] 重组表达载体可以是任何载体(例如,质粒或病毒),其能够方便地进行重组DNA步骤,并且能够产生多核苷酸的表达。载体的选择将通常依赖于载体与将引入该载体的宿主细胞的相容性。载体可以是线状或闭合环状质粒。

[0202] 载体可以是自主复制载体,即,作为染色体外实体(entity)存在的载体,其复制独立于染色体复制,例如,质粒、染色体外元件、微型染色体(minichromosome)或人工染色体。载体可以含有任何用于确保自复制的手段(means)。或者,载体可以是一种当被引入宿主细胞中时,整合到基因组中并且与整合了该载体的染色体一起复制的载体。此外,可以使用单独的载体或质粒或两个或更多个载体或质粒,其共同含有待引入宿主细胞基因组的完整DNA(total DNA),或可以使用转座子(transposon)。

[0203] 所述载体优选地含有一个或多个(几个)选择性标记,其允许简单选择经转化、转染、转导等的细胞。选择性标记是基因,其产物提供杀生物剂或病毒抗性、对重金属的抗性、对营养缺陷型的原养性(protothy to auxotrophs)等。

[0204] 细菌选择性标记的实例是来自枯草芽孢杆菌或地衣芽孢杆菌的dal基因,或赋予抗生素抗性的标记,所述抗生素抗性例如氨苄青霉素、氯霉素、卡那霉素或四环素抗性。对于酵母宿主细胞合适的标记是ADE2、HIS3、LEU2、LYS2、MET3、TRP1和URA3。用于丝状真菌宿主细胞的选择性标记包括但不限于amdS(乙酰胺酶)、argB(鸟氨酸氨甲酰基转移酶)、bar(草铵膦(phosphinothricin)乙酰转移酶)、hph(潮霉素磷酸转移酶)、niaD(硝酸还原酶)(nitrate reductase)、pyrG(乳清酸核苷-5'-磷酸脱羧酶)(orotidine-5'-phosphate decarboxylase)、sC(硫酸腺苷酰转移酶)和trpC(邻氨基苯甲酸合酶(anthranilate synthase))以及它们的等同物。优选用在曲霉属细胞中的是构巢曲霉或米曲霉的amdS和

pyrG基因和吸水链霉菌(*Streptomyces hygroscopicus*)的bar基因。

[0205] 所述载体优选含有元件,其允许载体整合入宿主细胞基因组或载体在细胞中独立于基因组的自主复制。

[0206] 为了整合入宿主细胞基因组,载体可依赖编码多肽的多核苷酸的序列或用于通过同源或非同源重组整合入基因组的任何其它载体元件。或者,载体可以含有额外的多核苷酸,用于指导通过同源重组整合入宿主细胞基因组染色体中的精确位置。为了增加在精确位置整合的可能性,整合元件应含有足够数量的核酸,如100至10,000碱基对,400至10,000碱基对,和800至10,000碱基对,其与相应的目标序列具有高度序列同一性以增强同源重组的概率。整合元件可以是任何序列,其与宿主细胞基因组中的目标序列同源。此外,整合元件可以是非编码或编码的多核苷酸。另一方面,可以将载体通过非同源重组整合到宿主细胞的基因组中。

[0207] 为了自主复制,载体可以进一步包含复制起点,其使载体能够在所述的宿主细胞中自主地复制。复制起点可以是介导自主复制的任何质粒复制子(replicator),其在细胞中发挥功能。术语“复制起点”或“质粒复制子”意指能够使质粒或载体体内复制的多核苷酸。

[0208] 细菌复制起点的实例是允许在大肠杆菌中复制的质粒pBR322、pUC19、pACYC177和pACYC184的复制起点,和允许在芽孢杆菌属中复制的质粒pUB110、pE194、pTA1060和pAMB1的复制起点。

[0209] 用于酵母宿主细胞中的复制起点的实例是2微米复制起点,ARS1,ARS4,ARS1和CEN3的组合,和ARS4和CEN6的组合。

[0210] 在丝状真菌细胞中有用的复制起点的实例是AMA1和ANS1(Gems等,1991,Gene 98:61-67;Cullen等,1987,Nucleic Acids Res.15:9163-9175;WO 00/24883)。分离AMA1基因和构建包含该基因的质粒或载体能够根据公开于WO 00/24883中的方法完成。

[0211] 可以将多于一个拷贝的本发明的多核苷酸插入宿主细胞以增加多肽的产生。多核苷酸拷贝数的增加可通过如下方法获得:将至少一个额外拷贝的序列整合入宿主细胞基因组,或将可扩增的选择性标记基因包括于多核苷酸,其中可通过在合适的选择剂(selectable agent)存在下培养细胞来选择含有选择性标记基因的扩增拷贝,且由此含有多核苷酸的额外拷贝的细胞。

[0212] 用于连接上述元件以构建本发明的重组表达载体的方法是本领域技术人员熟知的(参见,例如,Sambrook等,1989,见上文)。

[0213] 宿主细胞

[0214] 本发明还涉及重组宿主细胞,其包含本发明的多核苷酸可操作地连接于一个或多个(几个)指导本发明多肽的产生的调控序列。将包含多核苷酸的构建体或载体导入宿主细胞,使所述构建体或载体如前所述作为染色体整体或者作为自复制的染色体外载体维持。术语“宿主细胞”包括亲本细胞的任何后代,其由于复制过程中发生的突变而不同于亲本细胞。宿主细胞的选择将在很大程度上依赖于编码多肽的基因及其来源。

[0215] 宿主细胞可以是在本发明的多肽的重组产生中有用的任何细胞,例如,原核或真核细胞。

[0216] 原核宿主细胞可以是任何革兰氏阳性或革兰氏阴性细菌。革兰氏阳性细菌包括但

不限于,芽孢杆菌属、梭菌属、肠球菌属、地芽孢杆菌属、乳杆菌属、乳球菌属、海洋芽孢杆菌属、葡萄球菌属、链球菌属和链霉菌属。革兰氏阴性细菌包括但不限于,弯曲杆菌属、大肠杆菌属、黄杆菌属、梭杆菌属、螺杆菌属、泥杆菌属、奈瑟氏菌属、假单胞菌属、沙门氏菌属和脲原体属。

[0217] 细菌宿主细胞可以是任何芽孢杆菌属细胞,包括但不限于嗜碱芽孢杆菌、解淀粉芽孢杆菌、短芽孢杆菌、环状芽孢杆菌、克劳氏芽孢杆菌、凝结芽孢杆菌、坚强芽孢杆菌、灿烂芽孢杆菌、迟缓芽孢杆菌、地衣芽孢杆菌、巨大芽孢杆菌、短小芽孢杆菌、嗜热脂肪芽孢杆菌、枯草芽孢杆菌和苏云金芽孢杆菌细胞。

[0218] 细菌宿主细胞还可以是任何链球菌属细胞,包括但不限于,似马链球菌、酿脓链球菌、乳房链球菌(*Streptococcus uberis*)和马链球菌兽瘟亚种细胞。

[0219] 细菌宿主细胞还可以是任何链霉菌属细胞,包括但不限于,不产色链霉菌、除虫链霉菌、天蓝链霉菌、灰色链霉菌和浅青紫链霉菌细胞。

[0220] 可通过如下方法实现将DNA引入到芽孢杆菌属细胞:例如原生质体转化(参见,例如,Chang和Cohen,1979,*Mol.Gen.Genet.*168:111-115),使用感受态细胞(参见,例如,Young和Spizizen,1961,*J.Bacteriol.*81:823-829或Dubnau和Davidoff-Abelson,1971,*J.Mol.Biol.*56:209-221),电穿孔(参见,例如,Shigekawa和Dower,1988,*Biotechniques* 6:742-751)或接合(参见,例如,Koehler和Thorne,1987,*J.Bacteriol.*169:5771-5278)。可通过如下方法实现将DNA引入到大肠杆菌细胞:例如原生质体转化(参见,例如,Hanahan,1983,*J.Mol.Biol.*166:557-580)或电穿孔(参见,例如,Dower等,1988,*Nucleic Acids Res.*16:6127-6145)。可通过如下方法实现将DNA引入到链霉菌属细胞:例如原生质体转化和电穿孔(参见,例如,Gong等,2004,*Folia Microbiol.(Praha)*49:399-405),接合(参见,例如,Mazodier等,1989,*J.Bacteriol.*171:3583-3585),或转导(参见,例如,Burke等,2001,*Proc.Natl.Acad.Sci.USA* 98:6289-6294)。可通过如下方法实现将DNA引入到假单胞菌属细胞:例如电穿孔(参见,例如,Choi等,2006,*J.Microbiol.Methods* 64:391-397)或接合(参见,例如,Pinedo和Smets,2005,*Appl.Environ.Microbiol.*71:51-57)。可通过如下方法实现将DNA引入到链球菌属细胞:例如天然感受态(natural competence)(参见,例如,Perry和Kuramitsu,1981,*Infect.Immun.*32:1295-1297),原生质体转化(参见,例如,Catt和Jollick,1991,*Microbios.*68:189-207),电穿孔(参见,例如,Buckley等,1999,*Appl.Environ.Microbiol.*65:3800-3804)或接合(参见,例如,Clewell,1981,*Microbiol.Rev.*45:409-436)。然而,可以使用本领域已知的将DNA引入宿主细胞的任何方法。

[0221] 宿主细胞还可以是真核生物,如哺乳动物、昆虫、植物或真菌细胞。

[0222] 宿主细胞可为真菌细胞。“真菌”用在本文包括以下门:子囊菌门(Ascomycota)、担子菌门(Basidiomycota)、壶菌门(Chytridiomycota)和接合菌门(Zygomycota)(如由Hawksworth等,于Ainsworth and Bisby's Dictionary of The Fungi,第8版,1995,CAB International,University Press,Cambridge,UK中所定义)以及卵菌门(Oomycota)(如Hawksworth等,1995,见上,171页中所引用),和所有有丝分裂孢子真菌(mitosporic fungi)(Hawksworth等,1995,见上文)。

[0223] 真菌宿主细胞可为酵母细胞。“酵母”用在本文包括产子囊酵母(ascosporogenous

yeast)(内孢霉目(Endomycetales))、产担子酵母(basidiosporogenous yeast)和属于半知菌类(Fungi Imperfecti)(芽孢纲(Blastomycetes))的酵母。由于酵母的分类在未来可能改变,就本发明而言,将酵母定义为如Biology and Activities of Yeast(Skinner, F.A., Passmore, S.M., 和Davenport, R.R. 编, Soc.App.Bacteriol.Symposium Series No.9, 1980)中所述。

[0224] 酵母宿主细胞可为假丝酵母属、汉逊酵母属(Hansenula)、克鲁维酵母属、毕赤酵母属、酵母属、裂殖酵母属或西洋薯霉属细胞,如乳酸克鲁维酵母(Kluyveromyces lactis)、卡尔酵母、酿酒酵母、糖化酵母、道格拉氏酵母、克鲁弗酵母、诺地酵母、卵形酵母、或解脂西洋薯霉(Yarrowia lipolytica)细胞。

[0225] 真菌宿主细胞可为丝状真菌细胞。“丝状真菌”包括真菌门(Eumycota)和卵菌门的亚门(如由Hawksworth等, 1995, 见上文, 所定义)的所有丝状形式。丝状真菌通常的特征在于由壳多糖(chitin)、纤维素、葡聚糖、壳聚糖(chitosan)、甘露聚糖和其它复杂多糖组成的菌丝体壁。通过菌丝延伸进行营养生长, 而碳分解代谢是专性需氧的。相反, 酵母例如酿酒酵母的营养生长通过单细胞菌体的出芽生殖(budding)进行, 而碳分解代谢可以是发酵的。

[0226] 丝状真菌宿主细胞可为枝顶孢霉属、曲霉属、短梗霉属、烟管霉属(Bjerkandera)、拟蜡菌属、金孢子菌属、鬼伞属(Coprinus)、革盖菌属(Coriolus)、隐球菌属、Filibasidium、镰孢属、腐质霉属、梨孢菌属、毛霉属、毁丝霉属、新考玛脂霉属、脉孢菌属、拟青霉属、青霉属、平革菌属(Phanerochaete)、射脉菌属(Phlebia)、瘤胃壶菌属、侧耳属(Pleurotus)、裂褶菌属、踝节菌属、嗜热子囊菌属、梭孢壳属、弯颈霉属、栓菌属(Trametes)或木霉属细胞。

[0227] 例如, 丝状真菌宿主细胞可为泡盛曲霉、烟曲霉、臭曲霉、日本曲霉、构巢曲霉、黑曲霉、米曲霉、黑刺烟管菌(Bjerkandera adusta)、干拟蜡菌(Ceriporiopsis aneirina)、Ceriporiopsis caregiea、Ceriporiopsis gilvescens、Ceriporiopsis pannocinta、Ceriporiopsis rivulosa、Ceriporiopsis subrufa、虫拟蜡菌(Ceriporiopsis subvermispora)、Chrysosporium inops、嗜角质金孢子菌、Chrysosporium lucknowense、Chrysosporium merdarium、毡金孢子菌、Chrysosporium queenslandicum、热带金孢子菌、Chrysosporium zonatum、灰盖鬼伞(Coprinus cinereus)、毛革盖菌(Coriolus hirsutus)、杆孢状镰孢、禾谷镰孢、库威镰孢、大刀镰孢、禾本科镰孢、禾赤镰孢、异孢镰孢、合欢木镰孢、尖镰孢、多枝镰孢、粉红镰孢、接骨木镰孢、肤色镰孢、拟分枝孢镰孢、硫色镰孢、圆镰孢、拟丝孢镰孢、镶片镰孢、特异腐质霉、疏棉状腐质霉、米黑毛霉、嗜热毁丝霉、粗糙脉孢菌、产紫青霉、黄孢平革菌(Phanerochaete chrysosporium)、辐射射脉菌(Phlebia radiata)、刺芹侧耳(Pleurotus eryngii)、土生梭孢壳、长绒毛栓菌(Trametes villosa)、变色栓菌(Trametes versicolor)、哈茨木霉、康宁木霉、长枝木霉、里氏木霉或绿色木霉细胞。

[0228] 可将真菌细胞通过涉及原生质体形成、原生质体转化和细胞壁再生的方法以本身公知的方式转化。用于转化曲霉属和木霉属宿主细胞的合适方法在EP238 023, Yelton等, 1984, Proc.Natl.Acad.Sci.USA 81:1470-1474, 和Christensen等, 1988, Bio/Technology 6:1419-1422中描述。用于转化镰孢属菌种的合适方法由Malardier等, 1989, Gene 78:147-

156和WO 96/00787描述。可以使用由如下文献描述的方法转化酵母:Becker和Guarente,于Abelson,J.N.和Simon,M.I.编,Guide to Yeast Genetics and Molecular Biology, Methods in Enzymology,Volume 194,pp 182-187,Academic Press,Inc.,New York;Ito等,1983,J.Bacteriol.153:163;和Hinnen等,1978,Proc.Natl.Acad.Sci.USA 75:1920。

[0229] 产生方法

[0230] 本发明还涉及用于产生本发明多肽的方法,其包括:(a)在有助于产生所述多肽的条件下培养细胞,所述细胞以其野生型形式产生所述多肽;和(b)回收所述多肽。在一个优选的方面,所述细胞是土生梭孢壳的细胞。在一个更优选的方面,所述细胞是棘孢曲霉。在一个最优选的方面,所述细胞是土生梭孢壳NRRL 8126。

[0231] 本发明还涉及用于产生本发明的多肽的方法,其包括:(a)在有助于产生所述多肽的条件下培养本发明的重组宿主细胞;和(b)回收所述多肽。

[0232] 使用本领域熟知的方法在适合于产生所述多肽的营养培养基中培养细胞。例如,可以通过在合适培养基中和允许表达和/或分离所述多肽的条件下进行的摇瓶培养,和实验室或工业发酵罐中的小规模或大规模发酵(包括连续、分批、补料分批或固态发酵)来培养细胞。使用本领域已知的方法在合适的营养培养基中进行培养,所述营养培养基包含碳源和氮源和无机盐。合适的培养基能够从商业供应商获得或可以根据公开的组成制备(例如,在美国典型培养物保藏中心的目录中)。如果多肽分泌到营养培养基中,该多肽能够从所述培养基中直接回收。如果多肽不分泌,则其能够从细胞裂解物(lysate)回收。

[0233] 可以使用本领域已知的对于所述多肽是特异性的方法来检测多肽。这些检测方法可包括特异性抗体的使用、酶产物的形成或酶底物的消失。例如,酶试验(enzyme assay)可用于测定多肽的活性。

[0234] 多肽可以使用本领域已知的方法回收。例如,多肽可以通过常规方法从营养培养基中回收,所述常规方法包括但不限于离心、过滤、提取、喷雾干燥、蒸发或沉淀。

[0235] 多肽可以通过多种本领域已知的方法纯化以获得基本上纯的多肽,所述方法包括但不限于层析(例如,离子交换、亲和、疏水、层析聚焦和大小排阻)、电泳方法(例如,制备型(preparative)等电聚焦)、差示溶解度(例如,硫酸铵沉淀)、SDS-PAGE或提取(参见,例如,Protein Purification,J.-C.Janson和Lars Ryden,editors,VCH Publishers,New York,1989)。

[0236] 在可选的方面,不回收多肽,而将表达所述多肽的本发明宿主细胞作为多肽的来源。

[0237] 植物

[0238] 本发明还涉及分离的植物,例如,转基因植物、植物部分或植物细胞,其包含本发明的分离的多核苷酸,从而以可回收的量表达和产生所述多肽。多肽可从植物或植物部分回收。或者,同样可以将含有所述多肽的植物或植物部分用于改进食品或饲料的质量,例如,改进营养价值、适口性(palatability)和流变性质(rheological properties),或用于破坏抗营养因子。

[0239] 转基因植物可以是双子叶的(双子叶植物)或单子叶的(单子叶植物)。单子叶植物的实例是草(grasses),如草地早熟禾(meadow grass)(蓝草(blue grass),早熟禾属(Poa));饲用牧草(forage grass)如羊茅属(Festuca)、黑麦草属(Lolium);寒地型牧草

(temperate grass),如Agrostis(翦股颖属);和谷类,例如,小麦、燕麦、黑麦、大麦、稻(rice)、高粱和玉蜀黍(maize)(玉米)。

[0240] 双子叶植物的实例是烟草(tobacco),豆类(legumes),如羽扇豆(lupins),马铃薯,糖甜菜(sugar beet),豌豆,豆(bean)和大豆(soybean)和十字花科的(cruciferous)植物(十字花科(family Brassicaceae)),如花椰菜(cauliflower),油菜籽(rape seed)和紧密相关的模型生物体拟南芥(*Arabidopsis thaliana*)。

[0241] 植物部分的实例是茎(stem)、愈伤组织(callus)、叶(leaf)、根(root)、果实(fruit)、种子(seed)和块茎(tuber),以及包含这些部分的独立组织,例如,表皮(epidermis)、叶肉(mesophyll)、薄壁组织(parenchyme)、维管组织(vascular tissue)、分生组织(meristem)。具体的植物细胞区室(compartments),如叶绿体(chloroplast)、质外体(apoplast)、线粒体(mitochondria)、液泡(vacuole)、过氧化物酶体(peroxisome)和细胞质(cytoplasm)也被认为是植物部分。此外,任何植物细胞,无论什么组织来源,都被认为是植物部分。同样地,植物部分,如分离以促进本发明的应用的具体组织和细胞也被认为是植物部分,例如胚(embryo)、胚乳(endosperm)、糊粉(aleurone)和种皮(seed coat)。

[0242] 同样包含于本发明范围内的还有这些植物、植物部分和植物细胞的子代。

[0243] 表达多肽的转基因植物或植物细胞可以依照本领域已知方法构建。简而言之,通过如下方法构建所述植物或植物细胞:将编码多肽的一个或多个(几个)表达构建体并入植物宿主基因组或叶绿体基因组,并且将所得的修饰植物或植物细胞繁殖为转基因植物或植物细胞。

[0244] 表达构建体便利地是包含编码多肽的多核苷酸的核酸构建体,所述多核苷酸与在选择的植物或植物部分中表达该多核苷酸所需的适当的调节序列可操作地连接。此外,表达构建体可以包含对于鉴定宿主细胞有用的选择性标记,在所述宿主细胞中整合了表达构建体和将该构建体引入到所述植物中所必需的DNA序列(后者依赖于使用的DNA引入方法)。

[0245] 调节序列的选择,例如启动子和终止子序列和任选地信号或转运序列的选择,举例来说,基于期望何时、何处以及如何表达多肽而确定。例如,编码多肽的基因的表达可以是组成型的或诱导型的,或可以是发育、阶段或组织特异性的,并且基因产物可以靶向特定的组织或植物部分例如种子或叶。调节序列由例如Tague等,1988,Plant Physiology 86:506所述。

[0246] 对于组成性表达,可以使用35S-CaMV、玉米泛素1和稻肌动蛋白1启动子(Franck等,1980,Cell 21:285-294,Christensen等,1992,Plant Mo.Biol.18:675-689;Zhang等,1991,Plant Cell 3:1155-1165)。器官特异性启动子可以是例如来自贮藏库组织(storage sink tissue)例如种子、马铃薯块茎和果实的启动子(Edwards和Coruzzi,1990,Ann.Rev.Genet.24:275-303),或来自代谢库组织(metabolic sink tissue)例如分生组织的启动子(Ito等,1994,Plant Mol.Biol.24:863-878),种子特异性启动子诸如来自稻的谷蛋白(glutelin)、醇溶蛋白(prolamin)、球蛋白(globulin)或白蛋白(albumin)启动子(Wu等,1998,Plant Cell Physiol.39:885-889),来自豆球蛋白(legumin)B4和蚕豆(*Vicia faba*)的未知的种子蛋白基因的蚕豆启动子(Conrad等,1998,J.of Plant Physiol.152:708-711)、来自种子油体蛋白(oil body protein)的启动子(Chen等,1998,Plant Cell Physiol.39:935-941),来自欧洲油菜(*Brassica napus*)的贮藏蛋白napA启动子,或本技术

领域公知的任何其他种子特异性的启动子,例如,在WO 91/14772中所描述的。此外,启动子可为叶特异性的启动子,如来自稻或番茄的rbcs启动子(Kyozuka等,1993,Plant Physiology 102:991-1000),小球藻病毒(chlorella virus)腺嘌呤甲基转移酶(adenine methyltransferase)基因启动子(Mitra和Higgins,1994,Plant Mol.Biol.26:85-93),来自稻的aldP基因启动子(Kagaya等,1995,Mol.Gen.genet.248:668-674),或伤口诱导的启动子,如马铃薯pin2启动子(Xu等,1993,Plant Mol.Biol.22:573-588)。同样地,所述启动子可通过非生物的处理诱导,所述非生物的处理诸如温度、干旱或盐度变化,或通过外源施加的激活所述启动子的物质诱导,例如乙醇、雌激素(oestrogens)、植物激素(plant hormones)如乙烯、脱落酸(abscisic acid)和赤霉素(gibberellic acid),和重金属。

[0247] 启动子增强子元件也可以用于实现多肽在植物中的较高表达。例如,启动子增强子元件可以是内含子,其置于启动子和编码多肽的多核苷酸之间。例如Xu等,1993,见上,公开了使用稻肌动蛋白1基因的第一内含子以增强表达。

[0248] 选择性标记基因和表达构建体的任何其它部分可以选自本领域内可用的那些。

[0249] 将核酸构建体根据本领域已知的常规技术并入植物基因组,所述常规技术包括土壤杆菌属(Agrobacterium)介导的转化、病毒介导的转化、显微注射(microinjection)、粒子轰击、生物射弹转化和电穿孔(Gasser等,1990,Science 244:1293;Potrykus,1990,Bio/Technology 8:535;Shimamoto等,1989,Nature 338:274)。

[0250] 目前,根癌土壤杆菌(Agrobacterium tumefaciens)介导的基因转移(gene transfer),是产生转基因双子叶植物的优选方法(为了参考,见Hooykas和Schilperoort,1992,Plant Mol.Biol.19:15-38),而且它也可以用于转化单子叶植物,虽然对于这些植物其他的转化方法是常用的。目前,产生转基因单子叶植物的优选的方法,是用粒子(用转化DNA涂覆的微观的金或钨粒子)轰击胚愈伤组织(embryonic calli)或发育中的胚(developing embryos)(Christou,1992,Plant J.2:275-281;Shimamoto,1994,Current Opin.Biotech.5:158-162;Vasil等,1992,Bio/Technology 10:667-674)。转化单子叶植物的可供选择的方法是基于原生质体转化,如由Omirulleh等,1993,Plant Mol.Biol.21:415-428所描述的。其他用于依照本公开使用的转化方法包括那些描述于美国专利6,395,966和7,151,204的那些(两者均通过全文提述并入本文)。

[0251] 转化之后,根据本领域熟知的方法选择具有并入的表达构建体的转化体并且再生成为完整植物。通常设计转化方法用于通过如下方法在再生期间或在后续世代中选择性消除选择基因:例如,使用带有两个独立的T-DNA构建体的共转化或通过特异性重组酶位点特异性地切除选择基因。

[0252] 除了用根据本发明制备的构建体直接转化具体植物基因型之外,还可通过将具有所述构建体的植物与缺乏该构建体的第二植物杂交来制备转基因植物。举例而言,可将编码多肽的构建体通过杂交而引入特定植物品种,而根本无需直接转化该给定品种的植物。因此,本发明不仅涵盖从依照本发明经转化的细胞直接再生的植物,还包括此类植物的后裔(progeny)。如用于本文,后裔可指依照本发明制备的亲本植物任何世代的后代(offspring)。此种后裔可包含依据本发明制备的DNA构建体,或依据本发明制备的DNA构建体的一部分。杂交导致转基因通过将起始种系与供体植物种系交叉授粉而引入植物种系。此类步骤的非限制性实例进一步阐述于美国专利7,151,204号。

[0253] 植物可通过回交转化方法生成。举例而言,植物包括称作回交转化的基因型、种系、近交体(inbred)或杂合体(hybrid)的植物。

[0254] 可使用遗传标记以协助本发明的一种或多种转基因从一个遗传背景基因渗入(introgression)至另一个。标记协助的选择提供了相对于常规育种的优点,在于其可用于避免由表型变异导致的错误。进一步,遗传标记可在特定杂交的个体后裔中提供有关良种种质相对程度的数据。举例而言,当具有所需性状但除此之外(otherwise)具有非农艺学所需的遗传背景的植物与良种亲本杂交时,可使用遗传标记来选择不仅具有该目标性状,还具有相对较大比例所需种质的后裔。以此方式,使一种或多种性状基因渗入特定遗传背景所需的世代数得到最小化。

[0255] 本发明还涉及产生本发明多肽的方法,包括:(a)在有助于产生所述多肽的条件下培养包含编码所述多肽的多核苷酸的转基因植物或植物细胞;和(b)回收所述多肽。

[0256] 去除或减少纤维素分解增强活性

[0257] 本发明还涉及用于产生亲本细胞突变体的方法,其包括破坏或缺失编码本发明的多肽的多核苷酸或其部分,所述方法导致在相同条件下培养时,与亲本细胞相比突变的细胞产生较少的所述多肽。

[0258] 可以使用本领域熟知的方法(例如,插入、破坏、替代或缺失)通过减少或消除多核苷酸的表达来构建突变体细胞。在一个优选的方面,所述多核苷酸是失活的。待修饰或失活的多核苷酸可以是,例如,编码区或其对活性关键的部分,或表达编码区所需的调节元件。这种调节或调控序列的实例可以是启动子序列或其功能部分,即,足以影响多核苷酸表达的部分。用于可能的修饰的其它调控序列包括但不限于前导序列、聚腺苷酸化序列、前肽序列、信号肽序列、转录终止子和转录激活子。

[0259] 可以通过向亲本细胞施以诱变,并且选择其中已将多核苷酸的表达减少或消除的突变体细胞来进行多核苷酸的修饰或失活。诱变可能是特异性的或随机的,可以通过例如使用合适的物理或化学诱变剂进行,通过使用合适的寡核苷酸进行,或通过所述DNA序列进行PCR产生的诱变。此外,可以通过使用这些诱变剂的任何组合来进行诱变。

[0260] 适合于本发明目的的物理或化学诱变剂的实例包括紫外线(UV)照射、羟胺、N-甲基-N'-硝基-N-亚硝基胍(MNNG)、O-甲基羟胺、亚硝酸、乙基甲烷磺酸酯(ethyl methane sulphonate)(EMS)、亚硫酸氢钠、甲酸和核苷酸类似物。

[0261] 当使用这些试剂时,通常通过如下方法来进行所述诱变:在合适条件下存在优选的诱变剂时温育待诱变的亲本细胞,并筛选和/或选择显示基因表达减少的或无基因表达的突变体细胞。

[0262] 通过导入、取代或去除基因中的一个或多个(几个)核苷酸或其转录或翻译所需的调节元件可以实现所述多核苷酸的修饰或失活。例如,可以插入或删除核苷酸从而导致引入终止密码子,去除起始密码子,或改变开放阅读框。按照本领域已知的方法通过定位诱变或PCR产生的诱变可以实现这种修饰或失活。尽管在理论上所述修饰可以在体内进行,即,直接在表达待修饰的多核苷酸的细胞上进行,但优选如下面所示例的那样在体外进行所述修饰。

[0263] 消除或减少多核苷酸表达的便利方式的例子是基于基因取代,基因缺失,或基因破坏的技术。例如,在基因破坏方法中,将相应于内源多核苷酸的核酸序列在体外进行诱变

以产生缺陷性的核酸序列,然后将其转化入亲本细胞中以产生缺陷基因。通过同源重组,所述缺陷性核酸序列替代了内源性多核苷酸。可能理想的是所述缺陷性多核苷酸还编码标记,其可用于选择其中多核苷酸被修饰或破坏的转化体。在特别优选的方面,用选择性标记(如本文所述的那些)来破坏所述多核苷酸。

[0264] 本发明还涉及在细胞中抑制具有纤维素分解增强活性的多肽的表达的方法,包括施于细胞或在细胞中表达双链RNA(dsRNA)分子,其中所述dsRNA包含本发明的多核苷酸的亚序列。在一个优选的方面,所述dsRNA长度为约15、16、17、18、19、20、21、22、23、24、25或更多个双链核苷酸。

[0265] 所述dsRNA优选为小干扰RNA(siRNA)或微小RNA(miRNA)。在一个优选的方面,所述dsRNA是供抑制转录的小干扰RNA(siRNA)。在另一个优选的方面,所述dsRNA是供抑制翻译的微小RNA(miRNA)。

[0266] 本发明还涉及这样的双链RNA(dsRNA)分子,其包含SEQ ID NO:1,SEQ ID NO:3,SEQ ID NO:5,SEQ ID NO:7,SEQ ID NO:9,SEQ ID NO:11,SEQ ID NO:13,SEQ ID NO:15,SEQ ID NO:17,SEQ ID NO:19,或SEQ ID NO:21的成熟多肽编码序列的一部分,以供在细胞中抑制所述多肽的表达。尽管本发明并不限于任何具体的作用机制,但dsRNA可进入细胞并导致类似或相同序列的单链RNA(ssRNA)包括内源mRNA的降解。当细胞暴露于dsRNA时,来自同源基因的mRNA通过称作RNA干扰(RNAi)的过程受选择性降解。

[0267] 本发明的dsRNA可用于基因沉默。在一个方面,本发明提供了使用本发明的dsRNAi选择性降解RNA的方法。该过程可在体外、离体或体内实施。在一个方面,所述dsRNA分子可用于在细胞、器官或动物中生成功能缺失(loss-of-function)的突变。用于制备和使用dsRNA分子以选择性降解RNA的方法在本领域是公知的;参见,例如美国专利6489127、6506559、6511824和6515109号。

[0268] 本发明进一步涉及亲本细胞的突变体细胞,其包含编码多肽的多核苷酸或其调控序列的破坏或缺失,或沉默的编码该多肽的基因,这导致与亲本细胞相比突变体细胞产生更少的多肽或不产生多肽。

[0269] 多肽缺陷型突变体细胞作为表达天然和异源多肽的宿主细胞特别有用。所以,本发明进一步涉及生产天然或异源多肽的方法,其包括:(a)在有助于产生所述多肽的条件下培养突变体细胞;和(b)回收所述多肽。术语“异源多肽”意指对宿主细胞不是天然的多肽,例如天然蛋白质的变体。宿主细胞可包含多于一个拷贝的编码天然或异源多肽的多核苷酸。

[0270] 用于培养和纯化感兴趣的产物的方法可以通过本领域已知的方法进行。

[0271] 本发明用于产生基本上无纤维素分解增强活性的产物的方法在真核多肽,特别是真菌蛋白质例如酶的产生中是特别令人有兴趣的。纤维素分解增强-缺陷细胞也可以用于表达在制药上感兴趣的异源蛋白质例如激素、生长因子、受体等。术语“真核多肽”不仅包括天然多肽,也包括多肽例如酶,其通过氨基酸取代、缺失或添加或其它这样的修饰而被修饰以增强活性、热稳定性、pH耐受性等。

[0272] 在另外的方面,本发明涉及通过本发明方法产生的基本上无纤维素分解增强活性的蛋白质产物。

[0273] 组合物

[0274] 本发明还涉及包含本发明的多肽的组合物。优选地,所述组合物富含这种多肽。术语“富含”表示所述组合物的纤维素分解增强活性,例如,以至少1.1的富集因数(enrichment factor)增加。

[0275] 所述组合物可以包含本发明的多肽作为主要酶成分,例如,单成分组合物。或者,所述组合物可以包含多种酶活性,如一种或多种(几种)选自下组的酶:纤维素酶、半纤维素酶、棒曲霉素、酯酶、漆酶、木质素分解酶、果胶酶、过氧化物酶、蛋白酶和膨胀素。

[0276] 可以依照本领域内已知的方法制备多肽组合物,并且可以是液体或干组合物的形式。例如,所述多肽组合物可以是颗粒(granulate)或微粒(microgranulate)的形式。可以依照本领域内已知方法使纳入所述组合物中的多肽稳定化。

[0277] 下面给出本发明的多肽组合物的优选用途。本发明的多肽组合物的剂量和其他使用所述组合物的条件可基于本领域已知方法来确定。

[0278] 用途

[0279] 本发明还涉及下述使用具有纤维素分解增强活性的多肽或其组合物的方法。

[0280] 本发明还涉及降解或转化纤维素材料的方法,包括:在本发明的具有纤维素分解增强活性的多肽的存在下用酶组合物处理纤维素材料。在一个方面,所述方法还包括回收经降解或转化的纤维素材料。纤维素材料的降解或转化的可溶性产物可从不溶性纤维素材料使用本领域中公知的技术如例如离心、过滤和重力沉降来分离。

[0281] 本发明还涉及产生发酵产物的方法,其包括:(a)在存在本发明的具有纤维素分解增强活性的多肽的条件下,用酶组合物糖化纤维素材料;(b)用一种或多种(几种)发酵微生物发酵经糖化的纤维素材料以产生发酵产物;和(c)从发酵回收发酵产物。

[0282] 本发明还涉及发酵纤维素材料的方法,其包括:用一种或多种(几种)发酵微生物发酵纤维素材料,其中在存在本发明的具有纤维素分解增强活性的多肽的条件下,用酶组合物糖化纤维素材料。在一个方面,纤维素材料的发酵产生发酵产物。在另一个方面,所述方法进一步包括从发酵回收发酵产物。

[0283] 本发明的方法可用于将纤维素材料糖化为可发酵的糖,并将可发酵的糖转化为许多有用的物质,例如燃料,饮用乙醇,和/或发酵产物(例如酸、醇、酮、气体等)。从纤维素材料产生所需的发酵产物通常涉及预处理,酶水解(糖化),和发酵。

[0284] 根据本发明的纤维素材料的处理可以使用本领域的常规过程完成。此外,本发明的方法能使用经配置以依照发明操作的任何常规生物质处理设备进行处理。

[0285] 水解(糖化)和发酵,分别或同时,包括但不限于,分开的水解和发酵(SHF)、同步糖化和发酵(SSF)、同步糖化和共发酵(SSCF)、混合的水解和发酵(HHF)、分开的水解和共发酵(SHCF)、混合的水解和共发酵(HHCF),和直接微生物转化(DMC)。SHF使用分开的处理步骤以首先将纤维素材料酶水解为可发酵糖,例如,葡萄糖,纤维二糖、纤维三糖和戊糖,然后将可发酵糖发酵成为乙醇。在SSF中,纤维素材料的酶水解和糖变为乙醇的发酵组合在一个步骤中(Philippidis,G.P.,1996,Cellulose bioconversion technology,于Handbook on Bioethanol:Production and Utilization,Wyman,C.E编,Taylor&Francis,Washington,DC,179-212)。SSCF包括多种糖的共发酵(Sheehan,J.,和Himmel,M.,1999,Enzymes,energy and the environment:A strategic perspective on the U.S.Department of Energy's research and development activities for bioethanol,Biotechnol.Prog.15:817-

827)。HHF在同步糖化和水解步骤之外,还涉及单独的水解步骤,所述步骤可以在同一个反应器中进行。HHF过程中的步骤可以在不同的温度进行,即,高温酶糖化,然后在发酵菌株能够耐受的较低温度进行SSF。DMC在一个或多个(几个)步骤中组合了所有三个过程(酶产生、水解和发酵),其中使用相同的生物体产生用于将纤维素材料转化成可发酵糖并将可发酵糖转化成终产物的酶(Lynd,L.R.,Weimer,P.J.,van Zyl,W.H.,和Pretorius,I.S.,2002, *Microbial cellulose utilization:Fundamentals and biotechnology*, *Microbiol.Mol.Biol.Reviews* 66:506-577)。在本文可以理解的是,本领域中任何已知的方法,包括预处理、酶水解(糖化)、发酵,或它们的组合,可用于实施本发明的方法。

[0286] 常规设备包括补料分批搅拌反应器、分批搅拌反应器、具有超滤的连续流搅拌反应器和/或连续活塞流柱式反应器(Fernanda de Castilhos Corazza,Flávio Faria de Moraes,Gisella Maria Zanin和Ivo Neitzel,2003,Optimal control in fed-batch reactor for the cellobiose hydrolysis,*Acta Scientiarum.Technology*25:33-38;Gusakov,A.V.,和SinitSyn,A.P.,1985,Kinetics of the enzymatic hydrolysis of cellulose:1.A mathematical model for a batch reactor process, *Enz.Microb.Technol.*7:346-352)、研磨反应器(Ryu,S.K.,和Lee,J.M.,1983, Bioconversion of waste cellulose by using an attrition bioreactor, *Biotechnol.Bioeng.*25:53-65),或者具有由电磁场引起的强烈搅拌的反应器(Gusakov,A.V.,SinitSyn,A.P.,Davydkin,I.Y.,Davydkin,V.Y.,Protas,O.V.,1996,Enhancement of enzymatic cellulose hydrolysis using a novel type of bioreactor with intensive stirring induced by electromagnetic field, *Appl.Biochem.Biotechnol.*56:141-153)。其它反应器类型包括:流化床、升流层(upflow blanket)、固定化和挤出机型反应器以供水解和/或发酵。

[0287] 预处理。在本发明的方法的实施中,可以使用本领域已知的任何预处理过程破坏植物细胞壁的纤维素材料组分(Chandra等,2007,Substrate pretreatment:The key to effective enzymatic hydrolysis of lignocellulosics?*Adv.Biochem.Engin./Biotechnol.*108:67-93;Galbe和Zacchi,2007,Pretreatment of lignocellulosic materials for efficient bioethanol production,*Adv.Biochem.Engin./Biotechnol.*108:41-65;Hendriks和Zeeman,2009,Pretreatments to enhance the digestibility of lignocellulosic biomass,*Bioresource Technol.*100:10-18;Mosier等,2005,Features of promising technologies for pretreatment of lignocellulosic biomass,*Bioresource Technol.*96:673-686;Taherzadeh和Karimi,2008,Pretreatment of lignocellulosic wastes to improve ethanol and biogas production:A review,*Int.J.of Mol.Sci.*9:1621-1651;Yang和Wyman,2008, Pretreatment:the key to unlocking low-cost cellulosic ethanol,*Biofuels Bioproducts and Biorefining-Biofpr.*2:26-40)。

[0288] 纤维素材料也可以在预处理之前使用本领域中已知的方法进行粒度减小、预浸泡、润湿、洗涤和/或调节。

[0289] 常规的预处理包括但不限于,蒸汽预处理(伴随或不伴随爆炸)、稀酸预处理、热水预处理、碱性预处理、石灰预处理、湿氧化、湿爆炸、氨纤维爆炸、有机溶剂预处理和生物预

处理。其它预处理包括氨渗滤、超声、电穿孔、微波、超临界CO₂、超临界H₂O、臭氧和γ辐射预处理。

[0290] 可以在水解和/或发酵之前预处理纤维素材料。预处理优选在水解前进行。或者，预处理可以与酶水解同时进行以释放可发酵糖，如葡萄糖、木糖和/或纤维二糖。在大多数情况下，预处理步骤本身使一些生物质转化成可发酵糖(甚至在不存在酶的情况下)。

[0291] 蒸汽预处理:在蒸汽预处理中,加热纤维素材料以破坏植物细胞壁成分,包括木质素、半纤维素和纤维素,使酶可接触纤维素和其它部分,例如,半纤维素。使纤维素材料通过或穿过反应容器,其中注入蒸汽以增加温度至需要的温度和压力,并且在其中保持期望的反应时间。蒸汽预处理优选在140-230℃,更优选160-200℃,并且最优选170-190℃进行,其中最优的温度范围依赖于加入的任何化学催化剂。蒸汽预处理的停留时间优选1-15分钟,更优选3-12分钟,并且最优选4-10分钟,其中最优的停留时间依赖于温度范围和加入的任何化学催化剂。蒸汽预处理允许相对较高的固体加载量,使纤维素材料在预处理过程中通常仅仅变得潮湿。蒸汽预处理经常与预处理后的物质的爆炸放料(explosive discharge)组合,这称为蒸汽爆炸,即,快速闪变至大气压和物质的湍流,以通过破碎增加可接触的面积(Duff和Murray,1996,Bioresource Technology 85:1-33;Galbe和Zacchi,2002,Appl.Microbiol.Biotechnol.59:618-628;美国专利申请No.20020164730)。在蒸汽预处理过程中,切割半纤维素乙酰基团,并且得到的酸自催化半纤维素部分水解成单糖和寡糖。去除木质素至有限的程度。

[0292] 经常在蒸汽预处理之前加入催化剂如H₂SO₄或SO₂(通常0.3至3%w/w),其可减少时间,降低温度,增加回收率,并改进酶水解(Balasteros等,2006,Appl.Biochem.Biotechnol.129-132:496-508;Varga等,2004,Appl.Biochem.Biotechnol.113-116:509-523;Sassner等.,2006,Enzyme Microb.Technol.39:756-762)。

[0293] 化学预处理:术语“化学处理”指促进纤维素、半纤维素和/或木质素分离和/或释放的任何化学预处理。合适的化学预处理过程的实例包括例如稀酸预处理、石灰预处理、湿氧化、氨纤维/冷冻爆炸(AFEX)、氨渗滤(APR)和有机溶剂预处理。

[0294] 在稀酸预处理中,将纤维素材料与稀酸(通常是H₂SO₄)和水混合以形成浆料,由蒸汽加热至期望的温度,并在一段停留时间后闪变至大气压。可以用很多反应器设计进行稀酸预处理,例如,活塞流式反应器、逆流反应器或连续逆流收缩床反应器(Duff和Murray,1996, supra;Schell等,2004,Bioresource Technol.91:179-188;Lee等,1999,Adv.Biochem.Eng.Biotechnol.65:93-115)。

[0295] 还可以使用碱性条件下的几种预处理方法。这些碱预处理包括,但不限于,石灰预处理、湿氧化、氨渗滤(APR)和氨纤维/冷冻爆炸(AFEX)。

[0296] 用碳酸钙、氢氧化钠或氨,在85-150℃的低温进行石灰预处理,停留时间从1小时到几天(Wyman等,2005,Bioresource Technol.96:1959-1966;Mosier等,2005,Bioresource Technol.96:673-686)。WO 2006/110891、WO 2006/110899、WO 2006/110900和WO 2006/110901公开了使用氨的预处理方法。

[0297] 湿法氧化是热预处理,通常在180-200℃进行5-15分钟,加入氧化剂如过氧化氢或过压氧(Schmidt和Thomsen,1998,Bioresource Technol.64:139-151;Palonen等,2004,

Appl. Biochem. Biotechnol. 117:1-17; Varga等, 2004, Biotechnol. Bioeng. 88:567-574; Martin等, 2006, J. Chem. Technol. Biotechnol. 81:1669-1677)。预处理以优选1-40%干物质, 更优选2-30%干物质, 并且最优性5-20%干物质进行, 并且由于加入碱如碳酸钠, 初始pH经常会增加。

[0298] 湿法氧化预处理方法的修改方法, 称为湿爆炸(湿氧化和蒸汽爆炸的组合), 能够处理高达30%的干物质。在湿爆炸中, 在预处理过程中, 在一定的停留时间后引入氧化剂。然后通过闪变至大气压而结束预处理(WO 2006/032282)。

[0299] 氨纤维爆炸(AFEX)涉及在温和温度如90-100°C和高压如17-20bar, 用液体或气体氨将纤维素材料处理5-10分钟, 其中干物质含量可以高达60%(Gollapalli等, 2002, Appl. Biochem. Biotechnol. 98:23-35; Chundawat等, 2007, Biotechnol. Bioeng. 96:219-231; Alizadeh等, 2005, Appl. Biochem. Biotechnol. 121:1133-1141; Teymouri等, 2005, Bioresource Technol. 96:2014-2018)。AFEX预处理导致纤维素解聚, 和半纤维素的部分水解。木质素-糖复合物得到切割。

[0300] 有机溶剂预处理通过用含水乙醇(40-60%乙醇)在160-200°C提取30-60分钟而将纤维素材料去木质素化(Pan等, 2005, Biotechnol. Bioeng. 90:473-481; Pan等, 2006, Biotechnol. Bioeng. 94:851-861; Kurabi等, 2005, Appl. Biochem. Biotechnol. 121:219-230)。经常加入硫酸作为催化剂。在有机溶剂预处理中, 去除大部分半纤维素。

[0301] 合适的预处理方法的其他实例如Schell等, 2003, Appl. Biochem and Biotechnol. Vol. 105-108:69-85, 和Mosier等, 2005, Bioresource Technology 96:673-686, 和美国已公开的申请2002/0164730所述。

[0302] 在一个方面, 化学预处理优选作为酸处理, 并且更优选作为连续稀酸和/或弱酸(mild acid)处理进行。酸通常是硫酸, 但也可以使用其它酸, 如乙酸、柠檬酸、硝酸、磷酸、酒石酸、琥珀酸、氯化氢或其混合物。弱酸处理在优选1-5, 更优选1-4, 并且最优选1-3的pH范围内进行。在一个方面, 酸浓度在优选0.01至20wt%酸, 更优选0.05至10wt%酸, 甚至更优选0.1至5wt%酸, 并且最优选0.2至2.0wt%酸的范围内。酸与纤维素材料接触, 并在优选160-220°C, 和更优选165-195°C范围内的温度保持数秒到数分钟, 例如1秒-60分钟的时间。

[0303] 在另一个方面, 预处理作为氨纤维爆炸步骤(AFEX预处理步骤)进行。

[0304] 在另一个方面, 预处理发生在含水浆料中。在优选的方面, 在预处理过程中纤维素材料以优选10-80wt%, 更优选20-70wt%, 并且最优性30-60wt%, 如约50wt%的量存在。预处理的纤维素材料可以不洗涤或者使用本领域任何已知的方法洗涤, 例如, 用水洗涤。

[0305] 机械预处理: 术语“机械预处理”指各种类型的磨制(grinding)或粉碎(milling)(例如, 干磨、湿磨或振动球磨)。

[0306] 物理预处理: 术语“物理预处理”指促进纤维素、半纤维素和/或木质素从纤维素材料分离和/或释放的任何预处理。例如, 物理预处理可涉及辐射(例如, 微波辐射)、汽蒸/蒸汽爆炸、水热解(hydrothermolysis)和它们的组合。

[0307] 物理预处理可以涉及高压和/或高温(蒸汽爆炸)。在一个方面, 高压指范围在优选约300至约600psi, 更优选约350至约550psi, 并且最优选约400至约500psi, 如约450psi的压力。在另一个方面, 高温指范围在约100至约300°C, 优选约140至约235°C的温度。在一个优选的方面, 机械预处理在使用如上所定义的高温和高压的分批过程、蒸汽枪水解器系统,

例如来自Sunds Defibrator AB,Sweden的Sunds Hydrolyzer中进行。

[0308] 组合的物理和化学预处理:可以对纤维素材料进行物理和化学预处理。例如,预处理步骤可以涉及稀酸或弱酸处理和高的温度和/或压力处理。根据需要,可以顺序或同时进行物理和化学预处理。还可以包括机械预处理。

[0309] 因此,在一个优选的方面,对纤维素材料进行机械、化学或物理预处理,或者它们的任意组合,以促进纤维素、半纤维素和/或木质素的分离和/或释放。

[0310] 生物预处理:术语“生物预处理”指促进纤维素、半纤维素和/或木质素从纤维素材料分离和/或释放的任何生物预处理。生物预处理技术可以包括应用溶解木质素的微生物(参见,例如,Hsu,T.-A.,1996,Pretreatment of biomass,于Handbook on Bioethanol: Production and Utilization,Wyman,C.E编,Taylor&Francis,Washington,DC,179-212; Ghosh和Singh,1993,Physicochemical and biological treatments for enzymatic/microbial conversion of lignocellulosic biomass,Adv.Appl.Microbiol.39:295-333;McMillan,J.D.,1994,Pretreating lignocellulosic biomass:a review,于Enzymatic Conversion of Biomass for Fuels Production,Himmel,M.E.,Baker,J.O.,和Overend,R.P.,编,ACS Symposium Series 566,American Chemical Society,Washington,DC,第15章;Gong,C.S.,Cao,N.J.,Du,J.,和Tsao,G.T.,1999,Ethanol production from renewable resources,于Advances in Biochemical Engineering/Biotechnology,Scheper,T.,编,Springer-Verlag Berlin Heidelberg,Germany,65:207-241;Olsson和Hahn-Hagerdal,1996,Fermentation of lignocellulosic hydrolysates for ethanol production,Enz.Microb.Tech.18:312-331;和Vallander和Eriksson,1990,Production of ethanol from lignocellulosic materials:State of the art,Adv.Biochem.Eng./Biotechnol.42:63-95)。

[0311] 糖化。在水解(也称作糖化)步骤中,将纤维素材料(例如经预处理的)水解以将纤维素或亦将半纤维素分解成可发酵糖,如葡萄糖、纤维二糖、木糖、木酮糖、阿拉伯糖、甘露糖、半乳糖和/或可溶的寡糖。水解由酶组合物以酶法在本发明的具有纤维素分解增强活性的多肽的存在下进行。组合物的酶可以顺序加入。

[0312] 酶水解优选在易于由本领域技术人员确定的条件下,在合适的含水环境中进行。在一个方面,水解在适于酶的活性,即对于酶最优的条件下进行。水解可以以补料分批或连续的过程进行,其中将纤维素材料逐渐补入,例如,含酶的水解溶液中。

[0313] 糖化通常在搅拌釜反应器或发酵罐中,在受控的pH、温度和混合条件下进行。合适的处理时间、温度和pH条件可以由本领域技术人员容易地确定。例如,糖化可以持续长达200小时,但是通常优选进行约12至约96小时,更优选约16至约72小时,并且最优选约24至约48小时。温度的范围优选约25℃至约70℃,更优选约30℃至约65℃,并且更优选约40℃至约60℃,特别是约50℃。pH的范围优选约3至约8,更优选约3.5至约7,并且最优选约4至约6,特别是约pH 5。干燥固体含量优选约5至约50wt%,更优选约10至约40wt%,并且最优选约20至约30wt%。

[0314] 酶组合物可包含任何可用于降解或转化纤维素材料的蛋白。

[0315] 在一个方面,酶组合物包含或进一步包含一种或多种(几种)选自下组的酶:纤维素酶、具有纤维素分解增强活性的多肽,半纤维素酶,棒曲霉素(expansin),酯酶,漆酶,木

质素分解酶(ligninolytic enzyme),果胶酶,过氧化物酶,蛋白酶和膨胀素。在另一个方面,所述纤维素酶优选为一种或多种(几种)选自下组的酶:内切葡聚糖酶,纤维二糖水解酶和 β -葡糖苷酶。在另一个方面,所述半纤维素酶优选为一种或多种(几种)选自下组的酶:乙酰甘露聚糖酯酶、乙酰木聚糖酯酶、阿拉伯聚糖酶、阿拉伯呋喃糖苷酶、香豆酸酯酶、阿魏酸酯酶、半乳糖苷酶、葡糖醛酸糖苷酶、葡糖醛酸酯酶、甘露聚糖酶、甘露糖苷酶、木聚糖酶和木糖苷酶。

[0316] 在另一个方面,酶组合物包含一种或多种(几种)纤维素分解酶。在另一个方面,酶组合物包含或进一步包含一种或多种(几种)半纤维素分解酶。在另一个方面,酶组合物包含一种或多种(几种)纤维素分解酶和一种或多种(几种)半纤维素分解酶。在另一个方面,酶组合物包含一种或多种(几种)选自下组的酶:纤维素分解酶和半纤维素分解酶。在另一个方面,酶组合物包含内切葡聚糖酶。在另一个方面,酶组合物包含纤维二糖水解酶。在另一个方面,酶组合物包含 β -葡糖苷酶。在另一个方面,酶组合物包含具有纤维素分解增强活性的多肽。在另一个方面,酶组合物包含内切葡聚糖酶和纤维二糖水解酶。在另一个方面,酶组合物包含内切葡聚糖酶和 β -葡糖苷酶。在另一个方面,酶组合物包含纤维二糖水解酶和 β -葡糖苷酶。在另一个方面,酶组合物包含内切葡聚糖酶,纤维二糖水解酶,和 β -葡糖苷酶。在另一个方面,酶组合物包含内切葡聚糖酶,纤维二糖水解酶,和 β -葡糖苷酶,和具有纤维素分解增强活性的多肽。在另一个方面,酶组合物包含乙酰甘露聚糖酯酶。在另一个方面,酶组合物包含乙酰木聚糖酯酶。在另一个方面酶组合物包含阿拉伯聚糖酶(例如 α -L-阿拉伯聚糖酶)。在另一个方面,酶组合物包含阿拉伯呋喃糖苷酶(例如 α -L-阿拉伯呋喃糖苷酶)。在另一个方面,酶组合物包含香豆酸酯酶。在另一个方面,酶组合物包含阿魏酸酯酶。在另一个方面,酶组合物包含半乳糖苷酶(例如, α -半乳糖苷酶和/或 β -半乳糖苷酶)。在另一个方面,酶组合物包含葡糖醛酸糖苷酶(例如, α -D-葡糖醛酸糖苷酶)。在另一个方面,酶组合物包含葡糖醛酸酯酶。在另一个方面,酶组合物包含甘露聚糖酶。在另一个方面,酶组合物包含甘露糖苷酶(例如 β -甘露糖苷酶)。在另一个方面,酶组合物包含木聚糖酶。在一个优选的方面,所述木聚糖酶为家族10木聚糖酶。在另一个方面,酶组合物包含木糖苷酶(例如 β -木糖苷酶)。在另一个方面,酶组合物包含棒曲霉素。在另一个方面,酶组合物包含酯酶。在另一个方面,酶组合物包含漆酶。在另一个方面,酶组合物包含木质素分解酶。在一个优选的方面,所述木质素分解酶是锰过氧化物酶。在另一个优选的方面,所述木质素分解酶是木质素过氧化物酶。在另一个优选的方面,所述木质素分解酶是 H_2O_2 产生酶。在另一个方面,酶组合物包含果胶酶。在另一个方面,酶组合物包含过氧化物酶。在另一个方面,酶组合物包含蛋白酶。在另一个方面,酶组合物包含膨胀素。

[0317] 在本发明的方法中,酶可在发酵之前或过程中添加,例如在糖化过程中或在发酵微生物繁殖过程中或之后添加。

[0318] 所述酶组合物的一种或多种(几种)组分可为野生型蛋白、重组蛋白或野生型蛋白和重组蛋白的组合。举例而言,一种或多种(几种)组分可为细胞的天然蛋白,其用作宿主细胞以重组表达一种或多种(几种)酶组合物的其他组分。酶组合物的一种或多种(几种)组分可作为单组分产生,然后将其组合以形成酶组合物。所述酶组合物可为多组分和单组分蛋白制备物的组合。

[0319] 用于本发明方法中的酶可为任何适用的形式,如例如去除或未去除细胞的粗发酵

液,含或不含细胞碎片的细胞裂解物,半纯化或纯化的酶制备物,或宿主细胞,作为酶的来源。所述酶组合物可为干粉或颗粒,无粉尘的颗粒,液体,稳定化液体或稳定化受保护的酶。液体酶制备物可根据确立的工艺,例如通过添加稳定剂如糖、糖醇或其他多元醇,和/或乳酸或其他有机酸来稳定化。

[0320] 具有纤维素分解增强活性的酶和多肽的最适量取决于几个因素,其包括但不限于,组分纤维素分解酶的混合物、纤维素材料、纤维素材料的浓度、纤维素材料的预处理、温度、时间、pH和包括发酵生物体(例如,同步糖化和发酵的酵母)。

[0321] 在一个优选的方面,纤维素分解酶对于纤维素材料的有效量是约0.5至约50mg,优选约0.5至约40mg,更优选约0.5至约25mg,更优选约0.75至约20mg,更优选约0.75至约15mg,甚至更优选约0.5至约10mg,并且最优选约2.5至约10mg每g纤维素材料。

[0322] 在另一个优选的方面,具有纤维素分解增强活性的多肽对于纤维素材料的有效量是约0.01至约50.0mg,优选约0.01至约40mg,更优选约0.01至约30mg,更优选约0.01至约20mg,更优选约0.01至约10mg,更优选约0.01至约5mg,更优选约0.025至约1.5mg,更优选约0.05至约1.25mg,更优选约0.075至约1.25mg,更优选约0.1至约1.25mg,甚至更优选约0.15至约1.25mg,并且最优选约0.25至约1.0mg每g纤维素材料。

[0323] 在另一个优选的方面,具有纤维素分解增强活性的多肽对于纤维素分解酶蛋白的有效量是约0.005至约1.0g,优选约0.01至约1.0g,更优选约0.15至约0.75g,更优选约0.15至约0.5g,更优选约0.1至约0.5g,甚至更优选约0.1至约0.25g,并且最优选约0.05至约0.2g每g纤维素分解酶。

[0324] 具有纤维素分解酶活性或半纤维素分解酶活性的多肽以及其他可用于降解纤维素材料的蛋白/多肽,例如具有纤维素分解增强活性的多肽(在下文中称为“具有酶活性的多肽”)可源自或获得自任何合适的来源,包括细菌、真菌、酵母、植物或哺乳动物来源。术语“获得”在本文中意指该酶可从天然产生该酶作为天然酶的生物分离。术语“获得”在本文中还可意指该酶可在宿主生物中使用本文中所述的方法重组产生,其中经重组产生的酶对于宿主生物是天然的或异源的,或具有修饰的氨基酸序列,例如,具有一个或多个(几个)缺失、插入和/或取代的氨基酸,即重组产生的酶,其为天然氨基酸序列的片段和/或突变体或通过本领域已知的氨基酸改组方法产生的酶。天然酶的含义中涵盖的是天然变体,而外来酶的含义中涵盖的是重组(如通过定位诱变或重排)获得的变体。

[0325] 具有酶活性的多肽可以是细菌多肽。例如,所述多肽可以是具有酶活性的革兰氏阳性细菌多肽如芽孢杆菌属(*Bacillus*)、链球菌属(*Streptococcus*)、链霉菌属(*Streptomyces*)、葡萄球菌属(*Staphylococcus*)、肠球菌属(*Enterococcus*)、乳杆菌属(*Lactobacillus*)、乳球菌属(*Lactococcus*)、梭菌属(*Clostridium*)、地芽孢杆菌属(*Geobacillus*)或海洋芽孢杆菌属(*Oceanobacillus*)多肽;或具有酶活性的革兰氏阴性细菌多肽,如大肠杆菌、假单胞菌属(*Pseudomonas*)、沙门氏菌属(*Salmonella*)、弯曲杆菌属(*Campylobacter*)、螺杆菌属(*Helicobacter*)、黄杆菌属(*Flavobacterium*)、梭杆菌属(*Fusobacterium*)、泥杆菌属(*Ilyobacter*)、奈瑟氏菌属(*Neisseria*)或脲原体属(*Ureaplasma*)多肽。

[0326] 在一个优选的方面,所述多肽是具有酶活性的嗜碱芽孢杆菌、解淀粉芽孢杆菌、短芽孢杆菌、环状芽孢杆菌、克劳氏芽孢杆菌、凝结芽孢杆菌、坚强芽孢杆菌、灿烂芽孢杆菌、

迟缓芽孢杆菌、地衣芽孢杆菌、巨大芽孢杆菌、短小芽孢杆菌、嗜热脂肪芽孢杆菌、枯草芽孢杆菌或苏云金芽孢杆菌多肽。

[0327] 在另一个优选的方面,所述多肽是具有酶活性的似马链球菌、酿脓链球菌、乳房链球菌或马链球菌兽瘟亚种多肽。

[0328] 在另一个优选的方面,所述多肽是具有酶活性的不产色链霉菌、除虫链霉菌、天蓝链霉菌、灰色链霉菌或浅青紫链霉菌多肽。

[0329] 具有酶活性的多肽也可以是真菌多肽,并且更优选具有酶活性的酵母多肽如假丝酵母属、克鲁维酵母属、毕赤酵母属、酵母属、裂殖酵母属或西洋蓍霉属多肽;或更优选具有酶活性的丝状真菌多肽如枝顶孢霉属、伞菌属、链格孢属、曲霉属、短梗霉属、Botryosphaeria、拟蜡菌属、Chaetomidium、金孢子菌属、Claviceps、Cochliobolus、鬼伞属、Coptotermes、棒囊壳属、隐丛赤壳菌属、隐球菌属、色二孢属、黑耳属、Filibasidium、镰孢属、赤霉属、全鞭毛虫属、腐质霉属、耙齿菌属、蘑菇属、Leptosphaeria、梨孢菌属、Melanocarpus、多孔菌属、毛霉属、毁丝霉属、新考玛脂霉属、脉孢菌属、拟青霉属、青霉属、平革菌属、瘤胃壶菌属、Poitrasia、假黑盘菌属、Pseudotrichonympha、根毛霉属、裂褶菌属、柱顶孢属、踝节菌属、嗜热子囊菌属、梭孢壳属、弯颈霉属、木霉属、长毛盘菌属、轮枝孢属、包脚菇属或炭角菌属多肽。

[0330] 在一个优选的方面,所述多肽是具有酶活性的卡尔酵母、酿酒酵母、糖化酵母、道格拉氏酵母、克鲁弗酵母、诺地酵母或卵形酵母多肽。

[0331] 在另一个优选的方面,所述多肽是具有酶活性的解纤维枝顶孢霉、棘孢曲霉、泡盛曲霉、烟曲霉、臭曲霉、日本曲霉、构巢曲霉、黑曲霉、米曲霉、嗜角质金孢子菌、Chrysosporium lucknowense、热带金孢子菌、Chrysosporium merdarium、Chrysosporium inops、毡金孢子菌、Chrysosporium queenslandicum、Chrysosporium zonatum、杆孢状镰孢、禾谷镰孢、库威镰孢、大刀镰孢、禾本科镰孢、禾赤镰孢、异孢镰孢、合欢木镰孢、尖镰孢、多枝镰孢、粉红镰孢、接骨木镰孢、肤色镰孢、拟分枝孢镰孢、硫色镰孢、圆镰孢、拟丝孢镰孢、镶片镰孢、灰腐质霉、特异腐质霉、疏棉状腐质霉、白耙齿菌、米黑毛霉、嗜热毁丝霉、粗糙脉孢菌、绳状青霉、产紫青霉、黄孢平革菌、Thielavia achromatica、Thielavia albomyces、Thielavia albopilosa、澳洲梭孢壳、Thielavia fimeti、小孢梭孢壳、卵孢梭孢壳、Thielavia peruviana、瘤孢梭孢壳、毛梭孢壳、Thielavia subthermophila、土生梭孢壳、哈茨木霉、康宁木霉、长枝木霉、里氏木霉、绿色木霉或Trichophaea saccata多肽。

[0332] 还可以使用具有酶活性的多肽经化学修饰或蛋白质工程改造的突变体。

[0333] 所述酶组合物的一种或多种(几种)组分可以是重组组分,亦即,通过克隆编码所述单独组分的DNA序列并随后用该DNA序列转化细胞并在宿主中表达(参见,例如,W091/17243和W091/17244)产生。所述宿主优选是异源宿主(酶对宿主是异源的),但该宿主在一定条件下也可以是同源宿主(酶对宿主是天然的)。单组分纤维素分解蛋白还可以通过从发酵液中提纯这样的蛋白质来制备。

[0334] 在一个方面,所述一种或多种(几种)纤维素分解酶包括商业性纤维素分解酶制备物。适用于本发明的商业的纤维素分解蛋白制备物的实例包括,例如,CELLICTMCtec (Novozymes A/S)、CELLUCLASTTM(Novozymes A/S)、NOVOZYMTM188(Novozymes A/S)、CELLUZYMETM(Novozymes A/S)、CEREFLOTM(Novozymes A/S)和ULTRAFLOTM(Novozymes A/S),

ACCELERASE™(Genencor Int.)、LAMINEX™(Genencor Int.)、SPEZYME™CP(Genencor Int.)、ROHAMENT™7069W(Röhm GmbH)、FIBREZYME®LDI(Dyadic International, Inc.)、FIBREZYME®LBR(Dyadic International, Inc.)或 VISCOSTAR®150L(Dyadic International, Inc.)。所述纤维素酶以固体的约0.001到约5.0wt%，更优选固体的约0.025到约4.0wt%，且最优选固体的约0.005到约2.0wt%的有效量添加。

[0335] 可以用于本发明的方法的细菌内切葡聚糖酶的实例包括但不限于，解纤维热酸菌(*Acidothermus cellulolyticus*)内切葡聚糖酶(WO 91/05039;WO 93/15186;美国专利5,275,944;WO 96/02551;美国专利5,536,655,WO 00/70031,WO 05/093050);*Thermobifida fusca*内切葡聚糖酶III(WO 05/093050);和*Thermobifida fusca*内切葡聚糖酶V(WO 05/093050)。

[0336] 可以用于本发明的真菌内切葡聚糖酶的实例包括但不限于，里氏木霉内切葡聚糖酶I(Penttila等,1986,Gene 45:253-263;里氏木霉Ce17B内切葡聚糖酶I,GENBANK™登录号M15665);里氏木霉内切葡聚糖酶II(Saloheimo等,1988,Gene 63:11-22;里氏木霉Ce15A,GENBANK™登录号M19373);里氏木霉内切葡聚糖酶III(Okada等,1988,Appl.Environ.Microbiol.64:555-563;GENBANK™登录号AB003694);以及里氏木霉内切葡聚糖酶V(Saloheimo等,1994,Molecular Microbiology 13:219-228;GENBANK™登录号Z33381);棘孢曲霉内切葡聚糖酶(Ooi等,1990,Nucleic Acids Research 18:5884);川地曲霉(*Aspergillus kawachii*)内切葡聚糖酶(Sakamoto等,1995,Current Genetics 27:435-439);胡萝卜软腐欧文氏菌(*Erwinia carotovora*)内切葡聚糖酶(Saarilahti等,1990,Gene 90:9-14);尖镰孢内切葡聚糖酶(GENBANK™登录号L29381);灰腐质霉*thermoidea*变种内切葡聚糖酶(GENBANK™登录号AB003107);*Melanocarpus albomyces*内切葡聚糖酶(GENBANK™登录号MAL515703);粗糙脉孢菌内切葡聚糖酶(GENBANK™登录号XM_324477);特异腐质霉内切葡聚糖酶V;嗜热毁丝霉CBS 117.65内切葡聚糖酶;担子菌纲(basidiomycete)CBS 495.95内切葡聚糖酶;担子菌纲CBS 494.95内切葡聚糖酶;土生梭孢壳NRRL 8126 CEL6B内切葡聚糖酶;土生梭孢壳NRRL 8126 CEL6C内切葡聚糖酶;土生梭孢壳NRRL 8126 CEL7C内切葡聚糖酶;土生梭孢壳NRRL 8126 CEL7E内切葡聚糖酶;土生梭孢壳NRRL 8126 CEL7F内切葡聚糖酶;*Cladorrhinum foecundissimum* ATCC 62373 CEL7A内切葡聚糖酶;以及里氏木霉菌株VTT-D-80133内切葡聚糖酶(GENBANK™登录号M15665)。

[0337] 可用于本发明的纤维二糖水解酶的示例包括但不限于，里氏木霉纤维二糖水解酶I;里氏木霉纤维二糖水解酶II;特异腐质霉纤维二糖水解酶I、嗜热毁丝霉纤维二糖水解酶II、土生梭孢壳纤维二糖水解酶II(CEL6A)、嗜热毛壳菌(*Chaetomium thermophilum*)纤维二糖水解酶I以及嗜热毛壳菌纤维二糖水解酶II、烟曲霉纤维二糖水解酶I和烟曲霉纤维二糖水解酶II。

[0338] 可用于本发明的β-葡糖苷酶的实例包括但不限于米曲霉β-葡糖苷酶;烟曲霉β-葡糖苷酶;巴西青霉(*Penicillium brasilianum*)IBT 20888β-葡糖苷酶;黑曲霉β-葡糖苷酶;以及棘孢曲霉β-葡糖苷酶。

[0339] 米曲霉β-葡糖苷酶活性可根据WO 2002/095014获取。具有β-葡糖苷酶活性的烟曲霉多肽可根据WO 2005/047499获取。具有β-葡糖苷酶活性的巴西青霉多肽可根据WO 2007/019442获取。具有β-葡糖苷酶的黑曲霉多肽可根据Dan等,2000,J.Biol.Chem.275:4973-

4980获取。具有 β -葡糖苷酶活性的棘孢曲霉多肽可根据Kawaguchi等,1996, Gene 173:287-288获取。

[0340] 所述 β -葡糖苷酶可为融合蛋白。在一个方面,所述 β -葡糖苷酶为米曲霉 β -葡糖苷酶变体BG融合蛋白,或根据WO 2008/057637获得的米曲霉 β -葡糖苷酶融合蛋白。

[0341] 其它可用的内切葡聚糖酶、纤维二糖水解酶和 β -葡糖苷酶使用根据Henrissat B.,1991,A classification of glycosyl hydrolases based on amino-acid sequence similarities,Biochem.J.280:309-316和Henrissat B.和Bairoch A.,1996,Updating the sequence-based classification of glycosyl hydrolases,Biochem.J.316:695-696的分类公开于许多糖基水解酶家族中。

[0342] 其它可用于本发明的纤维素分解酶描述于EP 495,257、EP 531,315、EP 531,372、WO 89/09259、WO 94/07998、WO 95/24471、WO 96/11262、WO 96/29397、WO 96/034108、WO 97/14804、WO 98/08940、WO 98/012307、WO 98/13465、WO 98/015619、WO 98/015633、WO 98/028411、WO 99/06574、WO 99/10481、WO 99/025846、WO 99/025847、WO 99/031255、WO 2000/009707、WO 2002/050245、WO 2002/0076792、WO 2002/101078、WO 2003/027306、WO 2003/052054、WO 2003/052055、WO 2003/052056、WO 2003/052057、WO 2003/052118、WO 2004/016760、WO 2004/043980、WO 2004/048592、WO 2005/001065、WO 2005/028636、WO 2005/093050、WO 2005/093073、WO 2006/074005、WO 2006/117432、WO 2007/071818、WO 2007/071820、WO 2008/008070、WO 2008/008793、美国专利No.4,435,307、美国专利No.5,457,046、美国专利No.5,648,263、美国专利No.5,686,593、美国专利No.5,691,178、美国专利No.5,763,254以及美国专利No.5,776,757。

[0343] 在一个方面,所述一个或多个(几个)半纤维素分解酶包含商业性半纤维素分解酶制备物。适用于本发明的商业性半纤维素分解酶制备物的实例包括,例如SHEARZYME™(Novozymes A/S)、CELLIC™HTec(Novozymes A/S)、VISCOZYME®(Novozymes A/S)、ULTRAFLO®(Novozymes A/S)、PULPZYME®HC(Novozymes A/S)、MULTIFECT® Xylanase(Genencor)、ECOPULP®TX-200A(AB Enzymes)、HSP 6000Xylanase(DSM)、DEPOL™333P(Biocatalysts Limit,Wales,UK)、DEPOL™740L.(Biocatalysts Limit,Wales,UK)和DEPOL™762P(Biocatalysts Limit,Wales,UK)。

[0344] 可用于本发明方法的木聚糖酶的实例包括但不限于棘孢曲霉(*Aspergillus aculeatus*)木聚糖酶(GeneSeqP:AAR63790;WO 94/21785)、烟曲霉(*Aspergillus fumigatus*)木聚糖酶(WO 2006/078256)和土生梭孢壳(*Thielavia terrestris*)NRRL 8126木聚糖酶(WO 2009/079210)。

[0345] 可用于本发明方法的 β -木糖苷酶的实例包括但不限于里氏木霉(*Trichoderma reesei*) β -木糖苷酶(UniProtKB/TrEMBL登录号Q92458),埃默森踝节菌(*Talaromyces emersonii*)(SwissProt登录号Q8X212)和粗糙脉孢菌(*Neurospora crassa*)(SwissProt登录号Q7S0W4)。

[0346] 可用于本发明方法的乙酰木聚糖酯酶的实例包括但不限于红褐肉座菌(*Hypocrea jecorina*)乙酰木聚糖酯酶(WO 2005/001036)、粗糙脉孢菌乙酰木聚糖酯酶(UniProt登录号q7s259)、土生梭孢壳NRRL 8126乙酰木聚糖酯酶(WO 2009/042846)、球毛壳菌(*Chaetomium globosum*)乙酰木聚糖酯酶(Uniprot登录号Q2GWX4)、细丽毛壳菌

(*Chaetomium gracile*)乙酰木聚糖酯酶(GeneSeqP登录号AAB82124)、颖枯壳针孢(*Phaeosphaeria nodorum*)乙酰木聚糖酯酶(Uniprot登录号Q0UHJ1)和特异腐质霉(*Humicola insolens*)DSM 1800乙酰木聚糖酯酶(WO 2009/073709)。

[0347] 可用于本发明方法的阿魏酸酯酶的实例包括但不限于特异腐质霉DSM 1800阿魏酸酯酶(WO 2009/076122)、粗糙脉孢菌阿魏酸酯酶(UniProt登录号Q9HGR3)和费希新萨托菌(*Neosartorya fischeri*)阿魏酸酯酶(UniProt登录号A1D9T4)。

[0348] 可用于本发明方法的阿拉伯呋喃糖苷酶的实例包括但不限于特异腐质霉(*Humicola insolens*)DSM 1800阿拉伯呋喃糖苷酶(WO 2009/073383)和黑曲霉(*Aspergillus niger*)阿拉伯呋喃糖苷酶(GeneSeqP登录号AAR94170)。

[0349] 可用于本发明方法的 α -葡糖醛酸糖苷酶的实例包括但不限于棒曲霉(*Aspergillus clavatus*) α -葡糖醛酸糖苷酶(UniProt登录号alcc12)、里氏木霉(*Trichoderma reesei*) α -葡糖醛酸糖苷酶(Uniprot登录号Q99024)、埃默森踝节菌(*Talaromyces emersonii*) α -葡糖醛酸糖苷酶(UniProt登录号Q8X211)、黑曲霉(*Aspergillus niger*) α -葡糖醛酸糖苷酶(Uniprot登录号Q96WX9)、土曲霉(*Aspergillus terreus*) α -葡糖醛酸糖苷酶(SwissProt登录号Q0CJP9)和烟曲霉(*Aspergillus fumigatus*) α -葡糖醛酸糖苷酶(SwissProt登录号Q4WW45)。

[0350] 用于本发明方法的酶和蛋白可通过使用本领域已知方法(参见,例如Bennett, J.W.和LaSure,L.(编),*More Gene Manipulations in Fungi*,Academic Press,CA,1991),在含有合适碳源和氮源和无机盐的营养培养基上发酵上述指出的微生物菌株来产生。合适的培养基可从供应商获得,或可根据已公开的组成制备(例如美国典型培养物保藏中心的目录)。适于生长和酶产生的温度范围和其他条件在本领域是已知的(参见,例如Bailey, J.E.和Ollis,D.F.,*Biochemical Engineering Fundamentals*,McGraw-Hill Book Company,NY,1986)。

[0351] 所述发酵可以是任何其结果为酶或蛋白表达或分离的培养细胞的方法。因此,发酵可以理解为包括在合适的培养基中并在允许所述酶得以表达或分离的条件下进行的摇瓶培养,或在实验室或工业发酵罐中的小-或大规模发酵(包括连续、分批、补料分批或固态发酵)。通过上述方法产生的所得的酶可从发酵培养基回收并通过常规方法纯化。

[0352] 发酵。可通过一种或多种(几种)能将糖直接或间接发酵成所需发酵产物的发酵微生物发酵自经水解的纤维素材料获得可发酵糖。“发酵”或“发酵方法”指任何发酵方法或包含发酵步骤的任何方法。发酵方法还包括用于消费品醇工业(例如,啤酒和葡萄酒)、乳品业(例如,发酵乳制品)、皮革业和烟草业的发酵方法。发酵条件依赖于期望的发酵产物和发酵生物体,并且能由本领域的技术人员容易地确定。

[0353] 在发酵步骤中,作为预处理和酶水解步骤的结果从纤维素材料释放的糖,通过发酵生物体(如酵母)发酵成为产物,例如,乙醇。如上所述,水解(糖化)和发酵可以是分开的或同时的。

[0354] 在本发明的发酵步骤中可以使用任何合适的经水解的纤维素材料。通常根据所需的发酵产品(即,要从发酵获得的物质)和使用的方法来选择所述材料,如本领域中所公知的。

[0355] 术语“发酵培养基”在本文中可理解为指加入发酵微生物之前的培养基,如,由糖

化过程产生的培养基,以及同步的糖化和发酵方法(SSF)中使用的培养基。

[0356] “发酵微生物”指适用于理想的发酵方法产生发酵产物的任何微生物,包括细菌和真菌生物体。发酵生物体可以是C₆和/或C₅发酵生物体,或它们的组合。C₆和C₅发酵生物体均在本领域公知。合适的发酵微生物能将糖(如葡萄糖、木糖、木酮糖、阿拉伯糖、麦芽糖、甘露糖、半乳糖或寡糖)直接或间接地发酵(即,转化)成所需的发酵产品。

[0357] 产生乙醇的细菌和真菌发酵生物体的实例如Lin等,2006,Appl.Microbiol.Biotechnol.69:627-642所述。

[0358] 能发酵C₆糖的发酵微生物的实例包括细菌和真菌生物体,如酵母。优选的酵母包括酵母属菌种,优选酿酒酵母。

[0359] 能发酵C₅糖的发酵生物体的实例包括细菌和真菌生物体,如酵母。优选的C₅发酵酵母包括毕赤酵母属,优选树干毕赤酵母(*Pichia stipitis*)的菌株,如树干毕赤酵母CBS 5773;假丝酵母属,优选博伊丁假丝酵母(*Candida boidinii*)、芸薹假丝酵母(*Candida brassicae*)、休哈塔假丝酵母(*Candida sheatae*)、迪丹斯假丝酵母(*Candida diddensii*)、假热带假丝酵母(*Candida pseudotropicalis*)或产朊假丝酵母(*Candida utilis*)的菌株。

[0360] 其它发酵生物体包括发酵单胞菌属(*Zymomonas*),如运动发酵单胞菌(*Zymomonas mobilis*);汉逊酵母属,如异常汉逊酵母(*Hansenula anomala*);克鲁维酵母属,如脆壁克鲁维酵母;裂殖酵母属,如粟酒裂殖酵母(*S.pombe*);大肠杆菌,特别是已经经过遗传修饰而改进乙醇产量的大肠杆菌;梭菌属(*Clostridium*),如丙酮丁醇梭菌(*Clostridium acetobutylicum*),*Chlostridium thermocellum*,和*Chlostridium phytofermentans*;地芽孢杆菌属菌种(*Geobacillus sp.*);热厌氧杆菌属(*Thermoanaerobacter*),如解糖热厌氧杆菌(*Thermoanaerobacter saccharolyticum*);和芽孢杆菌属(*Bacillus*),如凝结芽孢杆菌(*Bacillus coagulans*)。

[0361] 在一个优选的方面,酵母是酵母属菌种。在一个更优选的方面,酵母是酿酒酵母。在另一个更优选的方面,酵母是糖化酵母(*Saccharomyces distaticus*)。在另一个更优选的方面,酵母是葡萄汁酵母(*Saccharomyces uvarum*)。在另一个优选的方面,酵母是克鲁维酵母属。在另一个更优选的方面,酵母是马克斯克鲁维酵母(*Kluyveromyces marxianus*)。在另一个更优选的方面,酵母是脆壁克鲁维酵母。在另一个优选的方面,酵母是假丝酵母属。在另一个更优选的方面,酵母是博伊丁假丝酵母。在另一个更优选的方面,酵母是芸薹假丝酵母。在另一个更优选的方面,酵母是迪丹斯假丝酵母。在另一个更优选的方面,酵母是假热带假丝酵母。在另一个更优选的方面,酵母是产朊假丝酵母。在另一个优选的方面,酵母是棒孢酵母属(*Clavispora*)。在另一个更优选的方面,酵母是葡萄牙棒孢酵母(*Clavispora lusitaniae*)。在另一个更优选的方面,酵母是仙人掌棒孢酵母(*Clavispora opuntiae*)。在另一个优选的方面,酵母是管囊酵母属(*Pachysolen*)。在另一个更优选的方面,酵母是嗜鞣管囊酵母(*Pachysolen tannophilus*)。在另一个优选的方面,酵母是毕赤酵母属。在另一个更优选的方面,酵母是树干毕赤酵母。在另一个优选的方面,酵母是酒香酵母属(*Bretannomyces*)。在另一个更优选的方面,酵母是克劳森酒香酵母(*Bretannomyces clausenii*)(Philippidis,G.P.,1996,Cellulose bioconversion technology,于Handbook on Bioethanol:Production and Utilization,Wyman,C.E.编,Taylor&Francis,Washington,DC,179-212)。

[0362] 能有效地将己糖和戊糖发酵成乙醇的细菌包括,例如,运动发酵单胞菌和丙酮丁醇梭菌,*Clostridium thermocellum*,*Chlostridium phytofermentans*,地芽孢杆菌属菌种,解糖热厌氧杆菌和凝结芽孢杆菌(Philippidis,1996,见上文)。

[0363] 在一个优选的方面,细菌是发酵单胞菌属。在更优选的方面,细菌是运动发酵单胞菌。在另一个优选的方面,细菌是梭菌属。在另一个更优选的方面,细菌是热纤维梭菌。

[0364] 商业上可得到的适合乙醇产生的酵母包括,例如ETHANOL RED™酵母(Red Star/Lesaffre,USA)、FALI™(Fleischmann's Yeast,USA)、SUPERSTART™和THERMOSACC™新鲜酵母(Ethanol Technology,WI,USA)、BIOFERM™AFT和XR(NABC-North American Bioproducts Corporation,GA,USA)、GERT STRAND™(Gert Strand AB,Sweden)和FERMIOL™(DSM Specialties)。

[0365] 在一个优选的方面,发酵微生物已经经过遗传修饰,提供发酵戊糖的能力,如利用木糖、利用阿拉伯糖和共同利用木糖和阿拉伯糖的微生物。

[0366] 通过将异源基因克隆入多种发酵微生物已经构建了能将己糖和戊糖转化成乙醇(共发酵)的生物体(Chen和Ho,1993,Cloning and improving the expression of *Pichia stipitis* xylose reductase gene in *Saccharomyces cerevisiae*, *Appl. Biochem. Biotechnol.* 39-40:135-147;Ho等,1998,Genetically engineered *Saccharomyces* yeast capable of effectively cofermenting glucose and xylose, *Appl. Environ. Microbiol.* 64:1852-1859;Kotter和Ciriacy,1993,Xylose fermentation by *Saccharomyces cerevisiae*, *Appl. Microbiol. Biotechnol.* 38:776-783;Walfridsson等,1995,Xylose-metabolizing *Saccharomyces cerevisiae* strains overexpressing the TKL1 and TAL1 genes encoding the pentose phosphate pathway enzymes transketolase and transaldolase, *Appl. Environ. Microbiol.* 61:4184-4190;Kuyper等,2004,Minimal metabolic engineering of *Saccharomyces cerevisiae* for efficient anaerobic xylose fermentation:a proof of principle, *FEMS Yeast Research* 4:655-664;Beall等,1991,Parametric studies of ethanol production from xylose and other sugars by recombinant *Escherichia coli*, *Biotech. Bioeng.* 38:296-303;Ingram等,1998,Metabolic engineering of bacteria for ethanol production, *Biotechnol. Bioeng.* 58:204-214;Zhang等,1995,Metabolic engineering of a pentose metabolism pathway in ethanologenic *Zymomonas mobilis*, *Science* 267:240-243;Deanda等,1996,Development of an arabinose-fermenting *Zymomonas mobilis* strain by metabolic pathway engineering, *Appl. Environ. Microbiol.* 62:4465-4470;WO 2003/062430,xylose isomerase)。

[0367] 在一个优选的方面,经过遗传修饰的发酵微生物是酿酒酵母。在另一个优选的方面,经过遗传修饰的发酵微生物是运动发酵单胞菌。在另一个优选的方面,经过遗传修饰的发酵微生物是大肠杆菌。在另一个优选的方面,经过遗传修饰的发酵微生物是产酸克雷伯氏菌(*Klebsiella oxytoca*)。在另一个优选的方面,所述经遗传修饰的发酵微生物是克鲁维酵母菌种。

[0368] 本领域中公知的是,上述生物体还能用于产生其它物质,如本文所述。

[0369] 通常向降解的木素纤维素或水解物加入发酵微生物,并进行约8至约96小时,如约

24至约60小时发酵。温度通常为约26℃至约60℃,特别是约32℃或50℃,并且在约pH 3至约pH 8,如约pH 4-5、6或7。

[0370] 在一个优选的方面,对降解的纤维素材料施用酵母和/或另一种微生物,并进行约12至约96小时,如通常为24-60小时发酵。在一个优选的方面,温度优选为约20℃至约60℃,更优选约25℃至约50℃,并且最优选约32℃至约50℃,特别是约32℃或50℃,并且pH通常为约pH 3至约pH 7,优选约pH 4-7。然而,一些发酵生物体例如细菌,具有更高的最适发酵温度。酵母或另一种微生物优选以约 10^5 - 10^{12} ,优选约 10^7 - 10^{10} ,特别是约 2×10^8 活细胞计数每ml发酵液的量施用。关于使用酵母进行发酵的进一步指导可以在例如“The Alcohol Textbook”(K.Jacques,T.P.Lyons和D.R.Kelsall编,Nottingham University Press, United Kingdom 1999)中找到,其通过提述并入本文。

[0371] 对于乙醇生产,在发酵后蒸馏发酵的浆料以提取乙醇。根据本发明的方法获得的乙醇可以用作,例如燃料乙醇;饮料乙醇,即,中性饮料酒,或工业乙醇。

[0372] 发酵刺激剂可以与本文所述的任何方法组合使用,以进一步改进发酵工艺,而且特定地,改进发酵微生物的性能,如,速率增加和乙醇得率。“发酵刺激剂”指用于发酵微生物(特别是酵母)生长的刺激剂。优选的用于生长的发酵刺激剂包括维生素和矿物质。维生素的实例包括多种维生素、生物素、泛酸(盐)、烟酸、内消旋肌醇(meso-inositol)、硫胺素、吡哆醇(pyridoxine)、对氨基苯甲酸、叶酸、核黄素和维生素A、B、C、D和E。参见,例如,Alfenore等,Improving ethanol production and viability of *Saccharomyces cerevisiae* by a vitamin feeding strategy during fed-batch process, Springer-Verlag(2002),其通过提述并入本文。矿物质的实例包括能够提供营养物的矿物质和矿物质盐,所述营养物包括P、K、Mg、S、Ca、Fe、Zn、Mn和Cu。

[0373] 发酵产物:发酵产物可以是源自发酵的任何物质。发酵产物可以是,不限于,醇(例如,阿拉伯醇、丁醇、乙醇、甘油、甲醇、1,3-丙二醇、山梨醇和木糖醇);有机酸(例如,乙酸、醋酮酸、己二酸、抗坏血酸、柠檬酸、2,5-二酮-D-葡萄糖酸、甲酸、反丁烯二酸、葡萄糖二酸、葡萄糖酸、葡萄糖醛酸、戊二酸、3-羟基丙酸、衣康酸、乳酸、苹果酸、丙二酸、草酸、草酰乙酸、丙酸、琥珀酸和木糖酸);酮(例如,丙酮);氨基酸(例如,天冬氨酸、谷氨酸、甘氨酸、赖氨酸、丝氨酸和苏氨酸);和气体(例如,甲烷、氢气(H_2)、二氧化碳(CO_2)和一氧化碳(CO))。发酵产物还可以是作为高价值产品的蛋白质。

[0374] 在一个优选的方面,发酵产物是醇。可理解的是,术语“醇”包括包含一个或多个羟基基团的物质。在更优选的方面,所述醇是阿拉伯醇。在另一个更优选的方面,所述醇是丁醇。在另一个更优选的方面,所述醇是乙醇。在另一个更优选的方面,所述醇是甘油。在另一个更优选的方面,所述醇是甲醇。在另一个更优选的方面,所述醇是1,3-丙二醇。在另一个更优选的方面,所述醇是山梨醇。在另一个更优选的方面,所述醇是木糖醇。参见,例如,Gong,C.S.,Cao,N.J.,Du,J.,和Tsao,G.T.,1999,Ethanol production from renewable resources,于Advances in Biochemical Engineering/Biotechnology,Scheper,T.编, Springer-Verlag Berlin Heidelberg,Germany,65:207-241;Silveira,M.M.,和Jonas,R.,2002,The biotechnological production of sorbitol, Appl.Microbiol.Biotechnol.59:400-408;Nigam,P.,和Singh,D.,1995,Processes for fermentative production of xylitol-a sugar substitute,Process Biochemistry 30

(2):117-124;Ezeji,T.C.,Qureshi,N.和Blaschek,H.P.,2003,Production of acetone, butanol and ethanol by *Clostridium beijerinckii* BA101 and in situ recovery by gas stripping,World Journal of Microbiology and Biotechnology 19(6):595-603。

[0375] 在另一个优选的方面,所述发酵产物是有机酸。在另一个更优选的方面,所述有机酸是乙酸。在另一个更优选的方面,所述有机酸是醋酮酸。在另一个更优选的方面,所述有机酸是己二酸。在另一个更优选的方面,所述有机酸是抗坏血酸。在另一个更优选的方面,所述有机酸是柠檬酸。在另一个更优选的方面,所述有机酸是2,5-二酮-D-葡糖酸。在另一个更优选的方面,所述有机酸是甲酸。在另一个更优选的方面,所述有机酸是反丁烯二酸。在另一个更优选的方面,所述有机酸是葡糖二酸。在另一个更优选的方面,所述有机酸是葡糖酸。在另一个更优选的方面,所述有机酸是葡糖醛酸。在另一个更优选的方面,所述有机酸是戊二酸。在另一个更优选的方面,所述有机酸是3-羟基丙酸。在另一个更优选的方面,所述有机酸是衣康酸。在另一个更优选的方面,所述有机酸是乳酸。在另一个更优选的方面,所述有机酸是苹果酸。在另一个更优选的方面,所述有机酸是丙二酸。在另一个更优选的方面,所述有机酸是草酸。在另一个更优选的方面,所述有机酸是丙酸。在另一个更优选的方面,所述有机酸是琥珀酸。在另一个更优选的方面,所述有机酸是木糖酸。参见,例如,Chen,R.,和Lee,Y.Y.,1997,Membrane-mediated extractive fermentation for lactic acid production from cellulosic biomass,Appl.Biochem.Biotechnol.63-65:435-448。

[0376] 在另一个优选的方面,所述发酵产物是酮。可理解的是术语“酮”涵盖含有一个或多个酮基的物质。在另一个更优选的方面,所述酮是丙酮。参见,例如,Qureshi和Blaschek,2003,见上文。

[0377] 在另一个优选的方面,所述发酵产物是氨基酸。在另一个更优选的方面,所述有机酸是天冬氨酸。在另一个更优选的方面,所述氨基酸是谷氨酸。在另一个更优选的方面,所述氨基酸是甘氨酸。在另一个更优选的方面,所述氨基酸是赖氨酸。在另一个更优选的方面,所述氨基酸是丝氨酸。在另一个更优选的方面,所述氨基酸是苏氨酸。参见,例如,Richard,A.,和Margaritis,A.,2004,Empirical modeling of batch fermentation kinetics for poly(glutamic acid)production and other microbial biopolymers,Biotechnology and Bioengineering 87(4):501-515。

[0378] 在另一个优选的方面,所述发酵产物是气体。在另一个更优选的方面,所述气体是甲烷。在另一个更优选的方面,所述气体是H₂。在另一个更优选的方面,所述气体是CO₂。在另一个更优选的方面,所述气体是CO。参见,例如,Kataoka,N.,A.Miya,和K.Kiriyama,1997,Studies on hydrogen production by continuous culture system of hydrogen-producing anaerobic bacteria,Water Science and Technology 36(6-7):41-47;和Gunaseelan V.N.于Biomass and Bioenergy,Vol.13(1-2),pp.83-114,1997,Anaerobic digestion of biomass for methane production:A review。

[0379] 回收。可以使用本领域已知的任何方法,任选地从发酵培养基回收发酵产物,所述方法包括,但不限于,层析、电泳方法、差示溶解度、蒸馏或提取。例如,通过常规蒸馏方法从发酵的纤维素材料分离并纯化醇。可以获得纯度高达约96vol%的乙醇,其能用作,例如,燃料乙醇、饮用乙醇,即,中性饮料酒,或工业乙醇。

[0380] 洗涤剂组合物

[0381] 本发明具有纤维素分解增强活性的多肽可添加至洗涤剂组合物而成为洗涤剂组合物的组分。

[0382] 本发明的洗涤剂组合物可配制为例如手洗或机洗的洗衣洗涤剂组合物,其包含适于预处理玷污的织物的洗衣添加剂组合物和漂洗添加的织物柔软剂(softener)组合物,或可配制为用于一般家用硬表面清洁操作的洗涤剂组合物,或可配制为用于手动或机械洗碗操作。因此,本发明还涉及用于清洗或洗涤硬表面或衣物的方法,包括将所述硬表面或衣物与本发明的洗涤剂组合物相接触。

[0383] 在一个具体方面,本发明提供了包含本发明多肽的洗涤剂添加剂。所述洗涤剂添加剂以及所述洗涤剂组合物还可包含一种或多种选自下组的酶:淀粉酶、阿拉伯糖酶、纤维素酶、角质酶、半乳聚糖酶、半纤维素酶、漆酶、脂质酶、甘露聚糖酶、氧化酶、果胶酶、蛋白酶、和木聚糖酶。

[0384] 一般而言所选酶的性质应与所选洗涤剂相容(即最适pH,与其他酶或非酶成分的相容性等),且所述酶应以有效量存在。

[0385] 纤维素酶:合适的纤维素酶包括那些细菌或真菌起源的。包括化学修饰的或蛋白质工程改造的突变体。合适的纤维素酶包括来自芽孢杆菌属、假单胞菌属、腐质霉属、镰孢属、梭孢壳属、枝顶孢霉属的纤维素酶,例如US 4,435,307、US 5,648,263、US 5,691,178、US 5,776,757号和WO 89/09259中公开的,由特异腐质酶、嗜热毁丝霉和尖镰孢产生的真菌纤维素酶。

[0386] 特别合适的纤维素酶是具有色彩维护(colour care)益处的碱性或中性纤维素酶。此类纤维素酶的例子有EP 0 495 257、EP 0 531 372、WO 96/11262、WO 96/29397、WO 98/08940中记载的纤维素酶。其它例子有纤维素酶变体,如WO 94/07998、EP 0 531 315、US 5,457,046、US 5,686,593、US 5,763,254、WO 95/24471、WO 98/12307和PCT/DK98/00299中记载的那些。

[0387] 商业性使用的纤维素酶包括CELLUZYME™和CAREZYME™(Novozymes A/S)、CLAZINASE™和PURADAX HA™(Genencor International Inc.)和KAC-500(B)™(Kao Corporation)。

[0388] 蛋白酶:合适的蛋白酶包括动物、植物或微生物来源的那些。优选微生物来源。包括化学修饰或蛋白质工程改造的突变体。蛋白酶可以是丝氨酸蛋白酶或金属蛋白酶,优选碱性微生物蛋白酶或胰蛋白酶样蛋白酶。碱性蛋白酶的实例是枯草杆菌蛋白酶,特别是源自芽孢杆菌属的那些,例如,枯草杆菌蛋白酶Novo,枯草杆菌蛋白酶Carlsberg,枯草杆菌蛋白酶309,枯草杆菌蛋白酶147和枯草杆菌蛋白酶168(在WO 89/06279中描述)。胰蛋白酶样蛋白酶的实例是胰蛋白酶(例如猪或牛来源的)和WO 89/06270和WO 94/25583中所述的镰孢属蛋白酶。

[0389] 有用的蛋白酶的实例是WO 92/19729、WO 98/20115、WO 98/20116和WO 98/34946中所述的变体,特别是在下述一个或多个位置有取代的变体:27、36、57、76、87、97、101、104、120、123、167、170、194、206、218、222、224、235和274。

[0390] 优选的商业上可获得的蛋白酶包括ALCALASE™、SAVINASE™、PRIMASE™、DURALASE™、ESPERASE™和KANNASE™(Novozymes A/S)、MAXATASE™、MAXACAL™、MAXAPEM™、PROPERASE™、PURAFACT™、PURAFACT OXP™、FN2™和FN3™(Genencor International Inc.)。

[0391] **脂肪酶**:合适的脂肪酶包括那些细菌或真菌起源的。包括化学修饰的或蛋白质工程改造的突变体。有用的脂肪酶的例子包括来自腐质霉属(同物异名嗜热霉属(*Thermomyces*))的脂肪酶,例如记载于EP 258 068和EP 305 216的来自疏棉状腐质霉(*H.lanuginosa*)(细毛嗜热霉(*T.lanuginosus*))或来自记载于WO 96/13580的特异腐质霉的脂肪酶,假单胞菌属脂肪酶,例如来自产碱假单胞菌(*P.alcaligenes*)和类产碱假单胞菌(*P.pseudoalcaligenes*)(EP 218 272)、洋葱假单胞菌(EP 331 376)、施氏假单胞菌(*P.stutzeri*)(GB 1,372,024)、荧光假单胞菌(*P.fluorescens*)、假单胞菌属菌种菌株SD 705(WO 95/06720和WO 96/27002)、威斯康星假单胞菌(*P.wisconsinensis*)(WO 96/12012)的脂肪酶,或芽孢杆菌属的脂肪酶,例如来自枯草芽孢杆菌(Dartois等,(1993), *Biochemica et Biophysica acta* 1131,253-260),嗜热脂肪芽孢杆菌(JP 64/744992)或短小芽孢杆菌(WO 91/16422)的脂肪酶。

[0392] 其它实例为脂肪酶变体,如WO 92/05249、WO 94/01541、EP 407 225、EP 260 105、WO 95/35381、WO 96/00292、WO 95/30744、WO 94/25578、WO 95/14783、WO 95/22615、WO 97/04079和WO 97/07202中记载的那些。

[0393] 优选的商业上可获得的脂肪酶包括LIPOLASE™和LIPOLASE ULTRA™(Novozymes A/S)。

[0394] **淀粉酶**:合适的淀粉酶(α 和/或 β)包括那些细菌或真菌起源的。包括化学修饰的或蛋白质工程改造的突变体。淀粉酶包括例如得自芽孢杆菌属例如在GB 1,296,839中更详细所述的地衣芽孢杆菌特定菌株的 α -淀粉酶。

[0395] 有用的淀粉酶的实例为WO 94/02597、WO 94/18314、WO 96/23873和WO 97/43424中所述的变体,尤其是在一个或多个以下位置具有取代的变体:15、23、105、106、124、128、133、154、156、181、188、190、197、202、208、209、243、264、304、305、391、408和444。

[0396] 商业上可获得的淀粉酶有DURAMYL™、TERMAMYL™、FUNGAMYL™和BAN™(Novozymes A/S)、RAPIDASE™和PURASTAR™(来自Genencor International Inc.)。

[0397] **过氧化物酶/氧化酶**:合适的过氧化物酶/氧化酶包括那些植物、细菌或真菌起源的。包括化学修饰的或蛋白质工程改造的突变体。有用的过氧化物酶的例子包括来自鬼伞属(*Coprinus*)的过氧化物酶,例如来自灰盖鬼伞(*C.cinereus*)的过氧化物酶及其变体,如WO 93/24618、WO 95/10602、和WO 98/15257中记载的那些。

[0398] 商业上可获得的过氧化物酶包括GUARDZYME™(Novozymes A/S)。

[0399] 可以通过添加分开的、含有一种或多种酶的添加剂或通过添加组合的、包含所有这些酶的添加剂将洗涤剂酶包含于洗涤剂组合物中。本发明的洗涤剂添加剂,即单独的添加剂或组合的添加剂,可配制为例如颗粒、液体、浆料等。优选的洗涤剂添加剂配制物是颗粒,特别是无粉尘的颗粒、液体特别是稳定化的液体,或浆料。

[0400] 无粉尘的颗粒可以例如如US 4,106,991和4,661,452中公开的产生,并且可以任选地通过本技术领域已知的方法进行包覆。蜡质包覆材料的实例是具有1000至20000的平均摩尔重量的聚(环氧乙烷)产品(聚乙二醇,PEG);具有16至50个环氧乙烷单元的乙氧基化壬基酚;乙氧基化脂肪醇,其中所述醇含有12至20个碳原子且其中存在15至80个环氧乙烷单元;脂肪醇;脂肪酸;和脂肪酸的甘油一酯、甘油二酯和甘油三酯。适于通过流化床技术施用的成膜包覆材料的实例在GB 1483591中给出。例如,可以通过根据已确立的方法添加多

元醇如丙二醇、糖或糖醇、乳酸或硼酸来使液体酶制备物稳定。受保护的酶可根据EP 238, 216中公开的方法来制备。

[0401] 本发明的洗涤剂组合物可以是任何方便的形式,例如,条、片剂、粉末、颗粒(granule)、糊剂(paste)或液体。液体洗涤剂可以是水性的,通常含有多至70%水和0-30%有机溶剂,或是非水性的。

[0402] 洗涤剂组合物包含一种或多种表面活性剂,所述表面活性剂可以是非离子的(包括半-极性的)和/或阴离子的和/或阳离子的和/或两性离子的。表面活性剂通常以按重量计0.1%至60%的水平存在。

[0403] 当包含于其中时,洗涤剂通常会含有约1%至约40%的阴离子表面活性剂,例如直链烷基苯磺酸酯(linear alkylbenzenesulfonate)、 α -链烯烃磺酸酯、烷基硫酸酯(脂肪醇硫酸酯)、醇乙氧基硫酸酯(alcohol ethoxysulfate)、仲烷基磺酸酯(secondary alkanesulfonate)、 α -磺基脂肪酸甲酯、烷基-或烯基琥珀酸或皂类(soap)。

[0404] 当包含于其中时,洗涤剂将通常含有约0.2%至约40%的非离子表面活性剂,例如醇乙氧基化物、壬基酚乙氧基化物、烷基聚苷(alkylpolyglycoside)、烷基二甲胺氧化物(alkyldimethylamineoxide)、乙氧基化脂肪酸单乙醇酰胺(ethoxylated fatty acid monoethanolamide)、脂肪酸单乙醇酰胺、多羟基烷基脂肪酸酰胺,或葡糖胺的N-酰基N-烷基衍生物(“葡糖酰胺(glucamides)”)。

[0405] 洗涤剂可以含有0-65%的洗涤剂增效剂(builder)或络合剂例如沸石、二磷酸盐/酯、三磷酸盐/酯、膦酸盐/酯(phosphonate)、碳酸盐/酯、柠檬酸盐/酯、氮川三乙酸(nitritotriacetic acid)、乙二胺四乙酸、二亚乙基三胺五乙酸、烷基-或烯基琥珀酸、可溶性硅酸盐或层状硅酸盐(例如来自Hoechst的SKS-6)。

[0406] 洗涤剂可包含一种或多种聚合物。实例是羧甲基纤维素、聚(乙烷基吡咯烷酮)、聚(乙二醇)、聚(乙烷基醇)、聚(乙烷基吡啶-N-氧化物)、聚(乙烷基咪唑)、聚羧酸盐/酯(polycarboxylate)例如聚丙烯酸盐/酯、马来酸/丙烯酸共聚物和甲基丙烯酸月桂酯/丙烯酸共聚物。

[0407] 洗涤剂可含有漂白体系,所述体系可包含H₂O₂源例如过硼酸盐或过碳酸盐,可将其与形成过酸的漂白活化剂例如四乙酰乙二胺或壬酰氧苯磺酸酯(nonanoyloxybenzenesulfonate)组合。或者,漂白体系可包含过氧酸,例如酰胺、二酰亚胺或砷型的过氧酸。

[0408] 可以使用常规稳定剂使本发明的洗涤剂组合物中的酶稳定化,所述稳定剂例如,多元醇如丙二醇或甘油,糖或糖醇,乳酸,硼酸或硼酸衍生物,例如,芳族硼酸酯,或苯基硼酸(phenyl boronic acid)衍生物例如4-甲酰苯基硼酸,并且所述组合物可按例如W0 92/19709和W0 92/19708中所述配制。

[0409] 洗涤剂也可含有其它常规洗涤剂成分例如织物调节剂,包括粘土、泡沫促进剂(foam booster)、抑泡剂(suds suppressor)、抗腐蚀剂、悬污剂(soil-suspending agent)、防污物再沉积剂(anti-soil redeposition agent)、染料、杀菌剂、光学增亮剂、助水溶物(hydrotrope)、晦暗抑制剂(tarnish inhibitors)或香料。

[0410] 在洗涤剂组合物中,可以以相当于0.01-100mg酶蛋白每升洗涤液,优选0.05-5mg酶蛋白每升洗涤液,特别是0.1-1mg酶蛋白每升洗涤液的量添加任何酶。

[0411] 在洗涤剂组合物中,可以以相当于每升洗涤液0.001-100mg蛋白、优选0.005-50mg蛋白,更优选0.01-25mg蛋白,甚至更优选0.05-10mg蛋白,最优选0.05-5mg蛋白,且甚至最优选0.01-1mg蛋白的量添加本发明具有纤维素分解增强活性的多肽。

[0412] 还可将本发明具有纤维素分解增强活性的多肽并入WO 97/07202中公开的洗涤剂配制物,将WO 97/07202通过提述并入本文。

[0413] 信号肽

[0414] 本发明还涉及编码信号肽的多核苷酸,所述信号肽包含或组成为SEQ ID NO:2的氨基酸1至17,SEQ ID NO:4的氨基酸1至19,SEQ ID NO:6的氨基酸1至17,SEQ ID NO:8的氨基酸1至19,SEQ ID NO:10的氨基酸1至21,SEQ ID NO:12的氨基酸1至24,SEQ ID NO:14的氨基酸1至16,SEQ ID NO:16的氨基酸1至18,SEQ ID NO:18的氨基酸1至22,SEQ ID NO:20的氨基酸1至16,或SEQ ID NO:22的氨基酸1至19。所述多核苷酸还可包含编码蛋白质的基因,其可操作地连接于所述信号肽。所述蛋白质优选对于所述信号肽是外源的。

[0415] 本发明还涉及包含此种多核苷酸的核酸构建体、表达载体和重组宿主细胞。

[0416] 本发明还涉及产生蛋白质的方法,包括:(a)培养包含此种多核苷酸的重组宿主细胞;和(b)回收所述蛋白质。

[0417] 所述蛋白质对于宿主细胞可以是天然的或异源的。术语“蛋白质”在本文的意思不是指特定长度的编码产物,并且因此包含肽、寡肽和多肽。术语“蛋白质”还包含组合以形成编码产物的两种或更多种多肽。所述蛋白质还包括杂合多肽和融合多肽。

[0418] 优选地,所述蛋白质是激素或其变体、酶、受体或其部分、抗体或其部分,或报道蛋白(reporter)。例如,所述蛋白质可为氧化还原酶、转移酶、水解酶、裂合酶(lyase)、异构酶或连接酶,如氨肽酶、淀粉酶、糖酶、羧肽酶、过氧化氢酶、纤维二糖水解酶、纤维素酶、壳多糖酶、角质酶、环糊精糖基转移酶、脱氧核糖核酸酶、内切葡聚糖酶、酯酶、 α -半乳糖苷酶、 β -半乳糖苷酶、葡糖淀粉酶、 α -葡萄糖苷酶、 β -葡萄糖苷酶、转化酶、漆酶、另外的脂肪酶、甘露糖苷酶、变聚糖酶(mutanase)、氧化酶、果胶分解酶、过氧化物酶、肌醇六磷酸酶、多酚氧化酶、蛋白水解酶、核糖核酸酶、转谷氨酰胺酶或木聚糖酶。

[0419] 基因可以从任何原核、真核或其它来源获得。

[0420] 通过以下实施例进一步对本发明进行描述,但不应将其理解为对本发明范围的限制。

实施例

[0421] 材料

[0422] 用作缓冲液和底物的化学品为至少试剂级的商品。

[0423] 菌株

[0424] 将土生梭孢壳NRRL 8126用作具有纤维素分解增强活性的家族61多肽的来源。将米曲霉JaL355菌株(WO 2002/40694)用于表达编码具有纤维素分解增强活性的多肽的土生梭孢壳家族61基因。

[0425] 培养基和溶液

[0426] PDA平板包含39g的马铃薯右旋糖琼脂,和蒸馏水加至1升。

[0427] NNCYP培养基包含5.0g的 NaNO_3 ,3.0g的 NH_4Cl ,2.0g的MES,2.5g的柠檬酸,0.2g的

CaCl₂ · 2H₂O, 1.0g的Bacto Peptone, 5.0g的酵母提取物, 0.2g的MgSO₄ · 7H₂O, 4.0g的K₂HPO₄, 1.0ml的COVE微量元素溶液, 2.5g的葡萄糖, 和蒸馏水加至1升。

[0428] 基本培养基(MM)平板包含6g的NaNO₃, 0.52g的KCl, 1.52g的KH₂PO₄, 1ml的COVE微量元素溶液, 20g的Noble琼脂, 20ml的50%葡萄糖, 2.5ml的MgSO₄ · 7H₂O, 20ml的0.02%生物素溶液, 和蒸馏水加至1升。

[0429] COVE微量元素溶液包含0.04g的Na₂B₄O₇ · 10H₂O, 0.4g的CuSO₄ · 5H₂O, 1.2g的FeSO₄ · 7H₂O, 0.7g的MnSO₄ · H₂O, 0.8g的Na₂MoO₂ · 2H₂O, 10g的ZnSO₄ · 7H₂O, 和蒸馏水加至1升。

[0430] M410培养基包含50g的麦芽糖, 50g的葡萄糖, 2g的MgSO₄ · 7H₂O, 4g的无水柠檬酸粉末, 2g的KH₂PO₄, 8g的酵母提取物, 2g的尿素, 0.5g的CaCl₂, 0.5ml的AMG微量金属溶液, 和蒸馏水加至1升。

[0431] AMG微量金属溶液包含14.3g的ZnSO₄ · 7H₂O, 2.5g的CuSO₄ · 5H₂O, 0.5g的NiCl₂ · 6H₂O, 13.8g的FeSO₄ · 7H₂O, 8.5g的MnSO₄ · 7H₂O, 3g的柠檬酸, 和蒸馏水加至1升。

[0432] 实施例1:关于土生梭孢壳NRRL 8126的DNA序列信息的来源

[0433] 基因组序列信息由U.S.Department of Energy Joint Genome Institute(JGI)生成。从JGI下载基因组的初级汇编(assembly), 并使用Pedant-Pro™ Sequence Analysis Suite(Biomax Informatics AG, Martinsried, Germany)分析。使用通过软件构建的基因模型作为起始点以供检测基因组中的GH61同源物。使用多种已知的GH61蛋白序列作为指引构建更精确的基因模型。

[0434] 实施例2:土生梭孢壳NRRL 8126基因组DNA提取

[0435] 为了生成供PCR扩增的基因组DNA, 将土生梭孢壳NRRL 8126在42°C和200rpm在带挡板的摇瓶中的补充1%葡萄糖的50ml的NNCYP培养基中生长24小时。菌丝体通过过滤收获, 在TE(10mM Tris-1mM EDTA)中洗涤两次, 并在液氮下冻结。将豌豆大小的冻结的菌丝体块悬于0.7ml的TE中的1%十二烷基硫酸锂, 并通过在FastPrep FP120(ThermoSavant, Holbrook, NY, USA)中用等体积的0.1mm氧化锆/二氧化硅珠(Biospec Products, Inc., Bartlesville, OK, USA)搅拌45秒来破坏。通过以13,000xg离心10分钟来去除碎块, 并将经清除的上清配至2.5M乙酸铵, 并在冰上温育20分钟。在温育期之后, 通过添加2体积的乙醇来沉淀核酸。在4°C在微型离心机中离心15分钟之后, 将沉淀在70%乙醇中洗涤, 并风干。将DNA重悬于120μl的0.1X TE, 并在37°C用1μl的不含DNase的RNase A温育20分钟。添加乙酸铵至2.5M, 并用2体积的乙醇沉淀DNA。将沉淀在70%乙醇中洗涤, 风干, 并重悬于TE缓冲液。

[0436] 实施例3:含有编码具有纤维素分解增强活性的家族GH61J多肽的土生梭孢壳NRRL 8126基因组序列的米曲霉表达载体的构建

[0437] 设计了如下所示的两种合成的寡核苷酸引物以从实施例2中制备的基因组DNA来PCR扩增土生梭孢壳NRRL 8126 gh61j基因。使用IN-FUSION™ Cloning Kit(BD Biosciences, Palo Alto, CA, USA)以将片段直接克隆入表达载体pAllLo2(WO 2004/099228), 而无需进行限制性消化和连接。

[0438] Ttgh1j-F(065367):

[0439] 5'-actggatttacc**ATGAAGTTCTCACTGGTGTG**-3'(SEQ ID NO:23)

[0440] Ttgh61j-R(065368):

[0441] 5'-TCACCTCTAGTTAATTAAT**TCAGCAGGAGATCGGGGCGG**-3' (SEQ ID NO:24)

[0442] 粗体字母代表编码序列。剩余序列同源pAlLo2的插入位点。

[0443] 将五十皮摩尔的上述每种引物用于PCR反应,所述反应包含100ng的土生梭孢壳NRRL 8126基因组DNA,Pfx Amplification Buffer(Invitrogen,Carlsbad,CA,USA),各0.4mM的dATP,dTTP,dGTP,和dCTP,1mM MgCl₂,和2.5单位的Pfx DNA聚合酶(Invitrogen,Carlsbad,CA,USA),最终体积为50μl。使用**EPPENDORF® MASTERCYCLER®5333**(Eppendorf Scientific,Inc.,Westbury,NY,USA)进行扩增,其程序为:1个循环,在98℃进行3分钟;和30个循环,每循环在98℃进行30秒,60℃进行30秒,和72℃进行1.5分钟。然后加热块进入4℃浸没循环。

[0444] 将反应产物在1.0%琼脂糖凝胶上使用40mM Tris碱-20mM乙酸钠-1mM EDTA二钠盐(TAE)缓冲液分离,其中将908bp的产物条带从凝胶切出,并使用**MINELUTE®Gel Extraction Kit**(QIAGEN Inc.,Valencia,CA,USA)根据生产商的指示纯化。然后将片段使用**IN-FUSION™Cloning Kit**克隆入Nco I和Pac I消化的pAlLo2,得到pSMai207,其中土生梭孢壳gh61j基因的转录处于NA2-tpi启动子(一种来自编码黑曲霉中性α-淀粉酶的基因的启动子,其中未翻译的先导序列由来自编码构巢曲霉中的丙糖磷酸异构酶的未翻译的先导序列所取代)的调控之下。连接反应(50μl)包含1X **IN-FUSION™Buffer**(BD Biosciences,Palo Alto,CA,USA),1X BSA(BD Biosciences,Palo Alto,CA,USA),1μl of **IN-FUSION™酶**(1:10稀释)(BD Biosciences,Palo Alto,CA,USA),100ng用Nco I and Pac I消化的pAlLo2,和50ng的土生梭孢壳gh61j纯化的PCR产物。将反应在室温温育30分钟。使用一μl的反应物转化大肠杆菌XL10**SOLOPACK®Gold Supercompetent**细胞(Stratagene,La Jolla,CA,USA)。含有pSMai207的大肠杆菌转化体通过限制性消化检测,而质粒DNA使用**BIOROBOT®9600**(QIAGEN Inc.,Valencia,CA,USA)制备。pSMai207中的土生梭孢壳gh61j插入通过DNA测序来确认。

[0445] 亦将相同的908bp土生梭孢壳gh61j PCR片段使用**TOPO®TA CLONING®Kit**(Invitrogen,Carlsbad,CA,USA)克隆入**pCR®2.1-TOPO**载体(Invitrogen,Carlsbad,CA,USA),以生成pSMai216。土生梭孢壳gh61j插入通过DNA测序确认。大肠杆菌pSMai216于2009年8月3日保藏于农业研究机构专利培养物保藏中心北区研究中心(Agricultural Research Service Patent Culture Collection,Northern Regional Research Center,Peoria,IL,USA),并分配登录号NRRL B-50301。

[0446] 实施例4:编码具有纤维素分解增强活性的GH61J多肽的土生梭孢壳NRRL 8126基因组序列的表征

[0447] 土生梭孢壳NRRL 8126 gh61j基因组克隆的DNA测序用Applied Biosystems Model 3700 Automated DNA Sequencer使用版本3.1BIG-DYE™终止子化学(Applied Biosystems,Inc.,Foster City,CA,USA)和dGTP化学(Applied Biosystems,Inc.,Foster City,CA,USA)和引物步移策略进行。就品质审视核苷酸序列数据,并以PHRED/PHRAP软件(University of Washington,Seattle,WA,USA)的协助将所有序列互相比较。获得的序列与来自JGI的序列相同。

[0448] 土生梭孢壳gh61j基因的核苷酸序列(SEQ ID NO:1)和推导的氨基酸序列(SEQ ID NO:2)示于图1。编码序列为878bp,包含终止密码子,并由66和71bp的内含子间隔。编码的预

测的蛋白为246个氨基酸。基因编码序列(包含内含子)的%G+C为63%G+C,而成熟多肽编码序列为63%。使用SignalP程序(Nielsen等,1997,Protein Engineering 10:1-6),预测了17个残基的信号肽。预测的成熟蛋白含有229个氨基酸,具有24.5kDa的预测的分子量和7.85的等电点pH。

[0449] 如EMBOSS的Needle程序中,以缺口开放罚分为10,缺口延伸罚分为0.5,和EBLOSUM62矩阵执行的,使用Needleman-Wunsch算法(Needleman和Wunsch,1970,J.Mol.Biol.48:443-453)确定了氨基酸序列的比较性逐对全局比对。比对显示编码具有纤维素分解增强活性的GH61J多肽的土生梭孢壳基因的推导的氨基酸序列与来自特异腐质霉的预测的GH61家族蛋白的推导的氨基酸序列(登录号geneseqp:ADM97935)分享57.7%的同一性(排除间隔)。

[0450] 实施例5:在米曲霉JaL355中表达土生梭孢壳NRRL 8126家族61糖基水解酶61j基因

[0451] 米曲霉JaL355(WO 2002/40694)原生质体根据Christensen等.,1988,Bio/Technology 6:1419-1422的方法制备。将大约2 μ g的pSMai207转化入米曲霉JaL355。

[0452] 用pSMai207转化米曲霉JaL355产生约30个转化体。将十个转化体分离至单个基本培养基平板。

[0453] 用5ml的0.01% **TWEEN**[®]20洗涤每个转化体的汇合的基本培养基平板,并分别接种入125ml玻璃摇瓶中的25ml的M410培养基,并在34 $^{\circ}$ C,250rpm温育。在5日温育之后,用**CRITERION**[®]Cell(Bio-Rad Laboratories,Hercules,CA,USA)在**CRITERION**[®]Tris-HCl凝胶(Bio-Rad Laboratories,Hercules,CA,USA)上根据生产商的指示分析5 μ l的来自每个培养的上清。将所得的凝胶用BIO-SAFE[™]Coomassie Stain(Bio-Rad Laboratories,Hercules,CA,USA)染色。培养物的SDS-PAGE概貌显示多数转化体具有预期的24KDA条带大小。用10ml的0.01% **TWEEN**[®]80洗涤转化体3的汇合的平板,并接种入含有500ml的M410培养基的2升Fernbach以生成用于表征酶的培养液。在第5日收获培养物,并使用0.22 μ m EXPRESS[™]PLUS Membrane(Millipore,Billerica,MA,USA)过滤。

[0454] 实施例6:含有编码具有纤维素分解增强活性的家族GH61K多肽的土生梭孢壳NRRL 8126基因组序列的米曲霉表达载体的构建

[0455] 设计了如下所示的两种合成的寡核苷酸引物以从实施例2中制备的基因组DNA来PCR扩增土生梭孢壳NRRL 8126 gh61k基因。使用IN-FUSION[™]Cloning Kit以将片段直接克隆入表达载体pAllo2,而无需进行限制性消化和连接。

[0456] Ttgh1k-F(065465):

[0457] 5'-actggatttacc**ATGAGGACGACATTCGCCGCCGCGT**-3'(SEQ ID NO:25)

[0458] Ttgh61k-R(065466):

[0459] 5'-TCACCTCTAGTTAATTA**ACTAAGAAGAAGGGGCGCACT**-3'(SEQ ID NO:26)

[0460] 粗体字母代表编码序列。剩余序列同源于pAllo2的插入位点。

[0461] 将五十皮摩尔的上述每种引物用于PCR反应,所述反应包含100ng的土生梭孢壳NRRL 8126基因组DNA,Pfx Amplification Buffer,各0.4mM的dATP,dTTP,dGTP,和dCTP,1mM MgCl₂,和2.5单位的Pfx DNA聚合酶,最终体积为50 μ l。使用EPPENDORF[®] MASTERCYCLER[®] 5333进行扩增,其程序为:1个循环,在98 $^{\circ}$ C进行3分钟;和30个循环,每循环在98 $^{\circ}$ C进行30

秒,60℃进行30秒,和72℃进行1.5分钟。然后加热块进入4℃浸没循环。

[0462] 将反应产物在1.0%琼脂糖凝胶上使用TAE缓冲液分离,其中将1283bp的产物条带从凝胶切出,并使用MINELUTE®Gel Extraction Kit根据生产商的指示纯化。然后将片段使用IN-FUSION™Cloning Kit克隆入Nco I和Pac I消化的pAllLo2,得到pSMai208,其中土生梭孢壳gh61k基因的转录处于NA2-tpi启动子的调控之下。连接反应(50μl)包含1X IN-FUSION™Buffer,1X BSA,1μl of IN-FUSION™酶(1:10稀释),100ng用Nco I and Pac I消化的pAllLo2,和50ng的土生梭孢壳gh61k纯化的PCR产物。将反应在室温温育30分钟。使用一μl的反应物转化大肠杆菌XL10SOLOPACK®Gold Supercompetent细胞。含有pSMai208的大肠杆菌转化体通过限制性消化检测,而质粒DNA使用BIOROBOT®9600制备。pSMai208中的土生梭孢壳gh61k插入通过DNA测序来确认。

[0463] 亦将相同的1283bp土生梭孢壳gh61k PCR片段使用TOPO TA CLONING®Kit克隆入pCR®2.1-TOPO载体,以生成pSMai217。土生梭孢壳gh61k插入通过DNA测序确认。大肠杆菌pSMai217于2009年8月3日保藏于农业研究机构专利培养物保藏中心北区研究中心(Agricultural Research Service Patent Culture Collection,Northern Regional Research Center,Peoria,IL,USA),并分配登录号NRRL B-50302。

[0464] 实施例7:编码具有纤维素分解增强活性的GH61K多肽的土生梭孢壳NRRL 8126基因组序列的表征

[0465] 土生梭孢壳NRRL 8126 gh61k基因组克隆的DNA测序用Applied Biosystems Model 3700 Automated DNA Sequencer使用版本3.1BIG-DYE™终止子化学和dGTP化学和引物步移策略进行。就品质审视核苷酸序列数据,并以PHRED/PHRAP软件的协助将所有序列互相比对。获得的序列与来自JGI的序列相同。

[0466] 土生梭孢壳gh61k基因的核苷酸序列(SEQ ID NO:3)和推导的氨基酸序列(SEQ ID NO:4)示于图2。编码序列为1253bp,包含终止密码子,并由96,84,和68bp的内含子间隔。编码的预测的蛋白为334个氨基酸。基因编码序列(包含内含子)的%G+C为66.6%G+C,而成熟多肽编码序列为69.3%。使用SignalP程序(Nielsen等,1997,见上),预测了19个残基的信号肽。预测的成熟蛋白含有315个氨基酸,具有31.7kDa的预测的分子量和6.68的等电点pH。

[0467] 如EMBOSS的Needle程序中,以缺口开放罚分为10,缺口延伸罚分为0.5,和EBLOSUM62矩阵执行的,使用Needleman-Wunsch算法(Needleman和Wunsch,1970,见上)确定了氨基酸序列的比较性逐对全局比对。比对显示编码具有纤维素分解增强活性的GH61K多肽的土生梭孢壳基因的推导的氨基酸序列与来自巴西青霉的预测的β-葡糖苷酶蛋白的推导的氨基酸序列(登录号geneseq AWW27060)分享64.8%的同一性(排除间隔)。

[0468] 实施例8:在米曲霉JaL355中表达土生梭孢壳NRRL 8126家族61糖基水解酶61k基因

[0469] 米曲霉JaL355(WO 2002/40694)原生质体根据Christensen等.,1988,见上的方法制备。将其用大约2μg的pSMai208转化。转化产生约25个转化体。将十个转化体分离至单个基本培养基平板。

[0470] 用5ml的0.01%TWEEN®20洗涤每个转化体的汇合的基本培养基平板,并分别接种入125ml玻璃摇瓶中的25ml的M410培养基,并在34℃,250rpm温育。在5日温育之后,用CRITERION®Cell在CRITERION®Tris-HCl凝胶上根据生产商的指示分析5μl的来

自每个培养的上清。将所得的凝胶用BIO-SAFE™Coomassie Stain染色。培养物的SDS-PAGE概貌显示多数转化体具有预期的32KDA条带大小。用10ml的0.01%TWEEN®80洗涤转化体5的汇合的平板,并接种入含有500ml的M410培养基的2升Fernbach以生成用于表征酶的培养液。在第5日收获培养物,并使用.22μm EXPRESS™PLUS Membrane过滤。

[0471] 实施例9:含有编码具有纤维素分解增强活性的家族GH61L多肽的土生梭孢壳NRRL 8126基因组序列的米曲霉表达载体的构建

[0472] 设计了如下所示的两种合成的寡核苷酸引物以从实施例2中制备的基因组DNA来PCR扩增土生梭孢壳NRRL 8126 gh611基因。使用IN-FUSION™Cloning Kit以将片段直接克隆入表达载体pA1Lo2,而无需进行限制性消化和连接。

[0473] Ttgh11-F1(066276):

[0474] 5'-actggatttacc**ATGAAGCTGAGCGTTGCCATCGCC**-3'(SEQ ID NO:27)

[0475] Ttgh611-R(065736):

[0476] 5'-TCACCTCTAGTTAATTA**ATTAGCACGTCTCAGCCGGCG**-3'(SEQ ID NO:28)

[0477] 粗体字母代表编码序列。剩余序列同源于pA1Lo2的插入位点。

[0478] 将五十皮摩尔的上述每种引物用于PCR反应,所述反应包含100ng的土生梭孢壳NRRL 8126基因组DNA,Pfx Amplification Buffer,各0.4mM的dATP,dTTP,dGTP,和dCTP,1mM MgCl₂,和2.5单位的Pfx DNA聚合酶,最终体积为50μl。使用EPPENDORF® MASTERCYCLER® 5333进行扩增,其程序为:1个循环,在98℃进行3分钟;和30个循环,每循环在98℃进行30秒,60℃进行30秒,和72℃进行1.5分钟。然后加热块进入4℃浸没循环。

[0479] 将反应产物在1.0%琼脂糖凝胶上使用TAE缓冲液分离,其中将828bp的产物条带从凝胶切出,并使用MINELUTE®Gel Extraction Kit根据生产商的指示纯化。然后将片段使用IN-FUSION™Cloning Kit克隆入Nco I和Pac I消化的pA1Lo2,得到pSMai209,其中土生梭孢壳gh611基因的转录处于NA2-tpi启动子的调控之下。连接反应(50μl)包含1X IN-FUSION™Buffer,1X BSA,1μl of IN-FUSION™酶(1:10稀释),100ng用Nco I and Pac I消化的pA1Lo2,和50ng的土生梭孢壳gh611纯化的PCR产物。将反应在室温温育30分钟。使用一μl的反应物转化大肠杆菌XL10SOLOPACK®Gold Supercompetent细胞。含有pSMai212的大肠杆菌转化体通过限制性消化检测,而质粒DNA使用BIOROBOT®9600制备。pSMai212中的土生梭孢壳gh611插入通过DNA测序来确认。

[0480] 亦将相同的828bp土生梭孢壳gh611 PCR片段使用TOPO TA CLONING®Kit克隆入pCR®2.1-TOPO载体,以生成pSMai218。土生梭孢壳gh611插入通过DNA测序确认。大肠杆菌pSMai218于2009年8月3日保藏于农业研究机构专利培养物保藏中心北区研究中心(Agricultural Research Service Patent Culture Collection,Northern Regional Research Center,Peoria,IL,USA),并分配登录号NRRL B-50303。

[0481] 实施例10:编码具有纤维素分解增强活性的GH61L多肽的土生梭孢壳NRRL 8126基因组序列的表征

[0482] 土生梭孢壳NRRL 8126 gh611基因组克隆的DNA测序用Applied Biosystems Model 3700 Automated DNA Sequencer使用版本3.1BIG-DYE™终止子化学和dGTP化学和引物步移策略进行。就品质审视核苷酸序列数据,并以PHRED/PHRAP软件的协助将所有序列互相比较。获得的序列与来自JGI的序列相同。

[0483] 土生梭孢壳gh611基因的核苷酸序列(SEQ ID NO:5)和推导的氨基酸序列(SEQ ID NO:6)示于图3。编码序列为798bp,包含终止密码子,并由55和59bp的内含子间隔。编码的预测的蛋白为227个氨基酸。基因编码序列(包含内含子)的%G+C为60.8%G+C,而成熟多肽编码序列为62.6%。使用SignalP程序(Nielsen等,1997,见上),预测了17个残基的信号肽。预测的成熟蛋白含有210个氨基酸,具有22.6kDa的预测的分子量和8.84的等电点pH。

[0484] 如EMBOSS的Needle程序中,以缺口开放罚分为10,缺口延伸罚分为0.5,和EBLOSUM62矩阵执行的,使用Needleman-Wunsch算法(Needleman和Wunsch,1970,见上)确定了氨基酸序列的比较性逐对全局比对。比对显示编码具有纤维素分解增强活性的GH61L多肽的土生梭孢壳基因的推导的氨基酸序列与来自土生梭孢霉的预测的 β -葡糖苷酶蛋白的推导的氨基酸序列(登录号geneseq ADM97933)分享59.2%的同一性(排除间隔)。

[0485] 实施例11:在米曲霉JaL355中表达土生梭孢壳NRRL 8126家族61糖基水解酶611基因

[0486] 米曲霉JaL355(WO 2002/40694)原生质体根据Christensen等.,1988,见上的方法制备。将其用大约2 μ g的pSMai212转化。转化产生约17个转化体。将十七个转化体分离至单个基本培养基平板。

[0487] 用5ml的0.01% TWEEN®20洗涤每个转化体的汇合的基本培养基平板,并分别接种入125ml玻璃摇瓶中的25ml的M410培养基,并在34°C,250rpm温育。在5日温育之后,用CRITERION®Cell在CRITERION®Tris-HCl凝胶上根据生产商的指示分析5 μ l的来自每个培养的上清。将所得的凝胶用BIO-SAFE™Coomassie Stain染色。培养物的SDS-PAGE概貌显示多数转化体具有预期的23KDA条带大小。用10ml的0.01% TWEEN®80洗涤转化体14的汇合的平板,并接种入含有500ml的M410培养基的2升Fernbach以生成用于表征酶的培养液。在第5日收获培养物,并使用.22 μ m EXPRESS™PLUS Membrane过滤。

[0488] 实施例12:经预处理的玉米秸秆的水解由具有纤维素分解增强活性的土生梭孢壳NRRL 8126 GH61J,GH61K,和GFH61L多肽增强

[0489] 培养液如实施例5,8,和11中所述制备,并使用Amicon超滤装置(Millipore, Bedford, MA, USA, 10kDa聚醚砜膜,40psi,4°C)浓缩大约20倍。在SDS-PAGE和考马斯蓝染色之后通过光密度法估计蛋白浓度。如WO 2005/074647中所述预处理玉米秸秆并将其作为测定底物制备,以生成经预处理玉米秸秆(PCS)。用于测定增强活性的基础(base)纤维素酶混合物从里氏木霉菌株SMA135(WO 2008/057637)制备。

[0490] PCS的水解使用1.6ml深孔板(Axygen,Santa Clara,CA,USA)使用1.0ml的总反应体积和1mM硫酸锰-50mM乙酸钠pH 5.0中的50mg/ml的PCS浓度来进行。将土生梭孢壳多肽(GH61J,GH61K和GFH61L)以0至25%或0至32%浓度范围的基础纤维素酶混合物的蛋白浓度分别添加至基础纤维素酶混合物。温育在50°C进行72小时。测定重复进行三次。离心等分试样,并通过使用平板离心机(SORVALL®RT7,Thermo Fisher Scientific,Waltham,MA, USA)以3000rpm离心10分钟(MULTISCREEN®HV 0.45 μ m,Millipore,Billerica,MA, USA)来过滤上清液体。当不立即使用时,将过滤的水解物等分试样冻结于-20°C。稀释于含0.05%w/w苯甲酸的0.005M H₂SO₄中的样品的糖浓度在用含0.05%w/w苯甲酸的0.005M H₂SO₄在60°C以0.6ml/分钟的流速从4.6x250mm AMINEX®HPX-87H柱(Bio-Rad Laboratories,Inc.,Hercules,CA,USA)洗脱之后测量,其中通过将纯糖样品(Absolute

Standards Inc., Hamden, CT, USA)校准的折光率检测(CHEMSTATION®, AGILENT®1100 HPLC, Agilent Technologies, Santa Clara, CA, USA)所得的葡萄糖和纤维二糖信号进行积分来进行定量。所得的当量用于计算对于每个反应的纤维素转化百分比。纤维素转化为葡萄糖加上纤维二糖的程度(转化,%)使用下式计算:

$$[0491] \quad \text{转化}(\%) = (\text{葡萄糖} + \text{纤维二糖} \times 1.053)(\text{mg/ml}) \times 100 \times 162 / (\text{纤维素}(\text{mg/ml}) \times 180) \\ = (\text{葡萄糖} + \text{纤维二糖} \times 1.053)(\text{mg/ml}) \times 100 / (\text{纤维素}(\text{mg/ml}) \times 1.111)$$

[0492] 在该式中,因子1.111反映了在纤维素转化为葡萄糖中的重量增加,而因子1.053反映了在纤维二糖转化为葡萄糖中的重量增加。PCS中的纤维素通过PCS的限制消化以释放葡萄糖和纤维二糖来确定。

[0493] 将增加量的土生梭孢壳多肽分别添加至基础纤维素酶混合物的结果示于图4。土生梭孢壳GH61J和GH61K在25%添加水平的添加分别提供了1.14和1.13的刺激因子。土生梭孢壳GH61L多肽在32%添加水平提供了1.13的刺激因子。

[0494] 实施例13:含有编码具有纤维素分解增强活性的家族GH61M多肽的土生梭孢壳NRRL 8126基因组序列的米曲霉表达载体的构建

[0495] 设计了如下所示的两种合成的寡核苷酸引物以从实施例2中制备的基因组DNA来PCR扩增土生梭孢壳NRRL 8126 gh61m基因。使用IN-FUSION™Cloning Kit以将片段直接克隆入表达载体pAllo2(WO 2004/099228),而无需进行限制性消化和连接。

[0496] Ttgh1m-F1(063567):

[0497] 5'-actggatttacc**ATGAAGCTGTCATCCCAGCTCGCC**-3'(SEQ ID NO:29)

[0498] Ttgh61m-R1(063568):

[0499] 5'-TCACCTCTAGTTAATTA**ACTAGCACTGAAAGACCGCCG**-3'(SEQ ID NO:30)

[0500] 粗体字母代表编码序列。剩余序列同源于pAllo2的插入位点。

[0501] 将五十皮摩尔的上述每种引物用于PCR反应,所述反应包含100ng的土生梭孢壳NRRL 8126基因组DNA,Pfx Amplification Buffer,各0.4mM的dATP,dTTP,dGTP,和dCTP,1mM MgCl₂,和2.5单位的Pfx DNA聚合酶,最终体积为50μl。使用EPENDORF® MASTERCYCLER® 5333进行扩增,其程序为:1个循环,在98℃进行3分钟;和30个循环,每循环在98℃进行30秒,60℃进行30秒,和72℃进行1.5分钟。然后加热块进入4℃浸没循环。

[0502] 将反应产物在1.0%琼脂糖凝胶上使用TAE缓冲液分离,其中将1007bp的产物条带从凝胶切出,并使用MINELUTE®Gel Extraction Kit根据生产商的指示纯化。然后将片段使用IN-FUSION™Cloning Kit克隆入Nco I和Pac I消化的pAllo2,得到pSMai197(图5),其中土生梭孢壳gh61m基因的转录处于NA2-tpi启动子的调控之下。连接反应(50μl)包含1X IN-FUSION™Buffer,1X BSA,1μl of IN-FUSION™酶(1:10稀释),100ng用Nco I and Pac I消化的pAllo2,和50ng的土生梭孢壳gh61m纯化的PCR产物。将反应在室温温育30分钟。使用一μl的反应物转化大肠杆菌XL10SOLOPACK®Gold Supercompetent细胞。含有pSMai197的大肠杆菌转化体通过限制性消化检测,而质粒DNA使用BIOROBOT®9600制备。pSMai197中的土生梭孢壳gh61m插入通过DNA测序来确认。

[0503] 亦将相同的1007bp土生梭孢壳gh61m PCR片段使用TOPO TA CLONING®Kit克隆入pCR®2.1-TOPO载体,以生成pSMai213。土生梭孢壳gh61m插入通过DNA测序确认。大肠杆菌pSMai213于2009年8月3日保藏于农业研究机构专利培养物保藏中心北区研究中心

(Agricultural Research Service Patent Culture Collection, Northern Regional Research Center, Peoria, IL, USA), 并分配登录号NRRL B-50300。

[0504] 实施例14: 编码具有纤维素分解增强活性的GH61M多肽的土生梭孢壳NRRL 8126基因组序列的表征

[0505] 土生梭孢壳NRRL 8126 gh61m基因组克隆的DNA测序用Applied Biosystems Model 3700 Automated DNA Sequencer使用版本3.1BIG-DYE™终止子化学和dGTP化学和引物步移策略进行。就品质审视核苷酸序列数据, 并以PHRED/PHRAP软件的协助将所有序列互相比较。获得的序列与来自JGI的序列相同。

[0506] 土生梭孢壳gh61m基因的核苷酸序列(SEQ ID NO:7)和推导的氨基酸序列(SEQ ID NO:8))示于图6。编码序列为977bp, 包含终止密码子, 并由85, 96, 和124bp的内含子间隔。编码的预测的蛋白为223个氨基酸。基因编码序列(包含内含子)的%G+C为62.6%G+C, 而成熟多肽编码序列为62.2%。使用SignalP程序(Nielsen等, 1997, 见上), 预测了19个残基的信号肽。预测的成熟蛋白含有204个氨基酸, 具有22.2kDa的预测的分子量和6.58的等电点pH。

[0507] 如EMBOSS的Needle程序中, 以缺口开放罚分为10, 缺口延伸罚分为0.5, 和EBLOSUM62矩阵执行的, 使用Needleman-Wunsch算法(Needleman和Wunsch, 1970, 见上)确定了氨基酸序列的比较性逐对全局比对。比对显示编码具有纤维素分解增强活性的GH61M多肽的土生梭孢壳基因的推导的氨基酸序列与来自Podospora anserina的预测的β-葡萄糖苷酶蛋白的推导的氨基酸序列(登录号UniProt B2ADY5)分享76.5%的同一性(排除间隔)。

[0508] 实施例15: 在米曲霉JaL355中表达土生梭孢壳NRRL 8126家族61糖基水解酶61m基因

[0509] 米曲霉JaL355(WO 2002/40694)原生质体根据Christensen等., 1988, 见上的方法制备。将其用大约2μg的pSMai197转化。转化产生约17个转化体。将十个转化体分离至单个基本培养基平板。

[0510] 用5ml的0.01% TWEEN®20洗涤每个转化体的汇合的基本培养基平板, 并分别接种入125ml玻璃摇瓶中的25ml的M410培养基, 并在34°C, 250rpm温育。在5日温育之后, 用CRITERION®Cell在CRITERION®Tris-HCl凝胶上根据生产商的指示分析5μl的来自每个培养的上清。将所得的凝胶用BIO-SAFE™Coomassie Stain染色。培养物的SDS-PAGE概貌显示多数转化体具有预期的22KDA条带大小。用10ml的0.01% TWEEN®80洗涤转化体9的汇合的平板, 并接种入含有500ml的M410培养基的2升Fernbach以生成用于表征酶的培养液。在第5日收获培养物, 并使用.22μm EXPRESS™PLUS Membrane过滤。

[0511] 实施例16: 经预处理的玉米秸秆的水解由具有纤维素分解增强活性的土生梭孢壳NRRL 8126 GH61M多肽增强

[0512] 培养液如实施例15中所述制备, 并使用Amicon超滤装置(Millipore, Bedford, MA, USA, 10kDa聚醚砜膜, 40psi, 4°C)浓缩大约20倍。在SDS-PAGE和考马斯蓝染色之后通过光密度法估计蛋白浓度。如WO 2005/074647中所述预处理玉米秸秆并将其作为测定底物制备, 以生成经预处理玉米秸秆(PCS)。用于测定增强活性的基础纤维素酶混合物从里氏木霉菌株SMA135(WO 2008/057637)制备。

[0513] PCS的水解使用1.6ml深孔板(Axygen, Santa Clara, CA, USA)使用1.0ml的总反应体积和1mM硫酸锰-50mM乙酸钠pH 5.0中的50mg/ml的PCS浓度来进行。将土生梭孢壳多肽

(GH61M)以0至25%浓度范围的基础纤维素酶混合物的蛋白浓度分别添加至基础纤维素酶混合物。温育在50℃进行72小时。测定重复进行三次。离心等分试样,并通过使用平板离心机(SORVALL®RT7,Thermo Fisher Scientific,Waltham,MA,USA)以3000rpm离心10分钟(MULTISCREEN®HV 0.45µm)来过滤上清液体。当不立即使用时,将过滤的水解物等分试样冻结于-20℃。稀释于含0.05%w/w苯甲酸的0.005M H₂SO₄中的样品的糖浓度在用含0.05%w/w苯甲酸的0.005M H₂SO₄在60℃以0.6ml/分钟的流速从4.6x250mmAMINEX®HPX-87H柱洗脱之后测量,其中通过将纯糖样品(Absolute Standards Inc.,Hamden,CT,USA)校准的折光率检测所得的葡萄糖和纤维二糖信号进行积分来进行定量。所得的当量用于计算对于每个反应的纤维素转化百分比。纤维素转化为葡萄糖加上纤维二糖的程度(转化,%)使用下式计算:

[0514] 转化(%)=(葡萄糖+纤维二糖x1.053)(mg/ml)x100x162/(纤维素(mg/ml)x180)
=(葡萄糖+纤维二糖x1.053)(mg/ml)x100/(纤维素(mg/ml)x1.111)

[0515] 在该式中,因子1.111反映了在纤维素转化为葡萄糖中的重量增加,而因子1.053反映了在纤维二糖转化为葡萄糖中的重量增加。PCS中的纤维素通过PCS的限制消化以释放葡萄糖和纤维二糖来确定。

[0516] 将增加量的土生梭孢壳GH61M多肽添加至基础纤维素酶混合物的结果示于图7。土生梭孢壳GH61M多肽在25%添加水平的添加提供了1.27的刺激因子。

[0517] 实施例17:含有用于土生梭孢壳家族GH61N基因的米曲霉表达载体的构建

[0518] 设计了如下所示的两种合成的寡核苷酸引物以从实施例2中制备的基因组DNA来PCR扩增土生梭孢壳家族GH61N基因。使用IN-FUSION™Cloning Kit以将片段直接克隆入表达载体pAllLo2,而无需进行限制性消化和连接。

[0519] 正向引物:

[0520] 5'-ACTGGATTTACC**ATGCCTTCTTTGCGCTCCAA**-3'(SEQ ID NO:31)

[0521] 反向引物:

[0522] 5'-TCACCTCTAGTTAATTAAT**TCAGTTTGCCTCCTCAGCCC**-3'(SEQ ID NO:32)

[0523] 粗体字母代表编码序列。剩余序列同源于pAllLo2的插入位点。

[0524] 将五十皮摩尔的上述每种引物用于PCR反应,所述反应含有1µg的土生梭孢壳基因组DNA,1X**ADVANTAGE**®GC-Melt LA Buffer(BD Biosciences,Palo Alto,CA,USA),1µl的dATP,dTTP,dGTP,和dCTP的10mM混合物,和1.25单位的**ADVANTAGE**®GC Genomic LA Polymerase Mix(BD Biosciences,Palo Alto,CA,USA),最终体积为25µl。扩增条件为:一个循环,在94℃进行1分钟;和30个循环,每循环在94℃进行30秒,60.5℃进行30秒,和72℃进行1分钟。然后将加热块维持在72℃进行5分钟,接着进行4℃浸没循环。

[0525] 将反应产物通过使用TAE缓冲液的1.0%琼脂糖凝胶电泳分离,其中将大约1.1kb的产物条带从凝胶切出,并使用**MINELUTE**®Gel Extraction Kit根据生产商的指示纯化。

[0526] 然后将片段使用IN-FUSION™Cloning Kit克隆入pAllLo2。用Nco I和Pac I消化载体。将片段通过使用TAE缓冲液的1.0%琼脂糖凝胶电泳纯化,从凝胶切出,并使用**QIAQUICK**®Gel Extraction Kit(QIAGEN Inc.,Valencia,CA,USA)纯化。在反应中将基因片段和消化的载体合并在一起,得到表达质粒pAG66,其中家族GH61N基因的转录处于

NA2-tpi启动子的调控之下。重组反应物(20 μ l)包含1X IN-FUSIONTMBuffer, 1X BSA, 1 μ l的IN-FUSIONTM酶(1:10稀释), 186ng用Nco I和Pac I消化的pAllo2, 和96.6ng的土生梭孢壳GH61N纯化PCR产物。反应在37 $^{\circ}$ C温育15分钟, 然后在50 $^{\circ}$ C温育15分钟用40 μ l的TE缓冲液稀释反应物, 并使用2.5 μ l的稀释反应物转化大肠杆菌Top10 Competent细胞(Stratagene, La Jolla, CA, USA)。通过限制酶消化鉴定了含有pAG66(GH61N基因)的大肠杆菌转化体, 并使用BIOROBOT[®]9600制备质粒DNA。

[0527] 亦使用TOPO[®]TACLONING[®]Kit将相同的1.1kb土生梭孢壳gh61n PCR片段克隆入pCR[®]2.1-TOPO载体, 以生成pAG68。通过DNA测序确认了土生梭孢壳gh61n插入。大肠杆菌pAG68于2009年9月18日保藏于农业研究机构专利培养物保藏中心北区研究中心(Agricultural Research Service Patent Culture Collection, Northern Regional Research Center, Peoria, IL, USA), 并分配登录号NRRL B-50320。

[0528] 实施例18: 编码具有纤维素分解增强活性的家族GH61N多肽的土生梭孢壳基因组序列的表征

[0529] 具有纤维素分解增强活性的土生梭孢壳GH61N多肽的核苷酸序列(SEQ ID NO:9)和推导的氨基酸序列(SEQ ID NO:10)示于图8。基因组多核苷酸为1107bp, 包含终止密码子, 并编码368个氨基酸的多肽。全长编码序列和成熟编码序列的%G+C分别为68.1%和68.3%。使用SignalP软件程序(Nielsen等, 1997, 见上), 预测了21个残基的信号肽。预测的成熟的蛋白含有347个氨基酸, 具有35.0kDa的分子量。

[0530] 用Interproscan程序(Mulder等, 2007, Nucleic Acids Res. 35:D224-D228)分析具有纤维素分解增强活性的GH61N多肽的推导的氨基酸序列显示所述GH61N多肽含有糖基水解酶家族61的特征序列(InterPro登录号IPR005103)。该特征序列见于成熟多肽(Pfam登录号PF03443)的大约残基1至221。

[0531] 如EMBOSS的Needle程序中, 以缺口开放罚分为10, 缺口延伸罚分为0.5, 和EBLOSUM62矩阵执行的, 使用Needleman-Wunsch算法(Needleman和Wunsch, 1970, 见上)确定了氨基酸序列的比较性逐对全局比对。比对显示土生梭孢壳GH61N多肽的推导的氨基酸序列与来自黑曲霉的另一种预测的家族61糖基水解酶的推导的氨基酸序列(UniProt登录号A2QZE1)分享72.3%的同一性(排除间隔)。

[0532] 实施例19: 在米曲霉JaL355中表达编码具有纤维素分解增强活性的GH61N多肽的土生梭孢壳基因组DNA。

[0533] 根据Christensen等, 1988, 见上的方法制备米曲霉JaL355原生质体, 将其用5 μ g的pAG43转化。将三个转化体分离至单独的PDA平板。

[0534] 从三个转化体每一个的起始转化平板取栓, 并分别添加至24孔板中的1ml的M410培养基, 将其在34 $^{\circ}$ C温育。在温育之后五日, 使用CRITERION[®]无染色, 8-16%梯度SDS-PAGE, (Bio-Rad Laboratories, Inc., Hercules, CA, USA)根据生产商的指示分析来自每个培养的7.5 μ l的上清。培养物的SDS-PAGE概貌显示几个转化体具有大约70kDa和35kDa的新的主要条带。

[0535] 用5ml的0.01% TWEEN[®]20洗涤两个转化体的汇合的PDA平板, 并接种入五个含有100ml的M410培养基的500ml Erlenmeyer烧瓶, 并温育以生成用于表征酶的培养液。在第3日和第5日收获烧瓶, 并使用0.22 μ m stericup吸滤器(Millipore, Bedford, MA, USA)过滤。

[0536] 实施例20:经预处理的玉米秸秆的水解由具有纤维素分解增强活性的土生梭孢壳GH61N多肽增强

[0537] 培养液如实施例19中所述制备,并使用Amicon超滤装置(Millipore, Bedford, MA, USA, 10kDa聚醚砜膜, 40psi, 4°C)浓缩大约20倍。在SDS-PAGE和考马斯蓝染色之后通过光密度法估计蛋白浓度。如WO 2005/074647中所述预处理玉米秸秆并将其作为测定底物制备,以生成经预处理玉米秸秆(PCS)。用于测定增强活性的基础纤维素酶混合物从里氏木霉菌株SMA135(WO 2008/057637)制备。

[0538] PCS的水解使用1.6ml深孔板(Axygen, Santa Clara, CA)使用1.0ml的总反应体积和1mM硫酸锰-50mM乙酸钠pH 5.0中的50mg/ml的PCS浓度来进行。将土生梭孢壳多肽(GH61N)分别以0至100%浓度范围的基础纤维素酶混合物的蛋白浓度分别添加至基础纤维素酶混合物。温育在50°C进行72小时。测定重复进行三次。离心等分试样,并通过使用平板离心机(SORVALL®RT7, Thermo Fisher Scientific, Waltham, MA, USA)以3000rpm离心10分钟(MULTISCREEN®HV 0.45µm, Millipore, Billerica, MA, USA)来过滤上清液体。当不立即使用时,将过滤的水解物等分试样冻结于-20°C。稀释于含0.05%w/w苯甲酸的0.005M H₂SO₄中的样品的糖浓度在用含0.05%w/w苯甲酸的0.005M H₂SO₄在60°C以0.6ml/分钟的流速从4.6x250mm AMINEX®HPX-87H柱(Bio-Rad Laboratories, Inc., Hercules, CA, USA)洗脱之后测量,其中通过将纯糖样品(Absolute Standards Inc., Hamden, CT, USA)校准的折光率检测(CHEMSTATION®, AGILENT®1100 HPLC, Agilent Technologies, Santa Clara, CA, USA)所得的葡萄糖和纤维二糖信号进行积分来进行定量。所得的当量用于计算对于每个反应的纤维素转化百分比。纤维素转化为葡萄糖加上纤维二糖的程度(转化,%)使用下式计算:

[0539] 转化(%)=(葡萄糖+纤维二糖x1.053)(mg/ml)x100x162/(纤维素(mg/ml)x180)
=(葡萄糖+纤维二糖x1.053)(mg/ml)x100/(纤维素(mg/ml)x1.111)

[0540] 在该式中,因子1.111反映了在纤维素转化为葡萄糖中的重量增加,而因子1.053反映了在纤维二糖转化为葡萄糖中的重量增加。PCS中的纤维素通过PCS的限制消化以释放葡萄糖和纤维二糖来确定。

[0541] 将增加量的土生梭孢壳多肽分别添加至基础纤维素酶混合物的结果示于图9。土生梭孢壳GH61N在50%添加百分比的添加提供了1.30的最大刺激优势(maximum stimulatory benefit)。

[0542] 实施例21:含有用于土生梭孢壳家族GH610基因的米曲霉表达载体的构建

[0543] 设计了如下所示的两种合成的寡核苷酸引物以从实施例2中制备的基因组DNA来PCR扩增土生梭孢壳家族GH610基因。使用IN-FUSION™Cloning Kit以将片段直接克隆入表达载体pAllo2,而无需进行限制性消化和连接。

[0544] 正向引物:

[0545] 5'-ACTGGATTTACC**ATGCCGCCCGCACTCCCTCA**-3'(SEQ ID NO:33)

[0546] 反向引物:

[0547] 5'-TCACCTCTAGTTAATTA**ACTAACCCCGCCGATCATA**CC-3'(SEQ ID NO:34)

[0548] 粗体字母代表编码序列。剩余序列同源於pAllo2的插入位点。

[0549] 将五十皮摩尔的上述每种引物用于PCR反应,所述反应含有1µg的土生梭孢壳基因

组DNA, 1X **ADVANTAGE**[®]GC-Melt LA Buffer, 1 μ l的dATP, dTTP, dGTP, 和dCTP的10mM混合物, 和1.25单位的**ADVANTAGE**[®]GC Genomic LA Polymerase Mix, 最终体积为25 μ l。扩增条件为: 一个循环, 在94 $^{\circ}$ C进行1分钟; 和30个循环, 每循环在94 $^{\circ}$ C进行30秒, 60.5 $^{\circ}$ C进行30秒, 和72 $^{\circ}$ C进行1分钟。然后将加热块维持在72 $^{\circ}$ C进行5分钟, 接着进行4 $^{\circ}$ C浸没循环。

[0550] 将反应产物在1.0%琼脂糖凝胶上使用TAE缓冲液分离, 其中将大约1kb的产物条带从凝胶切出, 并使用**MINELUTE**[®]Gel Extraction Kit根据生产商的指示纯化。

[0551] 然后将片段使用IN-FUSION[™]Cloning Kit克隆入pAllo2。用Nco I和Pac I消化载体。将片段通过琼脂糖凝胶电泳和**QIAQUICK**[®]Gel Extraction Kit纯化。在反应中将基因片段和消化的载体合并在一起, 得到表达质粒pAG67, 其中家族GH610基因的转录处于NA2-tpi启动子的调控之下。重组反应物(20 μ l)包含1X IN-FUSION[™]Buffer, 1X BSA, 1 μ l的IN-FUSION[™]酶(1:10稀释), 186ng用Nco I和Pac I消化的pAllo2, 和90.6ng的土生梭孢壳GH610纯化PCR产物。反应在37 $^{\circ}$ C温育15分钟, 然后在50 $^{\circ}$ C温育15分钟用40 μ l的TE缓冲液稀释反应物, 并使用2.5 μ l的稀释反应物转化大肠杆菌Top10 Competent细胞。通过限制酶消化鉴定了含有pAG67(GH610基因)的大肠杆菌转化体, 并使用**BIOROBOT**[®]9600制备质粒DNA。

[0552] 亦使用**TOPO**[®]**TACLONING**[®]Kit将相同的1kb土生梭孢壳gh61o PCR片段克隆入**pCR**[®]2.1-TOPO载体, 以生成pAG69。通过DNA测序确认了土生梭孢壳gh61o插入。大肠杆菌pAG69于2009年9月18日保藏于农业研究机构专利培养物保藏中心北区研究中心(Agricultural Research Service Patent Culture Collection, Northern Regional Research Center, Peoria, IL, USA), 并分配登录号NRRL B-50321。

[0553] 实施例22: 编码具有纤维素分解增强活性的家族GH610多肽的土生梭孢壳基因组序列的表征

[0554] 具有纤维素分解增强活性的土生梭孢壳GH610多肽的核苷酸序列(SEQ ID NO:11)和推导的氨基酸序列(SEQ ID NO:12)示于图10。基因组多核苷酸为993bp, 包含终止密码子, 并编码330个氨基酸的多肽。全长编码序列和成熟编码序列的%G+C均为69.4%。使用SignalP软件程序(Nielsen等, 1997, 见上), 预测了24个残基的信号肽。预测的成熟的蛋白含有306个氨基酸, 具有32.1kDa的分子量。

[0555] 用Interproscan程序(Mulder等, 2007, 见上)分析具有纤维素分解增强活性的GH610多肽的推导的氨基酸序列显示所述GH610多肽含有糖基水解酶家族61的特征序列(InterPro登录号IPR005103)。该特征序列见于成熟多肽(Pfam登录号PF03443)的大约残基1至245。

[0556] 如EMBOSS的Needle程序中, 以缺口开放罚分为10, 缺口延伸罚分为0.5, 和EBLOSUM62矩阵执行的, 使用Needleman-Wunsch算法(Needleman和Wunsch, 1970, 见上)确定了氨基酸序列的比较性逐对全局比对。比对显示土生梭孢壳GH610成熟多肽的推导的氨基酸序列与来自Podospira anserina的另一种预测的家族61糖基水解酶的推导的氨基酸序列(UniProt登录号B2AVC8)分享56.5%的同一性(排除间隔)。

[0557] 实施例23: 含有用于土生梭孢壳家族GH61P基因的米曲霉表达载体的构建

[0558] 设计了如下所示的两种合成的寡核苷酸引物以从实施例2中制备的基因组DNA来PCR扩增土生梭孢壳家族GH61P基因。使用IN-FUSION[™]Cloning Kit以将片段直接克隆入表

达载体pAllLo2(WO 2005/074647),而无需进行限制性消化和连接。

[0559] 正向引物:

[0560] 5'-ACTGGATTTACC**ATGAAGACATTCACCGCCCTCCTG**-3'(SEQ ID NO:35)

[0561] 反向引物:

[0562] 5'-TCACCTCTAGTTAATTAAT**TCAGCAAGTAAAGACCGCCG**-3'(SEQ ID NO:36)

[0563] 粗体字母代表编码序列。剩余序列同源于pAllLo2的插入位点。

[0564] 将五十皮摩尔的上述每种引物用于PCR反应,所述反应含有1 μ g的土生梭孢壳基因组DNA,1X**ADVANTAGE**[®]GC-Melt LA Buffer,1 μ l的dATP,dTTP,dGTP,和dCTP的10mM混合物,和1.25单位的**ADVANTAGE**[®]GC Genomic LA Polymerase Mix,最终体积为25 μ l。扩增条件为:一个循环,在94 $^{\circ}$ C进行1分钟;和30个循环,每循环在94 $^{\circ}$ C进行30秒,58.5 $^{\circ}$ C进行30秒,和72 $^{\circ}$ C进行5分钟。然后将加热块维持在72 $^{\circ}$ C进行5分钟,接着进行4 $^{\circ}$ C浸没循环。

[0565] 将反应产物在1.0%琼脂糖凝胶上使用TAE缓冲液分离,其中将大约1.2kb的产物条带从凝胶切出,并使用**MINELUTE**[®]Gel Extraction Kit根据生产商的指示纯化。

[0566] 然后将片段使用IN-FUSION[™]Cloning Kit克隆入pAllLo2。用Nco I和Pac I消化载体。将片段通过琼脂糖凝胶电泳和**QIAQUICK**[®]Gel Extraction Kit纯化。在反应中将基因片段和消化的载体合并在一起,得到表达质粒pAG70,其中家族GH61P基因的转录处于NA2-tpi启动子的调控之下。重组反应物(10 μ l)包含1X IN-FUSION[™]Buffer,1X BSA,0.5 μ l的IN-FUSION[™]酶(1:10稀释),93ng用Nco I和Pac I消化的pAllLo2,和2 μ l的土生梭孢壳GH61P纯化PCR产物。反应在37 $^{\circ}$ C温育15分钟,然后在50 $^{\circ}$ C温育15分钟用40 μ l的TE缓冲液稀释反应物,并使用2.5 μ l的稀释反应物转化大肠杆菌Top10 Competent细胞。通过限制酶消化鉴定了含有pAG70(GH61P基因)的大肠杆菌转化体,并使用**BIOROBOT**[®]9600制备质粒DNA。

[0567] 亦使用**TOPO**[®]**TACLONING**[®]Kit将相同的1.2kb土生梭孢壳gh61p PCR片段克隆入**pCR**[®]2.1-TOPO载体,以生成pAG75。通过DNA测序确认了土生梭孢壳gh61p插入。大肠杆菌pAG75于2009年9月18日保藏于农业研究机构专利培养物保藏中心北区研究中心(Agricultural Research Service Patent Culture Collection,Northern Regional Research Center,Peoria,IL,USA),并分配登录号NRRL B-50322。

[0568] 实施例24:编码具有纤维素分解增强活性的家族GH61P多肽的土生梭孢壳基因组序列的表征

[0569] 具有纤维素分解增强活性的土生梭孢壳GH61P多肽的核苷酸序列(SEQ ID NO:13)和推导的氨基酸序列(SEQ ID NO:14)示于图11。基因组多核苷酸为1221bp,包含终止密码子,且编码序列由231,75,和96bp的三个内含子间隔。预测的编码序列编码236个氨基酸的多肽。全长编码序列(包含内含子)和成熟编码序列的%G+C分别为60.2%和59.8%。使用SignalP软件程序(Nielsen等,1997,见上),预测了16个残基的信号肽。预测的成熟的蛋白含有220个氨基酸,具有23.6kDa的分子量。

[0570] 用Interproscan程序(Mulder等,2007,见上)分析具有纤维素分解增强活性的GH61P多肽的推导的氨基酸序列显示所述GH61P多肽含有糖基水解酶家族61的特征序列(InterPro登录号IPR005103)。该特征序列见于成熟多肽(Pfam登录号PF03443)的大约残基1至212。

[0571] 如EMBOSS的Needle程序中,以缺口开放罚分为10,缺口延伸罚分为0.5,和EBLOSUM62矩阵执行的,使用Needleman-Wunsch算法(Needleman和Wunsch,1970,见上)确定了氨基酸序列的比较性逐对全局比对。比对显示土生梭孢壳GH61P成熟多肽的推导的氨基酸序列与来自粗糙脉孢霉的另一种预测的家族61糖基水解酶的推导的氨基酸序列(UniProt登录号Q7SA19)分享80.3%的同一性(排除间隔)。

[0572] 实施例25:含有用于土生梭孢壳家族GH61R基因的米曲霉表达载体的构建

[0573] 设计了如下所示的两种合成的寡核苷酸引物以从基因组DNA来PCR扩增土生梭孢壳家族GH61R基因。使用IN-FUSION™Cloning Kit以将片段直接克隆入表达载体pA1Lo2(WO 2005/074647),而无需进行限制性消化和连接。

[0574] 正向引物:

[0575] 5'-ACTGGATTTACC**ATGGCCTTGCTGCTCTTGGCAGGC**-3'(SEQ ID NO:37)

[0576] 反向引物:

[0577] 5'-TCACCTCTAGTTAATTAAT**TCACCCATCCCATATCGGCC**-3'(SEQ ID NO:38)

[0578] 粗体字母代表编码序列。剩余序列同源于pA1Lo2的插入位点。

[0579] 将五十皮摩尔的上述每种引物用于PCR反应,所述反应含有1μg的土生梭孢壳基因组DNA,1X**ADVANTAGE**®GC-Melt LA Buffer,1μl的dATP,dTTP,dGTP,和dCTP的10mM混合物,和1.25单位的**ADVANTAGE**®GC Genomic LA Polymerase Mix,最终体积为25μl。扩增条件为:一个循环,在94℃进行1分钟;和30个循环,每循环在94℃进行30秒,59.4℃进行30秒,和72℃进行5分钟。然后将加热块维持在72℃进行5分钟,接着进行4℃浸没循环。

[0580] 将反应产物在1.0%琼脂糖凝胶上使用TAE缓冲液分离,其中将大约1kb的产物条带从凝胶切出,并使用**MINELUTE**®Gel Extraction Kit根据生产商的指示纯化。

[0581] 然后将片段使用IN-FUSION™Cloning Kit克隆入pA1Lo2。用Nco I和Pac I消化载体。将片段通过琼脂糖凝胶电泳和**QIAQUICK**®Gel Extraction Kit纯化。在反应中将基因片段和消化的载体合并在一起,得到表达质粒pAG71,其中家族GH61R基因的转录处于NA2-tpi启动子的调控之下。重组反应物(10μl)包含1X IN-FUSION™Buffer,1X BSA,0.5μl的IN-FUSION™酶(1:10稀释),93ng用Nco I和Pac I消化的pA1Lo2,和2μl的土生梭孢壳GH61R纯化PCR产物。反应在37℃温育15分钟,然后在50℃温育15分钟用40μl的TE缓冲液稀释反应物,并使用2.5μl的稀释反应物转化大肠杆菌Top10 Competent细胞。通过限制酶消化鉴定了含有pAG71(GH61R基因)的大肠杆菌转化体,并使用**BIROBOT**®9600制备质粒DNA。

[0582] 亦使用**TOPO**®**TACLONING**®Kit将相同的1kb土生梭孢壳gh61r PCR片段克隆入**pCR**®2.1-TOPO载体,以生成pAG76。通过DNA测序确认了土生梭孢壳gh61r插入。大肠杆菌pAG76于2009年9月18日保藏于农业研究机构专利培养物保藏中心北区研究中心(Agricultural Research Service Patent Culture Collection,Northern Regional Research Center,Peoria,IL,USA),并分配登录号NRRL B-50323。

[0583] 实施例26:编码具有纤维素分解增强活性的家族GH61R多肽的土生梭孢壳基因组序列的表征

[0584] 具有纤维素分解增强活性的土生梭孢壳GH61P多肽的核苷酸序列(SEQ ID NO:15)和推导的氨基酸序列(SEQ ID NO:16)示于图12。基因组多核苷酸为933bp,包含终止密码

子,且编码序列由72,53,和55bp的三个内含子间隔。预测的编码序列编码250个氨基酸的多肽。全长编码序列(包含内含子)和成熟编码序列的%G+C分别为61.8%和61.6%。使用SignalP软件程序(Nielsen等,1997,见上),预测了18个残基的信号肽。预测的成熟的蛋白含有232个氨基酸,具有26.0kDa的分子量。

[0585] 用Interproscan程序(Mulder等,2007,见上)分析具有纤维素分解增强活性的GH61R多肽的推导的氨基酸序列显示所述GH61R多肽含有糖基水解酶家族61的特征序列(InterPro登录号IPR005103)。该特征序列见于成熟多肽(Pfam登录号PF03443)的大约残基1至224。

[0586] 如EMBOSS的Needle程序中,以缺口开放罚分为10,缺口延伸罚分为0.5,和EBLOSUM62矩阵执行的,使用Needleman-Wunsch算法(Needleman和Wunsch,1970,见上)确定了氨基酸序列的比较性逐对全局比对。比对显示土生梭孢壳GH61R成熟多肽的推导的氨基酸序列与来自Chrysosporium lucknowense的另一种预测的家族61糖基水解酶的推导的氨基酸序列(GeneSeqP登录号AWI36233)分享72.8%的同一性(排除间隔)。

[0587] 实施例27:含有用于土生梭孢壳家族GH61S基因的米曲霉表达载体的构建

[0588] 设计了如下所示的两种合成的寡核苷酸引物以从实施例2中制备的基因组DNA来PCR扩增土生梭孢壳家族GH61S基因。使用IN-FUSION™Cloning Kit以将片段直接克隆入表达载体pAllLo2,而无需进行限制性消化和连接。

[0589] 正向引物:

[0590] 5'-ACTGGATTTACC**ATGATGCCGTCCCTTGTTTCGCTTC**-3'(SEQ ID NO:39)

[0591] 反向引物:

[0592] 5'-TCACCTCTAGTTAATTAAT**TCAACCATGTCTCCTGTCCC**-3'(SEQ ID NO:40)

[0593] 粗体字母代表编码序列。剩余序列同源源于pAllLo2的插入位点。

[0594] 将五十皮摩尔的上述每种引物用于PCR反应,所述反应含有1μg的土生梭孢壳基因组DNA,1X**ADVANTAGE**®GC-Melt LA Buffer,1μl的dATP,dTTP,dGTP,和dCTP的10mM混合物,和1.25单位的**ADVANTAGE**®GC Genomic LA Polymerase Mix,最终体积为25μl。扩增条件为:一个循环,在94℃进行1分钟;和30个循环,每循环在94℃进行30秒,58.5℃进行30秒,和72℃进行5分钟。然后将加热块维持在72℃进行5分钟,接着进行4℃浸没循环。

[0595] 将反应产物在1.0%琼脂糖凝胶上使用TAE缓冲液分离,其中将大约1.3kb的产物条带从凝胶切出,并使用**MINELUTE**®Gel Extraction Kit根据生产商的指示纯化。

[0596] 然后将片段使用IN-FUSION™Cloning Kit克隆入pAllLo2。用Nco I和Pac I消化载体。将片段通过琼脂糖凝胶电泳和**QIAQUICK**®Gel Extraction Kit纯化。在反应中将基因片段和消化的载体合并在一起,得到表达质粒pAG72,其中家族GH61S基因的转录处于NA2-tpi启动子的调控之下。重组反应物(10μl)包含1X IN-FUSION™Buffer,1X BSA,0.5μl的IN-FUSION™酶(1:10稀释),93ng用Nco I和Pac I消化的pAllLo2,和2μl的土生梭孢壳GH61S纯化PCR产物。反应在37℃温育15分钟,然后在50℃温育15分钟用40μl的TE缓冲液稀释反应物,并使用2.5μl的稀释反应物转化大肠杆菌Top10 Competent细胞。通过限制酶消化鉴定了含有pAG72(GH61S基因)的大肠杆菌转化体,并使用**BIOROBOT**®9600制备质粒DNA。

[0597] 亦使用**TOPO**®**TACLONING**®Kit将相同的1.3kb土生梭孢壳gh61s PCR片段克

隆入pCR®2.1-TOPO载体,以生成pAG77。通过DNA测序确认了土生梭孢壳gh61s插入。大肠杆菌pAG77于2009年9月18日保藏于农业研究机构专利培养物保藏中心北区研究中心(Agricultural Research Service Patent Culture Collection,Northern Regional Research Center,Peoria,IL,USA),并分配登录号NRRL B-50324。

[0598] 实施例28:编码具有纤维素分解增强活性的家族GH61S多肽的土生梭孢壳基因组序列的表征

[0599] 具有纤维素分解增强活性的土生梭孢壳GH61S多肽的核苷酸序列(SEQ ID NO:17)和推导的氨基酸序列(SEQ ID NO:18)示于图13。基因组多核苷酸为1584bp,包含终止密码子,且编码序列由64和83bp的两个内含子间隔。预测的编码序列编码478个氨基酸的多肽。全长编码序列(包含内含子)和成熟编码序列的%G+C分别为63.9%和64.0%。使用SignalP软件程序(Nielsen等,1997,见上),预测了信号肽,但其确切位置不明确。绝大多数GH61成熟多肽以组氨酸残基起始,因此最可能的信号肽是残基1至22。预测的成熟的蛋白含有456个氨基酸,具有48.7kDa的分子量。

[0600] 用Interproscan程序(Mulder等,2007,见上)分析具有纤维素分解增强活性的GH61S多肽的推导的氨基酸序列显示所述GH61S多肽含有糖基水解酶家族61的特征序列(InterPro登录号IPR005103)。该特征序列见于成熟多肽(Pfam登录号PF03443)的大约残基108至222。

[0601] 如EMBOSS的Needle程序中,以缺口开放罚分为10,缺口延伸罚分为0.5,和EBLOSUM62矩阵执行的,使用Needleman-Wunsch算法(Needleman和Wunsch,1970,见上)确定了氨基酸序列的比较性逐对全局比对。比对显示土生梭孢壳GH61S成熟多肽的推导的氨基酸序列与来自球毛壳菌的另一种预测的家族61糖基水解酶的推导的氨基酸序列(UniProt登录号Q2GZM2)分享65.1%的同一性(排除间隔)。

[0602] 实施例29:含有用于土生梭孢壳家族GH61T基因的米曲霉表达载体的构建

[0603] 设计了如下所示的两种合成的寡核苷酸引物以从实施例2中制备的基因组DNA来PCR扩增土生梭孢壳家族GH61T基因。使用IN-FUSION™Cloning Kit以将片段直接克隆入表达载体pAllo2,而无需进行限制性消化和连接。

[0604] 正向引物:

[0605] 5'-ACTGGATTTACC**ATGCAGCTCCTCGTGGGCTT**-3'(SEQ ID NO:41)

[0606] 反向引物:

[0607] 5'-TCACCTCTAGTTAATTAAT**TCAGCCACTCCACACCGGCG**-3'(SEQ ID NO:42)

[0608] 粗体字母代表编码序列。剩余序列同源於pAllo2的插入位点。

[0609] 将五十皮摩尔的上述每种引物用于PCR反应,所述反应含有1µg的土生梭孢壳基因组DNA,1X**ADVANTAGE**®GC-Melt LA Buffer,1µl的dATP,dTTP,dGTP,和dCTP的10mM混合物,和1.25单位的**ADVANTAGE**®GC Genomic LA Polymerase Mix,最终体积为25µl。扩增条件为:一个循环,在94°C进行1分钟;和30个循环,每循环在94°C进行30秒,60.5°C进行30秒,和72°C进行5分钟。然后将加热块维持在72°C进行5分钟,接着进行4°C浸没循环。

[0610] 将反应产物在1.0%琼脂糖凝胶上使用TAE缓冲液分离,其中将大约900bp的产物条带从凝胶切出,并使用**MINELUTE**®Gel Extraction Kit根据生产商的指示纯化。

[0611] 然后将片段使用IN-FUSION™Cloning Kit克隆入pAllo2。用Nco I和Pac I消化载

体。将片段通过琼脂糖凝胶电泳和QIAQUICK® Gel Extraction Kit纯化。在反应中将基因片段和消化的载体合并在一起,得到表达质粒pAG73,其中家族GH61T基因的转录处于NA2-tpi启动子的调控之下。重组反应物(10 μ l)包含1X IN-FUSION™ Buffer, 1X BSA, 0.5 μ l的IN-FUSION™酶(1:10稀释), 93ng用Nco I和Pac I消化的pAllo2, 和2 μ l的土生梭孢壳GH61T纯化PCR产物。反应在37°C温育15分钟,然后在50°C温育15分钟用40 μ l的TE缓冲液稀释反应物,并使用2.5 μ l的稀释反应物转化大肠杆菌Top10 Competent细胞。通过限制酶消化鉴定了含有pAG73(GH61T基因)的大肠杆菌转化体,并使用BIOROBOT®9600制备质粒DNA。

[0612] 亦使用TOPO®TA CLONING® Kit将相同的900bp土生梭孢壳gh61t PCR片段克隆入pCR®2.1-TOPO载体,以生成pAG78。通过DNA测序确认了土生梭孢壳gh61t插入。大肠杆菌pAG78于2009年9月18日保藏于农业研究机构专利培养物保藏中心北区研究中心(Agricultural Research Service Patent Culture Collection, Northern Regional Research Center, Peoria, IL, USA),并分配登录号NRRL B-50325。

[0613] 实施例30:编码具有纤维素分解增强活性的家族GH61T多肽的土生梭孢壳基因组序列的表征

[0614] 具有纤维素分解增强活性的土生梭孢壳GH61T多肽的核苷酸序列(SEQ ID NO:19)和推导的氨基酸序列(SEQ ID NO:20)示于图14。基因组多核苷酸为868bp,包含终止密码子,且编码序列由76和99bp的两个内含子间隔。预测的编码序列编码230个氨基酸的多肽。全长编码序列(包含内含子)和成熟编码序列的%G+C分别为61.5%和61.4%。使用SignalP软件程序(Nielsen等,1997,见上),预测了16个残基的信号肽。预测的成熟的蛋白含有214个氨基酸,具有23.1kDa的分子量。

[0615] 用Interproscan程序(Mulder等,2007,见上)分析具有纤维素分解增强活性的GH61T多肽的推导的氨基酸序列显示所述GH61T多肽含有糖基水解酶家族61的特征序列(InterPro登录号IPR005103)。该特征序列见于成熟多肽(Pfam登录号PF03443)的大约残基71至197。

[0616] 如EMBOSS的Needle程序中,以缺口开放罚分为10,缺口延伸罚分为0.5,和EBLOSUM62矩阵执行的,使用Needleman-Wunsch算法(Needleman和Wunsch,1970,见上)确定了氨基酸序列的比较性逐对全局比对。比对显示土生梭孢壳GH61T成熟多肽的推导的氨基酸序列与来自球毛壳菌的另一种预测的家族61糖基水解酶的推导的氨基酸序列(UniProt登录号Q2GUT0)分享87.4%的同一性(排除间隔)。

[0617] 实施例31:含有用于土生梭孢壳家族GH61U基因的米曲霉表达载体的构建

[0618] 设计了如下所示的两种合成的寡核苷酸引物以从实施例2中制备的基因组DNA来PCR扩增土生梭孢壳家族GH61U基因。使用IN-FUSION™ Cloning Kit以将片段直接克隆入表达载体pAllo2,而无需进行限制性消化和连接。

[0619] 正向引物:

[0620] 5'-ACTGGATTTACC**ATGAAGCTGTACCTGGCGGCCTTT**-3'(SEQ ID NO:43)

[0621] 反向引物:

[0622] 5'-TCACCTCTAGTTAATTA**TCAACCAGTCCACAGCGCTG**-3'(SEQ ID NO:44)

[0623] 粗体字母代表编码序列。剩余序列同源于pAllo2的插入位点。

[0624] 将五十皮摩尔的上述每种引物用于PCR反应,所述反应含有1 μ g的土生梭孢壳基因组DNA,1X **ADVANTAGE**[®]GC-Melt LA Buffer,1 μ l的dATP,dTTP,dGTP,和dCTP的10mM混合物,和1.25单位的**ADVANTAGE**[®]GC Genomic LA Polymerase Mix,最终体积为25 μ l。扩增条件为:一个循环,在94 $^{\circ}$ C进行1分钟;和30个循环,每循环在94 $^{\circ}$ C进行30秒,58.5 $^{\circ}$ C进行30秒,和72 $^{\circ}$ C进行5分钟。然后将加热块维持在72 $^{\circ}$ C进行5分钟,接着进行4 $^{\circ}$ C浸没循环。

[0625] 将反应产物在1.0%琼脂糖凝胶上使用TAE缓冲液分离,其中将大约1kb的产物条带从凝胶切出,并使用**MINELUTE**[®]Gel Extraction Kit根据生产商的指示纯化。

[0626] 然后将片段使用IN-FUSION[™]Cloning Kit克隆入pAllo2。用Nco I和Pac I消化载体。将片段通过琼脂糖凝胶电泳和**QIAQUICK**[®]Gel Extraction Kit纯化。在反应中将基因片段和消化的载体合并在一起,得到表达质粒pAG74,其中家族GH61U基因的转录处于NA2-tpi启动子的调控之下。重组反应物(10 μ l)包含1X IN-FUSION[™]Buffer,1X BSA,0.5 μ l的IN-FUSION[™]酶(1:10稀释),93ng用Nco I和Pac I消化的pAllo2,和2 μ l的土生梭孢壳GH61U纯化PCR产物。反应在37 $^{\circ}$ C温育15分钟,然后在50 $^{\circ}$ C温育15分钟用40 μ l的TE缓冲液稀释反应物,并使用2.5 μ l的稀释反应物转化大肠杆菌Top10 Competent细胞。通过限制酶消化鉴定了含有pAG74(GH61U基因)的大肠杆菌转化体,并使用**BIOROBOT**[®]9600制备质粒DNA。

[0627] 亦使用**TOPO**[®]**TACLONING**[®]Kit将相同的1kb土生梭孢壳gh61u PCR片段克隆入**pCR**[®]2.1-TOPO载体,以生成pAG79。通过DNA测序确认了土生梭孢壳gh61j插入。大肠杆菌pAG79于2009年9月18日保藏于农业研究机构专利培养物保藏中心北区研究中心(Agricultural Research Service Patent Culture Collection,Northern Regional Research Center,Peoria,IL,USA),并分配登录号NRRL B-50326。

[0628] 实施例32:编码具有纤维素分解增强活性的家族GH61U多肽的土生梭孢壳基因组序列的表征

[0629] 具有纤维素分解增强活性的土生梭孢壳GH61U多肽的核苷酸序列(SEQ ID NO:21)和推导的氨基酸序列(SEQ ID NO:22)示于图15。基因组多核苷酸为1068bp,包含终止密码子,且编码序列由64,52,96和82bp的四个内含子间隔。预测的编码序列编码257个氨基酸的多肽。全长编码序列(包含内含子)和成熟编码序列的%G+C分别为59.7%和59.3%。使用SignalP软件程序(Nielsen等,1997,见上),预测了19个残基的信号肽。预测的成熟的蛋白含有238个氨基酸,具有26.6kDa的分子量。

[0630] 用Interproscan程序(Mulder等,2007,见上)分析具有纤维素分解增强活性的GH61T多肽的推导的氨基酸序列未显示所述GH61U多肽含有糖基水解酶家族61的特征序列(InterPro登录号IPR005103)。然而,针对Pfam数据库的直接检索产生了与GH61家族(Pfam登录号PF03443)的明显命中(significant hit)(4.3×10^{-8} 的e值)。如EMBOSS的Needle程序中,以缺口开放罚分为10,缺口延伸罚分为0.5,和EBL0SUM62矩阵执行的,使用Needleman-Wunsch算法(Needleman和Wunsch,1970,见上)确定了氨基酸序列的比较性逐对全局比对。比对显示土生梭孢壳GH61U成熟多肽的推导的氨基酸序列与来自球毛壳菌的另一种预测的家族61糖基水解酶的推导的氨基酸序列(UniProt登录号Q2HHT1)分享74.4%的同一性(排除间隔)。

[0631] 生物材料的保藏

[0632] 下述的生物材料已经依据布达佩斯条约的条款保藏于农业研究机构专利培养物保藏中心北区研究中心(Agricultural Research Service Patent Culture Collection, Northern Regional Research Center), Peoria, IL, USA, 并给予下述的登录号:

[0633]

保藏	登录号	保藏日期
大肠杆菌(pSMai216)	NRRL B-50301	2009年7月15日
大肠杆菌(pSMai217)	NRRL B-50302	2009年7月15日
大肠杆菌(pSMai218)	NRRL B-50303	2009年7月15日
大肠杆菌(pSMai213)	NRRL B-50300	2009年7月15日
大肠杆菌(pAG68)	NRRL B-50320	2009年9月18日
大肠杆菌(pAG69)	NRRL B-50321	2009年9月18日
大肠杆菌(pAG75)	NRRL B-50322	2009年9月18日
大肠杆菌(pAG76)	NRRL B-50323	2009年9月18日
大肠杆菌(pAG77)	NRRL B-50324	2009年9月18日
大肠杆菌(pAG78)	NRRL B-50325	2009年9月18日
大肠杆菌(pAG79)	NRRL B-50326	2009年9月18日

[0634] 所述菌株于下述条件下保藏:确保在本专利申请未决期间,依据该外国专利法律的授权的人能够获得所述培养物。所述保藏物为所保藏菌株的基本上纯的培养物。在提交了该申请的副本,或其后续文本的国家,依据该外国专利法律可以获得所述保藏物。然而,应当理解,保藏物的获得并不构成对实施本发明的许可,实施本发明是对政府行为所授予的专利权的侵犯。

[0635] 通过下述段落进一步描述本发明:

[0636] [1]一种具有纤维素分解增强活性的分离的多肽,其选自下组:

[0637] (a)多肽,其具有与SEQ ID NO:2,SEQ ID NO:6,或SEQ ID NO:12的成熟多肽至少60%序列同一性;与SEQ ID NO:4的成熟多肽至少65%序列同一性;与SEQ ID NO:18的成熟多肽至少70%序列同一性;与SEQ ID NO:10,SEQ ID NO:16,或SEQ ID NO:22的成熟多肽至少75%序列同一性;与SEQ ID NO:8的成熟多肽至少80%序列同一性;与SEQ ID NO:14的成熟多肽至少85%序列同一性;或者与SEQ ID NO:20的成熟多肽至少90%序列同一性;

[0638] (b)多肽,其由多核苷酸编码,所述多核苷酸在中等-高,高或非常高严格条件下与以下杂交:(i)SEQ ID NO:1,SEQ ID NO:3,SEQ ID NO:5,SEQ ID NO:11,或SEQ ID NO:17的成熟多肽编码序列,(ii)包含于SEQ ID NO:1,SEQ ID NO:3,SEQ ID NO:5,SEQ ID NO:11,或SEQ ID NO:17的成熟多肽编码序列中的cDNA序列,或(iii)(i)或(ii)的全长互补链;在高或非常高严格条件下与以下杂交:(i)SEQ ID NO:7,SEQ ID NO:9,SEQ ID NO:13,SEQ ID NO:15,或SEQ ID NO:21的成熟多肽编码序列,(ii)包含于SEQ ID NO:7,SEQ ID NO:9,SEQ ID NO:13,SEQ ID NO:15,或SEQ ID NO:21的成熟多肽编码序列中的cDNA序列,或(iii)(i)或(ii)的全长互补链;或者在非常高严格条件下与以下杂交:i)SEQ ID NO:19的成熟多肽

编码序列,(ii)包含于SEQ ID NO:19的成熟多肽编码序列中的cDNA序列,或(iii)(i)或(ii)的全长互补链;

[0639] (c)多肽,其由多核苷酸编码,所述多核苷酸具有与SEQ ID NO:1,SEQ ID NO:5,或SEQ ID NO:11的成熟多肽编码序列至少60%序列同一性;与SEQ ID NO:3的成熟多肽编码序列至少65%序列同一性;与SEQ ID NO:17的成熟多肽编码序列至少70%序列同一性;与SEQ ID NO:9,SEQ ID NO:15,或SEQ ID NO:21的成熟多肽编码序列至少75%序列同一性;与SEQ ID NO:7的成熟多肽编码序列至少80%序列同一性;与SEQ ID NO:13的成熟多肽编码序列至少85%序列同一性;或者与SEQ ID NO:19的成熟多肽编码序列至少90%序列同一性;或者其cDNA序列;

[0640] (d)变体,其包含SEQ ID NO:2,SEQ ID NO:4,SEQ ID NO:6,SEQ ID NO:8,SEQ ID NO:10,SEQ ID NO:12,SEQ ID NO:14,SEQ ID NO:16,SEQ ID NO:18,SEQ ID NO:20,或SEQ ID NO:22的成熟多肽的一个或多个(几个)氨基酸的取代,缺失,和/或插入;和

[0641] (e)(a)、(b)、(c)或(d)的多肽具有纤维素分解增强活性的片段。

[0642] [2]段1的多肽,其与SEQ ID NO:2,SEQ ID NO:6,或SEQ ID NO:12的成熟多肽具有至少60%序列同一性。

[0643] [3]段2的多肽,其与SEQ ID NO:2,SEQ ID NO:6,或SEQ ID NO:12的成熟多肽具有至少65%序列同一性。

[0644] [4]段3的多肽,其与SEQ ID NO:2,SEQ ID NO:6,或SEQ ID NO:12的成熟多肽具有至少70%序列同一性。

[0645] [5]段4的多肽,其与SEQ ID NO:2,SEQ ID NO:6,或SEQ ID NO:12的成熟多肽具有至少75%序列同一性。

[0646] [6]段5的多肽,其与SEQ ID NO:2,SEQ ID NO:6,或SEQ ID NO:12的成熟多肽具有至少80%序列同一性。

[0647] [7]段6的多肽,其与SEQ ID NO:2,SEQ ID NO:6,或SEQ ID NO:12的成熟多肽具有至少85%序列同一性。

[0648] [8]段7的多肽,其与SEQ ID NO:2,SEQ ID NO:6,或SEQ ID NO:12的成熟多肽具有至少90%序列同一性。

[0649] [9]段8的多肽,其与SEQ ID NO:2,SEQ ID NO:6,或SEQ ID NO:12的成熟多肽具有至少95%序列同一性。

[0650] [10]段9的多肽,其与SEQ ID NO:2,SEQ ID NO:6,或SEQ ID NO:12的成熟多肽具有至少97%序列同一性。

[0651] [11]段1的多肽,其与SEQ ID NO:4的成熟多肽具有至少65%序列同一性。

[0652] [12]段11的多肽,其与SEQ ID NO:4的成熟多肽具有至少70%序列同一性。

[0653] [13]段12的多肽,其与SEQ ID NO:4的成熟多肽具有至少75%序列同一性。

[0654] [14]段13的多肽,其与SEQ ID NO:4的成熟多肽具有至少80%序列同一性。

[0655] [15]段14的多肽,其与SEQ ID NO:4的成熟多肽具有至少85%序列同一性。

[0656] [16]段15的多肽,其与SEQ ID NO:4的成熟多肽具有至少90%序列同一性。

[0657] [17]段16的多肽,其与SEQ ID NO:4的成熟多肽具有至少95%序列同一性。

[0658] [18]段17的多肽,其与SEQ ID NO:4的成熟多肽具有至少97%序列同一性。

- [0659] [19]段1的多肽,其与SEQ ID NO:18的成熟多肽具有至少70%序列同一性。
- [0660] [20]段19的多肽,其与SEQ ID NO:18的成熟多肽具有至少75%序列同一性。
- [0661] [21]段20的多肽,其与SEQ ID NO:18的成熟多肽具有至少80%序列同一性。
- [0662] [22]段21的多肽,其与SEQ ID NO:18的成熟多肽具有至少85%序列同一性。
- [0663] [23]段22的多肽,其与SEQ ID NO:18的成熟多肽具有至少90%序列同一性。
- [0664] [24]段23的多肽,其与SEQ ID NO:18的成熟多肽具有至少95%序列同一性。
- [0665] [25]段24的多肽,其与SEQ ID NO:18的成熟多肽具有至少97%序列同一性。
- [0666] [26]段1的多肽,其与SEQ ID NO:10,SEQ ID NO:16,或SEQ ID NO:22的成熟多肽具有至少75%序列同一性。
- [0667] [27]段26的多肽,其与SEQ ID NO:10,SEQ ID NO:16,或SEQ ID NO:22的成熟多肽具有至少80%序列同一性。
- [0668] [28]段27的多肽,其与SEQ ID NO:10,SEQ ID NO:16,或SEQ ID NO:22的成熟多肽具有至少85%序列同一性。
- [0669] [29]段28的多肽,其与SEQ ID NO:10,SEQ ID NO:16,或SEQ ID NO:22的成熟多肽具有至少90%序列同一性。
- [0670] [30]段29的多肽,其与SEQ ID NO:10,SEQ ID NO:16,或SEQ ID NO:22的成熟多肽具有至少95%序列同一性。
- [0671] [31]段30的多肽,其与SEQ ID NO:10,SEQ ID NO:16,或SEQ ID NO:22的成熟多肽具有至少97%序列同一性。
- [0672] [32]段1的多肽,其与SEQ ID NO:8的成熟多肽具有至少80%序列同一性。
- [0673] [33]段32的多肽,其与SEQ ID NO:8的成熟多肽具有至少85%序列同一性。
- [0674] [34]段33的多肽,其与SEQ ID NO:8的成熟多肽具有至少90%序列同一性。
- [0675] [35]段34的多肽,其与SEQ ID NO:8的成熟多肽具有至少95%序列同一性。
- [0676] [36]段35的多肽,其与SEQ ID NO:8的成熟多肽具有至少97%序列同一性。
- [0677] [37]段1的多肽,其与SEQ ID NO:14的成熟多肽具有至少85%序列同一性。
- [0678] [38]段37的多肽,其与SEQ ID NO:14的成熟多肽具有至少90%序列同一性。
- [0679] [39]段38的多肽,其与SEQ ID NO:14的成熟多肽具有至少95%序列同一性。
- [0680] [40]段39的多肽,其与SEQ ID NO:14的成熟多肽具有至少97%序列同一性。
- [0681] [41]段1的多肽,其与SEQ ID NO:20的成熟多肽具有至少90%序列同一性。
- [0682] [42]段41的多肽,其与SEQ ID NO:20的成熟多肽具有至少95%序列同一性。
- [0683] [43]段42的多肽,其与SEQ ID NO:20的成熟多肽具有至少97%序列同一性。
- [0684] [44]段1的多肽,其由多核苷酸编码,所述多核苷酸在中等-高严格条件下与以下杂交:(i)SEQ ID NO:1,SEQ ID NO:3,SEQ ID NO:5,SEQ ID NO:11,或SEQ ID NO:17的成熟多肽编码序列,(ii)包含于SEQ ID NO:1,SEQ ID NO:3,SEQ ID NO:5,SEQ ID NO:11,或SEQ ID NO:17的成熟多肽编码序列中的cDNA序列,或(iii)(i)或(ii)的全长互补链。
- [0685] [45]段44的多肽,其由多核苷酸编码,所述多核苷酸在高严格条件下与以下杂交:(i)SEQ ID NO:1,SEQ ID NO:3,SEQ ID NO:5,SEQ ID NO:11,或SEQ ID NO:17的成熟多肽编码序列,(ii)包含于SEQ ID NO:1,SEQ ID NO:3,SEQ ID NO:5,SEQ ID NO:11,或SEQ ID NO:17的成熟多肽编码序列中的cDNA序列,或(iii)(i)或(ii)的全长互补链。

[0686] [46]段45的多肽,其由多核苷酸编码,所述多核苷酸在非常高严格条件下与以下杂交:(i)SEQ ID NO:1,SEQ ID NO:3,SEQ ID NO:5,SEQ ID NO:11,或SEQ ID NO:17的成熟多肽编码序列,(ii)包含于SEQ ID NO:1,SEQ ID NO:3,SEQ ID NO:5,SEQ ID NO:11,或SEQ ID NO:17的成熟多肽编码序列中的cDNA序列,或(iii)(i)或(ii)的全长互补链。

[0687] [47]段1的多肽,其由多核苷酸编码,所述多核苷酸在高严格条件下与以下杂交:(i)SEQ ID NO:7,SEQ ID NO:9,SEQ ID NO:13,SEQ ID NO:15,或SEQ ID NO:21的成熟多肽编码序列,(ii)包含于SEQ ID NO:7,SEQ ID NO:9,SEQ ID NO:13,SEQ ID NO:15,或SEQ ID NO:21的成熟多肽编码序列中的cDNA序列,或(iii)(i)或(ii)的全长互补链。

[0688] [48]段47的多肽,其由多核苷酸编码,所述多核苷酸在非常高严格条件下与以下杂交:(i)SEQ ID NO:7,SEQ ID NO:9,SEQ ID NO:13,SEQ ID NO:15,或SEQ ID NO:21的成熟多肽编码序列,(ii)包含于SEQ ID NO:7,SEQ ID NO:9,SEQ ID NO:13,SEQ ID NO:15,或SEQ ID NO:21的成熟多肽编码序列中的cDNA序列,或(iii)(i)或(ii)的全长互补链。

[0689] [49]段1的多肽,其由多核苷酸编码,所述多核苷酸在非常高严格条件下与以下杂交:(i)SEQ ID NO:19的成熟多肽编码序列,(ii)包含于SEQ ID NO:19的成熟多肽编码序列中的cDNA序列,或(iii)(i)或(ii)的全长互补链。

[0690] [50]段1的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:1,SEQ ID NO:5,或SEQ ID NO:11的成熟多肽编码序列,或其cDNA序列具有至少60%序列同一性。

[0691] [51]段50的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:1,SEQ ID NO:5,或SEQ ID NO:11的成熟多肽编码序列,或其cDNA序列具有至少65%序列同一性。

[0692] [52]段51的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:1,SEQ ID NO:5,或SEQ ID NO:11的成熟多肽编码序列,或其cDNA序列具有至少70%序列同一性。

[0693] [53]段52的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:1,SEQ ID NO:5,或SEQ ID NO:11的成熟多肽编码序列,或其cDNA序列具有至少75%序列同一性。

[0694] [54]段53的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:1,SEQ ID NO:5,或SEQ ID NO:11的成熟多肽编码序列,或其cDNA序列具有至少80%序列同一性。

[0695] [55]段54的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:1,SEQ ID NO:5,或SEQ ID NO:11的成熟多肽编码序列,或其cDNA序列具有至少85%序列同一性。

[0696] [56]段55的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:1,SEQ ID NO:5,或SEQ ID NO:11的成熟多肽编码序列,或其cDNA序列具有至少90%序列同一性。

[0697] [57]段56的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:1,SEQ ID NO:5,或SEQ ID NO:11的成熟多肽编码序列,或其cDNA序列具有至少95%序列同一性。

[0698] [58]段57的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:1,SEQ ID NO:5,或SEQ ID NO:11的成熟多肽编码序列,或其cDNA序列具有至少97%序列同一性。

[0699] [59]段1的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:3的成熟多肽编码序列,或其cDNA序列具有至少65%序列同一性。

[0700] [60]段59的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:3的成熟多肽编码序列,或其cDNA序列具有至少70%序列同一性。

[0701] [61]段60的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:3的成熟多肽编码序列,或其cDNA序列具有至少75%序列同一性。

- [0702] [62]段61的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:3的成熟多肽编码序列,或其cDNA序列具有至少80%序列同一性。
- [0703] [63]段62的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:3的成熟多肽编码序列,或其cDNA序列具有至少85%序列同一性。
- [0704] [64]段63的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:3的成熟多肽编码序列,或其cDNA序列具有至少90%序列同一性。
- [0705] [65]段64的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:3的成熟多肽编码序列,或其cDNA序列具有至少95%序列同一性。
- [0706] [66]段65的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:3的成熟多肽编码序列,或其cDNA序列具有至少97%序列同一性。
- [0707] [67]段1的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:17的成熟多肽编码序列,或其cDNA序列具有至少70%序列同一性。
- [0708] [68]段67的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:17的成熟多肽编码序列,或其cDNA序列具有至少75%序列同一性。
- [0709] [69]段68的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:17的成熟多肽编码序列,或其cDNA序列具有至少80%序列同一性。
- [0710] [70]段69的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:17的成熟多肽编码序列,或其cDNA序列具有至少85%序列同一性。
- [0711] [71]段70的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:17的成熟多肽编码序列,或其cDNA序列具有至少90%序列同一性。
- [0712] [72]段71的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:17的成熟多肽编码序列,或其cDNA序列具有至少95%序列同一性。
- [0713] [73]段72的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:17的成熟多肽编码序列,或其cDNA序列具有至少97%序列同一性。
- [0714] [74]段1的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:9,SEQ ID NO:15,或SEQ ID NO:21的成熟多肽编码序列,或其cDNA序列具有至少75%序列同一性。
- [0715] [75]段74的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:9,SEQ ID NO:15,或SEQ ID NO:21的成熟多肽编码序列,或其cDNA序列具有至少80%序列同一性。
- [0716] [76]段75的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:9,SEQ ID NO:15,或SEQ ID NO:21的成熟多肽编码序列,或其cDNA序列具有至少85%序列同一性。
- [0717] [77]段76的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:9,SEQ ID NO:15,或SEQ ID NO:21的成熟多肽编码序列,或其cDNA序列具有至少90%序列同一性。
- [0718] [78]段77的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:9,SEQ ID NO:15,或SEQ ID NO:21的成熟多肽编码序列,或其cDNA序列具有至少95%序列同一性。
- [0719] [79]段78的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:9,SEQ ID NO:15,或SEQ ID NO:21的成熟多肽编码序列,或其cDNA序列具有至少97%序列同一性。
- [0720] [80]段1的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:7的成熟多肽编码序列,或其cDNA序列具有至少80%序列同一性。
- [0721] [81]段80的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:7的成熟多肽编

码序列,或其cDNA序列具有至少85%序列同一性。

[0722] [82]段81的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:7的成熟多肽编码序列,或其cDNA序列具有至少90%序列同一性。

[0723] [83]段82的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:7的成熟多肽编码序列,或其cDNA序列具有至少95%序列同一性。

[0724] [84]段83的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:7的成熟多肽编码序列,或其cDNA序列具有至少97%序列同一性。

[0725] [85]段1的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:13的成熟多肽编码序列,或其cDNA序列具有至少85%序列同一性。

[0726] [86]段85的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:13的成熟多肽编码序列,或其cDNA序列具有至少90%序列同一性。

[0727] [87]段86的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:13的成熟多肽编码序列,或其cDNA序列具有至少95%序列同一性。

[0728] [88]段87的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:13的成熟多肽编码序列,或其cDNA序列具有至少97%序列同一性。

[0729] [89]段1的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:19的成熟多肽编码序列,或其cDNA序列具有至少90%序列同一性。

[0730] [90]段89的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:19的成熟多肽编码序列,或其cDNA序列具有至少95%序列同一性。

[0731] [91]段90的多肽,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:19的成熟多肽编码序列,或其cDNA序列具有至少97%序列同一性。

[0732] [92]段1-91任一项的多肽,包含或组成为SEQ ID NO:2,SEQ ID NO:4,SEQ ID NO:6,SEQ ID NO:8,SEQ ID NO:10,SEQ ID NO:12,SEQ ID NO:14,SEQ ID NO:16,SEQ ID NO:18,SEQ ID NO:20,或SEQ ID NO:22。

[0733] [93]段1-91任一项的多肽,包含或组成为SEQ ID NO:2,SEQ ID NO:4,SEQ ID NO:6,SEQ ID NO:8,SEQ ID NO:10,SEQ ID NO:12,SEQ ID NO:14,SEQ ID NO:16,SEQ ID NO:18,SEQ ID NO:20,或SEQ ID NO:22的成熟多肽。

[0734] [94]段93的多肽,其中所述成熟多肽是SEQ ID NO:2的氨基酸18至246,SEQ ID NO:4的氨基酸20至334,SEQ ID NO:6的氨基酸18至227,SEQ ID NO:8的氨基酸20至223,SEQ ID NO:10的氨基酸22至368,SEQ ID NO:12的氨基酸25至330,SEQ ID NO:14的氨基酸17至236,SEQ ID NO:16的氨基酸19至250,SEQ ID NO:18的氨基酸23至478,SEQ ID NO:20的氨基酸17至230,或SEQ ID NO:22的氨基酸20至257。

[0735] [95]段1的多肽,其为SEQ ID NO:2,SEQ ID NO:4,SEQ ID NO:6,SEQ ID NO:8,SEQ ID NO:10,SEQ ID NO:12,SEQ ID NO:14,SEQ ID NO:16,SEQ ID NO:18,SEQ ID NO:20,或SEQ ID NO:22的片段,其中所述片段具有纤维素分解增强活性。

[0736] [96]段1的多肽,其由多核苷酸编码,所述多核苷酸包含于pSMai216,其包含于大肠杆菌NRRL B-50301;所述多核苷酸包含于pSMai217,其包含于大肠杆菌NRRL B-50302;所述多核苷酸包含于pSMai218,其包含于大肠杆菌NRRL B-50303;所述多核苷酸包含于pSMai213,其包含于大肠杆菌NRRL B-50300;所述多核苷酸包含于pAG68,其包含于大肠杆

菌NRRL B-50320;所述多核苷酸包含于pAG69,其包含于大肠杆菌NRRL B-50321;所述多核苷酸包含于pAG75,其包含于大肠杆菌NRRL B-50322;所述多核苷酸包含于pAG76,其包含于大肠杆菌NRRL B-50323;所述多核苷酸包含于pAG77,其包含于大肠杆菌NRRL B-50324;所述多核苷酸包含于pAG78,其包含于大肠杆菌NRRL B-50325;或所述多核苷酸包含于pAG79,其包含于大肠杆菌NRRL B-50326。

[0737] [97]一种组合物,其包含段1-96任一项的多肽。

[0738] [98]一种分离的多核苷酸,其编码段1-96任一项的多肽。

[0739] [99]一种核酸构建体或表达载体,其包含段98的多核苷酸,所述多核苷酸可操作地连接于一个或多个(几个)调控序列,所述调控序列指导所述多肽在表达宿主中的产生。

[0740] [100]一种重组宿主细胞,其包含段98的多核苷酸,所述多核苷酸可操作地连接于一个或多个调控序列,所述调控序列指导所述多肽的产生。

[0741] [101]一种产生段1-96任一项的多肽的方法,其包括:(a)在有助于所述多肽产生的条件下培养细胞,所述细胞以其野生型形式产生所述多肽;和(b)回收所述多肽。

[0742] [102]一种产生具有纤维素分解增强活性的多肽的方法,其包括:(a)在有助于所述多肽产生的条件下培养包含段98的多核苷酸的重组宿主细胞;和(b)回收所述多肽。

[0743] [103]一种转基因植物、植物部分或植物细胞,其用编码段1-96任一项的多肽的多核苷酸转化。

[0744] [104]一种产生具有纤维素分解增强活性的多肽的方法,其包括:(a)在有助于所述多肽产生的条件下培养段103的转基因植物或植物细胞;和(b)回收所述多肽。

[0745] [105]一种产生亲本细胞的突变体的方法,包括使编码段1-96任一项的多肽的多核苷酸失活,其导致所述突变体与亲本细胞相比较产生较少的多肽。

[0746] [106]一种突变体细胞,其由段105的方法产生。

[0747] [107]段106的突变体细胞,进一步包含编码天然或异源蛋白的基因。

[0748] [108]一种产生蛋白的方法,其包括:(a)在有助于所述蛋白产生的条件下培养段106或107的突变体细胞;和(b)回收所述蛋白。

[0749] [109]一种双链抑制RNA(dsRNA)分子,其包含段98的多核苷酸的亚序列,其中dsRNA任选地为siRNA或miRNA分子。

[0750] [110]段109的双链抑制RNA(dsRNA)分子,其为长约15、16、17、18、19、20、21、22、23、24、25或更长的双链体核苷酸。

[0751] [111]一种抑制具有纤维素分解增强活性的多肽在细胞中表达的方法,其包括对细胞施用或在细胞中表达段109或110的双链RNA(dsRNA)分子。

[0752] [112]一种通过段111的方法产生的细胞

[0753] [113]段111的细胞,进一步包含编码天然或异源蛋白的基因。

[0754] [114]一种产生蛋白的方法,其包括:(a)在有助于蛋白产生的条件下培养段112或113的细胞;和(b)回收所述蛋白。

[0755] [115]一种分离的多核苷酸,其编码信号肽,所述信号肽包含或组成为SEQ ID NO:2的氨基酸1至17,SEQ ID NO:4的氨基酸1至19,SEQ ID NO:6的氨基酸1至17,SEQ ID NO:8的氨基酸1至19,SEQ ID NO:10的氨基酸1至21,SEQ ID NO:12的氨基酸1至24,SEQ ID NO:14的氨基酸1至16,SEQ ID NO:16的氨基酸1至18,SEQ ID NO:18的氨基酸1至22,SEQ ID

NO:20的氨基酸1至16,或SEQ ID NO:22的氨基酸1至19。

[0756] [116]一种核酸构建体或表达载体,其包含编码蛋白的基因,所述基因可操作地连接于段115的多核苷酸,其中所述基因对于编码信号肽的多核苷酸是外源的。

[0757] [117]一种重组宿主细胞,其包含段115的多核苷酸,其中所述基因对于编码信号肽的多核苷酸是外源的。

[0758] [118]一种产生蛋白的方法,其包括:(a)在有助于所述蛋白产生的条件下培养包含段115的多核苷酸的重组宿主细胞,其中所述基因对于编码信号肽的多核苷酸是外源的;和(b)回收所述蛋白。

[0759] [119]一种用于降解或转化纤维素材料的方法,其包括:在段1-96任一项的具有纤维素分解增强活性的多肽的存在下用酶组合物处理所述纤维素材料。

[0760] [120]段119的方法,其中所述纤维素材料经预处理。

[0761] [121]段119或120的方法,其中所述酶组合物包含一种或多种(几种)选自下组的酶:纤维素酶、半纤维素酶、棒曲霉素、酯酶、漆酶、木质素分解酶、果胶酶、过氧化物酶、蛋白酶和膨胀素。

[0762] [122]段121的方法,其中所述纤维素酶是一种或多种(几种)选自下组的酶:内切葡聚糖酶、纤维二糖水解酶和 β -葡糖苷酶。

[0763] [123]段121的方法,其中所述半纤维素酶是一种或多种(几种)选自下组的酶:木聚糖酶、乙酰木聚糖酯酶、阿魏酸酯酶、阿拉伯呋喃糖苷酶、木糖苷酶和葡糖醛酸糖苷酶。

[0764] [124]段119-123任一项的方法,进一步包括回收经降解的纤维素材料。

[0765] [125]段124的方法,其中所述经降解的材料是糖。

[0766] [126]段125的方法,其中所述糖选自下组:葡萄糖、木糖、甘露糖、半乳糖和阿拉伯糖。

[0767] [127]一种用于产生发酵产物的方法,包括:(a)在段1-96任一项具有纤维素分解增强活性的多肽的存在下用酶组合物糖化纤维素材料;(b)用一种或多种发酵微生物发酵糖化的纤维素材料以产生发酵产物;和(c)从发酵回收所述发酵产物。

[0768] [128]段127的方法,其中所述纤维素材料经预处理。

[0769] [129]段127或128的方法,其中所述酶组合物包含一种或多种(几种)选自下组的酶:纤维素酶、半纤维素酶、棒曲霉素、酯酶、漆酶、木质素分解酶、果胶酶、过氧化物酶、蛋白酶和膨胀素。

[0770] [130]段129的方法,其中所述纤维素酶是一种或多种(几种)选自下组的酶:内切葡聚糖酶、纤维二糖水解酶和 β -葡糖苷酶。

[0771] [131]段129的方法,其中所述半纤维素酶是一种或多种(几种)选自下组的酶:木聚糖酶、乙酰木聚糖酯酶、阿魏酸酯酶、阿拉伯呋喃糖苷酶、木糖苷酶和葡糖醛酸糖苷酶。

[0772] [132]段127-131任一项的方法,其中步骤(a)和(b)在同时糖化和发酵中同时进行。

[0773] [133]段127-132任一项的方法,其中发酵产物是醇、有机酸、酮、氨基酸或气体。

[0774] [134]一种发酵纤维素材料的方法,包括:用一种或多种发酵微生物发酵所述纤维素材料,其中所述纤维素材料经在段1-96任一项的具有纤维素分解增强活性的多肽的存在下用酶组合物的糖化。

[0775] [135]段134的方法,其中所述纤维素材料的发酵产生发酵产物。

[0776] [136]段135的方法,进一步包括从发酵回收发酵产物。

[0777] [137]段134-136任一项的方法,其中所述纤维素材料在糖化之前经预处理。

[0778] [138]段134-137任一项的方法,其中所述酶组合物包含一种或多种(几种)选自下组的酶:纤维素酶、半纤维素酶、棒曲霉素、酯酶、漆酶、木质素分解酶、果胶酶、过氧化物酶、蛋白酶和膨胀素。

[0779] [139]段138的方法,其中所述纤维素酶是一种或多种(几种)选自下组的酶:内切葡聚糖酶、纤维二糖水解酶和 β -葡萄糖苷酶。

[0780] [140]段138的方法,其中所述半纤维素酶是一种或多种(几种)选自下组的酶:木聚糖酶、乙酰木聚糖酯酶、阿魏酸酯酶、阿拉伯呋喃糖苷酶、木糖苷酶和葡糖醛酸糖苷酶。

[0781] [141]段134-140任一项的方法,其中发酵产物是醇、有机酸、酮、氨基酸或气体。

[0782] [142]一种洗涤剂组合物,其包含段1-96任一项的多肽,和表面活性剂。

[0783] [143]段142的组合物,进一步包含一种或多种(几种)选自下组的酶:淀粉酶、阿拉伯糖酶、糖酶、纤维素酶、角质酶、半乳聚糖酶、半纤维素酶、漆酶、脂质酶、甘露聚糖酶、氧化酶、果胶酶、蛋白酶和木聚糖酶。

[0784] [144]段142或143的方法,其配制为棒型或条形剂(bar)、片剂、散剂、颗粒剂、糊剂或液体。

[0785] [145]一种用于清理或洗涤硬表面或衣物的方法,所述方法包括将所述硬表面或所述引物与段142-144任一项的组合物相接触。

[0786] 本文描述和要求保护的本发明并不局限于本文公开的具体方面的范围内,因为这些方面旨在作为本发明几个方面的说明。旨在将任何等同的方面包含于本发明的范围内。实际上,从前面的说明中,除本文所显示和描述的之外,本发明的多种修改对于本领域的技术人员来说是显而易见的。这些修改也旨在落入所附的权利要求的范围内。在冲突的情况下,将以包括定义部分的本公开为准。

序列表

<110> 诺维信股份有限公司(Novozymes, Inc.)
 S.梅于兰(Maiyuran, Suchindra)
 R.克雷默(Kramer, Randall)
 P.哈里斯(Harris, Paul)

<120> 具有纤维素分解增强活性的多肽及编码其的多核苷酸

<130> 11661-CN-PCD

<150> US 61/243,397

<151> 2009-09-17

<150> US 61/243,531

<151> 2009-09-18

<150> US 61/243,543

<151> 2009-09-18

<150> US 61/243,679

<151> 2009-09-18

<160> 44

<170> PatentIn version 3.5

<210> 1

<211> 878

<212> DNA

<213> 土生梭孢壳(*Thielavia terrestris*)

<400> 1

```

atgaagttct cactgggtc tctgtggtt tacggcctct cggctgaggc gcactccatc      60
ttccaggttc gtctcgaca tcacgetcaa ctcggtctgt ggcgtaaggc caaggattaa      120
cacggccggc agagagttc ggtcaacggc caagaccaag gcttctcac cggcctccgc      180
gtccaagca acaacaacce agtcaagat gtaacagcc agaacatgat ttggggccag      240
tcgggctcca agtegeagc cgttataac gtaaggccg gcgacaggat cggctcctc      300
tggcagcatg tcctggcgg cggccagttt tcgggtgacc cggacaacc gatcggccac      360
tcgcacaagg gcccgtgat ggcgtacct gctaaagtc acaatgccg gtccgggagc      420
caaacgggtc tgaagtgta agtagcggc gacgctcagg ggacgggat cggsggctg      480
ctccatccga gactaacacc gtggacaggt tcaagatctg gcaggacggg ttcgatacca      540
gcagcaagac atggggcgtc gacaacctg tcaagaacaa cggctgggtg tacttccacc      600
tgcccgagtg cctcgtccg gccagttat tccfgcgct cagagttctg gcgctgcact      660
cggcgtacca gcagggccg gccagttct accagtcctg cggccagatc aagcttccg      720
gtccgggtc cttcagccg tcccagcgg taagatccc gggegtctac agcgcaccg      780
acctgagcat cctcatcaac atctacggc gcacggggca gcccgacaac ggccgcaagg      840
cttacaacc ccttgaacc gccccgatc cctgtga      878

```

[0001]

<210> 2

<211> 246

<212> PRT

<213> 土生梭孢壳(*Thielavia terrestris*)

<400> 2

```

Met Lys Phe Ser Leu Val Ser Leu Len Ala Tyr Gly Leu Ser Val Glu
1           5           10          15
Ala His Ser Ile Phe Gln Arg Val Ser Val Asn Gly Gln Asp Gln Gly
20          25          30

```

Leu Leu Thr Gly Leu Arg Ala Pro Ser Asn Asn Asn Pro Val Gln Asp
35 40 45
Val Asn Ser Gln Asn Met Ile Cys Gly Gln Ser Gly Ser Lys Ser Gln
50 55 60
Thr Val Ile Asn Val Lys Ala Gly Asp Arg Ile Gly Ser Leu Trp Gln
65 70 75 80
His Val Ile Gly Gly Ala Gln Phe Ser Gly Asp Pro Asp Asn Pro Ile
85 90 95
Ala His Ser His Lys Gly Pro Val Met Ala Tyr Leu Ala Lys Val Asp
100 105 110
Asn Ala Ala Ser Ala Ser Gln Thr Gly Leu Lys Trp Phe Lys Ile Trp
115 120 125
Gln Asp Gly Phe Asp Thr Ser Ser Lys Thr Trp Gly Val Asp Asn Leu
130 135 140
Ile Lys Asn Asn Gly Trp Val Tyr Phe His Leu Pro Gln Cys Leu Ala
145 150 155 160
Pro Gly Gln Tyr Leu Leu Arg Val Glu Val Leu Ala Leu His Ser Ala
165 170 175
Tyr Gln Gln Gly Gln Ala Gln Phe Tyr Gln Ser Cys Ala Gln Ile Asn
180 185 190
Val Ser Gly Ser Gly Ser Phe Ser Pro Ser Gln Thr Val Ser Ile Pro
195 200 205
Gly Val Tyr Ser Ala Thr Asp Pro Ser Ile Leu Ile Asn Ile Tyr Gly
210 215 220
[0002] Ser Thr Gly Gln Pro Asp Asn Gly Gly Lys Ala Tyr Asn Pro Pro Gly
225 230 235 240
Pro Ala Pro Ile Ser Cys
245

<210> 3
<211> 1253
<212> DNA
<213> 土生梭孢壳(Thielavia terrestris)

<400> 3
atgaggacga cattgcgcgc cgcgttgcca gccttcgctg cgcaggaagt ggcaggccat 60
gcatattcc aacagctctg ggtggacggc accgaetata tacgtgctcc ccttttcett 120
ttgtgittgc ccatactoga ttgataaccc gaggccatcc aatgctgaact citacagcac 180
ggctcctcct gcgtccgcat gccgctgtcg aactcgcccg teacgaacgt cggcagcagg 240
gacatgatct gcaacgccgg cacgcgcccc gtcagcggga agtgcccct caagccggc 300
ggcaccgtga cggttgagat gcaccaggtg gctgatttc ctgagcgtcc tattctcctc 360
ggaagccct ttccatcct ttgccctggc taaccctcc gccctfecca geaacccggg 420
gatcggctgt gtaacaacga agccatcggc ggcgccact ggggaccgt gcaggtgtac 480
ctcagcaagg tggaggacgc gagcacggcg gacgggtcga eggctggtt caagatctc 540
gggacacgt ggiccaagaa ggcgggcagc tcggtgggg acgacgacaa ciggggcag 600
cgcgacctca acgctgtctg cggcaagatg caggtaaga tcccggcgga catcccgtc 660
ggcgactacc tctcgggc ggaggcctg gcctgcaca cggcgggcca ggtggcggc 720
gcgagttct acatgagctg ctaccagatc accgtgtcgg gggcggcag cgcagcccg 780

gccaccgtca agtfcaccgg cgcctacagc gccaacgacc cgggcaatcca catcaacatc 840
 caccgcggccg tgtccaacta cgtcgcgcgc gccccggccg tetattcccg cggcaacgacc 900
 aaggctggccg ggtccgggtg ccaaggctgc gagaacacgt gcaaggctcg ctcgtcgcgc 960
 acggcgacgg cgcctctggg caagagcggc gccggctccg acggcggcgc tgggaccgac 1020
 ggcgggtctt cgtcttcgag ccccgacacg gccagcgggt gcagcgtgca ggcctacggg 1080
 cagtcggcgc ggaacgggta ctccggcttc acccagtcgc cggtaagttc ggggtcgtct 1140
 gctctttgta ggaacatccg agaggcttgg ctgacgagge gttgtttag cccggclata 1200
 cttgcaagge ggtctctccg ccgtactatt cgcagtgctc cccttctct tag 1253

<210> 4
 <211> 334
 <212> PRT
 <213> 土牛梭孢壳(*Thielavia terrestris*)
 <400> 4

[0003]

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

275 280 285
 Asp Thr Gly Ser Ala Cys Ser Val Gln Ala Tyr Gly Gln Cys Gly Gly
 290 295 300
 Asn Gly Tyr Ser Gly Cys Thr Gln Cys Ala Pro Gly Tyr Thr Cys Lys
 305 310 315 320
 Ala Val Ser Pro Pro Tyr Tyr Ser Gln Cys Ala Pro Ser Ser
 325 330

<210> 5
 <211> 798
 <212> DNA
 <213> 土生梭孢壳(*Thielavia terrestris*)

<400> 5
 atgaagctga gcgttgccat cgcctgctcg cgtctggctc ttgccgagac tcactgtgag 60
 tgcctcgtct cactccagct actgcgaagc ttgctgacga tggctccctag aaaccttccc 120
 cagcatcgga aacaccctg actggcagta tgtgctgatt acaacgaact accagagcaa 180
 cgggcctgtg acggacgtca cctcggatca aattcggctc tacgaacgga acccaggcac 240
 gggagcgcag ggcatataca acgtcaccgc cggccagacc atcaactaca acgcgaagcc 300
 gtccatctcc caccctgggc ccatgtctt ctacattgct aaggctccc cggcccaaac 360
 cctctcgcacc tgggacggta agggggctgt gtggaccaag atctaccagg acatgcccac 420
 gttcggcagc agcctgacct ggcccacct gggttaagaat tctaccctg gaaatgaacg 480
 cacatttcca cagatctaac atggectaca ggcaccaagt ctgtcccct caccatcct 540
 ccttgcctcc agaacggcga ttacctctg cgagccgagc acatcctct acacagcgg 600
 agcagcgtcg gtggcgcca gttctacctc tegtgcgcc agcttactgt cagcggcgcc 660
 actgacctt ggaaccccaa gaaccggctc tcttcccgc gccttacaa ggcaacagac 720
 ccggcatct tgateaacat ctactacccc gtgccgaca gctactgcc gcccgcccc 780
 ccgctgaga cgtgctaa 798

[0004]

<210> 6
 <211> 227
 <212> PRT
 <213> 土生梭孢壳(*Thielavia terrestris*)

<400> 6
 Met Lys Leu Ser Val Ala Ile Ala Val Leu Ala Ser Ala Leu Ala Glu
 1 5 10 15
 Ala His Tyr Thr Phe Pro Ser Ile Gly Asn Thr Ala Asp Trp Gln Tyr
 20 25 30
 Val Arg Ile Thr Thr Asn Tyr Gln Ser Asn Gly Pro Val Thr Asp Val
 35 40 45
 Thr Ser Asp Gln Ile Arg Cys Tyr Glu Arg Asn Pro Gly Thr Gly Ala
 50 55 60
 Gln Gly Ile Tyr Asn Val Thr Ala Gly Gln Thr Ile Asn Tyr Asn Ala
 65 70 75 80
 Lys Ala Ser Ile Ser His Pro Gly Pro Met Ser Phe Tyr Ile Ala Lys
 85 90 95
 Val Pro Ala Gly Gln Thr Ala Ala Thr Trp Asp Gly Lys Gly Ala Val
 100 105 110
 Trp Thr Lys Ile Tyr Gln Asp Met Pro Lys Phe Gly Ser Ser Leu Thr

115 120 125
 Trp Pro Thr Met Gly Ala Lys Ser Val Pro Val Thr Ile Pro Arg Cys
 130 135 140
 Leu Gln Asn Gly Asp Tyr Leu Leu Arg Ala Glu His Ile Ala Leu His
 145 150 155 160
 Ser Ala Ser Ser Val Gly Gly Ala Gln Phe Tyr Leu Ser Cys Ala Gln
 165 170 175
 Leu Thr Val Ser Gly Gly Ser Gly Thr Trp Asn Pro Lys Asn Arg Val
 180 185 190
 Ser Phe Pro Gly Ala Tyr Lys Ala Thr Asp Pro Gly Ile Leu Ile Asn
 195 200 205
 Ile Tyr Tyr Pro Val Pro Thr Ser Tyr Ser Pro Pro Gly Pro Pro Ala
 210 215 220
 Glu Thr Cys
 225

<210> 7
 <211> 977
 <212> DNA
 <213> 土生梭孢壳(Thielavia terrestris)

[0005]

<400> 7
 atgaagctgt catcceagct cgcgcgccctc acgctggccg cggcctccgt gtcaggccac 60
 taactcttcc agcagattgc ceatggcggc accnagttcc cacccttaaga gtacatccga 120
 agaaacacga actataacag cctgtcacc agtctctcgt cgaacgacct gcgatgcaac 180
 gtaggcgggc agacggctgg caacacgacc gtctctgacg tgaagcgggg cgactccttc 240
 acctctact cggacgtggc cgtgtaccac caggggccca tctcactgtg cgtgccccgg 300
 gccaaacttg atcagtccca agcggactgt ccgctcgctt ggataaccac aattgactga 360
 cagcccgcac agctacatgt ccaaggctcc cggctccgtc gtggactacg acggtcccg 420
 cgactggttc aagatccacg actggggccc gaccttcagc aacggccagg cctcgtggcc 480
 gctgcggggt gctcccttc ccttccctc ccccttctc ccccttctc ccccccttc 540
 ccccccttc tgtctggfag cagccctgc tgaagtcctc gttagcaact accagtacaa 600
 catcccagac tgcattccga acggcgagta cctgctgcgc atccagtcgc tggcgatcca 660
 caaccgggac gccacgcgc agttctacat cagctgcgcg caggtccggg tctcgggcgg 720
 cggcagcgcc tccccctccc caacggccaa gatccccgc gggttaagg cagccgatcc 780
 csggataacc gcgaatgtga gtgccctatg ttccttgcgc tctttgttc ttgctccttg 840
 ctggcggtgc ttgaacgcta cggcctgtgg agggagggat gcatggatga ataggatgct 900
 gactgatggt gggacaccag attacaata acitccactc gtatacggtg ccgggtccgg 960
 cggcttcca gtgctag 977

<210> 8
 <211> 223
 <212> PRT
 <213> 土生梭孢壳(Thielavia terrestris)

<400> 8
 Met Lys Leu Ser Ser Gln Leu Ala Ala Leu Thr Leu Ala Ala Ser
 1 5 10 15
 Val Ser Gly His Tyr Ile Phe Glu Gln Ile Ala His Gly Gly Thr Lys

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

[0006]

<210> 9
 <211> 1107
 <212> DNA
 <213> 土牛梭孢壳(Thielavia terrestris)

<400> 9	
atgccttctt tcgcctccaa gactctcctt tccaccctgg cgggtgccc atccgtggcc	60
gccaccgggc acgtgtgcaa catcgtcacc aaccgggtct cgtaccaggg ttacgatecg	120
acctctctcc cttacaigca gaaccgcgcc atcgtgtctg gctggactgc cggcgacacg	180
gacnacggct ttgttgcccc ggatgccttc gccagtggeg atatcatctg ccacaagAAC	240
gccaccaacg ccaaggccca cgcctgtgtc gccgcgggag acaagatctt catccagttg	300
aacacatggc ccgagtccca ccaaggcccc gtcategact acctcgcgag ctgcggcagc	360
gcgtctctgg agaccctcga caagaccaag ctctagttct tcaagatcga ctaggtctgg	420
ctggtctcag gcagctctgg ccccggtgtg tggggtctcg accagctcat cggcaacaac	480
aactcgttgg tcgtctagat cccgccacc atcgcgccgg gaaactactg cctctcctac	540
gagatcatct cgtctcacag cgcctgaaac gccgacggcg ccagaaacta cccgcagttc	600
ttcaacctgc agatcaccgg caccggcacc gccacccctt cggcgtctcc cggcaacctg	660
ctctacaccg cgaccgacce gggcatctct gtcaaatctt acagcgcctc gatcacctac	720
acctctccgg ggccggccct catctccggc gccgtctcga tctccagtc ctctctccgc	780
atcaccgctt cggcaccgc cctgaccggc tctgccacc caccgcgcgc cggcctctct	840
accacaactt ccaccacca cgcctcggct gctctactct ctctctctgc tctctctgtt	900
acttccacca ccaccaccag cgcctcgggc gtgtctcaga cctctctctc ctctctctcc	960

gccccgtect etgccgcgcg cgcgcgccacc accaccgcgg etgccagcgc ccgcccgcacc 1020
 ggetgtect etggccgcgc caggaagcag cgcgcgccgc acgcgcggga taiggtggft 1080
 gcgcgagggg ctgaggagge aaactga 1107

<210> 10
 <211> 368
 <212> PRT
 <213> 土牛梭孢壳(*Thielavia terrestris*)

<400> 10

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

[0007]

325 330 335
 Ala Arg Pro Thr Gly Cys Ser Ser Gly Arg Ser Arg Lys Gln Pro Arg
 340 345 350
 Arg His Ala Arg Asp Met Val Val Ala Arg Gly Ala Glu Ala Asn
 355 360 365

<210> 11
 <211> 993
 <212> DNA
 <213> 土生梭孢壳(Thielavia terrestris)

<400> 11
 atgccgcccg cactcccica actcccaacc acggtccctga cegcccctcac cctcggttcc 60
 accgcactcg cccactcaca ectcgcttac attatcgtta acggcaagct ctaccagggc 120
 ttcgaccctc gccccgacca gcccaactac ccttcccggg tcgggtgttc caccggcctc 180
 gtcgacgacg gcttcgtcac gccggccaac tacctccacc cggacatcat ttgccacatc 240
 gccggacca gcccgcccg ccacgcgcc gtgcgcccg gcgaccgat ccaegtcag 300
 tggaaaggct ggcggctgg ccacatcggc cccgtgtgt cgtaccctgc ccgctgcgag 360
 tcggacacgg gctgcacggg ccagaacaag accgcctgc ggtggacca gatcgaagac 420
 tccagcccga ccatcgagaa cgtcgccgc gccggcacc agggcgagg cacccggc 480
 aagcctggg ccaccgacgt gctgatgcc gccacaaca gctggcaggt cgcctgccc 540
 gccggctgc gaaccgggc gtacgtgtg cgcaacgaga tcctcgcgt gcaactacgc 600
 gccggaaga accggcgca gaactatcc cctcgcata acctgtggg gacgccagt 660
 gglgalaala gtatgtgtgc tgcaacgac gccggtgta cggcggggg lctgcagaig 720
 gatcgatag accgcgccc gtctacaag gagaacgac cggcgctgt ggtcaatgc 780
 accgcccgc tctctcgtg tctctgccc gggccgacc tggcgccgg ccgccgccg 840
 gtgccgtac cgcagcagag cccgagcgt tcgacggcg cgggacgcc cgtctcgtt 900
 accaaggacta gccgacggc gcctacacg ggcgccatg acccgacgt tccggcgagg 960
 atgaaggga ggggtatga tcggcggt tag 993

[0008]

<210> 12
 <211> 330
 <212> PRT
 <213> 土生梭孢壳(Thielavia terrestris)

<400> 12
 Met Pro Pro Ala Leu Pro Gln Leu Leu Thr Thr Val Leu Thr Ala Leu
 1 5 10 15
 Thr Leu Gly Ser Thr Ala Leu Ala His Ser His Leu Ala Tyr Ile Ile
 20 25 30
 Val Asn Gly Lys Leu Tyr Gln Gly Phe Asp Pro Arg Pro His Gln Ala
 35 40 45
 Asn Tyr Pro Ser Arg Val Gly Trp Ser Thr Gly Ala Val Asp Asp Gly
 50 55 60
 Phe Val Thr Pro Ala Asn Tyr Ser Thr Pro Asp Ile Ile Cys His Ile
 65 70 75 80
 Ala Gly Thr Ser Pro Ala Gly His Ala Pro Val Arg Pro Gly Asp Arg
 85 90 95
 Ile His Val Gln Trp Asn Gly Trp Pro Val Gly His Ile Gly Pro Val

	100	105	110
Leu Ser Tyr	Leu Ala Arg Cys Glu Ser Asp Thr Gly Cys Thr Gly Gln		
	115	120	125
Asn Lys Thr	Ala Leu Arg Trp Thr Lys Ile Asp Asp Ser Ser Pro Thr		
	130	135	140
Met Gln Asn Val	Ala Gly Ala Gly Thr Gln Gly Glu Gly Thr Pro Gly		
	145	150	155
Lys Arg Trp	Ala Thr Asp Val Leu Ile Ala Ala Asn Asn Ser Trp Gln		
	165	170	175
Val Ala Val	Pro Ala Gly Leu Pro Thr Gly Ala Tyr Val Leu Arg Asn		
	180	185	190
Glu Ile Ile	Ala Leu His Tyr Ala Ala Arg Lys Asn Gly Ala Gln Asn		
	195	200	205
Tyr Pro Leu	Cys Met Asn Leu Trp Val Asp Ala Ser Gly Asp Asn Ser		
	210	215	220
Ser Val Ala	Ala Thr Thr Ala Ala Val Thr Ala Gly Gly Leu Gln Met		
	225	230	235
Asp Ala Tyr	Asp Ala Arg Gly Phe Tyr Lys Glu Asn Asp Pro Gly Val		
	245	250	255
Leu Val Asn	Val Thr Ala Ala Leu Ser Ser Tyr Val Val Pro Gly Pro		
	260	265	270
Thr Val Ala	Ala Gly Ala Thr Pro Val Pro Tyr Ala Gln Gln Ser Pro		
	275	280	285
Ser Val Ser	Thr Ala Ala Gly Thr Pro Val Val Val Thr Arg Thr Ser		
	290	295	300
Glu Thr Ala	Pro Tyr Thr Gly Ala Met Thr Pro Thr Val Ala Ala Arg		
	305	310	315
Met Lys Gly	Arg Gly Tyr Asp Arg Arg Gly		
	325	330	

[0009]

<210> 13
 <211> 1221
 <212> DNA
 <213> 土生梭孢壳(*Thielavia terrestris*)

<400> 13
 atgaagacat tcaaccgect cctggccgca gccggccctg tcgccggcca tggatatgtc 60
 gacaacgcca ccattggcgg ccagttttat caggtactct accgcttcaac ccaaggctccg 120
 ctggccacaa ctctataggt gtcataaait aacaagccac cgicccgcag ttetatcagg 180
 tgtgtcgtct accgacctg tggctccgfc tcagcaagcc acicacacgc ccatgatccc 240
 ctagccttac gtcgaccctt atttagcaac cttggcactg agtatttatt gtcccaaata 300
 ttgagctgaa ctgcacctcc ctagaatccc gcggtgctaa cattctttca gcccgacagg 360
 gctctctgat ccctccggg caacggcccg gtcacggacg tcactctcat cgaactgcag 420
 tgcaacgcca attccacccc ggccaagctc cagccactg ccgctgcccg ctgggactg 480
 attctccgt ggagctctg gctgagctg cactgtggcc ccgtctcaac ctacatggcc 540
 cgtgccccg acacgggctg ccaggactgg atgcccggca ctctctagga gcccatcttg 600
 caccatctcc atttcaaccg gccacacgca ctgacctata tctctgteta cccctgcagt 660
 gcggtctggt tcaagatcaa ggaggcggc cgcgacggca ctccaacac ctgggcccac 720
 gtactgttac cccgtcccag agagccaaag ccccccttc aacaagcaa acatctcaat 780

agcccagacc tacgcaactaa cccctctcct tccccctcga aaacacagac cccctgatg 840
 acggcgccca cctegtanac gtacacgata cectcctgcc tgaagaaggc ctactacctg 900
 gtccgccaag agatcattgc gctgcacgcc gectaacctt accccggcgc gcagttctac 960
 ccggcctgcc accagctcaa cgtcacgggc ggcgggtcca cegtaccgic gagcggcctg 1020
 gtggccttic ccggggcgta caaggcagc gaccccggga ttacgtacga tgcgtataaa 1080
 ggiggattgg ctggttggcc cagatcttgg tgaatgggga atgtggtgat gaggtttatt 1140
 atttgggatc ccgtggctaa cgtaacctcg ggtgtagcgc aaactacca gattcctggg 1200
 ccggcgtctt ttacttgcig a 1221

<210> 14
 <211> 236
 <212> PRT
 <213> 土生梭孢壳(*Thielavia terrestris*)

<<400> 14

[0010]

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

<210> 15
 <211> 933
 <212> DNA
 <213> 土生梭孢壳(*Thielavia terrestris*)


```

<400> 15
atggccttgc tgccttggc agccttggcc attctggccg ggccgctca tcccacggc      60
ggcctcgcca actacacagt gggcaacacc tggtataggg ggtagcgaag gggggcaccg      120
acaacgcctg cttagtaact ccaccatttc gagcgggcta acaccgggcg cagctacgac      180
cccttcacgc cgccggccga ccagattggc cagccgtgga tgatecaacg cgcgtgggac      240
tcgatcgacc cgacltcag cgtcaacgac aaggcgcctg cctgeaacac cccggccacg      300
gcgcgcacct cttacattcc caccgcgcg ggcgagaaca tcacggccgt gtactggtac      360
tggtgcacc cggtagggcc catgacggcg tggctggcgc ggtgcgacgg cgaactgccg      420
gagcccgacg tcaacgagge gcgcctggc aagatctggg agcccgccct gctcagcggg      480
ccgaacctgg ccgaggccat gtggtaccag aaggcgttcc agaactggga cggcagcccc      540
gacctgtggc ccgtcacgat cccggccggg ctgaagagcg gcctgtacat gatecggcac      600
gagactttgt cgateccagt caggataaa ccgcagtttt atcccagatg taccgatctg      660
aatgtgaccg gggatgggga cctgctgcgc ccgatgagt ttttggtaa gttcccgggc      720
gcttacaacg aagatagtga gtgaaacgcg aagcttcggt agccattggg ttgcctgat      780
ggaggttaga cccgtcgate aagatcaata tctactcgga ccagtagcgc aatacaacgg      840
tgagtgtaac aggtcgagca aaaccaanca gatcccgatg acgatgata tcagaattac      900
acaattcccg gagggccgat atgggatggg tga                                     933
    
```

```

<210> 16
<211> 250
<212> PRT
<213> 土生梭孢壳(Thielavia terrestris)
    
```

[0011]

```

<400> 16
Met Ala Leu Leu Leu Ala Gly Leu Ala Ile Leu Ala Gly Pro Ala
1      5      10     15
His Ala His Gly Gly Leu Ala Asn Tyr Thr Val Gly Asn Thr Trp Tyr
20     25     30
Arg Gly Tyr Asp Pro Phe Thr Pro Ala Ala Asp Gln Ile Gly Gln Pro
35     40     45
Trp Met Ile Gln Arg Ala Trp Asp Ser Ile Asp Pro Ile Phe Ser Val
50     55     60
Asn Asp Lys Ala Leu Ala Cys Asn Thr Pro Ala Thr Ala Pro Thr Ser
65     70     75     80
Tyr Ile Pro Ile Arg Ala Gly Glu Asn Ile Thr Ala Val Tyr Trp Tyr
85     90     95
Trp Leu His Pro Val Gly Pro Met Thr Ala Trp Leu Ala Arg Cys Asp
100    105    110
Gly Asp Cys Arg Asp Ala Asp Val Asn Glu Ala Arg Trp Phe Lys Ile
115    120    125
Trp Glu Ala Gly Leu Leu Ser Gly Pro Asn Leu Ala Glu Gly Met Trp
130    135    140
Tyr Gln Lys Ala Phe Gln Asn Trp Asp Gly Ser Pro Asp Leu Trp Pro
145    150    155    160
Val Thr Ile Pro Ala Gly Leu Lys Ser Gly Leu Tyr Met Ile Arg His
165    170    175
Glu Ile Leu Ser Ile His Val Glu Asp Lys Pro Gln Phe Tyr Pro Glu
180    185    190
    
```

Cys Ala His Leu Asn Val Thr Gly Gly Gly Asp Leu Leu Pro Pro Asp
 195 200 205
 Glu Phe Leu Val Lys Phe Pro Gly Ala Tyr Lys Glu Asp Asn Pro Ser
 210 215 220
 Ile Lys Ile Asn Ile Tyr Ser Asp Gln Tyr Ala Asn Thr Thr Asn Tyr
 225 230 235 240
 Thr Ile Pro Gly Gly Pro Ile Trp Asp Gly
 245 250

<210> 17
 <211> 1584
 <212> DNA
 <213> 土生梭孢壳(*Thielavia terrestris*)

[0012]

<400> 17
 atgatgccgt cccctgttgc ctctcaatg ggtctggoga ccgecttgcg cfcgctgtcc 60
 acagcacata ccgctttcac cacgcttttc atcaacggcg tegaccaagg ggacgggacc 120
 tgcatecgca tggccaagaa gggcagcgtt tgcaccatc ccattgctgg tggcctcgac 180
 agcccagaca tggcttggg tatgccctct gcgttcccc tgcgagagct ttctctgagc 240
 taacccaatg ccgcgittgc cagcccgaga cggacaacaa gccgtggcat tcacctgccc 300
 agcccggcg ggctccaagt tgagcttga gtcccgcatg tgggcccagc cctctcagcc 360
 cggctctatc gaccatccc aactcggctc gacggcaatc taccicnaac aagctccaa 420
 catcagctec gactcggctg ccggccctgg ctggttcaag atctacgcc aggctacga 480
 cacagccgc nagaagtgg ccacagagaa gctcctgac aacggcggcc tgcctgagcat 540
 cgagcttccg cccactctgc cggcgggata ctacctgcc cgcagcgaga tegtaccat 600
 ccagaacgtc accaacgacc acgtcgacc gcagttctac gttggctggc cacagctctt 660
 cgtccagggg cctccagcca ccccaccgt ccccccagac agactcgtct caatcccggg 720
 caacgtccat gctccgacc cggggctgac ctcaacatc tggcgcgacg accctccaa 780
 gacggctac accgtctgc gcccgcccc ctctccccc accgcccgc ccaccctcac 840
 ctccaccaac accaacgggc agcaacaaca acaacagcaa caggcgataa agcagacgga 900
 cggctgctac cccgcgact gccagctcan gaacccaac tggctcggcg ccgaggtgcc 960
 cgcgtacgc gacgagggcg gctgctgggc gctctcgcc gactgcttcg cccagctgga 1020
 cgcctgctac acgtcggcgc cgcaccggg cagcccgggc tgcaggctgt gggaggactg 1080
 gtgcaccgc attcagcagg gctgccgcgc gggcggtgg cggggccgc cgcctttca 1140
 tggggagggg gcagcagcgg aggtgtgaac ggttcgggga cgggtggcgg tgggtgtgt 1200
 ggtggtggg gcactggctc ttcttcggt ctctccccga cggagacggc ctctgtggc 1260
 cggggggcgc caagaatagc tgcctggcc gctcgggag gcgggacagg agacatggtt 1320
 gaagagttt tctctttta tgggacct tgcagcggct ggcgacggag ccgtggtgtt 1380
 ggttcgattc ttgcaggct tatcttctat gtccttctt cacttttgag accgaggcga 1440
 gcccctcgag tccatttact tctcttccac ctgtacctc acttctgta ccaggaacc 1500
 agtggttct ataatcgct gagcattaaa ctaggcatat ggccaagcaa aatgtcgct 1560
 gatgtagcc attactgaa ataa 1584

<210> 18

<211> 478
 <212> PRT
 <213> 土生梭孢壳(*Thielavia terrestris*)
 <400> 18
 Met Met Pro Ser Leu Val Arg Phe Ser Met Gly Leu Ala Thr Ala Phe
 1 5 10 15
 Ala Ser Leu Ser Thr Ala His Thr Val Phe Thr Thr Leu Phe Ile Asn
 20 25 30
 Gly Val Asp Gln Gly Asp Gly Thr Cys Ile Arg Met Ala Lys Lys Gly
 35 40 45
 Ser Val Cys Thr His Pro Ile Ala Gly Gly Leu Asp Ser Pro Asp Met
 50 55 60
 Ala Cys Gly Arg Asp Gly Gln Gln Ala Val Ala Phe Thr Cys Pro Ala
 65 70 75 80
 Pro Ala Gly Ser Lys Leu Ser Phe Glu Phe Arg Met Trp Ala Asp Ala
 85 90 95
 Ser Gln Pro Gly Ser Ile Asp Pro Ser His Leu Gly Ser Thr Ala Ile
 100 105 110
 Tyr Leu Lys Gln Val Ser Asn Ile Ser Ser Asp Ser Ala Ala Gly Pro
 115 120 125
 Gly Trp Phe Lys Ile Tyr Ala Glu Gly Tyr Asp Thr Ala Ala Lys Lys
 130 135 140
 Trp Ala Thr Glu Lys Leu Ile Asp Asn Gly Gly Leu Leu Ser Ile Glu
 145 150 155 160
 Leu Pro Pro Thr Leu Pro Ala Gly Tyr Tyr Leu Ala Arg Ser Glu Ile
 165 170 175
 Val Thr Ile Gln Asn Val Thr Asn Asp His Val Asp Pro Gln Phe Tyr
 180 185 190
 Val Gly Cys Ala Gln Leu Phe Val Gln Gly Pro Pro Thr Thr Pro Thr
 195 200 205
 Val Pro Pro Asp Arg Leu Val Ser Ile Pro Gly His Val His Ala Ser
 210 215 220
 Asp Pro Gly Leu Thr Phe Asn Ile Trp Arg Asp Asp Pro Ser Lys Thr
 225 230 235 240
 Ala Tyr Thr Val Val Gly Pro Ala Pro Phe Ser Pro Thr Ala Ala Pro
 245 250 255
 Thr Pro Thr Ser Thr Asn Thr Asn Gly Gln Gln Gln Gln Gln Gln
 260 265 270
 Gln Ala Ile Lys Gln Thr Asp Gly Val Ile Pro Ala Asp Cys Gln Leu
 275 280 285
 Lys Asn Ala Asn Trp Cys Gly Ala Glu Val Pro Ala Tyr Ala Asp Glu
 290 295 300
 Ala Gly Cys Trp Ala Ser Ser Ala Asp Cys Phe Ala Gln Leu Asp Ala
 305 310 315 320
 Cys Tyr Thr Ser Ala Pro Pro Thr Gly Ser Arg Gly Cys Arg Leu Trp
 325 330 335
 Glu Asp Trp Cys Thr Gly Ile Gln Gln Gly Cys Arg Ala Gly Arg Trp
 340 345 350
 Arg Gly Pro Pro Phe His Gly Glu Gly Ala Ala Ala Glu Thr Ala
 355 360 365
 Ser Ala Gly Arg Gly Gly Ala Arg Ile Ala Ala Val Ala Gly Cys Gly

[0013]

370 375 380
 Gly Gly Thr Gly Asp Met Val Glu Glu Val Phe Leu Phe Tyr Trp Asp
 385 390 395 400
 Ala Cys Ser Gly Trp Arg Arg Ser Arg Gly Gly Gly Ser Ile Leu Ala
 405 410 415
 Arg Leu Ile Leu His Val Leu Leu Pro Leu Leu Arg Pro Arg Arg Ala
 420 425 430
 Pro Arg Val His Leu Leu Leu Phe His Leu Tyr Leu Asn Phe Cys Tyr
 435 440 445
 Pro Gly Thr Ser Gly Phe Tyr Asn Arg Leu Ser Ile Lys Leu Gly Ile
 450 455 460
 Trp Pro Ser Lys Met Ser Pro Asp Val Ala His Tyr Val Lys
 465 470 475

<210> 19
 <211> 868
 <212> DNA
 <213> 土生梭孢壳(Thielavia terrestris)

[0014]

<400> 19
 atgcagctcc tcgtgggett getgettgea gcgctggctg ctcgagcaca ttgtatttet 60
 acccttttcc gctgctctcc cagctcaag gcaagaagac gcacgcagca gctaaeggac 120
 cctatcagac acatttccca gactctgtgt aaatgggcag cccgaggaca aggactggtc 180
 ggttacggcg atgaccaaga acgcgcagag caagcagga gtccaggacc cgaccagtcc 240
 cgacattcgc tctacacgt cgcagacggc gcctaacgtg gctacggtcc ctgcccggagc 300
 caccgtccat taatattcga ctacagagal caaccaaccg gcccggaccg agtactacct 360
 cgccaaggtt cccggcgggt cgtcgcccaa gacgtgggac gggtcagggg ccgtctggtt 420
 caagatctcg accaccatgc ctactttgga caacaacaag cagcttctct gzcccgaatca 480
 gactaggaac aattcccgtt ccaatcttct atttggcctt gactacggc cgatttcaatg 540
 ggagagaccg ttactgaag gggcaaccca accttcaata gacacgtaca cgacggtcaa 600
 caccgaccatc ccccccata cgcaccgtgg ggaatacctc ctccgggtcg agcagatcgc 660
 gctgcaactg gcctcgcagc ccaacgggac tcagttctac ctggcctgct cgcagatcca 720
 gattacgggc ggcggcaacg gcaaccccgg cccgctagtc gcgttgcccg gzcgtacaa 780
 gagcaacgac cccggcattt tggtaacat ctactctatg cagcccggcg attacaagcc 840
 gcccgggccc ccggtgtgga gtgctga 868

<210> 20
 <211> 230
 <212> PRT
 <213> 土生梭孢壳(Thielavia terrestris)

<400> 20
 Met Gln Leu Leu Val Gly Leu Leu Leu Ala Ala Val Ala Ala Arg Ala
 1 5 10 15
 His Tyr Thr Phe Pro Arg Leu Val Val Asn Gly Gln Pro Glu Asp Lys
 20 25 30
 Asp Trp Ser Val Thr Arg Met Thr Lys Asn Ala Gln Ser Lys Gln Gly
 35 40 45
 Val Gln Asp Pro Thr Ser Pro Asp Ile Arg Cys Tyr Thr Ser Gln Thr
 50 55 60

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

<210> 21
<211> 1068
<212> DNA
<213> 土生梭孢壳(*Thielavia terrestris*)

[0015]

<400> 21
atgaagctgt acctggcgcg cttcttaggc gccctgcera cccccggagc gttcgetcat 60
cgtaggttcc cegtctatct ccctaggggt agcaccacga ctaatttctc gtctgcccc 120
tgtagaatc cacgggattc tacttgtcaa cggcaccgaa acgccggaat ggaatatcgt 180
ccggtaatat ctaccittgt ctctttcttc cacaccaccg ctaacacatc atcagtgacg 240
tggctgggga gggcgcttac gaaccggaaa aataccccaa caccgagttc tttangacgc 300
ccccgcagac ggacateaac aaccggaaca tcacctgcgg caggaaacgg ttctactegg 360
ccagcaagac tgagacggcc gacatactgg ccggtcaga ggtcggcttc cggctctcgt 420
gggacggcaa cggcaagtac ggcgtgttct ggcattcccg gccggggcag atctacctct 480
ctcgtgctcc gaacgacgac ctggaggact acccgggcga cggagactgg ttcaagatcg 540
caaccggcgc ccctgtctcc aataccgagt ggcigtgtg gaacaagcat gactgagcc 600
ccaacattcc tgcaccaate gatccccaac ctggtaacca tggeggcgct cgggatgcaa 660
agagactaac tccagaggaa cctacctagt teaacttca cctccccaa agacagcgcg 720
cggcaagta cctgatgcgc atcgagcagt teatgccc caccgtcgaa tacagccagt 780
ggtactgcaa ctgcgcccac gtcaaatca tcggccccgg cggaggcaag ccgacgggct 840
ttgccaggtt tcccggcacc taacigtgt agatcccgg taagccggac ctaccggaca 900
cagaggcctc gggatagctt getancctt ttgtctct ctctttttct ctccccacta 960
ggcatcaagg tgccgttgaa ccagatcgtc aacagcggag agttgccgca ggaccaactg 1020
aggctgctcg agtacaagcc cccgggcca gcctgttga ctggttga 1068

<210> 22
 <211> 257
 <212> PRT
 <213> 土生梭孢壳(*Thielavia terrestris*)
 <400> 22
 Met Lys Leu Tyr Leu Ala Ala Phe Leu Gly Ala Val Ala Thr Pro Gly
 1 5 10 15
 Ala Phe Ala His Gln Ile His Gly Ile Leu Leu Val Asn Gly Thr Glu
 20 25 30
 Thr Pro Glu Trp Lys Tyr Val Arg Asp Val Ala Trp Glu Gly Ala Tyr
 35 40 45
 Glu Pro Glu Lys Tyr Pro Asn Thr Glu Phe Phe Lys Thr Pro Pro Gln
 50 55 60
 Thr Asp Ile Asn Asn Pro Asn Ile Thr Cys Gly Arg Asn Ala Phe Asp
 65 70 75 80
 Ser Ala Ser Lys Thr Glu Thr Ala Asp Ile Leu Ala Gly Ser Glu Val
 85 90 95
 Gly Phe Arg Val Ser Trp Asp Gly Asn Gly Lys Tyr Gly Val Phe Trp
 100 105 110
 His Pro Gly Pro Gly Gln Ile Tyr Leu Ser Arg Ala Pro Asn Asp Asp
 115 120 125
 Leu Glu Asp Tyr Arg Gly Asp Gly Asp Trp Phe Lys Ile Ala Thr Gly
 130 135 140
 Ala Ala Val Ser Asn Thr Glu Trp Leu Leu Trp Asn Lys His Asp Phe
 145 150 155 160
 Asn Phe Thr Ile Pro Lys Thr Thr Pro Pro Gly Lys Tyr Leu Met Arg
 165 170 175
 Ile Glu Gln Phe Met Pro Ser Thr Val Glu Tyr Ser Gln Trp Tyr Val
 180 185 190
 Asn Cys Ala His Val Asn Ile Ile Gly Pro Gly Gly Gly Thr Pro Thr
 195 200 205
 Gly Phe Ala Arg Phe Pro Gly Thr Tyr Thr Val Asp Asp Pro Gly Ile
 210 215 220
 Lys Val Pro Leu Asn Gln Ile Val Asn Ser Gly Glu Leu Pro Gln Asp
 225 230 235 240
 Gln Leu Arg Leu Leu Glu Tyr Lys Pro Pro Gly Pro Ala Leu Trp Thr
 245 250 255

[0016]

Gly

<210> 23
 <211> 32
 <212> DNA
 <213> 土生梭孢壳(*Thielavia terrestris*)

<400> 23
 actggattta ccatgaagtt ctcactggtg tc

32

<210> 24
 <211> 38
 <212> DNA
 <213> 土生梭孢壳(*Thielavia terrestris*)

<400> 24

	tcacctctag ttaattaatc agcaggagat cggggcgg	38
	<210> 25	
	<211> 37	
	<212> DNA	
	<213> 土生梭孢壳(<i>Thielavia terrestris</i>)	
	<400> 25	
	actggattta ccatgaggac gacattcgcc gccgcgt	37
	<210> 26	
	<211> 38	
	<212> DNA	
	<213> 土生梭孢壳(<i>Thielavia terrestris</i>)	
	<400> 26	
	tcacctctag ttaattaact aagaagaagg ggcgcact	38
	<210> 27	
	<211> 36	
	<212> DNA	
	<213> 土生梭孢壳(<i>Thielavia terrestris</i>)	
	<400> 27	
	actggattta ccatgaaget gagcgtgcc atcgcc	36
	<210> 28	
	<211> 38	
	<212> DNA	
	<213> 土生梭孢壳(<i>Thielavia terrestris</i>)	
	<400> 28	
	tcacctctag ttaattaatt agcacgtctc agccggcg	38
[0017]	<210> 29	
	<211> 36	
	<212> DNA	
	<213> 土生梭孢壳(<i>Thielavia terrestris</i>)	
	<400> 29	
	actggattta ccatgaaget gtcateccag ctcgcc	36
	<210> 30	
	<211> 38	
	<212> DNA	
	<213> 土生梭孢壳(<i>Thielavia terrestris</i>)	
	<400> 30	
	tcacctctag ttaattaact agcactgaan gaccgccg	38
	<210> 31	
	<211> 32	
	<212> DNA	
	<213> 土生梭孢壳(<i>Thielavia terrestris</i>)	
	<400> 31	
	actggattta ccatgccttc tttegcctcc aa	32
	<210> 32	
	<211> 38	
	<212> DNA	
	<213> 土生梭孢壳(<i>Thielavia terrestris</i>)	
	<400> 32	
	tcacctctag ttaattaatc agtttgcctc ctcagecc	38
	<210> 33	

	<211> 32	
	<212> DNA	
	<213> 土生梭孢壳(<i>Thielavia terrestris</i>)	
	<400> 33	
	actggattta ccatgccgcc cgcactccct ca	32
	<210> 34	
	<211> 38	
	<212> DNA	
	<213> 土生梭孢壳(<i>Thielavia terrestris</i>)	
	<400> 34	
	tcacctctag ttaattaact aaccccgcgc atcatacc	38
	<210> 35	
	<211> 36	
	<212> DNA	
	<213> 土生梭孢壳(<i>Thielavia terrestris</i>)	
	<400> 35	
	actggattta ccatgaagac attcacgcgc ctctctg	36
	<210> 36	
	<211> 38	
	<212> DNA	
	<213> 土生梭孢壳(<i>Thielavia terrestris</i>)	
	<400> 36	
	tcacctctag ttaattaatc agcaagtaaa gaccgcgc	38
[0018]	<210> 37	
	<211> 36	
	<212> DNA	
	<213> 土生梭孢壳(<i>Thielavia terrestris</i>)	
	<400> 37	
	actggattta ccatggcctt gctgctcttg gcaggc	36
	<210> 38	
	<211> 38	
	<212> DNA	
	<213> 土生梭孢壳(<i>Thielavia terrestris</i>)	
	<400> 38	
	tcacctctag ttaattaatc acccatccca tatcggcc	38
	<210> 39	
	<211> 36	
	<212> DNA	
	<213> 土生梭孢壳(<i>Thielavia terrestris</i>)	
	<400> 39	
	actggattta ccaigatgcc gtcccttggt cgcctc	36
	<210> 40	
	<211> 38	
	<212> DNA	
	<213> 土生梭孢壳(<i>Thielavia terrestris</i>)	
	<400> 40	
	tcacctctag ttaattaatc aaccaigtct cctgcccc	38
	<210> 41	
	<211> 32	
	<212> DNA	
	<213> 土生梭孢壳(<i>Thielavia terrestris</i>)	

	<400> 41	
	actggattta ccatgcagct cctcgtgggc tt	32
	<210> 42	
	<211> 38	
	<212> DNA	
	<213> 土生梭孢壳(<i>Thielavia terrestris</i>)	
	<400> 42	
	tcacctetag ttaattaatc agccaatcca caccggcg	38
[0019]	<210> 43	
	<211> 36	
	<212> DNA	
	<213> 土生梭孢壳(<i>Thielavia terrestris</i>)	
	<400> 43	
	actggattta ccatgaagct gtacctggcg gecitt	36
	<210> 44	
	<211> 38	
	<212> DNA	
	<213> 土生梭孢壳(<i>Thielavia terrestris</i>)	
	<400> 44	
	tcacctetag ttaattaatc aaccagtcca cagcgtg	38

M K F S L V S L L A Y S L S V E A H S I F Q R V
 1 ATGAAGTTCTCACTGGTGTCTCTGCTGGCTTACGGGCTCTCGGTCGAGGGCCACTCCATCTTCCAGAGAGTC
 S V N G Q D Q G L L T G L R A P S H H N F V Q D
 73 TCGGTCAACGGCCAAAGACCAAGGCTTGTCTACCGGGCTCCGGCTCCAAAGCAACAACACCCAGTGCAGAT
 V N S Q H M I C G Q S G S K S Q T V I N V K A G
 145 GTCAACAGCCAGAACATGATTTCGGGTCAGTCCGGGCTCCAAAGTCGACACCGTTATCAACGTCAGGGCCGGC
 D P I G S L W Q H V I G G A Q F S G D P E H F I
 217 GACAGGATCGGCTGCTCTGGCAGCAATGTCATCGGCGGCGCCAGTTTTCGGGTGACCCGGCAACCCGATC
 A H S H K G P V M A Y L A K V D H A A S A S Q T
 289 GCCCACTCGCACAAGGGCCCGTGTGTCGCTACCTTTCCTAAGGTCCACAAATGCCCGGCTCCGCGAGCCAAAG
 G L K W F K I W Q D G F D T S S K T W G V D H L
 361 GGTCTGAAGTGGTTCAGATCTGGCAGGACGGGTTTCGATACCAGCAGCAGACATGGGGCGTCCGACAACTG
 I K N H G W V Y F H L F Q C L A P G Q Y L L R V
 433 ATCAGAAACAACGGCTGGGTGTAATCCACTGCCGACAGTCCCTGCTCCGGCCASTATCTCCTGGGCGTC
 E V L A L H S A Y Q Q G Q A Q F Y Q S C A Q I H
 505 GAGGTCTGGGCGCTGCACTCGGCGTACCAGCAGGGCCAGGCCAGTTCTACCACTCCCTCCGCCCAGATCAAC
 V S G S G S F S F S Q T V S I F G V Y S A T D P
 577 GTCTCCGGCTCCGGGCTCTTCAGCCCTCCAGACGGTCCAGCATCCCGGGCTTACACCGCCACCCGACCCG
 S I L I H I Y G S T G Q P D H G G K A Y H F P G
 649 AGCATCTCATCAACATCTACGGCAGCACGGGGCAGCCCGACACCGCGGCAAGGCTTACAACCCCCCTGGA
 P A F I S C *
 721 CCGGCCCCGATCTCCTGCTGA

图1

N R T T F A A A L A A F A A Q E V A G H A I F Q
 1 ATGAGGACGACATTCGCGCGCGCTTGGCAGCCTTCCGTCGGCAGGAAGTGGCAAGCCATGCCATCTTCCAA
 Q L W H G S S C V R M P L S H S F V T H V G S R
 73 CAGCTCTGGCACGGCTCCTCCTGCGTCCGCATGCCGCTGTCSAACTCGCCCGTCACGAACCTCGGCAGCAGG
 D M I C H A G T R P V S G K C P V K A G G T V T
 145 GACATGATCTGCAACGCGGGCAGCGGCGCCGTCAGCGGSAAGTSCCCCGTCAAGGCGGGCGGCACCGTGAAG
 V E M R Q Q F G D R S C N N E A I G G A H W G F
 217 GTTGAATGCAACCAACCCGGGATCGGTCGTGTAACAACGAAGCCATCGGCGGCGCCCACTGGGGCCGG
 V Q V Y L S K V E D A S T A D G S T G F K I F
 289 GTGCAGGTGTACCTCAGCAAGGTGGAGBACCGGAGCACGGCGGACGGGTCGACGGGCTGTTCAAGATCTTC
 A D T W S K K A G S S V G D D D H W G T R D L H
 361 GCGGACAGGTGGTCCAAAGAGCGGGCAGCTCGGTTGGGGAGCAAGCAACTGGGGCCCGCGGACCTCAAC
 A C C G K M Q V K I P A D I P S G D Y L L K A E
 433 GCGTCTCGGCAAGATGCAGGTCAAGATCCCGGGGSACTCCCGTCCGGGCGACTACCTGCTCCGGGCGGAG
 A L A L H T A G Q V G G A Q F Y N S C Y Q I T V
 505 GCGCTGCGCTGCACACGGCGGGCCAGBTGGGCGGGCGGBCAGTTCTACATGAGCTGCTACCAGATCACCGTG
 S G G G S A S F A T V E F P G A Y S A N D F G I
 577 TCGGCGGGCGGCGAGCCAGCCCGGCGCCCGTCAAGTTCGCCGGGCGCTACAGCGCCACCGACCCCGGCTC
 H I N I H A A V S N Y V A P G P A V Y S G G T T
 649 CACATCAACATCCACGCGCGCGTGTCCAACTACGTCGCGGCGCGGCGGCGGCTCIBTTCGGGCGGCGACGACC
 K V A G S G C Q G C E H T C K V G S S F T A T A
 721 AAGGTGGCCGGTCCGGTGCAGGCTCGGAGAACACGTGCAAGGTCGGCTCGTCCGCCACGGCGACGGCG
 F S G K S G A G S D G G A G T D G G S S S S F
 793 CCGTCCGGCAAGAGCGGCGCGGTTCCGACGGCGGCGCTGGACCGACGGCGGGTCTTCGCTTCCGAGCCCC
 D T G S A C S V Q A Y G Q C G G N G Y S G C T Q
 865 GACACGGCAGCCTGTCAGCGTGCAGGCTTACGGGCGAGTCCGCGGGGACCGGTACTCCGGTTCACCCAG
 C A P G Y T C K A V S P F Y Y S Q C A F S S *
 937 TGCGGCGCGGCTATACTTGCAGGGGCTCTCCGCGTACTATTGCGAGTGGCGGCTCTCTCTTAS

图2

```

M K L S V A I A V L A S A L A E A H Y T F P G I
1 ATGAAGCTGAGCGTFTSCCATCGCCGTCCTGGCGTCTGCGGAGGCTCACTACACCTTCCCCAGGTC
G M T A D W Q Y V R I T T N Y Q S H G P V T D V
73 GGAAACACCCGCTGACTGGCAGTATGTSCGGATTACAACGAGCTACCAGAGCAACGGGCCGGTGACGGACGTC
T S D Q I R C Y E R N P G T G A Q G I Y N V T A
145 ACCTCGGATCAAATTCGGTGCTACGAAACCGAACCAGGACGGGAGCGCAGGGCATATACAACGTCACCGCC
G Q T I N Y N A K A S I S H P G P H S F Y I A K
217 GGCCAGACCATCAACTACAACGCGAAGGCGTCCATCTCCACCCGGGSCCATGTCTCTACATTGCTAAG
V F A G Q T A A T W D G K G A V W T R I Y Q D H
289 GTTCCCGCCGGCCAAACCGCTGCGACCTGGGACGGTAAGGGGGCTGTGTGGACCAGGATCTACCAGGACAIG
P K F G S S L T W P T M G A K S V P V T I P R C
361 CCCAAGTTGGCAGCAGCCTGACCTGGCCACCATTGGGCGCCAAGTCTGTCCCGTCACCATCCCTCGTTGC
L Q N G D Y L L P A E H I A L H S A S S V G G A
435 CTCCAGAAAGCGGATTACCTTCTGCEAGCCGAGCACATCGCTCTACACAGCGCGAGCAGCGTGGTGGCGCC
Q F Y L S C A Q L T V S G G S G T W R F K N R Y
505 CAGTTCTACCTCTCGTGCSCCCAGCTTACTGTGCAGCGGCGGCACTGGCACCTGGAACCCCAAGAACCAGGTC
S F P G A Y K A T D P G I L I H I Y Y F V P T S
577 TCCTTCCCCGGCGCTTACAAGGCAACAGACCCGGGCACTCTTGTATCAACATCTACTAACCCTGGCCGACGAGC
Y S P P G P A E T C *
649 TACTCGCCGCCCCGSCCCCGCTGAGACCTGCTAA

```

图3

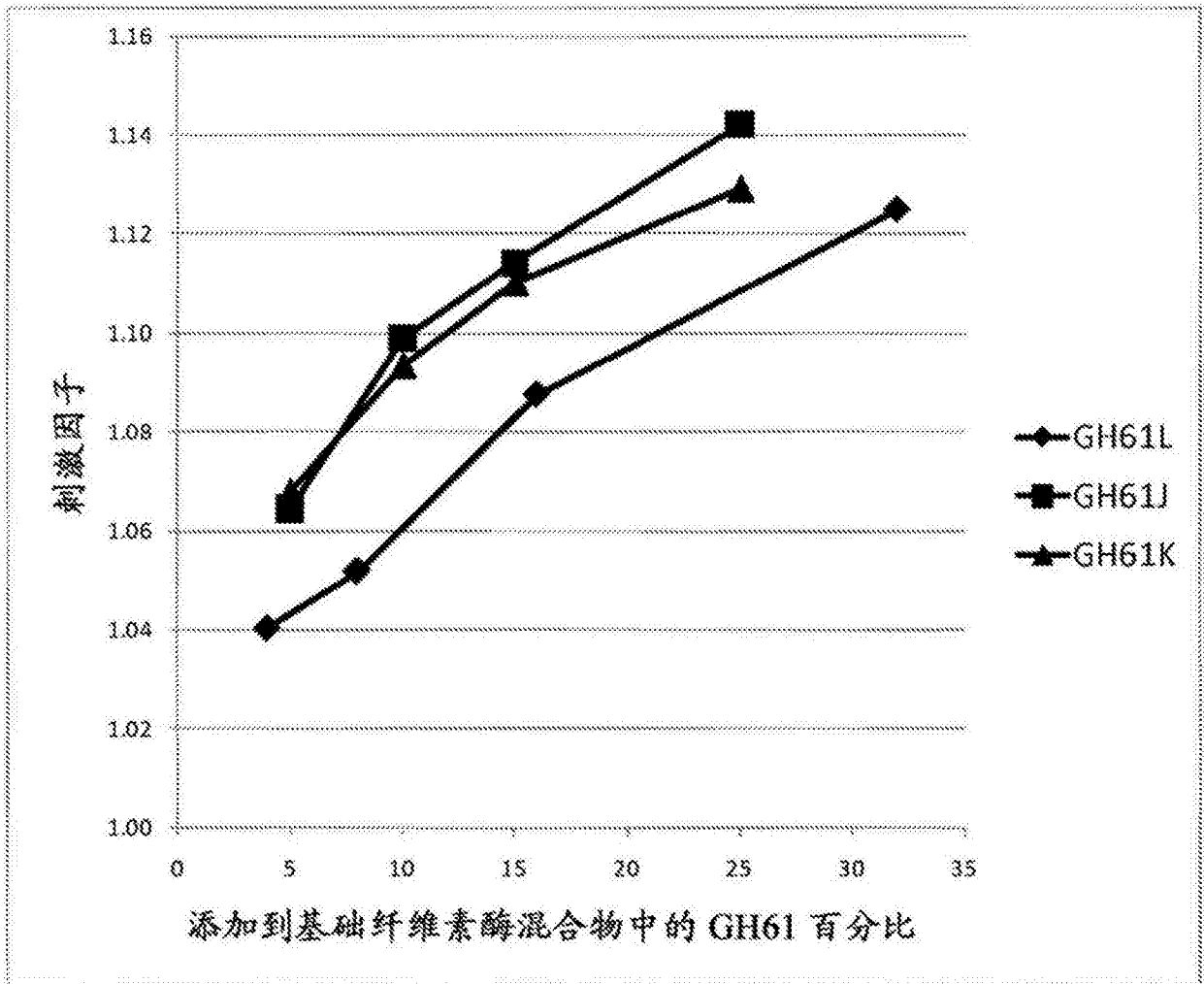


图4

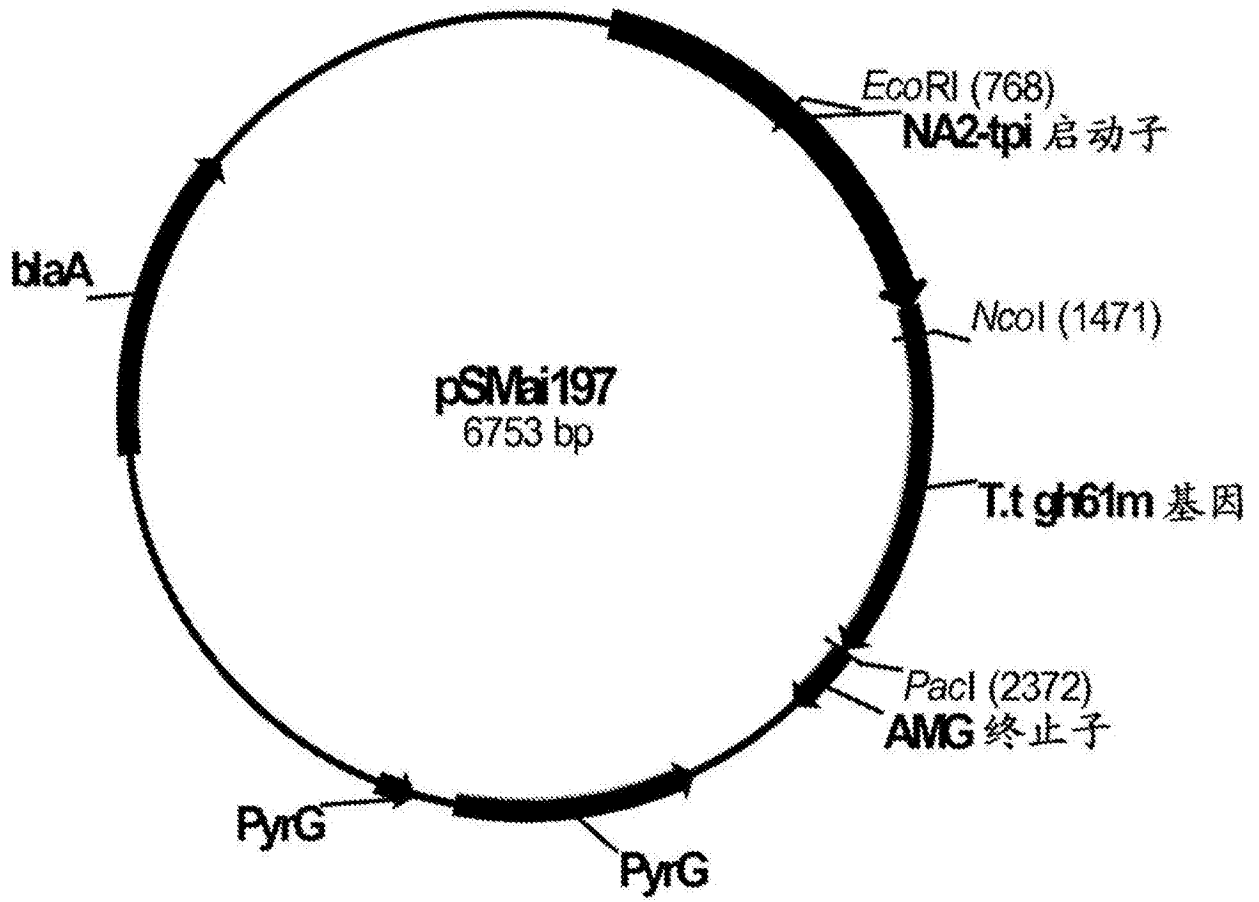


图5

M E L S S Q L A A L T L A A A S V S G H
 1 ATGAASCTGTCTATCCCAGCTGCGCCGCTCACGCTGGCCGCGGSCCTCCSTGTCCAGGCCAC
 Y I F E Q I A H G G T F F P P Y E Y I R
 61 TACATCTTGGAGCAGATTGCCCATGGCGGCACCAAGTTCCACCTTACGASTACATCCGA
 R N T H Y N S F V T S L S S N D L E C H
 121 AGAAACACGAACTATAACAGCCCTGTCACCAGTCTCTCGTGGAAACGACCTGGGATGCAAC
 Y G G E T A G H T T V L D V K A G D S F
 181 GTAGGCGGEGAGACGGCTGSCAACACGACCTCTCTCGACGTGAAGGCGGGGCGACTCCTTC
 T F Y S D V A V Y H Q G P I S L
 241 ACCTTCTACTCGGACGTGGCCGCTGTACCACCAGGGGCCCATCTCACTGTGCTCCCGCCGG
 301 GCCAACTTTGATCAGTCCCAAGCGGACTGTCCGCTCGCCTGGATAACACAAATTAAGTGA
 Y M S K A P G S V V D Y D G S G
 361 CAGCCCGCACAGCTACATGTCCAAAGGCTCCCGGCTCCGCTCGTGGACTACGACGGCTCCGG
 D W F K I H D W G F T F S N G Q A S W P
 421 CGACTGGTTCAGATCCACGACTGGGGCCCGACCTTCAGCAAAGGCCAGGCCCTCGTGGCC
 L R
 481 GCTGCGGGGTGCTCCCTTCCCTTCCCTCCCGCTTCCCTCCCGCTTCCCTCCCGCCCTTTC
 D N Y Q Y N
 541 CCCCCTTTTCTGTCTGGTGGCAAGCCCTGCTGACGTCCTCCCGTAGCAACTACCGTACAA
 I P T C I P N G E Y L L P I Q S L A I H
 601 CATCCCGAUGTGCATCCCGAAGCGGCGAGTSCCTGCTGCGCATCCAGTCCGCTGGCGATCCA
 N P G A T F Q P Y I S C A Q V R V S G G
 661 CAACCCGSSCGCCACGCCAGTTCATATCAGCTGCGCGCAGGTCCGGGTCTCCGGSCGG
 G S A S F S P T A K I F G A F K A T D P
 721 CGGCAGCGCCTCCCGCTCCCGACGGCCAGATCCCGGCGCCTTCAAGGGGACCGATCC
 G Y T A N
 781 CGGGTATACCGGAATGTGAGTGCCTATGTTCCTTGGCGCTCCTTGTTCCTTGTCTCTTG
 841 CTCGGCGTGTGTAACGCTACGGGCTGTGGAGGAGGGATGGATGGATGAATAGGATGCT
 I V N H F H S Y T V P G P
 901 GACTGATGGTGGGACACCAGATTTACRATAACTTCCACTCGTATACGGTGGCCGGGTCCGG
 A V F Q C *
 961 CGGTCTTTCAGTGCTAG

图6

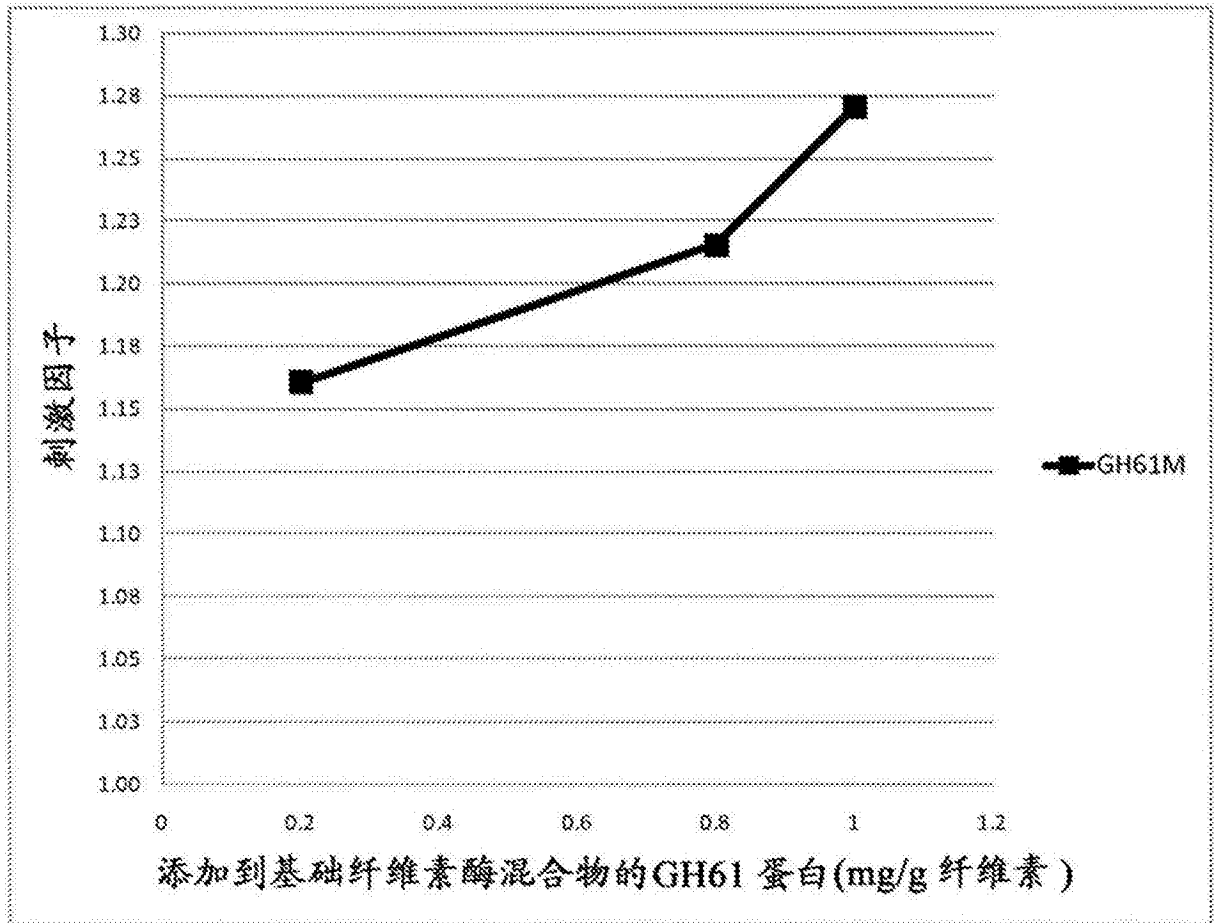


图7

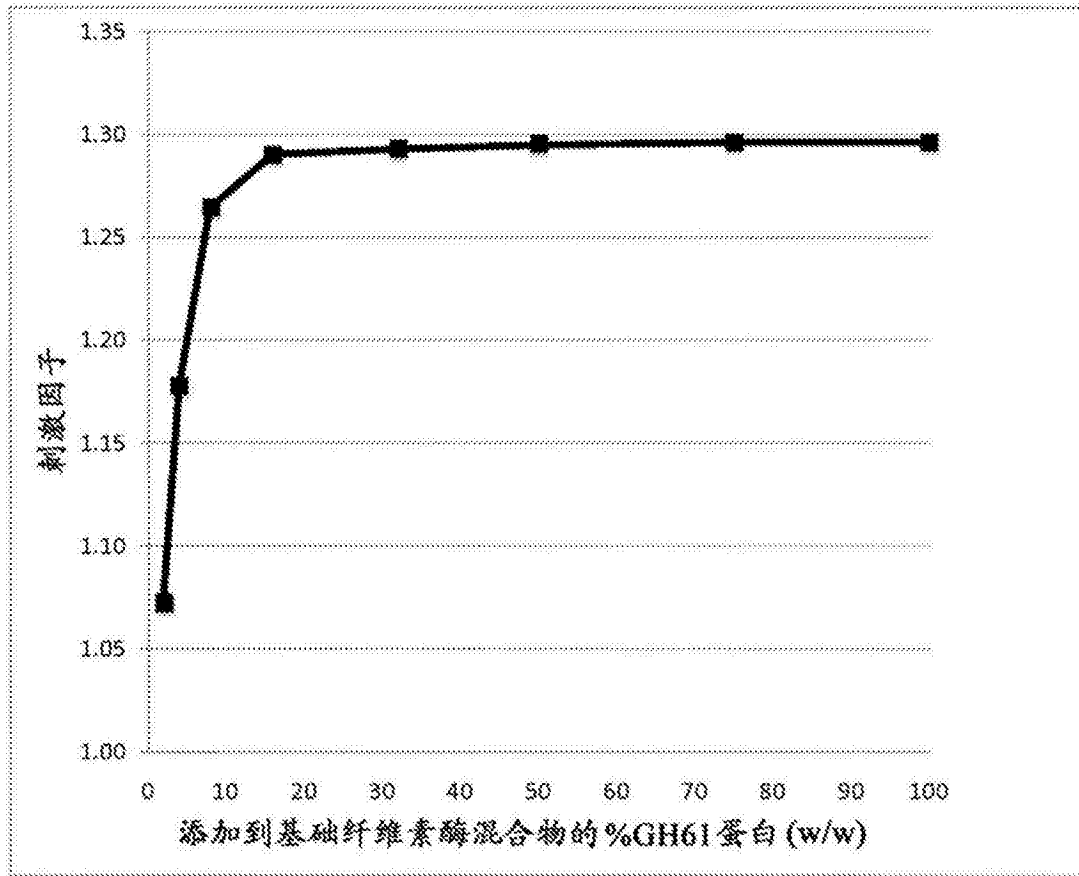


图9

M P P A L P Q L L T T V L T A L T L G E T A
 1 ATGCGCCCGCCCACTCCCTCAACTCCTAACCCACGGTCTGTGACCGCCCTCACCCTCGGTCCACCGCC
 L A H S H L A Y I I V N G K L Y Q G F D F R
 67 CTCGCCCACTCACACCTCGCGTACATTATCGTTAACGGCAAGCTCTACCAGGGCTTCGACCCCGCC
 F R Q A N Y P S R V G W S T S A V D D G F V
 133 CGCCACCCAGGCCAACTACCCTTCGCGGGTCCGGTGGTCCACCGGGCCCGTCCGACGACGGCTTCGTC
 T P A H Y S T F D I I C H I A G T S P A G H
 199 ACGCCGCCCACTACTCCACCCGACATCATTTGCCACATCGCCGGCACCAGCCCGCCGCCGCCAC
 A P V R P G D P I H V Q W N G W P V G H I G
 265 GCGCCCGTGCGCCCGGGCGACCCGATCCACGTCCAGTGGAAACGGCTGGCCCGTCCGCCACATCGGT
 F V L S Y L A R C E S D T G C T G Q H K T A
 331 CCCGTGTGTCTACCTCGCCCGCTGCGAGTCCGACACGGCTGCACGGGCCAGAACAGACCGCG
 L R W T K I D D S S P T H Q H V A G A G T Q
 397 CTGCGGTGGACCAAGATCGACACTCCAGCCACCATGCAGAACGTCCGCCGCCCGCCGCCACCCAG
 G E G T P G K R W A T D V L I A A N H S W Q
 463 GGCGAGGGCCACCCCGGCAAGCGCTGGGCCACCGAGCTGCTGATCGCCGCCAACACAGCTGGCAG
 V A V P A G L F T G A Y V L R H E I I A L H
 529 GTCCCGGTGCCCGCGGGCTGCGACCCGGCGCTACGTGCTGCSCAACGAGATCATCGCGCTGCAC
 Y A A R K N G A Q H Y P L C M N L W V D A S
 595 TACCGCCCGAGGAAGAACGGGGCCAGAACTATCCGCTCTGCATGACCTGTGGGTGGACGCCAST
 G D H S S V A A T T A A V T A G G L Q M D A
 661 GGTGATAATAGTATGTGGCTGCACCGACCGCGCGGTGACCGCGGGGGTCTGCAGATGGATGCC
 Y D A E G F Y K E N D P G V L V N V T A A L
 727 TATGACCGCGCGGGTCTTACAAGGAGAACGATCCGGCGTGTGGTCAATGTACAGCCCGCGCTG
 S S Y V V F G P T V A A G A T P V F Y A Q Q
 793 TCGTGSTATGTCTGCCCGGGCCGACGGTGGCGGCGGGGCGCACGCGCGGTGCGGTACGCGCAGCAG
 S P S V S T A A G T P V V V T R T S E T A F
 859 AGCCCGAGCGTGTCTGACCGCGGGGCGCACCCCGTCTGTTACAGGACTAGCCAGACCGCCCG
 Y T G A M T F T V A A R K K G R G Y D R R G
 925 TACACGGGCGCCATGACCGCCGACGGTTCGGCGAGGATGAAGGGGAGGGGTATGATCGCCGGGT
 *
 991 TAG

图10

M E T F F A L L A A A G L V A G H G Y V D N
 1 ATGAAGACATTACCCGCCCTCTGGCCGACGCCGCCCTCGTCGCCGGCCATGGATATGTCGACAAC
 A T I G G Q F Y Q
 67 GCCACCCATTGGCCGGCCASTTTTATCAGGTACTCTACCGCTTCACCCAAAGGTCCGCTGGCCACAAC
 133 CTATAGSTGTCATAAAATTAACAAGCCACCGTCCCGCAGTCTATCAGGTGTGCTCGCTACCGACCA
 199 TGTGGTCCCGTCTCAGCAAGCCACTCACACGCCCATGATCCCTAGCCTTAAGTGGACCCGTATTT
 N P
 265 AGCAACCTTGGCACGTAGTATTTATTGTCCCAAATATTTGAGCTGAACTGACACCTCCCTAGGATCCC
 A V L T F F Q P D R V S R S I P G N G P V T
 331 GCGGTGCTAACATTCTTTCAGCCCGACAGGGTCTCTCGATCCATCCCGGGCAACCGGCCCGGTCCAG
 D V T L I D L Q C N A N S T P A K L H A T A
 397 GACGTACTCTCATCGACCTGCAAGTSCAACGCCAATCCACCCCGGCCAAGCTCCACGCGCACTGCC
 A A G S D V I L R W T L W P E S H V G P V I
 463 GCTGCGGCTCGGACGTGATTCTCCGCTGGACGCTCTGGCCTGAGTCCGACGTTGGCCCGGTCATC
 T Y M A R C P D T G C Q D W M P G T S
 529 ACCTACATGGGCCCGCTGCCCGACACGGGCTGCCAGGACTGGATGCGGGCACTTCGTAGGAGCCC
 595 ATCTTGCACCATATCCAFFTTCAACCGGCCACAGCACTGACCCATATGTCGTCTACCCCTGCACT
 A V W F K I K E G G R D G T S H T R A D V R
 661 GCGGTCTGGTTCAAGATCAAGGAGGGCGGCGCGGCACTTCCAACACCTGGGGCCGCTACGT
 V F R P R E P E P P L Q Q S K H L N S P S L
 727 GTACCCCGTCCAGAGAGCCAAAGCCCGCCCTTCAACAAGCAAAACATCTCAATAGCCCGAGCCTA
 R T N P S P S P S K T Q T P L M T A P T S Y
 793 CGCACTAACCCCTCTCCTTCCCCCTCGAAAACACAGACCCCGCTGATGACGGCGCCACCTCGTAC
 T Y T I F S C L E K G Y Y L V R H E I I A L
 859 ACGTACACGATCCCCCTCTGCTGAAGRAGGGCTACTACCTGGTCCGCCACGAGATCATCGGCTG
 H A A Y T Y P G A Q F Y P G C H Q L N V T G
 925 CACGCGCCTACACCTACCCCGGCGCGCAGTCTACCCGGCTGCCACCAGCTCAACGTCACGGGC
 G G S T V F S S G L V A F P G A Y K G S D P
 991 GGCGGGTCCACCGTACCGTGGAGCGGCTGGTGGCCTTCCCGGGGGGTACAAGGGCCAGTGAACCC
 G I T Y D A Y K
 1057 GGGATTACGTACGATGCGIATAAAGGTGGGCTGGCTGTTGGCCAGCTCTTGGTGTGGGGGAAT
 A Q T Y
 1123 GTGGTGATGAGGTTTATTATTTGGGATCCCGTGGCTAACCTAACCCCTGGGTGTAGCGCAACGATC
 Q I F G F A V F T C *
 1189 CAGATTCTGGGCCGGCGCTTTACTTGGCTGA

图11

M A L L L L A G L A I L A G F A H A H G G L
 1 A T G G C C T T S C T G C T C T T G G C A G G C T T G G C C A T T C T G G C C S G G C G G C T C A T G C C C A C G G C G G C C T C
 A N Y T V G H T W Y R G
 67 G C C A A C T A C A C A G T G G G C A A C A C C T G G T A T A S G G G T G C G T A A S G G G G C A C C G A C A A C G C C T G G T
 Y D P F T F A A
 133 T A G T A A C T C C A C C A T T T C G A G C G G S C T A A C A C G G G C C A G C T A C G A C C C C T T C A G C C G G G C G G C
 D Q I G O P W N I Q R A W D S I D P I F S V
 199 G A C C A G A T C G G C C A G C C G T G G A T G A T C C A A C G E G C G T G G G A C T C G A T C G A C C C G A T C T T C A G G C T C
 N D K A L A C N T P A T A F T S Y I P I R A
 265 A A C G A C A A G G C C G C T C G C C T G C A A C A C C C C G G C C A C G C G C C G A C C T C T T A C A T T C C C A T C C G C G G
 G E H I T A V Y W Y W L H F V G F N T A W L
 331 G G C G A G A A C A T C A C G G C C G T G T A C T G G T A C T G G C T G C A C C C G G T G G G C C C C A T G A C G G C G T G G C T G
 A P C D G D C R D A D V N E A P W F K I W E
 397 G C G C G G T G C G A C G G C B A C T G C C C G A C G C C G A G D T C A A C G A G S C G C G T G G T T C A A G A T C T G G S A G
 A G L L S G F N L A E G M W Y Q K A F Q N W
 463 G O D G C C T G C T C A G C G G G C C G A A C C T G G C C G A G G G C A T G T G G T A C C A G A A G G C G T T C C A G A A C T G G
 D G S F D L W E V T I P A G L K S G G L Y N I
 529 G A C G C C A G C C C G A C C T G T G G C C C G T C A C G A T C C C G G C C G G C T G A A G S G C G G C C T G T A C A T G A T C
 R H E I L S I H V E D K P Q P Y P E C A H L
 595 C G G C A C G A G A T C T T G T C G A T C C A C G T C G A G G A T A A A C C G C A G T T T F A T C C C G A G T G C G C G A T C T G
 N V T G G G D L L P P D E F L V E F P G A Y
 661 A A T G T G A C C G G G G T G G G G A C C T G C T G C C G C C T G A T G A G T T T T T G G T G A A G T T C C C G G C G C T T A C
 K E D H
 727 A A G A A G A T A G T G A G T G A A A C G C G A A G C T T C G G T A G C C A T T G G G T T G C G C T G A T G G A G G T T A G A C C
 P S I K I N I Y S D Q Y A N T T
 793 C G T C G A T C A A G A T C A A T A T C T A C T C G G A C C A G T A C G C C A A T A C A A C G G T G A G T C T A A C A G G T C G A G
 N Y T I F G G P I W
 859 C A A A A C C A A A C A G A T G C C G A T G A C T G A T G A T C T C A G A A T T A C A C A A T T C C C G G A G S G C C G A T A T G G
 D S *
 925 G A T G G E T G A

图12

M H P G L V R F S M G L A T A F A S L S T A
 1 ATGATGCCGTCCCTTGTTCGCTTCTCAATGGGTCTGGGCGACCGCCTTCGGCTCGCTGTCCACAGCA
 H T V F T T L F I N G V D Q G D G T C I R M
 67 CATACCGTCTTACCACCGCTTTTCATCAACGGCGTCCGACCAAGGGGACGGGACCTGCATCCGCATG
 A K K G S V C T H P I A G G L D S P D M A C
 133 GCCAAGAAAGGGCAGCGTTTGCACCCATCCCATTTGCTGGTGGCCTCGACAGCCAGACATGGCTTGT
 G
 199 GGTATGCCCTCTGCGTFTCCCTTGGGAGAGCTTTCTCTCAGCTAACCCCAATGCCCGCTTGGCCAGG
 R D G Q Q A V A F T C P A P A G S K L S F E
 265 CCGAGGCGGACAACAAGCCGTGGCATTCACTGCCCAGCCCGGGCGGCTCCAAAGTTGAGCTTCGA
 F P R M W A D A S Q F G S I D P S H L G S T A
 331 GTTCCGATGTGGGCGGACGCTCTCAGCCCGGCTCTATCGACCCATCCACCTCGGCTCGAGCGG
 I Y L K Q V S H I S S D S A A G P G W F K I
 397 AATCTAOCITCAAACAAGTCTCCAACATCAGCTCCGACTCGGCTGCCGGCCCTGGCTGGTTCAGAT
 Y A E G Y D T A A K K W A T E K L I D N G G
 463 CTACGCCGAGGGCTACGACACAGCCGCCAAGAAGTGGGCCACAGAGAAGCTCATCGACAAGCGCGG
 L L S I E L P F T L P A G Y Y L A B S E I V
 529 CCTGCTGAGCATCGAGCTTCCGCCACTTGTGCCCGGGGTACTACTCCGCCCGCAGCGAGTGTGT
 T I Q N Y T N D H V D P Q F Y V G C A Q L F
 595 CACCATCCAGAACCTCACCAACGACCCTCGACCCCGCAGTTCTACGTTGGCTCGGCACACGCTCTT
 V Q G P F T T F T V P F D P L V S I F G H V
 661 CGTCCAGGGGCTCCGACCAACCCCAACCGTCCCGCCAGACGAGCTCGTCTCCATCCCGGGCCAGGT
 H A S D P G L T F N I W B D D F S E T A Y T
 727 CCATGCCCTCCGACCCGGGCTGACCTTCAACATCTGGCGCGACGACCCCTCCAAGACGGCTACAC
 V V G P A P F S F T A A P T P T S T H T R G
 793 CGTGTCCGGCCCGGCECCCTTCTCCGCCACCCGCCGCCACCCCACTCCACCAACCAACCAAGCG
 Q Q Q Q Q Q Q A I E Q T D G V I P A D C Q
 859 GCRGCAACBACARCAACAGCAACAGGCGATAAAGCAGACCGGCGGTGATCCCGCCGACTGCCA
 L K N A N W C G A E V P A Y A D E A G C W A
 925 GCTCAGGAACGCCAACTGGTGGCGGCGGAGGTGCCCGGTACGCCGACGAGGCCCGGCTGCTGGCC
 S S A D C F A Q L D A C Y T S A F P T G S R
 991 GTCGTCCGCGGACTGCTTCGCCAGCTGGACCGCTGCTACAGCTCGGGCCCGCCACCGCCAGCCG
 G C R L W E D W C T G I Q Q G C R A G P W R
 1057 CCGCTGCCCGCTGTGGGAGGACTGGTGCACCCGCCATTCACAGGCTGCCCGCGGGCGGCTGGCC
 G F P P F H G E G A A A E
 1123 GGGGCCGCCGCCCTTTTCATGGGGAGGGGCGACGAGGAGGTGTGAACGGTTCGGGGACGGGTGGC
 T A S
 1189 GGT
 A G R G G A R I A A V A G C G G G T G D N V
 1255 GCTGGCCGGGGGGGGGCAAGAAAGCTGCCGTGGCCGGCTGCGGAGGGGGGACAGGAGACATGGTT
 E E V F L F Y W D A C S G W R E S R G G G S
 1321 GAAGAGGTTTTCTCTTTATTGGGACGCTTSCAGCGGCTGGCGACGGAGCCGTGGTGGTGGTTCG
 I L A R L I L H V L L F L L E F R R A P R Y
 1387 ATTCCTGCGAGGCTTATCCCTCATGTCTCTTCCACTTTTGAGACCGAGGCGAGCCCTCGAGTC
 H L L L F H L Y L N F C Y P G T S G F Y N R
 1453 CATTACTTCTCTTCCACCTGTACCTCAACTTCTGTATCCAGGAACCAAGTGGTTTCTATATCGC
 L S I K L G I W F S K M S F D V A H Y V K *
 1519 CTGAGCATTAAGCTAGGCATATGCCAAGCAAAATGTCCGCTGATGTAGCCATTCCGTGAATAA

图13

M Q L L V G L L L A A V A A R A H
 1 ATGCAGCTCCTCGTGGGCTTGGCTTGCAGCGTGGCTGCTCGAGCACATTGTATTTCTACCCCT
 Y F
 67 TTCGGCCTGCCCTCCAGCCTCAAGGCAAGAAGACGCCACCCAGCAGCTAACGGACECTATCAGACAC
 F P R L V V N G Q P E D E D W S V T R M T E
 133 ATTTCCAGACTCCTGGTAAATGGSCAGCCCGAGGACAAAGGACTGGTCGGTTACGGGCATGACCAA
 N A Q S E Q G V Q D P T S P D I R C Y T S Q
 199 GAACGGCAGAGCAAGCAGGGAGTCCAGGAACCGACAGTCCCGACATTTCGCTGCTACACGTCSCA
 T A P N V A T V P A G A T V H Y I S T Q Q I
 265 GACGGCGCCTAACSTGGCTACGGTCCCTGCCGGAGCCACCGTCCATTACATATCGACTCAGCAGAT
 N H P G P T Q Y Y L A K V P A S S S A K T W
 331 CAACCCCGGGCCCGACCCAGTACTACCTCGCCAGGTACCGGGCGGGTCTGTCGGCCAGACSTG
 D G S G A V W F E I S T T H P Y L D N H K Q
 397 GGACGGGTCAGGGGCCCTCTGCTTCAAGATCTCGACCACCCTGCTTACTTGGACAAACAAGSCA
 L V W P N Q
 463 GCTTGTCTGGCCGAATCAGAGTAGGAACAATTCGCCCTCCAATCTTCGATTTGGCCCTTGAGCTACG
 H T Y T T
 529 GCCGATTGCATGGGAGAGACCGTGTACTGACGGGGCAACCCCAACCTTCATCAGACACGTCACACBAC
 V H T T I P A D T P S G E Y L L R V E Q I A
 595 GGTCACACGACCATCCCCGCGAATCGCCAGTGGGGAATACCTCCTCCGGGTCGAGCAGATCGC
 L H L A S Q P N G A Q P Y L A C S Q I Q I T
 661 GCTGCACCTGGCTCGCAACCCAAACGGGGCTCAGTTCCTACCTGGCTGCTCGCGATCCAGATTC
 G G S N G T P G P L V A L P G A Y K S N D P
 727 GGGCGCGGCAACCGCACGCCCGCCCGCTAGTCCGCTTGGCCGGGGCGTACAAGAGCAACGACCC
 G I L V N I Y S M Q P G D Y E P P G P P V W
 793 GGCATTTTGGTCAACATCTACTCTATGCAGCCCGCGGATTACAAGCCCGCCCGGGCCCGCGGTGTG
 S G *
 859 GAGTGGCTGA

图14

M K L Y L A A F L G A V A T P G A F A H
 1 ATGAAGCTGTACCTGGCGGCTTTCTAGGCGCGCTCGCCACCCCGGGAGCGTTCTGCTCATCGTASS
 Q I R
 67 TTCECCGCTATCTECCCTAGGGGTAGCACCACGACTAATTTCTCGTGGTCCCCCTGTAGAAATCCA
 G I L L V N G T E T P E W E Y V R
 133 CGGGATTCTACTTGTCAAAGCCACCCSAAAAGCCCGGAATGGAANTACGTCGGGTAATATCTACCTTG
 D V A W E G A Y E
 199 CTCTCCTTCTTCCACAACCAGCCTAACACATCATCAGTGGCGTGGCCTGGGAGGGGCGCTACGAAC
 P E K Y P N T E F F K T P P Q T D I N N P N
 265 CGGAAAANTACCCCAACACCGGAGTTCTTTAAGACGCCCCCGCAGACGGGACNTCAACAACCCGAAACA
 I T C G R H A F D S A S E T E T A D I L A G
 331 TCACCTGGCGCAGGAACGCTTTCGACTCGGACCAAGACTGAGACGGCCGACATACTGGCCGGCT
 S E V G F R Y S W D G H G K Y G V F W H P G
 397 CAGAGGTGGGCTTCCGCGTCTCGTGGGACGGCAACGGCAAGTACGGCGTSTTCTGGCATCCCGGGC
 P G Q I Y L S R A P H D D L E D Y R G D G D
 463 CGGGGCAGATCTACCTCTCTCGTGGCTCCGAACGACGACCTGGAGGACTACCGGGCGACGGAGACT
 W F K I A T G A A V S N T E W L L W H K H D
 529 GGTTCASRTGCRACCCGGCCGCGCTTCCAATACCGAGTGGCTGCTGTGGAAACAGCATGACC
 595 TGAGCCCCAACATTCCTCGCCCAATCGATCCCCAACCTGGTCCACCATGGCGGGCTCCGGGATGCAA
 F H F T I F E T T P F G
 661 AGAGACTAACTCCAGAGGAACCTACCTAGTTCAACTTCAACATCCCCAAGACGACGCCCGCGGGCA
 K Y L M R I E Q F M P S T V E Y S Q W Y V N
 727 AGTACCTGATGCGCATCGAGCBGTTTCATGCCCTCCACGGTCCGAATACAGCCAGTGGTACGTCAACT
 C A H V N I I G F G G G T F T G F A R F P G
 793 GCGCCCACTCAACATCATTCGGCCCGGGCGGAGGCACGCCGACGGGCTTFCAGGTTTCCCGGCA
 T Y T V D D P
 859 CCTACACTGTTGACGATCCCGGTAAGCCGGACCTACCGGACACAGAGGCTCCGGGATAGCTTCTA
 G I K V P L N Q I V
 925 ACCTTGTTFGCTCTCTCTCTTTTTCTCTCCCGACTAGGCATCAAGGTGCCGTTGAACCAGATCGTC
 H S G E L P Q D Q L R L L E Y K P P S P A L
 991 AACAGCGGAGGTTGCCGCRGGACCAACTGAGGCTGCTGAGGTACAGCCCCCGGGCCACGCGCTG
 W T G *
 1057 TGGACTGGTTGA

图15