



(12) 发明专利

(10) 授权公告号 CN 109790510 B

(45) 授权公告日 2024.04.12

(21) 申请号 201780059992.7

(22) 申请日 2017.10.03

(65) 同一申请的已公布的文献号

申请公布号 CN 109790510 A

(43) 申请公布日 2019.05.21

(30) 优先权数据

62/403,787 2016.10.04 US

(85) PCT国际申请进入国家阶段日

2019.03.28

(86) PCT国际申请的申请数据

PCT/US2017/054983 2017.10.03

(87) PCT国际申请的公布数据

W02018/067599 EN 2018.04.12

(73) 专利权人 丹尼斯科美国公司

地址 美国加利福尼亚州

专利权人 芬兰VTT技术研究中心有限公司

(72) 发明人 M·华德 Y·罗 F·O·本德苏

M·瓦尔科宁 M·萨罗海摩

N·阿罗 T·帕库拉

(74) 专利代理机构 北京市中咨律师事务所

11247

专利代理人 胡志君 黄革生

(54) 发明名称

在不存在诱导底物下丝状真菌细胞中的蛋白产生

(57) 摘要

本公开的某些实施例针对变体丝状真菌细胞、其组合物及其用于增加一种或多种目的蛋白的产生的方法。更特别地，在某些实施例中，本公开针对衍生自亲本丝状真菌细胞的变体丝状真菌(宿主)细胞，其中所述变体宿主细胞包含遗传修饰，所述遗传修饰使得能够在不存在诱导底物的情况下表达/产生目的蛋白(POI)。在某些实施例中，本公开的变体真菌宿主细胞包含遗传修饰，所述遗传修饰增加编码本文称为Ace3-L的

(51) Int.CI.

C12N 1/15 (2006.01)

C12N 9/42 (2006.01)

C12N 15/80 (2006.01)

C07K 14/37 (2006.01)

C12P 21/02 (2006.01)

C12N 1/14 (2006.01)

(56) 对比文件

CN 105980552 A, 2016.09.28

CA 2801799 A1, 2011.12.08

CN 102099460 A, 2011.06.15

CN 102939382 A, 2013.02.20

US 2007231819 A1, 2007.10.04

CN 113106114 A, 2021.07.13

CN 101831452 A, 2010.09.15

CN 102311951 A, 2012.01.11

CN 106978360 A, 2017.07.25

WO 2011151512 A2, 2011.12.08

AU 2005229437 A1, 2005.10.13

郑芳林.转录激活因子Xyr1在里氏木霉纤维素酶基因诱导表达中的作用机制研究.中国博士学位论文全文库.2017,全文. (续)

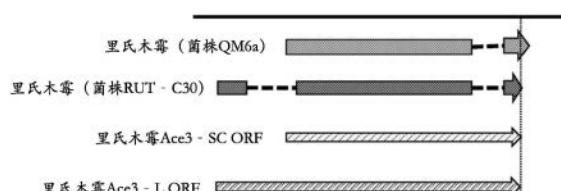
审查员 留盛鹏

权利要求书2页 说明书51页

序列表62页 附图28页

Ace3蛋白的纤维素酶表达激活子3(ace3)基因变体的表达。

来自里氏木霉株QM6a和RUT C30的Ace3蛋白编码区的示意图



[转续页]

[接上页]

(56) 对比文件

赵心清.里氏木霉代谢工程改造及纤维素合成调控.中国会议.2016,全文.

Robyn Peterson.Trichoderma reesei RUT-C30 - thirty years of strain improvement.《microbiology》.2011,第158卷(第1期),

M Carmen Limón 等.The effects of disruption of phosphoglucose isomerase gene on carbon utilisation and cellulase production in Trichoderma reesei Rut-C30.《Microb Cell Fact》.2011,第10卷(第40期),

吴鸿清.里氏木霉双基因同步定点整合体系的构建及提高 β -甘露聚糖酶表达的策略研究.《中国优秀硕士学位论文全文数据库(电子期刊)基础科学辑》.2015,

Xue Y 等.Promoting cellulase and hemicellulase production from Trichoderma orientalis EU7-22 by overexpression of

transcription factors Xyr1 and Ace3.

《Bioresour Technol》.2019,

Chen Y 等.Engineering of Trichoderma reesei for enhanced degradation of lignocellulosic biomass by truncation of the cellulase activator ACE3.《Biotechnol Biofuels》.2020,第13卷(第62期),

Luo Y 等.Modification of transcriptional factor ACE3 enhances protein production in Trichoderma reesei in the absence of cellulase gene inducer.《Biotechnol Biofuels》.2020,第13卷(第137期),

Kuo,A 等.hypothetical protein M419DRAFT_98455 [Trichoderma reesei RUT C-30]ETS02557.1.《Genbank Database》.2015,

Hideaki Koike 等.Comparative Genomics Analysis of Trichoderma reesei Strains.《INDUSTRIAL BIOTECHNOLOGY》.2013,

1. 一种在不存在诱导底物的情况下用于在木霉属物种(Trichodermasp.)真菌细胞中产生内源木质纤维素降解酶的方法,所述方法包括:

(i) 向所述真菌细胞中引入多核苷酸构建体,所述多核苷酸构建体在5'至3'方向包含:

(a) 包含启动子的第一核酸序列和(b)有效地连接到所述第一核酸序列的第二核酸序列,其中所述第二核酸序列编码SEQ ID NO:6或SEQ ID NO:12所示的Ace3蛋白,以及

(ii) 在适合于真菌细胞生长和蛋白产生的条件下发酵步骤(i)的细胞,其中此类适合的生长条件不包括诱导底物。

2. 如权利要求1所述的方法,其中所述多核苷酸构建体进一步包含在所述第二核酸3'端的并有效地连接到所述第二核酸的第三核酸序列,其中所述第三核酸序列包含天然ace3终止子序列。

3. 如权利要求1所述的方法,其中将所述多核苷酸构建体整合到所述真菌细胞基因组中。

4. 如权利要求1所述的方法,其中将所述多核苷酸构建体整合到所述真菌细胞基因组的端粒位点中。

5. 如权利要求1所述的方法,其中将所述多核苷酸构建体整合到所述真菌细胞基因组的葡糖淀粉酶gla1基因座中。

6. 如权利要求1所述的方法,其中所述多核苷酸构建体包含核苷酸序列,所述核苷酸序列为SEQ ID NO:4、SEQ ID NO:11或SEQ ID NO:13,或编码SEQ ID NO:6或SEQ ID NO:12的其他简并核苷酸序列。

7. 如权利要求1所述的方法,其中所述启动子选自由以下组成的组:SEQ ID NO:15所示的rev3启动子、SEQ ID NO:16所示的bx1启动子、SEQ ID NO:17所示的tk11启动子、SEQ ID NO:18所示的PID104295启动子、SEQ ID NO:19所示的d1d1启动子、SEQ ID NO:20所示的xyn4启动子、SEQ ID NO:21所示的PID72526启动子、SEQ ID NO:22所示的axe1启动子、SEQ ID NO:23所示的hxk1启动子、SEQ ID NO:24所示的dic1启动子、SEQ ID NO:25所示的opt启动子、SEQ ID NO:26所示的gut1启动子,和SEQ ID NO:27所示的pki1启动子。

8. 一种在不存在诱导底物的情况下用于在木霉属物种真菌细胞中产生异源目的蛋白的方法,所述方法包括:

(i) 向所述真菌细胞中引入多核苷酸构建体,所述多核苷酸构建体在5'至3'方向包含:

(a) 包含组成型启动子的第一核酸序列和(b)有效地连接到所述第一核酸序列的第二核酸序列,其中所述第二核酸序列编码SEQ ID NO:6或SEQ ID NO:12所示的Ace3蛋白,以及

(ii) 在适合于真菌细胞生长和蛋白产生的条件下发酵步骤(i)的细胞,其中此类适合的生长条件不包括诱导底物。

9. 如权利要求8所述的方法,其中所述真菌细胞包含编码异源目的蛋白的引入的多核苷酸构建体,其中在步骤(i)之前、在步骤(i)期间或在步骤(i)之后将所述构建体引入所述真菌细胞中。

10. 如权利要求8所述的方法,其中所述多核苷酸构建体包含在所述第二核酸3'端的并有效地连接到所述第二核酸的第三核酸序列,其中所述第三核酸序列包含天然ace3终止子序列。

11. 如权利要求8所述的方法,其中将所述多核苷酸构建体整合到所述真菌细胞基因组

中。

12. 如权利要求8所述的方法,其中将所述多核苷酸构建体整合到所述真菌细胞基因组的端粒位点中。

13. 如权利要求8所述的方法,其中将所述多核苷酸构建体整合到所述真菌细胞基因组的葡糖淀粉酶g1a1基因座中。

14. 如权利要求8所述的方法,其中所述多核苷酸构建体包含核苷酸序列,所述核苷酸序列包含与SEQ ID NO:4、SEQ ID NO:11或SEQ ID NO:13,或编码SEQ ID NO:6或SEQ ID NO:12的其他简并核苷酸序列。

15. 如权利要求8所述的方法,其中所述异源目的蛋白选自由以下组成的组: α -淀粉酶、碱性 α -淀粉酶、 β -淀粉酶、纤维素酶、葡聚糖酶、 α -葡糖苷酶、 α -半乳糖苷酶、葡糖淀粉酶、半纤维素酶、戊聚糖酶、木聚糖酶、转化酶、乳糖酶、柚皮苷酶、果胶酶、支链淀粉酶、酸性蛋白酶、碱性蛋白酶、菠萝蛋白酶、中性蛋白酶、木瓜蛋白酶、胃蛋白酶、肽酶、皱胃酶、高血压蛋白原酶、凝乳酶、枯草杆菌蛋白酶、嗜热菌蛋白酶、天冬氨酸蛋白酶、胰蛋白酶、脂肪酶、酯酶、磷脂酶、磷酸酶、植酸酶、酰胺酶、亚氨基酰化酶、谷氨酰胺酶、溶菌酶、青霉素酰化酶;异构酶、氧化还原酶、过氧化氢酶、氯过氧化物酶、葡萄糖氧化酶、羟基类固醇脱氢酶、过氧化物酶、裂解酶、天冬氨酸 β -脱羧酶、富马酸酶、组氨酸酶、转移酶、连接酶、氨肽酶、羧肽酶、几丁质酶、角质酶、脱氧核糖核酸酶、 α -半乳糖苷酶、 β -半乳糖苷酶、 β -葡糖苷酶、漆酶、甘露糖苷酶、齿斑葡聚糖酶、多酚氧化酶、核糖核酸酶和转谷氨酰胺酶。

16. 如权利要求8所述的方法,其中所述启动子选自由以下组成的组:SEQ ID NO:15所示的rev3启动子、SEQ ID NO:16所示的bx1启动子、SEQ ID NO:17所示的tk11启动子、SEQ ID NO:18所示的PID104295启动子、SEQ ID NO:19所示的d1d1启动子、SEQ ID NO:20所示的xyn4启动子、SEQ ID NO:21所示的PID72526启动子、SEQ ID NO:22所示的axe1启动子、SEQ ID NO:23所示的hxk1启动子、SEQ ID NO:24所示的dic1启动子、SEQ ID NO:25所示的opt启动子、SEQ ID NO:26所示的gut1启动子、和SEQ ID NO:27所示的pk11启动子。

17. 如权利要求8所述的方法,其中编码所述异源目的蛋白的所述多核苷酸构建体在纤维素诱导型基因启动子的控制下表达。

18. 一种对里氏木霉菌株进行遗传修饰以在不存在诱导底物的情况下增加内源木质纤维素降解酶的产生的方法,所述方法包括:

(i) 筛选和鉴定里氏木霉菌株,所述里氏木霉菌株包含编码SEQ ID NO:3的Ace3-S蛋白、SEQ ID NO:8的Ace3-SC蛋白、或SEQ ID NO:10的Ace3-LC蛋白的ace3基因的基因组拷贝,其中鉴定的所述里氏木霉菌株不包含编码SEQ ID NO:6的Ace3-L蛋白、SEQ ID NO:14的Ace3-LN蛋白、或SEQ ID NO:12的Ace3-EL蛋白的ace3基因的基因组拷贝,

(ii) 向步骤(i)中鉴定的里氏木霉菌株中引入多核苷酸构建体,所述多核苷酸构建体在5'到3'方向包含:(a) 包含启动子的第一核酸序列和

(b) 有效地连接到所述第一核酸序列的第二核酸序列,其中所述第二核酸序列编码SEQ ID NO:6或SEQ ID NO:12所示的Ace3蛋白,以及

(iii) 在适合于真菌细胞生长和蛋白产生的条件下发酵步骤(ii)的细胞,其中此类适合的生长条件不包括诱导底物。

在不存在诱导底物下丝状真菌细胞中的蛋白产生

[0001] 相关申请的交叉引用

[0002] 本申请要求2016年10月04日提交的美国临时申请号62/403,787的权益,该临时申请通过引用以其全文特此结合。

[0003] 序列表

[0004] 经由EFS同此提交的序列表文本文件包含于2017年10月02日创建的文件“NB41159WOPCT_SEQLISTING.txt”,其大小为157千字节。本序列表符合37C.F.R. §1.52(e),并以其全文通过引用结合在此。

技术领域

[0005] 本公开总体涉及分子生物学、生物化学、蛋白产生和丝状真菌的领域。本公开的某些实施例针对变体丝状真菌细胞、其组合物及其用于增加一种或多种目的蛋白的产生的方法。更特别地,在某些实施例中,本公开针对衍生自亲本丝状真菌细胞的变体丝状真菌(宿主)细胞,其中所述变体宿主细胞包含遗传修饰,该遗传修饰使得能够在不存在诱导底物的情况下表达/产生一种或多种目的蛋白(POI)。

背景技术

[0006] 纤维素(木质纤维素植物材料的组分)是自然界中发现的最丰富的多糖。同样,本领域已知丝状真菌是植物生物质的有效降解物,并且实际上是工业上相关的木质纤维素降解酶(下文统称为“纤维素酶”的酶)的主要来源。例如,已知丝状真菌产生细胞外纤维素酶(例如,纤维二糖水解酶、内切葡聚糖酶、 β -葡萄糖苷酶),其水解纤维素的 β -(1,4)-连接的糖苷键以产生葡萄糖(即,因此赋予这些丝状真菌利用纤维素生长的能力)。

[0007] 特别地,已知丝状真菌里氏木霉(*Trichoderma reesei*, *T. reesei*);真菌红褐肉座菌(*Hypocrea jecorina*)的无性型)是纤维素酶的有效产生者(参见,例如,PCT国际申请号WO 1998/15619、WO 2005/028636、WO 2006/074005、WO 1992/06221、WO 1992/06209、WO 1992/06183、WO 2002/12465等)。因此,已经利用如里氏木霉的丝状真菌产生酶的能力,所述酶在如纤维素衍生的乙醇、纺织品和衣服、洗涤剂、纤维、食品和饲料添加剂以及其他工业用途的商品的生产中是有价值的。

[0008] 已知这些工业上相关的酶在木霉属(*Trichoderma*)中的表达(和产生)取决于可用于生长的碳源。更特别地,丝状真菌产生纤维素酶是耗能的过程,因此,诱导和抑制机制都在丝状真菌中进化以确保这些酶的有效产生。例如,编码植物细胞壁材料降解所需酶(即纤维素酶/半纤维素酶)的各种基因在“诱导”底物存在下被“激活”,并且在易于代谢的碳源(例如,D-葡萄糖)存在下被“抑制”,经由称为“碳分解代谢物抑制”(下文称为“CCR”)的机制所述易于代谢的碳源优于植物生物质。因此,纤维素酶基因被葡萄糖紧紧抑制并被纤维素和某些二糖(例如槐糖、乳糖、龙胆二糖)诱导数千倍。例如,与含有葡萄糖的培养基相比,在含有诱导碳源(如纤维素或槐糖)的培养基上主要纤维二糖水解酶1(*cbh1*)的表达水平上被上调数千倍(Ilmen等人,1997)。此外,向“诱导的”里氏木霉培养物中添加“抑制”碳源显示

覆盖(override)(纤维素或槐糖)诱导,从而导致纤维素酶基因表达的下调(e1-Gogary等人,1989;Ilmen等人,1997)。因此,包含纤维素酶系统(酶)的基因的表达至少在转录水平上协调和调节,其中此系统的基因成员协同作用,并且如上所述是纤维素有效水解成可溶性寡糖所必需的。

[0009] 更具体地,全基因组分析显示在里氏木霉中存在至少十(10)个纤维素分解酶和十六(16)个木聚糖分解酶编码基因(Martinez等人,2008)。特别地,分泌的最丰富的酶是两种主要的纤维二糖水解酶(EC.3.2.1.91)、cbh1(纤维二糖水解酶1)和cbh2(纤维二糖水解酶2),以及两种主要的特异性内切- β -1,4-木聚糖酶(EC 3.3.1.8)、xyn1(内切-1,4- β -木聚糖酶xyn1)和xyn2(内切-1,4- β -木聚糖酶2),在本文中称为“主要的工业相关的半纤维素酶和纤维素酶”或“MIHC”。这些MIHC与另外的酶一起起作用以降解纤维素和木聚糖,这导致可溶性寡糖和单糖(如纤维二糖、D-葡萄糖、木二糖和D-木糖)的形成。此外,槐糖是这些酶中某些酶的转糖基活性的产物(Vaheri等人,1979)。更特别地,文献中已经报道了这些可溶性寡糖和单糖(即,纤维二糖、D-葡萄糖、木二糖、D-木糖和槐糖)影响里氏木霉中MICH的表达。例如,D-葡萄糖的存在引起CCR,该CCR导致少量MIHC的分泌。据信,槐糖是cbh1和cbh2表达最有效的诱导物。D-木糖以浓度依赖性方式调节xyn1和xyn2表达。

[0010] 通常,常常通过固体培养或浸没培养(包括分批、补料分批、和连续流动过程)进行丝状真菌(如木霉属)对酶/多肽的商业规模生产。例如,木霉属中工业纤维素酶生产中最成问题和最昂贵的方面之一是向木霉属宿主细胞提供合适的诱导物(即,诱导底物)。例如,与实验室规模实验的情况一样,商业规模上的纤维素酶(酶)生产通过在固体纤维素(即,诱导底物)上使真菌细胞生长或通过在二糖诱导物(如“乳糖”(即,诱导底物))存在下培养细胞来“诱导”。

[0011] 不幸的是,在工业规模上,两种“诱导”方法都具有缺点,这导致与纤维素酶生产相关的高成本。例如,如上所述,纤维素酶合成受制于纤维素诱导和葡萄糖抑制。因此,影响在诱导型启动子控制下纤维素酶产量的关键因素是维持纤维素底物与葡萄糖浓度之间的适当平衡(即,获得调控基因产物的合理的商业产量是至关重要的)。尽管纤维素是有效且廉价的诱导物,但是当丝状真菌细胞生长在固体纤维素上时,控制葡萄糖浓度可能是有问题的。在低浓度的纤维素下,葡萄糖产生可能太慢以至于不能满足活性细胞生长和功能的代谢需要。在另一方面,当葡萄糖生成快于其消耗时,葡萄糖抑制可以停止纤维素酶合成。因此,需要昂贵的工艺控制方案来提供缓慢的底物添加和葡萄糖浓度的监测(Ju和Afolabi,1999)。此外,由于纤维素材料的固体性质,底物的缓慢连续递送可能难以实现。

[0012] 通过使用可溶性“诱导底物”(如“乳糖”、“槐糖”或“龙胆二糖”)可以克服与使用纤维素作为“诱导底物”相关的一些问题。例如,当使用乳糖作为“诱导底物”时,乳糖必须以高浓度提供,以便起诱导物和碳源的作用(例如,参见Seiboth等人,2002)。槐糖是比纤维素更有效的诱导物,但是槐糖价格昂贵且难以制造。例如,葡萄糖、槐糖和其他二糖的混合物(即,通过葡萄糖的酶促转化产生)可以用于纤维素酶的有效生产,这导致比单独使用葡萄糖更高的(生产)成本。因此,虽然它比固体纤维素更容易处理和控制,但是使用槐糖作为诱导底物仍然使得生产纤维素酶的成本极其昂贵。

[0013] 基于前述内容,显然本领域对于以下仍然存在持续且未满足的需求:通过丝状真菌而不需要或不要求提供昂贵的诱导底物(例如,槐糖、乳糖等)进行酶/多肽的成本有效的

商业规模的生产。更特别地,本领域仍然存在通过丝状真菌宿主细胞商业规模生产一种或多种内源木质纤维素降解酶的需要,其中此类丝状真菌细胞能够在没有诱导底物的情况下表达这些基因中的一种或多种。此外,本领域对于以下还存在未满足的需求:成本有效地生产在此类丝状真菌宿主细胞中表达和产生的一种或多种异源蛋白产物,其中将编码此类蛋白的异源基因引入真菌宿主细胞中,该宿主细胞具有在不存在诱导底物的情况下表达这些异源基因的能力。

发明内容

[0014] 本公开的某些实施例涉及通过丝状真菌商业规模生产酶/多肽,而不需要或不要求提供用于此类生产的昂贵的诱导底物(例如,槐糖、乳糖等)。因此,某些其他实施例涉及变体丝状真菌细胞、其组合物及其用于增加一种或多种目的蛋白的产生的方法。例如,本公开的某些实施例针对衍生自亲本丝状真菌细胞的变体丝状真菌细胞,该变体细胞包含编码 Ace3 蛋白的引入的多核苷酸构建体,该 Ace3 蛋白包含与 SEQ ID NO:6 的 Ace3 蛋白约 90% 序列同一性,其中相对于亲本细胞,该变体细胞在不存在诱导底物的情况下产生增加量的目的蛋白(POI),其中该变体和亲本细胞在相似条件下培养。在某些其他实施例中,相对于亲本细胞,变体细胞在诱导底物存在下产生增加量的 POI,其中该变体和亲本细胞在相似条件下培养。

[0015] 在变体细胞的另外的实施例中,包含与 SEQ ID NO:6 约 90% 序列同一性的编码的 Ace3 蛋白包含“Lys-Ala-Ser-Asp”作为最后四个 C-末端氨基酸。在另外的实施例中,包含与 SEQ ID NO:6 约 90% 序列同一性的编码的 Ace3 蛋白包含有效地连接的并且在 SEQ ID NO:6 之前的 SEQ ID NO:98 的 N-末端氨基酸片段。在变体细胞的某些其他实施例中,Ace-3 蛋白包含与 SEQ ID NO:12 约 90% 序列同一性。在另外的实施例中,编码 Ace3 蛋白的引入的多核苷酸包含可读框 (ORF) 序列,该可读框 (ORF) 序列包含与 SEQ ID NO:5 约 90% 同一性的。

[0016] 在变体细胞的又其他实施例中,POI 是内源 POI 或异源 POI。因此,在某些实施例中,变体细胞包含编码异源 POI 的引入的多核苷酸构建体。在另外的实施例中,编码异源 POI 的多核苷酸构建体在纤维素诱导型基因启动子的控制下表达。在某些其他实施例中,内源 POI 是木质纤维素降解酶。因此,在其他实施例中,木质纤维素降解酶选自由纤维素酶、半纤维素酶、或其组合组成的组。在某些其他实施例中,木质纤维素降解酶选自由 cbh1、cbh2、eg11、eg12、eg13、eg14、eg15、eg16、bg11、bg12、xyn1、xyn2、xyn3、bx11、abf1、abf2、abf3、axe1、axe2、axe3、man1、agl11、agl12、agl13、g1r1、swo1、cip1 和 cip2 组成的组。在又其他实施例中,异源 POI 选自由以下组成的组: α -淀粉酶、碱性 α -淀粉酶、 β -淀粉酶、纤维素酶、葡聚糖酶、 α -葡萄糖苷酶、 α -半乳糖苷酶、葡糖淀粉酶、半纤维素酶、戊聚糖酶、木聚糖酶、转化酶、乳糖酶、柚皮苷酶(naringanase)、果胶酶、支链淀粉酶、酸性蛋白酶、碱性蛋白酶、菠萝蛋白酶、中性蛋白酶、木瓜蛋白酶、胃蛋白酶、肽酶、皱胃酶(rennet)、高血压蛋白原酶(rennin)、凝乳酶(chymosin)、枯草杆菌蛋白酶、嗜热菌蛋白酶、天冬氨酸蛋白酶、胰蛋白酶、脂肪酶、酯酶、磷脂酶、磷酸酶、植酸酶、酰胺酶、亚氨基酰化酶、谷氨酰胺酶、溶菌酶、青霉素酰化酶;异构酶、氧化还原酶、过氧化氢酶、氯过氧化物酶、葡萄糖氧化酶、羟基类固醇脱氢酶、过氧化物酶、裂解酶、天冬氨酸 β -脱羧酶、富马酸酶、组氨酸酶、转移酶、连接酶、氨肽酶、羧肽酶、几丁质酶、角质酶、脱氧核糖核酸酶、 α -半乳糖苷酶、 β -半乳糖苷酶、 β -葡萄糖苷酶、漆酶、甘露

糖苷酶、齿斑葡聚糖酶(mutanase)、多酚氧化酶、核糖核酸酶和转谷氨酰胺酶。

[0017] 在变体细胞的另外的实施例中,多核苷酸构建体包含启动子序列在编码Ace3蛋白的多核苷酸序列5'端并有效地连接到编码Ace3蛋白的多核苷酸序列,该Ace3蛋白包含与SEQ ID N0:6约90%序列同一性。在某些实施例中,多核苷酸构建体进一步包含在编码Ace3蛋白的多核苷酸序列3'端的并有效地连接到编码Ace3蛋白的多核苷酸序列的天然ace3终止子序列,该Ace3蛋白包含与SEQ ID N0:6约90%序列同一性。在某些其他实施例中,将多核苷酸构建体整合到真菌细胞基因组中。在某些实施例中,将多核苷酸构建体整合到真菌细胞基因组的端粒位点中。在某些其他实施例中,将多核苷酸构建体整合到真菌细胞基因组的葡糖淀粉酶(gla1)基因座中。在又其他实施例中,多核苷酸构建体包含核苷酸序列,该核苷酸序列包含与SEQ ID N0:4、SEQ ID N0:11或SEQ ID N0:13约90%序列同一性。在其他实施例中,编码的Ace3包含氨基酸序列,该氨基酸序列包含与SEQ ID N0:6、SEQ ID N0:12或SEQ ID N0:14蛋白95%序列同一性。因此,在其他实施例中,编码的Ace3蛋白包含SEQ ID N0:6、SEQ ID N0:12或SEQ ID N0:14的氨基酸序列。

[0018] 在其他实施例中,变体细胞包括遗传修饰,该遗传修饰表达编码SEQ ID N0:48的野生型木聚糖酶调节子1(Xyr1)蛋白或SEQ ID N0:46的变体木聚糖酶调节子1(Xyr1)蛋白的多核苷酸。在某些其他实施例中,变体细胞包含遗传修饰,其减少或阻止编码内源碳分解代谢物阻遏物1(Cre1)蛋白或Ace1蛋白的基因的表达。在又另外的实施例中,变体细胞包含遗传修饰,该遗传修饰包含表达Ace2蛋白。在其他实施例中,丝状真菌细胞是子囊菌(Ascomycota)亚门的盘菌亚门(Pezizomycotina)细胞。在某些其他实施例中,丝状真菌细胞是木霉属物种(*Trichoderma* sp.)细胞。

[0019] 在其他实施例中,本公开涉及编码Ace3蛋白的多核苷酸ORF,该Ace3蛋白包含与SEQ ID N0:6、SEQ ID N0:12或SEQ ID N0:14约90%序列同一性。

[0020] 在其他实施例中,本公开涉及由本公开的变体细胞产生的木质纤维素降解酶。在其他实施例中,本公开涉及由本公开的细胞的变体细胞产生的异源POI。

[0021] 在其他实施例中,本公开涉及在不存在诱导底物的情况下在木霉属物种真菌细胞中产生目的内源蛋白的方法,该方法包括(i)向该真菌细胞中引入多核苷酸构建体,该多核苷酸构建体在5'至3'方向包含(a)包含启动子的第一核酸序列和(b)有效地连接到该第一核酸序列的第二核酸序列,其中该第二核酸序列编码包含与SEQ ID N0:6约90%序列同一性的Ace3蛋白,并且包含“Lys-Ala-Ser-Asp”作为最后四个C-末端氨基酸,以及(ii)在适合于该真菌细胞生长和蛋白产生的条件下发酵步骤(i)的细胞,其中此类合适的生长条件不包括诱导底物。在该方法的某些实施例中,该多核苷酸构建体包含在所述第二核酸3'端的并有效地连接到第二核酸的第三核酸序列,其中所述第三核酸序列包含天然ace3终止子序列。在另外的实施例中,将多核苷酸构建体整合到真菌细胞基因组中。在某些实施例中,将多核苷酸构建体整合到真菌细胞基因组的端粒位点中。在某些其他实施例中,将多核苷酸构建体整合到真菌细胞基因组的葡糖淀粉酶(gla1)基因座中。在该方法的其他实施例中,多核苷酸构建体包含核苷酸序列,该核苷酸序列包含与SEQ ID N0:4、SEQ ID N0:11或SEQ ID N0:13约90%序列同一性。在其他实施例中,编码的Ace3蛋白包含氨基酸序列,该氨基酸序列包含与SEQ ID N0:6、SEQ ID N0:12或SEQ ID N0:14 95%序列同一性。在另外的实施例中,编码的Ace3蛋白包含SEQ ID N0:6、SEQ ID N0:12或SEQ ID N0:14的氨基酸序列。

[0022] 在该方法的某些实施例中,步骤(i)启动子选自由以下组成的组:rev3启动子(SEQ ID NO:15)、bx1启动子(SEQ ID NO:16)、tk11启动子(SEQ ID NO:17)、PID104295启动子(SEQ ID NO:18)、d1d1启动子(SEQ ID NO:19)、xyn4启动子(SEQ ID NO:20)、PID72526启动子(SEQ ID NO:21)、axe1启动子(SEQ ID NO:22)、hxk1启动子(SEQ ID NO:23)、dic1启动子(SEQ ID NO:24)、opt启动子(SEQ ID NO:25)、gut1启动子(SEQ ID NO:26)、和pki1启动子(SEQ ID NO:27)。

[0023] 在该方法的相关的实施例中,内源POI是木质纤维素降解酶。在某些实施例中,木质纤维素降解酶选自由纤维素酶、半纤维素酶、或其组合组成的组。在某些实施例中,纤维素酶选自由以下组成的组:cbh1、cbh2、egl11、egl12、egl13、egl14、egl15、egl16、bg11、bg12、swo1、cip1和cip2。在另外的实施例中,半纤维素酶选自由以下组成的组:xyn1、xyn2、xyn3、xyn4、bx11、abf1、abf2、abf3、axe1、axe2、axe3、man1、agl11、agl12、agl13和glr1。

[0024] 在方法的其他实施例中,步骤(i)进一步包含编码异源POI的引入的多核苷酸构建体。在某些实施例中,编码异源POI的多核苷酸构建体在纤维素诱导型基因启动子的控制下表达。在某些实施例中,纤维素诱导型基因启动子选自cbh1、cbh2、egl11、egl12、xyn2或stp1。

[0025] 在其他实施例中,该方法进一步包括遗传修饰,该遗传修饰表达编码SEQ ID NO:48的野生型木聚糖酶调节子1(Xyr1)蛋白或SEQ ID NO:46的变体木聚糖酶调节子1(Xyr1)蛋白的多核苷酸。在其他实施例中,该方法进一步包括遗传修饰,其减少或阻止编码内源碳分解代谢物阻遏物1(Cre1)蛋白或Ace1蛋白的基因的表达。在又另外的实施例中,该方法进一步包括遗传修饰,该遗传修饰包含表达Ace2蛋白。

[0026] 在其他实施例中,本公开涉及在不存在诱导底物的情况下在木霉属物种真菌细胞中产生目的内源蛋白的方法,该方法包括:(i)向该真菌细胞中引入多核苷酸构建体,该多核苷酸构建体在5'至3'方向包含(a)包含启动子的第一核酸序列和(b)有效地连接到该第一核酸序列的第二核酸序列,其中该第二核酸序列编码包含与SEQ ID NO:12约90%序列同一性的Ace3蛋白以及(ii)在适合于该真菌细胞生长和蛋白产生的条件下发酵步骤(i)的细胞,其中此类合适的生长条件不包括诱导底物。在其他实施例中,该多核苷酸构建体包含在所述第二核酸3'端的并有效地连接到第二核酸的第三核酸序列,其中所述第三核酸序列包含天然ace3终止子序列。在该方法的其他实施例中,将多核苷酸构建体整合到真菌细胞基因组中。在某些实施例中,将多核苷酸构建体整合到真菌细胞基因组的端粒位点中。在某些其他实施例中,将多核苷酸构建体整合到真菌细胞基因组的葡糖淀粉酶(glA1)基因座中。在该方法的其他实施例中,多核苷酸构建体包含核苷酸序列,该核苷酸序列包含与SEQ ID NO:4、SEQ ID NO:11或SEQ ID NO:13约90%序列同一性。在某些实施例中,编码的Ace3蛋白包含氨基酸序列,该氨基酸序列包含与SEQ ID NO:6、SEQ ID NO:12或SEQ ID NO:14 95%序列同一性。在某些其他实施例中,编码的Ace3蛋白包含SEQ ID NO:6、SEQ ID NO:12或SEQ ID NO:14的氨基酸序列。

[0027] 因此,在方法8的其他实施例中,内源POI是木质纤维素降解酶。在特别的实施例中,木质纤维素降解酶选自由纤维素酶、半纤维素酶、或其组合组成的组。在另外的实施例中,纤维素酶选自由以下组成的组:cbh1、cbh2、egl11、egl12、egl13、egl14、egl15、egl16、bg11、bg12、swo1、cip1和cip2。在其他实施例中,半纤维素酶选自由以下组成的组:xyn1、xyn2、xyn3、xyn4、bx11、abf1、abf2、abf3、axe1、axe2、axe3、man1、agl11、agl12、agl13和glr1。

[0028] 在该方法的另外的实施例中,步骤(i)进一步包含编码异源POI的引入的多核苷酸构建体。在某些实施例中,编码异源POI的多核苷酸构建体在纤维素诱导型基因启动子的控制下表达。在该方法的某些其他实施例中,步骤(i)启动子选自由以下组成的组:rev3启动子(SEQ ID NO:15)、bxl启动子(SEQ ID NO:16)、tkl1启动子(SEQ ID NO:17)、PID104295启动子(SEQ ID NO:18)、d1d1启动子(SEQ ID NO:19)、xyn4启动子(SEQ ID NO:20)、PID72526启动子(SEQ ID NO:21)、axe1启动子(SEQ ID NO:22)、hxk1启动子(SEQ ID NO:23)、dic1启动子(SEQ ID NO:24)、opt启动子(SEQ ID NO:25)、gut1启动子(SEQ ID NO:26)、和pki1启动子(SEQ ID NO:27)。在另外的实施例中,该方法进一步包括遗传修饰,该遗传修饰表达编码SEQ ID NO:48的野生型木聚糖酶调节子1(Xyr1)蛋白或SEQ ID NO:46的变体木聚糖酶调节子1(Xyr1)蛋白的多核苷酸。在某些其他实施例中,该方法进一步包括遗传修饰,其减少或阻止编码内源碳分解代谢物阻遏物1(Cre1)蛋白或Ace1蛋白的基因的表达。在又其他实施例中,该方法进一步包括遗传修饰,该遗传修饰包含表达Ace2蛋白。

[0029] 在某些实施例中,本公开涉及在不存在诱导底物的情况下在木霉属物种真菌细胞中产生目的异源蛋白的方法,该方法包括:(i)向该真菌细胞中引入多核苷酸构建体,该多核苷酸构建体在5'至3'方向包含(a)包含组成型启动子的第一核酸序列和(b)有效地连接到该第一核酸序列的第二核酸序列,其中该第二核酸序列编码包含与SEQ ID NO:6约90%序列同一性的Ace3蛋白,并且包含“Lys-Ala-Ser-Asp”作为最后四个C-末端氨基酸,以及(ii)在适合于该真菌细胞生长和蛋白产生的条件下发酵步骤(i)的细胞,其中此类合适的生长条件不包括诱导底物。因此,在该方法的某些实施例中,真菌细胞包含编码异源POI的引入的多核苷酸构建体,其中在步骤(i)之前、在步骤(i)期间或在步骤(i)之后将该构建体引入真菌细胞中。在另外的实施例中,该多核苷酸构建体包含在所述第二核酸3'端的并有效地连接到第二核酸的第三核酸序列,其中所述第三核酸序列包含天然ace3终止子序列。在该方法的其他实施例中,将多核苷酸构建体整合到真菌细胞基因组中。在某些实施例中,将多核苷酸构建体整合到真菌细胞基因组的端粒位点中。在其他实施例中,将多核苷酸构建体整合到真菌细胞基因组的葡糖淀粉酶(gla1)基因座中。

[0030] 在该方法的某些其他实施例中,多核苷酸构建体包含核苷酸序列,该核苷酸序列包含与SEQ ID NO:4、SEQ ID NO:11或SEQ ID NO:13约90%序列同一性。在另外的实施例中,编码的Ace3蛋白包含氨基酸序列,该氨基酸序列包含与SEQ ID NO:6、SEQ ID NO:12或SEQ ID NO:14 95%序列同一性。在其他实施例中,编码的Ace3蛋白包含SEQ ID NO:6、SEQ ID NO:12或SEQ ID NO:14的氨基酸序列。在其他实施例中,编码异源POI的多核苷酸构建体在纤维素诱导型基因启动子的控制下表达。在某些其他实施例中,异源POI选自由以下组成的组: α -淀粉酶、碱性 α -淀粉酶、 β -淀粉酶、纤维素酶、葡聚糖酶、 α -葡萄糖苷酶、 α -半乳糖苷酶、葡糖淀粉酶、半纤维素酶、戊聚糖酶、木聚糖酶、转化酶、乳糖酶、柚皮苷酶(naringanase)、果胶酶、支链淀粉酶、酸性蛋白酶、碱性蛋白酶、菠萝蛋白酶、中性蛋白酶、木瓜蛋白酶、胃蛋白酶、肽酶、皱胃酶(rennet)、高血压蛋白原酶(rennin)、凝乳酶(chymosin)、枯草杆菌蛋白酶、嗜热菌蛋白酶、天冬氨酸蛋白酶、胰蛋白酶、脂肪酶、酯酶、磷脂酶、磷酸酶、植酸酶、酰胺酶、亚氨基酰化酶、谷氨酰胺酶、溶菌酶、青霉素酰化酶;异构酶、氧化还原酶、过氧化氢酶、氯过氧化物酶、葡萄糖氧化酶、羟基类固醇脱氢酶、过氧化物酶、裂解酶、天冬氨酸 β -脱羧酶、富马酸酶、组氨酸酶、转移酶、连接酶、氨肽酶、羧肽酶、几丁质

酶、角质酶、脱氧核糖核酸酶、 α -半乳糖苷酶、 β -半乳糖苷酶、 β -葡糖苷酶、漆酶、甘露糖苷酶、齿斑葡聚糖酶(mutanase)、多酚氧化酶、核糖核酸酶和转谷氨酰胺酶。

[0031] 在该方法的其他实施例中,步骤(i)启动子选自由以下组成的组:rev3启动子(SEQ ID NO:15)、bx1启动子(SEQ ID NO:16)、tk11启动子(SEQ ID NO:17)、PID104295启动子(SEQ ID NO:18)、d1d1启动子(SEQ ID NO:19)、xyn4启动子(SEQ ID NO:20)、PID72526启动子(SEQ ID NO:21)、axe1启动子(SEQ ID NO:22)、hxk1启动子(SEQ ID NO:23)、dic1启动子(SEQ ID NO:24)、opt启动子(SEQ ID NO:25)、gut1启动子(SEQ ID NO:26)、和pki1启动子(SEQ ID NO:27)。在又其他实施例中,该方法进一步包括遗传修饰,该遗传修饰表达编码SEQ ID NO:48的野生型木聚糖酶调节子1(Xyr1)蛋白或SEQ ID NO:46的变体木聚糖酶调节子1(Xyr1)蛋白的多核苷酸。在另外的实施例中,该方法进一步包括遗传修饰,其减少或阻止编码内源碳分解代谢物阻遏物1(Cre1)蛋白或Ace1蛋白的基因的表达。在又其他实施例中,该方法进一步包括遗传修饰,该遗传修饰包含表达Ace2蛋白。

[0032] 在另外的实施例中,本公开涉及在不存在诱导底物的情况下在木霉属物种真菌细胞中产生目的异源蛋白的方法,该方法包括:(i)向该真菌细胞中引入多核苷酸构建体,该多核苷酸构建体在5'至3'方向包含(a)包含组成型启动子的第一核酸序列和(b)有效地连接到该第一核酸序列的第二核酸序列,其中该第二核酸序列编码包含与SEQ ID NO:12约90%序列同一性的Ace3蛋白以及(ii)在适合于该真菌细胞生长和蛋白产生的条件下发酵步骤(i)的细胞,其中此类合适的生长条件不包括诱导底物。在特别的实施例中,真菌细胞包含编码异源POI的引入的多核苷酸构建体,其中在步骤(i)之前、在步骤(i)期间或在步骤(i)之后将该构建体引入真菌细胞中。在其他实施例中,该多核苷酸构建体包含在所述第二核酸3'端的并有效地连接到第二核酸的第三核酸序列,其中所述第三核酸序列包含天然ace3终止子序列。在该方法的其他实施例中,将多核苷酸构建体整合到真菌细胞基因组中。在某些实施例中,将多核苷酸构建体整合到真菌细胞基因组的端粒位点中。在其他实施例中,将多核苷酸构建体整合到真菌细胞基因组的葡糖淀粉酶(gla1)基因座中。在又其他实施例中,多核苷酸构建体包含核苷酸序列,该核苷酸序列包含与SEQ ID NO:4、SEQ ID NO:11或SEQ ID NO:13约90%序列同一性。在某些实施例中,编码的Ace3蛋白包含氨基酸序列,该氨基酸序列包含与SEQ ID NO:6、SEQ ID NO:12或SEQ ID NO:14 95%序列同一性。在另外的实施例中,编码的Ace3蛋白包含SEQ ID NO:6、SEQ ID NO:12或SEQ ID NO:14的氨基酸序列。在该方法的另外的实施例中,编码异源POI的多核苷酸构建体在纤维素诱导型基因启动子的控制下表达。在某些实施例中,异源POI选自由以下组成的组: α -淀粉酶、碱性 α -淀粉酶、 β -淀粉酶、纤维素酶、葡聚糖酶、 α -葡糖苷酶、 α -半乳糖苷酶、葡糖淀粉酶、半纤维素酶、戊聚糖酶、木聚糖酶、转化酶、乳糖酶、柚皮苷酶(naringanase)、果胶酶、支链淀粉酶、酸性蛋白酶、碱性蛋白酶、菠萝蛋白酶、中性蛋白酶、木瓜蛋白酶、胃蛋白酶、肽酶、皱胃酶(rennet)、高血压蛋白原酶(rennin)、凝乳酶(chymosin)、枯草杆菌蛋白酶、嗜热菌蛋白酶、天冬氨酸蛋白酶、胰蛋白酶、脂肪酶、酯酶、磷脂酶、磷酸酶、植酸酶、酰胺酶、亚氨基酰化酶、谷氨酰胺酶、溶菌酶、青霉素酰化酶;异构酶、氧化还原酶、过氧化氢酶、氯过氧化物酶、葡萄糖氧化酶、羟基类固醇脱氢酶、过氧化物酶、裂解酶、天冬氨酸 β -脱羧酶、富马酸酶、组氨酸酶、转移酶、连接酶、氨肽酶、羧肽酶、几丁质酶、角质酶、脱氧核糖核酸酶、 α -半乳糖苷酶、 β -半乳糖苷酶、 β -葡糖苷酶、漆酶、甘露糖苷酶、齿斑葡聚糖酶(mutanase)、多酚氧化酶、核糖

核酸酶和转谷氨酰胺酶。

[0033] 在该方法的其他实施例中,步骤(i)启动子选自由以下组成的组:rev3启动子(SEQ ID NO:15)、bx1启动子(SEQ ID NO:16)、tk11启动子(SEQ ID NO:17)、PID104295启动子(SEQ ID NO:18)、d1d1启动子(SEQ ID NO:19)、xyn4启动子(SEQ ID NO:20)、PID72526启动子(SEQ ID NO:21)、axe1启动子(SEQ ID NO:22)、hxk1启动子(SEQ ID NO:23)、dic1启动子(SEQ ID NO:24)、opt启动子(SEQ ID NO:25)、gut1启动子(SEQ ID NO:26)、和pki1启动子(SEQ ID NO:27)。在又其他实施例中,该方法进一步包括遗传修饰,该遗传修饰表达编码SEQ ID NO:48的野生型木聚糖酶调节子1(Xyr1)蛋白或SEQ ID NO:46的变体木聚糖酶调节子1(Xyr1)蛋白的多核苷酸。在另外的实施例中,该方法进一步包括遗传修饰,其减少或阻止编码内源碳分解代谢物阻遏物1(Cre1)蛋白或Ace1蛋白的基因的表达。在又其他实施例中,该方法进一步包括遗传修饰,该遗传修饰包含表达Ace2蛋白。

[0034] 在某些其他实施例中,本公开涉及遗传修饰里氏木霉(*Trichoderma reesei*)菌株用于在不存在诱导底物的情况下增加内源蛋白的产生的方法,该方法包括:(i)筛选和鉴定里氏木霉菌株,所述里氏木霉菌株包含编码SEQ ID NO:3的Ace3-S蛋白、SEQ ID NO:8的Ace3-SC蛋白、或SEQ ID NO:10的Ace3-LC蛋白的ace3基因的基因组拷贝,其中鉴定的所述里氏木霉菌株不包含编码SEQ ID NO:6的Ace3-L蛋白、SEQ ID NO:14的Ace3-LN蛋白、或SEQ ID NO:12的Ace3-EL蛋白的ace3基因的基因组拷贝,(ii)向步骤(i)中鉴定的里氏木霉菌株中引入多核苷酸构建体,所述多核苷酸构建体在5'到3'方向包含:(a)包含启动子的第一核酸序列和(b)有效地连接到该第一核酸序列的第二核酸序列,其中该第二核酸序列编码包含与SEQ ID NO:6约90%序列同一性的Ace3蛋白并包含“Lys-Ala-Ser-Asp”作为最后四个C-末端氨基酸,或者该第二核酸序列编码包含与SEQ ID NO:12约90%序列同一性的Ace-3蛋白,以及(iii)在适于真菌细胞生长和蛋白产生的条件下发酵步骤(ii)的细胞,其中此类合适的生长条件不包括诱导底物。

[0035] 在某些其他实施例中,本公开针对衍生自亲本丝状真菌细胞的变体丝状真菌细胞,所述变体细胞包含天然ace3基因启动子被可替代的启动子替换的,其中相对于亲本细胞,所述变体细胞在不存在诱导底物的情况下产生增加量的目的蛋白(POI),其中所述变体和亲本细胞在相似条件下培养。在特别的实施例中,可替代的启动子是里氏木霉启动子。在另外的实施例中,可替代的启动子是选自由以下组成的组的启动子:rev3启动子(SEQ ID NO:15)、bx1启动子(SEQ ID NO:16)、tk11启动子(SEQ ID NO:17)、PID104295启动子(SEQ ID NO:18)、d1d1启动子(SEQ ID NO:19)、xyn4启动子(SEQ ID NO:20)、PID72526启动子(SEQ ID NO:21)、axe1启动子(SEQ ID NO:22)、hxk1启动子(SEQ ID NO:23)、dic1启动子(SEQ ID NO:24)、opt启动子(SEQ ID NO:25)、gut1启动子(SEQ ID NO:26)、和pki1启动子(SEQ ID NO:27)。

附图说明

[0036] 图1展示Ace3蛋白编码区的示意图。更特别地,图1A展示基于里氏木霉菌株QM6a和RUT-C30的注释的Ace3蛋白编码区的示意图,其与基因组中的相同DNA序列对齐(图1A)。预测的外显子和内含子分别显示为箭头和虚线。垂直虚线指示RUT-C30基因组中的无义突变。如本申请中所示筛选并测试SEQ ID NO:2的克隆的ace3-S(短)可读框和SEQ ID NO:5的克

隆的ace3-L(长)可读框。图1B中所示的是来自里氏木霉菌株RUT-C30的Ace3-L蛋白(SEQ ID NO:6)、来自里氏木霉菌株QM6a的Ace3-S蛋白(SEQ ID NO:3)、Ace3-SC蛋白(SEQ ID NO:8)、Ace3-LN蛋白(SEQ ID NO:14)、Ace3-LC蛋白(SEQ ID NO:10)、和Ace3-EL蛋白(SEQ ID NO:12)的氨基酸对齐。

[0037] 图2显示了Ace3表达载体pYL1(图2A)、pYL2(图2B)、pYL3(图2C)和pYL4(图2D)的示意图。

[0038] 图3展示里氏木霉亲本和变体细胞上清液的聚丙烯酰胺凝胶电泳(SDS-PAGE)，其中所述亲本和变体细胞在srMTP中的含有20%乳糖(lac)或20%葡萄糖(glu)的限定培养基中生长。在每个泳道中加载等体积的培养物上清液。M是分子量标记，并且将亲本里氏木霉菌株用作对照菌株。

[0039] 图4展示里氏木霉亲本和变体细胞上清液的SDS-PAGE，其中所述亲本和变体细胞在摇瓶中在补充有的1.5%葡萄糖/槐糖(sop)或1.5%葡萄糖(glu)的限定培养基中生长。在每个泳道中加载等体积的培养物上清液。培养物上清液的总蛋白浓度列于每个相应泳道的底部。M是分子量标记，并且KD是千道尔顿。

[0040] 图5展示在2L发酵罐中在限定培养基(其中葡萄糖/槐糖(sop)或葡萄糖(glu)作为碳源)中生长的里氏木霉亲本和变体细胞的SDS-PAGE。M是分子量标记，零(0)是种子培养物上清液。

[0041] 图6展示启动子替换构建体的示意图，其是通过将包含ace3基因座上的天然启动子的上游5'区的DNA片段、loxP-侧翼的潮霉素B抗性(选择性标记)盒、和包含有效融合(连接)至ace30RF的5'末端的目的启动子的DNA片段融合制成。

[0042] 图7显示了包含dicl1启动子的Ace3-表达载体pYL8的示意图。

[0043] 图8展示了里氏木霉亲本及其经修饰的(子代)细胞上清液的SDS-PAGE，其中所述亲本和经修饰的菌株在摇瓶中在补充有2.5%葡萄糖/槐糖(“Sop”，诱导条件)或2.5%葡萄糖(“Glu”，非诱导条件)的限定培养基中生长。在每个泳道中加载等体积的培养物上清液。M是分子量标记，并且KD是千道尔顿。

[0044] 图9显示了小规模(2L)发酵中的蛋白产生。里氏木霉亲本菌株和子代菌株“LT83”在限定培养基(其中葡萄糖/槐糖(Sop，诱导条件)或葡萄糖(Glu，非诱导条件)作为碳源)中生长。在发酵结束时由亲本菌株在葡萄糖/槐糖上产生的总蛋白任意地设定为100，并且绘制每个菌株在每个时间点产生的蛋白的相对量。

[0045] 图10展示了里氏木霉亲本及其经修饰的(子代)细胞上清液的SDS-PAGE，其中所述亲本和经修饰的菌株在摇瓶中在补充有2.5%葡萄糖/槐糖(“Sop”，诱导条件)或2.5%葡萄糖(“Glu”，非诱导条件)的限定培养基中生长。在每个泳道中加载等体积的培养物上清液。M是分子量标记，并且KD是千道尔顿。

[0046] 图11表示了ace3基因座的示意图。在ace3基因座的5'-端(N-末端)的箭头指示在cDNA序列(箭头1)建议的形式、RutC-30注释形式(箭头2)和QM6a注释形式(箭头3)中的不同转录起始位点。在ace3基因座的3'-端(C-末端)的箭头指示在RutC-30注释形式(箭头4)和QM6a注释形式(箭头5)中的不同终止密码子。

[0047] 图12显示了克隆的不同ace3形式的示意图。因此，如图12中表示以及如实例6中描述，克隆并测试了以下ace3形式：SEQ ID NO:1的“ace3-S”(包含1,713bp外显子3、148bp内

含子3和177bp外显子4)、SEQ ID N0:7的“ace3-SC”(包含1,713bp外显子3、148bp内含子3和144bp外显子4)、SEQ ID N0:4的“ace3-L”(包含258bp外显子2、423bp内含子2、1,635bp外显子3、148bp内含子3和144bp外显子4)、SEQ ID N0:9的“ace3-LC”(包含258bp外显子2、423bp内含子2、1,635bp外显子3、148bp内含子3和177bp外显子4)、SEQ ID N0:11的“ace3-EL”(包含61bp外显子1、142bp内含子1、332bp外显子2、423bp内含子2、1,635bp外显子3、148bp内含子3和144bp外显子4)、和SEQ ID N0:13的“ace3-LN”(包含258bp外显子2、1,635bp外显子3、148bp内含子3和144bp外显子4)。

[0048] 图13显示了ace3-SC基因形式的核酸序列(SEQ ID N0:7;包含1,713bp外显子3、148bp内含子3和144bp外显子4)蛋白和编码的Ace3-SC序列(SEQ ID N0:8)。如图13对于ace3-SC基因形式所示,以粗体黑色文本表示的核苷酸代表内含子序列。

[0049] 图14显示了ace3-S基因形式的核酸序列(SEQ ID N0:1;包含1,713bp外显子3、148bp内含子3和177bp外显子4)和编码的Ace3-S蛋白序列(SEQ ID N0:3)。如图14对于ace3-S基因形式的表示,以粗体黑色文本表示的核苷酸代表内含子序列。

[0050] 图15显示了ace3-L基因形式的核酸序列(SEQ ID N0:4,包含258bp外显子2、423bp内含子2、1,635bp外显子3、148bp内含子3、和144bp外显子4)和编码的Ace3-L蛋白序列(SEQ ID N0:6)。如图15对于ace3-L基因形式的表示,以粗体黑色文本表示的核苷酸代表内含子序列。

[0051] 图16显示了ace3-LC基因形式的核酸序列(SEQ ID N0:9,包含258bp外显子2、423bp内含子2、1,635bp外显子3、148bp内含子3、和177bp外显子4)和编码的Ace3-LC蛋白序列(SEQ ID N0:10)。如图16对于ace3-LC基因形式的表示,以粗体黑色文本表示的核苷酸代表内含子序列。

[0052] 图17显示了ace3-EL基因形式的核酸序列(SEQ ID N0:11,包含61bp外显子1、142bp内含子1、332bp外显子2、423bp内含子2、1,635bp外显子3、148bp内含子3、和144bp外显子4)和编码的Ace3-EL蛋白序列(SEQ ID N0:12)。如图17对于ace3-EL基因形式的表示,以粗体黑色文本表示的核苷酸代表内含子序列。

[0053] 图18显示了ace3-LN基因形式的核酸序列(SEQ ID N0:13,包含258bp外显子2、1,635bp外显子3、148bp内含子3、和144bp外显子4)和编码的Ace3-LN蛋白序列(SEQ ID N0:14)。如图18对于ace3-LN基因形式的表示,以粗体黑色文本表示的核苷酸代表内含子序列。

[0054] 图19展示设计用于在g1a1基因座处整合ace3形式的DNA片段的排列的示意图。

[0055] 图20显示了载体pMCM3282的示意图。

[0056] 图21显示了载体pCHL760的示意图。

[0057] 图22显示了载体pCHL761的示意图。

[0058] 图23展示在诱导(“Glu/Sop”)和非诱导(“Glu”)培养条件下,里氏木霉亲本细胞(图23,细胞ID 1275.8.1)和变体里氏木霉(子代)细胞(图23,细胞ID号2218、2219、2220、2222和2223)的浸没培养(即,摇瓶)产生的分泌蛋白的SDS-PAGE。

[0059] 生物学序列简述

[0060] 以下序列遵循37C.F.R.§§1.821-1.825(“含有核苷酸序列和/或氨基酸序列公开内容的专利申请的要求-序列规则”)并符合WIP标准ST.25(2009)和欧洲专利公约(EPC)以及专利合作条约(PCT)法规第5.2和49.5(a-bis)条,以及行政章程第208款和附件C关于序

列表的要求。用于核苷酸和氨基酸序列数据的符号和格式遵循如37C.F.R. §1.822中所述的条例。

- [0061] SEQ ID NO:1是包含编码SEQ ID NO:3的Ace3-S蛋白的基因的里氏木霉野生型菌株QM6a核酸序列。
- [0062] SEQ ID NO:2是编码SEQ ID NO:3的Ace3-S蛋白的核酸序列可读框(ORF)。
- [0063] SEQ ID NO:3是里氏木霉(菌株QM6a) Ace3蛋白(下文称为“Ace3-S”)的氨基酸序列。
- [0064] SEQ ID NO:4是包含编码SEQ ID NO:6的Ace3-L蛋白的基因的里氏木霉菌株Rut-C30核酸序列。
- [0065] SEQ ID NO:5是编码SEQ ID NO:6的Ace3-L蛋白的核酸序列ORF。
- [0066] SEQ ID NO:6是里氏木霉(菌株Rut-C30) Ace3蛋白(下文称为“Ace3-L”)的氨基酸序列。
- [0067] SEQ ID NO:7是包含编码SEQ ID NO:8的Ace3-SC蛋白的基因的里氏木霉核酸序列。
- [0068] SEQ ID NO:8是里氏木霉Ace3蛋白(下文称为“Ace3-SC”)的氨基酸序列。
- [0069] SEQ ID NO:9是包含编码SEQ ID NO:10的Ace3-LC蛋白的基因的里氏木霉核酸序列。
- [0070] SEQ ID NO:10是里氏木霉Ace3蛋白(下文称为“Ace3-LC”)的氨基酸序列。
- [0071] SEQ ID NO:11是包含编码SEQ ID NO:12的Ace3-EL蛋白的基因的里氏木霉核酸序列。
- [0072] SEQ ID NO:12是里氏木霉Ace3蛋白(下文称为“Ace3-EL”)的氨基酸序列。
- [0073] SEQ ID NO:13是包含编码SEQ ID NO:14的Ace3-LN蛋白的基因的里氏木霉核酸序列。
- [0074] SEQ ID NO:14是里氏木霉Ace3蛋白(下文称为“Ace3-LN”)的氨基酸序列。
- [0075] SEQ ID NO:15是包含rev3启动子序列的核酸序列。
- [0076] SEQ ID NO:16是包含 β -xyl启动子序列的核酸序列。
- [0077] SEQ ID NO:17是包含tki1启动子序列的核酸序列。
- [0078] SEQ ID NO:18是包含PID104295启动子序列的核酸序列。
- [0079] SEQ ID NO:19是包含d1d1启动子序列的核酸序列。
- [0080] SEQ ID NO:20是包含xyn4启动子序列的核酸序列。
- [0081] SEQ ID NO:21是包含PID72526启动子序列的核酸序列。
- [0082] SEQ ID NO:22是包含axe3启动子序列的核酸序列。
- [0083] SEQ ID NO:23是包含hxk1启动子序列的核酸序列。
- [0084] SEQ ID NO:24是包含dic1启动子序列的核酸序列。
- [0085] SEQ ID NO:25是包含opt启动子序列的核酸序列。
- [0086] SEQ ID NO:26是包含gut1启动子序列的核酸序列。
- [0087] SEQ ID NO:27是包含pki1启动子序列的核酸序列。
- [0088] SEQ ID NO:28是引物TP13的核酸序列。
- [0089] SEQ ID NO:29是引物TP14的核酸序列。

- [0090] SEQ ID NO:30是引物TP15的核酸序列。
- [0091] SEQ ID NO:31是引物TP16的核酸序列。
- [0092] SEQ ID NO:32是引物TP17的核酸序列。
- [0093] SEQ ID NO:33是引物TP18的核酸序列。
- [0094] SEQ ID NO:34是引物TP19的核酸序列。
- [0095] SEQ ID NO:35是引物TP20的核酸序列。
- [0096] SEQ ID NO:36是引物TP21的核酸序列。
- [0097] SEQ ID NO:37是引物TP22的核酸序列。
- [0098] SEQ ID NO:38是引物TP23的核酸序列。
- [0099] SEQ ID NO:39是引物TP24的核酸序列。
- [0100] SEQ ID NO:40是引物TP25的核酸序列。
- [0101] SEQ ID NO:41是引物TP26的核酸序列。
- [0102] SEQ ID NO:42是质粒pYL1的核酸序列。
- [0103] SEQ ID NO:43是质粒pYL2的核酸序列。
- [0104] SEQ ID NO:44是质粒pYL3的核酸序列。
- [0105] SEQ ID NO:45是质粒pYL4的核酸序列。
- [0106] SEQ ID NO:46是里氏木霉xyr1(A824V)变体蛋白的氨基酸序列。
- [0107] SEQ ID NO:47是里氏木霉Ace2蛋白的氨基酸序列。
- [0108] SEQ ID NO:48是里氏木霉野生型xyr1蛋白的氨基酸序列。
- [0109] SEQ ID NO:49是引物核酸序列。
- [0110] SEQ ID NO:50是引物核酸序列。
- [0111] SEQ ID NO:51是引物核酸序列。
- [0112] SEQ ID NO:52是引物核酸序列。
- [0113] SEQ ID NO:53是引物核酸序列。
- [0114] SEQ ID NO:54是引物核酸序列。
- [0115] SEQ ID NO:55是引物核酸序列。
- [0116] SEQ ID NO:56是引物核酸序列。
- [0117] SEQ ID NO:57是引物核酸序列。
- [0118] SEQ ID NO:58是引物核酸序列。
- [0119] SEQ ID NO:59是引物核酸序列。
- [0120] SEQ ID NO:60是引物核酸序列。
- [0121] SEQ ID NO:61是引物核酸序列。
- [0122] SEQ ID NO:62是引物核酸序列。
- [0123] SEQ ID NO:63是引物核酸序列。
- [0124] SEQ ID NO:64是引物核酸序列。
- [0125] SEQ ID NO:65是引物核酸序列。
- [0126] SEQ ID NO:66是引物核酸序列。
- [0127] SEQ ID NO:67是引物核酸序列。
- [0128] SEQ ID NO:68是引物核酸序列。

- [0129] SEQ ID NO:69是引物核酸序列。
- [0130] SEQ ID NO:70是引物核酸序列。
- [0131] SEQ ID NO:71是引物核酸序列。
- [0132] SEQ ID NO:72是引物核酸序列。
- [0133] SEQ ID NO:73是引物核酸序列。
- [0134] SEQ ID NO:74是引物核酸序列。
- [0135] SEQ ID NO:75是引物核酸序列。
- [0136] SEQ ID NO:76是引物核酸序列。
- [0137] SEQ ID NO:77是引物核酸序列。
- [0138] SEQ ID NO:78是引物核酸序列。
- [0139] SEQ ID NO:79是引物核酸序列。
- [0140] SEQ ID NO:80是引物核酸序列。
- [0141] SEQ ID NO:81是引物核酸序列。
- [0142] SEQ ID NO:82是人工核酸序列。
- [0143] SEQ ID NO:83是人工核酸序列。
- [0144] SEQ ID NO:84是人工核酸序列。
- [0145] SEQ ID NO:85是人工核酸序列。
- [0146] SEQ ID NO:86是人工核酸序列。
- [0147] SEQ ID NO:87是人工核酸序列。
- [0148] SEQ ID NO:88是人工核酸序列。
- [0149] SEQ ID NO:89是人工核酸序列。
- [0150] SEQ ID NO:90是人工核酸序列。
- [0151] SEQ ID NO:91是人工核酸序列。
- [0152] SEQ ID NO:92是引物核酸序列。
- [0153] SEQ ID NO:93是引物核酸序列。
- [0154] SEQ ID NO:94是引物核酸序列。
- [0155] SEQ ID NO:95是引物核酸序列。
- [0156] SEQ ID NO:96是引物核酸序列。
- [0157] SEQ ID NO:97是引物核酸序列。
- [0158] SEQ ID NO:98是包含SEQ ID NO:12的Ace3-EL蛋白的N-末端序列的四十五(45)个氨基酸的片段的氨基酸序列。
- [0159] SEQ ID NO:99是编码Ace3-SC蛋白的核酸序列ORF。
- [0160] SEQ ID NO:100是编码Ace3-LC蛋白的核酸序列ORF。
- [0161] SEQ ID NO:101是编码Ace3-EL蛋白的核酸序列ORF。
- [0162] SEQ ID NO:102是编码Ace3-LN蛋白的核酸序列ORF。

具体实施方式

- [0163] I. 概述
- [0164] 本公开的某些实施例针对变体丝状真菌细胞、其组合物及其用于增加一种或多种

目的蛋白的产生的方法。更特别地，本公开的某些实施例针对能够在不存在诱导饲料(即诱导底物，如乳糖、槐糖、龙胆二糖、纤维素等)的情况下产生一种或多种目的蛋白的变体丝状真菌细胞。因此，本公开的某些实施例针对衍生自亲本丝状真菌细胞的变体丝状真菌(宿主)细胞，其中所述变体宿主细胞包含遗传修饰，该遗传修饰使得能够在不存在诱导底物的情况下表达目的基因(编码目的蛋白)。目的基因(编码目的蛋白)可以是内源丝状真菌细胞基因(例如，cbh1、chb2、xyn1、xyn2、xyn3、egl11、egl12、egl13、bg11、bg12等)或与丝状真菌细胞异源的基因。

[0165] 因此，在某些其他实施例中，本公开的变体真菌宿主细胞包含遗传修饰，该遗传修饰增加“纤维素酶表达激活子3”(ace3)基因的表达，该基因编码选自由以下组成的组的Ace3蛋白：Ace3-L蛋白(SEQ ID N0:6)、Ace3-EL蛋白(SEQ ID N0:12)和Ace3-LN蛋白(SEQ ID N0:14)。在其他实施例中，本公开的变体真菌宿主细胞包含遗传修饰，该遗传修饰增加多核苷酸可读框(ORF)的表达，该可读框编码选自由以下组成的组的Ace3蛋白：Ace3-L蛋白(SEQ ID N0:6)、Ace3-EL蛋白(SEQ ID N0:12)和Ace3-LN蛋白(SEQ ID N0:14)。因此，在某些实施例中，本公开的变体真菌宿主细胞包含遗传修饰，该遗传修饰增加编码Ace3蛋白的Ace3基因或其ORF的表达/产生，该Ace3蛋白包含与选自由以下组成的组的Ace3蛋白约90%至约99%的序列同一性：Ace3-L蛋白(SEQ ID N0:6)、Ace3-EL蛋白(SEQ ID N0:12)和Ace3-LN蛋白(SEQ ID N0:14)。

[0166] 在某些实施例中，增加Ace3蛋白(即，Ace3-L、Ace3-EL或Ace3-LN蛋白)表达的遗传修饰是已整合到丝状真菌宿主细胞的基因组(染色体)中的ace3表达盒。在其他实施例中，增加丝状真菌细胞中Ace3蛋白表达的遗传修饰是游离维持的质粒构建体，该质粒构建体包含ace3表达盒(即，编码Ace3-L、Ace3-EL或Ace3-LN蛋白)。在其他实施例中，增加丝状真菌细胞中编码Ace3-L、Ace3-EL或Ace3-LN蛋白的ace3基因表达的遗传修饰是整合在端粒位点的端粒载体/质粒。在某些实施例中，此类表达盒或质粒以一个以上的拷贝存在。在某些其他实施例中，将ace3基因、或ace3ORF有效地连接到异源启动子。在其他实施例中，增加在丝状真菌细胞中编码Ace3-L、Ace3-EL或Ace3-LN蛋白的ace3基因(或其ORF)的表达的遗传修饰是天然ace3启动子的修饰(即，与ace3基因天然相关联的ace3启动子区域)，该修饰改变编码Ace3-L、Ace3-EL或Ace3-LN蛋白的ace3基因的表达。

[0167] II. 定义

[0168] 在进一步详细地描述本发明的组合物和方法之前，定义了以下术语和短语。未定义的术语应当符合本领域技术人员已知和使用的常规含义。

[0169] 在本说明书中引用的所有公开物以及专利都通过引用并入本文。

[0170] 在提供了一系列值的情况下，应理解每个中间值为到下限的十分之一单位(除非上下文清晰地另外指示)，该范围的上限与下限之间以及在该陈述范围内的任何其他陈述的或中间值均被涵盖在本发明的组合物和方法之内。这些较小范围的上限和下限可以被独立地包括在所述较小的范围内，并且也被涵盖在本发明的组合物和方法之内，服从所陈述范围中任何特别排除的限值。在所陈述的范围包括一个或者全部两个限值的情况下，排除那些包括的限值的任一个或者全部两个的范围也包括在本发明的组合物和方法中。

[0171] 本文提供了某些范围，其中数值前面是术语“约”。术语“约”在本文中用于为其后面的确切数字以及与该术语后面的数字接近或近似的数字提供文字支持。在确定数字是否

接近或近似于具体叙述的数字时,接近或近似的未列举的数字可以是在呈现其的上下文中提供具体叙述的数字的实质性等值的数字。例如,关于数值,术语“约”是指数值的^{-10%}至^{+10%}的范围,除非术语在上下文中另有具体定义。在另一个实例中,短语“约6的pH值”是指pH值为从5.4至6.6,除非pH值另有具体定义。

[0172] 本文提供的标题并非对本发明的组合物和方法的各个方面或实施例进行限制,这些方面或实施例可通过将说明书作为一个整体来参考而得到。因此,把说明书作为一个整体参考时,以下即将定义的术语得以更全面地定义。

[0173] 根据这一详细说明,以下缩写和定义适用。应当注意单数形式“一个/一种(a/an)”和“该(the)”包括复数个指示物,除非上下文中另有清楚地指示。因此,例如提及“酶”包括多个这样的酶,并且提及“剂量”包括提及一个或多个剂量以及本领域普通技术人员已知的其等效物,等等。

[0174] 进一步注意的是,权利要求书可以撰写以排除任何可选择的要素。因此,此陈述旨在作为使用与权利要求要素的叙述有关的排他性术语如“单独”、“仅”、“排除”、“不包括”等或使用“否定型”限定的前提基础。

[0175] 进一步注意的是,如本文所用,术语“包含”意指“包括但不限于”在术语“包含”之后的一种或多种组分。在术语“包含”之后的组分是必需的或强制性的,但是包含一种或多种组分的组合物可以进一步包括其他非强制性或可选择的一种或多种组分。

[0176] 还要注意的是,如本文所用,术语“由……组成”意指包括且限于术语“由……组成之后的一种或多种组分”。因此,术语“由……组成”之后的一种或多种组分是必需的或强制性的,且组合物中不存在一种或多种其他组分。

[0177] 如将对于本领域技术人员显而易见的是,在阅读本公开时,本文描述和展示的单独实施例中的每一个具有离散的组分和特征,这些组分和特征可以在不偏离本文所述的本发明组合物和方法的范围或精神的情况下容易地与任何其他若干个实施例的任何一个的特征分离或组合。可以按照所叙述的事件的顺序或按照逻辑上可行的任何其他顺序来进行任何叙述的方法。

[0178] 如本文所用,术语“子囊菌真菌细胞”是指真菌界中的子囊菌门中的任何生物体。子囊菌真菌细胞的实例包括但不限于盘菌亚门的丝状真菌,例如木霉属物种、曲霉属物种(*Aspergillus* spp.)、毁丝霉属物种(*Myceliophthora* spp.)和青霉属物种(*Penicillium* spp.)。

[0179] 如本文所用,术语“丝状真菌”是指真菌门和卵菌门的所有丝状形式。例如,丝状真菌包括但不限于枝顶孢属(*Acremonium*)、曲霉属、翘孢霉属(*Emericella*)、镰刀菌属(*Fusarium*)、腐质霉属(*Humicola*)、毛霉属(毛霉菌)、毁丝霉属、脉孢菌属(*Neurospora*)、青霉属、柱霉属(*Scytalidium*)、梭孢壳属(*Thielavia*)、弯颈霉属(*Tolypocladium*)或木霉属物种。在一些实施例中,丝状真菌可以是棘孢曲霉(*Aspergillus aculeatus*)、泡盛曲霉(*Aspergillus awamori*)、臭曲霉(*Aspergillus foetidus*)、日本曲霉(*Aspergillus japonicus*)、构巢曲霉(*Aspergillus nidulans*)、黑曲霉(*Aspergillus niger*)或米曲霉(*Aspergillus oryzae*)。

[0180] 在一些实施例中,丝状真菌是杆孢状镰孢(*Fusarium bactridioides*)、禾谷镰孢(*Fusarium cerealis*)、克地镰刀菌(*Fusarium crookwellense*)、大刀镰刀菌(*Fusarium*

culmorum)、禾谷镰刀菌 (*Fusarium graminearum*)、禾镰孢菌 (*Fusarium graminum*)、异孢镰刀菌 (*Fusarium heterosporum*)、合欢木镰孢 (*Fusarium negundi*)、尖孢镰刀菌 (*Fusarium oxysporum*)、多枝镰孢 (*Fusarium reticulatum*)、粉红镰刀菌 (*Fusarium roseum*)、接骨木镰刀菌 (*Fusarium sambucinum*)、肤色镰孢 (*Fusarium sarcochroum*)、拟分枝孢镰刀菌 (*Fusarium sporotrichioides*)、硫色镰刀菌 (*Fusarium sulphureum*)、圆镰孢 (*Fusarium torulosum*)、拟丝孢镰刀菌 (*Fusarium trichothecioides*)、或镰孢霉 (*Fusarium venenatum*)。在一些实施例中,丝状真菌是特异腐质霉 (*Humicola insolens*)、绵毛状腐质霉 (*Humicola lanuginosa*)、米黑毛霉 (*Mucor miehei*)、嗜热毁丝霉 (*Myceliophthora thermophila*)、粗糙链孢霉 (*Neurospora crassa*)、嗜热色串孢 (*Scytalidium thermophilum*)、或土生梭孢壳 (*Thielavia terrestris*)。

[0181] 在一些实施例中,丝状真菌是哈茨木霉 (*Trichoderma harzianum*)、康宁木霉 (*Trichoderma koningii*)、长枝木霉 (*Trichoderma longibrachiatum*)、里氏木霉或绿色木霉 (*Trichoderma viride*)。在一些实施例中,丝状真菌是衍生自里氏木霉菌株“Rut-C30”(其可获得自美国典型培养保藏中心(以里氏木霉ATCC保藏号56765))的里氏木霉细胞。在一些实施例中,丝状真菌是衍生自里氏木霉菌株“RL-P37”(其可获得自美国农业部北部地区研究中心的培养物保藏(作为NRRL号15709))的里氏木霉细胞。

[0182] 如本文所用,短语“一种或多种变体丝状真菌细胞”、“一种或多种变体真菌细胞”、“一种或多种变体细胞”等是指衍生自(即,获得自)属于盘菌亚门的亲本(对照)丝状真菌细胞的丝状真菌细胞。因此,如本文所定义的,“变体”丝状真菌细胞衍生自“亲本”丝状真菌细胞,其中“变体”细胞包含至少一个在“亲本”细胞中未发现的遗传修饰。例如,当将“变体丝状真菌细胞”与本公开的“亲本丝状真菌细胞”进行比较时,相比于包含至少一个遗传修饰的“变体”细胞,“亲本”细胞为遗传上未经修饰的(亲本)“对照”细胞。

[0183] 如本文所用,术语“遗传修饰”是指核酸序列的改变/变化。修饰可包括但不限于核酸序列中至少一个核苷酸的取代、缺失、插入或化学修饰。

[0184] 如本文所定义,短语“包含遗传修饰的一种或多种变体细胞”和“包含以下遗传修饰的一种或多种变体细胞,该遗传修饰增加编码Ace3-L蛋白、Ace3-EL蛋白和/或Ace3-LN蛋白的基因的表达”包括但不限于将编码Ace3-L蛋白、Ace3-EL蛋白和/或Ace3-LN蛋白的基因或ORF的至少一个拷贝引入丝状真菌细胞中。因此,包含外源地引入的编码Ace3-L蛋白、Ace3-EL蛋白和/或Ace3-LN蛋白的基因或ORF的至少一个拷贝的丝状真菌细胞是相对与亲本真菌细胞(未经修饰)包含遗传修饰的变体真菌细胞。

[0185] 在其他实施例中,针对编码本文公开的任何Ace3蛋白的内源ace3基因(即,编码Ace3-S蛋白、Ace3-SC蛋白、Ace3-L蛋白、Ace3-LC蛋白、Ace3-EL蛋白、和Ace3-LN蛋白的ace3基因)的存在,筛选本公开的亲本真菌细胞。例如,如果确定亲本真菌细胞包含编码Ace3-L蛋白、Ace3-EL蛋白、或Ace3-LN蛋白的ace3基因的内源拷贝,则可以通过遗传修饰(例如,通过用异源启动子替换ace3基因的内源启动子)产生其变体真菌细胞。同样,如果确定亲本真菌细胞包含编码Ace3-S蛋白、Ace3-SC蛋白或Ace3-LC蛋白的ace3基因的内源拷贝,则可通过遗传修饰产生其变体真菌细胞,例如通过向真菌细胞中引入编码本公开的Ace3-L蛋白、Ace3-EL蛋白和/或Ace3-LN蛋白的多核苷酸构建体,其可以进一步包括编码Ace3-S、Ace3-SC或其Ace3-LC蛋白的内源性ace3基因的遗传修饰。

[0186] 在其他实施例中,本公开的变体丝状真菌细胞将包含另外的遗传修饰。例如,在某些实施例中,此类变体丝状真菌细胞(即,包含外源地引入的编码本公开的Ace3-L蛋白、Ace3-EL和/或Ace3-LN蛋白的基因或ORF的拷贝)进一步包含遗传修饰,该遗传修饰减少编码碳分解代谢物阻遏蛋白“Cre1”或Ace1阻遏蛋白的基因的表达和/或活性。

[0187] 在其他实施例中,此类变体丝状真菌细胞(即,包含外源地引入的编码本公开的Ace3-L蛋白、Ace3-EL和/或Ace3-LN蛋白的基因或ORF的拷贝)进一步包含遗传修饰,该遗传修饰引入SEQ ID NO:25或SEQ ID NO:27中所示的木聚糖酶调节子1(Xyr1)的至少一个拷贝。

[0188] 如本文所用,“Ace3-L”蛋白形式(SEQ ID NO:6)和“Ace3-LN”蛋白形式(SEQ ID NO:14;参见,图1B)包含相同的氨基酸序列。然而,标题“Ace3-L”和“Ace3-LN”用于本公开的某些实施例中,用于与编码此类蛋白形式的某些基因进行比较,但决不意味着限制本公开。

[0189] 如本文所用,术语“宿主细胞”是指具有作为如本文所述的进入序列(即引入细胞多核苷酸序列)的宿主或表达载体的能力的丝状真菌细胞。

[0190] “异源的”核酸构建体或序列具有不是其被表达的细胞天然的或以天然形式存在的序列的一部分。关于控制序列,异源的是指在自然界中不起调节与其目前正在调节的基因的表达相同的基因的功能的控制序列(即启动子或增强子)。通常,异源的核酸序列对于它们存在的细胞或部分基因组不是内源的,并且已经通过感染、转染、转化、显微注射、电穿孔等添加到细胞中。“异源的”核酸构建体可以含有与在天然细胞中发现的控制序列/DNA编码序列组合相同或不同的控制序列/DNA编码序列组合。相似地,异源蛋白将通常是指在自然界中未发现彼此间的相同关系的两个或多个子序列(例如融合蛋白)。

[0191] 如本文所用,术语“DNA构建体”或“表达构建体”是指核酸序列,其包含至少两种DNA多核苷酸片段。DNA或表达构建体可用于将核酸序列引入真菌宿主细胞中。DNA可以在体外(例如,通过PCR)或通过任何其他合适的技术产生。在一些优选的实施例中,DNA构建体包含目的序列(例如,编码Ace3-L蛋白)。在某些实施例中,将目的多核苷酸序列有效地连接到启动子。在一些实施例中,该DNA构建体进一步包含至少一种选择性标记。在另外的实施例中,该DNA构建体包含与宿主细胞染色体同源的序列。在其他实施例中,DNA构建体包含与宿主细胞染色体非同源的序列。

[0192] 如本文所用,术语“纤维素酶(cellulase)”、“纤维素分解酶”或“纤维素酶(cellulase enzyme)”意指细菌或真菌酶,如外切葡聚糖酶、外切纤维二糖水解酶、内切葡聚糖酶和/或 β -葡萄糖苷酶。这些不同类型的纤维素酶协同作用来将纤维素及其衍生物转化为葡萄糖。例如,许多微生物制造水解纤维素的酶,包括木腐真菌木霉菌属、堆肥细菌高温单孢菌属(现在是嗜热双岐菌属(*Thermobifida*))、芽孢杆菌属和纤维单孢菌属;链霉菌属;以及真菌腐质霉属、曲霉属和镰孢菌属。由这些微生物制造的酶是以下有用于将纤维素转化成葡萄糖的三种作用类型的蛋白的混合物:内切葡聚糖酶(EG)、纤维二糖水解酶(CBH)、和 β -葡萄糖苷酶(BG)。如本文所定义,术语“内切葡聚糖酶”(EG)、“纤维二糖水解酶”(CBH)和“ β -葡萄糖苷酶”(BG)分别与它们的缩写“EG”、“CBH”和“BG”互换使用。

[0193] 如本文所用,术语“碳限制”是一种状态,其中微生物具有刚好足以产生所希望的蛋白产物,但是不足以完全满足生物体的需求的碳,例如维持生长。因此,最大量的碳朝向蛋白产生。

[0194] 如本文所用,术语“启动子”是指用于引导下游基因或其可读框(ORF)转录的核酸序列。通常启动子将适合于正在表达靶基因的宿主细胞。启动子与其他转录和翻译调控核酸序列(也称为“控制序列”)一起是表达给定的基因所必须的。通常,转录和翻译调控序列包括但不限于启动子序列、核糖体结合位点、转录起始和终止序列、翻译起始和终止序列、以及增强子或激活子序列。在某些实施例中,启动子是诱导型启动子。在其他实施例中,启动子是组成型启动子。

[0195] 如本文所用,“启动子序列”是出于表达目的而被特定丝状真菌识别的DNA序列。“启动子”被定义为引导核酸转录的一系列核酸控制序列。如本文所用,启动子包括转录起始位点附近的必需核酸序列,例如,在聚合酶II型启动子的情况下,是TATA元件。“组成型”启动子是一种在大多数环境和发育条件下具有活性的启动子。“诱导型”启动子是指在环境调节或发育调节下具有活性的启动子。

[0196] 术语“有效地连接的”启动子是指核酸表达控制序列(如启动子、或大量转录因子结合位点)和第二核酸序列之间的功能性连接,其中表达控制序列引导与第二序列相应的核酸序列的转录。因此,当核酸置于与另一个核酸序列的功能关系时,该核酸与另一个核酸序列“有效地连接”。例如,如果编码分泌性前导子(即信号肽)的DNA表达为参与多肽分泌的前蛋白,那么所述编码分泌性前导子的DNA有效地连接到编码所述多肽的DNA;如果启动子或增强子影响编码序列的转录,那么所述启动子或增强子有效地连接到所述序列;或者如果核糖体结合位点被定位以便促进翻译,那么所述核糖体结合位点有效地连接到编码序列。通常,“有效地连接”意指被连接的DNA序列是连续的,并且在分泌性前导子的情况下,是连续的并且处于阅读相中。然而,增强子不必是连续的。通过在方便的限制位点处连接来实现连接。如果这样的位点不存在,则按照常规实践使用这些合成的寡核苷酸衔接子或接头。在其他实施例中,通过无缝克隆方法完成连接,其中DNA以不依赖序列和无瘢痕的方式连接。无缝克隆通常用但不限于商业上可获得的系统进行,如Gibson组装(NEB)、NEBuilder HiFi DNA组装(NEB)、Golden Gate组装(NEB)、和GeneArt无缝克隆和组装系统(赛默飞世尔科技公司(ThermoFisher Scientific))。

[0197] 如本文所用的,术语“编码序列”是指直接确定其(编码的)蛋白产物氨基酸序列的核苷酸序列。编码序列的边界一般由通常以ATG起始密码子开始的可读框确定。编码序列通常包括DNA、cDNA和重组核苷酸序列。

[0198] 如本文定义的,术语“可读框”(下文称为“ORF”)意指包含不间断阅读框的核酸或核酸序列(无论是天然存在的、非天然存在的、或合成的),不间断阅读框由以下组成:(i)起始密码子,(ii)一系列表示氨基酸的更多密码子中的两(2)个,和(iii)终止密码子,该ORF以5'至3'方向阅读(或翻译)。

[0199] 如本文所用,术语“基因”意指涉及产生多肽链的DNA的区段,该区段可以包括或可以不包括编码区之前和之后的区域,例如,5'非翻译区(5'UTR)或“前导”序列以及3'UTR或“尾部”序列,以及个体编码区段(外显子)之间的插入序列(内含子)。基因可以编码商业上重要的工业蛋白或肽,如酶(例如蛋白酶、甘露聚糖酶、木聚糖酶、淀粉酶、葡糖淀粉酶、纤维素酶、氧化酶和脂肪酶等)。目的基因可以是天然存在的基因、突变基因或合成基因。

[0200] 如本文所用,当涉及细胞、核酸、蛋白、或载体使用时,术语“重组体”指示该细胞、核酸、蛋白或载体已经通过引入异源核酸或蛋白,或改变天然核酸或蛋白而进行修饰,或指

示该细胞是从此类修饰的细胞衍生的。因此,例如,这些重组细胞表达在细胞的天然形式(非重组)中未发现的基因或表达以其他方式异常表达的、低表达的或根本不表达的天然基因。

[0201] 术语“载体”在本文中定义为被设计用于携带待引入一种或多种细胞类型中的核酸序列的多核苷酸。载体包括克隆载体、表达载体、穿梭载体、质粒、噬菌体或病毒颗粒、DNA构建体、盒等。表达载体可包括调节序列,如启动子、信号序列、编码序列和转录终止子。

[0202] 如本文所用,“表达载体”意指包含编码序列的DNA构建体,所述编码序列与能够在合适的宿主中实现蛋白表达的合适的控制序列有效地连接。这样的控制序列可以包括影响转录的启动子,控制转录的任选的操纵子序列,编码mRNA上适合的核糖体结合位点的序列,增强子以及控制转录和翻译终止的序列。

[0203] 如本文所用,术语“分泌信号序列”表示编码多肽(即,“分泌肽”)的DNA序列,该多肽作为较大多肽的组分引导较大的多肽通过其中该较大多肽被合成的细胞的分泌途径。较大的多肽在通过分泌途径转运期间通常被切割以去除分泌肽。

[0204] 如本文所用,术语“诱导”是指响应于“诱导物”(即,诱导底物)的存在基因的转录增加,这导致以显著增加的速率在丝状真菌细胞中合成目的蛋白(下文称为“POI”)。为了测量编码POI的“目的基因”(下文称为“GOI”)或“目的ORF”的“诱导”,用候选诱导底物(诱导物)处理变体丝状真菌(宿主)细胞并与未用诱导底物(诱导物)处理的亲本丝状真菌(对照,未修饰的)细胞进行比较。因此,(未处理的)亲本(对照)细胞被指定为100%的相对蛋白活性值,其中当活性值(即,相对于对照细胞)大于100%、大于110%、更优选地150%、更优选地200%-500%(即,相对于对照高2至5倍)、或更优选高1000%-3000%时,实现在变体宿主细胞中编码POI的GOI的诱导。

[0205] 如本文所用,术语“诱导物(inducer、inducers)”、“诱导底物(inducing substrate、inducing substrates)”可互换地使用,并是指引起丝状真菌细胞产生“增加量”的多肽(例如,酶、受体、抗体等)的任何化合物,或如果不存在诱导底物时将产生的其他化合物/底物。诱导底物的实例包括但不限于槐糖、乳糖、龙胆糖和纤维素。

[0206] 如本文所用,术语“诱导饲料”是指至少包含“诱导底物”的组合物,其被喂养给丝状真菌细胞,其中此类诱导饲料诱导POI的产生。

[0207] 如本文所用,术语“分离的”或“纯化的”是指从其中去除与其天然相关的至少一种组分的核酸或氨基酸。

[0208] 如本文所定义,术语“目的蛋白”或“POI”是指希望在丝状真菌细胞中表达的多肽。此类蛋白可以是酶、底物结合蛋白、表面活性蛋白、结构蛋白等,并且可以高水平表达,并且可以用于商业化的目的。目的蛋白可以由内源基因或异源基因(即,相对于变体和/或亲本细胞)编码。目的蛋白可以在细胞内表达或作为分泌的蛋白(细胞外)表达。

[0209] 如本文所用,术语“多肽”和“蛋白”(和/或它们各自的复数形式)可互换地使用,是指包含通过肽键连接的氨基酸残基的任何长度的聚合物。本文中使用用于氨基酸残基的常规一字母或三字母代码。聚合物可以是线性的或支化的,它可以包含修饰的氨基酸,并且它可以被非氨基酸中断。该术语还涵盖天然修饰的或通过干预(例如,二硫键形成、糖基化、脂化、乙酰化、磷酸化、或任何其他操作或修饰,如与标记组分缀合)修饰的氨基酸聚合物。该定义中还包括,例如,含有一种或多种氨基酸类似物(包括例如,非天然氨基酸等)以及本领

域已知的其他修饰的多肽。

[0210] 如本文所用,功能上和/或结构上相似的蛋白被认为是“相关的蛋白”。此类蛋白可衍生自不同属和/或物种的生物体,或甚至不同纲的生物体(例如,细菌和真菌)。相关蛋白还涵盖通过一级序列分析确定的、通过二级或三级结构分析确定的、或者通过免疫交叉反应性确定的同源物。

[0211] 如本文所用,短语“基本上无活性”或相似短语意指特定活性在混合物中不可检测或以不会干扰混合物的预期目的的量存在。

[0212] 如本文所用,术语“衍生多肽”是指通过以下衍生的或可衍生的蛋白:在一个或多个N-末端和C-末端中的一个或两个添加一个或多个氨基酸、在氨基酸序列中一个或多个不同位点取代一个或多个氨基酸、在蛋白的一端或两端或者在氨基酸序列的一个或多个位点缺失一个或多个氨基酸、和/或在氨基酸序列中一个或多个位点插入一个或多个氨基酸。蛋白衍生物的制备可以通过修饰编码天然蛋白的DNA序列、将该DNA序列转化到合适的宿主中、以及表达经修饰的DNA序列以形成衍生蛋白来实现。

[0213] 相关的(和衍生)蛋白包括“变体蛋白”。变体蛋白通过在少量氨基酸残基处的取代、缺失和/或插入而不同于参考/亲本蛋白(例如,野生型蛋白)。变体与亲本蛋白之间的不同氨基酸残基的数量可以是一个或多个,例如,1、2、3、4、5、6、7、8、9、10、15、20、30、40、50、或更多个氨基酸残基。变体蛋白与参考蛋白可以共有至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、至少约91%、至少约92%、至少约93%、至少约94%、至少约95%、至少约96%、至少约97%、至少约98%、或甚至至少约99%、或更多氨基酸序列同一性。变体蛋白也可以与参考蛋白在选择的基序、结构域、表位、保守区域等方面不同。

[0214] 如本文所用,术语“同源蛋白”是指与参考蛋白具有相似活性和/或结构的蛋白。这并不旨在意味着同源物必定与进化相关。因此,该术语旨在涵盖从不同生物体获得的相同、相似、或相应(即,在结构和功能方面)的一种或多种酶。在一些实施例中,希望鉴定与参考蛋白具有类似的三级、二级和/或一级结构的同源物。在一些实施例中,同源蛋白诱导与参考蛋白相似的一种或多种免疫应答。在一些实施例中,将同源蛋白工程化以产生具有希望活性的酶。

[0215] 序列之间的同源性程度可以使用本领域已知的任何合适的方法确定(参见,例如,Smith和Waterman,1981;Needleman和Wunsch,1970;Pearson和Lipman,1988;威斯康星遗传学软件包(Wisconsin Genetics Software Package)(遗传学电脑集团(Genetics Computer Group,Madison,WI))中的如GAP、BESTFIT、FASTA、和TFASTA的程序;和Devereux等人,1984)。

[0216] 如本文所用,在至少两个核酸或多肽的上下文中,短语“基本上相似”和“基本上相同”典型地意指多核苷酸或多肽包含与参考(即,野生型)序列相比具有至少约70%同一性、至少约75%同一性、至少约80%同一性、至少约85%同一性、至少约90%同一性、至少约91%同一性、至少约92%同一性、至少约93%同一性、至少约94%同一性、至少约95%同一性、至少约96%同一性、至少约97%同一性、至少约98%同一性、或甚至至少约99%同一性、或更高同一性的序列。可以使用已知的程序(如BLAST、ALIGN和CLUSTAL)使用标准参数确定序列同一性。

[0217] 如本文所用,术语“基因”与术语“等位基因”同义,是指编码和指导蛋白或RNA表达

的核酸。丝状真菌的营养体形式通常是单倍体,因此指定基因的单拷贝(即单个等位基因)足以赋予特定的表型。

[0218] 如本文所用,术语“野生型”和“天然的”可互换使用,并是指天然发现的基因、蛋白、真菌细胞或菌株。

[0219] 如本文所用,“基因缺失”是指该基因从宿主细胞的基因组中去除。当基因包括与基因编码序列不紧邻的控制元件(例如,增强子元件)时,基因的缺失是指编码序列、以及任选地相邻的增强子元件(包括但不限于例如,启动子和/或终止子序列)的缺失。

[0220] 如本文所用,“基因的破坏”泛指任何实质上阻止宿主细胞中细胞产生功能性基因产物(例如,蛋白)的基因操作(即,突变)。示例性破坏方法包括基因的任何部分的完全或部分(包括多肽编码序列、启动子、增强子或另外的调节元件)缺失、或其诱变,其中诱变涵盖取代、插入、缺失、倒位、及其组合和变化,任何这些突变基本上阻止功能基因产物的产生。也可以使用RNAi、反义、CRISPR/Cas9或任何其他消除、减少或减轻基因表达的方法破坏基因。

[0221] 如本文所用,短语“包含‘遗传修饰’的变体[宿主]细胞,该‘遗传修饰’增加编码本公开的Ace3-L蛋白、Ace3-EL和/或Ace3-LN蛋白的基因的表达”包含向宿主细胞中引入(例如,经由转化)质粒或染色体整合盒,该质粒或染色体整合盒包含编码此类Ace3-L蛋白、Ace3-EL和/或Ace3-LN蛋白的ace3基因形式(或其ORF)。在某些其他实施例中,如当亲本真菌细胞天然地包含编码Ace3-L蛋白、Ace3-EL和/或Ace3-LN蛋白的内源基因形式时,短语“包含‘遗传修饰’的变体[宿主]细胞,该遗传修饰增加编码Ace3-L蛋白、Ace3-EL和/或Ace3-LN蛋白的基因的表达”包括用异源启动子替换天然/野生型ace3基因启动子。

[0222] 如本文所用,“需氧发酵”是指在氧存在下的生长。

[0223] 如本文所用,术语“细胞培养液”统称为液体/浸没培养中的培养基和细胞。

[0224] 如本文所用,术语“细胞团”是指存在于液体/浸没培养中的细胞组分(包括完整和裂解的细胞)。细胞质量可以干重或湿重表示。

[0225] 如本文所用,“功能性多肽/蛋白”是具有活性(如酶活性、结合活性、表面活性性质等)的蛋白,并且其未被诱变、截短或以其他方式修饰废除或减少该活性。如所指出的,功能性多肽可以是热稳定的或不耐热的。

[0226] 如本文所用,“功能性基因”是能够被细胞组分用于产生活性基因产物(通常是蛋白)的基因。功能性基因是破坏的基因的对立体,破坏的基因被修饰使得它们不能被细胞组分用于产生活性基因产物,或者具有降低的被细胞组分用于产生活性基因产物的能力。

[0227] 如本文所用,“目的蛋白”是期望在丝状真菌细胞的浸没培养中产生的蛋白。通常,目的蛋白对于工业、药物、动物健康、以及食品和饮料的用途具有商业重要性,使得它们需要大量生产。目的蛋白与由丝状真菌细胞表达的无数其他蛋白不同,其他蛋白通常不作为产品受关注,主要被认为是背景蛋白污染物。

[0228] 如本文所用,如果与亲本细胞相比,由变体细胞产生的蛋白的量增加至少5%、增加至少10%、增加至少15%、或更多,则“变体真菌宿主细胞”比“亲本真菌细胞”产生“实质上更多的蛋白/每单位量生物质”,其中蛋白的量被标准化为测量蛋白产生的细胞的生物质的总量,其中生物质可以以湿重(例如,细胞沉淀)或干重的形式表示。

[0229] III. 纤维素酶表达激活子3(ace3)

[0230] 最近,研究了来自里氏木霉培养物的转录概况(其中“诱导”纤维素酶/半纤维素酶产生(即,通过添加不同的诱导组合物;例如,镰刀菌,乳糖)分析数据(Hakkinen等人,2014)以鉴定纤维素酶和半纤维素酶基因表达的推定“调节子”,并鉴定编码调节蛋白的候选基因。Hakkinen等人(2014)将该候选基因(参见Hakkinen等人,图2和表2)鉴定为基因ID号77513(其中基因ID编号与里氏木霉数据库2.0中一样),并将该候选基因命名为“纤维素酶表达激活子3”(下文称“ace3”),并将编码的蛋白(即,候选转录因子)命名为“Ace3”。更特别地,Hakkinen等人(2014)研究使用基于公众可获得的里氏木霉菌株QM6a的基因组序列(参见,genome.jgi.doe.gov/Trire2/Trire2.home.html,其中QM6a预测的注释(基因ID 77513)由两个外显子和一个内含子组成(例如,参见图1))的预测的ace30RF(SEQ ID NO:2)。

[0231] 如本文所述并在下文实例部分中进一步所示,当相对于基于里氏木霉“RUT-C30菌株”注释的ace30RF,比较和评价Hakkinen等人,2014中描述的克隆的ace30RF时(即,基于Ace3的里氏木霉“QM6a菌株”注释),本公开的申请人发现了令人惊讶和意想不到的结果。例如,如以下实例1中所示,SEQ ID NO:1的里氏木霉“QM6a菌株”ace3基因(以及SEQ ID NO:2的ORF)编码SEQ ID NO:3的较短Ace3蛋白(本文称为“Ace3-S”) (相对于SEQ ID NO:4的里氏木霉“RUT-C30菌株”ace3基因(或SEQ ID NO:5的ORF),其编码SEQ ID NO:6的较长Ace3蛋白(本文称为“Ace3-L”))。

[0232] 相反,从公众可获得的里氏木霉菌株Rut-C30的基因组序列(参见,genome.jgi.doe.gov/TrireRUTC30_1/TrireRUTC30_1.home.html)(基因ID 98455)预测的ace30RF包含更长的蛋白序列(即,相对于来自里氏木霉QM6a的Ace3-S),该更长的蛋白序列包含三个外显子和两个内含子(图1A)。更特别地,由“RUT-C30”模型预测的起始密码子位于“QM6a”模型中的起始密码子的上游,并且在C-末端存在无义突变(Poggi-Parodi等人,2014),这产生更长的N-末端序列和更短的C-末端序列(例如参见,1B)。

[0233] 同样地,如下文实例6中所述,ace3基因编码区的5'端的位置不明显,因此,申请人进一步评估了如本文所述的ace3基因的5'末端。如上所述,即使DNA序列相同,联合基因组研究所(Joint Genome Institute)(JGI)的DNA序列的注释在突变菌株Rut-C30与野生型菌株QM6a之间也不同。在QM6a情况下,编码区的5'末端被建议在外显子3的上游和并且在内含子2内(如图11所示)。在Rut-C30情况下,编码区的5'末端在外显子2内(图11)。

[0234] 对基因组DNA序列和另外的cDNA序列的进一步分析表明可能存在“外显子1”和“内含子1”(如图11所示)。此外,Rut-C30中ace3编码区的3'末端包含产生过早终止密码子的突变(相对于野生型分离株QM6a的序列)(图11)。因此,如实例6中所述,申请人检查了如图12所示的这些不同可能形式的ace3基因的过表达的影响。

[0235] 此外,如下文实例中所述,通过用名为“pYL1”、“pYL2”、“pYL3”和“pYL4”的四种不同的ace3表达载体之一转化里氏木霉细胞,申请人构建(实例1)并测试(实例2-4)编码Ace3-S蛋白(SEQ ID NO:3)和Ace3-L蛋白(SEQ ID NO:6)的基因(参见,图2A-2D质粒图谱)。更具体地(实例1),这些表达载体含有载体骨架,其具有用于在大肠杆菌中复制和选择的细菌ColE1ori和AmpR基因。此外,表达载体(参见,图2A-2D)包含里氏木霉pyr2选择标记和与ace30RF编码序列(ace3-L或ace3-S)有效地连接的异源里氏木霉启动子序列(即,hxk1或pki1的启动子)、及其天然终止子。

[0236] 随后,在缓释微量滴定板(srMTP)(实例2)、摇瓶(实例3)和小规模发酵(实例4)中,在“诱导”和“非诱导”条件下测试/筛选稳定的里氏木霉转化体(即,变体宿主细胞A4-7、B2-1、C2-28和D3-1),并且收获培养物上清液并通过聚丙烯酰胺凝胶电泳分析(参见,图3-图5)。如这些实例中所示,所有测试的宿主细胞(即,亲本、变体A4-7、变体B2-1、变体C2-28和变体D3-1)在诱导底物(即,槐糖或乳糖)存在下分泌大量蛋白。相反,令人惊讶地观察到,在不存在诱导底物(即,槐糖或乳糖)的情况下,其中“葡萄糖”是唯一的碳源,只有表达ace3-L ORF(即,变体A4-7和C2-28)的变体宿主细胞能够产生分泌蛋白,而亲本(对照)细胞和表达ace3-S ORF(即,变体B2-1和D3-1)的变体宿主细胞不产生任何可检测的分泌蛋白。

[0237] 在其他实施例中,本公开进一步阐述使用十三(13)种不同的启动子(以驱动ace3-L(实例5)表达构建体的表达)在“非诱导”条件下增强的蛋白产生。例如,测试的十三种启动子包括(i)甲酰胺酶基因(rev3;蛋白ID 103041)启动子(SEQ ID NO:15)、(ii)β-木糖苷酶基因(bx1;蛋白ID 121127)启动子(SEQ ID NO:16)、(iii)转酮醇酶基因(tkl1;蛋白ID 2211)启动子(SEQ ID NO:17)、(iv)功能未知的基因(蛋白ID104295)启动子(SEQ ID NO:18)、(v)氧化还原酶基因(dld1;蛋白ID5345)启动子(SEQ ID NO:19)、(vi)木聚糖酶IV基因(xyn4;蛋白ID111849)启动子(SEQ ID NO:20)、(vii)α-葡糖醛酸糖苷酶基因(蛋白ID72526)启动子(SEQ ID NO:21)、(viii)乙酰木聚糖酯酶基因1(axe1;蛋白ID 73632)启动子(SEQ ID NO:22)、(ix)己糖激酶基因(hxk1;蛋白ID 73665)启动子(SEQ ID NO:23)、(x)线粒体载体蛋白基因(dic1;蛋白ID 47930)启动子(SEQ ID NO:24)、(xi)寡肽转运蛋白基因(opt;蛋白ID 44278)启动子(SEQ ID NO:25)、(xii)半乳糖激酶基因(gut1;蛋白ID 58356)启动子(SEQ ID NO:26)、和(xiii)丙酮酸激酶基因(pki1;蛋白ID 78439)启动子(SEQ ID NO:27)。如表3所示,亲本里氏木霉细胞仅在槐糖诱导物存在下产生分泌蛋白。相反,在诱导和非诱导条件下,包含并表达由十三种不同启动子中的任何一种驱动的Ace3-L的变体(子代)里氏木霉细胞均产生相似量的分泌蛋白。还在实例5中描述,在摇瓶实验和小规模发酵中进一步测试里氏木霉亲本菌株及其转化体。如图8所示,亲本(对照)里氏木霉细胞仅在槐糖(“Sop”)诱导物的存在下产生分泌的蛋白,而子代菌株LT83在诱导(“Sop”)和非诱导(“Glu”)条件下均产生相似量的分泌蛋白。同样,如图9所示,亲本(对照)里氏木霉菌株仅在槐糖诱导物(“Sop”)存在下产生分泌蛋白,而子代菌株LT83在诱导(“Sop”)和非诱导(“Glu”)条件下均产生相似量的蛋白。

[0238] 本公开的实例6描述了ace3基因的不同可能形式的过表达的效果的实验研究(例如,参见上文讨论的突变菌株Rut-C30/野生型菌株QM6a基因组序列注释,图11和图12)。因此,图12中描绘的不同形式的ace3在里氏木霉中过表达,其中里氏木霉ace3基因的过表达载体被设计成能够在里氏木霉中的葡糖淀粉酶基因座(gla1)上靶向整合ace3。因此,表5中呈现的构建体的不同之处在于具有不同形式的ace3基因。同样地,表7中的菌株在具有2%乳糖或2%葡萄糖作为碳源的液体培养基的24孔微量滴定板中生长,其中从培养上清液中测量总分泌蛋白的量。在两种培养基(即,2%乳糖或2%葡萄糖)中,ace3-L、ace3-EL和ace3-LN形式(即,具有RutC-30C-末端突变)的过表达提高总蛋白产量(表8)。在以乳糖为碳源的培养基中,所有形式的ace3基因的过表达在一些程度上提高总蛋白的产生,但在过表达ace3基因的ace3-L、ace3-EL和ace3-LN形式的菌株中,提高水平最高(表8)。因此,很明显,当过表达ace3基因的ace3-L、ace3-EL和ace3-LN形式时,在“非诱导条件”下(即,当将葡

葡萄糖用作碳源时)观察到高水平的分泌蛋白。

[0239] 本公开的实例7描述了启动子替换构建体(参见,图6),该启动子替换构建体通过将包含在ace3基因座上的天然启动子上游5'区的DNA片段、loxP-侧翼的潮霉素B抗性选择性标记盒、和包含有效融合(连接)至ace3可读框的5'末端的目的启动子的片段融合制成。例如,在某些实施例中,启动子替换构建体用于替换在具有可替代的启动子的里氏木霉细胞中的内源性ace3基因启动子。

[0240] 本公开的实例8描述了用木质纤维素目的基因启动子替换内源性非木质纤维素目的基因启动子。例如,里氏木霉葡萄糖淀粉酶表达构建体从DNA多核苷酸片段组装而成,其中编码里氏木霉葡萄糖淀粉酶的ORF序列有效地连接到5'(上游)里氏木霉cbh1启动子,并有效地连接到3'(下游)里氏木霉cbh1终止子,该构建体进一步包含作为选择性标记的里氏木霉pyr2基因。用葡萄糖淀粉酶表达构建体转化变体(子代)里氏木霉细胞(即,包含增加编码Ace3-L蛋白的基因表达的遗传修饰),并选择转化体并在具有葡萄糖作为碳源的液体培养基中(即,没有诱导底物,如槐糖或乳糖)培养,以鉴定能够在培养期间分泌里氏木霉葡萄糖淀粉酶的那些转化体。如图10所示,亲本里氏木霉细胞在具有葡萄糖/槐糖(诱导条件)的限定培养基中产生1,029μg/mL葡萄糖淀粉酶,并且在具有葡萄糖的限定培养基(非诱导条件)中仅产生38μg/mL葡萄糖淀粉酶,而经修饰的(子代)菌株“LT88”(包含由dic1启动子驱动的ace3-L)在“诱导”(“Sop”)条件下产生3倍更高的葡萄糖淀粉酶(即,相对于亲本(对照)菌株),并在“非诱导”(“Glu”)条件下产生2.5倍更高的葡萄糖淀粉酶(即,相对于亲本(对照)菌株)。因此,这些结果表明,包含Ace3-L ORF的修饰的(子代)细胞不仅在不存在诱导物的情况下产生细胞外蛋白,而且在此类诱导条件下这些变体细胞比亲本(对照)里氏木霉细胞产生更多的总蛋白。

[0241] 实例9描述了用木质纤维素目的基因启动子替换天然相关联的异源目的基因启动子。例如,编码布丘氏菌物种(*Buttiauxella* sp.)植酸酶(即,异源GOI)的ORF在5'末端有效地连接到里氏木霉cbh1启动子,并在3'末端有效地连接到里氏木霉cbh1终止子,其中DNA构建体进一步包含选择标记。用植酸酶表达构建体转化变体里氏木霉细胞(即,包含增加编码Ace3-L蛋白的基因表达的遗传修饰),选择转化体并在具有葡萄糖作为碳源的液体培养基中(即,没有诱导底物,如槐糖或乳糖)培养,以鉴定能够在培养期间分泌里氏木霉植酸酶的那些转化体。

[0242] 本公开的实例10描述了天然ace3启动子替换载体的构建,所述载体含有在里氏木霉pk1启动子下表达的酿脓链球菌(*Streptococcus pyogenes*)cas9基因和在U6启动子下表达的指导RNA。例如,将cas9介导的ace3启动子替换载体(pCHL760和pCHL761)转化到里氏木霉亲本细胞中,并且为了测试ace3启动子替换菌株的功能,在摇瓶中在50ml浸没培养中,在存在和不存在诱导物底物(槐糖)的情况下培养细胞。如在SDS-PAGE上所见,与葡萄糖/槐糖(诱导)相比,亲本细胞(图23, ID 1275.8.1)在具有葡萄糖(非诱导)的限定培养基中产生更少的分泌蛋白。相反,转化体2218、2219、2220、2222和2223在诱导和非诱导条件下产生相似量的分泌蛋白,这证明含有hxl1或dic1启动子(即,替代ace3基因座处的天然ace3启动子)的变体细胞在不存在诱导物的情况下产生细胞外蛋白。

[0243] 因此,如本文预期和描述的,本公开的某些方面针对一种或多种内源丝状真菌木质纤维素降解酶(即,纤维素分解酶,例如,纤维二糖水解酶、木聚糖酶、内切葡聚糖酶等)的

产生。更具体地,在完全没有诱导底物的情况下,本公开的某些实施例针对在本公开的变体宿主细胞(即,包含增加Ace3-L蛋白、Ace3-EL和/或Ace3-LN蛋白的表达的遗传修饰的变体宿主细胞)中以产生此类内源酶。本发明的变体宿主细胞、组合物和方法对于显著降低产生上述纤维素分解酶的成本/费用特别有用,特别是由于本公开的此类变体宿主细胞不需要诱导底物来产生此类纤维素分解酶(即,与在诱导底物存在下仅产生此类纤维素分解酶的亲本细胞相反)。

[0244] 例如,在某些实施例中,本公开针对变体真菌宿主细胞,其在不存在诱导底物的情况下表达/产生一种或多种内源目的蛋白、和/或在不存在诱导底物的情况下表达/产生一种或多种异源目的蛋白。因此,在某些实施例中,将本公开的变体真菌宿主细胞(即,包含增加Ace3-L蛋白、Ace3-EL和/或Ace3-LN的表达的遗传修饰)进一步修饰以表达内源、非木质纤维素目的蛋白和/或异源目的蛋白。例如,在某些实施例中,修饰变体真菌宿主细胞中编码内源、非木质纤维素目的蛋白的基因。因此,在某些实施例中,与编码内源、非木质纤维素目的蛋白的基因(或ORF)天然相关联的启动子被来自编码木质纤维素蛋白的丝状真菌基因的启动子(例如,5'-木质纤维素基因启动子,其有效地连接到编码非木质纤维素目的蛋白的内源基因)替换。同样地,在某些实施例中,将本公开的变体真菌宿主细胞(即,包含增加Ace3-L蛋白、Ace3-EL和/或Ace3-LN的表达的遗传修饰)进行修饰以表达异源目的蛋白。因此,在某些其他实施例中,与编码异源目的蛋白的基因天然地相关联的启动子被来自编码木质纤维素蛋白的丝状真菌基因的启动子(例如,5'-木质纤维素基因启动子,其有效地连接到编码异源目的蛋白的异源基因)替换。

[0245] 因此,在某些实施例中,本公开针对衍生自亲本丝状真菌细胞的变体丝状真菌细胞,其中相对于亲本细胞,该变体细胞包含增加编码Ace-L蛋白的基因的表达的遗传修饰,其中编码的Ace3-L蛋白包含与SEQ ID N0:6的Ace3-L蛋白约90%序列同一性。例如,在某些实施例中,包含与SEQ ID N0:6约90%序列同一性的编码的Ace3蛋白包含“Lys-Ala-Ser-Asp”作为最后四个C-末端氨基酸。在其他实施例中,包含与SEQ ID N0:6的约90%序列同一性的编码的Ace3蛋白进一步包含有效地连接的并且在SEQ ID N0:6之前的SEQ ID N0:98的N-末端氨基酸片段。在另外的实施例中,Ace-3蛋白包含与SEQ ID N0:12约90%序列同一性。

[0246] 在某些其他实施例中,本公开的多核苷酸包含核苷酸序列,该核苷酸序列包含与SEQ ID N0:4、SEQ ID N0:11或SEQ ID N0:13约90%序列同一性。在其他实施例中,本公开的多核苷酸包含核苷酸序列,该核苷酸序列包含与SEQ ID N0:5、SEQ ID N0:101或SEQ ID N0:102约90%序列同一性。

[0247] IV.丝状真菌宿主细胞

[0248] 在本公开的某些实施例中,提供了变体丝状真菌细胞(即,衍生自亲本丝状真菌细胞),其包含增加编码Ace3-L多肽的基因或ORF的表达的遗传修饰。更特别地,在某些实施例中,变体丝状真菌细胞(即,相对于亲本(对照)细胞)包含遗传修饰,该遗传修饰增加编码SEQ ID N0:6的Ace3-L蛋白的基因(或ORF)的表达。在优选的实施例中,在不存在诱导底物的情况下,包含遗传修饰(其增加编码SEQ ID N0:6的Ace3-L蛋白的基因(或ORF)的表达)的此类变体真菌细胞能够产生至少一种目的内源蛋白。在其他实施例中,在不存在诱导底物的情况下,包含遗传修饰(其增加编码SEQ ID N0:6的Ace3-L蛋白的基因(或ORF)的表达)的

此类变体真菌细胞能够产生至少一种目的异源蛋白。

[0249] 因此,在某些实施例中,用于在本公开内容中操作和使用的丝状真菌细胞包括来自子囊菌门、盘菌亚门的丝状真菌,特别是具有营养菌丝状态的真菌。此类生物体包括用于产生商业上重要的工业和药物蛋白的丝状真菌细胞,所述真菌细胞包括但不限于木霉属物种、曲霉属物种、镰孢属物种(*Fusarium spp.*)、丝孢菌属物种(*Scedosporium spp.*)、青霉菌物种、金孢子菌物种(*Chrysosporium spp.*)、头孢霉属物种(*Cephalosporium spp.*)、踝节菌属物种(*Talaromyces spp.*)、*Geosmithia spp.*、毁丝霉属物种、和脉孢菌属物种(*Neurospora spp.*)。

[0250] 特别的丝状真菌包括但不限于里氏木霉(先前分类为长枝木霉和红褐肉座菌)、黑曲霉、烟曲霉(*Aspergillus fumigatus*)、分解亚甲基丁二酸曲霉(*Aspergillus itaconicus*)、米曲霉、构巢曲霉、土曲霉(*Aspergillus terreus*)、酱油曲霉(*Aspergillus sojae*)、日本曲霉、赛多孢(*Scedosporium prolificans*)、粗糙链孢霉、绳状青霉(*Penicillium funiculosum*)、产黄青霉(*Penicillium chrysogenum*)、埃默森篮状菌(*Talaromyces (Geosmithia) emersonii*)、镰孢霉、嗜热毁丝霉和卢克诺文思金孢子菌(*Chrysosporium lucknowense*)。

[0251] V. 重组核酸和分子生物学

[0252] 在某些实施例中,本公开针对包含遗传修饰的变体丝状真菌宿主细胞,该遗传修饰增加编码Ace3-L、Ace3-EL或Ace3-LN蛋白的基因或ORF的表达。如上所述,此类变体宿主细胞在不存在诱导底物(即,与未经修饰的亲本(对照)细胞相反)的情况下,能够产生一种或多种目的蛋白。

[0253] 因此,在某些实施例中,本公开针对重组核酸,该重组核酸包含编码Ace3-L、Ace3-EL或Ace3-LN蛋白的基因或ORF。在某些实施例中,重组核酸包含用于在丝状真菌宿主细胞中产生Ace3-L、Ace3-EL或Ace3-LN蛋白的多核苷酸表达盒。在其他实施例中,多核苷酸表达盒包含在表达载体内。在某些实施例中,表达载体是质粒。在其他实施例中,重组核酸、其多核苷酸表达盒或表达载体包含核苷酸序列,该核苷酸序列包含与SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:11、SEQ ID NO:101、SEQ ID NO:13或SEQ ID NO:102至少85%序列同一性。在另外的实施例中,重组核酸、其多核苷酸表达盒或表达载体包含核苷酸序列,该核苷酸序列编码包含与SEQ ID NO:6约90%序列同一性的Ace3-L蛋白。

[0254] 在某些其他实施例中,重组核酸(或其多核苷酸表达盒或其表达载体)进一步包含一种或多种选择性标记。用于丝状真菌中的选择性标记包括但不限于 α ls1、 α mdS、 α ygR、 α yr2、 α yr4、 α yrG、 α sucA、博来霉素抗性标记、杀稻瘟菌素抗性标记、吡啶硫胺抗性标记、氯嘧磺隆乙酯抗性标记、新霉素抗性标记、腺嘌呤途径基因、色氨酸途径基因、和胸苷激酶标记等。在特别的实施例中,选择性标记是 α yr2,其组合物和使用方法通常在PCT公开号W0 2011/153449中示出。因此,在某些实施例中,编码本公开的Ace3蛋白的多核苷酸构建体包含编码与其有效地连接的选择性标记的核酸序列。

[0255] 在另外的实施例中,重组核酸、其多核苷酸构建体、多核苷酸表达盒或表达载体包含驱动编码Ace3-L、Ace3-EL或Ace3-LN蛋白的基因(或ORF)的表达的异源启动子。更特别地,在某些实施例中,异源启动子是组成型或诱导型启动子。在特别的实施例中,异源启动子选自由以下组成的组:rev3启动子、bx1启动子、tk11启动子、PID104295启动子、d1d1启动

子、xyn4启动子、PID72526启动子、axe1启动子、hxk1启动子、dic1启动子、opt启动子、gut1启动子、和pki1启动子。不希望受特定理论或作用机理的束缚,本文考虑启动子如rev3、bx1、tk1、PID104295、d1d1、xyn4、PID72526、axe1、hxk1、dic1、opt、gut1和pki1,其在葡萄糖限制条件下(即,相对于过量的葡萄糖浓度)产生更高的表达水平,在本公开中具有特别的用途。因此,在某些实施例中,重组核酸(或其多核苷酸构建体、多核苷酸表达盒或表达载体)包含启动子,该启动子在编码Ace3蛋白的核酸序列5'端的并有效地连接到编码Ace3蛋白的核酸序列。

[0256] 在另外的实施例中,重组核酸(或其多核苷酸构建体、多核苷酸表达盒或表达载体)进一步包含编码天然ace3终止子序列的核酸序列。因此,在某些实施例中,重组核酸(或其多核苷酸构建体、多核苷酸表达盒或表达载体)包含启动子(其在编码Ace3蛋白的核酸序列的5'端的并有效地连接到编码Ace3蛋白的核酸序列)、和天然ace3终止子序列(其在编码Ace3蛋白的核酸序列3'端的并有效地连接到编码Ace3蛋白的核酸序列)(例如,5'-Pro-ORF-Term-3',其中“Pro”是组成型启动子,“ORF”编码Ace3,并且“Term”是天然ace3终止子序列)。

[0257] 因此,在某些实施例中,转化丝状真菌和培养真菌的标准技术(本领域技术人员熟知的)用于转化本公开的真菌宿主细胞。因此,将DNA构建体或载体引入真菌宿主细胞中包括如下技术如:转化、电穿孔、核显微注射、转导、转染(例如脂质介导的转染和DEAE-糊精介导的转染)、用磷酸钙DNA沉淀孵育、用DNA包被的微粒高速轰击、基因枪或生物射弹转化、原生质体融合等。一般的转化技术是本领域已知的(参见,例如,Ausubel等人,1987,Sambrook等人,2001和2012,以及Campbell等人,1989)。异源蛋白在木霉属中的表达描述于例如美国专利号6,022,725;6,268,328;Harkki等人,1991以及Harkki等人,1989中。对于曲霉属菌株的转化,还参考了Cao等人(2000)。

[0258] 通常,使用已经进行了渗透性处理的原生质体或细胞转化木霉属的物种,通常以 10^5 至 10^7 /mL、特别是 2×10^6 /mL的密度进行。将100μL体积的在合适的溶液(例如1.2M山梨糖醇和50mM CaCl₂)中的这些原生质体或细胞与所希望的DNA混合。通常,向摄取溶液中添加高浓度的聚乙二醇(PEG)。也可以将添加剂例如二甲亚砜、肝素、亚精胺、氯化钾等添加到摄取溶液中以促进转化。类似的程序可用于其他真菌宿主细胞。参见,例如美国专利号6,022,725和6,268,328,其两者都通过引用并入本文。

[0259] 在某些实施例中,本公开针对一种或多种目的蛋白的表达和产生,所述目的蛋白对于丝状真菌宿主细胞是内源的(即,由包含遗传修饰的本公开的变体真菌宿主细胞产生这些内源蛋白,该遗传修饰增加Ace3-L的表达)。在其他实施例中,本公开针对表达和产生一种或多种目的蛋白,所述目的蛋白对于丝状真菌宿主细胞是异源的。因此,本公开总体上依赖于重组遗传学领域中的常规技术。公开了在本公开中使用的一般方法的基础文本包括Sambrook等人,(第2版,1989);Kriegler(1990)和Ausubel等人,(1994)。

[0260] 因此,在某些实施例中,将编码目的蛋白的异源基因或ORF引入丝状真菌(宿主)细胞中。在某些实施例中,通常将异源基因或ORF克隆到中间载体中,然后转化到丝状真菌(宿主)细胞中用于复制和/或表达。这些中间载体可以是原核载体,例如像质粒或穿梭载体。在某些实施例中,异源基因或ORF的表达处于其天然启动子的控制之下。在其他实施例中,将异源基因或ORF的表达置于异源启动子的控制下,该异源启动子可以是异源组成型启动子

或异源诱导型启动子。

[0261] 本领域技术人员清楚的是可以通过替换、取代、添加或消除一个或多个核苷酸而不改变启动子的功能来修饰天然(自然)的启动子。本发明的实践涵盖但不受对启动子的这些改变的限制。

[0262] 表达载体/构建体通常含有转录单元或表达盒,该转录单元或表达盒含有表达异源序列所需的所有附加元件。例如,典型的表达盒含有与编码目的蛋白的异源核酸序列有效地连接的5'启动子,并且还进一步包含转录物的有效聚腺苷酸化、核糖体结合位点和翻译终止所需的序列信号。该盒的附加元件可以包括增强子,并且如果使用基因组DNA作为结构基因,可以包括具有功能性剪接供体和受体位点的内含子。

[0263] 除了启动子序列之外,该表达盒还可以含有结构基因的录终止区下游转以提供有效的终止。该终止区可以从与启动子序列相同的基因获得或者可以从不同基因获得。尽管任何真菌终止子在本发明中可能是有功能的,但优选的终止子包括:来自木霉属cbhI基因的终止子,来自构巢曲霉trpC基因(Yelton等人,1984;Mullaney等人,1985)、泡盛曲霉或黑曲霉葡糖淀粉酶基因(Nunberg等人,1984;Boel等人,1984)、和/或米黑毛霉羧基蛋白酶基因(EPO公开号0215594)的终止子。

[0264] 用来将遗传信息运送到细胞中的具体表达载体不是特别关键。可以使用用于在真核细胞或原核细胞中表达的常规载体的任一种。标准的细菌表达载体包括噬菌体λ和M13,以及质粒如基于pBR322的质粒,pSKF,pET23D和融合表达系统如MBP、GST和LacZ。也可以将表位标签,例如c-myc,添加到重组蛋白中以提供便利的分离方法。

[0265] 可以包括在表达载体中的元件还可以是复制子,编码允许选择携带重组质粒的细菌的抗生素抗性的基因,或在质粒的非必需区中的允许插入异源序列的独特限制性位点。所选择的特定抗生素抗性基因也不是决定性的,因为本领域已知的许多抗性基因中的任一种都可以是合适的。优选地选择原核序列,使得它们不干扰DNA在里氏木霉中的复制或整合。

[0266] 本发明的转化的方法可以导致转化载体的全部或一部分稳定整合到丝状真菌的基因组中。然而,还考虑了导致维持自主复制的染色体外转化载体的转化。

[0267] 可以使用许多标准转染方法以产生表达大量异源蛋白质的里氏木霉细胞系。一些用于将DNA构建体引入木霉属的产纤维素酶菌株的公开的方法包括:Lorito,Hayes,DiPietro和Harman,1993,Curr.Genet.[现代遗传学]24:349-356;Goldman, VanMontagu和Herrera-Estrella,1990,Curr.Genet.[现代遗传学]17:169-174;Penttila,Nevalainen,Ratto,Salminen和Knowles,1987,Gene[基因]6:155-164,对于曲霉属Yelton、Hamer和Timberlake,1984,Proc.Natl.Acad.Sci.USA[美国国家科学院院刊]81:1470-1474,对于镰孢菌属Bajar,Podila和Kolattukudy,1991,Proc.Natl.Acad.Sci.USA[美国国家科学院院刊]88:8202-8212,对于链霉菌属Hopwood等人,1985,The John Innes Foundation[约翰英纳斯基金会],诺威奇,英国,以及对于芽孢杆菌属Brigidi,DeRossi,Bertarini,Riccardi和Matteuzzi,1990,FEMS Microbiol.Lett.[FEMS微生物学快报]55:135-138)。

[0268] 用于将外源核苷酸序列引入宿主细胞的任何已知程序都可以使用。这些包括使用磷酸钙转染、聚凝胺、原生质体融合、电穿孔、基因枪法、脂质体、显微注射、原生质载体、病毒载体,以及用于将克隆的基因组DNA、cDNA、合成DNA、或其他外源遗传材料引入宿主细胞

的任何其他已知方法(参见例如Sambrook等人,同上)。还使用的是农杆菌介导的转染方法,如在美国专利号6,255,115中描述的转染方法。只需要使用能够成功地将至少一个基因引入能够表达异源基因的宿主细胞中的具体遗传工程化程序。

[0269] 将表达载体引入细胞后,在有利于表达在纤维素酶基因启动子序列控制下的基因的条件下培养转染的细胞。大批转化细胞可以如本文所述进行培养。最后,使用标准技术从培养物回收产物。

[0270] 因此,本发明在此提供了表达在纤维素酶基因启动子序列控制下的所希望多肽的表达和增强的分泌,所述纤维素酶基因启动子序列包括天然存在的纤维素酶基因、融合DNA序列和各种异源构建体。本发明还提供了用于表达和分泌高水平的此类所希望物的方法。

[0271] VI. 目的蛋白

[0272] 如上所述,本公开的某些实施例针对变体丝状真菌细胞(衍生自亲本丝状真菌细胞),其中所述变体细胞包含增加编码Ace3-L、Ace3-EL或Ace3-LN蛋白质的基因的表达(相对于亲本细胞)的遗传修饰,其中编码的Ace3-L、Ace3-EL或Ace3-LN蛋白包含与SEQ ID NO:6的Ace3-L蛋白约90%序列同一性,其中所述变体细胞在不存在诱导底物的情况下表达至少一种目的蛋白(POI)。

[0273] 本公开的某些实施例特别有用于在不存在诱导底物的情况下增强蛋白(即,目的蛋白)的细胞内和/或细胞外产生。目的蛋白可以是内源蛋白(即,在宿主细胞中是内源的)或异源蛋白(即,在宿主细胞中不是天然的)。可根据本公开产生的蛋白包括但不限于激素、酶、生长因子、细胞因子、抗体等。

[0274] 例如,目的蛋白可包括但不限于半纤维素酶、过氧化物酶、蛋白酶、纤维素酶、木聚糖酶、脂肪酶、磷脂酶、酯酶、角质酶、果胶酶、角蛋白酶、还原酶、氧化酶、酚氧化酶、脂加氧酶、木质素酶、支链淀粉酶、鞣酸酶、戊聚糖酶、甘露聚糖酶、 β -葡聚糖酶、透明质酸酶、软骨素酶、漆酶、淀粉酶、葡糖淀粉酶、乙酰酯酶、氨基肽酶、淀粉酶、阿拉伯糖酶、阿拉伯糖苷酶、阿拉伯呋喃糖苷酶、羧肽酶、过氧化氢酶、脱氧核糖核酸酶、差向异构酶、 α -半乳糖苷酶、 β -半乳糖苷酶、 α -葡聚糖酶、葡聚糖裂解酶、内切- β -葡聚糖酶、葡萄糖氧化酶、葡糖醛酸糖苷酶、转化酶、异构酶等。

[0275] 在某些实施例中,目的蛋白包括但不限于以下中公开的酶:PCT申请公开号W0 03/027306、WO 200352118、WO 200352054、WO 200352057、WO 200352055、WO 200352056、WO 200416760、WO 9210581、WO 200448592、WO 200443980、WO 200528636、WO 200501065、WO 2005/001036、WO 2005/093050、WO 200593073、WO 200674005、WO 2009/149202、WO 2011/038019、WO 2010/141779、WO 2011/063308、WO 2012/125951、WO 2012/125925、WO 2012125937、WO/2011/153276、WO 2014/093275、WO 2014/070837、WO 2014/070841、WO 2014/070844、WO 2014/093281、WO 2014/093282、WO 2014/093287、WO 2014/093294、WO 2015/084596和WO 2016/069541。

[0276] 用于产生蛋白的最佳条件将随宿主细胞的选择以及待表达的一种或多种蛋白的选择而变化。本领域技术人员通过常规实验和/或优化可以容易地确定这样的条件。

[0277] 目的蛋白可以在表达后被纯化或分离。目的蛋白能以本领域技术人员已知的各种方式被分离或纯化,这取决于样品中存在的其他组分。标准纯化方法包括电泳、分子、免疫和色谱技术,包括离子交换、疏水、亲和以及反相HPLC色谱和色谱聚焦。例如,目的蛋白可以

使用标准抗目的蛋白抗体柱进行纯化。超滤和渗滤技术联合蛋白浓缩也是有用的。所需的纯化程度将取决于目的蛋白的预期用途而变化。在某些情况下，不需要纯化蛋白。

[0278] 在某些其他实施例中，为了证实本公开的经遗传修饰的真菌细胞(即，包含增加ace3-L的表达的遗传修饰的变体真菌宿主细胞)具有产生增加水平的目的蛋白的能力，可以进行各种筛选方法。表达载体可以编码用作可检测标签的与靶蛋白的多肽融合物，或者靶蛋白本身可以用作可选择或可筛选的标记。可以经由蛋白印迹、斑点印迹(可获自冷泉港试验方案网站(Cold Spring Harbor Protocols website)的方法)、ELISA、或者(如果标记为GFP)全细胞荧光或FACS来检测经标记的蛋白。例如，将包含6-组氨酸标签作为与靶蛋白的融合物，并且该标签将通过蛋白印迹检测。如果靶蛋白以足够高的水平表达，可以进行与考马斯/银染色组合的SDS-PAGE以检测与亲本(对照)细胞相比突变体宿主细胞表达的增加，在这种情况下不需要标记。此外，可以使用其他方法来确认目的蛋白的改进水平，例如，检测蛋白活性或每个细胞中的量、蛋白活性或每毫升培养基中的量的增加，从而允许培养或发酵有效地继续更长时间，或通过这些方法的组合。

[0279] 比生产力的检测是评估蛋白生产的另一种方法。比生产力(Q_p)可以通过以下公式确定：

$$[0280] Q_p = gP / gDCW \cdot hr$$

[0281] 其中“gP”是在罐中生产的蛋白的克数，“gDCW”是在罐中的干细胞重量(DCW)的克数，“hr”是从接种时间开始的几小时内的发酵时间，其包括生产时间以及生长时间。

[0282] 在其他实施例中，与(未经修饰的)亲本细胞相比，变体真菌宿主细胞能够产生至少约0.5%，例如，至少约0.5%、至少约0.7%、至少约1%、至少约1.5%、至少约2.0%、至少约2.5%、或甚至至少约3%、或更多的目的蛋白。

[0283] VII. 发酵

[0284] 在某些实施例中，本公开提供了产生目的蛋白的方法，该方法包括发酵变体真菌细胞，其中所述变体真菌细胞分泌目的蛋白。通常，本领域熟知的发酵方法用于发酵变体真菌细胞。在一些实施例中，真菌细胞在分批或连续发酵条件下生长。经典的分批发酵是封闭的系统，其中在发酵开始时设定培养基的组成，并且该组成在发酵期间不改变。在发酵开始时，用一种或多种所希望的生物体接种培养基。在这种方法中，允许发酵发生而不向系统添加任何组分。通常，分批发酵符合关于添加碳源的“分批”的资格，并且经常对控制因素(例如pH和氧浓度)进行尝试。分批系统的代谢物和生物质组成不断变化直到发酵停止时。在分批培养中，细胞通过静态停滞期进展到高生长对数期，最后进入生长速率减少或停止的稳定期。如果不经处理，处于稳定期的细胞最终死亡。通常，在对数期的细胞负责产物的大量生产。

[0285] 标准分批系统的合适的变体是“补料分批发酵”系统。在典型的分批系统的这种变化中，随着发酵的进展，将底物以增量添加。当分解代谢物抑制可能抑制细胞的代谢时并且在培养基中希望具有有限量的底物的情况下，补料分批系统是有用的。在补料分批系统中实际底物浓度的测量是困难的，并且因此基于可测量因素(例如pH、溶解的氧和废气(例如CO₂)的分压)的变化对其进行估计。分批和补料分批发酵是常用的并且在本领域中是熟知的。

[0286] 连续发酵是开放的系统，在该系统中，将定义的发酵培养基连续添加到生物反应

器中,同时去除等量的条件培养基以用于处理。连续发酵通常将培养物保持在恒定的高密度,其中细胞主要处于对数期生长。连续发酵允许对一种或多种影响细胞生长和/或产物浓度的因素进行调节。例如,在一个实施例中,将限制营养素(例如碳源或氮源)保持在固定的速度,并且允许调节所有其他参数。在其他系统中,影响生长的许多因素可以不断改变,而通过培养基浊度测量的细胞浓度保持不变。连续系统努力保持稳定态的生长条件。因此,由于转移培养基而引起的细胞损失应当与发酵中的细胞生长速率相平衡。调节用于连续发酵过程的营养素和生长因子的方法以及最大化产物形成速率的技术在工业微生物学领域中是众所周知的。

[0287] 本公开的某些实施例涉及用于培养真菌的发酵程序。用于产生纤维素酶的发酵程序是本领域已知的。例如,纤维素酶可以通过固体培养或浸没培养(包括分批、补料分批和连续流程)来产生。培养通常在生长培养基中完成,该生长培养基包含含水矿物盐培养基、有机生长因子、碳源和能源材料、分子氧以及当然还有待使用的丝状真菌宿主的起始接种物。

[0288] 除了碳源和能源、氧气、可同化氮和微生物的接种物外,还有必要以恰当比例供给合适量的矿质营养素来确保适当的微生物生长,使微生物转化过程中细胞对碳源和能源的同化最大化,并且在发酵培养基中获得最大细胞产量和最大细胞密度。

[0289] 含水矿质培养基的组成可在宽泛的范围内变化,这部分取决于所使用的微生物和底物,正如本领域所已知的。除了氮以外,该矿质培养基还应该包括处于合适的可溶性可同化离子形式和化合形式的合适量的磷、镁、钙、钾、硫和钠,并且还优选应存在也处于适合的可溶性可同化形式的某些痕量元素如酮、锰、钼、锌、铁、硼和碘以及其他,这些全部如本领域所已知的。

[0290] 该发酵反应是一需氧过程,其中所需的分子氧通过含分子氧气体如空气、富氧空气或甚至基本上纯的分子氧供给,只要维持发酵容器的内容物具有可有效用于帮助微生物物种以旺盛方式生长的合适的氧分压。

[0291] 发酵温度可以稍微变化,但对于丝状真菌如里氏木霉,温度通常将会在约20°C至40°C的范围内,一般优选在约25°C至34°C的范围内。

[0292] 微生物还需要可同化氮源。可同化氮源可以是任何含氮化合物或能够释放适于微生物进行代谢利用的形式的氮。虽然可采用多种有机氮源化合物如蛋白水解产物,但通常可以利用廉价的含氮化合物,如氨、氢氧化铵、尿素和多种铵盐(如磷酸铵、硫酸铵、焦磷酸铵、氯化铵或多种其他氮化合物)。氨气本身便于大规模操作,并且能以合适的量通过鼓泡穿过含水发酵物(发酵培养基)而使用。同时,还可采用这种氨来帮助进行pH控制。

[0293] 含水微生物发酵物(发酵混合物)中的pH范围应该是在约2.0至8.0的示例性范围内。在丝状真菌的情况下,pH通常在约2.5至8.0的范围内;在里氏木霉的情况下,pH通常在约3.0至7.0的范围内。微生物pH范围的偏好在一定程度上取决于所采用的培养基以及特定的微生物,并因而随着培养基的变化而稍微改变,如可以通过本领域技术人员容易地确定的。

[0294] 优选地,发酵以使得含碳底物可被控制为限制性因素的方式进行,从而为细胞提供了含碳底物的良好转化并避免这些细胞被相当量的未转化底物污染。后一种情况对水溶性底物而言不是问题,因为任何剩余的痕量物质均可容易地洗掉。然而,在非水溶性底物的

情况下这可能是个问题，并且需要增加的产物处理步骤如合适的洗涤步骤。

[0295] 如上所述，达到该水平的时间不是关键的，并且可随特定的微生物和所进行的发酵过程而变化。然而，本领域熟知如何确定发酵培养基中的碳源浓度以及所需的碳源水平是否已经达到。

[0296] 发酵可以分批或连续操作进行，为了易于控制、产生均一量的产物以及最经济地使用所有设备，分批补料操作是更为优选的。

[0297] 如果需要，在将含水矿质培养基进料至发酵罐之前，可以将碳源和能源材料的部分或全部和/或可同化氮源（如氨）的部分添加至该含水矿质培养基。

[0298] 优选以预定的速率，或者响应可通过监测例如碳和能量底物的浓度、pH、溶解氧、来自发酵罐的废气中的氧或二氧化碳、通过干细胞重量可测量的细胞密度、光透射比等而确定的需求来控制引入反应器的每一种料流。多种材料的进料速度可以变化以便获得与碳源和能源的有效利用相一致的尽可能快的细胞生长速率，以获得相对于底物变化尽可能高的微生物细胞产量。

[0299] 在分批操作或优选的补料分批操作中，一开始对所有的设备、反应器或发酵装置、器皿或容器、管道、附带的循环或冷却设备等进行灭菌，通常通过采用蒸汽例如在约121°C持续至少约15分钟。然后在存在所有所需的营养物，包括氧和含碳底物的情况下，用所选微生物的培养物接种灭菌过的反应器。所用的发酵罐类型并不重要。

[0300] 从发酵液收集和纯化（例如，纤维素酶）酶还可以通过本领域已知的程序进行。发酵液将通常含有细胞碎片，包括细胞、多种悬浮固体和其他生物质污染物（优选将它们通过本领域已知的手段从发酵液去除）、以及所希望的纤维素酶产物。

[0301] 适用于这样的去除的方法包括常规的固体-液体分离技术，例如像离心、过滤、透析、微量过滤、旋转真空过滤或其他已知的方法，以产生无细胞滤液。可以优选的是使用技术如超滤、蒸发或沉淀在结晶之前进一步浓缩发酵液或无细胞滤液。

[0302] 沉淀上清液或滤液的蛋白组分可以借助于盐（例如硫酸铵）来完成，然后通过各种层析程序（例如离子交换层析、亲和层析或类似的本领域公认的方法）进行纯化。

[0303] 实例

[0304] 应当理解的是，尽管以下实例说明了本公开的实施例，但仅是通过说明的方式给出的。从上述讨论和这些实例，本领域技术人员可以对本公开进行各种改变和修改以使其适应各种用途和条件。此类修改也旨在落入请求保护的发明的范围内。

[0305] 实例1

[0306] 在丝状真菌细胞中过表达产生ace3

[0307] 1A. 概述

[0308] 在本实例中，通过使用原生质体转化用含有pyr2基因、异源启动子和ace3基因的核酸转化亲本里氏木霉细胞，产生表达ace3基因的变体里氏木霉细胞（即，示例性丝状真菌）。如图1和图2中一般呈现的那样，构建了具有两种（2）不同的启动子和两种不同版本的ace30RF（图1；ace3-SC和ace3-L）的以四种不同组合的四（4）种不同的ace3-表达载体（图2A-2D）。选择hxk1（编码己糖激酶的基因）和pki1（编码丙酮酸激酶的基因）的启动子以驱动ace3的组成型表达，然而，技术人员也可以使用和选择其他启动子。

[0309] 1B. 里氏木霉宿主细胞

[0310] 以下实例中示出的里氏木霉亲本宿主细胞衍生自里氏木霉菌株RL-P37(NRRL保藏号15709),其中里氏木霉pyr2基因已被删除,如heir-Neiss和Montenecourt,1984中描述。

[0311] 1C.Ace3表达载体的构建

[0312] 如上文公开的具体实施方式中所示,Ace3是最近显示在诱导条件下(即,在乳糖存在下)产生纤维素酶和半纤维素酶所需的里氏木霉转录因子(Hakkinen等人,2014)。更特别地,Hakkinen等人(2014)使用基于公众可获得的里氏木霉菌株QM6a的基因组序列(参见,genome.jgi.doe.gov/Trire2/Trire2.home.html,其中QM6a预测的注释(蛋白ID 77513)由两个外显子和一个内含子组成(例如,参见图1))的预测的ace30RF。

[0313] 此外,从公众可获得的里氏木霉菌株Rut-C30的基因组序列(参见,(genome.jgi.doe.gov/TrireRUTC30_1/TrireRUTC30_1.home.html)(蛋白ID98455)预测的ace30RF包含更长的蛋白序列(即,相对于来自里氏木霉QM6a的(短)ace3),该更长的蛋白序列包含三个外显子和两个内含子(图1)。更特别地,由“RUT-C30”模型预测的起始密码子位于“QM6a”模型中的起始密码子的上游,并且在C-末端存在无义突变(Poggi-Parodi等人,2014),这产生更长的N-末端序列和更短的C-末端蛋白序列(图1)。

[0314] 在本实例中,克隆了短ace30RF(基于QM6a注释,但包括截短该蛋白的C-末端的RUT-C30无义突变(Ace3-S))和长ace30RF(基于RUT-C30注释(Ace3-L))。如图1所示,短ace3(Ace3-S)和长ace3(Ace3-L)ORF包含C-末端无义突变,如RUT-C30中所见(图1)。为了驱动ace30RF的表达,测试了异源己糖激酶(hxk1)启动子和异源丙酮酸激酶(pk1)启动子。

[0315] 因此,使用标准分子生物学程序构建了四种(4)Ace3-表达载体pYL1、pYL2、pYL3和pYL4(图2A-2D)。这些表达载体含有载体骨架,其具有用于在大肠杆菌中复制和选择的细菌ColE1ori和AmpR基因。除了里氏木霉pyr2选择标记,还存在里氏木霉启动子序列(即,hxk1或pk1启动子)、和具有其天然终止子的ace30RF(ace3-L或ace3-SC)。使用Q5高保真DNA聚合酶(新英格兰生物实验室(New England Biolabs))和下表1中示出的引物,从里氏木霉基因组DNA中PCR扩增里氏木霉启动子和ace30RF。

[0316] 用于PCR扩增每种载体的片段的特异性引物如下列出。为了构建载体pYL1,使用引物对TP13(SEQ ID NO:7)和TP14(SEQ ID NO:8)扩增hxk1启动子,使用引物TP15(SEQ ID NO:9)和TP16(SEQ ID NO:10)扩增Ace3-L ORF,并使用引物对TP17(SEQ ID NO:11)和TP18(SEQ ID NO:12)扩增载体骨架。质粒pYL1的完整序列如SEQ ID NO:21提供。

[0317] 为了构建载体pYL2,使用引物对TP13(SEQ ID NO:7)和TP19(SEQ ID NO:13)PCR扩增hxk1启动子,使用引物TP20(SEQ ID NO:14)和TP16(SEQ ID NO:10)扩增Ace3-SC ORF,并使用引物对TP17(SEQ ID NO:11)和TP18(SEQ ID NO:12)扩增载体骨架。质粒pYL2的完整序列如SEQ ID NO:22提供。

[0318] 为了构建载体pYL3,使用引物对TP21(SEQ ID NO:15)和TP22(SEQ ID NO:16)PCR扩增pk1启动子,使用引物TP23(SEQ ID NO:17)和TP16(SEQ ID NO:10)扩增Ace3-L ORF,并使用引物对TP17(SEQ ID NO:11)和TP24(SEQ ID NO:18)扩增载体骨架。质粒pYL3的完整序列如SEQ ID NO:23提供。

[0319] 为了构建载体pYL4,使用引物对TP21(SEQ ID NO:15)和TP25(SEQ ID NO:19)PCR扩增pk1启动子,使用引物TP26(SEQ ID NO:20)和TP16(SEQ ID NO:10)扩增Ace3-SC ORF,并使用引物对TP17(SEQ ID NO:11)和TP24(SEQ ID NO:18)扩增载体骨架。质粒pYL4的完整

序列如SEQ ID NO:24提供。

[0320] 对于每种载体,使用Gibson组装克隆试剂盒(新英格兰生物实验室(New England Biolabs);目录号:E5510S)根据制造商的方案组装上述三个PCR片段并转化到NEB DH5 α 感受态细胞中。使用Sanger测序对所得载体进行测序,并且它们的图谱显示在图2A-2D中。

[0321] 表1

[0322] 构建体组装引物

引物	序列	SEQ NO:
TP13	TCAGGGTTATTGTCTCATGGCCATTAGGCCTGGCAGGCACTGGCTGGACGACATGT	7
TP14	AGAGCCCTGGCCGGAGCTGCTGAGCCCATTGTTGAATTCTGGCGGGTAGCTGTTGA	8
TP15	TCAACAGCTACCCCGCCAGAATTCAACAATGGGCTCAGCAGCTCCGGCCCAGGGCTCT	9
TP16	TCGTAATAAACAAAGCGTAAGCTAGCTAGCGTAGGTTATGCGAGCAACATTGACGAAAC	10
TP17	GTTTGTGCAATGTTGCTCGCATAACCTACGCTAGCTAGTTACGCTTGTATTACGA	11
TP18	ACATGTCGTCCGAGCCAGTGCCAGGCCATAATGGCCATGAGACAATAACCTGA	12
TP19	AGGTGTAAGACGGGGAGTAGCGCAGCATTGTTGAATTCTGGCGGGTAGCTGTTGA	13
TP20	TCAACAGCTACCCCGCCAGAATTCAACAATGCTGCGCTACTCCCCGTCTTACACCT	14
TP21	TCAGGGTTATTGTCTCATGGCATTAGGCCTAGACTAGCGCCGGTCCCCTATCCCA	15
TP22	AGAGCCCTGGCCGGAGCTGCTGAGCCCATTGGAAGGGGGCGGCCGCGAGCCT	16
TP23	AGGCTCCCGGCCGCCCTTCACCATGGGCTCAGCAGCTCCGGCCCAGGGCTCT	17
TP24	TGGGATAAGGGGACCGGCCGCTAGTCTAGGCCTAAATGGCCATGAGACAATAACCTGA	18
TP25	TGTAAGACGGGGAGTAGCGCAGCATTGTTGAAGGGGGCGGCCGCGAGCCT	19
TP26	AGGCTCCCGGCCGCCCTTCACCATGCTGCGCTACTCCCCGTCTTACA	20

[0324] 1D. 里氏木霉的转化

[0325] 将pYL1、pYL2、pYL3和pYL4的表达载体使用Pac I酶(新英格兰生物实验室(New England Biolabs))线性化,并通过聚乙二醇(PEG)介导的原生质体转化转化入里氏木霉亲本宿主细胞中(Ouedraogo等人,2015;Penttila等人,1987)。使转化体在Vogel基本培养基琼脂板上生长,以选择由pyr2标记获得的尿苷原养型。通过在Vogel琼脂板上转移连续两轮获得稳定的转化体,之后通过铺板稀释孢子悬浮液获得单菌落。含有pYL1、pYL2、pYL3和pYL4的变体(即,经修饰的)宿主细胞被分别命名为变体A4-7、变体B2-1、变体C2-28、和变体D3-1。

[0326] 实例2

[0327] 缓释微量滴定板中的蛋白产生(srMTP)

[0328] 本实例描述了用于鉴定在非诱导条件下分泌酶的转化体(参见,实例1)的筛选方法。例如,在缓释微量滴定板(srMTP)中测试从实例1获得的稳定转化体。使用的srMTP是含有20%葡萄糖(wt/wt)或20%乳糖(wt/wt)的24孔PDMS弹性体板,其是如在PCT国际公开号W02014/047520中所述进行制备。

[0329] 在“非诱导”和“诱导”条件下测试实例1中描述的亲本和变体里氏木霉宿主细胞。在“非诱导条件”下,细胞在srMTP中(含有20%葡萄糖(wt/wt))的1.25ml限定培养基液体(补充2.5%葡萄糖(wt/vol))中生长。在“诱导条件”下,细胞在srMTP中(含有20%乳糖(wt/wt))的1.25ml限定培养基液体(补充2.5%葡萄糖/槐糖(wt/vol))中生长,其中槐糖和乳糖作为用于纤维素酶表达的强效诱导物。

[0330] 如美国专利US 7,713,725中所述进行葡萄糖/槐糖的制备。如在PCT国际公开号WO 2013/096056中一般描述的制备限定培养基,该限定培养基包含9g/L酪蛋白氨基酸、5g/L $(\text{NH}_4)_2\text{SO}_4$ 、4.5g/L KH_2PO_4 、1g/L $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ 、1g/L $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$ 、33g/L PIPPS缓冲液(在pH 5.5)、0.25mL/L里氏木霉微量元素。里氏木霉微量元素含有191.41g/1柠檬酸· H_2O 、200g/L $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ 、16g/L $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ 、0.56g/L $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ 、1.2g/L $\text{MnSO}_4 \cdot \text{H}_2\text{O}$ 和0.8g/L H_3BO_3 。将所有srMTP在28°C孵育大约120小时,伴随280rpm连续振荡。

[0331] 孵育后,收获所有培养物的上清液,并使用聚丙烯酰胺凝胶电泳(PAGE)进行分析。将等体积的培养物上清液在90°C经受还原环境持续十五(15)分钟,然后添加上样染料并用MOPS-SDS缓冲液在4%-12%NuPage™(英杰公司(Invitrogen),卡尔斯巴德CA)聚丙烯酰胺凝胶上分离。将凝胶用SimplyBlue™(Invitrogen)染色并成像(参见,图3)。

[0332] 如图3所示,所测试的所有宿主细胞在诱导物(即,培养基中的槐糖和srMTP中的乳糖)存在下分泌大量蛋白。然而,在不存在诱导物(即,槐糖或乳糖)的情况下,其中葡萄糖是唯一的碳源,仅表达产生Ace3-L ORF的变体宿主细胞(即,变体A4-7和变体C2-28)产生分泌蛋白,而表达Ace3-SC ORF的亲本宿主细胞或变体细胞(即,变体B1-1和变体D3-1)不产生高于PAGE检测极限的细胞外蛋白。该结果清楚地表明Ace3-L(即,与Ace3-S相反)使得能够在里氏木霉中实现无诱导物情况下的蛋白产生。

[0333] 使用纯化的酶作为参照,通过Zorbax C3反相(RP)分析确定分泌蛋白的相对浓度。例如,使用此方法分析上述宿主细胞的分泌蛋白谱,其中观察到所有宿主细胞(即,亲本和变体细胞)在诱导条件下产生相似的纤维素酶蛋白谱,其中纤维素酶由约40%CBH1、20%CBH2、10%EG1和7%EG2组成。在非诱导条件下,纤维素酶在亲本细胞和表达变体Ace3-S的宿主细胞中检测量低。相反,令人惊讶地发现表达变体Ace3-L的宿主细胞(即,变体A4-7和C2-28)产生与诱导条件下相似的纤维素酶比率。

[0334] 简而言之,这种分析方法如下进行:将上清液样品在50mM乙酸钠缓冲液(pH 5.0)中稀释,并通过添加20ppm EndoH进行去糖基化,在37°C孵育3小时。将十(10)μl 90%乙腈添加至100μL EndoH处理的样品中,并在注射前通过0.22μm过滤器。使用具有DAD检测器的Agilent 1290(安捷伦科技公司(Agilent Technologies)HPLC,其配备有Agilent Zorbax300SB C3RRHD 1.8um(2.1x 100mm)柱。该柱在60°C以1.0mL/min的流速操作,用在MiliQ水中的0.1%三氟乙酸(TFA)作为运行缓冲液A,以及在乙腈中0.07%TFA作为运行缓冲液B。DAD检测器在220nm和280nm下操作,具有4nm窗口。注射体积为10μL。

[0335] 另外,注意到表达变体Ace3-L的宿主细胞在具有葡萄糖的srMTP中产生约20%-30%的总细胞外蛋白(即,与在具有乳糖的srMTP中产生的总细胞外蛋白相比)。这种相对低的表达可能是由于具有葡萄糖的srMTP的高葡萄糖进料速率。例如,已经确定通常在低生长速率情况下表观察到最高的纤维素酶和半纤维素酶产生率(Arvas等人,2011)。然而,srMTP生长测定是相对高通量的测定,用于筛选用于蛋白产生的稳定的菌落。

[0336] 总之,表达Ace3-L ORF的变体里氏木霉宿主细胞能够在不存在诱导物的情况下产生纤维素酶和半纤维素酶,尽管蛋白产率较低。更特别地,这种低产生率与srMTP生长方法相关,而不是与宿主细胞的生产能力相关,如下文实例3和实例4中所示。

[0337] 实例3

[0338] 摆瓶中的蛋白产生

[0339] 为了进一步探索和验证上述srMTP结果,在摇瓶中在50-mL浸没培养中存在和不存在诱导物底物的情况下,使亲本宿主细胞和表达变体Ace3-L的宿主细胞生长。更特别地,亲本里氏木霉宿主细胞、变体A4-7细胞和变体C2-28细胞在诱导条件(即,葡萄糖/槐糖作为碳源)和非诱导条件(即,葡萄糖作为碳源)下、在浸没(液体)培养物中生长,并将比较它们各自的细胞外(分泌的)蛋白产生水平。简而言之,将每种宿主细胞(即,里氏木霉亲本宿主细胞、变体A4-7宿主细胞和变体C2-28宿主细胞)的菌丝体分别添加至带底部挡板的250-mL锥形瓶中的50-mL的YEG液体培养基中。YEG液体培养基含有5g/L酵母提取物和22g/L葡萄糖。使细胞培养物生长48小时,然后传代培养到新鲜的YEG中持续另外的24小时。然后在带底部挡板的250mL摇瓶中,将这些种子培养物接种到补充有1.5%葡萄糖(非诱导条件)的50mL限定培养基、或补充有1.5%葡萄糖/槐糖(诱导条件)的50mL限定培养基中。

[0340] 将所有摇瓶在28°C孵育,伴随200rpm连续振荡。孵育3天后,收获来自所有细胞培养物的上清液,并使用PAGE进行分析,如上文实例2中所述。使用Bio-Rad试剂(Thermo Scientific®;目录号:23236)和作为标准的牛血清白蛋白(BSA)的五个稀释液在595nm处通过Bradford染料结合测定法测量上清液中的总蛋白。通过高效液相色谱(HPLC)分析测量葡萄糖浓度,并且在孵育3天后在培养物中未检测到葡萄糖。

[0341] 如图4所示,亲本(对照)里氏木霉细胞在具有葡萄糖/槐糖(诱导)的限定培养基中产生464μg/mL总分泌蛋白,并且在具有葡萄糖(非诱导)的限定培养基中仅产生140μg/L的总分泌蛋白。相反,变体A4-7细胞和C2-28细胞在诱导和非诱导条件下均产生相似量的分泌蛋白,两者都高于在具有槐糖(诱导)的亲本(对照)细胞中产生的分泌蛋白。因此,这些结果证明含有Ace3-L ORF的变体细胞(即,变体A4-7和C2-28细胞)不仅在不存在诱导物的情况下产生细胞外蛋白,而且这些变体细胞在此类诱导条件下也比亲本(对照)里氏木霉细胞产生更多的总蛋白。

[0342] 实例4

[0343] 小规模补料分批发酵中的蛋白产生

[0344] 本实例显示变体Ace3-L表达细胞(即,变体A4-7和C2-28细胞)在小规模发酵中在存在和不存在诱导物底物的情况下产生相似量的纤维素酶和半纤维素酶。更特别地,里氏木霉发酵通常如美国专利号7,713,725中所述使用种子培养物在2L生物反应器中的柠檬酸盐基本培养基中进行。更具体地,在发酵过程中,在不同时间点收获来自所有培养物的上清液,并使等体积的培养物上清液经受PAGE分析。如图5所示,亲本(对照)里氏木霉细胞仅在槐糖诱导物存在下产生分泌蛋白。相反,变体A4-7和C2-28细胞(图5)在诱导和非诱导条件下产生相似量的蛋白。

[0345] 实例5

[0346] 用于Ace3表达的异源启动子

[0347] 本实例阐述使用驱动ace3-L表达的十三(13)种不同的启动子在“非诱导”条件下的增强的蛋白产生。更特别地,通过用含有pyr2基因、异源启动子和ace3-L基因的端粒载体转化亲本里氏木霉细胞(使用原生质体转化)产生表达ace3-L基因的里氏木霉细胞。

[0348] 因此,选择十三种里氏木霉启动子以驱动ace3-L ORF的表达,其中测试的十三种启动子包括但不限于:(i)甲酰胺酶基因(rev3;蛋白ID 103041)启动子(SEQ ID NO:15)、(ii)β-木糖苷酶基因(bx1;蛋白ID 121127)启动子(SEQ ID NO:16)、(iii)转酮醇酶基因

(tk11;蛋白ID 2211)启动子(SEQ ID NO:17)、(iv)功能未知的基因(蛋白ID 104295)启动子(SEQ ID NO:18)、(v)氧化还原酶基因(dld1;蛋白ID 5345)启动子(SEQ ID NO:19)、(vi)木聚糖酶IV基因(xyn4;蛋白ID 111849)启动子(SEQ ID NO:20)、(vii) α -葡糖醛酸糖苷酶基因(蛋白ID 72526)启动子(SEQ ID NO:21)、(viii)乙酰木聚糖酯酶基因1(axe1;蛋白ID 73632)启动子(SEQ ID NO:22)、(ix)己糖激酶基因(hxk1;蛋白ID 73665)启动子(SEQ ID NO:23)、(x)线粒体载体蛋白基因(dic1;蛋白ID 47930)启动子(SEQ ID NO:24)、(xi)寡肽转运蛋白基因(opt;蛋白ID 44278)启动子(SEQ ID NO:25)、(xii)半乳糖激酶基因(gut1;蛋白ID 58356)启动子(SEQ ID NO:26)、和(xiii)丙酮酸激酶基因(pki1;蛋白ID 78439)启动子(SEQ ID NO:27)。蛋白ID(PID)编号来自genome.jgi.doe.gov/Trire2/Trire2.home.html。因此,选择上述十三种启动子以驱动表达,因为其基因通常当葡萄糖浓度高时在生长期以低水平表达,并且当葡萄糖浓度低时或在槐糖诱导条件下以更高水平表达。

[0349] 以下表2总结了使用标准分子生物学方法构建的十三种启动子及其表达载体。更特别地,在本实例(表2)中测试的表达载体包含载体骨架,其具有用于在大肠杆菌中复制和选择的细菌ColE1ori和AmpR基因,以及用于在酿酒酵母中复制和选择的2 μ ori和Ura3基因。此外,存在里氏木霉端粒序列(“TrTEL”)、里氏木霉pyr2选择标记、里氏木霉启动子序列、和具有其天然终止子序列的ace3-L ORF。代表性的载体图显示在图7中,图7描绘了含有dic1启动子的载体pYL8。因此,其他载体(例如,pYL9、pYL12等)具有图7中呈现的相同的序列,除了不同的启动子序列。

[0350] 表2

[0351] 利用不同的真菌启动子来驱动ACE3-L表达的ACE3-L表达构建体

载体#	启动子	ace3 ORF	
pYL7	opt (SEQ ID NO: 25)	ace3-L	
pYL8	dic1 (SEQ ID NO: 24)	ace3-L	
pYL9	gut1 (SEQ ID NO: 26)	ace3-L	
pYL12	hxk1 (SEQ ID NO: 23)	ace3-L	
pYL13	pki1 (SEQ ID NO: 27)	ace3-L	
pYL22	rev3 (SEQ ID NO: 15)	ace3-L	
[0352]	pYL23	PID 104295 (SEQ ID NO: 18)	ace3-L
	pYL24	tkl1 (SEQ ID NO: 17)	ace3-L
	pYL25	bxl (SEQ ID NO: 16)	ace3-L
	pYL27	dld1 (SEQ ID NO: 19)	ace3-L
	pYL28	xyn4 (SEQ ID NO: 20)	ace3-L
	pYL29	PID 72526 (SEQ ID NO: 21)	ace3-L
	pYL30	axe1 (SEQ ID NO: 22)	ace3-L

[0353] 通过聚乙二醇(PEG)介导的原生质体转化将表达载体插入(转化)到里氏木霉亲本宿主菌株(包含非功能性pyr2基因)中(Ouedraogo等人,2015;Penttila等人,1987)。使转化体在Vogel基本培养基琼脂板上生长,以选择由pyr2标记获得的尿苷原养型。通过在Vogel琼脂板上连续两轮转移、然后在非选择性PDA板上连续两轮生长、以及在Vogel琼脂板上一

轮生长来获得稳定的转化体,之后通过铺板稀释孢子悬浮液获得单菌落。

[0354] 在“非诱导”和“诱导”条件下测试以上描述的亲本和转化的(子代)里氏木霉宿主细胞。例如,在“非诱导条件”下,细胞在常规24孔微量滴定板(MTP)中的1.25mL限定培养基液体(补充有2.5%葡萄糖(wt/vol))中生长。在“诱导条件”下,细胞在MTP中的1.25mL限定培养基的液体(补充有2.5%葡萄糖/槐糖(wt/vol))中生长,其中槐糖作为用于纤维素酶表达的强效诱导物。孵育后,收获来自所有培养物的上清液,并使用Bio-Rad试剂(Thermo Scientific®;目录号:23236)和作为标准的牛血清白蛋白(BSA)的五个稀释液在595nm处通过Bradford染料结合测定法测量总蛋白。

[0355] 如表3所示,亲本里氏木霉细胞仅在槐糖诱导物存在下产生高水平的分泌蛋白。相反,在诱导和非诱导条件下,包含并表达由十三种不同启动子驱动的Ace3-L的变体(子代)里氏木霉细胞均产生相似量的分泌蛋白。如表3所示,每种经修饰的(子代)菌株的蛋白水平表示为相对于由亲本菌株(LT4)在葡萄糖/槐糖(Glu/Sop)诱导条件下产生的蛋白(浓度)的比率。

[0356] 表3

[0357] 在诱导(“glu/sop”)和非诱导(“glu”)条件下,相对于经修饰的里氏木霉(子代)菌株,里氏木霉亲本菌株(LT4)的总分泌蛋白

[0358]

菌株ID	启动子	Glu/Sop ¹	Glu ²
LT4(亲本)	N/A	1.00	0.20
LT82	opt	1.16	1.07
LT83	dic1	1.35	1.12
LT85	gut1	0.95	1.08
LT86	hxk1	0.96	0.73
LT87	pkl1	1.16	0.98
LT149	rev3	0.96	0.76
LT150	PID 104295	0.94	0.41
LT151	tk11	1.02	1.01
LT152	bxl	1.00	1.04
LT154	dld1	1.00	0.58
LT155	xyn4	0.95	0.95
LT156	PID 72526	0.95	0.89
LT157	axe1	1.01	0.85

[0359] Glu/Sop¹是“葡萄糖/槐糖”的缩写;诱导条件。

[0360] Glu²是“葡萄糖”的缩写;非诱导条件。

[0361] 另外,如实例3中一般描述的,在摇瓶实验中进一步测试上述里氏木霉亲本菌株及其转化体。例如,里氏木霉子代菌株“LT83”的代表性结果显示在图8中,其中子代菌株LT83包含由dic1启动子(SEQ ID NO:28)驱动的ace3-L ORF。如图8所示,亲本(对照)里氏木霉细胞仅在槐糖(“Sop”)诱导物的存在下产生分泌的蛋白,而子代菌株LT83在诱导(“Sop”)和非诱导(“Glu”)条件下均产生相似量的分泌蛋白。

[0362] 同样地,如实例4中一般描述的那样进行小规模发酵。更特别地,如图9所示,亲本

(对照)里氏木霉菌株仅在槐糖诱导物(“Sop”)存在下产生分泌蛋白,而子代菌株LT83在诱导(“Sop”)和非诱导(“Glu”)条件下均产生相似量的蛋白。

[0363] 实例6

[0364] 里氏木霉ACE3过表达构建体的克隆

[0365] 尽管在联合基因组研究所(<https://genome.jgi.doe.gov/>)可公开获得里氏木霉的基因组序列,但ace3编码区5'末端的位置并不明显。例如,即使DNA序列相同,联合基因组研究所的DNA序列的注释在突变菌株Rut-C30 (genome.jgi.doe.gov/TrireUTC30_1/TrireUTC30_1.home.html)与野生型菌株QM6a (genome.jgi.doe.gov/Trire2/Trire2.home.html)之间也不同。在QM6a情况下,ace3编码区的5'末端被建议在外显子3的上游(5')和并且在内含子2内(如图11,箭头3所示)。在Rut-C30情况下,ace3编码区的5'末端在外显子2内(图11,箭头2)。对基因组DNA序列和另外的cDNA序列的进一步分析表明可能存在外显子1和内含子1(如图11所示)。此外,相对于野生型分离株QM6a的序列(图11,箭头5),Rut-C30中ace3编码区的3'末端包含产生过早终止密码子的突变(图11,箭头4)。因此,在本实例中,过表达ace3基因的这些不同可能形式的效果如本文所述进行了实验研究(例如,参见图11、图12、和图13-18)。

[0366] 因此,图12中描绘的不同形式的ace3在里氏木霉中过表达,其中里氏木霉ace3基因的过表达载体被设计成能够在里氏木霉中的葡糖淀粉酶基因座(g1a1)上靶向整合ace3。更特别地,在所有质粒(载体)构建体中,里氏木霉ace3基因在里氏木霉dic1启动子和天然ace3终止子的控制下表达。所述载体进一步包含用于选择里氏木霉转化体的pyr4标记与其天然启动子和终止子。包括pyr4启动子的重复使得能够在g1a1基因座整合后切除pyr4基因。使得能够在大肠杆菌中复制和扩增的载体骨架是EcoRI-XhoI消化的pRS426(Colot等人,2006)。使用表4中给出的引物通过PCR产生用于靶向整合所需的里氏木霉g1a1基因座的5'和3'侧翼、dic1启动子、不同形式的ace3编码区和终止子。侧翼片段的模板是来自野生型里氏木霉QM6a(ATCC保藏号No.13631)的基因组DNA。

[0367] 表4

[0368] 用于产生DNA片段的引物

引物	序列
Gla.5F (SID: 49)	GTAACGCCAGGGTTTCCCAGTCACGACGGTTAAACTCCATACGCAGCAAACATGGGCTTGGC
[0369]	Gla.5R (SID: 50) GTACAGTACTAGGTGTGAAGATTCCGTCAAGCTGGCGGAATGAAGGAGGATGTGTGAGAGG
	DICprom.F (SID: 51) CACACATCCTCCTTCATTCCGCCAAGCTTGACGGAATCTCACACCTAGTACTCGTAC
	Ace3RutC.R (SID: 52) TGACATTTTGTGTTCCAACACAGCATGCTTAGTCCGACGCCCTCGAGTCAGCC
	Ace3term.F (SID: 53) CTGGACTCGAAGGCGTCGGACTAACGATGCTGTGTTGAAACAACAAAAATGTC
	Ace3term.R (SID: 54) GCAGAGCAGCAGTAGTCGATGCTATTAAAGTAGGTTATGCGAGCAACATTG
	Gla.3F (SID: 55) CTCAGCCTCTCAGCCTCATCAGCCGCCGCTGAATCGGCAAGGGTAGTACTAG
	Gla.3R (SID: 56) GCGGATAACAATTACACAGGAAACAGCGTTAAACCACATGCCAGAGTTCGATGCGCAAG
	Ace3_nointron.F (SID: 57) GTACCTCAGCGCTGTCGATAGCTGCACGCAGTGCAGCTATCGACAGCGCTGAGTGCAC
	Ace3_nointron.R (SID: 58) GTGCACTGCACGTGGCATCGCGCAGTGCAGCTATCGACAGCGCTGAGGTACTC
	Ace3QM.F (SID: 59) GCGCGCTCCGCTGTCGTAACATATGCTGCGCTACTCCCCGTCTTAC
	DICprom_QM.R (SID: 60) GTAAGACGGGGAGTAGCGCAGCATAGTTACGACAGCGGAAGCGCCGCCTTATAAGTG
	Ace3RutC.F (SID: 61) GGCGGCGCTCCGCTGTCGTAACATATGGGCTCAGCAGCTCCGCCAGGGCTC
	DICprom_rutc.R (SID: 62) GCCCTGGGCCGGAGCTGCTGAGCCATAGTTACGACAGCGGAAGCGCCGCCTTATAAG
	Ace3cDNA.F (SID: 63) GGCGGCGCTCCGCTGTCGTAACATATGCCACAGCGGCCGCGCAGCAGCTGG
	DICprom_cDNA.R (SID: 64) CAGCTGCTGCCGCCGCGCTGTGGCCATAGTTACGACAGCGGAAGCGCCGCCTTATAAG

[0370] 上表中的“SID”是“序列识别号”的缩写,例如“SEQ ID NO”。

[0371] 对于dic1启动子、ace3基因和终止子,模板是携带这些片段的早期质粒pYL8(参见,实例5)。用Not I消化从早期质粒获得选择标记(pyr4)。用于产生所希望的DNA片段的PCR引物如表4所示。使用琼脂糖凝胶电泳分离PCR产物和消化的片段。根据制造商的方案,用凝胶提取试剂盒(凯杰公司(Qiagen))从凝胶中分离出正确的片段。使用如PCT/EP2013/050126(公布为W0 2013/102674)中所述的酵母同源重组方法,用上述片段构建质粒。从酵母中拯救质粒并转化到大肠杆菌中。选择若干个克隆,分离质粒DNA并测序。表5中呈现了质粒的概述。

[0372] 表5

[0373] 过表达质粒

质粒	密码子	克隆的基因	选择	基因座	注释	基因 SEQ ID	蛋白 SEQ ID	
[0374]	B7683	SC	ace3 SC 形式	PYR4	GLA	QM6a N-末端； RutC-30 C-末端	7	8
	B7684	S	ace3 S 形式	PYR4	GLA	QM6a N-末端； QM6a C-末端	1	3
	B7709	L	ace3 L 形式	PYR4	GLA	RutC-30 N-末端； RutC-30 C-末端	4	6
	B7752	LC	ace3 LC 形式	PYR4	GLA	RutC-30 N-末端； QM6a C-末端	9	10
	B7778	EL	ace3 EL 形式	PYR4	GLA	RutC-30 N-末端； RutC-30 C-末端	11	12
	B7779	LN	ace3 LN 形式	PYR4	GLA	RutC-30 N-末端； RutC-30 C-末端	13	14

[0375] 因此,表5中呈现的构建体的不同之处在于具有不同形式的ace3基因。SC形式是包含外显子3和4、以及内含子3的基因的短形式(参见,图12和图13,SEQ ID NO:7)。更特别地,SEQ ID NO:7的SC形式包含1,713bp外显子3、148bp内含子3和144bp外显子4。SC形式的(3'-末端)C-末端(即,外显子4)与里氏木霉RutC-30菌株具有相同的突变。

[0376] S形式是包含外显子3和4、以及内含子3但在外显子4的C-末端(3'末端)没有突变的基因的短形式(参见,图12和图14,SEQ ID NO:1)。更特别地,S形式包含1,713bp外显子3、148bp内含子3和177bp外显子4。在这两种形式中(即,“SC”和“S”),翻译起始密码子在长内含子2中(参见,图12),如对里氏木霉QM6a菌株所注释的,并且“SC”和“S”形式两者都缺失了推定的DNA结合结构域的编码区的部分。

[0377] L形式是包含外显子2、3和4,以及内含子2(长内含子)和内含子3的基因的长形式(例如,参见,图12和图15,SEQ ID NO:4)。更特别地,L形式包含258bp外显子2、423bp内含子2、1,635bp外显子3、148bp内含子3、和144bp外显子4。“L”形式的(3'-末端)C-末端与里氏木霉RutC-30菌株具有相同的突变。

[0378] LC形式是包含外显子2、3和4,以及内含子2(长内含子)和内含子3的基因的长形式(例如,参见,图12和图16,SEQ ID NO:9)。更特别地,LC形式包含258bp外显子2、423bp内含子2、1,635bp外显子3、148bp内含子3、和177bp外显子4。LC形式在C-末端没有突变。在“L”和“LC”形式中,翻译起始密码子在外显子2内,如对于Rut-C30菌株的JGI的注释。

[0379] EL形式是包含外显子1、2、3和4,以及内含子1、内含子2(长内含子)和内含子3的超长版本的ace3基因(例如,参见,图12和图17,SEQ ID NO:11)。更特别地,EL形式包含61bp外显子1、142bp内含子1、332bp外显子2、423bp内含子2、1,635bp外显子3、148bp内含子3、和144bp外显子4。EL形式的(3'-末端)C-末端与里氏木霉RutC-30菌株具有相同的突变。

[0380] LN形式是含有外显子2、3和4,以及内含子3,但缺乏内含子2的基因的长形式(例如,参见,图12和图18,SEQ ID NO:13)。LN形式的(3'-末端)C-末端与里氏木霉RutC-30菌株具有相同的突变。因此,如上所述,ace3的L、LC、LN和EL形式编码完整的推定DNA结合结构

域。

[0381] 转化至里氏木霉RL-P37菌株中

[0382] 用MssI消化表5中呈现的所有质粒以释放片段用于靶向整合并用琼脂糖凝胶电泳进行分离。例如,图19提供了显示用于转化里氏木霉的代表性片段内DNA序列排列的图。根据制造商的方案,用凝胶提取试剂盒(凯杰公司(Qiagen))从凝胶中分离出正确的片段。使用大约10 μ g纯化的片段以转化里氏木霉RL-P37菌株的pyr4⁻突变体的原生质体。如PCT公开号WO 2013/102674中所述,使用pyr4选择进行原生质体的制备和转化。

[0383] 将转化体划线到选择性培养基板上。筛选生长的克隆用于使用表6中列出的引物通过PCR正确整合。将给出预期信号的克隆纯化为单细胞克隆,并使用表6中列出的引物通过PCR、以及通过DNA印迹(数据未显示)重新筛选正确整合和克隆纯度。

[0384] 表6

[0385] PCR引物

引物	序列	SEQ ID NO:
Gla1_5creen.F	GCTGGAAGCTGCTGAGCAGATC	65
DICprom.R	GTGCCAGCATTCCCCAGACTCG	66
T061_pyr4_orf_screen	TTAGGCGACCTTTTCCA	67
Gla1_3creen.R	GCCGCTCAGGCATAACGAGCGAC	68
DICprom.F2	CTCTGGTCGGCCTGCCGTTG	69
ace3.R	TGAGTATAGCGGCTGACTTGTG	70

[0387] 不同ace3转化体的培养

[0388] 表7中的菌株在24孔微量滴定板中在具有2%乳糖或2%葡萄糖作为碳源的液体培养基中生长。培养基的其他组分是0.45% KH₂P_O4、0.5% (NH₄)₂S_O4、0.1% MgS_O4、0.1% CaC₁₂、0.9% 酪蛋白氨基酸、0.048% 柠檬酸xH₂O、0.05% FeS_O4x7H₂O、0.0003% MnS_O4xH₂O、0.004% ZnS_O4x7H₂O、0.0002% H₃B_O3和0.00014% CuS_O4x5H₂O。包括100mM PIPPS(Calbiochem)以将pH维持在5.5。

[0389] 表7

[0390] 里氏木霉菌株

密码子	菌株的名称	选择
M1904	RL-P37, 亲本菌株	Pyr4
M2015	<i>ace3</i> SC 克隆 2-1	
M2016	<i>ace3</i> SC 克隆 28-3	
M2017	<i>ace3</i> S 克隆 9-1	
M2018	<i>ace3</i> S 克隆 20-1	
M2019	<i>ace3</i> L 克隆 16-1	
M2020	<i>ace3</i> L 克隆 18-1	
M2021	<i>ace3</i> LC 克隆 52-1	
M2022	<i>ace3</i> LC 克隆 14-4	
M2023	<i>ace3</i> EL 克隆 3-3	
M2024	<i>ace3</i> EL 克隆 4-5	
M2025	<i>ace3</i> LN 克隆 3-3	

[0392]	M2026	<i>ace3</i> LN 克隆 4-1	Pyr4
--------	-------	-----------------------	------

[0393] 培养在28°C和800RPM下在具有80%湿度的Infors HT微量振荡器中进行。在第3-7天进行培养物的取样。根据制造商的方案,使用Bio Rad蛋白测定法从培养物上清液测量总分泌蛋白的量。在两种培养基中,具有RutC-30C-末端突变的*ace3*L EL和LN形式的过表达提高了总蛋白的产生。在以乳糖为碳源的培养基中,所有形式的*ace3*基因的过表达在一些程度上提高总蛋白的产生,但在过表达*ace3*基因的L、EL和LN形式的菌株中,提高水平最高。很明显,当葡萄糖(即,非诱导条件)用作转化体(其中*ace3*基因的L、EL或LN形式过表达)的碳源时,观察到高水平的分泌蛋白(表8)。

[0394] 表8

[0395] 在24孔板培养中由不同菌株产生的总蛋白

	菌株	5d, mg/ml	7d, mg/ml
[0396]	M2015	0,77	1,21
	M2016	0,54	0,91
	M2017	0,47	0,81
	M2018	1,23	1,12
	M2019	2,50	6,37
	M2020	1,97	6,53
	M2021	0,67	1,17
	M2022	1,19	1,09
	M2023	1,84	4,10
	M2024	2,05	4,09
	M2025	1,76	3,60
	M2026	1,93	4,04
	M1904	0,40	0,65
[0397]	M2015	0,59	0,89
	M2016	0,12	0,56
	M2017	0,00	0,19
	M2018	0,02	0,26
	M2019	1,57	2,74
	M2020	1,79	2,91
	M2021	0,03	0,40
	M2022	0,06	0,21
	M2023	1,40	2,44
	M2024	1,37	2,52
	M2025	1,24	2,26
	M2026	1,49	2,64
	M1904	0,02	0,37

[0398]

实例7

[0399]

内源Ace3异源启动子敲入

[0400] 启动子替换构建体(参见,图6)通过将包含ace3基因座处天然启动子的上游5'区的DNA片段、loxP-侧翼的潮霉素B抗性选择性标记盒、和包含目的启动子(其有效地融合至ace3可读框的5'末端)的片段融合制成(例如,参见国际PCT申请系列号PCT/US 2016/017113,其进一步描述了用于丝状真菌的基因/启动子替换盒)。

[0401] 因此,用上述启动子替换构建体转化里氏木霉细胞,其中分离转化体并提取基因组DNA用于诊断PCR以确认启动子替换构建体在天然ace3基因座处的同源重组。使用此方法,可以鉴定转化体,其中天然ace3启动子被潮霉素-B抗性盒和任何目的启动子替换。随后,通过cre重组酶的作用除去潮霉素B抗性盒(Nagy, 2000)。

[0402] 如在国际PCT公开号:WO 2016/100272、WO 2016/100571和WO 2016/100568中示例的,通过适当设计的指导RNA,针对天然ace3启动子的cas9的作用可以增强在ace3基因座处的同源整合效率。

[0403] 在Hu等人(2007)的出版物中描述了烟曲霉的条件型启动子替换(CPR)策略,其总

体上描述了使用烟曲霉NiiA氮可调节启动子(pNiiA)来删除和替换所选基因的内源启动子的策略。因此,在某些实施例中,可以使用类似方法用可替代启动子替换里氏木霉中ace3基因的内源启动子。

[0404] 实例8

[0405] 用木质纤维素目的启动子基因取代内源性非木质纤维素目的基因启动子

[0406] 从DNA多核苷酸片段组装里氏木霉葡糖淀粉酶表达构建体(例如,参见美国专利号7,413,879),其中编码里氏木霉葡糖淀粉酶的ORF序列有效地连接到5'(上游)里氏木霉cbh1启动子并有效地连接到3'(下游)里氏木霉cbh1终止子。DNA构建体进一步包含里氏木霉pyr2基因作为选择性标记。

[0407] 因此,用葡糖淀粉酶表达构建体转化本公开的变体(子代)里氏木霉细胞(即,包含增加编码Ace3-L蛋白的基因的表达的遗传修饰)。选择转化体并在具有葡萄糖作为碳源的液体培养基中培养(即,没有诱导底物,如槐糖或乳糖),以鉴定那些在培养期间能够分泌里氏木霉葡糖淀粉酶的转化体。

[0408] 因此,使表达里氏木霉葡糖淀粉酶的(亲本)菌株(对照)和经修饰的表达里氏木霉(子代)葡糖淀粉酶的菌株(即,包含并表达Ace3-L ORF)在摇瓶中生长,并且收获来自所有细胞培养物的上清液并使用PAGE进行分析,如上文实例2和实例3中一般描述的。

[0409] 更特别地,如图10所示,亲本里氏木霉细胞在具有葡萄糖/槐糖(诱导条件)的限定培养基中产生1,029 μ g/mL葡糖淀粉酶,并且在具有葡萄糖的限定培养基(非诱导条件)中仅产生38 μ g/mL葡糖淀粉酶,相反,经修饰的(子代)菌株“LT88”(包含dic1启动子驱动的ace3-L)在“诱导”(“Sop”)条件下产生3倍更高的葡糖淀粉酶(即,相对于在诱导条件下的亲本(对照)菌株),并在“非诱导”(“Glu”)条件下产生2.5倍更高的葡糖淀粉酶(即,相对于在诱导条件下的亲本(对照)菌株)、或在“非诱导”(“Glu”)条件下产生67倍更高的葡糖淀粉酶(相对于在非诱导条件下的亲本(对照)菌株)。因此,这些结果表明,包含Ace3-L ORF的修饰的(子代)细胞不仅在不存在诱导物的情况下产生细胞外蛋白,而且在此类诱导条件下这些变体细胞比亲本(对照)里氏木霉细胞产生更多的总蛋白。

[0410] 实例9

[0411] 用木质纤维素目的基因启动子替换天然相关联的异源目的基因启动子

[0412] 如以下,从DNA多核苷酸片段组装植酸酶表达构建体(例如,参见US专利号8,143,046)。编码布丘氏菌物种植酸酶的ORF在5'端有效地连接至里氏木霉cbh1启动子,并且在3'端有效地连接至里氏木霉cbh1终止子。DNA构建体进一步包含选择性标记、构巢曲霉amdS基因。用植酸酶表达构建体转化变体里氏木霉细胞(即,包含增加编码Ace3-L蛋白的基因的表达的遗传修饰)。选择转化体并在具有葡萄糖作为碳源的液体培养基中培养(即,没有诱导底物,如槐糖或乳糖),以鉴定那些在培养期间能够分泌布丘氏菌植酸酶的转化体。

[0413] 实例10

[0414] 天然ace3启动子替换载体的构建

[0415] 在本实例中,使用标准分子生物学方法构建两个ace3启动子替换载体pCHL760和pCHL761。载体骨架pMCM3282(图20)含有用于在大肠杆菌中复制和选择的pMB1ori和AmpR基因。另外,包括了在粗糙脉孢菌cpc1启动子和构巢曲霉trpC终止子下表达的对于里氏木霉的hph潮霉素选择标记。对于启动子替换,载体含有在里氏木霉pk1启动子下表达的针对玉

米优化的酿脓链球菌cas9密码子、和在U6启动子下表达的指导RNA(例如,参见,PCT公开号: WO 2016/100568和WO 2016/100272)。

[0416] 因此,将pMCM3282用EcoRV消化,并通过Gibson组装将3个片段(具有来自替换ace3天然启动子的里氏木霉hxk1或dic1启动子区域侧翼的里氏木霉ace3基因座的约1kb的5'和3'同源序列)克隆至pMCM3282/EcoRV,生成pCHL760(图21)和pCHL761(图22)。

[0417] 使用Q5高保真DNA聚合酶(新英格兰生物实验室(New England Biolabs))和下表9中列出的引物,从里氏木霉基因组DNA中PCR扩增5'和3'ace3同源序列以及hxk1或dic1启动子。

[0418] 表9

[0419] PCR引物

[0420]	CL1791(SID: 71)	TCTAGTATGTACGAGTACTAGGTGTGAAGATTCCGTCATTTCTCGACAT GCGAATGCG
	CL1792(SID: 72)	TGCCATGCAAACCCCCGCATTCGCATGTCGAGGAAATGACGGAATCTTCA CACCTAGTAC
	CL1793(SID: 73)	TGCAGCTACAGAGCCCTGGGCCGGAGCTGCTGAGCCCATAAGTTACGACA GCGGAAGCGC
	CL1794(SID: 74)	ATAGCACTTATAAGGC GGCGCTCCGCTGTCGTA ACTATGGGCTCAGCA GCTCCGGC
	CL1840(SID: 75)	TAAACAAAATAGGGGTTCCGCGCACATTCCCCGAAAAGTGCCACCTGGA TAGACTAGCA TCTGAGCCATTGCAGC
	CL1786(SID: 76)	AGTGGCACCGAGTCGGTGGTGCTTTCTATCGAGAGCATTGGTCAG TGGTGGCAAG
	CL1800(SID: 77)	ACCAATATACAAAACATGTCGTCGAGCCAGTGCCATTGCCATTCCCTCGAC ATGCGAATGC
	CL1801(SID: 78)	GTTGCCATGCAAACCCCCGCATTCGCATGTCGAGGAAATGGCAGGC ACTG GCTCGGACGAC
	CL1802(SID: 79)	AGCTACAGAGCCCTGGGCCGGAGCTGCTGAGCCCATTGTTGAATTCTGG CGGGGTAGCTG
	CL1803(SID: 80)	CTTTTACACTTTAACAGCTACCCCGCCAGAATTCAACAATGGGCTCA
[0421]		GCAGCTCCGGC
	CL1831(SID: 81)	TAGTCCGTTATCAACTGAAAAAGTGGCACCGAGTCGGTGGTGCTTTT TTTCTATCGAGATGTTCTGGATGGTGGAGAGG

[0422] 上表中的“SID”是“序列识别号”的缩写,例如“SEQ ID NO”。

[0423] 更特别地,用于PCR扩增每种载体的片段的特异性引物如下列出。为了构建pCHL760,使用引物对CL1840和CL1791扩增5'上游同源区,使用引物CL1794和CL1831扩增3'下游同源区,使用引物对CL1792和CL1793扩增dic1启动子。

[0424] 为了构建pCHL761,使用引物对CL1840和CL1800扩增5'上游同源区,使用引物CL1803和CL1831扩增3'下游同源区,使用引物对CL1801和CL1802扩增hxk1启动子。

[0425] 载体pMCM3282(图20)从5'至3'方向包括里氏木霉U6启动子、大肠杆菌ccdB盒、和涉及Cas9结合的单指导RNA(sgRNA)的结构区(包括来自U6基因的内含子)。用对木霉属ace3基因内的五个不同靶位点特异的序列替换ccdB盒。将指导RNA序列插入pCHL760(图21)和pCHL761(图22)以构建最终ace3启动子替换载体。

[0426] 因此,表10中呈现的具有AarI限制性位点的下列寡核苷酸被设计用于产生不同的sgRNA序列。

[0427] 表10

[0428] 寡核苷酸sgRNA序列

寡核苷酸 ID	寡核苷酸 说明	寡核苷酸序列	SEQ ID NO:
CL1821	TS1 的靠前 寡核苷酸	AGTCTATCGCAGCCTGCCTAGCTAATGTTT	82
CL1822	TS1 的殿后 寡核苷酸	TCTAAAACATTAGCTAAGGCAAGGCTGCGATA	
CL1823	TS4 的靠前 寡核苷酸	AGTCTATCGGCAGAGTCGCGTCTCCGGGTTT	
CL1824	TS4 的殿后 寡核苷酸	TCTAAAACCCGGAAGACGCGACTCTGCCGATA	
CL1825	TS5 的靠前 寡核苷酸	AGTCTATCGAATGAGTGTAGGTACGAGTAGTTT	
CL1826	TS5 的殿后 寡核苷酸	TCTAAAACACTCGTACCTACACTCATTGATA	
CL1827	TS8 的靠前 寡核苷酸	AGTCTATCGGCCGCAATAGCTTCTAATGTTT	88
CL1828	TS8 的殿后 寡核苷酸	TCTAAAACATTAGGAAGCTATTGCGGCCGATA	89
CL1829	TS10 的靠 前寡核苷酸	AGTCTATCGCAGCGCAATCAGTGCAGTGGTTT	90
CL1830	TS10 的殿 后寡核苷酸	TCTAAAACCACTGCACTGATTGCGCTGCGATA	91

[0431] 更特别地,将CL1821和CL1822、CL1823和CL1824、CL1825和CL1826、CL1827和CL1828、CL1829和CL1830退火以产生双链DNA,使用typeIIS无缝克隆方法将其分别克隆pCHL760和pCHL761的AarI位点。具有正确插入的指导RNA序列的最终质粒失去毒性ccdB基因。

[0432] 里氏木霉的转化

[0433] 通过聚乙二醇(PEG)介导的原生质体转化将pCHL760和pCHL761的cas9介导的ace3启动子替换载体转化到里氏木霉亲本细胞中。将转化体在具有潮霉素的Vogel的基本培养基琼脂上培养,以选择潮霉素抗性转化体。这些转化体中的一些不稳定,已经摄入质粒,但没有稳定整合到基因组DNA中。将转化体转移到Vogel非选择性琼脂培养基上以使质粒和潮霉素抗性标记丢失。

[0434] 为了筛选dic1启动子替换的转化体,提取基因组DNA并使用引物对CL1858和CL1848(期望整合的预期产物大小为2,412bp)、以及CL1853和CL1818(期望整合的预期产物大小为2,431bp)进行PCR,例如参见表11。随后通过DNA测序分析PCR产物以确认所希望的启动子整合。

[0435] 为筛选hxk1启动子替换的转化体,使用引物对CL1858和CL1898(期望整合的预期产物为大小1,784bp)、以及CL1853和CL1850(期望整合的预期产物大小为2,178bp)进行PCR,并且随后通过对PCR产物进行DNA测序确认正确的整合,例如参见表11。

[0436] 表11

[0437] PCR引物

	引物序列	SEQ ID NO:
引物 ID		
[0438] CL1858	TGGAGAGACTCGGAGAGGATAGG	92
CL1853	AGCGTGGAGGCAGTTGGAGTGG	93
[0439] CL1848	TGGACAAAGCCTGGGTCCCTGCTCC	94
CL1818	ATCCTGACTCGTCCTGTGTCGG	95
CL1898	AGTGCTTCGTTAGTGGACTTG	96
CL1850	CTCGGTAGCTGCTTGAATATAG	97

[0440] 摆瓶中的蛋白产生

[0441] 为了测试ace3启动子替换的菌株的功能,在摇瓶中在50ml浸没培养中存在和不存在诱导物底物(槐糖)的情况下培养细胞。使亲本里氏木霉宿主细胞(ID号1275.8.1)和变体细胞ID号2218、2219、2220、2221、2222和2223在液体培养物中在诱导条件(葡萄糖/槐糖作为碳源)和非诱导条件(葡萄糖作为碳源)两者下生长,并且比较它们各自的细胞外分泌蛋白产生水平。简而言之,将每种宿主细胞(即,里氏木霉亲本宿主细胞及其变体细胞)的菌丝体分别添加至在带底部挡板的250mL锥形瓶中的50mL的YEG液体培养基中。YEG液体培养基含有5g/L酵母提取物和22g/L葡萄糖。使细胞培养物生长48小时,然后传代培养到新鲜的YEG中持续另外的24小时。然后在带底部挡板的250mL摇瓶中,将这些种子培养物接种到补充有2.5%葡萄糖(非诱导条件)的50mL限定培养基、或补充有2.5%葡萄糖/槐糖(诱导条件)的50mL限定培养基中。将所有摇瓶在28℃孵育,伴随200rpm连续振荡。孵育4天后,收获所有细胞培养物的上清液,并使用SDS-PAGE分析,如图23所示。

[0442] 如以上在SDS-PAGE所见,与葡萄糖/槐糖(诱导)相比,亲本细胞(图23, ID 1275.8.1)在具有葡萄糖(非诱导)的限定培养基中产生更少的分泌蛋白。相反,转化体2218、2219、2220、2222和2223在诱导和非诱导条件下产生相似量的分泌蛋白。因此,这些结果证明含有替代ace3基因座处的天然ace3启动子的hxk1或dic1启动子的变体细胞在不存在诱导物的情况下产生细胞外蛋白。

[0443] 参考文献

- [0444] 欧洲专利申请号EP 215,594
- [0445] 欧洲专利申请号EP 244,234
- [0446] 欧洲申请号EP 0215594
- [0447] PCT国际申请序列号PCT/EP 2013/050126
- [0448] PCT国际申请序列号PCT/US 2016/017113
- [0449] PCT国际申请号WO 03/027306
- [0450] PCT国际申请号WO 1992/06183
- [0451] PCT国际申请号WO 1992/06209
- [0452] PCT国际申请号WO 1992/06221
- [0453] PCT国际申请号WO 1992/10581
- [0454] PCT国际申请号WO 1998/15619
- [0455] PCT国际申请号WO 2002/12465

- [0456] PCT国际申请号WO 2003/52054
- [0457] PCT国际申请号WO 2003/52055
- [0458] PCT国际申请号WO 2003/52056
- [0459] PCT国际申请号WO 2003/52057
- [0460] PCT国际申请号WO 2003/52118
- [0461] PCT国际申请号WO 2004/16760
- [0462] PCT国际申请号WO 2004/43980
- [0463] PCT国际申请号WO 2004/48592
- [0464] PCT国际申请号WO 2005/001036
- [0465] PCT国际申请号WO 2005/01065
- [0466] PCT国际申请号WO 2005/028636
- [0467] PCT国际申请号WO 2005/093050
- [0468] PCT国际申请号WO 2005/28636
- [0469] PCT国际申请号WO 2005/93073
- [0470] PCT国际申请号WO 2006/074005
- [0471] PCT国际申请号WO 2006/74005
- [0472] PCT国际申请号WO 2009/149202
- [0473] PCT国际申请号WO 2010/141779
- [0474] PCT国际申请号WO 2011/038019
- [0475] PCT国际申请号WO 2011/063308
- [0476] PCT国际申请号WO 2011/153276
- [0477] PCT国际申请号WO 2012/125925
- [0478] PCT国际申请号WO 2012/125951
- [0479] PCT国际申请号WO 2012125937
- [0480] PCT国际申请号WO 2013/102674
- [0481] PCT国际申请号WO 2014/047520
- [0482] PCT国际申请号WO 2014/070837
- [0483] PCT国际申请号WO 2014/070841
- [0484] PCT国际申请号WO 2014/070844
- [0485] PCT国际申请号WO 2014/093275
- [0486] PCT国际申请号WO 2014/093281
- [0487] PCT国际申请号WO 2014/093282
- [0488] PCT国际申请号WO 2014/093287
- [0489] PCT国际申请号WO 2014/093294
- [0490] PCT国际申请号WO 2015/084596
- [0491] PCT国际申请号WO 2016/069541
- [0492] PCT国际申请号WO 2016/100272
- [0493] PCT国际申请号WO 2016/100568
- [0494] PCT国际申请号WO 2016/100571

- [0495] 美国专利号6,022,725
- [0496] 美国专利号6,268,328
- [0497] 美国专利号7,413,879
- [0498] 美国专利号7,713,725
- [0499] 美国专利号8,143,046
- [0500] Alexopoulos,C.J.,Introductory Mycology,New York:Wiley,1962.
- [0501] Allen and Mortensen,Biotechnol.Bioeng.,2641-45,1981.
- [0502] Arvas,M.,Pakula,T.,Smit,B.,Rautio,J.,Koivistoinen,H.,Jouhten,P.,Lindfors,E.,Wiebe,M.,Penttila,M.,and Saloheimo,M.,“Correlation of gene expression and protein production rate-a system wide study”,BMC Genomics12,616,2011.
- [0503] Ausbel et al.,“Current Protocols in Molecular Biology”,Green Publishing Associates/Wiley Interscience,New York,1987.
- [0504] Ausubel et al.,Current Protocols in Molecular Biology,Green Publishing Associates/Wiley Interscience,New York,1994.
- [0505] Boel et al.,EMBO J.3:1581-1585,1984.
- [0506] Campbell et al.,Curr.Genet.,16:53-56,1989.
- [0507] Cao et al.,Science,9:991-1001,2000.
- [0508] Colot et al.,PNAS 103(27):10352-10357,2006.
- [0509] Devereux et al.,Nucleic Acids Res.12:387-395,1984.
- [0510] el-Gogary et al.,“Mechanism by which cellulose triggers cellobiohydrolase I gene expression in *Trichoderma reesei*”,PNAS,86(16)-6138-6141,1989.
- [0511] Hakkinen,M.,Valkonen,M.J.,Westerholm-Parvinen,A.,Aro,N.,Arvas,M.,Vitikainen,M.,Penttila,M.,Saloheimo,M.,and Pakula,T.M.,“Screening of candidate regulators for cellulase and hemicellulase production in *Trichoderma reesei* and identification of a factor essential for cellulase production”,Biotechnol Biofuels 7,14,2014.
- [0512] Harkki et al.,BioTechnol.,7:596-603,1989.
- [0513] Harkki et al.,Enzyme Microb.Technol.,13:227-233,1991.
- [0514] Hu et al.,“Essential gene identification and drug target prioritization in *Aspergillus fumigatus*”PLoS Pathog.,3(3),2007.
- [0515] Ilmen et al.,“Regulation of cellulase gene expression in the filamentous fungus *Trichoderma reesei*”,Applied and Environmental Microbiology,63(4)-1298-1306,1997.
- [0516] Ju and Afolabi,Biotechnol.Prog.,91-97,1999.
- [0517] Kriegler,Gene Transfer and Expression:A Laboratory Manual,1990.
- [0518] Martinez et al.,“Genome sequencing and analysis of the bio-mass-degrading fungi *Trichoderma reesei* (syn.*Hypocrea jecorina*)”,Nature

Biotechnology, 26:533-560, 2008.

[0519] Mullaney et al., MGG 199:37-45, 1985.

[0520] Nagy, "Cre recombinase:the universal reagent for genome tailoring", Genesis 26(2), 99-109, 2000.

[0521] Needleman and Wunsch, J.Mol.Biol., 48:443, 1970.

[0522] Nunberg et al., Mol.Cell Biol.4:2306, 1984.

[0523] Ouedraogo, J.P., Arentshorst, M., Nikolaev, I., Barends, S., and Ram, A.F., "I-SceI-mediated double-strand DNA breaks stimulate efficient gene targeting in the industrial fungus Trichoderma reesei" Applied microbiology and biotechnology 99, 10083-10095, 2015.

[0524] Pearson and Lipman, Proc.Natl.Acad.Sci.USA 85:2444, 1988.

[0525] Penttila, M., Nevalainen, H., Ratto, M., Salminen, E., and Knowles, J., "A versatile transformation system for the cellulolytic filamentous fungus Trichoderma reesei", Gene 61, 155-164, 1987.

[0526] Poggi-Parodi, D., Bidard, F., Pirayre, A., Portnoy, T., Blugeon, C., Seiboth, B., Kubicek, C.P., Le Crom, S., and Margeot, A., "Kinetic transcriptome analysis reveals an essentially intact induction system in a cellulase hyper-producer Trichoderma reesei strain", Biotechnol Biofuels 7, 173, 2014.

[0527] Sambrook et al., Molecular Cloning, A Laboratory Manual, 2nd Edition, Cold Spring Harbor Laboratory Press, Cold Spring, New York, 1989.

[0528] Sambrook et al., Molecular Cloning, A Laboratory Manual, 4th Edition, Cold Spring Harbor Laboratory Press, Cold Spring, New York, 2012.

[0529] Seiboth, et.al., Mol.Genet.Genomics, 124-32, 2002.

[0530] Sheir-Neiss and Montenecourt, "Characterization of the secreted cellulases of Trichoderma reesei wild type and mutants during controlled fermentations", Applied Microbiology and Biotechnology, 20(1):46-53, 1984.

[0531] Smith and Waterman, Adv.Appl.Math.2:482, 1981.

[0532] Vaheri et al., "Transglycosylation products of the cellulase system of Trichoderma reesei", Biotechnol.Lett., 1:41-46, 1979.

[0533] Yelton et al., PNAS USA 81:1470-1474, 1984.

<110> 丹尼斯科公司 (Danisco)
 <120> 在不存在诱导底物下丝状真菌细胞中的蛋白产生
 <130> NB41159-WO-PCT
 <150> US 62/240,787
 <151> 2016-10-04
 <160> 102
 <170> PatentIn 版本 3.5
 <210> 1
 <211> 2038
 <212> DNA
 <213> 里氏木霉
 <400> 1

atgtcgct actccccgt cttacacctg gatactctc cttgccacc actgaccaat	60
gcttcccc gcccaaagt cgagtaacctc agcgctgtcg atagctgcac gcactgccgc	120
gatggccacg tgcagtgac ttgcacctg cccctggcgc gacgcccggc caaaggcagg	180
aagaagagcg accagcccg ccagccgcct cctgatccga getcgcttc caccgcggct	240
cgaccggcc agatgccgccc gccgctgacc ttccggcc cccgacttagc cgccgtcgac	300
cccttcgcct cgtcgctcgct gtgcggcgc gcccggcggg agccgcgtga gcccgtcagc	360
attgacaacg gcctgccccg gcagccgcgtg ggcgacctgc cccgccttc caccatccag	420
aacatctcga cgcgccagcg atggatacac ctggccaacg ccatgacgct ggcacacacg	480
acgctagagc gctgtcgaa gctgttate gaccttttct tgcactacct ctacccctc	540
[0001] accccccctgg tgtacgagcc ggccctccgg gacgtgtcgctg catacattt ctcccgcccc	600
ttgcctggcg tcaaccaacc atgcgcgtc tcacagctca cgccagaccc gaccacggc	660
accacccccc tcaacgcgtc cgagtcgtgg gccggcttg gccagcccg cggctcgca	720
accgtcggtca gcaggctggc tccctggcc gactcgaccc tcaccctgtt cacggccgtc	780
tgcgcagagg cagcattcat gctaccaag gacattttcc ccgaaggaga atccgtctct	840
gagatcttgc tcgaaggcctc tcgggactgc ctgcaccaggc acctcgaggc cgacctggag	900
aatccgacgg ccaactcgat tgccattcgc tacttccact ccaactgcct ccacgctcg	960
gsgaagccca agtactcggt gcacatattt ggccggccca tccgcctggc gcaggctat	1020
cagctgcacg aggaggctgc cctcgagggg ctcgtccca tcgaggcaga gttccggcgt	1080
cgcgtcttt ggatcctgta cttggcgcac aagtcaaggc ctatactcaa caatggccc	1140
ataccatecc acaagtgactg ctgcgacgccc ggcatcacca cgctatacc gtcgggtatc	1200
gaggacgagt tcctgagcac ggcgtccgag ccgcggccga agagcttcat atccggcttc	1260
aacgcaaattg tgccgtctg gcagtcgcg gctgatttc tgctggaaat ccgcgtcg	1320
caagatcaga tgcgtcgac cttcgaggg accatcccc cgaaccatgt gtcgcctcc	1380
gccgacaggc agcatctcga ttctcttat gtcgcgttca tcacgtgtt ggacgtatc	1440
ccggcgtacc tccagtcgtg cactctggcg atggcagcga tggcagaagg caacgggtct	1500
gccgagtcca agcagtagt gatacgtgc atcaacgtgc aggtgacgtt tcactgtctg	1560
cgcgttca gaaagtcgga gattgtgcga gacatgtga gggtgatgaa cgaggccccc	1620
ttttggggcc tgcaggccaa tggcggccaa aacgtgagtc gtttccttgc ctcttc	1680
ttctgcacac cttttcttc gacgaccccc cctctcttt tatatccctg cggatatgtat	1740
tatcatcaag ctcggcact tggtgcta at ctgcgtat tatgttgc tatgttgc	1800
	1860

ggttgaaaaag attcgcccta tcggagctag tttgctggcc atcatccatc gcaaccagga	1920
ttcacccctt gctacgcgag ccaggagcga ctttccgtg cttttggata ttctcacg	1980
gctggactcg aaggcgtcg accaactgag gaatacgtcc actaccgttg ttggctaa	2038
<210> 2	
<211> 1890	
<212> DNA	
<213> 里氏木霉	
<400> 2	
atgctgcgct actccccgt cttacacctg gatactctc cttgccacc actgaccaat	60
gctttcccc gcccaaagtg cgagtacctc agcgctgtcg atagctgcac gcactgccgc	120
gatgcccacg tgcagtgcac ttgcacactg cccctggcgc gacgcggccc caaagcgagg	180
aagaagagcg accagcccg ccagccgcct cctgatccga gctcgcttc caccgcggct	240
cgaccggcc agatgccgc gccgctgacc ttctccggcc ccgcagtagc cgcgctgcag	300
cccttcgcct cgctgcgtgt gtcgcccgc acggcctggg agcccgctga gccgcetcage	360
attgacaacg gcctgcggcc gcagccgcgt ggccgactgc ccgccttc caccatccag	420
aacatctcga cgcgcgcgcg atggatacac ctggccaacg ccatgacgct ggcacacacg	480
acgctagagc gcgtctcgaa gcgtatgtac gacctttctc tcgactactt ctacccctc	540
accccccctgg tgtacgagcc ggccctccgg gacgtgtcg catacattt ctcccagccc	600
ttgcctggcg tcaaccaacc atgcgcgtg tcacagctca cgcagaccc gaccaccggc	660
accacccccc tcaacgcgtc cgagtcgtgg gccggctttg gccagcccg cggtcgcga	720
accgtcggca gcaggctggc tccctggcc gactcgacct tcacccttgt cacggcgctc	780
tgccgcaggc cagcattcat gtcacccaaag gacattttcc ccgaaggaga atccgtctc	840
gagatcttc tcgaagccctc tcggactgc ctgcaccaggc acctcgaggc cgacctggag	900
aatccgacgg ccaactcgat tgccattcgc tacttccact ccaactgcct ccacgtcgc	960
gccaaggccca agtactcgta gcacatattt ggccggccca tccgcctggc gcaggcatg	1020
cagctgcacg aggaggctgc cctcgagggg ctcgtccca tcgaggcaga gttccggcgt	1080
cgtctgtttt ggatcctgta cttggcgcac aagtcaaggc ctataactcaa caatggccc	1140
atcaccatcc acaagtactg ctgcacgcg ggcacatccca cgctataaccc gtcgggtatc	1200
gaggacgatg tcctgagcac ggcgtccgag ccgcggccga agagcttcat atccgttcc	1260
aacgcacaaatg tgcggctctg gcagtcgcg gtcgtttgc tgctggaaat ccgcgtgtcg	1320
caagatcaga tgcgtcgac cttcgaggg accatgcccc cgaaccatgt gtcgcctcc	1380
gccgacaggc agcatctcga ttctcttat gtccgcttca tcacgtgtt ggacgtatctc	1440
ccgcgttacc tcctcgatg cactctggcg atggcagcga tggcagaagg caacgggtt	1500
gccgagtcca agcagtacgt gatacgtgc atcaacactgc aggtgacgtt tcactgtctg	1560
cgcacatgttta ttacgcagaa attcgaagac ctctttatt ttgcctctgg cggtgagcag	1620
gctgtatctca gaaagtgcga gattgtgcga gacatgtga ggggtatgaa cgaggccc	1680
ttttggggcc tgcaggccaa tggcggccaa aacgttggaa agattgcct tatcggagct	1740
agtttgcgttgg ccatcatcca tcgcaaccag gattcaccct tggctacgcg agccaggagc	1800
gactttcccg tgctttggaa tatttcacg cggctggact cgaaggcgtc ggaccaactg	1860
agaataacgt ccactaccgt tggtggctaa	1890
<210> 3	
<211> 629	
<212> PRT	
<213> 里氏木霉	

<400> 3

Met Leu Arg Tyr Ser Pro Val Leu His Leu Asp Thr Leu Ser Leu Pro
 1 5 10 15

Pro Leu Thr Asn Ala Leu Pro Arg Pro Lys Cys Glu Tyr Leu Ser Ala
 20 25 30

Val Asp Ser Cys Thr His Cys Arg Asp Ala His Val Gln Cys Thr Phe
 35 40 45

Asp Leu Pro Leu Ala Arg Arg Gly Pro Lys Ala Arg Lys Lys Ser Asp
 50 55 60

Gln Pro Gly Gln Pro Pro Pro Asp Pro Ser Ser Leu Ser Thr Ala Ala
 65 70 75 80

Arg Pro Gly Gln Met Pro Pro Pro Leu Thr Phe Ser Gly Pro Ala Val
 85 90 95

Ala Ala Leu Gln Pro Phe Ala Ser Ser Leu Ser Pro Asp Ala Ala
 100 105 110

Trp Glu Pro Val Glu Pro Leu Ser Ile Asp Asn Gly Leu Pro Arg Gln
 115 120 125

Pro Leu Gly Asp Leu Pro Gly Leu Ser Thr Ile Gln Asn Ile Ser Thr
 130 135 140

Arg Gln Arg Trp Ile His Leu Ala Asn Ala Met Thr Leu Arg Asn Thr
 145 150 155 160

[0003] Thr Leu Glu Arg Val Ser Lys Arg Cys Ile Asp Leu Phe Phe Asp Tyr
 165 170 175

Leu Tyr Pro Leu Thr Pro Leu Val Tyr Glu Pro Ala Leu Arg Asp Val
 180 185 190

Leu Ala Tyr Ile Phe Ser Gln Pro Leu Pro Gly Val Asn Gln Pro Ser
 195 200 205

Pro Leu Ser Gln Leu Thr Pro Asp Pro Thr Thr Gly Thr Thr Pro Leu
 210 215 220

Asn Ala Ala Glu Ser Trp Ala Gly Phe Gly Gln Pro Ser Gly Ser Arg
 225 230 235 240

Thr Val Gly Ser Arg Leu Ala Pro Trp Ala Asp Ser Thr Phe Thr Leu
 245 250 255

Val Thr Ala Val Cys Ala Glu Ala Ala Phe Met Leu Pro Lys Asp Ile
 260 265 270

Phe Pro Glu Gly Glu Ser Val Ser Glu Ile Leu Leu Glu Ala Ser Arg
 275 280 285

Asp Cys Leu His Gln His Leu Glu Ala Asp Leu Glu Asn Pro Thr Ala
 290 295 300

Asn Ser Ile Ala Ile Arg Tyr Phe His Ser Asn Cys Leu His Ala Ala
 305 310 315 320

Gly Lys Pro Lys Tyr Ser Trp His Ile Phe Gly Glu Ala Ile Arg Leu
 325 330 335

Ala Gln Val Met Gln Leu His Glu Glu Ala Ala Leu Glu Gly Leu Val

	340	345	350
Pro Ile Glu Ala Glu Phe Arg Arg Arg Cys Phe Trp Ile Leu Tyr Leu			
355	360	365	
Gly Asp Lys Ser Ala Ala Ile Leu Asn Asn Arg Pro Ile Thr Ile His			
370	375	380	
Lys Tyr Cys Phe Asp Ala Gly Ile Thr Thr Leu Tyr Pro Ser Gly Ile			
385	390	395	400
Glu Asp Glu Phe Leu Ser Thr Ala Ser Glu Pro Pro Arg Lys Ser Phe			
405	410	415	
Ile Ser Gly Phe Asn Ala Asn Val Arg Leu Trp Gln Ser Ala Ala Asp			
420	425	430	
Leu Leu Leu Glu Ile Arg Val Leu Gln Asp Gln Met Met Gln His Phe			
435	440	445	
Arg Gly Thr Met Pro Pro Asn His Val Leu Pro Ser Ala Asp Arg Gln			
450	455	460	
His Leu Asp Ser Leu Tyr Val Arg Phe Ile Thr Cys Leu Asp Asp Leu			
465	470	475	480
Pro Pro Tyr Leu Gln Ser Cys Thr Leu Ala Met Ala Ala Met Ala Glu			
485	490	495	
Gly Asn Gly Ser Ala Glu Ser Lys Gln Tyr Val Ile Gln Cys Ile Asn			
500	505	510	
Leu Gln Val Thr Phe His Cys Leu Arg Met Val Ile Thr Gln Lys Phe			
[0004] 515	520	525	
Glu Asp Leu Ser Tyr Phe Ala Pro Gly Val Glu Gln Ala Asp Leu Arg			
530	535	540	
Lys Ser Glu Ile Val Arg Asp Met Leu Arg Val Met Asn Glu Ala Pro			
545	550	555	560
Phe Trp Gly Leu Gln Ala Asn Gly Glu Pro Asn Val Glu Lys Ile Arg			
565	570	575	
Leu Ile Gly Ala Ser Leu Leu Ala Ile Ile His Arg Asn Gln Asp Ser			
580	585	590	
Pro Leu Ala Thr Arg Ala Arg Ser Asp Phe Ser Val Leu Leu Asp Ile			
595	600	605	
Leu Thr Arg Leu Asp Ser Lys Ala Ser Asp Gln Leu Arg Asn Thr Ser			
610	615	620	
Thr Thr Val Val Gly			
625			
<210> 4			
<211> 2608			
<212> DNA			
<213> 里氏木霉			
<400> 4			
atgggctcag cagctccggc ccagggtct gttagtgcag ctgcaggcgg ccctccagct			60
gctggcgctg ggcgtggcgc tgtccacgcc ctcaccacct cgccccgagtc tgcctcgccc			120

[0005]	tcgcagcccc gctgccaaac cgccctcaacc acggccggc agaactcaact cgtgtcggt gcaaccttgt tccaccacca tcccagagge cgtctggtga gcagagectg cgaccgct cgccggcgca aggccaagggt cagtcttagcc cctttgtgt tgcttgcatc tctgtgtca ttgctccctcc tcctgctgt gctgatgctg ctgctccctcc tcctcctctt cctcccgctc tcctggtccc tggtccctgc tcttcataatg tccttactgc ccgtgtctcc tctcccggtt cccggtccccc ctcttcccggt cctcttctcc tgcgtgtctg tcatgcgtac aaagcataaca tacaatacat cagcatacat ggcaagcggtt gtgttgtgtt gagagttgtg tgtattgtat tgcaactgcct tcacaactcg ttcaatactgc tgcaagctca ccccaacacc gacctcgct tccatgtgc gctactcccc cgtcttacac ctggataactc tctccttgc accactgacc aatgtcttc cccgccccaaa gtgcgaggtac ctcaagcgctg tcgatagctg cacgcactgc cgcgatgccc acgtgcagtg cacttcgac ctgccccctgg cgcgacgcgg ccccaaagcg aggaagaaga ggcgaccagcc cggccagccg cctctgtatc cgagctcgct ctccaccgc gctcgaccccg gccagatgcc cccgcccgtg accttctccg gcccccgagt agccgcgt cagcccttcg cctcgctgc gctgtcgccc gacgcggctt gggagcccggt cgagccgc agcattgaca acggcctgcc cccggcagccg ctggggcacc tgcccgccct ctccaccate cagaacatct cgacgcgcca ggcgatggata cacctggcca acggcatgac gtcgcgcaac acgacgctag agcgcgtctc gaagcgatgt atcgacctct tcttcgacta cctcttacccc ctcacccccc tggtgtacga gcccggccctc cgggacgtgc tcgcatacat ctctccag cccttgctg gctcaacca accatcgccg ctgtcacagc tcacgccaga cccgaccacc ggcaccaccc ccctcaacgc tgccgagtc tggggccgct ttggccagcc cagcggctcg cgaaccgtcg gcagcaggctt ggtccctgg gccgactcga ctttacccct ggtcacggcc gtctgcgcag aggcagcatt catgetaccc aaggacattt tccccgaagg agaatccgtc tctgagatct tgctcgaaac ctctcggttgc accacctcgaa gggccgacccgt gagaatccga cggccaaactc gattggcatt cgctacttcc actccaaactg cctccacgt gccccggaa ccaagtactc gtggcacata tttggcgagg ccatccgcct ggccgaggc atgcagctgc acgaggaggc tgccctcgag gggctcgatcc ccatcgaggc agagttccgc cgctcgatgt tttggatct gtacttggc gacaagtcgac ccgtataact caacaatcg cccatcacca tccacaagta ctgcttcgac gcccgcataa ccacgtataa cccgtcggtt atcgaggacg agttcctgag cacggcgtcc gagccggccc ggaagagctt catatccgc ttcaacgcaa atgtcggttgc ctggcagtc gggctgtatt tgctgttgc aatccgcgt ctgcaagatc agatgatgca gcacttcga gggaccatgc ccccgaaacca tggctgtccc tccggccaca ggcagcatct cgattctctc tatgtccgt tcacccatcg ctggacgt ctcccgctt acctccagtc gtgcactctg gcatggcagc cgtatggcaga aggcaacggg tctgcccgtt ccaagcagta cgtgatgtacag tgcatcaacc tgcatggatc gtttactgt ctgcgcacggta taattacgca gaaattcgaa gacctctttt attttgcctc tggctgttgc caggctgtac tcagaaatgc ggagattgtg cggagatgc tgagggtgtt gaacgaggcg ccctttggg gcctgcaggc caatggcgg cccaaacgtga gtcgtttccct tggctgttgc ctttctgca cacccttttc ttgcacggacc cccctctctt ctatccatcc ctgcggatat gtatccatc aaggccggc actgttgc tttttgttgc tttttgttgc tttttgttgc gcagggttggaa aagattcgcc ttatcggttgc tagtttgc tttttgttgc tttttgttgc ggattccaccc ttggctacgc gagccaggag cggactttcc ttggctttgg atattctac ggccgtggac tcgaaggcgt cggactaa	180 240 300 360 420 480 540 600 660 720 780 840 900 960 1020 1080 1140 1200 1260 1320 1380 1440 1500 1560 1620 1680 1740 1800 1860 1920 1980 2040 2100 2160 2220 2280 2340 2400 2460 2520 2580 2608
--------	---	--

<210> 5

<211> 2037

<212>	DNA		
<213>	里氏木霉		
<400>	5		
	atgggctcag cagctccggc ccagggtct gtagctgcag ctgcaggcgg ccctccagct	60	
	gctggcgctg gcgcgtggcgc tgtccacgcc ctcaccacct cgccccgagtc tgcctcgcc	120	
	tgcagcccc gctgcacaac cgcctcaacc acgcccggc agaactcaact cgtgtcgct	180	
	gcaacacctgt tccaccacca tcccagaggg cgtctggta gcagagcctg cgaccgctgc	240	
	cggccggcga aggccaagtgc cgagttaccc agcgctgtcg atagctgcac gcactgccc	300	
	gatgcccacg tgcagtgcac ttgcacccctg cccctggcgc gacgcccggcc caaagcgagg	360	
	aagaagagcg accagccgg ccagccgcct cctgatccga gctcgctc caccgcggct	420	
	cgaccggcc agatgccccc ggcgcgtgacc ttctccggcc cgcgcgttagc cgcgcgtgc	480	
	cccttcgcct cgtcgctgcgt gtgcggcgc gggcctggg agccgcgtga gccgcgtc	540	
	attgacaacg gcctcccccg gcagccgcgt ggcgcacctgc cggcctc caccatccag	600	
	aacatctcga cgcgccagcg atggatacac ctggccaacg ccatgacgct ggcacacacg	660	
	acgctagagc gcgtctcgaa gcgatgtatc gaccttttct tcgactaccc ctacccctc	720	
	accccccctgg tgtacgagcc ggccctccgg gacgtgtcg catacattt ctccca	780	
	ttgcctggcg tcaaccaacc atgcgcgtgc tcacagctca cgcgcgtaccc gaccacggc	840	
	accacccccc tcaacgcgtc cgagtcgtgg gccggctttg gccagcccg cggcgtcgt	900	
	accgtcgca gcaggctggc tccctggcc gactgacact tcaccctgtt cacggccgtc	960	
	tgcgcagagg cagcattcat gctaccaag gacatccc ccgaaggaga atccgtct	1020	
	gagatcttc tcgaagccctc tcgggactgc ctgcaccaggc acctcgaggg cgacctggag	1080	
[0006]	aatccgcacgg ccaactcgat tgccattcgc tacttcact ccaactgcct ccacgcgtcg	1140	
	gsgaagccca agtactcgta gcacatattt ggcgaggcca tccgcctggc gcaggcatg	1200	
	cagctgcacg aggaggctgc cctcgagggg ctgcgtccca tcgaggcaga gttccggcgt	1260	
	cgctgtttt ggatcctgta cttggcgcac aagtcaagccg ctatactcaa caatccggcc	1320	
	atcaccatcc acaagttactg ctgcgacgcc ggcacatccca cgctataccgc gtcgggtatc	1380	
	gaggacgagt tcctgagcac ggcgtccgag ccgcggccgaa agagttcat atccgtctc	1440	
	aacgc当地 tgccgc当地 gcagttccgcg gctgattgc tgctggaaat ccgcgtgc当地	1500	
	caagatcaga tgc当地 gagccatccca cgc当地 accatgt gtc当地 ccgc当地	1560	
	gccgc当地 agc当地 tc当地 ttctctctat gtc当地 tc当地 ggacgatc当地	1620	
	ccgc当地 gta local tccagtc当地 cacttc当地 gatgc当地 tggc当地 aggc当地	1680	
	gccgagtc当地 agc当地 agtactgc当地 atcaacgc当地 aggtgacgat tcaactgc当地	1740	
	cgc当地 gta local ttacgc当地 attc当地 gagac ct当地 tt当地 tt当地 cgtt当地 gagc当地	1800	
	gctgatc当地 gaaatgc当地 gatgtgc当地 gacatgc当地 gggatgatgaa cgaggccccc	1860	
	ttttggcc当地 tgc当地 gagccaa tggc当地 gagccaa aacgtt当地 agatgc当地 tatcg当地 gagct	1920	
	agtttgc当地 ggcc当地 catcatccaa tc当地 gagccagg gattcaccct tggctacgc当地 agccaggagc	1980	
	gactttccg tgctttgga tatttc当地 acg cggctggact cgaaggc当地 ggactaa	2037	
<210>	6		
<211>	678		
<212>	PRT		
<213>	里氏木霉		
<400>	6		

Met Gly Ser Ala Ala Pro Ala Gln Gly Ser Val Ala Ala Ala Gly

1 5 10 15

Gly Pro Pro Ala Ala Gly Ala Gly Ala Gly Ala Val His Ala Leu Thr
 20 25 30

Thr Ser Pro Glu Ser Ala Ser Ala Ser Gln Pro Gly Ser Pro Thr Ala
 35 40 45

Ser Thr Thr Pro Pro Gln Asn Ser Leu Val Ser Ala Ala Thr Ser Phe
 50 55 60

His His His Pro Arg Gly Arg Leu Val Ser Arg Ala Cys Asp Arg Cys
 65 70 75 80

Arg Arg Arg Lys Ala Lys Cys Glu Tyr Leu Ser Ala Val Asp Ser Cys
 85 90 95

Thr His Cys Arg Asp Ala His Val Gln Cys Thr Phe Asp Leu Pro Leu
 100 105 110

Ala Arg Arg Gly Pro Lys Ala Arg Lys Ser Asp Gln Pro Gly Gln
 115 120 125

Pro Pro Pro Asp Pro Ser Ser Leu Ser Thr Ala Ala Arg Pro Gly Gln
 130 135 140

Met Pro Pro Pro Leu Thr Phe Ser Gly Pro Ala Val Ala Ala Leu Gln
 145 150 155 160

Pro Phe Ala Ser Ser Ser Leu Ser Pro Asp Ala Ala Trp Glu Pro Val
 165 170 175

Glu Pro Leu Ser Ile Asp Asn Gly Leu Pro Arg Gln Pro Leu Gly Asp
 180 185 190

[0007] Leu Pro Gly Leu Ser Thr Ile Gln Asn Ile Ser Thr Arg Gln Arg Trp
 195 200 205

Ile His Leu Ala Asn Ala Met Thr Leu Arg Asn Thr Thr Leu Glu Arg
 210 215 220

Val Ser Lys Arg Cys Ile Asp Leu Phe Phe Asp Tyr Leu Tyr Pro Leu
 225 230 235 240

Thr Pro Leu Val Tyr Glu Pro Ala Leu Arg Asp Val Leu Ala Tyr Ile
 245 250 255

Phe Ser Gln Pro Leu Pro Gly Val Asn Gln Pro Ser Pro Leu Ser Gln
 260 265 270

Leu Thr Pro Asp Pro Thr Thr Gly Thr Thr Pro Leu Asn Ala Ala Glu
 275 280 285

Ser Trp Ala Gly Phe Gly Gln Pro Ser Gly Ser Arg Thr Val Gly Ser
 290 295 300

Arg Leu Ala Pro Trp Ala Asp Ser Thr Phe Thr Leu Val Thr Ala Val
 305 310 315 320

Cys Ala Glu Ala Ala Phe Met Leu Pro Lys Asp Ile Phe Pro Glu Gly
 325 330 335

Glu Ser Val Ser Glu Ile Leu Leu Glu Ala Ser Arg Asp Cys Leu His
 340 345 350

Gln His Leu Glu Ala Asp Leu Glu Asn Pro Thr Ala Asn Ser Ile Ala
 355 360 365

Ile Arg Tyr Phe His Ser Asn Cys Leu His Ala Ala Gly Lys Pro Lys
 370 375 380
 Tyr Ser Trp His Ile Phe Gly Glu Ala Ile Arg Leu Ala Gln Val Met
 385 390 395 400
 Gln Leu His Glu Glu Ala Ala Leu Glu Gly Leu Val Pro Ile Glu Ala
 405 410 415
 Glu Phe Arg Arg Arg Cys Phe Trp Ile Leu Tyr Leu Gly Asp Lys Ser
 420 425 430
 Ala Ala Ile Leu Asn Asn Arg Pro Ile Thr Ile His Lys Tyr Cys Phe
 435 440 445
 Asp Ala Gly Ile Thr Thr Leu Tyr Pro Ser Gly Ile Glu Asp Glu Phe
 450 455 460
 Leu Ser Thr Ala Ser Glu Pro Pro Arg Lys Ser Phe Ile Ser Gly Phe
 465 470 475 480
 Asn Ala Asn Val Arg Leu Trp Gln Ser Ala Ala Asp Leu Leu Leu Glu
 485 490 495
 Ile Arg Val Leu Gln Asp Gln Met Met Gln His Phe Arg Gly Thr Met
 500 505 510
 Pro Pro Asn His Val Leu Pro Ser Ala Asp Arg Gln His Leu Asp Ser
 515 520 525
 Leu Tyr Val Arg Phe Ile Thr Cys Leu Asp Asp Leu Pro Pro Tyr Leu
 530 535 540
 [0008] Gln Ser Cys Thr Leu Ala Met Ala Ala Met Ala Glu Gly Asn Gly Ser
 545 550 555 560
 Ala Glu Ser Lys Gln Tyr Val Ile Gln Cys Ile Asn Leu Gln Val Thr
 565 570 575
 Phe His Cys Leu Arg Met Val Ile Thr Gln Lys Phe Glu Asp Leu Ser
 580 585 590
 Tyr Phe Ala Pro Gly Val Glu Gln Ala Asp Leu Arg Lys Ser Glu Ile
 595 600 605
 Val Arg Asp Met Leu Arg Val Met Asn Glu Ala Pro Phe Trp Gly Leu
 610 615 620
 Gln Ala Asn Gly Glu Pro Asn Val Glu Lys Ile Arg Leu Ile Gly Ala
 625 630 635 640
 Ser Leu Leu Ala Ile Ile His Arg Asn Gln Asp Ser Pro Leu Ala Thr
 645 650 655
 Arg Ala Arg Ser Asp Phe Ser Val Leu Leu Asp Ile Leu Thr Arg Leu
 660 665 670
 Asp Ser Lys Ala Ser Asp
 675
 <210> 7
 <211> 2005
 <212> DNA
 <213> 里氏木霉

<400>	7															
	atgctgcgt actccccgt cttacacctg gatactctt ccttgccacc actgaccaat	60														
	gctttcccc gcccaaagtg cgagtaacctc agcgctgtcg atagctgcac gcactgcccgc	120														
	gatgccacg tgcatgtcac ttgcacccgt cccctggcgc gacgcccggcc caaagcgagg	180														
	aagaagagcg accagccgg ccagccgcct cctgatccga gctgcttc caccgcggct	240														
	cgaccggcc agatgccgc gccgctgacc ttctccggcc ccgcagtagc cgccgtgcag	300														
	cccttcgcct cgtcgctgct gtcgcccac gggcctggg agcccgatcgaa gccgctcagc	360														
	attgacaacg gcctgcccc gcagccgctg ggcgacactc cccgccttc caccatccag	420														
	aacatctcgaa cgcgcacgatggatacac ctggccaacg ccatgacgct ggcacacacg	480														
	acgcttagagc gcgatcgaa gcgatgtatc gaccttttctc tgcactacccatccatccat	540														
	accccccgtt tgtacgagcc ggcctccgg gacgtgtcg catacatctt ctcccagccc	600														
	ttgcctggcg tcaaccaacc atcgccgctg tcacagctca cggccagaccc gaccaccggc	660														
	accacccccc tcaacgcgtc cgagtcgtgg gccggctttg gccagcccg cggctcgca	720														
	accgtcgca gcaggctggc tccctggcc gactcgaccc tcaaccctggt cacggccgtc	780														
	tgcgcagagg cagcattcat gctaccaag gacatccatccatccatccatccatccatccat	840														
	gagatcttgc tcaaggccctc tcggactgc ctgcaccagg acctcgaggc cgacctggag	900														
	aatccgacgg ccaactcgat tgccattcgat tacttccact ccaactgcct ccacgcgtcg	960														
	ggaaagccca agtactcgat gcacatattt ggcgaggcca tccgcctggc gcaggatcgat	1020														
	cagctgcacg aggaggctgc cctcgagggg ctcgtccca tcgaggcaga gttccggcgt	1080														
	cgctgtttt ggatctgtt cttggcgcac aagtcagccg ctatactcaa caatccggcc	1140														
[0009]	atcaccatcc acaagttactg ctgcacgccc ggcacatccatccatccatccatccatccat	1200														
	gaggacgagt tcctgagcac ggcgtccgag ccgcggccgaa agagttcat atccggcttc	1260														
	aacgcaaatg tgccgtctg gcagtcgcgc gctgatttgc tgctggaaat ccgcgtgtcg	1320														
	caagatcaga tgcgtcgatccatccatccatccatccatccatccatccatccatccatccat	1380														
	ccgcacaggc agcatctcgat ttctcttat gtccgttca tcacgtgtttt ggacgtatctc	1440														
	ccgcgttacc tccagtcgtt cacttcggcg atggcagcga tggcagaagg caacgggtct	1500														
	ccgcgttacc tccagtcgtt cacttcggcg atggcagcga tggcagaagg caacgggtct	1560														
	ccgcgttacc tccagtcgtt cacttcggcg atggcagcga tggcagaagg caacgggtct	1620														
	ccgcgttacc tccagtcgtt cacttcggcg atggcagcga tggcagaagg caacgggtct	1680														
	ttttggggcc tgcaggccaa tggcggccaa aacgttgatc gtttccctgt ctcttctt	1740														
	ttctgcacac cttttcttc gacgacccccc cctctctttt tatatccctg cggatatgtat	1800														
	tatcatcaag cctcggtact tggcgttacat ctgttctgtat tatgttgcgtt ggttgcgtca	1860														
	ggttggaaaat attcgcccta tcggagctt tttgtggcc atcatccatccatccatccat	1920														
	ttcaccccttgc tgcacgccc ccaggaggcga cttttccgtt cttttggata ttctcacgcg	1980														
	gctggactcg aaggcgtcg actaa	2005														
<210>	8															
<211>	618															
<212>	PRT															
<213>	里氏木霉															
<400>	8															
Met	Leu	Arg	Tyr	Ser	Pro	Val	Leu	His	Leu	Asp	Thr	Leu	Ser	Leu	Pro	
1																
	5															15
Pro	Leu	Thr	Asn	Ala	Leu	Pro	Arg	Pro	Lys	Cys	Glu	Tyr	Leu	Ser	Ala	
	20															30

Val Asp Ser Cys Thr His Cys Arg Asp Ala His Val Gln Cys Thr Phe
 35 40 45
 Asp Leu Pro Leu Ala Arg Arg Gly Pro Lys Ala Arg Lys Ser Asp
 50 55 60
 Gln Pro Gly Gln Pro Pro Pro Asp Pro Ser Ser Leu Ser Thr Ala Ala
 65 70 75 80
 Arg Pro Gly Gln Met Pro Pro Pro Leu Thr Phe Ser Gly Pro Ala Val
 85 90 95
 Ala Ala Leu Gln Pro Phe Ala Ser Ser Leu Ser Pro Asp Ala Ala
 100 105 110
 Trp Glu Pro Val Glu Pro Leu Ser Ile Asp Asn Gly Leu Pro Arg Gln
 115 120 125
 Pro Leu Gly Asp Leu Pro Gly Leu Ser Thr Ile Gln Asn Ile Ser Thr
 130 135 140
 Arg Gln Arg Trp Ile His Leu Ala Asn Ala Met Thr Leu Arg Asn Thr
 145 150 155 160
 Thr Leu Glu Arg Val Ser Lys Arg Cys Ile Asp Leu Phe Phe Asp Tyr
 165 170 175
 Leu Tyr Pro Leu Thr Pro Leu Val Tyr Glu Pro Ala Leu Arg Asp Val
 180 185 190
 Leu Ala Tyr Ile Phe Ser Gln Pro Leu Pro Gly Val Asn Gln Pro Ser
 195 200 205
[0010] Pro Leu Ser Gln Leu Thr Pro Asp Pro Thr Thr Gly Thr Thr Pro Leu
 210 215 220
 Asn Ala Ala Glu Ser Trp Ala Gly Phe Gly Gln Pro Ser Gly Ser Arg
 225 230 235 240
 Thr Val Gly Ser Arg Leu Ala Pro Trp Ala Asp Ser Thr Phe Thr Leu
 245 250 255
 Val Thr Ala Val Cys Ala Glu Ala Ala Phe Met Leu Pro Lys Asp Ile
 260 265 270
 Phe Pro Glu Gly Glu Ser Val Ser Glu Ile Leu Leu Glu Ala Ser Arg
 275 280 285
 Asp Cys Leu His Gln His Leu Glu Ala Asp Leu Glu Asn Pro Thr Ala
 290 295 300
 Asn Ser Ile Ala Ile Arg Tyr Phe His Ser Asn Cys Leu His Ala Ala
 305 310 315 320

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

	370	375	380
Lys Tyr Cys Phe Asp Ala Gly Ile Thr Thr Leu Tyr Pro Ser Gly Ile			
385	390	395	400
Glu Asp Glu Phe Leu Ser Thr Ala Ser Glu Pro Pro Arg Lys Ser Phe			
405	410	415	
Ile Ser Gly Phe Asn Ala Asn Val Arg Leu Trp Gln Ser Ala Ala Asp			
420	425	430	
Leu Leu Leu Glu Ile Arg Val Leu Gln Asp Gln Met Met Gln His Phe			
435	440	445	
Arg Gly Thr Met Pro Pro Asn His Val Leu Pro Ser Ala Asp Arg Gln			
450	455	460	
His Leu Asp Ser Leu Tyr Val Arg Phe Ile Thr Cys Leu Asp Asp Leu			
465	470	475	480
Pro Pro Tyr Leu Gln Ser Cys Thr Leu Ala Met Ala Ala Met Ala Glu			
485	490	495	
Gly Asn Gly Ser Ala Glu Ser Lys Gln Tyr Val Ile Gln Cys Ile Asn			
500	505	510	
Leu Gln Val Thr Phe His Cys Leu Arg Met Val Ile Thr Gln Lys Phe			
515	520	525	
Glu Asp Leu Ser Tyr Phe Ala Pro Gly Val Glu Gln Ala Asp Leu Arg			
530	535	540	
Lys Ser Glu Ile Val Arg Asp Met Leu Arg Val Met Asn Glu Ala Pro			
545	550	555	560
Phe Trp Gly Leu Gln Ala Asn Gly Glu Pro Asn Val Glu Lys Ile Arg			
565	570	575	
Leu Ile Gly Ala Ser Leu Leu Ala Ile Ile His Arg Asn Gln Asp Ser			
580	585	590	
Pro Leu Ala Thr Arg Ala Arg Ser Asp Phe Ser Val Leu Leu Asp Ile			
595	600	605	
Leu Thr Arg Leu Asp Ser Lys Ala Ser Asp			
610	615		
<210> 9			
<211> 2641			
<212> DNA			
<213> 里氏木霉			
<400> 9			
atgggctcag cagctccggc ccagggctct gtagctgcag ctgcaggcgg ccctccagct		60	
gctggcgctg gcgctggcgc tgtccacgcc ctcaccacct cgccccgagtc tgcctcgcc		120	
tgcagcccg gtcgccaac cgccctcaacc acgccggcgc agaactcact cgtgtcggct		180	
gcaacctcggt tccaccacca tcccagaggc cgtctggta gcagagcctg cgaccgctgc		240	
cggccggcga aggccaaggt cagtctagcc cctttgtgt tgcttgcatc tctgttgtca		300	
ttgtcttcc tcctgtgtc gctgtatgtc ctgtcttcc tccttccttc cctcccccgtc		360	
tcctggtccc tggtccctgc ttcatatgc tccttactgc ccgtgtctcc tctcccccgtt		420	
cccggtttccc tcctcccggt cctcttc tgcgtgtctg tcatgcgtac aaagcataaca		480	

	tacaatacat cagcatacat ggcaagcggtt gtgttgtgtt gagagtttg ttttttttat	540
	tgcactgcct tcacaactcg ttccatactgc tgccggctca ccccaacacc gacctcgct	600
	tccatgtgc gctactcccc cgttcttacac ctggatactc ttccttgcc accactgacc	660
	aatgcttttc cccgccccaa gtgcgagttac ctcagcgctg tcgatagctg cacgactgc	720
	cgcgatgccc acgtgcaggc cacttgcac ctggccctgg cgccgacgcgg ccccaaagcg	780
	aggaagaaga ggcggccggc cggccggccg cttccgtatc cgagctcgct ctccaccgcg	840
	gctcgaccgg gccagatgcc ggcggccgtg accttctcg gccccgcagt agccgcgtg	900
	cagcccttcg cttcgctgc gctgtcgccc gacgcggcctt gggagccgt cgagccgc	960
	agcattgaca acggcctgcc cccggcggccg ctggggcacc tggccggccttccaccate	1020
	cagaacatct cggatggata cacctggcca acgcccatac gctgcgcaac	1080
	acgacgcttag agcgcgtctc gaagcgatgt atcgaccctt tttcgacta cctctacccc	1140
	ctcacccccc tgggtacga gccggccctc cgggacgtgc tcgatccat cttctccag	1200
	cccttgcgtg gctcaacca accatcgccg ctgtcacagc tcacggcaga cccgaccacc	1260
	ggcaccaccc ccctcaacgc tgccgagtcg tggccggctt tggccagcc cagcgctcg	1320
	cgaacctcg gcagcaggctt ggtccctgg gccgactcga cttcacccctt gtcacggcc	1380
	gtctgcgcaag aggacgattt catgtaccc aaggacattt tccccgaagg agaattcg	1440
	tctgagatct tgctcaagc ctctcggttgc tggctgcacc agcacctcga gggccacctg	1500
	gagaatccga cggccaaactc gattggcatt cgctacttcc actccaaactg cttccacgt	1560
	gccccggaaactc ccaagttactc gtggcacata ttggcgagg ccatccgc ggcgcaggc	1620
	atgcagctgc acgaggaggc tggctcgag gggctcgcc ccatcgaggc agagttccgc	1680
	cgtcgttgc tttggatct gtacttggc gacaagtcg ccgtataact caacaatcgg	1740
	cccatcacca tccacaagta ctgttcgac gcccgcata ccaacgtata cccgtcggt	1800
	atcgaggacg agttcctgag cacggcgtcc gagccggcc ggaagagctt catatccgg	1860
	ttcaacgcaa atgtcggtt ctggcgttcc gggctgtatt tgctgttgc aatccgcgt	1920
	ctgcaagatc agatgtatcgc gactttcga gggaccatgc cccgaaacca tgtgtcgcc	1980
	tccggcaca ggcagcatct cgattcttc tatgtccgt tcatcacctt cttggacgt	2040
	ctccggcgtt acctccagtc gtgcactctg gcatggcag cgtatggcaga aggcaacggg	2100
	tctggcgagt ccaagcagta cgtgatacag tgcataacc tgcaggttgc gtttactgt	2160
	ctgcgcatttgg taattacgca gaaattcgaa gacctctttt attttgcctt tggcggtt	2220
	caggctgttgc tcaaaaatgcg ggagattgtt cggatgttgc tgagggttgc gaaaggcgc	2280
	cccttttggg gcctgcaggc caatggcgg cccaaacgtga gtcgtttcct tgcgttttct	2340
	cttttgcata cacccttttgc ttgcacgacc cccctcttctt ctatccatcc ctgcggat	2400
	gtatatcate aagcctcgac acttgggttgc aatctgttctt gattatgttgc tctggatgt	2460
	gcagggttgc aagattcgcc ttatcggttgc tagtttgc gccatcatcc atcgcaacca	2520
	ggattcaccc ttggctacgc gagccaggag cggatggcgtt gtcgttttgg atattctac	2580
	ggggctggac tcgaaggcgtt cggaccaactt gaggaaatcg tccactaccg ttgttggcta	2640
	a	2641
<210>	10	
<211>	689	
<212>	PRT	
<213>	里氏木霉	
<400>	10	

Met Gly Ser Ala Ala Pro Ala Gln Gly Ser Val Ala Ala Ala Gly

1 5 10 15

Gly Pro Pro Ala Ala Gly Ala Gly Ala Gly Ala Val His Ala Leu Thr
 20 25 30
 Thr Ser Pro Glu Ser Ala Ser Ala Ser Gln Pro Gly Ser Pro Thr Ala
 35 40 45
 Ser Thr Thr Pro Pro Gln Asn Ser Leu Val Ser Ala Ala Thr Ser Phe
 50 55 60
 His His His Pro Arg Gly Arg Leu Val Ser Arg Ala Cys Asp Arg Cys
 65 70 75 80
 Arg Arg Arg Lys Ala Lys Cys Glu Tyr Leu Ser Ala Val Asp Ser Cys
 85 90 95
 Thr His Cys Arg Asp Ala His Val Gln Cys Thr Phe Asp Leu Pro Leu
 100 105 110
 Ala Arg Arg Gly Pro Lys Ala Arg Lys Ser Asp Gln Pro Gly Gln
 115 120 125
 Pro Pro Pro Asp Pro Ser Ser Leu Ser Thr Ala Ala Arg Pro Gly Gln
 130 135 140
 Met Pro Pro Pro Leu Thr Phe Ser Gly Pro Ala Val Ala Ala Leu Gln
 145 150 155 160
 Pro Phe Ala Ser Ser Ser Leu Ser Pro Asp Ala Ala Trp Glu Pro Val
 165 170 175
 Glu Pro Leu Ser Ile Asp Asn Gly Leu Pro Arg Gln Pro Leu Gly Asp
 180 185 190
 [0013] Leu Pro Gly Leu Ser Thr Ile Gln Asn Ile Ser Thr Arg Gln Arg Trp
 195 200 205
 Ile His Leu Ala Asn Ala Met Thr Leu Arg Asn Thr Thr Leu Glu Arg
 210 215 220
 Val Ser Lys Arg Cys Ile Asp Leu Phe Phe Asp Tyr Leu Tyr Pro Leu
 225 230 235 240
 Thr Pro Leu Val Tyr Glu Pro Ala Leu Arg Asp Val Leu Ala Tyr Ile
 245 250 255
 Phe Ser Gln Pro Leu Pro Gly Val Asn Gln Pro Ser Pro Leu Ser Gln
 260 265 270
 Leu Thr Pro Asp Pro Thr Thr Gly Thr Thr Pro Leu Asn Ala Ala Glu
 275 280 285
 Ser Trp Ala Gly Phe Gly Gln Pro Ser Gly Ser Arg Thr Val Gly Ser
 290 295 300
 Arg Leu Ala Pro Trp Ala Asp Ser Thr Phe Thr Leu Val Thr Ala Val
 305 310 315 320
 Cys Ala Glu Ala Ala Phe Met Leu Pro Lys Asp Ile Phe Pro Glu Gly
 325 330 335
 Glu Ser Val Ser Glu Ile Leu Leu Glu Ala Ser Arg Asp Cys Leu His
 340 345 350
 Gln His Leu Glu Ala Asp Leu Glu Asn Pro Thr Ala Asn Ser Ile Ala
 355 360 365

Ile Arg Tyr Phe His Ser Asn Cys Leu His Ala Ala Gly Lys Pro Lys
 370 375 380
 Tyr Ser Trp His Ile Phe Gly Glu Ala Ile Arg Leu Ala Gln Val Met
 385 390 395 400
 Gln Leu His Glu Glu Ala Ala Leu Glu Gly Leu Val Pro Ile Glu Ala
 405 410 415
 Glu Phe Arg Arg Cys Phe Trp Ile Leu Tyr Leu Gly Asp Lys Ser
 420 425 430
 Ala Ala Ile Leu Asn Asn Arg Pro Ile Thr Ile His Lys Tyr Cys Phe
 435 440 445
 Asp Ala Gly Ile Thr Thr Leu Tyr Pro Ser Gly Ile Glu Asp Glu Phe
 450 455 460
 Leu Ser Thr Ala Ser Glu Pro Pro Arg Lys Ser Phe Ile Ser Gly Phe
 465 470 475 480
 Asn Ala Asn Val Arg Leu Trp Gln Ser Ala Ala Asp Leu Leu Glu
 485 490 495
 Ile Arg Val Leu Gln Asp Gln Met Met Gln His Phe Arg Gly Thr Met
 500 505 510
 Pro Pro Asn His Val Leu Pro Ser Ala Asp Arg Gln His Leu Asp Ser
 515 520 525
 Leu Tyr Val Arg Phe Ile Thr Cys Leu Asp Asp Leu Pro Pro Tyr Leu
 530 535 540
 [0014] Gln Ser Cys Thr Leu Ala Met Ala Ala Met Ala Glu Gly Asn Gly Ser
 545 550 555 560
 Ala Glu Ser Lys Gln Tyr Val Ile Gln Cys Ile Asn Leu Gln Val Thr
 565 570 575
 Phe His Cys Leu Arg Met Val Ile Thr Gln Lys Phe Glu Asp Leu Ser
 580 585 590
 Tyr Phe Ala Pro Gly Val Glu Gln Ala Asp Leu Arg Lys Ser Glu Ile
 595 600 605
 Val Arg Asp Met Leu Arg Val Met Asn Glu Ala Pro Phe Trp Gly Leu
 610 615 620
 Gln Ala Asn Gly Glu Pro Asn Val Glu Lys Ile Arg Leu Ile Gly Ala
 625 630 635 640
 Ser Leu Leu Ala Ile Ile His Arg Asn Gln Asp Ser Pro Leu Ala Thr
 645 650 655
 Arg Ala Arg Ser Asp Phe Ser Val Leu Leu Asp Ile Leu Thr Arg Leu
 660 665 670
 Asp Ser Lys Ala Ser Asp Gln Leu Arg Asn Thr Ser Thr Thr Val Val
 675 680 685
 Gly

<210> 11

<211> 2885

<212>	DNA	
<213>	里氏木霉	
<400>	11	
[0015]		
	atggccacag cggccgcggc agcagctggc ggcgcggcgg ttgctgcggg tgcagacaca	60
	ggtcgttga gtcccgtccc gtccgctcgc gttccctccc agctgccagc ccgcgtgggt	120
	ggcactggaa cgcaatcagt gcagtgccgc ccccccaact aacgctgccc	180
	cccggtggctc ctccggccaca caggcgctgc aggctccagc tctacaggcc ctccaggcct	240
	tccaggggctt ccaggcaccc ggacaggcgc cgtaggcgtat ggctcagcag ctccggccca	300
	gggcgtctgta gctgcagctg caggcgccccc tecagctgt ggctcggcgt ctggcgctgt	360
	ccacgcctc accacctcgc ccgagtctgc ctccggctcg cagccggct cggccaaccgc	420
	ctcaaccacg cggccgcaga actcactctg gtcggctgca acctcgttcc accaccatcc	480
	cagaggccgt ctggtgagca gagecgtgcga ccgcgtccgc cggcgcagg ccaaggtag	540
	tctagccctt ttgctgttgc ttgcattctt gttgttattt ctccctctt tgcgtgttct	600
	gatgctgttgc ctccctctt ctccctctt cccctcttcc tggcccttgg tccctgttct	660
	tcatatgttcc ttactgcctt tgcgttctt ccccttccctt gttcccttcc ctccctgttct	720
	cttctcttgc gtgtctgtca tgcttacaaa gcatatata aatacatcag catacatggc	780
	aagcgttgtt tggttgttgc agttgtgtt attgtatttgc actgccttca caactcggtt	840
	atactgttgc agccttaccc caacaccgc acgttcttcc atgctgttgc actccccgt	900
	cttacacctg gatactctt ctttgcacc actgaccaat gctttcccc gcccaaagtgc	960
	cgagtacctc agcgctgtcg atagctgcac gcactgcgc gatgcccacg tgcgtgttgc	1020
	tttcgacctg cccctggcgc gacgcccggc caaagcgagg aagaagagcg accagccgg	1080
	ccagccgcct cctgatccga gtcgttgc caccgcggct cggccggcc agatggcc	1140
	gccgctgacc ttctccggcc ccgcgttgc cgcgttgc cccctcgcc cgtgttgc	1200
	gtcgccgcac gcccgttgc cccctcgcc caccatccag aacatcttgc cgcgttgc	1260
	gcagccgtg ggcgacctgc cccctcgcc caccatccag aacatcttgc cgcgttgc	1320
	atggatacac ctggccaacg ccatgttgc ggcacacacg acgttagagc ggcgttgc	1380
	gcatgttgc gaccttttctt tgcacttgc ctacccttcc acccccttgg tgcgttgc	1440
	ggccctccgg gacgtgttgc catacatctt ctcccttgc ttgcctggc tcaaccaccc	1500
	atcgccgttgc tcacagcttca cggccggcc gaccaccggc accacccccc tcaacgttgc	1560
	cgagtgttgc gccggctttt gccagcccg cggcttgc caccgttgc gcagggttgc	1620
	tccctggcc gacttgcacct tcacccttgc caccgttgc tgcgttgc caccatccat	1680
	gttacccaaag gacattttcc cggaggaga atccgttgc gagattttgc tgcgttgc	1740
	tccggacttgc ctgcaccaggc accttgcaggc cgcgttgc aatccgttgc ccaactcgat	1800
	tgcctatcg ttttttttttcc ccaacttgc caccgttgc gggagccca agtactcg	1860
	gcacatattt ggcgaggcca tccgcgttgc gcagggttgc caccgttgc caccatccat	1920
	cctcgagggg ctcgttgc tgcgttgc gttccggcgt cgcgttgc tgcgttgc	1980
	cttggccgac aagtgcgttgc ctataacttca caatccggcc accacccatcc acaagacttgc	2040
	cttcgaccc ggcgttgc caccgttgc caccgttgc tgcgttgc caccatccat	2100
	ggcgttgc caccgttgc caccgttgc tgcgttgc caccgttgc tgcgttgc	2160
	gcagtccgttgc gtcgttgc tgcgttgc caccgttgc tgcgttgc caccatccat	2220
	cttgcgttgc accatccggcc caccgttgc tgcgttgc caccgttgc tgcgttgc	2280
	tttcgttgc tgcgttgc caccgttgc tgcgttgc caccgttgc tgcgttgc	2340
	cacttgcgttgc atggcgttgc tggcgttgc caacgggttgc ggcgttgc aaccgttgc	2400
	gatacgttgc atcaaccgttgc aggtgttgc tgcgttgc caccgttgc tgcgttgc	2460

	atcgaagac ctcttttatt ttgctcctgg cggtgagcag gctgatctca gaaagtgcga	2520
	gattgtgcga gacatgctga gggtgatgaa cgaggcgccc ttttggggcc tgcaggccaa	2580
	tggcgagcca aacgtgagtc gtttccctgt ctcttctt ttctgcacac cctttcttc	2640
	gacgaccccc cctctctt tatatecctg cgatatgta tatcatcaag cctcggcact	2700
	tgttgcataat ctgtcctgat tatgttgtct ggatgctgca ggttggaaaag attccctta	2760
	tcggagctag tttgctggcc atcatccatc gcaaccagga ttcaccctg gctacgcgag	2820
	ccaggagcga ctttccgtg cttttggata ttctcacgca gctggactcg aaggcgtcgg	2880
	actaa	2885
<210>	12	
<211>	723	
<212>	PRT	
<213>	里氏木霉	
<400>	12	
	Met Ala Thr Ala Ala Ala Ala Ala Gly Gly Ala Ala Val Ala Ala	
1	5	10
	Gly Ala Asp Thr Gly Ala Ala Gly Ser Ser Ser Thr Gly Pro Pro Gly	
	20	25
	30	
	Leu Pro Gly Leu Pro Gly Thr Arg Thr Gly Ser Val Ala Met Gly Ser	
	35	40
	45	
	Ala Ala Pro Ala Gln Gly Ser Val Ala Ala Ala Gly Gly Pro Pro	
	50	55
	60	
	Ala Ala Gly Ala Gly Ala Gly Ala Val His Ala Leu Thr Thr Ser Pro	
[0016]	65	70
	75	80
	Glu Ser Ala Ser Ala Ser Gln Pro Gly Ser Pro Thr Ala Ser Thr Thr	
	85	90
	95	
	Pro Pro Gln Asn Ser Leu Val Ser Ala Ala Thr Ser Phe His His His	
	100	105
	110	
	Pro Arg Gly Arg Leu Val Ser Arg Ala Cys Asp Arg Cys Arg Arg Arg	
	115	120
	125	
	Lys Ala Lys Cys Glu Tyr Leu Ser Ala Val Asp Ser Cys Thr His Cys	
	130	135
	140	
	Arg Asp Ala His Val Gln Cys Thr Phe Asp Leu Pro Leu Ala Arg Arg	
	145	150
	155	160
	Gly Pro Lys Ala Arg Lys Lys Ser Asp Gln Pro Gly Gln Pro Pro Pro	
	165	170
	175	
	Asp Pro Ser Ser Leu Ser Thr Ala Ala Arg Pro Gly Gln Met Pro Pro	
	180	185
	190	
	Pro Leu Thr Phe Ser Gly Pro Ala Val Ala Leu Gln Pro Phe Ala	
	195	200
	205	
	Ser Ser Ser Leu Ser Pro Asp Ala Ala Trp Glu Pro Val Glu Pro Leu	
	210	215
	220	
	Ser Ile Asp Asn Gly Leu Pro Arg Gln Pro Leu Gly Asp Leu Pro Gly	
	225	230
	235	240
	Leu Ser Thr Ile Gln Asn Ile Ser Thr Arg Gln Arg Trp Ile His Leu	

	245	250	255
Ala Asn Ala Met Thr Leu Arg Asn Thr Thr Leu Glu Arg Val Ser Lys			
260	265	270	
Arg Cys Ile Asp Leu Phe Phe Asp Tyr Leu Tyr Pro Leu Thr Pro Leu			
275	280	285	
Val Tyr Glu Pro Ala Leu Arg Asp Val Leu Ala Tyr Ile Phe Ser Gln			
290	295	300	
Pro Leu Pro Gly Val Asn Gln Pro Ser Pro Leu Ser Gln Leu Thr Pro			
305	310	315	320
Asp Pro Thr Thr Gly Thr Pro Leu Asn Ala Ala Glu Ser Trp Ala			
325	330	335	
Gly Phe Gly Gln Pro Ser Gly Ser Arg Thr Val Gly Ser Arg Leu Ala			
340	345	350	
Pro Trp Ala Asp Ser Thr Phe Thr Leu Val Thr Ala Val Cys Ala Glu			
355	360	365	
Ala Ala Phe Met Leu Pro Lys Asp Ile Phe Pro Glu Gly Glu Ser Val			
370	375	380	
Ser Glu Ile Leu Leu Glu Ala Ser Arg Asp Cys Leu His Gln His Leu			
385	390	395	400
Glu Ala Asp Leu Glu Asn Pro Thr Ala Asn Ser Ile Ala Ile Arg Tyr			
405	410	415	
[0017] Phe His Ser Asn Cys Leu His Ala Ala Gly Lys Pro Lys Tyr Ser Trp			
420	425	430	
His Ile Phe Gly Glu Ala Ile Arg Leu Ala Gln Val Met Gln Leu His			
435	440	445	
Glu Glu Ala Ala Leu Glu Gly Leu Val Pro Ile Glu Ala Glu Phe Arg			
450	455	460	
Arg Arg Cys Phe Trp Ile Leu Tyr Leu Gly Asp Lys Ser Ala Ala Ile			
465	470	475	480
Leu Asn Asn Arg Pro Ile Thr Ile His Lys Tyr Cys Phe Asp Ala Gly			
485	490	495	
Ile Thr Thr Leu Tyr Pro Ser Gly Ile Glu Asp Glu Phe Leu Ser Thr			
500	505	510	
Ala Ser Glu Pro Pro Arg Lys Ser Phe Ile Ser Gly Phe Asn Ala Asn			
515	520	525	
Val Arg Leu Trp Gln Ser Ala Ala Asp Leu Leu Leu Glu Ile Arg Val			
530	535	540	
Leu Gln Asp Gln Met Met Gln His Phe Arg Gly Thr Met Pro Pro Asn			
545	550	555	560
His Val Leu Pro Ser Ala Asp Arg Gln His Leu Asp Ser Leu Tyr Val			
565	570	575	
Arg Phe Ile Thr Cys Leu Asp Asp Leu Pro Pro Tyr Leu Gln Ser Cys			
580	585	590	

Thr Leu Ala Met Ala Ala Met Ala Glu Gly Asn Gly Ser Ala Glu Ser
 595 600 605
 Lys Gln Tyr Val Ile Gln Cys Ile Asn Leu Gln Val Thr Phe His Cys
 610 615 620
 Leu Arg Met Val Ile Thr Gln Lys Phe Glu Asp Leu Ser Tyr Phe Ala
 625 630 635 640
 Pro Gly Val Glu Gln Ala Asp Leu Arg Lys Ser Glu Ile Val Arg Asp
 645 650 655
 Met Leu Arg Val Met Asn Glu Ala Pro Phe Trp Gly Leu Gln Ala Asn
 660 665 670
 Gly Glu Pro Asn Val Glu Lys Ile Arg Leu Ile Gly Ala Ser Leu Leu
 675 680 685
 Ala Ile Ile His Arg Asn Gln Asp Ser Pro Leu Ala Thr Arg Ala Arg
 690 695 700
 Ser Asp Phe Ser Val Leu Leu Asp Ile Leu Thr Arg Leu Asp Ser Lys
 705 710 715 720
 Ala Ser Asp

<210> 13

<211> 2185

<212> DNA

<213> 里氏木霉

[0018]

<400> 13

atgggctcag cagctccggc ccagggtct gttagctgcag ctgcaggcg ccctccagct	60
gctggcgctg gcgctggcg tgcacacccc ctcaccaccc cgccccggc tgcctcgcc	120
tgcagcccg gctcgccaac cgccctcaacc acgcccggc agaactca cgtgtcggt	180
gcaacctcg tccaccacca tcccagaggc cgtctggta gcagagcctg cgaccgctgc	240
cggccggcgca aggccaagtg cgagtaccc acgcgtgtcg atagctgcac gcactgccgc	300
gatgcccacg tgcatgtcac ttgcacccgc cccctggcg gacgcggccc caaagcgagg	360
aagaagagcg accagcccg ccagccgcct cctgatccga gctcgctc caccggcgct	420
cgaccggcc agatgccgccc gccgctgacc ttctccggcc ccgcagtagc cgcgctgcag	480
ccctcgccct cgtcgctcg tgcggccgac gggcgttgg agcccgatcg gccgctcage	540
attgacaacg gctgtccccg gcagccgctg gggacactgc ccggctctc caccatccag	600
aacatctcgaa cggcccgatcg atggatacac ctggccaaacg ccgtacgtcg gcaaacacg	660
acgcttagagc gctgtcgaa gctgtgtatc gaccttttc tcgactacct ctacccctc	720
accccccctgg tgtacgagcc ggccctccgg gacgtgtcg catacatctt ctccagccc	780
ttgcctggcg tcaaccaacc atgcggctg tcacagctca cgccagaccc gaccaccggc	840
accacccccc tcaacgctgc cgagtcgtgg gccggcttg gccagcccg cggctcgca	900
accgtcgccca gcaggctggc tccctggcc gactcgaccc tccacccgtt cacggccgtc	960
tgcgcagagg cagcattcat gctaccaag gacatttcc ccgaaggaga atccgtctct	1020
gagatcttgc tcgaacgcctc tcggactgc ctgcaccagg acctcgaggc cgacccgttgg	1080
aatccgacgg ccaactcgat tgccattcgc tacttccact ccaactgcct ccacgcgtcg	1140
ggaaagccca agtactcgat gcacatattt ggcgaggcca tccgcctggc gcaggcatg	1200
cagctgcacg aggaggctgc cctcgagggg ctctggccca tcgaggcaga gttccggcg	1260

cgctgctttt ggatcctgtta cttggggcgcac aagtcaagccg ctataactcaa caatcgcccc atcaccatcc acaaggtaactg cttcgacgccc ggcataccca cgctatacccg gtcggttate gaggacgagt tcctgagcac ggcgtccgag ccgcggccgaa agagcttcat atccggcttc aacgc当地atg tgccggctcg gcagtcgcg gctgatttc tgctggaaat ccgcgtgt caagatcaga tgatgcagca ctttcgaggg accatgcggc cgaaccatgt gctgccctcc gccgacaggc agcatctcg ttctcttat gtccgcttca tcacctgctt ggacgatctc ccggcgtaacc tccagtcgtg cactctggcg atggcagcga tggcagaagg caacgggtct gccgagtcca agcagtagt gatacagtgc atcaacctgc aggtgacggt tcactgtctg cgccatggtaa ttacgcagaa attcgaagac ctcttttatt ttgctctgg cggtgagcag gctgatctca gaaagtcgga gattgtgcga gacatgtga gggtgatgaa cgaggcgccc tttggggcc tgcaggccaa tggcgagccaa aacgtgagtc gtttccttgt ctcttcctt ttctgcacac cttttcttc gacgacccccc cctctcttt tatatccctg cggatatgt tatcatcaag cctcggcact tggctaat ctgtcctgtat tatgttgctt ggatgtgc ggttggaaaag attcgcctta tcggagctag tttgctggcc atcatccatc gcaaccagga ttcaccccttgcgactcgccg ccaggagcga ctttccgtt ctttggata ttctcacgc gctggactcg aaggcgtcg actaa	1320 1380 1440 1500 1560 1620 1680 1740 1800 1860 1920 1980 2040 2100 2160 2185
<210> 14	
<211> 678	
<212> PRT	
<213> 里氏木霉	
<400> 14	
[0019] Met Gly Ser Ala Ala Pro Ala Gln Gly Ser Val Ala Ala Ala Gly 1 5 10 15	
Gly Pro Pro Ala Ala Gly Ala Gly Ala Val His Ala Leu Thr 20 25 30	
Thr Ser Pro Glu Ser Ala Ser Gln Pro Gly Ser Pro Thr Ala 35 40 45	
Ser Thr Thr Pro Pro Gln Asn Ser Leu Val Ser Ala Ala Thr Ser Phe 50 55 60	
His His His Pro Arg Gly Arg Leu Val Ser Arg Ala Cys Asp Arg Cys 65 70 75 80	
Arg Arg Arg Lys Ala Lys Cys Glu Tyr Leu Ser Ala Val Asp Ser Cys 85 90 95	
Thr His Cys Arg Asp Ala His Val Gln Cys Thr Phe Asp Leu Pro Leu 100 105 110	
Ala Arg Arg Gly Pro Lys Ala Arg Lys Ser Asp Gln Pro Gly Gln 115 120 125	
Pro Pro Pro Asp Pro Ser Ser Leu Ser Thr Ala Ala Arg Pro Gly Gln 130 135 140	
Met Pro Pro Pro Leu Thr Phe Ser Gly Pro Ala Val Ala Ala Leu Gln 145 150 155 160	
Pro Phe Ala Ser Ser Ser Leu Ser Pro Asp Ala Ala Trp Glu Pro Val 165 170 175	
Glu Pro Leu Ser Ile Asp Asn Gly Leu Pro Arg Gln Pro Leu Gly Asp	

	180	185	190
Leu Pro Gly Leu Ser Thr Ile Gln Asn Ile Ser Thr Arg Gln Arg Trp			
195	200	205	
Ile His Leu Ala Asn Ala Met Thr Leu Arg Asn Thr Thr Leu Glu Arg			
210	215	220	
Val Ser Lys Arg Cys Ile Asp Leu Phe Phe Asp Tyr Leu Tyr Pro Leu			
225	230	235	240
Thr Pro Leu Val Tyr Glu Pro Ala Leu Arg Asp Val Leu Ala Tyr Ile			
245	250	255	
Phe Ser Gln Pro Leu Pro Gly Val Asn Gln Pro Ser Pro Leu Ser Gln			
260	265	270	
Leu Thr Pro Asp Pro Thr Thr Gly Thr Thr Pro Leu Asn Ala Ala Glu			
275	280	285	
Ser Trp Ala Gly Phe Gly Gln Pro Ser Gly Ser Arg Thr Val Gly Ser			
290	295	300	
Arg Leu Ala Pro Trp Ala Asp Ser Thr Phe Thr Leu Val Thr Ala Val			
305	310	315	320
Cys Ala Glu Ala Ala Phe Met Leu Pro Lys Asp Ile Phe Pro Glu Gly			
325	330	335	
Glu Ser Val Ser Glu Ile Leu Leu Glu Ala Ser Arg Asp Cys Leu His			
340	345	350	
Gln His Leu Glu Ala Asp Leu Glu Asn Pro Thr Ala Asn Ser Ile Ala			
[0020] 355	360	365	
Ile Arg Tyr Phe His Ser Asn Cys Leu His Ala Ala Gly Lys Pro Lys			
370	375	380	
Tyr Ser Trp His Ile Phe Gly Glu Ala Ile Arg Leu Ala Gln Val Met			
385	390	395	400
Gln Leu His Glu Glu Ala Ala Leu Glu Gly Leu Val Pro Ile Glu Ala			
405	410	415	
Glu Phe Arg Arg Arg Cys Phe Trp Ile Leu Tyr Leu Gly Asp Lys Ser			
420	425	430	
Ala Ala Ile Leu Asn Asn Arg Pro Ile Thr Ile His Lys Tyr Cys Phe			
435	440	445	
Asp Ala Gly Ile Thr Thr Leu Tyr Pro Ser Gly Ile Glu Asp Glu Phe			
450	455	460	
Leu Ser Thr Ala Ser Glu Pro Pro Arg Lys Ser Phe Ile Ser Gly Phe			
465	470	475	480
Asn Ala Asn Val Arg Leu Trp Gln Ser Ala Ala Asp Leu Leu Glu			
485	490	495	
Ile Arg Val Leu Gln Asp Gln Met Met Gln His Phe Arg Gly Thr Met			
500	505	510	
Pro Pro Asn His Val Leu Pro Ser Ala Asp Arg Gln His Leu Asp Ser			
515	520	525	
Leu Tyr Val Arg Phe Ile Thr Cys Leu Asp Asp Leu Pro Pro Tyr Leu			

530	535	540
Gln Ser Cys Thr Leu Ala Met Ala Ala Met Ala Glu Gly Asn Gly Ser		
545	550	555
Ala Glu Ser Lys Gln Tyr Val Ile Gln Cys Ile Asn Leu Gln Val Thr		
565	570	575
Phe His Cys Leu Arg Met Val Ile Thr Gln Lys Phe Glu Asp Leu Ser		
580	585	590
Tyr Phe Ala Pro Gly Val Glu Gln Ala Asp Leu Arg Lys Ser Glu Ile		
595	600	605
Val Arg Asp Met Leu Arg Val Met Asn Glu Ala Pro Phe Trp Gly Leu		
610	615	620
Gln Ala Asn Gly Glu Pro Asn Val Glu Lys Ile Arg Leu Ile Gly Ala		
625	630	635
Ser Leu Leu Ala Ile Ile His Arg Asn Gln Asp Ser Pro Leu Ala Thr		
645	650	655
Arg Ala Arg Ser Asp Phe Ser Val Leu Leu Asp Ile Leu Thr Arg Leu		
660	665	670
Asp Ser Lys Ala Ser Asp		
675		
<210> 15		
<211> 930		
[0021] <212> DNA		
<213> 人工序列		
<220>		
<223> 启动子序列		
<400> 15		
agcagtggct tatagcaata tcgtgcttgt ctctgccctc gctgaagctta tccctccctc		60
gtgtccccct tcgtggagaa tgactgcagt aaaggagacg atacgctttt ggaaagggtcc		120
tgaagggggt ggctgttggg attgcaacctt ctggcatttg tcaaatggcg ctttatgcgt		180
cgcgcgcg agtgatgtt agagggaaatg caacgttggt acggcttcac aatttgcctt		240
ttcattaagg cttattgccc gctaaattac tataaggaag gtcggcagct gggattacgc		300
atgtgttgc agacacaatc tcgaacgaga cgctatcaga ggacaagttt tctgacaatc		360
tgtcgtggta ttgaacgctt ttctgtttt taaaaccagca ttcatgaggt tcgagggccc		420
acgcattcat tggatctca tcaatgaagc ggaatagtac aagaagacgc agcatccatt		480
acagaccta tgcagctaag gaaacggatc acccttaaag acgtggatgt agttccgtt		540
gctcgccac attagaaggc tggactata ggtgaggaa ggcgcgcgtt gttgtttatc		600
cggccatgata agaggagtgt aagccggctg catggacttt ttccatcaccc tgtctgcct		660
gtccatctt aaattggcaa gctccaccc ccggtaaaaa tcgcaggcgtt aaggctttct		720
aataaccact cagcatttgc tcaaccatgt ctgtgttctt ctcacccatgc ctgcacatc		780
tttagcctc caggagttgg cgaattgtt ccgatccatc cgtcaaatcg cattgagata		840
actcacgata ttctgttttccatcc aggctaaaga attgcctcca gcgttccagt tcccgtcgta		900
ggccgagacg acactgcctc actaccagcc		930
<210> 16		
<211> 930		

<212>	DNA	
<213>	人工序列	
<220>		
<223>	启动子序列	
<400>	16	
	accccaatct gaggttcagg tttggcttcc ttcgagttct ccaagttctc tgaatcgat	60
	gtccgcattc attggtatcg aagtttgta ttaatctcgaa gaatgtgcata acttcagtca	120
	cctcaacata cacggaatcc agcctttca tgaggaatcc ttactccttc atacccaag	180
	tgcggggca gataattgtt cccattcaca aatgaactta gactgtactc cgcaacttctt	240
	tcaggctctt ctctcctcac gatgccccaa tgcttgccac tccacaaggat acatgttagta	300
	atctcgagtt actccccccgc ttttgcttac tttagccggca agaatgagat gcaagttgt	360
	tcctgtcggg agtattggct aaatgaaac acaagttagaa gaagagaaac agaaaaggc	420
	caataccggc tattcacaac ggatctcgct ctgtgtctga tagaactagt aaaagtccgc	480
	agtatccgct gtcataaga tcataacta aaacgtaaac taaacgcaat ggcttgaaa	540
	aggggattga gacagaagat acaccaacga ccacccctgt tagtgacaga gatggcagtc	600
	attcgactag cggctggcaa ttgggtgtccg cctgttattc aggctaaata tctcgagg	660
	ccgggaaatc atggtgagga ggagttggca tgggtggagc cgtgtatgc gtcggaccac	720
	accaagaagc tgggttagca ttccgtccc gatgggggtt aacatgggtt aacatgtgg	780
	aattgcaaat gatgcattgg ggaagaatgc aacatcgta tcgtcataacc gctcaattt	840
	aatattggc ttttccggg atcagatgg aagaaggcaac agagagagag acaaggaaga	900
	ccgtgagcca ttgaaggaca gccggacgca	930
[0022]	<210> 17	
	<211> 930	
	<212>	DNA
	<213>	人工序列
	<220>	
	<223>	启动子序列
	<400> 17	
	tctatggcca ctctgtcgcc tggcatctg tatcagcacc aaagccatga ccgtatgt	60
	ttgatgtgcc atcttggcgt tcttcggca acacacccggg cttgagggtt gtcactccct	120
	ctccgcactt gagcacaacg ccggctgcct cgtgtcccag aatacagtct cttcgaa	180
	cgaggggagcc gattctccca gacttccaga aatggacatc cgatcttgc ctctgtcag	240
	ctgcagttga atgcgagtt gttgaagacc tttagacte acccgcatat gecagtaget	300
	ttgatttgca gcaagacttg tccgcgtct ggttataaca ctgggtgttc gaccagttc	360
	agtttggtt cggcagttgc ttgttagtgtt ggttgggtc gtggagttt aatgatccta	420
	tgtcaagat cccgagctac agaaacgttg ccgtggca tgggtgtga gcccattgtt	480
	taggtcgagg caatcgacag aagctgaaa gtgttagt ttgggtgtaa tagaatgt	540
	acggtgactt tgcgaatgcc caaatgtcaa gtgtaaagta ccaaactact gtcggcagg	600
	tgcgaagaaac gcaatattgc tggaaattt taaatcgaa gggtagacaa taagacgt	660
	tggcctctga ggagctcaa tgcgcgt ccaagagggt gctaaaactt cccgattcg	720
	caatcctcac ggcttcctcg atgcggatgc aataacctct attagcctat atctgcgg	780
	gctatagacc gaggatgtc gtttgcata tcttcgacat catttctctg atcctttct	840
	ttttattata tagaagctga agttgtacgg cacaactgtc aacacacccgt tgccggaa	900
	ttctccctta tacttgacca atccatcaag	930

<210>	18	
<211>	930	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	启动子序列	
<400>	18	
	tcaacatatg gttggactag gctcgctca cgtccagttt cttcaagtca tatgttccag	60
	gaccggaaag atgtctgcaa attggatagt ttttcttcg tagctgcctt cgataaacct	120
	gggacgcgat acgcctcagt tcaagtactt gatctgaagt accgtatgtt taggctgata	180
	tccgatggaa tgcccccaaa tacaaggctc gtcaacacctga ttatgagaca attttcac	240
	ctacgtcaat ctgcggcat tctttgcaag aaacacgatc agtccatctc tcaggagcct	300
	gtcgtgttt tctgctgaga ccattggctt tccctcgac agtatgttgc ccctggcta	360
	actctctcc acgcctacactt ggcctacactt acctaccagg gaggcggctt tcttacactac	420
	ctatcaaagg gcctccacgt ggctgcgtatc ctctaatcaa caaatccctt cctcgcaagc	480
	tttttttgc tgactgtctt ccaggttctg ctggccttgtt tcctcgccaa ccagagcgat	540
	cgggtcagga agcagcgaga agcgagctt caggagaata agcgagtctg tactccatcg	600
	gaagtctctc gcgcgtcgatc ctgccgatcc cgaggccact cgcttttgc cgaatgagcg	660
	tcacgacggc cctacagata taccgatagc gttcccgtc cacctgatca agttgtgcc	720
	tactatatgt ctgcaggcat gaccgtatcc aaagctgcag cttggatttctt cacctccatca	780
	gagccgaccc ttggaaatagca ggggacacat ttggccctt cagcaccgct ctttggcg	840
[0023]	tctcctctgc gtttgatttg ttcttccact tctctcttcc ctgatatctc ttcttgaact	900
	gccggctgcg ttgtcgaact cgtgctcactc	930
<210>	19	
<211>	930	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	启动子序列	
<400>	19	
	gcgatttagca cgcatcgctg ctgggaccgt gagctcgacgg ggcctcaacacg catggtggcg	60
	tactgaggga tatcgttcac atcggttgcac gaggcggaca tgagaaacga ctgttatct	120
	gtattattta ctgttacatg tacggcagtc actaataacct gtctcaatca atgaagattt	180
	atgttaataaa gcttaatcaa actgtctaaa atatgtttac attttaatgtc ttgtctctg	240
	gtttgtctgtataaaatgcgtt gaggttgcacat cgtccggaa tagtgcacg aagtggatc	300
	accaagtggaa gggactcattt attgttcatg cagtccttcat gtacaaggcc taaaatccg	360
	agctattccc cgtatgaccc gtcaataactc gatacgatc agtagctgtt tggacaagca	420
	agtactactg gtatgggtgtt aaccccgact gctacaaacc ctgggaagca gactctcaca	480
	acttattact aggtactctc gttcggttcc tcaaagttt gggcatctcg catgagagca	540
	agacgttgcg ggcacatcgctt cagctacttta tgatgcctt tcatcaagcc catgagacca	600
	cctatgcact gtcacaagag aacaaacttgc ctttttggaa acccgcaaaa acaggggaga	660
	tggccgttgg tctttgttgc tctttatcc gtttgatctg tggccacagg aacggggaa	720
	aaaaacaactc gggttcgat ttgcgggttg gcagcttcca acaggggtat ggaaaatcgaa	780
	ccttccca ctagcggttta atatccacaa gccggcgaaa tggtaaattt ataaatcgag	840

	catcattccc atctgggtgg atttcaaaca aaagaaaaagg aagcgcacatcg tggaaagggt	900
	gcaaggctcg gataaaagtcc ttcttcgc	930
<210>	20	
<211>	930	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	启动子序列	
<400>	20	
[0024]	tgcgaaactgg caatgagttt tcaaccatca aagtacggc gcagttgtcg tacatgggtgg tccgagaata aaaaggctac atggtaaatt cgttaaaggg caaggtaat aggagcacag tgatgcttcc ttggacagtg ttgcttacat ccgatcatcc caccttgaca tacactcgct tttagtattc gatcgtgcca acccgattga tacattacaa gcagaactga gccgctgtta tcgacacact tgccgttcc tcagtgagac gettctacct cgcttgcgtt gatatcacga ggcattgtat acgtcaacca cttcatatca agtctgattt ctgaatgcct tcaatcattc aggcccacaa atatccctac agcatgctca ctttgaatag agttttctc tcttttcgg ccagcaggcc cgtttggcgc gtatgagcgc cgaagctgac actccagcga cgtgctgatc tagattctgg acacgaatag attccttattt aagggagtga cgtccccacg atgaactccc gtcgagacag ttttgcatttgc atacagccgt ttaacagcca gcggtgtact ttatacaaag aacaacgtcg ctgtgttagt tgccccaact tggcgtgatt ggacggacgc cgactcaacc cctgcatcac aatccgcca gaagcgtcac cgcggataaa aaggctaaat ggatggtctt caactcttcg cgacatggaa cgcgagcagg tattacagaa tgatgaccgg ttaaatctgg cagtctccat ggtactgaag tcaaagcaag agggacaatt cacaatggcc ttgatataaa agcgataaca tgtctccgtt caaggcctgg atgagcagag ttttagagcaa actgcctgt tggaaatctt ctctctactt tctcctcgac	120 180 240 300 360 420 480 540 600 660 720 780 840 900 930
<210>	21	
<211>	930	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	启动子序列	
<400>	21	
	ctggggtaac agagctgttt gcggggatga agatcttgcat gaatatgggt aatgaggaat ttttccattt ggaggaagca ttgtcgaatc caagggtatg cgggctggg gaagggcatt tctccttagc tgtaaggctt ctagatactg tgatactcga tgatataataa gccaacccca agcaaaagatc cgctgggtgt ctgaacaaggc cgaatgccag gccccttttgcggagcccc tcatcgaaat ccgtgtgc agcggttcat gggttcaaca ttccgaaatag ttctccgc agtctgtgg gtagcagcag cactccggca gatgccactg tgcctgtgaa tgggggaga ggccagagta tggggacca tggggaggg agcacggaga cactcgaagg tcatgccttgc gagttgtgc tactttttt ttgcattgcct tggttccatc atcgaaagga gagcagcgg agatagagcc ttaagcatgg gtaaaaactcc ttccgacaac tctataacag acgttaggaa aacaacaagt tcccaggat taccaggagc ttctttacc catcgccaca acgtgcac gccaagaatc gctgccatc tcctcatctc cagccctca aagtctccaa gctgcattca gcatctccag gcggaccccttgc tggcccaaca cttaggcta aaggcttctg cccctggc	600 120 180 240 300 360 420 480 540 600 660 720 780 840 900 930

gcagtggcag aagcaacctg ctacttgacg acatgaaccc gtttgtaccg gtcggccat	780
gagatgatgc agtgttccaa tggcatagtt gagtcgaagc ggtgtttca ccccttaatg	840
atgtAACATG tatataacgg atgaccatgc cctcTTGAC ttcacttcaa cggtCTTATT	900
atCCTGCCTC gagCCACAAG cagCgcAGTC	930
<210> 22	
<211> 930	
<212> DNA	
<213> 人工序列	
<220>	
<223> 启动子序列	
<400> 22	
catccacate atcttgacga ggaaccatcg ggtggtcgga cggactctgc agatctgggg	60
tgcgtctgac cgacgagggtg tcaatttagcc tacgtgttag tacccatgtc gagaattgtat	120
acctatggat gcaagtgcct ctgaactatt actttatcca taggagcatc cgctaccaat	180
gtgcggtaa aattgcctt cagccaccct gtccTCGGTT aattgagtga gacgtgcata	240
acatggccgg cagcatggca tggttagatga tattgggtga acgtgtcaga agaaaaggct	300
agaatattcg agacagcttgc ctgatatgtc caaaaacttct caagatattt gatatgtgt	360
gagttactt tggcattata ctgttatgtc aatgttagagt gtacgctaca gtacccat	420
ccaagaactt ataacatacc tacataccta ctgagctaca ttttaggcag tgcCTCCGCT	480
gagaagctt gaattctact tttgtcgTT taactcgtga cgcattgact ggcgggtcat	540
cctgatacag aatcaggacc attgcatatg aaaaacagtc tgagaccaaa catccaatcg	600
ggagacaatg ctgccatgaa agctgatttgc ctgttccatcaca actcaaagag	660
atgcTTATG ttccggggaa ggaaatatga ctgtgacttagt aatgcattcg ggtcctgcca	720
aagaatgagt tgactatgag gaggcaaata tctgagtcgt gtgagtagac cagtaatga	780
cagctggggg taatcactta tgtgttcacc ggataatactt catattgata taaaatgcta	840
tgtatctccaa gtcccAACGA tactgagatg aacagcatct cttaacagtt gctttccaca	900
taaataattc ctccTGTAC aagagccaaa	930
<210> 23	
<211> 1059	
<212> DNA	
<213> 人工序列	
<220>	
<223> 启动子序列	
<400> 23	
ggcaggcaact ggctcgacg acatgttttgc tatattgggtt gggactgcgg ccgcagcggg	60
ggcgggagct ggtggcggcg atatgaattt ccggggcgttg ctacaacagg taccacttg	120
accacccatg gctgcccgtc ccctgctgg agctttcagg tcgcttcgg gcgttggcga	180
ggcaagttgg acgggtggga aatgacgaaa aatgggtgcatt cgcctttgtt ggtgtgtgt	240
agtagtagtt ctactatgag gtacgtatgtt agcagaaggaa tcgagactaga atctggcggc	300
attgcaaaagg ttatctggaa agagggaaaag ggcttgcattt ggcattatgaa tgcattttc	360
gtacgaacta ctatctgata acagtttaggtt actgttatcc atacaaagag tcttatagaa	420
acactgcattt gtaataaaat actcgttgc tgcTTGAATA tagtaataag atcaacatcc	480
tttcacctct agtctccgtt gattccagta aaagcgtca attctgactt ccgactctgt	540
tgtatccccgg tgcgtccca tcgggggtt ctagacgctg cctcaacgcc catgtaccgg	600

cctgatgggg cccttgggg caccacaagt ccactaaacg aagcactgg gacggactc	660
gataggccctg agcagcagcc ggtctcagca gccaaaccagg ccagctggaa gcateggcta	720
ggggaggggg gcccaactac tacgtgtact actaggta taatgaattt gatgggaccc	780
agccagccca acctaacttt ccagcctta tagtcgcage ctgcctcccc gtgcctcacg	840
cttttgctc ctctgctggc cggactcgga cctttgcga cctctgcgt accaacaatc	900
cctcttgtt caccctctcg cttttgtac ctgcacgcctc aattcctcgc tgccgcctca	960
cctaaccgcg tgtgcttgac tgccctcacg ctgcgcgc tcctgcctcc gcgagccctc	1020
ttttacactt ttcaacagct accccgccag aattcaaca	1059
<210> 24	
<211> 1963	
<212> DNA	
<213> 人工序列	
<220>	
<223> 启动子序列	
<400> 24	
[0026] gacggaaict tcacacccat tactcgtaca tactagatca gatgctggc gagtcgtggg	60
aatgctggca ctggatggta taaacacccg actgcttgac gccggatcag cggcgatgca	120
ccgagctgga cgtgtcaata ctacatgtat catgacgcag agctgtttct cctctggcac	180
atgaacccaa agacacaagg gtcacctcgat tcctgagcag atcaggtaca gactccata	240
tctcactggc cccatctctc gagcctaaac cggccccc cttgcacaa gatgccccaa	300
gccgttgcgc agagacgcag tggactcgac ctgcgtcagcc atcgatgcgc tctcagctgc	360
cgcttggcc aaccaacttg ctgcgaggca aaagtatttg ctgcattgtat gccaggcaac	420
ctctctctcc ttgcctccctc cattcagccg gcagcatcc aaccacccac ccaccggcgc	480
agcgccacgc aaggcacagt caggactaac actcttaggc tgcctggata catgtagaac	540
ggtctttgg tgggtgggt tgctgcacat aaggtatata ctgtacatgc atatagtatg	600
catacagtag acgctgcttg gggccctgt atgttagggta tacatcatcg tccatctgc	660
gctccccccgc aatccgctat ccggcatacc gagagctcag gagccggcgtc ttgcccgtcc	720
ttgtacgat tacacctcca agtctccccccat cacagctgcc tccgcacggat	780
gaccatcccc caagcaagtc gtccttgcc acttgggtga gggtcgccag tggacgagag	840
ctgctgccaa tgcccatatg cagcaccggc ctgcgtcggc tgactggata ctgcaagta	900
ttaaaaacggt cgacctcgcc ggttggctc gtgcgtcaact ctgcctctc gactctgact	960
ctgactctt ccgtatcccc gtcgcgtcgat gttggaaact gtcctcttg tctttctc	1020
tgccaaaccc tctcgtact aggtatctca acctttgtat acgagatcag tatcgccggc	1080
acgcagggtc tcatcctgac tgcgtcgatg tggaaatcgat ctcacgtctc actgcacaaca	1140
agcagttgc gatacgcaaa tcaatcgatc gcatcaggatg cattccagat ccggccctcg	1200
gacgtcgta tataatcgcc agccgccttc tcccaagag cccagcctt tggagcagga	1260
cccaaggctt gtccatattt tgggtcgatc gtcgtccctt tgccatcgcc atctgttcc	1320
tcccaacccaa cccacgcccc tcatcaatct ccatctcccc cgccaaaggc gactcgatcc	1380
ccgcgtgcac acacaaacca cccggggaa aacgagacca acgctatatac ctgtggcgc	1440
gcccgcctaa tccccctcc acgtcgcc tgcgtcgatg agttgttat ccgggtgatg	1500
cagctcccgat tccctccctc tctcgttgc tgcgtcgatg agtggcttgc tccacccccc	1560
aatctcgcc gggcagcgga gcccagcctg agccgcctc cgagctcggt gccgttgc	1620
gaccctatac ttcccgaggc cccaaaggcc gacgtcaaaa ctgcgtcggt atcctgcgg	1680
ctctggcggc ctcgcgttg ggcttgcgtt tgaaagaggt tgccgatgac gggatcgccag	1740

tccccaaagtt	caacccttc	tctggcccac	aatcatgac	ccccatgcc	gtcggttatgc	1800
cactcatgtc	atcgagtgt	gaacattgac	ggagcttgac	taatgtgaca	gagtcatc	1860
gggttggga	gagccaagca	gggtcctcag	ctagcgagge	ccgacccgag	gcmcagagtc	1920
ccgaccatag	cacttataag	gccccgcttc	cgctgtcgta	act		1963
<210>	25					
<211>	1067					
<212>	DNA					
<213>	人工序列					
<220>						
<223>	启动子序列					
<400>	25					
gtgctgcagt	tgctacgcgt	agtctcaact	ggcatgcgt	cgacaatact	aaaatcagca	60
aaggcctgagg	tccgttgggt	aaaaacttca	ttcgtactcg	tactcagacc	gcaaattctg	120
gtatggtttgc	atgggttgggt	tcatagtgtat	catccatatac	atgggaagcg	ccgcctgctc	180
cgagaagcaa	caagtagccca	acagcatcggt	cactatcgta	catcatgccc	caactctc	240
ctagccgttag	cggacctccc	tcaaccagac	agtgcctaag	atccttgac	gcactatcag	300
cgagtctatg	tacactgacc	aacatgcagc	cccggcgact	tccagatgg	gccaacttca	360
gtcgagacgg	cgttgttttg	cagttatttgc	ccggccgttgc	atatgcgtt	aatggaaaggc	420
gttccgcgtgg	aacttccctg	gcccacgttgc	gttggatctc	atgattgtgg	tatccctccg	480
tgcctcatca	gttctcgagg	aacttctgtat	caccgaatgc	catgtgaggc	ccccgctc	540
ccccagctgc	agcactccgt	accccatatac	ccaggcaaat	tatgcacccc	gattgcttgc	600
aatcaaaatc	gttggcgcc	gtcttgatatac	gcccgttgc	aaggaaaagg	caggatgtc	660
gttattgggtt	gctcaatac	gtcgctcgac	ggggaaatgtc	ccgcgttgc	ccaatatcct	720
gtggcgacaa	gggttatcaa	aaagttccctc	atccgcgtat	tgccaaacaac	cggtggcgt	780
gacaaagggtg	ccttccaaaa	gaggcgatatac	ctggatcgcc	ccccatgc	gttggatct	840
atcaaacggc	cagctggac	ttctgttctg	ttctggatgt	cgtgtatgg	gtgtgttgg	900
tgtaccttgc	ccgttgcgt	gccggtgctc	tcttgtgtat	cattgcattt	aaatatcatc	960
gtgcctgtcg	ccagtttttgc	ccgacttggat	gattctgtc	agctattcct	ctgcaagcgt	1020
tcaactgccc	ccggtgagga	acactcggtc	agtgcacaa	caacacc		1067
<210>	26					
<211>	1124					
<212>	DNA					
<213>	人工序列					
<220>						
<223>	启动子序列					
<400>	26					
cgagggcagat	accgagatgt	tgttagcacgg	aagtcatctc	gtctagctgc	ccgtggggaa	60
aaaaaaaaaaa	gtaagccgga	accgcaggc	caggcctgc	gctgtacaaa	ggagcaaaaca	120
cgcataaaac	acaataccac	cctatggtag	catcaataat	ccgtatcgt	aatagcatgg	180
gcgaatattt	gcctctcg	tacggcgctc	cgtataaaa	ataacacccc	cccaaatctc	240
ggtgccccc	attcggtct	tttctccac	cagttcgaa	ccctccctcc	tatgtactct	300
ctctctct	ctcggttgggt	gggggagagc	gagcgagcga	gagagagagg	cagtggaaagg	360
ctcaggctcg	gtctttgtcg	cgccaaaccct	ccggccatgtc	atctccctt	gattagcgcc	420
ctcacccatcg	tttctcgcccc	cctcgatggc	gttccgcct	ccacaatctg	gggagcgttgg	480

aattagcaac	tgtggaaactg	agggtagat	cccacgcgtc	cgcttagtaa	gtgcttaagc	540
gccgtgcgaa	agggtgcgag	caccggagtc	gatgtcgaaa	tactgatgtc	ctccctggaa	600
cccaattact	ccggggggcg	actaaagcag	cccagccagc	caaactagca	gcaaggcaca	660
cgctcacact	agtaatgccc	ggccggcgga	tgatgtgct	tgtctgcttgc	tttcttgtgt	720
cagggccagg	aagccaaatc	ttgtctgggg	gggcgacata	aagggccgt	cgtcgcccta	780
ccttgagcca	cgcaaccaca	cacctgtccg	gttcctcaa	ttccttatac	caccaaacga	840
accataacaa	caacgctctc	ctcctcttc	atcctcttc	atcattgtat	catcgctata	900
gcattgcgtt	gccagctctg	ttgtgtctgt	cagttcatat	aacaaaggcct	tccgtgtctc	960
tttcttcte	tcacactata	ttttttttt	ttcacctctc	atcacaaaga	cccacacett	1020
ttccgccaca	actgctctta	ccaaaggcac	actcgcacct	ttgaatgatc	tcctccttgc	1080
ccaaaagact	caagtccaga	agaagttcac	ggagttccaa	agcc		1124
<210>	27					
<211>	780					
<212>	DNA					
<213>	人工序列					
<220>						
<223>	启动子序列					
<400>	27					
agactagcgg	ccggccccct	tatccagct	gttccacgtt	ggcctgcccc	tcagtttagcg	60
ctcaactcaa	tggccctcac	tggcgaggcg	aggcaagga	tggaggggca	gcatgcctg	120
agttggagca	aagcgccgc	catggagca	gcaaccaac	ggagggatgc	cgtgtttgt	180
[0028]	cgtggctgt	gtggcaatc	cggcccttg	gttggctcac	agagcgttgc	240
tgagctatta	ttgcttagta	cagtatagag	agaggagaga	gagagagaga	gagagagggg	300
aaaaaaaggtg	aggttgaagt	gagaaaaaaaaa	aaaaaaaaaa	aatccaacc	actgacggct	360
gccggctctg	ccacccccc	ccctccaccc	cagacaacct	gcacactcag	cgcgcagcat	420
cacctaattct	tggctcgcc	tcccgagct	caggttgttt	ttttttctc	tccctcgat	480
cgaagccgcc	cttggccct	tatttatttc	cctctccatc	cttgcgttgc	tttggccat	540
ctgccccctt	gtctgcatct	ctttgcacg	catgcctta	tcgtcgctc	tttttcaact	600
cacgggagct	tgacgaagac	ctgactcg	agcctaacc	gtgtttct	ctccccccct	660
cccgaccggc	ttgacttttgc	tttctctcc	agtaccttat	cgcgaagccg	gaagaaccct	720
ctttaacccc	atcaaacaag	tttgtacaaa	aaagcaggct	ccgcggccgc	ccccttacc	780
<210>	28					
<211>	58					
<212>	DNA					
<213>	人工序列					
<220>						
<223>	引物					
<400>	28					
tcagggttat	tgtctcatgg	ccat taggc	ctggcaggca	ctggctcgga	cgacatgt	58
<210>	29					
<211>	58					
<212>	DNA					
<213>	人工序列					
<220>						

<223>	引物	
<400>	29	
	agagccctgg gccggagctg ctgagccat tggtaattc tggcggttga gctgttga	58
<210>	30	
<211>	58	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	30	
	tcaacagcta cccggcaga attcaacaat gggctcagca gctccggccc agggctct	58
<210>	31	
<211>	59	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	31	
	tcgtaaataa acaagcgtaa ctagctagcg tagtttatgc gagcaacatt gcacgaaac	59
<210>	32	
<211>	59	
[0029]	<212>	DNA
	<213>	人工序列
<220>		
<223>	引物	
<400>	32	
	gttcgtgca atgttgctcg cataacctac gctagctagt tacgcttgc tatttacga	59
<210>	33	
<211>	58	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	33	
	acatgtcgcc cgagccagtg cctgccaggc ctaaatggcc atgagacaat aaccctga	58
<210>	34	
<211>	57	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	34	
	aggtgtttaaga cgggggagta gcgcagcatt gtttaattct ggccgggttag ctgttga	57
<210>	35	

<211>	57	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	35	
tcaacagcta cccggccaga attcaacaat gctgcgtac tccccgtct tacacct		57
<210>	36	
<211>	59	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	36	
tcagggttat tgtctcatgg ccatttaggc ctagactagc ggccggccc cttatccca		59
<210>	37	
<211>	55	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	37	
[0030] agagccctgg gccggagctg ctgagccat ggtgaagggg gcgccgcgg agcct		55
<210>	38	
<211>	55	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	38	
aggtcccgccg gccggccct tcaccatggg ctcagcagct ccggcccagg gctct		55
<210>	39	
<211>	59	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	39	
tgggataagg ggaccggccg ctagtctagg cctaaatggc catgagacaa taaccctga		59
<210>	40	
<211>	51	
<212>	DNA	
<213>	人工序列	
<220>		

<223>	引物			
<400>	40			
tgtaagacgg	gggagtagcg cagcatggtg aagggggcgg ccgcggagcc	t		
<210>	41	51		
<211>	51			
<212>	DNA			
<213>	人工序列			
<220>				
<223>	引物			
<400>	41			
aggctccgcg	gccgccccct tcaccatgct gcgtactcc cccgtttac a	a		
<210>	42	51		
<211>	7882			
<212>	DNA			
<213>	人工序列			
<220>				
<223>	质粒 pYL1			
<400>	42			
ggcaggca	ct ggctcgga	c acatgtttt gatattggtt	ggactgcgg ccgcagcggg	60
ggcgggagct	gg tggcgccg	c atatgaattt ccggggcttg	ctacaacagg taccactttt	120
accacccatg	g ctgcgtcg	ccctgcttgg agcttcagg	tcgcttccgg gcgttggcga	180
ggcaagttgg	ac ggtgggga	aatgacgaaa aatggtcat	cgccttgtta ggtgtgtgt	240
agttagtagtt	ctactatgag	gtacgtatgt agcagaagga	tcgagctaga atctgcggc	300
attgcaaagg	ttatctggaa	agaggaaaag ggcctgaacc	ggcatatgg tgcattttc	360
gtacgaacta	ctatctgata	acagtttagt actgttatcc	atacaagag tctttagaa	420
acactgcate	gt aataaaaat	actcggtac tgcttgaata	tagtataag atcaacatcc	480
tttcacctct	ag tcccg	t gattccagta aaagcgtca	attctgactt cgcactctgt	540
t gatgc	cccg	t gatgc	ctagacgtg cctcaacgccc	600
cctgatgggg	cccttgggg	caccacaagt ccactaaacg	aagcaactggg gacgggactc	660
gatagccctg	agcagcagcc	ggtctcagca	gccaaccagg ccagctggaa	720
ggggaggggg	g cccaa	actac tacgtgtact	actaggta taatgaattt gatgggaccc	780
agccagccca	acctaacttt	ccagccttta tagctgcage	ctgcttcccc gtgcctcac	840
c ttttgc	ctctgcttgc	cg gactcgga	cctctgctcg accaacaatc	900
cctcttgtt	acc	cctctc	ctcgacgtc aattctcgc tgccgcctca	960
cctaaccgc	tg tgc	tttgcac	ctcggctcgc ctctgctcc	1020
ttttacactt	ttcaacagct	accccgccag	gcgagcctcc	1080
cagggctctg	tag	ctgcagggc	cctccagctg ctggcgttgg cgctggcgt	1140
gtccacgccc	tcaccac	ctc	ctgcgttgc	1200
gcctcaacca	gccc	gag	ccgcgtt ccaccac	1260
cccagaggcc	gt	ctgc	gttgc gac	1320
agtctagccc	tttgc	tttgc	ccgcgtt ccaccac	1380
ctgatgtgc	tg	tttgc	ccgcgtt ccaccac	1440
c ttcatatgt	c ttactg	tttgc	ccgcgtt ccaccac	1500
cttcttcct	gcgtgtctgt	catgcgtaca	aagcatacat acaatacatc	1560

[0032]	gcaagcgttg tgggtgtt agagttgtgt gtattgtatt gcactgcctt cacaactcgta tcatactgtc gcagccctcac cccaaacacccg acctcgctt ccatgctgcg ctactcccc gtcttacacc tggatactct ctccttgcca ccactgatcca atgctttcc cgccccaaag tgcgagtacc tcagecgctgt cgatagctgc acgcactgcc gcgatgccca cgtgcagtgc actttcgacc tgccctgjc gcgacgcggc cccaaaggcga ggaagaagag cgaccagccc ggccagccgc ctccgtatcc gagctcgctc tccaccgcgg ctcgaccgg ccagatgccg ccggccgtga cttctccgg ccccgatca gcccgcgtgc agcccttcgc ctgcgtcg ctgtcgcccg acgcggccgt ggagccgcgc gagccgcgtca gcattgacaa cggcctgccc cgccagccgc tggcgaccc gcccggctc tccaccatcc agaacatctc gacgcgccag cgatggatac acctggccaa cggcatgacg ctgcgcaca cgcgcgtaga ggcgcgtcg aagcgatgta tcgacccctt cttcgactac ctctacccccc tcacccccc ggtgtacgag ccggccctcc gggacgtgtc cgcatatc ttctccaggc cttgcctgg cgtcaaccaa ccatggccgc tgcacagct cacgcccac cccgaccaccc gcaccacccc cctcaacgc gccgagtcgt gggccggctt tggccagccc agccgcgtgc gaaccgtcgg cagcaggctg gctccctggg ccgactcgac cttcaccctg gtcacggccg tctgcgcaga ggcagcattc atgctaccca aggacattt ccccgagga gaatccgtt ctgagatctt gtcgaagcc tctcggtact gcctgcacca gcacctcgag gccacccgg agaatccgcac gccaactcg attgcattc gctacttcca ctccaaactgc ctccacgcgtc cggggaaagcc caagtactcg tggcacatat ttggcgaggc catccgcctg ggcaggtca tgcagctca cgaggaggct gccctcgagg ggctcgccc catcgaggca gagttccggc gtcgctgtt ttggatctg tacttggcg acaagtccgc cgctatactc aacaatcgcc ccatcaccat ccacaagttac tgcgtcgacg cggcatcactc acgcgtatac ccgtcggtt tcgaggacg gttcctgagc acggcgcccg agccggcccg gaagagctt atatccggct tcaacgcggaaa tgtgcggctc tggcagtcg cggctgattt gtcgtggaa atccgcgtgc tgcaagatca gatgtatcg caacttcgag ggaccatgcc cccgaaccat gtgtccctt ccgcgcacag gcagcatctc gattctctt atgtccgcctt catcacctgc ttggacgatc tccgcgttta cctccagtc tgcactctgg cgatggcagc gatggcagaa ggcaacgggt ctgcgcgtc caagcgtac gtgatacgt gcatcaacctt gcaggtgacg tttcactgtc tgccatgtt aattacgc aaattcgaag acctcttta ttttgcctt ggcgttgagc aggctgtatc cagaaagtcc gagattgtgc gagacatgtc gaggggtgatg aacggggcgc ccttttggg cctgcaggcc aatggcgagc caaacgtgag tcgtttcctt gtctttctc ttttgcac accctttct tcgacgaccc cccctctctc ttatatccc tgccgatatg tatatcatca agcctcgca cttgttgcata atctgtccctt attatgtt gtcgttgcgtc cagggtgaaa agattcgct tatcggtact agttgtgg ccacatcca tcgcaaccag gattcaccct tggctacgc agccaggaggc gactttccg tgctttggaa tatttcacg cggctggact cgaaggcg ggaccaactt aggaataactt ccactaccgt tgggtgtt aaatgtgtt gaaacaacaaa aaatgtcaaa gtcgggttaa atatggccag gatttgtt ttattccccc tttagcggtt ctgggtattt cccctttgtt tacttttc tgggtttcc agcacttggtt tttccagc tggggggaaac aaaaggcggtt tctttccctt atgcccgggg ttgtccgatt tagcattga gtgtacatct tccctacatt actaggtact taatgagatctt atggagatctt cccgtt cgatattca tcacgttggt gtatatacc gtcgttggct ttgaaacctt gaggatgggtt gcaatgcgtt gacgcctttt gcaaggacc aaaataagcg aaggatgaag tctgaatagg atacgaactt gctacctatg ggtgagcatg aaatgaagcg gtcggggaaa tggcggagaa acgcgtcgacg taacgctgtt gtttctcc gttcgttca atgttgcgtc cataacactac
--------	---

tcctttgate	ttttctacgg	ggtctgacgc	teagtggAAC	gaaaactcac	gttaaggcct	6840
gcagggccga	ttttggatcat	gagatttatca	aaaaggatct	tcaccttagat	cctttaaat	6900
taaaaaatgaa	gttttaaatc	aatctaaagt	atatatgagt	aaacttggc	tgacagttac	6960
caatgttAA	tcaagtggc	acctatctca	gcatctgtc	tatttcgttc	atccatagtt	7020
gcctgactcc	ccgtcggtgt	gataactacg	atacgggagg	gcttaccatc	tggccccagt	7080
gctgcaatga	taccgcgaga	cccacgctca	ccggctccag	atttatcagc	aataaaccag	7140
ccagccggaa	gggcccggc	cagaagtgg	cctgcaactt	tatccgcctc	catccagttc	7200
attaattgtt	gccgggaagc	tagagtaagt	agttcgccag	ttaatagttt	gchgcaacgtt	7260
gttgccattt	ctacaggcat	cgtggtgtca	cgctcgctgt	ttggtatggc	ttcattcagc	7320
tccgggttccc	aacgatcaag	gcatgttaca	tgtatcccc	tgttgtgcaa	aaaagcggtt	7380
agctccttcg	gtcctccgat	cgttgcaga	agtaagttgg	ccgcagtgtt	atcactcatg	7440
gttatggcag	cactgcataa	ttctcttact	gtcatgccat	ccgtaagatg	ctttctgtg	7500
actgggtgag	actcaaccaa	gtcattctga	gaatagtgt	tgccggcacc	gagttgtct	7560
tgcccggcgt	caatacggg	taatacccg	ccacatagca	gaactttaaa	agtgctcatc	7620
attggaaaac	gttcttcggg	gcaaaaactc	tcaaggatct	taccgctgtt	gagatccagt	7680
tcgatgtaac	ccactcggtc	acccaactga	tcttcagcat	cttttacttt	caccagcg	7740
tctgggtgag	caaaaacagg	aaggcaaaaat	gcccaaaaaa	aggaaataag	ggcgacacgg	7800
aaatgttga	tactcatact	tttcctttt	caatattatt	gaagcattta	tcagggttat	7860
tgtctcatgg	ccatTTAGGC	ct				7882

<210> 43

<211> 7279

<212> DNA

[0034]

<213> 人工序列

<220>

<223> 质粒 pYL2

<400> 43

ggcaggca	ggctcgga	acatgtttt	tatattgg	ttt gggactgcgg	ccgcaggcgg	60
ggcgggagct	ggtggcggcg	atataattt	ccgggcgtt	gctacaacagg	taccacttt	120
accacccatg	gctgcccgtc	ccctgttgg	agctttcagg	tgcgttccgg	gcgttggcga	180
ggcaagttgg	acgggtggg	aatgacgaaa	aatggtgcat	cgcctt	gttgcgtgt	240
agtagtagtt	ctactatgag	gtacgtatgt	agcagaagga	tgcagactaga	atctgcggc	300
attgcaaaagg	ttatctggaa	agagggaaaag	ggcctgaacc	ggcatatgg	tgcattttc	360
gtacgaacta	ctatctgata	acagtttagt	actgttatcc	atacaagag	tcttatagaa	420
acactgcac	gtataaaaat	actcggtac	tgcttgaata	tagtaataag	atcaacatcc	480
tttcacctt	agtctccgt	gattccagta	aaaggcgtca	attctgactt	ccgactctgt	540
tgtatcccc	tgtctgcca	tcgggtgg	ctagacgctg	cctcaacgcc	catgtaccgg	600
cctgtatggg	cccttgggg	caccacaagt	ccactaaacg	aagcactgg	gacgggactc	660
gatagccctg	agcagcagcc	ggtctcagca	gccaaccagc	ccagctggaa	gcatcggtca	720
ggggaggggg	gccccaaactac	tacgtgtact	actaggtaca	taatgaattt	gatgggaccc	780
agccagccca	acctaacttt	ccagccttta	tagtgcagc	ctgttcccc	gtgcctcacg	840
cttttgtctc	ctctgttggc	cggaactcgga	cctttgcga	cctctgtctcg	accaacaatc	900
cctcttgg	caccctctcg	cttttgcac	ctcgacgctc	aattcctcgc	tgccgcctca	960
cctaaccgcg	tgtgcttgc	tgccctcagc	ctcgacgctc	ctcctgttcc	gcgagctcc	1020
tttacactt	ttcaacagct	accccggc	aattcaacaa	tgcgtcgct	ctccccggc	1080

[0035]

ttacacctgg atactctctc cttgccacca ctgaccaatg ctctcccg cccaaagtgc	1140
gagtaccta gcgcgtcga tagctgcacg cactgccgcg atgcccacgt gcagtgca	1200
ttcgacctgc ccctggcgcc acgcggcccc aaagcgaggaa agaagagcga ccagccggc	1260
cagccgcctc ctgatccgag ctgcgtctcc accgcggctc gaccggcca gatgccggc	1320
ccgctgacct tcicggccc cgcaatgcgcgc ccttcgcctc gtgcgtcg	1380
tgcggcgcg cggcctggaa gcccgtcagca ttgacaacgg cctgccccgg	1440
cagccgctgg gcgacactgccc cggcctctcc accatccaga acatctcgac gcgcacgc	1500
tggatacacc tggccaacgc catgacgtcg cgcacacgcgc cgctagacgc cgtctcga	1560
cgatgtatcg accttcttctt cgactacctc taccctctca ccccccgtt gtacgagccg	1620
gcctccggg acgtgcgtc atacatcttc tccagccct tgcctggcgt caaccaacca	1680
tcgcccgtt cacagctcac gccagacccg accacggca ccacccctt caacgtgcc	1740
gagtcgtggg ccggctttgg ccagccgc ggctcgcaaa cgcgtggcag caggctggc	1800
ccctggggcc actcgacccctt caccctggc acggccgtct gcgcagaggc agcattcatg	1860
ctacccaagg acatttccc cgaaggagaa tccgtctctg agatcttgc cgaaggctct	1920
cgggactgcc tgaccaggca cctcgaggcc gacctggaga atccgacggc caactcgatt	1980
gcattcgct acttccactc caactgcctc cacgtgcgg ggaagcccaa gtactcg	2040
cacatatttgc gcgaggccat ccgcctggcg caggtcatgc agtgcacgc ggaggctgc	2100
ctcgaggggc tgcgtcccat cgaggcagag ttccggcgtc gctgttttgc gatccgtac	2160
ttggcgcaca agtcagccgc tatactcaac aatcgccca tcaccatcca caagtactgc	2220
ttcgacccgc gcatcaccac gctatacccg tcgggtatcg aggacgagg cctgagcacg	2280
gcgtccgagc cgcggggaa gagcttcata tccggcttca acgcaaatgt gcggctcg	2340
cagtccgccc ctgatttgcgt gtcggaaatc cgctgtgc aagatcagat gatgcacgc	2400
tttcgaggga ccatccccca gaaccatgtc ctgcctccg ccgcacaggca gcatctcgat	2460
tctctctatgc tccgcttcat cacctgttgc gacgtctcc cgcgttaccc ccagtcgtc	2520
actctggcga tggcagcgat ggcagaaggc aacgggtctg ccgagtcacaa gcagtcgt	2580
atacagtgc tcaacactgca ggtgacgttt cactgtctgc gcatggtaat tacgcagaaa	2640
ttcgaagacc tctcttattt tgctctggc gttgagcagg ctgatctcag aaagtccgg	2700
attgtgcgag acatgtcgag ggtgatgaac gaggccct tttgggcgc gcaggccaaat	2760
ggcgagccaa acgtgagtcg ttcccttgc tctctcttt tctgcacacc ctttttcgc	2820
acgacccccc ctctctctt atatccctgc ggatatgtat atcatcaagc ctggcactt	2880
gttgctaatac tgcgttgcatt atgttgtctg gatgtgcag gttgaaaaga ttgccttat	2940
cggagctgtt ttgcgtggca tcatccatgc caaccaggat tcacccttgc ctacgcgac	3000
caggagcgcac ttccctgtc ttggat tctcacgcgg ctggactcga aggctcg	3060
ccaaactgagg aatacgtcca ctaccgtgt tggttgcggaa caacaaaaaa	3120
tgtcaaagtc ggtgttgcggat tggccaggat ctttgttta ttcccttc agcgttgc	3180
gggttcccttcc ttgttgcgtt ttttccatgc acttgtttt ccagcgttgc	3240
ggggaaacaaa aggccgttct ttccctatgc ccagggttg tccgatttag cattgtgt	3300
tacatcttcc ctacattact aggtacttgc tggttgcggat gagatctccc gtcattccgg	3360
atattcatca ctttgcgttgc tatattccgtt gttggcttgc aaacccggat ttgggttgc	3420
atgcgttgc gccttttgcg aaggccaaa ataagcgaag gatgaagtc gataggata	3480
cgaactggct acctatgggt gaggcatggaa tgaagcggc gggaaatgg cggagaaacg	3540
ctcgacgtaa cgctgttgcgtt ttccctatgc tgcgttgcatt ttgcgttgc gatccgtt	3600
agcttagttac gtttgcgttgc ttacgacaag atctagaaga ttgcgttgc gatccgtt	3660
ataacaacaa ttgccttcc ttcccttccctt ctgcgttgc gatccgttgc	3720

tgcttaatca gtgaggcacc tatctcagcg atctgtctat ttcggtcatc catagttgcc	6420
tgactccccg tcgtgttagat aactacgata cgggagggt taccatctgg ccccagtgc	6480
gcaatgatac cgcgagaccc acgctcacgg gctccagatt tattcagcaat aaaccagcca	6540
gccggaaaggg ccgagcgcag aagtggctt gcaactttat ccgcctccat ccagtcatt	6600
aatttgtgcc gggaaagctag agtaagtatg tcgcccattt atagtttgcg caacgttgc	6660
gccattgcta caggcatcgt ggtgtcacgc tcgtcggtt gtagggcttc attcagtc	6720
ggttcccaac gatcaaggcg agttacatga tccccatgt tgtgcaaaaaa agcggttagc	6780
tccttcggtc ctccgatcgt tgtcagaagt aagttggccg cagtgttatac actcatgtt	6840
atggcagcac tgcataattc tcttactgtc atgcctatcc taagatgctt ttctgtact	6900
ggtgagttact caaccaagtc attctgagaa tagtgtatgc ggcgaccgag ttgtcttgc	6960
ccggcgtaa tacggataa taccggccca cataggcagaat cttttaaaagt gtcatcatt	7020
gaaaaaacgtt ctccggggcg aaaactctca aggatcttac cgctgtttagt atccagttcg	7080
atgttaaccca ctcgtgcacc caactgatct tcagcatctt ttactttcac cagcgttct	7140
gggtgagcaa aaacaggaag gcaaatgcc gcaaaaaagg gaataaggc gacacggaaa	7200
tgttgaatac tcatactttt ctttttcaa tattattgaa gcatttatca gggttattgt	7260
ctcatggcca tttaggcct	7279
<210> 44	
<211> 7603	
<212> DNA	
<213> 人工序列	
<220>	
<223> 质粒 pYL3	
[0037] <400> 44	
agactagcgg ccggccccct tatccagct gttccacgtt ggcctgcccc tcagttagcg	60
ctcaactcaa tggccctcac tggcgaggcg agggcaagga tggaggggca gcatcgctg	120
agttggagca aagcgccgc catggagca gcaaccaac ggagggatgc cgtgtttgt	180
cgtggctgtc gtggcaatc cggcccttg gttggctcac agagcgttgc tgtgagacca	240
ttagctatta ttgcttagta cagtatagag agaggagaga gagagagaga gagagagggg	300
aaaaaaaggta aggttgaagt gaaaaaaa aaaaaaaaaaa aaatccaacc actgacggct	360
gccggctctg ccacccccc ccctccaccc cagacaacct gcacactcag cgccgacat	420
cacctaattt tggctcgct tcccgacgt caggttgttt ttttttctc tctcccttgt	480
cgaagccgccc cttgtccct tatatttc cctctccatc cttgtctgcc tttggccat	540
ctggccctt gtctgcatct cttttgcacg catgcctta tggcgtctc ttttttact	600
cacggagct tgacgaagac ctgactcggt agcctaacct gctgatttct ctccccccct	660
cccgaccggc ttgacttttgc ttctcccttcc agtaccttac cgcgaagccg gaagaacct	720
ctttaacccc atcaaacaag ttgtacaaa aaagcaggct cgcggccgc ccccttcacc	780
atgggctcag cagctccggc ccagggctct gtagctgcag ctgcaggccg ccctccagct	840
gctggcgctg gcgctggcgcc tggccacgc ctcaccaccc cggccgagtc tgcctcgcc	900
tcgcaggccg gctcgccaaac cgcctcaacc acgcccggc agaactcact cgtgtcggt	960
gcaaccttgt tccaccacca tcccgaggc cgtctggta gcagagcctg cgaccgctgc	1020
cggccggcgca aggccaaggt cagtctagcc ctttgcgtgt tgcttgcac tctgttgtca	1080
ttgctcccttcc tccctgcgtct gctgatgcgt ctgctcccttcc tccctcccttcc cctcccttcc	1140
tcctggtccc tggccctgc tcttcatatg tccttactgc cgggtctcc tctcccttcc cccgttcccccc	1200
ctcctcccttcc tcccttcgtc tgcgtgtctg tcatgcgtac aaagcataca	1260

[0038]	tacaatacat cagcatacat ggcaagcggtt gtgttgtt gagagttgt ttttttttat tgcactgcct tcacaactcg ttccatctgc tgccggctca ccccaacacc gacccgtct tccatgtgc gctactcccc cgtttcacac ctggataactc ttccttgcc accactgacc aatgctcttc cccgcccaaa gtgcgagta ctcagcgctg tccatgtgc cacgcactgc cgccatgc acgtgcagtg cacttcgac ctggccctgg cgccacgcgg cccaaagcg aggaagaaga gcgaccagcc cggccagccg cctctgtate cgagctcgct ctccaccgcg gtcgaccccg ccacatgc cccggccgtg accttcctgg gcccccgagt agccgcgtg cagcccttcg ctcgtcgcte gctgtcgccc gacgcggctt gggagccgt cgagccgctc agcattgaca acggcctgcc cggcagccg ctggcgacc tgccggctt ctccaccatc cagaacatct cgacgcgcca gcgtggata cacctggca acggcatgac gtcgcgcaac acgacgctag agcgcgtctc gaagcgatgt atcgacctct tttcgacta cttctacccc ctcacccttc tgggtacga gcccggctc cggacgtgc tccatgtacat cttctccag cccttgcctg gcgtcaacca accatgcggc ctgtcacagc tcaacccaga cccgaccacc ggcaccaccc ccctcaacgc tgccgagtc tgccggctt tgccagcc cagcggctcg cgaaccgtcg gcagcaggct ggccggctgg gcccggctga cttccaccctt ggtcacggcc gtctgcgcag aggacgattt catgtacccc aaggacattt tccccgaagg agaatccgctc tctgagatct tgctcgaae ctctcgggac tgcctgcacc agcactcgaa gcccggactcg gagaatccga cggccaaactc gattggcatt cgctacttcc actccaaactg cttccacgct gcggggaaagc ccaagtactc gtggcacata tttggcgagg ccacccggctt ggcgcaggcc atgcagctgc acgaggaggc tgccctcgag gggctcgcc ccacggagcc agagttccgc cgtcgtcgct tttggatctt gtacttggc gacaagtcag ccgtataact caacaatcg cccatcaccatc tccacaagta ctgttcgac gcccggatca ccacgtata cccgtcggg atcgaggacg agttccttag caccgggtcc gagccggcc ggaagagctt catatccggc ttcaacgcaa atgtcggtt ctggcgttcc gcccggctt tgggtcgat aatccgcgt ctgcaagatc agatgtatgc gacatttcgaa gggaccatgc ccccaacca tgggtcgccc tccggcaca ggcagcatct cgattctctc tatgtccgtt tccatcacctg cttggacgat ctccggcgat acctcccgatc gtgcactctg gcgatggcag cgatggcaga aggcaacgg tctggcgat ccaagcgatc cgtgatcag tgcataacc tgcaggtgac gtttactgt ctgcgtcgatgg taattacgca gaaattcgaa gaccccttctt atttgtcc tggcggtttag caggctgtatc tcagaaagtc ggagattgtt cgagacatgc tgagggtat gaacggaggcg cccttttggg gcctgcagge caatggcgag ccaaacgtga gtcgttccct tggatcttct ctttctgca cacccttttcc ttgcacgacc cccctcttctt ctttatatcc ctgcggatata gtatatatc aaggcccgatc acttgggttctt aatctgtccct gattatgtt gtcgtatgt gcagggtgaa aagattcgcc ttatcgagc tagttgttgc gccatcatcc atgcgaacca ggattcaccc ttggctacgc gagccaggag cgactttcc gtgttttgg atattctcac cgccgtggac tcgaaggcgat cggaccaact gaggaaatcg tccactaccg ttgttggttca aatgtgtttt ggaacaacaa aaaatgtcaa agtgggttgc aatatggca ggttccat gttattttttt cttcagcgat gctgggttattt tccctttgtt ttactttttt ctgttttt cagcacttgtt tttcccgatc gtggggggaa caaaaggcgat ttctttttcc tggccagg gttggcgat ttagcattt agtgcacatc ttccctacat tactaggtac ttaatgagct tatggagatc tcccgatcattt ccggatattt atcactttgg tttatatatc cgtgggttgc tttggaaacctt ggagttgggt tgcaatgcag tgacccctt tgcaaggac caaaataagg gaaggatgaa gtctgaatag gatacgaact ggctaccat gggtgagcat gaaatgaagg ggtcggggaa atggcgagaa aacgcgtcgac gtaacgtgtt tggtttctc ctgtttcgat
--------	--

aaaaaaaaagga tctcaagaag atccttgat ctttctacg gggctgacg ctcatggaa	6540
cggaaaactca cgtaaggcc tgcaaggccg atttggtca tgagattatc aaaaaggata	6600
ttcacctaga tcctttaaa ttaaaaatga agtttaaat caatctaag tatatatgag	6660
taaacttgtt ctgacagttt ccaatgctt atcagtgagg cacctatctc agcgatctgt	6720
ctatccgtt catccatgt tgccctactc cccgtgtgt agataactac gatacggag	6780
ggcttaccat ctggcccccag tgctgcaatg ataccgcag acccacfctc accggctcca	6840
gatttatcag caataaacca gccagccgga agggccgagc gcagaagtgg tcctgcaact	6900
ttatccgcct ccattccatgt tattaattgt tgccggaaag cttagatgt tagttcgcca	6960
gttaatagtt tgcgcaacgt tggtgccatt gctacaggca tcgtgggtgc acgctcg	7020
tttggatatgg ctccatttcag ctccgttcc caacgtcaa ggcgagttac atgatcccc	7080
atgttgtgca aaaaagcgtt tagctccctt ggtccctcga tcgttgtcag aagtaatgg	7140
gccgcagtgt tatcactcat gtttatggca gcaactgcata attctttac tgtcatgca	7200
tccgttaagat gctttctgt gactggtag tactcaacca agtcattctg agaatagtgt	7260
atgcggcgac cgagttgctc ttgcccggcg tcaatacggg ataataccgc gccacatagc	7320
agaactttaa aagtgtcat cattggaaaa ctttcccg ggcgaaaact ctcaggatc	7380
ttaccgtgt tgagatccag ttcgatgtaa cccactcgta cacccaaactg atttcagca	7440
tctttactt tcaccagcgt ttctgggtga gcaaaaacag gaaggcaaaa tgcccaaaaa	7500
aaggaaataa gggcgacacg gaaatgttga atactcatac tcttccttt tcaatattat	7560
tgaaggcatt atcagggtt ttgtctcatg gccattnagg cct	7603

<210> 45

<211> 7000

<212> DNA

[0040]

<213> 人工序列

<220>

<223> 质粒 pYL4

<400> 45

agactagcgg ccggccccct tatccagct gttccacgtt ggcctgcccc tcagtttagcg	60
ctcaactcaa tggccctcac tggcgaggcg agggcaagga tggagggca gcatgcctg	120
agttggagca aagcggccgc catggagca gcaaccaac ggaggatgc cgtgcttgc	180
cgtggctgt gtggcaatc cggccctt gttggctcac agagcgttgc tgtgagacca	240
ttagcttata ttgcttagta cagtatagag agaggagaga gagagagaga gagagagggg	300
aaaaaaaggta aggttgaagt gagaaaaaaaa aaaaaaaaaaa aaatccaacc actgacggct	360
gcggcgtcg ccacccccc ccctccaccc cagacaacct gcacactcg cgcgcagcat	420
cacctaataatc tggctcgctt tcccgcagct caggttgc ttttttctc tctccctgt	480
cgaagccgcc cttgtccctt tatttatttc cctctccatc cttgtctgcc tttggccat	540
ctggcccttt gtctgcatct cttttgcacg catgcctta tcgtcgctc ttttttact	600
cacgggagct tgacgaagac ctgactcgta agcctaacct gctgatttct ctccccccct	660
cccgaccggc ttgacttttgc ttctccctcc agtaccttac cgcgaagccg gaagaacct	720
ctttaacccc atcaaacaag ttgtacaaa aaagcaggct ccgcggccgc ccccttccacc	780
atgctgcgt actccccctgt cttacacctg gatactctt cttgccacc actgaccaat	840
gcttctcccc gcccaaagtgc cgagtaccc agcgctgtcg atagctgcac gcactgccc	900
gatgcccacg tgcagtgacac ttgcacgttgc cccctggccgc gacgcccggcc caaaggcagg	960
aagaagagcg accagccgg ccagccgcct cctgatccga gctgctc caccggcgt	1020
cgaccggcc agatgcccgc gccgtgacc ttctccggcc ccgcagtagc cgcgcgtgcag	1080

[0041]	cccttcgcct cgtcgctcgct gtcgccccgac gcccgcggg agcccgctga gccgctcagc attgacaacg gcctgccccg gcagccgctg ggcgacctgc ccggcctctc caccatccag aacatctcga cgcgccagcg atggatacac ctggccaacg ccatgacgct ggcgaacacg acgctagagc gcgtctcgaa gcgtatgtatc gaccttttct tcgactacct ctacccctc accccccctgg tgtacgagcc ggccctccgg gacgtgctcg catacatttt ctcccaagccc ttgcctggcg tcaaccaacc atgcgcgtg tcacagctca cggccagaccc gaccacccggc accacccccc tcaacgcgtc cgagtcgtgg gcccggcttg gccagccagc cggctcgca accgtcgca gcaggctggc tccctggcc gactcgaccc tcacccttgt cacggccgtc tgcgcagagg cagcattcat gctacccaag gacattttcc cccaggaga atccgtct gagatcttgc tcgaaggctc tcgggactgc ctgcaccaggc acctcgaggc cgacctggag aatccgacgg ccaactcgat tgccattcgc tacttccact ccaactgcct ccacgcgtcg gggaagccca agtactcggt gcacatattt ggcgaggcca tccgcctggc gcaggcatg cagctgcacg aggaggctgc cctcgagggg ctgcgtccca tcgaggcaga gttccggcgt cgctgtttt ggatcctgta cttggcgcac aagttagccg ctataactcaa caatccggcc atcaccatcc acaagtagctg ctgcacgc ggcacatccca cgctataccg tcgggttac gaggacgagt tcctgagcac ggcgtccgag ccgcggccga agagttcat atccggcttc aacgcaaatg tgccgtctg gcagtcgcg gctgatttc tgctggaaat ccgcgtctg caagatcaga ttagtgcagca cttecgaggg accatcccc cgaaccatgt gtcgcctcc gccgacagggc agcatctcga ttctcttat gtccgcttca tcacctgtt ggacgatctc ccggcgtacc tccagtcgtg cacttcggcg atggcagcga tggcagaagg caacgggtct gccgagtcca agcagtagt gatacagtgc atcaacactgc aggtgacggt tcactgtct cgcatggtaa ttacgcagaa attcgaagac ctctttatt ttgcctctgg cgttgagcag gctgatctca gaaagtgcga gattgtgcga gacatgtca gggtgatgaa cgaggccccc ttttggggcc tgcaggccaa tggcggccaa aacgtgagtc gtttccttgt ctcttcctt ttctgcacac cctttcttc gacgacccccc ectctctt tatatccctg cggatatgt tatcatcaag cctcggcaact tggtgtaat ctgtctgtat tatgtgtct ggatgctgca ggttgaaaag attcgcctta tcggagctag tttgtggcc atcatccatc gcaaccagg ttcacccctt gctacgcgag ccaggagcga ctttccgtg ctggata ttctcacgcg gctggactcg aaggcgtcg accaactgag gaatacgtcc actaccgtt tggctaaat gtgtgttggaa acaacaaaaa atgtcaaagt cgggtaaat atggccagga tctttgttt attccccctt cagcgttgc gggattttcc ctttgcattt ctctttctg tttttccag cacttgcattt tccagcgtg ggggaaacaa aaggcgttcc ttcccttat gccagggtt gtccgattta gcatttgagt gtacatctt cctacattac taggtactta atgagctt ggagatctcc cgtcattccg gatattcattt acgttgggtt atatatccgt gttggctt gaaacctgga gttgggttgc aatgcgtga cgcctttgc gaaggccaa aataagcgaa ggatgaagtc tgaataggat acgaaactggc tacctatggg tgagcatgaa atgaagcggt cggggaaatg gcccggccaa gctcgacgtc acgctgttgg ttttctccgt ttctgcatt gttgctcgca taacctacgc tagctgttgc cgcctttgc ttacgacaa gatctagaag attcgagata gaataataat aataacaaca attgcctct tctttccacc ttttgcgt tactctccct tctgacattt aacgcctcaa tcagtcgtc gccttgcact tggcacggta atccctccgt tcttgcattt cctcagggtt agccaaagccc ttcatgcatt cgataatgtc atccagagtg aggtggccaa agatggggat gcccgtactcc ttccctcagct cgccaaatggc actcggtcca ggcttggagt cgtcggccatc cgcagcgggg agcttctcca tggcgtccag ggccacacg atgcggccga cgatggccgc ctccctgggtt atcttctcaa tggcgtccct 3720
--------	---

aatagttgc gcaacgttgt tgccattgct acaggcatcg tgggtc acg ctcgtcgtt	6420
ggtatggctt cattcagctc cggttccaa cgtcaaggc gagttacatg atccccatg	6480
tttgtcaaaa aagcggttag ctccctcggt cctccgatcg ttgtcagaag taagtggcc	6540
gcagtgttat cactcatgtt tatggcagca ctgcataatt ctcttactgt catgccatcc	6600
gtaagatgct ttctgtgac tggtagtac tcaaccaagt cattctgaga atagtgtatg	6660
cggcggaccga gttgctctg cccggcgtca atacgggata ataccgcgc acatagcaga	6720
actttaaaag tgctcatcat tggaaaacgt tcttcgggc gaaaactctc aaggatctta	6780
ccgctgttga gatccagttc gatgttaaccc actcggtcac ccaactgatc ttca gatct	6840
tttactttca ccagcggtt cgggtgagca aaaacaggaa ggcaaaatgc cgcaaaaaag	6900
ggaataaggc cgacacggaa atgttgaata ctcatactct tccttttca atattattga	6960
agcatttatac agggttatttgc tctcatggcc atttaggcct	7000
<210> 46	
<211> 940	
<212> PRT	
<213> 里氏木霉	
<400> 46	
Met Leu Ser Asn Pro Leu Arg Arg Tyr Ser Ala Tyr Pro Asp Ile Ser	
1 5 10 15	
Ser Ala Ser Phe Asp Pro Asn Tyr His Gly Ser Gln Ser His Leu His	
20 25 30	
Ser Ile Asn Val Asn Thr Phe Gly Asn Ser His Pro Tyr Pro Met Gln	
35 40 45	
[0043] His Leu Ala Gln His Ala Glu Leu Ser Ser Ser Arg Met Ile Arg Ala	
50 55 60	
Ser Pro Val Gln Pro Lys Gln Arg Gln Gly Ser Leu Ile Ala Ala Arg	
65 70 75 80	
Lys Asn Ser Thr Gly Thr Ala Gly Pro Ile Arg Arg Arg Ile Ser Arg	
85 90 95	
Ala Cys Asp Gln Cys Asn Gln Leu Arg Thr Lys Cys Asp Gly Leu His	
100 105 110	
Pro Cys Ala His Cys Ile Glu Phe Gly Leu Gly Cys Glu Tyr Val Arg	
115 120 125	
Glu Arg Lys Lys Arg Gly Lys Ala Ser Arg Lys Asp Ile Ala Ala Gln	
130 135 140	
Gln Ala Ala Ala Ala Ala Gln His Ser Gly Gln Val Gln Asp Gly	
145 150 155 160	
Pro Glu Asp Gln His Arg Lys Leu Ser Arg Gln Gln Ser Glu Ser Ser	
165 170 175	
Arg Gly Ser Ala Glu Leu Ala Gln Pro Ala His Asp Pro Pro His Gly	
180 185 190	
His Ile Glu Gly Ser Val Ser Ser Phe Ser Asp Asn Gly Leu Ser Gln	
195 200 205	
His Ala Ala Met Gly Gly Met Asp Gly Leu Glu Asp His His Gly His	
210 215 220	

Val Gly Val Asp Pro Ala Leu Gly Arg Thr Gln Leu Glu Ala Ser Ser
 225 230 235 240
 Ala Met Gly Leu Gly Ala Tyr Gly Glu Val His Pro Gly Tyr Glu Ser
 245 250 255
 Pro Gly Met Asn Gly His Val Met Val Pro Pro Ser Tyr Gly Ala Gln
 260 265 270
 Thr Thr Met Ala Gly Tyr Ser Gly Ile Ser Tyr Ala Ala Gln Ala Pro
 275 280 285
 Ser Pro Ala Thr Tyr Ser Ser Asp Gly Asn Phe Arg Leu Thr Gly His
 290 295 300
 Ile His Asp Tyr Pro Leu Ala Asn Gly Ser Ser Pro Ser Trp Gly Val
 305 310 315 320
 Ser Leu Ala Ser Pro Ser Asn Gln Phe Gln Leu Gln Leu Ser Gln Pro
 325 330 335
 Ile Phe Lys Gln Ser Asp Leu Arg Tyr Pro Val Leu Glu Pro Leu Leu
 340 345 350
 Pro His Leu Gly Asn Ile Leu Pro Val Ser Leu Ala Cys Asp Leu Ile
 355 360 365
 Asp Leu Tyr Phe Ser Ser Ser Ser Ala Gln Met His Pro Met Ser
 370 375 380
 Pro Tyr Val Leu Gly Phe Val Phe Arg Lys Arg Ser Phe Leu His Pro
 385 390 395 400
[0044] Thr Asn Pro Arg Arg Cys Gln Pro Ala Leu Leu Ala Ser Met Leu Trp
 405 410 415
 Val Ala Ala Gln Thr Ser Glu Ala Ser Phe Leu Thr Ser Leu Pro Ser
 420 425 430
 Ala Arg Ser Lys Val Cys Gln Lys Leu Leu Glu Leu Thr Val Gly Leu
 435 440 445
 Leu Gln Pro Leu Ile His Thr Gly Thr Asn Ser Pro Ser Pro Lys Thr
 450 455 460
 Ser Pro Val Val Gly Ala Ala Ala Leu Gly Val Leu Gly Val Ala Met
 465 470 475 480
 Pro Gly Ser Leu Asn Met Asp Ser Leu Ala Gly Glu Thr Gly Ala Phe
 485 490 495
 Gly Ala Ile Gly Ser Leu Asp Asp Val Ile Thr Tyr Val His Leu Ala
 500 505 510
 Thr Val Val Ser Ala Ser Glu Tyr Lys Gly Ala Ser Leu Arg Trp Trp
 515 520 525
 Gly Ala Ala Trp Ser Leu Ala Arg Glu Leu Lys Leu Gly Arg Glu Leu
 530 535 540
 Pro Pro Gly Asn Pro Pro Ala Asn Gln Glu Asp Gly Glu Gly Leu Ser
 545 550 555 560
 Glu Asp Val Asp Glu His Asp Leu Asn Arg Asn Asn Thr Arg Phe Val
 565 570 575

Thr Glu Glu Glu Arg Glu Glu Arg Arg Arg Ala Trp Trp Leu Val Tyr
 580 585 590
 Ile Val Asp Arg His Leu Ala Leu Cys Tyr Asn Arg Pro Leu Phe Leu
 595 600 605
 Leu Asp Ser Glu Cys Ser Asp Leu Tyr His Pro Met Asp Asp Ile Lys
 610 615 620
 Trp Gln Ala Gly Lys Phe Arg Ser His Asp Ala Gly Asn Ser Ser Ile
 625 630 635 640
 Asn Ile Asp Ser Ser Met Thr Asp Glu Phe Gly Asp Ser Pro Arg Ala
 645 650 655
 Ala Arg Gly Ala His Tyr Glu Cys Arg Gly Arg Ser Ile Phe Gly Tyr
 660 665 670
 Phe Leu Ser Leu Met Thr Ile Leu Gly Glu Ile Val Asp Val His His
 675 680 685
 Ala Lys Ser His Pro Arg Phe Gly Val Gly Phe Arg Ser Ala Arg Asp
 690 695 700
 Trp Asp Glu Gln Val Ala Glu Ile Thr Arg His Leu Asp Met Tyr Glu
 705 710 715 720
 Glu Ser Leu Lys Arg Phe Val Ala Lys His Leu Pro Leu Ser Ser Lys
 725 730 735
 Asp Lys Glu Gln His Glu Met His Asp Ser Gly Ala Val Thr Asp Met
 740 745 750
[0045]
 Gln Ser Pro Leu Ser Val Arg Thr Asn Ala Ser Ser Arg Met Thr Glu
 755 760 765
 Ser Glu Ile Gln Ala Ser Ile Val Val Ala Tyr Ser Thr His Val Met
 770 775 780
 His Val Leu His Ile Leu Leu Ala Asp Lys Trp Asp Pro Ile Asn Leu
 785 790 795 800
 Leu Asp Asp Asp Asp Leu Trp Ile Ser Ser Glu Gly Phe Val Thr Ala
 805 810 815
 Thr Ser His Ala Val Ser Ala Val Glu Ala Ile Ser Gln Ile Leu Glu
 820 825 830
 Phe Asp Pro Gly Leu Glu Phe Met Pro Phe Phe Tyr Gly Val Tyr Leu
 835 840 845
 Leu Gln Gly Ser Phe Leu Leu Leu Ile Ala Asp Lys Leu Gln Ala
 850 855 860
 Glu Ala Ser Pro Ser Val Ile Lys Ala Cys Glu Thr Ile Val Arg Ala
 865 870 875 880
 His Glu Ala Cys Val Val Thr Leu Ser Thr Glu Tyr Gln Arg Asn Phe
 885 890 895
 Ser Lys Val Met Arg Ser Ala Leu Ala Leu Ile Arg Gly Arg Val Pro
 900 905 910
 Glu Asp Leu Ala Glu Gln Gln Arg Arg Arg Glu Leu Leu Ala Leu
 915 920 925

Tyr Arg Trp Thr Gly Asn Gly Thr Gly Leu Ala Leu
 930 935 940
 <210> 47
 <211> 341
 <212> PRT
 <213> 里氏木霉
 <400> 47
 Met Asp Leu Arg Gln Ala Cys Asp Arg Cys His Asp Lys Lys Leu Arg
 1 5 10 15
 Cys Pro Arg Ile Ser Gly Ser Pro Cys Cys Ser Arg Cys Ala Lys Ala
 20 25 30
 Asn Val Ala Cys Val Phe Ser Pro Pro Ser Arg Pro Phe Arg Pro His
 35 40 45
 Glu Pro Leu Asn His Ser His Glu His Ser His Ser His Asn
 50 55 60
 His Asn Gly Val Gly Val Ser Phe Asp Trp Leu Asp Leu Met Ser Leu
 65 70 75 80
 Glu Gln Gln Gln Glu Gln Gln Gly Gln Pro Gln His Pro Pro Pro
 85 90 95
 Pro Val Gln Thr Leu Ser Glu Arg Leu Ala Ala Leu Leu Cys Ala Leu
 100 105 110
 [0046] Asp Arg Met Leu Gln Ala Val Pro Ser Ser Leu Asp Met His His Val
 115 120 125
 Ser Arg Gln Gln Leu Arg Glu Tyr Ala Asp Thr Val Gly Thr Gly Phe
 130 135 140
 Asp Leu Gln Ser Thr Leu Asp Ser Leu Leu His His Ala Gln Asp Leu
 145 150 155 160
 Ala Ser Leu Tyr Ser Glu Ala Val Pro Ala Ser Phe Asn Lys Arg Thr
 165 170 175
 Thr Ala Ala Glu Ala Asp Ala Leu Cys Ala Val Pro Asp Cys Val His
 180 185 190
 Gln Asp Arg Thr Ser Leu His Thr Thr Pro Leu Pro Lys Leu Asp His
 195 200 205
 Ala Leu Leu Asn Leu Val Met Ala Cys His Ile Arg Leu Leu Asp Val
 210 215 220
 Met Asp Thr Leu Ala Glu His Gly Arg Met Cys Ala Phe Met Val Ala
 225 230 235 240
 Thr Leu Pro Pro Asp Tyr Asp Pro Lys Phe Ala Val Pro Glu Ile Arg
 245 250 255
 Val Gly Thr Phe Val Ala Pro Thr Asp Thr Ala Ala Ser Met Leu Leu
 260 265 270
 Ser Val Val Val Glu Leu Gln Thr Val Leu Val Ala Arg Val Lys Asp
 275 280 285
 Leu Val Ala Met Val Asp Gln Val Lys Asp Asp Ala Arg Ala Ala Arg

	290	295	300
Glu Ala Lys Val Val Arg Leu Gln Cys Gly Ile Leu Leu Glu Arg Ala			
305	310	315	320
Glu Ser Thr Leu Gly Glu Trp Ser Arg Phe Lys Asp Gly Leu Val Ser			
325	330	335	
Ala Arg Leu Leu Lys			
340			
<210> 48			
<211> 940			
<212> PRT			
<213> 里氏木霉			
<400> 48			
Met Leu Ser Asn Pro Leu Arg Arg Tyr Ser Ala Tyr Pro Asp Ile Ser			
1	5	10	15
Ser Ala Ser Phe Asp Pro Asn Tyr His Gly Ser Gln Ser His Leu His			
20	25	30	
Ser Ile Asn Val Asn Thr Phe Gly Asn Ser His Pro Tyr Pro Met Gln			
35	40	45	
His Leu Ala Gln His Ala Glu Leu Ser Ser Ser Arg Met Ile Arg Ala			
50	55	60	
Ser Pro Val Gln Pro Lys Gln Arg Gln Gly Ser Leu Ile Ala Ala Arg			
65	70	75	80
[0047] Lys Asn Ser Thr Gly Thr Ala Gly Pro Ile Arg Arg Arg Ile Ser Arg			
85	90	95	
Ala Cys Asp Gln Cys Asn Gln Leu Arg Thr Lys Cys Asp Gly Leu His			
100	105	110	
Pro Cys Ala His Cys Ile Glu Phe Gly Leu Gly Cys Glu Tyr Val Arg			
115	120	125	
Glu Arg Lys Lys Arg Gly Lys Ala Ser Arg Lys Asp Ile Ala Ala Gln			
130	135	140	
Gln Ala Ala Ala Ala Ala Gln His Ser Gly Gln Val Gln Asp Gly			
145	150	155	160
Pro Glu Asp Gln His Arg Lys Leu Ser Arg Gln Gln Ser Glu Ser Ser			
165	170	175	
Arg Gly Ser Ala Glu Leu Ala Gln Pro Ala His Asp Pro Pro His Gly			
180	185	190	
His Ile Glu Gly Ser Val Ser Ser Phe Ser Asp Asn Gly Leu Ser Gln			
195	200	205	
His Ala Ala Met Gly Gly Met Asp Gly Leu Glu Asp His His Gly His			
210	215	220	
Val Gly Val Asp Pro Ala Leu Gly Arg Thr Gln Leu Glu Ala Ser Ser			
225	230	235	240
Ala Met Gly Leu Gly Ala Tyr Gly Glu Val His Pro Gly Tyr Glu Ser			
245	250	255	

Pro Gly Met Asn Gly His Val Met Val Pro Pro Ser Tyr Gly Ala Gln
 260 265 270
 Thr Thr Met Ala Gly Tyr Ser Gly Ile Ser Tyr Ala Ala Gln Ala Pro
 275 280 285
 Ser Pro Ala Thr Tyr Ser Ser Asp Gly Asn Phe Arg Leu Thr Gly His
 290 295 300
 Ile His Asp Tyr Pro Leu Ala Asn Gly Ser Ser Pro Ser Trp Gly Val
 305 310 315 320
 Ser Leu Ala Ser Pro Ser Asn Gln Phe Gln Leu Gln Leu Ser Gln Pro
 325 330 335
 Ile Phe Lys Gln Ser Asp Leu Arg Tyr Pro Val Leu Glu Pro Leu Leu
 340 345 350
 Pro His Leu Gly Asn Ile Leu Pro Val Ser Leu Ala Cys Asp Leu Ile
 355 360 365
 Asp Leu Tyr Phe Ser Ser Ser Ser Ala Gln Met His Pro Met Ser
 370 375 380
 Pro Tyr Val Leu Gly Phe Val Phe Arg Lys Arg Ser Phe Leu His Pro
 385 390 395 400
 Thr Asn Pro Arg Arg Cys Gln Pro Ala Leu Leu Ala Ser Met Leu Trp
 405 410 415
 Val Ala Ala Gln Thr Ser Glu Ala Ser Phe Leu Thr Ser Leu Pro Ser
 420 425 430
 [0048] Ala Arg Ser Lys Val Cys Gln Lys Leu Leu Glu Leu Thr Val Gly Leu
 435 440 445
 Leu Gln Pro Leu Ile His Thr Gly Thr Asn Ser Pro Ser Pro Lys Thr
 450 455 460
 Ser Pro Val Val Gly Ala Ala Ala Leu Gly Val Leu Gly Val Ala Met
 465 470 475 480
 Pro Gly Ser Leu Asn Met Asp Ser Leu Ala Gly Glu Thr Gly Ala Phe
 485 490 495
 Gly Ala Ile Gly Ser Leu Asp Asp Val Ile Thr Tyr Val His Leu Ala
 500 505 510
 Thr Val Val Ser Ala Ser Glu Tyr Lys Gly Ala Ser Leu Arg Trp Trp
 515 520 525
 Gly Ala Ala Trp Ser Leu Ala Arg Glu Leu Lys Leu Gly Arg Glu Leu
 530 535 540
 Pro Pro Gly Asn Pro Pro Ala Asn Gln Glu Asp Gly Glu Gly Leu Ser
 545 550 555 560
 Glu Asp Val Asp Glu His Asp Leu Asn Arg Asn Asn Thr Arg Phe Val
 565 570 575
 Thr Glu Glu Glu Arg Glu Arg Arg Ala Trp Trp Leu Val Tyr
 580 585 590
 Ile Val Asp Arg His Leu Ala Leu Cys Tyr Asn Arg Pro Leu Phe Leu
 595 600 605

Leu Asp Ser Glu Cys Ser Asp Leu Tyr His Pro Met Asp Asp Ile Lys
 610 615 620
 Trp Gln Ala Gly Lys Phe Arg Ser His Asp Ala Gly Asn Ser Ser Ile
 625 630 635 640
 Asn Ile Asp Ser Ser Met Thr Asp Glu Phe Gly Asp Ser Pro Arg Ala
 645 650 655
 Ala Arg Gly Ala His Tyr Glu Cys Arg Gly Arg Ser Ile Phe Gly Tyr
 660 665 670
 Phe Leu Ser Leu Met Thr Ile Leu Gly Glu Ile Val Asp Val His His
 675 680 685
 Ala Lys Ser His Pro Arg Phe Gly Val Gly Phe Arg Ser Ala Arg Asp
 690 695 700
 Trp Asp Glu Gln Val Ala Glu Ile Thr Arg His Leu Asp Met Tyr Glu
 705 710 715 720
 Glu Ser Leu Lys Arg Phe Val Ala Lys His Leu Pro Leu Ser Ser Lys
 725 730 735
 Asp Lys Glu Gln His Glu Met His Asp Ser Gly Ala Val Thr Asp Met
 740 745 750
 Gln Ser Pro Leu Ser Val Arg Thr Asn Ala Ser Ser Arg Met Thr Glu
 755 760 765
 Ser Glu Ile Gln Ala Ser Ile Val Val Ala Tyr Ser Thr His Val Met
 770 775 780
 [0049] His Val Leu His Ile Leu Leu Ala Asp Lys Trp Asp Pro Ile Asn Leu
 785 790 795 800
 Leu Asp Asp Asp Asp Leu Trp Ile Ser Ser Glu Gly Phe Val Thr Ala
 805 810 815
 Thr Ser His Ala Val Ser Ala Ala Glu Ala Ile Ser Gln Ile Leu Glu
 820 825 830
 Phe Asp Pro Gly Leu Glu Phe Met Pro Phe Phe Tyr Gly Val Tyr Leu
 835 840 845
 Leu Gln Gly Ser Phe Leu Leu Leu Ile Ala Asp Lys Leu Gln Ala
 850 855 860
 Glu Ala Ser Pro Ser Val Ile Lys Ala Cys Glu Thr Ile Val Arg Ala
 865 870 875 880
 His Glu Ala Cys Val Val Thr Leu Ser Thr Glu Tyr Gln Arg Asn Phe
 885 890 895
 Ser Lys Val Met Arg Ser Ala Leu Ala Leu Ile Arg Gly Arg Val Pro
 900 905 910
 Glu Asp Leu Ala Glu Gln Gln Arg Arg Arg Glu Leu Leu Ala Leu
 915 920 925
 Tyr Arg Trp Thr Gly Asn Gly Thr Gly Leu Ala Leu
 930 935 940
 <210> 49
 <211> 65

<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	49	
	gtaacgccag gttttccca gtcacgacgg tttaactcc atacgcaggca aacatggct	60
	tggc	65
<210>	50	
<211>	64	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	50	
	gtacgagtac tagtgtgaa gattccgtca agcttggcg gaatgaagga ggatgtgtga	60
	gagg	64
<210>	51	
<211>	59	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	51	
	cacacatct ctttcattcc gcccaagctt gacggaatct tcacacctag tactcgat	59
<210>	52	
<211>	57	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	52	
	tgacatttt tggttcca acacagcatg cttagtccga cgccctcgag tccagcc	57
<210>	53	
<211>	54	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	53	
	ctggactcga aggcgctcgga ctaagcatgc tgtgttggaa caacaaaaaa tgtc	54
<210>	54	
<211>	54	
<212>	DNA	
<213>	人工序列	

<220>		
<223> 引物		
<400> 54		
gcagagcgcg agtagtcgt gctattaatt aagttaggtt tgcgagcaac attg		54
<210> 55		
<211> 58		
<212> DNA		
<213> 人工序列		
<220>		
<223> 引物		
<400> 55		
ctcagccctc ctcagcccta tcagccgcgg ccgctgaatc ggcaagggtt agtactag		58
<210> 56		
<211> 62		
<212> DNA		
<213> 人工序列		
<220>		
<223> 引物		
<400> 56		
gcccataaca atttcacaca ggaaacagcg tttaaaccac atgccagagt tcgatgcga		60
ag		62
[0051]		
<210> 57		
<211> 57		
<212> DNA		
<213> 人工序列		
<220>		
<223> 引物		
<400> 57		
gtacccatc gctgtcgata gctgcacgca ctggcgcat gcccacgtgc agtgcac		57
<210> 58		
<211> 59		
<212> DNA		
<213> 人工序列		
<220>		
<223> 引物		
<400> 58		
gtgcactgca cgtggcata gcggcagtgc gtgcagctat cgacagcgct gaggtact		59
<210> 59		
<211> 48		
<212> DNA		
<213> 人工序列		
<220>		
<223> 引物		
<400> 59		

gccccgccttc cgctgtcgta actatgtgc gctactcccc cgtcttac	48
<210> 60	
<211> 58	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物	
<400> 60	
gtaagacggg ggagtagcgc agcatagtta cgacagcgga agcgccgcct tataagtg	58
<210> 61	
<211> 53	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物	
<400> 61	
ggcggcgctt ccgctgtcgt aactatggc tcagcagctc cggcccaggg ctc	53
<210> 62	
<211> 58	
<212> DNA	
<213> 人工序列	
<220>	
[0052] <223> 引物	
<400> 62	
gccctggccc ggagctgctg agcccatagt tacgacagcg gaagcgccgc cttataag	58
<210> 63	
<211> 53	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物	
<400> 63	
ggcggcgctt ccgctgtcgt aactatggcc acagcggccg cggcagcagc tgg	53
<210> 64	
<211> 59	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物	
<400> 64	
cagctgtgc cgccggccgct gtggccatag ttacgacagc ggaagcgccg cttataag	59
<210> 65	
<211> 22	
<212> DNA	

<213>	人工序列	
<220>		
<223>	引物	
<400>	65	
	gctggaaagct gctgaggcaga tc	22
<210>	66	
<211>	22	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	66	
	gtgccagcat tccccagact cg	22
<210>	67	
<211>	20	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	67	
	ttaggcgacc tcttttcca	20
[0053]		
<210>	68	
<211>	22	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	68	
	gccgctcagg catacgagcg ac	22
<210>	69	
<211>	20	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	69	
	ctctggtcgg cctgccgttg	20
<210>	70	
<211>	23	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	70	

ttagtatagc ggctgacttg tcg	23
<210> 71	
<211> 59	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物	
<400> 71	
tctagtagt acgagtacta ggtgtgaaga ttccgtcatt tcctcgacat gcgaatgcg	59
<210> 72	
<211> 59	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物	
<400> 72	
tgcctatcaa accccgcatt cgcatgtcga ggaaatgacg gaatttcac acctagta	59
<210> 73	
<211> 59	
<212> DNA	
<213> 人工序列	
<220>	
[0054] <223> 引物	
<400> 73	
tgcagctaca gagccctggg ccggagctgc tgagccata gttacgacag cggaagcgc	59
<210> 74	
<211> 57	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物	
<400> 74	
atagcactta taaggcggcg cttccgtgt cgtaactatg ggctcagcag ctccggc	57
<210> 75	
<211> 75	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物	
<400> 75	
taaacaaata ggggttccgc gcacatttcc ccgaaaatgt ccacctggat agactagcat	60
ctgagccatt gcagc	75
<210> 76	
<211> 60	

<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	76	
	agtggcaccc agtcgggtggt gctttttttt ctatcgagag cattggtag tggggcaag	60
<210>	77	
<211>	60	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	77	
	accaatatac aaaacatgtc gtccgagcca gtgcctgccca ttccctcgac atgcgaatgc	60
<210>	78	
<211>	60	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	78	
	gttgccatgc aaaccccgca ttgcgtatgtc gagaaatgg caggcactgg ctggacgac	60
[0055]	<210> 79	
	<211> 60	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 79	
	agctacagag ccctggcccg gagctgctga gcccattgtt gaattctggc gggtagctg	60
	<210> 80	
	<211> 60	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 80	
	cttttacact ttcaaacagc tacccggcca gaattcaaca atgggctcag cagctccggc	60
	<210> 81	
	<211> 81	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	

<400>	81	
tagtccgtta tcaaacttgaa aaagtggcac cgagtcggtg gtgcttttt ttctatcgag		60
atgttctgga tggtgagag g		81
<210>	82	
<211>	32	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	82	
agtctatcgc agccttgccc tagctaattt tt		32
<210>	83	
<211>	32	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	83	
tctaaaacat tagctaaggc aaggctgcga ta		32
<210>	84	
<211>	32	
[0056]		
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	84	
agtctatcg cagagtcgca tcttcgggt tt		32
<210>	85	
<211>	32	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	85	
tctaaaaccc ggaagacgca actctgccga ta		32
<210>	86	
<211>	33	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	86	
agtctatcga atgagtgttag gtacgagtag ttt		33
<210>	87	

<211>	33	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	87	
tctaaaacta ctcgtaccta cactcattcg ata		33
<210>	88	
<211>	32	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	88	
agtctatcg ccgcaatagc ttcctaattt tt		32
<210>	89	
<211>	32	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	89	
[0057]		
tctaaaacat taggaagcta ttgcggccga ta		32
<210>	90	
<211>	32	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	90	
agtctatcgc agcgcaatca gtgcagtggc tt		32
<210>	91	
<211>	32	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	91	
tctaaaacca ctgcactgat tgcgctgcga ta		32
<210>	92	
<211>	23	
<212>	DNA	
<213>	人工序列	
<220>		

<223>	引物	
<400>	92	
tgagagact	cggagaggat	agg 23
<210>	93	
<211>	22	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	93	
agcggtggagg	cagttggagt	gg 22
<210>	94	
<211>	24	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	94	
tggacaaagc	ctgggtcctg	ctcc 24
<210>	95	
<211>	22	
[0058]	22	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	95	
atccctgactc	gtcctgtgtc	gg 22
<210>	96	
<211>	22	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	96	
agtgttgtgt	ttagtggact	tg 22
<210>	97	
<211>	22	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物	
<400>	97	
ctcgtagct	gcttgaatat	ag 22
<210>	98	

<211>	45			
<212>	PRT			
<213>	人工序列			
<220>				
<223>	EL 形式 N-末端延伸			
<400>	98			
	Met Ala Thr Ala Ala Ala Ala Ala Gly Gly Ala Ala Val Ala Ala			
1	5	10	15	
Gly Ala Asp Thr Gly Ala Ala Gly Ser Ser Ser Thr Gly Pro Pro Gly				
20	25	30		
Leu Pro Gly Leu Pro Gly Thr Arg Thr Gly Ser Val Ala				
35	40	45		
<210>	99			
<211>	1857			
<212>	DNA			
<213>	里氏木霉			
<400>	99			
[0059]	atgctgcgt actccccgt cttacacctg gatactctc cttgccacc actgaccaat	60		
	gctttcccc gcccaaagtg cgagtagctc agcgctgtcg atagctgcac gcactgccc	120		
	gatgcccacg tgcagtgcac ttgcacctg cccctggcgc gacgcccggc caaagcgagg	180		
	aagaagagcg accagcccg ccagccgcct cctgatccga gtcgcgtc caccgcggct	240		
	cgaccggcc agatgccccc gccgctgacc ttctccggcc ccgcagtagc cgccgtgcag	300		
	cccttcgcct cgtgcgtcgt gtcgcccac gccgcctggg agccgcgtga gccgctcagc	360		
	attgacaacg gcctgcggcc gcagccgcgt ggcgacactgc ccggcctctc caccatccag	420		
	aacatctcga cgcgccagcg atggatacac ctggccaacg ccatgacgct ggcacacag	480		
	acgctagagc gcgtctcgaa gcgtatgtac gacctttct tcgactacct ctacccctc	540		
	accccccctgg tgtacgagcc ggccctccgg gacgtgtcg catacattt ctcccagccc	600		
	ttgcctggcg tcaaccaacc atgcgcgtc tcacagctca cgccagaccc gaccaccggc	660		
	accacccccc tcaacgcgtc cgagtcgtgg gccggctttg gccagcccg cggctcgca	720		
	accgtcgca gcaggctggc tccctgggcc gactcgacct tcaccctgtt cacggccgtc	780		
	tgcgcagagg cagcattcat gtcaccaag gacattttcc ccgaaggaga atccgtctct	840		
	gagatcttc tcaaggctc tcggactgc ctgcaccagg acctcgaggc cgacctggag	900		
	aatccgacgg ccaactcgat tgccattcgc tacttccact ccaactgcct ccacgtcg	960		
	gsgaagccca agtactcggt gcacatattt ggcgaggcca tccgcctggc gcaggcatg	1020		
	cagctgcacg aggaggctgc ctcgagggg ctgcgtccca tcgaggcaga gttccggcgt	1080		
	cgctgtttt ggatcctgta cttggcgcac aagtccgcg ctataactcaa caatccggcc	1140		
	atcaccatcc acaagactgt ctgcacgc ggcacatcca cgctataccc gtcgggtatc	1200		
	gaggacgagt tcctgagcac ggcgtccgag ccgcggccga agagcttcat atccggcttc	1260		
	aacgc当地 tgccgtctg gcagtcgcg gtcgttttc tgctggaaat ccgcgtcg	1320		
	caagatcaga tgcgtcgac cttcgaggaccatcccc cgaaccatgt gtcgcctcc	1380		
	ggccgacaggc agcatctcga ttctctctat gtccgcttca tcacgtcgat ggacgatctc	1440		
	ccggcggtacc tccagtcgt cactctggcg atggcagcga tggcagaagg caacgggtct	1500		
	ggccgacttca agcagtacgt gatacgtgc atcaacactgc aggtgacgat tcactgtctg	1560		

cgcatggtaa ttacgcagaa attcgaagac ctctttatt ttgctctgg cgttgaggcag	1620
gctgatctca gaaagtgcga gattgtgcga gacatgctga gggtgatgaa cgaggccc	1680
ttttgggcc tgcaggccaa tggcgagcca aacgtgaaa agattgcct tatcgagct	1740
agtttgcgg ccatcatcca tcgcaaccag gattcaccct tggctacgcg agccaggagc	1800
gactttccg tgctttgga tatttcacg cggtggact cgaaggcgtc ggactaa	1857
<210> 100	
<211> 2070	
<212> DNA	
<213> 里氏木霉	
<400> 100	
atgggctcag cagctccggc ccaggcgtc ttagctgcag ctgcaggcgg ccctccagct	60
gctggcgctg gcgcgtggcgc tgtccacgcc ctcaccacct cggccgagtc tgcctggcc	120
tgcagcccc gctgcacaac cgcctcaacc acgcccggc agaactcact cgtgtcggt	180
gcaacctcgt tccaccacca tcccagaggc cgctgtgtga gcagagcctg cgaccgctgc	240
cggccggcga aggccaagtg cgagtaccc acgcgtgtcg atagctgcac gcactgccc	300
gatgcccacg tgcagtgac tttcgacctg cccctggcgc gacgcccggc caaagcgagg	360
aagaagagcg accagcccg ccagccgcct cctgatccga gctgcgtc caccggcgt	420
cgaccggcc agatgccgccc gccgctgacc ttctccggcc cgcagtagc cgcgctgc	480
cccttcgcct cgctgtcgct gtgcggcgc gcccgtgg agccgtcga gccgctcagc	540
attgacaacg gcctgccccg gcagccgtg ggcacactgc cggcccttc caccatccag	600
aacatctcga cgcgccagcg atggatacac ctggccaacg ccatgacgt gcgeaacacg	660
acgctagagc gcgatcgaa gcgatgtatc gaccttttc tcgactaccc ctacccctc	720
accccccctgg tgcagtgaccc ggcctccgg gacgtgcgtc catacattt ctcccagccc	780
ttgcctggcg tcaaccaacc atgcgcgtg tcacagctca cgcagaccc gaccaccggc	840
accacccccc tcaacgctgc cgagtcgtgg gccggctttg gccagccag cggctcgcga	900
accgtcgca gcaggctggc tccctggcc gactcgaccc tcaccctgtt cacggccgtc	960
tgcgcagagg cagcattcat gctaccaag gacatttcc cccaggaga atccgtctc	1020
gagatcttc tgcagccctc tcggactgc ctgcaccaggc acctcgaggc cgacctggag	1080
aatccgcacgg ccaactcgat tgccattcgc tacttccact ccaactgcct ccacgtgcgc	1140
ggaaagccca agtactcggt gcacatattt ggcaggccca tccgcctggc gcaggtcatg	1200
cagctgcacg aggaggctgc ctcgagggg ctgcgtccca tcgaggcaga gttccggcgt	1260
cgcgtcttt ggtatcgatc cttggcgcac aagtgcgtg ctataactcaa caatggccc	1320
atcaccatcc acaagttactg cttcgacgccc ggcatcacca cgctataacc gtcgggtatc	1380
gaggacgatc tcctgagcac ggcgtccgag ccggccggaa agagttcat atccggcttc	1440
aacgcacaaatg tgcggctctg gcagtcgcg gtcgtttgc tgctggaaat ccgcgtgc	1500
caagatcaga tgcgtcgacca ctttcgaggaccatcccc cgaaccatgt gtcgcctcc	1560
gccgacagggc agcatctcga ttctctctat gtccgcttca tcacctgcgtt ggacgtctc	1620
ccggcgtacc tccagtcgtg cacttcggcg atggcagcga tggcagaagg caacgggtct	1680
gccgagtcca agcagtacgt gatacgtgc atcaacactgc aggtgacgtt tcactgtctg	1740
cgcgtttaa ttacgcagaa attcgaagac ctctttatt ttgctctgg cgttgaggcag	1800
gctgatctca gaaagtgcga gattgtgcga gacatgctga gggtgatgaa cgaggccc	1860
ttttgggcc tgcaggccaa tggcgagcca aacgtgaaa agattgcct tatcgagct	1920
agtttgcgg ccatcatcca tcgcaaccag gattcaccct tggctacgcg agccaggagc	1980
gactttccg tgctttgga tatttcacg cggtggact cgaaggcgtc ggaccaactg	2040

aggaatacgt ccactaccgt tgttggctaa	2070
<210> 101	
<211> 2172	
<212> DNA	
<213> 里氏木霉	
<400> 101	
atggccacag cggccgcggc agcagctggc ggcgcccgg ttgctgcggg tgcagacaca	60
ggcgctgcag gctccagctc tacaggccct ccaggccctc cagggcttcc aggacccgg	120
acaggctccg tggcgatggg ctcagcacgt ccggcccagg gctctgttagc tgcagctgca	180
ggcggccctc cagctgctgg cgctggcgct ggcgtgtcc acgcctcac cacctcgccc	240
gagtctgcct cggcctcgca gcccggctcg ccaaccgcct caaccacgcc gccgcagaac	300
tcactctgtgt cggctgcaac ctgcgttccac caccatccca gaggccgtct ggtgaggcaga	360
gcctcgacc gctgcccggc ggcgaaggcc aagtgcgagt acctcagcgc tgcgatagc	420
tgcacgcact gcccgcgtgc ccacgtgcag tgcacttgc acctgcctc ggccgcacgc	480
ggccccaag cgaggaagaa gagcgcaccag cccggccagg cgcctctga tccgagctcg	540
ctctccaccc cggtcgacc cggccagatg cccggccgc tgaccttctc cggccccgca	600
gtagccgcgc tgcagccctt cgcctctcg tcgctgtcgcc cgcacgcgc ctgggagccc	660
gtcgagccgc tcagcattga caacggcctg cccggcaggc cgctggcga cctgcccggc	720
ctctccacca tccagaacat ctgcacgcgc cagcgatgta tacacctggc caacccatg	780
acgctgcgca acacgacgct agagcgcgtc tcaagacgt gtatcgacct ttcttcgac	840
tacctctacc ccctcacccc cctgggtgtac gagccggccc tccgggacgt gtcgcatac	900
atcttctccc agcccttggc tggcgtaac caaccatgc cgctgtcaca gtcacgcca	960
gacccgacca cggcaccac cccctcaac gtcggcagg cgtggccgg ctggccag	1020
cccagcggtc cgcgaaccgt cggcagcagg ctggctccct gggccgactc gacccatcacc	1080
ctggtcacgg cgcgtcgcc agagggcagca ttcatgtac ccaaggacat ttccccgaa	1140
ggagaatccg tctctgagat ctgcgtcgaa gcctctcggt actgcctgca ccagcacctc	1200
gagggccacc tggagaatcc gacggccaac tcgattgcca ttgcgtactt ccactccaac	1260
tgccctccacg ctgcgggaa gcccaagtac tcgtggcaca tatttggcga ggccatccgc	1320
ctggcgccagg tcatgcagct gcacgaggagg gtcgcctcg aggggctcg cccatcgag	1380
gcagagtcc cccgtcgctg ctttggatc ctgtacttgg ggcacaagtc agccgtata	1440
ctcaacaatc gccccatcac catccacaag tactgttc acgcggcat caccacgcta	1500
tacccgtcgg gtatcgagga cgagttcctg acgcacggcgt ccgagccgcc ccggaaagagc	1560
ttcatatccg gttcaacgc aaatgtgcgg ctgcgtcgact ccgcggctga tttgcgtctg	1620
gaaatcccg tgcgtcaaga tcagatgtatc cagcacccgtt gagggaccat gccccgaaac	1680
catgtgtcg cctccggca caggcacgt ctgcatttc tctatgtccg cttcatcacc	1740
tgttggacg atctcccccgtc gtacccatcg tcgtgcactc tggcgatggc agcgatggca	1800
gaaggcaacg ggtctggca gtccaaaggc tacgtgatac agtgcataa cctgcagggt	1860
acgttcaact gtctgcgcac ggttaattacg cagaaattcg aagacccatc ttattttgt	1920
cctggcggtt agcaggctga tctcagaaag tcggagattg tgcgagacat gtcgggtgt	1980
atgaacgagg cggccctttt gggcctgcag gccaatggcg agccaaacgt tgaaaagatt	2040
cgccttatecg gagcttagtt gtcggccatc atccatcgca accaggattc acccttggct	2100
acgcgagcca ggagcgcactt ttccgtgttt ttggatattc tcacgcggct ggactcgaag	2160
gcgtcgact aa	2172
<210> 102	

<211>	2037	
<212>	DNA	
<213>	里氏木霉	
<400>	102	
	atgggctcag cagctccggc ccagggtctt gtagctgcag ctgcaggcgg ccctccagct	60
	gctggcgctg gcgctggcgc tgtccacgcc ctcaccaccc cgccccagtc tgccctggcc	120
	tgcagcccg gctcgccaaac cgccctcaacc acggccggcgc agaactcaact cgtgtcggt	180
	gcaacctcg tccaccacca tcccagaggc cgtctggta gcagagccctg cgaccgctgc	240
	cggccggcga aggccaagtg cgagttaccc tcgctgtcg atagctgcac gcactgcccgc	300
	gatgcccacg tgcagtgcac ttgcacccctg cccctggcgc gacgcccgc caaagcgagg	360
	aagaagagcg accagcccg ccagccgcct cctgatccga gctcgctc caccggcgt	420
	cgaccggcgc agatgccgc gccgctgacc ttctccggcc ccgcagtagc cgccgtcgac	480
	cccttcgcct cgtcgtcgct gtgcggccgac gcccgtggg agcccgctga gcccgtcagc	540
[0062]	attgacaacg gcctgccccg gcagccgcgt ggcgacactgc cccgcctc caccatccag	600
	aacatctcga cgccgcacgc atggatacac ctggccaaacg ccatgacgct ggcacacacg	660
	acgctagagc gcgtctcgaa gcgtatgtatc gacctttct tcgactaccc ctacccctc	720
	accccccctgg tgtacgagcc gcccctccgg gacgtgtcg catacatctt ctccccagccc	780
	ttgcctggcg tcaaccaacc atcggcgctg tcacagctca cgccagaccc gaccaccggc	840
	accacccccc tcaacgcgtgc cgagtcgtgg gccggctttg gccagcccg cggctcgac	900
	accgtcgca gcaggctggc tccctggcc gactcgaccc tcaccctggt cacggccgtc	960
	tgcgcagagg cagcattcat gtcacccaaag gacattttcc cccaaaggaga atccgtctct	1020
	gagatcttc tcgaagcctc tcggactgc ctgcaccaggc acctcgaggc cgacctggag	1080
	aatccgacgg ccaactcgat tgccattcgc tacttccact ccaactgcct ccacgctcg	1140
	gsgaagccca agtactcggt gcacatattt ggcgaggccca tccgcctggc gcaggtcatg	1200
	cagctgcacg aggaggctgc cctcgagggg ctgcgtccca tcgaggcaga gttccggcgt	1260
	cgctgtttt ggatcctgtt cttggcgac aagtcaaggcc ctataactcaa caatcgcccc	1320
	atcaccatcc acaagtactg ctgcacgccc ggcacatccca cgctataacc gtcgggtatc	1380
	gaggacgagt tcctgagcac ggcgtccgag ccggccggaa agagttcat atccggcttc	1440
	aacgcaaattg tgcggcttg gcaatccgcg gctgatttgc tgctggaaat ccgcgtgtcg	1500
	caagatcaga tgcacatccca ctggcggcc accatcccc cgaaccatgt gtcgcctcc	1560
	gccgacaggc agcatctcga ttctcttat gtcgccttca tcacctgtt ggacgatctc	1620
	ccggcgtacc tccagtcgt cactctggcg atggcagcga tggcagaagg caacgggtct	1680
	gccgagtcca agcagtacgt gatacagtgc atcaacactgc aggtgacgat tcactgtctg	1740
	cgcacatggtaa ttacgcagaa attcgaagac ctctttatt ttgcgtctgg cgttgcgcag	1800
	gctgatctca gaaagtcgga gattgtcgaa gacatgctga gggtgatgaa cgaggccccc	1860
	ttttggggcc tgcaggccaa tggcggccca aacgttgaaa agattgcct tatcgagct	1920
	agtttgctgg ccatcatcca tcgcaaccag gattcaccct tggctacgcg agccaggagc	1980
	gactttccg tgctttggta tattctcagc cggctggact cgaaggcgat ggactaa	2037

来自里氏木霉菌株QM6a和RUT-C30的Ace3蛋白编码区的示意图

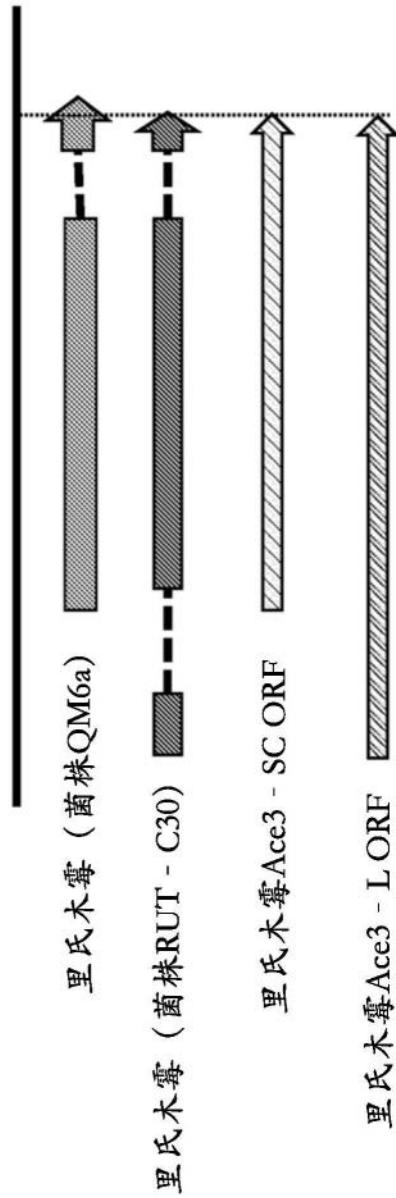


图1A

Ace3-S	1	10	20	30	40	50	60
Ace3-SC							
Ace3-L						-MGSAAAPAQGSVAAA	
Ace3-LN						-MGSAAAPAQGSVAAA	
Ace3-LC						-MGSAAAPAQGSVAAA	
Ace3-EL				MATAAAAGGAAVAAGGADTGAGSSSTGPPGLPGLPRTGSAAMGSAAAPAQGSVAAA			
Ace3-S			-MLRYSPLHLDTLSLPPLTNALP-				
Ace3-SC			-MLRYSPLHLDTLSLPPLTNALP-				
Ace3-L		GGPPAAGAGAGAVHALTSPESASASQPGSPTASTTPQNSLVSAAATSFHHPGRGLVSR					
Ace3-LN		GGPPAAGAGAGAVHALTSPESASASQPGSPTASTTPQNSLVSAAATSFHHPGRGLVSR					
Ace3-LC		GGPPAAGAGAGAVHALTSPESASASQPGSPTASTTPQNSLVSAAATSFHHPGRGLVSR					
Ace3-EL		GGPPAAGAGAGAVHALTSPESASASQPGSPTASTTPQNSLVSAAATSFHHPGRGLVSR					
Ace3-S			-RPKCEYLSAVDSCUTHCRDAHVQCTFDLPLARRGPKARRKSDQPGQPPPDESS				
Ace3-SC			-RPKCEYLSAVDSCUTHCRDAHVQCTFDLPLARRGPKARRKSDQPGQPPPDESS				
Ace3-L		ACDRCCRKKAKCEYLSAVDSCUTHCRDAHVQCTFDLPLARRGPKARRKSDQPGQPPPDESS					
Ace3-LN		ACDRCCRKKAKCEYLSAVDSCUTHCRDAHVQCTFDLPLARRGPKARRKSDQPGQPPPDESS					
Ace3-LC		ACDRCCRKKAKCEYLSAVDSCUTHCRDAHVQCTFDLPLARRGPKARRKSDQPGQPPPDESS					
Ace3-EL		ACDRCCRKKAKCEYLSAVDSCUTHCRDAHVQCTFDLPLARRGPKARRKSDQPGQPPPDESS					
Ace3-S			LSTAARPQMQPPPPLTFSGPAVAALQPFASSSSLSPDAAWEPVEPLSIDNGLPRQPLGLP				
Ace3-SC			LSTAARPQMQPPPPLTFSGPAVAALQPFASSSSLSPDAAWEPVEPLSIDNGLPRQPLGLP				
Ace3-L		LSTAARPQMQPPPPLTFSGPAVAALQPFASSSSLSPDAAWEPVEPLSIDNGLPRQPLGLP					
Ace3-LN		LSTAARPQMQPPPPLTFSGPAVAALQPFASSSSLSPDAAWEPVEPLSIDNGLPRQPLGLP					
Ace3-LC		LSTAARPQMQPPPPLTFSGPAVAALQPFASSSSLSPDAAWEPVEPLSIDNGLPRQPLGLP					
Ace3-EL		LSTAARPQMQPPPPLTFSGPAVAALQPFASSSSLSPDAAWEPVEPLSIDNGLPRQPLGLP					

Ace3-S	LSTIQNISTRQRWIHLANAMTLRNTTLERVKRCIDLFFFDDYPLTPLVYEPALRVDLAY
Ace3-SC	LSTIQNISTRQRWIHLANAMTLRNTTLERVKRCIDLFFFDDYPLTPLVYEPALRVDLAY
Ace3-L	LSTIQNISTRQRWIHLANAMTLRNTTLERVKRCIDLFFFDDYPLTPLVYEPALRVDLAY
Ace3-LN	LSTIQNISTRQRWIHLANAMTLRNTTLERVKRCIDLFFFDDYPLTPLVYEPALRVDLAY
Ace3-LC	LSTIQNISTRQRWIHLANAMTLRNTTLERVKRCIDLFFFDDYPLTPLVYEPALRVDLAY
Ace3-EL	LSTIQNISTRQRWIHLANAMTLRNTTLERVKRCIDLFFFDDYPLTPLVYEPALRVDLAY
Ace3-S	IFSQQLPGVNQPSPLSQLTPDPTGTTPLNAAESWAGFGQPSSRTVGSLAPWADSTFT
Ace3-SC	IFSQQLPGVNQPSPLSQLTPDPTGTTPLNAAESWAGFGQPSSRTVGSLAPWADSTFT
Ace3-L	IFSQQLPGVNQPSPLSQLTPDPTGTTPLNAAESWAGFGQPSSRTVGSLAPWADSTFT
Ace3-LN	IFSQQLPGVNQPSPLSQLTPDPTGTTPLNAAESWAGFGQPSSRTVGSLAPWADSTFT
Ace3-LC	IFSQQLPGVNQPSPLSQLTPDPTGTTPLNAAESWAGFGQPSSRTVGSLAPWADSTFT
Ace3-EL	IFSQQLPGVNQPSPLSQLTPDPTGTTPLNAAESWAGFGQPSSRTVGSLAPWADSTFT
Ace3-S	LVTAVCAEEAAFMPLPKDIFPEGESVSEILLEASRDCLHQHLEADLENPTANSIAIRYFHSN
Ace3-SC	LVTAVCAEEAAFMPLPKDIFPEGESVSEILLEASRDCLHQHLEADLENPTANSIAIRYFHSN
Ace3-L	LVTAVCAEEAAFMPLPKDIFPEGESVSEILLEASRDCLHQHLEADLENPTANSIAIRYFHSN
Ace3-LN	LVTAVCAEEAAFMPLPKDIFPEGESVSEILLEASRDCLHQHLEADLENPTANSIAIRYFHSN
Ace3-LC	LVTAVCAEEAAFMPLPKDIFPEGESVSEILLEASRDCLHQHLEADLENPTANSIAIRYFHSN
Ace3-EL	LVTAVCAEEAAFMPLPKDIFPEGESVSEILLEASRDCLHQHLEADLENPTANSIAIRYFHSN
Ace3-S	CLHAAGKPKYSWHIFGEAIRLAQVMQLHEEAALEGLVPIEAEFRRRCFWIILYLGDKSAAI
Ace3-SC	CLHAAGKPKYSWHIFGEAIRLAQVMQLHEEAALEGLVPIEAEFRRRCFWIILYLGDKSAAI
Ace3-L	CLHAAGKPKYSWHIFGEAIRLAQVMQLHEEAALEGLVPIEAEFRRRCFWIILYLGDKSAAI
Ace3-LN	CLHAAGKPKYSWHIFGEAIRLAQVMQLHEEAALEGLVPIEAEFRRRCFWIILYLGDKSAAI
Ace3-LC	CLHAAGKPKYSWHIFGEAIRLAQVMQLHEEAALEGLVPIEAEFRRRCFWIILYLGDKSAAI
Ace3-EL	CLHAAGKPKYSWHIFGEAIRLAQVMQLHEEAALEGLVPIEAEFRRRCFWIILYLGDKSAAI

Ace3-S LNNRPITIHKYCFDAGITTLYPSGIEDEFLSTASEPPRKSFISGFNANVRLWQSAADLL
 Ace3-SC LNNRPITIHKYCFDAGITTLYPSGIEDEFLSTASEPPRKSFISGFNANVRLWQSAADLL
 Ace3-L LNNRPITIHKYCFDAGITTLYPSGIEDEFLSTASEPPRKSFISGFNANVRLWQSAADLL
 Ace3-LN LNNRPITIHKYCFDAGITTLYPSGIEDEFLSTASEPPRKSFISGFNANVRLWQSAADLL
 Ace3-LC LNNRPITIHKYCFDAGITTLYPSGIEDEFLSTASEPPRKSFISGFNANVRLWQSAADLL
 Ace3-EL LNNRPITIHKYCFDAGITTLYPSGIEDEFLSTASEPPRKSFISGFNANVRLWQSAADLL

Ace3-S EIRVLQDQMMQHFRGTMPNHLVPSADRHQLDSLYVRFITCLDDLPPLQSQCTLAMAMA
 Ace3-SC EIRVLQDQMMQHFRGTMPNHLVPSADRHQLDSLYVRFITCLDDLPPLQSQCTLAMAMA
 Ace3-L EIRVLQDQMMQHFRGTMPNHLVPSADRHQLDSLYVRFITCLDDLPPLQSQCTLAMAMA
 Ace3-LN EIRVLQDQMMQHFRGTMPNHLVPSADRHQLDSLYVRFITCLDDLPPLQSQCTLAMAMA
 Ace3-LC EIRVLQDQMMQHFRGTMPNHLVPSADRHQLDSLYVRFITCLDDLPPLQSQCTLAMAMA
 Ace3-EL EIRVLQDQMMQHFRGTMPNHLVPSADRHQLDSLYVRFITCLDDLPPLQSQCTLAMAMA

Ace3-S EGNGSAESKQQVVIQCINLQVTFHCLRMVITQKFEDILSYFAPGVQEADLRKSEIVRDMLRV
 Ace3-SC EGNGSAESKQQVVIQCINLQVTFHCLRMVITQKFEDILSYFAPGVQEADLRKSEIVRDMLRV
 Ace3-L EGNGSAESKQQVVIQCINLQVTFHCLRMVITQKFEDILSYFAPGVQEADLRKSEIVRDMLRV
 Ace3-LN EGNGSAESKQQVVIQCINLQVTFHCLRMVITQKFEDILSYFAPGVQEADLRKSEIVRDMLRV
 Ace3-LC EGNGSAESKQQVVIQCINLQVTFHCLRMVITQKFEDILSYFAPGVQEADLRKSEIVRDMLRV
 Ace3-EL EGNGSAESKQQVVIQCINLQVTFHCLRMVITQKFEDILSYFAPGVQEADLRKSEIVRDMLRV

Ace3-S MNEAPFWGLQANGEPNVEKIRLIGASLLAIHRNQDSPLATRARSDFSVLILDILTRLDSK
 Ace3-SC MNEAPFWGLQANGEPNVEKIRLIGASLLAIHRNQDSPLATRARSDFSVLILDILTRLDSK
 Ace3-L MNEAPFWGLQANGEPNVEKIRLIGASLLAIHRNQDSPLATRARSDFSVLILDILTRLDSK
 Ace3-LN MNEAPFWGLQANGEPNVEKIRLIGASLLAIHRNQDSPLATRARSDFSVLILDILTRLDSK
 Ace3-LC MNEAPFWGLQANGEPNVEKIRLIGASLLAIHRNQDSPLATRARSDFSVLILDILTRLDSK
 Ace3-EL MNEAPFWGLQANGEPNVEKIRLIGASLLAIHRNQDSPLATRARSDFSVLILDILTRLDSK

Ace3-S ASDQLRNTSTVVVG (SID NO:3)
 Ace3-SC ASD----- (SID NO:8)
 Ace3-L ASD----- (SID NO:6)
 Ace3-LN ASD----- (SID NO:14)
 Ace3-LC ASDQLRNTSTVVVG (SID NO:10)
 Ace3-EL ASD----- (SID NO:12)

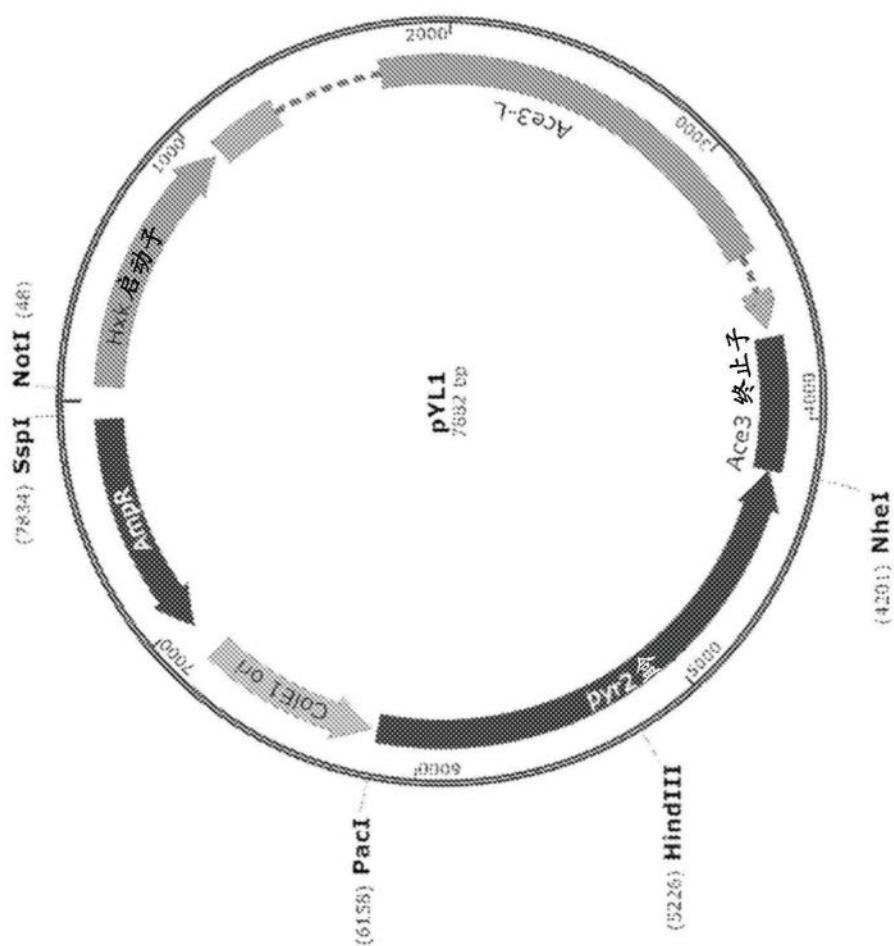


图2A

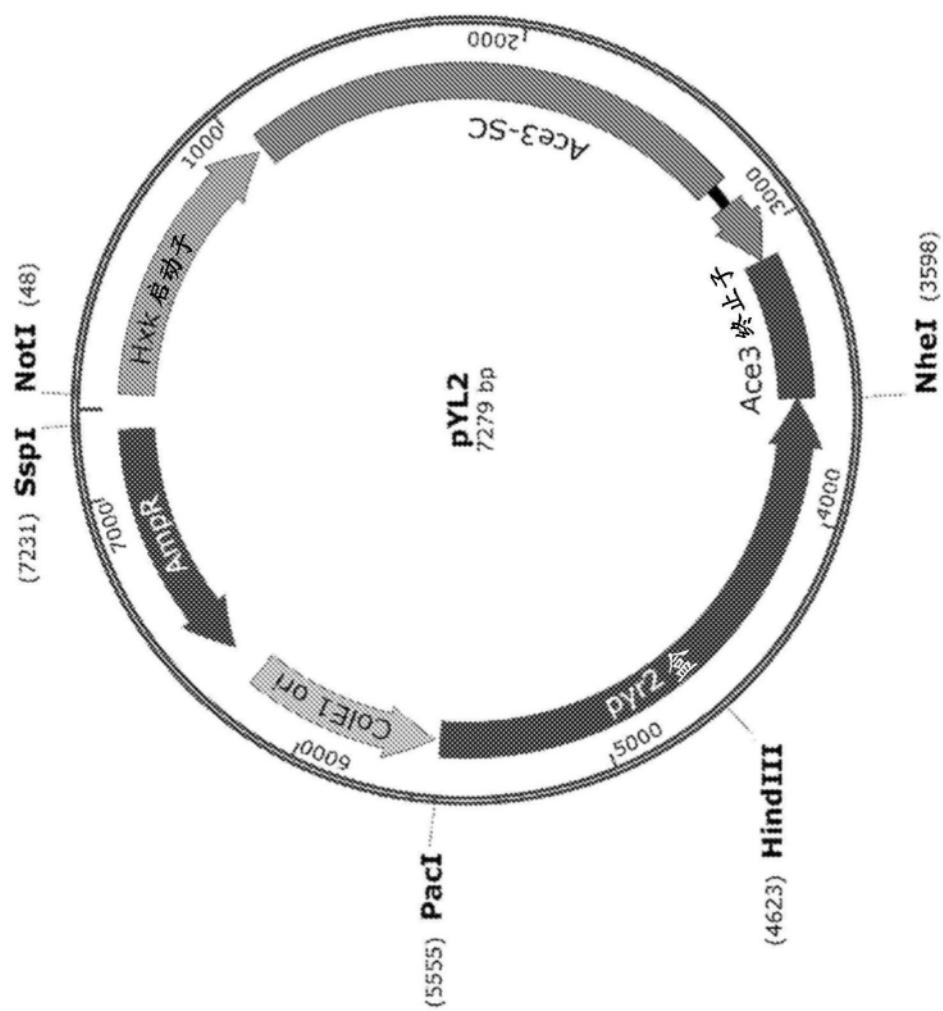


图2B

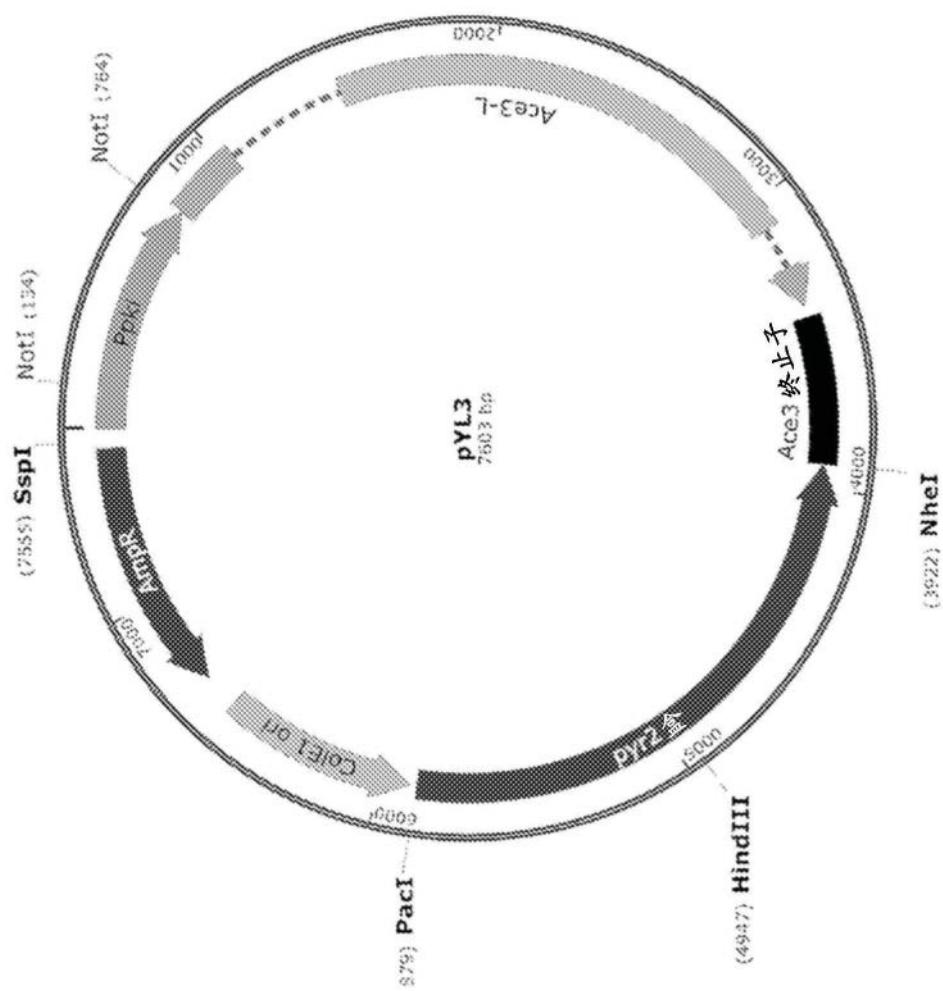


图2C

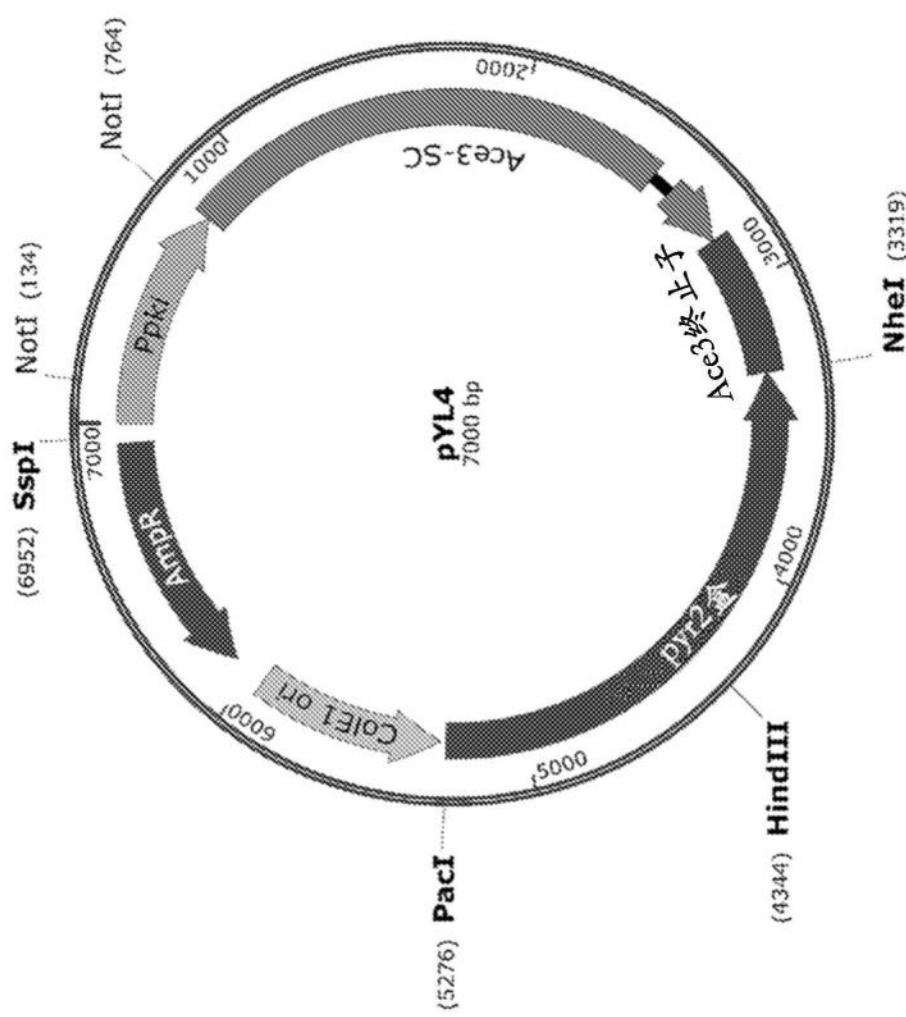


图2D

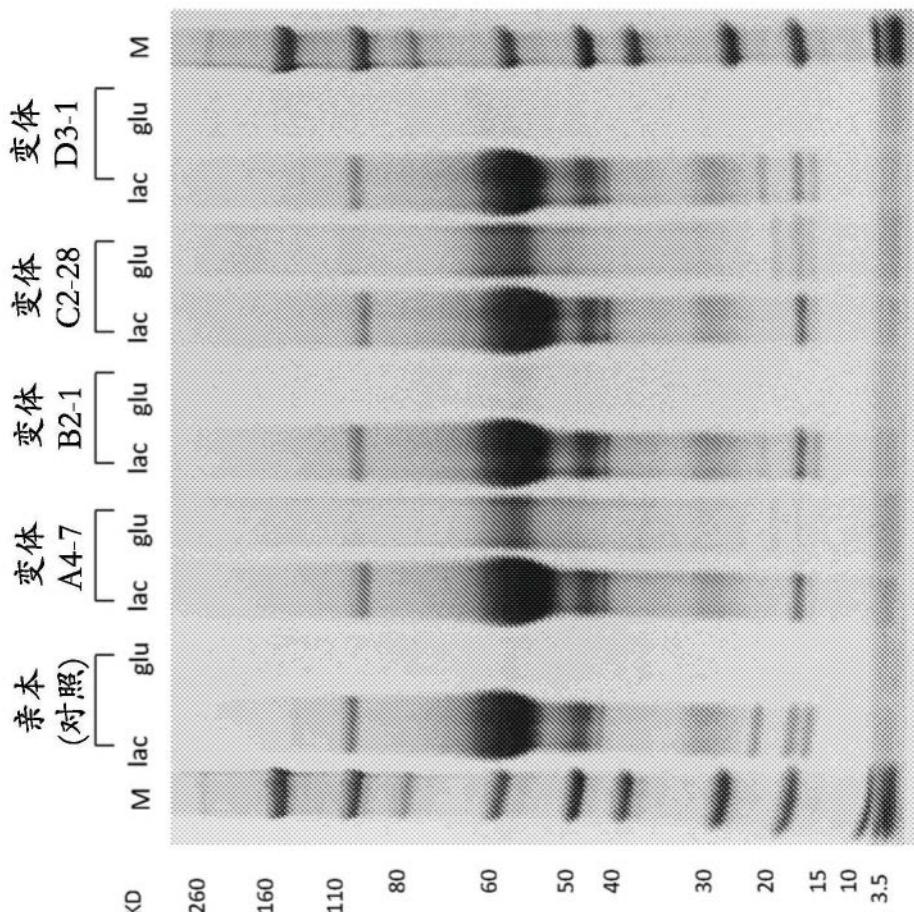


图3

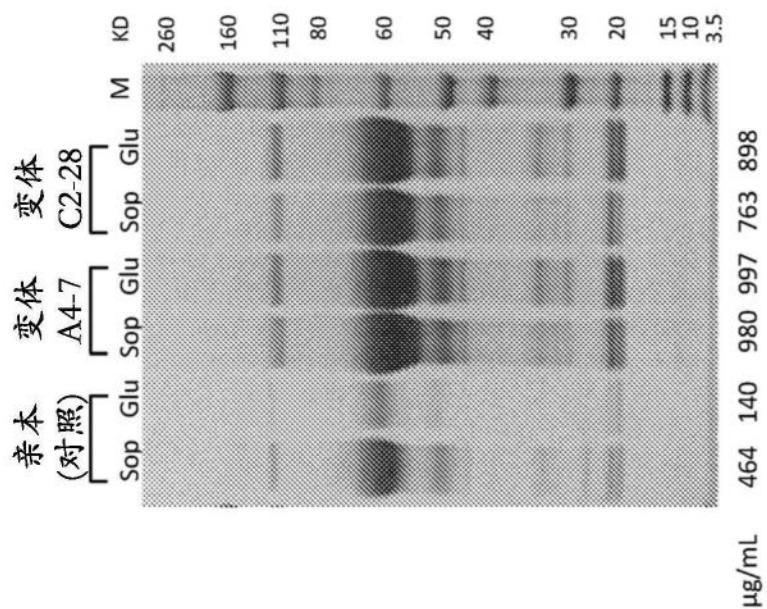


图4

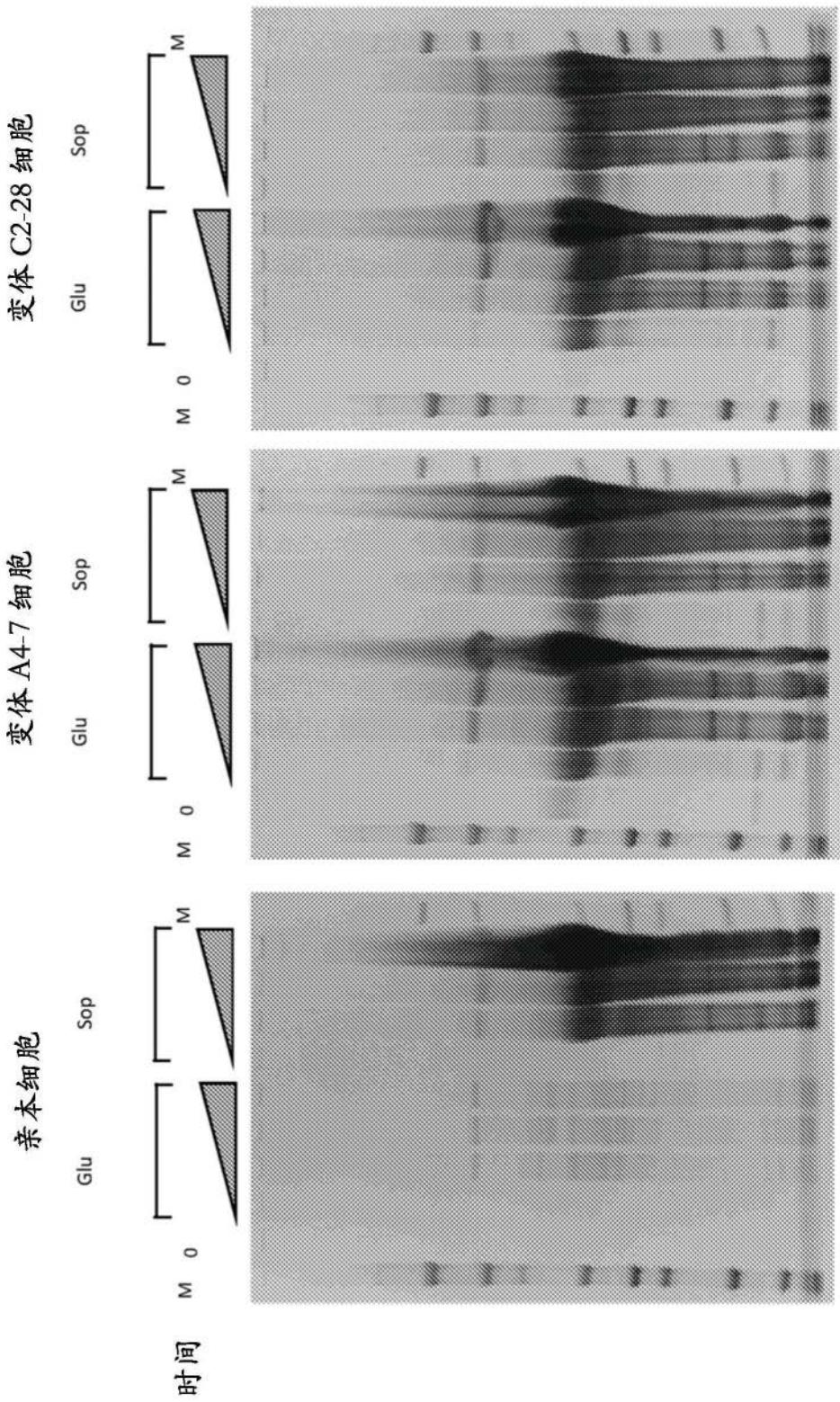


图5

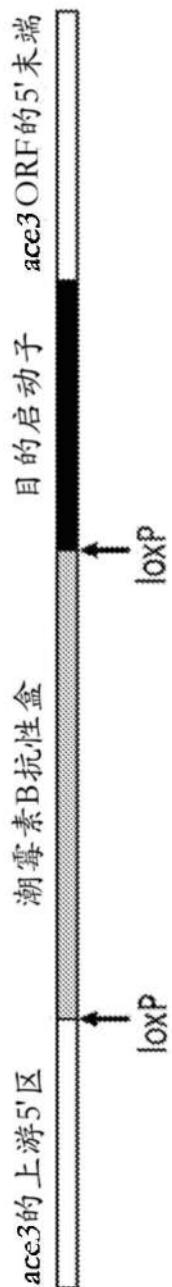
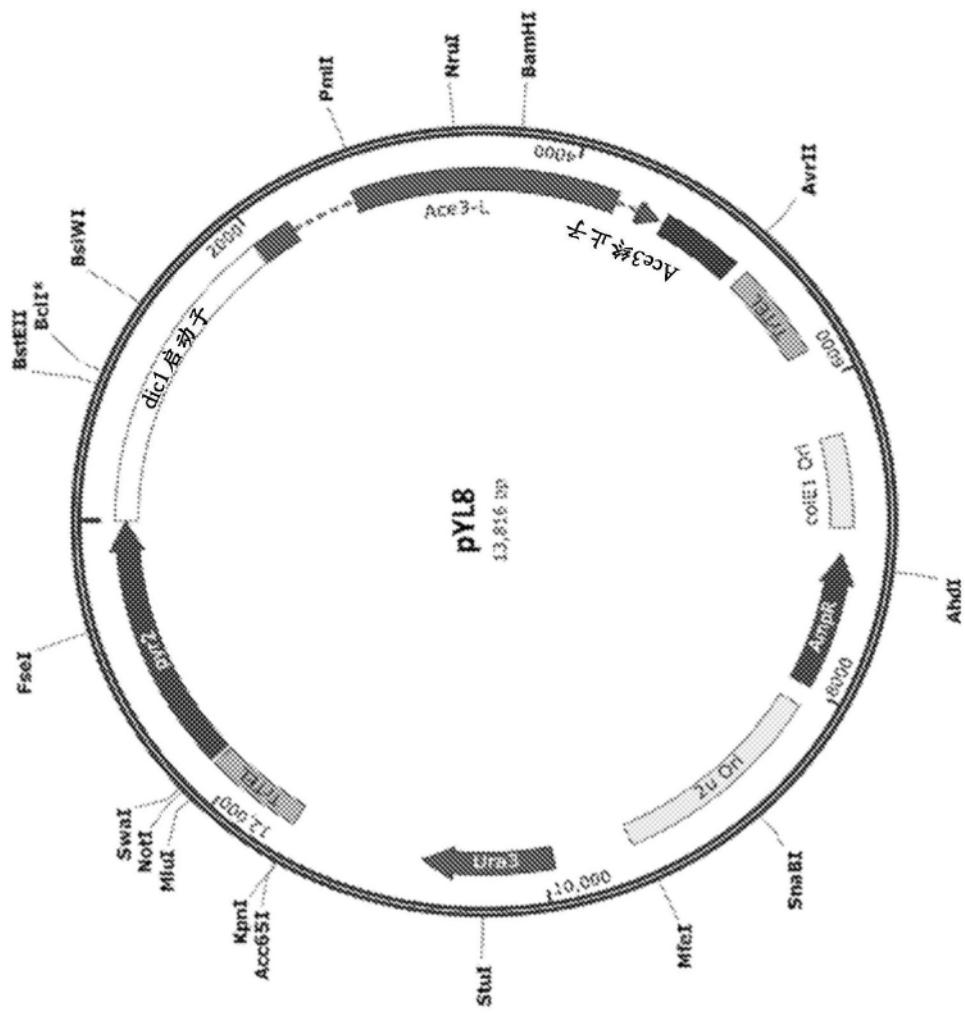


图6



冬7

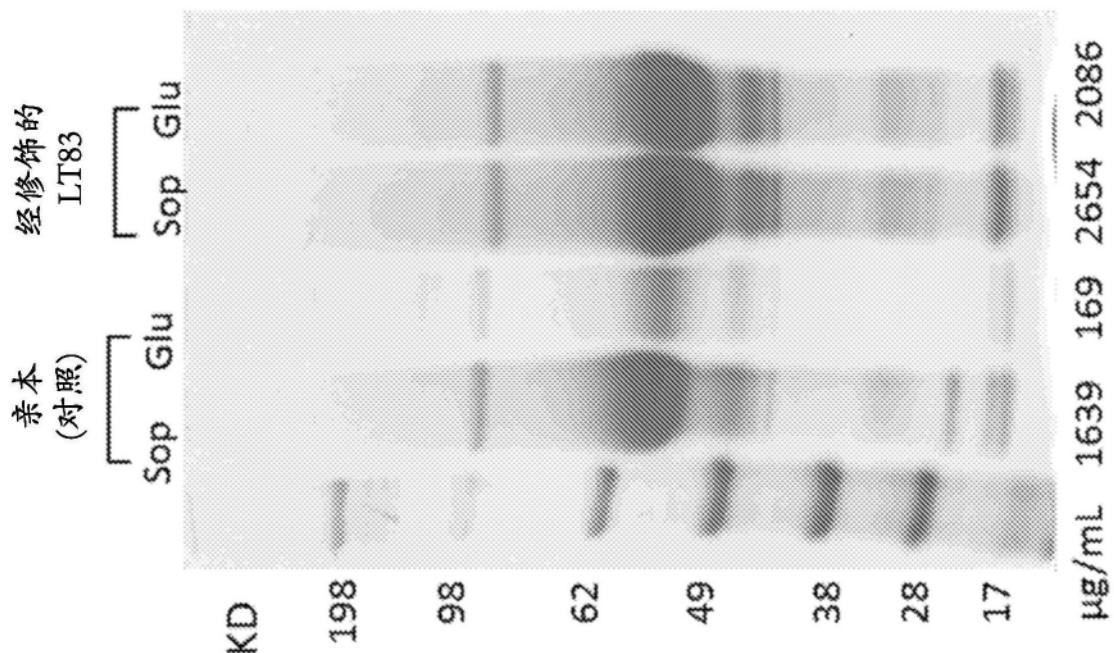


图8

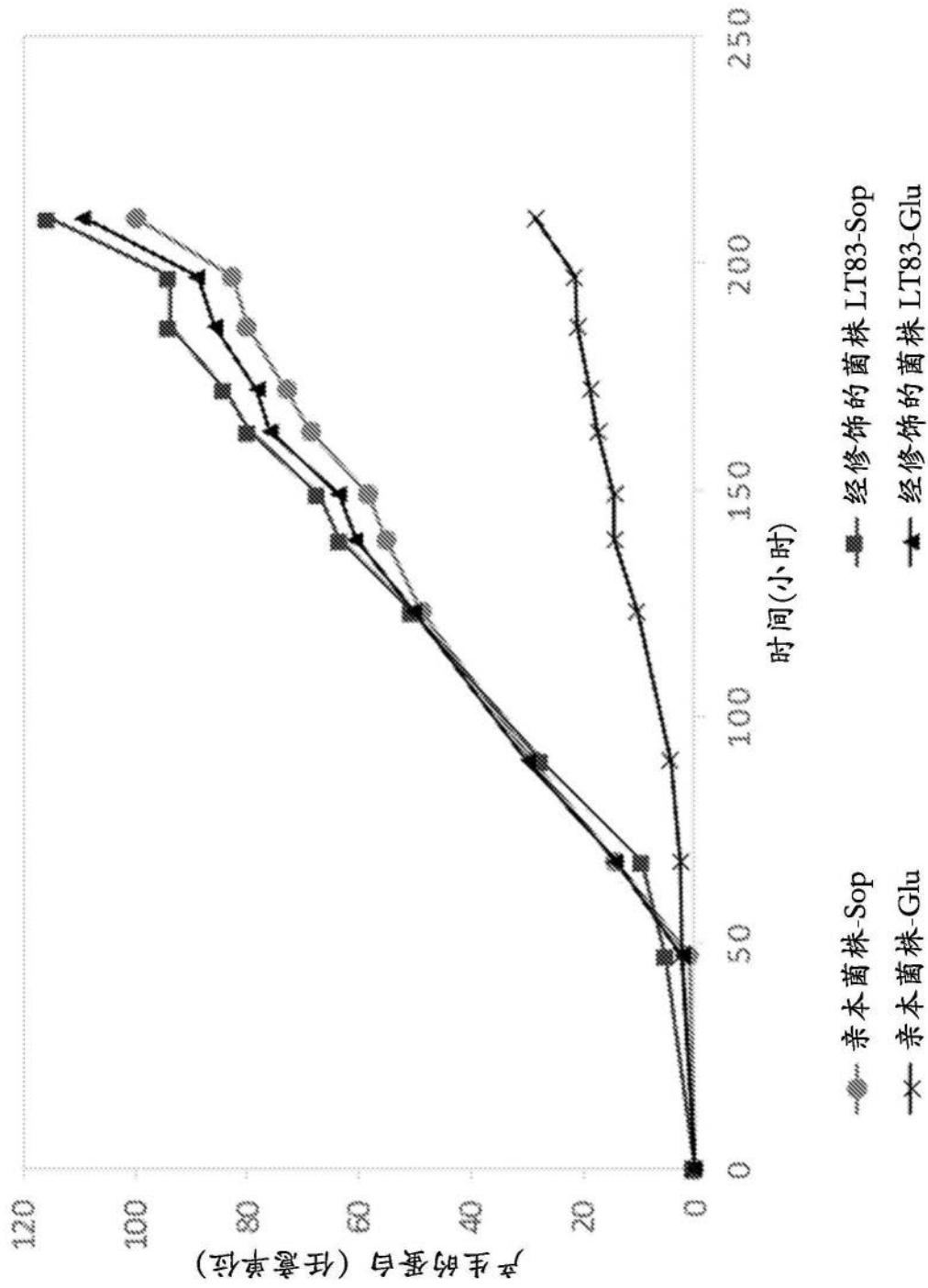


图9

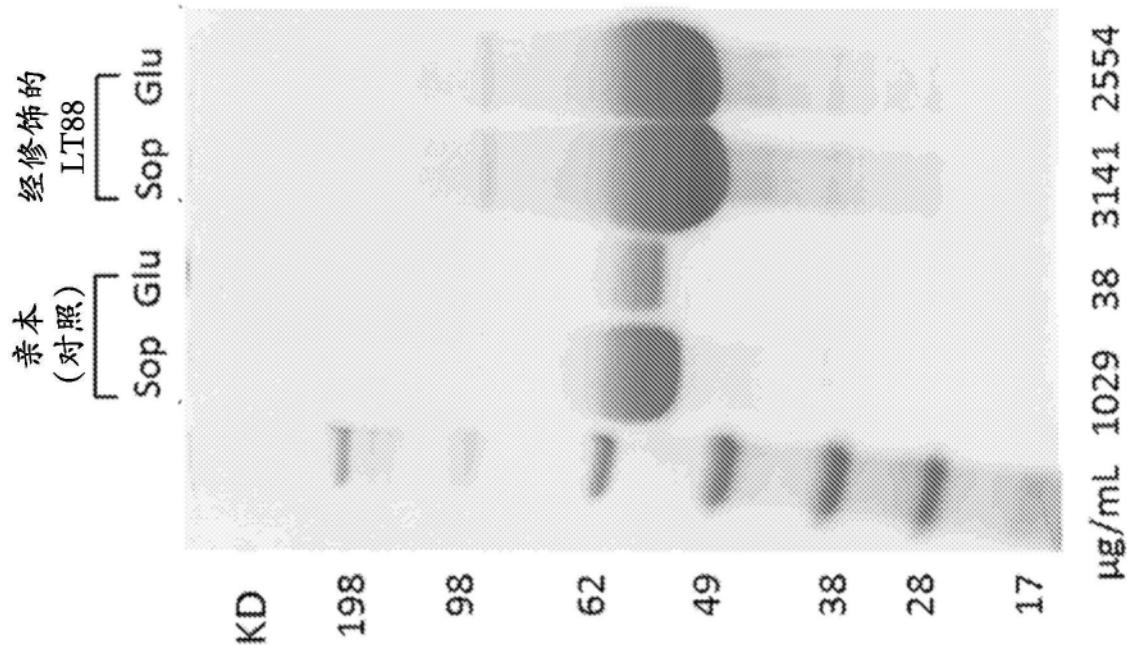


图10

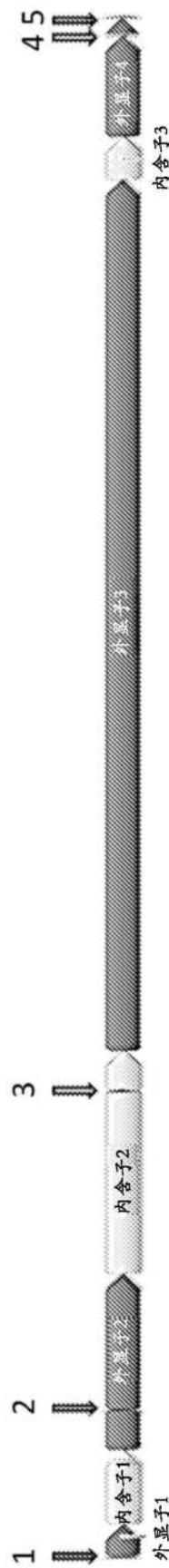


图11

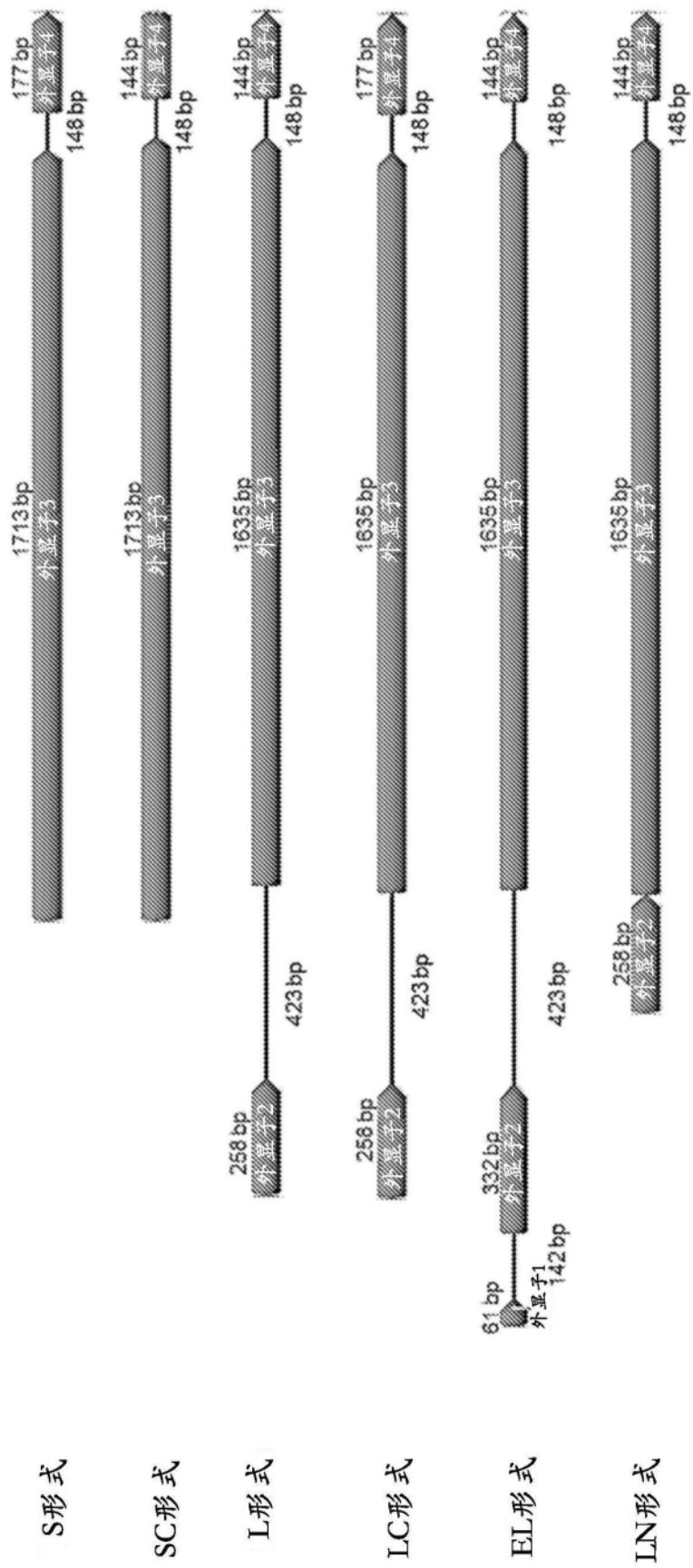


图12

Ace3 SC形式 (SEQ ID NO: 7)

Acetaminophen (SEQ ID NO: 8)

MLRYSFVILHLDLTLSSLPLTNALPRPKCEYLSAVDSCTHDRAHVOCTFDLPLARGPKARKKSQDQGPQPPDFSSSLSTAARPQGMPPLTFSGPAVAALQPFASSSSLSPDAAEWPVEPLSIDNGLPROFLGLPLGSTIONISTROWIHLANAMTLLRNNTTLEVSKRCIDLFFDYLYPLPVEPALRDVLAYIIFSQPLPGVNQFSPLSQLTPGTGTTPLNAAESWAGFGQPSGSRTVGSRLLAPWDSTFTLVTAUCAEEAFMLPKD1FPEGEVESSEILLEASRDCLHQHLEADLENPTANSIAIRYFHNSNCIHAAGKPKYSWHIFGEAIRLAQVMOLHEEAALEGIVPIAEEFRRCFWILYLGDKSAAILNNRPITIHKYCFDAGLTLYPSGIEDEFSLTASEEPRKSFISGENANVRIMOSAADLJLIEIRVLODOMMOHERGTMPPNHVLP SADROHLDLSLYVRFITCLDDLPYLQSCSTLAMAAMAEGNSAESKQVYTOCINLQVTFHCLRMVITQKFEDLSYFAPGVQEADLRKSETVRDMLRVNMNEAPFWGLQOANGEPMVEKIRLIGASILIAITHRNODSPLATRARSDESVILLDILTRLDKSASD



Acc3 S形式 (SEQ ID NO: 1)

Ace3蛋白S形式 (SEQ ID NO: 3)

MRLYSPVLLHDTLSSLPLTVALPRPKCEYLSAVIDSCTHCRAHVOQTFDLPLARRGPKARKKSQDQGQQPPFDSSSLSTAARPGQMPPFLTFSGPAAVAALQPFASSSLSPDAAEWPVEPL
SIDNGLPQLGLGJLSTQNISTRQWHLANAMTLRNTTLEVRSKRCIDLFFDLYPLTVYEPALRDVLAYIFSQPLPGVNQPSPLSQLTPDPTGTTPLNAAESWAGFQPSGS
SRTVGSRLAPWADSTFTLVTAUCAAAEMLPKDIFPEGESEVSIEILLEASRDCLHQHLEADLENPTANSIAIRYFHNSNCNLHAAGKPKYWHIFGEAIRLAQVMQLEHEEALEGLVPIAE
FRRRCFWIYLGDKSAIILNNRPITIHKYCFDAGITLYPSIEDFLSTASEPPRKSFIISGFNAVNRLWQSADLLEIRVLQDQMMHQFRGTMPPNHVLPSSADRHQLDSLYVRFITC
LDDLPYILOQSCTLAMAMAEGNGSAESKQYVIQCINLQVTFHCLRMVITQKFEDLSYFAPGVEQADLRKSEIVRDMLRVMNEAPFWGLQANGEPNVEKIRLIGASLLAIHRNQDSPLA
TRARSDESYVLLDLTRLDSSKAQDQLRNNTT'WVG



Ace3 L形式 (SEQ ID NO: 4)

Ac₃蛋白L形式 (SEQ ID NO: 6)

MGSAAAPAQGSVAAAAGGPPAAGAGAGAVHALTTSPESASAOFGSPTASTTPQNSLVSAAATSFHPRGRLLVSACDRCCRFRKACEYLSAVDSCTHCRDAHVGQCTFDLPLARRGPKARKSDQPGQPPPDPSSLSTAARPGMPPPLTFSGPAAVALQPFASSSLSPDAAEWPEVEPLSIDNGLPROPLGDLGLSTQNISTRQRWILANAMTLRNTTLERVSKRCIDLFDFDLYPLTPLVYEPALRDVDLAYIFSQPLPGVNQPSPLSQLTPDPTTGTPLNAEWSAGFGQPSGSRTVGSR LAPWADSTFTLVTA VCAEA A FMLKDIFPEGESVSEILLEASRDCLHQHLEADLENPTANSIAIRYFHSNCLHAAGKPYSWHI FGEATRIAQVMQLHEAALEGIVPIAEFRRCFWLIGDKSAAILNNRPITHKYCFDAGITLTPSGIEDEFLISTASEPPRKSFISGFENANVRWLQSOAADLLEIRVLQDMQHFRGTMMPNHVLP SADRHLDLSLYVRFITCDDLPPLQSOCTLAMAAMAEGNGSAESKQYVIQCINLQVTFHCLRMVITQKFEDLSYFAPGVQEADLRKSEIVRDLRVMMNEAFWGLOANGEPNVEKIRLIGASLLAIHRNODSPSPLATRARSDFSVLDDILTRLDKASD

Acetaminophen (SEQ ID NO: 9)

Acetaminophen (SEQ ID NO: 10)

NGSAAPAQGSVAAAAGGPPAAGAGAGAVHALTTSPESASASQPGSPTASTTPONSLVSAATSFHHPRGRILVSRACDRCRRIKAKCEYLSAVDSCTHCRAHVQCTFDLPLARRGGKAKSD
QPGQPPPDPSSLSSTAARPQGMPPPLTFSGPAAVALQPFASSSLSPDAAEWPVEPLSIDNGLPQLPGLSTIONI STRQWIHLANAMTLRNTTLERVSKRCDLFFDLYPLTPLYEPAP
1RDVLAYATISQPLPGVNQPSPLSOLTDPDTGTPLNAEWSWAGFGQPSGSRTVGSRLAPWADSTFTLVTAVCAEEAFMLPKDIFPEGEVSIEILLEASRDLHQLHEADLENPTANSIAIRYF
HENCLHAAGKPKYSWHIFGEAIRLAQVMQLEHEAALEGVP1EAEEFRRCFWLILGDGRSAA1LNRPITHKYCFDAGITLYPSGIEDEFLLSTASEPPRKSFISGFENANVRMOSAADLLE
1RVLQDOMMQHFRGTMPPNHVLPSSADRHQLDSLYVRFITCLDDLPYLOQSTLAMAMAEGNGSAEKQYVIOQINLQVTFHCLRMVITQKFEDLSYFAPGVEQADLRKSEIVRDMLRVNNEAP
FWGLOANGEENVEKIRLIGASLLAI1HRNODSPLATRARSDFSVLLDILTRLDKSASDOLENTSTTVG



Ace3 EL形式 (SEQ ID NO: 11)

Ace3蛋白EL形式 (SEQ ID NO: 12)

MATAAAAAGGAAVAAGADTGAGSSSTGGPPGLPGTRTGSYAMGSAAPAAQGSYAAAAGGPPAAGAGAGAVHALTTSPESASASQPGSPTASTTPQNSIVSAATSFHHPGRGLVSRACD
RCRRRKAKCEYLSAVDSCTHCRDAHVQCTFDLPLARRGPKARKKSDQPGQPDPSSLSTAARPQOMPPLTFSGEAVAALQPFASSSLSPDAAWEPVEPLSIDNGLPQLGDLPLGLSTION
ISTRQRWIHLANAMTLRNTTLERVKRKCIDLFFDLYPLTVYEALRDVLAYIFSQPLPGVNQPSPLSQLTPEDETTGTGTPLNAAESWAGFQPSGSRTVGSRSLAPWADSTFTLVTA
AEAFMLPKDIFPGEVESVSEILLEASRDCLHQHLEADLENPTNSIAIRYFHNSNLWAAGKPYSWHITEGEARLRAQVMQLEEEAALEGLVPIAEFRRRCFEWILYLGDKSAAILNNRPI
FDAGITLYPSGIEDEFLSTASEPPRSKSFISGENANVRWQSAADLLEIRVLUQDMQHFRGTMPPNHVLPSSADROHLDLSLYVRFTICLDDLPYLOQSCTLAMAEGNSA
EKGQYVIQCINLOVTFHCLRMOVITOKEFDLSSYFAPGVEAOADRKRSEIVRDMLRVNNEAPFWGLQANGEPVNEKIRLIGASLLAIIHRNODSPLATRARSDESVLLDILTRLD
SKASD

Ace3 LN形式 (SEQ ID NO: 13)

Ace3蛋白LN形式 (SEQ ID NO: 14)

MGSAAPAQGSVAAAAGGPPAAGAGAGAVHALTTSPESASASQPGSPTASTTPQNLSVSAATSFHIIHPRGRIVSPACDRCCRRAKCEYLSAATSDCCTHCRDAHVQCTFDLPLARRGPARKKSDSL
PGQQPPDPDESSLSTAARPQMPPLTFSGPVAALQFQASSSLSPDAAWEPVPLSIDNGLPROPLDLPGLSTIONISTRQRWILHANAMTLRLNTTLEVRVKRCIDLFFDYPLTPLVYEPAFLR
DVLAYIFSOPLPGVNQPSPLSQLTPDPTGTPLNAAESWAGFGQPSSGSRTVGSRLAPWDSTFTLVATCAEAAFMPLPKDIIFPEGEVSSEILLEASRDLCHOLEADLENPTANSIAIRYFHSN
CLHAAGKPKYSWHIFGEAIRLAQVMOLHEEAALLEGVLPIAEAFRRFCWILYLGDKSAAAILNRPRTIHKYCFDAGITLYPSGIEDEFFLSTASEPPRKSFISGFENANVRLWOSAADDLLEIRVL
QDOMMOHFRGTMPPNHVLPSSADRQHLDSSLVRFITCLDDLPYLOSCTLAMAAMAEGNGSAESKQVYIQCINLQVTFHCRLRMVITQKFEDLSYFAPGVEQADLRKSEIVRDMLRVNNEAPFWGLO
ANGENVEKIRLIGASLLAIHRNDOSPLATRARSDFSVLDDILTRDLSKASD

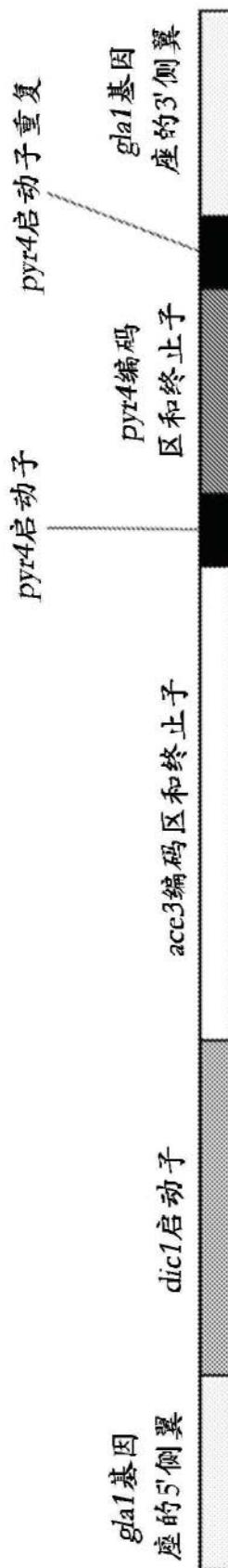


图19

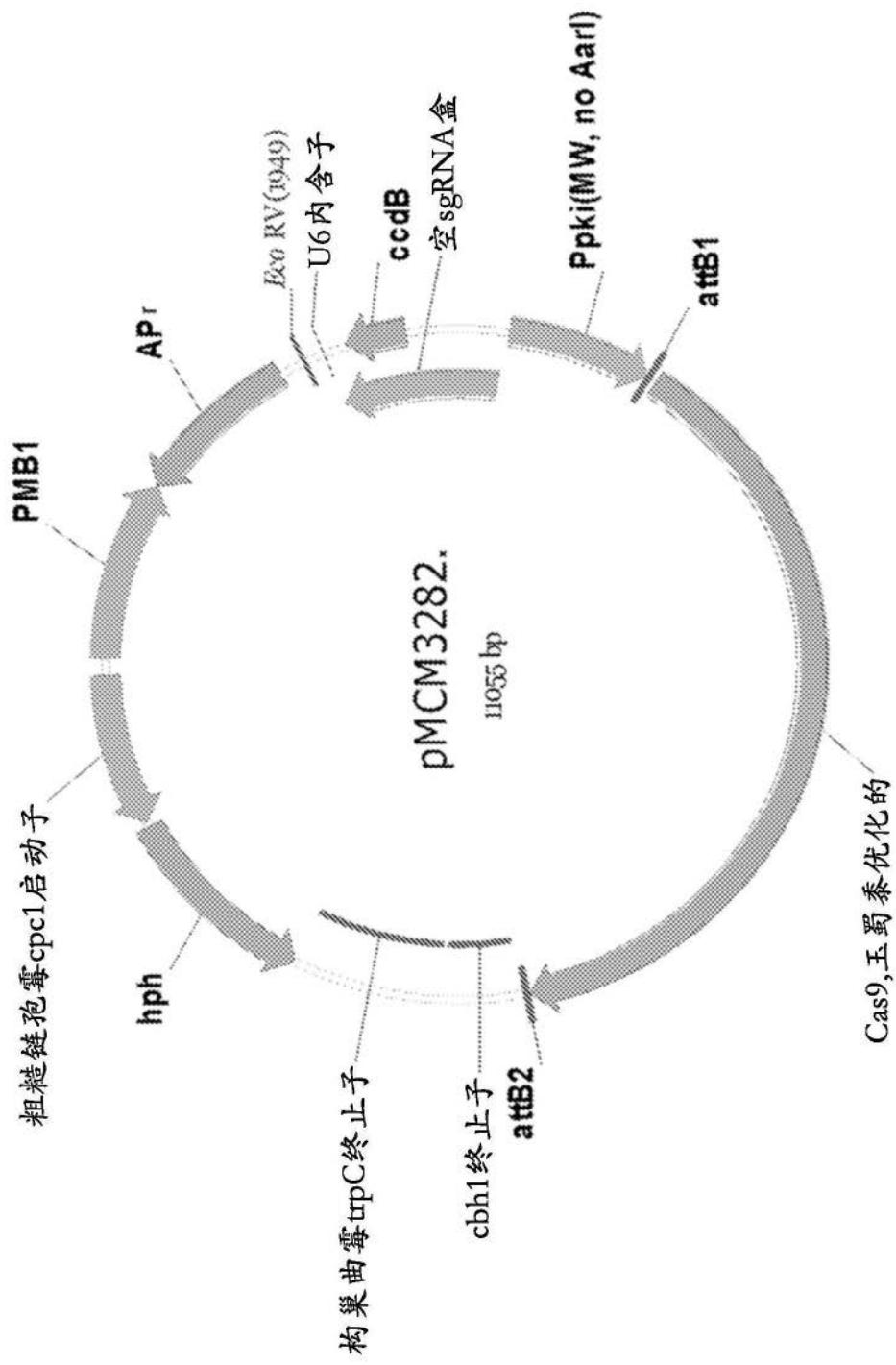


图20

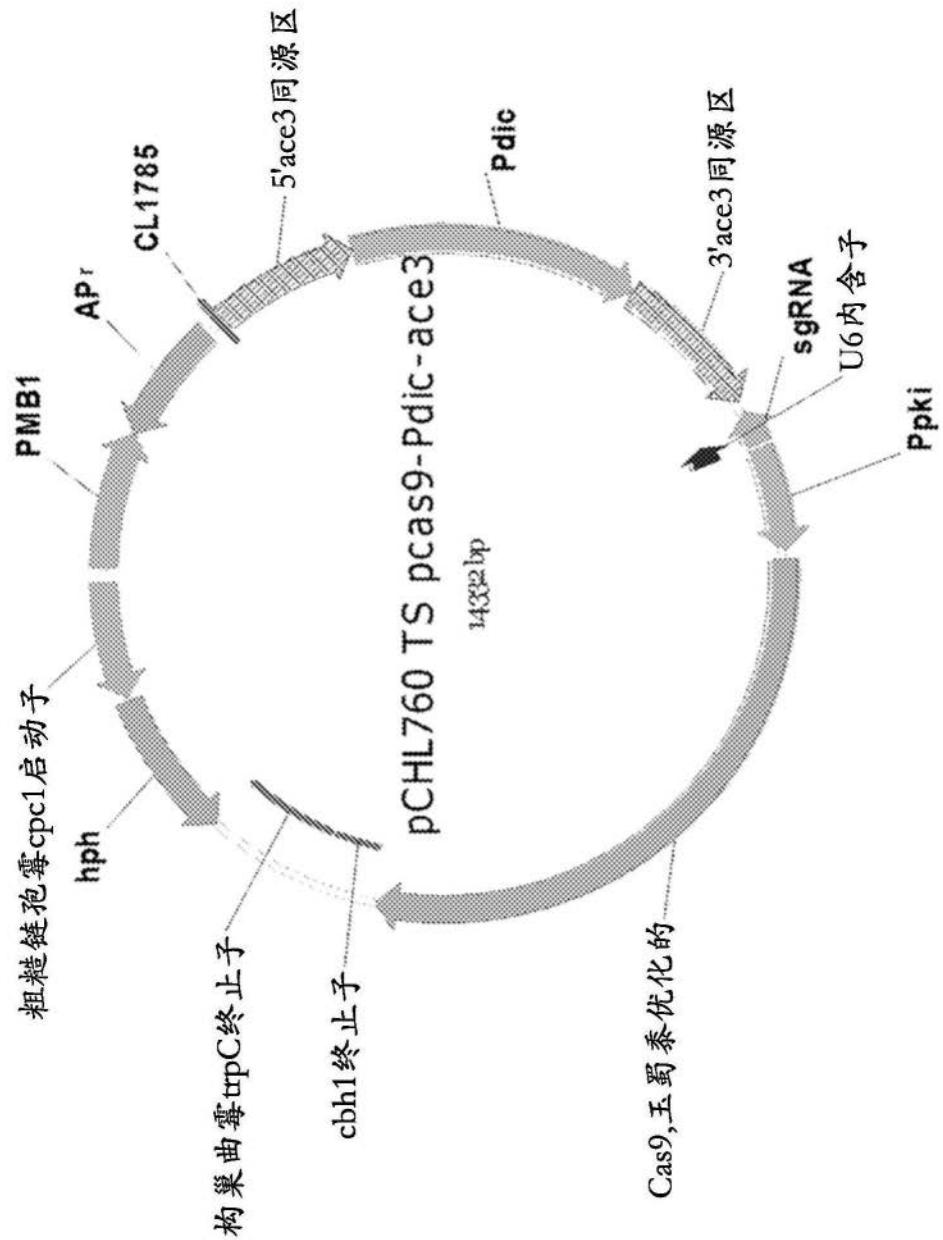


图21

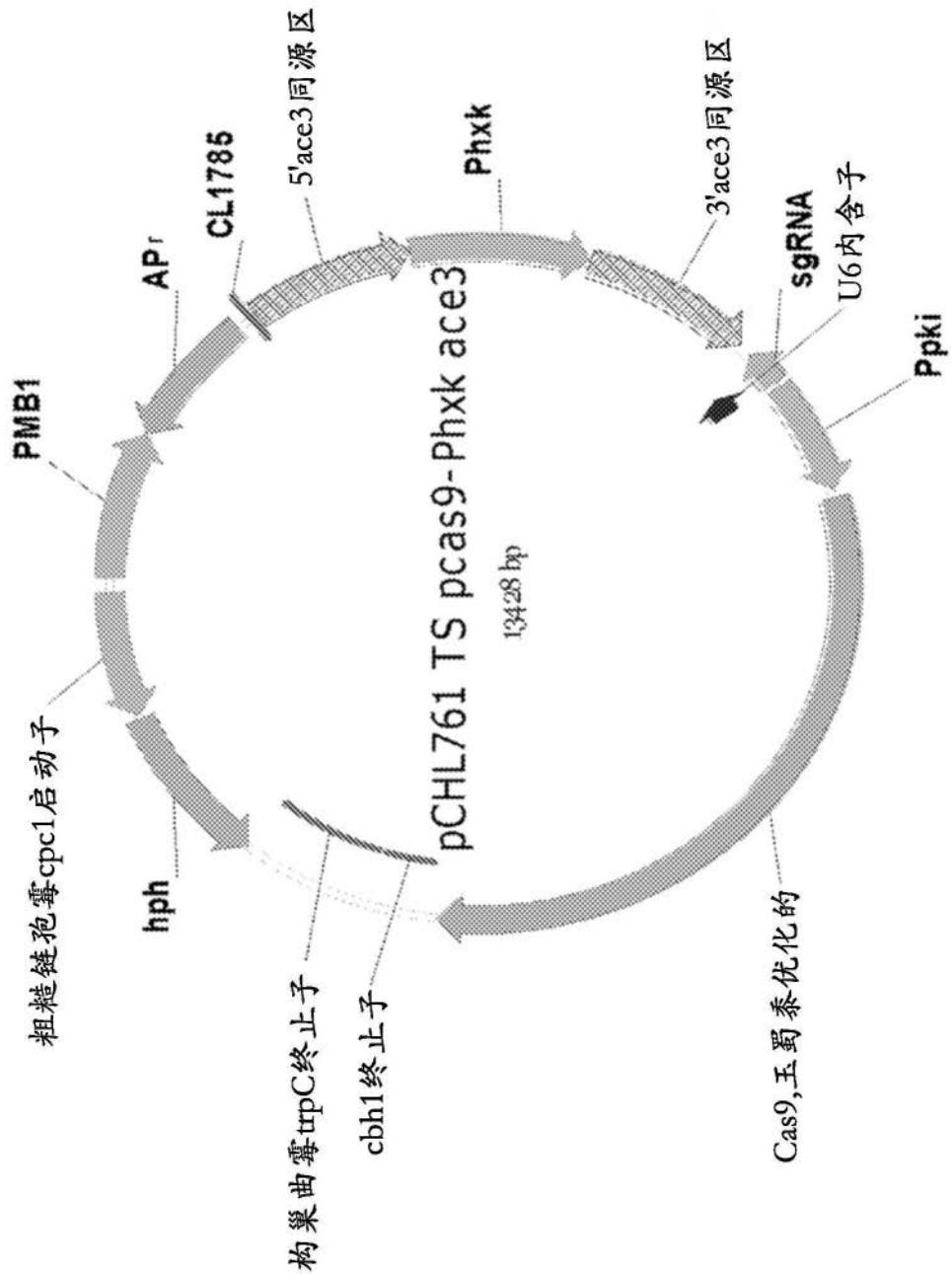


图22

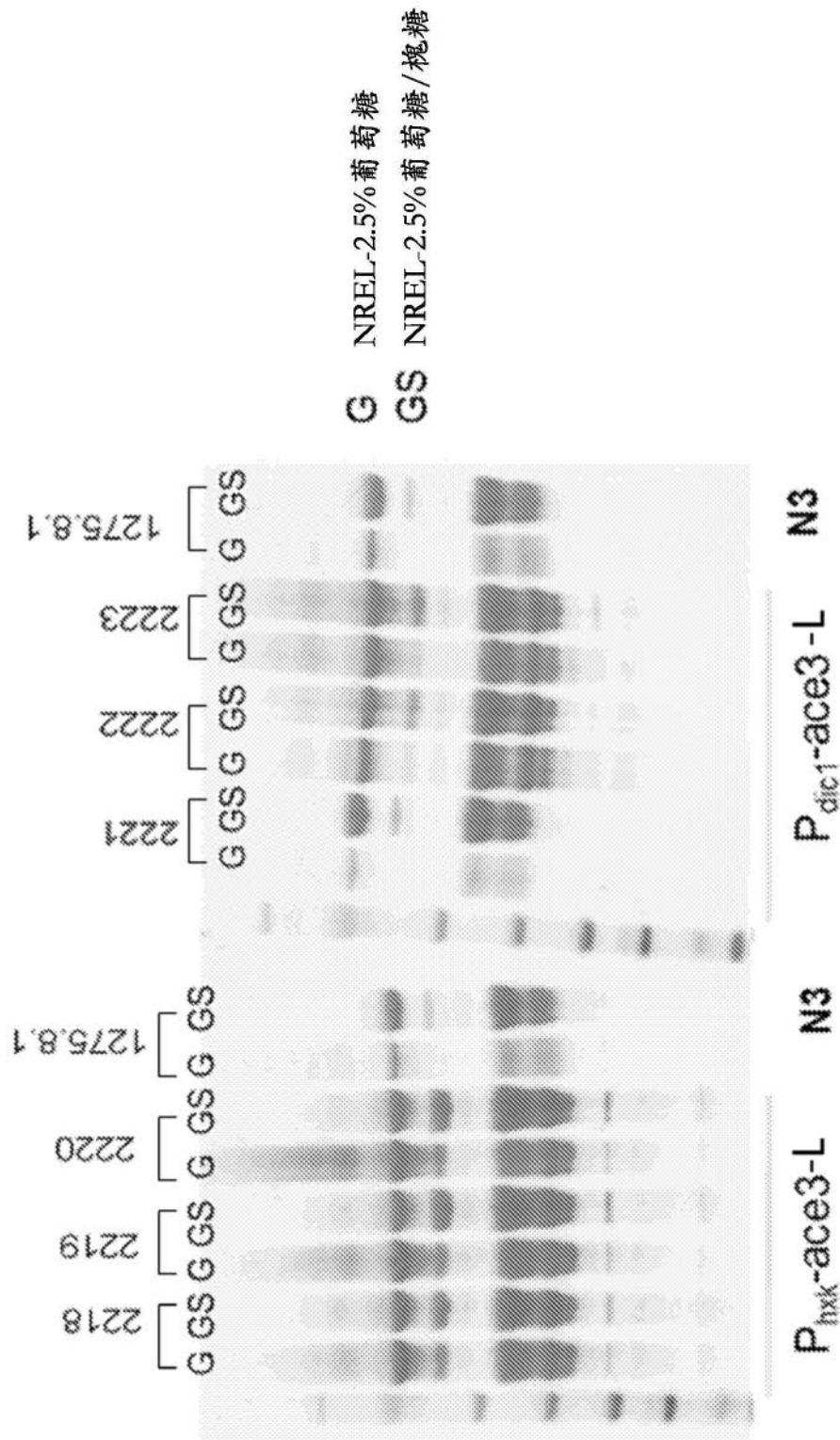


图23