



(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  
    专利代理师 胡志君 黄革生

(51) Int.Cl.  
    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数据库.2017,全文. (续)

审查员 留盛鹏

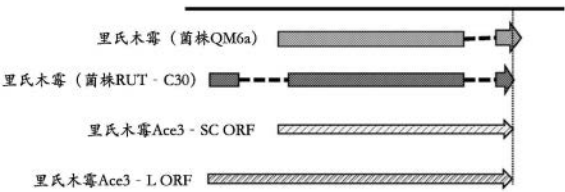
权利要求书2页 说明书51页  
序列表62页 附图28页

(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

Ace3蛋白的纤维素酶表达激活子3(ace3)基因变体的表达。

来自里氏木霉菌株QM6a和RUT-C30的Ace3蛋白编码区的示意图



CN 109790510 B

[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 $\beta$ -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 Xyl1 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(*Trichoderma* sp.) 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glal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,所述核苷酸序列是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dld1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hvk1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萄糖淀粉酶gla1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: $\alpha$ -淀粉酶、碱性 $\alpha$ -淀粉酶、 $\beta$ -淀粉酶、纤维素酶、葡聚糖酶、 $\alpha$ -葡萄糖苷酶、 $\alpha$ -半乳糖苷酶、葡萄糖淀粉酶、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 $\beta$ -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、 $\alpha$ -半乳糖苷酶、 $\beta$ -半乳糖苷酶、 $\beta$ -葡萄糖苷酶、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hxx1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。

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、 $\beta$ -葡萄糖苷酶),其水解纤维素的 $\beta$ -(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- $\beta$ -1,4-木聚糖酶(EC 3.3.1.8)、xyn1(内切-1,4- $\beta$ -木聚糖酶xyn1)和xyn2(内切-1,4- $\beta$ -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MIHC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 and 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、egl1、egl2、egl3、egl4、egl5、egl6、bgl1、bgl2、xyn1、xyn2、xyn3、bx11、abf1、abf2、abf3、axe1、axe2、axe3、man1、agl1、agl2、agl3、glr1、swol、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: $\alpha$ -淀粉酶、碱性 $\alpha$ -淀粉酶、 $\beta$ -淀粉酶、纤维素酶、葡聚糖酶、 $\alpha$ -葡萄糖苷酶、 $\alpha$ -半乳糖苷酶、葡萄糖淀粉酶、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 $\beta$ -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、 $\alpha$ -半乳糖苷酶、 $\beta$ -半乳糖苷酶、 $\beta$ -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 NO: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 NO: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萄糖淀粉酶(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 NO:6、SEQ ID NO:12或SEQ ID NO: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 NO:6、SEQ ID NO:12或SEQ ID NO: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 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,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 NO:6、SEQ ID NO:12或SEQ ID NO: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 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。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萄糖淀粉酶(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 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。

[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)、dld1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)、hxx1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、egl1、egl2、egl3、egl4、egl5、egl6、bgl1、bgl2、swol、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、bxl1、abf1、abf2、abf3、axe1、axe2、axe3、man1、agl1、agl2、agl3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、egl1、egl2、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、egl1、egl2、egl3、egl4、egl5、egl6、bgl1、bgl2、swol、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、bxl1、abf1、abf2、abf3、axe1、axe2、axe3、man1、agl1、agl2、agl3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)、bx1启动子(SEQ ID NO:16)、tk11启动子(SEQ ID NO:17)、PID104295启动子(SEQ ID NO:18)、dld1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)、hxx1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萄糖淀粉酶(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: $\alpha$ -淀粉酶、碱性 $\alpha$ -淀粉酶、 $\beta$ -淀粉酶、纤维素酶、葡聚糖酶、 $\alpha$ -葡萄糖苷酶、 $\alpha$ -半乳糖苷酶、葡萄糖淀粉酶、半纤维素酶、戊聚糖酶、木聚糖酶、转化酶、乳糖酶、柚皮苷酶(naringinase)、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 $\beta$ -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、 $\alpha$ -半乳糖苷酶、 $\beta$ -半乳糖苷酶、 $\beta$ -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。

[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)、dld1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)、hxx1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萄糖淀粉酶(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: $\alpha$ -淀粉酶、碱性 $\alpha$ -淀粉酶、 $\beta$ -淀粉酶、纤维素酶、葡聚糖酶、 $\alpha$ -葡萄糖苷酶、 $\alpha$ -半乳糖苷酶、葡萄糖淀粉酶、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 $\beta$ -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、 $\alpha$ -半乳糖苷酶、 $\beta$ -半乳糖苷酶、 $\beta$ -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。

[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)、hxx1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苷酸序列和(b)有效地连接到该第一核苷酸序列的第二核苷酸序列,其中该第二核苷酸序列编码包含与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,或者该第二核苷酸序列编码包含与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)、hxx1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ace3ORF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dic1启动子的Ace3-表达载体pYL8的示意图。

[0043] 图8显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显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 NO: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 NO: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 NO: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 NO: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 NO: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 NO: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 NO: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 NO: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 NO: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 NO: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 NO: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 NO: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 NO: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 NO: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 NO: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 NO: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 NO: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gla1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 $\beta$ -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dld1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hvk1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、egl1、egl2、egl3、bgl1、bgl2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 NO:6)、Ace3-EL蛋白(SEQ ID NO:12)和Ace3-LN蛋白(SEQ ID NO: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 NO:6)、Ace3-EL蛋白(SEQ ID NO:12)和Ace3-LN蛋白(SEQ ID NO: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 NO:6)、Ace3-EL蛋白(SEQ ID NO:12)和Ace3-LN蛋白(SEQ ID NO: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,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 $\beta$ -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(现在是嗜热双歧菌属(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助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 $\beta$ -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“ $\beta$ -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以不依赖序列和无疤痕的方式连接。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起始密码子开始的阅读框确定。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,控制转录的任意的操纵子序列,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

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所希望活性的酶。

[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ace3ORF(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ace3ORF,比较和评价Hakkinen等人,2014中描述的克隆的ace3ORF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ace3ORF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ace3ORF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) $\beta$ -木糖苷酶基因(bxl;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) $\alpha$ -葡萄糖醛酸糖苷酶基因(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(pk11;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萄糖淀粉酶基因座(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1,029 $\mu$ g/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38 $\mu$ g/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布丘氏菌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pki1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hvk1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 NO: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 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。在其他实施例中,包含与SEQ ID NO: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 NO:6之前的SEQ ID NO: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 NO: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 NO:4、SEQ ID NO:11或SEQ ID NO: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 NO:5、SEQ ID NO:101或SEQ ID NO: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 NO: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 NO: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 NO: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als1、amdS、hygR、pyr2、pyr4、pyrG、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、氯喹磺隆乙酯抗性标记、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pyr2,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、hxx1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、dld1、xyn4、PID72526、axe1、hxx1、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 $2 \times 10^6$ /mL的密度进行。将100 $\mu$ L体积的在合适的溶液(例如1.2M山梨糖醇和50mM CaCl<sub>2</sub>)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萄糖淀粉酶基因(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何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益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、 $\beta$ -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、氨基酸酶、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、 $\alpha$ -半乳糖苷酶、 $\beta$ -半乳糖苷酶、 $\alpha$ -葡聚糖酶、葡聚糖裂解酶、内切- $\beta$ -葡聚糖酶、葡萄糖氧化酶、葡萄糖醛酸糖苷酶、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WO 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:

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

[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CO<sub>2</sub>)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,将定义的发酶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)保持在固定的速率,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适量的磷、镁、钙、钾、硫和钠,并且还优选存在也处于合适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℃至40℃的范围内,一般优选在约25℃至34℃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℃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ace3ORF(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hvk1(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ace3ORF。

[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ace3ORF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ace3ORF(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ace3ORF(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ace3ORF的表达,测试了异源己糖激酶(hxk1)启动子和异源丙酮酸激酶(pki1)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pki1启动子)、和具有其天然终止子的ace3ORF(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ace3ORF。

[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pki1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pki1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 $\alpha$ 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	TCAGGGTTATTGTCTCATGGCCATTAGGCCTGGCAGGCACTGGCTCGGACGACATGT	7
TP14	AGAGCCCTGGGCCGGAGCTGCTGAGCCCATTTGTTGAATTCTGGCGGGGTAGCTGTTGA	8
TP15	TCAACAGCTACCCCGCCAGAATTCAACAATGGGCTCAGCAGCTCCGGCCCAGGGCTCT	9
TP16	TCGTAAATAAACAAGCGTAAGTACTAGCTAGCGTAGGTTATGCGAGCAACATTGCACGAAAC	10
TP17	GTTTCGTGCAATGTTGCTCGCATAACCTACGCTAGCTAGTTACGCTTGTATTATTACGA	11
TP18	ACATGTGCTCCGAGCCAGTGCCTGCCAGGCCTAAATGGCCATGAGACAATAACCCCTGA	12
TP19	AGGTGTAAGACGGGGGAGTAGCGCAGCATTGTTGAATTCTGGCGGGGTAGCTGTTGA	13
TP20	TCAACAGCTACCCCGCCAGAATTCAACAATGCTGCGCTACTCCCCGTCTTACACCT	14
TP21	TCAGGGTTATTGTCTCATGGCCATTAGGCCTAGACTAGCGGCCGGTCCCCTTATCCCA	15
TP22	AGAGCCCTGGGCCGGAGCTGCTGAGCCCATGGTGAAGGGGGCGGCCGCGGAGCCT	16
TP23	AGGCTCCGCGGCCGCCCCCTTACCATGGGCTCAGCAGCTCCGGCCCAGGGCTCT	17
TP24	TGGGATAAGGGGACCGGCCGCTAGTCTAGGCCTAAATGGCCATGAGACAATAACCCCTGA	18
TP25	TGTAAGACGGGGGAGTAGCGCAGCATGGTGAAGGGGGCGGCCGCGGAGCCT	19
TP26	AGGCTCCGCGGCCGCCCCCTTACCATGCTGCGCTACTCCCCGTCTTACA	20

[0324] 1D. 里氏木霉的转化

[0325] 将pYL1、pYL2、pYL3和pYL4的表达载体使用PacI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  $\text{KH}_2\text{PO}_4$ 、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 $\cdot\text{H}_2\text{O}$ 、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  $\text{H}_3\text{BO}_3$ 。将所有srMTP在28℃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℃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℃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℃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产生能力相关,如下文实例3和实例4中所示。

[0337] 实例3

[0338] 摇瓶中的蛋白产生

[0339] 为了进一步探索和验证上述srMTP结果,在摇瓶中在50-mL浸没培养中在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℃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 N0:15)、(ii) β-木糖苷酶基因(bx1;蛋白ID 121127)启动子(SEQ ID N0:16)、(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 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)  $\alpha$ -葡萄糖醛酸糖苷酶基因 (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 (*hvk1*; 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](http://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间以低水平表达,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 $\mu$ 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	<i>ace3</i> ORF
pYL7	<i>opt</i> (SEQ ID NO: 25)	<i>ace3</i> -L
pYL8	<i>dic1</i> (SEQ ID NO: 24)	<i>ace3</i> -L
pYL9	<i>gut1</i> (SEQ ID NO: 26)	<i>ace3</i> -L
pYL12	<i>hvk1</i> (SEQ ID NO: 23)	<i>ace3</i> -L
pYL13	<i>pki1</i> (SEQ ID NO: 27)	<i>ace3</i> -L
pYL22	<i>rev3</i> (SEQ ID NO: 15)	<i>ace3</i> -L
[0352] pYL23	PID 104295 (SEQ ID NO: 18)	<i>ace3</i> -L
pYL24	<i>tkl1</i> (SEQ ID NO: 17)	<i>ace3</i> -L
pYL25	<i>bxl</i> (SEQ ID NO: 16)	<i>ace3</i> -L
pYL27	<i>dld1</i> (SEQ ID NO: 19)	<i>ace3</i> -L
pYL28	<i>xyn4</i> (SEQ ID NO: 20)	<i>ace3</i> -L
pYL29	PID 72526 (SEQ ID NO: 21)	<i>ace3</i> -L
pYL30	<i>axe1</i> (SEQ ID NO: 22)	<i>ace3</i> -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

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 <sup>1</sup>	Glu <sup>2</sup>
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	hvk1	0.96	0.73
LT87	pki1	1.16	0.98
LT149	rev3	0.96	0.76
LT150	PID 104295	0.94	0.41
LT151	tkl1	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/Sop<sup>1</sup>是“葡萄糖/槐糖”的缩写;诱导条件。

[0360] Glu<sup>2</sup>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](https://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](https://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于外显子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(gla1)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gla1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gla1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

[0369]

引物	序列
Gla.5F (SID: 49)	GTAACGCCAGGGTTTTCCAGTCACGACGGTTTAACTCCATACGCAGCAAA CATGGGCTTGGGC
Gla.5R (SID: 50)	GTACGAGTACTAGGTGTGAAGATTCCGTCAAGCTTGGCGGAATGAAGGAGG ATGTGTGAGAGG
DICprom.F (SID: 51)	CACACATCCTCCTTCATTCCGCCCCAAGCTTGACGGAATCTTCACACCTAGTA CTCGTAC
Ace3RutC.R (SID: 52)	TGACATTTTTTTGTTGTTCCAACACAGCATGCTTAGTCCGACGCCTTCGAGTC CAGCC
Ace3term.F (SID: 53)	CTGGACTCGAAGGCGTCGGACTAAGCATGCTGTGTTGGAACAACAAAAAATG TC
Ace3term.R (SID: 54)	GCAGAGCAGCAGTAGTCGATGCTATTAATTAAGTAGGTTATGCGAGCAACAT TG
Gla.3F (SID: 55)	CTCAGCCTCTCTCAGCCTCATCAGCCGCGGCCGCTGAATCGGCAAGGGGTAG TACTAG
Gla.3R (SID: 56)	GCGGATAACAATTTACACAGGAAACAGCGTTTAAACCACATGCCAGAGTTC GATGCGCAAG
Ace3_nointron.F (SID: 57)	GTACCTCAGCGCTGTCGATAGCTGCACGCACTGCCGCGATGCCACGTGCAG TGCAC
Ace3_nointron.R (SID: 58)	GTGCACTGCACGTGGGCATCGCGGCAGTGCCTGCAGCTATCGACAGCGCTGAG GTACTC
Ace3QM.F (SID: 59)	GCGGCGCTTCCGCTGTCGTAACATATGCTGCGCTACTCCCCCGTCTTAC
DICprom_QM.R (SID: 60)	GTAAGACGGGGGAGTAGCGCAGCATAGTTACGACAGCGGAAGCGCCGCTTAT AAGTG
Ace3RutC.F (SID: 61)	GGCGGCGCTTCCGCTGTCGTAACATATGGGCTCAGCAGCTCCGGCCCAGGGCTC
DICprom_rutC.R (SID: 62)	GCCCTGGGCGGAGCTGCTGAGCCCATAGTTACGACAGCGGAAGCGCCGCTT ATAAG
Ace3cDNA.F (SID: 63)	GGCGGCGCTTCCGCTGTCGTAACATATGGCCACAGCGGCCGCGGCAGCAGCTGG
DICprom_cDNA.R (SID: 64)	CAGCTGCTGCCGCGGCCGCTGTGGCCATAGTTACGACAGCGGAAGCGCCGCTT TATAAG

[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NotI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



[0374]

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
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 $\mu$ g纯化的片段以转化里氏木霉RL-P37菌株的pyr4<sup>-</sup>突变体的原生质体。如PCT公开号W0 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_5screen.F	GCTGGAAGCTGCTGAGCAGATC	65
DICprom.R	GTGCCAGCATTCCCCAGACTCG	66
T061_pyr4_orf_screen	TTAGGCGACCTCTTTTCCA	67
Gla1_3screen.R	GCCGCTCAGGCATACGAGCGAC	68
DICprom.F2	CTCTGGTCGGCCTGCCGTTG	69
ace3.R	TGAGTATAGCGGCTGACTTGTCG	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% KH<sub>2</sub>PO<sub>4</sub>、0.5% (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>、0.1% MgSO<sub>4</sub>、0.1% CaCl<sub>2</sub>、0.9% 酪蛋白氨基酸、0.048% 柠檬酸xH<sub>2</sub>O、0.05% FeSO<sub>4</sub>x7H<sub>2</sub>O、0.0003% MnSO<sub>4</sub>xH<sub>2</sub>O、0.004% ZnSO<sub>4</sub>x7H<sub>2</sub>O、0.0002% H<sub>3</sub>BO<sub>3</sub>和0.00014% CuSO<sub>4</sub>x5H<sub>2</sub>O。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	
M2015	<i>ace3</i> SC 克隆 2-1	<i>Pyr4</i>
M2016	<i>ace3</i> SC 克隆 28-3	<i>Pyr4</i>
M2017	<i>ace3</i> S 克隆 9-1	<i>Pyr4</i>
M2018	<i>ace3</i> S 克隆 20-1	<i>Pyr4</i>
[0391] M2019	<i>ace3</i> L 克隆 16-1	<i>Pyr4</i>
M2020	<i>ace3</i> L 克隆 18-1	<i>Pyr4</i>
M2021	<i>ace3</i> LC 克隆 52-1	<i>Pyr4</i>
M2022	<i>ace3</i> LC 克隆 14-4	<i>Pyr4</i>
M2023	<i>ace3</i> EL 克隆 3-3	<i>Pyr4</i>
M2024	<i>ace3</i> EL 克隆 4-5	<i>Pyr4</i>
M2025	<i>ace3</i> LN 克隆 3-3	<i>Pyr4</i>
[0392] M2026	<i>ace3</i> LN 克隆 4-1	<i>Pyr4</i>

[0393] 培养在28℃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]	乳糖	M2015	0,77
		M2016	0,54
		M2017	0,47
		M2018	1,23
		M2019	2,50
		M2020	1,97
		M2021	0,67
		M2022	1,19
		M2023	1,84
		M2024	2,05
		M2025	1,76
		M2026	1,93
		M1904	0,40
	葡萄糖	M2015	0,59
		M2016	0,12
		M2017	0,00
		M2018	0,02
		M2019	1,57
		M2020	1,79
		M2021	0,03
		M2022	0,06
		M2023	1,40
		M2024	1,37
		M2025	1,24
		M2026	1,49
[0397]		M1904	0,02

[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萄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。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,用葡萄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萄糖淀粉酶的转化体。

[0408] 因此,使表达里氏木霉葡萄糖淀粉酶的(亲本)菌株(对照)和经修饰的表达里氏木霉(子代)葡萄糖淀粉酶的菌株(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 $\mu$ g/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38 $\mu$ g/mL葡萄糖淀粉酶,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倍更高的葡萄糖淀粉酶(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倍更高的葡萄糖淀粉酶(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倍更高的葡萄糖淀粉酶(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pki1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hvk1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hvk1或dic1启动子。

[0418] 表9

[0419] PCR引物

[0420]	CL1791 (SID: 71)	TCTAGTATGTACGAGTACTAGGTGTGAAGATTCCGTCATTTCTCGACATGCGAATGCG
	CL1792 (SID: 72)	TGCCATGCAAACCCCGCATTCGCATGTCGAGGAAATGACGGAATCTTCA CACCTAGTAC
	CL1793 (SID: 73)	TGCAGCTACAGAGCCCTGGGCCGAGCTGCTGAGCCCATAGTTACGACA GCGGAAGCGC
	CL1794 (SID: 74)	ATAGCACTTATAAGGCGGCGCTTCCGCTGTCGTA ACTATGGGCTCAGCA GCTCCGGC
	CL1840 (SID: 75)	TAAACAAATAGGGGTTCGCGCACATTTCCCCGAAAAGTGCCACCTGGA TAGACTAGCA TCTGAGCCATTGCAGC
	CL1786 (SID: 76)	AGTGGCACCGAGTCGGTGGTGCTTTTTTTTCTATCGAGAGCATTGGTCAG TGGTGGCAAG
	CL1800 (SID: 77)	ACCAATATACAAAACATGTCGTCCGAGCCAGTGCCTGCCATTTCTCTCGAC ATGCGAATGC
	CL1801 (SID: 78)	GTTGCCATGCAAACCCCGCATTCGCATGTCGAGGAAATGGCAGGCACTG GCTCGGACGAC
	CL1802 (SID: 79)	AGCTACAGAGCCCTGGGCCGAGCTGCTGAGCCCATTGTTGAATTCTGG CGGGGTAGCTG
	CL1803 (SID: 80)	CTTTTACACTTTTCAACAGCTACCCCGCCAGAATTCAACAATGGGCTCA
[0421]		GCAGTCCGGC
	CL1831 (SID: 81)	TAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGGTGCTTTTT 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hvk1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	AGTCTATCGCAGCCTTGCCCTTAGCTAATGTTT	82
CL1822	TS1 的殿后寡核苷酸	TCTAAAACATTAGCTAAGGCAAGGCTGCGATA	83
CL1823	TS4 的靠前寡核苷酸	AGTCTATCGGCAGAGTCGCGTCTTCCGGGTTT	84
CL1824	TS4 的殿后寡核苷酸	TCTAAAACCCGGAAGACGCGACTCTGCCGATA	85
CL1825	TS5 的靠前寡核苷酸	AGTCTATCGAATGAGTGTAGGTACGAGTAGTTT	86
CL1826	TS5 的殿后寡核苷酸	TCTAAAACACTCGTACCTACACTCATTCGATA	87
CL1827	TS8 的靠前寡核苷酸	AGTCTATCGGCCGCAATAGCTTCCTAATGTTT	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type IIS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hvk1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 ID	引物序列	SEQ ID NO:
[0438]	CL1858	TGGAGAGACTCGGAGAGGATAGG	92
	CL1853	AGCGTGGAGGCAGTTGGAGTGG	93
[0439]	CL1848	TGGACAAAGCCTGGGTCCTGCTCC	94
	CL1818	ATCCTGACTCGTCCTGTGTCGG	95
	CL1898	AGTGCTTCGTTTAGTGGACTTG	96
	CL1850	CTCGGTAGCTGCTTGAATATAG	97

## [0440] 摇瓶中的蛋白产生

[0441] 为了测试ace3启动子替换的菌株的功能,在摇瓶中在50ml浸没培养中在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hxxk1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W0 03/027306

[0450] PCT国际申请号W0 1992/06183

[0451] PCT国际申请号W0 1992/06209

[0452] PCT国际申请号W0 1992/06221

[0453] PCT国际申请号W0 1992/10581

[0454] PCT国际申请号W0 1998/15619

[0455] PCT国际申请号W0 2002/12465



[0456] PCT国际申请号W0 2003/52054  
[0457] PCT国际申请号W0 2003/52055  
[0458] PCT国际申请号W0 2003/52056  
[0459] PCT国际申请号W0 2003/52057  
[0460] PCT国际申请号W0 2003/52118  
[0461] PCT国际申请号W0 2004/16760  
[0462] PCT国际申请号W0 2004/43980  
[0463] PCT国际申请号W0 2004/48592  
[0464] PCT国际申请号W0 2005/001036  
[0465] PCT国际申请号W0 2005/01065  
[0466] PCT国际申请号W0 2005/028636  
[0467] PCT国际申请号W0 2005/093050  
[0468] PCT国际申请号W0 2005/28636  
[0469] PCT国际申请号W0 2005/93073  
[0470] PCT国际申请号W0 2006/074005  
[0471] PCT国际申请号W0 2006/74005  
[0472] PCT国际申请号W0 2009/149202  
[0473] PCT国际申请号W0 2010/141779  
[0474] PCT国际申请号W0 2011/038019  
[0475] PCT国际申请号W0 2011/063308  
[0476] PCT国际申请号W0 2011/153276  
[0477] PCT国际申请号W0 2012/125925  
[0478] PCT国际申请号W0 2012/125951  
[0479] PCT国际申请号W0 2012125937  
[0480] PCT国际申请号W0 2013/102674  
[0481] PCT国际申请号W0 2014/047520  
[0482] PCT国际申请号W0 2014/070837  
[0483] PCT国际申请号W0 2014/070841  
[0484] PCT国际申请号W0 2014/070844  
[0485] PCT国际申请号W0 2014/093275  
[0486] PCT国际申请号W0 2014/093281  
[0487] PCT国际申请号W0 2014/093282  
[0488] PCT国际申请号W0 2014/093287  
[0489] PCT国际申请号W0 2014/093294  
[0490] PCT国际申请号W0 2015/084596  
[0491] PCT国际申请号W0 2016/069541  
[0492] PCT国际申请号W0 2016/100272  
[0493] PCT国际申请号W0 2016/100568  
[0494] PCT国际申请号W0 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.,Salohelimo,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,2<sup>nd</sup> Edition,Cold Spring Harbor Laboratory Press,Cold Spring,New York,1989.

[0528] Sambrook et al.,Molecular Cloning,A Laboratory Manual,4<sup>th</sup> 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

[0001]

```

atgtgcgct actccccgt cttacacctg gatactctct cettgccacc actgaccaat      60
gctcttcccc gcccaaagtg cgagtacctc agcgtgtcg atagctgcac gactgcccgc      120
gatgcccacg tgcagtgcac ttctgacctg cccctggcgc gacgcggccc caaagcgagg      180
aagaagagcg accagcccgg ccagccgcct cctgatccga gctcgtctct caccgcggtc      240
cgaccgggcc agatgccgcc gccgtgacc ttctcggccc ccgcagtagc cgcgtgcag      300
cccttcgcct cgtcgtcgtc gtcgcccgc gcggcctggg agcccgtcga gccgtcagc      360
attgacaacg gcctgccccg gcagccgctg ggcgacctgc ccggcctctc caccatccag      420
aacatctcga cgcgccagcg atggatacac ctggccaacg ccatgacgct gcgcaaacag      480
acgctagagc gcgtctcgaa gcgatgtatc gacctcttct tcgactacct ctacccccctc      540
acccccctgg tgtacgagcc ggccctccgg gacgtgctcg catacatctt ctccagccc      600
ttgcctggcg tcaaccaacc atcgccgctg tcacagctca cgccagaccc gaccaccggc      660
accaccccc tcaacgctgc cgagtcgtgg gccggctttg gccagcccag cggctcgcga      720
accgtcggca gcaggtggc tccttgggcc gactcgacct tcaccttggc cagggccgtc      780
tgcgagaggc cagcattcat gctacccaag gacatcttcc ccgaaggaga atccgtctct      840
gagatcttgc tcgaagcctc tcgggactgc ctgcaccagc acctcgaggc cgacctggag      900
aatccgacgg ccaactcgat tgccattcgc tacttccact ccaactgcct ccacgtcgcg      960
gggaagccca agtactcgtg gcacataatt ggcgaggcca tccgcctggc gcaggtcatg      1020
cagctgcacg aggaggtgc cctcgagggg ctctgcccc tgcaggcaga gttccgcccgt      1080
cgctgctttt ggatcctgta cttgggcgac aagtcagccg ctatactcaa caatcgcccc      1140
atcaccatcc acaagtactg cttcgacgcc ggcatcacca cgctataccc gtcgggtatc      1200
gaggacgagt tcctgagcac ggcgccgag ccgccccgga agagcttcat atccggcttc      1260
aacgcaaattg tgcggctctg gcagtcgcgc gctgatttgc tgctggaaat ccgctgctg      1320
caagatcaga tgatgcagca ctttcgaggg accatgcccc cgaacctgtg gctgccctcc      1380
gccgacaggc agcatctcga ttctctctat gtccgcttca tcacctgctt ggacgatctc      1440
ccgccgtacc tccagtcgtg cactctggcg atggcagcga tggcagaagg caacgggtct      1500
gccgagtcga agcagtagct gatacagtgc atcaacctgc aggtgacgtt tcactgtctg      1560
cgcatggtaa ttacgcagaa attcgaagac ctctcttatt ttgctcctgg cgttgagcag      1620
gctgatctca gaaagtcgga gattgtgcga gacatgctga gggatgatga cgaggcggcc      1680
ttttggggcc tgcaggccaa tggcgagcca aacgtgagtc gtttccttgt ctcttctctt      1740
ttctgcacac ccttttcttc gacgaccccc cctctctctt tataatccctg cgatatgta      1800
tatcatcaag cctcggcact tgttgctaatt ctgtcctgat tatgttgtct ggatgctgca      1860

```

	ggttgaaaag attcgccctta tccgagctag ttgtctggcc atcatccatc gcaaccagga	1920
	ttcacccttg gctacgcgag ccaggagcga cttttccgtg cttttggata ttctcacgcg	1980
	gctggactcg aaggcgtcgg accaactgag gaatacgtcc actaccgttg ttggctaa	2038
	<210> 2	
	<211> 1890	
	<212> DNA	
	<213> 里氏木霉	
	<400> 2	
	atgtctgcgt actccccgt cttacacctg gatactctct ccttgccacc actgaccaat	60
	gctcttcccc gcccaaagtg cgagtacctc agcgtgtctg atagctgcac gcaactgccg	120
	gatgcccacg tgcagtgcac ttctgacctg cccttggcgc gacgcggccc caaagcgagg	180
	aagaagagcg accagcccgg ccagccgcct cctgatccga gctcgtctct caccgcggct	240
	cgacccggcc agatgccgcc gccgctgacc ttctccggcc ccgcagtagc cgcgctgcag	300
	cccttcgect cgtcgtcgtg gtgcgccgac gcggcctggg agcccgtcga gccgctcagc	360
	attgacaacg gcctgccccg gcagccgctg ggcgacctgc ccggcctctc caccatccag	420
	aacatctcga cgcgccagcg atggatacac ctggccaacg ccatgacgtg gcgcaacacg	480
	acgctagagc gcgtctcgaa gcgatgtatc gacctcttct tcgactacct ctacccccctc	540
	acccccctgg tgtacgagcc ggccctccgg gacgtgctcg catacatctt ctcccagccc	600
	ttgcctggcg tcaaccaacc atcgccgctg tcacagctca cgccagacce gaccaccggc	660
	accaccccc tcaacgtctg cgagtcgttg gccggctttg gccagcccag cggctcgcga	720
	accgtcggca gcaggtggc tccctgggcc gactcgacct tcacctgggt cagggccgtc	780
	tgcgagagg cagcattcat gctaccaag gacatcttcc ccgaaggaga atccgtctct	840
[0002]	gagatcttgc tcgaagcctc tcgggactgc ctgcaccagc acctcgaggc cgacctggag	900
	aatccgacgg ccaactcgat tgccattcgc tacttccact ccaactgcct ccacgtctcg	960
	gggaagccca agtactcgtg gcacatattt ggcgaggcca tccgcctggc gcaggtcatg	1020
	cagctgcacg aggaggtgc cctcgagggg ctctgccca tcgaggcaga gttccgccgt	1080
	cgtgctttt ggatcctgta ctggggcagc aagtcagccg ctatactcaa caatcgcccc	1140
	atcaccatcc acaagtactg ctctgacgcc ggcatcacca cgctataccc gtcgggtatc	1200
	gaggacgagt tectgagcac ggctccgag ccgccccgga agagcttcat atccggcttc	1260
	aacgcaaatg tgcggctctg gcagtcgcg gctgatttgc tgctggaaat ccgctgctg	1320
	caagatcaga tgatgcagca ctttcgaggg accatgcccc cgaacctatg gctgccctcc	1380
	gccgacaggc agcatctcga ttctctctat gtccgcttca tcacctgctt ggacgatctc	1440
	ccgccgtacc tccagtcgtg cactctggcg atggcagcga tggcagaagg caacgggtct	1500
	gccgagcca agcagtacgt gatacagtgc atcaacctgc aggtgacgtt tcaactgtctg	1560
	cgcatggtaa ttacgcagaa attcgaagac ctctcttatt ttgctcctgg cgttgagcag	1620
	gctgatctca gaaagtcgga gattgtgcga gacatgtga gggatgatga cgaggcggcc	1680
	ttttggggcc tgcaggccaa tggcgagcca aacgttgaaa agattcgctt tatcgagct	1740
	agtttgctgg ccatcatcca tcgcaaccag gattcaccct tggctacgcg agccaggagc	1800
	gacttttccg tgcttttga tattctcacg cggctggact cgaaggcgtc ggaccaactg	1860
	aggaatacgt ccactaccgt tgttggttaa	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 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 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  
 [0003] 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  
 [0004] 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 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 ccagggtctt gtagctgcag ctgcaggcgg cctccagct 60  
 gctggcgctg gcgctggcgc tgtccacgcc ctcaccacct cgccccagtc tgcctcgccc 120

	tcgcagcccg gctcgccaac cgcccaacc acgcgcgcgc agaactcact cgtgtcggt	180
	gcaacctcgt tccaccacca tcccagagge cgtctgggtga gcagagcctg cgaccgctgc	240
	cgccggcgca aggccaaagt cagtctagcc cctttgctgt tgcttgcatc tctgttgta	300
	ttgctcctcc tctgtctgt gctgatgctg ctgctcctcc tctcctcct cctccccgtc	360
	tcttggtccc tgggtccctgc tcttcataatg tcttactgc ccgtgtctcc tctccccgtt	420
	cccgttcccc ctctcccggt cctcttctcc tgcgtgtctg tcatgcgtac aaagcataca	480
	tacaatacat cagcatacat ggcaagcgtt gtgttgtgtt gagagtgtg tgtattgtat	540
	tgcactgcct tcacaactcg ttcatactgc tgcagcctca ccccaacacc gacctcgtct	600
	tccatgctgc gctactcccc cgtcttacac ctggatactc tctccttgcc accactgacc	660
	aatgctcttc cccgccccaa gtgcgagtac ctgagcgtg tcgatagctg cagcactgc	720
	cgcgatgccc acgtgcagtg cactttcgac ctgcccctgg cgcgacgcgg ccccaaagcg	780
	aggaagaaga gcgaccagcc cgccagccg cctcctgac cgagctcgt ctccaccgcg	840
	gctcgaccgg gccagatgcc cccgcccgtg accttctccg gccccgcagt agccgcgctg	900
	cagcccttcg cctcgtcgtc gctgtcgccc gacgcggcct gggagcccgt cgagccgctc	960
	agcattgaca acggcctgcc ccggcagccg ctgggcgacc tgcccgcct ctccaccatc	1020
	cagaacatct cgacgcgcca gcgatggata cacctggcca acgcatgac gctgcgcaac	1080
	acgacgctag agcgcgtctc gaagcgaigt atcgacctct tcttcgacta cctctacccc	1140
	ctcaccccc tggtgtacga gccggccctc cgggacgtgc tcgcatacat cttctcccag	1200
	cccttgctg gcgtcaacca accatgcccg ctgtcacagc tcacgccaga cccgaccacc	1260
	ggcaccaccc cctcaacgc tgccgagtgc tgggcgggt ttggccagcc cagcggctcg	1320
	cgaaccgtcg gcagcaggct ggctccctgg gccgactcga ccttcacctt ggtcacggcc	1380
[0005]	gtctgcgcag aggcagcatt catgctaccc aaggacattt tccccgaagg agaatccgtc	1440
	tctgagatct tgctcgaagc ctctcgggac tgctgcacc agcacctcga ggccgacctg	1500
	gagaatccga cggccaactc gattgccatt cgctacttcc actccaactg cctccacgt	1560
	gcggggaagc ccaagtactc gtggcacata tttggcgagg ccatccgcct ggccgaggtc	1620
	atgcagctgc acgaggagge tgccctcgag gggctcgtcc ccatcgagge agagtccgc	1680
	cgctcgtgct tttggatcct gtacttgggc gacaagtcag ccgctatact caacaatcgg	1740
	cccatcacca tccacaagta ctgcttcgac gccggcatca ccacgtata cccgtcgggt	1800
	atcgaggacg agttcttgag cacggcgtcc gagccgcccc ggaagagctt catatccggc	1860
	ttcaacgcaa atgtgcggct ctggcagtc gccgctgatt tgctgctgga aatccgcgtg	1920
	ctgcaagatc agatgatgca gcactttcga gggaccatgc cccgaacca tgtgtgccc	1980
	tccgccgaca ggcagcatct cgattctctc tatgtccgt tcatcacctg cttggacgat	2040
	ctcccgcgt acctccagtc gtgcactctg gcgatggcag cgatggcaga aggcaacggg	2100
	tctgccgagt ccaagcagta cgtgatacag tgcatcaacc tgcaggtgac gtttactgt	2160
	ctgcgatgg taattacgca gaaattcgaa gacctctctt attttgcctc tggcggtgag	2220
	cagcgtgatc tcagaaagtc ggagattgtg cgagacatgc tgagggtgat gaacgaggcg	2280
	cccttttggg gcctgcagge caatggcgag ccaaactga gtcgtttcct tgtctctct	2340
	cttttctgca cacccttttc ttcgacgacc cccctctct ctttatatcc ctgcggatat	2400
	gtatatcatc aagcctcggc acttgttgcct aatctgtcct gattatgttg tctggatgct	2460
	gcaggttgaa aagattcgcc ttatcggagc tagtttgcgt gccatcatcc atcgcaacca	2520
	ggattcacc ttggctacgc gagccaggag cgacttttcc gtgcttttgg atattctcac	2580
	gcggctggac tcgaaggcgt cggactaa	2608
	<210> 5	
	<211> 2037	



&lt;212&gt; DNA

&lt;213&gt; 里氏木霉

&lt;400&gt; 5

atgggctcag cagctccggc ccagggtctt gtagctgcag ctgcaggcgg ccctccagct 60  
 gctggcgctg gcgctggcgc tgtccacgcc ctcaccacct cgcccagatc tgcctcggcc 120  
 tcgcagcccg gctcgccaac cgcttcaacc acgcccgcgc agaactcaact cgtgtcggct 180  
 gcaacctcgt tccaccacca tcccagaggc cgtctgggtga gcagagcctg cgaccgctgc 240  
 cgccggcgca aggccaagtg cgagtacctc agcgtgtctg atagctgcac gcaactgccg 300  
 gatgcccacg tgcagtgcac ttctgacctg cccctggcgc gacgcggccc caaagcgagg 360  
 aagaagagcg accagcccgg ccagccgcct cctgatccga gctcgtcttc caccgcggct 420  
 cgaccgggcc agatgcgcgc gccgctgacc ttctccggcc ccgcagtagc cgcgctgcag 480  
 cccttcgctt cgtcgtcgtt gtcgcccgcg gcggcctggg agcccgtcga gccgctcagc 540  
 attgacaacg gcctgccccg gcagccgctg ggcgacctgc ccggcctctc caccatccag 600  
 aacatctcga cgcgccagcg atggatacac ctggccaacg ccatgacgct gcgcaacacg 660  
 acgctagagc gcgtctcgaa gcgatgtatc gacctcttct tcgactacct ctaccccttc 720  
 acccccctgg tgtacgagcc gcccttccgg gacgtgtctg catacatctt ctcccagccc 780  
 ttgcctggcg tcaaccaacc atcgccgctg tcacagctca cgccagacct gaccaccggc 840  
 accaccccc tcaacgctgc cgagtcgttg gccggctttg gccagcccag cggctcgcga 900  
 accgtcggca gcaggctggc tccctgggcc gactcgacct tcaccttggc caccggcgtc 960  
 tgcgcagagg cagcattcat gctacccaag gacatcttcc ccgaaggaga atccgtctct 1020  
 gagatcttgc tcgaagcctc tcgggactgc ctgcaccagc acctcgaggc cgacctggag 1080  
 aatccgacgg ccaactcgat tgccattcgc tacttccact ccaactgcct ccacgtgcg 1140  
 ggggaagcca agtactcgtg gcacatattt ggcgaggcca tccgcctggc gcaggtcatg 1200  
 cagctgcacg aggaggtgc cctcgagggg ctctgccca tcgaggcaga gttccgccgt 1260  
 cgctgctttt ggatcctgta cttgggcgac aagtcagccg ctatactcaa caatcgcccc 1320  
 atcaccatcc acaagtactg ctctgacgcc ggcatcacca cgtataccc gtcgggtatc 1380  
 gaggacgagt tcctgagcac ggcgctccgag ccgccccgga agagcttcat atccggttc 1440  
 aacgcaaatg tgcggctctg gcagtcgcg gctgatttgc tgctggaaat ccgctgtctg 1500  
 caagatcaga tgatgcagca ctttcgaggg accatgcccc cgaacctatg gctgccctcc 1560  
 gccgacaggc agcatctcga ttctctctat gtccgcttca tcacctgctt ggacgatctc 1620  
 ccgccgtacc tccagtcgtg cactctggcg atggcagcga tggcagaagg caacgggtct 1680  
 gccgagtcga agcagtacgt gatacagtgc atcaacctgc aggtgacgtt tcaactgtctg 1740  
 cgcatggtaa ttacgcagaa attcgaagac ctctcttatt ttgctcctgg cgttgagcag 1800  
 gctgatctca gaaagtcgga gattgtgcga gacatgtga ggggtgatgaa cgaggcggcc 1860  
 ttttggggcc tgcaggccaa tggcgagcca aacgttgaaa agattcgctt tatcgagct 1920  
 agtttgcctg ccatcatcca tcgcaaccag gattaccctt tggctacgcg agccaggagc 1980  
 gacttttccg tgcttttga tattctcagc cggtggact cgaaggcgtc ggactaa 2037

[0006]

&lt;210&gt; 6

&lt;211&gt; 678

&lt;212&gt; PRT

&lt;213&gt; 里氏木霉

&lt;400&gt; 6

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

1

5

10

15

[0007]

Gly	Pro	Pro	Ala	Ala	Gly	Ala	Gly	Ala	Gly	Ala	Val	His	Ala	Leu	Thr
Thr	Ser	Pro	Glu	Ser	Ala	Ser	Ala	Ser	Gln	Pro	Gly	Ser	Pro	Thr	Ala
Ser	Thr	Thr	Pro	Pro	Gln	Asn	Ser	Leu	Val	Ser	Ala	Ala	Thr	Ser	Phe
His	His	His	Pro	Arg	Gly	Arg	Leu	Val	Ser	Arg	Ala	Cys	Asp	Arg	Cys
Arg	Arg	Arg	Lys	Ala	Lys	Cys	Glu	Tyr	Leu	Ser	Ala	Val	Asp	Ser	Cys
Thr	His	Cys	Arg	Asp	Ala	His	Val	Gln	Cys	Thr	Phe	Asp	Leu	Pro	Leu
Ala	Arg	Arg	Gly	Pro	Lys	Ala	Arg	Lys	Lys	Ser	Asp	Gln	Pro	Gly	Gln
Pro	Pro	Pro	Asp	Pro	Ser	Ser	Leu	Ser	Thr	Ala	Ala	Arg	Pro	Gly	Gln
Met	Pro	Pro	Pro	Leu	Thr	Phe	Ser	Gly	Pro	Ala	Val	Ala	Ala	Leu	Gln
Pro	Phe	Ala	Ser	Ser	Ser	Leu	Ser	Pro	Asp	Ala	Ala	Trp	Glu	Pro	Val
Glu	Pro	Leu	Ser	Ile	Asp	Asn	Gly	Leu	Pro	Arg	Gln	Pro	Leu	Gly	Asp
Leu	Pro	Gly	Leu	Ser	Thr	Ile	Gln	Asn	Ile	Ser	Thr	Arg	Gln	Arg	Trp
Ile	His	Leu	Ala	Asn	Ala	Met	Thr	Leu	Arg	Asn	Thr	Thr	Leu	Glu	Arg
Val	Ser	Lys	Arg	Cys	Ile	Asp	Leu	Phe	Phe	Asp	Tyr	Leu	Tyr	Pro	Leu
Thr	Pro	Leu	Val	Tyr	Glu	Pro	Ala	Leu	Arg	Asp	Val	Leu	Ala	Tyr	Ile
Phe	Ser	Gln	Pro	Leu	Pro	Gly	Val	Asn	Gln	Pro	Ser	Pro	Leu	Ser	Gln
Leu	Thr	Pro	Asp	Pro	Thr	Thr	Gly	Thr	Thr	Pro	Leu	Asn	Ala	Ala	Glu
Ser	Trp	Ala	Gly	Phe	Gly	Gln	Pro	Ser	Gly	Ser	Arg	Thr	Val	Gly	Ser
Arg	Leu	Ala	Pro	Trp	Ala	Asp	Ser	Thr	Phe	Thr	Leu	Val	Thr	Ala	Val
Cys	Ala	Glu	Ala	Ala	Phe	Met	Leu	Pro	Lys	Asp	Ile	Phe	Pro	Glu	Gly
Glu	Ser	Val	Ser	Glu	Ile	Leu	Leu	Glu	Ala	Ser	Arg	Asp	Cys	Leu	His
Gln	His	Leu	Glu	Ala	Asp	Leu	Glu	Asn	Pro	Thr	Ala	Asn	Ser	Ile	Ala

[0008]

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  
 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

[0009]	<400> 7	
	atgctgcgct actccccgt cttacacctg gatactctct ccttgccacc actgaccaat	60
	gctcttcccc gcccaaagtg cgagtacctc agcgtgtctg atagctgcac gcaactgccg	120
	gatccccacg tgcagtgcac ttctgacctg cccctggcgc gacgcggccc caaagcgagg	180
	aagaagagcg accagcccgg ccagccgcct cctgatccga gctcgtcttc caccgcggct	240
	cgacccggcc agatgccgcc gccgctgacc ttctccggcc ccgcagtagc cgcgctgcag	300
	cccttcgcct cgtcgtcgtc gtcgcccgcg gcggcctggg agcccgtcga gccgctcagc	360
	attgacaacg gcctgccccg gcagccgctg ggcgacctgc ccggcctctc caccatccag	420
	aacatctcga cgcgccagcg atggatacac ctggccaacg ccatgacgct gcgcaacacg	480
	acgctagagc gcgtctcgaa gcgatgtatc gacctcttct tcgaactacct ctacccccctc	540
	acccccctgg tgtacgagcc ggccctccgg gacgtgctcg catacatctt ctcccagccc	600
	ttgcctggcg tcaaccaacc atcgccgctg tcacagctca cgccagaccc gaccaccggc	660
	accaccccc tcaacgctgc cgagtcgtgg gccggctttg gccagcccag cggctcgcga	720
	accgtcggca gcaggctggc tccctgggcc gactcgacct tcacctggt caccggcgtc	780
	tgcgcagagg cagcattcat gctacccaag gacattttcc ccgaaggaga atccgtctct	840
	gagatcttgc tcgaagcctc tcgggactgc ctgcaccagc acctcgaggc cgacctggag	900
	aatccgacgg ccaactcgat tgccattcgc tacttccact ccaactgcct ccacgtgcg	960
	gggaagccca agtactcgtg gcacatattt ggcgaggcca tccgectggc gcaggctcatg	1020
	cagctgcacg aggaggctgc cctcgagggg ctctgccccca tcgaggcaga gttccgccgt	1080
	cgctgctttt ggatcctgta cttgggcgac aagtcagccg ctatactcaa caatcggccc	1140
	atcaccatcc acaagtactg cttcgacgcc ggcatcacca cgctataccc gtcgggtatc	1200
	gaggacgagt tcctgagcac ggcgctccgag ccgccccgga agagcttcat atccggcttc	1260
	aacgcaaatg tgcggctctg gcagtcgagc gctgatttgc tgctggaaat ccgctgctg	1320
	caagatcaga tgatgcagca ctttcgaggg accatgcccc cgaacctagt gctgccctcc	1380
	gccgacaggc agcatctcga ttctctctat gtccgcttca tcacctgctt ggacgatctc	1440
	ccgccgtacc tccagtcgtg cactctggcg atggcagcga tggcagaagg caacgggtct	1500
	gccgagtcca agcagtagct gatacagtcg atcaacctgc aggtgacgtt tcactgtctg	1560
	cgcatggtaa ttacgcagaa attcgaagac ctctcttatt ttgctcctgg cgttgagcag	1620
	gctgatctca gaaagtcgga gattgtgcga gacatgctga ggggtgatgaa cgaggcggcc	1680
	ttttggggcc tgcaggccaa tggcgagcca aacgtgagtc gtttccttgt ctcttctctt	1740
	ttctgcacac cctttttctc gacgaccccc cctctctctt tatatccctg cggatatgta	1800
	tatcatcaag cctcggcact tgttgctaat ctgtctgat tatgttgtct ggatgctgca	1860
	ggttgaaaag attcgcttta tcggagctag tttgctggcc atcatccatc gcaaccagga	1920
	ttcaccttg gctacgcgag ccaggagcga cttttccgtg cttttggata ttctcacgcg	1980
	gctggactcg aaggcgtcgg 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 10 15	
	Pro Leu Thr Asn Ala Leu Pro Arg Pro Lys Cys Glu Tyr Leu Ser Ala	
	20 25 30	

[0010]

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	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
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
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			
	515	520	525	
	Glu Asp Leu Ser Tyr Phe Ala Pro Gly Val Glu Gln Ala Asp Leu Arg			
	530	535	540	
[0011]	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 cgcccagatc tgcctcggcc		120	
	tgcagcccg gctcgccaac cgctcaacc acgccgcgc agaactcact cgtgtcggt		180	
	gcaacctcgt tccaccacca tccagaggc cgtctggtga gcagagcctg cgaccgctgc		240	
	cgccggcgca aggccaaggt cagtctagcc ctttctgtgt tgccttcate tctgtgtca		300	
	ttgctectec tctgtctgt getgatgtg ctgtectec tectectec cctccccgtc		360	
	tcttggtccc tggctccctgc tctcatatg tcttactgc ccgtgtctec tctccccgtt		420	
	ccgttcccc ctcctccgt cctcttctcc tgcgtgtctg tcatgcgtac aaagcataca		480	

	tacaatacat cagcatacat ggcaagcggt gtgttgtgtt gagagttgtg tgtattgtat	540
	tgcactgcct tcacaactcg ttcatactgc tgcagcctca ccccaacacc gacctcgtct	600
	tccatgctgc gctactcccc cgtcttacac ctggatactc tctccttgcc accactgacc	660
	aatgctcttc cccgccc aaa gtgcgagtac ctgagcgctg tcgatatgctg cagcactgc	720
	cgcgatgccc acgtgcagtg cactttcgac ctgcccctgg cgcgacgcgg ccccaaagcg	780
	aggaagaaga ggcaccagcc cggccagccg cctcctgac cgagctcgtc ctccaccgcg	840
	gctcgacccg gccagatgcc gccgcgctg accttctccg gccccgcagt agccgcgctg	900
	cagcccttcg cctcgtcgtc gctgtgccc gacgcggcct gggagcccg cgagccgctc	960
	agcattgaca acggcctgcc ccggcagccg ctgggcgacc tgcccgccct ctccaccatc	1020
	cagaacatct cgacgcgcca gcgatggata cacctggcca acgccatgac gctgcgcaac	1080
	acgacgctag agcgcgtctc gaagcgaigt atcgacctct tcttcgacta cctctacccc	1140
	ctcaccccc tgggtgtacga gccggccctc cgggacgtgc tcgcatacat cttctcccag	1200
	cccttgccctg gcgtcaacca accatgcccg ctgtcacagc tcacgccaga cccgaccacc	1260
	ggcaccaccc cctcaacgc tgcagatcg tggcgccggt ttggccagcc cagcggtcgc	1320
	cgaaccgctg gcagcaggct ggctccctgg gccgactcga ccttcacctt ggtcacggcc	1380
	gtctgcgcag aggcagcatt catgctaccc aaggacattt tccccgaagg agaatccgtc	1440
	tctgagatct tgctcgaagc ctctcgggac tgcttgacc agcacctcga ggccgacctg	1500
	gagaatccga cggccaactc gattgccatt cgtacttcc actccaactg cctccacgct	1560
	gcggggaagc ccaagtactc ttggcacata tttggcgagg ccattccgct ggcgaggctc	1620
	atgcagctgc acgaggaggc tgccctcgag gggctcgtcc ccacgaggc agagttccgc	1680
	cgtcgtcgtc tttggatcct gtacttgggc gacaagtcag ccgtatact caacaatcgg	1740
[0012]	cccatcacca tccacaagta ctgcttcgac gccggcatca ccacgtata cccgtcgggt	1800
	atcaggagc agttcctgag cagcgctcc gagccgccc ggaagagctt catatccggc	1860
	ttcaacgcaa atgtgcggt ctggcagtc gccgctgatt tgcgtctgga aatccgcgtg	1920
	ctgcaagatc agatgatgca gcactttcga gggaccatgc cccgaacca tgtctgccc	1980
	tccgccgaca ggcagcatct cgattctctc tatgtccgt tcacacctg cttggacgat	2040
	ctcccgccgt acctccagtc gtgcaactcg gcgatggcag cgatggcaga aggcaacggg	2100
	tctgccgagt ccaagcagta cgtgatacag tgcatacacc tgcaggtgac gtttactgt	2160
	ctgcgcatgg taattacgca gaaattcgaa gacctctctt attttctcc tggcgttgag	2220
	caggctgac tcagaaagtc ggagattgtg cgagacatgc tgagggtgat gaacgaggcg	2280
	cccttttggg gcctgcagc caatggcgag ccaaactga gtcgtttcct tgtctcttct	2340
	ctttcttgca caccctttt ttcgacgacc cccctctct ctttatatcc ctgcggatat	2400
	gtatatcacc aagcctcgcc acttgttct aatctgtcct gattatgttg tctggatgct	2460
	gcaggttgaa aagattcgcc ttatcggagc tagtttctg gccatcatcc atcgcaacca	2520
	ggattcaccc ttggctacgc gagccaggag cgacttttcc gtgcttttgg atattctcac	2580
	gcgctggac tcgaaggcgt cggaccaact gaggaatacg 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 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	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		
355						360						365					



[0014]

```

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
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

```

&lt;210&gt; 11

&lt;211&gt; 2885

&lt;212&gt; DNA

&lt;213&gt; 里氏木霉

&lt;400&gt; 11

atggccacag cggccgcggc agcagctggc ggcgcggcgg ttgctgcggg tgcagacaca 60  
 ggtgcgttga gtcccgctcc gtccgctcgc gtccctccc agctgccagc ccgctgggt 120  
 ggcaactgaa cgcagtgca cgaatcagt gcagtcggc cccccaact aacgtgcc 180  
 cccgtggctc ctccggcaca caggcgctgc aggtccagc tctacaggcc ctccaggcct 240  
 tccagggtt ccaggcacc ggacaggctc cgtggcgatg ggctcagcag ctccggccca 300  
 gggctctgta gctgcagctg caggcggccc tccagctgct ggcgctggcg ctggcgctgt 360  
 ccacgccctc accacctcgc ccgagtctgc ctccggcctc cagcccggt cgccaaccgc 420  
 ctcaaccacg ccgcccgaga actcactcgt gtcggctgca acctcgctcc accaccatcc 480  
 cagaggccgt ctggtgagca gagcctgcga ccgtgccgc cggcgcaagg ccaaggtcag 540  
 tctagccct ttgctgtgc ttgcatctct gttgtcattg ctctctctcc tgctgtgct 600  
 gatgtgtctg ctctctctcc tctctctct cccgtctcc tggctccctg tccctgtct 660  
 tcatatgtcc ttactgcccg tgtctctct cccgttccc gttcccccct ctcccgtcct 720  
 ctctctctgc gtgtctgtca tgcgtacaaa gcatacatc aatacatcag catacatggc 780  
 aagcgtttgt ttgtgttgag agttgtgtgt attgtattgc actgccttca caactcgttc 840  
 atactgtgc agcctcacc caacaccgac ctctcttcc atgtgcgct actccccgt 900  
 cttacacctg gatactctct ccttgccacc actgaccaat gctcttccc gcccaaagt 960  
 cgagtacctc agcgtgtgc atagctgcac gactgccgc gatgccacg tgcagtcac 1020  
 tttcgacctg cccctggcg gacgcggccc caaagcgagg aagaagagcg accagcccgg 1080  
 ccagccgct cctgatccga gctcgtctc caccgcggct cgaccggcc agatgccgc 1140  
 gccgtgacc ttctccggc ccgcagtagc cgcgtgcag cccttcgct cgctcgtct 1200  
 gtcgccgac gcggcctggg agcccgtcga gccgtcagc attgacaac gcctgccccg 1260  
 gcagccgctg ggcgacctgc ccggcctctc caccatccag aacatctcga cgcgccagcg 1320  
 atggatacac ctggccaacg ccatgacgt gcgcaaacg acgctagagc gcgtctcga 1380  
 gcgatgtatc gacctctctc tcgactacct ctacccctc accccctgg tgtacagcc 1440  
 ggccctccgg gacgtgtcgc catacatctt ctccagccc ttgcctggcg tcaaccaacc 1500  
 atcgcgctg tcacagctca cgcagaccc gaccaccgg accaccccc tcaacgtgc 1560  
 cgagtcgtgg gccggctttg gccagcccag cggctcgcga accgtcggca gcaggctggc 1620  
 tccctgggccc gactcgaact tcacctggg cagcgccgc tgcgcagagg cagcattcat 1680  
 gctaccaag gacatttcc ccgaaggaga atccgtctct gagatcttgc tcgaagctc 1740  
 tcgggactgc ctgcaccagc acctcgaggc cgacctggag aatccgacgg ccaactcgat 1800  
 tgccattcgc tacttccact ccaactgcct ccacgtgcg gggaagccca agtactcgtg 1860  
 gcacatattt ggcgaggcca tccgctggc gcaggctcat gactgcacg aggaggtgc 1920  
 cctcgagggg ctctcccca tcgaggcaga gttccgccgt cgctgctttt ggatcctgta 1980  
 cttgggcgac aagtcagccg ctatactcaa caatcgccc atcaccatcc acaagtactg 2040  
 cttcgacgcc ggcatcacca cgctataccc gtcgggtatc gaggacgagt tcctgagcac 2100  
 ggcgtccgag ccgcccggga agagcttcat atccgcttc aacgcaaat tgcggctctg 2160  
 gcagtcgcg gctgatttgc tgctggaaat ccgctgtctg caagatcaga tgatgcagca 2220  
 ctttcgaggg accatcccc cgaacctgt gctgccctc gccgacaggc agcatctcga 2280  
 ttctctctat gtccgcttca tcacctgctt ggacgatctc ccgccgtacc tccagtcgtg 2340  
 cactctggcg atggcagcga tggcagaagg caacgggtct gccgagtcca agcagtacgt 2400  
 gatacagctg atcaacctgc aggtgacgtt tcaactgtct cgcatggtta ttacgcagaa 2460

	attcgaagac ctctcttatt ttgtcctcgg cgttgagcag gctgatctca gaaagtcgga	2520
	gattgtgcga gacatgctga gggatgatgaa cgagcgcccc ttttggggcc tgcaggccaa	2580
	tggcgagcca aacgtgagtc gtttccttgt ctctctctctt ttctgcacac cttttcttc	2640
	gacgaccccc cctctctctt tatatccctg cggatatgta tatcatcaag cctcggcact	2700
	tgttgctaatt ctgtcctgat tatgttgtct ggatgctgca gggtgaaaag attcgcccta	2760
	tcggagctag tttgctggcc atcatccatc gcaaccagga ttcacccttg gctacgcgag	2820
	ccaggagcga cttttccgtg cttttggata ttctcacgcg gctggactcg aaggcgctcg	2880
	actaa	2885
	<210> 12	
	<211> 723	
	<212> PRT	
	<213> 里氏木霉	
	<400> 12	
	Met Ala Thr Ala 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 Met Gly Ser	
	35 40 45	
	Ala Ala Pro Ala Gln Gly Ser Val Ala Ala Ala Ala Gly Gly Pro Pro	
	50 55 60	
[0016]	Ala Ala Gly Ala Gly Ala Gly Ala Val His Ala Leu Thr Thr Ser Pro	
	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 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	

[0017]

	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 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
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

<400> 13

atgggctcag cagctccggc ccagggtctt gtagctgcag ctgcaggcgg cctccagct 60  
 gctggcgtg gcgtggcgc tgtccacgcc ctcaccacct cgcccagatc tgccctggcc 120  
 tcgcagcccg gctcgccaac cgctcaacc acgccgccgc agaactcact cgtgtcggct 180  
 gcaacctcgt tccaccacca tccagaggc cgtctggtga gcagagcctg cgaccgtgc 240  
 cgccggcgca aggccaagtg cgagtacctc agcgtgtcgc atagctgcac gcaactgccg 300  
 gatgccacg tgcagtgcac ttctgacctg cccctggcgc gacgcggccc caaagcgagg 360  
 aagaagagcg accagccggg ccagccgctt cctgatccga gctcgtcttc caccgcggct 420  
 cgaccgggcc agatgccgcc gccgtgacc ttctccggcc ccgcagtagc cgcgtgcag 480  
 cccttcgctt cgtcgtcgtt gtcgcccgc gggcctggg agcccgctga gccgtcagc 540  
 attgacaacg gcctgccccg gcagccgctg ggcgacctgc ccggcctctc caccatccag 600  
 aacatctcga cgcgccagcg atggatacac ctggccaacg ccatgacgct gcgcaacacg 660  
 acgctagagc gcgtctcgaa gcgatgtatc gacctttctt tcgactacct ctaccctctc 720  
 acccccctgg tgtacgagcc ggccctccgg gacgtgctcg catacatctt ctcccagccc 780  
 ttgcttgccg tcaaccaacc atcgccgctg tcacagctca cgccagaccc gaccaccggc 840  
 accaccccc tcaacgtgc cgagtctgg gccggctttg gccagcccag cggctcgcga 900  
 accgtcgcca gcaggtggc tccctgggcc gactcgacct tcacctggt caccgcccgc 960  
 tgcgagagc cagcattcat gctacccaag gacattttcc ccgaaggaga atccgtctct 1020  
 gagatcttgc tcgaagctc tcgggactgc ctgcaccagc acctcgaggc cgacctggag 1080  
 aatccgacgg ccaactcgat tgccattcgc tacttccact ccaactgcct ccacgtcgcg 1140  
 gggaagccca agtactcgtg gcacatattt ggcgaggcca tccgcctggc gcaggtcatg 1200  
 cagctgcacg aggaggtgc cctcgagggg ctctgccccca tcgaggcaga gttccgccgt 1260

[0018]

```

cgctgctttt ggatcctgta cttgggcgac aagtcagccg ctatactcaa caatcggccc 1320
atcaccatcc acaagtactg cttcgacgcc ggcatcacca cgetataccc gtcgggtatc 1380
gaggacgagt tcctgagcac ggcgccgag ccgccccgga agagcttcat atccggttc 1440
aacgcaaatg tgcggctctg gcagtcgcgc gctgatttgc tgcgtgaaat ccgctgctg 1500
caagatcaga tgatgcagca ctttcgaggg accatgcccc cgaacatgt gctgccctcc 1560
gccgacaggc agcatctcga ttctctctat gtccgttca tcacctgctt ggacgatctc 1620
ccgccgtacc tccagtcgtg cactctggcg atggcagcga tggcagaagg caacgggtct 1680
gccgagtcca agcagtacgt gatacagtgc atcaacctgc aggtgacgtt tcactgtctg 1740
cgcatggtaa ttacgcagaa attcgaagac ctctcttatt ttgctcctgg cgttgagcag 1800
gctgatctca gaaagtcgga gattgtgcga gacatgctga gggatgatga cgaggcggcc 1860
ttttggggcc tgcaggccaa tggcgagcca aacgtgagtc gtttccttgt ctcttctctt 1920
ttctgcacac ccttttcttc gacgaccccc cctctctctt tatatccctg cgatatatga 1980
tatcatcaag cctcggcact tgttgctaatt ctgtcctgat tatgttgtct ggatgctgca 2040
ggttgaaaag attcgctta tcggagctag ttgtctggcc atcatccatc gcaaccagga 2100
ttcacccttg gctacgcgag ccaggagcga cttttccgtg cttttggata ttctcacgcg 2160
gctggactcg aaggcgtcgg actaa 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 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 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
	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
	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
[0020]	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
	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
	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
	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		
<212>	DNA		
<213>	人工序列		
<220>			
<223>	启动子序列		
<400>	15		
	agcagtggct tatagcaata tcgtgcttgt ctctgccctc gctgaagcta tccctccctc	60	
	gtgtccccct tcgtggagaa tgactgcagt aaaggagacg atacgctctt gggaaggtcc	120	
	tgaagggggt ggctgttggg attgcaacct ctggcatttg tcgaatggcg ctttatgcgt	180	
	cgcgctcgcg agtgatgttg agaggaaagt caacgttgtg acggcttcac aatttgcct	240	
	ttcattaagg cttattgccc gctaaattac tataaggaag gtcggcagct gggattacgc	300	
	atgtgttgcg agacacaatc tcgaacgaga cgctatcaga ggacaagttt tctgacaatc	360	
	tgctgttgta ttgaacgctt ttcgtgtttg taaaccagca ttcattgaggt tcgagggcc	420	
	acgcattcat tgggatctca tcaatgaagc ggaatagtagc aagaagaccg agcatccatt	480	
	acagacctca tgcagctaag gaaacggatc acccttaaag acgtggatgt agtttccgtt	540	
	gctcgcgcac attagaaggc tggctactata ggtgaggga ggcgatgctt gttgtttatc	600	
	cgccatgata agaggagtgt aagccggctg catggacttt ttccataccc tgtcttgcc	660	
	gtccatcttg aaattggcaa gctccacctt ccggttaaaa tcgcaggctt aaggtcttct	720	
	aataaccact cagcatttgc tcaacctatg ctttgcctct ctcacctgac ctgcattctc	780	
	ttttagcctc caggagtgg cgaattgatg ccgattttga cgtcaaatcg cattgagata	840	
	actcacgata ttcgttttac aggctaaaga attgcctcca gcgttccagt tcccgctgta	900	
	ggccgagacg acactgcctc actaccagcc	930	
<210>	16		
<211>	930		



&lt;212&gt; DNA

&lt;213&gt; 人工序列

&lt;220&gt;

&lt;223&gt; 启动子序列

&lt;400&gt; 16

```

accccaatct gaggttcagg ttiggettcc ttcgagttct ccaagttctc tgaatcgtat      60
gtccgcattc attggtatcg aagtttgtga ttaatctcga gaatgtgcat acttcagtc      120
cctcaacata cacggaatcc agcctcttca tgaggaatcc ttactccttc ataccccaag      180
tgccccgggca gataattgta ccatttcaca aatgaactta gactgtactc cgcacttctt      240
tcaggetcct ctctcctcac gatgcccaca tgcctgccac tccacaaggt acatgtagta      300
atctgcagtt actccccgcg ttttgcctac ttagccggca agaattgagat gcaagtttgt      360
tcctgtcggg agtattggct aaatggaaac acaagtagaa gaagagaaac agaaaaggtc      420
caataccggc tattcacaac ggatctgcgt ctgtgtctga tagaactagt aaaagtcggc      480
agtatccgct gtcataca tcatatacta aaacgtaagc taaacgcaat ggcttgga      540
aggggattga gacagaagat acaccaacga ccaccctgt tagtgacaga gatggcagtc      600
attcgactag cggctggcaa ttggtgtccg ccttgatttc aggctaaata tctcgaggtg      660
ccgggaaatc atggtgagga ggagttggca tgtgtggagc cgtgatgcag gtcggaccac      720
accaagaagc tggggtagca tttccgtccc gtatgggggt aacatggggg aacatgctgg      780
aattgcaaat gatgcatggg ggaagaatgc aacatcgtca tcgtcatacc gctcaattta      840
aatattgggc ttttccgggg atcagatgga agaaggcaac agagagagag acaaggaaga      900
ccgtgagcca ttgaaggaca gccggacgca      930

```

[0022]

&lt;210&gt; 17

&lt;211&gt; 930

&lt;212&gt; DNA

&lt;213&gt; 人工序列

&lt;220&gt;

&lt;223&gt; 启动子序列

&lt;400&gt; 17

```

tctatggcca ctctgtgcc ttgtcatctg tatcagcacc aaagccatga ccgtgatgta      60
ttgatgtgcc atcttggcgt tctttcggca acacaccggg cttgaggttt gtcactccct      120
ctccgcactt gagcacaacg ccggtgcct cgtgtcccag aatacagtct ccttcgaaga      180
cgaggagacc gattctccca gacttccaga aatggacatc cgatcttgca ctctgtcagc      240
ctgcagttga atgcgagttt gttgaagacc ttcagcactc acccgcatat gccagtagct      300
ttgatttgca gcaagacttg tccgcgtctt ggtttataca ctggtgcttc gaccagcttc      360
aggttgtggt cggcagttgc ttgtagtgat ggattgggtc gtggagcttg aatgatccta      420
tgctcaagat cccgagctac agaaacgttg ccgctgggca tggttgctga ggccatgggt      480
taggtcgagg caatcgacag aagctcgaaa gtagtgaagt ttggtgtaag tagaatgctg      540
acggtgactt tgcaaatgcc caaatgtcaa gtgtaaagta ccaaactact gtcggtcagg      600
tgcaagaaac gcaatatgct tgggaattat taaatcagaa gggtagacaa taagacgtga      660
tggcctctga ggagctcaaa tgatcgcgct ccaagagggt gctaaaactt cccgattcgg      720
caatcctcac ggcttctcgt atgcggatgc aataacctct attagcctat atctgcggaa      780
gctatagacc gagcatttgc gtttgtcata tcttcgacat catttctctg atccttttct      840
ttttattata tagaagctga agttgtacgg cacaactgtc aacacaccgt tgccgcggaa      900
ttctccccta 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 gctcgtctca cgtccagttt cttcaagtca tatgttccag	60
	gaccgggaag atgtctgcaa attggatagt ttttctttcg tagctgccct cgatcaacct	120
	gggacgcgat acgcctcagt tcaagtactt gatctgaagt accgtatgtt taggctgata	180
	tccgatggga tgcccccaaa tacaaggctc gtcaacctga ttatgagaca attctttcac	240
	ctacgtcaat ctcggtcat tctttgcaag aaacacgac agtccatctc tcaggagcct	300
	gtcgtgtttt tctgctgaga ccattgggct tcccctcgac agtatgttgc cccttggcta	360
	actctectec acgcctacct ggctacctg acctaccagg gaggcggctt tcttacctac	420
	ctatcaaagg gcctccacgt ggctgcgatg ctctaatcaa caaatctctt cctcgcaagc	480
	ttcttcttct tgactgtctt ccaggttctg ctggccttgt tctcggcaa ccagagcgat	540
	cggctcagga agcagcgaga agcagactct caggagaata agcagatcga tactccatcg	600
	gaagtctctc gcgcgtgcag ctgccgattc cgaggccact cgtcttttgc cgaatgagcg	660
	tcacgacggc cctacagata taccgatagc gttgccctgc cacctgatca agttgtcgcc	720
	tactatatgt ctcgaggcat gaccgtatcc aaagctgcag cttggattct cacctctca	780
	gagccgacct tggaatagca ggggacacat ttggcctcct cagcaccgct ctttggcgcg	840
	tctctcttgc gtttgatttg ttcttccact tctctcttcc ctgatatctc ttcttgaact	900
	gccggctgcg ttgtcgaact cgtgctcacg	930
<210>	19	
<211>	930	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	启动子序列	
<400>	19	
	gcgattagca cgcacgctg ctgggaccgt gagctcgacg ggctcaacag catggtggcg	60
	tactgaggga tatcgttcac atcgttgaac gaggcggaca tgagaaacga cttgatattct	120
	gtattattta cttgtacatg tacggcagtc actaatacct gtctcaatca atgaagattg	180
	atgtaaataa gcttaataca actgtctaaa atatgtttac attttaagtc ttgtctcctg	240
	gtttgtgcta ataaatgcgt gaggtgacat cgtccgggaa tagtgagacg aagtggatc	300
	accaagtgga gggactcatt attgttcatg cacgtcccat gtacaaggcc taaatccgcc	360
	agctattccc cgtatgacct gtcaatactc gatacgatac agtagctgta tggacaagca	420
	agtactactg gtatgggtgt aaccccagct gctacaaacc ctgggaagca gactctcaca	480
	acctattact aggtactctc gttcgttttc tcaaagtta gggcatctcg catgagagca	540
	agacgttgcc ggcatcgct cagctactta tgatgccatc tcatcaagcc catgagacca	600
	cctatgcact gtcacaagag aacaaacttc gccttttgga acccgcaaaa acaggggaga	660
	tgcccgttgg tctttgcttg tctttaatcc gcttgatctg tgtccacagg aacgggggaa	720
	aaaacaactc gggttcgtat ttgcggtgtg gcagcttcca acaggggtat ggaaaatcga	780
	ccttcccaca cgacgggtta atatccacaa gccggcgaaa tggtaaatat ataatcgag	840

[0023]

	catcattccc atctgggtgg atttcaaaca aaagaaaagg aagcgcacgc tggaaagggt	900
	gcaaggtcgc gataaagtc tcttctcgcc	930
	<210> 20	
	<211> 930	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 启动子序列	
	<400> 20	
	tgcgaactgg caatgagttt tcaacatca aagtcacggc gcagttgtcg tacatggtgg	60
	tccgagaata aaaaggctac atggtaaatt cgtaaaggg caaggtcaat aggagcacag	120
	tgatgcttcc ttggacagtg ttgcttacat ccgacatcc caccttgaca tacactcgt	180
	ttgagtattc gatcgtgcc acccgattga tacattacaa gcagaactga gccgctgtta	240
	tcgacacact tgcggtcttc tcagtgcagc gcttctacct cgcttgctt gatatcacga	300
	gggcattgat acgtcaacca ctccatatca agtctgattg ctgaatgcct tcaatcattc	360
	aggccacaa ataccctac agcatgctca ctttgaaatg agttttctc tctcttctgg	420
	ccagcaggcc cgtttggcgc gtatgagcgc cgaagctgac actccagcga cgtgctgac	480
	tagattctgg acacgaatag attcctattg aaggagtgga cgtcccccacg atgaactccc	540
	gtcgagacag ttttgcagc atacagccgt ttaacagcca gcggtgtact ttatacaaag	600
	aacaacgtcg ctgtgttagt tgcgccaact tggcgtgatt ggacggacgc cgactcaacc	660
	cctgcacac aatccgcaa gaagcgtcac cgcggataaa aaggctaaat ggatggtctt	720
	caactcttcg cgacatggaa cgcgagcagg tattacagaa tgatgaccgg ttaaatctgg	780
	cagtctccat ggtactgaag tcaaagcaag agggacaatt cacaatggcc ttgatataaa	840
	agcgataaca tgtctccggt caaggcctgg atgagcagag tttagagcaa actgcctgct	900
	tgggaatctt ctctctactt tctctcgac	930
	<210> 21	
	<211> 930	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 启动子序列	
	<400> 21	
	ctggggtaac agagctgttt gcggggatga agatcttgat gaatatggtg aatgaggaat	60
	ttttctatt ggaggaagca ttgtcgaatc caagggtatg cgggctcggg gaagggcac	120
	tctccttagc tgtcaaggct ctagatactg tgatactcga tgatatataa gccaaacca	180
	agcaaagatc cgctggttgt ctgaacaagc cgaatgccag gccctttttg ccggagcccc	240
	tcacgaaat ccgctgtgca agcggttcat gggttcaaca ttcggaatag cttctccgca	300
	agttctgtgg gtagcagcag cactccggca gatgccactg tgcctgtgaa tgtggggaga	360
	ggccagagta tggggaacca tgtgggagc agcacggaga cactcgaagg tcatgccttg	420
	gagttgtgcg tactttgttt ttgcatgcct tggttccatc atcgaaagga gagcagcgg	480
	agatagagcc ttaagcatgg gtaaaactcc ttccgacaac tctataacag acgttaggaa	540
	aacaacaagt tcccagttat tacccagagc ttcttttacc catcggcaca acgtgcaacg	600
	gccaaagaatc gctgccgac tcctcatctc cagccctca aagtctcaa gctgcatcca	660
	gcattctccag gcggacctg cttgccaca ctttaggcta aaggcttctg cccctggtcg	720

[0024]

	gcagtggcag aagcaacctg ctacttgacg acatgaaccc gtttgtaccg gctcggccat	780
	gagatgatgc agtggttcaa tggcatagtt gagtcgaagc ggtgttttca ccccttaatg	840
	atgtaacatg tatataacgg atgaccaatgc cctccttgac 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	
	catccacatc atcttgacga ggaaccatcg ggtggtcggc cggactctgc agatctgggg	60
	tgcgtctgac cgacgaggtg tcaattagcc tacgtgtgag taccatgct gagaattgat	120
	acctatggat gcaagtgcct ctgaactatt actttatcca taggagcatc cgtaccaat	180
	gtgcggttaa aattgccttt cagccaccct gtccctcggtt aattgagtga gacgtgcata	240
	acatggccgg cagcatggca tggtagatga tattgggtga acgtgtcaga agaaaaggct	300
	agaatattcg agacagcttg ctgatatgtg caaaacttct caagatattt gatatgtgta	360
	gagttactct tggcattata ctgtaatgtg aatgtagagt gtacgtaca gtacctaat	420
	ccaagaactt ataacatacc tacataccta ctgagctaca ttttaggcag tgctccgct	480
	gagaagcttt gaattctact tttgtcgttt taactcgtga cgcatcgact ggcggtcat	540
	cctgatacag aatcaggacc attgcatatg aaaaacagtc tgagaccaa catccaatcg	600
[0025]	ggagacaatg ctgcatgaa agctgattta gcctatttcg acttctcaca actcaaagag	660
	atgctttatg ttcgggggaa gggaatatga ctgtgactag aatgcatcgg ggtcctgcc	720
	aagaatgagt tgactatgag gaggcaaata tctgagtcac gtgagtagac cagtaaata	780
	cagctggggg taatcactta tgtgttcacc ggatatactt catattgata taaaatgcta	840
	tgatctccaa gtcccaacga tactgagatg aacagcatct cttaacagtt gctttccaca	900
	taaataattc ctctgttac aagagccaaa	930
	<210> 23	
	<211> 1059	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 启动子序列	
	<400> 23	
	ggcaggcact ggctcggacg acatgttttg tatattggtt gggactgcgg ccgcagcggg	60
	ggcgggagct ggtggcggcg atatgaattt ccgggcgttg ctacaacagg taccatttg	120
	accacccatg gctgccgtcg cctgtcttgg agctttcagg tcgcttcagg gcgttgccga	180
	ggcaagttag acggtgggga aatgacgaaa aatggtgcat cgcctttgta ggtgtgtgtg	240
	agtagtagtt ctactatgag gtacgtatgt agcagaagga tcgagctaga atctgccggc	300
	attgcaaagg ttatctggaa agaggaaaag ggctgaacc ggcatatgga tgcattcttc	360
	gtacgaacta ctatctgata acagtttagt actgttatcc atacaaagag tcttatagaa	420
	acactgcacg gtaataaaat actcggtagc tgcctgaata tagtaataag atcaacatcc	480
	ttcacctct agtctccgtg gattccagta aaagcgtca attctgactt ccgactctgt	540
	tgatgccccg tgtctgcccc tcgggggtgtg ctagacgctg cctcaacgcc catgtaccgg	600

	cctgatgggg cccttggggg caccacaagt ccactaaacg aagcactggg gacgggactc	660
	gatagccctg agcagcagcc ggtctcagca gcccaaccagc ccagctggaa gcatcggtta	720
	ggggaggggg gcccaactac tacgtgtact actaggtaca taatgaattg gatgggaccc	780
	agccagccca acctaacctt ccagccttta tagctgcagc ctgcttcccc gtgcctcacg	840
	ctttttgctc ctctgtctggc cggactcgga cctcttgca cctctgctcg accaacaatc	900
	cctcttggtg caccctctcg cttttgtac ctgcagctc aattcctcgc tgcgcctca	960
	cctaaccgcg tgtgtctgac tgccctcacg ctcggtctgc ctctgctcc gcgagcctcc	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	
	gacggaatct tcacacctag tactcgtaca tactagatca gatgctgggc gagtctgggg	60
	aatgctggca ctggatggta taaacacccg actgcttgac gccggatcag cggcgatgca	120
	ccgagctgga cgtgtcaata ctacatgtat catgacgcag agctgtttct cctctggcac	180
	atgaacccaa agacacaagc gtcacctcgt tctgagcag atcaggtaca gacttccata	240
	tctactgggt cccatctctc gagcctaaac cggcctccca ctttgcaaa gatgccccaa	300
	gccgttgccg agagacgacg tggactcgac ctctcagcc atcgatgccg tctcagctgc	360
	cgcttcggcc aaccaacttg ctgcgaggca aaagtatttg ctgcattgat gccagccaac	420
[0026]	ctctcctcc ttgcctcctc cattcagccg gcagcaatcc aaccacccac ccaccggcgc	480
	agcggcacgc aaggcacagt caggactaac actcttaggc tgcctggata catgtagaac	540
	ggtcttttgg tggttgggtg tctgcacat aaggtatata ctgtacatgc atatagtatg	600
	catacagtag acgctgcttg gcggccctgt atgtagggga tacatcatcg tccatctgct	660
	gtccccccgc aatccgctat ccggcatacc gagagctcag gagccggtcg ttgcccgctc	720
	ttgtacagat tacacctcca agtctccccc ggctcccat cacagctgcc tccgcacggt	780
	gaccatcccc caagcaagtc gctccttgcc acttggtgga gggctgccag tggacgagag	840
	ctgctgccaa tgcccatgat cagcaccggc ctgcgctggc tgactggata cttgcaagta	900
	ttaaacgggt cgacctgcc ggttggtctc gtcgtcaact cttgccctct gactctgact	960
	ctgactcttt ccgtatcccc gtgcacatcg gttagaagct gctcctcttg tcttttctc	1020
	tgccaaaccc tctcgtact aggtatctca acctttgtat acgagatcag tctcggggc	1080
	acgcagggtc tcatcctgac tcgtcctgtg tcggaatcgt ctcacgtctc actgccaaca	1140
	agcagtttgc gatacgcaaa tcaatcgctt gcatcaggtg cattccagat ccggcctccg	1200
	gacgtcggtt tataatcgcc agcgccttc tcccaagag ccagccttt tggagcagga	1260
	cccaggttt gtccatattt tgggtcgtac gcctgccttt tgccatgcc atctgcttc	1320
	tccacccaa ccacgcccc tcatcaatct ccctctccc cgccaaaggc gactcgatcc	1380
	ccgctgac acacaaacca ccccggggaa aacgagacca acgtatata ctgtggcagc	1440
	gcccgtcaa tccccctccc agcgtgcgcc ttgaccgcgg agtttggtat ccggtgagtg	1500
	cagtcctcgt cttccgacca tccctccctc tctgcttgtt agtggttgt cccaccgcc	1560
	aatctgcgcc ggccagcgga gccagcctg agcccgccct cgagctcggg gccgttctgc	1620
	gacctatac tttccgagg ccacaaaggcc gacgtcaaaa ctgctcgggt atcctgccgg	1680
	ctctggctgg cctgccgttg ggcttggtt tgaagaggt tgccgatgac gggatcgag	1740

	tccccaagtt caaccctctc tctgccccac caatcatgac ccccatgccc gtcgttatgc	1800
	cactcatgtc atcgcagtgt gaacattgac ggagcttgac taatgtgaca gagtcattea	1860
	ggttggttga gagccaagca gggtcctcag ctacgagagg cgcacccgag gcgcagagtc	1920
	ccgaccatag cacttataag gcggcgcttc 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 ggcattcgat cgacaataact aaaatcagca	60
	aagcctgagg tccgttgggt gaaaacttca ttcgtactcg tactcagacc gcaaattctg	120
	gtatggtttg atgggttgggt tcatagtgtat catccatata atgggaagcg ccgcctgtctc	180
	cgagaagcaa caagtagcca acagcatcgg cactatcgtg catcatgccc caactctctc	240
	ctagccgtag cggacctccc tcaaccagac agtgcceaag atcctttgac gcaatcatcag	300
	cgagtctatg tacactgacc aacatgcagc cccggcgact tcccagatgg gccaaattca	360
	gtcgagacgg cgttggtttg cagttaattg ccgccgcttg atatgcgctt aatggaaggc	420
	gttccgctgg aacttccctg gcgccacgtt gttggatctc atgattgtgg tatttttccg	480
	tgctcatca gttctcgagg aacttctgta caccgaatgc catgtgagge ccccgctcag	540
	ccccagctgc agcaactcgt accccatata ccaggcaaat tatcgacccc gattgcttgc	600
[0027]	aatcaaaatc gcttgccggc gtcttgatat gcccgctagc aaggaaaagg caggatgtctc	660
	gttatgtgtt gctcaatata gtctcgacg ggggaatgtc ccgcttctgt ccaatacct	720
	gtggcgacaa gggctatcaa aaagttctc atccgcgatg tgccaacaac cgtggcagta	780
	gacaaagggt ccttccaaaa gagcgataa ctggatcgcc ccccatgcat gcttgagtct	840
	atcaaacggc cagctggagc ttctgttctg ttctggagtt cgtgatgggt ggtgcttggg	900
	tgtaccttgc ccgttgcgtg gccggtgctc tcttgtgtga cattgcattt aaatatcatc	960
	gtgcctgtcg ccagttctta ccgacttga gattcttctc agctattcct ctgcaagcgt	1020
	tcaactgccg ccggtgagga acactcggtc agtgccaaca caacacc	1067
	<210> 26	
	<211> 1124	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 启动子序列	
	<400> 26	
	cgaggcagat accgagatgt ttagcacagg aagtcattct gtctagctgc ccgtggggaa	60
	aaaaaaaaa gtaagccgga accgcaggtc caggcctgcg gctgtacaaa ggagcaaaca	120
	cgatacaaac acaataccac cctatggttag catcaataat ccgtaatcgt aatagcatgg	180
	gcgaatattt gccctctcgg tacggcgctc cgtaataaaa ataacacccc cccaaatctc	240
	ggtgccccca attcgggtct ttttctccac cagttcgga cctctcctcc tatgtactct	300
	ctctctctct ctcgttgggt gggggagagc gagcgagcga gagagagagg cagtgaaggg	360
	ctcaggctcg gtctttgctg cgccaaccct cgcccatgtc atctccccct gattagcgcc	420
	ctcacctcag tttctcgccc cctcgtgacc gtttccgcct ccacaatctg gggagcctgg	480

	aattagcaac tgtggaactg aggggtagat cccacgcgtc ccgttagtaa gtgcttaagc	540
	gccgtgcgaa aggggtgcgag caccggagtc gatgtcgaaa tactgatgtc ctccctggaa	600
	cccaattact ccggggggcg actaaagcag ccagccagc caaactagca gcaaggcaca	660
	cgctcacact agtaatgcc ggccggcgga tgatgttgct tgtctgttg tttcttgtgt	720
	cagggccagg aagccaaatc ttgtctgggg gggcgacata aaggccgct cgtcggccta	780
	ccttgagcca cgcaaccaca caccgtccg gtttctcaa ttcttatac caccaaacga	840
	accataacaa caacgtctc ctctctctc atctctctc atcattgtat catcgtcata	900
	gcattgcgtt gccagctctg ttgtgtctgt cagtcatat aacaaagcct tccgtgtctc	960
	tttctctc tcacactata ttttttttt ttacctctc atcacaaaga cccacacctt	1020
	ttccgccaca actgtccta ccaaaggcac actcgcacct ttgaatgac tcctccttgt	1080
	ccaaagact 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 ccggtcccct tatcccagct gttccacgtt ggcctgcccc tcagttagcg	60
	ctcaactcaa tgcccctcac tggcgaggcg agggcaagga tggaggggca gcacgcctg	120
	agttggagca aagcgccgc catgggagca gcgaaccaac ggagggatgc cgtgctttgt	180
[0028]	cgtggctgct gtggccaatc cgggcccttg gttggctcac agagcgttg tgtgagacca	240
	tgagctatta ttgctaggta cagtatagag agaggagaga gagagagaga gagagagggg	300
	aaaaaaggtg aggttgaagt gaaaaaaaaa aaaaaaaaaa aaatccaacc actgacggct	360
	gccggctctg ccacccccct cctccaccc cagacaacct gcacactcag cgcgcagcat	420
	cacctaattc ttgctgcct tcccgcagct caggttggtt tttttttctc tctccctcgt	480
	cgaagccgcc cttgttccct tatttatttc cctctccatc cttgtctgcc ttgtgtccat	540
	ctgccccttt gtctgcatt cttttgcacg catgcctta tcgtcgtctc ttttttact	600
	cacgggagct tgacgaagac ctgactcgtg agcctaacct getgatttct cccccccct	660
	cccgaccggc ttgacttttg tttctctcc agtaccttat cgcgaagccg gaagaacct	720
	ctttaacccc atcaacaag ttgtacaaa aaagcaggct ccgcggccgc ccccttcacc	780
	<210> 28	
	<211> 58	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 28	
	tcagggttat tgtctcatgg ccatttaggc ctggcaggca ctggctcgga cgacatgt	58
	<210> 29	
	<211> 58	
	<212> DNA	
	<213> 人工序列	
	<220>	

[0029]	<223> 引物	
	<400> 29	
	agagccctgg gccggagctg ctgagcccat tgttgaattc tggcggggta gctgttga	58
	<210> 30	
	<211> 58	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 30	
	tcaacagcta ccccgccaga 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 taggttatgc gagcaacatt gcacgaaac	59
	<210> 32	
	<211> 59	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 32	
	gtttcgtgca atgttgctcg cataacctac gctagctagt tacgcttggt tattttacga	59
	<210> 33	
	<211> 58	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 33	
	acatgtcgtc cgagccagtg cctgccagge ctaaattgcc atgagacaat aacctga	58
	<210> 34	
	<211> 57	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 34	
	aggtgtaaga cgggggagta gcgcagcatt gttgaattct ggcggggtag ctgttga	57
	<210> 35	



	<211>	57	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	引物	
	<400>	35	
		tcaacagcta ccccgccaga attcaacaat gctgcgctac tccccgtct tacacct	57
	<210>	36	
	<211>	59	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	引物	
	<400>	36	
		tcagggttat tgtctcatgg ccatttaggc ctagactagc ggccgggtccc ctatccca	59
	<210>	37	
	<211>	55	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	引物	
	<400>	37	
[0030]		agagccctgg gccggagctg ctgagcccat ggtgaagggg gcggccgcgg agcct	55
	<210>	38	
	<211>	55	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	引物	
	<400>	38	
		aggetcecg gccgccccct 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 cagcatgggtg aagggggcgg ccgcggagcc t	51
	<210> 41	
	<211> 51	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 41	
	aggctccgcg gccgccccct tcaccatgct gcgtactcc cccgtcttac a	51
	<210> 42	
	<211> 7882	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 质粒 pYL1	
	<400> 42	
[0031]	ggcaggcact ggctcggacg acatgttttg tatattggtt gggactgcgg ccgcagcggg	60
	ggcgggagct ggtggcggcg atatgaattt ccgggcgttg ctacaacagg taccactttg	120
	accacccatg gctgccgtcg ccttgcttgg agctttcagg tcgcttcgg gcgttggcga	180
	ggcaagtttg acggtgggga aatgacgaaa aatggtgcat cgcctttgta ggtgtgtgtg	240
	agtagtagtt ctactatgag gtacgtatgt agcagaagga tcgagctaga atctgccggc	300
	attgcaaagg ttatctggaa agaggaaaag ggctgaacc ggcatatgga tgcattcttc	360
	gtacgaacta ctatctgata acagttaggt actgttatcc atacaaagag tcttatagaa	420
	acactgcata gtaataaaat actcggtagc tgcttgaata tagtaataag atcaacatcc	480
	tttcacctct agtctccgtg gattccagta aaagcgtca attctgactt ccgactctgt	540
	tgatgccccg tgtctgccc tgggggtggt ctagacgtg cctcaacgcc catgtaccgg	600
	cctgatgggg cccttggggg caccacaagt ccactaaacg aagcactggg gacgggactc	660
	gatagccctg agcagcagcc ggtctcagca gccaaaccag ccagctggaa gcatcggcta	720
	ggggaggggg gcccaactac tacgtgtact actaggtaca taatgaattg gatgggaccc	780
	agccagccca acctaaactt ccagccttta tagctgcagc ctgcttcccc gtgcctcacg	840
	ctttttgctc ctctgtggc cggactcgga cctcttgca cctctgctcg accaacaatc	900
	cctcttggtg caccctctcg cttttgctac ctgcagctc aattcctcgc tgccgcctca	960
	cctaaccgcg tgtgcttgac tgccctcacg ctcggtctgc ctctgctcc gcgagcctcc	1020
	ttttacactt ttcaacagct accccgccag aattcaacaa tgggctcagc agctccggcc	1080
	cagggtcttg tagctgcagc tgcaggcggc cctccagctg ctggcgctgg cgctggcgct	1140
	gtccacgccc tcaccacctc gcccgagtct gcctcggcct cgcagcccgg ctgcgaacc	1200
	gcctcaacca cgccgcgcga gaactcactc gtgtcggctg caacctcggt ccaccacat	1260
	cccagaggcc gtctggtgag cagagcctgc gaccgctgcc gccggcgcaa ggccaaggtc	1320
	agtctagccc ctttgtgtgt gcttgcattc ctgttgtcat tgctcctcct cctgtctgtg	1380
	ctgatgtctg tgctcctcct cctcctcctc ctccccgtct cctggtcctt ggteccctgt	1440
	cttcatatgt ccttactgcc cgtgtctcct ctccccgttc ccgttcccc tcttccgtc	1500
	ctcttctcct gcgtgtctgt catgcgtaca aagcatacat acaatacatc agcatacatg	1560

[0032]	gcaagcggttg	tgtttgtgttg	agagtttgtgt	gtattgtatt	gcactgcctt	cacaactcgt	1620
	tcatactgct	gcagcctcac	cccaacaccg	acctcgtctt	ccatgctgcg	ctactccccc	1680
	gtcttacacc	tggatactct	ctccttgcca	ccactgacca	atgctcttcc	ccgccc aaag	1740
	tgcgagtacc	tcagcgtgt	cgatagctgc	acgcaactgcc	gcgatgccc	cgtgcagtgc	1800
	actttcgacc	tgcccttggc	gcgacgcggc	cccaaagcga	ggaagaagag	cgaccagccc	1860
	ggccagccgc	ctcctgatec	gagctcgtc	tccaccgcgg	ctcgaccgcg	ccagatgccg	1920
	ccgccgctga	ccttctccgg	ccccgcagta	gccgcgctgc	agcccttcgc	ctcgtcgtcg	1980
	ctgtcgcccg	acgcggcctg	ggagcccgtc	gagccgctca	gcattgacaa	cggcctgccc	2040
	cggcagccgc	tgggcgacct	gcccggcctc	tccaccatcc	agaacatctc	gacgcgccag	2100
	cgatggatac	accttgccaa	cgccatgacg	ctgcgcaaca	cgacgctaga	gcgcgtctcg	2160
	aagcgatgta	tcgacctctt	cttcgactac	ctctaccccc	tcacccccct	ggtgtacgag	2220
	ccggccctcc	gggacgtgct	cgcatacatc	ttctcccagc	ccttgccctgg	cgtcaaccaa	2280
	ccatcgccgc	tgtcacagct	cagccagac	ccgaccaccg	gcaccacccc	cctcaacgct	2340
	gccgagtcgt	gggcccggctt	tggccagccc	agcggctcgc	gaaccgtcgg	cagcaggtcg	2400
	gctccctggg	ccgactcgac	cttcaccctg	gtcacggccg	tctgcgcaga	ggcagcattc	2460
	atgctaccca	aggacatctt	ccccgaagga	gaatccgtct	ctgagatctt	gctcgaagcc	2520
	tctcgggact	gcctgcacca	gcacctcgag	gccgacctgg	agaatccgac	ggccaactcg	2580
	attgccattc	gctacttcca	ctccaactgc	ctccacgtcg	cggggaagcc	caagtactcg	2640
	tggcacatat	ttggcgaggc	catccgcctg	gcgcaggtca	tgcagctgca	cgaggaggct	2700
	gccctcgagg	ggctcgtccc	catcgaggca	gagttccgcc	gtcgtgctt	ttggatcctg	2760
	tacttgggcg	acaagtcagc	cgctatactc	aacaatcggc	ccatcaccat	ccacaagtac	2820
	tgcttcgacg	ccggcatcac	cacgtatac	ccgtcgggta	tcgaggacga	gttcttgagc	2880
	acggcgctcg	agccgccccg	gaagagcttc	atatccggct	tcaacgcaaa	tgtgcggctc	2940
	tggcagtcctg	cggctgattt	gctgctggaa	atccgcgtgc	tgcaagatca	gatgatgcag	3000
	cactttcgag	ggaccatgcc	cccgaaccat	gtgctgcctt	ccgccgacag	gcagcatctc	3060
	gattctctct	atgtccgctt	catcacctgc	ttggacgac	tcccgccgta	cctccagtcg	3120
	tgcactctgg	cgatggcagc	gatggcagaa	ggcaacgggt	ctgccgagtc	caagcagtac	3180
	gtgatacagt	gcatcaacct	gcaggtgacg	tttactgtc	tgcgcatggt	aattacgcag	3240
	aaattcgaag	acctctctta	ttttgtcctt	ggcgttgagc	aggctgatct	cagaaagtgc	3300
	gagattgtgc	gagacatgct	gagggtgatg	aacgaggcgc	ccttttgggg	cctgcaggcc	3360
	aatggcgagc	caaacgtgag	tcgtttcctt	gtctcttctc	ttttctgcac	accttttct	3420
	tcgacgaccc	cccctctctc	tttataatcc	tgcggataatg	tatatcatca	agcctcggca	3480
	cttgttgcta	atctgtcctg	attatgttgt	ctggatgctg	caggttgaaa	agattcgctt	3540
	tatcgagct	agtttcttgg	ccatcatcca	tcgcaaccag	gattcaccct	tggctacgcg	3600
	agccaggagc	gacttttccg	tgtttttgga	tattctcacg	cggctggact	cgaaggcgtc	3660
	ggaccaactg	aggaatacgt	ccactaccgt	tgttggttaa	atgtgtgttg	gaacaacaaa	3720
	aaatgtcaaa	gtcgtgttaa	atatggccag	gatctttgtg	ttattccccc	ttcagcgttg	3780
	ctgggtatct	cccctttgtt	tactcttttc	tgttttttcc	agcacttggt	tttccagcag	3840
	tggggggaac	aaaaggcgtt	tctttccctt	atgccagggg	ttgtccgatt	tagcatttga	3900
	gtgtacatct	tccctacatt	actaggtact	taatgagctt	atggagatct	cccgtcatte	3960
	cggatattca	tcacgttggt	gtatataatcc	gtggttggtc	ttgaaacctg	gagttgggtt	4020
	gcaatgcagt	gacgcctttt	gcgaaggacc	aaaataagcg	aaggatgaag	tctgaatagg	4080
	atacgaactg	gctacctatg	ggtgagcatg	aaatgaagcg	gtcggggaaa	tggcggagaa	4140
	acgtcgcagc	taacgtgtgt	ggttttctcc	gtttcgtgca	atgttgctcg	cataacctac	4200

	gctagctagt tacgcttggt tatttacgac aagatctaga agattcgaga tagaataata	4260
	ataataacaa caatttgceet ctcttttcca ctttttcagt cttactctcc cttctgacat	4320
	tgaacgcctc aatcagtcag tcgccttgta ctggcacagg taatcctccg tggtcttgat	4380
	atcctcaggg gtagcaaagc ccttcatgcc atcgataatg tcattccagag tgaggatggc	4440
	aaagatgggg atgccgtact ccttcctcag ctgccaatg gcactcggtc caggcttgga	4500
	gtcgtcgcca tccgcagcgg ggagcttctc catgcgggcc agggccacga cgatgccggc	4560
	gacgatgccg ccctccttgg tgatcttctc aatggcgctc ctcttggcgg tgccggcggg	4620
	gatgacgtcg tcgacaatca ggacctctt gcccttgagc gaagcgcga cgatgttgcc	4680
	gccctcgccg tggctccttg cctccttgcg gtcaaacgag taggagacgc ggtccagggt	4740
	ctggggcgcc agctcgccga gcttgatggt gatggcggag cacagcggga tgcccttgta	4800
	ggccgggccc aagacgatgt cgaactctag gccggccttc tcctgggcct cgatgatggt	4860
	ctttgcaaag gcggaggcga tggcgccggc gaggcgcgcc gtgtggaatt cgcccgcgtt	4920
	gaagaagtag ggggatatcc gcttggactt gagctcgaag ctgccaaact tgaggacgcc	4980
	gccgtcgatg gcggatttga ggaagtcctg cttgtaggca ggcagctggg aggtggtagc	5040
	cattctgttg gatttgata gtgtccttat tctctgattt gaacagtaga tcaggacgag	5100
	tgagagggat gcagaggttg gattggagtg gttgagctat aaaatttaga ggcgcgccgt	5160
	atcgagtttt cacatggaag tcaaagcgta cagtgcgagc ttgtacgttg gtcttagtat	5220
	cccacaagct tctgtctagg tatgatgatg gtataagtc acccaaggca gaactcatct	5280
	tgaagattgt ctagagtgat tttaccgtg atgaaatgac tggactccct cctcctctc	5340
	ttatacga aaattgcctga ctctgcaaag gttgtttgtc ttggaagatg atgtgcccc	5400
[0033]	ccatcgtctt tatctcatac cccgccatct ttctagattc tcattctcaa caagaggggc	5460
	aatccatgat ctgcgatcca gatgtgcttc tggcctcata ctctgccttc aggttgatgt	5520
	tcacttaatt ggtgacgaat tcagctgatt tgctgcagta tgctttgtgt tggttcttctc	5580
	caggcttgtg ccagccatga gcgcttttag agcatgttgt cacctataaa ctcgagtaac	5640
	ggccacatat tgttcactac ttgaatcaca tacctaattt tgatagaatt gacatgttta	5700
	aagagctgag gtagctttaa tgccctgaa gtattgtgac acagcttctc acagagtgag	5760
	aatgaaaagt tggactcccc ctaatgaagt aaaagtttcg tctctgaacg gtgaagagca	5820
	tagatccggc atcaactacc ttgctagact acgacgtcaa ttctgcggcc ttttgacctt	5880
	tatatatgtc cattaatgca atagattctt tttttttttt tttttttttt tttttttttt	5940
	tttttttttg cccaatttcg cagatcaaag tggacgttat agcatcataa ctaagctcag	6000
	ttgtcagagg aagccgtcta ctaccttagc ccattccatcc agctccatac cttgatactt	6060
	tagacgtgaa gcaattcaca ctgtacgtct cgcagctctc cttcccgtc ttgcttcccc	6120
	actggggctc atgggtgcgtg tatcgtcccc tccttaatta aggccattta ggccgttgct	6180
	ggcgtttttc cataggtctc gccccctga cgagcatcac aaaaatcgac gctcaagtca	6240
	gaggtggcga aacccgacag gactataaag ataccaggcg tttccccctg gaagctccct	6300
	cgtgcgtctc cctgttccga ccctgccgtt taccggatac ctgtccgctt ttctcccttc	6360
	gggaagcgtg gcgctttctc atagctcacg ctgtaggtat ctcagttcgg tgtaggtcgt	6420
	tcgtccaag ctgggtgtg tgacgaacc cccggttcag cccgaccgt gcgcttctc	6480
	cggtaactat cgtcttgagt ccaaccgggt aagacacgac ttatcgccac tggcagcagc	6540
	cactggtaac aggattagca gagcgaggta ttaggcgggt gctacagagt tcttgaagtg	6600
	gtggcctaac tacggctaca ctagaaggac agtatattggt atctgcgctc tgctgaagcc	6660
	agttaccttc ggaaaaagag ttggtagctc ttgatccggc aaacaaacca ccgctggtag	6720
	cgggtggtttt tttgtttgca agcagcagat tacgcgcaga aaaaaaggat ctcaagaaga	6780

[0034]

tcctttgatac	ttttctacgg	ggtctgacgc	tcagtggaaac	gaaaactcac	gttaaggcct	6840
gcagggccga	ttttggtcat	gagattatca	aaaaggatct	tcacctagat	ccttttaaat	6900
taaaaaatgaa	gttttaaatc	aatctaaagt	atataatgagt	aaacttggtc	tgacagttac	6960
caatgcttaa	tcagtgaggc	acctatctca	gcgatctgtc	tatttcgttc	atccatagtt	7020
gcctgactcc	ccgtcgtgta	gataactacg	atacgggagg	gcttaccatc	tggccccagt	7080
gctgcaatga	taccgcgaga	cccacgctca	ccggtccag	atttatcagc	aataaaccag	7140
ccagccggaa	gggccgagcg	cagaagtggg	cctgcaactt	tatccgcctc	catccagtct	7200
attaattggt	gccgggaagc	tagagtaagt	agttcgccag	ttaatagttt	gcgcaacgtt	7260
gttgccattg	ctacaggcat	cgtaggtgtca	cgctcgtcgt	ttggtatggc	ttcattcagc	7320
tccggttccc	aacgatcaag	gcgagttaca	tgatcccca	tgttggtgaa	aaaagcgggt	7380
agtccttctg	gtcctccgat	cgtaggtcaga	agtaagtggg	ccgcagtgtt	atcactcatg	7440
gttatggcag	cactgcataa	ttctcttact	gtcatgccat	ccgtaagatg	cttttctgtg	7500
actggtgagt	actcaaccaa	gtcattctga	gaatagtgtg	tgccggcgacc	gagttgtctt	7560
tgcccgccgt	caatacggga	taataccgcg	ccacatagca	gaactttaaa	agtgtctatc	7620
attggaaaac	gttcttcggg	gcgaaaactc	tcaaggatct	taccgctgtt	gagatccagt	7680
tcgatgtaac	ccactcgtgc	acccaactga	tcctcagcat	cttttacttt	caccagcgtt	7740
tctgggtgag	caaaaacagg	aaggcaaaat	gccgcaaaaa	agggaataag	ggcgacacgg	7800
aaatgttgaa	tactcatact	cttctttttt	caatattatt	gaagcattta	tcagggttat	7860
tgtctcatgg	ccatttaggc	ct				7882
<210>	43					
<211>	7279					
<212>	DNA					
<213>	人工序列					
<220>						
<223>	质粒 pYL2					
<400>	43					
ggcaggcaact	ggctcggacg	acatgttttg	tatatgtgtt	gggactgcgg	ccgcagcggg	60
ggcgggagct	ggtggcgggc	atatgaattt	ccgggcgttg	ctacaacagg	taccactttg	120
accacccatg	gctgccgtcg	ccctgcttgg	agctttcagg	tcgcttcagg	gcgttggcga	180
ggcaagtgtg	acggtgggga	aatgacgaaa	aatggtgcat	cgccittgta	ggtgtgtgtg	240
agtagtagtt	ctactatgag	gtacgtatgt	agcagaagga	tcgagctaga	atctgccggc	300
attgcaaagg	ttatctggaa	agaggaaaag	ggcctgaacc	ggcatatgga	tgcattcttc	360
gtacgaacta	ctatctgata	acagttaggt	actgttatcc	atacaaagag	tcttatagaa	420
acactgcatac	gtaataaaaat	actcggtagc	tgcttgaata	tagtaataag	atcaacatcc	480
tttcaacctct	agtctccgtg	gattccagta	aaagcgtca	attctgactt	ccgactctgt	540
tgatgccccg	tgtctgcccc	tcgggggtgg	ctagacgctg	cctcaacgcc	catgtaccgg	600
cctgatgggg	cccttggggg	caccacaagt	ccactaaacg	aagcactggg	gacgggactc	660
gatagccctg	agcagcagcc	ggtctcagca	gccaaccagc	ccagctggaa	gcatcggtca	720
ggggaggggg	gcccactac	tacgtgtact	actaggtaca	taatgaattg	gatgggaccc	780
agccagccca	acctaacttt	ccagccttta	tagctgcagc	ctgcttcccc	gtgcctcacg	840
ctttttgctc	ctctgctggc	cggactcgga	cctcttgcca	cctctgctcg	accaacaatc	900
cctcttggtg	cacctctctg	cttttgctac	ctcgacgctc	aattcctcgc	tgcgcctca	960
cctaaccgcg	tgtgcttgac	tgccctcacg	ctcggtcgc	ctcctgctcc	gcgagcctcc	1020
ttttacactt	ttcaacagct	accccgccag	aattcaacaa	tgctgcgcta	ctccccgctc	1080

	ttacacctgg atactctctc cttgccacca ctgaccaatg ctcttccccg cccaaagtgc	1140
	gagtacctca gcgtgtcga tagctgcacg cactgccgcg atgcccacgt gcagtgcact	1200
	ttcgacctgc ccttggcgcg acgcggcccc aaagcgagga agaagagcga ccagccccgc	1260
	cagccgcctc ctgatccgag ctgcgtctcc accgcggctc gacccggcca gatgccccg	1320
	ccgtgacct tctccggccc cgcagtagcc gcgtgcagc ccttcgcctc gtcgtcgtg	1380
	tcgcccagc cggcctggga gcccgtcag ccgtcagca ttgacaacgg cctgccccgg	1440
	cagccgctgg gcgacctgcc cggcctctcc accatccaga acatctcgac gcgccagcga	1500
	tggatacacc tggccaacgc catgacgtg cgcaacacga cgctagagcg cgtctcgaag	1560
	cgaigtatcg acctcttctt cgactacctc taccctctca ccccttgggt gtacgagccg	1620
	gccctccggg acgtgtctgc atacatctc tcccagccct tgccctggcg caaccaacca	1680
	tcgccgctgt cacagctcac gccagaccg accaccggca ccacccccct caacgtgcc	1740
	gagtctggg ccggttttg ccagcccagc ggctcgcgaa ccgtcggcag caggctggct	1800
	ccctgggccc actcgacctt caccctggtc acggccgtct gcgcagagcg agcattcatg	1860
	ctaccaagg acattttccc cgaaggagaa tccgtctctg agatcttct cgaagcctct	1920
	cgggactgcc tgcaccagca cctcgaggcc gacctggaga atccgacggc caactcgatt	1980
	gccattcgt acttccact caactgcctc cacgtgcgg ggaagcccaa gtactcgtgg	2040
	cacatatttg gcgaggccat ccgctggcg caggctatgc agctgcacga ggaggctgcc	2100
	ctcgaggggc tcgtcccat cgaggcagag ttccgccgtc gctgttttg gatcctgtac	2160
	ttgggcgaca agtcaccgc tatactcaac aatcggccca tcaccatcca caagtactgc	2220
	ttcgacgcg gcataccac gctataccg tcgggtatcg aggacgagtt cctgagcacg	2280
	gcgtccgagc cgccccgaa gagcttcata tccggttca acgcaaatgt gcggctctgg	2340
[0035]	cagtcgcgg ctgatttct gctggaaatc cgcgtgtgc aagatcagat gatgcagcac	2400
	tttcgaggga ccattcccc gaacctgtg ctgcccctcg ccgacaggca gcattctgat	2460
	tctctctatg tccgttctat caccctgtg gacgatctc cgccgtacct ccagtcgtgc	2520
	actctggcga tggcagcgat ggcagaagcg aacgggtctg ccgagtccaa gcagtacgtg	2580
	atacagtga tcaacctga ggtgacgtt cactgtctgc gcatggtaat tacgcagaaa	2640
	ttcgaagacc tctcttattt tgctcctggc gttgagcagg ctgatctcag aaagtcggag	2700
	attgtgcgag acatgctgag ggtgatgaac gaggcgcctt tttggggcct gcaggccaat	2760
	ggcgagccaa acgtgagtcg ttctcttctc tctctcttt tctgcacacc cttttcttcg	2820
	acgaccccc ctctctcttt atatccctgc ggatatgtat atcatcaagc ctcggcactt	2880
	gttgctaate tgtcctgatt atgttgtctg gatgtgcag gttgaaaaga ttgccttat	2940
	cggagctagt ttgctggcca tcatccatcg caaccaggat tcacccttgg ctacgcgagc	3000
	caggagcgac ttttccgtgc ttttgatat tctcacgcgg ctggactcga aggcgtcgga	3060
	ccaactgagg aatacgtcca ctaccgttgt tggctaaatg tgtgttgga caacaaaaaa	3120
	tgtcaaagtc ggtgtaaata tggccaggat ctttgtgtta ttcccccttc agcgttctg	3180
	ggtatttccc ctttgtttac tctttctgt tttttccagc acttgttttt ccagcagtgg	3240
	ggggaacaaa aggcgtttct tccccctatg ccaggggttg tccgatttag catttgagtg	3300
	tacatcttcc ctacattact aggtacttaa tgagcttatg gagatctccc gtcattccgg	3360
	atattcatca cgttggtgta tatatccgtg gttggctttg aaacctggag ttgggttgca	3420
	atgcagtgc gccttttgcg aaggaccaaa ataagcgaa gatgaagtct gaataggata	3480
	cgaactggt acctatgggt gagcatgaaa tgaagcggtc ggggaaatgg cggagaaacg	3540
	ctcgacgtaa cgctgttggt tttctccgtt tcgtgcaatg ttgctcgcat aacctacgt	3600
	agctagttag gcttgtttat ttacgacaag atctagaaga ttcgagatag aataataata	3660
	ataacaacaa tttgcctctt ctttccacct tttcagttct actctccctt ctgacattga	3720

	acgcctcaat cagtcagtcg ccttgtactt ggacaggtaa tcctccgtgt tcttgatata	3780
	ctcaggggta gcaaagccct tcatgccatc gataatgtca tccagagtga ggatggcaaa	3840
	gatggggatg ccgctactct tcttcagctc gccaatggca ctcgggccag gcttggagtc	3900
	gtcgccatcc gcagcgggga gcttctccat gcggtccagg gccacgacga tgccggcgac	3960
	gatgccgccc tccttggatg tcttctcaat ggcgccctc ttggcgggtgc cggcgggtgat	4020
	gacgtcgtcg acaatcagga ccctcttgcc cttgagcgaa gcgccgacga tgttgcgcgc	4080
	ctcgccgtgg tccttggcct ccttgcggtc aaacgagtag gagacgcggt ccaggttctg	4140
	gggcgccagc tcgccgagct tgatggtgat ggccggagcac agcgggatgc cctttagaggc	4200
	cgggccgaag acgatgtcga actctaggcc ggcttctctc tgggcctcga tgatggtctt	4260
	tgcaaaggcg gaggcgatgg cgccggcgag gcgcgccgtg tggaattcgc ccgcgttgaa	4320
	gaagtagggg gatataccgt tggacttgag ctgaagctg ccaaacttga ggacgccgcc	4380
	gtcagtggtg gatttgagga agtcctgctt gtaggcaggc agctgggagg tggtagccat	4440
	tctgttgat ttggatagtg tccttattct ctgatttgaa cagtagatca ggacgagtga	4500
	gagggatgca gaggttgat tggagtgtt gagctataaa atttagaggc gcgccgtatc	4560
	gagttttcac atggaagtca aagcgtacag tgcgagcttg tacgttggtc ttagtatccc	4620
	acaagcttct gtctaggtat gatgatggtc ataatgcacc caaggcagaa ctcatcttga	4680
	agattgtcta gagtgatttt accgctgatg aaatgactgg actccctcct cctgctctta	4740
	tacgaaaaat tgcctgactc tgcaaagggt gtttgtcttg gaagatgatg tgcccccca	4800
	tcgtcttat ctcatacccc gccatcttct tagattctca tcttcaacaa gaggggcaat	4860
	ccatgatctg cgatccagat gtgcttcttg cctcatactc tgccttcagg ttgatgttca	4920
	cttaattggt gacgaattca gctgatttgc tgcagtatgc tttgtgttgg ttctttccag	4980
[0036]	gcttgtgcca gccatgagcg ctttgagagc atgttgtcac ctataaactc gagtaacggc	5040
	cacatattgt tcaactactg aatcacatac ctaattttga tagaattgac atgtttaaag	5100
	agctgaggta gctttaatgc ctctgaagta ttgtgacaca gcttctcaca gagtgagaat	5160
	gaaaagtgg actcccccata atgaagtaaa agtttcgtct ctgaacgggtg aagagcatag	5220
	atccggcatc aactacctgg ctagactacg acgtcaattc tgcggccttt tgacctttat	5280
	atatgtccat taatgcaata gattcttttt tttttttttt tttttttttt tttttttttt	5340
	ttttttgccc aatttcgcag atcaaagtgg acgttatagc atcataacta agctcagttg	5400
	ctgagggaag ccgtctacta ccttagccca tccatccagc tccataacct gatactttag	5460
	acgtgaagca attcacactg tacgtctcgc agctctcctt cccgctcttg cttccccact	5520
	gggtgccatg gtgcgtgtat cgccccctcc ttaattaagg ccatttaggc cgttgctggc	5580
	gtttttccat aggctccgcc cccctgacga gcatcacaaa aatcgacgct caagtcagag	5640
	gtggcgaaac ccgacaggac tataaagata ccaggcggtt ccccttgaa gctccctcgt	5700
	gcgctctcct gttccgacct tgcgccttac cggatacctg tccgccttct tcccttcggg	5760
	aagcgtggcg ctttctcata gctcacgtg taggtatctc agttcgggtg aggtcgttcg	5820
	ctccaagctg ggctgtgtgc acgaaccccc cgctcagccc gaccgctgcg ccttatccgg	5880
	taactatcgt cttgagtcca acccggttag acacgactta tcgccactgg cagcagccac	5940
	tggtaacagg attagcagag cgaggtatgt aggcgggtgt acagagttct tgaagtgggtg	6000
	gcctaactac ggctacacta gaaggacagt atttggtatc tgcgctctgc tgaagccagt	6060
	taccttcgga aaaagagttg ttagctcttg atccggcaaa caaaccaccg ctggtagcgg	6120
	tggttttttt gtttgcgaag agcagattac gcgcagaaaa aaaggatctc aagaagatcc	6180
	tttgatcttt tctacgggtg ctgacgtcga gtggaacgaa aactcacgtt aaggcctgca	6240
	gggccgattt tggatcatgag attatcaaaa aggatcttca cctagatcct tttaaattaa	6300
	aatgaagtt ttaaatcaat cttaaagtata tatgagtaaa cttggtctga cagttacca	6360

	tgcttaatca gtgaggcacc tatctcagcg atctgtctat ttcgttcac	6420
	tgactccccg tcgtgtagat aactacgata cgggagggct taccatctgg cccagtgct	6480
	gcaatgatac cgcgagaccc acgctcaccg gctccagatt tatcagcaat aaaccagcca	6540
	gccggaaggg ccgagcgcag aagtggtcct gcaactttat ccgcctccat ccagtctatt	6600
	aattgttgcc gggaagctag agtaagtagt tcgccagtta atagtttgcg caacgttggt	6660
	gccattgcta caggcatcgt ggtgtcacgc tcgtcgtttg gtatggcttc attcagctcc	6720
	ggttcccaac gatcaaggcg agttacatga tccccatgt tgtgcaaaaa agcgggttagc	6780
	tccttcggtc ctccgatcgt tgtcagaagt aagtggccg cagtgttate actcatggtt	6840
	atggcagcac tgcataattc tcttactgtc atgccatccg taagatgctt ttctgtgact	6900
	ggtgagtact caaccaagtc attctgagaa tagtgtatgc ggcgaccgag ttgctcttgc	6960
	ccggcgtcaa tacgggataa taccgcgcca catagcagaa ctttaaaagt gctcatcatt	7020
	ggaaaaagtt cttcggggcg aaaactctca aggatcttac cgctgttgag atccagttcg	7080
	atgtaacca ctcgtgcacc caactgatct tcagcatctt ttactttcac cagcgtttct	7140
	gggtgagcaa aaacaggaag gcaaaatgcc gcaaaaaagg gaataagggc gacacggaaa	7200
	tgttgaatac tcatactctt cctttttcaa 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 ccggtcccct tatcccagct gttccacgtt ggccctgcccc tcagttagcg	60
	ctcaactcaa tgcccctcac tggcgaggcg agggcaagga tggaggggca gcatcgctg	120
	agttggagca aagcggccgc catgggagca gcgaaccaac ggagggatgc cgtgctttgt	180
	cgtggctgct gtggccaate cgggcccttg gttggetcac agagcgttgc tgtgagacca	240
	tgagctatta ttgctaggta cagtatagag agaggagaga gagagagaga gagagagggg	300
	aaaaaagtg aggttgaagt gaaaaaaaaa aaaaaaaaaa aaatccaacc actgacggct	360
	gccggctctg ccacccccct cctccaccc cagacaacct gcacactcag cgcgcagcat	420
	cacctaatct tggtcgcct tcccgagct caggttggtt tttttttctc tctccctcgt	480
	cgaagccgcc cttgttccct tatttatttc cctctccatc cttgtctgcc tttggtccat	540
	ctgccccctt gtctgcctct cttttgcacg catcgctta tcgtcgtctc ttttttact	600
	cacgggagct tgacgaagac ctgactcgtg agcctaacct gctgatttct ctccccccct	660
	cccgaaccggc ttgacttttg tttctctcc agtaccttat cgcgaagccg gaagaacctt	720
	ctttaacccc atcaacaag ttgtacaaa aaagcaggct ccgcggccgc ccccttcacc	780
	atgggtcag cagctccggc ccagggtct gtagctgcag ctgcaggcgg cctccagct	840
	gctggcgtg gcgtggcg tgctccagcc ctaccacct cgcgcgagtc tgcctcgcc	900
	tcgcagccc gctcgccaac cgctcaacc acgcgcgcgc agaactcact cgtgtcggt	960
	gcaacctcgt tccaccacca tcccagaggc cgtctggtga gcagagcctg cgaccgtgc	1020
	cgcggcgca aggccaaggt cagtctagcc ctttgctgt tgcctgcac tctgtgtca	1080
	ttgtcctcc tctgtcgt gctgatgtg ctgctctcc tctcctcct cctccccgtc	1140
	tcttggtccc tggctccctgc tcttcatatg tcttactgc ccgtgtctc tctccccgtt	1200
	cccgttcccc ctctcccgct cctcttctcc tgcgtgtctg tcattcgctac aaagcataca	1260



[0038]	tacaatacat	cagcatacat	ggcaagcggt	gtgttgtgt	gagagttgtg	tgtattgtat	1320
	tgcactgcct	tcacaactcg	ttcatactgc	tgcagectca	ccccaacacc	gacctcgtct	1380
	tccatgctgc	gctactcccc	cgtcttacac	ctggatactc	tctccttgcc	accactgacc	1440
	aatgctcttc	cccgcccaaa	gtgcgagtag	ctcagecgtg	tcgatagctg	cacgcactgc	1500
	cgcgatgccc	acgtgcagtg	cacttttcgac	ctgcccctgg	cgcgacgcgg	ccccaaagcg	1560
	aggaagaaga	gcgaccagcc	cggccagccg	cctcctgata	cgagctcgtc	ctccaccgcg	1620
	gctcgaccgc	gccagatgcc	gccgcccgtg	accttctccg	gccccgcagt	agccgcgctg	1680
	cagcccttcg	cctcgtcgtc	gctgtcgcgc	gacgcggcct	gggagcccgt	cgagccgctc	1740
	agcattgaca	acggcctgcc	ccggcagccg	ctgggcgacc	tgcccgccct	ctccaccatc	1800
	cagaacatct	cgacgcgcca	gcgatggata	cacctggcca	acgccatgac	gctgcgcaac	1860
	acgacgctag	agcgcgtctc	gaagcgatgt	atcgacctct	tcttcgacta	cctctacccc	1920
	ctcacccccc	tgggtgtacga	gccggccctc	cgggacgtgc	tcgcatacat	cttctcccag	1980
	cccttgcctg	gcgtcaacca	accatcgccg	ctgtcacagc	tcacgccaga	cccgaccacc	2040
	ggcaccaccc	ccctcaacgc	tgccgagtcg	tgggcccggc	ttggccagcc	cagcggtctg	2100
	cgaaccgctg	gcagcaggct	ggctcccctg	gccgactcga	ccttcaccct	ggtcacggcc	2160
	gtctgcgcag	aggcagcatt	catgctaccc	aaggacattt	tccccgaagg	agaatccgtc	2220
	tctgagatct	tgctcgaagc	ctctcgggac	tgctgcacc	agcacctcga	ggccgacctg	2280
	gagaatccga	cggccaactc	gattgccatt	cgtacttcc	actccaactg	cctccacgct	2340
	gcggggaagc	ccaagtactc	gtggcacata	tttggcgagg	ccatccgcct	ggcgaggtgc	2400
	atgcagctgc	acgaggaggc	tgcctcgcag	gggctcgtcc	ccatcgaggc	agagttccgc	2460
	cgtcgtctgt	tttggatcct	gtacttgggc	gacaagtcag	ccgtataact	caacaatcgg	2520
	cccatcacca	tcacaagta	ctgcttcgac	gccggcatca	ccacgtata	cccgtcgggt	2580
	atcgaggacg	agttcctgag	cacggcgctc	gagccgcccc	ggaagagctt	catatccggc	2640
	ttcaacgcaa	atgtgcggct	ctggcagtc	gcggctgatt	tgctgctgga	aatccgcgtg	2700
	ctgcaagatc	agatgatgca	gcactttcga	gggaccatgc	ccccgaacca	tgtgctgccc	2760
	tccgccgaca	ggcagcatct	cgattctctc	tatgtccgct	tcacacactg	cttgacgat	2820
	ctcccgcgt	acctccagtc	gtgcaactct	gcgatggcag	cgatggcaga	aggcaacggg	2880
	tctgccgagt	ccaagcagta	cgtgatacag	tgcatcaacc	tgcaggtgac	gtttcactgt	2940
	ctgcgatg	taattacgca	gaaattcgaa	gacctctctt	attttgctcc	tggcggtgag	3000
	caggctgata	tcagaaagtc	ggagattgtg	cgagacatgc	tgagggtgat	gaacgaggcg	3060
	cccttttggg	gcctgcaggc	caatggcgag	ccaaacgtga	gtcgtttcct	tgtctcttct	3120
	cttttctgca	caccttttct	ttcgacgacc	ccccctctct	ctttatatcc	ctgcggatat	3180
	gtatatcacc	aagcctcggc	acttgttgc	aatctgtcct	gattatgttg	tctggatgct	3240
	gcaggttgaa	aagattcgcc	ttatcggagc	tagtttgcgt	gccatcatcc	atcgcaacca	3300
	ggattcaccc	ttggctacgc	gagccaggag	cgaactttcc	gtgcttttgg	atattctcac	3360
	gcggctggac	tcgaaggcgt	cggaccaact	gaggaatacg	tccactaccg	ttgttggtta	3420
	aatgtgtgtt	ggaacaacaa	aaaatgtcaa	agtcggtgta	aatatggcca	ggatctttgt	3480
	gttattcccc	cttcagcggt	gctgggtatt	tcccctttgt	ttactctttt	ctgttttttc	3540
	cagcacttgt	ttttccagca	gtggggggaa	caaaaggcgt	ttctttcccc	tatgccaggg	3600
	gttgccgat	ttagcatgtg	agtgtacatc	ttccctacat	tactaggtac	ttaatgagct	3660
	tatggagatc	tcccgtcatt	ccggatatct	atcacgttgg	tgtatatatc	cgtggttggc	3720
	tttgaaacct	ggagttgggt	tgcaatgcag	tgacgccttt	tgcaaggagc	caaaataagc	3780
	gaaggatgaa	gtctgaatag	gatacgaact	ggctacctat	gggtgagcat	gaaatgaagc	3840
	ggtcggggaa	atggcggaga	aacgctcgac	gtaacgctgt	tggttttctc	cgtttcgtgc	3900

	aatgttgctc gcataaccta cgctagctag ttacgcttgt ttatttacga caagatctag	3960
	aagattcgag atagaataat aataataaca acaatttgcc tcttctttcc accttttcag	4020
	tcttactctc ccttctgaca ttgaacgcct caatcagtca gtcgccttgt acttggcacg	4080
	gtaatcctcc gtgttcttga tatectcagg ggtagcaaag cccttcatgc catcgataat	4140
	gtcatccaga gtgaggatgg caaagatggg gatgccgtac tccttctca gctcgccaat	4200
	ggcactcggt ccaggcttgg agtcgtcgcc atccgcagcg gggagcttct ccatgcbgctc	4260
	cagggccacg acgatgccgg cgacgatgcc gccctccttg gtgatcttct caatggcgctc	4320
	cctcttggcg gtgccggcgg tgatgacgtc gtcgacaatc aggacctctc tgcccttgag	4380
	cgaagcgccg acgatgttgc cgccctcgcc gtggtccttg gccctccttg ggtcaaacga	4440
	gtaggagacg cgggtccagggt tctggggcgc cagctcgccg agcttgatgg tgatggcgga	4500
	gcacagcggg atgcccttgt aggccgggcc gaagacgatg tcgaactcta ggccggcctt	4560
	ctcctgggcc tcgatgatgg tctttgcaa ggccggaggcg atggcgccgg cgaggcgcg	4620
	cgtgtggaat tcgcccgcgt tgaagaagta gggggataac cgcttggaact tgagctcgaa	4680
	gctgcaaac ttgaggacgc cgccgtcgat ggccgatttg aggaagtcct gctttaggc	4740
	aggcagctgg gaggtgtag ccattctgtt ggatttggat agtgtcctta ttctctgatt	4800
	tgaacagtag atcaggacga gtgagaggga tgcagagggt ggattggagt ggttagccta	4860
	taaaatttag aggcgcgcgg tatcgagttt tcacatggaa gtcaaagcgt acagtgcgag	4920
	cttgtagctt ggtcttagta tcccacaagc ttctgtctag gtatgatgat ggctataagt	4980
	caccaaggc agaactcctc ttgaagattg tctagagtga tttaccgct gatgaaatga	5040
	ctggactccc tcctcctgct cttatacgaa aaattgcctg actctgcaa ggttgtttgt	5100
[0039]	cttgaagat gatgtgcccc cccatcgctc ttatctcata ccccgccatc tttctagatt	5160
	ctcatcttca acaagagggg caatccatga tctcgatcc agatgtgctt ctggcctcat	5220
	actctgcctt caggttgatg ttacttaat tggtagcgaa ttcagctgat ttgctgcagt	5280
	atgtcttgtg ttggttcttt ccaggcttgt gccagccatg agcgtcttga gagcatgttg	5340
	tcacctataa actcgagtaa cggccacata ttgttacta cttgaatcac atacctaatt	5400
	ttgatagaat tgacatgttt aaagagctga ggtagcttta atgcctctga agtattgtga	5460
	cacagcttct cacagagtga gaatgaaaag ttggactccc cctaataag taaaagtctc	5520
	gtctctgaac ggtgaagagc atagatccgg catcaactac ctggttagac tacgacgtca	5580
	attctgcggc cttttgacct ttatatatgt ccattaatgc aatagattct ttttttttt	5640
	tttttttttt tttttttttt tttttttttt gcccaatttc gcagatcaaa gtggacgtta	5700
	tagcatcata actaagctca gttgctgagg gaagccgtct actaccttag cccatccatc	5760
	cagctccata ccttgatact ttagacgtga agcaattcac actgtacgtc tcgcagctct	5820
	ccttcccgtc cttgtctccc cactggggtc catggtgcgt gtatcgctcc ctccttaatt	5880
	aaggccattt aggccgttgc tggcgttttt ccataggctc cgccccctg acgagcatca	5940
	caaaaatcga cgctcaagtc agaggtggcg aaaccgcaca ggactataaa gataccaggc	6000
	gtttccccct ggaagctccc tcgtgcgtc tcctgttccg accctgccgc ttaccggata	6060
	cctgtccgcc tttctccctt cgggaagcgt ggcgctttct catagctcac gctgtaggta	6120
	tctcagttcg gtgtaggtcg ttgcgtccaa gctgggctgt gtgcacgaac cccccgtca	6180
	gcccagccgc tgcgccttat ccgtaacta tcgtcttgag tccaaccgg taagacacga	6240
	cttatcgcca ctggcagcag ccactggtaa caggattagc agagcgaggt atgtaggcgg	6300
	tgctacagag ttcttgaagt ggtggcctaa ctacggctac actagaagga cagtatttgg	6360
	tatctgcgtc ctgctgaagc cagttacctt cggaataaga gttggtagct cttgatccgg	6420
	caaacaaacc accgctggta gcggtggttt tttgtttgc aagcagcaga ttacgcgcag	6480

[0040]

```

aaaaaaagga tctcaagaag atcctttgat cttttctacg gggctctgacg ctcagtggaa 6540
cgaaaaactca cgttaaggec tgcagggccg attttgggtca tgagattatc aaaaaggatc 6600
ttcacctaga tcctttttaa ttaaaaatga agttttaaat caatctaaag tatatatgag 6660
taaacttgggt ctgacagtta ccaatgctta atcagttagg cacctatctc agcgatctgt 6720
ctatttcgtt catccatagt tgcctgactc cccgtcgtgt agataactac gatacgggag 6780
ggcttaccat ctggccccag tctgcaatg ataccgcgag acccacgctc accggctcca 6840
gatttatcag caataaacca gccagccgga agggccgagc gcagaagtgg tcctgcaact 6900
ttatccgcct ccatccagtc tattaattgt tgcggggaag ctagagtaag tagttcgcca 6960
gttaatagtt tgcgcaacgt tgttgccatt gctacaggca tcgtgggtgtc acgctcgtcg 7020
tttggtatgg cttcattcag ctccggttcc caacgatcaa ggcgagttac atgatcccc 7080
atgttgtgca aaaaagcggg tagctccttc ggtcctccga tcgttgtcag aagtaagttg 7140
gccgcagtgt tatcactcat ggttatggca gcactgcata attctcttac tgtcatgcca 7200
tccgtaagat gcttttctgt gactgggtgag tactcaacca agtcattctg agaatagtgt 7260
atgcggcgac cgagttgctc ttgcccggcg tcaatacggg ataataccgc gccacatagc 7320
agaactttta aagtgtcat catttgaaaa cgttcttcgg ggcgaaaact ctcaaggatc 7380
ttaccgctgt tgagatccag ttgatgtaa cccactcgtg cacccaactg atcttcagca 7440
tcttttactt tcaccagcgt ttctgggtga gcaaaaacag gaaggcaaaa tgccgcaaaa 7500
aaggaataa gggcgacacg gaaatgttga atactcatac tcttctttt tcaatattat 7560
tgaagcattt atcagggta ttgtctcatg gccatttagg cct 7603

```

&lt;210&gt; 45

&lt;211&gt; 7000

&lt;212&gt; DNA

&lt;213&gt; 人工序列

&lt;220&gt;

&lt;223&gt; 质粒 pYL4

&lt;400&gt; 45

```

agactagcgg cgggtccctt tatccagct gttccacgtt ggctgcccc tcagttagcg 60
ctcaactcaa tgccctcac tggcgaggcg agggcaagga tggaggggca gcatcgctg 120
agttggagca aagcgccgc catgggagca gcgaaccaac ggagggatgc cgtgcttgt 180
cgtggctgct gtggccaatc cgggcccttg gttggctcac agagcgttgc tgtgagacca 240
tgagctatta ttgctaggta cagtatagag agaggagaga gagagagaga gagagagggg 300
aaaaaggtg aggttgaagt gagaaaaaaa aaaaaaaaaa aaatccaacc actgacggct 360
gccggetctg ccacccccct ccctccaccc cagacaacct gcacactcag cgcgcagcat 420
cacctaattc tggctcgctt tcccgagct caggttgttt ttttttctc tctccctcgt 480
cgaagccgcc cttgttccct tatttatctc cctctccatc cttgtctgcc tttggtccat 540
ctgccccctt gtctgcatct cttttgcacg catcgcttca tcgtcgtctc ttttttact 600
cacgggagct tgacgaagac ctgactcgtg agcctaacct gctgatttct ctccccct 660
cccgaccggc ttgacttttg tttctctcc agtaccttat cgcgaagcgc gaagaacct 720
ctttaacccc atcaacaag tttgtacaaa aaagcaggct ccgcggccgc ccccttcacc 780
atgtgcgct actccccct cttacacctg gatactctct ccttgccacc actgaccaat 840
gctcttcccc gcccaaagtg cgagtacctc agcgtgtctg atagctgcac gcactgccgc 900
gatgccacg tgcagtgcac ttctgacctg cccctggcgc gacgcggccc caaagcgagg 960
aagaagagcg accagcccgg ccagccgcct cctgatccga gctcgtctc caccgcggct 1020
cgacccggcc agatgccgc gccgctgacc ttctccggcc ccgcagtagc cgcgctgcag 1080

```

[0041]	cccttcgcct cgctcgtcgt gtcgcccac gcggcctggg agcccgtcga gccgctcagc	1140
	attgacaacg gcctgccccg gcagccgctg ggcgacctgc ccggcctctc caccatccag	1200
	aacatctcga cgcgccagcg atggatacac ctggccaacg ccatgacgct gcgcaacacg	1260
	acgctagagc gcgctctgaa gcgatgtatc gacctcttct tcgactacct ctacccccctc	1320
	acccccctgg tgtacgagcc ggccctccgg gacgtgctcg catacatctt ctcccagccc	1380
	ttgcctggcg tcaaccaacc atcgccgctg tcacagctca cgccagaccc gaccaccggc	1440
	accaccccc tcaacgctgc cgagtcgtgg gccggctttg gccagcccag cggctcgcga	1500
	accgtcggca gcaggctggc tccttgggcc gactcgacct tcaccttggc cagggccgtc	1560
	tgcgcagagg cagcattcat gctaccaag gacatcttcc ccgaaggaga atccgtctct	1620
	gagatcttgc tcgaagcctc tcgggactgc ctgcaccagc acctcgaggc cgacctggag	1680
	aatccgacgg ccaactcgat tgccattcgc tacttccact ccaactgcct ccacgtcgcg	1740
	gggaagccca agtactcgtg gcacatatct ggcgaggcca tccgcctggc gcaggctcatg	1800
	cagctgcacg aggaggtcgc cctcgagggg ctctgcccc a tcgaggcaga gttccgcccgt	1860
	cgctgctttt ggatcctgta cttgggagc aagtcagccg ctatactcaa caatcgcccc	1920
	atcaccatcc acaagtactg cttcgacgcc ggcatcacca cgctataccc gtcgggtatc	1980
	gaggacgagt tcctgagcac ggcgctccgag ccgccccgga agagcttcat atccggttc	2040
	aacgcaaatg tgcggctctg gcagtcgcgc gctgatttgc tgcgtgaaat ccgctgctg	2100
	caagatcaga tgatgcagca ctttcgaggg accatgcccc cgaacctagt gctgccctcc	2160
	gccgacaggc agcatctcga ttctctctat gtccgcttca tcacctgctt ggacgatctc	2220
	ccgccgtacc tccagtcgtg cactctggcg atggcagcga tggcagaagg caacgggtct	2280
	gccgagtcga agcagtagct gatacagtgc atcaacctgc aggtgacgtt tcaactgtctg	2340
	cgcatggtaa ttacgcagaa attcgaagac ctctcttatt ttgctcctgg cgttgagcag	2400
	gctgatctca gaaagtgcga gattgtgcga gacatgctga gggatgatga cgaggcgcgc	2460
	ttttggggcc tgcaggccaa tggcgagcca aacgtgagtc gtttccttgt ctcttctctt	2520
	ttctgcacac ccttttcttc gacgaccccc cctctctctt tataccctg cggatatgta	2580
	tatcatcaag cctcggcact tgttgctaat ctgtcctgat tatgttgtct ggatgctgca	2640
	ggttgaaaag attcgctta tcggagctag tttgctggcc atcatccatc gcaaccagga	2700
	ttcacccctt gctacgcgag ccaggagcga cttttccgtg cttttggata ttctcacgcg	2760
	gctggactcg aaggcgtcgg accaactgag gaatacgtcc actaccgttg ttggctaaat	2820
	gtgtgttga acaacaaaa atgtcaaagt cgggtgaaat atggccagga tctttgtgtt	2880
	attccccctt cagcgttget gggatattcc ccttgttta ctctttctg tttttccag	2940
	cacttgtttt tccagcagtg gggggaacaa aaggcgttcc tttccctat gccagggtt	3000
	gtccgattta gcatttgagt gtacatcttc cctacattac taggtactta atgagcttat	3060
	ggagatctcc cgtcattccg gatattcatc acgttgggtg atatatccgt ggttggcttt	3120
	gaaacctgga gttgggttgc aatgcagtga cgcttttgc gaaggaccaa aataagcgaa	3180
	ggatgaagtc tgaataggat acgaactggc tacctatggg tgagcatgaa atgaagcgg	3240
	cggggaaatg gcggagaaac gctcgacgta acgtgttgg tttctccgt ttcgtgcaat	3300
	gttgctcgca taacctacgc tagctagtta cgttgttta tttacgacaa gatctagaag	3360
	attcgagata gaataataat aataacaaca attgcctct tctttccacc ttttcagtct	3420
	tactctccct tctgacattg aacgcctcaa tcagtcagtc gccttgtaact tggcacggta	3480
	atctccgtg ttcttgatat cctcaggggt agcaaagccc ttcattgcat cgataatgtc	3540
	atccagagtg aggatggcaa agatggggat gccgtactcc ttcctcagct cgccaatggc	3600
	actcgtcca ggcttgaggt cgtcgccatc cgcagcgggg agcttctcca tgcggtccag	3660
	ggccacgacg atgccggcga cgatgccgcc ctcttgggtg atcttctcaa tggcgtccct	3720

	cttggcgggtg cggcggtga tgacgtcgtc gacaatcagg accctcttgc ccttgagcga	3780
	agcgccgacg atgttgccgc cctcgccgtg gtccctggcc tccttgccgt caaacgagta	3840
	ggagacgcgg tccaggttct gggcgccag ctgcccagc ttgatggtga tggcggagca	3900
	cagcgggatg cctttagg cggggccgaa gacgatgtcg aactctagc cggccttctc	3960
	ctgggcctcg atgatgtct ttgcaaaggc ggaggcgatg gcgccggcga ggcgcccggt	4020
	gtggaattcg ccgcgttga agaagtaggg ggatatccgc ttggacttga gctcgaagct	4080
	gccaaacttg aggacccgc cgtcgaatgg ggatttagg aagtcctgct tgtaggcagg	4140
	cagctgggag gtggtagcca ttctgttga ttggatagt gtccttattc tctgatttga	4200
	acagtagatc aggacgagt agagggatgc agaggttgga ttggagtgtg tgagctataa	4260
	aatttagagg cgcccgat cagattttca catggaagtc aaagcgtaca gtgcgagctt	4320
	gtacgttggc cttatgtacc cacaagcttc tctctaggta tgatgatggc tataagtcac	4380
	ccaaggcaga actcatcttg aagattgtct agagtattt taccgtgat gaaatgactg	4440
	gactccctcc tcctgtctt atacgaaaaa ttgcctgact ctgcaaaggc tgtttgtctt	4500
	ggaagatgat gtgcccccc atcgtcttca tctcataccc cgccatctt ctagattctc	4560
	atcttcaaca agaggggcaa tccatgatct gcgatccaga tgtgcttctg gcctcact	4620
	ctgccttcag gttgatgtc acttaattgg tgacgaattc agctgatttg ctgcagatg	4680
	ctttgtgttg gttctttcca ggcttgtgcc agccatgagc gctttgagag catgttgtca	4740
	cctataaact cgagtaacgg ccacatatg ttcactactt gaatcacata cctaattttg	4800
	atagaattga catgtttaaa gagctgaggt agctttaatg cctctgaagt attgtgacac	4860
	agcttctcac agagtgagaa tgaaggtg gactccccct aatgaagtaa aagtttcgtc	4920
	tctgaacggg gaagagcata gatccggcat caactacctg gctagactac gacgtcaatt	4980
[0042]	ctgcggcctt ttgacctta tatatgtcca ttaatgcaat agattctttt ttttttttt	5040
	ttttttttt ttttttttt tttttttgcc caatttcgca gatcaaagtg gacgttatag	5100
	catcataact aagctcagtt gctgaggga ggcgtctact acctagccc atccatccag	5160
	ctccatacct tgatacttta gacgtgaage aattcacact gtacgtctcg cagctctct	5220
	tcccgtctt gcttccccac tggggtccat ggtgcgtgta tcgtccctc cttaattaag	5280
	gccatttagg ccgttcttg cgtttttcca taggtccgc cccctgacg agcatcaca	5340
	aaatcgacgc tcaagtcaga ggtggcgaaa cccgacagga ctataaagat accaggcgtt	5400
	tccccctgga agctccctcg tgcgtctcc gtctccgacc ctgccgtta ccgatacct	5460
	gtccgcctt ctcccttcgg gaagcgtggc gctttctcat agctcacgct gtaggtatct	5520
	cagttcgggt taggtcgtc gctccaagct gggtgtgtg cacgaacccc ccgttcagcc	5580
	cgaccgctgc gccttatcc gtaactatcg tcttgagtcc aaccgggtaa gacacgactt	5640
	atgccactg gcagcagcca ctggtaacag gattagcaga gcgaggtatg taggcggtgc	5700
	tacagagttc ttgaagtgt ggcctaacta cggctacact agaaggacag tatttggtat	5760
	ctgcgtctg ctgaagccag ttaccttcgg aaaaagagtt ggtagctctt gatccggcaa	5820
	aaaaccacc gctggtagcg gtggtttttt tgtttgcaag cagcagatta cgcgcagaaa	5880
	aaaagatct caagaagatc ctttgatctt ttctacgggg tctgacgctc agtggaacga	5940
	aaactcacgt taaggcctgc agggccgatt ttggtcatga gattatcaaa aaggatcttc	6000
	acctagatcc ttttaaatta aaaatgaagt tttaaatcaa tctaaagtat atatgagtaa	6060
	acttggtctg acagttacca atgcttaac agtgaggcac ctatctcagc gatctgtcta	6120
	tttcttcat ccatagtgc ctgactcccc gtcgtgtaga taactacgat acgggagggc	6180
	ttaccatctg gccccagtgc tgcaatgata ccgcgagacc cacgtcacc ggctccagat	6240
	ttatcagcaa taaaccagcc agccggaagg gccgagcgca gaagtgttcc tgcaacttta	6300
	tccgcctcca tccagtctat taattgttgc cgggaagcta gagtaagtag ttcgccagtt	6360

```

aatagtttgc gcaacgttgt tgccattgct acaggcatcg tgggtgtcacg ctctgtcttt 6420
ggtatggctt cattcagctc cggttcccaa cgatcaaggc gagttacatg atcccccatg 6480
ttgtgcaaaa aagcggtag ctcttctcggc cctccgatcg ttgtcagaag taagttggcc 6540
gcagtgttat cactcatggt tatggcagca ctgcataatt ctcttactgt catgccatcc 6600
gtaagatgct tttctgtgac tgggtgagtac tcaaccaagt cattctgaga atagtgtatg 6660
cggcgaccga gttgctcttg cccggcgctca atacgggata ataccgcgcc acatagcaga 6720
actttaaaag tgctcatcat tggaaaacgt tcttcggggc gaaaactctc aaggatctta 6780
ccgctgttga gatccagttc gatgtaacct actcgtgcac ccaactgac ttcagcatct 6840
tttactttca ccagcgttcc tgggtgagca aaaacaggaa ggcaaaatgc cgcaaaaaag 6900
ggaataaggg cgacacggaa atgttgaata ctcatctct tctttttca atattattga 6960
agcatttacc agggttattg 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

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

[0043]

[0044]

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 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
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

[0045]

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	
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	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	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	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 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 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  
 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
	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		
[0047]	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		

[0048]

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
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 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
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
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
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

[0049]

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  
 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 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 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 ggttttccca gtcacgacgg tttaaactcc atacgcagca aacatgggct	60
	tgggc	65
	<210> 50	
	<211> 64	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 50	
	gtacgagtac taggtgtgaa gattccgtca agcttgggcg gaatgaagga ggatgtgtga	60
	gagg	64
	<210> 51	
	<211> 59	
	<212> DNA	
	<213> 人工序列	
	<220>	
[0050]	<223> 引物	
	<400> 51	
	cacacatcct ccttcattcc gccaagctt gacggaatct tcacacctag tactcgtac	59
	<210> 52	
	<211> 57	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 52	
	tgacattttt tgttgttcca acacagcatg cttagtccga cgccttcgag tccagcc	57
	<210> 53	
	<211> 54	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 53	
	ctggactcga aggcgtcgga ctaagcatgc tgtgttgga caacaaaaaa tgtc	54
	<210> 54	
	<211> 54	
	<212> DNA	
	<213> 人工序列	

	<220>		
	<223>	引物	
	<400>	54	
		gcagagcagc agtagtcgat gctattaatt aagtaggtta tgcgagcaac attg	54
	<210>	55	
	<211>	58	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	引物	
	<400>	55	
		ctcagcctct ctcagcctca tcagccgcgg ccgctgaatc ggcaaggggt agtactag	58
	<210>	56	
	<211>	62	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	引物	
	<400>	56	
		gcggataaca atttcacaca ggaaacagcg tttaaaccac atgccagagt tcgatgcgca	60
		ag	62
[0051]	<210>	57	
	<211>	57	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	引物	
	<400>	57	
		gtacctcagc gctgtcgata gctgcacgca ctgcccgcat gccacgtgc agtgcac	57
	<210>	58	
	<211>	59	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	引物	
	<400>	58	
		gtgcactgca cgtgggcata gcggcagtgc gtgcagctat cgacagcgct gaggtactc	59
	<210>	59	
	<211>	48	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	引物	
	<400>	59	

	gcggcgcttc cgctgtcgta actatgctgc 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 aactatgggc tcagcagctc cggcccaggg etc	53
	<210> 62	
	<211> 58	
	<212> DNA	
	<213> 人工序列	
	<220>	
[0052]	<223> 引物	
	<400> 62	
	gccctgggcc 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 acagcgcccg cggcagcagc tgg	53
	<210> 64	
	<211> 59	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 64	
	cagctgctgc cgcggccgct gtggccatag ttacgacagc ggaagcgccg cttataag	59
	<210> 65	
	<211> 22	
	<212> DNA	

	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 65	
	gctggaagct gctgagcaga 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 tctttttcca	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	



	tgagtatagc ggctgacttg tcg	23
	<210> 71	
	<211> 59	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 71	
	tctagtatgt acgagtacta ggtgtgaaga ttccgtcatt tcctcgacat gcgaatgcg	59
	<210> 72	
	<211> 59	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 72	
	tgccatgcaa accccgcatt cgcattgtcga ggaaatgacg gaatcttcac acctagtag	59
	<210> 73	
	<211> 59	
	<212> DNA	
	<213> 人工序列	
	<220>	
[0054]	<223> 引物	
	<400> 73	
	tgcagctaca gagecctggg ccggagctgc tgagcccata gttacgacag cggaagcgc	59
	<210> 74	
	<211> 57	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 74	
	atagcactta taaggcggcg cttccgctgt cgtaactatg ggctcagcag ctccggc	57
	<210> 75	
	<211> 75	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 75	
	taaacaata ggggttccgc gcacatttcc ccgaaaagtg ccacctggat agactagcat	60
	ctgagccatt gcagc	75
	<210> 76	
	<211> 60	

	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 76	
	agtggcaccg agtcggtggt gctttttttt ctatcgagag cattggtcag tggtaggcaag	60
	<210> 77	
	<211> 60	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 77	
	accaatatac aaaacatgtc gtccgagcca gtgcctgcca tttcctcgac atgcgaatgc	60
	<210> 78	
	<211> 60	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 78	
[0055]	gttgccatgc aaaccccgca ttcgcatgtc gaggaaatgg caggcactgg ctcggaacgac	60
	<210> 79	
	<211> 60	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 79	
	agctacagag ccctgggccg gagctgctga gccattgtt gaattctggc ggggtagctg	60
	<210> 80	
	<211> 60	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 80	
	cttttacact tttcaacagc taccccgcca gaattcaaca atgggctcag cagctccggc	60
	<210> 81	
	<211> 81	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	

[0056]	<400> 81	
	tagtccgtta tcaacttgaa aaagtggcac cgagtcggtg gtgctttttt ttctatcgag	60
	atgttctgga tggaggagag g	81
	<210> 82	
	<211> 32	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 82	
	agtctatcgc agccttgcct tagctaattg 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	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 84	
	agtctatcgg cagagtcgcg tcttccgggt tt	32
	<210> 85	
	<211> 32	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 85	
	tctaaaaccc ggaagacgcg actctgccga ta	32
	<210> 86	
	<211> 33	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 86	
	agtctatcga atgagtgtag 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	
	agtcctatcgg ccgcaatagc ttcctaattgt 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	
	agtcctatcgc agcgcaatca gtgcagtggg 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>	

[0058]	<223> 引物	
	<400> 92	
	tggagagact cggagaggat agg	23
	<210> 93	
	<211> 22	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 93	
	agcgtggagg 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	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 95	
	atcctgactc gtcctgtgtc gg	22
	<210> 96	
	<211> 22	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 96	
	agtgccttcgt ttagtggact tg	22
	<210> 97	
	<211> 22	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<400> 97	
	ctcggtagct 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]

atgctgcgct actccccgt cttacacctg gatactctct ccttgccacc actgaccaat 60  
 gctcttcccc gcccaaagt cgagtacctc agcgtgtcg atagctgcac gcaactgccgc 120  
 gatgcccacg tgcagtgcac ttctgacctg cccctggcgc gacgcggccc caaagcgagg 180  
 aagaagagcg accagcccgg ccagccgcct cctgatccga gctcgtcttc caccgcggct 240  
 cgacccggcc agatgccgcc gccgtgacc ttctccggcc ccgcagtagc cgcgtgcag 300  
 cccttcgctt cgctcgtcgt gtgcgccgac gcggcctggg agcccgtcga gccgtcagc 360  
 attgacaacg gcctgccccg gcagccgctg ggcgacctgc ccggcctctc caccatccag 420  
 aacatctcga cgcgccagcg atggatacac ctggccaacg ccatgacgct gcgcaacacg 480  
  
 acgctagagc gcgctctgaa gcgatgtatc gacctcttct tcgactacct ctacccccctc 540  
 acccccctgg tgtacgagcc ggccctccgg gacgtgctcg catacatctt ctcccagccc 600  
 ttgcctggcg tcaaccaacc atgcgcgtg tcacagctca cgccagaccc gaccaccggc 660  
 accaccccc tcaacgttgc cgagtcgtgg gccggctttg gccagcccag cggctcgcga 720  
 accgtcgga gcaggtggc tccctgggcc gactcgacct tcaccctggt cagggccgtc 780  
 tgcgcagagg cagcattcat gctacccaag gacatcttcc ccgaaggaga atccgtctct 840  
 gagatcttgc tcgaagcctc tcgggactgc ctgcaccagc acctcgaggc cgacctggag 900  
 aatccgacgg ccaactcgat tgccattcgc tacttccact ccaactgcct ccacgtgcg 960  
 gggaagccca agtactcgtg gcacataatt ggcgaggcca tccgcctggc gcaggtcatg 1020  
 cagctgcacg aggaggttgc cctcgagggg ctctcccca tcgaggcaga gttccgccgt 1080  
 cgctgctttt ggatcctgta cttgggcgac aagtcagccg ctatactcaa caatcgcccc 1140  
 atcaccatcc acaagtactg cttcgacgcc ggcatcacca cgctataccc gtcgggtatc 1200  
 gaggacgagt tcctgagcac ggcgctccgag ccgccccgga agagcttcat atccggttc 1260  
 aacgcaaatg tgcggctctg gcagtcgcg gctgatttgc tgcgggaaat ccgctgctg 1320  
 caagatcaga tgatgcagca ctttcgaggg accatgcccc cgaacctgt gctgccctcc 1380  
 gccgacaggc agcatctcga ttctctctat gtccgcttca tcacctgctt ggacgatctc 1440  
 ccgccgtacc tccagtcgtg cactctggcg atggcagcga tggcagaagg caacgggtct 1500  
 gccgagtcca agcagtacgt gatacagtgc atcaacctgc aggtgacgtt tcaactgtctg 1560

	cgcatggtaa ttacgcagaa attcgaagac ctctcttatt ttgctcctgg cggttgagcag	1620
	gctgatctca gaaagtccga gattgtgcga gacatgctga gggatgatgaa cgaggcgccc	1680
	ttttggggcc tgcaggccaa tggcgagcca aacgttgaaa agattcgctt tatcggagct	1740
	agtttgctgg ccatcatcca tcgcaaccag gattcacctt tggctacgcg agccaggagc	1800
	gacttttccg tgcttttga tattctcacg cggttggaact cgaaggcgct ggactaa	1857
	<210> 100	
	<211> 2070	
	<212> DNA	
	<213> 里氏木霉	
	<400> 100	
	atgggctcag cagctccggc ccagggtctt gtagctgcag ctgcaggcgg ccctccagct	60
	gctggcgctg gcgctggcgc tgtccacgcc ctaccacct cggccgagtc tgcctcgccc	120
	tcgcagcccg gctcgccaac cgctcaacc acgccggcg agaactcact cgtgtcggct	180
	gcaacctcgt tccaccacca tcccagaggc cgtctgggtga gcagagcctg cgaccgctgc	240
	cggcggcgca aggccaagtg cgagtacctc agcgtgtctg atagctgcac gactgccgc	300
	gatgccacg tgcagtgcac ttctgacctg cccctggcgc gacgcggccc caaagcgagg	360
	aagaagagcg accagcccg ccagccgctt cctgatccga gctcgtctc caccgaggct	420
	cgaccggcc agatgccgc gccgtgacc ttctccggcc ccgcagtagc cgcgtgcag	480
	cccttcgctt cgtcgtcgtt gtgcggcgac gggcgctggg agcccgctga gccgctcagc	540
	attgacaacg gcctgcccc gcagccgctg ggcgacctgc ccggcctctc caccatccag	600
	aacatctcga cgcgccagcg atggatacac ctggccaacg ccatgacgt gcgcaacacg	660
	acgctagagc gcgtctcgaa gcgatgtatc gacctcttct tcgactacct ctacccctc	720
	acccccctgg tgtacgagcc ggccctccgg gacgtgctcg catacatctt ctcccagccc	780
	ttgcctggcg tcaaccaacc atcgccgctg tcacagctca cgccagacc gaccaccggc	840
	accaccccc tcaacgtgc cgagtcgtgg gccggtttg gccagcccag cggtctcgga	900
	accgtcggca gcaggctggc tccctgggcc gactcgacct tcacctggt cacggcgctc	960
	tgcgcagagg cagcattcat gctacceaa gacatcttcc ccgaaggaga atccgtctct	1020
	gagatcttgc tcgaagcctc tcgggactgc ctgcaccagc acctcgaggc cgacctggag	1080
	aatccgacgg ccaactcgat tgccattcgc tacttccact ccaactgcct ccacgtgcg	1140
	gggaagccca agtactcgtg gcacatatct ggcgaggcca tccgcctggc gcaggctcatg	1200
	cagctgcacg aggaggtgc cctcgagggg ctctgcccc tcgaggcaga gttccgccgt	1260
	cgtcgtcttt ggatcctgta cttgggcgac aagtcagccg ctatactcaa caatcgcccc	1320
	atcaccatcc acaagtactg ctctgacgcc ggcacacca cgtataccc gtcgggtatc	1380
	gaggacgagt tcctgagcac ggcgtccgag ccgccccgga agagcttcat atccggcttc	1440
	aacgcaaata tgcggctctg gcagtcgcg gctgatttgc tgcctggaaat ccgctgctg	1500
	caagatcaga tgatgcagca ctttcgaggg accatgcccc cgaacctatg gctgccctcc	1560
	gccgacaggc agcatctcga ttctctctat gtccgcttca tcacctgctt ggacgatctc	1620
	ccgccgtacc tccagtcgtg cactctggcg atggcagcga tggcagaagg caacgggtct	1680
	gccgagtcca agcagtacgt gatacagtgc atcaacctgc aggtgacgtt tctgtctgtg	1740
	cgcatggtaa ttacgcagaa attcgaagac ctctcttatt ttgctcctgg cggttgagcag	1800
	gctgatctca gaaagtccga gattgtgcga gacatgctga gggatgatgaa cgaggcgccc	1860
	ttttggggcc tgcaggccaa tggcgagcca aacgttgaaa agattcgctt tatcggagct	1920
	agtttgctgg ccatcatcca tcgcaaccag gattcacctt tggctacgcg agccaggagc	1980
	gacttttccg tgcttttga tattctcacg cggttggaact cgaaggcgct ggaccaactg	2040

[0060]

	aggaatacgt ccactaccgt tgttggctaa	2070
	<210> 101	
	<211> 2172	
	<212> DNA	
	<213> 里氏木霉	
	<400> 101	
	atggccacag cggccgcggc agcagctggc ggcgcggcgg ttgctgcggg tgcagacaca	60
	ggcgtgcag gctccagctc tacaggccct ccaggccttc cagggttcc aggcacccgg	120
	acaggctccg tggcgtggg ctcagcagct ccggcccagg gctctgtagc tgcagctgca	180
	ggcggccctc cagctgctgg cgctggcgt ggcgtgtcc acgccctcac caccctgccc	240
	gagtctgcct cggcctcgca gcccggctcg ccaaccgcct caaccacgcc gccgcagaac	300
	tcactcgtgt cggctgcaac ctggttcac caccatcca gaggcgtct ggtgagcaga	360
	gcctgcgacc gctgccgccg gcgcaaggcc aagtgcgagt acctcagcgc tgtcgatagc	420
	tgcacgcaact gccgcgatgc ccacgtgcag tgcactttcg acctgcccct ggcgcgacgc	480
	ggccccaag cgaggaagaa gagcgaccag cccggccagc cgcctcctga tccgagctcg	540
	ctctccaccg cggtcgcacc cggccagatg ccgcccgcgc tgaccttctc cggccccgca	600
	gtagccgcgc tgcagccctt cgctcgtcg tcgtgtcgc ccgacgcggc ctgggagccc	660
	gtcagccgc tcagcattga caacggcctg cccggccagc cgctgggcca cctgcccggc	720
	ctctccacca tccagaacat ctgcacgcgc cagcgatgga tacacctggc caacgccatg	780
	acgtgcgca acacgacgt agagcgcgtc tcgaagcgat gtatcgacct cttcttcgac	840
	tacctctacc ccctacccc cctggtgtac gagccggccc tccgggacgt gctcgcatac	900
[0061]	atcttctccc agcccttgcc tggcgtcaac caaccatgc cgctgtcaca gtcacgcca	960
	gacccgacca ccggcaccac cccctcaac gctgccgagt cgtgggccgg ctttggccag	1020
	cccagcggt cgcgaaccgt cggcagcagg ctggtccct gggccgactc gaccttcacc	1080
	ctggtcacgg ccgtctgcgc agaggcagca ttcatgctac ccaaggacat ttccccgaa	1140
	ggagaatccg tctctgagat cttgctcgaa gcctctcggg actgcctgca ccagcacctc	1200
	gaggccgacc tggagaatcc gacggccaac tcgattgcca ttcgctactt ccactccaac	1260
	tgcctccacg ctgcggggaa gccaagtac tcgtggcaca tatttggcga ggccatccgc	1320
	ctggcgcagg tcatgcagct gcacgaggag gctgccctcg aggggctcgt ccccatcgag	1380
	gcagagtcc gccgtcgtg cttttggatc ctgtacttgg gcgacaagtc agccgtata	1440
	ctcaacaatc ggcccatcac catccacaag tactgtctcg acgccggcat caccacgcta	1500
	taccctcgg gtatcgagga cgagttcctg agcacggcgt ccgagccgcc ccggaagagc	1560
	ttcatatccg gcttcaacgc aaatgtgcgg ctctggcagt ccgcggtga ttgctgctg	1620
	gaaatccgcg tgctgcaaga tcagatgatg cagcactttc gagggacct gccccgaac	1680
	catgtgctgc cctccgccga caggcagcat ctcgattctc tctatgtccg cttcacacc	1740
	tgcttgacg atctccgcc gtacctccag tcgtgcactc tggcgatggc agcgatggca	1800
	gaaggcaacg ggtctgccga gtccaagcag tacgtgatac agtgcataa cctgcaggtg	1860
	acgtttcact gtctgcgcat ggtaattac cagaaattcg aagacctctc ttattttgct	1920
	cctggcggtg agcaggtga tctcagaaaag tcggagattg tgcgagacat gctgagggtg	1980
	atgaacgagg cgcccttttg gggcctgcag gccaatggcg agccaaacgt tgaaaagatt	2040
	cgccttateg gagctagttt gctggccatc atccatcgca accaggattc acccttggt	2100
	acgcgagcca ggagcgactt ttccgtgctt ttggatattc tcacgcggct ggactcgaag	2160
	gcgtcggact 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 ctccaccct cgcccagatc tgccctcgcc	120	
	tcgcagcccg gctcgccaac cgcctcaacc acgccgcgc agaactcact cgtgtcggt	180	
	gcaacctcgt tccaccacca tcccagaggc cgtctgggta gcagagcctg gcaccgctgc	240	
	cgccggcgca aggccaagtg cgagtacctc agcgtgtctg atagctgcac gcactgccgc	300	
	gatgccacg tcgagtgcac ttctgacctg cccctggcgc gacgcggccc caaagcgagg	360	
	aagaagagcg accagcccgg ccagccgcct cctgatccga gctcgtcttc caccgcggct	420	
	cgacccggcc agatgccgc gccgctgacc ttctccggcc ccgcagtagc cgcgctgcag	480	
	cccttcgect cgtcgtcgt gtcgcccga cgggcctggg agcccgtcga gccgctcagc	540	
	attgacaacg gcctgccccg gcagccgctg ggcgacctgc ccggcctctc caccatccag	600	
	aacatctcga cgcgccagcg atggatacac ctggccaacg ccatgacgt gcgcaacacg	660	
	acgctagagc gcgtctcgaa gcgatgtatc gacctcttct tcgactacct ctacccccctc	720	
	acccccctgg tgtacgagcc ggccctccgg gacgtgtctg catacatctt ctcccagccc	780	
	ttgcctggcg tcaaccaacc atcgccgctg tcacagctca cgccagacc gaccaccggc	840	
[0062]	accaccccc tcaacgctgc cgagtcgtgg gccggctttg gccagcccag cggctcgcga	900	
	accgtcgca gcaggtggc tccctgggccc gactcgacct tcacctggt cacggcgtc	960	
	tgcgagagg cagcattcat gctaccaag gacattttcc ccgaaggaga atccgtctct	1020	
	gagatcttgc tcgaagcctc tcgggactgc ctgcaccagc acctcgaggc gcacctggag	1080	
	aatccgacgg ccaactcgat tgccattcgc tacttccact ccaactgcct ccacgtgcg	1140	
	gggaagccca agtactcgtg gcacatattt ggcgaggcca tccgcctggc gcaggtcatg	1200	
	cagctgcacg aggaggtgc cctcgagggg ctctgcccc tcgaggcaga gtccgccgt	1260	
	cgctgctttt ggatcctgta cttgggcgac aagtcagccg ctatactcaa caatcgcccc	1320	
	atcaccatcc acaagtactg ctctgacgcc ggcatcacca cgctataccc gtcgggtatc	1380	
	gaggacgagt tcctgagcac ggcgccgag ccgccccgga agagcttcat atccggcttc	1440	
	aacgcaaatg tgcggctctg gcagtcgcg gctgatttgc tgctggaaat ccgctgctg	1500	
	caagatcaga tgatgcagca ctttcgaggg accatgcccc cgaacctgt gctgccctcc	1560	
	gccgacagge agcatctcga ttctctctat gtccgcttca tcacctgctt ggacgatctc	1620	
	ccgccgtacc tccagtcgtg cactctggcg atggcagcga tggcagaagg caacgggtct	1680	
	gccgagtcga agcagtagt gatacagtc atcaacctgc aggtgacgtt tcaactgtctg	1740	
	cgcatggtaa ttacgcagaa attcgaagac ctctcttatt ttgctcctgg cgttgagcag	1800	
	gctgatctca gaaagtcgga gattgtgcga gacatgctga gggtagtgaa cgaggcgcgc	1860	
	ttttggggcc tgcaggccaa tggcgagcca aacgttgaaa agattcgctt tatcgagct	1920	
	agtttgcctg ccatcatcca tcgcaaccag gattcaccct tggtacgcg agccaggagc	1980	
	gacttttccg tgcttttggg tattctcacg cggctggact cgaaggcgtc 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

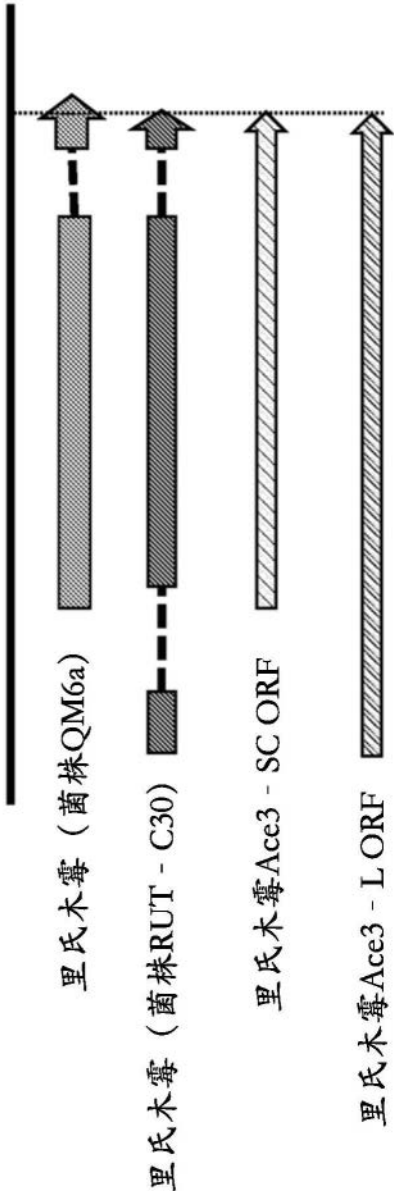


图1A

	1	10	20	30	40	50	60
Ace3-S							
Ace3-SC	-----	-----	-----	-----	-----	-----	-----
Ace3-L	-----	-----	-----	-----	-----	-----	-----
Ace3-LN	-----	-----	-----	-----	-----	-----	-----
Ace3-LC	-----	-----	-----	-----	-----	-----	-----
Ace3-EL	MATAAAAAGGA	AAAGADTGA	AGSSSTG	PPGLPGL	PGTRTGS	VAMGSAAP	QAQGSVAAAA
Ace3-S	-----	-----	-----	-----	-----	-----	-----
Ace3-SC	-----	-----	-----	-----	-----	-----	-----
Ace3-L	GGPPAAGAGAG	AVHALTTS	PESASASQ	PGSPTAST	TPPQNSL	VSAATSF	HHHPRGRLVSR
Ace3-LN	GGPPAAGAGAG	AVHALTTS	PESASASQ	PGSPTAST	TPPQNSL	VSAATSF	HHHPRGRLVSR
Ace3-LC	GGPPAAGAGAG	AVHALTTS	PESASASQ	PGSPTAST	TPPQNSL	VSAATSF	HHHPRGRLVSR
Ace3-EL	GGPPAAGAGAG	AVHALTTS	PESASASQ	PGSPTAST	TPPQNSL	VSAATSF	HHHPRGRLVSR
Ace3-S	-----	-----	-----	-----	-----	-----	-----
Ace3-SC	-----	-----	-----	-----	-----	-----	-----
Ace3-L	ACDRCRRRKAK	CEYLSAVD	SCTHCRDA	HVQCTFDL	PLARRGPK	KARKKSDQ	PGQPPDPSS
Ace3-LN	ACDRCRRRKAK	CEYLSAVD	SCTHCRDA	HVQCTFDL	PLARRGPK	KARKKSDQ	PGQPPDPSS
Ace3-LC	ACDRCRRRKAK	CEYLSAVD	SCTHCRDA	HVQCTFDL	PLARRGPK	KARKKSDQ	PGQPPDPSS
Ace3-EL	ACDRCRRRKAK	CEYLSAVD	SCTHCRDA	HVQCTFDL	PLARRGPK	KARKKSDQ	PGQPPDPSS
Ace3-S	-----	-----	-----	-----	-----	-----	-----
Ace3-SC	-----	-----	-----	-----	-----	-----	-----
Ace3-L	LSTAARPGQM	PPPLTFSG	PAVAALQ	PFASSSL	SPDAAWE	PVEPLSID	NGLPRQPLGDLPG
Ace3-LN	LSTAARPGQM	PPPLTFSG	PAVAALQ	PFASSSL	SPDAAWE	PVEPLSID	NGLPRQPLGDLPG
Ace3-LC	LSTAARPGQM	PPPLTFSG	PAVAALQ	PFASSSL	SPDAAWE	PVEPLSID	NGLPRQPLGDLPG
Ace3-EL	LSTAARPGQM	PPPLTFSG	PAVAALQ	PFASSSL	SPDAAWE	PVEPLSID	NGLPRQPLGDLPG

Ace3-S	LSTIQNISTRQRWIHLANAMTLRNTTLERSVKRCIDLFFDYLYPLTPLVYEPALRDVLAY
Ace3-SC	LSTIQNISTRQRWIHLANAMTLRNTTLERSVKRCIDLFFDYLYPLTPLVYEPALRDVLAY
Ace3-L	LSTIQNISTRQRWIHLANAMTLRNTTLERSVKRCIDLFFDYLYPLTPLVYEPALRDVLAY
Ace3-LN	LSTIQNISTRQRWIHLANAMTLRNTTLERSVKRCIDLFFDYLYPLTPLVYEPALRDVLAY
Ace3-LC	LSTIQNISTRQRWIHLANAMTLRNTTLERSVKRCIDLFFDYLYPLTPLVYEPALRDVLAY
Ace3-EL	LSTIQNISTRQRWIHLANAMTLRNTTLERSVKRCIDLFFDYLYPLTPLVYEPALRDVLAY
Ace3-S	IFSQPLPGVNPSPLSQLTPDPTTGTTPLNAAESWAGFGQPSGSRVTGSR LAPWADSTFT
Ace3-SC	IFSQPLPGVNPSPLSQLTPDPTTGTTPLNAAESWAGFGQPSGSRVTGSR LAPWADSTFT
Ace3-L	IFSQPLPGVNPSPLSQLTPDPTTGTTPLNAAESWAGFGQPSGSRVTGSR LAPWADSTFT
Ace3-LN	IFSQPLPGVNPSPLSQLTPDPTTGTTPLNAAESWAGFGQPSGSRVTGSR LAPWADSTFT
Ace3-LC	IFSQPLPGVNPSPLSQLTPDPTTGTTPLNAAESWAGFGQPSGSRVTGSR LAPWADSTFT
Ace3-EL	IFSQPLPGVNPSPLSQLTPDPTTGTTPLNAAESWAGFGQPSGSRVTGSR LAPWADSTFT
Ace3-S	LVTAVCAEAAAFMLPKDIFPEGESVSEILLEASRDC LHQHLEADLENPTANSIAIRYFHSN
Ace3-SC	LVTAVCAEAAAFMLPKDIFPEGESVSEILLEASRDC LHQHLEADLENPTANSIAIRYFHSN
Ace3-L	LVTAVCAEAAAFMLPKDIFPEGESVSEILLEASRDC LHQHLEADLENPTANSIAIRYFHSN
Ace3-LN	LVTAVCAEAAAFMLPKDIFPEGESVSEILLEASRDC LHQHLEADLENPTANSIAIRYFHSN
Ace3-LC	LVTAVCAEAAAFMLPKDIFPEGESVSEILLEASRDC LHQHLEADLENPTANSIAIRYFHSN
Ace3-EL	LVTAVCAEAAAFMLPKDIFPEGESVSEILLEASRDC LHQHLEADLENPTANSIAIRYFHSN
Ace3-S	CLHAAGKPKYSWHIFGEAIRLAQVMQLHEEAALEG LVP IEAEFRRCFWILYLGDKSAAI
Ace3-SC	CLHAAGKPKYSWHIFGEAIRLAQVMQLHEEAALEG LVP IEAEFRRCFWILYLGDKSAAI
Ace3-L	CLHAAGKPKYSWHIFGEAIRLAQVMQLHEEAALEG LVP IEAEFRRCFWILYLGDKSAAI
Ace3-LN	CLHAAGKPKYSWHIFGEAIRLAQVMQLHEEAALEG LVP IEAEFRRCFWILYLGDKSAAI
Ace3-LC	CLHAAGKPKYSWHIFGEAIRLAQVMQLHEEAALEG LVP IEAEFRRCFWILYLGDKSAAI
Ace3-EL	CLHAAGKPKYSWHIFGEAIRLAQVMQLHEEAALEG LVP IEAEFRRCFWILYLGDKSAAI

Ace3-S	LNRPITIHKYCFDAGITTLYPGIEDEFLSTASEPRKSFISGFNANVRLWQSAADLLL
Ace3-SC	LNRPITIHKYCFDAGITTLYPGIEDEFLSTASEPRKSFISGFNANVRLWQSAADLLL
Ace3-L	LNRPITIHKYCFDAGITTLYPGIEDEFLSTASEPRKSFISGFNANVRLWQSAADLLL
Ace3-LN	LNRPITIHKYCFDAGITTLYPGIEDEFLSTASEPRKSFISGFNANVRLWQSAADLLL
Ace3-LC	LNRPITIHKYCFDAGITTLYPGIEDEFLSTASEPRKSFISGFNANVRLWQSAADLLL
Ace3-EL	LNRPITIHKYCFDAGITTLYPGIEDEFLSTASEPRKSFISGFNANVRLWQSAADLLL
Ace3-S	EIRVLQDQMMQHFRGTMPPNHVLP SADRQHLDLSLYVRFITCLDDLPYQLQSCTLAAMA
Ace3-SC	EIRVLQDQMMQHFRGTMPPNHVLP SADRQHLDLSLYVRFITCLDDLPYQLQSCTLAAMA
Ace3-L	EIRVLQDQMMQHFRGTMPPNHVLP SADRQHLDLSLYVRFITCLDDLPYQLQSCTLAAMA
Ace3-LN	EIRVLQDQMMQHFRGTMPPNHVLP SADRQHLDLSLYVRFITCLDDLPYQLQSCTLAAMA
Ace3-LC	EIRVLQDQMMQHFRGTMPPNHVLP SADRQHLDLSLYVRFITCLDDLPYQLQSCTLAAMA
Ace3-EL	EIRVLQDQMMQHFRGTMPPNHVLP SADRQHLDLSLYVRFITCLDDLPYQLQSCTLAAMA
Ace3-S	EGNSAESKQYVIQINLQVTFHCLRMVITQKFEDLSYFAPGVEQADLRKSEIVRDMRLV
Ace3-SC	EGNSAESKQYVIQINLQVTFHCLRMVITQKFEDLSYFAPGVEQADLRKSEIVRDMRLV
Ace3-L	EGNSAESKQYVIQINLQVTFHCLRMVITQKFEDLSYFAPGVEQADLRKSEIVRDMRLV
Ace3-LN	EGNSAESKQYVIQINLQVTFHCLRMVITQKFEDLSYFAPGVEQADLRKSEIVRDMRLV
Ace3-LC	EGNSAESKQYVIQINLQVTFHCLRMVITQKFEDLSYFAPGVEQADLRKSEIVRDMRLV
Ace3-EL	EGNSAESKQYVIQINLQVTFHCLRMVITQKFEDLSYFAPGVEQADLRKSEIVRDMRLV
Ace3-S	MNEAPFWGLQANGEPNVEKIRLIGASLLAI IHRNQDSPLATRARSDFSVLDDILTRLDSK
Ace3-SC	MNEAPFWGLQANGEPNVEKIRLIGASLLAI IHRNQDSPLATRARSDFSVLDDILTRLDSK
Ace3-L	MNEAPFWGLQANGEPNVEKIRLIGASLLAI IHRNQDSPLATRARSDFSVLDDILTRLDSK
Ace3-LN	MNEAPFWGLQANGEPNVEKIRLIGASLLAI IHRNQDSPLATRARSDFSVLDDILTRLDSK
Ace3-LC	MNEAPFWGLQANGEPNVEKIRLIGASLLAI IHRNQDSPLATRARSDFSVLDDILTRLDSK
Ace3-EL	MNEAPFWGLQANGEPNVEKIRLIGASLLAI IHRNQDSPLATRARSDFSVLDDILTRLDSK
Ace3-S	ASDQLRNTSTTVVG (SID NO:3)
Ace3-SC	ASD----- (SID NO:8)
Ace3-L	ASD----- (SID NO:6)
Ace3-LN	ASD----- (SID NO:14)
Ace3-LC	ASDQLRNTSTTVVG (SID NO:10)
Ace3-EL	ASD----- (SID NO:12)

图1B

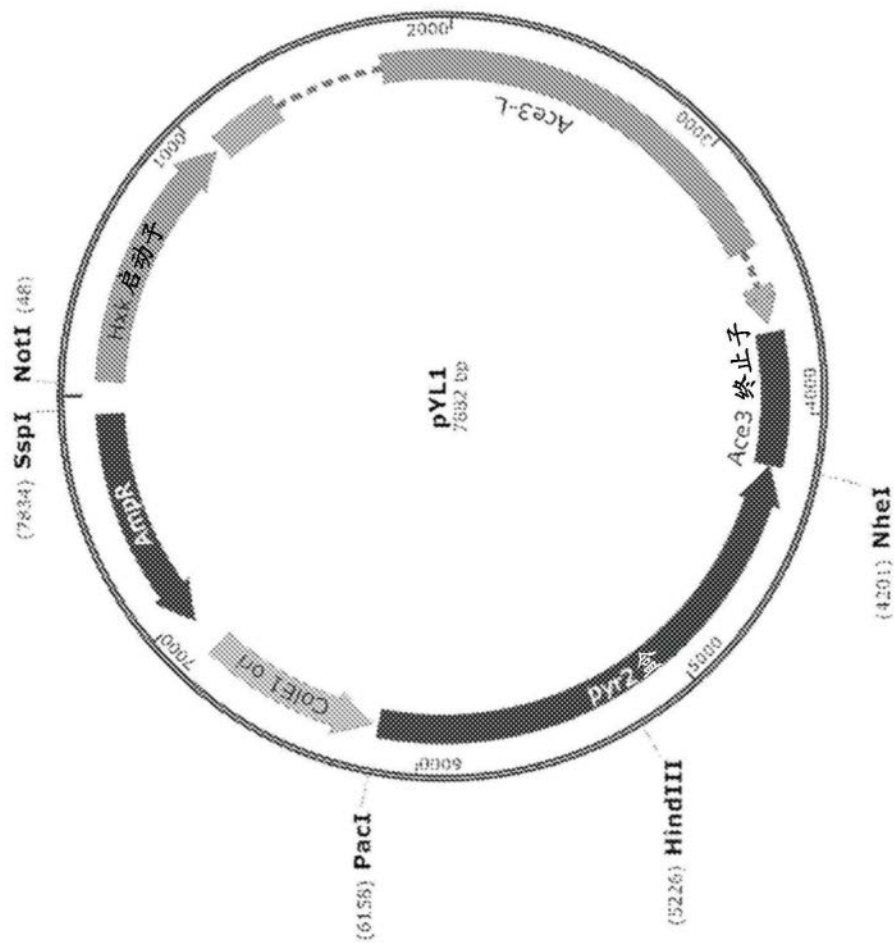


图2A

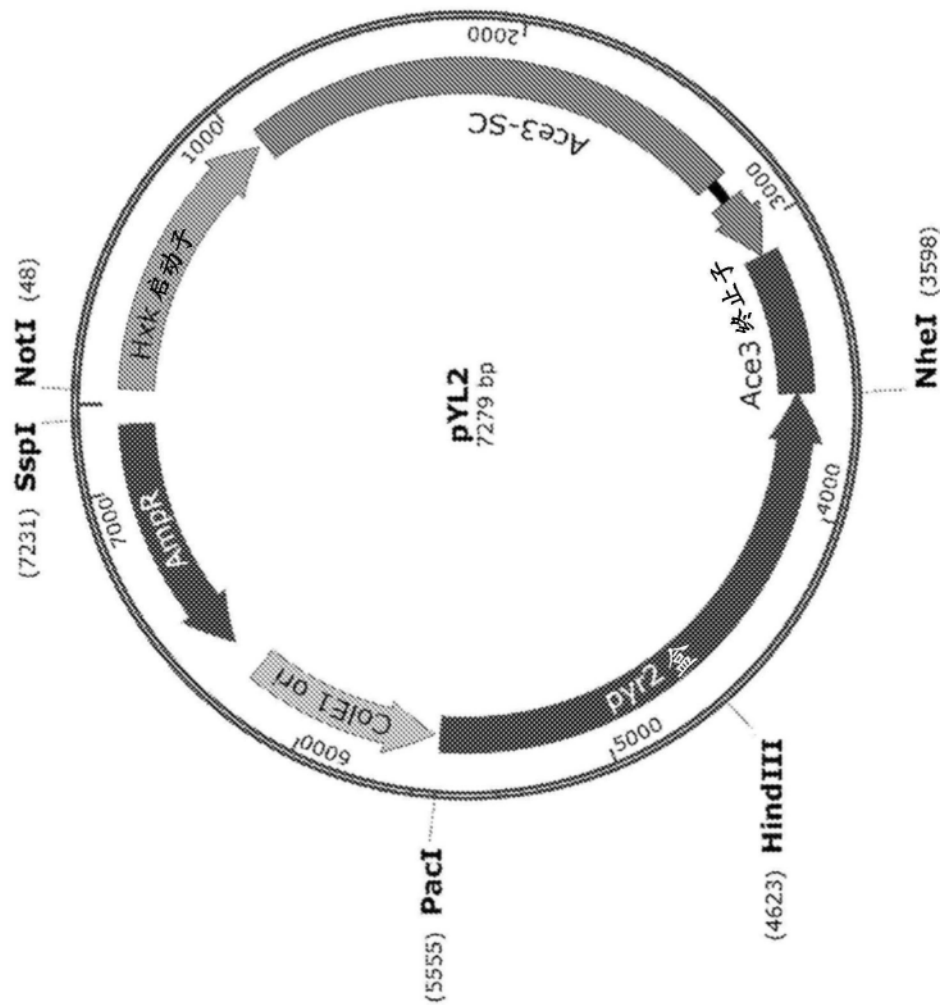


图2B

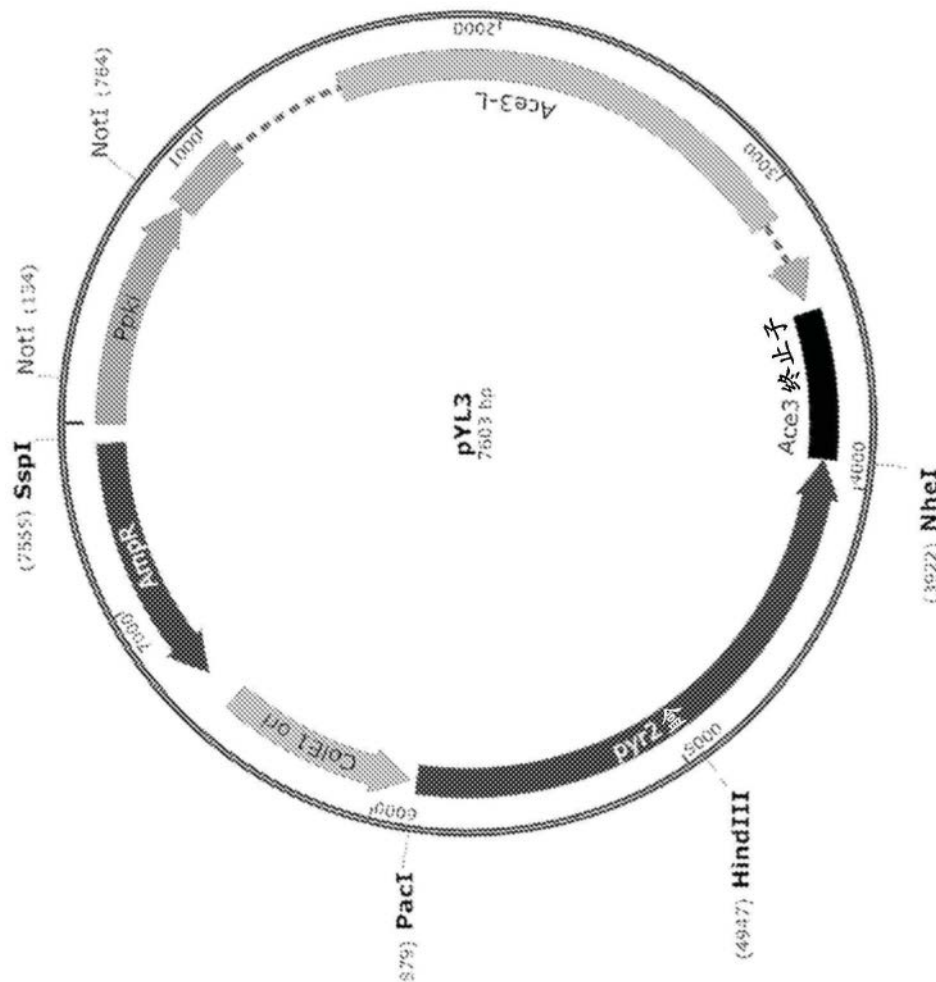


图2C



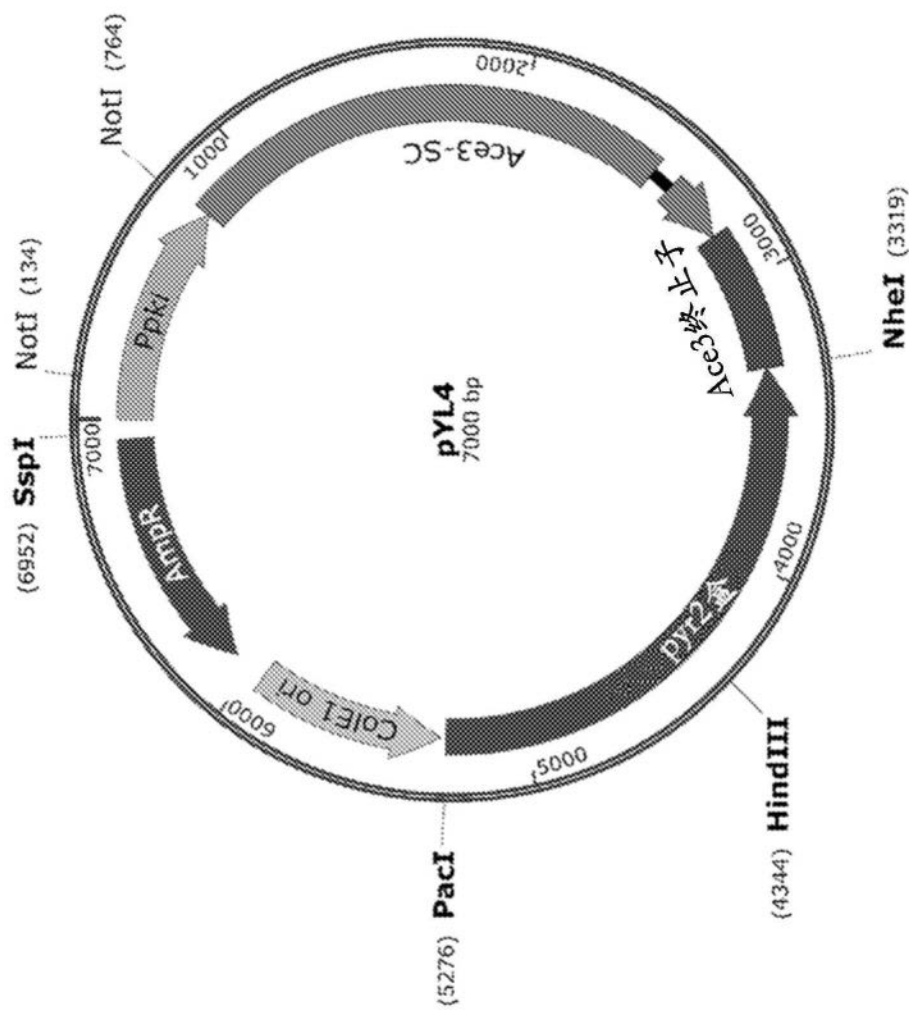


图2D

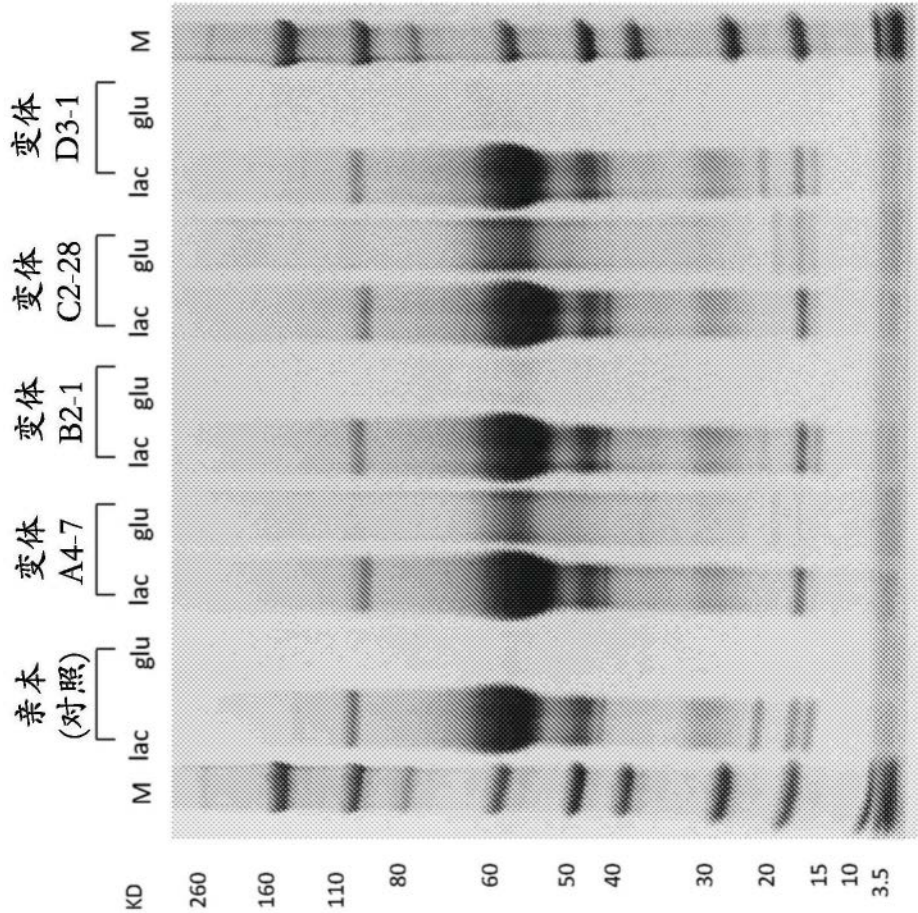


图3

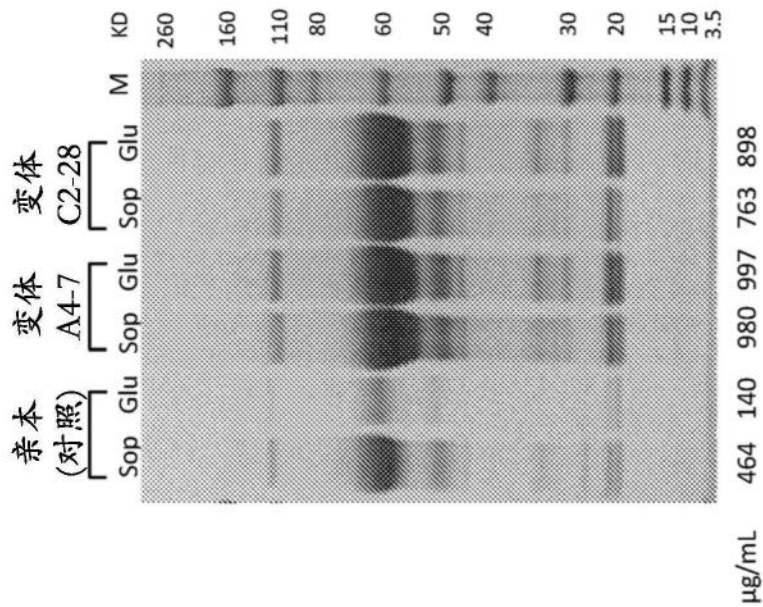
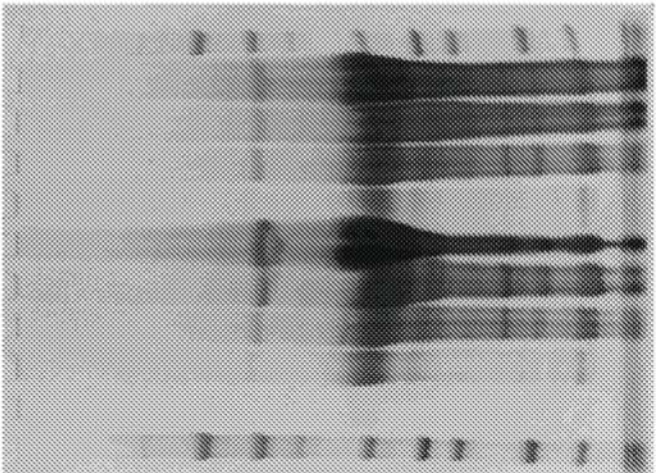
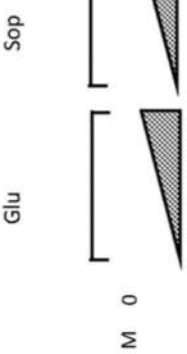


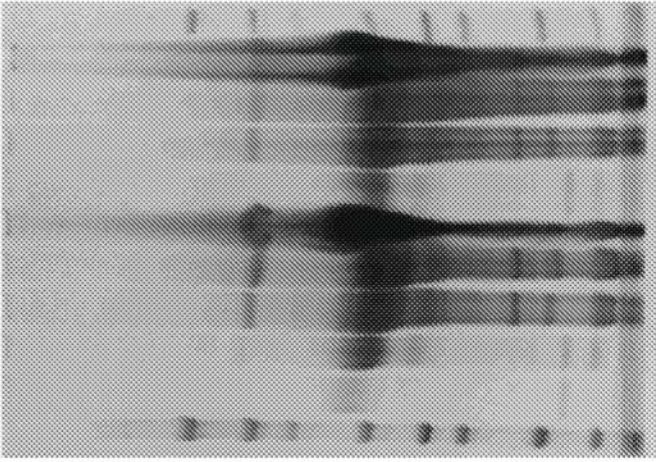
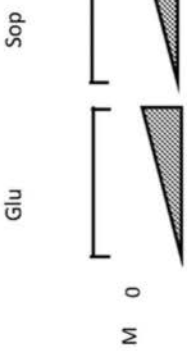
图4



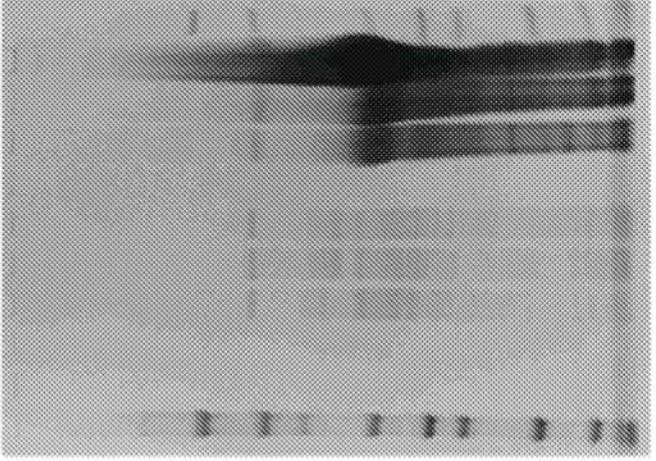
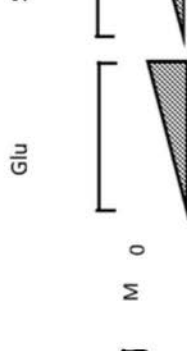
变体 C2-28 细胞



变体 A4-7 细胞



亲本细胞



时间

图5

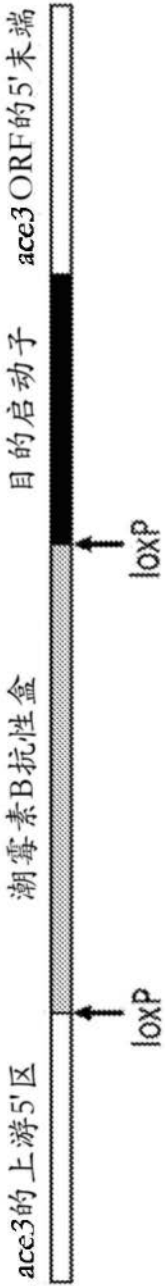


图6



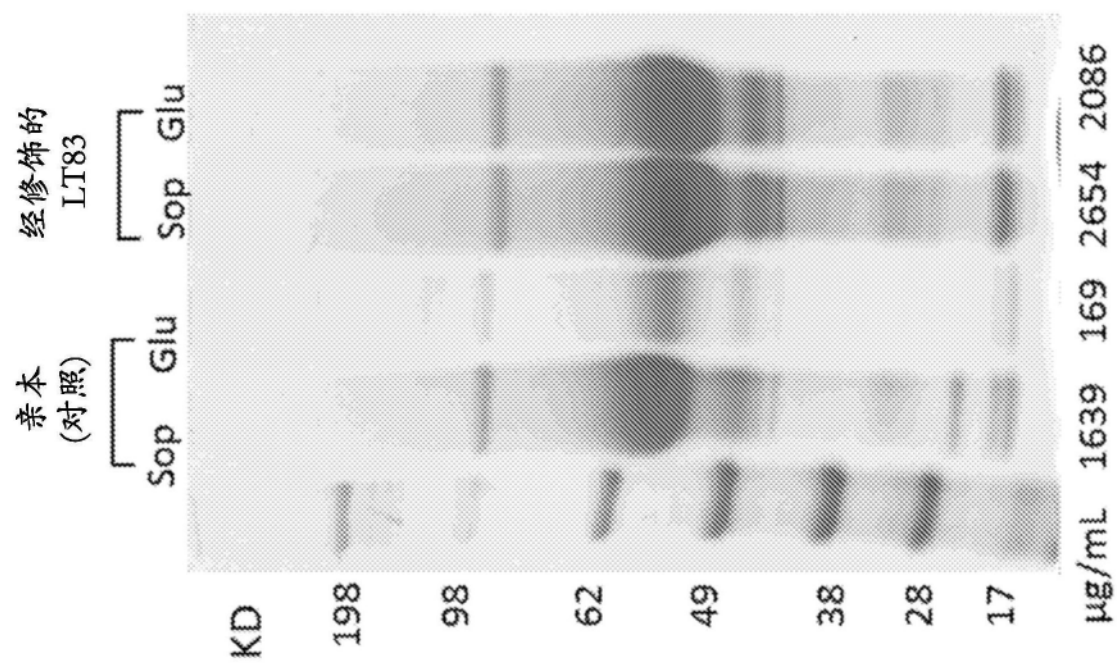


图8

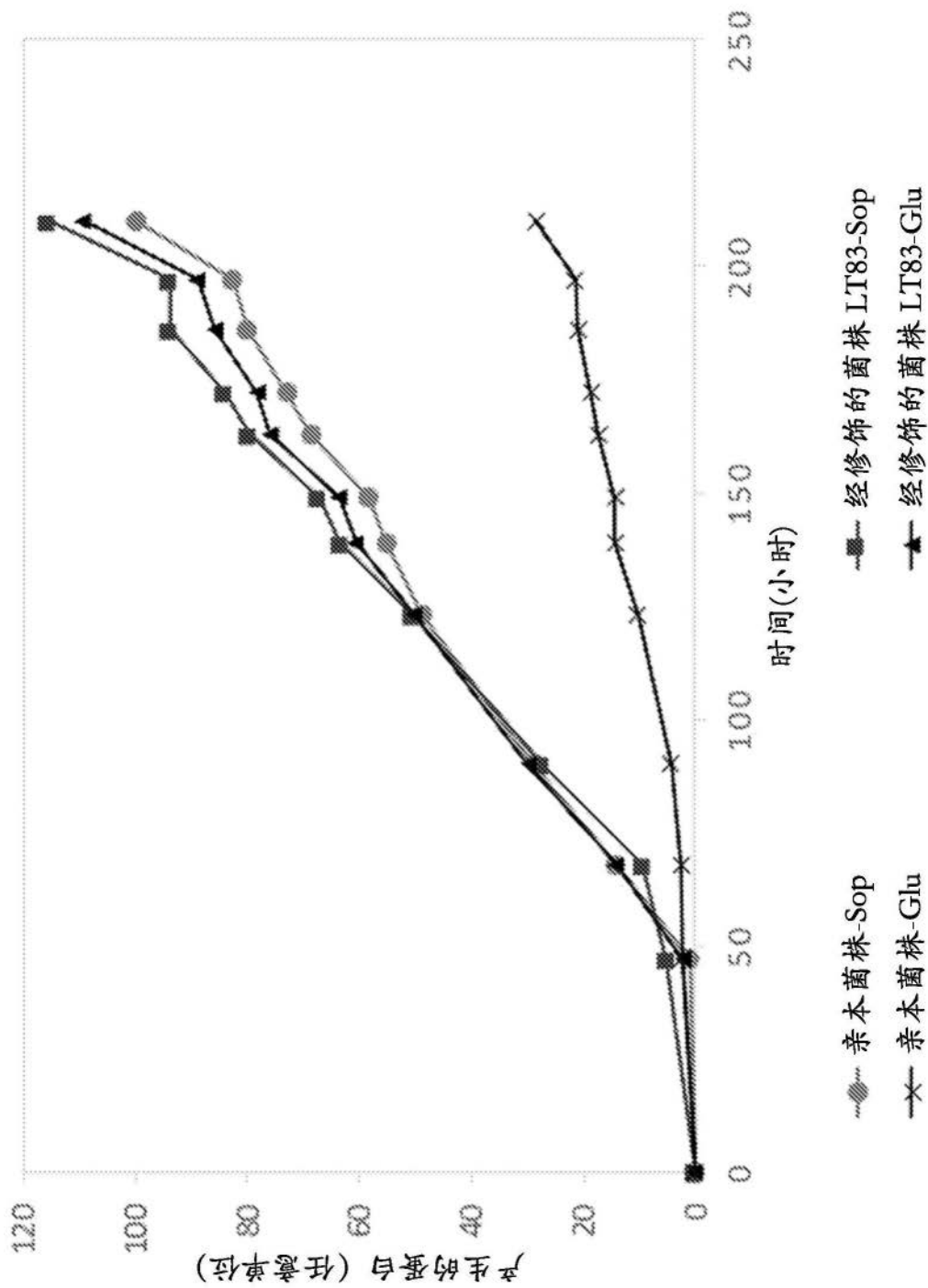


图9

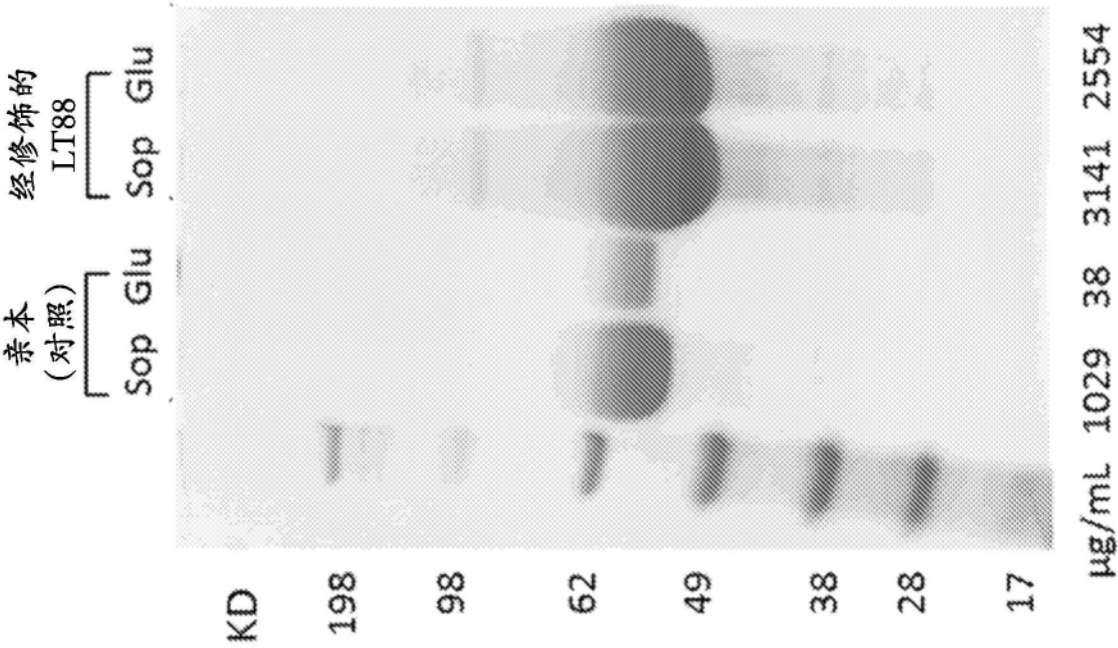


图10



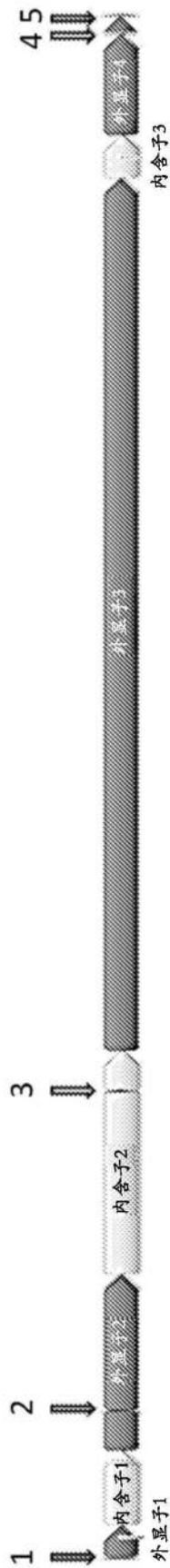


图11

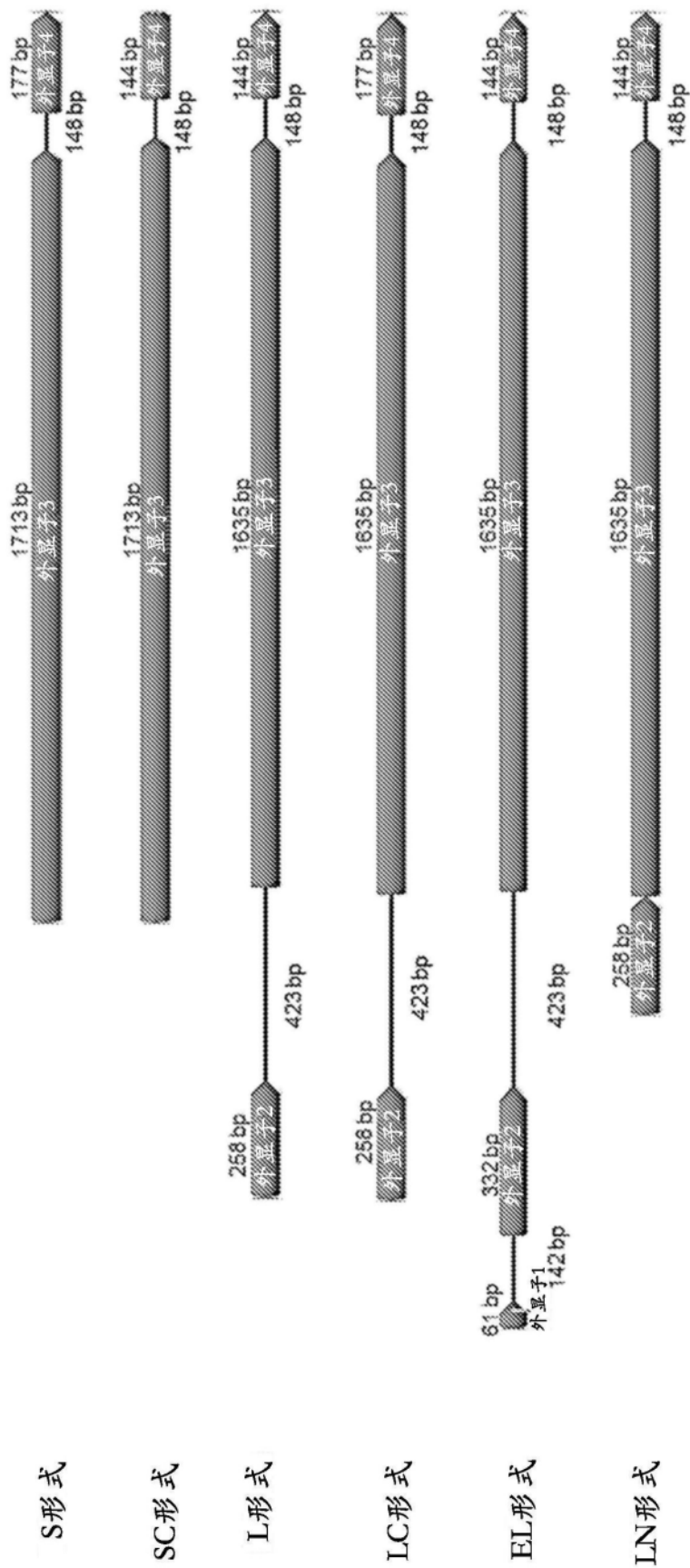


图12



### Ace3 S 形式 (SEQ ID NO: 1)

ATGCTGGCTACTCCCGCGTCTTACACCTGGATACTCTCTCTTGGCACCACTGACCAATGCTCTTCCCGCCCCAAAGTGCAGTACTCTCAGCGGTGTGATAGTGCACGCACTGCCGCGA  
TGCCACGTGCAGTCACTTTCGACCTGCCCTGGCGACGCGGCCCAAGACGAGGAAGAGCGACAGCCCCGGCCAGCCGCTCTGATCCGAGCTCGCTCTCCACCGCGGCTCGAC  
CCGCCAGATGCCGCCCGCGTACCTTCTCCGGCCCCGAGTAGCGCGCTGCAGCCCTTCCGCTCGTCTGCTGCTGCGCCCCGACGCGGCTGGAGCCCCGTCGAGCCGTCAGCATTGAC  
AACGGCTGCCCCGCGACCGCTGGCGACCTGCCCGCTCTCCACCATCCAGAACATCTCAGCGCGCCAGCGATGGATACACCTGGCCACAGCCATGACGTCGCGAACACGACGCTAGA  
GCGGCTCTCGAAGCATGTATCGACCTCTTCTTCGACTACCTCTACCCCTCACCCCTTACGAGCGCGCTTGGGTACGAGCGCGCTTGGCGACGCTCGCATACATCTTCTCCAGCCCTTGCCTGGCG  
TCAACCAACCATCGCCGCTGTACAGCTCAGCGCACCGCGACACCCCTCAAGCTCGCGAGTCTGGGCGCGCTTGGCCAGCCCAAGCGCTCGGCAACCGTCGGCAGC  
AGGCTGGCTCCCTGGCGGACTCGACCTTCAACCTGGTCAGGCGCTTGGCGAGAGGAGCATTCATGCTACCCAAAGACATTTTCCCGAAGAGAAATCCGCTCTCTGAGATCTTGTCTCGA  
AGCTCTGGGACTGCTGCACGACCTCGAGGCGGACTGGAGATCCGACGCGCAACTCGATTGCCATTCTCACTCCAACTGCCCTCCACGCTGGGGGAAGCCCAAGTACT  
CGTGGCACATATTTGGCGAGGCTCCGCTGGCGAGTCACTGAGTGCACGAGGAGTGGCTCGAGGGCTCGTCCCATCGAGGAGAGTTCGCGCTCGCTGCTTTGGATCTCTG  
TACTTGGCGCAAGTCAAGCTATATCAACAATCGGCCCCATCACTCCACAAGTACTGCTTCGACGCGCGCATCACACGCTATACCCGTCGGGTATCGAGGACGAGTTCCTGAGCAC  
GGCTCCGAGCGCGCCGGAAGCTTCATATCCGGCTTCAACGCAATGTGGGCTCTGGAGTCCGCGCTGATTTGCTGTGGAATCCGCGTCTGCAAGATCAGATGATGCAAGCACT  
TTCGAGGACCATGCCCGCCGACCATGTGCTGCCCTCCGCGCACAGGACGATCTCGATTCTCTATGTCGCTTCACTCACCTGCTTGGACGATCTCCGCGCTACCTCCAGTCTGTGCACT  
CTGGCGATGGCAGCATGGCAGGCAACGGGTCTGCCAGTCCAAAGCAGTACGTGATACAGTGCATCAACCTGCAAGTGACGTTTCACTGTCTGCGCATGGTAATACGACAGAAATTCGA  
AGACCTCTTATTTTGTCTGCGTGGAGGCTGATCTCAGAAAGTCGGAGATTGTGCGAGACATGTGAGGTGATGAACGAGCGCGCTTTTGGGCTCGCAGGCAATGGCGAGC  
CAACGAGTGGTCTTCTGCTCTTCTCTTCTTCTGACACCCCTTCTCTCTTATATCCCTGCGGATATGTATATCATCAAGCTCGGCACTTGTGCTTAAT  
**CTGTCTCGATTATGTTGCTGGATGCTGCAGGTTGAAAGATTGCGCTTATCGGAGCTAGTTTGTGCGCATCATCCATCGCAACGAGGATTCAACCTTGGCTACGCGAGCCAGGAGCACT**  
TTTCCGCTGCTTTTGGATATTCTCACGCGGCTGGACTCGAAGGCGTTCGACCAACTGAGGAATACGTCCACTACCGTTGTTGGCTAA

### Ace3 蛋白 S 形式 (SEQ ID NO: 3)

MLRYSFVLHDLTSLPPLTNALPRPKCEYLSAVDSCTHCRDAHVQCTFDLPLARRGPKARKKSDQPGQPPDPSSLSSTAARPGQMPPLTFSGPAVAALQPFASSSLSPPDAWEFVEPL  
SIDNGLPRQPLGDLPLSTIQNISTQRWTHLANAMTLRNTTLERVSKRCIDLFFDYLYPLPLVYEPALRDVLAYIFSQPLPGVNPSPLSQLTPDPTTGTTPFLNAESWAGFGQPSG  
SRTVGSRLAPWADSTFTLVAVCAEAAFMPLPKDIFPEGESVSEILLEASRDCILHQHLEADLENPTANSIAIRYFHSNCLHAAAGPKYSWHIFGEAIRLAQVMQHEEALEGLVPIEAE  
FRRRCFWILYLGDKSAAILNNRPIITHKYCFDAGITTLYPGSGIEDEFSTASEPPKRSFISGFNANVRLWQSAADLLLEIRVLQDMQHFRTGTFPNHVLFSADRQHLDSLIVRFTIC  
LDDLPPYLQSCSTLAWAAMAEAGSAESKQYVIQCINLQVTFHCLRMVITQKFEDLSYFAPGVEQADLRKSEIVRDMRLVMNEAFPFWGLQANGEPNVEKIRLIGASLLAI IHRNQDSPLA  
TRARDFSVLLDILTRLDKASDQLRNTSTTVVG

图14









### Ace3 LN 形式 (SEQ ID NO: 13)

ATGGGCTCAGCAGCTCCGGCCCAAGGCTCTGTAGCTGCAAGCGGCCCTCCAGCTGCTGGCGTGGCGTGGCGCTGCTCCACGCCCTCACCACCTCGCCCGAGTCTGCCTCGGCCCTCGCA  
GCCGGCTCGCCAAACCGCCTCAACCAACGCGCGCAGAACTCACTCTGTCTGGTGTCAACCTCGTTCCACCACCATCCAGAGGCCGCTGTGTGAGCAGAGCCTGCGACCGCTGCCGCCGGCGCA  
AGGCCAAGTGCAGTACCTCAGCGCTGTCGATAGCTGACGCACTGCCCGGATGCCCACTGTCAGTGCACCTTTTGACCTGCCCTGGCGCGACGCGGCCCCAAAGCGAGGAAGAGCGACCCAG  
CCCGCCAGCCGCTCCTGATCCGAGCTCGTCTCCACCGCGGCTCGACCCGGCCAGATGCCCGCGCGCTGACCTTCTCCGGCCCCGCGAGTAGCCGCGCTGCAGCCCTTCGCCCTCGTCTCGTCTCGCT  
GTCCGCCGACCGCGCCTGGGAGCCCTCGAGCCGCTCAGCAATTGACAAAGCGCTGCCCGCGCAGCCGCTGGCGACCTGCCCGCCCTCTCCACCATCCAGAACATCTCGACGCGCGCAGCGATGGA  
TACACTGGCCAAACGCCATGACGCTGCGCAACACGACGCTAGAGCGCTCTCGAAGCGATGATCGACCTCTTCTTCGACTACCTCTACCCCTCACCCTCGTGTGACGAGCCGCCCTCCGG  
GACGTGCTCGCATACATCTTCTCCAGCCCTTGCTGGCGTCAACCAACCATCCCGCTGTACAGCTCAGCGCAGACCCGACCAACCGGACCAACCCCTCAACGCTGCCGAGTCCCGGCCG  
CTTTGGCCAGCCAGCGGCTCGCGAACCGTGGCAGCAGGCTGGCTCCCTGGCGCGACTCGACCTTCAACCTGTGTCAGGCGCTCTGGCAGAGCAGCATTCATGCTACCCCAAGGACATTTTC  
CCGAAGGAGAAATCCGTCTGAGATCTTGTCTGAAGCCTCTCGGGACTGCCCTGCACAGCACTCGAGGCCGACTGGAGAAATCCGACGGCCAACTCGATTGCCATTCCTACTTCCACTCCCAAC  
TGCTCCACGCTCGGGGAAAGCCCAAGTACTCGTGGCACAATTTGGCGAGCCATCGGCTGGCGCAGGTCATGCACTGCACGAGGAGGCTGCCCTCGAGGGGCTCGTCCCATCGAGGCGAGA  
GTTCCCGCTCGCTCTTTTGGATCCTGTACTTTGGCGACAAGTCAAGTCAACAAATCGGCCCATCAACCATCCAGGCTGTGCTTCGACGCGCGCATCACACGCTATACCCGCTCGG  
GTATCGAGGACGAGTTCTGAGCACGCGCTCGAGCGCGCCCGGAGAGCTTCATATCGGCTTCAACGCAATGTGCGGCTCTGGAGTCCGCGCTGATTGCTGCTGGAAATCCGCGCTGCTG  
CAAGATCAGATGATGACGACTTTTCGAGGGGACCATGCCCGGAACCATGTGCTGCCCTCGCCGACAGGCGAGCATCTCGATTCTCTATGTCCGCTTCATCACCTGCTTGAGCGCATCCCGCC  
GTACCTCCAGTCTGCACTCTGGCGATGGCAGGATGGCAGAAAGGCAACGGGTCTGCCGAGTCCAAGCAGTACGTGATACAGTGCATCAACCTGCAGGTGACGTTTCACTGTCTGCGCATGGTAA  
TTACGCAGAAATTCGAAGACCTCTCTATTGCTCCTGGCGTTGACAGGCTGATCTCAGAAAGTCGAGATTTGTGCGAGACATGCTGAGGCTGATGAACGAGGCGCTTTTGGGGCTGCGAG  
GCCAATGGCGAGCCAAA**CGTGAGTCGTTTCTCTCTCTTTTCTGACGACCCCGCTCTCTCTTTTATATCCCTGCGGATATGTA**TATCATCAAGCT**TCGGCACT**  
**TGTTGCTAA**TCTGCTCTGATTA**TGTTGCTGATGCTGCAAGTTGAAAAGATT**CGGAGCTAGTTTGTGCTGGCCATCATCCATCGCAACGAGGATTACCCCTTGGCTACGCGAGCCAGG  
AGCGACTTTTCCGCTTTTGGATATTCTCACGCGGCTGGACTCGAAGGCGTCCGACTAA

### Ace3 蛋白 LN 形式 (SEQ ID NO: 14)

MGSAAQGSVAAAAGPPAAGAGAVHALTTPESASQPGSPASTTTPQNSLVSAATSFHHHPRGLVSRACDRRRKAKCEYLSAVDSCTHCRDAHVQCTFDLPLARRGPKARKSDQ  
FGQPPDPSSLSTAARPGOMPPPLTFSGFAVAALQPFASSLSLSPDAWEPVEPLSIDNGLPRQPLGLDPLGLSTIQNI STORWIIHLANAMTLRNTTLERVSKRCIDLFDYLYPLTFLVYEPALR  
DVLAYIFSQPLPGVNPQPSLSQLTPTDPTTGTTPLNAAESWAGFGQPSGSRVTGSR LAPWADSTFTLVAVCAEAAFMPLKDI FPEGESVSEI LLEASRDCILHQHLEADLENPTANSIAIRYFHSN  
CLHAAGKPKYSHWIFGEAIPLAQVMQIHEEALEGLVPIEAEFRRRCFWILYLGDKSAAILNNRPITIHKYCFDAGITTYLP SGIEDEFLSTASEPPRKSFISGFNANVRLWQSAADLLLEIRVL  
QDQMMQHFRGTMPPNHLPSADRQHLDSLVRFTTCLDDLPYQLQCTLAMAAAEAGNSAESKQYVIQICINLQVTFHCLRMVITQKFEDLSYFAPGVEQADLRKSEIVRDMRLRVNNEAPFWGLQ  
ANGEPNVEKIRLIGASLLAI IHRNQDSPLATRARSDFSLLDILTRLDKASD

图18



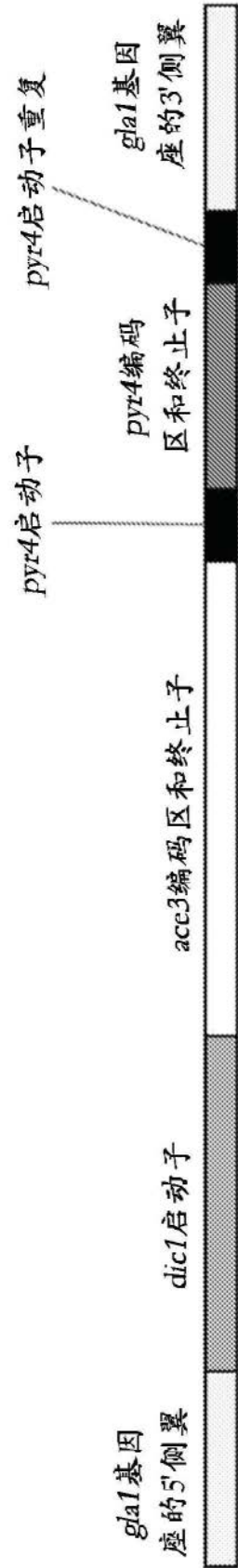


图19

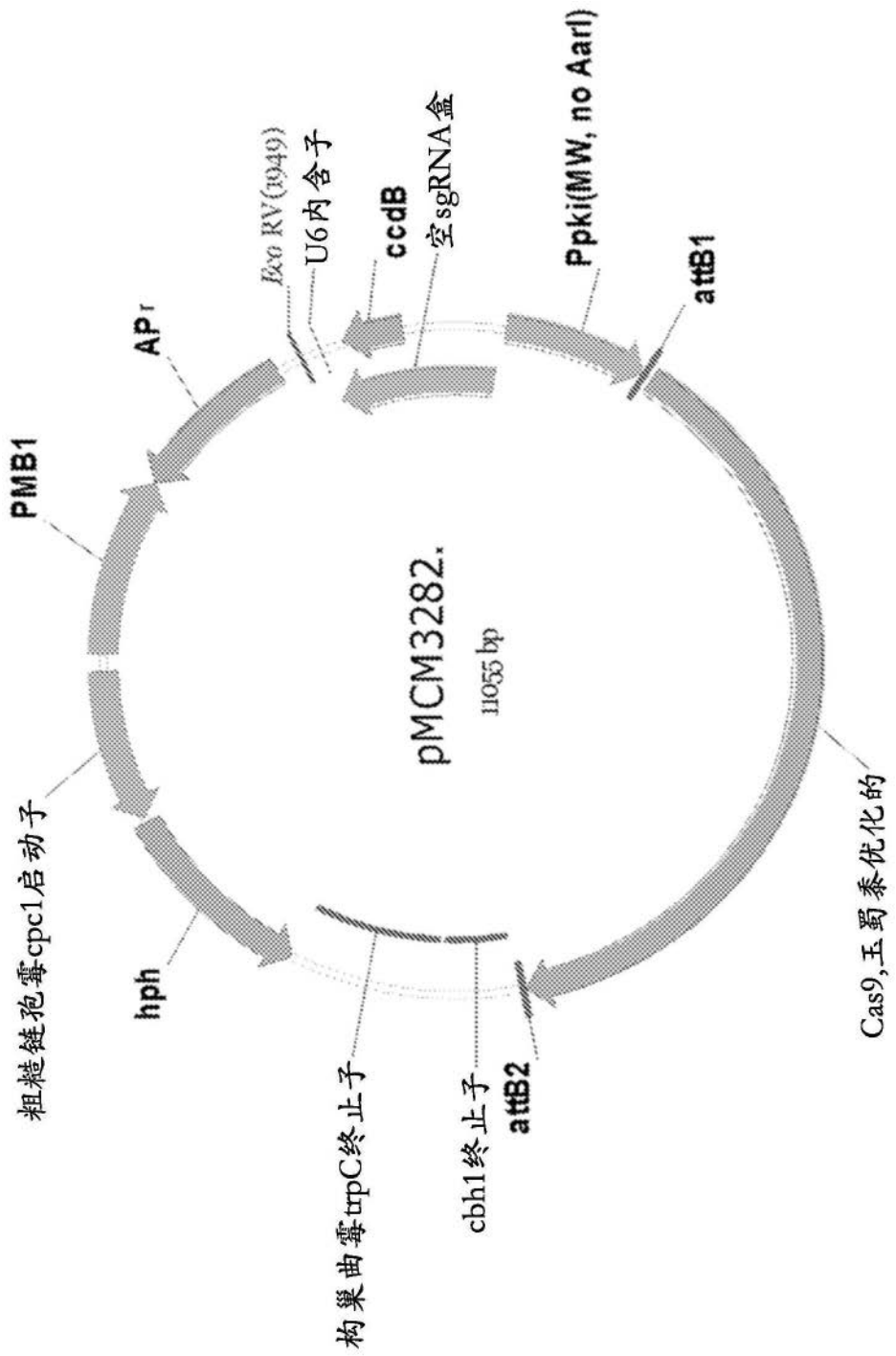


图20

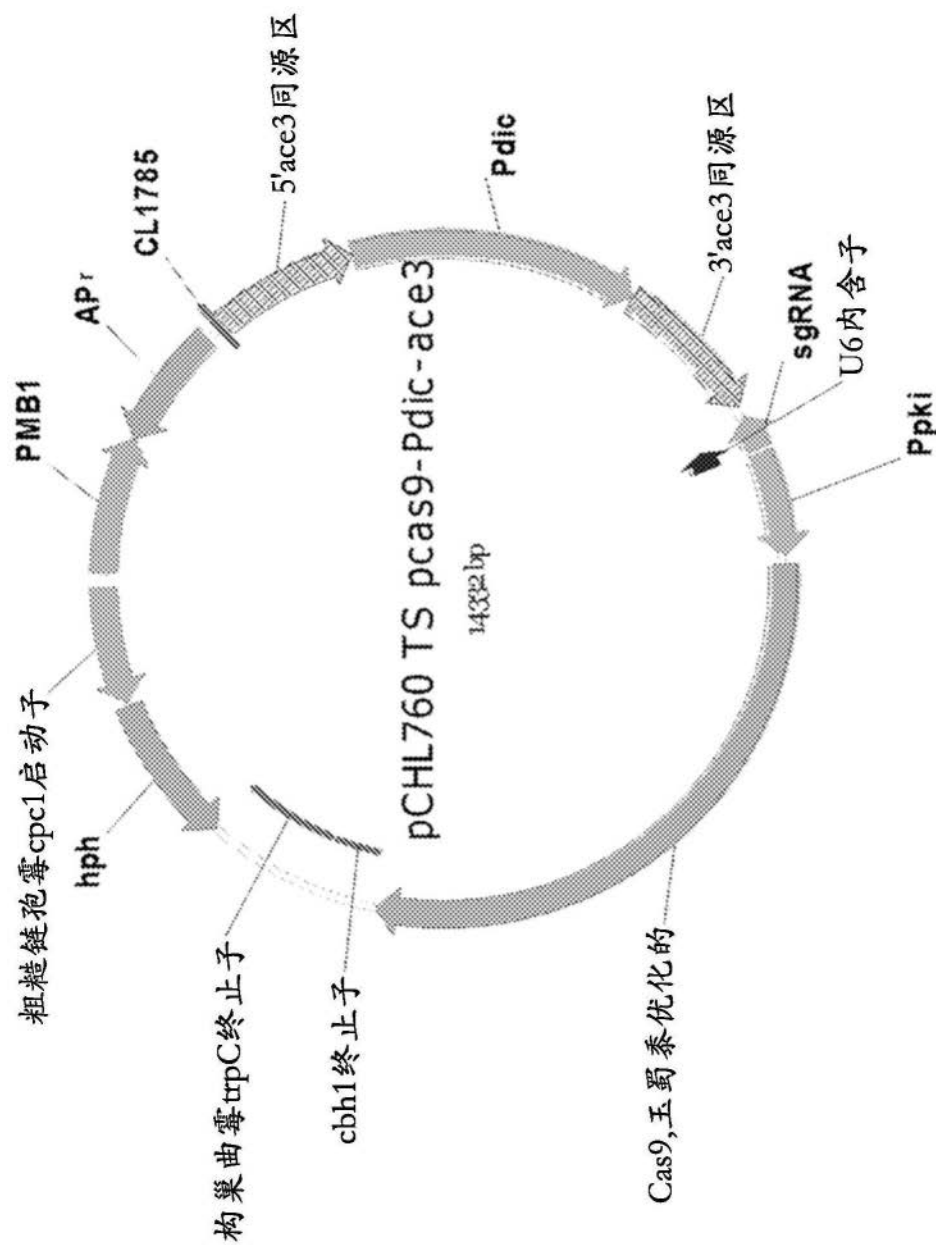


图21

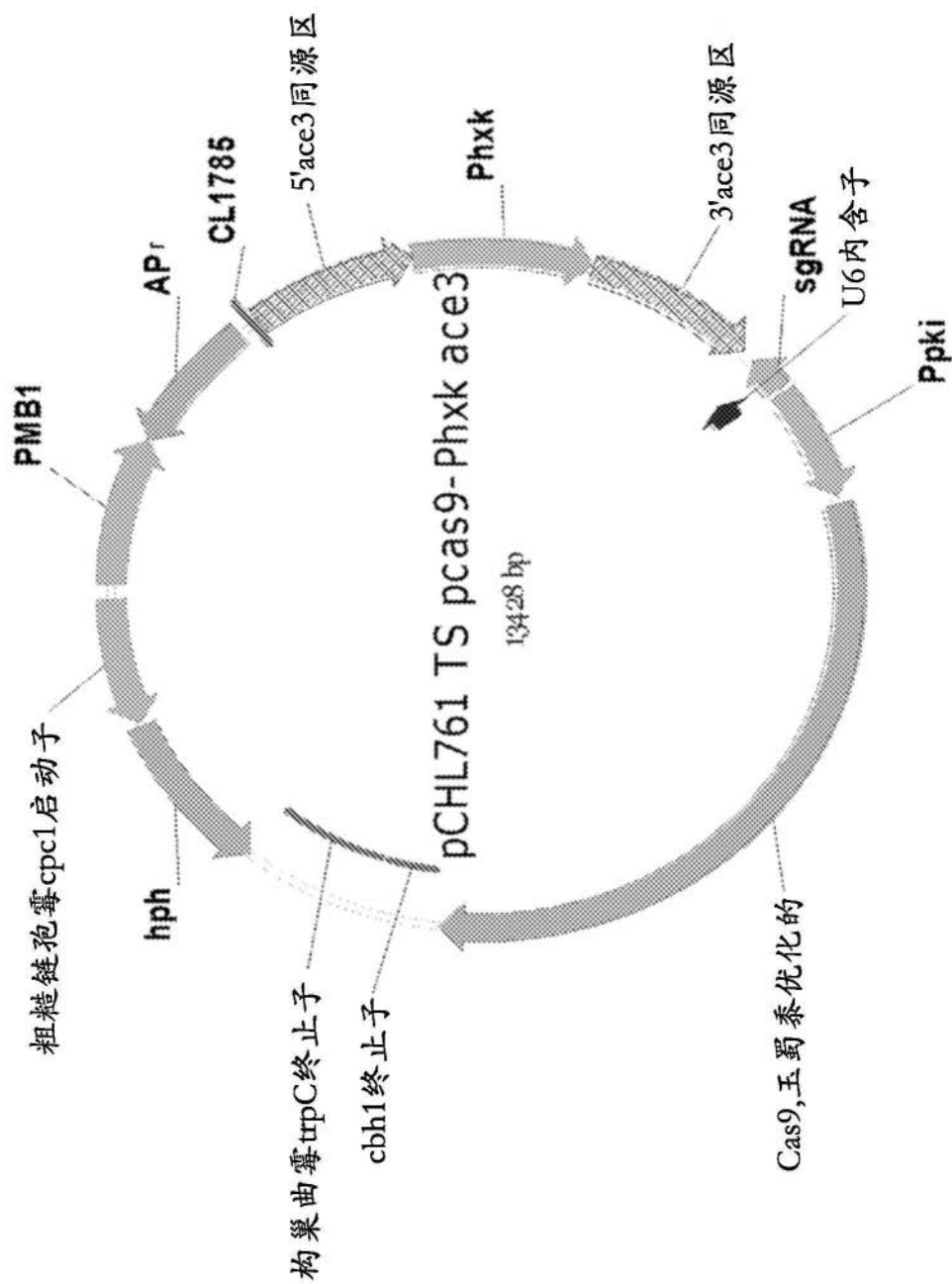


图22

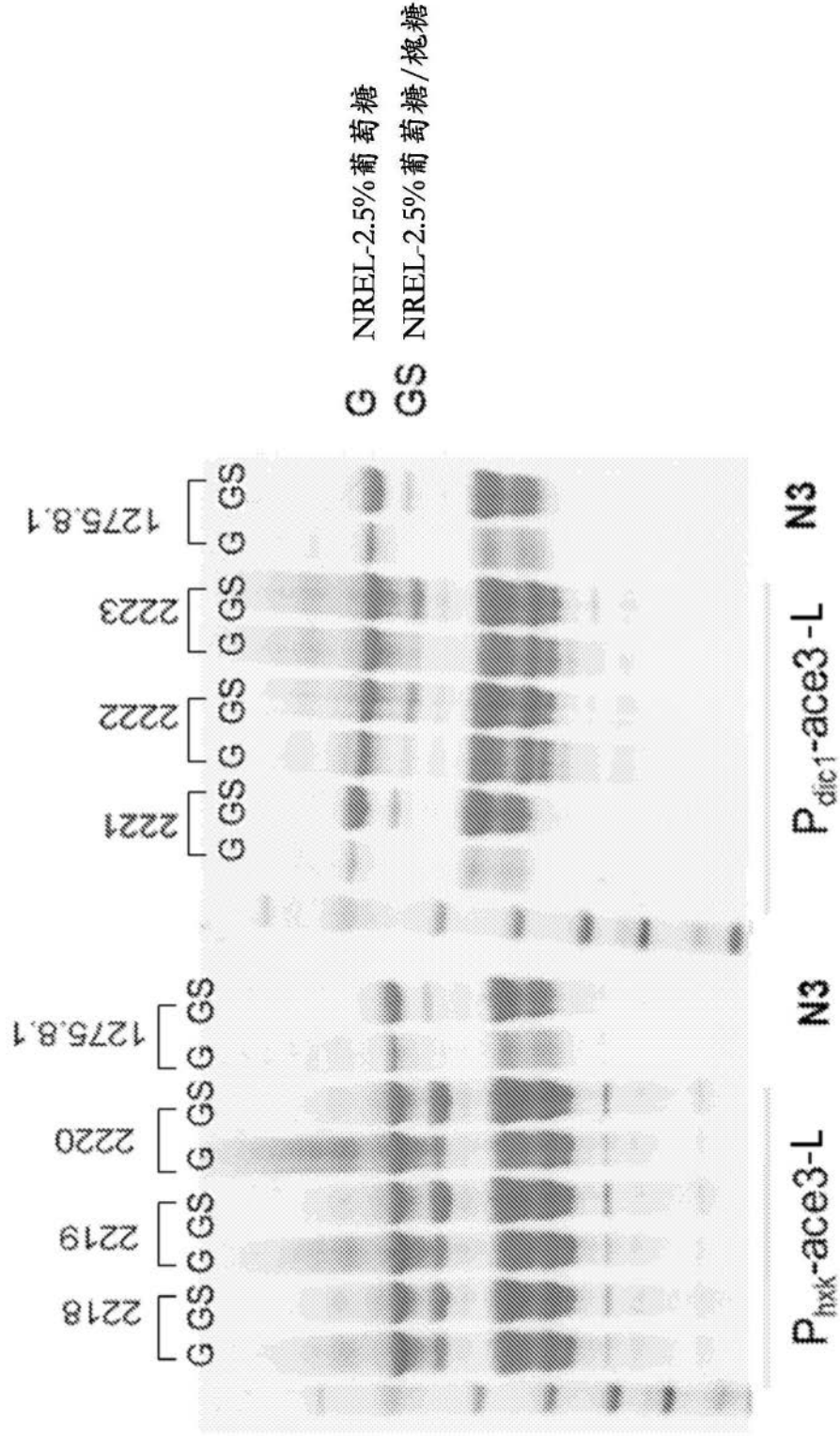


图23