



(12)发明专利申请

(10)申请公布号 CN 109661402 A

(43)申请公布日 2019.04.19

(21)申请号 201780012235.4

J·H·维瑟

(22)申请日 2017.02.22

(74)专利代理机构 北京市中咨律师事务所

(30)优先权数据

62/298,351 2016.02.22 US

11247

代理人 柴云峰 黄革生

(85)PCT国际申请进入国家阶段日

2018.08.20

(51)Int.Cl.

C07K 14/37(2006.01)

G12P 21/02(2006.01)

(86)PCT国际申请的申请数据

PCT/US2017/018901 2017.02.22

C12N 1/15(2006.01)

(87)PCT国际申请的公布数据

WO2017/147163 EN 2017.08.31

(71)申请人 丹尼斯科美国公司

地址 美国加利福尼亚州

(72)发明人 J·A·布鲁克霍夫 J·P·迈纳恩

C·B·迈克尔斯

I·A·范德布格特 J·维瑟

权利要求书3页 说明书36页

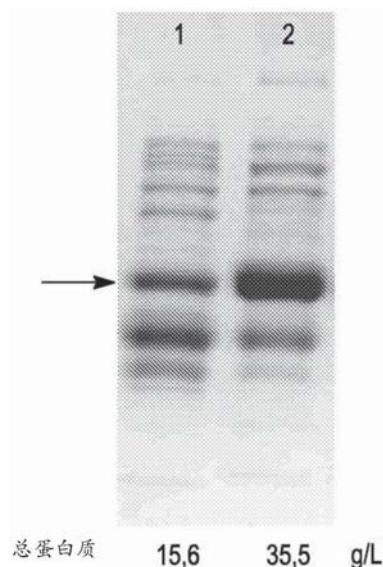
序列表55页 附图7页

(54)发明名称

真菌的高水平蛋白质生产系统

(57)摘要

提供了具有增加的蛋白质生产的细胞,其特征在于所述细胞包含经修饰的SUMO化;用于产生此类细胞或表达系统的方法;以及此类细胞在生产目的蛋白质中的用途。



1. 一种具有增加的蛋白质生产的经修饰的细胞,其中所述细胞已被修饰以导致改变的SUMO化。

2. 如权利要求1所述的经修饰的细胞,其中该蛋白质生产的增加是与缺乏SUMO化修饰的亲本细胞的总蛋白质生产相比,整体蛋白质生产增加至少1.1倍。

3. 如权利要求1所述的经修饰的细胞,其中该蛋白质生产的增加是与由缺乏SUMO化修饰的亲本细胞生产所述蛋白质相比,目的蛋白质生产增加至少1.1倍。

4. 如权利要求1-3中任一项所述的经修饰的细胞,其中该细胞已通过遗传修饰进行修饰以导致改变的SUMO化。

5. 如权利要求4所述的经修饰的细胞,其中该遗传修饰是点突变、插入和/或缺失。

6. 如权利要求4或权利要求5所述的经修饰的细胞,其中该遗传修饰是靶向遗传修饰。

7. 如权利要求1-6中任一项所述的经修饰的细胞,其中SUMO化已通过对编码SUMO化机器的内源性蛋白质的至少一个基因的遗传修饰而被修饰。

8. 如权利要求7所述的经修饰的细胞,其中该遗传修饰是表达调控序列或编码序列的修饰。

9. 如前述权利要求中任一项所述的经修饰的细胞,其中SUMO化已通过减少Ubc9生产和/或活性来修饰。

10. 如权利要求9所述的经修饰的细胞,其中Ubc9已经被位于ubc9基因的启动子序列中的基因破坏来修饰。

11. 如权利要求9或10中任一项所述的经修饰的细胞,其中该ubc9基因具有与SEQ ID NO:1[嗜热毁丝霉(*M. thermophila*)的Ubc9]具有至少50%序列同一性的野生型对应物。

12. 如权利要求9或10所述的经修饰的细胞,其中该ubc9基因具有编码蛋白质的野生型对应物,该蛋白质包含与以下序列基序的一个或多个具有至少75%序列同一性的序列:RLQEERKQWRKDHPFGF (SEQ ID NO:42) 和KPPKCKFTPPLFHPNVYPSGTVCLSIL (SEQ ID NO:43)。

13. 如权利要求1-12中任一项所述的经修饰的细胞,其中该细胞是真菌细胞,优选是丝状真菌细胞,更优选是选自以下组成的组的属的真菌细胞:金孢子菌属(*Chrysosporium*)、梭孢壳属(*Thielavia*)、脉孢菌属(*Neurospora*)、短梗霉属(*Aureobasidium*)、线黑粉菌属(*Filibasidium*)、单鞭毛菌属(*Piromyces*)、棒囊壳属(*Corynascus*)、隐球菌属(*Cryptococcus*)、支顶孢属(*Acremonium*)、弯颈霉属(*Tolyposcladium*)、柱顶孢霉(*Scytalidium*)、裂褶菌属(*Schizophyllum*)、孢子丝菌属(*Sporotrichum*)、青霉菌属(*Penicillium*)、赤霉菌属(*Gibberella*)、毁丝霉属(*Myceliophthora*)、毛霉菌属(*Mucor*)、曲霉属(*Aspergillus*)、镰刀菌属(*Fusarium*)、腐质霉属(*Humicola*)、木霉属(*Trichoderma*)、踝节菌属(*Talaromyces*)和*Rasamsonia*。

14. 如权利要求13所述的经修饰的细胞,其中该细胞是或衍生自嗜热毁丝霉。

15. 如权利要求14所述的经修饰的细胞,其中该细胞是或衍生自选自以下组成的组的菌株:C1(在登录号VKM F-3500D下保藏在俄罗斯科学院的全俄微生物保藏中心的国际保藏机构)、UV18-25(在登录号VKM F-3631D下保藏于VKM)、UV18#100f(在登录号CBS122188下保藏于CBS)、W1L(在登录号CBS122189下保藏于CBS)和W1L#100L(在登录号CBS122190下保藏于CBS的菌株)。

16. 如权利要求13所述的经修饰的细胞,其中该细胞是或衍生自里氏木霉

(*Trichoderma reesei*)。

17. 如权利要求13所述的经修饰的细胞,其中该细胞是或衍生自黑曲霉 (*Aspergillus niger*)。

18. 如权利要求1-17中任一项所述的经修饰的细胞,其中该细胞进一步包含编码至少一种目的蛋白质的外源表达构建体。

19. 如权利要求12所述的经修饰的细胞,其中该目的蛋白质是异源蛋白质。

20. 如权利要求18或19中任一项所述的经修饰的细胞,其中该目的蛋白质是乙酰酯酶、氨基肽酶、淀粉酶、阿拉伯糖酶 (arabinase)、阿拉伯呋喃糖苷酶、羧肽酶、过氧化氢酶、纤维素酶、几丁质酶、凝乳酶、角质酶、脱氧核糖核酸酶、差向异构酶、酯酶、 α -半乳糖苷酶、 β -半乳糖苷酶、 α -葡聚糖酶、葡聚糖裂合酶、内切- β -葡聚糖酶、葡糖淀粉酶、葡萄糖氧化酶、 α -葡糖苷酶、 β -葡糖苷酶、葡萄糖醛酸酶、半纤维素酶、己糖氧化酶、水解酶、转化酶、异构酶、漆酶、脂肪酶、裂合酶、甘露糖苷酶、氧化酶、氧化还原酶、果胶酸裂合酶、果胶乙酰酯酶、果胶解聚酶、果胶甲酯酶、果胶分解酶、过氧化物酶、酚氧化酶、植酸酶、聚半乳糖醛酸酶、蛋白酶、鼠李糖-半乳糖醛酸酶、核糖核酸酶、奇异果甜蛋白、转移酶、转运蛋白、转谷氨酰胺酶、木聚糖酶、己糖氧化酶、其功能片段、或其一种或多种的混合物。

21. 如权利要求18或19中任一项所述的经修饰的细胞,其中该目的蛋白是肽激素、生长因子、凝血因子、趋化因子、细胞因子、淋巴因子、抗体、受体、粘附分子、微生物抗原、其功能片段、或其一种或多种的混合物。

22. 如权利要求18或19中任一项所述的经修饰的细胞,其中所述蛋白质是用于降解木质纤维素材料的酶或其活性片段。

23. 如权利要求22所述的经修饰的细胞,其中所述酶是纤维二糖水解酶、木聚糖酶、内切葡聚糖酶、 β -葡糖苷酶、 β -木糖苷酶、辅酶、葡糖淀粉酶、 α -淀粉酶、 α -葡糖苷酶、植酸酶、蛋白酶、氨基肽酶、甘露聚糖酶、漆酶、过氧化氢酶、葡萄糖或己糖氧化酶、寡糖氧化酶、脂肪酶、或其一种或多种的混合物。

24. 一种用于产生与没有被修饰的亲本细胞相比具有增加的蛋白质生产的经修饰的细胞的方法,其中所述方法包括将至少一个遗传修饰引入编码SUMO化蛋白质的基因中的步骤。

25. 如权利要求24所述的方法,其中该遗传修饰是点突变、插入和/或缺失。

26. 如权利要求24或25所述的方法,其中该遗传修饰是靶向遗传修饰。

27. 如权利要求24-26中任一项所述的方法,其中所得遗传修饰是表达调控序列或编码序列的修饰。

28. 如权利要求27所述的方法,其中该遗传修饰导致减少的Ubc9生产和/或活性。

29. 如权利要求28所述的方法,其中该遗传修饰涵盖位于ubc9基因的启动子序列中的基因破坏。

30. 如权利要求28或29中任一项所述的方法,其中该ubc9基因具有与SEQ ID NO:1具有至少50%序列同一性的野生型对应物。

31. 如权利要求28或29中任一项所述的方法,其中该ubc9基因具有编码蛋白质的野生型对应物,该蛋白质包含与以下序列基序的一个或多个具有至少75%序列同一性的序列: RLQEERKQWRKDHPFGF (SEQ ID NO:42) 和KPPKCKFTPLFHPNVYPSGTVCLSIL (SEQ ID NO:43)。

32. 如权利要求24-31中任一项所述的方法,其中该细胞是真菌细胞,优选是丝状真菌细胞,更优选是选自由以下组成的组的属的真菌细胞:金孢子菌属、梭孢壳属、脉孢菌属、短梗霉属、线黑粉菌属、单鞭毛菌属、棒囊壳属、隐球菌属、支顶孢属、弯颈霉属、柱顶孢霉、裂褶菌属、孢子丝菌属、青霉菌属、赤霉菌属、毁丝霉属、毛霉菌属、曲霉属、镰刀菌属、腐质霉属、木霉属、踝节菌属和Rasamsonia。

33. 如权利要求24-32中任一项所述的方法,其中该细胞是或衍生自嗜热毁丝霉。

34. 如权利要求24-31中任一项所述的方法,其中该细胞是或衍生自选自由以下组成的组的菌株:C1(在登录号VKM F-3500D下保藏在俄罗斯科学院的全俄微生物保藏中心的国际保藏机构)、UV18-25(在登录号VKM F-3631 D下保藏于VKM)、UV18#100f(在登录号CBS122188下保藏于CBS)、W1L(在登录号CBS122189下保藏于CBS)和W1L#100L(在登录号CBS122190下保藏于CBS)。

35. 如权利要求24-32中任一项所述的方法,其中该细胞是或衍生自里氏木霉。

36. 如权利要求24-32中任一项所述的方法,其中该细胞是或衍生自黑曲霉。

37. 一种用于产生表达系统的方法,该方法包括以下步骤:用编码至少一种目的蛋白质的至少一种外源表达构建体转导如权利要求1-23中任一项所述的经修饰的细胞、或通过如权利要求24-36中任一项所述的方法获得的经修饰的细胞。

真菌的高水平蛋白质生产系统

[0001] 相关申请的交叉引用

[0002] 本申请要求于2016年2月22日提交的美国临时申请号62/298,351的权益,该临时申请通过引用以其全文结合在此。

技术领域

[0003] 本发明属于能够生产高水平的蛋白质的蛋白质生产系统、更具体地真菌的生产系统的领域。本发明是基于SUMO化的修饰导致蛋白质生产增加的发现。本发明使这些突变的真菌系统适合工业中的蛋白质生产。

[0004] 序列表

[0005] 与本文一起以电子方式提交的处于ASCII文本形式的序列表(文件名:NB36013WOPCT_SEQLIST_ST25.txt;大小143,201字节;以及创建日期:2017年2月22日)的内容形成说明书的一部分并且通过引用以其全文结合在此。

背景技术

[0006] 在工业应用中使用的酶的生产是不断增长的市场,这是由从基于化石的经济转向基于生物技术的经济的需求所驱动的。对这些酶的日益增长的需求使酶生产的成本成为重要的花费。酶生产成本的降低要求探索新颖的酶,连同用于高产量生产过程的可靠方法。为了建立成本效益好的酶生产过程,高水平的蛋白质生产和分泌是关键要求。

[0007] 真菌已被用作用于生产各种酶的宿主。属于但不限于金孢子菌属(*Chrysosporium*)、梭孢壳属(*Thielavia*)、踝节菌属(*Talaromyces*)、嗜热真菌属(*Thermomyces*)、热子囊菌属(*Thermoascus*)、脉孢菌属(*Neurospora*)、短梗霉属(*Aureobasidium*)、*Filivasidium*、单鞭毛菌属(*Piromyces*)、棒囊壳属(*Corynascus*)、隐球菌属(*Cryptococcus*)、支顶孢属(*Acremonium*)、弯颈霉属(*Tolyposcaldium*)、柱顶孢霉属(*Scytalidium*)、裂褶菌属(*Schizophyllum*)、孢子丝菌属(*Sporotrichum*)、青霉菌属(*Penicillium*)、赤霉菌属(*Gibberella*)、毁丝霉属(*Myceliophthora*)、毛霉菌属(*Mucor*)、曲霉属(*Aspergillus*)、镰刀菌属(*Fusarium*)、腐质霉属(*Humicola*)和木霉属(*Trichoderma*)的菌株加其无性型和有性型已经被用于大范围的酶的工业生产中。已经开发了在发酵液中分泌高至100g/L或更多蛋白质的菌株,参见例如Visser等人(Development of a mature fungal technology and production platform for industrial enzymes based on a *Myceliophthora thermophila* isolate, previously known as *Chrysosporium lucknowense* C1 [开发基于嗜热毁丝霉分离物,以前称为卢克诺文思金孢子菌(*Chrysosporium lucknowense*) C1的成熟真菌技术和工业酶生产平台], *Industrial Biotechnology* [工业生物技术], 2011.7(3):第214-223页),以及欧洲专利申请号2408910。这些真菌的蛋白质分泌能力使它们成为用于特定酶或酶混合物靶向生产的有吸引力的宿主。

发明内容

[0008] 本文提供了具有增加的蛋白质生产的经修饰的细胞。提供的是已被修饰以导致改变的SUMO化的经修饰的细胞。与缺乏SUMO化修饰的亲本细胞的总蛋白质生产相比,蛋白质生产的增加可以是整体蛋白质生产增加至少1.1倍。与由缺乏SUMO化的修饰的亲本细胞来生产所述蛋白质相比,蛋白质生产的增加可以是目的蛋白质生产增加至少1.1倍。

[0009] 本文提供的是经修饰的细胞,其中SUMO化已通过对编码SUMO化机器的内源性蛋白质的至少一个基因进行遗传修饰而被修饰。遗传修饰可以是点突变、插入和/或缺失。遗传修饰可以是靶向修饰。遗传修饰可以是表达调控序列或编码序列的修饰。在实施例中,通过减少Ubc9产生和/或活性来修饰SUMO化。在实施例中,通过位于ubc9基因的编码序列或位于启动子序列中的基因破坏来修饰Ubc9。在实施例中,该ubc9基因具有与SEQ ID NO:1[嗜热毁丝霉(*M. thermophila*)]具有至少50%序列同一性的野生型对应物。在实施例中,ubc9基因具有编码包含PFAM结构域PF00179的蛋白质的野生型对应物。在实施例中,ubc9基因具有编码包含与SEQ ID NO:42或43的一个或多个序列基序具有至少75%同一性的序列的蛋白质的野生型对应物。

[0010] 本文提供的经修饰的细胞可以是真菌细胞,如丝状真菌细胞。经修饰的细胞可以是选自由以下组成的组的属的真菌细胞:金孢子菌属、梭孢壳属、脉孢菌属、短梗霉属、线黑粉菌属(*Filibasidium*)、单鞭毛菌属、棒囊壳属、隐球菌属、支顶孢属、弯颈霉属(*Tolyposcladium*)、柱顶孢霉、裂褶菌属、孢子丝菌属、青霉菌属、赤霉菌属、毁丝霉属、毛霉菌属、曲霉属、镰刀菌属、腐质霉属、木霉属、踝节菌属和*Rasamsonia*。在实施例中,该细胞是或衍生自嗜热毁丝霉(*Myceliophthora thermophila*)。在实施例中,该细胞是或衍生自选自由以下组成的组的菌株:C1(在登录号VKM F-3500D下保藏在俄罗斯科学院的全俄微生物保藏中心的国际保藏机构)、UV18-25(在登录号VKM F-3631D下保藏于VKM)、UV18#100f(在登录号CBS122188下保藏于CBS)、W1L(在登录号CBS122189下保藏于CBS)和W1L#100L(在登录号CBS122190下保藏于CBS的菌株)。在实施例中,该细胞是嗜热毁丝霉细胞。在实施例中,该细胞不是衍生自C1、或不是衍生自UV18-25、或不是衍生自UV18#100f、或不是衍生自W1L、或不是衍生自W1L#100L。在实施例中,该细胞不是衍生自W1L#100.1 Δ alp1 Δ chil Δ pyr5。在实施例中,该细胞是或衍生自木霉属,并且在实施例中该细胞是或衍生自里氏木霉(*Trichoderma reesei*)。在实施例中,该细胞是或衍生自曲霉属,并且在实施例中,该细胞是或衍生自黑曲霉(*Aspergillus niger*)。

[0011] 本文提供的经修饰的细胞可以进一步包含编码至少一种目的蛋白质的外源表达构建体。在实施例中,该目的蛋白质是异源蛋白质。在实施例中,该目的蛋白质是乙酰酯酶、氨基肽酶、淀粉酶、阿拉伯糖酶(arabinase)、阿拉伯呋喃糖苷酶、羧肽酶、过氧化氢酶、纤维素酶、几丁质酶、凝乳酶、角质酶、脱氧核糖核酸酶、差向异构酶、酯酶、 α -半乳糖苷酶、 β -半乳糖苷酶、 α -葡聚糖酶、葡聚糖裂合酶、内切 β -葡聚糖酶、葡糖淀粉酶、葡萄糖氧化酶、 α -葡糖苷酶、 β -葡糖苷酶、葡萄糖醛酸酶、半纤维素酶、己糖氧化酶、水解酶、转化酶、异构酶、漆酶、脂肪酶、裂合酶、甘露糖苷酶、氧化酶、氧化还原酶、果胶酸裂合酶、果胶乙酰酯酶、果胶解聚酶、果胶甲酯酶、果胶分解酶、过氧化物酶、酚氧化酶、植酸酶、聚半乳糖醛酸酶、蛋白酶、鼠李糖-半乳糖醛酸酶、核糖核酸酶、奇异果甜蛋白、转移酶、转运蛋白、转谷氨酰胺酶、木聚糖酶、己糖氧化酶、其功能片段、或其一种或多种的混合物。在实施例中,该目的蛋白质

是肽激素、生长因子、凝血因子、趋化因子、细胞因子、淋巴因子、抗体、受体、粘附分子、微生物抗原、其功能片段、或其一种或多种的混合物。在实施例中，该蛋白质是用于降解木质纤维素材料的酶或其活性片段。因此，在一些实施例中，该目的蛋白质不是用于降解木质纤维素材料的酶或其活性片段。在实施例中，该目的蛋白质是纤维二糖水解酶、木聚糖酶、内切葡聚糖酶、 β -葡糖苷酶、 β -木糖苷酶、辅酶、葡糖淀粉酶、 α -淀粉酶、 α -葡糖苷酶、植酸酶、蛋白酶、氨肽酶、甘露聚糖酶、漆酶、过氧化氢酶、葡萄糖或己糖氧化酶、寡糖氧化酶、脂肪酶、或其一种或多种的混合物。

[0012] 本文还提供了用于生产与未经修饰的亲本细胞相比具有增加的蛋白质生产的经修饰的细胞的方法，该方法包括在编码SUMO化蛋白质的基因中引入至少一个遗传修饰。在实施例中，该遗传修饰是点突变、插入和/或缺失。在实施例中，该遗传修饰是靶向的。在实施例中，所得遗传修饰是表达调控序列或编码序列的修饰。在实施例中，表达调控序列是启动子序列。在实施例中，该遗传修饰导致减少的Ubc9生产和/或活性。在实施例中，该遗传修饰涵盖位于ubc9基因的启动子序列中的基因破坏。在实施例中，ubc9基因具有野生型对应物，该野生型对应物与SEQ ID NO:1[嗜热毁丝霉的ubc9]的启动子区域或编码区或两者具有至少50%、至少60%、至少70%、至少80%、至少85%、至少80%、至少95%、至少99%、或100%序列同一性。在实施例中，ubc9基因具有野生型对应物，该野生型对应物包含与SEQ ID NO:47或48具有至少50%、至少60%、至少70%、至少80%、至少85%、至少80%、至少95%、至少99%、或100%序列同一性的序列。在实施例中，ubc9基因具有野生型对应物，该野生型对应物包含与SEQ ID NO:45的启动子或编码部分具有至少50%、至少60%、至少70%、至少80%、至少85%、至少80%、至少95%、至少99%、或100%序列同一性的序列。

[0013] 本文还提供了用于产生表达系统的方法，该方法包括转导本文提供的经修饰的细胞或通过本文提供的方法获得的经修饰的细胞的步骤。在实施例中，该细胞用编码至少一种目的蛋白质的至少一种外源表达构建体转导。在实施例中，该至少一种目的蛋白质与该细胞是异源的。

[0014] 还提供了用于产生目的蛋白质的方法，该方法包括培养本文提供的经修饰的细胞、或者本文提供的或通过本文提供的方法获得的经修饰的细胞或表达系统的步骤。

[0015] 本文提供的是通过本文公开的经修饰的细胞产生的细胞肉汤、连同包含本文公开的经修饰的细胞或细胞肉汤的组合物。此类组合物可以任选地进一步包含底物，该底物包含木质纤维素材料。

[0016] 本文还提供了本文公开的经修饰的细胞在用于产生至少一种目的蛋白质的方法中的用途。在实施例中，该目的蛋白质通过异源表达构建体编码。在实施例中，该目的蛋白质与该细胞是异源的。

附图说明

[0017] 图1. 以下的最终发酵液的SDS-PAGE分析:

[0018] -W1L#100.1 Δ alp1 Δ chi1 Δ pyr5[eg2/pyr5] (泳道1); 和

[0019] -W1L#100.1 Δ alp1 Δ chi1 Δ pyr5[eg2/pyr5]ubc9- (泳道2)。

[0020] 每个泳道的蛋白质含量(约1g/L)的相等上样。箭头指示Eg2蛋白质。

[0021] 图2. 以下的微量滴定板肉汤的SDS-PAGE分析:

- [0022] -W1L#100.1 Δ alp1 Δ chi1 Δ pyr5[Twbx11/pyr5] (泳道1); 和,
- [0023] -W1L#100.1 Δ alp1 Δ chi1 Δ pyr5[TwBx11/pyr5] ubc9- (泳道2)
- [0024] TwBx11蛋白质由箭头指示。
- [0025] 图3. 以下的摇瓶培养物的SDS-PAGE分析:
- [0026] -W1L#100.1 Δ alp1 Δ chi1 Δ pyr5[AnPGII/pyr5]; 和,
- [0027] -W1L#100.1 Δ alp1 Δ chi1 Δ pyr5[AnPGII/pyr5] ubc9-
- [0028] 箭头指示AnPGII蛋白质。
- [0029] 图4. 在发酵期间的UV18-25和UV18-25ubc9-的总蛋白质浓度, 时间(h)。
- [0030] 图5. 在有或者没有将Bg11添加至酶混合物中的情况下, 在NREL PCS底物上孵育UV18-25和UV18-25ubc9-的最终发酵液68h之后的葡萄糖释放。
- [0031] 图6. 植酸酶活性测量证明在C1菌株中破坏ubc9基因启动子的作用。
- [0032] 图7. pH937的图表, 描述在实例8中。
- [0033] 图8. 来自以下的Ubc9蛋白质序列的多重序列对齐: 嗜热毁丝霉 (“Mycthe”; SEQ ID NO:2); 粗糙脉孢菌 (*N. crassa*) (“Neurcra”; SEQ ID NO:9); 构巢曲霉 (*A. nidulans*) (“Aspnid”; SEQ ID NO:11); 酿酒酵母 (*S. cerevisiae*) (“Saccer”; SEQ ID NO:15); 新型隐球菌 (*C. neoformans*) (“Cryneo”; SEQ ID NO:13); 以及保守基序 (基序1/共有区1: SEQ ID NO:42; 基序2: SEQ ID NO:43; 共有区2: SEQ ID NO:44)。
- [0034] 表4提供了SEQ ID No 1-7的概述。
- [0035] 生物保藏
- [0036] 本文提及的以下材料在“布达佩斯条约”下保藏, 该条约是关于专利程序目的中微生物保藏的国际规定, 并附有指定的保藏机构。

[0037]

菌株描述	保藏机构	保藏地址	保藏日期	登录号
嗜热毁丝霉 C1	微生物 RAS 的生物化学和生理学研究所 (IBFM)	普希诺 (Pushchino), 莫斯科地区, 142292, 俄罗斯	8/29/1996	VKM F-3500D
嗜热毁丝霉 UV18-25	微生物 RAS 的生物化学和生理学 G.K. Skryabin 研究所 (IBFM)	普希诺 (Pushchino), 莫斯科地区, 142290, 俄罗斯	9/2/1998	VKM F-3631D
嗜热毁丝霉 UV13-6	微生物 RAS 的生物化学和生理学 G.K. Skryabin 研究所 (IBFM)	普希诺 (Pushchino), 莫斯科地区, 142290, 俄罗斯	9/2/1998	VKM F-3632 D
嗜热毁丝霉 NG7C-19	微生物 RAS 的生物化学和生理学 G.K. Skryabin 研究所 (IBFM)	普希诺 (Pushchino), 莫斯科地区, 142290, 俄罗斯	9/2/1998	VKM F-3633 D
嗜热毁丝霉 UV18#100.f	真菌菌种保藏中心 (Centraal bureau voor Schimmelcultures)	Uppsalalaan 8 P.O. Box 85167 3508 AD 乌特勒支 (UTRECHT) 荷兰	12/5/2007	CBS 122188
嗜热毁丝霉 W1L	真菌菌种保藏中心 (Centraal bureau voor	Uppsalalaan 8 P.O. Box 85167 3508 AD 乌特勒支	12/5/2007	CBS 122189

[0038]

	Schimmelcultures)	(UTRECHT), 荷兰		
嗜热毁丝霉 W1L#100.1	真菌菌种保藏中心 (Centraal bureau voor Schimmelcultures)	Uppsalalaan 8 P.O. Box 85167 3508 AD 乌特勒支 (UTRECHT), 荷兰	12/5/2007	CBS 122190

[0039] 发明详述

[0040] 如本文所用,除非另有说明,提及百分比(%)同一性是指使用以下方式进行的同源性评估:1)使用采用标准默认参数的针对氨基酸检索的blastp和针对核酸检索的blastn进行BLAST 2.0Basic BLAST同源性检索,其中对于低复杂度区域,查询序列被默认过滤(描述于Altschul,S.F.,Madden,T.L.,Schäffer,A.A.,Zhang,J.,Zhang,Z.,Miller,W.和Lipman,D.J.(1997)“Gapped BLAST and PSI-BLAST:a new generation of protein database search programs.[空位BLAST和PSI-BLAST:新一代蛋白质数据库检索程序]”Nucleic Acids Res.[核酸研究]25:3389-3402);2)BLAST 2对齐(使用下面描述的参数);3)采用标准默认参数的PSI-BLAST(位置特异性迭代BLAST);和/或4)使用来自碳水化合物活性酶数据库的标准默认参数而确定的CAZy同源性(Coutinho,P.M.和Henrissat,B.(1999)Carbohydrate-active enzymes:an integrated database approach.[碳水化合物活性酶:综合数据库方法],在“Recent Advances in Carbohydrate Bioengineering[碳水化合物生物工程的最新进展]”中,H.J.Gilbert,G.Davies,B.Henrissat和B.Svensson编辑,The Royal Society of Chemistry[皇家化学学会],剑桥,第3-12页)和/或使用数据库如Foly数据库(网址:folly.esil.univ-mrs.fr)和PeroxiBase(网址:peroxibase.isb-sib.ch)而应用类似的策略。

[0041] 应注意,由于BLAST 2.0Basic BLAST和BLAST 2之间的标准参数的一些差异,可以使用BLAST 2程序将两个特异序列识别为具有显著的同源性,而使用这两个序列中的一个作为查询序列在BLAST 2.0Basic BLAST中进行的检索可能不识别顶部匹配中的第二个序列。此外,PSI-BLAST提供了一个自动的、易于使用的“轮廓”检索版本,这是查找序列同源物或变体的敏感方法。该程序首先执行空位BLAST数据库检索。PSI-BLAST程序使用返回的任何重要对齐的信息来构建特定位置的得分矩阵,该得分矩阵替换下一轮数据库检索的查询序列。因此,应当理解,通过使用这些程序中的任一个可以确定百分比同一性。

[0042] 可以使用如在Tatusova和Madden,(1999)，“Blast 2sequences—a new tool for comparing protein and nucleotide sequences[Blast 2序列-用于比较蛋白质和核苷酸序列的新工具]”，FEMS Microbiol Lett.[FEMS微生物通讯]174:247-250中描述的BLAST 2序列来将两个特异序列彼此对齐。使用BLAST 2.0算法在blastp或blastn中进行BLAST 2序列对齐,以在两个序列之间执行空位BLAST检索(BLAST 2.0),从而允许在所得对齐中引入空位(缺失和插入)。本文出于清楚的目的,使用如下的标准默认参数进行BLAST 2序列对

齐。

[0043] 对于blastn,使用0 BLOSUM62矩阵:

[0044] 匹配加分=1

[0045] 错配罚分=-2

[0046] 起始空位(5)和延伸空位(2)罚分

[0047] 空位x_下降(50)期望(10)字号(11)过滤器(开)

[0048] 对于blastp,使用0BLOSUM62矩阵:

[0049] 起始空位(11)和延伸空位(1)罚分

[0050] 空位x_下降(50)期望(10)字号(3)过滤(开)。

[0051] 除非本文另有说明,否则用给定的SEQ ID NO的同一性表示基于所述序列的全长和连续长度(即,其整个长度或整体)的同一性。

[0052] 如本文所用,关于本文所述的核酸或氨基酸序列,术语“连续(contiguous或consecutive)”是指以不间断的顺序连接。例如,对于第一序列包含第二序列的30个连续(contiguous或consecutive)氨基酸意味着第一序列包括与第二序列中的30个氨基酸残基的不间断序列100%同一的30个氨基酸残基的不间断序列。类似地,对于第一序列与第二序列具有“100%同一性”或“100%同一”意味着第一序列与第二序列精确匹配,在核苷酸或氨基酸之间没有空位。

[0053] 编码如本文所公开的蛋白质的核酸分子是指编码该蛋白质的核酸链的核苷酸序列。应当理解,编码给定氨基酸序列的双链DNA包含单链DNA及其互补链,该互补链具有与单链DNA互补的序列。如此,核酸分子可以是双链或单链的,并且包括那些互补链。推导互补序列的方法是本领域技术人员已知的。应当注意,由于核酸测序技术不是完全无错误的,因此本文中呈现的序列最多表示本文所披露的蛋白质的表现序列。

[0054] 如本文所用,参考杂交条件是指在使用核酸分子鉴定相似的核酸分子情况下的标准杂交条件。例如在Sambrook等人,Molecular Cloning:A Laboratory Manual[分子克隆:实验室手册],Cold Spring Harbor Labs Press[冷泉港实验室出版社],1989,Sambrook等人,同上(具体参见第9.31-9.62页)中公开了这样的标准条件。此外,例如,在Meinkoth等人,1984,Anal.Biochem[生物化学年鉴].138,267-284;Meinkoth等人,同上中公开了计算适当的杂交和洗涤条件以实现允许不同程度的核苷酸错配的杂交的公式。

[0055] 更具体地,如本文提到的中度严格杂交和洗涤条件是指允许与用于探测杂交反应的核酸分子具有至少约70%的核苷酸序列同一性的核酸分子分离的条件(即,允许约30%或更少的核苷酸错配的条件)。如本文提到的高严格杂交和洗涤条件是指允许与用于探测杂交反应的核酸分子具有至少约80%的核苷酸序列同一性的核酸分子分离的条件(即,允许约20%或更少的核苷酸错配的条件)。如本文提到的非常高严格杂交和洗涤条件是指允许与用于探测杂交反应的核酸分子具有至少约90%的核苷酸序列同一性的核酸分子分离的条件(即,允许约10%或更少的核苷酸错配的条件)。如上所讨论,本领域技术人员可以使用Meinkoth等人(同上)的公式来计算适当的杂交和洗涤条件以达到这些特定水平的核苷酸错配。这样的条件将会有所不同,这取决于是否正在形成DNA:RNA或DNA:DNA杂交体。DNA:DNA杂交体的计算的解链温度比DNA:RNA杂交体低10°C。优选地,DNA:DNA杂交体的严格杂交条件包括在6X SSC(0.9M Na⁺)的离子强度下,在约20°C至约35°C(较低严格)之间,更优选

地在约28°C至约40°C(更严格)之间,并且甚至更优选地在约35°C至约45°C(甚至更严格)之间的温度进行杂交,伴随适当的洗涤条件。优选地,DNA:RNA杂交体的严格杂交条件包括在6X SSC(0.9M Na⁺)的离子强度下,在约30°C至约45°C之间,更优选地在约38°C至约50°C之间,并且甚至更优选地在约45°C至约55°C之间的温度进行杂交,伴随类似严格的洗涤条件。这些值基于对大于约100个核苷酸、0%甲酰胺和G+C含量约40%的分子的解链温度(T_m)的计算。可替代地,可以根据经验计算T_m,如在Sambrook等人,同上,第9.31至9.62页中所述。通常,洗涤条件应尽可能严格,并且应适合所选择的杂交条件。例如,杂交条件可以包括比特定杂交体的计算的T_m低约20°C-25°C的盐和温度条件的组合,并且洗涤条件通常包括比该特定杂交体的计算的T_m低约12°C-20°C的盐和温度条件的组合。适合与DNA:DNA杂交体一起使用的杂交条件的一个实例包括在约42°C、在6X SSC(50%甲酰胺)中杂交2小时-24小时,随后是洗涤步骤,其包括一次或多次在室温在约2X SSC中洗涤,然后在较高温度和较低的离子强度下进行另外的洗涤(例如,在约37°C、在约0.1X-0.5X SSC中至少一次洗涤,然后在约68°C、在约0.1X-0.5X SSC中至少一次洗涤)。

[0056] 提及基因包括与天然(即野生型)基因相关的所有核苷酸序列,如控制由该基因编码的蛋白质的产生的调节区(例如但不限于转录、翻译或翻译后控制区域)以及编码区本身。基因可以包括或排除一个或多个内含子或其任何部分或任何其他序列,基因不包括在该蛋白质的cDNA中。当核酸分子包含如上所述的基因时,术语“核酸分子”和“基因”可以互换使用。

[0057] 修饰的基因包括通过取代、插入和/或缺失单个或多个核苷酸序列而修饰的天然基因,这些修饰的基因可以存在于包括编码多肽的区域的外显子的编码序列内,或存在于侧翼区,如通常在上游(例如,启动子、增强子和相关序列)、下游(例如,转录终止和聚(A)信号)或者影响多肽或目的调节分子的转录、翻译和/或活化的内部区域(例如,内含子)的调节区域。例如,多肽的活化可以需要去除一个或多个N-末端、C-末端或内部多肽区域,和/或需要特定氨基酸残基的翻译后修饰,如通过可以改变酶的靶向、降解、催化活性的糖基化、酰胺化等。

[0058] 使用重组DNA技术(例如,聚合酶链式反应(PCR)扩增、克隆等)或化学合成可以产生如本文所公开的核酸分子。核酸修饰可以使用本领域技术人员已知的多种方法产生(参见例如,Sambrook等人)。例如,核酸分子可以使用多种技术来修饰,包括但不限于通过经典诱变和重组DNA技术(例如,定点诱变、化学处理、限制酶切割、核酸片段的连接和/或PCR扩增)、或寡核苷酸混合物的合成和混合物组的连接以“构建”核酸分子的混合物及其组合。用于修饰编码蛋白质的重组核酸分子的另一种方法是基因改组(即,分子育种)(参见例如,Stemmer的美国专利号5,605,793,通过引用结合在此;Minshull和Stemmer;1999, Curr.Opin.Chem.Biol[化学生物学现行观点]3:284-290;Stemmer,1994,P.N.A.S.USA[美国科学院院报]91:10747-10751)。该技术可以用于有效地在该蛋白质中引入多个同时发生的变化。

[0059] 如本文所披露的核酸分子可以是重组核酸分子,其包含与至少一个表达控制序列有效地连接地上述核酸分子。更具体地,重组核酸分子通常包含重组载体和如本文所述的任何一种或多种核酸分子。如本文所用,重组载体是工程化(即,人工产生的)核酸分子,其用作操纵所选择的核苷酸序列和/或将这种核苷酸序列引入宿主细胞的工具。因此,重组载

体适合用于在克隆、测序和/或以其他方式操纵所选择的核苷酸序列中使用,如通过将所选择的核苷酸序列表达和/或递送到宿主细胞以形成重组细胞。这样的载体通常含有不被自然地发现与待克隆或递送的核苷酸序列相邻的核苷酸序列,尽管该载体还可以含有调节核苷酸序列(例如,启动子、非翻译区),这些调节核苷酸序列被自然地发现与本文所披露的核苷酸序列相邻或可用于本文所披露的核酸分子的表达(以下详细讨论)。该载体可以是原核或真核的RNA或DNA,并且通常是质粒。该载体可以作为染色体外元件(例如,质粒)被维持,或者它可以整合到重组宿主细胞的染色体中,尽管优选的是该载体与基因组保持分离。整个载体可以在宿主细胞内保留在适当位置,或者在某些条件下,可以缺失质粒DNA,留下本文所披露的核酸分子。整合的核酸分子可以处于染色体启动子控制下,天然或质粒启动子控制下或在若干个启动子控制的组合下。核酸分子的单拷贝或多拷贝可以整合到染色体中。本文所披露的重组载体可以含有至少一种选择性标记。

[0060] 用于本文所披露的重组核酸分子中的重组载体可以是表达载体。如本文所用,短语“表达载体”用于指适合于产生编码产物(例如,目的蛋白质,如本文所公开的酶)的载体。可以将编码待生产的产物的核苷酸序列插入重组载体中以产生重组核酸分子。将编码待生产的蛋白质的核苷酸序列以有效地将该核苷酸序列与载体中的调节序列连接的方式插入载体中,这使得能够在重组宿主细胞内转录和翻译该核苷酸序列。通常,重组核酸分子包括有效地连接到一个或多个表达控制序列(例如,转录控制序列或翻译控制序列)的如本文所公开的至少一种核酸分子。如本文所用,短语“重组分子”或“重组核酸分子”主要是指有效地与转录控制序列连接的核酸分子或核苷酸序列,但当这样的核酸分子是如本文所讨论的重组分子时,其可以与短语“核酸分子”互换使用。如本文所用,术语“有效地连接”是指以一种方式将核酸分子与表达控制序列连接使得当被转染(即转化、转导、转染、共轭或煅烧)进宿主细胞时该分子能够被表达。转录控制序列是控制转录的起始、延伸或终止的序列。尤其重要的转录控制序列是控制转录起始的那些序列,如启动子、增强子、操纵子和抑制子序列。适合的转录控制序列包括可在向其引入重组核酸分子的宿主细胞或生物体中起作用的任何转录控制序列。转录控制序列还可以包括前述任一项中的一种或多种的任何组合。

[0061] 重组核酸分子还可以含有另外的调节序列,如翻译调节序列、复制起点和与重组细胞相容的其他调节序列。重组分子,包括整合到宿主细胞染色体中的那些,优选地还包含分泌信号(即,信号区段核苷酸序列),以使表达的蛋白质能够从产生蛋白质的细胞分泌。适合的信号区段包括与待表达的蛋白质天然相关的信号区段或能够引导如本文所披露的蛋白质分泌的任何异源信号区段。重组分子可以包含前导序列,以使表达的蛋白质能够递送到并被插入到宿主细胞的膜中。适合的前导序列包括与该蛋白质天然相关的前导序列,或能够将蛋白质递送并插入到细胞膜的任何异源前导序列。

[0062] 术语“转染”通常用于意指可以将外源核酸分子(即,重组核酸分子)插入到细胞中的任何方法。术语“转化”可以与术语“转染”互换使用,当该术语用于意指将核酸分子引入微生物细胞或植物中并描述由于由微生物获得外源核酸而导致的遗传变化时,其基本上与术语“转染”同义。转染技术包括但不限于转化、粒子轰击、电穿孔、显微注射、脂质转染、吸附、感染和原生质体融合。

[0063] 术语“共转染”是指用两个单独的核酸分子同时转染。例如,共转染可以是指用一种包含具体基因的核酸分子和另一种包含具体标记基因的核酸分子同时转染。

[0064] 转基因在本文中理解为优选地经由本领域技术人员已知的重组技术已经引入细胞中的基因或经修饰的基因。转基因可以是同源的,即通常存在于细胞中,或者是异源的,即通常不存在于细胞中。优选地,转基因编码的目的蛋白质作为表达构建体的一部分,并且正在或将被转导至细胞或宿主细胞中用于所述目的蛋白质的重组生产。转基因可以编码异源或同源蛋白质。应当理解,编码同源蛋白质的转基因是异源转基因,该转基因作为通常不存在于细胞中的表达构建体的一部分(例如,包含与通常与蛋白质编码序列相关的启动子不同的启动子)。

[0065] 报道转基因或标记基因在本文中应理解为编码指示蛋白的转基因,即作为例如蛋白质表达水平的指示物来检测的蛋白质。

[0066] 异源序列在本文中应理解为在具体位置的序列,其在自然的所述位置不存在。换言之,包含异源序列的特定核酸序列是通常不会在自然界存在但经由随机或靶向遗传修饰被引入其中的核酸序列。

[0067] 异源蛋白质在本文中理解为不是由具体细胞天然产生的蛋白质,对于该具体细胞,该蛋白质被指示为异源的。

[0068] 同源蛋白质在本文中理解为由具体细胞天然产生的蛋白质,对于该具体细胞,该蛋白质被指示为同源的。同源蛋白质可以是细胞的内源性蛋白质、或外源性蛋白质,即在细胞已用编码同源蛋白质的表达载体转导的情况下由细胞重组地产生。

[0069] 应当理解,“工程化的”在本文中可以用于指人工产生的细胞(例如遗传修饰的细胞)或核酸或蛋白质序列。

[0070] 当与数值相关联地使用时(大约10、约10),词“大约(approximately)”或“约(about)”优选地意味着该值可以是值的或多或少10%的给定值。

[0071] 在本文件及其权利要求中,动词“包含”及其词形变化在其非限制意义上被用来表示包括该词后面的项目,但不排除未明确提到的项目。另外,通过不定冠词“一个(a)”或“一种(an)”提及一种要素时不排除多于一种要素存在的可能性,除非上下文清楚地要求存在一种并仅有一种要素。不定冠词“一个”或“一种”因此通常意指“至少一种”。

[0072] 在本说明书中引用的全部专利和参考文献通过引用以其全部内容结合在此。

[0073] 能够产生和分泌大量特定酶的微生物生产系统(特别是不存在高水平的其他蛋白质,包括对所希望的蛋白质的活性具有负面影响的蛋白质)对于研究和工业应用二者都具有实用性。它可以使能够简化筛选功能上表达所希望的酶的宿主。它还可以产生相对纯的酶。它还可以使能够简化所希望的酶的大规模纯化。这些优点将极大地有助于例如容易地生成针对不同应用定制的人工酶混合物,这些应用例如但不限于植物生物质水解(生物燃料和化学品)、纺织品整理、纸和纸浆中的应用、以及饲料和食品工业。使用所述方法产生的相对纯的酶也使能够在溶液(例如在水或混合溶剂中)或以固定形式设计有效的方法,这些方法用于但不限于生物催化、生物转化和生物修复。

[0074] 鉴定了具有出乎意料的高蛋白质生产能力、同时保持良好生长特征和对遗传修饰的适应性的真菌菌株的突变体。这些突变体作为微生物生产系统是有用的。本文公开了以下发现:SUMO化机器中酶的修饰是蛋白质生产的增加的原因。蛋白质或酶SUMO化是翻译后修饰机制,该翻译后修饰机制涉及SUMO-(小泛素样修饰剂)-蛋白质的成员与蛋白质或酶的赖氨酸残基的共价附接,该蛋白质或酶经由类似于但不同于泛素化途径的酶促级联而修饰

(SUMO化) (Wilkinson和Henley, *Biochem. J.* [生物化学杂志] 2010, 428 (2) :133-145)。SUMO蛋白质的实例是Smt3、SUMO-1、SUMO-2、SUMO-3和SUMO-4。据报道,底物蛋白质的蛋白质SUMO化的作用可导致底物蛋白质的活性、功能性和/或蛋白质相互作用的改变(增加或减少)。SUMO-缀合经由E1、E2和E3酶进行。经由E1“活化”酶, SUMO蛋白质以ATP-依赖性方式被活化。然后经由E2“缀合”酶(通常与E3“连接酶”缀合)将活化的SUMO蛋白转移至底物蛋白。SUMO蛋白酶在SUMO化的底物蛋白质的去SUMO化以及在前体SUMO蛋白质的活化中均起作用。已知多种E1活化酶、E3连接酶和SUMO蛋白酶,然而, Ubc9是唯一已知的E2缀合酶并且在生物体内高度保守(参见例如,图8)。已知的酵母E1活化酶是Aos1和Uba2。已知的哺乳动物E1活化酶是SAE1和SAE2。已知的酵母E3连接酶是Siz1、Siz2、Cst9和Mms21。已知的哺乳动物E3连接酶是PIAS1、PIAS3、PIAS4、PIASx α 、PIASx β 、PIASy、RanBP2、Pc2、Mms21、HDAC4、HDAC7、MUL1、Rhes、TOPORS、TLS、FUS、RSUME、ZMIZ1、NSE2、和TRAF7。已知的酵母SUMO蛋白酶是Ulp1和Ulp2。已知的哺乳动物SUMO蛋白酶是SENP-1、SENP-2、SENP-3、SENP-4、SENP-5、SENP-6、SENP-7、DESI-1、DESI-2和USPL1(参见例如, Chymokowitch等人, *Bioessays.* [生物学分析] 2015, 37 (10) : 1095-105;以及Wilkinson和Henley, *Biochem. J.* [生物化学杂志] 2010, 428 (2) :133-145)。

[0075] 因此,提供了具有增加的蛋白质生产的经修饰的细胞,其中所述细胞已被修饰而导致改变的SUMO化。SUMO化可以通过以下改变:修饰SUMO化机器和/或其编码基因的蛋白质,或暴露、孵化或插入SUMO化修饰剂,优选地导致细胞内SUMO化的减少。SUMO化修饰剂的实例是银杏酚酸、棣棠霉素B、大观霉素B1、毛壳色菌素(chaetochromin) A、紫黄素和/或来自其的衍生物。大观霉素B1和相关的天然产物公开在Hirohama等人的论文中(*ACS Chem, Biol.* [ACS化学生物学] 2013, 8, 2635-2642),通过抑制Ubc9来抑制SUMO化。

[0076] 可以使用可商购的试剂盒(例如来自艾博抗公司(Abcam), 剑桥, 马萨诸塞州)和/或本领域已知的方法确认细胞中SUMO化的修饰。本领域已知的方法包括例如ATP:PPi同位素交换测定,用薄层色谱法确定E1-催化的ATP:AMP交换速率、或通过基于凝胶的测定确定E1-SUMO缀合(Alontaga等人, 2012年7月, *Biochemical Analysis of Protein SUMOylation* [蛋白质SUMO化的生化分析]在*Curr Protoc Mol Biol* [当前分子生物学协议], 第10章:单元10.29)。

[0077] 在实施例中,与如在基本相同或相当条件下检测的其亲本细胞相比,经修饰的细胞显示增加的蛋白质生产。蛋白质生产的增加可以是总或全部蛋白质生产的增加。在实施例中,与亲本细胞相比,经修饰的细胞显示至少1.1、1.2、1.3、1.4、1.5、1.6、1.7、1.8、1.9、2、2.1、2.2、2.3、2.4、2.5、2.6、2.7、2.8、2.9、3、3.1、3.2、3.3、3.4、3.5、3.6、3.7、3.8、3.9、4、5、6、7、8、9、10、15、20、25、30、35、40、45或至少50倍的增加的蛋白质生产。在本文中“亲本细胞”因理解为祖先,从其中通过改变如本文描述的SUMO化直接衍生经修饰的细胞。优选地,经修饰的细胞通过以下直接衍生自亲本细胞:(i)通过遗传修饰编码SUMO化机器的内源性蛋白质的至少一个基因、和/或(ii)通过暴露、孵育或插入至少一种SUMO化修饰剂(优选地是SUMO化机器的至少一种内源性蛋白质的抑制剂)。

[0078] 蛋白质生产中的增加可以通过本领域已知的任何适合的技术测量。可以通过检测细胞培养的具体时期内细胞产生和/或分泌的蛋白质的总量或具体类型来测量蛋白质生产。可以使用可商购的测定(如比色Bradford和Lowry方法)、以及使用可商购的比色方法(如BCA测定(例如, Pierce; Biorad))的方法测量总或全部蛋白质生产。蛋白质生产的增加

可以通过检测编码的细胞外蛋白质的量、检测在限定时间后细胞培养物或发酵培养基中存在的蛋白质的量来测量。蛋白质生产的增加可以经由本领域已知的方法检测具体蛋白质(细胞或细胞外)的量或活性来测量。

[0079] 本领域技术人员还将理解,蛋白质生产的增加可以实现生产过程中的所希望的变化。可以通过本领域技术人员已知的方法测量的生产参数的改善来证明此类改善。例如,增加的蛋白质生产可以导致增加的滴度、增加的体积生产力、增加的比生产力、和/或增加的产量。例如可以通过在发酵期间在给定的时间点的每体积发酵液或发酵上清液的蛋白质或酶活性的量(例如克蛋白质/升或活性单位/升)来测量增加的滴度。例如可以通过每体积发酵液或发酵上清液的蛋白质或酶活性的速率(例如,克蛋白质/升/小时或活性单位/升/小时)来测量增加的体积生产力。例如可以通过每个细胞质量产生的蛋白质或酶活性的速率(例如,克蛋白质/克干细胞重量/小时或活性单位/克干细胞重量/小时)来测量增加的比生产力(包括例如最大比生产力)。例如可以通过在发酵期间消耗的每个量的碳或碳源(例如葡萄糖)产生的蛋白质或酶活性的量(例如,克蛋白质/克葡萄糖或活性单位/克葡萄糖或蛋白质中碳的克/克碳进料)来测量增加的产量。因此,本文提供的是生产过程,该生产过程包括发酵本文中的经修饰的细胞以产生其中至少一种生产参数增加的产物。此类增加可以是至少约1%、至少约5%、至少约10%、至少约25%、至少约30%、至少约40%、至少约50%、至少约60%、至少约70%、至少约80%、至少约90%、或至少约100%。

[0080] 与由缺乏SUMO化的修饰的亲本细胞来生产所述蛋白质相比,蛋白质生产的增加可以是目的蛋白质的生产增加至少1.1倍。优选地,与亲本细胞相比,经修饰的细胞显示至少约1.1、1.2、1.3、1.4、1.5、1.6、1.7、1.8、1.9、2、2.1、2.2、2.3、2.4、2.5、2.6、2.7、2.8、2.9、3、3.1、3.2、3.3、3.4、3.5、3.6、3.7、3.8、3.9、4、5、6、7、8、9、10、15、20、25、30、35、40、45或至少约50倍的目的蛋白质的增加的生产。此类目的蛋白质可以由作为充当报道基因的转基因编码的蛋白质(在本文中称为报道蛋白),其已经被引入经修饰的细胞中以检测经修饰的细胞的蛋白质生产能力。在建立通过修饰SUMO化引起的蛋白质生产的增加中,亲本细胞和经修饰的细胞二者都包含报道基因,并且这些细胞之间比较的编码蛋白质的产生保持在基本相同的培养条件下。可以通过本领域适合的方法测量报道基因生产以鉴定产生的报道基因的量,例如,在报道基因是荧光蛋白的情况下可以测量荧光,或者在报道基因是酶的情况下可以测量酶活性。

[0081] 适合报道转基因包含报道蛋白的编码序列,该编码序列有效地连接到启动子序列,该启动子序列允许在经修饰的细胞及其亲本细胞中表达报道转基因的编码序列。例如,在经修饰的细胞是菌株嗜热毁丝霉的真菌细胞的情况下,启动子序列可以是如在WO 2010/107303 A2(将其通过引用结合在此)中公开的启动子序列。该报道蛋白可以是分泌蛋白,并且可以通过检测胞外蛋白的水平测量蛋白质水平。报道基因可以连接(例如物理连接)到编码选择性标记的基因,该选择性标记如但不限于amdS选择性标记。本文例示的并且适合用于真菌宿主细胞(如菌株嗜热毁丝霉的真菌宿主细胞)的报道基因由SEQ ID NO:4代表,其包含chi1启动子序列和纤维素酶(Eg2)编码序列。本文例示的并且适合用于真菌宿主细胞(如菌株嗜热毁丝霉的真菌宿主细胞)的另外的报道基因由SEQ ID NO:5代表,其包含chi1启动子序列或cbh1启动子序列和聚半乳糖醛酸酶(AnPGII)编码序列。本文例示的并且适合用于真菌宿主细胞(如菌株嗜热毁丝霉的真菌宿主细胞)的另外的报道基因由SEQ ID NO:6

代表,其包含chi1启动子序列和 β -木糖苷酶(Bxl1)编码序列。本文例示的并且适合用于真菌宿主细胞(如菌株嗜热毁丝霉的真菌宿主细胞)的另外的报道基因包含chi1启动子序列或cbh1启动子序列和编码葡糖淀粉酶(如里氏木霉(*T.reesei*)葡糖淀粉酶)的序列。本文例示的并且适合用于真菌宿主细胞(如菌株嗜热毁丝霉的真菌宿主细胞)的另外的报道基因包含chi1启动子序列或cbh1启动子序列和编码植酸酶(例如布丘氏菌属物种(*Buttiauxella* sp.)植酸酶)变体的序列(分别为SEQ ID No 53和54)。编码适合在真菌细胞中表达的植酸酶的示例性序列可以包含SEQ ID NO:40的核苷酸1905-3184的序列。因此,适合的目的蛋白质包括但不限于Eg2、Bxl1、AnPGII、里氏木霉葡糖淀粉酶、和布丘氏菌属物种植酸酶变体。

[0082] 应当理解,还可以使用对蛋白质特异性的测定来测量目的蛋白质。此类测定法是本领域技术人员已知的,并且可包括例如结合测定、HPLC、ELISA、凝胶光密度测定或用于确定总蛋白质水平的方法。而且,如实例中所证明的,可以使用适合目的蛋白质的活性测定来测量目的酶蛋白质的产生的增加。本领域普通技术人员将能够容易地为所希望的蛋白质选择适当的测量方法。

[0083] 在一个实施例中,通过SUMO化途径的至少一个基因(如编码内源SUMO化酶的基因或编码内源SUMO蛋白质的基因)的遗传修饰,在经修饰的细胞中改变了SUMO化。在实施例中,遗传修饰可以是SUMO化途径的至少一个基因(如编码内源SUMO化酶的基因或编码内源SUMO蛋白质的基因)的靶向遗传修饰。SUMO化机器的蛋白质的实例是:Ubc9、Smt3、SUMO-1、SUMO-2、SUMO-3和SUMO-4、Siz1、Siz2、Cst9、Mms21、PIAS1、PIAS3、PIAS4、PIAS α 、PIAS β 、PIAS γ 、RanBP2、Pc2、Mms21、HDAC4、HDAC7、MUL1、Rhes、TOPORS、TLS、FUS、RSUME、ZMIZ1、NSE2、TRAF7、Ulp1、Ulp2、Aos1、Uba2、SENP-1、SENP-2、SENP-3、SENP-4、SENP-5、SENP-6、SENP-7、DESI-1、DESI-2和USPL1。该遗传修饰可以涵盖一个或多个点突变、异源序列的一个或多个插入和/或编码所述SUMO化酶或蛋白质的基因的内源序列(的一部分)的一个或多个缺失。该一个或多个点突变、一个或多个插入和/或一个或多个缺失可位于编码序列和/或调控序列中。表达调控序列的实例是启动子序列、终止子序列、启动子激活序列和编码转录因子的序列。该修饰可以导致去除或破坏编码内源SUMO化酶或蛋白质的全部或部分基因、或者例如用编码选择标记的基因替换全部或部分基因。适合的选择标记的实例是AmdS选择性标记。修饰可以涵盖“敲除”基因的内源拷贝。基因的“敲除”是指一种分子生物学技术,利用其使生物体中的基因不起作用,使得基因的表达大幅度降低或消除。进一步涵盖点突变,这些点突变导致酶或蛋白质生产和/或活性的降低或消除。

[0084] 遗传修饰可以导致编码SUMO化酶或蛋白质的产生、活性的改变或改变的产生和活性两者。遗传修饰可以导致编码的SUMO化酶或蛋白质的产生和/或活性的降低或甚至消除,并且可导致SUMO化的减少或消除。

[0085] 此类遗传修饰可以使用本领域技术人员已知的方法进行,这些方法配备有编码SUMO化酶或蛋白质的基因的序列。表达调控序列由本领域技术人员容易地鉴定。用于遗传修饰的适合方法包括本领域技术人员已知的那些并且包括但不限于本文示例的那些、同源重组、RNA沉默和CRISPR/Cas系统(Timberlake等人,1989,Science[科学]244(4910):1313-1317.; Moore,36-66页,在生物技术,第III卷:Fundamentals in Biotechnol.[基础生物技术]编辑Doelle,Robek,Berovic(2009); Singh等人,2017,Gene[基因]599:1-18)。适合的试

剂和方法同样是本领域已知的和/或可商购的(参见例如,WO 2016100568 A1、WO 2016100272 A1和WO 2016100571 A1,其中每个都通过引用结合在此)。例如,纯化的Cas9蛋白和指导RNA、或用来制备指导RNA的试剂盒可获得自PNA BIO (www.pnabio.com; 纽伯里公园(Newbury Park),加利福尼亚州)、NEB (www.neb.com; 伊普斯威奇(Ipswich),马萨诸塞州)、赛默飞世尔公司(ThermoFisher) (www.thermofisher.com; 沃尔瑟姆(Waltham),马萨诸塞州)、和IDT (www.idtdna.com或www.idtdna.com/site/order/oligoentry/index/crispr/2nm;Coralville,衣阿华州)。

[0086] 可以通过本领域已知的技术检测内源SUMO化酶或蛋白质的生产。经修饰的细胞的遗传修饰可以导致经修饰的SUMO化酶或蛋白质的转录物总量的减少,这可以通过本领域已知的适合的测定检测,例如经由如本文示例的RNA-seq(参考本公开的实例1)。与亲本细胞中的所述转录物的量相比,经修饰的细胞中的所述转录物的量可以减少至少约0.2、0.3、0.4、0.5、0.6、0.7、0.8、0.9、1、1.5、2、2.5、3、4、5、6、7、8、9、10、15、20、25、30、45、50、60、70、80、90、或至少100倍。与上文定义的指示的减少结合,遗传修饰可以使得所述转录物的量不减少至零,但是所述转录物的残留量为亲本细胞中存在的所述转录物的量的至少约0.1%、0.2%、0.3%、0.4%、0.5%、0.6%、0.7%、0.8%、0.9%、1%、2%、3%、4%、5%、6%、7%、8%、9%、10%。

[0087] 遗传修饰的细胞在结构上不同于亲代细胞,因为它包含遗传修饰。通过引入遗传修饰,可以直接从亲本细胞获得经修饰的细胞。修饰基因可以是编码选自以下组成的组的蛋白质的基因中的任一种:Ubc9、Smt3、SUMO-1、SUMO-2、SUMO-3和SUMO-4、Siz1、Siz2、Cst9、Mms21、PIAS1、PIAS3、PIAS4、PIAS α 、PIAS β 、PIAS γ 、RanBP2、Pc2、Mms21、HDAC4、HDAC7、MUL1、Rhes、TOPORS、TLS、FUS、RSUME、ZMIZ1、NSE2、TRAF7、Ulp1、Ulp2、SEN-1、SEN-2、SEN-3、SEN-4、SEN-5、SEN-6、SEN-7、DESI-1、DESI-2和USPL1、以及这些蛋白质或酶中的任一种的同系物。可以使用本领域已知的方法在具体目的细胞中鉴定同系物,例如序列对齐、系统发育树的产生和分析。遗传修饰可以涵盖编码一种或多种SUMO化蛋白质的基因的编码序列和/或表达调控序列的随机或定点突变、缺失、破坏、沉默,其导致通过包含如本文所定义的修饰的细胞的蛋白质生产的增加,即与在基本相同的条件下测试时通过其亲本细胞的蛋白质生产相比。

[0088] 该遗传修饰可以包含编码Ubc9的基因的修饰或由其组成。内源性ubc9基因可以是本领域技术人员已知的任何基因,例如,编码以下的ubc9基因:嗜热毁丝霉的Ubc9(SEQ ID NO:2)、粗糙脉孢菌(*Neurospora crassa*) Ubc9(核酸SEQ ID NO:8;蛋白质SEQ ID NO:9;登录号:NCU04302;XM_955999)、构巢曲霉Ubc9(核酸SEQ ID NO:10;蛋白质SEQ ID NO:11;登录号:AN_4399)、新型隐球菌(*Cryptococcus neoformans*) (核酸SEQ ID NO:12;蛋白质SEQ ID NO:13;登录号:CNAG_04328)、酿酒酵母(*Saccharomyces cerevisiae*) Ubc9(核酸SEQ ID NO:14;蛋白质SEQ ID NO:15;登录号:YDL064W;Z74112.1)、黑曲霉Ubc9(核酸SEQ ID NO:45;蛋白质SEQ ID NO:46)、和里氏木霉Ubc9(启动子序列SEQ ID NO:47;编码SEQ ID NO:48;cDNA SEQ ID NO:49;蛋白质SEQ ID NO:50)。序列信息可以例如在公共数据库中找到,如在JGI MycoCosm(genome.jgi.doe.gov/programs/fungi/index.jsf);国家生物技术信息中心("NCBI";www.ncbi.nlm.nih.gov);曲霉属基因组数据库("AsGD";www.aspergillusgenome.org);或酵母属基因组数据库("SGD";www.yeastgenome.org)中。

该遗传修饰可以在编码蛋白质的基因中,该蛋白质与以下序列基序中的一种或多种或两者的序列具有至少70%、至少75%、至少80%、至少85%、至少90%、或至少95%序列同一性,或包含以下序列基序中的一种或多种或两者的序列:RLQEERKQWRKDHPFGF (SEQ ID NO:42) 和KPPKCKFTPPLFHPNVYPSGTVCLSIL (SEQ ID NO:43)。遗传修饰可以是修饰ubc9基因的调控区,例如在ubc9基因的启动子序列中,例如通过在ubc9启动子区域(蛋白质编码序列的上游)插入表达盒。用于破坏嗜热毁丝霉中的ubc9启动子序列的插入盒的实例由SEQ ID NO:3表示。用于破坏里氏木霉中ubc9启动子序列的靶序列的实例由SEQ ID No:30、31、和32表示。用于破坏黑曲霉(*A.niger*)中ubc9启动子序列的插入盒的实例由SEQ ID NO:26表示。本文还考虑了导致ubc9表达、Ubc9生产或活性的修饰的其他遗传修饰,如Ubc9的失活突变体或ubc9编码序列中的基因破坏,优选地导致Ubc9产生和/或活性的降低或消除。

[0089] 经修饰的细胞特征在于其包含内源性ubc9(泛素缀合酶或E2酶)基因,该基因已经被修饰以增加蛋白质生产。包含经修饰的细胞中的修饰的内源性ubc9基因可以具有野生型(未经修饰的)对应物,该对应物在其整个长度上与SEQ ID NO:1具有至少40%、50%、60%、70%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%序列同一性,或该对应物在其整个长度上与SEQ ID NO:1(由位置1070和2195上的核苷酸限定的(即SEQ ID NO:1(1070-2195)))具有至少40%、50%、60%、70%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%序列同一性,或该对应物在其整个长度上与SEQ ID NO:1的序列(由位置1302和1994上的核苷酸限定的(即SEQ ID NO:1(1302-1994)),代表可读框)具有至少40%、50%、60%、70%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%序列同一性。包含经修饰的细胞中的修饰的内源性ubc9基因可以具有野生型对应物,该野生型对应物包含至少具有经修饰的细胞的序列,该序列可以具有野生型(未经修饰的)对应物,该野生型(未经修饰的)对应物在其整个长度上与SEQ ID NO:47或与SEQ ID NO:48具有至少40%、50%、60%、70%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%序列同一性。包含经修饰的细胞中的修饰的内源性ubc9基因可以具有野生型对应物,该野生型对应物包含至少具有经修饰的细胞的序列,该序列可以具有野生型(未经修饰的)对应物,该野生型(未经修饰的)对应物与SEQ ID NO:45的启动子或编码部分具有至少40%、50%、60%、70%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%序列同一性。

[0090] 包含经修饰的细胞中的遗传修饰的内源性ubc9基因可以编码Ubc9蛋白质,该Ubc9蛋白质具有序列,该序列在其整个长度上与SEQ ID NO:2具有至少40%、50%、60%、70%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%序列同一性。包含经修饰的细胞中的遗传修饰的内源性ubc9基因可以编码Ubc9蛋白质,该Ubc9蛋白质具有序列,该序列在其整个长度上与SEQ ID NO:50具有至少40%、50%、60%、70%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%序列同一性。包含经修饰的细胞中的遗传修饰的内源性ubc9基因可以编码Ubc9蛋白质,该Ubc9蛋白质具有序列,该序列在其整个长度上与SEQ ID NO:46具有至少40%、50%、60%、70%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%序列同一性。

[0091] 配备本公开,本领域技术人员可以容易地鉴定目的微生物中的ubc9序列。例如,超过600个公开可用的Ubc9直向同源物的对齐揭示了延伸的和强的序列保守的两个突出的延

伸,在本文中命名为基序1(17aa;SEQ ID NO:42)和基序2(27aa;SEQ ID NO:43)(参见图8)。两个延伸都位于表征ubc蛋白的“泛素缀合酶”蛋白质结构域(PF00179;pfam.xfam.org/)内。这些延伸可以用于从含有PF00179结构域的其他蛋白质中鉴定ubc9。

[0092] 经修饰的细胞可以具有内源性ubc9基因启动子序列的插入或破坏。可以通过插入表达盒来破坏ubc9启动子序列,例如,通过在与嗜热毁丝霉的ubc9基因的启动子序列内的核苷酸类似的位置上插入表达盒来破坏ubc9启动子序列,该位置由SEQ ID NO:1的位置1257和1258上的核苷酸表示。发现在该位置的启动子序列破坏导致总蛋白质生产的意外增加,更具体是在编码细胞外分泌蛋白的转导报道基因的表达中。不希望受任何理论束缚,启动子序列的破坏可能损害编码的Ubc9蛋白质的功能性表达,这可能经由其在翻译后修饰途径中的关键作用导致净蛋白质生产的增加。应当理解,内源性ubc9基因修饰不限于启动子序列破坏。本文考虑的是导致损害或消除编码的ubc9蛋白质的功能性表达的ubc9基因修饰。此外,本文考虑的是导致蛋白质生产的总体增加的SUMO化酶、蛋白质或基因修饰。

[0093] 在另一个实施例中,或与上文定义的实施例组合,将经修饰的细胞暴露于一种或多种改变SUMO化的生物或化学试剂,如SUMO化酶抑制剂或活化剂。可以将经修饰的细胞以有效抑制所述酶的活性的浓度暴露于SUMO化酶的至少一种抑制剂中。该抑制剂可以是选自自由以下组成的组中的任一个:银杏酚酸、棣棠霉素B、大观霉素B1、毛壳色菌素(chaeochromin)A、紫黄素和/或来自其的衍生物。大观霉素B1和相关的天然产物公开在Hirohama等人的论文中(ACS Chem,Biol.[ACS化学生物学]2013,8,2635-2642),通过抑制Ubc9来抑制SUMO化。因此,可以将经修饰的细胞以有效抑制所述酶的活性的浓度暴露于公开在Hirohama等人的论文中(ACS Chem,Biol.[ACS化学生物学]2013,8,2635-2642)中的大观霉素B1和/或相关的天然产物中。可以将经修饰的细胞以所述试剂的浓度暴露于改变SUMO化的试剂中,该经修饰的细胞,该浓度导致如本文所定义的蛋白质生产的增加,即与在基本相同的条件下测试时的其亲本细胞相比。暴露于如本文所定义的一种或多种SUMO化酶抑制剂或活化剂的经修饰的细胞可以在结构上与亲本细胞相似,并且仅与其不同之处在于它暴露于所述一种或多种抑制剂或一种或多种活化剂中。可以将经修饰的细胞暴露于抑制剂中以导致酶活性减少(与亲本细胞的酶活性相比)了至少约0.2、0.3、0.4、0.5、0.6、0.7、0.8、0.9、1、1.5、2、2.5、3、4、5、6、7、8、9、10、15、20、25、30、45、50、60、70、80、90、或至少约100倍。与上文定义的指示的减少结合,所得酶活性可以使得酶活性的量不是减少至零,而是残留酶活性是亲本细胞中的活性的至少约0.1%、0.2%、0.3%、0.4%、0.5%、0.6%、0.7%、0.8%、0.9%、1%、2%、3%、4%、5%、6%、7%、8%、9%、10%。

[0094] 由于本文提供的经修饰的细胞显示出高蛋白质生产能力,所以本文提供的经修饰的细胞可用于工业蛋白质生产,如但不限于生产药物蛋白质和/或工业酶(在本文中也称为“目的蛋白质”)。目的蛋白质可以是内源性蛋白质或外源性蛋白质,即由转基因编码。转基因可以编码同源或异源蛋白质。编码和产生目的蛋白质(或目的蛋白质混合物)的经修饰的细胞在本文中也称为表达系统。

[0095] 在其中目的蛋白质是转基因的本发明的实施例中,经修饰的细胞包含外源表达构建体或载体,该外源表达构建体或载体编码至少一种目的蛋白质(由包含在表达构建体或载体上的转基因编码)。表达构建体可以包含编码目的蛋白质的基因或多核苷酸的部分,其不是蛋白质编码区的部分(例如,编码蛋白质的基因的内含子或调控区),并且可以是双链

的、单链的,并且可以包括DNA、RNA、或DNA或RNA之一的衍生物(包括cDNA、探针和引物,包括指导序列)。任选地,表达构建体编码一种或多种工业酶。该表达构建体可以是编码至少两种目的蛋白质的嵌合表达构建体或载体。

[0096] 本领域技术人员应当理解,使用重组DNA技术可以通过操纵例如编码宿主细胞内的目的蛋白质的序列的拷贝数、转录编码目的蛋白质的那些序列的效率、翻译所得转录物的效率、以及翻译后经修饰的效率来影响对转染的表达载体(编码同源或异源蛋白质)的表达的控制。转基因可以包含翻译增强和/或调控序列,如适合在目的宿主细胞中表达的启动子序列。此类启动子序列可以是用语音在嗜热毁丝霉(如嗜热毁丝霉C1)中表达蛋白质的启动子序列,公开在WO 2010/107303A2中,将其通过引用结合在此。例如,此类启动子序列可以包含嗜热毁丝霉chi1启动子序列(如SEQ ID NO:51)、或嗜热毁丝霉cbh启动子序列(如SEQ ID NO:52)、或与SEQ ID NO:51或52具有至少约85%、至少约90%、至少约95%或至少约99%同一性的启动子序列。用于在木霉属中表达蛋白质的适合的启动子序列可以包括但不限于里氏木霉cbh1、cbh2、egl1、egl2、egl3、egl4、egl5、pki1、gpd1、xyn1、或xyn2启动子。此外,与天然启动子相比,该启动子序列可以被基因工程化以提高表达水平。有用于控制核酸分子表达的重组技术包括但不限于,将核酸分子整合到一个或多个宿主细胞染色体中,向质粒添加载体稳定性序列,转录控制信号的替换或修饰(例如,启动子、操纵子、增强子),翻译控制信号(例如,核糖体结合位点)的取代或修饰,对核酸分子进行修饰以对应于宿主细胞的密码子使用,以及使转录物不稳定的序列的缺失。

[0097] 该一种或多种目的蛋白质可以是酶混合物或多酶组合物。换言之,可以进一步修饰经修饰的细胞以产生对工业目的有益的其他酶或酶混合物或多酶组合物。

[0098] 目的蛋白质可以是例如半纤维素酶、过氧化物酶、蛋白酶、纤维素酶、木聚糖酶、脂肪酶、磷脂酶、酯酶、角质酶、果胶酶、角蛋白酶、还原酶、氧化酶、酚氧化酶、脂加氧酶、木质素酶、支链淀粉酶、鞣酸酶、戊聚糖酶、甘露聚糖酶、 β -葡聚糖酶、阿拉伯糖苷酶、透明质酸酶、软骨素酶、漆酶、淀粉酶、葡糖淀粉酶、其混合物、其功能片段、或这些酶及其功能片段中的一种或多种的混合物。蛋白质的非限制性实例可以进一步包括淀粉代谢中涉及的蛋白质或酶;糖原代谢中涉及的蛋白质或酶;乙酰酯酶、氨基肽酶、淀粉酶、阿拉伯糖酶(arabinases)、阿拉伯呋喃糖苷酶、羧肽酶、过氧化氢酶、纤维素酶、几丁质酶、凝乳酶、角质酶、脱氧核糖核酸酶、差向异构酶、酯酶、 α -半乳糖苷酶、 β -半乳糖苷酶、 α -葡聚糖酶、葡聚糖裂合酶、内切- β -葡聚糖酶、葡糖淀粉酶、葡萄糖氧化酶、 α -葡糖苷酶、 β -葡糖苷酶、葡萄糖醛酸酶、半纤维素酶、己糖氧化酶、水解酶、转化酶、异构酶、漆酶、脂肪酶、裂合酶、甘露糖苷酶、氧化酶、氧化还原酶、果胶酸裂合酶、果胶乙酰酯酶、果胶解聚酶、果胶甲酯酶、果胶分解酶、过氧化物酶、酚氧化酶、植酸酶、聚半乳糖醛酸酶、蛋白酶、鼠李糖-半乳糖醛酸酶、核糖核酸酶、奇异果甜蛋白、转移酶、转运蛋白、转谷氨酰胺酶、木聚糖酶、己糖氧化酶(D-己糖:02-氧化还原酶,EC 1.1.3.5)、其变体、其功能片段或其组合。目的蛋白质还可以是肽激素、生长因子、凝血因子、趋化因子、细胞因子、淋巴因子、抗体、受体、粘附分子、微生物抗原(例如,HBV表面抗原、HPV E7等)、或以上物质的变体、功能片段、或两种或更多种、三种或更多种、四种或更多种、五种或更多种、或六种或更多种的混合物。酶的混合物可以包括例如公开在PCT申请公开号WO 2008/153903、WO 2012/125951和/或WO 2012/125937中的组合物,其每个都通过引用结合在此。

[0099] 经修饰的细胞可以是编码(内源的和/或转基因的之一)和产生多酶组合物的细胞,该多酶组合物适合降解木质纤维素材料和/或半纤维素材料。此类多酶组合物可以包含至少一种纤维二糖水解酶、至少一种木聚糖酶、至少一种内切葡聚糖酶、至少一种 β -葡糖苷酶、至少一种 β -木糖苷酶、和至少一种辅酶。在实施例中,该多酶组合物可以缺乏导致形成纤维二糖内酯/纤维二糖酸和/或葡萄糖酸内酯/葡萄糖酸的酶。木聚糖酶可以是内切木聚糖酶、外切木聚糖酶、或 β -木糖苷酶。辅酶可以具有与酶核心组中的一种或多种酶相同或相似的功能或不同的功能。适合的辅酶已经在本文其他地方描述,并且通常可以包括例如纤维素酶、木聚糖酶、木质素酶、淀粉酶、脂质酶或葡萄糖醛酸酶。一些辅酶例如可以包括当在反应中与生物质接触时允许增加多酶组合物中酶(例如,半纤维素酶)活性的酶。辅酶或酶混合物可以由来自以下的酶构成:(1)商业供应商;(2)表达酶的克隆基因;(3)复合培养液(例如由培养基中的微生物菌株生长而产生的培养液,其中菌株将蛋白质和酶分泌到培养基中);(4)如(3)中生长的菌株的细胞裂解液;以及(5)表达能够降解木质纤维素的酶的植物材料。在一些实施例中,该辅酶是葡糖淀粉酶、果胶酶或木质素酶。辅酶包括选自自由以下组成的组的酶:纤维素酶、葡糖苷酶、溶解性多糖单加氧酶(或GH61或具有纤维素分解增强活性的多肽)、木聚糖酶、木糖苷酶、木质素酶、葡萄糖醛酸酶、阿拉伯呋喃糖苷酶、阿拉伯糖酶(arabinase)、阿拉伯半乳聚糖酶、阿魏酸酯酶、脂肪酶、果胶酶、葡甘聚糖酶、淀粉酶、昆布多糖酶、木葡聚糖酶、半乳聚糖酶、半乳糖苷酶、葡糖淀粉酶、果胶酸裂合酶、壳聚糖酶、外切- β -D-葡糖胺酶、纤维二糖脱氢酶、以及乙酰木聚糖酯酶。在实施例中,在缺乏导致形成纤维二糖内酯/纤维二糖酸和/或葡萄糖酸内酯/葡萄糖酸的酶情况下,辅酶存在于酶混合物中。该多酶组合物可以进一步包含至少一种半纤维素酶。半纤维素酶可以选自自由以下组成的组:木聚糖酶、阿拉伯呋喃糖苷酶、乙酰木聚糖酯酶、葡萄糖醛酸酶、内切半乳聚糖酶、甘露聚糖酶、内切阿拉伯糖酶(arabinase)、外切阿拉伯糖酶(arabinase)、外切半乳聚糖酶、阿魏酸酯酶、半乳甘露聚糖酶、木葡聚糖酶及其混合物。木聚糖酶可以选自自由以下组成的组:内切木聚糖酶、外切木聚糖酶、和 β -木糖苷酶。该多酶组合物可以进一步包含至少一种纤维素酶。

[0100] 本文考虑的多酶组合物可以作为粗发酵产物从经修饰的细胞中获得,任选地经受纯化步骤。该多酶组合物可以进一步包含一种或多种辅酶。辅酶可以包括至少一种选自自由以下组成的组的酶:纤维素酶、葡糖苷酶、溶解性多糖单加氧酶(或GH61或具有纤维素分解增强活性的多肽)、木聚糖酶、木糖苷酶、木质素酶、葡萄糖醛酸酶、阿拉伯呋喃糖苷酶、阿拉伯糖酶(arabinase)、阿拉伯半乳聚糖酶、阿魏酸酯酶、脂肪酶、果胶酶、葡甘聚糖酶、淀粉酶、昆布多糖酶、木葡聚糖酶、半乳聚糖酶、半乳糖苷酶、葡糖淀粉酶、果胶酸裂合酶、壳聚糖酶、外切- β -D-葡糖胺酶、纤维二糖脱氢酶、乙酰木聚糖酯酶以及乙酰酯酶。在一些实施例中,通过在底物上培养经修饰的细胞以产生酶来产生辅酶。该多酶组合物可以进一步包含用于降解含阿拉伯木聚糖的材料的至少一种蛋白质或其具有生物活性的片段。该多酶组合物可以进一步包含至少一种内切木聚糖酶、至少一种 β -木糖苷酶和至少一种阿拉伯呋喃糖苷酶。阿拉伯呋喃糖苷酶可以包含对单取代的木糖残基具有特异性的阿拉伯呋喃糖苷酶、对双取代的木糖残基具有特异性的阿拉伯呋喃糖苷酶或其组合。

[0101] 多酶组合物还可以包括纤维素酶、半纤维素酶(如木聚糖酶,包括内切木聚糖酶、外切木聚糖酶和 β -木糖苷酶;甘露聚糖酶,包括内切甘露聚糖酶、外切甘露聚糖酶和 β -甘露

糖苷酶)、木质素酶、淀粉酶、葡萄糖醛酸酶、蛋白酶、酯酶(包括阿魏酸酯酶)、脂肪酶、葡糖苷酶(如 β -葡糖苷酶)、以及木葡聚糖酶。

[0102] 还考虑的是产生包含酶的混合物的经修饰的细胞,这些酶能够降解细胞壁并且释放细胞内容物。此类细胞可以是细菌细胞,或藻类、真菌、或植物的细胞,这些细胞天然地产生酶或借助遗传修饰以表达一种或多种酶。

[0103] 在一个实施例中,在方法中采用多酶组合物来从降解生物质和/或木质纤维素材料(如本文所定义的富含半纤维素的生物质)获得可发酵产物,如糖。所述木质纤维素材料可以是例如农业废弃物或农业和林业产生的残余物。木质纤维素材料可以是例如玉米秸秆、甘蔗渣、或小麦秸秆。木质纤维素材料可以是预处理的木质纤维素材料。预处理的木质纤维素材料可以例如通过使生物质材料经受升高的温度并且添加稀酸、浓酸或稀碱溶液来生产。预处理的木质纤维素材料可以例如通过使生物质材料在高固体条件下经受低氨浓度来生产。木质纤维素材料可部分或完全降解成可发酵糖。考虑到商业上可行的成本降低的经济水平。部分由于包含生物质和木质纤维素材料的许多组分,所以需要能够降解木聚糖、木质素、蛋白质和糖类的酶或酶的混合物来实现糖化。该一种或多种另外的酶可以包括具有例如氧化还原酶、纤维二糖水解酶、内切葡聚糖酶、 β -葡糖苷酶、木聚糖酶和其他半纤维素酶活性的酶或其组合物。这些酶组合物适合降解生物质,如富含半纤维素的生物质,例如,如本文所定义的纸浆。

[0104] 多酶组合物是将纸浆溶解用于生产化学品,例如但不限于人造丝、玻璃纸以及若干化学品如纤维素酯(乙酸酯、硝酸酯、丙酸酯和丁酸酯)和纤维素醚(羧甲基纤维素以及甲基纤维素和乙基纤维素)以及乙醇(例如作为生物乙醇生物燃料)。还考虑了用于纸和纸浆漂白的多酶组合物。该多酶组合物可以用于改善纸浆和造纸工业中纸浆的漂白率。

[0105] 经修饰的细胞(例如,宿主细胞或生产生物)可以包括任何微生物,如原生生物、藻类、真菌或其他真核微生物、植物、昆虫、或动物细胞,并且可以是酵母或丝状真菌。经修饰的细胞是可以被转染的任何真菌(例如,丝状真菌或酵母或蘑菇)、藻类、植物、昆虫、或动物细胞。经修饰的细胞可以是培养的细胞。酵母的适合的属包括但不限于酵母属、裂殖酵母属、假丝酵母属、汉逊酵母属、毕赤酵母属、克鲁维酵母属和Phaffia属。适合的酵母物种包括但不限于酿酒酵母、粟酒裂殖酵母、白色念珠菌、多形汉逊酵母、巴斯德毕赤酵母、加拿大毕赤酵母、马克斯克鲁维酵母和红法夫酵母。

[0106] 适合的真菌属包括但不限于金孢子菌属、梭孢壳属、脉孢菌属、短梗霉属、线黑粉菌科(Filibasidium)、单鞭毛菌属、棒囊孢壳菌、隐球菌属、枝顶孢属、弯颈霉属、柱顶孢霉属、裂褶菌属、侧孢霉属、青霉属、赤霉属、毁丝霉属、毛霉属、曲霉属、镰孢霉属、腐质霉属、以及木霉属、踝节菌属、Rasamsonia以及其无性型和有性型。适合的真菌物种包括但不限于黑曲霉、米曲霉、构巢曲霉、日本曲霉、蓝色犁头霉(Absidia coerulea)、米根霉、嗜热毁丝霉、粗糙脉孢菌、间型脉孢菌(Neurospora intermedia)、里氏木霉、长枝木霉、变灰青霉(Penicillium canescens)、离生青霉、绳状青霉、亚拉巴马枝顶孢(Acremonium alabamense)、土生梭孢霉、嗜热侧孢霉(Sporotrichum thermophile)、Sporotrichum cellulophilum、球毛壳菌、Corynascus heterothallicus、埃默森篮状菌、Rasamsonia emersonii以及黄色蠕形霉(Talaromyces flavus)。

[0107] 经修饰的细胞可以是(进一步)遗传修饰的细胞或微生物,即除了编码SUMO化酶或

SUMO-蛋白质的基因的可能修饰和引入编码如本文详述的一种或多种目的蛋白质的转基因外,还包含(进一步)遗传修饰。可以进一步修饰细胞以产生减少的量的酶的生产或缺乏酶的生产,这些酶干扰/竞争目的蛋白质的表达和/或分泌和/或目的蛋白质的产生或酶活性。例如,可以修饰此类细胞以将干扰酶活性下调。下调可以例如通过引入酶活性的抑制剂(化学或生物学),通过操纵这些核酸分子被转录的效率、所得到的转录物被翻译的效率、和翻译后经修饰的效率,或通过基因突变、破坏或缺失来实现。可替代地,在酶生产或活性对目的酶的生产或活性具有正面影响的情况下,可以上调酶生产或活性的活性。此类酶可以是辅酶,例如在目的酶用于降解木质纤维素材料的情况下。本文还考虑下调一种或多种酶的活性,同时上调一种或多种酶的活性以实现期望的结果。

[0108] 经修饰的细胞可以是真菌细胞,或甚至是丝状真菌细胞。经修饰的细胞可以是选自自由以下组成的组的属的真菌细胞:金孢子菌属、梭孢壳属、脉孢菌属、短梗霉属、线黑粉菌属、单鞭毛菌属、棒囊壳属、隐球菌属、支顶孢属、弯颈霉属、柱顶孢霉、裂褶菌属、孢子丝菌属、青霉菌属、赤霉菌属、毁丝霉属、毛霉菌属、曲霉属、镰刀菌属、腐质霉属、木霉菌属、踝节菌属和*Rasamsonia*。经修饰的细胞可以是嗜热毁丝霉细胞,如以下的细胞:菌株C1(VKM F-3500 D)或由其衍生的突变体菌株(例如,UV13-6(登录号VKM F-3632 D);NG7C-19(登录号VKM F-3633 D);UV18-25(VKM F-3631D)、UV18#100f(CBS122188)、W1L(CBS122189)、或W1L#100L(CBS122190))。该菌株可以是具有蛋白酶和(半)纤维素酶的减少的表达的、并且可以甚至是不含蛋白酶和(半)纤维素酶表达和菌株。该菌株可以是W1L#100.1 Δ pyr5 Δ alp1,也称为LC菌株(Visser H等人,Ind.Biotechnol.[工业微生物学杂志]7:214-222,2011和WO 2010/107303,将其通过引用结合在此)。如在美国专利号6,015,707或美国专利号6,573,086中所述,称为C1(登录号VKM F-3500 D)的菌株分离自俄罗斯联邦远东索拉湖[Sola Lake, Far East of the Russian Federation]的森林碱性土壤样本。该菌株于1996年8月29日出于专利程序的目的,根据“布达佩斯条约”关于保藏微生物的国际规定的条款,被保藏在俄罗斯科学院所有俄罗斯微生物收藏馆(All-Russian Collection of Microorganisms of Russian Academy of Sciences (VKM)) (www.vkm.ru; Bakhurhina St.8, 莫斯科, 俄罗斯, 113184; Prospekt Nauki No.5, 普希诺, 莫斯科地区, 142290, 俄罗斯)(通过A.P.Sinitsyn, O.N.Okunev, I.V.Solov'eva, V.M.Chernoglasov, M.A.Emalfarb, A.Ben-Bassat; “FermTech” LTD Acad.Kapitsky str.32-2, 莫斯科, 117647, 俄罗斯), 称为卢克诺文思金孢子菌Garg 27K, VKM F-3500 D。已经生产了C1的各种突变菌株, 并且这些菌株于1998年9月2日出于专利程序的目的, 根据“布达佩斯条约”关于保藏微生物的国际规定的条款, 被保藏在俄罗斯科学院所有俄罗斯微生物收藏馆(All-Russian Collection of Microorganisms of Russian Academy of Sciences (VKM)), (Bakhurhina St.8, 莫斯科, 俄罗斯, 113184; Prospekt Nauki No.5, 普希诺, 莫斯科地区, 142290, 俄罗斯)(通过O.N.Okunev, A.P.Sinitsyn, V.M.Chernoglasov, 和M.A.Emalfarb; “FermTech” LTD, Acad.Kapitsy str., 32-2, 莫斯科, 117647, 俄罗斯), 或于2007年12月5日出于专利程序的目的被保藏在真菌菌种保藏中心(the Centraal Bureau voor Schimmelcultures (CBS)), Uppsalalaan 8, 3584CT Utrecht, 荷兰, (通过Dyadic Nederland B.V., Nieuwe Kanaal 7s, 6709PA瓦赫宁恩(Wageningen), 荷兰)。例如, 通过使菌株C1经受紫外光而被诱变以产生菌株UV13-6(登录号VKM F-3632 D; 1998年9月2日保藏于VKM)。随后用N-甲基-N'-硝基-N-

亚硝基胍进一步突变该菌株以产生菌株NG7C-19(登录号VKM F-3633 D;1998年9月2日保藏于VKM)。后一种菌株依次通过紫外光进行突变,产生菌株UV18-25(登录号VKM F-3631 D;1998年9月2日保藏于VKM)。UV18-25已经用紫外光突变并且选择低蛋白酶活性,在本文中表现为UV18#100f(CBS122188;2007年12月5日保藏于CBS)。菌株UV18-25也已经用紫外光突变并且选择低纤维素活性,产生菌株W1L(登录号CBS122189;2007年12月5日保藏于CBS),将该菌株W1L随后通过紫外光进行突变,并且选择低蛋白酶活性,产生菌株W1L#100L(登录号CBS122190;2007年12月5日保藏于CBS)。如在美国专利号6,015,707,美国专利号6,573,086和专利PCT/NL2010/000045中详细讨论的(每个通过引用结合在此),基于该微生物的形态特征和生长特征,菌株C1最初被分类为卢克诺文思金孢子菌。基于遗传测试,该C1菌株随后被重新分类为嗜热毁丝霉。在该文献中,卢克诺文思金孢子菌也作为嗜热侧孢霉(*Sporotrichum thermophile*)出现。在另外的实施例中,经修饰的细胞是显示高水平纤维素生产的菌株(HC菌株),如但不限于C1、UV13-6、NG7C-19、UV18-25或UV18#100f(登录号CBS122189),或其衍生物。如本文进一步详述的,衍生自UV18#100f并具有经修饰的SUMO化机器的经修饰的细胞(例如如本文定义的*ubc9*基因中的启动子插入)在纤维素酶的产生中特别受关注。如本文详述的具有经修饰的SUMO化机器的HC菌株的细胞可以进一步经遗传修饰以减少蛋白酶表达。经修饰的细胞可以进一步经遗传修饰,导致酶的活性降低或消除,这些酶导致形成纤维二糖内酯/纤维二糖酸和/或葡萄糖酸内酯/葡萄糖酸。此类进一步的遗传修饰的细胞或微生物可以是缺乏编码导致形成纤维二糖内酯/纤维二糖酸和/或葡萄糖酸内酯/葡萄糖酸的酶的功能基因的真菌细胞,例如,可以通过基因缺失、基因破坏、基因沉默或突变产生;或通过基因表达调控序列(如启动子序列、终止子序列、启动子激活序列和编码转录因子的序列)的缺失、破坏或突变;或者通过编码导致形成纤维二糖内酯/纤维二糖酸和/或葡萄糖酸内酯/葡萄糖酸的酶的基因的随机或定点突变。另外的经修饰的真菌可以是以下经修饰的真菌,其中编码负责产生选自纤维二糖内酯、纤维二糖酸、葡萄糖酸内酯和葡萄糖酸的一种或多种产物的酶(编码纤维二糖脱氢酶(CDH))的一种或多种基因被缺失、破坏或突变。该一种或多种基因可以编码具有纤维二糖脱氢酶(CDH)的氨基酸序列的酶,该纤维二糖脱氢酶选自与如W0 2013/159005 A2(将其通过引用结合在此)中公开的嗜热毁丝霉C1的内源性CDH1、CDH2和CDH3中的任一项具有至少90%、95%、或99%同一性的多肽的组。另外的经修饰的真菌可以是以下另外的经修饰的真菌,其中编码纤维二糖脱氢酶(CDH) CDH1、CDH2或CDH3的基因被缺失、破坏或突变。另外的经修饰的真菌可以是以下真菌,其中当通过如公开在W0 2013/159005 A2(将其通过引用结合在此)铁氰化物还原测定法测量时,CDH活性被降低约50%至约100%,或至少75%、90%或95%。可以敲除在嗜热毁丝霉C1中编码CDH1或编码CDH2的基因。在嗜热毁丝霉C1中编码CDH1和CDH2的基因可以被敲除(双敲除),如描述在W0 2013/159005 A2和W0 2012/061432 A1,将其通过引用结合在此。

[0109] 在实施例中,该宿主细胞是经修饰的里氏木霉细胞。经修饰的细胞可以包含*cbh1*、*cbh2*、*egl1*和*egl2*基因的一种或多种或全部的缺失(如描述在US 2015/0030717 A1中,将其通过引用结合在此)。经修饰的细胞可包含内切氨基葡萄糖苷酶基因的缺失。在实施例中,缺失防止分泌蛋白的去糖基化。

[0110] 另外的适合的宿主细胞包括昆虫细胞(最特别地是黑腹果蝇细胞、草地贪夜蛾Sf9和Sf21细胞以及粉纹夜蛾高-五细胞)、线虫细胞(特别是秀丽隐杆线虫细胞)、禽类细胞、两

栖动物细胞(特别是非洲爪蟾细胞)、爬行类动物细胞以及哺乳动物细胞(最特别是人、类人猿、犬、啮齿动物、牛或绵羊细胞,例如NIH3T3、CHO(中国仓鼠卵巢巢细胞)、COS、VERO、BHK、HEK及其他啮齿动物或人类细胞)。

[0111] 本发明还考虑遗传修饰的生物,如具有如本文详述的经修饰的SUMO化机器的藻类、和植物。这些植物可用于生产酶,和/或用作木质素、木质纤维素材料、纤维素材料和/或半纤维素材料(用作糖化方法的底物)。生产重组植物的方法是本领域已知的。例如,已经开发了用于植物转化的许多方法,包括生物和物理转化方案。参见例如,Miki等人,“Procedures for Introducing Foreign DNA into Plants”in *Methods in Plant Molecular Biology and Biotechnology*[在植物分子生物学和生物技术的方法中“用于将外源DNA引入植物的程序”],Glick,B.R.和Thompson,J.E.编辑(CRC Press,Inc.[CRC出版公司],Boca Raton[波卡拉顿],1993)第67-88页。此外,用于植物细胞或组织转化和植物再生的载体以及体外培养方法是可用的。参见例如,Gruber等人,“Vectors for Plant Transformation”in *Methods in Plant Molecular Biology and Biotechnology*[在植物分子生物学和生物技术的方法中“用于植物转化的载体”],Glick,B.R.和Thompson,J.E.编辑(CRC Press,Inc.[CRC出版公司],Boca Raton[波卡拉顿],1993)第89-119页。

[0112] 用于将表达载体引入植物的最广泛使用的方法是基于农杆菌属的天然转化系统。参见例如,Horsch等人,*Science*[科学]227:1229(1985)。用于农杆菌属介导的基因转移的农杆菌属载体系统和方法的描述由许多参考文献提供,包括Gruber等人,同上,Miki等人,同上,Moloney等人,*Plant Cell Reports*[植物细胞报道]8:238(1989),以及美国专利号4,940,838和5,464,763,两者均通过引用结合。

[0113] 植物转化的另一种普遍适用的方法是微粒介导的转化,参见例如,Sanford等人,*Part.Sci.Technol.*[部分科学技术]5:27(1987);Sanford,J.C.,*Trends Biotech.*[生物技术趋势]6:299(1988);Sanford,J.C.,*Physiol.Plant*[植物生理学]79:206(1990);Klein等人,*Biotechnology*[生物技术]10:268(1992)。用于将DNA物理递送到植物的另一种方法是对靶细胞进行超声处理。Zhang等人,*Bio/Technology*[生物/技术]9:996(1991)。可替代地,脂质体或原生质球融合已被用于将表达载体引入植物中。Deshayes等人,*EMBO J.*[欧洲分子生物学组织杂志]4:2731(1985);Christou等人,*Proc Natl.Acad.Sci.USA*[美国科学院院报]84:3962(1987)。还报道了使用CaCl₂沉淀、聚乙烯醇或聚-L-鸟氨酸将DNA直接吸收到原生质体中。Hain等人,*Mol.Gen.Genet.*[分子遗传学和基因组学]199:161(1985)和Draper等人,*Plant Cell Physiol.*[植物细胞生理学]23:451(1982)。还描述了对原生质体以及全细胞和组织进行的电穿孔。Donn等人,*In Abstracts of VIIth International Congress on Plant Cell and Tissue Culture IAPTC*[第七届国际植物细胞和组织培养大会摘要],A2-38,第53页(1990);D'Halluin等人,*Plant Cell*[植物细胞]4:1495-1505(1992)以及Spencer等人,*Plant Mol.Biol.*[植物分子生物学]24:51-61(1994)。

[0114] 进一步提供了用于生产如本文所定义的经修饰的细胞的方法。如上文详述,通过对SUMO化机器的至少一种酶或蛋白质进行遗传修饰或通过细胞暴露于SUMO化酶修饰剂来修饰经修饰的细胞以具有经修饰的SUMO化机器,以增加蛋白质生产。所述方法可包括通过插入、缺失或点突变破坏SUMO化机器蛋白的表达调控序列或编码序列(例如ubc9启动子序列)的步骤。所述方法可以包括靶向遗传修饰的至少一个步骤,其中所述步骤可以是插入

表达盒的步骤。应当理解，“靶向的”遗传修饰利用如此靶向的基因的序列信息来产生修饰。例如，靶向遗传修饰可以采用将具有足够程度的互补性或同源性的插入盒序列或反义序列或指导序列（如适合修饰策略）引入至靶基因序列以将修饰指向独特的目的基因。所述方法可以包括通过插入表达盒破坏SUMO化机器蛋白质的表达调控序列（例如ubc9启动子序列）或编码序列的步骤，其中通过在与嗜热毁丝霉的ubc9基因的启动子序列内的位置同源的位置上插入表达盒来破坏ubc9启动子序列，该位置由SEQ ID NO:1的位置1257和1258的核苷酸表示。同源位置在本文中理解为在与SEQ ID NO:1共享至少50%、60%、70%、80%、90%、或至少100%序列同一性的至少10、11、12、13、14、15、或20个核苷酸的连续序列或延伸内的位置，其中所述位置可以是连续延伸的中间或中心的位置。在本文详述的方法中用作起始材料的细胞或宿主细胞和在此方法中待修饰的任意的基因在上文详述。本发明的方法还可以包括将细胞孵育于或暴露于至少一种SUMO化酶或SUMO蛋白质的一种或多种修饰剂（例如抑制剂）中的步骤，如上文详述。

[0115] 还提供了包含如上文定义的经修饰的细胞的组合物。此类组合物可以是发酵液，该发酵液可以是粗发酵液。该发酵液可以进一步包含内源或外源（同源或异源）目的蛋白质或目的蛋白质的混合物。此类蛋白质或蛋白质的混合物可以是用于降解如上文详述的木质纤维素材料的一种或多种的酶。

[0116] 还提供了用于产生表达系统的方法，该方法包括转导如本文所定义的经修饰的细胞、或如本文所定义的用于产生此类经修饰的细胞的方法获得的细胞的步骤，该表达系统包含编码至少一种目的蛋白质的至少一种外源表达构建体。目的蛋白质可以是分泌蛋白，即，在由细胞产生后在细胞外环境中分泌的蛋白质。用每个表达不同的目的蛋白质的多个不同的表达构建体转导细胞和/或用编码多个不同目的蛋白质的单个表达构建体转导细胞。任选地，方法导致包含多个外源表达构建体和/或包含编码超过一种蛋白质的一种或多种表达构建体的细胞。编码的蛋白质可以是异源的或同源的。任选地，方法导致能够产生目的蛋白质的混合物的细胞。本领域技术人员将理解，使用重组DNA技术可以通过操纵例如宿主细胞内的核酸分子的拷贝数、那些核酸分子被转录的效率、所得到的转录物被翻译的效率以及翻译后修饰的效率来改善对转染的核酸分子的表达的控制。此外，与天然启动子相比，该启动子序列可能被基因工程化以提高表达水平。有用于控制核酸分子表达的重组技术包括但不限于，将核酸分子整合到一个或多个宿主细胞染色体中，向质粒添加载体稳定性序列，转录控制信号的替换或修饰（例如，启动子、操纵子、增强子），翻译控制信号（例如，核糖体结合位点）的取代或修饰，对核酸分子进行修饰以对应于宿主细胞的密码子使用，以及使转录物不稳定的序列的缺失。

[0117] 本发明不限于真菌，并且还考虑用本文公开的一种或多种核酸分子转化的遗传修饰的生物，如藻类和植物。植物可用于生产酶，和/或作为本发明的方法中用作底物的木质纤维素材料。生产重组植物的方法是本领域已知的。例如，已经开发了用于植物转化的许多方法，包括生物和物理转化方案。其参考文献在上文中指出。

[0118] 目的蛋白质或蛋白质混合物可以是如上文进一步详述的工业可应用蛋白质或蛋白质混合物。

[0119] 还提供的是用于产生目的蛋白质或目的蛋白质混合物的方法，该方法包括以下步骤：在有效产生蛋白质的条件下，培养如本文所定义的经修饰的细胞、或培养通过用于产生

如本文所定义的此类经修饰的细胞或表达系统的方法获得的经修饰的细胞。该蛋白质可以是内源性蛋白质或内源性蛋白质混合物、或者由如以上进一步详述的外源表达构建体编码的蛋白质或蛋白质混合物。当细胞已经用每个表达不同目的蛋白质的多个不同的表达构建体转导和/或当细胞已经用编码多个不同目的蛋白质的单个表达构建体转导时,用于产生如本文所定义的目的蛋白质或目的蛋白质混合物的方法将导致蛋白质混合物的生产。

[0120] 在一些情况下,可以回收(即,分离和/或纯化)该蛋白质,而在其他情况下,可以整体收获该细胞,该蛋白质和细胞中的任一种可以用于组合物中。本发明还提供了包含经修饰的细胞的组合物。在细胞是微生物的情况下,该细胞可以在适当的发酵培养基中培养,即其中微生物能够产生目的蛋白质的发酵培养基。微生物可以通过任何发酵方法进行培养,其包括但不限于分批、补料分批、细胞再循环和连续发酵。通常,真菌菌株在发酵罐中生长,任选地被离心或过滤以去除生物质,并任选地被浓缩、配制并干燥以产生作为粗发酵产物的蛋白质或多蛋白质组合物。用于培养丝状真菌的适合的条件描述于例如美国专利号6,015,707和美国专利号6,573,086中,将其通过引用结合在此。

[0121] 蛋白质回收是指收集含有蛋白质的整个培养基的过程,而不需要暗示另外的分离或纯化步骤。可以使用各种标准蛋白纯化技术来纯化产生的蛋白质,例如但不限于亲和层析、离子交换层析、过滤、电泳、疏水相互作用层析、凝胶过滤层析、反相色谱法、伴刀豆球蛋白A色谱法、聚焦色谱以及差异沉淀或增溶。蛋白质可以按“基本上纯的”形式回收、获得和/或使用。如本文所用,“基本上纯的”是指允许以根据本发明的任何方法有效使用该蛋白质的纯度。对于在其他应用中有益的蛋白质,它可以基本上不含可能干扰或会干扰其在进一步应用中使用的污染物、其他蛋白质和/或化学品。如本文所提及的“基本上纯的”蛋白质可以是可以通过任何方法(即,通过从天然来源直接纯化、重组地或合成地)产生的蛋白质,并且是已经从其他蛋白质组分纯化的蛋白质,使得该蛋白质在给定组合物中包含至少约80%重量/总蛋白重量(例如,目的蛋白质是溶液/组合物/缓冲液中的蛋白质的约80%),在给定的组合物中至少约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%、至少约99%或100%的重量/总蛋白重量。

[0122] 进一步提供了如本文定义的经修饰的细胞在用于产生如本文定义的至少一种目的蛋白质的方法中的用途。

[0123] 在以下实例中进一步解释本发明。这些实例不限制本发明的范围,而仅用于阐明。

[0124] 实例

[0125] 实例1:嗜热毁丝霉C1菌株的改善的蛋白质生产:破坏ubc9启动子区域

[0126] 在使用特定的纤维素酶表达盒(如在EP专利申请2408910中描述的),用W1L#100.1 Δ alp1 Δ chi1 Δ pyr5(产生第二代转化体)开始的两轮随后的共转化之后,诸位发明人发现意外高的蛋白质生产的嗜热毁丝霉C1菌株。转化体的遗传分析揭示了整合到基因的启动子区域中的整合事件,该基因注释为本文中由SEQ ID NO:1表示的泛素-缀合酶(E2(ubc9, NCBI登录号XP_003662268.1, Interpro ID IPR000608, Prosite Accession PS50127, 含有Pfam结构域PF00179)),其可能损害功能性表达。显示Ubc9在靶蛋白的SUMO化中起作用。

[0127] 为了更深入地了解第二代转化体的高蛋白质生产的表型,用第二代转化体进行差异基因表达(RNA-seq)分析并且与W1L#100.1 Δ alp1 Δ chi1 Δ pyr5进行比较。使用葡萄糖作

为碳进料进行具有两种菌株的碳限制补料分批培养。从这些培养物中,在开始进料后4小时、90小时和162小时取出菌丝体样品。从这些样品中,使用研磨的菌丝体作为起始材料,用SV总RNA分离系统(普洛麦格公司(Promega Corporation),美国)分离总RNA。随后使用Illumina平台将分离的总RNA送去测序。使用标准工具(TopHat,Cuffdiff)和定制脚本的组合分析和可视化配对末端链RNA测序数据。如在FPKM(每千碱基百万的片段)中测量并从126nt长的配对末端Illumina RNA-seq读数(每个总文库大小4.7M-8.2M末端配对)获得的表达呈现在表1中。来自差别基因表达分析的结果揭示了基因 $ubc9$ (涉及SUMO-缀合途径的SUMO缀合酶)的下调。在所有时间点始终显示 $ubc9$ 表达的降低,其中降低水平为95.4%至99%,对应于减少至少4.3倍(在 $t=4h$ 时)。

[0128] 表1. 使用RNA测序,用差别基因表达分析获得的FPKM值。

	W1L#100.1 $\Delta alp1\Delta chi1\Delta pyr5$	第二代转化体
[0129] 4 小时	83.0	3.8
90 小时	126.3	1.3
162 小时	95.4	2.9

[0130] 实例2:嗜热毁丝霉C1菌株的 $ubc9$ 基因启动子区域的靶向破坏。

[0131] 为了测试整合事件对蛋白质生产的影响,将 $amdS$ 选择标记插入嗜热毁丝霉C1菌株的相同基因座处,即插入SEQ ID NO:1的位置1257和1258上的核苷酸之间的 $ubc9$ 启动子区域。更具体地,使用的菌株衍生自W1L#100.1 $\Delta alp1\Delta chi1\Delta pyr5$ 。W1L#100.1 $\Delta alp1\Delta chi1\Delta pyr5$ 衍生自W1L#100.1(保藏在no CBS122190下),通过破坏 $alp1$ 和 $chi1$ 基因,如W0 2010/107303中描述的(通过引用结合在此,更具体地,在W02010/107303的实例5中)。

[0132] 通过将W1L#100.1 $\Delta alp1\Delta chi1\Delta pyr5$ 用编码Eg2的表达盒(在此由SEQ ID NO:4表示)和 $pyr5$ 选择标记(在此由SEQ ID NO:7表示)共转染产生转化体。将获得的转化体称为W1L#100.1 $\Delta alp1\Delta chi1\Delta pyr5$ [$eg2/pyr5$]。随后,将由SEQ ID NO:3代表的 $Ubc9-amdS$ 插入盒靶向至W1L#100.1 $\Delta alp1\Delta chi1\Delta pyr5$ [$eg2/pyr5$]中的特定的 $ubc9$ 基因座,类似于先前由Visser等人(2011)描述的产生基因缺失的程序。构建的 $amdS$ 插入盒由来自构巢曲霉的乙酰胺酶($amdS$)选择标记基因(包括其自身的启动子和终止子序列)组成,侧翼是 $ubc9$ 插入位点(SEQ ID NO:1的核苷酸位置1257-1258)的1257和1380bp(分别为上游和下游侧翼)的序列。这些插入盒进一步包含 cbh 重复序列,其与本实例无关,但可用于在将来的应用中去除任选的选择标记序列。 $amdS$ 的插入在携带与基因组中的插入位点相同的序列的侧翼序列的同源重组时发生。这样,将 $amdS$ 选择标记在特定基因座处插入W1L#100.1 $\Delta alp1\Delta chi1\Delta pyr5$ [$eg2/pyr5$]基因组中。

[0133] 使用菌落PCR筛选转化体,将其设置为扩增与插入位点的一部分结合的 $amdS$ 基因的一部分。只有当 $amdS$ 基因正确地插入正确的基因座时,才有可能进行扩增。这样,携带正确地插入的 $amdS$ 的转化体可以与具有在基因组中其他地方整合的选择标记的转化体区分开。

[0134] 确认的转化体菌株在微量滴定板和摇瓶中生长,以测试 $amdS$ 的插入对细胞外蛋白质生产的影响。通过SDS-PAGE目测观察证实了改善的蛋白质生产和分泌。随后的发酵实验(使用葡萄糖作为碳进料的碳限制分批培养)进一步证实了改善的蛋白质生产。用BCA测定

(Pierce BCA蛋白质测定试剂盒(赛默飞世尔科技公司(ThermoFisher Scientific)))针对细胞外蛋白质含量的蛋白质浓度测量显示转化体菌株的更高的蛋白质生产水平,其中蛋白质水平从15.6g/L增加至35.5g/L(图1)。此外,确定两种菌株的 Q_p 最大值(表2)。 Q_p 是特定的蛋白质生产速率(克蛋白质/克生物质/小时),其是通过将总蛋白质生产速率(以克/小时计)除以在那时反应器中存在的生物质的量(细胞干重)来计算。从总蛋白质生产与时间的斜率来计算总蛋白质生产速率。 Q_p 最大值从在亲本菌株中的0.011g/g/h增加至在 $ubc9$ 经修饰的菌株中的0.015g/g/h。

[0135] 表2. 在 $ubc9$ 启动子区域中有或没有破坏的总蛋白质和 Q_p

[0136]

菌株	总蛋白质 (g/L)	Q_p 最大 (克蛋白质/克生物质/小时)
W1L#100.1 $\Delta alp1\Delta chi1\Delta pyr5[eg2/pyr5]$	15.6	0.011
W1L#100.1 $\Delta alp1\Delta chi1\Delta pyr5[eg2/pyr5]$ $ubc9::amdS$	35.5	0.015

[0137] 实例3:用嗜热毁丝霉C1菌株的改善的异源蛋白质生产:破坏 $ubc9$ 启动子区域

[0138] 为了证明 $ubc9$ 启动子(Pubc9)的遗传修饰导致来自嗜热毁丝霉C1的以外的微生物的蛋白质的改进的生产水平,将Pubc9突变-如在实例1中描述的-引入从Talaromyces wortmanii W1L#100.1 $\Delta alp1 \Delta chi1 \Delta pyr5[TwBx11/pyr5]$ 表达异源蛋白质Bx11(β -木糖苷酶)和从黑曲霉(W1L#100.1 $\Delta chi1 \Delta pyr5[AnPGII/pyr5]$)表达PGII(聚半乳糖醛酸酶)的菌株W1L#100.1 $\Delta alp1 \Delta chi1 \Delta pyr5$ 中。

[0139] 通过在W1L#100.1 $\Delta alp1 \Delta chi1 \Delta pyr5$ 中,将 $pyr5$ 选择标记(在此由SEQ ID NO:7表示)与编码TwBx11或AnPGII的表达盒共转染,从而产生转化体W1L#100.1 $\Delta alp1 \Delta chi1 \Delta pyr5[TwBx11/pyr5]$ 或W1L#100.1 $\Delta alp1 \Delta chi1 \Delta pyr5[AnPGII/pyr5]$,产生转化体。此表达盒(在此分别由SEQ ID NO:6和5表示)进一步包含 $chi1$ 基因启动子和 $cbh1$ 终止子序列,它们通常用于获得相应基因的高表达水平(Visser等人,2011)。随后,如在实例2中进一步详述,将由SEQ ID NO:3代表的Ubc9- $amdS$ 插入盒靶向至菌株W1L#100.1 $\Delta alp1 \Delta chi1 \Delta pyr5[TwBx11/pyr5]$ 和W1L#100.1 $\Delta alp1 \Delta chi1 \Delta pyr5[AnPGII/pyr5]$ 中的 $ubc9$ 启动子。

[0140] 使用菌落PCR筛选转化体,将其设置为扩增与插入位点的一部分结合的 $amdS$ 基因的一部分。只有当 $amdS$ 基因正确地插入正确的基因座时,才有可能进行扩增。这样,携带正确地插入的 $amdS$ 的转化体可以与具有在基因组中其他地方整合的选择标记的转化体区分开。

[0141] 确认的转化体菌株在微量滴定板和摇瓶中生长,以测试 $amdS$ 的插入对细胞外蛋白质生产的影响。通过SDS-PAGE目测观察证实了改善的蛋白质生产和分泌(图2和图3)。

[0142] 实例4:嗜热毁丝霉HC C1菌株的改善的蛋白质生产:破坏 $ubc9$ 启动子区域

[0143] 为了证明在另一个嗜热毁丝霉世系中的Pubc9($ubc9$ 的启动子区域)的遗传修饰也导致改善的生产水平,将Pubc9突变-如实例1中描述的-引入HC-菌株UV18-25中。

[0144] 在实例2中进一步详述了,将由SEQ ID NO:3代表的 $amdS$ 选择标记引入菌株UV18-

25中的ubc9启动子区域。

[0145] 使用菌落PCR筛选转化体,将其设置为扩增与插入位点的一部分结合的amdS基因的一部分。只有当amdS基因正确地插入正确的基因座时,才有可能进行扩增。这样,携带正确地插入的amdS的转化体可以与具有在基因组中其他地方整合的选择标记的转化体区分开。

[0146] 确认的转化体菌株在微量滴定板和摇瓶中生长,以测试amdS的插入对细胞外蛋白质生产的影响。通过SDS-PAGE目测观察证实了改善的蛋白质生产和分泌。

[0147] 在发酵实验(如在实例2中描述的)中测试确认的转化体菌株以证实改善的蛋白质生产。用BCA测定(如在实例2中描述的)针对细胞外蛋白质含量的蛋白质浓度测量显示转化体菌株的更高的蛋白质生产水平,其中蛋白质水平从39.6g/L增加至47.5g/L。此外,如实例2中描述(表3),确定两种菌株的最大特定的蛋白质生产速率(Q_p最大值)。

[0148] 表3. 从Labfors (2.5L) 发酵获得的总蛋白质浓度(g/L)和Q_p最大(g/g/h)

[0149]

	总蛋白质 (g/L)	Q _p 最大 (克蛋白质/克生物质/小时)
UV18-25	39.6	0.016
UV18-25 ubc9::amdS	47.5	0.016

[0150] 测试来自菌株UV18-25和UV18-25ubc9::amdS的酶混合物的糖化效率。将NREL预处理的玉米秸秆(Schell等人,J Appl Biochem Biotechnol[应用生物化学与生物技术], 105:69-86,2003)用作底物。使用干物质含量为20%的底物,在50mL Greiner管中以10mL的总体积进行糖化反应。使用终浓度为100mM的乙酸盐缓冲液在pH 5下进行糖化反应。向每个反应中添加预定量的氢氧化钠,以将pH设为5。包括另外的反应(15mg/g)以每24小时监测pH。当此另外的反应的pH偏差超过0.1个pH单位时,添加2M NaOH以将pH校正至其初始起始pH。随后将调节pH所需的体积添加至其他反应管中并立即混合。叠氮化钠的剂量为0.02%(w/w)。在2.5mg/g DM、5mg/g DM、10mg/g DM和15mg/g DM的酶载量(剂量响应)下测试酶混合物。将来自嗜热毁丝霉C1(SEQ ID NO:16)纯化的Bg11添加至每个反应中(10%的基线加载)。使用上述BCA测定值,应用酶载量。将反应在52°C在300rpm孵育。在24h、和68h之后,取出0.2ml样品并使用微量板(pvdf)过滤,随后分析上清液。使用GPOD测定试剂盒(梅格泽姆公司(Megazyme),威克洛,爱尔兰)测量葡萄糖浓度。所有实验均一式两份进行。UV18-25ubc9::amdS的蛋白质混合物的糖化效率显示在图5中。

[0151] 表4 SEQ ID ON 1-7的序列概述

[0152]

SEQ ID NO	名称	特定区域	
		名称	位置
1	具有侧翼序列的 <i>ubc9</i> 基因	5'UTR	1070-1301
		<i>ubc9</i> ORF	1302-1994
		3'UTR	1995-2195
		整合位点	1257-1258
2	蛋白质序列 Ubc9		
3	<i>ubc9-amdS</i> 插入盒	<i>ubc9</i> 上游侧翼	7-1263
		<i>cbh</i> 重复序列	1269-1860
		<i>amdS</i>	1871-4592
		<i>cbh</i> 重复序列	4594-5184
		<i>ubc</i> 下游侧翼	5193-6572
4	表达盒 <i>eg2</i>	<i>Pchi-ext</i> :	31-1074
		<i>Pchi</i> :	1075-1849
		<i>eg2</i> :	1850-3467
		<i>Tcbh</i> :	3468-4483
5	表达盒 <i>AnPGII</i>	<i>Pchi-ext</i>	4-1047
		' <i>Pchi</i>	1048-1822
		<i>AnPGII</i>	1823-2913
		<i>Tcbh</i> '	2920-3935
6	表达盒 <i>TwBxl1</i>	<i>Pchi-ext</i>	31-1074
		' <i>Pchi</i>	1075-1848
		<i>TwBxl1</i>	1850-4245
		<i>Tcbh</i> '	4252-5267
7	<i>pyr5</i> 选择标记	<i>pyr5</i>	1173-1877

[0153] 实例5:表达异源葡糖淀粉酶的嗜热毁丝霉C1菌株的*ubc9*基因启动子区域的破坏

[0154] 通过用编码里氏木霉葡糖淀粉酶(指定为“DP1”)的表达盒(其DNA编码序列被修饰用于在C1菌株中表达)(在此分别由SEQ ID NO:17和SEQ ID NO:18表示)、以及*pyr5*选择标记(在此由SEQ ID NO:7表示)共转染菌株W1L#100.1 Δ *alp1* Δ *chi1* Δ *pyr5*和菌株UV18#100.f Δ *alp1* Δ *pyr5*,产生转化体。用于W1L#100.1 Δ *alp1* Δ *chi1* Δ *pyr5*菌株的表达盒包括*chi1*启动子序列(SEQ ID NO:51),并且用于UV18#100.f Δ *alp1* Δ *pyr5*菌株的表达盒包括*cbh1*启动子序列(SEQ ID NO:52)。使用来自梅格泽姆公司(Megazyme)(威克洛,爱尔兰)的 β -淀粉酶测定试剂盒(Betamyl-3)筛选获得的转化体的葡糖淀粉酶活性。选择具有最高葡糖淀粉酶活性的转化体并分别命名为W1L#100.1 Δ *alp1* Δ *chi1* Δ *pyr5*[DP1/*pyr5*]和UV18#100.f Δ *alp1* Δ *pyr5*[DP1/*pyr5*]。随后,类似于实例2中描述的,将在这些转换体中的*ubc9*启动子区域进行修饰。

[0155] 在通过菌落PCR(如实例2中描述的)确认*ubc9*启动子区域被修饰之后,将经修饰的

菌株及其表达葡糖淀粉酶的亲本菌株在Minifors (1L) 或Labfors (2.5L) 发酵罐(使用葡萄糖作为碳源的碳限制分批进料培养) 中生长, 以测试细胞外葡糖淀粉酶生产。使用对硝基苯基 α -D-吡喃葡萄糖苷 (pNPG; 西格玛公司 (Sigma) N-1377) 作为底物进行葡糖淀粉酶活性测定。通过将用pNPG测定测量的葡糖淀粉酶单位 (GAU) 转化为葡糖淀粉酶克/升 (g/L) 来计算由嗜热毁丝霉C1产生的葡糖淀粉酶的水平。为此, 通过对里氏木霉生产的葡糖淀粉酶的样品 (其中蛋白质浓度是已知的) 进行相同的pNPG测定来确定转换因子。此转换因子随后用于将在嗜热毁丝霉培养物中测量的将GAU转化为葡糖淀粉酶浓度。ubc9经修饰的嗜热毁丝霉菌株显示更高的葡糖淀粉酶生产水平 (表5B和6B)。

[0156] 来自W1L#100.1 Δ alp1 Δ chi1 Δ pyr5 [DP1/pyr5] ubc9经修饰的菌株的总蛋白质浓度 (用如实例2中描述的BCA测定测量的) 更高 (表5A)。来自UV18#100.f Δ alp1 Δ pyr5 [DP1/pyr5] ubc9经修饰的菌株的总蛋白质水平在Minifors (1L) 发酵中不增加, 但在Labfors (2.5L) 发酵中确实显示增加的蛋白质水平 (表6A)。例外, 计算 Q_p 最大值 (如实例2中描述的), 其报道在表7中。

[0157] 表5A: 用表达DP1的具有和不具有ubc9基因启动子破坏的菌株W1L#100.1 Δ alp1 Δ chi1 Δ pyr5产生的总蛋白质浓度

	1 升*	2.5 升
[0158] W1L#100.1 Δ alp1 Δ chi1 Δ pyr5 [DP1]	6.59 g/L	20.09 g/L
W1L#100.1 Δ alp1 Δ chi1 Δ pyr5 [DP1] ubc9::amdS	25.55 g/L	26.57 g/L

[0159] *是指用于评估菌株的发酵规模

[0160] 表5B: 用表达DP1的具有和不具有ubc9基因启动子破坏的菌株W1L#100.1 Δ alp1 Δ chi1 Δ pyr5产生的葡糖淀粉酶浓度

	1 升	2.5 升
[0161] W1L#100.1 Δ alp1 Δ chi1 Δ pyr5 [DP1]	4.3 g/L	16.4 g/L
W1L#100.1 Δ alp1 Δ chi1 Δ pyr5 [DP1] ubc9::amdS	19.2 g/L	25.6 g/L

[0162] 表6A: 用表达DP1的具有和不具有ubc9基因启动子破坏的菌株UV18#100.f Δ alp1 Δ pyr5产生的总蛋白质浓度

	1 升	2.5 升
[0163] UV18#100.f Δ alp1 Δ pyr5 [DP1]	53.37	72.61
UV18#100.f Δ alp1 Δ pyr5 [DP1] ubc9::amdS	52.08	80.01

[0164] 表6B用表达DP1的具有和不具有ubc9基因启动子破坏的菌株UV18#100.f Δ alp1 Δ pyr5产生的葡糖淀粉酶浓度

		1 升	2.5 升
[0165]	UV18#100.f $\Delta alp1 \Delta pyr5$ [DP1]	17.0	33.1
	UV18#100.f $\Delta alp1 \Delta pyr5$ [DP1] ubc9::amdS	24.7	34.4

[0166] 表7. 基于Labfors (2.5L) 发酵的总蛋白质浓度 (g/L) 和Qp最大值 (g/g/h)

菌株	总蛋白 质 (g/L)	Qp 最大 (克蛋白质 /克生物质/ 小时)
[0167] W1L#100.l $\Delta alp1 \Delta chi1 \Delta pyr5$ [DP1]	20.09	0.004
W1L#100.l $\Delta alp1 \Delta chi1 \Delta pyr5$ [DP1] ubc9::amdS	26.57	0.006
UV18#100.f $\Delta alp1 \Delta pyr5$ [DP1]	72.61	0.017
[0168] UV18#100.f $\Delta alp1 \Delta pyr5$ [DP1] ubc9::amdS	80.01	0.017

[0169] 实例6: 表达异源植酸酶的嗜热毁丝霉C1菌株的ubc9基因启动子区域的靶向破坏

[0170] 为了证明Pubc9 (ubc9的启动子区域) 的遗传修饰还导致异源植酸酶的改善的生产水平, 将Pubc9突变引入-如实例2中描述的-表达布丘氏菌属物种植酸酶变体 (蛋白质SEQ ID NO:21) 的C1菌株W1L#100.l $\Delta alp1 \Delta chi1 \Delta pyr5$ 和UV18#100.f $\Delta alp1 \Delta pyr5$ 中。

[0171] 通过将菌株W1L#100.l $\Delta alp1 \Delta chi1 \Delta pyr5$ 和UV18#100.f $\Delta alp1 \Delta pyr5$ 用编码植酸酶的表达盒 (在此分别由SEQ ID NO:53和SEQ ID NO:54表示) 和pyr5选择标记 (在此由SEQ ID NO:7表示) 共转染产生转化体。用于W1L#100.l $\Delta alp1 \Delta chi1 \Delta pyr5$ 菌株的表达盒包括chi1启动子序列 (SEQ ID NO:51), 并且用于UV18#100.f $\Delta alp1 \Delta pyr5$ 菌株的表达盒包括cbh1启动子序列 (SEQ ID NO:52)。将具有植酸酶活性升高比未转化亲本更高的获得的转化体分别称为W1L#100.l $\Delta alp1 \Delta chi1 \Delta pyr5$ [植酸酶/pyr5] 和UV18#100.f $\Delta alp1 \Delta pyr5$ [植酸酶/pyr5]。随后, 类似于实例2中描述的, 将在这些转换体中的ubc9启动子区域进行修饰。

[0172] 将确认的ubc9经修饰的菌株及其表达植酸酶的亲本菌株在发酵罐 (使用葡萄糖作为碳进料进行具有两种菌株的碳限制补料分批培养) 中生长以测试ubc9启动子修饰对细胞外植酸酶生产的影响。

[0173] 植酸酶活性测量 (US 2015/0030717, 通过引用结合在此) 显示ubc9经修饰的菌株的更高的活性水平, 其中对于W1L#100.l $\Delta alp1 \Delta chi1 \Delta pyr5$ [植酸酶/pyr5], 活性水平从2.1kU/g增加至5.9kU/g, 并且对于UV18#100.f $\Delta alp1 \Delta pyr5$ [植酸酶/pyr5], 活性水平从11.3kU/g增加至16.1kU/g (图6)。此外, 来自ubc9经修饰的菌株的总蛋白质水平更高, 其中对于W1L#100.l $\Delta alp1 \Delta chi1 \Delta pyr5$ [植酸酶/pyr5], 细胞外蛋白质水平从14g/Kg增加至20g/Kg, 并且对于UV18#100.f $\Delta alp1 \Delta pyr5$ [植酸酶/pyr5], 细胞外蛋白质水平从64g/Kg增加至75g/Kg (图6)。

[0174] 实例7:过表达异源聚半乳糖醛酸酶II的黑曲霉菌株的ubc9基因启动子区域的靶向破坏

[0175] 用于此研究的黑曲霉衍生自N402 (Bos等人,1988)。通过将菌株AB4.1 (N402的pyrG⁻衍生物) (van Hartingsveldt等人,1987)用5 μ g pAB4.1 (van Hartingsveldt等人,1987)和50 μ g pGpdA-pgaII^{C1}共转染来获得表达pGpdA-pgaII^{C1}的黑曲霉菌株。表达载体pGpdA-pgaII^{C1} (SEQ ID NO:23)含有用于来自黑曲霉的聚半乳糖醛酸酶II的编码序列(将其经密码子优化用于在嗜热毁丝霉C1中表达) (SEQ ID NO:24;蛋白质SEQ ID NO:25)和常用的GpdA启动子以驱动基因表达。如所述的(Arentshorst等人,2012)进行黑曲霉的转染。通过在选择性基本培养基上进行两轮纯化,来纯化转化体(Bennet和Lasure,1991)。在纯化之后,将各个转化体在微量滴定板孔(200 μ l基本培养基)中生长,并且使用PAHBAH测定(见下文)来测定培养基的半乳糖醛酸酶活性。显示PgaII表达的菌株在完全培养基上生长(Arentshorst等人,2012),并在分生孢子分离之前孵育4-5天。使用Biorad细胞计数器计数孢子并在4 $^{\circ}$ C在0.9%NaCl中储存。使用孢子悬浮液以 1.0×10^6 孢子/ml的密度接种摇瓶培养物(50ml基本培养基在300ml锥形瓶中),并且在37 $^{\circ}$ C、250rpm下生长36h。

[0176] 通过将ubc9敲入表达盒(SEQ ID NO:26)的amdS选择转染至黑曲霉菌株AN[PGaII]_高和AN[PGaII]_低中获得Ubc9-启动子(基因标识符An04g01350,在AspGD中)敲入菌株。设计敲入片段使得在ubc9起始-ATG的上游的45个碱基对处引入amdS选择标记,这类类似于用于在嗜热毁丝霉中敲入amdS选择标记的策略。敲入盒含有amdS标记,该amdS标记由5'ubc9上游区域的956bp和ubc9上游的983bp和编码区侧翼。将5863bp片段用于转染菌株AN[PGaII]_高和AN[PGaII]_低。如在Arentshorst等人,2012中描述的,在用于amdS选择的培养基上选择转化体菌株。将转化体在选择性基本乙酰胺板上纯化两次,然后将选择的菌株在完全培养基上生长以获得孢子,来接种液体培养物。

[0177] 如前所述分离基因组DNA(Meyer等人,2010),并且用于使用引物ubc9P1f和ubc9P4r的诊断PCR。用Phire Hot Start II DNA聚合酶或Phusion DNA聚合酶(赛默飞世尔科技公司(Thermo Scientific))进行PCR。将ubc9P1f (GATAGTGCTTAGCGACACCCG;SEQ ID NO:27)退火至amdS标记,并将ubc9P4r (TCGGCCACAAATCTCAGTGA;SEQ ID NO:28)退火至3'ubc9区域。预期在ubc9启动子中成功整合敲入片段会产生2222bp的PCR片段,而当片段异位插入时不会形成产物。

[0178] 根据(Sambrook和Russell 2001)进行Southern印迹分析。根据制造商的说明,使用Decalabel DNA标记试剂盒(赛默飞世尔科技公司(Thermo Scientific),沃尔瑟姆,马萨诸塞州)合成 α -³²P-dCTP-标记的探针。用BamHI消化黑曲霉菌株的基因组DNA,并用覆盖ubc9编码区的1000个核苷酸的探针标记。限制性内切酶从赛默飞世尔科技公司(Thermo Scientific)获得并根据制造商的说明书使用。

[0179] 如前所述进行生物反应器培养(Jorgensen等人,2010;Nitsche等人,2012),使用0.75% (w/v)葡萄糖作为碳源。

[0180] 使用聚半乳糖醛酸作为底物进行PAHBAH非还原糖测定以确定相对聚半乳糖醛酸酶活性(如M.Lever,1972所述)。样品的酶活性定义为在5分钟内每升释放的半乳糖醛酸的量除以每升反应体积的孵育时间(参见公式)。酶活性以单位每克蛋白质(U/g)显示。在稀释的样品上确定酶活性,这些样品在标准曲线(0mM-2mM半乳糖醛酸)的范围内。

$$[0181] \quad \text{酶活性} = \frac{X * V_r * D_e * 1000}{V_e * T}$$

[0182] X=mmol半乳糖醛酸/升

[0183] Vr=反应体积(以升计)

[0184] De=在添加至反应混合物之前的酶稀释

[0185] Ve=添加至反应混合物中的酶体积(以ml计)

[0186] Ec=储备溶液中的酶/蛋白质浓度(以mg/ml计)

[0187] T=孵育时间

[0188] 在使用PAHBAH非还原糖测定和PGA作为底物,在微量滴定板中培养后,测定用pAB4.1与pGpdA-pgaII^{C1}共转染之后获得的来自黑曲霉转化体的培养基样品的内切-聚半乳糖醛酸酶活性。选择与未转换的菌株相比,具有至少高3倍的活性的12个转化体,并在摇瓶培养物中再次生长用于可靠的定量活性测量。基于这些活性测量,选择两个菌株(An[PgaII]_高(110±10AU)和An[PgaII]_低(30±3AU)),其对PGA显示出远高于亲本菌株N402(4±1AU)的活性。选择菌株An[PgaII]_高和An[PgaII]_低作为高聚半乳糖醛酸酶和低半乳糖醛酸酶产生黑曲霉菌株,并且用于随后的实验。

[0189] 将菌株An[PgaII]_高和菌株An[PgaII]_低用ubc9::amdS启动子敲除构建体转化,并且选择和纯化能够在乙酰胺板上生长的转化体。分离转化体的基因组DNA,并且通过诊断PCR和Southern印迹分析以鉴定转化体,其中将ubc9::amdS启动子敲除片段整合在预期的位点。发现在An[PgaII]_高背景中的两个转化体(An[PgaII]_高ubc9::amdS#8和#17)和在An[PgaII]_低背景中的一个转化体(An[PgaII]_低ubc9::amdS#1)在预期的位点含有片段并且不含有另外的插入。

[0190] 选择的菌株An[PgaII]_高和An[PgaII]_低和衍生物An[PgaII]_高ubc9::amdS#8和#17和An[PgaII]_低ubc9::amdS#1分别在如之前描述的摇瓶中生长,并使用PAHBAH测定来测定内切-聚半乳糖醛酸酶活性。表8中给出了五种菌株对PGA的相对非还原活性。

[0191] 在生物反应器中培养An[PgaII]_高和衍生物An[PgaII]_高ubc9::amdS#17的酶生产显示在表9中。

[0192] 表8.在摇瓶培养后(48h,30°C),选择的转化体的培养基样品中对PGA的相对非还原活性

菌株	酶活性 (AU)	与亲本菌株相比的倍数增加
An[PgaII]_高	184 ± 31	不能提供
An[PgaII]_高	186 ± 28	1.01

[0193]

[0194]	ubc9::amdS #8		
	An[PgaII]_高 ubc9::amdS #17	179 ± 27	0.97
	An[PgaII]_低	30.2 ± 10.4	不能提供
	An[PgaII]_低 ubc9::amdS #1	32.3 ± 12.5	1.07

[0195] 表9. 在生物反应器培养 (pH3.0, 30°C) 中指数生长细胞期间, 选择的转化体的培养基样品中对PGA的相对非还原活性

时间点 *	An[PgaII]_高 _1 BR_JR1	An[PgaII]_高_2 BR_JR2	An[PgaII]_高 Ubc9::amdS #17_1 BR_JR3	An[PgaII]_高 Ubc9::amdS #17_2 BR_JR4
1	220	250	260	260
2	300	240	260	260

[0197] *为了能够比较不同的培养物, 在添加大约26ml (时间点1) 或30ml (时间点2) 的碱添加之后, 从生物反应器中取出样品。碱的添加与生物质积累成线性。

[0198] 实例8. 里氏木霉中的ubc9启动子的破坏

[0199] 构建包含以下DNA区段的质粒 (pLH937 (SEQ ID NO:29)) :

[0200] 1. 细菌载体, 其包含大肠杆菌ccdB基因、氨苄青霉素抗性基因和用于在大肠杆菌中选择和扩增的复制起点。

[0201] 2. 潮霉素磷酸转移酶可读框, 其有效地融合到粗糙脉孢菌cpc1启动子和构巢曲霉trpC终止子。

[0202] 选择里氏木霉ubc9启动子中的三种不同的Cas9靶位点。它们接近上述实例中用于嗜热毁丝霉 (C1) 的插入位点。

[0203] 靶位点1:GCAGTTCGACGCTTACCCACCGG (SEQ ID NO:30)

[0204] 靶位点2:CGACGCTTACCCACCGGGTGAGG (SEQ ID NO:31)

[0205] 靶位点3:GCGCGACTACCATCACGTCTCGG (SEQ ID NO:32)

[0206] 在每种情况下, 前20个核苷酸代表Cas9识别的原型间隔子区, 并且最后三个核苷酸是PAM (原型间隔子邻近基序)。基于这些靶序列产生用于与Cas9一起组装的指导RNA。

[0207] 将分泌布丘氏菌属物种植酸酶的工程化的变体 (BP17) (SEQ ID NO:33) 的里氏木霉菌株用于这些实验中。此菌株具有四种主要分泌纤维素酶基因 (cbh1、cbh2、egl1和egl2) 的缺失连同内切氨基葡萄糖苷酶基因的缺失, 以防止分泌蛋白的去糖基化。菌株构造描述于US 2015/0030717A1、实例1和2中。使用描述于US 2015/0030717 (将其通过引用结合在此) 中的方法制备此菌株的原生质体。

[0208] 将Cas9.RNP复合物和质粒pLH937共转染至里氏木霉原生质体中

[0209] 将Cas9蛋白质单独与三种不同的指导RNA中的每种混合, 并且在缓冲液中孵育, 以允许组装Cas9.RNP复合物。如WO 2016100568A1 (通过引用结合在此) 中描述的进行聚乙二醇 (PEG) 介导的里氏木霉原生质体的转化以使能够摄取Cas9.RNP和质粒DNA。

[0210] 将25ul Cas9.RNP、3ul pLH937质粒(0.3ug/uL)、1.2ul **Lipofectamine®** CRISPR-MAX转染试剂(赛默飞世尔公司(Thermo Fisher))混合并添加至原生质体中用于摄取。将Vogel琼脂培养基(具有1.1M山梨糖醇、2%葡萄糖和50ug/mL潮霉素B)倒在固化的Vogel琼脂培养基(具有1.1M山梨糖醇和20%葡萄糖但不含潮霉素)上,在转化程序后,将原生质体铺在其覆盖物上。

[0211] 将在潮霉素选择板上产生的转化的菌落转移到不含山梨糖醇或潮霉素的非选择性Vogel琼脂上,并在30°C孵育4天。然后将菌落在具有50ug/mL潮霉素的选择性Vogel琼脂上传代两次。那些生长良好的转化菌落被认为具有稳定的潮霉素抗性表型。预期,这些稳定转化体中的很多将具有整合在`ubc9`基因的启动子中的`cas9`靶位点处的pLH937、或其一些片段。从转化体分离基因组DNA。使用引物`nik1 2kb F`(5'-gccatgaacctcaccacacag;SEQ ID NO:34)和`nik1 2kb R`(5'-cggcgtggccctgttctcgag;SEQ ID NO:35)进行PCR以扩增里氏木霉`nik1`基因座的2kb区域。琼脂糖凝胶电泳证实所有样品都含有足够数量和质量的gDNA,来充当PCR中的模板。

[0212] 使用引物3078(5'-aagagatctctgccctcccaggg;SEQ ID NO:36)和3079(5'-gtgatggccggcttccagg;SEQ ID NO:37)进行PCR,以扩增跨越`cas9`靶位点的`ubc9`的启动子区域。根据制造商的指导,在50°C的退火温度和1分钟的延伸时间下,用Stratagene PfuUltra II Fusion HS DNA聚合酶(安捷伦科技公司(Agilent),圣克拉拉(Santa Clara),加利福尼亚州)进行PCR。使用这些条件,如果在`ubc9`启动子的`cas9`靶位点处存在大的插入物,则预期不会获得PCR产物。给出具有野生型`ubc9`基因座预期大小的PCR产物的那些转化体从进一步分析中消除。下面描述的所有转化体均未给出PCR产物,表明它们都具有插入在靶位点处的HygR载体。使用引物3078和3145(5'-ctttgccctcggacgagtgct;SEQ ID NO:38)进行PCR,以在`ubc9`启动子的侧翼区和潮霉素磷酸转移酶(HygR)基因的5'末端之间进行扩增。在50°C的退火温度和3分钟的延伸时间下进行PCR。只有在HygR载体插入`ubc9`启动子的靶位点时才能获得PCR产物。使用基因组DNA作为模板,以下转化体显示明显的PCR产物;3140-12、3140-34、3141-11、3141-25、3141-34、3142-15、3142-34、3140-2、3140-5和3141-1。

[0213] 使用引物3078和3146(5'-ctgaactcaccgacgtctgtc;SEQ ID NO:39)进行PCR,以在`ubc9`启动子的侧翼区和HygR基因的3'末端之间进行扩增。在50°C的退火温度和3分钟的延伸时间下进行PCR。只有在HygR载体插入`ubc9`启动子的靶位点时才能获得PCR产物。使用基因组DNA作为模板,以下转化体显示明显的PCR产物;3140-31、3141-15、3141-35、和3142-12。

[0214] 使用引物3078和3148(5'-gtgtgacagtgccgctcc;SEQ ID NO:41)进行PCR,以在`ubc9`启动子的侧翼区和粗糙脉孢菌`cpc1`启动子之间进行扩增。使用来自转化体3140-15的基因组DNA作为模板、50°C的退火温度和2分钟的延伸时间,扩增了大约3.5kb的DNA片段。只有在HygR载体插入`ubc9`启动子的靶位点时才能获得PCR产物。使用引物3078作为测序引物,在一个方向上对该扩增的DNA片段进行测序。序列分析证实pLH937 DNA插入在`ubc9`启动子中的`Cas9`靶位点处。总之,PCR分析证实表12中显示的25个转化体中的15个具有插入`ubc9`启动子区域的pLH937 DNA。其他转化体可能也具有插入`ubc9`启动子区域的pLH937 DNA,但由于插入的pLH937 DNA的程度或排列未知,因此这无法通过PCR验证。

[0215] 将转化体和亲本菌株在28°C一式三份在24孔缓释微量滴定板中培养4天,在该板

中并入20%乳糖(WO 2014047520A1)并且使用呈液体培养基的具有2.5%葡萄糖/槐糖的木霉属限定培养基(EP 1545217B1)。使用pNPP作为底物在上清液中测量植酸酶活性(US 2015/0030717,通过引用结合在此)。该测定作为终点测定进行,并在405nm处读取吸光度。来自每个菌株的一式三份培养物的平均吸光度与相关的标准偏差显示在表10中。

[0216] 表10:里氏木霉转化体的植酸酶活性

转化体	平均吸光度	标准偏差
3140-12	0.4	0.022
3140-15	0.544	0.027
3140-17	0.394	0.009
3140-18	0.456	0.028
3140-31	0.397	0.006
3140-33	0.455	0.007
3140-34	0.467	0.014
3140-35	0.407	0.025
3141-11	0.309	0.061
3141-13	0.343	0.02
3141-15	0.476	0.019
3141-20	0.447	0.011
3141-25	0.418	0.01
3141-34	0.507	0.018
3141-35	0.361	0.026
3142-12	0.501	0.029
3142-15	0.206	0.005
3142-25	0.359	0.017
3142-28	0.448	0.011
3142-32	0.429	0.027
3142-34	0.478	0.028
3142-35	0.39	0.022

[0217]

[0218]

3140-2	0.397	0.014
3140-5	0.351	0.006
3141-1	0.399	0.013
亲本菌株	0.474	0.004

[0219] 一个转化体(3140-15)比亲本菌株明显产生更多的分泌的植酸酶活性。将此转化体连同转化体3142-32和亲本菌株再次在微量滴定板中一式三份生长,并且确定培养上清液中的植酸酶活性(参见表11)。同样,转化体3140-15比亲本菌株产生大约15%更多的植酸酶。

[0220] 表11.里氏木霉转化体的植酸酶活性

[0221]

转化体	平均吸光度	标准偏差
亲本菌株	0.512	0.009
3142-32	0.494	0.015
3140-15	0.588	0.038

序列表

- <110> 丹尼斯科美国公司
- <120> 真菌的高水平蛋白质生产系统
- <130> NB36013
- <150> US62/298351
- <151> 2016-02-22
- <160> 54
- <170> PatentIn版本3.5
- <210> 1
- <211> 2470
- <212> DNA
- <213> 嗜热毁丝霉

[0001]

<400> 1	
gttaccagaa agcttgcgtt gggtaggtt atgagcgaaggccggggc acgcgtgtac	60
cgaggttaga tgggggggga atccccacgt ttaggatgca agagcagggt cgcaacggat	120
tgtttatcgc catcgagctc ttccaaaag agctggtcgg gcggcgtggg gtgacgaaag	180
taatccgtaa ggtaatattg gactgtcaac tagaaacaca cgttggagaa aatgctaaga	240
gttgggtaaa tgggggcaag agatgatgca tgccgtgggg atgtgcatgt gcagcatgtg	300
ctcgaccgg gcgtccctta tgctccacca ccgaggcccc ttgggatcg tcccgtccc	360
ggattctagg aaccgagaag agggctctgga agaaccacc tgtcccgcgt caaaggatcc	420
tggcctgctt gcctctttgt actccgtagc tatggcttca gcacaccgt cccagatcc	480
cgacgtggt cgatecaatc atccacctt tcgctctac cagcactgat cctcaacgtt	540
ggttctttt cgtatgacca gatgtagatg gtggggtgtg ccaccagccc ggactactat	600
gtaaggtacc gagtagtacc tcttctcgtc acattctggg cgcgcttgtt taaatgattc	660
tacagactcc ttatccgggc gtggcattca ttcaggagag ggtgagctcc taaagcaaac	720
cgaagagagt acctaactcc ctacctacc tattgattac acgacctcaa gcaagtacct	780
aaggtaaagg atttacctcc tttagataac cggcaccgcc cgaagagac cccggagagt	840
tgccccgcag tgaaggct cggaggaagg tagcgtacgg ggtaagaac ttctagttaa	900
ttgcagtgca tgcattgcat gcatgcatgg agaaccagct gttctccgcc ttccatttcg	960
gggctgtgac agaacaaca aattggtccc tgcgggcgcc cgaatttctg cttccaagg	1020
gcgttcccgt cagcgcgtc cgcagctccc ctactgcgt tccagaatc cggcgatat	1080
atcccgtgct tcccataca aggctcgtc tgcgtggcgc tcattgcatc gcaccctggc	1140
tcaggagtgc aaccgaacc ttgccagctg gcaccattt cctttgtctc gtatctagt	1200
ctgccacgg cttctggcgg cagccagca cctggccct ctgacatccg aaccattgga	1260
gaaaacccca tattgatcga agttcataa caactgcaa catgtcgeta tgccaaaatc	1320
gtctgcagga ggagcgttac gtcaaccgtc gctgtgcttc cggtagcccc gcgcgcgaca	1380
gcctcactaa cttggcgtcc acggtttagc aagcagtggc gcaaggacca tccgtttggc	1440
ttctatgcgc gaccgagaa gaatgcccag ggcgtgctt acctcaaggt ctgggagtgt	1500
ggtatccgg gaaaagagaa gaccatgtgg gaaggcggcc tgttcaagct ggtggtgaca	1560
tttcccagc gtacgcgca tcctttctg cgttctttt acgaaacgt tttaactaa	1620
ctaatactc gtctcataca gatatccca cgaaccccc caagtgtgaa tagccccgct	1680

gttgtgcgcc gatgttgaag cgcgggagaa ccacacgact aacattcgcc cttctgtag 1740
 gcaagtttac gccgcctctc ttcacacca acgtctaccc atceggcacc gtctgtttgt 1800
 cgattctgaa cgaagaagag gcgtggaagc ccgcgattac catcaagcag atcctgcttg 1860
 gcgtccagga cctcctcaac gaccccaacc cagaatctcc cgctcaggcc gaggcataca 1920
 acatgtacaa gaaagacaga gtgcagtacg agaggcgcat caggcagatt gtgcgggaga 1980
 atgccgcgcc atgaaaggtt cttgcaggac tacgtcttgc gcacggccga agggatagaa 2040
 ccgggtatgg aattctgatt acggagttaa gggataccaa tggcggtggg cgattacagc 2100
 atttggagtt cagacggcac gacatcatcg ggatcttctt ctactcgggg ttacgaaacg 2160
 cttgatcttg gatccggggg aggctgaggg gtgggatgag atacgccgta ggcagaatta 2220
 tgctgcgcga caacagatca ggtgtatgcg atgatttctc ccatcgtggc gtatgcacag 2280
 agacaactct tccgcagga acaagaggat cgagactatc acccccgtga tatgtcgccc 2340
 caaacattga agtgggtgtg ccgaagtgc acatcccttc ccgcggggtc ttggcaatgc 2400
 gctcaagcag ccattggctc gccccacc atggacttgg gggcagctta ctgcagtaag 2460
 cctcttgaga 2470

<210> 2
 <211> 157
 <212> PRT
 <213> 嗜热毁丝霉

<400> 2

[0002]

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

<210>	3	
<211>	6574	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	ubc9-amdS 插入表达盒	
<400>	3	
	ggccgcggta ccagaaagct tgcgttgggg taggttatga gcgcaaaggc cggggcacgc	60
	gtgtaccgag gttagatggg gggggaatcc ccacgttgag gatgcaagag cagggtcgca	120
	acggattggt ttatcgcatc gagctcttcc caaaagagct ggtcgggcgg cgtggggtga	180
	cgaaagtaat ccgtaaggta atattggact gtcaactaga aacacacggt ggagaaatg	240
	ctaagagttg ggtaaattgg ggcaagagat gatgcatgcc gtgggatgt gcatgtgag	300
	catgtgctcg acccgggcgt cccttatgct ccaccaccga ggcccctttg ggatcgtecc	360
	gtccccgat tctaggaacc gagaagaggg tctggaagaa ccaccctgtc ccgctcaaa	420
	ggatcctggc ctgcttgctt ctttgactc cgtagctatg gcttcagcac accggtcccc	480
	agatcccgac gctggtcgat ccaatcatcc accttttgc tctacgcagc actgatcctc	540
	aacgttggtt ctttttcgta tgaccagatg tagatggtgg ggtgtgccac cagcccggac	600
	tactatgtaa ggtaccgagt agtacctctt ctctcacat tctgggcgcg cttgtttaa	660
	tgattctaca gactccttat ccggcgtgg cattcattca ggagagggtg agctcctaaa	720
	gcaaaccgaa gagagtacct aactccctac ctccattt gattacacga cctcaagcaa	780
	gtacctaaag taaggcattt acctccttta gataaccggc accccgcgaa agagacccc	840
[0003]	gagagttgcc ccgcagttaa aaggctcggg ggaaggtagc gtacgggta aagaacttct	900
	agtgaattgc agtgcattga tgcatgtgat gcatggagaa ccagctgttc tccgccttc	960
	atttcggggc tgtgacagaa caaacaatt ggtccgtgcg ggcgcccga tttctgcctt	1020
	ccaagggcgt tcccgtcac ccgtctcgca gctcccctca ctgcgttcca gaattccggg	1080
	cgatatatcc cgtgcttccc atccaaaggc tcgacttgcg tggcgtcat tgcattcgac	1140
	cctggctcag gattgcaacc gaaccgttc cagctggcac cattttcctt tgtctcgta	1200
	tctagtctgc ccacgcttc tggcggcagc cgacgacctg gcccttctga catccgaacc	1260
	attttcgaag gtaccacgac caagaagatc acggtcgtca cccagttcct caagaactcg	1320
	gccggcagc tctccgagat caagcgttc tacgtccaga acggcaaggt catcccac	1380
	tccgagtcca ccatcccggg cgtcgagggc aactccatca cccaggactg gtgcgaccgc	1440
	cagaaggccc ccttcggcga cgtgaccgac ttccaggaca agggcggcat ggtccagatg	1500
	ggcaaggccc tcgctgggccc catggtctc gtcatgtcca tctgggacga ccacgctc	1560
	aacatgctct ggctcgactc cacctggccc atcgacggcg ccgcaagcc gggcggcag	1620
	cgcggtgctt gccccaccac ctctggcgtc cccgctgagg tcgaggccga gcccccaac	1680
	tccaactgca tcttctccaa catcctctt ggccccatcg gctccaccgt ctccgctg	1740
	cccgacggcg gcagcggcaa cccaaccgg cccgtcagct cgtccacccc ggtcccctcc	1800
	tcgtccacca catcctccgg ttctccggc ccgactggcg gcacgggtgt cgctaagcac	1860
	tatctagact ggaacgcaa cctgaaggg attcttctt tgagagatgg aagcgtgtca	1920
	tatctctteg gtctacggc aggtttttt ctgctcttc gtagcatggc atggtcactt	1980

[0004]

cagcgcttat ttacagttgc tggattgat ttcttgtgca aattgctatc tgacacttat	2040
tagctatgga gtcaccacat ttcccagcaa ctccccact tectctgcaa tgccaacgt	2100
ctctcttca ctgagctcc gtcgataac ctgcactgca accggtgcc catggtacgc	2160
ctccgataca tactcttct gcacgagggc atcaagetca ctaaccgct tgaactctc	2220
attcttctta tcgatgtct tatccgaaa ggtaaccgga acaaccacgc tcgtgaaate	2280
cagcaggttg atcacagagg cataccata gtaccggaac tggatcatgcc gtaccgcagc	2340
ggtaggcgta atcggcgca tgatggcgtc cagttccttc ccggcctttt cttcagcctc	2400
ccgccatttc tcaaggtact ccatctggtta attccacttc tggagatgcg tgtcccagag	2460
ctcgttcag ttaacagctt tgatgttcgg gttcagtagg tctttgatata ttggaatcgc	2520
cggctcgccg gatgcaactga tctcgccat tacgtcggcg ctgccgtcag ccgcgtagat	2580
atgggagatg agatcgtggc cgaatcgtg cttgtatggc gtccacgggg tcacggtgtg	2640
accgctttg gcgagtcgg cgacggtggt ttccacgcc cgcaggatag gagggtgtgg	2700
aaggacattg ccgtcgaagt ttagtagacc gatattgagc ccgccgttct tgatcttggg	2760
ggcaataatg tccgactcgg actggcgcca gggcatgggg atgacctgg agtcgtatct	2820
ccatggctcc tgaccgagga cggatttggg gaagagcgg aggtctaaca tacttcatca	2880
gtgactgcc gtctctgata tagtataaaa agcaagaaag gaggacagtg gagcctggt	2940
atagagcagg aaaagaagga agaggcgaag gactcacct caacagagtg cgtaatcggc	3000
ccgacaacgc tgtgcaccgt ctctgacc tccatgtgt tcgcatctt tgcatacggc	3060
agccgccat gactcggcct tagaccgtac aggaagtga acgcgccgg cactcgaatc	3120
gagccaccga tatccgttc tacaccgatg acccaccac gaatccaac gatcgcacc	3180
tcaccaccag aactgccgc gcacgaccag ttcttgttc gtgggttgac ggtgcgccg	3240
atgatgtgtg tgaactgtc gcagaccatc agggctcgc ggacagaggt cttgacgtag	3300
aagacggcac cggctttgcg gagcatggtt gtcagaacc agtccccctc gtcgacttg	3360
tttagccatg agatgtagcc cattgatgtt tcgtagcct ggtggcatat gttagctgac	3420
aaaaagggac atctaacgac ttaggggcaa cgggtgacct tgactcgaag ctggtcttg	3480
agagagatgg ggaggccatg gactggacca acgggtctct tgtgctttgc gtagtatca	3540
tcgagttccc ttgcctgcgc gagagcggcg tcagggaaga actcgtgggc gcagtttgc	3600
tgacagaag ccagcgtcag cttgatagtc ccataagggt gcgttgttac atctccctga	3660
gaggtagagg ggaccctact aactgctggg cgattgtgc ccgtttacag aatgctagc	3720
taacttccac cgaggtcaac tctccggcg ccagcttggg cacaagatct gcagcggagg	3780
ctctgtgat cttcagttcg gcctctgaaa ggatccccga tttctttggg aaatcaataa	3840
cgctgtcttc cgcaggcagc gtctggactt tccattcatc agggatggtt tttgcgagc	3900
gggcgcgctt atcagcgcc agttcttccc aggattgagg cattctgtgt tagcttatag	3960
tcaggatgtt ggctcgaca gtgtaactg ggagttggca tgagggttat gtagcttct	4020
ttagccccgc atccccctca ttctctcat tgatccccgg ggagcggatg gtgttgataa	4080
gagactaatt atagggttta gctgggtcct agctggtgat tgctggctt cgccgaattt	4140
tacgggcca gaaagctgc agaaccgcg cactggtaaa cggtaatata gctatcagcc	4200
ccatgctaac gaggttaaat tacgtgtatt gctgataaac accaacagag ctttactgaa	4260

	agatgggagt cacgggtggtg cttccccact gcgattattg cacaagcagc gagggcgaac	4320
	ttgactgtcg tcgctgagca gcctgcagtc aaacatacat atatatcaac cgcaagacg	4380
	tctggccttg tagaacacga cgtccctag caacacctgc cgtgtcagcc tctacggtg	4440
	ttacttgcatt tcaggatgct ctccagcggg cgagctattc aaaaattca aagcaggat	4500
	ctcgtattgc caggattcag ctgaagcaac agtgccaag gaaatctgcg tcggttctca	4560
	tctgggcttg ctcggtcctg gcgtagatct agaggtagca cgaccaagaa gatcacggtc	4620
	gtcaccagtc tctcaagaa ctgcggcggc gagctctccg agatcaagcg gttctacgtc	4680
	cagaacggca aggtcatccc caactccgag tccaccatcc cgggcgtcga gggcaactcc	4740
	atcaccagg actggtgcga ccgccagaag gccgccttcg gcgacgtgac cgacttccag	4800
	gacaaggcgg gcatggtcca gatgggcaag gccctcgcgg ggcccatggt cctcgtcatg	4860
	tccatctggg acgaccacgc cgtcaacatg ctctggctcg actccacctg gccatcgcac	4920
	ggcgcggca agccggggcgc cgagcgcggt gcctgcccc ccacctcggg cgtccccgt	4980
	gaggtcgagg ccgagggccc caactccaac gtcatcttct ccaacatccg cttcggcccc	5040
	atcggtcca ccgtctccgg cctgccgcac ggcggcagcg gcaaccccaa cccgcccgtc	5100
	agctcgtcca ccccggtccc ctctcgtcc accacatcct ccggttctc cggccccact	5160
	ggcggcagcg gtgtcgttaa gcaacttccg aaggagaaaa ccccatattg atcgaagttc	5220
	cataacaact gccaacatgt cgctatgcca aaatcgtctg caggaggagc ggtacgtcaa	5280
	ccgtcgtgt gttccggta cccccgcgc cgacagcctc actaacttgg cgtccacggt	5340
[0005]	ttagcaagca gtggcgcaag gaccatccgt ttggcttcta tgcgcgaccg cagaagaatg	5400
	cccagggcgt gcttgacctc aaggctctgg agtgtggtat cccgggcaaa gagaagacca	5460
	tgigggaagg cggcctgttc aagctggtg tgacatttcc cgacggtacg cggcatcctt	5520
	tcctcgggtt cttttacgga aacgttttac actaactaat caetcgtctc atacagagta	5580
	tcccacgaaa ccccccaagt gtaagtgcg cggctgttgt gcgccgatgt tgaagcggg	5640
	gagaaccaca cgactaacat tcgcccttcc ttaggcaag tttacgccgc ctctcttcca	5700
	ccccaacgtc taccatccg gcaccgtctg tttgtcgatt ctgaacgaag aagaggcgtg	5760
	gaagcccgcg attaccatca agcagatcct gcttggcgtc caggacctcc tcaacgacct	5820
	caaccagaa tctcccgcct agccgaggc atacaacatg tacaagaaag acagagtga	5880
	gtacgagagg cgcatcagc agattgtgcg ggagaatgcc gcgccatgaa aggtcttgc	5940
	aggactacgt cttgcgcac gccgaaggga tagaacggg tatggaattc tgattacgga	6000
	gttatgggat accaatggcg gtggcgatt acagcatttg gatttcagac ggcacgacat	6060
	catcgggact ttcttctact gcgggttacg aaacgttga tcttggatcc gggggaggct	6120
	gaggggtggg atgagatacg ccgtaggcag aattatgctg cgcgacaaca gatcaggtgt	6180
	atgcgatgat ttctccatc gtggcgtatg cacagagaca actcttccgc agggacaag	6240
	aggatcaga ctatcccc cgtgatatgt cgcaccaaac attgaagtgg tgttgccgaa	6300
	gtgtcacatc cttcccgcg ggtcttggc aatgcgctca agcagccatt ggctcgcgcc	6360
	ccacatgga cttgggggca gcttactgca gtaagcctct tgagagacaa atcgagactc	6420
	gcaacctcag cttctcagag atgcattgcg ttctcactc ccagagggcc agctcctcac	6480
	gtggtcgtt gtggtcctg cagacccca acgctgtcaa agccttctgt cttatcagtc	6540

atgттаaggt aatgacccat gtggaactgt tggc	6574
<210> 4 <211> 4483 <212> DNA <213> 人工序列	
<220> <223> 表达盒eg2	
<220> <221> 尚未归类的特征 <222> (1749)..(1749) <223> n是a、c、g、或t	
<220> <221> 尚未归类的特征 <222> (3668)..(3668) <223> n是a、c、g、或t	
<400> 4	
ggcccctcta gaactagtgg atccccggg ctgcagtecc ttacctatgg gctcctagtc	60
tcgttctct tttgataga ttgtatattt gcaacgttgc aaaatgagac atttcaatca	120
tatgtagccg ccagctactg ttagcgtact cagcgttgc caaacggcgg tttttctggg	180
tagcactgtg ccgctgccc ctgagccgtg cgtcgcggaa acccccttaa gtagcaagta	240
tgttaccgcc gagaccgaca atgctgttgg ttacctcgt ggtccatgat tgcaatctag	300
atatcgtgcg gggcttttgc aatcggtttt ccctaccac tttctcttt tggacacttt	360
ctcttttga aatgccgaa atgatcggc tcgctcacgc cccgaagtcc cgagctgggg	420
ctagatccgt gattgcaac cgggtcgaac gcgactgggg cagacctgc tcagccttgg	480
tcgtgccgga atggcgggta cctttaccag gtcgggatca attacatagg atgccatgtg	540
cgtggatttg attgcatcgc tgtcccttt gtatgtgtcc gagagcgaga catcaacgcg	600
aaaaccggaa tgcctccaac gtcgctctct gtctataggg tcttttttt tcttctgctc	660
catatcatct gtcttgaact aagtgatcat ctgctgtcac gtcccceca atgattgtaa	720
agaatgataa gtgatgctcg ccggggccag gctctgtgaa agttcctct ttggttgacg	780
atcaggtagc gccaacgttg attgggccgc ccgtaaaac cgacctgic tcctttcgtt	840
gcaagtctcc gcgagaccgt gccaaagcat ttctccggat cctcaatta cataaggttt	900
ggctccagg taggtctgga agctaccac ctcgccaag caaccaatca caaccagacc	960
tcgcccgtt tcgaccttcc tggtttgtct cagggttgc caacgtcct ccgtggcggg	1020
tgcttggta tcgacgtcg caggcagtg ccgggcacgc ggagccccg tcaaagctt	1080
accctttcag agctaggttt cattagcct tcgaaaaca cccaaggccc cgtcgaacc	1140
atcacaaccg gccgataacc agatctcgt aggtccgata aggatccaaa atggtgtcgg	1200
ctgacgttgc atgtgcccag gcaggaggat gatccccagg gttgttgcg gcagctccc	1260
cacgtcgggg agggggagg ggaggggaaa gccctaacta acgttcgttc tatcacgggc	1320
cgaccgggcc atgctttcgg cttgtgagcg gtgggtcaa gggcaacaag aatgctaag	1380
tgccggacga agacacgcgg gcatgaggtc tcagggtgac ctgcccacaa ccaagtccca	1440
ctgccatgc ctccagcagc aacgttgccc tagaagggtc agggggtttg ttgtagacc	1500
acgaccatgc tgccgcgag cggagggtt gcttgctaca ggcgctgaag ggtcaactcg	1560
gtgccaaaag tggtaccaa gcgtgccatc aagggaatg agatgatggt ggctcgtggg	1620

[0006]

caaagaaaag acaaggagg tgactctaga gagatgctct cgagttcacg ggtataagag	1680
cactgtgata gttcacaag ccggcgtact cctctagagc atctatcacc aacatcacca	1740
gaaaggtcnt agaccaggtg gttgccatat ccagtcgcaa aagagccaaa gagcgaagga	1800
gcacgaaagc acagcccaat cattccctgc tttgctactt cttctccacc atgaagtect	1860
ccatcctcgc cagcgtcttc gccacgggcg ccgtggctca aagtggtcgg tggcagcaat	1920
gtggtggcat cggatggcaa ggatcgaccg actgtgtgtc gggttaccac tgcgtctacc	1980
agaatgattg gtacagccag tgcgtgctg gcgcggcgtc gacaacgctc cagacatcta	2040
ccacgtccag gccaccgcc accagcaccg cccctccgtc gtccaccacc tcgcctagca	2100
agggcaagct caagtggctc ggacgcaacg agtcgggctc cgagttcggg gagggcaact	2160
accccgccct ctggggaag cacttcatct tcccgctgac ttcggcgatt caggtacggg	2220
ccaataataa tatattatta tagcaggcag gagggagcag gagaagaagg gagggcagg	2280
tggccaacaa tcggaagaag accgggaggc actgaccgtt gattcctttg tgtaatagac	2340
gctcatcaat gatggataca acatcttccg gatcgacttc tcgatggagc gtctggtgcc	2400
caaccagttg acgtgctct tcgacgaggc ctacctccgc aaacctgacc aggtggtcaa	2460
cttcgtgacg aacgcgggca agtacgccgt cctggaccgc cacaactacg gccggtacta	2520
cggcaacgct atcacggaca cgaacgcgtt ccggacctc tggaccaacc tggccaagca	2580
gttcgctec aactcgctcg tcattctcga caccaacaac gagtacaaca cgatggacca	2640
gaccctggtg ctcaacctca accaggccgc catcgacggc atccgggccg ccggcgcgac	2700
ctcgcagtac atcttctgctc agggcaacgc gtggagcggg gcctggagct ggaacacgac	2760
caacaccaac atggccgctt gacggaccgc cagaacaaga tcgtgtacga gatgcaccag	2820
tacctcgact cggacagctc gggcacccac gccgagtgcg tcagcagcaa catcggccgc	2880
cagcgcgtcg tcggagccac ccagtggctc cgcgccaacg gcaagctcgg cgtcctcggc	2940
gagttcgcgg gcggcgccaa cgcctctgca cagcaggccg tcaccggcct cctcgaccac	3000
ctccaggaca acagcgactg ctggctgggt gccctctggt gggccgccgg tcctggtgg	3060
ggcgactaca tgtactgctt cggtaagttt ctcccttgtt cttggctttc cccccagtaa	3120
gggagtcagg caacatgccc aagaccgctt cgcttcgct tcaaggcgtt cgttgctcac	3180
actgaagagt tccaacttc aacctgttc gtgtcctccg atcagcttcg acggggtgaa	3240
gggggaaggg atttgggagt gaggtggagg tcaaaaggag ggatatcccc agatctccac	3300
aaacggccct gagccaacaa cagcctctgg ggtcaaatg ggcaccaacc atacggtcat	3360
tcactcagga cacctgctaa cgcgtctctt tttttgtt ccagagcctc cttcgggcac	3420
cggctatgct aactacaact cgatcctaaa gaagtacttg ccgtaaggat cctaagtaag	3480
taaacgaacc tcctggaag aggttctgag acacgcgcga ttctctgta tatagtttta	3540
ttttcactc tggagtgtt cgetccacca gtacataaac ctttttttc acgtaacaaa	3600
atggttctt ttacagcat gtgaaccate ttgatgcctt gacctctca gttctcactt	3660
taacgtantt cgcgttagtc tgtatgtccc agttgcatgt agttgagata aatacccctg	3720
gaagtggctc tggcctttg tgggacggag cctctttct gtggtctgga gagcccgtc	3780
tctaccgctt acctcttac cacagtacac tactcacaca ttgctgaact gaccatcat	3840
accgtacttt atctgttaa ttcgtggtgc tctcactat tctatttct caaatggaga	3900

[0007]

gcacattcat cggcgcaggg atacacgggt tatggacccc aagagtgtaa ggactattat 3960
 tagtaatatt atatgcctct aggcgcctta acttcaacag gcgagcacta ctaatcaact 4020
 tttggtagac ccaattacaa acgaccatac gtgccggaaa ttttgggatt cegtcgcctc 4080
 tccccacca agctagaaga ggcaacgaac agccaatccc ggtgctaatt aaattatg 4140
 gttcattttt tttaaaaaaa tttttcttc ccattttct ctcgcttttc tttttcgc 4200
 cgtagttgat caaagtccaa gtcaagcgag ctatttgtgc tatagctcgg tggtataat 4260
 cagtacagct tagagaggt gtaaaggtat gataccacag cagtattcgc gctataagcg 4320
 gcactcctag actaattgtt acggtctaca gaagtaggta ataaaagcgt taattgttct 4380
 aaatactaga ggcacttaga gaagctatct aaatatatat tgaccctage ttattatccc 4440
 tattagtaag ttagttagct ctaacctata gatagccaaa tgc 4483

<210> 5
 <211> 3935
 <212> DNA
 <213> 人工序列

<220>
 <223> 表达盒AnPGII

<220>
 <221> 尚未归类的特征
 <222> (1722)..(1722)
 <223> n是a、c、g、或t

[0008]

<220>
 <221> 尚未归类的特征
 <222> (3120)..(3120)
 <223> n是a、c、g、或t

<400> 5
 gggctgcagt cccttaccta tgggctccta gtctcgttcc tctttttgat agatttgtat 60
 tttgcaacgt tgcaaaaata gacatttcaa tcatatgtag ccgccagcta ctgttagcgt 120
 actcagcgtt gcccaaacgg cggtttttct gggtagcact gtgccgcgtg cccctgagcc 180
 gtgcgtcgcg gaaacccct taagtagcaa gtatgttacc gccgagaccg acaatgctgt 240
 tggttacctc gctggtccat gattgcaatc tagatatcgt gcggggcttt tgcaatcgg 300
 tttccctacc cactttcttc ttttgacac tttctcttt ggaaaatgcc gaaatgatgc 360
 ggctcgtca cgcctcgaag tcccagctg gggctagatc cgtgattgca acgcggtgcg 420
 aacgcgactg gggcagacct cgtcagcct tggctcgtcc ggaatggcgg gtacctttac 480
 caggtcggga tcaattacat aggatgccat gtgcgtggat ttgattgcat cgtgtccct 540
 tttgatgtg tccgagagcg agacatcaac gcgaaaaccg gaatgctccc aacgtcgtc 600
 tctgttcata gggctttttt ttttctctc ctccatatca tctgtcttga actaagtgat 660
 catctgctgt cagtcctcgc ccaatgattg taaagaatga taagtgatgc tcgccggggc 720
 caggctctgt gaaagtccc tctttggtg acgatcaggt agcgcacaac ttgattgggc 780
 cgcccgtaaa atccgacct gtctccttc gttgcaagtc tccgcgagac cgtgccaagc 840
 atgttctccg gatecctcaa ttacataagg tttggctcca gggtaggtct ggaagctacc 900
 cacctcggcc aagcaaccaa tcacaaccag acctcgcggc gtttcgacct tcctggtttg 960
 tctcagggt ggccaacgct ctcccgtggc gggcgcctgg tgatcgagc tcgcaggcga 1020

gtgccgggca cgcggagccc ccgtcaaagc ttgacccttt cagagctagg tticattagg	1080
ccttcgaaaa caaccecaagg ccccgctcga accatcaca cgggccgata accagatctc	1140
ggtaggtccg ataaggatcc aaaatgggtg cggctgacgt tgcattgtcc caggcaggag	1200
gatgatcccc agggttgttg ccggcagctc ccgcacgtcg gggaggggga gggggagggg	1260
aaagccctaa ctaacgttcg ttctatcacg ggccgaccgg gccatgcttt cggtttgtga	1320
gcggtggggt caaggcaac aagaaatgct aagtgcggga cgaagacacg cgggcatgag	1380
gtctcagggt gacctgcga aaaccaagtc cactcgcca tgcctccagc agcaacgttg	1440
ccgtagaagg gtcagggggt ttgtttaga cccacacca tgctgccggc gagcggaggg	1500
ttggcttget acaggcctg aagggtcaac tcggtgccca aagtggctac caagcgtgcc	1560
atcaagggaa atgagatgat ggtggctcgt gggcaaagaa aagacaaggg aggtgactct	1620
agagagatgc tctcagttc acgggtataa gagcactgtg atcgttcaca aagccggcgt	1680
actcctctag agcatctatc atcaacatca ccagaaaggt cntagaccag gtggttgcca	1740
tatccagtcg caaaagagcc aaagagcgaa ggagcacgaa agcacagccc aatcattccc	1800
tgctttgcta cttctctcc accatgcatt cctttgett cctcctcgtc tacggcctcg	1860
tggttggtgc tacgttcgtc tcggcctccc ccatcgagge ccgcgacagc tgcacctca	1920
ctactccgcg tcccgcaag gccgtaagg ctaagtctc caccatcacc ctcaacaaca	1980
tcgaggtgcc cgctggcacg accctggacc tcaccggtct gacttcgggc accaaggtca	2040
tcttcagagg cagcaccact ttcagtacg aggagtgggc tggccccctc atctcagatg	2100
cggcgagca taccacgctc actggtgcct cgggccacct catcaactgc gacggcctc	2160
[0009] gctgggtgga tgtaagggc acctccgcta agaagaagcc gaagtcttc tacccccacg	2220
gcctggactc cagctcgatc accggcctca acatcaagaa cacgccccctc atggccttct	2280
ccgtgcagge caacgacatc accttcaccg acgtcactat caacaagcc gacggcgaca	2340
cccagggcgg tcacaacact gacgccttcg acgtgggtaa ctcggctggc gtcaacatca	2400
tcaagccgtg ggtccacaac caggacgatt gcctcgggt caacagcggc gagaacatct	2460
ggttactcgg cggcacgtgc atcggcggtc acggcctgag catcggctcc gtcggtgacc	2520
gcagcaaca cgtcgtcaag aacgtcacga tcgagcactc caccgtgtcc aacagcgaga	2580
acgccgtccg gatcaagacc atctccggcg ccacgggcag cgtcagcgag atcaactaca	2640
gaacatcgt catgagcggc atctccgact acggcgtcgt catccagcag gactacgagg	2700
acggcaagcc caccgtaag cccactaacg gtgtgactat ccaggacgtc aagctggaga	2760
gcgtgacggg tagcgtcgac agcggcgcca ccgagatcta cctcctgtgc ggttcgggca	2820
gtgtctccga ctggacctgg gatgacgta aggtgaccgg cggcaagaag agcactgcct	2880
gcaagaactt cccgtccgct gctagctgct aagaattcgg atcctaagta agtaaagca	2940
cctctctgaa ggagttctg agacacgcgc gattctctg tatatagttt tattttcac	3000
tctggagtgc ttcgtccac cagtacataa acctttttt tcacgtaaca aaatggcttc	3060
ttttcagacc atgtgaacca tctgatgcc ttgacctctt cagttctcac tttaacgtan	3120
ttcgcgttag tctgtatgct ccagttgcat gtagttgaga taaatacccc tggaagtggg	3180
tctggccctt tgtgggacgg agccctcttt ctgtggtctg gagagcccgc tctctaccgc	3240
ctaccttctt accacagtac actactcaca cattgctgaa ctgaccatc ataccgtact	3300

ttatectggt aattcgtggt gctgtcgact attctatgtg ctcaaatgga gagcacattc 3360
 atcggcgagc ggatacacgg tttatggacc ccaagagtgt aaggactatt attagtaata 3420
 ttatafgcct ctaggcgct taacttcaac aggcgagcac tactaatcaa cttttggtag 3480
 acccaattac aaacgacct acgtgccgga aatitttggga ttccgtecgc tctcccacac 3540
 caagctagaa gaggcaacga acagccaatc ccggtgctaa ttaaattata tggttcattt 3600
 tttttaaaaa aatittttct tcccatttct ctctcgcttt tctttttcgc atcgtagtgt 3660
 atcaaagtc aagtcaagcg agctatttgt gctatagctc ggtggctata atcagtacag 3720
 cttagagagg ctgtaaaggt atgataccac agcagtattc gcgctataag cggcactcct 3780
 agactaattg ttacggtcta cagaagtagg taataaaagc gtttaattgtt ctaaatacta 3840
 gaggcactta gagaagctat ctaaataat attgacccta gcttattatc cctattagta 3900
 agttagttag ctctaaccta tagatagatg catgc 3935

<210> 6
 <211> 5267
 <212> DNA
 <213> 人工序列

<220>
 <223> 表达盒TwBx11

<220>
 <221> 尚未归类的特征
 <222> (4452)..(4452)
 <223> n是a、c、g、或t

[0010]

<400> 6
 ggccgctcta gaactagtgg atccccggg ctgcagtccc ttacctatgg gctcctagtc 60
 tcgttctctt ttttgataga tttgtatgtt gcaacgttgc aaaaatgagac atttcaatca 120
 tatgtagccg ccagctactg ttagcgtact cagcgttgcc caaacggcgg tttttctggg 180
 tagcactgtg ccgctgctcc ctgagccgtg cgtcgcggaa acccccttaa gtagcaagta 240
 tgttaccgcc gagaccgaca atgctgttgg ttacctcgtt ggtccatgat tgcaatctag 300
 atatcgtgcg gggttttgc aatcggtttt ccctaccac tttcttcttt tggacacttt 360
 ctcttttggg aaatgccgaa atgatgcggc tcgctcacgc cccgaagtcc cgagctgggg 420
 ctagatccgt gattgcaacg cgggtcgaac gcgactgggg cagacctcgc tcagccttgg 480
 tcgtgccgga atggcgggta cttttaccag gtcgggatca attacatagg atgccatgtg 540
 cgtggatttg attgcatcgc tgcctctttt gtatgtgtcc gagagcgaga catcaacgcg 600
 aaaaccggaa tgcctccaac gtcgctctct gtccataggg tctttttttt tcttctgctc 660
 catatcatct gtcttgaact aagtgatcat ctgctgtcac gtcccgecca atgattgtaa 720
 agaatgataa gtgatgctcg ccggggccag gctctgtgaa agttccctct ttggttgacg 780
 atcaggtage gccaacgttg attgggccc cegtaaaatc cgacctgctc tctttctggt 840
 gcaagtctcc gcgagaccgt gccaaagcat ttctccggat cctcaatta cataagttt 900
 ggctccaggg taggtctgga agctaccac ctcggccaag caaccaatca caaccagacc 960
 tcgcgcgctt tcgaccttcc tggtttctct cagggtggc caacgtctc cctggcggg 1020
 tgcttggatg tcgaggtcgc caggcgagtg ccgggcacgc ggagcccccg tcaaagcttg 1080
 accctttcag agctaggttt cattaggcct tcgaaaacaa cccaaggccc cgtcgcaacc 1140

	atcacaccg gccgataacc agatctcgg aggtccgata aggatccaaa atggtgtcgg	1200
	ctgacgttgc atgtgccag gcaggaggat gatccccagg gttgttgccg gcagctccc	1260
	cacgtcgggg agggggagg ggaggggaaa gccctaacta acgttcgttc tatcacgggc	1320
	cgaccgggcc atgctttcgg cttgtgagcg gtgggtcaa gggcaacaag aaatgctaag	1380
	tgcgggacga agacacgcg gcacagagtc tcagggtgac ctgcgcaaaa ccaagtccca	1440
	ctcgccatgc ctccagcagc aacgttgccg tagaagggtc agggggtttg ttgtagacc	1500
	acgaccatgc tgccggcgag cggagggtt gcttctaca ggcctgaag ggtcaactcg	1560
	gtgccaaaag tggtacca gctgcccac aagggaatg agatgatggt ggctcgtggg	1620
	caaagaaaag acaaggagg tgactctaga gagatgctct cgagttcacg ggtataagag	1680
	caactgtgat gttcacaag ccggcgtact cctctagagc atctatcacc aacatcacca	1740
	gaaagtgcaa gaccaggtgg ttgcatatc cagtcgcaaa agagccaaa agcgaaggag	1800
	cacgaaagca cagcccaatc attccctgct ttgtacttc ttctccacca tgtacgcca	1860
	gttcgccacc ctgcgccccc tcgtgcggcg cggcgccgc cagaacgccg cccagaacaa	1920
	ccagacctac gccaaactaca gcagccagag ccagcccgac ctctaccccc agaccctgc	1980
	ccacctaac ttcagcttcc ccgactgcat caacggcccc ctcaagaca acatcgtctg	2040
	cgacaccagc gccaaactac tcgaccgagc cgagggcctg atcgccctct tcaccctcga	2100
	ggagctgatc aacaacacc agaacaccgc cccggcgctc cccgcctcg gctcccgc	2160
	ctaccaggtc tggtcggagg ccctccacgg cctcgaccgc gccaaactcg ccacgagcgg	2220
	cgacgagtg acctgggcca ccagcttccc catgcccacc ctacagatgg ccgcccctca	2280
[0011]	ccgaccctc atcaaccaga tcgccggcat catcggcacc caggcccgcg cttcaacaa	2340
	cgccggccgc tacggcctcg acgcctacgc cccaacatc aacggctacc gcaaccctc	2400
	ctggggccgc ggccaggaga cccccgca ggacgcaaac ttctcagca gcagctacgc	2460
	ctacgagtac atcaccggcc tcaggggcgg cgtcgacccc gaccacctca aggtcgtcgc	2520
	caccgccaag cacttcgccc gctacgacct cgagaactgg ggcggcaaca gccgctcgg	2580
	cttcgagccc agcatcacc agcaggacct cggcagtag tacaccccc agttctcgc	2640
	cgccagccgc tacgccaagg ccgagctt catgtgcagc tacaacagcg tcaacggcgt	2700
	ccccagctgc agcagcagct tctgtctcca gaccctctc cgcgacaact gggacttccc	2760
	cgagtacggc tacgtcagca gcgactgca cggcgtctac aacgtcttca acccccaggg	2820
	ctacgccagc aaccagagcg ccgcccgc gcacagcctc cgcgcccgca ccgacatcga	2880
	ctgcggccag acctaccctt ggaacctcaa ccagagcttc atcgagggca gcgtcacccg	2940
	cgcgagatc gagcgagca tcgtccgct ctacagcaac ctctgaagc tcggctactt	3000
	cgacggcgac aagagcgagt accgccagct cggctggaac gacgtcgtca ccaccgacg	3060
	ctggaacatc agctacgagg ccgcccgcga gggcatcgtc ctgctcaaga acgacggcat	3120
	ctcccctc agcaagcagc tcaagtcgat cgcctcacc gggccctggg ccaacgccac	3180
	cgagcagctc cagggcaact actacggcac cggcccctac ctcatcacc cgctccaggg	3240
	cgccagcgac gccggctaca aggtcaacta cgcctcggc accaacatcc tcggcaacac	3300
	caccgagggc ttcgcccagc ccctctcggc cggccagaag tcggagctca tcgtctacct	3360
	cgcgccatc gacaacacca tcgaggccga gggcaccgac cgcagtaac tcacctggcc	3420

	eggcaaccag ctcgacctca tccagcagct cagccagacc ggcaagcccc tegtctctct	3480
	ccagatgggc ggcggccagg tcgacagcag cagcatcaag gccaacagca aggtcaacgc	3540
	ctctgcttgg ggcgcttacc ccggccagag cggcggcacc gccatcttcg acatcctcag	3600
	eggcaagcgc gteccccgcc gccgcctcgt caccaccagc taccgcccg agtacgccac	3660
	ccagtcccc gccaccgaca tgaacctccg ccccgacggc gccagcaacc cgggccagac	3720
	gtatatgtgg tacaccgca cccccgtcta cgacttcggc tacggcctct tctacaccac	3780
	cttcaaggag accgcccaaga agctgggcag cagctcgttc gacatcagcg agatcgtcgc	3840
	cgccccccgc agccccagct acgagtacag cgagctggtc cccttcgtca acatcaccgc	3900
	caccatcaag aacaccggca agaccgccag ccctacacc gccatgtgtg tcgccaacac	3960
	caccaacgcc ggccccgcc cgtaccctca caagtggctc gtcggctacg accgcctcgc	4020
	cagcatcgag cccggcaaga gcgccacct cgtcatcccc gtccccatcg gcgccatcgc	4080
	ccgcgtcgac gagaacggca accgcatcgt ctaccggggc gactaccagc tggccctcaa	4140
	cgtcgagcgc tcggtcgtct gggacatcaa gtcaccggc gacgccgta ccatcgagaa	4200
	ctggccccct gacgagcagg agatccagcc cgacagcaac taatgaattc ggatcctaag	4260
	taagtaaagc aacctctctg aaggaggctc tgagacacgc gcgattcttc tgtatatagt	4320
	tttatttttc actctggagt gcttcgctcc accagtacat aaaccttttt tttcacgtaa	4380
	caaatggct tcttttcaga ccatgtgaac catcttgatg ccttgacctc ttcagttctc	4440
	actttaacgt anttcgcggt agtctgtatg tcccagttgc atgtagttga gataaatacc	4500
	ectggaagtg ggctcgggcc tttgtgggac ggagccctct tctgtggctc tggagagccc	4560
[0012]	gtctctacc gctaccttc ttaccacagt aactactca cacattgctg aactgaccca	4620
	tcataaccgta ctttatcctg ttaattcgtg gtgctgtcga ctattctatt tgcctaaatg	4680
	gagagcacat tcateggcgc agggatacac ggtttatgga ccccaagagt gtaaggacta	4740
	ttattagtaa tattatagc ctctaggcgc cttaacttca acaggcgagc actactaatc	4800
	aacttttggg agaccaatt acaaacgacc atacgtgccg gaaattttgg gattccgctc	4860
	gctctcccca accaagctag aagaggcaac gaacagccaa tcccgggtgct aattaaatta	4920
	tatggttcat tttttttaa aaaatttttt ctteccattt tectctcgtt tttcttttcc	4980
	gcatcgtagt tgatcaaagt ccaagtcaag cgagctatct gtgctatagc tcggtggcta	5040
	taatcagtac agcttagaga ggctgtaaag gtatgatacc acagcagtat tcgctgctata	5100
	agcggcactc ctagactaat tgttacggtc tacagaagta ggtaataaaa gcgttaattg	5160
	ttctaaatac tagaggcact tagagaagct atctaaatat atattgacct tagcttatta	5220
	tcctattag taagttagtt agctctaacc tatagataga tgcattgc	5267
	<210> 7	
	<211> 7794	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> pyr5选择标记	
	<400> 7	
	gatctggcca agggccacct tgaagtggtc ggcactcagc agtaaccgctc agcatagccc	60
	cattgcgcga tcatgtagcc tccaatccg tttccttcag gcgtggaggg gcgtaccata	120

	taggatgttc atcaaatcgc ggtggaaggg gtgctgatcc gtcaacacgg ggaagctcga	180
	gatgatggcg ccaaatttct cgctgcatgt ctctgtgta aatctgagca ccggggttag	240
	tgccgagtec gaaatcgaat catcaccagc ttcgtcgcaa acggtgtggt ggaatctggg	300
	gagatttgaa tcgatgagcc gggatcaccg acttgacttt gcgggtgtaa aacgccetaa	360
	ttctgctaac tgttcattga gcgtcagaaa atgtctgcga cctcagcag taattcggga	420
	ggcacatact ctgaaatccg ggacggactg taatggtgtt ctaggtcagc cagcatccca	480
	ccactccgct aaaacatcca acatactctg agtcggcagt cgcctctgtg tccgctcgcg	540
	gaacgtcagc ctttctcgtc ttgcaacctt tataacctag cttcaact tacgaaagga	600
	tgatgtcgat aaactctgt gcctgtacg aaaatgatca gcttagtca gaaaggaac	660
	agcatatata cgaattgaac cttacgtag gcacaggcgc gatatecttc caggccatgg	720
	caggcaacgc gcagggtgct gtccgaattt ggcatgggg gtgtttggga gttgtcagat	780
	agggagcacc tggcggggcg gtgtcgattt gccgtcccca agtcgccga aataattttt	840
	tgcggagagc acaaaatgat aagataaggc agcgggtgtg cgttatcaaa atatcccaag	900
	cccagctcga gaagcattgc aagtgggta cgacgtacgg agtactgtgt aactccgtag	960
	acatacaaaa aagttaaac acccttgggc cccacatga atccaacccc accaagaaat	1020
	tcgcatgga gggctctct tcagctctgc gatagggga cgtttgaca aactccctca	1080
	tttcttttt tcgaatcttc accaggagt cctccggtg acaataaac ttccagccca	1140
	agaaccgta agacacgtac gaaccgacag acatggcccc actcgttct tacaagccg	1200
[0013]	acttctccg gcggccatc gccgcaaca tctcaagtt tggcagcttc gagtcaagt	1260
	cgaagcgc atctgcctac ttctcaacg cgggagactt ctaccgggcc gatctgctcg	1320
	aggcgtcgc gacggcgtac gcgcacgcca tcatcgaggc gcaccggagc ggcgcgatcc	1380
	agttcgacat cgtcttcggc ccggcctaca aggggatccc gttggcgacg gcggccacca	1440
	ttcggctggg ccagctggat ccggccacgt acggccatac cacgtgctac tegtctgacc	1500
	gcaaggagc aaaggaccac ggtgaggcg gcaacattgt gggggcgcg ctcaagggca	1560
	agagggtgct tatcgtggat gatgtgatca cgctggcac ggcgaagcgt gaggcgatcg	1620
	ccaagattga gaaggaggc ggtatcgtag ccggcatcgt ggtcgcctg gaccgatgg	1680
	agaagctgcc ttcgccgat ggcgacgata gcaagccgat gccgagcgcg atcggtagc	1740
	tccggaagga gtatggtctg cccatcttg cgattctcac ctggacgac atcatcgagg	1800
	gtatcaagg gctcgcaccc gaggaggaca tcaggaggac ggaggagtac cggccaagt	1860
	acaaggcgc cgactaggcg ggaggaacaa atctagattg tgcaaaccgc tctgtgaata	1920
	caaaaaaaaa aaaaaaaaa ggcgtcggg tagtgacccc acgctgatt cgggggtaaa	1980
	gcgtgtgta gcctagcgt actttgtaa gtgcatttc tgattgaaa ggctcgacce	2040
	ccaatgtaa gcacaaagag gcaagtccag agtagaaag gccattatt acgtggcgc	2100
	accaccaga gacggcgcc cccaaggct acgaaatcaa ttcggcgta gcgcatagcc	2160
	gtcgagattc ccagtcacgg ccggtggcc gagtgaaat caacttctca tctgctcatg	2220
	tacgtccgat ccagaattcg gttgaccag gccgatgcc ccagaacccc gccttgttg	2280
	tttcgggct tctttgctc cgtgctctcg ttctagaaa gtggcgtctt ggtttggcgt	2340
	ggcaaatcgt agacctggag atggtgagat agtctgtaac agcatgcacg ggtagtccc	2400

gcctacacc cgaggcgtag gtcaagagct cttacatgt catgccacc gcccccicca	2460
aactgcgcct tgcteggcc gaggtgcgtg aaccccagcc gctcatagta agacacgagg	2520
tgctacagac atggctcgtc agccagcaga tccccctaa ccaatcetta gggcagaga	2580
ggtagtagt cactcacatc ctgacaaatc agtgccacc ggtccaccag gccgaattc	2640
ttcatctggt ccaaaaagc tttgataatc atttgccga ttccacatcg ctgtaggcgc	2700
gggagcacag cgagagagt caacccgacg gtacgcccg cctcctgatg gccgacgctc	2760
ttgtcaacgg tgcgtgcgc ggggttcgc cactccttc ggtacgcat gtcggcgtcg	2820
gtgacgacgt ccccgcgca tcgtgtcgag atgacgtgg cgagcaggac gtcacggcg	2880
ccgtcggcgc ggcccgtctc gaccggcttg gagtaggca gcgtctcgag gccgaggtt	2940
gcggcacggt ctgggacgac ggtgaggaaa acgccgagac tgagctcggg gcagacggtg	3000
aggcgtaag cgatctgctg gggttcgggt tagatggtt tcagcgtgct tgttgtgtgc	3060
tccttgactg gattgggctg aggattgctt gccttttcgg gcgacgccg atgttcgggc	3120
ttaggaaac aagcatttc cagctcagata caggatggga ggtcgttgat ggttagcggc	3180
ctgatattgg gtgaaaaggt tgagatgaac gggagtgcct tctggaggcg agattccggg	3240
ctgtccttgg ctgccctct cttctggcac agcatctctt gaagcgaagc aaagtctccg	3300
tcgacgtctg aggactcgtc aacggcctgg tcagcctctt ggataggaca aggactctgt	3360
tcctccgaca ggctttctc ggcgcctgcc ataactgcgc cgagtcgggt gtgtcttctc	3420
gcgactgcag ctagctaaga gatcgggccg tcgagccaaa ggagcttagg ttgaccgat	3480
caatgatcgc gctaggtgtg ccaaaaaccg ttaccgtctc ttgaataatt taatatgcag	3540
ctcggggctg aaacccttgc tgcaaacccg gaagtattgt ttctggggc tagccgggtt	3600
ctcaacggtc tcgttgccac acagctcgtc gtccgtaatc ggcgtggagc ttttagcgt	3660
ctaaagggtg gacgtccagg ctatgcactc agagaaacga gatgtgcgcg tagagatcca	3720
gtggcagggt tgcgccgca gatgtcggcg agctggaac agaagctgtg ttaccgtagt	3780
gctggaacgg tagttagaa tcgaatgggg ctaggagtag gagccaaggt gcaattcccc	3840
gccgcgtgc acgtctgcc gtgttaggaa atgctacctg caggtcacia ggctgccggg	3900
tagcatgcc tgggtgccc agcaacctat cggtatatgg agccgggcag ccggccagtc	3960
ggcagccgg gcacgccgg gtattttgaa agcctactc aacctcaatt acttgattag	4020
aaccgaagtt caggtaaagg gtgttaggtg tggctatggg attatggcgt gagaaatcag	4080
aatttaactc ttctttctac ccttgcatct ccccgtact ggctcagacc attcctgatg	4140
acgtgcgctg gaaatactaa atggcatagt acctttcgac ctctgccagc ctccaaggaa	4200
agccccaaaa gcgttctagc agtaggccc tggcgggccg tcgcttcgca actcgggcc	4260
cccgttgaag ctctgtggcg gggcaaagtt gccccagccc tctgacgctc accgcaagt	4320
gggtaactgaa tttcgtttga acctcatctt cagaagacgg aatacttctg tcagattgca	4380
catacacggt agctgtacct tcttccccg cagctcccc atcgtaggc cgttgagctt	4440
ggaatatgat caaagtcgta tgtatgtaca tacattgtag gactctcact gtccgcgaac	4500
agttacagt aaacaagaat ggccgcgtca agtcgtaggc acaagaagcg acggcaagaa	4560
acaacctcga gcaggggtat gatcttctcc gtactgctgt cagacatttg ctgcaccatt	4620
tgatgatgatt ctaccgcgtg ataggagcct agtgttttgg ccgagtcgga tagaagtaac	4680

[0014]

aaatttcgat	tgctgcagca	tggaatcgct	gagggtcaac	tgaatatccg	cccgaactag	4740
agttccccgt	tgtggtttca	ccggtcaacg	ccccctggcg	gatctctatt	ggattgcact	4800
gcaaaaaatct	gcacacgtag	actgctgctg	caccaaccgg	tggtcgagcg	ggttgaaaag	4860
gaacttctga	cgccaaccgc	aaatagaaca	gtcacccaac	acgaagetgt	tcagaatggg	4920
cgcatgcagg	gtgaagatgg	aagccagatc	agatggatcc	tcagagtcc	tcacatgtgc	4980
ggggcagagc	ctccgcgatg	gcttctctcg	cctactcaac	cttacctact	accacactag	5040
ggaagtattc	agatgtaata	ccttaaggta	ccttctctgt	agggtaaacg	cttccgatg	5100
gagggttact	tggagtctgt	tgagatcaca	catcccgcag	ccaattggtc	agcaaaactc	5160
tccccgtctc	tatgcagatg	ggagcagttg	aagtcttgcc	gcagtcgggg	caaaaggata	5220
ctccaagctg	cgtaaatact	tgacaagaca	gcaagacagt	cagtgtctca	ttcgcgtgat	5280
gtgtcgatga	agcgtacctg	aggttgtctc	gccgccaaag	ggicccagta	cttctgactg	5340
cctcgtgctg	aagggtcgtt	aaaaaaaaag	gtggctaccg	tatccggcca	ttgtccgttg	5400
cccagacttc	tagggcccct	ccgctcgtcg	ctccgcttct	tcatcccga	cttccgcgat	5460
gcgtgtgget	actcaaacag	tgcatcttac	ctaggtagaa	atittgtgta	caccggttgc	5520
tcggcaagtc	ttccagagtg	ttcattgaa	aagtggccga	ggagatagaa	tgcccgcate	5580
tcgccattgg	catgcaggca	cactgtttgg	gctctgtcgt	cgttaccctg	ttgttctaca	5640
ctaattgtag	gccgatctac	cccacgtcag	cccctgcct	agattacagg	tagtggggtg	5700
gcaatagccg	acagcgattc	ctcgcggtgg	cctagacca	attacgtgca	ggagggtagt	5760
gatccttgca	gcctcagaat	gctgggtagc	agcatacttc	agccttcttg	aaagccaccc	5820
gtgggtccctg	gatctatttg	caccttatta	tataacctct	ggagccctca	tactcggac	5880
ctcgatacag	agggacgcat	tcgtgcttgc	ataaccgcac	aatccaggcg	cggatgagac	5940
tcgggccttc	ttggttctct	ctacaatac	cgcaacaccc	ctccccccac	cccgaactt	6000
ccaacagata	ctgggcata	tccggccatc	cttcccttcc	ctcctctctg	caaaccaacg	6060
agaccgcccc	gagaagcaca	ccaaaatgta	cgcgaaaatc	cgacgccctg	atgtttcaga	6120
tcgaagcctg	agtactgact	ctccgagctc	aggcgcata	cgctcagtat	taccaattcg	6180
gagccccaga	gcgacgacca	ggacctgctg	tccttgaag	tgtaccccga	gatgacgatc	6240
gagaccttgc	gcagttccat	acaagccgaa	accaccaccc	acccagcgc	ccaacacctc	6300
taccacaatg	gccagctggt	cagcgacaac	tccaagaccc	tggccgagct	cgcgctgact	6360
gacggcgaca	tgctgcacct	ccacgtccgc	gacatgaggg	gcagcacgac	ggttccggca	6420
ggggggggca	ggtcaggacg	tcccgcggcg	cgccagcacc	agccggtgca	ggatcccga	6480
gtgatccgtt	tgcagattct	gggacacccc	aaactgaggg	gcgagttggc	caggtcgcgg	6540
cccgacttgg	tgccggcact	ggaggacccc	cagaggttcc	cacgctgttt	cgccgacagc	6600
ctggaccggg	agcggaggga	gcgcgaggag	cgccagcgc	agattcagct	gctgaattcg	6660
gacccttccg	atgtcgaggc	tcaggcga	atcgaggaga	tcatcccga	ggagcgggtc	6720
atggagaact	tgagaatgc	catggagcac	aaccccgaag	gtaacacagg	gaaacatagc	6780
ctcgcaccgc	acgacgtggt	cccttccaaa	cccgcgctaa	tccgcccctg	tcccagtttt	6840
tggtagcctg	catatgctgt	atatcgaggt	cgaagtcaac	ggatacaagg	tcaaggcgtt	6900
ggtcagctcg	ggcgcgcaag	ccaccatcat	gagccccag	tgcgcccagg	cctgcggcat	6960

[0015]

```

catgcggctc gtcgacaagc gctttgccgg catcgcacgg ggcgtgggaa cggccaacat 7020
catcggccgc gtgactcgg ccccgatcaa gatcggacc ctttccttc cttgcagctt 7080
caccgtcatg gagggcaagc aggtggaact gctgctcggc ctcgacatgc tgaagcgtca 7140
ccaggcgtgc atcgatcttg ccaaggacaa gctgattatc cagggagccg aggtgccgtt 7200
cctgggcccc gccgacattc cgaccgagac cgaggaggcc tatcagcagg agccgaccgt 7260
ccctgggccg gcgggcacga cgatcgcca gcgtccggt gccgtgcatg cgccgagcgc 7320
ggcagctcat gcggcccagt cgagcggtag tggtcctca ggtccgcaa gtgcagcag 7380
accgtcgttc ccagagaac acatcgacca attgatggcg ctagggcct ccgagcagag 7440
agccattcag gcgtggagg caaccggcg gaacgtcag tatcggcca gcttgatttt 7500
ccaggactga tgctattcaa catcctctgg cccatggctt gagacgacgt taacgggta 7560
cgaacttctc tctgaatcgc gaacgggta taaagaaaag atttctagc acgacctga 7620
aagcagctcc gcatggcatt tctggacat agctaagat cggtctgag actacgccga 7680
gagttcgtct ggtaacagat tggtagcgg ataccaaaa aactactgct aaggtgtgtg 7740
aaggtagtt tagagggcac cgggaccgg ccagcgcata gcctagaatt gaca 7794
    
```

```

<210> 8
<211> 691
<212> DNA
<213> 粗糙脉孢菌
    
```

[0016]

```

<400> 8
atggcgctgt gccagaacc cttgcaagag gaacggtacg ccgcaaagca ttgcgaaca 60
gtgtacctac gtccagcaga ctgacaatct tattctctcc ctctaggaag cagtggcgaa 120
aggaccatcc ctttggcttc tatgcccgtc cccagaagaa ccagcagggt gtgctcgact 180
tgaagatttg ggaatgtggt attccgggca aagaaaagac aatctgggag ggtggtctgt 240
tcaagttgac cgtcacgttc cctgatgta tgcgccctgc accggtcggg gtcttttagt 300
agaagaatac acgtgctgac ttgacgttcc ctgcatcgag aatatccac gaaaccccc 360
aagtgtacgc cgaagctag ccctctgcca tcccccttg ctacgatcac atggctaaca 420
tgttggcgtc tgagcaggca aattcgtccc ccctcttttc caceccaatg tctaccatc 480
aggcacggtc tgtctctcga tcctcaacga agaagaggca tggaagccgg ccatcacgat 540
gaagcagatc ctgctcgga tccaggactt getcaatgac cccaatccc agtccccagc 600
ccaggctgag gcgtacaate tcttcaagaa agatcgtcag gaatatgaac gtcggatcaa 660
gcgctcgtg cgggagaatg cgccccata g 691
    
```

```

<210> 9
<211> 157
<212> PRT
<213> 粗糙脉孢菌
    
```

```

<400> 9
Met Ala Leu Cys Gln Asn Arg Leu Gln Glu Glu Arg Lys Gln Trp Arg
1           5           10           15

Lys Asp His Pro Phe Gly Phe Tyr Ala Arg Pro Gln Lys Asn Gln Gln
20           25           30

Gly Val Leu Asp Leu Lys Ile Trp Glu Cys Gly Ile Pro Gly Lys Glu
    
```

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

<210> 10
 <211> 761
 <212> DNA
 <213> 构巢曲霉

[0017]

<400> 10
 atgtcgctct gtttgaatcg tctgactgag gaaaggtagg ctgattgact ccttttacca 60
 gcaactatat actctgttca gttctcgcag taagtcgagt tgagctgaca ggcacactcg 120
 cgcacaggaa gcaatggcga aaagatcate ctttcgcctt ttatgccaaag cctcatcgta 180
 cggcccaggg tgttttgat atgaaacgat gggaatgtgg cattcccggg aaaaaaggga 240
 caatctggga aggtggactt ttcaaattgg acgtcacttt tctgatggt gcgtggcaac 300
 tctataaaag ctcaaagca agttgtgtac tgactgttct ctttcagagt acctaccaa 360
 gccacccaaa tgtatgttcc ttctgctccg ctcatactca caaatatctt tcacgagaat 420
 gtcgcaagac tcttaaatg aataccgctt gcttagaagc taatgctgct gtgctgttct 480
 gggctaactt tatgtcatta tgaacaggca agttcgtgcc agctctgttt catcctaacg 540
 tctaccgctc tgggactggt tgtctgtcaa tcttgaatga agacgaagcg tggaagccgg 600
 caatcacgat taaacaaatc cttcttgta tccaagactt gctcgacgac ccaaatcccg 660
 agtctccagc gcaggcggaa gcatataata tgtacaagaa agacagagct gcttatgaga 720
 agcagtgtaa gcaggtgtc aaggaaaacc ccgctttata a 761

<210> 11
 <211> 157
 <212> PRT
 <213> 构巢曲霉

<400> 11
 Met Ser Leu Cys Leu Asn Arg Leu Thr Glu Glu Arg Lys Gln Trp Arg
 1 5 10 15
 Lys Asp His Pro Phe Ala Phe Tyr Ala Lys Pro His Arg Thr Ala Gln

20 25 30

Gly Val Leu Asp Met Lys Arg Trp Glu Cys Gly Ile Pro Gly Lys Lys
35 40 45

Gly Thr Ile Trp Glu Gly Gly Leu Phe Lys Leu Asp Val Thr Phe Pro
50 55 60

Asp Glu Tyr Pro Thr Lys Pro Pro Lys Cys Lys Phe Val Pro Ala Leu
65 70 75 80

Phe His Pro Asn Val Tyr Pro Ser Gly Thr Val Cys Leu Ser Ile Leu
85 90 95

Asn Glu Asp Glu Ala Trp Lys Pro Ala Ile Thr Ile Lys Gln Ile Leu
100 105 110

Leu Gly Ile Gln Asp Leu Leu Asp Asp Pro Asn Pro Glu Ser Pro Ala
115 120 125

Gln Ala Glu Ala Tyr Asn Met Tyr Lys Lys Asp Arg Ala Ala Tyr Glu
130 135 140

Lys Arg Val Lys Gln Val Val Lys Glu Asn Pro Ala Leu
145 150 155

[0018]

<210> 12
<211> 477
<212> DNA
<213> 新型隐球菌

<400> 12
atgagcaatt tggccaagc gcgacttcac gaggagagga agcaatggag aaaagaccat 60
ccatttgget tctatgccag accaacaag gcgctgatg gaaccttgaa cattatgagt 120
tggaagtgtg gaataccagg gaagctggt actgactggg aaggaggtat ctacgttgtc 180
aagatgaact ttccggacga attcctaca aagcctcaa agtgcaagtt tgatcctccc 240
ctcttcatc ccaacgtcta tcctctgga acaattgtc tctccathtt ggacgaagaa 300
aagtcttggg aaccctcgat caccgtcaag caaatctgtc ttggtatcca agatttgctt 360
gaacatgcta atgtaaata tectgtcag gtggaggcgt atcacatggt caagaacgat 420
cgcacatcct atgacaagc tatcaggcag caggctgttg aacgccgacc caagtag 477

<210> 13
<211> 158
<212> PRT
<213> 新型隐球菌

<400> 13

Met Ser Asn Leu Ala Gln Ala Arg Leu His Glu Glu Arg Lys Gln Trp
1 5 10 15

Arg Lys Asp His Pro Phe Gly Phe Tyr Ala Arg Pro Thr Lys Ala Ala
20 25 30

Asp Gly Thr Leu Asn Ile Met Ser Trp Glu Val Gly Ile Pro Gly Lys
35 40 45

Ala Gly Thr Asp Trp Glu Gly Gly Ile Tyr Val Val Lys Met Asn Phe
 50 55 60

Pro Asp Glu Phe Pro Thr Lys Pro Pro Lys Cys Lys Phe Asp Pro Pro
 65 70 75 80

Leu Phe His Pro Asn Val Tyr Pro Ser Gly Thr Ile Cys Leu Ser Ile
 85 90 95

Leu Asp Glu Glu Lys Ser Trp Lys Pro Ser Ile Thr Val Lys Gln Ile
 100 105 110

Cys Leu Gly Ile Gln Asp Leu Leu Glu His Ala Asn Val Asn Asp Pro
 115 120 125

Ala Gln Val Glu Ala Tyr His Met Phe Lys Asn Asp Arg Thr Ser Tyr
 130 135 140

Asp Lys Arg Ile Arg Gln Gln Ala Val Glu Arg Arg Pro Lys
 145 150 155

<210> 14
 <211> 584
 <212> DNA
 <213> 酿酒酵母

[0019]

<400> 14
 atgagttagtt tgtgtctaca gcgtcttcag gaagaaaggt aagtagtagt tttcctcctt 60
 ttatgcttac attctgtagg catacacaat ttcattccagc ggtatactaa caaatcgatg 120
 aacttaactt gttttacttg aataacagaa aaaaatggag aaaggatcat ccatttggat 180
 tttatgccaa accagttaag aaagctgatg ggtccatgga tttacagaaa tgggaagctg 240
 gstateccagg caaagaaggt aaaaactggg cgggtggtgt gtaccaatt acagtcgaat 300
 atccaaatga atatccttca aaacctccaa aggttaaatt tccagccgga ttttatcatc 360
 caaacgtgta tccaagtggc acaatatggt taagtatgtt aatgaagat caagattgga 420
 gaccgccat cacgttaaaa caaattgttc ttgggttca ggatctttta gactctccaa 480
 atccaaattc cctgctcaa gagcctgcat ggagatcatt ttcaagaaat aaggcggaaat 540
 atgacaagaa agttttgctt caagctaac agtactctaa atag 584

<210> 15
 <211> 157
 <212> PRT
 <213> 酿酒酵母

<400> 15
 Met Ser Ser Leu Cys Leu Gln Arg Leu Gln Glu Glu Arg Lys Lys Trp
 1 5 10 15

Arg Lys Asp His Pro Phe Gly Phe Tyr Ala Lys Pro Val Lys Lys Ala
 20 25 30

Asp Gly Ser Met Asp Leu Gln Lys Trp Glu Ala Gly Ile Pro Gly Lys
 35 40 45

Glu Gly Thr Asn Trp Ala Gly Gly Val Tyr Pro Ile Thr Val Glu Tyr

Thr Gly Ile Gly Met Ser Glu Thr Ile Lys Gly Ile Gln Asp Ala Gly
 180 185 190

Val Ile Ala Cys Ala Lys His Phe Ile Gly Asn Glu Gln Glu His Phe
 195 200 205

Arg Gln Val Pro Glu Ala Gln Gly Tyr Gly Tyr Asn Ile Ser Glu Thr
 210 215 220

Leu Ser Ser Asn Ile Asp Asp Lys Thr Met His Glu Leu Tyr Leu Trp
 225 230 235 240

Pro Phe Ala Asp Ala Val Arg Ala Gly Val Gly Ser Val Met Cys Ser
 245 250 255

Tyr Gln Gln Val Asn Asn Ser Tyr Ala Cys Gln Asn Ser Lys Leu Leu
 260 265 270

Asn Asp Leu Leu Lys Asn Glu Leu Gly Phe Gln Gly Phe Val Met Ser
 275 280 285

Asp Trp Gln Ala Gln His Thr Gly Ala Ala Ser Ala Val Ala Gly Leu
 290 295 300

Asp Met Ser Met Pro Gly Asp Thr Gln Phe Asn Thr Gly Val Ser Phe
 305 310 315 320

[0021] Trp Gly Ala Asn Leu Thr Leu Ala Val Leu Asn Gly Thr Val Pro Ala
 325 330 335

Tyr Arg Leu Asp Asp Met Ala Met Arg Ile Met Ala Ala Leu Phe Lys
 340 345 350

Val Thr Lys Thr Thr Asp Leu Glu Pro Ile Asn Phe Ser Phe Trp Thr
 355 360 365

Asp Asp Thr Tyr Gly Pro Ile His Trp Ala Ala Lys Gln Gly Tyr Gln
 370 375 380

Glu Ile Asn Ser His Val Asp Val Arg Ala Asp His Gly Asn Leu Ile
 385 390 395 400

Arg Glu Ile Ala Ala Lys Gly Thr Val Leu Leu Lys Asn Thr Gly Ser
 405 410 415

Leu Pro Leu Asn Lys Pro Lys Phe Val Ala Val Ile Gly Glu Asp Ala
 420 425 430

Gly Ser Ser Pro Asn Gly Pro Asn Gly Cys Ser Asp Arg Gly Cys Asn
 435 440 445

Glu Gly Thr Leu Ala Met Gly Trp Gly Ser Gly Thr Ala Asn Tyr Pro
 450 455 460

Tyr Leu Val Ser Pro Asp Ala Ala Leu Gln Ala Arg Ala Ile Gln Asp
 465 470 475 480

Gly Thr Arg Tyr Glu Ser Val Leu Ser Asn Tyr Ala Glu Glu Lys Thr
 485 490 495
 Lys Ala Leu Val Ser Gln Ala Asn Ala Thr Ala Ile Val Phe Val Asn
 500 505 510
 Ala Asp Ser Gly Glu Gly Tyr Ile Asn Val Asp Gly Asn Glu Gly Asp
 515 520 525
 Arg Lys Asn Leu Thr Leu Trp Asn Asn Gly Asp Thr Leu Val Lys Asn
 530 535 540
 Val Ser Ser Trp Cys Ser Asn Thr Ile Val Val Ile His Ser Val Gly
 545 550 555 560
 Pro Val Leu Leu Thr Asp Trp Tyr Asp Asn Pro Asn Ile Thr Ala Ile
 565 570 575
 Leu Trp Ala Gly Leu Pro Gly Gln Glu Ser Gly Asn Ser Ile Thr Asp
 580 585 590
 Val Leu Tyr Gly Lys Val Asn Pro Ala Ala Arg Ser Pro Phe Thr Trp
 595 600 605
 Gly Lys Thr Arg Glu Ser Tyr Gly Ala Asp Val Leu Tyr Lys Pro Asn
 610 615 620
 [0022] Asn Gly Asn Gly Ala Pro Gln Gln Asp Phe Thr Glu Gly Val Phe Ile
 625 630 635 640
 Asp Tyr Arg Tyr Phe Asp Lys Val Asp Asp Asp Ser Val Ile Tyr Glu
 645 650 655
 Phe Gly His Gly Leu Ser Tyr Thr Thr Phe Glu Tyr Ser Asn Ile Arg
 660 665 670
 Val Val Lys Ser Asn Val Ser Glu Tyr Arg Pro Thr Thr Gly Thr Thr
 675 680 685
 Ala Gln Ala Pro Thr Phe Gly Asn Phe Ser Thr Asp Leu Glu Asp Tyr
 690 695 700
 Leu Phe Pro Lys Asp Glu Phe Pro Tyr Ile Tyr Gln Tyr Ile Tyr Pro
 705 710 715 720
 Tyr Leu Asn Thr Thr Asp Pro Arg Arg Ala Ser Ala Asp Pro His Tyr
 725 730 735
 Gly Gln Thr Ala Glu Glu Phe Leu Pro Pro His Ala Thr Asp Asp Asp
 740 745 750
 Pro Gln Pro Leu Leu Arg Ser Ser Gly Gly Asn Ser Pro Gly Gly Asn
 755 760 765
 Arg Gln Leu Tyr Asp Ile Val Tyr Thr Ile Thr Ala Asp Ile Thr Asn
 770 775 780

Thr Gly Ser Val Val Gly Glu Glu Val Pro Gln Leu Tyr Val Ser Leu
785 790 795 800

Gly Gly Pro Glu Asp Pro Lys Val Gln Leu Arg Asp Phe Asp Arg Met
805 810 815

Arg Ile Glu Pro Gly Glu Thr Arg Gln Phe Thr Gly Arg Leu Thr Arg
820 825 830

Arg Asp Leu Ser Asn Trp Asp Val Thr Val Gln Asp Trp Val Ile Ser
835 840 845

Arg Tyr Pro Lys Thr Ala Tyr Val Gly Arg Ser Ser Arg Lys Leu Asp
850 855 860

Leu Lys Ile Glu Leu Pro
865 870

<210> 17
<211> 4699
<212> DNA
<213> 人工序列

<220>
<223> 表达盒pCBH-DP1-tCBH

<220>
<221> 尚未归类的特征
<222> (3884)..(3884)
<223> n是a、c、g、或t

[0023]

<400> 17
ggtatccgat ttggggaacg tcgatgaaag tattgcaaaa gtagcagagag ttgcgcaact 60
aactcgctgc cgaagaagct gcggaagaaa gagaacaccg aaagtggaat aacgttacgg 120
atgtcctgac ctcaaagtgg aaaccagccc ttctgtctct atttgggaaa gcggcttgcc 180
cttgaatgcg ctgcactgtg gcacgactac cagtgatcgg gaggagcaaa ctaccctggt 240
ccgttcccttg gtggggcggc actaggccca acttaggggtg atcggaggtc gatgccgagg 300
tctcgttgg tctgggtctt tctcatttcc cggtttgac cccccgttgc acctgctgat 360
cgcccccaa cgccgatgag gttgcgcca gaccgacaat caccgaggct gcattcccaa 420
gtatattgaa gatggcacca ggtaccgggt ttgctgtccc agtcgtttgg tgccaaattt 480
gggagttttt gagcctcaag atctggggaa atcgacctca acttcatac aagttaaagt 540
cgcacacacg gcgagtcca cgaagagaca catttttttc tgaaggctc tctccccgca 600
catcagaaac caccaaatac caagactgca gaagccgggg taagtgggcc accgggacta 660
cactaaaatg cggggagaag cgagatccgt tgcgaaggga agggatgggg tgtgctgagg 720
ctttctccgc tctcgtgccc cttttgcttg aatctagtgt acaccagggt aggtccgaa 780
ggagtatcta cggcagcgtt gttcgtgctg cggttgagagt caggcgggag acgagcaggc 840
gacaggagcc tcgcaccggc acttcggatc gcatttgcgc ggagcgtcaa atacgtcttt 900
ctgcggtcat cagagagcat cgtgaaccaa gtttcttccg caggcgggcc tgggcttcgc 960
agagtcgac tcggcggacg ctttccgtgt caccctgat aacctggctg ccgcgcccag 1020
actcctccaa tgaggtgtgt gtttgcctc gccgaccctt cagcaacctt aatcgttcc 1080

	atcgcacggc tccacgtcct cgaacgatgc cctcagtcgg tgccccggccg tggcaacat	1140
	aacgtgacat cgcgcgccag cctactagcc gctatcgacc ggtaggctt gtcaccgcag	1200
	cgccattct ccatcgggcc tctactctga tccacctcac ccaccgcaag cactagcgag	1260
	cctcaccaga gtgcaagcga cagaccgcc ttggcccttc gtccttgact atctcccaga	1320
	cctcttgcca tcttgccgac gccgccccct ttttttctc ctccccctgc cggcaggtcg	1380
	gtggccccag tcccagatg gcattgtcc gttgtccatg acgaccatc attcgatggc	1440
	tgactggcac actcgtcttg ttgagcacc gacggcccc ggcccgtctc ccacggtacg	1500
	gaacctcgtt gtacagtacc tctcgtaatg atacccaaca ccggggccga gcgctgggag	1560
	ggcggcgttc ccgagaagcc gggaaggcgg ctggccgct gacctttgtg acttggcgat	1620
	ggatcgggcc atggagaatg tccgtccgaa gcgacgcgac aattagcctg gctaccatcg	1680
	atataaattg ggtgattccc agctcttgat gggcgtgtct tctgcctggc agccctcgtc	1740
	ttcagatcaa gcaactgtgt gctgacctc ttccgtcatg cacgtcctct cgaccgctt	1800
	cctcctcggc agcgtgccg tccagaaggt cctcggcagg cccggcagca gcggcctcag	1860
	cgacgtcacc aagcgcagcg tcgacgactt catcagcacc gagacgccga tcgccctcaa	1920
	caacctcctc tgcaacgtcg gccccgacgg ctgcagggcc ttcggcacca gcgccggcgc	1980
	cgtgatgcc agccccagca ccatcgacc cgactactac tacatgtgga cccgcgacag	2040
	cgcctcgtg ttcaagaacc tcatcgacc cttcaccgag acgtacgacg ccggcctcca	2100
	gcgccgcatc gagcagtaca tcaccgcca ggtcacctc cagggcctca gcaacccag	2160
	cggcagcctc gccgacggca gcggcctggg cgagcccaag ttcgagctga ccctcaagcc	2220
[0024]	cttcaccggc aactggggca ggccccagag ggacggcccc gccctccgcg cgatcgccct	2280
	gatcggctac agcaagtggc tcatcaacaa caactaccag agcaccgtca gcaacgtcat	2340
	ctggccatc gtccgcaacg acctcaacta cgtcgcccag tattggaacc agaccggett	2400
	cgacctctg gaggaggtca acggcagcag cttcttcacc gtcgccaacc agcaccgcgc	2460
	cctggtcgag ggcgccacc tcgccgccac cctcggccag agcggcagcg ectacagcag	2520
	cgtcgcccc caggctctct gcttctcca gccttctgg gtcagcagcg gcggctacgt	2580
	cgacagcaac atcaacacca acgaggccg caccggcaag gacgtcaaca gcgtcctcac	2640
	cagcatccac accttcgacc ccaacctcgg ctgcgacgcc ggcaccttc agccctgcag	2700
	cgacaaggcc ctacgcaacc tcaagtgct cgtcgacagc ttccgagca tctacggcgt	2760
	caacaagggc atccccgcc gcgccgccgt cgccatcggc cgctacgccg aggacgtcta	2820
	ctacaacggc aaccctggt acctcgccac gttcgccgcc gccgagcagc tctacgacgc	2880
	catctacgtc tggaagaaga ccggcagcat caccgtcacc gccaccagcc tcgcttctt	2940
	ccaggagctg gtccccggcg tcaccgccg cacctacagc agcagcagct cgaccttac	3000
	caacatcacc aacgccgtgt cgacctacgc cgacgcttc ctcagcgagg ccgccaagta	3060
	cgtccccgcc gaggctcgc tcgccgagca gttcgaccgc aacagcggca cccccctcag	3120
	cgccctccac ctacctggt cgtacgccag cttcctcacc gccaccgccc gcaggccgg	3180
	catcgtcccc cccagctggg ccaacagcag cgcctcgacc atccctcga cctgcagcgg	3240
	cgccagcgtc gtcggcagct acagccgcc caccgccacc tcgttcccc cgagccagac	3300
	cccaagccc ggcgtccct cgggcacccc ctacacccc ctccctgcg ccacccccac	3360

ctcggtcgcc gtcaccttc acgagctggt gacacccag ttcggccaga ccgtcaaggt	3420
cgccggcaac gccgccccc tgggcaactg gtccaccagc gccgccgtgg ccctcgacgc	3480
cgtaactac gccgacaacc acccctctg gatcggcacc gtcaacctg aggccggcga	3540
cgctgctgag tacaagtaca tcaacgtcgg ccaggacggc agcgtcacct gggagagcga	3600
cccaaacac acctacacc tcccccgct cgctgcgtc acccaggtcg tcaaggagga	3660
cacctggcag agctaagaat tcggatccta agtaagtaaa cgaacctctc tgaaggaggt	3720
tctgagacac gcgcattct tctgtatata gttttatit tcaactctga gtgcttcgct	3780
ccaccagtac ataaccttt tttttcacgt aacaaaatgg cttctttca gaccatgtga	3840
accatcttga tccttgacc tcttcagttc tcaacttaac gtanttcgcg ttagtctgta	3900
tgccccagt gcatgtagtt gagataaata ccctggaag tgggtctggg cttttgtggg	3960
acggagccct ctttctgtgg tctggagagc ccgctctcta ccgctacct tcttaccaca	4020
gtacactact cacacattgc tgaactgacc catcataccg tactttatcc tgttaattcg	4080
tgggtctgtc gactattcta tttgctcaaa tggagagcac attcatcggc gcagggatac	4140
acggtttatg gacccaaga gtgtaaggac tattattagt aatattatat gcctctaggc	4200
gccttaactt caacaggcga gcaactacta tcaactttg gtagaccaa ttacaaacga	4260
ccatactgac cgaaaatit gggattccgt ccgctctccc caaccaagct agaagaggca	4320
acgaacagcc aatccccgtg ctaattaaat tatatggttc atttttitaa aaaaaatit	4380
ttcttccat tttcctctcg cttttctttt tcgcatcgta gttgatcaaa gtccaagtca	4440
agcgagctat ttgtctata gctcgggtgc tataatcagt acagcttaga gaggctgtaa	4500
aggtatgata ccacagcagt attcgcgcta taagcggcac tcctagacta attgttacgg	4560
tctacagaag taggtaataa aagcgttaat tgttctaaat actagaggca cttagagaag	4620
ctatctaaat atatatgac cctagcttat tatccctatt agtaagttag ttagctctaa	4680
cctatagata gatgcatgc	4699
<210> 18	
<211> 4741	
<212> DNA	
<213> 人工序列	
<220>	
<223> 表达盒pCHI1-DPI-tCBH	
<400> 18	
ctgcagtccc ttacctatgg gctcctagtc tcgttctct tttgataga tttgtatit	60
gcaacgttgc aaaatgagac atttcaatca tatgtagccg ccagctactg ttagcgtact	120
cagcgttgc caaacggcgg tttttctggg tagcaactgt ccgctgccc ctgagccgtg	180
cgctgcgaa accccttaa gtagcaagta tgttaccgcc gagaccgaca atgctgttgg	240
ttacctgct ggtccatgat tgcaatctag atatcgtcgg ggcttttgc aatcggtit	300
ccctaccac tttttcttt tggacacttt ctcttttga aaatgccgaa atgatgcggc	360
tcgctcacgc cccgaagtcc cgagctgggg ctagatccgt gattgcaacg cgggtcgaac	420
gcgactggg cagacctcgc tcagccttgg tcgtgccgga atggcgggta ctttaccag	480
gtcgggata attacatagg atgcatgtg cgtggattg attgcatcgc tgtcccttt	540
gtatgtgtcc gagagcgaga tatcaacgcg aaaaccgaa tgctcccaac gtcgctctc	600

gttcataggg tctttttttt tcttctgctc catatcatct gtcttgaact aagtgatcat	660
ctgctgtcac gteccgcca atgattgtaa agaatagataa gtgatgctcg ccggggccag	720
gctctgtgaa agttccctct ttggttgacg atcaggtagc gccaacgttg attgggccc	780
ccgtaaaatc cgacctgtc tcttttcgtt gcaagtctcc gcgagaccgt gccaaagcatg	840
ttctccggat cctcaatta cataagttt ggctccagg taggtctgga agctaccac	900
ctcgccaag caaccaatca caaccagacc tcgcgcggtt tcgaccttc tggtttgtct	960
cagggtgac caacgtcctc ccgtggcggg tgcctgtgga tcgaggtcg caggcgagt	1020
ccgggcacgc ggagccccc tcaaagcttg acccttcag agctaggtt cattagcct	1080
tcgaaacaa cccaagccc cgtcgcaacc atcacaacc gccgataacc agatctcgt	1140
aggtccgata aggatcaaaa atggtgtcgg ctgacgttc atgtgccag gcaggaggat	1200
gatccccagg gttgttcgg gcagctccc cagctcggg agggggagg ggaggggaaa	1260
gccctaacta acgttcgtt tctcacggc cgaccggcc atgcttcgg cttgtgagc	1320
gtgggtcaa gggcaacaag aaatgctaag tgcgggacga agacacgcg gcagtaggtc	1380
tcagggtgac ctgcaaaa ccaagtcca ctcgcatgc ctccagcag aacgttgcc	1440
tagaagggtc agggggttg ttgtagacc acgacctgc tgcggcgag cggagggtg	1500
gcttctaca ggcgtgaag ggtcaactc gtgcccagg tggctacca gcgtccatc	1560
aagggaatg agatgatgtt ggctcgtgg caaagaaa acaaggagg tgactctaga	1620
gagatgctc cgagttcac ggtataag cactgtgat gttcacaag ccggcgtact	1680
cctctagac atctatcacc aacatcacca gaaaggtaa gaccaggtg ttgcatatc	1740
cagtcgaaa agagccaaa agcgaaggag cacgaaagca cagcccaatc attcctgct	1800
ttgctactt tctccacca tgcacgtct ctcgaccgc gtctctctc gcagcgtcg	1860
ctccagaag gtctcggca ggcccggcag cagcggcctc agcgacgca ccaagcgag	1920
ctcgaagc acatcagca ccgagacgc gatcgccctc aacaacctc tctgcaact	1980
cgccccgac ggtcagagg ccttcggcac cagcggcgc gccgtgatc ccagccccag	2040
caccatgac ccgactact actacatgt gaccggcag agcggcctc tgttcaagaa	2100
cctcatgac cgttcaccg agacgtacga cggcgccctc cagcggcga tcgagcagta	2160
catcaccgc caggtcacc tccaggcct cagcaaccc agcggcagcc tcgcccagc	2220
cagcggcctg ggcgagcca agttcagct gacctcaag cccttcacc gcaactggg	2280
caggccccag agggacggc ccgcccctc cgcgatcgc ctgatcggt acagcaagt	2340
gtcatcaac aacaactacc agagcaccgt cagcaacgtc atctggccca tcgtccgca	2400
cgacctaac tacgtcgccc agtattgaa ccagaccgc ttcacctt gggaggagg	2460
caacggcagc agcttctca ccgtcgccaa ccagaccgc gccctggtc agggcgccac	2520
cctcggcgc acctcggc agagcggcag cgcctacag agcgtcgcc cccaggtct	2580
ctgttctc cagcgttct gggtagcag cggcggtac gtcgacgca acatcaaac	2640
caacgaggc cgcaccgca aggacgtcaa cagcgtcctc accagatcc acacctcga	2700
ccccaacctc ggtcgcgac ccggcacctt ccagccctc agcgacaag ccctcagca	2760
cctcaaggc gtcgtcaca gttccgcag catctacgc gtaacaagg gcatccccg	2820
cgcgccgccc gtcgcatcg gccctacgc caggacgtc tactacaag gcaaccctg	2880

[0026]

gtacctgcgc acgttcgceg cgcgagca gctctacgac gccatctacg tctggaagaa	2940
gaccggcagc atcaccgtea cgcaccagc cctcgcgttc ttccaggagc tggcctccgg	3000
cgtaaccgce ggcaacctaca gcagcagcag ctcgacctc accaacaatca tcaacgcegt	3060
gtcgacctac gccgacggct tcctcagcga ggccgccaag tacgtccccg ccgacggctc	3120
gctcggcgag cagttcgacc gcaacagcgg caccctcctc agcgcctcc acctcacctg	3180
gtcgtacgcc agcttcctca ccgccaccgc ccgacgggcc ggcatcgtcc cccccagctg	3240
ggccaacagc agcgcctcga ccatccctc gacctgcagc ggccagcagc tcgtcggcag	3300
ctacagccgc ccaccgcca cctcgttccc ccgagccag acccccaagc ccggcgtccc	3360
ctcgggcacc ccctacacc cctcccgtg cgccaccccc acctcggctg ccgtcacctt	3420
ccacgagctg gtgagcacc agttcggcca gaccgtcaag gtcgccgga acgcccgcgc	3480
cctgggcaac tggccacca gcgcccgct ggccctcgc gccgtcaact acgccgacaa	3540
ccaccctc tggatcgga ccgtcaacct cgaggccgga gacgtcgtc agtacaagta	3600
catcaacgtc ggccaggagc gcagcgtcac ctgggagagc gaccccaacc acacctacac	3660
cgccccgcc gtcgcctcgc tcaccaggt cgtcaaggag gacacctggc agagctaaga	3720
attcggatcc taagtaagta aacgaacctc tctgaaggag gttctgagac acgcgcgatt	3780
cttctgtata tagttttatt tttactctg gactgcttc ctcaccagt acataaacct	3840
ttttttcac gtaacaaaat ggcttctttt cagaccatgt gaaccatctt gatgccttga	3900
cctcttcagt tctacttta acgtagtctg cgtttgtctg tatgtcccag ttgcatgtag	3960
ttgagataaa taccctgga agtgggtctg ggcccttctg ggaccggagcc ctctttctgt	4020
ggctcggaga gcccgtctc taccgcctac ctcttaccac cagtacacta ctcacacatt	4080
gctgaactga cccatcatac cgtactttat cctgttaatt cgtgggtctg tcgactatc	4140
tatttgctca aatggagagc acattcatcg gccagggat acacggttta tggaccccaa	4200
gagtgtgaag actattatta gtaatattat atgcctctag gcgcctaac ttcaacaggc	4260
gagcactact aatcaacttt tggtagacc aattacaaac gaccatacgt gccggaatt	4320
ttggattcc gtcgcctc cccaaccaag ctagaagagg caacgaacag ccaatcccgg	4380
tgctaattaa attatatggt tcattttttt taaaaaatt tttcttccc attttcctc	4440
cgcttttctt tttcgcctc tagttgatca aagtccaagt caagcgagct atttgtgcta	4500
tagctcggtg gctataatca gtacagctta gagaggctgt aaaggtatga taccacagca	4560
gtattcgcgc tataagcggc actcctagac taattgttac ggtctacaga agtaggtaat	4620
aaaagcgtta attgttctaa atactagagg cacttagaga agctatctaa atatatattg	4680
accctagctt attatcccta ttagtaagtt agttagctct aacctataga tagatgcatg	4740
c	4741
<210> 19	
<211> 1899	
<212> DNA	
<213> 人工序列	
<220>	
<223> 编码里氏木霉的葡糖淀粉酶的序列， 密码子优化用于在嗜热毁丝霉中表达	
<400> 19	

Ala Gly Ala Ile Glu Ser Arg Lys Val His Gln Lys Pro Leu Ala Arg
 20 25 30

Ser Glu Pro Phe Tyr Pro Ser Pro Trp Met Asn Pro Asn Ala Asp Gly
 35 40 45

Trp Ala Glu Ala Tyr Ala Gln Ala Lys Ser Phe Val Ser Gln Met Thr
 50 55 60

Leu Leu Glu Lys Val Asn Leu Thr Thr Gly Val Gly Trp Gly Ala Glu
 65 70 75 80

Gln Cys Val Gly Gln Val Gly Ala Ile Pro Arg Leu Gly Leu Arg Ser
 85 90 95

Leu Cys Met His Asp Ser Pro Leu Gly Ile Arg Gly Ala Asp Tyr Asn
 100 105 110

Ser Ala Phe Pro Ser Gly Gln Thr Val Ala Ala Thr Trp Asp Arg Gly
 115 120 125

Leu Met Tyr Arg Arg Gly Tyr Ala Met Gly Gln Glu Ala Lys Gly Lys
 130 135 140

Gly Ile Asn Val Leu Leu Gly Pro Val Ala Gly Pro Leu Gly Arg Met
 145 150 155 160

[0029] Pro Glu Gly Gly Arg Asn Trp Glu Gly Phe Ala Pro Asp Pro Val Leu
 165 170 175

Thr Gly Ile Gly Met Ser Glu Thr Ile Lys Gly Ile Gln Asp Ala Gly
 180 185 190

Val Ile Ala Cys Ala Lys His Phe Ile Gly Asn Glu Gln Glu His Phe
 195 200 205

Arg Gln Val Pro Glu Ala Gln Gly Tyr Gly Tyr Asn Ile Ser Glu Thr
 210 215 220

Leu Ser Ser Asn Ile Asp Asp Lys Thr Met His Glu Leu Tyr Leu Trp
 225 230 235 240

Pro Phe Ala Asp Ala Val Arg Ala Gly Val Gly Ser Val Met Cys Ser
 245 250 255

Tyr Gln Gln Val Asn Asn Ser Tyr Ala Cys Gln Asn Ser Lys Leu Leu
 260 265 270

Asn Asp Leu Leu Lys Asn Glu Leu Gly Phe Gln Gly Phe Val Met Ser
 275 280 285

Asp Trp Gln Ala Gln His Thr Gly Ala Ala Ser Ala Val Ala Gly Leu
 290 295 300

Asp Met Ser Met Pro Gly Asp Thr Gln Phe Asn Thr Gly Val Ser Phe
 305 310 315 320

Trp Gly Ala Asn Leu Thr Leu Ala Val Leu Asn Gly Thr Val Pro Ala
 325 330 335

Tyr Arg Leu Asp Asp Met Ala Met Arg Ile Met Ala Ala Leu Phe Lys
 340 345 350

Val Thr Lys Thr Thr Asp Leu Glu Pro Ile Asn Phe Ser Phe Trp Thr
 355 360 365

Asp Asp Thr Tyr Gly Pro Ile His Trp Ala Ala Lys Gln Gly Tyr Gln
 370 375 380

Glu Ile Asn Ser His Val Asp Val Arg Ala Asp His Gly Asn Leu Ile
 385 390 395 400

Arg Glu Ile Ala Ala Lys Gly Thr Val Leu Leu Lys Asn Thr Gly Ser
 405 410 415

Leu Pro Leu Asn Lys Pro Lys Phe Val Ala Val Ile Gly Glu Asp Ala
 420 425 430

Gly Ser Ser Pro Asn Gly Pro Asn Gly Cys Ser Asp Arg Gly Cys Asn
 435 440 445

Glu Gly Thr Leu Ala Met Gly Trp Gly Ser Gly Thr Ala Asn Tyr Pro
 450 455 460

[0030] Tyr Leu Val Ser Pro Asp Ala Ala Leu Gln Ala Arg Ala Ile Gln Asp
 465 470 475 480

Gly Thr Arg Tyr Glu Ser Val Leu Ser Asn Tyr Ala Glu Glu Lys Thr
 485 490 495

Lys Ala Leu Val Ser Gln Ala Asn Ala Thr Ala Ile Val Phe Val Asn
 500 505 510

Ala Asp Ser Gly Glu Gly Tyr Ile Asn Val Asp Gly Asn Glu Gly Asp
 515 520 525

Arg Lys Asn Leu Thr Leu Trp Asn Asn Gly Asp Thr Leu Val Lys Asn
 530 535 540

Val Ser Ser Trp Cys Ser Asn Thr Ile Val Val Ile His Ser Val Gly
 545 550 555 560

Pro Val Leu Leu Thr Asp Trp Tyr Asp Asn Pro Asn Ile Thr Ala Ile
 565 570 575

Leu Trp Ala Gly Leu Pro Gly Gln Glu Ser Gly Asn Ser Ile Thr Asp
 580 585 590

Val Leu Tyr Gly Lys Val Asn Pro Ala Ala Arg Ser Pro Phe Thr Trp
 595 600 605

Gly Lys Thr Arg Glu Ser Tyr Gly Ala Asp Val Leu Tyr Lys Pro Asn
 610 615 620

Asn Gly Asn Gly Ala Pro Gln Gln Asp Phe Thr Glu Gly Val Phe Ile
 625 630 635 640
 Asp Tyr Arg Tyr Phe Asp Lys Val Asp Asp Asp Ser Val Ile Tyr Glu
 645 650 655
 Phe Gly His Gly Leu Ser Tyr Thr Thr Phe Glu Tyr Ser Asn Ile Arg
 660 665 670
 Val Val Lys Ser Asn Val Ser Glu Tyr Arg Pro Thr Thr Gly Thr Thr
 675 680 685
 Ala Gln Ala Pro Thr Phe Gly Asn Phe Ser Thr Asp Leu Glu Asp Tyr
 690 695 700
 Leu Phe Pro Lys Asp Glu Phe Pro Tyr Ile Tyr Gln Tyr Ile Tyr Pro
 705 710 715 720
 Tyr Leu Asn Thr Thr Asp Pro Arg Arg Ala Ser Ala Asp Pro His Tyr
 725 730 735
 Gly Gln Thr Ala Glu Glu Phe Leu Pro Pro His Ala Thr Asp Asp Asp
 740 745 750
 Pro Gln Pro Leu Leu Arg Ser Ser Gly Gly Asn Ser Pro Gly Gly Asn
 755 760 765
 Arg Gln Leu Tyr Asp Ile Val Tyr Thr Ile Thr Ala Asp Ile Thr Asn
 770 775 780
 Thr Gly Ser Val Val Gly Glu Glu Val Pro Gln Leu Tyr Val Ser Leu
 785 790 795 800
 Gly Gly Pro Glu Asp Pro Lys Val Gln Leu Arg Asp Phe Asp Arg Met
 805 810 815
 Arg Ile Glu Pro Gly Glu Thr Arg Gln Phe Thr Gly Arg Leu Thr Arg
 820 825 830
 Arg Asp Leu Ser Asn Trp Asp Val Thr Val Gln Asp Trp Val Ile Ser
 835 840 845
 Arg Tyr Pro Lys Thr Ala Tyr Val Gly Arg Ser Ser Arg Lys Leu Asp
 850 855 860
 Leu Lys Ile Glu Leu Pro
 865 870
 <210> 21
 <211> 413
 <212> PRT
 <213> 人工序列
 <220>
 <223> 布丘氏菌属物种植酸酶变体
 <400> 21
 Asn Asp Thr Pro Ala Ser Gly Tyr Gln Val Glu Lys Val Val Ile Leu
 1 5 10 15

Ser Arg His Gly Val Arg Ala Pro Thr Lys Met Thr Gln Thr Met Arg
 20 25 30
 Asp Val Thr Pro Tyr Thr Trp Pro Glu Trp Pro Val Lys Leu Gly Tyr
 35 40 45
 Ile Thr Pro Arg Gly Glu His Leu Ile Ser Leu Met Gly Gly Phe Tyr
 50 55 60
 Arg Gln Lys Phe Gln Gln Gln Gly Ile Leu Pro Arg Gly Ser Cys Pro
 65 70 75 80
 Thr Pro Asn Ser Ile Tyr Val Trp Thr Asp Val Ala Gln Arg Thr Leu
 85 90 95
 Lys Thr Gly Glu Ala Phe Leu Ala Gly Leu Ala Pro Gln Cys Gly Leu
 100 105 110
 Thr Ile His His Gln Gln Asn Leu Glu Lys Ala Asp Pro Leu Phe His
 115 120 125
 Pro Val Lys Ala Gly Ile Cys Ser Met Asp Lys Thr Gln Val Gln Gln
 130 135 140
 Ala Val Glu Lys Glu Ala Gln Thr Pro Ile Asp Asn Leu Asn Gln Arg
 145 150 155 160
 [0032] Tyr Ile Pro Glu Leu Ala Leu Met Asn Thr Ile Leu Asn Phe Ser Lys
 165 170 175
 Ser Pro Trp Cys Gln Lys His Ser Ala Asp Lys Pro Cys Asp Leu Ala
 180 185 190
 Leu Ser Met Pro Ser Lys Leu Ser Ile Lys Asp Asn Gly Asn Glu Val
 195 200 205
 Ser Leu Asp Gly Ala Ile Gly Leu Ser Ser Thr Leu Ala Glu Ile Phe
 210 215 220
 Leu Leu Glu Tyr Ala Gln Gly Met Pro Gln Val Ala Trp Gly Asn Ile
 225 230 235 240
 His Ser Glu Gln Glu Trp Ala Leu Leu Leu Lys Leu His Asn Val Tyr
 245 250 255
 Phe Asp Leu Met Glu Arg Thr Pro Tyr Ile Ala Arg His Lys Gly Thr
 260 265 270
 Pro Leu Leu Gln Ala Ile Ser Asn Ala Leu Asn Pro Asn Ala Thr Glu
 275 280 285
 Ser Lys Leu Pro Asp Ile Ser Pro Asp Asn Lys Ile Leu Phe Ile Ala
 290 295 300
 Gly His Asp Thr Asn Ile Ala Asn Ile Ala Gly Met Leu Asn Met Arg
 305 310 315 320

Trp Thr Leu Pro Gly Gln Pro Asp Asn Thr Pro Pro Gly Gly Ala Leu
 325 330 335

Val Phe Glu Arg Leu Ala Asp Lys Ser Gly Lys Gln Tyr Val Ser Val
 340 345 350

Ser Met Val Tyr Gln Thr Leu Glu Gln Leu Arg Ser Gln Thr Pro Leu
 355 360 365

Ser Leu Asn Gln Pro Pro Gly Ser Val Gln Leu Lys Ile Pro Gly Cys
 370 375 380

Asn Asp Gln Thr Ala Glu Gly Tyr Cys Pro Leu Ser Thr Phe Thr Arg
 385 390 395 400

Val Val Ser Gln Ser Val Glu Pro Gly Cys Gln Leu Gln
 405 410

<210> 22

<400> 22
 000

<210> 23

<211> 4179

<212> DNA

<213> 人工序列

<220>

<223> pGpdA-pgaIIC1

[0033]

<400> 23
 gaattcttgt atctctacac acaggetcaa atcaataaga agaacggttc gtctttttcg 60
 tttatatctt gcatcgtccc aaagetattg gcgggatatt ctgtttgcag ttgctgact 120
 tgaagtaatc tctgcagatc tttcgacct gaaatacgtc gagcctgctc cgcttggaag 180
 cggcgaggag cctcgtcctg tcacaactac caacatggag tacgataagg gccagttccg 240
 ccagctcatt aagagccagt tcatgggcgt tggcatgatg gccgtcatgc atctgtactt 300
 caagtacacc aaccctcttc tgatccagtc gatcatcccg ctgaagggcg ctttcgaatc 360
 gaatctggtt aagatccacg tcttcgggaa gccagcgact ggtgacctcc agcgtccctt 420
 taaggtgcc aacagctttc tcagccaggc ccagcccaag accgacaagg cctcctcca 480
 gaaccccgag aagaactgga ggggtggtgt caaggaggag taagctcctt attgaagtgc 540
 gaggacggag cgggtgcaag aggatattct tcgctctgta ttatagataa gatgatgagg 600
 aattggaggt agcatagctt catttgatt tgctttccag gctgagactc tagcttgag 660
 catagagggt ccctttggct ttcaatattc tcaagtatct cgagtttgaa cttattcccg 720
 tgaacctttt attcaccaat gagcattgga atgaacatga atctgaggac tgcaatgcc 780
 atgaggtttt cgaatacat ccggatgtcg aaggcttggg gcacctgcgt tggttgaatt 840
 tagaacgtgg cactattgat catccgatag ctctgcaaag ggcgttgcac aatgcaagtc 900
 aaacgttgct agcagttcca ggtggaatgt tatgatgagc attgtattaa atcaggagat 960
 atagcatgat ctctagttag ctaccacaaa aagtcagacg gcgtaaccaa aagtcacaca 1020
 acacaagctg taaggatttc ggcacggcta cggaagacgg agaagccac cttcagtgga 1080

[0034]

ctcgagtacc atttaattct atttgtgttt gatcgagacc taatacagcc cctacaacga	1140
ccatcaaagt cgtatagcta ccagtgagga agtggactca aatcgacttc agcaacatct	1200
cttgataaa cttaagcct aaactataca gaataagatg gtggagagct tataccgagc	1260
tcccaaatct gtccagatca tggttgaccg gtgectggat cttectatag aatcatcctt	1320
attcgttgac ctagctgatt ctggagtgac ccagagggtc atgacttgag cctaaaatcc	1380
gccgectcca ccattttagt aaaaatgtga cgaactcgtg agctctgtac agtgaccggt	1440
gactctttct ggcatgcgga gagacggacg gacgcagaga gaagggtga gtaataagcg	1500
ccactgcgcc agacagctct ggcggctctg aggtgcagtg gatgattatt aatccgggac	1560
cggcccccc tccgccccga agtggaaagg ctggtgtgcc cctcgttgac caagaatcta	1620
ttgcatcatc ggagaatatg gagcttcatc gaatcaccgg cagtaagcga aggagaatgt	1680
gaagccaggg gtgtatagcc gtcggcgaaa tagcatgcca ttaacctagg tacagaagtc	1740
caattgcttc cgatctgcta aaagattcac gagatagtac ctctccgaa gtaggtagag	1800
cgagtacccg gcgcgtaagc tccctaattg gcccatccgg catctgtagg gcgtccaaat	1860
atcgtgecte tectgctttg cccggtgtat gaaaccggaa aggccgctca ggagctggcc	1920
agcggcgag accgggaaca caagctggca gtcgacccat ceggtgctct gcactcgacc	1980
tgctgagtc cctcagtcce tggtaggcag ctttgccccg tctgtccgcc cgggtgtctg	2040
gcggggttga caaggtcgtt gcgtcagtc aacatttgtt gccatatatt cctgctctcc	2100
ccaccagctg ctcttttctt ttctctttct tttccatct tcagtatatt catcttccca	2160
tccaagaacc tttatttccc ctaagtaagt actttgtac atccatactc catccttccc	2220
atcccttatt cctttgaacc tttcagttcg agctttccca ctccatcgca gcttgactaa	2280
cagctacccc gcttgagcag acatcccat gcattccttt gcttcccctc tcgcttacgg	2340
ctctgtggtt ggtgctacgt tcgcttcggc ctccccatc gaggcccgg acagctgac	2400
cttactact gccgtgccc ccaaggccgg taaggctaag tgctccacca tcacctcaa	2460
caacatcgag gtcccgtg gcacgacctt ggacctcacc ggtctgactt cgggcaccaa	2520
ggtcatcttc gaggcacga ccacttttca gtacgaggag tgggctggcc ccctcatctc	2580
gatgtcggc gagcatatca ccgtcactgg tgctcgggc cacctcatca actgcgacgg	2640
cgctcgtgg tgggatgta agggcacctc cgtaagaag aagccgaagt tcttctacgc	2700
ccacggcctg gactccagct cgatcaccgg ctcaacatc aagaacacgc ccctcatggc	2760
cttctccgtg caggecaacg acatcacctt caccgacgtc actatcaaca acgcccagcg	2820
cgacaccag gccggtcaca aactgacgc cttcgacgtg ggtaactcgg tcggcgtcaa	2880
catcatcaag ccgtgggtcc acaaccagga cgattgcctc gcggtcaaca gcggcgagaa	2940
catctggttc actggcgca cgtgcatcgg cggtcacggc ctgagcatcg gctccgtcgg	3000
tgaccgcagc aacaacgtcg tcaagaacgt cacgatcgag cactccaccg tgtccaacag	3060
cgagaacgcc gtccgatca agaccatctc cgggccacg ggcagcgtca gcgagatcac	3120
ctacagcaac atcgtcatga gcggcatctc cgactacggc gtcgtcatcc agcaggacta	3180
cgaggacggc aagcccaccg gtaagcccac taacggtgtg actatccagg acgtcaagct	3240
ggagagcgtg acgggtagcg tcgacagcgg cggcaccgag atctacctcc tgtcgggttc	3300
gggcagctgc tccgactgga cctgggatga cgtcaaggtg accggcgca agaagagcac	3360

tgccctgaag aacttcccgt ccgtcgttag ctgctaagga tccacttaac gttactgaaa	3420
tcatcaaaaca gcttgacgaa tctggatata agatcgttgg tgtcgatgic agctccggag	3480
ttgagacaaa tgggtttcag gatctcgata agatacgttc atttgtccaa gcagcaaaga	3540
gtgccttcta gtgatttaat agctccatgt caacaagaat aaaacgcgtt tcgggtttac	3600
ctcttccaga tacagctcat ctgcaatgca ttaatgcatt ggacctcgca accctagtac	3660
gcccttcagg ctccgcgaa gcagaagaat agcttagcag agtctatitt cattttcggg	3720
agacgagatc aagcagatca acggctgca agagacctac gagactgagg aatccgctct	3780
tggtccacg cgactatata ttgtctctia attgtacttt gatcgtctcc tcttctttac	3840
tctgatagct tgactatgaa aattccgtca ccagcccctg ggttcgcaaa gataattgca	3900
ctgtttcttc ctggaactct caagcctaca ggacacacat tcatcgtagg tataaacctc	3960
gaaaatcatt cctaactaaga tgggtataca atagtaacca tgcattggtt cctagtgaat	4020
gtcctgtaac acccaatacg ccggccgaaa cttttttaca actctctat gagtcgttta	4080
cccagaatgc acaggtacac ttgttttagag gtaatccttc tttctagaag tcctcgtgta	4140
ctgtgtaagc gcccaactca catctccact cgaattccc	4179
<210> 24	
<211> 1089	
<212> DNA	
<213> 人工序列	
<220>	
<223> 来自黑曲霉的聚半乳糖醛酸酶II的编码序列, 其被 密码子优化用于在嗜热毁丝酶 C1中表达	
[0035] <400> 24	
atgcattcct ttgcttccct cctcgettac ggctcgtgg ctggtgctac gttcgttctg	60
gcctccccca tcgaggcccc cgacagctgc accttcaacta ctgccgctgc cgccaaggcc	120
ggtaaaggcta agtgctccac catcacctc aacaacatcg aggtccccgc tggcacgacc	180
ctggacctca ccggtctgac ttcgggcacc aaggtcatct tcgagggcac gaccactttt	240
cagtacgagg agtgggctgg cccctcatc tcgatgtcgg gcgagcatat caccgtcact	300
ggtgcctcgg gccacctcat caactcgcac ggcgctcgt ggtgggatgg taaggcacc	360
tccggtgaaga agaagccgaa gtcttcttac gccacggcc tggactccag ctcgatcacc	420
ggcctcaaca tcaagaacac gccctcatg gccttctccg tgcaggccaa cgacatcacc	480
ttaccgacg tcaactatcaa caacgccgac ggcgacacc agggcgggtca caaactgac	540
gccttcgacg tgggtaactc ggtcggcgtc aacatcatca agcctgggt ccacaaccag	600
gacgattgcc tcgcggtcaa cagcggcgag aacatctggt tcaactggcg cacgtgcatc	660
ggcggtcacg gctgagcat cggctcgtc ggtgaccgca gcaacaacgt cgtcaagaac	720
gtcacgatcg agcactccac cgtgtccaac agcgagaac ccgtccggat caagaccatc	780
tccggcgcca cgggcagcgt cagcgagatc acctacagca acatcgtcat gageggcatc	840
tccgactacg gcgtcgtcat ccagcaggac tacgaggacg gcaagccca cggtaagccc	900
actaacggtg tgactatcca ggacgtcaag ctggagagcg tgacgggtag cgtcgacagc	960
ggcggcaccg agatctacct cctgtcgggt tcgggcagct gctccgactg gacctgggat	1020
gacgtcaagg tgaccggcgg caagaagagc actgcctgca agaacttccc gtccgtcgtc	1080
agctgctaa	1089

<210> 25
 <211> 362
 <212> PRT
 <213> 黑曲霉
 <400> 25
 Met His Ser Phe Ala Ser Leu Leu Ala Tyr Gly Leu Val Ala Gly Ala
 1 5 10 15
 Thr Phe Ala Ser Ala Ser Pro Ile Glu Ala Arg Asp Ser Cys Thr Phe
 20 25 30
 Thr Thr Ala Ala Ala Lys Ala Gly Lys Ala Lys Cys Ser Thr Ile
 35 40 45
 Thr Leu Asn Asn Ile Glu Val Pro Ala Gly Thr Thr Leu Asp Leu Thr
 50 55 60
 Gly Leu Thr Ser Gly Thr Lys Val Ile Phe Glu Gly Thr Thr Thr Phe
 65 70 75 80
 Gln Tyr Glu Glu Trp Ala Gly Pro Leu Ile Ser Met Ser Gly Glu His
 85 90 95
 Ile Thr Val Thr Gly Ala Ser Gly His Leu Ile Asn Cys Asp Gly Ala
 100 105 110
 Arg Trp Trp Asp Gly Lys Gly Thr Ser Gly Lys Lys Lys Pro Lys Phe
 115 120 125
 Phe Tyr Ala His Gly Leu Asp Ser Ser Ser Ile Thr Gly Leu Asn Ile
 130 135 140
 Lys Asn Thr Pro Leu Met Ala Phe Ser Val Gln Ala Asn Asp Ile Thr
 145 150 155 160
 Phe Thr Asp Val Thr Ile Asn Asn Ala Asp Gly Asp Thr Gln Gly Gly
 165 170 175
 His Asn Thr Asp Ala Phe Asp Val Gly Asn Ser Val Gly Val Asn Ile
 180 185 190
 Ile Lys Pro Trp Val His Asn Gln Asp Asp Cys Leu Ala Val Asn Ser
 195 200 205
 Gly Glu Asn Ile Trp Phe Thr Gly Gly Thr Cys Ile Gly Gly His Gly
 210 215 220
 Leu Ser Ile Gly Ser Val Gly Asp Arg Ser Asn Asn Val Val Lys Asn
 225 230 235 240
 Val Thr Ile Glu His Ser Thr Val Ser Asn Ser Glu Asn Ala Val Arg
 245 250 255
 Ile Lys Thr Ile Ser Gly Ala Thr Gly Ser Val Ser Glu Ile Thr Tyr
 260 265 270

[0036]

Ser Asn Ile Val Met Ser Gly Ile Ser Asp Tyr Gly Val Val Ile Gln
 275 280 285

Gln Asp Tyr Glu Asp Gly Lys Pro Thr Gly Lys Pro Thr Asn Gly Val
 290 295 300

Thr Ile Gln Asp Val Lys Leu Glu Ser Val Thr Gly Ser Val Asp Ser
 305 310 315 320

Gly Ala Thr Glu Ile Tyr Leu Leu Cys Gly Ser Gly Ser Cys Ser Asp
 325 330 335

Trp Thr Trp Asp Asp Val Lys Val Thr Gly Gly Lys Lys Ser Thr Ala
 340 345 350

Cys Lys Asn Phe Pro Ser Val Ala Ser Cys
 355 360

<210> 26
 <211> 5855
 <212> DNA
 <213> 人工序列

<220>
 <223> ubc9-amds

<400> 26
 cccgggtggg tgtacttgcg ctctccctct tggaaggcct cccgtccttc taccccacgt 60
 gggaacacgt ggtctaagac tatcaagccg ccatatccgc cgtgaaattc gcagacatgg 120
 gatgagatcg ggggccgaat taaacacacg acgagcggga gagtagcttc atggttaagat 180
 aacaacttgt catgccctcg ttgagcggta ttaaagggcc gggttctccc tcacgacttg 240
 tetaactcgc ttctcatacg ctggccgatc cttcttgaac aagttagtagg cttcagcctg 300
 agctggtgac tcagggttgg ggtcgtctag caggctctga ataccaagaa ggatctgctt 360
 gatcgtgatt gcgggcttcc atgcctctc ctcgttcaag atcgaaagac acacagtgcc 420
 ggatggatag acgttggggt ggaaaagcgg aggcacgaat ttgcctgttg aatgtggatg 480
 atctggttag cttactgtgc gagaagtgcc acagggggga aaaccaata gcgcagaagg 540
 gtggcgttgg gaatagtacg gcagagaggg gaacctacac ttcggcggtc tagtggggta 600
 ttctgcgata gaacgcgggt tagcaggttt gtcgcataaa gcaccctcgc atagaacttca 660
 ctaccatcg gggaaagtaa cgtcaagctt gaaaagcccg ccttcccaga gggtttgcgc 720
 ctttccgggg ataccacact cccatcgttt gaggtcaagt gttccttgag gagtccgatg 780
 cggctttgca tagaagccaa aagggtggtc tcggcgccac tgctttctgt agatagtgcg 840
 gtttggggtt agtttgagga atgacacgcc ataagaggca tatgcaacgt tgtcagccac 900
 atacctctcc tccgtcagac gattctgcca aagcgacatt ttgggatgcg gtttcggagt 960
 acgtaggtgg gctggaaccg aagttcgaaa cgaccaagaa gatcacggtc gtcaccagat 1020
 tcctcaagaa ctcgccggc gagctctccg agatcaagcg gttctacgtc cagaacggca 1080
 aggtcatccc caactccgag tccaccatcc cgggcgtcga gggcaactcc atcaccagg 1140
 actggtgcga ccgccagaag gccgccttcg gcgacgtgac cgacttcag gacaagggcg 1200
 gcatggtcca gatgggcaag gccctcgcgg ggcccatggt cctcgtcatg tccatctggg 1260

[0037]

acgaccacgc cgtaacatg ctctggctcg actccacctg gcccatcgac ggcgccggca	1320
agccggggcg cgagcgcggt gcctgcccc ccacctcggg cgtccccgct gaggtcgagg	1380
ccgaggcccc caactccaac gtcatcttct ccaacatccg cttcggcccc atcggtcca	1440
ccgtctccgg cctgccccgac ggcggcagcg gcaaccccaa cccgccccgc agctcgtcca	1500
ccccggtccc ctctctgtec accacatcct ccggttctc cggccccact ggcggcacgg	1560
gtgtcgctaa gcaactatcta gactggaaac gcaaccctga agggattctt cctttgagag	1620
atggaagcgt gtcatatctc ttcggttcta cggcaggttt ttttctgctc tttcgtagca	1680
tggcatggtc acttcagcgc ttatttacag ttgctggtat tgatttcttg tgcaaatgca	1740
tatctgacac ttattagcta tggagtcacc acatttccca gcaacttccc cacttctct	1800
gcaatcgcca acgtectctc ttactgagt ctccgtccga taacctgcac tgcaaccggt	1860
gccccatggt acgctccgg atcatactct tectgcacga gggcatcaag ctactaacc	1920
gccttgaaac tctattctt cttatcgatg ttcttatccg caaaggtaac cggaacaacc	1980
acgctcgtga aatccagcag gttgatcaca gaggcatacc catagtaccg gaactggtca	2040
tgccgtaccg cagcggtagg cgtaatcggc gcgatgatgg cgtccagttc cttccccgcc	2100
tttctctcag cctcccccca tttctcaagg tactccatct ggtaattcca cttctggaga	2160
tgcggtccc agagctcgtt catgttaaca gctttgatgt tcgggttcag taggtctttg	2220
atatttggaa tcgccggctc gccggatgca ctgatatcgc gcattacgtc ggcgctgccg	2280
tcagcccgct agatatggga gatgagatcg tggccgaaat cgtgcttcta tggcgtccac	2340
ggggtcacgg tgtgaccggc tttggcaggt gcggcgacgg tggttccac gccgcgagg	2400
[0038] ataggagggt gtggaaggac attgccgctc aagtttagt agccgatatt gagccccccg	2460
ttcttgatct tggaggcaat aatgtccgac tcggactggc gccagggcat ggggatgacc	2520
ttggagtcgt atttccatgg ctctgaccg aggacggatt tggatgaagag gcggaggtct	2580
aaataacttc atcagtact gccggtctcg tatatagtat aaaaagcaag aaaggaggac	2640
agtggaggcc tggatatagag caggaaaaga aggaagagge gaaggactca ccctcaacag	2700
agtgcgtaat cggccccgaca acgctgtgca ccgtctctg accctccatg ctgttcgcca	2760
tctttgcata cggcagccgc ccatgactcg gccttagacc gtacaggaag ttgaacgcg	2820
ccggcactcg aatcgagcca ccgatatccg ttctacacc gatgacgcca ccacgaatcc	2880
caacgatcgc accctacca ccagaactgc cgccgcacga ccagttcttg ttgcgtgggt	2940
tgacggtcgc cccgatgatg ttgttactg tctcgagac catcagggtc tgcgggacag	3000
aggtcttgac gtagaagacg gcaccggctt tgcggagcat ggttgcaga accgagtccc	3060
cttcgtcgtc cttgtttagc catgagatgt agccattga tgtttcgtag cctggtggc	3120
atatgttagc tgacaaaaag ggacatctaa cgacttaggg gcaacggtgt acctgactc	3180
gaagctggtc tttgagagag atggggagge catggagtgg accaacgggt ctcttgtgct	3240
ttgcgtagta ttcacgagt tcccttgctt gcgcgagagc ggcgtcaggg aagaactcgt	3300
gggcgcagtt tgtctgcaca gaagccagcg tcagcttgat agtcccataa ggtggcggtg	3360
ttacatctcc ctgagaggta gaggggacc tactaactgc tggcgattg ctgcccgttt	3420
acagaatgct agcgtaactt ccaccgaggt caactctccg gccgccagct tggacacaag	3480
atctgcagcg gaggectctg tgatcttcag ttcggcctct gaaagatcc ccgattctt	3540

tgggaaatca ataacgctgt cttccgcagg cagcgtctgg actttccatt catcagggat	3600
ggtttttgcg aggccggcgc gcttatcagc ggccagtct tcccaggatt gaggcattct	3660
gtgttagctt atagtcagga tgttgctcg acgagtgtaa actgggagtt ggcatgaggg	3720
ttatgtaggc ttctttagcc ccgcatcccc ctcatctcc tcattgatcc cgggggagcg	3780
gatgggtgtg ataagagact aattataggg tttagctggt gcctagctgg tgattgctg	3840
gcttcgccga attttacggg ccaaggaaag ctgcagaacc gcggcactgg taaacggtaa	3900
ttaagctatc agccccatgc taacagttt aaattacgtg tattgctgat aaacaccaac	3960
agagctttac tgaagatgg gagtcacggt gtggcttccc cactgcgatt attgcacaag	4020
cagcgagggc gaacttgact gtcgtcctg agcagcctgc agtcaaacat acatataat	4080
caaccgcgaa gacgtctggc cttgtagaac acgacgctcc ctagcaaac ctgccgtgtc	4140
agcctctacg gttgttactt gcattcagga tgctctccag cgggcgagct attcaaaata	4200
ttcaaagcag gtatctcgta ttgccaggat tcagctgaag caacaggtgc caaggaaatc	4260
tgcgctggtt ctcatctggg cttgctcgtt cctggcgtag atctagaggt accacgacca	4320
agaagatcac ggtcgtcacc cagttcctca agaactcggc cggcgagctc tccgagatca	4380
agcggttcta cgtccagaac ggcaaggtea tcccactc cgagtccacc atcccgggcg	4440
tcgagggcaa ctccatcacc caggactggt gcgaccgcca gaaggccgcc ttcggcgacg	4500
tgaccgactt ccaggacaag ggcggcatgg tcagatggg caaggccctc gcggggccca	4560
tggtcctcgt catgtccatc tgggacgacc acgccgtcaa catgctctgg ctgactcca	4620
cttgcccat cgacggcgc ggcaagccgg gcgccgagcg cggctgcctgc cccaccact	4680
[0039] cggcgctccc cgctgaggtc gaggccgagg ccccaactc caacgtcatc ttctcaaca	4740
tcgcctcgg ccccatcgcc tcaccgtct cggcctgcc cgacggcggc agcggcaacc	4800
ccaaccgcc cgtcagctcg tcaccgccg tcccctctc gtccaccaca tcctccggtt	4860
ctccggccc gactggcggc acgggtgtcg ctaagcacta ttcgaaaagg ctgggttgac	4920
ggcgagaggg ttgagaaggt tgggagggag aggggaagag gagaagtag tgaggcgggg	4980
agaacaaggt ctgaggtatt taaagatag tgaggtagc aaattaaggt cggccgtgaa	5040
tgcgagacg aagtaacct gtcttggcgc tgcagctgct catcgtcggg tttgatccc	5100
cacagccaat cagtccccg cgccgctgca ggttatcgt gtcacatgac ctcatagatc	5160
cgcatcgaac tacgtacttg ggtcgatatg tccatcagcc cgtcttcgat ttactaaatt	5220
cctttgcctt gggttacagg accaatggaa tcttccatgt gtccttgcat tgtacattgt	5280
agtttattgg gtagacacag cttacatgca gtaaatagtt cagaagatcg ccgtgagcca	5340
gtcaaaagca catagtagaa gataccaatg cccaacgag ggaactttac tggcactaag	5400
tgacagctat tttagactct ccaaccacga cggcaactc ctacacttta ttgcgtagct	5460
gcgttatggt aacggcgaat atttttgtc aaggtgaata aatcatctcg tacaacttgc	5520
gttgaggttg tgaatttaga ccacaatcac tgtcaagga gccgtccgaa cccatttctg	5580
ggagccagtc taattcacgc ttgatgctac atgacttcaa agtagtatag ataccatag	5640
gataaacat gctacctggt tcttaagcac taatgtaat acaaacagtg acctggttaa	5700
taagtgggag ttgttgaag gtatccattc tcatagtagt cttgttggc cactccagac	5760
gaacatgcta gatatggata gatgtgagac aaatgtctct cgttgttatt cttcagttca	5820

	aaacttgtga ggtatcatat acagagttec ttccc	5855
	<210> 27 <211> 21 <212> DNA <213> 人工序列	
	<220> <223> ubc9P1f启动子	
	<400> 27 gatagtgtt agcgacacc g	21
	<210> 28 <211> 20 <212> DNA <213> 人工序列	
	<220> <223> ubc9P1r启动子	
	<400> 28 tcggccacaa atctcagtga	20
	<210> 29 <211> 5268 <212> DNA <213> 人工序列	
	<220> <223> pLH937质粒	
[0040]	<400> 29 cgtattgggc gctcttcgc ttctcgcctc actgactcgc tgcgctcggg cgttcggctg	60
	cgcgagcgg tatcagctca ctcaaaggcg gtaatacggg tatccacaga atcaggggat	120
	aacgcaggaa agaacatgtg agcaaaaggc cagcaaaagg ccaggaaccg taaaaaggcc	180
	gcgttgcctg cgtttttcca taggctccgc ccccctgacg agcatcacia aaatcgacgc	240
	tcaagtcaga ggtggcgaaa cccgacagga ctataaagat accaggegtt tccccctgga	300
	agctccctcg tgcgctctcc tgttccgacc ctgccgetta cggatacct gtccgccttt	360
	ctcccttcgg gaagegtggc gctttctcat agctcacgct gtaggtatct cagttcgggtg	420
	taggtcgttc gctccaagct gggtgtgtg cacgaacccc ccgttcagcc cgaccgctgc	480
	gccttatccg gtaactatcg tcttgagtec aaccggtaa gacacgactt atgccactg	540
	gcagcagcca ctggtaacag gattagcaga gcgaggtatg taggcggtgc tacagagttc	600
	ttgaagtggg gcctaacta cggtacact agaaggacag tatttggtat ctgcgctctg	660
	ctgaagccag ttaccttcgg aaaaagagtt ggtagctctt gatccggcaa aaaaaccacc	720
	gctggtagcg gtggtttttt tgtttgcaag cagcagatta cgcgagaaa aaaaggatct	780
	caagaagatc ctttgatctt ttctacgggg tctgacgctc agtggaaacga aaactcacgt	840
	taagggattt tggatcatgag attatcaaaa aggatcttca cctagatcct tttaaattaa	900
	aaatgaagtt ttaaatacaat ctaaagtata tatgagtaaa cttggtctga cagttaccaa	960
	tgcttaataca gtgaggcacc tatctcagcg atctgtctat ttcgttcate catagttgcc	1020
	tgactccccg tcgtgtagat aactacgata cgggagggct taccatctgg cccagtgct	1080
	gcaatgatac cgcgagaccc acgctcaccg gctccagatt tatcagcaat aaaccagcca	1140
	gccggaaggg ccgagcgag aagtggtcct gcaactttat ccgctccat ccagtcatt	1200

aattgttgc	gggaagctag	agtaagtagt	tcgccagtta	atagtttgcg	caacgttgtt	1260
gccattgcta	caggcatcgt	ggigtcacgc	tcgtcgtttg	gtatggcttc	attcagctcc	1320
ggttcccaac	gatcaaggcg	agttacatga	tccccatgt	tgtcaaaaa	agcggtttagc	1380
tccttcggtc	ctccgatcgt	gtcagaagt	aagtggccg	cagtgttate	actcatggtt	1440
atggcagcac	tgcataatc	tcttactgtc	atgccatccg	taagatgctt	ttctgtgact	1500
ggtgagtact	caaccaagtc	attctgagaa	tagtgtatgc	ggcgaccgag	ttgctcttgc	1560
ccggcgtcaa	tacgggataa	taccgcgcca	catagcagaa	ctttaaagt	gctcatcatt	1620
ggaaaacgtt	cttcggggcg	aaaactctca	aggatcttac	cgctgttgag	atccagttcg	1680
atgtaacca	ctcgtgcacc	caactgatct	tcagcatctt	ttactttcac	cagcgtttct	1740
gggtgagcaa	aaacaggaag	gcaaatgcc	gcaaaaaagg	gaataagggc	gacacggaaa	1800
tgttgaatac	tcatactctt	cttttttcaa	tattattgaa	gcatttatca	gggttattgt	1860
ctcatgagcg	gatacatatt	tgaatgtatt	tagaaaaata	aacaatagg	ggttccgcgc	1920
acatttcccc	gaaaagtgcc	acctggatat	cccagttgca	ggtggtcgac	ctgcagactg	1980
gctgtgtata	agggagcctg	acatttata	tcccagaac	atcaggttaa	tggcgttttt	2040
gatgtcattt	tcgggttgc	tgagatcagc	cacttcttcc	ccgataacgg	agaccggcac	2100
actggccata	tcggtgtca	tcatgcgcca	gctttcatcc	ccgatatgca	ccaccgggta	2160
aagttcacgg	gagactttat	ctgacagcag	acgtgactg	gccaggggga	tcaccatccg	2220
tcgcccggcg	gtgtcaataa	tactctctg	tacatccaca	aacagacgat	aacggctctc	2280
tctttiatag	gtgtaaacct	taaactgcat	ttcaccagcc	ccgttctcg	tcagcaaaag	2340
agccgttcat	ttcaataaac	cggcgacct	cagccatccc	ttctgtatt	tccgctttcc	2400
agcgttcggc	acgcagacga	cgggcttcat	ctgcatggt	tgtgcttacc	agaccggaga	2460
tattgacatc	atataatgct	tgagcaactg	atagctgtcg	ctgtcaactg	tactgtaat	2520
acgtgcttc	atagataacc	tctttttgac	atacttcggg	tatacatatc	agtatatatt	2580
cttataccgc	aaaaatcagc	ggatcctcta	gctagaaaga	aggattacct	ctaaacaagt	2640
gtacctgtgc	attctgggta	aacgactcat	aggagagttg	taaaaaagtt	tcggccggcg	2700
tattgggtgt	tacggagcat	tactaggca	accatggtta	ctattgtata	cccatcttag	2760
taggaatgat	tttcgaggtt	tatacctacg	atgaatgtgt	gtcctgtagg	cttgagagtt	2820
caaggaagaa	acagtgcaat	tatctttgcg	aaccagggg	ctggtgacgg	aattttcata	2880
gtcaagctat	cagagtaaag	aagaggagca	gtcaaaagta	caattagaga	caaatatata	2940
gtcgcgtgga	gccaagagcg	gattcctcag	tctcgtaggt	ctcttgacga	ccgttgatct	3000
gcttgatctc	gtctccgaa	aatgaaaata	gactctgcta	agctattctt	ctgcttcgcc	3060
ggagcctgaa	ggcgtacta	gggttgcgag	gtccaatgca	ttaatgcatt	gcagatgagc	3120
tgtatctgga	agaggtaaac	ccgaaacgcg	ttttattctt	gttgacatgg	agctattaaa	3180
tactagaag	gactctttg	ctgcttgac	aatgaacgt	atcttatcga	gatcctgaac	3240
accatttgtc	tcaactccgg	agctgacatc	gacaccaacg	atcttatatc	cagattcgtc	3300
aagctgtttg	atgatttcag	taacgttaag	tggatcccgg	tcggcatcta	ctctattctc	3360
ttgccctcgg	acgagtgtcg	ggcgtcgggt	ttcactatc	ggcagtgact	tctacacagc	3420
catcgttcca	gacggccgcg	cttctcggg	cgatttgtgt	acgcccgaca	gtcccggctc	3480

[0041]

[0042]

eggatcggac gattgctcg catcgacct gcgccaagc tgcatacgc aaattgccgt	3540
caaccaagct ctgatagagt tggtaagac caatgcggag catatacgc cggaggcgcg	3600
gcgatctgc aagctccgga tgcctcgcet cgaagtagcg cgtctgctgc tccatacaag	3660
ccaaccacgg cctccagaag aagatgttgg cgacctgta ttgggaatcc ccgaacatcg	3720
cctcgcctca gtcaatgacc gctgttatgc ggccattgtc cgtcaggaca ttgttgagc	3780
cgaatccgc gtgcacgagg tgccgactt cggggcagtc ctcggccca agcatcagct	3840
catcgagagc ctgcgcgac gacgcactga cgggtgctgc catcacagt ttgccagtgat	3900
acacatgggg atcagcaatc gcgcatatga aatcacgcca tgtagtgtat tgaccgattc	3960
cttgcggtcc gaatgggccc aaccgcctcg tctggctaag atcggcccga gcgatcgcac	4020
ccatggcctc cgcgaccgga tgcagaacag cgggcagttc ggtttcaggc aggtcttgca	4080
acgtgacacc ctgtgcacgg cgggagatgc aataggtcag gctctcgtg aattccccaa	4140
tgtaagcac ttccggaatc gggagcgcgg ccgatcaaa gtgccgataa acataacgat	4200
ctttgtagaa accatcggcg cagctattta cccgcaggac atatccacgc cctcctacat	4260
cgaagctgaa agcacgagat tcttcgccct ccgagagctg catcaggtcg gagacgctgt	4320
cgaacttttc gatcagaaac ttctcgacag acgtcgcggt gagttcaggc ttttcatat	4380
gggtacctga gaactcttg ttgccctgct ttccgtgca aatactaccg gtacttttgg	4440
gaaacaaggg aacaggaggg cgtctgctgt cgcggttctg agtgttcagg attgaagctg	4500
aagaaggtgc tgaggaagcg tagaactgtt gggacgcga gttctgagaa gagctgtacc	4560
gattggtgaa agccgaagaa gtgagttggt gcctgttgc ctggataatg ttgcaactc	4620
gctggttctg cagagacgga gacaaatgct ggctacgat ttgctgattc aggttgatac	4680
ctcggtcgag atactgtttt gttttgatag ggtggatttg gttgcagaga agaagaaagg	4740
aaggtcaaag agggaaaact gggcggaggg aaggattttg taccagcag caaactgcca	4800
ctgcagtggc cctggcagtg ccggcgagg caccacgca cggccgcgca accggttgg	4860
ccttgccac cacgaaacce ttctgaaagg tcagatgaa gtgtgcgaca gtgcgctcc	4920
ccaagccaat gcaggcgcca tggatccact cccaccgc aagatttcac tgtgcgttct	4980
tattggttgc cgcaaggcca gccaaagggg gaagtatgag tcacagcacc gatacaagaa	5040
aattgcagaa ctaacatatg gatgcgcgc ctattctgta gagctctggg caaagcacca	5100
atctgcggg tcggtacaca cactagcact gcccacctg aggcagtcag ccccgctgac	5160
cgaattgcca agagccaatg gagacgaaa gccaacgctg atggagcacc atctgaatgg	5220
acctcgtcgc cttgcctgga agggacaagg gacaccggag acgcgcc	5268

<210> 30
 <211> 23
 <212> DNA
 <213> 里氏木霉

<400> 30
 gcagttcgac gettaeccac cgg 23

<210> 31
 <211> 23
 <212> DNA
 <213> 里氏木霉

<400> 31

cgacgcttac ccaccgggtg agg 23

<210> 32
 <211> 23
 <212> DNA
 <213> 里氏木霉

<400> 32
 gcgcgactac catcacgtct egg 23

<210> 33
 <211> 413
 <212> PRT
 <213> 人工序列

<220>
 <223> 布丘氏菌属物种植酸酶变体

<400> 33

Asn Asp Thr Pro Ala Ser Gly Tyr Gln Val Glu Lys Val Val Ile Leu
 1 5 10 15

Ser Arg His Gly Val Arg Ala Pro Thr Lys Met Thr Gln Thr Met Arg
 20 25 30

Asp Val Thr Pro Asn Thr Trp Pro Glu Trp Pro Val Lys Leu Gly Tyr
 35 40 45

Ile Thr Pro Arg Gly Glu His Leu Ile Ser Leu Met Gly Gly Phe Tyr
 50 55 60

[0043] Arg Gln Lys Phe Gln Gln Gln Gly Ile Leu Ser Gln Gly Ser Cys Pro
 65 70 75 80

Thr Pro Asn Ser Ile Tyr Val Trp Thr Asp Val Ala Gln Arg Thr Leu
 85 90 95

Lys Thr Gly Glu Ala Phe Leu Ala Gly Leu Ala Pro Gln Cys Gly Leu
 100 105 110

Thr Ile His His Gln Gln Asn Leu Glu Lys Ala Asp Pro Leu Phe His
 115 120 125

Pro Val Lys Ala Gly Ile Cys Ser Met Asp Lys Thr Gln Val Gln Gln
 130 135 140

Ala Val Glu Lys Glu Ala Gln Thr Pro Ile Asp Asn Leu Asn Gln His
 145 150 155 160

Tyr Ile Pro Ser Leu Ala Leu Met Asn Thr Thr Leu Asn Phe Ser Lys
 165 170 175

Ser Pro Trp Cys Gln Lys His Ser Ala Asp Lys Ser Cys Asp Leu Gly
 180 185 190

Leu Ser Met Pro Ser Lys Leu Ser Ile Lys Asp Asn Gly Asn Glu Val
 195 200 205

Ser Leu Asp Gly Ala Ile Gly Leu Ser Ser Thr Leu Ala Glu Ile Phe
 210 215 220

	Leu Leu Glu Tyr Ala Gln Gly Met Pro Gln Ala Ala Trp Gly Asn Ile 225 230 235 240	
	His Ser Glu Gln Glu Trp Ala Leu Leu Leu Lys Leu His Asn Val Tyr 245 250 255	
	Phe Asp Leu Met Glu Arg Thr Pro Tyr Ile Ala Arg His Lys Gly Thr 260 265 270	
	Pro Leu Leu Gln Ala Ile Ser Asn Ala Leu Asn Pro Asn Ala Thr Glu 275 280 285	
	Ser Lys Leu Pro Asp Ile Ser Pro Asp Asn Lys Ile Leu Phe Ile Ala 290 295 300	
	Gly His Asp Thr Asn Ile Ala Asn Ile Ala Gly Met Leu Asn Met Arg 305 310 315 320	
	Trp Thr Leu Pro Gly Gln Pro Asp Asn Thr Pro Pro Gly Gly Ala Leu 325 330 335	
	Val Phe Glu Arg Leu Ala Asp Lys Ser Gly Lys Gln Tyr Val Ser Val 340 345 350	
	Ser Met Val Tyr Gln Thr Leu Glu Gln Leu Arg Ser Gln Thr Pro Leu 355 360 365	
[0044]	Ser Leu Asn Gln Pro Ala Gly Ser Val Gln Leu Lys Ile Pro Gly Cys 370 375 380	
	Asn Asp Gln Thr Ala Glu Gly Tyr Cys Pro Leu Ser Thr Phe Thr Arg 385 390 395 400	
	Val Val Ser Gln Ser Val Glu Pro Gly Cys Gln Leu Gln 405 410	
	<210> 34	
	<211> 21	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> niki 2kb F启动子	
	<400> 34	
	gccatgaacc tcaccacaca g	21
	<210> 35	
	<211> 21	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> niki 2kb R启动子	
	<400> 35	
	cggcgtggcc ctgttctcga g	21
	<210> 36	
	<211> 23	

Phe

<210> 43
 <211> 27
 <212> PRT
 <213> 人工序列

<220>
 <223> 基序2

<400> 43

Lys Pro Pro Lys Cys Lys Phe Thr Pro Pro Leu Phe His Pro Asn Val
 1 5 10 15

Tyr Pro Ser Gly Thr Val Cys Leu Ser Ile Leu
 20 25

<210> 44
 <211> 27
 <212> PRT
 <213> 人工序列

<220>
 <223> 共有区2

<400> 44

Lys Pro Pro Lys Cys Lys Phe Val Pro Pro Leu Phe His Pro Asn Val
 1 5 10 15

[0046]

Tyr Pro Ser Gly Thr Val Cys Leu Ser Ile Leu
 20 25

<210> 45
 <211> 2758
 <212> DNA
 <213> 黑曲霉

<400> 45

tagcgggaag gaactctgta tatgatacct cacaagtttt gaactgaaga atacaagcga 60
 gagacatttg tctcacatct atccatatct agcatgttcg tctggagtgc cacaacaaga 120
 ctactatgag aatggatacc ttccaacaac tcccacttat taaccaggtc actgtttgta 180
 ttaccattag tgcttaagaa ccaggtagca tggtttatcc tatgggtatc tatactactt 240
 tgaagtcatt tagcatcaag cgtgaattag actggctccc agaaatgggt tcggacggct 300
 ccttgacag tgattgtggt ctaaattcac aaactcaacg caagttgtac gagatgattt 360
 attcaccttg acaaaaaata ttcgccgtta ccataacgca gctacgcaat aaaggtgagg 420
 aagttgccgt cgtggttga gagtctaaaa tagctgtcac ttagtgccag taaagttccc 480
 tcgttggggc attggtatct tctactatgt gctttgagct ggctcacggc gatettctga 540
 actatttact gcatgtaagc tgtgtctacc caataaacta caatgtacaa tgcaaggaca 600
 catggaagat tccattggtc ctgtaaccca aggcaaagga atttagtaaa tcgaagacgg 660
 gctgatggac atatcgacce aagtacgtag ttcgatcggg atctatgagg tcatgtgaca 720
 gcgataacct gcagcggcgc cggggactga ttggctgtgg ggcatcaaac ccagcgatga 780
 gcagctgcac gcccaagaca gtgttacttc gtctgcgcac tcacggccga ccttaatttg 840
 cctacctcac tatecttaa atacctcaga cctgtttctc cccgcctcac taccttctcc 900

[0047]

```

tcttcccctc tcctcccaca ctttctcaac cctctcgccg tcaaccacage cttcttcggt      960
tccagcccac ctacgtactc cgaaccgcga tccc aaaatg tcgctttgcc agaactcgtct      1020
gacggaggag aggtatgtgg ctgacaacgt tgcatatgcc tcttatggcg tgtcattctc      1080
caaaactaac ccaaacgcga ctatctacag aaagcagtgg cgccgagacc acccttttgg      1140
cttctatgca aagccgcate ggactcctca aggaacactt gacctcaaac gatgggagtg      1200
tggtatcccc gaaaggcgc aaacctctg ggaaggcggg cttttcaagc ttgacgttac      1260
tttccccgat ggtgagttaa gtctatcgac ggtgtcttta tgcgacaaac ctgctaacce      1320
gcgttctatc gcagaatacc cactaagcc gccgaagtgt aggttcccct ctctgccgta      1380
ctattcccaa cgccaccctt ctgcctatt tggttttccc ccctgtggca cttctcgcac      1440
agtaagctaa ccagatcadc cacattcaac aggcaaatc gtgcctccgc tttccaccc      1500
caacgtctat ccacccgca ctgtgtgtct ttcgatcttg aacgaggagg aggcattgaa      1560
gcccgaate acgatcaagc agatccttct tggtattcag gacctgtag acgaccccaa      1620
ccctgagtea ccagctcagg ctgaagccta caacttgctc aagaaggatc ggccagcgta      1680
tgagaagcga gttagacaag tcgtgagga gaaccggcc ctttaatacc gctcaacgac      1740
ggcatgacaa gttgttatct taccatgaag ctactctccc gctcgtcgtg tgtttaatc      1800
ggccccgat ctcatccat gtctcgcaat ttcacggcgg atatggcggc ttgatagtct      1860
tagaccagct gttcccagc ggggtagaag gacgggaagc cttccaagag ggagagcgca      1920
agtacacca cccgggattt cacgataatg ctgagcgcgc cattcctgtc aaacgggttt      1980
cgaaatgtcg cattggcgca ggtgtgtctt cactaattcc tgtcttggag ggtgtcgaat      2040
actcgggttt cttcttagcg agcagcgtgg cgtacataga actgacaaaa ctggcggtt      2100
cgggacggcg tttgacactc taatgcgatt tatctagctt atgtctcttc tttctcttgg      2160
gatgctcgca cacttctacg tctaccatag ggggtgcatg ttggtgaccg attttcggat      2220
gatcatacca tagaaagacg gcttgtgccc cctgggtggg gggaagggtc ctctctaate      2280
gactccaact caaagtgtct ccaatctgtc atccttagtg cgtaaaagta aaacaagttc      2340
cagaagatc aaaaatctgg agaaagtact tttagtagaa aatgtggggg atgtgcaggc      2400
tgtgaaggga aaatggacaa cagttgaacg ttggaccag agaccagacc caacctgtca      2460
gggcacaacc aataaggatg acatagaaag aaagtaaccg ttgcagaaat tcagtttctg      2520
tgtacgagtt tcgcgctat cttatcgaa gtgagtcact gagatttgtg gccgaaatcg      2580
ccctcgggt tcgccccgc cgccgcgaac ggtgtcgcct tttccaatcc atatttaaga      2640
tttcccgtt catctctctc tttcccctt ggtctctat cttcgattag tcaattgctt      2700
ccttttctt ccttttctgt gtctacttga gtcggttgg caattgatcg tagtcagc      2758

```

<210> 46
 <211> 183
 <212> PRT
 <213> 黑曲霉

<400> 46

Met Ser Leu Cys Gln Asn Arg Leu Thr Glu Glu Arg Tyr Val Ala Asp
 1 5 10 15

Asn Val Ala Tyr Ala Ser Tyr Gly Val Ser Phe Leu Lys Leu Thr Pro

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

[0048]

Val Arg Glu Asn Pro Ala Leu	
180	
<210> 47	
<211> 500	
<212> DNA	
<213> 里氏木霉	
<400> 47	
gtccgctgta ggctcgtc cccaggcctg tgcagccctg tagcctggtc cctgcaccgc	60
tacaccgtcg cgcgtgccc cgcgcgcga ggtcatcaag tgaccgggac atggaggctc	120
ttttcgctac gattgactgc aagggaccga cgggactcaa gagatctctg ccctcccagg	180
gaatcgcact aggcagatcg cggcgtatc cgcagcctcg ccittatccc gggcttcttt	240
ttccatccat ttttaatccg actgtttgcc tttctatca tttccgcca ttctaattca	300
cgccctgggt ttgtgtgtc cctttcttt cttccctcac gagattctca acaacgacaa	360
ggcgtttttc ctttcttcat cctgtcaaga tccgtcgtca cctgcagcag gcagttcgac	420
gcttaccac cgggtgagge agcgcgacta ccatcacgtc tcgggatccg ttgttgaca	480
atattgtcta gcagtcacc	500
<210> 48	
<211> 672	
<212> DNA	
<213> 里氏木霉	
<400> 48	
atggctcttt gccagaacag gttacaggag gagcggtagg cgatgcagac tttcacagc	60

gcctcagatg gatccacgat gctgacggat acccaaccct tagcaagcaa tggcgacgtg 120
 accaccggtt cggattcttt gccaaaccgt cccggacaaa ggagggcgtg ctgcacctga 180
 agaactggga atgcggaatt ccgggcaagg aaaagacgat ctgggaggga gggttgttca 240
 agctcaacat tgcgtttccc gacggtgagt gctcgccgtg cctaataacg tcttggttgg 300
 gcagaaactg acatcagtec cgtttgtcca gagtacccaa caaagccgcc gaaatgtgag 360
 tacagacatc gatcccaggt gcittacggt gtcaatcggg atctaactgt actggcaggc 420
 aaatttgtcc ctccgctggt ccatccaac gtctaccct cgggcaccgt ctgcctgtcg 480
 atttcaacg aggaagaggc ctggaagccc gccatcacgg tcaagcagat tcttctgggc 540
 atccaagacc tgctcaacga cccaacccc gagtcgccc cccaagccga cgcctacaat 600
 ctgttcaaga aagacaagat cgagtacgaa aagcgcatcc ggcgcgtagt tcgcgagaat 660
 ccggcgccgt aa 672

<210> 49
 <211> 474
 <212> DNA
 <213> 里氏木霉

[0049]

<400> 49
 atggctcttt gccagaacag gttacaggag gagcgcaagc aatggcgacg tgaccacccg 60
 ttcgattctt ttgccaacac gtcccggaca aaggaggcgg tgctcgacct gaagaactgg 120
 gaatgcgaa ttccgggcaa ggaaaagacg atctgggagg gaggttgtt caagctcaac 180
 attgctttc ccgacagta cccaacaaag ccgccgaaat gcaaatttgt cctccgctg 240
 ttccatccca acgtctacc ctccggcacc gtctgctgt cgattctcaa cgaggaagag 300
 gcctggaagc cggccatcac ggtcaagcag attcttctgg gcattccaaga cctgctcaac 360
 gacccaaccc ccgagtcgcc cgccaagcc gacgctaca atctgttcaa gaaagacaag 420
 atcgagtacg aaaagcgcac ccggcgcgta gttcgcgaga atccggcgcc gtaa 474

<210> 50
 <211> 157
 <212> PRT
 <213> 里氏木霉

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

Asn Glu Glu Glu Ala Trp Lys Pro Ala Ile Thr Val Lys Gln Ile Leu
 100 105 110
 Leu Gly Ile Gln Asp Leu Leu Asn Asp Pro Asn Pro Glu Ser Pro Ala
 115 120 125
 Gln Ala Asp Ala Tyr Asn Leu Phe Lys Lys Asp Lys Ile Glu Tyr Glu
 130 135 140
 Lys Arg Ile Arg Arg Val Val Arg Glu Asn Pro Ala Pro
 145 150 155

<210> 51
 <211> 1820
 <212> DNA
 <213> 嗜热毁丝酶

<220>
 <221> 尚未归类的特征
 <222> (1720)..(1720)
 <223> n是a、c、g、或t

[0050]

<400> 51
 cctgcagtc cttacctatg ggtcctcctagc ctcgttcctc tttttgatag atttgtattt 60
 tgcaacgttg caaatgaga catttcaatc atatgtagcc gccagctact gttagcgtac 120
 tcagcgttgc ccaaacggcg gtttttctgg gtagcactgt gccgcgtgcc cctgagccgt 180
 gcgtcgcgga aaccctcta agtagcaagt atgttaccgc cgagaccgac aatgctgttg 240
 gttacctcgc tggccatga ttgcaatcta gatatcgtgc ggggcttttg caatcggttt 300
 tcctaccca ctttcttctt ttggacactt tctcttttgg aaaatgccga aatgatgcgg 360
 ctcgctcagc ccccgaagtc ccgagctggg gctagatccg tgattgcaac gcggtgcgaa 420
 cgcgactggg gcagacctcg ctcagccttg gtcgtgccgg aatggcgggt acctttacca 480
 ggtcgggata aattacatag gatgccatgt gcgtggattt gattgcatcg ctgtcccttt 540
 tgtatgtgtc cgagagcgag acatcaacgc gaaaaccgga atgetcccaa egtegetetc 600
 tgttcatagg gtcttttttt ttcttctgct ccatatcacc tgtcttgaac taagtatca 660
 tctgtgtgca cgtcccgcc aatgattgta aagaatgata agtgatgctc gccggggcca 720
 ggctctgtga aagttcctc tttggttgac gatcagtag cgccaacgtt gattgggccc 780
 cccgtaaaat ccgaccctgt ctctttctgt tgcaagtctc cgcgagaccg tgccaagcat 840
 gttctccgga tcctcaatt acataaggtt tggctccagg gtaggtctgg aagctacca 900
 cctcggccaa gcaaccaate acaaccagac ctcgcgcgct ttcgaccttc ctggtttgtc 960
 tcagggcttg ccaacgtcct cccgtggcgg gtgcctgggt atcgcaggtc gcaggcgagt 1020
 gccgggcaag cggagcccc gtcaaagctt gacctttca gagctagggt tcattaggcc 1080
 ttcgaaaaca acccaagccc ccgctgcaac catcacaacc ggccgataac cagatctcgg 1140
 taggtccgat aaggatccaa aatggtgtcg gctgacgttg catgtgccc ggcaggagga 1200
 tgatccccag gttgttgc ggcagctccc gcacgtcggg gagggggagg gggaggggaa 1260
 agccetaact aacgttctgt ctatcacggg ccgaccgggc catgctttcg gcttgtgagc 1320
 ggtgggtgca agggcaacaa gaaatgctaa gtgcgggacg aagacacgcg ggcattgaggt 1380
 ctcagggtga cctgcgcaaa accaagtccc actcgccatg cctccagcag caacgttgcc 1440

gtagaagggt cagggggttt gttgtagacc caccacatg ctgccggcga gcggagggtt	1500
ggcttgctac aggcgctgaa gggcaactc ggtgccccaa gtggctacca agcgtgccat	1560
caagggaaat gagatgatgg tggctcgtgg gcaaagaaaa gacaaggag gtgactctag	1620
agagatgctc tcgagttcac gggataaaga gcaactgtgat cgttcacaaa gccggcgtac	1680
tcctctagag catctatcat caacatcacc agaaaggtcn tagaccaggt ggttgccata	1740
tccagtcgca aaagagccaa agagcgaagg agcacgaaag cacagcccaa tcattccctg	1800
ctttgctact tcttctccac	1820
<210> 52	
<211> 1776	
<212> DNA	
<213> 嗜热毁丝酶	
<400> 52	
ggtatccgat ttggggaacg tcgatgaaag tattgcaaaa gtgacgagag ttgcgcaact	60
aactcgtgc cgaagaagct gcggaagaaa gagaacaccg aaagtggaat aacgttacgg	120
atgtcctgac ctcaaagttg aaaccagccc ttctgtctt atttgggaaa gcgcttgcc	180
cttgaatgcg ctgcaactgtg gcacgactac cagtgatcgg gaggagcaaa ctaccctggt	240
ccgttccttg gtggggcggc actaggccca acttagggtg atcggaggtc gatcccggg	300
tcctcgttg tctggctct tctcatttc cggtttgac cccccgttc acctgctgat	360
cgcccgcaa cgcgatgag gttgcgcca gaccgacaat caccgcgct gcattccaa	420
gtatattgaa gatggcaca ggtaccgggt tttgcgtccc agtcgtttg tgccaaattt	480
[0051] gggagtttt gagcctcaag atctgggaa atcgacctc acttcatac aagttaaagt	540
cgcacacacg gcgagttcca cgaagagaca cattttttc tgaaggcctc tctccccga	600
catcagaaac caccaaatac caagactgca gaagccgggg taagtgggcc accgggacta	660
cactaaaatg cggggagaag cgagatccgt tgcgaaggga agggatgggg tgtgctgagg	720
ctttctccg tctcgtgctg cttttgctt aatctagtgt acaccagggt aggctccgaa	780
ggagtatcta cggcagcgtt gttcgtgctg cgttagagat caggcggag acgagcaggc	840
gacaggagcc tcgaccgct acttcggatc gcatttgcgc ggagcgtcaa atacgtctt	900
ctcgggtcat cagagagcat cgtgaaccaa gtttcttcc caggcggcc tggcctcgc	960
agagtcgac tcggcggacg ccttcctgtt caccctgat aacctggtg ccgcgccag	1020
actctccaa tgagggtgtt ggttgcctc gccgacctt cagcaacctt aatcgttcc	1080
atcgacggc tccagctct cgaacgatc cctcagtcg tccccgccg tggcaacct	1140
aacgtgacat cgcgcccag cctactagcc gctatcgacc ggttaggctt gtcaccgag	1200
cgccattct ccatcgggcc tctactctga tccacctac ccaccgaaag cactagcgag	1260
cctcaccaga gtcaagcga cagaccgc ttggccctt gccttgact atctccaga	1320
cctcttgcca tcttgccgac gcccccctt ttttttctc ctccccctgc cggcaggtcg	1380
gtggccccag tcccagatg gcattgctc gttgtccatg acgacctc attcgtggc	1440
tgactggcac actcgtctt tttgagcat gacggcccgc ggcccgtctc ccacgttacg	1500
gaacctcgtt gtacagtacc tctcgtaatg atacccaaca cggggccga gcgtgggag	1560
ggcggcgttc ccgagaagcc gggaaggcgg ctggccggct gacctttgt acttgccgat	1620

ggatgcggcc atggagaatg tccgtccgaa gcgacgcgac aattagcctg gctaccatcg	1680
atataaattg ggtgattccc agctcttgat gggcgtgtct tctgcctggc agccctcgtc	1740
ttcagatcaa gcaactgtgt gctgactctc ttcctg	1776
<210> 53	
<211> 4164	
<212> DNA	
<213> 人工序列	
<220>	
<223> pCHI-植酸酶-tCBH	
<400> 53	
gtcccttacc tatgggctcc tagtctcgtt cctctttttg atagatttgt attttgcaac	60
gttgcaaaat gagacatttc aatcatatgt agccgccagc tactgttagc gtactcagcg	120
ttgccaaaac ggcggttttt ctgggtagca ctgtgccgcg tgcccctgag ccgtgcgtcg	180
cggaaacccc cttaagtagc aagtatgtta ccgccgagac cgacaatgct gttggttacc	240
tgcgtggtcc atgattgcaa tctagatata gtgcggggct tttgcaatcg gttttcccta	300
cccactttct tcttttgagc actttctctt ttggaaaatg ccgaaatgat gcggctcgtc	360
cacgccccga agtcccagc tggggctaga tccgtgattg caacgcggtg cgaacgcgac	420
tggggcagac ctgcctcagc cttggctcgt ccggaatggc ggttaccttt accaggtcgg	480
gatcaattac ataggatgcc atgtgcgtgg atttgattgc atcgtgtcc cttttgatg	540
tgcccgagag cgagatatca acgcgaaaac cggaatgctc ccaacgtcgc tctctgttca	600
tagggicttt tttttcttc tgctccatat catctgtctt gaactaagtg atcatctgct	660
gtcacgtccc gcccaatgat tgtaaagaat gataagtgat gctcgcgggg gccaggetct	720
gtgaaagtcc cctctttggt tgacgatcag gtagcgccaa cgttgattgg gccgcccgtg	780
aaateccgac ctgtctcctt tcgttgcaag tctccgcgag accgtgccaa gcatgttctc	840
cggatccctc aattacataa ggtttggctc cagggtaggt ctggaagcta cccacctcgg	900
ccaagcaacc aatcacaacc agacctcgcg gcgtttcgac cttcctggtt tgtctcaggg	960
ctggccaacg tctctccggt gcgggtgctt ggtgatcgca ggtcgcagge gaggccggg	1020
cacgcggagc ccccgtaaaa gcttgacctt ttcagagcta ggtttcatta ggccttcgaa	1080
aacaacccaa ggccccgtcg caaccatcac aaccggccga taaccagatc tcggtaggtc	1140
cgataaggat ccaaaatggt gtcggctgac gttgcatgtg cccaggcagg aggatgatcc	1200
ccagggttgt tgccgcgagc tcccgcacgt cggggagggg gagggggagg gaaagccct	1260
aactaacgtt cgttctatca cgggccgacc gggccatgct ttcggcttgt gacgggtggg	1320
gtcaagggca acaagaaatg ctaagtgcgg gacgaagaca cgcgggcatg aggtctcagg	1380
gtgacctcgc caaaaccaag tcccactcgc catgctcca gcagcaactg tgccgtagaa	1440
gggtcagggg gtttgttcta gaccacgac catgctccc gcgagcggag ggttgcttg	1500
ctacagggcg tgaagggtca actcgggtgc caaagtggct accaagcgtg ccatcaaggg	1560
aaatgagatg atggtggctc gtgggcaaag aaaagacaag ggaggtgact ctagagagat	1620
gctctcagat tcacgggtat aagagcactg tgatcgttca caaagccggc gtactcctct	1680
agagcatcta tcatcaacat caccagaaaag gtcaagacca ggtggttgcc atatccagtc	1740
gcaaaagagc caaagagcga aggagcacga aagcacagcc caatcattcc ctgctttgct	1800

[0052]

actttcttc	caccatgcag	accttcggtg	ctttctcgt	ttccttctc	gccgccaggt	1860
aagttggcct	tgatgaacca	tatcatatat	cgccgagaag	tgaccgcgt	gctgagactg	1920
agacagcggc	ctggcccgcc	ccaacgatac	ccctgccagc	ggetaccagg	tcgagaaggt	1980
cgctacctc	agccgccacg	gcgtccgcgc	ccctaccaag	atgaccacaga	ccatgcgcga	2040
cgtaacccc	tacacctggc	ccgagtggcc	cgtaagctc	ggctacatca	cccctcgcgg	2100
cgagcacctc	atcacctca	tgggcggctt	ctaccgccag	aagttccagc	agcagggcat	2160
ctcctctcgc	ggctcgtgcc	ccacccccaa	cagcatctac	gtctggaccg	acgtcgccea	2220
gcgcacctc	aagaccggcg	aggccttctc	cgccggcctc	gccccccagt	gcggcctcac	2280
catccaccac	cagcagaacc	tcgagaaggc	cgacccctc	ttccaccccg	tcaaggccgg	2340
catctgcagc	atggacaaga	cccaggtcca	gcaggccgctc	gagaaggagg	cccagacccc	2400
catcgacaac	ctcaaccagc	gtacatccc	cgagctcgcc	ctcatgaaca	ccatcctcaa	2460
cttcagcaag	agcccttggg	gccagaagca	cagcgcgcgac	aagccctgcg	acctcgcctt	2520
cagcatgcc	agcaagctca	gcatcaagga	caacggcaac	gaggtctccc	tcgacggcgc	2580
tatcggcctc	agetccacc	tcgccgagat	cttctctctc	gagtacccc	agggcatgcc	2640
tcaggtgcc	tgggcaaca	tcacagcga	gcaggagtgg	gcctctctcc	tcaagctcca	2700
caacgtctac	ttcgacctca	tggagcgac	cccctacatc	gcccgccaca	agggcacccc	2760
ctcctccag	gccatcagca	acgccctcaa	ccccaacgcc	accgagagca	agctccccga	2820
catcagcccc	gacaacaaga	tcctcttcat	cgccggccac	gacaccaaca	tcgccaacat	2880
cgccggcatg	ctcaaatgc	gtggaccct	ccccggccag	cccgacaaca	ccccccctgg	2940
cggcgctctc	gtctttgagc	gcctcgcgca	caagtccggc	aagcagtacg	tcagcgtcag	3000
catggctctac	cagacctcgc	agcagctccg	cagccagacc	cccctcagcc	tcaaccagcc	3060
tcccggcagc	gtccagctca	agatccccgg	ctgcaacgac	cagaccgccg	agggctactg	3120
ccccctcagc	accttcacc	gcgtcgtcag	ccagagcgtc	gagccccggt	gccagctcca	3180
gtaataaacg	aacctctctg	aaggaggttc	tgagacacgc	gcgattcttc	tgtatatagt	3240
ttttttttt	actctggagt	gcttcgctcc	accagtacat	aaaccttttt	tttcacgtaa	3300
caaatggct	tcctttcaga	ccatgtgaac	catcttgatg	cttgacctc	ttcagttctc	3360
actttaacgt	agttcgcggt	tgtctgtatg	tcccagttgc	atgtagtga	gataaatacc	3420
cctggaagtg	ggctcgggcc	ttgtgggac	ggagccctct	ttctgtggtc	tggagagccc	3480
gtctctacc	gctaccttc	ttaccacagt	acactactca	cacattgctg	aactgaccca	3540
tcataaccgta	ctttatcctg	ttaattcgtg	gtgctgtcga	ctattctatt	tgctcaaatg	3600
gagagcacat	tcacgcgcgc	agggatacac	ggtttatgga	cccgaagagt	gtaaggacta	3660
ttattagtaa	tattatagc	ctctaggcgc	cttaacttca	acaggcgagc	actactaatc	3720
aacttttggg	agaccaatt	acaacgacc	atacgtccc	gaaattttgg	gattccgtcc	3780
gctctcccca	accaagctag	aagaggcaac	gaacagccaa	tcccgtgct	aattaaatta	3840
tatggttcca	tttttttaaa	aaaatttttt	tcitccatt	ttcctctcgc	ttttcttttt	3900
cgcatcgtag	ttgatcaaac	tccaagtcaa	gcgagctatt	tgtgctatag	ctcgggtgct	3960
ataatcagta	cagcttagag	aggctgtaaa	ggtatgatac	cacagcagta	ttcgcgctat	4020
aagcggcact	cctagactaa	ttgttacggt	ctacagaagt	aggttaataaa	agcgttaatt	4080

[0053]

gttctaaata ctagaggcac ttagagaagc tatctaaata tatattgacc ctagcttatt	4140
atccctatta gtaagttagt tagc	4164
<210> 54	
<211> 4128	
<212> DNA	
<213> 人工序列	
<220>	
<223> pCBH-植酸酶-tCBH	
<400> 54	
aggtatccga ttggggaac gtcgatgaaa gtattgcaaa agtgacgaga gttgcgcaac	60
taactcgctg ccgaagaagc tgcggaagaa agagaacacc gaaagtggaa taacgttacg	120
gatgtcctga cctcaaagtt gaaaccagcc cttcctgctc tatttgggaa agcggcttgc	180
ccttgaatgc gctgcactgt ggcacgacta ccagtgatcg ggaggagcaa actaccctgg	240
tccgttcctt ggtggggcgg cactaggccc aacttagggt gatcggagggt cgatgccgcg	300
gtcctcgttg gtctgggctc ttctcatttc ccggtttgca cccccgttg cacctgctga	360
tcgccccca acgccgatga ggttgcgcc agaccgacaa tcaccgcggc tgcattccca	420
agtatattga agatggcacc aggtaccggg ttttgcgtcc cagtcgtttg gtccaaatt	480
tgggagtttt tgagcctcaa gatctgggga aatcgacctc aacttccata caagttaaag	540
tcgcacacac ggcgagttcc acgaagagac acattttttt ctgaaggcct ctctccccgc	600
acatcagaaa ccaccaaata ccaagactgc agaagccggg gtaagtgggc caccgggact	660
acactaaat gcggggagaa gcgagatccg ttgcgaaggg aagggatggg gtgtgctgcg	720
gctttctccg ctctcgtgcg ccttttgctt gaatctagt tacaccaggg taggctccga	780
aggagtatct acggcagcgc tgttcgtgct gcgttgagag tcagggcgga gacgagcagg	840
cgacaggagc ctgcaccgg cacttcggat cgcatttgcg cggagcgtca aatacgtct	900
tctgcggtea tcagagagca tcgtgaacca aggttcttcc gcagggcggc ctgggcttcg	960
cagagtcgca ctcgccggac gccttccgtg tcaccctga taacctggct gccgcgcca	1020
gactcctcca atgaggtgtg tggttgccct cgccgacctc tcagcaacct taatcgcttc	1080
catcgcacgg ctccacgtcc tcgaacgatg ccctcagtc gtgcccggcc gtggcaacca	1140
taacgtgaca tcgcccca gcctactagc cgctatcgac cggtaggct tgcaccgca	1200
gcgccattc tccatcgggc ctctactctg atccacctca ccaccgcaa gactagcga	1260
gcctcaccag agtgcaagcg acacgaccgg cttggccctt cgtccttgac tatctcccag	1320
acctcttgc atcttgccga cgccgcccc tttttttct cctccccctg ccggcaggtc	1380
ggtggccca gteccagat ggcattgctc cgttgtccat gacgacctc cattcgatg	1440
ctgactgga cactcgtctt gtttagcat cgacggcccg cggcccgtct cccacgtac	1500
ggaacctcgt tgtacgtac ctctcgtaat gataccaac accggggccg agcgtggga	1560
ggcggcgctt cccgagaagc cggaaggcg gctggccggc tgaccttgt gacttggcga	1620
tggatcggc catggagaat gtccgtccga agcgacgca caattagcct ggctaccatc	1680
gatataaatt gggtgattcc cagctcttga tggcgtgtc ttctgcctgg cagccctcgt	1740
cttcagatca agcaactgtg tgctgatcct ctccaccat gcagacctc ggtgcttttc	1800
tcgttccctt cctcgcgcc aggtaagttg gccttgatga accatatcat atatcgcca	1860

[0054]

gaagtggacc gcgtgctgag actgagacag cggcctggcc gcggccaacg ataccctgc	1920
cagcggctac caggtcgaga aggtcgtcat ctcagccgc cacggcgtcc gcgcccctac	1980
caagatgacc cagaccatgc gcgacgtcac cccctacacc tggccccagt ggccccgtaa	2040
gctcggctac ataccctc gcggcgagca cctcatcagc ctcatggcg gcttctaccg	2100
ccagaagtcc cagcagcagg gcatcctccc tcgggctcg tgccccaccc ccaacagcat	2160
ctacgtctgg accgacgtcg cccagcgcac cctcaagacc ggcgaggcct tctcgcgg	2220
cctcgcctcc cagtgcggcc tcaccatcca ccaccagcag aacctcgaga aggccgacct	2280
cctcttccac cccgtcaagg ccggcatctg cagcatggac aagaccagc tccagcaggc	2340
cgtcgagaag gaggcccaga cccccatga caacctcaac cagcgtaca tccccgagct	2400
cgccctcatg aacaccatcc tcaacttcag caagagcccc tggtgccaga agcacagcgc	2460
cgacaagccc tgcgacctcg cctcagcat gccagcaag ctcagcatca aggacaacgg	2520
caacgaggtc tcctcgaag gcgctatcgg ctcagctcc acctcgcgg agatcttct	2580
cctcagtag gccaggcca tgctcaggt cgctggggc aacatccaca gcgagcagga	2640
gtggccctc ctctcaagc tcacaacgt ctacttcgac ctcatggagc gcacccccta	2700
catcgcctgc cacaaggcca cccccctct ccaggccatc agcaacgccc tcaacccaa	2760
cgccaccgag agcaagctcc ccgacatcag ccccacaac aagatcctct tcctcgcgg	2820
ccacgacacc aacatcgcca acatcgcgg catgctcaac atgcgctgga cctccccgg	2880
ccagcccgac aacaccccc ctggcggcgc tctcgtcttt gagcgcctcg ccgacaagtc	2940
[0055] cggaagcag tacgtcagc tcagcatggt ctaccagacc ctgagcagc tccgagcca	3000
gacccccctc agcctcaacc agcctcccg cagcgtccag ctcaagatcc ccgctgcaa	3060
cgaccagacc gccgagggtc actgccccct cagcacctc acccgctcg tcagccagag	3120
cgtcgagccc ggtgcccagc tcagtaata aacgaacctc tctgaaggag gttctgagc	3180
acgcgcgatt ctctgtata tagttttatt ttctactctg gactgcttcg ctccaccagt	3240
acataaacct tttttttcac gtaacaaaat ggcttctttt cagaccatgt gaacctctt	3300
gatgcttga cctcttcagt tctacttta acgtagtctg cgtttgtctg tatgtcccag	3360
ttgcatgtag ttgagataaa taccctgga agtgggtctg gccctttgtg ggacggagcc	3420
ctctttctgt ggtctggaga gccgctctc taccgcctac cttcttacc cagtacacta	3480
ctcacacatt gctgaactga cccatcatac cgtactttat cctgttaatt cgtggtgctg	3540
tcgactatc tatttgctca aatggagagc acattcatc gcgagggat acacggttta	3600
tgaccccaa gagtgtaagg actattatta gtaatattat atgcctctag gcgcctaac	3660
ttcaacagc gagcactact aatcaacttt tggtagacc aattacaac gaccatacgt	3720
gccggaatt ttggattcc gtcgctctc cccaaccaag ctagaagagg caacgaacag	3780
ccaatcccgg tgetaattaa attatatggt tccattttt taaaaaatt ttttcttcc	3840
cattttctc tcgctttct ttttcgcat gtagttgatc aaagtccaag tcaagcagc	3900
tatttgctc atagctcggg ggctataatc agtacagctt agagaggctg taaaggtatg	3960
ataccacagc agtattcgcg ctataagcgg cactcctaga ctaattgta cgtctacag	4020
aagtaggtaa taaaagcgtt aattgttcta aactactagag gcacttagag aagctatcta	4080
aatatatatt gaccctagc tattatccct attagtaagt tagttagc	4128

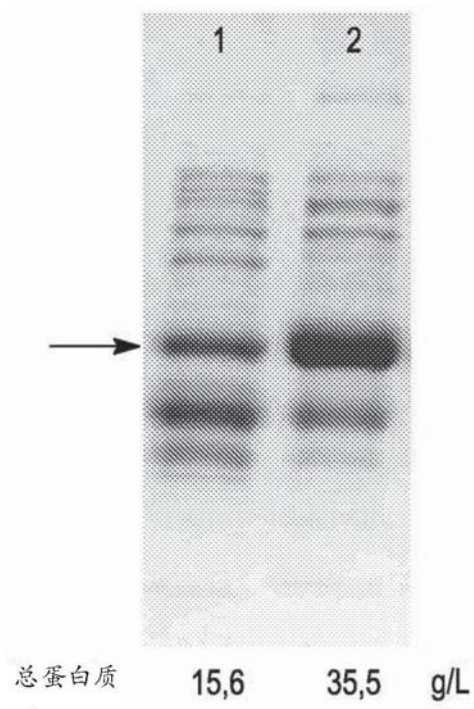


图1

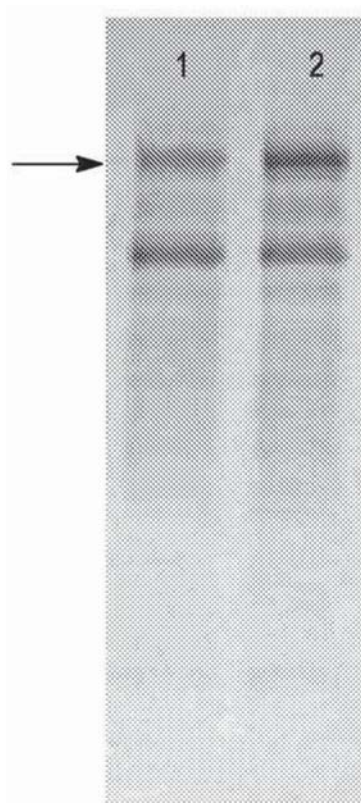


图2

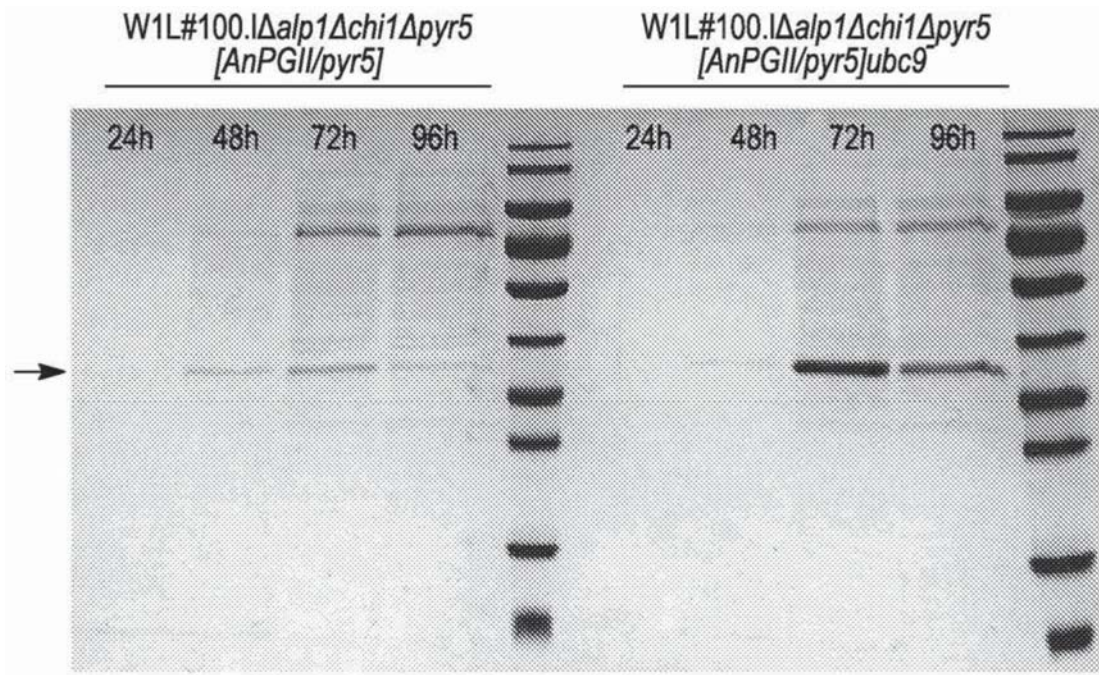


图3

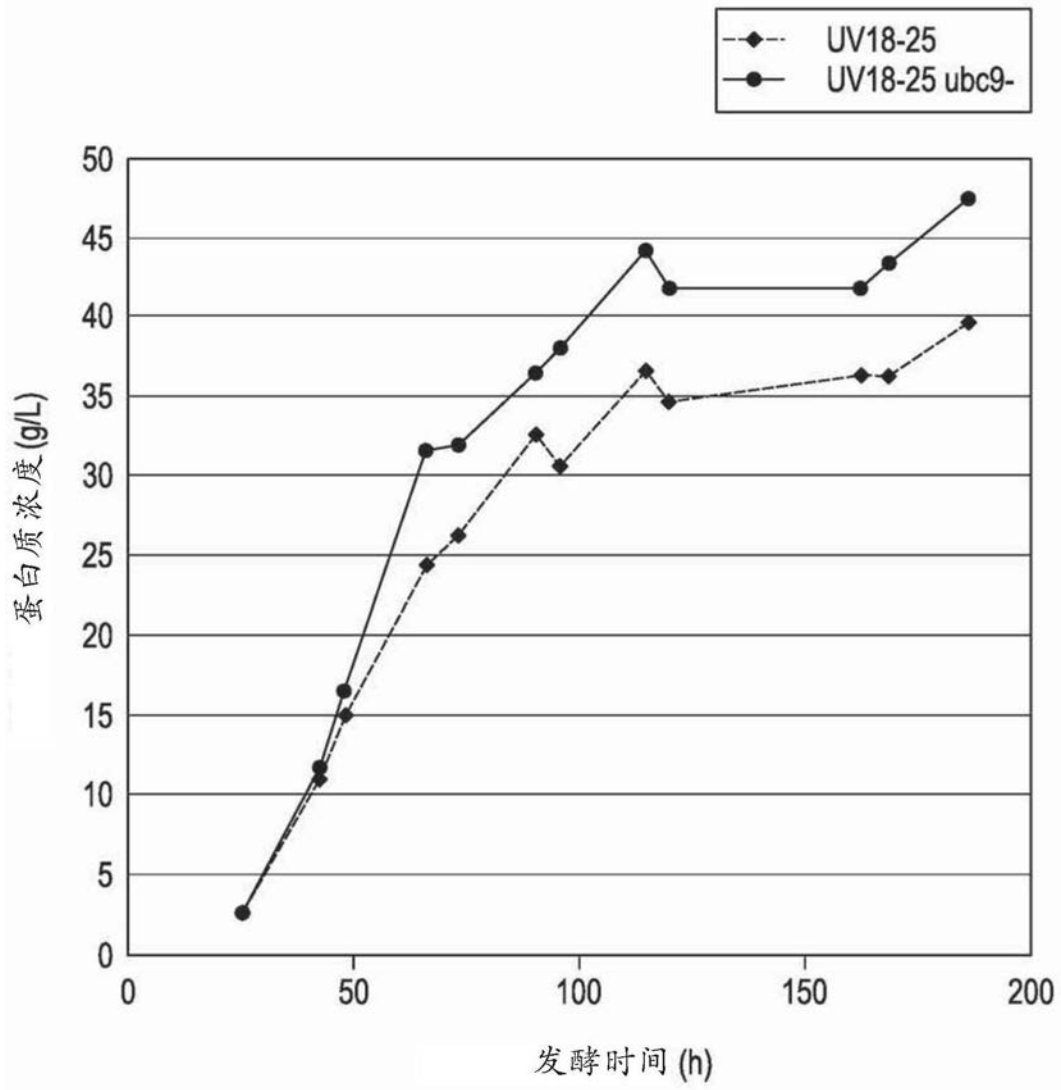


图4

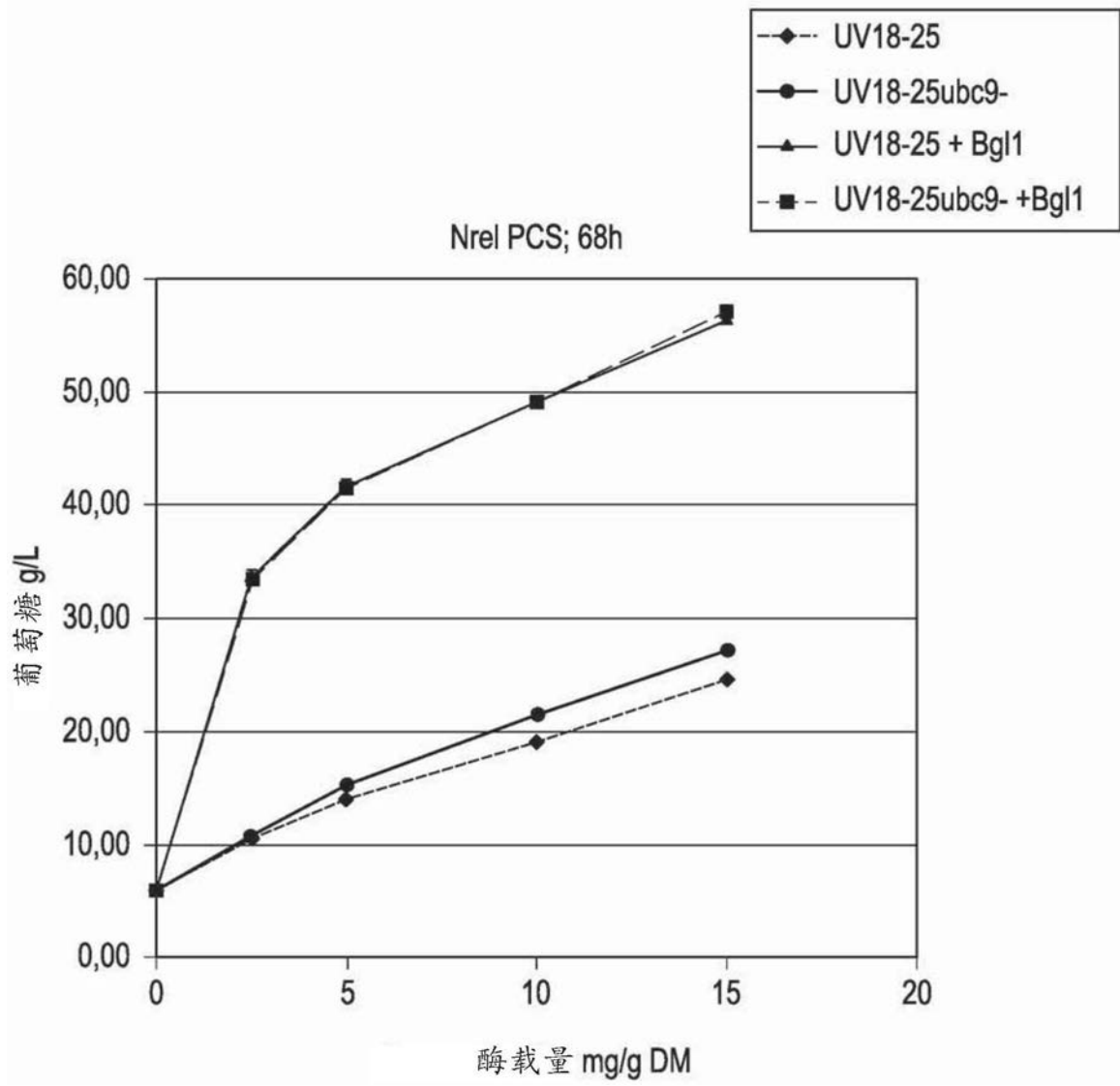


图5

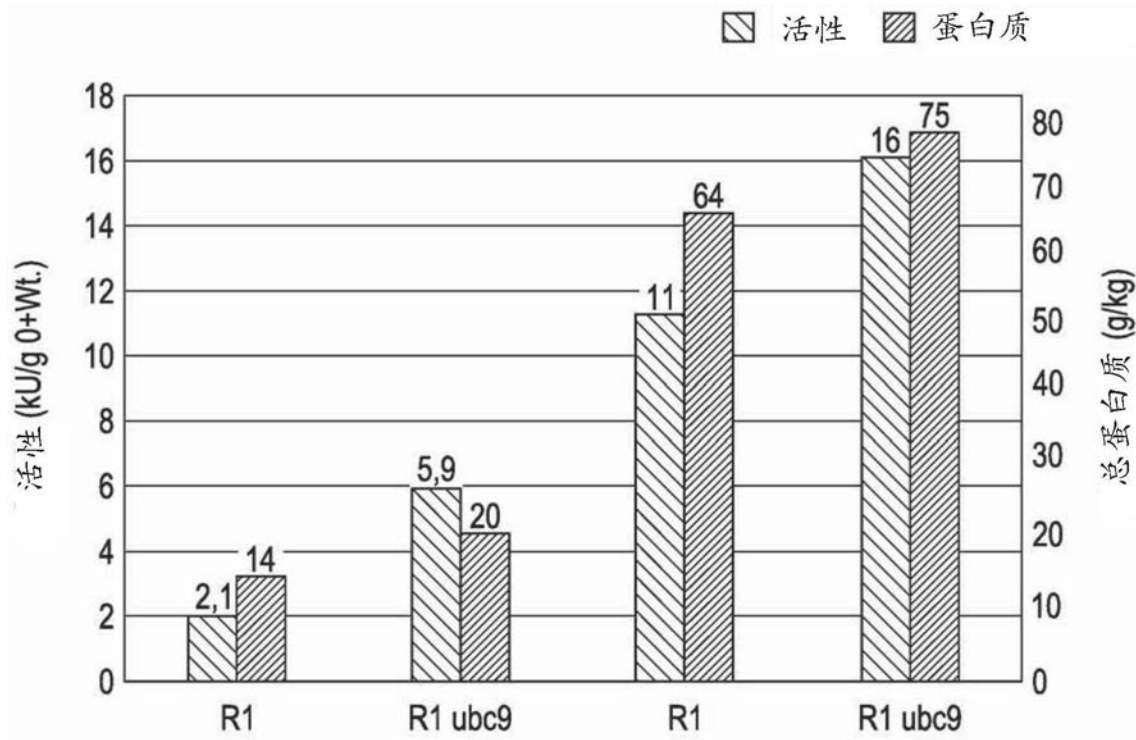


图6

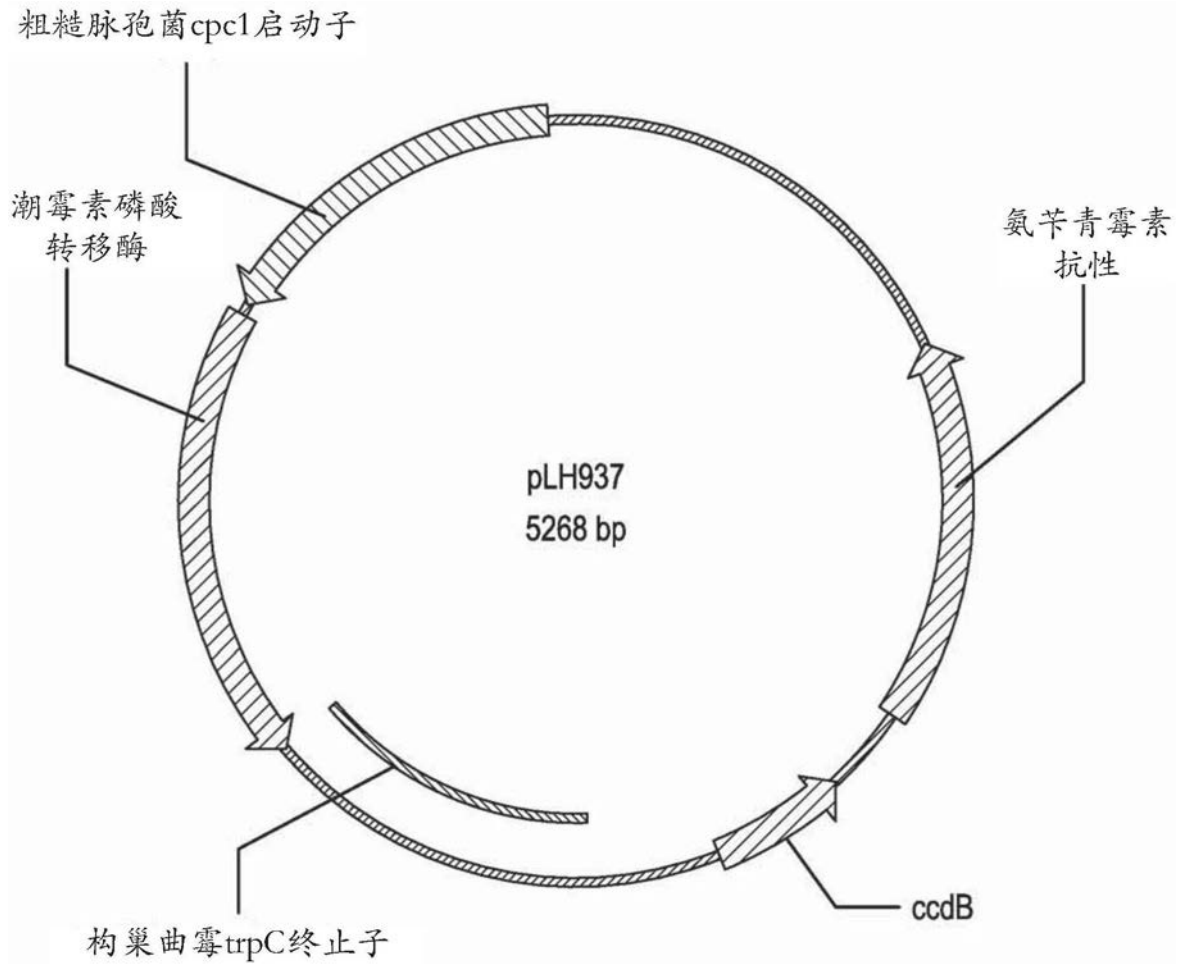


图7

```

<多蛋白序列对齐 (5x158)>
. 001 . 002 . 003 . 004 . 005
-MSLCQNr1qeerkqwrkdhpfgyarpqknaqgvldlkvwecgipgkectmw Mycthe
-MALCQNr1qeerkqwrkdhpfgyarpqknqgvldlkiwecgipgkectiw Neutra
-MSLCLNr1teerkqwrkdhpfafyakphrtaqgvldmkrwecgipgkctgiw Aspnid
MSSLCLQr1qeerkqwrkdhpfgyakpvkkadgsmldlkweagipgkectnw Saccar
MSNLAQAr1heerkqwrkdhpfgyarptkaadgtlnimswevigipgkagtdw Cryneo
*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*
. 001 . 002 . 003 . 004 . 005
RLQEERKQWRKDHFFGF 基序1
RLqEERKqWRkDHFFgF 共有区1
. 006 . 007 . 008 . 009 . 010
EGGLFKLVVTFPDEYPTkpkckftpplfhpnvypsgtvcslsilineeawkpa Mycthe
EGGLFKLTVTFPDEYPTkpkckfvpplfhpnvypsgtvcslsilineeawkpa Neutra
EGGLFKLDVTFPDEYPTkpkckfvpalhfnvypsgtvcslsilineeawkpa Aspnid
AGGVYPIITVEYIPNEYPSkpkvkfpagfyhpnvypsgticlsilnedqdrpa Saccar
EGGIYVVKMNFDFEFTkpkckfdpplfhpnvypsgticlsildeeekswkps Cryneo
*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*
. 006 . 007 . 008 . 009 . 010
KPPCKFTPPLFHPNVYPSGTVCLSIL 基序2
KPPKckFvpplfhPNVYPSGTVCLSIL 共有区2
011 . 012 . 013 . 014 . 015 . 0
ITIKQILLGVQDILLNDPNPESPAQAEAYNMYKKDRVQYERRIRQIVRENAAP Mycthe
ITMKQILLGIQDILLNDPNPESPAQAEAYNLFKKDRQYERRIKRVVRENAAP Neutra
ITIKQILLGIQDILLDDPNPESPAQAEAYNMYKKDRAAYEKRVKQVVENPAL Aspnid
ITLKQIVLVGVQDILLDSPNPNSPAQEPAWRSFSRNKAIEDKVKVLLQAKQYSK- Saccar
ITVKQICLGIQDILLEHANVNDPAQVEAYHMFKNDRTSYDKRIRQQAVERRPK Cryneo
*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*
011 . 012 . 013 . 014 . 015 . 0

```

图8