



(12) 发明专利申请

(10) 申请公布号 CN 103103171 A

(43) 申请公布日 2013. 05. 15

(21) 申请号 201210350893. 3

(51) Int. Cl.

(22) 申请日 2004. 12. 07

C12N 9/42 (2006. 01)

(30) 优先权数据

C12N 15/56 (2006. 01)

2003-409692 2003. 12. 08 JP

C12N 15/63 (2006. 01)

(62) 分案原申请数据

C12N 1/19 (2006. 01)

200480036453. 4 2004. 12. 07

C12N 1/15 (2006. 01)

C12R 1/885 (2006. 01)

C12R 1/645 (2006. 01)

(83) 生物保藏信息

FERM BP - 5977 1997. 06. 13

FERM BP - 5973 1997. 06. 13

FERM BP-10127 2003. 12. 01

FERM BP - 6047 1997. 08. 11

FERM BP - 6045 1997. 08. 11

FERM BP - 5974 1997. 06. 13

(71) 申请人 明治制果药业株式会社

地址 日本东京都

(72) 发明人 渡边学 矢内耕二 露木由美子

(74) 专利代理机构 中国专利代理(香港)有限公司
72001

代理人 郭文洁 权陆军

权利要求书1页 说明书20页

序列表26页 附图2页

(54) 发明名称

耐表面活性剂的纤维素酶及其修饰方法

(57) 摘要

公开了一种抑制在存在表面活性剂时内切糖苷酶活性降低的方法,其特征通过通过修饰具有内切糖苷酶活性, N 末端氨基酸不是焦谷氨酸的蛋白,使其转变为 N 末端具有焦谷氨酸的蛋白。另外,公开了一种具有内切糖苷酶活性,其 N 末端通过氨基酸修饰转变为焦谷氨酸的修饰蛋白,编码该蛋白的多核苷酸,包含该多核苷酸的表达载体,用该表达载体转化的宿主细胞和通过培养宿主细胞而生产蛋白的方法。

1. 选自下述的、N 末端为焦谷氨酸的蛋白质：

(a) 由在 SEQ ID NO :2、4、38 或 40 所示的氨基酸序列中缺失、替换、插入或添加 1 ~ 30 个氨基酸所得的氨基酸序列组成、所述氨基酸序列的 N 末端被转变为焦谷氨酸、且其内切葡聚糖酶活性在存在表面活性剂时降低很小的修饰蛋白质；

(b) 由在 SEQ ID NO :2、4、38 或 40 所示的氨基酸序列中保守替换 1 ~ 30 个氨基酸所得的氨基酸序列组成、所述氨基酸序列的 N 末端被转变为焦谷氨酸、且其内切葡聚糖酶活性在存在表面活性剂时降低很小的修饰蛋白质；及

(c) 同源蛋白质，其由与包含 SEQ ID NO :2、4、38 或 40 所示的氨基酸序列的蛋白质的同源性至少为 85% 的氨基酸序列组成、所述氨基酸序列的 N 末端被转变为焦谷氨酸、且其内切葡聚糖酶活性在存在表面活性剂时降低很小。

2. 选自下述的、编码 N 末端为焦谷氨酸的蛋白质的多核苷酸：

(a) 多核苷酸，其由在 SEQ ID NO :1、3、37 或 39 所示的碱基序列中缺失、替换、插入或添加 1 ~ 90 个碱基所得的碱基序列组成、所述碱基序列的 N 末端密码子编码谷氨酸或谷氨酰胺、且所述多核苷酸编码蛋白质的内切葡聚糖酶活性在存在表面活性剂时降低很小；及

(b) 多核苷酸，其在严紧条件下与 SEQ ID NO :1、3、37 或 39 的碱基序列组成的多核苷酸序列杂交、所述碱基序列的 N 末端密码子编码谷氨酸或谷氨酰胺、且所述多核苷酸编码蛋白质的内切葡聚糖酶活性在存在表面活性剂时降低很小。

3. 包含权利要求 2 的多核苷酸的表达载体。

4. 用权利要求 3 的表达载体转化的宿主细胞。

5. 权利要求 4 的宿主细胞，其中所述宿主细胞是酵母或丝状真菌。

6. 权利要求 5 的宿主细胞，其中，所述丝状真菌是属于腐质霉(*Humicola*) 或木霉属(*Trichoderma*) 的微生物。

7. 权利要求 6 中的宿主细胞，其中，所述丝状真菌是 *Humicola insolens* 或绿色木霉(*Trichoderma viride*)。

8. 生产前述蛋白质的方法，包括：培养权利要求 5~7 任一项中的宿主细胞的步骤，以及从培养获得的宿主细胞或其培养物中回收权利要求 1 的蛋白质的步骤。

9. 根据权利要求 8 的方法生产的蛋白质。

耐表面活性剂的纤维素酶及其修饰方法

[0001] 本申请是申请日为 2004 年 12 月 7 日,申请号为 200480036453.4 的、发明名称和本发明相同的发明专利申请的分案申请。

技术领域

[0002] 本发明涉及一种修饰具有内切葡聚糖酶活性的蛋白(特别是属于家族 45 并具有内切葡聚糖酶活性的纤维素酶)的 N-末端为焦谷氨酸,从而使其成为在表面活性剂时内切葡聚糖酶活性降低很小的纤维素酶的方法,并涉及该纤维素酶。

技术背景

[0003] 纤维素是自然资源中最丰富的资源,因此可有效分解纤维素的纤维素酶系统在各个领域具有广泛用途。在本发明过程中,纯化和测定了多种纤维素酶,进一步,克隆了多种纤维素酶的基因,并通过序列同源性分析而进行家族分类(参见非专利参考文献 1)。

[0004] 另一方面,根据其性质,可将纤维素酶应用于多个工业领域,特别是纺织品加工领域。例如,纤维素酶处理可改善含纤维素的织物的手感和/或外观,或通过“酵素洗涤”,赋予含纤维素的有色织物“石磨”外观,从而使织物的色彩局部性变化。另外,在莱塞尔加工过程中,可用纤维素酶除去加工过程中在织物表面产生的绒毛。在本文中,莱塞尔是一种来源于木浆的再生纤维素织物,最近由于其本身特性,例如高强度或吸水性,并且其生产过程造成的环境污染小而倍受关注。

[0005] 目前认为纤维素酶是通过多个酶的共同作用,例如协同作用来分解纤维素的。在由多个酶组成的纤维素酶组中含有可降低织物韧性的酶等不适于纺织品加工领域的酶。因此,需要通过蛋白分离技术和/或基因工程技术从纤维素酶组中分离适于纺织品加工的酶组分,并生产这些酶组分。特别地,对来源于丝状真菌例如木霉属(*Trichoderma*)或腐质霉属(*Humicola*)的微生物的纤维素酶已进行了深入研究。例如,已分离到腐质霉属中的纤维素酶组分 CBH I、EG V(参见专利文献 1)、NCE2、NCE4 和 NCE5,及木霉属中的 CBH I、CBH II、EG II 和 EG III 等,因此,可通过基因工程技术制备过表达的酶或单一组分的酶等,从而来生产含一种或多种特定的纤维素酶组分为主要组分,适于不同目的的纤维素酶制品。另外,已表明属于家族 45 的纤维素酶,例如 NCE4(参见专利文献 2)、NCE5(参见专利文献 3)、RCE1(参见专利文献 4)或 STCE1(参见国际专利申请 PCT/JP2004/15733)在上述领域是非常有用的。

[0006] 另一方面,当纤维素酶用于衣物洗涤剂时,则不仅要改善纤维素酶组分的量,而且要改善其性质。更具体地,衣物洗涤剂含多种表面活性剂,而且将其溶解于水所获得的溶液是碱性的(pH10-pH 11)。因此,包含在衣物洗涤剂中的纤维素酶要在碱性条件下可耐受多种表面活性剂的作用。Otzen, D. E. 等人已报道向来源于 *Humicola insolens* 的 Cel45 的内部氨基酸序列中引入一个突变,可使其在 pH7,存在线性烷基苯磺酸盐(LAS)的条件下表现出约 3.3 倍于野生型 Cel45 的活性(参见非专利文献 2)。但由该突变提供的在存在表面活性剂时抑制活性降低的作用仅限于 Cel45 或其同源蛋白,并不适于与 Cel45 同源性较低

的,属于家族 45 的内切葡聚糖酶。

[0007] (专利文献 1) 国际专利 W091/17243

[0008] (专利文献 2) 国际专利 W098/03667

[0009] (专利文献 3) 国际专利 W001/90375

[0010] (专利文献 4) 国际专利 W000/24879

[0011] 非专利文献 1:Henrissat B., Bairoch A.Updating the sequence-based classification of glycosyl hydrolases.Biochem. J. 316 :695-696(1996)

[0012] 非专利文献 2:Daniel E.Otzen, Lars Christiansen, Martin Schulein.A comparative study of the unfolding of the endoglucanase Cel45 from Humicola insolens in denaturant and surfactant.Protein Sci. 8 :1878-1887(1999)

[0013] 发明描述

[0014] 本发明要解决的问题

[0015] 本发明的一个目标是提供将具有内切葡聚糖酶活性的蛋白(特别是属于家族 45,并具有内切葡聚糖酶活性的蛋白)转变为具有内切葡聚糖酶活性,而且其活性在存在表面活性剂时降低很小的蛋白的方法;该方法中使用的载体;具有内切葡聚糖酶活性,而且其活性在存在表面活性剂时降低很小的蛋白;及其编码多核苷酸。本发明的另一个目标是用上述微生物提供一种可有效地生产用于洗涤衣物的酶蛋白的微生物。

[0016] 解决问题的方法

[0017] 本发明人进行了广泛研究,发现将焦谷氨酸(以下有时以 pQ 表示)或含 pQ 的肽加到属于家族 45,并具有内切葡聚糖酶活性的蛋白的 N 末端的蛋白,例如 N 末端添加型纤维素酶,与野生型纤维素酶相比较,其内切葡聚糖酶活性在表面活性剂存在时没有显著降低。

[0018] 在存在阴离子表面活性剂时保持高内切葡聚糖酶活性的性质对于衣物洗涤的酶来说是非常有用的,但没有发现这样的纤维素酶。另外,也没有关于通过添加焦谷氨酸或含焦谷氨酸的肽而在存在表面活性剂时保持酶活性的作用的报道。

[0019] 本发明的另一个方面是关于所有 N 末端未被保护而在存在表面活性剂时保持酶活性的纤维素酶,可通过向纤维素酶中加入焦谷氨酸或含焦谷氨酸的肽来抑制在表面活性剂存在时活性降低。可添加焦谷氨酸或含焦谷氨酸的肽的纤维素酶无特别的限定,但优选属于家族 45 的纤维素酶。

[0020] 因此,本发明包括以下方面:

[0021] (1) 一种抑制在存在表面活性剂时内切葡聚糖酶活性降低的方法,其特征在于修饰具有内切葡聚糖酶活性、而 N 末端不是焦谷氨酸的蛋白的 N 末端,使其成为 N 末端具有焦谷氨酸的蛋白;

[0022] (2) (1) 中的方法,其中通过向具有内切葡聚糖酶活性、而 N 末端不是焦谷氨酸的蛋白的 N 末端加入焦谷氨酸或可转变为焦谷氨酸的氨基酸,或 N 末端含有焦谷氨酸或可转变为焦谷氨酸的氨基酸的肽来进行修饰;

[0023] (3) (1) 中的方法,其中通过用焦谷氨酸或可转变为焦谷氨酸的氨基酸,或 N 末端含有焦谷氨酸或可转变为焦谷氨酸的氨基酸的肽替换具有内切葡聚糖酶活性、而 N 末端不是焦谷氨酸的蛋白的 N 末端氨基酸或包含 N 末端的区域来进行修饰;

[0024] (4) (1) 到 (3) 任一项中的方法,其中具有内切葡聚糖酶活性、而 N 末端不是焦谷氨

酸的蛋白是属于家族 45 的纤维素酶；

[0025] (5) 一种具有内切葡聚糖酶活性的修饰蛋白，其 N 末端氨基酸通过氨基酸修饰转变为焦谷氨酸；

[0026] (6) (5) 中的修饰蛋白，其是通过 (1) 到 (4) 任一项中的方法获得的；

[0027] (7) 一种从以下组中选择的蛋白：

[0028] (a) 包含 SEQ ID NO :2、4、38 或 40 所示的氨基酸序列的蛋白；

[0029] (b) 包含在 SEQ ID NO :2、4、38 或 40 所示的氨基酸序列中缺失、替换、插入或添加一个或多个氨基酸所得的氨基酸序列，而且其内切葡聚糖酶活性在存在表面活性剂时降低很小的修饰蛋白；及

[0030] (c) 同源蛋白，其氨基酸序列与包含 SEQ ID NO :2、4、38 或 40 所示的氨基酸序列的蛋白的同源性至少为 85%，并且，其内切葡聚糖酶活性在存在表面活性剂时降低很小；

[0031] (8) 编码 (5) 到 (7) 任一项中的蛋白的多核苷酸；

[0032] (9) 从以下组中选择的多核苷酸：

[0033] (a) 包含 SEQ ID NO :1、3、37 或 39 所示的核酸序列的多核苷酸；

[0034] (b) 包含在 SEQ ID NO :1、3、37 或 39 所示的核酸序列中缺失、替换、插入或添加一个或多个碱基所得的碱基序列，而且其所编码蛋白的内切葡聚糖酶活性在存在表面活性剂时降低很小的多核苷酸；及

[0035] (c) 与 SEQ ID NO :1、3、37 或 39 所示的核酸序列在严紧条件下杂交，而且编码内切葡聚糖酶活性在存在表面活性剂时降低很小的蛋白的多核苷酸；

[0036] (10) 含 (8) 或 (9) 所述的多核苷酸的表达载体；

[0037] (11) 用 (10) 所述的表达载体转化的宿主细胞；

[0038] (12) (11) 所述的宿主细胞，其中所述宿主细胞是酵母或丝状真菌；

[0039] (13) (12) 所述的宿主细胞，其中，所述丝状真菌是属于腐质霉属 (*Humicola*) 或木霉属的微生物；

[0040] (14) (13) 中的宿主细胞，其中，所述丝状真菌是 *Humicola insolens* 或绿色木霉 (*Trichoderma viride*)；

[0041] (15) 一种生产 (5) 到 (7) 中任何一种蛋白的方法，包括：培养 (11) 到 (14) 任一项中的宿主细胞，从培养的宿主细胞或培养物中回收蛋白；及

[0042] (16) 一种由 (15) 所述的方法产生的蛋白。

[0043] 发明效果

[0044] 根据本发明可有效产生用于衣物洗涤，而且其内切葡聚糖酶活性在存在表面活性剂时降低很小的新型纤维素酶。

[0045] 发明的优选实施方案

[0046] 属于家族 45 并具有内切葡聚糖酶活性的蛋白（此后有时指“属于家族 45 的纤维素酶”）

[0047] 家族 45

[0048] 此处所用术语“家族 45”是指根据 Henrissat B. 等人. 关于碳水化合物活性酶的疏水簇分析而划分为家族 45 的蛋白 [Henrissat B., Bairoch A. Updating the sequence-based classification of glycosyl hydrolases. *Biochem. J.* 316 :

695-696(1996)]。

[0049] 具有内切葡聚糖酶活性的蛋白

[0050] 此处所用术语“具有内切葡聚糖酶活性的蛋白”是指具有内切葡聚糖酶活性的酶，例如内切-1,4- β -葡聚糖酶 (EC3.2.1.4)，它可水解 β -1,4-葡聚糖中的 β -1,4-吡喃型葡萄糖基键。

[0051] 表面活性剂

[0052] 此处所用术语“表面活性剂”是一种含在衣物洗涤剂中的洗涤剂成分，可广泛地分为阴离子、阳离子和非离子表面活性剂。通常使用的是阴离子表面活性剂。本发明中所用的优选阴离子表面活性剂可以是，例如线性烷基苯磺酸盐（以下有时以 LAS 表示）。

[0053] 内切葡聚糖酶活性

[0054] 此处所用术语“内切葡聚糖酶（以下以“EGU”表示）活性”定义为通过以下流程测定的羧甲基纤维素溶液粘度的降低而得到的酶活性。

[0055] 更具体地，作为底物溶液，将羧甲基纤维素 (Hercules) 溶解在 0.1mol/L 的 Tris-HCl 缓冲液 (pH10.0) 中，使终浓度为 3.5%。预先将底物溶液 (5mL) 在 40℃ 加热 10 分钟，然后加入 0.15mL 酶溶液，将整个溶液混匀，在 40℃ 反应 30 分钟。通过 R 型粘度计 (RE100 ; TOKI SANGYO CO., LTD.) 在 40℃ 测定混合液的粘度。“1 单位”酶活性定义为在每个反应条件下，将最初粘度降低到 1/2 所用的酶量。使用的阴离子表面活性剂是线性烷基苯磺酸盐 (Wako Pure Chemical Industries, Co., Ltd.)，将其加到羧甲基纤维素溶液中，至终浓度为 800ppm。

[0056] 存在表面活性剂时内切葡聚糖酶活性降低的抑制

[0057] 此处所用术语“存在表面活性剂时其内切葡聚糖酶活性降低较小”是指将其 N 末端根据本发明进行修饰的蛋白 (N 末端修饰型蛋白) 与修饰前的原始蛋白 (以下简称为原始蛋白) 在存在表面活性剂时，对其内切葡聚糖酶活性进行比较，N 末端修饰型蛋白的内切葡聚糖酶活性比原始蛋白高。

[0058] 原始蛋白

[0059] 本发明方法中所用的原始蛋白并无特别限定，只要 N 末端是除焦谷氨酸外的氨基酸，并具有内切葡聚糖酶活性的蛋白即可。原始蛋白优选地可以包括，例如纤维素酶（例如，内切葡聚糖酶、纤维二糖水解酶或 β -葡萄糖苷酶），及属于家族 45 的纤维素酶。在本文中，只要原始蛋白至少具有内切葡聚糖酶活性即可，因此可以是只具有内切葡聚糖酶活性的蛋白，或具有除内切葡聚糖酶活性外其它一种或多种酶活性的蛋白。另外，原始蛋白可以是天然蛋白或经遗传修饰的蛋白。

[0060] 属于家族 45 的纤维素酶的来源

[0061] 属于家族 45 的纤维素酶可通过基因工程技术，例如重组 DNA 技术或多肽合成技术产生，或从分离的野生型菌株获得。另外，纤维素酶包括属于家族 45 的野生型纤维素酶的突变体。

[0062] 属于家族 45 的纤维素酶可从微生物，例如丝状真菌或接合菌获得。丝状真菌包括，例如属于腐质霉属（例如 *Humicola insolens*）、木霉属（例如绿色木霉）、圆孢霉属（例如 *Staphylotrichum coccosporum*）或 *Myriococcum*（例如 *Myriococcum thermophilum*）的微生物。更具体地，来源于腐质霉属的纤维素酶包括，例如 CBH I、EG V、NCE2、NCE4 或

NCE5 等;来源于木霉属的纤维素酶包括,例如 CBH I、CBH II、EG II 和 EG III 等;来源于 *Staphylotrichum* 属的纤维素酶包括,例如 STCE1 和 STCE3 等;及来源于 *Myriococcum* 属的纤维素酶包括,例如 MTE1 等。接合菌包括,例如属于根霉属 (*Rhizopus*) (例如米曲霉 (*Rhizopus oryzae*))、毛霉属 (*Mucor*) (例如卷枝毛霉 (*Mucor circinelloides*)) 或须霉属 (*Phycomyces*) (例如闪光须霉 (*Phycomyces nitens*)) 的微生物。更具体地,包括来源于米曲霉的 RCE I、RCE II 或 RCE III;来源于卷枝毛霉的 MCEI 或 MCE II;或来源于闪光须霉的 PCE I (W000/24879)。

[0063] 焦谷氨酸或含焦谷氨酸的肽

[0064] 此处所用术语“焦谷氨酸”是指通过成熟蛋白 N 末端谷氨酰胺或谷氨酸的环化而产生的焦谷氨酸。焦谷氨酸具有 N 末端氨基不暴露在外性质。可在体内或体外进行焦谷氨酸的环化。体内,通过将编码经基因工程设计而使成熟蛋白的 N 末端为谷氨酰胺或谷氨酸的修饰蛋白的多核苷酸在宿主中进行表达,从而得到焦谷氨酸环化的蛋白。体外,将 N 末端具有谷氨酰胺或谷氨酸的蛋白用酸性溶液,例如甲酸进行处理,从而得到焦谷氨酸环化的蛋白。

[0065] 此处所用术语“肽”是指含一个或多个氨基酸,而且氨基酸通过肽键聚合的化合物。因此,此处所用术语“含焦谷氨酸的肽”是指其 N 末端为焦谷氨酸的肽。含焦谷氨酸的肽包括两个或多个交联的氨基酸,例如 2 到 40 个氨基酸,优选地为 2 到 30 个氨基酸,更优选地为 2 到 20 个氨基酸,更优选地为 2 到 10 个氨基酸,更优选地为 2 到 5 个氨基酸,最优选地为 2 到 4 个氨基酸交联。氨基酸并无特别限定,只要他们可被本领域技术人员用于所述目的即可。

[0066] 在本发明的方法中,对将原始蛋白修饰为 N 末端具有焦谷氨酸的蛋白的方法不进行限定,只要可进行蛋白修饰即可。这些方法包括,例如基因工程技术或化学技术。

[0067] 根据基因工程技术,原始蛋白可通过,例如基因工程水平上的添加和/或替代合适的氨基酸或氨基酸序列而进行修饰。

[0068] 在一个通过基因工程添加的实施例中,通过向原始蛋白(例如,具有内切葡聚糖酶活性,且 N 末端不是焦谷氨酸的蛋白)的 N 末端加入焦谷氨酸或 N 末端含焦谷氨酸的肽而进行蛋白修饰。

[0069] 例如,更具体地,通过基因工程技术进行添加的实施例可包括以下步骤:

[0070] (1) 向编码原始蛋白的多核苷酸的 5' 末端添加编码可转变为焦谷氨酸的氨基酸(谷氨酰胺或谷氨酸)的多核苷酸,或编码 N 末端具有可转变为焦谷氨酸的氨基酸的多肽的多核苷酸;及

[0071] (2) 在可使 N 末端氨基酸进行焦谷氨酸环化的宿主中表达 (1) 中所得的多核苷酸。

[0072] 在通过基因工程替代的实施例中,通过将原始蛋白的 N 末端氨基酸或 N 末端区替代为焦谷氨酸或 N 末端含焦谷氨酸的肽而进行蛋白修饰。

[0073] 例如,更具体地,通过基因工程替代的实施例可包括以下步骤:

[0074] (1) 将编码原始蛋白的 5' 末端或含 5' 末端区的多核苷酸用编码可转变为焦谷氨酸的氨基酸(谷氨酰胺或谷氨酸)的多核苷酸,或编码 N 末端具有可转变为焦谷氨酸的氨基酸的肽的多核苷酸替代;及

[0075] (2) 在可使N末端氨基酸进行焦谷氨酸环化的宿主中表达(1)中所得的多核苷酸。

[0076] 使用化学技术的实施例可包括,例如,

[0077] (a) 将焦谷氨酸(或N末端含焦谷氨酸的肽)直接加到原始蛋白的N末端的实施例;

[0078] (b) 将可转变为焦谷氨酸的氨基酸(或N末端具有可转变为焦谷氨酸的氨基酸的肽)通过化学方法加到原始蛋白N末端,并通过化学方法进行N末端氨基酸焦谷氨酸化的实施例;或

[0079] (c) 将N末端具有可转变为焦谷氨酸的氨基酸的原始蛋白的N末端氨基酸通过化学方法焦谷氨酸化的实施例。

[0080] 在属于家族45的纤维素酶的N末端加入焦谷氨酸或合焦谷氨酸的肽的方法

[0081] 修饰方法可通过用属于家族45的纤维素酶在实施例中进行进一步阐述。在属于家族45的纤维素酶的N末端加入焦谷氨酸或含焦谷氨酸的肽的方法可通过基因工程技术进行。在通常所用的纤维素酶的生产中,编码所需纤维素酶的多核苷酸的编码区可连接到在丝状真菌等宿主中有功能的启动子和终止子之间,然后将得到的表达盒转入宿主中。另外,编码在宿主细胞中具有分泌功能的信号肽序列的多核苷酸也可以加到表达盒中。当将表达盒引入到宿主细胞中时,所需的纤维素酶可分泌到培养基中,从而可很容易地收集所需的纤维素酶。在这种情况下,可通过向紧邻分泌信号肽序列的下游加入编码氨基酸的多核苷酸而将所需氨基酸加到纤维素酶的N末端。另外,也可通过用宿主中的分泌信号序列进行N末端氨基酸的氨基基团的修饰。例如,在来源于绿色木霉的cbh1或cbh2,或来源于Humicola insolens的NCE2或NCE5中,N末端是焦谷氨酸化的,这种修饰可通过用来源于这种微生物的分泌信号序列并在绿色木霉或Humicola insolens中表达来制备。根据优选实施例,如上所述制备的属于家族45的修饰纤维素酶具有在存在表面活性剂时其内切葡聚糖酶活性降低很小的优势性质。

[0082] 根据另一个实施例,整个序列可根据本领域技术人员常用的技术进行化学合成。在这种情况下,可用部分天然蛋白进行合成。

[0083] 本发明中的蛋白

[0084] 本发明中的蛋白通过将具有内切葡聚糖酶活性的原始蛋白的N末端修饰为焦谷氨酸而进行制备(例如,通过获得属于家族45的纤维素酶,并向该成熟蛋白的N末端氨基酸中链加入焦谷氨酸(pQ)或含焦谷氨酸的肽而制备),而且其内切葡聚糖酶活性在存在表面活性剂时降低很小。另外,本发明包括通过上述方法制备的,其内切葡聚糖酶活性在存在表面活性剂时降低很小的蛋白。

[0085] 更具体地,本发明中的蛋白包括从由以下蛋白组成的组中选择的蛋白:

[0086] (a) 包含SEQ ID NO:2、4、38或40所示的氨基酸序列的蛋白;

[0087] (b) 含有在SEQ ID NO:2、4、38或40所示的氨基酸序列中缺失、替换、插入或添加一个或多个氨基酸所得的氨基酸序列,而且其内切葡聚糖酶活性在存在表面活性剂时降低很小的修饰蛋白;及

[0088] (c) 同源蛋白,其氨基酸序列与SEQ ID NO:2、4、38或40所示的氨基酸序列的蛋白的同源性至少为85%,而且其内切葡聚糖酶活性在存在表面活性剂时降低很小。

[0089] SEQ ID NO:2的氨基酸序列是N末端添加型NCE4的氨基酸序列(参见实施例A2),

其中含有 5 个氨基酸 (N 末端为焦谷氨酸) 的肽加到来源于 *Humicola insolens* MN200-1 的内切葡聚糖酶 NCE4 的 N 末端。

[0090] SEQ ID NO:4 的氨基酸序列是 N 末端替代型 STCE1 的氨基酸序列 (参见实施例 B2), 其中含有 4 个氨基酸 (N 末端为焦谷氨酸) 的肽替代来源于 *Staphylotrichum coccosporum* IF0 31817 的内切葡聚糖酶 STCE1 的 N 末端氨基酸 (Ala)。

[0091] SEQ ID NO:38 的氨基酸序列是 pQ 添加型 STCE1 的氨基酸序列 (参见实施例 B3), 其中焦谷氨酸加到来源于 *Staphylotrichum coccosporum* IF0 31817 的内切葡聚糖酶 STCE1 的 N 末端。

[0092] SEQ ID NO:40 的氨基酸序列是 N 末端添加型 STCE1 的氨基酸序列 (参见实施例 B4), 其中含有 4 个氨基酸 (N 末端为焦谷氨酸) 的肽加到来源于 *Staphylotrichum coccosporum* IF0 31817 的内切葡聚糖酶 STCE1 的 N 末端。

[0093] 此处所用术语“修饰蛋白”是指包含在 SEQ ID NO:2、4、38 或 40 所示的氨基酸序列中缺失、替代、插入或添加一个或多个氨基酸 (例如, 一个或几个氨基酸) 所得的氨基酸序列, 而且其内切葡聚糖酶活性在存在表面活性剂时降低很小的蛋白。修饰例如“缺失、替代、插入或添加”的氨基酸的数目优选地是 1 到 30, 更优选地是 1 到 10, 最优选地是 1 到 6。

[0094] 另外, 修饰蛋白是指包含在 SEQ ID NO:2、4、38 或 40 所示的氨基酸序列中保守替代一个或多个氨基酸, 而且其内切葡聚糖酶活性在存在表面活性剂时降低很小的蛋白。此处所用术语“保守替代”是指蛋白中一个或多个氨基酸残基可被具有相同化学性质的不同氨基酸替代, 从而使其蛋白活性不发生实质性改变。保守替代可包括, 例如将疏水残基替代为另一个疏水残基, 或将极性残基替代为另一个具有相同电荷的极性残基。具有相同化学性质并可互相保守替代的氨基酸是本领域技术人员已知的。更具体地, 非极性 (疏水) 氨基酸包括, 例如丙氨酸、缬氨酸、异亮氨酸、亮氨酸、脯氨酸、色氨酸、苯丙氨酸或蛋氨酸。极性 (中性) 氨基酸包括, 例如甘氨酸、丝氨酸、苏氨酸、酪氨酸、谷氨酰胺、天冬酰胺或半胱氨酸。带正电荷的碱性氨基酸包括, 例如精氨酸、组氨酸或赖氨酸。带负电荷的酸性氨基酸包括, 例如天冬氨酸或谷氨酸。

[0095] 此处所用术语“同源蛋白”是指其与包含 SEQ ID NO:2、4、38 或 40 所示的氨基酸序列的同源性 (序列相似性) 至少为 85% (优选地是 90% 或以上, 最优选地是 95% 或以上), 且其内切葡聚糖酶活性在存在表面活性剂时降低很小的蛋白。此处所用同源性以用已知的同源性分析软件 FASTA3, 根据缺省参数计算出的值 (相似性) 表示 [Science, 227, 1438-1441 (1985); Proc. Natl. Acad. Sci. USA, 85, 2444-2448 (1988); <http://www.ddbj.nig.ac.jp/E-mail/homology-j.html>]。

[0096] 如上所述, 在本发明的“由氨基酸序列 SEQ ID NO:2 所示的蛋白”中, 除了加到 N 末端的由 5 个氨基酸组成的肽外, 其它部分来源于内切葡聚糖酶 NCE4。另外, 在本发明的“氨基酸序列 SEQ ID NO:4、38 或 40 所示的蛋白”中, 除了 N 末端氨基酸或肽外, 其它部分来源于内切葡聚糖酶 STCE1。内切葡聚糖酶 NCE4 和 STCE1 属于家族 45。已知属于家族 45 的内切葡聚糖酶包括, 例如来源于腐质霉属的 NCE5 (W001/90375)。

[0097] 在图 1 和 2 中, 所示的是内切葡聚糖酶 STCE1 的氨基酸序列 [信号肽 (SEQ ID NO:43), 成熟蛋白 (SEQ ID NO:44)]、内切葡聚糖酶 NCE4 的氨基酸序列 [信号肽 (SEQ ID NO:45), 成熟蛋白 (SEQ ID NO:46)] 和内切葡聚糖酶 NCE5 的氨基酸序列 [信号肽 (SEQ ID NO:

47),成熟蛋白 (SEQ ID NO :48)] 的比对结果。

[0098] 图 1 所示的是前半部分 (即 N 末端部分) 的比对结果,图 2 所示的是后半部分 (即 C 末端部分) 的比对结果。图 1 和 2 中的符号 “*” 表示与 STCE1 一致的氨基酸。

[0099] 如图 1 和 2 所示,属于家族 45 的内切葡聚糖酶包括作为共有结构域的催化结构域 (1 到 207 位),有些含有接头 (208 到 258 位) 和 / 或纤维素结合结构域 (CBD) (259 到 295 位)。在本文中,以上结构域之后的括号中的数字表示内切葡聚糖酶 STCE1 的氨基酸序列 (SEQ ID NO :44) 中的氨基酸编号。

[0100] 在内切葡聚糖酶的催化结构域和纤维素结合结构域均存在保守氨基酸,但在接头没有明显的保守区存在。含多个保守氨基酸或在该区域所含的共有氨基酸的区域 (例如,催化结构域或纤维素结合结构域,特别是催化结构域) 是对内切葡聚糖酶 (例如 STCE1 或 NCE4) 酶活性重要的区域或氨基酸。因此,在除了这种重要区域或氨基酸的其它区域或氨基酸处进行氨基酸修饰 (例如缺失、替代、插入和 / 或添加,特别是保守替代),可高效率获得该酶活性得到维持的修饰蛋白质或同源蛋白质,而无需过度试验。。

[0101] 另外,即使在内切葡聚糖酶含多个保守氨基酸的区域中,用不同 的氨基酸对非共有氨基酸进行修饰 (优选地是相同的氨基酸进行保守替代) 也可以保持酶活性。因此,通过这种修饰,可高效率获得该酶活性得到维持的修饰蛋白质或同源蛋白质,而无需过度试验。

[0102] 即使是含有大量保守氨基酸的区域内的共有氨基酸,改变为其它氨基酸时,有时也可维持其酶活性,尤其是改变为可保守取代类似氨基酸的时候,其可能性增高。本发明的修饰蛋白质或同源蛋白质,只有是具有内切葡聚糖酶活性,可以是任意区域,例如催化区域、接头或纤维素结合区域所含的氨基酸改变的蛋白质。

[0103] 编码本发明中蛋白的多核苷酸

[0104] 根据本发明,提供编码包含 SEQ ID NO :2、4、38 或 40 所示的氨基酸序列的蛋白,或其修饰或同源蛋白 (以下统称为本发明中的蛋白) 的多核苷酸。当给定一个蛋白的氨基酸序列时,则可容易地选择编码该氨基酸序列的核苷酸序列,因此可选择多种编码本发明中蛋白的核苷酸序列。此处所用术语“多核苷酸”包括 DNA 和 RNA, DNA 是优选的。

[0105] 本发明中的多核苷酸包括从由以下多核苷酸组成的组中选择的多核苷酸:

[0106] (a) 包含核苷酸序列 SEQ ID NO :1、3、37 或 39 的多核苷酸;

[0107] (b) 包含在核苷酸序列 SEQ ID NO :1、3、37 或 39 中缺失、替代、插入或添加一个或多个碱基所得的碱基序列,而且其所编码蛋白的内切葡聚糖酶活性在存在表面活性剂时降低很小的多核苷酸;及

[0108] (c) 在严紧条件下与核苷酸序列 SEQ ID NO :1、3、37 或 39 所示的多核苷酸杂交,而且其所编码蛋白的内切葡聚糖酶活性在存在表面活性剂时降低很小的多核苷酸。

[0109] 在上述 (b) 中的核苷酸序列中,被缺失、替代、插入或添加的核苷酸的数目可以是例如 1 到 90,优选地是 1 到 30,更优选地是 1 到 18,最优选地是 1 到 9。

[0110] 此处所用术语“在严紧条件下”是指以下条件。根据 ECL 直接 DNA/RNA 标记和检测系统 (Amersham) 的说明书所述,在 42℃ 预杂交 1 小时后,加入核苷酸序列 SEQ ID NO :1、3、37 或 39 所示的多核苷酸,并在 42℃ 杂交 14 到 16 小时。杂交后,用含 0.4% SDS 和 6mol/L 尿素的 0.5×SSC (1×SSC ;15mmol/L 柠檬酸钠,150mmol/L 氯化钠) 在 42℃ 洗涤 20 分钟,并重复 2 次,然后用 5×SSC 在室温下洗涤 5 分钟,并重复 2 次。

[0111] 本发明中的多核苷酸包括天然多核苷酸。也可以是全合成的多核苷酸。而且可用部分天然多核苷酸进行合成。典型地,本发明中的多核苷酸可从来源于微生物的基因组文库,通过基因工程的普通方法,例如用根据部分氨基酸序列信息设计的合适的 DNA 探针,进行筛选而获得。

[0112] 本发明中的蛋白的产生

[0113] 本发明中的蛋白可通过用含编码本发明蛋白的多核苷酸片段的多核苷酸分子(优选以表达载体形式)转化宿主细胞,从而使多核苷酸分子在宿主细胞中复制,并进行基因表达而在宿主细胞中产生。

[0114] 根据本发明,提供了含编码本发明中的蛋白的多核苷酸片段,从而使多核苷酸片段可在宿主微生物中复制并表达蛋白的表达载体。

[0115] 本发明中的表达载体可基于存在于染色体外并能独立于染色体复制而进行复制的自主复制载体(例如质粒)进行构建。选择性地,本发明中的表达载体可以用载体转化宿主时将其整合到宿主微生物基因组中,与染色体一起复制的载体。本发明中载体的构建可通过基因工程中通用流程或方法进行。

[0116] 为了用本发明中的表达载体转化宿主微生物而表达具有所需活性的蛋白。优选地是表达载体除了本发明中的多核苷酸片段外还包括,例如可调控表达的多核苷酸,或筛选转化子的基因标记。

[0117] 在本发明中可使用多种调控转录或翻译的信号作为可调控表达的多核苷酸,例如启动子、终止子或编码分泌信号肽的多核苷酸。可通过普通方法将调控的多核苷酸与其插入片段进行连接。

[0118] 本发明中所用的启动子并无特别限定,只要其在宿主微生物中具有转录活性即可。启动子可以是调控来源于相同或不同宿主微生物的编码蛋白的基因表达的多核苷酸。例如,可在大肠杆菌中使用乳糖操纵子或色氨酸操纵子的启动子;在酵母中使用乙醇脱氢酶基因、酸性磷酸酶基因、半乳糖利用基因或甘油醛-3-磷酸脱氢酶基因的启动子;在真菌例如丝状真菌中使用 α -淀粉酶基因、葡糖淀粉酶基因、纤维二糖水解酶基因或甘油醛-3-磷酸脱氢酶基因的启动子。

[0119] 信号肽并无特别限定,只要其可在宿主微生物中提供蛋白分泌作用即可。信号肽多核苷酸可从与编码蛋白的基因相同或不同来源的宿主微生物中获得。

[0120] 基因标记可根据筛选转化子的方法进行合适的选择。例如,在本发明中可使用抗药基因或与营养缺陷型突变互补的基因作为基因标记。例如,当宿主是细菌时,可使用氨苄青霉素抗性基因、卡那霉素抗性基因或四环素抗性基因。当宿主是酵母时,可使用色氨酸合成基因(trpI, trpC)、尿嘧啶合成基因(ura3)、亮氨酸合成基因(leu2)或组氨酸合成基因(his3)。当宿主是真菌例如丝状真菌时,可使用越霉素抗性基因、硝酸盐利用基因(niaD)、精氨酸合成基因(argB)、尿嘧啶合成基因(pyr4)、潮霉素抗性基因、除草剂抗性基因、博来霉素抗性基因或 aureobasidin 抗性基因。

[0121] 根据本发明,提供用表达载体转化的微生物。本发明中所用的宿主-载体系统并无特别限定。例如,可使用大肠杆菌、放线菌、酵母或丝状真菌系统,或利用这种微生物进行融合蛋白表达的系统。当用酵母作为宿主时,包括,例如,使用酿酒酵母属(Saccharomyces)、汉逊酵母属(Hansenula)、或毕赤酵母属(Pichia)的微生物,优选地是

酿酒酵母。当使用丝状真菌作为宿主时,可以使用任何丝状真菌,优选地是属于腐质霉属、木霉属、曲霉属 (*Aspergillus*)、*Acremonium* 或镰刀菌属 (*Fusarium*) 的微生物,更优选地是属于腐质霉或木霉属的微生物。更优选地是 *Humicola insolens*、*Humicola thermoidea*、绿色木霉、里氏木霉 (*Trichoderma reesei*)、长梗木霉 (*Trichoderma longibrachiatum*)、*Staphylotrichum coccosporum*、黑曲霉 (*Aspergillus niger*)、米曲霉 (*Aspergillus oryzae*)、尖孢镰孢 (*Fusarium oxysporum*) 或 *Acremonium cellulolyticus*, 优选地是 *Humicola insolens* 或绿色木霉。

[0122] 可通过普通方法,例如钙离子法、锂离子法、电穿孔法、PEG 法、土壤杆菌法或基因枪法用表达载体转化微生物,根据待转化宿主选择合适的方法。

[0123] 在本发明中,将本发明中的转化子进行培养,得到的培养物用来获得本发明的目的蛋白。本发明中的转化子可通过普通方法选择合适的培养基或培养条件进行培养。

[0124] 可向培养基中补充例如可被本发明中的转化子代谢或利用的碳源或氮源、无机盐、各种维生素、各种氨基酸例如谷氨酸或天冬酰胺、微量营养物质例如核苷酸、或选择试剂例如抗生素。另外,可适当添加利于本发明中的转化子生长或促进本发明中的蛋白产生的有机和/或无机物质。另外,如果需要,可使用含其它营养物质的天然或合成培养基。

[0125] 可在培养基中使用任何碳源和氮源,只要它可被本发明中的转化子利用即可。可利用碳源可使用例如蔗糖、淀粉、甘油、葡萄糖、果糖、麦芽糖、乳糖、纤维素、纤维二糖、糊精、动物油或植物油或其水解物。一般来说,其浓度在培养基中优选地是 0.1% 到 5%。可利用氮源可使用例如动植物组分或其提取物、例如蛋白胨、肉提取物、玉米浸出液或脱脂豆粉、有机酸铵盐例如琥珀酸铵盐或酒石酸铵盐、尿素或其它多种含氮化合物例如无机酸或有机酸。无机盐可使用例如那些可产生钠、钾、钙、镁、钴、磷酸、硫酸或其它离子的无机盐。

[0126] 可适当使用任何含其它组分,例如细胞、分泌液或微生物例如酵母提取物、或细小植物粉的培养基,只要这种培养基不干预本发明中的转化子生长和本发明中的蛋白的产生和积累即可。当培养具有营养需求的突变株时,通常向培养基中加入满足其营养需求的物质。在本文中,如果培养基中含天然底物,则不需加入这种营养物质。

[0127] 根据所用的转化子和外部条件,对培养条件例如培养基成分、培养基流动性、培养温度、搅拌速率或通气速率进行合适的选择和控制在,从而得到优选的结果。如果在液体培养基中产生泡沫,则可使用消泡剂例如硅树脂油、植物油、矿物油或表面活性剂。

[0128] 在获得的培养物中积累的本发明中的蛋白包含于本发明中的转化子和培养物滤液中。因此,可通过离心分离培养物组分而从培养物滤液和转化子细胞中获得本发明中的蛋白。

[0129] 本发明中的蛋白可通过普通方法从培养物滤液中回收。从培养物中回收本发明中的蛋白的流程可按单独或按任意顺序组合,或重复进行。例如,可使用萃取过滤、离心、透析、浓缩、干燥、冷冻、吸附、解吸附、基于在各种溶剂中溶解度不同的分离方法(例如沉淀、盐析、结晶、重结晶、反向溶解或层析)。

[0130] 另外,本发明中的蛋白可从本发明中的转化子的培养物中获得。例如,通过普通方法从培养物中提取(例如,研磨处理或压力破碎)、回收(例如,过滤或离心)、或纯化(例如,盐析或溶剂沉淀)。

[0131] 可根据普通方法对得到的粗物质进行纯化,例如用载体物质例如葡聚糖、纤维素、

琼脂糖、合成多聚物或硅胶进行柱层析。本发明中的目的蛋白的纯品可从本发明中的转化子培养物中通过上述方法单独或以合适组合而获得。

[0132] 如上所述,根据本发明的另一个实施例,提供本发明中的新型蛋白的生产过程。可以与制备转化子的原始宿主实质上相同的方式进行转化子的培养(包括培养条件)。转化子培养后的目的蛋白的回收方法可通过常用的方法进行。

[0133] 保藏菌株

[0134] 用本发明的表达载体 pEGD01 转化的大肠杆菌 JM109 (*Escherichia coli*/pEGD01) (FERM BP-5973) 在 1996 年 7 月 12 日国内保藏于独立行政法人产业技术综合研究所特许生物寄托中心[(以前名称) 通商产业省工业技术院生命工学工业技术研究所(地址:〒305-8566 日本国茨城县筑波东 1 丁目 1 番地 1 中央第 6)], 并于 1997 年 6 月 13 日转为国际保藏。国际保藏号(国际保藏号后的括号 [] 中的数字是国内保藏号) 是 FERM BP-5973 [FERM P-15729]。

[0135] 用本发明的表达载体 pMKD01 转化的大肠杆菌 JM109 (*Escherichia coli*/pMKD01) (FERM BP-5974) 在 1996 年 7 月 12 日在国内保藏于独立行政法人产业技术综合研究所特许生物寄托中心[(以前名称) 通商产业省工业技术院生命工学工业技术研究所(地址: 〒305-8566 日本国茨城县筑波东 1 丁目 1 番地 1 中央第 6)], 并于 1997 年 6 月 13 日转为国际保藏。国际保藏号(国际保藏号后的括号 [] 中的数字是国内保藏号) 是 FERM BP-5974 [FERM P-15730]。

[0136] 用本发明的表达载体 pCB1-M2XR 转化的大肠杆菌 JM109 (*Escherichia coli*/pCB1-M2XR) (FERM BP-6045) 在 1996 年 9 月 9 日在国内保藏于独立行政法人产业技术综合研究所特许生物寄托中心[(以前名称) 通商产业省工业技术院生命工学工业技术研究所(地址: 〒305-8566 日本国茨城县筑波东 1 丁目 1 番地 1 中央第 6)], 并于 1997 年 8 月 11 日转为国际保藏。国际保藏号(国际保藏号后的括号 [] 中的数字是国内保藏号) 是 FERM BP-6045 [FERM P-15840]。

[0137] 在本文中,本发明中的质粒 pCB1-M2 不仅可通过实施例中描述的方法获得,而且也可以通过预先用限制性酶 XbaI 消化的质粒 pCB1-M2XR 的自连而获得。

[0138] 作为本发明中的表达载体的宿主的 *Humicola insolens* MN200-1 (FERM BP-5977) 在 1996 年 7 月 15 日在国内保藏于独立行政法人产业技术综合研究所特许生物寄托中心[(以前名称) 通商产业省工业技术院生命工学工业技术研究所(地址: 〒305-8566 日本国茨城县筑波东 1 丁目 1 番地 1 中央第 6)], 并于 1997 年 6 月 13 日转为国际保藏。国际保藏号(国际保藏号后的括号 [] 中的数字是国内保藏号) 是 FERMBP-5977 [FERM P-15736]。

[0139] 作为本发明中的表达载体的宿主的绿色木霉 MC300-1 (*Trichoderma viride* MC300-1) (FERM BP-6047) 在 1996 年 9 月 9 日在国内保藏于独立行政法人产业技术综合研究所特许生物寄托中心[(以前名称) 通商产业省工业技术院生命工学工业技术研究所(地址: 〒305-8566 日本国茨城县筑波东 1 丁目 1 番地 1 中央第 6)], 并于 1997 年 8 月 11 日转为国际保藏。国际保藏号(国际保藏号后的括号 [] 中的数字是国内保藏号) 是 FERM BP-6047 [FERM P-15842]。

[0140] 用本发明的表达载体 pUC118-STCEex 转化的大肠杆菌 JM109 (*E. coli* DH5 α /pUC118-STCEex) (FERM BP-10127) 在 2003 年 12 月 1 日在国内保藏于独立行政法人产业技

术综合研究所特许生物寄托中心（地址：〒305-8566 日本国茨城县筑波东 1 丁目 1 番地 1 中央第 6），并于 2004 年 9 月 15 日转为国际保藏。国际保藏号（国际保藏号后的括号 [] 中的数字是国内保藏号）是 FERM BP-10127 [FERM P-19602]。在本文中，质粒 pUC118-STCEex 是通过将 STCE 基因插入到质粒 pUC118 的 BamHI 位点而得到的质粒。包含在 BamHI 片段中的内切葡聚糖酶 STCE1 基因包括一个内含子，其序列为 SEQ ID NO :5。

实施例

[0141] 本发明将通过以下实施例进行进一步描述，但不作为对本发明的限制。

[0142] 实施例 A1：在 Humicola insolens 中表达 N 末端添加型纤维素酶和野生型纤维素酶的表达载体的构建

[0143] (1) 用于表达 N 末端添加型纤维素酶的表达载体 pMKD01 的构建

[0144] 用 BamHI 消化质粒 pUC118 (Takara Bio)。得到的片段用 DNA 平端化试剂盒 (Takara Bio) 进行平端化，然后用 DNA 连接试剂盒 (Takara Bio) 进行自连，从而得到 pUC118BN。将所得的 pUC118BN 用 SphI 消化，平末端化后自连得到 pUC118BSN。然后，根据特开平 No. 8-126492 中所述的方法从 Humicola insolens MN200-1 (FERM BP-5977) 获得含纤维素酶 NCE2 基因全长的 3.4kb PstI-XbaI 片段。将该片段连接到预先用相同的限制性酶消化的 pUC118BSN 中，得到 pUC118BSN-PX。用 Sculptor 体外突变系统 (Amersham Bioscience) 在 pUC118BSN-PX 引入定点突变，得到质粒 pM21。用于突变的寡核苷酸使用的是寡核苷酸 MNC-02 和 MNC-03。

[0145] MNC-02 :5' -GAGCGCCAGAACTGTGGATCCACTTGGTGAGCAAT G-3' (SEQ ID NO :6)

[0146] MNC-03 :5' -TCCGCCGTTCTGAGCGGATCCAGGCGTTTGGCGCG-3' (SEQ ID NO :7)

[0147] 用 BamHI 消化质粒 pM21，并与来源于 Humicola insolens MN200-1 的纤维二糖水解酶基因 (NCE3) 的 BamHI 消化的 PCR 片段连接，得到质粒 pKM04。用来源于 Humicola insolens MN200-1 (FERM BP-5977) 的基因组 DNA (W098/03640) 为模板，以下列引物 MKA-05 和 MKA-06 扩增 NCE3 的 PCR 片段。

[0148] MKA-05 :5' -GCCGCCAGCAGGCGGGATCCCTCACCACCGAGAG G-3' (SEQ ID NO :8)

[0149] MKA-06 :5' -TGATCGTCGAGTCAGGGATCCAGAATTTACAGGCA C-3' (SEQ ID NO :9)

[0150] 然后，用来源于构巢曲霉 (Aspergillus nidulans) 的 trpC 基因的启动子和终止子 (Mullaney, E. J., et.al., Mol. Gen. Genet., 199, 37-45, 1985) 制备可在 Humicola insolens 中表达特开昭 59-175889 号公报记载的越霉素抗性基因的基因。将得到的基因连接到用 XbaI 消化的质粒 pKM04 中，得到表达 N 末端添加型纤维素酶的表达载体 pKMD01 (FERM BP-5974)。在 pKMD01 中，在距 NCE2 基因的成熟蛋白 N 末端下游 10bp 和其终止密码子下游 3bp 处引入 BamHI 识别位点，从而在属于家族 45 的纤维素酶的 N 末端加入 5 个氨基酸残基。

[0151] 表 1

[0152]

BamHI	BamHI
5'-GAGCGCCAGAACTGTGGATCCCTC-	-TGCCTGTAAgaggatccagg-3' (SEQ ID NO: 10)
GluArgGlnAsnCysGlySerLeu-	-CysLeu (SEQ ID NO: 11)
↑ 成熟蛋白质的N末端	

[0153] (2) 表达野生型纤维素酶的表达载体 pJD01 的构建

[0154] 向以上 (1) 中获得的质粒 pM21 中用寡核苷酸 pMN-Bam 引入定点突变, 得到引入突变的质粒 pM21-m-A1。

[0155] pMN-Bam : 5' -GGTCAAACAAGTCTGTGCGGATCCTGGGACAAGA TGGCCAAGTTCTTCTTAC-3' (SEQ ID NO:12)

[0156] 用限制性酶 HindIII 和 BamHI 消化质粒 pM21-m-A1, 回收约 1kb 的片段。然后, 将实施例 A1 (1) 中获得的 pMKD01 用 HindIII 和 BamHI 消化, 回收约 7kb 的片段。将这些片段连接, 得到质粒 pJD01。在该载体中, 在距 NCE2 基因翻译起始位点上游 15bp 和其终止密码子下游 3bp 处引入 BamHI 识别位点, 从而将属于家族 45 的纤维素酶从转录起始位点进行表达。

[0157] 表 2

[0158]

<u>BamHI</u>	<u>BamHI</u>
5'- <u>tgcggatcctgggacaag</u> ATGGCC-	-CCGTTCTGA <u>gcggatccagg</u> -3'
MetAla--	--ProPhe
	(SEQ ID NO: 14)

[0159] 实施例 A2: 表达野生型 NCE4 和 N 末端添加型 NCE4 的表达载体的构建, Humicola insolens 的转化及活性评估

[0160] (1) 表达载体的构建

[0161] 将全长 NCE4 基因编码区连接到实施例 A1 (2) 中获得的 pJD01 中, 构建 pN2EX-N4 作为野生型 NCE4 的表达载体。

[0162] 首先, 用来源于 Humicola insolens MN200-1 的基因组 DNA 为模板, 以引物组合 NCE4-Ne 和 NCE4-Ce 进行 PCR 扩增, 得到 NCE4 基因的全长编码区。将得到的约 0.9kb 的 PCR 产物用 BamHI 消化, 并连接到预先用相同的限制性酶消化的 pJD01 中, 得到 pN2EX-N4。

[0163] NCE4-Ne : 5' -GGGGGGATCCTGGGACAAGATGCGTTCTCCCTC TC-3' (SEQ ID NO:15)

[0164] NCE4-Ce : 5' -GGGGGGATCCCTGCGTTTACAGGCACTGATGG-3' (SEQ ID NO:16)

[0165] 将已去除分泌信号序列的 NCE4 基因成熟蛋白的编码区连接到实施例 A1 (1) 中获得的 pMKD01 (FREM BP-5974) 中, 得到 pEGD01 (FREM BP-5973) 作为 N 末端添加型 NCE4 的表达载体。首先, 用来源于 Humicola insolens MN200-1 的基因组 DNA 为模板, 以引物组合 NCE4-Ns 和 NCE4-Cs 进行 PCR 扩增, 得到 NCE4 基因的成熟蛋白编码区。将得到的约 0.8kb 的 PCR 产物用 BamHI 消化, 并连接到预先用相同的限制性酶消化的 pMKD01 中, 得到 PEG01。

[0166] NCE4-Ns : 5' -CCGGTGTGGCCGGATCCGCTGATGGCAAG-3' (SEQ ID NO:17)

[0167] NCE4-Cs : 5' -TAAGGCCCTCAAGGATCCCTGCGTCTACAG-3' (SEQ ID NO:18)

[0168] (2) 含野生型 NCE4 和 N 末端添加型 NCE4 的转化子

[0169] 将 Humicola insolens MN200-1 在 S 培养基 (3.0% 葡萄糖, 2.0% 酵母提取物, 0.1% 聚脲, 0.03% 氯化钙和 0.03% 硫酸镁 pH6.8) 中 37°C 培养 24 小时。通过离心收集得到的菌丝体, 用 0.5mol/L 蔗糖洗涤, 并悬浮在含 3mg/mL β -葡萄糖醛酸苷酶 (Sigma)、1mg/mL 几丁质酶和 1mg/mL 消解酶 20T (Seikagaku Corp.) 的 0.5mol/L 蔗糖溶液中。悬浮液在 30°C 轻轻振荡 60 到 90 分钟以得到原生质体。通过过滤和离心收集得到的原生质体, 用 SUTC 缓冲液 (0.5mol/L 蔗糖, 10mmol/L 氯化钙和 10mmol/L Tris-HCl, pH7.5) 洗涤, 并重悬到 SUTC 中而得到原生质体悬浮液。

[0170] 向 100 μ L 原生质体悬浮液中加入 10 μ L DNA 溶液 (pEGD01 或 pN2EX-N4)。每个在冰上静置 5 分钟, 然后加入 400 μ L PEG 溶液 (60% 聚乙二醇 4000, 10mmol/L 氯化钙和 10mmol/L Tris-HCl, pH7.5), 整个溶液在冰上静置 20 分钟。将得到的原生质体悬浮液用 SUTC 洗涤后, 用含 200 μ g/mL 潮霉素 B 的 YMG 培养基 (1% 葡萄糖, 0.4% 酵母提取物, 0.2% 麦芽提取物及 1% 琼脂; pH6.8) 与 YMG 软琼脂一起覆盖在原生质体悬浮液上。平板在 37°C 培养 5 天, 以获得转化子克隆。

[0171] (3) 培养基上清中野生型 NCE4 和 N 末端添加型 NCE4 的 EGU 活性比较

[0172] 根据 W098/03667 中所述方法培养得到的转化子, 得到培养上清。将培养上清进行 SDS-PAGE, 选择可清晰检测到约 43kDa 的 NCE4 带的培养上清。测定培养上清在 pH10, 存在或不存在 LAS 时的 EGU 活性, 并比较活性的降低程度。结果如表 3 所示。在表 3 中, 以 pH10 时的 EGU 活性为 100, 用相应的百分比来表示 EGU 活性。

[0173] 表 3

[0174]

	表达载体	EGU (%)	
		无 LAS	有 LAS
野生型 NCE4	pN2EX-N4	100	27.1
N 末端添加型 NCE4	pEGD01	100	55.6

[0175] 结果表明, 从 N 末端添加型 NCE4 转化子获得的培养上清在存在 LAS 时活性降低很小。与存在 LAS 时的 EGU 活性相比较发现, N 末端添加型 NCE4 的活性是野生型 NCE4 的 2.1 倍。另外, *Humicola insolens* MN200-1 的 EGU 活性约占转化子 4%, 从而表明几乎所有 EGU 活性均来源于整合表达载体表达的重组酶。

[0176] (4) 野生型 NCE4 和 N 末端添加型 NCE4N 末端氨基酸序列分析

[0177] 将上述转化子的培养上清进行 SDS-PAGE, 并将电泳的蛋白电转移到 PVDF 膜 (Immobilon-PSQ; Millipore) 上。从 blot 上剪切下约 43kDa 的 NCE4 带, 并用蛋白测序仪 Model 492 (Applied Biosystems) 分析氨基酸序列。

[0178] 结果表明, 从 pN2EX-N4 转化子获得的野生型 NCE4 序列与 W098/03667 所述的来源于 *Humicola insolens* MN200-1 的 NCE4 的 DNA 序列一致。

[0179] 但从 pEGD01 转化子获得的 N 末端添加型 NCE4 氨基酸序列不能用上述方法测定。用 pfu 焦谷氨酸氨肽酶 (Takara Bio) 除去 N 末端添加型 NCE4 修饰的 N 末端, 然后测定其氨基酸序列 (SEQ NO:2)。

[0180] 结果总结于表 4 中, 比较野生型 NCE4 和 N 末端添加型 NCE4 的 N 末端氨基酸序列表明, 每个 NCE4 的 N 末端氨基酸序列与从构建的相应表达载体预测的序列一致。

[0181] 表 4

[0182]

	N 末端氨基酸	N 末端氨基酸序列
野生型 NCE4	丙氨酸 (未修饰)	ADGKSTR (SEQ ID NO: 19)
N 末端添加型 NCE4	焦谷氨酸 (修饰)	pQNCGSADGKSTR (SEQ ID NO: 20)

[0183] 实施例 B1: 在绿色木霉中表达 N 末端替代型纤维素酶和野生型纤维素酶的表达载体的构建

[0184] (1) 表达载体的构建

[0185] 将来源于绿色木霉 MC300-1 (根据 W098/11239 获得) 含全长 cbh1 基因的 7kb PstI 片段、含 cbh1 启动子区的 3.1kb HindIII 片段和含 cbh1 终止子区的 2.7kb SalI 片段克隆到质粒 pUC119 中, 分别构建质粒 pCB1-H4 和质粒 pCB1-S1。用辅助噬菌体 M13K07 转染含每种质粒的大肠杆菌 JM109, 从而获得单链 DNA。以质粒 pCB1-H4 和 pCB1-S1 中得到的单链 DNAs 为模板, 将磷酸化的寡核苷酸 CBn-Stu 和 Cbc-Xho 退火, 从而通过 Sculptor 体外突变系统 (Amersham Bioscience) 引入突变分别得到质粒 pCB1H4-19 和质粒 pCB1S1-17。将通过用 XbaI 和 XhoI 消化质粒 pCB1H4-19 而获得的约 6kb 的片段与用 XbaI 消化, 然后用 XhoI 部分消化质粒 pCB1S1-17 而获得的约 1.2kb 片段连接, 得到 pCB1-M。

[0186] CBn-Stu :5' -GATACATGATGCGCAGGCCTTAGTCGACTAGAATG C-3' (SEQ ID NO :21)

[0187] Cbc-Xho :5' -GATCCTCAAGCTTTTGTCTGAGTACCTTACAAGCA C-3' (SEQ ID NO :22)

[0188] 然后将用 SalI 消化 pCB1-M 而得到的约 2.7kb 片段克隆到质粒 pUC119 中, 得到质粒 pCB1-SalM。对从质粒 pCB1-SalM 获得的单链 DNA 用磷酸化寡核苷酸 CB1-SmSph、CB1-Bam 和 CB1-Pst 退火, 从而通过 Sculptor 体外突变系统引入突变。

[0189] CB1-SmSph :5' -GGAGGGTGCATGCCGACTGAGCCCGGGCAGTAG CC-3' (SEQ ID NO :23)

[0190] CB1-Bam :5' -GCCGGGAGAGGATCCAGTGGAGG-3' (SEQ ID NO :24)

[0191] CB1-Pst :5' -GCTCGAGTACCTTACTGCAGGCACTGAGAG-3' (SEQ ID NO :25)

[0192] 然后用 XbaI 和 EcoRI 消化质粒 pUC118, 用 DNA 平端化试剂盒进行平端化并自连, 从而得到质粒 pUC118-SBN。将预先用相同的限制性酶剪切的约 1.4kb 的 cbh1 启动子片段克隆到用 HindIII 和 SalI 消化的 pUC118-SBN 中。将得到的质粒用 SalI 消化, 并与约 2.8kb 引入突变的 cbh1 编码区和终止子区连接, 得到 pCB2-M2。引入的突变 (除 BamHI 突变外) 如表 5 所示。

[0193] 表 5

[0194]

SalI StuI

5'-ctagtcgactaaggccctgcgcatcATGTATCAAAAGTGGCCCTCATCTCGGCTTCTTGGCTACT (SEQ ID NO: 26)

MetTyrGlnLysLeuAlaLeuIleSerAlaPheLeuAlaThr (SEQ ID NO: 27)

SmaI SphI PstI XhoI HindIII

5'-GCGCGGCTCAGTGGCATTGCAACC -- CAGTGGCTGCAAGTAAggtactcgagcaaaagcttgag-3' (SEQ ID NO: 28)

AlaArgAlaGlnSerAlaCysThr -- GlnCysLeuGln (SEQ ID NO: 29)

↑ 成熟蛋白质的N末端

[0195] 实施例 B2 :表达 N 末端替代型 STCE1 和野生型 STCE1 的表达载体的构建, 绿色木霉的转化和活性评估

[0196] (1) 表达载体的构建

[0197] 将 STCE1 基因编码区连接到实施例 1 中获得的质粒 pCB1-M2 中, 得到 pCB-Ste 作为野生型 STCE1 的表达载体。

[0198] 首先, 以质粒 pUC118-STCEex (FERM BP-10127) 为模板, 用引物组合 STCE1-TNERV 和 STCE1-TCET22I 进行 PCR, 扩增 STCE1 基因的编码区。将得到的约 1kbp 的 PCR 产物用 EcoRV 和 EcoT22I 消化, 并与预先用 StuI 和 PstI 消化的 pCB1-M2 连接, 得到 pCB-Ste。

[0199] STCE1-TNERV :GGGGATATCGCGCATCATGCGTTCCTCCCCCGT CCTC (SEQ ID NO :30)

[0200] STCE1-TCET22I :GGGATGCATTTAAAGGCACTGCGAGTACCAGT C (SEQ ID NO :31)

[0201] 将去除分泌信号序列的 STCE1 基因成熟蛋白的编码区连接到实施例 B1 中获得的 pCB1-M2 中,得到 pCB-Sts 作为 N 末端替代型 STCE1 的表达载体。

[0202] 首先,以质粒 pUC118-STCEex (FERM BP-10127) 为模板,用引物组合 STCE1-TmNSph 和 STCE1-TCET22I 进行 PCR,扩增 STCE1 基因成熟蛋白的编码区。将得到的约 1kbp 的 PCR 产物用 SphI 和 EcoT22I 消化,并与预先用 SphI 和 PstI 消化的 pCB1-M2 连接,得到 pCB-Sts。

[0203] STCE1-TmNSph :GGGGCATGCGATGGCAAGTCGACCCGCTAC (SEQ ID NO :32)

[0204] (2) 绿色木霉菌株 2 的产生

[0205] 将绿色木霉 MC300-1 (FERM BP-6047) 孢子悬液 (约 10^9 CFU/mL) 用位于 30cm 处的两盏 UV 灯的 UV 进行照射,并轻轻振荡。将悬液铺在选择性培养基上,28℃培养 7 天。筛选长出的菌株作为具有尿嘧啶需求的绿色木霉菌株。

[0206] 在本文中,选择培养基是补充 $10 \mu\text{g/mL}$ 尿嘧啶和 1mg/mL 5-氟乳清酸的基本培养基 [0.2%磷酸二氢钾,0.4%硫酸铵,0.03%尿素,0.03%七水硫酸镁,0.03%氯化钙,0.5%葡萄糖,2.5%琼脂和 0.01%痕量元素 (通过在 1L 水中溶解 5mg 七水硫酸铁,1.56mg 七水硫酸镁,1.4mg 七水硫酸锌和 2mg 氯化钴而制备)]。

[0207] (3) 含野生型 STCE1 或 N 末端替代型 STCE1 的转化子

[0208] 用实施例 B2(1) 中获得的 pCB-Ste 或 pCB-Sts 通过共转化方法,用实施例 B2(2) 中获得的绿色木霉菌株 2 和来源于 *Neurospora crassa* 的含 pyr4 基因的标记质粒 pPYR4 转化绿色木霉,得到在基本培养基上生长的菌株作为转化子。

[0209] 首先,将绿色木霉菌株 2 接种到含 50mL 菌丝形成培养基 [1%酵母提取物,1%麦芽提取物,2%聚脲,2.5%葡萄糖,0.1%磷酸氢二钾,0.05%七水硫酸镁和 0.0001%尿嘧啶 (pH7.0)] 的 200mL 锥形瓶中,并在 28℃培养 2 天。通过离心收集得到的细胞,并用 0.5mol/L 蔗糖洗涤,并悬浮在含 3mg/mL β -葡糖醛酸糖苷酶 (Sigma)、 1mg/mL 几丁质酶和 1mg/mL 消解酶 20T (Seikagaku Corp.) 的 0.5mol/L 蔗糖溶液中。悬浮液在 30℃轻轻振荡 60 到 90 分钟以得到原生质体。通过过滤和离心收集得到的原生质体,用 SUTC 缓冲液洗涤,并重悬到 SUTC 中而得到原生质体悬浮液。

[0210] 向 $100 \mu\text{L}$ 原生质体悬浮液中加入 $10 \mu\text{L}$ DNA 溶液 (pCB-Ste 或 pCB-Sts)。在冰上静置 5 分钟,加入 $400 \mu\text{L}$ PEG 溶液,然后整个溶液在冰上静置 20 分钟。将得到的原生质体悬浮液用 SUTC 洗涤后,用含 0.5mol/L 蔗糖的基本培养基和软琼脂一起覆盖在原生质体悬浮液上。平板在 28℃培养 5 天,然后将长出的克隆转移到基本培养基上,在其上获得转化子克隆。

[0211] (4) 培养基上清中野生型 STCE1 和 N 末端替代型 STCE1 的 EGU 活性比较

[0212] 根据 W098/11239 中所述方法培养得到的转化子,得到培养上清。将培养上清进行 SDS-PAGE,选择可清晰检测到约 45kDa 的 STCE1 带的培养上清。测定培养上清在 pH10,存在或不存在 LAS 时的 EGU 活性,并比较存在 LAS 时活性的降低程度。结果如表 6 所示。在表 6 中,以 pH10 时的 EGU 活性为 100,用相应的百分比来表示 EGU 活性。

[0213] 表 6

[0214]

	表达载体	EGU (%)	
		无LAS	有LAS
野生型STCE1	pCB-Ste	100	14.5
N末端替代型STCE1	pCB-Sts	100	25.9

[0215] 结果表明,比较存在LAS时的EGU活性发现,N末端替代型STCE1的EGU活性约是野生型STCE1的1.8倍。另外,不能检测到绿色木霉菌株2的EGU活性,因此表明EGU活性均来源于整合表达载体表达的重组酶。

[0216] (5) 野生型STCE1和N末端替代型STCE1 N末端氨基酸分析

[0217] 用以上转化子的培养上清制备含0.05mol/L Tris-HCl(pH8.0)/1mol/L硫酸钠的溶液,并吸附在具有相等床体积(BV)的TOYOPEARL Butyl-650S(Tosoh Corporation)上。用3个BV的0.05mol/L Tris-HCl(pH8.0)/1mol/L硫酸钠洗去非吸附性成分,以除去非吸附性成分。然后用3个BV的0.05mol/L Tris-HCl(pH8.0)/0.5mol/L硫酸钠进行洗脱,并将洗脱液用Pellicon XL(截留5000;Millipore)和Ultrafree-15(截留5000;Millipore)进行浓缩。用获得的浓缩液制备含0.1mol/L磷酸钠(pH5.8)/1mol/L硫酸铵溶液。将得到的溶液用Resource PHE(6mL;Amersham Bioscience)进行疏水层析,以回收非吸附组分。将非吸附组分浓缩,并用PD-10(Amersham Bioscience)脱盐,并用该组分制备含0.05mol/L Tris-HCl(pH7.5)的溶液。将该溶液通过Resource Q(6mL;Amersham Bioscience),回收非吸附组分。将非吸附组分脱盐,并制备含0.05mol/L醋酸钠(pH5.0)的溶液。将该溶液通过Resource S(6mL;Amersham Bioscience),回收非吸附组分。将非吸附组分浓缩,并用Superdex 75pg(16×60mm;Amersham Bioscience)进行凝胶过滤,以回收分子量约为45000的组分。每个组分(来源于以上转化子培养上清)仅含一条45kDa的带。

[0218] 将纯化的组分进行SDS-PAGE,并将电泳的蛋白电转移到PVDF膜上。从blot上剪切下约45kDa的STCE1带,并用蛋白测序仪Model492(Applied Biosystems)分析氨基酸序列。

[0219] 结果表明,从pCB1-Ste转化子获得的野生型STCE1序列(SEQ ID NO.5)与来源于Staphylotrichum coccosporum IF0(目前名称:NBRC)31817的DNA序列一致。

[0220] 但从N末端替代型STCE1转化子获得的STCE1的氨基酸序列不能用上述方法测定。用pfu焦谷酰胺氨肽酶(Takara Bio)除去N末端替代型STCE1修饰的N末端,然后测定其氨基酸序列(SEQ NO:4)。

[0221] 结果总结于表7中,比较野生型STCE1和N末端替代型STCE1的N末端氨基酸序列。结果表明,每个STCE1的N末端氨基酸序列与从构建的相应表达载体预测的序列一致。

[0222] 表7

[0223]

	N末端氨基酸	N末端氨基酸序列
野生型STCE1	丙氨酸(未修饰)	ADGKSTR (SEQ ID NO: 33)
N末端替代型STCE1	焦谷氨酸(修饰)	pQSACDGKSTR (SEQ ID NO: 34)

[0224] (6) 纯化的野生型STCE1和纯化的N末端替代型STCE1的EGU活性比较

[0225] 测定实施例B2(5)中获得的纯化的野生型STCE1和纯化的N末端替代型STCE1在pH10,不存在或存在LAS时的EGU活性,并比较存在LAS时活性的降低程度。结果如表8所

示。在表 8 中,以 pH10 时的 EGU 活性为 100,用相应的百分比来表示 EGU 活性。

[0226] 表 8

[0227]

	EGU (%)	
	无LAS	有LAS
纯化的野生型STCE1	100	16.2
纯化的N末端替代型STCE1	100	30.0

[0228] 结果表明,比较存在 LAS 时的 EGU 活性发现,纯化的 N 末端替代型 STCE1 的 EGU 活性约是纯化的野生型 STCE1 的 1.9 倍。

[0229] 实施例 B3:添加到 STCE1 的 N 末端序列的改变,载体的构建,绿色木霉的转化及活性评估

[0230] (1) 仅向成熟蛋白 N 末端添加焦谷氨酸的表达载体的构建

[0231] 将去除分泌信号序列的 STCE1 基因成熟蛋白的编码区连接到实施例 B1 获得的质粒 pCB1-M2 中,得到 pCB-Stm11 作为仅添加焦谷氨酸的 STCE1 的表达载体。

[0232] 首先,以质粒 pUC118-STCEex (FERM BP-10127) 为模板,用引物组合 STCE-TmNSma 和 STCE1-TCET22I 进行 PCR,扩增 STCE1 基因成熟蛋白的编码区。将得到的约 1kbp 的 PCR 产物用 SmaI 和 EcoT22I 消化,并与预先用 SmaI 和 PstI 消化的 pCB1-M2 连接,得到 pCB-Stm11。用上述载体 pCB-Stm11 转化的绿色木霉可产生 N 末端氨基酸序列为 pQADGKSTR (SEQ ID NO: 41) 的 STCE1 [即 1 个焦谷氨酸 (pQ) 加到野生型 STCE1N 末端的序列]。

[0233] STCE-TmNSma:GGCCCGGGCTCAGGCCGATGGCAAGTCGACCCG (SEQ ID NO:35)

[0234] (2) 添加焦谷氨酸的 STCE1 的 EGU 活性比较

[0235] 用 pCB-Stm11 根据实施例 B2(3) 所述的方法转化绿色木霉菌株 2。根据 W098/11239 培养得到的转化子,得到培养上清。将培养上清进行 SDS-PAGE,选择可清晰检测到约 45kDa 的 STCE1 带的培养上清。测定培养上清在 pH10,存在或不存在 LAS 时的 EGU 活性,并与实施例 B2(4) 中野生型 STCE1 比较存在 LAS 时活性的降低程度。结果如表 9 所示。在表 9 中,以 pH10 时的 EGU 活性为 100,用相应的百分比来表示 EGU 活性。

[0236] 表 9

[0237]

	表达载体	EGU (%)	
		无LAS	有LAS
野生型STCE1	pCB-Stm	100	14.5
pQ添加型STCE1	pCB-Stm11	100	23.0

[0238] 结果表明,比较存在 LAS 时的 EGU 活性发现,焦谷氨酸添加型 STCE1 的 EGU 活性约是野生型 STCE1 的 1.6 倍。

[0239] 实施例 B4:添加到 STCE1 的 N 末端序列的改变,载体的构建,绿色木霉的转化及活性评估

[0240] (1) 向成熟蛋白 N 末端添加含 pQ 的 4 个氨基酸的表达载体的构建

[0241] 将去除分泌信号序列的 STCE1 基因成熟蛋白的编码区连接到实施例 B1 获得的质

粒 pCB1-M2 中,得到 pCB-Stm12 作为添加含焦谷氨酸的 4 个氨基酸的 STCE1 的表达载体。

[0242] 首先,以质粒 pUC118-STCEex(FERM BP-10127)为模板,用引物组合 STCE-TmNSph-2 和 STCE1-TCET22I 进行 PCR,扩增 STCE1 基因成熟蛋白的编码区。将得到的约 1kbp 的 PCR 产物用 SphI 和 EcoT22I 消化,并与预先用 SphI 和 PstI 消化的 pCB1-M2 连接,得到 pCB-Stm12。用上述载体 pCB-Stm12 转化的绿色木霉可产生 N 末端氨基酸序列为 pQSACADGKSTR(SEQ ID NO:42) [即含 4 个氨基酸(N 末端为焦谷氨酸)的肽加到野生型 STCE1N 末端的序列)的 STCE1。

[0243] STCE-TmNSph-2:GGGCATGCGCCGATGGCAAGTCGACCCGC (SEQ ID NO:36)

[0244] (2) 含 4 个氨基酸的 pQ-STCE1 的 EGU 活性比较

[0245] 用 pCB-Stm12 根据实施例 B2(3)所述的方法转化绿色木霉菌株 2。根据 W098/11239 培养得到的转化子,得到培养上清。将培养上清进行 SDS-PAGE,选择可清晰检测到约 45kDa 的 STCE1 带的培养上清。测定培养上清在 pH10,存在或不存在 LAS 时的 EGU 活性,并与实施例 B2(4)中野生型 STCE1 比较存在 LAS 时活性的降低程度。结果如表 10 所示。在表 10 中,以 pH10 时的 EGU 活性为 100,用相应的百分比来表示 EGU 活性。

[0246] 表 10

[0247]

	表达载体	EGU (%)	
		无LAS	有LAS
野生型STCE1	pCB-Stm12	100	14.5
pQ-4-AAAs添加型STCE1	pCB-Stm12	100	22.4

[0248] 结果表明,比较存在 LAS 时的 EGU 活性发现,含焦谷氨酸的 4 个氨基酸的添加型 STCE1 的 EGU 活性约是野生型 STCE1 的 1.5 倍。

[0249] 工业应用

[0250] 与野生型纤维素酶相比较,本发明中的蛋白的内切葡聚糖酶活性 在存在表面活性剂时降低较小,因此可作为衣物洗涤剂的成分。

[0251] 虽然本发明通过特定实施例进行描述,但各种对本领域技术人员明显的改变和修饰并不偏离所附权利要求的范围。

[0252] 附图的简要描述

[0253] 图 1 所示的是内切葡聚糖酶 STCE1 [信号肽 (SEQ ID NO:43),成熟蛋白 (SEQ ID NO:44)] 的氨基酸序列与已知属于家族 45 的内切葡聚糖酶 NCE4 [信号肽 (SEQ ID NO:45),成熟蛋白 (SEQ ID NO:46)] 和 NCE5 [信号肽 (SEQ ID NO:47),成熟蛋白 (SEQ ID NO:48)] 的氨基酸序列关于 N 末端序列的比较结果。

[0254] 图 2 所示的是图 1 描述的关于 C 末端序列的比较结果。

[0255] 序列列表文本文件

[0256] “人工序列”的特征在序列列表中以数字标识进行描述。每个核苷酸序列 SEQ ID NOS:6-10、12-13、15-18、21-26、28、30-32 和 35-36 是引物 MNC-02、引物 MNC-03、引物 MKA-05、引物 MKA-06、pMKD01、引物 pMN-Bam、pJD01、引物 NCE4-Ne、引物 NCE4-Ce、引物 NCE4-Ns、引物 NCE4-Cs、引物 CBn-Stu、引物 CBc-Xho、引物 CB1-SmSph、引物 CB1-Bam、引物

CB1-Pst、pCB1-M2、pCB1-M2、引物 STCE1-TNERV、引物 STCE1-TCET22I、引物 STCE1-TmNSph、引物 STCE1-TmNSma 和引物 STCE1-TmNSph-2。每个氨基酸序列 SEQ ID NOS :11、14、27 和 29 是 pMKD01、pJD01、pCB1-M2 和 pCB1-M2。

23

CPC1262719D. 120917. 序列表. TXT

tct att gcc ggc agc aat gag gcg ggc tgg tgc tgc gcc tgc tac gag Ser Ile Ala Gly Ser Asn Glu Ala Gly Trp Cys Cys Ala Cys Tyr Glu 85 90 95	288
ctc acc ttc aca tcc ggt cct gtt get ggc aag aag atg gtc gtc cag Leu Thr Phe Thr Ser Gly Pro Val Ala Gly Lys Lys Met Val Val Gln 100 105 110	336
tcc acc agc act ggc ggt gat ctt ggc agc aac cac ttc gat ctc aac Ser Thr Ser Thr Gly Gly Asp Leu Gly Ser Asn His Phe Asp Leu Asn 115 120 125	384
atc ccc ggc ggc ggc gtc ggc atc ttc gac gga tgc act ccc cag ttc Ile Pro Gly Gly Gly Val Gly Ile Phe Asp Gly Cys Thr Pro Gln Phe 130 135 140	432
ggc ggt ctg ccc ggc cag cgc tac ggc ggc atc tcg tcc cgc aac gag Gly Gly Leu Pro Gly Gln Arg Tyr Gly Gly Ile Ser Ser Arg Asn Glu 145 150 155 160	480
tgc gat cgg ttc ccc gac gcc ctc aag ccc ggc tgc tac tgg cgc ttc Cys Asp Arg Phe Pro Asp Ala Leu Lys Pro Gly Cys Tyr Trp Arg Phe 165 170 175	528
gac tgg ttc aag aac gcc gac aac ccg age ttc age ttc cgt cag gtc Asp Trp Phe Lys Asn Ala Asp Asn Pro Ser Phe Ser Phe Arg Gln Val 180 185 190	576
caa tgc cca gcc gag ctc gtc get cgc acc gga tgc cgc cgc aac gac Gln Cys Pro Ala Glu Leu Val Ala Arg Thr Gly Cys Arg Arg Asn Asp 195 200 205	624
gac ggc aac ttc cct gcc gtc cag atc ccc tcc age age acc age tet Asp Gly Asn Phe Pro Ala Val Gln Ile Pro Ser Ser Ser Thr Ser Ser 210 215 220	672
ccg gtc ggc cag cct acc agt acc agc acc acc tcc acc tcc acc acc Pro Val Gly Gln Pro Thr Ser Thr Ser Thr Thr Ser Thr Ser Thr Thr 225 230 235 240	720
tcg age ccg ccc gtc cag cct acg act ccc age ggc tgc act get gag Ser Ser Pro Pro Val Gln Pro Thr Thr Pro Ser Gly Cys Thr Ala Glu 245 250 255	768
agg tgg gct cag tgc ggc ggc aat ggc tgg age ggc tgc acc acc tgc Arg Trp Ala Gln Cys Gly Gly Asn Gly Trp Ser Gly Cys Thr Thr Cys 260 265 270	816
gtc gct ggc agc acc tgc acg aag att aat gac tgg tac cat cag tgc Val Ala Gly Ser Thr Cys Thr Lys Ile Asn Asp Trp Tyr His Gln Cys 275 280 285	864
ctg taa Leu	870

[0003]

CPCH1262719D. 120917. 序列表. TXT

```

<210> 2
<211> 289
<212> PRT
<213> Humicola insolens MN200-1

<220>
<221> misc_feature
<222> (1)..(3)
<223> 焦谷氨酸

<400> 2

Gln Asn Cys Gly Ser Ala Asp Gly Lys Ser Thr Arg Tyr Trp Asp Cys
1          5          10          15

Cys Lys Pro Ser Cys Gly Trp Ala Lys Lys Ala Pro Val Asn Gln Pro
20          25          30

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

Lys Ser Gly Cys Glu Pro Gly Gly Val Ala Tyr Ser Cys Ala Asp Gln
50          55          60

Thr Pro Trp Ala Val Asn Asp Asp Phe Ala Phe Gly Phe Ala Ala Thr
65          70          75          80

Ser Ile Ala Gly Ser Asn Glu Ala Gly Trp Cys Cys Ala Cys Tyr Glu
85          90          95

Leu Thr Phe Thr Ser Gly Pro Val Ala Gly Lys Lys Met Val Val Gln
100         105         110

Ser Thr Ser Thr Gly Gly Asp Leu Gly Ser Asn His Phe Asp Leu Asn
115         120         125

Ile Pro Gly Gly Gly Val Gly Ile Phe Asp Gly Cys Thr Pro Gln Phe
130         135         140

Gly Gly Leu Pro Gly Gln Arg Tyr Gly Gly Ile Ser Ser Arg Asn Glu
145         150         155         160

Cys Asp Arg Phe Pro Asp Ala Leu Lys Pro Gly Cys Tyr Trp Arg Phe
165         170         175

Asp Trp Phe Lys Asn Ala Asp Asn Pro Ser Phe Ser Phe Arg Gln Val
180         185         190

Gln Cys Pro Ala Glu Leu Val Ala Arg Thr Gly Cys Arg Arg Asn Asp
195         200         205

Asp Gly Asn Phe Pro Ala Val Gln Ile Pro Ser Ser Ser Thr Ser Ser
210         215         220

Pro Val Gly Gln Pro Thr Ser Thr Ser Thr Thr Ser Thr Ser Thr Thr
225         230         235         240

```

[0004]

CPCH1262719D.120917.序列表.TXT

Ser Ser Pro Pro Val Gln Pro Thr Thr Pro Ser Gly Cys Thr Ala Glu
245 250 255

Arg Trp Ala Gln Cys Gly Gly Asn Gly Trp Ser Gly Cys Thr Thr Cys
260 265 270

Val Ala Gly Ser Thr Cys Thr Lys Ile Asn Asp Trp Tyr His Gln Cys
275 280 285

Leu

<210> 3

<211> 897

<212> DNA

<213> Staphylotrichum coccosporum IFO 31817

<220>

<221> CDS

<222> (1)..(897)

<223>

<220>

<221> misc_feature

<222> (1)..(3)

<223> 焦谷氨酸

<220>

<221> 来源

<222> (13)..(897)

<223> Staphylotrichum coccosporum IFO 31817

<400> 3

cag tcg gca tgc gat ggc aag tcc acc cgc tac tgg gac tgc tgc aag 48
Gln Ser Ala Cys Asp Gly Lys Ser Thr Arg Tyr Trp Asp Cys Cys Lys
1 5 10 15

cct tcg tgc teg tgg ccc ggc aag gcc tcg gtg aac cag ccc gtc ttc 96
Pro Ser Cys Ser Trp Pro Gly Lys Ala Ser Val Asn Gln Pro Val Phe
20 25 30

gcc tgc agc gcc aac ttc cag cgc atc agc gac ccc aac gtc aag tcg 144
Ala Cys Ser Ala Asn Phe Gln Arg Ile Ser Asp Pro Asn Val Lys Ser
35 40 45

ggc tgc gac ggc ggc tcc gcc tac gcc tgc gcc gac cag acc ccg tgg 192
Gly Cys Asp Gly Gly Ser Ala Tyr Ala Cys Ala Asp Gln Thr Pro Trp
50 55 60

gcc gtc aac gac aac ttc teg tac ggc ttc gcc gcc acg tcc atc tcg 240
Ala Val Asn Asp Asn Phe Ser Tyr Gly Phe Ala Ala Thr Ser Ile Ser
65 70 75 80

ggc ggc aac gag gcc tcg tgg tgc tgt ggc tgc tac gag ctg acc ttc 288
Gly Gly Asn Glu Ala Ser Trp Cys Cys Gly Cys Tyr Glu Leu Thr Phe

[0005]

CPCH1262719D. 120917. 序列表. TXT

85	90	95	
acc tcg ggc ccc gtc gct ggc aag acc atg gtt gtc cag tcc acc tcg Thr Ser Gly Pro Val Ala Gly Lys Thr Met Val Val Gln Ser Thr Ser 100 105 110			336
acc ggc ggc gac ctc ggc acc aac cac ttc gac ctg gcc atg ccc ggt Thr Gly Gly Asp Leu Gly Thr Asn His Phe Asp Leu Ala Met Pro Gly 115 120 125			384
ggt ggt gtc ggc atc ttc gac ggc tgc tgc ccc cag ttc ggc ggc ctc Gly Gly Val Gly Ile Phe Asp Gly Cys Ser Pro Gln Phe Gly Gly Leu 130 135 140			432
gcc ggc gac cgc tac ggc ggc gtc tgc tgc cgc agc cag tgc gac tcg Ala Gly Asp Arg Tyr Gly Gly Val Ser Ser Arg Ser Gln Cys Asp Ser 145 150 155 160			480
ttc ccc gcc gcc ctc aag ccc ggc tgc tac tgg cgc ttc gac tgg ttc Phe Pro Ala Ala Leu Lys Pro Gly Cys Tyr Trp Arg Phe Asp Trp Phe 165 170 175			528
aag aac gcc gac aac ccg acc ttc acc ttc cgc cag gtc cag tgc ccg Lys Asn Ala Asp Asn Pro Thr Phe Thr Phe Arg Gln Val Gln Cys Pro 180 185 190			576
tcg gag ctc gtc gcc cgc acc ggc tgc cgc cgc aac gac gac ggc aac Ser Glu Leu Val Ala Arg Thr Gly Cys Arg Arg Asn Asp Asp Gly Asn 195 200 205			624
ttc ccc gtc ttc acc cct ccc tcg ggc ggt cag tcc tcc tcg tct tcc Phe Pro Val Phe Thr Pro Pro Ser Gly Gly Gln Ser Ser Ser Ser Ser 210 215 220			672
tcc tcc agc agc gcc aag ccc acc tcc acc tcc acc tcg acc acc tcc Ser Ser Ser Ser Ala Lys Pro Thr Ser Thr Ser Thr Ser Thr Thr Ser 225 230 235 240			720
acc aag gct acc tcc acc acc tcg acc gcc tcc agc cag acc tcg tcg Thr Lys Ala Thr Ser Thr Thr Ser Thr Ala Ser Ser Gln Thr Ser Ser 245 250 255			768
tcc acc ggc ggc ggc tgc gcc gcc cag cgc tgg geg cag tgc ggc ggc Ser Thr Gly Gly Gly Cys Ala Ala Gln Arg Trp Ala Gln Cys Gly Gly 260 265 270			816
atc ggg ttc tcg ggc tgc acc acg tgc gtc agc ggc acc acc tgc aac Ile Gly Phe Ser Gly Cys Thr Thr Cys Val Ser Gly Thr Thr Cys Asn 275 280 285			864
aag cag aac gac tgg tac tcg cag tgc ctt taa Lys Gln Asn Asp Trp Tyr Ser Gln Cys Leu 290 295			897
<210> 4			
<211> 298			

[0006]

CPC1262719D. 120917. 序列表. TXT

<212> PRT

<213> *Staphylotrichum coccosporum* IFO 31817

<220>

<221> misc_feature

<222> (1)..(3)

<223> 焦谷氨酸

<400> 4

Gln Ser Ala Cys Asp Gly Lys Ser Thr Arg Tyr Trp Asp Cys Cys Lys
1 5 10 15

Pro Ser Cys Ser Trp Pro Gly Lys Ala Ser Val Asn Gln Pro Val Phe
20 25 30

Ala Cys Ser Ala Asn Phe Gln Arg Ile Ser Asp Pro Asn Val Lys Ser
35 40 45

Gly Cys Asp Gly Gly Ser Ala Tyr Ala Cys Ala Asp Gln Thr Pro Trp
50 55 60

Ala Val Asn Asp Asn Phe Ser Tyr Gly Phe Ala Ala Thr Ser Ile Ser
65 70 75 80

Gly Gly Asn Glu Ala Ser Trp Cys Cys Gly Cys Tyr Glu Leu Thr Phe
85 90 95

Thr Ser Gly Pro Val Ala Gly Lys Thr Met Val Val Gln Ser Thr Ser
100 105 110

Thr Gly Gly Asp Leu Gly Thr Asn His Phe Asp Leu Ala Met Pro Gly
115 120 125

Gly Gly Val Gly Ile Phe Asp Gly Cys Ser Pro Gln Phe Gly Gly Leu
130 135 140

Ala Gly Asp Arg Tyr Gly Gly Val Ser Ser Arg Ser Gln Cys Asp Ser
145 150 155 160

Phe Pro Ala Ala Leu Lys Pro Gly Cys Tyr Trp Arg Phe Asp Trp Phe
165 170 175

Lys Asn Ala Asp Asn Pro Thr Phe Thr Phe Arg Gln Val Gln Cys Pro
180 185 190

Ser Glu Leu Val Ala Arg Thr Gly Cys Arg Arg Asn Asp Asp Gly Asn
195 200 205

Phe Pro Val Phe Thr Pro Pro Ser Gly Gly Gln Ser Ser Ser Ser Ser
210 215 220

Ser Ser Ser Ser Ala Lys Pro Thr Ser Thr Ser Thr Ser Thr Thr Ser
225 230 235 240

Thr Lys Ala Thr Ser Thr Thr Ser Thr Ala Ser Ser Gln Thr Ser Ser

[0007]

CPCH1262719D. 120917. 序列表. TXT

245	250	255
Ser Thr Gly Gly Gly Cys Ala Ala Gln Arg Trp Ala Gln Cys Gly Gly		
260	265	270
Ile Gly Phe Ser Gly Cys Thr Thr Cys Val Ser Gly Thr Thr Cys Asn		
275	280	285
Lys Gln Asn Asp Trp Tyr Ser Gln Cys Leu		
290	295	

<210> 5
 <211> 1037
 <212> DNA
 <213> Staphylotrichum coccosporum IF0 31817

<220>
 <221> 信号肽
 <222> (1).. (63)
 <223>

<220>
 <221> 外显子
 <222> (64).. (333)
 <223>

<220>
 <221> 内含子
 <222> (334).. (419)
 <223>

<220>
 <221> 外显子
 <222> (420).. (1037)
 <223>

<400> 5

atgcgttct cccccgtct cgcacggcc ctggcctg cctccccct ggccgcctc	60
gct gcc gat ggc aag tcg acc cgc tac tgg gac tgt tgc aag ceg tcg	108
Ala Asp Gly Lys Ser Thr Arg Tyr Trp Asp Cys Cys Lys Pro Ser	
1 5 10 15	
tgc tcg tgg ccc ggc aag gcc tcg gtg aac cag ccc gtc ttc gcc tgc	156
Cys Ser Trp Pro Gly Lys Ala Ser Val Asn Gln Pro Val Phe Ala Cys	
20 25 30	
agc gcc aac ttc cag cgc atc agc gac ccc aac gtc aag tcg ggc tgc	204
Ser Ala Asn Phe Gln Arg Ile Ser Asp Pro Asn Val Lys Ser Gly Cys	
35 40 45	
gac ggc ggc tcc gcc tac gcc tgc gcc gac cag acc ccg tgg gcc gtc	252
Asp Gly Gly Ser Ala Tyr Ala Cys Ala Asp Gln Thr Pro Trp Ala Val	
50 55 60	

[0008]

CPCH1262719D. 120917. 序列表. TXT

aac gac aac ttc tgc tac ggc ttc gcc gcc acg tcc atc tgc ggc ggc Asn Asp Asn Phe Ser Tyr Gly Phe Ala Ala Thr Ser Ile Ser Gly Gly 65 70 75	300
aac gag gcc tgc tgg tgc tgt ggc tgc tac gag tgagtgttc ccccccccc Asn Glu Ala Ser Trp Cys Cys Gly Cys Tyr Glu 80 85 90	353
cccccccaac ccccggttcg gtcccttgcc gtgccttctt cataactaacc gcctaccccc	413
tccagg ctg acc ttc acc tgc ggc ccc gtc gct ggc aag acc atg gtt Leu Thr Phe Thr Ser Gly Pro Val Ala Gly Lys Thr Met Val 95 100	461
gtc cag tcc acc tgc acc ggc ggc gac ctc ggc acc aac cac ttc gac Val Gln Ser Thr Ser Thr Gly Gly Asp Leu Gly Thr Asn His Phe Asp 105 110 115 120	509
ctg gcc atg ccc ggt ggt ggt gtc ggc atc ttc gac ggc tgc tgc ccc Leu Ala Met Pro Gly Gly Gly Val Gly Ile Phe Asp Gly Cys Ser Pro 125 130 135	557
cag ttc ggc ggc ctc gcc ggc gac cgc tac ggc ggc gtc tgc tgc cgc Gln Phe Gly Gly Leu Ala Gly Asp Arg Tyr Gly Gly Val Ser Ser Arg 140 145 150	605
agc cag tgc gac tgc ttc ccc gcc gcc ctc aag ccc ggc tgc tac tgg Ser Gln Cys Asp Ser Phe Pro Ala Ala Leu Lys Pro Gly Cys Tyr Trp 155 160 165	653
cgc ttc gac tgg ttc aag aac gcc gac aac ccg acc ttc acc ttc cgc Arg Phe Asp Trp Phe Lys Asn Ala Asp Asn Pro Thr Phe Thr Phe Arg 170 175 180	701
cag gtc cag tgc ccg tgc gag ctc gtc gcc cgc acc ggc tgc cgc cgc Gln Val Gln Cys Pro Ser Glu Leu Val Ala Arg Thr Gly Cys Arg Arg 185 190 195 200	749
aac gac gac ggc aac ttc ccc gtc ttc acc cct ccc tgc ggc ggt cag Asn Asp Asp Gly Asn Phe Pro Val Phe Thr Pro Pro Ser Gly Gly Gln 205 210 215	797
tec tec tgc tet tcc tcc tcc agc agc gcc aag ccc acc tcc acc tcc Ser Ser Ser Ser Ser Ser Ser Ser Ala Lys Pro Thr Ser Thr Ser 220 225 230	845
acc tgc acc acc tcc acc aag gct acc tcc acc acc tgc acc gcc tcc Thr Ser Thr Thr Ser Thr Lys Ala Thr Ser Thr Thr Ser Thr Ala Ser 235 240 245	893
agc cag acc tgc tgc tcc acc ggc ggc ggc tgc gcc gcc cag cgc tgg Ser Gln Thr Ser Ser Ser Thr Gly Gly Gly Cys Ala Ala Gln Arg Trp 250 255 260	941
gcg cag tgc ggc ggc atc ggg ttc tgc ggc tgc acc acg tgc gtc agc Ala Gln Cys Gly Gly Ile Gly Phe Ser Gly Cys Thr Thr Cys Val Ser	989

[0009]

CPCH1262719D.120917.序列表.TXT

265	270	275	280	
ggc acc acc tgc aac aag cag aac gac tgg tac teg cag tgc ctt tga				1037
Gly Thr Thr Cys Asn Lys Gln Asn Asp Trp Tyr Ser Gln Cys Leu				
	285	290	295	
<210> 6				
<211> 36				
<212> DNA				
<213> 人工序列				
<220>				
<223> 人工序列序列描述: 引物 MNC-02				
<400> 6				
gagcgccaga actgtggatc cacttgggtga gcaatg				36
<210> 7				
<211> 35				
<212> DNA				
<213> 人工序列				
<220>				
<223> 人工序列序列描述: 引物 MNC-03				
<400> 7				
tccgccgttc tgagcggatc caggcgtttg gcgcg				35
<210> 8				
<211> 36				
<212> DNA				
<213> 人工序列				
<220>				
<223> 人工序列序列描述: 引物 MKA-05				
<400> 8				
gccgccccagc aggcgggatc cctcaccacc gagagg				36
<210> 9				
<211> 36				
<212> DNA				
<213> 人工序列				
<220>				
<223> 人工序列序列描述: 引物 MKA-06				
<400> 9				
tgatcgtega gtcagggatc cagaatttac aggcac				36
<210> 10				
<211> 44				
<212> DNA				
<213> 人工序列				
<220>				

[0010]

CPC11262719D. 120917. 序列表. TXT

<223> 人工序列序列描述: pMKD01

<220>

<221> misc_feature

<222> (24)..(25)

<223> 不连续的碱基

<400> 10

gagcgccaga actgtggatc cctctgcctg taageggatc cagg

44

<210> 11

<211> 10

<212> PRT

<213> 人工序列

<220>

<223> 人工序列序列描述: pMKD01

<220>

<221> NON_CONS

<222> (8)..(9)

<223>

<220>

<221> MOD_RES

<222> (1)..(1)

<223> 吡咯烷酮羧酸

<400> 11

Glu Arg Gln Asn Cys Gly Ser Leu Cys Leu
1 5 10

<210> 12

<211> 53

<212> DNA

<213> 人工序列

<220>

<223> 人工序列序列描述: 引物 pMN-Bam

<400> 12

gggtcaaacaa gtctgtgcgg atcctgggac aagatggcca agttcttcct tac

53

<210> 13

<211> 44

<212> DNA

<213> 人工序列

<220>

<223> 人工序列序列描述: pJD01

<220>

<221> misc_feature

<222> (24)..(25)

<223> 不连续的碱基

[0011]

CPCH1262719D.120917.序列表.TXT

<400> 13
tgcggatcct gggacaagat ggccccgttc tgagcggatc cagg 44

<210> 14
<211> 4
<212> PRT
<213> 人工序列

<220>
<223> 人工序列序列描述: pJD01

<220>
<221> NON_CONS
<222> (2).. (3)
<223>

<400> 14
Met Ala Pro Phe
1

<210> 15
<211> 37
<212> DNA
<213> 人工序列

<220>
<223> 人工序列序列描述: 引物 NCE4-Ne

<400> 15
ggggggatcc tgggacaaga tgcgttcctc ccctctc 37

<210> 16
<211> 32
<212> DNA
<213> 人工序列

<220>
<223> 人工序列序列描述: 引物 NCE4-Ce

<400> 16
ggggggatcc ctgcgtttac aggcactgat gg 32

<210> 17
<211> 30
<212> DNA
<213> 人工序列

<220>
<223> 人工序列序列描述: 引物 NCE4-Ns

<400> 17
ccggtgttgg ccggtaccgc tgatggcaag 30

<210> 18

[0012]

CPCH1262719D. 120917. 序列表. TXT

<211> 30
<212> DNA
<213> 人工序列

<220>
<223> 人工序列序列描述: 引物 NCE4-Cs

<400> 18
taaggccctc aaggatccct gcgtctacag

30

<210> 19
<211> 7
<212> PRT
<213> Humicola insolens MN200-1

<400> 19

Ala Asp Gly Lys Ser Thr Arg
1 5

<210> 20
<211> 12
<212> PRT
<213> Humicola insolens MN200-1

<220>
<221> MOD_RES
<222> (1)..(1)
<223> 吡咯烷酮羧酸

<220>
<221> MUTAGEN
<222> (1)..(5)
<223>

<400> 20

Gln Asn Cys Gly Ser Ala Asp Gly Lys Ser Thr Arg
1 5 10

<210> 21
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 人工序列序列描述: 引物 CBn-Stu

<400> 21
gatacatgat gcgcaggcct tagtcgacta gaatgc

36

<210> 22
<211> 36
<212> DNA
<213> 人工序列

[0013]

CPC1262719D. 120917. 序列表. TXT

<220>
<223> 人工序列序列描述: 引物 CBe-Xho

<400> 22
gatcctcaag cttttgctcg agtaccttac aagcac 36

<210> 23
<211> 35
<212> DNA
<213> 人工序列

<220>
<223> 人工序列序列描述: 引物 CB1-SmSph

<400> 23
ggagggtgca tgccgactga gcccgggcag tagcc 35

<210> 24

<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 人工序列序列描述: 引物 CB1-Bam

<400> 24
gccgggagag gatccagtgg agg 23

<210> 25
<211> 30
<212> DNA
<213> 人工序列

<220>
<223> 人工序列序列描述: 引物 CB1-Pst

<400> 25
gctcgagtac cttactgcag gcactgagag 30

<210> 26
<211> 66
<212> DNA
<213> 人工序列

<220>
<223> 人工序列序列描述: pCB1-M2

<400> 26
ctagtcgact aaggcctgcg catcatgtat caaaagttgg ccctcatctc ggccttcttg 60
getact 66

<210> 27
<211> 14
<212> PRT

[0014]

CPC11262719D. 120917. 序列表. TXT

<213> 人工序列

<220>

<223> 人工序列序列描述: pCB1-M2

<400> 27

Met	Tyr	Gln	Lys	Leu	Ala	Leu	Ile	Ser	Ala	Phe	Leu	Ala	Thr
1				5								10	

<210> 28

<211> 61

<212> DNA

<213> 人工序列

<220>

<223> 人工序列序列描述: pCB1-M2

<220>

<221> misc_feature

<222> (24)..(25)

<223> 不连续的碱基

<400> 28

gcccgggctc agtcggcatg caccagtgct ctgcagtaag gtactcgagc aaaagcttga 60

g 61

<210> 29

<211> 12

<212> PRT

<213> 人工序列

<220>

<223> 人工序列序列描述: pCB1-M2

<220>

<221> NON_CONS

<222> (8)..(9)

<223>

<400> 29

Ala	Arg	Ala	Gln	Ser	Ala	Cys	Thr	Gln	Cys	Leu	Gln
1				5							10

<210> 30

<211> 37

<212> DNA

<213> 人工序列

<220>

<223> 人工序列序列描述: 引物 STCE1-TNERV

<400> 30

ggggatatcg cgcatcatgc gtctctcccc cgtctctc 37

[0015]

CPCH1262719D. 120917. 序列表. TXT

<210> 31
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 人工序列序列描述: 引物 STCE1-TCET221

<400> 31
gggatgcatt taaaggcact gcgagtacca gtc 33

<210> 32
<211> 30
<212> DNA
<213> 人工序列

<220>
<223> 人工序列序列描述: 引物 STCE1-TmNSph

<400> 32
ggggcatgcg atggcaagtc gaccgctac 30

<210> 33
<211> 7
<212> PRT
<213> Staphylotrichum coccosporum IF0 31817

<400> 33
Ala Asp Gly Lys Ser Thr Arg
1 5

<210> 34
<211> 10
<212> PRT
<213> Staphylotrichum coccosporum IF0 31817

<220>
<221> MOD_RES
<222> (1).. (1)
<223> 吡咯烷酮羧酸

<220>
<221> MUTAGEN
<222> (1).. (4)
<223>

<400> 34
Gln Ser Ala Cys Asp Gly Lys Ser Thr Arg
1 5 10

<210> 35
<211> 33
<212> DNA

[0016]

CPCH1262719D. 120917. 序列表. TXT

<213> 人工序列

<220>

<223> 人工序列序列描述: 引物 STCE1-TmNSma

<400> 35

ggcccgggct caggccgatg gcaagtcgac ccg

33

<210> 36

<211> 29

<212> DNA

<213> 人工序列

<220>

<223> 人工序列序列描述: 引物 STCE1-TmNSph-2

<400> 36

gggcatgcgc cgatggcaag tcgaccgc

29

<210> 37

<211> 891

<212> DNA

<213> Staphylotrichum coccosporum IF0 31817

<220>

<221> CDS

<222> (1)..(891)

<223>

<220>

<221> misc_feature

<222> (1)..(3)

<223> 焦谷氨酸

<220>

<221> 来源

<222> (4)..(891)

<223> Staphylotrichum coccosporum IF0 31817

<400> 37

cag gcc gat ggc aag tcc acc cgc tac tgg gac tgc tgc aag cct tcg 48
 Gln Ala Asp Gly Lys Ser Thr Arg Tyr Trp Asp Cys Cys Lys Pro Ser
 1 5 10 15

tgc tcg tgg ccc ggc aag gcc tcg gtg aac cag ccc gtc ttc gcc tgc 96
 Cys Ser Trp Pro Gly Lys Ala Ser Val Asn Gln Pro Val Phe Ala Cys
 20 25 30

agc gcc aac ttc cag cgc atc agc gac ccc aac gtc aag tcg ggc tgc 144
 Ser Ala Asn Phe Gln Arg Ile Ser Asp Pro Asn Val Lys Ser Gly Cys
 35 40 45

gac ggc ggc tcc gcc tac gcc tgc gcc gac cag acc ccg tgg gcc gtc 192
 Asp Gly Gly Ser Ala Tyr Ala Cys Ala Asp Gln Thr Pro Trp Ala Val
 50 55 60

[0017]

CPCH1262719D. 120917. 序列表. TXT

aac gac aac ttc tgc tac ggc ttc gcc gcc acg tcc atc tgc ggc ggc Asn Asp Asn Phe Ser Tyr Gly Phe Ala Ala Thr Ser Ile Ser Gly Gly 65 70 75 80	240
aac gag gcc tgc tgg tgc tgt ggc tgc tac gag ctg acc ttc acc tgc Asn Glu Ala Ser Trp Cys Cys Gly Cys Tyr Glu Leu Thr Phe Thr Ser 85 90 95	288
ggc ccc gtc gct ggc aag acc atg gtt gtc cag tcc acc tgc acc ggc Gly Pro Val Ala Gly Lys Thr Met Val Val Gln Ser Thr Ser Thr Gly 100 105 110	336
ggc gac ctc ggc acc aac cac ttc gac ctg gcc atg ccc ggt ggt ggt Gly Asp Leu Gly Thr Asn His Phe Asp Leu Ala Met Pro Gly Gly Gly 115 120 125	384
gtc ggc atc ttc gac ggc tgc tgc ccc cag ttc ggc ggc ctc gcc ggc Val Gly Ile Phe Asp Gly Cys Ser Pro Gln Phe Gly Gly Leu Ala Gly 130 135 140	432
gac cgc tac ggc ggc gtc tgc tgc cgc agc cag tgc gac tgc ttc ccc Asp Arg Tyr Gly Gly Val Ser Ser Arg Ser Gln Cys Asp Ser Phe Pro 145 150 155 160	480
gcc gcc ctc aag ccc ggc tgc tac tgg cgc ttc gac tgg ttc aag aac Ala Ala Leu Lys Pro Gly Cys Tyr Trp Arg Phe Asp Trp Phe Lys Asn 165 170 175	528
gcc gac aac ccg acc ttc acc ttc cgc cag gtc cag tgc ccg tgc gag Ala Asp Asn Pro Thr Phe Thr Phe Arg Gln Val Gln Cys Pro Ser Glu 180 185 190	576
ctc gtc gcc cgc acc ggc tgc cgc cgc aac gac gac ggc aac ttc ccc Leu Val Ala Arg Thr Gly Cys Arg Arg Asn Asp Asp Gly Asn Phe Pro 195 200 205	624
gtc ttc acc cct ccc tgc ggc ggt cag tcc tcc tgc tet tcc tcc tcc Val Phe Thr Pro Pro Ser Gly Gly Gln Ser Ser Ser Ser Ser Ser 210 215 220	672
agc agc gcc aag ccc acc tcc acc tcc acc tgc acc acc tcc acc aag Ser Ser Ala Lys Pro Thr Ser Thr Ser Thr Ser Thr Thr Ser Thr Lys 225 230 235 240	720
gct acc tcc acc acc tgc acc gcc tcc agc cag acc tgc tgc tcc acc Ala Thr Ser Thr Thr Ser Thr Ala Ser Ser Gln Thr Ser Ser Ser Thr 245 250 255	768
ggc ggc ggc tgc gcc gcc cag cgc tgg gcg cag tgc ggc ggc atc ggg Gly Gly Gly Cys Ala Ala Gln Arg Trp Ala Gln Cys Gly Gly Ile Gly 260 265 270	816
ttc tgc ggc tgc acc acg tgc gtc agc ggc acc acc tgc aac aag cag Phe Ser Gly Cys Thr Thr Cys Val Ser Gly Thr Thr Cys Asn Lys Gln 275 280 285	864

[0018]

CPCH1262719D. 120917. 序列表. TXT

aac gac tgg tac tcg cag tgc ctt taa 891
 Asn Asp Trp Tyr Ser Gln Cys Leu
 290 295

 <210> 38
 <211> 296
 <212> PRT
 <213> *Staphylotrichum coccosporum* IF0 31817

 <220>
 <221> misc_feature
 <222> (1)..(3)
 <223> 焦谷氨酸

 <400> 38

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

[0019]

CPCH1262719D.120917. 序列表.TXT

```

210                215                220
Ser Ser Ala Lys Pro Thr Ser Thr Ser Thr Ser Thr Thr Ser Thr Lys
225                230                235                240

Ala Thr Ser Thr Thr Ser Thr Ala Ser Ser Gln Thr Ser Ser Ser Thr
245                250                255

Gly Gly Gly Cys Ala Ala Gln Arg Trp Ala Gln Cys Gly Gly Ile Gly
260                265                270

Phe Ser Gly Cys Thr Thr Cys Val Ser Gly Thr Thr Cys Asn Lys Gln
275                280                285

Asn Asp Trp Tyr Ser Gln Cys Leu
290                295

<210> 39
<211> 900
<212> DNA
<213> Staphylotrichum coccosporum IPO 31817

<220>
<221> CDS
<222> (1).. (900)
<223>

<220>
<221> 来源
<222> (13).. (900)
<223> 焦谷氨酸

<220>
<221> misc feature
<222> (1).. (3)
<223> 焦谷氨酸

<400> 39
cag teg gca tgc gcc gat ggc aag tcc acc cgc tac tgg gac tgc tgc      48
Gln Ser Ala Cys Ala Asp Gly Lys Ser Thr Arg Tyr Trp Asp Cys Cys
1          5          10          15

aag cct teg tgc teg tgg ccc ggc aag gcc teg gtg aac cag ccc gtc      96
Lys Pro Ser Cys Ser Trp Pro Gly Lys Ala Ser Val Asn Gln Pro Val
20        25        30

ttc gcc tgc agc gcc aac ttc cag cgc atc agc gac ccc aac gtc aag      144
Phe Ala Cys Ser Ala Asn Phe Gln Arg Ile Ser Asp Pro Asn Val Lys
35        40        45

tcg ggc tgc gac ggc ggc tcc gcc tac gcc tgc gcc gac cag acc cag      192
Ser Gly Cys Asp Gly Gly Ser Ala Tyr Ala Cys Ala Asp Gln Thr Pro
50        55        60

tgg gcc gtc aac gac aac ttc teg tac ggc ttc gcc gcc acg tcc atc      240
Trp Ala Val Asn Asp Asn Phe Ser Tyr Gly Phe Ala Ala Thr Ser Ile

```

[0020]

CPCH1262719D. 120917. 序列表. TXT

65	70	75	80	
tgc ggc ggc aac gag gcc tgc tgg tgc tgt ggc tgc tac gag ctg acc				288
Ser Gly Gly Asn Glu Ala Ser Trp Cys Cys Gly Cys Tyr Glu Leu Thr	85	90	95	
ttc acc tgc ggc ccc gtc gct ggc aag acc atg gtt gtc cag tcc acc				336
Phe Thr Ser Gly Pro Val Ala Gly Lys Thr Met Val Val Gln Ser Thr	100	105	110	
tgc acc ggc ggc gac ctc ggc acc aac cac ttc gac ctg gcc atg ccc				384
Ser Thr Gly Gly Asp Leu Gly Thr Asn His Phe Asp Leu Ala Met Pro	115	120	125	
ggc ggt ggt gtc ggc atc ttc gac ggc tgc tgc ccc cag ttc ggc ggc				432
Gly Gly Gly Val Gly Ile Phe Asp Gly Cys Ser Pro Gln Phe Gly Gly	130	135	140	
ctc gcc ggc gac cgc tac ggc ggc gtc tgc tgc cgc agc cag tgc gac				480
Leu Ala Gly Asp Arg Tyr Gly Gly Val Ser Ser Arg Ser Gln Cys Asp	145	150	155	160
tgc ttc ccc gcc gcc ctc aag ccc ggc tgc tac tgg cgc ttc gac tgg				528
Ser Phe Pro Ala Ala Leu Lys Pro Gly Cys Tyr Trp Arg Phe Asp Trp	165	170	175	
ttc aag aac gcc gac aac ccg acc ttc acc ttc cgc cag gtc cag tgc				576
Phe Lys Asn Ala Asp Asn Pro Thr Phe Thr Phe Arg Gln Val Gln Cys	180	185	190	
ccg tgc gag ctc gtc gcc cgc acc ggc tgc cgc cgc aac gac gac ggc				624
Pro Ser Glu Leu Val Ala Arg Thr Gly Cys Arg Arg Asn Asp Asp Gly	195	200	205	
aac ttc ccc gtc ttc acc cct ccc tgc ggc ggt cag tcc tcc tgc tct				672
Asn Phe Pro Val Phe Thr Pro Pro Ser Gly Gly Gln Ser Ser Ser Ser	210	215	220	
tcc tcc tcc agc agc gcc aag ccc acc tcc acc tcc acc tgc acc acc				720
Ser Ser Ser Ser Ser Ala Lys Pro Thr Ser Thr Ser Thr Ser Thr Thr	225	230	235	240
tcc acc aag gct acc tcc acc acc tgc acc gcc tcc agc cag acc tgc				768
Ser Thr Lys Ala Thr Ser Thr Thr Ser Thr Ala Ser Ser Gln Thr Ser	245	250	255	
tgc tcc acc ggc ggc ggc tgc gcc gcc cag cgc tgg gcg cag tgc ggc				816
Ser Ser Thr Gly Gly Gly Cys Ala Ala Gln Arg Trp Ala Gln Cys Gly	260	265	270	
ggc atc ggg ttc tgc ggc tgc acc acg tgc gtc agc ggc acc acc tgc				864
Gly Ile Gly Phe Ser Gly Cys Thr Thr Cys Val Ser Gly Thr Thr Cys	275	280	285	
aac aag cag aac gac tgg tac tgc cag tgc ctt taa				900
Asn Lys Gln Asn Asp Trp Tyr Ser Gln Cys Leu				

[0021]

CPCH1262719D. 120917. 序列表. TXT

290

295

<210> 40
 <211> 299
 <212> PRT
 <213> *Staphylotrichum coccosporum* IF0 31817

 <220>
 <221> misc_feature
 <222> (1)..(3)
 <223> 焦谷氨酸

 <400> 40

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

[0022]

CPCH1262719D, 120917, 序列表, TXT

Ser Ser Ser Ser Ser Ala Lys Pro Thr Ser Thr Ser Thr Ser Thr Thr
225 230 235 240

Ser Thr Lys Ala Thr Ser Thr Thr Ser Thr Ala Ser Ser Gln Thr Ser
245 250 255

Ser Ser Thr Gly Gly Gly Cys Ala Ala Gln Arg Trp Ala Gln Cys Gly
260 265 270

Gly Ile Gly Phe Ser Gly Cys Thr Thr Cys Val Ser Gly Thr Thr Cys
275 280 285

Asn Lys Gln Asn Asp Trp Tyr Ser Gln Cys Leu
290 295

<210> 41

<211> 8

<212> PRT

<213> Staphylotrichum coccosporum IFO 31817

<220>

<221> MOD_RES

<222> (1).. (1)

<223> 吡咯烷酮羧酸

<220>

<221> MUTAGEN

<222> (1).. (1)

<223>

<400> 41

Gln Ala Asp Gly Lys Ser Thr Arg
1 5

<210> 42

<211> 11

<212> PRT

<213> Staphylotrichum coccosporum IFO 31817

<220>

<221> MOD_RES

<222> (1).. (1)

<223> 吡咯烷酮羧酸

<220>

<221> MUTAGEN

<222> (1).. (4)

<223>

<400> 42

Gln Ser Ala Cys Ala Asp Gly Lys Ser Thr Arg
1 5 10

<210> 43

[0023]

CPCH1262719D, 120917, 序列表, TXT

<211> 21

<212> PRT

<213> *Staphyлотrichum coccosporum* IF0 31817

<400> 43

Met Arg Ser Ser Pro Val Leu Arg Thr Ala Leu Ala Ala Ala Leu Pro
1 5 10 15

Leu Ala Ala Leu Ala
20

<210> 44

<211> 295

<212> PRT

<213> *Staphyлотrichum coccosporum* IF0 31817

<400> 44

Ala Asp Gly Lys Ser Thr Arg Tyr Trp Asp Cys Cys Lys Pro Ser Cys
1 5 10 15

Ser Trp Pro Gly Lys Ala Ser Val Asn Gln Pro Val Phe Ala Cys Ser
20 25 30

Ala Asn Phe Gln Arg Ile Ser Asp Pro Asn Val Lys Ser Gly Cys Asp
35 40 45

Gly Gly Ser Ala Tyr Ala Cys Ala Asp Gln Thr Pro Trp Ala Val Asn
50 55 60

Asp Asn Phe Ser Tyr Gly Phe Ala Ala Thr Ser Ile Ser Gly Gly Asn
65 70 75 80

Glu Ala Ser Trp Cys Cys Gly Cys Tyr Glu Leu Thr Phe Thr Ser Gly
85 90 95

Pro Val Ala Gly Lys Thr Met Val Val Gln Ser Thr Ser Thr Gly Gly
100 105 110

Asp Leu Gly Thr Asn His Phe Asp Leu Ala Met Pro Gly Gly Gly Val
115 120 125

Gly Ile Phe Asp Gly Cys Ser Pro Gln Phe Gly Gly Leu Ala Gly Asp
130 135 140

Arg Tyr Gly Gly Val Ser Ser Arg Ser Gln Cys Asp Ser Phe Pro Ala
145 150 155 160

Ala Leu Lys Pro Gly Cys Tyr Trp Arg Phe Asp Trp Phe Lys Asn Ala
165 170 175

Asp Asn Pro Thr Phe Thr Phe Arg Gln Val Gln Cys Pro Ser Glu Leu
180 185 190

Val Ala Arg Thr Gly Cys Arg Arg Asn Asp Asp Gly Asn Phe Pro Val

[0024]

CPCH1262719D. 120917. 序列表. TXT

```

195                200                205
Phe Thr Pro Pro Ser Gly Gly Gln Ser Ser Ser Ser Ser Ser Ser
210                215                220

Ser Ala Lys Pro Thr Ser Thr Ser Thr Ser Thr Thr Ser Thr Lys Ala
225                230                235                240

Thr Ser Thr Thr Ser Thr Ala Ser Ser Gln Thr Ser Ser Ser Thr Gly
245                250                255

Gly Gly Cys Ala Ala Gln Arg Trp Ala Gln Cys Gly Gly Ile Gly Phe
260                265                270

Ser Gly Cys Thr Thr Cys Val Ser Gly Thr Thr Cys Asn Lys Gln Asn
275                280                285

Asp Trp Tyr Ser Gln Cys Leu
290                295

<210> 45
<211> 20
<212> PRT
<213> Humicola insolens

<400> 45

Met Arg Ser Ser Pro Leu Leu Arg Ser Ala Val Val Ala Ala Leu Pro
1          5          10          15

Val Leu Ala Leu
20

<210> 46
<211> 286
<212> PRT
<213> Humicola insolens

<400> 46

Ala Ala Asp Gly Lys Ser Thr Arg Tyr Trp Asp Cys Cys Lys Pro Ser
1          5          10          15

Cys Gly Trp Ala Lys Lys Ala Pro Val Asn Gln Pro Val Phe Ser Cys
20          25          30

Asn Ala Asn Phe Gln Arg Leu Thr Asp Phe Asp Ala Lys Ser Gly Cys
35          40          45

Glu Pro Gly Gly Val Ala Tyr Ser Cys Ala Asp Gln Thr Pro Trp Ala
50          55          60

Val Asn Asp Asp Phe Ala Phe Gly Phe Ala Ala Thr Ser Ile Ala Gly
65          70          75          80

Ser Asn Glu Ala Gly Trp Cys Cys Ala Cys Tyr Glu Leu Thr Phe Thr

```

[0025]

CPCH1262719D, 120917, 序列表, TXT

<400> 48

Ala Gln Ser Gly Ser Gly Arg Thr Thr Arg Tyr Trp Asp Cys Cys Lys
1 5 10 15

Pro Ser Cys Ala Trp Pro Gly Lys Gly Pro Ala Pro Val Arg Thr Cys
20 25 30

Asp Arg Trp Asp Asn Pro Leu Phe Asp Gly Gly Asn Thr Arg Ser Gly
35 40 45

Cys Asp Ala Gly Gly Gly Ala Tyr Met Cys Ser Asp Gln Ser Pro Trp
50 55 60

Ala Val Ser Asp Asp Leu Ala Tyr Gly Trp Ala Ala Val Asn Ile Ala
65 70 75 80

Gly Ser Asn Glu Arg Gln Trp Cys Cys Ala Cys Tyr Glu Leu Thr Phe
85 90 95

Thr Ser Gly Pro Val Ala Gly Lys Arg Met Ile Val Gln Ala Ser Asn
100 105 110

Thr Gly Gly Asp Leu Gly Asn Asn His Phe Asp Ile Ala Met Pro Gly
115 120 125

Gly Gly Val Gly Ile Phe Asn Ala Cys Thr Asp Gln Tyr Gly Ala Pro
130 135 140

Pro Asn Gly Trp Gly Gln Arg Tyr Gly Gly Ile Ser Gln Arg His Glu
145 150 155 160

Cys Asp Ala Phe Pro Glu Lys Leu Lys Pro Gly Cys Tyr Trp Arg Phe
165 170 175

Asp Trp Phe Leu Asn Ala Asp Asn Pro Ser Val Asn Trp Arg Gln Val
180 185 190

Ser Cys Pro Ala Glu Ile Val Ala Lys Ser Gly Cys Ser Arg
195 200 205

1

33

NCE4	MRSSPLLRSAVVAALPVLAL————AADGKSTRYWDCCCKPSCGNWAK	21
	***** *	
STCE1	MRSSPVLRTALAAALPLAALA————ADGKSTRYWDCCCKPSCSNWPG	20
	* ***** **	
NCE5	MQLPLTLLTLLPALAA————AQS GSGRTTRYWDCCCKPSCANWPG	23
	信号肽 → 催化区域 →	
NCE4	KAPVNOPVFSCNANFQRLTDF—DAKSGCEPGGVAYSCADQTPWAVND DFA	70
	** ***** * ***** * ***** ** ** ***** *	
STCE1	KASVNOPVFAC SANFORISDP—NVKS GCD—GGSAYACADQTPWAVND NFS	68
	* ** * * * ***** ** ** * ** ***** *	
NCE5	KGPA—PVRTC DRWDNPLFDGGNTRSGCDAGGGAYMCS DQSPWAVSDDLA—	71
NCE4	FGFAATS IAGSNEAGWCCACYELTFTSGPVAGKMMVQSTSTGGDLGSNH	120
	***** * *** ** ***** ***** *	
STCE1	YGFAATS ISGGNEASWCCG CYELTFTSGPVAGKTMVQSTSTGGDLGTNH	118
	** ** * * ** *** ***** * ** ***** **	
NCE5	YGWAAVN IAGSNERQWCCACYELTFTSGPVAGKRM IVOASNTGGDLGNH	121
NCE4	FDLNIPGGGVGIFDGCTPQFGGLP—GQRYGGISSRNECDRFPDALKPG	167
	*** ***** ***** * ***** ** ** *****	
STCE1	FDLAMPGGGVGIFDGCSPQFGGLA—GDRYGGVSSRSQCD SFPAALKPG	165
	** ***** * * * * ***** * ** ** *****	
NCE5	FDIAMPGGGVGIFNACTDQYGAPPNGWNGQRYGGISQRHECDAFPEKLKPG	171

图 1

NCE4	CYWRFDWFKNADNPSFSFRQVQCPAELVARTGCCRNDGDFPAVQIPSSS	217
	***** * ***** ***** **	
STCE1	CYWRFDWFKNADNPTFTFRQVQCPSELVARTGCCRNDGDFPVFTPPSGG	215
	***** ***** *** ** * ** *	
NCE5	CYWRFDWFLNADNPSVNRQVSCPAEIVAKSGCSR	206
	接头 →	
NCE4	TSSPVGQPTSTSTTSTTTSSPPVQPTTPS	255
	** * ***** ** ** * ** *	
STCE1	QSSSSSSSSAKPTSTSTTSTTKATSTTSTASSQTSSSTGGCAAQRWA	265
NCE5		
	CBD →	
NCE4	CQCGGNGWSGCTTCVAGSTCTKINDWYHQCL	286
	***** * ***** * ** * *****	
STCE1	CQCGGIGFSGCTTCVSGTTCNKONDWYSQCL	295
NCE5		

图 2