(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 101094918 B
(45)授权公告日 2014.05.14

(21)申请号 200580045498.2
(22)申请日 2005.12.20
(30)优先权数据
60/640,399 2004.12.30 US
60/648,233 2005.01.27 US
(85)PCT国际申请进入国家阶段日
2007.06.29
(86)PCT国际申请的申请数据
PCT/US2005/046435 2005.12.20
(87)PCT国际申请的公布数据
WO2006/073839 EN 2006.07.13
(73)专利权人 金克克国际有限公司
地址 美国加利福尼亚州
(72)发明人 K·A·克拉森 N·邓恩-克莱曼
S·E·兰特斯 C·E·皮尔格林
P·范索林格恩 M·沃德
(74)专利代理机构 北京市中咨律师事务所 11247
代理人 黄革生 刘金辉
(51)Int.Cl.
C12N 9/58(2006.01)
C12N 15/63(2006.01)
C12N 5/10(2006.01)
C08B 31/00(2006.01)

(56)对比文件
朱桂兰、童群义.β-呋喃果糖苷酶合成低聚乳果糖的工艺研究.《食品工业科技》.2008,第29卷(第12期),第188-190页.
Berka R.M.,et al.Molecular Cloning and Deletion of the Gene Encoding Aspergillopepsin a from Aspergillus-awamori.《GENE》.1990,第86卷(第2期),153-162.
Viterbo A.,et al.Isolation of two aspartyl proteases from Trichoderma asperellum expressed during colonization of cucumber roots.《FEMS MICROBIOLOGY LETTERS》.2004,第238卷(第1期),151-158.

审查员 王璟

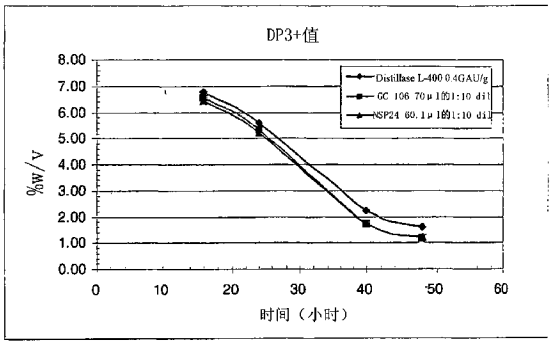
权利要求书1页 说明书25页
序列表21页 附图18页

(54)发明名称

酸性真菌蛋白酶

(57)摘要

本发明涉及新型酸性蛋白酶,更具体地涉及NSP24家族蛋白酶和NSP25家族蛋白酶,包括它们的生物学活性片段,并涉及编码所述蛋白酶的核酸分子。同样提供了包括编码所述蛋白酶的核酸序列的载体和宿主细胞,生产所述蛋白酶的方法,酶组合物以及使用所述蛋白酶的方法。



CN 101094918 B

1. 一种分离的 NSP24 家族蛋白酶，其能够在酸性条件下水解蛋白质，并且其为 SEQ ID NO :10 所示氨基酸序列。

2. 权利要求 1 的分离的 NSP24 家族蛋白酶，其能够在酸性条件下水解蛋白质，其中蛋白酶活性的最适 pH 范围为 pH3.0 至 4.5。

3. 母体 NSP24 家族蛋白酶，其包含且通过蛋白水解切割产生权利要求 1-2 之任一项的酸性蛋白酶，其为 SEQ ID NO :2 所示的氨基酸序列。

4. 一种分离的多核苷酸，其编码权利要求 1-2 之任一项的 NSP24 家族蛋白酶或权利要求 3 的母体 NSP24 家族蛋白酶。

5. 权利要求 4 所述的分离的多核苷酸，其为 SEQ ID NO :8 的序列。

6. 一种载体，其包括权利要求 4 或 5 所述的多核苷酸。

7. 一种宿主细胞，其用权利要求 4 或 5 所述的多核苷酸转化。

8. 权利要求 7 所述的宿主细胞，其中所述宿主细胞是丝状真菌细胞。

9. 权利要求 8 所述的宿主细胞，其中所述丝状真菌细胞是曲霉属物种 (Aspergillus spp.)、镰刀菌属物种 (Fusarium spp.) 或者木霉属物种 (Trichoderma spp.)。

10. 权利要求 9 所述的宿主细胞，其中所述曲霉是黑曲霉 (A.niger)、米曲霉 (A.oryzae)、构巢曲霉 (A.nidulans) 或泡盛曲霉 (A.awamori)。

11. 权利要求 9 所述的宿主细胞，其中所述木霉是里氏木霉 (T.reesei)。

12. 权利要求 8 所述的宿主细胞，其中所述宿主细胞是四重缺失的宿主细胞。

13. 一种产生蛋白酶的方法，包括

a) 向宿主细胞中导入多核苷酸，其包括被可操纵地连接于权利要求 4 或 5 的多核苷酸的启动子，

b) 将所述宿主细胞在适于表达和产生 NSP24 家族蛋白酶的培养条件下培养，并

c) 产生所述 NSP24 家族蛋白酶。

14. 根据权利要求 13 所述的方法，进一步包括回收所述产生的蛋白酶。

15. 一种酶组合物，其包括权利要求 1-2 之任一项所述的 NSP24 家族蛋白酶。

16. 权利要求 15 所述的酶组合物，其中所述组合物是清洁组合物。

17. 权利要求 16 所述的酶组合物，其中所述清洁组合物是去污剂组合物。

18. 权利要求 15 所述的酶组合物，其中所述组合物是淀粉水解组合物。

19. 权利要求 15 所述的酶组合物，其中所述组合物是动物饲料组合物。

20. 权利要求 15 所述的酶组合物，其中所述组合物被用于乙醇生产过程。

21. 权利要求 15 所述的酶组合物，其中所述组合物被用于淀粉糖化过程。

22. 权利要求 15 所述的酶组合物，其中所述组合物被用于麦芽糖或果糖的生产。

23. 权利要求 15 所述的酶组合物，其中所述组合物是个人护理组合物。

24. 权利要求 15 所述的酶组合物，进一步包括葡糖淀粉酶。

25. 权利要求 15 所述的酶组合物，进一步包括 α 淀粉酶。

26. 权利要求 25 所述的酶组合物，进一步包括葡糖淀粉酶。

27. 一种水解淀粉的方法，包括将含淀粉底物与权利要求 15 所述的酶组合物在适于淀粉水解的条件下接触，并获得水解的淀粉。

酸性真菌蛋白酶

[0001] 相关申请

[0002] 本申请要求2004年12月30日提交的题为Acid FungalProteases(酸性真菌蛋白酶)的美国临时专利申请号60/640,399和2005年1月27日提交的题为Acid Fungal Proteases(酸性真菌蛋白酶)的美国临时专利申请号60/648,233的优先权,其内容在此被全部并入本文作为参考。

发明领域

[0003] 本发明涉及编码被命名为NSP24家族蛋白酶、NSP25家族蛋白酶和PepA蛋白酶的酸性蛋白酶的多核苷酸;NSP24和NSP25家族蛋白酶多肽;包括所述蛋白酶的组合物以及它们的使用。

[0004] 发明背景

[0005] 蛋白酶是能够切开肽键的酶。酸性蛋白酶(例如具有酸性最适pH的蛋白酶)是由大量不同的生物,包括哺乳动物和微生物产生的。例如,微生物酸性蛋白酶是由细菌菌株诸如芽孢杆菌属某种(Bacillus sp.)的菌株(JP01240184)和真菌菌株诸如根霉属某种(Rhizopus sp.)(EP72978)、Schytalidium sp.(JP48091273)、Sulpholobus sp.、热原体属某种(Thermoplasma sp.)(WO/9010072)以及曲霉属某种(Aspergillus sp.)(JP50121486和EP82395)的菌株产生的。

[0006] Berka等(Gene(1990)96:313)公开了编码来自泡盛曲霉(Aspergillus awamori)的天冬氨酸蛋白酶曲霉胃蛋白酶(aspergillopepsin)A的基因。对编码来自米曲霉(Aspergillus oryzae)的天冬氨酸蛋白酶曲霉胃蛋白酶(aspergillopepsin)O的基因的克隆由Berka等(Gene(1993)125:195-198)加以描述。对编码来自米曲霉的酸性蛋白酶(PepA)的基因 的克隆由Gomi等(Biosci.Biotech.Biochem.(1993)57(7):1095-1100)加以描述。

[0007] 蛋白酶,具体而言是酸性蛋白酶,被广泛地应用工业应用中,例如用于制备食品和饲料、用于皮革工业(例如使皮革脱毛)、用于生产蛋白质水解产物、以及用于生产醇类诸如乙醇生产、酒类生产与酿造。

[0008] 对于许多不同的应用,特别是在食品和饲料工业中,对蛋白酶仍然有着持续的需求。

[0009] 发明概述

[0010] 申请人已发现大量新型蛋白酶基因,其包括编码NSP24蛋白酶(SEQ ID NO:2或SEQ ID NO:10)的新型nsp24基因;编码NSP25蛋白酶(SEQ ID NO:9)的新型nsp25基因;和编码新型PepA蛋白酶(SEQID NO:7)的新型pepA变体基因。

[0011] 因此,本发明的特征是NSP24蛋白酶、NSP25蛋白酶或PepA蛋白酶以及它们的变体的重组的或者基本上纯的制剂。

[0012] 在本发明的一些方面,蛋白酶是NSP24家族蛋白酶多肽,其包括与SEQ ID NO:2或SEQ ID NO:10中的氨基酸序列基本相同的氨基酸序列(下文图6所示)。在一些实施方式

中，NSP24 家族蛋白酶多肽是由 SEQ ID NO:8 中的核酸编码（下文图 5 所示），或者由具有与 SEQ IDNO:8 的核酸基本相同的核酸序列的核酸编码。

[0013] 在本发明的其它方面，NSP24 家族蛋白酶多肽与 SEQ ID NO:10 中的序列在氨基酸序列上具有多达 10 个残基的差异。在一些实施方式中，NSP24 家族蛋白酶多肽与 SEQ ID NO:10 中的序列在氨基酸序列上具有多达 10％残基的差异。在一些实施方式中，具有如此的差异，以至于 NSP24 家族蛋白酶多肽表现出 NSP24 蛋白酶的生物学活性，例如 NSP24 蛋白酶保留了天然存在的 NSP24 蛋白酶的生物学活性。

[0014] 在本发明进一步的方面，NSP24 家族蛋白酶多肽包括此处所述的 NSP24 蛋白酶序列以及其它的 N- 末端和 / 或 C- 末端氨基酸序列。

[0015] 在本发明的其它方面，NSP24 家族蛋白酶多肽包括来自 SEQ IDNO:2 或 SEQ ID NO:10 的氨基酸序列的全部或片段，在阅读框中，其与其它氨基酸残基融合，优选地与由基因组 DNA5′ 至编码来自 SEQ ID NO:1 或 SEQ ID NO:8 的序列的基因组 DNA 所编码的残基融合。

[0016] 仍在本发明的其它方面，NSP24 家族蛋白酶是重组体融合蛋白，其具有第一 NSP24 家族蛋白酶部分和第二多肽部分，例如具有与 NSP24 家族蛋白酶不相关的氨基酸序列的第二多肽部分。该第二多肽部分可以是 DNA 结合结构域或者聚合酶活化结构域。本发明的多肽包括那些由于多基因的存在、选择性转录事件、选择性 RNA 剪接事件以及选择性翻译和翻译后事件所产生的多肽。该多肽可以在系统例如培养的细胞中表达，该系统产生与表达的 NSP24 蛋白酶在天然细胞中表达时基本上相同的翻译后修饰，或者可以在这样的系统中表达，该系统导致当在天然细胞中表达时存在的翻译后修饰的省略。

[0017] 仍在其它方面，本发明涉及酶组合物，其包括 NSP24 家族蛋白酶以及一种或多种其它组分，例如载体、稀释剂或溶剂。所述其它组分可以是赋予该组合物体外用途、体内用途、药物用途或兽医用途的组分。在本方面的一些实施方式中，酶组合物将包括其它酶类。在优选的实施方式中，其它酶将会是葡糖淀粉酶、α 淀粉酶或它们的组合。

[0018] 仍在进一步的方面，本发明提供了基本上纯的核酸，其具有或包括编码 NSP24 家族蛋白酶多肽的核苷酸序列，所述多肽包括与 SEQ ID NO:2 或 SEQ ID NO:10 的氨基酸序列具有至少 80％序列同一性的氨基酸序列。

[0019] 在一些方面，NSP24 家族蛋白酶核酸将包括转录调控序列，例如可操纵地连接于 NSP24 家族蛋白酶基因序列的转录启动子或转录增强子序列的至少其中之一，例如以使 NSP24 家族蛋白酶基因序列适合作为表达载体使用。

[0020] 仍在其它方面，编码本发明的 NSP24 蛋白酶多肽（例如 SEQ IDNO:2）的核酸在严格条件下与核酸探针杂交，该探针对应于来自 SEQID NO:8 的至少 12 个连续的核苷酸，更优选地对应于来自 SEQ ID NO:8 的至少 20 个连续的核苷酸。

[0021] 本发明的另一方面提供了 NSP24 家族蛋白酶（例如 NSP24）在各种工业设施方面的应用。例如，NSP24 家族蛋白酶可以被用于农业废料的酶促降解以生产醇类燃料和其它重要的工业化学品，用于生产动物或人类食品，或者作为去污剂组合物的组分，用于皮革加工以及 蛋白质基的纤维加工（诸如毛料或丝绸），用于生物量应用，用于个人护理应用（皮肤、头发、口腔护理等），用于药物和保健护理应用以及用于生产适用于以上应用的新型肽。

[0022] 在进一步的方面，本发明涉及多核苷酸，其编码具有 SEQ IDNO:7 的 pepA 变体蛋白

酶 L388M。在一些实施方式中，该多核苷酸具有 SEQ ID NO:5 的序列。

[0023] 仍在另一方面，本发明涉及 NSP25 家族蛋白酶。在一些实施方式中，NSP25 家族蛋白酶将具有与 SEQ ID NO:9 至少 85％的序列同一性。在其它实施方式中，NSP25 家族蛋白酶将由与 SEQ ID NO:4 具有至少 85％序列同一性的多核苷酸所编码。仍在其它实施方式中，NSP25 家族蛋白酶将是母体 NSP25 家族蛋白酶的生物学活性片段。

附图说明

[0024] 图 1 举例说明了糖类的降解 (DP+3)％ w/v，使用了 1)NSP24、2) 可商业获得的蛋白酶，GC106 和 3)DISTILLASE，其不包括蛋白酶 (参见实施例 5)。

[0025] 图 2 描述了糖类的降解 (DP2)％ w/v，使用了 NSP24、GC106 和 DISTILLASE。

[0026] 图 3 举例说明了葡萄糖的形成 (DP1)，使用了 NSP24、GC106 和 DISTILLASE。对于 NSP24 和 GC106 样品，在 40 小时末剩余的葡萄糖量为 0.2％ w/v 以下，而在 48 小时末为 0.1％ w/v 以下。相反，对于 DISTILLASE，在 48 小时末以％ w/v 测量的葡萄糖量稍大于 1.0％ w/v。

[0027] 图 4 举例说明了 NSP24、GC106 和 DISTILLASE 的乙醇生产 (％ v/v)。通过使用两种蛋白酶样品产生乙醇的速率和量基本相同。相反，DISTILLASE 产生较少的乙醇且速度较慢。

[0028] 图 5A-D 举例说明了由里氏木霉 (Trichodermal reesei) 获得的 pTrex3g_NSP24cDNA 克隆的核苷酸序列 (SEQ ID NO:1)。NSP24 基因序列加下划线，推定的基因内含子序列用黑体加以标识。编码该蛋白酶的核酸序列用 SEQ ID NO:8 的序列加以表示。

[0029] 图 6A-B 举例说明了来自里氏木霉的 NSP24 的预测的氨基酸 序列 (407 个氨基酸) (SEQ ID NO:2)(图 6A)，以及 NSP24 核苷酸序列，其带有用粗体字母标识的推定内含子 (图 6B)(SEQ ID NO:8)。在图 6A 中，信号肽为粗体，前原序列 (prepro sequence) 为粗体加下划线，而成熟的 NSP24 蛋白以 KYGAPIS... 开始并用 SEQ ID NO:10 表示。

[0030] 图 7 举例说明了 pTrex3g_NSP24 载体以及图 5 核苷酸序列上的限制性酶切位点的位置。

[0031] 图 8 举例说明了 pepA 蛋白酶的核酸序列 (SEQ ID NO:3)。推定的内含子为粗体。

[0032] 图 9A-B 举例说明了编码新型 NSP25 蛋白酶 (399 个氨基酸)(SEQ ID NO:9) 的核酸序列 (SEQ ID NO:4)。信号序列为粗体。

[0033] 图 10 举例说明了新型 pepA 蛋白酶变体 (L388M)(SEQ ID NO:7) 的核酸序列 (SEQ ID NO:5)，其中图中加下划线的′ A′ 是由图 8pepA 中的′ C′ 改变而成。

[0034] 图 11 举例说明了表达载体，pSL899_pepA。

[0035] 图 12A-E 举例说明了表达载体 pSL899_pepA 的核苷酸序列 (SEQ ID NO:6)。Xho I 切割位点用 ^ 表示，而 Xba I 位点用 * 表示。pepA 的编码序列以粗体表示。内含子被加下划线。

[0036] 图 13 举例说明了由 SEQ ID NO:5 编码的蛋白的 PepA 变体 L388M 的氨基酸序列 (SEQ ID NO:7)。

[0037] 发明详述

[0038] 现在本发明将仅仅通过参考的方式使用下列定义和实施例来加以详细描述。本文

中所参考的全部专利和出版物，包括在此类专利和出版物中公开的全部序列，被明确并入本文作为参考。

[0039] 除非另作说明，本发明的实践将使用常规的细胞生物学、细胞培养、分子生物学、转基因生物学、微生物学、重组 DNA 以及免疫学技术，其在本领域技术人员的能力范围内。此类技术在文献中有所描述。参见例如 Sambrook、Fritsch 和 Maniatis 编辑的 Molecular CloningALaboratory Manual，第二版 (Cold Spring Harbor Laboratory Press :1989) ;Ausubel 等编辑的 Short Protocols in Molecular Biology(5thEd. 2002) ; DNA Cloning,Volumes I 和 II(D.N.Glover ed.,1985) ; Oligonucleotide Synthesis(M.J.Gait ed.,1984) ;Mullis 等美国专利号 :4,683,195 ;Nucleic Acid Hybridization(B.D.Hames&S.J.Higgins eds.1984) ;Transcription And Translation(B.D.Hames&S.J.Higgins eds.1984) ;Culture Of Animal Cells(R.I.Freshney, Alan R.Liss, Inc., 1987) ;Immobilized Cells And Enzymes(IRL Press,1986) ;B.Perbal,APractical-Guide To Molecular Cloning(1984) ; 论 文，Methods InEnzymology(Academic Press, Inc., N.Y.) ;Gene Transfer Vectors ForMammalian Cells(J.H.Miller 和 M.P.Calos eds., 1987, Cold SpringHarbor Laboratory) ;Methods In Enzymology, Vols.154 和 155(Wu 等 eds.),Immunochemical Methods In Cell And Molecular Biology(Mayer 和 Walker, eds., Academic Press,London,1987) ;Handbook Of ExperimentalImmunology,Volumes I-IV(D.M.Weir 和 C.C.Blackwell, eds.,1986) ;Manipulating the Mouse Embryo, (Cold Spring Harbor Laboratory Press,Cold Spring Harbor,N.Y.,1986)。同样，有关蛋白酶的制备、表达、分离和使用的方法的信息可以通过阅读美国专利号 6,768,001 而获得，该专利在此以其整体并入本文作为参考。

[0040] 通过下列详细说明并通过权利要求，本发明的其它特征和优势将是显而易见的。尽管与本文所述相似或等同的任何方法和材料都可以被用于本发明的实践或测试，但优选的方法和材料还是被加以描述。

[0041] 除非在本文中另作定义，此处所用的所有科学技术术语具有与本发明所属领域普通技术人员所通常理解的相同的含义。Singleton 等 DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY,2D ED., John Wiley and Sons, New York(1994) 和 Hale & Marham, THEHARPER COLLINS DICTIONARY OF BIOLOGY,Harper Perennial,NY(1991) 为技术人员提供了本发明中所使用的许多术语的通用字典。

[0042] 此处所提供的标题并非是限定本发明的各个方面或实施方式，其可以作为本说明书的参考以整体包含于其中。因此，以下即刻被定义的术语以整体作为本说明书的参考而被更充分地定义。

[0043] 数字范围包括限定该范围的数字。

[0044] 除非另作说明，核酸是按照 5′ 至 3′ 方向从左至右书写 ;氨基酸 序列是按照氨基至羧基方向从左至右书写。

[0045] 应当注意，如本说明书和所附权利要求中所用的，单数形式“一 (a)”、“一 (an)”和“该 (the)”，除非将其内容另作清楚的描述，包括复数参考。因此，例如，提及含“ (一种) 化合物 (a compound)”的组合物时，则包括了两种或更多种化合物的混合物。同样应当注意的是，术语″ 或 (or)″ ，除非将其内容另作清楚的描述，一般用作包括“和 / 或 (and/or)”

的含义。

[0046] 定义

[0047] “蛋白酶”表示来源于微生物，例如真菌、细菌，或者来自植物或动物的蛋白或者蛋白的多肽结构域或者多肽，其具有在蛋白质骨架的一个或多个不同位置催化切开肽键的能力（例如 E.C.3.4）。

[0048] “酸性蛋白酶”指在酸性条件下具有水解蛋白的能力的蛋白酶。

[0049] 如此处所用，“NSP24 家族蛋白酶”表示以其天然或野生型形式存在时具有蛋白酶活性的酶（例如图 6 的蛋白）；与 SEQ ID NO:2 或 SEQ ID NO:10 的氨基酸序列具有至少 50％、至少 60％、至少 70％、至少 80％、至少 85％、至少 90％、至少 93％、至少 95％、至少 97％、至少 98％和至少 99％序列同一性的蛋白酶蛋白；SEQ ID NO:2 或 SEQ IDNO:10 的氨基酸序列的衍生物，以及蛋白酶序列的生物学活性片段。

[0050] 如此处所用，“衍生物”表示来源于前体或母体蛋白（例如天然蛋白）的蛋白质，是通过在 C- 末端和 N- 末端之一或者两者上加入一个和多个氨基酸，在氨基酸序列中的一个或大量不同位点置换一个或多个氨基酸，在该蛋白的单个或两个末端或者在该氨基酸序列的一个或多个位点缺失一个或多个氨基酸，或者在该氨基酸序列中的一个或多个位点插入一个或多个氨基酸而获得的。

[0051] 如此处所用，“天然序列 NSP24”或“野生型 NSP24 序列”包括与来源于自然的 NSP24 家族蛋白酶具有相同氨基酸序列的多肽。

[0052] “生物学活性片段”（例如具有 SEQ ID NO:10 序列的 NSP24 家族蛋白酶的生物学活性片段）表示 NSP24 家族蛋白酶或 NSP25 家族蛋白酶，其具有蛋白酶活性，但所包含的序列少于 NSP24 家族蛋白酶或 NSP25 家族蛋白酶的前体或母体的全序列。

[0053] 如此处所用，术语“分离的”或“纯化的”指由于将蛋白酶从与其天然相关的一种或多种或所有天然存在的组分中分离出来而改变其天然状态的蛋白酶。

[0054] “PepA”指酸性蛋白酶，具有与 SEQ ID NO:7 至少 95％的序列同一性。

[0055] “L388M”指变体 PepA，其具有 SEQ ID NO:7 序列。

[0056] 如此处所用，“NSP25 家族蛋白酶”表示与 SEQ ID NO:9 具有至少 85％序列同一性的蛋白酶及其生物学活性片段。

[0057] “与 NSP24 家族蛋白酶不相关的”表示具有一种氨基酸序列，其与 SEQ ID NO110 的 NSP24 蛋白酶具有 30％以下的同源性、20％以下的同源性或者 10％以下的同源性。

[0058] 术语“肽”、“蛋白质”和“多肽”在本文中可以被互换使用。

[0059] 如此处所用，有关鉴定的氨基酸或核苷酸序列的“序列同一性百分比（％）”被定义为与感兴趣序列（例如 NSP24 家族蛋白酶序列）的氨基酸残基或核苷酸相一致的候选序列的氨基酸残基或核苷酸的百分比，这是经序列比对之后获得，且如有必要，引入空位，以便获得最大百分比的序列同一性，而不考虑作为部分序列同一性的任何保守取代。

[0060] 如此处所用，术语“α 淀粉酶（例如 E.C. 类 3.2.1.1）”指催化 α-1,4- 糖苷键水解的酶。此类酶也被描述为实现含有 1,4-α- 连接 D- 葡萄糖单位的多糖中 1,4-α-D- 糖苷键的外切或内切水解的酶。用于描述这些酶的另一个术语是“糖原酶”。示例性的酶包括 α-1,4- 葡聚糖 -4- 葡聚糖水化酶葡聚糖水解酶。

[0061] 如此处所用，术语“葡糖淀粉酶”指葡糖淀粉酶类的酶（例如 E.C.3.2.1.3，葡糖

淀粉酶，1,4-α-D-葡聚糖葡萄糖水化酶）。这些酶为外切作用酶，其从直链淀粉和支链淀粉分子的非还原性末端释放出葡糖基残基。尽管速度大大低于 α-1,4 键，但此酶也水解 α-1,6 键和 α-1,3 键。

[0062] 术语“启动子”表示一种调控序列，其涉及结合 RNA 聚合酶以启动基因转录。

[0063] 如此处所用，“异源启动子”是与基因或纯化的核酸非天然相 关的启动子。

[0064] 如此处所用，多肽的“纯化的制剂”或“基本上纯的制剂”表示多肽已被从细胞、与其天然存在的其它蛋白质、脂类或核酸分离出来。

[0065] 如此处所用，在植物或动物细胞的情况下，“纯化的细胞制品”指细胞的体外制品，而非整个完整的植物或动物。在培养细胞或微生物细胞的情况下，其由至少 10％且更优选地 50％的受试细胞的制品组成。

[0066] “基本上纯的核酸”，例如基本上纯的 DNA，是这样的核酸，该核酸是以下的一种或者两者：其与序列例如编码序列的之一或两者并非紧密相邻，其在所述核酸的来源生物中天然存在的基因组中与所述序列是紧密相邻的（例如 5′ 末端的序列和 3′ 末端的序列）；或者其基本不含在所述核酸的来源生物中存在的核酸序列。该术语包括，例如重组 DNA，其被整合入载体，例如整合入自我复制的质粒或病毒，或者整合入原核生物或真核生物的基因组 DNA，或者其以独立于其它 DNA 序列的单独分子（例如通过 PCR 或限制性内切核酸酶处理而产生的 cDNA 或基因组 DNA 片段）存在。基本上纯的 DNA 还包括重组 DNA，其为编码其它 NSP24 蛋白酶序列的杂合基因的一部分。

[0067] 如此处所用，“同源的”指两种多肽分子之间或者两种核酸分子之间的序列相似性。当两条比较序列中的位置都被相同的碱基或氨基酸单体亚基占据时，例如，若两 DNA 分子中每个分子的一个位置都被腺嘌呤所占据，那么所述分子在该位置上就是同源的。两序列间的同源性百分比，是两序列所共有的匹配位置或同源位置数除以所比较的位置数再乘以 100 的函数。例如，如果两序列中 10 个位置中的 6 个位置匹配或同源，那么两序列则为 60％同源。通过举例的方式，DNA 序列 ATTGCC 和 TATGGC 具有 50％的同源性。一般地，在两序列比对时进行比较以提供最大同源性。

[0068] 如此处所用，术语“载体”指被设计为将核酸导入一种或多种细胞类型的多核苷酸序列。载体包括克隆载体、表达载体、穿梭载体、质粒、噬菌粒、盒以及类似载体。

[0069] 如此处所用，“表达载体”表示一种 DNA 构建物，其包含被 可操纵地连接于合适的调控序列的 DNA 序列，所述调控序列能够影响该 DNA 在合适宿主中的表达。

[0070] 术语“表达”表示根据基因的核酸序列产生多肽的过程。

[0071] 如此处所用，术语“可操纵地连接”表示将调控区诸如启动子、终止子、分泌信号或增强子区结合于 (attached to) 或连接于 (linked to) 结构基因并控制该基因的表达。

[0072] 如此处所用，“来源于”微生物的物质（例如多核苷酸或蛋白质）表示该物质对于该微生物而言是天然的。

[0073] 如此处所用，“微生物”指细菌、真菌、病毒、原生动物以及其它的微生物或微观生物 (microscopic organisms)。

[0074] 如此处所用，“宿主菌株”或“宿主细胞”表示一种适合于表达载体的宿主，且包括所述细胞的后代，其中所述表达载体包括根据本发明的 DNA。

[0075] 术语“丝状真菌”指所有丝状形式的真菌亚门 (Eumycotina) 生物（参见

Alexopoulos, C. J. (1962), INTRODUCTORY MYCOLOGY, Wiley, New York and AINSWORTH AND BlSBY DICTIONARY OF THEFUNGI, 9thEd. (2001) Kirk 等, EdS., CAB International University Press, Cambridge UK)。这些真菌是通过带有细胞壁的营养菌丝体来加以表征的，其细胞壁包括几丁质、纤维素和其它复杂多糖。本发明的丝状真菌在形态学、生理学和遗传学上不同于酵母菌。丝状真菌的营养生长是通过菌丝的延伸进行的，碳的分解代谢是专性好氧的。

[0076] 如此处所用，术语“木霉属”或“木霉属某种”指以前或目前被分类为木霉属的任何真菌属。

[0077] 如此处所用，术语“四重缺失 (quad-delete)”或“四重缺失的 (quad-deleted)”宿主细胞指由木霉属宿主菌株细胞产生的细胞和原生质体，其缺少至少两个编码功能性内切葡聚糖酶 (endoglucanases) 的基因和至少两个编码功能性纤维二糖水解酶 (cellobiohydrolases) 的基因。

[0078] 如此处所用，术语“培养”指使微生物细胞群落在适当条件下在液体或固体培养基中生长。在一种实施方式中，培养指将淀粉底物诸如包含颗粒淀粉的底物经发酵生物转化为终产物（一般在容器或反 应器中进行）。发酵是通过微生物将有机物质酶促厌氧降解以产生较简单的有机化合物。当发酵过程在厌氧条件下进行时，并非意欲将该术语仅仅限定在严格的厌氧条件，因为发酵也在氧存在下进行。

[0079] 如此处所用，术语“接触”指将各种酶（一种或多种）放置于足够接近各种底物之处，以使所述酶（一种或多种）能够将该底物转变成为终产物。本领域的技术人员将认识到酶与各种底物的混合溶液可以实现接触。

[0080] 在关于将核酸序列中插入细胞的上下文中，术语“导入的”表示“转染”、或“转化”或“转导”，且包括涉及将核酸序列整合入真核或原核细胞，其中该核酸序列可以被整合入该细胞的基因组（例如染色体、质粒、质体或线粒体 DNA）中，转变成自主复制元或者是被瞬时表达（例如，转染的 mRNA）。

[0081] 如此处所用，有关细胞所用的术语“转化的”、“稳定转化的”和“转基因的”表示该细胞含有非天然的（例如异源的）核酸序列，该核酸序列或者整合入其基因组或者作为染色体外质粒被保持多代。

[0082] 如此处所用，有关多肽或多核苷酸的术语“异源的”指在宿主细胞中并非天然存在的多肽或多核苷酸。

[0083] 术语“过量表达”表示在宿主细胞中表达多肽的过程，其中多核苷酸已被导入该宿主细胞中。

[0084] 如此处所述，本发明的一个方面的特征是“基本上纯的”（或重组的）核酸，其包括编码 NSP24 家族蛋白酶或 NSP25 家族蛋白酶的核苷酸序列、和 / 或此类核酸的等价物。

[0085] 术语“等价物”指编码功能上等价的多肽的核苷酸序列。等价的核苷酸序列将包括由于一个或多个核苷酸置换、添加或缺失而有所不同的序列，诸如等位基因变体。例如在一些实施方式中，由于遗传密码的简并性，等价的核苷酸序列包括不同于 SEQ ID NO:8 核苷酸序列的序列，所述 SEQ ID NO:8 核苷酸序列编码 SEQ ID NO:2 中所示的 NSP24 蛋白酶。

[0086] 如此处所用，术语“糖化作用”指将淀粉酶促转化成为葡萄糖。

[0087] 如此处所用，“淀粉”指由植物的复合多糖碳水化合物组成的 任何物质，包括具有

式 $(C_6H_{10}O_5)_x$ 的直链淀粉和支链淀粉，其中 X 可以是任何数字。

[0088] 术语“颗粒淀粉”指生的未熟化的（生的）淀粉（例如尚未经过糊化的淀粉）。

[0089] 如此处所用，术语“凝胶化作用”表示通过熟化将淀粉分子溶解以形成粘性悬浮液。

[0090] 如此处所用，术语“液化作用”指淀粉转化的阶段，其间凝胶化的淀粉被水解以提供低分子量的可溶性糊精。

[0091] 如此处所用，术语“可溶性淀粉水解物”指由淀粉水解产生的可溶性产物，其可以包含单糖、二糖和寡糖（例如葡萄糖、麦芽糖和更高级的糖）。

[0092] 术语“单糖”表示聚合物诸如淀粉的一个单体单位，其中聚合度 (DP) 是 1（例如葡萄糖、甘露糖、果糖和半乳糖）。

[0093] 术语“二糖”表示一种包含两个共价连接的单糖单位的化合物 (DP2)（例如蔗糖、乳糖和麦芽糖）。

[0094] 术语“DP3+”表示各种聚合度在 3 以上的聚合物。

[0095] 蛋白酶和编码它们的多核苷酸

[0096] 本发明涉及 NSP24 家族蛋白酶，诸如酸性蛋白酶，并且还涉及酸性真菌蛋白酶，其与 SEQ ID NO:2 蛋白酶或 SEQ ID NO:10 蛋白酶（图 6）具有至少 50％、至少 60％、至少 70％、至少 80％、至少 85％、至少 90％、至少 93％、至少 95％、至少 97％、至少 98％、至少 99％的序列同一性。在一些实施方式中，NSP24 家族蛋白酶被命名为 NSP24，其包括 SEQ ID NO:10 的序列（成熟蛋白序列）或者还包括 SEQ ID NO:2 的前蛋白序列。

[0097] 在一些实施方式中，本发明涉及 NSP24 家族蛋白酶的生物学活性片段。在一些实施方式中，生物学活性片段包括具有至少 250 个氨基酸残基、至少 300 个氨基酸残基、至少 350 个氨基酸残基、至少 375 个氨基酸残基以及至少 400 个氨基酸残基的蛋白酶。

[0098] 在其它实施方式中，生物学活性片段包括多肽序列的至少 60％、至少 70％、至少 80％、至少 85％、至少 90％、至少 93％、至少 95％、至少 97％、至少 98％、至少 99％，所述多肽序列与图 6 中的蛋白质序列 (SEQ ID NO:2 或 SEQ ID NO:10) 具有至少 80％、至少 85％、至少 90％、至少 93％、至少 95％、至少 97％、至少 98％和至少 99％的序列同一性。在一些实施方式中，生物学活性片段将包括多肽序列的至少 60％、至少 70％、至少 80％、至少 85％、至少 90％、至少 95％和至少 98％，所述多肽序列与具有 SEQ ID NO:2 或 SEQ ID NO:10 的母体 NSP24 蛋白酶具有至少 95％的序列同一性。在一些实施方式中，生物学活性片段将包括多肽序列的至少 60％、至少 70％、至少 80％、至少 85％、至少 90％、至少 95％和至少 98％，所述多肽序列与具有 SEQID NO:2 或 SEQ ID NO:10 的母体 NSP24 蛋白酶具有至少 99％的序列同一性。

[0099] 在一些实施方式中，生物学活性片段为体内存在的多种片段，例如由转录后加工所产生的片段或者由选择性剪接的 RNA 的翻译所产生的片段。片段包括那些在天然的或内源性的细胞中表达的片段，例如由于翻译后加工、例如由于氨基末端信号序列的去除产生的片段，并包括在表达系统例如 CHO 细胞中产生那些片段。一些优选的片段是这样的片段，例如活性片段，其由蛋白水解切割或选择性剪接事件而产生。因为肽诸如 NSP24 家族蛋白酶通常表现出一定范围的生理学特性，并且因为此类特性可以被归因于由该分子的不同部分引起的，所以有用的 NSP24 家族蛋白酶片段或 NSP24 家族蛋白酶类似物是在任何 NSP24

蛋白酶活性生物学试验中表现出生物学活性的那些。

[0100] 在一些实施方式中，生物学活性片段包括具有 SEQ ID NO:2 或 SEQ ID NO:10 的 NSP24 的蛋白酶活性的至少 20％、至少 30％、至少 40％、至少 50％、至少 60％、至少 70％、至少 80％、至少 90％、至少 95％和至少 100％。在一些优选的实施方式中，片段或类似物在任何体内或体外 NSP24 蛋白酶试验中具有 NSP24 蛋白酶 (SEQ ID NO:2 或 SEQ ID NO:10) 的蛋白酶活性的至少 40％或至少 90％。

[0101] NSP24 家族蛋白酶或 NSP25 家族蛋白酶的片段可以通过本领域技术人员所已知的方法生成。表现出蛋白酶生物学活性的候选片段的能力可以通过如此处所述的本领域技术人员已知的方法加以评估。同样包括在内的是 NSP24 家族蛋白酶或 NSP25 家族蛋白酶，其含有对 于该肽的生物学活性非必需的残基或者由于选择性 mRNA 剪接或选择性蛋白质加工事件所产生的残基。

[0102] 在一些实施方式中，本发明所包括的蛋白酶是具有 SEQ IDNO:2 或 SEQ ID NO:10 的蛋白酶的衍生物。衍生物可以与 SEQ IDNO:10 具有至少 80％、85％、90％、93％、95％、97％、98％和 99％的序列同一性。

[0103] 本发明也包括多种蛋白酶类似物。蛋白酶类似物是带有增加肽稳定性的修饰的那些；此类类似物可以在肽序列中含有例如一个或多个非肽键（其取代了肽键）。同样包括在内的是：包括除天然存在的 L- 氨基酸以外的残基例如 D- 氨基酸，或者非天然存在的或合成氨基酸，例如 b 或氨基酸的类似物；以及环状类似物。类似物在氨基酸序列方面或者不涉及序列的方面或者两个方面，都不同于天然存在的蛋白酶诸如 NSP24 或 NSP25 蛋白。非序列修饰包括本发明所包括的蛋白酶的体内或体外的化学衍生作用。非序列修饰作用包括乙酰化作用、甲基化作用、磷酸化作用、羧化作用或糖基化作用的变化。

[0104] 在进一步的实施方式中，本发明包括 NSP25 家族蛋白酶。NSP25 家族蛋白酶是与 SEQ ID NO:9 的成熟蛋白序列（图 9）具有至少 85％、至少 90％、至少 93％、至少 95％、至少 97％、至少 98％和至少 99％氨基酸序列同一性的酸性蛋白酶，或者它们的生物学活性片段。一种具体的 NSP25 家族蛋白酶是具有 SEQ ID NO:9 的被命名为 NSP25 的蛋白酶。在一些实施方式中，NSP25 家族蛋白酶将是一种蛋白酶的生物学活性片段，其包括与 SEQ ID NO:9 具有至少 90％序列同一性的序列的至少 75％、至少 80％、至少 85％、至少 90％和至少 95％。在另一个实施方式中，NSP25 家族蛋白酶将是一种蛋白酶的生物学活性片段，其包括与 SEQ ID NO:9 具有至少 95％序列同一性的序列的至少 75％、至少 80％、至少 85％、至少 90％和至少 95％。

[0105] 尽管根据本发明的酸性蛋白酶能够在酸性条件下水解蛋白，但在一些实施方式中，蛋白酶活性的最适 pH 是在 pH3.0 至 5.5 的范围内。在一些实施方式中，蛋白酶活性的最适 pH 范围是在 pH3.0 至 5.0 之间，而在其它实施方式中，蛋白酶活性的最适 pH 范围是在 pH3.0 至 4.5 之间。

[0106] 根据本发明的蛋白酶，诸如 NSP24 家族蛋白酶或 NSP25 家族蛋白酶，可以包括氨基酸置换诸如使用 L- 氨基酸进行“保守氨基酸的置换”，其中一种氨基酸被另一种生物学相似的氨基酸置换。保守氨基酸的置换是那些保持了被置换的氨基酸的总电荷、疏水性 / 亲水性、和 / 或空间体积 (steric bulk) 的置换。保守置换的实例是在下列各组间进行的置换：Gly/Ala、Val/Ile/Leu、Lys/Arg、Asn/Gln、Glu/Asp、Ser/Cys/Thr 和 Phe/Trp/Tyr。衍

生物可以例如有少至 1 至 10 个氨基酸残基不同，诸如 6-10 个、少至 5 个、少至 4 个、3 个、2 个或者甚至 1 个氨基酸残基不同。表 1 举例说明了本领域所公认的示例性的氨基酸置换。此外，置换可以通过一种或多种不破坏蛋白酶生物学活性的非保守性氨基酸置换、缺失或插入来完成。

[0107] 表 1 保守氨基酸的置换

[0108]

对于氨基酸	符号	用其中任何一个置换
丙氨酸	A	D-Ala、Gly、β-Ala、L-Cys、D-Cys
精氨酸	R	D-Arg、Lys、D-Lys、homo-Arg、D-homo-Arg、Met、Ile、D-Met、D-Ile、Orn、D-Orn
天冬酰胺	N	D-Asn、Asp、D-Asp、Glu、D-Glu、Gln、D-Gln
天冬氨酸	D	D-Asp、D-Asn、Asn、Glu、D-Glu、Gln、D-Gln
半胱氨酸	C	D-Cys、S-Me-Cys、Met、D-Met、Thr、D-Thr
谷酰胺	Q	D-Gln、Asn、D-Asn、Glu、D-Glu、Asp、D-Asp
谷氨酸	E	D-Glu、D-Asp、Asp、Asn、D-Asn、Gln、D-Gln
甘氨酸	G	Ala、D-Ala、Pro、D-Pro、b-Ala、Acp
异亮氨酸	I	D-Ile、Val、D-Val、Leu、D-Leu、Met、D-Met
亮氨酸	L	D-Leu、Val、D-Val、Leu、D-Leu、Met、D-Met
赖氨酸	K	D-Lys、Arg、D-Arg、homo-Arg、D-homo-Arg、Met、D-Met、Ile、D-Ile、Orn、D-Orn
甲硫氨酸	M	D-Met、S-Me-Cys、Ile、D-Ile、Leu、D-Leu、Val、D-Val
苯丙氨酸	F	D-Phe、Tyr、D-Thr、L-多巴、His、D-His、Trp、D-Trp、反-3,4 或 5-苯基脯氨酸、顺-3,4 或 5-苯基脯氨酸 D-Pro、L-I-噻唑烷-4-羧酸、D-或 L-1-噁唑烷-4-
脯氨酸	P	羧酸
丝氨酸	S	D-Ser、Thr、D-Thr、allo-Thr、Met、D-Met、Met(O)、D-Met(O)、L-Cys、D-Cys
苏氨酸	T	D-Thr、Ser、D-Ser、allo-Thr、Met、D-Met、
		Met(O)、D-Met(O)、Val、D-Val
酪氨酸	Y	D-Tyr、Phe、D-Phe、L-多巴、His、D-His
缬氨酸	V	D-Val、Leu、D-Leu、Ile、D-Ile、Met、D-Met

[0109] 在一些实施方式中，本发明的蛋白酶是天然的序列。此类天然序列可以从自然界分离或者可以通过重组或合成的方式产生。术语“天然序列”具体地包括天然存在的截短形式或分泌形式的 NSP24 或 NSP25 家族蛋白酶（例如生物学活性片段），以及天然存在的变体形式（例如选择性剪接的形式）。

[0110] 在一些实施方式中，本发明的酸性蛋白酶是 PepA 蛋白酶，其与 SEQ ID NO:7 具有至少 97％、至少 98％和至少 99％的序列同一性。在一些实施方式中，该蛋白酶具有 SEQ ID NO:7 序列并被命名为“L388M”。在进一步的实施方式中，该蛋白酶由具有 SEQ ID NO:5 或 SEQ ID NO:3 序列的核苷酸序列所编码。

[0111] 本发明还涉及编码本发明所包括的蛋白酶的多核苷酸序列。一些此类多核苷酸包括：

[0112] a) 编码 NSP24 家族蛋白酶的多核苷酸，该蛋白酶与 SEQ IDNO:2 或 SEQ ID NO:10 具有至少 80％、至少 85％、至少 90％、至少 93％、至少 95％、至少 97％、至少 98％和至少 99％的序列同一性；

[0113] b) 编码 SEQ ID NO:2 序列的多核苷酸；

[0114] c) 具有 SEQ ID NO:8 序列的多核苷酸；

[0115] d) 编码 NSP24 家族蛋白酶生物学活性片段的多核苷酸；

[0116] e) 与 SEQ ID NO:8 序列具有至少 80％、至少 85％、至少 90％、至少 95％、至少 97％、至少 98％和至少 99％的序列同一性的多核苷酸；

[0117] f) 与对应于 SEQ ID NO:4、SEQ ID NO:8 或者 SEQ ID NO:4 或 SEQ ID NO:8 片段的 DNA 序列的核酸探针相杂交的多核苷酸，所述片段具有至少 10、15、20、30、40、50、60、70、

80、90、100 或 150 个连续的核苷酸；

[0118] g) 编码 NSP25 家族蛋白酶的多核苷酸，其与 SEQ ID NO:4 具有至少 85%、至少 90%、至少 93%、至少 95%、至少 97%、至少 98%和至少 99%的序列同一性；

[0119] h) 编码 SEQ ID NO:9 蛋白酶的多核苷酸；

[0120] i) 具有 SEQ ID NO:4 序列的多核苷酸；

[0121] j) 编码 NSP25 家族蛋白酶的生物学活性片段的多核苷酸；

[0122] k) 编码 SEQ ID NO:7 的序列及其生物学活性片段的多核苷酸；和

[0123] 1) 具有 SEQ ID NO:3 或 SEQ ID NO:5 序列的多核苷酸。

[0124] 因为遗传密码的简并性，一个以上的密码子可以被用于编码一具体氨基酸。因此，不同的 DNA 序列可以编码与例如 SEQ IDNO:2 多肽具有相同氨基酸序列的多肽。本发明包括编码此同一多肽的多种多核苷酸。

[0125] 当单链形式的核酸能够在适宜的温度和溶液离子强度条件下与其它核酸进行退火时，则该核酸可与另一核酸序列杂交。对于低、中、高以及极高严紧性条件的杂交而言，杂交和清洗条件为本领域所熟知（参见例如 Sambrook(1989)supra，具体是第 9 章和第 11 章）。一般而言，杂交涉及核苷酸探针和同源 DNA 序列，其通过互补多核苷酸广泛的碱基配对形成稳定的双链杂合体（同样参见第 8 章，Gene Cloning，An Introduction，T.A.Brown(1995)Chapman and Hall London)。在一些实施方式中，可以将带有探针和同源序列的滤膜在 2× 氯化钠 / 柠檬酸钠 (SSC)、0.5% SDS 中，在大约 60℃（中等严紧性）、65℃（中 / 高严紧性）、70℃（高严紧性）和大约 75℃（非极高严紧性）下进行洗涤 (CurrentProtocols in Molecular Biology，John Wiley & Sons，New York，1989，6.3.1-6.3.6，在此被并入本文作为参考）。

[0126] 包括在本发明内的是：等位基因变异；天然突变体；诱导突变体；由在高或低严紧性条件下与编码 SEQ ID NO:2、SEQ ID NO:8、SEQ ID NO:9 和 SEQ ID NO:10 的多肽的核酸杂交的 DNA 所编码的蛋白；以及被抗血清特异性结合至具有 SEQ ID NO:2 或 SEQ ID NO:10 的 NSP24 蛋白酶的多肽，尤其是被抗血清特异性结合至 NSP24 蛋白酶的活性位点或结合结构域的多肽。在一些实施方式中，编码本发明的 NSP24 家族蛋白酶的核酸，诸如编码 SEQ ID NO:2 的 NSP24 蛋白酶的核酸，在高严紧性条件下，与对应于来自 SEQ ID NO:8 的至少 12、15 或 20 个连续核苷酸的核酸杂交。

[0127] 本发明的核酸和多肽包括由于所公开序列中的测序错误而不同于本文所公开的序列的那些核酸和多肽。

[0128] DNA 序列的同源性由两 DNA 序列间的同一性程度确定。可以用计算机程序来确定多肽序列或核苷酸序列的同源性或者同一性百分比。进行序列联配以及确定序列同一性的方法是技术人员已知的，可以不经过过多的实验而进行，并且可以明确地计算出同一性的值。参见，例如 Ausubel 等 eds.(1995)Current Protocols in Molecular Biology，Chapter19(Greene Publishing and Wiley-Interscience，New York)；和 ALIGN 程序 (Dayhoff(1978)in Atlas of Protein Sequence and Structure5:Suppl.3(National Biomedical Research Foundation，Washington，D.C.)。有大量用于比对序列和确定序列同一性的算法是可利用的，包括，例如 Needleman 等 (1970)J.Mol.Biol.48:443 的同源性联配算法；Smith 等 (1981)Adv.Appl.Math.2:482 的局部同源性联配算法；Pearson

等 (1988)Proc.Natl.Acad.Sci.85:2444 的相似性搜索法 ;Smith-Waterman 算法 (Meth.Mol.Biol.70:173-187(1997) ; 以及 BLASTP、BLASTN 和 BLASTX 算法 (参见 Altschul 等 (1990)J.Mol.Biol.215:403-410)。使用这些算法的计算机化程序也是可利用的,包括但不限于 :ALIGN 或 Megalign(DNASTAR) 软件、或者 WU-BLAST-2(Altschul 等, Meth.Enzym., 266:460-480(1996)) ; 或者 GAP、BESTFIT、BLAST Altschul 等, supra、FASTA 和 TFASTA, 可获自 Genetics Computing Group(GCG) 包, Version8, Madison, Wis., USA ; 和 PC/Gene 程序中的 CLUSTAL, Intelligenetics, Mountain View, Calif。本领域的技术人员可以确定用于测量联配的合适参数,包括用于在所比较的序列长度内获得最大联配所需的算法。优选地,使用程序确定的默认参数来测定序列同一性。具体而言,序列同一性可以通过 Smith-Waterman 同源性搜索算法加以测定 (Meth.Mol.Biol.70:173-187(1997)),如在 MSPRCH 程序 (OxfordMolecular) 中使用拟合空位搜索 (affine gap search),并使用下列搜索参数所进行的 : 空位开放罚分为 12,空位延伸罚分为 1。优选地,成对的氨基酸比较可以使用 Genetics Computer Group, Inc., Madison, Wis. 的 GCG 序列分析软件包中的 GAP- 程序进行,其使用 blosum62 氨基酸置换矩阵,空位权重为 12,长度权重为 2。关于两氨基酸序列的最适联 配,变体氨基酸序列的连续片段相对于参比氨基酸序列而言可以含有额外的氨基酸残基或缺失的氨基酸残基。被用于与参比氨基酸序列进行比较的连续片段将包括至少 20 个连续的氨基酸残基,并可以是 30、40、50、或更多个氨基酸残基。可以通过指定空位罚分,对衍生物的氨基酸序列中与包含空位有关的增加的序列同一性进行校正。

[0129] 在一些实施方式中,本发明所包括的蛋白酶 (例如与 SEQ IDNO:2 序列具有至少 80% 序列同一性的 NSP24 家族蛋白酶),来源于细菌或真菌诸如丝状真菌。一些优选的丝状真菌包括曲霉属某些种 (Aspergillus spp.) 和木霉属某些种 (Trichoderma spp.)。一种优选的木霉为里氏木霉 (T.reesei)。然而,根据本发明的蛋白酶和 / 或编码蛋白酶的 DNA 可以是来源于真菌,诸如犁头霉属某些种 (Absidia spp.) ; 枝顶孢霉属某些种 (Acremonium spp.) ; 伞菌属某些种 (Agaricus spp.) ;Anaeromyces spp. ; 曲霉属某些种 (Aspergillus spp.), 包括刺孢曲霉 (A.aculeatus)、泡盛曲霉 (A.awamori)、黄曲霉 (A.flavus)、臭曲霉 (A.foetidus)、A.fumaricus、烟曲霉 (A.fumigatus)、构巢曲霉 (A.nidulans)、黑曲霉 (A.niger)、米曲霉 (A.oryzae)、土曲霉 (A.terreus) 和杂色曲霉 (A.versicolor) ; Aeurobasidium spp. ;Cephalosporum spp. ; 毛壳菌属某些种 (Chaetomium spp.) ; 鬼伞属某些种 (Coprinus spp.) ;Dactyllum spp. ; 镰刀菌属某些种 (Fusarium spp.), 包括 F conglomerans、砖红镰刀菌 (F.decemcellulare)、爪哇镰刀菌 (F.javanicum)、亚麻镰刀菌 (F.lini)、尖镰孢菌 (F.oxysporum) 和茄腐镰刀菌 (F.solani) ; 粘帚霉属某些种 (Gliocladium spp.) ; 腐质霉属某些种 (Humicola spp.), 包括特异腐质霉 (H.insolens) 和疏棉状腐质霉 (H.lanuginose) ; 毛霉属某些种 (Mucorspp.) ; 脉孢菌属某些种 (Neurospora spp.), 包括粗糙脉孢菌 (N.crassa) 和好食脉孢霉 (N.sitophila) ; Neocallimastix spp. ;Orpinomyces spp. ; 青霉属某些种 (Penicillium spp.) ; 显革菌属某些种 (Phanerochaete spp.) ;Phlebia spp. ;Piromyces spp. ; 根霉属某些种 (Rhizopus spp.) ; 裂褶菌属某些种 (Schizophyllum spp.) ; 栓菌属某些种 (Trametes spp.) ; 木霉属某些种 (Trichoderma spp.), 包括里氏木霉 (T.reesei)、里氏木霉 (T.reesei)(长枝木霉 (T.longibrachiatum)) 和绿色木霉 (T.viride) ; 和接合霉属某些种 (Zygorhynchus

spp.)。

[0130] 宿主细胞

[0131] 在一些实施方式中,本发明提供了用本文所述的DNA构建物和载体转化的宿主细胞。在一些实施方式中,被导入宿主细胞的编码本发明所包括的蛋白酶(例如与SEQ ID NO:2具有至少95%序列同一性的NSP24家族蛋白酶)的多核苷酸编码异源性蛋白酶,而在其它实施方式中,该多核苷酸编码在宿主细胞中过量表达的内源性蛋白酶。在一些实施方式中,本发明提供了在宿主细胞诸如细菌和真菌宿主细胞中发挥功能的基因启动子控制下,异源性蛋白酶基因的表达或者蛋白酶基因的过量表达。

[0132] 一些优选的宿主细胞包括丝状真菌细胞。丝状真菌宿主细胞的非限定性实例包括木霉属某些种(Trichoderma spp.)(例如绿色木霉(T.viride)和里氏木霉(T.reesei)、红褐肉座菌(Hypocreajecorina)的无性型,以前被分类为长枝木霉(T.longibrachiatum))、青霉属某些种(Penicillium spp.)、腐质霉属某些种(Humicola spp.)(例如特异腐质霉(H.insolens)和灰色腐质霉(H.grisea))、曲霉属某些种(Aspergillus spp.)(例如黑曲霉(A.niger)、构巢曲霉(A.nidulans)、米曲霉(A.orzyae)和泡盛曲霉(A.awamori))、镰刀菌属某些种(Fusarium spp.)(禾赤镰孢菌(F.graminum))、脉孢菌属某些种(Neurospora spp.)、肉座菌属某些种(Hypocrea spp.)和毛霉属某些种(Mucor spp.)。进一步地,宿主细胞可以包括芽孢杆菌属某些种(Bacillus spp.)(例如枯草芽孢杆菌(B.subtilis)、地衣芽孢杆菌(B.licheniformis)、缓慢芽孢杆菌(B.Lentus)、脂肪嗜热芽孢杆菌(B.stearothremophilus)和短小芽胞杆菌(B.brevis))以及链霉菌属某些种(Streptomyces spp.)(例如杂色链霉菌(S.coelicolor)和变铅青链霉菌(S.lividans)(TK23和TK21))。

[0133] 分子生物学

[0134] 本发明依靠重组遗传领域的常规技术。公开本发明中的一般使用方法的基本教科书包括Sambrook等Molecular Cloning,ALaboratory Manual(2nd ed.1989);Kriegler,Gene Transfer and Expression:A Laboratory Manual(1990);和Ausubel等,eds.,Current Protocols in Molecular Biology(1994))。

[0135] 异源基因,包括例如丝状真菌的基因启动子序列,一般在转化进入宿主细胞诸如里氏木霉细胞进行复制和/或表达以前,被克隆进中间载体。这些中间载体一般是原核载体,例如质粒或穿梭载体。

[0136] 为获得克隆基因的高水平表达,异源基因与启动子间的距离优选地同天然存在的基因与启动子间的距离大约相同。然而,如本领域已知,在不损失启动子功能的条件下,对此距离的一些变化是可以接受的。

[0137] 本领域的技术人员应认识到,在不改变其功能的条件下,天然启动子可以通过置换(replacement)、取代、添加或消除一个或多个核苷酸而被修饰。本发明实践包括而不限于对启动子的此类改变。

[0138] 表达载体/构建物一般包含转录单位或表达盒,其包含表达异源序列所需的全部额外的元件。因此,典型的表达盒包含被可操纵地连接于异源核酸序列的启动子和对转录物进行有效的多聚腺苷酸化、核糖体结合位点以及翻译终止所需的信号。盒的其它元件可以包括增强子,且如果基因组DNA被用作结构基因,则还包括携带功能性剪接供体位点和

受体位点的内含子。

[0139] 本发明实践并不受到对遗传构建物中启动子的选择所限制。但是，示例性的启动子为里氏木霉 cbh1、cbh2、egl、eg2、eg3、eg5、xln1 和 xln2 启动子。同样，来自泡盛曲霉和黑曲霉葡糖淀粉酶基因 (glaA) 的启动子 (Nunberg 等 (1984)Mol.Cell Biol.4:2306-2315) 以及来自构巢曲霉乙酰胺酶的启动子可以用于所述的载体。被用于枯草芽孢杆菌所用载体的优选启动子为 AprE 启动子；大肠杆菌所用的优选启动子为 Lac 启动子，酿酒酵母 (Saccharomyces cerevisiae) 所用的优选启动子为 PGK1，黑曲霉 (Aspergillusniger) 所用的优选启动子为 glaA，而里氏木霉 (Trichoderma reesei) 的优选启动子为 cbh1。

[0140] 除启动子序列以外，表达盒应当还包含结构基因的转录终止区下游以提供有效的终止作用。终止区可以从与启动子序列相同的基因中获得，或者可以从不同的基因中获得。

[0141] 尽管任何真菌终止子在本发明中都有可能是有用的，但是一些优选的终止子包括：来自构巢曲霉 trpC 基因 (Yelton，M. 等 (1984) PNAS USA81:1470-1474，Mullaney，E.J. 等 (1985)MGG199:37-45)、泡盛曲霉或黑曲霉葡糖淀粉酶基因 (Nunberg，J.H. 等 (1984)Mol.Cell Biol.4:2306，Boel，E. 等 (1984)EMBO J.3:1581-1585)、米曲霉 TAKA 淀粉酶基因和米氏毛霉 (Mucor miehei) 羧基蛋白酶基因 (EPO 公布号 .0215594) 的终止子。

[0142] 用于将遗传信息传送入细胞的具体表达载体并非特别关键。用于在真核或原核细胞表达的任何常规载体都可以被使用。标准的细菌表达载体包括噬菌体 λ 和 M13，以及质粒诸如基于 pBR322 的质粒、pSKF、pET23D，和融合表达系统诸如 MBP、GST 和 LacZ。也可以给重组蛋白添加表位标签，以提供简便的分离方法，例如 c-myc。适当的表达和 / 或整合载体实例在 Sambrook 等 (1989)supra、Bennett 和 Lasure(Eds.)More Gene Manipulations in Fungi，(1991)Academic Press pp.70-76 与 pp.396-428 及其所引用的文章；USP5,874,276 以及真菌遗传保藏中心菌株目录 (Fungal Genetic Stock Center Catalogue ofStrains) (FGSC，www.fgsc.net.) 中提供。可以从 Promega 和 Invitrogen 获得有用的载体。一些具体的有用载体包括 pBR322、pUC18、pUC100、pDON™201、pENTR™、pGEN®3Z和pGEN®4Z。然而，本发明意欲包括其它形式的表达载体，其具有发挥着等效的功能并且已被或正在被本领域所知晓。因此，大量的各种宿主 / 表达载体的组合可以被用于表达本发明的 DNA 序列。有用的表达载体，例如，可以包括染色体的、非染色体的和合成的 DNA 序列的片段，诸如各种已知的 SV40 衍生物和已知的细菌质粒，例如来自大肠杆菌的质粒，包括 col E1、pCR1、pBR322、pMb9、pUC19 以及它们的衍生物、宿主范围更广的质粒，例如 RP4、噬菌体 DNA 例如大量的 λ 噬菌体的衍生物，例如 NM989，以及其它 DNA 噬菌体例如 M13 和丝状单链 DNA 噬菌体、酵母质粒诸如 2μ 质粒或其衍生物。

[0143] 在一些实施方式中，表达载体包括选择性标记。选择性标记的实例包括赋予其抗微生物抗性的标记。营养型标记也可用于本发明中，包括那些本领域已知的标记如 amdS、argB 和 pyr4。用于转化木霉的有用标记是本领域已知的 (参见例如 Finkelstein，chapter6，inBiotechnology of Filamentous Fungi，Finkelstein 等，EDS Butterworth-Heinemann，Boston MA(1992) 和 Kinghorn 等，(1992)AppliedMolecular Genetics of Filamentous Fungi，Blackie Academic andProfessional，Chapman and Hall，London)。在一些实施方式中，表达载体还将包括复制元、编码抗生素抗性的基因，这使得能够对包容重组质粒的细菌进行选择，以及包括质粒非必需区中独特的限制性位点，

以使得能够插入异源序列。所选的具体抗生素抗性基因并非关键性的，本领域已知的许多抗性基因中的任何一个都是适用的。优选地选择原核序列，这样，它们便无法干扰 DNA 在里氏木霉中的复制和整合。

[0144] 本发明的转化方法可以导致将全部或部分转化载体稳定整合入宿主细胞诸如丝状真菌宿主细胞的基因组中。但是，还考虑的转化作用是，其导致了自我复制的染色体外转化载体的维持。

[0145] 许多标准的转染方法可以被用于产生表达大量蛋白酶的细菌或丝状真菌（例如曲霉属或木霉属）细胞系。一些公布的将 DNA 构建物导入产纤维素木霉属菌株的方法包括 Lorito，Hayes，DiPietro 和 Harman，(1993)Curr. Genet. 24 :349-356 ；Goldman，VanMontagu 和 Herrera-Estrella，(1990)Curr. Genet. 17:169-174 ；和 Penttila，Nevalainen，Ratto，Salminen 和 Knowles，(1987)Gene6 :155-164，也参见 USP6,022,725 ；USP6,268,328 和 Nevalainen 等，" The Molecular Biology ofTrichoderma and its Application to the Expression of Both Homologousand Heterologous Genes" in Molecular Industrial Mycology，Eds，Leongand Berka，Marcel Dekker Inc.，NY(1992)pp129-148 ；对于曲霉属而言，包括 Yelton、Hamer 和 Timberlake，(1984)Proc. Natl. Acad. Sci. USA81 :1470-1474，对于镰刀菌属而言，包括 Baj ar，Podila 和 Kolattukudy，(1991)Proc. Natl. Acad. Sci. USA88 :8202-8212，对于链霉菌属而言，包括 Hopwood 等，1985，Genetic Manipulation of Streptomyces :LaboratoryManual，The John Innes Foundation，Norwich，UK 和 Fernandez-Abalos 等，Microbiol149:1623-1632(2003)，而对于芽孢杆菌属而言，包括 Brigidi，DeRossi，Bertarini，Riccardi 和 Matteuzzi，(1990)FEMS Microbiol. Lett. 55:135-138。

[0146] 然而，将外源核苷酸序列导入宿主细胞的任何熟知的方法都可以被使用。这些方法包括使用磷酸钙 - 转染法、polybrene 法、原生质 体融合、电穿孔、生物射弹法 (biolistics)、脂质体、显微注射、原生质载体、病毒载体以及任何其它熟知的将克隆的基因组 DNA、cDNA、合成 DNA 或其它外源遗传物质导入宿主细胞的方法（参见例如 Sambrook 等，supra)。同样使用的是美国专利号 6,255,115 中所述的农杆菌介导的转染法。唯一必不可少的是所用的具体遗传工程方法应当是能够将至少一种基因成功地导入能够表达该基因的宿主细胞中。在一种实施方式中，本发明涉及一种生产本发明所包括的蛋白酶（例如 NSP24 家族蛋白酶）的方法，其包括向宿主细胞中导入包含启动子的多核苷酸，该启动子被可操纵地连接于编码蛋白酶诸如 NSP 家族蛋白酶的核酸 ；在适于蛋白酶表达和产生的合适培养条件下培养该宿主细胞 ；并产生所述的蛋白酶。在一些优选的实施方式中，蛋白酶是与 SEQ ID NO :2 或 SEQ ID NO :10 具有至少 95％序列同一性的 NSP24 家族蛋白酶或者其生物学活性片段。

[0147] 当表达载体被导入细胞后，转染的细胞或转化的细胞被培养在适于基因表达的条件下，并受到蛋白酶基因启动子序列的控制。大批量的转化的细胞可以如下文实施例 3 中所述进行培养。最后，使用标准技术从培养物中回收产物。

[0148] 因此，本发明在此提供了所需多肽的表达和增强的分泌，该多肽的表达是在基因启动子序列的控制之下，所述基因启动子序列包括天然存在的蛋白酶基因、融合 DNA 序列以及多种异源构建物。本发明还提供了表达和分泌高水平的此类所需多肽的方法。

[0149] 蛋白表达

[0150] 本发明的蛋白质是由用载体诸如含有基因的表达载体转化的培养细胞产生的，所述基因的表达是在基因启动子序列控制之下。本发明对于增强蛋白质诸如本发明所包括的蛋白酶的细胞内和/或细胞外产生特别有用。蛋白质可以是同源的或者异源的。对于所述基因表达的适宜条件包括向培养物中提供本发明的诱导供给组合物(inducingfeed composition)。生产该蛋白的最适条件将随对宿主细胞的选择以及随对将要表达的蛋白酶蛋白的选择而变化。通过常规的实验和优化作用，本领域的技术人员将会容易地确定此类条件。

[0151] 感兴趣的蛋白酶蛋白可以在表达后被分离或回收并纯化。根据样品中存在的其它组分，感兴趣的蛋白可以按本领域技术人员已知的各种方法加以分离或纯化。标准的纯化方法包括电泳技术、分子技术、免疫技术和层析技术，所述层析技术包括离子交换层析、疏水层析、亲合层析以及反相HPLC色谱和色谱聚焦。例如，感兴趣的蛋白可以用标准的抗感兴趣蛋白的抗体柱加以纯化。结合蛋白质浓缩，超滤和透析技术也是有用的。合适纯化技术的一般指南，参见Scopes，Protein Purification(1982)。必需的纯化程度将根据感兴趣蛋白质的用途而变化。在一些实例中，将不必进行纯化。

[0152] 细胞培养

[0153] 宿主细胞和转化细胞可以培养在常规的营养培养基中。转化的宿主细胞的培养基可以根据适合于活化启动子以及选择转化体来进行修改。具体的培养条件诸如温度、pH及类似的条件可以是用于被选择用来表达的宿主细胞的那些条件，并且对于本领域的技术人员而言将是显而易见的。此外，优选的培养条件可以在科学文献诸如Sambrook，(1982)supra；Kieser，T，MJ.Bibb，MJ.Buttner，KF Chater，and D.A.Hopwood(2000)PRACTICAL STREPTOMYCES GENETICS.JohnInnes Foundation，Norwich UK；Harwood. 等，(1990)MOLECULARBIOLOGICAL METHODS FOR BACILLUS，John Wiley中和/或从美国典型培养物保藏中心(ATCC；www.atcc.org)中找到。真菌宿主细胞诸如木霉细胞的稳定转化体一般可以根据其较快的生长速度或者在固体培养基上形成带有光滑而非粗糙轮廓的圆形克隆而区别于非稳定转化体。

[0154] 表达多肽的回收和纯化蛋白酶的方法

[0155] 本发明所包括的多肽，诸如由转化宿主细胞所产生的与SEQ IDNO：10具有至少80％序列同一性的多肽，可以从培养基中回收，这通过多种常规方法进行，包括通过离心或过滤将宿主细胞从培养基中分离出来，或者如有必要，裂解细胞并从细胞级分或细胞残骸移取上清。在一些情况下，经过澄清之后，上清或滤液中的蛋白质组分利用盐例 如硫酸铵进行沉淀。然后，将沉淀的蛋白质溶解并可以通过各种层析方法例如离子交换层析、凝胶过滤层析、亲合层析以及其它本领域所认知的方法进行纯化。肽和蛋白质的抗体可以通过免疫动物例如兔或小鼠来进行制备，并通过现有技术的方法回收抗-NSP24蛋白酶抗体。

[0156] 用于本发明的试验包括，但不限于WO9934011和USP6,605,458中所述的那些试验。

[0157] 组合物和应用

[0158] 在一些实施方式中，本发明涉及包含如此处所述的本发明的蛋白酶的组合物。在根据本发明的组合物以及应用中有用的蛋白酶的一些非限制性实例包括例如NSP24家族

蛋白酶或者NSP25家族蛋白酶，更具体地与SEQ ID NO:2具有至少85%序列同一性的NSP24家族蛋白酶或者其生物学活性片段，诸如与SEQ ID NO:10序列具有至少90%序列同一性的蛋白酶。在一些实施方式中，酶组合物为单组分蛋白酶组合物。在一些实施方式中，本发明涉及在工业和商业应用中使用本发明的蛋白酶的方法。组合物与工业应用的下列描述意欲为示例性的而非包括性的。

[0159] 包括本发明蛋白酶的组合物可以进一步包括其它酶类，诸如，但不限于，葡糖淀粉酶、α 淀粉酶、颗粒淀粉水解酶、纤维素酶、脂酶、木聚糖酶、角质酶、半纤维素酶、氧化酶以及它们的组合。

[0160] 在一些优选的实施方式中，组合物将包括本发明的蛋白酶和葡糖淀粉酶，所述蛋白酶与SEQ ID NO:10序列具有至少85%的序列同一性。葡糖淀粉酶可以是从丝状真菌来源诸如曲霉属、木霉属或根霉属菌株获得的野生型葡糖淀粉酶，或者葡糖淀粉酶可以是蛋白质工程化葡糖淀粉酶，诸如黑曲霉葡糖淀粉酶的变体。在其它优选的实施方式中，组合物将包括本发明的蛋白酶和 α 淀粉酶。在一些实施方式中，α 淀粉酶可以从细菌来源诸如芽孢杆菌属某些种或者从真菌来源诸如曲霉属某些种获得。在一些实施方式中，组合物可以包括根据本发明的蛋白酶以及葡糖淀粉酶和 α 淀粉酶。这些酶类的商业来源是已知的，并且可从例如Genencor International，Inc. 和Novozymes A/S获得。

[0161] 在若干实施方式中，本发明已考虑用于乙醇生产、烘焙、果 汁生产、酿造、蒸馏、酒类制造、皮革、油类与脂肪、纸与浆以及动物饲料生产。

[0162] 在其它实施方式中，本发明被考虑作为去污剂和清洁产品的活性“生物学”组分。在此，蛋白酶、淀粉酶和脂酶被用于分解蛋白质、淀粉和脂肪污渍。本发明的实施方式包括通过进行稳定性研究测试酶类与去污剂组分的相容性以及在多种配方中测试它们。

[0163] 仍在另一实施方式中，本发明考虑了将淀粉液化并糖化为葡萄糖并异构化为果糖的酶促用途。本发明可以被用于将大体积的植物底物，诸如粮食（例如玉米、小麦、蜀黍、黑麦及类似物）转变为增甜剂，如高果糖玉米糖浆和麦芽糖糖浆。

[0164] 本发明的酶（一种或多种）可应用于食品和饲料工业以提高蛋白质的可消化性。蛋白酶还被用于多种工业应用方面，具体而言，用于纺织品、平版印刷、化学工艺、农业、环境废弃物转化、生物浆加工、生物量转化为燃料以及其它化学方法（一种或多种）。进而，该蛋白酶被应用于保健及个人护理产品诸如化妆品、皮肤护理品、牙膏及类似物。

[0165] 饲料

[0166] 本文所述的本发明酶被用于动物饲料。所述饲料可以包括植物材料诸如玉米、小麦、高粱、大豆、油菜、向日葵以及任何此类植物材料的混合物或者植物蛋白来源，用于家禽、猪、反刍动物、水产养殖和宠物。应当考虑，各种性能参数诸如生长、饲料摄取和饲料效率，以及改善的均匀性、动物棚舍中降低的氨浓度、随之提高的动物福利和健康状况都将得到改善。

[0167] 食品

[0168] 食用蛋白质水解产物代表了一个小而重要的市场部分。此类制品被用于术后患者或用于消化系统损伤的个体。此水解产物可以作为本身相对粗的制品被施用(Clegg，1978In″ Biochemical Aspects ofNew Protein Food″，J，Adler-Nissen，B，O，Eggum，L，Munck & H.S.Olseneds.，p.109-117，Pergamon，Oxford)，或者作为高度纯化的氨基酸混合

物用于静脉内施用。乳蛋白的酶水解产物已被用作食用制品。

[0169] 瘦肉食品，特别是肉类的酶促嫩化，代表了一个巨大的市场部分，其目前由植物蛋白酶和某些微生物酶类所主导。鱼类瘦肉的酶促成熟和嫩化在许多国家也是相当重要的。因此，现在所描述的酶类被用于食品的多种用途中。

[0170] 进一步，本发明的酶或酶组合物可以用于由例如植物蛋白如大豆、豌豆、羽扇豆或油菜籽蛋白、奶如酪蛋白、肉类蛋白或鱼蛋白来制作蛋白质水解产物。此处所述的酶（一种或多种）可以被用于蛋白质水解产物以改善溶解性、均一性或可发酵性，减少抗原性、减少水解产物的苦味，或者用于制造食品、饲料或医药产品的其它目的。此处所述的酶（一种或多种）可以被单独使用或者与其它肽酶一起或者与其它酶类如外切肽酶一起使用。此处所述的酶（一种或多种）与富含外切肽酶的酶制剂一起使用，将会改善蛋白质水解产物的口味。

[0171] 而且，该酶或酶组合物可以被用于鱼或肉的加工，例如改变质地和 / 或粘性。

[0172] 皮革

[0173] 工业皮革制造有赖于涉及皮革的清洁、脱毛以及最终的鞣制和干燥的一系列步骤。酶处理在脱毛步骤中发挥了重要的作用，该步骤是通过应用蛋白水解酶，如本发明的肽水解酶而实现的；由于它们具有高蛋白水解活性和在低 pH 下具有功效，从而能够提供皮革制造中目前所用的哺乳动物蛋白酶的有效替代品。

[0174] 毛料与丝绸

[0175] 此处所述的蛋白酶被用于工业处理毛料制品以赋予所期望的特性。在一种实施方式中，本发明提供了处理纺织品的组合物。该组合物可以被用于处理例如毛料与丝绸（参见例如 RE216,034；EP134,267；US4,533,359；和 EP344,259）。

[0176] 本发明的方法可以被用于处理含蛋白的纤维，例如角蛋白纤维。其适于处理毛料、毛料纤维或动物毛发，诸如安哥拉山羊毛 (angora)、马海毛 (mohair)、开司米毛 (cashmere)、alpacca 或其它商业上有用的动物毛发产品，其可以来源于绵羊、山羊、羊驼、骆驼、兔等。 蚕丝、蜘蛛丝或人类毛发也可以用本发明的方法处理。所述纤维可以是纤维或顶绒 (top)、纱线 (yarn) 或者纺织品或针织品或服装的形式。

[0177] 清洁

[0178] 本发明还涉及含有本发明蛋白酶（一种或多种）的清洁组合物。该清洁组合物可以额外含有通常用于清洁组合物的多种添加剂。它们可以选自，但不限于漂白剂、表面活性剂、助洗剂、酶和漂白催化剂。对于本领域的普通技术人员而言，什么添加剂适合于包含在组合物中将会是显而易见的。此处所提供的清单并非详尽，而应当被视为合适添加剂的实例。对于本领域的普通技术人员而言，仅使用那些与组合物中的酶和其它组分相容的添加剂，例如表面活性剂，也将会是显而易见的。

[0179] 蛋白质，具体而言是本发明的那些蛋白质，可以按重量被配制成大约 0.01 至大约 5％（优选地 0.1％至 0.5％）水平的已知的粉末去污剂和液体去污剂，其具有 3.5 至 7.0 之间的酸性 pH。在一些实施方式中，此类清洁组合物进一步包括其它酶类诸如淀粉酶、其它蛋白酶、纤维素酶、脂酶或内切葡萄糖苷酶、以及助洗剂和稳定剂。在一些实施方式中，pH 是在 4.0 至 6.5 之间，优选地在 4.0 至 5.6 之间。根据所需的活性水平，尽管由于其最适 pH 的缘故，这些酶被称为酸性蛋白酶，但是在 pH7-9 下使用这些酶也是有可能的。

[0180] 将蛋白质加入常规清洁组合物并没有任何特殊的使用限制。换言之，只要 pH 在以上范围内，而温度在所述蛋白质变性温度以下，适于去污剂的任何温度与 pH 也都适于本组合物。此外，本发明的蛋白质被用在无去污剂的清洁组合物中，再者单独使用或者与助洗剂和稳定剂组合使用。

[0181] 蛋白质加工

[0182] 蛋白质原材料的酶促水解经常导致苦味肽 (bitter peptides) 的形成 (Clegg，1978)。存在于蛋白水解产物中的苦味肽可能代表了相当多的实际问题，如这样的情况，例如在不同类型奶酪的熟化过程中以及在食用蛋白质水解产物的生产过程中。水解产物的苦味通常是由于特定 的肽，且尤其是含有高比例疏水氨基酸的那些肽引起的。通过苦味肽的完全或者部分水解可以有效地减少苦味。因此，此处所述的酶类可用于去除食品的苦味。本发明的酶或酶组合物可以被用于减少食物中蛋白质和 / 或蛋白质水解产物的苦味。

[0183] 根据本发明还考虑到从蛋白质和 / 或蛋白质水解产物生产游离氨基酸。当游离氨基酸为谷氨酸的情况下，其强化了食品的风味。

[0184] 所述蛋白质或蛋白质水解产物可以是动物或植物来源的。在本发明的实施方式中，将要被水解的蛋白是酪蛋白或大豆蛋白。

[0185] 该蛋白可以被用于生产食品诸如奶酪和含可可的食品。

[0186] 尽管此处所述的酶（一种或多种）以及富集了本发明酶的酶制品在用于生产无苦味的蛋白质或蛋白质水解产物中可能特别是有利的，但是此处所述的酶（一种或多种）还可以被用于大量工业应用，包括含蛋白质物质诸如细胞壁的降解或改性。一些蛋白质如伸展蛋白是植物细胞壁的组分。因此，此处所述的酶（一种或多种）将有助于植物细胞壁的降解和改性。

[0187] 本发明酶制品的剂量以及使用该制品时的其它条件可以根据本领域已知的方法加以确定。

[0188] 蛋白质沉淀物在某些产品诸如例如啤酒中也提出了相当大的问题，原因在于沉淀导致产品混浊。在啤酒的冷却储藏期间，当可溶性蛋白沉淀时，会在啤酒中产生混浊。这一问题有着相当大的经济重要性，而且除挑选合适的原材料用于制造啤酒外，目前避免此问题的主要方法是向啤酒中加入蛋白水解酶。

[0189] 个人护理

[0190] 在一些实施方式中，一旦此处所述的蛋白酶被合成并纯化，则将有效的量加入个人护理组合物（一种或多种），其可用于个人护理产品。个人护理产品可以被划分 / 描述为化妆品、非处方（“OTC”）化合物，其可用于个人护理应用（例如化妆品、皮肤护理、口腔护理、头发护理、指甲护理）。在一些实施方式中，此处所述的蛋白酶被加入个人护理组合物诸如头发护理组合物、皮肤护理组合物、指甲护理组合物、化妆品组合物或者它们的任何组合。因此，可以在例如用于清洗隐形 眼镜的溶液、牙膏、化妆品和皮肤护理产品中使用该酶或酶制品。

[0191] 增甜剂

[0192] 此处所述的蛋白酶可用于生产高麦芽糖或高果糖糖浆以及其它增甜剂。含可发酵糖类或可被转化为糖类的组分的原材料通常为含淀粉的植物材料，包括但不限于粮食作物（例如玉米、小麦、蜀黍、燕麦和黑麦）的块茎、根、茎、穗和谷粒以及含糖原材料诸如甜菜、

甘蔗、水果材料和糖蜜。

[0193] 益生素（调节肠道微生态平衡用的寡糖类等有机物，Prebiotics）

[0194] 酶制品可以用于从蛋白质生产肽，其中使用基本无其它蛋白水解活性的克隆的酶是有利的。

[0195] 通过使用此处所描述的酶（一种或多种）（例如纯化的酶）来水解合适的蛋白来源，产生高度适于作为微生物底物的游离氨基酸和肽的粗制品是可能的，所述微生物对用于生长的氨基酸有着特殊的要求。

[0196] 这是用在工业发酵中的相当大量的微生物的情况。必需氨基酸的供应通常代表了此类发酵过程中的流程节约的一项重要因素。通过应用酶所产生的氨基酸制品不但适合用作实验室的底物而且适合用作大规模工业发酵的底物。

[0197] 此处所述的酶（一种或多种）也可以被用于功能肽、益生素和类似物的原位产生。术语“益生素”指食品或饲料组分，其通过选择性地刺激消化道内，优选地为结肠内的一种或有限数量的细菌的生长和 / 或活性而对宿主产生有益的影响。

[0198] 发酵与生物乙醇

[0199] 使用本发明的蛋白酶组合物由含淀粉底物的发酵来生产乙醇，可以包括生产燃料酒精或便携酒精 (portable alcohol)。在一些实施方式中，酶组合物也可以被用于促进燕麦、麦芽以及其它原材料的酵母发酵以生产例如啤酒。

[0200] 淀粉酶是酿造和烘焙工业的基本酶类。在麦芽制备过程中以及在缺乏添加的糖类或其它碳水化合物的条件下进行的某些烘焙过程 中，需要淀粉酶来降解淀粉。获得此类酶的充分活性是成问题的，尤其在麦芽制备工业。一种用生理学可接受的系统充分提高淀粉酶活性的方法，导致产生了更加迅速的麦芽制备方法，并且由于提高的糖类可利用性，从而导致产生了具有降低的碳水化合物含量的酒精饮料诸如啤酒。

[0201] 在一些实施方式中，含淀粉底物诸如谷粒（例如玉米、小麦和高粱）、穗以及其它植物残留物的水解将会产生酒精诸如乙醇。酒精的生产方法在 The Alcohol Textbook，A Reference for the Beverage，Fueland Industrial Alcohol Industries，3rd Ed.，Eds.，K.A.Jacques 等，(1999)Nottingham University Press，UK 中有所描述。在本发明的一些实施方式中，蛋白酶将与葡糖淀粉酶以及任选的 α 淀粉酶一起以组合物方式被应用于合并的糖化发酵步骤，也被称为同步糖化和发酵。同样参考了 Chapter2.1，Fermentation Alcohol，S.Lewis in Industrial Enzymology，2nd.Ed.Eds.，T.Godfrey 和 S.West，(1996) Stockton Press，NY。使用酸性真菌蛋白酶与葡糖淀粉酶相组合通过发酵生产乙醇的方法是已知的。

[0202] 例如，USP5,231,017 公开了使用来源于黑曲霉的蛋白酶生产乙醇的方法，包括获得液化的醪液并在糖化步骤期间将蛋白酶引入此液化醪液中，所述糖化步骤可以与发酵步骤合并在一起。在一些实施方式中，本发明的蛋白酶组合物将被用于在非熟化过程中用颗粒淀粉底物生产醇类例如乙醇，其中该过程是在低于用于生产醇类的底物中淀粉的糊化温度的温度下进行。而淀粉水解过程中所用的蛋白酶的数量将决定于蛋白酶的酶促活性。在一些实施方式中，该量将是加入到 450g 浆液中的范围为 0.001 至 2.0ml 的 2％蛋白酶溶液，所述浆液被调节至 20-33％干固体，其中该浆液在糖化作用期间和 / 或在水解的淀粉中为液化的醪。其它有用的范围包括 0.005 至 1.5ml 以及还有 0.01 至 1.0ml。

[0203] 用蛋白酶处理种子和谷粒为通过发酵过程生产麦芽和饮料提供了有利条件。

[0204] 同样期望在糖化作用期间使用蛋白酶以便水解面粉中的蛋白质，并因此在对随后的醇类发酵阶段的预期中富集含有可溶性氮的麦芽汁。谷物中增强的淀粉酶活性提高了发芽的速度和效率，这在麦芽制备过程中是重要的，其中生产的麦芽具有提高的酶促活性，从而导 致淀粉被强化水解为可发酵的碳水化合物，因而，改善了醇类饮料例如啤酒和苏格兰威士忌生产过程中的发酵效率。

[0205] 在以下的实验公开内容中，使用了下列缩略语 :eq（当量）;M（摩尔浓度）;μM（微摩尔浓度）;N（正常的）;mol（摩尔）;mmol（毫摩尔）;μmol（微摩尔）;nmol（纳摩尔）;g（克）;mg（毫克）;kg（千克）;μg（微克）;L（升）;ml（毫升）;μl（微升）;cm（厘米）;mm（毫米）;μm（微米）;nm（纳米）;℃（摄氏度）;h（小时）;min（分钟）;sec（秒）;msec（毫秒）;Ci（居里）;mCi（毫居里）;μCi（微居里）;TLC（薄层色谱）;Ts（甲苯磺酰基）;Bn（苯甲基）;Ph（苯基）;Ms（甲磺酰基）;Et（乙基）、Me（甲基）、ds 或 DS（干固体含量）、SAPU（酸性蛋白酶分光光度单位，其中 1SAPU 是在试验条件下每分钟从酪蛋白底物释放 1 微摩尔酪氨酸的蛋白酶活性的量）以及 GAU（葡糖淀粉酶单位，其被定义为在 pH4.2 和 60℃条件下每小时从可溶性淀粉底物产生以葡萄糖计算的 1g 还原糖的酶量）。

[0206] 实施例

[0207] 下面的实施例进一步详细描述了本发明，这些实施例的意图并不在于以任何方式限制要求保护的本发明的范围。附图被认为是本发明的说明书和描述的一个完整部分。所有引用的参考文献在此特意地并入作为参考，用于在此描述的所有内容。下列实施例被提供用于举例说明所要求保护的发明，而并非限定所要求保护的发明。

[0208] 实施例 1

[0209] 新型蛋白酶 NSP24 的里氏木霉 DNA 克隆

[0210] 基因组 DNA 是从里氏木霉菌株 QM6a 中提取的。根据在里氏木霉基因组的叠连群 1-5500 中所发现的推定的蛋白酶序列，设计 PCR 引物 (Joint Genome Institute(JGI) T.reesei genome v1.0)。正向引物包含用于定向克隆入 pENTR/D 载体 (Invitrogen) 的基序。

[0211] afp6f 引物的序列为 CACCATGCAGACCTTTGGAGCT(SEQID NO:11)，而 afp7r 引物的序列为 TTATTTCTGAGCCCAGCCCAG(SEQ ID NO:12)。根据 Invitrogen Gateway 系统方案，通过凝胶提取方 法来纯化 1.3kb 的 PCR 产物（凝胶纯化试剂盒 (Gel Purification kit)，Qiagen)，并克隆进 pENTR/D。

[0212] 然后，将载体转化进带有卡那霉素选择标记的化学感受态 Top10 大肠杆菌 (Invitrogen) 中。将来自若干独立克隆的质粒 DNA 用限制性酶消化，从而确认正确大小的插入片段。对来自若干克隆的蛋白酶基因插入片段进行测序 (Sequetech,Mountain View，CA)。将来自一个克隆的质粒 DNA，pENTR/D_55.3 与 pTrex3g/amdS 目的载体 DNA 加入 LR 克隆酶反应 (Invitrogen Gateway system) 中。pTrex3g 载体基于大肠杆菌 pSL1180 (Pharmacia Inc.，NJ)，其为 pUC118 噬菌粒基载体并在 WO05/001036 中有所描述。在 LR 克隆酶反应中进行的重组，用来自 pENTR/D_55.3 的里氏木霉蛋白酶置换目的载体的 CmR 和 ccdB 基因。此重组将蛋白酶定向插入目的载体的 cbh1 启动子和终止子之间。44 和 50bp 的重组位点序列分别保留在蛋白酶基因的上游和下游。将 LR 克隆酶反应的等份试样转化进化学感受态 Top10

大肠杆菌中并过夜生长，用羧苄青霉素选择。将来自若干克隆的质粒 DNA 用限制性酶消化以确认正确的插入片段大小。将来自克隆的质粒 DNA pTrex3g_55.3.1 用 XbaI 消化以释放出包括 cbh1 启动子 :NSP24 蛋白酶：终止子 :amdS 的表达盒。通过琼脂糖凝胶提取法用标准技术纯化此 5.8kb 盒，并将其转化进里氏木霉菌株，其来自于公众可获得的菌株 QM6a（参见 WO05/001036）。参考图 5、6 和 7。

[0213] 实施例 2

[0214] 新型蛋白酶 NSP25 的里氏木霉 DNA 克隆

[0215] 基因组 DNA 是从里氏木霉菌株 QM6a 中提取的。根据在里氏木霉基因组 (JGI T.reesei genome v1.0) 的叠连群 22-263400 中所发现的推定的蛋白酶序列，设计 PCR 引物。正向引物包含用于定向克隆入 pENTR/D 载体 (Invitrogen) 的基序。

[0216] afp8f 引物的序列为 CACCATGCAGCCCTCATTTGGCAG(SEQIDNO:13)，而 afp9r 引物的序列为 CTATTTCTTCTGCGCCCAGCCAAC(SEQ ID NO:14)。根据 InvitrogenGateway 系统方案，通过凝胶提取法来纯化 1.2kb 的 PCR 产物（凝胶 纯化试剂盒 (Gel Purification kit)，Qiagen）并克隆进 pENTR/D。然后，将载体转化进带有卡那霉素选择标记的化学感受态 Top10 大肠杆菌 (Invitrogen) 中。将来自若干独立克隆的质粒 DNA 用限制性酶消化，从而确认正确大小的插入片段。对来自若干克隆的蛋白酶基因插入片段进行测序 (Sequetech，Mountain View，CA)。将来自一个克隆的质粒 DNA，pENTR/D_22.2 与 pTrex3g/amdS 目的载体 DNA 一起加入 LR 克隆酶反应 (Invitrogen Gateway system) 中。在 LR 克隆酶反应中进行的重组，用来自 pENTR/D_22.2 的里氏木霉蛋白酶置换目的载体中的 CmR 和 ccdB 基因。此重组将蛋白酶定向插入目的载体的 cbh1 启动子和终止子之间。44 和 50bp 的重组位点序列分别保留在蛋白酶基因的上游和下游。将 LR 克隆酶反应的等份试样转化进化学感受态 Top10 大肠杆菌中并过夜生长，用羧苄青霉素选择。将来自若干克隆的质粒 DNA 用限制性酶消化，以确认正确的插入片段大小。将来自克隆的质粒 DNApTrex3g_22.2#1 用 XbaI（并用 EcoRI 消化以便将细菌骨架消化成小片段，其在电泳过程中与盒迁移开来）消化，以释放出包括 cbh1 启动子 :NSP25 蛋白酶：终止子 :amdS 的表达盒。通过琼脂糖凝胶提取法用标准技术纯化此 5.7kb 盒，并将其转化进里氏木霉菌株，该菌株来自于公众可获得的菌株 QM6a。除 NSP24 插入片段用 NSP25 序列替换外，用于转化的质粒与图 7 中举例说明的质粒基本相同。

[0217] 实施例 3

[0218] 木霉的 PEG 真菌转化

[0219] 将来自形成孢子的菌丝体的平板的 $2cm^2$ 琼脂塞 (agar plug) 接种于 250ml4- 挡板摇瓶中的 50ml YEG 肉汤中，并以 200rpm 在 37℃培养 16-20 小时。通过将液体体积转移入 50ml 锥形管中并在 2500rpm 下离心 10 分钟来回收菌丝体。吸去上清。将菌丝体沉淀转移入装有 40ml 过滤除菌的 β-D- 葡聚糖酶 (InterSpex Products，Inc.) 溶液的 250ml0.22μm CA Coming 过滤瓶中，并在 200rpm，30℃下温育 2 小时。通过无菌 Miracloth(CalBiochem，LaJolla，CA)，将菌丝体收集入 50ml 锥形离心管中，在 2000rpm 下离心 5 分钟，抽气。将沉淀用 50ml1.2M 山梨醇洗涤一次，再次离心，抽气，并用 25ml 山梨醇 /$CaCl_2$ 洗涤。将 原生质体用血细胞计数器计数、离心、抽气，并重悬于山梨醇 /$CaCl_2$ 中，其体积足以产生 1.25×10^8/ml 的原生质体浓度。每一转化反应使用 200μl 的等份试样。将 20μg DNA(>1μg/ul) 装

入 15ml 锥形管，并将锥形管放置于冰上。加入 200μl 原生质体。加入 50μl PEG 混合物，并轻轻混匀，在冰上温育 20 分钟。将 2ml PEG 混合物加入锥形管中，并在室温下温育 5 分钟。将 4ml 山梨醇 /$CaCl_2$（总共 6.25ml）加入到锥形管中。此转化混合物被分为 3 等份，每一覆盖物 (overlay) ～ 2ml。将 2ml 加入到熔化的乙酰胺山梨醇顶层琼脂的试管中，并将覆盖物混合物 (overlay mixture) 倒在乙酰胺山梨醇平皿上，用于选择能够依靠乙酰胺作为唯一的氮源来生长的转化体。将平皿在 28-30℃温育，直至克隆出现。通过在乙酰胺培养基（无山梨醇的乙酰胺山梨醇配方）上对单个克隆重复传代，转化体得以纯化。

[0220] 材料

[0221] 40ml β-D- 葡聚糖酶溶液：600mg β-D- 葡聚糖酶；400mg$MgSO_4 \cdot 7H_2O$ 和 40ml1.2M 山梨醇。

[0222] 200ml PEG 混合物：50g PEG4000(BDH Laboratory SuppliesPoole, England) 和 1.47g $CaCl_2 \cdot 2H_2O$，在 MilIi Q 水中制成。

[0223] 山梨醇 /$CaCl_2$：1.2M 山梨醇和 50mM $CaCl_2$

[0224] 为了进行 amdS 选择，使用乙酰胺山梨醇平皿和覆盖物。为对孢子进行纯化，使用同样的平皿，但不含山梨醇。

[0225] 乙酰胺山梨醇琼脂（培养板和顶层琼脂）

[0226] 乙酰胺 (Aldrich99％经升华的)0.6g/L；CsCl1.68g/L；葡萄糖 20g/L；$KH_2PO_4$20g/L；$MgSO_4 \cdot 7H_2O$0.6g/L；CaCl2 • 2H200.6g/L；1000× 盐（参见下文）1ml。将 pH 调至 5.5 并将体积定容至 300ml。用 0.22 微米滤膜过滤除菌并在烘箱中加热至 55℃。

[0227] 向 700ml 水中加入 Noble 琼脂（低熔点用于顶层琼脂）20g 和山梨醇 218g，然后高压灭菌。将此混合物冷却至 55℃，并加入滤膜除菌的乙酰胺混合物。倒入平皿和试管。

[0228] 1000× 盐 ——$FeSO_4 \cdot 7H_2O$(0.5g/100ml)；$MnSO_4 \cdot H_2O$(0.16g/100ml)；$ZnSO_4 \cdot 7H_2O$(0.14g/100ml)；$CoCl_2 \cdot 6H_2O$(0.1g/100ml)，并用 0.22 微米滤膜过滤除菌。

[0229] 马铃薯葡萄糖琼脂 (PDA, Difco Dehydrated Culture Media)——马铃薯，来自 200g/L 的浸液；葡萄糖 20g/L 和琼脂 15g/L 在 50-80％终体积的 dH_2O 中充分混合，然后定容至 100％终体积。将此混合物高压灭菌，冷却至 55℃并倾倒。

[0230] 为了制成用于 pH3.5 培养基的 1％脱脂奶琼脂，如上制备 PDA，并向 100ml 熔化的 PDA 中加入 1.8ml 的 10％酒石酸和 12.5ml 灭菌的 8％脱脂奶，并铺板。为了对脱脂奶进行预灭菌，将 8％脱脂奶 (Difco) 在 122-123℃下高压灭菌 10 分钟，暴露期间的室压为 32-35psi。取出混合物，冷却至室温并在室温下储存。

[0231] 在摇瓶中生长 3 天后的转化体中，评估蛋白酶的表达。用琼脂塞接种里氏木霉培养基 (Davis 等，(1970)Methods Enzymol.17:79-143)。将培养物在 30℃下摇动培养 3 天。使培养肉汤通过 0.22 微米滤膜，并将滤液点到 1％脱脂奶琼脂上。经过在室温下过夜培养，观察到清澈的区带。

[0232] 实施例 4

[0233] NSP24、NSP25 和 L388M PeoA 的 PH 活性图

[0234] PepA（野生型和 L388M）、NSP24 和 NSR25 在里氏木霉菌株中均过量表达，它们的 pH 活性图，使用获自 Molecular Probes(EnzChekPortease Kit-Green fluorescence) 的荧光标记的酪蛋白试验进行测定。PepA（野生型和 L388M）和 NSP 是全发酵样品，而 NSP24 是在

50％甘油中稳定化的纯化蛋白。将该酶稀释为 1.0mg/ml、0.5mg/ml 和 0.25mg/ml。荧光标记的底物在 DI H_2O 中被稀释为 0.1mg/ml。将 10ml 底物加入 50ml 各种 pH 的缓冲液和 30ul DI H_2O 中。通过加入 10ml 酶启动反应，并在加入 100ul1.0M pH10 的磷酸盐中止反应前使其继续进行不同的时段。用 485 激发光在 538nm 发射处测量样品的荧光，并在 SpectraMAX EM 荧光板读数器中在 530nm 将发射光用滤镜截止。 NSP24 在 pH3.7 下具有最适活性，野生型 PepA 在 pH3.4 下具有最适活性，而 L388M pepA 在 pH3.5 下具有最适活性。NSP25 在 pH4.6 下具有最适活性。

[0235] 实施例 5

[0236] 在实验室发酵中里氏木霉 NSP24 蛋白酶与 GC106 的比较

[0237] 当今乙醇工业中所使用的标准蛋白酶是由 GenencorInternational，Inc. 商业销售的蛋白酶 GC106。在糖类降解、葡萄糖形成以及乙醇生产方面，将 NSP24 的功能性与 GC106 进行比较。

[0238] 材料

[0239] Distillase L-400(Lot#107-04057-901，372GAU/g)

[0240] GC106(Lot#A01-01300-001，1010SAPU/g)

[0241] NSP24(Lot#20040423，1165SAPU/g)

[0242] Red Star Red Yeast

[0243] 来自乙醇生产商的醪液和稀釜馏物 (Mash and Thin Stillage)(玉米)

[0244] 方法

[0245] 从乙醇生产商获得醪液和稀釜馏物 (Mash and Thin Stillage)(在发酵前，也被称为大浅盒 (backset))，并混合至 26.5 白利 (brix)。将 pH 用 1N HCL 调至 pH4.3。然后将样品分为 3-300 克的等份试样并放入 32℃水浴中。平衡后，加入下列酶组合：

[0246] 表 2

[0247]

酶	水平	酶	水平
-	-	Distillase L-400	0.4GAU/g
GC106	70μl 的 1:10dil	Distillase L-400	0.4GAU/g
NSP24	60μl 的 1:10dil	Distillase L-400	0.4GAU/g

[0248] DISTILLASEL-400 是来源于黑曲霉的液体葡糖淀粉酶，其可以从 Genencor International Inc. 购得。在加入酶后，加入 1.00 克 / 瓶的 Red Star Red Yeast。在 16、24、40 和 48 小时取样并离心。将 500ul 每一样品放入装有 50ul1.1NH_2SO_4 的试管中以停止反应。2 分钟后，将样品用 4.5ml DI H_2O 稀释并混合。混匀后，使样品通过 0.45 微米滤膜并放入 HPLC 瓶中进行分析。通过 HPLC(Phenomenex Rezex8u) 分析样品。结果在图 1-4 中列出。NSP24 的表现与 GC106 相似。

[0249] 实验室 6

[0250] 在非熟化过程中 NSP24 对碎玉米的乙醇产量的影响

[0251] 用 DI H_2O 制成碾碎的玉米的 30％ DS 浆液。该碾碎的玉米是典型的用于乙醇工业的 #2 黄色马齿形玉米样品，其被粉碎以使 70％以上将通过 30 目筛。谷粒的含水量用 OHAUS，MB35Halogen moisturebalance(MB35 卤素水份平衡)(NJ) 进行测量。用 6N H_2SO_4 将 pH 调至 4.2。在 125ml 装有 100g 醪液的烧瓶中进行发酵，其中含有 1.0GAU/g 剂量的

STARGEN001 并含有或不含有 0.5kg/MT 剂量的 NSP24。

[0252] 在 45ml 水中制备 5g Red Star Ethanol Red 干酵母 (Lesaffre yeastCorporation, Milwaukee, WI)，并在接种至发酵罐前在 32℃水浴中混合 1 小时。向每一 125ml 烧瓶中加入 0.5ml 酵母浆液。将烧瓶放入 32℃水浴中并温和混合此醪液。在发酵期间，取出样品进行 HPLC 分析。(HPLC 柱 :Phenomenex Rezex 有机酸柱 (RHM-Monosaccharide)#00H0132-KO ;柱温 :60℃ ;流动相 :0.01N H_2SO_4 ;流速 :0.6ml/min ;检测器 :RI ;注射体积 :20uL。72 小时后停止发酵。在不同采样间隔的化合物产量，包括糖类、乳酸、甘油和乙醇，在下表 3 中表示，其中 + 表示 NSP24 被加入烧瓶中，而 -- 表示 NSP24 未被加入烧瓶中。经测量，所有样品的乳酸在大约 0.01 至 0.02% w/v 之间，而 DP-2 则测定为 0.0。在 24 小时，乙酸被测定为大约 0，而在 71 小时，对于所有样品在 0.03 至 0.04 之间。

[0253] 表 3

[0254]

NSP24	小时	% w/v DP > 4	% w/v DP-2	% w/v DP-1	% w/v 甘油	% w/v 乙醇
--	24	0.44	0.04	0.96	0.73	11.23
--	48	0.52	0.11	1.66	0.86	15.47
--	71	0.54	0.14	2.27	0.90	16.82
+	24	0.62	0.04	0.32	0.69	14.24
+	48	0.59	0.13	0.95	0.80	17.81
+	71	0.58	0.16	1.64	0.82	18.03

[0255] 实施例 7

[0256] 使用玉米胚乳进行乙醇生产中不同蛋白酶的比较

[0257] 使用胚乳 (除菌玉米，75.8% 淀粉，99.5% 的颗粒大小 < 30 目) 作为颗粒淀粉底物制得 29.5% 的 DS 醪液。将一百克每 - 醪液转移至 125ml 烧瓶，并将培养基的 pH 调至 pH 4.5。将蛋白酶 (NSP24 ;中性蛋白酶 (MULTIFECT NEUTRAL, PROTEINASE-T) 和碱性蛋白酶 (SPEZYME FAN, PROTEX 6L MULTIFECT P-3000 和 PROTEASE 899(Genencor International)) 以 0.5kg/MT 加入，随后以 2.5Kgs/MT 加入 STARGEN 001(Genencor International) 淀粉。然后用 0.5ml20% 酵母 (Red Star Ethanol Red) 接种烧瓶，并放入水浴中保持在 32℃。在温育过程中将烧瓶的内容物连续搅拌以获得均匀的混合。在不同的时间间隔取样进行 HPLC 分析。测定 72 小时发酵罐肉汤中 DDGS 的残留淀粉和蛋白质含量。乙醇生产的结果在下表 4 中表示。

[0258] 表 4

[0259]

蛋白酶	% 乙醇, v/v 20 小时	% 乙醇, v/v 27 小时	% 乙醇, v/v 43 小时	% 乙醇, v/v 51 小时	% 乙醇, v/v 75 小时
NSP24	11.50	14.42	17.39	17.97	18.61
MULTIFECT P3000	9.58	12.12	15.04	15.75	16.96
PROTEX 6L	9.94	12.54	15.46	15.95	17.29
SPEZYME FAN	9.78	12.27	15.03	15.88	17.19
PROTEINASE T	9.29	11.73	15.01	15.87	17.28
PROTEINASE899	9.62	11.90	14.66	15.37	17.95
MULTIFECTNEUTRAL	9.63	11.91	14.73	15.31	16.85

<120> 酸性真菌蛋白酶

<130>GC863-PCT

<140>PCT/US2005/046435
<141>2005-12-20

<150>US 60/640,399
<151>2004-12-30

<150>US 60/648,233
<151>2005-01-27

<160>14

<170>PatentIn version 3.2

<210>1
<211>9931
<212>DNA
<213> 人工序列

<220>
<223> 合成的 pTrex3g_NSP24 质粒

<400>1
ctgcagccac ttgcagtccc gtggaattct cacggtgaat gtaggccttt tgtagggtag 60
gaattgtcac tcaagcaccc ccaacctcca ttacgcctcc cccatagagt tcccaatcag 120
tgagtcatgg cactgttctc aaatagattg gggagaagtt gacttccgcc cagagctgaa 180
ggtcgcacaa ccgcatgata tagggtcggc aacggcaaaa aagcacgtgg ctcaccgaaa 240
agcaagatgt ttgcgatcta acatccagga acctggatac atccatcatc acgcacgacc 300
actttgatct gctggtaaac tcgtattcgc cctaaaccga agtgcgtggt aaatctacac 360
gtgggcccct ttcggtatac tgcgtgtgtc ttctctaggt gccattcttt tcccttcctc 420
tagtgttgaa ttgtttgtgt tggagtccga gctgtaacta cctctgaatc tctggagaat 480
ggtggactaa cgactaccgt gcacctgcat catgtatata atagtgatcc tgagaagggg 540
ggtttggagc aatgtgggac tttgatggtc atcaaacaaa gaacgaagac gcctcttttg 600
caaagttttg tttcggctac ggtgaagaac tggatacttg ttgtgtcttc tgtgtatttt 660

tgtggcaaca agaggccaga gacaatctat tcaaacacca agcttgctct tttgagctac 720
aagaacctgt ggggtatata tctagagttg tgaagtcggt aatcccgctg tatagtaata 780
cgagtcgcat ctaaatactc cgaagctgct gcgaacccgg agaatcgaga tgtgctggaa 840
agcttctagc gagcggctaa attagcatga aaggctatga gaaattctgg agacggcttg 900
ttgaatcatg gcgttccatt cttcgacaag caaagcgttc cgtcgcagta gcaggcactc 960
attcccgaaa aaactcggag attcctaagt agcgatggaa ccggaataat ataataggca 1020
atacattgag ttgcctcgac ggttgcaatg caggggtact gagcttggac ataactgttc 1080
cgtaccccac ctcttctcaa cctttggcgt ttccctgatt cagcgtaccc gtacaagtcg 1140
taatcactat taacccagac tgaccggacg tgttttgccc ttcatttgga gaaataatgt 1200
cattgcgatg tgtaatttgc ctgcttgacc gactggggct gttcgaagcc cgaatgtagg 1260
attgttatcc gaactctgct cgtagaggca tgttgtgaat ctgtgtcggg caggacacgc 1320
ctcgaaggtt cacggcaagg gaaaccaccg atagcagtgt ctagtagcaa cctgtaaagc 1380
cgcaatgcag catcactgga aaatacaaac caatggctaa aagtacataa gttaatgcct 1440
aaagaagtca tataccagcg gctaataatt gtacaatcaa gtggctaaac gtaccgtaat 1500
ttgccaacgg cttgtggggt tgcagaagca acggcaaagc cccacttccc cacgtttgtt 1560
tcttcactca gtccaatctc agctggtgat cccccaattg ggtcgcttgt ttgttccggt 1620
gaagtgaaag aagacagagg taagaatgtc tgactcggag cgttttgcat acaaccaagg 1680
gcagtgatgg aagacagtga aatgttgaca ttcaaggagt atttagccag ggatgcttga 1740
gtgtatcgtg taaggaggtt tgtctgccga tacgacgaat actgtatagt cacttctgat 1800
gaagtggtcc atattgaaat gtaagtcggc actgaacagg caaaagattg agttgaaact 1860
gcctaagatc tcgggccctc gggccttcgg cctttgggtg tacatgtttg tgctccgggc 1920
aaatgcaaag tgtggtagga tcgaacacac tgctgccttt accaagcagc tgagggtatg 1980
tgataggcaa atgttcaggg gccactgcat ggtttcgaat agaaagagaa gcttagccaa 2040
gaacaatagc cgataaagat agcctcatta aacggaatga gctagtaggc aaagtcagcg 2100
aatgtgtata tataaaggtt cgaggtccgt gcctccctca tgctctcccc atctactcat 2160
caactcagat cctccaggag acttgtacac catcttttga ggcacagaaa cccaatagtc 2220
aaccatcaca agtttgtaca aaaaagcagg ctccgcggcc gccccccttca ccatgcagac 2280
ctttggagct tttctcgttt ccttcctcgc cgccagcggc ctggccgcgg ccctccccac 2340
cgagggtcag aagacggctt ccgtcgaggt ccagtacaac aagaactacg tcccccacgg 2400
ccctactgct ctcttcaagg ccaagagaaa gtatggcgct cccatcagcg acaacctgaa 2460
gtctctcgtg gctgccaggc aggccaagca ggctctcgcc aagcgccaga ccggctcggc 2520
gcccaaccac cccagtgaca gcgccgattc ggagtacatc acctccgtct ccatcggcac 2580
tccggctcag gtcctccccc tggactttga caccggctcc tccgacctgt gggtctttag 2640
ctccgagacg cccaagtctt cggccaccgg ccacgccatc tacacgccct ccaagtcgtc 2700
cacctccaag aaggtgtctg gcgccagctg gtccatcagc tacggcgacg gcagcagctc 2760
cagcggcgat gtctacaccg acaaggtcac catcggaggc ttcagcgtca acacccaggg 2820
cgtcgagtct gccacccgcg tgtccaccga gttcgtccag gacacggtca tctctggcct 2880
cgtcggcctt gcctttgaca gcggcaacca ggtcaggccg cacccgcaga agacgtggtt 2940
ctccaacgcc gccagcagcc tggctgagcc ccttttcact gccgacctga ggcacggaca 3000

gagtaagtag acactcactg gaattcgttc ctttcccgat catcatgaaa gcaagtagac 3060
tgactgaacc aaacaactag acggcagcta caactttggc tacatcgaca ccagcgtcgc 3120
caagggcccc gttgcctaca cccccgttga caacagccag ggcttctggg agttcactgc 3180
ctcgggctac tctgtcggcg gcggcaagct caaccgcaac tccatcgacg gcattgccga 3240
caccggcacc accctgctcc tcctcgacga caacgtcgtc gatgcctact acgccaacgt 3300
ccagtcggcc cagtacgaca accagcagga gggtgtcgtc ttcgactgcg acgaggacct 3360
cccttcgttc agcttcggtg ttggaagctc caccatcacc atccctggcg atctgctgaa 3420
cctgactccc ctcgaggagg gcagctccac ctgcttcggt ggcctccaga gcagctccgg 3480
cattggcatc aacatctttg gtgacgttgc cctcaaggct gccctggttg tctttgacct 3540
cggcaacgag cgcctgggct gggctcagaa ataaaagggt gggcgcgccg acccagcttt 3600
cttgtacaaa gtggtgatcg cgccagctcc gtgcgaaagc ctgacgcacc ggtagattct 3660
tggtgagccc gtatcatgac ggcggcggga gctacatggc cccgggtgat ttatttttt 3720
tgtatctact tctgaccctt ttcaaatata cggtcaactc atctttcact ggagatgcgg 3780
cctgcttggt attgcgatgt tgtcagcttg gcaaattgtg gctttcgaaa acacaaaacg 3840
attccttagt agccatgcat tttaagataa cggaatagaa gaaagaggaa attaaaaaaa 3900
aaaaaaaaac aaacatcccg ttcataaccc gtagaatcgc cgctcttcgt gtatcccagt 3960
accagtttat tttgaatagc tcgcccgctg gagagcatcc tgaatgcaag taacaaccgt 4020
agaggctgac acggcaggtg ttgctaggga gcgtcgtgtt ctacaaggcc agacgtcttc 4080
gcggttgata tatatgtatg tttgactgca ggctgctcag cgacgacagt caagttcgcc 4140
ctcgctgctt gtgcaataat cgcagtgggg aagccacacc gtgactccca tctttcagta 4200
aagctctgtt ggtgtttatc agcaatacac gtaatttaaa ctcgttagca tggggctgat 4260
agcttaatta ccgtttacca gtgccatggt tctgcagctt tccttggccc gtaaaattcg 4320
gcgaagccag ccaatcacca gctaggcacc agctaaaccc tataattagt ctcttatcaa 4380
caccatccgc tcccccggga tcaatgagga gaatgagggg gatgcggggc taaagaagcc 4440
tacataaccc tcatgccaac tcccagttta cactcgtcga gccaacatcc tgactataag 4500
ctaacacaga atgcctcaat cctgggaaga actggccgct gataagcgcg cccgcctcgc 4560
aaaaaccatc cctgatgaat ggaaagtcca gacgctgcct gcggaagaca gcgttattga 4620
tttcccaaag aaatcgggga tcctttcaga ggccgaactg aagatcacag aggcctccgc 4680
tgcagatctt gtgtccaagc tggcggccgg agagttgacc tcggtggaag ttacgctagc 4740
attctgtaaa cgggcagcaa tcgcccagca gttagtaggg tcccctctac ctctcaggga 4800
gatgtaacaa cgccacctta tgggactatc aagctgacgc tggcttctgt gcagacaaac 4860
tgcgcccacg agttcttccc tgacgccgct ctcgcgcagg caagggaact cgatgaatac 4920
tacgcaaagc acaagagacc cgttggtcca ctccatggcc tccccatctc tctcaaagac 4980
cagcttcgag tcaaggtaca ccgttgcccc taagtcgtta gatgtccctt tttgtcagct 5040
aacatatgcc accagggcta cgaaacatca atgggctaca tctcatggct aaacaagtac 5100
gacgaagggg actcggttct gacaaccatg ctccgcaaag ccggtgccgt cttctacgtc 5160
aagacctctg tcccgcagac cctgatggtc tgcgagacag tcaacaacat catcgggcgc 5220
accgtcaacc cacgcaacaa gaactggtcg tgcggcggca gttctggtgg tgagggtgcg 5280
atcgttggga ttcgtggtgg cgtcatcggt gtaggaacgg atatcggtgg ctcgattcga 5340

gtgccggccg cgttcaactt cctgtacggt ctaaggccga gtcatgggcg gctgccgtat 5400
gcaaagatgg cgaacagcat ggagggtcag gagacggtgc acagcgttgt cgggccgatt 5460
acgcactctg ttgagggtga gtccttcgcc tcttccttct tttcctgctc tataccaggc 5520
ctccactgtc ctcctttctt gctttttata ctatatacga gaccggcagt cactgatgaa 5580
gtatgttaga cctccgcctc ttcaccaaat ccgtcctcgg tcaggagcca tggaaatacg 5640
actccaaggt catccccatg ccctggcgcc agtccgagtc ggacattatt gcctccaaga 5700
tcaagaacgg cgggctcaat atcggctact acaacttcga cggcaatgtc cttccacacc 5760
ctcctatcct gcgcggcgtg gaaaccaccg tcgccgcact cgccaaagcc ggtcacaccg 5820
tgaccccgtg gacgccatac aagcacgatt tcggccacga tctcatctcc catatctacg 5880
cggctgacgg cagcgccgac gtaatgcgcg atatcagtgc atccggcgag ccggcgattc 5940
caaatatcaa agacctactg aacccgaaca tcaaagctgt taacatgaac gagctctggg 6000
acacgcatct ccagaagtgg aattaccaga tggagtacct tgagaaatgg cgggaggctg 6060
aagaaaaggc cgggaaggaa ctggacgcca tcatcgcgcc gattacgcct accgctgcgg 6120
tacggcatga ccagttccgg tactatgggt atgcctctgt gatcaacctg ctggatttca 6180
cgagcgtggt tgttccggtt acctttgcgg ataagaacat cgataagaag aatgagagtt 6240
tcaaggcggt tagtgagctt gatgccctcg tgcaggaaga gtatgatccg gaggcgtacc 6300
atggggcacc ggttgcagtg caggttatcg gacggagact cagtgaagag aggacgttgg 6360
cgattgcaga ggaagtgggg aagttgctgg gaaatgtggt gactccatag ctaataagtg 6420
tcagatagca atttgcacaa gaaatcaata ccagcaactg taaataagcg ctgaagtgac 6480
catgccatgc tacgaaagag cagaaaaaaa cctgccgtag aaccgaagag atatgacacg 6540
cttccatctc tcaaaggaag aatcccttca gggttgcgtt tccagtctag acacgtataa 6600
cggcacaagt gtctctcacc aaatgggtta tatctcaaat gtgatctaag gatggaaagc 6660
ccagaatatc gatcgcgcgc agatccatat atagggcccg ggttataatt acctcaggtc 6720
gacgtcccat ggccattcga attcgtaatc atggtcatag ctgtttcctg tgtgaaattg 6780
ttatccgctc acaattccac acaacatacg agccggaagc ataaagtgta aagcctgggg 6840
tgcctaatga gtgagctaac tcacattaat tgcgttgcgc tcactgcccg ctttccagtc 6900
gggaaacctg tcgtgccagc tgcattaatg aatcggccaa cgcgcgggga gaggcggttt 6960
gcgtattggg cgctcttccg cttcctcgct cactgactcg ctgcgctcgg tcgttcggct 7020
gcggcgagcg gtatcagctc actcaaaggc ggtaatacgg ttatccacag aatcagggga 7080
taacgcagga aagaacatgt gagcaaaagg ccagcaaaag gccaggaacc gtaaaaaggc 7140
cgcgttgctg gcgtttttcc ataggctccg ccccccctgac gagcatcaca aaaatcgacg 7200
ctcaagtcag aggtggcgaa acccgacagg actataaaga taccaggcgt ttcccccctgg 7260
aagctccctc gtgcgctctc ctgttccgac cctgccgctt accggatacc tgtccgcctt 7320
tctcccttcg ggaagcgtgg cgctttctca tagctcacgc tgtaggtatc tcagttcggt 7380
gtaggtcgtt cgctccaagc tgggctgtgt gcacgaaccc cccgttcagc ccgaccgctg 7440
cgccttatcc ggtaactatc gtcttgagtc caacccggta agacacgact tatcgccact 7500
ggcagcagcc actggtaaca ggattagcag agcgaggtat gtaggcggtg ctacagagtt 7560
cttgaagtgg tggcctaact acggctacac tagaagaaca gtatttggta tctgcgctct 7620
gctgaagcca gttaccttcg gaaaaagagt tggtagctct tgatccggca aacaaaccac 7680

cgctggtagc ggtggttttt ttgtttgcaa gcagcagatt acgcgcagaa aaaaaggatc 7740
tcaagaagat cctttgatct tttctacggg gtctgacgct cagtggaacg aaaactcacg 7800
ttaagggatt ttggtcatga gattatcaaa aaggatcttc acctagatcc ttttaaatta 7860
aaaatgaagt tttaaatcaa tctaaagtat atatgagtaa acttggtctg acagttacca 7920
atgcttaatc agtgaggcac ctatctcagc gatctgtcta tttcgttcat ccatagttgc 7980
ctgactcccc gtcgtgtaga taactacgat acgggagggc ttaccatctg gccccagtgc 8040
tgcaatgata ccgcgagacc cacgctcacc ggctccagat ttatcagcaa taaaccagcc 8100
agccggaagg gccgagcgca gaagtggtcc tgcaacttta tccgcctcca tccagtctat 8160
taattgttgc cgggaagcta gagtaagtag ttcgccagtt aatagtttgc gcaacgttgt 8220
tgccattgct acaggcatcg tggtgtcacg ctcgtcgttt ggtatggctt cattcagctc 8280
cggttcccaa cgatcaaggc gagttacatg atcccccatg ttgtgcaaaa aagcggttag 8340
ctccttcggt cctccgatcg ttgtcagaag taagttggcc gcagtgttat cactcatggt 8400
tatggcagca ctgcataatt ctcttactgt catgccatcc gtaagatgct tttctgtgac 8460
tggtgagtac tcaaccaagt cattctgaga atagtgtatg cggcgaccga gttgctcttg 8520
cccggcgtca atacgggata ataccgcgcc acatagcaga actttaaaag tgctcatcat 8580
tggaaaacgt tcttcggggc gaaaactctc aaggatctta ccgctgttga gatccagttc 8640
gatgtaaccc actcgtgcac ccaactgatc ttcagcatct tttactttca ccagcgtttc 8700
tgggtgagca aaaacaggaa ggcaaaatgc cgcaaaaaag ggaataaggg cgacacggaa 8760
atgttgaata ctcatactct tccttttca atattattga agcatttatc agggttattg 8820
tctcatgagc ggatacatat ttgaatgtat ttagaaaaat aaacaaatag gggttccgcg 8880
cacatttccc cgaaaagtgc cacctgacgt ctaagaaacc attattatca tgacattaac 8940
ctataaaaat aggcgtatca cgaggccctt tcgtctcgcg cgtttcggtg atgacggtga 9000
aaacctctga cacatgcagc tcccggagac ggtcacagct tgtctgtaag cggatgccgg 9060
gagcagacaa gcccgtcagg gcgcgtcagc gggtgttggc gggtgtcggg gctggcttaa 9120
ctatgcggca tcagagcaga ttgtactgag agtgcaccat aaaattgtaa acgttaatat 9180
tttgttaaaa ttcgcgttaa atttttgtta aatcagctca ttttttaacc aataggccga 9240
aatcggcaaa atcccttata aatcaaaaga atagcccgag atagggttga gtgttgttcc 9300
agtttggaac aagagtccac tattaaagaa cgtggactcc aacgtcaaag ggcgaaaaac 9360
cgtctatcag ggcgatggcc cactacgtga accatcaccc aaatcaagtt ttttggggtc 9420
gaggtgccgt aaagcactaa atcggaaccc taaagggagc ccccgattta gagcttgacg 9480
gggaaagccg gcgaacgtgg cgagaaagga agggaagaaa gcgaaaggag cgggcgctag 9540
ggcgctggca agtgtagcgg tcacgctgcg cgtaaccacc acacccgccg cgcttaatgc 9600
gccgctacag ggcgcgtact atggttgctt tgacgtatgc ggtgtgaaat accgcacaga 9660
tgcgtaagga gaaaataccg catcaggcgc cattcgccat tcaggctgcg caactgttgg 9720
gaagggcgat cggtgcgggc ctcttcgcta ttacgccagc tggcgaaagg gggatgtgct 9780
gcaaggcgat taagttgggt aacgccaggg ttttcccagt cacgacgttg taaaacgacg 9840
gccagtgccc aagcttacta gtacttctcg agctctgtac atgtccggtc gcgacgtacg 9900
cgtatcgatg gcgccagctg caggcggccg c 9931

<210>2
<211>407
<212>PRT
<213> 里氏木霉 (Trichoderma reesei)
<400>2

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

260 265 270

Trp Glu Phe Thr Ala Ser Gly Tyr Ser Val Gly Gly Gly Lys Leu Asn

275 280 285

Arg Asn Ser Ile Asp Gly Ile Ala Asp Thr Gly Thr Thr Leu Leu Leu

290 295 300

Leu Asp Asp Asn Val Val Asp Ala Tyr Tyr Ala Asn Val Gln Ser Ala

305 310 315 320

Gln Tyr Asp Asn Gln Gln Glu Gly Val Val Phe Asp Cys Asp Glu Asp

325 330 335

Leu Pro Ser Phe Ser Phe Gly Val Gly Ser Ser Thr Ile Thr Ile Pro

340 345 350

Gly Asp Leu Leu Asn Leu Thr Pro Leu Glu Glu Gly Ser Ser Thr Cys

355 360 365

Phe Gly Gly Leu Gln Ser Ser Ser Gly Ile Gly Ile Asn Ile Phe Gly

370 375 380

Asp Val Ala Leu Lys Ala Ala Leu Val Val Phe Asp Leu Gly Asn Glu

385 390 395 400

Arg Leu Gly Trp Ala Gln Lys

405

<210>3

<211>1347

<212>DNA

<213> 人工序列

<220>

<223> 合成的 pepA 核苷酸

<400>3

atggtcgtct tcagcaaaac cgctgccctc gttctgggtc tgtcctccgc cgtctctgcg 60

gcgccggctc ctactcgcaa gggcttcacc atcaaccaga ttgcccggcc tgccaacaag 120

acccgcacca tcaacctgcc aggcatgtac gcccgttccc tggccaagtt tggcggtacg 180

gtgccccaga gcgtgaagga ggctgccagc aagggtagtg ccgtgaccac gccccagaac 240

aatgacgagg agtacctgac tcccgtcact gtcggaaagt ccaccctcca tctggacttt 300

gacaccggat ctgcagatct gtaagcttcc ctgctcgggt gttcgggcaa atcgtgacta 360

acctggacta gctgggtctt ctcggacgag ctcccttcct cggagcagac cggtcacgat 420

ctgtacacgc ctagctccag cgcgaccaag ctgagcggct acacttggga catctcctac 480

ggtgacggca gctcggccag cggagacgtg taccgggata ctgtcactgt cggcggtgtc 540

accaccaaca agcaggctgt tgaagcagcc agcaagatca gctccgagtt cgttcagaac 600

acggccaatg acggcctttt gggactggcc tttagctcca tcaacactgg tgagtcaatc 660
ctacatcagc cgggttgacc tacctgctga ccgatacaca gtccagccca aggcgcagac 720
caccttcttc gacaccgtca agtcccagct ggactctccc cttttcgccg tgcagctgaa 780
gcacgacgcc cccggtgttt acgactttgg ctacatcgat gactccaagt acaccggttc 840
tatcacctac acggatgccg atagctccca gggttactgg ggcttcagca ccgacggcta 900
cagtatcggt gacggcagct ccagctccag cggcttcagc gccattgctg gtaagaaccg 960
ccttcattta acacacaact tgtccacctc tttactaact agtgtataga caccggtacc 1020
accctcatcc tcctcgatga cgaaatcgtc tccgcctact acgagcaggt ttctggcgct 1080
caggagagcg aggaagccgg tggctacgtt ttctcttgct cgaccaaccc ccctgacttc 1140
actgtcgtga ttggcgacta caaggccgtt gttccgggca agtacatcaa ctacgctccc 1200
atctcgactg gcagctccac ctgctttggc ggtatccaga gcaacagcgg tctgggactg 1260
tccatcctgg gtgatgtttt cttgaagagc cagtacgtgg tcttcaactc tgagggccct 1320
aagctgggat tcgccgctca ggcttag 1347

<210>4
<211>1199
<212>DNA
<213> 人工序列

<220>
<223> 合成的 NSP25 核苷酸

<400>4
atgcagccct catttggcag cttcctcgtc accgtcctgt ctgcctccat ggcagcaggc 60
agtgtcattc ccagcacaaa cgccaaccct ggctccttcg agatcaagag atccgccaac 120
aaagccttca caggccgcaa tggccctcta gcattagccc gtacatacgc caagtacggt 180
gttgaagtcc ccaaaactct ggtcgatgct attcaactcg tcaagtccat ccagctcgca 240
aagcgggaca gcgccaccgt cactgccacg ccggaccacg acgacatcga gtatcttgtc 300
cccgtcaaga tcggaactcc tccccaaaca cttaacctgg attttgacac gggcagctcc 360
gatctctggg tcttctcatc agatgtcgac ccgacctcct cccagggcca tgacatctac 420
accccgtcca agagcacatc ttccaaaaag ttggaaggag cctcatggaa catcacatat 480
ggagaccgct catcatcatc cggcgatgtc taccacgata ttgtctccgt cggaaacctg 540
acagtaaagt cccaagccgt cgagtccgct cgaaacgtct cggccagttc acccagggca 600
acaacgacgg cctcgtcggc ctggcgttta gctccatcaa cacagtcaag cccacgccgc 660
aaaagacgtg gtacgacaac atcgtcggca gccttgactc tcccgtcttt gttgctgatc 720
tgcgccacga cacgcccggc agctaccact tcggctccat cccctccgaa gcaagcaaag 780
ccttctacgc ccccatcgac aacagcaagg gcttctggca attcagcacg agcagcaaca 840
ttagcggcca gttcaacgcc gttgcagaca ctggcactac tctgctgctc gccagcgacg 900
acctcgtcaa ggcctactac gcaaaggtcc agggcgcccg tgtgaacgtc ttcctgggcg 960

gctacgtctt caactgcacc actcagctgc ccgactttac ctttactgtt ggagagggca 1020
acatcactgt ccccggtacc ttgataaact attccgaggc tggcaacggc cagtgttttg 1080
gcggtattca gccgtcgggg ggtcttcctt ttgctatctt tggtgacatt gctcttaagg 1140
ctgcgtatgt tatttttgac agtggcaaca agcaggttgg ctgggcgcag aagaaatag 1199

<210>5
<211>1347
<212>DNA
<213> 人工序列

<220>
<223> 合成的 L388M pepA 变体核苷酸

<400>5
atggtcgtct tcagcaaaac cgctgccctc gttctgggtc tgtcctccgc cgtctctgcg 60
gcgccggctc ctactcgcaa gggcttcacc atcaaccaga ttgcccggcc tgccaacaag 120
acccgcacca tcaacctgcc aggcatgtac gcccgttccc tggccaagtt tggcggtacg 180
gtgccccaga gcgtgaagga ggctgccagc aagggtagtg ccgtgaccac gccccagaac 240
aatgacgagg agtacctgac tcccgtcact gtcggaaagt ccaccctcca tctggacttt 300
gacaccggat ctgcagatct gtaagcttcc ctgctcgggt gttcgggcaa atcgtgacta 360
acctggacta gctgggtctt ctcggacgag ctcccttcct cggagcagac cggtcacgat 420
ctgtacacgc ctagctccag cgcgaccaag ctgagcggct acacttggga catctcctac 480
ggtgacggca gctcggccag cggagacgtg taccgggata ctgtcactgt cggcggtgtc 540
accaccaaca agcaggctgt tgaagcagcc agcaagatca gctccgagtt cgttcagaac 600
acggccaatg acggcctttt gggactggcc tttagctcca tcaacactgg tgagtcaatc 660
ctacatcagc cgggttgacc tacctgctga ccgatacaca gtccagccca aggcgcagac 720
caccttcttc gacaccgtca agtcccagct ggactctccc cttttcgccg tgcagctgaa 780
gcacgacgcc cccggtgttt acgactttgg ctacatcgat gactccaagt acaccggttc 840
tatcacctac acggatgccg atagctccca gggttactgg ggcttcagca ccgacggcta 900
cagtatcggt gacggcagct ccagctccag cggcttcagc gccattgctg gtaagaaccg 960
ccttcattta acacacaact tgtccacctc tttactaact agtgtataga caccggtacc 1020
accctcatcc tcctcgatga cgaaatcgtc tccgcctact acgagcaggt ttctggcgct 1080
caggagagcg aggaagccgg tggctacgtt ttctcttgct cgaccaaccc ccctgacttc 1140
actgtcgtga ttggcgacta caaggccgtt gttccgggca agtacatcaa ctacgctccc 1200
atctcgactg gcagctccac ctgctttggc ggtatccaga gcaacagcgg tctgggactg 1260
tccatcctgg gtgatgtttt cttgaagagc cagtacgtgg tcttcaactc tgagggccct 1320
aagatgggat tcgccgctca ggcttag 1347

<210>6

<211>10286
<212>DNA
<213> 人工序列

<220>
<223> 合成的 pSL899_pepA 质粒

<400>6

tacgtatttt gaatagctcg cccgctggag agcatcctga atgcaagtaa caaccgtaga 60
ggctgacacg gcaggtgttg ctagggagcg tcgtgttcta caaggccaga cgtcttcgcg 120
gttgatatat atgtatgttt gactgcaggc tgctcagcga cgacagtcaa gttcgccctc 180
gctgcttgtg caataatcgc agtggggaag ccacaccgtg actcccatct ttcagtaaag 240
ctctgttggt gtttatcagc aatacacgta atttaaactc gttagcatgg ggctgatagc 300
ttaattaccg tttaccagtg ccgcggttct gcagctttcc ttggcccgta aaattcggcg 360
aagccagcca atcaccagct aggcaccagc taaaccctat aattagtctc ttatcaacac 420
catccgctcc cccgggatca atgaggagaa tgaggggat gcggggctaa agaagcctac 480
ataaccctca tgccaactcc cagtttacac tcgtcgagcc aacatcctga ctataagcta 540
acacagaatg cctcaatcct gggaagaact ggccgctgat aagcgcgccc gcctcgcaaa 600
aaccatccct gatgaatgga aagtccagac gctgcctgcg gaagacagcg ttattgattt 660
cccaaagaaa tcggggatcc tttcagaggc cgaactgaag atcacagagg cctccgctgc 720
agatcttgtg tccaagctgg cggccggaga gttgacctcg gtggaagtta cgctagcatt 780
ctgtaaacgg gcagcaatcg cccagcagtt agtagggtcc cctctacctc tcagggagat 840
gtaacaacgc caccttatgg gactatcaag ctgacgctgg cttctgtgca gacaaactgc 900
gcccacgagt tcttccctga cgccgctctc gcgcaggcaa gggaactcga tgaatactac 960
gcaaagcaca agagacccgt tggtccactc catggcctcc ccatctctct caaagaccag 1020
cttcgagtca aggtacaccg ttgcccctaa gtcgttagat gtccctttt gtcagctaac 1080
atatgccacc agggctacga aacatcaatg ggctacatct catggctaaa caagtacgac 1140
gaaggggact cggttctgac aaccatgctc cgcaaagccg gtgccgtctt ctacgtcaag 1200
acctctgtcc cgcagaccct gatggtctgc gagacagtca acaacatcat cgggcgcacc 1260
gtcaacccac gcaacaagaa ctggtcgtgc ggcggcagtt ctggtggtga gggtgcgatc 1320
gttgggattc gtggtggcgt catcggtgta ggaacggata tcggtggctc gattcgagtg 1380
ccggccgcgt tcaacttcct gtacggtcta aggccgagtc atgggcggct gccgtatgca 1440
aagatggcga acagcatgga gggtcaggag acggtgcaca gcgttgtcgg gccgattacg 1500
cactctgttg agggtgagtc cttcgcctct tccttctttt cctgctctat accaggcctc 1560
cactgtcctc ctttcttgct ttttatacta tatacgagac cggcagtcac tgatgaagta 1620
tgttagacct ccgcctcttc accaaatccg tcctcggtca ggagccatgg aaatacgact 1680
ccaaggtcat ccccatgccc tggcgccagt ccgagtcgga cattattgcc tccaagatca 1740
agaacggcgg gctcaatatc ggctactaca acttcgacgg caatgtcctt ccacaccctc 1800
ctatcctgcg cggcgtggaa accaccgtcg ccgcactcgc caaagccggt cacaccgtga 1860

ccccgtggac gccatacaag cacgatttcg gccacgatct catctcccat atctacgcgg 1920
ctgacggcag cgccgacgta atgcgcgata tcagtgcatc cggcgagccg gcgattccaa 1980
atatcaaaga cctactgaac ccgaacatca aagctgttaa catgaacgag ctctgggaca 2040
cgcatctcca gaagtggaat taccagatgg agtaccttga gaaatggcgg gaggctgaag 2100
aaaaggccgg gaaggaactg gacgccatca tcgcgccgat tacgcctacc gctgcggtac 2160
ggcatgacca gttccggtac tatgggtatg cctctgtgat caacctgctg gatttcacga 2220
gcgtggttgt tccggttacc tttgcggata agaacatcga taagaagaat gagagtttca 2280
aggcggttag tgagcttgat gccctcgtgc aggaagagta tgatccggag gcgtaccatg 2340
gggcaccggt tgcagtgcag gttatcggac ggagactcag tgaagagagg acgttggcga 2400
ttgcagagga agtggggaag ttgctgggaa atgtggtgac tccatagcta ataagtgtca 2460
gatagcaatt tgcacaagaa atcaatacca gcaactgtaa ataagcgctg aagtgaccat 2520
gccatgctac gaaagagcag aaaaaaacct gccgtagaac cgaagagata tgacacgctt 2580
ccatctctca aaggaagaat cccttcaggg ttgcgtttcc agtctagcta gagtcgagga 2640
ttgcctgaac attgacattc ggcgtccggc cgggaccacc gcggactcga agctgcctgt 2700
gctggtctgg atctttggcg gaggctttga acttggttca aaggcgatgt atgatggtac 2760
aacgatggta tcatcgtcga tagacaagaa catgcctatc gtgtttgtag caatgaatta 2820
tcgcgtggga ggtttcgggt tcttgcccgg aaaggagatc ctggaggacg ggtccgcgaa 2880
cctagggctc ctggaccaac gccttgccct gcagtgggtt gccgacaaca tcgaggcctt 2940
tggtggagac ccggacaagg tgacgatttg gggagaatca gcaggagcca tttccgtttt 3000
tgatcagatg atcttgtacg acggaaacat cacttacaag gataagccct tgttccgggg 3060
ggccatcatg gactccggta gtgttgttcc cgcagacccc gtcgatgggg tcaagggaca 3120
gcaagtatat gatgcggtag tggaatctgc aggctgttcc tcttctaacg acaccctagc 3180
ttgtctgcgt gaactagact acaccgactt cctcaatgcg gcaaactccg tgccaggcat 3240
tttaagctac cattctgtgg cgttatcata tgtgcctcga ccggacggga cggcgttgtc 3300
ggcatcaccg gacgttttgg gcaaagcagg gaaatatgct cgggtcccgt tcatcgtggg 3360
cgaccaagag gatgagggga ccttattcgc cttgtttcag tccaacatta cgacgatcga 3420
cgaggtggtc gactacctgg cctcatactt cttctatgac gctagccgag agcagcttga 3480
agaactagtg gccctgtacc cagacaccac cacgtacggg tctccgttca ggacaggcgc 3540
ggccaacaac tggtatccgc aatttaagcg attggccgcc attctcggcg acttggtctt 3600
caccattacc cggcgggcat tcctctcgta tgcagaggaa atctcccctg atcttccgaa 3660
ctggtcgtac ctggcgacct atgactatgg caccccagtt ctggggacct tccacggaag 3720
tgacctgctg caggtgttct atgggatcaa gccaaactat gcagctagtt ctagccacac 3780
gtactatctg agctttgtgt atacgctgga tccgaactcc aaccgggggg agtacattga 3840
gtggccgcag tggaaggaat cgcggcagtt gatgaatttc ggagcgaacg acgccagtct 3900
ccttacggat gatttccgca acgggacata tgagttcatc ctgcagaata ccgcggcgtt 3960
ccacatctga tgccattggc ggaggggtcc ggacggtcag gaacttagcc ttatgagatg 4020
aatgatggac gtgtctggcc tcggaaaagg atatatgggg atcatgatag tactagccat 4080
attaatgaag ggcatatacc acgcgttgga cctgcgttat agcttcccgt tagttatagt 4140
accatcgtta taccagccaa tcaagtcacc acgcacgacc ggggacggcg aatccccggg 4200

aattgaaaga aattgcatcc caggccagtg aggccagcga ttggccacct ctccaaggca 4260
cagggccatt ctgcagcgct ggtggattca tcgcaatttc ccccggcccg gcccgacacc 4320
gctataggct ggttctccca caccatcgga gattcgtcgc ctaatgtctc gtccgttcac 4380
aagctgaaga gcttgaagtg gcgagatgtc tctgcaggaa ttcaagctag atgctaagcg 4440
atattgcatg gcaatatgtg ttgatgcatg tgcttcttcc ttcagcttcc cctcgtgcag 4500
atgaggtttg gctataaatt gaagtggttg gtcggggttc cgtgaggggc tgaagtgctt 4560
cctccctttt agacgcaact gagagcctga gcttcatccc cagcatcatt acacctcgag 4620
atggtcgtct tcagcaaaac cgctgccctc gttctgggtc tgtcctccgc cgtctctgcg 4680
gcgccggctc ctactcgcaa gggcttcacc atcaaccaga ttgcccggcc tgccaacaag 4740
acccgcacca tcaacctgcc aggcatgtac gcccgttccc tggccaagtt tggcggtacg 4800
gtgccccaga gcgtgaagga ggctgccagc aagggtagtg ccgtgaccac gccccagaac 4860
aatgacgagg agtacctgac tcccgtcact gtcggaaagt ccaccctcca tctggacttt 4920
gacaccggat ctgcagatct gtaagcttcc ctgctcgggt gttcgggcaa atcgtgacta 4980
acctggacta gctgggtctt ctcggacgag ctcccttcct cggagcagac cggtcacgat 5040
ctgtacacgc ctagctccag cgcgaccaag ctgagcggct acacttggga catctcctac 5100
ggtgacggca gctcggccag cggagacgtg taccgggata ctgtcactgt cggcggtgtc 5160
accaccaaca agcaggctgt tgaagcagcc agcaagatca gctccgagtt cgttcagaac 5220
acggccaatg acggcctttt gggactggcc tttagctcca tcaacactgg tgagtcaatc 5280
ctacatcagc cgggttgacc tacctgctga ccgatacaca gtccagccca aggcgcagac 5340
caccttcttc gacaccgtca agtcccagct ggactctccc cttttcgccg tgcagctgaa 5400
gcacgacgcc cccggtgttt acgactttgg ctacatcgat gactccaagt acaccggttc 5460
tatcacctac acggatgccg atagctccca gggttactgg ggcttcagca ccgacggcta 5520
cagtatcggt gacggcagct ccagctccag cggcttcagc gccattgctg gtaagaaccg 5580
ccttcattta acacacaact tgtccacctc tttactaact agtgtataga caccggtacc 5640
accctcatcc tcctcgatga cgaaatcgtc tccgcctact acgagcaggt ttctggcgct 5700
caggagagcg aggaagccgg tggctacgtt ttctcttgct cgaccaaccc ccctgacttc 5760
actgtcgtga ttggcgacta caaggccgtt gttccgggca agtacatcaa ctacgctccc 5820
atctcgactg gcagctccac ctgctttggc ggtatccaga gcaacagcgg tctgggactg 5880
tccatcctgg gtgatgtttt cttgaagagc cagtacgtgg tcttcaactc tgagggccct 5940
aagctgggat tcgccgctca ggcttagtct agagtcgacc gcgacggtga ccgacacctg 6000
gcggtagact atttattcct gttgatatga aggatgagca tgagggtaat tgctcatata 6060
atcatgtatg tagtggatgt gcataagagc aacgaaatgg aagcctgatc atgtgattgt 6120
attgcgaccg acggaattga ggatatgcgg agatacggac agtgccagag ccattgtctt 6180
cacgtaaagt accagacggt ccctgatttc ttcttgcaca tagcattagg caattgacat 6240
gttgtcgctc tactgatatc actgtccctc aaagcatagc catgagctca tcttagatcc 6300
aagcacgtaa ttccatagcc gaggtccaca gtggagcaac agcagcatcc atcattgctt 6360
ctcccccagg ggcctcttag cgactaaacc tggagtatgt ctcaaccagc caatgaatcg 6420
tcttcgcttc aatgtccttg acacttctga gagggtcccc atccctcaat gctaattcaa 6480
aatatagccg agatgcatgg tggagtccaa agtagacagt attgccggaa tgacggggcc 6540

agttgcgccg aggtcattgg ccggctgtga tgccatctgc cactaaatcc gatcattgat 6600
ccaccgccca cgagggccgt ctttgctttt gcgctgcgtc caggttcaca catctctctc 6660
tctgcagctc cagactgacc agactattct acttactggt ctgatcggct ccatcagagc 6720
tatggcgtta tcccgtgccg ttgctgcgcc atcgctatct tgatcgcgag ctcgaactca 6780
cttcttgttt taatagttgt tctcggtgac tgagtgtcgg tgagtgacag accacaacac 6840
cattgttgca gggggtaaat ttattcaatt caggaattgg attgttcgtc ccgccatgat 6900
gttcttgccg gctttgttgg ccctgtttgt cggatgcgac gccctcgctg tgcagcaggc 6960
aggtactgct ggatgatgag ccgtcggtct ccgcgcgcaa gcctaacttc ctcttcattc 7020
ttacggatga tcaggatctg cagatcgaat tccaccggcg tatatgccgt atacacaggc 7080
gagaatcaag gagaagggta ctgagttttg aatcatttgt tactactggc tctgtgctgt 7140
ccgtcgcgcg tgagtctttg gacggaagac aggctcataa tactaatgtg acggatgtga 7200
acccgcctta tggtatgaat acctctcaga tcggtcatgt ttcttcggtg taaaattgct 7260
aatgcagcat aggcggatac cccaagttcg tcgcccaagg cttcaacgaa aacttcctcc 7320
ccgtttggct gcagtccgcc ggttacaata ccttctacac ggggaaactg ttcaactgcc 7380
acagcgtcgc tacctataat gcaccgtttg tgaacggctt caatggctct gatttcctcc 7440
tcgatcccca cacctattcc tactggaacg cgacgtacca acgaaaccat gagcctccgc 7500
ggagctacga gggacaatac acaacggatg tgatgcggga gaaggcatcg gggttgttgg 7560
cagatgcgct ggacaggacg cgccgttctt ctgacggtgc ctatccgccg cacacgaaca 7620
tcgacctgca ggcatgcaag cttggcactg gccgtcgttt tacaacgtcg tgactgggaa 7680
aaccctggcg ttacccaact taatcgcctt gcagcacatc ccccctttcgc cagctggcgt 7740
aatagcgaag aggcccgcac cgatcgccct tcccaacagt tgcgcagcct gaatggcgaa 7800
tggcgcctga tgcggtattt tctccttacg catctgtgcg gtatttcaca ccgcatatgg 7860
tgcactctca gtacaatctg ctctgatgcc gcatagttaa gccagccccg acacccgcca 7920
acacccgctg acgcgccctg acgggcttgt ctgctcccgg catccgctta cagacaagct 7980
gtgaccgtct ccgggagctg catgtgtcag aggttttcac cgtcatcacc gaaacgcgcg 8040
agacgaaagg gcctcgtgat acgcctattt ttataggtta atgtcatgat aataatggtt 8100
tcttagacgt caggtggcac ttttcgggga aatgtgcgcg gaacccctat ttgtttattt 8160
ttctaaatac attcaaatat gtatccgctc atgagacaat aaccctgata aatgcttcaa 8220
taatattgaa aaaggaagag tatgagtatt caacatttcc gtgtcgccct tattcccttt 8280
tttgcggcat tttgccttcc tgtttttgct cacccagaaa cgctggtgaa agtaaaagat 8340
gctgaagatc agttgggtgc acgagtgggt tacatcgaac tggatctcaa cagcggtaag 8400
atccttgaga gttttcgccc cgaagaacgt tttccaatga tgagcacttt taaagttctg 8460
ctatgtggcg cggtattatc ccgtattgac gccgggcaag agcaactcgg tcgccgcata 8520
cactattctc agaatgactt ggttgagtac tcaccagtca cagaaaagca tcttacggat 8580
ggcatgacag taagagaatt atgcagtgct gccataacca tgagtgataa cactgcggcc 8640
aacttacttc tgacaacgat cggaggaccg aaggagctaa ccgcttttt gcacaacatg 8700
ggggatcatg taactcgcct tgatcgttgg gaaccggagc tgaatgaagc cataccaaac 8760
gacgagcgtg acaccacgat gcctgtagca atggcaacaa cgttgcgcaa actattaact 8820
ggcgaactac ttactctagc ttcccggcaa caattaatag actggatgga ggcggataaa 8880

gttgcaggac cacttctgcg ctcggccctt ccggctggct ggtttattgc tgataaatct 8940
ggagccggtg agcgtgggtc tcgcggtatc attgcagcac tggggccaga tggtaagccc 9000
tcccgtatcg tagttatcta cacgacgggg agtcaggcaa ctatggatga acgaaataga 9060
cagatcgctg agataggtgc ctcactgatt aagcattggt aactgtcaga ccaagtttac 9120
tcatatatac tttagattga tttaaaactt catttttaat ttaaaaggat ctaggtgaag 9180
atccttttg ataatctcat gaccaaaatc ccttaacgtg agttttcgtt ccactgagcg 9240
tcagaccccg tagaaaagat caaaggatct tcttgagatc ctttttttct gcgcgtaatc 9300
tgctgcttgc aaacaaaaaa accaccgcta ccagcggtgg tttgtttgcc ggatcaagag 9360
ctaccaactc tttttccgaa ggtaactggc ttcagcagag cgcagatacc aaatactgtc 9420
cttctagtgt agccgtagtt aggccaccac ttcaagaact ctgtagcacc gcctacatac 9480
ctcgctctgc taatcctgtt accagtggct gctgccagtg gcgataagtc gtgtcttacc 9540
gggttggact caagacgata gttaccggat aaggcgcagc ggtcgggctg aacgggggt 9600
tcgtgcacac agcccagctt ggagcgaacg acctacaccg aactgagata cctacagcgt 9660
gagctatgag aaagcgccac gcttcccgaa gggagaaagg cggacaggta tccggtaagc 9720
ggcagggtcg gaacaggaga gcgcacgagg gagcttccag ggggaaacgc ctggtatctt 9780
tatagtcctg tcgggtttcg ccacctctga cttgagcgtc gatttttgtg atgctcgtca 9840
gggggggcgga gcctatggaa aaacgccagc aacgcggcct ttttacggtt cctggccttt 9900
tgctggcctt ttgctcacat gttctttcct gcgttatccc ctgattctgt ggataaccgt 9960
attaccgcct ttgagtgagc tgataccgct cgccgcagcc gaacgaccga gcgcagcgag 10020
tcagtgagcg aggaagcgga agagcgccca atacgcaaac cgcctctccc cgcgcgttgg 10080
ccgattcatt aatgcagctg gcacgacagg tttcccgact ggaaagcggg cagtgagcgc 10140
aacgcaatta atgtgagtta gctcactcat taggcacccc aggctttaca ctttatgctt 10200
ccggctcgta tgttgtgtgg aattgtgagc ggataacaat ttcacacagg aaacagctat 10260
gaccatgatt acgaattcga gctcgg 10286

<210>7
<211>394
<212>PRT
<213> 人工序列

<220>
<223> 合成的 L388M pepA 变体肽

<400>7

Met	Val	Val	Phe	Ser	Lys	Thr	Ala	Ala	Leu	Val	Leu	Gly	Leu	Ser	Ser
1				5				10					15		
Ala	Val	Ser	Ala	Ala	Pro	Ala	Pro	Thr	Arg	Lys	Gly	Phe	Thr	Ile	Asn
			20					25				30			
Gln	Ile	Ala	Arg	Pro	Ala	Asn	Lys	Thr	Arg	Thr	Ile	Asn	Leu	Pro	Gly

35 40 45

Met Tyr Ala Arg Ser Leu Ala Lys Phe Gly Gly Thr Val Pro Gln Ser
50 55 60

Val Lys Glu Ala Ala Ser Lys Gly Ser Ala Val Thr Thr Pro Gln Asn
65 70 75 80

Asn Asp Glu Glu Tyr Leu Thr Pro Val Thr Val Gly Lys Ser Thr Leu
85 90 95

His Leu Asp Phe Asp Thr Gly Ser Ala Asp Leu Trp Val Phe Ser Asp
100 105 110

Glu Leu Pro Ser Ser Glu Gln Thr Gly His Asp Leu Tyr Thr Pro Ser
115 120 125

Ser Ser Ala Thr Lys Leu Ser Gly Tyr Thr Trp Asp Ile Ser Tyr Gly
1301 35 140

Asp Gly Ser Ser Ala Ser Gly Asp Val Tyr Arg Asp Thr Val Thr Val
145 150 155 160

Gly Gly Val Thr Thr Asn Lys Gln Ala Val Glu Ala Ala Ser Lys Ile
165 170 175

Ser Ser Glu Phe Val Gln Asn Thr Ala Asn Asp Gly Leu Leu Gly Leu
180 185 190

Ala Phe Ser Ser Ile Asn Thr Val Gln Pro Lys Ala Gln Thr Thr Phe
195 200 205

Phe Asp Thr Val Lys Ser Gln Leu Asp Ser Pro Leu Phe Ala Val Gln
210 215 220

Leu Lys His Asp Ala Pro Gly Val Tyr Asp Phe Gly Tyr Ile Asp Asp
225 230 235 240

Ser Lys Tyr Thr Gly Ser Ile Thr Tyr Thr Asp Ala Asp Ser Ser Gln
245 250 255

Gly Tyr Trp Gly Phe Ser Thr Asp Gly Tyr Ser Ile Gly Asp Gly Ser
260 265 270

Ser Ser Ser Ser Gly Phe Ser Ala Ile Ala Asp Thr Gly Thr Thr Leu
275 280 285

Ile Leu Leu Asp Asp Glu Ile Val Ser Ala Tyr Tyr Glu Gln Val Ser
290 295 300

Gly Ala Ser Gly Glu Thr Glu Ala Gly Gly Tyr Val Phe Ser Cys Ser
305 310 315 320

Thr Asn Pro Pro Asp Phe Thr Val Val Ile Gly Asp Tyr Lys Ala Val
325 330 335

Val Pro Gly Lys Tyr Ile Asn Tyr Ala Pro Ile Ser Thr Gly Ser Ser
340 345 350

Thr Cys Phe Gly Gly Ile Gln Ser Asn Ser Gly Leu Gly Leu Ser Ile
355 360 365
Leu Gly Asp Val Phe Leu Lys Ser Gln Tyr Val Val Phe Asn Ser Glu
370 375 380
Gly Pro Lys Met Gly Phe Ala Ala Gln Ala
385 390

<210>8
<211>1302
<212>DNA
<213> 里氏木霉

<400>8
atgcagacct ttggagcttt tctcgtttcc ttcctcgccg ccagcggcct ggccgcggcc 60
ctccccaccg agggtcagaa gacggcttcc gtcgaggtcc agtacaacaa gaactacgtc 120
ccccacggcc ctactgctct cttcaaggcc aagagaaagt atggcgctcc catcagcgac 180
aacctgaagt ctctcgtggc tgccaggcag gccaagcagg ctctcgccaa gcgccagacc 240
ggctcggcgc ccaaccaccc cagtgacagc gccgattcgg agtacatcac ctccgtctcc 300
atcggcactc cggctcaggt cctccccctg gactttgaca ccggctcctc cgacctgtgg 360
gtctttagct ccgagacgcc caagtcttcg gccaccggcc acgccatcta cacgccctcc 420
aagtcgtcca cctccaagaa ggtgtctggc gccagctggt ccatcagcta cggcgacggc 480
agcagctcca gcggcgatgt ctacaccgac aaggtcacca tcggaggctt cagcgtcaac 540
acccagggcg tcgagtctgc cacccgcgtg tccaccgagt tcgtccagga cacggtcatc 600
tctggcctcg tcggccttgc ctttgacagc ggcaaccagg tcaggccgca cccgcagaag 660
acgtggttct ccaacgccgc cagcagcctg gctgagcccc ttttcactgc cgacctgagg 720
cacggacaga gtaagtagac actcactgga attcgttcct ttcccgatca tcatgaaagc 780
aagtagactg actgaaccaa acaactagac ggcagctaca actttggcta catcgacacc 840
agcgtcgcca agggccccgt tgcctacacc cccgttgaca acagccaggg cttctgggag 900
ttcactgcct cgggctactc tgtcggcggc ggcaagctca accgcaactc catcgacggc 960
attgccgaca ccggcaccac cctgctcctc ctcgacgaca acgtcgtcga tgcctactac 1020
gccaacgtcc agtcggccca gtacgacaac cagcaggagg gtgtcgtctt cgactgcgac 1080
gaggacctcc cttcgttcag cttcggtgtt ggaagctcca ccatcaccat ccctggcgat 1140
ctgctgaacc tgactccccc cgaggagggc agctccacct gcttcggtgg cctccagagc 1200
agctccggca ttggcatcaa catctttggt gacgttgccc tcaaggctgc cctggttgtc 1260
tttgacctcg gcaacgagcg cctgggctgg gctcagaaat aa 1302
<210>9
<211>399
<212>PRT
<213> 人工序列

<220>
<223> 合成的 NSP25 肽

<220>
<221>misc_feature
<222>(195)..(195)
<223>Xaa 可以是任何天然氨基酸

<400>9

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

225 230 235 240
Leu Arg His Asp Thr Pro Gly Ser Tyr His Phe Gly Ser Ile Pro Ser
245 250 255
Glu Ala Ser Lys Ala Phe Tyr Ala Pro Ile Asp Asn Ser Lys Gly Phe
260 265 270
Trp Gln Phe Ser Thr Ser Ser Asn Ile Ser Gly Gln Phe Asn Ala Val
275 280 285
Ala Asp Thr Gly Thr Thr Leu Leu Leu Ala Ser Asp Asp Leu Val Lys
290 295 300
Ala Tyr Tyr Ala Lys Val Gln Gly Ala Arg Val Asn Val Phe Leu Gly
305 310 315 320
Gly Tyr Val Phe Asn Cys Thr Thr Gln Leu Pro Asp Phe Thr Phe Thr
325 330 335
Val Gly Glu Gly Asn Ile Thr Val Pro Gly Thr Leu Ile Asn Tyr Ser
340 345 350
Glu Ala Gly Asn Gly Gln Cys Phe Gly Gly Ile Gln Pro Ser Gly Gly
355 360 365
Leu Pro Phe Ala Ile Phe Gly Asp Ile Ala Leu Lys Ala Ala Tyr Val
370 375 380
Ile Phe Asp Ser Gly Asn Lys Gln Val Gly Trp Ala Gln Lys Lys
385 390 395

<210>10
<211>355
<212>PRT
<213> 里氏木霉

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

85 90 95

Ser Gly Ala Ser Trp Ser Ile Ser Tyr Gly Asp Gly Ser Ser Ser Ser

100 105 110

Gly Asp Val Tyr Thr Asp Lys Val Thr Ile Gly Gly Phe Ser Val Asn

115 120 125

Thr Gln Gly Val Glu Ser Ala Thr Arg Val Ser Thr Glu Phe Val Gln

130 135 140

Asp Thr Val Ile Ser Gly Leu Val Gly Leu Ala Phe Asp Ser Gly Asn

145 150 155 160

Gln Val Arg Pro His Pro Gln Lys Thr Trp Phe Ser Asn Ala Ala Ser

165 170 175

Ser Leu Ala Glu Pro Leu Phe Thr Ala Asp Leu Arg His Gly Gln Asn

180 185 190

Gly Ser Tyr Asn Phe Gly Tyr Ile Asp Thr Ser Val Ala Lys Gly Pro

195 200 205

Val Ala Tyr Thr Pro Val Asp Asn Ser Gln Gly Phe Trp Glu Phe Thr

210 215 220

Ala Ser Gly Tyr Ser Val Gly Gly Gly Lys Leu Asn Arg Asn Ser Ile

225 230 235 240

Asp Gly Ile Ala Asp Thr Gly Thr Thr Leu Leu Leu Leu Asp Asp Asn

245 250 255

Val Val Asp Ala Tyr Tyr Ala Asn Val Gln Ser Ala Gln Tyr Asp Asn

260 265 270

Gln Gln Glu Gly Val Val Phe Asp Cys Asp Glu Asp Leu Pro Ser Phe

275 280 285

Ser Phe Gly Val Gly Ser Ser Thr Ile Thr Ile Pro Gly Asp Leu Leu

290 295 300

Asn Leu Thr Pro Leu Glu Glu Gly Ser Ser Thr Cys Phe Gly Gly Leu

305 310 315 320

Gln Ser Ser Ser Gly Ile Gly Ile Asn Ile Phe Gly Asp Val Ala Leu

325 330 335

Lys Ala Ala Leu Val Val Phe Asp Leu Gly Asn Glu Arg Leu Gly Trp

340 345 350

Ala Gln Lys

355

<210>11

<211>22

<212>DNA

<213> 人工序列

<220>
<223> 引物

<400>11
caccatgcag acctttggag ct 22

<210>12
<211>21
<212>DNA
<213> 人工序列

<220>
<223> 引物

<400>12
ttatttctga gcccagccca g 21

<210>13
<211>24
<212>DNA
<213> 人工序列

<220>
<223> 引物

<400>13
caccatgcag ccctcatttg gcag 24

<210>14
<211>24
<212>DNA
<213> 人工序列

<220>
<223> 引物

<400>14

ctatttcttc tgcgcccagc caac 24

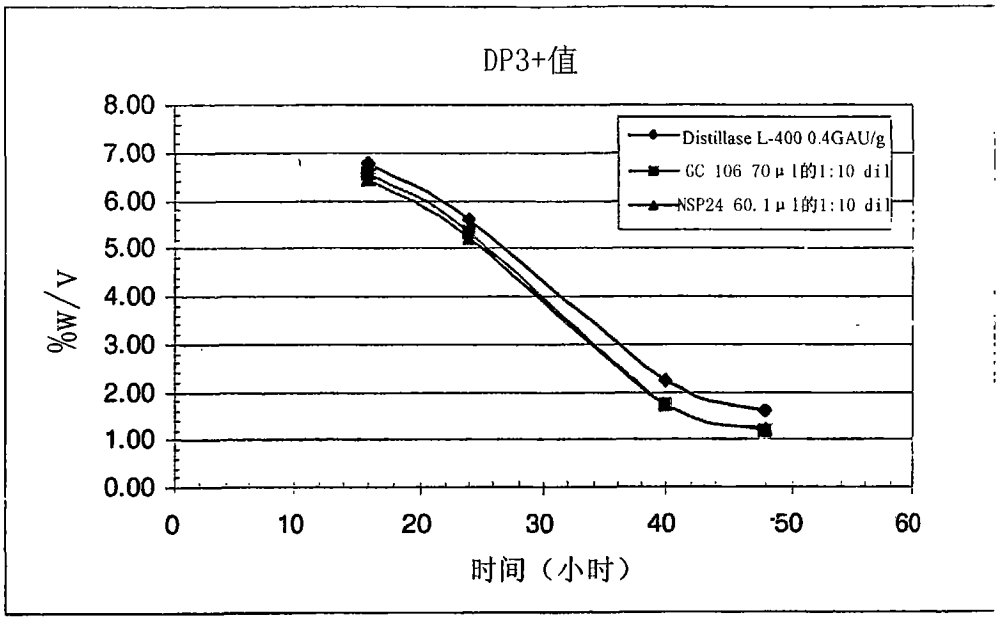

DP3+值
8.00
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
%W/V
0
10
20
30
40
50
60
时间（小时）
Distillase L-400 0.4GAU/g
GC 106 70μl的1:10 dil
NSP24 60.1μl的1:10 dil

图 1

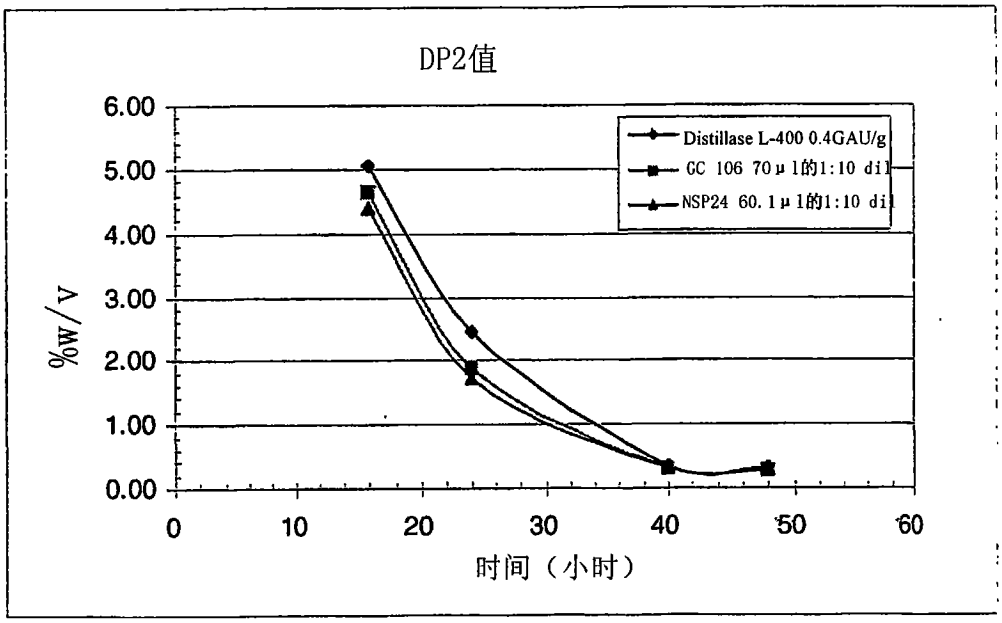

DP2值
6.00
5.00
4.00
3.00
2.00
1.00
0.00
%W/V
0
10
20
30
40
50
60
时间（小时）
Distillase L-400 0.4GAU/g
GC 106 70μl的1:10 dil
NSP24 60.1μl的1:10 dil

图 2

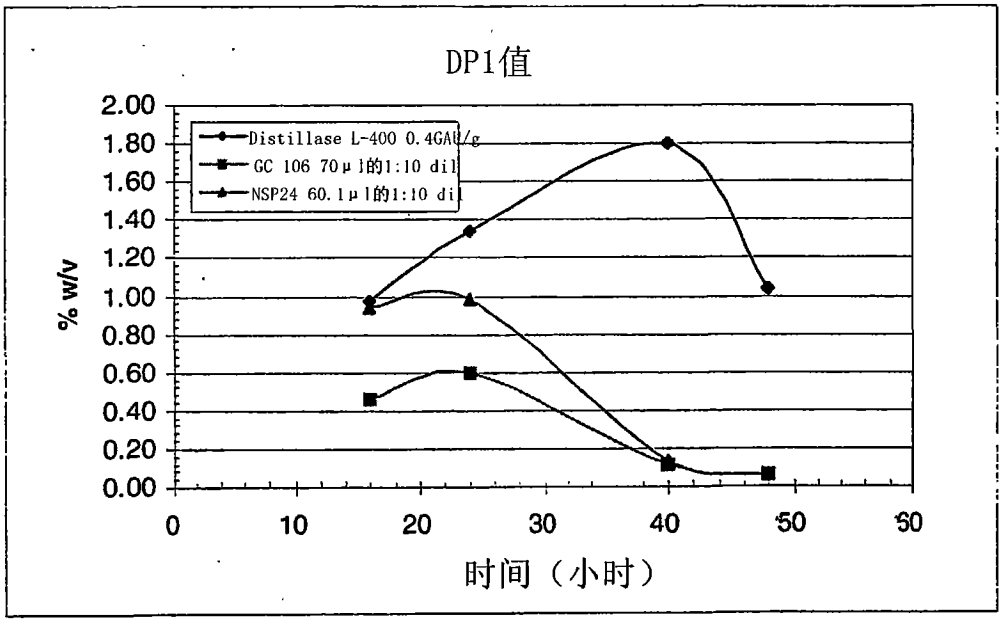

DP1值
2.00
1.80
1.60
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
% w/v
Distillase L-400 0.4GAU/g
GC 106 70μl的1:10 dil
NSP24 60.1μl的1:10 dil
0
10
20
30
40
50
60
时间（小时）

图 3

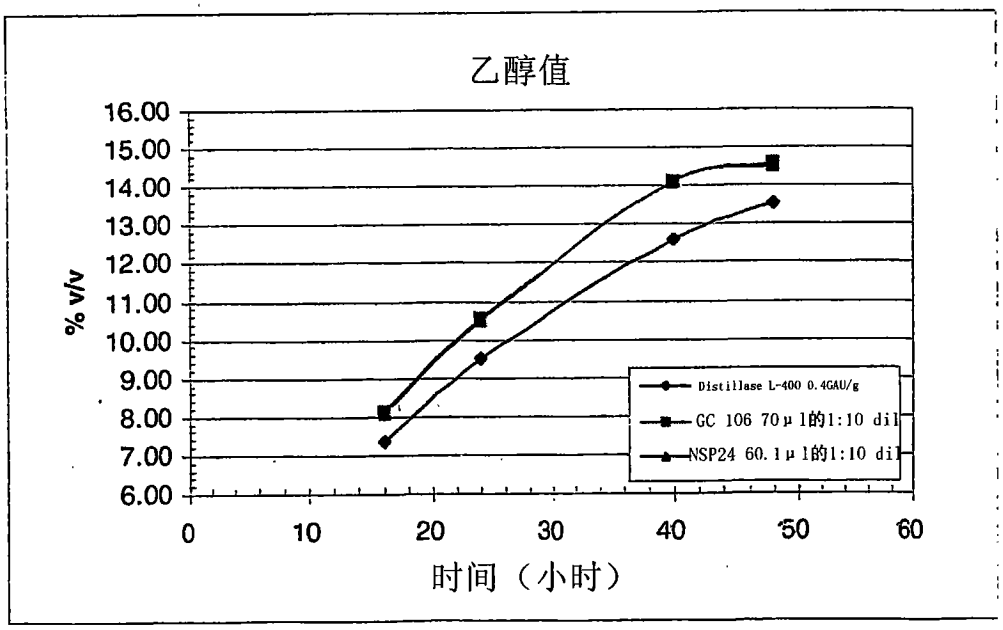

乙醇值
16.00
15.00
14.00
13.00
12.00
11.00
10.00
9.00
8.00
7.00
6.00
% v/v
Distillase L-400 0.4GAU/g
GC 106 70μl的1:10 dil
NSP24 60.1μl的1:10 dil
0
10
20
30
40
50
60
时间（小时）

图 4

pTrex3g_NSP24核酸序列(SEQ ID NO:1)

CTGCAGCCACTTGCAGTCCCGTGGAATTCTCACGGTGAATGTAGGCCTTTTGTAGGGTAGGA
ATTGTCACTCAAGCACCCCCAACCTCCATTACGCCTCCCCCATAGAGTTCCCAATCAGTGAG
TCATGGCACTGTTCTCAAATAGATTGGGGAGAAGTTGACTTCCGCCCAGAGCTGAAGGTCGC
ACAACCGCATGATATAGGGTCGGCAACGGCAAAAAAGCACGTGGCTCACCGAAAAGCAAGAT
GTTTGCGATCTAACATCCAGGAACCTGGATACATCCATCATCACGCACGACCACTTTGATCT
GCTGGTAAACTCGTATTCGCCCTAAACCGAAGTGCGTGGTAAATCTACACGTGGGCCCCTTT
CGGTATACTGCGTGTGTCTTCTCTAGGTGCCATTCTTTTCCCTTCCTCTAGTGTTGAATTGT
TTGTGTTGGAGTCCGAGCTGTAACTACCTCTGAATCTCTGGAGAATGGTGGACTAACGACTA
CCGTGCACCTGCATCATGTATATAATAGTGATCCTGAGAAGGGGGGTTTGGAGCAATGTGGG
ACTTTGATGGTCATCAAACAAAGAACGAAGACGCCTCTTTTGCAAAGTTTTGTTTCGGCTAC
GGTGAAGAACTGGATACTTGTTGTGTCTTCTGTGTATTTTTGTGGCAACAAGAGGCCAGAGA
CAATCTATTCAAACACCAAGCTTGCTCTTTTGAGCTACAAGAACCTGTGGGGTATATATCTA
GAGTTGTGAAGTCGGTAATCCCGCTGTATAGTAATACGAGTCGCATCTAAATACTCCGAAGC
TGCTGCGAACCCGGAGAATCGAGATGTGCTGGAAAGCTTCTAGCGAGCGGCTAAATTAGCAT
GAAAGGCTATGAGAAATTCTGGAGACGGCTTGTTGAATCATGGCGTTCCATTCTTCGACAAG
CAAAGCGTTCCGTCGCAGTAGCAGGCACTCATTCCCGAAAAAACTCGGAGATTCCTAAGTAG
CGATGGAACCGGAATAATATAATAGGCAATACATTGAGTTGCCTCGACGGTTGCAATGCAGG
GGTACTGAGCTTGGACATAACTGTTCCGTACCCCACCTCTTCTCAACCTTTGGCGTTTCCCT
GATTCAGCGTACCCGTACAAGTCGTAATCACTATTAACCCAGACTGACCGGACGTGTTTTGC
CCTTCATTTGGAGAAATAATGTCATTGCGATGTGTAATTTGCCTGCTTGACCGACTGGGGCT
GTTCGAAGCCCGAATGTAGGATTGTTATCCGAACTCTGCTCGTAGAGGCATGTTGTGAATCT
GTGTCGGGCAGGACACGCCTCGAAGGTTCACGGCAAGGGAAACCACCGATAGCAGTGTCTAG
TAGCAACCTGTAAAGCCGCAATGCAGCATCACTGGAAAATACAAACCAATGGCTAAAAGTAC
ATAAGTTAATGCCTAAAGAAGTCATATACCAGCGGCTAATAATTGTACAATCAAGTGGCTAA
ACGTACCGTAATTTGCCAACGGCTTGTGGGGTTGCAGAAGCAACGGCAAAGCCCCACTTCCC
CACGTTTGTTTCTTCACTCAGTCCAATCTCAGCTGGTGATCCCCCAATTGGGTCGCTTGTTT
GTTCCGGTGAAGTGAAAGAAGACAGAGGTAAGAATGTCTGACTCGGAGCGTTTTGCATACAA
CCAAGGGCAGTGATGGAAGACAGTGAAATGTTGACATTCAAGGAGTATTTAGCCAGGGATGC
TTGAGTGTATCGTGTAAGGAGGTTTGTCTGCCGATACGACGAATACTGTATAGTCACTTCTG
ATGAAGTGGTCCATATTGAAATGTAAGTCGGCACTGAACAGGCAAAAGATTGAGTTGAAACT
GCCTAAGATCTCGGGCCCTCGGGCCTTCGGCCTTTGGGTGTACATGTTTGTGCTCCGGGCAA
ATGCAAAGTGTGGTAGGATCGAACACACTGCTGCCTTTACCAAGCAGCTGAGGGTATGTGAT
AGGCAAATGTTCAGGGGCCACTGCATGGTTTCGAATAGAAAGAGAAGCTTAGCCAAGAACAA
TAGCCGATAAAGATAGCCTCATTAAACGGAATGAGCTAGTAGGCAAAGTCAGCGAATGTGTA
TATATAAAGGTTCGAGGTCCGTGCCTCCCTCATGCTCTCCCCATCTACTCATCAACTCAGAT
CCTCCAGGAGACTTGTACACCATCTTTTGAGGCACAGAAACCCAATAGTCAACCATCACAAG
TTTGTACAAAAAAGCAGGCTCCGCGGCCGCCCCCTTCACCATGCAGACCTTTGGAGCTTTTC
TCGTTTCCTTCCTCGCCGCCAGCGGCCTGGCCGCGGCCCTCCCCACCGAGGGTCAGAAGACG
GCTTCCGTCGAGGTCCAGTACAACAAGAACTACGTCCCCCACGGCCCTACTGCTCTCTTCAA
GGCCAAGAGAAAGTATGGCGCTCCCATCAGCGACAACCTGAAGTCTCTCGTGGCTGCCAGGC
AGGCCAAGCAGGCTCTCGCCAAGCGCCAGACCGGCTCGGCGCCCAACCACCCCAGTGACAGC
GCCGATTCGGAGTACATCACCTCCGTCTCCATCGGCACTCCGGCTCAGGTCCTCCCCCTGGA
CTTTGACACCGGCTCCTCCGACCTGTGGGTCTTTAGCTCCGAGACGCCCAAGTCTTCGGCCA
CCGGCCACGCCATCTACACGCCCTCCAAGTCGTCCACCTCCAAGAAGGTGTCTGGCGCCAGC
TG

图 5A

GTCCATCAGCTACGGCGACGGCAGCAGCTCCAGCGGCGATGTCTACACCGACAAGGTCACCA
TCGGAGGCTTCAGCGTCAACACCCAGGGCGTCGAGTCTGCCACCCGCGTGTCCACCGAGTTC
GTCCAGGACACGGTCATCTCTGGCCTCGTCGGCCTTGCCTTTGACAGCGGCAACCAGGTCAG
GCCGCACCCGCAGAAGACGTGGTTCTCCAACGCCGCCAGCAGCCTGGCTGAGCCCCTTTTCA
CTGCCGACCTGAGGCACGGACAGA**GTAAGTAGACACTCACTGGAATTCGTTCCTTTCCCGAT**
CATCATGAAAGCAAGTAGACTGACTGAACCAAACAACTAGACGGCAGCTACAACTTTGGCTA
CATCGACACCAGCGTCGCCAAGGGCCCCGTTGCCTACACCCCCGTTGACAACAGCCAGGGCT
TCTGGGAGTTCACTGCCTCGGGCTACTCTGTCGGCGGCGGCAAGCTCAACCGCAACTCCATC
GACGGCATTGCCGACACCGGCACCACCCTGCTCCTCCTCGACGACAACGTCGTCGATGCCTA
CTACGCCAACGTCCAGTCGGCCCAGTACGACAACCAGCAGGAGGGTGTCGTCTTCGACTGCG
ACGAGGACCTCCCTTCGTTCAGCTTCGGTGTTGGAAGCTCCACCATCACCATCCCTGGCGAT
CTGCTGAACCTGACTCCCCTCGAGGAGGGCAGCTCCACCTGCTTCGGTGGCCTCCAGAGCAG
CTCCGGCATTGGCATCAACATCTTTGGTGACGTTGCCCTCAAGGCTGCCCTGGTTGTCTTTG
ACCTCGGCAACGAGCGCCTGGGCTGGGCTCAGAAATAAAAGGGTGGGCGCGCCGACCCAGCT
TTCTTGTACAAAGTGGTGATCGCGCCAGCTCCGTGCGAAAGCCTGACGCACCGGTAGATTCT
TGGTGAGCCCGTATCATGACGGCGGCGGGAGCTACATGGCCCCGGGTGATTTATTTTTTTTG
TATCTACTTCTGACCCTTTTCAAATATACGGTCAACTCATCTTTCACTGGAGATGCGGCCTG
CTTGGTATTGCGATGTTGTCAGCTTGGCAAATTGTGGCTTTCGAAAACACAAAACGATTCCT
TAGTAGCCATGCATTTTAAGATAACGGAATAGAAGAAAGAGGAAATTAAAAAAAAAAAAAAA
ACAAACATCCCGTTCATAACCCGTAGAATCGCCGCTCTTCGTGTATCCCAGTACCAGTTTAT
TTTGAATAGCTCGCCCGCTGGAGAGCATCCTGAATGCAAGTAACAACCGTAGAGGCTGACAC
GGCAGGTGTTGCTAGGGAGCGTCGTGTTCTACAAGGCCAGACGTCTTCGCGGTTGATATATA
TGTATGTTTGACTGCAGGCTGCTCAGCGACGACAGTCAAGTTCGCCCTCGCTGCTTGTGCAA
TAATCGCAGTGGGGAAGCCACACCGTGACTCCCATCTTTCAGTAAAGCTCTGTTGGTGTTTA
TCAGCAATACACGTAATTTAAACTCGTTAGCATGGGGCTGATAGCTTAATTACCGTTTACCA
GTGCCATGGTTCTGCAGCTTTCCTTGGCCCGTAAAATTCGGCGAAGCCAGCCAATCACCAGC
TAGGCACCAGCTAAACCCTATAATTAGTCTCTTATCAACACCATCCGCTCCCCCGGGATCAA
TGAGGAGAATGAGGGGGATGCGGGGCTAAAGAAGCCTACATAACCCTCATGCCAACTCCCAG
TTTACACTCGTCGAGCCAACATCCTGACTATAAGCTAACACAGAATGCCTCAATCCTGGGAA
GAACTGGCCGCTGATAAGCGCGCCCGCCTCGCAAAAACCATCCCTGATGAATGGAAAGTCCA
GACGCTGCCTGCGGAAGACAGCGTTATTGATTTCCCAAAGAAATCGGGGATCCTTTCAGAGG
CCGAACTGAAGATCACAGAGGCCTCCGCTGCAGATCTTGTGTCCAAGCTGGCGGCCGGAGAG
TTGACCTCGGTGGAAGTTACGCTAGCATTCTGTAAACGGGCAGCAATCGCCCAGCAGTTAGT
AGGGTCCCCTCTACCTCTCAGGGAGATGTAACAACGCCACCTTATGGGACTATCAAGCTGAC
GCTGGCTTCTGTGCAGACAAACTGCGCCCACGAGTTCTTCCCTGACGCCGCTCTCGCGCAGG
CAAGGGAACTCGATGAATACTACGCAAAGCACAAGAGACCCGTTGGTCCACTCCATGGCCTC
CCCATCTCTCTCAAAGACCAGCTTCGAGTCAAGGTACACCGTTGCCCCTAAGTCGTTAGATG
TCCCTTTTTGTCAGCTAACATATGCCACCAGGGCTACGAAACATCAATGGGCTACATCTCAT
GGCTAAACAAGTACGACGAAGGGGACTCGGTTCTGACAACCATGCTCCGCAAAGCCGGTGCC
GTCTTCTACGTCAAGACCTCTGTCCCGCAGACCCTGATGGTCTGCGAGACAGTCAACAACAT
CATCGGGCGCACCGTCAACCCACGCAACAAGAACTGGTCGTGCGGCGGCAGTTCTGGTGGTG
AGGGTGCGATCGTTGGGATTCGTGGTGGCGTCATCGGTGTAGGAACGGATATCGGTGGCTCG
ATTCGAGTGCCGGCCGCGTTCAACTTCCTGTACGGTCTAAGGCCGAGTCATGGGCGGCTGCC
GTATGCAAAGATGGCGAACAGCATGGAGGGTCAGGAGACGGTGCACAGCGTTGTCGGGCCGA
TTACGCACTCTGTTGAGGGTGAGTCCTTCGCCTCTTCCTTCTTTTCCTGCTCTATACCAGGC
CTCCACTGTCCTCCTTTCTTGCTTTTTATACTATATACGAGACCGGCAGTCACTGATGAAGT
ATGTTAGACCTCCGCCTCTTCACCAAATCCGTCCTCGGTCAGGAGCCATGGAAATACGACTC
CAAGGTCATCCCCATGCCCTGGCGCCAGTCCGAGTCGGACATTATTGCCTCCAAGATCAAGA
ACGGCGGGCTCAATATCGGCTACTACAACTTCGACGGCAATGTCCTTCCACACCCTCCTATC
CTGCGCGGCGTGGAAACCACCGTCGCCGCACTCGCCAAAGCCGGTCACACCGTGACCCCGTG
GACGCCATACAAGCACGATTTCGGCCACGATCTCATCTCCCATATCTACGCGGCTGACGGCA
GCGCCGACGTAATGCGCGATATC

图 5B

AGTGCATCCGGCGAGCCGGCGATTCCAAATATCAAAGACCTACTGAACCCGAACATCAAAGC
TGTTAACATGAACGAGCTCTGGGACACGCATCTCCAGAAGTGGAATTACCAGATGGAGTACC
TTGAGAAATGGCGGGAGGCTGAAGAAAAGGCCGGGAAGGAACTGGACGCCATCATCGCGCCG
ATTACGCCTACCGCTGCGGTACGGCATGACCAGTTCCGGTACTATGGGTATGCCTCTGTGAT
CAACCTGCTGGATTTCACGAGCGTGGTTGTTCCGGTTACCTTTGCGGATAAGAACATCGATA
AGAAGAATGAGAGTTTCAAGGCGGTTAGTGAGCTTGATGCCCTCGTGCAGGAAGAGTATGAT
CCGGAGGCGTACCATGGGGCACCGGTTGCAGTGCAGGTTATCGGACGGAGACTCAGTGAAGA
GAGGACGTTGGCGATTGCAGAGGAAGTGGGGAAGTTGCTGGGAAATGTGGTGACTCCATAGC
TAATAAGTGTCAGATAGCAATTTGCACAAGAAATCAATACCAGCAACTGTAAATAAGCGCTG
AAGTGACCATGCCATGCTACGAAAGAGCAGAAAAAAACCTGCCGTAGAACCGAAGAGATATG
ACACGCTTCCATCTCTCAAAGGAAGAATCCCTTCAGGGTTGCGTTTCCAGTCTAGACACGTA
TAACGGCACAAGTGTCTCTCACCAAATGGGTTATATCTCAAATGTGATCTAAGGATGGAAAG
CCCAGAATATCGATCGCGCGCAGATCCATATATAGGGCCCGGGTTATAATTACCTCAGGTCG
ACGTCCCATGGCCATTCGAATTCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTA
TCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCT
AATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAAC
CTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGG
GCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGG
TATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAG
AACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTT
TTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGC
GAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCT
CCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGC
GCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGG
GCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTT
GAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAG
CAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACA
CTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTT
GGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCA
GCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTG
ACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATC
TTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTA
AACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTAT
TTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTA
CCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATC
AGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCT
CCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTG
CGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTC
ATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAG
CGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTC
ATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGT
GACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTT
GCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATT
GGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGAT
GTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGT
GAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGA
ATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAG
CGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCC
GAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGG
CGTATCACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACAT
GCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTC
AGGGCGCGTCAGCGGGTGTTGGC

图 5C

GGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATAA
AATTGTAAACGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTT
TTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGCCCGAGATAGGG
TTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAA
AGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCAAATCAAGTT
TTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGA
GCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGG
CGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTA
ATGCGCCGCTACAGGGCGCGTACTATGGTTGCTTTGACGTATGCGGTGTGAAATACCGCACA
GATGCGTAAGGAGAAAATACCGCATCAGGCGCCATTCGCCATTCAGGCTGCGCAACTGTTGG
GAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGC
AAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCA
GTGCCCAAGCTTACTAGTACTTCTCGAGCTCTGTACATGTCCGGTCGCGACGTACGCGTATC
GATGGCGCCAGCTGCAGGCGGCCGC

图 5D

NSP24氨基酸序列—407氨基酸(SEQ ID NO:2):

MQTFGAFLVSFLAASGLAAALPTEGQKTASVEVQYNKNYVPHGPTALFKAKRKYGAPISDNL
KSLVAARQAKQALAKRQTGSAPNHPSDSADSEYITSVSIGTPAQVLPLDFDTGSSDLWVFSS
ETPKSSATGHAIYTPSKSSTSKKVSGASWSISYGDGSSSSGDVYTDKVTIGGFSVNTQGVES
ATRVSTEFVQDTVISGLVGLAFDSGNQVRPHPQKTWFSNAASSLAEPLFTADLRHGQNGSYN
FGYIDTSVAKGPVAYTPVDNSQGFWEFTASGYSVGGGKLNRNSIDGIADTGTTLLLLDDNVV
DAYYANVQSAQYDNQQEGVVFDCDEDLPSFSFGVGSSTITIPGDLLNLTPLEEGSSTCFGGL
QSSSGIGINIFGDVALKAALVVFDLGNERLGWAQK

图 6A

NSP24多核苷酸序列(SEQ ID NO:8)

ATGCAGACCTTTGGAGCTTTTCTCGTTTCCTTCCTCGCCGCCAGCGGCCTGGCCGCGGCCCT
CCCCACCGAGGGTCAGAAGACGGCTTCCGTCGAGGTCCAGTACAACAAGAACTACGTCCCCC
ACGGCCCTACTGCTCTCTTCAAGGCCAAGAGAAAGTATGGCGCTCCCATCAGCGACAACCTG
AAGTCTCTCGTGGCTGCCAGGCAGGCCAAGCAGGCTCTCGCCAAGCGCCAGACCGGCTCGGC
GCCCAACCACCCCAGTGACAGCGCCGATTCGGAGTACATCACCTCCGTCTCCATCGGCACTC
CGGCTCAGGTCCTCCCCCTGGACTTTGACACCGGCTCCTCCGACCTGTGGGTCTTTAGCTCC
GAGACGCCCAAGTCTTCGGCCACCGGCCACGCCATCTACACGCCCTCCAAGTCGTCCACCTC
CAAGAAGGTGTCTGGCGCCAGCTGGTCCATCAGCTACGGCGACGGCAGCAGCTCCAGCGGCG
ATGTCTACACCGACAAGGTCACCATCGGAGGCTTCAGCGTCAACACCCAGGGCGTCGAGTCT
GCCACCCGCGTGTCCACCGAGTTCGTCCAGGACACGGTCATCTCTGGCCTCGTCGGCCTTGC
CTTTGACAGCGGCAACCAGGTCAGGCCGCACCCGCAGAAGACGTGGTTCTCCAACGCCGCCA
GCAGCCTGGCTGAGCCCCTTTTCACTGCCGACCTGAGGCACGGACAGA**GTAAGTAGACACTC**
ACTGGAATTCGTTCCTTTCCCGATCATCATGAAAGCAAGTAGACTGACTGAACCAAACAACT
AGACGGCAGCTACAACTTTGGCTACATCGACACCAGCGTCGCCAAGGGCCCCGTTGCCTACA
CCCCCGTTGACAACAGCCAGGGCTTCTGGGAGTTCACTGCCTCGGGCTACTCTGTCGGCGGC
GGCAAGCTCAACCGCAACTCCATCGACGGCATTGCCGACACCGGCACCACCCTGCTCCTCCT
CGACGACAACGTCGTCGATGCCTACTACGCCAACGTCCAGTCGGCCCAGTACGACAACCAGC
AGGAGGGTGTCGTCTTCGACTGCGACGAGGACCTCCCTTCGTTCAGCTTCGGTGTTGGAAGC
TCCACCATCACCATCCCTGGCGATCTGCTGAACCTGACTCCCCTCGAGGAGGGCAGCTCCAC
CTGCTTCGGTGGCCTCCAGAGCAGCTCCGGCATTGGCATCAACATCTTTGGTGACGTTGCCC
TCAAGGCTGCCCTGGTTGTCTTTGACCTCGGCAACGAGCGCCTGGGCTGGGCTCAGAAATAA

图 6B

pTrex3g_NSP24载体，其包括pTrex3g_NSP24序列（图1，如上）

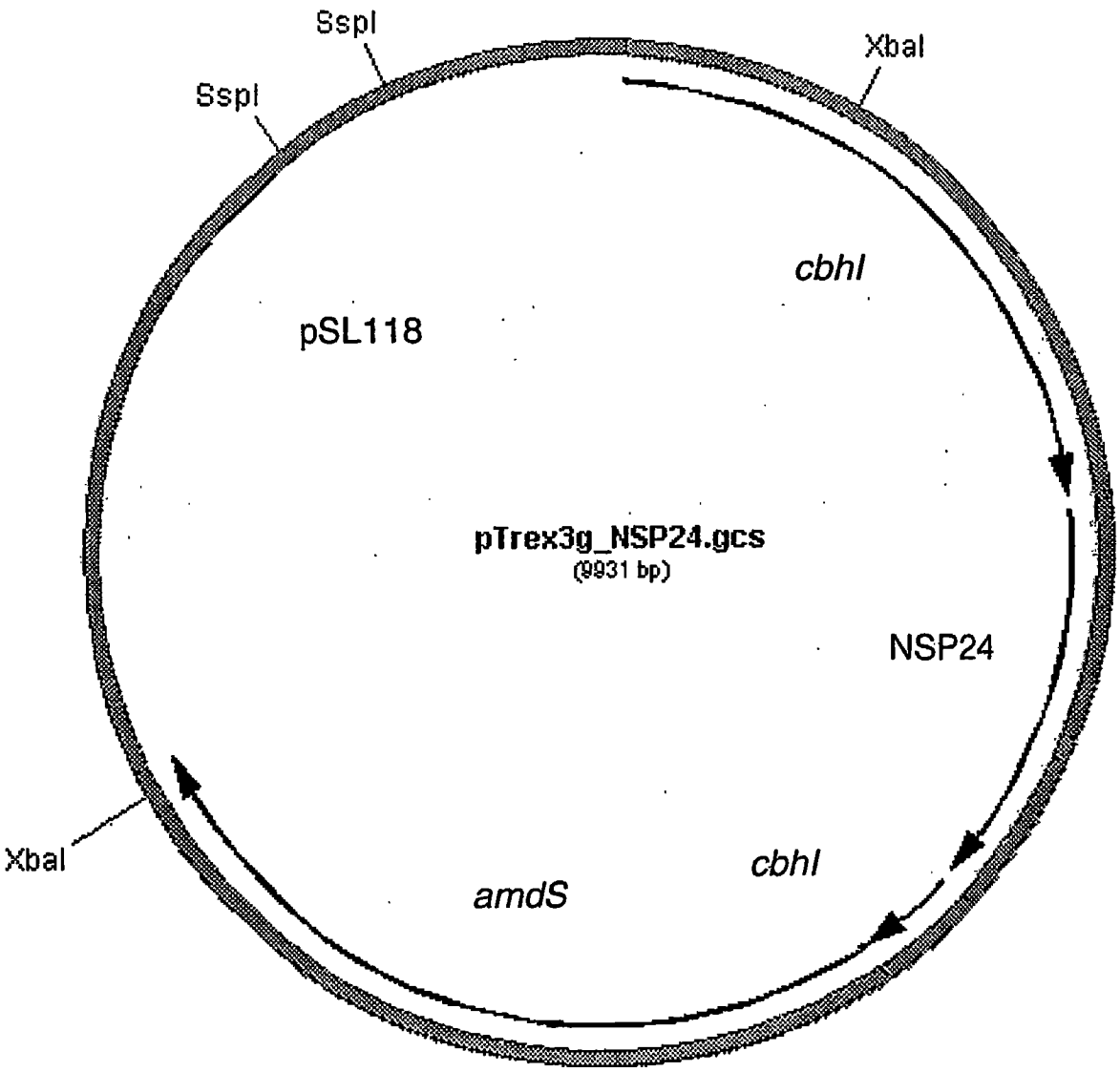


图 7

(SEQ ID NO: 3)-编码pepA蛋白酶的核酸序列，含粗体表示的推定内含子

ATGGTCGTCTTCAGCAAAACCGCTGCCCTCGTTCTGGGTCTGTCCTCCGCCGTCTCTGCGGC
GCCGGCTCCTACTCGCAAGGGCTTCACCATCAACCAGATTGCCGGGCCTGCCAACAAGACCC
GCACCATCAACCTGCCAGGCATGTACGCCCGTTCCCTGGCCAAGTTTGGCGGTACGGTGCCC
CAGAGCGTGAAGGAGGCTGCCAGCAAGGGTAGTGCCGTGACCACGCCCCAGAACAATGACGA
GGAGTACCTGACTCCCGTCACTGTCGGAAAGTCCACCCTCCATCTGGACTTTGACACCGGAT
CTGCAGATCT**GTAAGCTTCCCTGCTCGGGTGTTCGGGCAAATCGTGACTAACCTGGACTAG**C
TGGGTCTTCTCGGACGAGCTCCCTTCCTCGGAGCAGACCGGTCACGATCTGTACACGCCTAG
CTCCAGCGCGACCAAGCTGAGCGGCTACACTTGGGACATCTCCTACGGTGACGGCAGCTCGG
CCAGCGGAGACGTGTACCGGGATACTGTCACTGTCGGCGGTGTCACCACCAACAAGCAGGCT
GTTGAAGCAGCCAGCAAGATCAGCTCCGAGTTCGTTCAGAACACGGCCAATGACGGCCTTTT
GGGACTGGCCTTTAGCTCCATCAACACTG**GTGAGTCAATCCTACATCAGCCGGGTTGACCTA**
CCTGCTGACCGATACACAGTCCAGCCCAAGGCGCAGACCACCTTCTTCGACACCGTCAAGTC
CCAGCTGGACTCTCCCCTTTTCGCCGTGCAGCTGAAGCACGACGCCCCCGGTGTTTACGACT
TTGGCTACATCGATGACTCCAAGTACACCGGTTCTATCACCTACACGGATGCCGATAGCTCC
CAGGGTTACTGGGGCTTCAGCACCGACGGCTACAGTATCGGTGACGGCAGCTCCAGCTCCAG
CGGCTTCAGCGCCATTGCTG**GTAAGAACCGCCTTCATTTAACACACAACTTGTCCACCTCTT**
TACTAACTAGTGTATAGACACCGGTACCACCCTCATCCTCCTCGATGACGAAATCGTCTCCG
CCTACTACGAGCAGGTTTCTGGCGCTCAGGAGAGCGAGGAAGCCGGTGGCTACGTTTTCTCT
TGCTCGACCAACCCCCCTGACTTCACTGTCGTGATTGGCGACTACAAGGCCGTTGTTCCGGG
CAAGTACATCAACTACGCTCCCATCTCGACTGGCAGCTCCACCTGCTTTGGCGGTATCCAGA
GCAACAGCGGTCTGGGACTGTCCATCCTGGGTGATGTTTTCTTGAAGAGCCAGTACGTGGTC
TTCAACTCTGAGGGCCCTAAGCTGGGATTCGCCGCTCAGGCTTAG

图 8

(SEQ ID NO: 4) 编码NSP25蛋白酶的核酸序列

ATGCAGCCCTCATTTGGCAGCTTCCTCGTCACCGTCCTGTCTGCCTCCATGGCAGCAGGCAG
TGTCATTCCCAGCACAAACGCCAACCCTGGCTCCTTCGAGATCAAGAGATCCGCCAACAAAG
CCTTCACAGGCCGCAATGGCCCTCTAGCATTAGCCCGTACATACGCCAAGTACGGTGTTGAA
GTCCCCAAAACTCTGGTCGATGCTATTCAACTCGTCAAGTCCATCCAGCTCGCAAAGCGGGA
CAGCGCCACCGTCACTGCCACGCCGGACCACGACGACATCGAGTATCTTGTCCCCGTCAAGA
TCGGAACTCCTCCCCAAACACTTAACCTGGATTTTGACACGGGCAGCTCCGATCTCTGGGTC
TTCTCATCAGATGTCGACCCGACCTCCTCCCAGGGCCATGACATCTACACCCCGTCCAAGAG
CACATCTTCCAAAAAGTTGGAAGGAGCCTCATGGAACATCACATATGGAGACCGCTCATCAT
CATCCGGCGATGTCTACCACGATATTGTCTCCGTCGGAAACCTGACAGTAAAGTCCCAAGCC
GTCGAGTCCGCTCGAAACGTCTCGGCCAGTTCACCCAGGGCAACAACGACGGCCTCGTCGGC
CTGGCGTTTAGCTCCATCAACACAGTCAAGCCCACGCCGCAAAAGACGTGGTACGACAACAT
CGTCGGCAGCCTTGACTCTCCCGTCTTTGTTGCTGATCTGCGCCACGACACGCCCGGCAGCT
ACCACTTCGGCTCCATCCCCTCCGAAGCAAGCAAAGCCTTCTACGCCCCCATCGACAACAGC
AAGGGCTTCTGGCAATTCAGCACGAGCAGCAACATTAGCGGCCAGTTCAACGCCGTTGCAGA
CACTGGCACTACTCTGCTGCTCGCCAGCGACGACCTCGTCAAGGCCTACTACGCAAAGGTCC
AGGGCGCCCGTGTGAACGTCTTCCTGGGCGGCTACGTCTTCAACTGCACCACTCAGCTGCCC
GACTTTACCTTTACTGTTGGAGAGGGCAACATCACTGTCCCCGGTACCTTGATAAACTATTC
CGAGGCTGGCAACGGCCAGTGTTTTGGCGGTATTCAGCCGTCGGGGGGTCTTCCTTTTGCTA
TCTTTGGTGACATTGCTCTTAAGGCTGCGTATGTTATTTTTGACAGTGGCAACAAGCAGGTT
GGCTGGGCGCAGAAGAAATAG

图 9A

B. SEQ ID NO: 9 - 推导的NSP25氨基酸序列(399个氨基酸)

MQPSFGSFLVTVLSASMAAGSVIPSTNANPGSFEIKRSANKAFTGRNGPLALARTYAKYG
VEVPKTLVDAIQLVKSIQLAKRDSATVTATPDHDDIEYLVPVKIGTPPQTLNLDFDTGSS
DLWVFSSDVDPTSSQGHDIYTPSKSTSSKKLEGASWNITYGDRSSSSGDVYHDIVSVGNL
TVKSQAVESARNVSXQFTQGNNDGLVGLAFSSINTVKPTPQKTWYDNIVGSLDSPVFVAD
LRHDTPGSYHFGSIPSEASKAFYAPIDNSKGFWQFSTSSNISGQFNAVADTGTTLLLASD
DLVKAYYAKVQGARVNVFLGGYVFNCTTQLPDFTFTVGEGNITVPGTLINYSEAGNGQCF
GGIQPSGGLPFAIFGDIALKAAYVIFDSGNKQVGWAQKK

图 9B

SEQ ID NO: 5 - 新型pepA蛋白酶(L388M)的核酸序列

ATGGTCGTCTTCAGCAAAACCGCTGCCCTCGTTCTGGGTCTGTCCTCCGCCGTCTCTGCGGC
GCCGGCTCCTACTCGCAAGGGCTTCACCATCAACCAGATTGCCCGGCCTGCCAACAAGACCC
GCACCATCAACCTGCCAGGCATGTACGCCCGTTCCCTGGCCAAGTTTGGCGGTACGGTGCCC
CAGAGCGTGAAGGAGGCTGCCAGCAAGGGTAGTGCCGTGACCACGCCCCAGAACAATGACGA
GGAGTACCTGACTCCCGTCACTGTCGGAAAGTCCACCCTCCATCTGGACTTTGACACCGGAT
CTGCAGATCT**GTAAGCTTCCCTGCTCGGGTGTTCGGGCAAATCGTGACTAACCTGGACTAGC**
TGGGTCTTCTCGGACGAGCTCCCTTCCTCGGAGCAGACCGGTCACGATCTGTACACGCCTAG
CTCCAGCGCGACCAAGCTGAGCGGCTACACTTGGGACATCTCCTACGGTGACGGCAGCTCGG
CCAGCGGAGACGTGTACCGGGATACTGTCACTGTCGGCGGTGTCACCACCAACAAGCAGGCT
GTTGAAGCAGCCAGCAAGATCAGCTCCGAGTTCGTTCAGAACACGGCCAATGACGGCCTTTT
GGGACTGGCCTTTAGCTCCATCAACACTG**GTGAGTCAATCCTACATCAGCCGGGTTGACCTA**
CCTGCTGACCGATACACAGTCCAGCCCAAGGCGCAGACCACCTTCTTCGACACCGTCAAGTC
CCAGCTGGACTCTCCCCTTTTCGCCGTGCAGCTGAAGCACGACGCCCCCGGTGTTTACGACT
TTGGCTACATCGATGACTCCAAGTACACCGGTTCTATCACCTACACGGATGCCGATAGCTCC
CAGGGTTACTGGGGCTTCAGCACCGACGGCTACAGTATCGGTGACGGCAGCTCCAGCTCCAG
CGGCTTCAGCGCCATTGCTG**GTAAGAACCGCCTTCATTTAACACACAACTTGTCCACCTCTT**
TACTAACTAGTGTATAGACACCGGTACCACCCTCATCCTCCTCGATGACGAAATCGTCTCCG
CCTACTACGAGCAGGTTTCTGGCGCTCAGGAGAGCGAGGAAGCCGGTGGCTACGTTTTCTCT
TGCTCGACCAACCCCCCTGACTTCACTGTCGTGATTGGCGACTACAAGGCCGTTGTTCCGGG
CAAGTACATCAACTACGCTCCCATCTCGACTGGCAGCTCCACCTGCTTTGGCGGTATCCAGA
GCAACAGCGGTCTGGGACTGTCCATCCTGGGTGATGTTTTCTTGAAGAGCCAGTACGTGGTC
TTCAACTCTGAGGGCCCTAAGATGGGATTCGCCGCTCAGGCTTAG

图 10

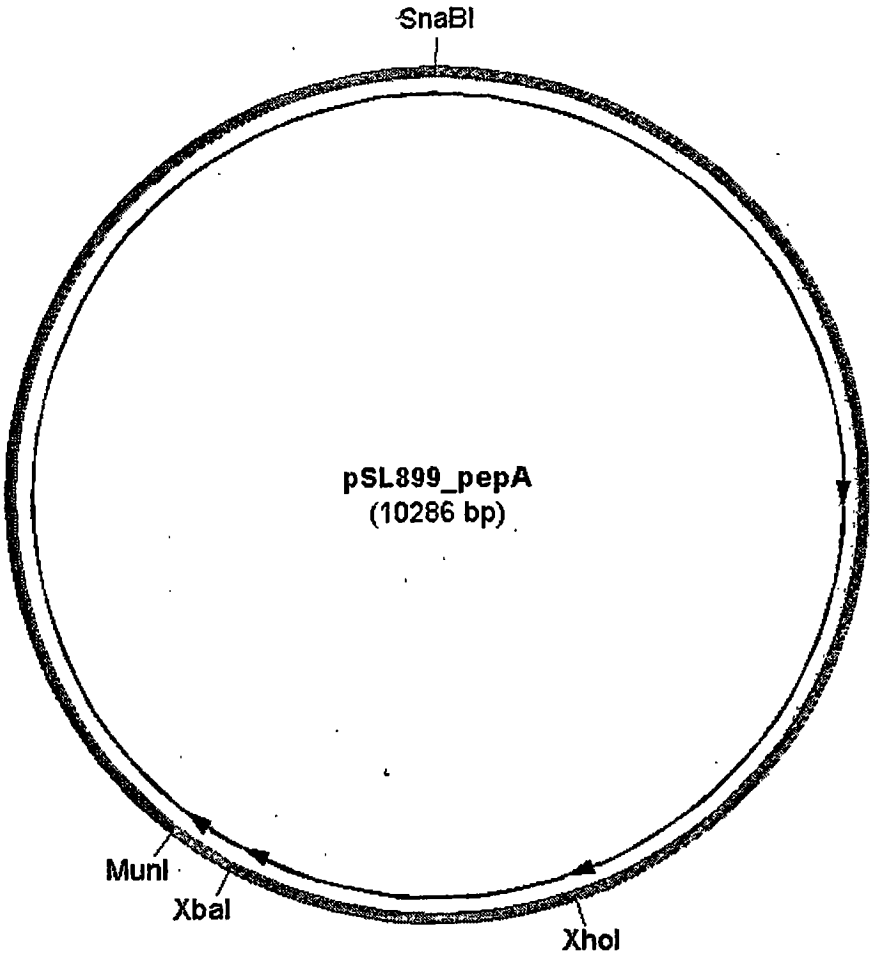


图 11

SEQ ID NO: 6

TACGTATTTT GAATAGCTCG CCCGCTGGAG AGCATCCTGA ATGCAAGTAA
CAACCGTAGA GGCTGACACG GCAGGTGTTG CTAGGGAGCG TCGTGTTCTA
CAAGGCCAGA CGTCTTCGCG GTTGATATAT ATGTATGTTT GACTGCAGGC
TGCTCAGCGA CGACAGTCAA GTTCGCCCTC GCTGCTTGTG CAATAATCGC
AGTGGGGAAG CCACACCGTG ACTCCCATCT TTCAGTAAAG CTCTGTTGGT
GTTTATCAGC AATACACGTA ATTTAAACTC GTTAGCATGG GGCTGATAGC
TTAATTACCG TTTACCAGTG CCGCGGTTCT GCAGCTTTCC TTGGCCCGTA
AAATTCGGCG AAGCCAGCCA ATCACCAGCT AGGCACCAGC TAAACCCTAT
AATTAGTCTC TTATCAACAC CATCCGCTCC CCCGGGATCA ATGAGGAGAA
TGAGGGGGAT GCGGGGCTAA AGAAGCCTAC ATAACCCTCA TGCCAACTCC
CAGTTTACAC TCGTCGAGCC AACATCCTGA CTATAAGCTA ACACAGAATG
CCTCAATCCT GGGAAGAACT GGCCGCTGAT AAGCGCGCCC GCCTCGCAAA
AACCATCCCT GATGAATGGA AAGTCCAGAC GCTGCCTGCG GAAGACAGCG
TTATTGATTT CCCAAAGAAA TCGGGGATCC TTTCAGAGGC CGAACTGAAG
ATCACAGAGG CCTCCGCTGC AGATCTTGTG TCCAAGCTGG CGGCCGGAGA
GTTGACCTCG GTGGAAGTTA CGCTAGCATT CTGTAAACGG GCAGCAATCG
CCCAGCAGTT AGTAGGGTCC CCTCTACCTC TCAGGGAGAT GTAACAACGC
CACCTTATGG GACTATCAAG CTGACGCTGG CTTCTGTGCA GACAAACTGC
GCCCACGAGT TCTTCCCTGA CGCCGCTCTC GCGCAGGCAA GGGAACTCGA
TGAATACTAC GCAAAGCACA AGAGACCCGT TGGTCCACTC CATGGCCTCC
CCATCTCTCT CAAAGACCAG CTTCGAGTCA AGGTACACCG TTGCCCCTAA
GTCGTTAGAT GTCCCTTTTT GTCAGCTAAC ATATGCCACC AGGGCTACGA
AACATCAATG GGCTACATCT CATGGCTAAA CAAGTACGAC GAAGGGGACT
CGGTTCTGAC AACCATGCTC CGCAAAGCCG GTGCCGTCTT CTACGTCAAG
ACCTCTGTCC CGCAGACCCT GATGGTCTGC GAGACAGTCA ACAACATCAT
CGGGCGCACC GTCAACCCAC GCAACAAGAA CTGGTCGTGC GGCGGCAGTT
CTGGTGGTGA GGGTGCGATC GTTGGGATTC GTGGTGGCGT CATCGGTGTA
GGAACGGATA TCGGTGGCTC GATTCGAGTG CCGGCCGCGT TCAACTTCCT
GTACGGTCTA AGGCCGAGTC ATGGGCGGCT GCCGTATGCA AAGATGGCGA
ACAGCATGGA GGGTCAGGAG ACGGTGCACA GCGTTGTCGG GCCGATTACG
CACTCTGTTG AGGGTGAGTC CTTCGCCTCT TCCTTCTTTT CCTGCTCTAT
ACCAGGCCTC CACTGTCCTC CTTTCTTGCT TTTTATACTA TATACGAGAC
CGGCAGTCAC TGATGAAGTA TGTTAGACCT CCGCCTCTTC ACCAAATCCG
TCCTCGGTCA GGAGCCATGG AAATACGACT CCAAGGTCAT CCCCATGCCC
TGGCGCCAGT CCGAGTCGGA CATTATTGCC TCCAAGATCA AGAACGGCGG
GCTCAATATC GGCTACTACA ACTTCGACGG CAATGTCCTT CCACACCCTC
CTATCCTGCG CGGCGTGGAA ACCACCGTCG CCGCACTCGC CAAAGCCGGT
CACACCGTGA CCCCGTGGAC GCCATACAAG CACGATTTCG GCCACGATCT
CATCTCCCAT ATCTACGCGG CTGACGGCAG CGCCGACGTA ATGCGCGATA
TCAGTGCATC CGGCGAGCCG GCGATTCCAA ATATCAAAGA CCTACTGAAC
CCGAACATCA AAGCTGTTAA CATGAACGAG CTCTGGGACA CGCATCTCCA
GAAGTGGAAT TACCAGATGG AGTACCTTGA GAAATGGCGG GAGGCTGAAG
AAAAGGCCGG GAAGGAACTG GACGCCATCA TCGCGCCGAT TACGCCTACC
GCTGCGGTAC GGCATGACCA GTTCCGGTAC TATGGGTATG CCTCTGTGAT
CAACCTGCTG GATTTCACGA GCGTGGTTGT TCCGGTTACC TTTGCGGATA

图 12A

AGAACATCGA TAAGAAGAAT GAGAGTTTCA AGGCGGTTAG TGAGCTTGAT
GCCCTCGTGC AGGAAGAGTA TGATCCGGAG GCGTACCATG GGGCACCGGT
TGCAGTGCAG GTTATCGGAC GGAGACTCAG TGAAGAGAGG ACGTTGGCGA
TTGCAGAGGA AGTGGGGAAG TTGCTGGGAA ATGTGGTGAC TCCATAGCTA
ATAAGTGTCA GATAGCAATT TGCACAAGAA ATCAATACCA GCAACTGTAA
ATAAGCGCTG AAGTGACCAT GCCATGCTAC GAAAGAGCAG AAAAAAACCT
GCCGTAGAAC CGAAGAGATA TGACACGCTT CCATCTCTCA AAGGAAGAAT
CCCTTCAGGG TTGCGTTTCC AGTCTAGCTA GAGTCGAGGA TTGCCTGAAC
ATTGACATTC GGCGTCCGGC CGGGACCACC GCGGACTCGA AGCTGCCTGT
GCTGGTCTGG ATCTTTGGCG GAGGCTTTGA ACTTGGTTCA AAGGCGATGT
ATGATGGTAC AACGATGGTA TCATCGTCGA TAGACAAGAA CATGCCTATC
GTGTTTGTAG CAATGAATTA TCGCGTGGGA GGTTTCGGGT TCTTGCCCGG
AAAGGAGATC CTGGAGGACG GGTCCGCGAA CCTAGGGCTC CTGGACCAAC
GCCTTGCCCT GCAGTGGGTT GCCGACAACA TCGAGGCCTT TGGTGGAGAC
CCGGACAAGG TGACGATTTG GGGAGAATCA GCAGGAGCCA TTTCCGTTTT
TGATCAGATG ATCTTGTACG ACGGAAACAT CACTTACAAG GATAAGCCCT
TGTTCCGGGG GGCCATCATG GACTCCGGTA GTGTTGTTCC CGCAGACCCC
GTCGATGGGG TCAAGGGACA GCAAGTATAT GATGCGGTAG TGGAATCTGC
AGGCTGTTCC TCTTCTAACG ACACCCTAGC TTGTCTGCGT GAACTAGACT
ACACCGACTT CCTCAATGCG GCAAACTCCG TGCCAGGCAT TTTAAGCTAC
CATTCTGTGG CGTTATCATA TGTGCCTCGA CCGGACGGGA CGGCGTTGTC
GGCATCACCG GACGTTTTGG GCAAAGCAGG GAAATATGCT CGGGTCCCGT
TCATCGTGGG CGACCAAGAG GATGAGGGGA CCTTATTCGC CTTGTTTCAG
TCCAACATTA CGACGATCGA CGAGGTGGTC GACTACCTGG CCTCATACTT
CTTCTATGAC GCTAGCCGAG AGCAGCTTGA AGAACTAGTG GCCCTGTACC
CAGACACCAC CACGTACGGG TCTCCGTTCA GGACAGGCGC GGCCAACAAC
TGGTATCCGC AATTTAAGCG ATTGGCCGCC ATTCTCGGCG ACTTGGTCTT
CACCATTACC CGGCGGGCAT TCCTCTCGTA TGCAGAGGAA ATCTCCCCTG
ATCTTCCGAA CTGGTCGTAC CTGGCGACCT ATGACTATGG CACCCCAGTT
CTGGGGACCT TCCACGGAAG TGACCTGCTG CAGGTGTTCT ATGGGATCAA
GCCAAACTAT GCAGCTAGTT CTAGCCACAC GTACTATCTG AGCTTTGTGT
ATACGCTGGA TCCGAACTCC AACCGGGGGG AGTACATTGA GTGGCCGCAG
TGGAAGGAAT CGCGGCAGTT GATGAATTTC GGAGCGAACG ACGCCAGTCT
CCTTACGGAT GATTTCCGCA ACGGGACATA TGAGTTCATC CTGCAGAATA
CCGCGGCGTT CCACATCTGA TGCCATTGGC GGAGGGGTCC GGACGGTCAG
GAACTTAGCC TTATGAGATG AATGATGGAC GTGTCTGGCC TCGGAAAAGG
ATATATGGGG ATCATGATAG TACTAGCCAT ATTAATGAAG GGCATATACC
ACGCGTTGGA CCTGCGTTAT AGCTTCCCGT TAGTTATAGT ACCATCGTTA
TACCAGCCAA TCAAGTCACC ACGCACGACC GGGGACGGCG AATCCCCGGG
AATTGAAAGA AATTGCATCC CAGGCCAGTG AGGCCAGCGA TTGGCCACCT
CTCCAAGGCA CAGGGCCATT CTGCAGCGCT GGTGGATTCA TCGCAATTTC
CCCCGGCCCG GCCCGACACC GCTATAGGCT GGTTCTCCCA CACCATCGGA
GATTCGTCGC CTAATGTCTC GTCCGTTCAC AAGCTGAAGA GCTTGAAGTG
GCGAGATGTC TCTGCAGGAA TTCAAGCTAG ATGCTAAGCG ATATTGCATG
GCAATATGTG TTGATGCATG TGCTTCTTCC TTCAGCTTCC CCTCGTGCAG

图 12B

```
ATGAGGTTTG GCTATAAATT GAAGTGGTTG GTCGGGGTTC CGTGAGGGGC
TGAAGTGCTT CCTCCCTTTT AGACGCAACT GAGAGCCTGA GCTTCATCCC
CAGCATCATT ACACCTCGAG ATGGTCGTCT TCAGCAAAAC CGCTGCCCTC
                  ^^^^^^
                  Xho I
GTTCTGGGTC TGTCCTCCGC CGTCTCTGCG GCGCCGGCTC CTACTCGCAA
GGGCTTCACC ATCAACCAGA TTGCCCGGCC TGCCAACAAG ACCCGCACCA
TCAACCTGCC AGGCATGTAC GCCCGTTCCC TGGCCAAGTT TGGCGGTACG
GTGCCCCAGA GCGTGAAGGA GGCTGCCAGC AAGGGTAGTG CCGTGACCAC
GCCCCAGAAC AATGACGAGG AGTACCTGAC TCCCGTCACT GTCGGAAAGT
CCACCCTCCA TCTGGACTTT GACACCGGAT CTGCAGATCT GTAAGCTTCC
CTGCTCGGGT GTTCGGGCAA ATCGTGACTA ACCTGGACTA GCTGGGTCTT
CTCGGACGAG CTCCCTTCCT CGGAGCAGAC CGGTCACGAT CTGTACACGC
CTAGCTCCAG CGCGACCAAG CTGAGCGGCT ACACTTGGGA CATCTCCTAC
GGTGACGGCA GCTCGGCCAG CGGAGACGTG TACCGGGATA CTGTCACTGT
CGGCGGTGTC ACCACCAACA AGCAGGCTGT TGAAGCAGCC AGCAAGATCA
GCTCCGAGTT CGTTCAGAAC ACGGCCAATG ACGGCCTTTT GGGACTGGCC
TTTAGCTCCA TCAACACTGG TGAGTCAATC CTACATCAGC CGGGTTGACC
TACCTGCTGA CCGATACACA GTCCAGCCCA AGGCGCAGAC CACCTTCTTC
GACACCGTCA AGTCCCAGCT GGACTCTCCC CTTTTCGCCG TGCAGCTGAA
GCACGACGCC CCCGGTGTTT ACGACTTTGG CTACATCGAT GACTCCAAGT
ACACCGGTTC TATCACCTAC ACGGATGCCG ATAGCTCCCA GGGTTACTGG
GGCTTCAGCA CCGACGGCTA CAGTATCGGT GACGGCAGCT CCAGCTCCAG
CGGCTTCAGC GCCATTGCTG GTAAGAACCG CCTTCATTTA ACACACAACT
TGTCCACCTC TTTACTAACT AGTGTATAGA CACCGGTACC ACCCTCATCC
TCCTCGATGA CGAAATCGTC TCCGCCTACT ACGAGCAGGT TTCTGGCGCT
CAGGAGAGCG AGGAAGCCGG TGGCTACGTT TTCTCTTGCT CGACCAACCC
CCCTGACTTC ACTGTCGTGA TTGGCGACTA CAAGGCCGTT GTTCCGGGCA
AGTACATCAA CTACGCTCCC ATCTCGACTG GCAGCTCCAC CTGCTTTGGC
GGTATCCAGA GCAACAGCGG TCTGGGACTG TCCATCCTGG GTGATGTTTT
CTTGAAGAGC CAGTACGTGG TCTTCAACTC TGAGGGCCCT AAGCTGGGAT
TCGCCGCTCA GGCTTAGTCT AGAGTCGACC GCGACGGTGA CCGACACCTG
                  *** ***
                  XbaI
GCGGTAGACT ATTTATTCCT GTTGATATGA AGGATGAGCA TGAGGGTAAT
TGCTCATATA ATCATGTATG TAGTGGATGT GCATAAGAGC AACGAAATGG
AAGCCTGATC ATGTGATTGT ATTGCGACCG ACGGAATTGA GGATATGCGG
AGATACGGAC AGTGCCAGAG CCATTGTCTT CACGTAAAGT ACCAGACGGT
CCCTGATTTC TTCTTGCACA TAGCATTAGG CAATTGACAT GTTGTCGCTC
TACTGATATC ACTGTCCCTC AAAGCATAGC CATGAGCTCA TCTTAGATCC
AAGCACGTAA TTCCATAGCC GAGGTCCACA GTGGAGCAAC AGCAGCATCC
ATCATTGCTT CTCCCCCAGG GGCCTCTTAG CGACTAAACC TGGAGTATGT
CTCAACCAGC CAATGAATCG TCTTCGCTTC AATGTCCTTG ACACTTCTGA
GAGGGTCCCC ATCCCTCAAT GCTAATTCAA AATATAGCCG AGATGCATGG
TGGAGTCCAA AGTAGACAGT ATTGCCGGAA TGACGGGGCC AGTTGCGCCG
```

图 12C

AGGTCATTGG CCGGCTGTGA TGCCATCTGC CACTAAATCC GATCATTGAT
CCACCGCCCA CGAGGGCCGT CTTTGCTTTT GCGCTGCGTC CAGGTTCACA
CATCTCTCTC TCTGCAGCTC CAGACTGACC AGACTATTCT ACTTACTGGT
CTGATCGGCT CCATCAGAGC TATGGCGTTA TCCCGTGCCG TTGCTGCGCC
ATGGCTATCT TGATCGCGAG CTCGAACTCA CTTCTTGTTT TAATAGTTGT
TCTCGGTGAC TGAGTGTCGG TGAGTGACAG ACCACAACAC CATTGTTGCA
GGGGGTAAAT TTATTCAATT CAGGAATTGG ATTGTTCGTC CCGCCATGAT
GTTCTTGCCG GCTTTGTTGG CCCTGTTTGT CGGATGCGAC GCCCTCGCTG
TGCAGCAGGC AGGTACTGCT GGATGATGAG CCGTCGGTCT CCGCGCGCAA
GCCTAACTTC CTCTTCATTC TTACGGATGA TCAGGATCTG CAGATCGAAT
TCCACCGGCG TATATGCCGT ATACACAGGC GAGAATCAAG GAGAAGGGTA
CTGAGTTTTG AATCATTTGT TACTACTGGC TCTGTGCTGT CCGTCGCGCG
TGAGTCTTTG GACGGAAGAC AGGCTCATAA TACTAATGTG ACGGATGTGA
ACCCGCCTTA TGGTATGAAT ACCTCTCAGA TCGGTCATGT TTCTTCGGTG
TAAAATTGCT AATGCAGCAT AGGCGGATAC CCCAAGTTCG TCGCCCAAGG
CTTCAACGAA AACTTCCTCC CCGTTTGGCT GCAGTCCGCC GGTTACAATA
CCTTCTACAC GGGGAAACTG TTCAACTGCC ACAGCGTCGC TACCTATAAT
GCACCGTTTG TGAACGGCTT CAATGGCTCT GATTTCCTCC TCGATCCCCA
CACCTATTCC TACTGGAACG CGACGTACCA ACGAAACCAT GAGCCTCCGC
GGAGCTACGA GGGACAATAC ACAACGGATG TGATGCGGGA GAAGGCATCG
GGGTTGTTGG CAGATGCGCT GGACAGGACG CGCCGTTCTT CTGACGGTGC
CTATCCGCCG CACACGAACA TCGACCTGCA GGCATGCAAG CTTGGCACTG
GCCGTCGTTT TACAACGTCG TGACTGGGAA AACCCTGGCG TTACCCAACT
TAATCGCCTT GCAGCACATC CCCCTTTCGC CAGCTGGCGT AATAGCGAAG
AGGCCCGCAC CGATCGCCCT TCCCAACAGT TGCGCAGCCT GAATGGCGAA
TGGCGCCTGA TGCGGTATTT TCTCCTTACG CATCTGTGCG GTATTTCACA
CCGCATATGG TGCACTCTCA GTACAATCTG CTCTGATGCC GCATAGTTAA
GCCAGCCCCG ACACCCGCCA ACACCCGCTG ACGCGCCCTG ACGGGCTTGT
CTGCTCCCGG CATCCGCTTA CAGACAAGCT GTGACCGTCT CCGGGAGCTG
CATGTGTCAG AGGTTTTCAC CGTCATCACC GAAACGCGCG AGACGAAAGG
GCCTCGTGAT ACGCCTATTT TTATAGGTTA ATGTCATGAT AATAATGGTT
TCTTAGACGT CAGGTGGCAC TTTTCGGGGA AATGTGCGCG GAACCCCTAT
TTGTTTATTT TTCTAAATAC ATTCAAATAT GTATCCGCTC ATGAGACAAT
AACCCTGATA AATGCTTCAA TAATATTGAA AAAGGAAGAG TATGAGTATT
CAACATTTCC GTGTCGCCCT TATTCCCTTT TTTGCGGCAT TTTGCCTTCC
TGTTTTTGCT CACCCAGAAA CGCTGGTGAA AGTAAAAGAT GCTGAAGATC
AGTTGGGTGC ACGAGTGGGT TACATCGAAC TGGATCTCAA CAGCGGTAAG
ATCCTTGAGA GTTTTCGCCC CGAAGAACGT TTTCCAATGA TGAGCACTTT
TAAAGTTCTG CTATGTGGCG CGGTATTATC CCGTATTGAC GCCGGGCAAG
AGCAACTCGG TCGCCGCATA CACTATTCTC AGAATGACTT GGTTGAGTAC
TCACCAGTCA CAGAAAAGCA TCTTACGGAT GGCATGACAG TAAGAGAATT
ATGCAGTGCT GCCATAACCA TGAGTGATAA CACTGCGGCC AACTTACTTC
TGACAACGAT CGGAGGACCG AAGGAGCTAA CCGCTTTTTT GCACAACATG
GGGGATCATG TAACTCGCCT TGATCGTTGG GAACCGGAGC TGAATGAAGC
CATACCAAAC GACGAGCGTG ACACCACGAT GCCTGTAGCA ATGGCAACAA

图 12D

CGTTGCGCAA ACTATTAACT GGCGAACTAC TTACTCTAGC TTCCCGGCAA
CAATTAATAG ACTGGATGGA GGCGGATAAA GTTGCAGGAC CACTTCTGCG
CTCGGCCCTT CCGGCTGGCT GGTTTATTGC TGATAAATCT GGAGCCGGTG
AGCGTGGGTC TCGCGGTATC ATTGCAGCAC TGGGGCCAGA TGGTAAGCCC
TCCCGTATCG TAGTTATCTA CACGACGGGG AGTCAGGCAA CTATGGATGA
ACGAAATAGA CAGATCGCTG AGATAGGTGC CTCACTGATT AAGCATTGGT
AACTGTCAGA CCAAGTTTAC TCATATATAC TTTAGATTGA TTTAAAACTT
CATTTTTAAT TTAAAAGGAT CTAGGTGAAG ATCCTTTTTG ATAATCTCAT
GACCAAAATC CCTTAACGTG AGTTTTCGTT CCACTGAGCG TCAGACCCCG
TAGAAAAGAT CAAAGGATCT TCTTGAGATC CTTTTTTTCT GCGCGTAATC
TGCTGCTTGC AAACAAAAAA ACCACCGCTA CCAGCGGTGG TTTGTTTGCC
GGATCAAGAG CTACCAACTC TTTTTCCGAA GGTAACTGGC TTCAGCAGAG
CGCAGATACC AAATACTGTC CTTCTAGTGT AGCCGTAGTT AGGCCACCAC
TTCAAGAACT CTGTAGCACC GCCTACATAC CTCGCTCTGC TAATCCTGTT
ACCAGTGGCT GCTGCCAGTG GCGATAAGTC GTGTCTTACC GGGTTGGACT
CAAGACGATA GTTACCGGAT AAGGCGCAGC GGTCGGGCTG AACGGGGGGT
TCGTGCACAC AGCCCAGCTT GGAGCGAACG ACCTACACCG AACTGAGATA
CCTACAGCGT GAGCTATGAG AAAGCGCCAC GCTTCCCGAA GGGAGAAAGG
CGGACAGGTA TCCGGTAAGC GGCAGGGTCG GAACAGGAGA GCGCACGAGG
GAGCTTCCAG GGGGAAACGC CTGGTATCTT TATAGTCCTG TCGGGTTTCG
CCACCTCTGA CTTGAGCGTC GATTTTTGTG ATGCTCGTCA GGGGGGCGGA
GCCTATGGAA AAACGCCAGC AACGCGGCCT TTTTACGGTT CCTGGCCTTT
TGCTGGCCTT TTGCTCACAT GTTCTTTCCT GCGTTATCCC CTGATTCTGT
GGATAACCGT ATTACCGCCT TTGAGTGAGC TGATACCGCT CGCCGCAGCC
GAACGACCGA GCGCAGCGAG TCAGTGAGCG AGGAAGCGGA AGAGCGCCCA
ATACGCAAAC CGCCTCTCCC CGCGCGTTGG CCGATTCATT AATGCAGCTG
GCACGACAGG TTTCCCGACT GGAAAGCGGG CAGTGAGCGC AACGCAATTA
ATGTGAGTTA GCTCACTCAT TAGGCACCCC AGGCTTTACA CTTTATGCTT
CCGGCTCGTA TGTTGTGTGG AATTGTGAGC GGATAACAAT TTCACACAGG
AAACAGCTAT GACCATGATT ACGAATTCGA GCTCGG

图 12E

L388M pepA的氨基酸序列(SEQ ID NO: 7)

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

图 13