

[19] 中华人民共和国国家知识产权局

[51] Int. Cl⁷



[12] 发明专利说明书

专利号 ZL 98805175.3

C12N 15/57

C12N 9/48

C12N 1/15

C12N 1/19

C12P 21/06

A23J 3/30

[45] 授权公告日 2005 年 11 月 16 日

[11] 授权公告号 CN 1227363C

[22] 申请日 1998.5.12 [21] 申请号 98805175.3

[30] 优先权

[32] 1997.5.16 [33] US [31] 08/857,884

[32] 1997.10.20 [33] US [31] 06/062,892

[86] 国际申请 PCT/US1998/009629 1998.5.12

[87] 国际公布 WO1998/051803 英 1998.11.19

[85] 进入国家阶段日期 1999.11.16

[71] 专利权人 诺沃奇梅兹生物技术有限公司

地址 美国加利福尼亚

[72] 发明人 A·布林克沃斯凯 K·布朗

M·W·雷伊 A·科洛茨

T·博恩

审查员 唐 莉

[74] 专利代理机构 中国国际贸易促进委员会专利
商标事务所

代理人 樊卫民

权利要求书 2 页 说明书 72 页 附图 4 页

[54] 发明名称 具有脯氨酰二肽氨肽酶活性的多肽
和编码该多肽的核酸

[57] 摘要

本发明涉及从米曲霉分离的具有脯氨酰二肽氨肽酶活性的多肽，以及编码该多肽的分离的核酸序列。本发明也涉及包含该核酸序列的核酸构建体、载体和宿主细胞以及产生和使用这些多肽的方法。

I S S N 1 0 0 8 - 4 2 7 4

1. 一种具有二肽氨肽酶活性的分离的多肽，其含有 SEQ ID NO: 2 的序列或一个或多个 SEQ ID NO: 2 的保守性取代、插入和缺失。
2. 权利要求 1 的多肽，其含有 SEQ ID NO: 2 的氨基酸序列。
3. 权利要求 1 的多肽，其由 SEQ ID NO: 2 的氨基酸序列组成。
4. 权利要求 1 的多肽，其含有 SEQ ID NO: 2 的一个或多个保守性取代、插入和缺失。
5. 权利要求 1 的多肽，其具有下列物理化学性质：(i) 活性范围从 pH4.4-pH9.8，在 pH8.7 下有最佳活性，在 Ala-Pro-对硝基苯胺存在下于室温下温育 5 分钟后测定；(ii) pH7.5 时，相对于初始活性具有 90%或更高活性的温度稳定性，在不含底物时于 65°C 温育 20 分钟后测定；以及 (iii) 针对 Xaa-Pro-对硝基苯胺或 Xaa-Ala-对硝基苯胺的活性，其中 Xaa 选自 Ala、Arg、Asp、Gly 和 Val。
6. 权利要求 1 的多肽，它从曲霉属菌株中获得。
7. 权利要求 1 的多肽，它从米曲霉菌株中获得。
8. 权利要求 1 的多肽，它由大肠杆菌 NRRL B-21682 所含质粒 pMWR52 中包含的核酸序列所编码。
9. 权利要求 1 的多肽用于与一种氨肽酶协同作用水解蛋白质的用途。
10. 一种分离的核酸序列，其编码权利要求 1-9 中任一项的多肽。
11. 一种核酸构建体，其包含权利要求 10 的核酸序列，该核酸序列与一种或多种引导多肽在合适的表达宿主中产生的控制序列有效连接。
12. 一种重组表达载体，其包含权利要求 11 的核酸构建体、启动子和转录及翻译终止信号。
13. 一种包含权利要求 11 的核酸构建体的重组宿主细胞。
14. 一种产生具有二肽氨肽酶活性的多肽的方法，其包括 (a) 培养权利要求 13 的宿主细胞以产生具有二肽氨肽酶活性的多肽；和 (b)

回收该多肽。

15. 一种产生突变的核酸序列的方法，其包括(a)在 SEQ ID NO: 1 核酸序列中导入至少一个突变，其中该突变的核酸序列编码具有 SEQ ID NO: 2 的氨基酸序列，和(b)回收该突变的核酸序列。

16. 通过权利要求 15 的方法产生的突变核酸序列。

17. 一种产生具有二肽氨肽酶活性的多肽的方法，其包括(a)培养含有权利要求 16 的突变核酸序列的菌株以产生具有二肽氨肽酶活性的多肽；和(b)回收该多肽。

18. 一种产生亲本细胞突变体的方法，其包含在亲本细胞中破坏或删除编码权利要求 1 的多肽之核酸序列以产生突变细胞，其导致突变细胞产生比亲本细胞更少的具有二肽氨肽酶活性的多肽。

19. 由权利要求 18 的方法产生的突变体。

20. 一种由蛋白质底物产生水解产物的方法，其包括对底物施用权利要求 1 的多肽和一种内肽酶。

21. 权利要求 20 的方法，其中水解产物富含 Ala、Arg、Asp、Gly 和/或 Val。

22. 一种通过权利要求 21 的方法产生的蛋白质水解产物。

23. 一种含有权利要求 22 的蛋白质水解产物的食品。

24. 一种从蛋白质底物获得富含游离谷氨酸和/或肽结合谷氨酸残基的水解产物的方法，其包括对底物施行脱酰胺处理并施用权利要求 1 的多肽。

25. 权利要求 24 的方法，其进一步包括对底物施用一种或多种非特异作用的内肽酶和/或外肽酶。

26. 一种通过权利要求 25 的方法获得的蛋白质水解产物。

27. 一种含有权利要求 26 的蛋白质水解产物的食品。

28. 一种含有权利要求 1 的多肽和载体的增香组合物。

29. 一种用于生面团的预混合物，其含有权利要求 1 的多肽和焙烤配料。

具有脯氨酰二肽氨肽酶活性的多肽和编码该多肽的核酸

相关申请的交叉引用

本申请是于 1997 年 5 月 16 日提交的未决美国专利申请系列号 08/857,884 和于 1997 年 10 月 20 日提交的未决美国专利申请系列号 60/062,892 的部分继续申请, 这些申请在此完全引入作为参考。

发明背景

发明领域

本发明涉及具有二肽氨肽酶活性的多肽和编码该多肽的分离的核酸序列。本发明也涉及包含该核酸序列的核酸构建体、载体和宿主细胞以及产生该多肽的方法。本发明进一步涉及获得可用作增香剂的蛋白质水解产物的方法。

相关技术描述

多种食物产品如汤、酱油和调料含有通过蛋白质材料的水解获得的香味剂。通常使用强盐酸实现这种水解, 随后用氢氧化钠中和。然而, 这种化学水解导致水解中获得的氨基酸严重降解, 也导致在该化学反应过程中形成有害的副产物。对通过化学水解获得的增香剂应用的越来越多的关注导致酶水解方法的发展。

蛋白质材料的酶水解方法目的在于获得高度的水解 (DH), 通常使用非特异作用的蛋白水解酶 (即非特异作用的内肽酶和外肽酶) 的复合物来实现。例如, WO 94/25580 描述了一种利用从米曲霉 (*Aspergillus oryzae*) 中获得的非特异作用的酶制剂水解蛋白质的方法。特异作用的蛋白水解酶不用于此目的, 因为这些酶仅导致程度不足的水解。

具有二肽氨肽酶活性的多肽催化从肽、多肽和蛋白质的 N 端去除

二肽。这些多肽被归类于国际生物化学与分子生物学联合会 (International Union of Biochemistry and Molecular Biology) 的酶分类号 E.C. 3.4.14. 之下。

Beauvais 等人 (1997, 生物化学杂志 (Journal of Biological Chemistry) 272: 6238-6244) 公开了一种来源于烟曲霉 (*Aspergillus fumigatus*) 的二肽-肽酶, 根据 SDS-PAGE 其具有 88kDa 的分子量, 以及限于 X-Ala、His-Ser 和 Ser-Tyr 二肽在中性最适 pH 水解的底物特异性。Tachi 等人 (1992, 植物化学 (Phytochemistry) 31: 3707-3709) 公开了一种来源于米曲霉的 X-脯氨酰二肽氨肽酶, 根据 SDS-PAGE 其具有 145kDa 的分子量, 以及在中性最适 pH 时针对 N 端游离二肽和酰胺的倒数第二个位置中脯氨酸残基羧基位点处肽键的底物特异性。

具有希望的感官性质和高水解度的蛋白质水解产物的产生通常需要使用肽酶活性的混合物。提供单一组分的肽酶是所希望的, 该酶具有的活性可用于提高 (单独地或与其它酶结合地) 用在食物产品中的蛋白质水解产物的感官性质和水解程度。

本发明的一个目的在于提供具有二肽氨肽酶活性的改良的多肽, 以及获得具有希望的感官性质和高水解度的蛋白质水解产物的方法。

发明概述

本发明涉及具有二肽氨肽酶活性的分离的多肽, 其选自:

- (a) 一种多肽, 其具有与 SEQ ID NO: 2 的氨基酸序列至少有 50% 同一性的氨基酸序列;
- (b) 由一种核酸序列编码的多肽, 该核酸序列在中度严格条件下可与 (i) SEQ ID NO: 1 的核酸序列、(ii) 其互补链或 (iii) 其亚序列杂交;
- (c) (a) 或 (b) 的等位变体; 以及
- (d) (a)、(b) 或 (c) 的一种片段, 其中该片段具有二肽氨肽酶活性; 以及

- (e) 一种具有二肽氨肽酶活性的多肽，其具有下列物理化学性质：(i) 约 pH4.4-约 pH9.8 的最适 pH，在 Ala-Pro-对硝基苯胺存在下于环境温度下温育 5 分钟后测定；(ii) pH7.5 时 90%或更高的温度稳定性（相对于初始活性），在不含底物时于 65℃ 温育 20 分钟后测定；以及 (iii) 对 Xaa-Pro-对硝基苯胺或 Xaa-Ala-对硝基苯胺的活性，其中 Xaa 选自 Ala、Arg、Asp、Gly 和 Val。

本发明也涉及编码该多肽的分离的核酸序列，并涉及包含该核酸序列的核酸构建体、载体和宿主细胞以及产生该多肽的方法。

本发明也涉及从蛋白质底物获得水解产物的方法，其包括对蛋白质材料施用（单独地或与一种内肽酶结合地）具有二肽氨肽酶活性的多肽，并涉及通过该方法获得的水解产物。

本发明也涉及从蛋白质底物获得一种富含游离谷氨酸和/或肽结合谷氨酸残基的水解产物的方法，该方法包括对底物施行一种脱酰胺处理，以及经过具有二肽氨肽酶活性的多肽的处理。

本发明进一步涉及含有一种具有二肽氨肽酶活性的多肽的增香组合物。

在最后一方面，本发明的方法可用于食品相关应用如烘焙中以改进香味。另外，可通过添加由本发明的方法获得的水解产物实现食品中香味的改进。

附图简述

图 1 显示米曲霉 ATCC 20386 二肽氨肽酶的核酸序列和推定的氨基酸序列（分别为 SEQ ID NO: 1 和 SEQ ID NO: 2）。

图 2 显示 pDM181 的限制性酶切图。

图 3 显示 pMWR54 的限制性酶切图。

发明详述

具有二肽氨肽酶活性的多肽

术语“二肽氨肽酶活性”在此定义为从一种肽、多肽或蛋白质序列的N端切下二肽的肽酶活性。以广义方式定义，二肽氨肽酶能从肽、多肽或蛋白质的未取代N端氨基切下二肽XY，其中X或Y可代表选自Ala、Arg、Asn、Asp、Cys、Gln、Glu、Gly、His、Ile、Leu、Lys、Met、Phe、Pro、Ser、Thr、Trp、Tyr和Val的任何氨基酸残基，但至少是Ala、Arg、Asp、Gly和/或Val。所有X和Y既可以不同也可以相同。应当理解，本发明的具有二肽氨肽酶活性的分离的多肽对待切割的二肽的氨基酸序列可以是非特异的。

在第一种实施方案中，本发明涉及具有二肽氨肽酶活性的分离的多肽，该多肽具有与SEQ ID NO: 2的氨基酸序列有一定程度同一性的氨基酸序列，至少约50%，优选地至少约60%，优选地至少约70%，更优选地至少约80%，甚至更优选地至少约90%，最优选地至少约95%，甚至最优选地至少约97%（以下称“同源多肽”）。在一种优选实施方案中，同源多肽具有与SEQ ID NO: 2的氨基酸序列相差5个氨基酸、优选地4个氨基酸、更优选地3个氨基酸、甚至更优选地2个氨基酸、最优选地1个氨基酸的氨基酸序列。对于本发明而言，应用同一性表、罚区间为10、罚长度为10，通过Clustal法（Higgins, 1989, CABIOS 5: 151-153）测定两种氨基酸序列间同一性的程度。

优选地，本发明的多肽包含SEQ ID NO: 2的氨基酸序列或等位变体；以及其片段，其中该片段具有二肽氨肽酶活性。在一种更优选的实施方案中，本发明的多肽包含SEQ ID NO: 2的氨基酸序列。在另一种优选实施方案中，本发明的多肽具有SEQ ID NO: 2的氨基酸序列或其片段，其中该片段具有二肽氨肽酶活性。SEQ ID NO: 2的片段是从该氨基酸序列的氨基端和/或羧基端删除一个或多个氨基酸的一种多肽。在一种最优选的实施方案中，该多肽具有SEQ ID NO: 2的氨基酸序列。

优选地，一种片段含有至少455个氨基酸残基、更优选地至少555个氨基酸残基、最优选地至少655个氨基酸残基。

等位变体表示占据相同染色体位点的一种基因的任何两种或多种

备择形式。等位变异通过突变自然发生，可导致种群内的表型多态性。基因突变可以是沉默的（编码的多肽中无变化），或者可以编码具有改变的氨基酸序列的多肽。术语等位变体也用于表示由基因的等位变体所编码的蛋白质。

由于一个或多个氨基酸残基的插入或缺失和/或一个或多个氨基酸残基被不同氨基酸残基的置换，同源多肽的氨基酸序列可不同于 SEQ ID NO: 2 的氨基酸序列。优选地，氨基酸变化只有微小性质的变化，即不显著影响蛋白质的折叠和/或活性的保守性氨基酸置换；一般为 1 到约 30 个氨基酸的小量缺失；小量氨基端或羧基端延伸，如一个氨基端甲硫氨酸残基；可达约 20-25 个残基的小接头肽；或者通过改变净电荷或另一种功能如聚组氨酸序列段、抗原表位或结合结构域便于纯化的小量延伸。

保守性置换的实例在下列范围之内：碱性氨基酸（如精氨酸、赖氨酸和组氨酸）、酸性氨基酸（如谷氨酸和天冬氨酸）、极性氨基酸（如谷氨酰胺和天冬酰胺）、疏水性氨基酸（如亮氨酸、异亮氨酸和缬氨酸）、芳族氨基酸（如苯丙氨酸、色氨酸和酪氨酸）和小氨基酸（如甘氨酸、丙氨酸、丝氨酸、苏氨酸和甲硫氨酸）。一般不改变特异活性的氨基酸置换为本领域所知，描述于例如，H.Neurath 和 R.L.Hill, 1979, 《蛋白质》（The Proteins），Academic Press, 纽约。最常发生的交换是 Ala/Ser、Val/Ile、Asp/Glu、Thr/Ser、Ala/Gly、Ala/Thr、Ser/Asn、Ala/Val、Ser/Gly、Tyr/Phe、Ala/Pro、Lys/Arg、Asp/Asn、Leu/Ile、Leu/Val、Ala/Glu 和 Asp/Gly，以及相反的交换。

在第二种实施方案中，本发明涉及具有二肽氨肽酶活性的分离的多肽，其由在低度严格条件下、更优选地在中度严格条件下、最优选地在高度严格条件下可与一种寡核苷酸探针杂交的核酸序列所编码，该寡核苷酸探针在相同条件下可与 SEQ ID NO: 1 的核酸序列或其互补链杂交（J. Sambrook, E.F.Fritsch 和 T.Maniatus, 1989, 《分子克隆：实验室手册》（Molecular Cloning, A Laboratory Manual）第 2 版，冷泉港，纽约）；或者等位变体和该多肽的片段，其中该片段具

有二肽氨肽酶活性。

杂交表明，按照标准 Southern 印迹程序，在低度到高度的严格条件下（即，在 $5\times$ SSPE、0.3%SDS、 $200\mu\text{g/ml}$ 剪切并变性的鲑精 DNA 和对于低度、中度和高度严格条件分别为 25%、35%或 50%的甲酰胺中，于 42°C 预杂交和杂交），该核酸序列与对应于 SEQ ID NO: 1 所示核酸序列的多肽编码部分的寡核苷酸探针杂交。

SEQ ID NO: 2 的氨基酸序列或其部分序列可用来设计寡核苷酸探针，或者编码本发明的多肽的核酸序列，如 SEQ ID NO: 1 的核酸序列，或其亚序列，可用来根据本领域所周知的方法从不同属或种的菌株中鉴定并克隆编码具有二肽氨肽酶活性的多肽的 DNA。特别地，按照标准 Southern 印迹程序，这些探针能用于与目标属或种的基因组或 cDNA 杂交，以鉴定并分离其中相应的基因。这些探针可明显短于完整的序列，但长度至少应为 15 个、优选地至少 25 个、更优选地至少 40 个核苷酸。也可使用更长的探针。DNA 和 RNA 两种探针都能使用。为了检测相应的基因，探针一般被标记（例如，用 ^{32}P 、 ^3H 、 ^{35}S 、生物素或抗生物素蛋白标记）。

因此，可筛查由这些其它生物制备的基因组文库、cDNA 文库或组合化学文库，筛选与上述探针杂交并且编码具有二肽氨肽酶活性的多肽的 DNA。可通过琼脂糖或聚丙烯酰胺凝胶电泳或其它分离技术分离这些其它生物的基因组 DNA 或其它 DNA。可将来源于文库的 DNA 或分离的 DNA 转移并固定于硝酸纤维素膜或其它合适的载体材料上。为了鉴定一个克隆或与 SEQ ID NO: 1 同源 DNA，在 Southern 印迹中使用载体材料，其中最后用 $2\times$ SSC、0.2%SDS 将该载体材料洗涤 3 次，每次 30 分钟，洗涤温度优选地至少 50°C ，更优选地至少 55°C ，更优选地至少 60°C ，更优选地至少 65°C ，甚至更优选地至少 70°C ，最优选地至少 75°C 。用 X 光胶片检测在这些条件下与寡核苷酸探针杂交的分子。

在第三种实施方案中，本发明涉及具有下列物理化学性质的分离的多肽：（a）约 pH4.4 到约 pH9.8 的最适 pH，在 Ala-Pro-对硝基苯

胺存在下于环境温度下温育 5 分钟后测定；(b) pH7.5 时 90%或更高的温度稳定性（相对于初始活性），在不含底物时于 65℃温育 20 分钟后测定；以及(c) 针对 Xaa-Pro-对硝基苯胺或 Xaa-Ala-对硝基苯胺的活性，其中 Xaa 选自 Ala、Arg、Asp、Gly 和 Val。本发明的多肽也具有水解其它底物的能力。

在一种优选实施方案中，最适 pH 范围为约 pH4.4 到约 pH9.8，更优选地约 pH5.8 到约 pH9.8，最优选地约 pH7.5 到约 pH9.3，这是在 Ala-Pro-对硝基苯胺存在下于环境温度下温育 5 分钟后测定的。

在另一种优选实施方案中，本发明的多肽与一种氨肽酶协同作用水解另一种多肽。术语“与一种氨肽酶协同作用水解另一种多肽”在此定义为一种二肽氨肽酶与一种氨肽酶的结合，相对于任一单独的酶，这种结合使肽或多肽的水解提高至少 5 倍，更优选地至少 10 倍，甚至更优选地至少 25 倍，最优选地至少 50 倍，甚至最优选地至少 100 倍。该氨肽酶可以是任何氨肽酶，但优选地是从米曲霉中获得的氨肽酶，更优选地是如 WO 96/28542 所述从米曲霉中获得的氨肽酶 I。该多肽优选地是一种三肽，但可以是任何更大的肽或蛋白质。

在第四种实施方案中，本发明涉及与具有 SEQ ID NO: 2 的氨基酸序列的多肽有免疫化学同一性或部分免疫化学同一性的分离的多肽。通过众所周知的 Ouchterlony 双向免疫扩散法经免疫交叉反应同一性试验测定免疫化学性质。特别地，根据 Harboe 和 Ingild 在 N.H.Axelsen, J.Kroll 和 B.Weeks 所编的《定量免疫电泳手册》(A Manual of Quantitative Immunoelectrophoresis), Blackwell Scientific Publications, 1973, 第 23 章或者 Johnstone 和 Thorpe, 《实践免疫化学》(Immunochemistry in Practice), Blackwell Scientific Publications, 1982 (特别是第 27-31 页) 中所述的方法，通过免疫兔（或其它啮齿类动物）制备一种抗血清，其所含的抗体是免疫活性的或者可与具有 SEQ ID NO: 2 的氨基酸序列的多肽的表位相结合。具有免疫化学同一性的多肽是使用特定免疫化学技术以相同方式与该抗血清反应的多肽，如沉淀物的总融合、相同的沉淀物形态学和/或相

同的电泳迁移率。Axelsen, Bock 和 Kroll 在 N.H.Axelsen, J.Kroll 和 B.Weeks 所编的《定量免疫电泳手册》，Blackwell Scientific Publications, 1973, 第 10 章中描述了免疫化学同一性的进一步解释。具有部分免疫化学同一性的多肽是一种以部分相同的方式与抗血清反应的多肽，如沉淀物的部分融合、部分相同的沉淀物形态学和/或部分相同的电泳迁移率，使用特异的免疫化学技术。Bock 和 Axelsen 在 N.H.Axelsen, J.Kroll 和 B.Weeks 所编的《定量免疫电泳手册》，Blackwell Scientific Publications, 1973, 第 11 章中描述了部分免疫化学同一性的进一步解释。

可与一种寡核苷酸探针杂交的核苷酸序列所编码的多肽，该探针可与 SEQ ID NO: 1 的核苷酸序列、其互补链或 SEQ ID NO: 1 的等位变体和亚序列杂交；该多肽的等位变体和片段；或同源多肽和具有相同或部分相同的免疫学性质的多肽，可从任何属的微生物中获得。

在一种优选实施方案中，这些多肽可从细菌来源获得。例如，这些多肽可从革兰氏阳性菌获得，如芽孢杆菌属菌株，例如：嗜碱芽孢杆菌（*Bacillus alkalophilus*）、解淀粉芽孢杆菌（*Bacillus amyloliquefaciens*）、短芽孢杆菌（*Bacillus brevis*）、环状芽孢杆菌（*Bacillus circulans*）、凝结芽孢杆菌（*Bacillus coagulans*）、灿烂芽孢杆菌（*Bacillus lautus*）、迟缓芽孢杆菌（*Bacillus lentus*）、地衣芽孢杆菌（*Bacillus licheniformis*）、巨大芽孢杆菌（*Bacillus megaterium*）、嗜热脂肪芽孢杆菌（*Bacillus stearothermophilus*）、枯草芽孢杆菌（*Bacillus subtilis*）或苏云金芽孢杆菌（*Bacillus thuringiensis*）；或链霉菌属菌株，例如，浅青紫链霉菌（*Streptomyces lividans*）或鼠灰链霉菌（*Streptomyces murinus*）；或从革兰氏阴性菌获得，例如大肠杆菌或假单胞菌属的种（*Pseudomonas sp.*）。

该多肽可从真菌来源获得，更优选地从酵母株获得，如假丝酵母属（*Candida*）、克鲁维酵母属（*Kluyveromyces*）、毕赤酵母属（*Pichia*）、酵母属（*Saccharomyces*）、裂殖酵母属（*Schizosaccharomyces*）或 *Yarrowia* 菌株；或丝状真菌株，如枝顶孢属（*Acremonium*）、曲霉

属 (*Aspergillus*)、出芽短梗霉属 (*Aureobasidium*)、隐球酵母属 (*Cryptococcus*)、*Filibasidium*、镰孢属 (*Fusarium*)、腐质霉属 (*Humicola*)、*Magnaporthe*、毛霉属 (*Mucor*)、毁丝霉属 (*Myceliophthora*)、*Neocallimastix*、脉孢霉属 (*Neurospora*)、拟青霉属 (*Paecilomyces*)、青霉属 (*Penicillium*)、*Piromyces*、裂褶菌属 (*Schizophyllum*)、篮霉属 (*Talaromyces*)、嗜热子囊菌属 (*Thermoascus*)、梭孢壳属 (*Thielavia*)、*Tolypocladium* 或木霉属 (*Trichoderma*) 菌株。

在一种优选实施方案中，该多肽获自卡尔酵母 (*Saccharomyces carlsbergensis*)、酿酒酵母 (*Saccharomyces cerevisiae*)、糖化酵母 (*Saccharomyces diastaticus*)、*Saccharomyces douglasii*、克鲁弗酵母 (*Saccharomyces kluyveri*)、诺地酵母 (*Saccharomyces norbensis*) 或 *Saccharomyces oviformis* 菌株。

在另一种优选实施方案中，该多肽获自杆孢状镰孢 (*Fusarium bactridioides*)、*Fusarium cerealis*、*Fusarium crookwellense*、黄色镰孢 (*Fusarium culmorum*)、禾谷镰孢 (*Fusarium graminearum*)、禾赤镰孢 (*Fusarium graminum*)、异孢镰孢 (*Fusarium heterosporum*)、合欢木镰孢 (*Fusarium negundi*)、尖镰孢 (*Fusarium oxysporum*)、网状镰孢 (*Fusarium reticulatum*)、粉红镰孢 (*Fusarium roseum*)、接骨木镰孢 (*Fusarium sambucinum*)、肤色镰孢 (*Fusarium sarcochromum*)、腐皮镰孢 (*Fusarium solani*)、拟枝孢镰孢 (*Fusarium sporotrichioides*)、*Fusarium sulphureum*、*Fusarium torulosum*、*Fusarium trichothecioides*、*Fusarium venenatum*、*Humicola insolens*、柔毛腐质霉 (*Humicola lanuginosa*)、米黑毛霉 (*Mucor miehei*)、嗜热毁丝霉 (*Myceliophthora thermophila*)、粗糙脉孢霉 (*Neurospora crassa*)、产紫青霉 (*Penicillium purpurogenum*)、*Trichoderma harzianum*、康宁木霉 (*Trichoderma koningii*)、*Trichoderma longibrachiatum*、*Trichoderma reesei* 或绿色木霉 (*Trichoderma viride*) 菌株。

本发明的多肽优选地获自曲霉属的种，包括但不限于：棘孢曲霉（*Aspergillus aculeatus*）、泡盛曲霉（*Aspergillus awamori*）、臭曲霉（*Aspergillus foetidus*）、日本曲霉（*Aspergillus japonicus*）、构巢曲霉、黑曲霉（*Aspergillus niger*）或米曲霉。

在一种更优选的实施方案中，本发明的多肽从米曲霉株中获得，最优选地从米曲霉 ATCC 20386 或其突变株中获得，例如，具有 SEQ ID NO: 2 的氨基酸序列的多肽。

应当理解，对于上述种，本发明包括完全的和不完全的两种状态，以及其它分类学等同物，例如无性型，不管其种名如何。本领域的技术人员将方便地认识到合适的等同物的同一性。本发明的多肽也可获自如 Raper, K.D. 和 Fennel D.I., 1965, 《曲霉属》（*The Genus Aspergillus*）, The Wilkins Company, Baltimore 所定义的是曲霉属同义的微生物。曲霉是有丝分裂孢子真菌，其特征在于曲霉(*aspergillum*)由终止于泡囊的、无已知有性状态的分生孢子柄组成，其随后产生一层或两层不同地被称为小梗或瓶梗的同时形成的特化细胞，和被称为分生孢子的无性形成的孢子。曲霉属的已知有性型包括散囊菌属（*Eurotium*）、*Neosartorya* 和 *Emericella*。曲霉属菌株及其有性型在许多培养物保藏所如美国典型培养物保藏中心（ATCC）、德意志微生物保藏中心（DSM）、真菌菌种保藏中心（CBS）和农业研究机构保藏中心（NRRL）处易于为公众所获得。

此外，可以使用上述探针从其它来源包括从自然界（例如，土壤、堆肥、水等等）分离的微生物中鉴定并获得这些多肽。从自然环境分离微生物的技术在本领域中众所周知。然后类似地通过筛查另一种微生物的基因组文库或 cDNA 文库可得到核酸序列。一旦用探针检测到编码该多肽的核酸序列，可使用本领域普通技术人员所知的技术分离或克隆该序列（见，例如，Sambrook 等人，1989，同前）。

对于本发明来说，在此使用的与特定来源有关的术语“获自”意思是，多肽由该来源产生或由一种插入了该来源的基因的细胞产生。

此处所定义的“分离的”多肽是基本上不含其它非二肽氨肽酶多

肽的多肽，例如，经 SDS-PAGE 测定，纯度至少约 20%、优选地至少约 40%、更优选地约 60%、甚至更优选地约 80%、最优选地约 90%、甚至最优选地约 95%。

核酸序列

本发明也涉及编码本发明多肽的分离的核酸序列。在一种优选实施方案中，该核酸序列编码一种获自曲霉如米曲霉的多肽，在一种更优选的实施方案中，该核酸序列获自米曲霉 ATCC 20386，例如 SEQ ID NO: 1 的核酸序列。在另一种更优选的实施方案中，该核酸序列是含于大肠杆菌 NRRL B-21682 所含质粒 pMWR52 中的序列。本发明也包括编码具有 SEQ ID NO: 2 的氨基酸序列的多肽的核酸序列，它由于遗传密码的简并性而与 SEQ ID NO: 1 不同。本发明也涉及 SEQ ID NO: 1 的亚序列，其编码具有二肽氨肽酶活性的 SEQ ID NO: 2 的片段。除了自 5'端和/或 3'端缺失一个或多个核苷酸外，SEQ ID NO: 1 的亚序列是 SEQ ID NO: 1 所含的核酸序列。优选地，亚序列含有至少 990 个核苷酸、更优选地至少 1140 个核苷酸、最优选地至少 1290 个核苷酸。

该核酸序列可以获自如 Raper, K.D.和 Fennel D.I., 1965, 同前所定义的作为曲霉的分类学等同物的微生物，不管已知的其种名如何。

用于分离或克隆编码一种多肽的核酸序列的技术在本领域众所周知，包括从基因组 DNA 的分离、从 cDNA 的制备或其组合。例如，使用众所周知的聚合酶链反应 (PCR) 或为了检测具有共同结构特征的克隆 DNA 片段对表达文库的抗体筛查，能实现从这种基因组 DNA 中克隆本发明的核酸序列。参见，例如，Innis 等人，1990，《PCR: 方法与应用指南》(PCR: A Guide to Methods and Application)，Academic Press，纽约。也可应用其它核酸扩增方法如连接酶链反应 (LCR)、连接的活化转录 (LAT) 和基于核酸序列的扩增 (NASBA)。可从曲霉株或另一种或相关生物中克隆核酸序列，因而，例如，该核酸序列可以是该核酸序列多肽编码区的等位变体或种变体。

在此所用的术语“分离的核酸序列”是指基本上不含其它核酸序列的一种核酸序列，例如，经琼脂糖电泳测定，纯度至少约 20%、优选地至少约 40%、更优选地至少约 60%、甚至更优选地至少约 80%、最优选地至少约 90%。例如，可通过基因工程中使用的标准克隆方法获得分离的核酸序列，该方法将核酸序列从其天然位置再定位到将再产生该序列的不同位点。克隆方法可包括：包含编码该多肽的核酸序列的期望核酸片段的切割和分离、该片段向载体分子中的插入、以及该重组载体向宿主细胞中的掺入，在宿主细胞中该核酸序列的多拷贝或克隆将被复制。该核酸序列可以是基因组、cDNA、RNA、半合成、合成的来源或其任一组合。

本发明也涉及编码一种活性多肽的核酸序列，其与 SEQ ID NO: 1 的核酸序列有一定程度的同源性，同源性至少约为 50%、优选地约 60%、优选地约 70%、优选地约 80%、更优选地约 90%、甚至更优选地约 95%、最优选地约 97%。对于本发明而言，可以应用同一性表、罚区间为 10、罚区间长度为 10，通过 Clustal 方法（Higgins, 1989, 同前）测定两种核酸序列之间的同源性程度。

编码本发明多肽的核酸序列的修饰，可能是合成基本上类似于该多肽的多肽所必需的。术语“基本上类似于”该多肽是指该多肽的非天然存在的形式。这些多肽也许以某些基因工程方式与自其天然来源分离的多肽不同。例如，使用如定点诱变合成多肽的变体可能是有意义的，这些变体在比活性、热稳定性、最适 pH 等方面有所不同。根据作为 SEQ ID NO: 1 的多肽编码部分存在的核酸序列，例如其亚序列和/或通过核苷酸置换的引入，可构建类似的序列，该核苷酸置换的引入不产生核酸序列所编码的多肽的另一种氨基酸序列，但相应于准备用于产生酶的宿主生物的密码子选择偏好，或者该核苷酸置换的引入可产生一种不同的氨基酸序列。关于核苷酸置换的一般性描述参见，例如，Ford 等人，1991，《蛋白质表达与纯化》（Protein Expression and Purification）2: 95-107。

对本领域的技术人员而言显然地，这种置换能在分子功能的关键

区之外进行，而仍产生活性多肽。可根据本领域已知的方法如定点诱变或丙氨酸扫描诱变（见，例如，Cunningham 和 Wells, 1989, 科学 (Science) 244: 1080-1085）鉴定氨基酸残基，这些氨基酸残基对于本发明的分离核酸序列所编码的多肽的活性是必需的，因此优选地不予置换。在后一种方法中，在分子中每一带正电荷的残基处均引入突变，测定产生的突变分子的二肽氨肽酶活性，以鉴定对于分子活性关键的氨基酸残基。底物-酶相互作用的位点也能通过三维结构分析来确定，这些测定技术如核磁共振分析、结晶学或光亲和标记（见，例如，de Vos 等人, 1992, 科学 255: 306-312; Smith 等人, 1992, 分子生物学杂志 (Journal of Molecular Biology) 224:899-904; Wlodaver 等人, 1992, FEBS Letters 309:59-64）。

本发明的多肽也包括融合多肽或可切开的融合多肽，其中另一种多肽在该多肽或其片段的 N 端或 C 端与之融合。融合多肽的产生方法是将编码另一种多肽的核酸序列（或其一部分）与本发明的核酸序列（或其一部分）相融合。产生融合多肽的技术在本领域中已知，包括连接编码多肽的编码序列，使得它们符合读框，并使融合多肽的表达位于相同启动子和终止子的控制之下。

本发明也涉及编码本发明多肽的分离的核酸序列，其在低度严格条件下、更优选地中度严格条件下、最优选地高度严格条件下可与一种寡核苷酸探针杂交，该寡核苷酸探针在相同条件下可与 SEQ ID NO: 1 的核酸序列或其互补链；或其等位变体及亚序列杂交 (Sambrook 等人, 1989, 同前)。

核酸构建体

本发明也涉及包含与一种或多种控制序列有效连接的本发明的核酸序列的核酸构建体，该控制序列在其适宜的条件下引导编码序列在合适的宿主细胞中表达。应当理解，表达包括参与多肽产生的任何步骤，包括但不限于转录、转录后修饰、翻译、翻译后修饰和分泌。

在此将“核酸构建体”定义为单链或双链的核酸分子，其分离自

天然存在的基因，或经修饰含有以自然界中不会存在的方式组合及并列的核酸片段。当核酸构建体中含有表达本发明的编码序列所需的所有控制序列时，术语核酸构建体与术语表达盒同义。术语“编码序列”在此定义为被转录为 mRNA 并被翻译为本发明的多肽的序列。编码序列的边界一般由核糖体结合位点（原核生物）或 ATG 起始密码子（真核生物）和转录终止序列所限定，核糖体结合位点或 ATG 起始密码子恰好位于 mRNA 5'端开放阅读框的上游，转录终止序列恰好位于 mRNA 3'端开放阅读框的下游。编码序列可包括但不限于 DNA、cDNA 和重组核酸序列。

可以以多种方式操作编码本发明多肽的分离的核酸序列，以准备该多肽的表达。在插入到载体之前对编码一种多肽的核酸序列的操作可以是希望的或必需的，这取决于表达载体。应用克隆方法修饰核酸序列的技术在本领域中众所周知。

在此将术语“控制序列”定义为包括对于本发明的多肽表达所必需的或有利的的所有组分。每种控制序列对于编码多肽的核酸序列而言可以是自身的或外源的。这样的控制序列包括但不限于：前导序列、聚腺苷酸化序列、前肽序列、启动子、信号序列和转录终止子。至少，控制序列包括启动子和转录及翻译终止信号。控制序列可具有接头，这是为了引入特定的限制性位点，以便于控制序列与编码多肽的核酸序列编码区的连接。术语“有效地连接”在此定义为一种构型，其中将一种控制序列适当地置于相对于 DNA 序列编码序列的位置，使得该控制序列引导多肽的产生。

控制序列可以是一种合适的启动子序列，即为了核酸序列的表达而被宿主细胞识别的一种核酸序列。启动子序列中含有介导多肽表达的转录控制序列。启动子可以是在所选宿主细胞中显示转录活性的任何核酸序列，包括突变型、截短型或杂合型启动子，并可从编码与宿主细胞同源或异源的胞外或胞内多肽的基因中获得。

特别是在细菌宿主细胞中，引导本发明的核酸构建体转录的合适启动子的实例，是从下列基因中获得的启动子：大肠杆菌乳糖（lac）

操纵子、天蓝色链霉菌 (*Streptomyces coelicolor*) 琼脂糖酶基因 (*dagA*)、枯草芽孢杆菌果聚糖蔗糖酶基因 (*sacB*)、地衣芽孢杆菌 α -淀粉酶基因 (*amyL*)、嗜热脂肪芽孢杆菌的产麦芽淀粉酶基因 (*amyM*)、解淀粉芽孢杆菌 α -淀粉酶基因 (*amyQ*)、地衣芽孢杆菌青霉素酶基因 (*penP*)、枯草芽孢杆菌 *xyl A* 和 *xyl B* 基因, 和原核生物的 β -内酰胺酶基因 (Villa-Kamaroff 等人, 1978, 美国国家科学院院报 (Proceedings of the National Academy of Sciences USA) 75: 3727-3731), 以及 *tac* 启动子 (DeBoer 等人, 1983, 美国国家科学院院报 80: 21-25)。更多的启动子描述于《科学美国人》(Scientific American), 1980, 242: 74-94 中的“来源于重组细菌的有用蛋白质”; 和 Sambrook 等人, 1989, 同前。

在丝状真菌宿主细胞中引导本发明的核酸构建体转录的合适启动子的实例是从编码下列酶的基因中获得的启动子: 米曲霉 TAKA 淀粉酶、米黑根毛霉 (*Rhizomucor miehei*) 天冬氨酸蛋白酶、黑曲霉中性 α -淀粉酶、黑曲霉酸稳定性 α -淀粉酶、黑曲霉或泡盛曲霉葡糖淀粉酶 (*gla A*)、米黑根毛霉脂肪酶、米曲霉碱性蛋白酶、米曲霉丙糖磷酸异构酶、构巢曲霉乙酰胺酶、尖镰孢胰蛋白酶样蛋白酶 (美国专利号 4,288,627), 及其突变型、截短型和杂合型启动子。用于丝状真菌宿主细胞的特别优选的启动子是 TAKA 淀粉酶、NA2-*tpi* (来源于编码黑曲霉中性 α -淀粉酶和米曲霉丙糖磷酸异构酶的基因的杂合型启动子) 和 *glaA* 启动子。

在酵母宿主中, 有用的启动子可获自酿酒酵母烯醇化酶 (*ENO-1*) 基因、酿酒酵母半乳糖激酶基因 (*GAL1*)、酿酒酵母乙醇脱氢酶/甘油醛-3-磷酸脱氢酶基因 (*ADH2/GAP*) 和酿酒酵母 3-磷酸甘油酸激酶基因。Romanos 等人, 1992, 酵母 (Yeast) 8: 423-488 描述了用于酵母宿主细胞的其它有用的启动子。在哺乳动物宿主细胞中, 有用的启动子包括病毒启动子, 如来源于猿猴病毒 40 (SV40)、劳斯肉瘤病毒 (RSV)、腺病毒、牛乳头瘤病毒 (BPV) 和人巨细胞病毒 (CMV) 的启动子。

控制序列也可以是合适的转录终止序列，即被宿主细胞识别以终止转录的一种序列。终止序列被有效地连接到编码多肽的核酸序列的3'端。在所选宿主细胞中有功能的任何终止子均可用于本发明。

用于丝状真菌宿主细胞的优选的终止子从编码米曲霉 TAKA 淀粉酶、黑曲霉葡糖淀粉酶、构巢曲霉邻氨基苯甲酸合酶、黑曲霉 α -葡糖苷酶和尖镰孢胰蛋白酶样蛋白酶的基因中获得。

用于酵母宿主细胞的优选的终止子可从编码酿酒酵母烯醇化酶、酿酒酵母细胞色素 C (CYC1) 或酿酒酵母甘油醛-3-磷酸脱氢酶的基因中获得。Romanos 等人, 1992, 同前, 描述了用于酵母宿主细胞的其它有用的终止子。用于哺乳动物宿主细胞的终止序列在本领域中众所周知。

控制序列也可以是合适的前导序列，即对于宿主细胞翻译重要的 mRNA 的一种非翻译区。前导序列被有效地连接到编码多肽的核酸序列的5'端。在所选宿主细胞中有功能的任何前导序列均可用于本发明。

用于丝状真菌宿主细胞的优选的前导序列可从编码米曲霉 TAKA 淀粉酶和构巢曲霉丙糖磷酸异构酶的基因中获得。

用于酵母宿主细胞的合适的前导序列从酿酒酵母烯醇化酶 (ENO-1) 基因、酿酒酵母 3-磷酸甘油酸激酶基因、酿酒酵母 α -因子和酿酒酵母乙醇脱氢酶/甘油醛-3-磷酸脱氢酶基因 (ADH2/GAP) 中获得。

控制序列也可以是聚腺苷酸化序列，即被有效地连接到核酸序列3'端的一种序列，在转录时它可被宿主细胞作为信号识别，以将聚腺苷残基加至转录的 mRNA。在所选宿主细胞中有功能的任何聚腺苷酸化序列均可用于本发明。

用于丝状真菌宿主细胞的优选的聚腺苷酸化序列可从编码米曲霉 TAKA 淀粉酶、黑曲霉葡糖淀粉酶、构巢曲霉邻氨基苯甲酸合酶和黑曲霉 α -葡糖苷酶的基因中获得。

Guo 和 Sherman, 1995, 分子细胞生物学 (Molecular Cellular Biology) 15:5983-5990 描述了用于酵母宿主细胞的有用的聚腺苷酸化序列。用于哺乳动物宿主细胞的聚腺苷酸化序列在本领域中众所周

知。

控制序列也可以是信号肽编码区，其编码与多肽的氨基端相连接的氨基酸序列，该氨基酸序列能引导编码的多肽进入细胞的分泌途径。核酸序列编码序列的 5'端可固有地含有一种信号肽编码区，该编码区与编码分泌多肽的编码区片段在翻译读框内天然连接。此外，编码序列的 5'端可含有对于编码序列外源的信号肽编码区。当编码序列不是正常地含有信号肽编码区时，可能需要外源的信号肽编码区。此外，外源信号肽编码区可简单地替换天然信号肽编码区以获得增强的多肽分泌。信号肽编码区可从曲霉属的种的葡糖淀粉酶或淀粉酶基因、根霉菌属的种的脂酶或蛋白酶基因、酿酒酵母 α -因子的基因、芽孢杆菌属的种的淀粉酶或蛋白酶基因或小牛前凝乳酶原基因中获得。然而，能引导表达的多肽进入所选宿主细胞分泌途径的任何信号肽编码区均可用于本发明。

用于细菌宿主细胞的有效信号肽编码区是从芽孢杆菌 NCIB 11837 的产麦芽淀粉酶基因、嗜热脂肪芽孢杆菌 α -淀粉酶基因、地衣芽孢杆菌枯草蛋白酶基因、地衣芽孢杆菌 β -内酰胺酶基因、嗜热脂肪芽孢杆菌中性蛋白酶基因(nprT、nprS、nprM)或枯草芽孢杆菌 PrsA 基因中获得的信号肽编码区。Simonen 和 Palva, 1993, 微生物综述 (Microbiological Reviews) 57: 109-137 描述了更多的信号肽。

用于丝状真菌宿主细胞的有效信号肽编码区是从米曲霉 TAKA 淀粉酶基因、黑曲霉中性淀粉酶基因、米黑根毛霉天冬氨酸蛋白酶基因、柔毛腐质霉纤维素酶基因或柔毛腐质霉脂酶基因中获得的信号肽编码区。

用于酵母宿主细胞的有用的信号肽从酿酒酵母 α -因子和酿酒酵母转化酶的基因中获得。Romanos 等人, 1992, 同前, 描述了其它有用的信号肽编码区。

控制序列也可以是前肽编码区，其编码位于多肽氨基端的一种氨基酸序列。所产生的多肽被称为酶原 (proenzyme) 或多肽原 (或者有时称为酶原 (zymogen))。多肽原通常无活性，通过从多肽原上催

化或自催化地切割前肽能转化为成熟的活性多肽。前肽编码区可从枯草芽孢杆菌碱性蛋白酶基因 (aprE)、枯草芽孢杆菌中性蛋白酶基因 (nprT)、酿酒酵母 α -因子基因、米黑根毛霉天冬氨酸蛋白酶基因或嗜热毁丝霉漆酶基因 (WO 95/33836) 中获得。

当信号肽和前肽区两者都存在于多肽的氨基端时, 前肽区紧接多肽的氨基端, 信号肽区紧接前肽区的氨基端。

本发明的核酸构建体也可包含一种或多种核酸序列, 其编码有利于引导多肽表达的一种或多种因子, 例如, 转录激活物 (例如反式作用因子)、陪伴分子和加工蛋白酶。在所选宿主细胞中有功能的任何因子均可用于本发明。编码一种或多种这类因子的核酸不一定与编码多肽的核酸序列串联。

转录激活物是激活编码多肽的核酸序列转录的一种蛋白质 (Kudla 等人, 1990, EMBO 杂志 (EMBO Journal) 9:1355-1364; Jarai 和 Buxton, 1994, 当代遗传学 (Current Genetics) 26:2238-244; Verdier, 1990, 酵母 6: 271-297)。编码激活物的核酸序列可从编码嗜热脂肪芽孢杆菌 NprA (npr A)、酿酒酵母血红素激活蛋白 1 (hap1)、酿酒酵母半乳糖代谢蛋白 4 (gal4)、构巢曲霉氨调节蛋白 (areA) 和米曲霉 α -淀粉酶激活物 (amyR) 的基因中获得。至于更多的实例, 见 Verdier, 1990, 同前, 和 MacKenzie 等人, 1993, 普通微生物学杂志 (Journal of General Microbiology) 139: 2295-2307。

陪伴分子是一种辅助另一种多肽正确折叠的蛋白质 (Hartl 等人, 1994, TIBS 19:20-25; Bergeron 等人, 1994, TIBS 19:124-128; Demolder 等人, 1994, 生物技术杂志 (Journal of Biotechnology) 32:179-189; Craig, 1993, 科学 260: 1902-1903; Gething 和 Sambrook, 1992, 自然 355: 33-45; Puig 和 Gilbert, 1994, 生物化学杂志 (Journal of Biological Chemistry) 269: 7764-7771; Wang 和 Tsou, 1993, The FASEB Journal 7: 1515-11157; Robinson 等人, 1994, 生物技术 (Bio/Technology) 1: 381-384; Jacobs 等人, 1993, 分子微生物学 (Molecular Microbiology) 8:957-966)。编码陪伴分子的核酸序列

可从编码枯草芽孢杆菌 GroE 蛋白、枯草芽孢杆菌 PrsA、米曲霉蛋白质二硫键异构酶、酿酒酵母钙联接蛋白、酿酒酵母 BiP/GRP78 和酿酒酵母 Hsp70 的基因中获得。至于更多的实例，见 Gething 和 Sambrook, 1992, 同前, 和 Hartl 等人, 1994, 同前。

加工蛋白酶是切割前肽以产生有生物化学活性的成熟多肽的一种蛋白酶 (Enderlin 和 Ogrydziak, 1994, 酵母 10: 67-79; Fuller 等人, 1989, 美国国家科学院院报 86: 1434-1438; Julius 等人, 1984, 细胞 (Cell) 37: 1075-1089; Julius 等人, 1983, 细胞 32: 839-852; 美国专利号 5,702,934)。编码加工蛋白酶的核酸序列可从编码酿酒酵母二肽氨肽酶、酿酒酵母 Kex2、*Yarrowia lipolytica* 二价加工内切蛋白酶 (xpr6) 和尖镰孢金属蛋白酶 (p45 基因) 的基因中获得。

也可希望加入能够调节与宿主细胞生长相关的多肽表达的调节序列。调节系统的实例是那些响应包括存在调节化合物在内的化学或物理刺激而引起基因表达被打开或关闭的调节系统。原核系统中的调节系统包括 lac、tac 和 trp 操纵子系统。在酵母中，可使用 ADH2 系统或 GAL1 系统。在丝状真菌中，TAKA α -淀粉酶启动子、黑曲霉葡萄糖淀粉酶启动子和米曲霉葡萄糖淀粉酶启动子可用作调节序列。调节序列的其它实例是允许基因扩增的调节序列。在真核系统中，包括在氨甲蝶呤存在下扩增的二氢叶酸还原酶基因和有重金属时扩增的金属硫蛋白基因。在这些情况中，编码多肽的核酸序列常与调节序列有效地连接。

本发明也涉及用于改变编码本发明的多肽的内源基因表达的核酸构建体。构建体可含有改变内源基因表达所需的最小量的组分。在一种实施方案中，核酸构建体优选地包含 (a) 一种导向序列、(b) 一种调节序列、(c) 一种外显子和 (d) 一种剪接供体位点。一旦核酸构建体引入细胞中后，该构建体通过同源重组在内源基因位点处插入细胞基因组中。导向序列引导元件 (a) - (d) 向内源基因中的整合，以使元件 (b) - (d) 与内源基因有效地连接。在另一种实施方案中，核酸构建体包含 (a) 一种导向序列、(b) 一种调节序列、(c) 一种外显子、

(d) 一种剪接供体位点、(e) 一种内含子和 (f) 一种剪接受体位点，其中导向序列引导元件 (a) - (f) 的整合，以使元件 (b) - (f) 与内源基因有效地连接。然而，构建体可含有其它组分，如选择性标记。

在这两种实施方案中，这些组分的引入导致一种新转录单位的产生，其改变了内源基因的表达。实质上，这种新转录单位是由导向构建体引入的序列与内源基因的融合产物。在一种内源基因已改变的实施方案中，该基因被激活。在该实施方案中，用同源重组来替换、破坏或使与亲代细胞的内源基因正常连接的调节区失能，方法是插入使基因以比相应亲代细胞更高的水平表达的调节序列。使用本领域众所周知的方法（见，例如，美国专利号 5,641,670），通过在构建体中包含可扩增的选择性标记基因，能进一步扩增活化的基因。在内源基因已改变的另一种实施方案中，基因的表达降低。

导向序列可在内源基因之内、与基因直接邻接、在上游基因之内或在内源基因的上游及距其一定距离处。能使用一种或多种导向序列。例如，环状质粒或 DNA 片段优选地使用单个导向序列，而线性质粒或 DNA 片段优选地使用两种导向序列。

构建体的调节序列可包含一种或多种启动子、增强子、支架附着区或基质附着部位、负调节元件、转录结合位点或这些序列的组合。

构建体此外含有内源基因的一种或多种外显子。外显子被定义为一种 DNA 序列，其拷贝为 RNA 并存在于成熟 mRNA 分子中，使得外显子序列与内源基因的编码区符合读框。外显子可任选地含有编码一种或多种氨基酸和/或部分地编码一种氨基酸的 DNA。另外，外显子含有对应于 5' 非编码区的 DNA。外源外显子编码一种或多种氨基酸和/或氨基酸的一部分时，设计核酸构建体，使得在转录和剪接后阅读框架与内源基因的编码区符合读框，以使来源于第二种外显子的 mRNA 部分的适当读框未改变。

构建体的剪接供体位点引导一种外显子向另一种外显子的剪接。一般地，第一种外显子位于第二种外显子的 5'，在 3' 侧重叠并位于第一种外显子侧面的剪接供体位点识别在第二种外显子 5' 侧位于第二种

外显子侧面的剪接受体位点。象剪接供体位点一样，剪接受体位点是引导一种外显子向另一种外显子剪接的序列。与剪接供体位点一起作用，剪接机制使用剪接受体位点来实现内含子的去除。

表达载体

本发明也涉及含有本发明的核酸序列、启动子和转录及翻译终止信号的重组表达载体。上述多种核酸和控制序列可以连接在一起产生一种重组表达载体，该重组表达载体中可包括一种或多种适宜的限制性位点，以便于编码多肽的核酸序列在这些位点的插入或置换。此外，可通过将核酸序列或含有该序列的核酸构建体插入到合适的表达载体中来表达本发明的核酸序列。在构建表达载体中，应使编码序列位于载体中，以使编码序列与用于表达（及可能的分泌）的合适的控制序列有效地连接。

重组表达载体可以是能方便地进行重组 DNA 方法并能导致核酸序列表达的任何载体（例如质粒或病毒）。载体的选择一般取决于载体与待引入载体的宿主细胞之间的相容性。载体可以是线性或闭合环状的质粒。载体可以是自主复制性载体，即，作为一种染色体外实体存在的载体，其复制不依赖于染色体的复制，例如，质粒、染色体外元件、微型染色体或人工染色体。载体可含有保证自复制的任何工具。另外，载体可以是这样一种载体，当被引入宿主细胞时，它整合到基因组中并与所整合的染色体一起复制。载体系统可以是一种载体或质粒或是两种或多种载体或质粒，它们一起含有被引入到宿主细胞基因组中的总 DNA，或者可以是一种转座子。

本发明的载体优选地含有允许容易地筛选转化细胞的一种或多种选择性标记。选择性标记是这样一种基因，其产物提供杀生物剂或病毒抗性、重金属抗性、原养型到营养缺陷型的转变等等。细菌选择性标记的实例是枯草芽孢杆菌或地衣芽孢杆菌的 *dal* 基因，或是赋予抗生素抗性如氨基青霉素、卡那霉素、氯霉素或四环素抗性的标记。用于哺乳动物细胞的合适的标记是二氢叶酸还原酶（*dhfr*）、潮霉素磷酸

转移酶 (hygB)、氨基糖苷磷酸转移酶 II、和腐草霉素抗性基因。用于酵母宿主细胞的合适的标记是 ADE2、HIS3、LEU2、LYS2、MET3、TRP1 和 URA3。用于丝状真菌宿主细胞的选择性标记可以选自但不限于: amdS (乙酰胺酶)、argB (鸟氨酸氨甲酰基转移酶)、bar (链霉素乙酰转移酶)、hygB (潮霉素磷酸转移酶)、niaD (硝酸还原酶)、pyrG (乳清酸核苷-5'-磷酸脱羧酶)、sC (硫酸腺苷酰转移酶)、trpC (邻氨基苯甲酸合酶), 以及来源于其它种的相当物。优选的用于曲霉属细胞的是构巢曲霉或米曲霉的 amdS 和 pyrG 基因以及吸水链霉菌 (*Streptomyces hygroscopicus*) 的 bar 基因。另外, 当选择性标记位于分离的载体上时, 可通过共转化完成筛选, 例如, 如 WO 91/17243 所述。

本发明的载体优选地含有这样一种元件, 其允许载体向宿主细胞基因组中的稳定整合或载体在细胞中不依赖于细胞基因组的自主复制。

对于向宿主细胞基因组中的整合而言, 载体可依赖编码多肽的核酸序列, 或用于通过同源或非同源重组使载体稳定地整合到基因组中的载体的任何其它元件。此外, 载体可以含有用于引导通过同源重组向宿主细胞基因组中整合的另外的核酸序列。这些另外的核酸序列使载体能在染色体的精确位置整合到宿主细胞基因组中。为了增加在精确位置整合的可能性, 整合元件优选地应含有足够数量的与相应靶序列高度同源的核酸, 如 100-1,500 个碱基对, 优选地 400-1,500 个碱基对, 最优选地 800-1,500 个碱基对, 以增加同源重组的可能性。整合元件可以是与宿主细胞基因组中的靶序列同源的任何序列。此外, 整合元件可以是非编码的或编码的核酸序列。另一方面, 也可通过非同源重组将载体整合到宿主细胞的基因组中。

对于自主复制而言, 载体可进一步包含使载体能在所述宿主细胞中自主复制的复制起点。细菌复制起点的实例是允许在大肠杆菌中复制的质粒 pBR322、pUC19、pACYC177 和 pACYC184 的复制起点, 以及允许在芽孢杆菌中复制的质粒 pUB110、pE194、pTA1060 和

pAMβ1 的复制起点。用于酵母宿主细胞的复制起点的实例是 2 微米复制起点、ARS1、ARS4、ARS1 与 CEN3 的组合、以及 ARS4 与 CEN6 的组合。复制起点可以具有使其在宿主细胞内的作用成为温度敏感型的突变（见，例如，Ehrlich, 1978, 美国国家科学院院报 75: 1433）。

可以将多于一个拷贝的编码本发明的多肽的核酸序列插入宿主细胞中以增强该核酸序列的表达。获得核酸序列稳定扩增的方法是，将至少核酸序列的另一个拷贝整合到宿主细胞基因组中，或者将核酸序列包含有可扩增的选择性标记基因，通过在合适的选择剂的存在下培养细胞，能筛选出含有扩增拷贝的选择性标记基因、并因此筛选含有另外拷贝的核酸序列的细胞。

用于连接上述元件以构建本发明的重组表达载体的方法为本领域的技术人员所熟知（见，例如，Sambrook 等人，1989，同前）。

宿主细胞

本发明也涉及包含本发明的核酸序列的重组宿主细胞，其方便地用于多肽的重组产生。术语“宿主细胞”包括由于在复制期间发生突变而与亲代细胞不同的亲代细胞的任何子代。

将含有本发明的核酸序列的载体引入宿主细胞，以使载体作为染色体整合体或作为自复制的染色体外载体而维持在胞内。一般认为整合是有利的，因为核酸序列更可能在细胞内稳定地保持。如上所述可通过同源或非同源重组进行载体向宿主染色体中的整合。

宿主细胞的选择很大程度上取决于编码多肽的基因及其来源。宿主细胞可以是单细胞微生物如原核生物或非单细胞微生物如真核生物。有用的单细胞是细菌细胞，如革兰氏阳性菌，包括但不限于：芽孢杆菌细胞，例如，嗜碱芽孢杆菌、解淀粉芽孢杆菌、短芽孢杆菌、环状芽孢杆菌、*Bacillus clausii*、凝结芽孢杆菌、灿烂芽孢杆菌、迟缓芽孢杆菌、地衣芽孢杆菌、巨大芽孢杆菌、嗜热脂肪芽孢杆菌、枯草芽孢杆菌和苏云金芽孢杆菌；或链霉菌细胞，例如，浅青紫链霉菌或鼠灰链霉菌，或者是革兰氏阴性菌如大肠杆菌和假单胞菌属的种。在

一种优选实施方案中，细菌宿主细胞是迟缓芽孢杆菌、地衣芽孢杆菌、嗜热脂肪芽孢杆菌或枯草芽孢杆菌细胞。载体向细菌宿主细胞中的引入可通过例如原生质体转化（见，例如，Chang 和 Cohen, 1979, 分子普通遗传学 (Molecular General Genetics) 168: 111-115)、使用感受态细胞（见，例如，Young 和 Spizizin, 1961, 细菌学杂志 (Journal of Bacteriology) 81: 823-829, 或 Dubnau 和 Davidoff-Abelson, 1971, 分子生物学杂志 56: 209-211)、电穿孔法（见，例如，Shigekawa 和 Dower, 1988, 生物技术 (Biotechniques) 6: 742-751) 或接合法（见，例如，Koehler 和 Thorne, 1987, 细菌学杂志 169: 5771-5278）来实现。

宿主细胞可以是真核生物，如哺乳动物细胞、昆虫细胞、植物细胞或真菌细胞。有用的哺乳动物细胞包括中国仓鼠卵巢 (CHO) 细胞、HeLa 细胞、仔仓鼠肾 (BHK) 细胞、COS 细胞或其它多种可得的无限增殖化细胞系，例如，可从美国典型培养物保藏中心得到的细胞系。

在一种优选实施方案中，宿主细胞是真菌细胞。在此所用的“真菌”包括子囊菌门 (Ascomycota)、担子菌门 (Basidiomycota)、壶菌门 (Chytridiomycota) 和接合菌门 (Zygomycota) (如 Hawksworth 等人, 于《Ainsworth 和 Bisby's 真菌词典》(Ainsworth and Bisby's Dictionary of The Fungi), 第 8 版, 1995, CAB International, University Press, 剑桥, 英国所定义的), 以及卵菌门 (Oomycota) (如 Hawksworth 等人, 1995, 同前, 第 171 页所述) 和所有有丝分裂孢子真菌 (Hawksworth 等人, 1995, 同前)。子囊菌门的代表性组群包括, 例如, 脉孢霉属 (Neurospora)、正青霉属 (Eupenicillium) (=青霉属)、Emericella (=曲霉属)、散囊菌属 (Eurotium) (=曲霉属) 和下面列出的真酵母。担子菌门的实例包括蘑菇、锈菌和黑粉菌。壶菌门的代表性组群包括, 例如, 异水霉属 (Allomyces)、小芽枝霉属 (Blastocladiella)、雕蚀菌属 (Coelomomyces) 和水生真菌。卵菌门的代表性组群包括, 例如, 水霉属 (Saprolegniomycetous) 水生真菌 (水霉) 如绵霉属 (Achlya)。有丝分裂孢子真菌的实例包括曲霉属、

青霉属、假丝酵母 (*Candida*) 和链格孢属 (*Alternaria*)。接合菌门的代表性组群包括, 例如, 根霉属 (*Rhizopus*) 和毛霉属 (*Mucor*)。

在一种更优选的实施方案中, 真菌宿主细胞是酵母细胞。在此所用的“酵母”包括产子囊酵母 (*ascosporogenous yeast*) (内孢霉目 (*Endomycetales*))、产担孢子囊酵母 (*basidiosporogenous yeast*) 和属于半知菌类 (*Fungi Imperfecti*) (芽酵母 (*Blastomycetes*)) 的酵母。产子囊酵母分为蚀精霉科 (*Spermophthoraceae*) 和酵母科 (*Saccharomycetaceae*)。后者由四个亚科组成: 裂殖酵母亚科 (*Schizosaccharomycoideae*) (例如, 裂殖酵母属)、拿逊酵母亚科 (*Nadsonioideae*)、油脂酵母亚科 (*Lipomycoideae*) 和酵母亚科 (*Saccharomycoideae*) (例如, 克鲁维酵母属、毕赤酵母属和酵母属)。产担孢子囊酵母包括白冬孢酵母属 (*Leucosporidium*)、红冬孢酵母属 (*Rhodosporidium*)、锁掷酵母属 (*Sporidiobolus*)、*Filobasidium* 和 *Filobasidiella*。属于半知菌类的酵母分为两个科: 掷孢酵母科 (*Sporobolomycetaceae*) (例如, 掷孢酵母属 (*Sorobolomyces*) 和布勒弹孢酵母属 (*Bullera*)) 和隐球酵母科 (*Cryptococcaceae*) (例如, 假丝酵母属)。由于酵母的分类在将来可能改变, 对于本发明而言, 应当如《酵母的生物学及活性》(*Biology and Activities of Yeast*) (Skinner, F.A., Passmore, S.M.和 Davenport, R.R.编, Soc. App. Bacteriol. Symposium Series No. 9, 1980) 所述定义酵母。酵母的生物学和酵母遗传学的操作在本领域中众所周知 (见, 例如, 《酵母的生物化学和遗传学》(*Biochemistry and Genetics of Yeast*), Bacil, M., Horecker, B.J.和 Stopani, A.O.M.编, 第 2 版, 1987; 《酵母》(*The Yeasts*), Rose, A.H.和 Harrison, J. S.编, 第 2 版, 1987; 以及《酵母属酵母的分子生物学》(*The Molecular Biology of the Yeast Saccharomyces*), Strathern 等人编, 1981)。

在一种甚至更优选的实施方案中, 酵母宿主细胞是假丝酵母属、汉逊酵母属 (*Hansenula*)、克鲁维酵母属、毕赤酵母属、酵母属、裂殖酵母属或 *Yarrowia* 的种的细胞。

在一种最优选的实施方案中，酵母宿主细胞是卡尔酵母、酿酒酵母、糖化酵母、*Saccharomyces douglasii*、克鲁弗酵母、诺地酵母或 *Saccharomyces oviformis* 细胞。在另一种最优选的实施方案中，酵母宿主细胞是乳酸克鲁维酵母 (*Kluyveromyces lactis*) 细胞。在另一种最优选的实施方案中，酵母宿主细胞是 *Yarrowia Lipolytica* 细胞。

在另一种更优选的实施方案中，真菌宿主细胞是丝状真菌细胞。“丝状真菌”包括真菌亚门 (*Eumycota*) 和卵菌亚门的所有丝状形式 (如 Hawksworth 等人, 1995, 同前, 所述)。丝状真菌的特征在于由几丁质、纤维素、葡聚糖、脱乙酰壳多糖、甘露聚糖和其它复杂多糖组成的菌丝壁。营养生长是通过菌丝延伸，碳分解代谢是专性需氧的。相反，酵母如酿酒酵母的营养生长是通过单细胞菌体的芽殖，碳分解代谢可以是发酵的。在一种更优选的实施方案中，丝状真菌宿主细胞是但不限于：枝顶孢属、曲霉属、镰孢属、腐质霉属、毛霉属、毁丝霉属、脉孢霉属、青霉属、梭孢壳属、*Tolypocladium* 和木霉属的种的细胞。

在一种甚至更优选的实施方案中，丝状真菌宿主细胞是曲霉属细胞。在另一种甚至更优选的实施方案中，丝状真菌宿主细胞是枝顶孢属细胞。在另一种甚至更优选的实施方案中，丝状真菌宿主细胞是镰孢属细胞。在另一种甚至更优选的实施方案中，丝状真菌宿主细胞是腐质霉属细胞。在另一种甚至更优选的实施方案中，丝状真菌宿主细胞是毛霉属细胞。在另一种甚至更优选的实施方案中，丝状真菌宿主细胞是毁丝霉属细胞。在另一种甚至更优选的实施方案中，丝状真菌宿主细胞是脉孢霉属细胞。在另一种甚至更优选的实施方案中，丝状真菌宿主细胞是青霉属细胞。在另一种甚至更优选的实施方案中，丝状真菌宿主细胞是梭孢壳属细胞。在另一种甚至更优选的实施方案中，丝状真菌宿主细胞是 *Tolypocladium* 细胞。在另一种甚至更优选的实施方案中，丝状真菌宿主细胞是木霉属细胞。

在一种最优选的实施方案中，丝状真菌宿主细胞是泡盛曲霉、臭曲霉、日本曲霉、构巢曲霉、黑曲霉或米曲霉细胞。在另一种最优选

的实施方案中，丝状真菌宿主细胞是杆孢状镰孢、*Fusarium cerealis*、*Fusarium crookwellense*、黄色镰孢、禾谷镰孢、禾赤镰孢、异孢镰孢、合欢木镰孢、尖镰孢、网状镰孢、粉红镰孢、接骨木镰孢、肤色镰孢、腐皮镰孢、拟枝孢镰孢、*Fusarium sulphureum*、*Fusarium torulosum*、*Fusarium trichothecioides*、*Fusarium venenatum* 细胞。在一种甚至最优选的实施方案中，丝状真菌亲代细胞是 *Fusarium venenatum* (*Nirenberg sp. nov.*) 细胞。在另一种最优选的实施方案中，丝状真菌宿主细胞是 *Humicola insolens* 或柔毛腐质霉细胞。在另一种最优选的实施方案中，丝状真菌宿主细胞是米黑毛霉细胞。在另一种最优选的实施方案中，丝状真菌宿主细胞是嗜热毁丝霉细胞。在另一种最优选的实施方案中，丝状真菌宿主是粗糙脉孢霉细胞。在另一种最优选的实施方案中，丝状真菌宿主细胞是产紫青霉细胞。在另一种最优选的实施方案中，丝状真菌宿主细胞是 *Thielavia terrestris* 细胞。在另一种最优选的实施方案中，木霉属细胞是 *Trichoderma harzianum*、康宁木霉、*Trichoderma longibrachiatum*、*Trichoderma reesei* 或绿色木霉细胞。

真菌细胞的转化方法包括原生质体形成、原生质体转化和以本来已知的方式进行的细胞壁再生。用于曲霉属宿主细胞转化的合适方法描述于 EP 238 023 和 Yelton 等人，1984，美国国家科学院院报 81：1470-1474。Malardier 等人，1989，基因 78：147-156 和 WO 96/00787 描述了转化镰孢属的种的合适方法。可用 Becker 和 Guarente 在 Abelson, J. N. 和 Simon, M.I. 编的《酵母遗传学和分子生物学指南》(Guide to Yeast Genetics and Molecular Biology)，酶学方法 (Methods in Enzymology)，194 卷，182-187 页，Academic Press，纽约；Ito 等人，1983，细菌学杂志 (Journal of Bacteriology) 153：163；和 Hinnen 等人，1978，美国国家科学院院报 75：1920 中所述的方法转化酵母。哺乳动物细胞可用 Graham 和 Van der Eb (1978，病毒学 (Virology) 52：546) 的磷酸钙沉淀法通过直接摄取而转化。

产生方法

本发明也涉及产生本发明的多肽的方法，包括（a）培养其以野生型形式能产生多肽的菌株，产生含有该多肽的上清液；和（b）回收该多肽。优选地该菌株属于曲霉属。

本发明也涉及产生本发明的多肽的方法，包括（a）在有利于多肽产生的条件下培养宿主细胞；和（b）回收该多肽。

本发明进一步涉及产生本发明的多肽的方法，包括（a）在有利于多肽产生的条件下培养引入了一种新转录单位的同源重组细胞，该转录单位包含调节序列、外显子、和/或与编码该多肽的内源核酸序列第二种外显子有效连接的剪接供体位点；以及（b）回收该多肽。该方法基于基因激活技术的应用，例如，如美国专利号 5,641,670 所述。基因激活技术基于激活在细胞中通常不表达的基因，或提高在细胞中以很低水平表达的基因的表达。基因激活技术包括下列方法：将含有调节序列、外显子和/或剪接供体位点的外源 DNA 构建体插入细胞的基因组 DNA 中，这样，插入导致一种新转录单位的产生，该转录单位中调节序列、外显子和/或剪接供体位点与内源基因有效地连接，并激活内源基因的表达。

在本发明的产生方法中，用本领域中已知的方法在适于多肽产生的营养培养基中培养细胞。例如，可在实验室或工业发酵罐中通过摇瓶培养、小规模或大规模发酵（包括连续发酵、分批发酵、分批补料发酵或固态发酵）培养细胞，这在合适的培养基中和使多肽能被表达和/或分离的条件下进行。用本领域中已知的方法（参见，例如，细菌和酵母的参考文献；Bennett, J.W.和 LaSure, L., 编，《真菌中的多基因操作》（*More Gene Manipulations in Fungi*），Academic Press, CA, 1991）在含有碳源和氮源和无机盐的合适营养培养基中进行培养。合适的培养基可从商业供应商得到或根据公开的组成（例如，于美国典型培养物保藏中心的目录中）制备。如果多肽分泌到营养培养基中，则可直接从培养基中回收该多肽。如果该多肽不被分泌，则可从细胞裂解产物中回收。

可使用本领域中已知的、对多肽特异的方法检测多肽。这些检测方法可包括特异性抗体的使用、酶产物的形成或酶底物的消失。例如，可使用酶测定法检测多肽的活性。测定二肽氨肽酶活性的方法在本领域中已知，包括例如，测定 405nm 处对硝基苯胺水解的初始速率。

产生的多肽可通过本领域中已知的方法回收。例如，可用常规方法从营养培养基中回收多肽，这些方法包括但不限于：离心、过滤、萃取、喷雾干燥、蒸发或沉淀。

可用本领域中已知的多种方法纯化本发明的多肽，这些方法包括但不限于：层析（例如，离子交换、亲和、疏水、层析聚焦和大小排阻）、电泳方法（例如，制备型等电聚焦）、差示溶解度法（例如，硫酸铵沉淀）、SDS-PAGE 或萃取（见，例如，《蛋白质纯化》（Protein Purification） J.-C. Janson 和 Lars Ryden 编，VCH Publishers, 纽约，1989）。

二肽氨肽酶活性的去除或降低

本发明也涉及产生亲代细胞的突变细胞的方法，其包括破坏或删除编码多肽的核酸序列或其控制序列，这导致突变细胞比亲代细胞产生较少的多肽。

通过具有二肽氨肽酶活性的多肽在细胞中表达所需核酸序列的修饰和失活，可方便地实现二肽氨肽酶活性降低的菌株的构建。待修饰或失活的核酸序列可以是，例如，编码显示二肽氨肽酶活性所必需多肽的核酸序列或其部分，或者核酸序列可具有从核酸序列的编码序列表达多肽所需的调节功能。这种调节或控制序列的一个实例可以是启动子序列或其功能性部分，即，足以影响多肽表达的部分。用于可能的修饰的其它控制序列包括但不限于：前导序列、聚腺苷酸化序列、前肽序列、信号序列和终止终止子。

进行核酸序列修饰或失活的方法包括对细胞诱变和筛选二肽氨肽酶产生能力降低或消除的细胞。可以进行特定或随机诱变包括，例如，应用合适的物理或化学诱变剂、应用合适的寡核苷酸或对 DNA 序列施

行 PCR 产生的诱变。此外，可使用这些诱变剂的任何组合进行诱变。

适于本目的的物理或化学诱变剂的实例包括：紫外线（UV）照射、羟胺、N-甲基-N'-硝基-N-亚硝基胍（MNNG）、O-甲基羟胺、亚硝酸、甲基磺酸乙酯（EMS）、亚硫酸氢钠、甲酸和核苷酸类似物。

使用这些试剂时，施行诱变的方法一般是：在合适条件下和所选诱变剂存在下温育待诱变的细胞，以及筛选显示降低的或无二肽氨肽酶活性表达的细胞。

通过在编码多肽的核酸序列中或其转录或翻译所需的调节元件中一个或多个核苷酸的引入、置换或删除，可实现本发明的多肽产生的修饰或失活。例如，可插入或删除核苷酸，以引起终止密码子的引入、起始密码子的去除或开放阅读框的改变。可按照本领域中已知的方法通过定点诱变或 PCR 产生的诱变实现这种修饰或失活。尽管，原则上可在体内进行修饰，即，直接在表达待修饰的核酸序列的细胞上进行，但如下例证的在体外进行修饰是优选的。

灭活或降低所选宿主细胞产生该酶合适方法的实例基于基因置换或基因遮断(interruption)技术。例如，在基因遮断法中，在体外诱变对应于目标内源基因或基因片段的核酸序列，产生缺损核酸序列，然后将其转化入宿主细胞产生缺损基因。通过同源重组，缺损核酸序列置换内源基因或基因片段。期望地，缺损基因或基因片段也编码一种标记，该标记可用于筛选编码多肽的基因已被修饰或破坏的转化体。

另外，可使用与多肽编码序列互补的核苷酸序列通过已建立的反义技术进行编码本发明多肽的核酸序列的修饰或失活。尤其是，可通过引入与编码多肽的核酸序列互补的核苷酸序列降低或消除细胞对多肽的产生，所引入的核酸序列可在细胞中转录并能与细胞中产生的多肽 mRNA 杂交。在可使互补反义核苷酸序列与多肽 mRNA 杂交的条件下，翻译的多肽的量因而降低或消除。

优选地，按照本发明的方法修饰的细胞是微生物来源的，例如，与细胞同源或异源的适用于产生希望的蛋白质产物的真菌株。

本发明进一步涉及亲代细胞的突变细胞，其包含编码多肽的核酸

序列或其控制序列的破坏或缺失，这导致突变细胞比亲代细胞产生更少的多肽。

如此产生的多肽缺陷型突变细胞，尤其可作为宿主细胞用于同源和/或异源多肽的表达。因此，本发明进一步涉及产生同源或异源多肽的方法，包括（a）在有利于多肽产生的条件下培养突变细胞；和（b）回收该多肽。在本文内容中，术语“异源多肽”在此被定义为对于宿主细胞非自身的多肽、经修饰已改变天然序列的天然蛋白质，或由于用重组 DNA 技术对宿主细胞操作而定量改变其表达的天然蛋白质。

在又另一方面，本发明涉及通过一种细胞的发酵产生基本上无二肽氨肽酶活性的蛋白质产物的方法，这种细胞产生本发明的多肽及目标蛋白质产物。该方法包括：在发酵过程中或完成之后向发酵液中加入有效量的能抑制二肽氨肽酶活性的试剂，从发酵液中回收目标产物，以及任选地对回收的产物进一步纯化。

在又另一方面，本发明涉及产生基本上无二肽氨肽酶活性的蛋白质产物的方法，其中目标蛋白质产物由细胞中存在的编码本发明的多肽的 DNA 序列所编码。该方法包括：在允许产物表达的条件下培养细胞，对产生的培养液进行 pH 和温度组合处理以基本上降低二肽氨肽酶活性，以及从培养液中回收该产物。另外，可对从培养液中回收的酶制剂进行 pH 和温度组合处理。pH 和温度组合处理可任选地与用二肽氨肽酶抑制剂处理结合使用。

pH 和温度组合处理优选地在 pH9-11 和 40-75℃ 下进行足够的时间，以达到希望的效果，一般 30-60 分钟足够。

依照本发明的这一方面，去除至少 60%、优选地至少 75%、更优选地至少 85%、更优选地至少 95%、最优选地至少 99% 的二肽氨肽酶活性是可能的。预期用这些方法可获得二肽氨肽酶活性的完全去除。

可通过本领域中已知的方法进行用于目标产物的培养和纯化的方法。

在真核多肽尤其是真菌蛋白质如酶的产生中，用于产生基本上不含二肽氨肽酶产物的本发明的方法是特别有意义的。这些酶可选自，

例如，淀粉分解酶、脂肪分解酶、蛋白水解酶、溶胞酶、氧化还原酶或植物细胞壁降解酶。这些酶的实例包括：氨肽酶、淀粉酶、淀粉葡糖苷酶、糖酶、羧肽酶、过氧化氢酶、纤维素酶、几丁质酶、角质酶、环化糊精糖基转移酶、脱氧核糖核酸酶、酯酶、半乳糖苷酶、 β -半乳糖苷酶、葡糖淀粉酶、葡糖氧化酶、葡糖苷酶、盐过氧化物酶(haloperoxidase)、半纤维素酶、转化酶、异构酶、漆酶、连接酶、脂肪酶、裂合酶、甘露糖苷酶、氧化酶、果胶降解酶、过氧化物酶、肌醇六磷酸酶、酚氧化酶、多酚氧化酶、蛋白水解酶、核糖核酸酶、转移酶、谷氨酰胺转移酶或木聚糖酶。二肽氨肽酶缺陷的细胞也可用于表达有药学意义的异源蛋白质，如激素、生长因子、受体等等。

应当理解，术语“真核多肽”不仅包括天然多肽也包括为了提高活性、热稳定性、pH 耐受性等而通过氨基酸置换、缺失或加成或其它这类修饰方法修饰的多肽，如酶。

在另一方面，本发明涉及通过本发明的方法产生的、基本上无二肽氨肽酶活性的蛋白质产物。

产生蛋白质水解产物的方法

本发明的多肽可用于蛋白质水解产物的产生，用于提高水解程度和香味产生。

本发明进一步涉及结合内肽酶使用本发明的多肽引起富含蛋白质材料高度水解的方法。该方法包括用该多肽和一种内肽酶对蛋白质底物的处理。可用酶同时或连续地处理底物。

以蛋白质水解方法中常规使用的有效量向蛋白质底物加入本发明的多肽，优选地为每 100g 蛋白质约 0.1-约 100,000 二肽氨肽酶单位(DPAPU)，更优选地每 100g 蛋白质约 1-约 10,000 二肽氨肽酶单位(DPAPU)。在此定义，一个二肽氨肽酶单位(DPAPU)是在特定条件下每分钟从 Ala-Pro-对硝基苯胺(Sigma Chemical Co., St. Louis MO)中释放 1 微摩尔对硝基苯胺所需的酶量。

内肽酶可获自：芽孢杆菌属菌株，优选地地衣芽孢杆菌或枯草芽

孢杆菌、葡萄球菌属 (*Staphylococcus*) 菌株, 优选地金黄色葡萄球菌 (*Staphylococcus aureus*)、链霉菌属菌株, 优选地热普通链霉菌 (*Streptomyces thermovularis*) 或灰色链霉菌 (*Streptomyces griseus*)、放线菌属的种的菌株、曲霉属菌株, 优选地棘孢曲霉、泡盛曲霉、臭曲霉、构巢曲霉、黑曲霉或米曲霉, 或者镰孢属菌株, 优选地 *Fusarium venenatum*。

以蛋白质水解方法中常规使用的有效量向蛋白质底物加入内肽酶, 优选地为约 0.05-约 15 AU/100g 蛋白质, 更优选地约 0.1-约 8 AU/100g 蛋白质。一个 AU (Anson 单位) 被定义为这样的酶量, 在标准条件 (即, 25°C、pH7.5 和 10 分钟反应时间) 下以初始速率消化血红蛋白, 使得每分钟释放一定量的 TCA 可溶性产物, 该产物与酚试剂产生与一个毫当量酪氨酸相同的颜色。在请求丹麦的 Novo Nordisk A/S 后, 分析方法 AF 4/5 是可用的, 该方法在此引入作为参考。

可在酶制剂不失活的任何适宜温度下, 优选地约 20°C-约 70°C 下, 进行酶处理, 即底物与酶制剂的温育。依照公认的实践, 适当灭活酶制剂的方法包括: 将温育混合物的温度提高到酶失活的温度, 例如约 70°C 以上, 或者类似地, 将温育混合物的 pH 降低至酶失活的值, 例如约 4.0 以下。

此外, 本发明的方法导致蛋白质底物水解程度的提高。在此使用时, 水解程度 (DH) 是蛋白水解酶所水解的蛋白质中氨键总数的百分数。

在本发明的另一方面, 水解产物具有含量升高的 Ala、Arg、Asp、Gly 和/或 Val, 例如 1.1 倍高。

在本发明的另一方面, 本发明的多肽与一种氨肽酶协同作用, 在单独的任一酶活性不能水解三肽时可水解该三肽。在一种优选实施方案中, 该氨肽酶是如 WO 96/28542 所述从米曲霉中获得的氨肽酶 I。

本发明也涉及获得富含游离谷氨酸和/或肽结合谷氨酸残基的蛋白质水解产物的方法, 该方法包括:

(a) 对底物施行脱酰胺方法; 和

(b) 对底物施行具有二肽氨肽酶活性的多肽的作用。

这两个步骤可同时进行，或者第二个步骤可在第一个步骤之后进行。

本发明的这些方法产生味道极好的蛋白质水解产物，因为无论是游离的或是肽结合的谷氨酸（Glu），都在蛋白质水解产物的香味和适口性中起重要作用。这些方法也产生具有改善的功能性尤其是改善的溶解度、改善的乳化特性、提高的水解程度及改善的泡沫特性的蛋白质水解产物。

酰胺（谷氨酰胺或天冬酰胺）通过氨的释放向荷电酸（谷氨酸或天冬氨酸）的转化被称为脱酰胺作用。脱酰胺作用可作为非酶法或酶法脱酰胺过程发生。

在一种优选实施方案中，脱酰胺作用以酶法脱酰胺过程进行，例如，对底物施用一种谷氨酰胺转移酶和/或肽谷氨酰胺酶。

谷氨酰胺转移酶可以是任何适宜来源的，包括哺乳动物，参见例如，JP 1050382 和 JP 5023182，包括活化的因子 XIII，参见例如，WO 93/15234；来源于鱼，参见例如，EP 555,649；从微生物获得，参见例如，EP 379,606、WO 96/06931 和 WO 96/22366。在一种优选实施方案中，谷氨酰胺转移酶从卵菌（Oomycete）获得，包括疫霉属（*Phytophthora*）菌株，优选地恶疫霉（*Phytophthora cactorum*），或腐霉属（*Pythium*）菌株，优选地畸雌腐霉（*Pythium irregulare*）、*Pythium* sp.、间型腐霉（*Pythium intermedium*）、终极腐霉（*Pythium ultimum*）或周雄腐霉（*Pythium peritremum*）（或缠器腐霉（*Pythium periplocum*））。在另一种优选实施方案中，谷氨酰胺转移酶是细菌来源的，获自芽孢杆菌属菌株，优选地枯草芽孢杆菌、链轮丝菌属（*Streptoverticillium*）菌株，优选地茂原链轮丝菌（*Streptoverticillium mobaraensis*）、灰肉色链轮丝菌（*Streptoverticillium griseocarneum*）或肉桂链轮丝菌（*Streptoverticillium cinnamomeum*），以及链霉菌属菌株，优选地利迪链霉菌（*Streptomyces lydicus*）。

肽谷氨酰胺酶可以是肽谷氨酰胺酶 I（肽酰谷氨酰胺酶； EC

3.5.1.43) 或肽谷氨酰胺酶 II (蛋白质-谷氨酰胺谷氨酰胺酶; EC 3.5.1.44) 或其任何混合物。肽谷氨酰胺酶可获自: 曲霉属菌株, 优选地日本曲霉、芽孢杆菌属菌株, 优选地环状芽孢杆菌、隐球酵母属菌株, 优选地浅白隐球酵母 (*Cryptococcus albidus*)、或德巴利酵母属 (*Debaryomyces*) 菌株, 优选地克洛德巴利酵母 (*Debaryomyces kloecheri*)。

以脱酰胺方法中常规使用的有效量向蛋白质底物加入谷氨酰胺转移酶, 优选地约 0.01%-约 5% (w/w), 更优选地约 0.1%-约 1% (w/w) 的酶制剂, 与底物量有关。

以脱酰胺方法中常规使用的有效量向蛋白质底物加入肽谷氨酰胺酶, 优选地每 100g 底物约 0.01-约 100,000 PGase 单位, 更优选地每 100g 底物约 0.1-约 10,000 PGase 单位。

可根据 Cedrangoro 等人的方法 (1965, 酶学 (*Enzymologia*) 29: 143) 测定肽谷氨酰胺酶活性。根据该方法, 将用 1N NaOH 调节至 pH6.5 的 0.5ml 酶样品装入一个小瓶。然后向瓶中加入 1ml 硼酸盐 pH10.8 缓冲液。用 5N 硫酸吸收释放的氨, 用奈氏 (Nessler's) 试剂使混合物显色, 于 420nm 测定。一个 PGase 单位是在这些条件下每分钟能产生 1 微摩尔氨的酶量。

另外, 可根据 US 3,857,967 或以下实施例 17 所述的方法测定肽谷氨酰胺酶活性。

在本发明方法的步骤 (b) 中, 对底物施用本发明的多肽。以蛋白质水解方法中常规使用的有效量向蛋白质底物加入本发明的多肽, 优选地范围约 0.001-约 0.5 AU/100g 底物, 更优选地约 0.01-约 0.1 AU/100g 底物。

在另一种实施方案中, 可用本发明的方法产生富含游离谷氨酸和/或肽结合谷氨酸残基的水解产物, 该方法进一步包括:

(c) 对底物施用一种或多种非特异作用的内肽酶和/或外肽酶。

该步骤可与步骤 (a) 和 (b) 同时进行, 或可在步骤 (a) 和 (b) 之后进行。

在一种优选实施方案中，非特异作用的内肽酶和/或外肽酶获自：曲霉属菌株，优选地黑曲霉、米曲霉或酱油曲霉（*Aspergillus sojiae*），或芽孢杆菌属菌株，优选地解淀粉芽孢杆菌、迟缓芽孢杆菌、地衣芽孢杆菌或枯草芽孢杆菌。

以蛋白质水解方法中常规使用的有效量向底物加入非特异作用的内肽酶和/或外肽酶，优选地约 0.05-约 15 CPU/100g 底物，更优选地约 0.1-约 5 CPU/100g 底物。一个 CPU（酪蛋白蛋白酶单位）被定义为在标准条件下，即于 25℃ 和 pH9.5 温育 30 分钟时，每分钟从酪蛋白释放 1 微摩尔伯氨基（与丝氨酸标准对比测定）的酶量。在请求丹麦 Bagsvaerd 的 Novo Nordisk A/S 后可使用分析方法 AF 228/1，该方法在此引入作为参考。

可在酶制剂不失活的任何温度、优选地约 20℃-约 70℃ 下进行每种酶处理。然后可灭活酶制剂，方法是提高温度，例如至约 70℃ 以上，或降低 pH，例如降至约 4.0 以下。

本发明的方法中使用的蛋白质底物可包括完整的蛋白质、预水解的蛋白质（即肽）或其混合物。蛋白质底物可以是植物或动物来源的。优选地，蛋白质底物是植物来源的，例如：大豆蛋白、谷类蛋白，例如小麦面筋、玉米面筋、大麦、黑麦、燕麦、水稻、玉米醇溶蛋白、羽扇豆、棉籽蛋白、油菜籽蛋白、花生、苜蓿蛋白、豌豆蛋白、蚕豆蛋白、黑芝麻蛋白或向日葵。动物来源的蛋白质底物可以是乳清蛋白、酪蛋白、肉类蛋白、鱼类蛋白、红细胞、蛋清、明胶或乳白蛋白。

本发明也涉及通过这些方法产生的蛋白质水解产物。

其它应用

本发明也涉及用本发明多肽使酶失活的方法。

此外，本发明的多肽可用于希望特异性切割肽序列的多种目的。例如，某些蛋白质或肽以无活性的前体的形式合成，其在成熟蛋白质的 N 端含有许多另外的氨基酸残基。本发明的多肽能提供必要的翻译后加工以激活这些前体蛋白质。

组合物

在又另一方面，本发明涉及含有本发明的多肽的多肽组合物。优选地，该组合物富含本发明的多肽。在本文内容中，术语“富含”是指多肽组合物的二肽氨肽酶活性提高，例如，富含系数为 1.1。

多肽组合物可含有本发明的多肽作为主要的酶组分，例如，单组分多肽组合物。此外，组合物可包含多种酶活性，如氨肽酶、淀粉酶、糖酶、羧肽酶、过氧化氢酶、纤维素酶、几丁质酶、角质酶、环化糊精糖基转移酶、脱氧核糖核酸酶、酯酶、 α -半乳糖苷酶、 β -半乳糖苷酶、葡糖淀粉酶、 α -葡糖苷酶、 β -葡糖苷酶、盐过氧化物酶、转化酶、漆酶、脂肪酶、甘露糖苷酶、氧化酶、果胶降解酶、肽谷氨酰胺酶、过氧化物酶、肌醇六磷酸酶、多酚氧化酶、蛋白水解酶、核糖核酸酶、谷氨酰胺转移酶或木聚糖酶。其它酶可用下列微生物产生：属于曲霉属，优选地棘孢曲霉、泡盛曲霉、黑曲霉或米曲霉，或者木霉属、腐质霉属，优选地 *Humicola insolens*，或者镰孢属，优选地杆孢状镰孢、*Fusarium cerealis*、*Fusarium crookwellense*、黄色镰孢、禾谷镰孢、禾赤镰孢、异孢镰孢、合欢木镰孢、尖镰孢、网状镰孢、粉红镰孢、接骨木镰孢、肤色镰孢、拟枝孢镰孢、*Fusarium sulphureum*、*Fusarium torulosum*、*Fusarium trichothecioides* 或 *Fusarium venenatum*。

在一种优选实施方案中，本发明涉及一种增香组合物，其含有具有二肽氨肽酶活性的多肽和一种载体。可使用本领域已知的任何合适的载体。在另一种优选实施方案中，该增香组合物进一步包含一种或多种非特异作用的内肽酶和/或外肽酶。在另一种优选实施方案中，该增香组合物进一步包含一种或多种特异作用的内肽酶和/或外肽酶。

在一种优选实施方案中，特异作用的蛋白水解酶是一种内肽酶如谷氨酰内肽酶 (EC 3.4.21.19)；赖氨酰内肽酶 (EC 3.4.21.50)；亮氨酰内肽酶 (EC 3.4.21.57)；甘氨酰内肽酶 (EC 3.4.22.25)；脯氨酰内肽酶 (EC 3.4.21.26)；胰蛋白酶 (EC 3.4.21.4) 或胰蛋白酶样 (赖氨酸/精氨酸特异的) 内肽酶；或肽酰-Asp 金属内肽酶 (EC 3.4.24.33)。

谷氨酰内肽酶 (EC 3.4.21.19) 优选地可获自: 芽孢杆菌属菌株尤其是地衣芽孢杆菌和枯草芽孢杆菌、葡萄球菌属菌株尤其是金黄色葡萄球菌、链霉菌属菌株尤其是热普通链霉菌和灰色链霉菌, 或放线菌属菌株。

赖氨酰内肽酶 (EC 3.4.21.50) 优选地可获自: 无色杆菌属 (*Achromobacter*) 菌株, 尤其是水解无色杆菌 (*Achromobacter lyticus*), 溶杆菌属 (*Lysobacter*) 菌株, 尤其是产酶溶杆菌 (*Lysobacter enzymogenes*), 或假单胞菌属菌株, 尤其是铜绿假单胞菌 (*Pseudomonas aeruginosa*)。

亮氨酰内肽酶 (EC 3.4.21.57) 可以是植物来源的。

甘氨酰内肽酶 (EC 3.4.22.25) 优选地可从木瓜植物 (番木瓜 (*Carica papaya*)) 获得。

脯氨酰内肽酶 (EC 3.4.21.26) 优选地可从黄杆菌属 (*Flavobacterium*) 菌株获得, 或者可以是植物来源的。

胰蛋白酶样内肽酶优选地可从镰孢属菌株尤其是尖镰孢获得, 例如, 如 WO 89/06270 或 WO 94/25583 所述。

肽酰-Asp 金属内肽酶 (EC 3.4.24.33) 优选地可从假单胞菌属菌株尤其是莓实假单胞菌 (*Pseudomonas fragi*) 获得。

在另一种优选实施方案中, 特异作用的蛋白水解酶是一种可从肽的任一端作用的外肽酶。

在一种优选实施方案中, 特异作用的蛋白水解酶是一种氨肽酶如亮氨酰氨肽酶 (EC 3.4.11.1); 或三肽氨肽酶 (EC 3.4.11.4)。

在另一种优选实施方案中, 特异作用的蛋白水解酶是一种羧肽酶如脯氨酸羧肽酶 (EC 3.4.16.2); 羧肽酶 A (EC 3.4.17.1); 羧肽酶 B (EC 3.4.17.2); 羧肽酶 C (EC 3.4.16.5); 羧肽酶 D (EC 3.4.16.6); 赖氨酸 (精氨酸) 羧肽酶 (EC 3.4.17.3); 甘氨酸羧肽酶 (EC 3.4.17.4); 丙氨酸羧肽酶 (EC 3.4.17.6); 谷氨酸羧肽酶 (EC 3.4.17.11); 肽酰-二肽酶 A (EC 3.4.15.1); 或肽酰-二肽酶 A (EC 3.4.15.5)。

多肽组合物可按照本领域已知的方法制备, 并且可以是液体或干

燥组合物的形式。可通过本领域已知的方法稳定该多肽。

本发明也涉及含有通过本发明的方法获得的蛋白质水解产物的食品，例如焙烤产品。这类食品显示提高的感官质量，如香味、适口性、口感、香气和表层颜色的改善。

在本文内容中，术语“焙烤产品”包括从生面团制备的、具有软或脆的特征的任何食品。可通过本发明方便地产生的焙烤产品，无论是白色、浅色或深色类型，其实例包括：面包，尤其是白面包、全麦面包或黑麦面包，一般为块或卷的形式；法国棍形面包；皮塔面包；玉米面豆卷；蛋糕；薄烤饼；饼干；脆面包等等。

这类焙烤产品通常从含有面粉和水并且一般要发酵的生面团制备。可以以多种方式发酵生面团，如加入碳酸氢钠等，或加入发酵剂（发酵的生面团），但优选地通过加入合适的酵母培养物如酿酒酵母培养物（面包酵母）发酵生面团。可使用任何可商业获得的酿酒酵母株。

另外，在焙烤产品制备中使用的生面团可以是新鲜的或冷冻的。K. Kulp 和 K. Lorenz 在“冷冻和冷藏的生面团与奶蛋糊”中描述了冷冻生面团的制备。本发明的增香组合物一般以 0.01-5%、更优选地 0.1-3% 的量包含于生面团中。

在本发明的方法中，可将本发明的多肽、内肽酶、谷氨酰胺转移酶、肽谷氨酰胺酶、一种或多种特异和/或非特异作用的内肽酶和/或外肽酶、和/或一种或多种上述酶，分开地或同时地加入制备生面团的混合物中，或加入制备生面团的任何配料如面粉中。

本发明进一步涉及一种预混合物，例如为面粉组合物的形式，用于生面团和/或由生面团制成的焙烤产品，其中预混合物含有本发明的多肽或增香组合物及一种载体或焙烤配料，以及任选地上述一种或多种其它的酶。

在另一种实施方案中，预混合物包含通过本发明的方法获得的水解产物。

可通过将相关的酶与一种合适的载体如面粉、淀粉、糖或盐相混合，制备预混合物。预混合物可含有其它的生面团改善添加剂和/或面

包改善添加剂。

在本文内容中，术语“预混合物”是通常含有面粉的发酵粉的混合物，其制备是为了允许在指定条件下贮存，以及为生面团制备过程中的处理提供方便。这种预混合物可方便地用于工业和商业面包烤制工厂和设施，以及零售面包房。

本发明也涉及通过本发明的方法产生的水解产物的应用，其作为食品如焙烤食品的添加剂，提高感官质量如香味、适口性和香气。

通过本发明的方法获得的、富含游离谷氨酸和/或肽结合谷氨酸残基的水解产物可用于多种工业用途，尤其是，需要掺入功能性蛋白质之处。

例如，本发明也涉及食品，其含有通过本发明的方法获得的、富含游离谷氨酸和/或肽结合谷氨酸残基的水解产物，并且涉及动物饲料添加剂，其含有通过本发明的方法获得的、富含游离谷氨酸和/或肽结合谷氨酸残基的水解产物。

下列实施例进一步描述了本发明，不应把这些实施例看作是限制本发明的范围。

实施例

实施例 1: FLAVOURZYME™ 二肽氨肽酶 I 的纯化

从 FLAVOURZYME™ 培养液 (Novo Nordisk A/S, Bagsvaerd, 丹麦) 中纯化二肽氨肽酶 I。通过在由碳源和氮源和痕量金属组成的培养基中培养米曲霉株 1568 (ATCC 20386) 产生 FLAVOURZYME™ 培养液。首先，用 180ml 20mM 磷酸钠 (pH7.0) 缓冲液稀释培养液 (20ml, 含 720mg 蛋白质)，并用具有 0.45 μ m 滤器的 Nalgene Filterware 过滤。将过滤的溶液加样于含 31ml Q-Sepharose, Big Beads (Pharmacia Biotech AB, Uppsala, 瑞典) 并用 20mM 磷酸钠 (pH7.0) 缓冲液预平衡的 24 \times 130mm 柱上。用 20mM 乙酸钠缓冲液 (pH3.5) 洗脱二肽氨肽酶 I。在 50mM 磷酸钠 (pH7.5) 缓冲液中用 1mg/ml Ala-Pro-对硝基苯胺作为底物于 405nm 监测二肽氨肽酶 I 活性。通过用 PM10 膜 (Amicon, New

Bedford, MA) 超滤将获得的含二肽氨肽酶 I 活性的溶液浓缩为 38ml, 然后用 20mM Na_2HPO_4 溶液调节 pH 至 pH7.0。

将获得的溶液加样于含 31ml Q-Sepharose Big Beads (Pharmacia Biotech AB, Uppsala, 瑞典) 并用 20mM 磷酸钠(pH7.0)缓冲液预平衡的 24×130mm 柱上。用 20mM 磷酸钠(pH7.0)缓冲液中的 0-0.3M NaCl 梯度洗脱二肽氨肽酶 I。如上所述监测级分的二肽氨肽酶 I 活性。使用对 20mM 磷酸钠(pH7.0)缓冲液的超滤收集、合并、脱盐并浓缩含二肽氨肽酶 I 活性的级分。

将获得的溶液加样于用 20mM 磷酸钠(pH7.0)缓冲液预平衡的 MonoQ 16/10 (20ml) 预装柱 (Pharmacia Biotech AB, Uppsala, 瑞典) 上。用 20mM 磷酸钠(pH7.0)缓冲液中的 0-0.27M NaCl 梯度洗脱二肽氨肽酶 I。如上所述监测级分的二肽氨肽酶 I 活性。如上所述使用超滤收集、合并 0.200-0.212M NaCl 之间的级分, 并将其用 1.7M $(\text{NH}_4)_2\text{SO}_4$ /20mM 磷酸钠(pH7.0)缓冲液再缓冲。

将获得的溶液加样于含苯基 Superose 树脂(Pharmacia Biotech AB, Uppsala, 瑞典) 并用 1.7M $(\text{NH}_4)_2\text{SO}_4$ /20mM 磷酸钠(pH7.0)缓冲液预平衡的 7×50mm 柱上。用 20mM 磷酸钠(pH7.0)缓冲液中的反向 1.7-0M $(\text{NH}_4)_2\text{SO}_4$ 梯度洗脱二肽氨肽酶 I。如上所述监测级分的二肽氨肽酶 I 活性。发现对 Ala-Pro-对硝基苯胺具有最高活性的两种级分根据 SDS-PAGE 分析至少 95%同源。主要带具有大约 95kDa (93-96kDa) 的分子量。

实施例 2: 蛋白质序列测定和氨基酸分析方法

用配有联机 HPLC 和液相三氟乙酸 (TFA) 输送的 Applied Biosystems 476A 蛋白质测序仪 (Perkin Elmer/ Applied Biosystems Division, Foster City, CA) 对实施例 1 所述部分纯化的二肽氨肽酶 I 和二肽氨肽酶 I 的消化片段进行 N-端序列测定。将纯化的二肽氨肽酶 I 样品从 SDS-PAGE 凝胶转印至 Novex PVDF 膜 (Novex, San Diego, CA) 上, 并用测序试剂 (Perkin Elmer/ Applied Biosystems Division,

Foster City, CA) 从印迹柱上测序。通过联机 HPLC 实现乙内酰苯硫脲-氨基酸的检测, 其中使用含有 3.5% 溶于水的四氢呋喃及 15ml 含乙酸、乙酸钠和己烷磺酸钠的预混合浓缩液 (Perkin Elmer/ Applied Biosystems Division, Foster City, CA) 的缓冲液 A, 以及含乙腈的缓冲液 B。收集数据并在 Macintosh IIsi 上用 Applied Biosystems 610 数据分析软件分析。

也用胰蛋白酶对部分纯化的二肽氨肽酶 I (实施例 1 中的 MonoQ 峰) 进行凝胶内消化, 产生酶的肽片段用于蛋白质序列测定。使用 8-16% Novex Tris-甘氨酸凝胶 (Novex, San Diego, CA) 通过 SDS-PAGE 电泳分离胰蛋白酶消化的二肽氨肽酶 I。切下对应于 95kDa 分子量的八个考马斯蓝染色凝胶片, 并用 50% 乙腈中的 200mM NH_4HCO_3 充分洗涤。随后将凝胶片在 1mM 二硫苏糖醇 (DTT) 中于 37°C 还原 20 分钟, 并在暗处于室温下用等体积的 100mM 碘乙酸烷基化 20 分钟。用 50% 乙腈中的 200mM NH_4HCO_3 反复洗涤凝胶片。移除上清液, 在 Speed-Vac (Savant Instruments, Farmingdale, NY) 上干燥凝胶片。在每毫升 135mM NH_4HCO_3 中含 0.033mg 测序级修饰的猪胰蛋白酶 (Promega, Madison, WI) 的溶液中再水合凝胶片。该溶液的制备方法是将 1 份每毫升胰蛋白酶重悬缓冲液 (Promega, Madison, WI) 中的 0.1mg 胰蛋白酶稀释至 2 份 200mM NH_4HCO_3 中。凝胶片在 37°C 温育 20 小时。通过用 60% 乙腈中的 0.1% TFA 重复洗涤, 每次 1 小时, 从凝胶中提取肽片段。干燥提取的肽, 在 25% 乙腈中的 0.05% TFA 中重建, 然后进一步稀释于 0.05% TFA 中。用 Micropure 0.45 μm 过滤单元 (Amicon, Inc., Beverly, MA) 过滤肽片段。然后使用配备有 2.1 \times 250mm Vydac C18-RP 柱 (5 微米) 的 Hewlett-Packard 1090L HPLC 通过反相 HPLC 分离肽片段。使用分级梯度与 80% 乙腈中的 0.1% TFA 作为洗脱液。手工收集肽样品, 然后进行 N 端序列测定。

也对部分纯化的二肽氨肽酶 I 施用溴化氰, 产生酶的肽片段用于序列测定。用溴化氰消化二肽氨肽酶 I 的方法是, 用少量溴化氰晶体重建部分纯化的溶于 70% 甲酸的干燥样品, 并在暗处于室温下温育 18

小时。使用 10-20% Novex Tricine 凝胶 (Novex, San Diego, CA) 通过 SDS-PAGE 电泳分离肽片段, 并如上所述测序。

二肽氨肽酶 I 的 N 端序列测定显示 N 端被明显地阻断。获得如下的弱序列, 其中括号中的氨基酸残基未 100% 确定, 以 X 标记的残基无法确定:

肽 1: XEGSKRLTFXETVVKQAIT(P) (SEQ ID NO: 3)

溴化氰降解的片段具有下列氨基酸序列, 其中加下划线的氨基酸残基与酿酒酵母的二肽氨肽酶 I (Anna-Arriola 和 Herskowitz, 1994, 酵母 10: 801-810; Galisson 和 Dujon, 1996, 酵母 12: 877-885) 100% 匹配:

肽 2: QRLPPGFSPDKKYPILFTPYGG (SEQ ID NO: 4)

肽 3: KYIGPIK (SEQ ID NO: 5)

肽 4: GEGSKRL (SEQ ID NO: 6)

用胰蛋白酶凝胶内消化产生下列肽, 其中加下划线的氨基酸残基与实施例 7 所述米曲霉二肽氨肽酶 I 核酸序列的推定氨基酸序列 100% 匹配:

肽 5: XPILETPY (SEQ ID NO: 7)

肽 6: XVPLMPDQ(Q)GDIQYAQ (SEQ ID NO: 8)

实施例 3: 基因组 DNA 的提取

将米曲霉 1568 在 25ml 0.5% 酵母提取物-2% 葡萄糖 (YEG) 培养基中于 37°C 和 250 转/分下培养 24 小时。然后通过微孔布 (Miracloth) (Calbiochem, La Jolla, CA) 过滤收集菌丝体, 并用 25ml 10mM Tris-1mM EDTA (TE) 缓冲液洗涤一次。从菌丝体制剂中排出多余的缓冲液, 随后冷冻于液氮中。在电动咖啡研磨机中将冷冻的菌丝体制剂研磨成细粉末, 将粉末加入含有 20ml TE 缓冲液和 5ml 20% w/v 十二烷基硫酸钠 (SDS) 的一次性塑料离心管中。轻轻颠倒混合物数次确保其混匀, 用等体积的酚: 氯仿: 异戊醇 (25: 24: 1, v/v/v) 抽提两次。向抽提的样品中加入乙酸钠 (3M 溶液) 至终浓度为 0.3M, 随后加入

2.5 倍体积用冰预冷的乙醇沉淀 DNA。以 15,000×g 将管离心 30 分钟沉淀 DNA。使 DNA 沉淀风干 30 分钟，之后重悬浮于 0.5ml TE 缓冲液中。向重悬浮的 DNA 沉淀中加入不含 DNase 的核糖核酸酶 A 至终浓度为 100µg/ml，然后于 37℃ 温育该混合物 30 分钟。加入蛋白酶 K (200µg/ml)，将管于 37℃ 再温育 1 小时。最后，用酚：氯仿：异戊醇抽提样品两次，并用乙醇沉淀 DNA。沉淀的 DNA 以 70%乙醇洗涤，真空下干燥，重悬浮于 TE 缓冲液中，4℃ 贮存。

实施例 4: 米曲霉 1568 二肽氨肽酶 I 的 PCR 扩增

设计正向简并寡核苷酸引物，对应于肽序列 DW(I/V)YEEE，这是在公布的大多数二肽氨肽酶 I 蛋白质序列中发现的一种保守基元。设计反向简并寡核苷酸引物，对应于实施例 2 所述的肽 2 (SEQ ID NO: 4) 的部分肽 PPGFSDKKYP。按照厂商说明书用 Applied Biosystems 394 型 DNA/RNA 合成仪合成如下所示的简并寡核苷酸引物，用于从米曲霉 1568 基因组 DNA 中 PCR 扩增二肽氨肽酶 I 基因片段：

正向引物：5'-GAYTGGITITAYGARGARGAR-3' (SEQ ID NO: 9)

反向引物：5'-GGRTAYTTYTTRTCIGGISWRAAICCIGGIGG-3' (SEQ ID NO: 10)

(R=A 或 G, Y=C 或 T, S=G 或 C, W=A 或 T, I=肌苷)

使用大约 1µg 如实施例 3 所述从米曲霉 1568 中分离的基因组 DNA 作为模板准备扩增反应物 (100µl)。每一反应物含有下列成分：1µg 基因组 DNA、40pmol 正向引物、40pmol 反向引物、200µM 每种 dATP、dCTP、dGTP 和 dTTP、1×Taq 聚合酶缓冲液 (Perkin-Elmer 公司, Branchburg, NJ) 和 2.5 单位的 Taq 聚合酶 (Perkin-Elmer 公司, Branchburg, NJ)。在 Perkin-Elmer 480 型热循环仪中温育该反应物，程序设定如下：第 1 循环，94℃ 2 分钟、45℃ 1 分钟和 72℃ 1 分钟；第 2-30 循环，每次循环 94℃ 1 分钟、45℃ 1 分钟和 72℃ 1 分钟。在 1% 琼脂糖凝胶 (Eastman Kodak, Rochester, NY) 中分离反应产物。从凝胶中切下大约 1.0kb 的产物带，并按照厂商说明书用 GenElute 旋转

柱 (Supelco, Bellefonte, PA) 纯化。随后将纯化的 PCR 产物克隆至 pCRII 载体 (Invitrogen, San Diego, CA) 中, 使用 lac 正向和反向引物 (New England BioLabs, Beverly, MA) 测定 DNA 序列。

用上述二肽氨肽酶 I PCR 引物从米曲霉 1568 中扩增由大约 321 个密码子 (963bp) 组成的二肽氨肽酶 I 基因片段。DNA 序列分析显示, 扩增的基因片段编码一部分相应的米曲霉 1568 二肽氨肽酶 I 基因。二肽氨肽酶 I 基因片段用于探查米曲霉 1568 基因组 DNA 文库。

实施例 5: DNA 文库的构建

在噬菌体克隆载体 λ ZipLox (Life Technologies, Gaithersburg, MD) 中构建基因组 DNA 文库。首先, 用 Tsp509I 部分消化总细胞 DNA, 并在 1% 琼脂糖凝胶中按大小分级分离。切下以 3-7kb 大小迁移的 DNA 片段, 用 Prep-a-Gene 试剂 (BioRad Laboratories, Hercules, CA) 从凝胶中洗脱。将洗脱的 DNA 片段与 EcoRI 酶切并去磷酸化的 λ ZipLox 载体臂 (Life Technologies, Gaithersburg, MD) 连接, 用商业包装提取物 (Stratagene, La Jolla, CA) 包装该连接混合物。在大肠杆菌 Y1090ZL 细胞 (Life Technologies, Gaithersburg, MD) 中平板接种并扩增包装的 DNA 文库。未扩增的基因组 DNA 文库含有 3.1×10^6 pfu/ml (不含 DNA 的背景值为 2.0×10^4 pfu/ml)。

实施例 6: 二肽氨肽酶 I 克隆的鉴定

使用来源于米曲霉 1568 的二肽氨肽酶 I PCR 片段作为探针, 通过噬菌斑杂交从实施例 5 所述文库中筛查出大约 10000 个噬菌斑。用 UV Stratalinker (Stratagene, La Jolla, CA) 将 DNA 交联至膜 (Hybond N+, Amersham, Arlington Heights, IL) 上。该膜在含 5 \times SSPE、0.3% SDS、50% 甲酰胺和 10 μ g/ml 变性并剪切的鲱精 DNA 的杂交液中 45 $^{\circ}$ C 浸泡 3 小时。用随机引物 DNA 标记试剂盒 (Boehringer Mannheim, Mannheim, 德国) 放射性标记如实施例 2 所述从米曲霉 1568 基因组 DNA 中分离的二肽氨肽酶 I 基因片段, 加入 NaOH 至终浓度为 0.1M

使之变性，并以每毫升杂交液大约 1×10^6 cpm 的活性加至杂交液中。将该混合液在振荡水浴中 45°C 温育过夜。温育后，在 55°C 下以含 0.2% SDS 的 $2 \times \text{SSC}$ 洗膜一次，随后在相同温度下以 $2 \times \text{SSC}$ 洗两次。将膜在印迹纸上干燥 15 分钟，包裹于 SaranWrap™ 中，并使用增感屏（Kodak, Rochester, NY）将其在 -70°C 下暴露于 X 光胶片过夜。

根据与探针的强杂交信号的产生，筛选出三个噬菌斑作进一步研究，分别被命名为大肠杆菌 DH5 α MWR52A、DH5 α MWR52B 和 DH5 α MWR52C。这三个噬菌斑在大肠杆菌 Y1090ZL 细胞中纯化两次，随后通过大肠杆菌 DH10BZL 细胞（Life Technologies, Gaithersburg, MD）的感染用体内剪切法从 pZL1 衍生物 λ ZipLox 载体（D'Alessio 等人, 1992, 焦点（Focus®）14:76）中剪切下二肽氨肽酶 I 基因。将含有三种质粒的菌落接种于 3ml LB+50 $\mu\text{g}/\text{ml}$ 羧苄青霉素培养基中，并于 37°C 培养过夜。使用 Wizard 373 DNA 纯化试剂盒（Promega, Madison, WI）从每种培养物中小量制备 DNA。通过 DNA 序列测定确定二肽氨肽酶 I 编码质粒（pMWR52）。

实施例 7: 米曲霉 1568 二肽氨肽酶 I 基因的 DNA 序列分析

使用染料终止子化学的引物步移技术（Giesecke 等人, 1992, 病毒学方法杂志 (Journal of Virology Methods) 38:47-60），用 Applied Biosystems 377 型自动 DNA 测序仪（Applied Biosystems, Inc., Foster City, CA）对实施例 6 所述大肠杆菌 DH5 α MWR52 中 pMWR52 所含的二肽氨肽酶 I 基因的两条链进行 DNA 序列测定。设计寡核苷酸测序引物，对应于二肽氨肽酶 I 基因中的互补序列，并按照厂商说明书在 Applied Biosystems 394 型 DNA/RNA 合成仪中合成。

编码米曲霉 1568 二肽氨肽酶 I 的基因的核苷酸序列显示于图 1 中（SEQ ID NO: 1）。克隆的插入片段序列分析显示出一个被 83bp 内含子隔断的 2396 个核苷酸的开放阅读框（不包括终止密码子）。该开放阅读框的 G+C 含量为 55.3%。推断的氨基酸序列编码 771 个氨基酸的蛋白质（SEQ ID NO: 2）。根据 van Heijne 原则（van Heijne, 1984,

分子生物学杂志 173: 243-251), 前 16 个氨基酸可能组成一种引导新生多肽进入内质网的分泌型信号肽(在图 1 中框出)。

如实施例 2 所述来源于纯化二肽氨肽酶 I 的部分肽的氨基酸序列在图 1 中用下划线标出, 并且与在米曲霉 1568 二肽氨肽酶 I cDNA 的推定氨基酸序列 (SEQ ID NO: 2) 中所发现的序列一致。

使用 Clustal 序列对比程序 (Higgins, 1989, CABIOS 5: 151-153) 将米曲霉 1568 二肽氨肽酶 I 的推定氨基酸序列与酿酒酵母二肽氨肽酶 I 的序列 (Anna-Arriola 和 Herskowitz, 1994, 酵母 10: 801-810) (SEQ ID NO: 11) 相比较, 发现 23.2% 的同一性。

实施例 8: 米曲霉 1568 二肽氨肽酶镰孢表达载体的构建

扩增米曲霉二肽氨肽酶 I 的编码区, 将产生的片段克隆至 pDM181 中用于在镰孢属中表达。pDM181 提供镰孢胰蛋白酶 (SP387) 启动子和终止子, 以及 bar 选择性标记基因。具体地, 使用针对第一个读框内 ATG 设计并向下游延伸 13bp 的有义引物 (P1) 和针对转录终止密码子区设计并向下游延伸 10bp 的反义引物 (P2), 通过 PCR 扩增该片段。为便于扩增片段的克隆, 有义和反义引物分别含有一个 *Swa*I 和一个 *Pac*I 限制性位点。按照厂商说明书用 ABI 394 型 DNA/RNA 合成仪 (Applied Biosystems, Inc., Foster City, CA) 合成以下所示的寡核酸引物。

*Swa*I

P1: 5'-GATTTAAATCACCATGAAGGTACGTCAATTCCACTG-3' (SEQ ID NO:12)

*Pac*I

P2: 5'-GTTAATTAATCTACTCCTCCAAGTCCTTCTTAGTCC-3' (SEQ ID NO:13)

50 μ l PCR 溶液 (10mM Tris-HCl pH8.3、50mM KCl、1.5mM MgCl₂、0.01% w/v 明胶) 含有大约 200ng pMWR52 DNA、200 μ M 每种 dATP、dCTP、dGTP 和 dTTP 以及 50pmol 上述每种 PCR 引物。加入 5 单位的 PWO 聚合酶 (Boehringer Mannheim, Indianapolis, IN)。于 95 $^{\circ}$ C 温育反应物 3 分钟, 并冷却至 80 $^{\circ}$ C。然后在 Perkin-Elmer

9600 热循环仪中对反应物进行 30 次循环，每一循环 95℃ 30 秒、57℃ 1 分钟和 72℃ 1 分钟。最后一次循环后，将反应物于 72℃ 温育 5 分钟。

通过用 *Swa*I 和 *Pac*I 消化分离预期的 2.4kb 片段，并克隆至用相同限制性内切核酸酶消化的 pDM181（图 2）中，生成 pMWR54（图 3）。为检验克隆的 PCR 片段的保真性，根据 Hattori 和 Sakaki 的方法（1986，分析生物化学 (Analytical Biochemistry) 152: 232-237），按照厂商说明书使用 Applied Biosystems 373A 型自动测序仪（Applied Biosystems, Inc., Foster City, CA）对该片段测序。

对 pMWR54 中克隆的二肽氨肽酶扩增插入片段的序列测定证实，与 SEQ ID NO: 1 所述的序列无差异。

实施例 9: 镰孢 CC1-3 的转化和转化体的分析

将镰孢 CC1-3 株是镰孢 A3/5 株的高度分支的形态突变体（ATCC 20334）（Wiebe 等人，1992，真菌学研究 (Mycological Research) 96: 555-562; Wiebe 等人，1991，真菌学研究 95: 1284-1288; Wiebe 等人，1991，真菌学研究 96: 555-562），在含 Vogel's 盐（Vogel, 1964，美国自然 (Am. Nature) 98: 435-446）、25mM NaNO₃ 和 1.5% 葡萄糖的液体培养基中于 28℃ 和 150 转/分下培养 4 天。通过 4 层粗棉布、最后通过 1 层微孔布过滤纯化分生孢子。离心浓缩分生孢子悬液。用约 10⁸ 个分生孢子接种 50ml 含 1% 酵母提取物、2% 细菌蛋白胨和 2% 葡萄糖的 YPG 培养基，在 24℃ 和 150 转/分下温育 14 小时。将产生的菌丝捕获于无菌 0.4mm 滤器上，用无菌蒸馏水和 1.0M MgSO₄ 连续冲洗。在 10ml NOVOZYM 234TM 溶液（2-10mg/ml 于 1.0M MgSO₄ 中）中重悬浮菌丝，于 80 转/分摇动下 34℃ 消化 15-30 分钟。通过 4 层粗棉布并通过微孔布连续过滤从产生的原生质体悬液中去掉未消化的菌丝物质。20ml 1M 山梨糖醇与原生质体溶液合并。混合后，离心沉淀原生质体，并通过重悬浮和离心在 20ml 1M 山梨糖醇和 20ml STC（0.8M 山梨糖醇、0.05M Tris pH8.0、0.05M CaCl₂）中连续洗涤。将洗涤的原生质体以 5×10⁷/ml 的浓度重悬浮于 4 份 STC 和 1 份 SPTC

中(0.8M 山梨糖醇、40% PEG 4000、0.05M Tris pH8.0、0.05M CaCl_2)。

将 100ml 原生质体悬液加入聚丙烯试管(17×100mm)中的 10 μg pMWR54 中,混合并在冰上孵育 30 分钟。将 1ml SPTC 轻轻混合入原生质体悬液中,于室温下继续温育 20 分钟。12.5ml 含 1×Vogel's 盐、25mM NaNO_3 、0.8M 蔗糖和 1%低熔点琼脂糖(Sigma Chemical Company, St. Louis, MO)的融化溶液(冷却至 40℃)与原生质体混合,然后涂板于空 100mm 培养板上。于室温下继续温育 10-14 天。在室温温育 24 小时后,将 12.5ml 相同培养基+每毫升 10mg BASTA™(Hoechst Schering, Rodovre, 丹麦)覆盖于培养板上。其中用酚:氯仿:异戊醇(25:24:1)抽提 BASTA™ 两次,使用前再用氯仿:异戊醇(24:1)抽提一次。

两周后,显现 21 个转化体。将每个转化体边缘的菌丝体片段转移至含 Vogel's /BASTA™ 培养基的 24 孔板的各个孔中。培养基每升包含 25g 蔗糖、25g Noble 琼脂、20ml 50×Vogel's 盐(Vogel, 1964, 同前)、25mM NaNO_3 和 10g BASTA™。用塑料袋密封平板以保持湿度,并在室温下温育大约 1 周。

实施例 10: 米曲霉 1568 二肽氨肽酶 I 在镰孢中的表达

将实施例 9 所述 21 个镰孢 CC1-3 转化体边缘的菌丝体片段接种于 20ml M400Da 培养基中,并在 30℃ 和 200 转/分下温育 5 天,该培养基每升包含 50g 麦芽糖糊精、2.0g $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ 、2.0g KH_2PO_4 、4.0g 柠檬酸、8.0g 酵母提取物、2.0g 尿素和 0.5ml 痕量金属溶液。用 5N NaOH 将培养基调节为 pH6.0。痕量金属溶液每升包含 14.3g $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ 、2.5g $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ 、0.5g $\text{NiCl}_2 \cdot 6\text{H}_2\text{O}$ 、13.8g $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ 、8.5g $\text{MnSO}_4 \cdot \text{H}_2\text{O}$ 和 3.0g 柠檬酸。未转化的宿主也用作对照。在第 5 天收获 1ml 每种培养上清液,并贮存于 4℃。测定二肽氨肽酶 I 活性,方法为,将 1 μl 酶上清液与 200 μl 每毫升 50mM 磷酸钠(pH7.5)含 2mg Ala-Pro-pNA 的 Ala-Pro-pNA 底物贮存液混合,监测 405nm 和环境温度下吸光度的改变。

当测定二肽氨肽酶 I 活性时, 21 个 pMWR54 原代转化体中有 19 个的培养上清液为阳性。

在 125ml 摇瓶中, 将镰孢原代转化体#1 和#16 在 25ml M400Da 培养基中 30℃ 培养 5 天。用双层微孔布过滤镰孢原代转化体的全培养液。回收滤过液, 然后贮存于-20℃。

实施例 11: 镰孢中产生的重组米曲霉 1568 二肽氨肽酶 I 的纯化

从实施例 10 所述原代转化体#1 的镰孢培养液中纯化重组二肽氨肽酶 I。将培养液 (20ml) 通过配备有 0.45 μ m 滤器的 Nalgene 过滤装置 (Nalgene, Rochester, NY) 过滤。用 20mM 磷酸钠(pH7.0)缓冲液将样品稀释 10 倍, 并用使用 PM10 超滤膜的超滤系统 (Amicon, Beverly, MA) 浓缩。样品的导电率为 2.5mS。

然后将样品加样于含有 60ml Q-Sepharose Big Beads 并用 400ml 20mM 磷酸钠(pH7.0)缓冲液预平衡的柱 (XK-26) 上。洗柱直到达到基线。以 5ml/分钟的流速, 用 20mM 磷酸钠(pH7.0)缓冲液中 0-0.40M NaCl 的线性梯度洗脱二肽氨肽酶 I, 共 10 个柱体积。在~0.24M NaCl 处洗脱下二肽氨肽酶 I。按照下列方案对可作用于 Ala-Pro-pNA 的级分进行 SDS-PAGE。通过在 20 μ l DMSO 中溶解 2mg Ala-Pro-pNA (Bachem, Torrance, CA) 制备底物。然后, 加入 980 μ l 50mM 磷酸钠(pH7.5)缓冲液。在 96 孔微量滴定板中, 向 100 μ l 米曲霉二肽氨肽酶 I (用 50mM 磷酸钠缓冲液(pH7.5)稀释 100 倍) 中加入 100 μ l 底物溶液, 并使用 SpectroMax 340 读板仪 (Molecular Devices, Sunnyvale, Ca) 对 405nm 吸光度的增加测定 4 分钟。然后合并同类的级分。

实施例 12: 重组米曲霉 1568 二肽氨肽酶 I 的表征

就最适 pH、温度稳定性、底物特异性和动力学参数, 表征了实施例 11 所述的纯化的二肽氨肽酶 I。

使用 Ala-Pro-pNA (HCl 盐) 作为底物在广域缓冲液中测定最适 pH, 广域缓冲液由 0.125M 柠檬酸、0.125M 单价磷酸钠和 0.125M 硼

酸组成，用 10N NaOH 将 pH 调节为 4.35-9.83，增量为 0.5pH。Ala-Pro-pNA 底物的制备方法是，将 100mg Ala-Pro-pNA 溶于 1ml DMSO 中，向在环境温度下不同 pH 值的 980 μ l 广域缓冲液中加入 20 μ l Ala-Pro-pNA-DMSO 溶液。通过向不同 pH 值的 200 μ l 2mg/ml Ala-Pro-pNA 中加入 10 μ l 等份用水稀释 20 倍的二肽氨肽酶 I 进行测定。监测 405nm 的吸光度 5 分钟。通过向不同 pH 值的 200 μ l 2mg/ml Ala-Pro-pNA 中加入 10 μ l 水测定底物的自水解作为对照。

下面表 1 显示的结果证明，在所测定的 pH 范围 4.35-9.83 中，二肽氨肽酶 I 对作为底物的 Ala-Pro-pNA 具有活性，最佳活性在 pH~8.7。在 pH 值大于 7 时观察到底物的自水解。

表 1

pH	平均活性	平均背景	平均活性-背景	相对活性
4.35	8.75 mOD/分钟	0 mOD/分钟	8.75 mOD/分钟	0.023
4.87	32.5	0	32.5	0.088
5.36	75.75	0	75.75	0.20
5.86	113.8	0	113.8	0.307
6.38	135.48	0	135.48	0.365
6.85	168.45	2	166.45	0.45
7.2	188.86	2	186.86	0.503
7.51	230.97	3	227.97	0.615
7.97	308.52	5.8	302.72	0.817
8.71	383.15	12.5	370.65	1
9.32	247.68	29.8	217.88	0.588
9.83	171	74	97	0.261

使用下列方案测定二肽氨肽酶 I 的温度稳定性：490 μ l 50mM 磷酸钠缓冲液(pH7.0)在 1.7ml Eppendorf 管中于 37 $^{\circ}$ C、45 $^{\circ}$ C、55 $^{\circ}$ C、60 $^{\circ}$ C、65 $^{\circ}$ C、70 $^{\circ}$ C 和 75 $^{\circ}$ C 预温育 30 分钟。然后加入 10 μ l 纯化的二肽氨肽酶 I，样品另外温育 20 分钟。然后将样品置于冰上。一旦所有温度的温育都完成，即用 Ala-Pro-pNA 作为底物测定活性。

进行测定的方法是，在环境温度下将 30 μ l 不同温度的温育混合物与 200 μ l 2mg/ml 溶于 50mM 磷酸钠(pH7.5)缓冲液中的 Ala-Pro-pNA 相混合。监测 405nm 的吸光度 5 分钟。

表 2 显示的结果证明，在 65 $^{\circ}$ C、pH7.5 温育 20 分钟后，二肽氨肽酶 I 保留 90% 的活性。

表 2

温度 (°C)	相对于 37°C 的百分活性
37	100
45	103
55	103
65	92.5
70	37.7
75	0.7

用 50mM 磷酸钠(pH7.5)缓冲液中稀释 100 倍的纯化二肽氨肽酶 I 测定不同单肽、二肽和三肽对硝基苯胺底物相对于 Ala-Pro-pNA 的相对活性。将每种底物溶解于 DMSO 中至 100mg/ml 的终浓度, 然后用 50mM 磷酸钠(pH7.5)缓冲液稀释 50 倍至 2mg/ml。进行测定的方法是, 将 100 μ l 底物溶液与 100 μ l 二肽氨肽酶 I 溶液混合, 监测 405nm 和环境温度下吸光度的改变。

表 3 显示的结果证明, 二肽氨肽酶 I 优选地水解 Xaa-Pro-pNA 和 Xaa-Ala-pNA 底物, 其中 Xaa 对应于任何天然氨基酸。

表 3

底物	活性 (mOD/分钟)	相对于 Ala-Pro-pNA 的百分活性
Gly-Arg-pNA	0	0
Gly-Pro-pNA	125	51.25
Arg-Pro-pNA	103	42.17
Val-Ala-pNA	49.5	20.3
Gly-Glu-pNA	<2	0
Ala-Pro-pNA	244	100
Ala-Ala-pNA	30.2	12.4
Asp-Pro-pNA	70.95	29
Leu-pNA	0	0
Ala-pNA	0	0
Ala-Ala-Pro-pNA	10.6	4.3

用下列方案测定不同二肽氨肽酶 I 底物的动力学参数。这些底物包括 Ala-Pro-pNA、Asp-Pro-pNA 和 Ala-Ala-pNA。对于 Ala-Pro-pNA 将 A_{280} 为 1.521 的纯化二肽氨肽酶 I 用水稀释 25 倍, 对于 Asp-Pro-pNA 稀释 20 倍, 对于 Ala-Ala-pNA 稀释 10 倍。将每种底物溶解于 DMSO

中至 100mg/ml 的终浓度，然后用 50mM 磷酸钠(pH7.5)缓冲液稀释 50 倍至 2mg/ml。在 96 孔微量滴定板中，除了不对 Asp-Pro-pNA 底物进行 200 μ l 底物测定外，将 10 μ l 纯化的二肽氨肽酶 I 与如下每种底物温育，并测定 405nm 的吸光度 3 分钟：

1. 200 μ l 2mg/ml 底物+0 μ l 50mM 磷酸钠缓冲液 pH7.5
2. 100 μ l 2mg/ml 底物+100 μ l 50mM 磷酸钠缓冲液 pH7.5
3. 50 μ l 2mg/ml 底物+150 μ l 50mM 磷酸钠缓冲液 pH7.5
4. 25 μ l 2mg/ml 底物+175 μ l 50mM 磷酸钠缓冲液 pH7.5
5. 10 μ l 2mg/ml 底物+190 μ l 50mM 磷酸钠缓冲液 pH7.5
6. 5 μ l 2mg/ml 底物+195 μ l 50mM 磷酸钠缓冲液 pH7.5

对于不同糖基化形式使用 97kDa 的平均分子量，建立 Lineweaver-Burke 试验以测定对于每种底物的 K_m 和 k_{cat} 。

对于 Ala-Pro-pNA，测定的 K_m 和 k_{cat} 分别为 0.140mM 和 576.3/分钟。

对于 Asp-Pro-pNA，测定的 K_m 和 k_{cat} 分别为 0.632mM 和 244.3/分钟。

对于 Ala-Ala-pNA，测定的 K_m 和 k_{cat} 分别为 1.08mM 和 106.5/分钟。

实施例 13: 米曲霉 1568 氨肽酶 I 的纯化

以 10000 转 / 分离心 50ml 体积的实施例 1 所述的 FLAVOURZYME™ 制剂 10 分钟。用 0.2 μ m Nalgene 滤器过滤上清液，然后用 20mM 磷酸钠(pH7.5)缓冲液将滤过液稀释为 350ml。用配备有 PM10 膜的 Amicon 超滤单元浓缩稀释的滤过液。加入更多的 20mM 磷酸盐(pH7.5)缓冲液，浓缩 3 倍，将样品调节为适当的 pH 和传导率。

然后将酶溶液加样于含有 Q-Sepharose Big Beads 并用 20mM 磷酸盐(pH7.5)缓冲液预平衡的 XK-26 柱上。施行 0-300mM NaCl 的梯度，然后用 350ml 300mM NaCl 冲洗。按照下列方案测定级分的 Leu-pNA 活性。通过在 20 μ l DMSO 中溶解 2mg Leu-pNA(Sigma Chemical

Co., St. Louis, MO) 制备底物。然后, 加入 980 μ l 50mM 磷酸钠(pH7.5) 缓冲液。在 96 孔微量滴定板中, 向 100 μ l 米曲霉氨肽酶 I (用 50mM 磷酸钠(pH7.5) 缓冲液稀释 100 倍) 中加入 100 μ l 底物溶液, 用 SpectroMax 340 读板仪对 405nm 吸光度的增加测定 4 分钟。合并最具活性的级分, 并如上用 PM10 超滤浓缩。

用 20mM 磷酸盐(pH7.5) 缓冲液将浓缩的样品稀释为 250ml, 加样于 Mono-Q 16/10 柱 (Pharmacia Biotech AB, Uppsala, 瑞典)。施行 0-400mM NaCl 的梯度, 如上测定级分的 Leu-pNA 活性。合并最具活性且纯化的级分 (用 SDS-PAGE), 并如上用 PM10 超滤浓缩。

在含 1.7M 硫酸铵的 50mM 磷酸盐(pH7.0) 缓冲液中稀释该样品, 并如上浓缩。重复此步骤三次。将样品加于用含 1.7M 硫酸铵的 50mM 磷酸钠(pH7.0) 缓冲液预平衡的苯基 Sepharose 柱 (Pharmacia Biotech AB, Uppsala, 瑞典)。施行 1.7M-0M 硫酸铵的梯度, 然后用 50mM 磷酸钠(pH7.0) 缓冲液冲洗。测定级分的 Leu-pNA 活性, 通过 SDS-PAGE 检查其纯度。仍存在靠近 70kDa 的两条小带。

用 Microcon 10 微量浓缩器 (Amicon, New Bedford, MA) 浓缩 1ml 体积的最纯的级分。将 100 μ l 体积的浓缩液加于用 20mM 磷酸盐 pH7.0 缓冲液预平衡的 Superose 12 柱 (Pharmacia Biotech AB, Uppsala, 瑞典), 并以 30ml 相同缓冲液洗脱。只发现一个峰。SDS-PAGE 分析显示靠近 70kDa 的两条小带。这些带或许是二肽氨肽酶的聚集物, 没有进一步纯化的必要。

使用 PM10 超滤和 20mM 磷酸盐(pH7.0) 缓冲液, 将苯基 Sepharose 柱的所有活性级分合并、浓缩并脱盐。

实施例 14: 米曲霉 1568 氨肽酶 I 与米曲霉 1568 二肽氨肽酶 I 之间的协同作用

通过在 20 μ l DMSO 中溶解 2mg Ala-Phe-Pro-pNA (Bachem, Torrance, CA) 制备底物。然后, 加入 980 μ l 50 mM 磷酸钠缓冲液(pH7.5) 和 100 μ l 乙醇。在 96 孔微量滴定板中, 向 100 μ l 100 倍稀释的氨肽酶

I (实施例 12) 中加入 100 μ l Ala-Phe-Pro-pNA 溶液; 向 100 μ l 100 倍稀释的二肽氨肽酶 I (实施例 11) 中加入 100 μ l Ala-Phe-Pro-pNA 溶液; 向 50 μ l 100 倍稀释的氨肽酶 I 与 50 μ l 100 倍稀释的二肽氨肽酶 I 的混合液中加入 100 μ l Ala-Phe-Pro-pNA 溶液。测定 405nm 的吸光度 10 分钟。

使用 405nm 处对于对硝基苯胺为 10000 的消光系数, 测定纯化的氨肽酶 I 对 Leu-pNA 具有 26.822 LAPU/ml 的活性。LAPU 被定义为如 AF 298/1-GB (请求丹麦的 Novo Nordisk A/S 后可使用) 所述测定的亮氨酸氨肽酶活性。同样使用对于对硝基苯胺为 10000 的消光系数, 测定纯化的二肽氨肽酶 I 对 Ala-Pro-pNA 具有 11.54 DPAPU/ml 的活性。100 倍稀释时, 氨肽酶 I 对 Ala-Pro-pNA 显示低于 2mOD/分钟的速度, 而二肽氨肽酶 I 对 Leu-pNA 显示低于 2mOD/分钟的速度。100 倍稀释的氨肽酶 I 对 Ala-Phe-Pro-pNA 经 10 分钟也不显示活性。100 倍稀释的二肽氨肽酶 I 对 Ala-Phe-Pro-pNA 经 10 分钟也不显示活性。当应用 Ala-Phe-Pro-pNA 测定 100 倍稀释的二肽氨肽酶 I 和 100 倍稀释的氨肽酶 I 的混合物时, 10 分钟以上观察到 122mOD/分钟的速度。

实施例 15: 应用米曲霉 1568 二肽氨肽酶 I 对蛋白质水解产物的制备

按照下列方法, 使用明胶、大豆、面筋和酪蛋白作为底物, 在水解程度测定中检测实施例 11 所述的纯化的二肽氨肽酶 I。

水解程度 (DH) 测定在 50 $^{\circ}$ C 下进行 18 小时作为 10ml 规模的小量水解, 其中使用 2% 浓度的明胶、大豆粉片、面筋和酪蛋白酸钠, 调节至 pH7, 必要时在水解过程中不进行 pH 调节。水解产物在水浴中 85 $^{\circ}$ C 灭活 3 分钟。所用的酶是 FLAVOURZYMETM 与 ALCALASETM 2.4L (Novo Nordisk A/S, Bagsvaerd, 丹麦) 以及二肽氨肽酶 I 与 FLAVOURZYMETM 和 ALCALASETM 2.4L。每毫升每 200mg 蛋白质中, 分别以 3mg、3mg、0.125mg 剂量使用 FLAVOURZYMETM、ALCALASETM 和二肽氨肽酶 I。

通过上清液与 OPA (邻苯二醛, Sigma Chemical Co., St. Louis,

MO)的反应测定如 Adler-Nissen(1986, 食物蛋白质的酶水解(Enzymic Hydrolysis of Food Proteins), Elsevier Applied Science Publishers)所定义的 DH。对于 OPA 试剂, 将 160mg OPA 溶解于 4ml 乙醇中, 并转移至 200ml 容量瓶中, 该容量瓶中含有 7.62g 十水合四硼酸二钠、200mg 十二烷基硫酸钠和 176mg 二硫苏糖醇的溶液, 用水填充容量瓶至 200ml。

将 25 μ l 体积的适当稀释的上清液与 200 μ l OPA 试剂在微量滴定板孔中混合, 并使之在 25 $^{\circ}$ C 反应恰好 2 分钟。用微量滴定板读板仪测定 340nm 的吸光度, 并在减去空白值(与 OPA 试剂反应的水)后与 95mM L-丝氨酸标准溶液的吸光度相比较。为了确定真实 DH, 用 Adler-Nissen 所建议的用于三硝基苯磺酸法的因数 (Adler-Nissen, 1979, 农业与食品化学(Agricultural and Food Chemistry) 17: 1256) 修正上清液中所测定的丝氨酸等同物, 三硝基苯磺酸法与所述 OPA 法有相同的反应。根据水解混合物中蛋白质的总量(不是根据可溶性蛋白质)计算水解程度。

结果显示, 对于所试的所有蛋白质, 二肽氨肽酶使水解程度比只用 FLAVOURZYMETM和 ALCALASETM的样品提高了 5%。

实施例 16: 脱酰胺作用引起提高的蛋白质溶解性及谷氨酸的释放

小麦面筋(WG)从 Cargill (JOB 5141) 获得, 脱酰胺的小麦面筋(DWG)获自 StaPro Consultancy B.V., Lemdijk 32, 9422 TH Smilde, NL。通过将 11g 面筋与 89g 水混合制备 8%蛋白质悬液。用 NaOH 调节 pH 至 6.5。向该悬液中加入如 WO 91/13554 所述可获得的谷氨酸/天冬氨酸特异蛋白酶(SP446), 或者如 WO 89/06270 所述可获得的赖氨酸/精氨酸特异蛋白酶(SP387)。对于 SP446 剂量为 0.01AU/g 蛋白质, 对于 SP387 为 0.006 AU/g 蛋白质。以 20 LAPU/g 蛋白质的剂量向某些水解产物中加入 FLAVOURZYMETM(一种非特异作用的蛋白酶制剂, 可从丹麦 Bagsvaerd 的 Novo Nordisk A/S 获得, 含有内肽酶和外肽酶活性, 通过米曲霉的发酵获得)。一个 LAPU(亮氨酸氨肽酶

单位)是在下列条件下每分钟分解 1 微摩尔 L-亮氨酸-对硝基苯胺的酶量: 26mM L-亮氨酸-对硝基苯胺于 0.1M Tris pH8.0 缓冲液中, 40℃ 10 分钟。水解后释放对硝基苯胺使溶液变为黄色, 于 405nm 监测。

不经另外的 pH 调节, 在 50℃ 下进行水解 18 小时。通过 85℃ 加热 15 分钟灭活酶。调节 pH 至 5, 离心水解产物。测定上清液中蛋白质和游离谷氨酸的含量。

使用 6.25 的 Kjeldahl 因数, 通过 Kjeldahl 分析测定蛋白质含量。

使用谷氨酸测定试剂盒按照厂商说明书 (Boehringer-Mannheim, Indianapolis, IN) 测定游离谷氨酸的含量。该方法适于在微量滴定板中使用。

当将小麦面筋 (WG) 与脱酰胺的小麦面筋 (DWG) 相比较时, 表 4 所示的结果证明, 脱酰胺作用提高了面筋对特异蛋白酶的敏感性, 使得更多的蛋白质成为可溶的。通过加入 FLAVOURZYME™ 与一种特异蛋白酶, 谷氨酸的释放由于脱酰胺作用而加倍。

表 4

水解产物	蛋白质溶解度 (%)		谷氨酸含量 mg/l	
	WG	DWG	WG	DWG
SP446	18	54	0	0
SP387	35	44	0	0
SP446+ FLAVOURZYME™	34	87	1000	2000

实施例 17: 酶脱酰胺作用与谷氨酸的释放

肽谷氨酰胺酶 II 的产生方法是, 将环状芽孢杆菌株 ATCC 21590 在含 200ml 培养基的摇瓶 (400ml) 中在 270 转/分混合下 30℃ 培养 20 小时, 该培养基中含有 1% 聚脲、0.5% 乳糖、0.025% $MgSO_4 \cdot 7H_2O$ 、0.005% $FeSO_4 \cdot 7H_2O$ 、0.025% KH_2PO_4 和 17% $Na_2PO_4 \cdot 12 H_2O$ (pH 调节为 7.2)。在 1 升摇瓶中 4000 转/分离心收获细胞。然后冷冻细胞。

在室温下从环状芽孢杆菌中纯化肽谷氨酰胺酶 II。融化冷冻的环

状芽孢杆菌细胞，并悬浮于裂解缓冲液（50mM Tris/HCl; 25%(w/v) 蔗糖; 1mM EDTA, pH8.0）中，直到获得均匀的悬液—每升裂解缓冲液 100g 湿细胞。在裂解缓冲液中溶解溶菌酶（10mg/ml）和 DNase I（Sigma DN-25, 10mg/ml）。然后每升细胞悬液中加入 100ml 溶菌酶溶液、10ml 1.0M MgCl₂ 和 1ml DNase I 溶液。使该酶作用 1 小时。

将悬液通过赛氏深层过滤板过滤，在 Sephadex G25 柱上（Pharmacia）将滤过液转移到 10mM KH₂PO₄/NaOH, pH8.0（缓冲液 A）中。将酶溶液加样于用缓冲液 A 平衡的 SOURCE Q 柱（Pharmacia）上，并用缓冲液 A 中的线性 NaCl 梯度（0→500mM）洗脱。如下所述分析该柱级分的肽谷氨酰胺酶 II 活性，合并有活性的级分。合并级分的 280nm 吸光度为 1.78，因此估计蛋白质含量为 1.8mg/ml。

从 SDS-PAGE 凝胶判断，肽谷氨酰胺酶 II 合并物中蛋白质的纯度大约为 25%。因此，该制剂中含有大约 0.5mg/ml 的纯肽谷氨酰胺酶 II。

使用 Boehringer-Mannheim 氨测定试剂盒（目录号 1112732）通过测定 N-叔丁氧羰基-Gln-Pro（N-t-BOC-Gln-Pro; SIGMA B-4403 号）的 γ -羧基酰胺水解过程中形成的氨来测定肽谷氨酰胺酶活性。在该试剂盒中，通过测定谷氨酸脱氢酶对 NADH 的消耗测定氨，也使用不加入 N-t-BOC-Gln-Pro 的空白，以减去其它 NADH 消耗酶的作用。

将总共 200mg 的小麦面筋蛋白质加入 9ml 沸水中，冷却后将 pH 调节为 7.0。然后加入 250 μ l 上述肽谷氨酰胺酶 II 制剂（PEP）。以 0.04 AU/g 蛋白质的量加入实施例 16 所述的谷氨酸/天冬氨酸特异蛋白酶（SP446），并以 20 LAPU/g 蛋白质的量加入实施例 16 所述的 FLAVOURZYME™。

不经 pH 调节地在 50℃ 进行水解 18 小时。也同样处理不加肽谷氨酰胺酶的对照。离心水解产物，如实施例 16 所述测定谷氨酸。如实施例 15 所述测定 DH。

以下表 5 所示的结果证明，用肽谷氨酰胺酶制剂水解提高了 DH 以及谷氨酸的释放。

表 5

水解	DH %	谷氨酸 mg/l
- PEP	40	131
+ PEP	43	171

生物材料的保藏

下列生物材料已按照布达佩斯条约的条款保藏于农业研究机构保藏中心，1815 University Street, Peoria, Illinois, 61604, 并予以下列保藏号:

保藏物	保藏号	保藏日期
大肠杆菌 DH5 α pMWR52	NRRL B-21682	1997年4月18日

序列表

(1) 一般信息:

(i) 申请人:

(A) 姓名: Novo Nordisk Biotech, Inc.

(B) 街道: 1445 drew Avenue

(C) 城市: Davis, California

(D) 国家: 美国

(E) 邮政编码: 95616-4880

(F) 电话: (530) 757-8100

(G) 传真: (530) 758-0317

(ii) 发明名称: 具有二肽氨肽酶活性的多肽及编码该多肽的核酸

(iii) 序列数目: 13

(iv) 联系地址:

(A) 收件人: Novo Nordisk of North America, Inc.

(B) 街道: 405 Lexington Avenue

(C) 城市: 纽约

(D) 州: NY

(E) 国家: 美国

(F) 邮政编码: 10174

(v) 计算机可读形式:

(A) 介质类型: 磁盘

(B) 计算机: IBM 兼容机

(C) 操作系统: DOS

(D) 软件: FastSEQ for Windows Version 2.0

(vi) 当前申请信息:

(A) 申请号: 待分配

(B) 递交日期: 1998年5月15日

(C) 分类:

(viii) 律师/代理机构信息:

(A)名称:Valeta A. Gregg

(B)登记号:41,324

(C)文献/案卷号: 5254.204-WO

(ix)通讯信息:

(A)电话:212-867-0123

(B)传真:212-878-9655

(2) SEQ. ID. NO.1 的信息:

(i)序列特征;

(A)长度: 2313 个碱基对

(B)类型: 核酸

(C)链型: 单链

(D)拓扑结构: 线性

(ix)特征;

(A)名称/关键词: 编码序列

(B)位置: 1...2313

(D)其他信息:

(xi) SEQ. ID. NO.1 的序列描述;

ATG AAG TAC TCC AAG CTT CTG CTG CTC CTG GTC AGT GTG GTC CAG GCC	48
Met Lys Tyr Ser Lys Leu Leu Leu Leu Leu Val Ser Val Val Gln Ala	
1 5 10 15	
CTG GAT GTG CCT CGG AAA CCA CAC GCG CCC ACC GGA GAA GGC AGT AAG	96
Leu Asp Val Pro Arg Lys Pro His Ala Pro Thr Gly Glu Gly Ser Lys	
20 25 30	
CGT CTC ACC TTC AAT GAG ACC GTA GTC AAG CAA GCA ATT ACG CCG ACC	144
Arg Leu Thr Phe Asn Glu Thr Val Val Lys Gln Ala Ile Thr Pro Thr	
35 40 45	
TCT CGC TCG GTG CAA TGG CTC TCG GGC GCA GAG GAT GGA TCC TAC GTG	192
Ser Arg Ser Val Gln Trp Leu Ser Gly Ala Glu Asp Gly Ser Tyr Val	
50 55 60	
TAC GCG GCG GAA GAC GGC AGT CTC ACC ATC GAG AAC ATC GTC ACC AAC	240
Tyr Ala Ala Glu Asp Gly Ser Leu Thr Ile Glu Asn Ile Val Thr Asn	
65 70 75 80	

GAG Glu	TCA Ser	CGC Arg	ACG Thr	CTC Leu	ATC Ile	CCT Pro	GCG Ala	GAC Asp	AAG Lys	ATT Ile	CCG Pro	ACA Thr	GGG Gly	AAG Lys	GAA Glu	288
				85					90					95		
GCG Ala	TTC Phe	AAT Asn	TAC Tyr	TGG Trp	ATC Ile	CAT His	CCC Pro	GAC Asp	TTG Leu	TCG Ser	TCG Ser	GTG Val	CTG Leu	TGG Trp	GCG Ala	336
			100					105					110			
TCC Ser	AAC Asn	CAC His	ACC Thr	AAG Lys	CAG Gln	TAT Tyr	CGG Arg	CAT His	TCG Ser	TTC Phe	TTT Phe	GCC Ala	GAT Asp	TAT Tyr	TAC Tyr	384
		115					120					125				
GTC Val	CAG Gln	GAT Asp	GTG Val	GAG Glu	TCA Ser	CTC Leu	AAG Lys	TCC Ser	GTG Val	CCC Pro	CTG Leu	ATG Met	CCC Pro	GAT Asp	CAG Gln	432
		130				135					140					
GAA Glu	GGT Gly	GAT Asp	ATT Ile	CAA Gln	TAT Tyr	GCC Ala	CAA Gln	TGG Trp	AGC Ser	CCC Pro	GTG Val	GGC Gly	AAT Asn	ACC Thr	ATC Ile	480
145					150					155					160	
GCT Ala	TTT Phe	GTT Val	CGC Arg	GAG Glu	AAT Asn	GAC Asp	CTT Leu	TAT Tyr	GTC Val	TGG Trp	GAT Asp	AAT Asn	GGT Gly	ACC Thr	GTT Val	528
				165					170					175		
ACT Thr	CGC Arg	ATT Ile	ACT Thr	GAT Asp	GAT Asp	GGT Gly	GGC Gly	CCC Pro	GAC Asp	ATG Met	TTC Phe	CAC His	GGC Gly	GTG Val	CCG Pro	576
			180					185					190			
GAC Asp	TGG Trp	ATC Ile	TAT Tyr	GAA Glu	GAG Glu	GAG Glu	ATC Ile	CTC Leu	GGC Gly	GAT Asp	CGC Arg	TAC Tyr	GCG Ala	TTG Leu	TGG Trp	624
			195				200					205				
TTC Phe	TCG Ser	CCA Pro	GAT Asp	GGT Gly	GAA Glu	TAT Tyr	CTG Leu	GCT Ala	TAC Tyr	TTG Leu	AGC Ser	TTC Phe	AAT Asn	GAG Glu	ACT Thr	672
		210				215					220					
GGG Gly	GTT Val	CCG Pro	ACC Thr	TAC Tyr	ACC Thr	GTT Val	CAG Gln	TAT Tyr	TAT Tyr	ATG Met	GAT Asp	AAC Asn	CAA Gln	GAG Glu	ATC Ile	720
225					230					235					240	
GCT Ala	CCG Pro	GCG Ala	TAT Tyr	CCA Pro	TGG Trp	GAG Glu	CTG Leu	AAG Lys	ATA Ile	AGG Arg	TAT Tyr	CCC Pro	AAG Lys	GTG Val	TCG Ser	768
				245					250					255		
CAG Gln	ACG Thr	AAT Asn	CCG Pro	ACC Thr	GTG Val	ACG Thr	TTG Leu	AGT Ser	CTG Leu	CTT Leu	AAC Asn	ATC Ile	GCT Ala	AGC Ser	AAG Lys	816
			260					265					270			
GAG Glu	GTG Val	AAG Lys	CAG Gln	GCG Ala	CCG Pro	ATC Ile	GAC Asp	GCG Ala	TTC Phe	GAG Glu	TCA Ser	ACT Thr	GAC Asp	TTG Leu	ATC Ile	864
		275					280					285				
ATT Ile	GGC Gly	GAG Glu	GTT Val	GCT Ala	TGG Trp	CTC Leu	ACT Thr	GAT Asp	ACT Thr	CAC His	ACC Thr	ACC Thr	GTT Val	GCT Ala	GCT Ala	912
		290				295					300					
AAG Lys	GCG Ala	TTC Phe	AAC Asn	CGT Arg	GTC Val	CAG Gln	GAC Asp	CAG Gln	CAA Gln	AAG Lys	GTC Val	GTC Val	GCG Ala	GTC Val	GAT Asp	960
305					310					315					320	
ACT Thr	GCC Ala	TCG Ser	AAC Asn	AAG Lys	GCT Ala	ACT Thr	GTC Val	ATC Ile	AGC Ser	GAC Asp	CGA Arg	GAT Asp	GGG Gly	ACC Thr	GAT Asp	1008
				325					330					335		

GGA TGG CTC GAT AAC CTT CTT TCA ATG AAG TAT ATT GGC CCT ATC AAG	1056
Gly Trp Leu Asp Asn Leu Leu Ser Met Lys Tyr Ile Gly Pro Ile Lys	
340 345 350	
CCG TCC GAC AAG GAT GCC TAC TAC ATC GAC ATC TCT GAC CAT TCG GGA	1104
Pro Ser Asp Lys Asp Ala Tyr Tyr Ile Asp Ile Ser Asp His Ser Gly	
355 360 365	
TGG GCG CAT CTG TAT CTC TTC CCC GTT TCG GGC GGC GAA CCT ATC CCA	1152
Trp Ala His Leu Tyr Leu Phe Pro Val Ser Gly Gly Glu Pro Ile Pro	
370 375 380	
CTA ACC AAA GGC GAC TGG GAG GTC ACG TCT ATT CTG AGT ATT GAT CAG	1200
Leu Thr Lys Gly Asp Trp Glu Val Thr Ser Ile Leu Ser Ile Asp Gln	
385 390 395 400	
GAA CGC CAG TTG GTG TAC TAC CTG TCG ACT CAA CAC CAC AGC ACC GAG	1248
Glu Arg Gln Leu Val Tyr Tyr Leu Ser Thr Gln His His Ser Thr Glu	
405 410 415	
CGC CAT CTC TAC TCC GTC TCC TAT TCC ACG TTT GCG GTC ACC CCG CTC	1296
Arg His Leu Tyr Ser Val Ser Tyr Ser Thr Phe Ala Val Thr Pro Leu	
420 425 430	
GTC GAC GAC ACC GTT GCC GCG TAC TGG TCT GCT TCC TTC TCC GCG AAC	1344
Val Asp Asp Thr Val Ala Ala Tyr Trp Ser Ala Ser Phe Ser Ala Asn	
435 440 445	
TCG GGC TAC TAC ATC CTC ACA TAC GGA GGC CCA GAC GTA CCC TAC CAG	1392
Ser Gly Tyr Tyr Ile Leu Thr Tyr Gly Gly Pro Asp Val Pro Tyr Gln	
450 455 460	
GAA CTC TAC ACG ACC AAC AGT ACC AAA CCA CTC CGC ACA ATC ACC GAC	1440
Glu Leu Tyr Thr Thr Asn Ser Thr Lys Pro Leu Arg Thr Ile Thr Asp	
465 470 475 480	
AAC GCC AAA GTA CTC GAG CAA ATC AAG GAC TAT GCA TTG CCC AAC ATC	1488
Asn Ala Lys Val Leu Glu Gln Ile Lys Asp Tyr Ala Leu Pro Asn Ile	
485 490 495	
ACC TAC TTC GAG CTT CCC CTC CCC TCC GGA GAA ACC CTC AAT GTG ATG	1536
Thr Tyr Phe Glu Leu Pro Leu Pro Ser Gly Glu Thr Leu Asn Val Met	
500 505 510	
CAG CGC TTA CCC CCC GGG TTC TCC CCG GAT AAG AAG TAC CCC ATA CTT	1584
Gln Arg Leu Pro Pro Gly Phe Ser Pro Asp Lys Lys Tyr Pro Ile Leu	
515 520 525	
TTC ACC CCA TAC GGC GGC CCA GGC GCC CAA GAA GTG ACC AAG AGA TGG	1632
Phe Thr Pro Tyr Gly Gly Pro Gly Ala Gln Glu Val Thr Lys Arg Trp	
530 535 540	
CAA GCC CTG AAT TTC AAG GCC TAT GTC GCC TCC GAC AGC GAA CTC GAG	1680
Gln Ala Leu Asn Phe Lys Ala Tyr Val Ala Ser Asp Ser Glu Leu Glu	
545 550 555 560	
TAC GTA ACC TGG ACT GTC GAC AAC CGC GGC ACA GGT TTC AAA GGA CGC	1728
Tyr Val Thr Trp Thr Val Asp Asn Arg Gly Thr Gly Phe Lys Gly Arg	
565 570 575	
AAG TTC CGC TCC GCC GTC ACG CGC CAA CTC GGC CTC CTC GAA GCA GAA	1776
Lys Phe Arg Ser Ala Val Thr Arg Gln Leu Gly Leu Leu Glu Ala Glu	
580 585 590	

GAC CAG ATC TAC GCC GCG CAA CAG GCG GCC AAC ATC CCC TGG ATC GAT	1824
Asp Gln Ile Tyr Ala Ala Gln Gln Ala Ala Asn Ile Pro Trp Ile Asp	
595 600 605	
GCA GAC CAC ATC GGC ATC TGG GGC TGG AGT TTC GGA GGC TAC TTG ACC	1872
Ala Asp His Ile Gly Ile Trp Gly Trp Ser Phe Gly Gly Tyr Leu Thr	
610 615 620	
AGC AAG GTC CTG GAG AAG GAC AGC GGT GCT TTC ACA TTA GGA GTC ATC	1920
Ser Lys Val Leu Glu Lys Asp Ser Gly Ala Phe Thr Leu Gly Val Ile	
625 630 635 640	
ACC GCC CCT GTT TCT GAC TGG CGT TTC TAC GAC TCA ATG TAC ACG GAG	1968
Thr Ala Pro Val Ser Asp Trp Arg Phe Tyr Asp Ser Met Tyr Thr Glu	
645 650 655	
CGC TAC ATG AAG ACC CTC TCG ACC AAT GAG GAG GGC TAC GAG ACC AGC	2016
Arg Tyr Met Lys Thr Leu Ser Thr Asn Glu Glu Gly Tyr Glu Thr Ser	
660 665 670	
GCC GTC CGC AAG ACT GAC GGG TTC AAG AAC GTC GAG GGC GGA TTC TTG	2064
Ala Val Arg Lys Thr Asp Gly Phe Lys Asn Val Glu Gly Gly Phe Leu	
675 680 685	
ATC CAG CAC GGA ACG GGC GAC GAT AAC GTC CAT TTC CAG AAC TCG GCT	2112
Ile Gln His Gly Thr Gly Asp Asn Val His Phe Gln Asn Ser Ala	
690 695 700	
GCG CTG GTG GAT CTC CTG ATG GGC GAT GGC GTC TCT CCT GAG AAG CTC	2160
Ala Leu Val Asp Leu Leu Met Gly Asp Gly Val Ser Pro Glu Lys Leu	
705 710 715 720	
CAT TCG CAA TGG TTC ACA GAC TCA GAC CAC GGA ATC AGC TAC CAT GGT	2208
His Ser Gln Trp Phe Thr Asp Ser Asp His Gly Ile Ser Tyr His Gly	
725 730 735	
GGC GGC GTG TTC CTG TAC AAG CAA CTG GCC CGG AAG CTC TAC CAG GAG	2256
Gly Gly Val Phe Leu Tyr Lys Gln Leu Ala Arg Lys Leu Tyr Gln Glu	
740 745 750	
AAG AAC CGA CAG ACG CAG GTG CTG ATG CAC CAG TGG ACT AAG AAG GAC	2304
Lys Asn Arg Gln Thr Gln Val Leu Met His Gln Trp Thr Lys Lys Asp	
755 760 765	
TTG GAG GAG	2313
Leu Glu Glu	
770	

(2) SEQ. ID. NO.2 的信息:

(i) 序列特征;

(A) 长度: 771 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(ii) 分子类型: 蛋白质

(v)片段类型; 内部

(xi) SEQ. ID. NO.2 的序列描述;

```

Met Lys Tyr Ser Lys Leu Leu Leu Leu Leu Val Ser Val Val Gln Ala
1      5      10
Leu Asp Val Pro Arg Lys Pro His Ala Pro Thr Gly Glu Gly Ser Lys
20
Arg Leu Thr Phe Asn Glu Thr Val Val Lys Gln Ala Ile Thr Pro Thr
35      40      45
Ser Arg Ser Val Gln Trp Leu Ser Gly Ala Glu Asp Gly Ser Tyr Val
50      55      60
Tyr Ala Ala Glu Asp Gly Ser Leu Thr Ile Glu Asn Ile Val Thr Asn
65      70      75      80
Glu Ser Arg Thr Leu Ile Pro Ala Asp Lys Ile Pro Thr Gly Lys Glu
85      90      95
Ala Phe Asn Tyr Trp Ile His Pro Asp Leu Ser Ser Val Leu Trp Ala
100      105      110
Ser Asn His Thr Lys Gln Tyr Arg His Ser Phe Phe Ala Asp Tyr Tyr
115      120      125
Val Gln Asp Val Glu Ser Leu Lys Ser Val Pro Leu Met Pro Asp Gln
130      135      140
Glu Gly Asp Ile Gln Tyr Ala Gln Trp Ser Pro Val Gly Asn Thr Ile
145      150      155      160
Ala Phe Val Arg Glu Asn Asp Leu Tyr Val Trp Asp Asn Gly Thr Val
165      170      175
Thr Arg Ile Thr Asp Asp Gly Gly Pro Asp Met Phe His Gly Val Pro
180      185      190
Asp Trp Ile Tyr Glu Glu Glu Ile Leu Gly Asp Arg Tyr Ala Leu Trp
195      200      205
Phe Ser Pro Asp Gly Glu Tyr Leu Ala Tyr Leu Ser Phe Asn Glu Thr
210      215      220
Gly Val Pro Thr Tyr Thr Val Gln Tyr Tyr Met Asp Asn Gln Glu Ile
225      230      235      240
Ala Pro Ala Tyr Pro Trp Glu Leu Lys Ile Arg Tyr Pro Lys Val Ser
245      250      255
Gln Thr Asn Pro Thr Val Thr Leu Ser Leu Leu Asn Ile Ala Ser Lys
260      265      270
Glu Val Lys Gln Ala Pro Ile Asp Ala Phe Glu Ser Thr Asp Leu Ile
275      280      285
Ile Gly Glu Val Ala Trp Leu Thr Asp Thr His Thr Thr Val Ala Ala
290      295      300
Lys Ala Phe Asn Arg Val Gln Asp Gln Gln Lys Val Val Ala Val Asp
305      310      315      320
Thr Ala Ser Asn Lys Ala Thr Val Ile Ser Asp Arg Asp Gly Thr Asp
325      330      335
Gly Trp Leu Asp Asn Leu Leu Ser Met Lys Tyr Ile Gly Pro Ile Lys
340      345      350
Pro Ser Asp Lys Asp Ala Tyr Tyr Ile Asp Ile Ser Asp His Ser Gly
355      360      365
Trp Ala His Leu Tyr Leu Phe Pro Val Ser Gly Gly Glu Pro Ile Pro
370      375      380
Leu Thr Lys Gly Asp Trp Glu Val Thr Ser Ile Leu Ser Ile Asp Gln
385      390      395      400
Glu Arg Gln Leu Val Tyr Tyr Leu Ser Thr Gln His His Ser Thr Glu
405      410      415
Arg His Leu Tyr Ser Val Ser Tyr Ser Thr Phe Ala Val Thr Pro Leu
420      425      430
Val Asp Asp Thr Val Ala Ala Tyr Trp Ser Ala Ser Phe Ser Ala Asn
435      440      445
Ser Gly Tyr Tyr Ile Leu Thr Tyr Gly Gly Pro Asp Val Pro Tyr Gln
450      455      460
Glu Leu Tyr Thr Thr Asn Ser Thr Lys Pro Leu Arg Thr Ile Thr Asp
465      470      475      480

```

```

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

```

(2) SEQ. ID. NO.3 的信息:

(i)序列特征;

(A)长度: 20 个氨基酸

(B)类型: 氨基酸

(C)链型: 单链

(D)拓扑结构: 线性

(ii)分子类型; 非

(xi) SEQ. ID. NO.3 的序列描述;

```

Xaa Glu Gly Ser Lys Arg Leu Thr Phe Xaa Glu Thr Val Val Lys Gln
 1                               5           10           15
Ala Ile Thr Pro
      20

```

(2) SEQ. ID. NO.4 的信息:**(i)序列特征;****(A)长度: 22 个氨基酸****(B)类型: 氨基酸****(C)链型: 单链****(D)拓扑结构: 线性****(ii)分子类型; 非****(xi) SEQ. ID. NO.4 的序列描述;**

Gln	Arg	Leu	Pro	Pro	Gly	Phe	Ser	Pro	Asp	Lys	Lys	Tyr	Pro	Ile	Leu
1				5					10					15	
Phe	Thr	Pro	Tyr	Gly	Gly										
			20												

(2) SEQ. ID. NO.5 的信息:**(i)序列特征;****(A)长度: 7 个氨基酸****(B)类型: 氨基酸****(C)链型: 单链****(D)拓扑结构: 线性****(ii)分子类型; 非****(xi) SEQ. ID. NO.5 的序列描述;**

Lys	Tyr	Ile	Gly	Pro	Ile	Lys
1				5		

(2) SEQ. ID. NO.6 的信息:**(i)序列特征;****(A)长度: 7 个氨基酸****(B)类型: 氨基酸****(C)链型: 单链****(D)拓扑结构: 线性**

(ii)分子类型; 非

(xi) SEQ. ID. NO.6 的序列描述;

Gly Glu Gly Ser Lys Arg Leu
1 5

(2) SEQ. ID. NO.7 的信息:

(i)序列特征;

(A)长度: 8 个氨基酸

(B)类型: 氨基酸

(C)链型: 单链

(D)拓扑结构: 线性

(ii)分子类型; 非

(xi) SEQ. ID. NO.7 的序列描述;

Xaa Pro Ile Leu Phe Thr Pro Tyr
1 5

(2) SEQ. ID. NO.8 的信息:

(i)序列特征;

(A)长度: 16 个氨基酸

(B)类型: 氨基酸

(C)链型: 单链

(D)拓扑结构: 线性

(ii)分子类型; 非

(xi) SEQ. ID. NO.8 的序列描述;

Xaa Val Pro Leu Met Pro Asp Gln Gln Gly Asp Ile Gln Tyr Ala Gln
1 5 10 15

(2) SEQ. ID. NO.9 的信息:**(i)序列特征;****(A)长度: 21 个氨基酸****(B)类型: 氨基酸****(C)链型: 单链****(D)拓扑结构: 线性****(ii)分子类型; 非****(xi) SEQ. ID. NO.9 的序列描述;**

```

Gly Ala Tyr Thr Gly Gly Ile Thr Ile Thr Ala Tyr Gly Ala Arg Gly
 1           5           10           15
Ala Arg Gly Ala Arg
                20

```

(2) SEQ. ID. NO.10 的信息:**(i)序列特征;****(A)长度: 32 个氨基酸****(B)类型: 氨基酸****(C)链型: 单链****(D)拓扑结构: 线性****(ii)分子类型; 非****(xi) SEQ. ID. NO.10 的序列描述;**

```

Gly Gly Arg Thr Ala Tyr Thr Thr Tyr Thr Thr Arg Thr Cys Ile Gly
 1           5           10           15
Gly Ile Ser Trp Arg Ala Ala Ile Cys Cys Ile Gly Gly Ile Gly Gly
                20           25           30

```

(2) SEQ. ID. NO.11 的信息:**(i)序列特征;****(A)长度: 931 个氨基酸**

(B)类型: 氨基酸

(C)链型: 单链

(D)拓扑结构: 线性

(xi) SEQ. ID. NO.11 的序列描述;

```

Met Ser Ala Ser Thr His Ser His Lys Arg Lys Asn Ser His Leu Phe
 1          5          10          15
Pro Gln Arg Lys Ser Ser Asn Ser Ser Met Asp Lys Pro Phe Phe Pro
 20          25          30
Asn Asn Asp Ser Val Ala Asn Thr Asp Pro Gln Ser Asn Glu Asn Gly
 35          40          45
His Thr Ile Asn Glu Ile Arg Pro Thr Glu Ala Thr Ile Asp Val Thr
 50          55          60
Asp Val Pro Gln Thr Pro Phe Leu Gln Glu Gln Tyr Ser Met Arg Pro
 65          70          75          80
Arg Arg Glu Ser Phe Gln Phe Asn Asp Ile Glu Asn Gln His His Thr
 85          90          95
His Ser Phe Phe Ser Val Asn Lys Phe Asn Arg Arg Trp Gly Glu Trp
 100         105         110
Ser Leu Pro Glu Lys Arg Ser Tyr Val Leu Val Phe Thr Leu Ile Ala
 115         120         125
Leu Ser Val Leu Val Leu Leu Val Ile Leu Ile Pro Ser Lys Leu Leu
 130         135         140
Pro Thr Lys Ile Thr Arg Pro Lys Thr Ser Ala Gly Asp Ser Ser Leu
 145         150         155         160
Gly Lys Arg Ser Phe Ser Ile Glu Asn Val Leu Asn Gly Asp Phe Ala
 165         170         175
Ile Pro Glu Asp Thr Phe His Phe Ile Asp Pro Pro Gln Arg Leu Leu
 180         185         190
Gly Gln Asp Ser Asp Pro Gly Leu Tyr Phe Thr Thr Lys Glu Ile Asp
 195         200         205
Gly His Thr Asn Phe Ile Ala Lys Gln Leu Phe Asp Glu Thr Phe Glu
 210         215         220
Val Asn Leu Gly Gly Asn Arg Phe Leu Tyr Glu Gly Val Glu Phe Thr
 225         230         235         240
Val Ser Thr Val Gln Ile Asn Tyr Lys Leu Asp Lys Leu Ile Phe Gly
 245         250         255
Thr Asn Leu Glu Ser Glu Phe Arg His Ser Ser Lys Gly Phe Tyr Trp
 260         265         270
Ile Lys Asp Leu Asn Thr Gly Asn Ile Glu Pro Ile Leu Pro Pro Glu
 275         280         285
Lys Ser Asp Asp Asn Tyr Glu Leu Gly Leu Ser Lys Leu Ser Tyr Ala
 290         295         300

```

His Phe Ser Pro Ala Tyr Asn Tyr Ile Tyr Phe Val Tyr Glu Asn Asn
 305 310 315 320
 Leu Phe Leu Gln Gln Val Asn Ser Gly Val Ala Lys Lys Val Thr Glu
 325 330 335
 Asp Gly Ser Lys Asp Ile Phe Asn Ala Lys Pro Asp Trp Ile Tyr Glu
 340 345 350
 Glu Glu Val Leu Ala Ser Asp Gln Ala Ile Trp Trp Ala Pro Asp Asp
 355 360 365
 Ser Lys Ala Val Phe Ala Arg Phe Asn Asp Thr Ser Val Asp Asp Ile
 370 375 380
 Arg Leu Asn Arg Tyr Thr Asn Met Asn Glu Ala Tyr Leu Ser Asp Thr
 385 390 395 400
 Lys Ile Lys Tyr Pro Lys Pro Gly Phe Gln Asn Pro Gln Phe Asp Leu
 405 410 415
 Phe Leu Val Asn Leu Gln Asn Gly Ile Ile Tyr Ser Ile Asn Thr Gly
 420 425 430
 Gly Gln Lys Asp Ser Ile Leu Tyr Asn Gly Lys Trp Ile Ser Pro Asp
 435 440 445
 Thr Phe Arg Phe Glu Ile Thr Asp Arg Asn Ser Lys Ile Leu Asp Val
 450 455 460
 Lys Val Tyr Asp Ile Pro Ser Ser Gln Met Leu Thr Val Arg Asn Thr
 465 470 475 480
 Asn Ser Asn Leu Phe Asn Gly Trp Ile Glu Lys Thr Lys Asp Ile Leu
 485 490 495
 Ser Ile Pro Pro Lys Pro Glu Leu Lys Arg Met Asp Tyr Gly Tyr Ile
 500 505 510
 Asp Ile His Ala Asp Ser Arg Gly Phe Ser His Leu Phe Tyr Tyr Pro
 515 520 525
 Thr Val Phe Ala Lys Glu Pro Ile Gln Leu Thr Lys Gly Asn Trp Glu
 530 535 540
 Val Thr Gly Asn Gly Ile Val Gly Tyr Glu Tyr Glu Thr Asp Thr Ile
 545 550 555 560
 Phe Phe Thr Ala Asn Glu Ile Gly Val Met Ser Gln His Leu Tyr Ser
 565 570 575
 Ile Ser Leu Thr Asp Ser Thr Thr Gln Asn Thr Phe Gln Ser Leu Gln
 580 585 590
 Asn Pro Ser Asp Lys Tyr Asp Phe Tyr Asp Phe Glu Leu Ser Ser Ser
 595 600 605
 Ala Arg Tyr Ala Ile Ser Lys Lys Leu Gly Pro Asp Thr Pro Ile Lys
 610 615 620
 Val Ala Gly Pro Leu Thr Arg Val Leu Asn Val Ala Glu Ile His Asp
 625 630 635 640
 Asp Ser Ile Leu Gln Leu Thr Lys Asp Glu Lys Phe Lys Glu Lys Ile
 645 650 655
 Lys Asn Tyr Asp Leu Pro Ile Thr Ser Tyr Lys Thr Met Val Leu Asp
 660 665 670
 Asp Gly Val Glu Ile Asn Tyr Ile Glu Ile Lys Pro Ala Asn Leu Asn
 675 680 685
 Pro Lys Lys Lys Tyr Pro Ile Leu Val Asn Ile Tyr Gly Gly Pro Gly
 690 695 700
 Ser Gln Thr Phe Thr Thr Lys Ser Ser Leu Ala Phe Glu Gln Ala Val
 705 710 715 720
 Val Ser Gly Leu Asp Val Ile Val Leu Gln Ile Glu Pro Arg Gly Thr
 725 730 735
 Gly Gly Lys Gly Trp Ser Phe Arg Ser Trp Ala Arg Glu Lys Leu Gly
 740 745 750
 Tyr Trp Glu Pro Arg Asp Ile Thr Glu Val Thr Lys Lys Phe Ile Gln
 755 760 765
 Arg Asn Ser Gln His Ile Asp Glu Ser Lys Ile Ala Ile Trp Gly Trp
 770 775 780

```

Ser Tyr Gly Gly Phe Thr Ser Leu Lys Thr Val Glu Leu Asp Asn Gly
785                               790                               795                               800
Asp Thr Phe Lys Tyr Ala Met Ala Val Ala Pro Val Thr Asn Trp Thr
                               805                               810                               815
Leu Tyr Asp Ser Val Tyr Thr Glu Arg Tyr Met Asn Gln Pro Ser Glu
                               820                               825                               830
Asn His Glu Gly Tyr Phe Glu Val Ser Thr Ile Gln Asn Phe Lys Ser
                               835                               840                               845
Phe Glu Ser Leu Lys Arg Leu Phe Ile Val His Gly Thr Phe Asp Asp
                               850                               855                               860
Asn Val His Ile Gln Asn Thr Phe Arg Leu Val Asp Gln Leu Asn Leu
865                               870                               875                               880
Leu Gly Leu Thr Asn Tyr Asp Met His Ile Phe Pro Asp Ser Asp His
                               885                               890                               895
Ser Ile Arg Tyr His Asn Ala Gln Arg Ile Val Phe Gln Lys Leu Tyr
                               900                               905                               910
Tyr Trp Leu Arg Asp Ala Phe Ala Glu Arg Phe Asp Asn Thr Glu Val
                               915                               920                               925
Leu His Leu
930

```

(2) SEQ. ID. NO.12 的信息:

(i)序列特征;

(A)长度: 36 个碱基对

(B)类型: 核酸

(C)链型: 单链

(D)拓扑结构: 线性

(xi) SEQ. ID. NO.12 的序列描述;

GATTTAAATC ACCATGAAGG TACGTCAATT CCACTG

36

(2) SEQ. ID. NO.13 的信息:

(i)序列特征;

(A)长度: 36 个碱基对

(B)类型: 核酸

(C)链型: 单链

(D)拓扑结构: 线性

(ii)分子类型; cDNA

(xi) SEQ. ID. NO.13 的序列描述;

GTTAATTAATCTACTCCTCCAAGTCCTTCTTAGTCC

90 ATGAGGTCGCTTTTGTGGCTTCGTTGCTTTCGGGGGTGTGGTGGGAGGGCGCTTGTTCGCCGGATGAGTTCGCCGAGGATATTCAG
 M R S L L W A S L L S G V L A G R A L V S P D E F P E D I Q
 180 TTGGAAGATCTGCTGGAGGATCCCAACAGCTTGAGGACTTCGCCTATGCCTACCCCGAGCGCAATCGCGTCTTTGGTGGTAAAGCCAC
 L E D L L E G S Q Q L E D F A Y A Y P E R N R V F G G K A H
 270 GACGACACGGTTAACTATCTACGAGGAGCTGAAGAAGACTGGCTACTATGATGTCTACAAGCAGCCTCAGGTGCAGGTGGAGCAAT
 D D T V N Y L Y E E L K K T G Y Y D V Y K Q P Q V H L W S N
 360 GCCGACCAGCGCTCAGGTGGCGATGAGGAAATCGAGGCGAAGACCATGACCTACAGTCCCAGCGTCGAGGTACCCGCCGATGTAGCC
 A D D T L K V G D E E I E A K T M T Y S P S V E V T A D V A
 450 GTCGTCAAGAACCTGGGATGCAGCGAGCGGATTACCCATCCGATGTGGAGGGCAAGCTCGCCCTGATCAAGCGTGGAGAAATGCCCGTTC
 V V K N L G C S E A D Y P S D V E G K V A L I K R G E C P F
 540 GGGCACAGTCCGGTTCGCTGCCAAGCCAGGGCCGGCTTCGATTGTCTATAACAATGTGGCCGGATCCATGGCCGGCACCCCTTGGC
 G D K S V L A A K A K A A S I V Y N N V A G S M A G T L G
 630 GCGGGCAGAGTATAGGGACCGTATTCGGCCATTGTCGGTATCAGCTTGGAGGATGGCCAGAACTGATCAAGCTTCTGAGGCTGGA
 A A Q S D K G P Y S A I V G I S L E D G Q K L I K L A E A G
 720 TCGGTATCTGTGGATCTGTGGTGGATAGTAAGCAGGAGAACCGTACGACGTATAACCGTTGTCGGCCAGACGAAAGGCGCGATCCGAAC
 S V S V D L W V D S K Q E N R T T Y N V V A Q T K G G D P N
 810 AACGTCGTCGGCTGGGTGCCACACGGACTCAGTCGAGGGGGCCCTGGTATCAACGACGATGGCTCGGGCATTTAGCAACTTGGTC
 N V V A L G G H T D S V E A G P G I N D D G S G I I S N L V

图 1A

ATTGCCAAAGCGCTCAGGCAGTACTCCGTC AAGAA TGCCGGTGGCCTTCCCTCTTCTGGACAGCAGAGGAGTTCGGTCTGCTGGGCAGCAAC 900
 I A K A L T Q Y S V K N A V R F L F W T A E E F G I L G S N
 TACTAGTCTCCCATCTGAATGCCACCAGCTGAACAAGATCCGACTGTACCTGAAC TTCGACATGATCGCCTCACCTAACTACGCCCTC 990
 Y Y V S H L N A T E L N K I R L Y L N F D M I A S P N Y A L
 ATGATCTATGACGGTGTGGATCGGGGTTCAACCAGAGCGGACCGGGCGGTTCCGGCCAGATCGAGAAAC TGTTCGAGGACTACTACGAC 1080
 M I Y D G D G S A F N Q S G P A G S A Q I E K L F E D Y Y D
 TCCATCGACCTGCCCTCATATCCCACCCAGTTTGACGGACGTTCCGACTACGAGGCC TTTATCCTGAAAGGCA TTCCGTCCGGTGGACTC 1170
 S I D L P H I P T Q F D G R S D Y E A F I L N G I P S G G L
 TTCACGGGGCCGAGGGCATCATGTCCGAAGAGAACGCAAGCCGCTGGGAGGTCAAGCCGGGTGGCTACGAGCCCAACTACCCACGCC 1260
 F T G A E G I M S E E N A S R W G G Q A G V A Y D A N Y H A
 GCGGGAGACAACATGACCAACCTCAACCATGAAGCCTTCCCTGATCAACTCCAAGCCACCGCCTTCGCCGTGGCCACTACGCCCAACGAC 1350
 A G D N M T N L N H E A F L I N S K A T A F A V A T Y A N D
 CTCTCTCGATCCCCAANGGAATACCACATCCTCTTGCACCGAGCCCGCACCATGGACCA TTTCGGCAAGAGAGCTCCGMAAGACA 1440
 L S S I P K R N T T S S L H R R A R T M R P F G K R A P K T
 CACGCTCACGTATCAGGATCCGGATGCTGGCATTTCTCAAGTCGAGGCATAG 1491
 H A H V S G S G C W H S Q V E A

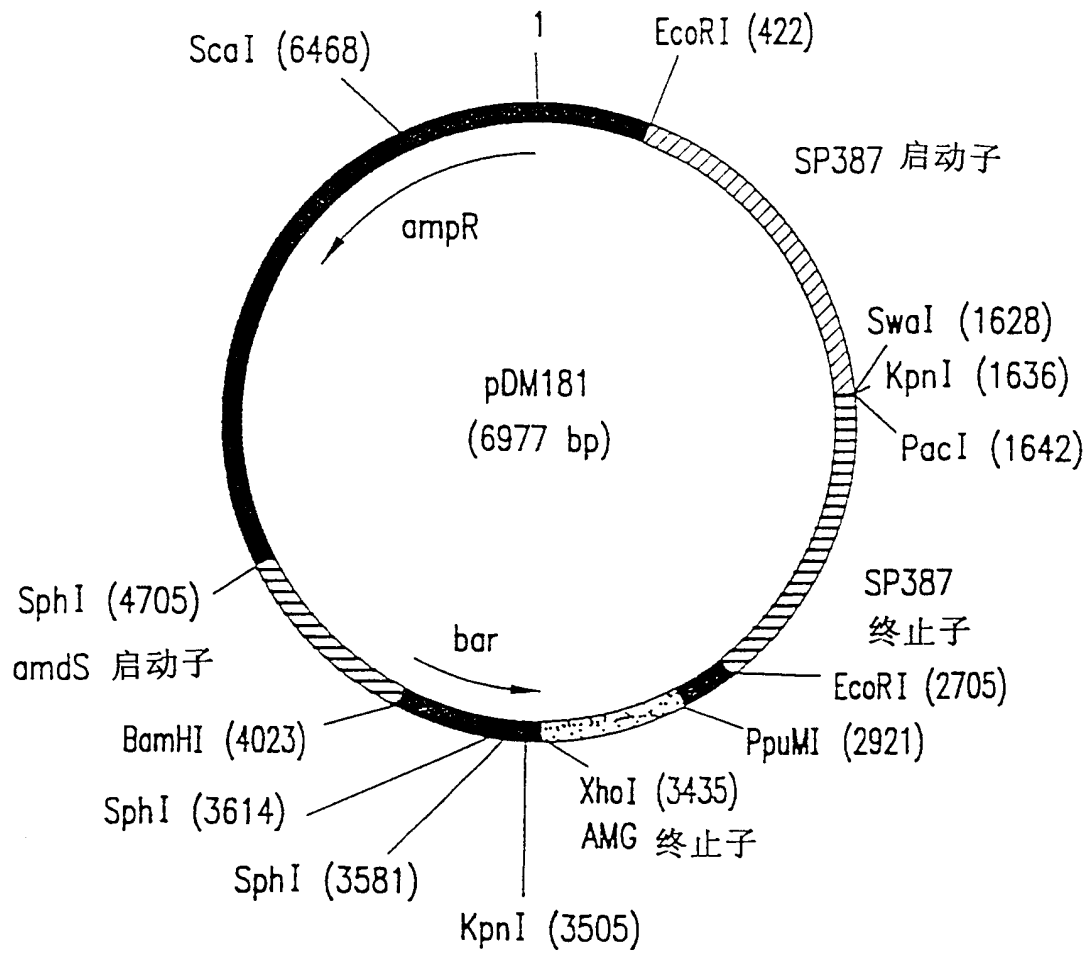


图 2

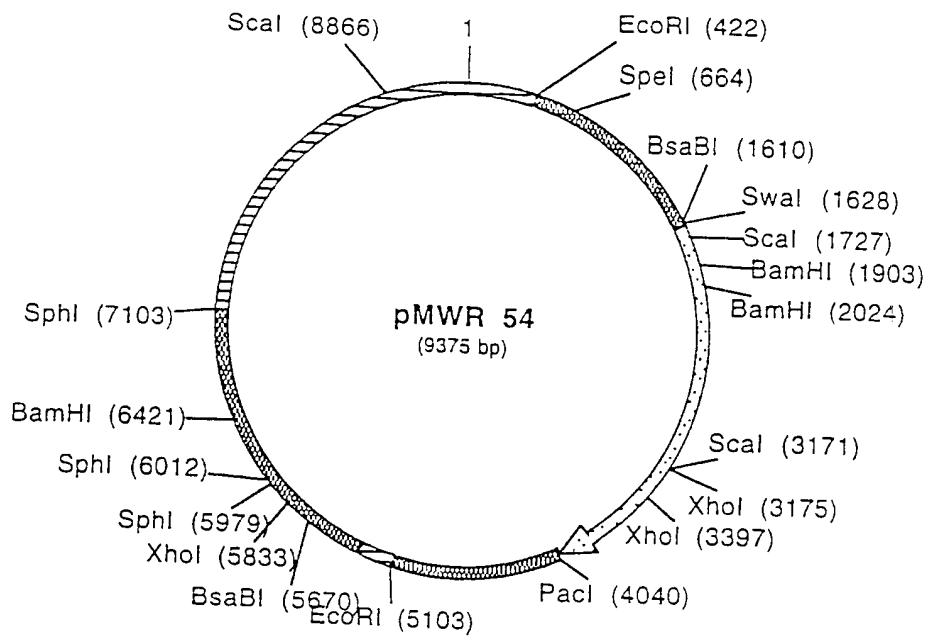


图 3