

[19] 中华人民共和国国家知识产权局



[12] 发明专利说明书

专利号 ZL 97191622.5

[51] Int. Cl.

C12N 15/57 (2006.01)

C12N 15/70 (2006.01)

C12N 9/48 (2006.01)

C12N 1/21 (2006.01)

A23J 3/34 (2006.01)

C12R 1/19 (2006.01)

[45] 授权公告日 2008 年 3 月 26 日

[11] 授权公告号 CN 100376678C

[22] 申请日 1997.10.3 [21] 申请号 97191622.5

[30] 优先权

[32] 1996.10.4 [33] US [31] 08/726,880

[32] 1996.11.27 [33] US [31] 08/757,534

[86] 国际申请 PCT/US1997/017977 1997.10.3

[87] 国际公布 WO1998/014599 英 1998.4.9

[85] 进入国家阶段日期 1998.7.8

[73] 专利权人 诺沃奇梅兹有限公司

地址 美国加利福尼亚州

共同专利权人 诺沃奇梅兹有限公司

[72] 发明人 A·布林科夫斯基 R·伯卡
M·雷 E·格莱特利 A·克罗茨
T·E·马希森 C·达姆伯曼
K·布朗

[56] 参考文献

WO9609397A1 1996.3.28

US3645850A 1972.2.29

审查员 丁海

[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 李瑛

权利要求书 3 页 说明书 59 页 附图 6 页

[54] 发明名称

来自米曲霉的羧肽酶和编码该酶的核酸

[57] 摘要

本发明涉及具有羧肽酶活性的多肽和编码多肽的分离的核酸序列。本发明也涉及包含所说的核酸序列的核酸构建体、载体和宿主细胞以及产生所说的多肽的方法。本发明还涉及获得作为香味改良剂有用的蛋白质水解产物的方法。

1.一种具有羧肽酶活性的分离的多肽，其中所述多肽选自下组：

(a) 具有羧肽酶活性的多肽，其具有通过对SEQ ID NO:2的氨基酸进行保守取代而获得的氨基酸序列；

(b) 由核酸序列编码的具有羧肽酶活性的多肽，所述核酸序列是通过对SEQ ID NO:1进行一个或多个核苷酸的添加、删除、插入或取代而获得的；和

(c) 由在中等严格条件下，与(i)SEQ ID NO:1的核酸序列，或(ii)它的互补链杂交的核酸序列编码的具有羧肽酶活性的多肽，其中中等严格条件被定义为在42℃下，在5×SSPE，0.3% SDS，200mg/ml剪切并变性的鲑精DNA和35% 甲酰胺中预杂交和杂交。

2. 权利要求1的多肽，其具有下列特征：(i) 在25℃，pH4.0-pH5.0范围内具有最佳pH活性；(ii) 在pH4，55℃-60℃的范围内具有最佳温度活性；(iii)在pH4.0和60℃下于50mM乙酸盐缓冲液中温育酶10分钟后，具有至少65%的剩余活性；和(iv)具有水解N-CBZ-Ala-X中的X的能力，其中N-CBZ是N-苄酯基，X选自Ile、Glu、Lys、Arg、Asp、Asn、Phe和Tyr。

3. 权利要求1的多肽，该多肽具有通过对SEQ ID NO: 2的氨基酸进行保守取代而获得的氨基酸序列。

4. 权利要求1的多肽，该多肽包含SEQ ID NO: 2的氨基酸序列。

5. 权利要求1的多肽，该多肽由SEQ ID NO: 2的氨基酸序列组成。

6. 权利要求1的多肽，其由核酸序列编码，所述核酸序列是通过对SEQ ID NO:1进行一个或多个核苷酸的添加、删除、插入或取代而获得的。

7. 权利要求1的多肽，该多肽由在中等严格条件下，与(i)SEQ ID NO: 1的核酸序列，或(ii)它的互补链杂交的核酸序列编码，其中中等严格条件被定义为在42℃，在5×SSPE，0.3% SDS，200mg/ml剪切并变性的鲑精DNA和35% 甲酰胺中预杂交和杂交。

8. 权利要求7的多肽，该多肽由在高度严格条件下，与(i)SEQ ID NO: 1的核酸序列，或(ii)它的互补链杂交的核酸序列编码，其中高度严格条件被定义为在42℃，在5×SSPE，0.3% SDS，200mg/ml剪切并变性的鲑精DNA和50% 甲酰胺中预杂交和杂交。

9. 权利要求1的多肽，该多肽得自曲霉属的菌株。

10. 权利要求9的多肽，该多肽得自米曲霉的菌株。

11. 权利要求1的多肽，该多肽由大肠杆菌NRRL B-21616中包含的质粒pEJG12中包含的核酸序列编码。

12. 一种编码权利要求1多肽的分离的核酸序列。

13. 一种核酸构建体，其包含可操作连接到一种或多种调控序列上的权利要求12的核酸序列，所述调控序列指导多肽在适合的表达宿主中的表达。

14. 一种重组表达载体，其包含权利要求13的核酸构建体、启动子以及转录和翻译终止信号。

15. 一种包含权利要求13的核酸构建体的重组宿主细胞。

16. 一种用于制备权利要求1的多肽的方法，该方法包括(a)培养产生权利要求1所述多肽的野生型菌株，以产生所述多肽；和(b)回收该多肽。

17. 一种用于制备权利要求1的多肽的方法，该方法包括(a)在利于产生权利要求1所述多肽的条件下培养包含核酸构建体的宿主细胞，所说的核酸构建体包含编码所述多肽的核酸序列；和(b)回收所述多肽。

18. 一种用于产生亲本细胞的突变体细胞的方法，该方法包括在能够产生权利要求1所述多肽的野生型菌株ATCC 20386中，破坏或缺失编码权利要求1的多肽的核酸序列或其调控序列，其导致突变体细胞产生的多肽比所说亲本细胞产生的少。

19. 由权利要求18的方法产生的突变体细胞。

20. 一种用于制备异源多肽的方法，该方法包括(a)在利于产生权利要求1所述多肽的条件下培养权利要求18的突变体细胞；和(b)回收

该多肽。

21. 一种由蛋白质底物制备蛋白水解产物的方法，该方法包括使所述底物受到权利要求1的所述多肽和内肽酶作用。

22. 一种从蛋白质底物获得富含游离谷氨酸和/或结合肽的谷氨酸残基的蛋白水解产物的方法，该方法包括使底物经历脱酰胺过程及使底物受到权利要求1的所述多肽作用。

23. 权利要求22的方法，该方法还包括使底物经受一种或多种具有非特异性作用的内肽酶和/或外肽酶作用。

来自米曲霉的羧肽酶和编码该酶的核酸

发明背景

发明领域

本发明涉及具有羧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所说多肽的方法。本发明还涉及获得作为香味改良剂有用的蛋白质水解产物的方法。

相关技术的描述

各种食品(例如,汤、调味料和调味品)包含由蛋白质物质水解获得的调味剂。这种水解通常利用强盐酸,接着用氢氧化钠中和完成。然而,这类化学水解导致水解期间所获得的氨基酸的严重降解,也导致在这一化学反应过程中形成危险的副产物。对利用由化学水解获得的调味剂的日益关注已导致酶促水解方法的发展。

酶促水解方法旨在获得高水解度(DH),这通常利用非特异性作用的蛋白水解酶复合物(例如非特异性作用的内和外肽酶)达到。例如,WO 94/25580描述了通过利用得自米曲霉的非特异性作用的酶制剂水解蛋白质的方法。特异性作用的蛋白水解酶还没有用于这一目的,因为这类酶仅导致不完全水解。

酸性羧肽酶(EC 3.4.16)是丝氨酸外肽酶,该酶催化从肽、寡肽或蛋白质的C末端除去氨基酸。这些羧肽酶通常具有窄的底物特异性,即,它们仅可切割较少的氨基酸。

米曲霉的酸性羧肽酶以前已报道了。例如,Nakadai, Nasuno和Iguchi, 1972, 农业和生物化学36: 1343-1352, 公开了分子量为120kDa(凝胶过滤)并在pH为3.0至4.0范围有最大活性的羧肽酶I。Nakadai, Nasuno和Iguchi, 1972, 农业和生物化学36: 1473-1480, 公开了分子量为105kDa(凝胶过滤)并在pH为3.0有最大活性的羧肽酶II。Nakadai, Nasuno和Iguchi, 1972,

农业和生物化学36: 1481-1488, 公开了分子量为61kDa(凝胶过滤)和3.0的pH为最佳值的羧肽酶III。Nakadai, Nasuno和Iguchi, 1972, 农业和生物化学37: 1237-1251公开了分子量为43kDa(凝胶过滤)并在pH为3.0有最大活性的羧肽酶IV。Tekeuchi和Ichishima, 1986, 农业和生物化学50: 633-638公开了分子量为73kDa (SDS-PAGE)的羧肽酶O。Tekeuchi, Ushijima和Ichishima, 1982, 现代微生物学7: 19-23公开了分子量都为63kDa(凝胶过滤)并在pH为3.7至4.0范围有最大活性的羧肽酶O-1和羧肽酶O-2。Ichishima等, 1972, 生物化学杂志72: 1045-1048公开了几种曲霉属酸性羧肽酶的酶促性质的比较。Azarenkova等, 1976, Biokhimiya 41: 20-27公开了来自米曲霉的分子量为37kDa(SDS-PAGE)并具有4至5的pH最佳值的一种酸性羧肽酶的分离物。

具有理想的感官特性和高水解度的蛋白质水解产物的产生一般需要混合利用肽酶活性。理想的是提供一种单一组分的肽酶, 该酶具有利于改进单独或与其它酶结合用于食品的蛋白质水解产物的感官特性和水解度的活性。

本发明的目的是提供具有羧肽酶活性的改良的多肽以及获得具有理想的感官特性和高水解度的蛋白质水解产物的方法。

发 明 概 述

本发明涉及具有羧肽酶活性的分离的多肽, 该多肽选自下组:

(a) 包含与SEQ ID NO:2的氨基酸序列有至少50%等同性的氨基酸序列的多肽;

(b) 由核酸序列编码的多肽, 该核酸序列在中等严格条件下, 与(i)SEQ ID NO:1的核酸序列, (ii)它的互补链, (iii)其子序列杂交;

(c) 一种多肽, 其(i)在25℃下, 约pH3.0-约pH7.5的范围有最大活性; (ii)在pH 4, 约55℃-约60℃的范围有最大活性; (iii)在pH4.0和60℃30分钟后, 有至少约65.5%的剩余活性; (iv)具有水解N-CBZ-Ala-X的X的能力, 其中N-CBZ是N-苄酯基, X是任何氨基酸;

(d) (a)或(b)的等位基因变体;

(e) (a)、(b)或(d)的片段, 其中该片段保持羧肽酶活性。

本发明也涉及编码所说多肽的分离的核酸序列和包含核酸序列的核酸

构建体、载体和宿主细胞以及产生多肽的方法。

本发明也涉及从蛋白质底物获得水解产物的方法，该方法包括使蛋白质物质受到只具有羧肽酶活性或既具有羧肽酶活性又具有内肽酶活性的多肽的作用，本发明也涉及由此法获得的水解产物。

本发明也涉及从蛋白质底物获得富含游离谷氨酸和/或结合肽的谷氨酸残基的水解产物的方法，该方法包括使底物经历脱酰胺过程和使底物受到具有羧肽酶活性的多肽的作用。

本发明还涉及包含具有羧肽酶活性的多肽的香味改良组合物。该组合物还可包含附加的酶促活性。

最后一个方面，本发明的方法可用于与食品相关的应用以改良香味，如烘烤。另外，食品的香味改良可以通过添加用本发明的方法所获得的水解产物达到。

附图的简要描述

图1显示米曲霉 ATCC 20386 羧肽酶 I 活性对pH的依赖性。

图2显示米曲霉 ATCC 20386 羧肽酶 I 活性对温度的依赖性。

图3显示米曲霉 ATCC 20386羧肽酶 I 的核酸序列和推定的氨基酸序列 (SEQ ID NOS:1和2)。

图4显示米曲霉 ATCC 20386羧肽酶 I 的序列和其它已知的羧肽酶序列的比较。

发明详述

具有羧肽酶活性的多肽

术语“羧肽酶活性”本文定义为肽酶活性，该肽酶活性催化除去肽、寡肽或蛋白质的C-末端氨基酸。以通常的方式定义的羧肽酶活性能够从肽、多肽或蛋白质的C-末端切割氨基酸X，其中X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组成的组的任何氨基酸残基。可以理解的是：具有本发明的羧肽酶活性的分离的多肽对切割肽、多肽或蛋白质的氨基酸序列是非特异性的。

在第一实施方案中，本发明涉及具有某种氨基酸序列的分离的多肽，所说的氨基酸序列与SEQ ID NO:2的氨基酸序列具有至少约50%的等同程

度, 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%, 此多肽保持羧肽酶活性(下文的“同源多肽”)。在一个优选的实施方案中, 同源多肽具有的氨基酸序列有五个氨基酸, 优选地四个氨基酸, 更优选地三个氨基酸, 更优选地两个氨基酸, 最优选地一个氨基酸不同于SEQ ID NO:2的氨基酸序列。对于本发明的目的, 两种氨基酸序列的等同程度通过Clustal方法(Higgins,1989,CABIOS 5:151-153)以等同性表(一个缺口罚10, 一个缺口长度罚10)测定。

优选地, 本发明的多肽包含SEQ ID NO:2的氨基酸序列或等位基因变体及其片段, 其中片段保持羧肽酶活性。在更为优选的实施方案中, 本发明的多肽包含SEQ ID NO:2的氨基酸序列。在另一个更为优选的实施方案中, 本发明的多肽具有SEQ ID NO:2的氨基酸序列或其片段, 其中片段保持羧肽酶活性。在最优选的实施方案中, 多肽具有SEQ ID NO: 2的氨基酸序列。

优选地, 所说片段包含至少300个氨基酸残基, 更优选地至少400个氨基酸残基, 最优选地至少500个氨基酸残基。

等位基因变体表明占据相同染色体基因座的任意两种或多种其它基因形式。等位基因变异通过突变天然产生, 并可导致群体内表型的多态性。基因突变可以是沉默的(编码的多肽没有变化)或可编码具有改变的氨基酸序列的多肽。术语等位基因变体也用来表明由基因的等位基因变体编码的蛋白质。

同源多肽的氨基酸序列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由不同的氨基酸残基取代而不同于SEQ ID NO:2的氨基酸序列。优选地, 氨基酸变化是较小的, 即这是不会明显地影响蛋白质的折叠和/或活性的保守氨基酸的取代、小的缺失, 典型地为1个至约30个氨基酸的缺失、小的氨基或羧基末端延长, 如氨基末端甲硫氨酸残基、达约20-25残基的小接头肽或通过改变净电荷或其它功能利于纯化的小延长(如多组氨酸段、抗原表位或结合区)。

保守取代的例子在碱性氨基酸(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)、疏水氨基

酸(如亮氨酸、异亮氨酸和缬氨酸)、芳香族氨基酸(如苯丙氨酸、色氨酸和酪氨酸)以及小氨基酸(如甘氨酸、丙氨酸、丝氨酸、苏氨酸和甲硫氨酸)的组中。氨基酸取代一般不改变特异性活性在本领域已知并且已描述,例如,由H.Neurath和R.L.Hill,1979,在蛋白质,纽约学院出版社中。存在的最普通的交换是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 以及它们的逆转。

在第二个实施方案中,本发明涉及由核酸序列编码的具有羧肽酶活性的分离的多肽或该多肽的等位基因变体和片段,其中所说的核酸序列在低严格条件,更优选地在中等严格条件,最优选地,在高严格条件下与在相同条件下和SEQ ID NO: 1的核酸序列或它的互补链杂交的寡核苷酸探针杂交(J.Sambrook, E.F.Fritsch和T.Manias,1989, 分子克隆, 实验室手册, 第二版, 冷泉港, 纽约)。

杂交表明: 核酸序列和对应于SEQ ID NO: 1显示的核酸序列的多肽编码部分的寡核苷酸探针在由低到高的严格条件下杂交(即, 42℃下在5X SSPE, 0.3% SDS, 200mg/ml剪切和变性的鲑精DNA和25、35或50%甲酰胺中分别在低、中和高严格下预杂交和杂交), 是按照标准Southern印迹方法进行的。

SEQ ID NO:2的氨基酸序列或其部分氨基酸序列可用于设计寡核苷酸探针, 或者编码本发明多肽的核酸序列, 如SEQ ID NO:1的核酸序列或其子序列可用于从不同属或种的菌株按照本领域众所周知的方法鉴定和克隆编码具有羧肽酶活性的多肽的DNA。尤其是, 这类探针可用于与兴趣属或种的基因组或cDNA杂交, 按照标准Southern印迹方法进行, 以鉴定和分离其中对应的基因。这类探针可以比整个序列短得多, 但长度应该至少为15, 优选地至少25, 更优选地至少40个核苷酸。更长的探针也可以利用。DNA和RNA探针都可以利用。典型地, 探针是为检测对应的基因标记的(例如, 用³²P、³H、³⁵S、生物素或抗生物素蛋白)。

因此, 为获得可与上述的探针杂交且编码有羧肽酶活性的多肽的DNA, 可筛选从这类其它有机体制备的基因组、cDNA或组合的化学文库。这类其它有机体的基因组或其它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。文库的DNA或分离的DNA可转移到并固定在硝酸纤维素或其它合适的载体物质上。为鉴定与SEQ ID NO:1同源的克隆或DNA,载体物质用于Southern印迹,其中载体物质每次用 $2 \times \text{SSC}$ 、0.2%SDS在至少50℃,更优选地至少55℃,更优选地至少60℃,更优选地至少65℃,更优选地至少70℃,最优选地至少75℃的温度最后洗涤三次,达30分钟。在这些条件下,与寡核苷酸探针杂交的分子利用X-射线胶片检测。

在第三个实施方案中,本发明涉及一些分离的多肽:它们(i)在25℃下约pH3.0至约pH7.5范围有最大活性;(ii)在pH 4约55℃至约60℃范围有最大活性;(iii)在pH 4.0和60℃下30分钟后,至少有约65.5%的残余活性;(iv)具有水解N-CBZ-Ala-X的能力,其中X是任何氨基酸。优选地,多肽在大约4.0至大约6.0的pH范围和25℃下,最优选地在大约4.0至大约5.0的pH范围下具有最佳活性。多肽也优选具有通过SDS-PAGE的约66kDa-约70kDa的分子量范围。此外,优选地多肽具有至少约60%,最优选地至少65%的剩余活性,最优选地,在pH4.0和60℃下10分钟后,剩余活性在50-85%的范围。优选地,本发明的多肽也有至少约70%,最优选地至少75%的剩余活性,最优选地,在pH4.0和55℃下30分钟后,剩余活性在65-90%的范围。优选地,多肽能够水解N-CBZ-Ala-X,其中X是Ile、Glu、Lys、Arg、Asp、Asn、Phe或Tyr。

在第四个实施方案中,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等同性或部分免疫化学等同性的分离的多肽。免疫化学性质由通过众所周知的Ouchterlony双向免疫扩散方法的免疫学交叉反应等同性试验测定。具体地说,抗血清按照由Harboe和Ingild,在N.H. Axelsen, J. Kroll和B. Weeks编辑的定量免疫电泳手册, Blackwell科学出版社, 1973, 第23章或Johnstone和Thorpe, 实用免疫化学, Blackwell科学出版社, 1982(更具体地在27-31页)中描述的方法通过免疫兔(或其它啮齿类)制备,其中抗血清含有免疫活性的或结合到具有SEQ ID NO: 2的氨基酸序列的多肽表位上的抗体。具有免疫化学等同性的多肽是一种以相同的方式(如利用特异性免疫化学技术,沉淀的全部融合、相同的沉淀的形态学和/或相同的电泳迁移率)与抗血清反应的多肽。免疫化学等同性的进一步解释由Axelsen、Bock和Kroll,在N.H.Axelsen, J. Kroll和B.Weeks编辑的定量

免疫电泳手册, Blackwell科学出版社, 1973, 第10章中描述。有部分免疫化学等同性的多肽是以部分相同的方式(如利用特异性免疫化学技术, 沉淀的部分融合、部分相同的沉淀的形态学和/或部分相同的电泳迁移率)与抗血清反应的多肽。部分的免疫化学等同性的进一步解释由Bock和Axelsen在N.H.Axelsen, J. Kroll和B. Weeks编辑的定量免疫电泳手册, Blackwell科学出版社, 1973, 第11章中描述。

由与寡核苷酸探针杂交的核酸序列编码的多肽可从任何属的微生物获得, 其中寡核苷酸探针可与SEQ ID NO:1的核酸序列或它的互补链或多肽的等位基因变体和片段、同源多肽和具有相同或部分相同免疫性质的多肽杂交。

在一个更优选的实施方案中, 这些多肽从细菌源获得。例如, 这些多肽可从革兰氏阳性细菌(如芽孢杆菌菌株(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、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、枯草芽孢杆菌以及苏云金芽孢杆菌)或链霉菌属(例如, 变铅青链霉菌或鼠灰链霉菌)菌株)或从革兰氏阴性细菌(如大肠杆菌或假单胞菌某种)获得。

在另一个优选的实施方案中, 这些多肽从真菌源获得。例如, 多肽可获自酵母菌株, 例如, 假丝酵母属、克鲁维酵母属、毕赤酵母、酵母属、裂殖酵母属或洋蓍草属菌株。在一个优选的实施方案中, 多肽获自卡尔酵母、酿酒酵母、糖化酵母、*Saccharomyces douglasii*、克鲁弗酵母、诺地酵母或*Saccharomyces oviformis*菌株。多肽也可获自丝状真菌菌株。例如, 多肽可获自顶孢霉属、曲霉属、短梗霉属、隐球酵母属、*Filibasidium*、镰刀菌属、腐质霉属、*Magnaporthe*、毛霉属、毁丝霉属、*Neocallimastix*、脉孢菌属、拟青霉属、青霉属、*Piromyces*、裂褶菌属、*Talaromyces*、嗜热子囊菌属、梭孢壳属、*Tolypocladium*或木霉属菌株。在最优选的实施方案中, 多肽获自棘孢曲霉、泡盛曲霉、*Aspergillus foetidus*、*Aspergillus japonicus*、构巢曲霉、黑曲霉、米曲霉、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、檲叶槭镰孢、尖镰孢、多枝镰孢、粉红镰孢、接骨木镰孢、肤色镰孢、*Fusarium sulphureum*、*Fusarium toruloseum*、*Fusarium trichothecioides*、*Fusarium*

venenatum、*Humicola insolens*、*Humicola lanuginosa*、米黑毛霉、*Myceliophthora thermophila*、粗糙脉孢菌、产紫青霉、*Thielavia terrestris*、*Trichoderma harzianum*、康宁木霉、*Trichoderma longibrachiatum*、*Trichoderma reesei*或绿色木霉菌株。

此外，这类多肽可利用上述的探针从包括天然分离的微生物的其它源(例如，土壤、堆肥、水等)中鉴定和获得。用于把微生物从天然生境分离的技术在本领域是众所周知的。然后，核酸序列可以通过类似地筛选另一个微生物的基因组或cDNA 文库获得。一旦把编码多肽的核酸序列用探针检测出来，可通过利用对本领域普通技术人员已知的技术分离或克隆序列(参见，例如，Sambrook等，1989，同上)。

优选地，本发明的多肽得自曲霉属物种，包括但不限于棘孢曲霉、泡盛曲霉、*Aspergillus foeridus*、*Aspergillus japonicus*、构巢曲霉、黑曲霉或米曲霉。公众可从许多培养物保藏中心(如美国典型培养物保藏中心(ATCC)、Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH(DSM)、Centraalbureau Voor Schimmelcultures(CBS)和农业研究服务机构专利培养物保藏中心、北方区域研究中心(NRRL))得到这些物种的菌株。

在一个更优选的实施方案中，本发明的多肽(例如，具有SEQ ID NO:2的氨基酸序列的多肽)得自米曲霉菌株，最优选地，得自米曲霉ATCC 20386或其突变菌株。

本发明的多肽也可得自微生物，此微生物与由Raper, K.D.和Fennel, D.I., 1965, 曲霉属, Wilkins公司, Baltimore定义的曲霉属是异名的同类。曲霉是有丝分裂孢子的真菌，以具有由分生孢子菌柄组成的洒水器为特征，没有已知的以泡囊结束的有性型状态，此真菌反过来具有一层或两层同时形成的不同地称作小梗或瓶梗的特化细胞和无性生殖形成的称作分生孢子的孢子。已知的曲霉属的有性型包括散囊菌属、*Neosartorya*和裸孢壳(*Emericella*)。曲霉属及它们有性型的菌株在许多培养物保藏中心(如美国典型培养物保藏中心(ATCC)、Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH(DSM)、Centraalbureau Voor Schimmelcultures(CBS)和农业研究服务机构专利培养物保藏中心、北方区

域研究中心(NRRL))易为公众得到。

对于本发明的目的，用于本文的，与一个给定的源相关的术语“得自”将意味着多肽由该源或由其中源基因已经插入的细胞产生。

如本文定义的，“分离的”多肽是实质上无其它非羧肽酶多肽，例如，如由SDS-PAGE测定的，纯度至少约20%，优选地纯度至少约40%，更优选地纯度约60%，更优选地纯度约80%，最优选地纯度约90%，最优选地纯度约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。在一个优选的实施方案中，核酸序列编码得自曲霉属(例如，米曲霉)的多肽，在一个更优选的实施方案中，核酸序列得自米曲霉ATCC 20386(例如，SEQ ID NO:1的核酸序列)。在另一个更优选的实施方案中，核酸序列是质粒pEJG12中所包含的序列，其中该质粒包含在大肠杆菌NRRL B-21616中。本发明也包含编码多肽的核酸序列，其中，多肽具有SEQ ID NO:2的氨基酸序列，且由于遗传密码的简并性不同于SEQ ID No:1。本发明也涉及编码保持羧肽酶活性的SEQ ID NO: 2片段的SEQ ID No:1子序列。优选地，子序列包含至少900个核苷酸，更优选地至少1200个核苷酸，最优选地至少1500个核苷酸。

如上所述，核酸序列可得自微生物，此微生物由 Raper,K.D. 和 Fennel,D.I.,1965，同上定义为米曲霉的异名的同类或有性型。

用于分离或克隆编码多肽的核酸序列的技术在本领域是已知的，且包括从基因组DNA、cDNA制剂、或它们的组合物中分离。可以影响本发明的这类基因组DNA的核酸序列的克隆，例如，通过利用众所周知的聚合酶链反应(PCR)或筛选表达文库的抗体以检测有共用结构特征的克隆的DNA片段。参见，例如，Innis等，1990，PCR：方法和应用指导，纽约学院出版社。可利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)。核酸序列可从曲霉属菌株或另一种或相关的有机体克隆，因此，例如，可以是核酸序列的多肽编码区的等位基因或物种变体。

用于本文的术语“分离的核酸序列”指实质上无其它核酸序列的核酸序列，例如，如通过琼脂糖电泳测定的，纯度至少约20%，优选地纯度至少

约40%，更优选地纯度至少约60%，更优选地纯度至少约80%，最优选地纯度至少约90%。例如，分离的核酸序列可通过用于基因工程的标准克隆方法获得，以将核酸序列从它的天然位置重新定位到其将在此再生的不同位点。克隆过程包括切除和分离包含编码多肽的核酸序列的要求的核酸片段，把片段插入载体分子，把重组载体导入宿主细胞，在其中复制核酸序列的多个拷贝或克隆。核酸序列可以是基因组的、cDNA、RNA、半合成的、合成来源的或它们的任何组合。

本发明也涉及编码活性多肽的，与核酸序列SEQ ID NO:1有至少约50%，优选地约60%，优选地约70%，优选地约80%，更优选地约90%，更优选地约95%，最优选地约97%的同源性的核酸序列。对于本发明的目的，两种核酸序列之间的同源程度通过Clustal方法(Higgins,1989,同上)以等同性表(一个缺口罚10，一个缺口长度罚10)测定。

编码本发明的多肽的核酸序列的修饰对合成实质上类似于该多肽的多肽是必要的。术语与多肽“实质上类似”指多肽的非天然存在形式。这些多肽以某些工程方法可以不同于从其天然源分离的多肽。例如，利用例如位点特异性诱变，它可对合成多肽变体是有益的，其中，变体在特异性活性、热稳定性、最适pH或类似的方面不同。在如编码SEQ ID NO:1部分序列(例如，它的子序列)的多肽代表的核酸序列基础上可构建类似的序列，和/或通过导入核苷酸取代构建，该取代不产生由核酸序列编码的多肽的另一种氨基酸序列，但符合为产生酶计划的宿主有机体的密码子用法，或通过导入可产生不同氨基酸序列的核苷酸取代构建。对于核苷酸取代的一般描述，参见，例如，Ford等，1991，蛋白质的表达和纯化 2:95-107。

对本领域技术人员明显的是：这类取代可在对分子功能关键性的区域外部进行，并仍产生活性多肽。对由本发明的分离的核酸序列编码的多肽活性必要的，因此，优选地不易被取代的氨基酸残基可按照本领域已知的方法(如位点特异性诱变或丙氨酸扫描诱变)鉴定(参见，例如，Cunningham和Wells, 1989，科学244:1081-1085)。在后一种技术中，在分子的每个带正电荷残基中导入突变，检测产生的突变体分子的羧肽酶活性以确定对分子活性关键性的氨基酸残基。底物-酶的作用位点也可以通过三维结构的分析(如核磁共振分析、晶体学或光亲和标记技术)测定(参见，例如，de Vos

等, 1992, 科学255:306-312; Smith等, 1992, 分子生物学杂志224:899-904; Wlodaver等, 1992, FEBS Letters 309:59-64)。

本发明的多肽也包括融合的多肽或可切割的融合多肽, 其中另一种多肽在多肽或其片段的N末端或C末端融合。融合的多肽通过融合编码另一种多肽的核酸序列(或其部分)到本发明的核酸序列(或其部分)上产生。产生融合多肽的技术在本领域是已知的, 且包括: 连接编码多肽的编码序列使它们符合读框, 且使融合的多肽在相同的启动子和终止子控制下表达。

本发明也涉及编码本发明多肽的分离的核酸序列, 该序列在低严格的条件, 更优选地在中等严格条件, 最优选地在高严格条件下与寡核苷酸探针杂交, 该探针在相同条件下与SEQ ID NO:1核酸序列或它的互补链; 或其等位基因变体和子序列杂交(Sambrook等, 1989, 同上)。

核酸构建体

本发明也涉及包含本发明可操作连接到一种或多种调控序列上的核酸序列的核酸构建体, 这些调控序列在与控制序列相容的条件下, 在合适的宿主细胞中指导编码序列的表达。表达过程将理解为包括任何多肽产生中所涉及的步骤, 包括但不限于转录、转录后修饰、翻译、翻译后修饰和分泌。

“核酸构建体”本文定义为单链或双链的核酸分子, 该核酸分子从天然存在的基因分离或为包含以某种天然存在的方式结合和并列的核酸片段已经修饰。当核酸构建体包含表达本发明的编码序列所要求的所有调控序列时, 术语核酸构建体与术语表达盒同义。本文所定义的术语“编码序列”是置于上述的调控序列的控制下, 转录为mRNA且翻译为本发明的多肽的序列。一般地, 编码序列的界线由5'末端的翻译起始密码子ATG和3'末端的翻译终止密码子确定。编码序列可以包括但不限于DNA、cDNA和重组核酸序列。

为使编码本发明的多肽的分离的核酸序列表达多肽, 可以通过各种方式操作。编码多肽的核酸序列插入载体前的操作要求或必须依赖于表达载体。利用克隆方法修改核酸序列的技术在本领域是众所周知的。

本文定义的术语“调控序列”包括所有对本发明的多肽表达必要或有利的组分。每一种调控序列可以是天然的或与编码多肽的核酸序列无关。这

类调控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。最少地，调控序列包括启动子和转录和翻译终止信号。为引入利于调控序列与编码多肽的核酸序列的编码区连接的特异性限制位点，可以为调控序列提供接头。术语“可操作地连接”本文定义为构型，其中调控序列适当地置于与DNA序列的编码序列相关的位置，这样调控序列指导多肽的产生。

调控序列可以是适合的启动子序列，即表达核酸序列的宿主细胞可识别的核酸序列。启动子序列包含调节多肽表达的转录调控序列。启动子可以是任何核酸序列，该核酸序列在选择的宿主细胞中显示转录活性(包括突变的、截短的和杂合启动子)，且可从编码胞外或胞内任意与宿主细胞同源或异源的多肽的基因获得。

用于指导(尤其在细菌宿主细胞中)转录本发明的核酸构建体的适合启动子的实例，是得自大肠杆菌乳糖操纵子、天蓝色链霉菌琼脂糖酶基因(dagA)、枯草芽孢杆菌果聚糖生成酶基因(sacB)、地衣芽孢杆菌 α -淀粉酶基因(amyL)、嗜热脂肪芽孢杆菌产麦芽糖淀粉酶基因(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xylA和xylB基因以及原核 β -内酰胺酶基因的启动子(Villa-Kamaroff等，1978，美国国家科学院论文集75:3727-3731)，以及tac启动子(DeBoer等，1983，美国国家科学院论文集80:21-25)。在科学的美国人的“重组细菌的有用蛋白质”，1980,242:74-94和Sambrook等，1989，同上中描述了另外的启动子。

用于指导在丝状真菌宿主细胞中转录本发明的核酸构建体的适合启动子的实例，是得自编码米曲霉TAKA淀粉酶、*Rhizomucor miehei*天冬氨酸蛋白酶、黑曲霉中性 α -淀粉酶、黑曲霉酸稳定的 α -淀粉酶、黑曲霉或泡盛曲霉葡糖淀粉酶(glaA)、*Rhizomucor miehei*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的基因(如美国专利No. 4,288,627中描述的，本文一并参考)和它们的突变的、截短的以及杂合启动子。用于丝状真菌宿主细胞中的特别优选的启动子是TAKA淀粉酶、NA2-tpi(得自编码黑曲霉的中性 α -淀粉酶和米曲霉丙糖磷酸异构酶的基因的杂合启动子)和glaA启动子。

在酵母宿主中，有用的启动子得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)、酿酒酵母醇脱氢酶/3-磷酸甘油醛脱氢酶基因(ADH2/GAP)和酿酒酵母3-磷酸甘油酸激酶基因。酵母宿主细胞的另一个有用的启动子由Romanos等，1992，酵母8:423-488描述。在哺乳动物宿主细胞中，有用的启动子包括病毒启动子(如得自猴病毒40(SV40)、Rous肉瘤病毒(RSV)、腺病毒和牛乳头瘤病毒(BPV)的病毒启动子)。

调控序列也可以是适合的转录终止序列，即由宿主细胞识别的终止转录的序列。终止序列可操作连接到编码多肽的核酸序列的3'末端。在选择的宿主细胞中任何有功能的终止子都可用于本发明。

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

酵母宿主细胞的优选的终止子得自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。酵母宿主细胞的其它有用的终止子由Romanos等描述，1992，同上。哺乳动物宿主细胞的终止子序列在本领域是众所周知的。

调控序列也可以是适合的前导序列，即对宿主细胞翻译重要的mRNA的非翻译区。前导序列可操作连接到编码多肽的核酸序列的5'末端。在选择的宿主细胞中任何有功能的前导序列都可用于本发明。

丝状真菌宿主细胞的优选的启动子得自编码米曲霉TAKA淀粉酶和米曲霉丙糖磷酸异构酶的基因。

酵母宿主细胞的适合启动子得自酿酒酵母烯醇化酶(ENO-1)基因、酿酒酵母3-磷酸甘油酸激酶基因、酿酒酵母 α -因子和酿酒酵母醇脱氢酶/甘油醛-3-磷酸脱氢酶基因(ADH2/GAP)。

调控序列也可以是聚腺苷酸化序列，该序列可操作连接到核酸序列的3'末端，转录时，由宿主细胞作为信号识别以增加聚腺苷残基至所转录的mRNA中。任何选择的宿主细胞中有功能的聚腺苷酸化序列都可用于本发明。

丝状真菌宿主细胞的优选的聚腺苷酸化序列得自编码米曲霉 TAKA 淀粉酶、黑曲霉葡萄糖淀粉酶、构巢曲霉邻氨基苯甲酸合酶以及黑曲霉 α -糖

苷酶的基因。

对酵母宿主细胞有用的聚腺苷酸化序列由Guo和Sherman, 1995, 分子细胞生物学15:5983-5990描述。哺乳动物宿主细胞的聚腺苷酸化序列在本领域是众所周知的。

调控序列也可以是信号肽编码区, 该区编码连接到多肽氨基末端的氨基酸序列, 其可以指导所编码的多肽进入细胞的分泌途径。核酸序列的编码序列5'末端可以内在地包含信号肽编码区, 该区在翻译读框与编码分泌多肽的编码区段天然连接。另外, 编码序列的5'末端可以包含与编码序列无关的信号肽编码区。通常, 编码序列不包含信号肽编码区时, 可能需要外源信号肽编码区。另外, 外源信号肽编码区只可以取代天然信号肽编码区以获得多肽的分泌。信号肽编码区可以得自曲霉属物种的葡糖淀粉酶或淀粉酶基因、*Rhizomucor*物种的脂肪酶或蛋白酶基因、酿酒酵母的 α -因子基因、芽孢杆菌属物种的淀粉酶或蛋白酶基因或小牛前凝乳酶原基因。然而, 指导表达的多肽进入选择的宿主细胞的分泌途径的任何信号肽编码区都可用于本发明。

细菌宿主细胞的有效信号肽编码区是得自芽孢杆菌NCIB11837的产麦芽淀粉酶基因、嗜热脂肪芽孢杆菌 α -淀粉酶基因、地衣芽孢杆菌枯草杆菌素基因、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)以及枯草芽孢杆菌PrsA基因的信号肽编码区。另外的信号肽由Simonen和Palva, 1993, 微生物学综述57:109-137描述。

丝状真菌宿主细胞的有效信号肽编码区是得自米曲霉TAKA淀粉酶基因、黑曲霉中性淀粉酶基因、*Rhizomucor miehei*天冬氨酸蛋白酶基因、*Humicola lanuginosa*纤维素酶基因或*Rhizomucor miehei*脂肪酶基因的信号肽编码区。

酵母宿主细胞的有用信号肽得自酿酒酵母 α -因子和酿酒酵母转化酶的基因。其它有用的信号肽的编码区由Romanos等, 1992, 同上描述。

调控序列也可以是前肽的编码区, 编码定位在多肽氨基末端的氨基酸序列。产生的多肽作为前酶或前多肽(或某些情况下作为酶原)是已知的。前多肽通常是无活性的, 并可从前多肽通过催化或自动催化切割前肽转化为成熟的活性多肽。前肽的编码区可得自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 -因子基因或*Myceliophthora thermophila*漆酶基因(WO 95/33836)。

本发明的核酸构建体也可包含一种或多种核酸序列,该核酸序列编码一种或多种对指导多肽表达有利的因子((例如,激活物(例如,反式作用因子)、陪伴分子以及加工蛋白酶)。任何在选择的宿主细胞中有功能的因子都可用于本发明。编码一种或多种这些因子的核酸与编码多肽的核酸序列未必是一前一后的。

激活物是激活编码多肽的核酸序列转录的蛋白质(Kudla等, 1990, EMBO杂志 9:1355-1364; Jarai和Buxton, 1994, 现代遗传学 26: 2238-244; Verdier, 1990, 酵母6: 271-297)。编码激活物的核酸序列可得自编码嗜热脂肪芽孢杆菌NprA(nprA)、酿酒酵母血红蛋白激活物蛋白质1(hapl)、酿酒酵母半乳糖代谢蛋白质4(gal 4)以及构巢曲霉氮调节蛋白质(areA)的基因。对于进一步的例子,参见Verdier, 1990, 同上和MacKenzie等, 1993, 普通微生物学杂志139: 2295-2307。

陪伴分子是协助另一种多肽适当折叠的蛋白质(Hartl等, 1994, TIBS 19: 20-25; Bergeron等, 1994, TIBS 19: 124-128; Demolder等, 1994, 生物技术杂志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269: 7764-7771; Wang和Tsou, 1993, FASEB杂志 7: 1515-11157; Robinson等, 1994, 生物/技术 1: 381-384)。编码陪伴分子的核酸序列可得自编码枯草芽孢杆菌GroE蛋白质、米曲霉蛋白质二硫化物异构酶、酿酒酵母钙联接蛋白、酿酒酵母BiP/GRP78以及酿酒酵母Hsp70的基因。进一步的实施例参见Gething和Sambrook, 1992, 同上和Hartl等, 1994, 同上。

加工蛋白酶是切割前肽以产生成熟的有生物化学活性的多肽的蛋白酶(Enderlin和Ogrydziak, 1994, 酵母10:67-79; Fuller等, 1989, 美国国家科学院论文集86:1434-1438; Julius等, 1984, 细胞37:1075-1089; Julius等, 1983, 细胞32:839-852)。编码加工蛋白酶的核酸序列可得自编码酿酒酵母二肽基氨基酸酶、酿酒酵母Kex2和*Yarrowia lipolytica* 二元加工内蛋白酶(xpr6)的基因。

理想的也可加入使得可相对于宿主细胞的生长调节多肽表达的调节序

列。调节系统的实例是使基因表达打开或关闭以响应化学或物理刺激的调节系统，包括调节化合物的存在。原核生物系统中的调节系统将包括lac、tac和trp操纵子系统。在酵母中，可以利用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使得基因扩增的序列。在真核生物系统中，这些调节序列包括氨甲蝶呤存在时扩增的二氢叶酸还原酶基因和以重金属扩增的金属硫蛋白基因。在这些情况下，编码多肽的核酸序列将与调节序列可操作地连接。

表达载体

本发明也涉及包含本发明的核酸序列、启动子以及转录和翻译终止信号的重组表达载体。上述的各种核酸和调控序列可连接在一起以产生重组表达载体，此载体可包括一个或多个方便的限制酶切位点使得在这些位点插入或取代编码多肽的核酸序列。另外，本发明的核酸序列可通过将核酸序列或包含该序列的核酸构建体适合表达的载体中来表达。在构建表达载体时，编码序列定位在载体中使得编码序列与用于表达及可能用于分泌的适合的控制序列可操作地连接。

重组表达载体可以是可便利地受重组DNA过程的作用并可引起核酸序列表达的任何载体(例如，质粒或病毒)。载体的选择典型地取决于载体与载体将引入的宿主细胞的相容性。载体可以是线型的或闭环状质粒。载体可以是自主复制载体，即，作为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、染色体外因子、小型染色体或人工染色体。载体可以包含任何用以保证自我复制的手段。载体也可以在引入宿主细胞时整合进入基因组，并与其整合进入的染色体一起复制。载体系统可以是单个的载体或质粒或两个或多个载体或质粒，它们共同包含将被引入宿主细胞基因组或转座子中的全部DNA。

本发明的载体优选地包含一种或多种选择性标记，这类标记使得转化的细胞容易选择。选择性标记是一种基因，其产物具有生物杀伤剂或病毒抗性、对重金属的抗性、针对营养缺陷型的原养型及其类似的性质。细菌的可选择标记的例子是来自枯草芽孢杆菌或地衣芽孢杆菌的dal基因，或给与抗生素抗性(如氨苄青霉素、卡那霉素、氯霉素或四环素抗性)的标记。常用

的哺乳动物的标记是二氢叶酸还原酶基因。酵母宿主细胞的适合的标记是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选自包括但不限于amdS(乙酰胺酶)、argB(鸟氨酸氨甲酰转移酶)、bar(蒴丝菌素乙酰转移酶)、hygB(潮霉素磷酸转移酶)、niaD(硝酸盐还原酶)、pyrG(乳清酸核苷-5'-磷酸脱羧酶)、sC(硫酸腺苷转移酶)、trpC(邻氨基苯甲酸合酶)的组和glufosinate抗性标记以及其它物种的等价物。优选的用于曲霉属细胞的标记是构巢曲霉或米曲霉的amdS和pyrG基因以及吸水链霉菌的bar基因。此外,选择可以通过共转化完成,例如,如WO 91/17243中所述的,其中选择性标记在单独的载体上。

本发明的载体优选地包含一种元件,此元件使得载体稳定的整合进入宿主细胞基因组,或使细胞中载体的自我复制不依赖于细胞的基因组。

当把本发明的载体引入宿主细胞时,它可以整合进入宿主细胞的基因组。对于整合,载体可依赖于编码多肽的核酸序列或载体的用于通过同源或异源的重组把载体稳定整合进入基因组的其它元件。另外,载体可包含用于指导通过同源重组整合进入宿主细胞基因组的附加核酸序列。附加的核酸序列能使载体在染色体的精确位置整合进入宿主细胞基因组。为了增加在精确位置整合的可能性,整合元件优选地应该包含足够数量的核酸,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,且和对应的靶序列是高度同源的以提高同源重组的概率。整合元件可以是与宿主细胞的基因组的靶序列同源的任何序列。此外,整合元件可以是非编码或编码核酸序列。另一方面,载体可以通过异源重组整合进入宿主细胞的基因组。这些核酸序列可以是与宿主细胞基因组的靶序列同源的任何序列,此外,可以是非编码或编码序列。

为了自主复制,载体还可以包含使载体能够在正讨论的宿主细胞中自主复制的复制起点。细菌的复制起点的例子是使得在大肠杆菌中复制的质粒pBR322、pUC19、pACYC177和pACYC184,以及使得在芽孢杆菌中复制的pUB110、pE194、pTA1060和pAM β 1的复制起点。用于酵母宿主细胞中的复制起点的例子是2微米的复制起点、ARS1、ARS4、ARS1和CEN3的组合体和ARS4和CEN6的组合体。复制起点可以是具有突变的起点,该突变使其在宿主细胞中对温度敏感(参见,例如, Ehrlich, 1978, 美国国家科学

院论文集75:1433)。

一个以上编码本发明的多肽的核酸序列的拷贝可以插入到宿主细胞中以扩增核酸序列的表达。核酸序列的稳定扩增可以通过把至少一种附加的序列的拷贝整合进入宿主细胞基因组或通过包含具有核酸序列的可扩增的选择性标记基因获得,其中,在适当的选择性试剂存在时,通过培养细胞,可选择包含选择性标记基因的扩增拷贝及它的核酸序列的附加拷贝的细胞。

连接上述的元件构建本发明的重组表达载体所利用的方法对本领域技术人员是众所周知的(参见,例如, Sambrook等, 1989, 同上)。

宿主细胞

本发明也涉及重组宿主细胞,此细胞包含本发明的核酸序列,并方便地用于多肽的重组产生。术语“宿主细胞”包含亲本细胞的任何子代,由于复制过程中发生突变,子代与其亲本细胞不同。

优选地细胞以包含本发明的核酸序列的载体转化,其后载体整合进入宿主细胞的染色体中。“转化”意味着:包含本发明核酸序列的载体导入宿主细胞,因此,载体作为染色体整合体或作为自我复制的额外染色体载体保持。由于核酸序列更可能稳定地保持在细胞中,通常认为整合是有利的。进入宿主染色体的载体的整合如上所述的可通过同源或异源重组发生。

对宿主细胞的选择很大程度上将取决于编码多肽和其源的基因。宿主细胞可以是单细胞微生物(例如,原核生物)或多细胞微生物(例如,真核生物)。有用的单细胞是细菌细胞(如包括但不限于芽孢杆菌细胞(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、凝结芽孢杆菌、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)或链霉菌属细胞(如变铅青链霉菌或鼠灰链霉菌)的革兰氏阳性细菌或革兰氏阴性细菌(如大肠杆菌和假单胞菌某种))。在一个优选的实施方案中,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。例如,细菌宿主细胞的转化可通过原生质体的转化(参见,例如, Chang 和Cohen, 1979, 普通分子遗传学 168:111-115)、通过利用感受态细胞(参见,例如, Young and Spizizin, 1961, 细菌学杂志 81:823-829, 或Dubnau和Davidoff-Abelson, 1971, 分

子生物学杂志56:209-221)、通过电穿孔法(参见,例如, Shigekawa 和 Dower, 1988, 生物技术6:742-751)或通过缀合(参见,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、微型仓鼠肾(BHK)细胞、COS细胞或其它可供使用的任何数量的无限增殖化细胞系(如来自美国典型培养物保藏中心)。

在一个优选的实施方案中,宿主细胞是真菌细胞。本文利用的“真菌”包括子囊菌亚门、担子菌亚门、壶菌亚门以及接合菌亚门(如由Hawksworth等,在真菌的Ainsworth和Bisby辞典,第8版,1995,英国剑桥国际民用航空局,大学出版社)以及卵菌亚门(如在Hawksworth等,1995,同上,171页中引用)和所有有丝分裂孢子的真菌(Hawksworth等,1995,同上)。子囊菌亚门代表类型包括:例如,链孢霉属、正青霉属(=Penicillium)、裸孢壳属(=Aspergillus)、散囊菌属(=Aspergillus)以及以下所列出的真正的酵母。担子菌亚门的例子包括蘑菇、锈斑病和黑穗病。壶菌亚门的代表类型包括:例如,异水霉属、小芽枝霉属、雕蚀菌属以及水生真菌。卵菌亚门的代表类型包括:例如,水霉性水生真菌(水霉菌)(如绵霉属)。有丝分裂孢子的真菌的例子包括曲霉属、青霉属、念珠菌属、和链格孢属。接合菌亚门的代表类型包括:例如,根霉属和毛霉属。

在一个优选的实施方案中,真菌宿主细胞是酵母细胞。本文所利用的“酵母”包括产子囊酵母(内孢霉目)、产担孢子酵母以及属于半知菌类(芽孢菌纲)的酵母。产子囊酵母分为蚀精霉科和酵母科。后者包含四个亚科:裂殖酵母亚科(如,裂殖酵母属)、拿逊酵母亚科、油脂酵母亚科以及酵母菌亚科(如克鲁维酵母属、毕赤酵母属和酵母属)。产担孢子酵母包括白冬孢酵母属、紫红孢酵母属、锁掷酵母属、线担子酵母属以及Filobasidiella。属于半知菌类的酵母分为两个科:掷孢酵母科(例如, Sorobolomyces和布勒掷孢酵母属)和隐球酵母科(例如,念珠菌属)。由于酵母的分类将来可能变化,为了本发明的目的,将如在酵母的生物学和活性中描述的(Skinner, F.A., Passmore, S.M.和Davenport, R.R.编辑, Soc. App. Bacteriol.专题研讨会系列 No.9,1989)定义酵母。酵母生物学和酵母遗传学的操作在本领域是众

所周知的(参见, 例如, 酵母的生物化学和遗传学, Bacil, M., Horecker, B.J.和Stopani, A.O.M.编辑, 第二版, 1987; 酵母, Rose, A.H.和Harrison, J.S.编辑, 第二版, 1987; 和酵母属酵母的分子生物学, Strathern等, 编辑, 1981)。

在一个更优选的实施方案中, 酵母宿主细胞是念珠菌属、克鲁维酵母属、毕赤酵母、酵母属、或洋著草属的一个种的细胞。

在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。在另一个最优选的实施方案中, 酵母宿主细胞是乳酸克鲁维酵母细胞。在另一个最优选的实施方案中, 酵母宿主细胞是*Yarrowia lipolytica*细胞。

在一个优选的实施方案中, 真菌宿主细胞是丝状的真菌细胞。“丝状真菌”包括真菌门和卵菌门的亚门的所有丝状形式(如由Hawksworth等, 1995, 同上定义)。丝状真菌以具有由壳多糖、纤维素、葡聚糖、脱乙酰壳多糖、甘露聚糖和其它复合多糖组成的菌丝体壁为特征。营养体的生长通过菌丝延长, 碳分解代谢是专性需氧的。相对地, 通过酵母(如酿酒酵母)的营养体生长是由于萌发了单细胞的叶状体, 碳分解代谢可以是发酵。在一个更优选的实施方案中, 丝状真菌宿主细胞是但不限于支顶孢属、曲霉属、镰孢属、腐质霉属、毛霉属、毁丝霉属、链孢霉属、青霉属、梭孢壳属、弯劲霉属和木霉属的一个种的细胞。

在一个更优选的实施方案中, 丝状真菌宿主细胞是曲霉属细胞。在另一个更优选的实施方案中, 丝状真菌宿主细胞是支顶孢属细胞。在另一个更优选的实施方案中, 丝状真菌宿主细胞是镰孢属细胞。在另一个更优选的实施方案中, 丝状真菌宿主细胞是腐质霉属细胞。在另一个更优选的实施方案中, 丝状真菌宿主细胞是毛霉属细胞。在另一个更优选的实施方案中, 丝状真菌宿主细胞是毁丝霉属细胞。在另一个更优选的实施方案中, 丝状真菌宿主细胞是链孢霉属细胞。在另一个更优选的实施方案中, 丝状真菌宿主细胞是青霉属细胞。在另一个更优选的实施方案中, 丝状真菌宿主细胞是梭孢壳属细胞。在另一个更优选的实施方案中, 丝状真菌宿主细胞是弯劲霉属细胞。在另一个更优选的实施方案中, 丝状真菌宿主细胞是木霉

属细胞。

在最优选的实施方案中，丝状真菌宿主细胞是泡盛曲霉、*Aspergillus foetidus*、*Aspergillus japonicus*、构巢曲霉、黑曲霉或米曲霉细胞。在另一个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、檣叶槭镰孢、尖镰孢、多枝镰孢、粉红镰孢、接骨木镰孢、肤色镰孢、*Fusarium sulphureum*、*Fusarium toruloseum*、*Fusarium trichothecioides*或*Fusarium venenatum*细胞。在最优选的实施方案中，丝状真菌亲本细胞是*Fusarium venenatum*(Nirenberg sp.nov.)。在另一个最优选的实施方案中，丝状真菌宿主细胞是*Humicola insolens*或*Humicola lanuginosa*细胞。在另一个最优选的实施方案中，丝状真菌宿主细胞是米黑毛霉细胞。在另一个最优选的实施方案中，丝状真菌宿主细胞是*Myceliophthora thermophilum*细胞。在另一个最优选的实施方案中，丝状真菌宿主细胞是粗糙脉孢菌细胞。在另一个最优选的实施方案中，丝状真菌宿主细胞是产紫青霉细胞。在另一个最优选的实施方案中，丝状真菌宿主细胞是*Thielavia terrestris*细胞。在另一个最优选的实施方案中，木霉细胞是*Trichoderma harzianum*、康宁木霉、*Trichoderma longibrachiatum*、*Trichoderma reesei*或绿色木霉细胞。

真菌细胞可以通过包括以本身的已知的方式的原生质体形成、原生质体转化以及细胞壁的再生转化。转化曲霉属宿主细胞的适合的方法在EP238 023和Yelton等，1984，美国国家科学院论文集 81:1470-1474中描述。转化镰孢属某种的适合的方法由Malardier等，1989，基因 78:147-156或在WO 96/00787中描述。酵母可利用由Becker和Guarente，在Abelson中，J.N.和Simon，M.I.编辑，酵母遗传学和分子生物学指南，酶学方法，第194卷，pp.182-187，纽约学院出版社公司；Ito等，1983，细菌学杂志 153:163和Hinnen等，1978，美国国家科学院论文集 75:1920描述的方法转化。哺乳动物的细胞可以利用Graham和Van der Eb的磷酸钙沉淀方法(1978，病毒学 52:546)通过直接摄取转化。

产生方法

本发明也涉及用于产生本发明的多肽的方法，此方法包括(a)培养其野生

型形式能够产生多肽的菌株，以产生包含多肽的上清液；和(b)回收多肽。优选地，该菌株是曲霉属的。

本发明也涉及用于产生本发明的多肽的方法，此方法包括(a)在对产生多肽有益条件下培养宿主细胞；和(b)回收多肽。

在两种方法中，细胞在产生多肽的适合营养培养基中利用本领域已知的方法培养。例如，细胞可以通过摇瓶培养、小规模或大规模发酵作用(包括连续的、分批的、补料分批或固态发酵)培养，其中发酵是在实验室或工业发酵罐中，在适合的培养基中或在使多肽表达和/或分离的条件下进行的。利用本领域已知的方法，培养发生在包含碳和氮源以及无机盐的适合营养培养基中(参见，例如，细菌和酵母的参考；Bennett, J.W.和LaSure, L. 编辑，真菌中的更多基因操作，学院出版社，CA，1991)。适合的培养基由市售获得或可以按照公开的组合物(例如，在美国典型培养物保藏中心的目录中)制备。如果多肽分泌进入营养培养基，则多肽可从培养基直接回收。如果不分泌多肽，则多肽从细胞裂解物回收。

多肽可利用本领域技术人员已知的、对多肽特异的方法检测。这些检测方法可包括：特异性抗体的利用、酶产物的形成或酶底物的消失。例如，酶测定可用于测定多肽的活性。用于测定羧肽酶活性的方法在本领域是已知的，且包括，例如，根据Roth, 1971，分析化学 43:880的方法，邻苯二醛与二硫苏糖醇监测酶水解下游离氨基酸的释放。

产生的多肽可通过本领域已知的方法回收。例如，多肽可通过包括但不限于离心、过滤、萃取、喷雾干燥、蒸发或沉淀的常规方法从营养培养基回收。

本发明的多肽可通过本领域已知的各种方法纯化，这些方法包括但不限于：色谱(例如，离子交换色谱、亲和色谱、疏水色谱、色谱聚焦以及大小排阻色谱)、电泳法(例如，制备性等电聚焦(IEF))、不同溶解性(例如，硫酸铵沉淀)、SDS-PAGE或萃取(参见，例如，蛋白质纯化，J.-C. Janson和Lars Ryden编辑，纽约VCH出版社，1989)。

羧肽酶活性的除去或降低

本发明也涉及用于产生亲本细胞的突变细胞的方法，该方法包括破坏或缺失编码多肽的核酸序列或其调控序列，这导致突变细胞产生的多肽比亲

本细胞产生的多肽少。

构建已降低羧肽酶活性的菌株可通过核酸序列的修饰或灭活方便地完成，此核酸序列对细胞中有羧肽酶活性的多肽的表达是必要的。例如，要修饰或灭活的核酸序列可以是编码对显示羧肽酶活性必要的多肽或其部分的核酸序列，或此核酸序列可具有来自核酸序列的编码序列的多肽表达所要求的调节功能。这类调节或调控序列的例子可以是启动子序列或其功能部分(即，足以影响多肽表达的部分)。其它可能用于修饰的调控序列在上文已描述。

核酸序列的修饰或灭活可通过使细胞诱变并且选择其中产生羧肽酶能力已降低的细胞进行。可以是特异性的或随机的诱变，例如，可通过利用适合的物理或化学诱变剂、通过利用适合的寡核苷酸或通过使DNA序列受到PCR产生的诱变进行。此外，诱变可通过任意结合利用这些诱变剂进行。

用于本发明目的的适合的物理或化学诱变剂的例子包括：紫外(UV)线、羟胺、N-甲基-N'-硝基-N-亚硝基胍(MNNG)、O-甲基羟胺、亚硝酸、甲磺酸乙酯、亚硫酸氢钠、甲酸以及核苷酸类似物。

当利用这类试剂时，典型地，诱变是通过在适合条件下，在选择的诱变剂存在时，培养细胞使发生诱变，并选择显示降低的羧肽酶活性的细胞或产物进行。

本发明的修饰或灭活多肽的产生可通过引入、取代或除去编码多肽的核酸序列或其转录或翻译所要求的调节元件中的一个或多个核苷酸完成。例如，可插入或除去核苷酸以导致引入终止密码子、除去起始密码子或开放读框的变化。这类修饰或灭活可按照本领域已知的方法，通过位点特异性诱变或PCR产生的诱变完成。虽然，原则上，修饰可在活体内完成，即，直接在细胞上表达核酸序列以使修饰，但优选的是：如以下举例说明的，修饰在活体外进行。

通过选择的宿主细胞灭活或降低产物的方便方法的例子是基于基因置换或基因中断的技术。例如，在基因中断方法中，对应于内源性基因或兴趣基因片段的核酸序列在体外诱变以产生缺损的核酸序列，然后，把此序列转化进入宿主细胞以产生缺损的基因。通过同源重组，缺损核酸序列取代内源性基因或基因片段。理想的是：缺损基因或基因片段也编码用于选

择转化体的标记物，其中，该转化体中编码多肽的基因已修饰或破坏。

另外，编码本发明多肽的核酸序列的修饰或灭活可利用与编码多肽的序列互补的核苷酸序列，通过建立的反义技术进行。更具体地说，利用细胞的产生多肽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除去，其中，该多肽可在细胞中转录，并可与细胞中产生的多肽mRNA杂交。因此，在允许互补的反义核苷酸序列与多肽mRNA杂交的条件下，可降低或除去翻译的多肽的量。

优选的是：按照本发明的方法，要修饰的细胞是微生物来源的，例如，真菌菌株适合于所要求的蛋白质产物、细胞的同源物或异源物的产生。

本发明还涉及亲本细胞的突变细胞，其包含编码多肽的核酸序列或其调控序列的破坏或缺失，这导致突变细胞产生的多肽比亲本细胞产生的多肽少。

这样构建的缺乏多肽的突变细胞作为表达同源和/或异源多肽的宿主细胞是特别有用的。因此，本发明也涉及产生同源或异源多肽的方法，该方法包括(a)在有利于产生多肽的条件下，培养突变细胞；和(b)回收多肽。在本文中，术语“异源多肽”定义为不是宿主细胞的天然多肽，其中已修饰改变了天然序列的天然蛋白质或其表达由于用重组DNA技术操纵宿主细胞定量改变的天然蛋白质。

另一方面，本发明也涉及通过产生本发明的多肽以及兴趣蛋白质产物的细胞发酵，产生实质上无羧肽酶活性的蛋白质产物的方法。此方法包含在发酵期间或发酵完成后向发酵肉汤加入能够抑制羧肽酶活性的有效量的试剂，从发酵肉汤回收兴趣产物以及有选择地使回收的产物进一步纯化。此方法在以下实施例中进一步说明。

另一方面，本发明涉及产生实质上无羧肽酶活性的蛋白质产物的方法，其中兴趣蛋白质产物由细胞中存在的编码本发明多肽的DNA序列编码。此方法包括在允许产物表达的条件下培养细胞，使产生的培养肉汤受到联合的pH和温度处理以实质上降低羧肽酶活性并从培养肉汤回收产物。另外，联合的pH和温度处理可在由培养肉汤回收的酶制剂上进行。联合的pH和温度处理可选择性地与以羧肽酶抑制剂的处理结合利用。

依据本发明的这方面，可能除去至少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%的羧肽酶活性。人们期望通过利用这种方法，可完全除去羧肽酶活性。

优选地，联合的pH和温度处理以6.5-7的pH范围和25-40℃的温度范围进行足够时间以达到要求的效果，典型地，30-60分钟是足够的。

用于培养和纯化兴趣产物的方法可通过本领域已知的方法进行。

本发明的产生实质上无羧肽酶活性产物的方法在真核生物多肽，尤其真菌蛋白质(如酶)的产生中特别有利。酶可选自，例如，淀粉酶、脂酶、蛋白酶、溶纤酶、氧化还原酶或植物细胞壁降解酶。这类酶的例子包括氨肽酶、淀粉酶、淀粉葡萄糖苷酶、糖酶、羧肽酶、过氧化氢酶、纤维素酶、壳多糖酶、角质酶、环糊精糖基转移酶、脱氧核糖核酸酶、酯酶、半乳糖苷酶、 β -半乳糖苷酶、葡萄糖淀粉酶、葡萄糖氧化酶、葡萄糖苷酶、卤代过氧化物酶、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、果胶溶解酶、过氧化物酶、植酸酶、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。缺乏羧肽酶的细胞也可用于表达有药物作用的异源蛋白质，如激素、生长因子、受体和类似物。

要理解的是：术语“真核生物多肽”不仅包括天然多肽，而且包含那些已通过氨基酸取代、缺失或添加或其它这类的修饰以提高活性、耐热性、pH耐受性和类似性质的修饰的多肽(例如，酶)。

另一方面，本发明涉及通过本发明的方法产生的实质上无羧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以提高水解度和香味发展。

本发明也涉及与内肽酶结合利用本发明的多肽以使富含蛋白质物质的高度水解。此方法包括以多肽和内肽酶处理蛋白质底物。底物可用酶同时或连续处理。

通常，把本发明的多肽以蛋白质水解过程中常规利用的有效量(优选地，为每100g蛋白质约0.1-约100,000 CPDU的范围，更优选地，为每100g蛋白质约1-约10,000 CPDU的范围)加至蛋白质底物中。如本文所定义的，

一个CPDU(羧肽酶单位)是在pH 4.5和25℃下, 每分钟从0.5 mM 的N-CBZ-Ala-Glu(Sigma化学公司, St.Louis MO)溶液中释放1微摩尔谷氨酸的羧肽酶量。

内肽酶可得自芽孢杆菌属(优选地为地衣芽孢杆菌或枯草芽孢杆菌)菌株、葡萄球菌属(优选地为金黄色葡萄球菌)菌株、链霉菌属(优选地为*Streptomyces thermovularis*或*Streptomyces griseus*)菌株、放线菌属物种的菌株、曲霉属(优选地为棘孢曲霉、泡盛曲霉、*Aspergillus foetidus*、构巢曲霉、黑曲霉或米曲霉)菌株或镰孢属(优选地为*Fusarium venenatum*)菌株。

通常, 把内肽酶以蛋白质水解过程中常规利用的有效量(优选地, 为每100g蛋白质约0.05-约15AU的范围, 更优选地, 为每100g蛋白质约0.1-约8AU的范围)加至蛋白质底物中。一个AU(Anson单位)定义为: 标准条件(即25℃、pH 7.5和10分钟的反应时间)下, 以初始速率消化血红蛋白, 使每分钟释放的TCA可溶性产物的量以苯酚试剂显示的颜色与一毫当量的酪氨酸以苯酚试剂显示的颜色相同的酶量。分析方法AF 4/5经请求可以从Novo Nordisk A/S, 丹麦得到, 这在本文一并参考。

酶处理(即与酶制剂温育底物)可发生在合适的温度(该温度下, 酶制剂不失活), 优选地, 在约20℃-约70℃的范围。依据所形成的惯例, 酶制剂可通过适当地增加温育混合物的温度到酶失活的温度(例如, 约70℃以上), 或类似地通过降低温育混合物的pH到酶失活的pH(例如, 在约4.0以下)而失活。

此外, 本发明的方法导致蛋白质底物的水解度提高。如本文所使用的, 水解度(DH)是: 在已由蛋白酶水解的蛋白质中氨基键总数的百分数。

本发明的另一方面, 水解产物中Leu、Gly、Ala和/或Pro的含量增加, 例如, 增大1.1倍。

本发明也涉及获得富含游离谷氨酸和/或结合肽的谷氨酸残基的水解产物的方法, 这一方法包括:

- (a) 使底物经历脱酰胺过程; 和
- (b) 使具有羧肽酶活性的多肽作用于底物。

两个步骤可同时进行或第二步骤可以在第一个步骤后进行。

本发明的这些方法产生极好香味的蛋白质水解产物, 因为, 不论游离的还是肽结合的谷氨酸(Glu)在蛋白质水解产物的香味和适口性中都起重要的

作用。这些方法也产生官能度已提高(尤其是,提高溶解性、提高乳化性质、增加水解度及提高起沫性质)的蛋白质水解产物。

通过氨的释放,由酰胺(谷氨酰胺或天冬酰胺)转化为荷电酸(谷氨酸或天冬氨酸)被认为是脱酰胺作用。脱酰胺作用可作为非酶促或酶促脱酰胺作用过程发生。

在一个优选的实施方案中,脱酰胺作用作为酶促脱酰胺作用过程(例如,通过使底物受转谷氨酰胺酶和/或肽谷氨酰胺酶作用)进行。

转谷氨酰胺酶可是任何方便来源(包括哺乳动物(参见,例如,JP 1050382和JP 5023182)、包括活化因子XIII(参见,例如,WO 93/15234))的;或衍生自鱼的转谷氨酰胺酶(参见,例如,EP 555,649)和得自微生物的转谷氨酰胺酶(参见,例如,EP 379,606,WO 96/06931和WO 96/22366)。在一个优选的实施方案中,转谷氨酰胺酶得自卵菌纲(包括*Phytophthora*(优选地为*Phytophthora cactorum*)菌株或包括*Pythium*(优选地为*Pythium irregulare*、*Pythium sp.*、*Pythium intermedium*、*Pythium ultimum*或*Pythium peritum*(或*Pythium periplocum*)菌株。在另一个优选的实施方案中,转谷氨酰胺酶是细菌来源的,并得自芽孢杆菌菌株(优选地为枯草芽孢杆菌)、链霉菌属(优选地为*Streptomyces*属(优选地为*Streptomyces mobaraensis*、*Streptomyces griseocarneum*或*Streptomyces cinnamomeus*)菌株和链霉菌属(优选地为*Streptomyces lydicus*)菌株。

肽谷氨酰胺酶可是肽谷氨酰胺酶I(肽基谷氨酰胺酶; 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)或其任意混合物。肽谷氨酰胺酶可得自曲霉属(优选地为*Aspergillus japonicus*)菌株、芽孢杆菌(优选地为环状芽孢杆菌)菌株、隐球菌属(优选地为*Cryptococcus albidus*)菌株或德巴利酵母属(优选地为克洛德巴利酵母)菌株。

通常,把转谷氨酰胺酶以脱酰胺过程中常规利用的有效量(优选地,为酶制剂相对于底物的量约0.01-约5%(w/w)的范围,更优选地,为约0.1-约1%(w/w)的范围)加至蛋白质底物中。

通常,把肽谷氨酰胺酶以脱酰胺过程中常规利用的有效量(优选地,为每100g底物约0.01-约100,000PG酶单位的范围,更优选地,为每100g底物约0.1-约10,000PG酶单位的范围)加至蛋白质底物中。

按照Cedrangoro等的方法(1965, 酶学29:143), 可测定肽谷氨酰胺酶的活性。按照此方法, 把0.5ml以1N NaOH调整pH到6.5的酶样品装入一小容器中。然后, 把1ml pH为10.8的硼酸盐缓冲液加至容器中。排出的氨由5N硫酸吸收, 并通过利用Nessler试剂, 使混合物形成在420nm测定的颜色。一个PG酶单位是在这些条件下每分钟能够产生1微摩尔氨的酶量。

另外, 肽谷氨酰胺酶活性可按照US 3,857,967或以下实施例11描述的方法测定。

在本发明方法的(b)步骤中, 使本发明的多肽作用于底物。通常, 把本发明的多肽以蛋白质水解过程中常规利用的有效量(优选地, 为每100g底物约0.001-约0.5AU的范围, 更优选地, 为每100g底物约0.01-约0.1AU的范围)加至蛋白质底物中。

在另一个实施方案中, 本发明的产生富含游离谷氨酸和/或结合肽的谷氨酸残基的水解产物的方法还包括:

(c) 使一种或多种非特异性的内和/或外肽酶作用于底物。

此步骤可与步骤(a)和(b)同时进行, 或可接着步骤(a)和(b)进行。

在一个优选的实施方案中, 非特异性作用的内和/或外肽酶得自曲霉属(优选地为黑曲霉、米曲霉、或*Aspergillus sojae*)菌株、或芽孢杆菌属(优选地和解淀粉芽孢杆菌、迟缓芽孢杆菌、地衣芽孢杆菌或枯草芽孢杆菌)菌株。

通常, 把非特异性作用的内和/或外肽酶以蛋白质水解过程中常规利用的有效量(优选地, 为每100g底物约0.05-约15CPU的范围, 更优选地, 为每100g底物约0.1-约5CPU的范围)加至蛋白质底物中。一个CPU(酪蛋白酶单位)定义为: 在标准条件(即, 在25℃和pH 9.5下温育30分钟)下, 每分钟从酪蛋白释放1微摩尔伯氨基(通过与丝氨酸标准比较测定)的酶量。本文一并参考的分析方法AF 228/1经请求可从Novo Nordisk A/S, Bagsværd, 丹麦得到。

每种酶处理可在酶制剂未失活的任意温度(优选地, 在约20℃-约70℃的范围)进行。然后, 酶制剂可通过增加温度(例如, 到约70℃之上)或降低pH(例如, 在约4.0之下)灭活。

用于本发明方法的蛋白质底物可由完整的蛋白质、预先水解的蛋白质(即, 肽)或其混合物组成。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,大豆蛋白质、谷物蛋白质(例如,小麦谷蛋白、玉米谷蛋白、大麦、黑麦、燕麦、稻米、玉米醇溶蛋白、羽扇豆属植物、棉花种子蛋白质、油菜种子蛋白质、花生、苜蓿蛋白质、豌豆蛋白质、大豆蛋白质、芝麻种子蛋白质或向日葵)。动物来源的蛋白质底物可以是乳清蛋白质、酪蛋白、肉蛋白质、鱼蛋白质、红细胞、卵清蛋白、明胶或乳白蛋白。

本发明也涉及由这些方法产生的蛋白质水解产物。

其它用途

本发明也涉及用本发明的多肽灭活酶的方法。

此外,本发明的多肽可对许多要求肽序列特异性裂解的目的是有用的。例如,一些蛋白质或肽以成熟蛋白质的N端包含许多附加氨基酸残基的无活性前体形式合成。本发明的多肽可提供必要的翻译后加工以激活这类前体蛋白质。

多肽组合物

另一方面,本发明涉及包含本发明多肽的多肽组合物。优选地,组合物富含本发明的多肽。在本文中,术语“富含”表明:多肽组合物的羧肽酶活性已增加,例如,富集因子为1.1。

多肽组合物可包含本发明的多肽作为主要的酶促组分(例如,单成分的多肽组合物)。另外,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、壳多糖酶、角质酶、环糊精糖基转移酶、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、卤代过氧化物酶、转化酶、漆酶、脂肪酶、甘露糖苷酶、氧化酶、果胶分解酶、肽谷氨酰胺酶、过氧化物酶、植酸酶、多酚氧化酶、蛋白酶、核糖核酸酶、转谷氨酰胺酶或木聚糖酶。附加酶借助于属于曲霉属(优选地为棘孢曲霉、泡盛曲霉、黑曲霉或米曲霉)或木霉属、腐质霉属(优选地为 *Humicola insolens*)或镰孢属(优选地为杆孢状镰孢、*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、多枝镰孢、粉红镰孢、接骨木镰孢、肤色镰孢、*Fusarium sulphureum*、*Fusarium toruloseum*、*Fusarium trichothecioides*或*Fusarium venenatum*)的微生物是可生产的。

在一个优选的实施方案中,本发明涉及包含具有羧肽酶活性的多肽的香味改良组合物。在另一个优选的实施方案中,香味改良组合物还包含一种或多种非特异性作用的内和/或外肽酶。在另一个优选的实施方案中,调味组合物还包含一种或多种特异性作用的内和/或外肽酶。

在一个优选的实施方案中,特异性作用的蛋白酶是一种内肽酶(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/精氨酸特异性)内肽酶或肽基-天冬氨酸金属内肽酶(EC 3.4.24.33))。

优选地,谷氨酰内肽酶(EC 3.4.21.19)可得自芽孢杆菌属(尤其是地衣芽孢杆菌和枯草芽孢杆菌)菌株、葡萄球菌属(尤其是金黄色葡萄球菌)菌株、链霉菌属(尤其是*Streptomyces thermovulgaris*和*Streptomyces griseus*)菌株或Acrinomyces sp.菌株。

优选地,赖氨酰内肽酶(EC 3.4.21.50)可得自无色杆菌属(尤其是*Achromobacter lyticus*)菌株、溶杆菌属(尤其是产酶溶杆菌)菌株或假单胞菌属(尤其是铜绿假单胞菌)菌株。

亮氨酰内肽酶(EC 3.4.21.57)可是植物来源的。

优选地,甘氨酰内肽酶(EC 3.4.22.25)可得自番木瓜植物(番木瓜树)。

优选地,脯氨酰内肽酶(EC 3.4.21.26)可得自产黄菌属菌株,或它可以是植物来源的。

优选地,胰蛋白酶样内肽酶可得自镰孢属,尤其是尖镰孢菌株,例如,如WO 89/06270或WO 94/25583中描述的。

优选地,肽基-天冬氨酸金属内肽酶(EC 3.4.24.33)可得自假单胞菌属菌株,尤其是*Pseudomonas fragi*菌株。

在另一个优选的实施方案中,特异性作用的蛋白酶是可从肽两端作用的外肽酶。

在一个优选的实施方案中,特异性作用的蛋白酶是氨肽酶(如亮氨酰氨肽酶(EC 3.4.11.1)或三肽氨肽酶(EC 3.4.11.4))。

在另一个优选的实施方案中,特异性作用的蛋白酶是羧肽酶(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)、肽基-二肽酶A(EC 3.4.15.1)或肽基-二肽酶(EC 3.4.15.5)。

按照本领域已知的方法，可制备多肽组合物，且可以是液体或干燥组合物的形式。多肽可用本领域已知的方法稳定化。

本发明也涉及包含用本发明方法获得的蛋白水解产物的食品产品(例如，烘烤产品)。这类食品产品显示了改善的器官感觉质量(如在香味、适口性、口感、香气和外壳颜色方面的改善)。

在本文中，术语“烘烤的产品”包括由生面团(软或脆性的)制备的任何食品。烘烤的产品的例子，不论白的、亮或暗的类型，可由本发明便利地产生的，是面包，尤其是白的、全麦或黑麦面包，典型地，是条或卷、法国baguette型面包、pita面包、tacos、蛋糕、薄煎饼、饼干、脆面包和类似形式的面包。

通常，这类烘烤的产品从包含面粉和水，且典型地是发酵的生面团制备。生面团可以各种方法(如通过加入小苏打或类似物或通过加入曲(发酵的生面团)发酵，但优选地，生面团通过加入适合的酵母培养物(如酿酒酵母(面包酵母)的培养物)发酵。可使用任何市售的酿酒酵母菌株。

此外，用于制备烘烤产品的生面团可是新鲜或冷冻的。冷冻生面团的制备由K.Kulp和K.Lorenz在“冷冻和冷藏的生面团和糊状物”中描述。典型地，本发明的香味改良组合物以0.01-5%，优选地0.1-3%的范围的量包括在生面团中。

在本发明的方法中，把本发明的多肽、内肽酶、转谷氨酰胺酶、肽谷氨酰胺酶、一种或多种特异性和/或非特异性作用的内和/或外肽酶和/或一种或多种以上详细说明书的酶单独或同时加至制备生面团的混合物或制备生面团的任何组分(例如，面粉)中。

本发明还涉及生面团和/或由生面团制备的烘烤产品的预混合物(例如，面粉组合物的形式)，其中预混合物包含本发明的香味改良组合物并选择性地包含一种或多种以上详细说明书的其它酶。

在另一个实施方案中，预混合物包含由本发明的方法获得的水解产物。

预混合物可通过把有关酶与适合的载体(如面粉、淀粉、食糖或盐)混合制备。预混合物可包含其它生面团改良/或面包改良添加剂。

在本文中,术语“预混合物”是烘烤剂的混合物,通常包括面粉,是已制备的能在指定的条件下保存,且在生面团制备过程中为处理提供方便。这类预混合物在工业和商业烘烤面包的工厂和设施以及零售面包店可有便利的应用。

本发明也涉及由本发明的方法产生的水解产物作为食品产品(如烘烤食品)添加剂的利用以改善器官感觉的质量(如香味、适口性和香气)。

由本发明的方法获得的富含游离谷氨酸和/或结合肽的谷氨酸残基的水解产物可用于各种工业应用,尤其是需要功能蛋白质的掺入时。

例如,本发明也涉及包含由本发明的方法获得的富含游离谷氨酸和/或结合肽的谷氨酸残基的水解产物的食品产品和本发明也涉及包含由本发明的方法获得的富含游离谷氨酸和/或结合肽的谷氨酸残基的水解产物的动物饲料添加剂。

本发明进一步通过下列实施例描述,该实施例不应该限制本发明的范围。

实施例

材料

用作缓冲液和底物的化学药品至少是试剂级的市售产品。

实施例1: FLAVOURZYMETM羧肽酶I的纯化

羧肽酶是从FLAVOURZYMETM肉汤(Novo Nordisk A/S, Bagsværd, 丹麦)纯化的。首先,把肉汤(20ml;720mg蛋白质)以180ml的20mM pH7.0磷酸钠缓冲液稀释,并利用装有0.45mM过滤器的Nalgene Filterware过滤。把过滤的溶液装载在用20mM pH7.0的磷酸钠缓冲液预平衡的包含31ml的Q-Sepharose大珠的24 × 130mm柱(“Pharmacia Biotech AB”, Uppsala, 瑞典)上。利用从7.0(20mM磷酸钠缓冲液)至5.0(20mM乙酸钠缓冲液)的pH梯度,然后利用从5.0-3.6(20mM乙酸钠缓冲液)的pH梯度,洗脱羧肽酶。羧肽酶活性在340nm,利用1.5mM 吡喃基丙烯酰丙氨酸作为底物在50mM pH6.0的磷酸钠缓冲液中监测。收集、集中并利用超滤

(Diaflo膜, 10PK)浓缩pH4.3和3.6间的组分至25ml。

浓缩的溶液用100ml 20mM pH7.0的磷酸钠缓冲液稀释, 然后, 装载在用20mM pH7.0的磷酸钠缓冲液预平衡的包含MonoQ小珠的20 × 100mm柱(“Pharmacia Biotech AB”, Uppsala, 瑞典)上。把羧肽酶在pH7.0 20mM的磷酸钠缓冲液中以0-1M NaCl梯度洗脱。如上述, 监测组分的羧肽酶活性。收集、集中并利用对20mM pH4.0乙酸钠缓冲液的超滤浓缩0.10-0.13M NaCl的组分。

基于SDS PAGE分析, 纯化的制剂估测为至少95%是同源的。发现主要带有大约68kDa(66-70kDa范围)的分子量。

实施例2: 蛋白质测序和氨基酸分析的方法

来自实施例1肉汤的纯化的羧肽酶和羧肽酶的降解片段的N端测序在具有联机HPLC和液相三氟醋酸(TFA)分送的应用生物系统476A蛋白质测序仪(Perkin Elmer/Applied Biosystems Division, Foster City, CA)上进行。把纯化的羧肽酶样品转印到SDS-PAGE凝胶的Novex PVDF膜(Novex, San Diego, CA)上, 并利用测序试剂从blott药筒测序(Perkin Elmer/Applied Biosystems Division, Foster City, CA)。苯乙内酰硫脲氨基酸的检测通过利用水中包含3.5%四氢呋喃和15-30ml包含乙酸、乙酸钠和己磺酸钠的预浓缩物(Perkin Elmer/Applied Biosystems Division, Foster City, CA)的缓冲液A和包含乙腈的缓冲液B的联机HPLC完成。收集并利用应用生物系统610数据分析软件, 在Macintosh IIsi上分析数据。

纯化的羧肽酶也易受溴化氰作用产生用于测序的酶的肽片段。把羧肽酶通过在70%甲酸中用几粒溴化氰晶体重建干燥的纯化羧肽酶样品, 并在暗处室温下温育18小时, 以溴化氰消化羧肽酶。利用10-20% Novex Tricine凝胶(Novex, San Diego, CA)通过SDS-PAGE电泳分离肽片段, 并如上述测序。

羧肽酶的N端测序表明: 已阻断N端。降解的片段有下列氨基酸序列, 其中, 括号中的氨基酸残基不是100%确定的, 以?标记的残基不能测定。划线的氨基酸残基与微紫青霉的羧肽酶S1 100%对应:

肽1: ? YGGHYGPAF(F)NH(F)(Y)(E)(Q)(N)E(R)(SEQ ID NO:3)

溴化氰片段的测序显示: 下列肽片段序列与微紫青霉的羧肽酶有30-

40%是同源的(Svendsen等, 1993, FEBS通讯333: 39-43):

肽2(50kDa): DAIGVNI?YTQ?NNDVYYA (SEQ ID NO:4)

肽3(42kDa): DAIGVNI(N)YTQSN(D)VYYAFQQTGDFVWPNFIEDL(SEQ ID NO:5)

肽4(17kDa): (?C)RDNVEGP(?)YAFAGRGVYDIRHPYD(P)(D)(T) (SEQ ID NO:6)

实施例3: 纯化的羧肽酶I的表征

1,10-菲咯啉和苯甲基磺酰氟化物对羧肽酶的抑制作用在pH7.0 20mM磷酸盐缓冲液(其中, 水解在340nm监测)中利用吡喃基丙烯酰丙氨酸赖氨酸作为底物估测。

结果表明: 1,10-菲咯啉不抑制羧肽酶活性。另一方面, 在pH7.0, 利用吡喃基丙烯酰丙氨酸赖氨酸作为底物, 苯甲基磺酰氟化物完全抑制羧肽酶活性。这些结果表明羧肽酶是丝氨酸蛋白酶。

利用N-CBZ-Ala-Ile作为底物, 在50mM乙酸钠缓冲液(pH 4-7.5)和50mM柠檬酸/KH₂PO₄缓冲液(pH 2.9和3.9)中, 25℃下双分子常数K_{cat}/K_m的pH依赖性是在pH 4.0-4.5时, 具有明显最佳值的钟形曲线(图1)。曲线的右和左支斜率分别显示为-1和1, 这间接显示酶活性中心包含两个pK值为3.3和5.7的离子生成基团。

米曲霉ATCC 20386羧肽酶I的活性对温度的依赖性, 在50mM pH4的乙酸钠缓冲液中利用N-CBZ-Ala-Ile作为底物测定。如图2显示的结果表明: 在pH为4, 约55℃-约60℃的温度范围观察到最佳活性。

羧肽酶的底物特异性通过在pH4.0 50mM乙酸盐缓冲液中, 测定表1中列出的底物水解来测定。按照Roth方法, M., 1971, 分析化学 43: 880, 邻苯二醛与二硫苏糖醇一起用于监测在酶促水解条件下, 游离Ile、Glu、Lys、Arg、Asp、Asn、Gly、Phe和Tyr的释放。N-CBZ-Ala-X或N-CBZ-Glu-X水解的K_{cat}值是基于羧肽酶的分子量是68 kD和酶制剂均匀的假设测定的。

已测定羧肽酶具有十分广的底物特异性, 没有观察到对芳香族、庞大脂肪族或极性羧基末端氨基酸残基的优先选择。羧肽酶、N-CBZ-Ala-Glu、N-CBZ-Ala-Ile、N-CBZ-Ala-Lys、N-CBZ-Ala-Arg及N-CBZ-Ala-Asn的最佳底物有十分不同的末端氨基酸残基, 但在pH4.0下, 所有特异性水解的动力学参数值是类似的(表1)。

表1. 羧肽酶的底物特异性

底物	$K_m 10^3 (M)$	$K_{cat} (s^{-1})$	$K_{cat}/K_m 10^3 (M^{-1}s^{-1})$
N-CBZ-Ala-Ile	0.11	24	218
N-CBZ-Ala-Glu	0.10	32	320
N-CBZ-Ala-Lys	0.12	32	270
N-CBZ-Ala-Arg	0.11	27	246
N-CBZ-Ala-Asp	0.27	27	100
N-CBZ-Ala-Asn	0.09	19	216
N-CBZ-Ala-Gly	0.92	5.5	6
N-CBZ-Ala-Phe	0.24	17	72
N-CBZ-Ala-Tyr	0.09	17.5	194
N-CBZ-Glu-Tyr	0.11	4.7	43

羧肽酶也能够从N-CBZ-Ala-Pro切割Pro。然而，对这种底物，酶的特异性是低的。与很高浓度($4.9 \times 10^{-7} M$)的羧肽酶温育2小时后，仅16%的2mM N-CBZ-Ala-Pro溶液水解。N-CBZ-Ala-Phe是羧肽酶的一种好的底物。 K_{cat} 值比来自米曲霉的酸性羧肽酶O-1和O-2至少高8.5倍(Takeuchi和Ichishima, 1981, 农业生物化学45: 1033, Takeuchi等, 1982, 现代微生物学7: 19)。

羧肽酶对吡喃基丙烯酰丙氨酰赖氨酸的 K_m 值在pH 4.0, 25℃下是0.4mM。

羧肽酶也具有高的热稳定性。在pH 4.0和60℃下，在50mM乙酸盐缓冲液中温育酶10分钟与来自米曲霉的37kDa羧肽酶(在这种温度完全失活)相比导致65%的剩余活性(Azarenkova等, 1976, Biokhimia 41: 20)。在pH 4.0和55℃下，在50mM乙酸盐缓冲液中温育酶10分钟导致74%的剩余活性。

羧肽酶活性通过测定N-CBZ-Ala-Glu的特异性水解速率测定。一个羧肽酶单位(CPDU)是在pH4.5和25℃下，每分钟从0.5mM N-CBZ-Ala-Glu溶液释放1微摩尔谷氨酸的羧肽酶量。按照Roth, 1971, 同上的方法，包含邻苯二醛的试剂与二硫苏糖醇结合用于监测酶促水解下游离谷氨酸的释放。邻

苯二醛试剂是0.1M四硼酸钠中的邻苯二醛(6mM)和二硫苏糖醇(6.5mM)溶液。简言之, 20mM pH4.5的乙酸盐缓冲液中的0.05ml酶溶液与1ml底物温育。以一定的时间间隔, 抽取0.3ml的小份, 并与0.9ml的邻苯二醛试剂混合, 利用1cm的路径长度, 测定产物在A₃₄₀的吸光度。A₃₄₀与温育时间是线性关系时, 酶的特异性活性利用公式: 特异性活性=斜率(每分钟光密度) × 11.35计算。按照这种方法, 纯化羧肽酶的特异性活性测定为每毫克蛋白质11个CPDU, 它等同于每分钟每毫克蛋白质11微摩尔谷氨酸。

实施例4: RNA的分离

米曲霉菌株ATCC 20386在每升由7.5g土豆淀粉、10g大豆粗粉、2g KH₂PO₄、5g Na₂HPO₄·2H₂O和0.1g ZnSO₄·7H₂O组成的培养基中, 在发酵罐中培养。在30℃下生长五天后, 取两升样品, 并收集菌丝体, 在液态N₂中冷冻, 并在-80℃保存。全部RNA从冷冻的粉末状米曲霉ATCC 20386菌丝体通过由硫氰酸胍萃取, 其后通过穿过5.7M氯化铯超速离心制备(Chirgwin等, 1979, 生物化学18: 5294-5299)。按照Aviv和Leder(1972, 美国国家科学院论文集69: 1408-1412), 通过寡(dT)-纤维素亲和性色谱分离聚(A)+RNA。

实施例5: cDNA文库的构建

除寡(dT)-NotI锚定引物替代寡(dT)12-18引物用于第一个链反应外, 如由Gubler和Hoffman(1983, 基因25: 263-269)和Sambrook等(1989, 分子克隆: 实验室手册, 冷泉港出版社, 冷泉港, 纽约)描述的, 从5μg实施例4的米曲霉ATCC 20386聚(A)+RNA合成双链cDNA。合成后, 把cDNA利用绿豆核酸酶(生命技术, Gaithersburg, MD)处理, 以T4 DNA聚合酶钝化末端(Boehringer Mannheim, Indianapolis, IN), 并利用约50倍摩尔过量的连接物, 连接到非复发的BstXI连接物(Invitrogen, San Diego, CA)上。适应的cDNA以NotI消化, 通过琼脂糖凝胶电泳进行1.2-3.0kb的cDNA的大小分级分离, 并连接进入BstXI/NotI切开的pYES2.0载体(Invitrogen, San Diego, CA)。按照制造厂商的指导, 连接混合物用于转化电感受态大肠杆菌DH10B细胞(生命技术, Gaithersburg, MD)。把由 1×10^6 独立克隆组成的文库作为单个库(25,000-30,000集落形成单位/库)在-80℃下在20%甘油中

保存，并作为双链cDNA和连接混合物在-20℃下保存。

实施例6：米曲霉ATCC 20386羧肽酶I的PCR扩增

基于在实施例2中描述的，米曲霉ATCC 20386羧肽酶I部分肽的氨基酸序列，按照制造厂商的指导，以应用的生物系统394型DNA/RNA合成仪，合成以下显示的简并寡核苷酸引物以用于PCR扩增米曲霉ATCC 20386羧肽酶I的基因片段：

正向引物：5'-TAYGGNGGICAYTAYGGICCNG-3' (SEQ ID NO:7)

反向引物：5'-ATRAARTTIGGCCAIACRAARTC-3' (SEQ ID NO: 8)

(R=A或G, Y=C或T, N=G或A或C或T, I=肌苷)

利用约1μg从实施例5的米曲霉ATCC 20386 cDNA文库分离的质粒作为模板制备扩增反应物(100μl)。每种反应物含有下列组分：1μg质粒、40pmol正向引物、40pmol反向引物、各200μM dATP、dCTP、dGTP和dTTP、1×Taq聚合酶缓冲液(Perkin-Elmer公司, Branchburg, NJ)和2.5个单位Taq聚合酶(Perkin-Elmer公司, Branchburg, NJ)。把反应物在Perkin-Elmer480型热循环仪中温育：循环设计如下：循环1 -95℃下5分钟，45℃下2分钟及67℃下2分钟；循环2-30 -95℃下2分钟；45℃下一分钟及67℃下2分钟。反应产物在1%琼脂糖凝胶(Eastman Kodak, Rochester, NY)上分离。按照制造厂商的指导，把550bp产物带从凝胶中切除，并利用GenElute自旋柱(Supelco, Bellefonte, PA)纯化。其后，把纯化的PCR产物克隆进入pCRII载体(Invitrogen, San Diego, CA)，并利用lac正向和反向引物测定DNA序列(新英格兰BioLabs, Beverly, MA)。

由约186个密码子组成的羧肽酶I基因节段(550bp)如图1显示用以上所述的羧肽酶特异性PCR引物从米曲霉ATCC 20386扩增。DNA序列分析显示：扩增的基因节段编码部分相应的米曲霉ATCC 20386羧肽酶I基因。羧肽酶I基因节段用于探测米曲霉ATCC 20386的cDNA文库。

实施例7：羧肽酶I克隆的鉴定

把米曲霉ATCC 20386的cDNA文库铺在以50mg/ml羧苄青霉素补充的Luria琼脂平板上。集落转移(Maniatis等, 1982, 分子克隆, 实验室手册, 冷泉港出版社, 冷泉港, 纽约)在约5,000个集落上进行，把DNA利用UV Stratalinker(Stratagene, La Jolla, CA)交联到膜(Hybond N+, Amersham,

Arlington Heights, IL)上。把膜在45℃下,在包含5×SSPE、0.3%SDS、50%甲酰胺以及10mg/ml变性且剪切的鲑精子DNA的杂交溶液中,浸三小时。如实施例1描述的,从米曲霉ATCC 20386分离的羧肽酶I基因片段利用随机引物DNA标记试剂盒(Boehringer Mannheim, Mannheim,德国)进行放射性标记,通过加入NaOH至0.1M的终浓度变性,并以每毫升杂交溶液约 1×10^6 cpm的活性,加至杂交溶液中。混合物在45℃下,以摇动水浴温育过夜。温育后,把膜在55℃在2×SSC中以0.2%SDS洗涤一次,接着,在2×SSC中,在相同温度下洗涤2次。把膜在印迹纸上干燥15分钟,以SaranWrapTM包裹,并在-70℃曝露于具有增感屏的X-射线胶卷(Kodak, Rochester, NY)。

六个集落(指定的大肠杆菌DH5a克隆EJG12、EJG12A、EJC12B、EJG12C、EJG12D以及EJG12E)与探针产生强的杂交信号。把六个集落接种到3ml的LB+50μg/ml羧苄青霉素培养基中,并在37℃下生长过夜。小量制剂DNA利用Wizard 373 DNA纯化试剂盒(Promega, Madison, WI)从这些克隆的每个制备。编码羧肽酶的质粒通过DNA测序确证。

实施例8: 米曲霉ATCC 20386羧肽酶I基因的DNA序列分析

实施例2中描述的羧肽酶I克隆(大肠杆菌DH5a EJG12)的DNA测序以应用生物系统373A型自动化DNA测序仪(应用生物系统公司, Foster City, CA),利用引物步行技术,以染料终止化学(Giesecke等, 1992, 病毒学方法杂志38: 47-60)在双链上进行。寡核苷酸测序引物设计为羧肽酶I基因的互补序列,并按照制造厂商的指导,在应用生物系统394型 DNA/RNA合成仪上合成。

编码米曲霉ATCC 20386羧肽酶I cDNA的基因的核苷酸序列在图3(SEQ ID NO:1)显示。克隆的插入序列分析显示了编码555个氨基酸的蛋白质的1665个核苷酸的大的开放读框(终止密码子除外)。这种开放读框的G+C含量是52.1%。基于van Heijne规则(van Heijne, 1984, 分子生物学杂志173: 243-251),头18个氨基酸可能包含分泌信号肽,此肽指导新生多肽进入内质网组织(在图3中画双线)。

如图3(SEQ ID NO:2)显示的推断的米曲霉ATCC 20386羧肽酶I的氨基酸序列表明: 计算的初级翻译产物的分子量是61.2kDa, 这与基于SDS-

PAGE上纯化的蛋白质迁移率的估计值68kDa一致。如实施例2描述的,由纯化的羧肽酶I获得的肽的氨基酸序列在图3中是划线的,且与推断的米曲霉ATCC 20386羧肽酶I cDNA的氨基酸序列中发现的一致。

利用Clustal序列对比程序(Higgins, 1989, CABIOS 5:151-153)比较推断的米曲霉ATCC 20386羧肽酶I的氨基酸序列与文献中所报导的其它真菌羧肽酶的氨基酸序列,如图4显示,与微紫青霉羧肽酶S1相比,观察到38.8%的等同性(Svendsen等, 1993, 同上)(SEQ ID NO:9),相对于微紫青霉羧肽酶S3,有13.3%的等同性(Svendsen和Day, 1995, FEBS通讯371:1-3)(SEQ ID NO:10),相对于*Aspergillus phoenicis*羧肽酶有16.4%等同性(Chiba等, 1995, 生物化学杂志308:405-409)(SEQ ID NO:11),相对于黑曲霉羧肽酶有14.7%的等同性(van den Hombergh等, 1994, 基因151: 73-79)(SEQ ID NO:12)。推断的米曲霉羧肽酶I的氨基酸序列具有包含丝氨酸羧肽酶中保守的Asp-His-Ser(图4的星标记)的催化三联体。

实施例9: 利用其它真菌的基因组DNA的交叉杂交研究

克隆的米曲霉ATCC 20386羧肽酶基因在与各种真菌属的基因组DNA样品的Southern杂交实验中用作探针。在低严格性(45℃下, 25%甲酰胺, 5 × SSPE, 0.3%SDS)、中等严格性(45℃下, 35%甲酰胺, 5 × SSPE, 0.3%SDS)和高严格性(45℃下, 50%甲酰胺, 5 × SSPE, 0.3%SDS)的条件下探测Southern印迹。基因组DNA样品利用以下概述的方案从下列物种中分离: 黑曲霉(Bo-95)、米曲霉(A1560)、产紫青霉(A3191)、红色青霉(CBS 433.62)、灰腐质霉变种(*Humicola grisea* var.)*themoidea*(ATCC 16453)、灰葡萄孢(ATCC 11542)、*Curvularia verruculosa*(CBS 147.63)、立枯丝核菌(IMI 358730)、*Trichoderma harzianum*(CBS 819.68)、*Absidia griseola*(ATCC 22618)、疣孢漆斑菌(ATCC 9095)、*Myceliophthora thermophila*(CBS 117.65)和变异青霉(CBS 386.48)。

把上述的每种菌株在32℃和250rpm下,在25ml的0.5%酵母提取物-2%葡萄糖(YEG)培养基中生长24小时。然后,通过过Miracloth(Calbiochem, La Jolla, CA)的过滤,收集菌丝体,并以25ml 10mM Tris-1mM EDTA(TE)缓冲液洗涤一次。由菌丝体排出过量缓冲液,其后这些菌丝体在液氮中冷冻。冷冻的菌丝体在电咖啡研磨机中磨成细粉,并把粉加至一次性塑料离

心管中20ml的TE缓冲液和5ml的20%w/v十二烷基硫酸钠(SDS)中。把混合物轻轻地倒转几次以确保混合,并以等体积的苯酚:氯仿:异戊醇(25:24:1 v/v/v)提取两次。加入乙酸钠(3M溶液)至0.3M的终浓度,其后加入2.5体积的冰冷乙醇以沉淀核酸。然后,通过以 $15,000 \times g$ 离心试管30分钟,使核酸沉淀为粒状。使颗粒重新悬浮于0.5m TE缓冲液中前,在空气中干燥30分钟。加入无DNA酶的核糖核酸酶A至100mg/ml的浓度,并把混合物在37℃下温育30分钟。然后,把蛋白酶K以200mg/ml的浓度加入,并把混合物在37℃下,再温育1小时。最后,如前所述,在用乙酸钠和乙醇沉淀DNA前,把混合物以苯酚:氯仿:异戊醇(25:24:1 v/v/v)提取两次。把DNA颗粒在真空下干燥,重新悬浮于TE缓冲液,在4℃下保存直到再次利用。

把每一种DNA样品(约5 μ g)在1%琼脂糖凝胶上电泳前,以BamHI和EcoRI消化。把DNA印迹到Zeta-Probe尼龙膜(BioRad实验室, Hercules, CA)上,并以包含羧肽酶I基因的、翻译切口的DNA探针探测。把印迹在45℃下,在 $2 \times \text{SSC} + 0.2\% \text{SDS}$ 中洗涤30分钟,并在 $2 \times \text{SSC}$ 中(没有SDS)洗涤30分钟。把洗涤的滤器在-70℃下,曝露在具有增感屏的X-射线胶卷下过夜。

把来自米曲霉ATCC 20286的羧肽酶基因与几个其它真菌物种的可探测的羧肽酶基因序列互交(表2)。在低严格条件下,强杂交信号在来自米曲霉A1560、黑曲霉和灰葡萄孢的DNA中是明显的。在*Curvularia verruculosa*、灰腐质霉变种(*Humicola grisea* var.)*thermoidea*、疣孢漆斑菌和*Trichoderma harzianum*中检测到较弱的信号。在*Absidia griseola*、*Myceliophthora thermophila*、产紫青霉、红色青霉、变异青霉或立枯丝核菌中未检测到任何杂交。利用中等严格条件,仅与米曲霉A1560和灰葡萄孢杂交的强信号是可见的。观察到与来自黑曲霉、*Curvularia verruculosa*、灰腐质霉变种(*Humicola grisea* var.)*thermoidea*及疣孢漆斑菌的DNA的弱杂交。在高严格条件下,仅米曲霉A1560的DNA与羧肽酶基因探针杂交。这些数据表明:米曲霉ATCC 20286羧肽酶基因可用作探针克隆其它丝状真菌的羧肽酶基因。

表2. 来自各种真菌的基因组DNA样品与克隆的米曲霉ATCC 20286羧肽酶基因杂交探测。A+++表示强阳性杂交信号, +表示弱信号, -表示没有

可检测到的杂交。

基因组DNA源	低严格性	中等严格性	高严格性
<i>Absidia griseola</i>	-	-	-
黑曲霉	+++	+	-
米曲霉A1560	+++	+++	+++
灰葡萄孢	+++	+++	-
<i>Curvularia verruculosa</i>	+	+	-
<i>Humicola grisea var.thermoidea</i>	+	+	-
<i>Myceliophthora thermophia</i>	-	-	-
疣孢漆斑菌	+	+	-
<i>Penicillium purpurogenum</i>	-	-	-
红色青霉	-	-	-
变异青霉	-	-	-
立枯丝核菌	-	-	-
<i>Trichoderma harzianum</i>	+	-	-

实施例10：由具有羧肽酶活性的多肽的蛋白质水解

商品级的大豆粗粉、小麦麸质、酪蛋白及明胶用作蛋白质底物。大豆粗粉从脱脂的烘烤的去壳大豆20/80/20(Cargill BV, NL)制备；活体小麦麸质得自Crespel和Dieters，丹麦；酪蛋白作为酪蛋白酸钠Miprodan(MD食品，丹麦)获得；明胶作为骨明胶获得(DGF商店，丹麦)。

底物作为2%(w/v)水溶液制备，其中蛋白质浓度用Kjeldahl方法测定。以干燥组合物形式获得的各200mg的每种底物，溶解在9ml自来水中，并把pH调整为7.0。加入包含FLAVOURZYME™(3LAPU)，羧肽酶(11.5CPDU)，或这两种酶的组合的1ml酶溶液。把反应混合物在50℃下温育18小时。然后把酶通过加热到85℃达3分钟灭活。

把水解产物冷却，并利用OPA(邻苯二醛，Sigma化学公司，St. Louis, MO)作为检测试剂，在微滴定板上测定水解度(DH)。如Adler-Nissen(1986，食品蛋白质的酶促水解，Elsevier应用科学出版社)所述定义的DH通过上清

液与OPA(邻苯二醛, Sigma化学公司, St. Louis, MO)的反应测定。对于OPA试剂, 把160mg OPA溶解在4ml乙醇中, 并转移到包含7.62g四硼酸钠十水合物、200mg十二烷基硫酸钠和176mg二硫苏糖醇溶液的200ml容量瓶中, 把烧瓶用水充满到200ml。

把25 μ l体积适当稀释的上清液与200 μ l的OPA试剂在微滴定板孔中混合, 并使得在25℃下恰好反应2分钟。在微滴定板读数仪上测定340nm处的吸光度, 并与减少空白值(水与OPA试剂反应)后, 95mM的L-丝氨酸标准溶液的吸光度对比。为测定真实的DH, 在上清液中测定的丝氨酸当量以由Adler-Nissen建议的用于三硝基苯磺酸法(Adler-Nissen, 1979, 农业和食品化学17: 1256)(此法与所述的OPA方法给出相同的反应)的因子校正。水解度基于水解混合物中蛋白质总量(不是基于可溶性蛋白质)计算。

以下表3总结的结果表明: 对于所有底物, 羧肽酶单独没有影响, 但当加入FLAVOURZYMETM时, 提高了由FLAVOURZYMETM获得的DH。

表 3

底 物	酪蛋白	明胶	大豆粗粉	小麦麸质
酶	DH	DH	DH	DH
FLAVOURZYME	44.8	20.2	36	41
羧肽酶	0	0	0	0
Flavourzyme+	50.2	22.9	40	45.6
羧肽酶				

实施例11: 增加的蛋白质溶解性和通过脱酰胺作用的谷氨酸的释放

小麦麸质(WG)得自Cargill(JOB 5141), 脱去酰氨基的小麦麸质(DWG)得自StaPro Consultancy B.V., Lemdijk 32, 9422 TH Smilde, NL。8%的蛋白质悬浮液通过把11g麸质与89g水混合制备。把pH以NaOH调整至6.5。把如在WO 91/13554中描述的可获得的谷氨酸/天冬氨酸特异性蛋白酶(SP446)或如在WO 89/06270中描述的可获得的赖氨酸/精氨酸特异性蛋白酶(SP387)加至悬浮液中。剂量是: 对于SP446为0.01AU/g蛋白质, 对于SP387为0.006AU/g蛋白质。把FLAVOURZYMETM以20LAPU/g蛋白质的

剂量加至一些水解产物中。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-对硝基酰基苯胺的酶量,此条件为40℃下,26mM的L-亮氨酸-对硝基酰基苯胺在pH 8.0的0.1M Tris 缓冲液中反应10分钟。当水解时,释放对硝基酰基苯胺,使溶液变为在405nm下监测的黄色。

无须再调整pH,使水解在50℃下进行18小时。通过在85℃下加热15分钟使酶失活。把pH调整到5,并离心水解产物。测定上清液中蛋白质和游离谷氨酸的含量。

通过Kjeldahl分析,利用6.25的Kjeldahl 因子,测定蛋白质的含量。

按照制造厂商的指导(Boehringer-Mannheim, Indianapolis IN),通过利用谷氨酸测定试剂盒,测定游离谷氨酸的含量。此方法适用于微量滴定板中。

比较小麦麸质(WG)和脱去酰氨基的小麦麸质(DWG),表4显示的结果表明:脱酰胺作用增加了麸质对特异性蛋白酶的敏感性,因此,更多蛋白质变得可溶。通过加入具有特异性蛋白酶的FLAVOURZYMETM,由于脱酰胺作用,谷氨酸的释放增加了一倍。

表 4

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

实施例12: 酶促脱酰胺作用和谷氨酸的释放

肽谷氨酰胺酶II通过在30℃下在摇瓶(400ml)中以270rpm使环状芽孢杆菌菌株ATCC 21590混合生长20小时,此瓶包含200ml由1%多肽、0.5%乳糖、0.025%MgSO₄·7H₂O、0.005%FeSO₄·7H₂O、0.025%KH₂PO₄及17%Na₂HPO₄·12H₂O组成的培养基(pH调整到7.2)。把细胞通过以

4000rpm离心收集在1升烧瓶中。然后冷冻细胞。

来自环状芽孢杆菌的肽谷氨酰胺酶II的纯化在室温下进行。使冷冻的环状芽孢杆菌细胞解冻，并悬浮在裂解缓冲液(50mM Tris/HCl; 25%蔗糖(w/v); 1mM EDTA, pH8.0)中，直到获得每升裂解缓冲液100g湿细胞的均相悬浮液。把溶菌酶(10mg/ml)和DNA酶I(Sigma DN-25, 10mg/ml)溶解在裂解缓冲液中。然后，以每升100ml溶菌酶溶液、10ml 1.0M的MgCl₂及1ml DNA酶I溶液加至细胞悬浮液中。使得酶反应1小时。

悬浮液通过Seitz深度过滤平板过滤，并把滤液转移到葡聚糖凝胶G25柱(Pharmacia)上pH8.0的10mM KH₂PO₄/NaOH缓冲液A中。把酶溶液应用于以缓冲液A平衡的SOURCE Q柱(Pharmacia)，并以缓冲液A中线性NaCl梯度(0→500mM)洗脱。如下述，分析柱上组分的肽谷氨酰胺酶II的活性，并集中有活性的组分。集中的组分在280nm的吸光度是1.78，因此，蛋白质含量估计为1.8mg/ml。

肽谷氨酰胺酶II库中蛋白质的纯度如由SDS-PAGE凝胶判断的大约是25%。因此，制剂包含大约0.5mg/ml纯化的肽谷氨酰胺酶II。

肽谷氨酰胺酶的活性通过利用Boehringer-Mannheim氨测定试剂盒(Cat.No. 1112732)测定N-叔丁氧羰基-Gln-Pro(N-t-BOC-Gln-Pro; SIGMA No.B-4403)的γ-甲酰胺的水解期间形成的氨测定。在这种试剂盒中，通过测定由谷氨酸脱氢酶消耗的NADH测定氨，也应用未加入N-t-BOC-Gln-Pro的空白以减去其它NADH消耗酶的影响。

把共200mg的小麦麸质蛋白质加至9ml沸水中，冷却后，把pH调整到7.0。然后，如上述的，把250μl的肽谷氨酰胺酶II制剂(PEP)加入。把实施例11中所述的谷氨酸/天冬氨酸特异性蛋白酶(SP446)以0.04AU/g蛋白质的量加入，实施例11中描述的FLAVOURZYMETM以20LAPU/g蛋白质的量加入。

使水解在50℃下，无须调整pH，进行18小时。未加入肽谷氨酰胺酶的对照也进行。离心水解产物，并如实施例11中描述的，测定谷氨酸。

如实施例10中描述的，测定小麦麸质蛋白质的DH。

如以下表5中显示的结果表明：以肽谷氨酰胺酶制剂的水解增加了DH以

及谷氨酸的释放。

表 5

水解	DH%	谷氨酸(mg/l)
缺PEP	40	131
加PEP	43	171

实施例13：大豆水解产物的香味改良

大豆水解产物利用8%的大豆粗粉蛋白质浓度(1升或0.2升)在50℃下进行18小时制备，其中，没有以ALCALASE™2.4L(Novo Nordisk Bagsverd, 丹麦)调整pH，如实施例1中所述的，制备纯化的羧肽酶溶液(大约40CPDU/ml)。剂量是每升1200mg ALCALASE™或每升1200mg ALCALASE™及5ml羧肽酶。

然后，使酶在93℃下灭活15分钟，如实施例10和11中所述的，分别分析DH和谷氨酸。在125℃下成熟30分钟前，把pH调整到5。成熟后，把所有的水解产物离心。

所产生的大豆粗粉水解产物通过一连串香味实验对象评价。把上清液稀释5.3倍，并以随机顺序给十一位评委献两次。评委在0-9的等级上估价下列香味的属性：苦味、味美、肉汤、肉、小鸡、烟及MSG。

DH结果和不同的香味属性的平均得分在表6中显示。

表 6

处理	DH	苦味	味美	肉汤	肉	蔬菜	小鸡	烟	MSG
碱酶	17.3	4.45	1.68	1.95	1.14	1.86	1.82	2.77	1.41
碱酶 +CPD	33.0	2.77	2.18	2.59	1.41	2.23	2.36	3.14	1.86

苦味结果显示：把羧肽酶加至ALCALASE™明显地降低苦味。关于全部的味美性，通过加羧肽酶至ALCALASE™，看到有改善。

肉汤、小鸡及可能的MSG香味也通过把羧肽酶加至ALCALASE™得以改善。肉、烟及蔬菜香味也通过加入羧肽酶改善。

生物材料的保藏

下列生物材料已根据布达佩斯条约保藏在农业研究服务机构专利培养

物保藏中心，北方地区研究中心，学院街1815，Peoria，伊利诺州，61604，
且给出下列保藏号：

保藏物	保藏号	保藏日期
大肠杆菌DH5a pEJG12	NRRL B-21616	1996年8月28日

序 列 表

(1) 一般信息:

(i) 申请人:

(A) 名称: Novo Nordisk Biotech公司

(B) 街道: 1445 Drew Avenue

(C) 城市: Davis, california

(D) 国家: 美利坚合众国

(E) 邮政编码(ZIP): 95616-4880

(F) 电话: (916)757-8100

(G) 传真: (916)758-0317

(i) 申请人:

(A) 名称: Novo Nordisk A/S

(B) 街道: Novo Alle

(C) 城市: Bagsværd

(D) 国家: 丹麦

(E) 邮政编码(ZIP): DK-2880

(F) 电话: +45 4444 8888

(G) 传真: +45 4449 3256

(ii) 发明名称: 羧肽酶和编码该酶的核酸

(iii) 序列数: 12

(iv) 通信地址:

(A) 收信人: 北美Novo Nordisk公司

(B) 街道: Lexington路405

(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: 软盘

(B) 计算机: IBM兼容机

(C) 操作系统: DOS

(D) 软件: Windows FastSEQ 2.0版

(vi) 当前申请的数据:

(A) 申请号: 待指定

(B) 申请日: 1997年10月3日

(C) 分类:

(viii) 律师/代理人信息:

(A) 姓名: Lambiris, Elias J

(B) 登记号: 33,728

(C) 参考/证书号: 4990.204-WO

(ix) 电讯信息:

(A) 电话: 212-867-0123

(B) 传真: 212-878-9655

(C) 电传:

(2) SEQ ID NO: 1的信息:

(i) 序列特征:

(A) 长度: 1662个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线型

(ii) 分子类型: 基因组DNA

(ix) 特征:

(A) 名称/关键词: 编码序列

(B) 位置: 1...1662

(D) 其它信息:

(xi) 序列描述: SEQ ID NO:1:

ATG CGT GGC TAC GAA TTT CTC TCA GTG CTA CCC TTG GTT GCA GCC AGT Met Arg Gly Tyr Glu Phe Leu Ser Val Leu Pro Leu Val Ala Ala Ser 1 5 10 15	48
TGG GCC CTT CCA GGA AGT ACA CCG GCG TCC GTC GGT AGA AGA CAG CTA Trp Ala Leu Pro Gly Ser Thr Pro Ala Ser Val Gly Arg Arg Gln Leu 20 25 30	96
CCC AAG AAC CCC ACC GGG GTC AAG ACT CTT ACA ACC GCA AAC AAT GTC Pro Lys Asn Pro Thr Gly Val Lys Thr Leu Thr Thr Ala Asn Asn Val 35 40 45	144
ACC ATC CGG TAC AAG GAA CCC GGG GCA GAG GGC GTC TGC GAG ACT ACC Thr Ile Arg Tyr Lys Glu Pro Gly Ala Glu Gly Val Cys Glu Thr Thr 50 55 60	192
CCG GGT GTC AAA TCC TAC TCT GGA TAT GTC GAC ACC TCT CCC GAG TCC Pro Gly Val Lys Ser Tyr Ser Gly Tyr Val Asp Thr Ser Pro Glu Ser 65 70 75 80	240
CAT ACC TTC TTC TGG TTC TTC GAA GCC AGA CAT AAC CCA GAA ACT GCA His Thr Phe Phe Trp Phe Phe Glu Ala Arg His Asn Pro Glu Thr Ala 85 90 95	288
CCT ATC ACA TTG TGG TTG AAT GGT GGC CCT GGA AGC GAT TCT TTG ATC Pro Ile Thr Leu Trp Leu Asn Gly Gly Pro Gly Ser Asp Ser Leu Ile 100 105 110	336
GGT CTC TTC GAA GAG TTG GGC CCT TGC CAT GTC AAT TCG ACT TTT GAT Gly Leu Phe Glu Glu Leu Gly Pro Cys His Val Asn Ser Thr Phe Asp 115 120 125	384
GAC TAC ATC AAC CCT CAC TCG TGG AAC GAG GTC TCC AAT TTA CTA TTC Asp Tyr Ile Asn Pro His Ser Trp Asn Glu Val Ser Asn Leu Leu Phe 130 135 140	432
CTG TCC CAG CCA TTG GGA GTC GGC TTT TCA TAT AGT GAT ACG GTT GAT Leu Ser Gln Pro Leu Gly Val Gly Phe Ser Tyr Ser Asp Thr Val Asp 145 150 155 160	480
GGG TCC ATT AAC CCT GTA ACT GGG GTC GTC GAA AAT TCG AGC TTT GCA Gly Ser Ile Asn Pro Val Thr Gly Val Val Glu Asn Ser Ser Phe Ala 165 170 175	528
GGA GTT CAG GGC CGG TAC CCA ACC ATT GAT GCC ACT CTG ATC GAT ACT Gly Val Gln Gly Arg Tyr Pro Thr Ile Asp Ala Thr Leu Ile Asp Thr 180 185 190	576
ACC AAT CTT GCC GCA GAG GCC GCT TGG GAG ATC CTG CAA GGA TTC CTT Thr Asn Leu Ala Ala Glu Ala Ala Trp Glu Ile Leu Gln Gly Phe Leu 195 200 205	624
AGT GGA CTA CCT AGC TTG GAC TCT AGG GTG CAG TCT AAG GAC TTC AGT Ser Gly Leu Pro Ser Leu Asp Ser Arg Val Gln Ser Lys Asp Phe Ser 210 215 220	672
CTA TGG ACG GAG AGC TAT GGA GGG CAC TAT GGT CCT GCA TTC TTC AAT Leu Trp Thr Glu Ser Tyr Gly Gly His Tyr Gly Pro Ala Phe Phe Asn 225 230 235 240	720

CAT TTT TAC GAG CAG AAT GAG AGA ATT GCC AAC GGT AGT GTT AAT GGT	768
His Phe Tyr Glu Gln Asn Glu Arg Ile Ala Asn Gly Ser Val Asn Gly	
245 250 255	
GTT CAG CTT AAT TTC AAC TCT CTG GGA ATT ATT AAC GGC ATC ATC GAC	816
Val Gln Leu Asn Phe Asn Ser Leu Gly Ile Ile Asn Gly Ile Ile Asp	
260 265 270	
GAG GCG ATC CAG GCC CCT TAC TAC CCT GAA TTC GCT GTG AAC AAT ACC	864
Glu Ala Ile Gln Ala Pro Tyr Tyr Pro Glu Phe Ala Val Asn Asn Thr	
275 280 285	
TAC GGT ATC AAG GCT GTC AAC GAG ACC GTC TAC AAC TAC ATG AAG TTT	912
Tyr Gly Ile Lys Ala Val Asn Glu Thr Val Tyr Asn Tyr Met Lys Phe	
290 295 300	
GCC AAC CAA ATG CCA AAT GGT TGC CAG GAT TTG ATT TCC ACC TGC AAA	960
Ala Asn Gln Met Pro Asn Gly Cys Gln Asp Leu Ile Ser Thr Cys Lys	
305 310 315 320	
CAG ACA AAC CGC ACC GCA TTA GCT GAC TAC GCC CTC TGC GCC GAA GCC	1008
Gln Thr Asn Arg Thr Ala Leu Ala Asp Tyr Ala Leu Cys Ala Glu Ala	
325 330 335	
ACC AAC ATG TGC AGG GAC AAT GTT GAG GGG CCA TAC TAC GCC TTT GCT	1056
Thr Asn Met Cys Arg Asp Asn Val Glu Gly Pro Tyr Tyr Ala Phe Ala	
340 345 350	
GGT CGT GGT GTG TAT GAT ATT CGG CAT CCA TAT GAT GAC CCG ACT CCG	1104
Gly Arg Gly Val Tyr Asp Ile Arg His Pro Tyr Asp Asp Pro Thr Pro	
355 360 365	
CCA AGT TAT TAC AAC AAA TTT CTG GCA AAG GAC TCT GTC ATG GAC GCT	1152
Pro Ser Tyr Tyr Asn Lys Phe Leu Ala Lys Asp Ser Val Met Asp Ala	
370 375 380	
ATC GGC GTC AAC ATC AAC TAC ACC CAG TCC AAT AAT GAC GTC TAC TAC	1200
Ile Gly Val Asn Ile Asn Tyr Thr Gln Ser Asn Asn Asp Val Tyr Tyr	
385 390 395 400	
GCT TTC CAG CAA ACA GGC GAC TTT GTC TGG CCC AAC TTC ATC GAA GAC	1248
Ala Phe Gln Gln Thr Gly Asp Phe Val Trp Pro Asn Phe Ile Glu Asp	
405 410 415	
CTC GAG GAG ATC CTT GCT CTC CCC GTG CGT GTC TCC CTC ATC TAT GGC	1296
Leu Glu Glu Ile Leu Ala Leu Pro Val Arg Val Ser Leu Ile Tyr Gly	
420 425 430	
GAC GCC GAT TAC ATC TGC AAC TGG TTC GGC GGT CAG GCC GTT TCC CTC	1344
Asp Ala Asp Tyr Ile Cys Asn Trp Phe Gly Gly Gln Ala Val Ser Leu	
435 440 445	
GCT GCG AAC TAC TCC CAA GCC GCC CAG TTC CGA AGC GCA GGG TAC ACG	1392
Ala Ala Asn Tyr Ser Gln Ala Ala Gln Phe Arg Ser Ala Gly Tyr Thr	
450 455 460	
CCC CTG AAA GTC AAC GGC GTC GAG TAT GGG GAA ACT CGC GAG TAT GGT	1440
Pro Leu Lys Val Asn Gly Val Glu Tyr Gly Glu Thr Arg Glu Tyr Gly	
465 470 475 480	
AAT TTC TCC TTC ACT CGC GTC TAT GAG GCA GGC CAT GAA GTC CCA TAC	1488
Asn Phe Ser Phe Thr Arg Val Tyr Glu Ala Gly His Glu Val Pro Tyr	
485 490 495	
TAC CAG CCC ATC GCC TCC CTG CAA TTG TTT AAC CGG ACT ATC TTC GGT	1536
Tyr Gln Pro Ile Ala Ser Leu Gln Leu Phe Asn Arg Thr Ile Phe Gly	
500 505 510	

```

TGG GAT ATC GCA GAG GGC CAG AAG AAG ATC TGG CCC AGC TAC AAG ACG      1584
Trp Asp Ile Ala Glu Gly Gln Lys Lys Ile Trp Pro Ser Tyr Lys Thr
      515                      520                      525

AAT GGA ACG GCT ACA GCT ACG CAT ACA CAG TCG TCC GTG CCG CTG CCT      1632
Asn Gly Thr Ala Thr Ala Thr His Thr Gln Ser Ser Val Pro Leu Pro
      530                      535                      540

ACG GCT ACC AGC ATG TCC AGT GTT GGT ATG                                1662
Thr Ala Thr Ser Met Ser Ser Val Gly Met
      545                      550

```

(2) SEQ ID NO:2的信息:

(i) 序列特征:

(A) 长度: 554个氨基酸

(B) 类型: 氨基酸

(C) 链型: 单链

(D) 拓扑结构: 线型

(ii) 分子类型: 蛋白质

(v) 片段类型: 内部

(xi) 序列描述: SEQ ID NO:2:

```

Met Arg Gly Tyr Glu Phe Leu Ser Val Leu Pro Leu Val Ala Ala Ser
 1          5          10          15

Trp Ala Leu Pro Gly Ser Thr Pro Ala Ser Val Gly Arg Arg Gln Leu
      20          25          30

Pro Lys Asn Pro Thr Gly Val Lys Thr Leu Thr Thr Ala Asn Asn Val
      35          40          45

Thr Ile Arg Tyr Lys Glu Pro Gly Ala Glu Gly Val Cys Glu Thr Thr
      50          55          60

Pro Gly Val Lys Ser Tyr Ser Gly Tyr Val Asp Thr Ser Pro Glu Ser
      65          70          75          80

His Thr Phe Phe Trp Phe Phe Glu Ala Arg His Asn Pro Glu Thr Ala
      85          90          95

Pro Ile Thr Leu Trp Leu Asn Gly Gly Pro Gly Ser Asp Ser Leu Ile
      100         105         110

Gly Leu Phe Glu Glu Leu Gly Pro Cys His Val Asn Ser Thr Phe Asp
      115         120         125

Asp Tyr Ile Asn Pro His Ser Trp Asn Glu Val Ser Asn Leu Leu Phe
      130         135         140

Leu Ser Gln Pro Leu Gly Val Gly Phe Ser Tyr Ser Asp Thr Val Asp
      145         150         155         160

Gly Ser Ile Asn Pro Val Thr Gly Val Val Glu Asn Ser Ser Phe Ala
      165         170         175

Gly Val Gln Gly Arg Tyr Pro Thr Ile Asp Ala Thr Leu Ile Asp Thr
      180         185         190

```

Thr Asn Leu Ala Ala Glu Ala Ala Trp Glu Ile Leu Gln Gly Phe Leu
 195 200 205
 Ser Gly Leu Pro Ser Leu Asp Ser Arg Val Gln Ser Lys Asp Phe Ser
 210 215 220
 Leu Trp Thr Glu Ser Tyr Gly Gly His Tyr Gly Pro Ala Phe Phe Asn
 225 230 235 240
 His Phe Tyr Glu Gln Asn Glu Arg Ile Ala Asn Gly Ser Val Asn Gly
 245 250 255
 Val Gln Leu Asn Phe Asn Ser Leu Gly Ile Ile Asn Gly Ile Ile Asp
 260 265 270
 Glu Ala Ile Gln Ala Pro Tyr Tyr Pro Glu Phe Ala Val Asn Asn Thr
 275 280 285
 Tyr Gly Ile Lys Ala Val Asn Glu Thr Val Tyr Asn Tyr Met Lys Phe
 290 295 300
 Ala Asn Gln Met Pro Asn Gly Cys Gln Asp Leu Ile Ser Thr Cys Lys
 305 310 315 320
 Gln Thr Asn Arg Thr Ala Leu Ala Asp Tyr Ala Leu Cys Ala Glu Ala
 325 330 335
 Thr Asn Met Cys Arg Asp Asn Val Glu Gly Pro Tyr Tyr Ala Phe Ala
 340 345 350
 Gly Arg Gly Val Tyr Asp Ile Arg His Pro Tyr Asp Asp Pro Thr Pro
 355 360 365
 Pro Ser Tyr Tyr Asn Lys Phe Leu Ala Lys Asp Ser Val Met Asp Ala
 370 375 380
 Ile Gly Val Asn Ile Asn Tyr Thr Gln Ser Asn Asn Asp Val Tyr Tyr
 385 390 395 400
 Ala Phe Gln Gln Thr Gly Asp Phe Val Trp Pro Asn Phe Ile Glu Asp
 405 410 415
 Leu Glu Glu Ile Leu Ala Leu Pro Val Arg Val Ser Leu Ile Tyr Gly
 420 425 430
 Asp Ala Asp Tyr Ile Cys Asn Trp Phe Gly Gly Gln Ala Val Ser Leu
 435 440 445
 Ala Ala Asn Tyr Ser Gln Ala Ala Gln Phe Arg Ser Ala Gly Tyr Thr
 450 455 460
 Pro Leu Lys Val Asn Gly Val Glu Tyr Gly Glu Thr Arg Glu Tyr Gly
 465 470 475 480
 Asn Phe Ser Phe Thr Arg Val Tyr Glu Ala Gly His Glu Val Pro Tyr
 485 490 495
 Tyr Gln Pro Ile Ala Ser Leu Gln Leu Phe Asn Arg Thr Ile Phe Gly
 500 505 510
 Trp Asp Ile Ala Glu Gly Gln Lys Lys Ile Trp Pro Ser Tyr Lys Thr
 515 520 525
 Asn Gly Thr Ala Thr Ala Thr His Thr Gln Ser Ser Val Pro Leu Pro
 530 535 540
 Thr Ala Thr Ser Met Ser Ser Val Gly Met
 545 550

(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: 线型****(ii) 分子类型: 无****(xi) 序列描述: SEQ ID NO:3:**

Xaa	Tyr	Gly	Gly	His	Tyr	Gly	Pro	Ala	Phe	Phe	Asn	His	Phe	Tyr	Glu
1				5					10					15	
Gln	Asn	Glu	Arg												
			20												

(2) SEQ ID NO: 4的信息:**(i) 序列特征:****(A) 长度: 19个氨基酸****(B) 类型: 氨基酸****(C) 链型: 单链****(D) 拓扑结构: 线型****(ii) 分子类型: 无****(xi) 序列描述: SEQ ID NO: 4:**

Asp	Ala	Ile	Gly	Val	Asn	Ile	Xaa	Tyr	Thr	Gln	Xaa	Asn	Asn	Asp	Val
1				5					10					15	
Tyr	Tyr	Ala													

(2) SEQ ID NO:5的信息:**(i) 序列特征:****(A) 长度: 35个氨基酸****(B) 类型: 氨基酸****(C) 链型: 单链****(D) 拓扑结构: 线型****(ii) 分子类型: 无****(xi) 序列描述: SEQ ID NO:5:**

```

Asp Ala Ile Gly Val Asn Ile Asn Tyr Thr Gln Ser Asn Asn Asp Val
 1           5           10           15
Tyr Tyr Ala Phe Gln Gln Thr Gly Asp Phe Val Trp Pro Asn Phe Ile
          20           25           30
Glu Asp Leu
          35

```

(2) SEQ ID NO:6的信息:

(i) 序列特征:

(A) 长度: 29个氨基酸

(B) 类型: 氨基酸

(C) 链型: 单链

(D) 拓扑结构: 线型

(ii) 分子类型: 无

(xi) 序列描述: SEQ ID NO:6

```

Xaa Cys Arg Asp Asn Val Glu Gly Pro Xaa Tyr Ala Phe Ala Gly Arg
 1           5           10           15
Gly Val Tyr Asp Ile Arg His Pro Tyr Asp Pro Asp Thr
          20           25

```

(2) SEQ ID NO:7的信息:

(i) 序列特征:

(A) 长度: 20个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线型

(xi) 序列描述: SEQ ID NO:7:

TAYGGNGGCA YTAYGGCCNG

20

(2) SEQ ID NO:8 的信息:

(i) 序列特征:

(A) 长度: 21个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线型

(ii) 分子类型: cDNA

(xi) 序列描述: SEQ ID NO: 8:

ATRAARTTGG CCAACRAART C

21

(2) SEQ ID NO:9的信息:**(i) 序列特征:****(A) 长度: 423个氨基酸****(B) 类型: 氨基酸****(C) 链型: 单链****(D) 拓扑结构: 线型****(xi) 序列描述: SEQ ID NO:9:**

```

Phe Val Lys Asn Ser Gly Ile Cys Glu Thr Thr Pro Gly Val Asn Gln
 1          5          10          15
Tyr Ser Gly Tyr Leu Ser Val Gly Ser Asn Met Asn Met Trp Phe Trp
          20          25          30
Phe Phe Glu Ala Arg Asn Asn Pro Gln Gln Ala Pro Leu Ala Ala Trp
          35          40          45
Phe Asn Gly Gly Pro Gly Cys Ser Ser Met Ile Gly Leu Phe Gln Glu
 50          55          60
Asn Gly Pro Cys His Phe Val Asn Gly Asp Ser Thr Pro Ser Leu Asn
 65          70          75          80
Glu Asn Ser Trp Asn Asn Tyr Ala Asn Met Ile Tyr Ile Asp Gln Pro
          85          90          95
Ile Gly Val Gly Phe Ser Tyr Gly Thr Asp Asp Val Thr Ser Thr Val
          100          105          110
Thr Ala Ala Pro Tyr Val Trp Asn Leu Leu Gln Ala Phe Tyr Ala Gln
          115          120          125
Arg Pro Glu Tyr Glu Ser Arg Asp Phe Ala Ile Phe Thr Glu Ser Tyr
          130          135          140
Gly Gly His Tyr Gly Pro Glu Phe Ala Ser Tyr Ile Glu Gln Gln Asn
          145          150          155          160
Ala Ala Ile Lys Ala Gly Ser Val Thr Gly Gln Asn Val Asn Ile Val
          165          170          175
Ala Leu Gly Val Asn Asn Gly Trp Ile Asp Ser Thr Ile Gln Glu Lys
          180          185          190
Ala Tyr Ile Asp Phe Ser Tyr Asn Asn Ser Tyr Gln Gln Ile Ile Asp
          195          200          205
Ser Ser Thr Arg Asp Ser Leu Leu Asp Ala Tyr Asn Asn Gln Cys Leu
          210          215          220
Pro Ala Leu Gln Gln Cys Ser Gln Ser Gly Ser Thr Ser Asp Cys Thr
          225          230          235          240
Asn Ala Asp Ser Val Cys Tyr Gln Asn Ile Glu Gly Pro Ile Ser Ser
          245          250          255
Ser Gly Asp Phe Asp Val Tyr Asp Ile Arg Glu Pro Ser Asn Asp Pro
          260          265          270
Tyr Pro Pro Lys Thr Tyr Ser Thr Tyr Leu Ser Asp Pro Thr Val Val
          275          280          285
Lys Ala Ile Gly Ala Arg Thr Asn Tyr Gln Glu Cys Pro Asn Gly Pro
          290          295          300
Tyr Asn Lys Phe Ala Ser Thr Gly Asp Asn Pro Arg Ser Phe Leu Ser
          305          310          315          320
Thr Leu Ser Ser Val Val Gln Ser Gly Ile Asn Val Leu Val Trp Ala
          325          330          335
Gly Asp Ala Asp Trp Ile Cys Asn Trp Leu Gly Asn Tyr Glu Val Ala
          340          345          350
Asn Ala Val Asp Phe Pro Gly Asn Ala Gln Phe Ser Ala Leu Asp Leu
          355          360          365

```

```

Ala Pro Tyr Thr Val Asn Gly Val Glu Lys Gly Gln Phe Lys Thr Val
  370          375          380
Asp Asn Phe Ser Phe Leu Lys Val Tyr Gly Ala Gly His Glu Val Pro
 385          390          395          400
Tyr Tyr Gln Pro Asp Thr Ala Leu Gln Ala Phe Lys Gln Ile Ile Gln
          405          410          415
Lys Lys Pro Ile Ser Ser Thr
          420

```

(2) SEQ ID NO:10 的信息:

(i) 序列特征:

(A) 长度: 481个氨基酸

(B) 类型: 氨基酸

(C) 链型: 单链

(D) 拓扑结构: 线型

(xi) 序列描述: SEQ ID NO:10:

```

Ser Thr Lys Asn Tyr Arg Phe Leu Asn Glu Lys Thr Lys Ala Asn Leu
 1          5          10          15
Val His His Leu Pro Asp Val Pro Tyr Asp Ile Gly Glu Met Tyr Ser
          20          25          30
Gly Leu Met Pro Ile Asp Met His Asn Glu Ser Arg Ala Leu Phe Tyr
          35          40          45
Ile Phe Gln Pro Thr Ile Gly Glu Pro Val Asp Glu Val Thr Ile Trp
          50          55          60
Met Asn Gly Gly Pro Gly Cys Ser Ser Met Glu Ser Phe Leu Gln Glu
          65          70          75          80
Thr Gly Arg Phe Leu Trp Gln Pro Gly Thr Tyr Ala Pro Val Glu Asn
          85          90          95
Pro Tyr Ser Trp Val Val Leu Thr Asn Val Leu Trp Val Asp Gln Pro
          100          105          110
Val Gly Thr Gly Tyr Ser Ile Gly Thr Pro Thr Ala Thr Ser Gln Glu
          115          120          125
Glu Thr Ala Gln Asp Phe Val Lys Phe Phe Lys Asn Phe Gln Lys Thr
          130          135          140
Tyr Gly Ile Lys Asn Phe Lys Ile Tyr Val Thr Gly Glu Ser Tyr Ala
          145          150          155          160
Gly Arg Tyr Val Pro Tyr Ile Ser Ala Ala Met Leu Asp Glu Lys Asp
          165          170          175
Lys Glu Tyr Phe Asp Leu Gln Gly Ala Leu Ala Tyr Asp Pro Cys Ile
          180          185          190
Gly Gln Phe Asp Tyr Val Gln Glu Glu Ile Pro Val Val Pro Phe Val
          195          200          205
Lys Glu Asn Ala Asn Leu Phe Asn Phe Asn Glu Thr Phe Met Ala Glu
          210          215          220
Leu Glu His Leu His Lys Ser Cys Gly Tyr Ala Asp Phe Ile Asp Lys
          225          230          235          240
Tyr Leu Thr Phe Pro Pro Pro Lys Glu Gln Pro Pro Leu Phe Phe Asn
          245          250          255
Tyr Thr Ser Met Ala Asn Glu Asp Val Phe Asp Met Val Tyr Asn Glu
          260          265          270
Val Phe Lys Ile Asn Pro Cys Phe Asp Leu Tyr Glu Val Asn Leu Met
          275          280          285
Cys Pro Leu Gln Trp Asp Val Leu Ala Phe Pro Thr Ser Leu Val Tyr
          290          295          300
Gln Pro Ala Gly Ala Thr Val Tyr Phe Asp Arg Ala Asp Val Lys Lys
          305          310          315          320

```

```

Ala Leu His Ala Pro Asn Val Thr Trp Ala Glu Cys Ser Asn Asn Pro
      325      330      335
Val Phe Val Gly Gly Ser Ser Gly Pro Glu Gln Glu Gly Asp Thr Ser
      340      345      350
Ala Asn Pro Ile Glu His Val Leu Pro Gln Val Ile Glu Ala Thr Asn
      355      360      365
Arg Val Leu Ile Ser Asn Gly Asp Phe Asp Met Val Ile Leu Thr Asn
      370      375      380
Gly Thr Leu Leu Ala Ile Gln Asn Met Thr Trp Asn Gly His Leu Gly
385      390      395      400
Phe Gln Lys Lys Pro Ser Ala Pro Ile Asp Ile Lys Ile Pro Asp Leu
      405      410      415
Gln Tyr Lys Glu Val Phe Ala Glu Asn Gly Ala Ser Ser Leu Asp Gly
      420      425      430
Ala Gln Gly Ile Met Gly Val Gln His Tyr Glu Arg Gly Leu Met Lys
      435      440      445
Ala Gln Thr Tyr Gln Ser Gly His Met Gln Pro Gln Tyr Gln Pro Arg
      450      455      460
Val Ala Tyr Arg His Leu Glu Trp Leu Leu Lys Arg Thr Asp Glu Leu
465      470      475      480
Gln

```

(2) SEQ ID NO:11的信息:

(i) 序列特征:

(A) 长度: 523个氨基酸

(B) 类型: 氨基酸

(C) 链型: 单链

(D) 拓扑结构: 线型

(xi) 序列描述: SEQ ID NO:11:

```

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

Asn Gly Gly Pro Gly Cys Ser Ser Leu Asp Gly Leu Leu Gln Glu Asn
115      120      125
Gly Pro Phe Leu Trp Gln Pro Gly Thr Tyr Lys Pro Val Pro Asn Pro
130      135      140
Tyr Ser Trp Thr Asn Leu Thr Asn Val Val Tyr Ile Asp Gln Pro Ala
145      150      155      160
Gly Thr Gly Phe Ser Pro Gly Pro Ser Thr Val Asn Asp Glu Glu Asp
165      170      175
Val Ala Ala Gln Phe Asn Ser Trp Phe Lys His Phe Val Asp Thr Phe
180      185      190
Asp Leu His Gly Arg Lys Val Tyr Ile Thr Gly Glu Ser Tyr Ala Gly
195      200      205
Met Tyr Val Pro Tyr Ile Ala Asp Ala Met Leu Asn Glu Glu Asp Thr
210      215      220

```

```

Thr Tyr Phe Asn Leu Lys Gly Ile Gln Ile Asn Asp Pro Ser Ile Asn
225                230                235                240
Ser Asp Ser Val Met Met Tyr Ser Pro Ala Val Arg His Leu Asn His
                245                250                255
Tyr Asn Asn Ile Phe Arg Leu Asn Ser Thr Phe Leu Ser Tyr Ile Asn
                260                265                270
Gly Lys Ala Asp Lys Cys Gly Tyr Asn Ala Phe Leu Asp Lys Ala Ile
                275                280                285
Thr Tyr Pro Pro Pro Thr Pro Phe Pro Thr Ala Pro Glu Ile Thr Glu
                290                295                300
Asp Cys Gln Val Trp Asp Glu Val Val Met Ala Ala Tyr Asp Ile Asn
305                310                315                320
Pro Cys Phe Asn Tyr Tyr His Leu Ile Asp Phe Cys Pro Tyr Leu Trp
                325                330                335
Asp Val Leu Gly Phe Pro Ser Leu Gly Phe Gly Pro Asp Asn Tyr Phe
                340                345                350
Asn Arg Ser Asp Val Gln Lys Ile Leu His Val Pro Pro Thr Asp Tyr
                355                360                365
Ser Val Cys Ser Glu Thr Val Ile Phe Ala Asn Gly Asp Gly Ser Asp
                370                375                380
Pro Ser Ser Trp Gly Pro Leu Pro Ser Val Ile Glu Arg Thr Asn Asn
385                390                395                400
Thr Ile Ile Gly His Gly Trp Leu Asp Tyr Leu Leu Phe Leu Asn Gly
                405                410                415
Ser Leu Ala Thr Ile Gln Asn Met Thr Trp Asn Gly Lys Gln Gly Phe
                420                425                430
Gln Ser Pro Pro Val Glu Pro Leu Phe Val Pro Tyr His Tyr Gly Leu
                435                440                445
Ala Glu Leu Tyr Trp Gly Asp Glu Pro Asp Pro Tyr Asn Leu Asp Ala
                450                455                460
Gly Ala Gly Tyr Leu Gly Thr Ala His Thr Glu Arg Gly Leu Thr Phe
465                470                475                480
Ser Ser Val Tyr Leu Ser Gly His Glu Ile Pro Gln Tyr Val Pro Gly
                485                490                495
Ala Leu Thr Ala Ser Trp Ser Ser Cys Leu Val Glu Leu Ile Val Phe
                500                505                510
Pro Arg Arg Gly Thr Thr Pro Leu Asn Phe Ser
                515                520

```

(2) SEQ ID NO:12的信息:

(i) 序列特征:

(A) 长度: 530个氨基酸

(B) 类型: 氨基酸

(C) 链型: 单链

(D) 拓扑结构: 线型

(xi) 序列描述: SEQ ID NO:12:

```

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

```

Cys	Thr	Asp	Asn	Ala	Ser	Ala	Ala	Lys	His	Gly	Arg	Phe	Gly	Gln	Lys
			20					25					30		
Ala	Arg	Asp	Ala	Met	Asn	Ile	Ala	Asn	Gly	Ser	Ala	Asn	Ala	Val	Lys
		35					40					45			
His	Ser	Leu	Lys	Ile	Pro	Val	Glu	Asp	Tyr	Gln	Phe	Leu	Asn	Asn	Lys
	50					55					60				
Thr	Lys	Pro	Tyr	Arg	Val	Glu	Ser	Leu	Pro	Asp	Val	His	Phe	Asp	Leu
65					70					75					80
Gly	Glu	Met	Tyr	Ser	Gly	Leu	Val	Pro	Ile	Glu	Lys	Gly	Asn	Val	Ser
				85					90					95	
Arg	Ser	Leu	Phe	Phe	Val	Phe	Gln	Pro	Thr	Ile	Gly	Glu	Pro	Val	Asp
			100					105					110		
Glu	Thr	Thr	Ile	Trp	Leu	Asn	Gly	Gly	Pro	Gly	Cys	Ser	Ser	Leu	Glu
	115						120					125			
Ala	Leu	Ser	Pro	Gly	Glu	Cys	Arg	Phe	Val	Trp	Gln	Pro	Gly	Thr	Tyr
	130					135					140				
Gln	Pro	Val	Glu	Asn	Pro	Tyr	Ser	Trp	Val	Asn	Leu	Thr	Asn	Val	Leu
145					150					155					160
Trp	Val	Asp	Gln	Pro	Val	Gly	Thr	Gly	Phe	Ser	Leu	Gly	Val	Pro	Thr
			165						170					175	
Ala	Thr	Ser	Glu	Glu	Glu	Ile	Ala	Glu	Asp	Phe	Val	Lys	Phe	Phe	Lys
			180					185					190		
Asn	Trp	Gln	Gln	Ile	Phe	Gly	Ile	Lys	Asn	Phe	Lys	Ile	Tyr	Val	Thr
	195						200					205			
Gly	Glu	Ser	Tyr	Ala	Gly	Arg	Tyr	Val	Pro	Tyr	Ile	Ser	Ala	Ala	Phe
	210					215					220				
Leu	Asp	Gln	Asn	Asp	Thr	Glu	His	Phe	Asn	Leu	Lys	Gly	Ala	Leu	Ala
225					230					235					240
Tyr	Asp	Pro	Cys	Ile	Gly	Gln	Phe	Asp	Tyr	Val	Gln	Glu	Glu	Ala	Pro
			245						250					255	
Val	Val	Pro	Phe	Val	Gln	Lys	Asn	Asn	Ala	Leu	Phe	Asn	Phe	Asn	Ala
			260					265					270		
Ser	Phe	Leu	Ala	Glu	Leu	Glu	Ser	Ile	His	Glu	Gln	Cys	Gly	Tyr	Lys
	275						280					285			
Asp	Phe	Ile	Asp	Gln	Tyr	Leu	Val	Phe	Pro	Ala	Ser	Gly	Val	Gln	Pro
	290					295					300				
Pro	Lys	Ala	Met	Asn	Trp	Ser	Asp	Pro	Thr	Cys	Asp	Val	Tyr	Asp	Ile
305					310					315					320
Val	Asn	Asn	Ala	Val	Leu	Asp	Pro	Asn	Pro	Cys	Phe	Asn	Pro	Tyr	Glu
			325						330					335	
Ile	Asn	Glu	Met	Cys	Pro	Ile	Leu	Trp	Asp	Val	Leu	Gly	Phe	Pro	Thr
			340					345					350		
Glu	Val	Asp	Tyr	Leu	Pro	Ala	Ala	Pro	Ala	Ser	Thr	Leu	Thr	Ala	Leu
	355						360					365			
Ile	Lys	Arg	Ala	Met	His	Ala	Pro	Asn	Ile	Thr	Trp	Ser	Glu	Cys	Ser
	370					375					380				
Val	Glu	Ser	Val	Phe	Val	Gly	Gly	Asp	Gly	Gly	Pro	Glu	Gln	Glu	Gly
385					390					395					400
Asp	Tyr	Ser	Ala	Asn	Pro	Ile	Glu	His	Val	Leu	Pro	Gln	Val	Ile	Glu
			405						410					415	
Gly	Thr	Asn	Arg	Val	Leu	Ile	Gly	Asn	Gly	Asp	Tyr	Asp	Met	Val	Ile
		420						425					430		
Leu	Thr	Asn	Gly	Thr	Leu	Leu	Ser	Ile	Gln	Asn	Met	Thr	Trp	Asn	Gly
		435					440					445			
Lys	Leu	Gly	Phe	Asp	Thr	Ala	Pro	Ser	Thr	Pro	Ile	Asn	Ile	Asp	Ile
	450					455					460				
Pro	Asp	Leu	Met	Tyr	Asn	Glu	Val	Phe	Ile	Glu	Asn	Gly	Tyr	Asp	Pro
465					470					475					480
Gln	Gly	Gly	Gln	Gly	Val	Met	Gly	Ile	Gln	His	Tyr	Glu	Arg	Gly	Leu
			485						490					495	
Met	Trp	Ala	Glu	Thr	Phe	Gln	Ser	Gly	His	Met	Gln	Pro	Gln	Phe	Gln
		500						505					510		
Pro	Arg	Val	Ser	Tyr	Arg	His	Leu	Glu	Trp	Leu	Leu	Gly	Arg	Arg	Asp
		515					520					525			
Thr	Leu														
	530														

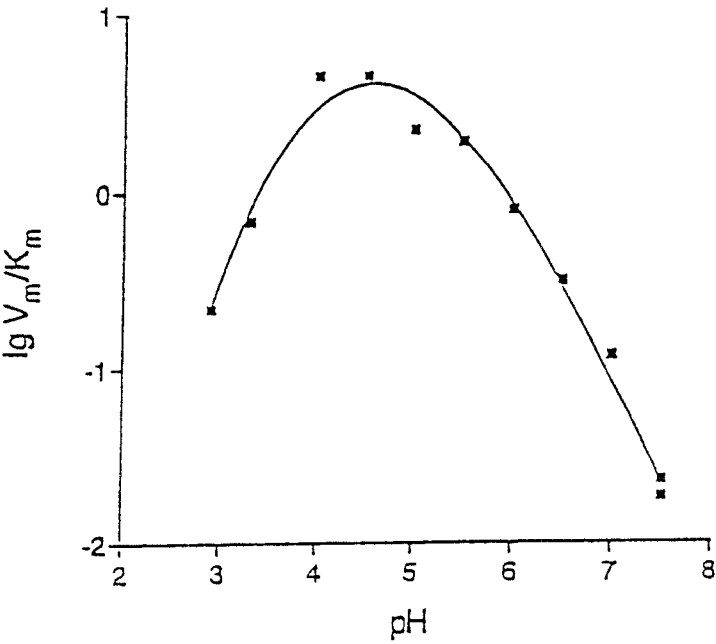


图 1

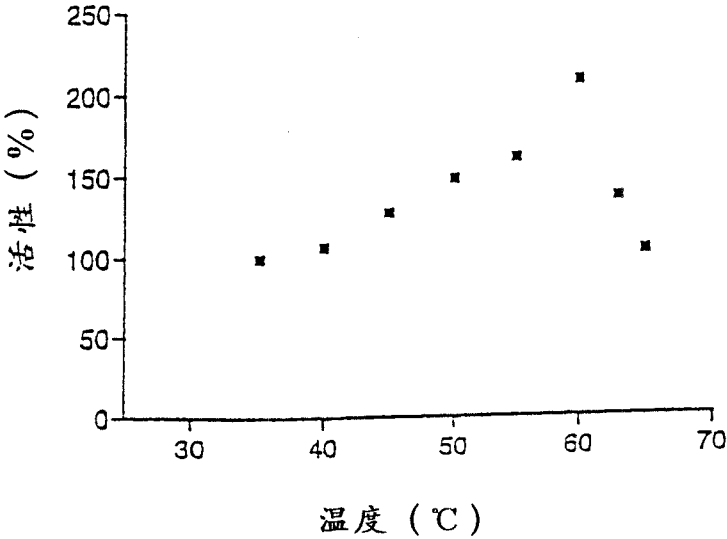


图 2

ATGGGTGGTACTGATTTCTCTCAGTCTACCGTTGGTTCAGCGAGTTGCGCCCTTCAGCGAGTACAGCGAGTGGTGGT
 MRGYEPLSVLPVAAASWALPGSTPASVGRRL
 ACCGAGAACCCGACCGCGTCAAGCTCTTACAGCGCGAGCAATGTGCGATCCGTTACAGCGAACCGCGCGAGCGCGT
 PKNPTGVKTLTTANNVTIRYKEPGAEGVCE
 CCGCGGTGTCAATCTACTCTCGATATGTCCAGCTCTCCGAGTCCCATACCTTCTTCTGTTCTTCGAGCGCGACATA
 TPGVKSYSGYVOTSPESHTEFWFFEARHNPE
 CCACCTATCATTGTGTTGAATGGTCCCGCTCGAGCGATTCCTTTCATCGGTCTCTTCGAGCGAGTTGCGCCCTTCGATGTC
 APITLWLNCGPGSOSLIGLPSELGPGCHVNSTF
 TGATGACTACATCAACCGTCACTGTCCAGCGAGTCTCAATTTACTATTCTGTCCGAGCGAGTTGCGAGTCCCGTTTCATA
 DDYINPHSWNEVSNLLPLSQPLGVGFSYSOT
 TTGATCGGTCCATTACCGTGTAACTCGCGTGTCCAAATTCAGCTTTCAGCGAGTTCAGCGCGGTACCGAGCGATTGATG
 VOGSINPVTGVVENSSPAGVQGRYPPTIDATLI
 GATACTACCAATCTTCCCGAGCGCGCGTTCGAGAGTCTCCAGCGATTCCTTACTCGAGTACCTACCTTCGAGTCTACCGT
 DTTNLAABAAWEILGGLSLGLPSLDSPVQSKD
 CTTCAGTCTATCGAGCGAGCTATCCAGCGCGCTATCGTCTCTCAATTTTCATCATTTTACGAGCGAGTACGAGATTCC
 FSLWTESYGGHYGPAFFNHFFYEONERIANGS
 TTAACTGTGTCACTTAATTTCACTCTCTCCGATTAATACCGCATCTCCAGCGAGCGCGATCCAGCGCGCTTACTACCGT
 VNGVQLNPNLSLGIINGIIDEAIQAPYYPEFAV
 AACATACCTACCGTATCAACCGTGTCAAGCGAGCGTCTACAGTACATGAAGTTTCCAGCGCAATCCGAAATGTTCCGAG
 NNTYGIKAVNETVYNYMKFANQMPNGCQDLIS
 CAGCTCCAGAGAGCAACCGCGCGCGATTACCTCACTACCGCTCTCCCGCGAGCGCGAGCATGTCCAGCGAGATGTTGA
 TCKQTNR TALADYALCAEATNMCRDNVEGPY
 ACCGCTTTCGTCGTGTGTGTATGATATTCGCGATCCATATGATGACCGGAGTCCCGCGAGTTATTACAGCAATTTCTCG
 YAFAGRGVYDIRHPYDOPTPPSYXKFLAKDS
 GTGATCGAGCTATCCCGGTCAACATCACTACAGCGAGTCAATTAAGCTCTACTACCGTTTCAGCGAGCGAGCTTT
 VMDAIGVNNINYTQSNNOVYYAFQQTGDFVWPN
 CTTCATCGAGAGCTCCAGCGAGTCTCTCTCCCGGTGTCTCCCTCATCTATCCGAGCGCGGATTACATCTCCAGT
 FIEDLEILALPVRVSLIYGDAOYICNWPGG
 ACGCGTTTCCTCTCTCCGAGTACTCCGAGCGCGCGAGTTCCGAGCGCGAGTACAGCGCGCTGAAAGTCAACCGCGTGG
 QAVSLAANYSQAAQFRSAGYTPLKXVNGVEYGE
 ACTCCGAGTATGTAATTTCTCTTCACTCCGTCTATGAGCGCGCGAGTCCGATACTACGAGCGCGATCCGCTCTCTG
 TREYGNFSPFTRVYEAGHEVPYYQPIASLQLFN
 CCGGACTATCTTCGTTCCGATATCCAGCGCGCGAGAGCATCTCCCGAGTACAGAGCGAATCCGAGCGCTACAGCTAC
 RTIFGWDIABGQKKIWPSYKTNGTATATHTQ
 CGTCCGTCCCGCTCCCTACCGCTACGAGCATGTCAGTGTTCGATG 1648
 SSVPLPTATSMSSVGMAL
 ..

图 3

